

# ATR analysis

TFM

Montserrat Domingo Ayllón

4 de junio, 2024

## Índice

<b>1 Variables Demográficas y clínicas</b>	<b>2</b>
1.1 Cálculo de medias y desviaciones estándar por subgrupos para variables demográficas y clínicas . . . . .	2
1.2 Cálculo de medias y desviaciones estándar global en <i>gene carriers</i> para variables demográficas y clínicas . . . . .	2
1.3 Cálculo de medias y desviaciones estándar en preHD y HD para PBA-Enroll, Stroop-interferences y TMT(B-A) Direct . . . . .	2
1.4 Cálculo de medias y desviaciones estándar global para GC para PBA-Enroll, Stroop-interferences y TMT(B-A) Direct . . . . .	2
1.5 T-test para variables numéricas CT-GC, CT-preHD y CT-HD . . . . .	3
1.6 T-test para variables numéricas pHD-HD . . . . .	4
1.7 Wilcoxon Mann Whitney test para variables ordinales . . . . .	5
1.8 Chi-square/Fisher test para variable nominal (gender) . . . . .	5
<b>2 ATR Average: RATR y LATR por separado</b>	<b>7</b>
2.1 Creación de DATAFRAMES . . . . .	7
2.2 Automated Exploratory Data Analysis . . . . .	9
2.3 Estadística descriptiva gráfica . . . . .	20
2.3.1 BARPLOTS . . . . .	20
2.3.1.1 IRON . . . . .	20
2.3.1.2 AD . . . . .	21
2.3.1.3 MD . . . . .	22
2.3.1.4 RD . . . . .	23
2.3.1.5 FA . . . . .	24
2.3.2 HISTOGRAMAS y QQ-PLOTS . . . . .	25
2.3.2.1 IRON - RATR . . . . .	25
2.3.2.2 IRON - LATR . . . . .	29
2.3.2.3 AD - RATR . . . . .	32
2.3.2.4 AD - LATR . . . . .	35

2.3.2.5	MD - RATR . . . . .	38
2.3.2.6	MD - LATR . . . . .	41
2.3.2.7	RD - RATR . . . . .	44
2.3.2.8	RD - LATR . . . . .	47
2.3.2.9	FA - RATR . . . . .	50
2.3.2.10	FA - LATR . . . . .	53
2.3.3	PLOT de CORRELACIONES . . . . .	56
2.4	Estadística descriptiva matemática . . . . .	57
2.4.1	Media, mediana, desviación estándar, kurtosis y skewness . . . . .	57
2.4.2	Normalidad Univariante - Homocedasticidad de cada métrica para cada factor . . . . .	58
2.4.3	Normalidad multivariante . . . . .	60
2.4.4	Homocedasticidad de la matriz de varianza-covarianza . . . . .	75
<b>3</b>	<b>ANÁLISIS del average: RATR y LATR por separado</b> . . . . .	<b>77</b>
3.1	MANOVA . . . . .	77
3.1.1	RATR . . . . .	78
3.1.2	LATR . . . . .	85
3.1.3	Comparaciones pareadas . . . . .	91
3.2	PERMANOVA . . . . .	114
3.2.1	RATR . . . . .	116
3.2.2	LATR . . . . .	145
3.2.3	Comparaciones por pares . . . . .	170
3.3	ANOVA . . . . .	177
3.3.1	IRON - RATR . . . . .	177
3.3.2	IRON - LATR . . . . .	180
3.3.3	AD - RATR . . . . .	184
3.3.4	AD - LATR . . . . .	187
3.3.5	MD - RATR . . . . .	190
3.3.6	MD - LATR . . . . .	193
3.3.7	RD - RATR . . . . .	196
3.3.8	RD - LATR . . . . .	199
3.3.9	FA - RATR . . . . .	202
3.3.10	FA - LATR . . . . .	205
3.3.11	Eliminación de outliers - Nuevos ANOVAS . . . . .	208
3.3.12	Comparaciones pareadas por subgrupos . . . . .	215

<b>4 ATR Average: ambos ATR conjuntamente</b>	<b>217</b>
4.1 Creación de DATAFRAMES . . . . .	217
4.2 Automated Exploratory Data Analysis . . . . .	218
4.3 Estadística descriptiva gráfica . . . . .	227
4.3.1 BARPLOTS . . . . .	227
4.3.1.1 IRON . . . . .	227
4.3.1.2 AD . . . . .	228
4.3.1.3 MD . . . . .	229
4.3.1.4 RD . . . . .	230
4.3.1.5 FA . . . . .	231
4.3.2 HISTOGRAMAS y QQ-PLOTS . . . . .	231
4.3.2.1 IRON . . . . .	232
4.3.2.2 AD . . . . .	235
4.3.2.3 MD . . . . .	238
4.3.2.4 RD . . . . .	241
4.3.2.5 FA . . . . .	244
4.3.3 PLOT de CORRELACIONES . . . . .	247
4.4 Estadística descriptiva matemática . . . . .	248
4.4.1 Media, mediana, desviación estándar, kurtosis y skewness . . . . .	248
4.4.2 Normalidad Univariante - Homocedasticidad de cada métrica para cada factor . . . . .	249
4.4.3 Normalidad multivariante . . . . .	250
4.4.4 Homocedasticidad de la matriz de varianza-covarianza . . . . .	258
<b>5 ANÁLISIS del average: ambos ATR conjuntamente</b>	<b>260</b>
5.1 MANOVA . . . . .	260
5.1.1 5 Métricas . . . . .	261
5.1.2 3 Métricas de difusividad . . . . .	263
5.1.3 2 Métricas Iron-FA . . . . .	267
5.1.4 Comparaciones pareadas . . . . .	269
5.2 PERMANOVA . . . . .	282
5.2.1 Matriz de distancias . . . . .	282
5.2.2 Modelos . . . . .	300
5.2.3 Valoración de la dispersión . . . . .	307
5.2.4 Comparaciones por pares . . . . .	317
5.3 ANOVA . . . . .	321
5.3.1 IRON . . . . .	322
5.3.2 AD . . . . .	326
5.3.3 MD . . . . .	328
5.3.4 RD . . . . .	333

5.3.5	FA . . . . .	337
5.3.6	Comparaciones pareadas por subgrupos . . . . .	340
5.4	KRUSKAL WALLIS TEST . . . . .	340
5.5	ANOVA PERMUTACIONAL . . . . .	342
5.5.1	IRON . . . . .	344
5.5.2	AD . . . . .	345
5.5.3	MD . . . . .	347
5.5.4	RD . . . . .	348
5.5.5	FA . . . . .	349
5.5.6	Comparaciones pareadas por subgrupos - Pairwise permutation test . . . . .	350
<b>6</b>	<b>Regresión logística</b>	<b>353</b>
6.1	ONE vs REST . . . . .	353
6.1.1	RATR . . . . .	353
6.1.2	LATR . . . . .	362
6.1.3	Ambos ATR . . . . .	370
6.2	ONE vs ONE . . . . .	379
6.2.1	RATR . . . . .	379
6.2.2	LATR . . . . .	384
6.2.3	Both ATR . . . . .	390
<b>7</b>	<b>Correlaciones clínicas</b>	<b>397</b>
7.1	Correlaciones parciales . . . . .	397
7.1.1	CAP controlado por género . . . . .	397
7.1.2	Age_onset controlado por género . . . . .	397
7.1.3	UHDRS functional score controlado por género y CAP . . . . .	397
7.1.4	UHDRS functional score controlado por género . . . . .	398
7.1.5	UHDRS cognitive score controlado por género y CAP . . . . .	398
7.1.6	UHDRS cognitive score controlado por género . . . . .	398
7.1.7	UHDRS motor score controlado por género y CAP . . . . .	398
7.1.8	UHDRS motor score controlado por género . . . . .	399
7.1.9	Stroop_interferences controlado por género y CAP . . . . .	399
7.1.10	Stroop_interferences controlado por género . . . . .	399
7.1.11	Direct TMT(B-A) controlado por género y CAP . . . . .	399
7.1.12	Direct TMT(B-A) controlado por género . . . . .	400
7.1.13	PBA-ENROLL-Depression controlado por género y CAP . . . . .	400
7.1.14	PBA-ENROLL-Depression controlado por género . . . . .	400
7.1.15	PBA-ENROLL-Irritability controlado por género y CAP . . . . .	400
7.1.16	PBA-ENROLL-Irritability controlado por género . . . . .	401

7.1.17 PBA-ENROLL-Psychosis controlado por género y CAP . . . . .	401
7.1.18 PBA-ENROLL-Psychosis controlado por género . . . . .	401
7.1.19 PBA-ENROLL-Apathy controlado por género y CAP . . . . .	401
7.1.20 PBA-ENROLL-Apathy controlado por género . . . . .	402
7.1.21 PBA-ENROLL-Dysexecutive controlado por género y CAP . . . . .	402
7.1.22 PBA-ENROLL-Dysexecutive controlado por género . . . . .	402

# 1 Variables Demográficas y clínicas

## 1.1 Cálculo de medias y desviaciones estándar por subgrupos para variables demográficas y clínicas

Cuadro 1: Edad y Años de escolarización (mean +/- sd)

Subgrupo	Edad	Años educativos
Control	46.42+/-11.14	12.54+/-2.89
PreHD	40.62+/-9.73	13.54+/-2.67
HD	51+/-10.23	11.33+/-2.87

Cuadro 2: Variables clínicas (mean +/- sd) en PreHD y HD

Subgrupo	CAG	CAP	Cognitive Reserve	Age of Onset	UHDRS-motor	UHDRS-cognitive	UHDRS-functional
PreHD	42.77+/-2.17	81.3+/-16.97	15.54+/-3.78	50.17+/-8.15	0.38+/-0.87	328+/-46.25	24.92+/-0.28
HD	44.17+/-3.13	112.13+/-18.86	11.56+/-3.71	45.98+/-9.59	22.94+/-13.78	192+/-62.22	19.18+/-5.43

## 1.2 Cálculo de medias y desviaciones estándar global en gene carriers para variables demográficas y clínicas

Cuadro 3: Variables demográficas y clínicas (mean +/- sd) en gene carriers

Edad	Educación	CAG	CAP	Age of Onset	Cognitive Reserve	UHDRS-motor	UHDRS-cognitive	UHDRS-functional
46.65+/-11.15	12.26+/-2.95	43.58+/-2.81	99.2+/-23.58	47.74+/-9.12	13.23+/-4.18	13.17+/-15.31	255.14+/-87.91	21.67+/-4.97

## 1.3 Cálculo de medias y desviaciones estándar en preHD y HD para PBA-Enroll, Stroop-interferences y TMT(B-A) Direct

Cuadro 4: Variables psiquiátricas PBA-Enroll (mean +/- sd) en PreHD y HD

Subgrupo	Depression	Irritability	Psychosis	Apathy	Dysexecutive
PreHD	1.58+/-3	0.67+/-1.78	0.17+/-0.58	1.42+/-2.71	1.42+/-2.02
HD	2.76+/-3.35	2.06+/-4.39	0.31+/-1.01	3+/-3.32	2.94+/-2.73

Cuadro 5: Variables de función ejecutiva (mean +/- sd) en PreHD y HD

Subgrupo	Stroop	TMTdirect
PreHD	49.08+/-8.57	36.23+/-36.53
HD	28.8+/-11.93	105.88+/-63.5

## 1.4 Cálculo de medias y desviaciones estándar global para GC para PBA-Enroll, Stroop-interferences y TMT(B-A) Direct

Cuadro 6: Variables psiquiátricas PBA-Enroll (mean +/- sd) en gene carriers

Depression	Irritability	Psychosis	Apathy	Dysexecutive
2.28+/-3.21	1.48+/-3.57	0.25+/-0.84	2.34+/-3.13	2.31+/-2.54

Cuadro 7: Variables de función ejecutiva (mean +/- sd) en gene carriers

Stroop	TMTdirect
38.21 +/-14.58	75.7 +/-63.32

## 1.5 T-test para variables numéricas CT-GC, CT-preHD y CT-HD

Se genera dos t-tests para comprobar si existen **diferencias entre controles y gene carriers en relación a la edad (en años) y a los años de escolarización**. Consulta del siguiente recurso web: ANOVA in R.

Shapiro-Wilk normality test

```
data: demo_attr$`Age(years)`
W = 0.96119, p-value = 0.07323

Shapiro-Wilk normality test

data: demo_attr$Years_education
W = 0.88039, p-value = 5.533e-05
```

Cuadro 8: T-test controles vs gene carriers

Variable	T-Statistic	df	p.value	estimate_CT	estimate_GC
Years_Education	0.3574749	50.13185	0.7222369	12.54167	12.25806
Age	-0.07542425	49.61952	0.9401804	46.41667	46.64516

Como se observa en la tabla, no existen diferencias significativas entre controles y gene carriers en cuanto a la edad (en años) y a los años de escolarización.

Se genera dos anovas para comprobar si existen **diferencias entre los 3 subgrupos en relación a la edad (en años) y a los años de escolarización**; posteriormente, se ejecutan los **post-hoc tests**. En este contexto, se pueden utilizar dos métodos:

- La función `anova_test` del paquete `rstatix` y posteriormente ejecutar los t-tests por pares;
- La función `aov` del paquete `stats` y posteriormente ejecutar los post-hoc Tukey tests ya sea con la función `HSD.test` del paquete `agricolae` o la función `TukeyHSD` del paquete `stats`.

Se han generado todas las posibilidades. A continuación se muestran los resultados de la combinación `anova_test-t.test`.

Cuadro 9: ANOVA para edad y años de escolarización con la función `anova_test`

Variables	Effect	DFn	DFd	F	p	p<.05	ges
3G - Age	Subgroup	2	52	3.673	0.032	*	0.124
3G - Years of education	Subgroup	2	52	2.355	0.105		0.083

Cuadro 10: T-test para edad y años de escolarización con la función `t.test`

Vars	statistic	df	p-value
Age CT vs PHD	1.64439805036553	27.7909825451394	0.111360485247531

CT vs HD	-1.38315695316033	38.2995156869118	0.17462742221389
PHD vs HD	-2.87062948214092	26.7683613333542	0.00790796297882448
<b>Years of Education</b>			
CT vs PHD	-1.05417408935518	26.5298169793182	0.301309799063886
CT vs HD	1.34653243902194	36.9051310615085	0.186342996029099
PHD vs HD	2.20087645004456	27.0988113604117	0.0364452873286952

Con este método, el ANOVA muestra diferencias significativas en la edad para la variable subgrupo, pero no así para los años de escolarización. Respecto al t-test, las diferencias de edad se objetivan entre preHD y HD en relación a una edad mayor en los HD respecto a los preHD. A pesar de no ser significativo en el ANOVA, también se ha generado el t-test de los años de escolarización por el hecho de ser exhaustivo y éste muestra diferencias en los años de educación entre preHD y HD (tal vez por ser los HD mayores pueden presentar menor tasa de estudios medios y/o universitarios). De todas formas, este resultado de años de escolarización no se tendrá en cuenta dado que el ANOVA no ha sido significativo.

A continuación se muestran los resultados de la combinación `aov-HSDtest/TukeyHSD`.

Cuadro 11: ANOVA para edad y años de escolarización con la función `aov`

name	Df	Sum Sq	Mean Sq	F value	Pr(>F)
<b>Age</b>					
Subgroup	2	814.73	407.36	3.67	0.03
Residuals	52	5766.91	110.90	NA	NA
<b>Years of Education</b>					
Subgroup	2	37.79	18.90	2.36	0.10
Residuals	52	417.19	8.02	NA	NA

Cuadro 12: HSD Tukey test para edad y años de escolarización con las funciones `HSD.test` y `TukeyHSD`

	Media	Grupos	Comparación	Diferencia de medias	p-valor ajustado
<b>Age</b>					
HD	51.00	a	PreHD vs Controles (2-1)	-5.801	0.255
Control	46.42	ab	HD vs Controles (3-1)	4.583	0.350
preHD	40.62	b	HD vs PreHD (3-2)	10.385	0.024
<b>Years of Education</b>					
preHD	13.54	a	PreHD vs Controles (2-1)	0.997	0.567
Control	12.54	a	HD vs Controles (3-1)	-1.208	0.365
HD	11.33	a	HD vs PreHD (3-2)	-2.205	0.092

Al igual que con el método anterior, el ANOVA muestra diferencias significativas en la edad para la variable subgrupo, pero no así para los años de escolarización. Respecto a los post-hoc Tukey tests, las diferencias de edad también se objetivan entre preHD y HD en relación a una edad mayor en los HD respecto a los preHD, y, a diferencia de los t-tests anteriores, no muestran diferencias entre grupos en los años de educación.

## 1.6 T-test para variables numéricas pHD-HD

Se generan t-tests para cada variable numérica para determinar la existencia o la ausencia de **diferencias entre preHD y HD en relación a edad (en años), años de escolarización, edad de inicio de la enfermedad, CAG, CAP, reserva cognitiva, UHDRS funcional, UHDRS motor, UHDRS cognitivo, Stroop-interference y TMT-direct**.

Cuadro 13: T-tests preHD vs HD para diferentes variables numéricas

Variables	Statistic	df	p.value	estimate_pHD	estimate_HD
Age	-2.870629	26.76836	0.007907963	40.61538	51

Years of Education	2.200876	27.09881	0.03644529	13.53846	11.33333
CAG	-1.4689	28.97318	0.1526349	42.76923	44.16667
CAP	-4.760798	27.50418	5.553655e-05	81.30291	112.1301
Age of onset	1.310268	28.1371	0.2007011	50.17291	45.98436
UHDRS-motor	-6.730824	16.1665	4.574068e-06	0.3846154	22.94118
UHDRS-functional	4.352813	16.10894	0.0004861086	24.92308	19.17647
UHDRS-cognitive	6.615483	25.46697	5.688455e-07	328	192
Cognitive reserve	2.917436	25.74475	0.007227162	15.53846	11.55556
Stroop interference	5.212153	25.20698	2.103128e-05	49.07692	28.8
TMTdirect	-3.77815	26.28114	0.0008204783	36.23077	105.8824

Existen diferencias entre pHD y HD en las siguientes variables:

- Edad (los HD son más mayores);
- Años de educación (más años en pHD);
- CAP (mayor en HD debido a que tienen más edad);
- UHDRS-motor score (mayor en HD);
- Reserva cognitiva (mayor en pHD);
- UHDRS-functional score (mayor en pHD);
- UHDRS-cognitive score (mayor en pHD);
- Stroop-interferences (mayor en pHD);
- TMT direct (mayor en HD).

No existen diferencias entre pHD y HD en las siguientes variables: age-of-onset, CAG.

## 1.7 Wilcoxon Mann Whitney test para variables ordinales

Se generan **tests de Wilcoxon** para determinar **diferencias entre preHD y HD en las variables psiquiátricas PBA-Enroll** (ordinales).

Cuadro 14: Wilcoxon Tests entre preHD y HD en Variables psiquiátricas PBA-Enroll

Variables	statistic	p-value
Depression	79	0.2737173
Irritability	78	0.2037743
Psychosis	92	0.7622456
Apathy	60.5	0.05780502
Dysexecutive	65	0.08831126

No existen diferencias entre pHD y HD en relación a los valores de PBA-ENROLL, aunque cabe mencionar que tanto apatía como disexecutivo presentan p-valores menores a 0.10.

## 1.8 Chi-square/Fisher test para variable nominal (gender)

El **test Chi-square** se utiliza para determinar la **diferencia entre controles y gene carriers** mediante una de las dos funciones siguientes: **chisq\_test** del paquete **rstatix** o **chisq.test** del paquete **stats**.

Para determinar la **diferencia entre los 3 subgrupos** o las **diferencias pareadas entre subgrupos**, se pueden utilizar el **test Chi-square** con una de las dos funciones especificadas anteriormente o el **test de Fisher** con la función **fisher.test** del paquete **stats**.

Dado que el subgrupo de preHD (presintomáticos) sólo tiene un varón, R especifica que la aproximación de Chi-square puede ser incorrecta en las comparaciones de los 3 subgrupos a la vez y en las comparaciones controles-PreHD y PreHD-HD.

Cuadro 15: Número de hombres y mujeres por subgrupo

	Hombres	Mujeres
Controles	13	11
PreHD	1	12
HD	7	11

Cuadro 16: Chi-square test para género

Variables	statistic	df	p-value
Controls vs Gene Carriers	3.486291	1	0.06187906
3 Subgroups	7.721927	2	0.02104771
Controls vs preHD	5.893487	1	0.01519698
Controls vs HD	0.4474432	1	0.5035517
PreHD vs HD	2.380475	1	0.1228602

Cuadro 17: Fisher test para género

Variables	p-value
3 Subgroups	0.01753407
Controls vs preHD	0.01113493
Controls vs HD	0.3664521
PreHD vs HD	0.09535039

No existen diferencias de género entre controles y *gene carriers*, pero sí existen diferencias de género entre los tres subgrupos.

Al realizar la comparativa dos a dos, se evidencia que no existen diferencias de género entre preHD y HD y entre controles y HD, pero sí existen diferencias de género entre controles y preHD.

## 2 ATR Average: RATR y LATR por separado

### 2.1 Creación de DATAFRAMES

Se cargan los datos.

Se recodifica la variable *age* a factor mediante la información encontrada en la web siguiente Etiquetar y recodificar variables en R.

Se han generado dos edades factorizadas, una con 7 niveles (división de 5 en 5 años) y otra con 3 niveles (división de 10 en 10 años); la razón de ello es debido a la ausencia de sujetos en la combinación de algunos de los niveles de edad procedentes de la edad factorizada con 7 niveles con las variables de género y/o subgrupo, hecho que impide la comparación estadística por pares (ver apartado ANOVA permutacional - pairwise permutational test en *along-the-tract*).

```
# Lectura de los datos
library("readxl")

# Valores medios de cada una de las métricas de RM en cada sujeto
iron_avg <- read_excel("D:/Masters/MasterBioInformaticaBioEstadistica/TFM/Scripts_ATR/Tables_allpoints/mean")
diff_avg <- read_excel("D:/Masters/MasterBioInformaticaBioEstadistica/TFM/Scripts_ATR/Tables_allpoints/mean")
ATR_avg<-cbind(iron_avg[,c(1,2,13,5)], diff_avg[,c(43:46,11:14)])

# Datos demográficos: código, grupo, subgrupo, género y edad
demographics <- read_excel('D:/Masters/MasterBioInformaticaBioEstadistica/TFM/INFO_CLÍNICA/demos_bis.xlsx')
demographics$Group = factor(demographics$Group,
                             levels=c('1','2'), labels=c('Control', 'Carriers'))
demographics$Subgroup = factor(demographics$Subgroup,
                                levels=c('1','2','3'),
                                labels=c('Control', 'PreHD', 'HD'))
demographics$Gender = factor(demographics$Gender,
                             levels=c('1','2'), labels=c('Male', 'Female'))

#Factorización de la edad en 7 niveles (de 5 en 5 años)
agef<-c()
agef[demographics$`Age(years)`>60]<-7 # 8 individuos
agef[demographics$`Age(years)`>55 & demographics$`Age(years)`<=60]<-6 # 3 individuos
agef[demographics$`Age(years)`>50 & demographics$`Age(years)`<=55]<-5 # 4 individuos
agef[demographics$`Age(years)`>45 & demographics$`Age(years)`<=50]<-4 # 15 individuos
agef[demographics$`Age(years)`>40 & demographics$`Age(years)`<=45]<-3 # 7 individuos
agef[demographics$`Age(years)`>35 & demographics$`Age(years)`<=40]<-2 # 9 individuos
agef[demographics$`Age(years)`<=35]<-1 # 9 individuos
agef<-as.factor(agef)

#Factorización de la edad en 3 niveles (de 10 en 10 años)
agef2<-c()
agef2[demographics$`Age(years)`>50]<-3 # 15 individuos
agef2[demographics$`Age(years)`>40 & demographics$`Age(years)`<=50]<-2 # 22 individuos
agef2[demographics$`Age(years)`<=40]<-1 # 18 individuos
agef2<-as.factor(agef2)

# Generación del dataframe
ATR_avg_demo<-cbind(demographics, agef, agef2, ATR_avg[,-c(1,2)]) 

# Summary statistics del dataframe
sumATR_avg_demo<-summary(ATR_avg_demo[,-c(1,5,6)])
kable(list(sumATR_avg_demo[,1:4], sumATR_avg_demo[,c(5,7:10)],sumATR_avg_demo[,c(6,11:14)]),
```

Cuadro 18: Summary statistics

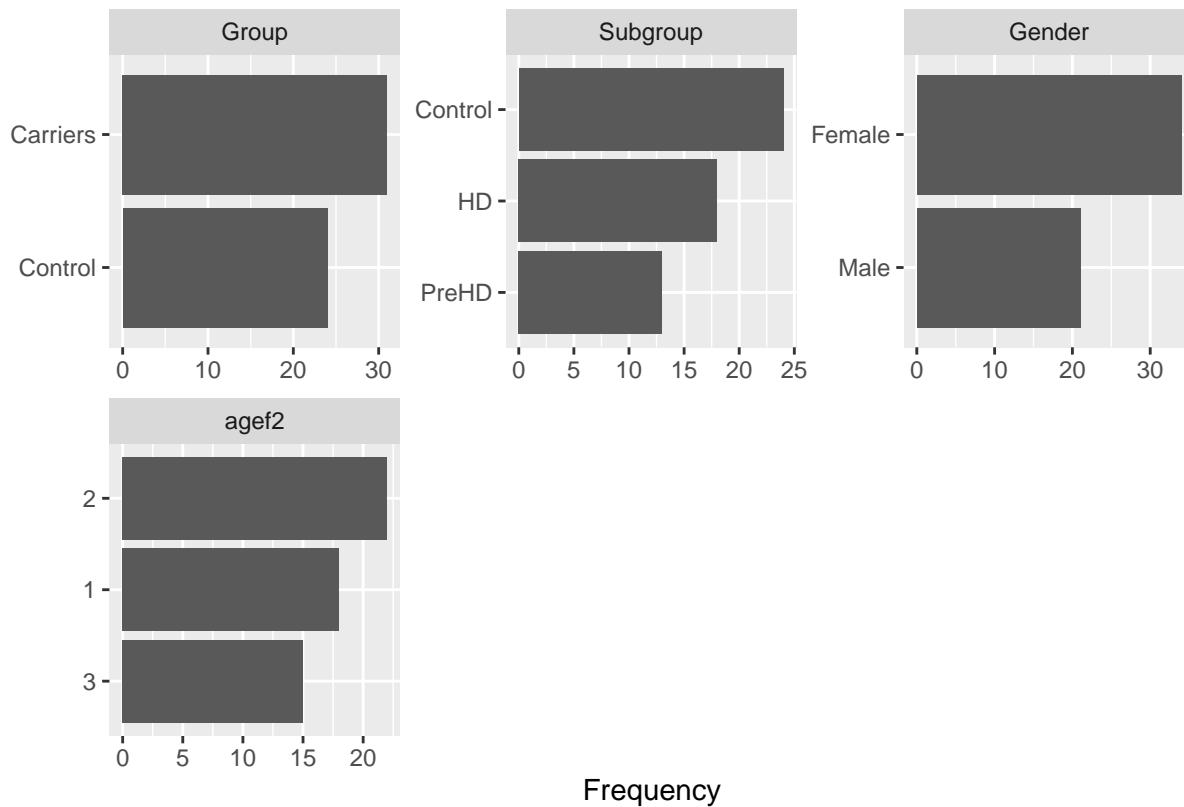
Group	Subgroup	Gender	agef2	
Control :24	Control:24	Male :21	1:18	
Carriers:31	PreHD :13	Female:34	2:22	
NA	HD :18	NA	3:15	
NA	NA	NA	NA	
NA	NA	NA	NA	
NA	NA	NA	NA	
rhatr_Iron	rhatr_AD	rhatr_RD	rhatr_MD	
Min. :0.02014	Min. :0.0008239	Min. :0.0004432	Min. :0.0005788	Min. :0.3170
1st Qu.:0.02272	1st Qu.:0.0008899	1st Qu.:0.0004883	1st Qu.:0.0006251	1st Qu.:0.3646
Median :0.02354	Median :0.0009164	Median :0.0005108	Median :0.0006406	Median :0.3784
Mean :0.02359	Mean :0.0009186	Mean :0.0005092	Mean :0.0006457	Mean :0.3802
3rd Qu.:0.02445	3rd Qu.:0.0009368	3rd Qu.:0.0005292	3rd Qu.:0.0006628	3rd Qu.:0.3963
Max. :0.02996	Max. :0.0010076	Max. :0.0006190	Max. :0.0007424	Max. :0.4443
lhatr_Iron	lhatr_AD	lhatr_RD	lhatr_MD	
Min. :0.01911	Min. :0.0008042	Min. :0.0004411	Min. :0.0005621	Min. :0.3170
1st Qu.:0.02339	1st Qu.:0.0008752	1st Qu.:0.0004791	1st Qu.:0.0006117	1st Qu.:0.3695
Median :0.02458	Median :0.0009017	Median :0.0004991	Median :0.0006281	Median :0.3850
Mean :0.02454	Mean :0.0008997	Mean :0.0004978	Mean :0.0006318	Mean :0.3816
3rd Qu.:0.02572	3rd Qu.:0.0009222	3rd Qu.:0.0005173	3rd Qu.:0.0006485	3rd Qu.:0.3991
Max. :0.02909	Max. :0.0009925	Max. :0.0005766	Max. :0.0007040	Max. :0.4224

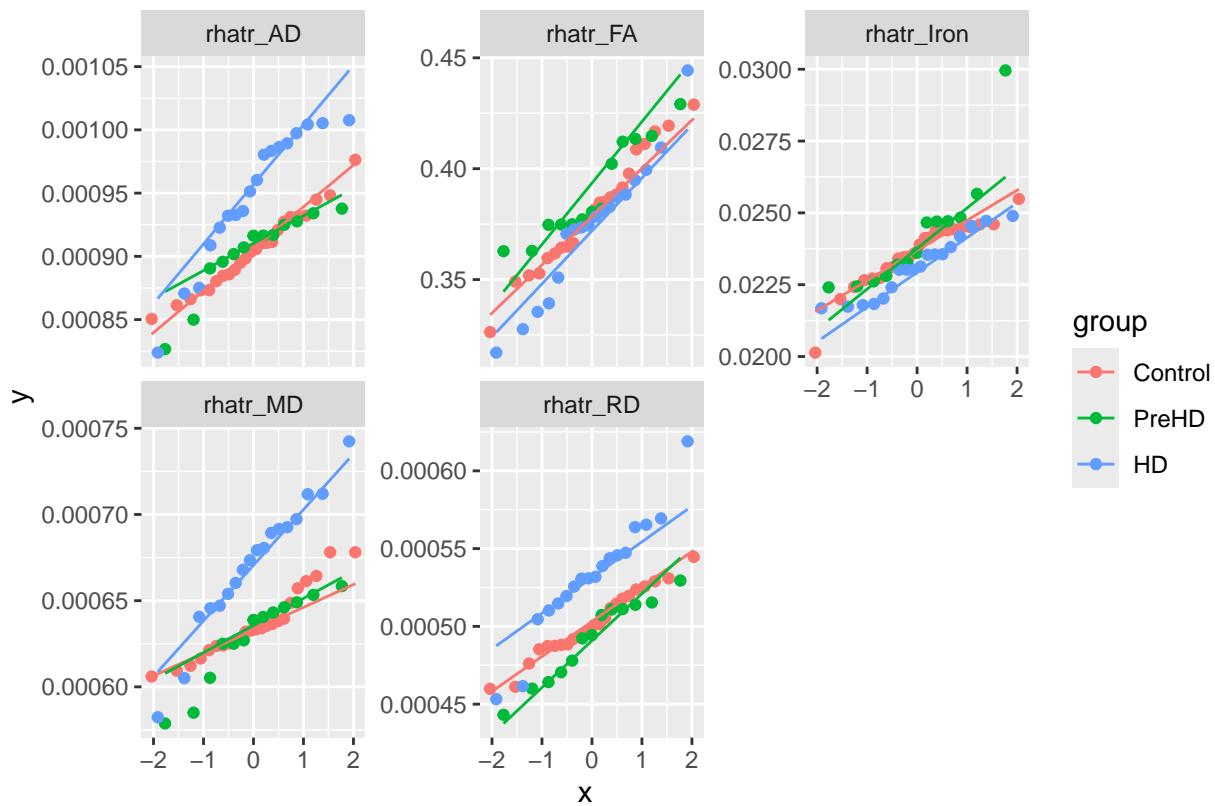
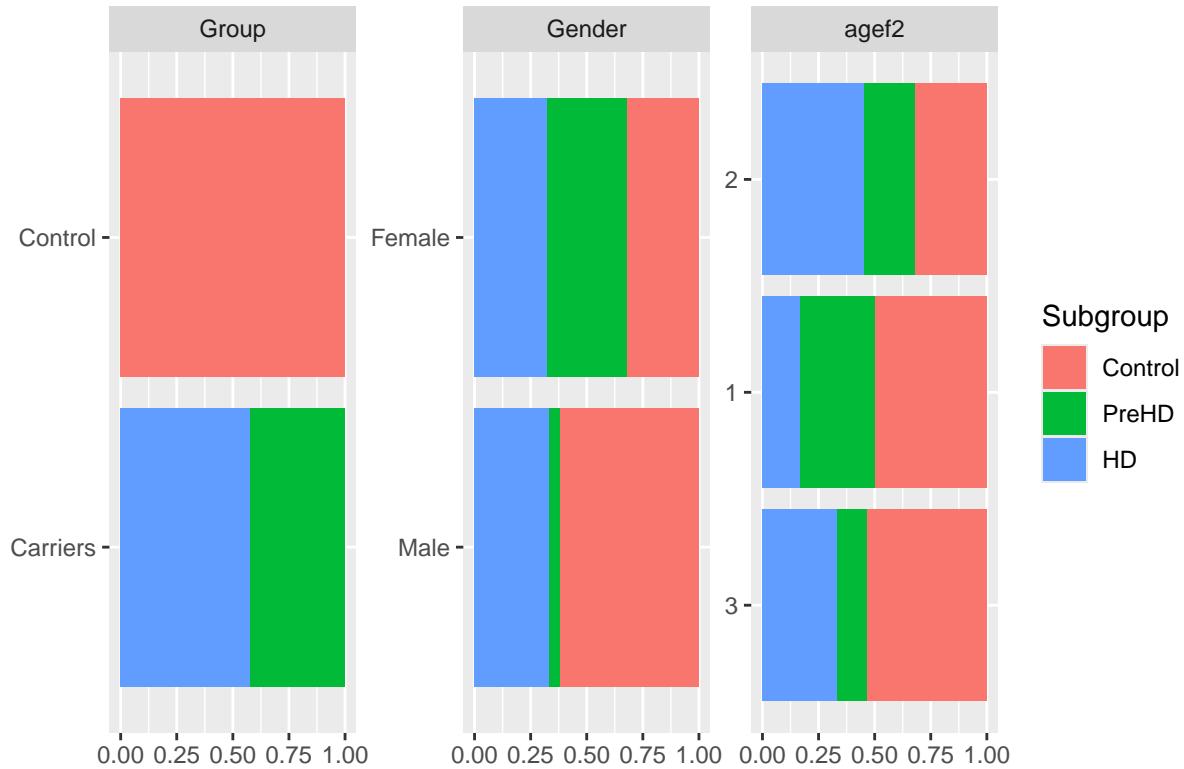
```
caption='Summary statistics', booktabs=TRUE)
```

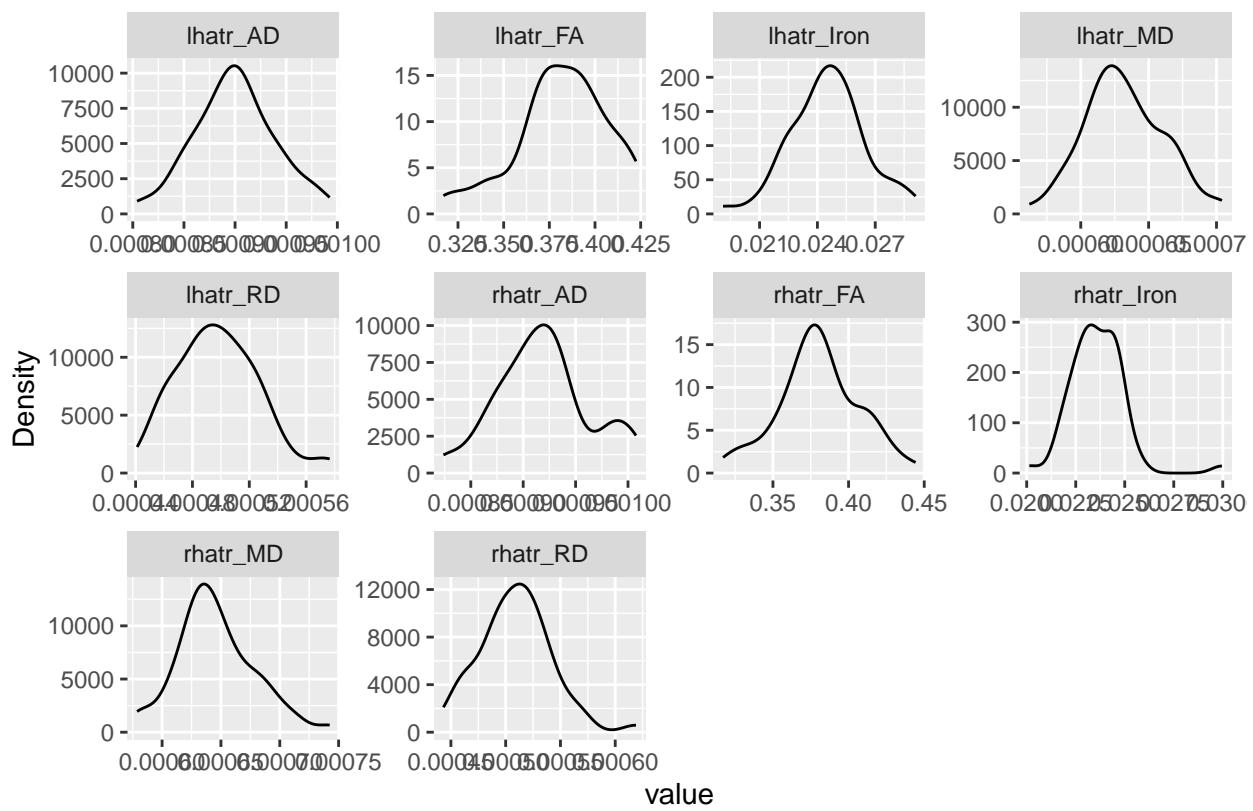
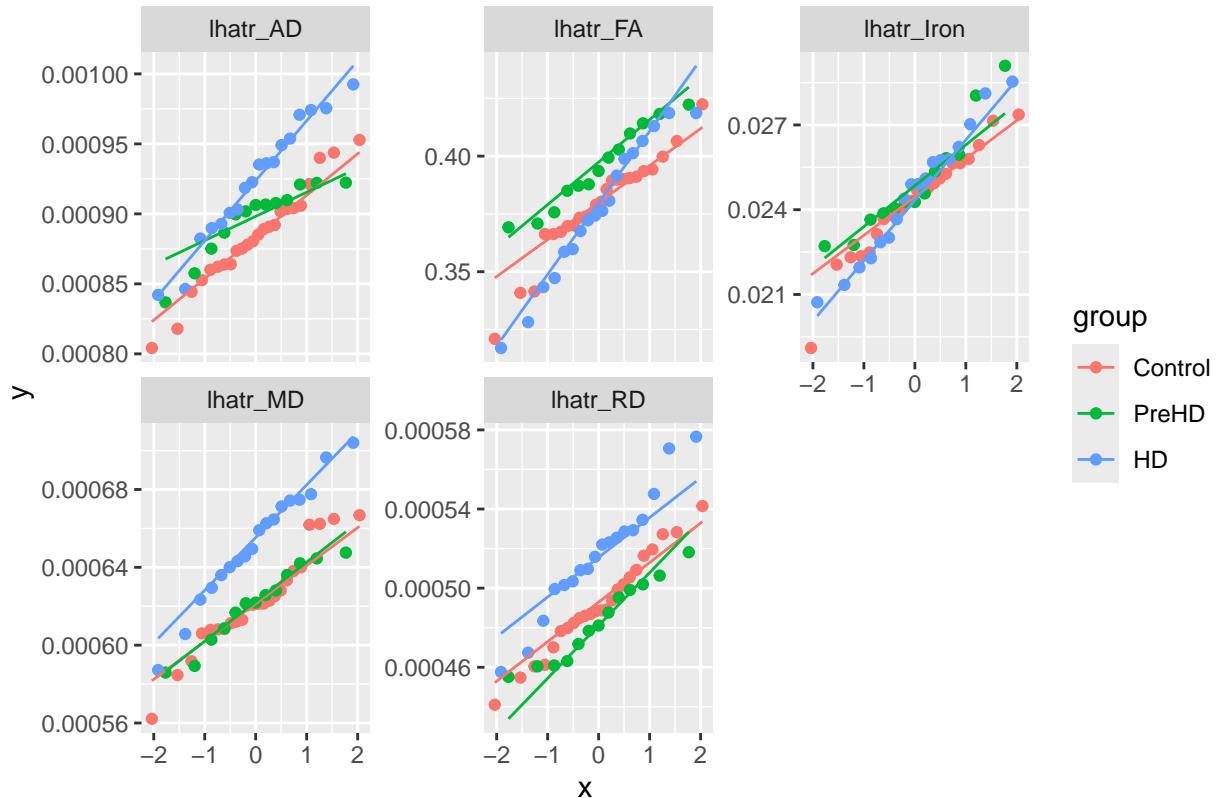
## 2.2 Automated Exploratory Data Analysis

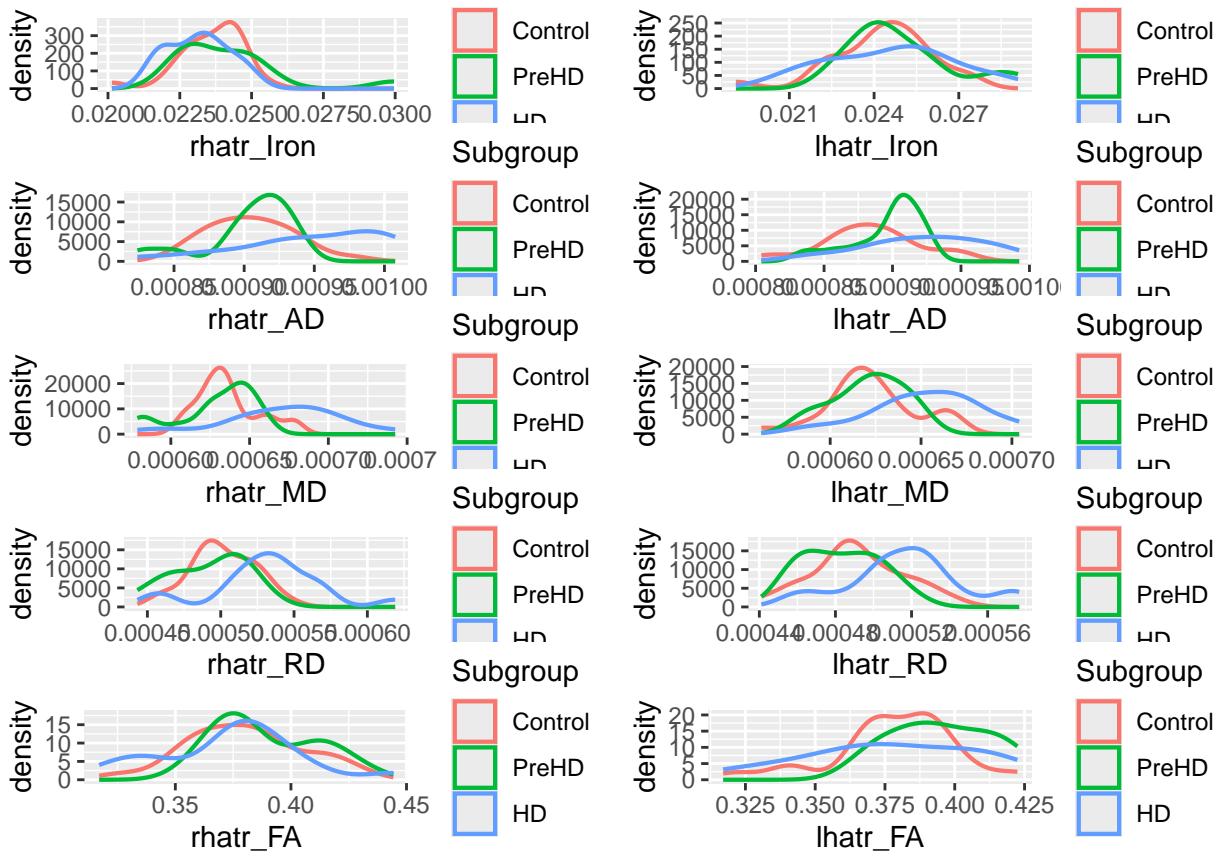
```
library(DataExplorer)
library(dplyr)
ATR_avg_demo<-ATR_avg_demo[,-c(1,5,6)]
ATR_avg_demo %>%
  create_report(
    output_file = paste("Report"),
    report_title = "EDA Report - Both ATR",
    y = "Subgroup"
)
```

Se generan los diferentes plots que se consideran necesarios con las funciones del paquete `DataExplorer`.

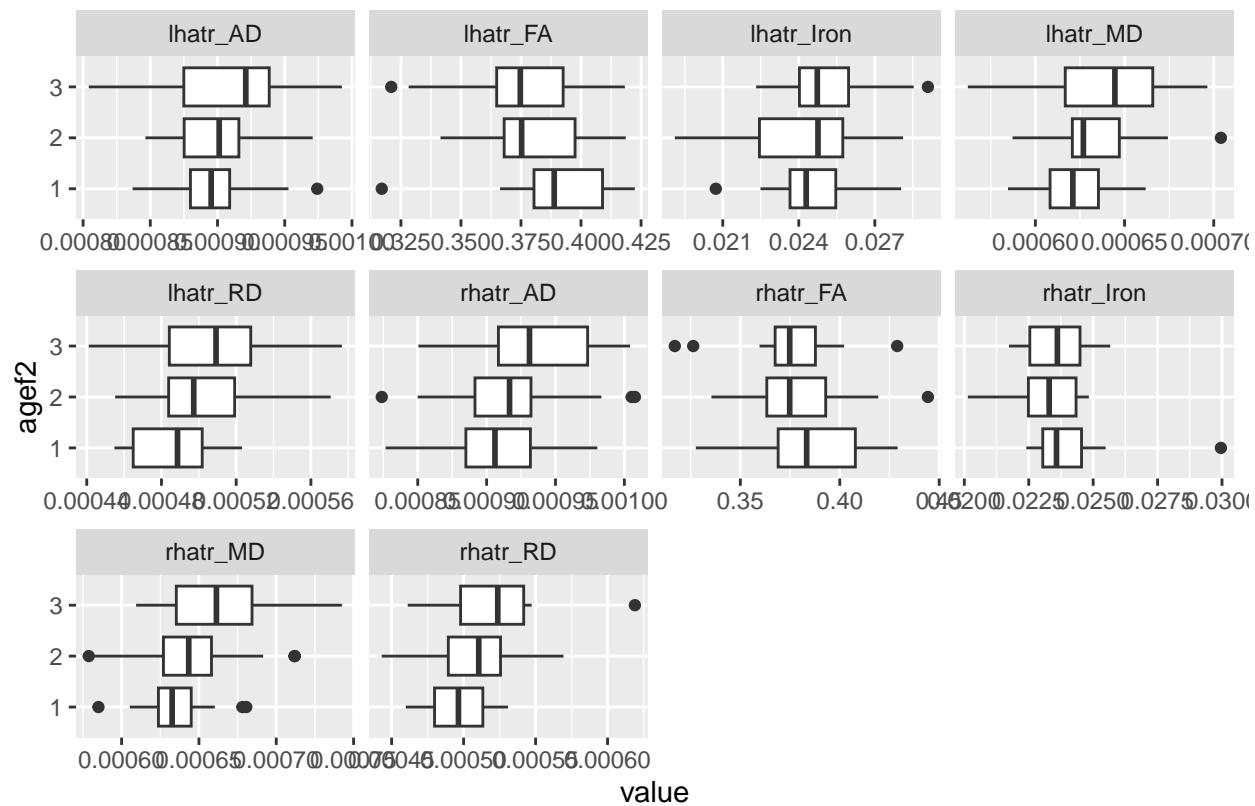
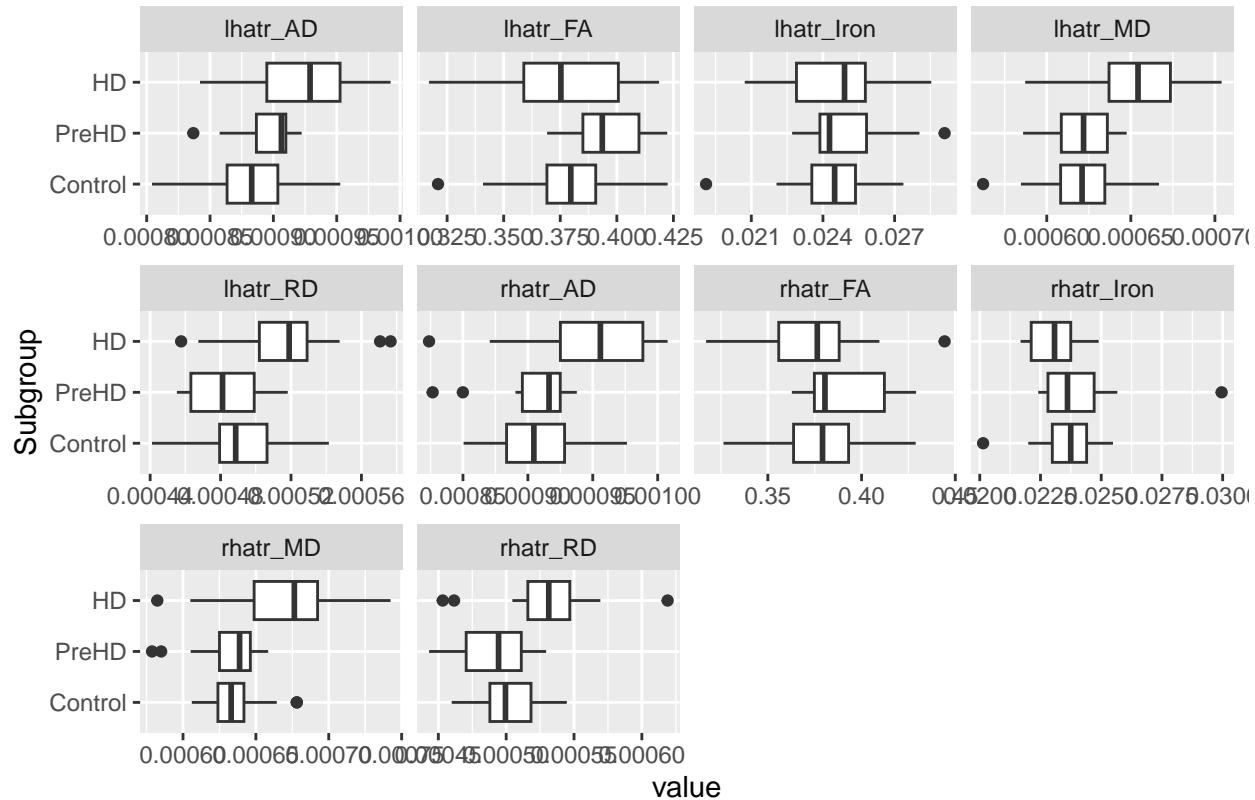


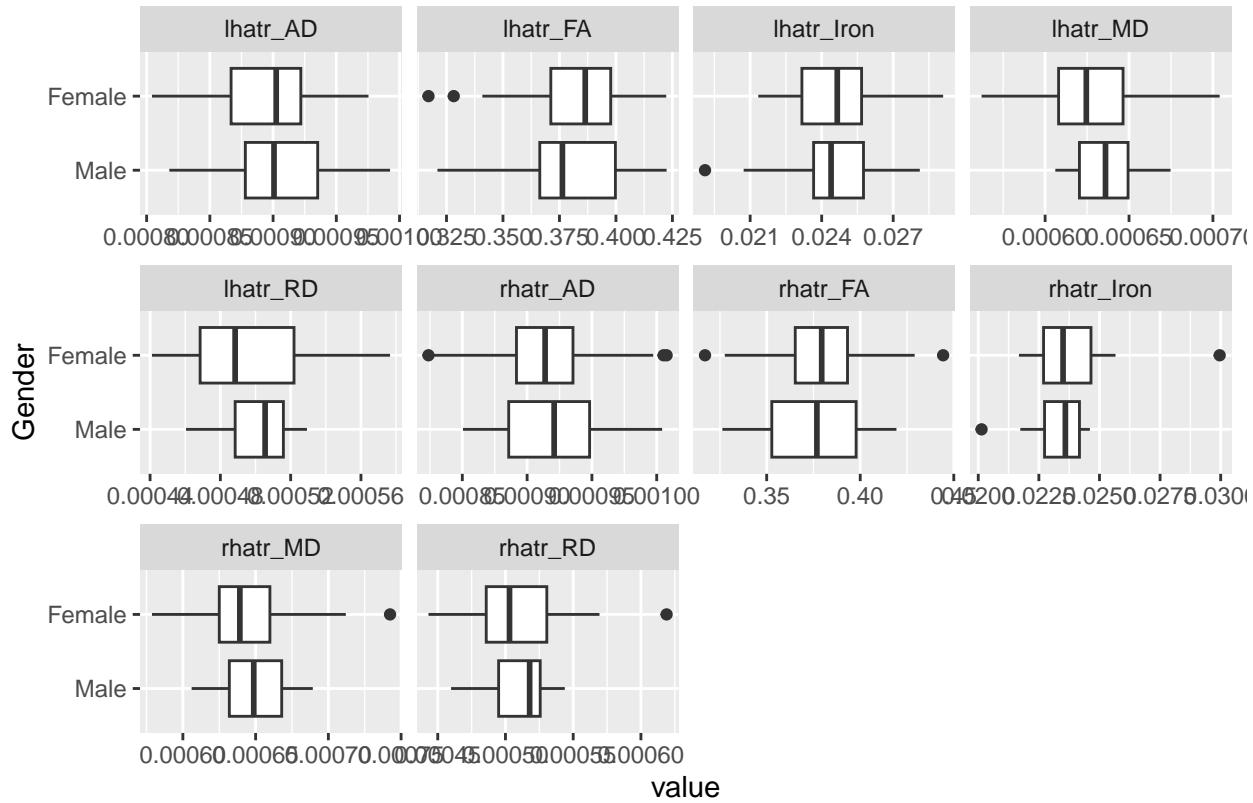






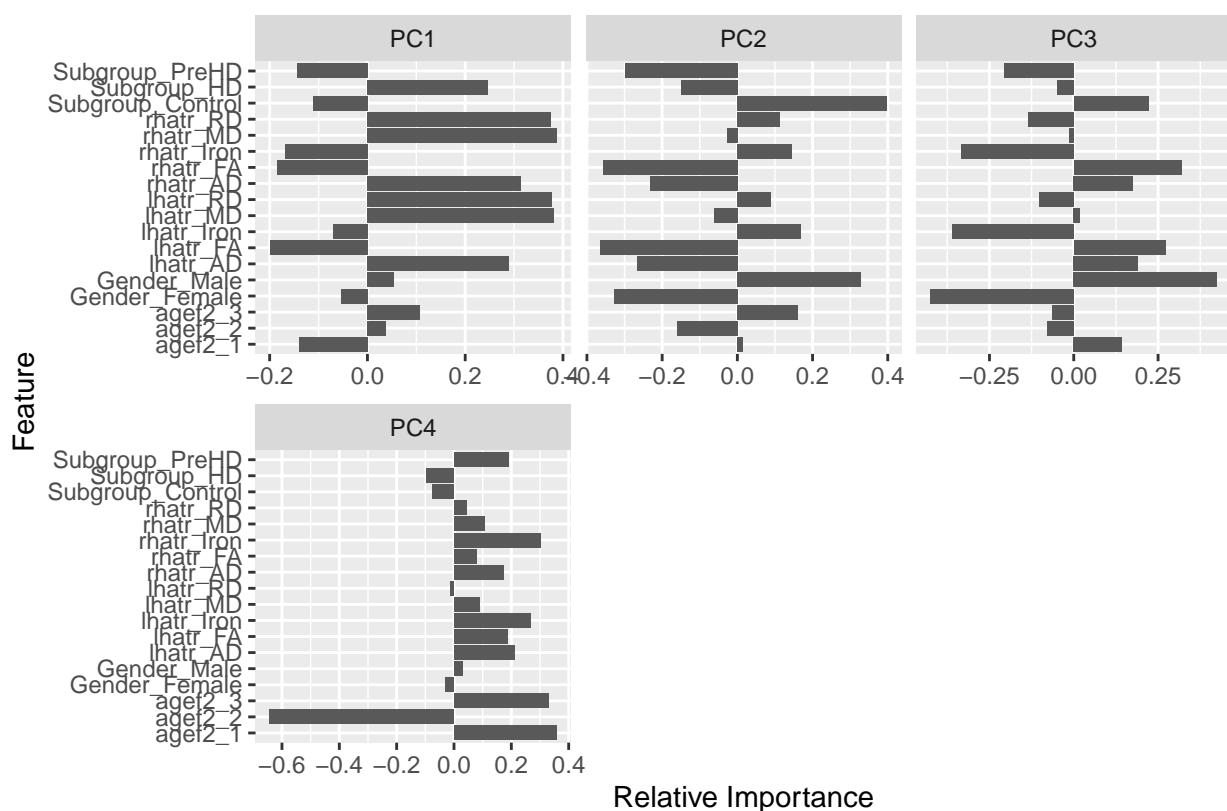
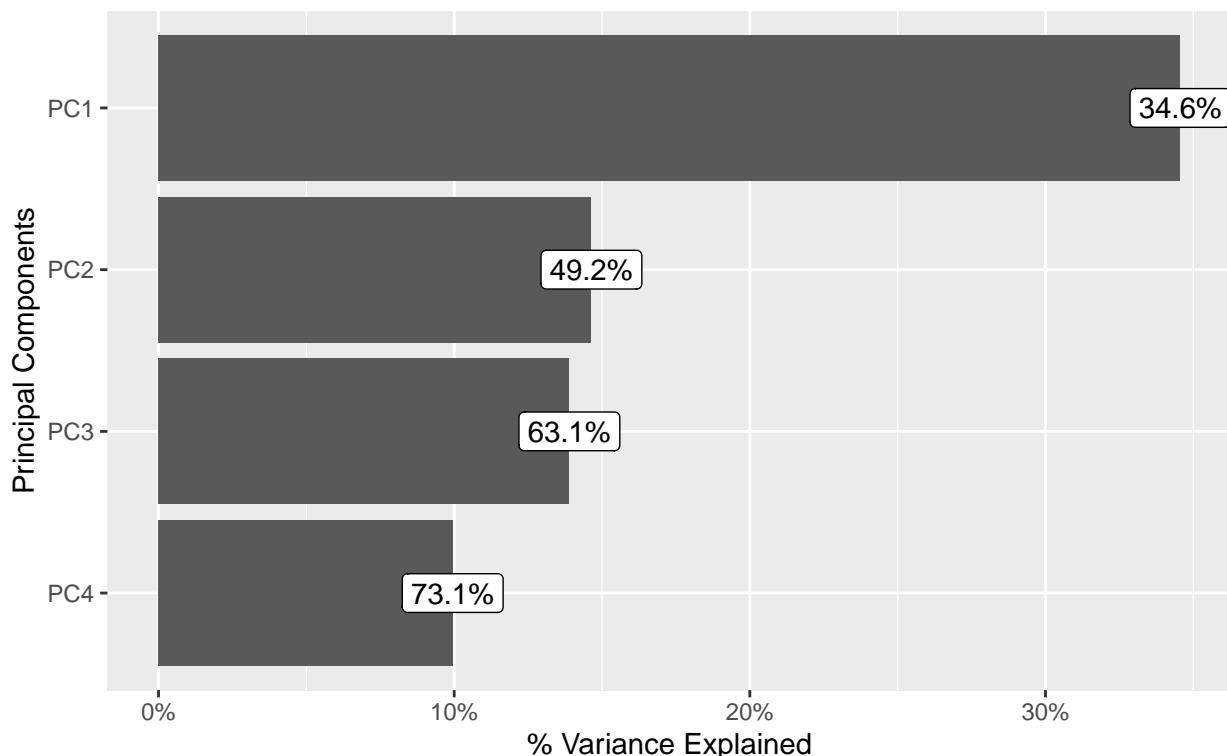
Con los gráficos qq-plot y los density plots, se observa que los pacientes HD tienen mayores valores de difusividad axial, media y radial y menores valores de hierro y de anisotropía fraccional, y, los sujetos presintomáticos tienen mayores valores de anisotropía fraccional y una tendencia a mayores valores de hierro.



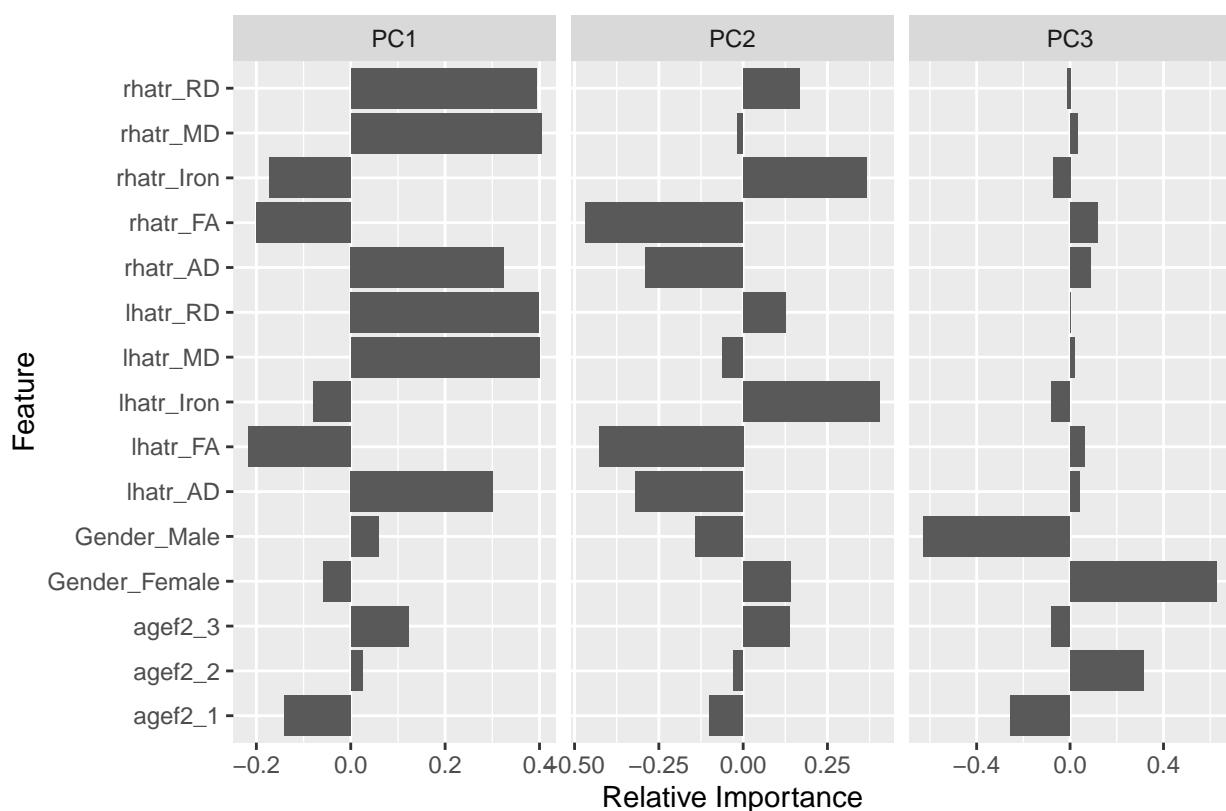
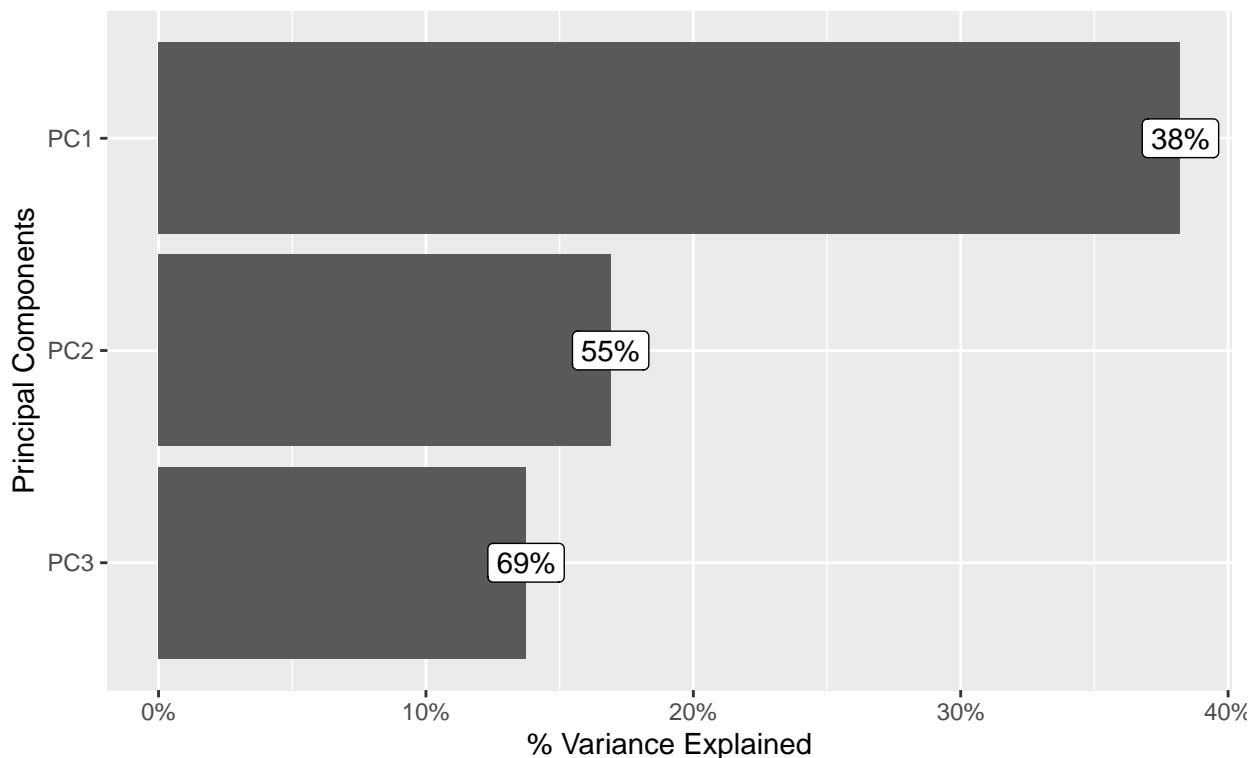


Los boxplots sirven para determinar si existen diferencias entre las variables de RM y los niveles de las variables categóricas (subgrupo, género y edad factorizada).

% Variance Explained By Principal Components  
 (Note: Labels indicate cumulative % explained variance)



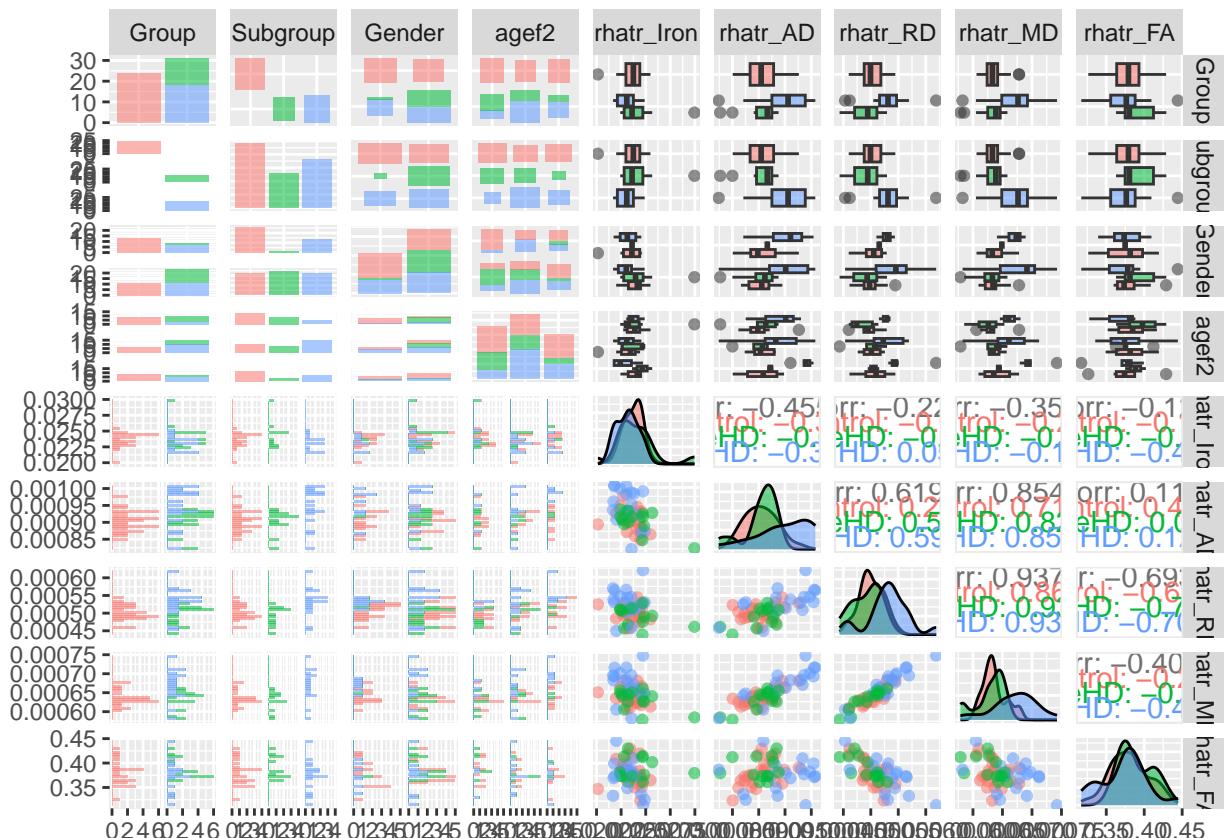
% Variance Explained By Principal Components  
 (Note: Labels indicate cumulative % explained variance)



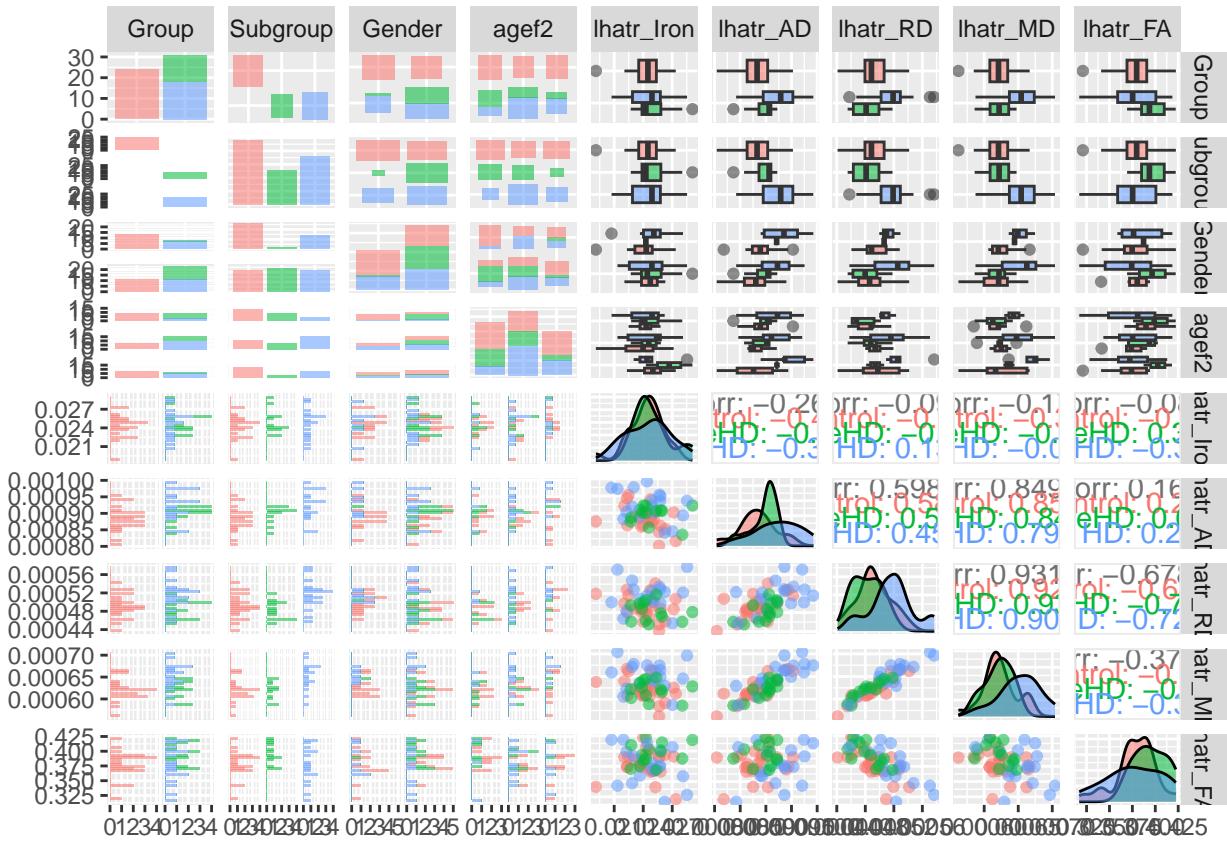
El gráfico del PCA muestra 3 componentes principales, los dos primeros parecen diferenciar los 3 subgrupos y el tercer componente describe las covariables demográficas (género y edad factorizada).

Se genera un plot para cada ATR con la función `ggpairs` del paquete `GGally`.

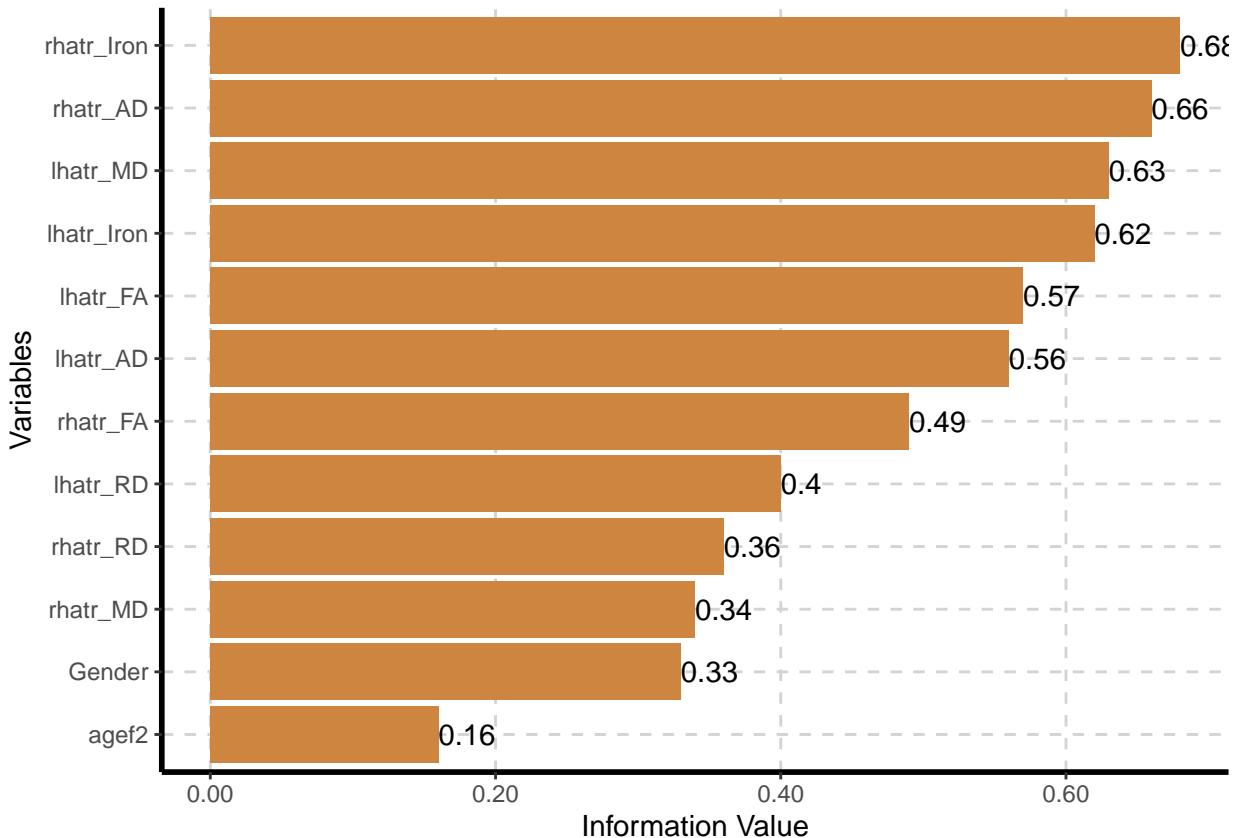
```
ATR_avg_demo<-cbind(demographics, agef, agef2, ATR_avg[,-c(1,2)])
ATR_avg_demo<-ATR_avg_demo[,-c(1,5,6)]
library(GGally)
RATRAvg_demo<-ATR_avg_demo[,c(1:5,7:10)]
RATRAvg_demo %>%
  ggpairs(mapping = aes(color=ATR_avg_demo$Subgroup, alpha=0.5))
```



```
library(GGally)
LATRAvg_demo<-ATR_avg_demo[,c(1:4,6,11:14)]
LATRAvg_demo %>%
  ggpairs(mapping = aes(color=ATR_avg_demo$Subgroup, alpha=0.5))
```



Se genera una tabla mediante la función `ExpCatStat` del paquete `SmartEDA` para determinar la importancia de las variables en función de su información. Esta función combina resultados de la ponderación de la evidencia, el valor de la información y las estadísticas resumidas. Remarcarse que el IV value depende del número de bins que modifica el valor informativo de las variables.



Cuadro 19: Asociaciones y Poder predictivo de las variables respecto a Subgrupo

Variable	Target	Unique	Chi-squared	p-value	df	IV Value	Cramers V	Degree of Association	Predictive Power
Gender	Subgroup	2	7.722	0.019	NA	0.33	0.37	Strong	Highly Predictive
agef2	Subgroup	3	5.101	0.274	NA	0.16	0.22	Moderate	Somewhat Predictive
rhatr_Iron	Subgroup	15	33.410	0.216	NA	0.68	0.55	Strong	Highly Predictive
lhatr_Iron	Subgroup	15	28.151	0.509	NA	0.62	0.51	Strong	Highly Predictive
rhatr_AD	Subgroup	15	40.746	0.036	NA	0.66	0.61	Strong	Highly Predictive
rhatr_RD	Subgroup	15	36.598	0.116	NA	0.36	0.58	Strong	Highly Predictive
rhatr_MD	Subgroup	15	43.610	0.022	NA	0.34	0.63	Strong	Highly Predictive
rhatr_FA	Subgroup	15	29.679	0.397	NA	0.49	0.52	Strong	Highly Predictive
lhatr_AD	Subgroup	15	36.838	0.100	NA	0.56	0.58	Strong	Highly Predictive
lhatr_RD	Subgroup	15	31.417	0.315	NA	0.40	0.53	Strong	Highly Predictive
lhatr_MD	Subgroup	15	30.227	0.366	NA	0.63	0.52	Strong	Highly Predictive
lhatr_FA	Subgroup	15	19.572	0.929	NA	0.57	0.42	Strong	Highly Predictive

La tabla muestra el grado de asociación y el poder predictivo de cada variable (métricas de RM, género y edad factorizada) con respecto a la variable diana (en este caso, subgrupo). La función `ExpCatStat` del paquete `SmartEDA` combina los resultados del peso de la evidencia, el valor informativo y el resumen estadístico de cada variable categórica introducida. Para las variables numéricas, la función las discretiza y, en función del número de bins, los resultados pueden variar.

## 2.3 Estadística descriptiva gráfica

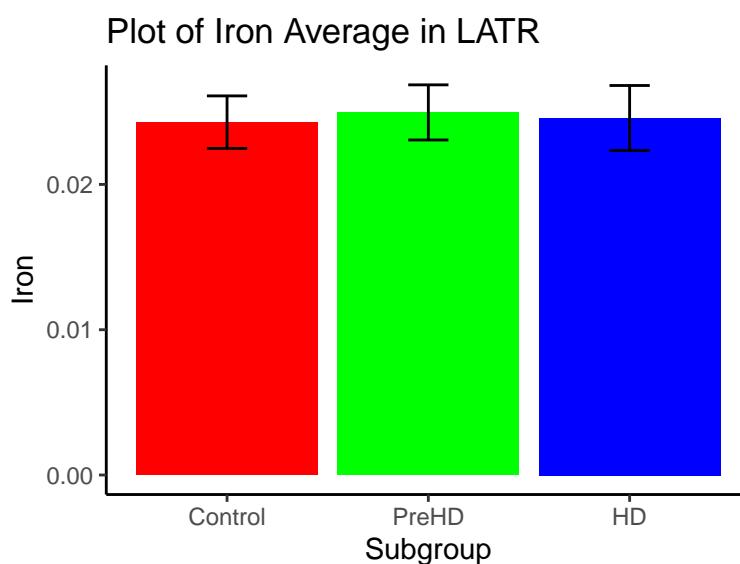
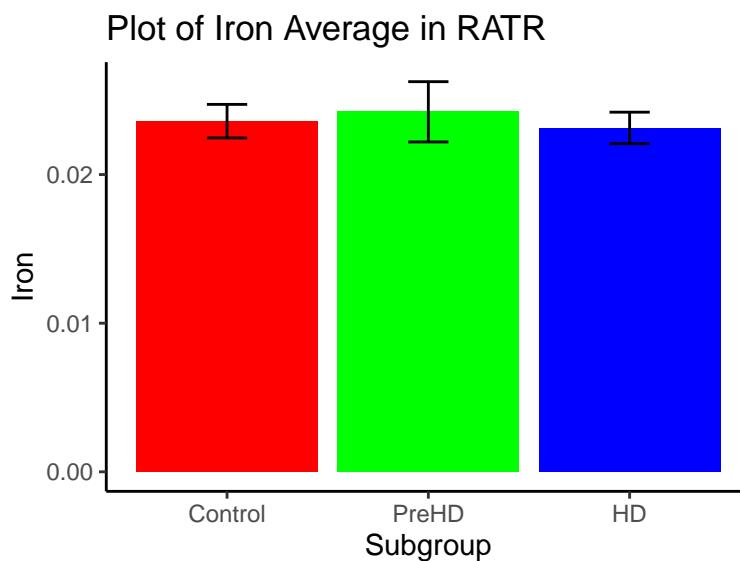
Para determinar el **comportamiento de los datos (normales o no normales)**, se procede a la visualización gráfica de los datos así como a la determinación de medidas estadísticas.

### 2.3.1 BARPLOTS

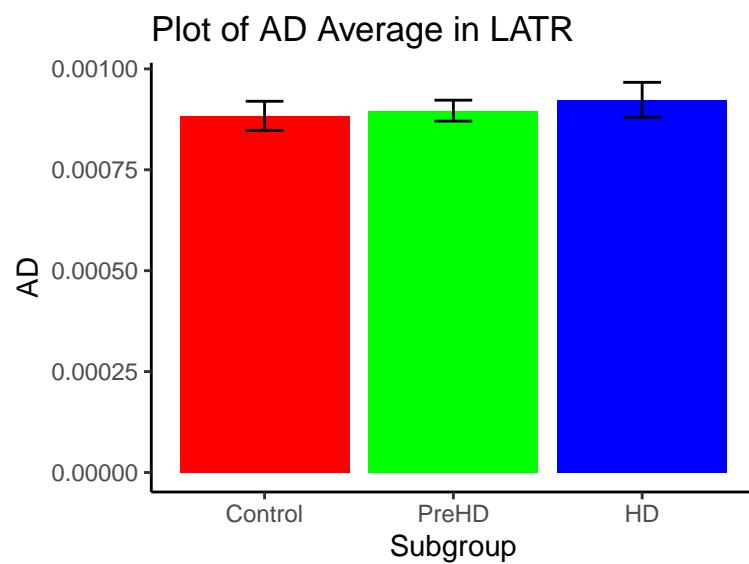
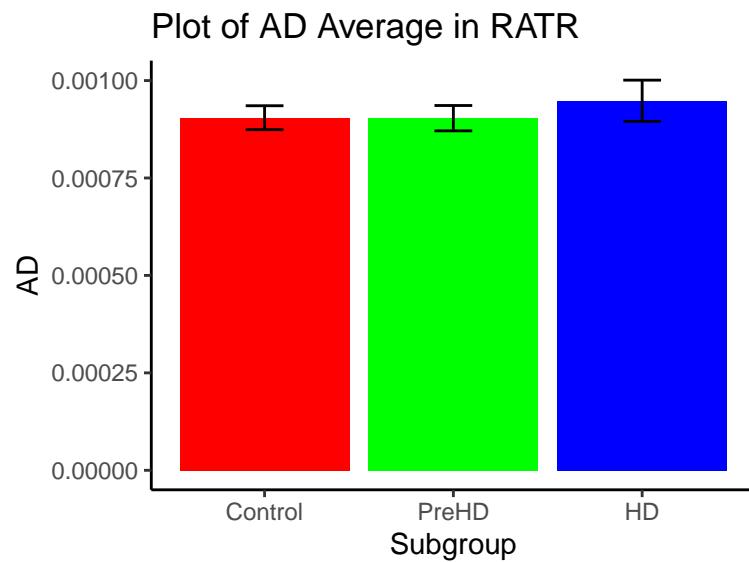
Se procede a la visualización de los datos mediante barplots (la altura del barplot es la media de la medida y se añade un errorbar que es la media +/- desviación estándar) que da una idea de las posibles diferencias entre subgrupos.

Se consulta la siguiente referencia: ggplot2 barplots : Quick start guide - R software and data visualization.

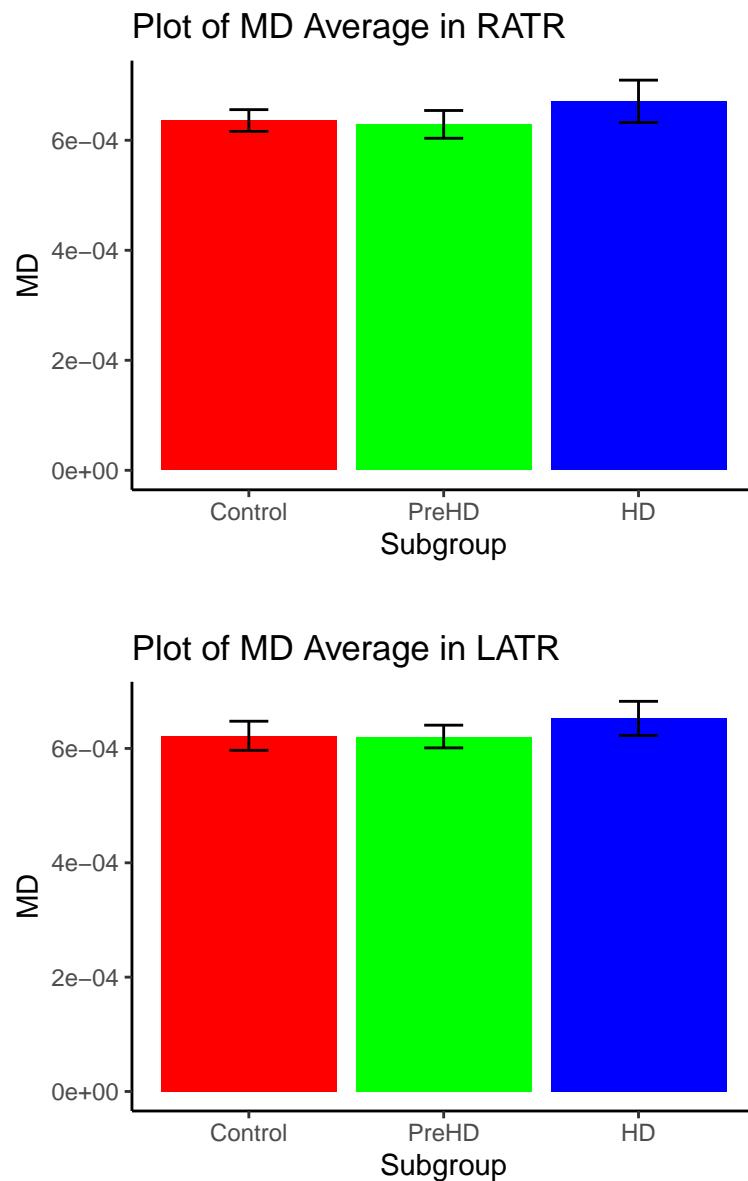
#### 2.3.1.1 IRON



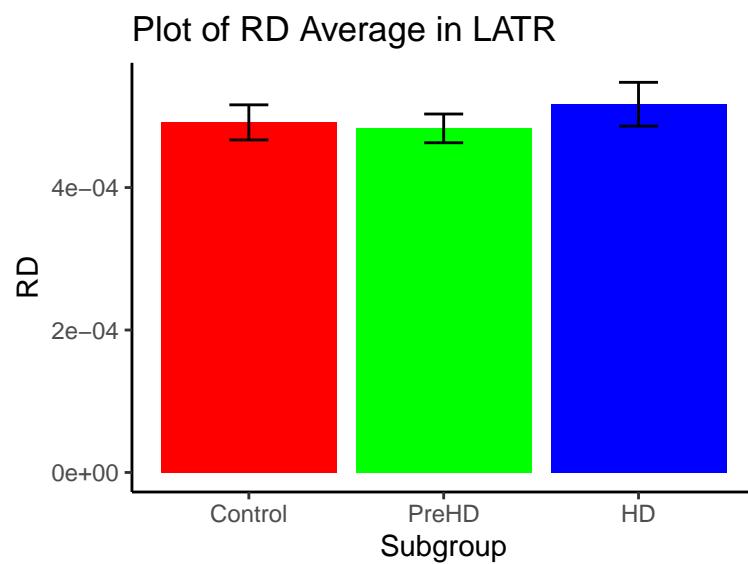
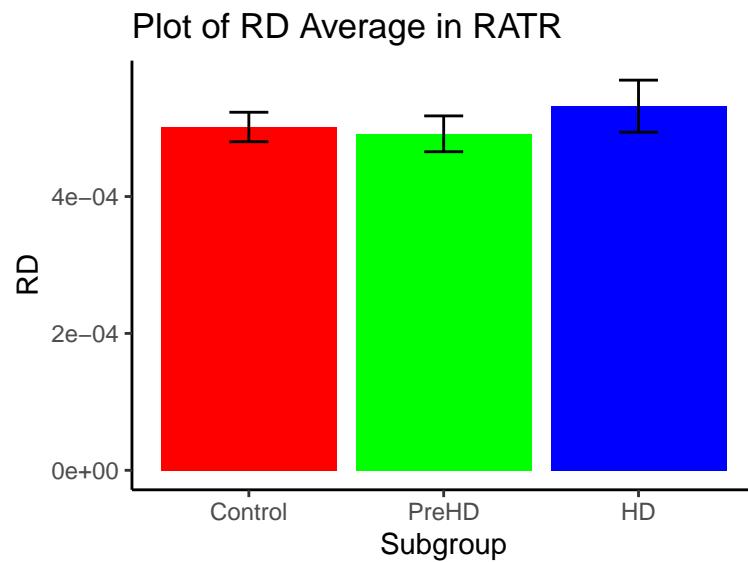
### 2.3.1.2 AD



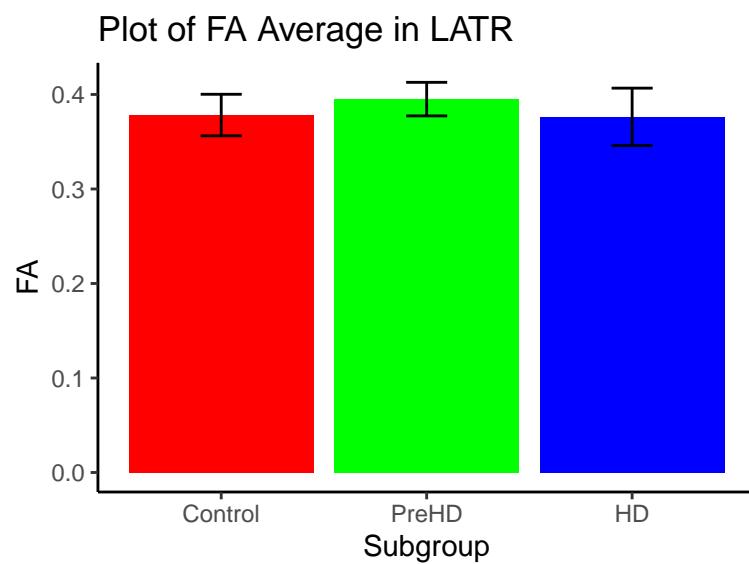
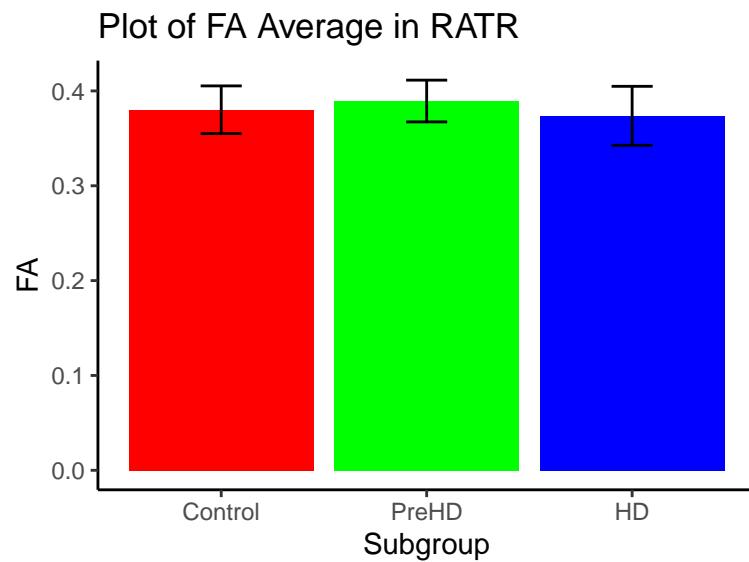
### 2.3.1.3 MD



#### 2.3.1.4 RD



### 2.3.1.5 FA



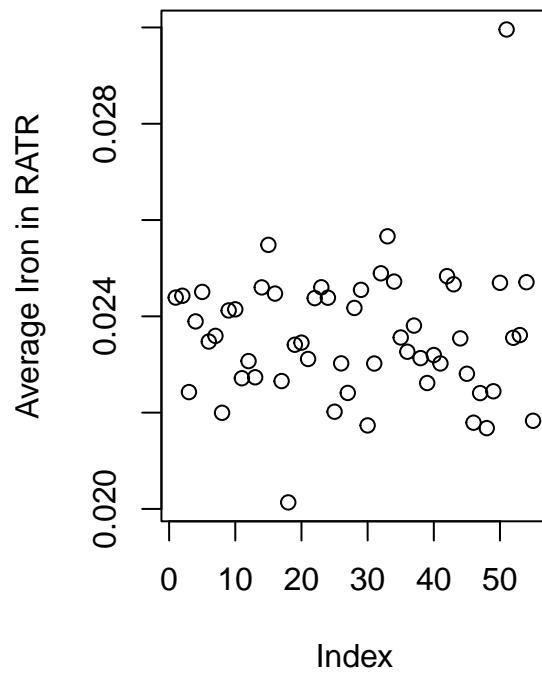
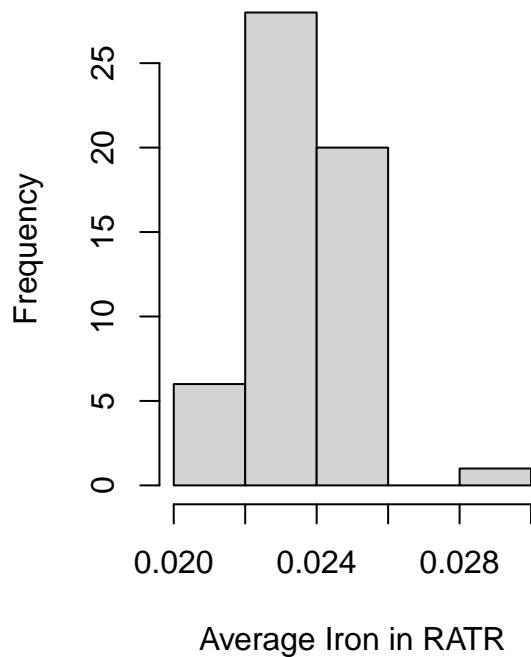
### 2.3.2 HISTOGRAMAS y QQ-PLOTS

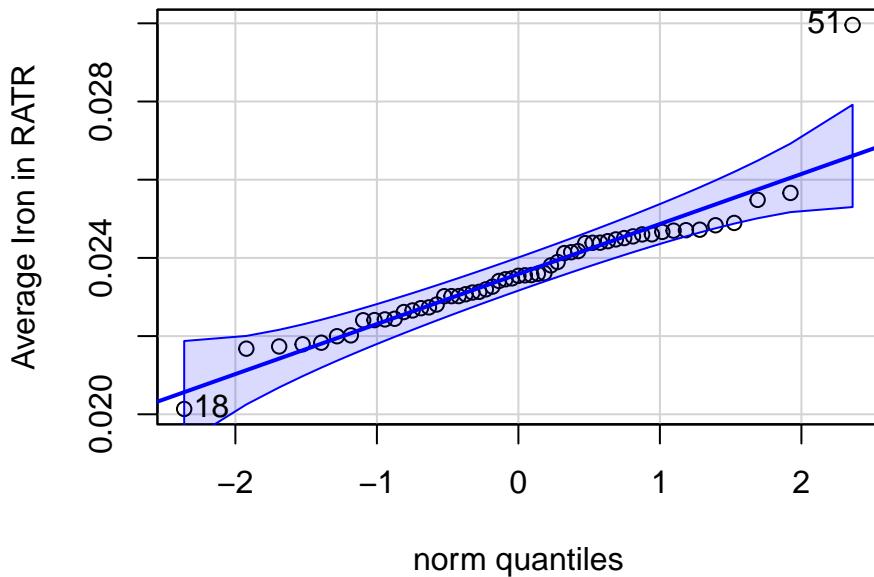
Para obtener una idea de la **normalidad de los datos y de la presencia/ausencia de outliers**, para cada métrica de RM, se genera un histograma, un plot y cuatro qq-plots (un qqplot sin distinción entre subgrupos y un qqplot para cada subgrupo).

Para valorar posibles **diferencias en variabilidad de las métricas de RM en los niveles de los diferentes factores de interés** subgrupo, género y edad\_factorizada, también se generan plots de las métricas de RM para cada subgrupo, género y grupo de edad (tanto para la edad factorizada de 7 niveles como la edad factorizada de 3 niveles).

#### 2.3.2.1 IRON - RATR

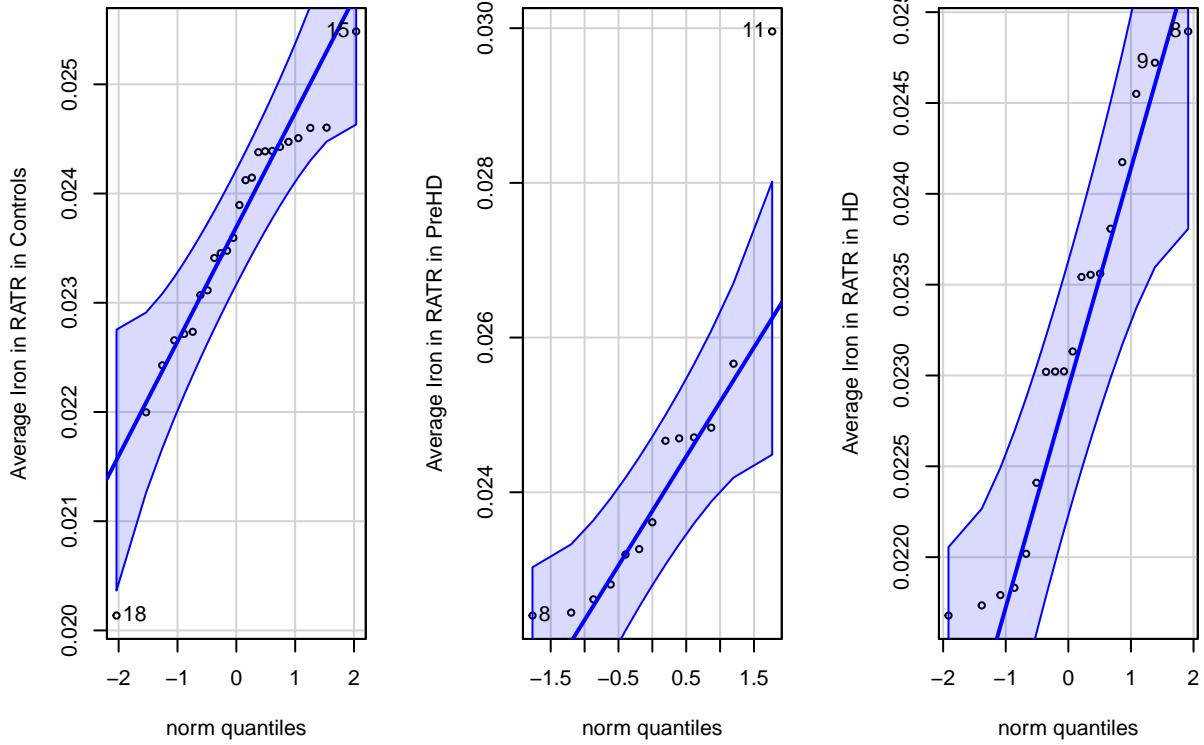
Histogram of RATR Average Iron



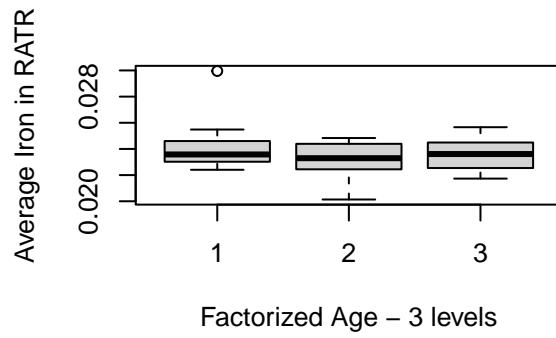
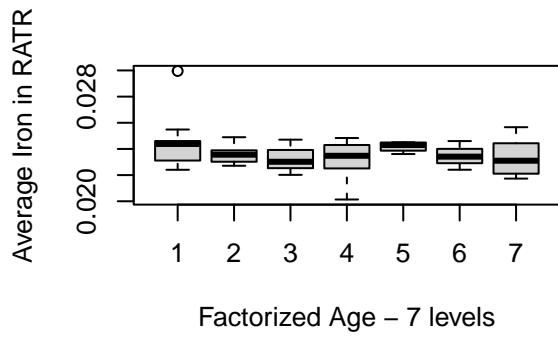
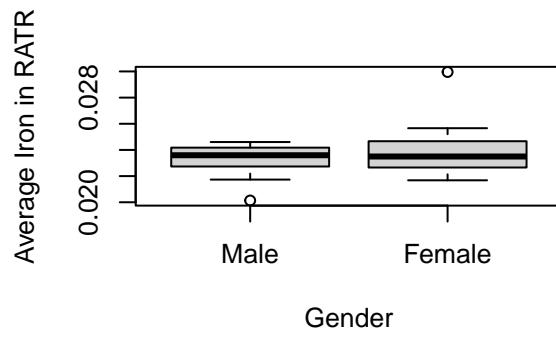
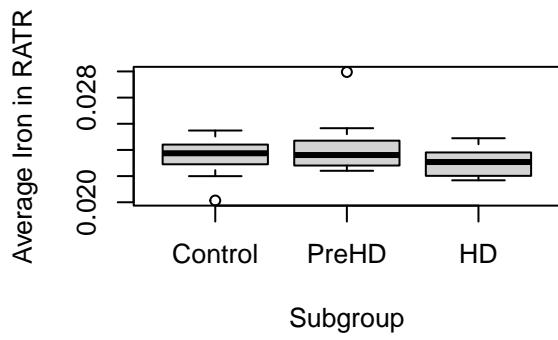


[1] 51 18

[1] 18 15  
[1] 11 8

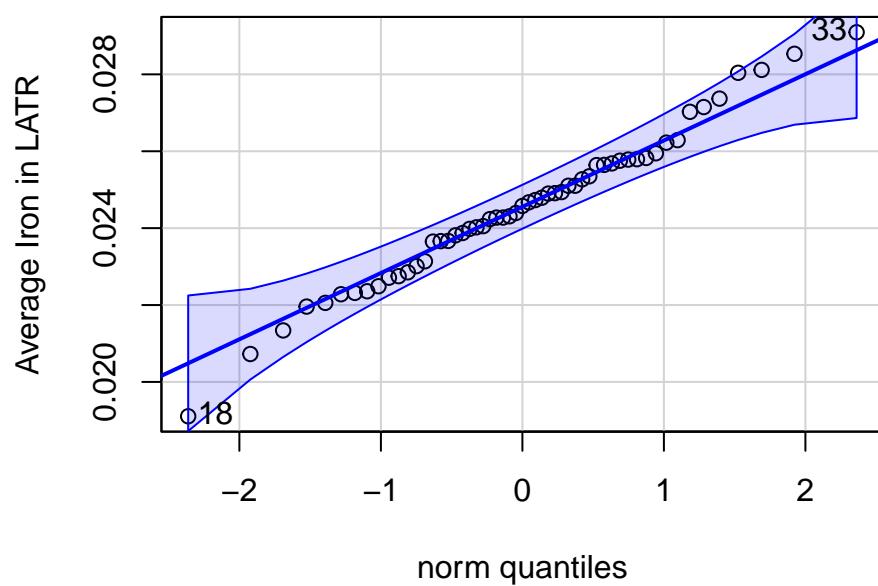
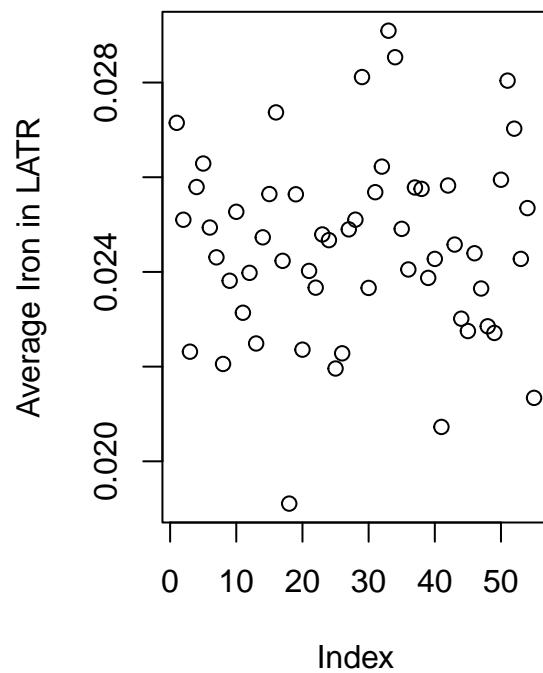
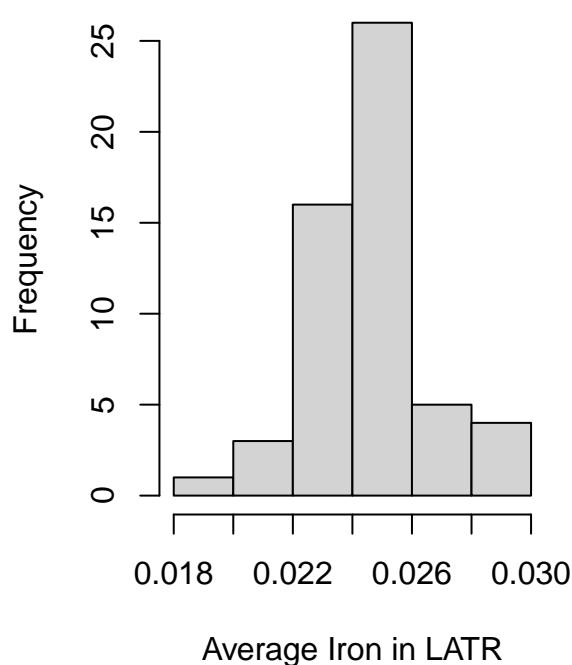


[1] 8 9



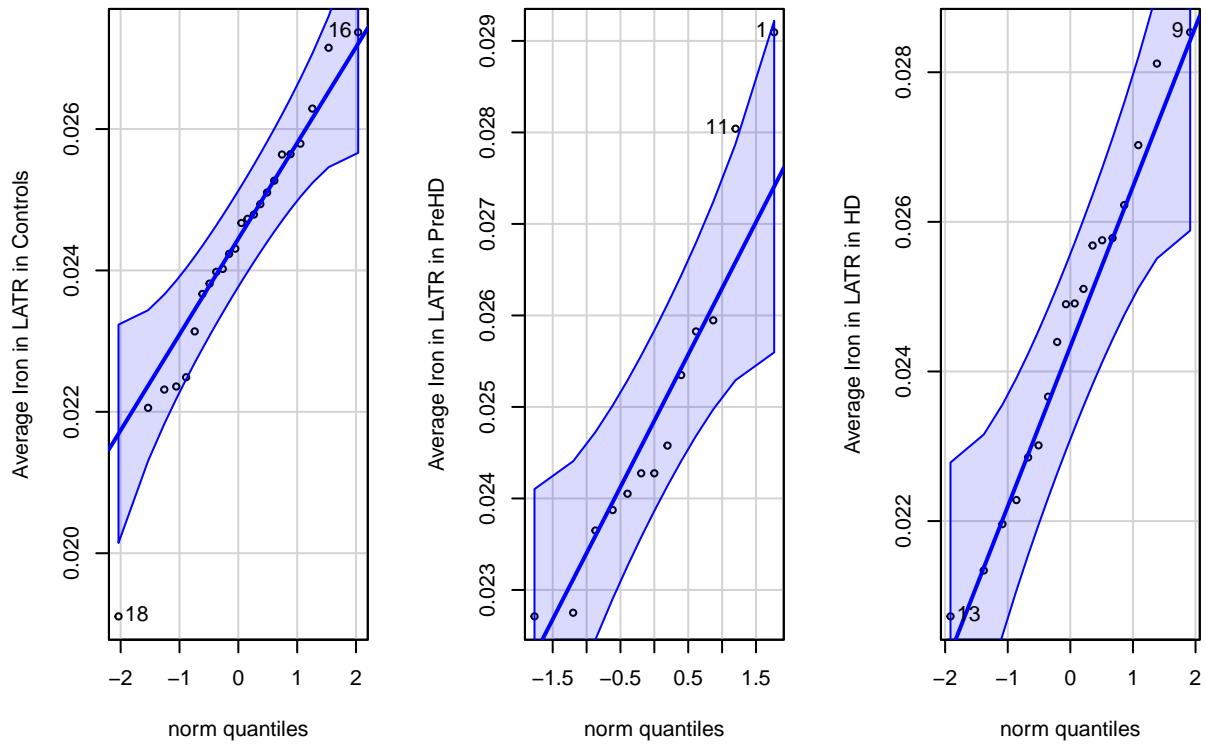
### 2.3.2.2 IRON - LATR

Histogram of LATR Average Iron

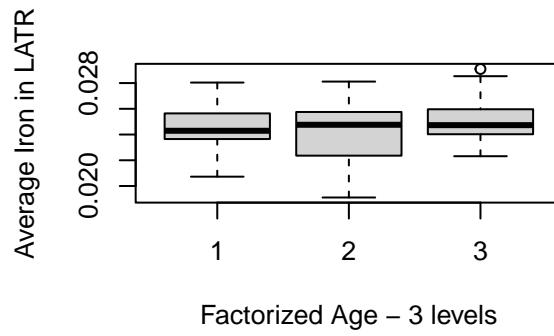
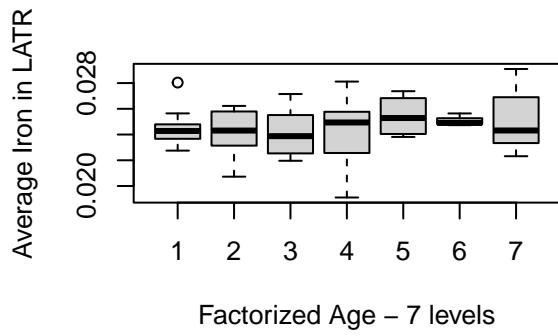
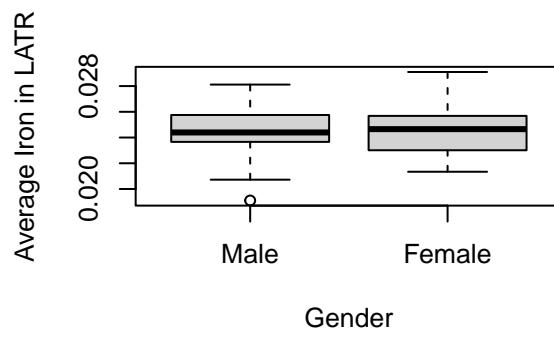
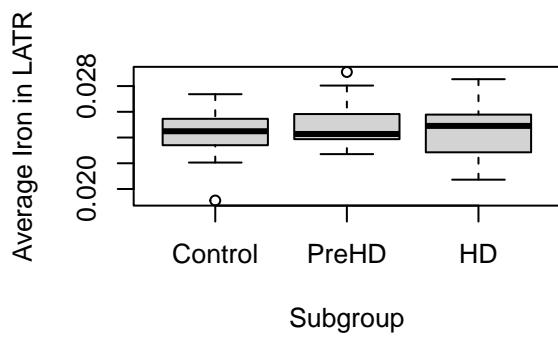


[1] 18 33

```
[1] 18 16  
[1] 1 11
```

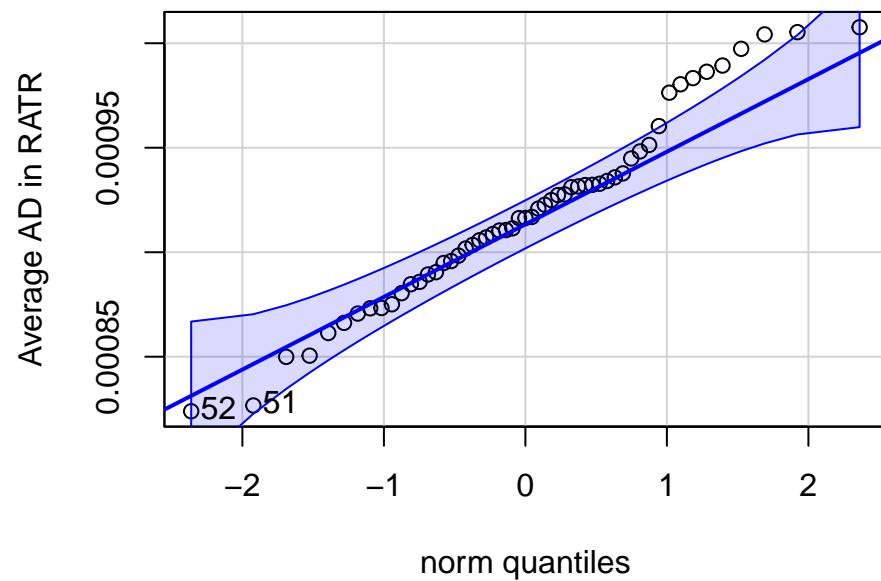
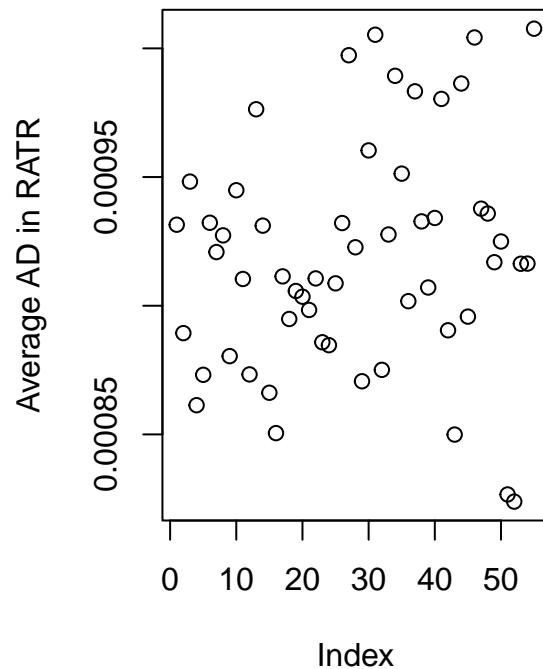
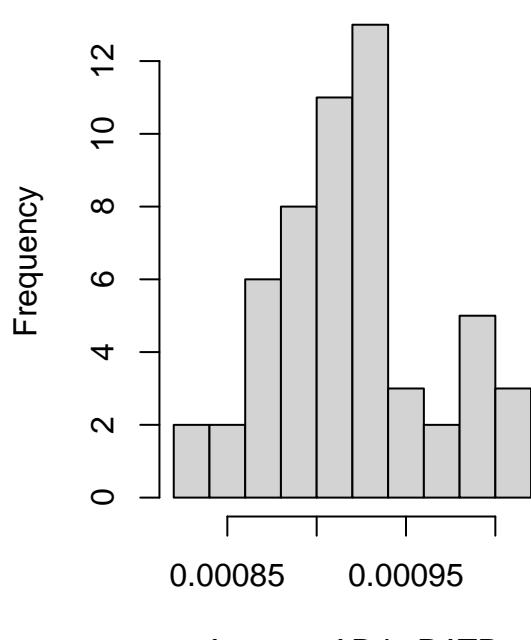


```
[1] 9 13
```



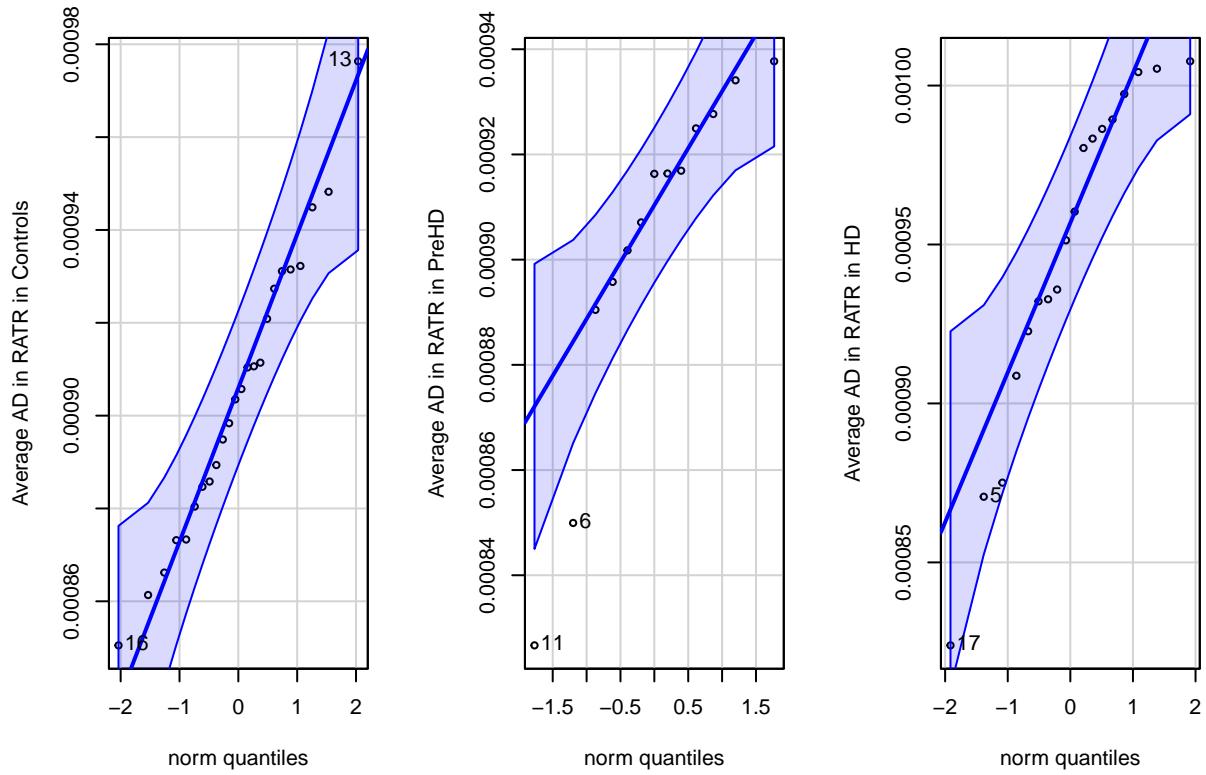
### 2.3.2.3 AD - RATR

Histogram of RATR Average AD

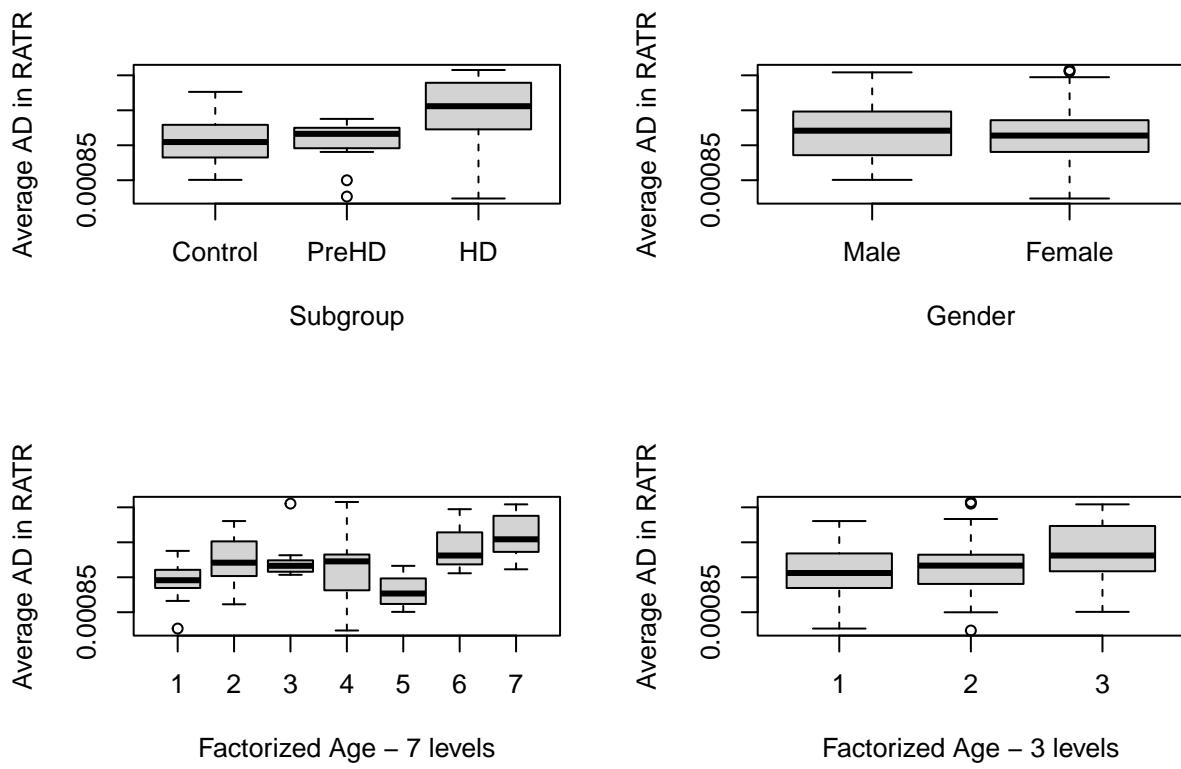


[1] 52 51

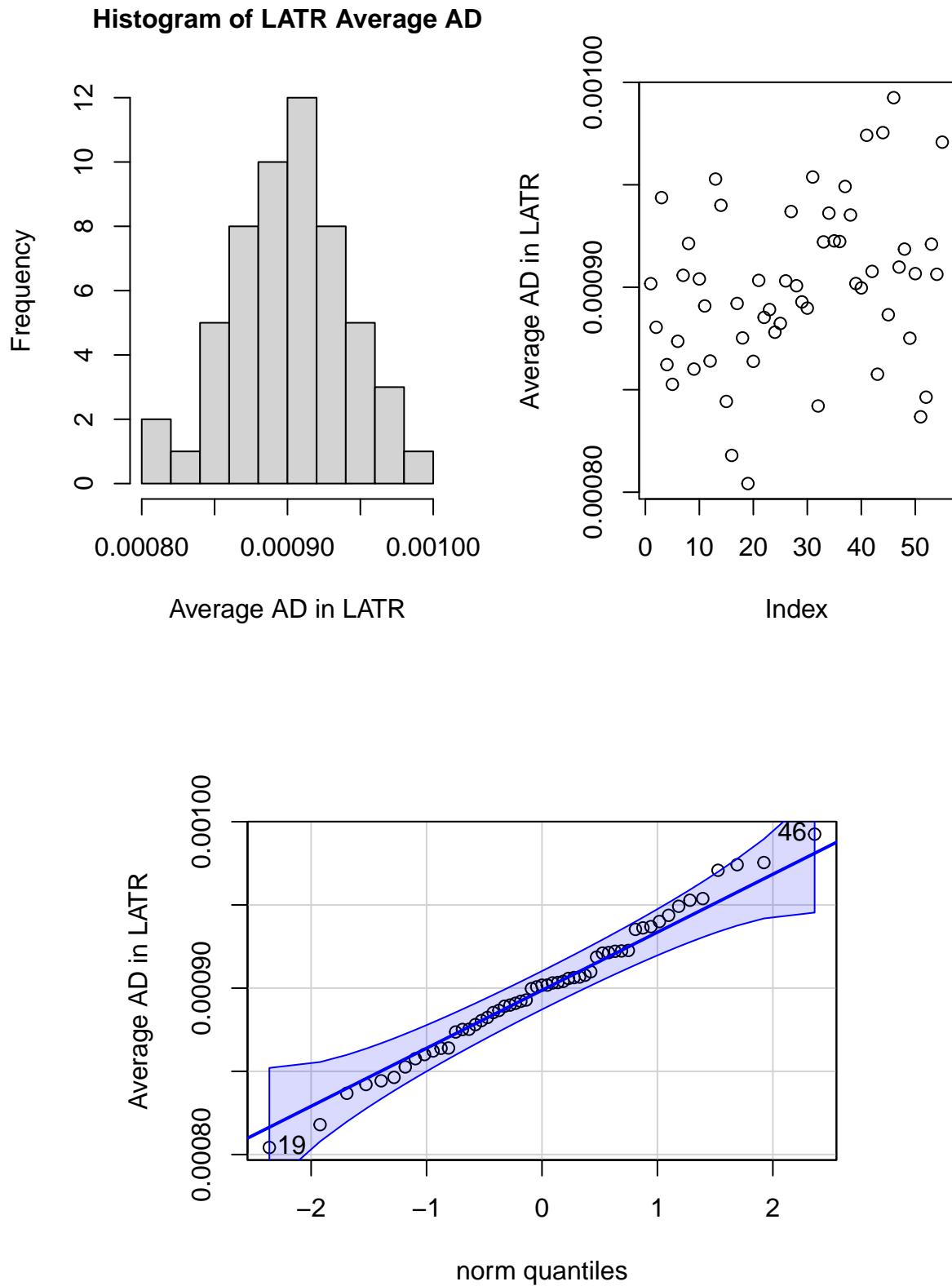
```
[1] 13 16  
[1] 11 6
```



```
[1] 17 5
```

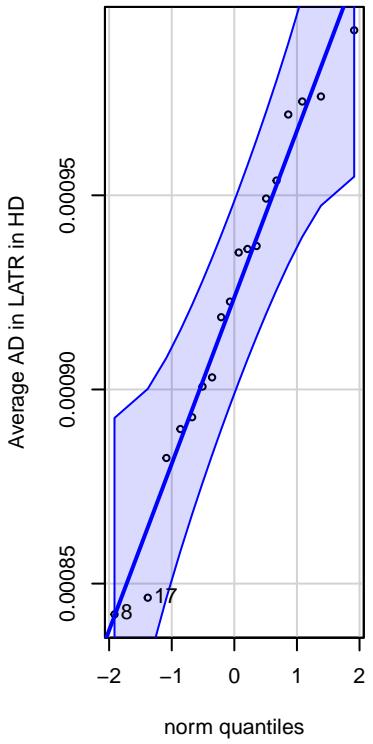
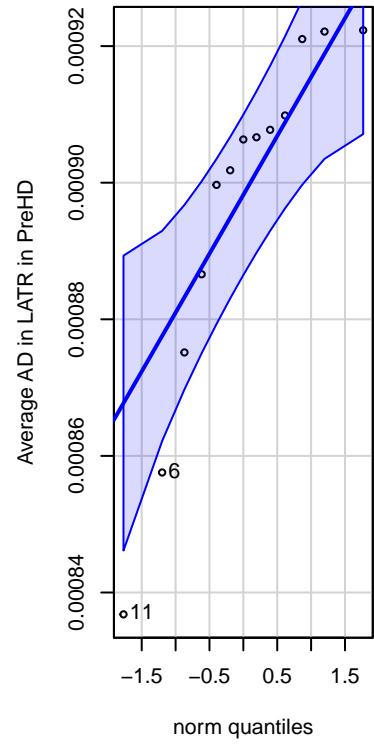
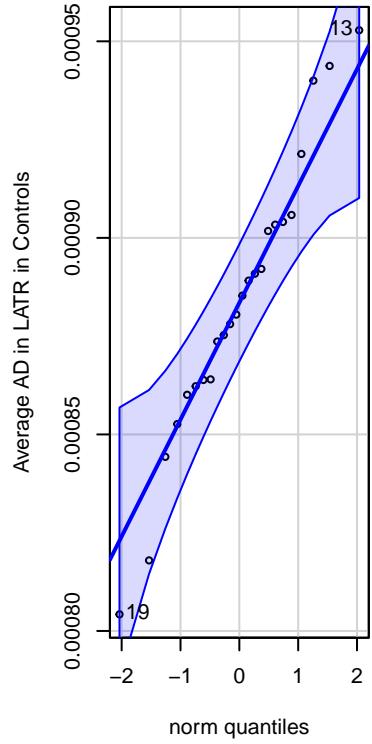


#### 2.3.2.4 AD - LATR

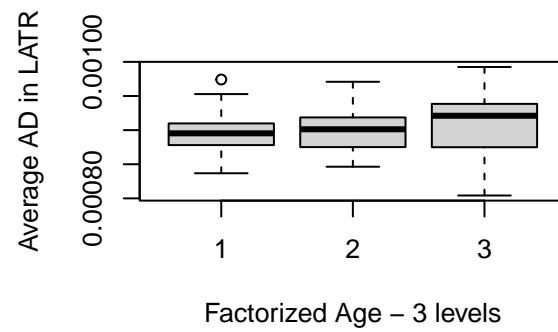
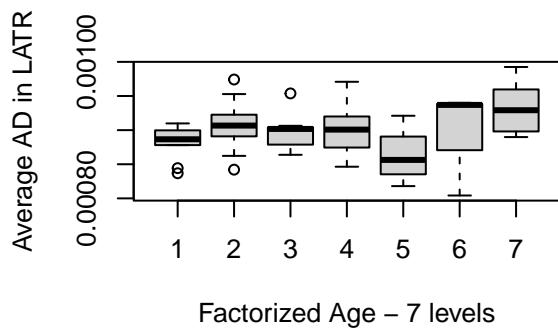
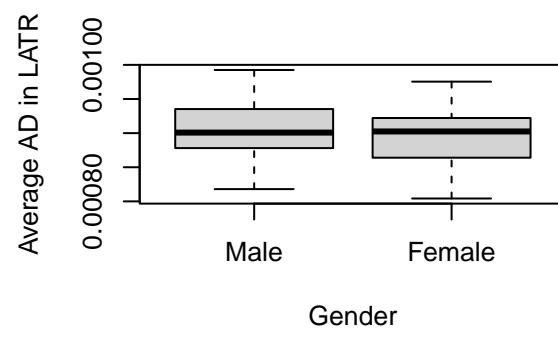
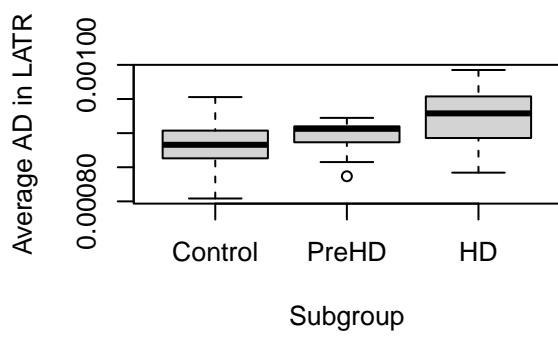


[1] 19 46

```
[1] 19 13  
[1] 11 6
```

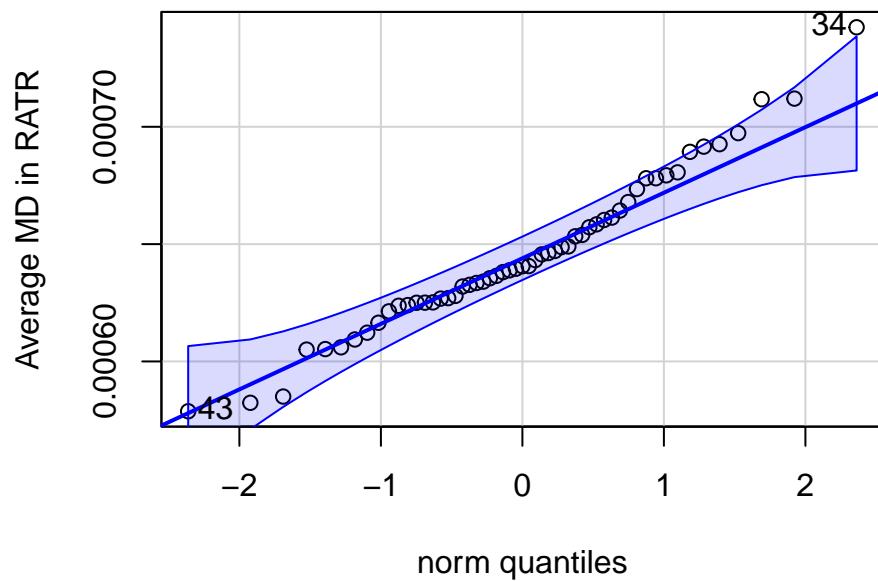
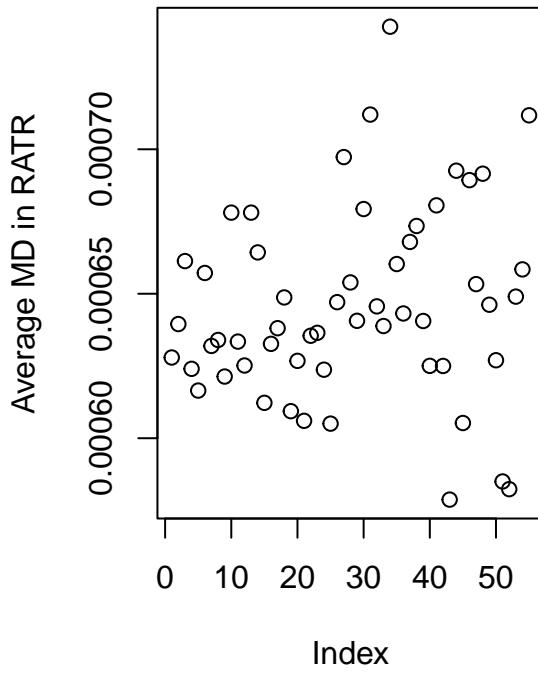
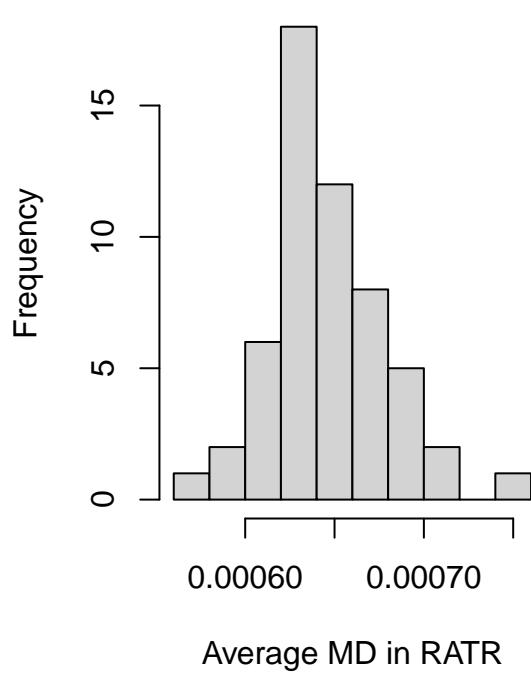


```
[1] 8 17
```



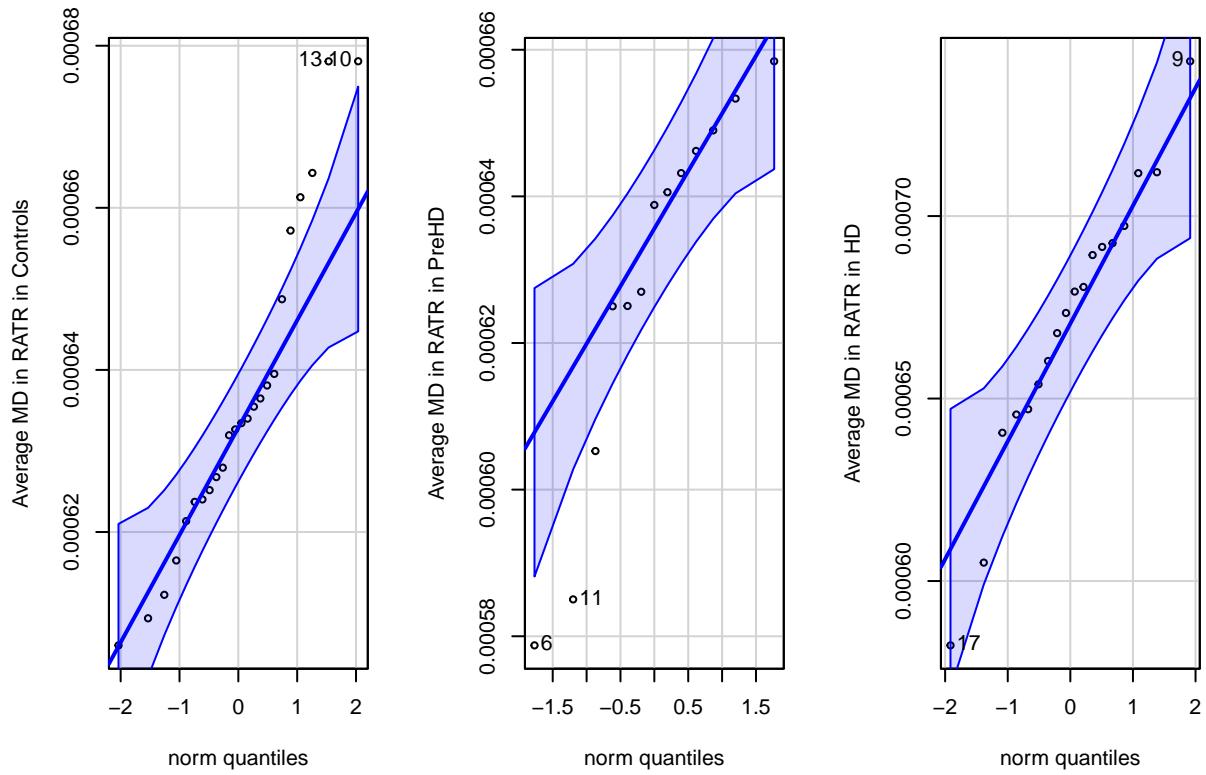
#### 2.3.2.5 MD - RATR

Histogram of RATR Average MD

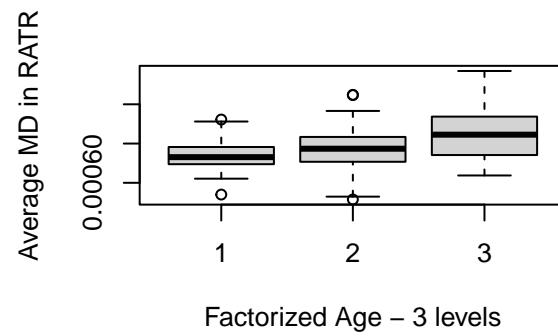
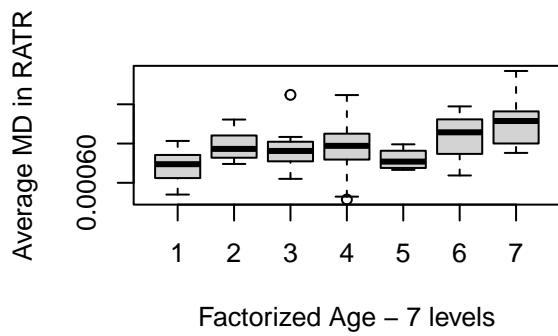
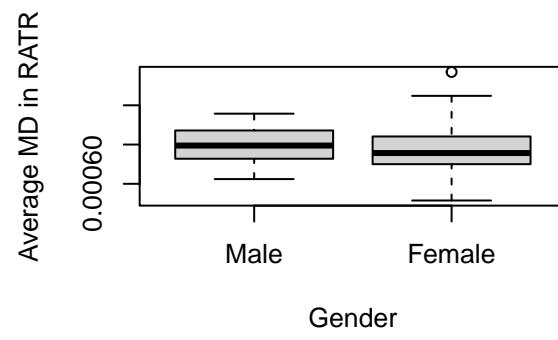
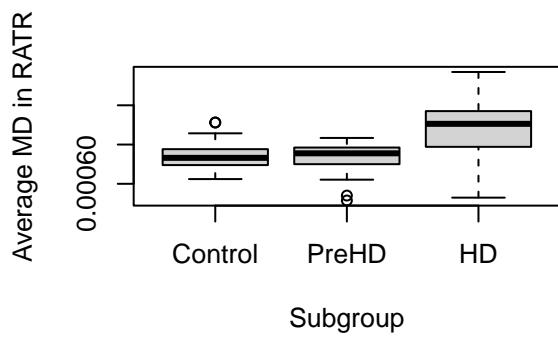


[1] 34 43

```
[1] 10 13  
[1] 6 11
```

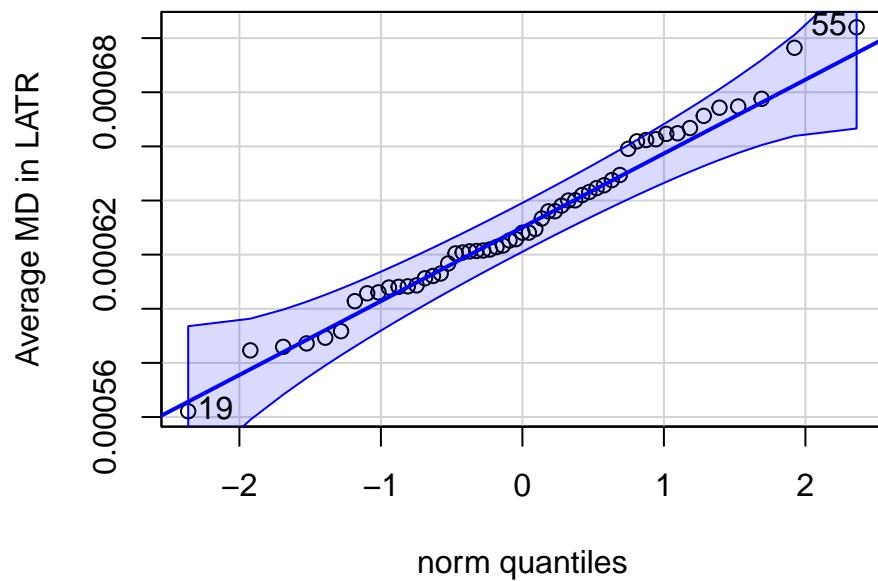
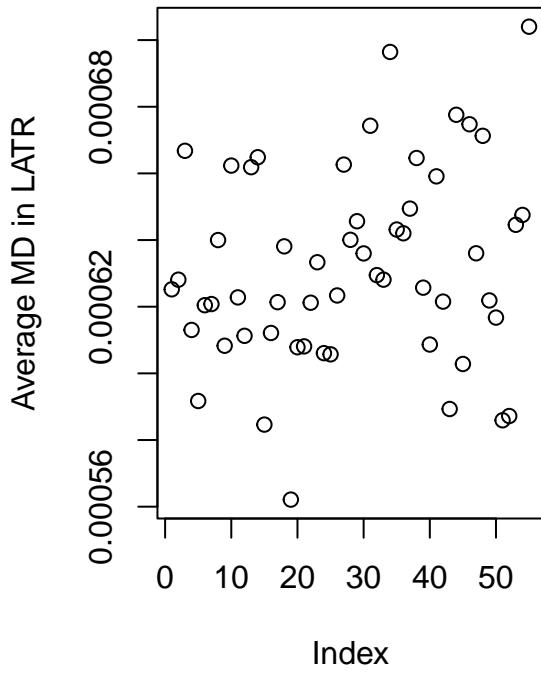
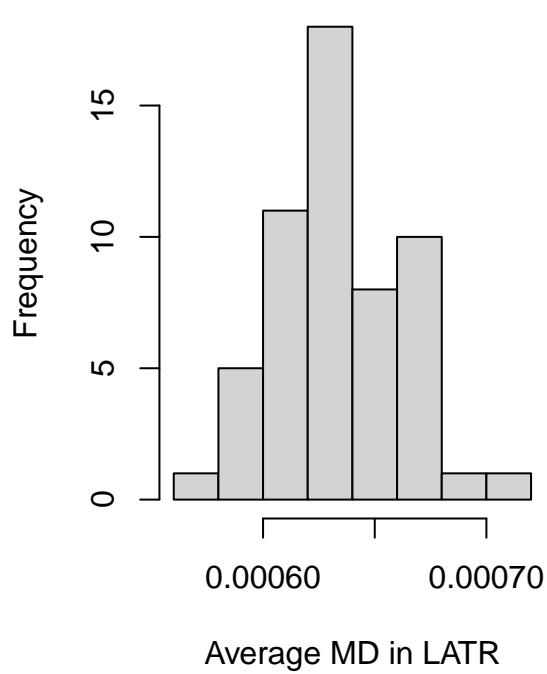


```
[1] 17 9
```



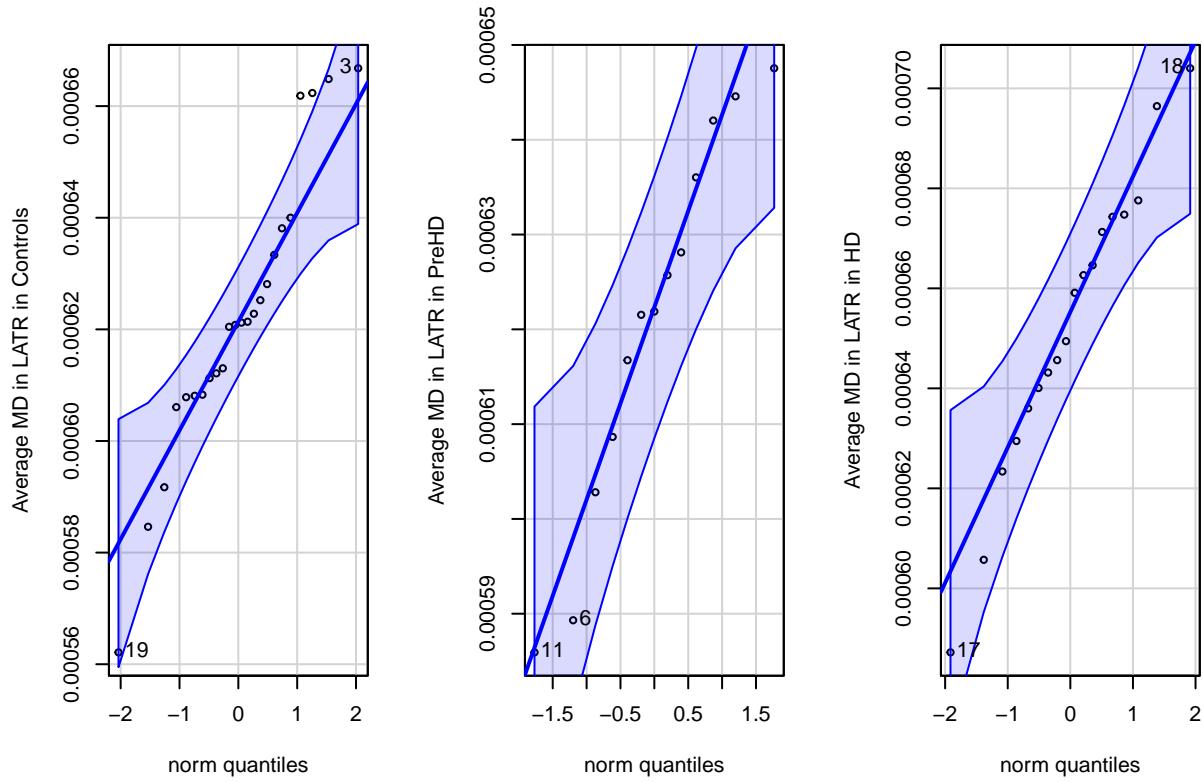
#### 2.3.2.6 MD - LATR

Histogram of LATR Average MD

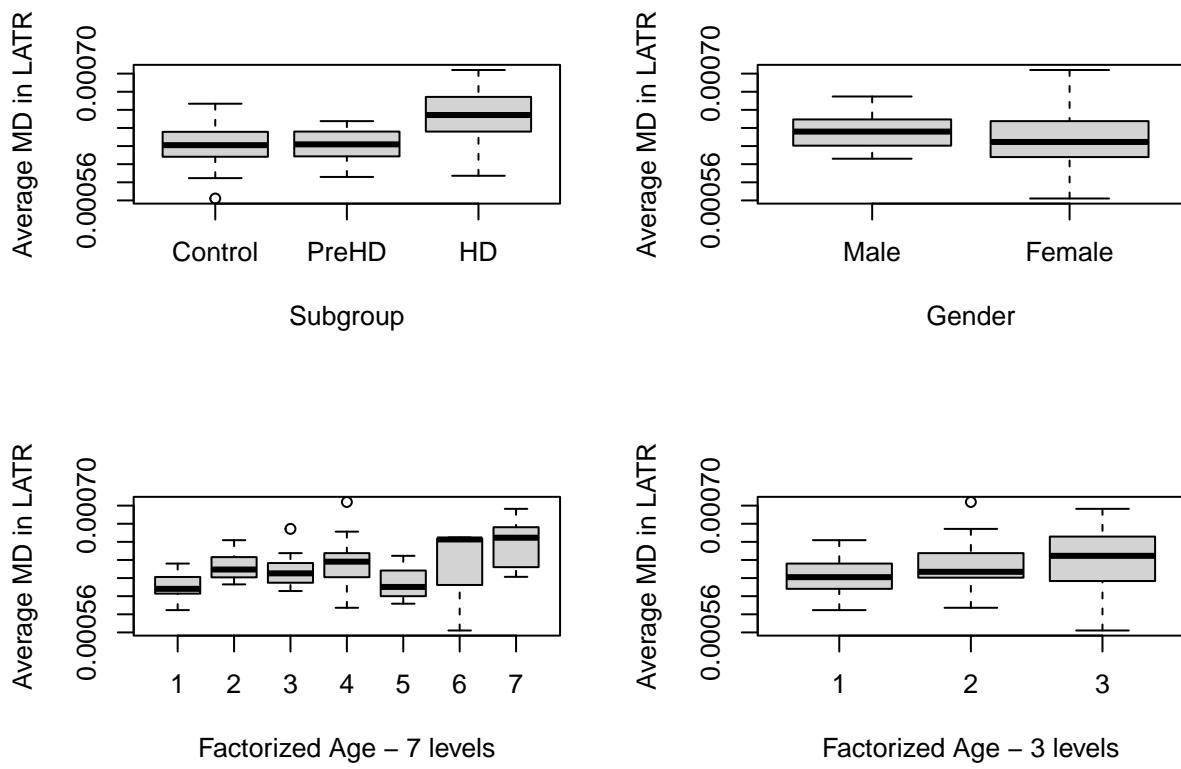


[1] 55 19

```
[1] 19 3  
[1] 11 6
```

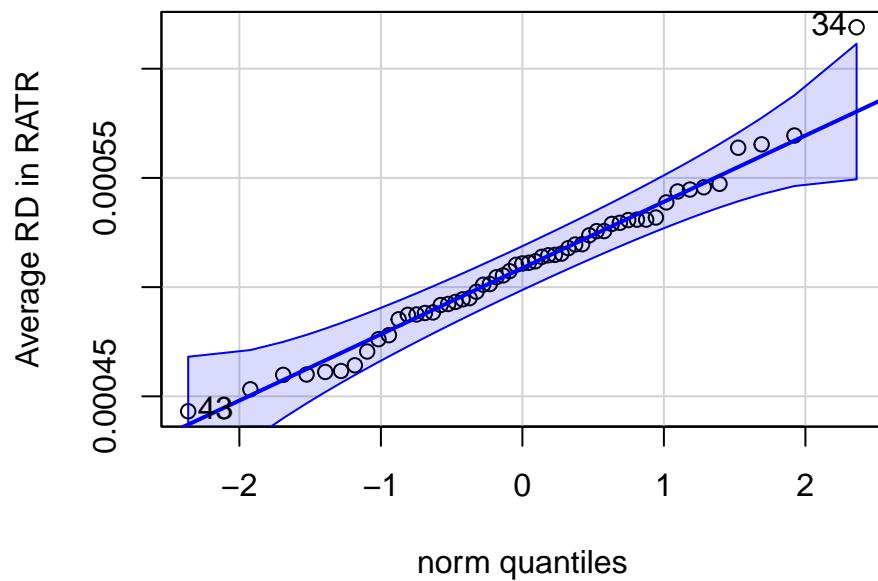
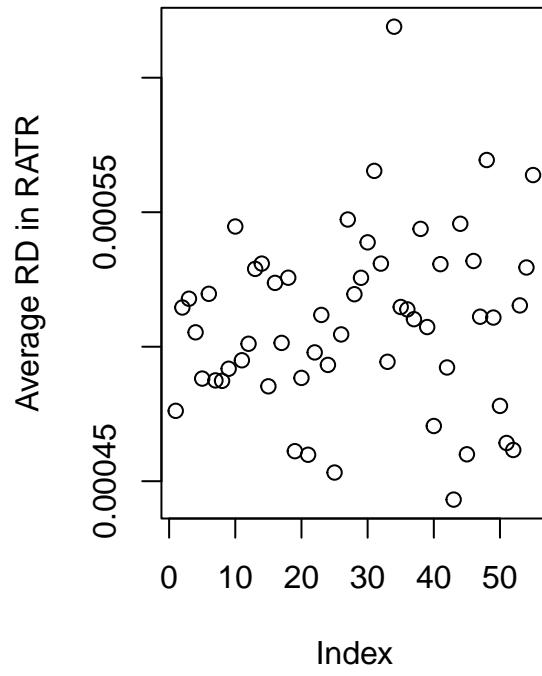
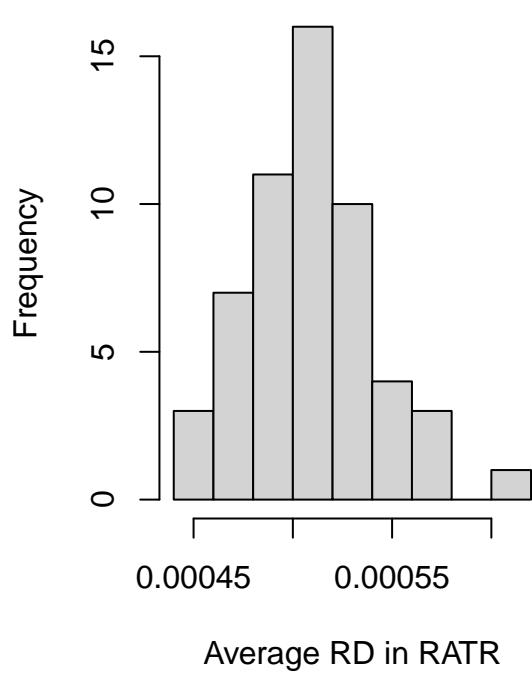


```
[1] 17 18
```



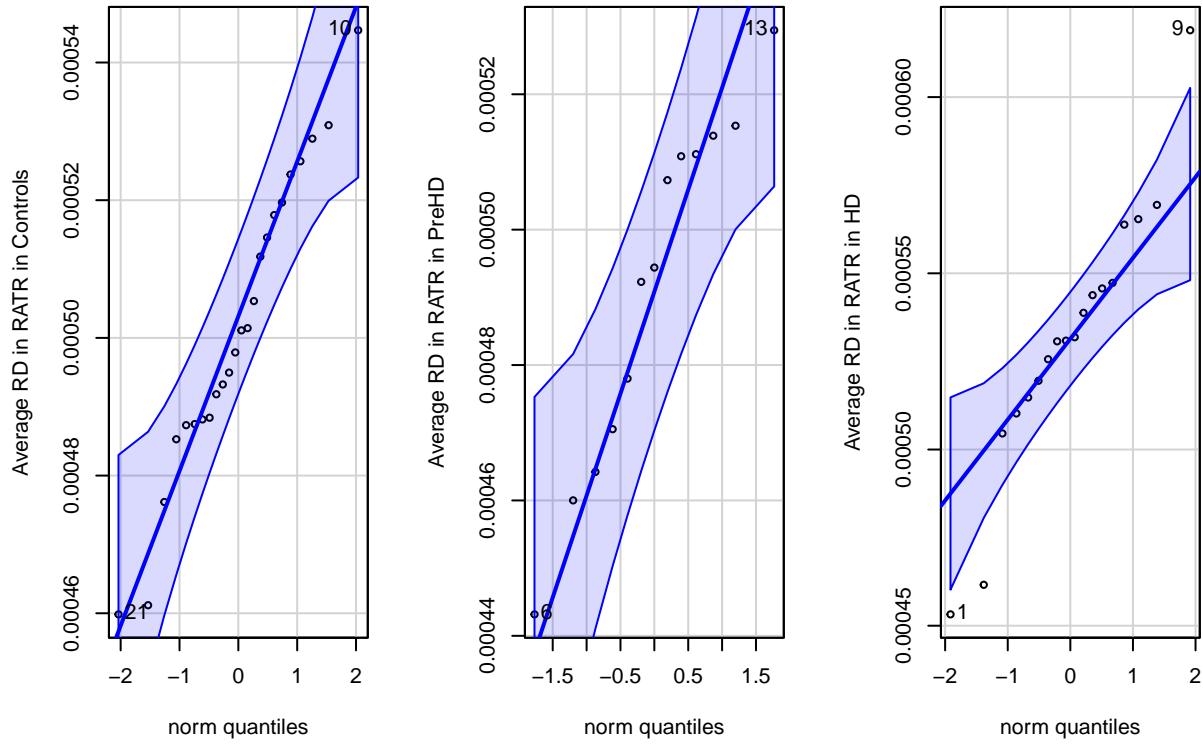
#### 2.3.2.7 RD - RATR

Histogram of RATR Average RD

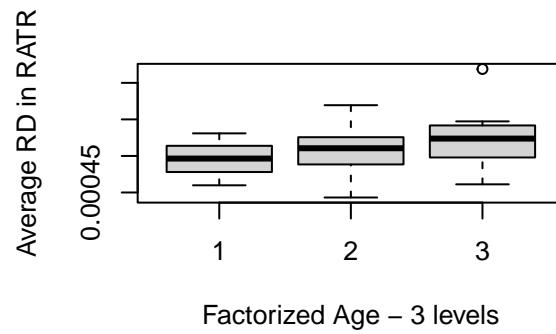
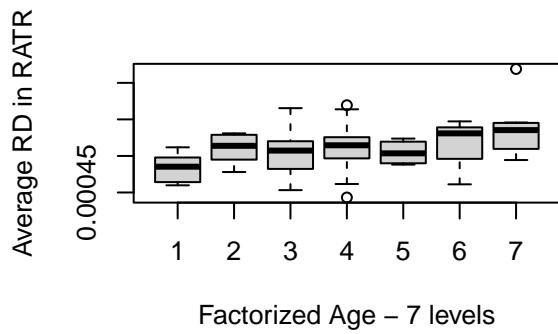
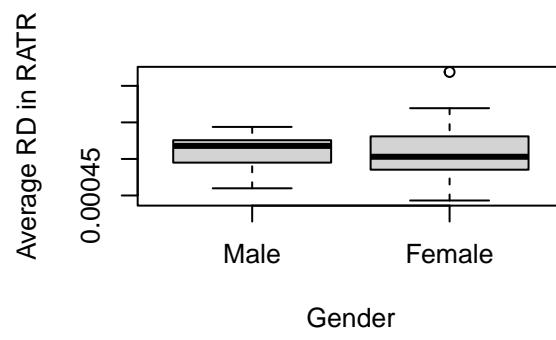
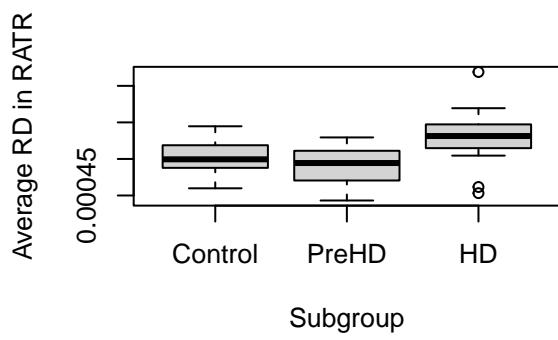


[1] 34 43

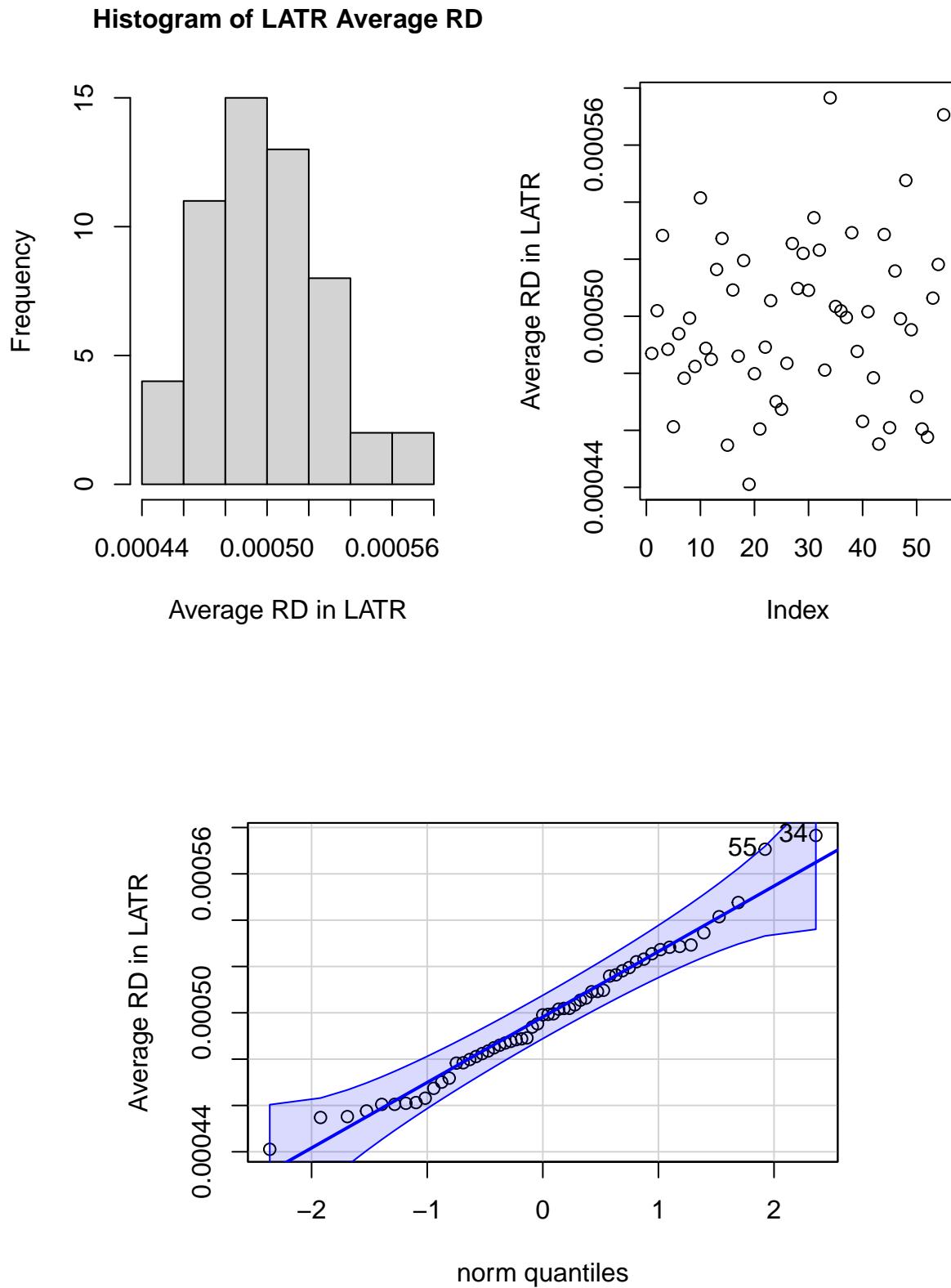
```
[1] 10 21  
[1] 6 13
```



```
[1] 9 1
```

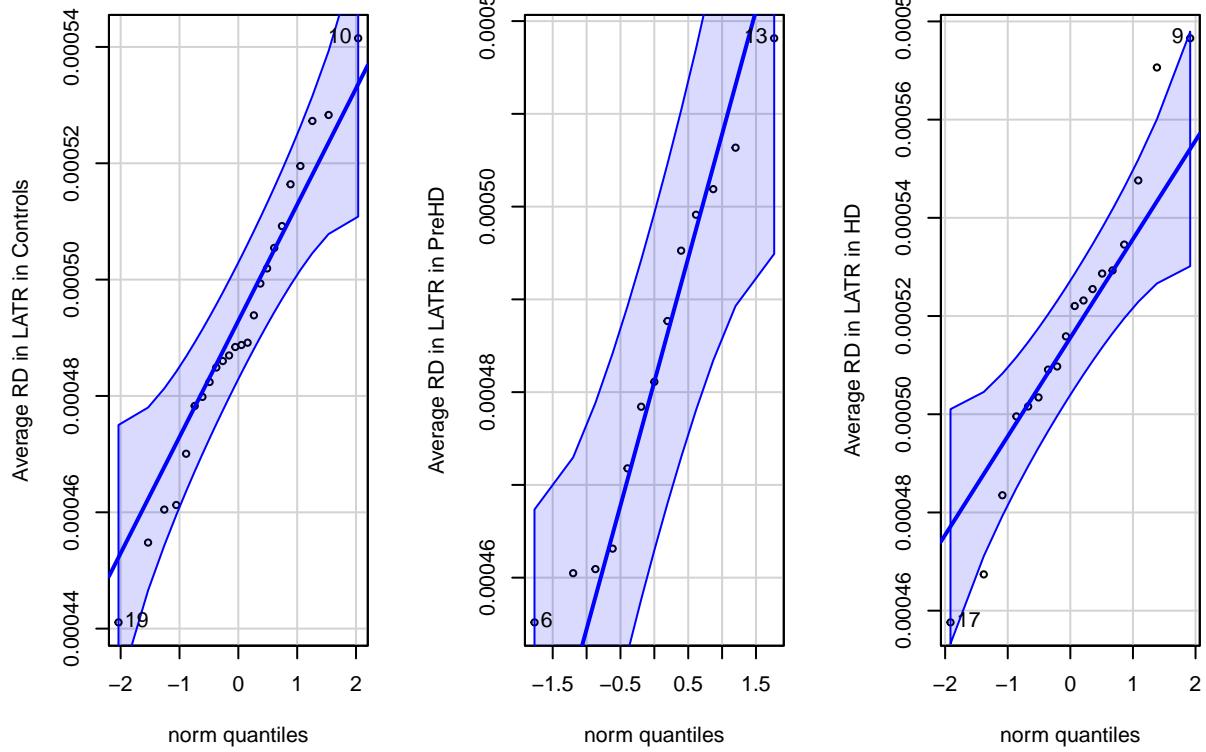


#### 2.3.2.8 RD - LATR

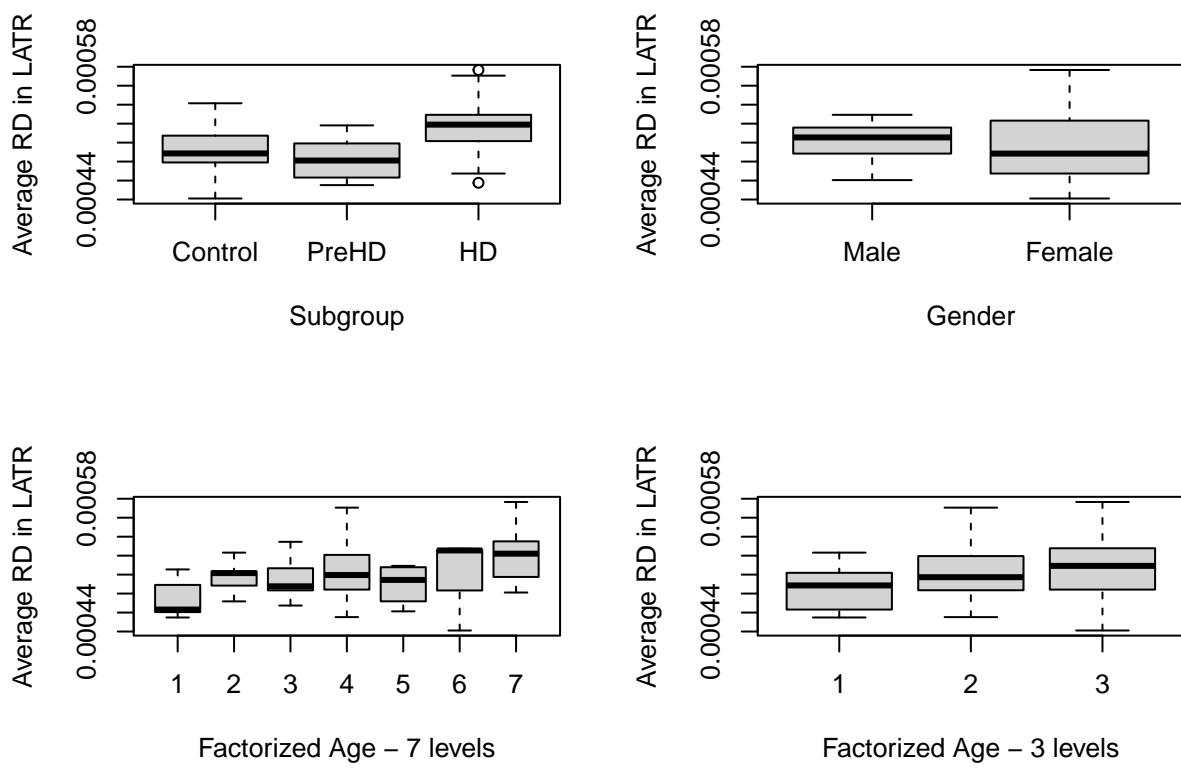


[1] 34 55

```
[1] 19 10  
[1] 13 6
```

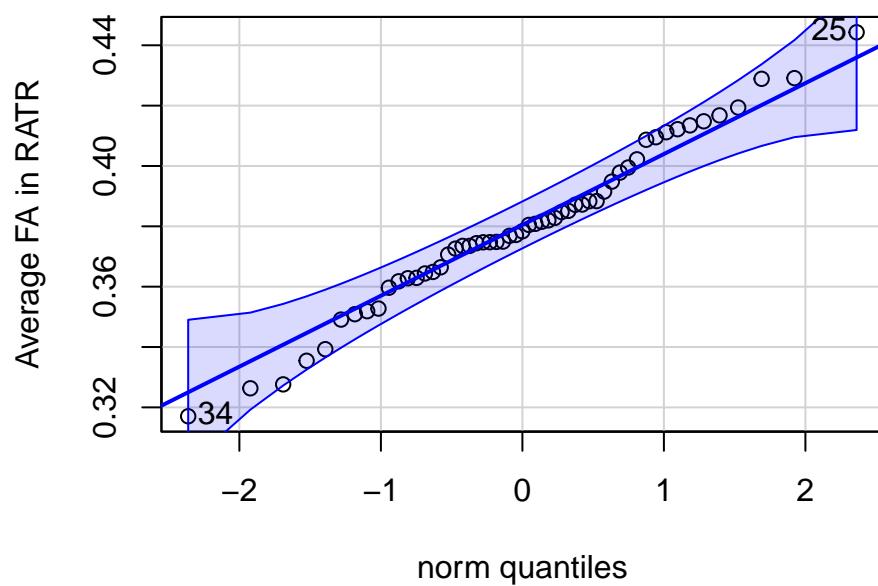
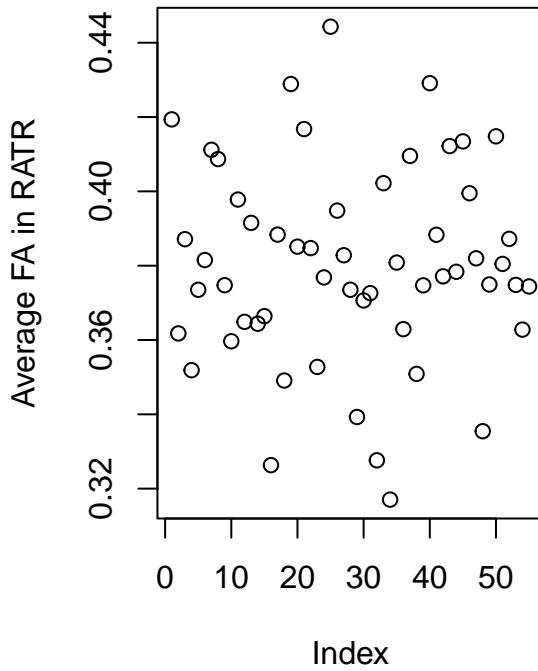
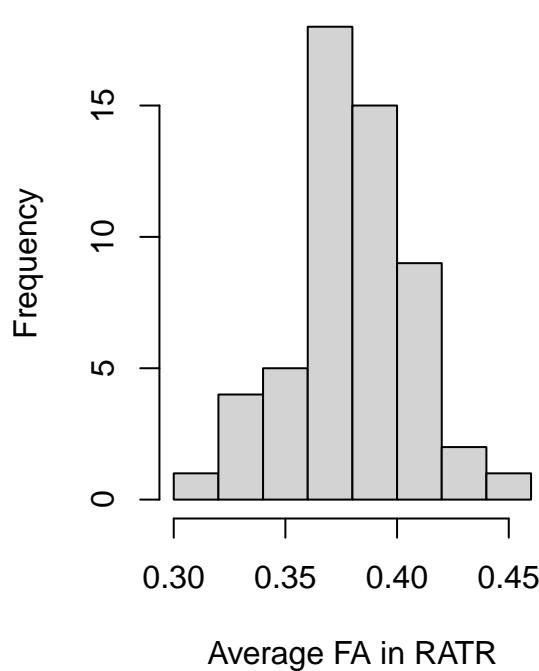


```
[1] 9 17
```



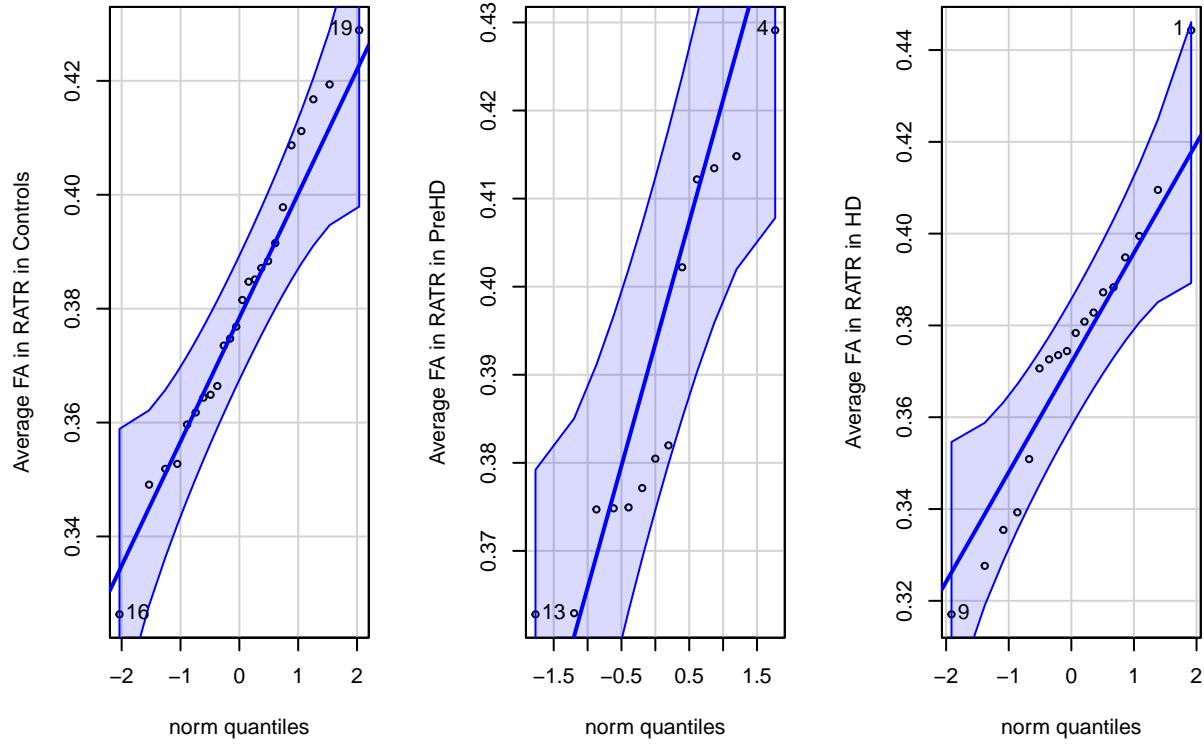
#### 2.3.2.9 FA - RATR

Histogram of RATR Average FA

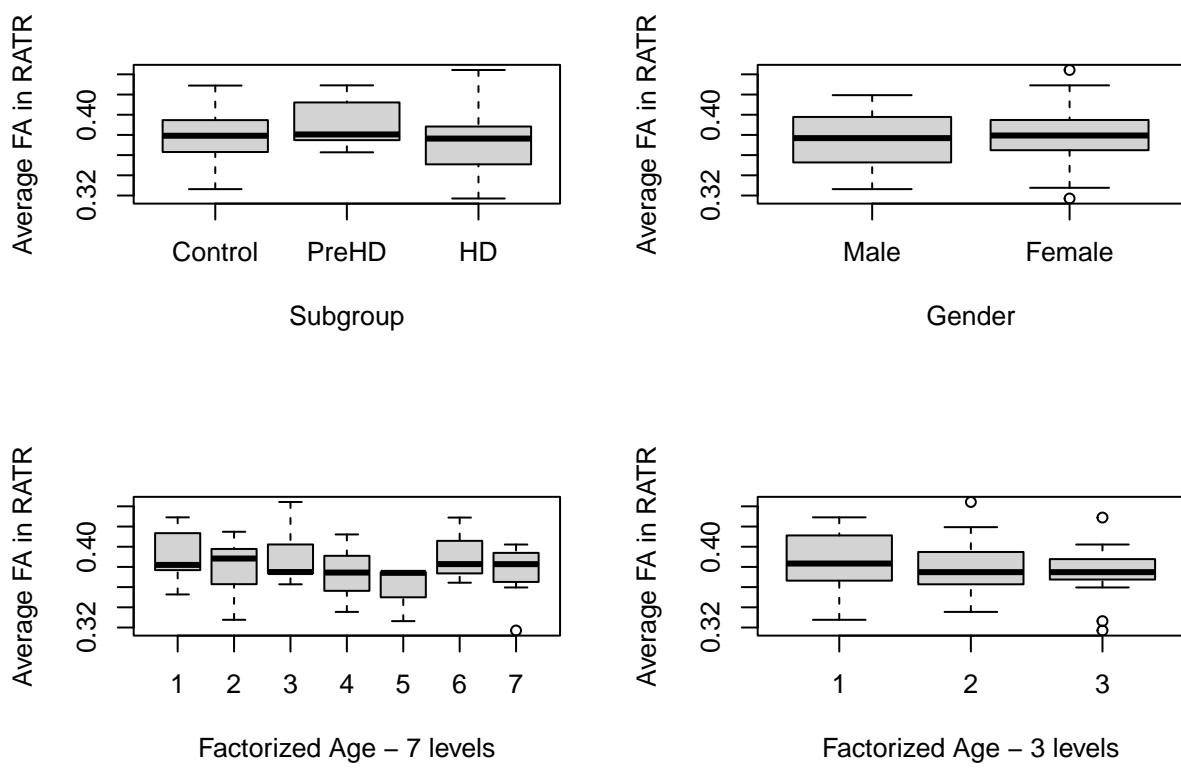


[1] 25 34

```
[1] 16 19  
[1] 4 13
```

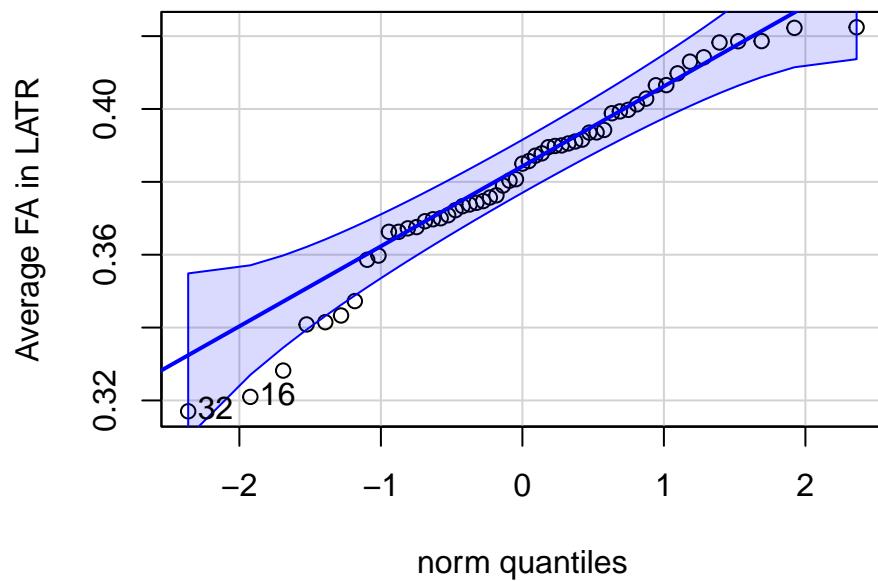
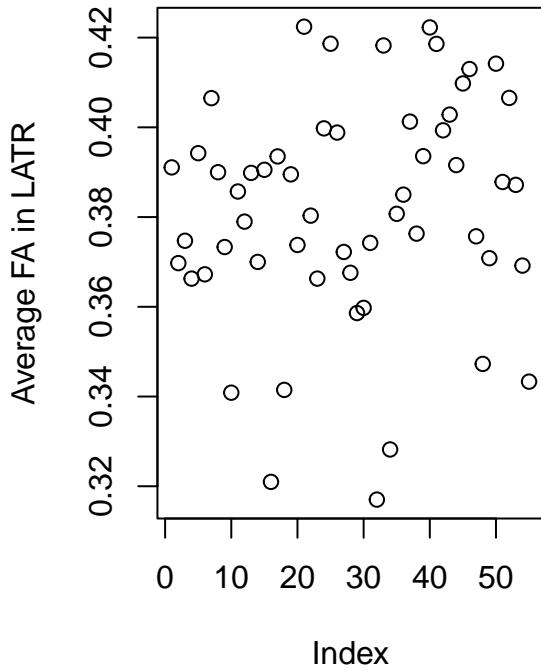
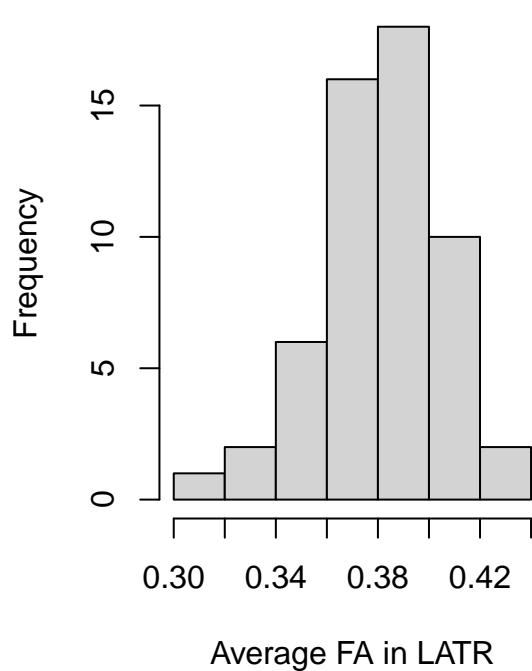


```
[1] 1 9
```



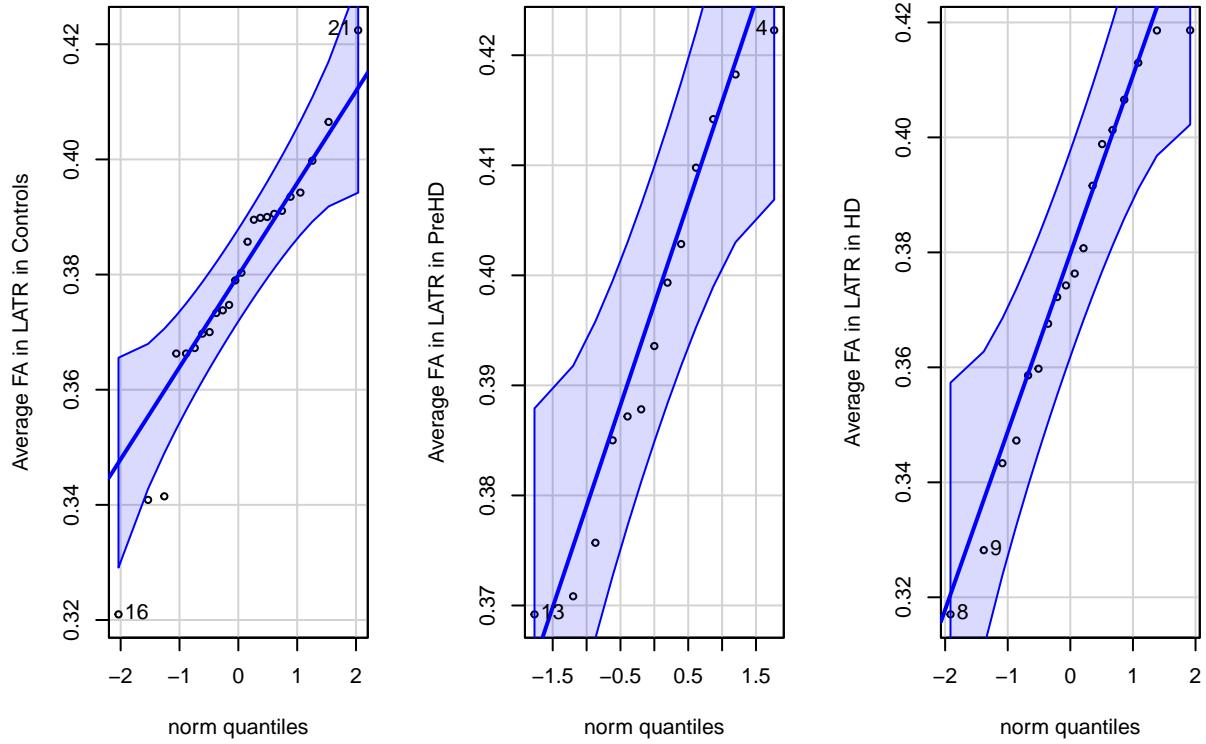
#### 2.3.2.10 FA - LATR

Histogram of LATR Average FA

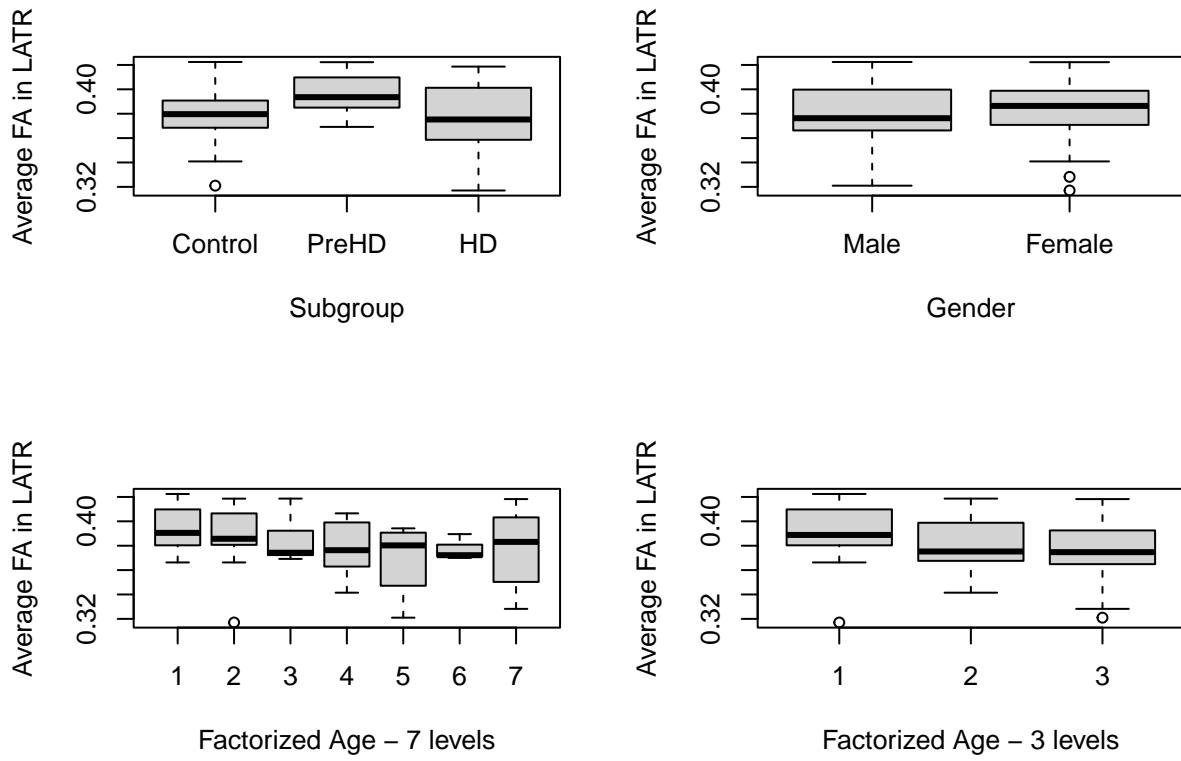


[1] 32 16

```
[1] 16 21  
[1] 4 13
```



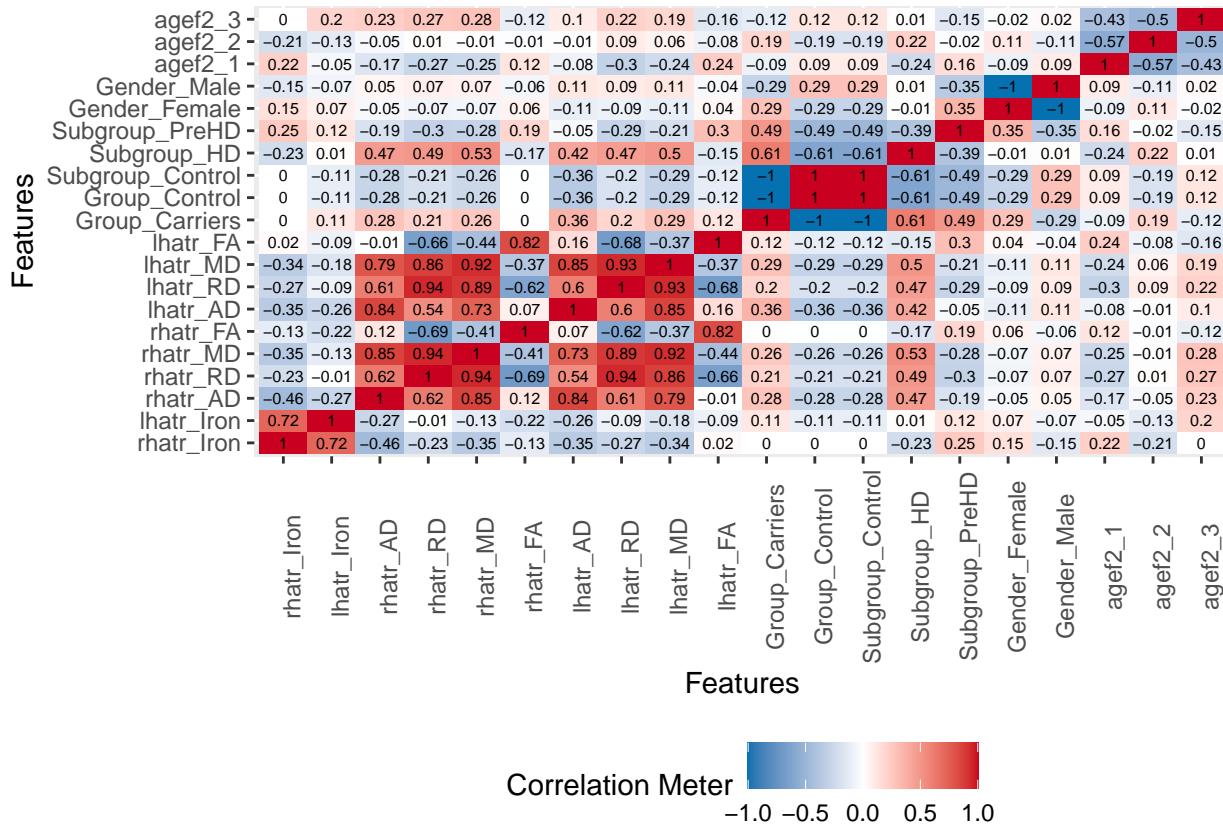
```
[1] 8 9
```



Con estos gráficos, globalmente se observa que los sujetos más atípicos para preHD son los índices 6 y 11 y para HD son los índices 9 y 17. Sin embargo, es importante remarcar que los posibles sujetos outliers varían de métrica a métrica, hecho que dificultaría generar muestras homogéneas en los modelos estadísticos de las diferentes métricas, fenómeno que conceptualmente no me agrada. Además es una muestra con pocos sujetos y eliminar sujetos reduciría potencia estadística.

### 2.3.3 PLOT de CORRELACIONES

Previamente, a generar los diferentes modelos estadísticos, se genera un plot de correlaciones entre todas las variables.



En este plot, se objetivan los siguientes fenómenos:

- Las métricas AD, RD y MD están altamente correlacionadas tanto en el mismo hemisferio como con las métricas del hemisferio contralateral;
- Las métricas de FA entre ambos hemisferios están altamente correlacionadas;
- Las métricas de hierro entre ambos hemisferios están altamente correlacionadas;
- El subgrupo preHD se correlaciona con las métricas de hierro y FA (tanto del hemisferio derecho como del hemisferio izquierdo), con el género femenino y con los grupos de edad 1 (de los 3 niveles) y 1 y 3 (de los 7 niveles);
- El subgrupo HD se correlaciona con las métricas AD, RD y MD (de ambos hemisferios), con los grupos de edad 2 (de los 3 niveles) y 4 y 7 (de los 7 niveles);
- El género femenino se correlaciona con las métricas de hierro y FA (de ambos hemisferios), con el grupo carriers, el subgrupo PreHD, y con los grupos de edad 2 (de los 3 niveles) y 3 y 6 (de los 7 niveles);
- El género masculino se correlaciona con las métricas de AD, RD y MD (de ambos hemisferios), con el grupo y el subgrupo Control, y con los grupos de edad 1 (de los 3 niveles) y 2 y 5 (de los 7 niveles);
- El subgrupo de edad 3 (de los 3 niveles) se correlaciona con hierro del hemisferio izquierdo, AD, RD y MD de ambos hemisferios, con el grupo y el subgrupo Control;
- El subgrupo de edad 1 (de los 3 niveles) se correlaciona con hierro del hemisferio derecho, FA de ambos hemisferios, con los subgrupos Control y preHD y con el género masculino.

Habiendo expuesto lo anterior, se objetiva que tanto género como edad son covariables o factores confusores en la relación entre las métricas y los subgrupos, por tanto dichas variables se van a introducir en los modelos para controlar su efecto. Se utilizará la edad factorizada de 3 niveles dado que presenta una n por nivel más comparable entre niveles.

## 2.4 Estadística descriptiva matemática

### 2.4.1 Media, mediana, desviación estándar, kurtosis y skewness

Se procede a determinar la **media**, la **mediana**, la **desviación estándar**, la **kurtosis** y el **skewness** (estas dos últimas medidas pueden calcularse mediante tres librerías `e1071`, `moments` y `agricolae` y se ha consultado la siguiente web para su interpretación Skewness and Kurtosis in Statistics).

En función de los valores de *skewness*, una distribución puede ser:

- **Simétrica**, con valores de *skewness* entre -0.5 y 0.5, con la media y la mediana siendo similares entre sí;
- **Asimétrica positiva** (*right-skewed* o *right-tailed*), cuyo histograma muestra una cola derecha más larga concentrándose la mayoría de las observaciones en la cola izquierda y con una media mayor que la mediana.
- **Asimétrica negativa** (*left-skewed* o *left-tailed*), cuyo histograma muestra una cola izquierda más larga concentrándose la mayoría de las observaciones en la cola derecha y con una mediana mayor que la media.

La asimetría se gradará en función de los valores de *skewness* como moderada con valores entre 0.5 y 1 o -0.5 y -1 y como alta/severa con valores mayores a 1 o inferiores a -1.

En función de los valores de *kurtosis*, una distribución puede ser:

- **Mesokúrtica**, en el caso de la distribución normal, con valores de kurtosis próximos a 3;
- **Leptokúrtica** (*kurtosis* positiva con valores mayores de 3), con una distribución concentrada de valores generando colas gruesas y pico afilado;
- **Platikúrtica** (*kurtosis* negativa con valores menores de 3), con una distribución que muestra un pico más bajo y ancho y colas más delgadas.

Cuadro 20: Media, mediana, desviación estándar, skewness y kurtosis

Variables	Iron		AD		MD		RD		FA	
	RATR	LATR	RATR	LATR	RATR	LATR	RATR	LATR	RATR	LATR
Mean	0.02359	0.02454	0.00092	9e-04	0.00065	0.00063	0.00051	5e-04	0.38022	0.38158
Median	0.02354	0.02458	0.00092	9e-04	0.00064	0.00063	0.00051	5e-04	0.37836	0.385
Standard Deviation	0.00141	0.00196	4e-05	4e-05	3e-05	3e-05	3e-05	3e-05	0.02676	0.02498
Skewness	1.34597	-0.04429	0.16977	0.04384	0.46919	0.20533	0.49439	0.43044	-0.02574	-0.54421
Kurtosis	5.87476	0.19307	-0.39719	-0.22622	0.26177	-0.24064	0.86246	-0.03208	-0.11756	-0.01365

Los valores de **hierro en el ATR derecho** presentan una asimetría marcada/severa positiva con valores de *skewness* mayores de 1 (exactamente valor de 1.35) y son leptokúrticos (*kurtosis* positiva con valores superiores a 3, exactamente valor de 5.87).

Los valores de **hierro en el ATR izquierdo** presentan una simetría con valores de *skewness* entre -0.5 y 0.5 (exactamente valor de -0.04) y son platikúrticos (*kurtosis* negativa con valores inferiores a 3, exactamente valor de 0.19).

Los valores de **difusividad axial en el ATR derecho** presentan una simetría con valores de *skewness* entre -0.5 y 0.5 (exactamente valor de 0.17) y son platikúrticos (*kurtosis* negativa con valores inferiores a 3, exactamente valor de -0.4).

Los valores de **difusividad axial en el ATR izquierdo** presentan una simetría con valores de *skewness* entre -0.5 y 0.5 (exactamente valor de 0.04) y son platikúrticos (*kurtosis* negativa con valores inferiores a 3, exactamente valor de -0.23).

Los valores de **difusividad media en el ATR derecho** presentan una simetría con valores de *skewness* entre -0.5 y 0.5 (exactamente valor de 0.47) y son platikúrticos (*kurtosis* negativa con valores inferiores a 3, exactamente valor de 0.26).

Los valores de **difusividad media en el ATR izquierdo** presentan una simetría con valores de *skewness* entre -0.5 y 0.5 (exactamente valor de 0.21) y son platikúrticos (*kurtosis* negativa con valores inferiores a 3, exactamente valor de -0.24).

Los valores de **difusividad radial en el ATR derecho** presentan una simetría con valores de *skewness* entre -0.5 y 0.5 (exactamente valor de 0.49) y son platikúrticos (*kurtosis* negativa con valores inferiores a 3, exactamente valor de 0.86).

Los valores de **difusividad radial en el ATR derecho** presentan una simetría con valores de *skewness* entre -0.5 y 0.5 (exactamente valor de 0.43) y son platikúrticos (*kurtosis* negativa con valores inferiores a 3, exactamente valor de -0.03).

Los valores de **anisotropía fraccional en el ATR derecho** presentan una simetría con valores de *skewness* entre -0.5 y 0.5 (exactamente valor de -0.03) y son platikúrticos (*kurtosis* negativa con valores inferiores a 3, exactamente valor de -0.12).

Los valores de **anisotropía fraccional en el ATR izquierdo** presentan una asimetría moderada negativa con valores de *skewness* entre -0.5 y -1.0 (exactamente valor de -0.54) y son platikúrticos (*kurtosis* negativa con valores inferiores a 3, exactamente valor de -0.01).

#### 2.4.2 Normalidad Univariante - Homocedasticidad de cada métrica para cada factor

Se procede a **valorar la normalidad** de los valores de las diferentes métricas de RM **y la homocedasticidad** de dichas métricas en los niveles de las variables categóricas de interés (subgrupo, edad\_factorizada, género).

La normalidad se puede obtener de una de las dos funciones siguientes: `shapiro.test` del paquete `stats` o `shapiro_test` del paquete `rstatix`. Los resultados son coincidentes.

La homocedasticidad se obtiene de la función `fligner.test` del paquete `stats`.

Cuadro 21: Valoración de la normalidad de las métricas de RM con la función shapiro test

	Iron		AD		MD		RD		FA	
	RATR	LATR								
W-Statistic	0.8860	0.9885	0.9758	0.9924	0.9755	0.9860	0.9730	0.9773	0.9854	0.9622
p-value	0.0001	0.8755	0.3312	0.9790	0.3204	0.7694	0.2507	0.3822	0.7395	0.0816

Cuadro 22: Valoración de la homocedasticidad de las métricas de RM para la variable subgrupo

	Iron		AD		MD		RD		FA	
	RATR	LATR								
Statistic	2.0133	1.4370	7.0202	5.0139	5.6891	2.4385	1.5150	0.9829	0.8320	3.9997
df	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000
p-value	0.3654	0.4875	0.0299	0.0815	0.0582	0.2955	0.4688	0.6117	0.6597	0.1354

Cuadro 23: Valoración de la homocedasticidad de las métricas de RM para la variable edad\_factorizada con 7 niveles

	Iron		AD		MD		RD		FA	
	RATR	LATR								
Statistic	7.3695	4.9139	3.8852	2.3624	2.9386	2.5454	2.1029	2.6573	0.6316	5.7039
df	6.0000	6.0000	6.0000	6.0000	6.0000	6.0000	6.0000	6.0000	6.0000	6.0000
p-value	0.2880	0.5549	0.6922	0.8835	0.8165	0.8634	0.9100	0.8505	0.9959	0.4572

Cuadro 24: Valoración de la homocedasticidad de las métricas de RM para la variable edad factorizada con 3 niveles

	Iron		AD		MD		RD		FA	
	RATR	LATR								
Statistic	0.3657	1.8066	0.3362	3.3915	2.9109	4.2921	0.6548	1.0952	0.0739	1.0465
df	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000
p-value	0.8329	0.4052	0.8453	0.1835	0.2333	0.1169	0.7208	0.5783	0.9637	0.5926

Cuadro 25: Valoración de la homocedasticidad de las métricas de RM para la variable género

	Iron		AD		MD		RD		FA	
	RATR	LATR								
Statistic	1.7764	0.0011	0.0527	0.0162	1.3061	2.4105	5.2712	4.6815	0.0185	0.0913
df	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
p-value	0.1826	0.9736	0.8184	0.8988	0.2531	0.1205	0.0217	0.0305	0.8917	0.7626

Los valores de **hierro en el ATR derecho** no son normales, pero presentan homocedasticidad de varianza entre los niveles de las variables categóricas de interés (subgrupo, edad\_factorizada, género).

Los valores de **hierro en el ATR izquierdo** son normales, con homocedasticidad de varianza entre los niveles de las variables categóricas de interés (subgrupo, edad\_factorizada, género).

Los valores de **difusividad axial en el ATR derecho** son normales, con homocedasticidad de varianza entre los niveles de las variables categóricas edad\_factorizada y género, pero con heterocedasticidad de varianza entre subgrupos (p valor de 0.03). Posteriormente, en la generación del modelo, se valorará el comportamiento de los residuos para decidir sobre la eliminación de outliers.

Los valores de **difusividad axial en el ATR izquierdo** son normales, con homocedasticidad de varianza entre los niveles de las variables categóricas de interés (subgrupo, edad\_factorizada, género).

Los valores de **difusividad media en el ATR derecho** son normales, con homocedasticidad de varianza entre los niveles de las variables categóricas de interés (subgrupo, edad\_factorizada, género).

Los valores de **difusividad media en el ATR izquierdo** son normales, con homocedasticidad de varianza entre los niveles de las variables categóricas de interés (subgrupo, edad\_factorizada, género).

Los valores de **difusividad radial en el ATR derecho** son normales, con homocedasticidad de varianza entre los niveles de las variables categóricas subgrupo y edad\_factorizada, pero con heterocedasticidad de varianza entre géneros (p valor de 0.022). Posteriormente, en la generación del modelo, se valorará el comportamiento de los residuos para decidir sobre la eliminación de outliers.

Los valores de **difusividad radial en el ATR izquierdo** son normales, con homocedasticidad de varianza entre los niveles de las variables categóricas subgrupo y edad\_factorizada, pero con heterocedasticidad de varianza entre géneros (p valor de 0.03). Posteriormente, en la generación del modelo, se valorará el comportamiento de los residuos para decidir sobre la eliminación de outliers.

Los valores de **anisotropía fraccional en el ATR derecho** son normales, con homocedasticidad de varianza entre los niveles de las variables categóricas de interés (subgrupo, edad\_factorizada, género).

Los valores de **anisotropía fraccional en el ATR izquierdo** son normales, con homocedasticidad de varianza entre los niveles de las variables categóricas de interés (subgrupo, edad\_factorizada, género).

**Por tanto, de todas las métricas, la variable rhatr-iron es la única sin distribución normal, las variables rhatr-RD y lhatr-RD presentan heterocedasticidad para género, y la variable rhatr-AD presenta heterocedasticidad para subgrupo.**

Se intentará trabajar con todos los sujetos en los diferentes modelos porque mi interés es tener muestras homogéneas entre las diferentes métricas. En el caso que no se cumplan los criterios para aplicar métodos paramétricos, se procederá a generar métodos no paramétricos. De todas formas, se generarán ambos métodos para todas las métricas dado que también es de mi interés que los métodos sean homogéneos entre todas las métricas.

### 2.4.3 Normalidad multivariante

Se consultan las siguientes fuentes de información: Package ‘MVN’, Testing for Multivariate Normality, How to Perform Multivariate Normality Tests in R.

La función `mvn` del paquete MVN genera tests de normalidad multivariante. Dicha función tiene diferentes argumentos, de los cuales detallo los que he utilizado:

- `data`: especificar el dataframe o la matriz numérica;
- `mvnTest`: existen diferentes posibilidades; yo he elegido el test de Royston de normalidad multivariante. Este test calcula ya sea el estadístico W del Shapiro-Wilk para muestras platikúrticas o el Shapiro-Francia para muestras leptokúrticas;
- `covariance`: para las opciones de `mardia` o `royston` en `mvnTest`; TRUE para normalizar la matriz de covarianza por `n` o FALSE para normalizarla por `n-1`;
- `desc`: TRUE para calcular los estadísticos descriptivos;
- `univariateTest`: se selecciona el test de normalidad univariada de interés;
- `univariatePlot`: se seleccionan los plots de normalidad univariada;
- `multivariatePlot`: se selecciona los plots de normalidad multivariada;
- `multivariateOutlierMethod`: se selecciona uno de los dos métodos de detección de outliers multivariados ('`quan`' es el método cuantil por defecto, '`adj`' es el método cuantil ajustado), ambos métodos basados en la distancia de Mahalanobis;
- `showOutliers`: TRUE para mostrar los outliers multivariados;
- `showNewData`: TRUE para mostrar los datos sin outliers.

El output es una lista con los siguientes argumentos:

- `multivariateNormality`: dataframe especificando el estadístico de normalidad multivariante y su p-valor;
- `univariateNormality`: dataframe con el estadístico de normalidad univariante y su p-valor para cada métrica de RM;
- `Descriptives`: dataframe con los diferentes estadísticos descriptivos (media, mediana, desviación estándar, mínimo, máximo, cuartil 25, cuartil 75, skew y kurtosis);
- `multivariateOutliers`: dataframe con la distancia de Mahalanobis para cada outlier;
- `newData`: un dataframe con los nuevos datos habiendo eliminado los outliers.

Se va testear la normalidad multivariante en 3 supuestos:

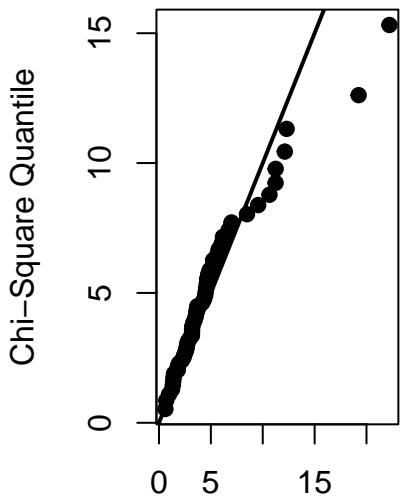
- Las 5 métricas de RM conjuntamente;
- Las 3 métricas de RM (AD, RD, MD) que parecen correlacionarse con el subgrupo HD;
- Las 2 métricas de RM (iron, FA) que parecen correlacionarse con el subgrupo preHD.

Resultados para el RATR:

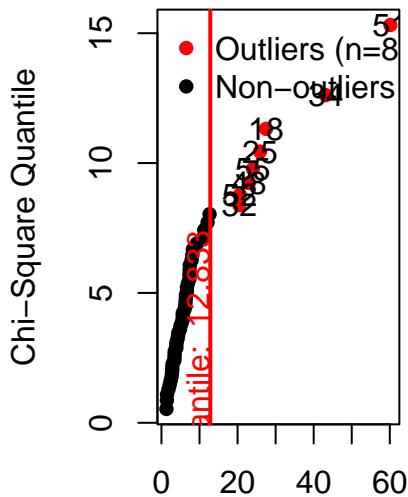
```
# Para las 5 métricas de RM conjuntamente
library(MVN)
par(mfrow = c(1, 2))
#RATR_avg<-ATR_avg_demo[,c(3,8, 10:13)]
RATR_avg<-ATR_avg_demo[,c('Subgroup','rhatr_Iron','rhatr_AD','rhatr_RD',
                           'rhatr_MD','rhatr_FA')]

# Valoración de la normalidad multivariante con todos los sujetos
RATR_mvnb<-mvn(RATR_avg[-1], mvnTest = 'royston', covariance = TRUE, desc = TRUE,
                 univariateTest = 'SW', univariatePlot = 'qq',
                 multivariatePlot = 'qq', multivariateOutlierMethod = 'quan',
                 showOutliers = TRUE, showNewData = TRUE)
```

Chi-Square Q-Q Plot



Chi-Square Q-Q Plot



Squared Mahalanobis Distance      Robust Squared Mahalanobis Distance

```

index<-rownames(RATR_mvnm$newData)
RATR_avg_WoOut<-RATR_avg[c(index),]
table(RATR_avg$Subgroup)

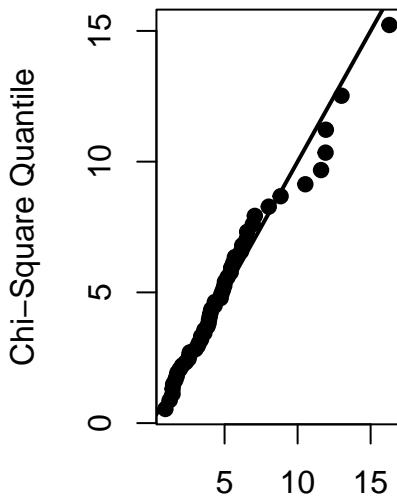
Control    PreHD      HD
      24        13       18
table(RATR_avg_WoOut$Subgroup)

Control    PreHD      HD
      23        12       12
  
```

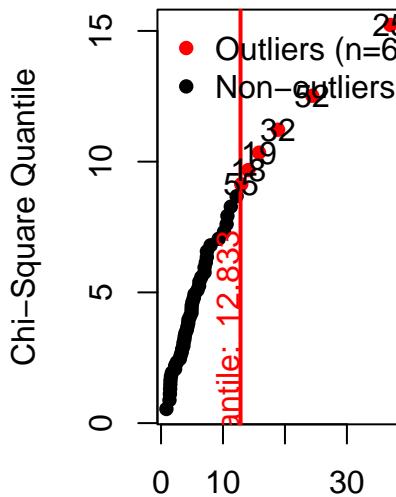
```

# Valoración de la normalidad multivariante con eliminación de los dos outliers
# más extremos
par(mfrow = c(1, 2))
print(RATR_mvnm$multivariateOutliers,max = 6)
  Observation Mahalanobis Distance Outlier
51           51          60.163   TRUE
34           34          43.182   TRUE
[ reached 'max' / getOption("max.print") -- omitted 6 rows ]
newRATR_avg<-RATR_avg[-c(34,51),]
RATR_avg[c(34,51),]
  Subgroup rhatr_Iron    rhatr_AD    rhatr_RD    rhatr_MD rhatr_FA
34       HD  0.0247215 0.000989331 0.000618966 0.00074242 0.317049
51     PreHD  0.0299576 0.000826666 0.000464213 0.00058503 0.380471
newRATR_mvnm<-mvn(newRATR_avg[-1], mvnTest = 'royston', covariance = TRUE,
                     desc = TRUE, univariateTest = 'SW', univariatePlot = 'qq',
                     multivariatePlot = 'qq', multivariateOutlierMethod = 'quan',
                     showOutliers = TRUE, showNewData = TRUE)
  
```

**Chi-Square Q-Q Plot**



**Chi-Square Q-Q Plot**



Squared Mahalanobis Distance Robust Squared Mahalanobis Distance

```
RATR_avg_demo<-ATR_avg_demo[-c(34,51),-c(9,14:17)]
#RATR_avg_demo_WoOut<-ATR_avg_demo[-c(34,51),-c('lhatr_Iron','lhatr_AD','lhatr_RD',
#                                         'lhatr_MD','lhatr_FA')]
RATR_avg_demo<-ATR_avg_demo[,-c(9,14:17)]
```

Cuadro 26: Comparación de la normalidad multivariante sin y con eliminación de outliers para el ATR derecho habiendo considerado las 5 métricas conjuntamente

	Test	H	p value	MVN
Without elimination of outliers	Royston	15.451419	0.0023417	NO
With elimination of outliers	Royston	2.118924	0.6607944	YES

Cuadro 27: Comparación de la normalidad univariante sin y con eliminación de outliers para las 5 métricas de ATR derecho

Without elimination of outliers					With elimination of outliers				
Test	Variable	Statistic	p value	Normality	Test	Variable	Statistic	p value	Normality
Shapiro-Wilk	rhatr_Iron	0.8860	0.0001	NO	Shapiro-Wilk	rhatr_Iron	0.9749	0.3253	YES
Shapiro-Wilk	rhatr_AD	0.9758	0.3312	YES	Shapiro-Wilk	rhatr_AD	0.9745	0.3143	YES
Shapiro-Wilk	rhatr_RD	0.9730	0.2507	YES	Shapiro-Wilk	rhatr_RD	0.9866	0.8138	YES
Shapiro-Wilk	rhatr_MD	0.9755	0.3204	YES	Shapiro-Wilk	rhatr_MD	0.9792	0.4792	YES
Shapiro-Wilk	rhatr_FA	0.9854	0.7395	YES	Shapiro-Wilk	rhatr_FA	0.9849	0.7351	YES

Cuadro 28: Comparación de los estadísticos descriptivos sin y con eliminación de outliers para las 5 métricas de ATR derecho

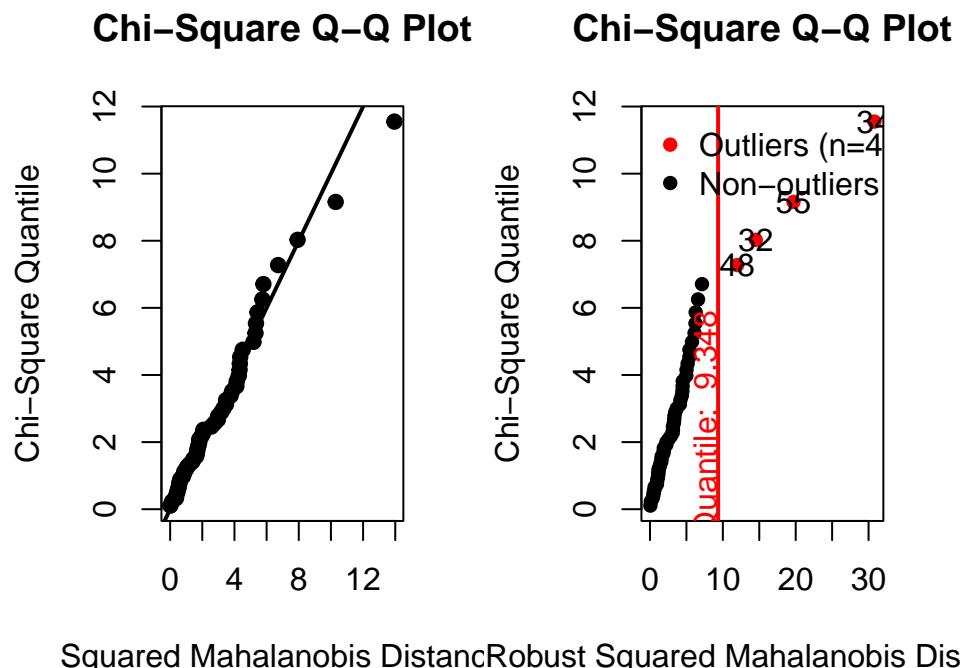
Variables	n	Mean	Std.Dev	Median	Min	Max	25th	75th	Skew	Kurtosis
<b>Without elimination of outliers</b>										

rhatr_Iron	55	0.0235915	0.0014052	0.0235422	0.0201360	0.0299576	0.0227244	0.0244503	1.3459722	5.8747627
rhatr_AD	55	0.0009186	0.0000442	0.0009164	0.0008239	0.0010076	0.0008899	0.0009368	0.1697689	-0.3971935
rhatr_RD	55	0.0005092	0.0000329	0.0005108	0.0004432	0.0006190	0.0004883	0.0005292	0.4943887	0.8624593
rhatr_MD	55	0.0006457	0.0000331	0.0006406	0.0005788	0.0007424	0.0006251	0.0006628	0.4691852	0.2617652
rhatr_FA	55	0.3802175	0.0267579	0.3783580	0.3170490	0.4443100	0.3646285	0.3963130	-0.0257356	-0.1175621
<b>With elimination of outliers</b>										
rhatr_Iron	53	0.0234501	0.0011073	0.0234765	0.0201360	0.0256621	0.0227145	0.0243902	-0.3731448	-0.0380987
rhatr_AD	53	0.0009190	0.0000420	0.0009164	0.0008239	0.0010076	0.0008905	0.0009358	0.2841794	-0.3022323
rhatr_RD	53	0.0005079	0.0000291	0.0005108	0.0004432	0.0005694	0.0004884	0.0005289	-0.0813874	-0.4551572
rhatr_MD	53	0.0006450	0.0000298	0.0006406	0.0005788	0.0007120	0.0006252	0.0006613	0.2478718	-0.2404239
rhatr_FA	53	0.3814046	0.0257944	0.3783580	0.3263210	0.4443100	0.3648810	0.3978020	0.1117419	-0.2478463

**Considerando las 5 métricas conjuntamente**, se observa que no existe normalidad multivariante ni normalidad univariante para el hierro. Si se eliminan todos los outliers detectados por la función `mvn`, se elimina un tercio de los pacientes, un sujeto preHD y un sujeto control. Considero no correcto eliminar un tercio de los HD (pacientes sintomáticos), por lo que decido eliminar exclusivamente los dos sujetos (un HD y un preHD) con la mayor distancia de Mahalanobis, y, con ello, se consigue una normalidad multivariante y una normalidad univariante para todas las métricas (previamente a la eliminación de los outliers, el hierro no presentaba normalidad).

```
# Para las 3 métricas de RM correlacionadas con HD: AD, RD, MD
library(MVN)
par(mfrow = c(1, 2))
#RATR_avg<-ATR_avg_demo[,c(3,8, 10:13)]
RATR_avg<-ATR_avg_demo[,c('Subgroup','rhatr_Iron','rhatr_AD','rhatr_RD',
                           'rhatr_MD','rhatr_FA')]

# Valoración de la normalidad multivariante con todos los sujetos
RATR_mvnmv(RATR_avg[-c(1,2,6)], mvnTest = 'royston', covariance = TRUE, desc = TRUE,
            univariateTest = 'SW', univariatePlot = 'qq',
            multivariatePlot = 'qq', multivariateOutlierMethod = 'quan',
            showOutliers = TRUE, showNewData = TRUE)
```



```

index<-rownames(RATR_mvnm$newData)
RATR_avg_WoOut<-RATR_avg[c(index),]
table(RATR_avg$Subgroup)

```

Control	PreHD	HD
24	13	18

```
table(RATR_avg_WoOut$Subgroup)
```

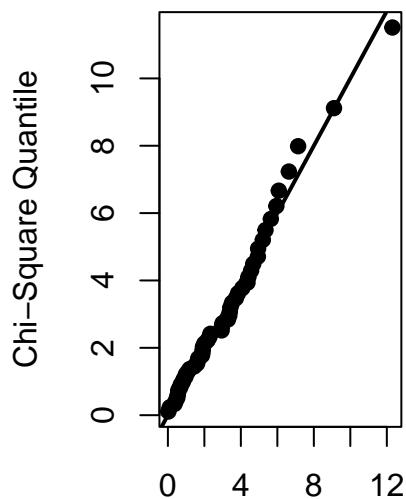
Control	PreHD	HD
24	13	14

```

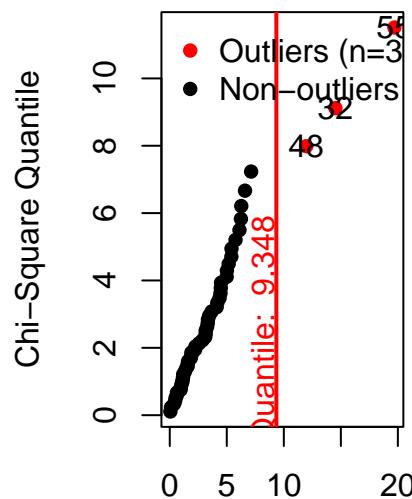
# Valoración de la normalidad multivariante con eliminación del outlier más
# extremo
par(mfrow = c(1, 2))
print(RATR_mvnm$multivariateOutliers,max = 6)
Observation Mahalanobis Distance Outlier
34           34          30.802    TRUE
55           55          19.696    TRUE
[ reached 'max' / getOption("max.print") -- omitted 2 rows ]
newRATR_avg<-RATR_avg[-c(34),]
RATR_avg[c(34),]
Subgroup rhatr_Iron    rhatr_AD    rhatr_RD    rhatr_MD rhatr_FA
34       HD 0.0247215 0.000989331 0.000618966 0.00074242 0.317049
newRATR_mvnm<-mvn(newRATR_avg[-c(1,2,6)], mvnTest = 'royston', covariance = TRUE,
desc = TRUE, univariateTest = 'SW', univariatePlot = 'qq',
multivariatePlot = 'qq', multivariateOutlierMethod = 'quan',
showOutliers = TRUE, showNewData = TRUE)

```

**Chi-Square Q-Q Plot**



**Chi-Square Q-Q Plot**



Squared Mahalanobis Distance      Robust Squared Mahalanobis Dist

```

RATR_avg_demo_WoOut_ADRDMD<-ATR_avg_demo[-c(34),-c(8,9,13:17)]
#RATR_avg_demo_WoOut_ADRDMD<-ATR_avg_demo[-c(34),-c('rhatr_Iron','lhatr_Iron','rhatr_FA','lhatr_AD',

```

```
#'lhatr_RD', 'lhatr_MD', 'lhatr_FA')]

RATR_avg_demo_ADRDMD<-ATR_avg_demo[, -c(8, 9, 13:17)]
```

Cuadro 29: Comparación de la normalidad multivariante sin y con eliminación de outliers para el ATR derecho habiendo considerado las métricas AD, RD, MD conjuntamente

	Test	H	p value	MVN
Without elimination of outliers	Royston	2.8342477	0.2469134	YES
With elimination of outliers	Royston	0.8900952	0.6590264	YES

Cuadro 30: Comparación de la normalidad univariante sin y con eliminación de outliers para las métricas AD, RD, MD del ATR derecho

Without elimination of outliers					With elimination of outliers				
Test	Variable	Statistic	p value	Normality	Test	Variable	Statistic	p value	Normality
Shapiro-Wilk	rhatr_AD	0.9758	0.3312	YES	Shapiro-Wilk	rhatr_AD	0.9766	0.3687	YES
Shapiro-Wilk	rhatr_RD	0.9730	0.2507	YES	Shapiro-Wilk	rhatr_RD	0.9857	0.7657	YES
Shapiro-Wilk	rhatr_MD	0.9755	0.3204	YES	Shapiro-Wilk	rhatr_MD	0.9807	0.5328	YES

Cuadro 31: Comparación de los estadísticos descriptivos sin y con eliminación de outliers para las métricas AD, RD, MD del ATR derecho

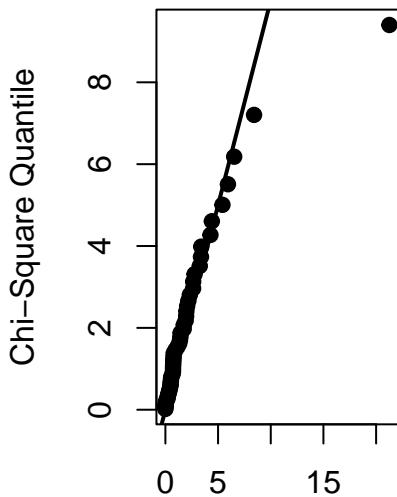
Variables	n	Mean	Std.Dev	Median	Min	Max	25th	75th	Skew	Kurtosis
<b>Without elimination of outliers</b>										
rhatr_AD	55	0.0009186	4.42e-05	0.0009164	0.0008239	0.0010076	0.0008899	0.0009368	0.1697689	-0.3971935
rhatr_RD	55	0.0005092	3.29e-05	0.0005108	0.0004432	0.0006190	0.0004883	0.0005292	0.4943887	0.8624593
rhatr_MD	55	0.0006457	3.31e-05	0.0006406	0.0005788	0.0007424	0.0006251	0.0006628	0.4691852	0.2617652
<b>With elimination of outliers</b>										
rhatr_AD	54	0.0009173	4.35e-05	0.0009163	0.0008239	0.0010076	0.0008896	0.0009354	0.1902316	-0.2920651
rhatr_RD	54	0.0005071	2.95e-05	0.0005105	0.0004432	0.0005694	0.0004882	0.0005281	-0.0568160	-0.5335057
rhatr_MD	54	0.0006439	3.06e-05	0.0006400	0.0005788	0.0007120	0.0006251	0.0006611	0.1917191	-0.2792079

**Considerando las métricas AD, RD y MD conjuntamente**, se observa que existe normalidad multivariante con todos los sujetos y, si se eliminan todos los outliers detectados por la función `mvn`, se eliminan 4 pacientes. Para generar una similitud con el análisis anterior (considerando las 5 métricas del ATR derecho), se procede a eliminar el outlier con mayor distancia de Mahalanobis (sujeto 34, un HD), y, con ello, se mantiene la normalidad multivariante y la normalidad univariante para todas las métricas.

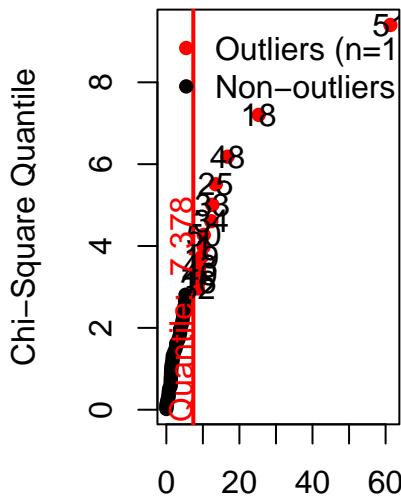
```
# Para las 2 métricas de RM correlacionadas con preHD: Iron,FA
library(MVN)
par(mfrow = c(1, 2))
#RATR_avg<-ATR_avg_demo[, c(3,8, 10:13)]
RATR_avg<-ATR_avg_demo[,c('Subgroup', 'rhatr_Iron', 'rhatr_AD', 'rhatr_RD',
                           'rhatr_MD', 'rhatr_FA')]

# Valoración de la normalidad multivariante con todos los sujetos
RATR_mvnmvn(RATR_avg[-c(1,3:5)], mvnTest = 'royston', covariance = TRUE, desc = TRUE,
              univariateTest = 'SW', univariatePlot = 'qq',
              multivariatePlot = 'qq', multivariateOutlierMethod = 'quan',
              showOutliers = TRUE, showNewData = TRUE)
```

**Chi-Square Q-Q Plot**



**Chi-Square Q-Q Plot**



Squared Mahalanobis Distance      Robust Squared Mahalanobis Distance

```

index<-rownames(RATR_mvnm$newData)
RATR_avg_WoOut<-RATR_avg[c(index),]
table(RATR_avg$Subgroup)

Control    PreHD      HD
      24       13      18
table(RATR_avg_WoOut$Subgroup)

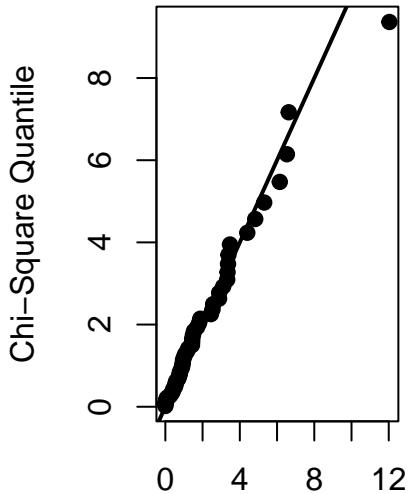
Control    PreHD      HD
      20        8      14

# Valoración de la normalidad multivariante con eliminación del outlier más
# extremo
par(mfrow = c(1, 2))
print(RATR_mvnm$multivariateOutliers,max = 6)
  Observation Mahalanobis Distance Outlier
51           51       61.314     TRUE
18           18       25.280     TRUE
 [ reached 'max' / getOption("max.print") -- omitted 11 rows ]
newRATR_avg<-RATR_avg[-c(51),]
RATR_avg[c(51),]

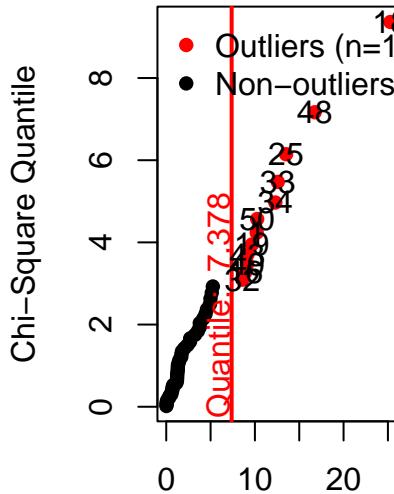
  Subgroup rhatr_Iron    rhatr_AD    rhatr_RD    rhatr_MD rhatr_FA
51    PreHD  0.0299576 0.000826666 0.000464213 0.00058503 0.380471
newRATR_mvnm<-mvn(newRATR_avg[-c(1,3:5)], mvnTest = 'royston', covariance = TRUE,
                     desc = TRUE, univariateTest = 'SW', univariatePlot = 'qq',
                     multivariatePlot = 'qq', multivariateOutlierMethod = 'quan',
                     showOutliers = TRUE, showNewData = TRUE)

```

**Chi-Square Q-Q Plot**



**Chi-Square Q-Q Plot**



Squared Mahalanobis Distance      Robust Squared Mahalanobis Distance

```
RATR_avg_demo_WoOut_IronFA<-ATR_avg_demo[-c(51), -c(9:12, 14:17)]
#RATR_avg_demo_WoOut_IronFA<-ATR_avg_demo[-c(51), -c('rhatr_AD', 'rhatr_RD',
#                                         'rhatr_MD', 'lhatr_Iron',
#                                         'lhatr_AD', 'lhatr_RD',
#                                         'lhatr_MD', 'lhatr_FA')]
RATR_avg_demo_IronFA<-ATR_avg_demo[, -c(9:12, 14:17)]
```

Cuadro 32: Comparación de la normalidad multivariante sin y con eliminación de outliers para el ATR derecho habiendo considerado las métricas Iron y FA conjuntamente

	Test	H	p value	MVN
Without elimination of outliers	Royston	18.136071	0.0001153	NO
With elimination of outliers	Royston	1.700364	0.4272644	YES

Cuadro 33: Comparación de la normalidad univariante sin y con eliminación de outliers para las métricas Iron y FA del ATR derecho

Without elimination of outliers					With elimination of outliers				
Test	Variable	Statistic	p value	Normality	Test	Variable	Statistic	p value	Normality
Shapiro-Wilk	rhatr_Iron	0.8860	0.0001	NO	Shapiro-Wilk	rhatr_Iron	0.9722	0.2397	YES
Shapiro-Wilk	rhatr_FA	0.9854	0.7395	YES	Shapiro-Wilk	rhatr_FA	0.9864	0.7953	YES

Cuadro 34: Comparación de los estadísticos descriptivos sin y con eliminación de outliers para las métricas Iron y FA del ATR derecho

Variables	n	Mean	Std.Dev	Median	Min	Max	25th	75th	Skew	Kurtosis
<b>Without elimination of outliers</b>										
rhatr_Iron	55	0.0235915	0.0014052	0.0235422	0.020136	0.0299576	0.0227244	0.0244503	1.3459722	5.8747627

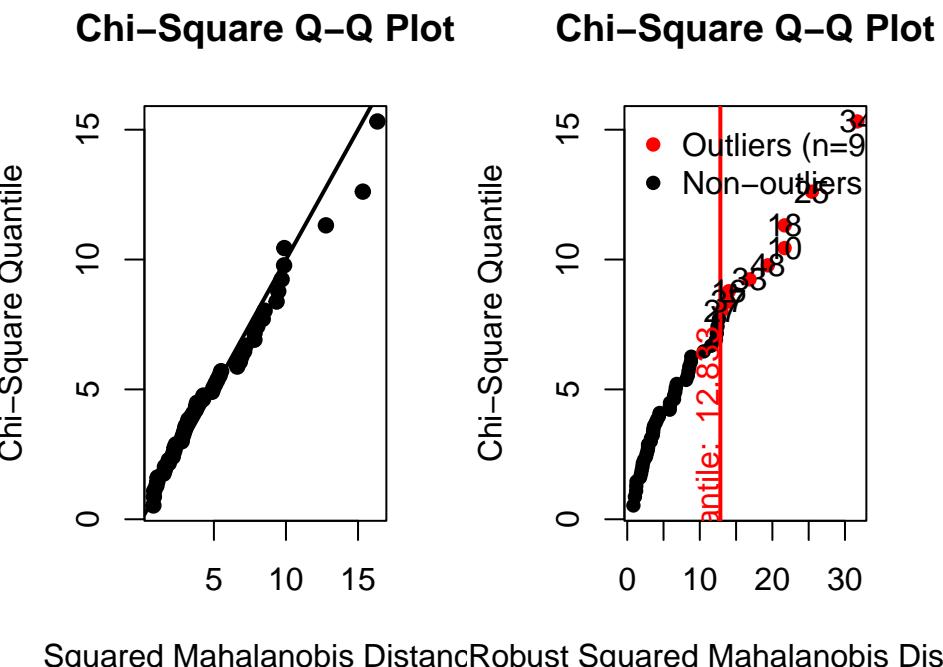
rhatr_FA	55	0.3802175	0.0267579	0.3783580	0.317049	0.4443100	0.3646285	0.3963130	-0.0257356	-0.1175621
<b>With elimination of outliers</b>										
rhatr_Iron	54	0.0234736	0.0011104	0.0235093	0.020136	0.0256621	0.0227194	0.0244182	-0.3978523	-0.0619997
rhatr_FA	54	0.3802129	0.0270091	0.3777460	0.317049	0.4443100	0.3645022	0.3970575	-0.0249758	-0.1719187

Considerando las métricas **iron** y **FA** conjuntamente, se observa que no existe normalidad multivariante ni normalidad univariante para el hierro. Si se eliminaran todos los outliers detectados por la función **mvn**, se eliminarían 4 controles, 5 preHD y 4 HD. Para generar una similitud con los dos análisis anteriores, se procede a eliminar el outlier con mayor distancia de Mahalanobis (sujeto 51, un preHD), y, con ello, se consigue la normalidad multivariante y la normalidad univariante para ambas métricas.

#### Resultados para el LATR:

```
library(MVN)
par(mfrow = c(1, 2))
# Para las 5 métricas de RM conjuntamente
#LATR_avg<-ATR_avg_demo[,c(3,9, 14:17)]
LATR_avg<-ATR_avg_demo[,c('Subgroup','lhatr_Iron','lhatr_AD','lhatr_RD',
                           'lhatr_MD','lhatr_FA')]

# Valoración de la normalidad multivariante con todos los sujetos
LATR_mvnmv<-mvn(LATR_avg[-1], mvnTest = 'royston', covariance = TRUE, desc = TRUE,
                  univariateTest = 'SW', univariatePlot = 'qq',
                  multivariatePlot = 'qq', multivariateOutlierMethod = 'quan',
                  showOutliers = TRUE, showNewData = TRUE)
```



```
index<-rownames(LATR_mvnmv$newData)
LATR_avg_WoOut<-LATR_avg[c(index),]
table(LATR_avg$Subgroup)
```

```

Control    PreHD      HD
     24       13      18
table(LATR_avg_WoOut$Subgroup)

```

```

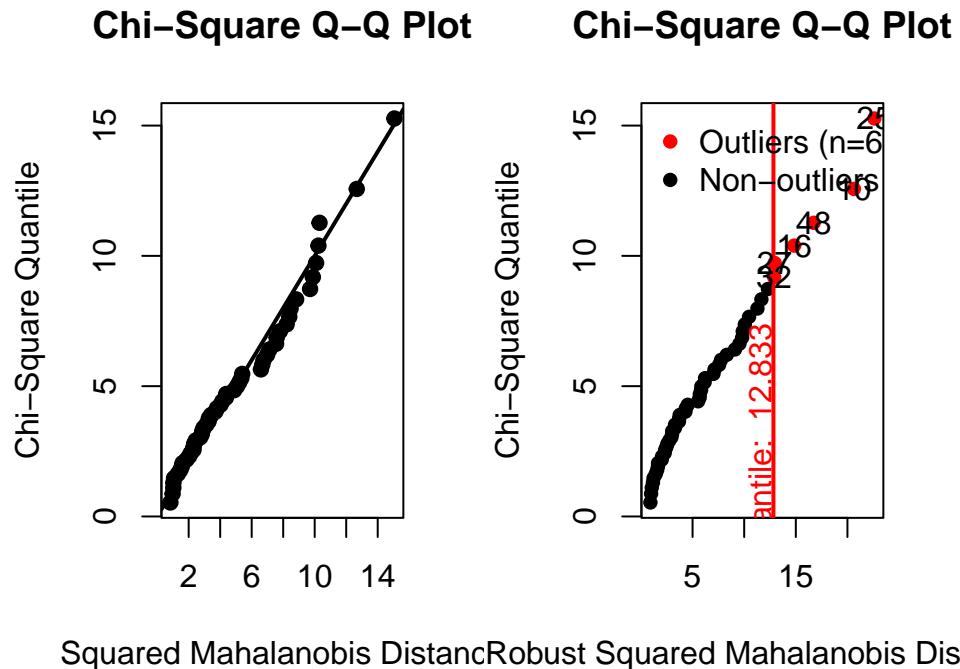
Control    PreHD      HD
     21       12      13

```

```

# Valoración de la normalidad multivariante con eliminación del outlier más
# extremo
par(mfrow = c(1, 2))
print(LATR_mvnmultivariateOutliers,max = 6)
  Observation Mahalanobis Distance Outlier
34          34         31.713   TRUE
25          25         25.442   TRUE
[ reached 'max' / getOption("max.print") -- omitted 7 rows ]
newLATR_avg<-LATR_avg[-c(34),]
LATR_avg[c(34),]
  Subgroup lhatr_Iron   lhatr_AD   lhatr_RD   lhatr_MD lhatr_FA
34      HD  0.0285345 0.00093618 0.000576567 0.000696437  0.3282
newLATR_mvnmvnm(newLATR_avg[-1], mvnTest = 'royston', covariance = TRUE,
                 desc = TRUE, univariateTest = 'SW', univariatePlot = 'qq',
                 multivariatePlot = 'qq', multivariateOutlierMethod = 'quan',
                 showOutliers = TRUE, showNewData = TRUE)

```



```

LATR_avg_demo_WoOut<-LATR_avg_demo[-c(34),-c(8, 10:13)]
#LATR_avg_demo_WoOut<-LATR_avg_demo[-c(34),-c('lhatr_Iron', 'lhatr_AD', 'lhatr_RD',
#                                         'lhatr_MD', 'lhatr_FA')]
#
LATR_avg_demo<-LATR_avg_demo[,-c(8, 10:13)]

```

Cuadro 35: Comparación de la normalidad multivariante sin y con eliminación de outliers para ATR izquierdo habiendo considerado las 5 métricas conjuntamente

		Test	H	p value	MVN
Without elimination of outliers		Royston	2.356691	0.6056566	YES
With elimination of outliers		Royston	1.816905	0.7246743	YES

Cuadro 36: Comparación de la normalidad univariante sin y con eliminación de outliers para las 5 métricas de ATR izquierdo

Without elimination of outliers					With elimination of outliers				
Test	Variable	Statistic	p value	Normality	Test	Variable	Statistic	p value	Normality
Shapiro-Wilk	lhatr_Iron	0.9885	0.8755	YES	Shapiro-Wilk	lhatr_Iron	0.9895	0.9165	YES
Shapiro-Wilk	lhatr_AD	0.9924	0.9790	YES	Shapiro-Wilk	lhatr_AD	0.9909	0.9546	YES
Shapiro-Wilk	lhatr_RD	0.9773	0.3822	YES	Shapiro-Wilk	lhatr_RD	0.9857	0.7637	YES
Shapiro-Wilk	lhatr_MD	0.9860	0.7694	YES	Shapiro-Wilk	lhatr_MD	0.9869	0.8184	YES
Shapiro-Wilk	lhatr_FA	0.9622	0.0816	YES	Shapiro-Wilk	lhatr_FA	0.9642	0.1064	YES

Cuadro 37: Comparación de los estadísticos descriptivos sin y con eliminación de outliers para las 5 métricas de ATR izquierdo

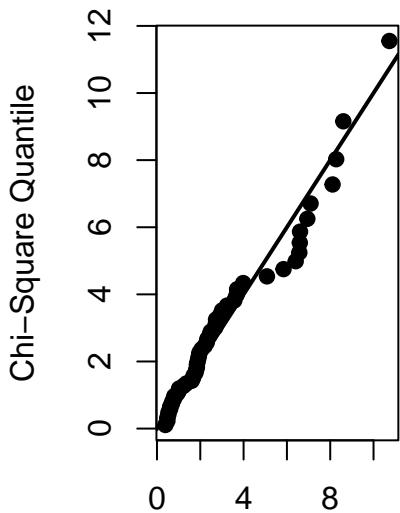
Variables	n	Mean	Std.Dev	Median	Min	Max	25th	75th	Skew	Kurtosis
<b>Without elimination of outliers</b>										
lhatr_Iron	55	0.0245373	0.0019590	0.0245776	0.0191068	0.0290945	0.0233934	0.0257187	-0.0442936	0.1930655
lhatr_AD	55	0.0008997	0.0000399	0.0009018	0.0008042	0.0009925	0.0008752	0.0009222	0.0438449	-0.2262237
lhatr_RD	55	0.0004978	0.0000290	0.0004991	0.0004411	0.0005766	0.0004791	0.0005173	0.4304379	-0.0320765
lhatr_MD	55	0.0006318	0.0000293	0.0006281	0.0005621	0.0007040	0.0006117	0.0006485	0.2053286	-0.2406366
lhatr_FA	55	0.3815795	0.0249843	0.3850040	0.3170270	0.4224300	0.3694610	0.3990850	-0.5442118	-0.0136489
<b>With elimination of outliers</b>										
lhatr_Iron	54	0.0244633	0.0018981	0.0244858	0.0191068	0.0290945	0.0232654	0.0256739	-0.1076489	0.2996520
lhatr_AD	54	0.0008990	0.0000400	0.0009012	0.0008042	0.0009925	0.0008752	0.0009219	0.0801523	-0.1999189
lhatr_RD	54	0.0004964	0.0000271	0.0004972	0.0004411	0.0005706	0.0004788	0.0005163	0.2393167	-0.3492653
lhatr_MD	54	0.0006306	0.0000282	0.0006269	0.0005621	0.0007040	0.0006115	0.0006471	0.1362887	-0.2181905
lhatr_FA	54	0.3825680	0.0241089	0.3853515	0.3170270	0.4224300	0.3697998	0.3992100	-0.5366882	0.1411953

En el caso del ATR izquierdo, **considerando las 5 métricas conjuntamente**, se observa que existe normalidad multivariante sin eliminación de outliers. En el caso que se eliminaran todos los outliers definidos por la distancia de Mahalanobis, se eliminarían 3 controles, un preHD y 5 HD, con una mayor eliminación relativa de pacientes sintomáticos, considerando no correcto eliminar los outliers. Para generar una similitud con el ATR derecho, se elimina el outlier con mayor distancia de Mahalanobis (sujeto 34, un HD, siendo el mismo sujeto eliminado también en el ATR derecho), y, con ello, se mantiene la normalidad multivariante y la normalidad univariante para todas las métricas.

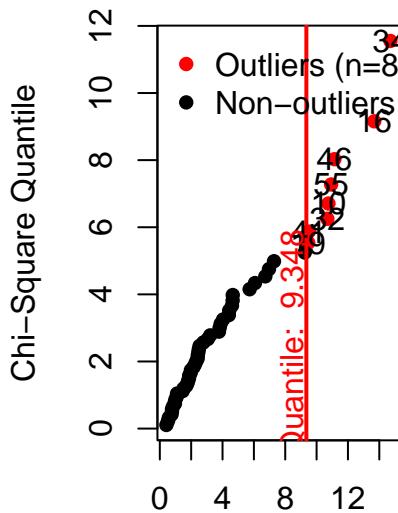
```
library(MVN)
par(mfrow = c(1, 2))
# Para las 3 métricas de RM correlacionadas con HD: AD, RD, MD
#LATR_avg<-ATR_avg_demo[,c(3,9, 14:17)]
LATR_avg<-ATR_avg_demo[,c('Subgroup','lhatr_Iron','lhatr_AD','lhatr_RD',
                           'lhatr_MD','lhatr_FA')]

# Valoración de la normalidad multivariante con todos los sujetos
LATR_mvnmv<-mvn(LATR_avg[-c(1,2,6)], mvnTest = 'royston', covariance = TRUE, desc = TRUE,
                   univariateTest = 'SW', univariatePlot = 'qq',
                   multivariatePlot = 'qq', multivariateOutlierMethod = 'quan',
                   showOutliers = TRUE, showNewData = TRUE)
```

**Chi-Square Q-Q Plot**



**Chi-Square Q-Q Plot**



Squared Mahalanobis Distance Robust Squared Mahalanobis Distance

```

index<-rownames(LATR_mvnm$newData)
LATR_avg_WoOut<-LATR_avg[c(index),]
table(LATR_avg$Subgroup)

Control    PreHD      HD
      24        13       18
table(LATR_avg_WoOut$Subgroup)

Control    PreHD      HD
      21        13       13

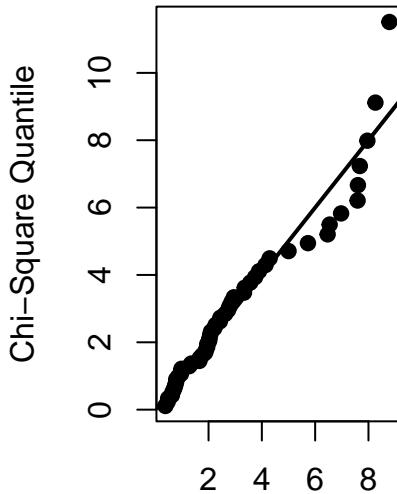
```

```

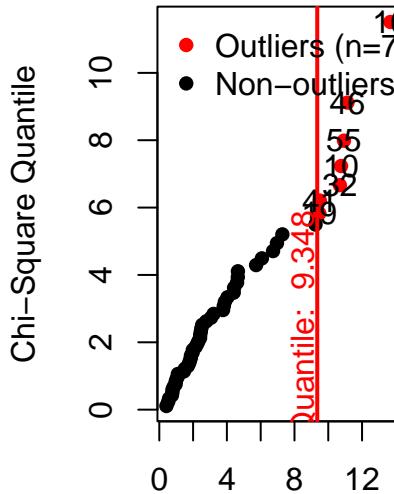
# Valoración de la normalidad multivariante con eliminación del outlier más
# extremo
par(mfrow = c(1, 2))
print(LATR_mvnm$multivariateOutliers,max = 6)
  Observation Mahalanobis Distance Outlier
34          34         14.708     TRUE
16          16         13.677     TRUE
 [ reached 'max' / getOption("max.print") -- omitted 6 rows ]
newLATR_avg<-LATR_avg[-c(34),]
LATR_avg[c(34),]
  Subgroup lhatr_Iron   lhatr_AD   lhatr_RD   lhatr_MD lhatr_FA
34      HD  0.0285345 0.00093618 0.000576567 0.000696437  0.3282
newLATR_mvnm<-mvn(newLATR_avg[-c(1,2,6)], mvnTest = 'royston', covariance = TRUE,
                     desc = TRUE, univariateTest = 'SW', univariatePlot = 'qq',
                     multivariatePlot = 'qq', multivariateOutlierMethod = 'quan',
                     showOutliers = TRUE, showNewData = TRUE)

```

**Chi-Square Q-Q Plot**



**Chi-Square Q-Q Plot**



Squared Mahalanobis Distance Robust Squared Mahalanobis Distance

```
LATR_avg_demo_WoOut_ADRDMD<-ATR_avg_demo[-c(34), -c(8:13,17)]
#LATR_avg_demo_WoOut_ADRDMD<-ATR_avg_demo[-c(34), -c('rhatr_Iron', 'lhatr_Iron',
#                                         'rhatr_AD', 'rhatr_RD',
#                                         'rhatr_MD', 'rhatr_FA', 'lhatr_FA')]
LATR_avg_demo_ADRDMD<-ATR_avg_demo[, -c(8:13,17)]
```

Cuadro 38: Comparación de la normalidad multivariante sin y con eliminación de outliers para ATR izquierdo habiendo considerado las métricas AD, RD y MD conjuntamente

	Test	H	p value	MVN
Without elimination of outliers	Royston	0.5495921	0.7737979	YES
With elimination of outliers	Royston	0.1011854	0.9565806	YES

Cuadro 39: Comparación de la normalidad univariante sin y con eliminación de outliers para las métricas AD, RD y MD del ATR izquierdo

Without elimination of outliers					With elimination of outliers				
Test	Variable	Statistic	p value	Normality	Test	Variable	Statistic	p value	Normality
Shapiro-Wilk	lhatr_AD	0.9924	0.9790	YES	Shapiro-Wilk	lhatr_AD	0.9909	0.9546	YES
Shapiro-Wilk	lhatr_RD	0.9773	0.3822	YES	Shapiro-Wilk	lhatr_RD	0.9857	0.7637	YES
Shapiro-Wilk	lhatr_MD	0.9860	0.7694	YES	Shapiro-Wilk	lhatr_MD	0.9869	0.8184	YES

Cuadro 40: Comparación de los estadísticos descriptivos sin y con eliminación de outliers para las métricas AD, RD y MD del ATR izquierdo

Variables	n	Mean	Std.Dev	Median	Min	Max	25th	75th	Skew	Kurtosis
<b>Without elimination of outliers</b>										
lhatr_AD	55	0.0008997	3.99e-05	0.0009018	0.0008042	0.0009925	0.0008752	0.0009222	0.0438449	-0.2262237

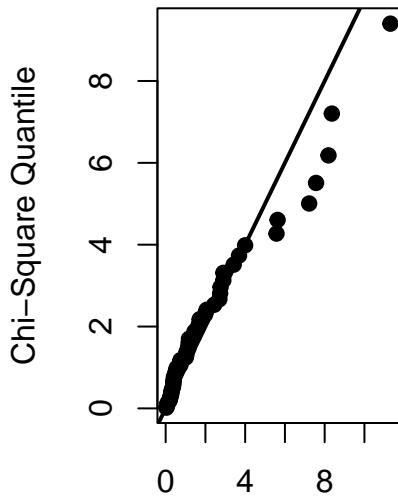
lhatr_RD	55	0.0004978	2.90e-05	0.0004991	0.0004411	0.0005766	0.0004791	0.0005173	0.4304379	-0.0320765
lhatr_MD	55	0.0006318	2.93e-05	0.0006281	0.0005621	0.0007040	0.0006117	0.0006485	0.2053286	-0.2406366
<b>With elimination of outliers</b>										
lhatr_AD	54	0.0008990	4.00e-05	0.0009012	0.0008042	0.0009925	0.0008752	0.0009219	0.0801523	-0.1999189
lhatr_RD	54	0.0004964	2.71e-05	0.0004972	0.0004411	0.0005706	0.0004788	0.0005163	0.2393167	-0.3492653
lhatr_MD	54	0.0006306	2.82e-05	0.0006269	0.0005621	0.0007040	0.0006115	0.0006471	0.1362887	-0.2181905

Considerando las métricas **AD**, **RD** y **MD** conjuntamente, se observa que existe normalidad multivariante y univariante para las 3 métricas sin eliminación de outliers. En el caso que se eliminaran todos los outliers definidos por la distancia de Mahalanobis, se eliminarían 3 controles y 5 HD, con una mayor eliminación relativa de pacientes sintomáticos, considerando no correcto eliminar los outliers. Para generar una similitud con los análisis previos, se elimina el outlier con mayor distancia de Mahalanobis (sujeto 34, un HD), y, con ello, se mantiene la normalidad multivariante y la normalidad univariante para todas las métricas.

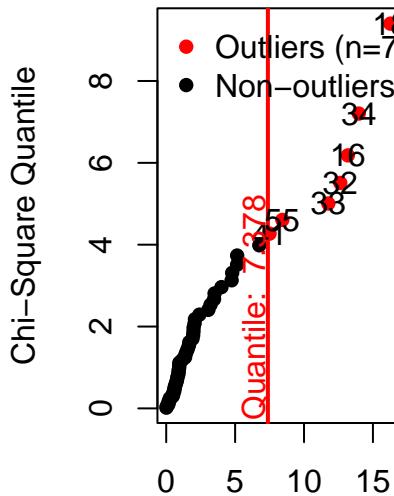
```
library(MVN)
par(mfrow = c(1, 2))
# Para las 2 métricas de RM correlacionadas con preHD: Iron y FA
#LATR_avg<-ATR_avg_demo[,c(3,9, 14:17)]
LATR_avg<-ATR_avg_demo[,c('Subgroup','lhatr_Iron','lhatr_AD','lhatr_RD',
                           'lhatr_MD','lhatr_FA')]

# Valoración de la normalidad multivariante con todos los sujetos
LATR_mvnmvn(LATR_avg[-c(1,3:5)], mvnTest = 'royston', covariance = TRUE, desc = TRUE,
             univariateTest = 'SW', univariatePlot = 'qq',
             multivariatePlot = 'qq', multivariateOutlierMethod = 'quan',
             showOutliers = TRUE, showNewData = TRUE)
```

Chi-Square Q-Q Plot



Chi-Square Q-Q Plot



Squared Mahalanobis Distance Robust Squared Mahalanobis Distance

```
index<-rownames(LATR_mvnmvn$newData)
LATR_avg_WoOut<-LATR_avg[c(index),]
table(LATR_avg$Subgroup)
```

```

Control    PreHD      HD
  24        13       18
table(LATR_avg_WoOut$Subgroup)

```

```

Control    PreHD      HD
  22        12       14

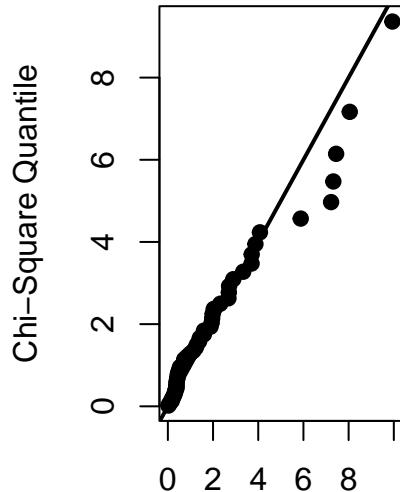
```

```

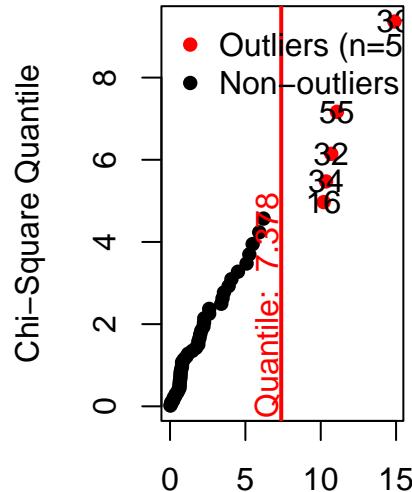
# Valoración de la normalidad multivariante con eliminación del outlier más
# extremo
par(mfrow = c(1, 2))
print(LATR_mvnmultivariateOutliers,max = 6)
  Observation Mahalanobis Distance Outlier
18          18           16.268     TRUE
34          34           13.976     TRUE
[ reached 'max' / getOption("max.print") -- omitted 5 rows ]
newLATR_avg<-LATR_avg[-c(18),]
LATR_avg[c(18),]
  Subgroup lhatr_Iron   lhatr_AD   lhatr_RD   lhatr_MD lhatr_FA
18  Control  0.0191068 0.000875271 0.000519527 0.000638108 0.341479
newLATR_mvnmvnm(newLATR_avg[-c(1,3:5)], mvnTest = 'royston', covariance = TRUE,
  desc = TRUE, univariateTest = 'SW', univariatePlot = 'qq',
  multivariatePlot = 'qq', multivariateOutlierMethod = 'quan',
  showOutliers = TRUE, showNewData = TRUE)

```

Chi-Square Q-Q Plot



Chi-Square Q-Q Plot



Squared Mahalanobis Distance Robust Squared Mahalanobis Dist

```

LATR_avg_demo_WoOut_IronFA<-LATR_avg_demo[-c(18),-c(8,10:16)]
#LATR_avg_demo_WoOut_IronFA<-LATR_avg_demo[-c(18),-c('lhatr_Iron','lhatr_AD','lhatr_RD',
#                                              'lhatr_MD','lhatr_FA','lhatr_AD',
#                                              'lhatr_RD','lhatr_MD')]
LATR_avg_demo_IronFA<-LATR_avg_demo[,-c(8,10:16)]

```

Cuadro 41: Comparación de la normalidad multivariante sin y con eliminación de outliers para ATR izquierdo habiendo considerado las métricas Iron y FA conjuntamente

		Test	H	p value	MVN
Without elimination of outliers		Royston	2.485619	0.2885702	YES
With elimination of outliers		Royston	3.111499	0.2109326	YES

Cuadro 42: Comparación de la normalidad univariante sin y con eliminación de outliers para las métricas Iron y FA del ATR izquierdo

Without elimination of outliers					With elimination of outliers				
Test	Variable	Statistic	p value	Normality	Test	Variable	Statistic	p value	Normality
Shapiro-Wilk	lhatr_Iron	0.9885	0.8755	YES	Shapiro-Wilk	lhatr_Iron	0.9851	0.7357	YES
Shapiro-Wilk	lhatr_FA	0.9622	0.0816	YES	Shapiro-Wilk	lhatr_FA	0.9587	0.0605	YES

Cuadro 43: Comparación de los estadísticos descriptivos sin y con eliminación de outliers para las métricas Iron y FA del ATR izquierdo

Variables	n	Mean	Std.Dev	Median	Min	Max	25th	75th	Skew	Kurtosis
<b>Without elimination of outliers</b>										
lhatr_Iron	55	0.0245373	0.0019590	0.0245776	0.0191068	0.0290945	0.0233934	0.0257187	-0.0442936	0.1930655
lhatr_FA	55	0.3815795	0.0249843	0.3850040	0.3170270	0.4224300	0.3694610	0.3990850	-0.5442118	-0.0136489
<b>With elimination of outliers</b>										
lhatr_Iron	54	0.0246379	0.0018284	0.0246244	0.0207260	0.0290945	0.0236526	0.0257363	0.2675624	-0.2325812
lhatr_FA	54	0.3823221	0.0245986	0.3853515	0.3170270	0.4224300	0.3697998	0.3992100	-0.5894641	0.1719903

**Considerando las métricas Iron y FA conjuntamente**, se observa que existe normalidad multivariante y univariante para las 2 métricas sin eliminación de outliers. En el caso que se eliminaran todos los outliers definidos por la distancia de Mahalanobis, se eliminarían 2 controles, un preHD y 4 HD. Para generar una similitud con los análisis previos, se elimina el outlier con mayor distancia de Mahalanobis (sujeto 18, un control), y, con ello, se mantiene la normalidad multivariante y la normalidad univariante para ambas métricas.

#### 2.4.4 Homocedasticidad de la matriz de varianza-covarianza

Se consultan las siguientes fuentes de información: Homogeneidad de la covarianza, Box's M Test: Definition, boxM {heplots}.

El **test de homogeneidad de las matrices de covarianza o suposición de homocedasticidad** es una suposición que debe cumplirse en el análisis multivariante, y, consiste en que múltiples grupos en un diseño experimental o test estadístico tengan matrices de covarianza iguales, es decir, las matrices de covarianza poblacional para las variables dependientes de cada grupo deben ser iguales.

Las variables pueden no ser homocedásticas por dos razones principales:

- Una variable puede ser no normal;
- Una variable puede tener una relación con la transformación de otra variable.

El **test M de Box** (o test de Box para la equivalencia de matrices de covarianza que utiliza la distribución F) es un test paramétrico que se utiliza para comparar la variación en muestras multivariadas, es decir, determina si dos o más matrices de covarianza son iguales (homogéneas).

Un **inconveniente** de dicho test es su extremada sensibilidad a las desviaciones de la normalidad, es decir, los datos deben tener una distribución normal multivariada. Además, tiene muy poca potencia para tamaños de muestra pequeños (resultados no significativos en el contexto de muestras pequeñas no indican necesariamente que las

matrices de covarianza sean iguales) y es demasiado sensible para muestras de gran tamaño (resultados falsamente significativos en muestras de gran tamaño, por lo que en este contexto se recomienda un alfa más pequeño).

La **hipótesis nula** p-valor superior a 0.05 es que las matrices de covarianza observadas para las variables dependientes son iguales en todos los grupos: un estadístico M de Box no significativo indica que las matrices de covarianza son iguales.

La **hipótesis alternativa** con p-valor inferior a 0.05 indica que las matrices de covarianza son significativamente diferentes. El incumplimiento de la suposición de homogeneidad de las matrices de covarianza no siempre significa que no se puede ejecutar un análisis estadístico. En MANOVA, aparecen  $n$  desiguales a medida que disminuyen los tamaños de muestra, violando la suposición de homogeneidad de covarianza; en estos casos, se debe utilizar la traza de Pillai por ser más robusta que otros tests como la traza de Hotelling o la Lambda de Wilks.

Como se hizo en el caso de los tests de homocedasticidad univariante, se procede a calcular la homocedasticidad multivariante (para las 5 métricas, para las métricas AD-RD-MD y para las métricas Iron-FA) mediante los test M de Box para las variables independientes de subgrupo, género y edad factorizada, con los dataframes sin y con la eliminación de outliers.

#### Resultados para el RATR:

Cuadro 44: Homocedasticidad de las matrices de covarianza del ATR derecho

Variables	For 5 metrics			For diffusivity metrics			For iron-FA metrics		
	Subgroup	Gender	Age	Subgroup	Gender	Age	Subgroup	Gender	Age
<b>Without elimination of outliers</b>									
Chi-square	51.038	24.367	25.255	24.514	17.786	9.83	11.865	2.214	5.057
Df	30	15	30	12	6	12	6	3	6
P-value	0.01	0.059	0.713	0.017	0.007	0.631	0.065	0.529	0.537
<b>With elimination of outliers</b>									
Chi-square	40.488	17.524	28.209	20.035	13.321	14.033	3.813	0.267	5.815
Df	30	15	30	12	6	12	6	3	6
P-value	0.096	0.288	0.559	0.066	0.038	0.299	0.702	0.966	0.444

**Sin eliminación de outliers**, se objetiva una falta de homocedasticidad multivariante para subgrupo en los casos de las 5 métricas conjuntamente y de las métricas AD-RD-MD, y, una falta de homocedasticidad multivariante para género en el caso de las métricas AD-RD-MD.

**Con eliminación de outliers**, se objetiva una falta de homocedasticidad multivariante para género en el caso de las métricas AD-RD-MD.

#### Resultados para el LATR:

Cuadro 45: Homocedasticidad de las matrices de covarianza del ATR izquierdo

Variables	For 5 metrics			For diffusivity metrics			For iron-FA metrics		
	Subgroup	Gender	Age	Subgroup	Gender	Age	Subgroup	Gender	Age
<b>Without elimination of outliers</b>									
Chi-square	31.801	20.239	47.7	10.247	14.583	17.76	7.843	0.54	7.301
Df	30	15	30	12	6	12	6	3	6
P-value	0.377	0.163	0.021	0.594	0.024	0.123	0.25	0.91	0.294
<b>With elimination of outliers</b>									
Chi-square	29.625	19.185	48.859	10.354	12.912	18.483	10.85	3.337	4.638
Df	30	15	30	12	6	12	6	3	6
P-value	0.485	0.205	0.016	0.585	0.044	0.102	0.093	0.343	0.591

**Sin eliminación de outliers**, se objetiva una falta de homocedasticidad multivariante para edad factorizada en el caso de las 5 métricas conjuntamente y para género en el caso de las métricas AD-RD-MD.

**Con eliminación de outliers**, se objetiva una falta de homocedasticidad multivariante para género en el caso de las métricas AD-RD-MD.

### 3 ANÁLISIS del average: RATR y LATR por separado

#### 3.1 MANOVA

Se consultan las siguientes fuentes de información: 14. ANOVA / MANOVA, MANOVA in R – How To Implement and Interpret One-Way MANOVA.

MANOVA (*Multivariate ANOVA o Multivariate ANalysis Of VAriance*, análisis multivariante de la varianza) es una extensión del ANOVA, pero MANOVA incluye al menos dos variables dependientes para analizar diferencias de las matrices de covarianzas procedentes de las variables dependientes entre múltiples grupos (factores).

MANOVA en R utiliza por defecto el **test de Pillai**, que es el de mayor poder estadístico, para los cálculos, aunque se pueden utilizar otros tests (lambda de Wilk, raíz mayor de Roy o test de Hotelling-Lawley), y, luego se convierte en un estadístico F para determinar la significancia de las diferencias de medias grupales.

El ANOVA suele sufrir de errores de tipo I que se producen con la realización de múltiples ANOVA para cada variable dependiente. En cambio, MANOVA determina la existencia de diferencias de grupo basándose en la información combinada de múltiples variables dependientes. Debido a que MANOVA utiliza más de una variable dependiente, las hipótesis nula y alternativa cambian ligeramente:

- **H<sub>0</sub>**: Los vectores medios de grupo son los mismos para todos los grupos o no difieren significativamente;
- **H<sub>1</sub>**: Al menos uno de los vectores medios del grupo es diferente del resto.

MANOVA es una técnica paramétrica que implica ciertos **supuestos** sobre los parámetros de la población de la que se toman las muestras y dichas suposiciones son más restrictivas que en el caso del ANOVA:

- **Normalidad multivariada**: cada combinación de variables independientes o dependientes debe tener una distribución normal multivariada. Los errores también deben ajustarse a la **normalidad multivariada** (es decir, se distribuyen normalmente en todas las dimensiones);
- Toda la **matriz de varianza-covarianza es homogénea/homocedástica** (es decir, las varianzas de todas las variables y las covarianzas entre todos los pares de variables son iguales en todos los grupos);
- **Sin multicolinealidad**: las variables dependientes no deben tener correlaciones muy altas.
- **Sin valores atípicos**: no debería haber valores atípicos en las variables dependientes.
- **Linealidad**: las variables dependientes deben tener una relación lineal con cada grupo (factor) de la variable independiente.
- **Independencia de las observaciones**: las unidades son intercambiables entre sí.

De todas ellas, la multicolinealidad es el supuesto que no se cumple entre las métricas de difusión, con una alta correlación entre AD-RD-MD y entre FA-RD.

```
# Correlación entre las métricas del ATR derecho
cor(RATR_avg_demo[,8:12])
    rhatr_Iron   rhatr_AD   rhatr_RD   rhatr_MD   rhatr_FA
rhatr_Iron  1.0000000 -0.4552380 -0.2290054 -0.3540628 -0.1272555
rhatr_AD   -0.4552380  1.0000000  0.6185212  0.8541651  0.1186825
rhatr_RD   -0.2290054  0.6185212  1.0000000  0.9369202 -0.6934039
rhatr_MD   -0.3540628  0.8541651  0.9369202  1.0000000 -0.4060830
rhatr_FA   -0.1272555  0.1186825 -0.6934039 -0.4060830  1.0000000
# Correlación entre las métricas del ATR izquierdo
cor(LATR_avg_demo[,8:12])
    lhatr_Iron   lhatr_AD   lhatr_RD   lhatr_MD   lhatr_FA
lhatr_Iron  1.0000000 -0.2635346 -0.08969432 -0.1789768 -0.0853358
lhatr_AD   -0.26353463  1.0000000  0.59763448  0.8488016  0.1630650
lhatr_RD   -0.08969432  0.5976345  1.00000000  0.9311774 -0.6779453
lhatr_MD   -0.17897684  0.8488016  0.93117739  1.0000000 -0.3729132
lhatr_FA   -0.08533580  0.1630650 -0.67794527 -0.3729132  1.0000000
```

Para aplicar MANOVA en R, se debe cumplir una **condición crítica**: el número de observaciones (filas) por grupo de la variable independiente debe ser mayor al número de variables dependientes.

Si el MANOVA es significativo, se puede medir el **tamaño del efecto** mediante la métrica Eta cuadrado parcial, la cual mide el efecto que tiene la variable independiente sobre las variables dependientes: si el valor es 0,14 o mayor, se puede afirmar que el tamaño del efecto es grande.

Para ejecutar MANOVA, se utiliza la función `manova` del paquete `stats`. Se utilizan los dataframes con eliminación de outliers dado que son los que presentan normalidad multivariante. Se generan MANOVAs para las 5 métricas conjuntas, para las 3 métricas de difusividad (AD, RD, MD) y para las métricas de hierro y anistropía fraccional. El primer modelo generado consiste en las 3 variables independientes en interacción y los modelos subsiguientes con simplificación de las interacciones y/o del número de factores en función de los resultados del primer modelo.

Como se verá a continuación, algunos de los modelos especifican que los residuos son rango-deficientes y se ha de especificar el argumento `tol=0` para poder ver el resultado/resumen del modelo. El problema de especificar tolerancia 0 es que los resultados pueden ser inexactos, con lo que transformar las respuestas/variables dependientes para eliminar la alta correlación es mejor opción. Además, cuando los residuos son rango-deficientes, no se permite calcular el *effect size*.

En nuestro caso, las variables con mayor correlación son las difusividades media, axial y radial (las 3 métricas con la misma unidad de medida  $mm^2/s$ , con valores muy próximos a cero dado que los valores se hallan en el rango de  $10^{-4} - 10^{-5}$ ). Se han intentado varias transformaciones (logarítmica con la función `log`, isométrica log-ratio con la función `ilr`, log-ratio centrada con la función `clr`, arcsine con la función `asin(sqrt)`, raíz cuadrada con la función `sqrt`), ya sea de las tres variables dependientes altamente correlacionadas o de las cinco conjuntamente. Finalmente, se opta por generar una transformación logarítmica de las variables de difusividad por estar muy correlacionadas y medirse en la misma unidad; además, dicha transformación es la que obtiene unos coeficientes de discriminantes lineales menos dispares entre las 5 métricas.

Se consultan las webs siguientes: manova: Multivariate Analysis of Variance, Error in summary.manova - residuals have rank order deficiency, clr: Centred Log-Ratio (clr) transformation, Isometric log ratio transform, How to perform isometric log-ratio transformation.

### 3.1.1 RATR

```
# Para las 5 métricas conjuntamente - Sin transformación de variables
# Generar los MANOVAS
RATR_manova5_1<-manova(cbind(rhatr_Iron,rhatr_AD,rhatr_RD,rhatr_MD,rhatr_FA)~Subgroup*Gender*agef2,
                           data=RATR_avg_demo_WoOut)
summary(RATR_manova5_1)
Error in summary.manova(RATR_manova5_1): residuals have rank 4 < 5
summary(RATR_manova5_1,tol=0)
      Df Pillai approx F num Df den Df Pr(>F)
Subgroup        2  0.37832  1.58638     10    68  0.1295
Gender          1  0.02436  0.16477      5    33  0.9737
agef2           2  0.35358  1.46037     10    68  0.1737
Subgroup:Gender 2  0.22059  0.84299     10    68  0.5895
Subgroup:agef2   4  0.60904  1.29318     20   144  0.1926
Gender:agef2     2  0.32670  1.32765     10    68  0.2339
Subgroup:Gender:agef2 2  0.18417  0.68967     10    68  0.7304
Residuals       37
RATR_manova5_2<-manova(cbind(rhatr_Iron,rhatr_AD,rhatr_RD,rhatr_MD,rhatr_FA)~Subgroup+Gender+agef2,
                           data=RATR_avg_demo_WoOut)
summary(RATR_manova5_2)
      Df Pillai approx F num Df den Df Pr(>F)
Subgroup     2  0.33094  1.74483     10    88  0.08314 .
Gender       1  0.02088  0.18336      5    43  0.96738
agef2        2  0.27741  1.41716     10    88  0.18599
```

```

Residuals 47
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RATR_manova5_3<-manova(cbind(rhatr_Iron,rhatr_AD,rhatr_RD,rhatr_MD,rhatr_FA)~Subgroup,
                           data=RATR_avg_demo_WoOut)
summary(RATR_manova5_3)
      Df Pillai approx F num Df den Df Pr(>F)
Subgroup   2 0.31222  1.7389     10     94 0.08328 .
Residuals 50
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Generar effect size
library(effectsize)
RATR_SizeMan5_2<-effectsize::eta_squared(RATR_manova5_2);RATR_SizeMan5_2
# Effect Size for ANOVA (Type I)

Parameter | Eta2 (partial) |      95% CI
-----
Subgroup |          0.17 | [0.00, 1.00]
Gender   |          0.02 | [0.00, 1.00]
agef2    |          0.14 | [0.00, 1.00]

- One-sided CIs: upper bound fixed at [1.00].
RATR_SizeMan5_3<-effectsize::eta_squared(RATR_manova5_3);RATR_SizeMan5_3
# Effect Size for ANOVA (Type I)

Parameter | Eta2 (partial) |      95% CI
-----
Subgroup |          0.16 | [0.00, 1.00]

- One-sided CIs: upper bound fixed at [1.00].

```

En el caso de las **5 métricas conjuntas sin transformación logarítmica**, se objetiva que subgrupo es el único factor casi significativo con un p valor de 0.083, con un tamaño de efecto de 0.16.

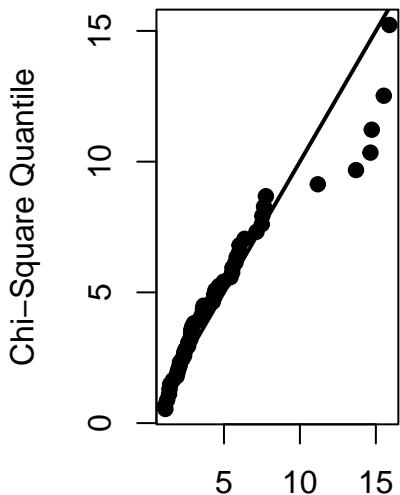
```

# Para las 5 métricas conjuntamente - Con transformación logarítmica de las métricas de difusividad
# Primero transformación de las variables
RATR_avg_demo_WoOut_log<-cbind(RATR_avg_demo_WoOut[,1:8],log(cbind(RATR_avg_demo_WoOut[,9:11])),
                                  RATR_avg_demo_WoOut[,12])
colnames(RATR_avg_demo_WoOut_log)<-colnames(RATR_avg_demo_WoOut)

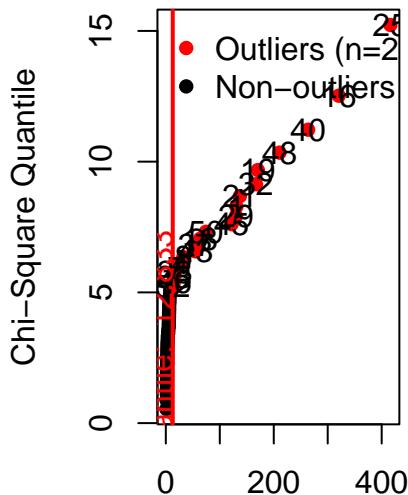
# Comprobar normalidad multivariante
par(mfrow = c(1, 2))
RATR_mvnllog5<-mvn(RATR_avg_demo_WoOut_log[-c(1:7)], mvnTest = 'royston', covariance = TRUE,
                      desc = TRUE, univariateTest = 'SW', univariatePlot = 'qq',
                      multivariatePlot = 'qq', multivariateOutlierMethod = 'quan',
                      showOutliers = TRUE, showNewData = TRUE)

```

**Chi-Square Q-Q Plot**



**Chi-Square Q-Q Plot**



Squared Mahalanobis Distance      Robust Squared Mahalanobis Distance

```
RATR_mvnllog5$multivariateNormality
  Test      H   p value MVN
1 Royston 1.663924 0.7506743 YES
RATR_mvnllog5$univariateNormality
  Test Variable Statistic   p value Normality
1 Shapiro-Wilk rhatr_Iron    0.9749  0.3253   YES
2 Shapiro-Wilk rhatr_AD     0.9790  0.4704   YES
3 Shapiro-Wilk rhatr_RD     0.9839  0.6880   YES
4 Shapiro-Wilk rhatr_MD     0.9821  0.6084   YES
5 Shapiro-Wilk rhatr_FA     0.9849  0.7351   YES

# Comprobar homocedasticidad de la matriz de covarianzas
RATR_MBox_SG5log<-boxM(cbind(rhatr_Iron,rhatr_AD,rhatr_RD,rhatr_MD,rhatr_FA)~Subgroup,
                         data=RATR_avg_demo_WoOut_log, cov=TRUE);RATR_MBox_SG5log

  Box's M-test for Homogeneity of Covariance Matrices

data: Y
Chi-Sq (approx.) = 41.806, df = 30, p-value = 0.07436
RATR_MBox_G5log<-boxM(cbind(rhatr_Iron,rhatr_AD,rhatr_RD,rhatr_MD,rhatr_FA)~Gender,
                         data=RATR_avg_demo_WoOut_log, cov=TRUE);RATR_MBox_G5log

  Box's M-test for Homogeneity of Covariance Matrices

data: Y
Chi-Sq (approx.) = 15.371, df = 15, p-value = 0.4251
RATR_MBox_AGE5log<-boxM(cbind(rhatr_Iron,rhatr_AD,rhatr_RD,rhatr_MD,rhatr_FA)~agef2,
                           data=RATR_avg_demo_WoOut_log, cov=TRUE);RATR_MBox_AGE5log

  Box's M-test for Homogeneity of Covariance Matrices
```

```

data: Y
Chi-Sq (approx.) = 24.286, df = 30, p-value = 0.7589

# Generar los MANOVAS
RATR_manova5_1log<-manova(cbind(rhatr_Iron,rhatr_AD,rhatr_RD,rhatr_MD,rhatr_FA)~Subgroup*Gender*agef2,
                             data=RATR_avg_demo_WoOut_log)
summary(RATR_manova5_1log)
  Df Pillai approx F num Df den Df Pr(>F)
Subgroup      2 0.40774  1.74132     10    68 0.08912 .
Gender        1 0.02189  0.14772      5    33 0.97933
agef2         2 0.33305  1.35863     10    68 0.21847
Subgroup:Gender 2 0.15917  0.58798     10    68 0.81826
Subgroup:agef2  4 0.61871  1.31746     20   144 0.17697
Gender:agef2   2 0.31469  1.26973     10    68 0.26504
Subgroup:Gender:agef2 2 0.16882  0.62690     10    68 0.78577
Residuals      37
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RATR_manova5_2log<-manova(cbind(rhatr_Iron,rhatr_AD,rhatr_RD,rhatr_MD,rhatr_FA)~Subgroup+Gender+agef2,
                             data=RATR_avg_demo_WoOut_log)
summary(RATR_manova5_2log)
  Df Pillai approx F num Df den Df Pr(>F)
Subgroup      2 0.36840  1.98693     10    88 0.04415 *
Gender        1 0.01831  0.16037      5    43 0.97562
agef2         2 0.26005  1.31523     10    88 0.23494
Residuals      47
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RATR_manova5_3log<-manova(cbind(rhatr_Iron,rhatr_AD,rhatr_RD,rhatr_MD,rhatr_FA)~Subgroup,
                             data=RATR_avg_demo_WoOut_log)
summary(RATR_manova5_3log)
  Df Pillai approx F num Df den Df Pr(>F)
Subgroup      2 0.3475   1.9767     10    94 0.04447 *
Residuals      50
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Generar effect size
library(effectsize)
RATR_SizeMan5_2log<-effectsize::eta_squared(RATR_manova5_2log);RATR_SizeMan5_2log
# Effect Size for ANOVA (Type I)

Parameter | Eta2 (partial) |      95% CI
-----
Subgroup |          0.18 | [0.00, 1.00]
Gender   |          0.02 | [0.00, 1.00]
agef2   |          0.13 | [0.00, 1.00]

- One-sided CIs: upper bound fixed at [1.00].
RATR_SizeMan5_3log<-effectsize::eta_squared(RATR_manova5_3log);RATR_SizeMan5_3log
# Effect Size for ANOVA (Type I)

Parameter | Eta2 (partial) |      95% CI
-----
Subgroup |          0.17 | [0.00, 1.00]

```

- One-sided CIs: upper bound fixed at [1.00].

En el caso de las **5 métricas conjuntas con transformación logarítmica de las 3 variables de difusividad**, se objetiva que subgrupo es el único factor significativo (p-valor de 0.044). con un tamaño de efecto de 0.17. Dicha transformación no condiciona una pérdida de la normalidad multivariante ni de la homocedasticidad de la matriz de covarianzas.

```
# Para las 3 métricas AD-RD-MD - Sin transformación de variables
# Generar los MANOVAS
RATR_manova3_1<-manova(cbind(rhatr_AD,rhatr_RD,rhatr_MD)~Subgroup*Gender*agef2,
                           data=RATR_avg_demo_WoOut_ADRDMD)
summary(RATR_manova3_1)
Error in summary.manova(RATR_manova3_1): residuals have rank 2 < 3
summary(RATR_manova3_1,tol=0)
      Df Pillai approx F num Df den Df Pr(>F)
Subgroup        2 0.32080  2.35623     6    74 0.03881 *
Gender          1 0.01344  0.16351     3    36 0.92023
agef2           2 0.11882  0.77898     6    74 0.58903
Subgroup:Gender 2 0.06852  0.43756     6    74 0.85146
Subgroup:agef2   4 0.18138  0.61134    12   114 0.82888
Gender:agef2     2 0.06127  0.38979     6    74 0.88336
Subgroup:Gender:agef2 2 0.05642  0.35802     6    74 0.90301
Residuals       38
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RATR_manova3_2<-manova(cbind(rhatr_AD,rhatr_RD,rhatr_MD)~Subgroup+Gender+agef2,
                           data=RATR_avg_demo_WoOut_ADRDMD)
summary(RATR_manova3_2)
Error in summary.manova(RATR_manova3_2): residuals have rank 2 < 3
summary(RATR_manova3_2,tol=0)
      Df Pillai approx F num Df den Df Pr(>F)
Subgroup        2 0.300906  2.77454     6    94 0.01578 *
Gender          1 0.012677  0.19688     3    46 0.89799
agef2           2 0.112269  0.93175     6    94 0.47607
Residuals       48
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RATR_manova3_3<-manova(cbind(rhatr_AD,rhatr_RD,rhatr_MD)~Subgroup,
                           data=RATR_avg_demo_WoOut_ADRDMD)
summary(RATR_manova3_3)
Error in summary.manova(RATR_manova3_3): residuals have rank 2 < 3
summary(RATR_manova3_3,tol=0)
      Df Pillai approx F num Df den Df Pr(>F)
Subgroup        2 0.2864  2.7855     6   100 0.01511 *
Residuals       51
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

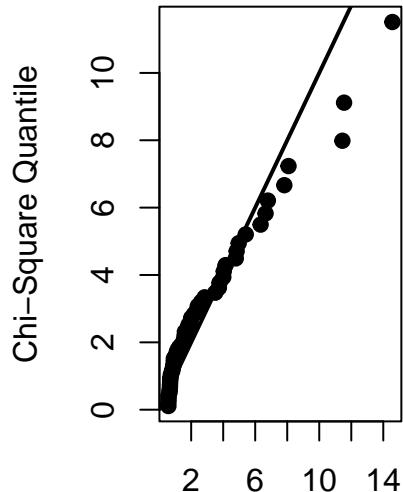
# Generar effect size
RATR_SizeMan3_2<-effectsize::eta_squared(RATR_manova3_2);RATR_SizeMan3_2
Error in summary.manova(model): residuals have rank 2 < 3
Error in eval(expr, envir, enclos): objeto 'RATR_SizeMan3_2' no encontrado
RATR_SizeMan3_3<-effectsize::eta_squared(RATR_manova3_3);RATR_SizeMan3_3
Error in summary.manova(model): residuals have rank 2 < 3
Error in eval(expr, envir, enclos): objeto 'RATR_SizeMan3_3' no encontrado
```

En el caso de las **3 métricas AD-RD-MD sin transformación logarítmica**, se objetiva que subgrupo es el único factor significativo; debido a que los residuos son rango deficientes, no se permite obtener el *effect size*.

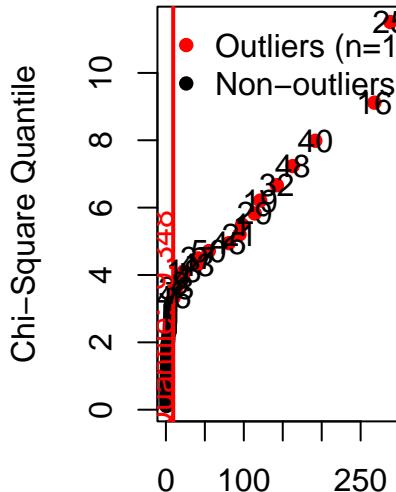
```
# Para las 3 métricas AD-RD-MD - Con transformación logarítmica de las métricas
# Primero transformación de las variables
RATR_avg_demo_WoOut_ADRDMD_log<-cbind(RATR_avg_demo_WoOut_ADRDMD[,1:7],
                                             log(cbind(RATR_avg_demo_WoOut_ADRDMD[,8:10])))

# Comprobar normalidad multivariante
par(mfrow = c(1, 2))
RATR_mvnllog3<-mvn(RATR_avg_demo_WoOut_ADRDMD_log[,-c(1:7)], mvnTest = 'royston',
                      covariance = TRUE, desc = TRUE, univariateTest = 'SW',
                      univariatePlot = 'qq', multivariatePlot = 'qq',
                      multivariateOutlierMethod = 'quan', showOutliers = TRUE,
                      showNewData = TRUE)
```

**Chi-Square Q-Q Plot**



**Chi-Square Q-Q Plot**



Squared Mahalanobis Distance      Robust Squared Mahalanobis Distance

```
RATR_mvnllog3$multivariateNormality
  Test      H   p value MVN
1 Royston 0.6581312 0.7396902 YES
RATR_mvnllog3$univariateNormality
  Test Variable Statistic   p value Normality
1 Shapiro-Wilk rhatr_AD    0.9794    0.4766    YES
2 Shapiro-Wilk rhatr_RD    0.9834    0.6557    YES
3 Shapiro-Wilk rhatr_MD    0.9828    0.6252    YES

# Comprobar homocedasticidad de la matriz de covarianzas
RATR_MBox_SG3log<-boxM(cbind(rhatr_AD,rhatr_RD,rhatr_MD)~Subgroup,
                         data=RATR_avg_demo_WoOut_ADRDMD_log, cov=TRUE); RATR_MBox_SG3log

Box's M-test for Homogeneity of Covariance Matrices
```

```

data: Y
Chi-Sq (approx.) = 22.781, df = 12, p-value = 0.02964
RATR_MBox_G3log<-boxM(cbind(rhatr_AD,rhatr_RD,rhatr_MD)~Gender,
                         data=RATR_avg_demo_WoOut_ADRDMD_log, cov=TRUE);RATR_MBox_G3log

  Box's M-test for Homogeneity of Covariance Matrices

data: Y
Chi-Sq (approx.) = 9.5094, df = 6, p-value = 0.1469
RATR_MBox_AGE3log<-boxM(cbind(rhatr_AD,rhatr_RD,rhatr_MD)~agef2,
                           data=RATR_avg_demo_WoOut_ADRDMD_log, cov=TRUE);RATR_MBox_AGE3log

  Box's M-test for Homogeneity of Covariance Matrices

data: Y
Chi-Sq (approx.) = 6.0424, df = 12, p-value = 0.9139

# Generar los MANOVAS
RATR_manova3_1log<-manova(cbind(rhatr_AD,rhatr_RD,rhatr_MD)~Subgroup*Gender*agef2,
                             data=RATR_avg_demo_WoOut_ADRDMD_log)
summary(RATR_manova3_1log)
  Df Pillai approx F num Df den Df Pr(>F)
Subgroup      2 0.34440  2.56559      6    74 0.02593 *
Gender        1 0.01721  0.21016      3    36 0.88870
agef2         2 0.09259  0.59871      6    74 0.73043
Subgroup:Gender 2 0.03356  0.21051      6    74 0.97243
Subgroup:agef2  4 0.19193  0.64932     12   114 0.79577
Gender:agef2   2 0.05750  0.36509      6    74 0.89876
Subgroup:Gender:agef2 2 0.05714  0.36275      6    74 0.90017
Residuals     38
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RATR_manova3_2log<-manova(cbind(rhatr_AD,rhatr_RD,rhatr_MD)~Subgroup+Gender+agef2,
                             data=RATR_avg_demo_WoOut_ADRDMD_log)
summary(RATR_manova3_2log)
  Df Pillai approx F num Df den Df Pr(>F)
Subgroup      2 0.33272  3.12640      6    94 0.007735 **
Gender        1 0.01651  0.25742      3    46 0.855653
agef2         2 0.08446  0.69078      6    94 0.657560
Residuals     48
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RATR_manova3_3log<-manova(cbind(rhatr_AD,rhatr_RD,rhatr_MD)~Subgroup,
                             data=RATR_avg_demo_WoOut_ADRDMD_log)
summary(RATR_manova3_3log)
  Df Pillai approx F num Df den Df Pr(>F)
Subgroup      2 0.31811  3.1523       6   100 0.007134 **
Residuals    51
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Generar effect size
RATR_SizeMan3_2log<-effectsize::eta_squared(RATR_manova3_2log);RATR_SizeMan3_2log
# Effect Size for ANOVA (Type I)

Parameter | Eta2 (partial) |      95% CI

```

```

Subgroup | 0.17 | [0.03, 1.00]
Gender | 0.02 | [0.00, 1.00]
agef2 | 0.04 | [0.00, 1.00]

- One-sided CIs: upper bound fixed at [1.00].
RATR_SizeMan3_3log<-effectsize::eta_squared(RATR_manova3_3log);RATR_SizeMan3_3log
# Effect Size for ANOVA (Type I)

Parameter | Eta2 (partial) | 95% CI
-----
Subgroup | 0.16 | [0.03, 1.00]

- One-sided CIs: upper bound fixed at [1.00].

```

En el caso de las **3 métricas AD-RD-MD con transformación logarítmica**, se objetiva que subgrupo es el único factor significativo, con un *effect size* de 0.16. Cabe destacar que la normalidad multivariante se mantiene, pero se pierde la homocedasticidad de la matriz de covarianzas para subgrupo, hecho que puede sesgar el test de MANOVA.

```

# Para las 2 métricas Iron-FA
# Generar los MANOVAS
RATR_manova2_1<-manova(cbind(rhatr_Iron,rhatr_FA)~Subgroup*Gender*agef2,
                         data=RATR_avg_demo_WoOut_IronFA)
summary(RATR_manova2_1)
      Df Pillai approx F num Df den Df Pr(>F)
Subgroup 2 0.143958 1.47367 4 76 0.2185
Gender 1 0.014850 0.27887 2 37 0.7582
agef2 2 0.056469 0.55204 4 76 0.6981
Subgroup:Gender 2 0.031209 0.30119 4 76 0.8763
Subgroup:agef2 4 0.246025 1.33253 8 76 0.2406
Gender:agef2 2 0.138095 1.40921 4 76 0.2390
Subgroup:Gender:agef2 2 0.151935 1.56205 4 76 0.1931
Residuals 38

RATR_manova2_2<-manova(cbind(rhatr_Iron,rhatr_FA)~Subgroup+Gender+agef2,
                         data=RATR_avg_demo_WoOut_IronFA)
summary(RATR_manova2_2)
      Df Pillai approx F num Df den Df Pr(>F)
Subgroup 2 0.120667 1.54098 4 96 0.1965
Gender 1 0.010156 0.24111 2 47 0.7867
agef2 2 0.041663 0.51060 4 96 0.7281
Residuals 48

RATR_manova2_3<-manova(cbind(rhatr_Iron,rhatr_FA)~Subgroup,
                         data=RATR_avg_demo_WoOut_IronFA)
summary(RATR_manova2_3)
      Df Pillai approx F num Df den Df Pr(>F)
Subgroup 2 0.11721 1.5875 4 102 0.1833
Residuals 51

```

En el caso de las **2 métricas Iron-FA**, ningún factor es significativo, por lo que no se calcula el *effect size*.

### 3.1.2 LATR

```

# Para las 5 métricas conjuntamente
# Generar los MANOVAS
LATR_manova5_1<-manova(cbind(lhatr_Iron,lhatr_AD,lhatr_RD,lhatr_MD,lhatr_FA)~Subgroup*Gender*agef2,
                         data=LATR_avg_demo_WoOut)
summary(LATR_manova5_1)
      Df Pillai approx F num Df den Df Pr(>F)
Subgroup        2 0.52060  2.46331     10    70 0.01381 *
Gender          1 0.22797  2.00796      5    34 0.10246
agef2           2 0.35032  1.48648     10    70 0.16289
Subgroup:Gender 2 0.11349  0.42111     10    70 0.93183
Subgroup:agef2   4 0.65648  1.45294     20   148 0.10700
Gender:agef2    2 0.35225  1.49646     10    70 0.15916
Subgroup:Gender:agef2 2 0.25431  1.01976     10    70 0.43608
Residuals       38
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
LATR_manova5_2<-manova(cbind(lhatr_Iron,lhatr_AD,lhatr_RD,lhatr_MD,lhatr_FA)~Subgroup+Gender+agef2,
                         data=LATR_avg_demo_WoOut)
summary(LATR_manova5_2)
      Df Pillai approx F num Df den Df Pr(>F)
Subgroup    2 0.44297  2.5604     10    90 0.009036 **
Gender      1 0.18265  1.9665      5    44 0.102554
agef2       2 0.28790  1.5134     10    90 0.147518
Residuals   48
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
LATR_manova5_3<-manova(cbind(lhatr_Iron,lhatr_AD,lhatr_RD,lhatr_MD,lhatr_FA)~Subgroup,
                         data=LATR_avg_demo_WoOut)
summary(LATR_manova5_3)
      Df Pillai approx F num Df den Df Pr(>F)
Subgroup    2 0.40692  2.4521     10    96 0.01182 *
Residuals   51
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Generar effect size
LATR_SizeMan5_2<-effectsize::eta_squared(RATR_manova5_2);LATR_SizeMan5_2
# Effect Size for ANOVA (Type I)

Parameter | Eta2 (partial) |      95% CI
-----
Subgroup |             0.17 | [0.00, 1.00]
Gender   |             0.02 | [0.00, 1.00]
agef2   |             0.14 | [0.00, 1.00]

- One-sided CIs: upper bound fixed at [1.00].
LATR_SizeMan5_3<-effectsize::eta_squared(RATR_manova5_3);LATR_SizeMan5_3
# Effect Size for ANOVA (Type I)

Parameter | Eta2 (partial) |      95% CI
-----
Subgroup |             0.16 | [0.00, 1.00]

- One-sided CIs: upper bound fixed at [1.00].

```

En el caso de las **5 métricas conjuntas**, se objetiva que subgrupo es el único factor significativo. con un tamaño de efecto de 0.16.

```
# Para las 3 métricas AD-RD-MD - Sin transformación de variables
# Generar los MANOVAS
LATR_manova3_1<-manova(cbind(lhatr_AD,lhatr_RD,lhatr_MD)~Subgroup*Gender*agef2,
                         data=LATR_avg_demo_WoOut_ADRDMD)
summary(LATR_manova3_1)
Error in summary.manova(LATR_manova3_1): residuals have rank 2 < 3
summary(LATR_manova3_1,tol=0)
      Df Pillai approx F num Df den Df Pr(>F)
Subgroup        2 0.41200   3.1998     6    74 0.007556 ** 
Gender          1 0.16918   2.4436     3    36 0.079866 .  
agef2           2 0.22640   1.5744     6    74 0.166616
Subgroup:Gender 2 0.06757   0.4312     6    74 0.855830
Subgroup:agef2  4 0.28907   1.0130    12   114 0.441824
Gender:agef2    2 0.15127   1.0091     6    74 0.426034
Subgroup:Gender:agef2 2 0.11837   0.7759     6    74 0.591396
Residuals       38
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
LATR_manova3_2<-manova(cbind(lhatr_AD,lhatr_RD,lhatr_MD)~Subgroup+Gender+agef2,
                         data=LATR_avg_demo_WoOut_ADRDMD)
summary(LATR_manova3_2)
Error in summary.manova(LATR_manova3_2): residuals have rank 2 < 3
summary(LATR_manova3_2,tol=0)
      Df Pillai approx F num Df den Df Pr(>F)
Subgroup        2 0.36002   3.4392     6    94 0.004092 ** 
Gender          1 0.14216   2.5410     3    46 0.067914 .
agef2           2 0.17682   1.5194     6    94 0.180229
Residuals       48
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
LATR_manova3_3<-manova(cbind(lhatr_AD,lhatr_RD,lhatr_MD)~Subgroup,
                         data=LATR_avg_demo_WoOut_ADRDMD)
summary(LATR_manova3_3)
Error in summary.manova(LATR_manova3_3): residuals have rank 2 < 3
summary(LATR_manova3_3,tol=0)
      Df Pillai approx F num Df den Df Pr(>F)
Subgroup        2 0.33171   3.3139     6   100 0.005118 ** 
Residuals       51
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Generar effect size
LATR_SizeMan3_2<-effectsize::eta_squared(LATR_manova3_2);LATR_SizeMan3_2
Error in summary.manova(model): residuals have rank 2 < 3
Error in eval(expr, envir, enclos): objeto 'LATR_SizeMan3_2' no encontrado
LATR_SizeMan3_3<-effectsize::eta_squared(LATR_manova3_3);LATR_SizeMan3_3
Error in summary.manova(model): residuals have rank 2 < 3
Error in eval(expr, envir, enclos): objeto 'LATR_SizeMan3_3' no encontrado
```

En el caso de las **3 métricas AD-RD-MD sin transformación logarítmica**, se objetiva que subgrupo es el único factor significativo; debido a que los residuos son rango deficientes, no se permite obtener el *effect size*.

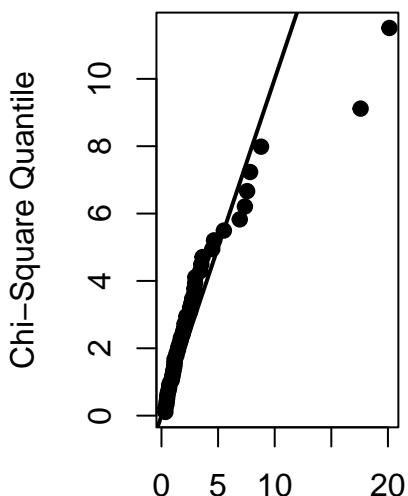
```

# Para las 3 métricas AD-RD-MD - Con transformación logarítmica de las métricas
# Primero transformación de las variables
LATR_avg_demo_WoOut_ADRDMD_log<-cbind(LATR_avg_demo_WoOut_ADRDMD[,1:7],
                                           log(cbind(LATR_avg_demo_WoOut_ADRDMD[,8:10])))

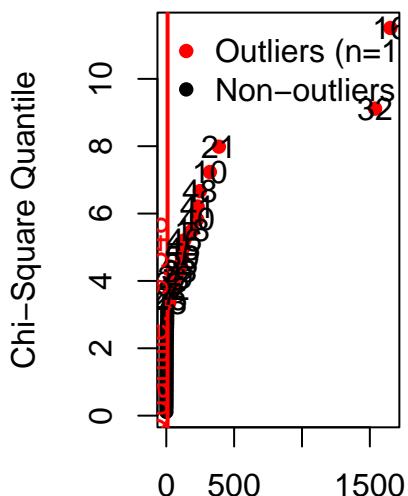
# Comprobar normalidad multivariante
par(mfrow = c(1, 2))
LATR_mvnllog3<-mvn(LATR_avg_demo_WoOut_ADRDMD_log[-c(1:7)], mvnTest = 'royston',
                      covariance = TRUE, desc = TRUE, univariateTest = 'SW',
                      univariatePlot = 'qq', multivariatePlot = 'qq',
                      multivariateOutlierMethod = 'quan', showOutliers = TRUE, showNewData = TRUE)

```

**Chi-Square Q-Q Plot**



**Chi-Square Q-Q Plot**



Squared Mahalanobis Distance      Robust Squared Mahalanobis Distance

```

LATR_mvnllog3$multivariateNormality
  Test      H   p value MVN
1 Royston 0.03854733 0.9836486 YES
LATR_mvnllog3$univariateNormality
  Test Variable Statistic   p value Normality
1 Shapiro-Wilk lhattr_AD    0.9914    0.9651     YES
2 Shapiro-Wilk lhattr_RD    0.9882    0.8719     YES
3 Shapiro-Wilk lhattr_MD    0.9881    0.8672     YES

# Comprobar homocedasticidad de la matriz de covarianzas
LATR_MBox_SG3log<-boxM(cbind(lhattr_AD,lhattr_RD,lhattr_MD)~Subgroup,
                         data=LATR_avg_demo_WoOut_ADRDMD_log, cov=TRUE);LATR_MBox_SG3log

```

Box's M-test for Homogeneity of Covariance Matrices

```

data: Y
Chi-Sq (approx.) = 26.023, df = 12, p-value = 0.01065
LATR_MBox_G3log<-boxM(cbind(lhattr_AD,lhattr_RD,lhattr_MD)~Gender,

```

```

data=LATR_avg_demo_WoOut_ADRDMD_log, cov=TRUE);LATR_MBox_G3log

Box's M-test for Homogeneity of Covariance Matrices

data: Y
Chi-Sq (approx.) = 10.92, df = 6, p-value = 0.09087
LATR_MBox_AGE3log<-boxM(cbind(lhatr_AD,lhatr_RD,lhatr_MD)~agef2,
                           data=LATR_avg_demo_WoOut_ADRDMD_log, cov=TRUE);LATR_MBox_AGE3log

Box's M-test for Homogeneity of Covariance Matrices

data: Y
Chi-Sq (approx.) = 26.496, df = 12, p-value = 0.009126

# Generar los MANOVAS
LATR_manova3_1log<-manova(cbind(lhatr_AD,lhatr_RD,lhatr_MD)~Subgroup*Gender*agef2,
                             data=LATR_avg_demo_WoOut_ADRDMD_log)
summary(LATR_manova3_1log)
      Df Pillai approx F num Df den Df Pr(>F)
Subgroup        2 0.37964  2.88962      6    74 0.01382 *
Gender          1 0.06817  0.87788      3    36 0.46163
agef2           2 0.18115  1.22835      6    74 0.30165
Subgroup:Gender 2 0.07402  0.47402      6    74 0.82555
Subgroup:agef2   4 0.22752  0.77960     12   114 0.67027
Gender:agef2     2 0.15204  1.01473      6    74 0.42245
Subgroup:Gender:agef2 2 0.06820  0.43539      6    74 0.85297
Residuals       38
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
LATR_manova3_2log<-manova(cbind(lhatr_AD,lhatr_RD,lhatr_MD)~Subgroup+Gender+agef2,
                             data=LATR_avg_demo_WoOut_ADRDMD_log)
summary(LATR_manova3_2log)
      Df Pillai approx F num Df den Df Pr(>F)
Subgroup      2 0.34548  3.2714      6    94 0.005759 **
Gender        1 0.05967  0.9730      3    46 0.413639
agef2         2 0.15948  1.3575      6    94 0.239939
Residuals     48
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
LATR_manova3_3log<-manova(cbind(lhatr_AD,lhatr_RD,lhatr_MD)~Subgroup,
                             data=LATR_avg_demo_WoOut_ADRDMD_log)
summary(LATR_manova3_3log)
      Df Pillai approx F num Df den Df Pr(>F)
Subgroup      2 0.32215  3.2001      6   100 0.006468 **
Residuals    51
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Generar effect size
LATR_SizeMan3_2log<-effectsize::eta_squared(LATR_manova3_2log);LATR_SizeMan3_2log
# Effect Size for ANOVA (Type I)

Parameter | Eta2 (partial) |      95% CI
-----
Subgroup |             0.17 | [0.03, 1.00]
Gender   |             0.06 | [0.00, 1.00]

```

```

agef2 | 0.08 | [0.00, 1.00]

- One-sided CIs: upper bound fixed at [1.00].
LATR_SizeMan3_3log<-effectsize::eta_squared(LATR_manova3_3log);LATR_SizeMan3_3log
# Effect Size for ANOVA (Type I)

Parameter | Eta2 (partial) | 95% CI
-----
Subgroup | 0.16 | [0.03, 1.00]

- One-sided CIs: upper bound fixed at [1.00].

```

En el caso de las **3 métricas AD-RD-MD con transformación logarítmica**, se objetiva que subgrupo es el único factor significativo, con un *effect size* de 0.16. Cabe destacar que la normalidad multivariante se mantiene, pero se pierde la homocedasticidad de la matriz de covarianzas para subgrupo y edad factorizada, hecho que puede sesgar el test de MANOVA.

```

# Para las 2 métricas Iron-FA
# Generar los MANOVAS
LATR_manova2_1<-manova(cbind(lhatr_Iron,lhatr_FA)~Subgroup*Gender*agef2,
                         data=LATR_avg_demo_WoOut_IronFA)
summary(LATR_manova2_1)
      Df Pillai approx F num Df den Df Pr(>F)
Subgroup        2 0.15152   1.5574     4    76 0.194320
Gender          1 0.01976   0.3730     2    37 0.691258
agef2           2 0.09105   0.9062     4    76 0.464753
Subgroup:Gender 2 0.06744   0.6630     4    76 0.619620
Subgroup:agef2  4 0.21221   1.1276     8    76 0.354726
Gender:agef2    2 0.31818   3.5945     4    76 0.009737 ** 
Subgroup:Gender:agef2 2 0.11354   1.1435     4    76 0.342597
Residuals       38
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
LATR_manova2_2<-manova(cbind(lhatr_Iron,lhatr_FA)~Subgroup+Gender:agef2,
                         data=LATR_avg_demo_WoOut_IronFA)
summary(LATR_manova2_2)
      Df Pillai approx F num Df den Df Pr(>F)
Subgroup        2 0.12921   1.5885     4    92 0.18399
Gender:agef2    5 0.35796   2.0056    10    92 0.04141 *
Residuals       46
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
LATR_manova2_3<-manova(cbind(lhatr_Iron,lhatr_FA)~Subgroup,
                         data=LATR_avg_demo_WoOut_IronFA)
summary(LATR_manova2_3)
      Df Pillai approx F num Df den Df Pr(>F)
Subgroup        2 0.11524   1.5591     4    102 0.1909
Residuals       51

# Generar effect size
LATR_SizeMan2_2<-effectsize::eta_squared(LATR_manova2_2);LATR_SizeMan2_2
# Effect Size for ANOVA (Type I)

Parameter | Eta2 (partial) | 95% CI
-----
Subgroup | 0.06 | [0.00, 1.00]

```

```
Gender:agef2 | 0.18 | [0.00, 1.00]
```

```
- One-sided CIs: upper bound fixed at [1.00].
```

En el caso de las **2 métricas Iron-FA**, la interacción género-edad factorizada es significativa, con un tamaño de efecto del 0.18.

### 3.1.3 Comparaciones pareadas

Se genera un **Análisis Discriminante Lineal (LDA)**, el cual tiene como objetivo encontrar qué combinación lineal de variables individuales dependientes (*features*) conduce a la máxima separabilidad de subgrupos e interpretar esta combinación lineal. Se utiliza la función `lda` del paquete MASS.

Se van a analizar aquellos MANOVA que han detectado el subgrupo como factor significativo y se van a utilizar los dataframes con transformaciones logarítmicas dado que sin dichas transformaciones la función `lda` expone que las variables de difusividad media, radial y axial son constantes (probablemente debido a sus valores muy próximos a cero), como se verá a continuación. En el caso del ATR izquierdo para las 5 métricas, se realiza la transformación logarítmica de las 3 métricas de difusividad y se comprueba la normalidad multivariante como la homocedasticidad de la matriz de covarianzas, dado que no se había ejecutado en el MANOVA anterior (no se ejecutó porque los residuos no eran rango-deficientes).

Dos de las premisas del LDA es que los *features* de cada subgrupo o clase sigan una distribución normal y que las matrices de covarianzas sean iguales, por tanto, se van a utilizar los dataframes reducidos (con la eliminación de outliers).

Los pasos son: generación de los datasets de entrenamiento y de testeo, generación del modelo LDA con el dataset de entrenamiento, predicción de las clases o de las probabilidades posteriores del dataset de testeo, y evaluación del modelo mediante curvas ROC, matriz de confusión o coeficiente kappa.

La interpretación del coeficiente kappa es la siguiente según Fleiss et al. 2003 (KAPPA COEFFICIENT INTERPRETATION):

- Valores mayores de 0.75 tienen un excelente acuerdo;
- Valores menores de 0.40 tienen un pobre acuerdo;
- Valores entre 0.40 y 0.75 tiene un acuerdo de débil a bueno.

Según McHugh 2012:

Value of K	Level of agreement	% of data that are reliable
0 - 0.20	None	0 - 4%
0.21 - 0.39	Minimal	4 - 15%
0.40 - 0.59	Weak	15 - 35%
0.60 - 0.79	Moderate	35 - 63%
0.80 - 0.90	Strong	64 - 81%
Above 0.90	Almost Perfect	82 - 100%

### Resultados para el RATR

```
library(caret)
library(MASS)
library(pROC)
library(multiROC)
# Para las 5 métricas

# Primero, división del dataframe sin transformación logarítmica en los datasets
```

```

# de entrenamiento y testeo
set.seed(123456)
train_index<-createDataPartition(RATR_avg_demo_WoOut$Subgroup, p=0.8, list=FALSE)
train<-RATR_avg_demo_WoOut[train_index,]
test<-RATR_avg_demo_WoOut[-train_index,]

# Segundo, se genera el modelo LDA
RATR_LDA5model<-lda(Subgroup~cbind(rhatr_Iron,rhatr_AD,rhatr_RD,rhatr_MD,rhatr_FA),
                      data=train)
Error in lda.default(x, grouping, ...): variables 2 3 4 appear to be constant within groups

# Se repiten el primer y segundo paso con el dataframe con transformación logarítmica
set.seed(123456)

train_index<-createDataPartition(RATR_avg_demo_WoOut_log$Subgroup, p=0.8, list=FALSE)
train<-RATR_avg_demo_WoOut_log[train_index,]
test<-RATR_avg_demo_WoOut_log[-train_index,]

RATR_LDAlog5model<-lda(Subgroup~cbind(rhatr_Iron,rhatr_AD,rhatr_RD,rhatr_MD,rhatr_FA),
                         data=train)
print(RATR_LDAlog5model)
Call:
lda(Subgroup ~ cbind(rhatr_Iron, rhatr_AD, rhatr_RD, rhatr_MD,
                      rhatr_FA), data = train)

Prior probabilities of groups:
Control      PreHD       HD
0.4545455  0.2272727  0.3181818

Group means:
            cbind(rhatr_Iron, rhatr_AD, rhatr_RD, rhatr_MD, rhatr_FA)rhatr_Iron
Control                               0.02364404
PreHD                                0.02368070
HD                                     0.02300306
            cbind(rhatr_Iron, rhatr_AD, rhatr_RD, rhatr_MD, rhatr_FA)rhatr_AD
Control                               -7.005847
PreHD                                -7.005238
HD                                     -6.955647
            cbind(rhatr_Iron, rhatr_AD, rhatr_RD, rhatr_MD, rhatr_FA)rhatr_RD
Control                               -7.592971
PreHD                                -7.618293
HD                                     -7.542063
            cbind(rhatr_Iron, rhatr_AD, rhatr_RD, rhatr_MD, rhatr_FA)rhatr_MD
Control                               -7.356628
PreHD                                -7.369588
HD                                     -7.306095
            cbind(rhatr_Iron, rhatr_AD, rhatr_RD, rhatr_MD, rhatr_FA)rhatr_FA
Control                               0.3778487
PreHD                                 0.3903923
HD                                    0.3768309

Coefficients of linear discriminants:
                                         LD1
cbind(rhatr_Iron, rhatr_AD, rhatr_RD, rhatr_MD, rhatr_FA)rhatr_Iron -62.54342
cbind(rhatr_Iron, rhatr_AD, rhatr_RD, rhatr_MD, rhatr_FA)rhatr_AD -112.17002
cbind(rhatr_Iron, rhatr_AD, rhatr_RD, rhatr_MD, rhatr_FA)rhatr_RD -106.12930

```

```

cbind(rhatr_Iron, rhatr_AD, rhatr_RD, rhatr_MD, rhatr_FA)rhatr_MD    242.79516
cbind(rhatr_Iron, rhatr_AD, rhatr_RD, rhatr_MD, rhatr_FA)rhatr_FA    17.44913
                                         LD2
cbind(rhatr_Iron, rhatr_AD, rhatr_RD, rhatr_MD, rhatr_FA)rhatr_Iron  33.47382
cbind(rhatr_Iron, rhatr_AD, rhatr_RD, rhatr_MD, rhatr_FA)rhatr_AD  358.04029
cbind(rhatr_Iron, rhatr_AD, rhatr_RD, rhatr_MD, rhatr_FA)rhatr_RD  481.93784
cbind(rhatr_Iron, rhatr_AD, rhatr_RD, rhatr_MD, rhatr_FA)rhatr_MD -845.19078
cbind(rhatr_Iron, rhatr_AD, rhatr_RD, rhatr_MD, rhatr_FA)rhatr_FA  39.51893

```

Proportion of trace:

```

LD1   LD2
0.9028 0.0972

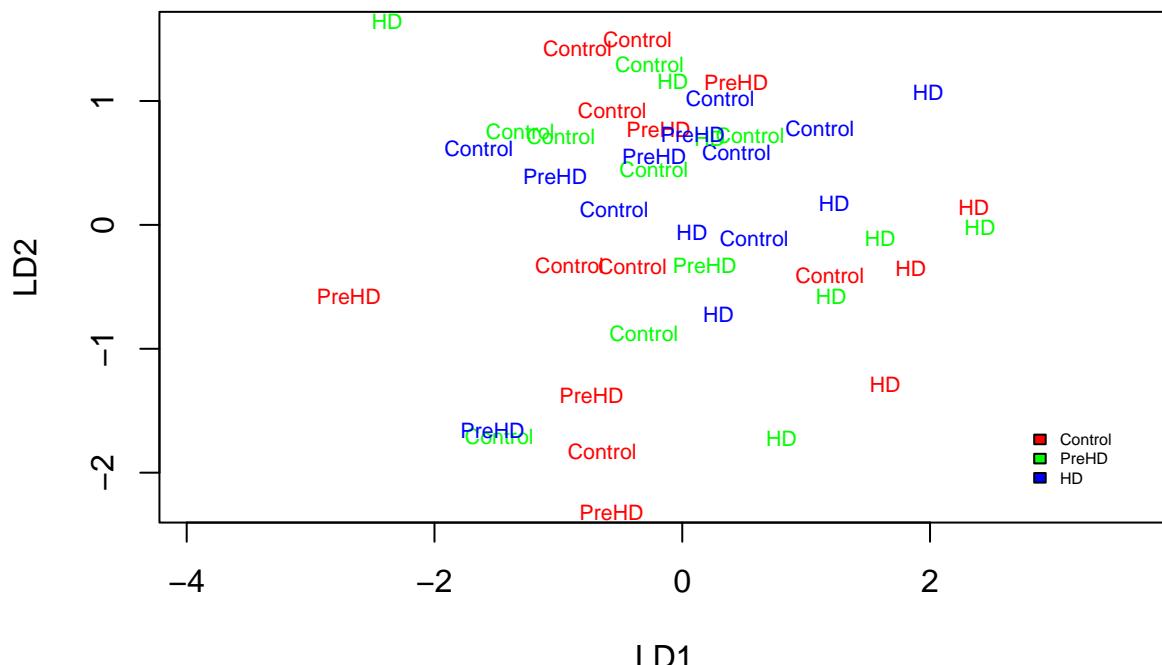
```

```

# Se genera el plot
## Define colors for each subgroup
subgroup_colors <- c("Control" = "red", "PreHD" = "green", "HD" = "blue")
## Plot the model with custom colors for species
plot(RATR_LDAlog5model, col = subgroup_colors,
      main="Visualization of 5log-LDA model of the right ATR")
legend('bottomright', inset=0.05, bty='n', cex=0.5, legend=c('Control','PreHD','HD'),
       fill=c('red','green','blue'))

```

## Visualization of 5log-LDA model of the right ATR



```

# Se predicen las clases
predClas_RATRLDA5<-predict(RATR_LDAlog5model, newdata = test, type='class')
test_subgroup<-as.character(test$Subgroup)
test_subgroup<-factor(test_subgroup, levels = levels(predClas_RATRLDA5$class))
table(predClas_RATRLDA5$class, test_subgroup)

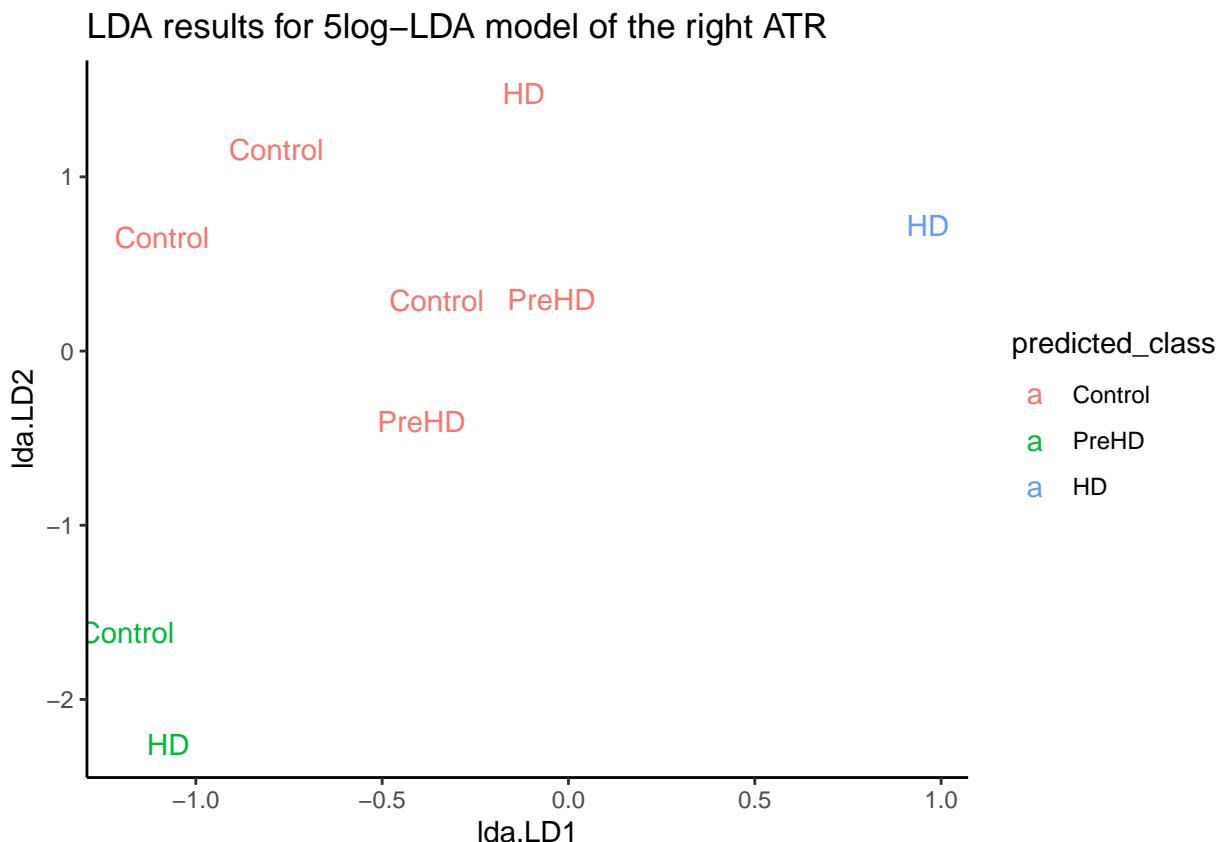
```

```

test_subgroup
  Control PreHD HD
Control      3     2   1
PreHD        1     0   1
HD          0     0   1

RATR_LDA5log_DF<-data.frame(
  predicted_class=predClas_RATRLDA5$class,
  subgroup=test[, 'Subgroup'],
  lda=predClas_RATRLDA5$x
)
str(RATR_LDA5log_DF)
'data.frame': 9 obs. of 4 variables:
$ predicted_class: Factor w/ 3 levels "Control","PreHD",...: 1 1 1 2 2 1 1 3 1
$ subgroup       : Factor w/ 3 levels "Control","PreHD",...: 1 1 1 1 3 3 2 3 2
$ lda.LD1        : num -1.091 -0.354 -0.784 -1.185 -1.073 ...
$ lda.LD2        : num 0.655 0.292 1.159 -1.615 -2.261 ...
ggplot(RATR_LDA5log_DF) +
  geom_text(aes(x=lda.LD1,y=lda.LD2,label=subgroup, color = predicted_class)) +
  theme_classic() + ggtitle('LDA results for 5log-LDA model of the right ATR')

```



```

# Evaluación del modelo

## Accuracy
accuracyRATR_LDAlog5<-sum(diag(table(predClas_RATRLDA5$class, test$Subgroup)))/nrow(test)
accuracyRATR_LDAlog5
[1] 0.4444444

```

```

## Confusion Matrix
confMX_RATR_LDA5log<-confusionMatrix(predClas_RATRLDA5$class, test$Subgroup)
confMX_RATR_LDA5log
Confusion Matrix and Statistics

Reference
Prediction Control PreHD HD
Control      3     2   1
PreHD        1     0   1
HD           0     0   1

Overall Statistics

    Accuracy : 0.4444
    95% CI : (0.137, 0.788)
    No Information Rate : 0.4444
    P-Value [Acc > NIR] : 0.6257

    Kappa : 0.1

McNemar's Test P-Value : 0.5062

Statistics by Class:

          Class: Control Class: PreHD Class: HD
Sensitivity          0.7500     0.0000   0.3333
Specificity          0.4000     0.7143   1.0000
Pos Pred Value       0.5000     0.0000   1.0000
Neg Pred Value       0.6667     0.7143   0.7500
Prevalence           0.4444     0.2222   0.3333
Detection Rate       0.3333     0.0000   0.1111
Detection Prevalence 0.6667     0.2222   0.1111
Balanced Accuracy    0.5750     0.3571   0.6667

## ROC curves one vs the rest by multiROC package
Control_pred_RATRLDA5log<- predClas_RATRLDA5$posterior[, 'Control']
PreHD_pred_RATRLDA5log<- predClas_RATRLDA5$posterior[, 'PreHD']
HD_pred_RATRLDA5log<- predClas_RATRLDA5$posterior[, 'HD']
Control_true<-ifelse(test$Subgroup=='Control',1,0)
PreHD_true<-ifelse(test$Subgroup=='PreHD',1,0)
HD_true<-ifelse(test$Subgroup=='HD',1,0)
DF_ROC_RATRLDA5log<-cbind(Control_true,PreHD_true,HD_true,Control_pred_RATRLDA5log,
                               PreHD_pred_RATRLDA5log,HD_pred_RATRLDA5log)
multiROC_RATR_LDA5log<-multi_roc(DF_ROC_RATRLDA5log)
multiROC_RATR_LDA5log_DF<-plot_roc_data(multiROC_RATR_LDA5log)
ggplot(multiROC_RATR_LDA5log_DF, aes(x=1-Specificity, y=Sensitivity, color = Group)) +
  geom_step() + geom_point() + labs(title= "5-LDA model of RATR ROC curve: One vs Rest",
  x = "False Positive Rate (1-Specificity)",
  y = "True Positive Rate (Sensitivity)") + geom_abline(intercept = 0, slope=1,linetype = 'dotted') +
  annotate('text',x=0.75, y=0.25,
  label=paste('AUC for Controls ROC curve = ',
  mean(multiROC_RATR_LDA5log_DF$AUC[multiROC_RATR_LDA5log_DF$Group=='Control'])),
  color = '#FF9900',size = 8/.pt) +
  annotate('text',x=0.75, y=0.20,
  label=paste('AUC for PreHD ROC curve = ',

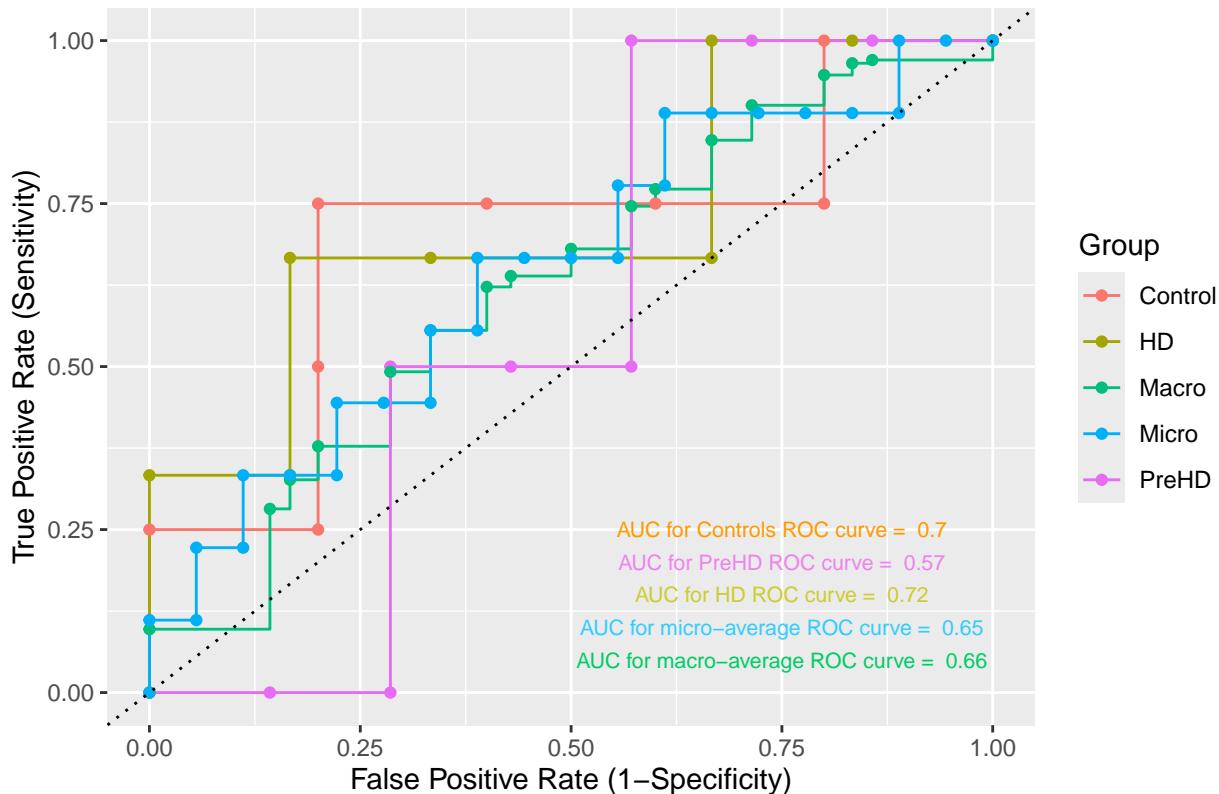
```

```

        round(mean(multiROC_RATR_LDA5log_DF$AUC[multiROC_RATR_LDA5log_DF$Group=='PreHD']),2)
color = 'violet',size = 8/.pt) +
annotate('text',x=0.75, y=0.15,
label=paste('AUC for HD ROC curve = ',
round(mean(multiROC_RATR_LDA5log_DF$AUC[multiROC_RATR_LDA5log_DF$Group=='HD']),2)),
color = '#CCCC33',size = 8/.pt) +
annotate('text',x=0.75, y=0.10,
label=paste('AUC for micro-average ROC curve = ',
round(mean(multiROC_RATR_LDA5log_DF$AUC[multiROC_RATR_LDA5log_DF$Group=='Micro']),2),
color = '#33CCFF',size = 8/.pt)+
annotate('text',x=0.75, y=0.05,
label=paste('AUC for macro-average ROC curve = ',
round(mean(multiROC_RATR_LDA5log_DF$AUC[multiROC_RATR_LDA5log_DF$Group=='Macro']),2),
color = '#00CC66',size = 8/.pt)

```

### 5-LDA model of RATR ROC curve: One vs Rest



```

## Kappa coefficient
library(irr)
kappa_RATR_LDA5log<-kappam.fleiss(data.frame(predClas_RATRLDA5$class,test$Subgroup))
kappa_RATR_LDA5log
Fleiss' Kappa for m Raters

Subjects = 9
Raters = 2
Kappa = 0.0625

z = 0.257
p-value = 0.797

```

```

library(MASS)
# Para las 3 métricas

# Primero, división del dataframe sin transformación logarítmica en los datasets
# de entrenamiento y testeo
set.seed(123456)
train_index<-createDataPartition(RATR_avg_demo_WoOut_ADRDMD$Subgroup, p=0.8, list=FALSE)
train<-RATR_avg_demo_WoOut_ADRDMD[train_index,]
test<-RATR_avg_demo_WoOut_ADRDMD[-train_index]

# Segundo, se genera el modelo LDA
RATR_LDA3model<-lda(Subgroup~cbind(rhatr_AD,rhatr_RD,rhatr_MD), data=train)
Error in lda.default(x, grouping, ...): variables 1 2 3 appear to be constant within groups

# Se repiten el primer y segundo paso con el dataframe con transformación logarítmica
set.seed(123456)

train_index<-createDataPartition(RATR_avg_demo_WoOut_ADRDMD_log$Subgroup, p=0.8, list=FALSE)
train<-RATR_avg_demo_WoOut_ADRDMD_log[train_index,]
test<-RATR_avg_demo_WoOut_ADRDMD_log[-train_index,]

RATR_LDAlog3model<-lda(Subgroup~cbind(rhatr_AD,rhatr_RD,rhatr_MD), data=train)
print(RATR_LDAlog3model)
Call:
lda(Subgroup ~ cbind(rhatr_AD, rhatr_RD, rhatr_MD), data = train)

Prior probabilities of groups:
  Control      PreHD       HD 
0.4444444 0.2444444 0.3111111 

Group means:
  cbind(rhatr_AD, rhatr_RD, rhatr_MD)rhatr_AD
Control                  -7.005847
PreHD                   -7.013681
HD                      -6.955647
  cbind(rhatr_AD, rhatr_RD, rhatr_MD)rhatr_RD
Control                  -7.592971
PreHD                   -7.623463
HD                      -7.542063
  cbind(rhatr_AD, rhatr_RD, rhatr_MD)rhatr_MD
Control                  -7.356628
PreHD                   -7.376339
HD                      -7.306095

Coefficients of linear discriminants:
                               LD1          LD2
cbind(rhatr_AD, rhatr_RD, rhatr_MD)rhatr_AD -78.42251  183.6945
cbind(rhatr_AD, rhatr_RD, rhatr_MD)rhatr_RD -85.46848  249.2872
cbind(rhatr_AD, rhatr_RD, rhatr_MD)rhatr_MD 187.94599 -438.4480

Proportion of trace:
LD1     LD2
0.9425 0.0575

# Se genera el plot
##Define colors for each subgroup

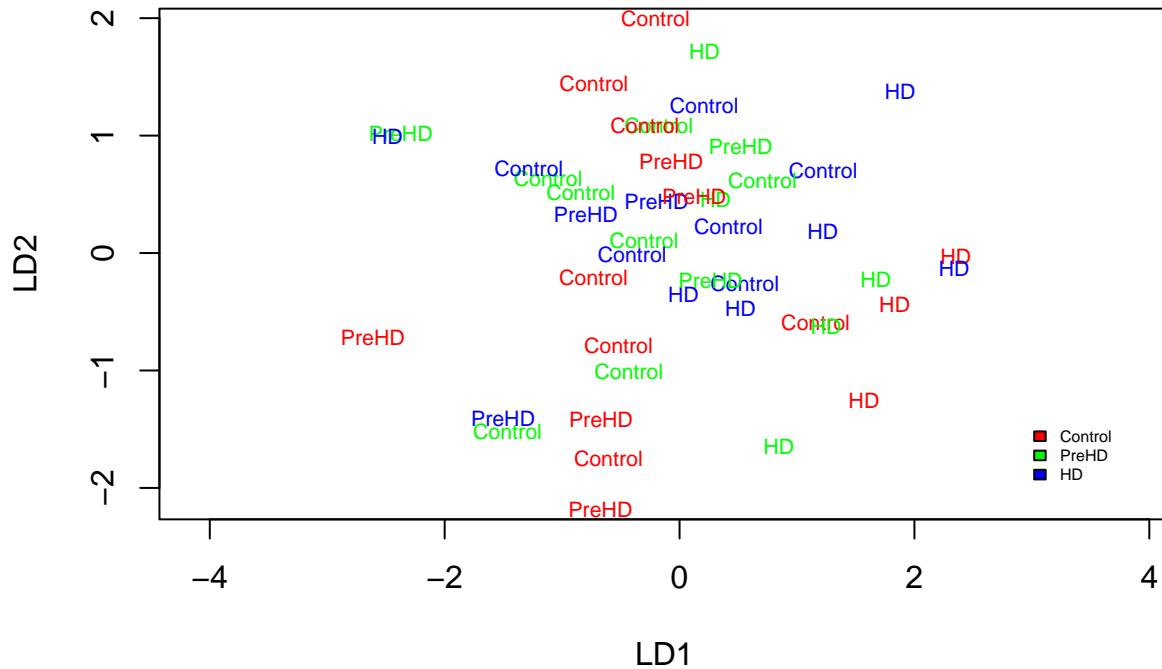
```

```

subgroup_colors <- c("Control" = "red", "PreHD" = "green", "HD" = "blue")
## Plot the model with custom colors for species
plot(RATR_LDAlgo3model, col = subgroup_colors,
      main="Visualization of 3log-LDA model of the right ATR")
legend('bottomright',inset=0.05, bty='n', cex=0.5, legend=c('Control','PreHD','HD'),
      fill=c('red','green','blue'))

```

## Visualization of 3log-LDA model of the right ATR



```

# Se predicen las clases
predClas_RATRLDA3<-predict(RATR_LDAlgo3model, newdata = test, type='class')
test_subgroup<-as.character(test$Subgroup)
test_subgroup<-factor(test_subgroup, levels = levels(predClas_RATRLDA3$class))
table(predClas_RATRLDA3$class, test_subgroup)
  test_subgroup
    Control PreHD HD
Control      3   2   1
PreHD        1   0   1
HD           0   0   1

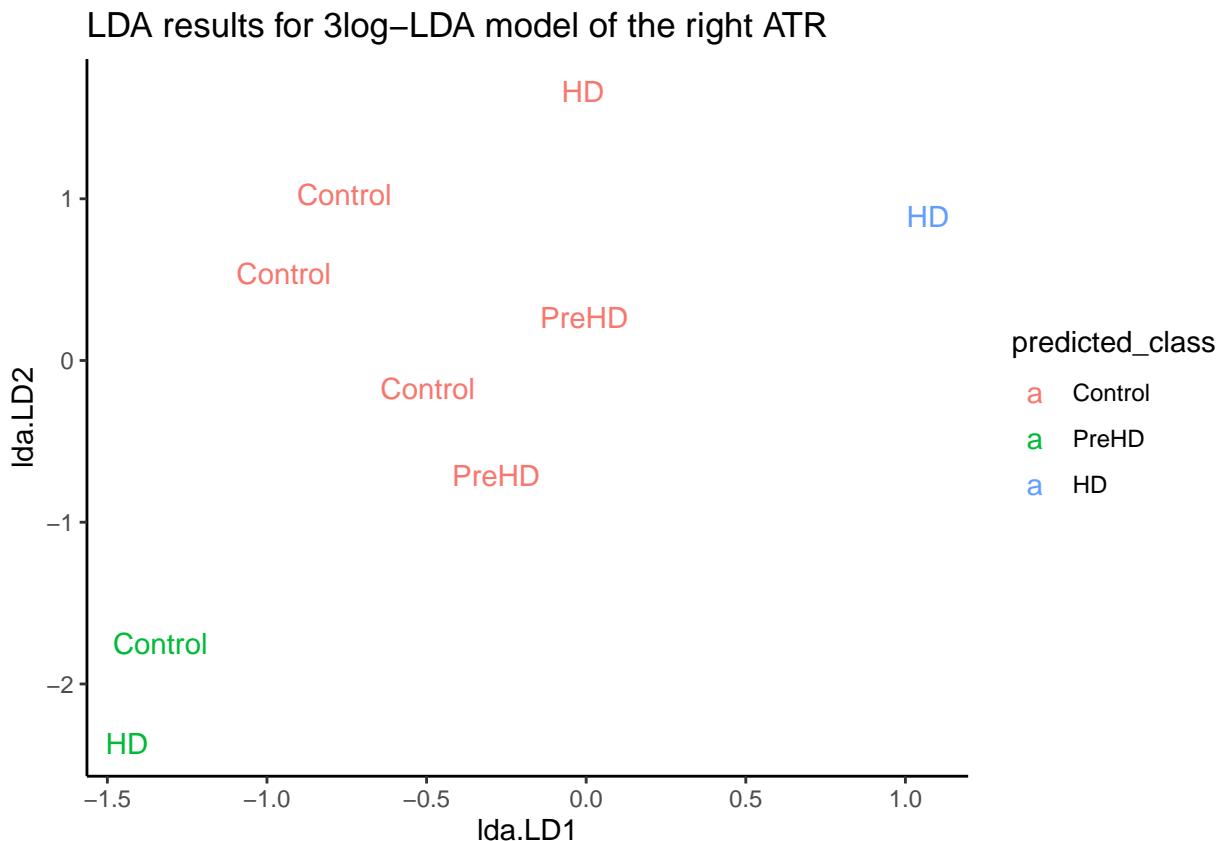
RATR_LDA3log_DF<-data.frame(
  predicted_class=predClas_RATRLDA3$class,
  subgroup=test[, 'Subgroup'],
  lda=predClas_RATRLDA3$x
)
str(RATR_LDA3log_DF)
'data.frame':  9 obs. of  4 variables:
 $ predicted_class: Factor w/ 3 levels "Control","PreHD",...: 1 1 1 2 2 1 1 3 1
 $ subgroup       : Factor w/ 3 levels "Control","PreHD",...: 1 1 1 1 3 3 2 3 2

```

```

$ lda.LD1      : num -0.947 -0.496 -0.758 -1.334 -1.438 ...
$ lda.LD2      : num 0.539 -0.17 1.024 -1.747 -2.369 ...
ggplot(RATR_LDA3log_DF) +
  geom_text(aes(x=lda.LD1,y=lda.LD2,label=subgroup, color = predicted_class)) +
  theme_classic() + ggtitle('LDA results for 3log-LDA model of the right ATR')

```



```

# Evaluación del modelo

## Accuracy
accuracyRATR_LDAlog3<-sum(diag(table(predClas_RATRLDA3$class, test$Subgroup)))/nrow(test)
accuracyRATR_LDAlog3
[1] 0.4444444

## Confusion Matrix
confMX_RATR_LDA3log<-confusionMatrix(predClas_RATRLDA3$class, test$Subgroup)
confMX_RATR_LDA3log
Confusion Matrix and Statistics

Reference
Prediction Control PreHD HD
  Control     3     2   1
  PreHD       1     0   1
  HD          0     0   1

Overall Statistics

Accuracy : 0.4444

```

```

95% CI : (0.137, 0.788)
No Information Rate : 0.4444
P-Value [Acc > NIR] : 0.6257

Kappa : 0.1

```

Mcnemar's Test P-Value : 0.5062

Statistics by Class:

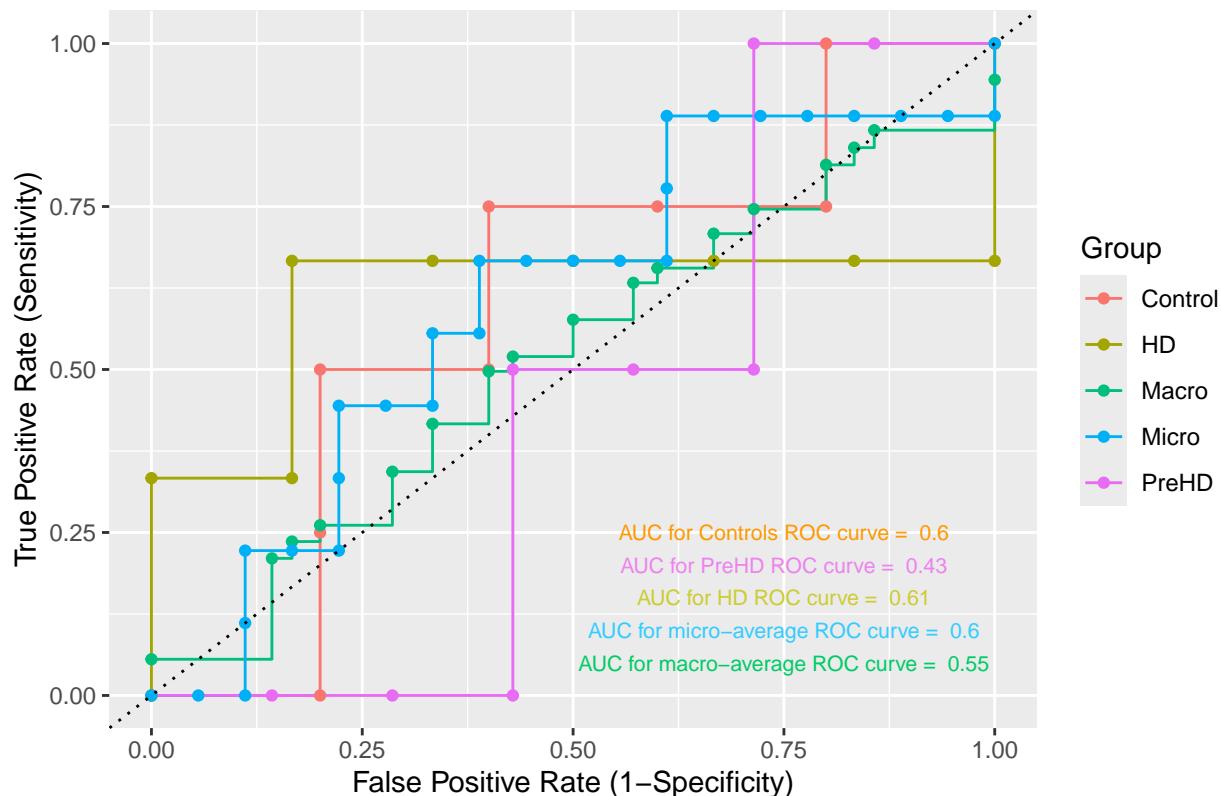
	Class: Control	Class: PreHD	Class: HD
Sensitivity	0.7500	0.0000	0.3333
Specificity	0.4000	0.7143	1.0000
Pos Pred Value	0.5000	0.0000	1.0000
Neg Pred Value	0.6667	0.7143	0.7500
Prevalence	0.4444	0.2222	0.3333
Detection Rate	0.3333	0.0000	0.1111
Detection Prevalence	0.6667	0.2222	0.1111
Balanced Accuracy	0.5750	0.3571	0.6667

```

## ROC curves one vs the rest by multiROC package
Control_pred_RATRLDA3log<- predClas_RATRLDA3$posterior[, 'Control']
PreHD_pred_RATRLDA3log<- predClas_RATRLDA3$posterior[, 'PreHD']
HD_pred_RATRLDA3log<- predClas_RATRLDA3$posterior[, 'HD']
Control_true<-ifelse(test$Subgroup=='Control',1,0)
PreHD_true<-ifelse(test$Subgroup=='PreHD',1,0)
HD_true<-ifelse(test$Subgroup=='HD',1,0)
DF_ROC_RATRLDA3log<-cbind(Control_true,PreHD_true,HD_true,Control_pred_RATRLDA3log,
                               PreHD_pred_RATRLDA3log,HD_pred_RATRLDA3log)
multiROC_RATR_LDA3log<-multi_roc(DF_ROC_RATRLDA3log)
multiROC_RATR_LDA3log_DF<-plot_roc_data(multiROC_RATR_LDA3log)
ggplot(multiROC_RATR_LDA3log_DF, aes(x=1-Specificity, y=Sensitivity, color = Group)) +
  geom_step() + geom_point() + labs(title= "3-LDA model of RATR ROC curve: One vs Rest",
  x = "False Positive Rate (1-Specificity)",
  y = "True Positive Rate (Sensitivity)") + geom_abline(intercept = 0, slope=1,linetype = 'dotted') +
  annotate('text',x=0.75, y=0.25,
  label=paste('AUC for Controls ROC curve = ',
  mean(multiROC_RATR_LDA3log_DF$AUC[multiROC_RATR_LDA3log_DF$Group=='Control']), 
  color = '#FF9900',size = 8/.pt) +
  annotate('text',x=0.75, y=0.20,
  label=paste('AUC for PreHD ROC curve = ',
  round(mean(multiROC_RATR_LDA3log_DF$AUC[multiROC_RATR_LDA3log_DF$Group=='PreHD']),2),
  color = 'violet',size = 8/.pt) +
  annotate('text',x=0.75, y=0.15,
  label=paste('AUC for HD ROC curve = ',
  round(mean(multiROC_RATR_LDA3log_DF$AUC[multiROC_RATR_LDA3log_DF$Group=='HD']),2)),
  color = '#CCCC33',size = 8/.pt) +
  annotate('text',x=0.75, y=0.10,
  label=paste('AUC for micro-average ROC curve = ',
  round(mean(multiROC_RATR_LDA3log_DF$AUC[multiROC_RATR_LDA3log_DF$Group=='Micro']),2),
  color = '#33CCFF',size = 8/.pt)+
  annotate('text',x=0.75, y=0.05,
  label=paste('AUC for macro-average ROC curve = ',
  round(mean(multiROC_RATR_LDA3log_DF$AUC[multiROC_RATR_LDA3log_DF$Group=='Macro']),2),
  color = '#00CC66',size = 8/.pt)

```

### 3-LDA model of RATR ROC curve: One vs Rest



```
## Kappa coefficient
library(irr)
kappa_RATR_LDA3log<-kappam.fleiss(data.frame(predClas_RATRLDA3$class,test$Subgroup))
kappa_RATR_LDA3log
Fleiss' Kappa for m Raters

Subjects = 9
Raters = 2
Kappa = 0.0625

z = 0.257
p-value = 0.797
```

### Resultados para el LATR

```
library(MASS)
# Para las 5 métricas

# Primero, división del dataframe sin transformación logarítmica en los datasets
# de entrenamiento y testeo
set.seed(123456)
train_index<-createDataPartition(LATR_avg_demo_WoOut$Subgroup, p=0.8, list=FALSE)
train<-LATR_avg_demo_WoOut[train_index,]
test<-LATR_avg_demo_WoOut[-train_index]

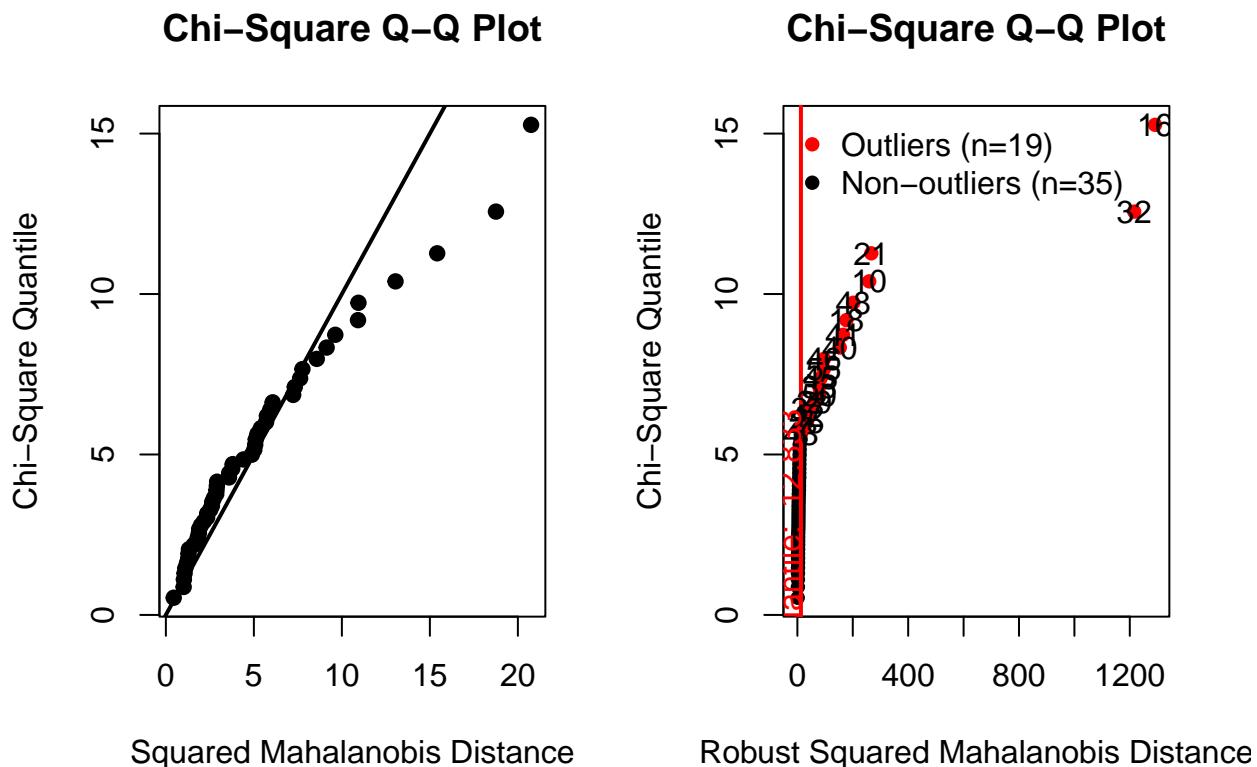
# Segundo, se genera el modelo LDA
LATR_LDA5model<-lda(Subgroup~cbind(lhatr_Iron,lhatr_AD,lhatr_RD,lhatr_MD,lhatr_FA),
```

```

            data=train)
Error in lda.default(x, grouping, ...): variables 2 3 4 appear to be constant within groups

# Segundo, con el dataframe con transformación logarítmica
## Primero transformación de las variables
LATR_avg_demo_WoOut_log<-cbind(LATR_avg_demo_WoOut[,1:8],log(cbind(LATR_avg_demo_WoOut[,9:11])),
                                 LATR_avg_demo_WoOut[,12])
colnames(LATR_avg_demo_WoOut_log)<-colnames(LATR_avg_demo_WoOut)
## Comprobar normalidad multivariante
par(mfrow = c(1, 2))
LATR_mvnllog5<-mvn(LATR_avg_demo_WoOut_log[-c(1:7)], mvnTest = 'royston', covariance = TRUE,
                     desc = TRUE, univariateTest = 'SW', univariatePlot = 'qq',
                     multivariatePlot = 'qq', multivariateOutlierMethod = 'quan',
                     showOutliers = TRUE, showNewData = TRUE)

```



```

LATR_mvnllog5$multivariateNormality
  Test      H   p value MVN
1 Royston 1.749787 0.7377654 YES
LATR_mvnllog5$univariateNormality
  Test Variable Statistic   p value Normality
1 Shapiro-Wilk lhatr_Iron    0.9895   0.9165   YES
2 Shapiro-Wilk lhatr_AD     0.9914   0.9651   YES
3 Shapiro-Wilk lhatr_RD     0.9882   0.8719   YES
4 Shapiro-Wilk lhatr_MD     0.9881   0.8672   YES
5 Shapiro-Wilk lhatr_FA     0.9642   0.1064   YES
## Comprobar homocedasticidad de la matriz de covarianzas
LATR_MBox_SG5log<-boxM(cbind(lhatr_Iron,lhatr_AD,lhatr_RD,lhatr_MD,lhatr_FA))~Subgroup,

```

```

    data=LATR_avg_demo_WoOut_log, cov=TRUE);LATR_MBox_SG5log

  Box's M-test for Homogeneity of Covariance Matrices

data: Y
Chi-Sq (approx.) = 44.913, df = 30, p-value = 0.03933
LATR_MBox_G5log<-boxM(cbind(lhatr_Iron,lhatr_AD,lhatr_RD,lhatr_MD,lhatr_FA)~Gender,
                        data=LATR_avg_demo_WoOut_log, cov=TRUE);LATR_MBox_G5log

  Box's M-test for Homogeneity of Covariance Matrices

data: Y
Chi-Sq (approx.) = 21.027, df = 15, p-value = 0.136
LATR_MBox_AGE5log<-boxM(cbind(lhatr_Iron,lhatr_AD,lhatr_RD,lhatr_MD,lhatr_FA)~agef2,
                           data=LATR_avg_demo_WoOut_log, cov=TRUE);LATR_MBox_AGE5log

  Box's M-test for Homogeneity of Covariance Matrices

data: Y
Chi-Sq (approx.) = 64.093, df = 30, p-value = 0.0002862

```

Con la transformación logarítmica, se mantiene la normalidad multivariante, pero se pierde la homocedasticidad de la matriz de covarianzas para los factores subgrupo y edad factorizada, hecho que podría sesgar los resultados en la ejecución de un MANOVA (no se ha ejecutado dado que en el paso anterior el modelo no generaba residuos rango-deficientes).

```

# Se repiten el primer y segundo paso con el dataframe con transformación logarítmica
set.seed(123456)

train_index<-createDataPartition(LATR_avg_demo_WoOut_log$Subgroup, p=0.8, list=FALSE)
train<-LATR_avg_demo_WoOut_log[train_index,]
test<-LATR_avg_demo_WoOut_log[-train_index,]

LATR_LDAlog5model<-lda(Subgroup~cbind(lhatr_Iron,lhatr_AD,lhatr_RD,lhatr_MD,lhatr_FA),
                         data=train)
print(LATR_LDAlog5model)
Call:
lda(Subgroup ~ cbind(lhatr_Iron, lhatr_AD, lhatr_RD, lhatr_MD,
                      lhatr_FA), data = train)

Prior probabilities of groups:
Control      PreHD       HD
0.4444444 0.2444444 0.3111111

Group means:
  cbind(lhatr_Iron, lhatr_AD, lhatr_RD, lhatr_MD, lhatr_FA)lhatr_Iron
Control                               0.02431511
PreHD                                0.02478308
HD                                    0.02413531
  cbind(lhatr_Iron, lhatr_AD, lhatr_RD, lhatr_MD, lhatr_FA)lhatr_AD
Control                               -7.025666
PreHD                                -7.017937
HD                                    -6.984902
  cbind(lhatr_Iron, lhatr_AD, lhatr_RD, lhatr_MD, lhatr_FA)lhatr_RD
Control                               -7.612283
PreHD                                -7.638276

```

```

HD -7.572606
  cbind(lhatr_Iron, lhatr_AD, lhatr_RD, lhatr_MD, lhatr_FA)lhatr_MD
Control -7.376196
PreHD -7.386221
HD -7.335917
  cbind(lhatr_Iron, lhatr_AD, lhatr_RD, lhatr_MD, lhatr_FA)lhatr_FA
Control 0.3774804
PreHD 0.3951712
HD 0.3780013

Coefficients of linear discriminants:

LD1
cbind(lhatr_Iron, lhatr_AD, lhatr_RD, lhatr_MD, lhatr_FA)lhatr_Iron 148.03876
cbind(lhatr_Iron, lhatr_AD, lhatr_RD, lhatr_MD, lhatr_FA)lhatr_AD -409.55038
cbind(lhatr_Iron, lhatr_AD, lhatr_RD, lhatr_MD, lhatr_FA)lhatr_RD -412.12957
cbind(lhatr_Iron, lhatr_AD, lhatr_RD, lhatr_MD, lhatr_FA)lhatr_MD 850.48479
cbind(lhatr_Iron, lhatr_AD, lhatr_RD, lhatr_MD, lhatr_FA)lhatr_FA 45.91251

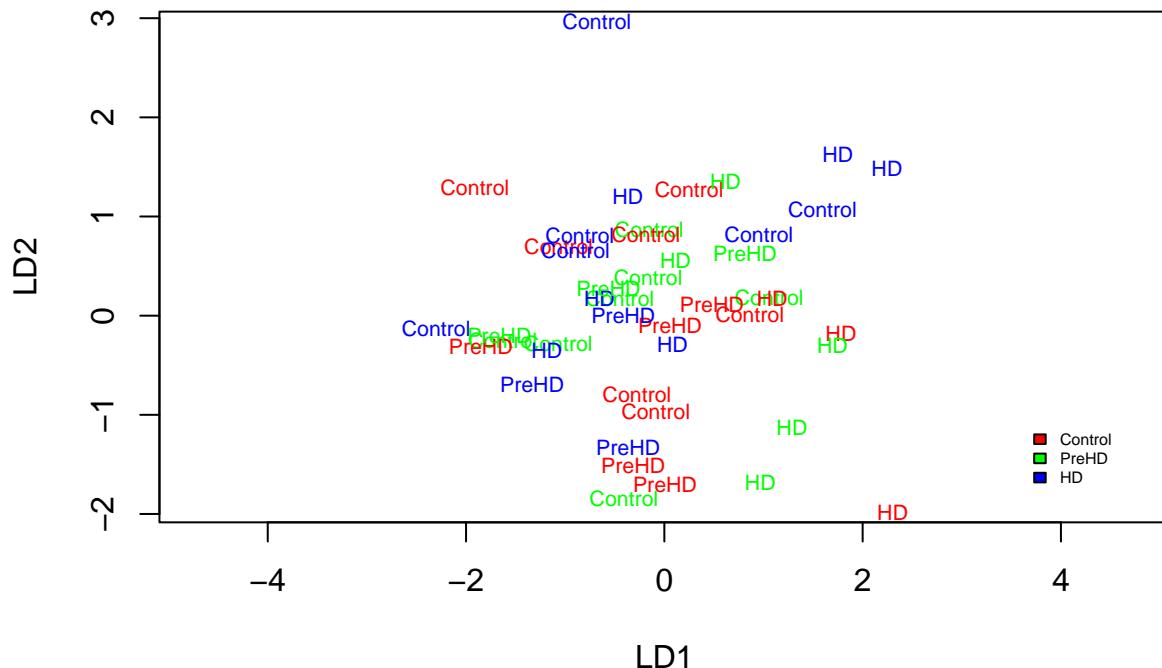
LD2
cbind(lhatr_Iron, lhatr_AD, lhatr_RD, lhatr_MD, lhatr_FA)lhatr_Iron -279.262884
cbind(lhatr_Iron, lhatr_AD, lhatr_RD, lhatr_MD, lhatr_FA)lhatr_AD 337.361659
cbind(lhatr_Iron, lhatr_AD, lhatr_RD, lhatr_MD, lhatr_FA)lhatr_RD 429.155786
cbind(lhatr_Iron, lhatr_AD, lhatr_RD, lhatr_MD, lhatr_FA)lhatr_MD -773.967514
cbind(lhatr_Iron, lhatr_AD, lhatr_RD, lhatr_MD, lhatr_FA)lhatr_FA 5.076428

Proportion of trace:
LD1 LD2
0.7531 0.2469

# Se genera el plot
##Define colors for each subgroup
subgroup_colors <- c("Control" = "red", "PreHD" = "green", "HD" = "blue")
## Plot the model with custom colors for species
plot(LATR_LDAlog5model, col = subgroup_colors,
      main="Visualization of 5log-LDA model of the left ATR")
legend('bottomright', inset=0.05, bty='n', cex=0.5, legend=c('Control','PreHD','HD'),
       fill=c('red','green','blue'))

```

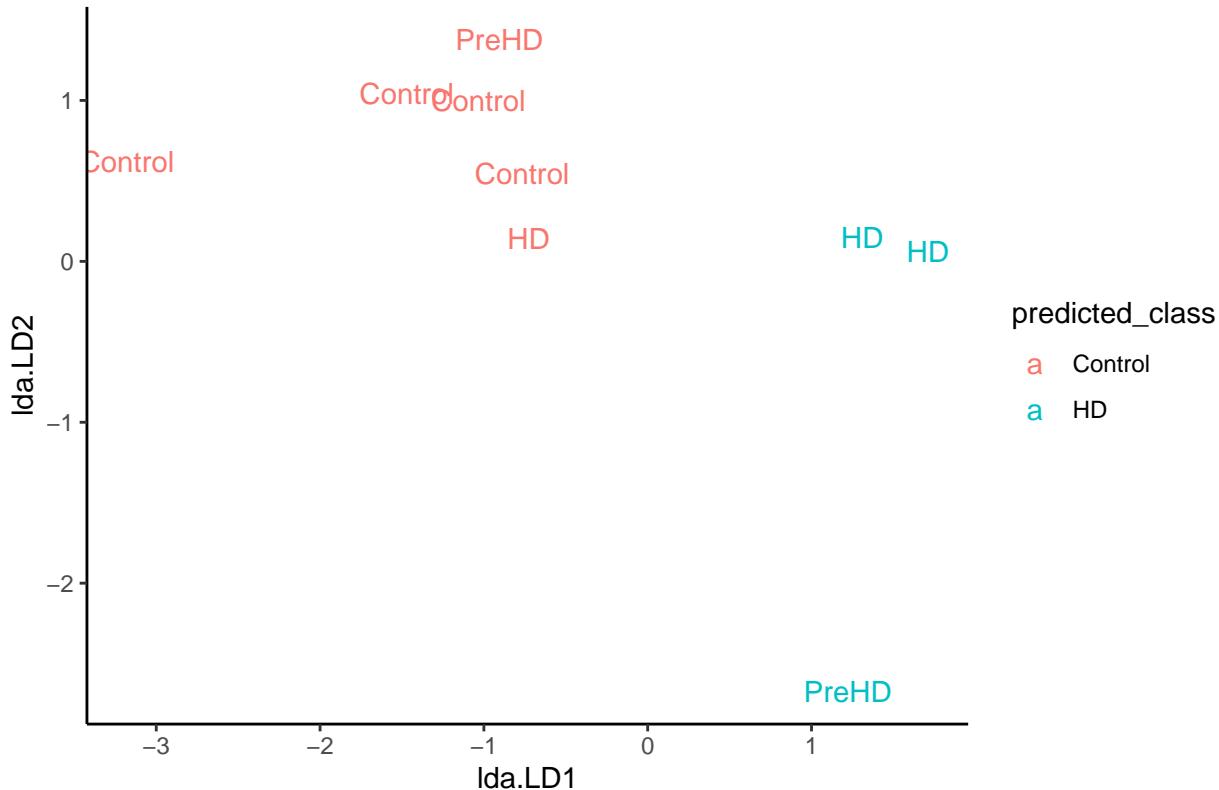
## Visualization of 5log-LDA model of the left ATR



```
# Se predicen las clases
predClas_LATRLDA5<-predict(LATR_LDAlong5model, newdata = test, type='class')
test_subgroup<-as.character(test$Subgroup)
test_subgroup<-factor(test_subgroup, levels = levels(predClas_LATRLDA5$class))
table(predClas_LATRLDA5$class, test_subgroup)
  test_subgroup
  Control PreHD HD
Control      4   1  1
PreHD        0   0  0
HD           0   1  2

LATR_LDA5log_DF<-data.frame(
  predicted_class=predClas_LATRLDA5$class,
  subgroup=test[, 'Subgroup'],
  lda=predClas_LATRLDA5$x
)
str(LATR_LDA5log_DF)
'data.frame':  9 obs. of  4 variables:
 $ predicted_class: Factor w/ 3 levels "Control","PreHD",...: 1 1 1 1 1 3 3 3 1
 $ subgroup      : Factor w/ 3 levels "Control","PreHD",...: 1 1 1 1 3 3 2 3 2
 $ lda.LD1       : num  -1.47 -0.766 -1.033 -3.179 -0.726 ...
 $ lda.LD2       : num  1.041 0.544 1.003 0.619 0.142 ...
ggplot(LATR_LDA5log_DF) +
  geom_text(aes(x=lda.LD1,y=lda.LD2,label=subgroup, color = predicted_class)) +
  theme_classic() + ggtitle('LDA results for 5log-LDA model of the left ATR')
```

## LDA results for 5log-LDA model of the left ATR



```
# Evaluación del modelo
## Accuracy
accuracyLATR_LDAlog5<-sum(diag(table(predClas_LATRLDA5$class, test$Subgroup)))/nrow(test)
accuracyLATR_LDAlog5
[1] 0.6666667

## Confusion Matrix
confMX_LATR_LDA5log<-confusionMatrix(predClas_LATRLDA5$class, test$Subgroup)
confMX_LATR_LDA5log
Confusion Matrix and Statistics

Reference
Prediction Control PreHD HD
Control      4     1   1
PreHD        0     0   0
HD           0     1   2

Overall Statistics

Accuracy : 0.6667
95% CI  : (0.2993, 0.9251)
No Information Rate : 0.4444
P-Value [Acc > NIR] : 0.1574

Kappa : 0.4375

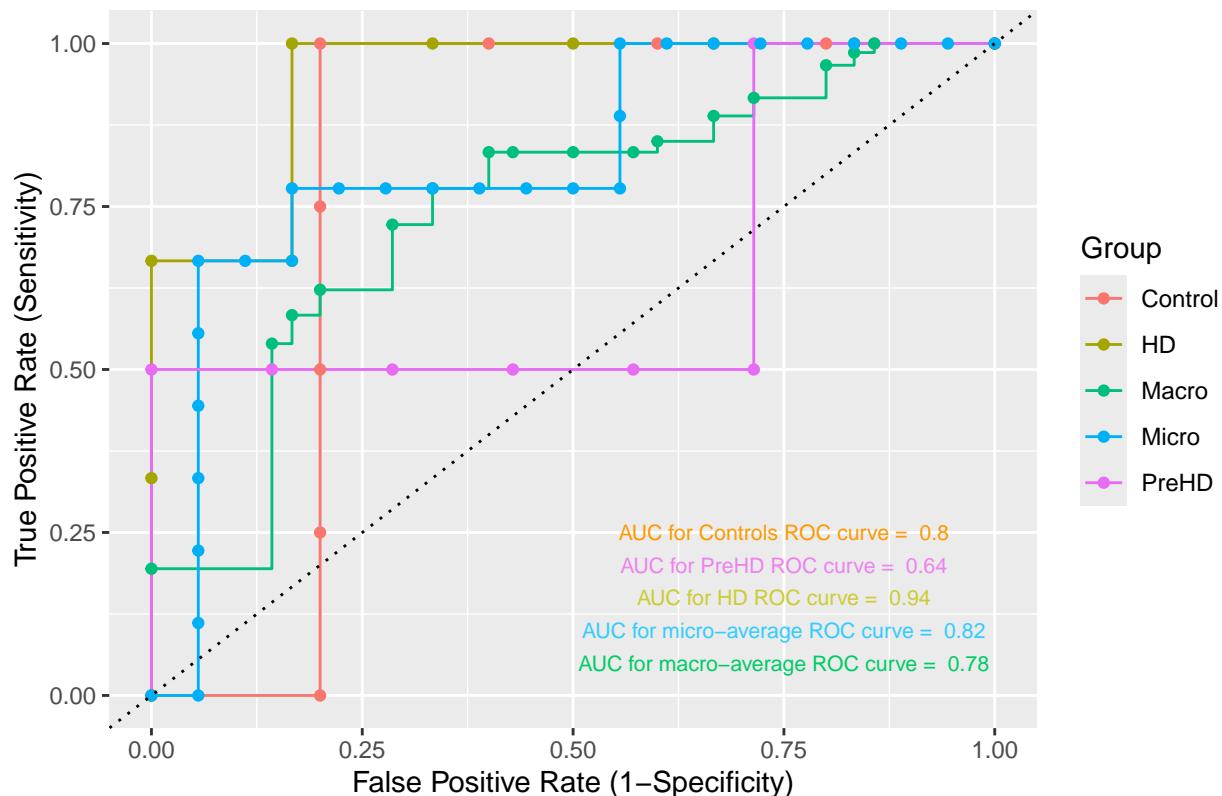
McNemar's Test P-Value : 0.3916
```

Statistics by Class:

	Class: Control	Class: PreHD	Class: HD
Sensitivity	1.0000	0.0000	0.6667
Specificity	0.6000	1.0000	0.8333
Pos Pred Value	0.6667	NaN	0.6667
Neg Pred Value	1.0000	0.7778	0.8333
Prevalence	0.4444	0.2222	0.3333
Detection Rate	0.4444	0.0000	0.2222
Detection Prevalence	0.6667	0.0000	0.3333
Balanced Accuracy	0.8000	0.5000	0.7500

```
## ROC curves one vs the rest by multiROC package
Control_pred_LATRLDA5log<- predClas_LATRLDA5$posterior[, 'Control']
PreHD_pred_LATRLDA5log<- predClas_LATRLDA5$posterior[, 'PreHD']
HD_pred_LATRLDA5log<- predClas_LATRLDA5$posterior[, 'HD']
Control_true<-ifelse(test$Subgroup=='Control',1,0)
PreHD_true<-ifelse(test$Subgroup=='PreHD',1,0)
HD_true<-ifelse(test$Subgroup=='HD',1,0)
DF_ROC_LATRLDA5log<-cbind(Control_true,PreHD_true,HD_true,Control_pred_LATRLDA5log,
                             PreHD_pred_LATRLDA5log,HD_pred_LATRLDA5log)
multiROC_LATR_LDA5log<-multi_roc(DF_ROC_LATRLDA5log)
multiROC_LATR_LDA5log_DF<-plot_roc_data(multiROC_LATR_LDA5log)
ggplot(multiROC_LATR_LDA5log_DF, aes(x=1-Specificity, y=Sensitivity, color = Group)) +
  geom_step() + geom_point() + labs(title= "5-LDA model of LATR ROC curve: One vs Rest",
  x = "False Positive Rate (1-Specificity)",
  y = "True Positive Rate (Sensitivity)") + geom_abline(intercept = 0, slope=1,linetype = 'dotted') +
  annotate('text',x=0.75, y=0.25,
           label=paste('AUC for Controls ROC curve = ',
                       mean(multiROC_LATR_LDA5log_DF$AUC[multiROC_LATR_LDA5log$Group=='Control'])),
           color = '#FF9900',size = 8/.pt) +
  annotate('text',x=0.75, y=0.20,
           label=paste('AUC for PreHD ROC curve = ',
                       round(mean(multiROC_LATR_LDA5log_DF$AUC[multiROC_LATR_LDA5log$Group=='PreHD'])),
           color = 'violet',size = 8/.pt) +
  annotate('text',x=0.75, y=0.15,
           label=paste('AUC for HD ROC curve = ',
                       round(mean(multiROC_LATR_LDA5log_DF$AUC[multiROC_LATR_LDA5log$Group=='HD'])),
           color = '#CCCC33',size = 8/.pt) +
  annotate('text',x=0.75, y=0.10,
           label=paste('AUC for micro-average ROC curve = ',
                       round(mean(multiROC_LATR_LDA5log_DF$AUC[multiROC_LATR_LDA5log$Group=='All'])),
           color = '#33CCFF',size = 8/.pt)+
  annotate('text',x=0.75, y=0.05,
           label=paste('AUC for macro-average ROC curve = ',
                       round(mean(multiROC_LATR_LDA5log_DF$AUC[multiROC_LATR_LDA5log$Group=='All'])),
           color = '#00CC66',size = 8/.pt)
```

## 5-LDA model of LATR ROC curve: One vs Rest



```
## Kappa coefficient
library(irr)
kappa_LATR_LDA5log<-kappam.fleiss(data.frame(predClas_LATRLDA5$class,test$Subgroup))
kappa_LATR_LDA5log
Fleiss' Kappa for m Raters

Subjects = 9
Raters = 2
Kappa = 0.413

z = 1.58
p-value = 0.115
```

```
# Primero, división del dataframe sin transformación logarítmica en los datasets
# de entrenamiento y testeo
set.seed(123456)
train_index<-createDataPartition(LATR_avg_demo_WoOut_ADRDMD$Subgroup, p=0.8, list=FALSE)
train<-LATR_avg_demo_WoOut_ADRDMD[train_index,]
test<-LATR_avg_demo_WoOut_ADRDMD[-train_index,]
# Segundo, se genera el modelo LDA
LATR_LDA3model<-lda(Subgroup~cbind(lhatr_AD,lhatr_RD,lhatr_MD), data=train)
Error in lda.default(x, grouping, ...): variables 1 2 3 appear to be constant within groups

# Se repiten el primer y segundo paso con el dataframe con transformación logarítmica
set.seed(123456)

train_index<-createDataPartition(LATR_avg_demo_WoOut_ADRDMD_log$Subgroup, p=0.8, list=FALSE)
train<-LATR_avg_demo_WoOut_ADRDMD_log[train_index,]
```

```

test<-LATR_avg_demo_WoOut_ADRDMD_log[-train_index,]

LATR_LDAlog3model<-lda(Subgroup~cbind(lhatr_AD,lhatr_RD,lhatr_MD),data=train)
print(LATR_LDAlog3model)
Call:
lda(Subgroup ~ cbind(lhatr_AD, lhatr_RD, lhatr_MD), data = train)

Prior probabilities of groups:
  Control      PreHD       HD
0.4444444 0.2444444 0.3111111

Group means:
  cbind(lhatr_AD, lhatr_RD, lhatr_MD)lhatr_AD
Control                  -7.025666
PreHD                   -7.017937
HD                      -6.984902
  cbind(lhatr_AD, lhatr_RD, lhatr_MD)lhatr_RD
Control                  -7.612283
PreHD                   -7.638276
HD                      -7.572606
  cbind(lhatr_AD, lhatr_RD, lhatr_MD)lhatr_MD
Control                  -7.376196
PreHD                   -7.386221
HD                      -7.335917

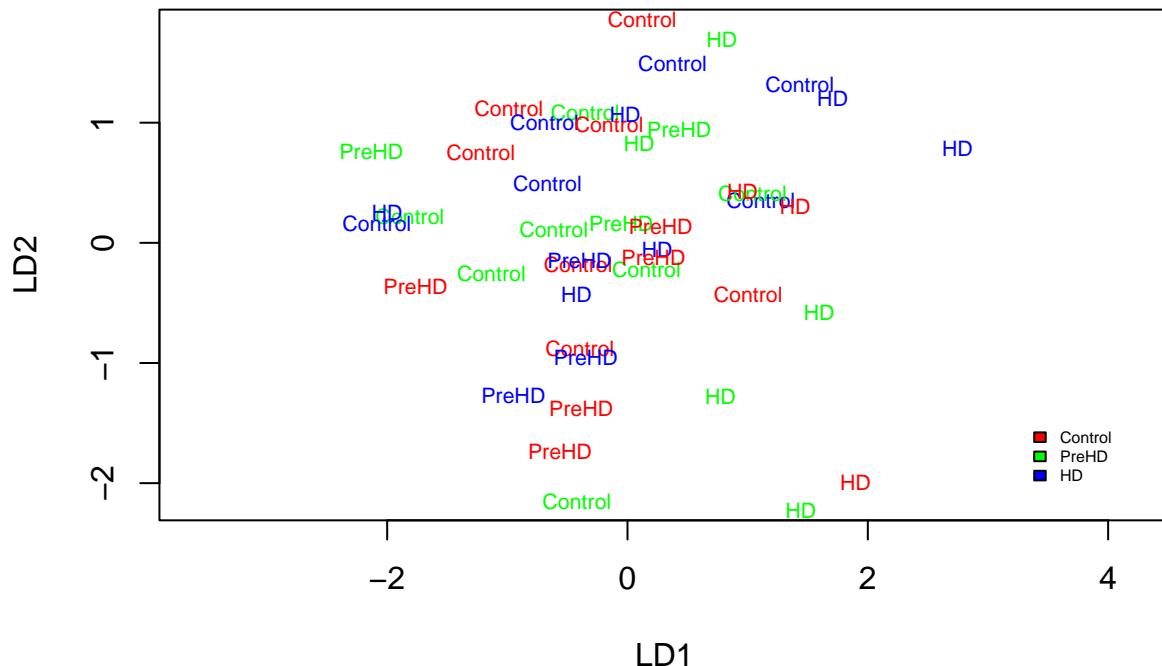
Coefficients of linear discriminants:
                               LD1          LD2
cbind(lhatr_AD, lhatr_RD, lhatr_MD)lhatr_AD -369.2345  428.9971
cbind(lhatr_AD, lhatr_RD, lhatr_MD)lhatr_RD -411.9426  526.3884
cbind(lhatr_AD, lhatr_RD, lhatr_MD)lhatr_MD  806.3915 -961.6843

Proportion of trace:
  LD1     LD2
0.7785 0.2215

# Se genera el plot
##Define colors for each subgroup
subgroup_colors <- c("Control" = "red", "PreHD" = "green", "HD" = "blue")
## Plot the model with custom colors for species
plot(LATR_LDAlog3model, col = subgroup_colors,
      main="Visualization of 3log-LDA model of the left ATR")
legend('bottomright',inset=0.05, bty='n', cex=0.5, legend=c('Control','PreHD','HD'),
      fill=c('red','green','blue'))

```

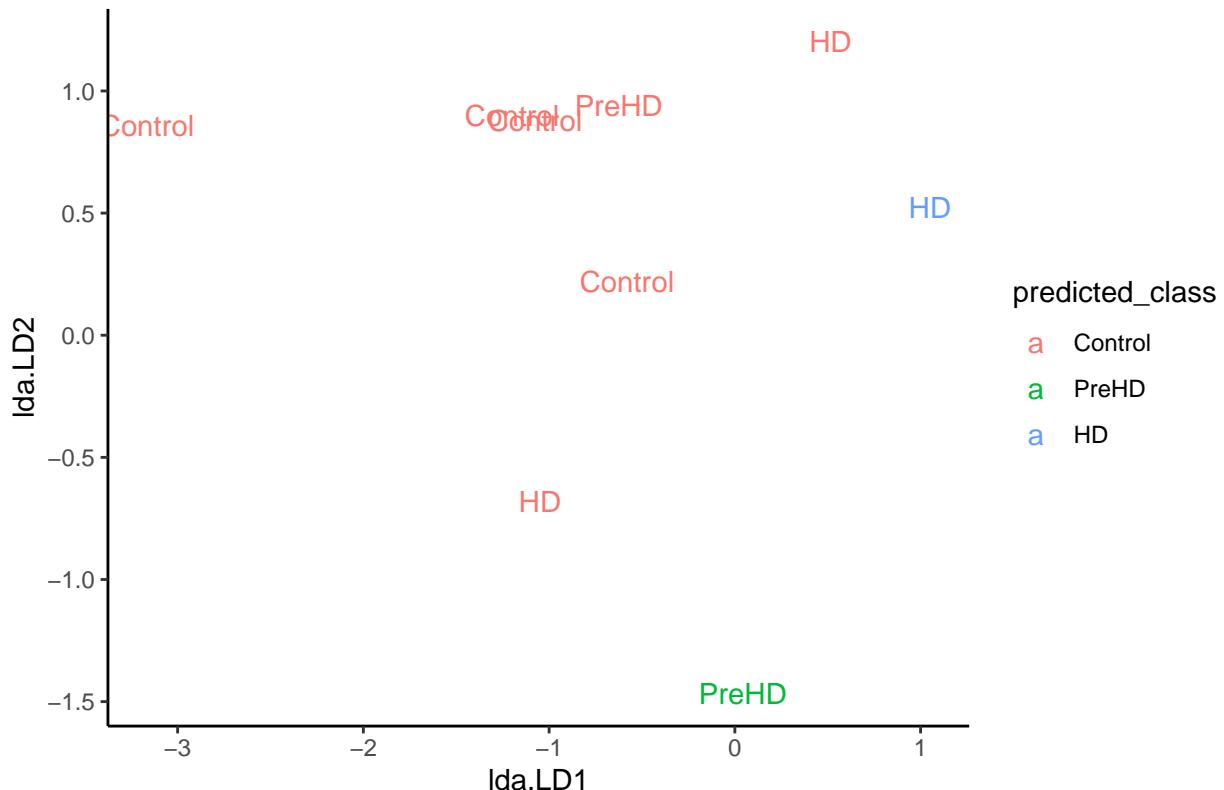
## Visualization of 3log-LDA model of the left ATR



```
# Se predicen las clases
predClas_LATRLDA3<-predict(LATR_LDAlong3model, newdata = test, type='class')
test_subgroup<-as.character(test$Subgroup)
test_subgroup<-factor(test_subgroup, levels = levels(predClas_LATRLDA3$class))
table(predClas_LATRLDA3$class, test_subgroup)
  test_subgroup
    Control PreHD HD
Control      4   1  2
PreHD        0   1  0
HD           0   0  1

LATR_LDA3log_DF<-data.frame(
  predicted_class=predClas_LATRLDA3$class,
  subgroup=test[, 'Subgroup'],
  lda=predClas_LATRLDA3$x
)
str(LATR_LDA3log_DF)
'data.frame':  9 obs. of  4 variables:
 $ predicted_class: Factor w/ 3 levels "Control","PreHD",...: 1 1 1 1 1 1 2 3 1
 $ subgroup      : Factor w/ 3 levels "Control","PreHD",...: 1 1 1 1 3 3 2 3 2
 $ lda.LD1       : num  -1.199 -0.581 -1.076 -3.164 -1.049 ...
 $ lda.LD2       : num  0.898 0.222 0.881 0.859 -0.681 ...
ggplot(LATR_LDA3log_DF) +
  geom_text(aes(x=lda.LD1,y=lda.LD2,label=subgroup, color = predicted_class)) +
  theme_classic() + ggtitle('LDA results for 3log-LDA model of the left ATR')
```

### LDA results for 3log-LDA model of the left ATR



```
# Evaluación del modelo
## Accuracy
accuracyLATR_LDAlog3<-sum(diag(table(predClas_LATRLDA3$class, test$Subgroup)))/nrow(test)
accuracyLATR_LDAlog3
[1] 0.6666667

## Confusion Matrix
confMX_LATR_LDA3log<-confusionMatrix(predClas_LATRLDA3$class, test$Subgroup)
confMX_LATR_LDA3log
Confusion Matrix and Statistics

Reference
Prediction Control PreHD HD
Control      4     1   2
PreHD        0     1   0
HD           0     0   1

Overall Statistics

Accuracy : 0.66667
95% CI  : (0.2993, 0.9251)
No Information Rate : 0.4444
P-Value [Acc > NIR] : 0.1574

Kappa : 0.4375

McNemar's Test P-Value : NA
```

### Statistics by Class:

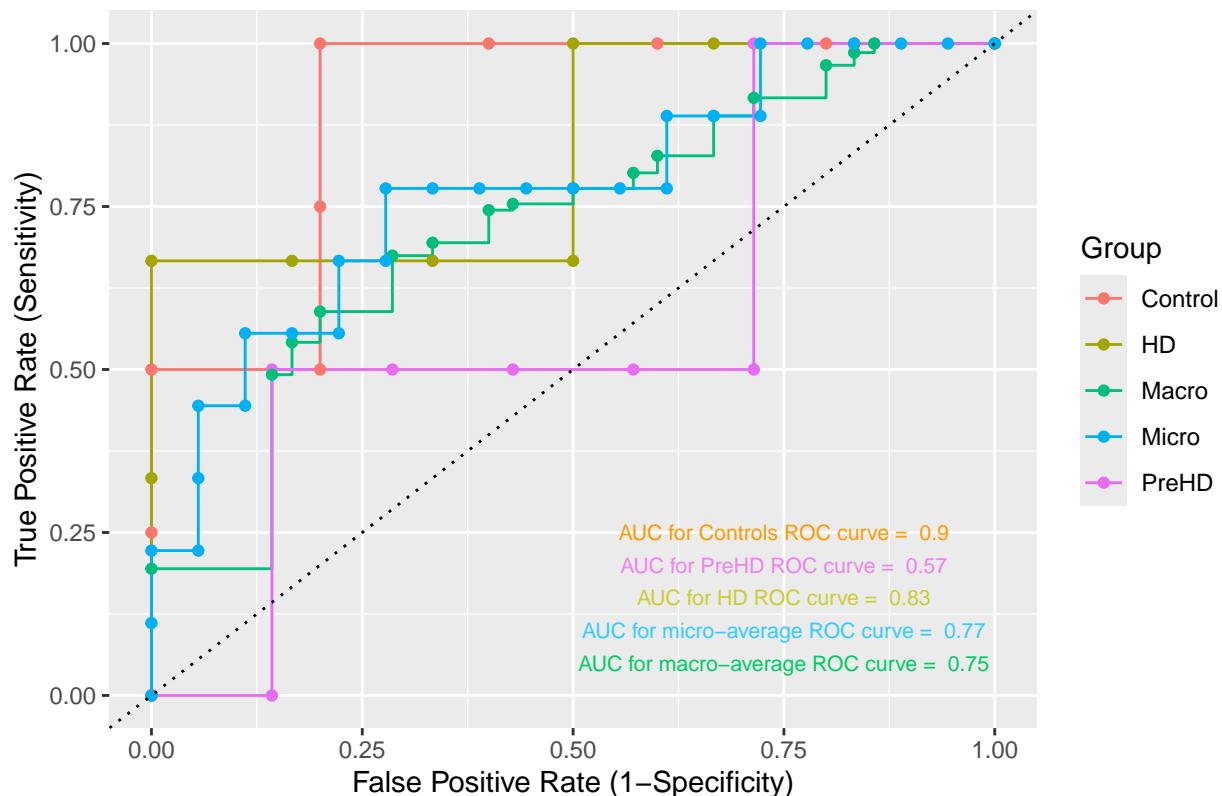
	Class: Control	Class: PreHD	Class: HD
Sensitivity	1.0000	0.5000	0.3333
Specificity	0.4000	1.0000	1.0000
Pos Pred Value	0.5714	1.0000	1.0000
Neg Pred Value	1.0000	0.8750	0.7500
Prevalence	0.4444	0.2222	0.3333
Detection Rate	0.4444	0.1111	0.1111
Detection Prevalence	0.7778	0.1111	0.1111
Balanced Accuracy	0.7000	0.7500	0.6667

```

## ROC curves one vs the rest by multiROC package
Control_pred_LATRLDA3log<- predClas_LATRLDA3$posterior[, 'Control']
PreHD_pred_LATRLDA3log<- predClas_LATRLDA3$posterior[, 'PreHD']
HD_pred_LATRLDA3log<- predClas_LATRLDA3$posterior[, 'HD']
Control_true<-ifelse(test$Subgroup=='Control',1,0)
PreHD_true<-ifelse(test$Subgroup=='PreHD',1,0)
HD_true<-ifelse(test$Subgroup=='HD',1,0)
DF_ROC_LATRLDA3log<-cbind(Control_true,PreHD_true,HD_true,Control_pred_LATRLDA3log,
                             PreHD_pred_LATRLDA3log,HD_pred_LATRLDA3log)
multiROC_LATR_LDA3log<-multi_roc(DF_ROC_LATRLDA3log)
multiROC_LATR_LDA3log_DF<-plot_roc_data(multiROC_LATR_LDA3log)
ggplot(multiROC_LATR_LDA3log_DF, aes(x=1-Specificity, y=Sensitivity, color = Group)) +
  geom_step() + geom_point() + labs(title= "3-LDA model of LATR ROC curve: One vs Rest",
  x = "False Positive Rate (1-Specificity)",
  y = "True Positive Rate (Sensitivity)") + geom_abline(intercept = 0, slope=1,linetype = 'dotted') +
  annotate('text',x=0.75, y=0.25,
           label=paste('AUC for Controls ROC curve = ',
                       mean(multiROC_LATR_LDA3log_DF$AUC[multiROC_LATR_LDA3log_DF$Group=='Control'])),
           color = '#FF9900',size = 8/.pt) +
  annotate('text',x=0.75, y=0.20,
           label=paste('AUC for PreHD ROC curve = ',
                       round(mean(multiROC_LATR_LDA3log_DF$AUC[multiROC_LATR_LDA3log_DF$Group=='PreHD']),2)),
           color = 'violet',size = 8/.pt) +
  annotate('text',x=0.75, y=0.15,
           label=paste('AUC for HD ROC curve = ',
                       round(mean(multiROC_LATR_LDA3log_DF$AUC[multiROC_LATR_LDA3log_DF$Group=='HD']),2)),
           color = '#CCCC33',size = 8/.pt) +
  annotate('text',x=0.75, y=0.10,
           label=paste('AUC for micro-average ROC curve = ',
                       round(mean(multiROC_LATR_LDA3log_DF$AUC[multiROC_LATR_LDA3log_DF$Group=='Micro']),2)),
           color = '#33CCFF',size = 8/.pt) +
  annotate('text',x=0.75, y=0.05,
           label=paste('AUC for macro-average ROC curve = ',
                       round(mean(multiROC_LATR_LDA3log_DF$AUC[multiROC_LATR_LDA3log_DF$Group=='Macro']),2)),
           color = '#00CC66',size = 8/.pt)

```

### 3-LDA model of LATR ROC curve: One vs Rest



```
## Kappa coefficient
library(irr)
kappa_LATR_LDA3log<-kappam.fleiss(data.frame(predClas_LATRLDA3$class,test$Subgroup))
kappa_LATR_LDA3log
Fleiss' Kappa for m Raters

Subjects = 9
Raters = 2
Kappa = 0.393

z = 1.59
p-value = 0.112
```

Se objetiva un solapamiento entre los tres subgrupos mediante el análisis discriminante lineal, ya sea considerando las 5 métricas (iron, AD, RD, MD, FA) o las 3 métricas de difusividad (AD, RD, MD), tanto en el ATR derecho como en el ATR izquierdo.

### 3.2 PERMANOVA

Se ha consultado la siguiente fuente de información: 20. PERMANOVA, 16. Permutation Tests, 25. Controlling Permutations, 26. Restricting Permutations.

**PERMANOVA (PER)mutational Multivariate ANalysis of VAriance, análisis multivariante permutacional de la varianza**) es una técnica analítica basada en la permutación (sin suposiciones distribucionales sobre la normalidad multivariante o la homogeneidad de las varianzas) para datos multivariantes. Se decide utilizar también PERMANOVA debido a la multicolinealidad de las variables dependientes, especialmente entre AD, RD y MD, y debido a la pérdida de homocedasticidad en las matrices de covarianza en ambos ATRs con los MANOVA obtenidos a partir de las 3 métricas de difusividad con transformación logarítmica.

PERMANOVA se basa en el **teorema de Huygens**: la suma de las diferencias al cuadrado entre los puntos y su centroide es igual a la suma de las distancias entre puntos al cuadrado dividida por el número de puntos, por tanto, la variación dentro de un grupo se puede calcular directamente a partir de una matriz de distancia, sin necesidad de conocer la localización de los centroides grupales. Por tanto, si se puede calcular la variación a partir de la matriz de distancia, eso significa que también se puede dividir esa variación entre diferentes fuentes (variables independientes). Dado que se basa en la matriz de distancias, PERMANOVA es una técnica extremadamente potente y flexible, **aplicable a datos de cualquier dimensionalidad** (útil tanto para datos univariantes como para datos multivariantes), puede expresarse a través de cualquier medida de distancia y se puede utilizar para **analizar modelos complejos**, incluidos múltiples factores, covariables (predictores distribuidos continuamente) e interacciones entre términos.

Como es una técnica basada en permutaciones, **no requiere suposiciones sobre la normalidad y/o la homogeneidad de las varianzas**. La **única suposición** del PERMANOVA (compartida con el resto de pruebas de permutación) es que las **observaciones** sean **intercambiables** bajo la hipótesis nula y la justificación para usar permutaciones para evaluar la significación.

El **estadístico** de PERMANOVA, **pseudo-F**, se modela directamente a partir del estadístico F ANOVA convencional: se calcula como una relación entre la cantidad de variación entre los grupos y dentro de ellos (between group variation/within group variation), con el numerador y el denominador ponderados cada uno por sus grados de libertad. El valor que toma el estadístico es 0 o positivo, correspondiendo los valores más grandes a una mayor importancia proporcional del factor de agrupación. Al igual que con otras técnicas de permutación, la **distribución del estadístico** se evalúa mediante el reetiquetado repetido de las unidades de observación (es decir, la permutación de las identidades de grupo) y el recálculo del estadístico. Remarcar que el estadístico es idéntico al estadístico F convencional cuando se calcula utilizando la distancia euclídea para una sola variable.

Para aplicar un PERMANOVA, se requieren los siguientes **pasos**:

- Primero, se calcula la matriz de distancias utilizando una medida de distancia apropiada;
- Segundo, se eleva al cuadrado la matriz de distancias;
- Tercero, se calcula las tres particiones de la variación: – Variación total (*total sum of squares SST*): la suma de las distancias al cuadrado dividido por el número de observaciones. – Variación dentro del grupo (*sum of squares within groups SSW*): se calcula igual que SST pero separadamente para cada grupo y se suma el valor de cada grupo para obtener el SSW global; – Variación entre grupos (*sum of squares between groups SSB*): se calcula con la substracción entre SST-SSW.
- Cuarto, se calcula el estadístico:  $PseudoF = \frac{SSB/(t-1)}{SSW/(N-1)}$ , donde t es el número de grupos y N es el número total de observaciones/sujetos.
- Quinto, se evalúa la significación estadística a través del test de permutación.
- Sexto, se calcula el p-valor como la proporción de permutaciones con un valor de pseudo-F igual o mayor al de los datos.

Posteriormente, en el caso que el permanova sea significativo, se calcula entre qué grupos se encuentran dichas diferencias mediante un test pareado.

Para cada ATR, se va a proceder de la manera siguiente:

- Generación de las matrices de distancias;
- Generación de los modelos PERMANOVA;

- Valoración de la dispersión de los modelos PERMANOVA;
- Comparación por pares.

Para determinar **qué distancia elegir**, se han consultado estas fuentes de información: Clustering Distance Measures, Distance/similarity measures, 12. Common Distance Measures.

	Species data (asymmetric)	Non-species data (symmetric)
Quantitative (continuous)	Bray-Curtis Chi-Square UniFrac	Euclidean Manhattan
Mixed, including categorical		Gower
Binary	Jaccard Sorenson UniFrac	Simple matching

La matriz de distancias se puede calcular con las funciones `dist`, `daisy` y `vegdist` de los paquetes `stats`, `cluster` y `vegan`.

Existen diferentes paquetes que ofrecen funciones para ejecutar/calcular PERMANOVA en R. En este caso, se va a utilizar la función `adonis2` del paquete `vegan`. Existen **diferentes parámetros a elegir**:

- *Permutations*: número de permutaciones y cómo realizar la permutación (existencia de bloques, dentro de los niveles de un factor...);
- *Method*: elección de la medida de distancia;
- *By*: método de evaluación de la significancia de los términos (secuencial, marginal, global);
- *Strata*: grupos en los que se restringen las permutaciones.

A pesar de corresponder a una técnica no paramétrica, se decide generar PERMANOVAS con los dataframes originales (`RATR_avg_demo`, `LATR_avg_demo`) y con la eliminación de los outliers que se ha realizado en la normalidad multivariante (`RATR_avg_demo_WoOut`, `LATR_avg_demo_WoOut`).

En cuanto a la **valoración de la dispersión**, se consulta la siguiente fuente: 21. PERMDISP. Como ya se comentó previamente, las técnicas de permutación no suponen homogeneidad de varianzas, pero los **patrones de dispersión** (variabilidad entre unidades de la muestra) pueden influir en la interpretación del estadístico debido a que las diferencias significativas entre grupos pueden deberse a diferencias en las medias (centroídes) y/o diferencias en la dispersión de las unidades de la muestra alrededor de esos centroídes. Esta sensibilidad a la dispersión en los análisis multivariados es idéntica a la sensibilidad a la varianza en los análisis univariados (los grupos en un análisis univariado pueden diferir con respecto a sus valores medios y/o la variación alrededor de esas medias).

La dispersión **se puede examinar** mediante los tests de Levene o Fligner en los análisis univariados y mediante **PERMDISP** que es una extensión multivariada del test de Levene para examinar si los grupos difieren en la variabilidad entre ellos. PERMDISP calcula la distancia de cada punto al centroide de su grupo y luego examina si esas distancias difieren entre los grupos. Como no siempre se puede calcular directamente el centroide (como en el caso de las distancias semimétricas), se aplica un **Análisis de Coordenadas Principales (PCoA)** a la matriz de distancias y los centroídes se calculan sobre la base de las coordenadas en el espacio PCoA.

Un **resultado no significativo** indicaría que los grupos no difieren en dispersión y que, por lo tanto, las diferencias se deben exclusivamente a diferencias de localización, y, un **resultado significativo** indicaría que los grupos pueden diferir exclusivamente en dispersión o tanto en dispersión como en localización.

Es importante recordar que la dispersión se ve fuertemente **afectada por las transformaciones aplicadas a los datos originales y por la elección de la medida de distancia**, por lo que si PERMDISP se utiliza tras un PERMANOVA, es esencial que se utilicen los mismos ajustes de datos y medidas de distancia en ambos casos, de lo contrario los tests no serían directamente comparables.

El **procedimiento básico** para PERMDISP es el siguiente:

1. Se parte de una matriz de distancias y se ejecuta una ordenación del Análisis de Coordenadas Principales (PCoA) sobre dicha matriz.
2. Se superpone el factor de agrupación en el PCoA para codificar cada unidad de muestra según el grupo al que pertenece.
3. Se calcula el centroide (puntuación promedio) para cada nivel del factor de agrupación en el espacio de ordenación.
4. Se calcula la distancia entre cada observación y el centroide del grupo al que pertenece.

**Esta técnica no implica un test estadístico, es decir, produce un conjunto de datos univariados** (la distancia desde cada observación al centroide de su grupo) con una distribución relativamente normal, independientemente de las características de los datos multivariados que generaron la matriz de distancia, y que puede analizarse posteriormente mediante otros tests estadísticos como ANOVA, PERMANOVA u otras técnicas basadas en permutación.

El PERMDISP (o análisis de homogeneidad multivariada de dispersiones/varianzas grupales) se va a **ejecutar mediante la función betadisper del paquete vegan**. Esta función realiza un PCoA de la matriz de distancias, calcula la distancia desde cada unidad de muestra al centroide para su nivel del factor de agrupación y guarda estas distancias en un objeto de la clase 'betadisper'. Se pueden aplicar diferentes funciones a dicho objeto con diferentes objetivos:

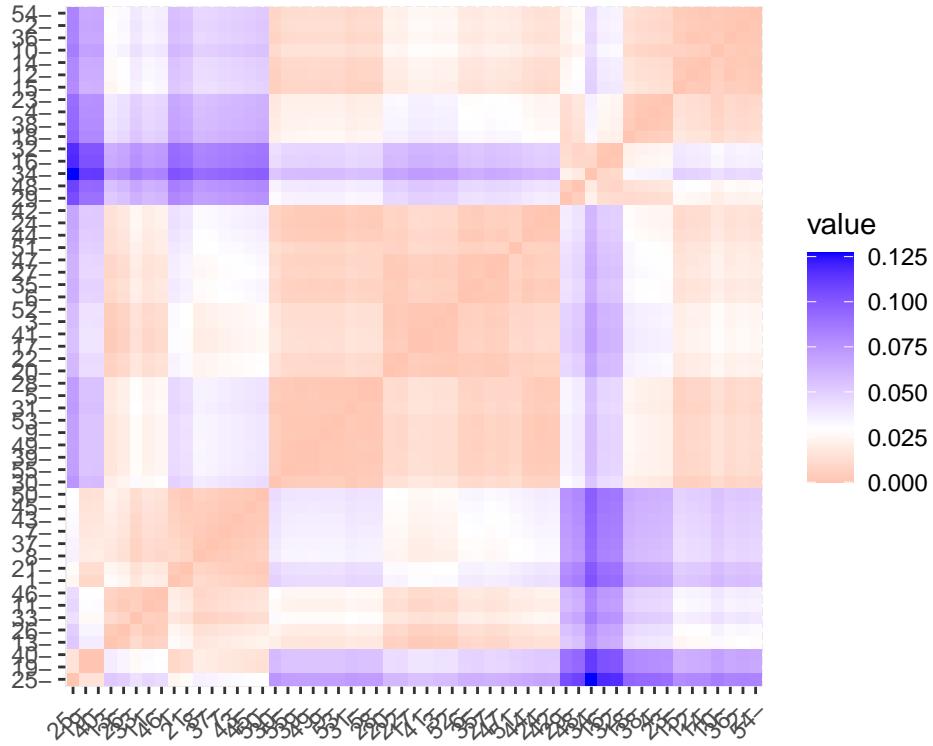
- Ver datos gráficamente: funciones `plot` o `boxplot`;
- Realizar tests estadísticos: funciones `anova`, `TukeyHSD` o `permute`.

### 3.2.1 RATR

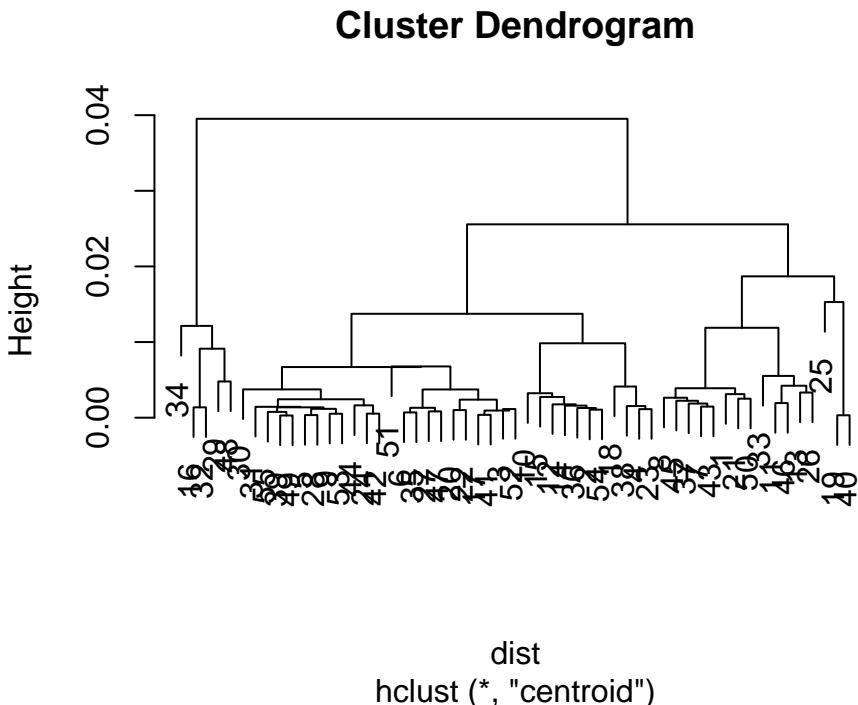
Se ha procedido a **calcular la matriz de distancias** con la función `daisy` del paquete `cluster` para poder visualizar las similaridades/disimilaridades entre subgrupos mediante MDS y tSNE. Se consulta esta web: Rtsne: Perplexity is too large. Las métricas de RM son variables numéricas continuas y no procedentes de especies, por lo que se decide utilizar la **distancia euclídea** en la generación de la matriz de distancias.

```
library(cluster)
# Matriz de distancias para las 5 métricas - Sin eliminación de outliers
dist<-daisy(RATR_avg_demo[,-c(1:7)],metric='euclidean')
distmx<-as.matrix(dist)

library(factoextra)
fviz_dist(dist)
```



```
clust<-hclust(dist, method = "centroid", members = NULL)
plot(clust)
```

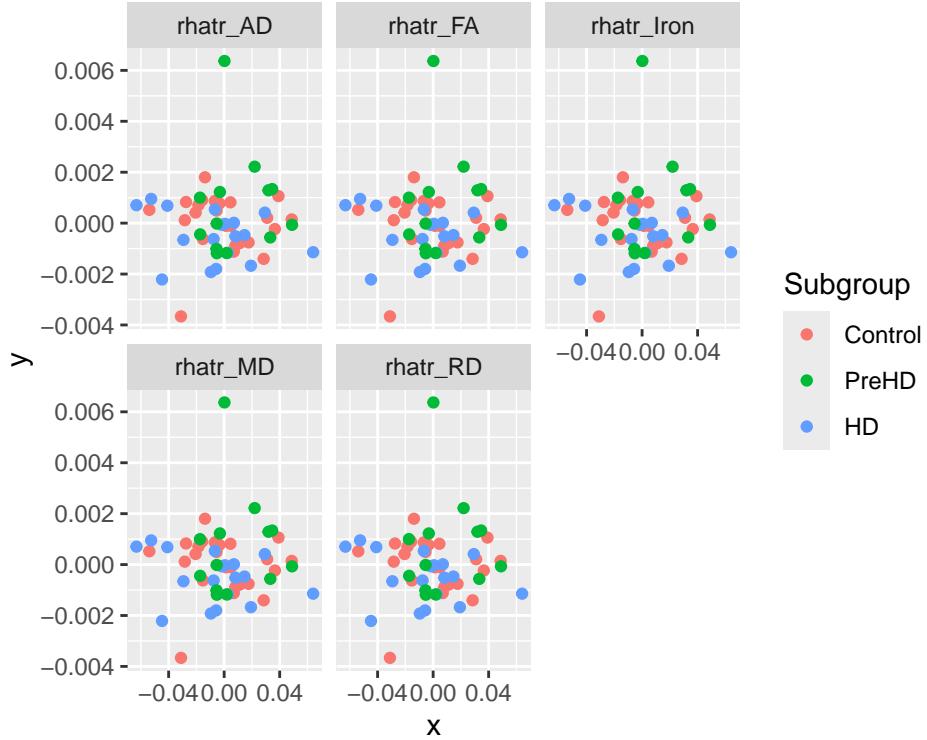


```
mds_dist<-cmdscale(distmx, eig=TRUE, k=2)
df_mds_dist<-data.frame(RATR_avg_demo, x=mds_dist$points[,1],
```

```

y=mds_dist$points[,2])
df_mds_dist2<-df_mds_dist %>%
  dplyr::select(x, y, Subgroup, Gender, agef2, rhatr_Iron, rhatr_AD, rhatr_RD,
                rhatr_MD, rhatr_FA) %>%
  tidyr::gather(key = "variable", value = "values", -x, -y, -Subgroup, -Gender, -agef2)
library(ggplot2)
ggplot(df_mds_dist2, aes(x, y)) + geom_jitter(aes(col=Subgroup)) + facet_wrap(~variable)

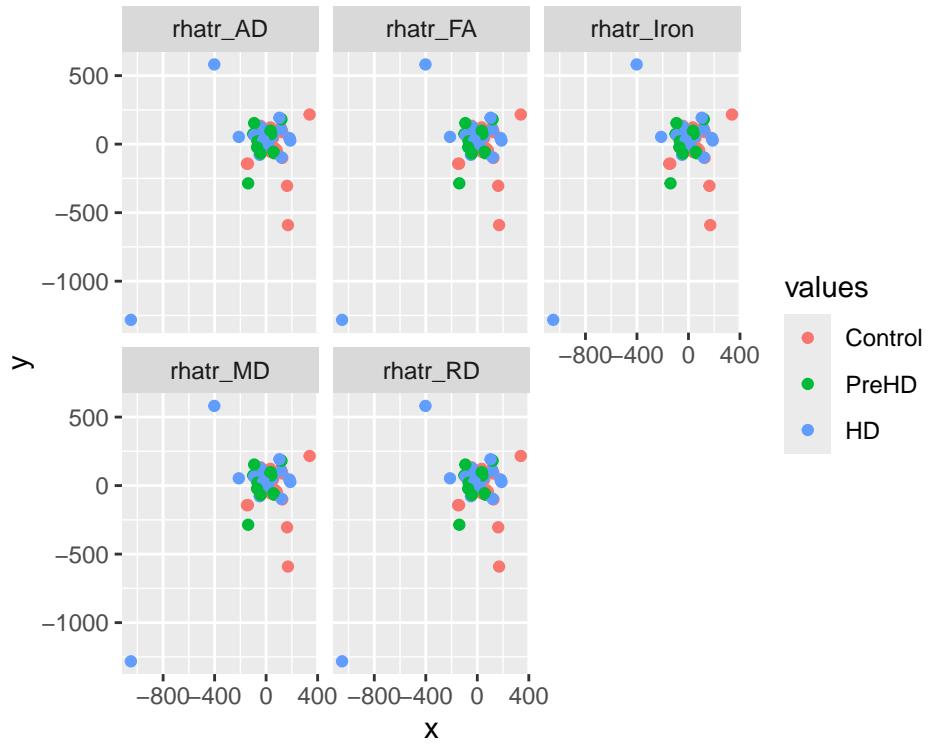
```



```

library(Rtsne)
my_Rtsne<-Rtsne(X= distmx,perplexity= floor((nrow(RATR_avg_demo) - 1) / 3),dims = 2)
# Maximum perplexity
my_Rtsne$perplexity
[1] 18
#fit_tsne<-Rtsne(dist,is_distance=TRUE,check_duplicates=FALSE,pca=FALSE,dims=2)
set.seed(123)
fit_tsne<-Rtsne(dist,perplexity = 1, is_distance = TRUE, check_duplicates = FALSE,
                  pca = FALSE, dims = 2, initial_dims = 50)
df_tsne_vars <- as_tibble(cbind(RATR_avg_demo,
                                    x = fit_tsne$Y[, 1], y = fit_tsne$Y[, 2]))
df_tsne <- df_tsne_vars %>%
  dplyr::select(x, y, Subgroup, Gender, agef2, rhatr_Iron, rhatr_AD, rhatr_RD,
                rhatr_MD, rhatr_FA) %>%
  tidyr::gather(key = "variable", value = "values", -x, -y, -Subgroup, -Gender, -agef2)
ggplot(df_tsne,aes(x, y, col=values)) + geom_point(aes(col=Subgroup)) + facet_wrap(~variable)

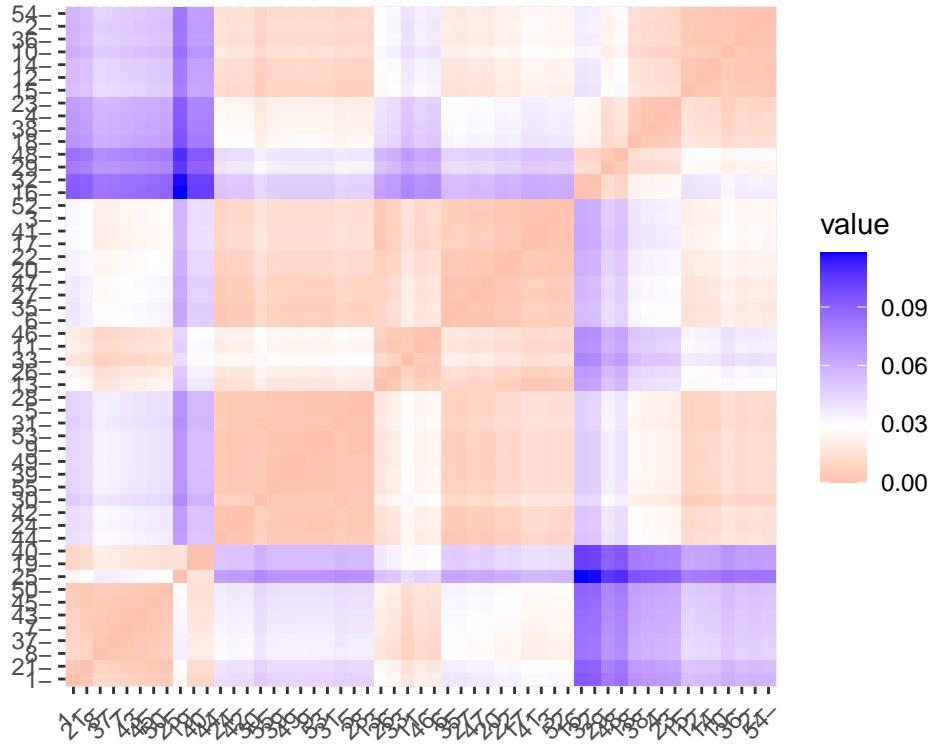
```



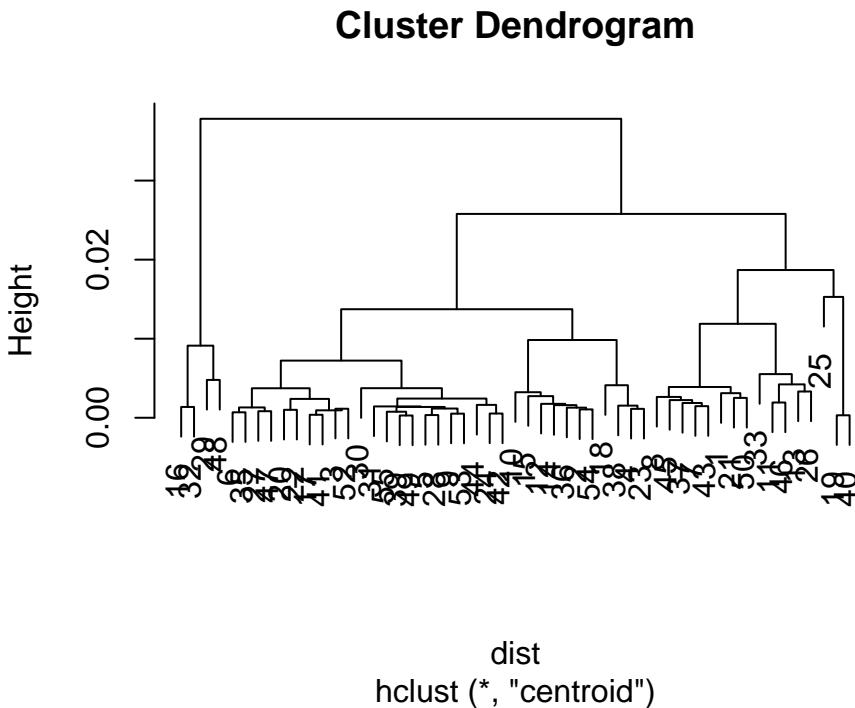
Según los gráficos obtenidos con el dataframe con todos los sujetos (sin eliminación de outliers) y con la matriz de distancias calculada con las 5 métricas de RM, los subgrupos no parecen separarse con dicha matriz de distancias.

```
# Matriz de distancias para las 5 métricas - Con eliminación de outliers
dist<-daisy(RATR_avg_demo_WoOut[,-c(1:7)],metric='euclidean')
distmx<-as.matrix(dist)

fviz_dist(dist)
```



```
clust<-hclust(dist, method = "centroid", members = NULL)
plot(clust)
```

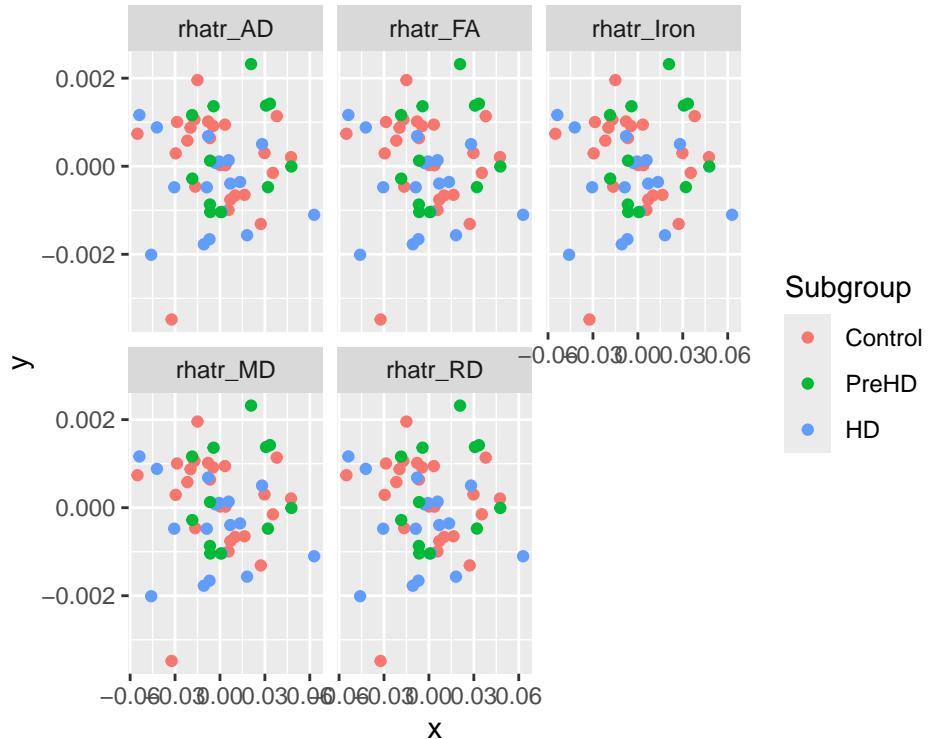


```
mds_dist<-cmdscale(distmx, eig=TRUE, k=2)
df_mds_dist<-data.frame(RATR_avg_demo_WoOut, x=mds_dist$points[,1],
```

```

y=mds_dist$points[,2])
df_mds_dist2<-df_mds_dist %>%
  dplyr::select(x, y, Subgroup, Gender, agef2, rhatr_Iron, rhatr_AD, rhatr_RD,
                rhatr_MD, rhatr_FA) %>%
  tidyverse::gather(key = "variable", value = "values", -x, -y, -Subgroup, -Gender, -agef2)
ggplot(df_mds_dist2, aes(x, y)) + geom_jitter(aes(col=Subgroup)) + facet_wrap(~variable)

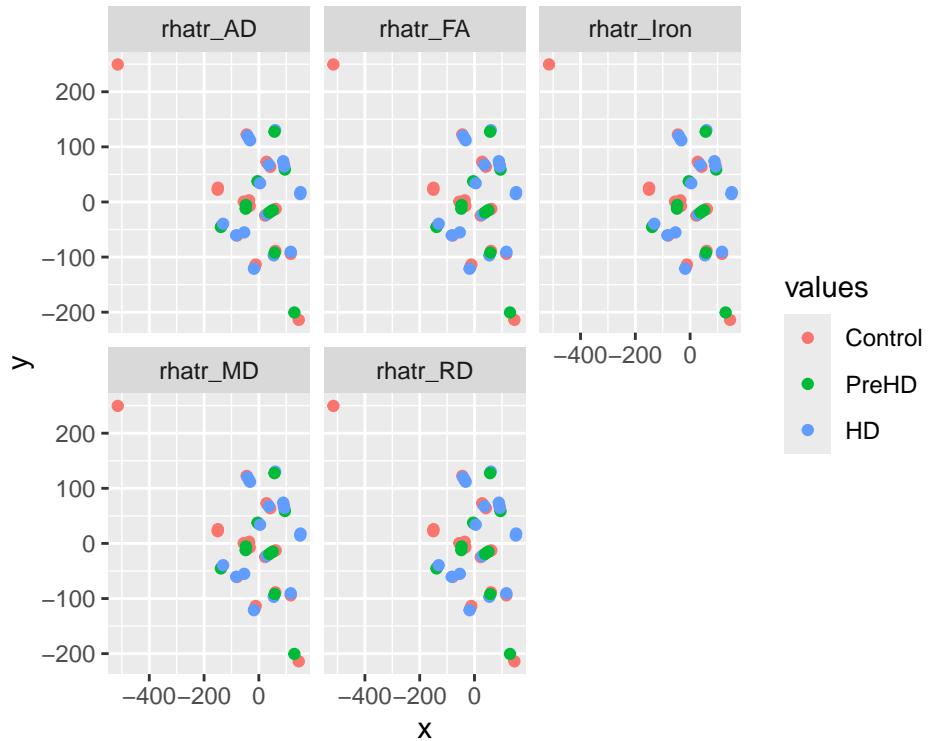
```



```

my_Rtsne<-Rtsne(X= distmx,perplexity= floor((nrow(RATR_avg_demo_WoOut) - 1) / 3),dims = 2)
# Maximum perplexity
my_Rtsne$perplexity
[1] 17
set.seed(123)
fit_tsne<-Rtsne(dist,perplexity = 1, is_distance = TRUE,
                  check_duplicates = FALSE, pca = FALSE, dims = 2, initial_dims = 50)
df_tsne_vars <- as_tibble(cbind(RATR_avg_demo_WoOut,
                                  x = fit_tsne$Y[, 1], y = fit_tsne$Y[, 2]))
df_tsne <- df_tsne_vars %>%
  dplyr::select(x, y, Subgroup, Gender, agef2, rhatr_Iron, rhatr_AD, rhatr_RD,
                rhatr_MD, rhatr_FA) %>%
  tidyverse::gather(key = "variable", value = "values", -x, -y, -Subgroup, -Gender, -agef2)
ggplot(df_tsne, aes(x, y, col=values)) + geom_point(aes(col=Subgroup)) + facet_wrap(~variable)

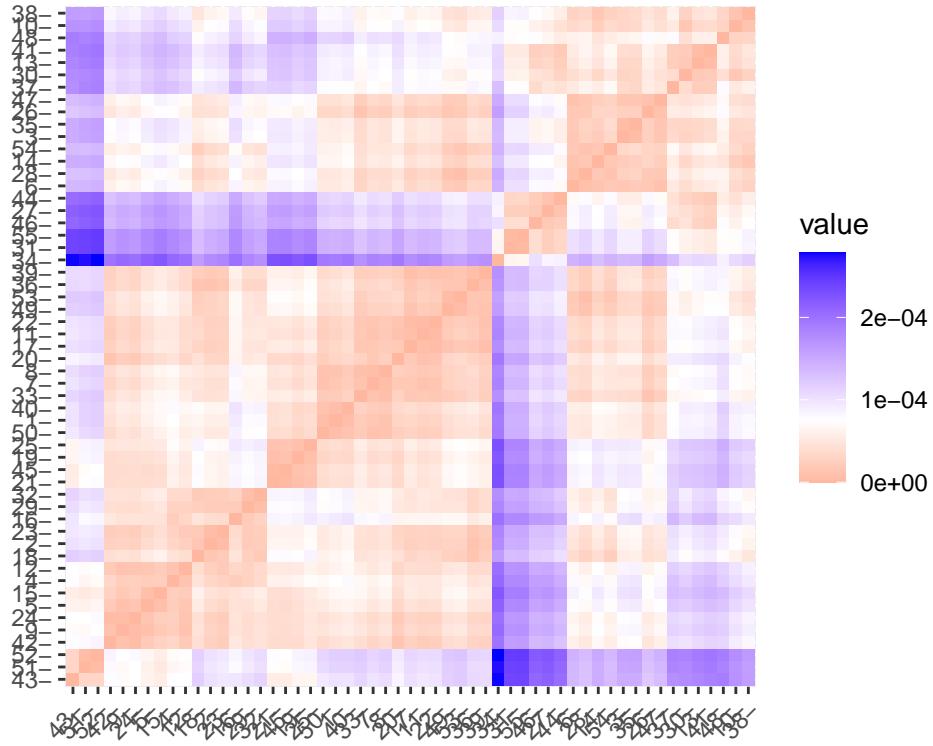
```



Según los gráficos obtenidos con el dataframe con eliminación de outliers y con las matriz de distancias calculada con las 5 métricas de RM, los subgrupos tampoco parecen separarse con dicha matriz de distancias.

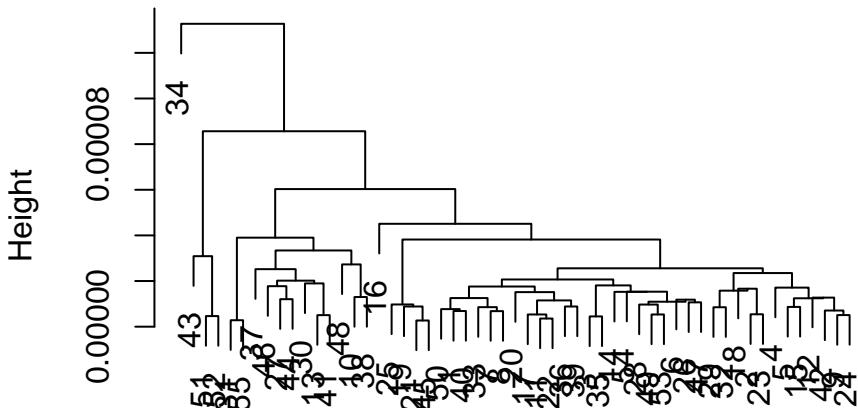
```
# Matriz de distancias para las 3 métricas de difusividad
dist<-daisy(RATR_avg_ADRDMD[,-c(1:7)],metric='euclidean')
distmx<-as.matrix(dist)

fviz_dist(dist)
```



```
clust<-hclust(dist, method = "centroid", members = NULL)
plot(clust)
```

**Cluster Dendrogram**



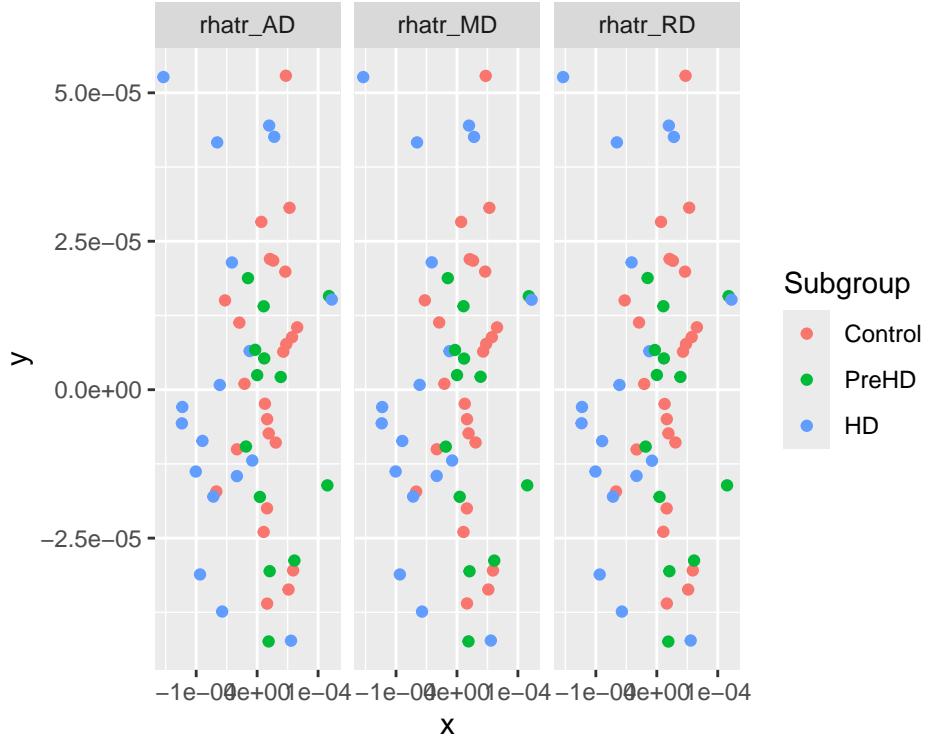
dist  
hclust (\*, "centroid")

```
mds_dist<-cmdscale(distmx, eig=TRUE, k=2)
df_mds_dist<-data.frame(RATR_avg_demo_ADRDMD, x=mds_dist$points[,1],
```

```

y=mds_dist$points[,2])
df_mds_dist2<-df_mds_dist %>%
  dplyr::select(x, y, Subgroup, Gender, agef2, rhatr_AD, rhatr_RD, rhatr_MD) %>%
  tidyverse::gather(key = "variable", value = "values", -x, -y, -Subgroup, -Gender, -agef2)
ggplot(df_mds_dist2, aes(x,y)) + geom_jitter(aes(col=Subgroup)) + facet_wrap(~variable)

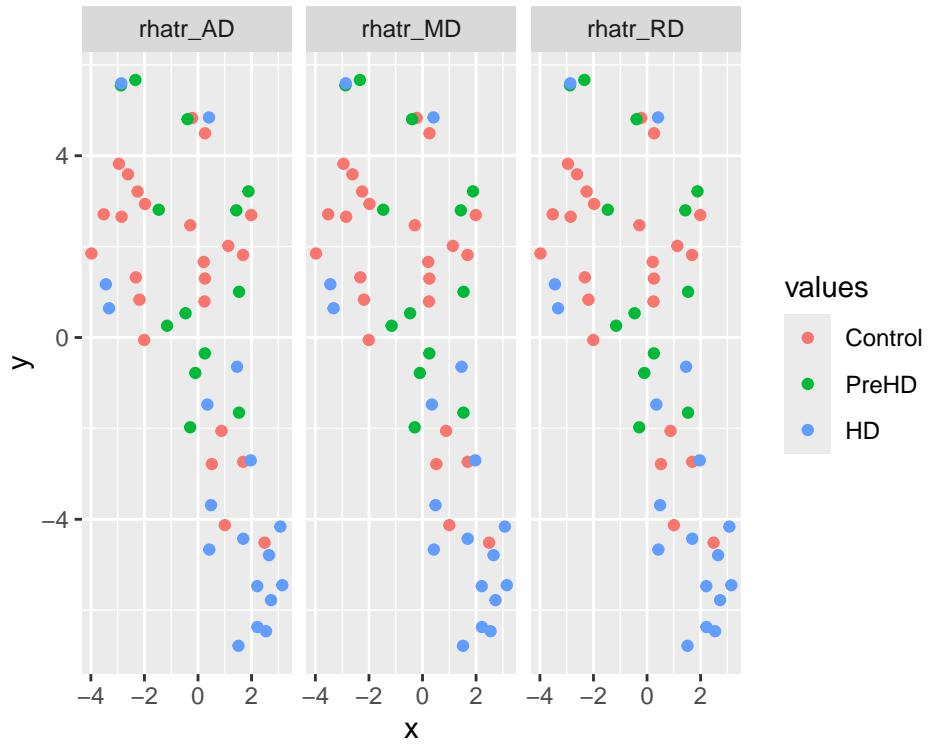
```



```

my_Rtsne<-Rtsne(X= distmx,perplexity= floor((nrow(RATR_avg_demo_ADRDMD) - 1) / 3),dims = 2)
# Maximum perplexity
my_Rtsne$perplexity
[1] 18
set.seed(123)
fit_tsne<-Rtsne(dist,perplexity = my_Rtsne$perplexity, is_distance = TRUE, check_duplicates = FALSE,
                  pca = FALSE, dims = 2, initial_dims = 50)
df_tsne_vars <- as_tibble(cbind(RATR_avg_demo_ADRDMD,
                                    x = fit_tsne$Y[, 1], y = fit_tsne$Y[, 2]))
df_tsne <- df_tsne_vars %>%
  dplyr::select(x, y, Subgroup, Gender, agef2, rhatr_AD, rhatr_RD, rhatr_MD) %>%
  tidyverse::gather(key = "variable", value = "values", -x, -y, -Subgroup, -Gender, -agef2)
ggplot(df_tsne, aes(x, y, col=values)) + geom_point(aes(col=Subgroup)) + facet_wrap(~variable)

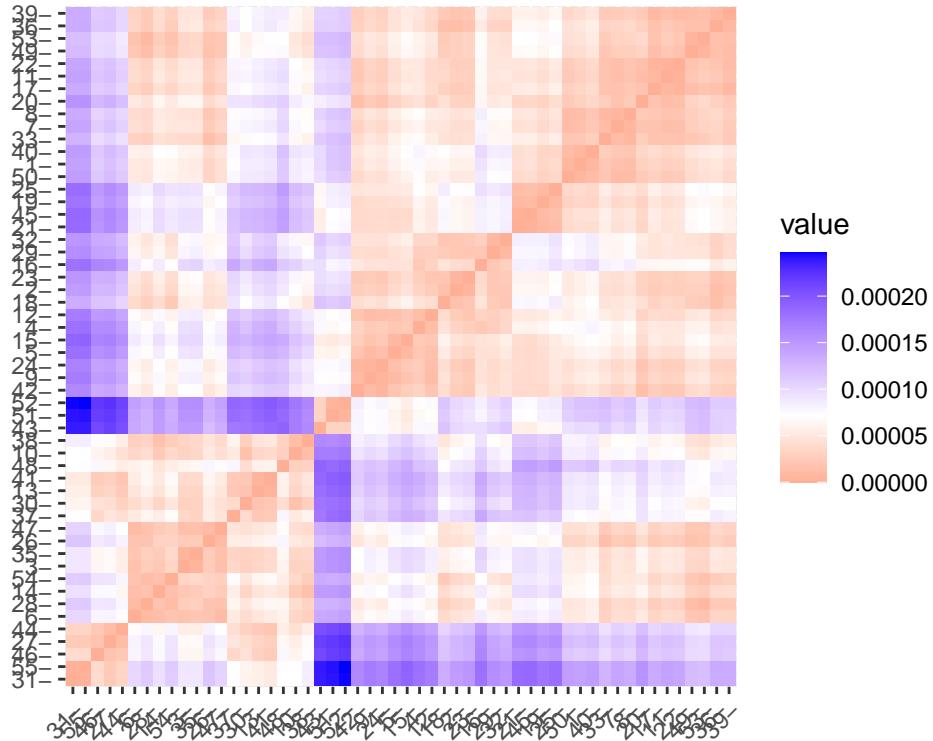
```



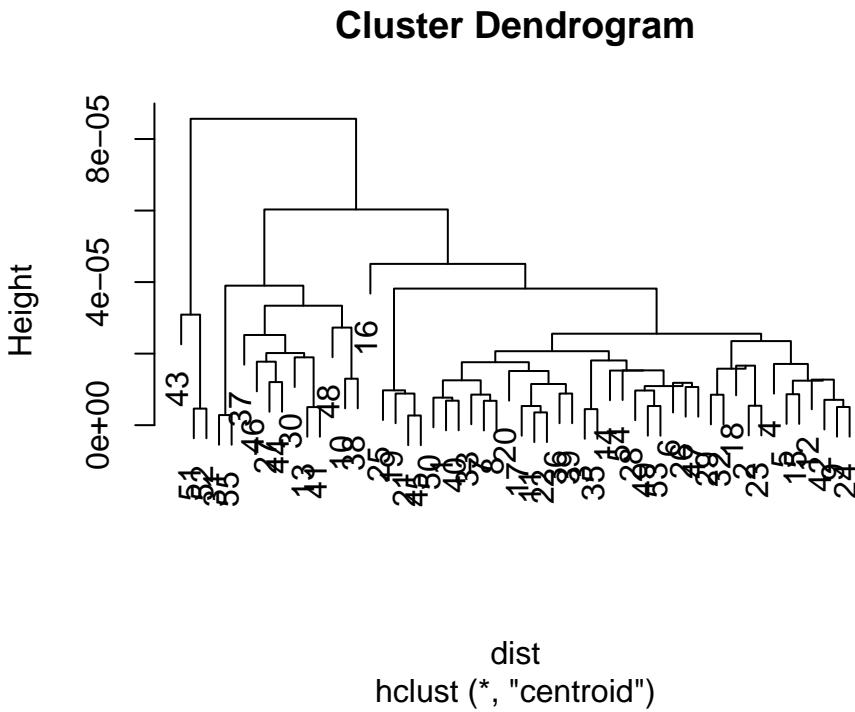
Según los gráficos obtenidos con el dataframe con todos los sujetos (sin eliminación de outliers) y con la matriz de distancias obtenida con las 3 métricas de difusividad, los subgrupos HD y preHD vs controles parecen separarse con dicha matriz de distancias, pero existe gran solapamiento entre los tres subgrupos.

```
# Matriz de distancias para las 3 métricas de difusividad
dist<-daisy(RATR_avg_WoOut_ADRDMD[,-c(1:7)],metric='euclidean')
distmx<-as.matrix(dist)

fviz_dist(dist)
```



```
clust<-hclust(dist, method = "centroid", members = NULL)
plot(clust)
```

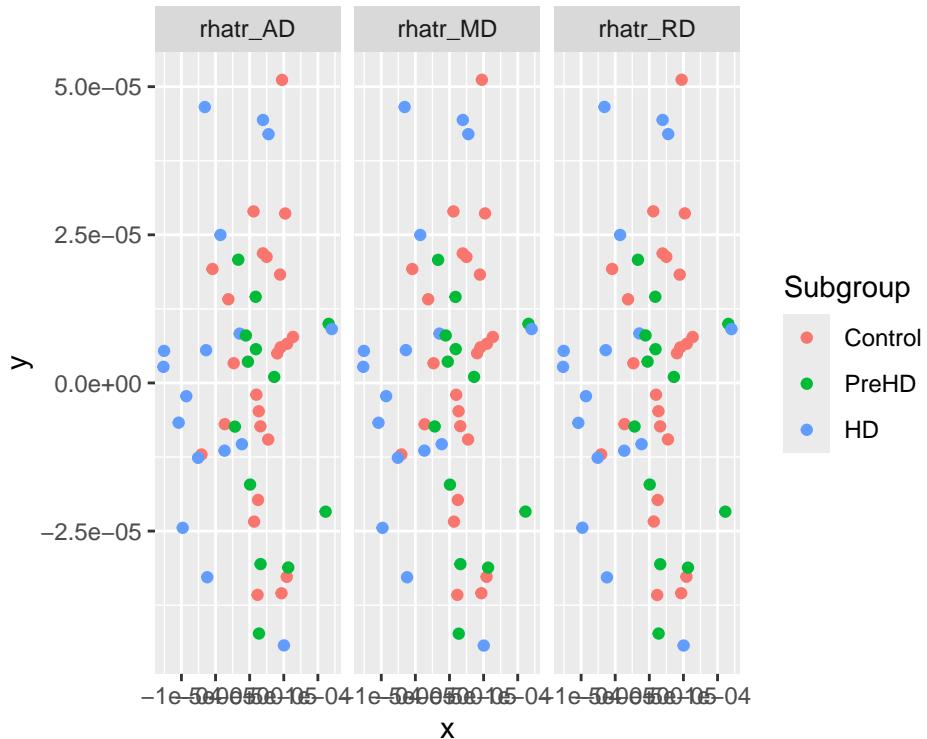


```
mds_dist<-cmdscale(distmx, eig=TRUE, k=2)
df_mds_dist<-data.frame(RATR_avg_demo_WoOut_ADRDMD, x=mds_dist$points[,1],
```

```

y=mds_dist$points[,2])
df_mds_dist2<-df_mds_dist %>%
  dplyr::select(x, y, Subgroup, Gender, agef2, rhatr_AD, rhatr_RD, rhatr_MD) %>%
  tidyverse::gather(key = "variable", value = "values", -x, -y, -Subgroup, -Gender, -agef2)
ggplot(df_mds_dist2, aes(x,y)) + geom_jitter(aes(col=Subgroup)) + facet_wrap(~variable)

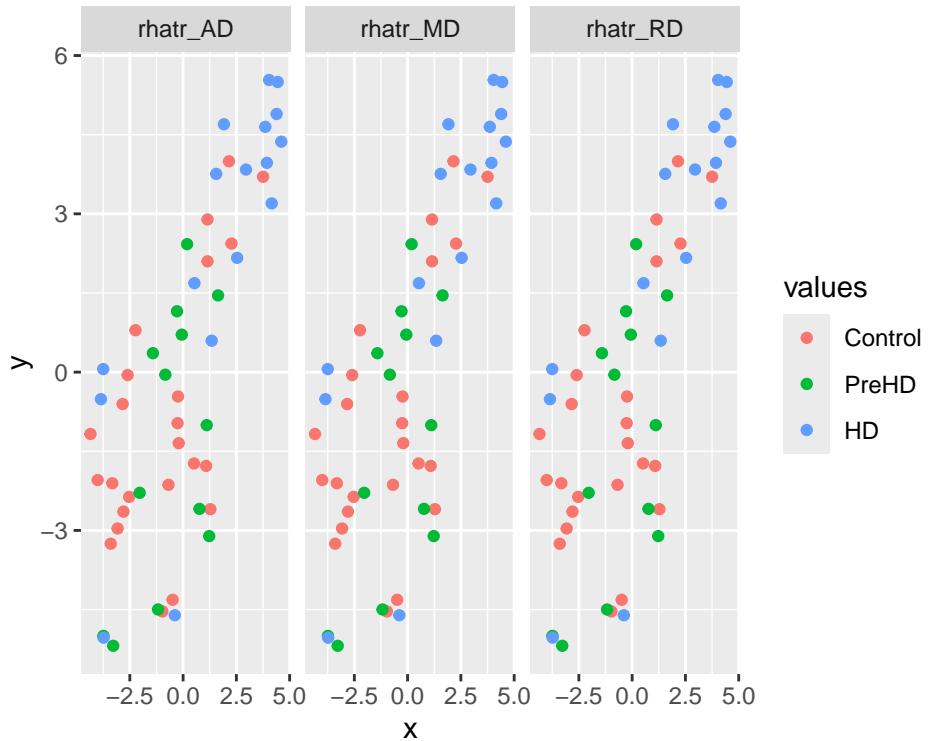
```



```

my_Rtsne<-Rtsne(X= distmx,perplexity=floor((nrow(RATR_avg_demo_WoOut_ADRDMD)-1)/3), dims=2)
# Maximum perplexity
my_Rtsne$perplexity
[1] 17
set.seed(123)
fit_tsne<-Rtsne(dist,perplexity = my_Rtsne$perplexity, is_distance = TRUE, check_duplicates = FALSE,
                  pca = FALSE, dims = 2, initial_dims = 50)
df_tsne_vars <- as_tibble(cbind(RATR_avg_demo_WoOut_ADRDMD,
                                  x = fit_tsne$Y[, 1], y = fit_tsne$Y[, 2]))
df_tsne <- df_tsne_vars %>%
  dplyr::select(x, y, Subgroup, Gender, agef2, rhatr_AD, rhatr_RD, rhatr_MD) %>%
  tidyverse::gather(key = "variable", value = "values", -x, -y, -Subgroup, -Gender, -agef2)
ggplot(df_tsne, aes(x, y, col=values)) + geom_point(aes(col=Subgroup)) + facet_wrap(~variable)

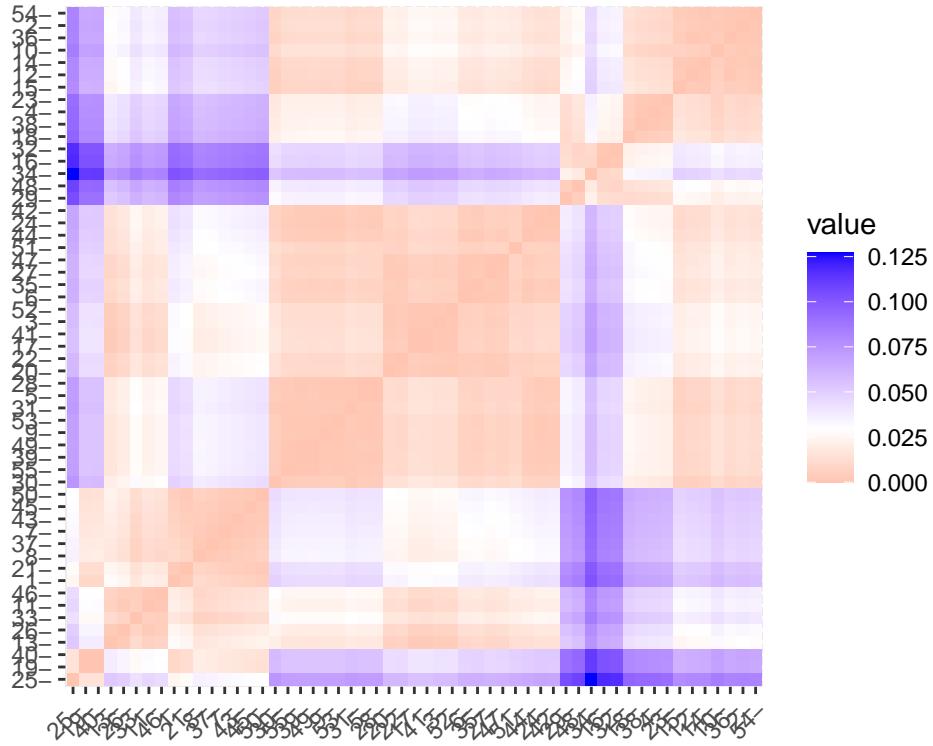
```



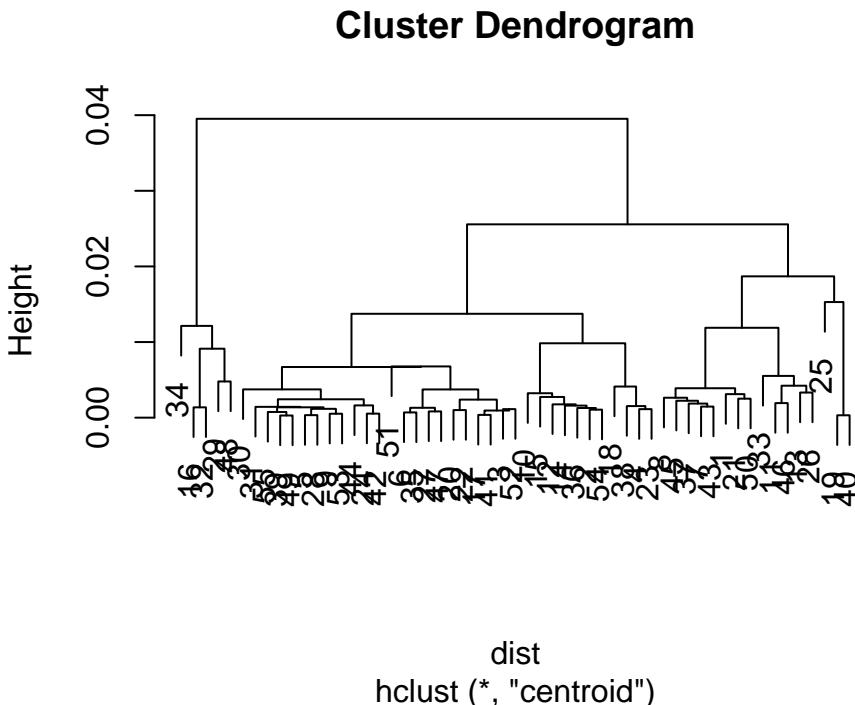
Según los gráficos obtenidos con el dataframe con eliminación de outliers y con la matriz de distancias obtenida con las 3 métricas de difusividad, los subgrupos HD y preHD vs controles también parecen separarse con dicha matriz de distancias, pero existe gran solapamiento entre los tres subgrupos.

```
# Matriz de distancias para las métricas de iron y FA
dist<-daisy(RATR_avg_IronFA[,-c(1:7)],metric='euclidean')
distmx<-as.matrix(dist)

fviz_dist(dist)
```



```
clust<-hclust(dist, method = "centroid", members = NULL)
plot(clust)
```

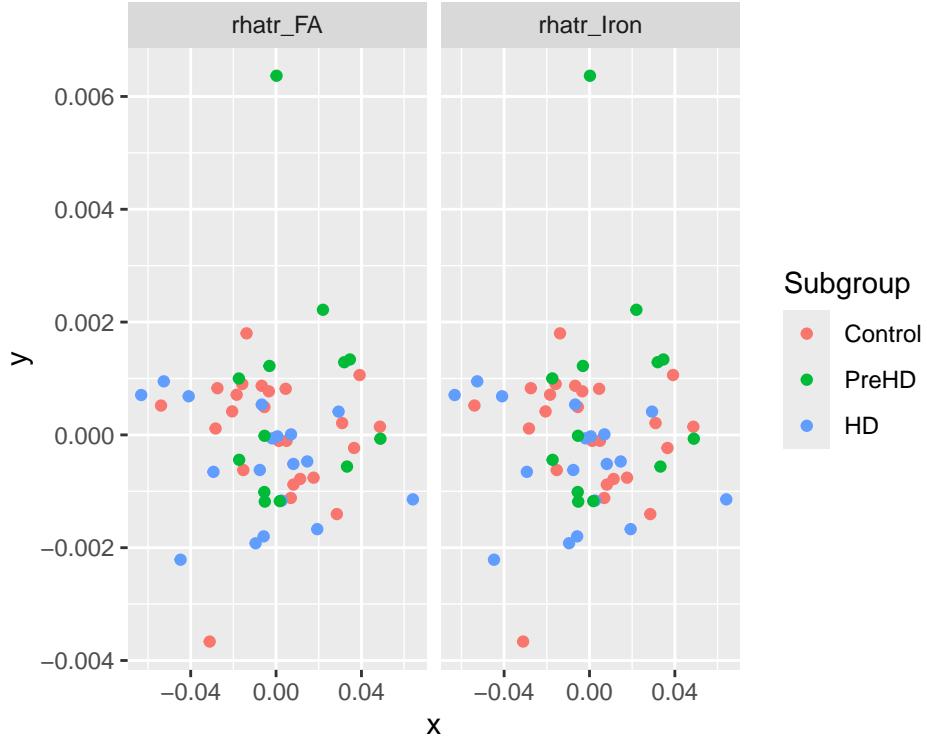


```
mds_dist<-cmdscale(distmx, eig=TRUE, k=2)
df_mds_dist<-data.frame(RATR_avg_demo_IronFA, x=mds_dist$points[,1],
```

```

y=mds_dist$points[,2])
df_mds_dist2<-df_mds_dist %>%
  dplyr::select(x, y, Subgroup, Gender, agef2, rhatr_Iron, rhatr_FA) %>%
  tidyrr::gather(key = "variable", value = "values", -x, -y, -Subgroup, -Gender, -agef2)
ggplot(df_mds_dist2, aes(x,y)) + geom_jitter(aes(col=Subgroup)) + facet_wrap(~variable)

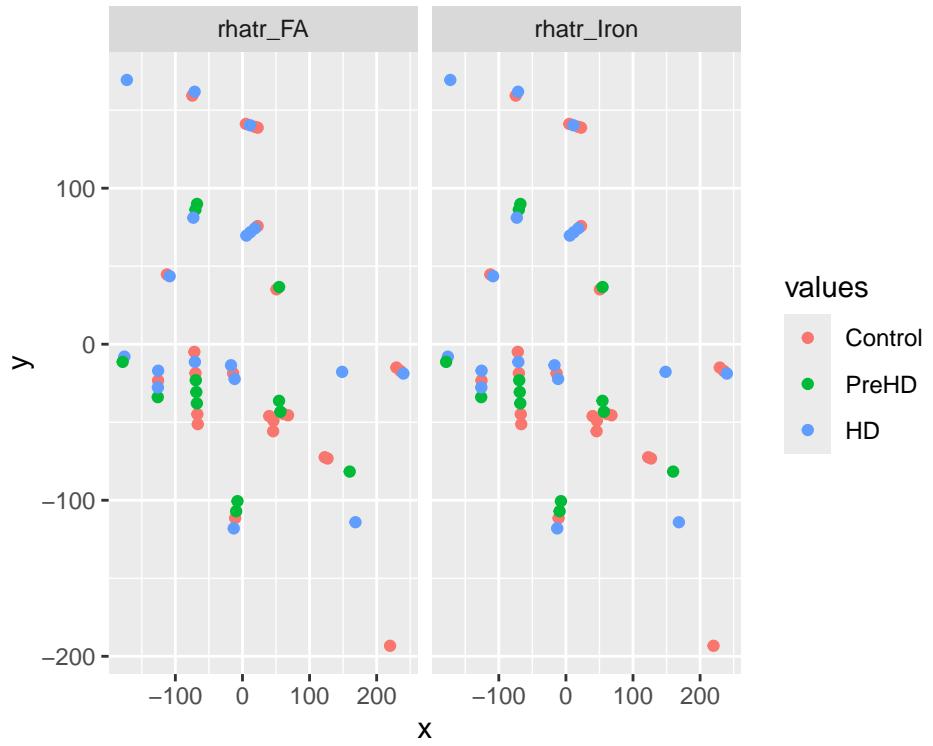
```



```

my_Rtsne<-Rtsne(X= distmx,perplexity= floor((nrow(RATR_avg_demo_IronFA) - 1) / 3),dims = 2)
# Maximum perplexity
my_Rtsne$perplexity
[1] 18
set.seed(123)
fit_tsne<-Rtsne(dist,perplexity = 1, is_distance = TRUE,
                  check_duplicates = FALSE, pca = FALSE, dims = 2, initial_dims = 50)
df_tsne_vars<-as_tibble(cbind(RATR_avg_demo_IronFA,x=fit_tsne$Y[,1],y=fit_tsne$Y[,2]))
df_tsne <- df_tsne_vars %>%
  dplyr::select(x, y, Subgroup, Gender, agef2, rhatr_Iron, rhatr_FA) %>%
  tidyrr::gather(key = "variable", value = "values", -x, -y, -Subgroup, -Gender, -agef2)
ggplot(df_tsne, aes(x, y, col=values)) + geom_point(aes(col=Subgroup)) + facet_wrap(~variable)

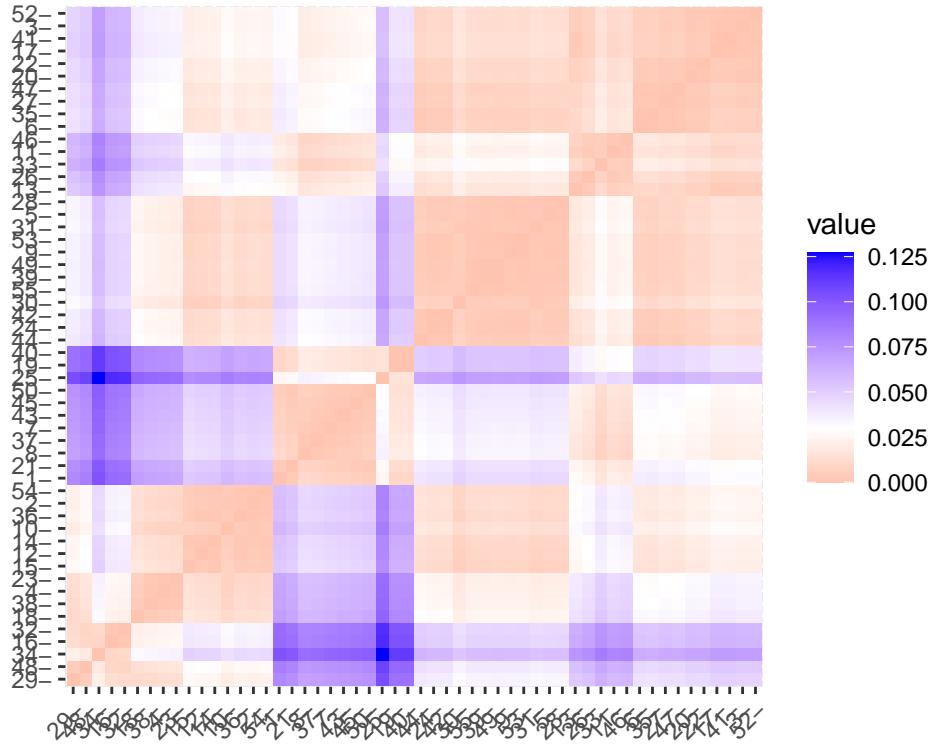
```



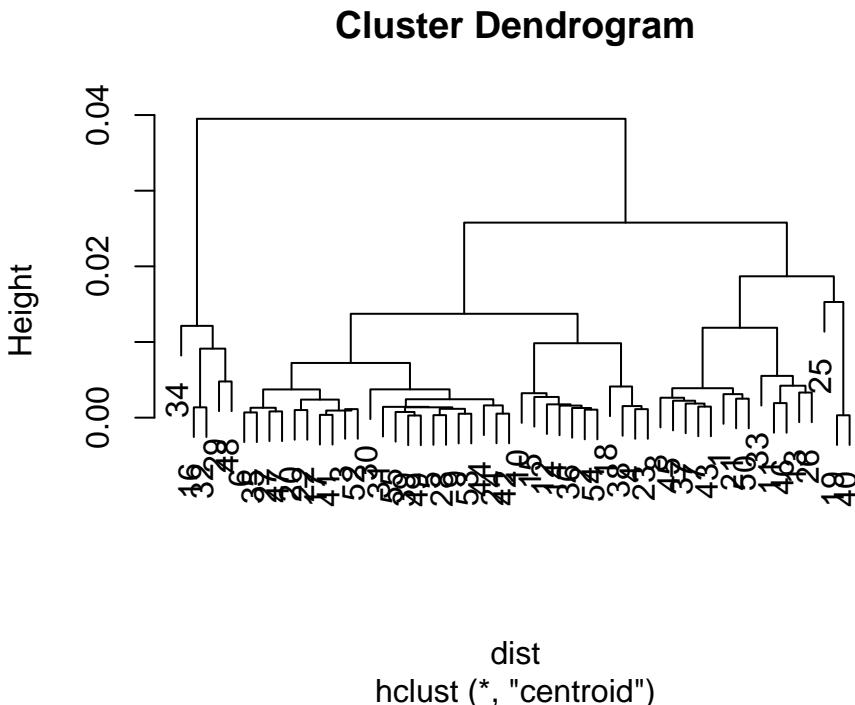
Según los gráficos obtenidos con el dataframe con todos los sujetos (sin eliminación de outliers) y la matriz de distancias obtenida de las métricas de iron y FA, los subgrupos no parecen separarse con dicha matriz de distancias.

```
# Matriz de distancias para las métricas de iron y FA
dist<-daisy(RATR_avg_demo_WoOut_IronFA[,-c(1:7)],metric='euclidean')
distmx<-as.matrix(dist)

fviz_dist(dist)
```



```
clust<-hclust(dist, method = "centroid", members = NULL)
plot(clust)
```

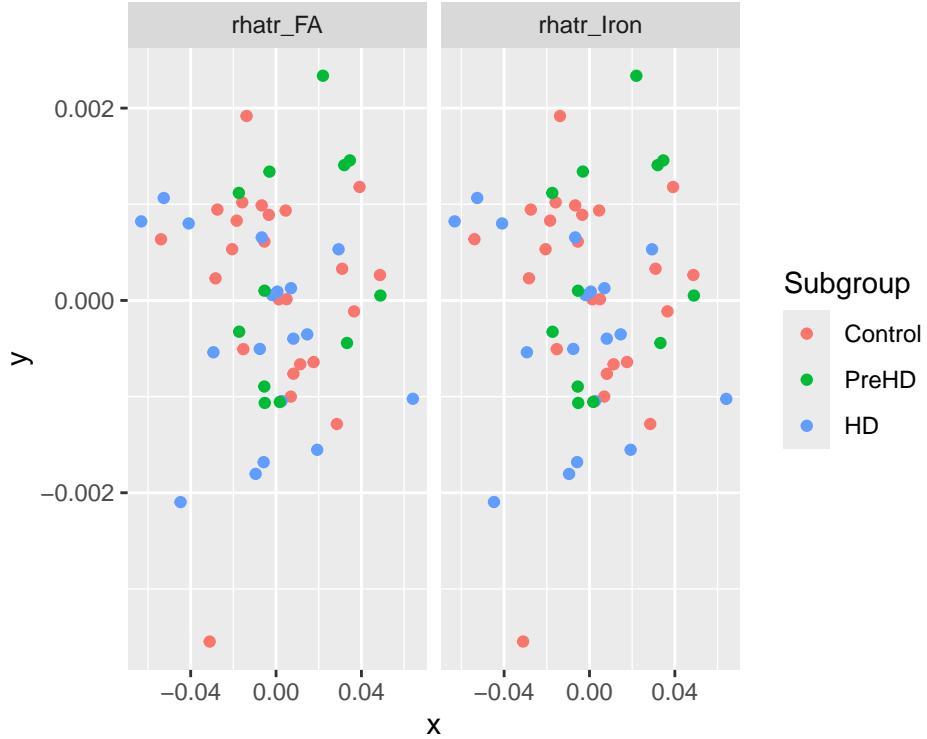


```
mds_dist<-cmdscale(distmx, eig=TRUE, k=2)
df_mds_dist<-data.frame(RATR_avg_demo_WoOut_IronFA, x=mds_dist$points[,1],
```

```

y=mds_dist$points[,2])
df_mds_dist2<-df_mds_dist %>%
  dplyr::select(x, y, Subgroup, Gender, agef2, rhatr_Iron, rhatr_FA) %>%
  tidyr::gather(key = "variable", value = "values", -x, -y, -Subgroup, -Gender, -agef2)
ggplot(df_mds_dist2, aes(x,y)) + geom_jitter(aes(col=Subgroup)) + facet_wrap(~variable)

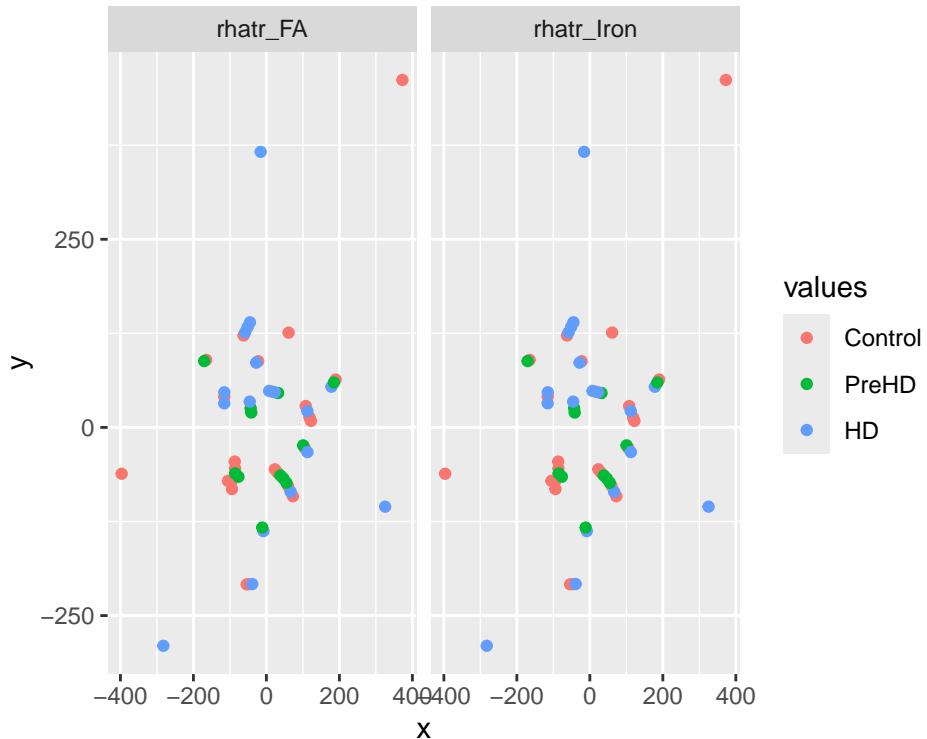
```



```

my_Rtsne<-Rtsne(X= distmx,perplexity= floor((nrow(RATR_avg_demo_WoOut_IronFA)-1)/3),dims = 2)
# Maximum perplexity
my_Rtsne$perplexity
[1] 17
set.seed(123)
fit_tsne<-Rtsne(dist,perplexity = 1, is_distance = TRUE,
                  check_duplicates = FALSE, pca = FALSE, dims = 2, initial_dims = 50)
df_tsne_vars<-as_tibble(cbind(RATR_avg_demo_WoOut_IronFA,x=fit_tsne$Y[,1],
                                y=fit_tsne$Y[,2]))
df_tsne <- df_tsne_vars %>%
  dplyr::select(x, y, Subgroup, Gender, agef2, rhatr_Iron, rhatr_FA) %>%
  tidyr::gather(key = "variable", value = "values", -x, -y, -Subgroup, -Gender, -agef2)
ggplot(df_tsne, aes(x, y, col=values)) + geom_point(aes(col=Subgroup)) + facet_wrap(~variable)

```



Según los gráficos obtenidos con el dataframe con eliminación de outliers y la matriz de distancias obtenida de las métricas de iron y FA, los subgrupos no parecen separarse con dicha matriz de distancias.

También se calculan las matrices de distancia con la función `vegdist` del paquete `vegan` para la combinación de métricas Iron-FA (por poder correlacionarse con el subgrupo preHD), AD-MD-RD (por poder correlacionarse con el subgrupo HD), y, para las 5 métricas en su conjunto (método euclídeo porque son variables numéricas continuas, no derivadas de especies). Dichas matrices se utilizarán para generar los modelos PERMANOVA. Las variables categóricas no se introducen en la matriz de distancias porque se utilizan como variables independientes en los modelos PERMANOVA.

Como se objetiva a continuación, se han calculado las matrices para ambos supuestos, con y sin eliminación de outliers. Teóricamente, PERMANOVA, al ser una técnica no paramétrica, podría ejecutarse con todos los sujetos ya que no necesita normalidad multivariante ni homocedasticidad de la matriz de covarianzas, pero, al basarse en permutaciones, requiere verificar una betadispersión homogénea entre grupos. Así que se generarán modelos PERMANOVA con y sin eliminación de outliers.

```
library(vegan)
# Matriz de distancias para las 5 métricas
distDifIron<-vegdist(RATR_avg_demo[,-c(1:7)],method='euclidean')
distDifIron_WoOut<-vegdist(RATR_avg_demo_WoOut[,-c(1:7)],method='euclidean')

# Matriz de distancias para las 3 métricas de difusividad (AD, RD, MD)
distADMRD<-vegdist(RATR_avg_demo_ADRDMD[,c('rhatr_AD','rhatr_RD','rhatr_MD')], 
                     method='euclidean')
distADMRD_WoOut<-vegdist(RATR_avg_demo_WoOut_ADRDMD[,c('rhatr_AD','rhatr_RD',
                           'rhatr_MD')], 
                           method='euclidean')

# Matriz de distancias para las métricas de Iron y FA
distIronFA<-vegdist(RATR_avg_demo_IronFA[,c('rhatr_Iron','rhatr_FA')], 
                     method='euclidean')
distIronFA_WoOut<-vegdist(RATR_avg_demo_WoOut_IronFA[,c('rhatr_Iron','rhatr_FA')],
```

```
method='euclidean')
```

A continuación, se generan los diferentes modelos PERMANOVA (remarcar que no se puede añadir el efecto aleatorio). Primero, los modelos se ejecutan con las 3 variables independientes en interacción. Segundo, se eliminan las interacciones no significativas para simplificar los modelos.

```
library(vegan)
# Establecer seed para que el resultado de las permutaciones sea siempre el mismo
set.seed(123)

# Modelos para las 5 métricas
permDifIron1<-adonis2(distDifIron~Subgroup*agef2*Gender, data=RATR_avg_demo,
                        permutations=999, by='terms');permDifIron1[1:5]
  Df SumOfSqs      R2      F Pr(>F)
Subgroup        2 0.001850 0.04773 1.1336  0.333
agef2           2 0.000406 0.01047 0.2487  0.786
Gender          1 0.000010 0.00025 0.0121  0.922
Subgroup:agef2  4 0.001156 0.02983 0.3542  0.835
Subgroup:Gender 2 0.000450 0.01161 0.2759  0.761
agef2:Gender    2 0.001148 0.02961 0.7033  0.493
Subgroup:agef2:Gender 2 0.001920 0.04951 1.1760  0.287
Residual        39 0.031830 0.82098
Total           54 0.038770 1.00000
permDifIron2<-adonis2(distDifIron~Subgroup+agef2+Gender, data=RATR_avg_demo,
                        permutations=999, by='terms');permDifIron2[1:5]
  Df SumOfSqs      R2      F Pr(>F)
Subgroup        2 0.001850 0.04773 1.2419  0.281
agef2           2 0.000406 0.01047 0.2724  0.788
Gender          1 0.000010 0.00025 0.0132  0.930
Residual        49 0.036504 0.94155
Total           54 0.038770 1.00000
permDifIron3<-adonis2(distDifIron~Subgroup, data=RATR_avg_demo, permutations=999,
                        by='terms');permDifIron3[1:5]
  Df SumOfSqs      R2      F Pr(>F)
Subgroup        2 0.00185 0.04773 1.3031  0.287
Residual        52 0.03692 0.95227
Total           54 0.03877 1.00000

permDifIron1_WoOut<-adonis2(distDifIron_WoOut~Subgroup*agef2*Gender,
                             data=RATR_avg_demo_WoOut, permutations=999,
                             by='terms');permDifIron1_WoOut[1:5]
  Df SumOfSqs      R2      F Pr(>F)
Subgroup        2 0.001268 0.03658 0.8155  0.470
agef2           2 0.000194 0.00558 0.1245  0.903
Gender          1 0.000109 0.00314 0.1398  0.726
Subgroup:agef2  4 0.001742 0.05024 0.5600  0.680
Subgroup:Gender 2 0.000281 0.00812 0.1810  0.848
agef2:Gender    2 0.001268 0.03657 0.8153  0.462
Subgroup:agef2:Gender 2 0.001035 0.02985 0.6654  0.519
Residual        37 0.028767 0.82992
Total           52 0.034662 1.00000
permDifIron2_WoOut<-adonis2(distDifIron_WoOut~Subgroup+agef2+Gender,
                             data=RATR_avg_demo_WoOut, permutations=999,
                             by='terms');permDifIron2_WoOut[1:5]
  Df SumOfSqs      R2      F Pr(>F)
Subgroup        2 0.001268 0.03658 0.9005  0.414
```

```

agef2      2 0.000194 0.00558 0.1375  0.868
Gender     1 0.000109 0.00314 0.1543  0.669
Residual   47 0.033092 0.95470
Total      52 0.034662 1.00000

permDifIron3_WoOut<-adonis2(distDifIron_WoOut~Subgroup, data=RATR_avg_demo_WoOut,
                               permutations=999, by='terms');permDifIron3_WoOut[1:5]
    Df SumOfSqs      R2      F Pr(>F)
Subgroup   2 0.001268 0.03658 0.9493  0.393
Residual   50 0.033394 0.96342
Total      52 0.034662 1.00000

# Modelos para las 3 métricas de difusividad (AD, RD, MD)

permADMDRD1<-adonis2(distADMDRD~Subgroup*agef2*Gender, data=RATR_avg_demo_ADRDMD,
                        permutations=999, by='terms');permADMDRD1[1:5]
    Df SumOfSqs      R2      F Pr(>F)
Subgroup   2 5.5411e-08 0.24873 8.0897  0.001 ***
agef2      2 1.3646e-08 0.06126 1.9923  0.107
Gender     1 4.0700e-10 0.00183 0.1188  0.845
Subgroup:agef2 4 1.0722e-08 0.04813 0.7827  0.550
Subgroup:Gender 2 9.0700e-10 0.00407 0.1325  0.950
agef2:Gender 2 4.7200e-09 0.02119 0.6891  0.534
Subgroup:agef2:Gender 2 3.3970e-09 0.01525 0.4959  0.660
Residual    39 1.3357e-07 0.59955
Total       54 2.2278e-07 1.00000
---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

permADMDRD2<-adonis2(distADMDRD~Subgroup+agef2+Gender, data=RATR_avg_demo_ADRDMD,
                        permutations=999, by='terms');permADMDRD2[1:5]
    Df SumOfSqs      R2      F Pr(>F)
Subgroup   2 5.5411e-08 0.24873 8.8549  0.001 ***
agef2      2 1.3646e-08 0.06126 2.1807  0.107
Gender     1 4.0700e-10 0.00183 0.1300  0.834
Residual   49 1.5331e-07 0.68819
Total      54 2.2278e-07 1.00000
---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

permADMDRD3<-adonis2(distADMDRD~Subgroup, data=RATR_avg_demo_ADRDMD, permutations=999,
                        by='terms');permADMDRD3[1:5]
    Df SumOfSqs      R2      F Pr(>F)
Subgroup   2 5.5411e-08 0.24873 8.608  0.001 ***
Residual   52 1.6736e-07 0.75127
Total      54 2.2278e-07 1.00000
---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

permADMDRD1_WoOut<-adonis2(distADMDRD_WoOut~Subgroup*agef2*Gender,
                               data=RATR_avg_demo_WoOut_ADRDMD, permutations=999,
                               by='terms');permADMDRD1_WoOut[1:5]
    Df SumOfSqs      R2      F Pr(>F)
Subgroup   2 4.3745e-08 0.22334 6.4653  0.002 **
agef2      2 8.9120e-09 0.04550 1.3172  0.277
Gender     1 9.9700e-10 0.00509 0.2946  0.664
Subgroup:agef2 4 7.1410e-09 0.03646 0.5277  0.791
Subgroup:Gender 2 3.6400e-10 0.00186 0.0538  0.992
agef2:Gender 2 3.6200e-09 0.01848 0.5350  0.619
Subgroup:agef2:Gender 2 2.5330e-09 0.01293 0.3744  0.754

```

```

Residual           38 1.2856e-07 0.65634
Total             53 1.9587e-07 1.00000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
permADMDRD2_WoOut<-adonis2(distADMDRD_WoOut~Subgroup+agef2+Gender,
                               data=RATR_avg_demo_WoOut_ADRDMD, permutations=999,
                               by='terms');permADMDRD2_WoOut[1:5]
      Df   SumOfSqs      R2      F Pr(>F)
Subgroup    2 4.3745e-08 0.22334 7.3824  0.001 ***
agef2       2 8.9120e-09 0.04550 1.5040  0.208
Gender      1 9.9700e-10 0.00509 0.3363  0.649
Residual    48 1.4222e-07 0.72607
Total       53 1.9587e-07 1.00000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
permADMDRD3_WoOut<-adonis2(distADMDRD_WoOut~Subgroup, data=RATR_avg_demo_WoOut_ADRDMD,
                               permutations=999, by='terms');permADMDRD3_WoOut[1:5]
      Df   SumOfSqs      R2      F Pr(>F)
Subgroup    2 4.3745e-08 0.22334 7.3328  0.002 **
Residual   51 1.5212e-07 0.77666
Total       53 1.9587e-07 1.00000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Modelos para las métricas de iron y FA
permIronFA1<-adonis2(distIronFA~Subgroup*agef2*Gender, data=RATR_avg_demo_IronFA,
                       permutations=999, by='terms');permIronFA1[1:5]
      Df   SumOfSqs      R2      F Pr(>F)
Subgroup    2 0.001850 0.04773 1.1336  0.348
agef2       2 0.000406 0.01047 0.2487  0.771
Gender      1 0.000010 0.00025 0.0121  0.922
Subgroup:agef2  4 0.001156 0.02983 0.3542  0.826
Subgroup:Gender  2 0.000450 0.01161 0.2759  0.763
agef2:Gender   2 0.001148 0.02961 0.7033  0.516
Subgroup:agef2:Gender  2 0.001920 0.04951 1.1760  0.314
Residual     39 0.031829 0.82099
Total        54 0.038770 1.00000
permIronFA2<-adonis2(distIronFA~Subgroup+agef2+Gender, data=RATR_avg_demo_IronFA,
                       permutations=999, by='terms');permIronFA2[1:5]
      Df   SumOfSqs      R2      F Pr(>F)
Subgroup    2 0.001850 0.04773 1.2419  0.316
agef2       2 0.000406 0.01047 0.2724  0.770
Gender      1 0.000010 0.00025 0.0132  0.912
Residual   49 0.036504 0.94155
Total       54 0.038770 1.00000
permIronFA3<-adonis2(distIronFA~Subgroup, data=RATR_avg_demo_IronFA,
                       permutations=999, by='terms');permIronFA3[1:5]
      Df   SumOfSqs      R2      F Pr(>F)
Subgroup    2 0.001850 0.04773 1.3031  0.276
Residual   52 0.036919 0.95227
Total       54 0.038770 1.00000

permIronFA1_WoOut<-adonis2(distIronFA_WoOut~Subgroup*agef2*Gender,
                             data=RATR_avg_demo_WoOut_IronFA, permutations=999,
                             by='terms');permIronFA1_WoOut[1:5]
      Df   SumOfSqs      R2      F Pr(>F)

```

```

Subgroup          2 0.001930 0.04984 1.1657  0.345
agef2            2 0.000452 0.01167 0.2730  0.760
Gender           1 0.000013 0.00033 0.0156  0.906
Subgroup:agef2   4 0.001356 0.03502 0.4095  0.814
Subgroup:Gender  2 0.000450 0.01163 0.2720  0.771
agef2:Gender     2 0.001148 0.02964 0.6934  0.498
Subgroup:agef2:Gender 2 0.001920 0.04956 1.1593  0.340
Residual         38 0.031459 0.81231
Total            53 0.038729 1.00000
permIronFA2_WoOut<-adonis2(distIronFA_WoOut~Subgroup+agef2+Gender,
                             data=RATR_avg_demo_WoOut_IronFA, permutations=999,
                             by='terms');permIronFA2_WoOut[1:5]
  Df SumOfSqs      R2      F Pr(>F)
Subgroup    2 0.001930 0.04984 1.2749  0.298
agef2       2 0.000452 0.01167 0.2986  0.737
Gender       1 0.000013 0.00033 0.0170  0.908
Residual    48 0.036334 0.93816
Total        53 0.038729 1.00000
permIronFA3_WoOut<-adonis2(distIronFA_WoOut~Subgroup, data=RATR_avg_demo_WoOut_IronFA,
                             permutations=999, by='terms');permIronFA3_WoOut[1:5]
  Df SumOfSqs      R2      F Pr(>F)
Subgroup    2 0.001930 0.04984 1.3375  0.266
Residual    51 0.036798 0.95016
Total        53 0.038729 1.00000

```

Los **modelos PERMANOVA generados con las 5 métricas**, ya sea con el dataframe completo o con el dataframe reducido eliminando los outliers, no evidencian factores significativos.

Los **modelos PERMANOVA generados con las 3 métricas de difusividad**, ya sea con el dataframe completo o con el dataframe reducido eliminando los outliers, evidencian el subgrupo como único factor significativo.

Los **modelos PERMANOVA generados con las métricas de iron y FA**, ya sea con el dataframe completo o con el dataframe reducido eliminando los outliers, no evidencian factores significativos.

**Por tanto, PERMANOVA ha sido menos sensible que MANOVA para detectar significación estadística en la evaluación de las 5 métricas conjuntamente.**

Se calculan las **dispersiones** para cada uno de los modelos que se han generado en el paso anterior y tan sólo se han calculado para el factor subgrupo dado que es el único que ha dado significativo.

```

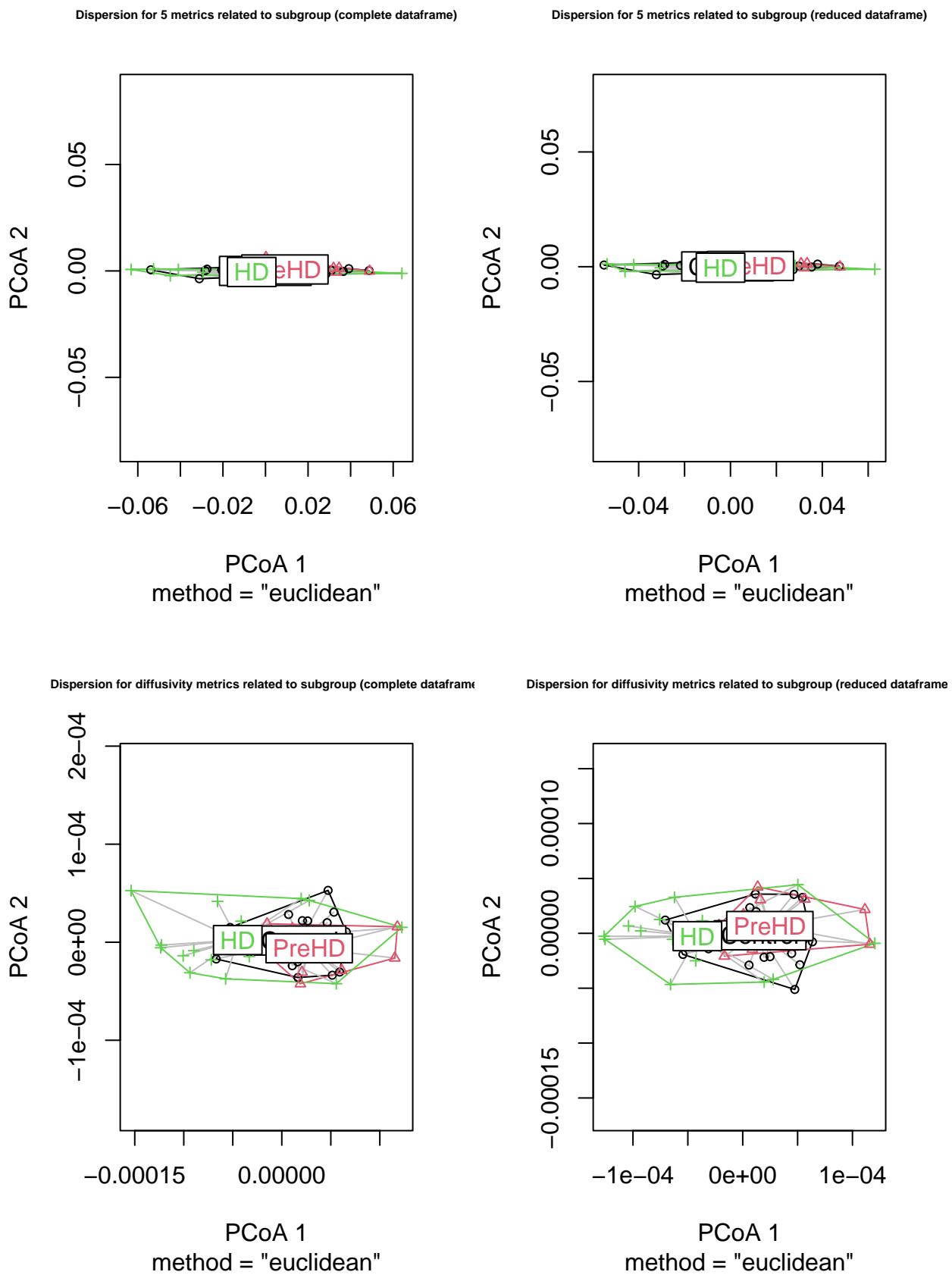
dispersionDifIron<-betadisper(d=distDifIron, group=RATR_avg_demo$Subgroup,
                                 type="centroid", bias.adjust=TRUE)
dispersionDifIron_WoOut<-betadisper(d=distDifIron_WoOut,
                                      group=RATR_avg_demo_WoOut$Subgroup,
                                      type="centroid", bias.adjust=TRUE)

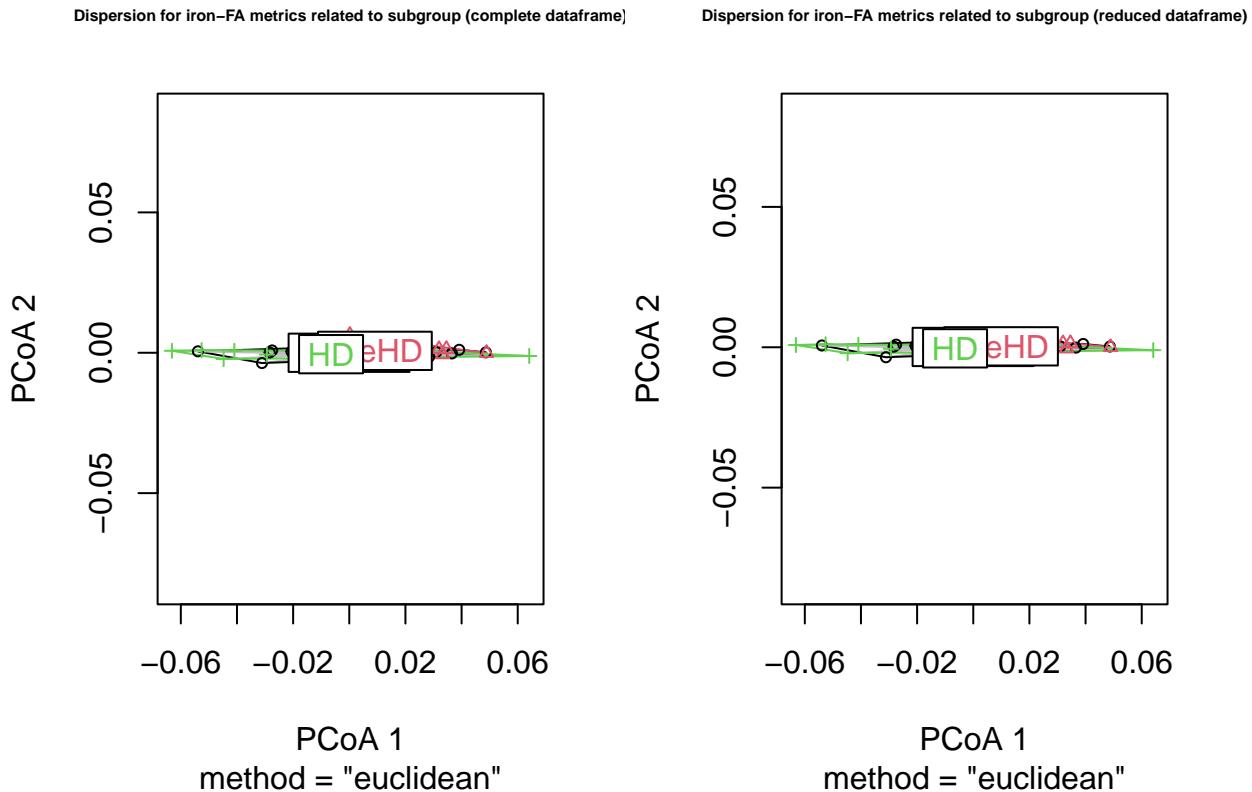
dispersionADMDRD<-betadisper(d=distADMDRD, group=RATR_avg_demo_ADRDMD$Subgroup,
                                 type="centroid", bias.adjust=TRUE)
dispersionADMDRD_WoOut<-betadisper(d=distADMDRD_WoOut,
                                      group=RATR_avg_demo_WoOut_ADRDMD$Subgroup,
                                      type="centroid", bias.adjust=TRUE)

dispersionIronFA<-betadisper(d=distIronFA, group=RATR_avg_demo_IronFA$Subgroup,
                               type="centroid", bias.adjust=TRUE)
dispersionIronFA_WoOut<-betadisper(d=distIronFA_WoOut,
                                      group=RATR_avg_demo_WoOut_IronFA$Subgroup,
                                      type="centroid", bias.adjust=TRUE)

```

Se generan **plots** para cada una de las dispersiones.





Cabe mencionar que el subgrupo HD presenta una mayor dispersión respecto a los subgrupos preHD y control, y el subgrupo preHD presenta una menor dispersión respecto a los subgrupos HD y control.

Se generan **modelos anova** para cada una de las dispersiones.

```
summary(aov(dispersionDifIron$distances~RATR_avg_demo$Subgroup))
      Df   Sum Sq   Mean Sq F value Pr(>F)
RATR_avg_demo$Subgroup  2 0.000107 5.351e-05     0.2   0.819
Residuals                 52 0.013916 2.676e-04

summary(aov(dispersionDifIron_WoOut$distances~RATR_avg_demo_WoOut$Subgroup))
      Df   Sum Sq   Mean Sq F value Pr(>F)
RATR_avg_demo_WoOut$Subgroup  2 0.000005 2.570e-06     0.01   0.99
Residuals                     50 0.012542 2.508e-04

summary(aov(dispersionADMDRD$distances~RATR_avg_demo_ADRDMD$Subgroup))
      Df   Sum Sq   Mean Sq F value Pr(>F)
RATR_avg_demo_ADRDMD$Subgroup  2 8.000e-09 3.999e-09    4.537 0.0153 *
Residuals                     52 4.583e-08 8.810e-10
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
summary(aov(dispersionADMDRD_WoOut$distances~RATR_avg_demo_WoOut_ADRDMD$Subgroup))
      Df   Sum Sq   Mean Sq F value Pr(>F)
RATR_avg_demo_WoOut_ADRDMD$Subgroup  2 6.070e-09 3.034e-09    3.826 0.0283 *
Residuals                     51 4.044e-08 7.930e-10
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
summary(aov(dispersionIronFA$distances~RATR_avg_demo_IronFA$Subgroup))
      Df   Sum Sq   Mean Sq F value Pr(>F)
```

```

RATR_avg_demo_IronFA$Subgroup 2 0.000107 0.0000535      0.2  0.819
Residuals                      52 0.013917 0.0002676
summary(aov(dispersionIronFA_WoOut$distances~RATR_avg_demo_WoOut_IronFA$Subgroup))
                                         Df Sum Sq Mean Sq F value Pr(>F)
RATR_avg_demo_WoOut_IronFA$Subgroup 2 0.000088 4.413e-05   0.164  0.85
Residuals                          51 0.013762 2.698e-04

```

Las dispersiones de las 3 métricas de difusividad presentan significación estadística, con lo que no se puede determinar si las diferencias observadas en PERMANOVA son debidas a diferencias en el posicionamiento de los centroides y/o a las diferencias en dispersión. Como se mencionó más arriba, no se puede determinar cuánta diferencia en el PERMANOVA es debida al posicionamiento de centroides y cuánta debida a la variabilidad en dispersión.

Se decide intentar la **transformación logarítmica de las 3 métricas de difusividad para determinar si se mejora la dispersión**. A continuación, se exponen los resultados para el **dataframe completo**.

```

# Se genera la transformación logarítmica
RATR_avg_demo_ADRDMD_log<-cbind(RATR_avg_demo_ADRDMD[,1:7],
                                    log(RATR_avg_demo_ADRDMD[,8:10]))

# Se calculan las distancias
distADMDRD_log<-vegdist(RATR_avg_demo_ADRDMD_log[,c('rhatr_AD','rhatr_MD','rhatr_RD')],
                           method='euclidean')

# Se generan los modelos PERMANOVA
set.seed(123)
permADMDRD_log1<-adonis2(distADMDRD_log~Subgroup*agef2*Gender,
                           data=RATR_avg_demo_ADRDMD_log, permutations=999, by='terms')
permADMDRD_log1[1:5]

```

	Df	SumOfSqs	R2	F	Pr(>F)
Subgroup	2	0.11955	0.24689	7.8582	0.002 **
agef2	2	0.02999	0.06194	1.9715	0.147
Gender	1	0.00122	0.00253	0.1609	0.803
Subgroup:agef2	4	0.02011	0.04154	0.6610	0.653

```

permADMDRD_log1[1:5]

```

	Df	SumOfSqs	R2	F	Pr(>F)
Subgroup	2	0.11955	0.24689	7.8582	0.002 **
agef2	2	0.02999	0.06194	1.9715	0.147
Gender	1	0.00122	0.00253	0.1609	0.803
Subgroup:agef2	4	0.02011	0.04154	0.6610	0.653
Subgroup:Gender	2	0.00222	0.00459	0.1460	0.931
agef2:Gender	2	0.00831	0.01717	0.5464	0.632
Subgroup:agef2:Gender	2	0.00615	0.01270	0.4043	0.709
Residual	39	0.29665	0.61265		
Total	54	0.48421	1.00000		

```

---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
permADMDRD_log2<-adonis2(distADMDRD_log~Subgroup,data=RATR_avg_demo_ADRDMD_log,
                           permutations=999, by='terms')
permADMDRD_log2[1:5]

```

	Df	SumOfSqs	R2	F	Pr(>F)
Subgroup	2	0.11955	0.24689	8.5235	0.002 **
Residual	52	0.36466	0.75311		
Total	54	0.48421	1.00000		

```

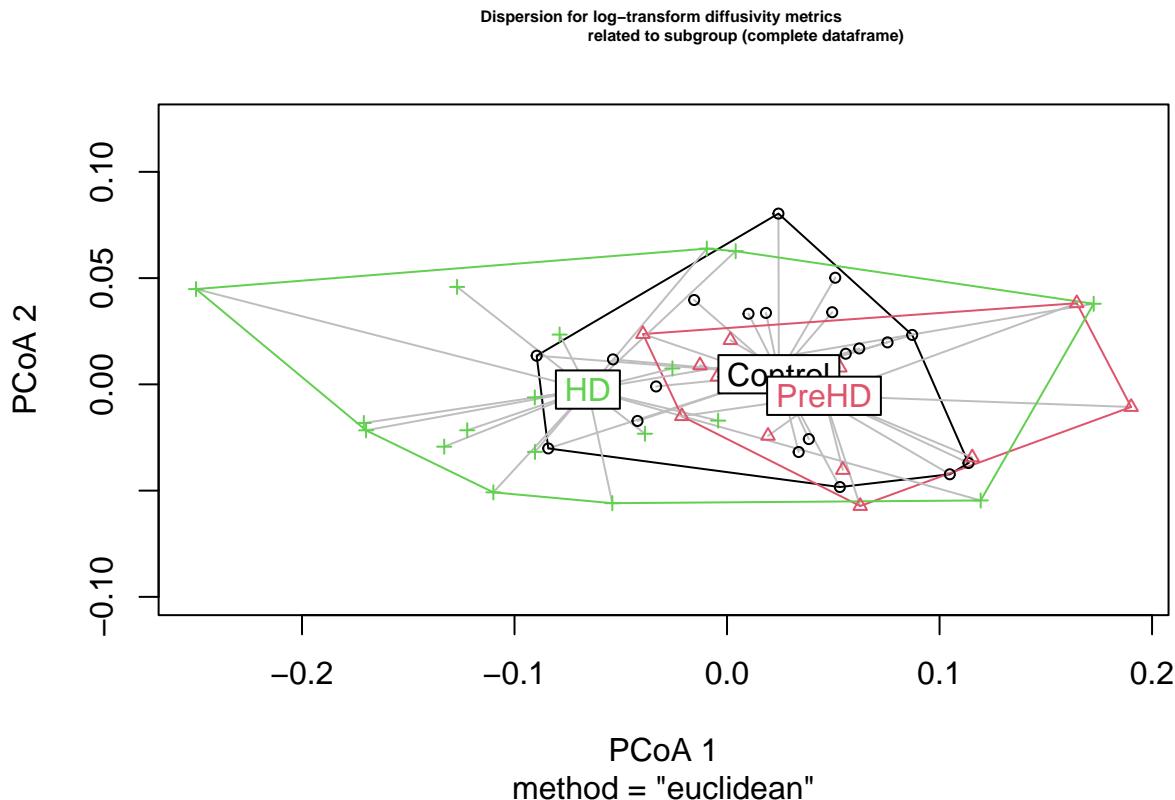
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Se determinan las dispersiones
dispersionADMDRD_log<-betadisper(d=distADMDRD_log, group=RATR_avg_demo_ADRDMD_log$Subgroup,
                                     type="centroid", bias.adjust=TRUE)

# Se generan los plots
par(mfrow = c(1, 1))

```

```
plot(dispersionADMDRD_log, main=list('Dispersion for log-transform diffusivity metrics  
related to subgroup (complete dataframe)', cex=0.5))
```



```
# Modelos ANOVA para las 3 métricas - Dataframe completo con transformación logarítmica
summary(aov(dispersionADMDRD_log$distances~RATR_avg_demo_ADRDMD_log$Subgroup))
      Df Sum Sq Mean Sq F value Pr(>F)
RATR_avg_demo_ADRDMD_log$Subgroup  2 0.01291 0.006456   3.121 0.0525 .
Residuals                          52 0.10756 0.002068
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

En el caso de la **transformación logarítmica de las 3 métricas de difusividad para el dataframe completo**, se objetiva que el factor subgrupo es el único término significativo en el PERMANOVA (como ocurría sin la transformación logarítmica). Como se visualiza en el plot, persiste una mayor dispersión para el subgrupo HD y unas dispersiones similares para los subgrupos preHD y control, pero, a nivel de ANOVA, la dispersión no es estadísticamente significativa, por tanto, las diferencias observadas en PERMANOVA son debidas a diferencias en el posicionamiento de los centroides y no a las diferencias en dispersión entre subgrupos. La transformación logarítmica obtiene unos valores F y p-valor de 3.121 y de 0.052 respectivamente comparado con los valores F y p-valor para el PERMANOVA sin transformación logarítmica de 4.537 y 0.015 respectivamente.

A continuación, se exponen los resultados para el **dataframe reducido** (con eliminación de outliers).

```
# Se genera la transformación logarítmica
RATR_avg_demo_ADRDMD_W0_log<-cbind(RATR_avg_demo_WoOut_ADRDMD[,1:7],
                                         log(RATR_avg_demo_WoOut_ADRDMD[,8:10]))
```

```
# Se calculan las distancias
```

```

distADMDRD_W0_log<-vegdist(RATR_avg_demo_ADRDMD_W0_log[,c('rhatr_AD','rhatr_MD','rhatr_RD')],  

                             method='euclidean')

# Se generan los modelos PERMANOVA  

set.seed(123)
permADMDRD_W0_log1<-adonis2(distADMDRD_W0_log~Subgroup*agef2*Gender,  

                               data=RATR_avg_demo_ADRDMD_W0_log, permutations=999, by='terms')
permADMDRD_W0_log1[1:5]
  Df SumOfSqs      R2      F Pr(>F)  

Subgroup    2  0.09228 0.22046 6.1865  0.007 **  

agef2       2  0.01855 0.04431 1.2435  0.322  

Gender      1  0.00298 0.00713 0.3999  0.556  

Subgroup:agef2  4  0.01175 0.02807 0.3939  0.858  

Subgroup:Gender  2  0.00084 0.00200 0.0560  0.988  

agef2:Gender   2  0.00527 0.01259 0.3533  0.786  

Subgroup:agef2:Gender 2  0.00351 0.00838 0.2353  0.861  

Residual     38  0.28340 0.67707  

Total        53  0.41857 1.00000
--  

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
permADMDRD_W0_log2<-adonis2(distADMDRD_W0_log~Subgroup,data=RATR_avg_demo_ADRDMD_W0_log,  

                                permutations=999, by='terms')
permADMDRD_W0_log2[1:5]
  Df SumOfSqs      R2      F Pr(>F)  

Subgroup    2  0.09228 0.22046 7.2115  0.001 ***  

Residual   51  0.32629 0.77954  

Total      53  0.41857 1.00000
--  

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

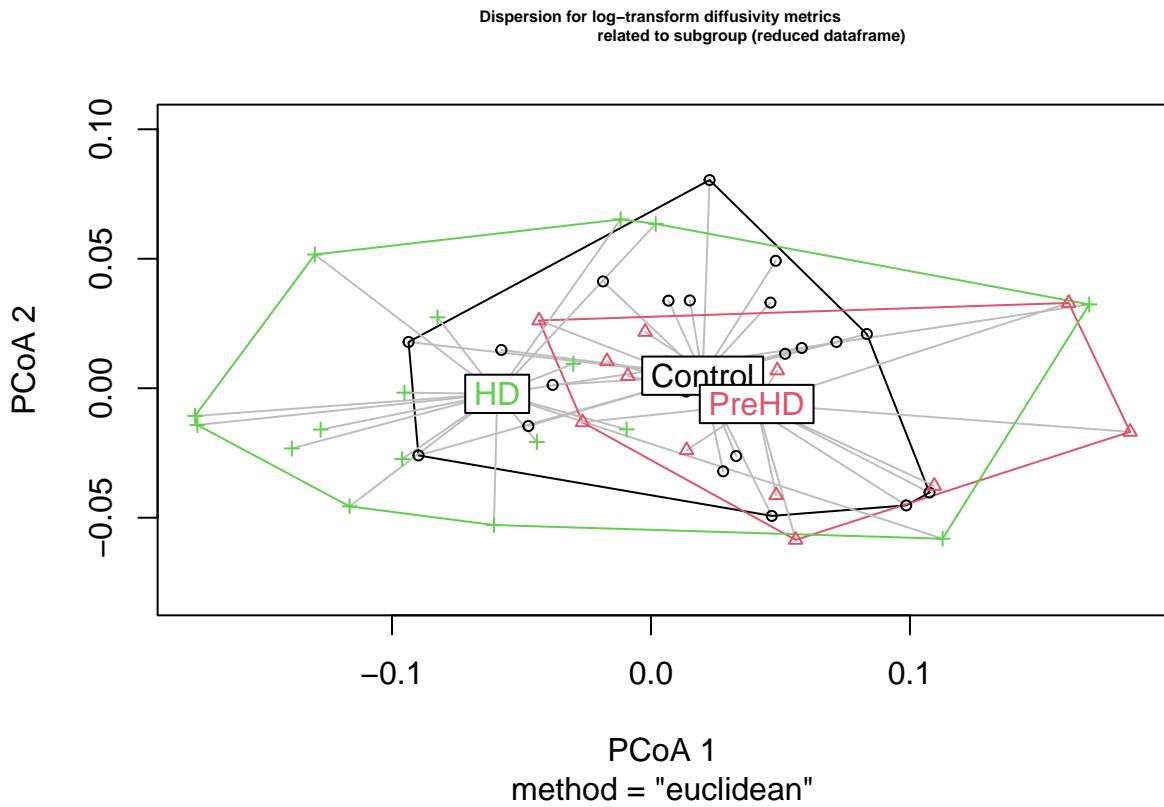
# Se determinan las dispersiones
dispersionADMDRD_W0_log<-betadisper(d=distADMDRD_W0_log, group=RATR_avg_demo_ADRDMD_W0_log$Subgroup,  

                                         type="centroid",bias.adjust=TRUE)

# Se generan los plots
par(mfrow = c(1, 1))
plot(dispersionADMDRD_W0_log, main=list('Dispersion for log-transform diffusivity metrics  

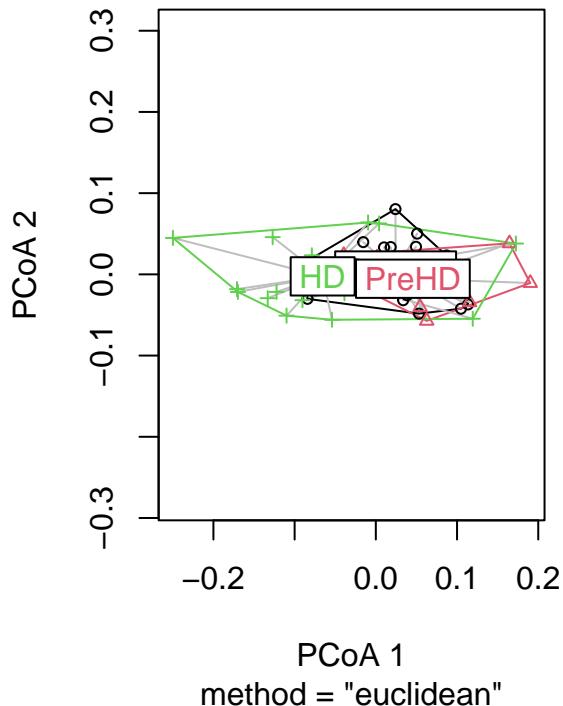
                                         related to subgroup (reduced dataframe)',cex=0.5))

```

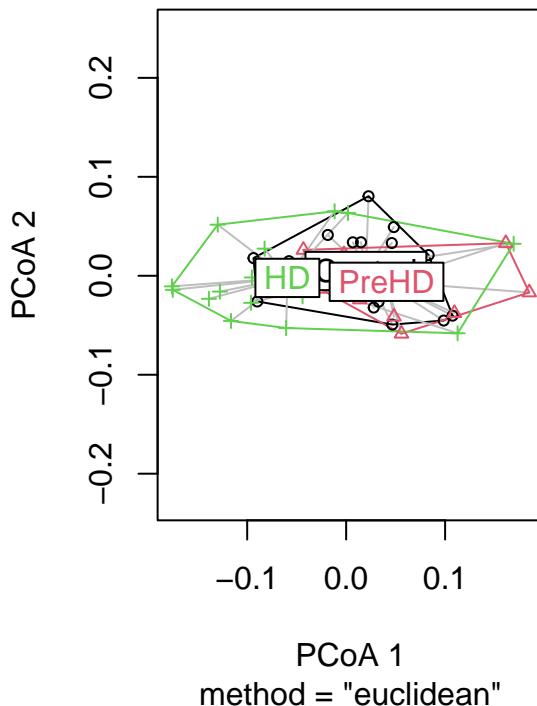


```
# Modelos ANOVA para las 3 métricas - Dataframe reducido con transformación logarítmica
summary(aov(dispersionADMDRD_W0_log$distances~RATR_avg_demo_ADRDMD_W0_log$Subgroup))
      Df  Sum Sq Mean Sq F value Pr(>F)
RATR_avg_demo_ADRDMD_W0_log$Subgroup  2 0.00908 0.004540   2.544 0.0885 .
Residuals                           51 0.09101 0.001785
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Dispersion for log-transform diffusivity related to subgroup (complete datafr:



Dispersion for log-transform diffusivity related to subgroup (reduced datafr:



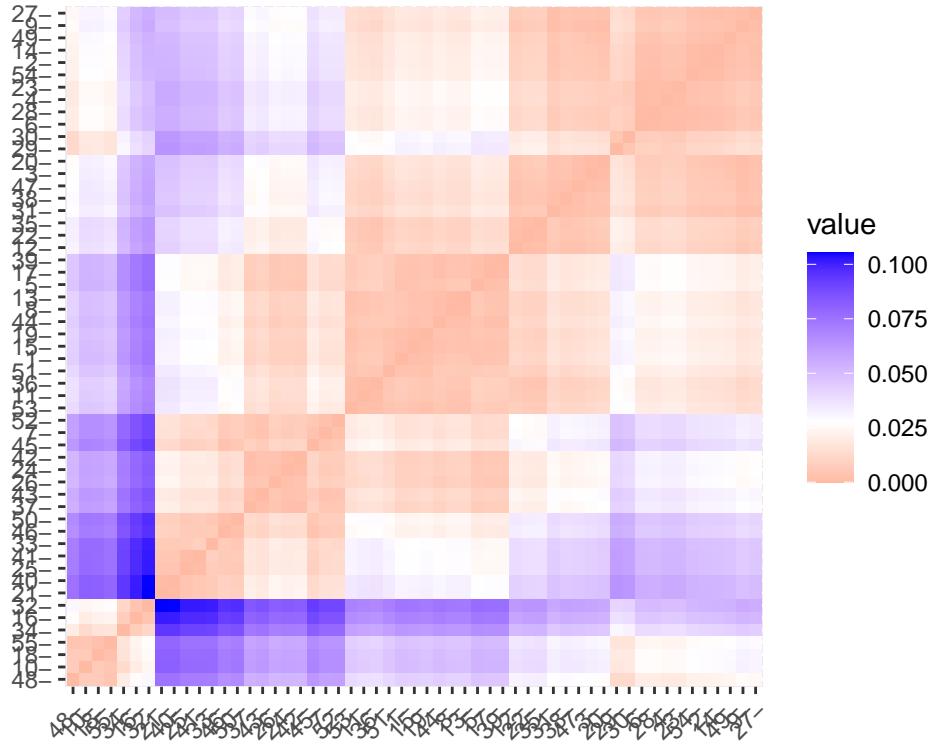
En el caso de la **transformación logarítmica de las 3 métricas de difusividad para el dataframe reducido**, se objetiva que el factor subgrupo es el único término significativo en el PERMANOVA (como ocurría sin la transformación logarítmica). Como se visualiza en el plot, persiste una mayor dispersión para el subgrupo HD y unas dispersiones similares para los subgrupos preHD y control, pero, a nivel de ANOVA, la dispersión no es estadísticamente significativa, por tanto, las diferencias observadas en PERMANOVA son debidas a diferencias en el posicionamiento de los centroides y no a las diferencias en dispersión entre subgrupos. La transformación logarítmica obtiene unos valores F y p-valor de 2.544 y de 0.088 respectivamente comparado con los valores F y p-valor para el PERMANOVA sin transformación logarítmica de 3.826 y 0.028 respectivamente.

### 3.2.2 LATR

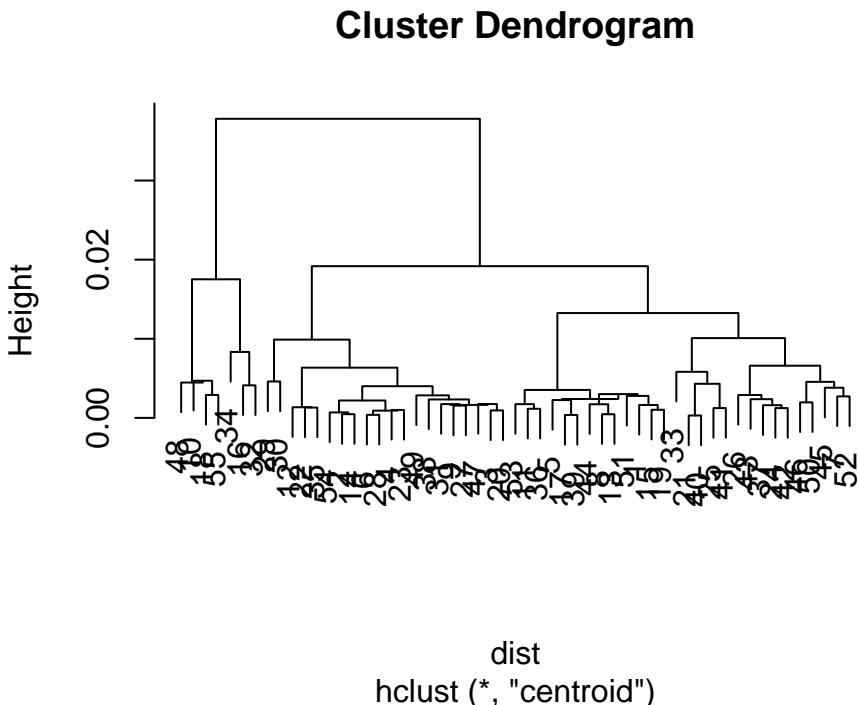
Las métricas de RM son variables numéricas continuas y no procedentes de especies, por lo que se decide utilizar la **distancia euclídea** en la generación de la matriz de distancias. Se ha procedido a **calcular la matriz de distancias** con la función `daisy` del paquete `cluster` para poder visualizar las similaridades/disimilaridades entre subgrupos mediante MDS y tSNE.

```
# Matriz de distancias para las 5 métricas - Sin eliminación de outliers
dist<-daisy(LATR_avg_demo[,-c(1:7)],metric='euclidean')
distmx<-as.matrix(dist)

fviz_dist(dist)
```



```
clust<-hclust(dist, method = "centroid", members = NULL)
plot(clust)
```

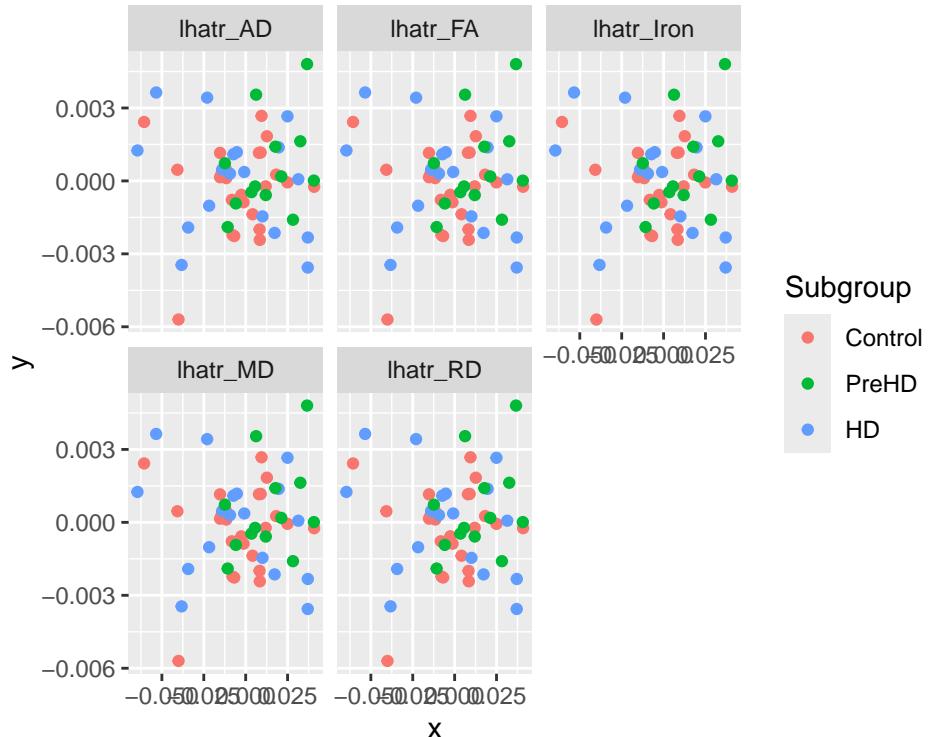


```
mds_dist<-cmdscale(distmx, eig=TRUE, k=2)
df_mds_dist<-data.frame(LATR_avg_demo, x=mds_dist$points[,1],
```

```

y=mds_dist$points[,2])
df_mds_dist2<-df_mds_dist %>%
  dplyr::select(x, y, Subgroup, Gender, agef2, lhatr_Iron, lhatr_AD, lhatr_RD,
                lhatr_MD, lhatr_FA) %>%
  tidyr::gather(key = "variable", value = "values", -x, -y, -Subgroup, -Gender, -agef2)
ggplot(df_mds_dist2,aes(x, y)) + geom_jitter(aes(col=Subgroup)) + facet_wrap(~variable)

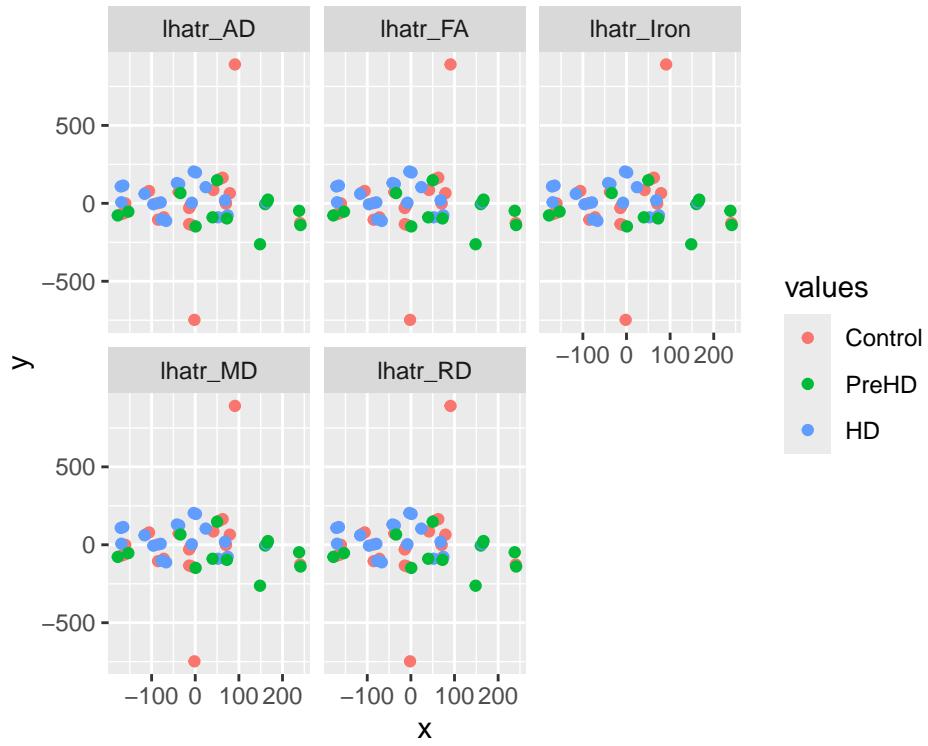
```



```

my_Rtsne<-Rtsne(X= distmx,perplexity= floor((nrow(LATR_avg_demo) - 1) / 3),dims = 2)
# Maximum perplexity
my_Rtsne$perplexity
[1] 18
fit_tsne<-Rtsne(dist,is_distance=TRUE,check_duplicates=FALSE,pca=FALSE,dims=2)
Error in .check_tsne_params(nrow(X), dims = dims, perplexity = perplexity, : perplexity is too large for t
set.seed(123)
fit_tsne<-Rtsne(dist,perplexity = 1, is_distance = TRUE, check_duplicates = FALSE,
                 pca = FALSE, dims = 2, initial_dims = 50)
df_tsne_vars <- as_tibble(cbind(LATR_avg_demo,
                                    x = fit_tsne$Y[, 1], y = fit_tsne$Y[, 2]))
df_tsne <- df_tsne_vars %>%
  dplyr::select(x, y, Subgroup, Gender, agef2, lhatr_Iron, lhatr_AD, lhatr_RD,
                lhatr_MD, lhatr_FA) %>%
  tidyr::gather(key = "variable", value = "values", -x, -y, -Subgroup, -Gender, -agef2)
ggplot(df_tsne,aes(x, y, col=values)) + geom_point(aes(col=Subgroup)) + facet_wrap(~variable)

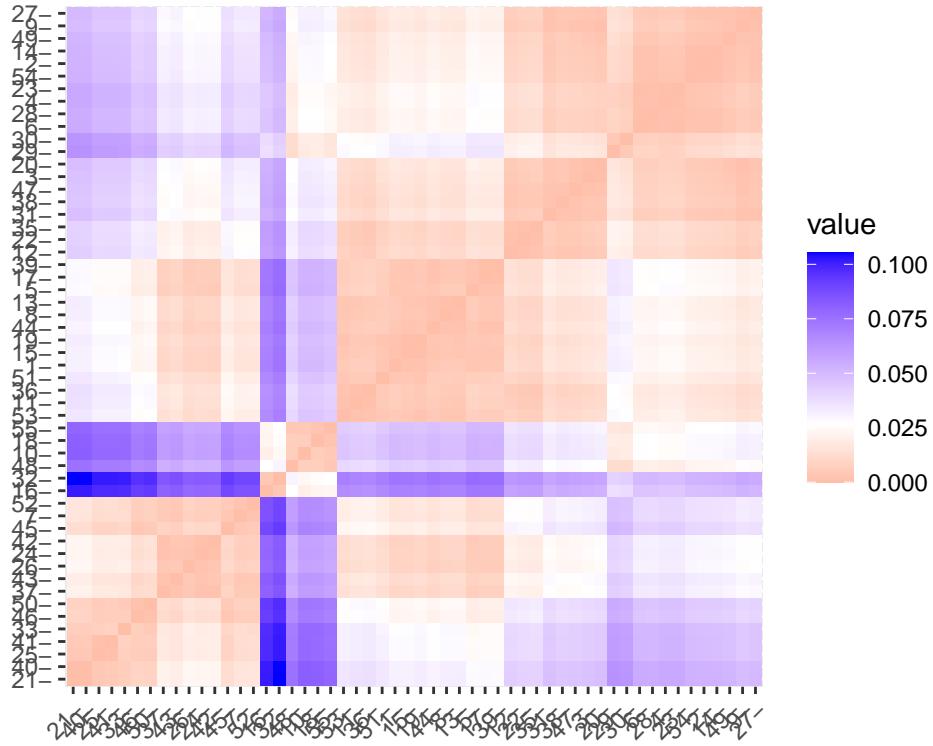
```



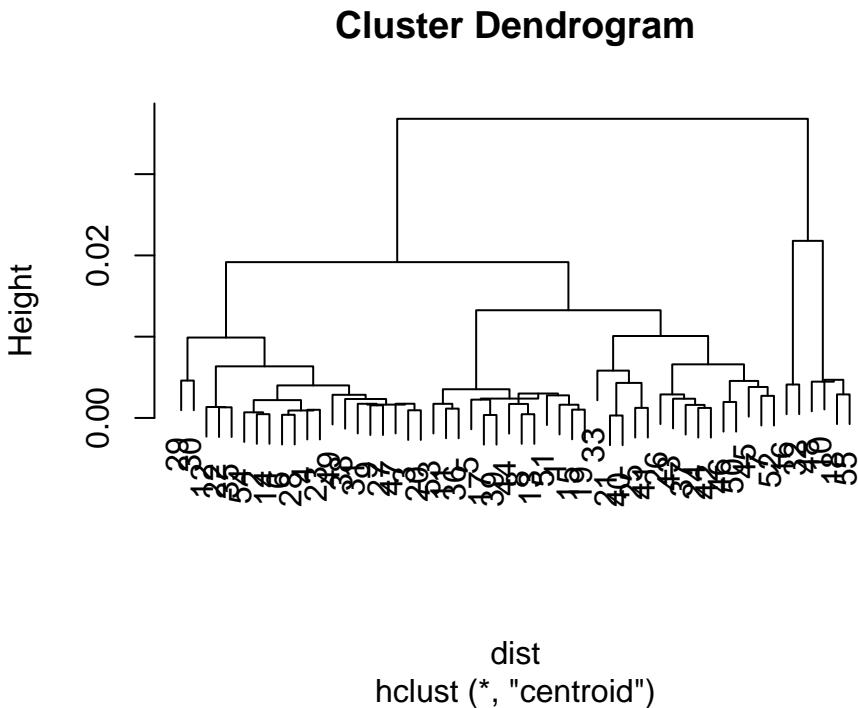
Según los gráficos obtenidos con el dataframe con todos los sujetos (sin eliminación de outliers) y con las matriz de distancias calculada con las 5 métricas de RM, los subgrupos no parecen separarse con dicha matriz de distancias.

```
# Matriz de distancias para las 5 métricas - Con eliminación de outliers
dist<-daisy(LATR_avg_WoOut[,-c(1:7)],metric='euclidean')
distmx<-as.matrix(dist)

fviz_dist(dist)
```



```
clust<-hclust(dist, method = "centroid", members = NULL)
plot(clust)
```

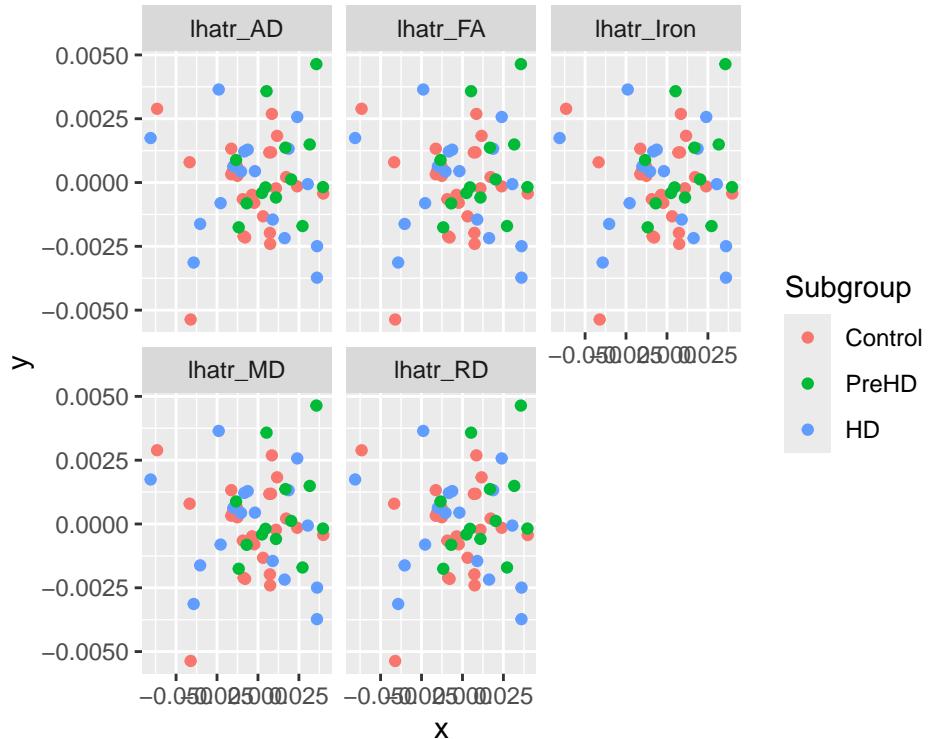


```
mds_dist<-cmdscale(distmx, eig=TRUE, k=2)
df_mds_dist<-data.frame(LATR_avg_demo_WoOut, x=mds_dist$points[,1],
```

```

y=mds_dist$points[,2])
df_mds_dist2<-df_mds_dist %>%
  dplyr::select(x, y, Subgroup, Gender, agef2, lhatr_Iron, lhatr_AD, lhatr_RD,
                lhatr_MD, lhatr_FA) %>%
  tidyverse::gather(key = "variable", value = "values", -x, -y, -Subgroup, -Gender, -agef2)
ggplot(df_mds_dist2, aes(x, y)) + geom_jitter(aes(col=Subgroup)) + facet_wrap(~variable)

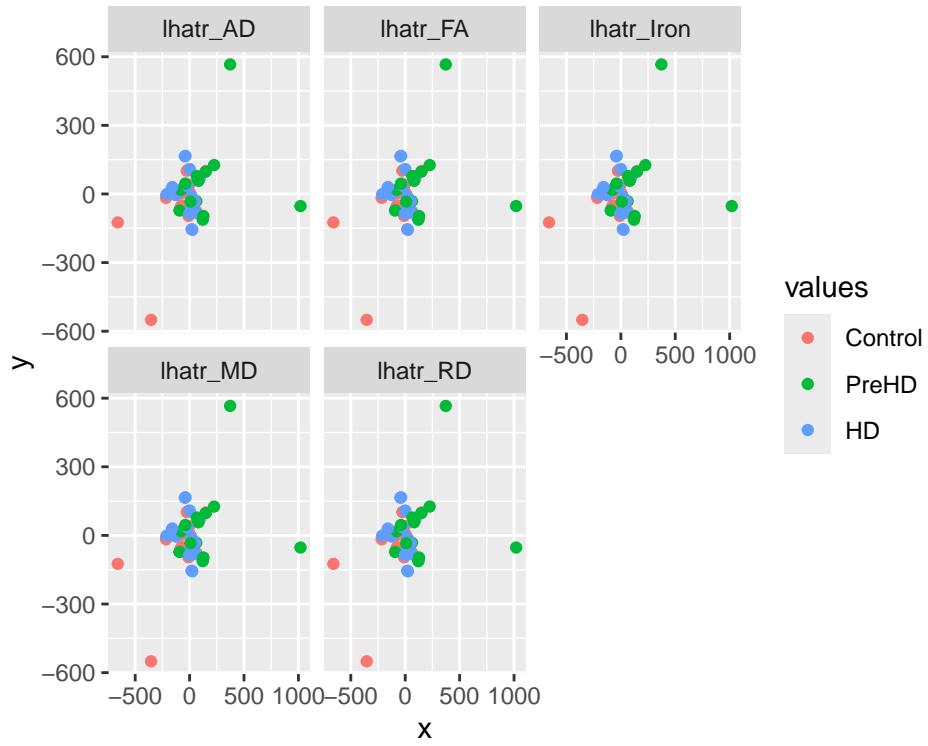
```



```

my_Rtsne<-Rtsne(X= distmx,perplexity= floor((nrow(LATR_avg_demo_WoOut) - 1) / 3),dims = 2)
# Maximum perplexity
my_Rtsne$perplexity
[1] 17
set.seed(123)
fit_tsne<-Rtsne(dist,perplexity = 1, is_distance = TRUE,
                  check_duplicates = FALSE, pca = FALSE, dims = 2, initial_dims = 50)
df_tsne_vars <- as_tibble(cbind(LATR_avg_demo_WoOut,
                                    x = fit_tsne$Y[, 1], y = fit_tsne$Y[, 2]))
df_tsne <- df_tsne_vars %>%
  dplyr::select(x, y, Subgroup, Gender, agef2, lhatr_Iron, lhatr_AD, lhatr_RD,
                lhatr_MD, lhatr_FA) %>%
  tidyverse::gather(key = "variable", value = "values", -x, -y, -Subgroup, -Gender, -agef2)
ggplot(df_tsne, aes(x, y, col=values)) + geom_point(aes(col=Subgroup)) + facet_wrap(~variable)

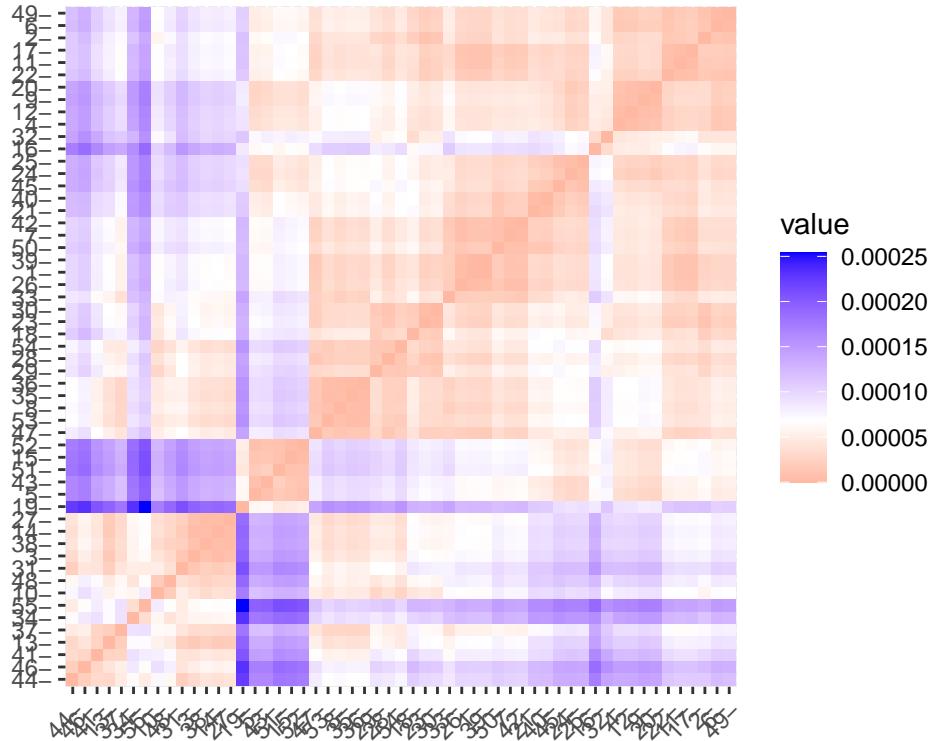
```



Según los gráficos obtenidos con el dataframe con eliminación de outliers y con las matriz de distancias calculada con las 5 métricas de RM, los subgrupos tampoco parecen separarse con dicha matriz de distancias.

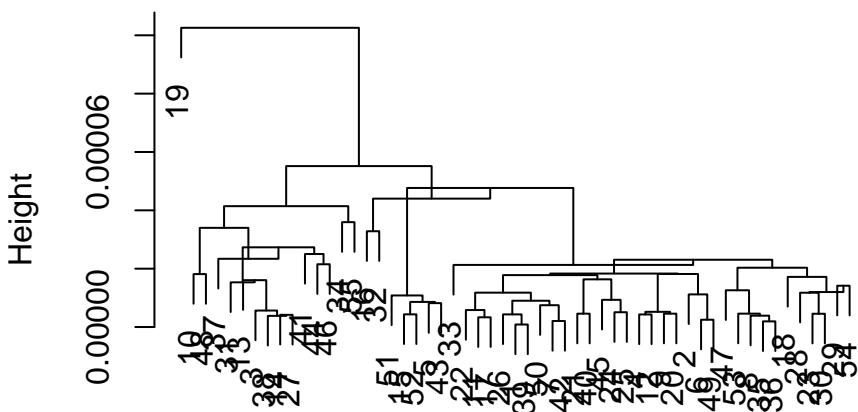
```
# Matriz de distancias para las 3 métricas de difusividad
dist<-daisy(LATR_avg_demo_ADRDMD[,-c(1:7)],metric='euclidean')
distmx<-as.matrix(dist)

fviz_dist(dist)
```



```
clust<-hclust(dist, method = "centroid", members = NULL)
plot(clust)
```

### Cluster Dendrogram



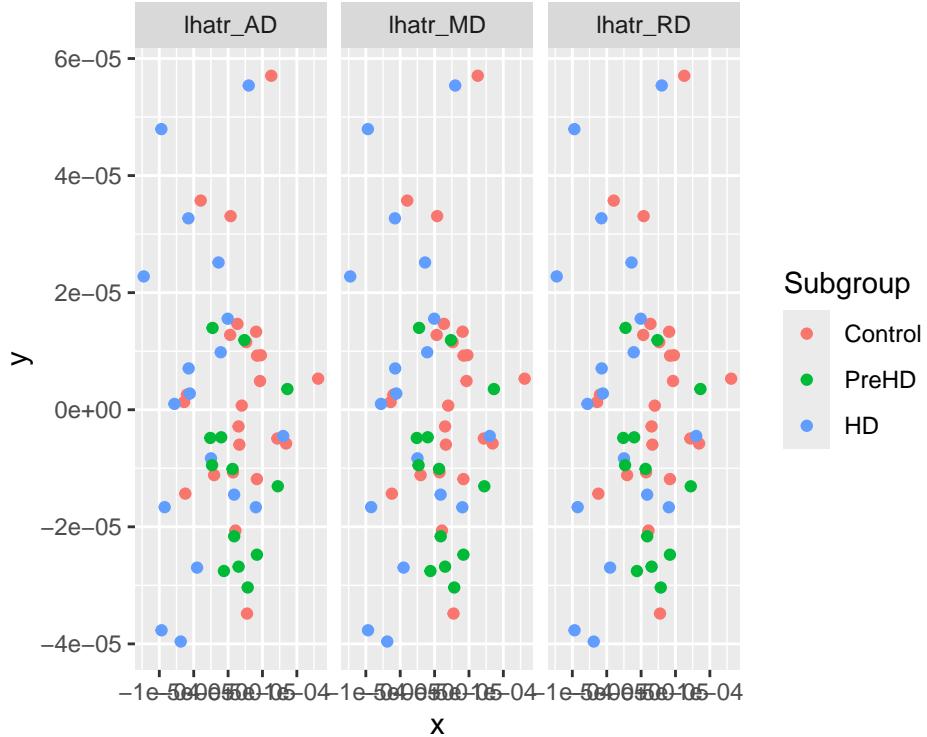
```
dist
hclust (*, "centroid")
```

```
mds_dist<-cmdscale(distmx, eig=TRUE, k=2)
df_mds_dist<-data.frame(LATR_avg_demo_ADRDMD, x=mds_dist$points[,1],
```

```

y=mds_dist$points[,2])
df_mds_dist2<-df_mds_dist %>%
  dplyr::select(x, y, Subgroup, Gender, agef2, lhatr_AD, lhatr_RD, lhatr_MD) %>%
  tidyverse::gather(key = "variable", value = "values", -x, -y, -Subgroup, -Gender, -agef2)
ggplot(df_mds_dist2, aes(x,y)) + geom_jitter(aes(col=Subgroup)) + facet_wrap(~variable)

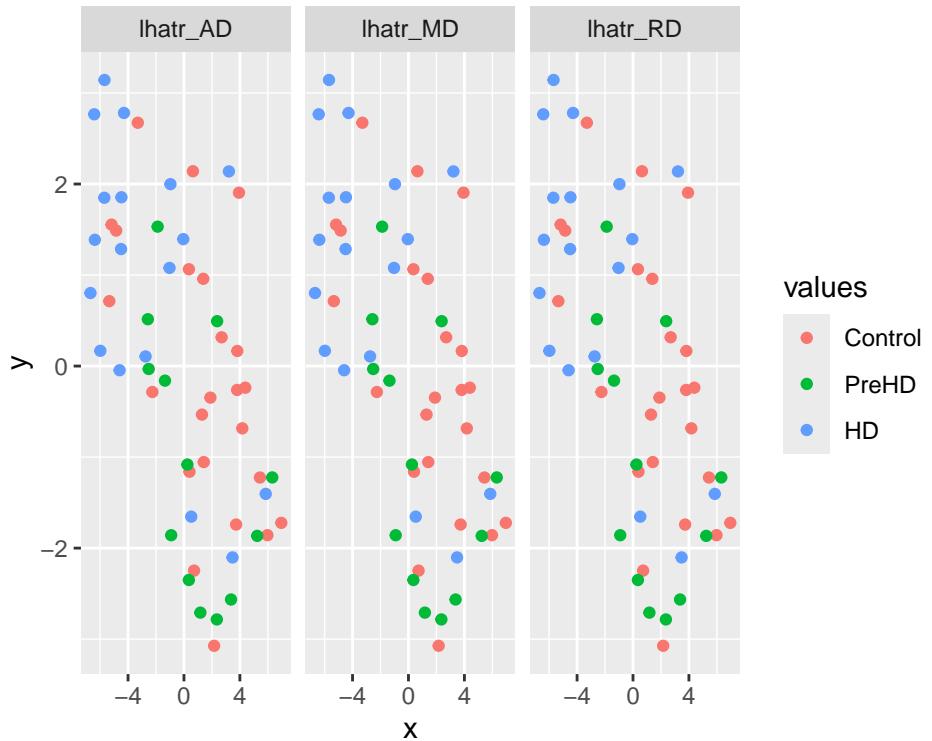
```



```

my_Rtsne<-Rtsne(X= distmx,perplexity= floor((nrow(LATR_avg_demo_ADRDMD) - 1) / 3),dims = 2)
# Maximum perplexity
my_Rtsne$perplexity
[1] 18
set.seed(123)
fit_tsne<-Rtsne(dist,perplexity = my_Rtsne$perplexity, is_distance = TRUE, check_duplicates = FALSE,
                  pca = FALSE, dims = 2, initial_dims = 50)
df_tsne_vars <- as_tibble(cbind(LATR_avg_demo_ADRDMD,
                                    x = fit_tsne$Y[, 1], y = fit_tsne$Y[, 2]))
df_tsne <- df_tsne_vars %>%
  dplyr::select(x, y, Subgroup, Gender, agef2, lhatr_AD, lhatr_RD, lhatr_MD) %>%
  tidyverse::gather(key = "variable", value = "values", -x, -y, -Subgroup, -Gender, -agef2)
ggplot(df_tsne, aes(x, y, col=values)) + geom_point(aes(col=Subgroup)) + facet_wrap(~variable)

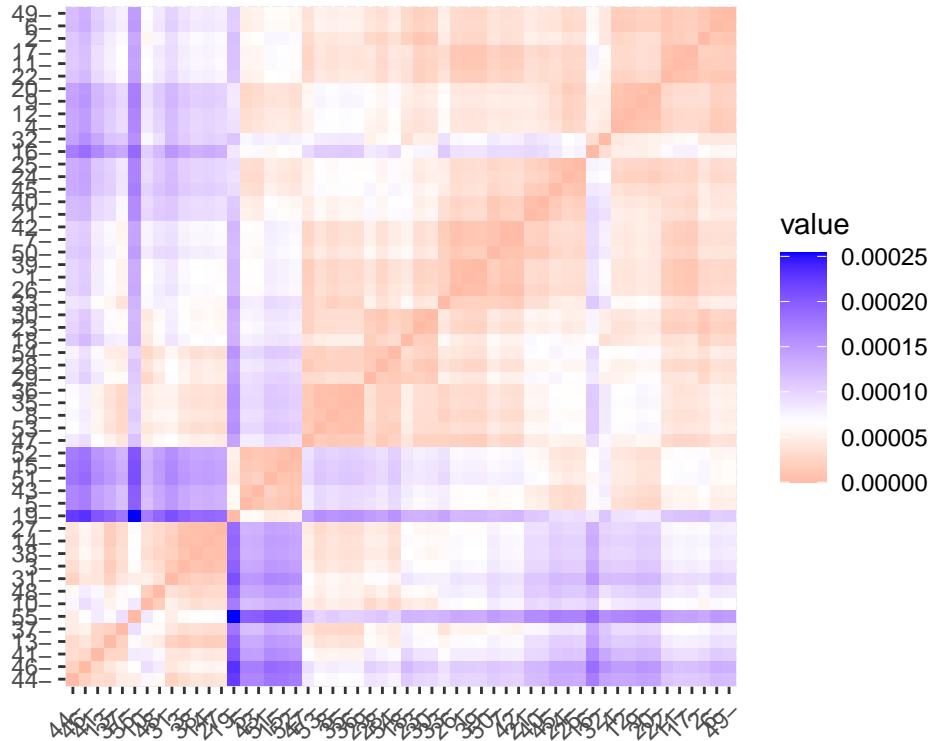
```



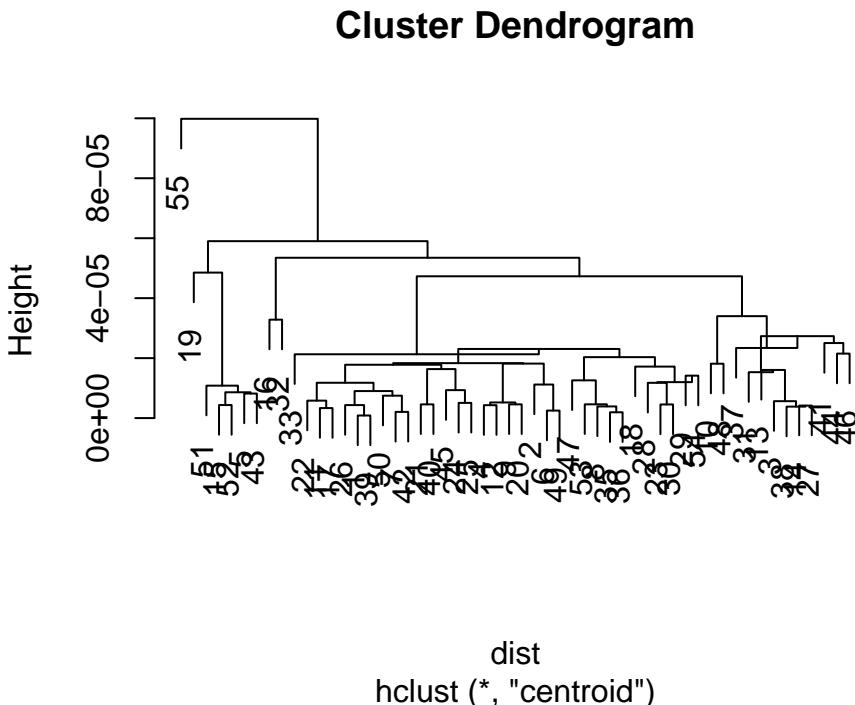
Según los gráficos obtenidos con el dataframe con todos los sujetos (sin eliminación de outliers) y con la matriz de distancias obtenida con las 3 métricas de difusividad, los subgrupos HD y preHD vs controles parecen separarse con dicha matriz de distancias, pero existe gran solapamiento entre los tres subgrupos.

```
# Matriz de distancias para las 3 métricas de difusividad
dist<-daisy(LATR_avg_WoOut_ADRDMD[,-c(1:7)],metric='euclidean')
distmx<-as.matrix(dist)

fviz_dist(dist)
```



```
clust<-hclust(dist, method = "centroid", members = NULL)
plot(clust)
```

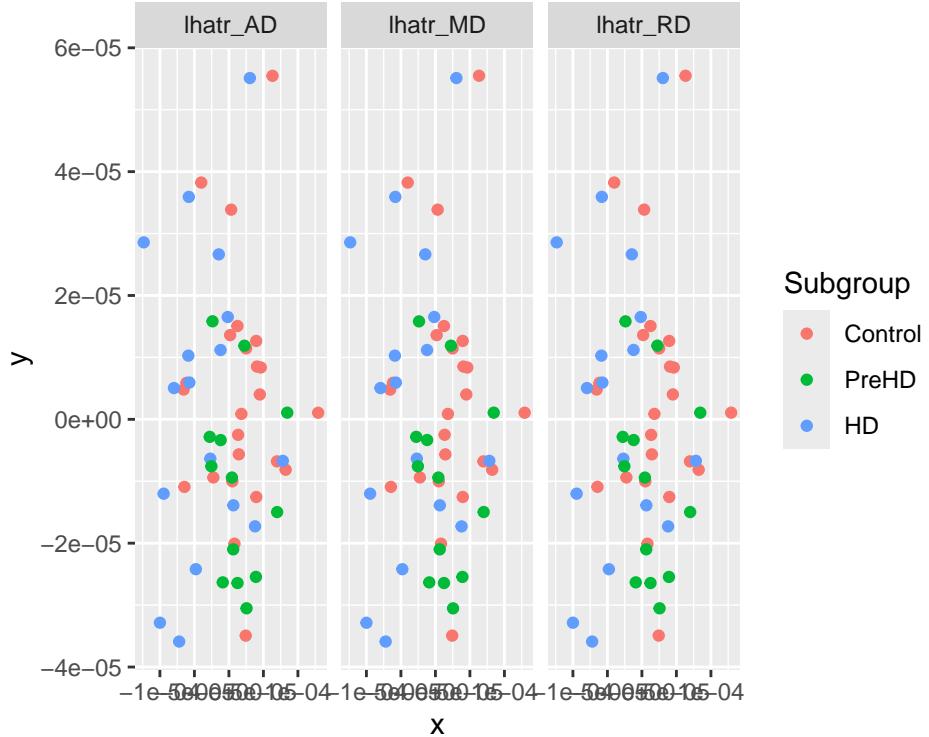


```
mds_dist<-cmdscale(distmx, eig=TRUE, k=2)
df_mds_dist<-data.frame(LATR_avg_demo_WoOut_ADRDMD, x=mds_dist$points[,1],
```

```

y=mds_dist$points[,2])
df_mds_dist2<-df_mds_dist %>%
  dplyr::select(x, y, Subgroup, Gender, agef2, lhatr_AD, lhatr_RD, lhatr_MD) %>%
  tidyverse::gather(key = "variable", value = "values", -x, -y, -Subgroup, -Gender, -agef2)
ggplot(df_mds_dist2, aes(x,y)) + geom_jitter(aes(col=Subgroup)) + facet_wrap(~variable)

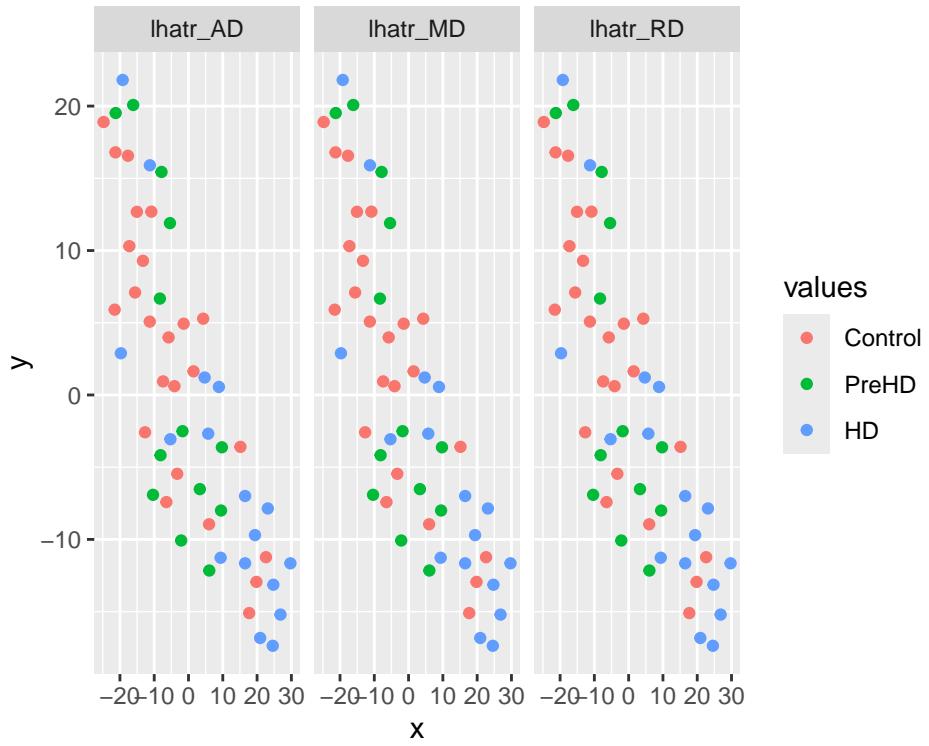
```



```

my_Rtsne<-Rtsne(X= distmx,perplexity=floor((nrow(LATR_avg_demo_WoOut_ADRDMD)-1)/3),dims=2)
# Maximum perplexity
my_Rtsne$perplexity
[1] 17
set.seed(123)
fit_tsne<-Rtsne(dist,perplexity = my_Rtsne$perplexity, is_distance = TRUE, check_duplicates = FALSE,
                  pca = FALSE, dims = 2, initial_dims = 50)
df_tsne_vars <- as_tibble(cbind(LATR_avg_demo_WoOut_ADRDMD,
                                    x = fit_tsne$Y[, 1], y = fit_tsne$Y[, 2]))
df_tsne <- df_tsne_vars %>%
  dplyr::select(x, y, Subgroup, Gender, agef2, lhatr_AD, lhatr_RD, lhatr_MD) %>%
  tidyverse::gather(key = "variable", value = "values", -x, -y, -Subgroup, -Gender, -agef2)
ggplot(df_tsne, aes(x, y, col=values)) + geom_point(aes(col=Subgroup)) + facet_wrap(~variable)

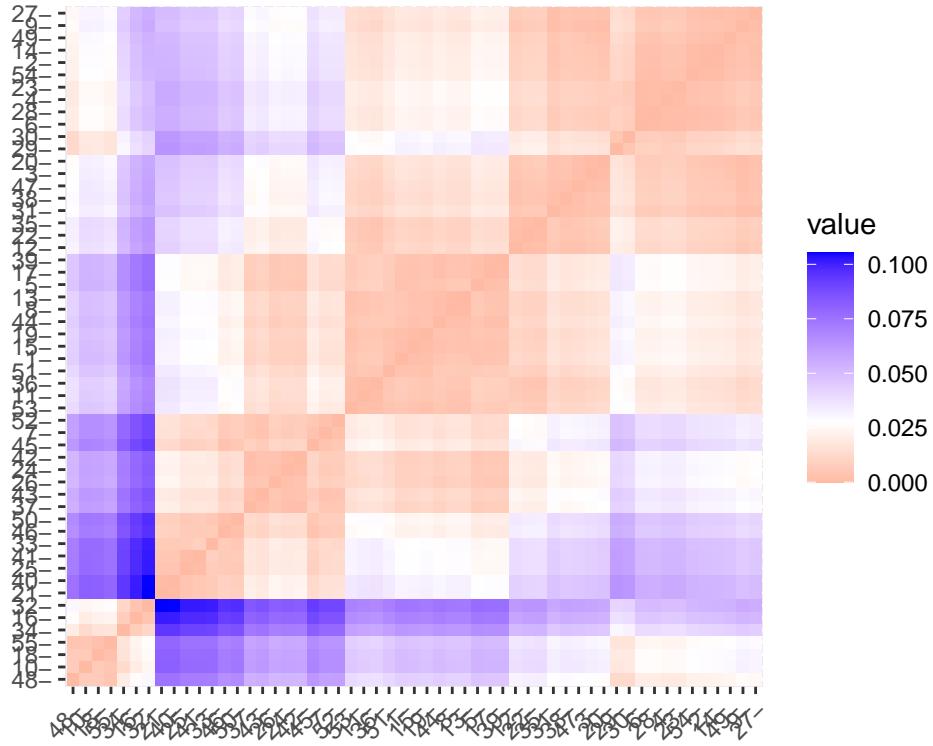
```



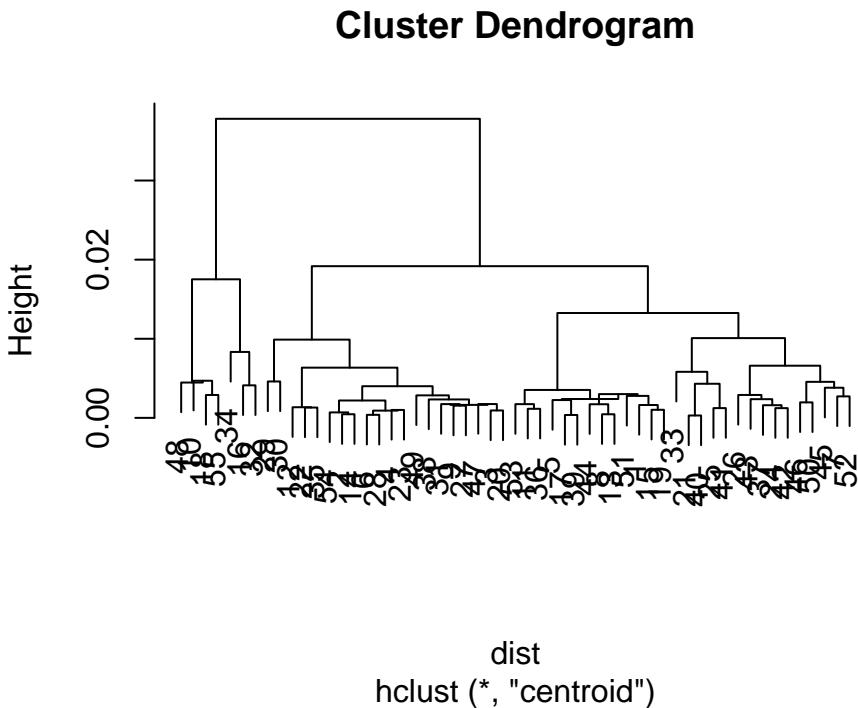
Según los gráficos obtenidos con el dataframe con eliminación de outliers y con la matriz de distancias obtenida con las 3 métricas de difusividad, los subgrupos HD y preHD vs controles también parecen separarse con dicha matriz de distancias, pero existe gran solapamiento entre los tres subgrupos.

```
# Matriz de distancias para las métricas de iron y FA
dist<-daisy(LATR_avg_demo_IronFA[,-c(1:7)],metric='euclidean')
distmx<-as.matrix(dist)

fviz_dist(dist)
```



```
clust<-hclust(dist, method = "centroid", members = NULL)
plot(clust)
```

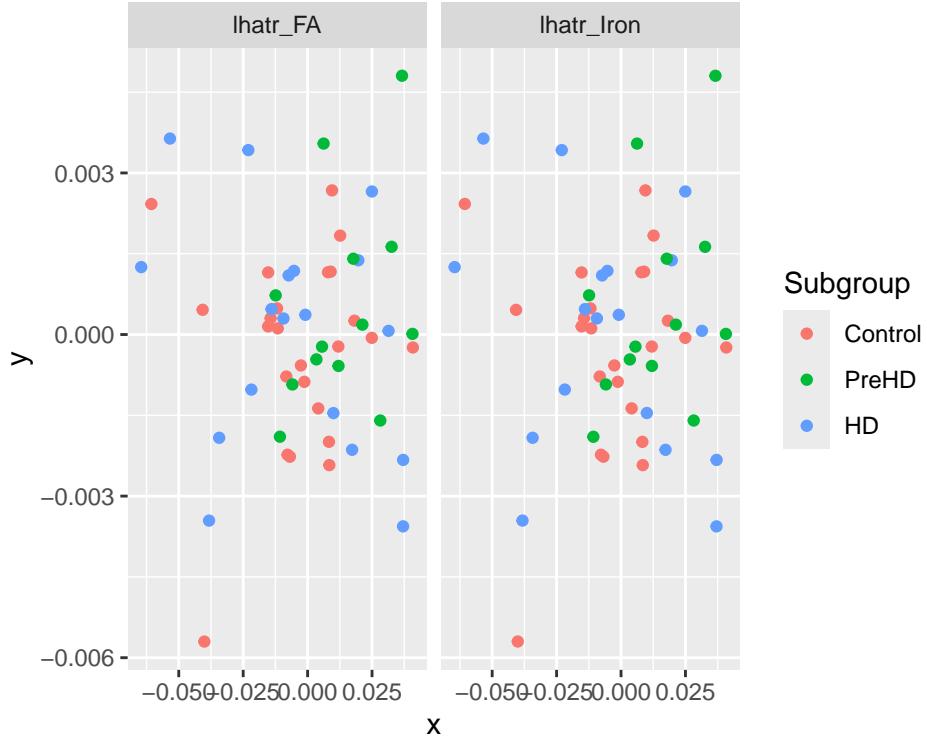


```
mds_dist<-cmdscale(distmx, eig=TRUE, k=2)
df_mds_dist<-data.frame(LATR_avg_demo_IronFA, x=mds_dist$points[,1],
```

```

y=mds_dist$points[,2])
df_mds_dist2<-df_mds_dist %>%
  dplyr::select(x, y, Subgroup, Gender, agef2, lhatr_Iron, lhatr_FA) %>%
  tidyverse::gather(key = "variable", value = "values", -x, -y, -Subgroup, -Gender, -agef2)
ggplot(df_mds_dist2, aes(x,y)) + geom_jitter(aes(col=Subgroup)) + facet_wrap(~variable)

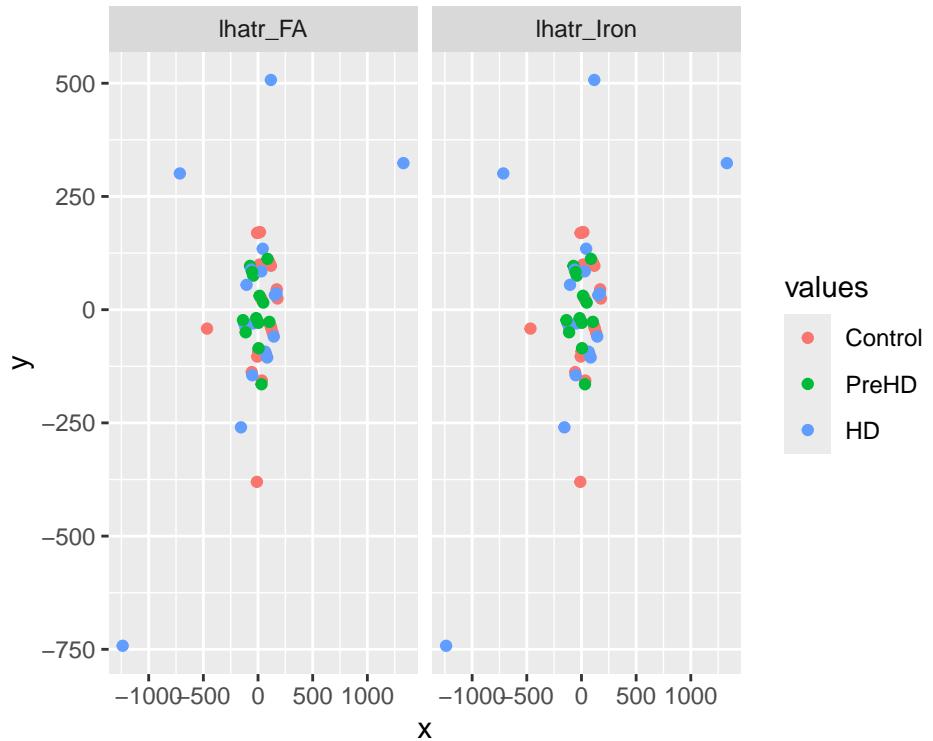
```



```

my_Rtsne<-Rtsne(X= distmx,perplexity= floor((nrow(LATR_avg_demo_IronFA) - 1) / 3),dims = 2)
# Maximum perplexity
my_Rtsne$perplexity
[1] 18
set.seed(123)
fit_tsne<-Rtsne(dist,perplexity = 1, is_distance = TRUE,
                  check_duplicates = FALSE, pca = FALSE, dims = 2, initial_dims = 50)
df_tsne_vars<-as_tibble(cbind(LATR_avg_demo_IronFA,x=fit_tsne$Y[,1],y=fit_tsne$Y[,2]))
df_tsne <- df_tsne_vars %>%
  dplyr::select(x, y, Subgroup, Gender, agef2, lhatr_Iron, lhatr_FA) %>%
  tidyverse::gather(key = "variable", value = "values", -x, -y, -Subgroup, -Gender, -agef2)
ggplot(df_tsne, aes(x, y, col=values)) + geom_point(aes(col=Subgroup)) + facet_wrap(~variable)

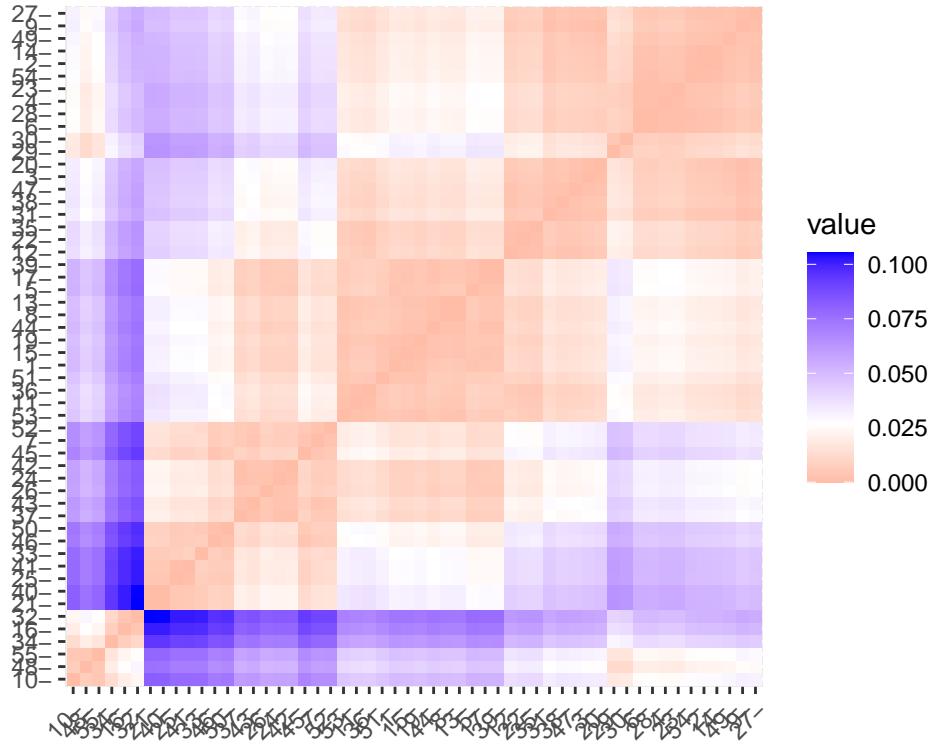
```



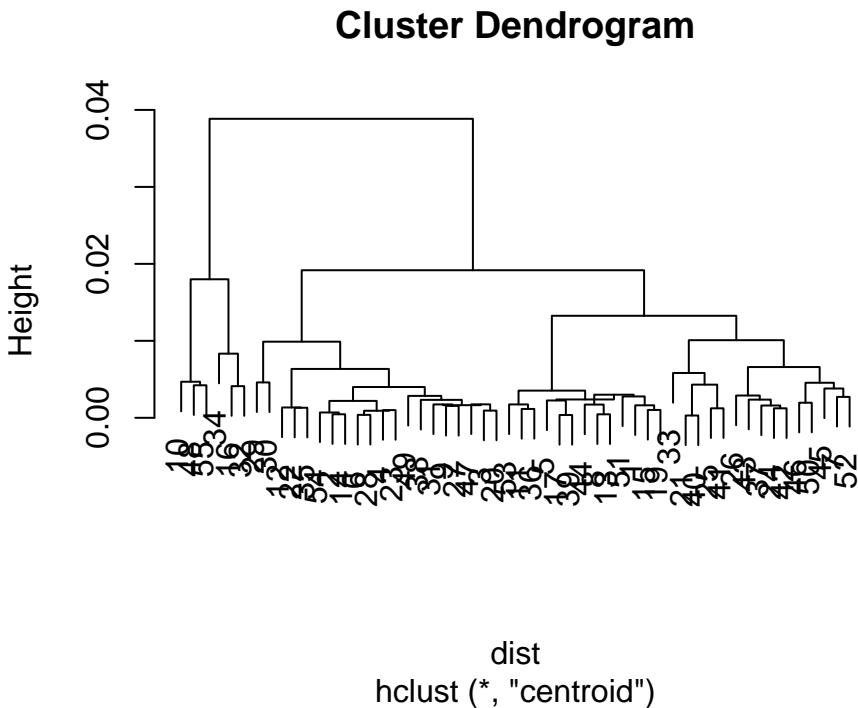
Según los gráficos obtenidos con el dataframe con todos los sujetos (sin eliminación de outliers) y la matriz de distancias obtenida de las métricas de iron y FA, los subgrupos no parecen separarse con las matrices de distancias.

```
# Matriz de distancias para las métricas de iron y FA
dist<-daisy(LATR_avg_WoOut_IronFA[,-c(1:7)],metric='euclidean')
distmx<-as.matrix(dist)

fviz_dist(dist)
```



```
clust<-hclust(dist, method = "centroid", members = NULL)
plot(clust)
```

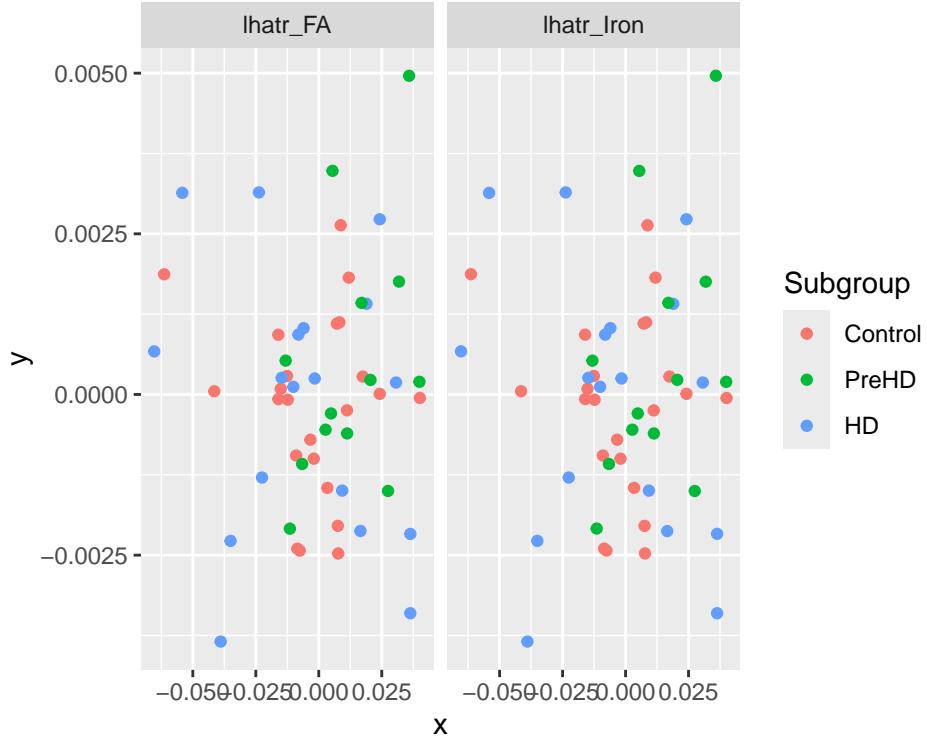


```
mds_dist<-cmdscale(distmx, eig=TRUE, k=2)
df_mds_dist<-data.frame(LATR_avg_demo_WoOut_IronFA, x=mds_dist$points[,1],
```

```

y=mds_dist$points[,2])
df_mds_dist2<-df_mds_dist %>%
  dplyr::select(x, y, Subgroup, Gender, agef2, lhatr_Iron, lhatr_FA) %>%
  tidyverse::gather(key = "variable", value = "values", -x, -y, -Subgroup, -Gender, -agef2)
ggplot(df_mds_dist2, aes(x,y)) + geom_jitter(aes(col=Subgroup)) + facet_wrap(~variable)

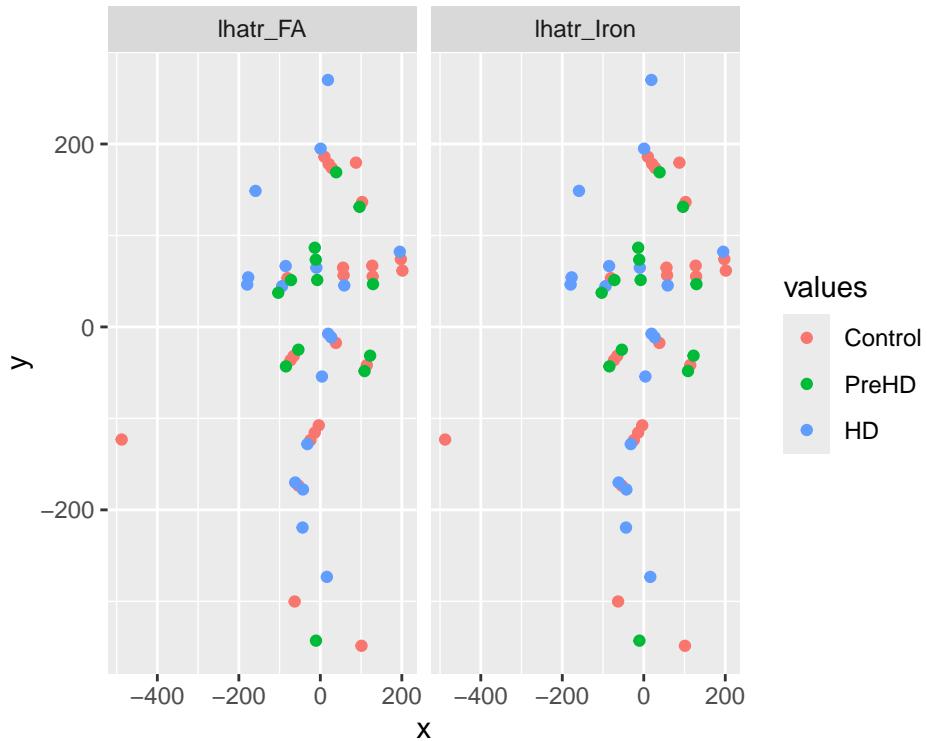
```



```

my_Rtsne<-Rtsne(X= distmx,perplexity= floor((nrow(LATR_avg_demo_WoOut_IronFA)-1)/3),dims = 2)
# Maximum perplexity
my_Rtsne$perplexity
[1] 17
set.seed(123)
fit_tsne<-Rtsne(dist,perplexity = 1, is_distance = TRUE,
                  check_duplicates = FALSE, pca = FALSE, dims = 2, initial_dims = 50)
df_tsne_vars<-as_tibble(cbind(LATR_avg_demo_WoOut_IronFA,x=fit_tsne$Y[,1],
                                y=fit_tsne$Y[,2]))
df_tsne <- df_tsne_vars %>%
  dplyr::select(x, y, Subgroup, Gender, agef2, lhatr_Iron, lhatr_FA) %>%
  tidyverse::gather(key = "variable", value = "values", -x, -y, -Subgroup, -Gender, -agef2)
ggplot(df_tsne, aes(x, y, col=values)) + geom_point(aes(col=Subgroup)) + facet_wrap(~variable)

```



Según los gráficos obtenidos con el dataframe con eliminación de outliers y la matriz de distancias obtenida de las métricas de iron y FA, los subgrupos tampoco parecen separarse con las matrices de distancias.

También se calculan las matrices de distancia con la función `vegdist` del paquete `vegan` para la combinación de métricas Iron-FA (por poder correlacionarse con el subgrupo preHD), AD-MD-RD (por poder correlacionarse con el subgrupo HD), y, para las 5 métricas en su conjunto. Dichas matrices se utilizarán para generar los modelos PERMANOVA. Las variables categóricas no se introducen en la matriz de distancias porque se utilizan como variables independientes en los modelos PERMANOVA.

Como se objetiva a continuación, se han calculado las matrices para ambos supuestos, con y sin eliminación de outliers. Teóricamente, PERMANOVA, al ser una técnica no paramétrica, podría ejecutarse con todos los sujetos ya que no necesita normalidad multivariante ni homocedasticidad de la matriz de covarianzas, pero, al basarse en permutaciones, requiere verificar una betadispersión homogénea entre grupos. Así que se generarán modelos PERMANOVA con y sin eliminación de outliers.

```
library(vegan)
# Matriz de distancias para las 5 métricas
distDifIron<-vegdist(LATR_avg_demo[,-c(1:7)],method='euclidean')
distDifIron_WoOut<-vegdist(LATR_avg_demo_WoOut[,-c(1:7)],method='euclidean')

# Matriz de distancias para las 3 métricas de difusividad (AD, RD, MD)
distADMDRD<-vegdist(LATR_avg_demo_ADRDMD[,c('lhatr_AD','lhatr_RD','lhatr_MD')], 
                      method='euclidean')
distADMDRD_WoOut<-vegdist(LATR_avg_demo_WoOut_ADRDMD[,c('lhatr_AD','lhatr_RD',
                           'lhatr_MD')], 
                           method='euclidean')

# Matriz de distancias para las métricas de Iron y FA
distIronFA<-vegdist(LATR_avg_demo_IronFA[,c('lhatr_Iron','lhatr_FA')], 
                     method='euclidean')
distIronFA_WoOut<-vegdist(LATR_avg_demo_WoOut_IronFA[,c('lhatr_Iron','lhatr_FA')], 
                           method='euclidean')
```

A continuación, se generan los diferentes modelos PERMANOVA (remarcar que no se puede añadir el efecto aleatorio). Primero, los modelos se ejecutan con las 3 variables independientes en interacción. Segundo, se eliminan las interacciones no significativas para simplificar los modelos.

```
# Establecer seed para que el resultado de las permutaciones sea siempre el mismo
set.seed(123)

library(vegan)
# Modelos para las 5 métricas
permDifIron1<-adonis2(distDifIron~Subgroup*agef2*Gender, data=LATR_avg_demo,
                        permutations=999, by='terms');permDifIron1[1:5]
  Df SumOfSqs      R2      F Pr(>F)
Subgroup        2 0.003141 0.09261 2.7221  0.078 .
agef2           2 0.001240 0.03655 1.0742  0.370
Gender          1 0.000067 0.00198 0.1165  0.752
Subgroup:agef2  4 0.001401 0.04130 0.6070  0.665
Subgroup:Gender 2 0.001842 0.05432 1.5965  0.236
agef2:Gender    2 0.002454 0.07235 2.1267  0.121
Subgroup:agef2:Gender 2 0.001271 0.03747 1.1013  0.332
Residual        39 0.022500 0.66342
Total           54 0.033915 1.00000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
permDifIron2<-adonis2(distDifIron~Subgroup+agef2+Gender, data=LATR_avg_demo,
                        permutations=999, by='terms');permDifIron2[1:5]
  Df SumOfSqs      R2      F Pr(>F)
Subgroup        2 0.003141 0.09261 2.6114  0.066 .
agef2           2 0.001240 0.03655 1.0306  0.374
Gender          1 0.000067 0.00198 0.1118  0.750
Residual        49 0.029468 0.86886
Total           54 0.033915 1.00000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
permDifIron3<-adonis2(distDifIron~Subgroup, data=LATR_avg_demo, permutations=999,
                        by='terms');permDifIron3[1:5]
  Df SumOfSqs      R2      F Pr(>F)
Subgroup        2 0.003141 0.09261 2.6536  0.075 .
Residual        52 0.030774 0.90739
Total           54 0.033915 1.00000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

permDifIron1_WoOut<-adonis2(distDifIron_WoOut~Subgroup*agef2*Gender,
                               data=LATR_avg_demo_WoOut, permutations=999,
                               by='terms');permDifIron1_WoOut[1:5]
  Df SumOfSqs      R2      F Pr(>F)
Subgroup        2 0.0026909 0.08681 2.4863  0.106
agef2           2 0.0010294 0.03321 0.9511  0.384
Gender          1 0.0000059 0.00019 0.0110  0.937
Subgroup:agef2  4 0.0018051 0.05824 0.8340  0.534
Subgroup:Gender 2 0.0012669 0.04087 1.1706  0.316
agef2:Gender    2 0.0025433 0.08205 2.3500  0.097 .
Subgroup:agef2:Gender 2 0.0010922 0.03523 1.0091  0.369
Residual        38 0.0205630 0.66340
Total           53 0.0309967 1.00000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```

permDifIron2_WoOut<-adonis2(distDifIron_WoOut~Subgroup+agef2+Gender,
                               data=LATR_avg_demo_WoOut, permutations=999,
                               by='terms');permDifIron2_WoOut[1:5]
    Df   SumOfSqs      R2      F Pr(>F)
Subgroup  2  0.0026909  0.08681  2.3682  0.092 .
agef2     2  0.0010294  0.03321  0.9059  0.413
Gender     1  0.0000059  0.00019  0.0105  0.945
Residual  48  0.0272705  0.87979
Total     53  0.0309967  1.00000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
permDifIron3_WoOut<-adonis2(distDifIron_WoOut~Subgroup, data=LATR_avg_demo_WoOut,
                               permutations=999, by='terms');permDifIron3_WoOut[1:5]
    Df   SumOfSqs      R2      F Pr(>F)
Subgroup  2  0.0026909  0.08681  2.4241  0.092 .
Residual  51  0.0283058  0.91319
Total     53  0.0309967  1.00000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Modelos para las 3 métricas de difusividad (AD, RD, MD)
permADMDRD1<-adonis2(distADMDRD~Subgroup*agef2*Gender, data=LATR_avg_demo_ADRDMD,
                        permutations=999, by='terms');permADMDRD1[1:5]
    Df   SumOfSqs      R2      F Pr(>F)
Subgroup  2  3.8439e-08 0.21621  6.4845  0.005 **
agef2     2  5.5130e-09 0.03101  0.9300  0.430
Gender     1  3.1250e-09 0.01758  1.0544  0.329
Subgroup:agef2  4  4.8850e-09 0.02748  0.4120  0.857
Subgroup:Gender  2  1.7800e-09 0.01001  0.3003  0.827
agef2:Gender    2  5.9140e-09 0.03326  0.9976  0.399
Subgroup:agef2:Gender  2  2.5370e-09 0.01427  0.4280  0.731
Residual          39  1.1559e-07 0.65018
Total             54  1.7779e-07 1.00000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
permADMDRD2<-adonis2(distADMDRD~Subgroup+agef2+Gender, data=LATR_avg_demo_ADRDMD,
                        permutations=999, by='terms');permADMDRD2[1:5]
    Df   SumOfSqs      R2      F Pr(>F)
Subgroup  2  3.8439e-08 0.21621  7.2049  0.002 **
agef2     2  5.5130e-09 0.03101  1.0333  0.356
Gender     1  3.1250e-09 0.01758  1.1716  0.279
Residual  49  1.3071e-07 0.73520
Total     54  1.7779e-07 1.00000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
permADMDRD3<-adonis2(distADMDRD~Subgroup, data=LATR_avg_demo_ADRDMD, permutations=999,
                        by='terms');permADMDRD3[1:5]
    Df   SumOfSqs      R2      F Pr(>F)
Subgroup  2  3.8439e-08 0.21621  7.1721  0.003 **
Residual  52  1.3935e-07 0.78379
Total     54  1.7779e-07 1.00000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

permADMDRD1_WoOut<-adonis2(distADMDRD_WoOut~Subgroup*agef2*Gender,
                               data=LATR_avg_demo_WoOut_ADRDMD, permutations=999,

```

```

                by='terms');permADMDRD1_WoOut[1:5]
Df   SumOfSqs      R2      F Pr(>F)
Subgroup 2 3.2485e-08 0.19586 5.4509  0.004 ** 
agef2    2 3.9600e-09 0.02388 0.6645  0.586
Gender    1 3.7200e-09 0.02243 1.2484  0.277
Subgroup:agef2 4 3.8840e-09 0.02342 0.3259  0.925
Subgroup:Gender 2 1.0870e-09 0.00655 0.1824  0.906
agef2:Gender  2 5.5100e-09 0.03322 0.9246  0.411
Subgroup:agef2:Gender 2 1.9810e-09 0.01195 0.3324  0.789
Residual     38 1.1323e-07 0.68270
Total        53 1.6586e-07 1.00000
---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
permADMDRD2_WoOut<-adonis2(distADMDRD_WoOut~Subgroup+agef2+Gender,
                               data=LATR_avg_demo_WoOut_ADRDMD, permutations=999,
                               by='terms');permADMDRD2_WoOut[1:5]
Df   SumOfSqs      R2      F Pr(>F)
Subgroup 2 3.2485e-08 0.19586 6.2027  0.002 ** 
agef2    2 3.9600e-09 0.02388 0.7562  0.465
Gender    1 3.7200e-09 0.02243 1.4206  0.205
Residual 48 1.2569e-07 0.75783
Total     53 1.6586e-07 1.00000
---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
permADMDRD3_WoOut<-adonis2(distADMDRD_WoOut~Subgroup, data=LATR_avg_demo_WoOut_ADRDMD,
                               permutations=999, by='terms');permADMDRD3_WoOut[1:5]
Df   SumOfSqs      R2      F Pr(>F)
Subgroup 2 3.2485e-08 0.19586 6.2109  0.002 ** 
Residual 51 1.3337e-07 0.80414
Total     53 1.6586e-07 1.00000
---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Modelos para las métricas de iron y FA
permIronFA1<-adonis2(distIronFA~Subgroup*agef2*Gender, data=LATR_avg_demo_IronFA,
                       permutations=999, by='terms');permIronFA1[1:5]
Df   SumOfSqs      R2      F Pr(>F)
Subgroup 2 0.003141 0.09261 2.7220  0.086 .
agef2    2 0.001240 0.03655 1.0742  0.328
Gender    1 0.000067 0.00198 0.1165  0.736
Subgroup:agef2 4 0.001401 0.04130 0.6070  0.648
Subgroup:Gender 2 0.001842 0.05432 1.5965  0.203
agef2:Gender  2 0.002454 0.07235 2.1267  0.145
Subgroup:agef2:Gender 2 0.001271 0.03747 1.1013  0.344
Residual     39 0.022500 0.66342
Total        54 0.033915 1.00000
---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
permIronFA2<-adonis2(distIronFA~Subgroup+agef2+Gender, data=LATR_avg_demo_IronFA,
                       permutations=999, by='terms');permIronFA2[1:5]
Df   SumOfSqs      R2      F Pr(>F)
Subgroup 2 0.003141 0.09261 2.6113  0.076 .
agef2    2 0.001240 0.03655 1.0306  0.360
Gender    1 0.000067 0.00198 0.1118  0.767
Residual 49 0.029467 0.86886
Total     54 0.033915 1.00000

```

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
permIronFA3<-adonis2(distIronFA~Subgroup, data=LATR_avg_demo_IronFA,
                        permutations=999, by='terms');permIronFA3[1:5]
      Df SumOfSqs      R2      F Pr(>F)
Subgroup  2 0.003141 0.09261 2.6536  0.092 .
Residual 52 0.030774 0.90739
Total     54 0.033915 1.00000

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

permIronFA1_WoOut<-adonis2(distIronFA_WoOut~Subgroup*agef2*Gender,
                             data=LATR_avg_demo_WoOut_IronFA, permutations=999,
                             by='terms');permIronFA1_WoOut[1:5]
      Df SumOfSqs      R2      F Pr(>F)
Subgroup  2 0.002909 0.09021 2.5682  0.086 .
agef2     2 0.001099 0.03407 0.9698  0.388
Gender    1 0.000202 0.00628 0.3575  0.555
Subgroup:agef2 4 0.001183 0.03670 0.5224  0.729
Subgroup:Gender 2 0.001506 0.04670 1.3295  0.281
agef2:Gender   2 0.002007 0.06223 1.7716  0.199
Subgroup:agef2:Gender 2 0.001819 0.05642 1.6062  0.232
Residual      38 0.021522 0.66740
Total        53 0.032247 1.00000

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
permIronFA2_WoOut<-adonis2(distIronFA_WoOut~Subgroup+agef2+Gender,
                             data=LATR_avg_demo_WoOut_IronFA, permutations=999,
                             by='terms');permIronFA2_WoOut[1:5]
      Df SumOfSqs      R2      F Pr(>F)
Subgroup  2 0.002909 0.09021 2.4902  0.099 .
agef2     2 0.001099 0.03407 0.9404  0.403
Gender    1 0.000202 0.00628 0.3466  0.565
Residual 48 0.028037 0.86944
Total     53 0.032247 1.00000

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
permIronFA3_WoOut<-adonis2(distIronFA_WoOut~Subgroup, data=LATR_avg_demo_WoOut_IronFA,
                             permutations=999, by='terms');permIronFA3_WoOut[1:5]
      Df SumOfSqs      R2      F Pr(>F)
Subgroup  2 0.002909 0.09021 2.5285  0.082 .
Residual 51 0.029338 0.90979
Total     53 0.032247 1.00000

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Los **modelos PERMANOVA generados con las 5 métricas**, ya sea con el dataframe completo o con el dataframe reducido eliminando los outliers, no evidencian factores significativos, pero, a diferencia del ATR derecho, el subgrupo es casi significativo (con p valores menores 0.10) en los tres modelos (desde el modelo más complejo con los factores en interacción hasta el modelo más simple con subgrupo como único factor).

Los **modelos PERMANOVA generados con las 3 métricas de difusividad**, ya sea con el dataframe completo o con el dataframe reducido eliminando los outliers, evidencian el subgrupo como único factor significativo.

Los **modelos PERMANOVA generados con las métricas de iron y FA**, ya sea con el dataframe completo o con el dataframe reducido eliminando los outliers, no evidencian factores significativos, pero, a diferencia del ATR derecho, el subgrupo es casi significativo (con p valores menores 0.10) en los tres modelos (desde el modelo más

complejo con los factores en interacción hasta el modelo más simple con subgrupo como único factor).

Por tanto, PERMANOVA ha sido menos sensible que MANOVA para detectar significación estadística en la evaluación de las 5 métricas conjuntamente, pero, presenta una casi significación estadística tanto para las 5 métricas conjuntamente como para las métricas iron-FA.

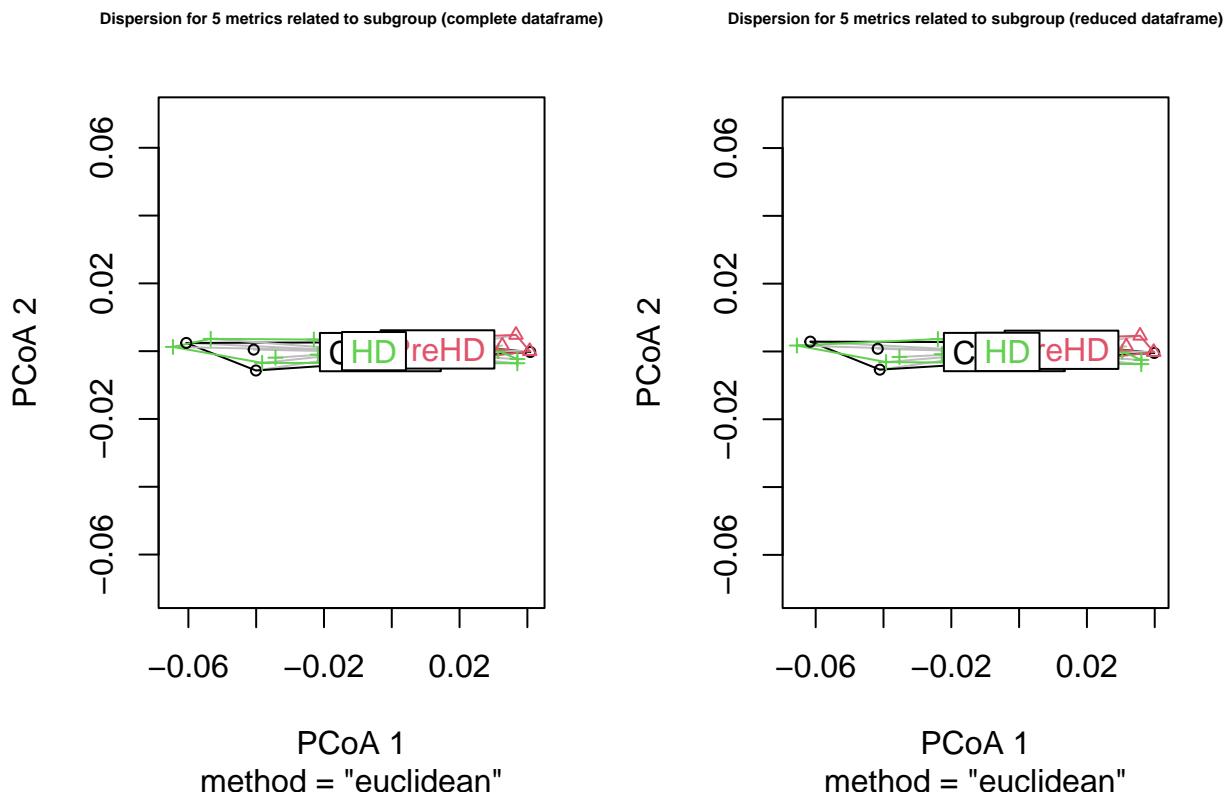
Se calculan las **dispersiones** para cada uno de los modelos que se han generado en el paso anterior y tan sólo se han calculado para el factor subgrupo dado que es el único que ha dado significativo.

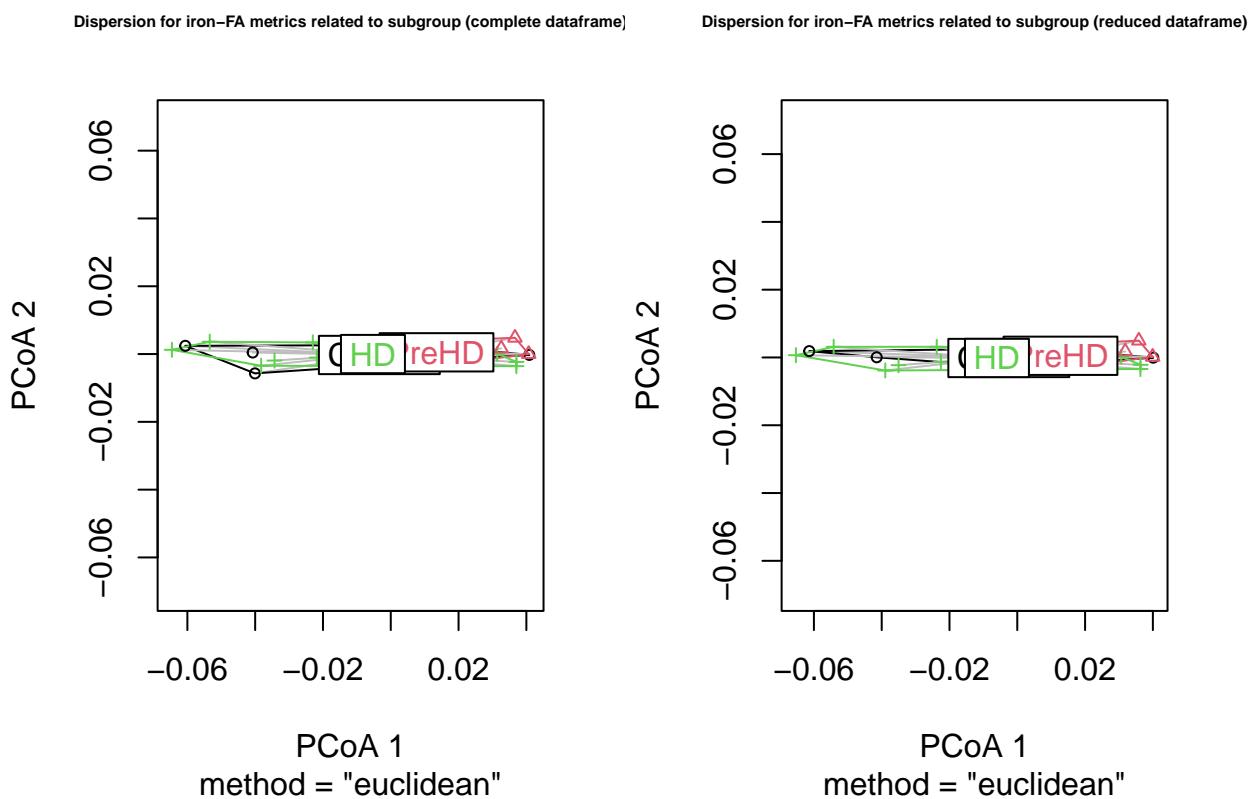
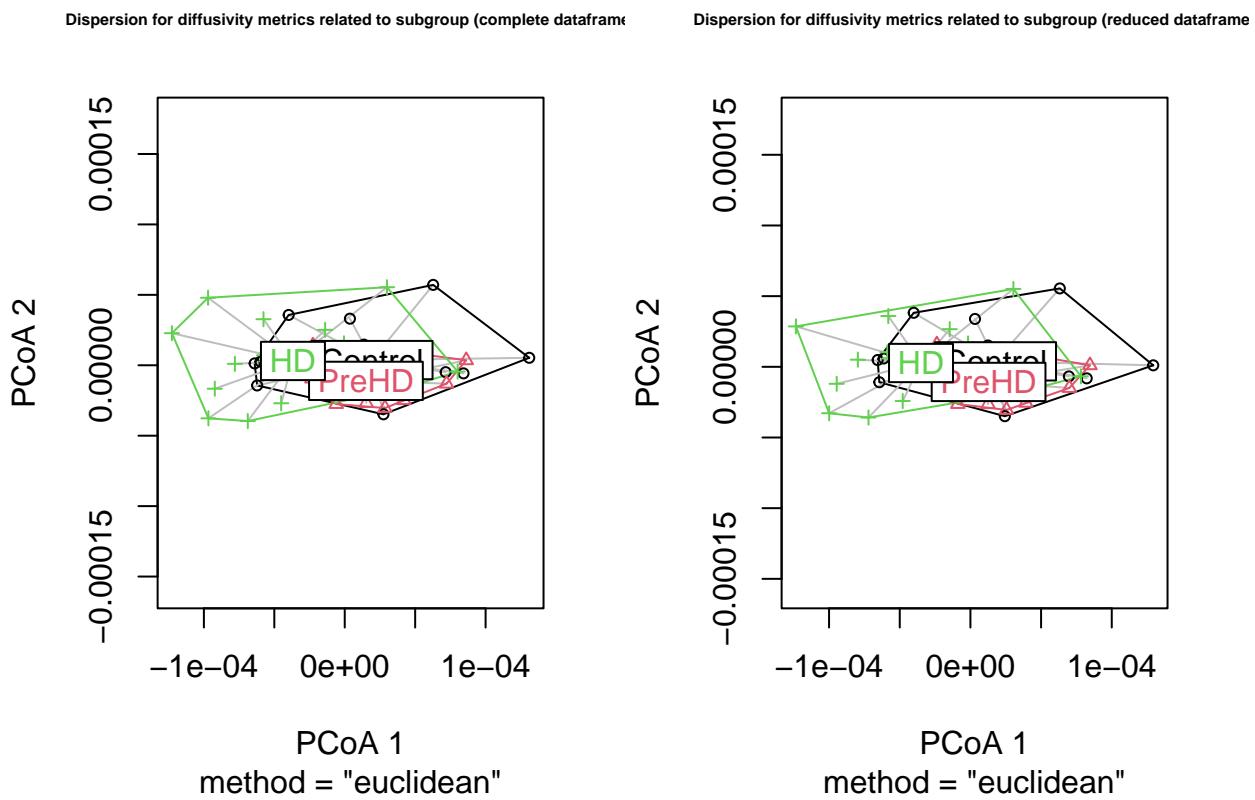
```
dispersionDifIron<-betadisper(d=distDifIron, group=LATR_avg_demo$Subgroup,
                                 type="centroid", bias.adjust=TRUE)
dispersionDifIron_WoOut<-betadisper(d=distDifIron_WoOut,
                                      group=LATR_avg_demo_WoOut$Subgroup,
                                      type="centroid", bias.adjust=TRUE)

dispersionADMDRD<-betadisper(d=distADMDRD, group=LATR_avg_demo_ADRDMD$Subgroup,
                                 type="centroid", bias.adjust=TRUE)
dispersionADMDRD_WoOut<-betadisper(d=distADMDRD_WoOut,
                                      group=LATR_avg_demo_WoOut_ADRDMD$Subgroup,
                                      type="centroid", bias.adjust=TRUE)

dispersionIronFA<-betadisper(d=distIronFA, group=LATR_avg_demo_IronFA$Subgroup,
                               type="centroid", bias.adjust=TRUE)
dispersionIronFA_WoOut<-betadisper(d=distIronFA_WoOut,
                                      group=LATR_avg_demo_WoOut_IronFA$Subgroup,
                                      type="centroid", bias.adjust=TRUE)
```

Se generan **plots** para cada una de las dispersiones.





Cabe mencionar que el subgrupo preHD presenta una menor dispersión respecto a los subgrupos HD y control, y,

que el subgrupo HD presenta una dispersión similar respecto al subgrupo control.

Se generan **modelos anova** para cada una de las dispersiones.

```
summary(aov(dispersionDifIron$distances~LATR_avg_demo$Subgroup))
          Df   Sum Sq   Mean Sq F value Pr(>F)
LATR_avg_demo$Subgroup  2 0.000954 0.0004768   2.211   0.12
Residuals                 52 0.011212 0.0002156

summary(aov(dispersionDifIron_WoOut$distances~LATR_avg_demo_WoOut$Subgroup))
          Df   Sum Sq   Mean Sq F value Pr(>F)
LATR_avg_demo_WoOut$Subgroup  2 0.000673 0.0003367   1.659   0.2
Residuals                 51 0.010350 0.0002029

summary(aov(dispersionADMDRD$distances~LATR_avg_demo_ADRDMD$Subgroup))
          Df   Sum Sq   Mean Sq F value Pr(>F)
LATR_avg_demo_ADRDMD$Subgroup  2 3.260e-09 1.628e-09   2.133   0.129
Residuals                 52 3.969e-08 7.632e-10

summary(aov(dispersionADMDRD_WoOut$distances~LATR_avg_demo_WoOut_ADRDMD$Subgroup))
          Df   Sum Sq   Mean Sq F value Pr(>F)
LATR_avg_demo_WoOut_ADRDMD$Subgroup  2 2.710e-09 1.353e-09   1.778   0.179
Residuals                 51 3.881e-08 7.611e-10

summary(aov(dispersionIronFA$distances~LATR_avg_demo_IronFA$Subgroup))
          Df   Sum Sq   Mean Sq F value Pr(>F)
LATR_avg_demo_IronFA$Subgroup  2 0.000954 0.0004768   2.211   0.12
Residuals                 52 0.011212 0.0002156

summary(aov(dispersionIronFA_WoOut$distances~LATR_avg_demo_WoOut_IronFA$Subgroup))
          Df   Sum Sq   Mean Sq F value Pr(>F)
LATR_avg_demo_WoOut_IronFA$Subgroup  2 0.001079 0.0005395   2.553 0.0878 .
Residuals                 51 0.010779 0.0002113

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Ninguna de las dispersiones presentan significación estadística, con lo que las diferencias observadas en PERMANOVA son debidas a diferencias en el posicionamiento de los centroides.

### 3.2.3 Comparaciones por pares

Se procede a generar las **comparaciones por pares** de los modelos PERMANOVA que han obtenido significación estadística para subgrupo y cuyas dispersiones no hayan resultado estadísticamente significativas. Por tanto, se van a realizar las comparaciones de los siguientes PERMANOVA: permADMDRD\_log2 y permADMDRD\_WO\_log2 (PERMANOVA de las transformaciones logarítmicas de las 3 métricas de difusividad del ATR derecho con el dataframe completo y reducido, respectivamente), y, permADMDRD3 y permADMDRD3\_WoOut (PERMANOVA de las 3 métricas de difusividad del ATR izquierdo con el dataframe completo y reducido, respectivamente).

Para la comparación por pares, se utiliza la función pairwise.adonis2 del paquete pairwiseAdonis. Esta función admite los mismos argumentos que la función adonis2. En la comparación por pares, tan sólo se ha puesto como variable independiente el subgrupo, dado que el resto de factores no han sido significativos en los PERMANOVA.

Para el **ATR derecho**, se generan las comparaciones por pares para las transformaciones logarítmicas de las 3 métricas de difusividad, tanto con el dataframe completo (sin eliminación de outliers) como con el dataframe con eliminación de outliers.

```
# Establecer seed para que el resultado de las permutaciones sea siempre el mismo
set.seed(123)

# Matriz de distancias para las 3 métricas de difusividad (AD, RD, MD)
```

```

distADMDRD_log<-vegdist(RATR_avg_demo_ADRDMD_log[,c('rhatr_AD','rhatr_MD','rhatr_RD')],  

                           method='euclidean')  

distADMDRD_WO_log<-vegdist(RATR_avg_demo_ADRDMD_WO_log[,c('rhatr_AD','rhatr_MD','rhatr_RD')],  

                           method='euclidean')

#Test a posteriori
library(devtools)
install_github("pmartinezarbizu/pairwiseAdonis/pairwiseAdonis", force=TRUE)
vegan (2.6-4 -> 2.6-6.1) [CRAN]
-- R CMD build -----
  v  checking for file 'C:/Users/monts/AppData/Local/Temp/Rtmp2b1GyB/remotes8d282ced7af8/pmartinezarbi...
    - preparing 'pairwiseAdonis':  

  checking DESCRIPTION meta-information ...      checking DESCRIPTION meta-information ...      v  checking D...
    - checking for LF line-endings in source and make files and shell scripts  

  - checking for empty or unneeded directories  

  Omitted 'LazyData' from DESCRIPTION  

    - building 'pairwiseAdonis_0.4.1.tar.gz'

library(pairwiseAdonis)
pairwise_permanova_RATR_ADMDRD_log<-pairwiseAdonis::pairwise.adonis2(distADMDRD_log~  

                        Subgroup,  

                        data=RATR_avg_demo_ADRDMD_log,  

                        permutations=999, by='terms')  

pairwise_permanova_RATR_ADMDRD_WO_log<-pairwiseAdonis::pairwise.adonis2(distADMDRD_WO_log~  

                        Subgroup,  

                        data=RATR_avg_demo_ADRDMD_WO_log,  

                        permutations=999, by='terms')

pairwise_permanova_RATR_ADMDRD_log
$parent_call
[1] "distADMDRD_log ~ Subgroup , strata = Null , permutations 999"

$Control_vs_HD
  Df SumOfSqs      R2      F Pr(>F)
Subgroup  1  0.08361 0.22172 11.396  0.005 ***
Residual 40  0.29347 0.77828
Total     41  0.37708 1.00000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

$Control_vs_PreHD
  Df SumOfSqs      R2      F Pr(>F)
Subgroup  1  0.00467 0.02803 1.0093  0.335
Residual 35  0.16194 0.97197
Total     36  0.16661 1.00000

$HD_vs_PreHD
  Df SumOfSqs      R2      F Pr(>F)
Subgroup  1  0.09325 0.25397 9.8727  0.001 ***
Residual 29  0.27392 0.74603
Total     30  0.36717 1.00000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

attr("class")

```

```

[1] "pwadstrata" "list"
pairwise_permanova_RATR_ADMRD_WO_log
$parent_call
[1] "distADMRD_WO_log ~ Subgroup , strata = Null , permutations 999"

$Control_vs_HD
      Df SumOfSqs      R2      F Pr(>F)
Subgroup  1  0.06313 0.19839 9.652  0.003 **
Residual 39  0.25510 0.80161
Total     40  0.31824 1.00000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

$Control_vs_PreHD
      Df SumOfSqs      R2      F Pr(>F)
Subgroup  1  0.00467 0.02803 1.0093  0.341
Residual 35  0.16194 0.97197
Total     36  0.16661 1.00000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

$HD_vs_PreHD
      Df SumOfSqs      R2      F Pr(>F)
Subgroup  1  0.074054 0.23919 8.8031  0.002 **
Residual 28  0.235544 0.76081
Total     29  0.309598 1.00000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

attr(,"class")
[1] "pwadstrata" "list"

```

Las comparaciones por pares ponen de relieve que **las diferencias significativas entre subgrupos en cuanto a las métricas AD-RD-MD se hallan entre HD vs controles y entre HD vs preHD para ambos dataframes** (con y sin eliminación de outliers), con valores de difusividad mayores para los pacientes HD sintomáticos.

Para el **ATR izquierdo**, se generan las comparaciones por pares para las 5 métricas conjuntas, para las 3 métricas de difusividad, y, para las métricas Iron-FA.

```

# Establecer seed para que el resultado de las permutaciones sea siempre el mismo
set.seed(123)

# Matriz de distancias para las 5 métricas
distDifIron<-vegdist(LATR_avg_demo[, -c(1:7)], method='euclidean')
distDifIron_WoOut<-vegdist(LATR_avg_demo_WoOut[, -c(1:7)], method='euclidean')

# Matriz de distancias para las 3 métricas de difusividad (AD, RD, MD)
distADMRD<-vegdist(LATR_avg_demo_ADRMD[, c('lhatr_AD', 'lhatr_RD', 'lhatr_MD')], 
                     method='euclidean')
distADMRD_WoOut<-vegdist(LATR_avg_demo_WoOut_ADRMD[, c('lhatr_AD', 'lhatr_RD',
                                                       'lhatr_MD')], 
                           method='euclidean')

# Matriz de distancias para las métricas de Iron y FA
distIronFA<-vegdist(LATR_avg_demo_IronFA[, c('lhatr_Iron', 'lhatr_FA')], 
                     method='euclidean')
distIronFA_WoOut<-vegdist(LATR_avg_demo_WoOut_IronFA[, c('lhatr_Iron', 'lhatr_FA')], 
                           method='euclidean')

```

```

#Test a posteriori para DifIron
pairwise_permanova_LATR_DifIron<-pairwiseAdonis::pairwise.adonis2(distDifIron~Subgroup,
                                                               data=LATR_avg_demo,
                                                               permutations=999,
                                                               by='terms')
pairwise_permanova_LATR_DifIron_WoOut<-pairwiseAdonis::pairwise.adonis2(distDifIron_WoOut~
                                                               Subgroup,
                                                               data=LATR_avg_demo_WoOut,
                                                               permutations=999,
                                                               by='terms')

#Test a posteriori para AD-MD-RD
pairwise_permanova_LATR_ADMRD<-pairwiseAdonis::pairwise.adonis2(distADMRD~Subgroup,
                                                               data=LATR_avg_demo_ADRDMD,
                                                               permutations=999,
                                                               by='terms')
pairwise_permanova_LATR_ADMRD_WoOut<-pairwiseAdonis::pairwise.adonis2(distADMRD_WoOut~
                                                               Subgroup,
                                                               data=LATR_avg_demo_WoOut_ADRDMD,
                                                               permutations=999,
                                                               by='terms')

#Test a posteriori para Iron-FA
pairwise_permanova_LATR_IronFA<-pairwiseAdonis::pairwise.adonis2(distIronFA~Subgroup,
                                                               data=LATR_avg_demo_IronFA,
                                                               permutations=999,
                                                               by='terms')
pairwise_permanova_LATR_IronFA_WoOut<-pairwiseAdonis::pairwise.adonis2(distIronFA_WoOut~
                                                               Subgroup,
                                                               data=LATR_avg_demo_WoOut_IronFA,
                                                               permutations=999,
                                                               by='terms')

```

```

pairwise_permanova_LATR_DifIron
$parent_call
[1] "distDifIron ~ Subgroup , strata = Null , permutations 999"

```

```

$Control_vs_HD
  Df  SumOfSqs      R2      F Pr(>F)
Subgroup  1 0.0000377 0.0014 0.056  0.817
Residual 40 0.0269425 0.9986
Total    41 0.0269802 1.0000

```

```

$Control_vs_PreHD
  Df  SumOfSqs      R2      F Pr(>F)
Subgroup  1 0.0024008 0.13782 5.5946  0.025 *
Residual 35 0.0150195 0.86218
Total    36 0.0174203 1.00000
---
```

```

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

$HD_vs_PreHD
  Df  SumOfSqs      R2      F Pr(>F)
Subgroup  1 0.0026558 0.1194 3.9323  0.06 .
Residual 29 0.0195865 0.8806

```

```

Total      30 0.0222423 1.00000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

attr("class")
[1] "pwadstrata" "list"
pairwise_permanova_LATR_DifIron_WoOut
$parent_call
[1] "distDifIron_WoOut ~ Subgroup , strata = Null , permutations 999"

$Control_vs_HD
    Df   SumOfSqs       R2       F Pr(>F)
Subgroup  1 0.0000088 0.00036 0.014  0.935
Residual 39 0.0244741 0.99964
Total     40 0.0244828 1.00000

$Control_vs_PreHD
    Df   SumOfSqs       R2       F Pr(>F)
Subgroup  1 0.0024008 0.13782 5.5946  0.026 *
Residual 35 0.0150195 0.86218
Total     36 0.0174203 1.00000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

$HD_vs_PreHD
    Df   SumOfSqs       R2       F Pr(>F)
Subgroup  1 0.0018705 0.09851 3.0597  0.111
Residual 28 0.0171180 0.90149
Total     29 0.0189886 1.00000

attr("class")
[1] "pwadstrata" "list"

```

Las comparaciones por pares ponen de relieve que **las diferencias significativas entre subgrupos en cuanto a las 5 métricas analizadas conjuntamente** se hallan entre **preHD vs controles** para ambos dataframes (con y sin eliminación de outliers).

```

pairwise_permanova_LATR_ADMRD
$parent_call
[1] "distADMRD ~ Subgroup , strata = Null , permutations 999"

$Control_vs_HD
    Df   SumOfSqs       R2       F Pr(>F)
Subgroup  1 3.2494e-08 0.21084 10.687  0.003 **
Residual 40 1.2163e-07 0.78916
Total     41 1.5412e-07 1.00000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

$Control_vs_PreHD
    Df   SumOfSqs       R2       F Pr(>F)
Subgroup  1 1.9970e-09 0.02546 0.9146  0.363
Residual 35 7.6443e-08 0.97454
Total     36 7.8441e-08 1.00000

$HD_vs_PreHD

```

```

      Df   SumOfSqs      R2      F Pr(>F)
Subgroup  1 2.1790e-08 0.21276 7.8374  0.004 ** 
Residual  29 8.0626e-08 0.78724
Total     30 1.0242e-07 1.00000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

attr(,"class")
[1] "pwadstrata" "list"
pairwise_permanova_LATR_ADMRD_WoOut
$parent_call
[1] "distADMRD_WoOut ~ Subgroup , strata = Null , permutations 999"

$Control_vs_HD
      Df   SumOfSqs      R2      F Pr(>F)
Subgroup  1 2.7700e-08 0.19323 9.341  0.006 ** 
Residual  39 1.1565e-07 0.80677
Total     40 1.4335e-07 1.00000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

$Control_vs_PreHD
      Df   SumOfSqs      R2      F Pr(>F)
Subgroup  1 1.9970e-09 0.02546 0.9146  0.378
Residual  35 7.6443e-08 0.97454
Total     36 7.8441e-08 1.00000

$HD_vs_PreHD
      Df   SumOfSqs      R2      F Pr(>F)
Subgroup  1 1.8153e-08 0.1956  6.8086  0.007 ** 
Residual  28 7.4652e-08 0.8044
Total     29 9.2805e-08 1.00000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

attr(,"class")
[1] "pwadstrata" "list"

```

Al igual que en el ATR derecho, las comparaciones por pares ponen de relieve que **las diferencias significativas entre subgrupos en cuanto a las métricas AD-RD-MD se hallan entre HD vs controles y entre HD vs preHD para ambos dataframes** (con y sin eliminación de outliers), con valores de difusividad mayores para los pacientes HD sintomáticos.

```

pairwise_permanova_LATR_IronFA
$parent_call
[1] "distIronFA ~ Subgroup , strata = Null , permutations 999"

$Control_vs_HD
      Df   SumOfSqs      R2      F Pr(>F)
Subgroup  1 0.0000377 0.0014  0.056  0.826
Residual  40 0.0269424 0.9986
Total     41 0.0269801 1.00000

$Control_vs_PreHD
      Df   SumOfSqs      R2      F Pr(>F)
Subgroup  1 0.0024008 0.13782 5.5947  0.025 *

```

```

Residual 35 0.0150194 0.86218
Total     36 0.0174202 1.00000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

$HD_vs_PreHD
  Df  SumOfSqs      R2      F Pr(>F)
Subgroup 1 0.0026558 0.1194 3.9323  0.059 .
Residual 29 0.0195864 0.8806
Total    30 0.0222422 1.0000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

attr("class")
[1] "pwadstrata" "list"
pairwise_permanova_LATR_IronFA_WoOut
$parent_call
[1] "distIronFA_WoOut ~ Subgroup , strata = Null , permutations 999"

$Control_vs_HD
  Df  SumOfSqs      R2      F Pr(>F)
Subgroup 1 0.0001231 0.0048 0.1882  0.683
Residual 39 0.0255063 0.9952
Total    40 0.0256293 1.0000

$Control_vs_PreHD
  Df  SumOfSqs      R2      F Pr(>F)
Subgroup 1 0.0019365 0.12478 4.8472  0.027 *
Residual 34 0.0135833 0.87522
Total    35 0.0155198 1.00000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

$HD_vs_PreHD
  Df  SumOfSqs      R2      F Pr(>F)
Subgroup 1 0.0026558 0.1194 3.9323  0.083 .
Residual 29 0.0195864 0.8806
Total    30 0.0222422 1.0000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

attr("class")
[1] "pwadstrata" "list"

```

Las comparaciones por pares para ambos dataframes (con y sin eliminación de outliers) ponen de relieve que **las diferencias significativas entre subgrupos en cuanto a las métricas Iron-FA se hallan entre preHD vs controles**, y, las diferencias son casi significativas con p-valores inferiores a 0.10 entre preHD y HD.

### 3.3 ANOVA

El **ANOVA** (**ANalysis Of VAriance**) se realiza para determinar qué métricas (individuales) de RM difieren entre subgrupos (variable independiente de interés), controlando por género y edad factorizada (covariables). Por tanto, se generan 5 **anovas**, uno para cada métrica de RM (iron, ad, rd, md, fa), en los que:

- Las métricas de RM son las variables dependientes;
- El subgrupo es la variable independiente de interés; y
- El género y la edad factorizada son las covariables a controlar.

Se generan los modelos con la **función aov** del paquete **stats**. Se procede a ejecutar **dos comprobaciones sobre los residuos** (siendo el residuo de cada observación la diferencia entre la observación real y la media del grupo al que pertenece):

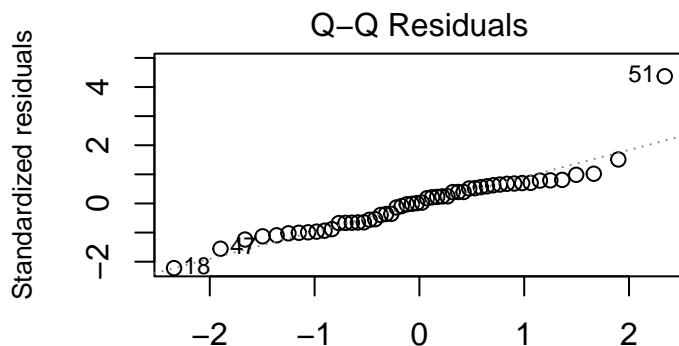
- Residuos presentan una **distribución normal**? Esto se comprueba mediante gráficos qqplot (la normalidad de los residuos se valora en función de la proximidad de éstos a la línea diagonal) y mediante el **test estadístico Shapiro-Wilk**.
- Residuos presentan **homocedasticidad**? Esto se comprueba mediante plots *fitted\_values vs Residuals* y mediante el **test de Breusch-Pagan**, el cual genera una regresión de los residuos sobre los valores predichos para determinar si ésta explica la varianza residual (valores de p menores de 0.05 indican heterocedasticidad).

En este caso, métrica por métrica, se determinan qué sujetos son posibles sujetos outliers/influyentes, y, en función del cumplimiento o no de los supuestos de ANOVA, se decide la eliminación o no de los posibles sujetos outliers/influyentes.

#### 3.3.1 IRON - RATR

Se procede a generar los modelos ANOVA. En el **primer modelo** generado **anova1\_RATRavg\_iron**, se introducen las **3 variables independientes con interacción**.

```
anova1_RATRavg_iron<-aov(rhatr_Iron~Subgroup*agef2*Gender, data=ATR_avg_demo)
summary(anova1_RATRavg_iron)
Df      Sum Sq   Mean Sq F value Pr(>F)
Subgroup        2 8.870e-06 4.433e-06   2.275  0.116
agef2           2 4.090e-06 2.044e-06   1.049  0.360
Gender          1 1.320e-06 1.317e-06   0.676  0.416
Subgroup:agef2  4 2.930e-06 7.320e-07   0.375  0.825
Subgroup:Gender 2 2.760e-06 1.381e-06   0.709  0.498
agef2:Gender    2 5.230e-06 2.614e-06   1.341  0.273
Subgroup:agef2:Gender 2 5.420e-06 2.710e-06   1.390  0.261
Residuals       39 7.601e-05 1.949e-06
plot(anova1_RATRavg_iron, which=2, cex.lab=0.8, cex.sub=0.8)
```

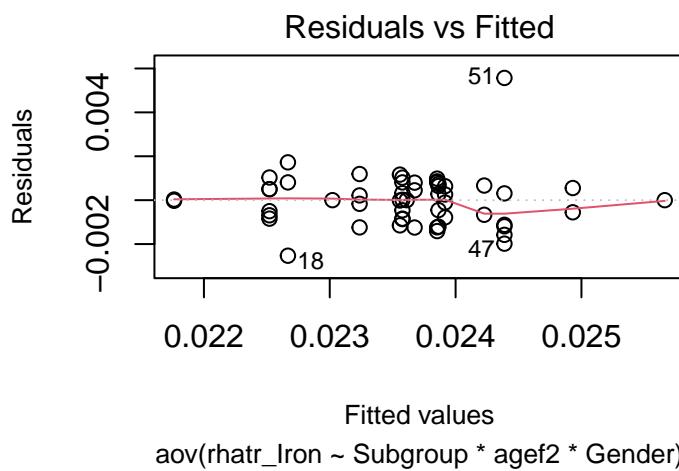


Theoretical Quantiles  
 $\text{aov}(\text{rhatr\_Iron} \sim \text{Subgroup} * \text{agef2} * \text{Gender})$

```
shapiro.test(residuals(anova1_RATRavg_iron))
```

```
Shapiro-Wilk normality test

data: residuals(anova1_RATRavg_iron)
W = 0.86812, p-value = 2.298e-05
plot(anova1_RATRavg_iron, which=1, cex.lab=0.8, cex.sub=0.8)
```



Fitted values  
 $\text{aov}(\text{rhatr\_Iron} \sim \text{Subgroup} * \text{agef2} * \text{Gender})$

```
library(lmtest)
bptest(anova1_RATRavg_iron)

studentized Breusch-Pagan test

data: anova1_RATRavg_iron
BP = 12.442, df = 15, p-value = 0.6453
```

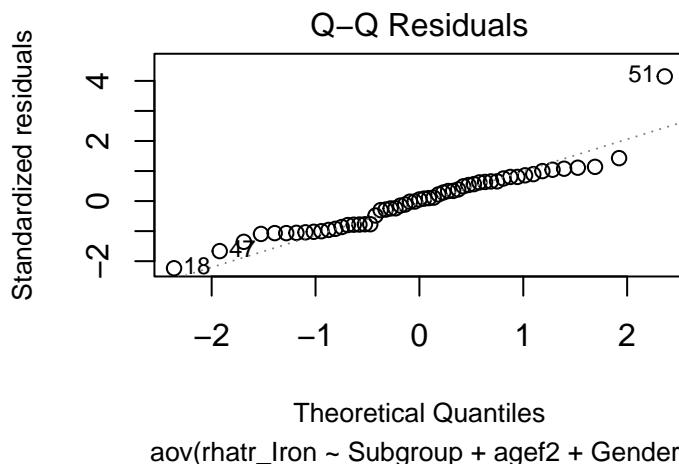
En este modelo, ningún factor o interacción es significativo; además, se evidencia que los residuos no presentan una

distribución normal, pero sí una homocedasticidad de varianzas. Los posibles sujetos outliers e influyentes son los individuos con índices 18, 47 y 51.

Se genera un **segundo modelo** `anova2_RATRavg_iron` con las **3 variables independientes sin interacción** dado que las interacciones no han sido significativas.

```
anova2_RATRavg_iron<-aov(rhatr_Iron~Subgroup+agef2+Gender, data=ATR_avg_demo)
summary(anova2_RATRavg_iron)
Df      Sum Sq  Mean Sq F value Pr(>F)
Subgroup    2 8.870e-06 4.433e-06   2.352  0.106
agef2       2 4.090e-06 2.044e-06   1.084  0.346
Gender       1 1.320e-06 1.317e-06   0.699  0.407
Residuals   49 9.235e-05 1.885e-06

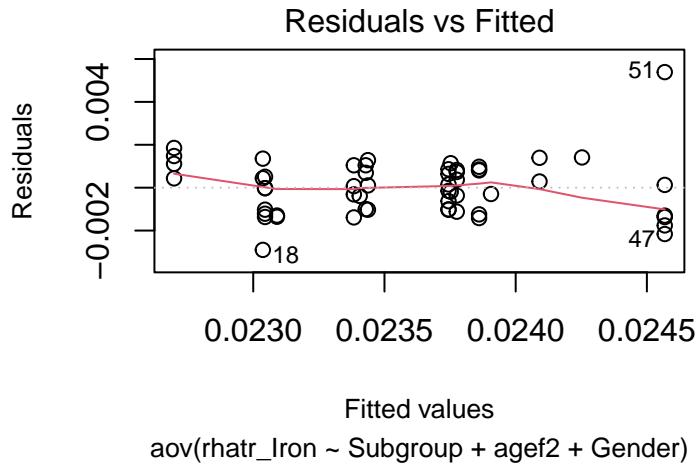
plot(anova2_RATRavg_iron, which=2, cex.lab=0.8, cex.sub=0.8)
```



```
shapiro.test(residuals(anova2_RATRavg_iron))

Shapiro-Wilk normality test

data: residuals(anova2_RATRavg_iron)
W = 0.91756, p-value = 0.001078
plot(anova2_RATRavg_iron, which=1, cex.lab=0.8, cex.sub=0.8)
```



```
bptest(anova2_RATRavg_iron)
```

```
studentized Breusch-Pagan test
```

```
data: anova2_RATRavg_iron
BP = 5.4667, df = 5, p-value = 0.3616
```

En este modelo, ningún factor es significativo; además, se evidencia que los residuos no presentan una distribución normal, pero sí una homocedasticidad de varianzas. Los posibles sujetos outliers e influyentes son los individuos con índices 18, 47 y 51.

Cuadro 47: ANOVA Models for Iron of the Right ATR

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Subgroup	2	8.9e-06	4.4e-06	2.2746959	0.1163078
agef2	2	4.1e-06	2.0e-06	1.0485426	0.3601181
Gender	1	1.3e-06	1.3e-06	0.6759003	0.4160002
Subgroup:agef2	4	2.9e-06	7.0e-07	0.3753292	0.8248114
Subgroup:Gender	2	2.8e-06	1.4e-06	0.7087785	0.4984754
agef2:Gender	2	5.2e-06	2.6e-06	1.3414154	0.2732715
Subgroup:agef2:Gender	2	5.4e-06	2.7e-06	1.3904795	0.2610243
Residuals	39	7.6e-05	1.9e-06	NA	NA

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Subgroup	2	8.90e-06	4.4e-06	2.3523392	0.1058052
agef2	2	4.10e-06	2.0e-06	1.0843330	0.3461016
Gender	1	1.30e-06	1.3e-06	0.6989712	0.4071882
Residuals	49	9.23e-05	1.9e-06	NA	NA

### 3.3.2 IRON - LATR

Se procede a generar los modelos ANOVA. En el **primer modelo** generado `anova1_LATRavg_iron`, se introducen las **3 variables independientes con interacción**.

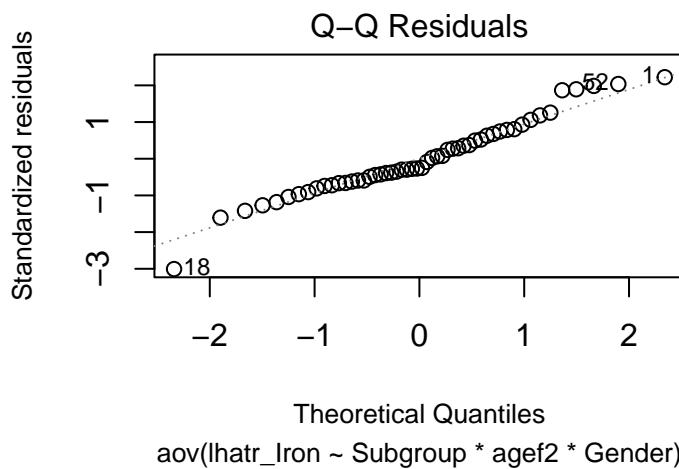
```
anova1_LATRavg_iron<-aov(lhatr_Iron~Subgroup*agef2*Gender, data=ATR_avg_demo)
summary(anova1_LATRavg_iron)
```

```

          Df    Sum Sq  Mean Sq F value Pr(>F)
Subgroup        2 3.800e-06 1.901e-06   0.536  0.589
agef2           2 1.066e-05 5.331e-06   1.503  0.235
Gender          1 5.000e-08 5.500e-08   0.016  0.902
Subgroup:agef2  4 6.590e-06 1.647e-06   0.464  0.761
Subgroup:Gender 2 1.191e-05 5.954e-06   1.679  0.200
agef2:Gender    2 2.134e-05 1.067e-05   3.008  0.061 .
Subgroup:agef2:Gender 2 1.454e-05 7.271e-06   2.050  0.142
Residuals      39 1.383e-04 3.547e-06

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
plot(anova1_LATRavg_iron, which=2, cex.lab=0.8, cex.sub=0.8)

```



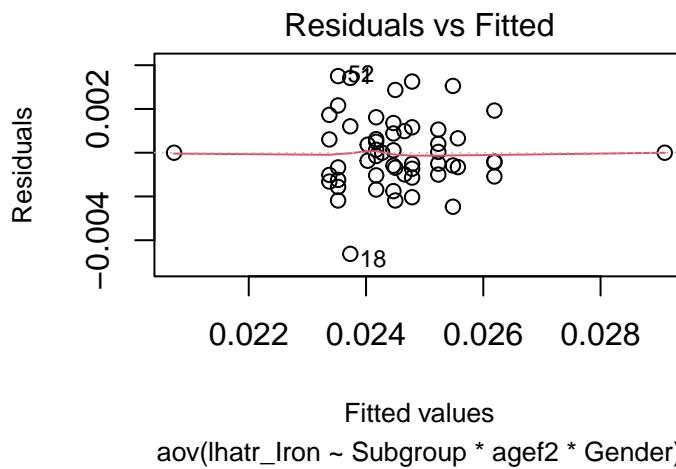
```

shapiro.test(residuals(anova1_LATRavg_iron))

Shapiro-Wilk normality test

data:  residuals(anova1_LATRavg_iron)
W = 0.97103, p-value = 0.2047
plot(anova1_LATRavg_iron, which=1, cex.lab=0.8, cex.sub=0.8)

```



```
bptest(anova1_LATRavg_iron)
```

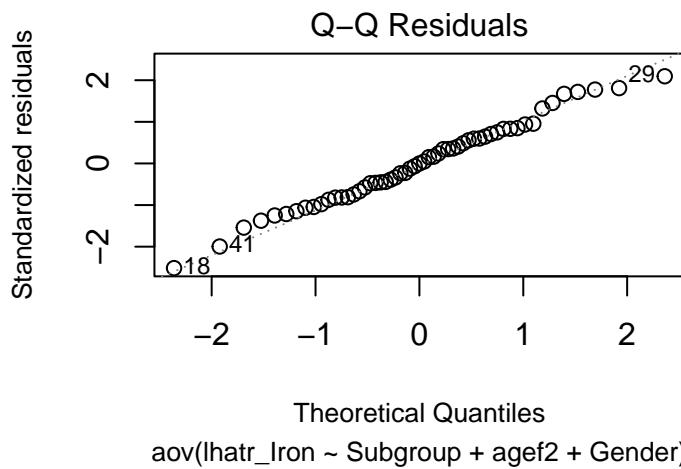
```
studentized Breusch-Pagan test
```

```
data: anova1_LATRavg_iron
BP = 25.263, df = 15, p-value = 0.04652
```

En este modelo, ningún factor o interacción es significativo; además, se evidencia que los residuos presentan una distribución normal y heterocedasticidad de varianzas. Los posibles sujetos outliers e influyentes son los individuos con índices 1, 18 y 52.

Se genera un **segundo modelo** `anova2_LATRavg_iron` con las **3 variables independientes sin interacción** dado que las interacciones no han sido significativas.

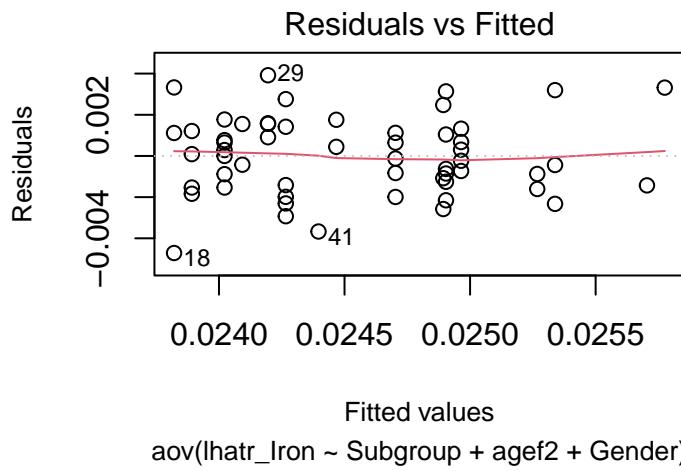
```
anova2_LATRavg_iron<-aov(lhatr_Iron~Subgroup+agef2+Gender, data=ATR_avg_demo)
summary(anova2_LATRavg_iron)
  Df   Sum Sq  Mean Sq F value Pr(>F)
Subgroup     2 3.800e-06 1.901e-06   0.483  0.620
agef2        2 1.066e-05 5.331e-06   1.355  0.267
Gender       1 5.000e-08 5.500e-08   0.014  0.906
Residuals   49 1.927e-04 3.933e-06
plot(anova2_LATRavg_iron, which=2, cex.lab=0.8, cex.sub=0.8)
```



```
shapiro.test(residuals(anova2_LATRavg_iron))
```

Shapiro-Wilk normality test

```
data: residuals(anova2_LATRavg_iron)
W = 0.98859, p-value = 0.8798
plot(anova2_LATRavg_iron, which=1, cex.lab=0.8, cex.sub=0.8)
```



```
bptest(anova2_LATRavg_iron)
```

studentized Breusch-Pagan test

```
data: anova2_LATRavg_iron
BP = 6.0476, df = 5, p-value = 0.3016
```

En este modelo, ningún factor es significativo y los residuos presentan una distribución normal y una homocedasticidad de varianzas. Los posibles sujetos outliers e influyentes son los individuos con índices 18, 29 y 41.

Cuadro 48: ANOVA Models for Iron of the Left ATR

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Subgroup	2	0.0000038	1.90e-06	0.5359929	0.5893360
agef2	2	0.0000107	5.30e-06	1.5028537	0.2350974
Gender	1	0.0000001	1.00e-07	0.0155001	0.9015595
Subgroup:agef2	4	0.0000066	1.60e-06	0.4642820	0.7614842
Subgroup:Gender	2	0.0000119	6.00e-06	1.6785428	0.1998488
agef2:Gender	2	0.0000213	1.07e-05	3.0076795	0.0609854
Subgroup:agef2:Gender	2	0.0000145	7.30e-06	2.0499927	0.1423823
Residuals	39	0.0001383	3.50e-06	NA	NA

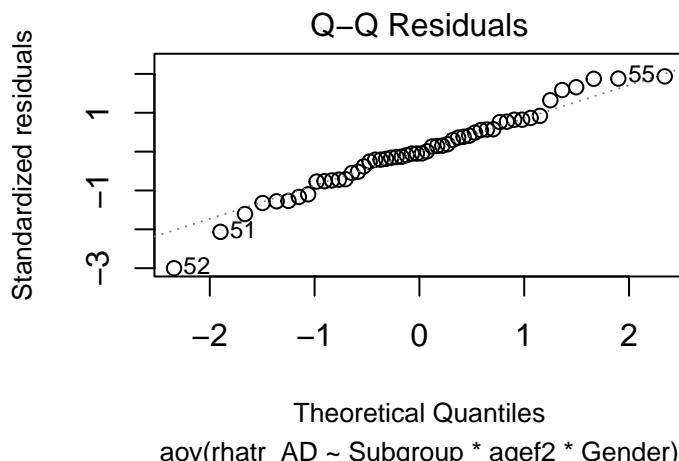
  

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Subgroup	2	0.0000038	1.9e-06	0.4834137	0.6195844
agef2	2	0.0000107	5.3e-06	1.3554285	0.2673323
Gender	1	0.0000001	1.0e-07	0.0139796	0.9063647
Residuals	49	0.0001927	3.9e-06	NA	NA

### 3.3.3 AD - RATR

Se procede a generar los modelos ANOVA. En el **primer modelo** generado `anova1_RATRavg_ad`, se introducen las **3 variables independientes con interacción**.

```
anova1_RATRavg_ad<-aov(rhatr_AD~Subgroup*agef2*Gender, data=ATR_avg_demo)
summary(anova1_RATRavg_ad)
Df Sum Sq Mean Sq F value Pr(>F)
Subgroup 2 2.337e-08 1.168e-08 7.028 0.00247 **
agef2 2 5.690e-09 2.845e-09 1.712 0.19386
Gender 1 2.100e-10 2.140e-10 0.129 0.72161
Subgroup:agef2 4 6.050e-09 1.512e-09 0.909 0.46799
Subgroup:Gender 2 1.200e-10 6.000e-11 0.036 0.96474
agef2:Gender 2 2.940e-09 1.471e-09 0.885 0.42090
Subgroup:agef2:Gender 2 2.130e-09 1.065e-09 0.641 0.53227
Residuals 39 6.483e-08 1.662e-09
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
plot(anova1_RATRavg_ad, which=2, cex.lab=0.8, cex.sub=0.8)
```



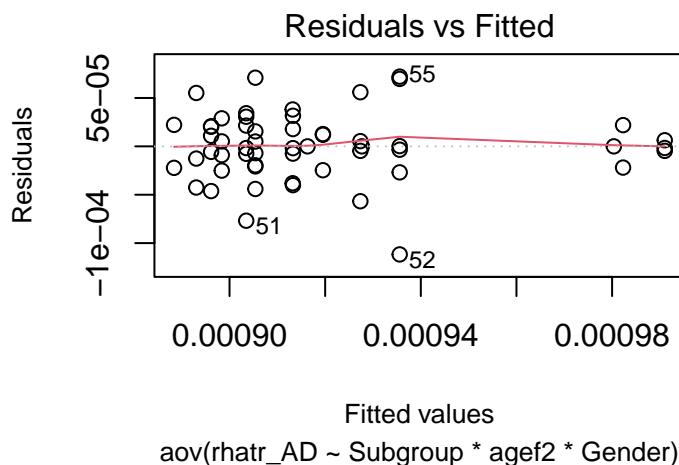
```

shapiro.test(residuals(anova1_RATRavg_ad))

Shapiro-Wilk normality test

data: residuals(anova1_RATRavg_ad)
W = 0.96634, p-value = 0.1258
plot(anova1_RATRavg_ad, which=1, cex.lab=0.8, cex.sub=0.8)

```



```

bptest(anova1_RATRavg_ad)

studentized Breusch-Pagan test

data: anova1_RATRavg_ad
BP = 14.293, df = 15, p-value = 0.5034

```

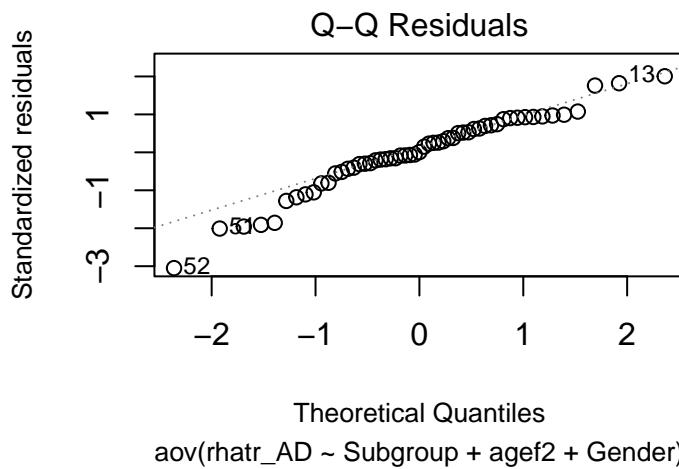
En este modelo, la variable subgrupo es significativa y los residuos presentan una distribución normal y homocedasticidad de varianzas. Los posibles sujetos outliers e influyentes son los individuos con índices 51, 52 y 55.

Se genera un **segundo modelo** anova2\_RATRavg\_ad con las **3 variables independientes sin interacción** dado que las interacciones no han sido significativas.

```

anova2_RATRavg_ad<-aov(rhatr_AD~Subgroup+agef2+Gender, data=ATR_avg_demo)
summary(anova2_RATRavg_ad)
  Df   Sum Sq   Mean Sq F value    Pr(>F)
Subgroup     2 2.337e-08 1.168e-08   7.526 0.00141 ** 
agef2        2 5.690e-09 2.845e-09   1.833 0.17077  
Gender       1 2.100e-10 2.140e-10   0.138 0.71196  
Residuals   49 7.607e-08 1.552e-09
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
plot(anova2_RATRavg_ad, which=2, cex.lab=0.8, cex.sub=0.8)

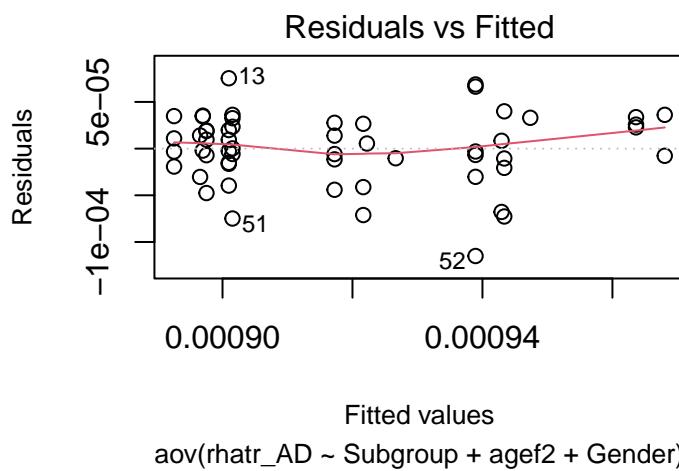
```



```
shapiro.test(residuals(anova2_RATRavg_ad))
```

Shapiro-Wilk normality test

```
data: residuals(anova2_RATRavg_ad)
W = 0.95971, p-value = 0.06274
plot(anova2_RATRavg_ad, which=1, cex.lab=0.8, cex.sub=0.8)
```



```
bptest(anova2_RATRavg_ad)
```

studentized Breusch-Pagan test

```
data: anova2_RATRavg_ad
BP = 4.6269, df = 5, p-value = 0.4631
```

En este modelo, la variable subgrupo es significativa, los residuos presentan una distribución normal y homocedasticidad de varianzas, y los posibles sujetos outliers e influyentes son los individuos con índices 13, 51 y 52.

Cuadro 49: ANOVA Models for AD of the Right ATR

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Subgroup	2	0e+00	0	7.0283991	0.0024735
agef2	2	0e+00	0	1.7116102	0.1938603
Gender	1	0e+00	0	0.1288074	0.7216075
Subgroup:agef2	4	0e+00	0	0.9094804	0.4679863
Subgroup:Gender	2	0e+00	0	0.0359287	0.9647410
agef2:Gender	2	0e+00	0	0.8848573	0.4208964
Subgroup:agef2:Gender	2	0e+00	0	0.6409109	0.5322704
Residuals	39	1e-07	0	NA	NA

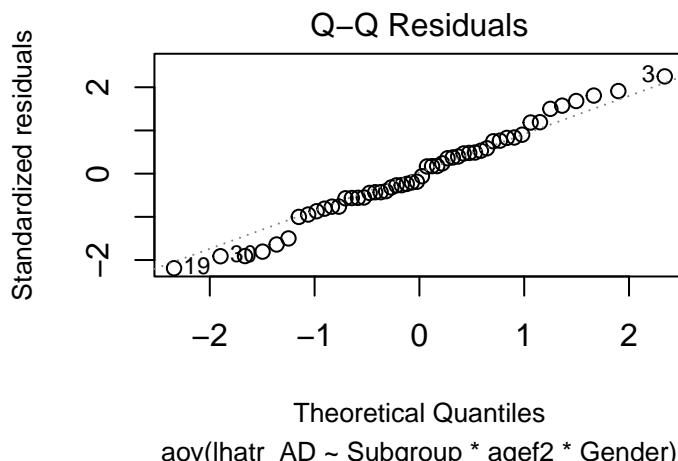
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Subgroup	2	0e+00	0	7.5258230	0.0014119
agef2	2	0e+00	0	1.8327467	0.1707706
Gender	1	0e+00	0	0.1379235	0.7119551
Residuals	49	1e-07	0	NA	NA

### 3.3.4 AD - LATR

Se procede a generar los modelos ANOVA. En el **primer modelo** generado `anova1_LATRavg_ad`, se introducen las **3 variables independientes con interacción**.

```
anova1_LATRavg_ad<-aov(lhatr_AD~Subgroup*agef2*Gender, data=ATR_avg_demo)
summary(anova1_LATRavg_ad)

Df      Sum Sq   Mean Sq F value    Pr(>F)
Subgroup        2 1.653e-08 8.264e-09  5.731 0.00658 ***
agef2           2 1.520e-09 7.580e-10  0.526 0.59532
Gender          1 2.290e-09 2.286e-09  1.585 0.21551
Subgroup:agef2  4 3.330e-09 8.310e-10  0.576 0.68138
Subgroup:Gender 2 2.100e-10 1.040e-10  0.072 0.93081
agef2:Gender     2 4.820e-09 2.410e-09  1.671 0.20118
Subgroup:agef2:Gender 2 1.250e-09 6.230e-10  0.432 0.65230
Residuals       39 5.624e-08 1.442e-09
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
plot(anova1_LATRavg_ad, which=2, cex.lab=0.8, cex.sub=0.8)
```



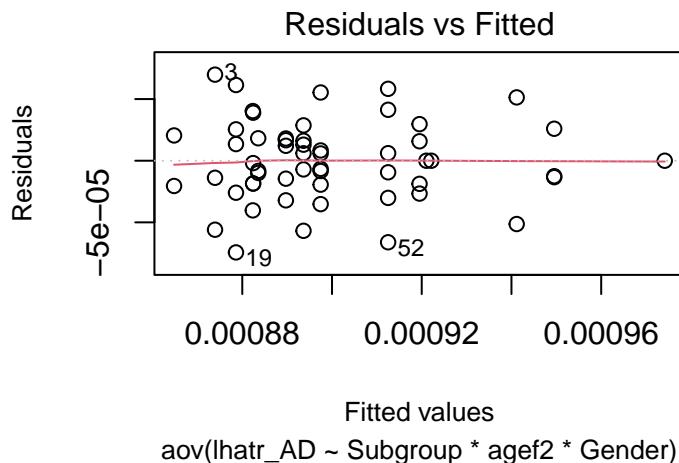
```

shapiro.test(residuals(anova1_LATRavg_ad))

Shapiro-Wilk normality test

data: residuals(anova1_LATRavg_ad)
W = 0.9889, p-value = 0.8911
plot(anova1_LATRavg_ad, which=1, cex.lab=0.8, cex.sub=0.8)

```



Fitted values  
aov(lhatr\_AD ~ Subgroup \* agef2 \* Gender)

```

bptest(anova1_LATRavg_ad)

studentized Breusch-Pagan test

data: anova1_LATRavg_ad
BP = 19.425, df = 15, p-value = 0.1951

```

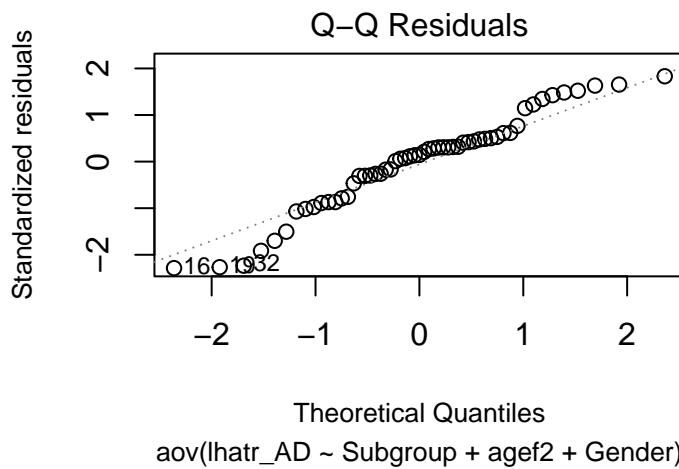
En este modelo, la variable subgrupo es significativa y los residuos presentan una distribución normal y homocedasticidad de varianzas. Los posibles sujetos outliers son los individuos con índices 3, 19 y 30 y los posibles sujetos influyentes son los individuos con índices 3, 19 y 52.

Se genera un **segundo modelo** anova2\_LATRavg\_ad con las **3 variables independientes sin interacción** dado que las interacciones no han sido significativas.

```

anova2_LATRavg_ad<-aov(lhatr_AD~Subgroup+agef2+Gender, data=ATR_avg_demo)
summary(anova2_LATRavg_ad)
      Df   Sum Sq  Mean Sq F value    Pr(>F)
Subgroup     2 1.653e-08 8.264e-09  6.151 0.00414 ***
agef2        2 1.520e-09 7.580e-10  0.564 0.57251
Gender        1 2.290e-09 2.286e-09  1.701 0.19822
Residuals    49 6.584e-08 1.344e-09
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
plot(anova2_LATRavg_ad, which=2, cex.lab=0.8, cex.sub=0.8)

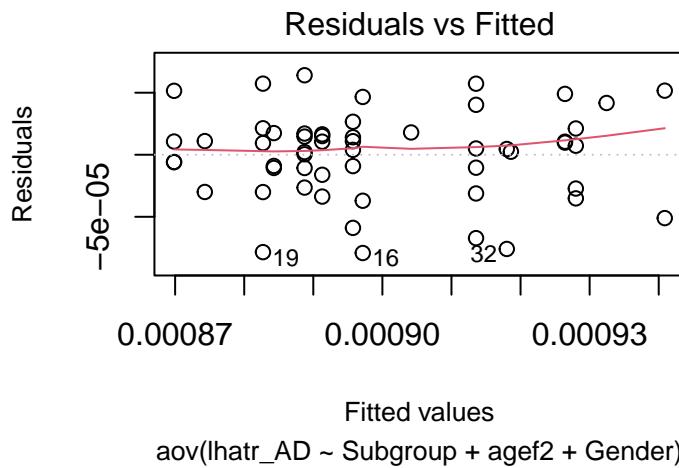
```



```
shapiro.test(residuals(anova2_LATRavg_ad))
```

Shapiro-Wilk normality test

```
data: residuals(anova2_LATRavg_ad)
W = 0.95927, p-value = 0.0599
plot(anova2_LATRavg_ad, which=1, cex.lab=0.8, cex.sub=0.8)
```



```
bptest(anova2_LATRavg_ad)
```

studentized Breusch-Pagan test

```
data: anova2_LATRavg_ad
BP = 7.2413, df = 5, p-value = 0.2033
```

En este modelo, la variable subgrupo es significativa, los residuos presentan una distribución normal y homocedasticidad de varianzas, y los posibles sujetos outliers e influyentes son los individuos con índices 16, 19 y 32.

Cuadro 50: ANOVA Models for AD of the Left ATR

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Subgroup	2	0e+00	0	5.7310501	0.0065756
agef2	2	0e+00	0	0.5256186	0.5953181
Gender	1	0e+00	0	1.5851108	0.2155100
Subgroup:agef2	4	0e+00	0	0.5764519	0.6813762
Subgroup:Gender	2	0e+00	0	0.0718337	0.9308085
agef2:Gender	2	0e+00	0	1.6713159	0.2011832
Subgroup:agef2:Gender	2	0e+00	0	0.4319635	0.6523014
Residuals	39	1e-07	0	NA	NA

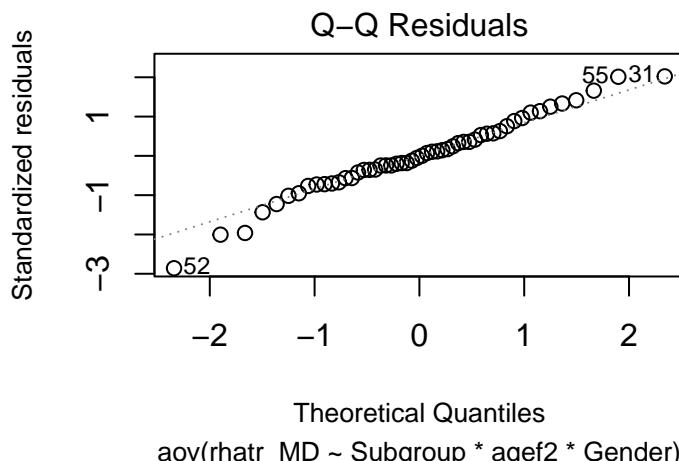
  

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Subgroup	2	0e+00	0	6.1508070	0.0041375
agef2	2	0e+00	0	0.5641163	0.5725130
Gender	1	0e+00	0	1.7012085	0.1982236
Residuals	49	1e-07	0	NA	NA

### 3.3.5 MD - RATR

Se procede a generar los modelos ANOVA. En el **primer modelo** generado `anova1_RATRavg_md`, se introducen las **3 variables independientes con interacción**.

```
anova1_RATRavg_md<-aov(rhatr_MD~Subgroup*agef2*Gender, data=ATR_avg_demo)
summary(anova1_RATRavg_md)
Df      Sum Sq   Mean Sq F value    Pr(>F)
Subgroup        2 1.724e-08 8.622e-09 10.188 0.000276 ***
agef2           2 4.210e-09 2.106e-09  2.489 0.096121 .
Gender          1 1.200e-10 1.150e-10  0.136 0.714041
Subgroup:agef2  4 2.780e-09 6.950e-10  0.821 0.519553
Subgroup:Gender 2 2.200e-10 1.100e-10  0.130 0.878707
agef2:Gender     2 1.060e-09 5.310e-10  0.628 0.539224
Subgroup:agef2:Gender 2 5.100e-10 2.550e-10  0.302 0.741128
Residuals       39 3.301e-08 8.460e-10
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
plot(anova1_RATRavg_md, which=2, cex.lab=0.8, cex.sub=0.8)
```



Theoretical Quantiles  
`aov(rhatr_MD ~ Subgroup * agef2 * Gender)`

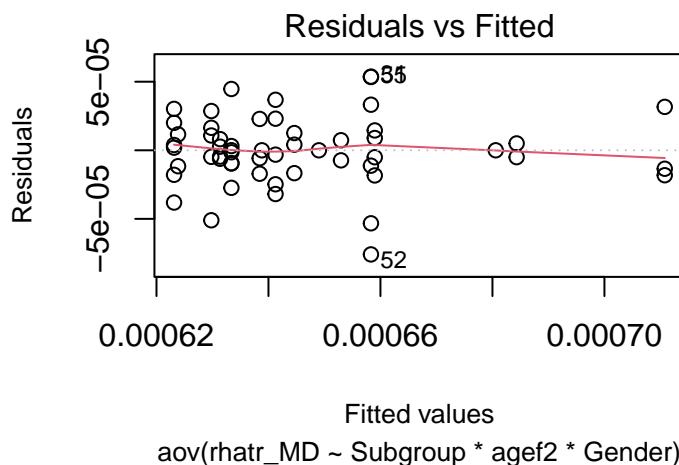
```

shapiro.test(residuals(anova1_RATRavg_md))

Shapiro-Wilk normality test

data: residuals(anova1_RATRavg_md)
W = 0.96979, p-value = 0.1802
plot(anova1_RATRavg_md, which=1, cex.lab=0.8, cex.sub=0.8)

```



```

bptest(anova1_RATRavg_md)

studentized Breusch-Pagan test

data: anova1_RATRavg_md
BP = 27.665, df = 15, p-value = 0.02376

```

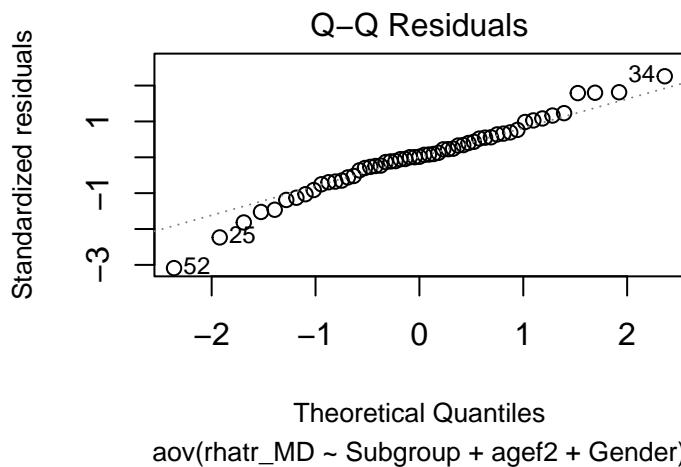
En este modelo, la variable subgrupo es significativa y los residuos presentan una distribución normal y heterocedasticidad de varianzas. Los posibles sujetos outliers e influyentes son los individuos con índices 31, 52 y 55.

Se genera un **segundo modelo** anova2\_RATRavg\_md con las **3 variables independientes sin interacción** dado que las interacciones no han sido significativas.

```

anova2_RATRavg_md<-aov(rhatr_MD~Subgroup+agef2+Gender, data=ATR_avg_demo)
summary(anova2_RATRavg_md)
Df      Sum Sq  Mean Sq F value    Pr(>F)
Subgroup   2 1.724e-08 8.622e-09 11.242 9.59e-05 ***
agef2      2 4.210e-09 2.106e-09  2.746  0.0741 .
Gender      1 1.200e-10 1.150e-10  0.150  0.6999
Residuals  49 3.758e-08 7.670e-10
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
plot(anova2_RATRavg_md, which=2, cex.lab=0.8, cex.sub=0.8)

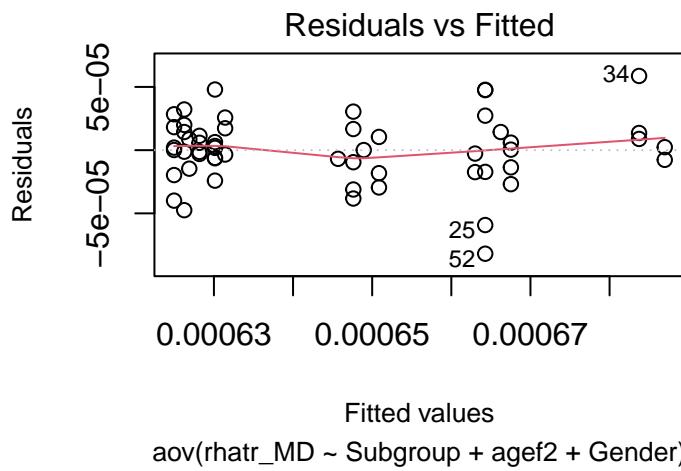
```



```
shapiro.test(residuals(anova2_RATRavg_md))
```

Shapiro-Wilk normality test

```
data: residuals(anova2_RATRavg_md)
W = 0.97492, p-value = 0.3031
plot(anova2_RATRavg_md, which=1, cex.lab=0.8, cex.sub=0.8)
```



```
bptest(anova2_RATRavg_md)
```

studentized Breusch-Pagan test

```
data: anova2_RATRavg_md
BP = 9.0718, df = 5, p-value = 0.1062
```

En este modelo, la variable subgrupo es significativa y los residuos presentan una distribución normal y homocedasticidad de varianzas. Los posibles sujetos outliers e influyentes son los individuos con índices 25, 34 y 52.

Cuadro 51: ANOVA Models for MD of the Right ATR

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Subgroup	2	0	0	10.1875145	0.0002757
agef2	2	0	0	2.4886086	0.0961211
Gender	1	0	0	0.1362438	0.7140405
Subgroup:agef2	4	0	0	0.8212831	0.5195532
Subgroup:Gender	2	0	0	0.1297337	0.8787068
agef2:Gender	2	0	0	0.6275086	0.5392244
Subgroup:agef2:Gender	2	0	0	0.3018957	0.7411275
Residuals	39	0	0	NA	NA

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Subgroup	2	0	0	11.2421147	0.0000959
agef2	2	0	0	2.7462266	0.0740565
Gender	1	0	0	0.1503477	0.6998827
Residuals	49	0	0	NA	NA

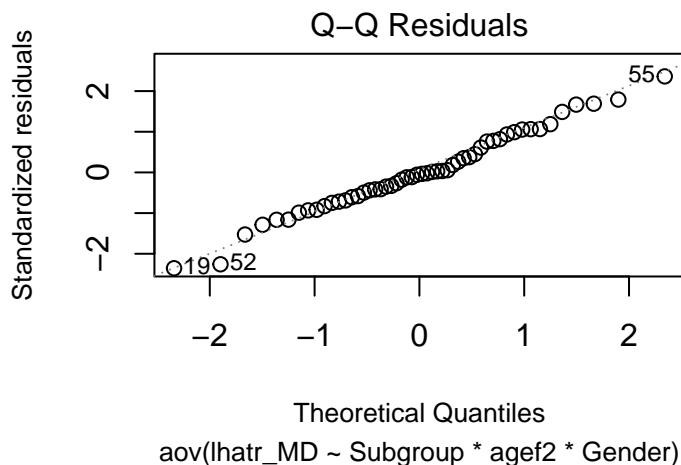
### 3.3.6 MD - LATR

Se procede a generar los modelos ANOVA. En el **primer modelo** generado `anova1_LATRavg_md`, se introducen las **3 variables independientes con interacción**.

```
anova1_LATRavg_md<-aov(lhatr_MD~Subgroup*agef2*Gender, data=ATR_avg_demo)
summary(anova1_LATRavg_md)

Df      Sum Sq   Mean Sq F value    Pr(>F)
Subgroup        2 1.149e-08 5.743e-09   7.489 0.00177 ***
agef2           2 1.635e-09 8.170e-10   1.066 0.35427
Gender          1 6.410e-10 6.410e-10   0.836 0.36612
Subgroup:agef2  4 1.002e-09 2.510e-10   0.327 0.85829
Subgroup:Gender 2 3.490e-10 1.750e-10   0.228 0.79748
agef2:Gender     2 7.720e-10 3.860e-10   0.503 0.60853
Subgroup:agef2:Gender 2 5.150e-10 2.570e-10   0.336 0.71702
Residuals       39 2.991e-08 7.670e-10

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
plot(anova1_LATRavg_md, which=2, cex.lab=0.8, cex.sub=0.8)
```



Theoretical Quantiles

`aov(lhatr_MD ~ Subgroup * agef2 * Gender)`

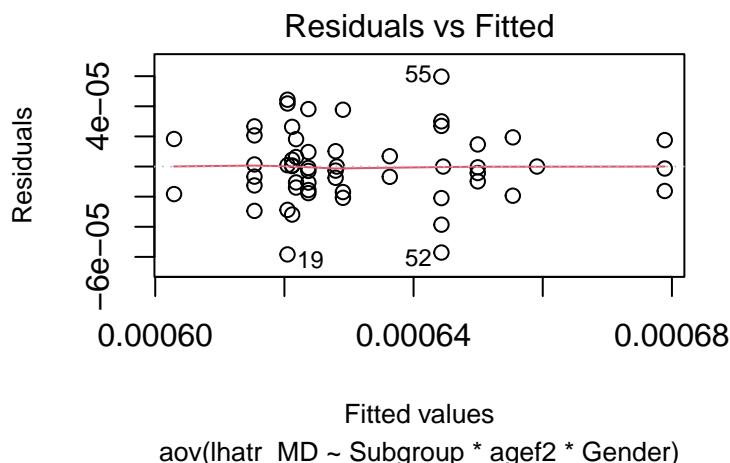
```

shapiro.test(residuals(anova1_LATRavg_md))

Shapiro-Wilk normality test

data: residuals(anova1_LATRavg_md)
W = 0.98201, p-value = 0.578
plot(anova1_LATRavg_md, which=1, cex.lab=0.8, cex.sub=0.8)

```



```
bptest(anova1_LATRavg_md)
```

```
studentized Breusch-Pagan test
```

```

data: anova1_LATRavg_md
BP = 26.798, df = 15, p-value = 0.03042

```

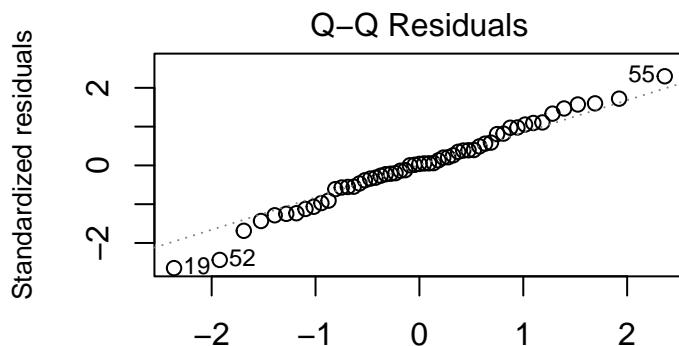
En este modelo, la variable subgrupo es significativa y los residuos presentan una distribución normal y heterocedasticidad de varianzas. Los posibles sujetos outliers e influyentes son los individuos con índices 19, 52 y 55.

Se genera un **segundo modelo** anova2\_LATRavg\_md con las **3 variables independientes sin interacción** dado que las interacciones no han sido significativas.

```

anova2_LATRavg_md<-aov(lhatr_MD~Subgroup+agef2+Gender, data=ATR_avg_demo)
summary(anova2_LATRavg_md)
      Df   Sum Sq  Mean Sq F value    Pr(>F)
Subgroup     2 1.149e-08 5.743e-09  8.647 0.000608 ***
agef2        2 1.630e-09 8.170e-10  1.231 0.300994
Gender        1 6.400e-10 6.410e-10  0.965 0.330656
Residuals    49 3.255e-08 6.640e-10
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
plot(anova2_LATRavg_md, which=2, cex.lab=0.8, cex.sub=0.8)

```

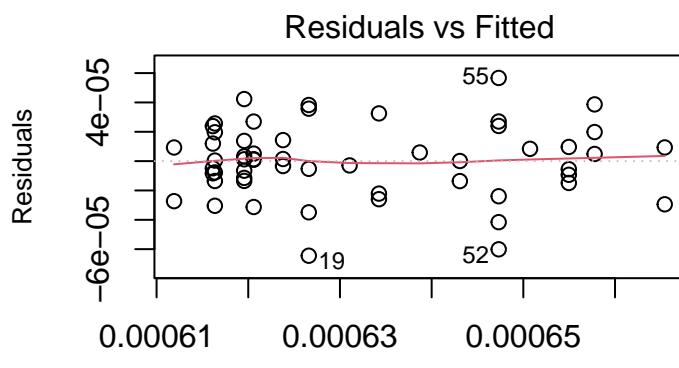


aov(lhatr\_MD ~ Subgroup + agef2 + Gender)

```
shapiro.test(residuals(anova2_LATRavg_md))
```

Shapiro-Wilk normality test

```
data: residuals(anova2_LATRavg_md)
W = 0.98685, p-value = 0.8075
plot(anova2_LATRavg_md, which=1, cex.lab=0.8, cex.sub=0.8)
```



aov(lhatr\_MD ~ Subgroup + agef2 + Gender)

```
bptest(anova2_LATRavg_md)
```

studentized Breusch-Pagan test

```
data: anova2_LATRavg_md
BP = 10.182, df = 5, p-value = 0.07025
```

En este modelo, la variable subgrupo es significativa y los residuos presentan una distribución normal y homocedasticidad de varianzas. Los posibles sujetos outliers e influyentes son los individuos con índices 19, 52 y 55.

Cuadro 52: ANOVA Models for MD of the Left ATR

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Subgroup	2	0	0	7.4888898	0.0017683
agef2	2	0	0	1.0658154	0.3542658
Gender	1	0	0	0.8361494	0.3661195
Subgroup:agef2	4	0	0	0.3267733	0.8582921
Subgroup:Gender	2	0	0	0.2276172	0.7974796
agef2:Gender	2	0	0	0.5030925	0.6085281
Subgroup:agef2:Gender	2	0	0	0.3355084	0.7170177
Residuals	39	0	0	NA	NA

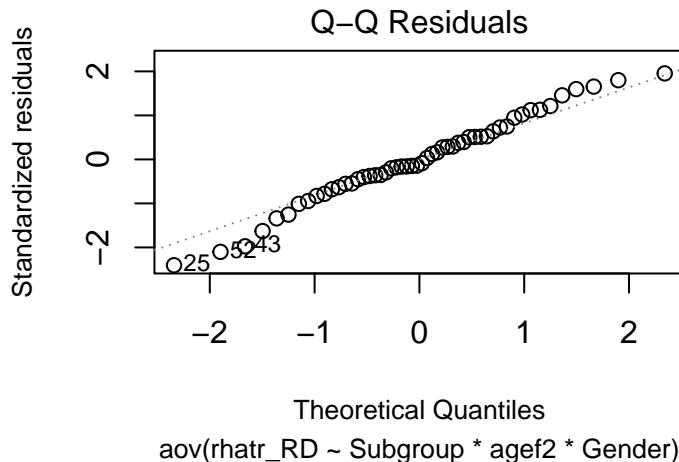
  

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Subgroup	2	0	0	8.6465520	0.0006079
agef2	2	0	0	1.2305734	0.3009936
Gender	1	0	0	0.9654047	0.3306557
Residuals	49	0	0	NA	NA

### 3.3.7 RD - RATR

Se procede a generar los modelos ANOVA. En el **primer modelo** generado `anova1_RATRavg_rd`, se introducen las **3 variables independientes con interacción**.

```
anova1_RATRavg_rd<-aov(rhatr_RD~Subgroup*agef2*Gender, data=ATR_avg_demo)
summary(anova1_RATRavg_rd)
Df Sum Sq Mean Sq F value Pr(>F)
Subgroup 2 1.480e-08 7.400e-09 8.078 0.00116 **
agef2 2 3.740e-09 1.872e-09 2.043 0.14328
Gender 1 8.000e-11 7.700e-11 0.084 0.77303
Subgroup:agef2 4 1.890e-09 4.740e-10 0.517 0.72366
Subgroup:Gender 2 5.700e-10 2.840e-10 0.310 0.73510
agef2:Gender 2 7.200e-10 3.580e-10 0.391 0.67914
Subgroup:agef2:Gender 2 7.500e-10 3.770e-10 0.412 0.66518
Residuals 39 3.573e-08 9.160e-10
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
plot(anova1_RATRavg_rd, which=2, cex.lab=0.8, cex.sub=0.8)
```



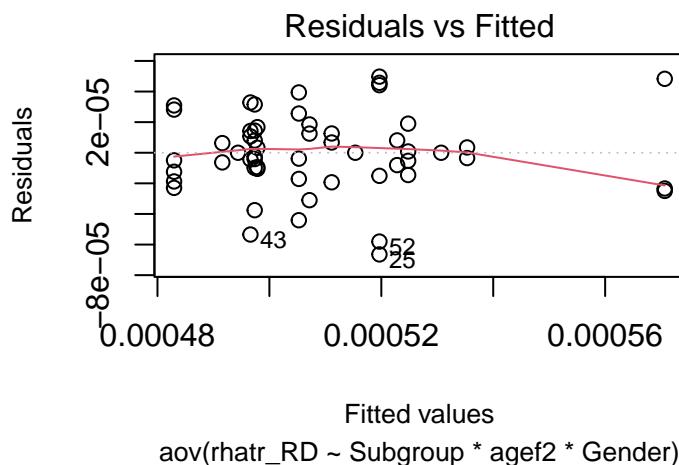
```

shapiro.test(residuals(anova1_RATRavg_rd))

Shapiro-Wilk normality test

data: residuals(anova1_RATRavg_rd)
W = 0.97754, p-value = 0.3901
plot(anova1_RATRavg_rd, which=1, cex.lab=0.8, cex.sub=0.8)

```



```

bptest(anova1_RATRavg_rd)

studentized Breusch-Pagan test

data: anova1_RATRavg_rd
BP = 29.375, df = 15, p-value = 0.01439

```

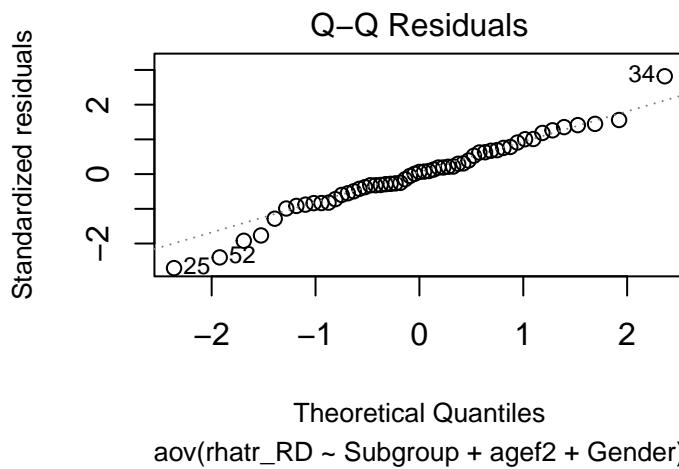
En este modelo, la variable subgrupo es significativa y los residuos presentan una distribución normal y heterocedasticidad de varianzas. Los posibles sujetos outliers e influyentes son los individuos con índices 25, 43 y 52.

Se genera un **segundo modelo** anova2\_RATRavg\_rd con las **3 variables independientes sin interacción** dado que las interacciones no han sido significativas.

```

anova2_RATRavg_rd<-aov(rhatr_RD~Subgroup+agef2+Gender, data=ATR_avg_demo)
summary(anova2_RATRavg_rd)
      Df   Sum Sq  Mean Sq F value    Pr(>F)
Subgroup     2 1.480e-08 7.400e-09  9.142 0.000423 ***
agef2        2 3.740e-09 1.872e-09  2.312 0.109743
Gender        1 8.000e-11 7.700e-11  0.095 0.758653
Residuals    49 3.966e-08 8.090e-10
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
plot(anova2_RATRavg_rd, which=2, cex.lab=0.8, cex.sub=0.8)

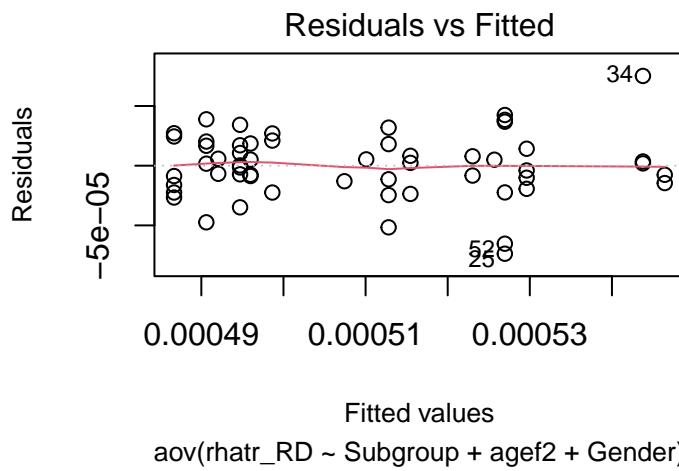
```



```
shapiro.test(residuals(anova2_RATRavg_rd))
```

```
Shapiro-Wilk normality test

data: residuals(anova2_RATRavg_rd)
W = 0.97859, p-value = 0.4301
plot(anova2_RATRavg_rd, which=1, cex.lab=0.8, cex.sub=0.8)
```



```
bptest(anova2_RATRavg_rd)
```

```
studentized Breusch-Pagan test

data: anova2_RATRavg_rd
BP = 9.7022, df = 5, p-value = 0.08413
```

En este modelo, la variable subgrupo es significativa y los residuos presentan una distribución normal y homocedasticidad de varianzas. Los posibles sujetos outliers e influyentes son los individuos con índices 25, 34 y 52.

Cuadro 53: ANOVA Models for RD of the Right ATR

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Subgroup	2	0	0	8.0775972	0.0011610
agef2	2	0	0	2.0430487	0.1432799
Gender	1	0	0	0.0843468	0.7730297
Subgroup:agef2	4	0	0	0.5169837	0.7236630
Subgroup:Gender	2	0	0	0.3101850	0.7351036
agef2:Gender	2	0	0	0.3907946	0.6791385
Subgroup:agef2:Gender	2	0	0	0.4119868	0.6651817
Residuals	39	0	0	NA	NA

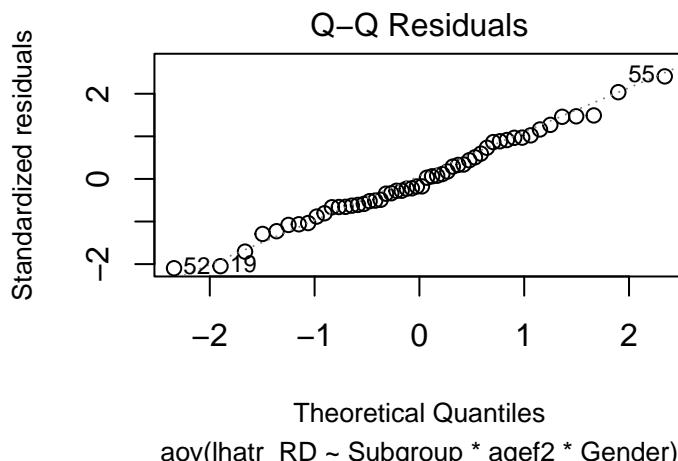
  

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Subgroup	2	0	0	9.1422246	0.0004226
agef2	2	0	0	2.3123226	0.1097426
Gender	1	0	0	0.0954637	0.7586528
Residuals	49	0	0	NA	NA

### 3.3.8 RD - LATR

Se procede a generar los modelos ANOVA. En el **primer modelo** generado `anova1_LATRavg_rd`, se introducen las **3 variables independientes con interacción**.

```
anova1_LATRavg_rd<-aov(lhatr_RD~Subgroup*agef2*Gender, data=ATR_avg_demo)
summary(anova1_LATRavg_rd)
Df      Sum Sq   Mean Sq F value    Pr(>F)
Subgroup        2 1.042e-08 5.212e-09   6.903 0.00271 ***
agef2           2 2.362e-09 1.181e-09   1.564 0.22206
Gender          1 1.980e-10 1.980e-10   0.263 0.61124
Subgroup:agef2  4 5.570e-10 1.390e-10   0.185 0.94502
Subgroup:Gender 2 1.224e-09 6.120e-10   0.811 0.45190
agef2:Gender     2 3.220e-10 1.610e-10   0.213 0.80899
Subgroup:agef2:Gender 2 7.770e-10 3.880e-10   0.515 0.60180
Residuals       39 2.944e-08 7.550e-10
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
plot(anova1_LATRavg_rd, which=2, cex.lab=0.8, cex.sub=0.8)
```



Q-Q Residuals

aov(lhatr\_RD ~ Subgroup \* agef2 \* Gender)

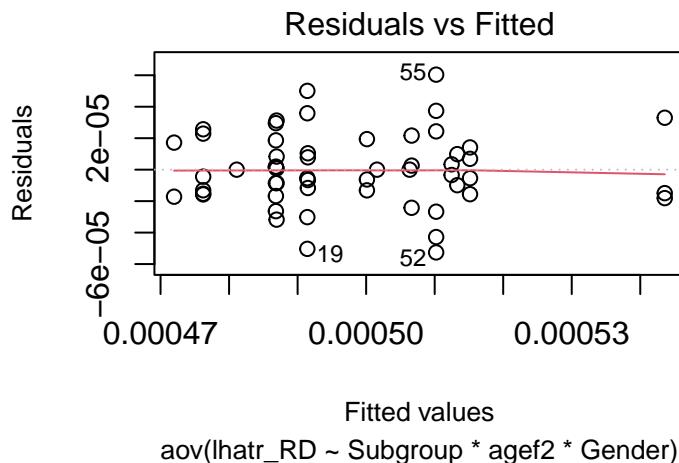
```

shapiro.test(residuals(anova1_LATRavg_rd))

Shapiro-Wilk normality test

data: residuals(anova1_LATRavg_rd)
W = 0.9883, p-value = 0.8684
plot(anova1_LATRavg_rd, which=1, cex.lab=0.8, cex.sub=0.8)

```



```

bptest(anova1_LATRavg_rd)

studentized Breusch-Pagan test

data: anova1_LATRavg_rd
BP = 30.899, df = 15, p-value = 0.009062

```

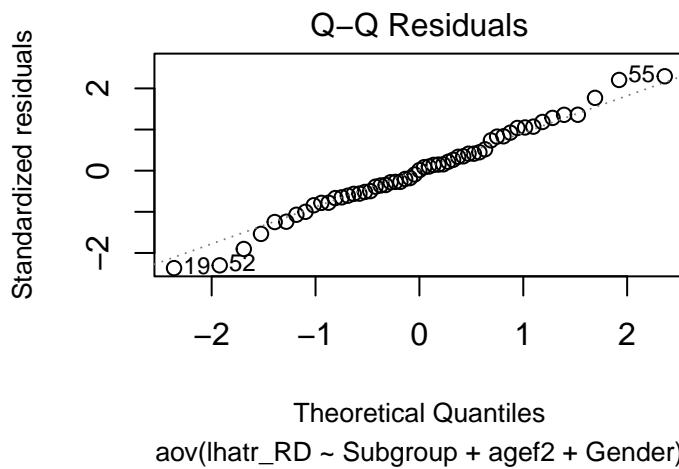
En este modelo, la variable subgrupo es significativa y los residuos presentan una distribución normal y heterocedasticidad de varianzas. Los posibles sujetos outliers e influyentes son los individuos con índices 19, 52 y 55.

Se genera un **segundo modelo** anova2\_LATRavg\_rd con las **3 variables independientes sin interacción** dado que las interacciones no han sido significativas.

```

anova2_LATRavg_rd<-aov(lhatr_RD~Subgroup+agef2+Gender, data=ATR_avg_demo)
summary(anova2_LATRavg_rd)
  Df   Sum Sq  Mean Sq F value    Pr(>F)
Subgroup     2 1.042e-08 5.212e-09  7.900 0.00106 ***
agef2        2 2.360e-09 1.181e-09  1.790 0.17764
Gender       1 2.000e-10 1.980e-10  0.301 0.58604
Residuals    49 3.232e-08 6.600e-10
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
plot(anova2_LATRavg_rd, which=2, cex.lab=0.8, cex.sub=0.8)

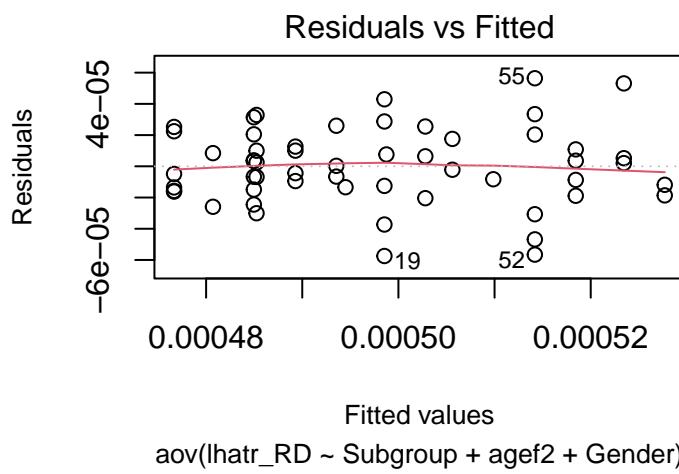
```



```
shapiro.test(residuals(anova2_LATRavg_rd))
```

Shapiro-Wilk normality test

```
data: residuals(anova2_LATRavg_rd)
W = 0.98815, p-value = 0.8627
plot(anova2_LATRavg_rd, which=1, cex.lab=0.8, cex.sub=0.8)
```



```
bptest(anova2_LATRavg_rd)
```

studentized Breusch-Pagan test

```
data: anova2_LATRavg_rd
BP = 12.242, df = 5, p-value = 0.03161
```

En este modelo, la variable subgrupo es significativa y los residuos presentan una distribución normal y heterocedasticidad de varianzas. Los posibles sujetos outliers e influyentes son los individuos con índices 19, 52 y 55.

Cuadro 54: ANOVA Models for RD of the Left ATR

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Subgroup	2	0	0	6.9032104	0.0027125
agef2	2	0	0	1.5643908	0.2220605
Gender	1	0	0	0.2625884	0.6112375
Subgroup:agef2	4	0	0	0.1846018	0.9450232
Subgroup:Gender	2	0	0	0.8106996	0.4518973
agef2:Gender	2	0	0	0.2131300	0.8089860
Subgroup:agef2:Gender	2	0	0	0.5145059	0.6017968
Residuals	39	0	0	NA	NA

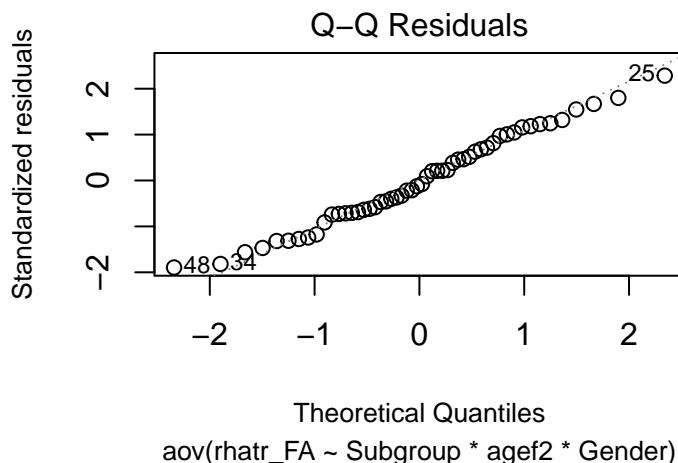
  

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Subgroup	2	0	0	7.9004249	0.0010619
agef2	2	0	0	1.7903775	0.1776425
Gender	1	0	0	0.3005211	0.5860449
Residuals	49	0	0	NA	NA

### 3.3.9 FA - RATR

Se procede a generar los modelos ANOVA. En el **primer modelo** generado `anova1_RATRavg_fa`, se introducen las **3 variables independientes con interacción**.

```
anova1_RATRavg_fa<-aov(rhatr_FA~Subgroup*agef2*Gender, data=ATR_avg_demo)
summary(anova1_RATRavg_fa)
Df Sum Sq Mean Sq F value Pr(>F)
Subgroup 2 0.00184 0.0009207 1.131 0.333
agef2 2 0.00040 0.0002009 0.247 0.783
Gender 1 0.00001 0.0000086 0.011 0.919
Subgroup:agef2 4 0.00115 0.0002884 0.354 0.840
Subgroup:Gender 2 0.00045 0.0002238 0.275 0.761
agef2:Gender 2 0.00114 0.0005714 0.702 0.502
Subgroup:agef2:Gender 2 0.00191 0.0009570 1.175 0.319
Residuals 39 0.03175 0.0008142
plot(anova1_RATRavg_fa, which=2, cex.lab=0.8, cex.sub=0.8)
```



Q-Q Residuals

aov(rhatr\_FA ~ Subgroup \* agef2 \* Gender)

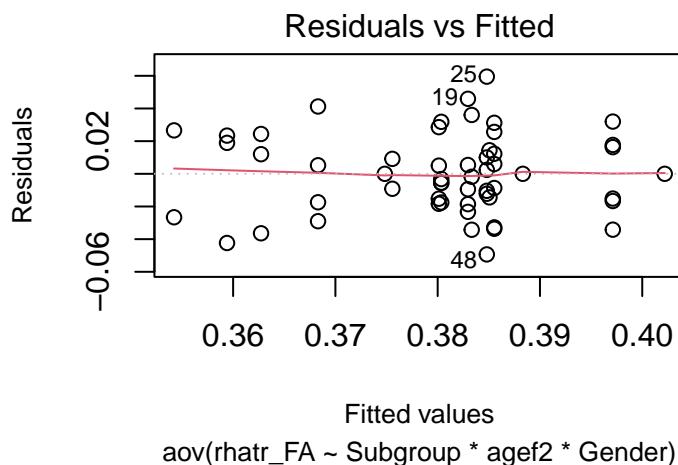
```

shapiro.test(residuals(anova1_RATRavg_fa))

Shapiro-Wilk normality test

data: residuals(anova1_RATRavg_fa)
W = 0.98943, p-value = 0.9095
plot(anova1_RATRavg_fa, which=1, cex.lab=0.8, cex.sub=0.8)

```



```

bptest(anova1_RATRavg_fa)

studentized Breusch-Pagan test

data: anova1_RATRavg_fa
BP = 8.9935, df = 15, p-value = 0.8779

```

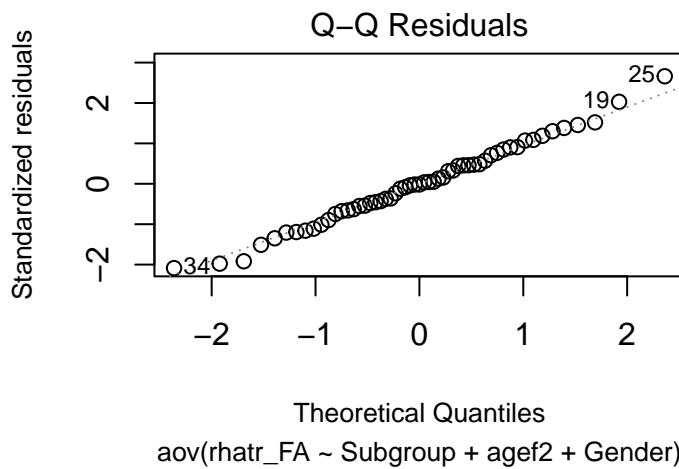
En este modelo, ningún factor o interacción es significativo y los residuos presentan una distribución normal y homocedasticidad de varianzas. Los posibles sujetos outliers son los individuos con índices 25, 34 y 48, y, los posibles sujetos influyentes son los individuos con índices 19, 25 y 48.

Se genera un **segundo modelo** `anova2_RATRavg_fa` con las **3 variables independientes sin interacción** dado que las interacciones no han sido significativas.

```

anova2_RATRavg_fa<-aov(rhatr_FA~Subgroup+agef2+Gender, data=ATR_avg_demo)
summary(anova2_RATRavg_fa)
   Df  Sum Sq  Mean Sq F value Pr(>F)
Subgroup     2 0.00184 0.0009207   1.239  0.299
agef2        2 0.00040 0.0002009   0.270  0.764
Gender       1 0.00001 0.0000086   0.012  0.915
Residuals    49 0.03641 0.0007431
plot(anova2_RATRavg_fa, which=2, cex.lab=0.8, cex.sub=0.8)

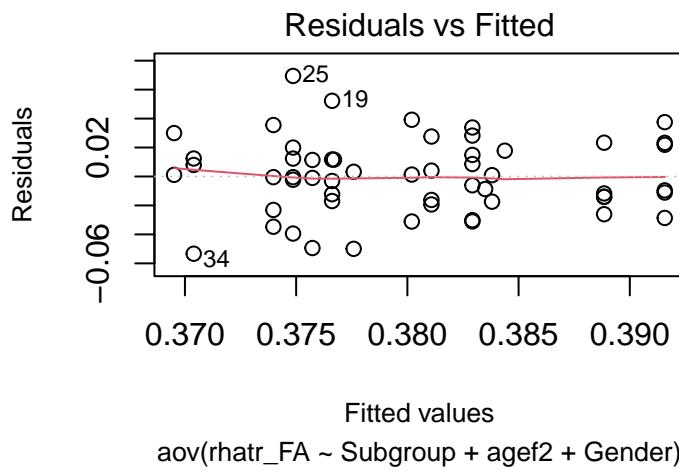
```



```
shapiro.test(residuals(anova2_RATRavg_fa))
```

```
Shapiro-Wilk normality test

data: residuals(anova2_RATRavg_fa)
W = 0.9921, p-value = 0.9751
plot(anova2_RATRavg_fa, which=1, cex.lab=0.8, cex.sub=0.8)
```



```
bptest(anova2_RATRavg_fa)

studentized Breusch-Pagan test

data: anova2_RATRavg_fa
BP = 2.5484, df = 5, p-value = 0.7692
```

En este modelo, ningún factor es significativo y los residuos presentan una distribución normal y homocedasticidad de varianzas. Los posibles sujetos outliers e influyentes son los individuos con índices 19, 25 y 34.

Cuadro 55: ANOVA Models for FA of the Right ATR

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Subgroup	2	0.0018415	0.0009207	1.1308736	0.3331052
agef2	2	0.0004018	0.0002009	0.2467685	0.7825322
Gender	1	0.0000086	0.0000086	0.0105047	0.9188908
Subgroup:agef2	4	0.0011534	0.0002884	0.3541556	0.8395540
Subgroup:Gender	2	0.0004475	0.0002238	0.2748370	0.7611550
agef2:Gender	2	0.0011428	0.0005714	0.7018185	0.5018349
Subgroup:agef2:Gender	2	0.0019141	0.0009570	1.1754571	0.3193744
Residuals	39	0.0317535	0.0008142	NA	NA

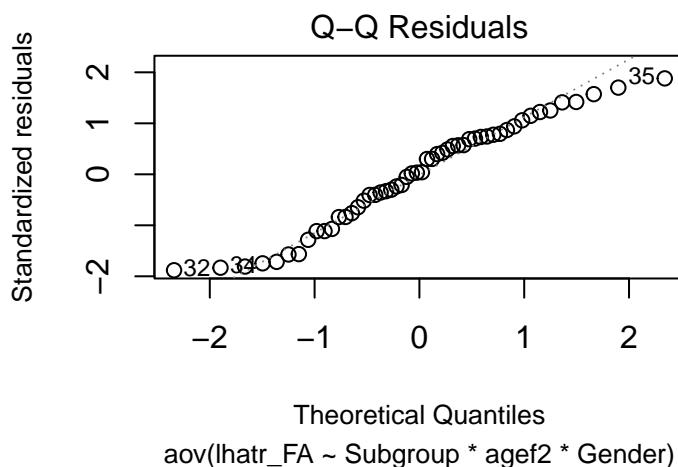
  

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Subgroup	2	0.0018415	0.0009207	1.2390822	0.2985652
agef2	2	0.0004018	0.0002009	0.2703808	0.7642199
Gender	1	0.0000086	0.0000086	0.0115099	0.9150016
Residuals	49	0.0364113	0.0007431	NA	NA

### 3.3.10 FA - LATR

Se procede a generar los modelos ANOVA. En el **primer modelo** generado `anova1_LATRavg_fa`, se introducen las **3 variables independientes con interacción**.

```
anova1_LATRavg_fa<-aov(lhatr_FA~Subgroup*agef2*Gender, data=ATR_avg_demo)
summary(anova1_LATRavg_fa)
Df Sum Sq Mean Sq F value Pr(>F)
Subgroup 2 0.003137 0.0015685 2.736 0.0773 .
agef2 2 0.001229 0.0006144 1.072 0.3523
Gender 1 0.000067 0.0000672 0.117 0.7340
Subgroup:agef2 4 0.001394 0.0003486 0.608 0.6593
Subgroup:Gender 2 0.001830 0.0009151 1.596 0.2157
agef2:Gender 2 0.002433 0.0012163 2.121 0.1335
Subgroup:agef2:Gender 2 0.001256 0.0006281 1.095 0.3445
Residuals 39 0.022362 0.0005734
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
plot(anova1_LATRavg_fa, which=2, cex.lab=0.8, cex.sub=0.8)
```



Theoretical Quantiles

aov(lhatr\_FA ~ Subgroup \* agef2 \* Gender)

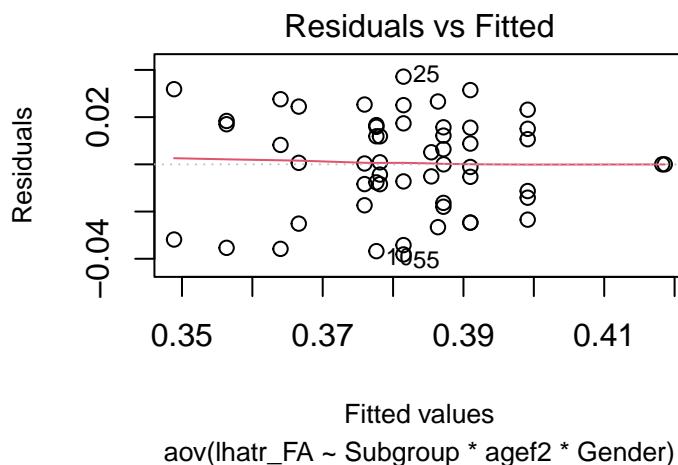
```

shapiro.test(residuals(anova1_LATRavg_fa))

Shapiro-Wilk normality test

data: residuals(anova1_LATRavg_fa)
W = 0.9639, p-value = 0.09731
plot(anova1_LATRavg_fa, which=1, cex.lab=0.8, cex.sub=0.8)

```



```

bptest(anova1_LATRavg_fa)

studentized Breusch-Pagan test

data: anova1_LATRavg_fa
BP = 22.679, df = 15, p-value = 0.0912

```

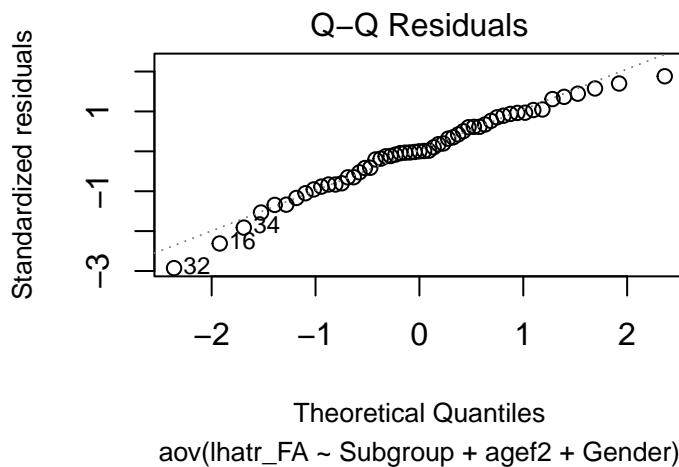
En este modelo, ningún factor o interacción es significativo y los residuos presentan una distribución normal y homocedasticidad de varianzas. Los posibles sujetos outliers son los individuos con índices 32, 34 y 35, y, los sujetos influyentes los individuos con índices 10, 25 y 55.

Se genera un **segundo modelo** `anova2_LATRavg_fa` con las **3 variables independientes sin interacción** dado que las interacciones no han sido significativas.

```

anova2_LATRavg_fa<-aov(lhatr_FA~Subgroup+agef2+Gender, data=ATR_avg_demo)
summary(anova2_LATRavg_fa)
      Df   Sum Sq   Mean Sq F value Pr(>F)
Subgroup     2 0.003137 0.0015685   2.625 0.0826 .
agef2        2 0.001229 0.0006144   1.028 0.3652
Gender        1 0.0000067 0.00000672  0.112 0.7389
Residuals    49 0.029275 0.0005974
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
plot(anova2_LATRavg_fa, which=2, cex.lab=0.8, cex.sub=0.8)

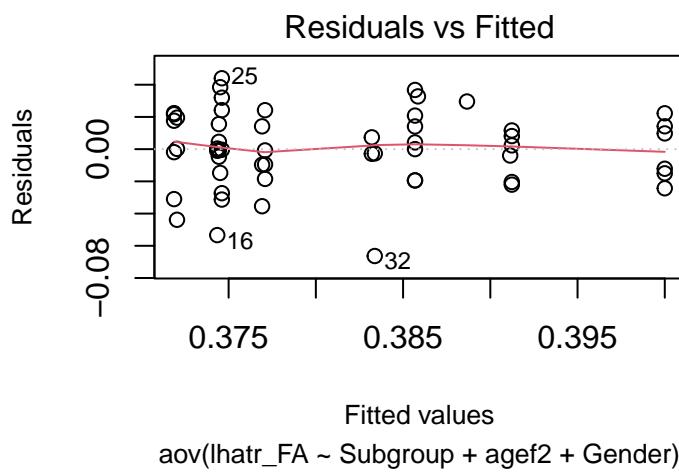
```



```
shapiro.test(residuals(anova2_LATRavg_fa))
```

Shapiro-Wilk normality test

```
data: residuals(anova2_LATRavg_fa)
W = 0.97951, p-value = 0.4674
plot(anova2_LATRavg_fa, which=1, cex.lab=0.8, cex.sub=0.8)
```



```
bptest(anova2_LATRavg_fa)
```

studentized Breusch-Pagan test

```
data: anova2_LATRavg_fa
BP = 7.4363, df = 5, p-value = 0.1902
```

En este modelo, ningún factor es significativo y los residuos presentan una distribución normal y homocedasticidad de varianzas. Los posibles sujetos outliers son los individuos con índices 16, 32 y 34, y los posibles sujetos influyentes son los individuos con índices 16, 25 y 32.

Cuadro 56: ANOVA Models for FA of the Left ATR

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Subgroup	2	0.0031370	0.0015685	2.7355611	0.0773103
agef2	2	0.0012288	0.0006144	1.0715942	0.3523302
Gender	1	0.0000672	0.0000672	0.1171237	0.7340128
Subgroup:agef2	4	0.0013942	0.0003486	0.6079034	0.6593333
Subgroup:Gender	2	0.0018302	0.0009151	1.5960035	0.2156608
agef2:Gender	2	0.0024325	0.0012163	2.1212480	0.1335058
Subgroup:agef2:Gender	2	0.0012561	0.0006281	1.0953971	0.3444741
Residuals	39	0.0223616	0.0005734	NA	NA

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Subgroup	2	0.0031370	0.0015685	2.6253546	0.0825793
agef2	2	0.0012288	0.0006144	1.0284233	0.3651584
Gender	1	0.0000672	0.0000672	0.1124052	0.7388535
Residuals	49	0.0292747	0.0005974	NA	NA

En resumen, los modelos anova de interés están formados por las variables independientes sin interacción.

Se presenta una **tabla resumen** para objetivar la significancia o no de la variable subgrupo, de la normalidad y homocedasticidad/heterocedasticidad de los residuos.

Cuadro 57: Modelos Anova para valores Average

ATR-Metrics	Subgroup		Normality		Homoscedasticity		
	F	p-value	W statistic	p-value	BP statistic	df	p-value
<b>Iron</b>							
RATR	2.352	0.106	0.918	0.001	5.467	5	0.362
LATR	0.483	0.62	0.989	0.88	6.048	5	0.302
<b>AD</b>							
RATR	7.526	0.001	0.96	0.063	4.627	5	0.463
LATR	6.151	0.004	0.959	0.06	7.241	5	0.203
<b>MD</b>							
RATR	11.242	0	0.975	0.303	9.072	5	0.106
LATR	8.647	0.001	0.987	0.808	10.182	5	0.07
<b>RD</b>							
RATR	9.142	0	0.979	0.43	9.702	5	0.084
LATR	7.9	0.001	0.988	0.863	12.242	5	0.032
<b>FA</b>							
RATR	1.239	0.299	0.992	0.975	2.548	5	0.769
LATR	2.625	0.083	0.98	0.467	7.436	5	0.19

Como se observa en la tabla:

- La variable *rhatr\_iron* no tiene normalidad;
- La variable *lhatr\_rd* no tiene homocedasticidad.

### 3.3.11 Eliminación de outliers - Nuevos ANOVAS

Se procede a la eliminación de outliers en dichas variables para poder generar nuevos modelos anova y así determinar si se cumplen los criterios requeridos por ANOVA.

A continuación, para la variable ***rhatr-iron***: primero, se procede a la **eliminación de outliers**.

```

# Extracción de outlier considerando todos los sujetos de la muestra en conjunto
ATR_avg_demo_out<-ATR_avg_demo
resumen<-data.frame()
resumen<-ATR_avg_demo_out |>
  dplyr::summarize(across(rhatr_Iron, list(iqr= ~ IQR(.), q1= ~quantile(.,probs = .25),
                            q3= ~quantile(.,probs = .75))))
index<-which(ATR_avg_demo_out$rhatr_Iron<(resumen[[2]]-1.5*resumen[[1]]) | 
              ATR_avg_demo_out$rhatr_Iron>(resumen[[3]]+1.5*resumen[[1]]))
index
[1] 51

ATR_avg_demo_out$rhatr_Iron<-replace(ATR_avg_demo_out$rhatr_Iron, index, NA)

# Normalidad y homocedasticidad entre subgrupos, géneros y edad factorizada
stats::shapiro.test(ATR_avg_demo_out$rhatr_Iron)

Shapiro-Wilk normality test

data: ATR_avg_demo_out$rhatr_Iron
W = 0.97216, p-value = 0.2397
fligner.test(rhatr_Iron~Subgroup, data=ATR_avg_demo_out)

Fligner-Killeen test of homogeneity of variances

data: rhatr_Iron by Subgroup
Fligner-Killeen:med chi-squared = 0.51254, df = 2, p-value = 0.7739
fligner.test(rhatr_Iron~agef, data=ATR_avg_demo_out)

Fligner-Killeen test of homogeneity of variances

data: rhatr_Iron by agef
Fligner-Killeen:med chi-squared = 7.31, df = 6, p-value = 0.2931
fligner.test(rhatr_Iron~agef2, data=ATR_avg_demo_out)

Fligner-Killeen test of homogeneity of variances

data: rhatr_Iron by agef2
Fligner-Killeen:med chi-squared = 1.6366, df = 2, p-value = 0.4412
fligner.test(rhatr_Iron~Gender, data=ATR_avg_demo_out)

Fligner-Killeen test of homogeneity of variances

data: rhatr_Iron by Gender
Fligner-Killeen:med chi-squared = 0.82933, df = 1, p-value = 0.3625

```

Con la eliminación del outlier, los valores de la variable *rhatr\_iron* presentan una distribución normal y homocedasticidad de varianzas entre los niveles de las variables categóricas de interés (subgrupo, edad\_factorizada, género).

Se generan nuevamente los modelos anova, el **primer modelo** *anova3\_RATRavg\_iron* con las **3 variables independientes en interacción** y el **segundo modelo** *anova4\_RATRavg\_iron* con las **3 variables independientes sin interacción**.

```

anova3_RATRavg_iron<-aov(rhatr_Iron~Subgroup*agef2*Gender, data=ATR_avg_demo_out)
summary(anova3_RATRavg_iron)

```

```

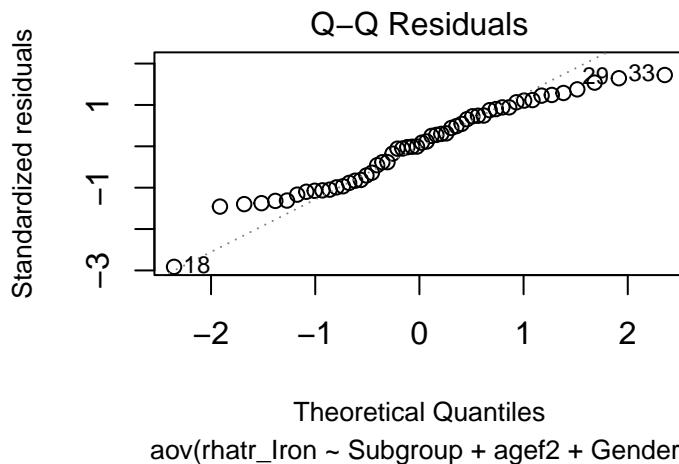
Df      Sum Sq   Mean Sq F value Pr(>F)
Subgroup        2 3.240e-06 1.622e-06   1.589 0.2175
agef2           2 1.440e-06 7.219e-07   0.707 0.4994
Gender          1 7.700e-07 7.652e-07   0.750 0.3920
Subgroup:agef2  4 7.690e-06 1.923e-06   1.884 0.1333
Subgroup:Gender 2 2.760e-06 1.381e-06   1.353 0.2705
agef2:Gender    2 5.230e-06 2.614e-06   2.561 0.0905 .
Subgroup:agef2:Gender 2 5.420e-06 2.710e-06   2.655 0.0833 .
Residuals       38 3.879e-05 1.021e-06

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
1 observation deleted due to missingness

anova4_RATRavg_iron<-aov(rhatr_Iron~Subgroup+agef2+Gender, data=ATR_avg_demo_out)
summary(anova4_RATRavg_iron)
Df      Sum Sq   Mean Sq F value Pr(>F)
Subgroup        2 3.240e-06 1.622e-06   1.300 0.282
agef2           2 1.440e-06 7.219e-07   0.579 0.565
Gender          1 7.700e-07 7.652e-07   0.613 0.437
Residuals       48 5.989e-05 1.248e-06

1 observation deleted due to missingness
plot(anova4_RATRavg_iron, which=2, cex.lab=0.8, cex.sub=0.8)

```



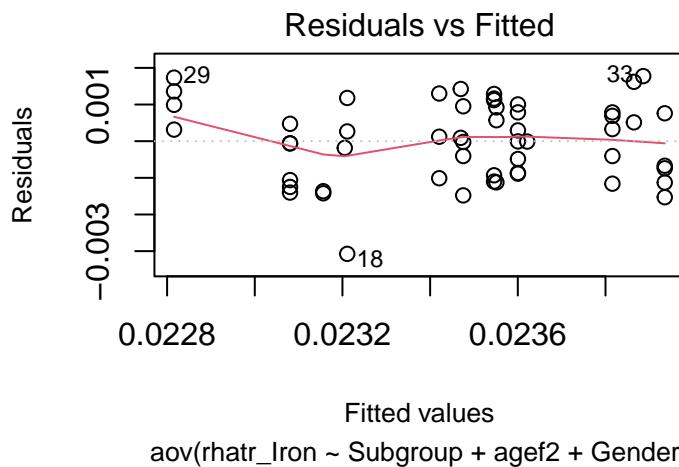
```

shapiro.test(residuals(anova4_RATRavg_iron))

Shapiro-Wilk normality test

data:  residuals(anova4_RATRavg_iron)
W = 0.96134, p-value = 0.07915
plot(anova4_RATRavg_iron, which=1, cex.lab=0.8, cex.sub=0.8)

```



```
bptest(anova4_RATRavg_iron)

studentized Breusch-Pagan test

data: anova4_RATRavg_iron
BP = 3.3051, df = 5, p-value = 0.6531
```

Como el modelo anova con interacción no presenta ningún factor o interacción significativo, se genera el modelo sin interacción y éste tampoco muestra ningún factor significativo, pero los residuos siguen una distribución normal con homocedasticidad de varianzas.

Cuadro 58: New ANOVA Models for Iron of the Right ATR after elimination of outlier (index 51)

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Subgroup	2	3.20e-06	1.6e-06	1.5885890	0.2174770
agef2	2	1.40e-06	7.0e-07	0.7072293	0.4993817
Gender	1	8.00e-07	8.0e-07	0.7496039	0.3920353
Subgroup:agef2	4	7.70e-06	1.9e-06	1.8836918	0.1332868
Subgroup:Gender	2	2.80e-06	1.4e-06	1.3533340	0.2705470
agef2:Gender	2	5.20e-06	2.6e-06	2.5612844	0.0904691
Subgroup:agef2:Gender	2	5.40e-06	2.7e-06	2.6549667	0.0833154
Residuals	38	3.88e-05	1.0e-06	NA	NA

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Subgroup	2	3.20e-06	1.6e-06	1.2995936	0.2820621
agef2	2	1.40e-06	7.0e-07	0.5785704	0.5645611
Gender	1	8.00e-07	8.0e-07	0.6132363	0.4374187
Residuals	48	5.99e-05	1.2e-06	NA	NA

A continuación, para la variable *lhatr-rd*: primero, se procede a la **eliminación de outliers**.

```
# Extracción de outlier considerando todos los sujetos de la muestra en conjunto
ATR_avg_demo_out<-ATR_avg_demo
resumen<-data.frame()
resumen<-ATR_avg_demo_out |>
  dplyr::summarize(across(lhatr_RD, list(iqr= ~ IQR(.),
                                         q1= ~quantile(.,probs = .25),
                                         q3= ~quantile(.,probs = .75))))
```

```

index<-which(ATR_avg_demo_out$lhatr_RD<(resumen[[2]]-1.5*resumen[[1]]) |
             ATR_avg_demo_out$lhatr_RD>(resumen[[3]]+1.5*resumen[[1]]))
index
[1] 34

ATR_avg_demo_out$lhatr_RD<-replace(ATR_avg_demo_out$lhatr_RD, index, NA)

# Normalidad y homocedasticidad entre subgrupos, géneros y edad factorizada
stats::shapiro.test(ATR_avg_demo_out$lhatr_RD)

  Shapiro-Wilk normality test

data: ATR_avg_demo_out$lhatr_RD
W = 0.98569, p-value = 0.7637
fligner.test(lhatr_RD~Subgroup, data=ATR_avg_demo_out)

  Fligner-Killeen test of homogeneity of variances

data: lhatr_RD by Subgroup
Fligner-Killeen:med chi-squared = 0.38014, df = 2, p-value = 0.8269
fligner.test(lhatr_RD~agef, data=ATR_avg_demo_out)

  Fligner-Killeen test of homogeneity of variances

data: lhatr_RD by agef
Fligner-Killeen:med chi-squared = 2.135, df = 6, p-value = 0.9069
fligner.test(lhatr_RD~agef2, data=ATR_avg_demo_out)

  Fligner-Killeen test of homogeneity of variances

data: lhatr_RD by agef2
Fligner-Killeen:med chi-squared = 0.71163, df = 2, p-value = 0.7006
fligner.test(lhatr_RD~Gender, data=ATR_avg_demo_out)

  Fligner-Killeen test of homogeneity of variances

data: lhatr_RD by Gender
Fligner-Killeen:med chi-squared = 4.3084, df = 1, p-value = 0.03792

```

Con la eliminación del outlier, los valores de la variable *lhatr\_rd* presentan una distribución normal y homocedasticidad de varianzas para subgrupo y edad factorizada pero no para género.

Se generan nuevamente los modelos anova, el **primer modelo anova3\_LATRavg\_rd** con las **3 variables independientes en interacción** y el **segundo modelo anova4\_LATRavg\_rd** con las **3 variables independientes sin interacción**.

```

anova3_LATRavg_rd<-aov(lhatr_RD~Subgroup*agef2*Gender, data=ATR_avg_demo_out)
summary(anova3_LATRavg_rd)

  Df   Sum Sq  Mean Sq F value Pr(>F)
Subgroup        2 7.867e-09 3.934e-09  5.375 0.0088 ***
agef2          2 1.574e-09 7.870e-10  1.075 0.3513
Gender          1 4.360e-10 4.360e-10  0.596 0.4448
Subgroup:agef2  4 2.010e-10 5.000e-11  0.069 0.9910
Subgroup:Gender 2 7.200e-10 3.600e-10  0.492 0.6151
agef2:Gender     2 5.600e-11 2.800e-11  0.038 0.9625
Subgroup:agef2:Gender 2 3.270e-10 1.640e-10  0.224 0.8007

```

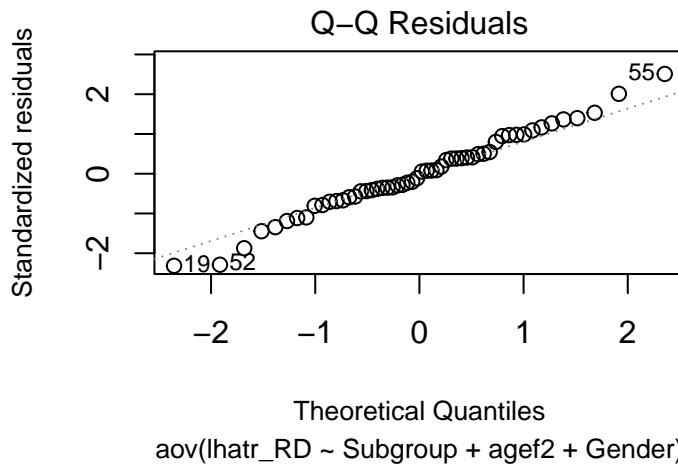
```

Residuals           38 2.781e-08 7.320e-10
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
1 observation deleted due to missingness

anova4_LATRavg_rd<-aov(lhatr_RD~Subgroup+agef2+Gender, data=ATR_avg_demo_out)
summary(anova4_LATRavg_rd)
  Df   Sum Sq  Mean Sq F value    Pr(>F)
Subgroup     2 7.867e-09 3.934e-09   6.485 0.00321 ** 
agef2        2 1.574e-09 7.870e-10   1.297 0.28264  
Gender       1 4.360e-10 4.360e-10   0.719 0.40061  
Residuals    48 2.912e-08 6.070e-10

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
1 observation deleted due to missingness
plot(anova4_LATRavg_rd, which=2, cex.lab=0.8, cex.sub=0.8)

```



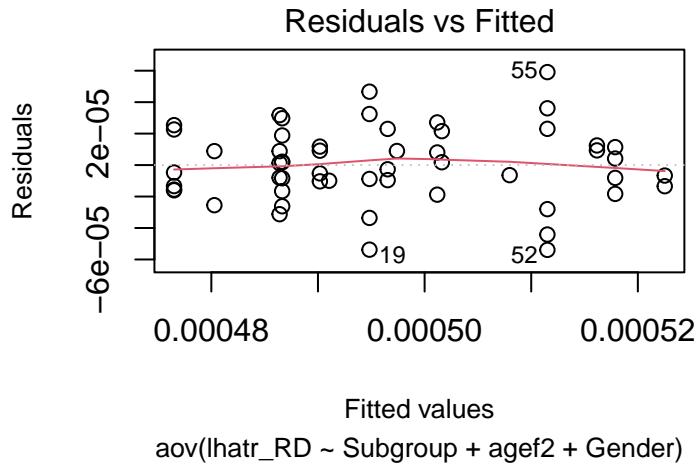
```

shapiro.test(residuals(anova4_LATRavg_rd))

Shapiro-Wilk normality test

data:  residuals(anova4_LATRavg_rd)
W = 0.98889, p-value = 0.8963
plot(anova4_LATRavg_rd, which=1, cex.lab=0.8, cex.sub=0.8)

```



```
bptest(anova4_LATRavg_rd)
```

```
studentized Breusch-Pagan test
```

```
data: anova4_LATRavg_rd
BP = 10.895, df = 5, p-value = 0.05349
```

Como el modelo con interacción no presenta ninguna interacción significativa, se genera el modelo sin interacción y éste muestra el subgrupo como variable significativa y los residuos del modelo siguen una distribución normal con homocedasticidad de varianzas.

Cuadro 59: New ANOVA Models for Radial Diffusivity of the Left ATR after elimination of outlier (index 34)

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Subgroup	2	0	0	5.3745727	0.0088007
agef2	2	0	0	1.0753247	0.3513415
Gender	1	0	0	0.5960893	0.4448493
Subgroup:agef2	4	0	0	0.0687222	0.9910068
Subgroup:Gender	2	0	0	0.4921790	0.6151358
agef2:Gender	2	0	0	0.0383108	0.9624509
Subgroup:agef2:Gender	2	0	0	0.2235747	0.8006996
Residuals	38	0	0	NA	NA
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Subgroup	2	0	0	6.4846636	0.0032145
agef2	2	0	0	1.2974276	0.2826423
Gender	1	0	0	0.7192086	0.4006129
Residuals	48	0	0	NA	NA

Cuadro 60: Nuevos Modelos Anova para valores Average tras eliminación de outliers

ATR-Metrics	Subgroup		Normality		Homoscedasticity		
	F	p-value	W statistic	p-value	BP statistic	df	p-value
Iron-RATR	1.3	0.282	0.961	0.079	3.305	5	0.653
RD-LATR	6.485	0.003	0.989	0.896	10.895	5	0.053

### 3.3.12 Comparaciones pareadas por subgrupos

Con las funciones `HSD.test` y `TukeyHSD` de los paquetes `agricolae` y `stats` respectivamente, se generan comparaciones pareadas por subgrupos (post-hoc Tukey tests) en los modelos ANOVA (`anova2_RATRavg_ad`, `anova2_LATRavg_ad`, `anova2_RATRavg_md`, `anova2_LATRavg_md`, `anova2_RATRavg_rd` y `anova4_LATRavg_rd`) que han detectado significancia entre la variable dependiente (métrica de RM) y la variable independiente subgrupo (las dos covariables no han mostrado significación).

Fórmula del Tukey HSD test:

$$\frac{\max(\bar{y}_i - \mu_i) - \min(\bar{y}_j - \mu_j)}{\sqrt{\frac{\hat{S}_E^2}{n}}} = q_{t, N-t}$$

$$HSD = q_{t, N-t}(\alpha) * \sqrt{\frac{\hat{S}_E^2}{n}}$$

$|\bar{y}_i - \bar{y}_j| > HSD \rightarrow$  stastitical significant difference, being

N = total number of observations

t = number of levels of the factor

n = sample size of each level of the factor

$\hat{S}_E^2$  = estimate of the error or residual variance

$\bar{y}_i, \bar{y}_j$  = sample means of the levels/populations i and j

$q_{t, N-t}(\alpha)$  = distribution of the studentized range with the following parameters: t groups, N-t degrees of freedom and  $\alpha$  level of significance

$$n_h = \frac{2}{\sum_{i=1}^2 \frac{1}{n_i}}$$

Cuadro 61: HSD Tukey test para las métricas AD, MD, RD de ambos ATR en relación a la variable subgrupo

	Mean	Groups	Comparison	Difference of means	Adjusted p-value
<b>AD-RATR</b>					
HD	0.0009482	a	PreHD-Control	-1.20e-06	0.9960133
Control	0.0009047	b	HD-Control	4.35e-05	0.0025019
PreHD	0.0009035	b	HD-PreHD	4.47e-05	0.0084889
<b>AD-LATR</b>					
HD	0.0009235	a	PreHD-Control	1.28e-05	0.5712423
PreHD	0.0008964	ab	HD-Control	3.99e-05	0.0029315
Control	0.0008836	b	HD-PreHD	2.71e-05	0.1162008
<b>MD-RATR</b>					
HD	0.0006707	a	PreHD-Control	-7.00e-06	0.7424743
Control	0.0006359	b	HD-Control	3.48e-05	0.0005589
PreHD	0.0006289	b	HD-PreHD	4.18e-05	0.0003815
<b>MD-LATR</b>					
HD	0.0006525	a	PreHD-Control	-1.30e-06	0.9873615
Control	0.0006222	b	HD-Control	3.03e-05	0.0012535
PreHD	0.0006208	b	HD-PreHD	3.17e-05	0.0040823
<b>RD-RATR</b>					
HD	0.0005320	a	PreHD-Control	-1.00e-05	0.5692045
Control	0.0005016	b	HD-Control	3.05e-05	0.0034472
PreHD	0.0004916	b	HD-PreHD	4.04e-05	0.0008324
<b>RD-LATR</b>					
HD	0.0005135	a	PreHD-Control	-8.40e-06	0.5844837
Control	0.0004915	b	HD-Control	2.20e-05	0.0187306

PreHD	0.0004830	b	HD-PreHD	3.04e-05	0.0043635
-------	-----------	---	----------	----------	-----------

En el caso de la **difusividad axial en ATR derecho**, existen diferencias significativas entre controles-HD y preHD-HD con unos valores aumentados de difusividad axial en HD respecto a preHD y controles.

En el caso de la **difusividad axial en ATR izquierdo**, existen diferencias significativas entre controles-HD con unos valores aumentados de difusividad axial en HD respecto a controles.

En el caso de la **difusividad media para ambos ATR**, existen diferencias significativas entre controles-HD y preHD-HD con unos valores aumentados de difusividad media en HD respecto a preHD y controles.

En el caso de la **difusividad radial para ambos ATR**, existen diferencias significativas entre controles-HD y preHD-HD con unos valores aumentados de difusividad radial en HD respecto a preHD y controles.

## 4 ATR Average: ambos ATR conjuntamente

### 4.1 Creación de DATAFRAMES

```
RATR_metrics<-cbind(ATR_avg_demo[,c(1,8,10:13)])
LATR_metrics<-cbind(ATR_avg_demo[,c(1,9,14:17)])
ATR_metrics<-cbind(ATR_avg_demo[,c(1,8:17)])

ATR_Iron<-c(ATR_metrics$rhatr_Iron, ATR_metrics$lhatr_Iron)
ATR_AD<-c(ATR_metrics$rhatr_AD, ATR_metrics$lhatr_AD)
ATR_RD<-c(ATR_metrics$rhatr_RD, ATR_metrics$lhatr_RD)
ATR_MD<-c(ATR_metrics$rhatr_MD, ATR_metrics$lhatr_MD)
ATR_FA<-c(ATR_metrics$rhatr_FA, ATR_metrics$lhatr_FA)

ATR_codi<-c(ATR_metrics$codi, ATR_metrics$codi)
ATR_subgroup<-c(ATR_avg_demo$Subgroup, ATR_avg_demo$Subgroup)
ATR_Gender<-c(ATR_avg_demo$Gender, ATR_avg_demo$Gender)
ATR_agef<-c(ATR_avg_demo$agef, ATR_avg_demo$agef)
ATR_agef2<-c(ATR_avg_demo$agef2, ATR_avg_demo$agef2)
ATR_hemisphere<-c(rep("right",nrow(ATR_avg_demo)), rep("left",nrow(ATR_avg_demo)))

ATR_DifIron<-cbind(ATR_codi, ATR_hemisphere, ATR_subgroup, ATR_Gender,
                     ATR_agef, ATR_agef2, ATR_Iron, ATR_AD, ATR_RD, ATR_MD, ATR_FA)
ATR_DifIron_Df<-as.data.frame(ATR_DifIron)
ATR_DifIron_Df$ATR_hemisphere<-as.factor(ATR_DifIron_Df$ATR_hemisphere)
ATR_DifIron_Df$ATR_subgroup<-as.factor(ATR_DifIron_Df$ATR_subgroup)
ATR_DifIron_Df$ATR_Gender<-as.factor(ATR_DifIron_Df$ATR_Gender)
ATR_DifIron_Df$ATR_agef<-as.factor(ATR_DifIron_Df$ATR_agef)
ATR_DifIron_Df$ATR_agef2<-as.factor(ATR_DifIron_Df$ATR_agef2)
ATR_DifIron_Df[,7:11]<-as.numeric(unlist(ATR_DifIron_Df[,7:11]))

# Summary statistics del dataframe
sumATR_DifIron_Df<-summary(ATR_DifIron_Df[,-c(1,5)])
kable(list(sumATR_DifIron_Df[,1:4], sumATR_DifIron_Df[,5:9]),
      caption='Summary statistics for both ATR conjointly', booktabs=TRUE)
```

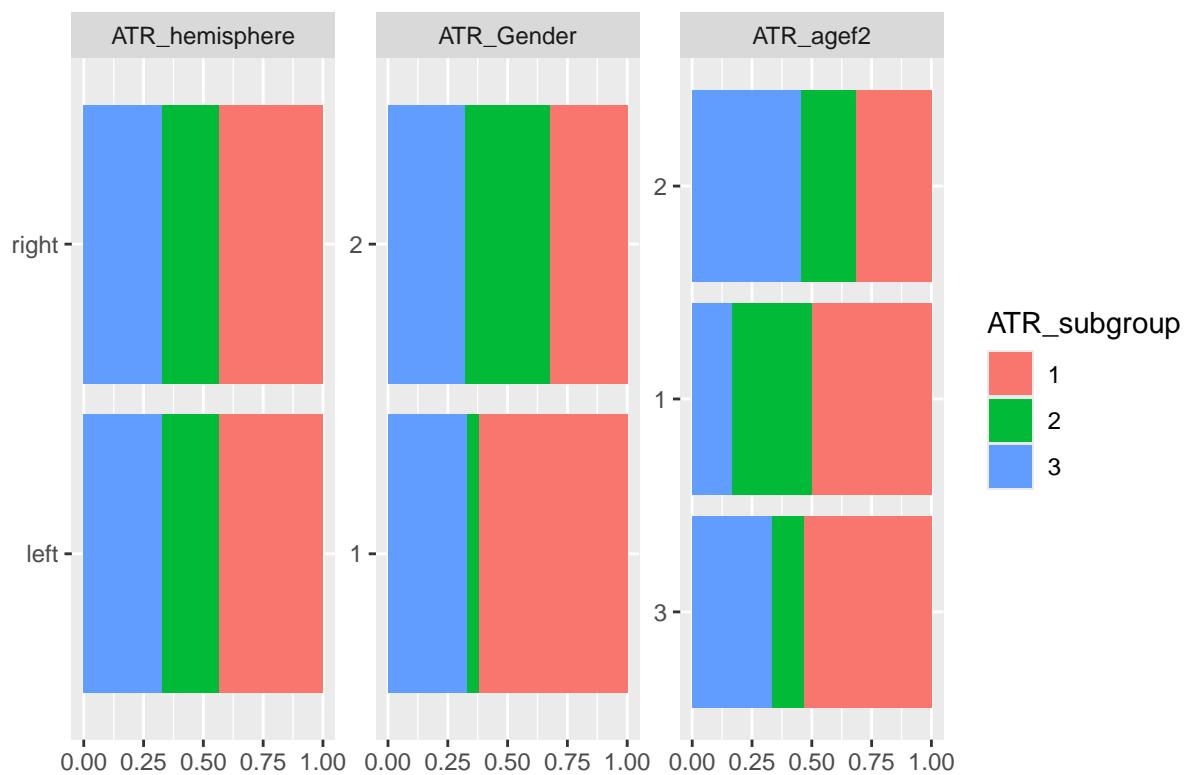
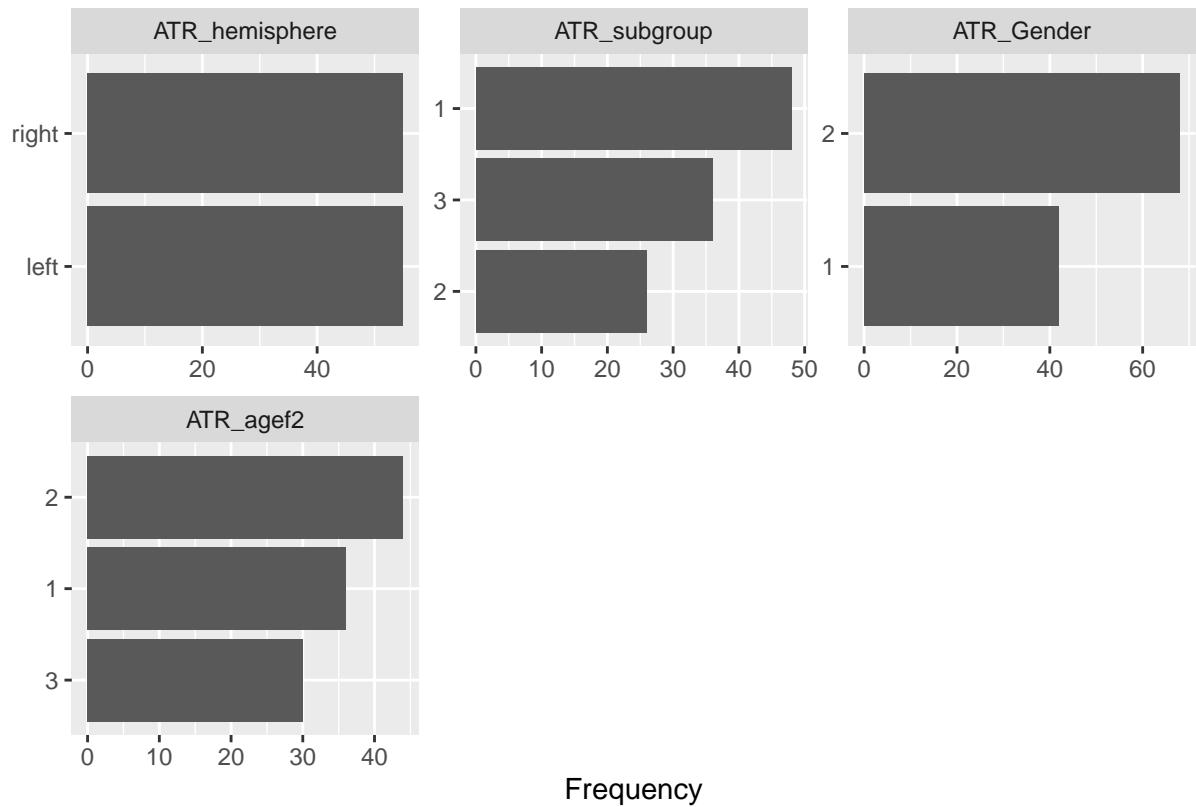
Cuadro 62: Summary statistics for both ATR conjointly

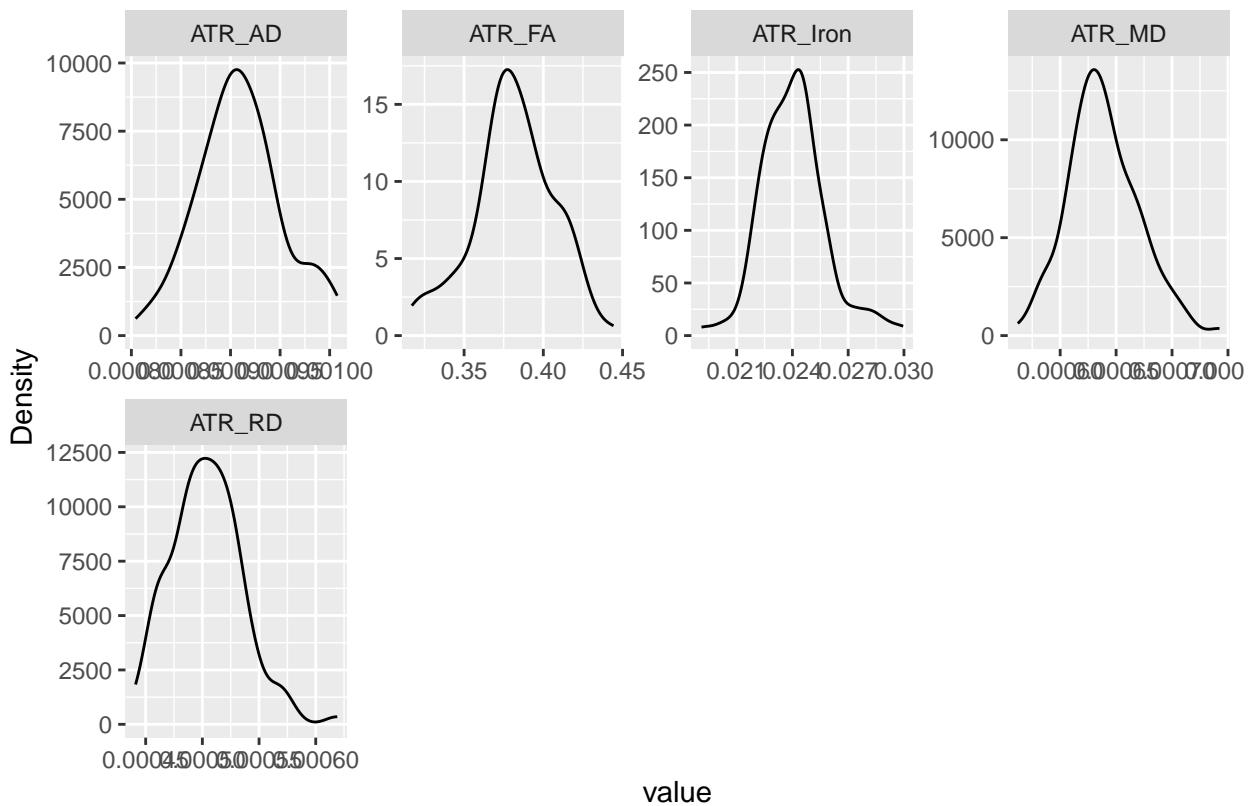
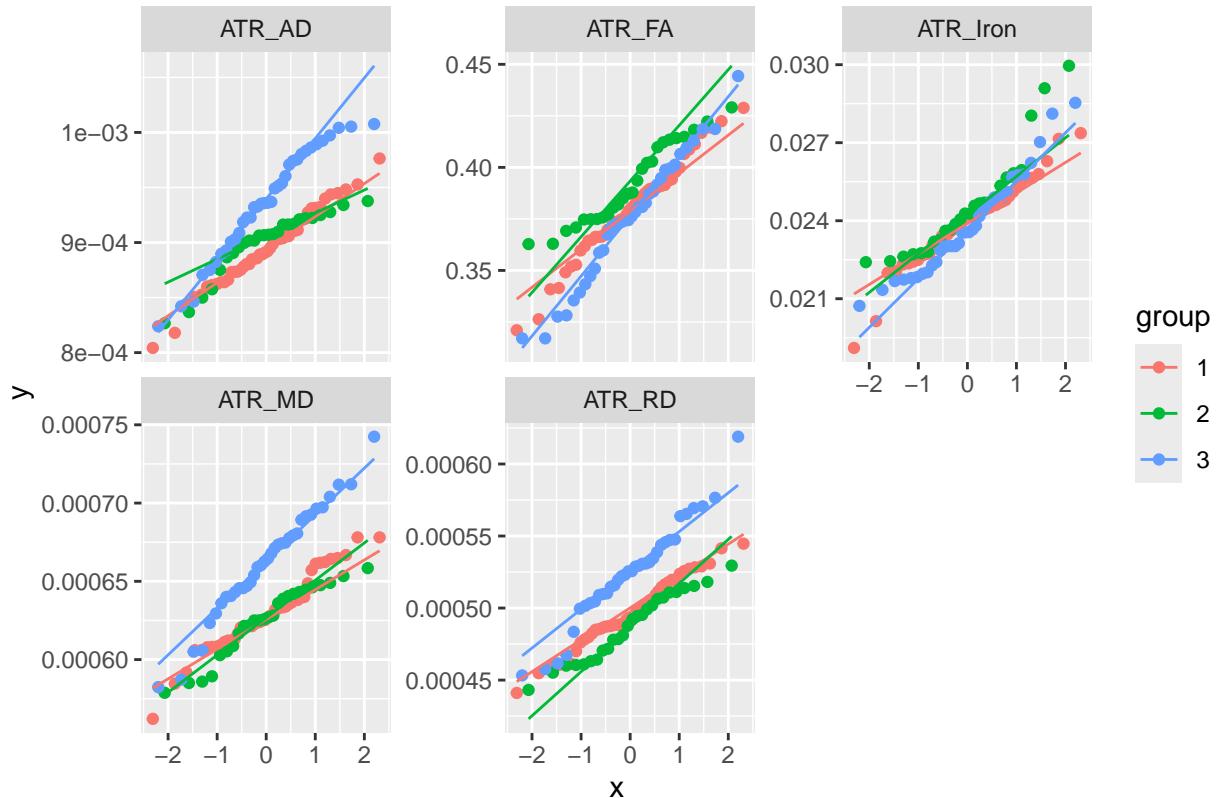
ATR_hemisphere	ATR_subgroup	ATR_Gender	ATR_agef2	
left :55	1:48	1:42	1:36	
right:55	2:26	2:68	2:44	
NA	3:36	NA	3:30	
NA	NA	NA	NA	
NA	NA	NA	NA	
NA	NA	NA	NA	
ATR_Iron	ATR_AD	ATR_RD	ATR_MD	ATR_FA
Min. :0.01911	Min. :0.0008042	Min. :0.0004411	Min. :0.0005621	Min. :0.3170
1st Qu.:0.02289	1st Qu.:0.0008809	1st Qu.:0.0004839	1st Qu.:0.0006209	1st Qu.:0.3673
Median :0.02404	Median :0.0009065	Median :0.0005019	Median :0.0006360	Median :0.3806
Mean :0.02406	Mean :0.0009092	Mean :0.0005035	Mean :0.0006387	Mean :0.3809
3rd Qu.:0.02488	3rd Qu.:0.0009338	3rd Qu.:0.0005236	3rd Qu.:0.0006600	3rd Qu.:0.3986
Max. :0.02996	Max. :0.0010076	Max. :0.0006190	Max. :0.0007424	Max. :0.4443

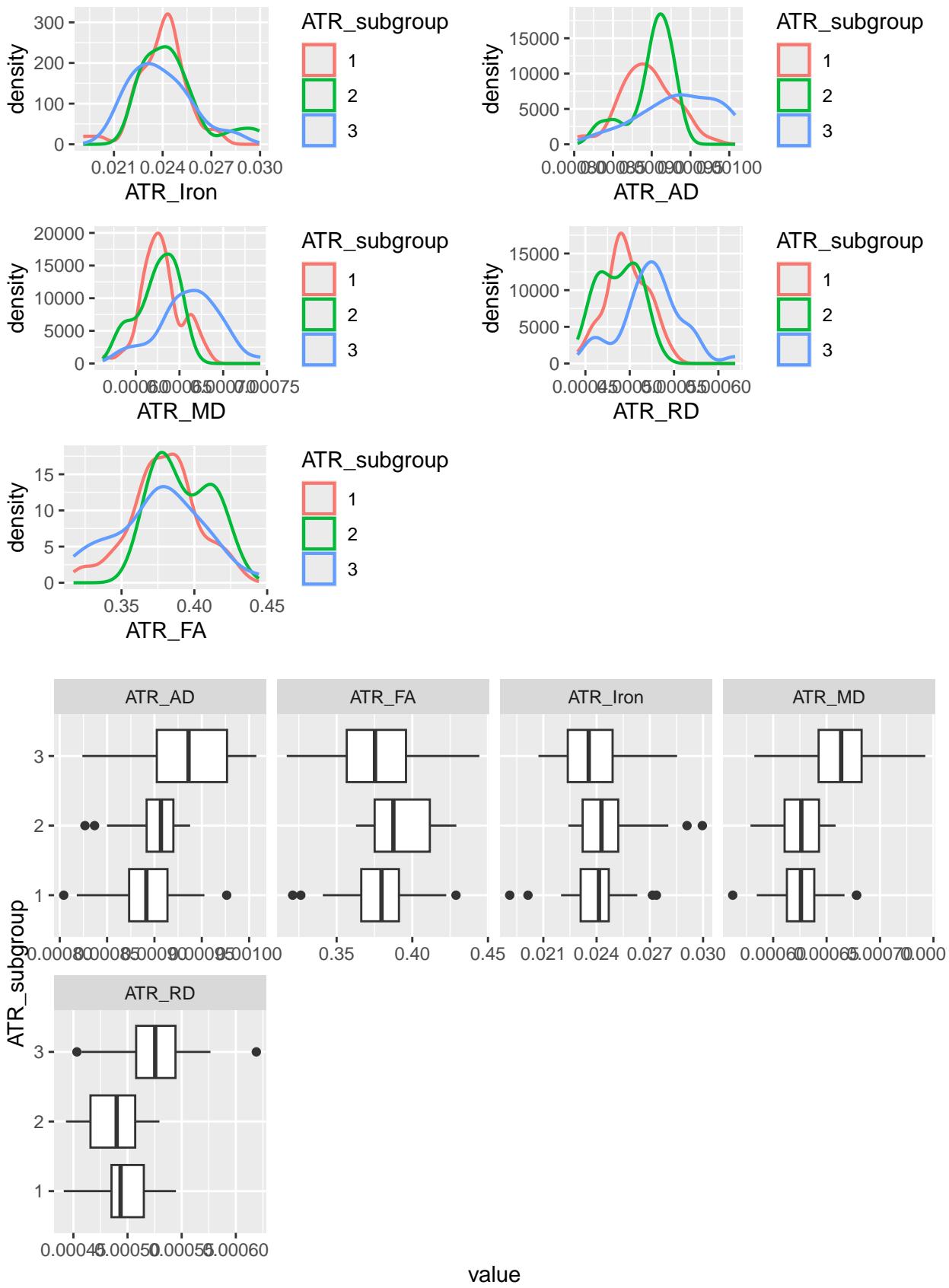
## 4.2 Automated Exploratory Data Analysis

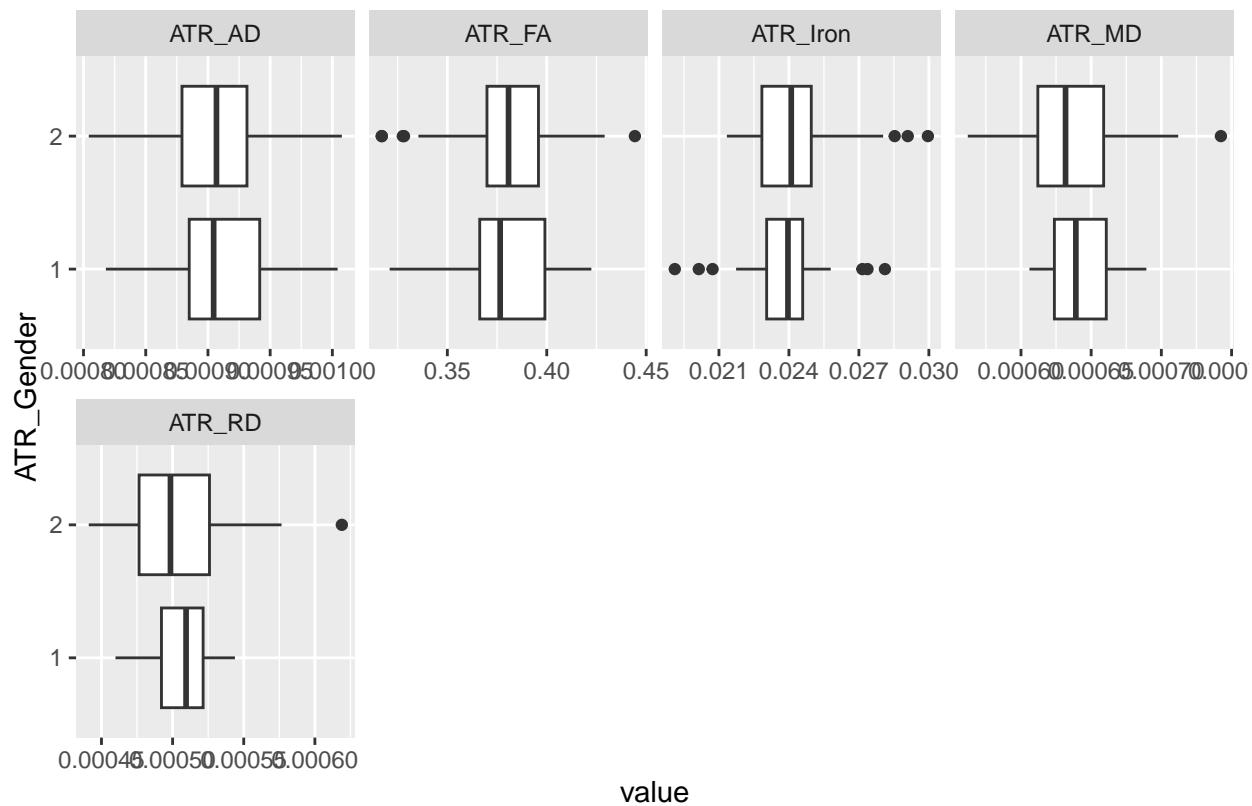
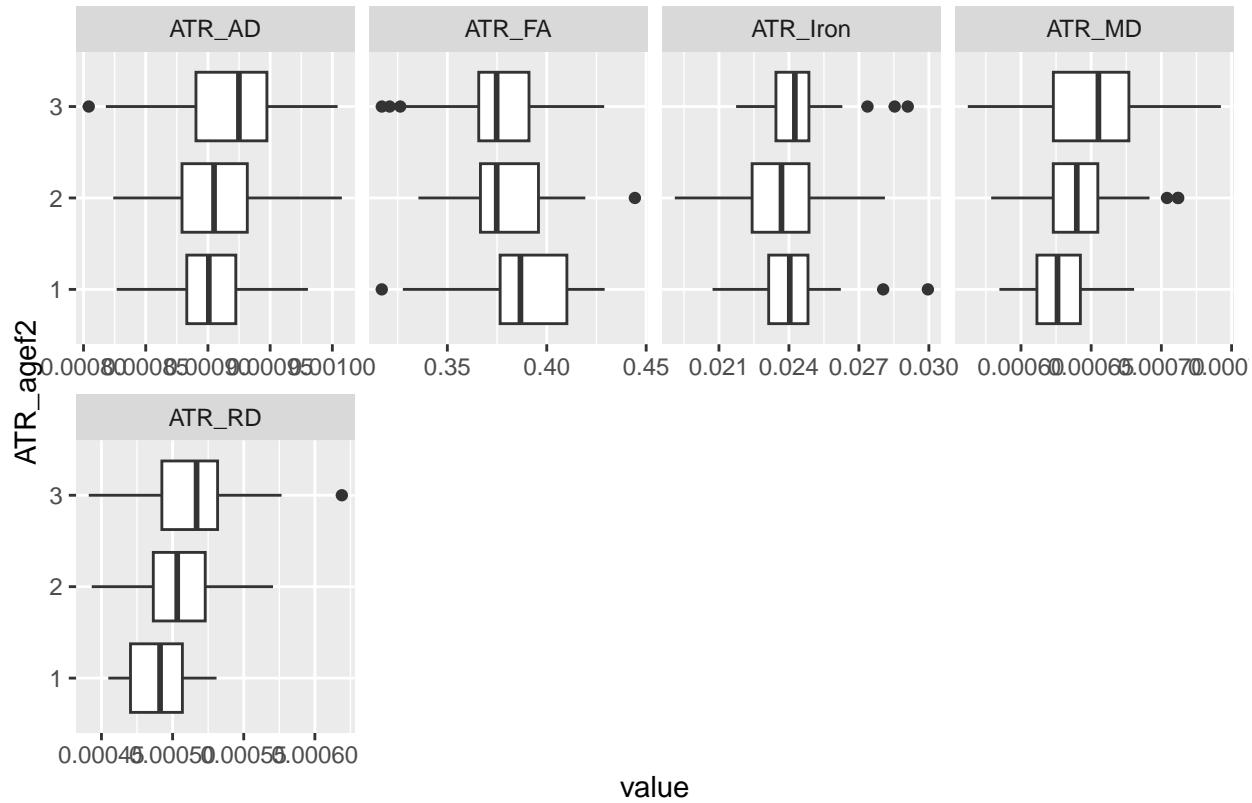
```
library(DataExplorer)
library(dplyr)
ATR_DifIron_EDA<-ATR_DifIron_Df[,-c(1,5)]
ATR_DifIron_EDA %>%
  create_report(
    output_file = paste("Report"),
    report_title = "EDA Report - ATR",
    y = "ATR_subgroup"
  )
```

Se generan los diferentes plots que se consideran necesarios con las funciones del paquete **DataExplorer**.

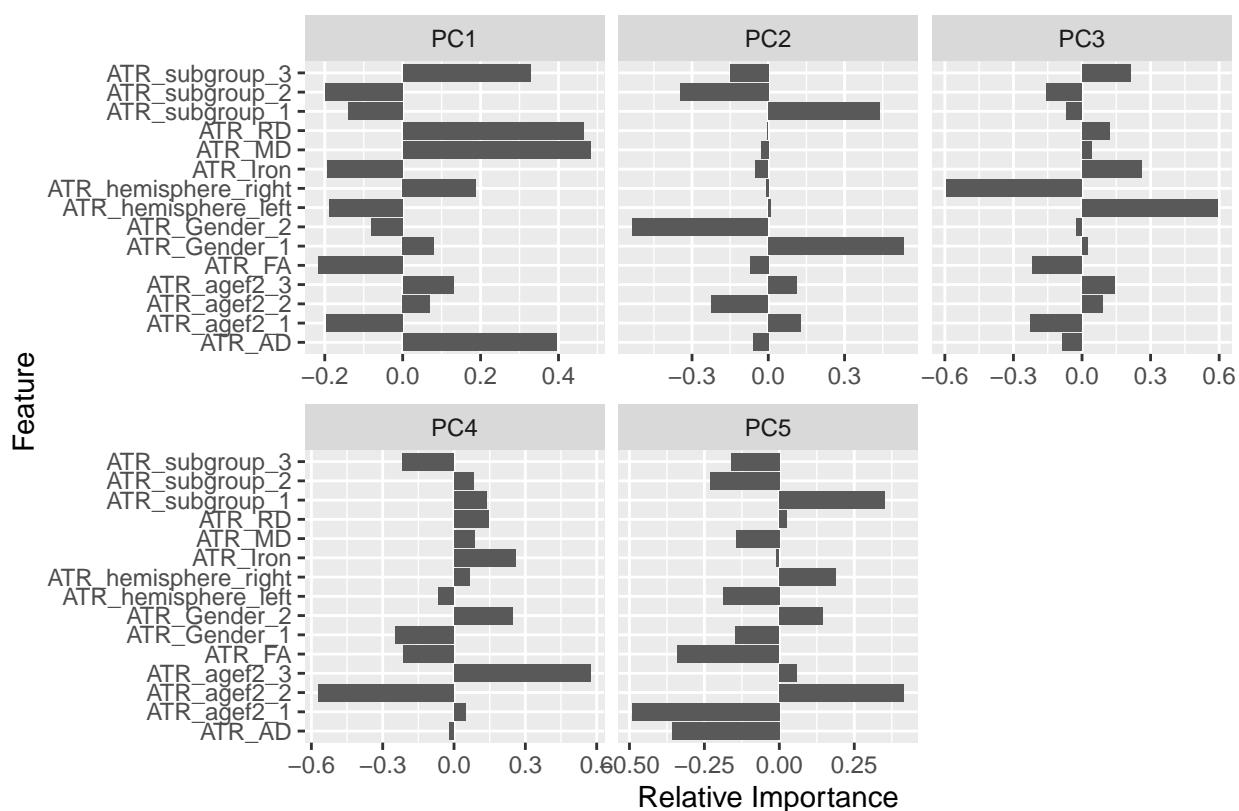
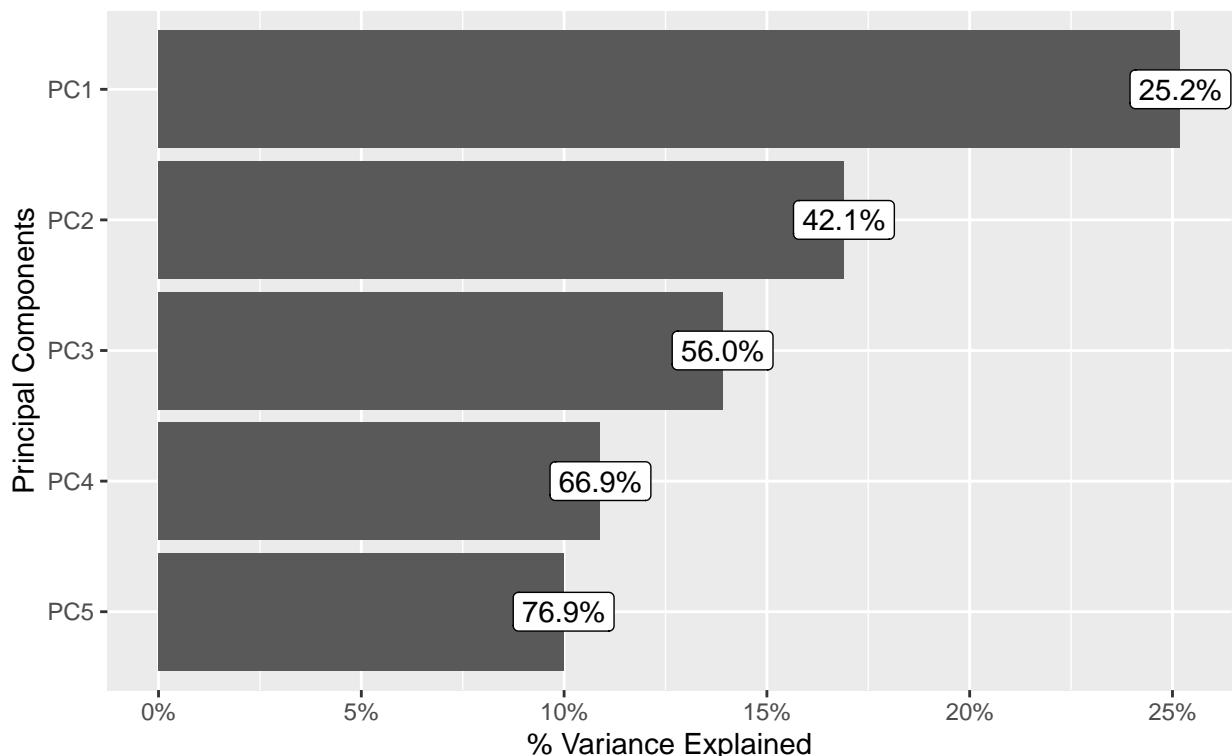




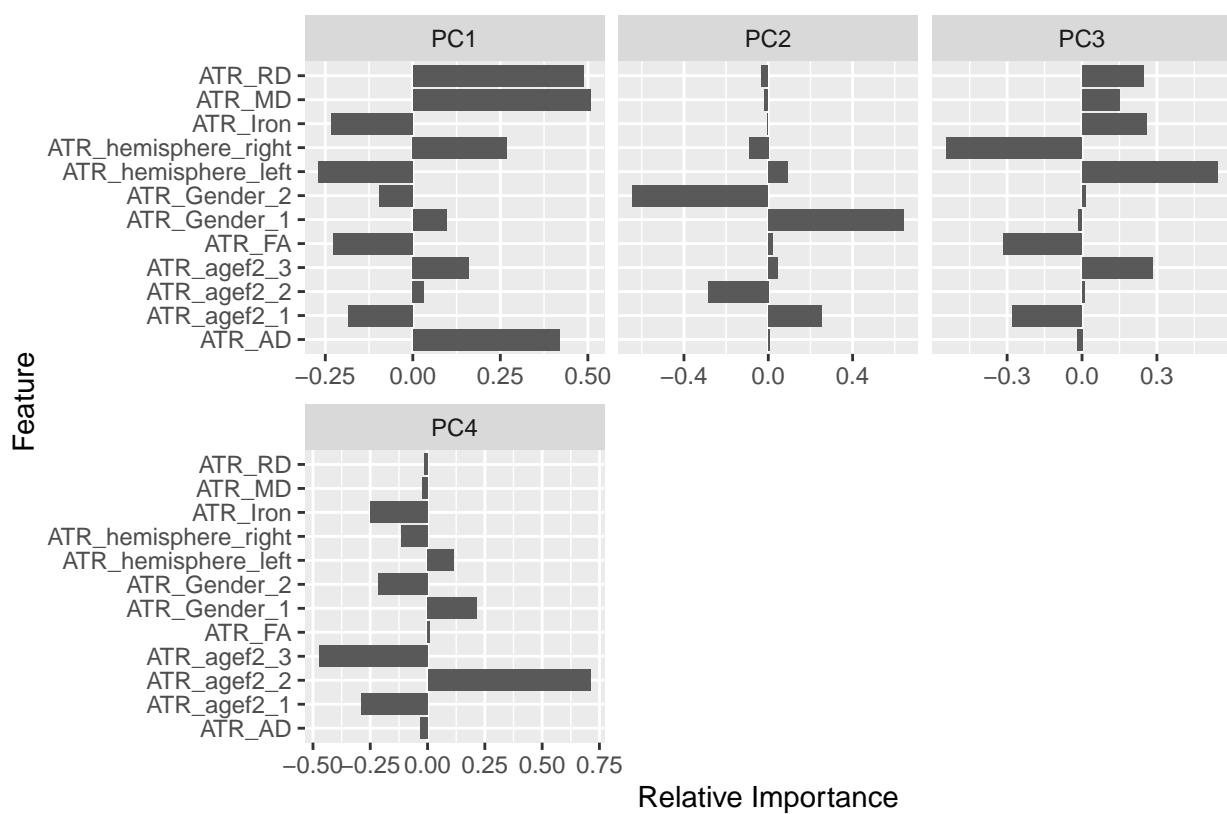
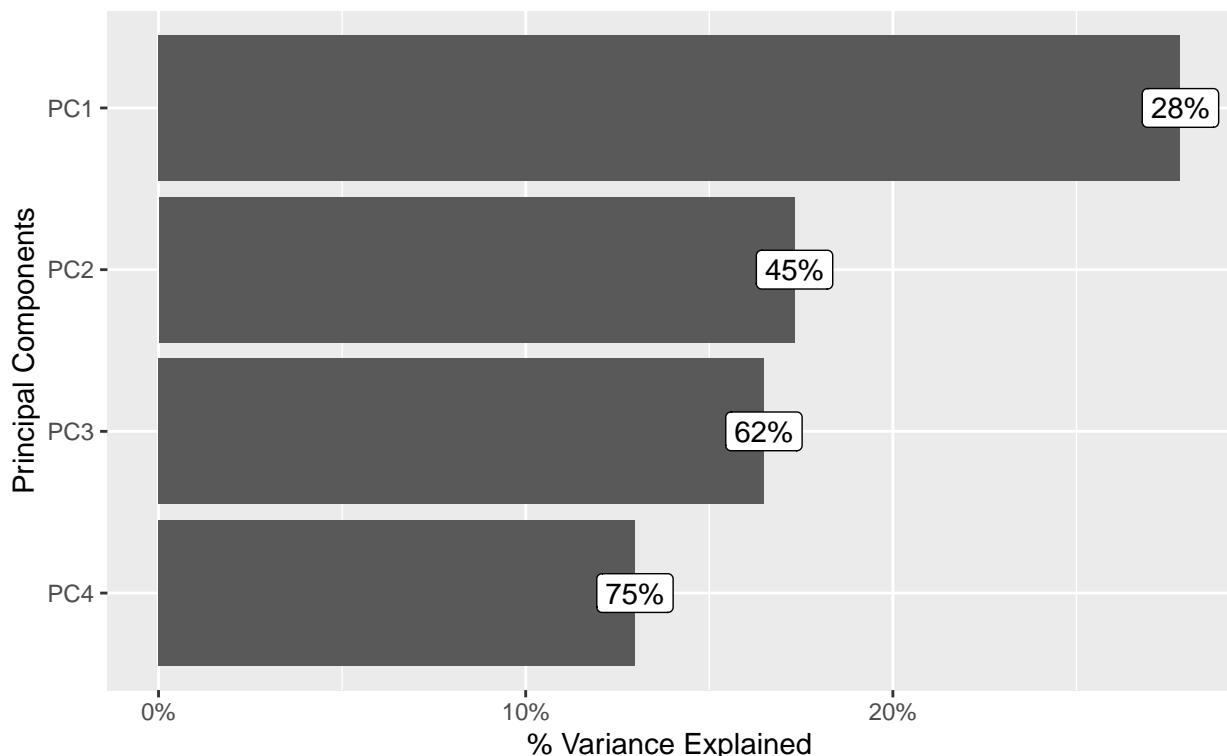




% Variance Explained By Principal Components  
 (Note: Labels indicate cumulative % explained variance)



% Variance Explained By Principal Components  
 (Note: Labels indicate cumulative % explained variance)

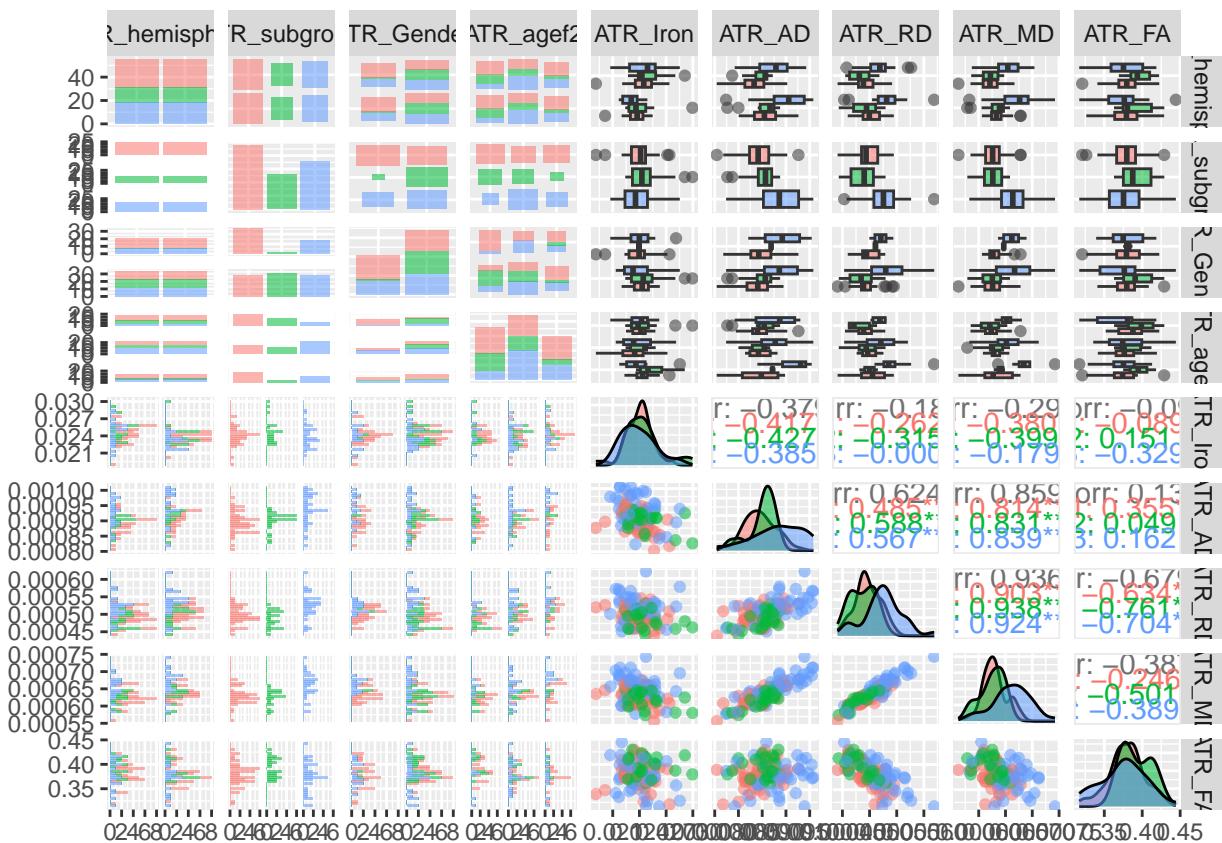


Se genera un plot con la función `ggpairs` del paquete `GGally`.

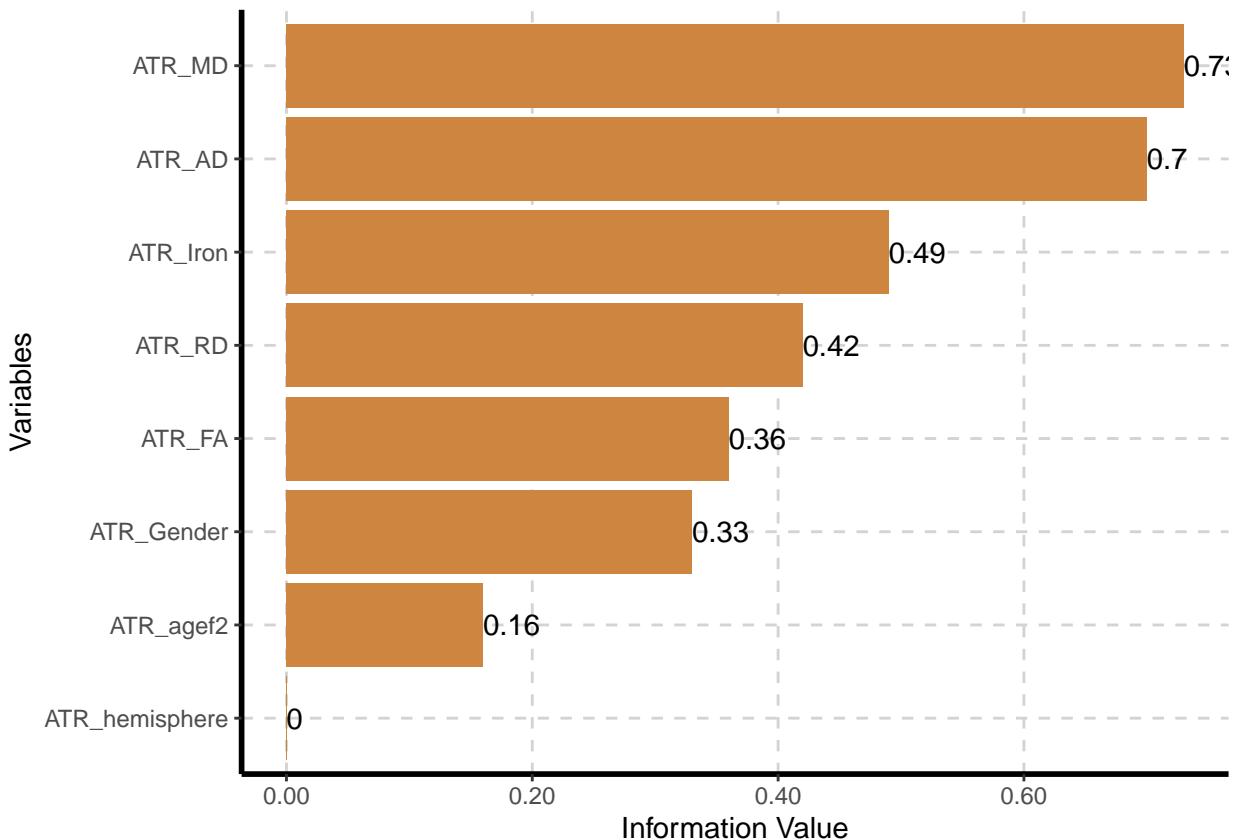
```

library(GGally)
ATR_DifIron_EDA %>%
  ggpairs(mapping = aes(color=ATR_DifIron_EDA$ATR_subgroup, alpha=0.5))

```



Se genera una tabla mediante la función ExpCatStat del paquete SmartEDA para determinar la importancia de las variables en función de su información. Esta función combina resultados de la ponderación de la evidencia, el valor de la información y las estadísticas resumidas. La función discretiza las variables numéricas y que el número de bins modifica el valor informativo de las variables y los IV values.



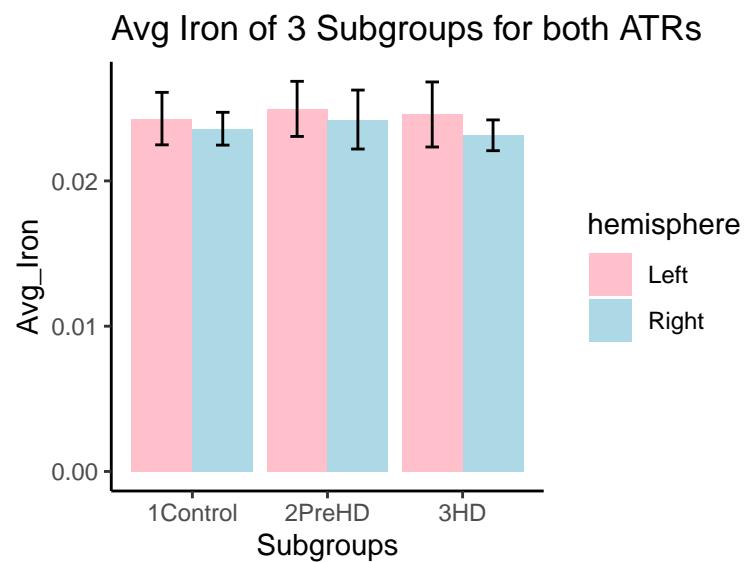
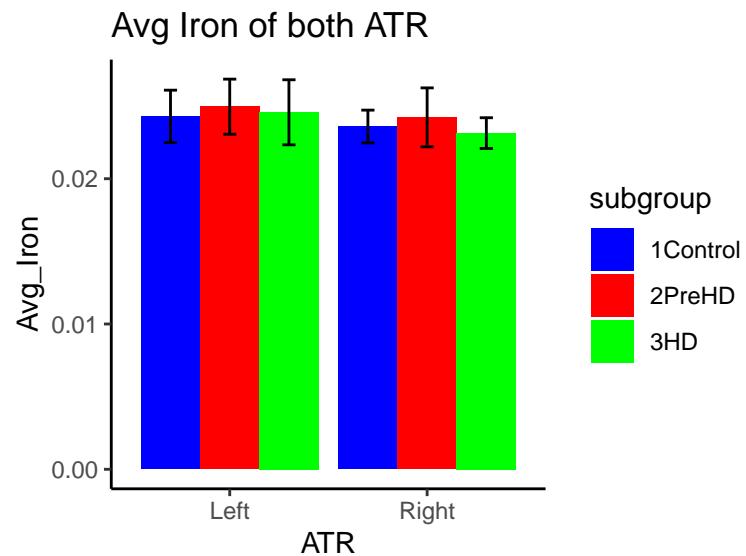
Cuadro 63: Asociaciones y Poder predictivo de las variables respecto a Subgrupo

Variable	Target	Unique	Chi-squared	p-value	df	IV Value	Cramers V	Degree of Association	Predictive Power
ATR_hemisphere	ATR_subgroup	2	0.000	1.000	NA	0.00	0.00	Very Weak	Not Predictive
ATR_Gender	ATR_subgroup	2	15.444	0.001	NA	0.33	0.37	Strong	Highly Predictive
ATR_agef2	ATR_subgroup	3	10.202	0.029	NA	0.16	0.22	Moderate	Somewhat Predictive
ATR_Iron	ATR_subgroup	20	43.062	0.248	NA	0.49	0.44	Strong	Highly Predictive
ATR_AD	ATR_subgroup	20	60.811	0.006	NA	0.70	0.53	Strong	Highly Predictive
ATR_RD	ATR_subgroup	20	59.912	0.008	NA	0.42	0.52	Strong	Highly Predictive
ATR_MD	ATR_subgroup	20	66.398	0.000	NA	0.73	0.55	Strong	Highly Predictive
ATR_FA	ATR_subgroup	20	34.933	0.644	NA	0.36	0.40	Strong	Highly Predictive

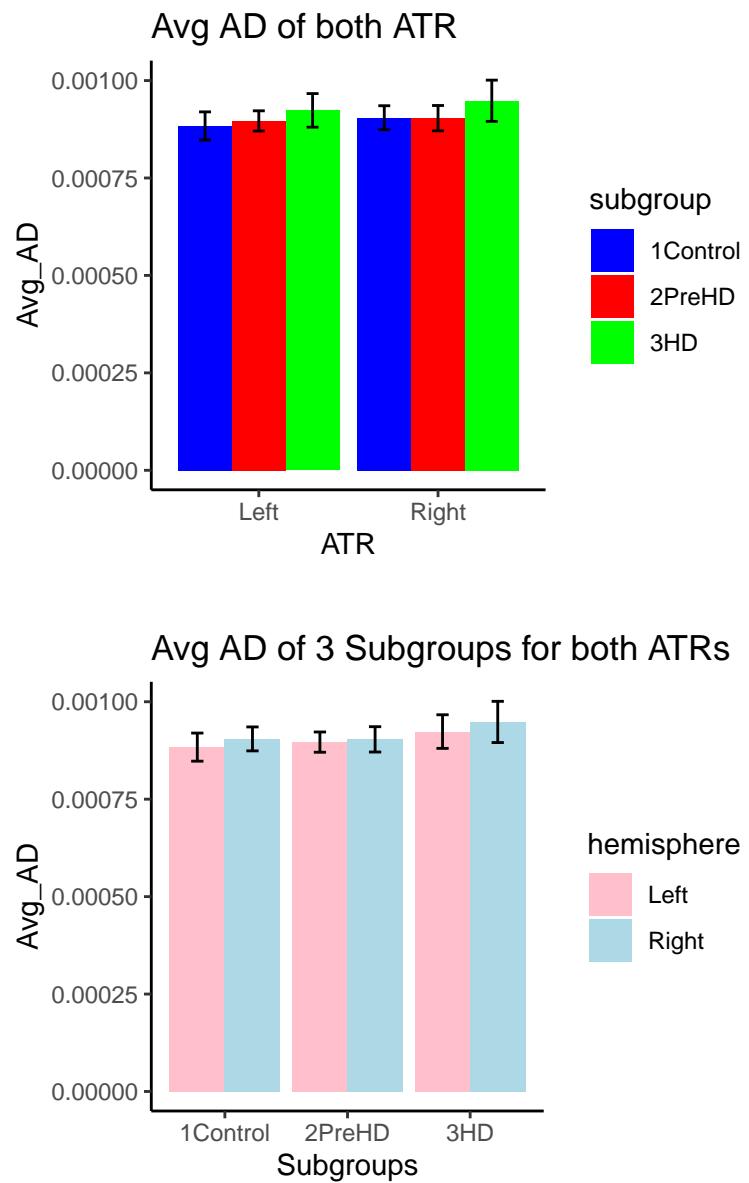
## 4.3 Estadística descriptiva gráfica

### 4.3.1 BARPLOTS

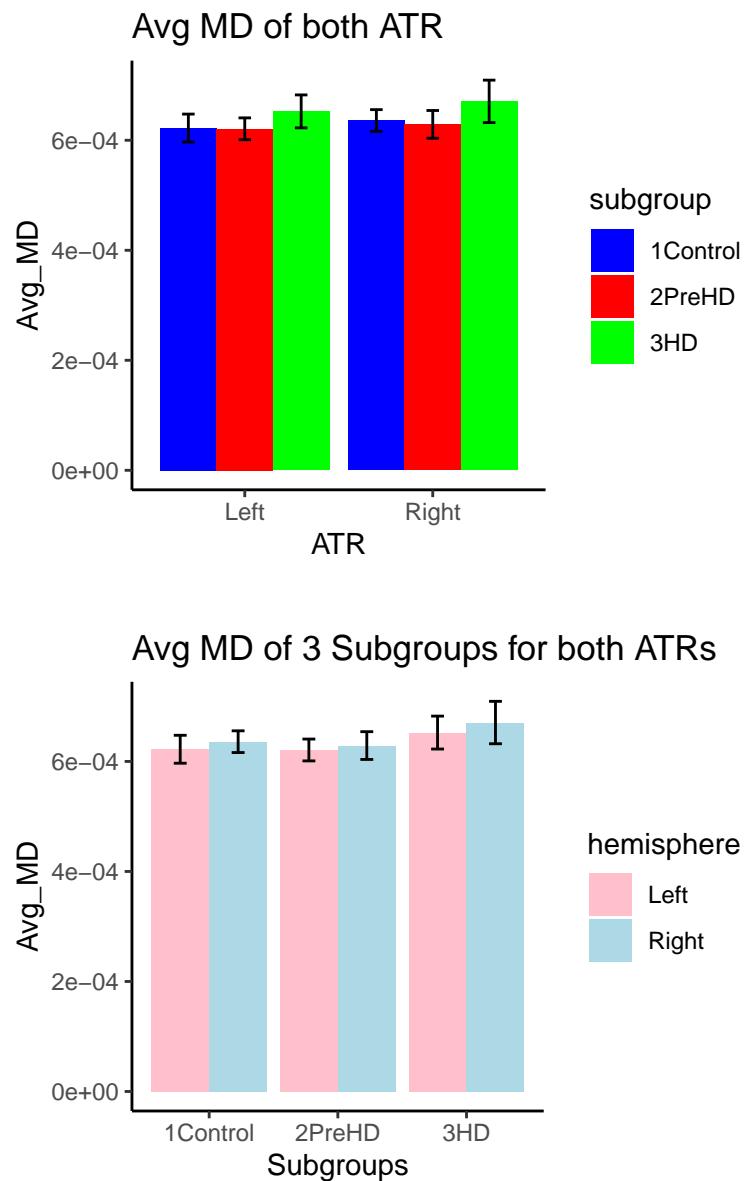
#### 4.3.1.1 IRON



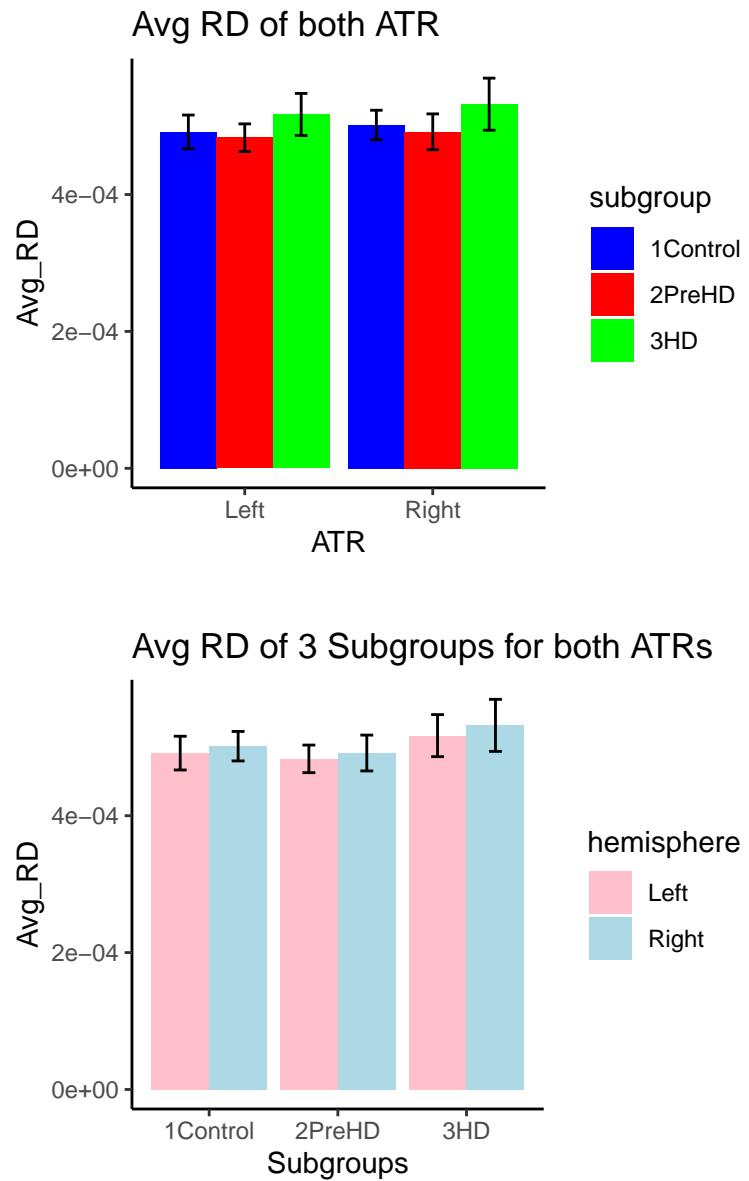
#### 4.3.1.2 AD



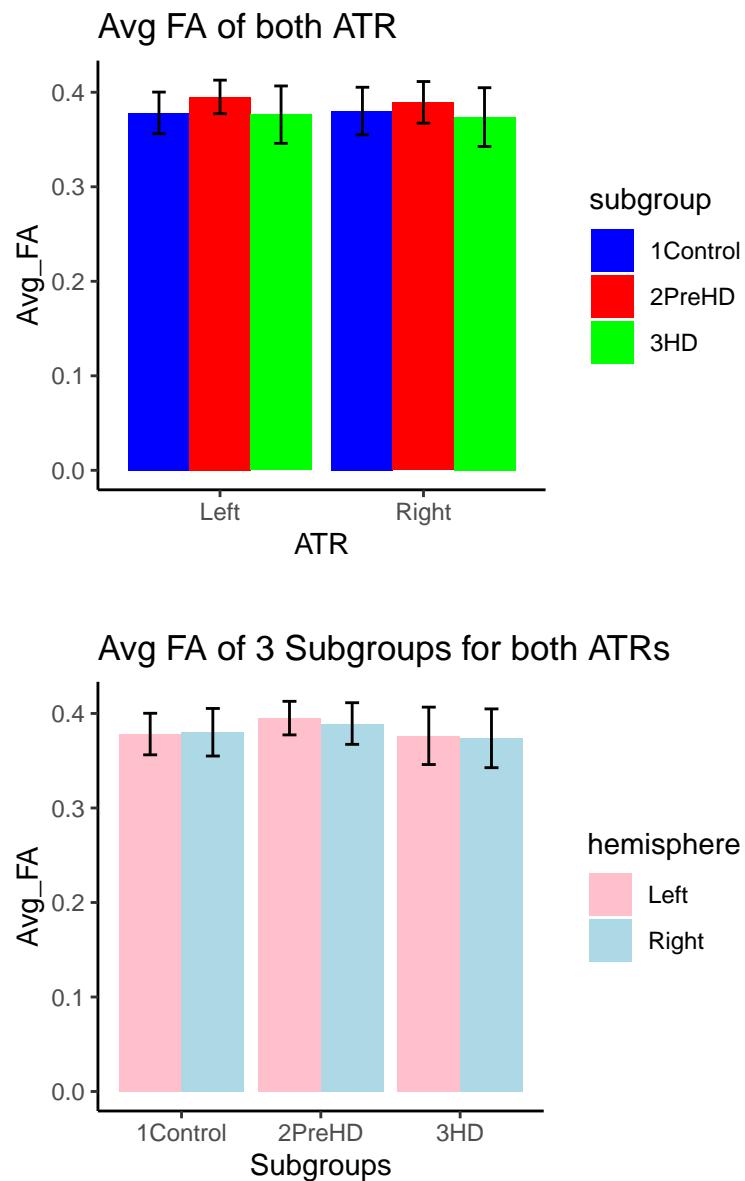
#### 4.3.1.3 MD



#### 4.3.1.4 RD



#### 4.3.1.5 FA



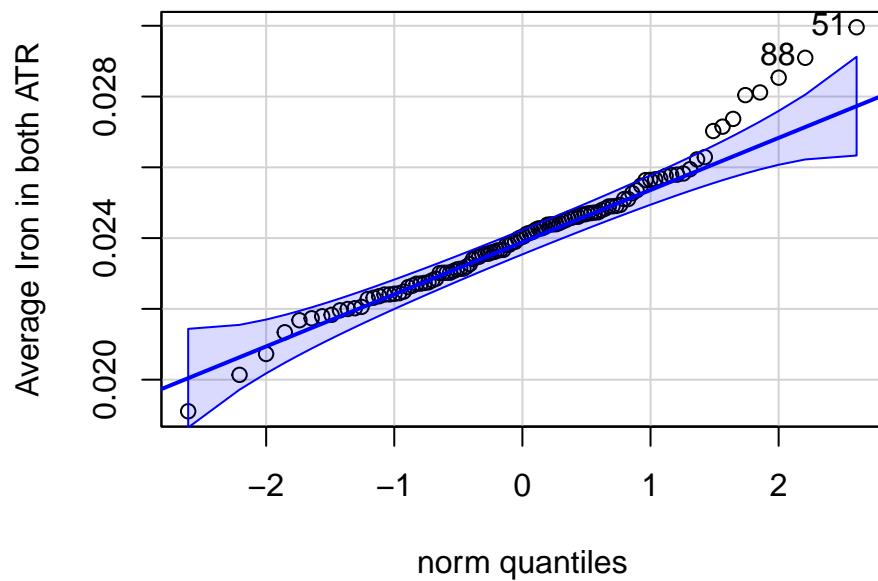
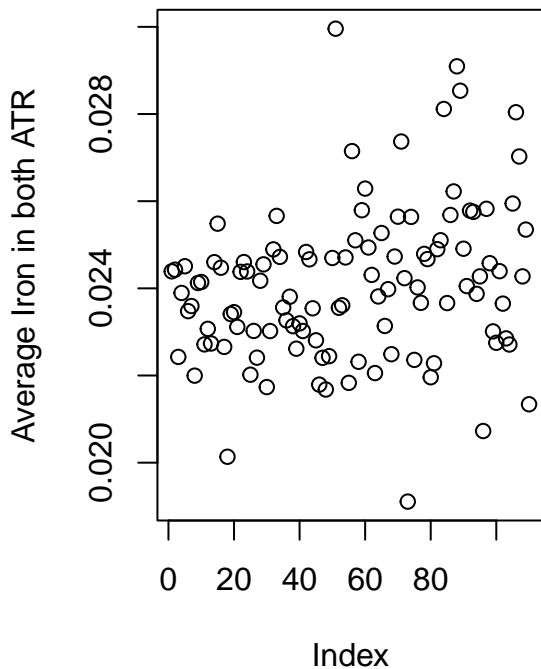
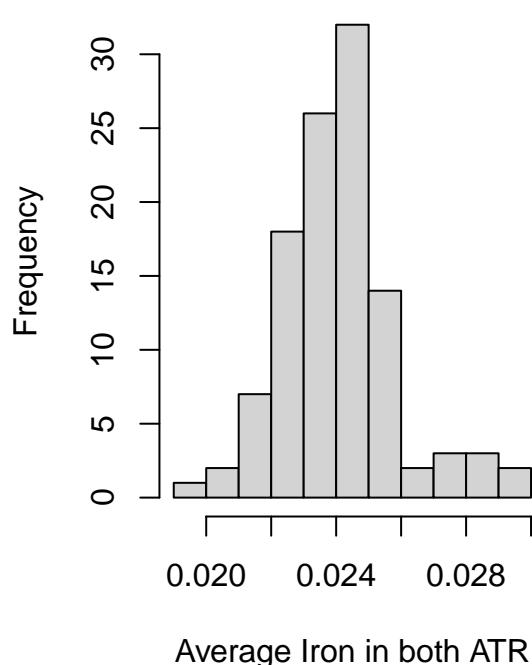
#### 4.3.2 HISTOGRAMAS y QQ-PLOTS

Para obtener una idea de la **normalidad de los datos y de la presencia/ausencia de outliers**, para cada métrica de RM, se genera un histograma, un plot y cuatro qq-plots (un qqplot sin distinción entre subgrupos y un qqplot para cada subgrupo). No se genera un qqplot para cada hemisferio dado que ya se ha realizado dicho gráfico en el apartado de análisis de los ATRs derecho e izquierdo por separado.

Para valorar posibles **diferencias en variabilidad de las métricas de RM en los niveles de los diferentes factores de interés** subgrupo, hemisferio, género y edad factorizada, también se generan plots de las métricas de RM para cada subgrupo, hemisferio, género y grupo de edad (tanto para la edad factorizada de 7 niveles como la edad factorizada de 3 niveles).

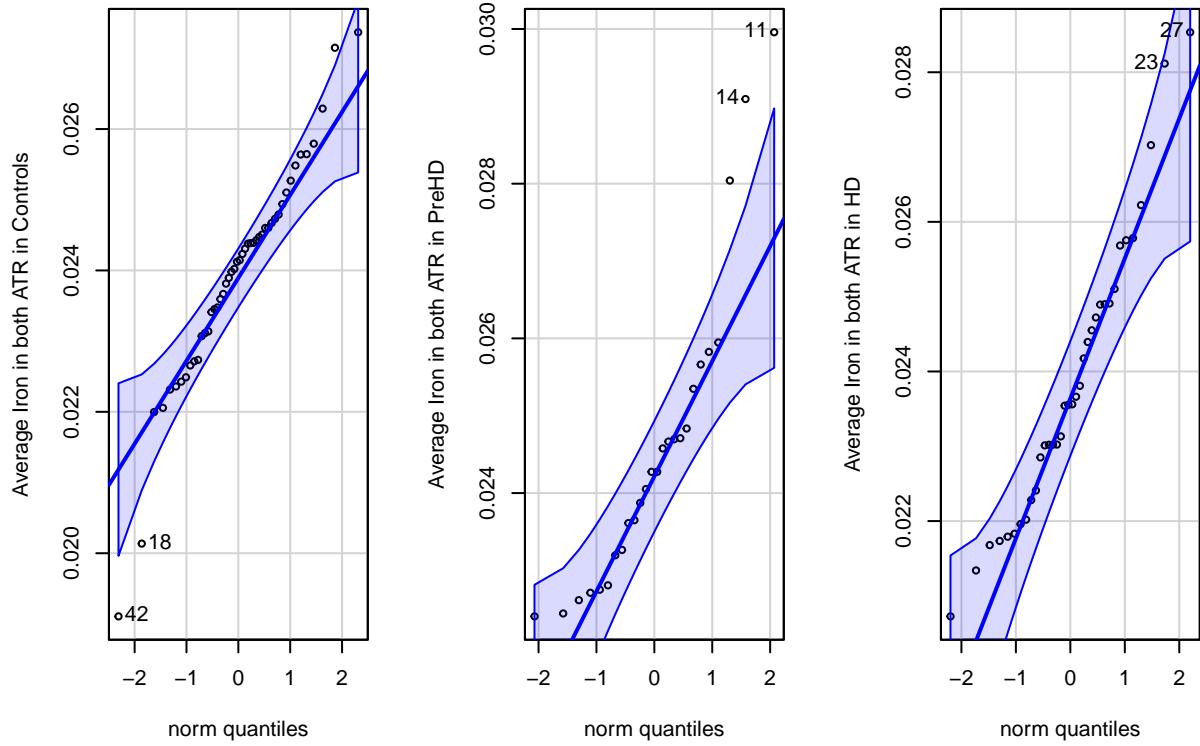
#### 4.3.2.1 IRON

Histogram of ATR Average Iron

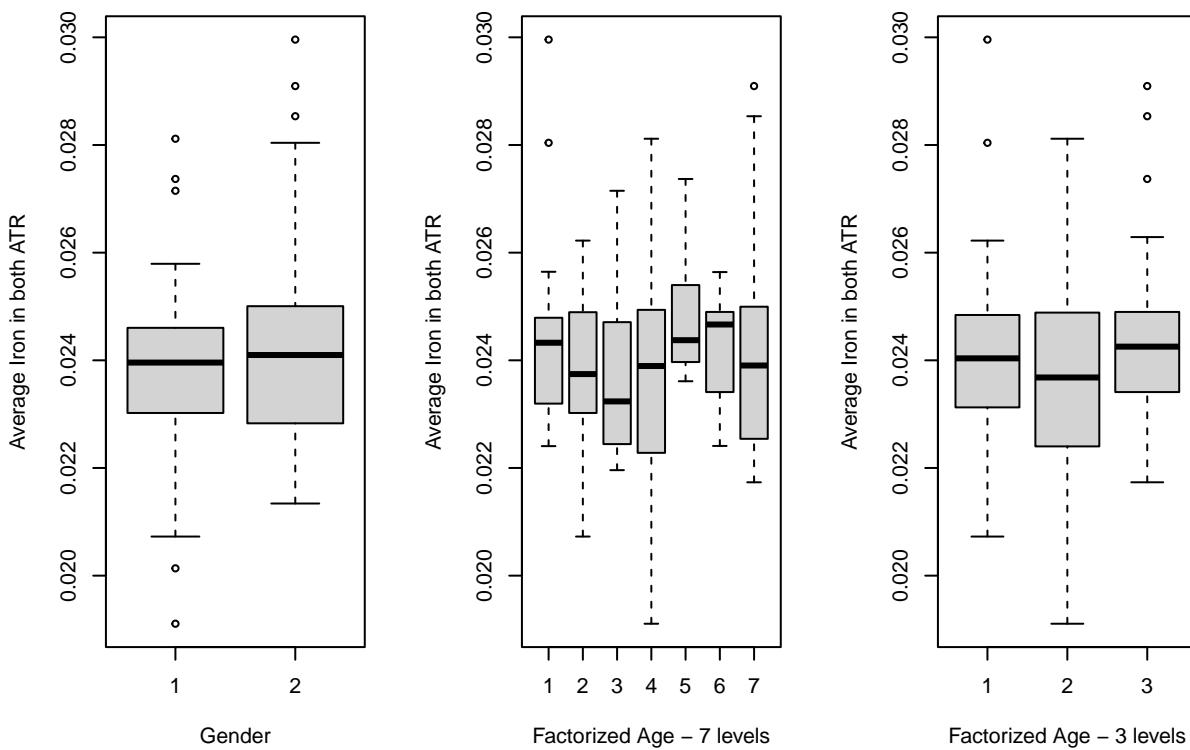
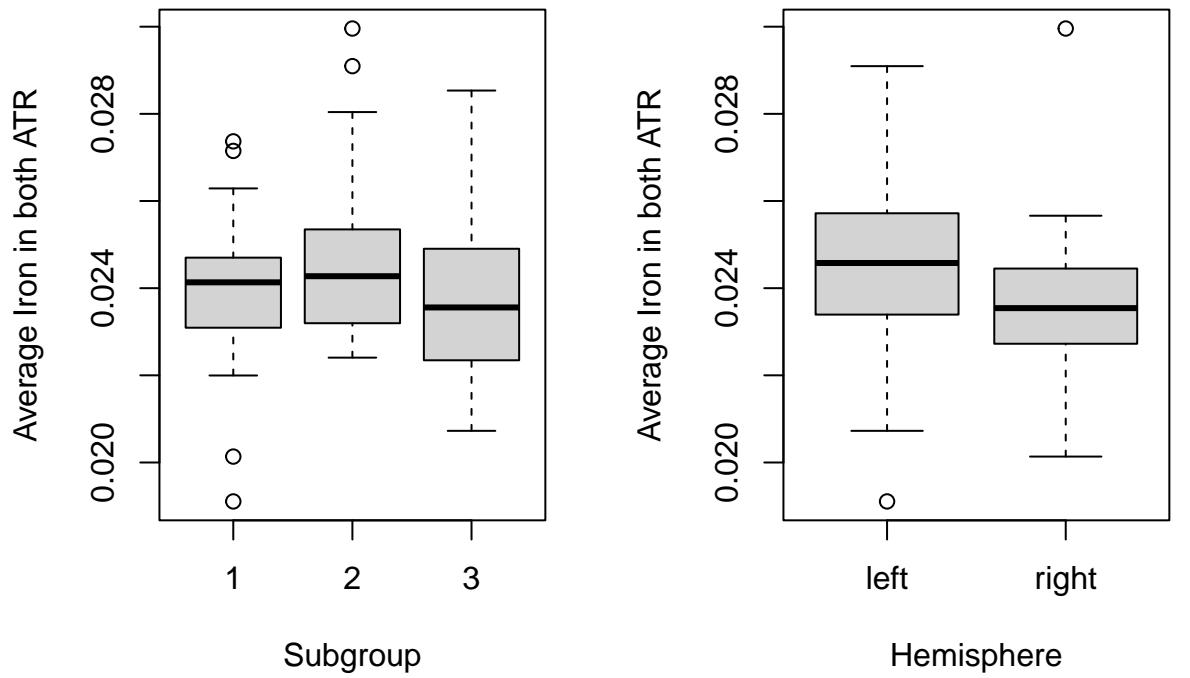


[1] 51 88

```
[1] 42 18  
[1] 11 14
```

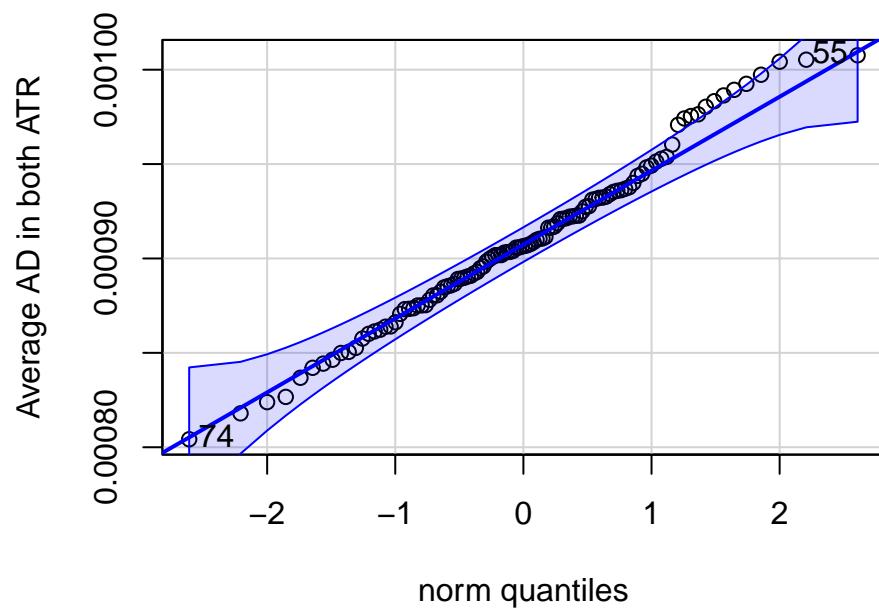
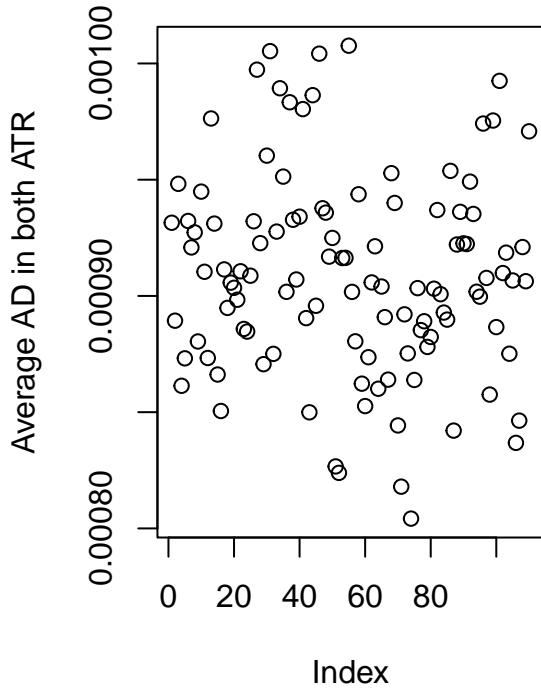
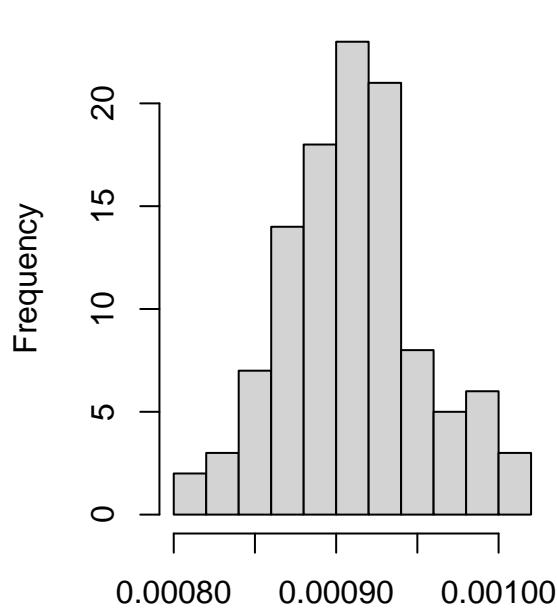


```
[1] 27 23
```



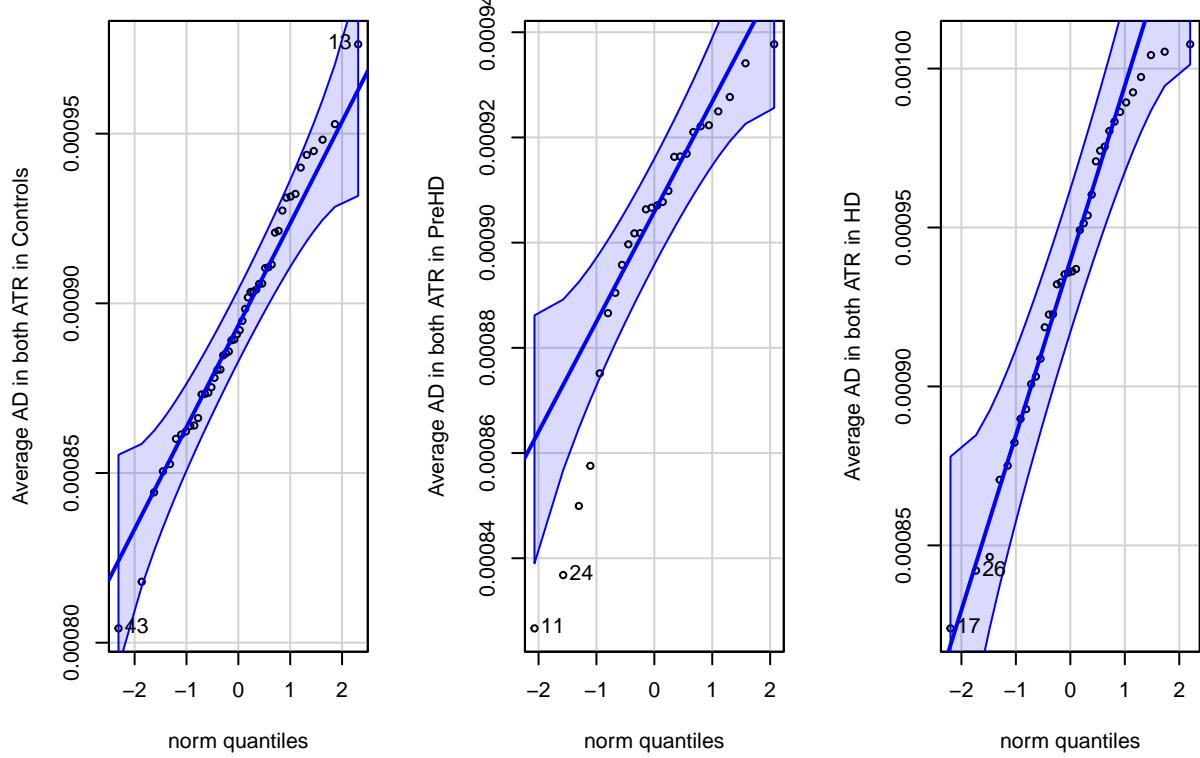
#### 4.3.2.2 AD

Histogram of ATR Average AD

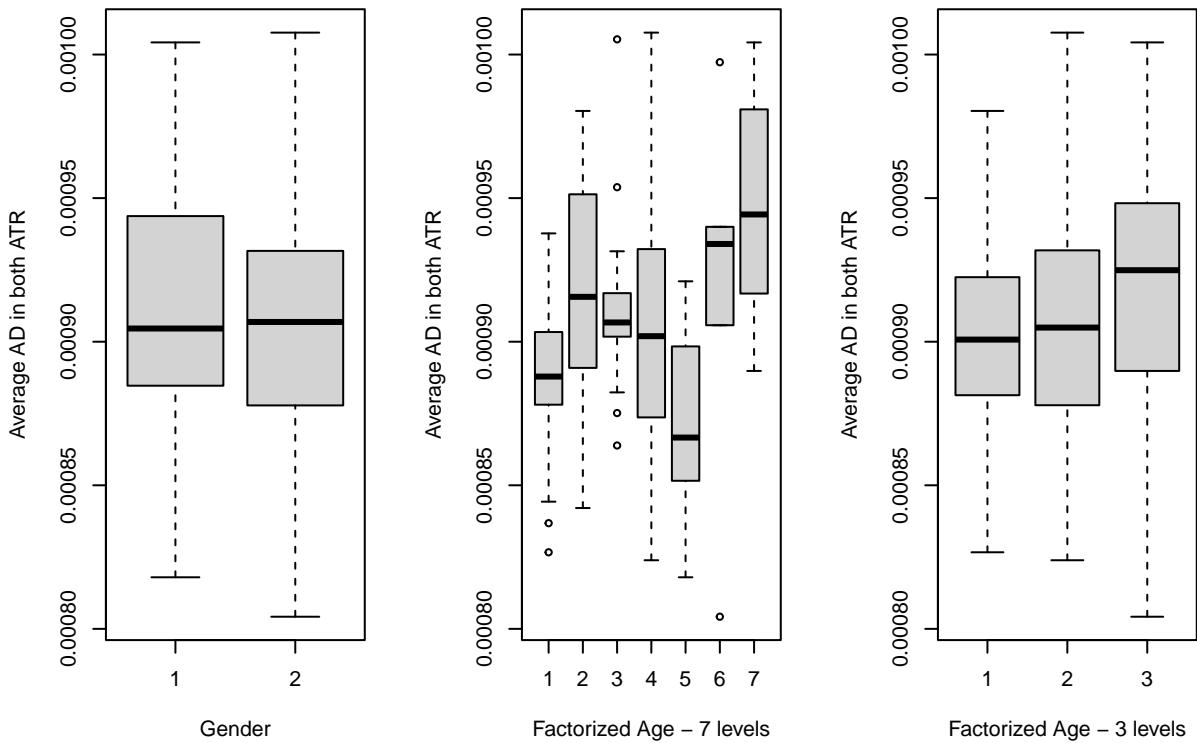
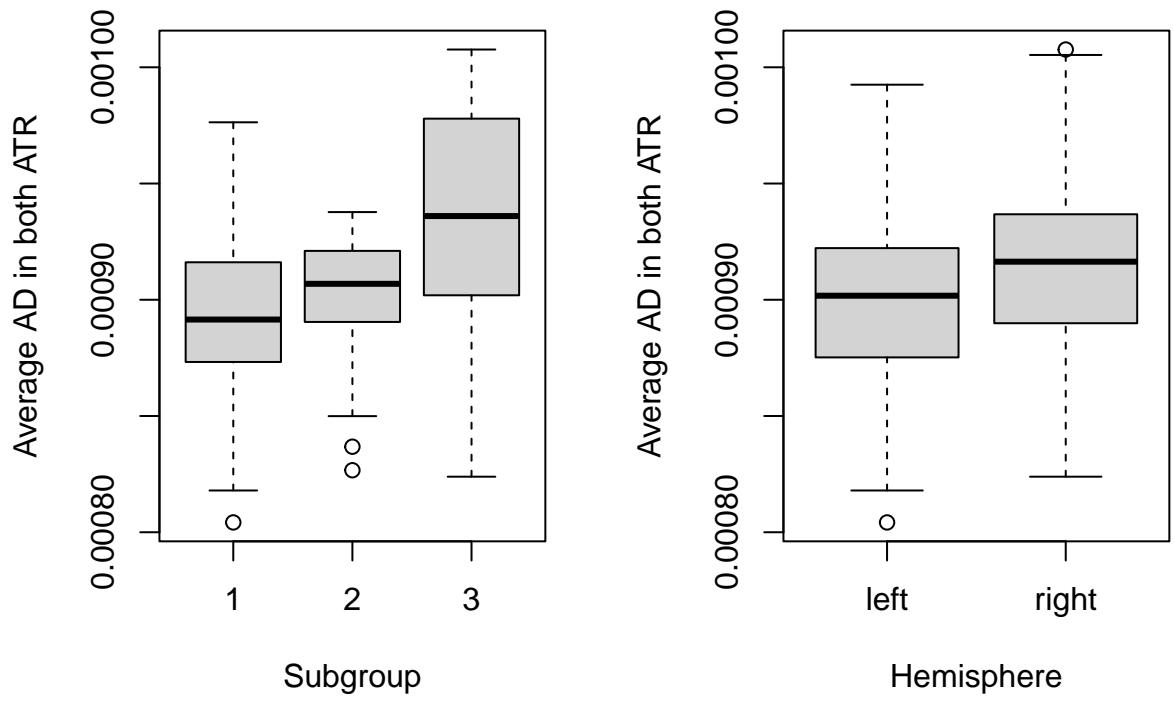


[1] 74 55

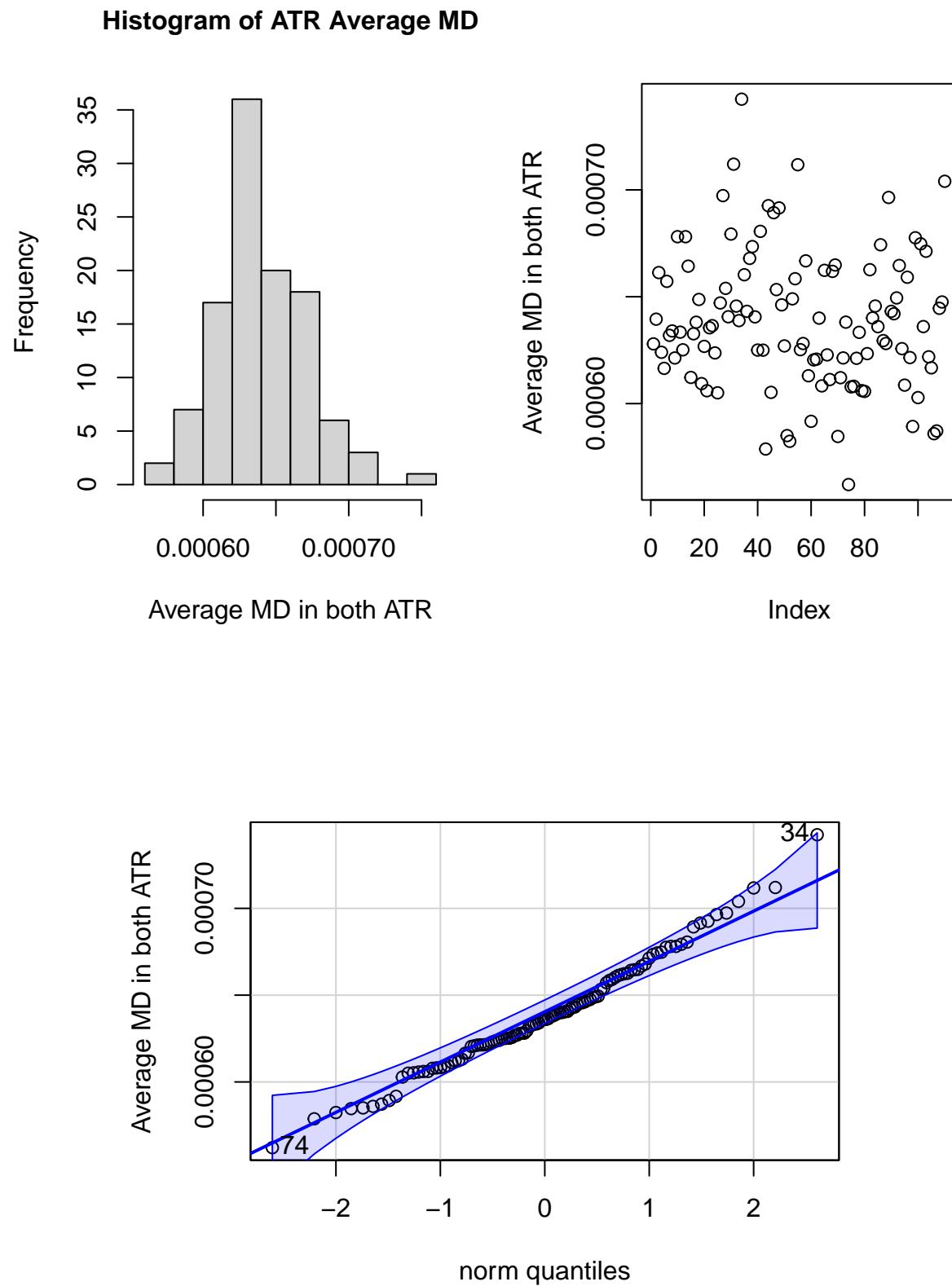
```
[1] 43 13  
[1] 11 24
```



```
[1] 17 26
```

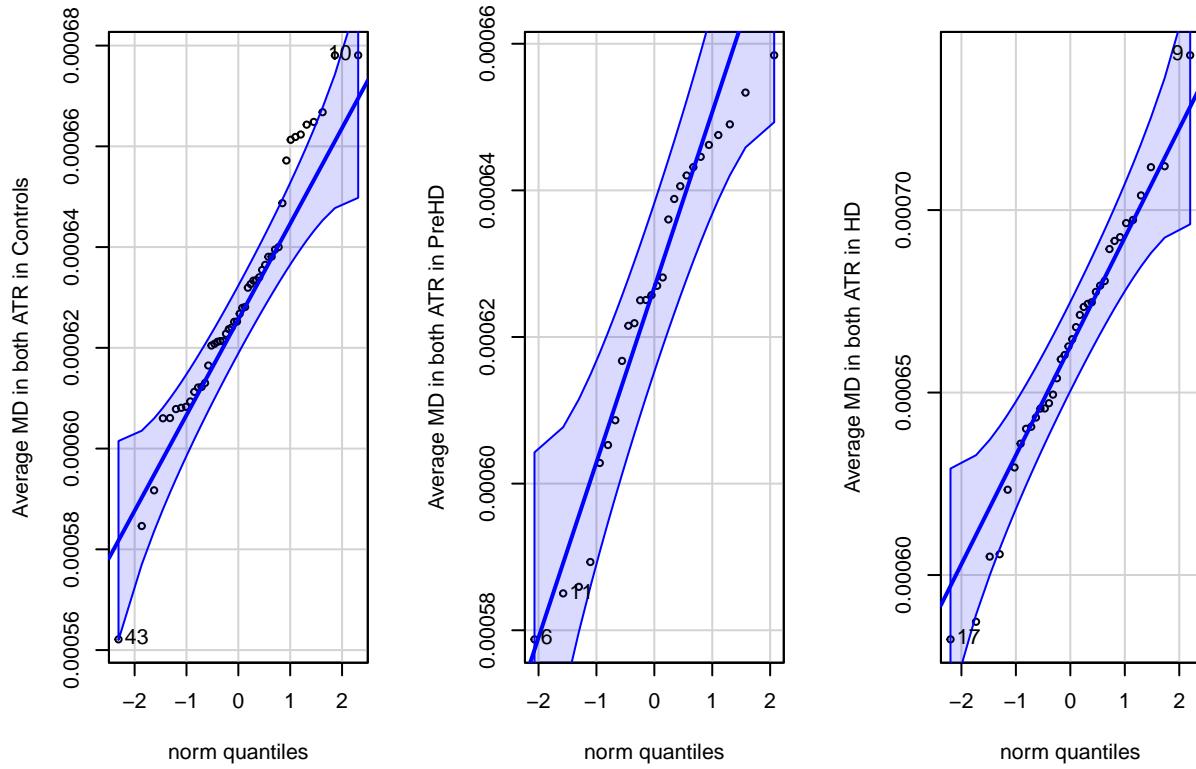


#### 4.3.2.3 MD

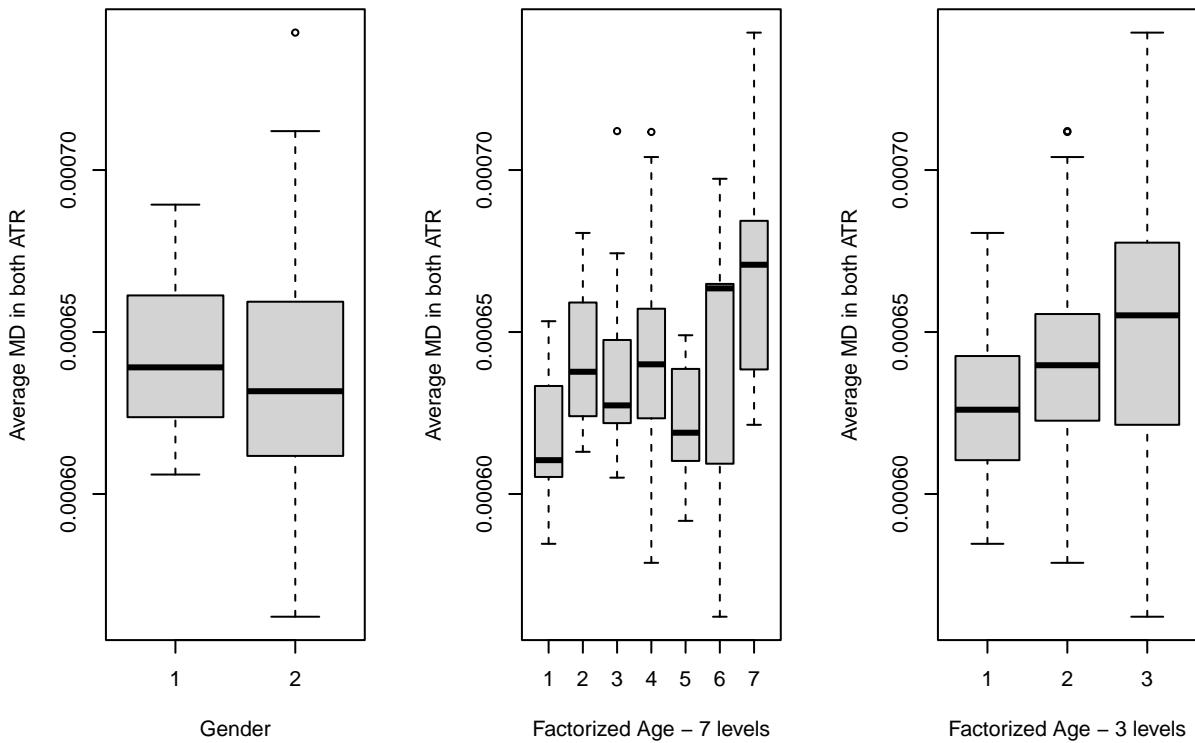
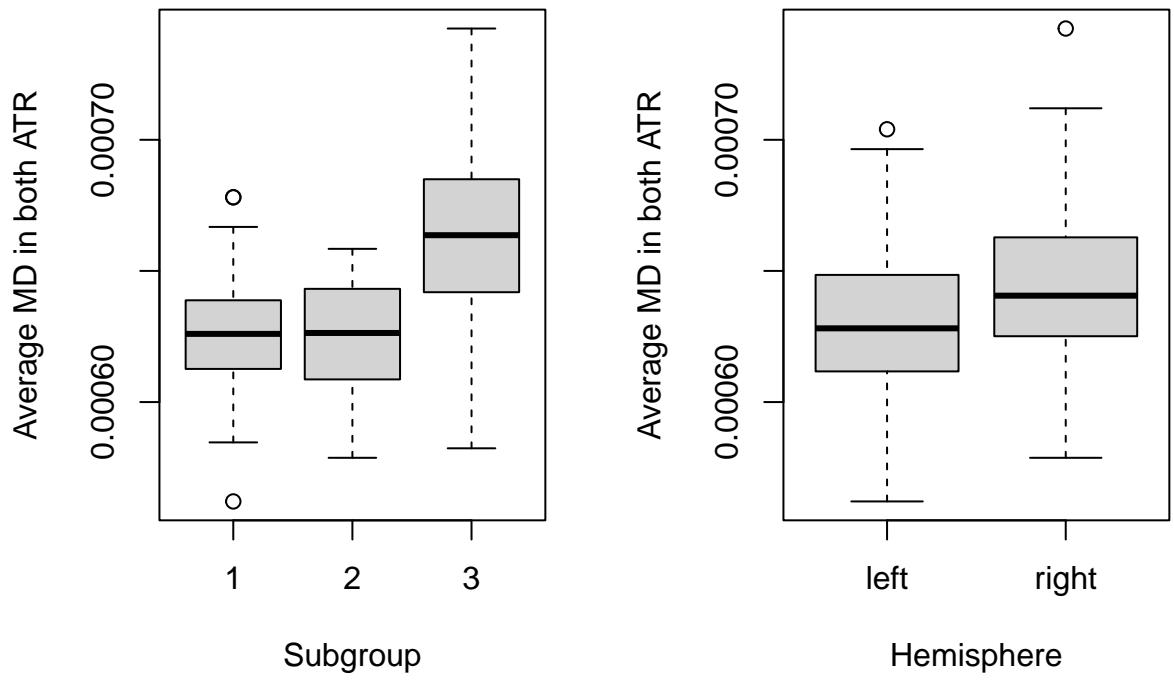


[1] 34 74

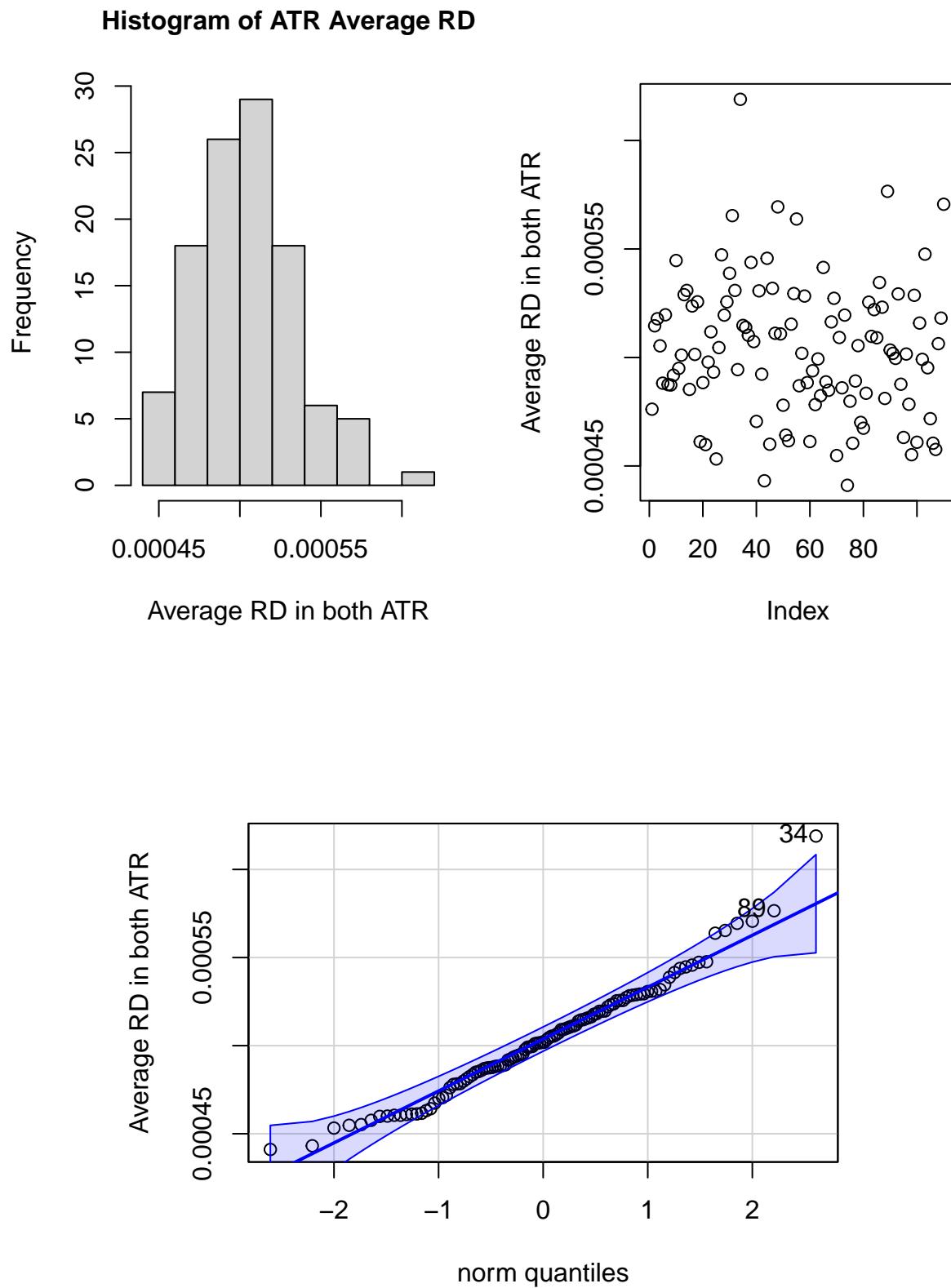
```
[1] 43 10  
[1] 6 11
```



```
[1] 9 17
```

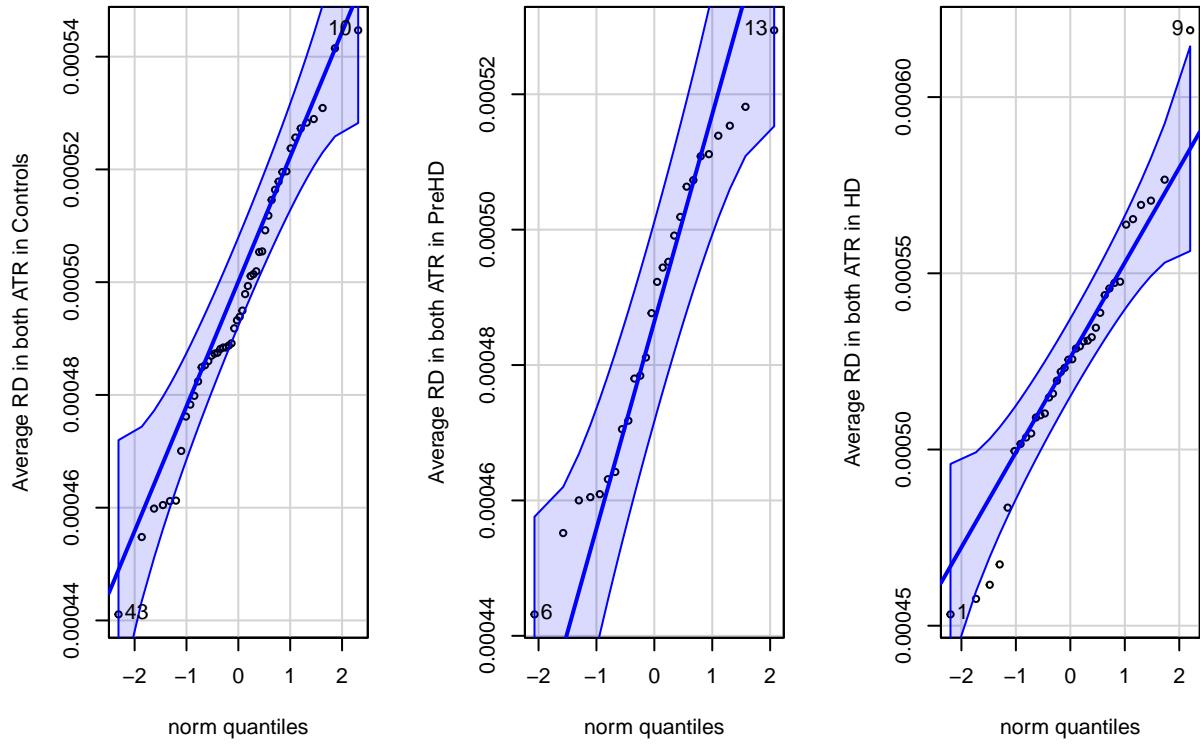


#### 4.3.2.4 RD

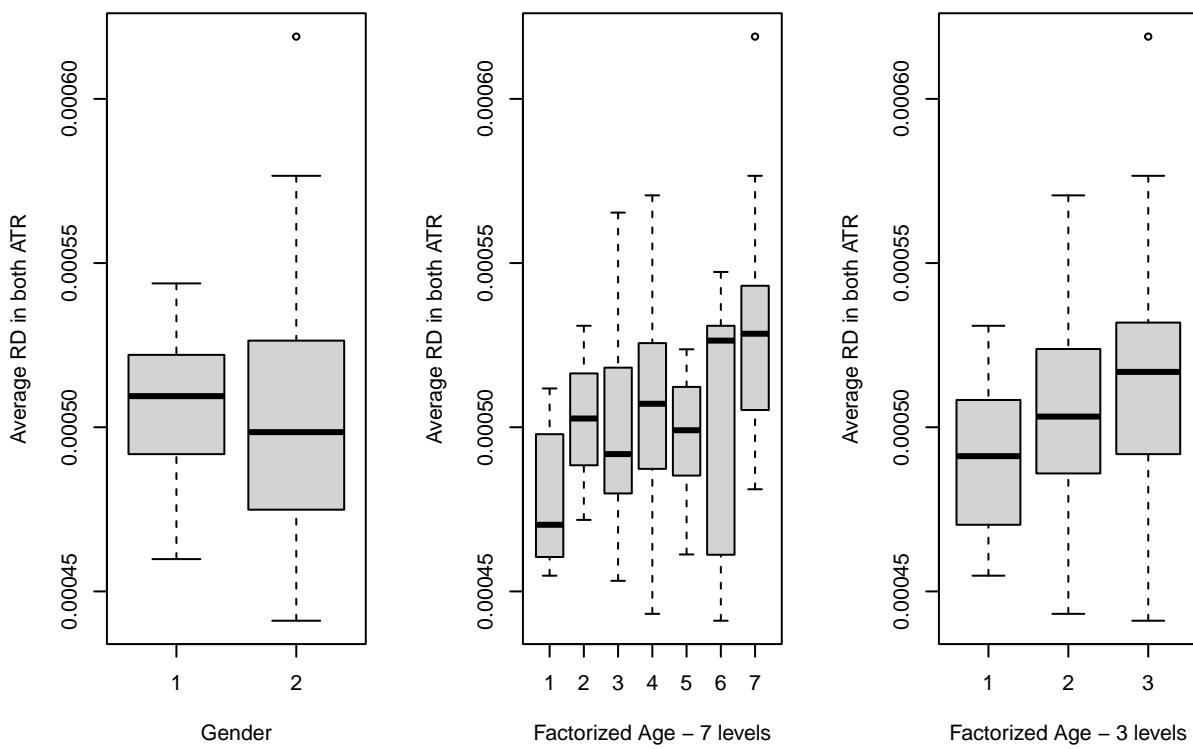
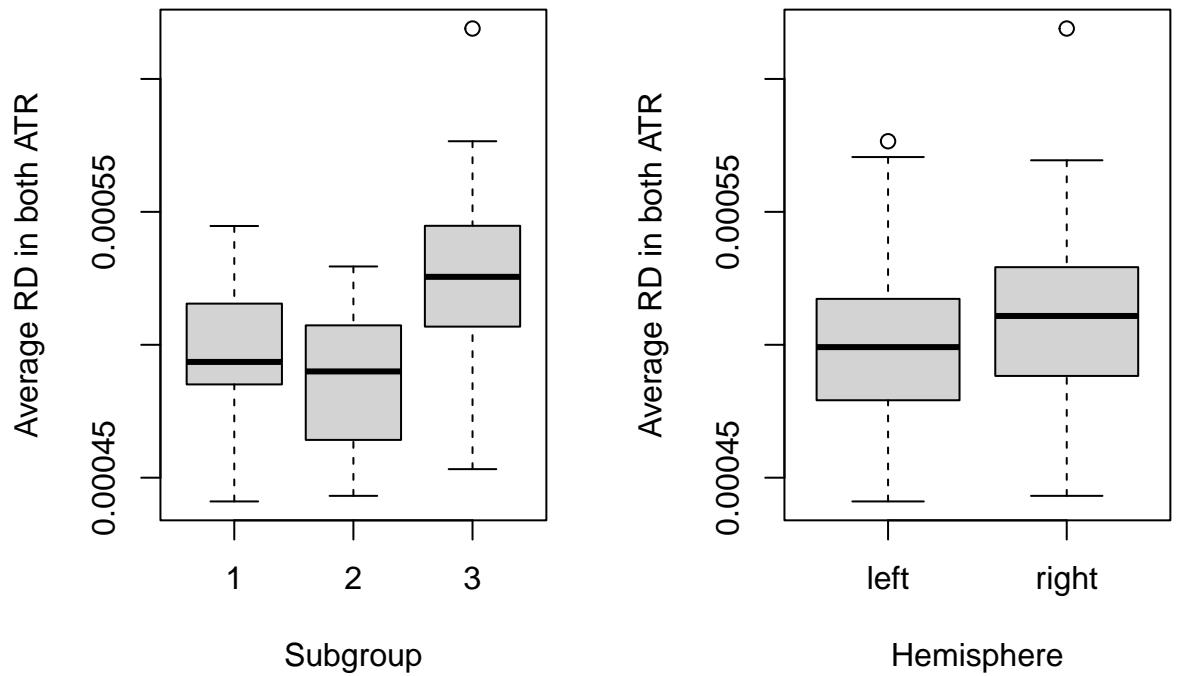


[1] 34 89

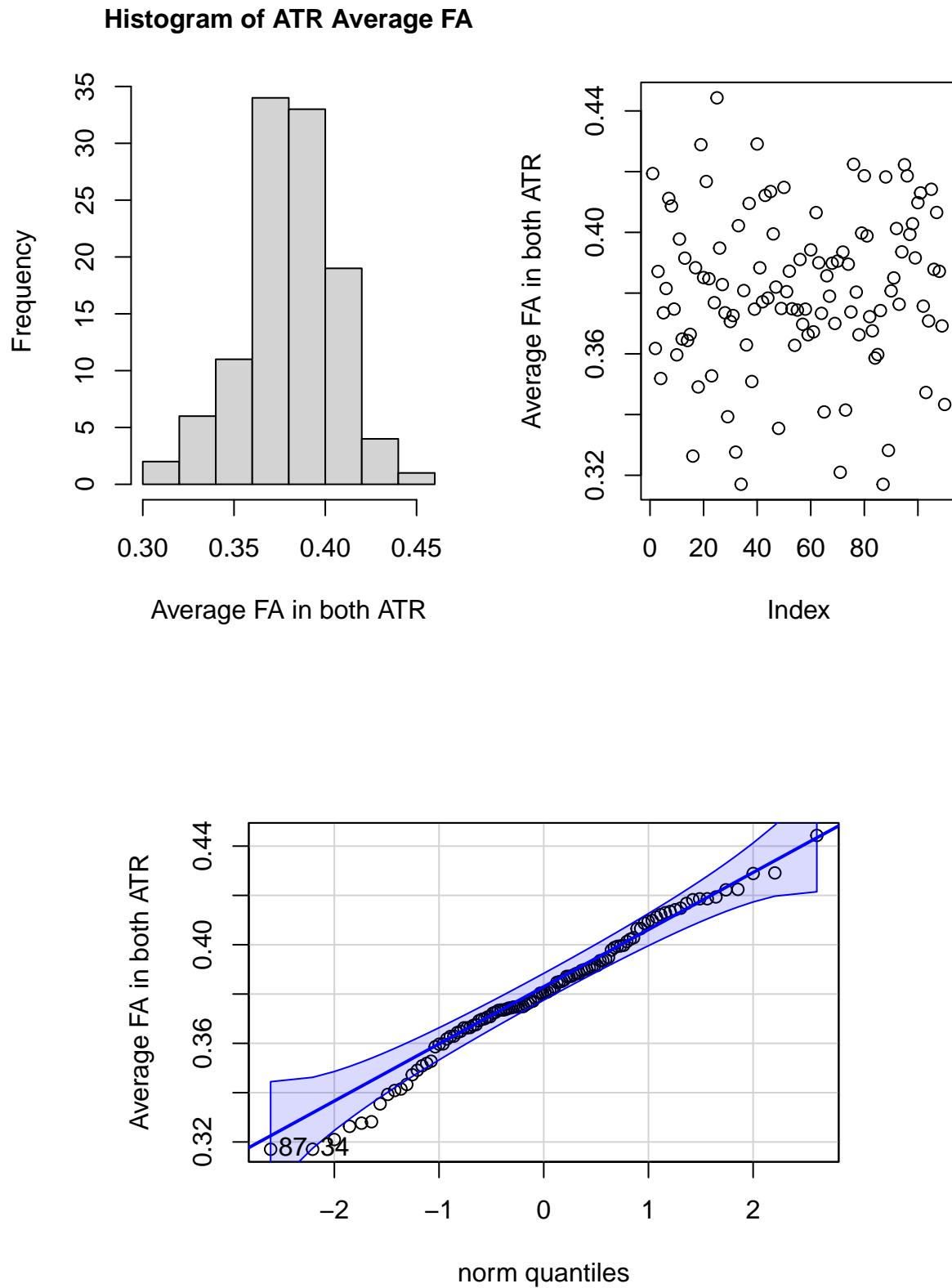
```
[1] 43 10  
[1] 6 13
```



```
[1] 9 1
```

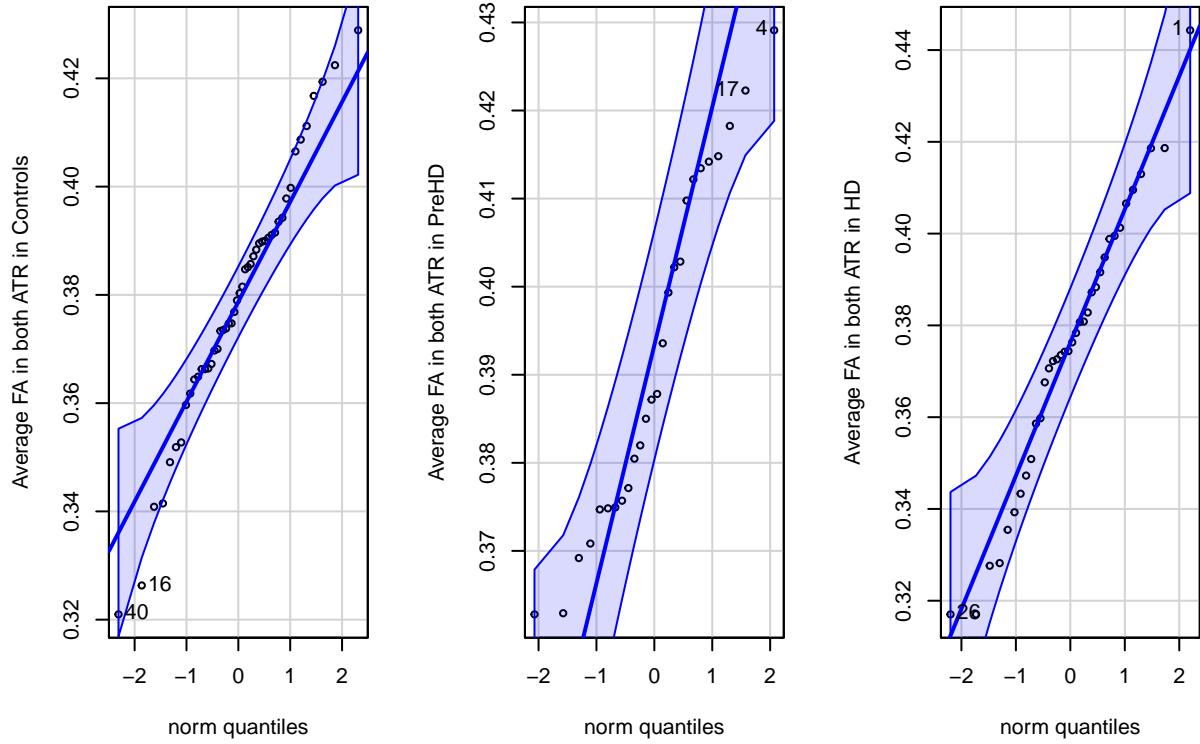


#### 4.3.2.5 FA

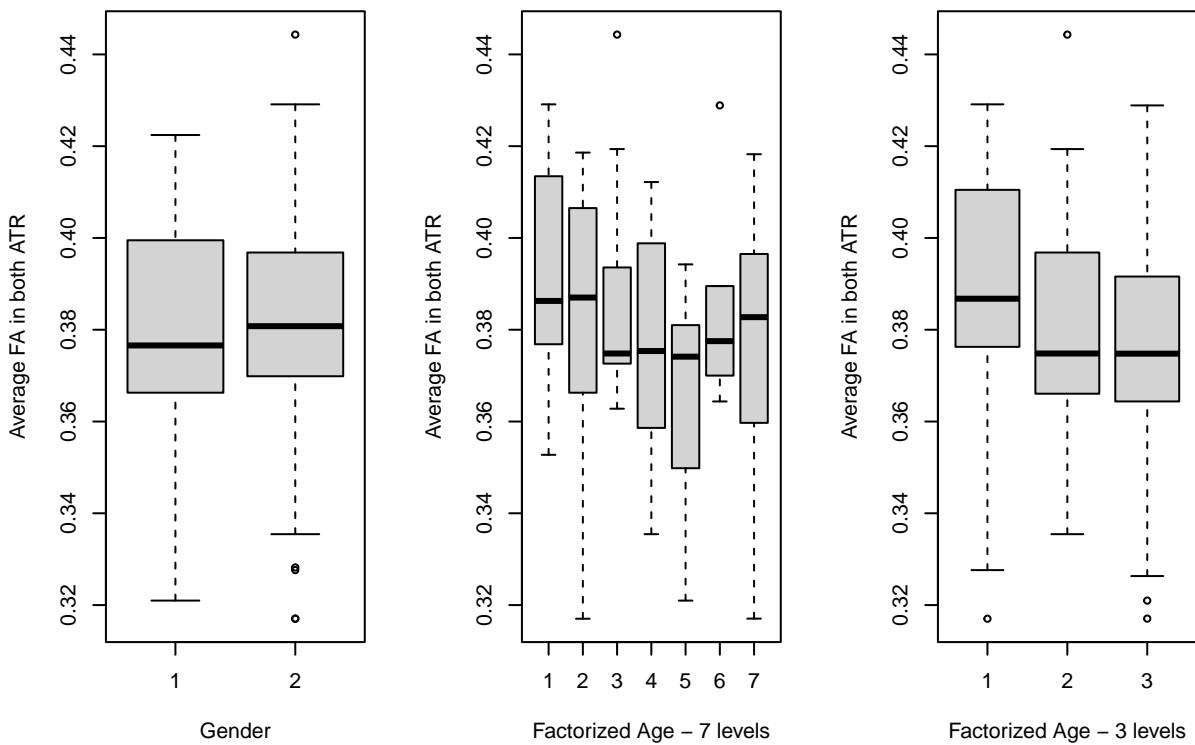
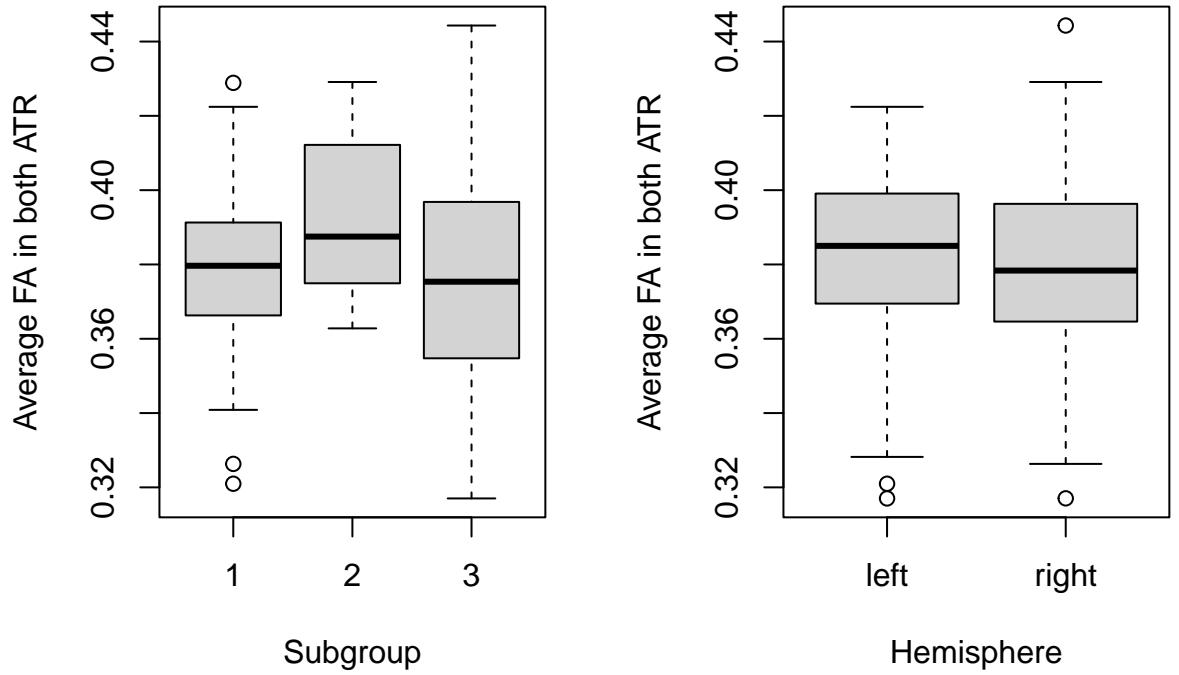


[1] 87 34

```
[1] 40 16  
[1] 4 17
```

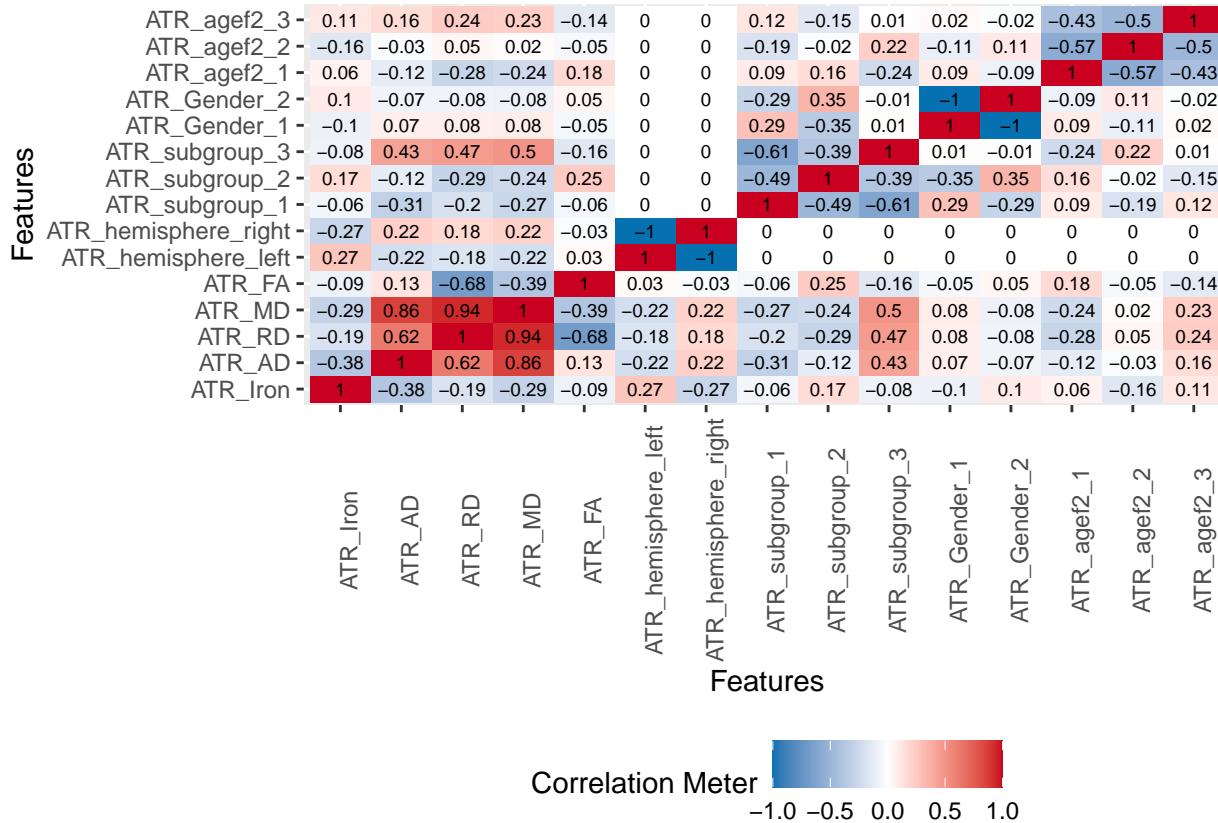


```
[1] 1 26
```



### 4.3.3 PLOT de CORRELACIONES

Previamente, a generar los diferentes modelos estadísticos, se genera un plot de correlaciones entre todas las variables.



En este plot, se objetivan los siguientes fenómenos:

- Las métricas AD, RD y MD están altamente correlacionadas entre sí;
- El ATR derecho se correlaciona con las métricas AD, RD y MD;
- El ATR izquierdo se correlaciona con la métrica de hierro;
- El subgrupo control se correlaciona con el género masculino, con los grupos de edad 1 y 3 (de los 3 niveles) y 1, 5 y 6 (de los 7 niveles);
- El subgrupo preHD se correlaciona con las métricas de hierro y FA, género femenino, con los grupos de edad 1 (de los 3 niveles) y 1 y 3 (de los 7 niveles);
- El subgrupo HD se correlaciona con las métricas AD, RD y MD, con los grupos de edad 2 (de los 3 niveles) y 4 y 7 (de los 7 niveles);

Habiendo expuesto lo anterior, se objetiva que tanto género como edad son covariables o factores confusores en la relación entre las métricas y los subgrupos, por tanto dichas variables se van a introducir en los modelos para controlar su efecto. Se utilizará la edad factorizada de 3 niveles dado que presenta una n por nivel más comparable entre niveles.

En relación a la lateralidad (ATR derecho e izquierdo), la correlación entre ATR derecho y métricas de difusividad y entre ATR izquierdo y métrica de hierro puede ser debido a que los pacientes HD sintomáticos tengan mayor afectación del ATR derecho y que los sujetos preHD presintomáticos tengan mayor acúmulo de hierro en el ATR izquierdo, pero, también puede ser debido a una fisiología diferencial entre ambos ATRs. Por tanto, es una relación que se debe de valorar detalladamente, tanto como factor independiente (hemisferio) como en interacción con subgrupo.

## 4.4 Estadística descriptiva matemática

### 4.4.1 Media, mediana, desviación estándar, kurtosis y skewness

Se procede a determinar la **media**, la **mediana**, la **desviación estándar**, la **kurtosis** y el **skewness** (se ha consultado la siguiente web para su interpretación Skewness and Kurtosis in Statistics).

En función de los valores de *skewness*, una distribución puede ser:

- **Simétrica**, con valores de *skewness* entre -0.5 y 0.5, con la media y la mediana siendo similares entre sí;
- **Asimétrica positiva** (*right-skewed* o *right-tailed*), cuyo histograma muestra una cola derecha más larga concentrándose la mayoría de las observaciones en la cola izquierda y con una media mayor que la mediana.
- **Asimétrica negativa** (*left-skewed* o *left-tailed*), cuyo histograma muestra una cola izquierda más larga concentrándose la mayoría de las observaciones en la cola derecha y con una mediana mayor que la media.

La asimetría se gradará en función de los valores de *skewness* como moderada con valores entre 0.5 y 1 o -0.5 y -1 y como alta/severa con valores mayores a 1 o inferiores a -1.

En función de los valores de *kurtosis*, una distribución puede ser:

- **Mesokúrtica**, en el caso de la distribución normal, con valores de kurtosis próximos a 3;
- **Leptokúrtica** (*kurtosis* positiva con valores mayores de 3), con una distribución concentrada de valores generando colas gruesas y pico afilado;
- **Platikúrtica** (*kurtosis* negativa con valores menores de 3), con una distribución que muestra un pico más bajo y ancho y colas más delgadas.

Cuadro 64: Media, mediana, desviación estándar, skewness y kurtosis

Variables	Iron	AD	MD	RD	FA
Mean	0.02406	0.00091	0.00064	5e-04	0.3809
Median	0.02404	0.00091	0.00064	5e-04	0.3806
Standard Deviation	0.00176	4e-05	3e-05	3e-05	0.02578
Skewness	0.54797	0.17205	0.4172	0.51936	-0.26755
Kurtosis	1.23628	-0.17188	0.30782	0.68273	-0.02898

Remarcar que los valores medios y desviaciones estándar de las 3 métricas de difusividad presentan valores cercanos a cero (unidad de medida  $mm^2/s$ ) y esto puede conllevar problemas con algunos de los métodos de análisis estadístico.

Los valores de **hierro** presentan una asimetría moderada positiva con valores de *skewness* entorno a 0.5 (exactamente valor de 0.55) y son platikúrticos (*kurtosis* negativa con valores de inferiores a 3, exactamente valor de 1.24).

Los valores de **difusividad axial** presenta una simetría con valores de *skewness* entre -0.5 y 0.5 (exactamente valor de 0.17) y son platikúrticos (*kurtosis* negativa con valores de inferiores a 3, exactamente valor de -0.17).

Los valores de **difusividad media** presenta una simetría con valores de *skewness* entre -0.5 y 0.5 (exactamente valor de 0.42) y son platikúrticos (*kurtosis* negativa con valores de inferiores a 3, exactamente valor de 0.31).

Los valores de **difusividad radial** presenta una asimetría moderada positiva con valores de *skewness* entorno a 0.5 (exactamente valor de 0.52) y son platikúrticos (*kurtosis* negativa con valores de inferiores a 3, exactamente valor de 0.68).

Los valores de **anisotropía fraccional** presenta una simetría con valores de *skewness* entre -0.5 y 0.5 (exactamente valor de -0.27) y son platikúrticos (*kurtosis* negativa con valores de inferiores a 3, exactamente valor de -0.03).

#### 4.4.2 Normalidad Univariante - Homocedasticidad de cada métrica para cada factor

Se procede a **valorar la normalidad** de los valores de las diferentes métricas de RM y la **homocedasticidad** de dichas métricas en los niveles de las variables categóricas de interés (subgrupo, hemisferio, edad\_factorizada, género).

La normalidad se puede obtener de una de las dos funciones siguientes: `shapiro.test` del paquete `stats` o `shapiro_test` del paquete `rstatix`. Los resultados son coincidentes.

La homocedasticidad se obtiene de la función `fligner.test` del paquete `stats`.

Cuadro 65: Valoración de la normalidad de las métricas de RM con la función shapiro test

	Iron	AD	MD	RD	FA
W-Statistic	0.9685	0.9883	0.9848	0.9764	0.9837
p-value	0.0104	0.4572	0.2474	0.0476	0.2008

Cuadro 66: Valoración de la homocedasticidad de las métricas de RM para las variables independientes

Variables	Iron	AD	MD	RD	FA
<b>Subgroup</b>					
Statistic	2.4862	9.0067	6.4606	2.3326	2.9282
df	2.0000	2.0000	2.0000	2.0000	2.0000
p-value	0.2885	0.0111	0.0395	0.3115	0.2313
<b>Hemisphere</b>					
Statistic	6.0899	0.3478	0.1501	0.2007	0.0001
df	1.0000	1.0000	1.0000	1.0000	1.0000
p-value	0.0136	0.5554	0.6985	0.6541	0.9917
<b>Gender</b>					
Statistic	0.6202	0.0554	4.1329	8.7985	0.0783
df	1.0000	1.0000	1.0000	1.0000	1.0000
p-value	0.4310	0.8139	0.0421	0.0030	0.7796
<b>Factorized age</b>					
Statistic	2.0986	3.2632	5.7135	1.7968	0.2172
df	2.0000	2.0000	2.0000	2.0000	2.0000
p-value	0.3502	0.1956	0.0575	0.4072	0.8971

Cuadro 67: Valoración de la homocedasticidad de las métricas de RM para las variables independientes

Variables	Iron	AD	MD	RD	FA
<b>Subgrupo</b>					
Statistic	2.4862	9.0067	6.4606	2.3326	2.9282
df	2.0000	2.0000	2.0000	2.0000	2.0000
p-value	0.2885	0.0111	0.0395	0.3115	0.2313
<b>Hemisferio</b>					
Statistic	6.0899	0.3478	0.1501	0.2007	0.0001
df	1.0000	1.0000	1.0000	1.0000	1.0000
p-value	0.0136	0.5554	0.6985	0.6541	0.9917
<b>Género</b>					
Statistic	0.6202	0.0554	4.1329	8.7985	0.0783
df	1.0000	1.0000	1.0000	1.0000	1.0000
p-value	0.4310	0.8139	0.0421	0.0030	0.7796
<b>Edad factorizada (7 niveles)</b>					
Statistic	5.1629	5.2749	4.2930	4.4586	2.0373
df	6.0000	6.0000	6.0000	6.0000	6.0000
p-value	0.5231	0.5091	0.6371	0.6149	0.9162
<b>Edad factorizada (3 niveles)</b>					

Statistic	2.0986	3.2632	5.7135	1.7968	0.2172
df	2.0000	2.0000	2.0000	2.0000	2.0000
p-value	0.3502	0.1956	0.0575	0.4072	0.8971

Los valores de **hierro** no son normales, pero presentan homocedasticidad de varianza entre los niveles de las variables categóricas de interés (subgrupo, edad factorizada, género), a excepción del hemisferio.

Los valores de **difusividad axial** son normales, pero presentan homocedasticidad de varianza entre los niveles de las variables categóricas de interés (hemisferio, edad factorizada, género), a excepción del subgrupo.

Los valores de **difusividad media** son normales, con homocedasticidad de varianza entre los niveles de las variables categóricas hemisferio y edad factorizada y con heterocedasticidad de varianza entre los niveles de las variables categóricas subgrupo y género.

Los valores de **difusividad radial** no son normales, pero presentan homocedasticidad de varianza entre los niveles de las variables categóricas de interés (subgrupo, hemisferio, edad factorizada), a excepción del género.

Los valores de **anisotropía fraccional** son normales, con homocedasticidad de varianza entre los niveles de las variables categóricas de interés (subgrupo, hemisferio, edad factorizada, género).

Por tanto, de todas las métricas, el hierro y la difusividad radial no siguen una distribución normal, y, todas las métricas, a excepción de la anisotropía fraccional, presentan heterocedasticidad para alguna de las variables categóricas independientes de interés. Posteriormente, en la generación de los modelos, se valorará el comportamiento de los residuos para decidir sobre la eliminación de outliers.

Se intentará trabajar con todos los sujetos en los diferentes modelos porque mi interés es tener muestras homogéneas entre las diferentes métricas. En el caso que no se cumplan los criterios para aplicar métodos paramétricos, se procederá a generar métodos no paramétricos. De todas formas, se generarán ambos métodos para todas las métricas dado que también es de mi interés que los métodos sean homogéneos entre todas las métricas.

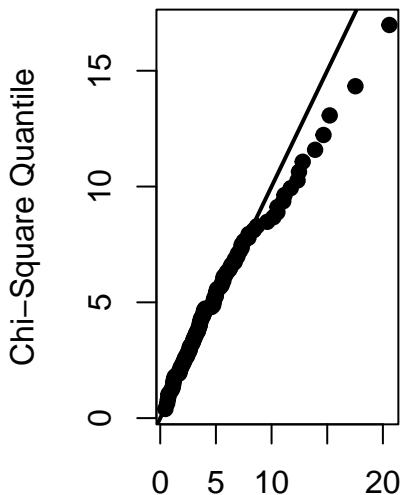
#### 4.4.3 Normalidad multivariante

Siguiendo el patrón del análisis de normalidad multivariante aplicado para cada ATR por separado, se va a utilizar la función `mvn` del paquete MVN para generar los tests de normalidad multivariante. Se va testear la normalidad multivariante en 3 supuestos:

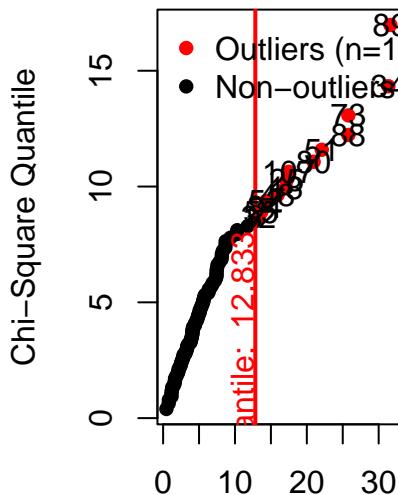
- Las 5 métricas de RM conjuntamente;
- Las 3 métricas de RM (AD, RD, MD) que parecen correlacionarse con el subgrupo HD;
- Las 2 métricas de RM (iron, FA) que parecen correlacionarse con el subgrupo preHD.

```
# Para las 5 métricas de RM conjuntamente
library(MVN)
par(mfrow = c(1, 2))
ATR_avg<-ATR_DifIron_Df[,c('ATR_subgroup','ATR_hemisphere','ATR_Gender','ATR_agef2',
                           'ATR_Iron','ATR_AD','ATR_RD','ATR_MD','ATR_FA')]
# Valoración de la normalidad multivariante con todos los sujetos
ATR_mvnmv<-mvn(ATR_avg[-c(1:4)], mvnTest = 'royston', covariance = TRUE, desc = TRUE,
                  univariateTest = 'SW', univariatePlot = 'qq',
                  multivariatePlot = 'qq', multivariateOutlierMethod = 'quan',
                  showOutliers = TRUE, showNewData = TRUE)
```

**Chi-Square Q-Q Plot**



**Chi-Square Q-Q Plot**



Squared Mahalanobis Distance Robust Squared Mahalanobis Distance

```

index<-rownames(ATR_mvnm$newData)
ATR_avg_WoOut<-ATR_avg[c(index),]
table(ATR_avg$ATR_subgroup)

1 2 3
48 26 36
table(ATR_avg_WoOut$ATR_subgroup)

1 2 3
46 24 26

```

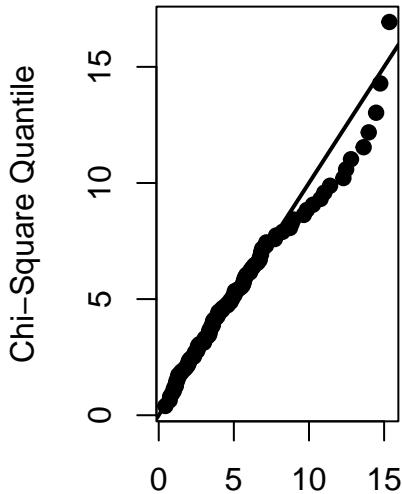
```

# Valoración de la normalidad multivariante con eliminación de los dos outliers
# más extremos, con distancias de Mahalanobis mayores de 30
par(mfrow = c(1, 2))
print(ATR_mvnm$multivariateOutliers,max = 6)
  Observation Mahalanobis Distance Outlier
89         89        31.632    TRUE
34         34        31.349    TRUE
[ reached 'max' / getOption("max.print") -- omitted 12 rows ]
newATR_avg<-ATR_avg[-c(34,89),]
ATR_avg[c(34,89),]

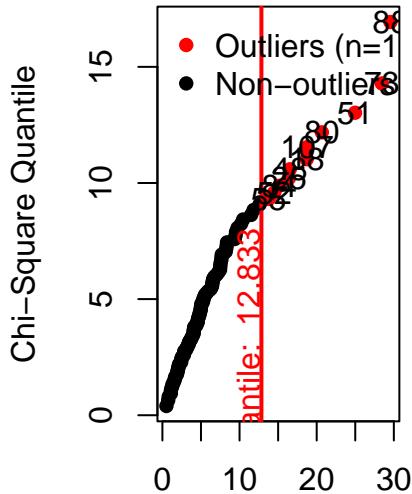
  ATR_subgroup ATR_hemisphere ATR_Gender ATR_ägef2 ATR_Iron       ATR_AD
34            3           right          2        3 0.0247215 0.000989331
89            3           left          2        3 0.0285345 0.000936180
  ATR_RD      ATR_MD      ATR_FA
34 0.000618966 0.000742420 0.317049
89 0.000576567 0.000696437 0.328200
newATR_mvnm<-mvn(newATR_avg[-c(1:4)], mvnTest = 'royston', covariance = TRUE,
desc = TRUE, univariateTest = 'SW', univariatePlot = 'qq',
multivariatePlot = 'qg', multivariateOutlierMethod = 'quan',
showOutliers = TRUE, showNewData = TRUE)

```

### Chi-Square Q-Q Plot



### Chi-Square Q-Q Plot



Squared Mahalanobis Distance Robust Squared Mahalanobis Distance

```
ATR_DifIron_Df_WoOut <- ATR_DifIron_Df [ -c(34, 89), ]
```

Cuadro 68: Comparación de la normalidad multivariante sin y con eliminación de outliers habiendo considerado las 5 métricas conjuntamente

	Test	H	p value	MVN
Without elimination of outliers	Royston	11.248908	0.0163310	NO
With elimination of outliers	Royston	6.674028	0.1236385	YES

Cuadro 69: Comparación de la normalidad univariante sin y con eliminación de outliers para las 5 métricas

Without elimination of outliers					With elimination of outliers				
Test	Variable	Statistic	p value	Normality	Test	Variable	Statistic	p value	Normality
Shapiro-Wilk	ATR_Iron	0.9685	0.0104	NO	Shapiro-Wilk	ATR_Iron	0.9711	0.0186	NO
Shapiro-Wilk	ATR_AD	0.9883	0.4572	YES	Shapiro-Wilk	ATR_AD	0.9886	0.4960	YES
Shapiro-Wilk	ATR_RD	0.9764	0.0476	NO	Shapiro-Wilk	ATR_RD	0.9867	0.3625	YES
Shapiro-Wilk	ATR_MD	0.9848	0.2474	YES	Shapiro-Wilk	ATR_MD	0.9894	0.5576	YES
Shapiro-Wilk	ATR_FA	0.9837	0.2008	YES	Shapiro-Wilk	ATR_FA	0.9882	0.4665	YES

Cuadro 70: Comparación de los estadísticos descriptivos sin y con eliminación de outliers para las 5 métricas

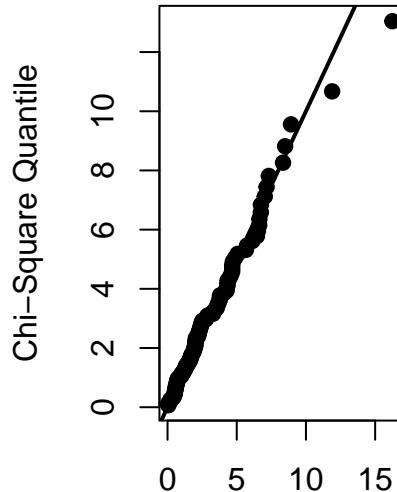
Variables	n	Mean	Std.Dev	Median	Min	Max	25th	75th	Skew	Kurtosis
<b>Without elimination of outliers</b>										
ATR_Iron	110	0.0240644	0.0017621	0.0240366	0.0191068	0.0299576	0.0228914	0.0248790	0.5479732	1.2362771
ATR_AD	110	0.0009092	0.0000430	0.0009065	0.0008042	0.0010076	0.0008809	0.0009338	0.1720546	-0.1718767
ATR_RD	110	0.0005035	0.0000314	0.0005019	0.0004411	0.0006190	0.0004839	0.0005236	0.5193648	0.6827261
ATR_MD	110	0.0006387	0.0000319	0.0006360	0.0005621	0.0007424	0.0006209	0.0006600	0.4171961	0.3078151
ATR_FA	110	0.3808985	0.0257764	0.3805980	0.3170270	0.4443100	0.3673255	0.3985767	-0.2675536	-0.0289816

With elimination of outliers											
ATR_Iron	108	0.0240169	0.0017234	0.0240009	0.0191068	0.0299576	0.0228403	0.0248494	0.5164108	1.3490253	
ATR_AD	108	0.0009082	0.0000426	0.0009061	0.0008042	0.0010076	0.0008805	0.0009324	0.1852667	-0.1163038	
ATR_RD	108	0.0005017	0.0000287	0.0005017	0.0004411	0.0005706	0.0004832	0.0005223	0.1146438	-0.4685656	
ATR_MD	108	0.0006372	0.0000300	0.0006357	0.0005621	0.0007120	0.0006207	0.0006586	0.2093360	-0.1351374	
ATR_FA	108	0.3819777	0.0247307	0.3807780	0.3170270	0.4443100	0.3687887	0.3989600	-0.1893995	-0.0093789	

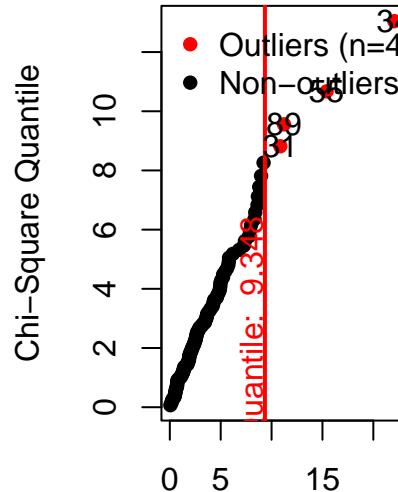
Considerando las 5 métricas conjuntamente, se observa que no existe normalidad multivariante ni normalidad univariante para el hierro y la difusividad radial. Si se eliminaran todos los outliers detectados por la función mvn, se eliminarían 2 controles, 2 preHD y 10 HD, siendo más de un 10% de las observaciones. Decido eliminar exclusivamente los dos sujetos (dos HD, el mismo sujeto con su medida para cada ATR) con mayor distancia de Mahalanobis (mayor de 30), y, con ello, se consigue una normalidad multivariante y una normalidad univariante para todas las métricas a excepción del hierro.

```
# Para las 3 métricas de RM correlacionadas con HD: AD, RD, MD
par(mfrow = c(1, 2))
ATR_avg<-ATR_DifIron_Df[,c('ATR_subgroup', 'ATR_hemisphere', 'ATR_Gender', 'ATR_agef2',
                           'ATR_Iron', 'ATR_AD', 'ATR_RD', 'ATR_MD', 'ATR_FA')]
# Valoración de la normalidad multivariante con todos los sujetos
ATR_mvnmv(ATTR_avg[-c(1:5,9)], mvnTest = 'royston', covariance = TRUE, desc = TRUE,
           univariateTest = 'SW', univariatePlot = 'qq',
           multivariatePlot = 'qq', multivariateOutlierMethod = 'quan',
           showOutliers = TRUE, showNewData = TRUE)
```

Chi-Square Q-Q Plot



Chi-Square Q-Q Plot



Squared Mahalanobis Distance Robust Squared Mahalanobis Distance

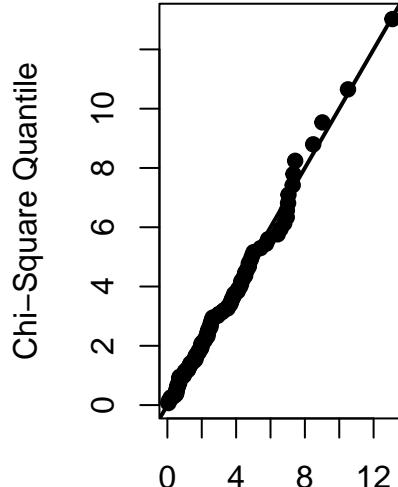
```
index<-rownames(ATR_mvnmv$newData)
ATR_avg_WoOut<-ATR_avg[c(index),]
table(ATR_avg$ATR_subgroup)

 1  2  3
48 26 36
table(ATR_avg_WoOut$ATR_subgroup)
```

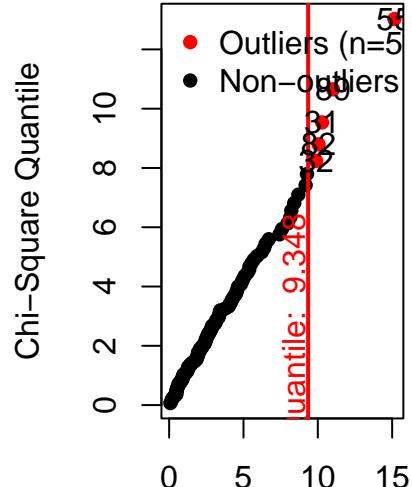
```
1 2 3
48 26 32
```

```
# Valoración de la normalidad multivariante con eliminación del outlier más
# extremo, con distancia de Mahalanobis mayor de 20
par(mfrow = c(1, 2))
print(ATR_mvnmultivariateOutliers,max = 6)
Observation Mahalanobis Distance Outlier
34      34          22.071    TRUE
55      55          15.405    TRUE
[ reached 'max' / getOption("max.print") -- omitted 2 rows ]
newATR_avg<-ATR_avg[-c(34),]
ATR_avg[c(34),]
  ATR_subgroup ATR_hemisphere ATR_Gender ATR_agef2 ATR_Iron       ATR_AD
34            3           right         2        3 0.0247215 0.000989331
  ATR_RD     ATR_MD     ATR_FA
34 0.000618966 0.00074242 0.317049
newATR_mvnmv(newATR_avg[-c(1:5,9)], mvnTest = 'royston', covariance = TRUE,
  desc = TRUE, univariateTest = 'SW', univariatePlot = 'qq',
  multivariatePlot = 'qq', multivariateOutlierMethod = 'quan',
  showOutliers = TRUE, showNewData = TRUE)
```

**Chi-Square Q-Q Plot**



**Chi-Square Q-Q Plot**



Squared Mahalanobis Distance Robust Squared Mahalanobis Distance

```
ATR_DifIron_Df_WoOut_ADRDMD<-ATR_DifIron_Df[-c(34), -c(7,11,12)]
ATR_DifIron_Df_ADRDMD<-ATR_DifIron_Df[-c(7,11,12)]
```

Cuadro 71: Comparación de la normalidad multivariante sin y con eliminación de outliers habiendo considerado las métricas AD, RD, MD conjuntamente

	Test	H	p value	MVN
Without elimination of outliers	Royston	4.502439	0.1044047	YES
With elimination of outliers	Royston	1.249133	0.5409830	YES

Cuadro 72: Comparación de la normalidad univariante sin y con eliminación de outliers para las métricas AD, RD, MD

Without elimination of outliers					With elimination of outliers				
Test	Variable	Statistic	p value	Normality	Test	Variable	Statistic	p value	Normality
Shapiro-Wilk	ATR_AD	0.9883	0.4572	YES	Shapiro-Wilk	ATR_AD	0.9892	0.5368	YES
Shapiro-Wilk	ATR_RD	0.9764	0.0476	NO	Shapiro-Wilk	ATR_RD	0.9861	0.3206	YES
Shapiro-Wilk	ATR_MD	0.9848	0.2474	YES	Shapiro-Wilk	ATR_MD	0.9886	0.4863	YES

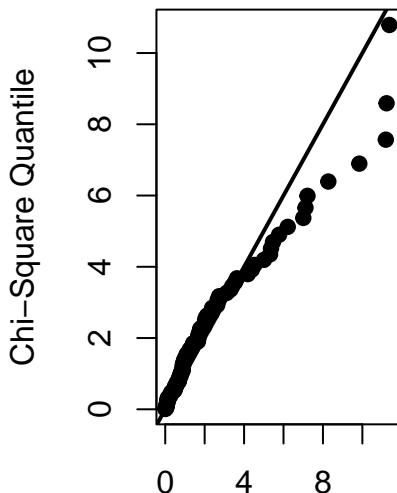
Cuadro 73: Comparación de los estadísticos descriptivos sin y con eliminación de outliers para las métricas AD, RD, MD

Variables	n	Mean	Std.Dev	Median	Min	Max	25th	75th	Skew	Kurtosis
<b>Without elimination of outliers</b>										
ATR_AD	110	0.0009092	4.30e-05	0.0009065	0.0008042	0.0010076	0.0008809	0.0009338	0.1720546	-0.1718767
ATR_RD	110	0.0005035	3.14e-05	0.0005019	0.0004411	0.0006190	0.0004839	0.0005236	0.5193648	0.6827261
ATR_MD	110	0.0006387	3.19e-05	0.0006360	0.0005621	0.0007424	0.0006209	0.0006600	0.4171961	0.3078151
<b>With elimination of outliers</b>										
ATR_AD	109	0.0009084	4.25e-05	0.0009063	0.0008042	0.0010076	0.0008805	0.0009328	0.1696858	-0.1148554
ATR_RD	109	0.0005024	2.95e-05	0.0005019	0.0004411	0.0005766	0.0004835	0.0005232	0.1862341	-0.3767993
ATR_MD	109	0.0006378	3.04e-05	0.0006360	0.0005621	0.0007120	0.0006208	0.0006591	0.2140573	-0.1911274

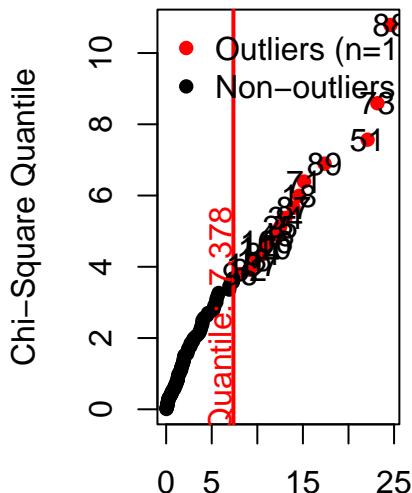
**Considerando las métricas AD, RD y MD conjuntamente**, con todos los sujetos se observa que existe normalidad multivariante y normalidad univariante para las difusividades axial y media (no así para la difusividad radial), y, si se eliminaran todos los outliers detectados por la función `mvn`, se eliminarían 4 pacientes HD. Para generar una similitud con el análisis anterior (considerando las 5 métricas conjuntamente), se procede a eliminar el outlier con mayor distancia de Mahalanobis (sujeto 34, un HD, con una distancia mayor de 20), y, con ello, se conserva la normalidad multivariante y se consigue normalidad univariante para todas las métricas.

```
# Para las 2 métricas de RM correlacionadas con preHD: Iron, FA
par(mfrow = c(1, 2))
ATR_avg<-ATR_DifIron_Df[,c('ATR_subgroup', 'ATR_hemisphere', 'ATR_Gender', 'ATR_agef2',
                           'ATR_Iron', 'ATR_AD', 'ATR_RD', 'ATR_MD', 'ATR_FA')]
# Valoración de la normalidad multivariante con todos los sujetos
ATR_mvnmv<-mvn(ATR_avg[-c(1:4,6:8)], mvnTest = 'royston', covariance = TRUE, desc = TRUE,
                  univariateTest = 'SW', univariatePlot = 'qq',
                  multivariatePlot = 'qq', multivariateOutlierMethod = 'quan',
                  showOutliers = TRUE, showNewData = TRUE)
```

### Chi-Square Q-Q Plot



### Chi-Square Q-Q Plot



Squared Mahalanobis Distance      Robust Squared Mahalanobis Distance

```
index<-rownames(ATR_mvnm$newData)
ATR_avg_WoOut<-ATR_avg[c(index),]
table(ATR_avg$ATR_subgroup)

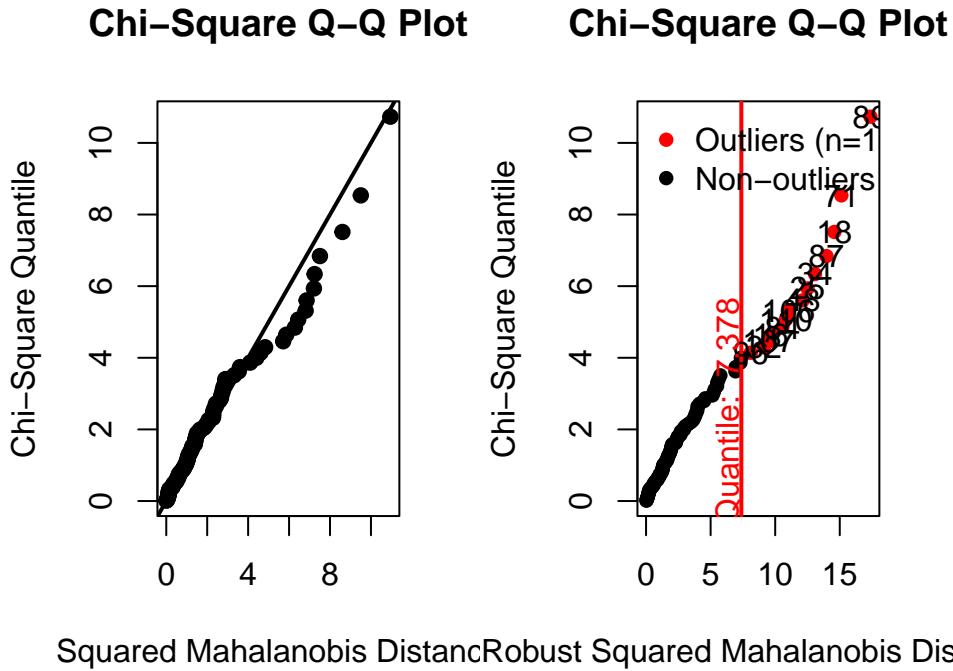
1 2 3
48 26 36
table(ATR_avg_WoOut$ATR_subgroup)

1 2 3
44 23 26
```

```
# Valoración de la normalidad multivariante con eliminación de los 3 outliers más
# extremos, con distancias de Mahalanobis mayores de 20
par(mfrow = c(1, 2))
print(ATR_mvnm$multivariateOutliers,max = 9)
  Observation Mahalanobis Distance Outlier
88          88        24.595    TRUE
73          73        23.165    TRUE
51          51        22.096    TRUE
[ reached 'max' / getOption("max.print") -- omitted 14 rows ]
newATR_avg<-ATR_avg[-c(88,73,51),]
ATR_avg[c(88,73,51),]

  ATR_subgroup ATR_hemisphere ATR_Gender ATR_aged2 ATR_Iron       ATR_AD
88            2           left         2        3 0.0290945 0.000922127
73            1           left         1        2 0.0191068 0.000875271
51            2           right        2        1 0.0299576 0.000826666
  ATR_RD      ATR_MD      ATR_FA
88 0.000481115 0.000628119 0.418249
73 0.000519527 0.000638108 0.341479
51 0.000464213 0.000585030 0.380471
newATR_mvnm<-mvn(newATR_avg[-c(1:4,6:8)], mvnTest = 'royston', covariance = TRUE,
```

```
desc = TRUE, univariateTest = 'SW', univariatePlot = 'qq',
multivariatePlot = 'qq', multivariateOutlierMethod = 'quan',
showOutliers = TRUE, showNewData = TRUE)
```



```
ATR_DifIron_Df_WoOut_IronFA<-ATR_DifIron_Df[-c(88,73,51), -c(8:10,12)]
ATR_DifIron_Df_IronFA<-ATR_DifIron_Df[,-c(8:10,12)]
```

Cuadro 74: Comparación de la normalidad multivariante sin y con eliminación de outliers habiendo considerado las métricas Iron y FA conjuntamente

	Test	H	p value	MVN
Without elimination of outliers	Royston	9.232284	0.0098907	NO
With elimination of outliers	Royston	3.576269	0.1671735	YES

Cuadro 75: Comparación de la normalidad univariante sin y con eliminación de outliers para las métricas Iron y FA

Without elimination of outliers					With elimination of outliers				
Test	Variable	Statistic	p value	Normality	Test	Variable	Statistic	p value	Normality
Shapiro-Wilk	ATR_Iron	0.9685	0.0104	NO	Shapiro-Wilk	ATR_Iron	0.9824	0.1696	YES
Shapiro-Wilk	ATR_FA	0.9837	0.2008	YES	Shapiro-Wilk	ATR_FA	0.9831	0.1934	YES

Cuadro 76: Comparación de los estadísticos descriptivos sin y con eliminación de outliers para las métricas Iron y FA

Variables	n	Mean	Std.Dev	Median	Min	Max	25th	75th	Skew	Kurtosis
<b>Without elimination of outliers</b>										
ATR_Iron	110	0.0240644	0.0017621	0.0240366	0.0191068	0.0299576	0.0228914	0.0248790	0.5479732	1.2362771

ATR_FA	110	0.3808985	0.0257764	0.3805980	0.3170270	0.4443100	0.3673255	0.3985767	-0.2675536	-0.0289816
<b>With elimination of outliers</b>										
ATR_Iron	107	0.0240087	0.0015465	0.0240214	0.0201360	0.0285345	0.0229316	0.0248128	0.4096840	0.3884127
ATR_FA	107	0.3809219	0.0256008	0.3807250	0.3170270	0.4443100	0.3674120	0.3983185	-0.2783649	0.0450823

Considerando las métricas iron y FA conjuntamente, se observa que no existe normalidad multivariante ni normalidad univariante para el hierro. Si se eliminaran todos los outliers detectados por la función `mvn`, se eliminarían 4 controles, 3 preHD y 10 HD, siendo más de un 15% de las observaciones. Para generar una similitud con los dos análisis anteriores, se procede a eliminar los outliers con mayor distancia de Mahalanobis (dos sujetos preHD 51 y 88, y, un sujeto control 73, los tres con distancias mayores a 20), y, con ello, se consigue la normalidad multivariante y la normalidad univariante para ambas métricas.

#### 4.4.4 Homocedasticidad de la matriz de varianza-covarianza

El **test de homogeneidad de las matrices de covarianza o suposición de homocedasticidad** es una suposición que debe cumplirse en el análisis multivariante, y, consiste en que múltiples grupos en un diseño experimental o test estadístico tengan matrices de covarianza iguales, es decir, las matrices de covarianza poblacional para las variables dependientes de cada grupo deben ser iguales.

Las variables pueden no ser homocedásticas por dos razones principales:

- Una variable puede ser no normal;
- Una variable puede tener una relación con la transformación de otra variable.

El **test M de Box** (o test de Box para la equivalencia de matrices de covarianza que utiliza la distribución F) es un test paramétrico que se utiliza para comparar la variación en muestras multivariadas, es decir, determina si dos o más matrices de covarianza son iguales (homogéneas).

Un **inconveniente** de dicho test es su extremada sensibilidad a las desviaciones de la normalidad, es decir, los datos deben tener una distribución normal multivariada. Además, tiene muy poca potencia para tamaños de muestra pequeños (resultados no significativos en el contexto de muestras pequeñas no indican necesariamente que las matrices de covarianza sean iguales) y es demasiado sensible para muestras de gran tamaño (resultados falsamente significativos en muestras de gran tamaño, por lo que en este contexto se recomienda un alfa más pequeño).

La **hipótesis nula** p-valor superior a 0.05 es que las matrices de covarianza observadas para las variables dependientes son iguales en todos los grupos: un estadístico M de Box no significativo indica que las matrices de covarianza son iguales.

La **hipótesis alternativa** con p-valor inferior a 0.05 indica que las matrices de covarianza son significativamente diferentes. El incumplimiento de la suposición de homogeneidad de las matrices de covarianza no siempre significa que no se puede ejecutar un análisis estadístico paramétrico. En MANOVA, aparecen **n** desiguales a medida que disminuyen los tamaños de muestra, violando la suposición de homogeneidad de covarianza; en estos casos, se debe utilizar la traza de Pillai por ser más robusta que otros tests como la traza de Hotelling o la Lambda de Wilks.

Como se hizo en el caso de los tests de homocedasticidad univariante, se procede a calcular la homocedasticidad multivariante (para las 5 métricas, para las métricas AD-RD-MD y para las métricas Iron-FA) mediante los test M de Box para las variables independientes subgrupo, hemisferio, género y edad factorizada, con los dataframes sin y con la eliminación de outliers.

Cuadro 77: Homocedasticidad de las matrices de covarianza, SIN eliminación de outliers

VARIABLES	SUBGRUPO	HEMISFERIO	GÉNERO	EDAD
<b>Para las 5 métricas</b>				
Chi-square	55.198	15.054	35.815	50.223
Df	30	15	15	30
P-value	0.003	0.448	0.002	0.012
<b>Para las métricas AD-RD-MD</b>				

Chi-square	24.436	1.467	25.375	12.132
Df	12	6	6	12
P-value	0.018	0.962	0	0.435
<b>Para las métricas Iron-FA</b>				
Chi-square	10.807	6.14	0.083	6.32
Df	6	3	3	6
P-value	0.095	0.105	0.994	0.388

Cuadro 78: Homocedasticidad de las matrices de covarianza, CON eliminación de outliers

Variables	Subgrupo	Hemisferio	Género	Edad
<b>Para las 5 métricas</b>				
Chi-square	47.417	13.741	27.87	56.596
Df	30	15	15	30
P-value	0.023	0.545	0.022	0.002
<b>Para las métricas AD-RD-MD</b>				
Chi-square	19.946	1.124	20.829	11.163
Df	12	6	6	12
P-value	0.068	0.98	0.002	0.515
<b>Para las métricas Iron-FA</b>				
Chi-square	11.953	10.89	0.426	9.214
Df	6	3	3	6
P-value	0.063	0.012	0.935	0.162

Cuadro 79: Homocedasticidad de las matrices de covarianza, CON eliminación de outliers

Without elimination of outliers					With elimination of outliers				
Variables	Subgroup	Hemisphere	Gender	Age	Variables	Subgroup	Hemisphere	Gender	Age
<b>For 5 metrics</b>									
Chi-square	55.198	15.054	35.815	50.223	Chi-square	47.417	13.741	27.87	56.596
Df	30	15	15	30	Df	30	15	15	30
P-value	0.003	0.448	0.002	0.012	P-value	0.023	0.545	0.022	0.002
<b>For diffusivity metrics</b>									
Chi-square	24.436	1.467	25.375	12.132	Chi-square	19.946	1.124	20.829	11.163
Df	12	6	6	12	Df	12	6	6	12
P-value	0.018	0.962	0	0.435	P-value	0.068	0.98	0.002	0.515
<b>For Iron-FA metrics</b>									
Chi-square	10.807	6.14	0.083	6.32	Chi-square	11.953	10.89	0.426	9.214
Df	6	3	3	6	Df	6	3	3	6
P-value	0.095	0.105	0.994	0.388	P-value	0.063	0.012	0.935	0.162

En el caso de las **5 métricas conjuntas**, tanto sin como con eliminación de outliers, se objetiva una falta de homocedasticidad multivariante para subgrupo, género y edad factorizada.

En el caso de las **3 métricas de difusividad**, con el dataframe completo (sin eliminación de outliers), se objetiva una falta de homocedasticidad multivariante para subgrupo y género, y, con la eliminación de outliers, se objetiva una falta de homocedasticidad multivariante exclusivamente para género.

En el caso de las **métricas de hierro y FA**, con el dataframe completo (sin eliminación de outliers), se objetiva homocedasticidad multivariante para las 4 variables independientes (subgrupo, hemisferio, género y edad factorizada), y, con la eliminación de outliers, se objetiva una falta de homocedasticidad multivariante exclusivamente para hemisferio.

## 5 ANÁLISIS del average: ambos ATR conjuntamente

El interés de este análisis radica en dos puntos:

- Determinar la existencia o la ausencia de diferencias entre hemisferios (la lateralidad derecha/izquierda puede ser un factor confusor en la relación de las métricas de RM con el efecto de la enfermedad, ó, en relación a la fisiopatología de la enfermedad con una afectación diferencial temporo-espacial de los ATRs);
- Valorar conjuntamente ambos ATR con el objetivo de ganar potencia estadística dado que se ha observado una correlación entre los valores de AD-RD-MD con el grupo HD y entre los valores iron-FA con el grupo preHD.

### 5.1 MANOVA

MANOVA (*Multivariate ANOVA* o *Multivariate ANalysis Of VAriance*, análisis multivariante de la varianza) es una extensión del ANOVA, pero MANOVA incluye al menos dos variables dependientes para analizar diferencias de las matrices de covarianzas procedentes de las variables dependientes entre múltiples grupos (factores). MANOVA en R utiliza por defecto el **test de Pillai**, que es el de mayor poder estadístico, para los cálculos, aunque se pueden utilizar otros tests (lambda de Wilk, raíz mayor de Roy o test de Hotelling-Lawley), y, luego se convierte en un estadístico F para determinar la significancia de las diferencias de medias grupales.

El ANOVA suele sufrir de errores de tipo I que se producen con la realización de múltiples ANOVA para cada variable dependiente. En cambio, MANOVA determina la exsistencia de diferencias de grupo basándose en la información combinada de múltiples variables dependientes. Debido a que MANOVA utiliza más de una variable dependiente, las hipótesis nula y alternativa cambian ligeramente:

- **H<sub>0</sub>:** Los vectores medios de grupo son los mismos para todos los grupos o no difieren significativamente;
- **H<sub>1</sub>:** Al menos uno de los vectores medios del grupo es diferente del resto.

MANOVA es una técnica paramétrica que implica ciertos **supuestos** sobre los parámetros de la población de la que se toman las muestras y dichas suposiciones son más restrictivas que en el caso del ANOVA:

- **Normalidad multivariada:** cada combinación de variables independientes o dependientes debe tener una distribución normal multivariada. Los errores también deben ajustarse a la **normalidad multivariada** (es decir, se distribuyen normalmente en todas las dimensiones);
- Toda la **matriz de varianza-covarianza es homogénea/homocedástica** (es decir, las varianzas de todas las variables y las covarianzas entre todos los pares de variables son iguales en todos los grupos);
- **Sin multicolinealidad:** las variables dependientes no deben tener correlaciones muy altas.
- **Sin valores atípicos:** no debería haber valores atípicos en las variables dependientes.
- **Linealidad:** las variables dependientes deben tener una relación lineal con cada grupo de la variable independiente (factor).
- **Independencia de las observaciones:** las unidades son intercambiables entre sí.

El supuesto de normalidad multivariada se ha comprobado en los pasos anteriores y se cumple.

El supuesto de homocedasticidad de las matrices de varianza no se cumple para todas las variables independientes (factores), pero, como se comentó en el paso anterior, el incumplimiento de esta suposición no siempre significa que no se puedan ejecutar análisis estadísticos paramétricos, sobretodo cuando aparecen **n** desiguales a medida que disminuyen los tamaños de muestra (siendo este el caso para la variable género en el subgrupo preHD), y, en estos casos, se debe utilizar la traza de Pillai por ser más robusta que otros tests como la traza de Hotelling o la Lambda de Wilks, siendo la traza de Pillai la de defecto en MANOVA.

El supuesto de multicolinealidad tampoco se cumple entre las métricas de difusión, con una alta correlación entre las métricas de difusividad (AD-RD-MD) y entre la difusividad radial (RD) y la anisotropía fraccional (FA).

```
# Correlación entre las métricas de ambos ATR
cor(ATR_DifIron_Df[,7:11])
    ATR_Iron   ATR_AD   ATR_RD   ATR_MD   ATR_FA
ATR_Iron  1.0000000 -0.3793633 -0.1894373 -0.2947272 -0.0905882
ATR_AD   -0.3793633  1.0000000  0.6244238  0.8589011  0.1296171
ATR_RD   -0.1894373  0.6244238  1.0000000  0.9363447 -0.6793069
ATR_MD   -0.2947272  0.8589011  0.9363447  1.0000000 -0.3871477
ATR_FA   -0.0905882  0.1296171 -0.6793069 -0.3871477  1.0000000
```

La existencia de outliers se ha solucionado en el paso de evaluación de normalidad multivariante, pero he sido conservadora a la hora de eliminar los outliers, es decir, no he eliminado todos los sujetos detectados por la función `mvn` como outliers.

La linealidad de las variables dependientes con cada grupo de cada factor se ha comprobado gráficamente mediante qqplots.

Y, por último, las observaciones son independientes entre sí.

Para aplicar MANOVA en R, se debe cumplir una **condición crítica**: el número de observaciones (filas) por grupo de la variable independiente debe ser mayor al número de variables dependientes. Este punto puede ser conflictivo en el género masculino para el subgrupo preHD, donde sólo existe un varón.

Si el MANOVA es significativo, se puede medir el **tamaño del efecto** mediante la métrica `Eta cuadrado parcial`, la cual mide el efecto que tiene la variable independiente sobre las variables dependientes: si el valor es 0,14 o mayor, se puede afirmar que el tamaño del efecto es grande.

Para ejecutar MANOVA, se utiliza la función `manova` del paquete `stats`. Se utilizan los dataframes con eliminación de outliers dado que son los que presentan normalidad multivariante. Se generan MANOVAs para las 5 métricas conjuntas, para las 3 métricas de difusividad (AD, RD, MD) y para las métricas de hierro y anistropía fraccional. El primer modelo generado consiste en las 4 variables independientes en interacción y los modelos subsiguientes con simplificación de las interacciones y/o del número de factores en función de los resultados del primer modelo.

### 5.1.1 5 Métricas

```
# Para las 5 métricas conjuntamente
# Generar los MANOVAS
ATR_manova5_1<-manova(cbind(ATR_Iron,ATR_AD,ATR_RD,ATR_MD,ATR_FA) ~
                        ATR_subgroup*ATR_hemisphere*ATR_Gender*ATR_agef2,
                        data=ATR_DifIron_Df_WoOut)
summary(ATR_manova5_1)
Df Pillai approx F num Df
ATR_subgroup          2 0.42755  3.9698   10
ATR_hemisphere        1 0.12904  2.1334    5
ATR_Gender             1 0.07340  1.1407    5
ATR_agef2              2 0.28734  2.4495   10
ATR_subgroup:ATR_hemisphere  2 0.05412  0.4061   10
ATR_subgroup:ATR_Gender   2 0.13767  1.0793   10
ATR_hemisphere:ATR_Gender 1 0.02939  0.4360    5
ATR_subgroup:ATR_agef2     4 0.46050  1.9516   20
ATR_hemisphere:ATR_agef2   2 0.09968  0.7659   10
ATR_Gender:ATR_agef2      2 0.31986  2.7795   10
ATR_subgroup:ATR_hemisphere:ATR_Gender 2 0.02707  0.2003   10
ATR_subgroup:ATR_hemisphere:ATR_agef2   4 0.10836  0.4177   20
ATR_subgroup:ATR_Gender:ATR_agef2      2 0.13579  1.0634   10
ATR_hemisphere:ATR_Gender:ATR_agef2    2 0.01528  0.1124   10
ATR_subgroup:ATR_hemisphere:ATR_Gender:ATR_agef2 2 0.06266  0.4722   10
Residuals
```

```

den Df      Pr(>F)
ATR_subgroup
ATR_hemisphere
ATR_Gender
ATR_agef2
ATR_subgroup:ATR_hemisphere
ATR_subgroup:ATR_Gender
ATR_hemisphere:ATR_Gender
ATR_subgroup:ATR_agef2
ATR_hemisphere:ATR_agef2
ATR_Gender:ATR_agef2
ATR_subgroup:ATR_hemisphere:ATR_Gender
ATR_subgroup:ATR_hemisphere:ATR_agef2
ATR_subgroup:ATR_Gender:ATR_agef2
ATR_hemisphere:ATR_Gender:ATR_agef2
ATR_subgroup:ATR_hemisphere:ATR_Gender:ATR_agef2
Residuals
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

ATR_manova5_2<-manova(cbind(ATR_Iron,ATR_AD,ATR_RD,ATR_MD,ATR_FA)~ATR_subgroup+
                         ATR_hemisphere+ATR_agef2+ATR_subgroup:ATR_agef2+
                         ATR_Gender:ATR_agef2, data=ATR_DifIron_Df_WoOut)
summary(ATR_manova5_2)
Df  Pillai approx F num Df den Df      Pr(>F)
ATR_subgroup   2 0.40466  4.6671    10   184 6.166e-06 ***
ATR_hemisphere 1 0.11517  2.3689     5    91  0.045474 *
ATR_agef2      2 0.27014  2.8734    10   184  0.002360 **
ATR_subgroup:ATR_agef2 4 0.38720  2.0149    20   376  0.006322 **
ATR_agef2:ATR_Gender 3 0.40338  2.8895    15   279  0.000297 ***
Residuals      95
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Generar effect size
library(effectsize)
ATR_SizeMan5_2<-effectsize::eta_squared(ATR_manova5_2);ATR_SizeMan5_2
# Effect Size for ANOVA (Type I)

Parameter | Eta2 (partial) |      95% CI
-----
ATR_subgroup |          0.20 | [0.09, 1.00]
ATR_hemisphere |        0.12 | [0.00, 1.00]
ATR_agef2 |        0.14 | [0.03, 1.00]
ATR_subgroup:ATR_agef2 |  0.10 | [0.01, 1.00]
ATR_agef2:ATR_Gender |        0.13 | [0.04, 1.00]

- One-sided CIs: upper bound fixed at [1.00].

```

En el caso de las **5 métricas conjuntas**, se objetiva que los factores subgrupo y edad factorizada, y, las interacciones subgrupo-edad factorizada, edad-factorizada-género son los términos significativos, destacando un *size effect* de 0.2 para subgrupo y de 0.14 para la edad factorizada. Remarcar que la interacción subgrupo-hemisferio no ha resultado significativa y que el factor hemisferio es casi significativo (con un pvalor de 0.071).

### 5.1.2 3 Métricas de difusividad

Como se verá a continuación, los modelos MANOVA generados con las 3 métricas de difusividad especifican que los residuos son rango-deficientes y se ha de especificar el argumento `tol=0` para poder ver el resultado/resumen del modelo. El problema de especificar tolerancia 0 es que los resultados pueden ser inexactos, con lo que transformar las respuestas/variables dependientes para eliminar la alta correlación es mejor opción. Además, cuando los residuos son rango-deficientes, no se permite calcular el *effect size*.

```
# Para las 3 métricas AD-RD-MD - Sin transformación de variables
# Generar los MANOVAS
ATR_manova3_1<-manova(cbind(ATR_AD,ATR_RD,ATR_MD)~ATR_subgroup*
                         ATR_hemisphere*ATR_Gender*ATR_agef2,
                         data=ATR_DifIron_Df_WoOut_ADRDMD)

summary(ATR_manova3_1)
Error in summary.manova(ATR_manova3_1): residuals have rank 2 < 3
summary(ATR_manova3_1,tol=0)

ATR_subgroup          Df  Pillai approx F num Df
ATR_hemisphere        1  0.07307   1.9707   3
ATR_Gender             1  0.05012   1.3191   3
ATR_agef2              2  0.09583   1.2749   6
ATR_subgroup:ATR_hemisphere  2  0.03935   0.5085   6
ATR_subgroup:ATR_Gender  2  0.06063   0.7920   6
ATR_hemisphere:ATR_Gender 1  0.01882   0.4796   3
ATR_subgroup:ATR_agef2    4  0.12829   0.8600  12
ATR_hemisphere:ATR_agef2 2  0.05156   0.6704   6
ATR_Gender:ATR_agef2     2  0.08804   1.1665   6
ATR_subgroup:ATR_hemisphere:ATR_Gender 2  0.01676   0.2141   6
ATR_subgroup:ATR_hemisphere:ATR_agef2  4  0.06833   0.4486  12
ATR_subgroup:ATR_Gender:ATR_agef2  2  0.05983   0.7812   6
ATR_hemisphere:ATR_Gender:ATR_agef2 2  0.00738   0.0939   6
ATR_subgroup:ATR_hemisphere:ATR_Gender:ATR_agef2 2  0.02486   0.3188   6
Residuals               77

den Df      Pr(>F)
ATR_subgroup          152 7.658e-05 ***
ATR_hemisphere        75  0.1256
ATR_Gender             75  0.2745
ATR_agef2              152  0.2721
ATR_subgroup:ATR_hemisphere 152  0.8013
ATR_subgroup:ATR_Gender  152  0.5775
ATR_hemisphere:ATR_Gender 75  0.6975
ATR_subgroup:ATR_agef2    231  0.5886
ATR_hemisphere:ATR_agef2 152  0.6737
ATR_Gender:ATR_agef2     152  0.3272
ATR_subgroup:ATR_hemisphere:ATR_Gender 152  0.9719
ATR_subgroup:ATR_hemisphere:ATR_agef2  231  0.9418
ATR_subgroup:ATR_Gender:ATR_agef2  152  0.5859
ATR_hemisphere:ATR_Gender:ATR_agef2 152  0.9969
ATR_subgroup:ATR_hemisphere:ATR_Gender:ATR_agef2 152  0.9264
Residuals

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

ATR_manova3_2<-manova(cbind(ATR_AD,ATR_RD,ATR_MD)~ATR_subgroup+
                         ATR_hemisphere+ATR_Gender+ATR_agef2,
                         data=ATR_DifIron_Df_WoOut_ADRDMD)

summary(ATR_manova3_2)
```

```
Error in summary.manova(ATR_manova3_2): residuals have rank 2 < 3
summary(ATR_manova3_2,tol=0)
      Df Pillai approx F num Df den Df Pr(>F)
ATR_subgroup   2 0.311888   6.2201     6    202 5.198e-06 ***
ATR_hemisphere 1 0.065314   2.3293     3    100  0.07898 .
ATR_Gender     1 0.041948   1.4595     3    100  0.23021
ATR_agef2       2 0.085679   1.5068     6    202  0.17747
Residuals      102
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

ATR_manova3_3<-manova(cbind(ATR_AD,ATR_RD,ATR_MD)~ATR_subgroup+ATR_hemisphere,
                         data=ATR_DifIron_Df_WoOut_ADRDMD)
summary(ATR_manova3_3)
Error in summary.manova(ATR_manova3_3): residuals have rank 2 < 3
summary(ATR_manova3_3,tol=0)
      Df Pillai approx F num Df den Df Pr(>F)
ATR_subgroup   2 0.295635   6.0132     6    208 8.089e-06 ***
ATR_hemisphere 1 0.061297   2.2419     3    103  0.08784 .
Residuals      105
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Generar effect size
ATR_SizeMan3_3<-effectsize::eta_squared(ATR_manova3_3);ATR_SizeMan3_3
Error in summary.manova(model): residuals have rank 2 < 3
Error in eval(expr, envir, enclos): objeto 'ATR_SizeMan3_3' no encontrado
```

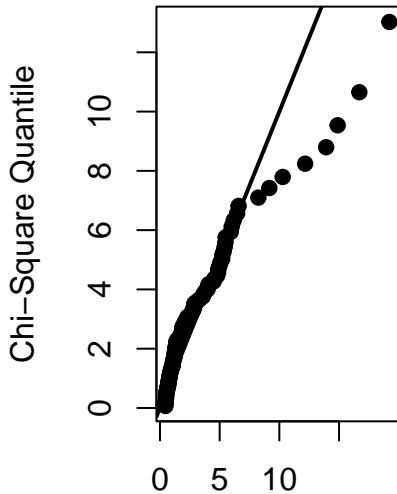
En el caso de las **3 métricas AD-RD-MD sin transformación logarítmica**, se objetiva que subgrupo es el único término significativo y que hemisferio es casi significativo (con un p-valor de 0.08); debido a que los residuos son rango deficientes, no se permite obtener el *effect size*.

Las difusividades media, axial y radial (las 3 métricas con la misma unidad de medida  $mm^2/s$ , con valores muy próximos a cero dado que los valores se hallan en el rango de  $10^{-4}$ – $10^{-5}$ ) son las variables con mayor correlación entre sí. Se procede a la transformación logarítmica con la función `log`, por seguir un paralelismo con la transformación realizada en el análisis MANOVA de cada ATR por separado.

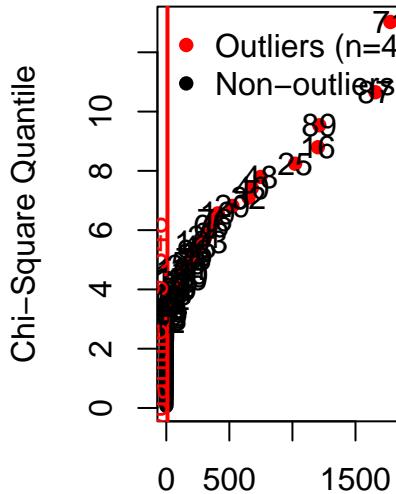
```
# Para las 3 métricas AD-RD-MD - Con transformación logarítmica de las métricas
# Primero transformación de las variables
ATR_DifIron_Df_WoOut_ADRDMD_log<-cbind(ATR_DifIron_Df_WoOut_ADRDMD[,1:6],
                                           log(cbind(ATR_DifIron_Df_WoOut_ADRDMD[,7:9])))

# Comprobar normalidad multivariante
par(mfrow = c(1, 2))
ATR_mvnllog3<-mvn(ATR_DifIron_Df_WoOut_ADRDMD_log[,-c(1:6)], mvnTest = 'royston',
                     covariance = TRUE, desc = TRUE, univariateTest = 'SW',
                     univariatePlot = 'qq', multivariatePlot = 'qq',
                     multivariateOutlierMethod = 'quan', showOutliers = TRUE,
                     showNewData = TRUE)
```

### Chi-Square Q-Q Plot



### Chi-Square Q-Q Plot



Squared Mahalanobis Distance      Robust Squared Mahalanobis Distance

```

ATR_mvnllog3$multivariateNormality
  Test      H   p value MVN
1 Royston 0.5937104 0.7488434 YES
ATR_mvnllog3$univariateNormality
  Test Variable Statistic   p value Normality
1 Shapiro-Wilk ATR_AD     0.9913    0.7175   YES
2 Shapiro-Wilk ATR_RD     0.9880    0.4409   YES
3 Shapiro-Wilk ATR_MD     0.9910    0.6945   YES

# Comprobar homocedasticidad de la matriz de covarianzas
ATR_MBox_SG3log<-boxM(cbind(ATR_AD,ATR_RD,ATR_MD)~ATR_subgroup,
                         data=ATR_DifIron_Df_WoOut_ADRDMD_log, cov=TRUE);ATR_MBox_SG3log

  Box's M-test for Homogeneity of Covariance Matrices

data: Y
Chi-Sq (approx.) = 48.292, df = 12, p-value = 2.781e-06
ATR_MBox_HEM3log<-boxM(cbind(ATR_AD,ATR_RD,ATR_MD)~ATR_hemisphere,
                         data=ATR_DifIron_Df_WoOut_ADRDMD_log, cov=TRUE);ATR_MBox_HEM3log

  Box's M-test for Homogeneity of Covariance Matrices

data: Y
Chi-Sq (approx.) = 8.8094, df = 6, p-value = 0.1846
ATR_MBox_G3log<-boxM(cbind(ATR_AD,ATR_RD,ATR_MD)~ATR_Gender,
                         data=ATR_DifIron_Df_WoOut_ADRDMD_log, cov=TRUE);ATR_MBox_G3log

  Box's M-test for Homogeneity of Covariance Matrices

data: Y
Chi-Sq (approx.) = 16.176, df = 6, p-value = 0.01284

```

```

ATR_MBox_AGE3log<-boxM(cbind(ATR_AD,ATR_RD,ATR_MD)~ATR_agef2,
                         data=ATR_DifIron_Df_WoOut_ADRDMD_log, cov=TRUE);ATR_MBox_AGE3log

  Box's M-test for Homogeneity of Covariance Matrices

data: Y
Chi-Sq (approx.) = 20.975, df = 12, p-value = 0.05075

# Generar los MANOVAS
ATR_manova3_1log<-manova(cbind(ATR_AD,ATR_RD,ATR_MD)~ATR_subgroup*ATR_hemisphere*
                           ATR_Gender*ATR_agef2,data=ATR_DifIron_Df_WoOut_ADRDMD_log)
summary(ATR_manova3_1log)

Df Pillai approx F num Df
ATR_subgroup 2 0.35493 5.4658 6
ATR_hemisphere 1 0.07692 2.0834 3
ATR_Gender 1 0.02434 0.6237 3
ATR_agef2 2 0.10647 1.4244 6
ATR_subgroup:ATR_hemisphere 2 0.01372 0.1749 6
ATR_subgroup:ATR_Gender 2 0.05851 0.7634 6
ATR_hemisphere:ATR_Gender 1 0.00737 0.1855 3
ATR_subgroup:ATR_agef2 4 0.14098 0.9492 12
ATR_hemisphere:ATR_agef2 2 0.03403 0.4385 6
ATR_Gender:ATR_agef2 2 0.07870 1.0377 6
ATR_subgroup:ATR_hemisphere:ATR_Gender 2 0.01005 0.1280 6
ATR_subgroup:ATR_hemisphere:ATR_agef2 4 0.04523 0.2947 12
ATR_subgroup:ATR_Gender:ATR_agef2 2 0.05029 0.6534 6
ATR_hemisphere:ATR_Gender:ATR_agef2 2 0.00634 0.0806 6
ATR_subgroup:ATR_hemisphere:ATR_Gender:ATR_agef2 2 0.01490 0.1902 6
Residuals 77
den Df Pr(>F)
ATR_subgroup 152 3.798e-05 ***
ATR_hemisphere 75 0.1095
ATR_Gender 75 0.6019
ATR_agef2 152 0.2087
ATR_subgroup:ATR_hemisphere 152 0.9832
ATR_subgroup:ATR_Gender 152 0.5998
ATR_hemisphere:ATR_Gender 75 0.9059
ATR_subgroup:ATR_agef2 231 0.4986
ATR_hemisphere:ATR_agef2 152 0.8522
ATR_Gender:ATR_agef2 152 0.4031
ATR_subgroup:ATR_hemisphere:ATR_Gender 152 0.9927
ATR_subgroup:ATR_hemisphere:ATR_agef2 231 0.9898
ATR_subgroup:ATR_Gender:ATR_agef2 152 0.6873
ATR_hemisphere:ATR_Gender:ATR_agef2 152 0.9980
ATR_subgroup:ATR_hemisphere:ATR_Gender:ATR_agef2 152 0.9792
Residuals

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

ATR_manova3_2log<-manova(cbind(ATR_AD,ATR_RD,ATR_MD)~ATR_subgroup+ATR_hemisphere+
                           ATR_Gender+ATR_agef2,data=ATR_DifIron_Df_WoOut_ADRDMD_log)
summary(ATR_manova3_2log)

Df Pillai approx F num Df den Df Pr(>F)
ATR_subgroup 2 0.33063 6.6679 6 202 1.878e-06 ***
ATR_hemisphere 1 0.06963 2.4948 3 100 0.06425 .
ATR_Gender 1 0.02161 0.7364 3 100 0.53273

```

```

ATR_agef2      2  0.09726   1.7208      6     202    0.11773
Residuals     102
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

ATR_manova3_3log<-manova(cbind(ATR_AD,ATR_RD,ATR_MD)~ATR_subgroup+ATR_hemisphere,
                           data=ATR_DifIron_Df_WoOut_ADRDMD_log)
summary(ATR_manova3_3log)
Df   Pillai approx F num Df den Df Pr(>F)
ATR_subgroup   2  0.314523   6.4691      6     208 2.851e-06 ***
ATR_hemisphere 1  0.065309   2.3989      3     103  0.07222 .
Residuals     105
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Generar effect size
ATR_SizeMan3_3log<-effectsize::eta_squared(ATR_manova3_3log);ATR_SizeMan3_3log
# Effect Size for ANOVA (Type I)

Parameter | Eta2 (partial) |      95% CI
-----
ATR_subgroup |          0.16 | [0.07, 1.00]
ATR_hemisphere |        0.07 | [0.00, 1.00]

- One-sided CIs: upper bound fixed at [1.00].

```

En el caso de las **3 métricas AD-RD-MD con transformación logarítmica**, se objetiva que subgrupo es el único factor significativo con un *effect size* de 0.16, y, que hemisferio es casi significativo (con un p-valor de 0.07) pero con *effect size* no significativo. Cabe destacar que la normalidad multivariante y la homocedasticidad de la matriz de covarianzas para hemisferio y edad factorizada se mantienen, pero se pierde la homocedasticidad de la matriz de covarianzas para subgrupo, hecho que puede sesgar el test de MANOVA, y, la heterocedasticidad de la matriz de covarianzas persiste para género.

### 5.1.3 2 Métricas Iron-FA

```

# Para las 2 métricas Iron-FA
# Generar los MANOVAS
ATR_manova2_1<-manova(cbind(ATR_Iron,ATR_FA)~ATR_subgroup*ATR_hemisphere*
                        ATR_Gender*ATR_agef2,data=ATR_DifIron_Df_WoOut_IronFA)
summary(ATR_manova2_1)
Df   Pillai approx F num Df
ATR_subgroup      2  0.090424   1.7994      4
ATR_hemisphere    1  0.185454   8.5379      2
ATR_Gender         1  0.002796   0.1051      2
ATR_agef2          2  0.046396   0.9025      4
ATR_subgroup:ATR_hemisphere  2  0.013430   0.2569      4
ATR_subgroup:ATR_Gender       2  0.026745   0.5150      4
ATR_hemisphere:ATR_Gender    1  0.011001   0.4171      2
ATR_subgroup:ATR_agef2        4  0.077612   0.7671      8
ATR_hemisphere:ATR_agef2     2  0.015728   0.3012      4
ATR_Gender:ATR_agef2          2  0.222738   4.7624      4
ATR_subgroup:ATR_hemisphere:ATR_Gender  2  0.013758   0.2632      4
ATR_subgroup:ATR_hemisphere:ATR_agef2   3  0.038723   0.5002      6
ATR_subgroup:ATR_Gender:ATR_agef2     2  0.094301   1.8804      4

```

```

ATR_hemisphere:ATR_Gender:ATR_agef2      2  0.054903   1.0726    4
ATR_subgroup:ATR_hemisphere:ATR_Gender:ATR_agef2 2  0.030433   0.5872    4
Residuals                                     76
                                         den Df     Pr(>F)
ATR_subgroup                                152 0.1318065
ATR_hemisphere                             75  0.0004564 *** 
ATR_Gender                                  75  0.9003418
ATR_agef2                                    152 0.4641774
ATR_subgroup:ATR_hemisphere                152 0.9050803
ATR_subgroup:ATR_Gender                     152 0.7247767
ATR_hemisphere:ATR_Gender                  75  0.6604530
ATR_subgroup:ATR_agef2                      152 0.6322573
ATR_hemisphere:ATR_agef2                   152 0.8767914
ATR_Gender:ATR_agef2                       152 0.0011992 ** 
ATR_subgroup:ATR_hemisphere:ATR_Gender     152 0.9011797
ATR_subgroup:ATR_hemisphere:ATR_agef2       152 0.8075042
ATR_subgroup:ATR_Gender:ATR_agef2           152 0.1166919
ATR_hemisphere:ATR_Gender:ATR_agef2         152 0.3721309
ATR_subgroup:ATR_hemisphere:ATR_Gender:ATR_agef2 152 0.6724211
Residuals
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

ATR_manova2_2<-manova(cbind(ATR_Iron,ATR_FA)~ATR_subgroup+ATR_hemisphere+
                         ATR_Gender:ATR_agef2,data=ATR_DifIron_Df_WoOut_IronFA)
summary(ATR_manova2_2)
   Df Pillai approx F num Df den Df     Pr(>F)
ATR_subgroup      2 0.082087   2.0972      4    196 0.0826419 .
ATR_hemisphere   1 0.153348   8.7844      2     97 0.0003117 ***
ATR_Gender:ATR_agef2 5 0.180629   1.9459     10    196 0.0412077 *
Residuals          98
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Generar effect size
library(effects)
ATR_SizeMan2_2<-effectsize::eta_squared(ATR_manova2_2);ATR_SizeMan2_2
# Effect Size for ANOVA (Type I)

Parameter | Eta2 (partial) |      95% CI
-----
ATR_subgroup |          0.04 | [0.00, 1.00]
ATR_hemisphere |        0.15 | [0.05, 1.00]
ATR_Gender:ATR_agef2 |  0.09 | [0.00, 1.00]

- One-sided CIs: upper bound fixed at [1.00].

```

En el caso de las **2 métricas Iron-FA**, el hemisferio y la interacción género-edad factorizada son los dos términos significativos, y, el subgrupo es casi significativo con p-valor de 0.08, destacando tan sólo el *effect size* para hemisferio (0.15). Remarcando que la interacción subgrupo-hemisferio no ha resultado significativa.

En resumen, la interacción subgrupo-hemisferio no es significativa en ninguna de las combinaciones métricas. Los términos significativos han sido los siguientes:

- Con las 5 métricas conjuntas, subgrupo, edad factorizada y las interacciones subgrupo-edad y edad-género;

- Con las 3 métricas de difusividad, subgrupo;
- Con las métricas de iron y FA, hemisferio y la interacción género-edad.

También es importante remarcar que tanto el análisis conjunto de ambos ATRs como el análisis de cada ATR por separado detectan el subgrupo como factor significativo con la evaluación de las 5 métricas conjuntamente y con las 3 métricas de difusividad, y, el análisis conjunto de ambos ATRs no aumenta la sensibilidad para detectar diferencias entre subgrupos en el análisis conjunto de las métricas iron-FA.

#### 5.1.4 Comparaciones pareadas

Para determinar qué subgrupo difiere del resto, se genera un test post hoc mediante el **Análisis Discriminante Lineal (LDA)**, el cual encuentra una combinación lineal de las variables dependientes (*features*) que mejor separados o más grupos. Se utiliza la función `lda` del paquete MASS.

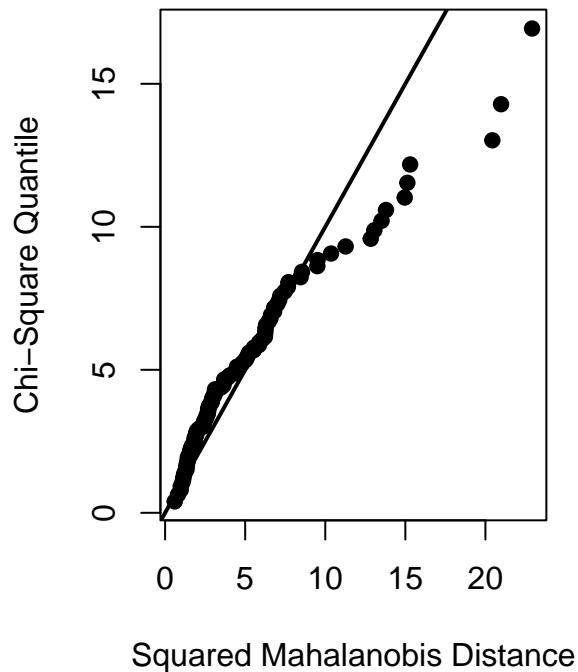
Se van a analizar aquellos MANOVA que han detectado el subgrupo como factor significativo y se van a utilizar los dataframes con transformaciones logarítmicas dado que sin dichas transformaciones la función `lda` expone que las variables de difusividad media, radial y axial son constantes (probablemente debido a sus valores muy próximos a cero), como se verá a continuación. En el caso de las 5 métricas evaluadas conjuntamente, se realiza la transformación logarítmica de las 3 métricas de difusividad y se comprueba la normalidad multivariante como la homocedasticidad de la matriz de covarianzas, dado que no se había ejecutado en el MANOVA anterior (no se ejecutó porque los residuos no eran rango-deficientes).

```
library(MASS)
# Para las 5 métricas

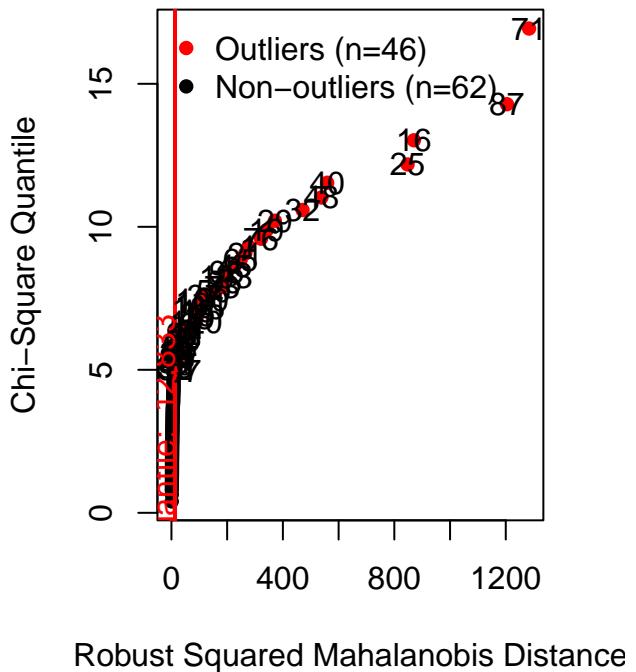
# Primero, división del dataframe sin transformación logarítmica en los datasets
# de entrenamiento y testeo
set.seed(123456)
train_index<-createDataPartition(ATR_DifIron_Df_WoOut$ATR_subgroup, p=0.8, list=FALSE)
train<-ATR_DifIron_Df_WoOut[train_index,]
test<-ATR_DifIron_Df_WoOut[-train_index,]
# Segundo, se genera el modelo LDA
ATR_LDA5model<-lda(ATR_subgroup~cbind(ATR_Iron,ATR_AD,ATR_RD,ATR_MD,ATR_FA), data=train)
Error in lda.default(x, grouping, ...): variables 2 3 4 appear to be constant within groups

# Segundo, con el dataframe con transformación logarítmica
## Primero transformación de las variables
ATR_DifIron_Df_WoOut_log<-cbind(ATR_DifIron_Df_WoOut[,1:7],log(cbind(ATR_DifIron_Df_WoOut[,8:10])),
                                   ATR_DifIron_Df_WoOut[,11])
colnames(ATR_DifIron_Df_WoOut_log)<-colnames(ATR_DifIron_Df_WoOut[-12])
## Comprobar normalidad multivariante
par(mfrow = c(1, 2))
ATR_mvnllog5<-mvn(ATR_DifIron_Df_WoOut_log[-c(1:6)], mvnTest = 'royston', covariance = TRUE,
                     desc = TRUE, univariateTest = 'SW', univariatePlot = 'qq',
                     multivariatePlot = 'qq', multivariateOutlierMethod = 'quan',
                     showOutliers = TRUE, showNewData = TRUE)
```

### Chi-Square Q-Q Plot



### Chi-Square Q-Q Plot



```

ATR_mvnllog5$multivariateNormality
  Test      H   p value MVN
1 Royston 6.185779 0.1506936 YES
ATR_mvnllog5$univariateNormality
  Test Variable Statistic   p value Normality
1 Shapiro-Wilk ATR_Iron    0.9711  0.0186   NO
2 Shapiro-Wilk ATR_AD     0.9910  0.6944   YES
3 Shapiro-Wilk ATR_RD     0.9873  0.4008   YES
4 Shapiro-Wilk ATR_MD     0.9917  0.7562   YES
5 Shapiro-Wilk ATR_FA     0.9882  0.4665   YES
## Comprobar homocedasticidad de la matriz de covarianzas
ATR_MBox_SG5log<-boxM(cbind(ATR_Iron,ATR_AD,ATR_RD,ATR_MD,ATR_FA)~ATR_subgroup,
                        data=ATR_DifIron_Df_WoOut_log, cov=TRUE);ATR_MBox_SG5log

Box's M-test for Homogeneity of Covariance Matrices

data: Y
Chi-Sq (approx.) = 77.675, df = 30, p-value = 4.242e-06
ATR_MBox_HEM5log<-boxM(cbind(ATR_Iron,ATR_AD,ATR_RD,ATR_MD,ATR_FA)~ATR_hemisphere,
                          data=ATR_DifIron_Df_WoOut_log, cov=TRUE);ATR_MBox_HEM5log

Box's M-test for Homogeneity of Covariance Matrices

data: Y
Chi-Sq (approx.) = 17.106, df = 15, p-value = 0.3126
ATR_MBox_G5log<-boxM(cbind(ATR_Iron,ATR_AD,ATR_RD,ATR_MD,ATR_FA)~ATR_Gender,
                      data=ATR_DifIron_Df_WoOut_log, cov=TRUE);ATR_MBox_G5log

```

```

Box's M-test for Homogeneity of Covariance Matrices

data: Y
Chi-Sq (approx.) = 26.382, df = 15, p-value = 0.0342
ATR_MBox_AGE5log<-boxM(cbind(ATR_Iron,ATR_AD,ATR_RD,ATR_MD,ATR_FA)~ATR_agef2,
                           data=ATR_DifIron_Df_WoOut_log, cov=TRUE);ATR_MBox_AGE5log

Box's M-test for Homogeneity of Covariance Matrices

data: Y
Chi-Sq (approx.) = 64.786, df = 30, p-value = 0.0002336

```

Con la transformación logarítmica, se mantiene la variabilidad multivariante, persiste la heterocedasticidad de la matriz de covarianzas para los factores subgrupo, género y edad factorizada y se mantiene la homocedasticidad de la matriz de covarianzas para el factor hemisferio. Las heterocedasticidades podrían sesgar los resultados en la ejecución de un MANOVA (no se ha ejecutado dado que en el paso anterior el modelo no generaba residuos rango-deficientes).

```

# Se repiten el primer y segundo paso con el dataframe con transformación logarítmica

## Se generan los dos datasets
set.seed(123456)

train_index<-createDataPartition(ATR_DifIron_Df_WoOut_log$ATR_subgroup, p=0.8, list=FALSE)
train<-ATR_DifIron_Df_WoOut_log[train_index,]
test<-ATR_DifIron_Df_WoOut_log[-train_index,]

## Se genera el LDA
ATR_LDAlog5model<-lda(ATR_subgroup~cbind(ATR_Iron,ATR_AD,ATR_RD,ATR_MD,ATR_FA),data=train)
print(ATR_LDAlog5model)
Call:
lda(ATR_subgroup ~ cbind(ATR_Iron, ATR_AD, ATR_RD, ATR_MD, ATR_FA),
     data = train)

Prior probabilities of groups:
      1       2       3 
0.4431818 0.2386364 0.3181818

Group means:
  cbind(ATR_Iron, ATR_AD, ATR_RD, ATR_MD, ATR_FA) ATR_Iron
1                               0.02406834
2                               0.02459784
3                               0.02348842
  cbind(ATR_Iron, ATR_AD, ATR_RD, ATR_MD, ATR_FA) ATR_AD
1                             -7.020846
2                             -7.015410
3                             -6.972645
  cbind(ATR_Iron, ATR_AD, ATR_RD, ATR_MD, ATR_FA) ATR_RD
1                             -7.608945
2                             -7.631564
3                             -7.562340
  cbind(ATR_Iron, ATR_AD, ATR_RD, ATR_MD, ATR_FA) ATR_MD
1                             -7.372135
2                             -7.381458
3                             -7.324745
  cbind(ATR_Iron, ATR_AD, ATR_RD, ATR_MD, ATR_FA) ATR_FA

```

```

1          0.3786882
2          0.3929521
3          0.3800975

Coefficients of linear discriminants:

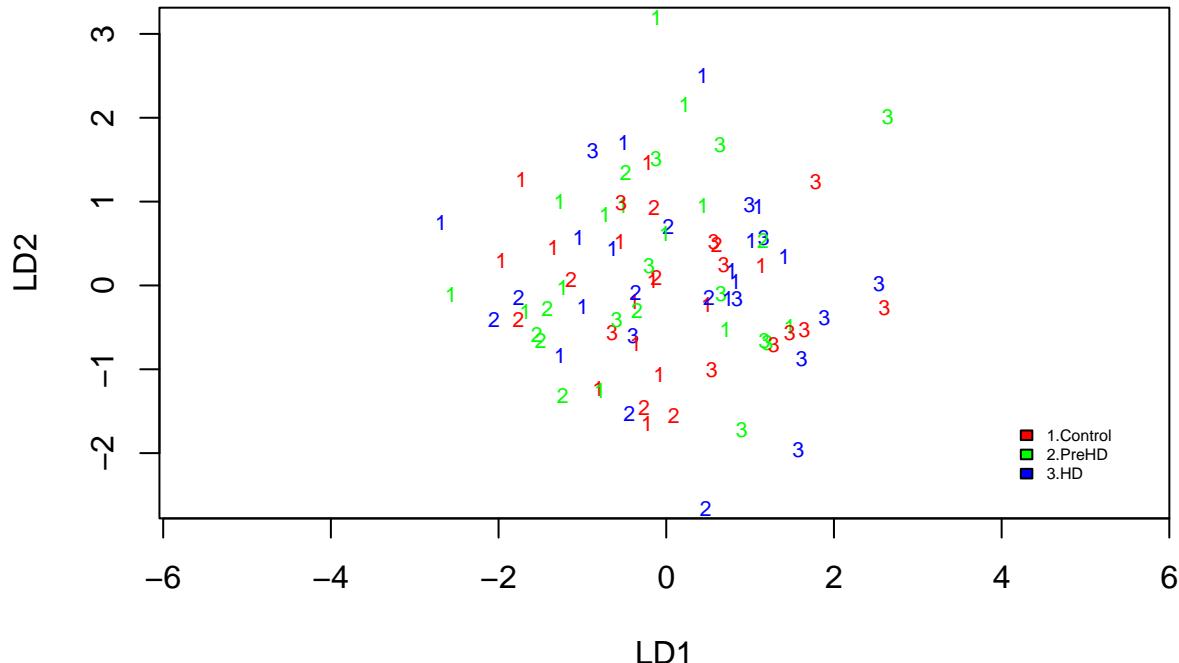
LD1           LD2
cbind(ATR_Iron, ATR_AD, ATR_RD, ATR_MD, ATR_FA)ATR_Iron -27.51108 -335.66417
cbind(ATR_Iron, ATR_AD, ATR_RD, ATR_MD, ATR_FA)ATR_AD   -258.68421   55.93589
cbind(ATR_Iron, ATR_AD, ATR_RD, ATR_MD, ATR_FA)ATR_RD   -173.31637  114.90177
cbind(ATR_Iron, ATR_AD, ATR_RD, ATR_MD, ATR_FA)ATR_MD   456.01076 -180.99893
cbind(ATR_Iron, ATR_AD, ATR_RD, ATR_MD, ATR_FA)ATR_FA    107.31069   7.14558

Proportion of trace:
LD1      LD2
0.8368  0.1632

# Se genera el plot
##Define colors for each subgroup
subgroup_colors <- c("1" = "red", "2" = "green", "3" = "blue")
## Plot the model with custom colors for species
plot(ATR_LDAlog5model, col = subgroup_colors,
     main="Visualization of 5log-LDA model of both ATR")
legend('bottomright', inset=0.05, bty='n', cex=0.5, legend=c('1.Control','2.PreHD','3.HD'),
       fill=c('red','green','blue'))

```

## Visualization of 5log-LDA model of both ATR



```

# Se predicen las clases
predClas_ATRLDA5<-predict(ATR_LDAlog5model, newdata = test, type='class')

```

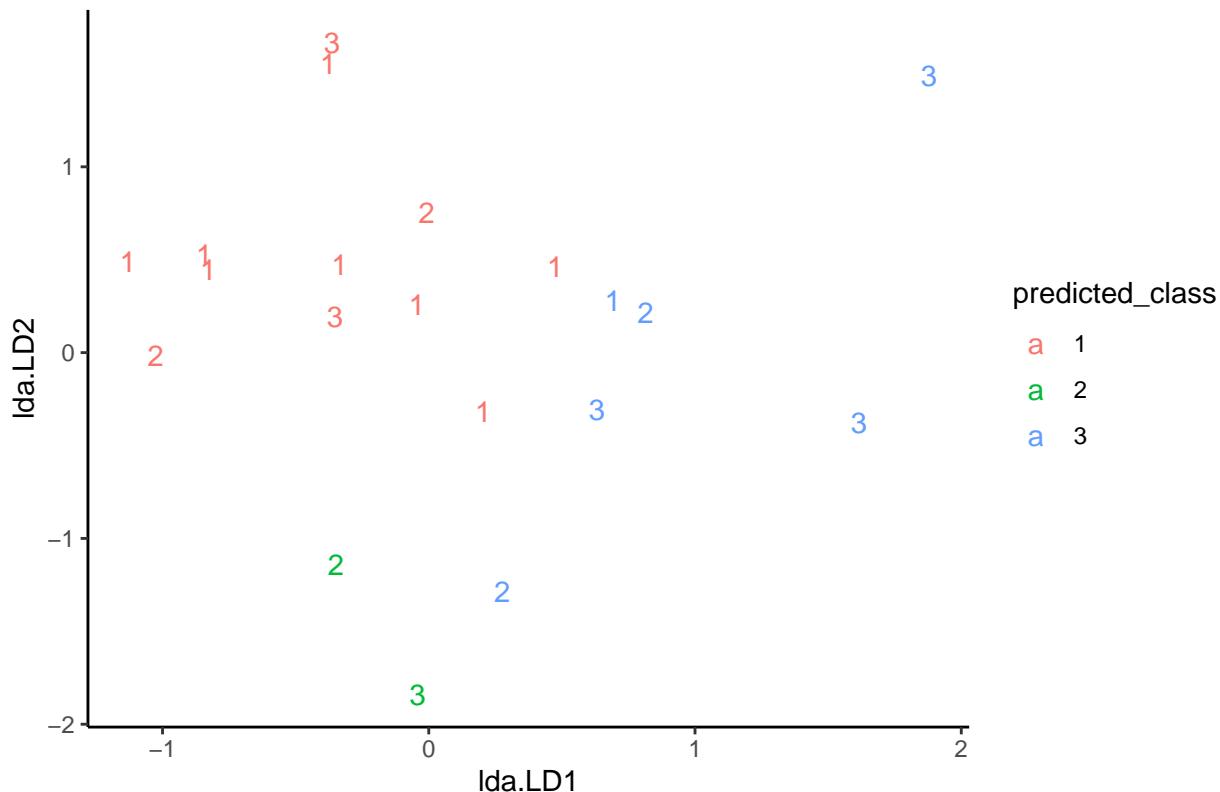
```

test_subgroup<-as.character(test$ATR_subgroup)
test_subgroup<-factor(test_subgroup, levels = levels(predClas_ATRLDA5$class))
table(predClas_ATRLDA5$class, test_subgroup)
  test_subgroup
  1 2 3
1 8 2 2
2 0 1 1
3 1 2 3

ATR_LDA5log_DF<-data.frame(
  predicted_class=predClas_ATRLDA5$class,
  subgroup=test[, 'ATR_subgroup'],
  lda=predClas_ATRLDA5$x
)
str(ATR_LDA5log_DF)
'data.frame':   20 obs. of  4 variables:
 $ predicted_class: Factor w/ 3 levels "1","2","3": 1 1 1 1 3 1 3 1 1 1 ...
 $ subgroup       : Factor w/ 3 levels "1","2","3": 1 1 1 1 1 1 2 2 2 1 ...
 $ lda.LD1        : num  -0.377 -1.13 0.205 -0.843 0.692 ...
 $ lda.LD2        : num  1.559 0.489 -0.314 0.527 0.282 ...
ggplot(ATR_LDA5log_DF) +
  geom_text(aes(x=lda.LD1,y=lda.LD2,label=subgroup, color = predicted_class)) +
  theme_classic() + ggtitle('LDA results for 5log-LDA model of both ATR')

```

LDA results for 5log-LDA model of both ATR



```

# Evaluación del modelo
## Accuracy
accuracyATR_LDAlog5<-sum(diag(table(predClas_ATRLDA5$class, test$ATR_subgroup)))/nrow(test)

```

```

accuracyATR_LDAlog5
[1] 0.6

## Confusion Matrix
confMX_ATR_LDA5log<-confusionMatrix(predClas_ATRLDA5$class, test$ATR_subgroup)
confMX_ATR_LDA5log
Confusion Matrix and Statistics

Reference
Prediction 1 2 3
      1 8 2 2
      2 0 1 1
      3 1 2 3

Overall Statistics

    Accuracy : 0.6
    95% CI : (0.3605, 0.8088)
    No Information Rate : 0.45
    P-Value [Acc > NIR] : 0.1308

    Kappa : 0.3496

McNemar's Test P-Value : 0.4459

Statistics by Class:

          Class: 1 Class: 2 Class: 3
Sensitivity       0.8889   0.2000   0.5000
Specificity        0.6364   0.9333   0.7857
Pos Pred Value     0.6667   0.5000   0.5000
Neg Pred Value     0.8750   0.7778   0.7857
Prevalence         0.4500   0.2500   0.3000
Detection Rate     0.4000   0.0500   0.1500
Detection Prevalence 0.6000   0.1000   0.3000
Balanced Accuracy  0.7626   0.5667   0.6429

## ROC curves one vs the rest by multiROC package
Control_pred_ATRLDA5log<- predClas_ATRLDA5$posterior[, '1']
PreHD_pred_ATRLDA5log<- predClas_ATRLDA5$posterior[, '2']
HD_pred_ATRLDA5log<- predClas_ATRLDA5$posterior[, '3']
Control_true<-ifelse(test$ATR_subgroup=='1',1,0)
PreHD_true<-ifelse(test$ATR_subgroup=='2',1,0)
HD_true<-ifelse(test$ATR_subgroup=='3',1,0)
DF_ROC_ATRLDA5log<-cbind(Control_true,PreHD_true,HD_true,Control_pred_ATRLDA5log,
                           PreHD_pred_ATRLDA5log,HD_pred_ATRLDA5log)
multiROC_ATR_LDA5log<-multi_roc(DF_ROC_ATRLDA5log)
multiROC_ATR_LDA5log_DF<-plot_roc_data(multiROC_ATR_LDA5log)
ggplot(multiROC_ATR_LDA5log_DF, aes(x=1-Specificity, y=Sensitivity, color = Group)) +
  geom_step() + geom_point() + labs(title= "5-LDA model of both ATR ROC curve: One vs Rest",
  x = "False Positive Rate (1-Specificity)",
  y = "True Positive Rate (Sensitivity)") + geom_abline(intercept = 0, slope=1,linetype = 'dotted') +
  annotate('text',x=0.75, y=0.25,
  label=paste('AUC for Controls ROC curve = ',
  round(mean(multiROC_ATR_LDA5log_DF$AUC[multiROC_ATR_LDA5log_DF$Group=='Control']),2),
  color = '#FF9900',size = 8/.pt) +

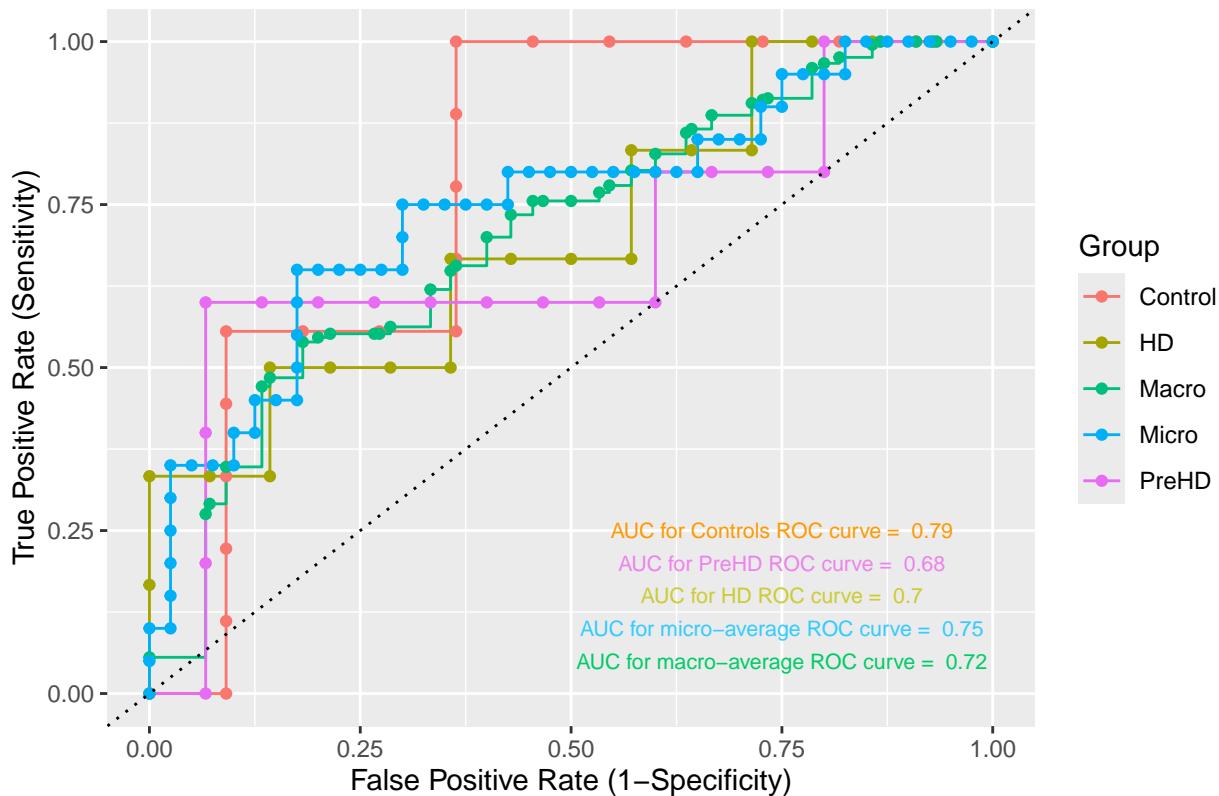
```

```

annotate('text',x=0.75, y=0.20,
        label=paste('AUC for PreHD ROC curve = ',
                    round(mean(multiROC_ATR_LDA5log_DF$AUC[multiROC_ATR_LDA5log_DF$Group=='PreHD']),2)),
        color = 'violet',size = 8/.pt) +
annotate('text',x=0.75, y=0.15,
        label=paste('AUC for HD ROC curve = ',
                    round(mean(multiROC_ATR_LDA5log_DF$AUC[multiROC_ATR_LDA5log_DF$Group=='HD']),2)),
        color = '#CCCC33',size = 8/.pt) +
annotate('text',x=0.75, y=0.10,
        label=paste('AUC for micro-average ROC curve = ',
                    round(mean(multiROC_ATR_LDA5log_DF$AUC[multiROC_ATR_LDA5log_DF$Group=='Micro']),2)),
        color = '#33CCFF',size = 8/.pt) +
annotate('text',x=0.75, y=0.05,
        label=paste('AUC for macro-average ROC curve = ',
                    round(mean(multiROC_ATR_LDA5log_DF$AUC[multiROC_ATR_LDA5log_DF$Group=='Macro']),2)),
        color = '#00CC66',size = 8/.pt)

```

## 5-LDA model of both ATR ROC curve: One vs Rest



```

## Kappa coefficient
library(irr)
kappa_ATR_LDA5log<-kappam.fleiss(data.frame(predClas_ATRLDA5$class,test$ATR_subgroup))
kappa_ATR_LDA5log
Fleiss' Kappa for m Raters

Subjects = 20
Raters = 2
Kappa = 0.337

```

```

z = 2.04
p-value = 0.0412

library(MASS)
# Para las 3 métricas con el dataframe sin transformación logarítmica
## Primero, se generan los dos datasets (entrenamiento y testeo)
set.seed(123456)
train_index<-createDataPartition(ATR_DifIron_Df_WoOut_ADRDMD$ATR_subgroup, p=0.8, list=FALSE)
train<-ATR_DifIron_Df_WoOut_ADRDMD[train_index,]
test<-ATR_DifIron_Df_WoOut_ADRDMD[-train_index,]
## Se genera el LDA
ATR_LDA3model<-lda(ATR_subgroup~cbind(ATR_AD,ATR_RD,ATR_MD),data=train)
Error in lda.default(x, grouping, ...): variables 1 2 3 appear to be constant within groups

# Para las 3 métricas con el dataframe con transformación logarítmica
## Primero, se generan los dos datasets (entrenamiento y testeo)
set.seed(123456)
train_index<-createDataPartition(ATR_DifIron_Df_WoOut_ADRDMD_log$ATR_subgroup, p=0.8, list=FALSE)
train<-ATR_DifIron_Df_WoOut_ADRDMD_log[train_index,]
test<-ATR_DifIron_Df_WoOut_ADRDMD_log[-train_index,]
## Se genera el LDA
ATR_LDAlog3model<-lda(ATR_subgroup~cbind(ATR_AD,ATR_RD,ATR_MD),data=train)
print(ATR_LDAlog3model)
Call:
lda(ATR_subgroup ~ cbind(ATR_AD, ATR_RD, ATR_MD), data = train)

Prior probabilities of groups:
      1       2       3
0.4431818 0.2386364 0.3181818

Group means:
  cbind(ATR_AD, ATR_RD, ATR_MD)ATR_AD cbind(ATR_AD, ATR_RD, ATR_MD)ATR_RD
1           -7.020846           -7.608945
2           -7.015410           -7.631564
3           -6.978911           -7.557112
  cbind(ATR_AD, ATR_RD, ATR_MD)ATR_MD
1            -7.372135
2            -7.381458
3            -7.324853

Coefficients of linear discriminants:
                               LD1        LD2
cbind(ATR_AD, ATR_RD, ATR_MD)ATR_AD -451.7661  325.3276
cbind(ATR_AD, ATR_RD, ATR_MD)ATR_RD -497.4898  405.7847
cbind(ATR_AD, ATR_RD, ATR_MD)ATR_MD  969.6680 -738.8309

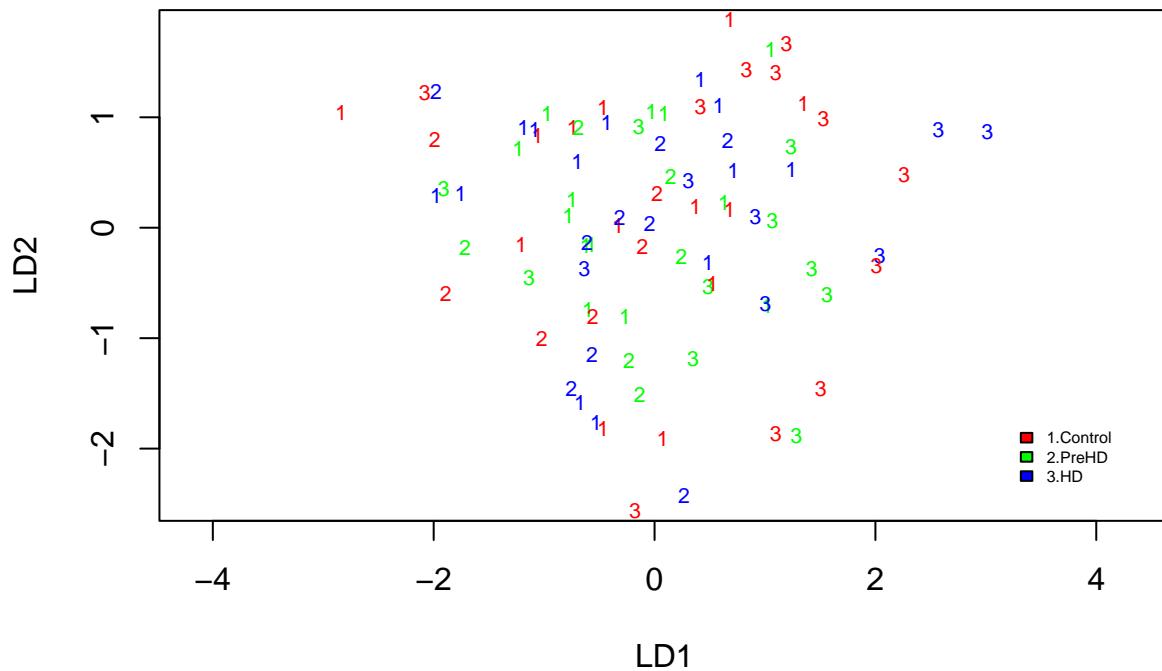
Proportion of trace:
LD1   LD2
0.881 0.119

# Se genera el plot del modelo
## Define colors for each subgroup
subgroup_colors <- c("1" = "red", "2" = "green", "3" = "blue")
## Plot the model with custom colors for species
plot(ATR_LDAlog3model, col = subgroup_colors,
     main="Visualization of 3log-LDA model of both ATR")

```

```
legend('bottomright', inset=0.05, bty='n', cex=0.5, legend=c('1.Control', '2.PreHD', '3.HD'), fill=c('red', 'green', 'blue'))
```

## Visualization of 3log-LDA model of both ATR



```
# Se predicen las clases
predClas_ATRLDA3<-predict(ATR_LDAlog3model, newdata = test, type='class')
test_subgroup<-as.character(test$ATR_subgroup)
test_subgroup<-factor(test_subgroup, levels = levels(predClas_ATRLDA3$class))
table(predClas_ATRLDA3$class, test_subgroup)
  test_subgroup
  1 2 3
1 9 4 4
2 0 1 0
3 0 0 3

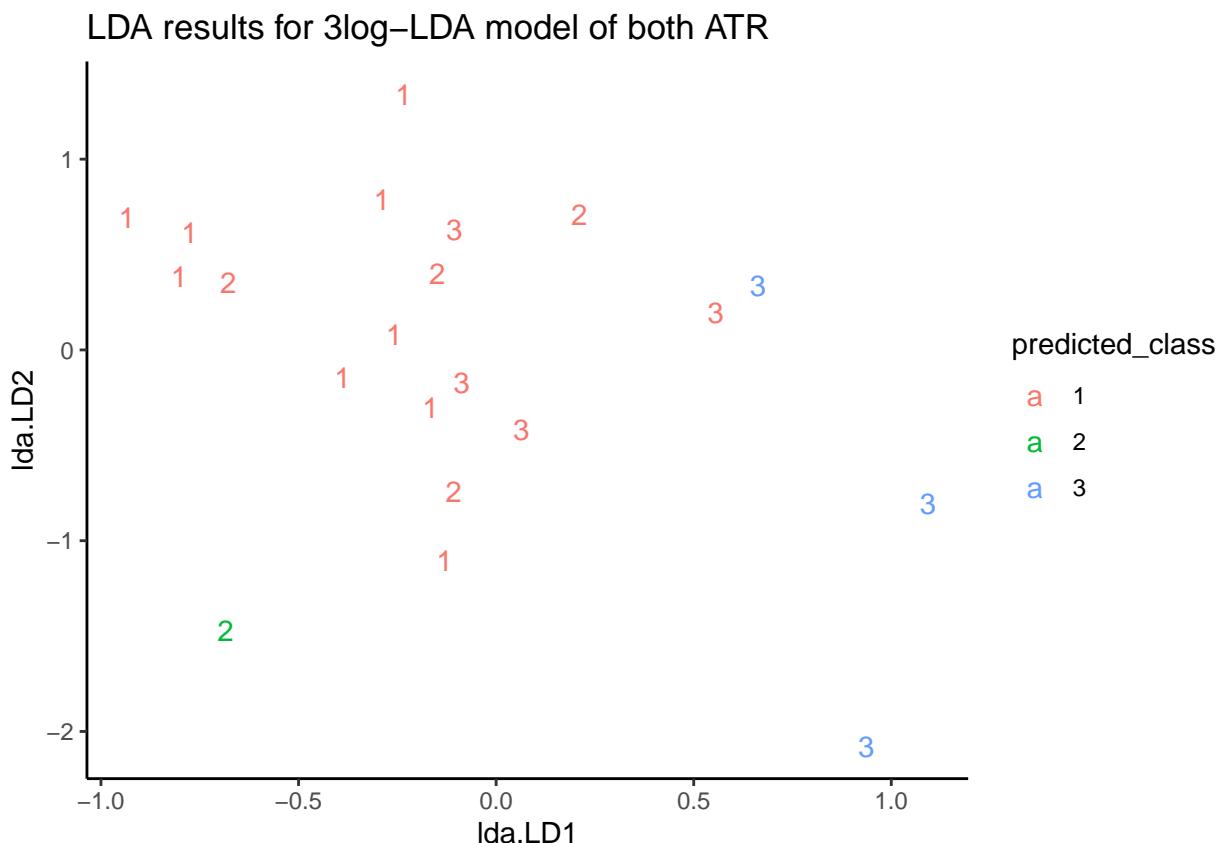
# Se predicen las clases
predClas_ATRLDA3<-predict(ATR_LDAlog3model, newdata = test, type='class')
test_subgroup<-as.character(test$ATR_subgroup)
test_subgroup<-factor(test_subgroup, levels = levels(predClas_ATRLDA3$class))
table(predClas_ATRLDA3$class, test_subgroup)
  test_subgroup
  1 2 3
1 9 4 4
2 0 1 0
3 0 0 3

ATR_LDA3log_DF<-data.frame(
  predicted_class=predClas_ATRLDA3$class,
  subgroup=test[, 'ATR_subgroup'],
```

```

lda=predClas_ATRLDA3$x
)
str(ATR_LDA3log_DF)
'data.frame': 21 obs. of 4 variables:
$ predicted_class: Factor w/ 3 levels "1","2","3": 1 1 1 1 1 1 1 1 1 1 ...
$ subgroup      : Factor w/ 3 levels "1","2","3": 1 1 1 1 1 1 1 1 1 1 ...
$ lda.LD1       : num -0.234 -0.934 -0.131 -0.776 -0.388 ...
$ lda.LD2       : num 1.341 0.692 -1.104 0.614 -0.143 ...
ggplot(ATR_LDA3log_DF) +
  geom_text(aes(x=lda.LD1,y=lda.LD2,label=subgroup, color = predicted_class)) +
  theme_classic() + ggtitle('LDA results for 3log-LDA model of both ATR')

```



```

# Evaluación del modelo
## Accuracy
accuracyATR_LDAlog3<-sum(diag(table(predClas_ATRLDA3$class, test$ATR_subgroup)))/nrow(test)
accuracyATR_LDAlog3
[1] 0.6190476

## Confusion Matrix
confMX_ATR_LDA3log<-confusionMatrix(predClas_ATRLDA3$class, test$ATR_subgroup)
confMX_ATR_LDA3log
Confusion Matrix and Statistics

Reference
Prediction 1 2 3
      1 9 4 4
      2 0 1 0

```

3 0 0 3

#### Overall Statistics

```
Accuracy : 0.619
 95% CI : (0.3844, 0.8189)
No Information Rate : 0.4286
P-Value [Acc > NIR] : 0.06212

Kappa : 0.3588
```

Mcnemar's Test P-Value : NA

#### Statistics by Class:

	Class: 1	Class: 2	Class: 3
Sensitivity	1.0000	0.20000	0.4286
Specificity	0.3333	1.00000	1.0000
Pos Pred Value	0.5294	1.00000	1.0000
Neg Pred Value	1.0000	0.80000	0.7778
Prevalence	0.4286	0.23810	0.3333
Detection Rate	0.4286	0.04762	0.1429
Detection Prevalence	0.8095	0.04762	0.1429
Balanced Accuracy	0.6667	0.60000	0.7143

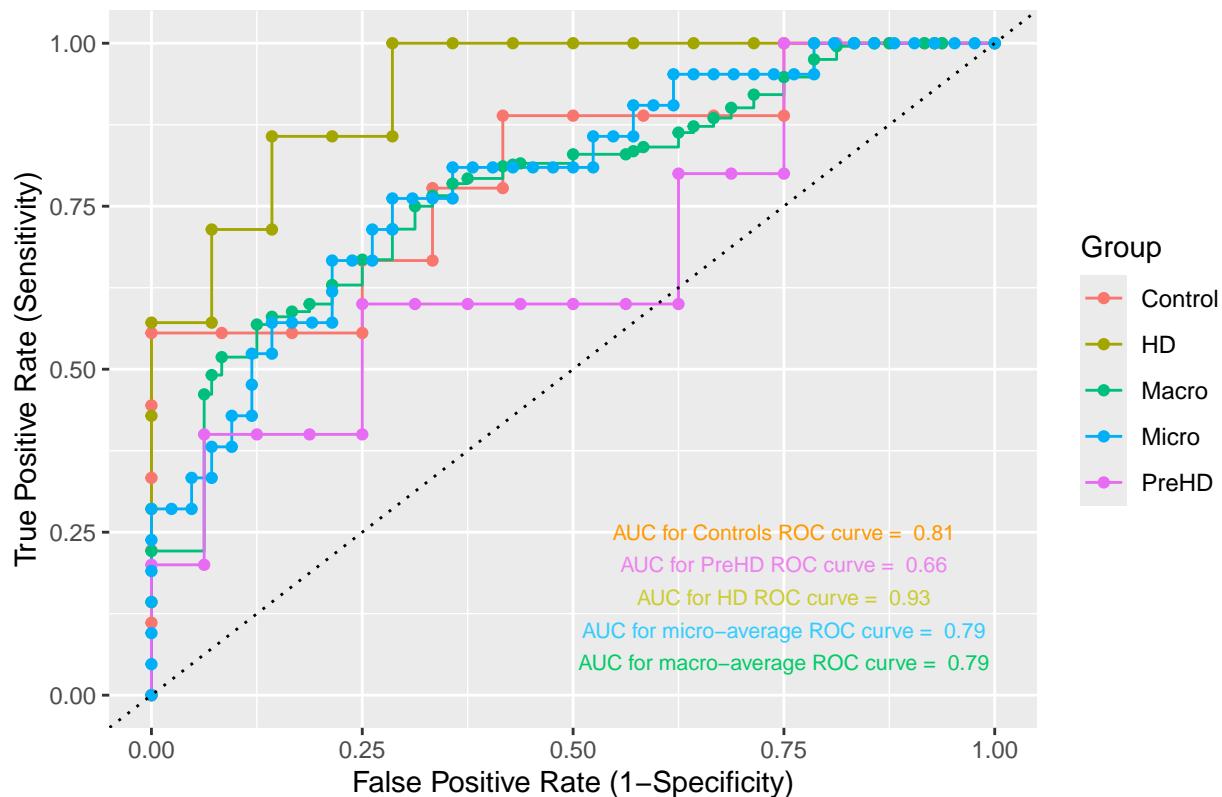
```
## ROC curves one vs the rest by multiROC package
Control_pred_ATRLDA3log<- predClas_ATRLDA3$posterior[, '1']
PreHD_pred_ATRLDA3log<- predClas_ATRLDA3$posterior[, '2']
HD_pred_ATRLDA3log<- predClas_ATRLDA3$posterior[, '3']
Control_true<-ifelse(test$ATR_subgroup=='1',1,0)
PreHD_true<-ifelse(test$ATR_subgroup=='2',1,0)
HD_true<-ifelse(test$ATR_subgroup=='3',1,0)
DF_ROC_ATRLDA3log<-cbind(Control_true,PreHD_true,HD_true,Control_pred_ATRLDA3log,
                           PreHD_pred_ATRLDA3log,HD_pred_ATRLDA3log)
multiROC_ATR_LDA3log<-multi_roc(DF_ROC_ATRLDA3log)
multiROC_ATR_LDA3log_DF<-plot_roc_data(multiROC_ATR_LDA3log)
ggplot(multiROC_ATR_LDA3log_DF, aes(x=1-Specificity, y=Sensitivity, color = Group)) +
  geom_step() + geom_point() + labs(title= "3-LDA model of both ATR ROC curve: One vs Rest",
  x = "False Positive Rate (1-Specificity)",
  y = "True Positive Rate (Sensitivity)") + geom_abline(intercept = 0, slope=1,linetype = 'dotted') +
  annotate('text',x=0.75, y=0.25,
  label=paste('AUC for Controls ROC curve = ',
  round(mean(multiROC_ATR_LDA3log_DF$AUC[multiROC_ATR_LDA3log_DF$Group=='Control']),2),
  color = '#FF9900',size = 8/.pt) +
  annotate('text',x=0.75, y=0.20,
  label=paste('AUC for PreHD ROC curve = ',
  round(mean(multiROC_ATR_LDA3log_DF$AUC[multiROC_ATR_LDA3log_DF$Group=='PreHD']),2),
  color = 'violet',size = 8/.pt) +
  annotate('text',x=0.75, y=0.15,
  label=paste('AUC for HD ROC curve = ',
  round(mean(multiROC_ATR_LDA3log_DF$AUC[multiROC_ATR_LDA3log_DF$Group=='HD']),2)),
  color = '#CCCC33',size = 8/.pt) +
  annotate('text',x=0.75, y=0.10,
  label=paste('AUC for micro-average ROC curve = ',
  round(mean(multiROC_ATR_LDA3log_DF$AUC[multiROC_ATR_LDA3log_DF$Group=='Micro']),2)),
  color = '#33CCFF',size = 8/.pt)+
```

```

annotate('text',x=0.75, y=0.05,
        label=paste('AUC for macro-average ROC curve = ',
                    round(mean(multiROC_ATR_LDA3log_DF$AUC[multiROC_ATR_LDA3log_DF$Group=='Macro']),2)),
        color = '#00CC66',size = 8/.pt)

```

### 3-LDA model of both ATR ROC curve: One vs Rest



```

## Kappa coefficient
library(irr)
kappa_ATR_LDA3log<-kappam.fleiss(data.frame(predClas_ATRLDA3$class,test$ATR_subgroup))
kappa_ATR_LDA3log
Fleiss' Kappa for m Raters

Subjects = 21
Raters = 2
Kappa = 0.294

z = 1.79
p-value = 0.0733

```

A diferencia del análisis realizado con los ATRs derecho e izquierdo por separado, los subgrupos logran diferenciarse entre sí, especialmente con el uso de las 5 métricas, a pesar de que existe solapamiento entre los 3 subgrupos.

Cuadro 80: Overall Statistics of Confusion Matrices

	LDA-5log-RATR	LDA-3log-RATR	LDA-5log-LATR	LDA-3log-LATR	LDA-5log-both ATR	LDA-3log-both ATR
Accuracy		0.444	0.444	0.667	0.667	0.600

Kappa	0.100	0.100	0.438	0.438	0.350	0.359
AccuracyLower	0.137	0.137	0.299	0.299	0.361	0.384
AccuracyUpper	0.788	0.788	0.925	0.925	0.809	0.819
AccuracyNull	0.444	0.444	0.444	0.444	0.450	0.429
AccuracyPValue	0.626	0.626	0.157	0.157	0.131	0.062
McnemarPValue	0.506	0.506	0.392	NaN	0.446	NaN
Fleiss_Kappa	0.062	0.062	0.413	0.393	0.337	0.294

Cuadro 81: Statistics by Class of Confusion Matrices - Table 1

	Sensitivity	Specificity	Pos Pred Value	Neg Pred Value	Precision	Recall	F1
<b>LDA-5log for RATR</b>							
Class: Control	0.750	0.400	0.500	0.667	0.500	0.750	0.600
Class: PreHD	0.000	0.714	0.000	0.714	0.000	0.000	NaN
Class: HD	0.333	1.000	1.000	0.750	1.000	0.333	0.500
<b>LDA-3log for RATR</b>							
Class: Control	0.750	0.400	0.500	0.667	0.500	0.750	0.600
Class: PreHD	0.000	0.714	0.000	0.714	0.000	0.000	NaN
Class: HD	0.333	1.000	1.000	0.750	1.000	0.333	0.500
<b>LDA-5log for LATR</b>							
Class: Control	1.000	0.600	0.667	1.000	0.667	1.000	0.800
Class: PreHD	0.000	1.000	NAN	0.778	NA	0.000	NA
Class: HD	0.667	0.833	0.667	0.833	0.667	0.667	0.667
<b>LDA-3log for LATR</b>							
Class: Control	1.000	0.400	0.571	1.000	0.571	1.000	0.727
Class: PreHD	0.500	1.000	1.000	0.875	1.000	0.500	0.667
Class: HD	0.333	1.000	1.000	0.750	1.000	0.333	0.500
<b>LDA-5log for both ATR</b>							
Class: Control	0.889	0.636	0.667	0.875	0.667	0.889	0.762
Class: PreHD	0.200	0.933	0.500	0.778	0.500	0.200	0.286
Class: HD	0.500	0.786	0.500	0.786	0.500	0.500	0.500
<b>LDA-3log for both ATR</b>							
Class: Control	1.000	0.333	0.529	1.000	0.529	1.000	0.692
Class: PreHD	0.200	1.000	1.000	0.800	1.000	0.200	0.333
Class: HD	0.429	1.000	1.000	0.778	1.000	0.429	0.600

Cuadro 82: Statistics by Class of Confusion Matrices - Table 2

	Prevalence	Detection Rate	Detection Prevalence	Balanced Accuracy
<b>LDA-5log for RATR</b>				
Class: Control	0.444	0.333	0.667	0.575
Class: PreHD	0.222	0.000	0.222	0.357
Class: HD	0.333	0.111	0.111	0.667
<b>LDA-3log for RATR</b>				
Class: Control	0.444	0.333	0.667	0.575
Class: PreHD	0.222	0.000	0.222	0.357
Class: HD	0.333	0.111	0.111	0.667
<b>LDA-5log for LATR</b>				
Class: Control	0.444	0.444	0.667	0.800
Class: PreHD	0.222	0.000	0.000	0.500
Class: HD	0.333	0.222	0.333	0.750
<b>LDA-3log for LATR</b>				
Class: Control	0.444	0.444	0.778	0.700
Class: PreHD	0.222	0.111	0.111	0.750
Class: HD	0.333	0.111	0.111	0.667
<b>LDA-5log for both ATR</b>				
Class: Control	0.450	0.400	0.600	0.763
Class: PreHD	0.250	0.050	0.100	0.567
Class: HD	0.300	0.150	0.300	0.643
<b>LDA-3log for both ATR</b>				

Class: Control	0.429	0.429	0.810	0.667
Class: PreHD	0.238	0.048	0.048	0.600
Class: HD	0.333	0.143	0.143	0.714

## 5.2 PERMANOVA

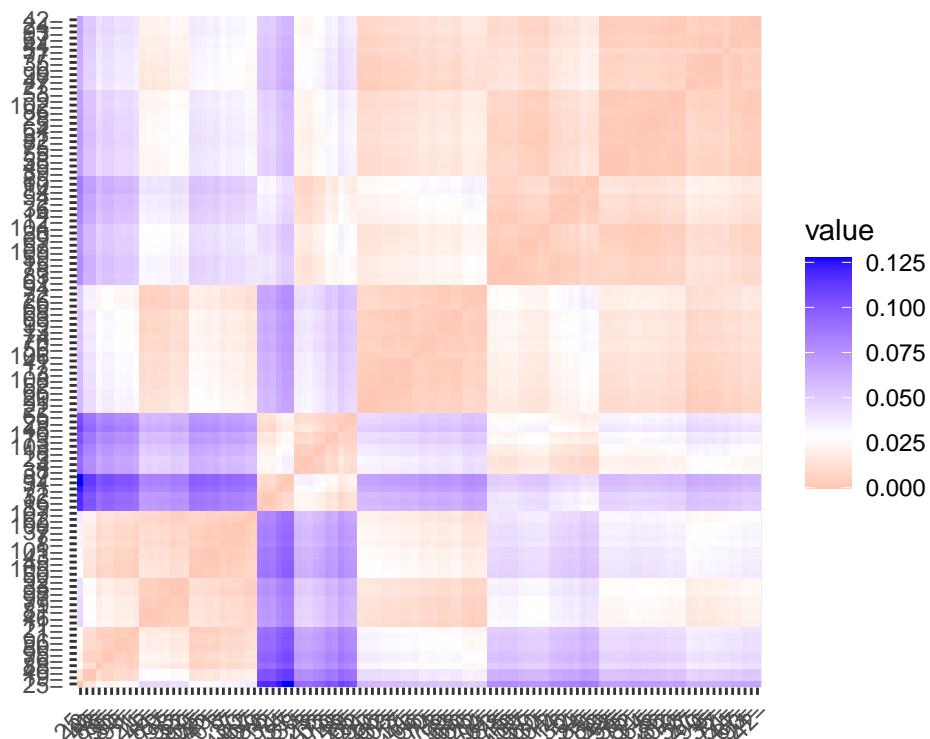
Como en el caso de los PERMANOVAS generados para los ATR derecho e izquierdo por separado, a pesar que PERMANOVA corresponde a una técnica no paramétrica, se decide generar PERMANOVAS con los dataframes originales (ATR\_DifIron\_Df, ATR\_DifIron\_ADRDMD, ATR\_DifIron\_IronFA) y con los dataframes generados con la eliminación de los outliers que se ha realizado en la normalidad multivariante (ATR\_DifIron\_Df\_WoOut, ATR\_DifIron\_WoOut\_ADRDMD, ATR\_DifIron\_WoOut\_IronFA).

### 5.2.1 Matriz de distancias

Se ha procedido a **calcular la matriz de distancias** con la función `daisy` del paquete `cluster` para poder visualizar las similaridades/disimilaridades entre subgrupos mediante MDS y tSNE. La **medida de distancia utilizada** ha sido *euclidean* dado que las variables dependientes (métricas de RM) son numéricas y no procedentes de especies.

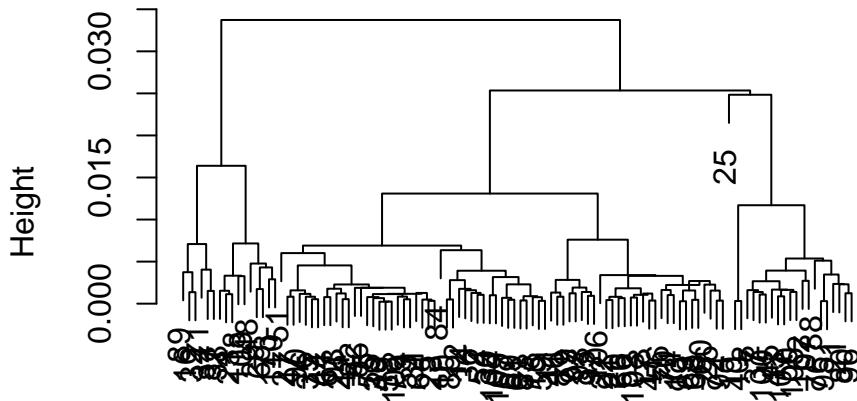
```
# Matriz de distancias para las 5 métricas - Sin eliminación de outliers
library(cluster)
dist<-daisy(ATR_DifIron_Df[,-c(1:6,12)],metric='euclidean')
distmx<-as.matrix(dist)

library(factoextra)
fviz_dist(dist)
```



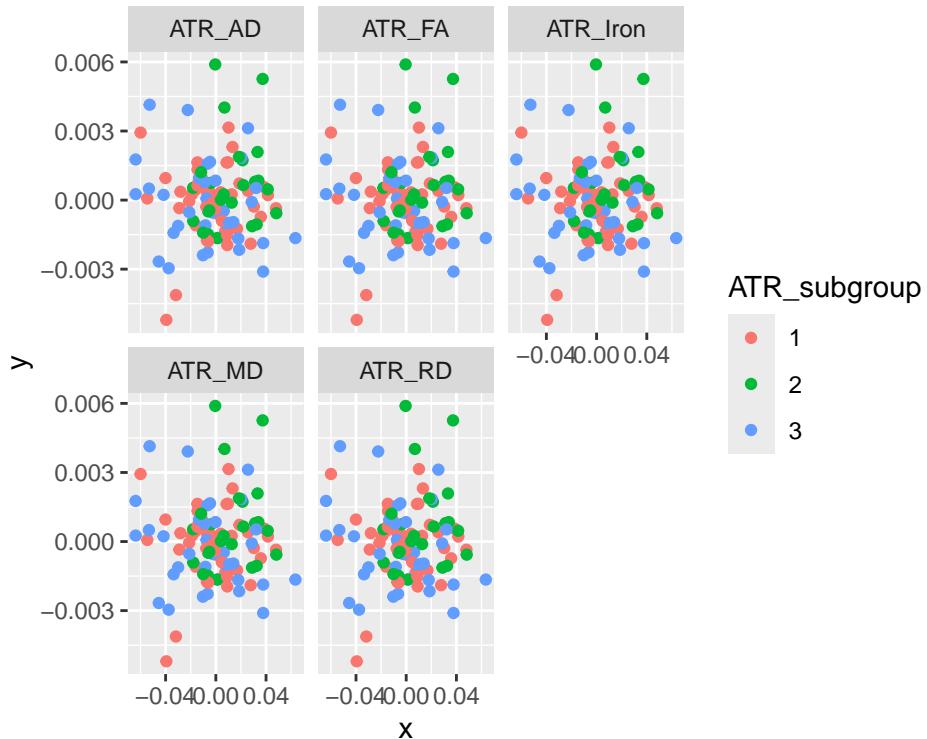
```
clust<-hclust(dist, method = "centroid", members = NULL)
plot(clust)
```

## Cluster Dendrogram



```
dist  
hclust (*, "centroid")
```

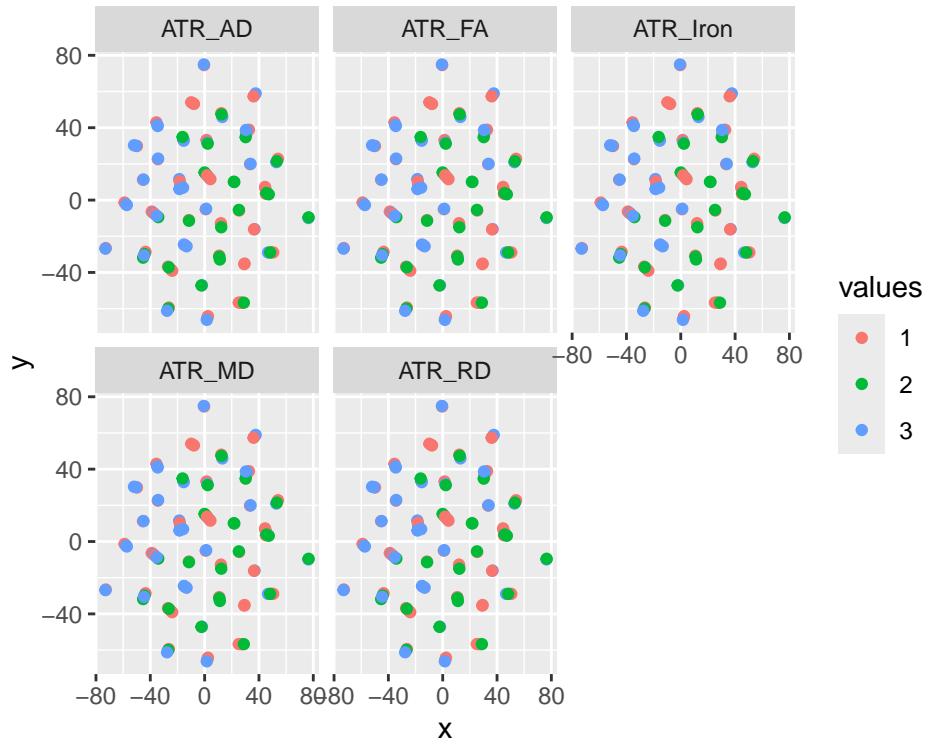
```
mds_dist<-cmdscale(distmx, eig=TRUE, k=2)  
df_mds_dist<-data.frame(ATR_DifIron_Df, x=mds_dist$points[,1],  
                           y=mds_dist$points[,2])  
df_mds_dist2<-df_mds_dist %>%  
  dplyr::select(x, y, ATR_subgroup, ATR_hemisphere, ATR_Gender, ATR_ägef2,  
                ATR_Iron, ATR_AD, ATR_RD, ATR_MD, ATR_FA) %>%  
  tidyverse::gather (key='variable', value='values', -x, -y, -ATR_subgroup,  
                    -ATR_hemisphere, -ATR_Gender, -ATR_ägef2)  
library(ggplot2)  
ggplot(df_mds_dist2,aes(x,y)) + geom_jitter(aes(col=ATR_subgroup)) + facet_wrap(~variable)
```



```

library(Rtsne)
my_Rtsne<-Rtsne(X= distmx,perplexity= floor((nrow(ATR_DifIron_Df) - 1) / 3),dims = 2)
# Maximum perplexity
my_Rtsne$perplexity
[1] 36
fit_tsne <- Rtsne(dist,is_distance = TRUE, perplexity=1, check_duplicates = FALSE,
                     pca = FALSE, dims = 2, initial_dims = 50)
df_tsne_vars <- as_tibble(cbind(ATR_DifIron_Df,
                                  x = fit_tsne$Y[, 1], y = fit_tsne$Y[, 2]))
df_tsne <- df_tsne_vars %>%
  dplyr::select(x, y, ATR_subgroup, ATR_hemisphere, ATR_Gender, ATR_agef2,
                ATR_Iron, ATR_AD, ATR_RD, ATR_MD, ATR_FA) %>%
  tidyverse::gather(key = "variable", value = "values", -x, -y, -ATR_subgroup,
                    -ATR_hemisphere, -ATR_Gender, -ATR_agef2)
ggplot(df_tsne,aes(x, y, col=values)) +  geom_point(aes(col=ATR_subgroup)) + facet_wrap(~variable)

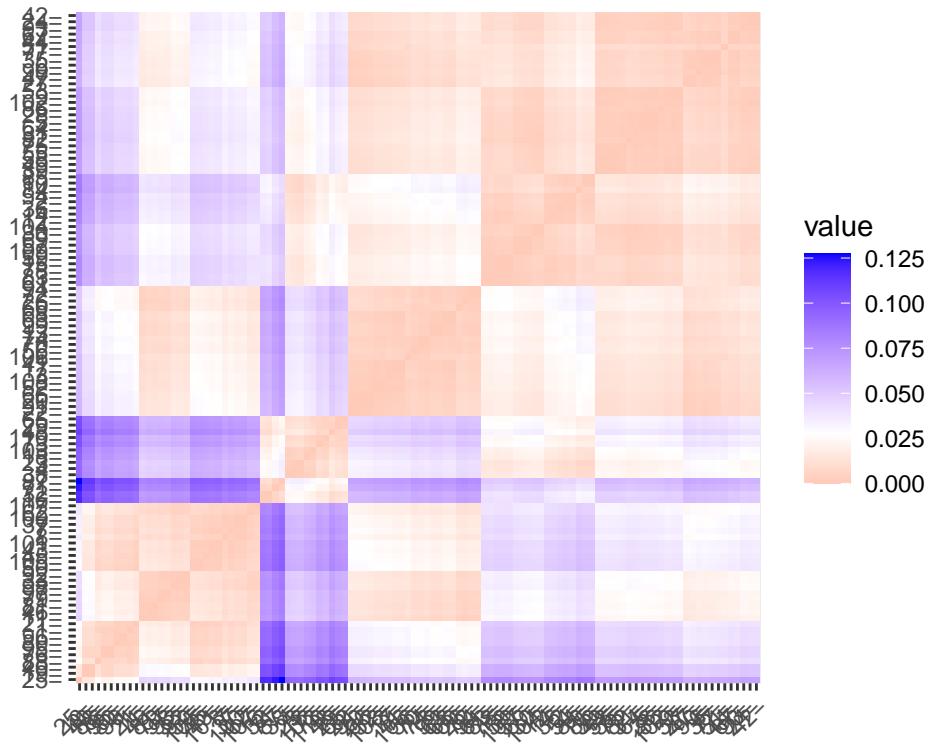
```



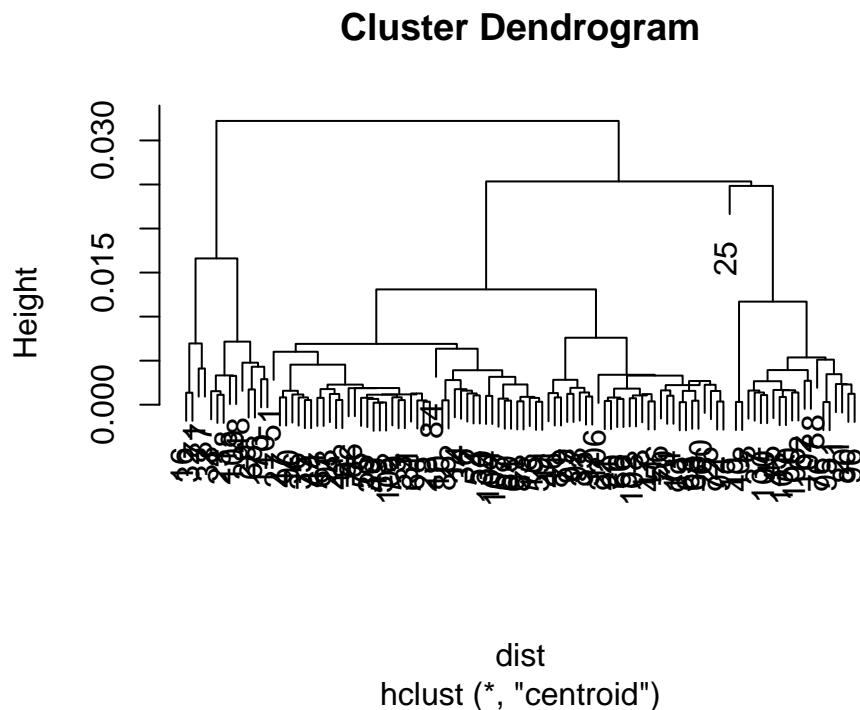
Según los gráficos obtenidos con el dataframe con todos los sujetos (sin eliminación de outliers) y con las matriz de distancias calculada con las 5 métricas de RM, no se consigue una separación entre subgrupos.

```
# Matriz de distancias para las 5 métricas - Con eliminación de outliers
library(cluster)
dist<-daisy(ATR_DifIron_Df_WoOut[,-c(1:6,12)],metric='euclidean')
distmx<-as.matrix(dist)

library(factoextra)
fviz_dist(dist)
```



```
clust<-hclust(dist, method = "centroid", members = NULL)
plot(clust)
```

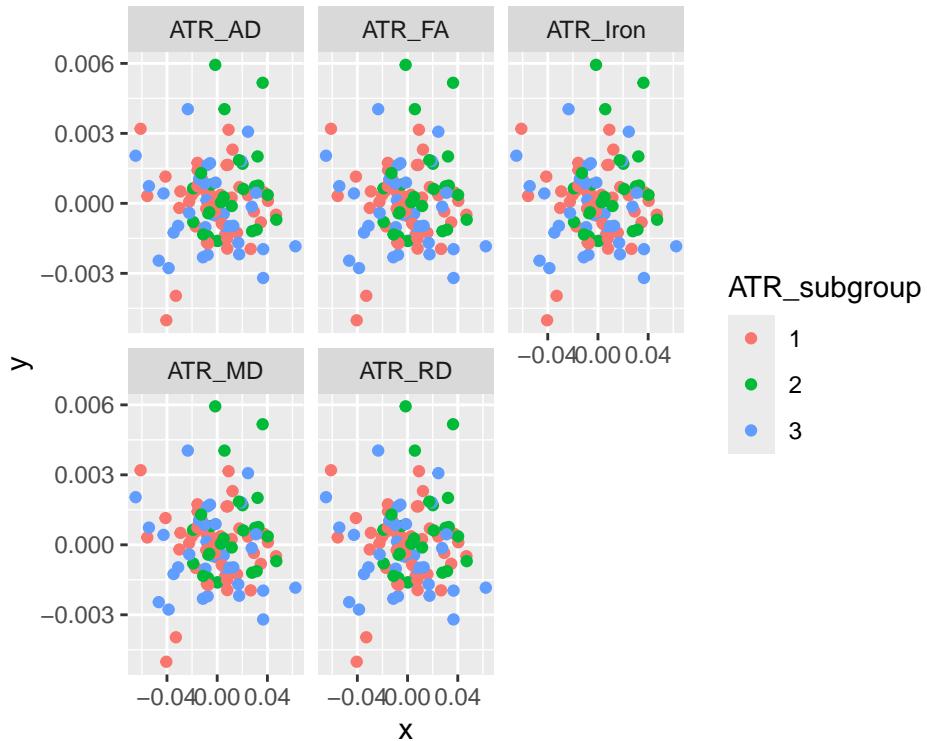


```
mds_dist<-cmdscale(distmx, eig=TRUE, k=2)
df_mds_dist<-data.frame(ATR_DifIron_Df_WoOut, x=mds_dist$points[,1],
```

```

y=mds_dist$points[,2])
df_mds_dist2<-df_mds_dist %>%
  dplyr::select(x, y, ATR_subgroup, ATR_hemisphere, ATR_Gender, ATR_ägef2,
                ATR_Iron, ATR_AD, ATR_RD, ATR_MD, ATR_FA) %>%
  tidyverse::gather(key='variable', value='values', -x, -y, -ATR_subgroup,
                    -ATR_hemisphere, -ATR_Gender, -ATR_ägef2)
library(ggplot2)
ggplot(df_mds_dist2,aes(x,y)) + geom_jitter(aes(col=ATR_subgroup)) + facet_wrap(~variable)

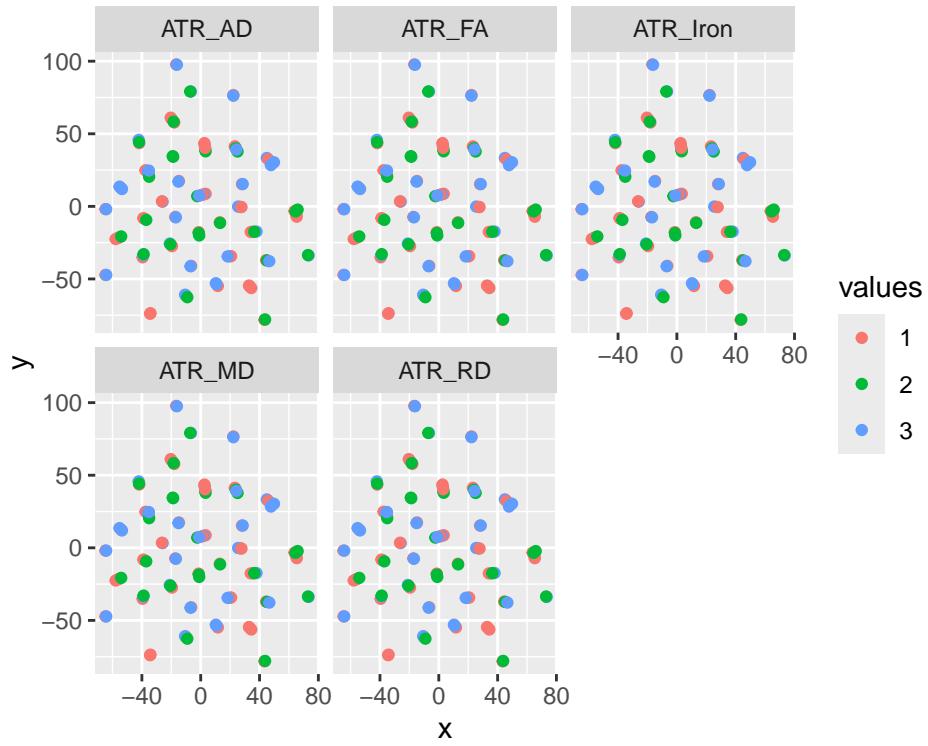
```



```

library(Rtsne)
my_Rtsne<-Rtsne(X= distmx,perplexity= floor((nrow(ATR_DifIron_Df_WoOut) - 1) / 3),dims = 2)
# Maximum perplexity
my_Rtsne$perplexity
[1] 35
fit_tsne <- Rtsne(dist,is_distance = TRUE, perplexity=1, check_duplicates = FALSE,
                     pca = FALSE, dims = 2, initial_dims = 50)
df_tsne_vars <- as_tibble(cbind(ATR_DifIron_Df_WoOut,
                                    x = fit_tsne$Y[, 1], y = fit_tsne$Y[, 2]))
df_tsne <- df_tsne_vars %>%
  dplyr::select(x, y, ATR_subgroup, ATR_hemisphere, ATR_Gender, ATR_ägef2,
                ATR_Iron, ATR_AD, ATR_RD, ATR_MD, ATR_FA) %>%
  tidyverse::gather(key = "variable", value = "values", -x, -y, -ATR_subgroup,
                    -ATR_hemisphere, -ATR_Gender, -ATR_ägef2)
ggplot(df_tsne,aes(x, y, col=values)) + geom_point(aes(col=ATR_subgroup)) + facet_wrap(~variable)

```

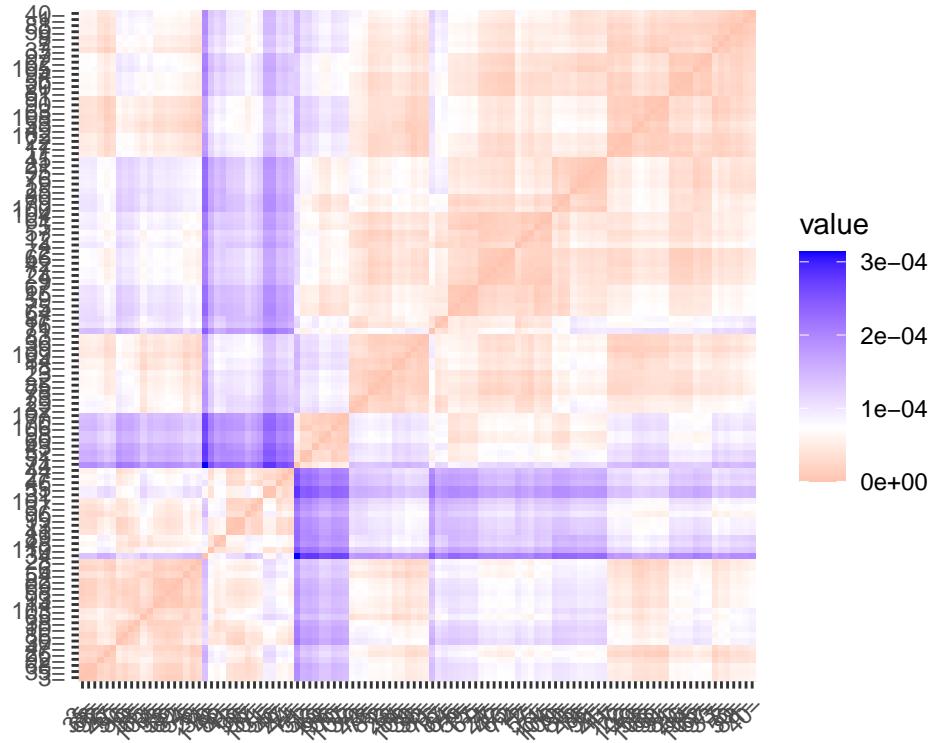


Según los gráficos obtenidos con el dataframe con eliminación de outliers y con las matriz de distancias calculada con las 5 métricas de RM, tampoco se consigue una separación entre subgrupos.

Se procederá de la misma forma para las 3 métricas de difusividad y las métricas iron-FA.

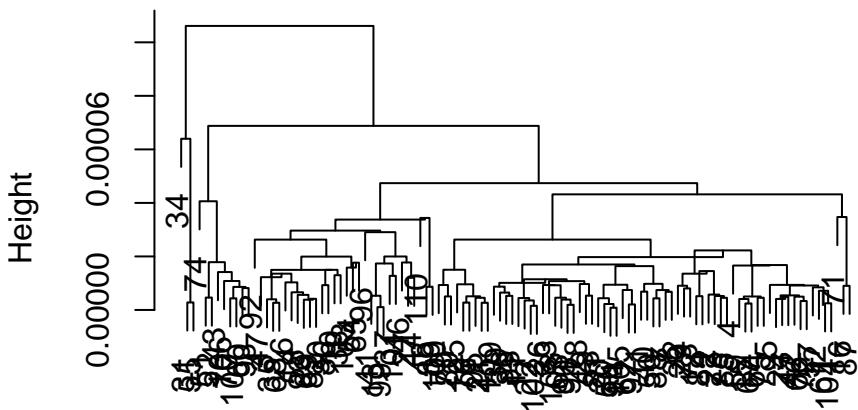
```
# Matriz de distancias para las 3 métricas - Sin eliminación de outliers
library(cluster)
dist<-daisy(ATR_DifIron_Df_ADRDMD[,-c(1:6,12)],metric='euclidean')
distmx<-as.matrix(dist)

library(factoextra)
fviz_dist(dist)
```



```
clust<-hclust(dist, method = "centroid", members = NULL)
plot(clust)
```

**Cluster Dendrogram**



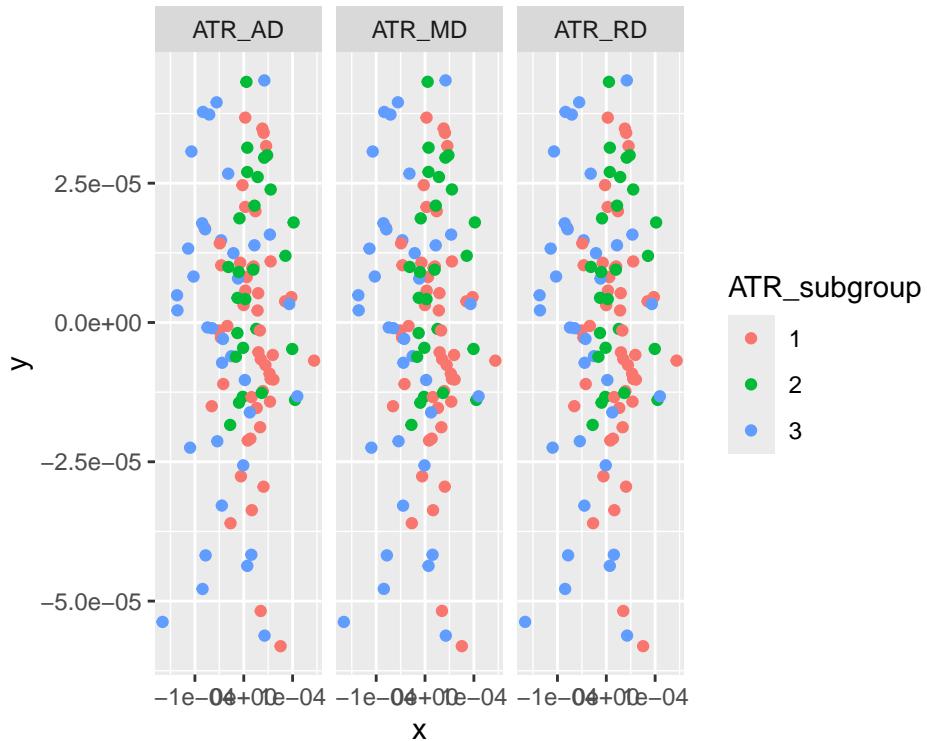
dist  
hclust (\*, "centroid")

```
mds_dist<-cmdscale(distmx, eig=TRUE, k=2)
df_mds_dist<-data.frame(ATR_DifIron_Df_ADRDMD, x=mds_dist$points[,1],
```

```

y=mds_dist$points[,2])
df_mds_dist2<-df_mds_dist %>%
  dplyr::select(x, y, ATR_subgroup, ATR_hemisphere, ATR_Gender, ATR_agef2,
                ATR_AD, ATR_RD, ATR_MD) %>%
  tidyverse::gather(key='variable', value='values', -x, -y, -ATR_subgroup,
                    -ATR_hemisphere, -ATR_Gender, -ATR_agef2)
library(ggplot2)
ggplot(df_mds_dist2,aes(x,y)) + geom_jitter(aes(col=ATR_subgroup)) + facet_wrap(~variable)

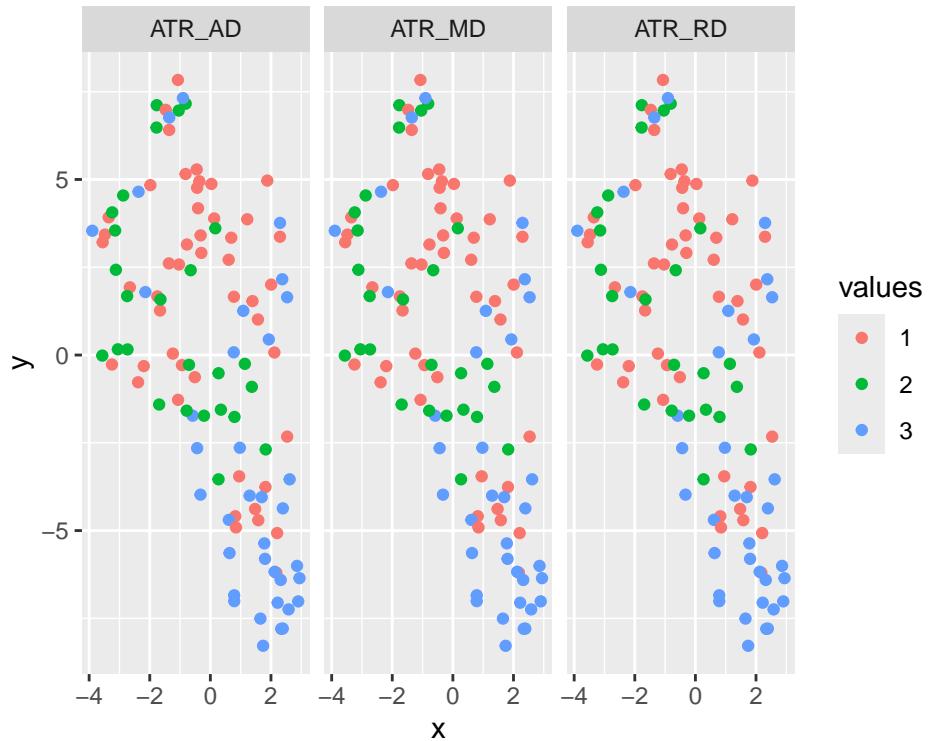
```



```

library(Rtsne)
my_Rtsne<-Rtsne(X= distmx,perplexity= floor((nrow(ATR_DifIron_Df_ADRMD) - 1) / 3),dims = 2)
# Maximum perplexity
my_Rtsne$perplexity
[1] 36
fit_tsne <- Rtsne(dist,is_distance = TRUE,check_duplicates = FALSE,pca = FALSE,dims = 2)
df_tsne_vars <- as_tibble(cbind(ATR_DifIron_Df_ADRMD,
                                    x = fit_tsne$Y[, 1], y = fit_tsne$Y[, 2]))
df_tsne <- df_tsne_vars %>%
  dplyr::select(x, y, ATR_subgroup, ATR_hemisphere, ATR_Gender, ATR_agef2,
                ATR_AD, ATR_RD, ATR_MD) %>%
  tidyverse::gather(key = "variable", value = "values", -x, -y, -ATR_subgroup,
                    -ATR_hemisphere, -ATR_Gender, -ATR_agef2)
ggplot(df_tsne,aes(x, y, col=values)) + geom_point(aes(col=ATR_subgroup)) + facet_wrap(~variable)

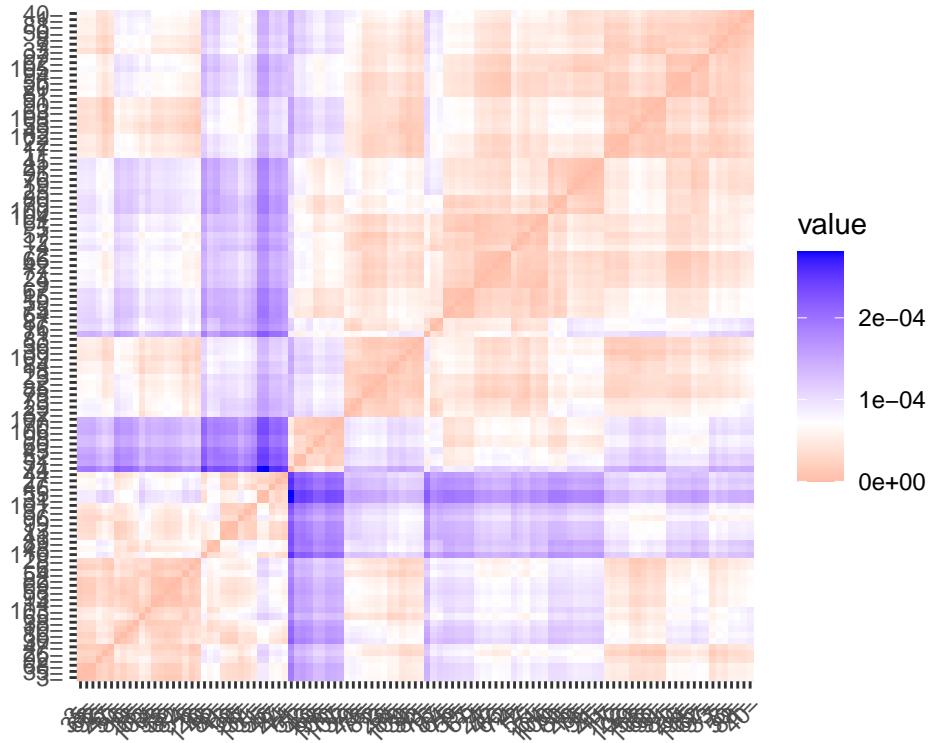
```



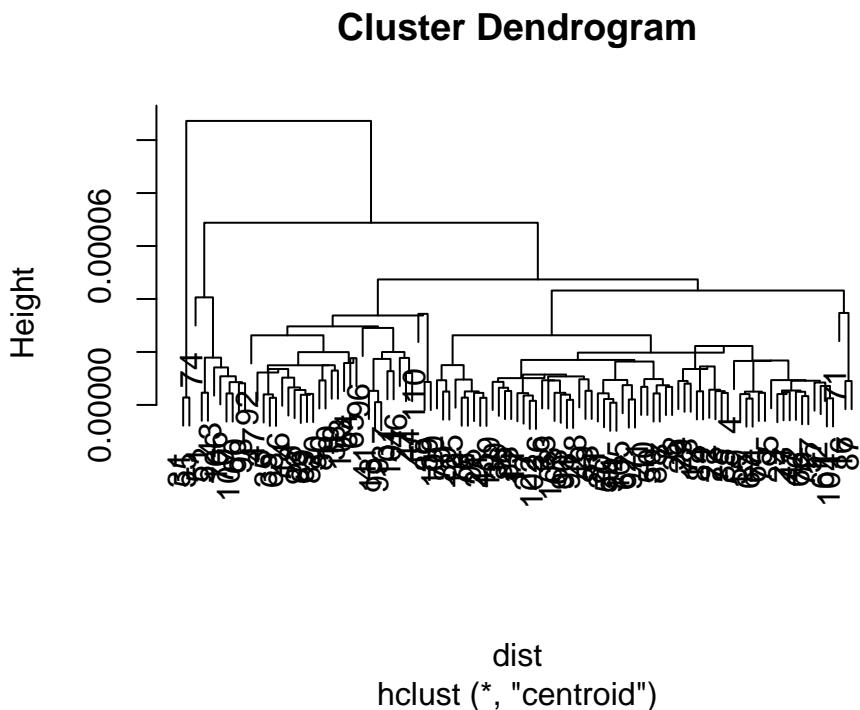
Según los gráficos obtenidos con el dataframe con todos los sujetos (sin eliminación de outliers) y con las matriz de distancias calculada con las 3 métricas de difusividad, se consigue una separación entre el subgrupo de pacientes sintomático (HD) y los otros dos subgrupos (controles y sujetos presintomáticos preHD).

```
# Matriz de distancias para las 3 métricas - Con eliminación de outliers
library(cluster)
dist<-daisy(ATR_DifIron_Df_WoOut_ADRDMD[,-c(1:6,12)],metric='euclidean')
distmx<-as.matrix(dist)

library(factoextra)
fviz_dist(dist)
```



```
clust<-hclust(dist, method = "centroid", members = NULL)
plot(clust)
```

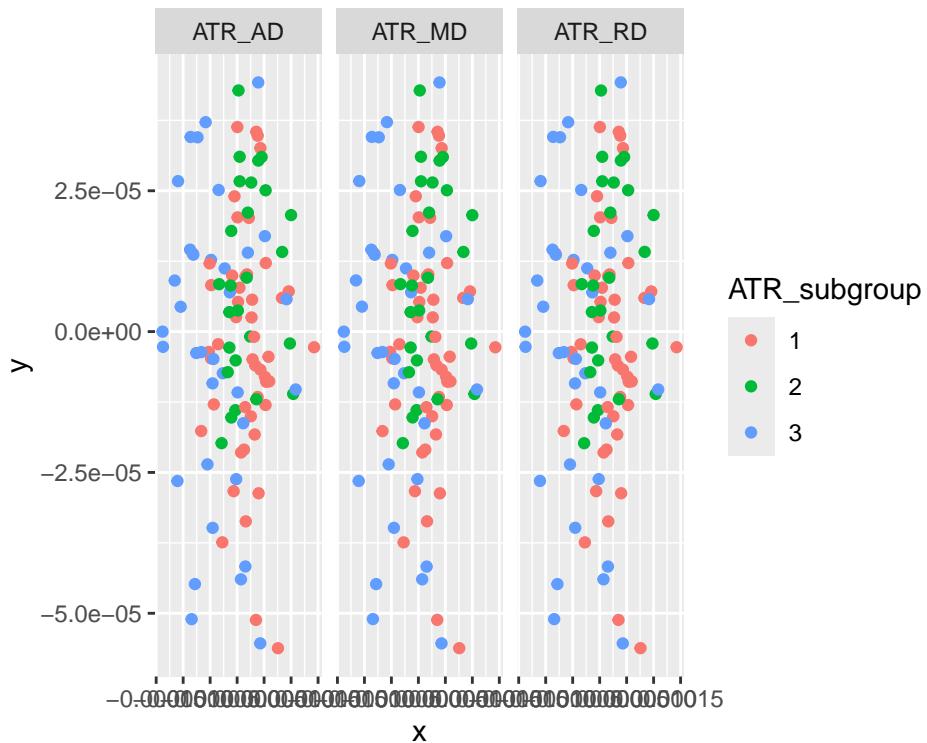


```
mds_dist<-cmdscale(distmx, eig=TRUE, k=2)
df_mds_dist<-data.frame(ATR_DifIron_Df_WoOut_ADRDMD, x=mds_dist$points[,1],
```

```

y=mds_dist$points[,2])
df_mds_dist2<-df_mds_dist %>%
  dplyr::select(x, y, ATR_subgroup, ATR_hemisphere, ATR_Gender, ATR_agef2,
                ATR_AD, ATR_RD, ATR_MD) %>%
  tidyverse::gather(key='variable', value='values', -x, -y, -ATR_subgroup,
                    -ATR_hemisphere, -ATR_Gender, -ATR_agef2)
library(ggplot2)
ggplot(df_mds_dist2,aes(x,y)) + geom_jitter(aes(col=ATR_subgroup)) + facet_wrap(~variable)

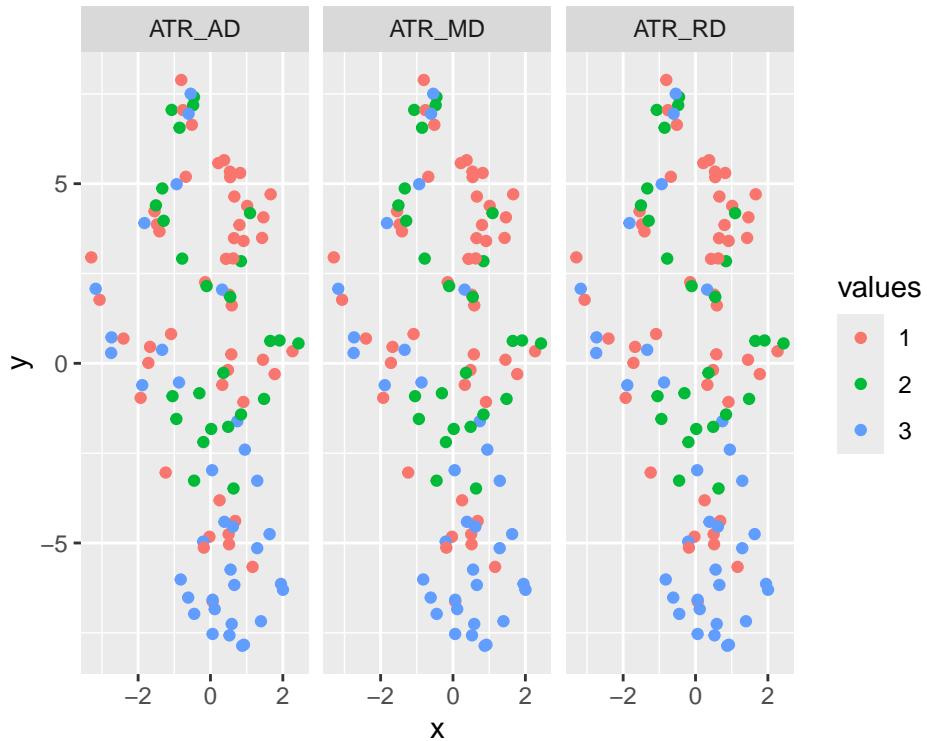
```



```

library(Rtsne)
my_Rtsne<-Rtsne(X= distmx,perplexity= floor((nrow(ATR_DifIron_Df_WoOut_ADRDMD) - 1) / 3),dims = 2)
# Maximum perplexity
my_Rtsne$perplexity
[1] 36
fit_tsne <- Rtsne(dist,is_distance = TRUE,check_duplicates = FALSE,pca = FALSE,dims = 2)
df_tsne_vars <- as_tibble(cbind(ATR_DifIron_Df_WoOut_ADRDMD,
                                    x = fit_tsne$Y[, 1], y = fit_tsne$Y[, 2]))
df_tsne <- df_tsne_vars %>%
  dplyr::select(x, y, ATR_subgroup, ATR_hemisphere, ATR_Gender, ATR_agef2,
                ATR_AD, ATR_RD, ATR_MD) %>%
  tidyverse::gather(key = "variable", value = "values", -x, -y, -ATR_subgroup,
                    -ATR_hemisphere, -ATR_Gender, -ATR_agef2)
ggplot(df_tsne,aes(x, y, col=values)) + geom_point(aes(col=ATR_subgroup)) + facet_wrap(~variable)

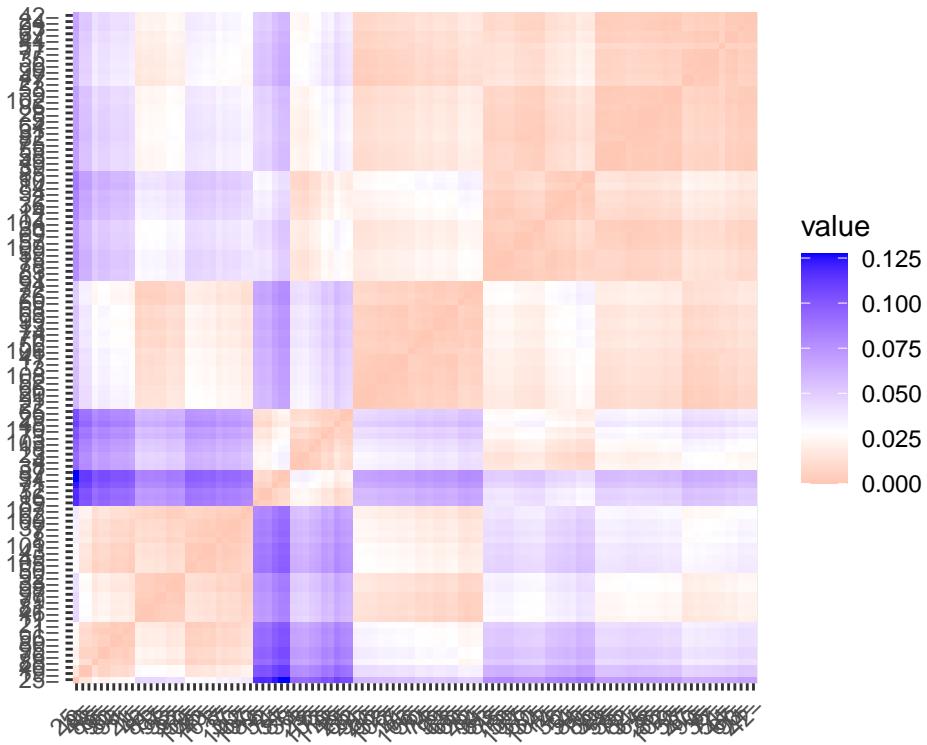
```



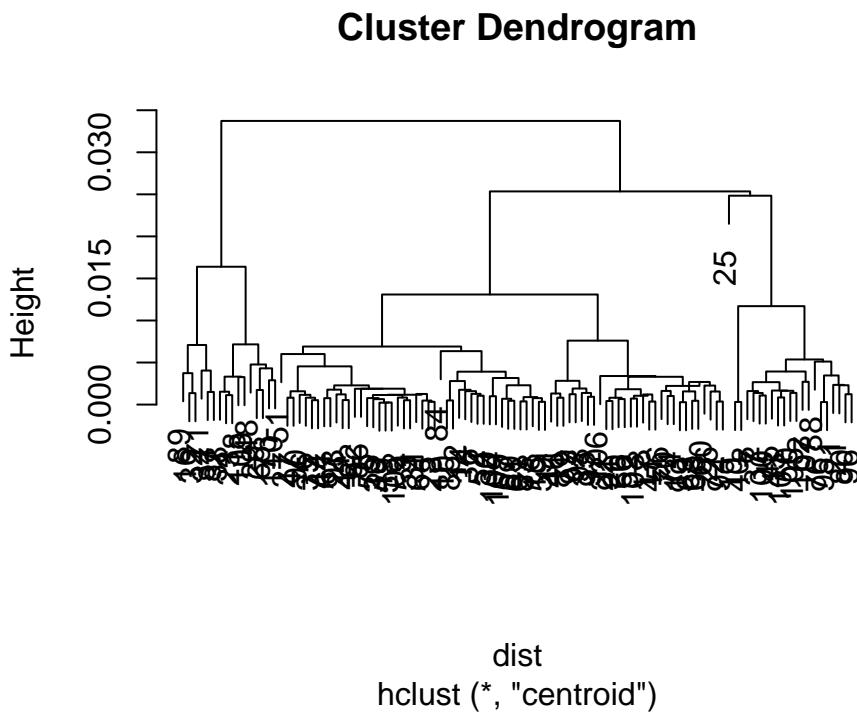
Según los gráficos obtenidos con eliminación de outliers y con las matrices de distancias calculada con las 3 métricas de difusividad, también se consigue una separación entre el subgrupo de pacientes sintomático (HD) y los otros dos subgrupos (controles y sujetos presintomáticos preHD), pero el uso del dataframe reducido (eliminando los outliers) no mejora la separación.

```
# Matriz de distancias para las métricas de iron y FA - Sin eliminación de outliers
library(cluster)
dist<-daisy(ATR_DifIron_Df_IronFA[,-c(1:6,12)],metric='euclidean')
distmx<-as.matrix(dist)

library(factoextra)
fviz_dist(dist)
```



```
clust<-hclust(dist, method = "centroid", members = NULL)
plot(clust)
```

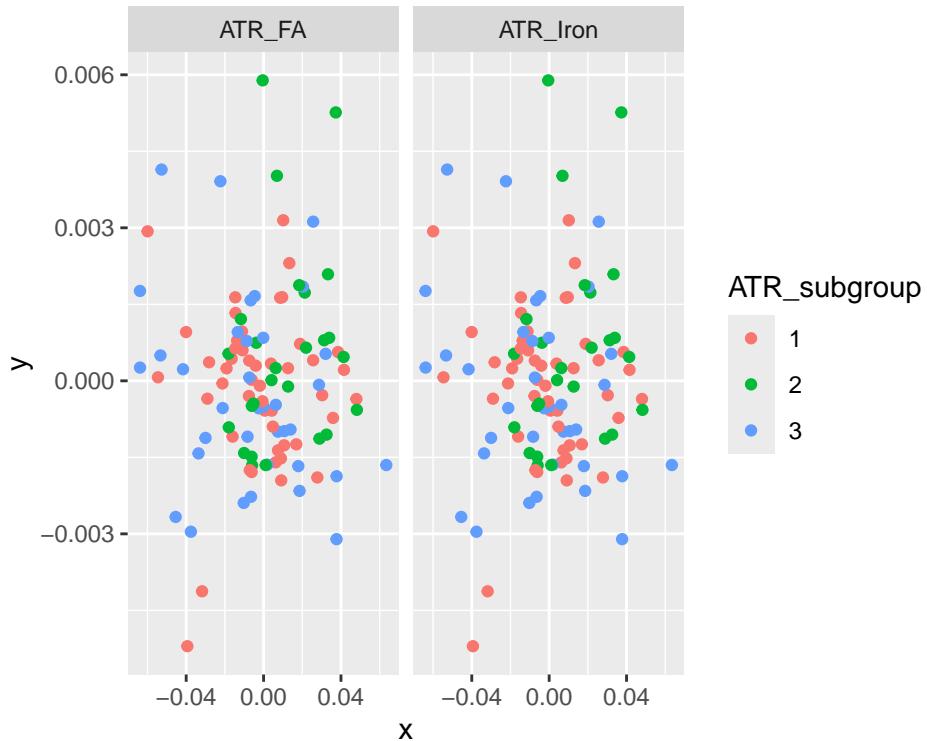


```
mds_dist<-cmdscale(distmx, eig=TRUE, k=2)
df_mds_dist<-data.frame(ATR_DifIron_Df_IronFA, x=mds_dist$points[,1],
```

```

y=mds_dist$points[,2])
df_mds_dist2<-df_mds_dist %>%
  dplyr::select(x, y, ATR_subgroup, ATR_hemisphere, ATR_Gender, ATR_agef2,
                ATR_Iron, ATR_FA) %>%
  tidyrr::gather(key='variable', value='values', -x, -y, -ATR_subgroup,
                 -ATR_hemisphere, -ATR_Gender, -ATR_agef2)
library(ggplot2)
ggplot(df_mds_dist2,aes(x,y)) + geom_jitter(aes(col=ATR_subgroup)) + facet_wrap(~variable)

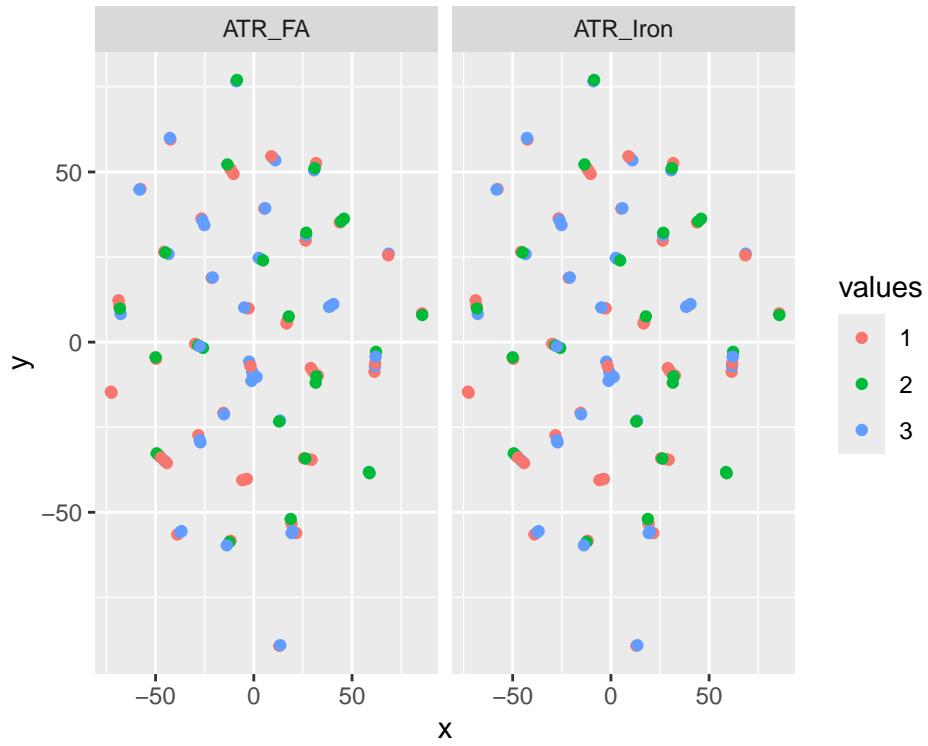
```



```

library(Rtsne)
my_Rtsne<-Rtsne(X= distmx,perplexity= floor((nrow(ATR_DifIron_Df_IronFA) - 1) / 3),dims = 2)
# Maximum perplexity
my_Rtsne$perplexity
[1] 36
fit_tsne <- Rtsne(dist,is_distance = TRUE, perplexity=1, check_duplicates = FALSE,
                     pca = FALSE,dims = 2, initial_dims=50)
df_tsne_vars <- as_tibble(cbind(ATR_DifIron_Df_IronFA,
                                   x = fit_tsne$Y[, 1], y = fit_tsne$Y[, 2]))
df_tsne <- df_tsne_vars %>%
  dplyr::select(x, y, ATR_subgroup, ATR_hemisphere, ATR_Gender, ATR_agef2,
                ATR_Iron, ATR_FA) %>%
  tidyrr::gather(key = "variable", value = "values", -x, -y, -ATR_subgroup,
                 -ATR_hemisphere, -ATR_Gender, -ATR_agef2)
ggplot(df_tsne,aes(x, y, col=values)) + geom_point(aes(col=ATR_subgroup)) + facet_wrap(~variable)

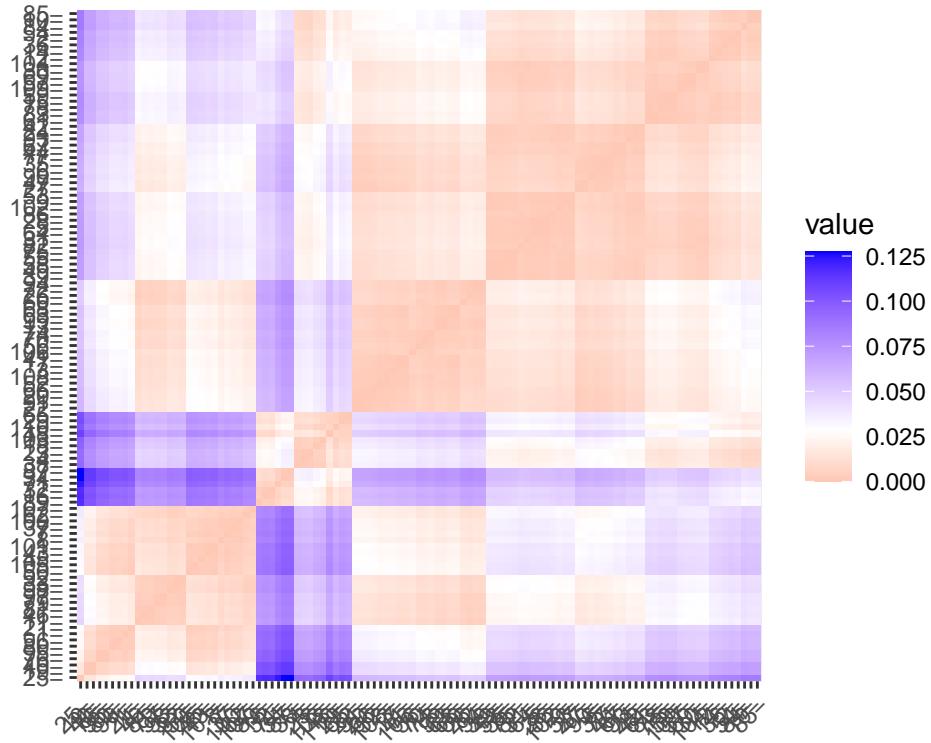
```



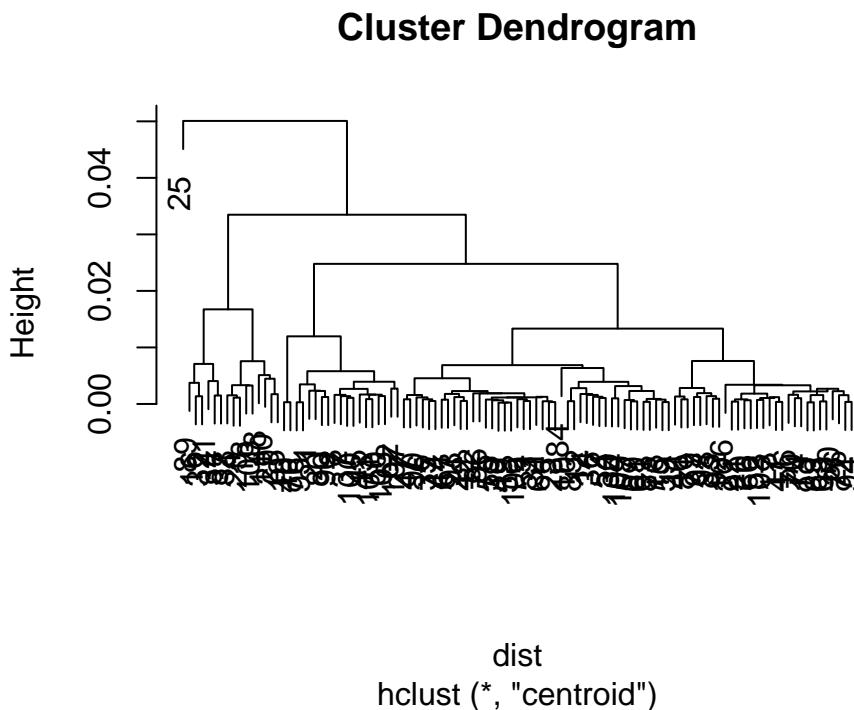
Según los gráficos obtenidos con el dataframe con todos los sujetos (sin eliminación de outliers) y con las matriz de distancias calculada con las métricas de iron y FA, no se consigue una separación entre subgrupos.

```
# Matriz de distancias para las métricas de iron y FA - Con eliminación de outliers
library(cluster)
dist<-daisy(ATR_DifIron_Df_WoOut_IronFA[,-c(1:6,12)],metric='euclidean')
distmx<-as.matrix(dist)

library(factoextra)
fviz_dist(dist)
```



```
clust<-hclust(dist, method = "centroid", members = NULL)
plot(clust)
```

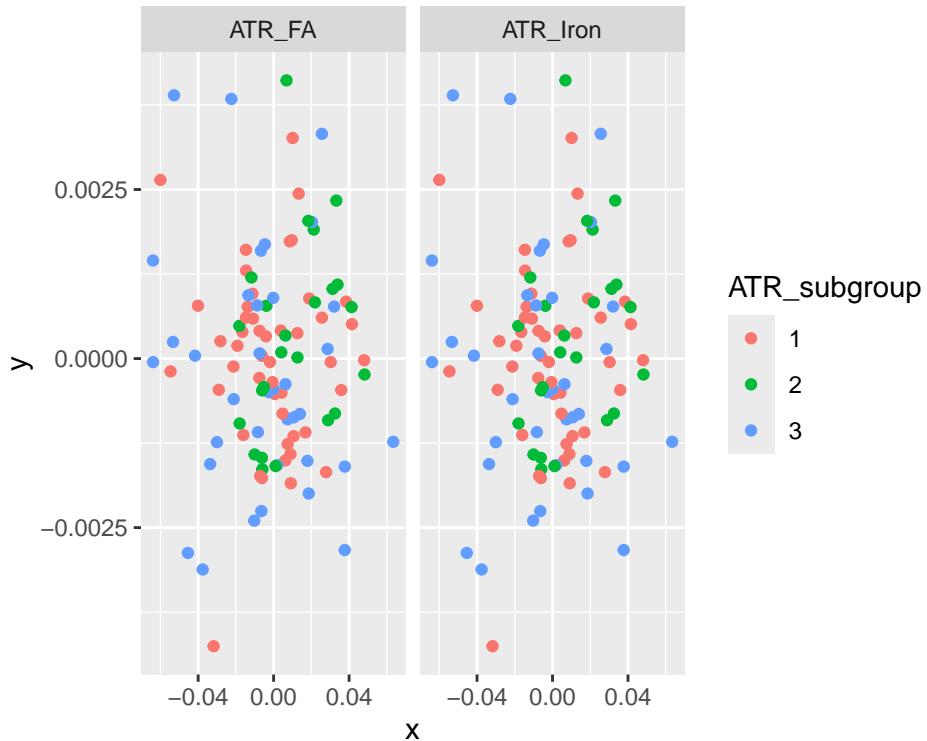


```
mds_dist<-cmdscale(distmx, eig=TRUE, k=2)
df_mds_dist<-data.frame(ATR_DifIron_Df_WoOut_IronFA, x=mds_dist$points[,1],
```

```

y=mds_dist$points[,2])
df_mds_dist2<-df_mds_dist %>%
  dplyr::select(x, y, ATR_subgroup, ATR_hemisphere, ATR_Gender, ATR_agef2,
                ATR_Iron, ATR_FA) %>%
  tidyrr::gather(key='variable', value='values', -x, -y, -ATR_subgroup,
                 -ATR_hemisphere, -ATR_Gender, -ATR_agef2)
library(ggplot2)
ggplot(df_mds_dist2,aes(x,y)) + geom_jitter(aes(col=ATR_subgroup)) + facet_wrap(~variable)

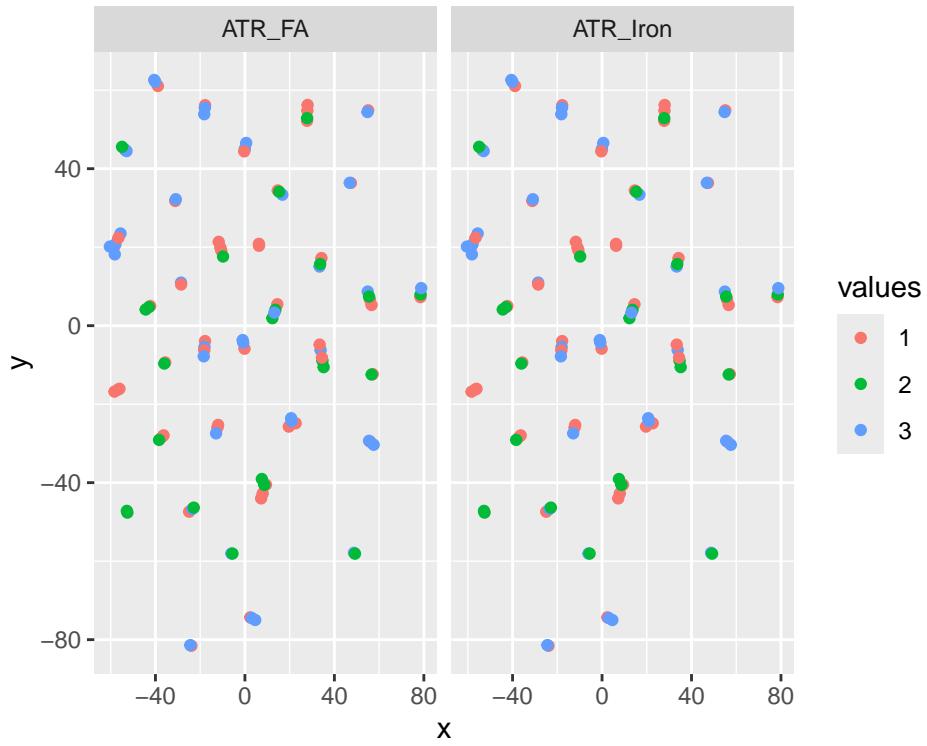
```



```

library(Rtsne)
my_Rtsne<-Rtsne(X= distmx,perplexity= floor((nrow(ATR_DifIron_Df_WoOut_IronFA) - 1) / 3),dims = 2)
# Maximum perplexity
my_Rtsne$perplexity
[1] 35
fit_tsne <- Rtsne(dist,is_distance = TRUE, perplexity=1, check_duplicates = FALSE,
                     pca = FALSE,dims = 2, initial_dims=50)
df_tsne_vars <- as_tibble(cbind(ATR_DifIron_Df_WoOut_IronFA,
                                  x = fit_tsne$Y[, 1], y = fit_tsne$Y[, 2]))
df_tsne <- df_tsne_vars %>%
  dplyr::select(x, y, ATR_subgroup, ATR_hemisphere, ATR_Gender, ATR_agef2,
                ATR_Iron, ATR_FA) %>%
  tidyrr::gather(key = "variable", value = "values", -x, -y, -ATR_subgroup,
                 -ATR_hemisphere, -ATR_Gender, -ATR_agef2)
ggplot(df_tsne,aes(x, y, col=values)) + geom_point(aes(col=ATR_subgroup)) + facet_wrap(~variable)

```



Según los gráficos obtenidos con el dataframe con eliminación de outliers y con las matriz de distancias calculada con las métricas de iron y FA, tampoco se consigue una separación entre subgrupos.

También se calculan las matrices de distancia con la función `vegdist` del paquete `vegan` para la combinación de métricas Iron-FA (por poder correlacionarse con el subgrupo preHD), AD-MD-RD (por poder correlacionarse con el subgrupo HD), y, para las 5 métricas en su conjunto, utilizando la distancia `euclidean` en el argumento `method` porque son variables numéricas continuas, no derivadas de especies. Dichas matrices se utilizarán para generar los modelos PERMANOVA.

Como se objetiva a continuación, se han calculado las matrices para ambos supuestos, con y sin eliminación de outliers. Teóricamente, PERMANOVA, al ser una técnica no paramétrica, podría ejecutarse con todos los sujetos ya que no necesita normalidad multivariante ni homocedasticidad de la matriz de covarianzas, pero, al basarse en permutaciones, requiere verificar una betadispersión homogénea entre grupos. Así que se generarán modelos PERMANOVA con y sin eliminación de outliers para determinar si la eliminación de outliers mejora la dispersión.

```
# Matriz de distancias para el dataframe completo
distDifIron<-vegdist(ATR_DifIron_Df[,-c(1:6,12)],method='euclidean')
distADMDRD<-vegdist(ATR_DifIron_Df_ADRMD[,c('ATR_AD','ATR_MD','ATR_RD')],method='euclidean')
distIronFA<-vegdist(ATR_DifIron_Df_IronFA[,c('ATR_Iron','ATR_FA')],method='euclidean')

# Matriz de distancias para el dataframe con eliminación de outliers
distDifIron_W0<-vegdist(ATR_DifIron_Df_WoOut[,-c(1:6,12)],method='euclidean')
distADMDRD_W0<-vegdist(ATR_DifIron_Df_WoOut_ADRMD[,c('ATR_AD','ATR_MD','ATR_RD')],method='euclidean')
distIronFA_W0<-vegdist(ATR_DifIron_Df_WoOut_IronFA[,c('ATR_Iron','ATR_FA')],method='euclidean')
```

### 5.2.2 Modelos

A continuación, se generan los diferentes modelos PERMANOVA (remarcar que no se puede añadir el efecto aleatorio). Primero, los modelos se ejecutan con las 4 variables independientes en interacción. Segundo, se eliminan los términos (interacciones y/o factores) no significativos para simplificar los modelos.

```

# Establecer seed para que el resultado de las permutaciones sea siempre el mismo
set.seed(123)

# Modelos para el dataframe completo
## Difusión-Iron
permDifIron1<-adonis2(distDifIron~ATR_subgroup*ATR_hemisphere*ATR_agef2*ATR_Gender,
                        data=ATR_DifIron_Df, permutations=999, by='terms')
permDifIron1[1:5]

```

	Df	SumOfSqs	R2	F
ATR_subgroup	2	0.004721	0.06489	3.3891
ATR_hemisphere	1	0.000076	0.00104	0.1086
ATR_agef2	2	0.001440	0.01979	1.0334
ATR_Gender	1	0.000015	0.00020	0.0213
ATR_subgroup:ATR_hemisphere	2	0.000270	0.00371	0.1938
ATR_subgroup:ATR_agef2	4	0.002146	0.02950	0.7704
ATR_hemisphere:ATR_agef2	2	0.000206	0.00283	0.1478
ATR_subgroup:ATR_Gender	2	0.001929	0.02652	1.3850
ATR_hemisphere:ATR_Gender	1	0.000062	0.00086	0.0894
ATR_agef2:ATR_Gender	2	0.003476	0.04777	2.4951
ATR_subgroup:ATR_hemisphere:ATR_agef2	4	0.000411	0.00565	0.1474
ATR_subgroup:ATR_hemisphere:ATR_Gender	2	0.000363	0.00499	0.2606
ATR_subgroup:ATR_agef2:ATR_Gender	2	0.002895	0.03979	2.0781
ATR_hemisphere:ATR_agef2:ATR_Gender	2	0.000126	0.00173	0.0905
ATR_subgroup:ATR_hemisphere:ATR_agef2:ATR_Gender	2	0.000295	0.00406	0.2120
Residual	78	0.054330	0.74669	
Total	109	0.072761	1.00000	

Pr(>F)

	Pr(>F)
ATR_subgroup	0.030 *
ATR_hemisphere	0.752
ATR_agef2	0.347
ATR_Gender	0.895
ATR_subgroup:ATR_hemisphere	0.837
ATR_subgroup:ATR_agef2	0.560
ATR_hemisphere:ATR_agef2	0.861
ATR_subgroup:ATR_Gender	0.232
ATR_hemisphere:ATR_Gender	0.782
ATR_agef2:ATR_Gender	0.072 .
ATR_subgroup:ATR_hemisphere:ATR_agef2	0.963
ATR_subgroup:ATR_hemisphere:ATR_Gender	0.786
ATR_subgroup:ATR_agef2:ATR_Gender	0.129
ATR_hemisphere:ATR_agef2:ATR_Gender	0.930
ATR_subgroup:ATR_hemisphere:ATR_agef2:ATR_Gender	0.806
Residual	
Total	

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

```

permDifIron2<-adonis2(distDifIron~ATR_subgroup+ATR_agef2:ATR_Gender,
                        data=ATR_DifIron_Df, permutations=999, by='terms')
permDifIron2[1:5]

```

	Df	SumOfSqs	R2	F	Pr(>F)
ATR_subgroup	2	0.004721	0.06489	3.7271	0.029 *
ATR_agef2:ATR_Gender	5	0.003435	0.04721	1.0847	0.361
Residual	102	0.064604	0.88790		
Total	109	0.072761	1.00000		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

```

permDifIron3<-adonis2(distDifIron~ATR_subgroup,data=ATR_DifIron_Df,
                         permutations=999, by='terms'); permDifIron3[1:5]
      Df SumOfSqs      R2      F Pr(>F)
ATR_subgroup   2 0.004721 0.06489 3.7124   0.02 *
Residual     107 0.068040 0.93511
Total        109 0.072761 1.00000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## AD-RD-MD
permADMDRD1<-adonis2(distADMDRD~ATR_subgroup*ATR_hemisphere*ATR_agef2*ATR_Gender,
                         data=ATR_DifIron_Df_ADRDMD, permutations=999, by='terms')
permADMDRD1[1:5]
      Df SumOfSqs      R2      F
ATR_subgroup    2 9.2000e-08 0.21944 14.4007
ATR_hemisphere  1 1.8690e-08 0.04459  5.8523
ATR_agef2       2 1.7720e-08 0.04225  2.7729
ATR_Gender       1 2.8600e-09 0.00682  0.8957
ATR_subgroup:ATR_hemisphere  2 1.8500e-09 0.00441  0.2893
ATR_subgroup:ATR_agef2      4 1.4390e-08 0.03432  1.1262
ATR_hemisphere:ATR_agef2   2 1.4400e-09 0.00344  0.2259
ATR_subgroup:ATR_Gender     2 2.5200e-09 0.00600  0.3937
ATR_hemisphere:ATR_Gender  1 6.7000e-10 0.00160  0.2100
ATR_agef2:ATR_Gender       2 1.0190e-08 0.02432  1.5957
ATR_subgroup:ATR_hemisphere:ATR_agef2  4 1.2200e-09 0.00290  0.0953
ATR_subgroup:ATR_hemisphere:ATR_Gender  2 1.7000e-10 0.00041  0.0270
ATR_subgroup:ATR_agef2:ATR_Gender      2 5.0900e-09 0.01215  0.7974
ATR_hemisphere:ATR_agef2:ATR_Gender   2 4.4000e-10 0.00105  0.0687
ATR_subgroup:ATR_hemisphere:ATR_agef2:ATR_Gender  2 8.4000e-10 0.00200  0.1314
Residual          78 2.4916e-07 0.59429
Total             109 4.1926e-07 1.00000
Pr(>F)
0.001 ***
0.015 *
0.054 .
0.353
0.825
0.342
0.879
0.730
0.754
0.176
0.997
0.995
0.452
0.977
0.948
Residual
Total
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
permADMDRD2<-adonis2(distADMDRD~ATR_subgroup+ATR_hemisphere+ATR_agef2,
                         data=ATR_DifIron_Df_ADRDMD, permutations=999, by='terms')
permADMDRD2[1:5]
      Df SumOfSqs      R2      F Pr(>F)
ATR_subgroup    2 9.2000e-08 0.21944 16.4489   0.001 ***

```

```

ATR_hemisphere   1 1.8690e-08 0.04459  6.6847  0.008 ** 
ATR_agef2        2 1.7720e-08 0.04225  3.1673  0.026 *  
Residual         104 2.9085e-07 0.69372
Total            109 4.1926e-07 1.00000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Iron-FA
permIronFA1<-adonis2(distIronFA~ATR_subgroup*ATR_hemisphere*ATR_agef2*ATR_Gender,
                        data=ATR_DifIron_Df_IronFA, permutations=999, by='terms')
permIronFA1[1:5]

```

	Df	SumOfSqs	R2	F
ATR_subgroup	2	0.004721	0.06489	3.3891
ATR_hemisphere	1	0.000076	0.00104	0.1086
ATR_agef2	2	0.001440	0.01979	1.0334
ATR_Gender	1	0.000015	0.00020	0.0213
ATR_subgroup:ATR_hemisphere	2	0.000270	0.00371	0.1938
ATR_subgroup:ATR_agef2	4	0.002146	0.02950	0.7704
ATR_hemisphere:ATR_agef2	2	0.000206	0.00283	0.1478
ATR_subgroup:ATR_Gender	2	0.001929	0.02652	1.3850
ATR_hemisphere:ATR_Gender	1	0.000062	0.00086	0.0894
ATR_agef2:ATR_Gender	2	0.003476	0.04777	2.4951
ATR_subgroup:ATR_hemisphere:ATR_agef2	4	0.000411	0.00565	0.1474
ATR_subgroup:ATR_hemisphere:ATR_Gender	2	0.000363	0.00499	0.2606
ATR_subgroup:ATR_agef2:ATR_Gender	2	0.002895	0.03979	2.0781
ATR_hemisphere:ATR_agef2:ATR_Gender	2	0.000126	0.00173	0.0905
ATR_subgroup:ATR_hemisphere:ATR_agef2:ATR_Gender	2	0.000295	0.00406	0.2120
Residual	78	0.054329	0.74669	
Total	109	0.072760	1.00000	
		Pr(>F)		
ATR_subgroup		0.045 *		
ATR_hemisphere		0.732		
ATR_agef2		0.356		
ATR_Gender		0.893		
ATR_subgroup:ATR_hemisphere		0.833		
ATR_subgroup:ATR_agef2		0.548		
ATR_hemisphere:ATR_agef2		0.878		
ATR_subgroup:ATR_Gender		0.219		
ATR_hemisphere:ATR_Gender		0.759		
ATR_agef2:ATR_Gender		0.091 .		
ATR_subgroup:ATR_hemisphere:ATR_agef2		0.959		
ATR_subgroup:ATR_hemisphere:ATR_Gender		0.790		
ATR_subgroup:ATR_agef2:ATR_Gender		0.123		
ATR_hemisphere:ATR_agef2:ATR_Gender		0.897		
ATR_subgroup:ATR_hemisphere:ATR_agef2:ATR_Gender		0.802		
Residual				
Total				

```

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
permIronFA2<-adonis2(distIronFA~ATR_subgroup+ATR_agef2:ATR_Gender,
                        data=ATR_DifIron_Df_IronFA, permutations=999, by='terms')
permIronFA2[1:5]

```

	Df	SumOfSqs	R2	F	Pr(>F)
ATR_subgroup	2	0.004721	0.06489	3.7270	0.034 *
ATR_agef2:ATR_Gender	5	0.003435	0.04721	1.0847	0.384
Residual	102	0.064604	0.88790		

```

Total           109 0.072760 1.000000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
permIronFA3<-adonis2(distIronFA~ATR_subgroup, data=ATR_DifIron_Df_IronFA,
                       permutations=999, by='terms')
permIronFA3[1:5]
  Df SumOfSqs      R2      F Pr(>F)
ATR_subgroup   2 0.004721 0.06489 3.7123  0.031 *
Residual      107 0.068039 0.93511
Total          109 0.072760 1.000000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

# Establecer seed para que el resultado de las permutaciones sea siempre el mismo
set.seed(123)

# Modelos para el dataframe generado con la eliminación de outliers
## Difusión-Iron
permDifIron1_WoOut<-adonis2(distDifIron_WO~ATR_subgroup*ATR_hemisphere*ATR_agef2*ATR_Gender,
                               data=ATR_DifIron_Df_WoOut, permutations=999, by='terms')
permDifIron1_WoOut[1:5]
  Df SumOfSqs      R2      F
ATR_subgroup   2 0.003621 0.05507 2.7689
ATR_hemisphere 1 0.000059 0.00090 0.0905
ATR_agef2       2 0.000977 0.01485 0.7468
ATR_Gender       1 0.000028 0.00042 0.0425
ATR_subgroup:ATR_hemisphere 2 0.000259 0.00394 0.1979
ATR_subgroup:ATR_agef2 4 0.002948 0.04483 1.1269
ATR_hemisphere:ATR_agef2 2 0.000204 0.00310 0.1558
ATR_subgroup:ATR_Gender 2 0.001131 0.01720 0.8650
ATR_hemisphere:ATR_Gender 1 0.000076 0.00116 0.1162
ATR_agef2:ATR_Gender 2 0.003686 0.05605 2.8180
ATR_subgroup:ATR_hemisphere:ATR_agef2 4 0.000403 0.00613 0.1541
ATR_subgroup:ATR_hemisphere:ATR_Gender 2 0.000417 0.00634 0.3188
ATR_subgroup:ATR_agef2:ATR_Gender 2 0.001896 0.02883 1.4493
ATR_hemisphere:ATR_agef2:ATR_Gender 2 0.000125 0.00191 0.0958
ATR_subgroup:ATR_hemisphere:ATR_agef2:ATR_Gender 2 0.000231 0.00352 0.1768
Residual        76 0.049700 0.75577
Total          107 0.065760 1.000000
Pr(>F)
ATR_subgroup    0.066 .
ATR_hemisphere 0.779
ATR_agef2       0.499
ATR_Gender       0.836
ATR_subgroup:ATR_hemisphere 0.843
ATR_subgroup:ATR_agef2 0.339
ATR_hemisphere:ATR_agef2 0.870
ATR_subgroup:ATR_Gender 0.439
ATR_hemisphere:ATR_Gender 0.725
ATR_agef2:ATR_Gender 0.068 .
ATR_subgroup:ATR_hemisphere:ATR_agef2 0.959
ATR_subgroup:ATR_hemisphere:ATR_Gender 0.730
ATR_subgroup:ATR_agef2:ATR_Gender 0.252
ATR_hemisphere:ATR_agef2:ATR_Gender 0.928
ATR_subgroup:ATR_hemisphere:ATR_agef2:ATR_Gender 0.848
Residual

```

```

Total
-----
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
permDifIron2_WoOut<-adonis2(distDifIron_WO~ATR_subgroup+ATR_agef2:ATR_Gender,
                               data=ATR_DifIron_Df_WoOut, permutations=999, by='terms')
permDifIron2_WoOut[1:5]
      Df SumOfSqs      R2      F Pr(>F)
ATR_subgroup    2 0.003621 0.05507 3.1199  0.039 *
ATR_agef2:ATR_Gender  5 0.004100 0.06235 1.4129  0.221
Residual       100 0.058039 0.88258
Total          107 0.065760 1.00000
-----
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
permDifIron3_WoOut<-adonis2(distDifIron_WO~ATR_subgroup,data=ATR_DifIron_Df_WoOut,
                               permutations=999, by='terms'); permDifIron3_WoOut[1:5]
      Df SumOfSqs      R2      F Pr(>F)
ATR_subgroup    2 0.003621 0.05507 3.0597  0.063 .
Residual       105 0.062139 0.94493
Total          107 0.065760 1.00000
-----
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## AD-RD-MD
permADMDRD1_WoOut<-adonis2(distADMDRD_WO~ATR_subgroup*ATR_hemisphere*ATR_agef2*ATR_Gender,
                               data=ATR_DifIron_Df_WoOut_ADRDMD, permutations=999, by='terms')
permADMDRD1_WoOut[1:5]
      Df SumOfSqs      R2      F
ATR_subgroup    2 8.0050e-08 0.20606 12.6227
ATR_hemisphere 1 1.5720e-08 0.04047 4.9578
ATR_agef2       2 1.3630e-08 0.03509 2.1495
ATR_Gender       1 3.7100e-09 0.00955 1.1699
ATR_subgroup:ATR_hemisphere 2 1.2700e-09 0.00328 0.2008
ATR_subgroup:ATR_agef2  4 1.0970e-08 0.02825 0.8652
ATR_hemisphere:ATR_agef2 2 8.7000e-10 0.00223 0.1365
ATR_subgroup:ATR_Gender  2 1.7300e-09 0.00446 0.2735
ATR_hemisphere:ATR_Gender 1 5.5000e-10 0.00142 0.1740
ATR_agef2:ATR_Gender   2 9.1000e-09 0.02342 1.4347
ATR_subgroup:ATR_hemisphere:ATR_agef2 4 8.4000e-10 0.00215 0.0659
ATR_subgroup:ATR_hemisphere:ATR_Gender 2 4.2000e-10 0.00108 0.0662
ATR_subgroup:ATR_agef2:ATR_Gender  2 4.0300e-09 0.01037 0.6353
ATR_hemisphere:ATR_agef2:ATR_Gender 2 3.8000e-10 0.00097 0.0597
ATR_subgroup:ATR_hemisphere:ATR_agef2:ATR_Gender 2 1.0500e-09 0.00269 0.1648
Residual        77 2.4415e-07 0.62850
Total          108 3.8846e-07 1.00000
      Pr(>F)
ATR_subgroup    0.001 ***
ATR_hemisphere 0.020 *
ATR_agef2       0.099 .
ATR_Gender       0.307
ATR_subgroup:ATR_hemisphere 0.917
ATR_subgroup:ATR_agef2  0.505
ATR_hemisphere:ATR_agef2 0.954
ATR_subgroup:ATR_Gender  0.837
ATR_hemisphere:ATR_Gender 0.812
ATR_agef2:ATR_Gender   0.255
ATR_subgroup:ATR_hemisphere:ATR_agef2 1.000

```

```

ATR_subgroup:ATR_hemisphere:ATR_Gender          0.984
ATR_subgroup:ATR_agef2:ATR_Gender                0.584
ATR_hemisphere:ATR_agef2:ATR_Gender            0.982
ATR_subgroup:ATR_hemisphere:ATR_agef2:ATR_Gender 0.944
Residual
Total
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
permADMDRD2_WoOut<-adonis2(distADMDRD_WO~ATR_subgroup+ATR_hemisphere,
                               data=ATR_DifIron_Df_WoOut_ADRDMD, permutations=999, by='terms')
permADMDRD2_WoOut[1:5]
  Df   SumOfSqs      R2       F Pr(>F)
ATR_subgroup    2 8.0050e-08 0.20606 14.3579  0.001 ***
ATR_hemisphere 1 1.5720e-08 0.04047  5.6394  0.014 *
Residual        105 2.9270e-07 0.75347
Total           108 3.8846e-07 1.00000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Iron-FA
permIronFA1_WoOut<-adonis2(distIronFA_WO~ATR_subgroup*ATR_hemisphere*ATR_agef2*ATR_Gender,
                               data=ATR_DifIron_Df_WoOut_IronFA, permutations=999, by='terms')
permIronFA1_WoOut[1:5]
  Df   SumOfSqs      R2       F
ATR_subgroup     2 0.004040 0.05794 2.8974
ATR_hemisphere   1 0.000085 0.00121 0.1215
ATR_agef2         2 0.001863 0.02671 1.3360
ATR_Gender        1 0.000062 0.00089 0.0892
ATR_subgroup:ATR_hemisphere 2 0.000050 0.00072 0.0362
ATR_subgroup:ATR_agef2   4 0.001763 0.02528 0.6321
ATR_hemisphere:ATR_agef2 2 0.000165 0.00236 0.1182
ATR_subgroup:ATR_Gender  2 0.001138 0.01632 0.8162
ATR_hemisphere:ATR_Gender 1 0.000156 0.00224 0.2241
ATR_agef2:ATR_Gender   2 0.003006 0.04311 2.1560
ATR_subgroup:ATR_hemisphere:ATR_agef2 4 0.000337 0.00483 0.1208
ATR_subgroup:ATR_hemisphere:ATR_Gender 1 0.000252 0.00361 0.3610
ATR_subgroup:ATR_agef2:ATR_Gender   2 0.003488 0.05002 2.5015
ATR_hemisphere:ATR_agef2:ATR_Gender 2 0.000083 0.00118 0.0592
ATR_subgroup:ATR_hemisphere:ATR_agef2:ATR_Gender 2 0.000259 0.00372 0.1859
Residual
Total
  Df   SumOfSqs      R2       F
  76 0.052981 0.75984
  106 0.069726 1.00000
  Pr(>F)
  ATR_subgroup      0.070 .
  ATR_hemisphere   0.739
  ATR_agef2         0.272
  ATR_Gender        0.783
  ATR_subgroup:ATR_hemisphere 0.969
  ATR_subgroup:ATR_agef2   0.662
  ATR_hemisphere:ATR_agef2 0.907
  ATR_subgroup:ATR_Gender  0.452
  ATR_hemisphere:ATR_Gender 0.646
  ATR_agef2:ATR_Gender   0.132
  ATR_subgroup:ATR_hemisphere:ATR_agef2 0.982
  ATR_subgroup:ATR_hemisphere:ATR_Gender 0.562
  ATR_subgroup:ATR_agef2:ATR_Gender   0.082 .
  ATR_hemisphere:ATR_agef2:ATR_Gender 0.951

```

```

ATR_subgroup:ATR_hemisphere:ATR_agef2:ATR_Gender  0.831
Residual
Total
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
permIronFA2_WoOut<-adonis2(distIronFA_WO~ATR_subgroup+ATR_subgroup:ATR_agef2:ATR_Gender,
                             data=ATR_DifIron_Df_WoOut_IronFA, permutations=999, by='terms')
permIronFA2_WoOut[1:5]
      Df SumOfSqs      R2      F Pr(>F)
ATR_subgroup    2 0.004040 0.05794 3.3870  0.035 *
ATR_subgroup:ATR_agef2:ATR_Gender  13 0.011420 0.16378 1.4731  0.141
Residual        91 0.054267 0.77828
Total          106 0.069726 1.00000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
permIronFA3_WoOut<-adonis2(distIronFA_WO~ATR_subgroup, data=ATR_DifIron_Df_WoOut_IronFA,
                             permutations=999, by='terms')
permIronFA3_WoOut[1:5]
      Df SumOfSqs      R2      F Pr(>F)
ATR_subgroup    2 0.004040 0.05794 3.1979  0.033 *
Residual       104 0.065687 0.94206
Total          106 0.069726 1.00000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

A diferencia de los PERMANOVA de los ATR derecho e izquierdo por separado que no evidenciaban factores significativos, en el **modelo PERMANOVA generado con las 5 métricas** con el dataframe completo, se detecta subgrupo como el único término significativo. Remarcar que PERMANOVA es menos sensible que MANOVA para detectar significación para el resto de factores (MANOVA detectó significación estadística para los factores subgrupo, hemisferio y edad factorizada y para las interacciones subgrupo-edad factorizada y edad factorizada-género).

En el **modelo PERMANOVA de las 3 métricas de difusividad con el dataframe completo**, los términos significativos son el subgrupo, el hemisferio y la edad factorizada, y, en el **modelo PERMANOVA de las 3 métricas de difusividad con el dataframe reducido eliminando los outliers**, los términos significativos son el subgrupo y el hemisferio. Remarcar que PERMANOVA es más sensible que MANOVA para detectar significación para el resto de factores (MANOVA detectó significación estadística exclusivamente para el factor subgrupo).

A diferencia de los PERMANOVA de los ATR derecho e izquierdo por separado que no evidenciaban factores significativos, en los **modelos PERMANOVA generados con las métricas iron-FA**, ya sea con el dataframe completo o con el dataframe reducido eliminando los outliers, el único término significativo es el subgrupo. Remarcar que PERMANOVA ha conseguido detectar significación para subgrupo a diferencia del MANOVA que sólo detectó significación para el factor hemisferio y la interacción edad factorizada-género.

**En ninguno de los modelos, se objetiva una interacción subgrupo-hemisferio significativa.**

Recordar que los resultados son idénticos tanto si se utilizada la matriz de distancias calculada en el paso anterior, como si se utiliza la fórmula con la variable dependiente de interés ~ variables/factores independientes y se especifica el término `method` del tipo de distancia a calcular. De todas formas, el tener las distancias calculadas es de relevancia para poder calcular posteriormente las dispersiones.

### 5.2.3 Valoración de la dispersión

Se calculan las **dispersiones** para los factores significativos de cada uno de los modelos que se han generado en el paso anterior.

```

# Dispersiones para el dataframe completo
dispersionDifIron<-betadisper(d=distDifIron, group=ATR_DifIron_Df$ATR_subgroup,
```

```

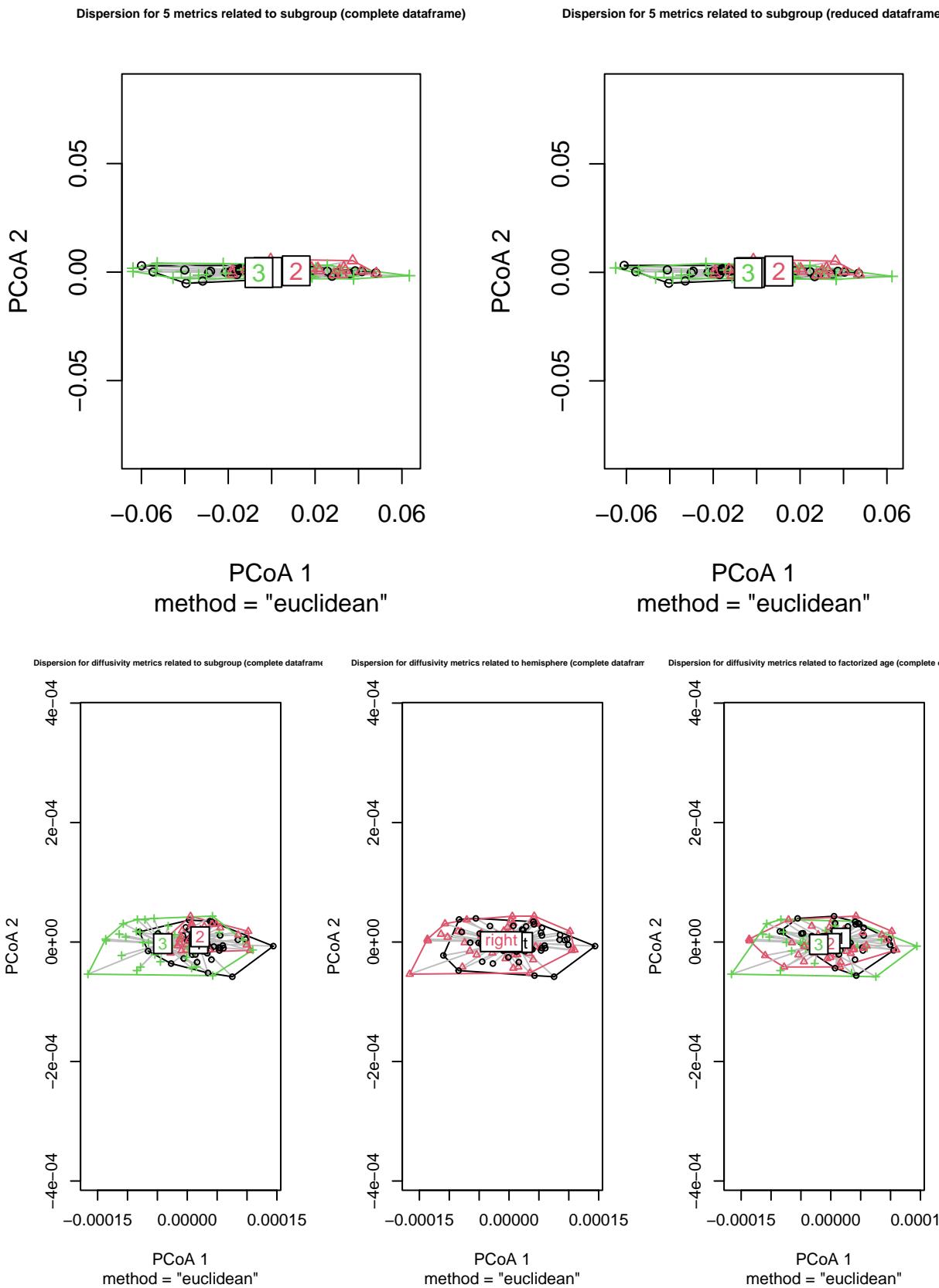
type="centroid",bias.adjust=TRUE)
dispersionADMDRD1<-betadisper(d=distADMDRD, group=ATR_DifIron_Df_ADRDMD$ATR_subgroup,
                                type="centroid",bias.adjust=TRUE)
dispersionADMDRD2<-betadisper(d=distADMDRD, group=ATR_DifIron_Df_ADRDMD$ATR_hemisphere,
                                type="centroid",bias.adjust=TRUE)
dispersionADMDRD3<-betadisper(d=distADMDRD, group=ATR_DifIron_Df_ADRDMD$ATR_agef2,
                                type="centroid",bias.adjust=TRUE)
dispersionIronFA<-betadisper(d=distIronFA, group=ATR_DifIron_Df_IronFA$ATR_subgroup,
                               type="centroid",bias.adjust=TRUE)

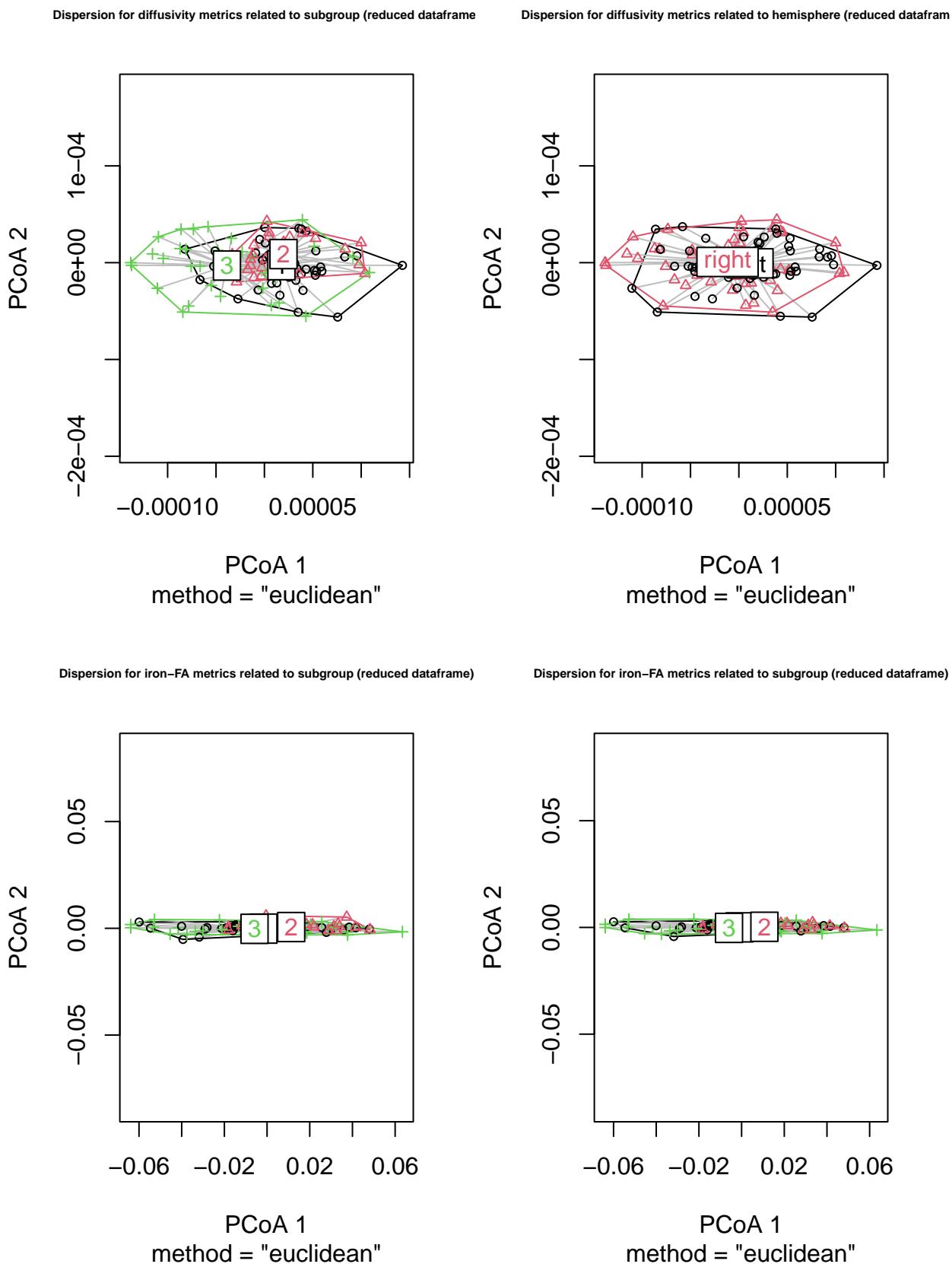
# Dispersiones para el dataframe reducido
dispersionDifIron_W0<-betadisper(d=distDifIron_W0,
                                   group=ATR_DifIron_Df_WoOut$ATR_subgroup,
                                   type="centroid",bias.adjust=TRUE)
dispersionADMDRD_W01<-betadisper(d=distADMDRD_W0,
                                   group=ATR_DifIron_Df_WoOut_ADRDMD$ATR_subgroup,
                                   type="centroid",bias.adjust=TRUE)
dispersionADMDRD_W02<-betadisper(d=distADMDRD_W0,
                                   group=ATR_DifIron_Df_WoOut_ADRDMD$ATR_hemisphere,
                                   type="centroid",bias.adjust=TRUE)
dispersionIronFA_W0<-betadisper(d=distIronFA_W0,
                                   group=ATR_DifIron_Df_WoOut_IronFA$ATR_subgroup,
                                   type="centroid",bias.adjust=TRUE)

```

No se ha requerido utilizar el argumento `add` que sirve para añadir una constante a las disimilaridades no diagonales con el fin de que todos los eigenvalues sean no-negativos en el análisis de coordenadas principales (PCoA), ya que si existe algún eigenvalue negativo R emite un mensaje de aviso `Warning message: some squared distances are negative and changed to zero.`

Se generan **plots** para cada una de las dispersiones.





Cabe mencionar que el subgrupo HD presenta una mayor dispersión respecto a los subgrupos preHD y control, y el subgrupo preHD presenta una menor dispersión respecto a los subgrupos HD y control, y este comportamiento

se da en las 5 métricas, en las 3 métricas de difusividad y en las métricas iron-FA.

En las 3 métricas de difusividad, también es importante remarcar que ambos ATR presentan una dispersión similar, y, en cuanto a la edad factorizada, el nivel 1 (menores de 45) tiene menor dispersión que los otros dos niveles y el nivel 3 (mayores de 55) presenta mayor dispersión que los otros dos niveles.

Se generan **modelos anova** para cada una de las dispersiones.

```
# Modelos ANOVA para las 5 métricas
summary(aov(dispersionDifIron$distances~ATR_DifIron_Df$ATR_subgroup))
      Df   Sum Sq   Mean Sq F value Pr(>F)
ATR_DifIron_Df$ATR_subgroup  2 0.000795 0.0003975   1.709  0.186
Residuals                      107 0.024897 0.0002327

summary(aov(dispersionDifIron_W0$distances~ATR_DifIron_Df_WoOut$ATR_subgroup))
      Df   Sum Sq   Mean Sq F value Pr(>F)
ATR_DifIron_Df_WoOut$ATR_subgroup  2 0.000353 0.0001763   0.808  0.448
Residuals                      105 0.022902 0.0002181

# Modelos ANOVA para las 3 métricas - Dataframe completo
summary(aov(dispersionADMDRD1$distances~ATR_DifIron_Df_ADRDMD$ATR_subgroup))
      Df   Sum Sq   Mean Sq F value Pr(>F)
ATR_DifIron_Df_ADRDMD$ATR_subgroup  2 1.004e-08 5.021e-09   5.949 0.00355 **
Residuals                      107 9.032e-08 8.440e-10
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
summary(aov(dispersionADMDRD2$distances~ATR_DifIron_Df_ADRDMD$ATR_hemisphere))
      Df   Sum Sq   Mean Sq F value Pr(>F)
ATR_DifIron_Df_ADRDMD$ATR_hemisphere  1 6.200e-10 6.226e-10   0.564  0.454
Residuals                      108 1.192e-07 1.104e-09
summary(aov(dispersionADMDRD3$distances~ATR_DifIron_Df_ADRDMD$ATR_agef2))
      Df   Sum Sq   Mean Sq F value Pr(>F)
ATR_DifIron_Df_ADRDMD$ATR_agef2  2 5.950e-09 2.976e-09   2.662 0.0744 .
Residuals                      107 1.196e-07 1.118e-09
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Modelos ANOVA para las 3 métricas - Dataframe reducido
summary(aov(dispersionADMDRD_W01$distances~ATR_DifIron_Df_WoOut_ADRDMD$ATR_subgroup))
      Df   Sum Sq   Mean Sq F value
ATR_DifIron_Df_WoOut_ADRDMD$ATR_subgroup  2 8.100e-09 4.049e-09   5.197
Residuals                      106 8.258e-08 7.790e-10
Pr(>F)
ATR_DifIron_Df_WoOut_ADRDMD$ATR_subgroup 0.00703 **
Residuals
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
summary(aov(dispersionADMDRD_W02$distances~ATR_DifIron_Df_WoOut_ADRDMD$ATR_hemisphere))
      Df   Sum Sq   Mean Sq F value
ATR_DifIron_Df_WoOut_ADRDMD$ATR_hemisphere  1 1.700e-10 1.681e-10   0.167
Residuals                      107 1.076e-07 1.006e-09
Pr(>F)
ATR_DifIron_Df_WoOut_ADRDMD$ATR_hemisphere 0.683
Residuals

# Modelos ANOVA para las métricas Iron-FA
summary(aov(dispersionIronFA$distances~ATR_DifIron_Df_IronFA$ATR_subgroup))
      Df   Sum Sq   Mean Sq F value Pr(>F)
ATR_DifIron_Df_IronFA$ATR_subgroup  2 0.000795 0.0003975   1.708  0.186
```

```

Residuals 107 0.024898 0.0002327
summary(aov(dispersionIronFA_W0$distances~ATR_DifIron_Df_WoOut_IronFA$ATR_subgroup))
          Df Sum Sq Mean Sq F value Pr(>F)
ATR_DifIron_Df_WoOut_IronFA$ATR_subgroup 2 0.000879 0.0004396 1.873 0.159
Residuals 104 0.024408 0.0002347

```

Las dispersiones para subgrupo de las 3 métricas de difusividad presentan significación estadística, con lo que no se puede determinar si las diferencias observadas en PERMANOVA son debidas a diferencias en el posicionamiento de los centroides y/o a las diferencias en dispersión. Como se mencionó más arriba, no se puede determinar cuánta diferencia en el PERMANOVA es debida al posicionamiento de centroides y cuánta debida a la variabilidad en dispersión. El resto de dispersiones no son significativas.

Se decide intentar la **transformación logarítmica de las 3 métricas de difusividad para determinar si se mejora la dispersión**. A continuación, se exponen los resultados para el **dataframe completo**.

```

# Se genera la transformación logarítmica
ATR_DifIron_Df_ADRDMD_log<-cbind(ATR_DifIron_Df_ADRDMD[,1:6],
                                      log(ATR_DifIron_Df_ADRDMD[,7:9]))

# Se calculan las distancias
distADMDRD_log<-vegdist(ATR_DifIron_Df_ADRDMD_log[,c('ATR_AD','ATR_MD','ATR_RD')],
                           method='euclidean')

# Se generan los modelos PERMANOVA
set.seed(123)
permADMDRD_log1<-adonis2(distADMDRD_log~ATR_subgroup*ATR_hemisphere*ATR_agef2*ATR_Gender,
                           data=ATR_DifIron_Df_ADRDMD_log, permutations=999, by='terms')
permADMDRD_log1[1:5]

Df SumOfSqs R2 F
ATR_subgroup 2 0.20521 0.22209 14.2941
ATR_hemisphere 1 0.03795 0.04108 5.2875
ATR_agef2 2 0.04157 0.04499 2.8957
ATR_Gender 1 0.00614 0.00665 0.8554
ATR_subgroup:ATR_hemisphere 2 0.00275 0.00298 0.1915
ATR_subgroup:ATR_agef2 4 0.02737 0.02962 0.9531
ATR_hemisphere:ATR_agef2 2 0.00283 0.00306 0.1971
ATR_subgroup:ATR_Gender 2 0.00726 0.00786 0.5060
ATR_hemisphere:ATR_Gender 1 0.00102 0.00110 0.1416
ATR_agef2:ATR_Gender 2 0.01650 0.01785 1.1491
ATR_subgroup:ATR_hemisphere:ATR_agef2 4 0.00214 0.00232 0.0746
ATR_subgroup:ATR_hemisphere:ATR_Gender 2 0.00053 0.00058 0.0371
ATR_subgroup:ATR_agef2:ATR_Gender 2 0.01050 0.01136 0.7313
ATR_hemisphere:ATR_agef2:ATR_Gender 2 0.00073 0.00079 0.0510
ATR_subgroup:ATR_hemisphere:ATR_agef2:ATR_Gender 2 0.00159 0.00172 0.1109
Residual 78 0.55990 0.60595
Total 109 0.92400 1.00000
Pr(>F)
ATR_subgroup 0.001 ***
ATR_hemisphere 0.017 *
ATR_agef2 0.046 *
ATR_Gender 0.376
ATR_subgroup:ATR_hemisphere 0.890
ATR_subgroup:ATR_agef2 0.453
ATR_hemisphere:ATR_agef2 0.900
ATR_subgroup:ATR_Gender 0.655
ATR_hemisphere:ATR_Gender 0.828
ATR_agef2:ATR_Gender 0.315

```

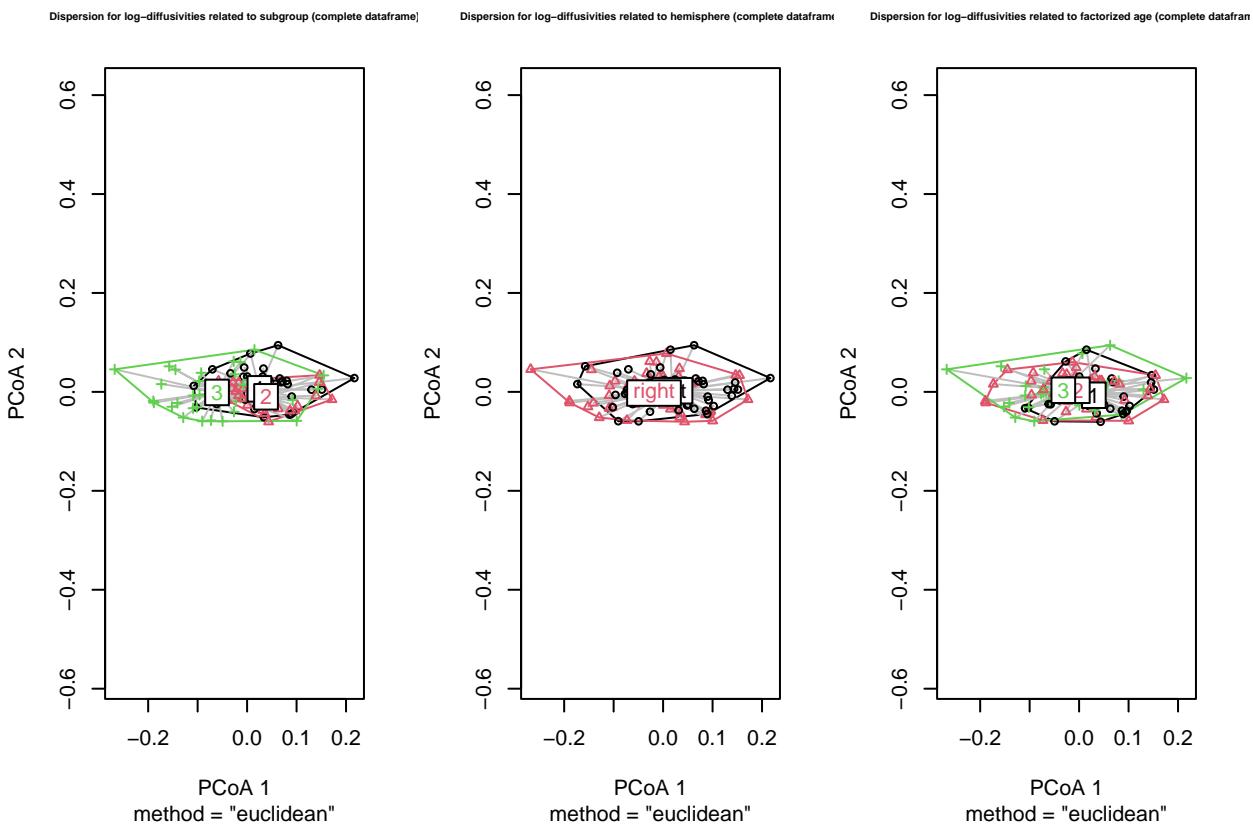
```

ATR_subgroup:ATR_hemisphere:ATR_agef2           1.000
ATR_subgroup:ATR_hemisphere:ATR_Gender          0.996
ATR_subgroup:ATR_agef2:ATR_Gender                0.497
ATR_hemisphere:ATR_agef2:ATR_Gender             0.992
ATR_subgroup:ATR_hemisphere:ATR_agef2:ATR_Gender 0.957
Residual
Total
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
permADMDRD_log2<-adonis2(distADMDRD_log~ATR_subgroup+ATR_hemisphere+ATR_agef2,
                           data=ATR_DifIron_Df_ADRDMD_log, permutations=999, by='terms')
permADMDRD_log2[1:5]
  Df SumOfSqs      R2      F Pr(>F)
ATR_subgroup    2  0.20521 0.22209 16.6928  0.001 ***
ATR_hemisphere  1  0.03795 0.04108  6.1748  0.007 **
ATR_agef2       2  0.04157 0.04499  3.3817  0.026 *
Residual        104 0.63926 0.69184
Total           109 0.92400 1.00000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Se determinan las dispersiones
dispersionADMDRD_log1<-betadisper(d=distADMDRD_log, group=ATR_DifIron_Df_ADRDMD_log$ATR_subgroup,
                                      type="centroid", bias.adjust=TRUE)
dispersionADMDRD_log2<-betadisper(d=distADMDRD_log, group=ATR_DifIron_Df_ADRDMD_log$ATR_hemisphere,
                                      type="centroid", bias.adjust=TRUE)
dispersionADMDRD_log3<-betadisper(d=distADMDRD_log, group=ATR_DifIron_Df_ADRDMD_log$ATR_agef2,
                                      type="centroid", bias.adjust=TRUE)

# Se generan los plots
par(mfrow = c(1, 3))
plot(dispersionADMDRD_log1, main=list('Dispersion for log-diffusivities related to subgroup (complete data)')
plot(dispersionADMDRD_log2, main=list('Dispersion for log-diffusivities related to hemisphere (complete data)')
plot(dispersionADMDRD_log3, main=list('Dispersion for log-diffusivities related to factorized age (complete data)'))

```



```
# Modelos ANOVA para las 3 métricas - Dataframe completo con transformación logarítmica
summary(aov(dispersionADMDRD_log1$distances~ATR_DifIron_Df_ADRDMD_log$ATR_subgroup))
Df Sum Sq Mean Sq F value Pr(>F)
ATR_DifIron_Df_ADRDMD_log$ATR_subgroup 2 0.01459 0.007296 3.749 0.0267 *
Residuals 107 0.20822 0.001946
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
summary(aov(dispersionADMDRD_log2$distances~ATR_DifIron_Df_ADRDMD_log$ATR_hemisphere))
Df Sum Sq Mean Sq F value Pr(>F)
ATR_DifIron_Df_ADRDMD_log$ATR_hemisphere 1 0.00073 0.0007346 0.3 0.585
Residuals 108 0.26408 0.0024452
summary(aov(dispersionADMDRD_log3$distances~ATR_DifIron_Df_ADRDMD_log$ATR_ägef2))
Df Sum Sq Mean Sq F value Pr(>F)
ATR_DifIron_Df_ADRDMD_log$ATR_ägef2 2 0.01036 0.005182 2.16 0.12
Residuals 107 0.25669 0.002399
```

En el caso de la **transformación logarítmica de las 3 métricas de difusividad para el dataframe completo**, se objetiva que los factores subgrupo, hemisferio y edad factorizada son los términos significativos en el PERMANOVA (como ocurría sin la transformación logarítmica). Persiste una mayor dispersión para el subgrupo HD y una menor dispersión para el subgrupo preHD visualizada en el plot y significativa en el anova. La transformación logarítmica disminuye el valor de F y aumenta el valor de p, pero no lo suficiente; así, los valores F y p-valor para el PERMANOVA con transformación logarítmica son de 3.749 y de 0.027 respectivamente comparado con los valores F y p-valor para el PERMANOVA sin transformación logarítmica de 5.949 y 0.004 respectivamente.

A continuación, se exponen los resultados para el **dataframe reducido** (con eliminación de outliers).

```
# Se genera la transformación logarítmica
ATR_DifIron_Df_ADRDMD_WO_log<-cbind(ATR_DifIron_Df_WoOut_ADRDMD[,1:6],
```

```

log(ATR_DifIron_Df_WoOut_ADRDMD[,7:9]))
```

*# Se calculan las distancias*

```
distADMDRD_WO_log<-vegdist(ATR_DifIron_Df_ADRDMD_WO_log[,c('ATR_AD','ATR_MD','ATR_RD')],  
                             method='euclidean')
```

*# Se generan los modelos PERMANOVA*

```
set.seed(123)  
permADMDRD_WO_log1<-adonis2(distADMDRD_WO_log~ATR_subgroup*ATR_hemisphere*ATR_agef2*ATR_Gender,  
                               data=ATR_DifIron_Df_ADRDMD_WO_log, permutations=999, by='terms')  
permADMDRD_WO_log1[1:5]
```

	Df	SumOfSqs	R2	F			
ATR_subgroup	2	0.17707	0.20842	12.4707			
ATR_hemisphere	1	0.03106	0.03657	4.3758			
ATR_agef2	2	0.03171	0.03732	2.2330			
ATR_Gender	1	0.00836	0.00984	1.1772			
ATR_subgroup:ATR_hemisphere	2	0.00178	0.00209	0.1251			
ATR_subgroup:ATR_agef2	4	0.01952	0.02297	0.6873			
ATR_hemisphere:ATR_agef2	2	0.00139	0.00164	0.0981			
ATR_subgroup:ATR_Gender	2	0.00500	0.00589	0.3522			
ATR_hemisphere:ATR_Gender	1	0.00075	0.00089	0.1062			
ATR_agef2:ATR_Gender	2	0.01362	0.01603	0.9589			
ATR_subgroup:ATR_hemisphere:ATR_agef2	4	0.00127	0.00150	0.0448			
ATR_subgroup:ATR_hemisphere:ATR_Gender	2	0.00138	0.00163	0.0974			
ATR_subgroup:ATR_agef2:ATR_Gender	2	0.00725	0.00853	0.5104			
ATR_hemisphere:ATR_agef2:ATR_Gender	2	0.00055	0.00065	0.0389			
ATR_subgroup:ATR_hemisphere:ATR_agef2:ATR_Gender	2	0.00222	0.00261	0.1563			
Residual	77	0.54665	0.64344				
Total	108	0.84957	1.00000				
			Pr(>F)				
ATR_subgroup			0.001 ***				
ATR_hemisphere			0.041 *				
ATR_agef2			0.093 .				
ATR_Gender			0.260				
ATR_subgroup:ATR_hemisphere			0.954				
ATR_subgroup:ATR_agef2			0.636				
ATR_hemisphere:ATR_agef2			0.964				
ATR_subgroup:ATR_Gender			0.792				
ATR_hemisphere:ATR_Gender			0.849				
ATR_agef2:ATR_Gender			0.408				
ATR_subgroup:ATR_hemisphere:ATR_agef2			1.000				
ATR_subgroup:ATR_hemisphere:ATR_Gender			0.964				
ATR_subgroup:ATR_agef2:ATR_Gender			0.677				
ATR_hemisphere:ATR_agef2:ATR_Gender			0.996				
ATR_subgroup:ATR_hemisphere:ATR_agef2:ATR_Gender			0.935				
Residual							
Total							
---							
Signif. codes:	0	'***'	0.001 '**'	0.01 '*'	0.05 '.'	0.1 ' '	1
permADMDRD_WO_log2<-adonis2(distADMDRD_WO_log~ATR_subgroup+ATR_hemisphere+ATR_agef2, data=ATR_DifIron_Df_ADRDMD_WO_log, permutations=999, by='terms')							
permADMDRD_WO_log2[1:5]	Df	SumOfSqs	R2	F	Pr(>F)		
ATR_subgroup	2	0.17707	0.20842	14.9556	0.001 ***		
ATR_hemisphere	1	0.03106	0.03657	5.2477	0.017 *		
ATR_agef2	2	0.03171	0.03732	2.6780	0.043 *		

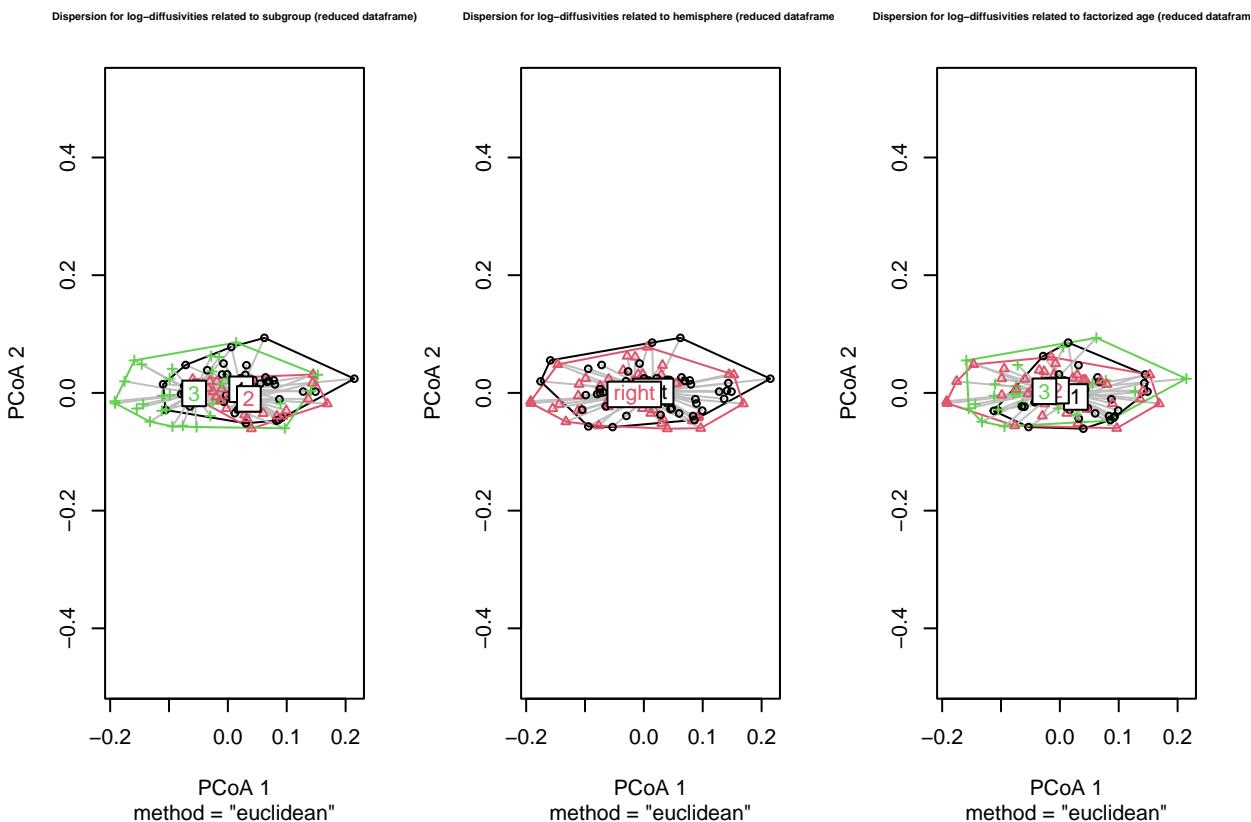
```

Residual      103  0.60973  0.71770
Total         108  0.84957  1.00000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Se determinan las dispersiones
dispersionADMDRD_WO_log1<-betadisper(d=distADMDRD_WO_log,
                                         group=ATR_DifIron_Df_ADRDMD_WO_log$ATR_subgroup,
                                         type="centroid",bias.adjust=TRUE)
dispersionADMDRD_WO_log2<-betadisper(d=distADMDRD_WO_log,
                                         group=ATR_DifIron_Df_ADRDMD_WO_log$ATR_hemisphere,
                                         type="centroid",bias.adjust=TRUE)
dispersionADMDRD_WO_log3<-betadisper(d=distADMDRD_WO_log,
                                         group=ATR_DifIron_Df_ADRDMD_WO_log$ATR_ägef2,
                                         type="centroid",bias.adjust=TRUE)

# Se generan los plots
par(mfrow = c(1, 3))
plot(dispersionADMDRD_WO_log1, main=list('Dispersion for log-diffusivities related to subgroup (reduced data frame)', 'ATR_subgroup'))
plot(dispersionADMDRD_WO_log2, main=list('Dispersion for log-diffusivities related to hemisphere (reduced data frame)', 'ATR_hemisphere'))
plot(dispersionADMDRD_WO_log3, main=list('Dispersion for log-diffusivities related to factorized age (reduced data frame)', 'ATR_ägef2'))

```



```

# Modelos ANOVA para las 3 métricas - Dataframe completo con transformación logarítmica
summary(aov(dispersionADMDRD_WO_log1$distances~ATR_DifIron_Df_ADRDMD_WO_log$ATR_subgroup))
Df  Sum Sq  Mean Sq F value Pr(>F)
ATR_DifIron_Df_ADRDMD_WO_log$ATR_subgroup  2  0.01076  0.005380   3.047  0.0517 .
Residuals                               106  0.18716  0.001766

```

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
summary(aov(dispersionADMDRD_W0_log2$distances~ATR_DifIron_Df_ADRDMD_W0_log$ATR_hemisphere))
              Df  Sum Sq  Mean Sq F value Pr(>F)
ATR_DifIron_Df_ADRDMD_W0_log$ATR_hemisphere  1 0.00008 8.11e-05   0.037  0.847
Residuals                               107 0.23220 2.17e-03
summary(aov(dispersionADMDRD_W0_log3$distances~ATR_DifIron_Df_ADRDMD_W0_log$ATR_ägef2))
              Df  Sum Sq  Mean Sq F value Pr(>F)
ATR_DifIron_Df_ADRDMD_W0_log$ATR_ägef2    2 0.00662 0.003309   1.537  0.22
Residuals                               106 0.22830 0.002154

```

En el caso de la **transformación logarítmica de las 3 métricas de difusividad para el dataframe reducido**, se objetiva que los factores subgrupo, hemisferio y edad factorizada son los términos significativos en el PERMANOVA (a diferencia del PERMANOVA obtenido sin transformación logarítmica, en el que sólo subgrupo y hemisferio son significativos). Persiste una mayor dispersión para el subgrupo HD y una menor dispersión para el subgrupo preHD visualizada en el plot, pero sin significación estadística en el anova. La transformación logarítmica disminuye el valor de F y aumenta el valor de p lo suficiente como para perder la significación estadística; así, los valores F y p-valor para el PERMANOVA con transformación logarítmica son de 3.047 y de 0.052 respectivamente comparado con los valores F y p-valor para el PERMANOVA sin transformación logarítmica de 5.197 y 0.007 respectivamente.

En resumen, la interacción subgrupo-hemisferio no es significativa en ninguna de las combinaciones métricas, pero los PERMANOVA obtenidos con las métricas de ambos ATR han permitido aumentar la sensibilidad para detectar diferencias entre subgrupos, especialmente en la evaluación de las 5 métricas conjuntas y de las métricas Iron-FA.

#### 5.2.4 Comparaciones por pares

Se procede a generar las comparaciones por pares de los PERMANOVA que han obtenido significación estadística para subgrupo y cuyas dispersiones no hayan resultados estadísticamente significativas. Por tanto, se van a realizar las comparaciones de los siguientes PERMANOVA: permDifIron3 (PERMANOVA de las 5 métricas con el dataframe completo), permADMDRD\_log2 y permADMDRD\_W0\_log2 (PERMANOVA de la transformación logarítmica de las 3 métricas de difusividad con los dataframes completo y reducido, respectivamente), y, permIronFA3 y permIronFA3\_WoOut (PERMANOVA de métricas iron-FA con el dataframe completo y reducido, respectivamente).

Para la comparación por pares, se utiliza la función pairwise.adonis2 del paquete pairwiseAdonis. Esta función admite los mismos argumentos que la función adonis2. No se realiza la comparación por pares de los factores hemisferio ni edad factorizada para el PERMANOVA de las 3 métricas de difusividad dado que no es de interés para el estudio.

```

# Se carga la librería
library(devtools)
install_github("pmartinezarbizu/pairwiseAdonis/pairwiseAdonis", force=TRUE)
vegan (2.6-4 -> 2.6-6.1) [CRAN]
-- R CMD build
  checking for file 'C:\Users\monts\AppData\Local\Temp\Rtmp2b1GyB\remotes8d2843791a1c\pmartinezarbi...
  - preparing 'pairwiseAdonis':
checking DESCRIPTION meta-information ... v  checking DESCRIPTION meta-information
  - checking for LF line-endings in source and make files and shell scripts
  - checking for empty or unneeded directories
  Omitted 'LazyData' from DESCRIPTION
  - building 'pairwiseAdonis_0.4.1.tar.gz'

library(pairwiseAdonis)

```

```

# Establecer seed para que el resultado de las permutaciones sea siempre el mismo
set.seed(123)

# Test a posteriori para las 5 métricas
pairwise_permanova_DifIron<-pairwiseAdonis::pairwise.adonis2(distIronFA~ATR_subgroup,
                                                               data=ATR_DifIron_Df,
                                                               permutations=999, by='terms')
pairwise_permanova_DifIron_W0<-pairwiseAdonis::pairwise.adonis2(distDifIron_W0~ATR_subgroup,
                                                               data=ATR_DifIron_Df_WoOut,
                                                               permutations=999, by='terms')

# Test a posteriori para las 3 métricas de difusividad
pairwise_permanova_ADMRD_log<-pairwiseAdonis::pairwise.adonis2(distADMRD_log~
                                                               ATR_subgroup,
                                                               data=ATR_DifIron_Df_ADRDMD_log,
                                                               permutations=999, by='terms')

pairwise_permanova_ADMRD_W0_log<-pairwiseAdonis::pairwise.adonis2(distADMRD_W0_log~
                                                               ATR_subgroup,
                                                               data=ATR_DifIron_Df_ADRDMD_W0_log,
                                                               permutations=999, by='terms')

# Test a posteriori para las métricas de Iron y FA
pairwise_permanova_IronFA<-pairwiseAdonis::pairwise.adonis2(distIronFA~ATR_subgroup,
                                                               data=ATR_DifIron_Df_IronFA,
                                                               permutations=999, by='terms')
pairwise_permanova_IronFA_W0<-pairwiseAdonis::pairwise.adonis2(distIronFA_W0~ATR_subgroup,
                                                               data=ATR_DifIron_Df_WoOut_IronFA,
                                                               permutations=999, by='terms')

pairwise_permanova_DifIron
$parent_call
[1] "distIronFA ~ ATR_subgroup , strata = Null , permutations 999"

$`1_vs_3`
  Df SumOfSqs      R2      F Pr(>F)
ATR_subgroup  1 0.000354 0.00606 0.5  0.489
Residual     82 0.058126 0.99394
Total        83 0.058480 1.00000

$`1_vs_2`
  Df SumOfSqs      R2      F Pr(>F)
ATR_subgroup  1 0.002873 0.0744 5.7877  0.021 *
Residual     72 0.035741 0.9256
Total        73 0.038614 1.00000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

$`3_vs_2`
  Df SumOfSqs      R2      F Pr(>F)
ATR_subgroup  1 0.004467 0.0957 6.3495  0.016 *
Residual     60 0.042211 0.9043
Total        61 0.046678 1.00000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

attr("class")
[1] "pwadstrata" "list"
pairwise_permanova_DifIron_W0
$parent_call
[1] "distDifIron_W0 ~ ATR_subgroup , strata = Null , permutations 999"

$`1_vs_3`
      Df SumOfSqs      R2      F Pr(>F)
ATR_subgroup 1 0.000024 0.00046 0.0367 0.872
Residual     80 0.052226 0.99954
Total        81 0.052250 1.00000

$`1_vs_2`
      Df SumOfSqs      R2      F Pr(>F)
ATR_subgroup 1 0.002873 0.0744 5.7877 0.022 *
Residual     72 0.035741 0.9256
Total        73 0.038615 1.00000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

$`3_vs_2`
      Df SumOfSqs      R2      F Pr(>F)
ATR_subgroup 1 0.002942 0.07496 4.7 0.035 *
Residual     58 0.036311 0.92504
Total        59 0.039253 1.00000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

attr("class")
[1] "pwadstrata" "list"

```

Las comparaciones por pares ponen de relieve que **las diferencias significativas entre subgrupos en cuanto a las 5 métricas se hallan entre preHD vs controles y entre preHD vs HD para ambos dataframes (completo y reducido, sin y con eliminación de outliers, respectivamente)**.

```

pairwise_permanova_ADMRD_log
$parent_call
[1] "distADMRD_log ~ ATR_subgroup , strata = Null , permutations 999"

$`1_vs_3`
      Df SumOfSqs      R2      F Pr(>F)
ATR_subgroup 1 0.15196 0.20192 20.747 0.001 ***
Residual     82 0.60058 0.79808
Total        83 0.75254 1.00000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

$`1_vs_2`
      Df SumOfSqs      R2      F Pr(>F)
ATR_subgroup 1 0.00741 0.02006 1.4742 0.24
Residual     72 0.36188 0.97994
Total        73 0.36929 1.00000

$`3_vs_2`
      Df SumOfSqs      R2      F Pr(>F)
ATR_subgroup 1 0.14937 0.23919 18.863 0.001 ***

```

```

Residual      60  0.47511  0.76081
Total         61  0.62448  1.00000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

attr("class")
[1] "pwadstrata" "list"
pairwise_permanova_ADMRD_W0_log
$parent_call
[1] "distADMRD_W0_log ~ ATR_subgroup , strata = Null , permutations 999"

$`1_vs_3`
      Df SumOfSqs      R2      F Pr(>F)
ATR_subgroup  1  0.13015 0.19015 19.019  0.001 ***
Residual     81  0.55430 0.80985
Total        82  0.68446 1.00000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

$`1_vs_2`
      Df SumOfSqs      R2      F Pr(>F)
ATR_subgroup  1  0.00741 0.02006 1.4742   0.21
Residual     72  0.36188 0.97994
Total        73  0.36929 1.00000

$`3_vs_2`
      Df SumOfSqs      R2      F Pr(>F)
ATR_subgroup  1  0.13032 0.23307 17.93   0.001 ***
Residual     59  0.42883 0.76693
Total        60  0.55915 1.00000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

attr("class")
[1] "pwadstrata" "list"

```

Las comparaciones por pares ponen de relieve que las diferencias significativas entre subgrupos en cuanto a las 3 métricas de difusividad se hallan entre HD vs controles y entre HD vs preHD para el dataframes reducido (con eliminación de outliers) con la transformación logarítmica de las 3 métricas de difusividad.

```

pairwise_permanova_IronFA
$parent_call
[1] "distIronFA ~ ATR_subgroup , strata = Null , permutations 999"

$`1_vs_3`
      Df SumOfSqs      R2      F Pr(>F)
ATR_subgroup  1  0.000354 0.00606 0.5  0.484
Residual     82  0.058126 0.99394
Total        83  0.058480 1.00000

$`1_vs_2`
      Df SumOfSqs      R2      F Pr(>F)
ATR_subgroup  1  0.002873 0.0744 5.7877  0.016 *
Residual     72  0.035741 0.9256
Total        73  0.038614 1.0000
---
```

```

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

$`3_vs_2`
      Df SumOfSqs      R2      F Pr(>F)
ATR_subgroup 1 0.004467 0.0957 6.3495  0.018 *
Residual     60 0.042211 0.9043
Total        61 0.046678 1.0000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

attr("class")
[1] "pwadstrata" "list"
pairwise_permanova_IronFA_W0
$parent_call
[1] "distIronFA_W0 ~ ATR_subgroup , strata = Null , permutations 999"

$`1_vs_3`
      Df SumOfSqs      R2      F Pr(>F)
ATR_subgroup 1 0.000501 0.00876 0.7157   0.41
Residual     81 0.056651 0.99124
Total        82 0.057152 1.00000
---
$`1_vs_2`
      Df SumOfSqs      R2      F Pr(>F)
ATR_subgroup 1 0.002152 0.06056 4.4477  0.047 *
Residual     69 0.033389 0.93944
Total        70 0.035541 1.00000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

$`3_vs_2`
      Df SumOfSqs      R2      F Pr(>F)
ATR_subgroup 1 0.003965 0.08753 5.5635  0.027 *
Residual     58 0.041333 0.91247
Total        59 0.045298 1.00000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

attr("class")
[1] "pwadstrata" "list"

```

Las comparaciones por pares ponen de relieve que **las diferencias significativas entre subgrupos en cuanto a las métricas Iron-FA se hallan entre preHD vs controles y entre preHD vs HD para ambos dataframes** (completo y reducido, sin y con eliminación de outliers, respectivamente).

### 5.3 ANOVA

Como los modelos MANOVA y PERMANOVA han detectado diferencias entre subgrupos en relación a las 5 métricas de RM, se procede a generar los ANOVA para valorar cada métrica por separado. En este caso, los ANOVA se van a generar con 4 factores: subgrupo, hemisferio, edad factorizada y género. Se empezará con un modelo con interacción de los 4 factores, posteriormente se irán eliminando los términos (interacciones) hasta llegar al modelo más simple. Remarcarse que tanto el hemisferio, la edad como el género son covariables que me interesa mantener en el modelo final aunque no sean significativas para eliminar su influencia en la relación entre métricas-subgrupos.

### 5.3.1 IRON

```

anova1_ATRavg_iron<-aov(ATR_Iron~ATR_subgroup*ATR_hemisphere*ATR_agef2*ATR_Gender,
                           data=ATR_DifIron_Df)
summary(anova1_ATRavg_iron)

Df      Sum Sq   Mean Sq F value
ATR_subgroup          2 9.490e-06 4.744e-06  1.726
ATR_hemisphere        1 2.460e-05 2.460e-05  8.951
ATR_agef2              2 9.480e-06 4.738e-06  1.724
ATR_Gender             1 9.600e-07 9.550e-07  0.348
ATR_subgroup:ATR_hemisphere  2 3.180e-06 1.591e-06  0.579
ATR_subgroup:ATR_agef2    4 8.050e-06 2.011e-06  0.732
ATR_hemisphere:ATR_agef2 2 5.270e-06 2.636e-06  0.959
ATR_subgroup:ATR_Gender   2 1.239e-05 6.197e-06  2.255
ATR_hemisphere:ATR_Gender 1 4.200e-07 4.170e-07  0.152
ATR_agef2:ATR_Gender     2 2.384e-05 1.192e-05  4.337
ATR_subgroup:ATR_hemisphere:ATR_agef2 4 1.470e-06 3.670e-07  0.134
ATR_subgroup:ATR_hemisphere:ATR_Gender 2 2.280e-06 1.138e-06  0.414
ATR_subgroup:ATR_agef2:ATR_Gender       2 1.516e-05 7.581e-06  2.759
ATR_hemisphere:ATR_agef2:ATR_Gender    2 2.730e-06 1.363e-06  0.496
ATR_subgroup:ATR_hemisphere:ATR_agef2:ATR_Gender 2 4.800e-06 2.401e-06  0.874
Residuals                  78 2.143e-04 2.748e-06
Pr(>F)
ATR_subgroup          0.18469
ATR_hemisphere        0.00371 **
ATR_agef2              0.18504
ATR_Gender             0.55716
ATR_subgroup:ATR_hemisphere 0.56287
ATR_subgroup:ATR_agef2    0.57290
ATR_hemisphere:ATR_agef2 0.38764
ATR_subgroup:ATR_Gender   0.11167
ATR_hemisphere:ATR_Gender 0.69792
ATR_agef2:ATR_Gender     0.01636 *
ATR_subgroup:ATR_hemisphere:ATR_agef2 0.96958
ATR_subgroup:ATR_hemisphere:ATR_Gender 0.66226
ATR_subgroup:ATR_agef2:ATR_Gender       0.06957 .
ATR_hemisphere:ATR_agef2:ATR_Gender    0.61080
ATR_subgroup:ATR_hemisphere:ATR_agef2:ATR_Gender 0.42148
Residuals
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova2_ATRavg_iron<-aov(ATR_Iron~ATR_subgroup+ATR_hemisphere+ATR_agef2*ATR_Gender,
                           data=ATR_DifIron_Df)
summary(anova2_ATRavg_iron)

Df      Sum Sq   Mean Sq F value Pr(>F)
ATR_subgroup          2 9.490e-06 4.744e-06  1.771 0.1753
ATR_hemisphere        1 2.460e-05 2.460e-05  9.185 0.0031 **
ATR_agef2              2 9.480e-06 4.738e-06  1.769 0.1757
ATR_Gender             1 9.600e-07 9.550e-07  0.357 0.5517
ATR_agef2:ATR_Gender   2 2.345e-05 1.173e-05  4.379 0.0150 *
Residuals                 101 2.705e-04 2.678e-06
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

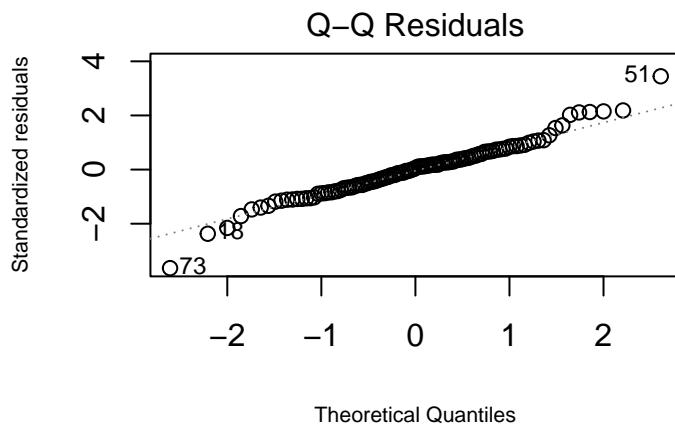
```

shapiro.test(residuals(anova2_ATRavg_iron))

Shapiro-Wilk normality test

data: residuals(anova2_ATRavg_iron)
W = 0.97283, p-value = 0.02387
plot(anova2_ATRavg_iron, which=2, cex.lab=0.7, cex.sub=0.7)

```



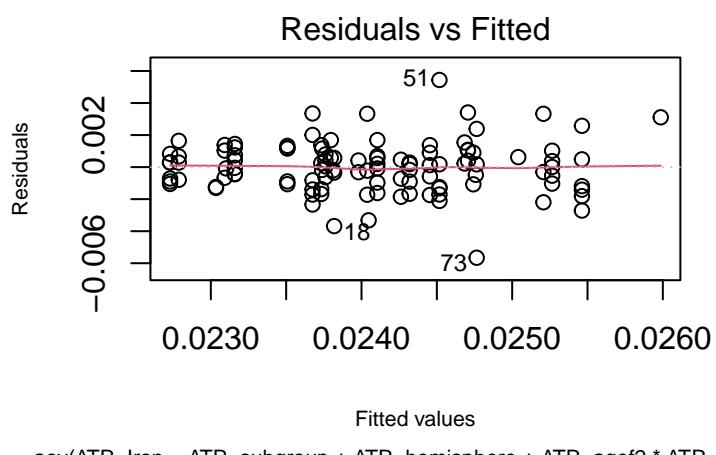
```

bptest(anova2_ATRavg_iron)

studentized Breusch-Pagan test

data: anova2_ATRavg_iron
BP = 9.6193, df = 8, p-value = 0.2928
plot(anova2_ATRavg_iron, which=1, cex.lab=0.7, cex.sub=0.7)

```



En el modelo con interacción `anova1_ATRavg_iron`, se objetiva interacción significativa entre edad factorizada y género y el factor significativo es el hemisferio, por lo que se genera el modelo `anova2_ATRavg_iron` con la

eliminación de las interacciones no significativas. Los residuos del modelo `anova2_ATRavg_iron` no siguen una distribución normal, pero presentan homocedasticidad de varianzas. Los posibles sujetos outliers e influyentes son los individuos con los índices 18, 51 y 73.

Se procede a eliminar los outliers con la fórmula  $q1 - 1.5 * IQR$  y  $q3 + 1.5 * IQR$ .

```

ATR_DifIron_Df_out<-ATR_DifIron_Df
resumen<-data.frame()
resumen<-ATR_DifIron_Df_out |>
  dplyr:::summarize(across(ATR_Iron, list(iqr= ~ IQR(.), q1= ~quantile(.,probs = .25),
                            q3= ~quantile(.,probs = .75))))
index<-which(ATR_DifIron_Df_out$ATR_Iron<(resumen[[2]]-1.5*resumen[[1]]) | 
             ATR_DifIron_Df_out$ATR_Iron>(resumen[[3]]+1.5*resumen[[1]]))
index
[1] 51 73 84 88 89 106

ATR_DifIron_Df_out$ATR_Iron<-replace(ATR_DifIron_Df_out$ATR_Iron, index, NA)

anova3_ATRavg_iron<-aov(ATR_Iron~ATR_subgroup*ATR_hemisphere*ATR_agef2*ATR_Gender,
                           data=ATR_DifIron_Df_out)
summary(anova3_ATRavg_iron)

Df      Sum Sq   Mean Sq F value
ATR_subgroup          2 4.460e-06 2.229e-06  1.435
ATR_hemisphere        1 1.891e-05 1.891e-05 12.176
ATR_agef2              2 1.430e-06 7.130e-07  0.459
ATR_Gender             1 8.000e-08 7.600e-08  0.049
ATR_subgroup:ATR_hemisphere 2 6.000e-07 2.990e-07  0.193
ATR_subgroup:ATR_agef2  4 6.100e-06 1.524e-06  0.981
ATR_hemisphere:ATR_agef2 2 1.520e-06 7.610e-07  0.490
ATR_subgroup:ATR_Gender  2 3.530e-06 1.763e-06  1.135
ATR_hemisphere:ATR_Gender 1 1.020e-06 1.019e-06  0.656
ATR_agef2:ATR_Gender   2 2.632e-05 1.316e-05  8.472
ATR_subgroup:ATR_hemisphere:ATR_agef2 4 2.340e-06 5.840e-07  0.376
ATR_subgroup:ATR_hemisphere:ATR_Gender 1 8.700e-07 8.710e-07  0.561
ATR_subgroup:ATR_agef2:ATR_Gender   2 5.170e-06 2.584e-06  1.664
ATR_hemisphere:ATR_agef2:ATR_Gender 2 4.700e-06 2.352e-06  1.514
ATR_subgroup:ATR_hemisphere:ATR_agef2:ATR_Gender 2 7.910e-06 3.954e-06  2.545
Residuals            73 1.134e-04 1.553e-06
Pr(>F)
ATR_subgroup          0.244827
ATR_hemisphere        0.000825 ***
ATR_agef2              0.633941
ATR_Gender             0.825588
ATR_subgroup:ATR_hemisphere 0.825217
ATR_subgroup:ATR_agef2  0.423317
ATR_hemisphere:ATR_agef2 0.614855
ATR_subgroup:ATR_Gender 0.327014
ATR_hemisphere:ATR_Gender 0.420643
ATR_agef2:ATR_Gender   0.000491 ***
ATR_subgroup:ATR_hemisphere:ATR_agef2 0.824904
ATR_subgroup:ATR_hemisphere:ATR_Gender 0.456372
ATR_subgroup:ATR_agef2:ATR_Gender   0.196555
ATR_hemisphere:ATR_agef2:ATR_Gender 0.226792
ATR_subgroup:ATR_hemisphere:ATR_agef2:ATR_Gender 0.085410 .
Residuals

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

6 observations deleted due to missingness

anova4_ATRavg_iron<-aov(ATR_Iron~ATR_subgroup+ATR_hemisphere+ATR_agef2*ATR_Gender,
                           data=ATR_DifIron_Df_out)
summary(anova4_ATRavg_iron)
  Df   Sum Sq  Mean Sq F value    Pr(>F)
ATR_subgroup      2 4.460e-06 2.229e-06   1.348 0.26469
ATR_hemisphere    1 1.891e-05 1.891e-05  11.440 0.00105 ** 
ATR_agef2         2 1.430e-06 7.130e-07   0.431 0.65117
ATR_Gender         1 8.000e-08 7.600e-08   0.046 0.83072
ATR_agef2:ATR_Gender 2 1.640e-05 8.202e-06   4.960 0.00893 ** 
Residuals          95 1.571e-04 1.653e-06

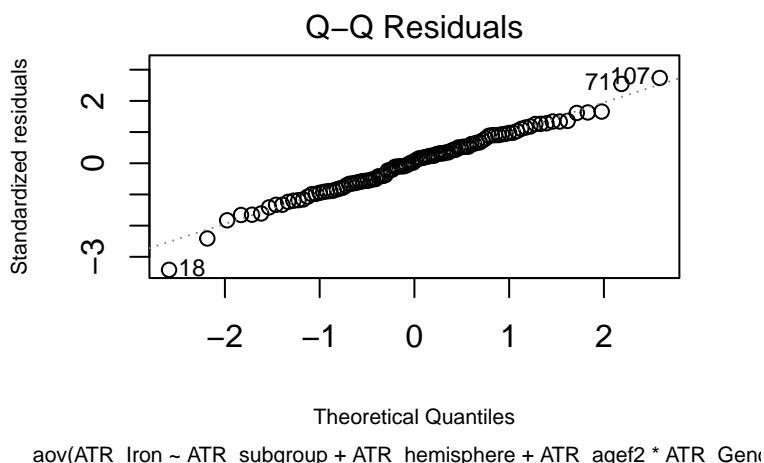
-----
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
6 observations deleted due to missingness

shapiro.test(residuals(anova4_ATRavg_iron))

Shapiro-Wilk normality test

data:  residuals(anova4_ATRavg_iron)
W = 0.98969, p-value = 0.6124
plot(anova4_ATRavg_iron, which=2, cex.lab=0.7, cex.sub=0.7)

```



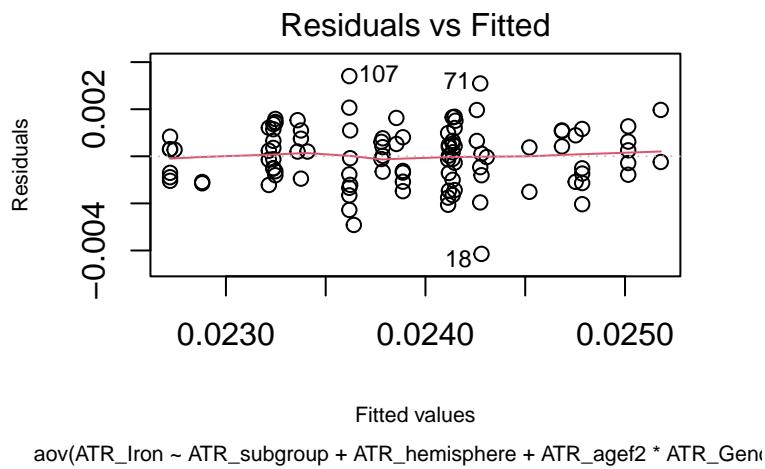
```

bptest(anova4_ATRavg_iron)

studentized Breusch-Pagan test

data:  anova4_ATRavg_iron
BP = 6.165, df = 8, p-value = 0.6288
plot(anova4_ATRavg_iron, which=1, cex.lab=0.7, cex.sub=0.7)

```



Una vez eliminado los outliers (individuos con índices 51, 73, 84, 88, 89, 106), los residuos del modelo `anova4_ATRavg_iron` presentan una distribución normal y una homocedasticidad de varianzas. Los términos significativos son la interacción entre edad factorizada y género y el factor hemisferio.

### 5.3.2 AD

```
anova1_ATRavg_ad<-aov(ATR_AD~ATR_subgroup*ATR_hemisphere*ATR_agef2*ATR_Gender,
                           data=ATR_DifIron_Df)
summary(anova1_ATRavg_ad)
```

	Df	Sum Sq	Mean Sq	F value
ATR_subgroup	2	3.863e-08	1.932e-08	12.444
ATR_hemisphere	1	9.860e-09	9.864e-09	6.355
ATR_agef2	2	6.370e-09	3.184e-09	2.052
ATR_Gender	1	1.950e-09	1.950e-09	1.256
ATR_subgroup:ATR_hemisphere	2	1.260e-09	6.320e-10	0.407
ATR_subgroup:ATR_agef2	4	8.690e-09	2.172e-09	1.399
ATR_hemisphere:ATR_agef2	2	8.400e-10	4.190e-10	0.270
ATR_subgroup:ATR_Gender	2	2.400e-10	1.180e-10	0.076
ATR_hemisphere:ATR_Gender	1	5.500e-10	5.500e-10	0.355
ATR_agef2:ATR_Gender	2	7.430e-09	3.716e-09	2.394
ATR_subgroup:ATR_hemisphere:ATR_agef2	4	6.800e-10	1.710e-10	0.110
ATR_subgroup:ATR_hemisphere:ATR_Gender	2	9.000e-11	4.600e-11	0.029
ATR_subgroup:ATR_agef2:ATR_Gender	2	2.830e-09	1.417e-09	0.913
ATR_hemisphere:ATR_agef2:ATR_Gender	2	3.300e-10	1.650e-10	0.106
ATR_subgroup:ATR_hemisphere:ATR_agef2:ATR_Gender	2	5.400e-10	2.710e-10	0.175
Residuals	78	1.211e-07	1.552e-09	
			Pr(>F)	
ATR_subgroup		2.04e-05	***	
ATR_hemisphere		0.0137	*	
ATR_agef2		0.1354		
ATR_Gender		0.2658		
ATR_subgroup:ATR_hemisphere		0.6668		
ATR_subgroup:ATR_agef2		0.2420		
ATR_hemisphere:ATR_agef2		0.7642		
ATR_subgroup:ATR_Gender		0.9271		

```

ATR_hemisphere:ATR_Gender           0.5533
ATR_agef2:ATR_Gender                 0.0979 .
ATR_subgroup:ATR_hemisphere:ATR_agef2 0.9786
ATR_subgroup:ATR_hemisphere:ATR_Gender 0.9710
ATR_subgroup:ATR_agef2:ATR_Gender      0.4055
ATR_hemisphere:ATR_agef2:ATR_Gender   0.8992
ATR_subgroup:ATR_hemisphere:ATR_agef2:ATR_Gender 0.8402
Residuals

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova2_ATRavg_ad<-aov(ATR_AD~ATR_subgroup+ATR_hemisphere+ATR_agef2+ATR_Gender,
                         data=ATR_DifIron_Df)
summary(anova2_ATRavg_ad)
  Df    Sum Sq  Mean Sq F value    Pr(>F)
ATR_subgroup     2 3.863e-08 1.932e-08 13.762 5.05e-06 ***
ATR_hemisphere   1 9.860e-09 9.864e-09  7.028  0.00929 **
ATR_agef2         2 6.370e-09 3.184e-09  2.269  0.10858
ATR_Gender        1 1.950e-09 1.950e-09  1.389  0.24128
Residuals       103 1.446e-07 1.404e-09

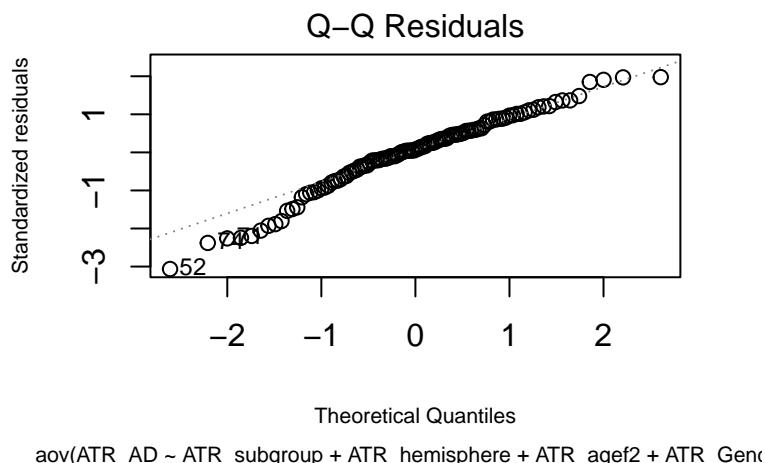
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

shapiro.test(residuals(anova2_ATRavg_ad))

Shapiro-Wilk normality test

data:  residuals(anova2_ATRavg_ad)
W = 0.97049, p-value = 0.01523
plot(anova2_ATRavg_ad, which=2, cex.lab=0.7, cex.sub=0.7)

```



```

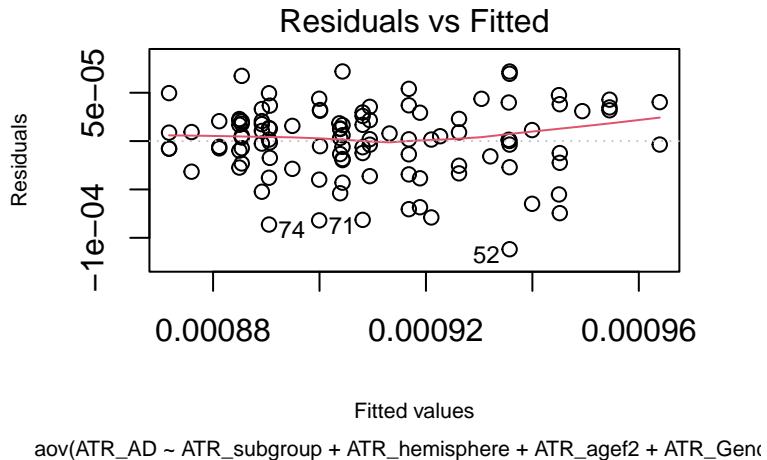
bptest(anova2_ATRavg_ad)

studentized Breusch-Pagan test

data:  anova2_ATRavg_ad

```

```
BP = 6.7483, df = 6, p-value = 0.3448
plot(anova2_ATRavg_ad, which=1, cex.lab=0.7, cex.sub=0.7)
```



```
aov(ATR_AD ~ ATR_subgroup + ATR_hemisphere + ATR_agef2 + ATR_Gender)
```

En el modelo con interacción `anova1_ATRavg_ad`, no se objetiva interacción significativa y los factores significativos son el subgrupo y el hemisferio, por lo que se genera el modelo `anova2_ATRavg_ad` con eliminación de las interacciones. Los residuos del modelo `anova2_ATRavg_ad` no siguen una distribución normal, pero presentan homocedasticidad de varianzas. Los posibles sujetos outliers e influyentes son los individuos con los índices 52, 71 y 74.

Se procede a la eliminación de outliers.

```
ATR_DifIron_Df_out<-ATR_DifIron_Df
resumen<-data.frame()
resumen<-ATR_DifIron_Df_out |>
  dplyr::summarize(across(ATR_AD, list(iqr= ~ IQR(.),
                                         q1= ~quantile(.,probs = .25),
                                         q3= ~quantile(.,probs = .75))))
index <-which(ATR_DifIron_Df_out$ATR_AD<(resumen[[2]]-1.5*resumen[[1]]) |>
  ATR_DifIron_Df_out$ATR_AD>(resumen[[3]]+1.5*resumen[[1]]))
index
integer(0)
```

No se detectan outliers. Por tanto, en este caso, no se pueden generar nuevos modelos.

### 5.3.3 MD

```
anova1_ATRavg_md<-aov(ATR_MD~ATR_subgroup*ATR_hemisphere*ATR_agef2*ATR_Gender,
                           data=ATR_DifIron_Df)
summary(anova1_ATRavg_md)
```

	Df	Sum Sq	Mean Sq	F value
ATR_subgroup	2	2.834e-08	1.417e-08	17.566
ATR_hemisphere	1	5.290e-09	5.293e-09	6.563
ATR_agef2	2	5.490e-09	2.747e-09	3.406
ATR_Gender	1	6.500e-10	6.500e-10	0.806

```

ATR_subgroup:ATR_hemisphere          2 3.900e-10 1.960e-10  0.243
ATR_subgroup:ATR_agef2                4 3.520e-09 8.800e-10  1.092
ATR_hemisphere:ATR_agef2            2 3.500e-10 1.760e-10  0.218
ATR_subgroup:ATR_Gender              2 5.600e-10 2.810e-10  0.348
ATR_hemisphere:ATR_Gender           1 1.100e-10 1.060e-10  0.132
ATR_agef2:ATR_Gender                 2 1.770e-09 8.860e-10  1.099
ATR_subgroup:ATR_hemisphere:ATR_agef2 4 2.600e-10 6.500e-11  0.081
ATR_subgroup:ATR_hemisphere:ATR_Gender 2 1.000e-11 4.000e-12  0.005
ATR_subgroup:ATR_agef2:ATR_Gender      2 8.700e-10 4.330e-10  0.537
ATR_hemisphere:ATR_agef2:ATR_Gender   2 6.000e-11 3.000e-11  0.038
ATR_subgroup:ATR_hemisphere:ATR_agef2:ATR_Gender 2 1.600e-10 8.000e-11  0.099
Residuals                           78 6.291e-08 8.070e-10

Pr(>F)
5.03e-07 ***
0.0123 *
0.0382 *
0.3720
0.7849
0.3666
0.8045
0.7073
0.7175
0.3383
0.9880
0.9954
0.5866
0.9629
0.9061

Residuals
---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova2_ATRavg_md<-aov(ATR_MD~ATR_subgroup+ATR_hemisphere+ATR_agef2+ATR_Gender,
                         data=ATR_DifIron_Df)
summary(anova2_ATRavg_md)
  Df  Sum Sq  Mean Sq F value    Pr(>F)
ATR_subgroup     2 2.834e-08 1.417e-08  20.562 3.06e-08 ***
ATR_hemisphere   1 5.290e-09 5.293e-09   7.682  0.00662 **
ATR_agef2        2 5.490e-09 2.747e-09   3.987  0.02149 *
ATR_Gender        1 6.500e-10 6.500e-10   0.944  0.33364
Residuals       103 7.097e-08 6.890e-10

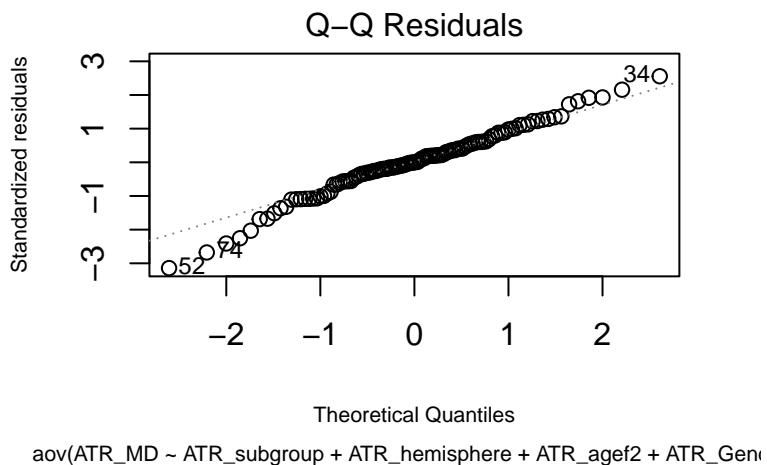
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

shapiro.test(residuals(anova2_ATRavg_md))

Shapiro-Wilk normality test

data:  residuals(anova2_ATRavg_md)
W = 0.98227, p-value = 0.152
plot(anova2_ATRavg_md, which=2, cex.lab=0.7, cex.sub=0.7)

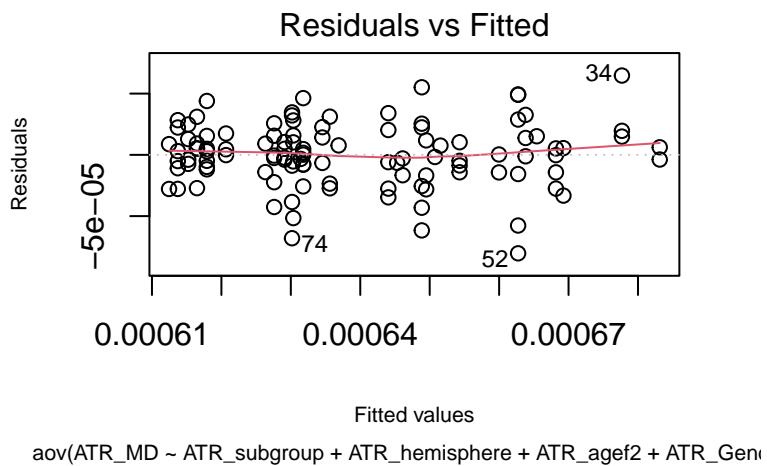
```



```
bptest(anova2_ATRavg_md)
```

```
studentized Breusch-Pagan test
```

```
data: anova2_ATRavg_md
BP = 15.575, df = 6, p-value = 0.01622
plot(anova2_ATRavg_md, which=1, cex.lab=0.7, cex.sub=0.7)
```



En el modelo con interacción `anova1_ATRavg_md`, no se objetiva interacción significativa, por lo que se genera el modelo `anova2_ATRavg_md` con eliminación de las interacciones.

En el modelo `anova2_ATRavg_md`, los factores significativos son el subgrupo, el hemisferio y la edad factorizada. Los residuos del modelo `anova2_ATRavg_md` siguen una distribución normal, pero no presentan homocedasticidad de varianzas. Los posibles sujetos outliers e influyentes son los individuos con los índices 34, 52 y 74.

Se procede a la eliminación de outliers.

```

ATR_DifIron_Df_out<-ATR_DifIron_Df
resumen<-data.frame()
resumen<-ATR_DifIron_Df_out |>
  dplyr::summarize(across(ATR_MD, list(iqr= ~ IQR(.), q1= ~quantile(.,probs = .25),
                           q3= ~quantile(.,probs = .75))))
index <-which(ATR_DifIron_Df_out$ATR_MD<(resumen[[2]]-1.5*resumen[[1]]) |  

              ATR_DifIron_Df_out$ATR_MD>(resumen[[3]]+1.5*resumen[[1]]))
index
[1] 34 74

ATR_DifIron_Df_out$ATR_MD<-replace(ATR_DifIron_Df_out$ATR_MD, index, NA)

anova3_ATRavg_md<-aov(ATR_MD~ATR_subgroup*ATR_hemisphere*ATR_agef2*ATR_Gender,  

                        data=ATR_DifIron_Df_out)
summary(anova3_ATRavg_md)

Df      Sum Sq   Mean Sq F value
ATR_subgroup          2 2.300e-08 1.150e-08 15.293
ATR_hemisphere         1 3.460e-09 3.465e-09  4.607
ATR_agef2              2 5.370e-09 2.687e-09  3.573
ATR_Gender              1 5.400e-10 5.360e-10  0.713
ATR_subgroup:ATR_hemisphere 2 1.600e-10 8.200e-11  0.110
ATR_subgroup:ATR_agef2    4 1.510e-09 3.780e-10  0.502
ATR_hemisphere:ATR_agef2 2 2.000e-11 9.000e-12  0.011
ATR_subgroup:ATR_Gender   2 2.000e-10 1.000e-10  0.134
ATR_hemisphere:ATR_Gender 1 0.000e+00 2.000e-12  0.003
ATR_agef2:ATR_Gender     2 1.900e-09 9.510e-10  1.265
ATR_subgroup:ATR_hemisphere:ATR_agef2 4 1.900e-10 4.900e-11  0.065
ATR_subgroup:ATR_hemisphere:ATR_Gender 2 1.000e-11 7.000e-12  0.009
ATR_subgroup:ATR_agef2:ATR_Gender       2 3.100e-10 1.540e-10  0.205
ATR_hemisphere:ATR_agef2:ATR_Gender    2 1.100e-10 5.600e-11  0.074
ATR_subgroup:ATR_hemisphere:ATR_agef2:ATR_Gender 2 1.600e-10 8.200e-11  0.109
Residuals                  76 5.715e-08 7.520e-10
Pr(>F)
ATR_subgroup          2.62e-06 ***
ATR_hemisphere        0.0350 *
ATR_agef2              0.0329 *
ATR_Gender              0.4010
ATR_subgroup:ATR_hemisphere 0.8964
ATR_subgroup:ATR_agef2    0.7342
ATR_hemisphere:ATR_agef2 0.9887
ATR_subgroup:ATR_Gender   0.8752
ATR_hemisphere:ATR_Gender 0.9597
ATR_agef2:ATR_Gender     0.2882
ATR_subgroup:ATR_hemisphere:ATR_agef2 0.9922
ATR_subgroup:ATR_hemisphere:ATR_Gender 0.9907
ATR_subgroup:ATR_agef2:ATR_Gender       0.8147
ATR_hemisphere:ATR_agef2:ATR_Gender    0.9288
ATR_subgroup:ATR_hemisphere:ATR_agef2:ATR_Gender 0.8965
Residuals

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
2 observations deleted due to missingness

anova4_ATRavg_md<-aov(ATR_MD~ATR_subgroup+ATR_hemisphere+ATR_agef2+ATR_Gender,  

                        data=ATR_DifIron_Df_out)

```

```

summary(anova4_ATRavg_md)
  Df Sum Sq Mean Sq F value Pr(>F)
ATR_subgroup  2 2.300e-08 1.150e-08 18.813 1.14e-07 ***
ATR_hemisphere 1 3.460e-09 3.465e-09 5.668 0.0192 *
ATR_agef2     2 5.370e-09 2.687e-09 4.395 0.0148 *
ATR_Gender    1 5.400e-10 5.360e-10 0.877 0.3511
Residuals     101 6.175e-08 6.110e-10

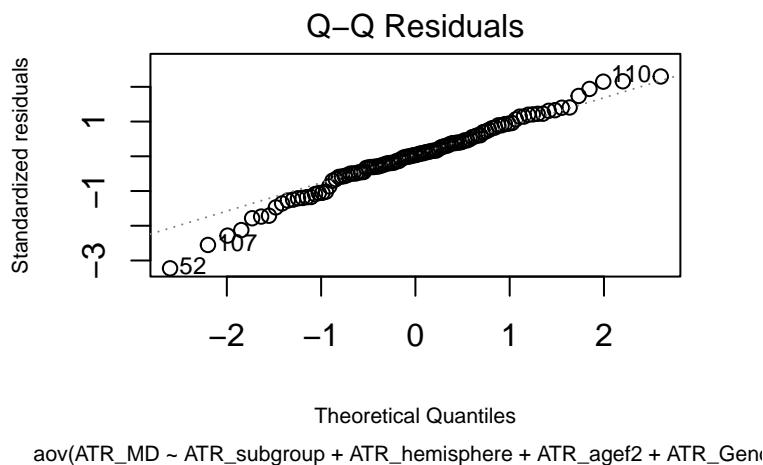
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
2 observations deleted due to missingness

shapiro.test(residuals(anova4_ATRavg_md))

Shapiro-Wilk normality test

data: residuals(anova4_ATRavg_md)
W = 0.98482, p-value = 0.2591
plot(anova4_ATRavg_md, which=2, cex.lab=0.7, cex.sub=0.7)

```



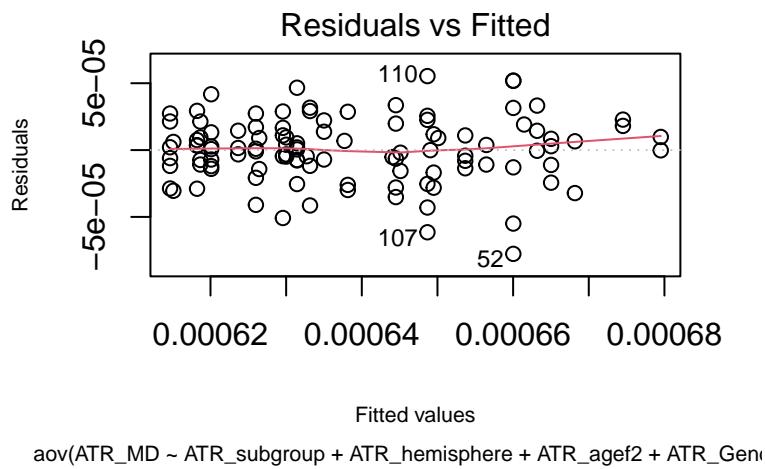
```

bptest(anova4_ATRavg_md)

studentized Breusch-Pagan test

data: anova4_ATRavg_md
BP = 15.325, df = 6, p-value = 0.01787
plot(anova4_ATRavg_md, which=1, cex.lab=0.7, cex.sub=0.7)

```



Con la eliminación de outliers (individuos con índices 34, 74), no se ha mejorado la homocedasticidad de varianzas. Los residuos del modelo `anova4_ATRavg_md` siguen presentando una distribución normal y los elementos significativos son el subgrupo, el hemisferio y la edad factorizada.

#### 5.3.4 RD

```
anova1_ATRavg_rd<-aov(ATR_RD~ATR_subgroup*ATR_hemisphere*ATR_agef2*ATR_Gender,
                           data=ATR_DifIron_Df)
summary(anova1_ATRavg_rd)

Df      Sum Sq   Mean Sq F value
ATR_subgroup          2 2.503e-08 1.252e-08 14.980
ATR_hemisphere        1 3.540e-09 3.537e-09  4.233
ATR_agef2              2 5.850e-09 2.926e-09  3.502
ATR_Gender              1 2.600e-10 2.620e-10  0.313
ATR_subgroup:ATR_hemisphere  2 1.900e-10 9.600e-11  0.114
ATR_subgroup:ATR_agef2    4 2.180e-09 5.450e-10  0.652
ATR_hemisphere:ATR_agef2 2 2.500e-10 1.270e-10  0.152
ATR_subgroup:ATR_Gender   2 1.720e-09 8.590e-10  1.029
ATR_hemisphere:ATR_Gender 1 1.000e-11 1.400e-11  0.017
ATR_agef2:ATR_Gender     2 9.900e-10 4.950e-10  0.592
ATR_subgroup:ATR_hemisphere:ATR_agef2 4 2.700e-10 6.800e-11  0.081
ATR_subgroup:ATR_hemisphere:ATR_Gender 2 7.000e-11 3.700e-11  0.044
ATR_subgroup:ATR_agef2:ATR_Gender       2 1.390e-09 6.970e-10  0.834
ATR_hemisphere:ATR_agef2:ATR_Gender    2 5.000e-11 2.400e-11  0.029
ATR_subgroup:ATR_hemisphere:ATR_agef2:ATR_Gender 2 1.400e-10 6.900e-11  0.083
Residuals                78 6.517e-08 8.360e-10

Pr(>F)
ATR_subgroup            3.12e-06 ***
ATR_hemisphere          0.043 *
ATR_agef2                0.035 *
ATR_Gender                 0.577
ATR_subgroup:ATR_hemisphere 0.892
ATR_subgroup:ATR_agef2      0.627
ATR_hemisphere:ATR_agef2    0.859
ATR_subgroup:ATR_Gender      0.362
```

```

ATR_hemisphere:ATR_Gender          0.897
ATR_agef2:ATR_Gender               0.555
ATR_subgroup:ATR_hemisphere:ATR_agef2 0.988
ATR_subgroup:ATR_hemisphere:ATR_Gender 0.957
ATR_subgroup:ATR_agef2:ATR_Gender    0.438
ATR_hemisphere:ATR_agef2:ATR_Gender 0.972
ATR_subgroup:ATR_hemisphere:ATR_agef2:ATR_Gender 0.921
Residuals

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova2_ATRavg_rd<-aov(ATR_RD~ATR_subgroup+ATR_hemisphere+ATR_agef2+ATR_Gender,
                           data=ATR_DifIron_Df)
summary(anova2_ATRavg_rd)
  Df   Sum Sq  Mean Sq F value Pr(>F)
ATR_subgroup     2 2.503e-08 1.252e-08 17.795 2.3e-07 ***
ATR_hemisphere   1 3.540e-09 3.537e-09  5.029  0.0271 *
ATR_agef2         2 5.850e-09 2.926e-09  4.160  0.0183 *
ATR_Gender        1 2.600e-10 2.620e-10  0.372  0.5434
Residuals       103 7.245e-08 7.030e-10

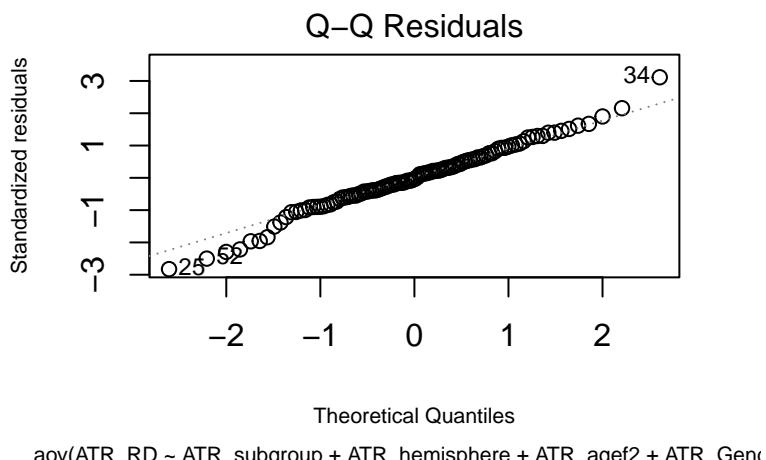
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

shapiro.test(residuals(anova2_ATRavg_rd))

Shapiro-Wilk normality test

data:  residuals(anova2_ATRavg_rd)
W = 0.98784, p-value = 0.4258
plot(anova2_ATRavg_rd, which=2, cex.lab=0.7, cex.sub=0.7)

```



```

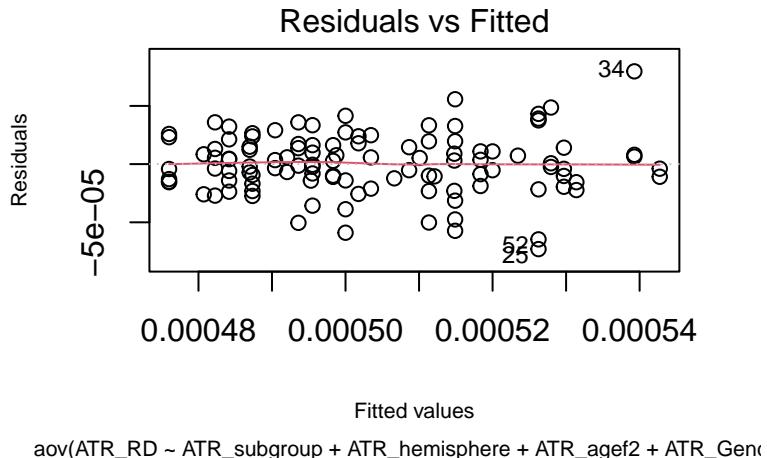
bptest(anova2_ATRavg_rd)

studentized Breusch-Pagan test

data:  anova2_ATRavg_rd

```

```
BP = 19.384, df = 6, p-value = 0.003562
plot(anova2_ATRavg_rd, which=1, cex.lab=0.7, cex.sub=0.7)
```



```
aov(ATR_RD ~ ATR_subgroup + ATR_hemisphere + ATR_aged2 + ATR_Gender)
```

En el modelo con interacción `anova1_ATRavg_rd`, no se objetiva interacción significativa, por lo que se genera el modelo `anova2_ATRavg_rd` con eliminación de las interacciones.

En el modelo `anova2_ATRavg_rd`, los factores significativos son el subgrupo, el hemisferio y la edad factorizada. Los residuos del modelo `anova2_ATRavg_rd` siguen una distribución normal, pero no presentan homocedasticidad de varianzas. Los posibles sujetos outliers e influyentes son los individuos con los índices 25, 34 y 52.

Se procede a la eliminación de outliers.

```
ATR_DifIron_Df_out<-ATR_DifIron_Df
resumen<-data.frame()
resumen<-ATR_DifIron_Df_out |>
  dplyr::summarize(across(ATR_RD, list(iqr= ~ IQR(.), q1= ~quantile(.,probs = .25),
                           q3= ~quantile(.,probs = .75))))
index <-which(ATR_DifIron_Df_out$ATR_RD<(resumen[[2]]-1.5*resumen[[1]]) | 
              ATR_DifIron_Df_out$ATR_RD>(resumen[[3]]+1.5*resumen[[1]]))
index
[1] 34

ATR_DifIron_Df_out$ATR_RD<-replace(ATR_DifIron_Df_out$ATR_RD, index, NA)

anova3_ATRavg_rd<-aov(ATR_RD~ATR_subgroup*ATR_hemisphere*ATR_aged2*ATR_Gender,
                        data=ATR_DifIron_Df_out)
summary(anova3_ATRavg_rd)
Df      Sum Sq   Mean Sq F value
ATR_subgroup                      2 2.076e-08 1.038e-08 12.957
ATR_hemisphere                     1 2.540e-09 2.545e-09  3.177
ATR_aged2                          2 4.220e-09 2.108e-09  2.632
ATR_Gender                          1 5.300e-10 5.320e-10  0.665
ATR_subgroup:ATR_hemisphere        2 1.000e-11 7.000e-12  0.008
ATR_subgroup:ATR_aged2             4 1.040e-09 2.610e-10  0.325
ATR_hemisphere:ATR_aged2          2 4.000e-11 2.200e-11  0.027
ATR_subgroup:ATR_Gender            2 1.150e-09 5.750e-10  0.718
```

```

ATR_hemisphere:ATR_Gender           1 1.000e-11 6.000e-12  0.007
ATR_agef2:ATR_Gender                2 4.400e-10 2.180e-10  0.272
ATR_subgroup:ATR_hemisphere:ATR_agef2 4 9.000e-11 2.100e-11  0.027
ATR_subgroup:ATR_hemisphere:ATR_Gender 2 2.600e-10 1.310e-10  0.164
ATR_subgroup:ATR_agef2:ATR_Gender     2 6.600e-10 3.310e-10  0.413
ATR_hemisphere:ATR_agef2:ATR_Gender 2 2.000e-11 9.000e-12  0.011
ATR_subgroup:ATR_hemisphere:ATR_agef2:ATR_Gender 2 2.300e-10 1.170e-10  0.146
Residuals                           77 6.167e-08 8.010e-10
                                         Pr(>F)
ATR_subgroup                         1.41e-05 ***
ATR_hemisphere                      0.0786 .
ATR_agef2                            0.0784 .
ATR_Gender                           0.4175
ATR_subgroup:ATR_hemisphere          0.9918
ATR_subgroup:ATR_agef2               0.8601
ATR_hemisphere:ATR_agef2            0.9731
ATR_subgroup:ATR_Gender              0.4911
ATR_hemisphere:ATR_Gender           0.9318
ATR_agef2:ATR_Gender                 0.7624
ATR_subgroup:ATR_hemisphere:ATR_agef2 0.9986
ATR_subgroup:ATR_hemisphere:ATR_Gender 0.8493
ATR_subgroup:ATR_agef2:ATR_Gender    0.6632
ATR_hemisphere:ATR_agef2:ATR_Gender 0.9892
ATR_subgroup:ATR_hemisphere:ATR_agef2:ATR_Gender 0.8644
Residuals
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
1 observation deleted due to missingness

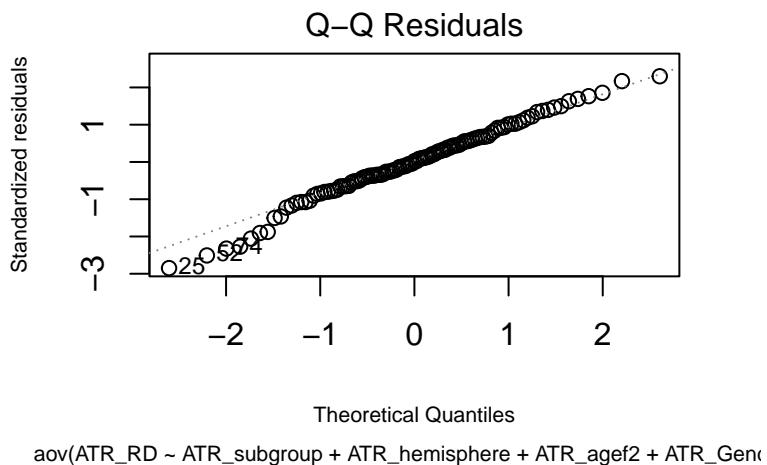
anova4_ATRavg_rd<-aov(ATR_RD~ATR_subgroup+ATR_hemisphere+ATR_agef2+ATR_Gender,
                         data=ATR_DifIron_Df_out)
summary(anova4_ATRavg_rd)
   Df   Sum Sq  Mean Sq F value    Pr(>F)
ATR_subgroup      2 2.076e-08 1.038e-08 16.131 8.19e-07 ***
ATR_hemisphere   1 2.540e-09 2.545e-09  3.955  0.0494 *
ATR_agef2         2 4.220e-09 2.108e-09  3.277  0.0417 *
ATR_Gender        1 5.300e-10 5.320e-10  0.827  0.3652
Residuals         102 6.562e-08 6.430e-10
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
1 observation deleted due to missingness

shapiro.test(residuals(anova4_ATRavg_rd))

  Shapiro-Wilk normality test

data:  residuals(anova4_ATRavg_rd)
W = 0.98759, p-value = 0.4147
plot(anova4_ATRavg_rd, which=2, cex.lab=0.7, cex.sub=0.7)

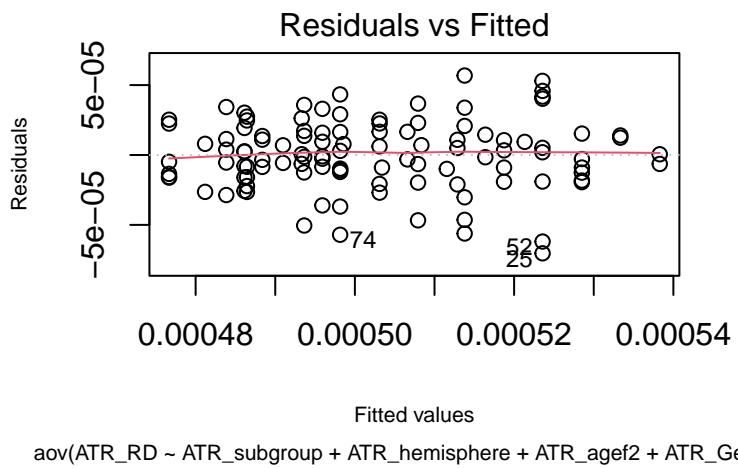
```



```
bptest(anova4_ATRavg_rd)

studentized Breusch-Pagan test

data: anova4_ATRavg_rd
BP = 19.492, df = 6, p-value = 0.003409
plot(anova4_ATRavg_rd, which=1, cex.lab=0.7, cex.sub=0.7)
```



Con la eliminación del outlier (individuo con índice 34), no se ha mejorado la homocedasticidad de varianzas. Los residuos del modelo `anova4_ATRavg_rd` siguen presentando una distribución normal y los elementos significativos son el subgrupo, el hemisferio y la edad factorizada.

### 5.3.5 FA

```

anova1_ATRavg_fa<-aov(ATR_FA~ATR_subgroup*ATR_hemisphere*ATR_agef2*ATR_Gender,
                         data=ATR_DifIron_Df)
summary(anova1_ATRavg_fa)

Df   Sum Sq  Mean Sq F value
ATR_subgroup          2 0.00471 0.0023558 3.396
ATR_hemisphere        1 0.00005 0.0000510 0.074
ATR_agef2              2 0.00143 0.0007150 1.031
ATR_Gender             1 0.00001 0.0000139 0.020
ATR_subgroup:ATR_hemisphere 2 0.00027 0.0001334 0.192
ATR_subgroup:ATR_agef2    4 0.00214 0.0005346 0.771
ATR_hemisphere:ATR_agef2 2 0.00020 0.0001003 0.145
ATR_subgroup:ATR_Gender   2 0.00192 0.0009585 1.382
ATR_hemisphere:ATR_Gender 1 0.00006 0.0000618 0.089
ATR_agef2:ATR_Gender     2 0.00345 0.0017260 2.488
ATR_subgroup:ATR_hemisphere:ATR_agef2 4 0.00041 0.0001023 0.147
ATR_subgroup:ATR_hemisphere:ATR_Gender   2 0.00036 0.0001804 0.260
ATR_subgroup:ATR_agef2:ATR_Gender        2 0.00288 0.0014399 2.075
ATR_hemisphere:ATR_agef2:ATR_Gender      2 0.00012 0.0000617 0.089
ATR_subgroup:ATR_hemisphere:ATR_agef2:ATR_Gender 2 0.00029 0.0001453 0.209
Residuals                78 0.05412 0.0006938

Pr(>F)
0.0385 *
0.7870
0.3616
0.8879
0.8255
0.5476
0.8656
0.2573
0.7661
0.0897 .
0.9636
0.7717
0.1324
0.9150
0.8116

Residuals

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova2_ATRavg_fa<-aov(ATR_FA~ATR_subgroup+ATR_hemisphere+ATR_agef2+ATR_Gender,
                         data=ATR_DifIron_Df)
summary(anova2_ATRavg_fa)

Df   Sum Sq  Mean Sq F value Pr(>F)
ATR_subgroup          2 0.00471 0.0023558 3.665 0.029 *
ATR_hemisphere        1 0.00005 0.0000510 0.079 0.779
ATR_agef2              2 0.00143 0.0007150 1.112 0.333
ATR_Gender             1 0.00001 0.0000139 0.022 0.883
Residuals              103 0.06622 0.0006429

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

shapiro.test(residuals(anova2_ATRavg_fa))

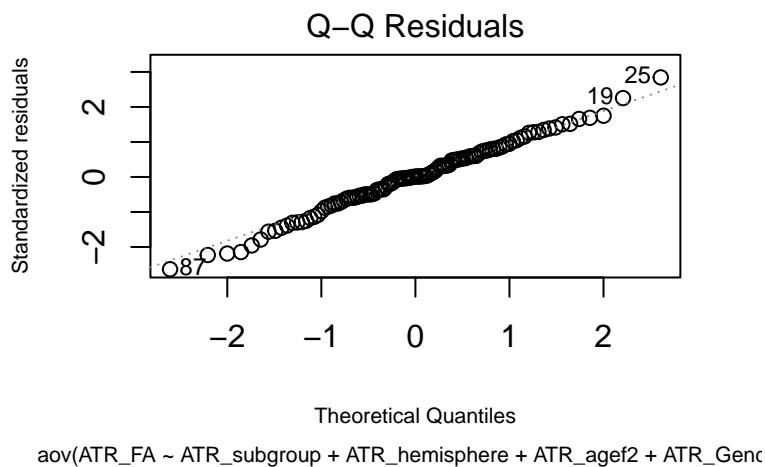
Shapiro-Wilk normality test

```

```

data: residuals(anova2_ATRavg_fa)
W = 0.99434, p-value = 0.9355
plot(anova2_ATRavg_fa, which=2, cex.lab=0.7, cex.sub=0.7)

```



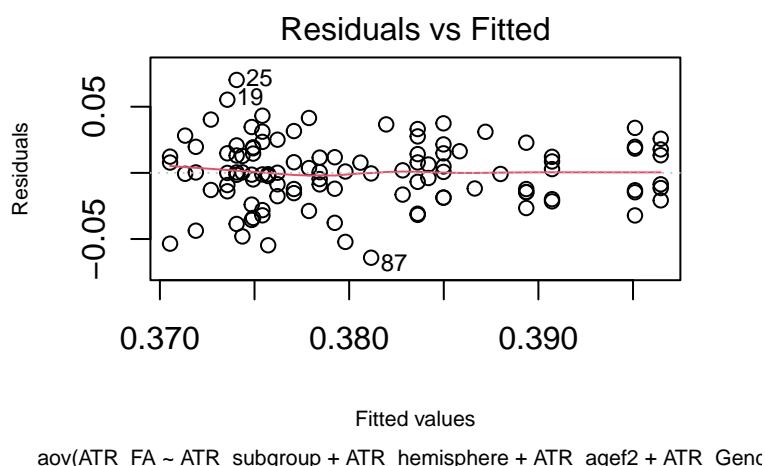
```

bptest(anova2_ATRavg_fa)

studentized Breusch-Pagan test

data: anova2_ATRavg_fa
BP = 8.4379, df = 6, p-value = 0.2077
plot(anova2_ATRavg_fa, which=1, cex.lab=0.7, cex.sub=0.7)

```



En el modelo con interacción `anova1_ATRavg_fa`, no se objetiva interacción significativa, por lo que se genera el modelo `anova2_ATRavg_fa` con eliminación de las interacciones.

En el modelo `anova2_ATRavg_fa`, el factor significativo es el subgrupo. Los residuos del modelo `anova2_ATRavg_fa` siguen una distribución normal y presentan homocedasticidad de varianzas.

**En resumen:**

- El **modelo de hierro** presenta unos residuos con normalidad y homocedasticidad de varianzas, pero el factor subgrupo no es significativo;
- El **modelo de AD** presenta unos residuos sin normalidad y con homocedasticidad de varianzas, y, los factores subgrupo y hemisferio son significativos;
- Los **modelos de MD y RD** presentan unos residuos con normalidad y sin homocedasticidad de varianzas, y, los factores subgrupo, hemisferio y edad factorizada son significativos;
- El **modelo de FA** presenta unos residuos con normalidad y homocedasticidad de varianzas, y, el factor subgrupo es significativo.

También remarcar que los modelos anova de interés están formados por las variables independientes sin interacción.

Se presenta una **tabla resumen** para objetivar la significancia o no de la variable subgrupo, de la normalidad y homocedasticidad/heterocedasticidad de los residuos antes y después de la eliminación de outliers.

Cuadro 83: Modelos Anova para valores Average

ATR-Metrics	Subgroup		Normality		Homoscedasticity		
	F	p-value	W statistic	p-value	BP statistic	df	p-value
Iron	1.771	0.175	0.973	0.024	9.619	8	0.293
Iron-Without Outliers	1.348	0.265	0.99	0.612	6.165	8	0.629
AD	13.762	0	0.97	0.015	6.748	6	0.345
MD	20.562	0	0.982	0.152	15.575	6	0.016
MD-Without Outliers	18.813	0	0.985	0.259	15.325	6	0.018
RD	17.795	0	0.988	0.426	19.384	6	0.004
RD-Without Outliers	16.131	0	0.988	0.415	19.492	6	0.003
FA	3.665	0.029	0.994	0.935	8.438	6	0.208

Dada esta variabilidad de resultados entre métricas y que la eliminación de outliers en cada métrica corresponde a diferentes individuos, se decide probar con métodos no paramétricos.

De todas formas, dado que para la métrica de FA no se ha tenido que eliminar ningún sujeto y que el factor de interés (subgrupo) ha resultado significativo, se genera para esta métrica la comparación por pares.

### 5.3.6 Comparaciones pareadas por subgrupos

Como se ha mencionado, sólo se realizará para la métrica FA.

Cuadro 84: HSD Tukey test para la métrica FA en relación a la variable subgrupo

	Variables	FA(mean)	Groups	Comparison	Difference of means	Adjusted p-value
2	PreHD	0.3922114	a	2-1	0.0130360	0.0924830
1	Control	0.3791754	ab	3-1	-0.0041497	0.7389281
3	HD	0.3750257	b	3-2	-0.0171857	0.0261657

Como se observa, las **diferencias entre subgrupos en la métrica FA de ambos ATRs radica en la comparativa entre HD vs preHD**, con un valor aumentado de FA para preHD en relación a HD.

## 5.4 KRUSKAL WALLIS TEST

El **test de Kruskal Wallis** (generalización de la prueba U de Mann-Whitney para k muestras independientes) es un test no paramétrico (alternativa no paramétrica del ANOVA) para evaluar diferencias estadísticamente significativas entre k grupos independientes con respecto a sus medianas utilizando rangos.

Los **supuestos** que deben cumplirse son los siguientes (Test de Kruskal Wallis (H test) en R):

- **Independencia:** las observaciones dentro de los grupos y entre ellos son independientes;
- **Datos ordinales o continuos:** la variable objeto de estudio debe medirse en una escala ordinal o continua;
- **Homogeneidad de varianzas:** la varianza de la variable dependiente debe ser aproximadamente igual en todos los grupos.
- **Muestreo aleatorio:** los datos deben obtenerse a partir de una muestra aleatoria de la población;
- **No se requiere normalidad.**

Las observaciones (sujetos) son independientes entre sí. Las variables dependientes (métricas de RM) son continuas. El muestreo no es aleatorio ya que es un estudio de casos y controles, con lo que este supuesto no se cumpliría. Respecto a la **homocedasticidad de varianzas**, se procede a su comprobación.

Cuadro 85: Valoración de la homocedasticidad de las métricas de RM con la función fligner test

	Subgroup			Hemisphere			Gender			Factorized Age		
	Statistic	df	p-value	Statistic	df	p-value	Statistic	df	p-value	Statistic	df	p-value
Iron	2.4862	2	0.2885	6.0899	1	0.0136	0.6202	1	0.4310	2.0986	2	0.3502
AD	9.0067	2	0.0111	0.3478	1	0.5554	0.0554	1	0.8139	3.2632	2	0.1956
MD	6.4606	2	0.0395	0.1501	1	0.6985	4.1329	1	0.0421	5.7135	2	0.0575
RD	2.3326	2	0.3115	0.2007	1	0.6541	8.7985	1	0.0030	1.7968	2	0.4072
FA	2.9282	2	0.2313	0.0001	1	0.9917	0.0783	1	0.7796	0.2172	2	0.8971

En relación a los **subgrupos**, las métricas de hierro, difusividad radial y anistropía fraccional presentan homocedasticidad de varianzas, y, las métricas de difusividad axial y difusividad media presentan heterocedasticidad de varianzas.

En relación a los **hemisferios**, las métricas de difusividad axial, difusividad media, difusividad radial y anistropía fraccional presentan homocedasticidad de varianzas, y, la métrica de hierro presenta heterocedasticidad de varianzas.

En relación a los **géneros**, las métricas de hierro, difusividad axial y anistropía fraccional presentan homocedasticidad de varianzas, y, las métricas de difusividad media y difusividad radial presentan heterocedasticidad de varianzas.

En relación a la **edad factorizada**, todas las métricas presentan homocedasticidad de varianzas.

La homocedasticidad o heterocedasticidad de varianzas de las diferentes métricas en relación a las diferentes variables categóricas independientes varía en función de cada combinación (métrica-variable categórica). La variable categórica subgrupo es el factor de interés primario en este estudio, y, las **métricas de hierro, difusividad radial y de anisotropía fraccional cumplen el criterio de homocedasticidad de varianzas para subgrupo**, así que decidí realizar el test de Kruskal Wallis para dichas métricas a pesar de que no se cumple el criterio de aleatoriedad. Para las métricas de difusividad axial y difusividad media, no se generan los tests de Kruskal Wallis dado que presentan heterocedasticidad de varianzas para subgrupo. Si uno decidiera ser estadísticamente estricto, **no se deberían generar los tests de Kruskal Wallis dado que no se cumplen los criterios de aleatoriedad**.

La función `kruskal.test` del paquete `stats` se utiliza para realizar el test de Kruskal-Wallis, el cual evalúa si existen diferencias entre grupos en las distribuciones en función de los parámetros de localización (Cómo realizar fácilmente una prueba de Kruskal-Wallis en R).

- $H_0$ : las medianas de todos los grupos son iguales;
- $H_1$ : desigualdad entre las medianas (al menos la mediana de un grupo es diferente de las demás).

```
kw_iron<-kruskal.test(ATR_Iron~ATR_subgroup, data=ATR_DifIron_Df);kw_iron

Kruskal-Wallis rank sum test

data: ATR_Iron by ATR_subgroup
Kruskal-Wallis chi-squared = 2.1092, df = 2, p-value = 0.3483
kw_rd<-kruskal.test(ATR_RD~ATR_subgroup, data=ATR_DifIron_Df);kw_rd
```

```

Kruskal-Wallis rank sum test

data: ATR_RD by ATR_subgroup
Kruskal-Wallis chi-squared = 24.454, df = 2, p-value = 4.896e-06
kw_fa<-kruskal.test(ATR_FA~ATR_subgroup, data=ATR_DifIron_Df);kw_fa

Kruskal-Wallis rank sum test

data: ATR_FA by ATR_subgroup
Kruskal-Wallis chi-squared = 6.3937, df = 2, p-value = 0.04089

```

Cuadro 86: Tests de Kruskal Wallis para las métricas de hierro, difusividad radial y anisotropía fraccional en relación a la variable independiente subgrupo

	Iron	Radial Diffusivity	Fractional Anisotropy
Kruskal-Wallis chi-squared	2.109	24.454	6.394
df	2	2	2
p-value	0.348	4.9e-06	0.041

Como se evidencia en la tabla, **existen diferencias estadísticamente significativas entre subgrupos para las métricas de difusividad radial y anisotropía fraccional.**

Se ejecuta un test post-hoc con la función `pairwise.wilcox.test` del paquete `stats` para determinar entre qué pares de subgrupos radican las diferencias observadas para difusividad radial y anisotropía fraccional. Se especifica `p.adjust.method='fdr'` como método de ajuste de los p-valores por comparaciones múltiples.

```

PW_RD<-pairwise.wilcox.test(x=ATR_DifIron_Df$ATR_RD,
                               g=ATR_DifIron_Df$ATR_subgroup,
                               p.adjust.method = "fdr")
PW_FA<-pairwise.wilcox.test(x=ATR_DifIron_Df$ATR_FA,
                               g=ATR_DifIron_Df$ATR_subgroup,
                               p.adjust.method = "fdr")

```

Cuadro 87: Comparaciones Post-hoc con los p-valores ajustados mediante fdr para las métricas de difusividad radial y anisotropía fraccional en relación a la variable independiente subgrupo

Comparison	Radial Diffusivity	Fractional Anisotropy
1-2	0.17359	0.04552
1-3	4e-05	0.60508
2-3	2e-05	0.04552

Las **diferencias entre subgrupos en la métrica RD de ambos ATRs radica en la comparativa entre HD vs controles y entre HD vs preHD** (con un valor aumentado de RD para HD en relación a preHD y controles).

Las **diferencias entre subgrupos en la métrica FA de ambos ATRs radica en la comparativa entre preHD vs controles y entre preHD vs HD** (con un valor aumentado de FA para preHD en relación a HD y controles), a diferencia de los resultados obtenidos con ANOVA-post hoc Tukey en que sólo se detectaron diferencias para FA entre preHD vs HD.

## 5.5 ANOVA PERMUTACIONAL

Se consultan las siguientes webs: `aovp`: Fitting and testing ANOVA using permutation tests, `lmp`: Fitting and testing linear models with permutation tests., `Permutation` tests in R.

Los **tests de permutación** (también llamadas **tests de aleatorización o de re-aleatorización**) pueden resultar especialmente útiles con datos obtenidos a partir de distribuciones desconocidas (especialmente no normales), con tamaños muestrales pequeños (que no permiten confiar en resultados asintóticos) o con la existencia de valores atípicos. R tiene dos paquetes potentes para generar tests de permutación, el paquete **coin** y el paquete **lmPerm**. El **paquete lmPerm**, particularmente fácil de implementar, proporciona tests de permutación para modelos lineales, pudiéndose utilizar para todo tipo de diseños ANOVA/ANCOVA, así como para regresión simple, polinómica y múltiple, gracias a las funciones **aovp()** y **lmp()**, respectivamente.

La función **aovp** del paquete **lmPerm** (Package‘lmPerm’) es una extensión del modelo ANOVA tradicional, cuyo objetivo principal es ajustar y testear modelos ANOVA utilizando tests de permutación en lugar de tests normales/paramétricos habituales y así las pruebas de hipótesis están basadas en permutaciones, siendo una buena alternativa al ANOVA con la violación de los supuestos de normalidad u homocedasticidad. La función **aovp** ajusta el modelo ANOVA llamando a **lmp** (*linear model permutation based*, modelo lineal basado en tests de permutación) para cada estrato del factor independiente.

El **uso de dicha función** es el siguiente:

```
aovp(formula, data = NULL, perm = "Exact", seqs = FALSE, center = TRUE, projections = FALSE, qr = TRUE, contrasts = NULL, ...)
```

Los argumentos de la función son los siguientes:

- **formula**: especificación de la fórmula del modelo;
- **data**: el dataframe que contiene las variables especificadas en la fórmula;
- **perm**: ‘Exact’, ‘Prob’ o ‘SPR’ para calcular las probabilidades de permutación, o ‘ ’ para producir probabilidades de F-test.
- **seqs**: TRUE para calcular SS (*Sum of Squares*) secuencial (type I SS), FALSE para calcular un único SS (type III SS);
- **center**: TRUE para centrar las variables numéricas;
- **projections**: valor lógico (TRUE/FALSE) para determinar si se obtienen o no las proyecciones;
- **qr**: valor lógico (TRUE/FALSE) para determinar si se obtiene o no la descomposición QR;
- **contrasts**: lista de contrastes para algunos de los factores de la fórmula (no se utilizan para términos de error ya que si se proporcionan contrastes para factores presentes sólo en el término de error, R generará un *warning*).
- **...**: argumentos adicionales a pasar a la función **aovp**, siendo iguales que para la función **lmp**.

Importante saber qué tipo de cálculo se va a utilizar para el SS: el tipo I es secuencial por lo que el orden de los factores (variables independientes) es importante (primero se calcula el SS del primer factor, segundo se calcula el SS del segundo factor y así sucesivamente), y, el tipo III es único por lo que el orden de los factores no es importante (cada efecto se ajusta a todos los demás efectos).

El **resultado de aovp** es similar al de **aov**, pero con p-valores basados en permutación en lugar de p-valores derivados de la teoría normal.

El **modelo asumido por aovp** es  $Y = Xb + Zg + e$ , donde **X** es la matriz de incidencia de efectos fijos y **Z** es la matriz de incidencia de efectos aleatorios (estratos de error) con columnas que representan los distintos estratos de error. El **algoritmo aovp()** proyecta Y en estratos de modo que cada estrato definido por bloques completos tenga un único término de error. X también se proyecta de modo que el modelo en este estrato sea  $P(Y) = P(X)bi + ei$ . El vector  $b_i$  se divide en fuentes con **dfj** grados de libertad para la j-ésima fuente, y **summary(aovp())** producirá una tabla ANOVA para estas fuentes en el i-ésimo estrato. Se generarán los p-valores del test permutación o del test F en función de la especificación en el argumento **perm**.

Sobre el tipo de SS a elegir, se consulta la web Anova – Type I/II/III SS explained. Con **datos desbalanceados**, existen **3 formas diferentes de calcular las sumas de cuadrados** para ANOVA (tipo I, II y III) en función de las hipótesis a testear:

- El **tipo I** testea el efecto principal del primer factor (A) sin controlar por el resto de factores, seguido del efecto principal del segundo factor (B) después del efecto principal de A (también sin controlar por el resto de factores), seguido del efecto de interacción AB después de los efectos principales. Debido a la naturaleza

secuencial y al hecho de que los dos factores principales se testean en un orden particular, este tipo de sumas de cuadrados dará resultados diferentes para datos no balanceados dependiendo de qué efecto principal se considere primero. Así pues, para datos no balanceados, este enfoque testea una diferencia en las medias marginales ponderadas. En términos prácticos, esto significa que los resultados dependen de los tamaños de muestra obtenidos, es decir, de las proporciones en el conjunto de datos particular. Por tanto, este tipo de SS no suele ser la hipótesis que interesa cuando se trabaja con datos desequilibrados.

- El **tipo II** testea cada efecto principal después de controlar por el resto de efectos principales, suponiendo que ninguna interacción es significativa, por tanto, primero se debe testear que las interacciones no sean significativas para poder continuar con el análisis de los efectos principales. Si realmente no hay interacción, el tipo II es estadísticamente más poderoso que el tipo III. Computacionalmente, esto equivale a ejecutar un análisis de tipo I con diferentes órdenes de los factores y obtener el resultado apropiado.
- El **tipo III** testea la presencia de un efecto principal después de controlar por el otro efecto principal e interacción, por tanto, este enfoque es válido en presencia de interacciones significativas. Si las interacciones no son significativas, el tipo II ofrece un test más potente. Importante destacar que no es interesante interpretar un efecto principal si hay interacciones presentes (en términos generales, si hay una interacción significativa, los efectos principales no deberían analizarse).

Por último, si los **datos están balanceados**, los factores son ortogonales y los tipos I, II y III dan todos los mismos resultados.

**Por lo general, la hipótesis de interés trata sobre la importancia de un factor mientras se controla el nivel de los otros factores** y esto equivale a utilizar SS de tipo II o III. En general, si no hay un efecto de interacción significativo, entonces el tipo II es más potente y sigue el principio de marginalidad. Si hay interacción, entonces el tipo II es inapropiado, mientras que el tipo III aún puede usarse, pero los resultados deben interpretarse con precaución (en presencia de interacciones, los efectos principales rara vez son interpretables).

Como se hizo en el apartado del ANOVA, dado que los modelos MANOVA y PERMANOVA han detectado diferencias entre subgrupos en relación a las 5 métricas de RM, **se procede a generar los ANOVA permutacionales para valorar cada métrica por separado**. En este caso, los ANOVA permutacionales también se van a generar con 4 factores: subgrupo, hemisferio, edad factorizada y género. Se empezará con un **modelo con interacción de los 4 factores evaluado con un SS del tipo III**; posteriormente se irán eliminando los términos (interacciones) hasta llegar al **modelo más simple evaluado con un SS del tipo I**. Remarcar que tanto el hemisferio, la edad como el género son covariables que me interesa mantener en el modelo final aunque no sean significativas para eliminar su influencia en la relación entre métricas-subgrupos.

### 5.5.1 IRON

```
# Establecer seed para que el resultado de las permutaciones sea siempre el mismo
set.seed(123)

library(lmPerm)
anovaP_iron1<-aovp(ATR_Iron~ATR_subgroup*ATR_hemisphere*ATR_agef2*ATR_Gender,
                     data=ATR_DifIron_Df, perm="", seqs=F)
[1] "Settings: unique SS"
summary(anovaP_iron1)

ATR_subgroup          Df  Sum Sq  Mean Sq F value
2 1.060e-06 5.320e-07  0.193
ATR_hemisphere        1 3.220e-06 3.222e-06  1.173
ATR_subgroup:ATR_hemisphere 2 1.560e-06 7.800e-07  0.284
ATR_agef2              2 1.156e-05 5.780e-06  2.103
ATR_subgroup:ATR_agef2 4 2.287e-05 5.719e-06  2.081
ATR_hemisphere:ATR_agef2 2 1.172e-05 5.862e-06  2.133
ATR_subgroup:ATR_hemisphere:ATR_agef2 4 3.370e-06 8.430e-07  0.307
ATR_Gender              1 1.418e-05 1.418e-05  5.160
ATR_subgroup:ATR_Gender 2 5.620e-06 2.810e-06  1.023
ATR_hemisphere:ATR_Gender 1 1.670e-06 1.672e-06  0.609
```

```

ATR_subgroup:ATR_hemisphere:ATR_Gender          2 2.840e-06 1.420e-06 0.517
ATR_agef2:ATR_Gender                           2 2.638e-05 1.319e-05 4.800
ATR_subgroup:ATR_agef2:ATR_Gender              2 1.516e-05 7.581e-06 2.759
ATR_hemisphere:ATR_agef2:ATR_Gender           2 4.080e-06 2.040e-06 0.742
ATR_subgroup:ATR_hemisphere:ATR_agef2:ATR_Gender 2 4.800e-06 2.401e-06 0.874
Residuals                                         78 2.143e-04 2.748e-06
Pr(>F)
0.8245
0.2822
0.7535
0.1289
0.0912 .
0.1253
0.8726
0.0259 *
0.3644
0.4377
0.5984
0.0108 *
0.0696 .
0.4793
0.4215
Residuals
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

En el modelo `anovaP_iron1` generado con interacciones y con SS del tipo III, se evidencia que la interacción edad factorizada-género es significativa. En el siguiente modelo, se mantiene dicha interacción y se calcula la SS de tipo I para cada factor.

```

set.seed(123)
anovaP_iron2<-aovp(ATR_Iron~ATR_subgroup+ATR_hemisphere+ATR_agef2*ATR_Gender,
                     data=ATR_DifIron_Df, perm="", seqs=T)
[1] "Settings: sequential SS"
summary(anovaP_iron2)
      Df   Sum Sq  Mean Sq F value Pr(>F)
ATR_subgroup    2 9.490e-06 4.744e-06  1.771 0.1753
ATR_hemisphere  1 2.460e-05 2.460e-05  9.185 0.0031 **
ATR_agef2       2 9.480e-06 4.738e-06  1.769 0.1757
ATR_Gender       1 9.600e-07 9.550e-07  0.357 0.5517
ATR_agef2:ATR_Gender 2 2.345e-05 1.173e-05  4.379 0.0150 *
Residuals      101 2.705e-04 2.678e-06
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Con este modelo `anovaP_iron2`, los términos significativos son el factor hemisferio y la interacción edad factorizada-género.

### 5.5.2 AD

```

set.seed(123)
anovaP_ad1<-aovp(ATR_AD~ATR_subgroup*ATR_hemisphere*ATR_agef2*ATR_Gender,
                   data=ATR_DifIron_Df, perm="", seqs=F)

```

```
[1] "Settings: unique SS "
summary(anovaP_ad1)

ATR_subgroup
ATR_hemisphere
ATR_subgroup:ATR_hemisphere
ATR_agef2
ATR_subgroup:ATR_agef2
ATR_hemisphere:ATR_agef2
ATR_subgroup:ATR_hemisphere:ATR_agef2
ATR_Gender
ATR_subgroup:ATR_Gender
ATR_hemisphere:ATR_Gender
ATR_subgroup:ATR_hemisphere:ATR_Gender
ATR_agef2:ATR_Gender
ATR_subgroup:ATR_agef2:ATR_Gender
ATR_hemisphere:ATR_agef2:ATR_Gender
ATR_subgroup:ATR_hemisphere:ATR_agef2:ATR_Gender
Residuals
Df      Sum Sq   Mean Sq F value
2 4.129e-08 2.065e-08 13.301
1 2.500e-09 2.503e-09  1.612
2 6.100e-10 3.050e-10  0.197
2 4.120e-09 2.060e-09  1.327
4 9.570e-09 2.391e-09  1.541
2 2.700e-10 1.350e-10  0.087
4 6.600e-10 1.640e-10  0.106
1 1.880e-09 1.881e-09  1.212
2 1.000e-09 5.000e-10  0.322
1 1.400e-10 1.410e-10  0.091
2 1.000e-10 4.800e-11  0.031
2 9.270e-09 4.637e-09  2.987
2 2.830e-09 1.417e-09  0.913
2 3.300e-10 1.670e-10  0.108
2 5.400e-10 2.710e-10  0.175
78 1.211e-07 1.552e-09
Pr(>F)
1.07e-05 ***
0.2079
0.8218
0.2711
0.1987
0.9167
0.9802
0.2743
0.7255
0.7638
0.9696
0.0562 .
0.4055
0.8982
0.8402
Residuals
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

En el modelo `anovaP_ad1` generado con interacciones y con SS del tipo III, no se evidencia ninguna interacción significativa (la interacción edad factorizada-género es casi significativa con un p-valor de 0.056). En el siguiente modelo, se eliminan todas las interacciones y se calcula la SS de tipo I para cada factor.

```
set.seed(123)
anovaP_ad2<-aovp(ATR_AD~ATR_subgroup+ATR_hemisphere+ATR_agef2+ATR_Gender,
                   data=ATR_DifIron_Df, perm="", seqs=T)
[1] "Settings: sequential SS "
summary(anovaP_ad2)

Df      Sum Sq   Mean Sq F value    Pr(>F)
ATR_subgroup  2 3.863e-08 1.932e-08 13.762 5.05e-06 ***
ATR_hemisphere 1 9.860e-09 9.864e-09  7.028  0.00929 **
ATR_agef2      2 6.370e-09 3.184e-09  2.269  0.10858
ATR_Gender      1 1.950e-09 1.950e-09  1.389  0.24128
Residuals     103 1.446e-07 1.404e-09
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Con este modelo `anovaP_md2`, los términos significativos son los factores subgrupo y hemisferio.

### 5.5.3 MD

```
set.seed(123)
anovaP_md1<-aovp(ATR_MD~ATR_subgroup*ATR_hemisphere*ATR_agef2*ATR_Gender,
                    data=ATR_DifIron_Df, perm="", seqs=F)
[1] "Settings: unique SS "
summary(anovaP_md1)

Df      Sum Sq   Mean Sq F value
ATR_subgroup                                2 2.070e-08 1.035e-08 12.832
ATR_hemisphere                               1 1.840e-09 1.844e-09  2.287
ATR_subgroup:ATR_hemisphere                 2 2.400e-10 1.210e-10  0.150
ATR_agef2                                    2 1.130e-09 5.670e-10  0.703
ATR_subgroup:ATR_agef2                      4 3.220e-09 8.050e-10  0.998
ATR_hemisphere:ATR_agef2                   2 9.000e-11 4.400e-11  0.054
ATR_subgroup:ATR_hemisphere:ATR_agef2       4 3.200e-10 8.000e-11  0.100
ATR_Gender                                    1 1.140e-09 1.145e-09  1.419
ATR_subgroup:ATR_Gender                     2 7.400e-10 3.690e-10  0.458
ATR_hemisphere:ATR_Gender                  1 1.000e-11 1.300e-11  0.016
ATR_subgroup:ATR_hemisphere:ATR_Gender     2 2.000e-11 9.000e-12  0.011
ATR_agef2:ATR_Gender                        2 2.350e-09 1.174e-09  1.455
ATR_subgroup:ATR_agef2:ATR_Gender           2 8.700e-10 4.330e-10  0.537
ATR_hemisphere:ATR_agef2:ATR_Gender        2 6.000e-11 2.900e-11  0.036
ATR_subgroup:ATR_hemisphere:ATR_agef2:ATR_Gender 2 1.600e-10 8.000e-11  0.099
Residuals                                     78 6.291e-08 8.070e-10

Pr(>F)
ATR_subgroup                                1.52e-05 ***
ATR_hemisphere                               0.135
ATR_subgroup:ATR_hemisphere                 0.861
ATR_agef2                                    0.498
ATR_subgroup:ATR_agef2                      0.414
ATR_hemisphere:ATR_agef2                   0.947
ATR_subgroup:ATR_hemisphere:ATR_agef2       0.982
ATR_Gender                                    0.237
ATR_subgroup:ATR_Gender                     0.634
ATR_hemisphere:ATR_Gender                  0.901
ATR_subgroup:ATR_hemisphere:ATR_Gender     0.989
ATR_agef2:ATR_Gender                        0.240
ATR_subgroup:ATR_agef2:ATR_Gender           0.587
ATR_hemisphere:ATR_agef2:ATR_Gender        0.965
ATR_subgroup:ATR_hemisphere:ATR_agef2:ATR_Gender 0.906
Residuals                                     ---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

En el modelo `anovaP_md1` generado con interacciones y con SS del tipo III, no se evidencia ninguna interacción significativa. En el siguiente modelo, se eliminan todas las interacciones y se calcula la SS de tipo I para cada factor.

```
set.seed(123)
anovaP_md2<-aovp(ATR_MD~ATR_subgroup+ATR_hemisphere+ATR_agef2+ATR_Gender,
                    data=ATR_DifIron_Df, perm="", seqs=T)
[1] "Settings: sequential SS "
summary(anovaP_md2)
```

```

          Df   Sum Sq  Mean Sq F value    Pr(>F)
ATR_subgroup     2 2.834e-08 1.417e-08  20.562 3.06e-08 ***
ATR_hemisphere   1 5.290e-09 5.293e-09   7.682  0.00662 **
ATR_agef2         2 5.490e-09 2.747e-09   3.987  0.02149 *
ATR_Gender        1 6.500e-10 6.500e-10   0.944  0.33364
Residuals       103 7.097e-08 6.890e-10
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Con este modelo anovaP\_md2, los términos significativos son los factores subgrupo, hemisferio y edad factorizada.

#### 5.5.4 RD

```

set.seed(123)
anovaP_rd1<-aovp(ATR_RD~ATR_subgroup*ATR_hemisphere*ATR_agef2*ATR_Gender,
                     data=ATR_DifIron_Df, perm="", seqs=F)
[1] "Settings: unique SS"
summary(anovaP_rd1)

          Df   Sum Sq  Mean Sq F value    Pr(>F)
ATR_subgroup     2 1.314e-08 6.568e-09  7.861
ATR_hemisphere   1 1.550e-09 1.553e-09  1.859
ATR_subgroup:ATR_hemisphere 2 2.400e-10 1.220e-10  0.146
ATR_agef2         2 1.610e-09 8.050e-10  0.964
ATR_subgroup:ATR_agef2 4 2.230e-09 5.560e-10  0.666
ATR_hemisphere:ATR_agef2 2 1.400e-10 7.100e-11  0.085
ATR_subgroup:ATR_hemisphere:ATR_agef2 4 2.300e-10 5.800e-11  0.069
ATR_Gender        1 8.400e-10 8.450e-10  1.011
ATR_subgroup:ATR_Gender 2 2.490e-09 1.247e-09  1.492
ATR_hemisphere:ATR_Gender 1 0.000e+00 0.000e+00  0.000
ATR_subgroup:ATR_hemisphere:ATR_Gender 2 1.300e-10 6.300e-11  0.075
ATR_agef2:ATR_Gender 2 1.210e-09 6.040e-10  0.723
ATR_subgroup:ATR_agef2:ATR_Gender 2 1.390e-09 6.970e-10  0.834
ATR_hemisphere:ATR_agef2:ATR_Gender 2 9.000e-11 4.500e-11  0.054
ATR_subgroup:ATR_hemisphere:ATR_agef2:ATR_Gender 2 1.400e-10 6.900e-11  0.083
Residuals       78 6.517e-08 8.360e-10
Pr(>F)
ATR_subgroup 0.000776 ***
ATR_hemisphere 0.176701
ATR_subgroup:ATR_hemisphere 0.864758
ATR_agef2 0.386042
ATR_subgroup:ATR_agef2 0.617612
ATR_hemisphere:ATR_agef2 0.918612
ATR_subgroup:ATR_hemisphere:ATR_agef2 0.991137
ATR_Gender 0.317831
ATR_subgroup:ATR_Gender 0.231247
ATR_hemisphere:ATR_Gender 0.983535
ATR_subgroup:ATR_hemisphere:ATR_Gender 0.927675
ATR_agef2:ATR_Gender 0.488470
ATR_subgroup:ATR_agef2:ATR_Gender 0.438249
ATR_hemisphere:ATR_agef2:ATR_Gender 0.947789
ATR_subgroup:ATR_hemisphere:ATR_agef2:ATR_Gender 0.920570
Residuals
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

En el modelo `anovaP_rd1` generado con interacciones y con SS del tipo III, no se evidencia ninguna interacción significativa. En el siguiente modelo, se eliminan todas las interacciones y se calcula la SS de tipo I para cada factor.

```
set.seed(123)
anovaP_rd2<-aovp(ATR_RD~ATR_subgroup+ATR_hemisphere+ATR_agef2+ATR_Gender,
                    data=ATR_DifIron_Df, perm="", seqs=T)
[1] "Settings: sequential SS"
summary(anovaP_rd2)
  Df   Sum Sq  Mean Sq F value Pr(>F)
ATR_subgroup  2 2.503e-08 1.252e-08 17.795 2.3e-07 ***
ATR_hemisphere 1 3.540e-09 3.537e-09  5.029  0.0271 *
ATR_agef2      2 5.850e-09 2.926e-09  4.160  0.0183 *
ATR_Gender      1 2.600e-10 2.620e-10  0.372  0.5434
Residuals     103 7.245e-08 7.030e-10
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Con este modelo `anovaP_rd2`, los términos significativos son los factores subgrupo, hemisferio y edad factorizada.

### 5.5.5 FA

```
set.seed(123)
anovaP_fa1<-aovp(ATR_FA~ATR_subgroup*ATR_hemisphere*ATR_agef2*ATR_Gender,
                    data=ATR_DifIron_Df, perm="", seqs=F)
[1] "Settings: unique SS"
summary(anovaP_fa1)
  Df   Sum Sq  Mean Sq F value Pr(>F)
ATR_subgroup          2 0.00042 0.0002104  0.303
ATR_hemisphere        1 0.00016 0.0001649  0.238
ATR_subgroup:ATR_hemisphere  2 0.00036 0.0001786  0.257
ATR_agef2              2 0.00230 0.0011488  1.656
ATR_subgroup:ATR_agef2    4 0.00361 0.0009014  1.299
ATR_hemisphere:ATR_agef2  2 0.00034 0.0001715  0.247
ATR_subgroup:ATR_hemisphere:ATR_agef2  4 0.00014 0.0000348  0.050
ATR_Gender              1 0.00003 0.0000349  0.050
ATR_subgroup:ATR_Gender    2 0.00366 0.0018322  2.641
ATR_hemisphere:ATR_Gender  1 0.00002 0.0000180  0.026
ATR_subgroup:ATR_hemisphere:ATR_Gender  2 0.00045 0.0002237  0.322
ATR_agef2:ATR_Gender       2 0.00359 0.0017963  2.589
ATR_subgroup:ATR_agef2:ATR_Gender  2 0.00288 0.0014399  2.075
ATR_hemisphere:ATR_agef2:ATR_Gender  2 0.00023 0.0001141  0.164
ATR_subgroup:ATR_hemisphere:ATR_agef2:ATR_Gender 2 0.00029 0.0001453  0.209
Residuals                78 0.05412 0.0006938
Pr(>F)
ATR_subgroup            0.7393
ATR_hemisphere          0.6273
ATR_subgroup:ATR_hemisphere 0.7737
ATR_agef2                0.1976
ATR_subgroup:ATR_agef2    0.2777
ATR_hemisphere:ATR_agef2 0.7815
ATR_subgroup:ATR_hemisphere:ATR_agef2 0.9952
ATR_Gender                0.8230
ATR_subgroup:ATR_Gender    0.0777 .
ATR_hemisphere:ATR_Gender 0.8724
```

```

ATR_subgroup:ATR_hemisphere:ATR_Gender      0.7253
ATR_agef2:ATR_Gender                         0.0815 .
ATR_subgroup:ATR_agef2:ATR_Gender             0.1324
ATR_hemisphere:ATR_agef2:ATR_Gender          0.8487
ATR_subgroup:ATR_hemisphere:ATR_agef2:ATR_Gender 0.8116
Residuals
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

En el modelo `anovaP_fa1` generado con interacciones y con SS del tipo III, no se evidencia ninguna interacción significativa. En el siguiente modelo, se eliminan todas las interacciones y se calcula la SS de tipo I para cada factor.

```

set.seed(123)
anovaP_fa2<-aovp(ATR_FA~ATR_subgroup+ATR_hemisphere+ATR_agef2+ATR_Gender,
                   data=ATR_DifIron_Df, perm="", seqs=T)
[1] "Settings: sequential SS"
summary(anovaP_fa2)
  Df Sum Sq Mean Sq F value Pr(>F)
ATR_subgroup   2 0.00471 0.0023558  3.665  0.029 *
ATR_hemisphere 1 0.00005 0.0000510  0.079  0.779
ATR_agef2       2 0.00143 0.0007150  1.112  0.333
ATR_Gender       1 0.00001 0.0000139  0.022  0.883
Residuals      103 0.06622 0.0006429
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Con este modelo `anovaP_fa2`, el término significativo es el factor subgrupo.

**En resumen:**

- El **modelo de hierro** presenta el factor hemisferio y la interacción edad factorizada-género como términos significativos;
- El **modelo de difusividad axial** presenta los factores subgrupo y hemisferio como términos significativos;
- Los **modelo de difusividad media** y **difusividad radial** presentan los factores subgrupo, hemisferio y edad factorizada como términos significativos;
- El **modelo de anistropía fraccional** presenta el factor subgrupo como término significativo.

Remarcar que, a pesar de utilizar un test de permutaciones, no se consigue poder estadístico para detectar diferencias entre subgrupos para la métrica de hierro.

Es importante destacar que la **interacción subgrupo-hemisferio no es significativa**, por tanto, **no se puede concluir que la enfermedad de Huntington condicione una mayor degeneración de un ATR que de otro ATR**, es decir, **no se puede concluir una mayor vulnerabilidad de un hemisferio a la enfermedad de Huntington**. Así, las diferencias entre hemisferios detectadas para el hierro y para las difusividades axial y media puede considerarse inherentes a la fisiología diferencial de los hemisferios (diferentes funciones ejecutadas por cada hemisferio pueden condicionar diferencias en la estructura).

### 5.5.6 Comparaciones pareadas por subgrupos - Pairwise permutation test

A continuación se generan las comparaciones pareadas por subgrupos para las métricas de difusividad (axial, media y radial) y de anistropía fraccional.

Se va a utilizar la función `pairwisePermutationTest` del paquete `rcompanion`, habiendo consultado la web `pairwisePermutationTest: Pairwise two-sample independence tests`. Dicha función genera tests de independencia por pares entre grupos.

El uso de dicha función es el siguiente:

```
pairwisePermutationTest(formula = NULL, data = NULL, x = NULL, g = NULL, method = "fdr", ...)
```

Los argumentos de la función son los siguientes:

- **formula**: una fórmula indicando la variable de medida y la variable de agrupamiento (ejemplo,  $y \sim \text{grupo}$ );
- **data**: el dataframe a utilizar;
- **x**: la variable respuesta introducida como vector;
- **g**: la variable de agrupamiento introducida como vector;
- **method**: el método de ajuste de los p-valores por tests múltiples;
- **...**: argumentos adicionales utilizados en la función `coin::independence_test`.

En este caso, el **método de ajuste por múltiples comparaciones** que se va a utilizar es el **fdr**.

El resultado de `pairwisePermutationTest` es un dataframe que contiene los siguientes ítems:

- **Comparison**: las comparaciones realizadas entre los pares de grupos;
- **Stat**: el estadístico calculado por permutación;
- **p.value**: el p-valor sin ajustar,
- **p.adjust**: el p-valor ajustado por comparaciones múltiples.

```
library(rcompanion)
# Comparaciones pareadas por subgrupos para AD
PairPerm_ad<-pairwisePermutationTest(ATR_AD~ATR_subgroup,
                                       data=ATR_DifIron_Df, method='fdr')

# Comparaciones pareadas por subgrupos para MD
PairPerm_md<-pairwisePermutationTest(ATR_MD~ATR_subgroup,
                                       data=ATR_DifIron_Df, method='fdr')
Warning in .local(.Object, ...): The conditional covariance matrix has zero
diagonal elements

# Comparaciones pareadas por subgrupos para RD
PairPerm_rd<-pairwisePermutationTest(ATR_RD~ATR_subgroup,
                                       data=ATR_DifIron_Df, method='fdr')
Warning in .local(.Object, ...): The conditional covariance matrix has zero
diagonal elements

# Comparaciones pareadas por subgrupos para FA
PairPerm_fa<-pairwisePermutationTest(ATR_FA~ATR_subgroup,
                                       data=ATR_DifIron_Df, method='fdr')
```

Para las métricas de difusividad radial y difusividad media se evidencia el mensaje **warning** especificando **The conditional covariance matrix has zero diagonal elements**, por lo que se decide trabajar con la transformación logarítmica de dichas métricas, misma transformación que la utilizada en la ejecución del MANOVA.

```
ATR_DifIron_Log_DF<-cbind(ATR_DifIron_Df[1:11],log(ATR_DifIron_Df[, (8:10)]))
colnames(ATR_DifIron_Log_DF)<-c(colnames(ATR_DifIron_Df[1:11]),paste('Log',c(colnames(ATR_DifIron_Df[8:10]))))

# Comparaciones pareadas por subgrupos para MD con transformación logarítmica
PairPerm_mdLog<-pairwisePermutationTest(Log_ATR_MD~ATR_subgroup,
                                         data=ATR_DifIron_Log_DF, method='fdr')

# Comparaciones pareadas por subgrupos para RD
PairPerm_rdLog<-pairwisePermutationTest(Log_ATR_RD~ATR_subgroup,
                                         data=ATR_DifIron_Log_DF, method='fdr')
```

Cuadro 88: Pairwise Permutation Test Results

Comparison	Stat	p.value	p.adjust
<b>AD</b>			
1 - 2 = 0	-0.7292	0.4659	0.4659000
1 - 3 = 0	-4.09	4.312e-05	0.0001294
2 - 3 = 0	-3.075	0.002103	0.0031540
<b>LogMD</b>			
1 - 2 = 0	0.7344	0.4627	0.4627000
1 - 3 = 0	-4.397	1.095e-05	0.0000328
2 - 3 = 0	-4.001	6.314e-05	0.0000947
<b>LogRD</b>			
1 - 2 = 0	1.596	0.1105	0.1105000
1 - 3 = 0	-3.92	8.859e-05	0.0001329
2 - 3 = 0	-4.064	4.831e-05	0.0001329
<b>FA</b>			
1 - 2 = 0	-2.334	0.01959	0.0293800
1 - 3 = 0	0.7105	0.4774	0.4774000
2 - 3 = 0	2.42	0.01553	0.0293800

Las diferencias entre subgrupos en la métricas de difusividad axial, media (con transformación logarítmica) y radial (con transformación logarítmica) de ambos ATRs radica en la comparativa entre HD vs controles y entre HD vs preHD (con un valor aumentado de AD, RD y MD para HD en relación a preHD y controles).

Las diferencias entre subgrupos en la métrica FA de ambos ATRs radica en la comparativa entre preHD vs controles y entre preHD vs HD (con un valor aumentado de FA para preHD en relación a HD y controles), resultado superponible al obtenido con Kruskal-Wallis y diferenciando del obtenido con ANOVA-post hoc Tukey en el que tan sólo se detectaron diferencias para FA entre preHD vs HD.

## 6 Regresión logística

### 6.1 ONE vs REST

En la regresión logística, se van a utilizar los dataframes con todos los sujetos dado que se quiere observar el poder de clasificación de los biomarcadores (métricas de RM) en poblaciones heterogéneas para así poder extrapolar los resultados a la clínica diaria.

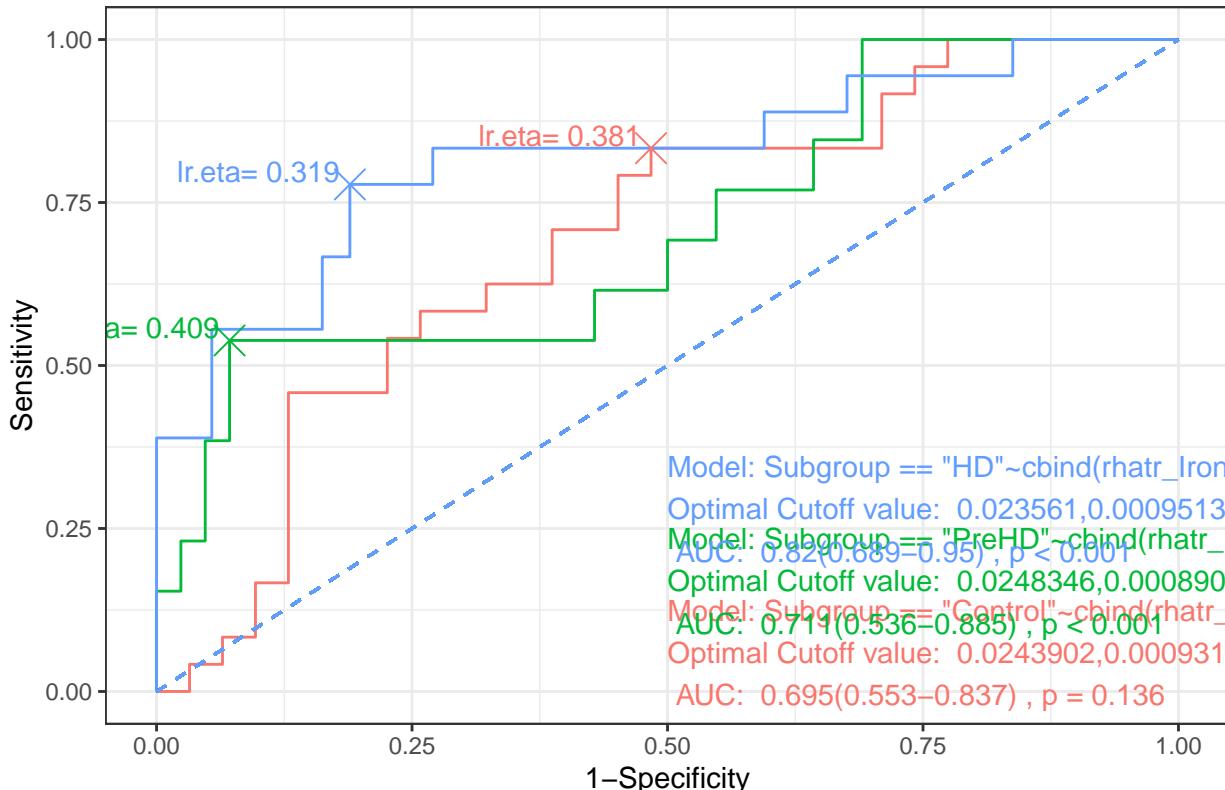
Se utiliza la función `multipleROC` del paquete `multipleROC`.

#### 6.1.1 RATR

```
# ROC curves one vs the rest by multipleROC function from multipleROC package
library(multipleROC)
#multipleROC(Subgroup=='Control'~cbind(rhatr_Iron,rhatr_AD,rhatr_RD,rhatr_MD,rhatr_FA),
#             data=RATR_avg_demo, plot=TRUE)
#multipleROC(Subgroup=='PreHD'~cbind(rhatr_Iron,rhatr_AD,rhatr_RD,rhatr_MD,rhatr_FA),
#             data=RATR_avg_demo, plot=TRUE)
#multipleROC(Subgroup=='HD'~cbind(rhatr_Iron,rhatr_AD,rhatr_RD,rhatr_MD,rhatr_FA),
#             data=RATR_avg_demo, plot=TRUE)
#Global_ROC_RATR5<-multipleROC(Subgroup~cbind(rhatr_Iron,rhatr_AD,rhatr_RD,rhatr_MD,rhatr_FA),
#                                   data=RATR_avg_demo, plot=TRUE)

Global_ROC_RATR5<-multipleROC(Subgroup~cbind(rhatr_Iron,rhatr_AD,rhatr_RD,rhatr_MD,rhatr_FA),
                                  data=RATR_avg_demo, plot=FALSE)
CT_ROC_RATR5<-multipleROC(Subgroup=='Control'~cbind(rhatr_Iron,rhatr_AD,rhatr_RD,rhatr_MD,
                                                       rhatr_FA),
                            data=RATR_avg_demo, plot=FALSE)
PH_ROC_RATR5<-multipleROC(Subgroup=='PreHD'~cbind(rhatr_Iron,rhatr_AD,rhatr_RD,rhatr_MD,
                                                       rhatr_FA), data=RATR_avg_demo, plot=FALSE)
HD_ROC_RATR5<-multipleROC(Subgroup=='HD'~cbind(rhatr_Iron,rhatr_AD,rhatr_RD,rhatr_MD,rhatr_FA),
                           data=RATR_avg_demo, plot=FALSE)
plot_ROC(list(CT_ROC_RATR5,PH_ROC_RATR5,HD_ROC_RATR5),show.eta = TRUE,show.sens = FALSE) +
  labs(title= "Log Regression of RATR")
```

## Log Regression of RATR



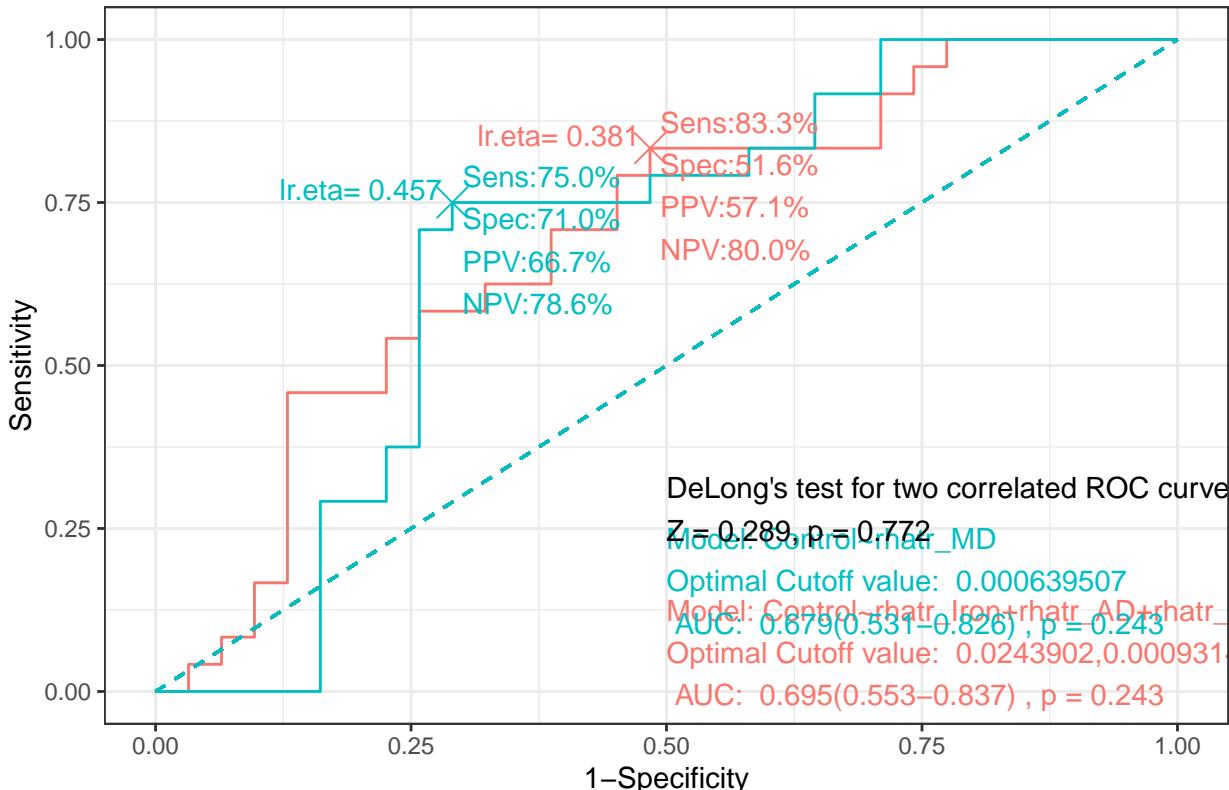
```
#plot_ROC(list(CT_ROC_RATR5,PH_ROC_RATR5,HD_ROC_RATR5),show.eta = TRUE,show.sens = TRUE)+  
# facet_grid(~.) + labs(title= "Log Regression of RATR")  
cutoff_RATR5<-rbind(CT_ROC_RATR5$cutoff[1,1],PH_ROC_RATR5$cutoff[1,1],HD_ROC_RATR5$cutoff[1,1])  
rownames(cutoff_RATR5)<-c('Control','PreHD','HD')  
kableExtra::kable(cutoff_RATR5,  
                  caption = 'Cutoff values of MRI metrics for subgroups in the 5-metrics LOGISTIC REGRESSION model of RATR',  
                  booktabs = TRUE) %>%  
kable_styling(position = 'center', font_size = 8, latex_options='hold_position')
```

Cuadro 89: Cutoff values of MRI metrics for subgroups in the 5-metrics LOGISTIC REGRESSION model of RATR

	rhatr_Iron	rhatr_AD	rhatr_RD	rhatr_MD	rhatr_FA
Control	0.0243902	0.0009315	0.0004762	0.0006279	0.419371
PreHD	0.0248346	0.0008905	0.0004923	0.0006250	0.377134
HD	0.0235610	0.0009513	0.0005148	0.0006603	0.380831

```
# STEP ROC curves one vs the rest by step_ROC function from multipleROC package  
## CONTROLS  
RATR_avg_demo$Control<-ifelse(RATR_avg_demo$Subgroup=='Control',1,0)  
step_ROC(Control~rhatr_Iron+rhatr_AD+rhatr_RD+rhatr_MD+rhatr_FA,data=RATR_avg_demo)+  
labs(title= "Step Log Regression of RATR for Controls")
```

## Step Log Regression of RATR for Controls



```
step_ROC(Control~rhatr_Iron+rhatr_AD+rhatr_RD+rhatr_MD+rhatr_FA, data=RATR_avg_demo,
          plot=FALSE)
```

Analysis of Deviance Table

```
Model 1: Control ~ rhatr_MD
Model 2: Control ~ rhatr_Iron + rhatr_AD + rhatr_RD + rhatr_MD + rhatr_FA
Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      53    71.409
2      49   68.989  4    2.4197  0.6591
```

### This has the same results

```
fitLogRegRATR_CT<-glm(Control~rhatr_Iron+rhatr_AD+rhatr_RD+rhatr_MD+rhatr_FA,
                         data=RATR_avg_demo, family=binomial)
summary(fitLogRegRATR_CT)
```

Call:

```
glm(formula = Control ~ rhatr_Iron + rhatr_AD + rhatr_RD + rhatr_MD +
    rhatr_FA, family = binomial, data = RATR_avg_demo)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	3.758e+01	2.834e+01	1.326	0.185
rhatr_Iron	-2.925e+02	2.421e+02	-1.208	0.227
rhatr_AD	9.016e+07	1.813e+08	0.497	0.619
rhatr_RD	1.803e+08	3.625e+08	0.497	0.619
rhatr_MD	-2.705e+08	5.438e+08	-0.497	0.619
rhatr_FA	-3.330e+01	6.065e+01	-0.549	0.583

(Dispersion parameter for binomial family taken to be 1)

```

Null deviance: 75.353 on 54 degrees of freedom
Residual deviance: 68.989 on 49 degrees of freedom
AIC: 80.989

Number of Fisher Scoring iterations: 4
finalLogRegRATR_CT=step(fitLogRegRATR_CT, trace=0)
summary(finalLogRegRATR_CT)

Call:
glm(formula = Control ~ rhatr_MD, family = binomial, data = RATR_avg_demo)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   11.15     6.11   1.824   0.0681 .
rhatr_MD   -17700.30  9497.68  -1.864   0.0624 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 75.353 on 54 degrees of freedom
Residual deviance: 71.409 on 53 degrees of freedom
AIC: 75.409

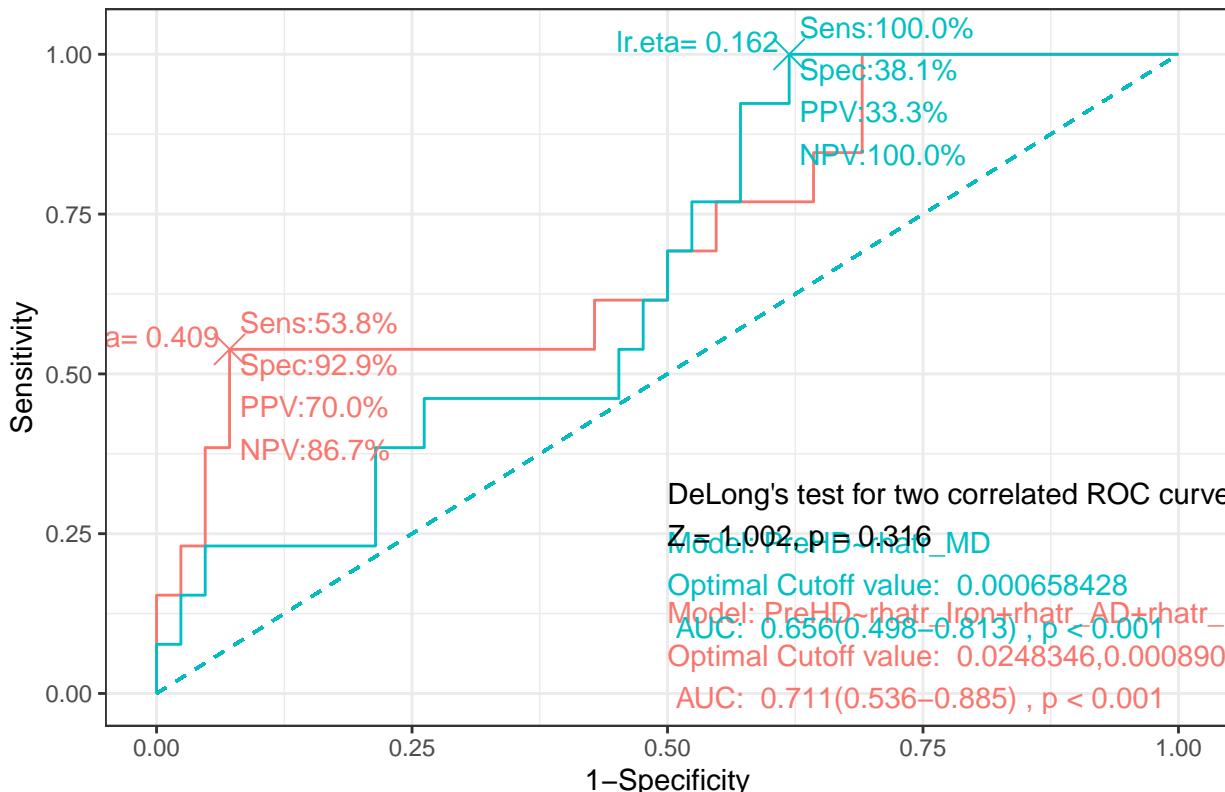
Number of Fisher Scoring iterations: 4
anova(finalLogRegRATR_CT, fitLogRegRATR_CT, test='Chisq')
Analysis of Deviance Table

Model 1: Control ~ rhatr_MD
Model 2: Control ~ rhatr_Iron + rhatr_AD + rhatr_RD + rhatr_MD + rhatr_FA
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1       53      71.409
2       49      68.989  4    2.4197   0.6591

## PREHD
RATR_avg_demo$PreHD<-ifelse(RATR_avg_demo$Subgroup=='PreHD',1,0)
step_ROC(PreHD~rhatr_Iron+rhatr_AD+rhatr_RD+rhatr_MD+rhatr_FA,data=RATR_avg_demo) +
  labs(title= "Step Log Regression of RATR for PreHD")

```

## Step Log Regression of RATR for PreHD



```
step_ROC(PreHD~rhatr_Iron+rhatr_AD+rhatr_RD+rhatr_MD+rhatr_FA, data=RATR_avg_demo, plot=FALSE)
Analysis of Deviance Table

Model 1: PreHD ~ rhatr_MD
Model 2: PreHD ~ rhatr_Iron + rhatr_AD + rhatr_RD + rhatr_MD + rhatr_FA
Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      53      55.196
2      49      52.027  4   3.1695  0.5299
### This has the same results
fitLogRegRATR_PH<-glm(PreHD~rhatr_Iron+rhatr_AD+rhatr_RD+rhatr_MD+rhatr_FA, data=RATR_avg_demo, family=binomial)
summary(fitLogRegRATR_PH)

Call:
glm(formula = PreHD ~ rhatr_Iron + rhatr_AD + rhatr_RD + rhatr_MD +
    rhatr_FA, family = binomial, data = RATR_avg_demo)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) 1.637e+01 3.459e+01  0.473   0.636
rhatr_Iron  4.007e+02 3.201e+02  1.252   0.211
rhatr_AD   -5.479e+07 2.355e+08 -0.233   0.816
rhatr_RD   -1.097e+08 4.711e+08 -0.233   0.816
rhatr_MD   1.645e+08 7.066e+08  0.233   0.816
rhatr_FA   -4.434e+01 7.124e+01 -0.622   0.534

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 60.154 on 54 degrees of freedom
```

```

Residual deviance: 52.027 on 49 degrees of freedom
AIC: 64.027

Number of Fisher Scoring iterations: 5
finalLogRegRATR_PH=step(fitLogRegRATR_PH, trace=0)
summary(finalLogRegRATR_PH)

Call:
glm(formula = PreHD ~ rhatr_RD, family = binomial, data = RATR_avg_demo)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) 12.043     6.063   1.986   0.0470 *
rhatr_RD   -26293.818 12173.386  -2.160   0.0308 *
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 60.154 on 54 degrees of freedom
Residual deviance: 54.594 on 53 degrees of freedom
AIC: 58.594

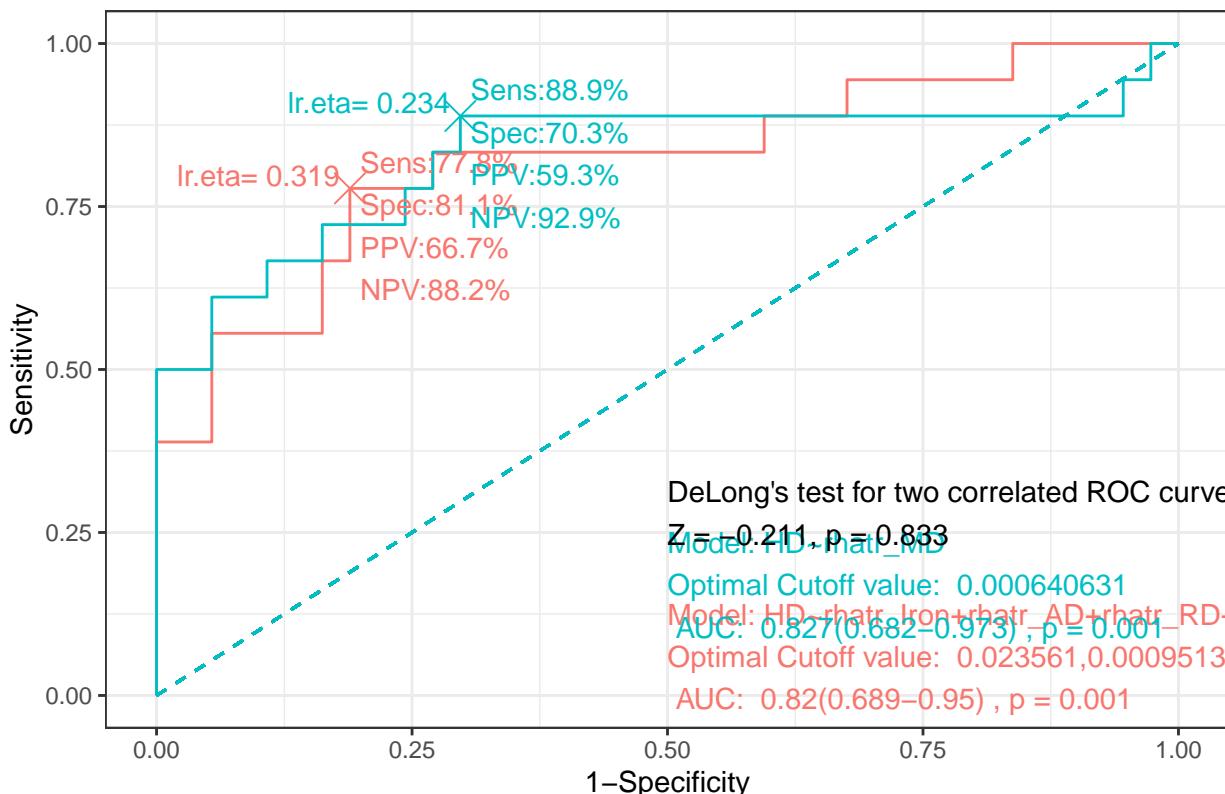
Number of Fisher Scoring iterations: 4
anova(finalLogRegRATR_PH, fitLogRegRATR_PH, test='Chisq')
Analysis of Deviance Table

Model 1: PreHD ~ rhatr_RD
Model 2: PreHD ~ rhatr_Iron + rhatr_AD + rhatr_RD + rhatr_MD + rhatr_FA
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1         53      54.594
2         49      52.027  4    2.5673  0.6326

## HD
RATR_avg_demo$HD<-ifelse(RATR_avg_demo$Subgroup=='HD',1,0)
step_ROC(HD~rhatr_Iron+rhatr_AD+rhatr_RD+rhatr_MD+rhatr_FA, data=RATR_avg_demo) +
  labs(title= "Step Log Regression of RATR for HD")

```

## Step Log Regression of RATR for HD



```
step_ROC(HD ~ rhatr_Iron + rhatr_AD + rhatr_RD + rhatr_MD + rhatr_FA, data=RATR_avg_demo, plot=FALSE)
Analysis of Deviance Table
```

```
Model 1: HD ~ rhatr_MD
```

```
Model 2: HD ~ rhatr_Iron + rhatr_AD + rhatr_RD + rhatr_MD + rhatr_FA
```

```
Resid. Df Resid. Dev Df Deviance Pr(>Chi)
```

```
1      53      51.843
2      49      50.592  4   1.2503   0.8698
```

```
### This has the same results
```

```
fitLogRegRATR_HD <- glm(HD ~ rhatr_Iron + rhatr_AD + rhatr_RD + rhatr_MD + rhatr_FA,
                           data=RATR_avg_demo, family=binomial)
```

```
summary(fitLogRegRATR_HD)
```

```
Call:
```

```
glm(formula = HD ~ rhatr_Iron + rhatr_AD + rhatr_RD + rhatr_MD +
    rhatr_FA, family = binomial, data = RATR_avg_demo)
```

```
Coefficients:
```

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-5.857e+01	3.911e+01	-1.498	0.134
rhatr_Iron	-1.008e+02	3.285e+02	-0.307	0.759
rhatr_AD	-1.319e+08	2.247e+08	-0.587	0.557
rhatr_RD	-2.637e+08	4.493e+08	-0.587	0.557
rhatr_MD	3.957e+08	6.740e+08	0.587	0.557
rhatr_FA	7.557e+01	8.355e+01	0.904	0.366

```
(Dispersion parameter for binomial family taken to be 1)
```

```

Null deviance: 69.545 on 54 degrees of freedom
Residual deviance: 50.592 on 49 degrees of freedom
AIC: 62.592

Number of Fisher Scoring iterations: 5
finalLogRegRATR_HD=step(fitLogRegRATR_HD, trace=0)
summary(finalLogRegRATR_HD)

Call:
glm(formula = HD ~ rhatr_MD, family = binomial, data = RATR_avg_demo)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -31.10      9.14  -3.403 0.000667 ***
rhatr_MD    46690.81 13991.52   3.337 0.000847 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 69.545 on 54 degrees of freedom
Residual deviance: 51.843 on 53 degrees of freedom
AIC: 55.843

Number of Fisher Scoring iterations: 5
anova(finalLogRegRATR_HD, fitLogRegRATR_HD, test='Chisq')
Analysis of Deviance Table

Model 1: HD ~ rhatr_MD
Model 2: HD ~ rhatr_Iron + rhatr_AD + rhatr_RD + rhatr_MD + rhatr_FA
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1       53     51.843
2       49     50.592  4    1.2503  0.8698

```

Cuadro 90: Binomial Logistic Regression for Controls vs Rest in Right ATR

	Estimate	Std.Error	z value	p value
<b>Model 2: Control ~ rhatr_Iron + rhatr_AD + rhatr_RD + rhatr_MD + rhatr_FA</b>				
(Intercept)	3.758020e+01	2.833601e+01	1.3262348	0.1847619
rhatr_Iron	-2.925063e+02	2.420719e+02	-1.2083447	0.2269147
rhatr_AD	9.015858e+07	1.812664e+08	0.4973817	0.6189199
rhatr_RD	1.802765e+08	3.625131e+08	0.4972966	0.6189799
rhatr_MD	-2.704721e+08	5.437848e+08	-0.4973882	0.6189154
rhatr_FA	-3.329527e+01	6.065375e+01	-0.5489400	0.5830466
<b>Model 1: Control ~ rhatr_MD</b>				
(Intercept)	1.114648e+01	6.110217e+00	1.8242363	0.0681164
rhatr_MD	-1.770030e+04	9.497677e+03	-1.8636452	0.0623715

Cuadro 91: Comparison of two Binomial Logistic Regressions for Controls vs Rest in Right ATR

	Resid. Df	Resid. Dev	Resid. Df	Deviance	Pr(>Chi)
Model 1: Control ~ rhatr_MD	53	71.40887	NA	NA	NA

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
Model 2: Control ~ rhatr_Iron + rhatr_AD + rhatr_RD + rhatr_MD + rhatr_FA	49	68.98919	4	2.419681	0.6590735

Cuadro 92: Binomial Logistic Regression for PreHD vs Rest in Right ATR

	Estimate	Std.Error	z value	p value
<b>Model 2: PreHD ~ rhatr_Iron + rhatr_AD + rhatr_RD + rhatr_MD + rhatr_FA</b>				
(Intercept)	1.637282e+01	3.459053e+01	0.4733325	0.6359760
rhatr_Iron	4.006652e+02	3.200762e+02	1.2517806	0.2106498
rhatr_AD	-5.479180e+07	2.355489e+08	-0.2326132	0.8160617
rhatr_RD	-1.097179e+08	4.710878e+08	-0.2329034	0.8158364
rhatr_MD	1.644653e+08	7.066377e+08	0.2327434	0.8159606
rhatr_FA	-4.434159e+01	7.123769e+01	-0.6224456	0.5336489
<b>Model 1: PreHD ~ rhatr_RD</b>				
(Intercept)	1.204260e+01	6.062605e+00	1.9863735	0.0469919
rhatr_RD	-2.629382e+04	1.217339e+04	-2.1599428	0.0307771

Cuadro 93: Comparison of two Binomial Logistic Regressions for PreHD vs Rest in Right ATR

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
Model 1: PreHD ~ rhatr_RD	53	54.59394	NA	NA	NA
Model 2: PreHD ~ rhatr_Iron + rhatr_AD + rhatr_RD + rhatr_MD + rhatr_FA	49	52.02663	4	2.567312	0.6326245

Cuadro 94: Binomial Logistic Regression for HD vs Rest in Right ATR

	Estimate	Std.Error	z value	p value
<b>Model 2: HD ~ rhatr_Iron + rhatr_AD + rhatr_RD + rhatr_MD + rhatr_FA</b>				
(Intercept)	-5.856590e+01	3.910585e+01	-1.4976248	0.1342307
rhatr_Iron	-1.007533e+02	3.284834e+02	-0.3067228	0.7590544
rhatr_AD	-1.319214e+08	2.246925e+08	-0.5871196	0.5571234
rhatr_RD	-2.636833e+08	4.493421e+08	-0.5868208	0.5573241
rhatr_MD	3.956868e+08	6.740457e+08	0.5870326	0.5571818
rhatr_FA	7.556812e+01	8.354744e+01	0.9044935	0.3657338
<b>Model 1: HD ~ rhatr_MD</b>				
(Intercept)	-3.110010e+01	9.139919e+00	-3.4026671	0.0006673
rhatr_MD	4.669081e+04	1.399152e+04	3.3370796	0.0008466

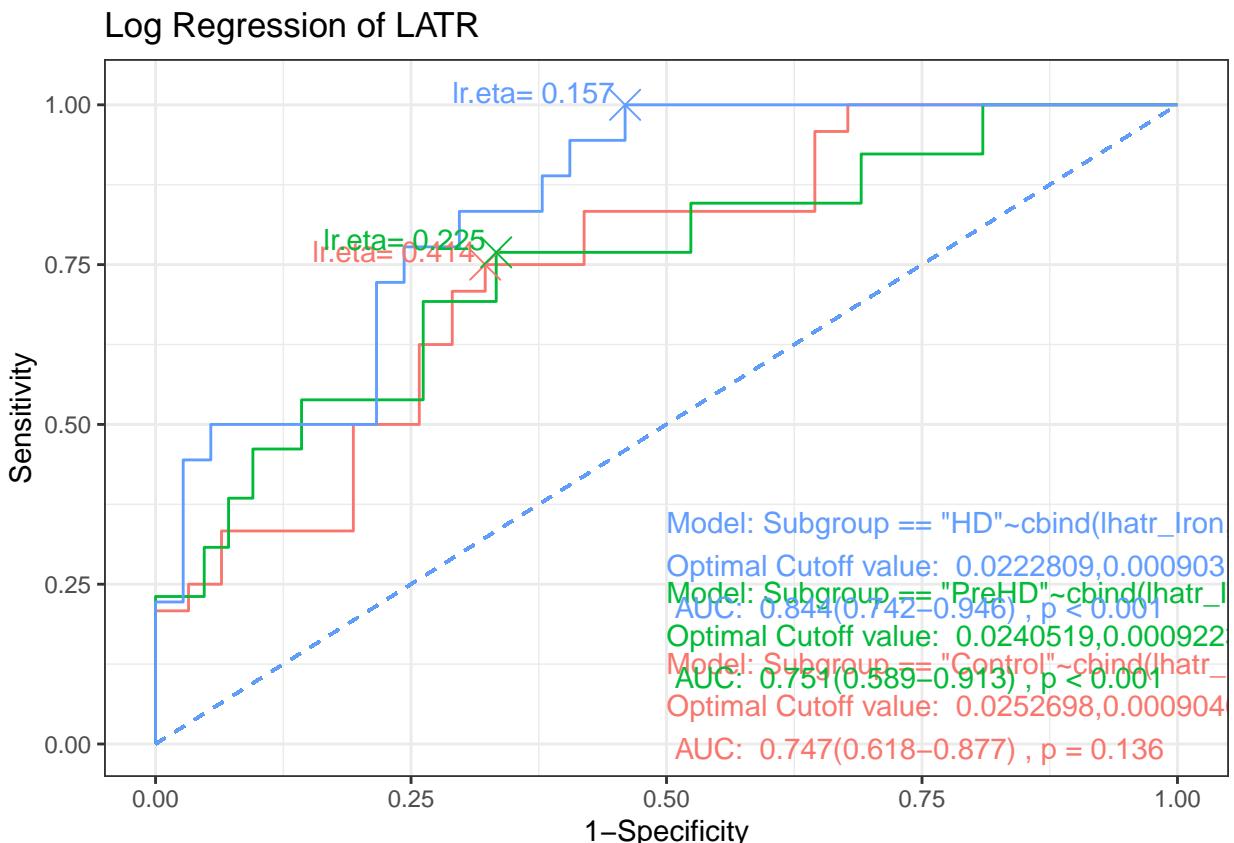
Cuadro 95: Comparison of two Binomial Logistic Regressions for HD vs Rest in Right ATR

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
Model 1: HD ~ rhatr_MD	53	51.84262	NA	NA	NA
Model 2: HD ~ rhatr_Iron + rhatr_AD + rhatr_RD + rhatr_MD + rhatr_FA	49	50.59233	4	1.250289	0.8697515

### 6.1.2 LATR

```
# ROC curves one vs the rest by multipleROC function from multipleROC package
library(multipleROC)
#multipleROC(Subgroup=='Control'~cbind(lhatr_Iron,lhatr_AD,lhatr_RD,lhatr_MD,lhatr_FA),
#               data=LATR_avg_demo, plot=TRUE)
#multipleROC(Subgroup=='PreHD'~cbind(lhatr_Iron,lhatr_AD,lhatr_RD,lhatr_MD,lhatr_FA),
#               data=LATR_avg_demo, plot=TRUE)
#multipleROC(Subgroup=='HD'~cbind(lhatr_Iron,lhatr_AD,lhatr_RD,lhatr_MD,lhatr_FA),
#               data=LATR_avg_demo, plot=TRUE)
#Global_ROC_LATR5<-multipleROC(Subgroup~cbind(lhatr_Iron,lhatr_AD,lhatr_RD,lhatr_MD,
#                                         lhatr_FA),data=LATR_avg_demo, plot=TRUE)

Global_ROC_LATR5<-multipleROC(Subgroup~cbind(lhatr_Iron,lhatr_AD,lhatr_RD,lhatr_MD,
                                              lhatr_FA),data=LATR_avg_demo, plot=FALSE)
CT_ROC_LATR5<-multipleROC(Subgroup=='Control'~cbind(lhatr_Iron,lhatr_AD,lhatr_RD,
                                                       lhatr_MD,lhatr_FA),
                           data=LATR_avg_demo, plot=FALSE)
PH_ROC_LATR5<-multipleROC(Subgroup=='PreHD'~cbind(lhatr_Iron,lhatr_AD,lhatr_RD,
                                                       lhatr_MD,lhatr_FA),
                           data=LATR_avg_demo, plot=FALSE)
HD_ROC_LATR5<-multipleROC(Subgroup=='HD'~cbind(lhatr_Iron,lhatr_AD,lhatr_RD,
                                                 lhatr_MD,lhatr_FA),
                           data=LATR_avg_demo, plot=FALSE)
plot_ROC(list(CT_ROC_LATR5,PH_ROC_LATR5,HD_ROC_LATR5),show.eta = TRUE,
         show.sens = FALSE) + labs(title= "Log Regression of LATR")
```



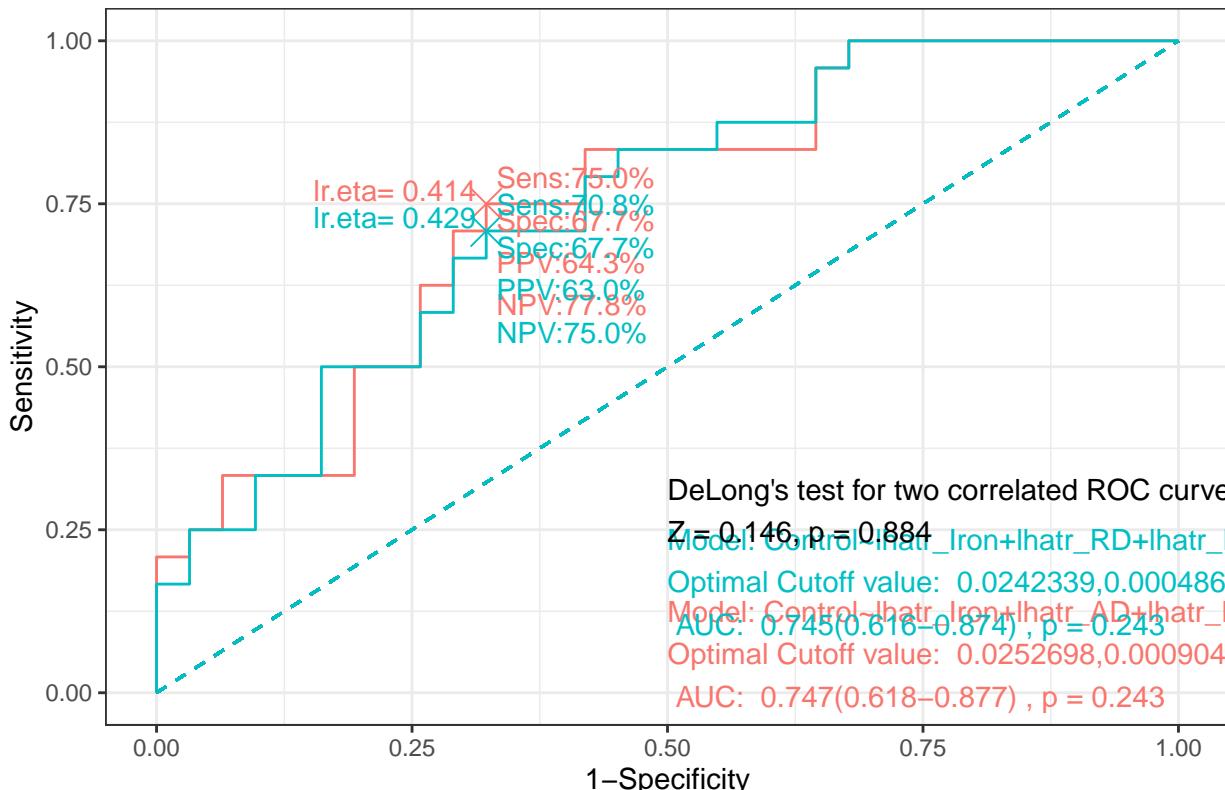
```
#plot_ROC(list(CT_ROC_LATR5,PH_ROC_LATR5,HD_ROC_LATR5),show.eta = TRUE,
#           show.sens = TRUE)+ facet_grid(~.) + labs(title= "Log Regression of LATR")
cutoff_LATR5<-rbind(CT_ROC_LATR5$cutoff[1,1],PH_ROC_LATR5$cutoff[1,1],
                      HD_ROC_LATR5$cutoff[1,1])
rownames(cutoff_LATR5)<-c('Control','PreHD','HD')
kableExtra::kable(cutoff_LATR5,
                  caption = 'Cutoff values of MRI metrics for subgroups in the 5-metrics LOGISTIC REGRESSION',
                  booktabs = TRUE) %>%
kable_styling(position = 'center', font_size = 8, latex_options='hold_position')
```

Cuadro 96: Cutoff values of MRI metrics for subgroups in the 5-metrics LOGISTIC REGRESSION model of LATR

	lhatr_Iron	lhatr_AD	lhatr_RD	lhatr_MD	lhatr_FA
Control	0.0252698	0.0009041	0.0005415	0.0006623	0.340858
PreHD	0.0240519	0.0009223	0.0005019	0.0006420	0.385004
HD	0.0222809	0.0009031	0.0004835	0.0006234	0.398835

```
# STEP ROC curves one vs the rest by step_ROC function from multipleROC package
## CONTROLS
LATR_avg_demo$Control<-ifelse(LATR_avg_demo$Subgroup=='Control',1,0)
step_ROC(Control~lhatr_Iron+lhatr_AD+lhatr_RD+lhatr_MD+lhatr_FA,data=LATR_avg_demo)+
  labs(title= "Step Log Regression of LATR for Controls")
```

### Step Log Regression of LATR for Controls



```
step_ROC(Control~lhatr_Iron+lhatr_AD+lhatr_RD+lhatr_MD+lhatr_FA,data=LATR_avg_demo,
          plot=FALSE)
```

## Analysis of Deviance Table

```

Model 1: Control ~ lhatr_Iron + lhatr_RD + lhatr_FA
Model 2: Control ~ lhatr_Iron + lhatr_AD + lhatr_RD + lhatr_MD + lhatr_FA
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      51    63.246
2      49    62.783  2  0.46278   0.7934
### This has the same results
fitLogRegLATR_CT<-glm(Control~lhatr_Iron+lhatr_AD+lhatr_RD+lhatr_MD+lhatr_FA,
                         data=LATR_avg_demo, family=binomial)
summary(fitLogRegLATR_CT)

Call:
glm(formula = Control ~ lhatr_Iron + lhatr_AD + lhatr_RD + lhatr_MD +
     lhatr_FA, family = binomial, data = LATR_avg_demo)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) 6.028e+01 2.925e+01 2.061  0.0393 *
lhatr_Iron -2.952e+02 1.988e+02 -1.485  0.1376
lhatr_AD   -9.760e+07 1.984e+08 -0.492  0.6228
lhatr_RD   -1.953e+08 3.968e+08 -0.492  0.6225
lhatr_MD    2.928e+08 5.952e+08  0.492  0.6227
lhatr_FA   -7.474e+01 6.623e+01 -1.128  0.2591
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 75.353 on 54 degrees of freedom
Residual deviance: 62.783 on 49 degrees of freedom
AIC: 74.783

Number of Fisher Scoring iterations: 4
finalLogRegLATR_CT=step(fitLogRegLATR_CT, trace=0)
summary(finalLogRegLATR_CT)

Call:
glm(formula = Control ~ lhatr_Iron + lhatr_RD + lhatr_FA, family = binomial,
     data = LATR_avg_demo)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) 51.98      17.78   2.924  0.00346 **
lhatr_Iron -299.81     191.39  -1.566  0.11724
lhatr_RD   -49574.45   17464.46  -2.839  0.00453 **
lhatr_FA    -53.13      20.31  -2.615  0.00892 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 75.353 on 54 degrees of freedom
Residual deviance: 63.246 on 51 degrees of freedom
AIC: 71.246

```

```

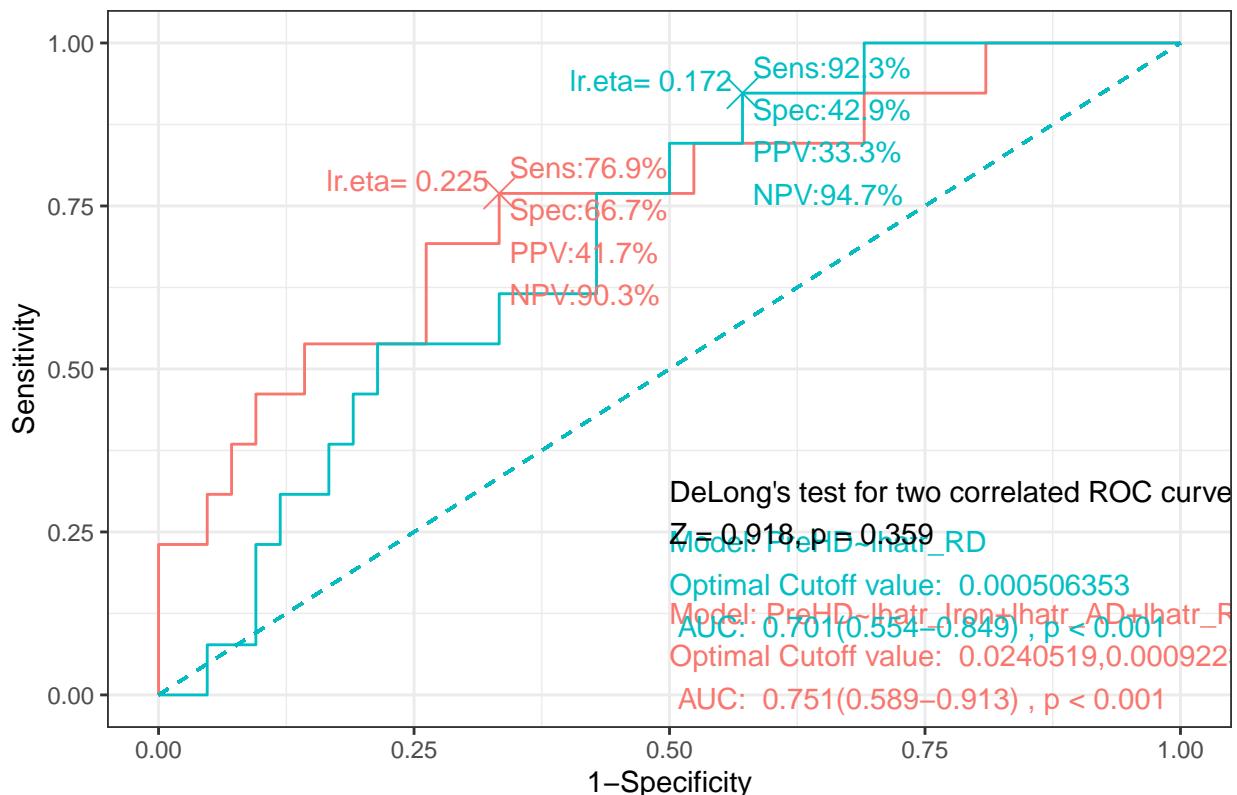
Number of Fisher Scoring iterations: 4
anova(finalLogRegLATR_CT, fitLogRegLATR_CT, test='Chisq')
Analysis of Deviance Table

Model 1: Control ~ lhatr_Iron + lhatr_RD + lhatr_FA
Model 2: Control ~ lhatr_Iron + lhatr_AD + lhatr_RD + lhatr_MD + lhatr_FA
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      51     63.246
2      49     62.783  2   0.46278   0.7934

## PREHD
LATR_avg_demo$PreHD<-ifelse(LATR_avg_demo$Subgroup=='PreHD',1,0)
step_ROC(PreHD~lhatr_Iron+lhatr_AD+lhatr_RD+lhatr_MD+lhatr_FA,data=LATR_avg_demo) +
  labs(title= "Step Log Regression of LATR for PreHD")

```

## Step Log Regression of LATR for PreHD



```

step_ROC(PreHD~lhatr_Iron+lhatr_AD+lhatr_RD+lhatr_MD+lhatr_FA,data=LATR_avg_demo,
         plot=FALSE)
Analysis of Deviance Table

Model 1: PreHD ~ lhatr_RD
Model 2: PreHD ~ lhatr_Iron + lhatr_AD + lhatr_RD + lhatr_MD + lhatr_FA
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      53     55.156
2      49     51.454  4   3.7028   0.4477

### This has the same results
fitLogRegLATR_PH<-glm(PreHD~lhatr_Iron+lhatr_AD+lhatr_RD+lhatr_MD+lhatr_FA,
                       data=LATR_avg_demo, family=binomial)
summary(fitLogRegLATR_PH)

```

```

Call:
glm(formula = PreHD ~ lhatr_Iron + lhatr_AD + lhatr_RD + lhatr_MD +
    lhatr_FA, family = binomial, data = LATR_avg_demo)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) 7.469e+00 3.270e+01  0.228   0.819
lhatr_Iron  2.062e+02 2.092e+02  0.986   0.324
lhatr_AD   -1.487e+08 2.288e+08 -0.650   0.516
lhatr_RD   -2.975e+08 4.575e+08 -0.650   0.515
lhatr_MD   4.462e+08 6.862e+08  0.650   0.516
lhatr_FA   -1.994e+01 7.631e+01 -0.261   0.794

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 60.154 on 54 degrees of freedom
Residual deviance: 51.454 on 49 degrees of freedom
AIC: 63.454

Number of Fisher Scoring iterations: 5
finalLogRegLATR_PH=step(fitLogRegLATR_PH, trace=0)
summary(finalLogRegLATR_PH)

Call:
glm(formula = PreHD ~ lhatr_RD + lhatr_MD, family = binomial,
     data = LATR_avg_demo)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)      6.299      7.993  0.788   0.4307
lhatr_RD   -78678.007 40935.016 -1.922   0.0546 .
lhatr_MD      49754.168 36196.440  1.375   0.1693
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 60.154 on 54 degrees of freedom
Residual deviance: 53.134 on 52 degrees of freedom
AIC: 59.134

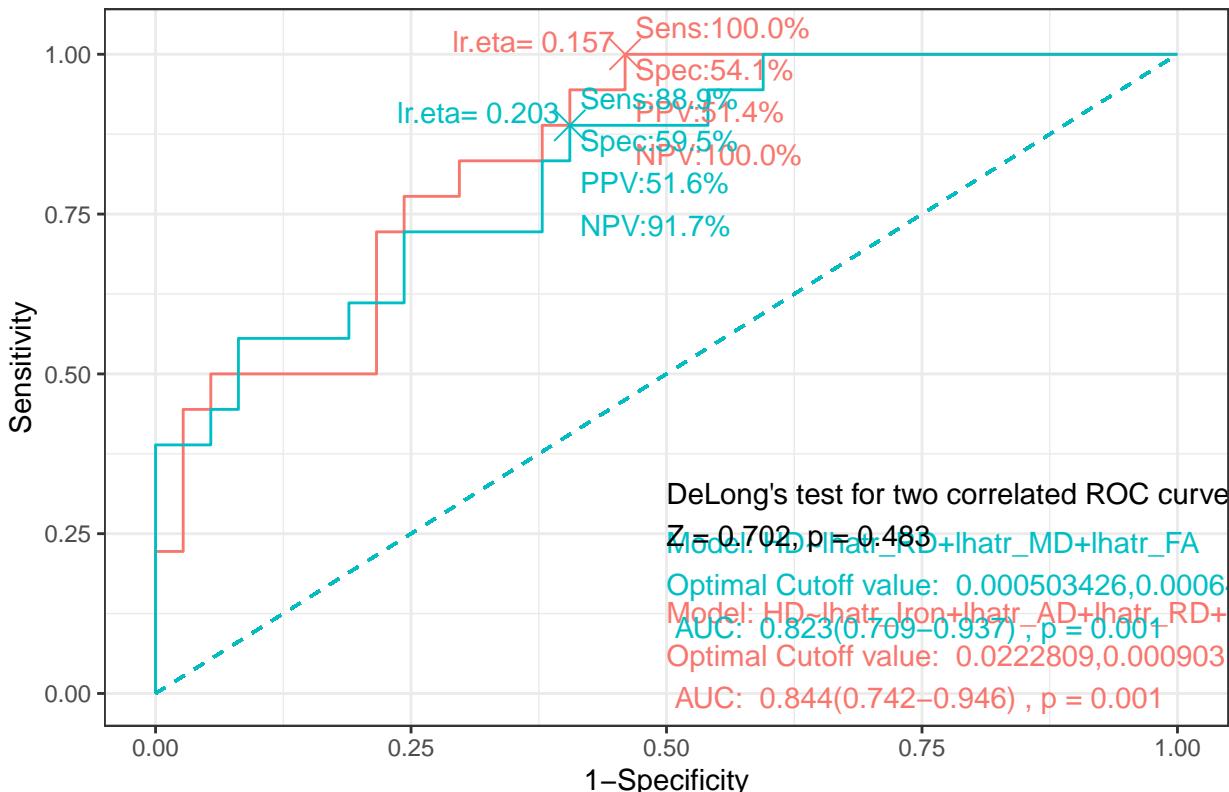
Number of Fisher Scoring iterations: 5
anova(finalLogRegLATR_PH, fitLogRegLATR_PH, test='Chisq')
Analysis of Deviance Table

Model 1: PreHD ~ lhatr_RD + lhatr_MD
Model 2: PreHD ~ lhatr_Iron + lhatr_AD + lhatr_RD + lhatr_MD + lhatr_FA
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1        52      53.134
2        49      51.454  3     1.6801   0.6414

## HD
LATR_avg_demo$HD<-ifelse(LATR_avg_demo$Subgroup=='HD', 1, 0)
step_ROC(HD~lhatr_Iron+lhatr_AD+lhatr_RD+lhatr_MD+lhatr_FA, data=LATR_avg_demo)+
  labs(title= "Step Log Regression of LATR for HD")

```

## Step Log Regression of LATR for HD



```
step_ROC(HD ~ lhatr_Iron + lhatr_AD + lhatr_RD + lhatr_MD + lhatr_FA, data=LATR_avg_demo,
         plot=FALSE)
```

Analysis of Deviance Table

```
Model 1: HD ~ lhatr_RD + lhatr_MD + lhatr_FA
Model 2: HD ~ lhatr_Iron + lhatr_AD + lhatr_RD + lhatr_MD + lhatr_FA
Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      51      51.402
2      49      49.169  2     2.2336   0.3273
```

### This has the same results

```
fitLogRegLATR_HD <- glm(HD ~ lhatr_Iron + lhatr_AD + lhatr_RD + lhatr_MD + lhatr_FA,
                           data=LATR_avg_demo, family=binomial)
summary(fitLogRegLATR_HD)
```

Call:

```
glm(formula = HD ~ lhatr_Iron + lhatr_AD + lhatr_RD + lhatr_MD +
    lhatr_FA, family = binomial, data = LATR_avg_demo)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-8.137e+01	3.400e+01	-2.393	0.0167 *
lhatr_Iron	1.198e+02	1.912e+02	0.626	0.5311
lhatr_AD	3.345e+08	2.458e+08	1.361	0.1736
lhatr_RD	6.693e+08	4.916e+08	1.361	0.1734
lhatr_MD	-1.004e+09	7.375e+08	-1.361	0.1735
lhatr_FA	1.081e+02	7.498e+01	1.442	0.1493

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

```

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 69.545 on 54 degrees of freedom
Residual deviance: 49.169 on 49 degrees of freedom
AIC: 61.169

Number of Fisher Scoring iterations: 5
finalLogRegLATR_HD=step(fitLogRegLATR_HD, trace=0)
summary(finalLogRegLATR_HD)

Call:
glm(formula = HD ~ lhatr_RD + lhatr_FA, family = binomial, data = LATR_avg_demo)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -45.89     15.90  -2.886  0.00390 **
lhatr_RD    64213.25 19540.77   3.286  0.00102 **
lhatr_FA     33.98     19.61   1.733  0.08316 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 69.545 on 54 degrees of freedom
Residual deviance: 53.063 on 52 degrees of freedom
AIC: 59.063

Number of Fisher Scoring iterations: 4
anova(finalLogRegLATR_HD, fitLogRegLATR_HD, test='Chisq')
Analysis of Deviance Table

Model 1: HD ~ lhatr_RD + lhatr_FA
Model 2: HD ~ lhatr_Iron + lhatr_AD + lhatr_RD + lhatr_MD + lhatr_FA
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      52      53.063
2      49      49.169  3   3.8949   0.273

```

Cuadro 97: Binomial Logistic Regression for Controls vs Rest in Left ATR

	Estimate	Std.Error	z value	p value
<b>Model 2: Control ~ lhatr_Iron + lhatr_AD + lhatr_RD + lhatr_MD + lhatr_FA</b>				
(Intercept)	6.028409e+01	2.925237e+01	2.0608275	0.0393195
lhatr_Iron	-2.952404e+02	1.988438e+02	-1.4847858	0.1376006
lhatr_AD	-9.760322e+07	1.983983e+08	-0.4919560	0.6227504
lhatr_RD	-1.953053e+08	3.967639e+08	-0.4922456	0.6225457
lhatr_MD	2.928482e+08	5.951678e+08	0.4920431	0.6226889
lhatr_FA	-7.474197e+01	6.623175e+01	-1.1284916	0.2591124
<b>Model 1: Control ~ lhatr_Iron + lhatr_RD + lhatr_FA</b>				
(Intercept)	5.197832e+01	1.777647e+01	2.9239955	0.0034557
lhatr_Iron	-2.998062e+02	1.913923e+02	-1.5664487	0.1172436
lhatr_RD	-4.957445e+04	1.746446e+04	-2.8385908	0.0045313
lhatr_FA	-5.312647e+01	2.031465e+01	-2.6151798	0.0089180

Cuadro 98: Comparison of two Binomial Logistic Regressions for Controls vs Rest in Left ATR

	Resid.	Resid.			
	Df	Dev	Df	Deviance	Pr(>Chi)
Model 1: Control ~ lhatr_Iron + lhatr_RD + lhatr_FA	51	63.24580	NA	NA	NA
Model 2: Control ~ lhatr_Iron + lhatr_AD + lhatr_RD + lhatr_MD + lhatr_FA	49	62.78302	2	0.4627793	0.7934302

Cuadro 99: Binomial Logistic Regression for PreHD vs Rest in Left ATR

	Estimate	Std.Error	z value	p value
<b>Model 2: PreHD ~ lhatr_Iron + lhatr_AD + lhatr_RD + lhatr_MD + lhatr_FA</b>				
(Intercept)	7.468704e+00	3.270183e+01	0.2283879	0.8193447
lhatr_Iron	2.062424e+02	2.091854e+02	0.9859310	0.3241670
lhatr_AD	-1.486993e+08	2.287552e+08	-0.6500369	0.5156684
lhatr_RD	-2.975249e+08	4.574660e+08	-0.6503760	0.5154494
lhatr_MD	4.461873e+08	6.862306e+08	0.6502002	0.5155629
lhatr_FA	-1.994037e+01	7.630619e+01	-0.2613204	0.7938454
<b>Model 1: PreHD ~ lhatr_RD + lhatr_MD</b>				
(Intercept)	6.298510e+00	7.992566e+00	0.7880461	0.4306698
lhatr_RD	-7.867801e+04	4.093502e+04	-1.9220221	0.0546030
lhatr_MD	4.975417e+04	3.619644e+04	1.3745597	0.1692680

Cuadro 100: Comparison of two Binomial Logistic Regressions for PreHD vs Rest in Left ATR

	Resid.	Resid.			
	Df	Dev	Df	Deviance	Pr(>Chi)
Model 1: PreHD ~ lhatr_RD + lhatr_MD	52	53.1336	NA	NA	NA
Model 2: PreHD ~ lhatr_Iron + lhatr_AD + lhatr_RD + lhatr_MD + lhatr_FA	49	51.4535	3	1.680093	0.6413686

Cuadro 101: Binomial Logistic Regression for HD vs Rest in Left ATR

	Estimate	Std.Error	z value	p value
<b>Model 2: HD ~ lhatr_Iron + lhatr_AD + lhatr_RD + lhatr_MD + lhatr_FA</b>				
(Intercept)	-8.136790e+01	3.400424e+01	-2.3928748	0.0167169
lhatr_Iron	1.197633e+02	1.912195e+02	0.6263133	0.5311095
lhatr_AD	3.345195e+08	2.458268e+08	1.3607932	0.1735790
lhatr_RD	6.692673e+08	4.916463e+08	1.3612782	0.1734258
lhatr_MD	-1.003681e+09	7.374701e+08	-1.3609788	0.1735204
lhatr_FA	1.081109e+02	7.498029e+01	1.4418578	0.1493425
<b>Model 1: HD ~ lhatr_RD + lhatr_FA</b>				
(Intercept)	-4.589269e+01	1.589984e+01	-2.8863610	0.0038972
lhatr_RD	6.421325e+04	1.954077e+04	3.2861166	0.0010158
lhatr_FA	3.398260e+01	1.961336e+01	1.7326250	0.0831623

Cuadro 102: Comparison of two Binomial Logistic Regressions for HD vs Rest in Left ATR

	Resid.	Resid.			
	Df	Dev	Df	Deviance	Pr(>Chi)
Model 1: HD ~ lhatr_RD + lhatr_FA	52	53.06347	NA	NA	NA

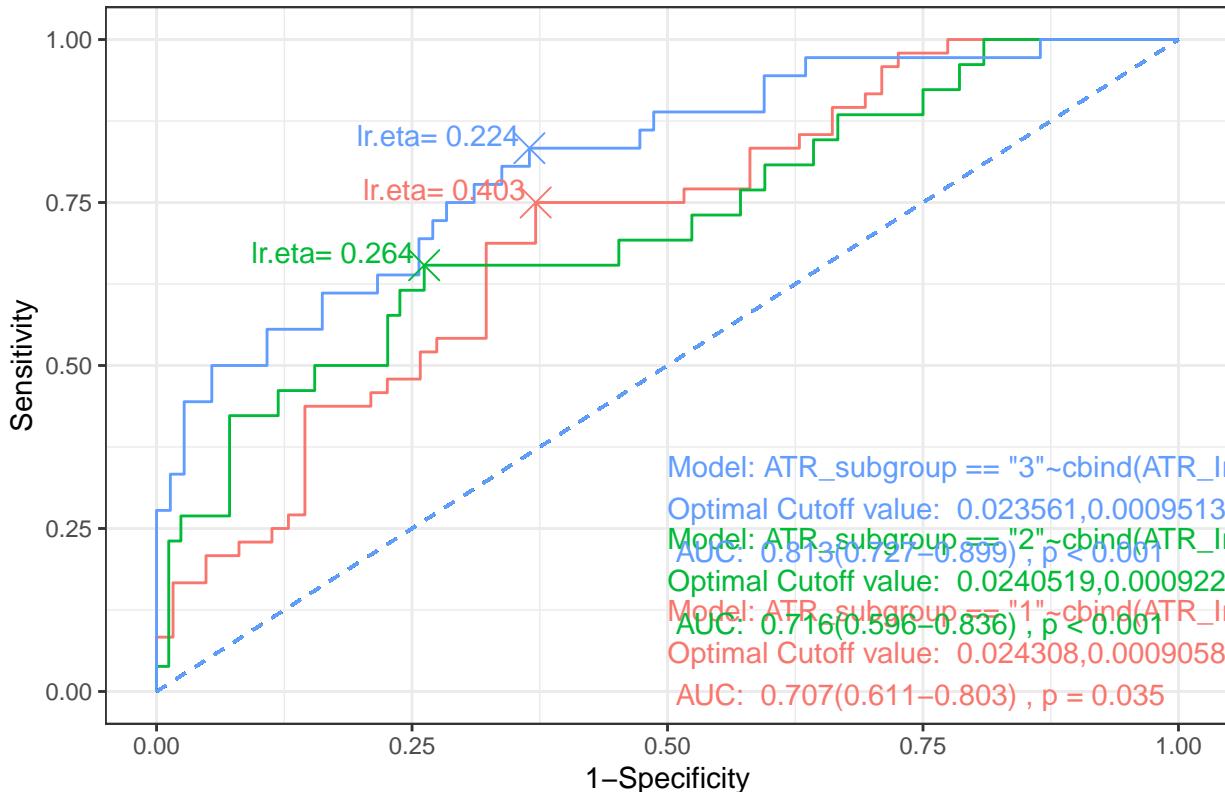
	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
Model 2: HD ~ lhatr_Iron + lhatr_AD + lhatr_RD + lhatr_MD + lhatr_FA	49	49.16853	3	3.89494	0.2730344

### 6.1.3 Ambos ATR

```
# ROC curves one vs the rest by multipleROC function from multipleROC package
library(multipleROC)
#multipleROC(ATR_subgroup=='1'~cbind(ATR_Iron,ATR_AD,ATR_RD,ATR_MD,ATR_FA),
#           data=ATR_avg, plot=TRUE)
#multipleROC(ATR_subgroup=='2'~cbind(ATR_Iron,ATR_AD,ATR_RD,ATR_MD,ATR_FA),
#           data=ATR_avg, plot=TRUE)
#multipleROC(ATR_subgroup=='3'~cbind(ATR_Iron,ATR_AD,ATR_RD,ATR_MD,ATR_FA),
#           data=ATR_avg, plot=TRUE)
#Global_ROC_ATR5<-multipleROC(ATR_subgroup~cbind(ATR_Iron,ATR_AD,ATR_RD,ATR_MD,ATR_FA),
#                                 data=ATR_avg, plot=TRUE)

Global_ROC_ATR5<-multipleROC(ATR_subgroup~cbind(ATR_Iron,ATR_AD,ATR_RD,ATR_MD,ATR_FA),
                                data=ATR_avg, plot=FALSE)
CT_ROC_ATR5<-multipleROC(ATR_subgroup=='1'~cbind(ATR_Iron,ATR_AD,ATR_RD,ATR_MD,ATR_FA),
                           data=ATR_avg, plot=FALSE)
PH_ROC_ATR5<-multipleROC(ATR_subgroup=='2'~cbind(ATR_Iron,ATR_AD,ATR_RD,ATR_MD,ATR_FA),
                           data=ATR_avg, plot=FALSE)
HD_ROC_ATR5<-multipleROC(ATR_subgroup=='3'~cbind(ATR_Iron,ATR_AD,ATR_RD,ATR_MD,ATR_FA),
                           data=ATR_avg, plot=FALSE)
plot_ROC(list(CT_ROC_ATR5,PH_ROC_ATR5,HD_ROC_ATR5),show.eta = TRUE,show.sens = FALSE) +
  labs(title= "Log Regression of Both ATR")
```

## Log Regression of Both ATR



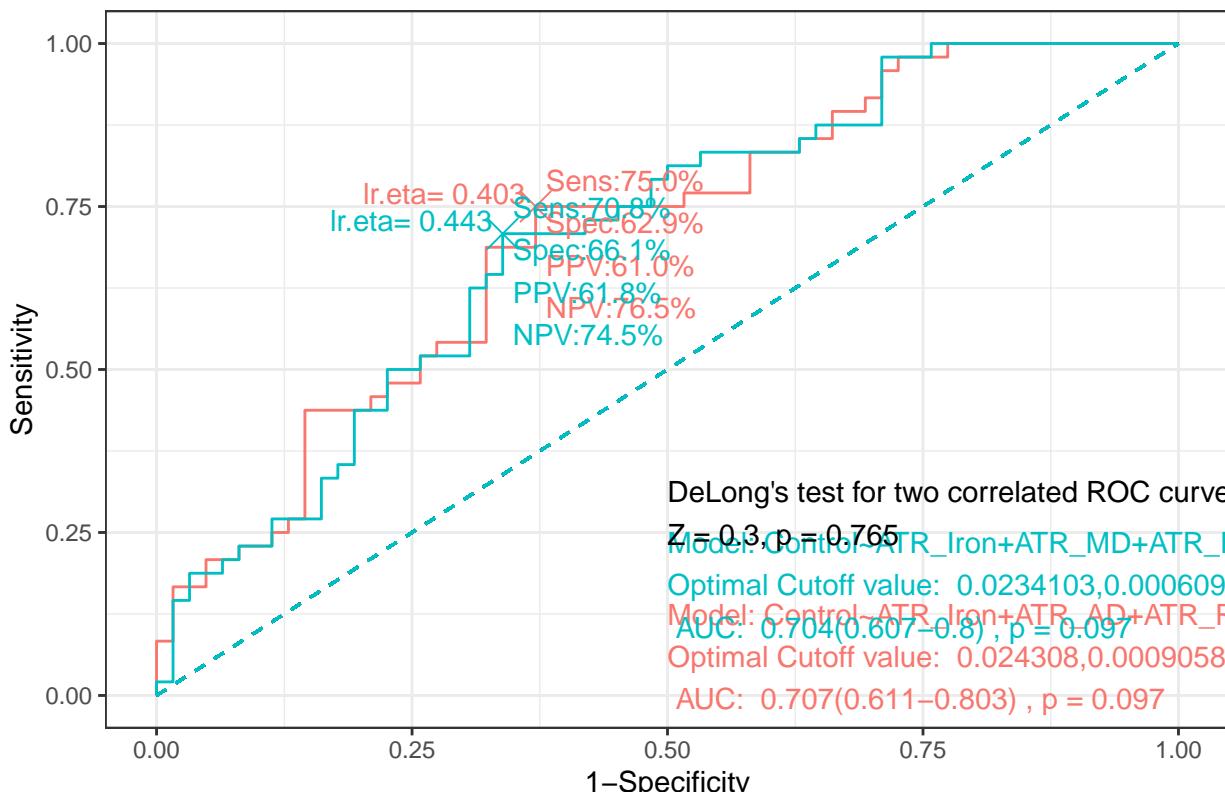
```
#plot_ROC(list(CT_ROC_ATR5,PH_ROC_ATR5,HD_ROC_ATR5),show.eta = TRUE,show.sens = TRUE)+  
# facet_grid(~.) + labs(title= "Log Regression of Both ATR")  
cutoff_ATR5<-rbind(CT_ROC_ATR5$cutoff[1,1],PH_ROC_ATR5$cutoff[1,1],HD_ROC_ATR5$cutoff[1,1])  
rownames(cutoff_ATR5)<-c('Control','PreHD','HD')  
kableExtra::kable(cutoff_ATR5,  
                  caption = 'Cutoff values of MRI metrics for subgroups in the 5-metrics LOGISTIC REGRESSION',  
                  booktabs = TRUE) %>%  
kable_styling(position = 'center', font_size = 8, latex_options='hold_position')
```

Cuadro 103: Cutoff values of MRI metrics for subgroups in the 5-metrics LOGISTIC REGRESSION model of Both ATR

	ATR_Iron	ATR_AD	ATR_RD	ATR_MD	ATR_FA
Control	0.0243080	0.0009058	0.0004783	0.0006208	0.406508
PreHD	0.0240519	0.0009223	0.0005019	0.0006420	0.385004
HD	0.0235610	0.0009513	0.0005148	0.0006603	0.380831

```
# STEP ROC curves one vs the rest by step_ROC function from multipleROC package  
## CONTROLS  
ATR_avg$Control<-ifelse(ATR_avg$ATR_subgroup=='1',1,0)  
step_ROC(Control~ATR_Iron+ATR_AD+ATR_RD+ATR_MD+ATR_FA,data=ATR_avg)+  
  labs(title= "Step Log Regression of Both ATR for Controls")
```

## Step Log Regression of Both ATR for Controls



```
step_ROC(Control~ATR_Iron+ATR_AD+ATR_RD+ATR_MD+ATR_FA, data=ATR_avg, plot=FALSE)
Analysis of Deviance Table
```

```
Model 1: Control ~ ATR_Iron + ATR_MD + ATR_FA
Model 2: Control ~ ATR_Iron + ATR_AD + ATR_RD + ATR_MD + ATR_FA
Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      106     133.88
2      104     133.26  2   0.6152   0.7352
```

### This has the same results

```
fitLogRegATR_CT<-glm(Control~ATR_Iron+ATR_AD+ATR_RD+ATR_MD+ATR_FA, data=ATR_avg,
                        family=binomial)
summary(fitLogRegATR_CT)
```

```
Call:
glm(formula = Control ~ ATR_Iron + ATR_AD + ATR_RD + ATR_MD +
    ATR_FA, family = binomial, data = ATR_avg)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	4.828e+01	1.974e+01	2.446	0.0144 *
ATR_Iron	-2.875e+02	1.444e+02	-1.990	0.0465 *
ATR_AD	2.876e+06	1.324e+08	0.022	0.9827
ATR_RD	5.677e+06	2.647e+08	0.021	0.9829
ATR_MD	-8.601e+06	3.970e+08	-0.022	0.9827
ATR_FA	-5.487e+01	4.391e+01	-1.250	0.2114

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

```

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 150.71 on 109 degrees of freedom
Residual deviance: 133.26 on 104 degrees of freedom
AIC: 145.26

Number of Fisher Scoring iterations: 4
finalLogRegCTR_CT=step(fitLogRegCTR_CT, trace=0)
summary(finalLogRegCTR_CT)

Call:
glm(formula = Control ~ ATR_Iron + ATR_MD + ATR_FA, family = binomial,
     data = ATR_avg)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) 35.359     10.102   3.500 0.000465 ***
ATR_Iron    -294.733    142.249  -2.072 0.038270 *
ATR_MD      -31768.523   8884.547  -3.576 0.000349 ***
ATR_FA       -21.755     9.482   -2.294 0.021771 *
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 150.71 on 109 degrees of freedom
Residual deviance: 133.88 on 106 degrees of freedom
AIC: 141.88

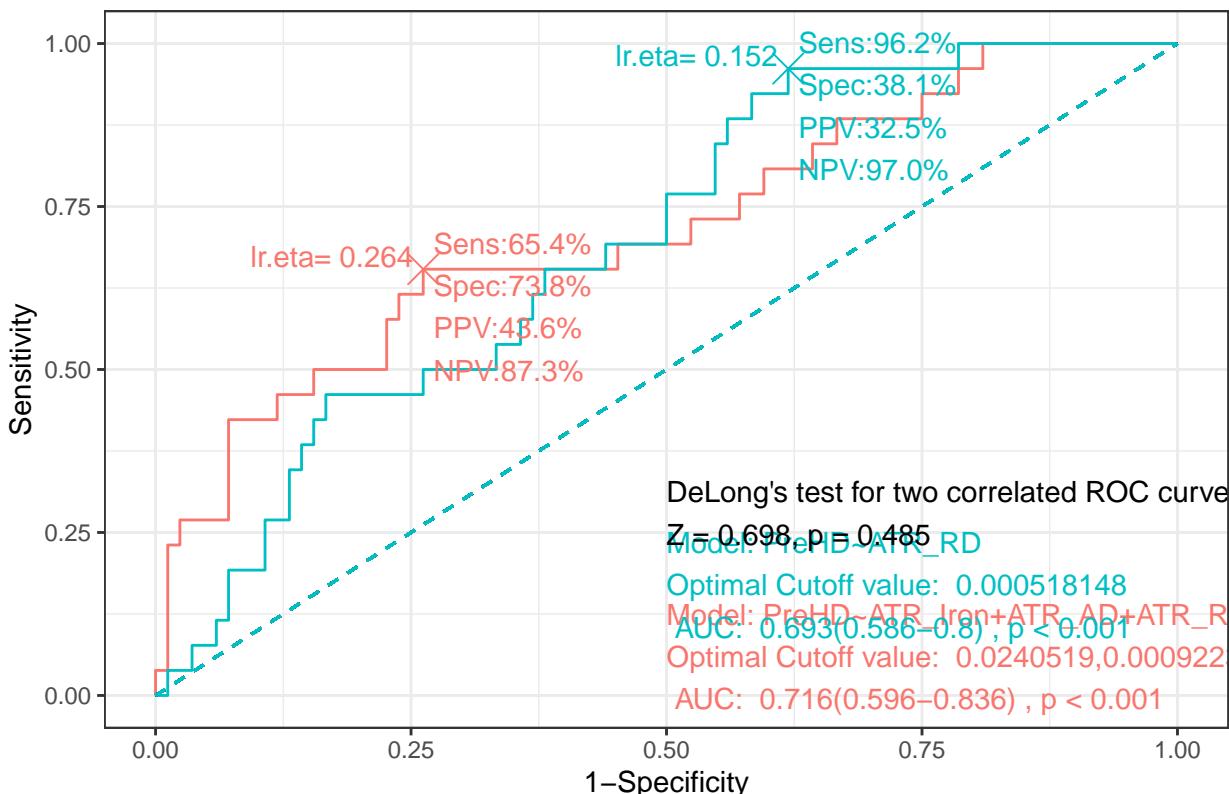
Number of Fisher Scoring iterations: 3
anova(finalLogRegCTR_CT, fitLogRegCTR_CT, test='Chisq')
Analysis of Deviance Table

Model 1: Control ~ ATR_Iron + ATR_MD + ATR_FA
Model 2: Control ~ ATR_Iron + ATR_AD + ATR_RD + ATR_MD + ATR_FA
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1       106      133.88
2       104      133.26  2    0.6152   0.7352

## PREHD
ATR_avg$PreHD<-ifelse(ATR_avg$ATR_subgroup=='2',1,0)
step_ROC(PreHD~ATR_Iron+ATR_AD+ATR_RD+ATR_MD+ATR_FA,data=ATR_avg)+
  labs(title= "Step Log Regression of Both ATR for PreHD")

```

## Step Log Regression of Both ATR for PreHD



```
step_ROC(PreHD~ATR_Iron+ATR_AD+ATR_RD+ATR_MD+ATR_FA, data=ATR_avg, plot=FALSE)
Analysis of Deviance Table
```

```
Model 1: PreHD ~ ATR_RD
Model 2: PreHD ~ ATR_Iron + ATR_AD + ATR_RD + ATR_MD + ATR_FA
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1       108     110.08
2       104    105.84  4     4.249   0.3734
```

### This has the same results

```
fitLogRegATR_PH<-glm(PreHD~ATR_Iron+ATR_AD+ATR_RD+ATR_MD+ATR_FA, data=ATR_avg,
                      family=binomial)
summary(fitLogRegATR_PH)
```

```
Call:
glm(formula = PreHD ~ ATR_Iron + ATR_AD + ATR_RD + ATR_MD + ATR_FA,
     family = binomial, data = ATR_avg)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	1.530e+01	2.239e+01	0.683	0.494
ATR_Iron	2.506e+02	1.570e+02	1.596	0.110
ATR_AD	-7.385e+07	1.620e+08	-0.456	0.648
ATR_RD	-1.478e+08	3.239e+08	-0.456	0.648
ATR_MD	2.216e+08	4.859e+08	0.456	0.648
ATR_FA	-3.801e+01	5.080e+01	-0.748	0.454

(Dispersion parameter for binomial family taken to be 1)

```

Null deviance: 120.31  on 109  degrees of freedom
Residual deviance: 105.84  on 104  degrees of freedom
AIC: 117.84

Number of Fisher Scoring iterations: 5
finalLogRegATR_PH=step(fitLogRegATR_PH, trace=0)
summary(finalLogRegATR_PH)

Call:
glm(formula = PreHD ~ ATR_Iron + ATR_RD + ATR_MD, family = binomial,
     data = ATR_avg)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -5.384e-01 8.084e+00 -0.067  0.9469
ATR_Iron      2.384e+02 1.544e+02  1.544  0.1226
ATR_RD       -5.638e+04 2.519e+04 -2.238  0.0252 *
ATR_MD        3.414e+04 2.454e+04  1.391  0.1642
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 120.31  on 109  degrees of freedom
Residual deviance: 106.76  on 106  degrees of freedom
AIC: 114.76

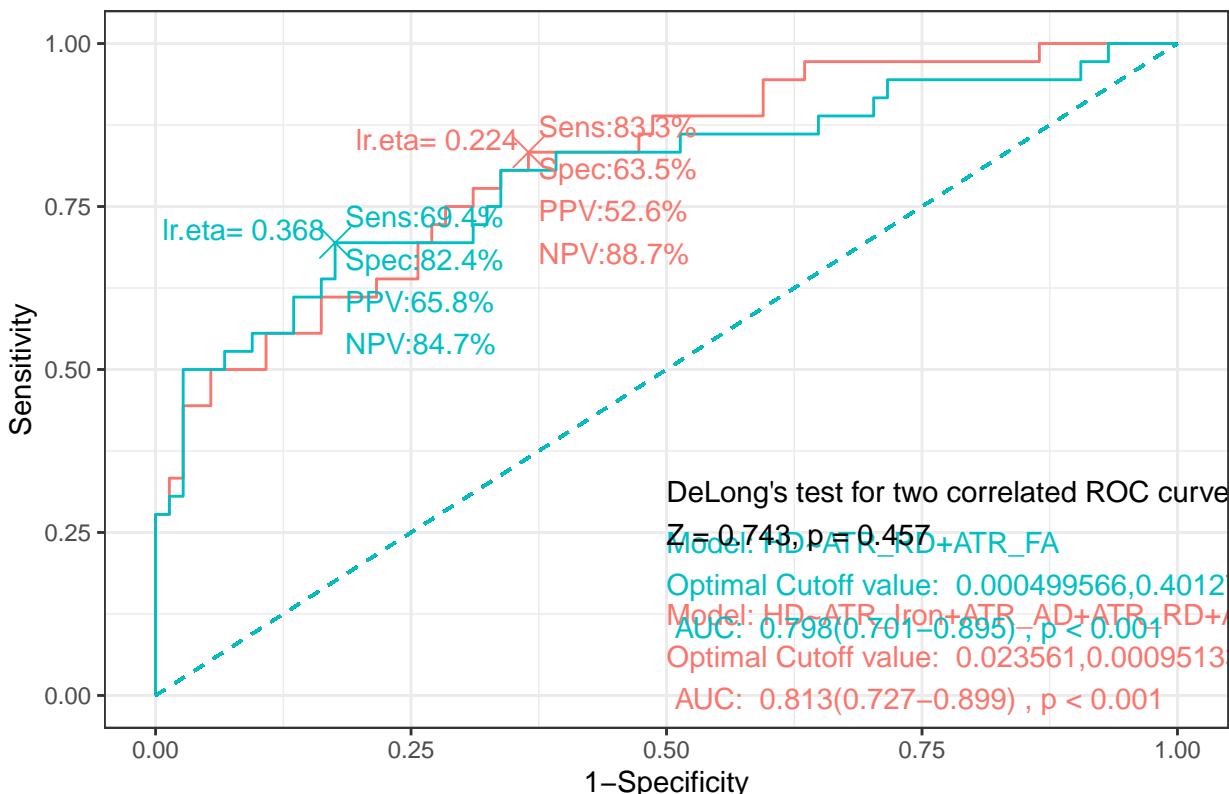
Number of Fisher Scoring iterations: 4
anova(finalLogRegATR_PH, fitLogRegATR_PH, test='Chisq')
Analysis of Deviance Table

Model 1: PreHD ~ ATR_Iron + ATR_RD + ATR_MD
Model 2: PreHD ~ ATR_Iron + ATR_AD + ATR_RD + ATR_MD + ATR_FA
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1       106      106.76
2       104      105.84  2   0.92747   0.6289

## HD
ATR_avg$HD<-ifelse(ATR_avg$ATR_subgroup=='3',1,0)
step_ROC(HD~ATR_Iron+ATR_AD+ATR_RD+ATR_MD+ATR_FA,data=ATR_avg)+
  labs(title= "Step Log Regression of Both ATR for HD")

```

## Step Log Regression of Both ATR for HD



```
step_ROC(HD ~ ATR_Iron + ATR_AD + ATR_RD + ATR_MD + ATR_FA, data=ATR_avg, plot=FALSE)
Analysis of Deviance Table
```

```
Model 1: HD ~ ATR_RD + ATR_FA
Model 2: HD ~ ATR_Iron + ATR_AD + ATR_RD + ATR_MD + ATR_FA
Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      107     106.21
2      104     103.89  3    2.3208  0.5085
```

### This has the same results

```
fitLogRegATR_HD <- glm(HD ~ ATR_Iron + ATR_AD + ATR_RD + ATR_MD + ATR_FA, data=ATR_avg,
                         family=binomial)
summary(fitLogRegATR_HD)
```

```
Call:
glm(formula = HD ~ ATR_Iron + ATR_AD + ATR_RD + ATR_MD + ATR_FA,
     family = binomial, data = ATR_avg)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-7.109e+01	2.421e+01	-2.936	0.00332 **
ATR_Iron	8.631e+01	1.536e+02	0.562	0.57421
ATR_AD	7.336e+07	1.551e+08	0.473	0.63634
ATR_RD	1.469e+08	3.103e+08	0.473	0.63588
ATR_MD	-2.202e+08	4.654e+08	-0.473	0.63617
ATR_FA	9.394e+01	5.433e+01	1.729	0.08382 .

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

```

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 139.09 on 109 degrees of freedom
Residual deviance: 103.89 on 104 degrees of freedom
AIC: 115.89

Number of Fisher Scoring iterations: 5
finalLogRegATR_HD=step(fitLogRegATR_HD, trace=0)
summary(finalLogRegATR_HD)

Call:
glm(formula = HD ~ ATR_RD + ATR_FA, family = binomial, data = ATR_avg)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -42.77     10.80  -3.960 7.49e-05 ***
ATR_RD      59640.55   13161.88   4.531 5.86e-06 ***
ATR_FA       30.99      13.40    2.313   0.0207 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 139.09 on 109 degrees of freedom
Residual deviance: 106.21 on 107 degrees of freedom
AIC: 112.21

Number of Fisher Scoring iterations: 5
anova(finalLogRegATR_HD, fitLogRegATR_HD, test='Chisq')
Analysis of Deviance Table

Model 1: HD ~ ATR_RD + ATR_FA
Model 2: HD ~ ATR_Iron + ATR_AD + ATR_RD + ATR_MD + ATR_FA
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      107      106.21
2      104      103.89  3    2.3208  0.5085

```

Cuadro 104: Binomial Logistic Regression for Controls vs Rest in Both ATR

	Estimate	Std.Error	z value	p value
<b>Model 2: Control ~ ATR_Iron + ATR_AD + ATR_RD + ATR_MD + ATR_FA</b>				
(Intercept)	4.827775e+01	1.973549e+01	2.4462402	0.0144355
ATR_Iron	-2.874788e+02	1.444332e+02	-1.9903927	0.0465477
ATR_AD	2.875823e+06	1.323500e+08	0.0217289	0.9826642
ATR_RD	5.677200e+06	2.646832e+08	0.0214490	0.9828875
ATR_MD	-8.601310e+06	3.970367e+08	-0.0216638	0.9827162
ATR_FA	-5.487245e+01	4.390912e+01	-1.2496824	0.2114156
<b>Model 1: Control ~ ATR_Iron + ATR_MD + ATR_FA</b>				
(Intercept)	3.535931e+01	1.010187e+01	3.5002748	0.0004648
ATR_Iron	-2.947330e+02	1.422491e+02	-2.0719496	0.0382701
ATR_MD	-3.176852e+04	8.884547e+03	-3.5757055	0.0003493
ATR_FA	-2.175458e+01	9.481845e+00	-2.2943402	0.0217710

Cuadro 105: Comparison of two Binomial Logistic Regressions for Controls vs Rest in Both ATR

	Resid.	Resid.	Df	Dev	Df	Deviance	Pr(>Chi)
Model 1: Control ~ ATR_Iron + ATR_MD + ATR_FA		106	133.8762	NA	NA	NA	NA
Model 2: Control ~ ATR_Iron + ATR_AD + ATR_RD + ATR_MD + ATR_FA		104	133.2610	2	0.6152042	0.7352078	

Cuadro 106: Binomial Logistic Regression for PreHD vs Rest in Both ATR

	Estimate	Std.Error	z value	p value
<b>Model 2: PreHD ~ ATR_Iron + ATR_AD + ATR_RD + ATR_MD + ATR_FA</b>				
(Intercept)	1.529561e+01	2.238802e+01	0.6832052	0.4944772
ATR_Iron	2.506118e+02	1.570252e+02	1.5959979	0.1104893
ATR_AD	-7.384848e+07	1.619740e+08	-0.4559279	0.6484418
ATR_RD	-1.478349e+08	3.239282e+08	-0.4563816	0.6481156
ATR_MD	2.216412e+08	4.859058e+08	0.4561403	0.6482891
ATR_FA	-3.801049e+01	5.079704e+01	-0.7482816	0.4542903
<b>Model 1: PreHD ~ ATR_Iron + ATR_RD + ATR_MD</b>				
(Intercept)	-5.383611e-01	8.084437e+00	-0.0665923	0.9469063
ATR_Iron	2.384033e+02	1.544116e+02	1.5439471	0.1226011
ATR_RD	-5.637715e+04	2.519207e+04	-2.2378926	0.0252281
ATR_MD	3.413951e+04	2.454149e+04	1.3910933	0.1641971

Cuadro 107: Comparison of two Binomial Logistic Regressions for PreHD vs Rest in Both ATR

	Resid.	Resid.	Df	Dev	Df	Deviance	Pr(>Chi)
Model 1: PreHD ~ ATR_Iron + ATR_RD + ATR_MD		106	106.7634	NA	NA	NA	NA
Model 2: PreHD ~ ATR_Iron + ATR_AD + ATR_RD + ATR_MD + ATR_FA		104	105.8359	2	0.9274737	0.628929	

Cuadro 108: Binomial Logistic Regression for HD vs Rest in Both ATR

	Estimate	Std.Error	z value	p value
<b>Model 2: HD ~ ATR_Iron + ATR_AD + ATR_RD + ATR_MD + ATR_FA</b>				
(Intercept)	-7.108793e+01	2.421164e+01	-2.9361055	0.0033236
ATR_Iron	8.631486e+01	1.536224e+02	0.5618637	0.5742089
ATR_AD	7.335629e+07	1.551466e+08	0.4728191	0.6363422
ATR_RD	1.469088e+08	3.102801e+08	0.4734717	0.6358767
ATR_MD	-2.201730e+08	4.654282e+08	-0.4730548	0.6361741
ATR_FA	9.393842e+01	5.433223e+01	1.7289632	0.0838157
<b>Model 1: HD ~ ATR_RD + ATR_FA</b>				
(Intercept)	-4.277051e+01	1.080010e+01	-3.9601957	0.0000749
ATR_RD	5.964055e+04	1.316188e+04	4.5313078	0.0000059
ATR_FA	3.099380e+01	1.339956e+01	2.3130456	0.0207201

Cuadro 109: Comparison of two Binomial Logistic Regressions for HD vs Rest in Both ATR

	Resid.	Df	Resid.	Dev	Df	Deviance	Pr(>Chi)
Model 1: HD ~ ATR_RD + ATR_FA		107	106.2077	NA	NA	NA	NA

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
Model 2: HD ~ ATR_Iron + ATR_AD + ATR_RD + ATR_MD + ATR_FA	104	103.8869	3	2.320796	0.5085486

## 6.2 ONE vs ONE

### 6.2.1 RATR

Primero, sin transformación logarítmica de AD, RD y MD

```
library(nnet)
# Level of reference in the outcome variable
RATR_avg_demo$Subgroup2<-relevel(RATR_avg_demo$Subgroup, ref = 'Control')
# Fit the model
fitMultinomialLogRegRATR<-multinom(Subgroup2~rhatr_Iron+rhatr_AD+rhatr_RD+rhatr_MD+rhatr_FA,
                                         data=RATR_avg_demo)
# weights: 21 (12 variable)
initial value 60.423676
iter  10 value 57.420057
iter  20 value 57.334254
iter  30 value 57.203182
iter  40 value 57.100553
iter  50 value 56.306650
iter  60 value 55.906207
iter  70 value 55.862853
iter  80 value 55.768614
iter  90 value 55.706428
iter 100 value 55.290940
final value 55.290940
stopped after 100 iterations
summary(fitMultinomialLogRegRATR)
Call:
multinom(formula = Subgroup2 ~ rhatr_Iron + rhatr_AD + rhatr_RD +
rhatr_MD + rhatr_FA, data = RATR_avg_demo)

Coefficients:
            (Intercept) rhatr_Iron  rhatr_AD   rhatr_RD   rhatr_MD   rhatr_FA
PreHD      -7.892092    174.0295  -4.612754  -2.742870  -3.366172   8.199155
HD         9.256841   -169.2366  12.646613   7.134243   8.971722 -14.935289

Std. Errors:
            (Intercept) rhatr_Iron  rhatr_AD   rhatr_RD   rhatr_MD   rhatr_FA
PreHD      8.372293    241.6216  3.610948  1.903002  2.472296 13.49741
HD        8.332402    260.2548  3.698061  2.042170  2.594133 12.54130

Residual Deviance: 110.5819
AIC: 134.5819
# Wald test (z-test)
z_fitMultinomialLogRegRATR<-summary(fitMultinomialLogRegRATR)$coefficients/summary(fitMultinomialLogRegRATR)
round(z_fitMultinomialLogRegRATR,4)
            (Intercept) rhatr_Iron  rhatr_AD   rhatr_RD   rhatr_MD   rhatr_FA
PreHD      -0.9426     0.7203  -1.2774  -1.4413  -1.3616   0.6075
HD        1.1109    -0.6503   3.4198   3.4935   3.4585  -1.1909
# p-value
```

```

p_fitMultinomialLogRegRATR<-(1-pnorm(abs(z_fitMultinomialLogRegRATR),0,1))*2
round(p_fitMultinomialLogRegRATR,4)
  (Intercept) rhatr_Iron rhatr_AD rhatr_RD rhatr_MD rhatr_FA
PreHD      0.3459     0.4714   0.2014   0.1495   0.1733   0.5435
HD        0.2666     0.5155   0.0006   0.0005   0.0005   0.2337
# Tabla
sum_fitMultinomialLogRegRATR<-summary(fitMultinomialLogRegRATR)
tbl_fitMultinomialLogRegRATR<-rbind(sum_fitMultinomialLogRegRATR$coefficients,
                                         sum_fitMultinomialLogRegRATR$standard.errors,
                                         round(z_fitMultinomialLogRegRATR,4),
                                         round(p_fitMultinomialLogRegRATR,4))
kableExtra::kable(tbl_fitMultinomialLogRegRATR, caption='Multinomial Logistic Regression for Right ATR (controls as a reference)', pack_rows(index=c('Coefficients'=2, 'Standard Errors'=2, 'Z-test'=2, 'p-value'=2)) %>%
  kable_styling(position = 'center', font_size = 8, latex_options='hold_position')

```

Cuadro 110: Multinomial Logistic Regression for Right ATR (controls as a reference)

	(Intercept)	rhatr_Iron	rhatr_AD	rhatr_RD	rhatr_MD	rhatr_FA
<b>Coefficients</b>						
PreHD	-7.892092	174.0295	-4.612754	-2.742870	-3.366172	8.199155
HD	9.256841	-169.2366	12.646612	7.134243	8.971722	-14.935289
<b>Standard Errors</b>						
PreHD	8.372293	241.6216	3.610948	1.903002	2.472296	13.497413
HD	8.332402	260.2548	3.698061	2.042170	2.594133	12.541300
<b>Z-test</b>						
PreHD	-0.942600	0.7203	-1.277400	-1.441300	-1.361600	0.607500
HD	1.110900	-0.6503	3.419800	3.493500	3.458500	-1.190900
<b>p-value</b>						
PreHD	0.345900	0.4714	0.201400	0.149500	0.173300	0.543500
HD	0.266600	0.5155	0.000600	0.000500	0.000500	0.233700

```

# Evaluación del modelo
## Predicted probabilities
PP_MultinomialLogRegRATR<-fitMultinomialLogRegRATR$fitted.values
PP_MultinomialLogRegRATR_DF<-as.data.frame(PP_MultinomialLogRegRATR)
## Predicted classes
PC_LogRegRATR<-matrix(ncol=ncol(PP_MultinomialLogRegRATR_DF),
                        nrow=nrow(PP_MultinomialLogRegRATR_DF))
for (i in 1:nrow(PP_MultinomialLogRegRATR_DF)){
  PC_LogRegRATR[i,]<-ifelse(PP_MultinomialLogRegRATR_DF[i,]==max(PP_MultinomialLogRegRATR_DF[i,]),1,0)
}
colnames(PC_LogRegRATR)<-colnames(PP_MultinomialLogRegRATR)
## ROC curves one vs the rest by multiROC package
Control_pred_LogRegRATR<- PP_MultinomialLogRegRATR_DF$Control
PreHD_pred_LogRegRATR<- PP_MultinomialLogRegRATR_DF$PreHD
HD_pred_LogRegRATR<- PP_MultinomialLogRegRATR_DF$HD
Control_true<-ifelse(RATR_avg_demo$Subgroup=='Control',1,0)
PreHD_true<-ifelse(RATR_avg_demo$Subgroup=='PreHD',1,0)
HD_true<-ifelse(RATR_avg_demo$Subgroup=='HD',1,0)
DF_ROC_LogRegRATR<-cbind(Control_true,PreHD_true,HD_true,Control_pred_LogRegRATR,
                           PreHD_pred_LogRegRATR,HD_pred_LogRegRATR)
multiROC_LogRegRATR<-multiROC::multi_roc(DF_ROC_LogRegRATR)
multiROC_LogRegRATR_DF<-plot_roc_data(multiROC_LogRegRATR)
ggplot(multiROC_LogRegRATR_DF, aes(x=1-Specificity, y=Sensitivity,
                                      color = Group)) +

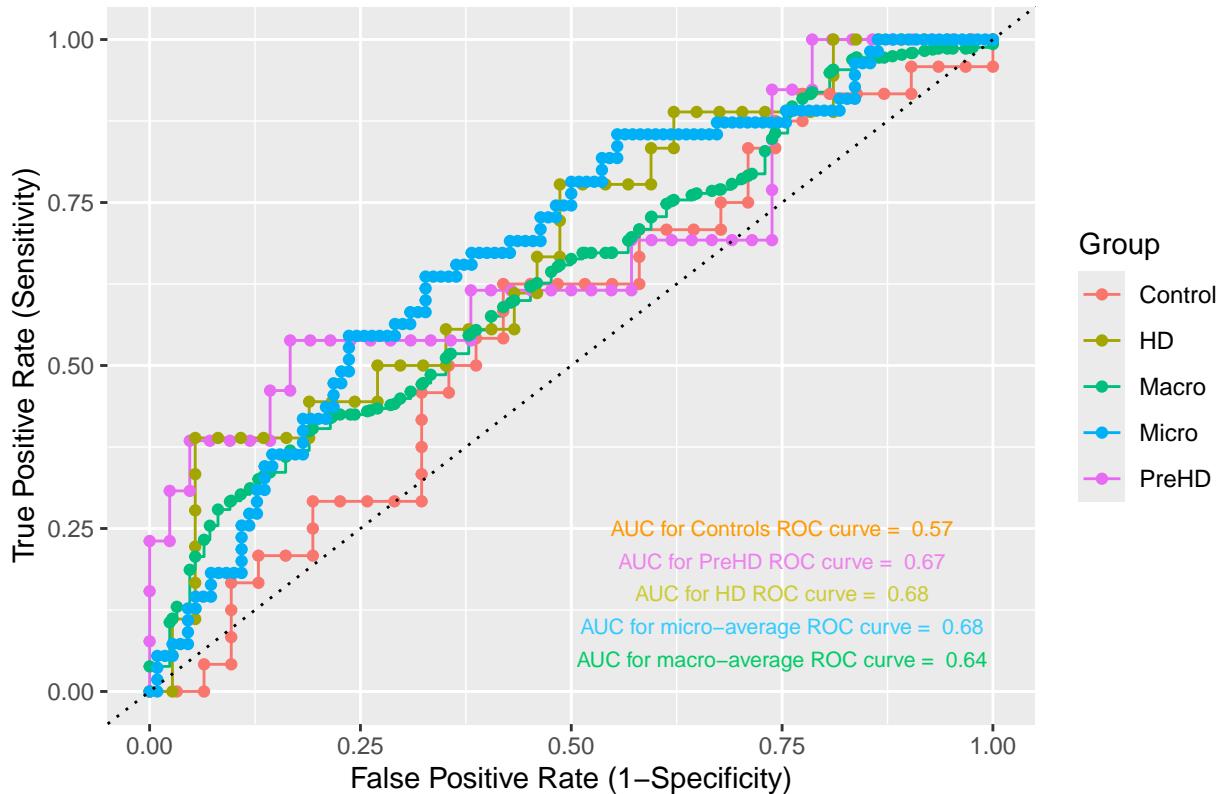
```

```

geom_step() + geom_point() + labs(title= "ROC curve for Logistic Regression model of Right ATR: One vs One",
  x = "False Positive Rate (1-Specificity)",
  y = "True Positive Rate (Sensitivity)") + geom_abline(intercept = 0,
                                                       slope=1,linetype = 'dotted') +
  annotate('text',x=0.75, y=0.25,
           label=paste('AUC for Controls ROC curve = ',
                      round(mean(multiROC_LogRegRATR_DF$AUC[multiROC_LogRegRATR_DF$Group=='Control']),2)),
           color = '#FF9900',size = 8/.pt) +
  annotate('text',x=0.75, y=0.20,
           label=paste('AUC for PreHD ROC curve = ',
                      round(mean(multiROC_LogRegRATR_DF$AUC[multiROC_LogRegRATR_DF$Group=='PreHD']),2)),
           color = 'violet',size = 8/.pt) +
  annotate('text',x=0.75, y=0.15,
           label=paste('AUC for HD ROC curve = ',
                      round(mean(multiROC_LogRegRATR_DF$AUC[multiROC_LogRegRATR_DF$Group=='HD']),2)),
           color = '#CCCC33',size = 8/.pt) +
  annotate('text',x=0.75, y=0.10,
           label=paste('AUC for micro-average ROC curve = ',
                      round(mean(multiROC_LogRegRATR_DF$AUC[multiROC_LogRegRATR_DF$Group=='Micro']),2)),
           color = '#33CCFF',size = 8/.pt) +
  annotate('text',x=0.75, y=0.05,
           label=paste('AUC for macro-average ROC curve = ',
                      round(mean(multiROC_LogRegRATR_DF$AUC[multiROC_LogRegRATR_DF$Group=='Macro']),2)),
           color = '#00CC66',size = 8/.pt)

```

ROC curve for Logistic Regression model of Right ATR: One vs One



Segundo, con transformación logarítmica de AD, RD, MD

```

library(nnet)
# Fit the model
fitMultinomialLogRegRATRLog<-multinom(Subgroup2~rhatr_Iron+log(rhatr_AD)+log(rhatr_RD)+log(rhatr_MD)+rhatr_FA,
                                         data=RATR_avg_demo)

# weights: 21 (12 variable)
initial value 60.423676
iter 10 value 57.335449
iter 20 value 54.285848
iter 30 value 49.456548
iter 40 value 49.338690
iter 50 value 48.754929
final value 48.754841
converged
summary(fitMultinomialLogRegRATRLog)
Call:
multinom(formula = Subgroup2 ~ rhatr_Iron + log(rhatr_AD) + log(rhatr_RD) +
    log(rhatr_MD) + rhatr_FA, data = RATR_avg_demo)

Coefficients:
(Intercept) rhatr_Iron log(rhatr_AD) log(rhatr_RD) log(rhatr_MD)
PreHD      -31.18636   44.31412    34.10618   -30.52739    -7.499958
HD        198.70890  -23.67815   -10.32841    22.67540   15.371862
rhatr_FA
PreHD -49.25265
HD     36.80009

Std. Errors:
(Intercept) rhatr_Iron log(rhatr_AD) log(rhatr_RD) log(rhatr_MD) rhatr_FA
PreHD      66.10494   8.376241   38.48403   38.76309   2.196443  73.86691
HD        67.50784   4.948671   45.88861   45.66604   2.684594  88.69891

Residual Deviance: 97.50968
AIC: 121.5097
# Wald test (z-test)
z_fitMultinomialLogRegRATRLog<-summary(fitMultinomialLogRegRATRLog)$coefficients/summary(fitMultinomialLogRegRATRLog)
round(z_fitMultinomialLogRegRATRLog,4)
(Intercept) rhatr_Iron log(rhatr_AD) log(rhatr_RD) log(rhatr_MD) rhatr_FA
PreHD      -0.4718     5.2905     0.8862    -0.7875    -3.4146   -0.6668
HD        2.9435    -4.7847    -0.2251     0.4965     5.7260    0.4149

# p-value
p_fitMultinomialLogRegRATRLog<-(1-pnorm(abs(z_fitMultinomialLogRegRATRLog),0,1))*2
round(p_fitMultinomialLogRegRATRLog,4)
(Intercept) rhatr_Iron log(rhatr_AD) log(rhatr_RD) log(rhatr_MD) rhatr_FA
PreHD      0.6371       0     0.3755     0.4310     6e-04    0.5049
HD        0.0032       0     0.8219     0.6195     0e+00    0.6782

# Tabla
sum_fitMultinomialLogRegRATRLog<-summary(fitMultinomialLogRegRATRLog)
tbl_fitMultinomialLogRegRATRLog<-rbind(sum_fitMultinomialLogRegRATRLog$coefficients,
                                         sum_fitMultinomialLogRegRATRLog$standard.errors,
                                         round(z_fitMultinomialLogRegRATRLog,4),
                                         round(p_fitMultinomialLogRegRATRLog,4))

kableExtra::kable(tbl_fitMultinomialLogRegRATRLog, caption='Multinomial Logistic Regression for Right ATR',
                 pack_rows(index=c('Coefficients'=2, 'Standard Errors'=2, 'Z-test'=2, 'p-value'=2)) %>%
kable_styling(position = 'center', font_size = 8, latex_options='hold_position')

```

Cuadro 111: Multinomial Logistic Regression for Right ATR (controls as a reference) with log transform of diffusivities

	(Intercept)	rhatr_Iron	log(rhatr_AD)	log(rhatr_RD)	log(rhatr_MD)	rhatr_FA
<b>Coefficients</b>						
PreHD	-31.18636	44.314117	34.10618	-30.52739	-7.499958	-49.25265
HD	198.70890	-23.678145	-10.32841	22.67540	15.371862	36.80009
<b>Standard Errors</b>						
PreHD	66.10494	8.376242	38.48403	38.76309	2.196443	73.86691
HD	67.50784	4.948671	45.88861	45.66604	2.684594	88.69891
<b>Z-test</b>						
PreHD	-0.47180	5.290500	0.88620	-0.78750	-3.414600	-0.66680
HD	2.94350	-4.784700	-0.22510	0.49650	5.726000	0.41490
<b>p-value</b>						
PreHD	0.63710	0.000000	0.37550	0.43100	0.000600	0.50490
HD	0.00320	0.000000	0.82190	0.61950	0.000000	0.67820

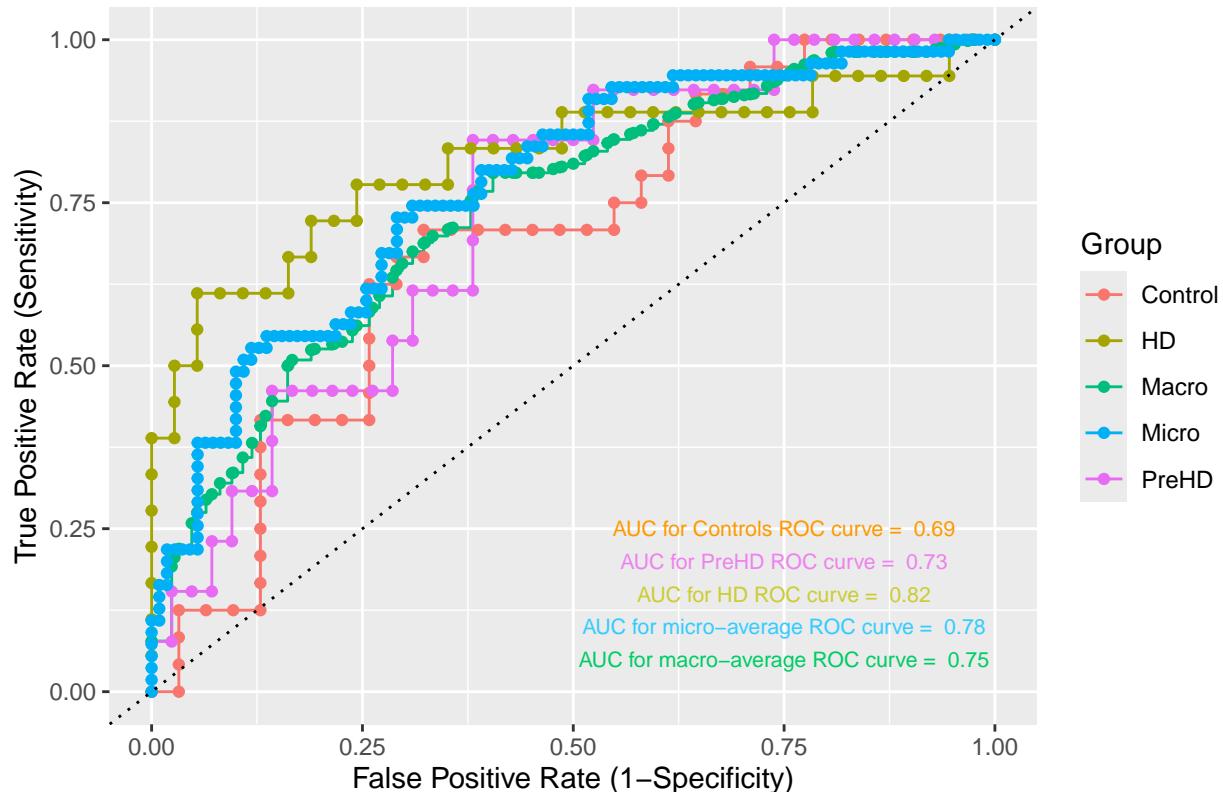
```
# Evaluación del modelo
## Predicted probabilities
PP_MultinomialLogRegLATR<-fitMultinomialLogRegRATRLog$fitted.values
PP_MultinomialLogRegLATR_DF<-as.data.frame(PP_MultinomialLogRegLATR)
## Predicted classes
PC_LogRegLATR<-matrix(ncol=ncol(PP_MultinomialLogRegLATR_DF),
                       nrow=nrow(PP_MultinomialLogRegLATR_DF))
for (i in 1:nrow(PP_MultinomialLogRegLATR_DF)){
  PC_LogRegLATR[i,]<-ifelse(PP_MultinomialLogRegLATR_DF[i,]==max(PP_MultinomialLogRegLATR_DF[i,]),1,0)
}
colnames(PC_LogRegLATR)<-colnames(PP_MultinomialLogRegLATR_DF)
## ROC curves one vs the rest by multiROC package
Control_pred_LogRegLATR<- PP_MultinomialLogRegLATR_DF$Control
PreHD_pred_LogRegLATR<- PP_MultinomialLogRegLATR_DF$PreHD
HD_pred_LogRegLATR<- PP_MultinomialLogRegLATR_DF$HD
Control_true<-ifelse(LATR_avg_demo$Subgroup=='Control',1,0)
PreHD_true<-ifelse(LATR_avg_demo$Subgroup=='PreHD',1,0)
HD_true<-ifelse(LATR_avg_demo$Subgroup=='HD',1,0)
DF_ROC_LogRegLATR<-cbind(Control_true,PreHD_true,HD_true,Control_pred_LogRegLATR,
                           PreHD_pred_LogRegLATR,HD_pred_LogRegLATR)
multiROC_LogRegLATR<-multiROC::multi_roc(DF_ROC_LogRegLATR)
multiROC_LogRegLATR_DF<-plot_roc_data(multiROC_LogRegLATR)
ggplot(multiROC_LogRegLATR_DF, aes(x=1-Specificity, y=Sensitivity, color = Group)) +
  geom_step() + geom_point() + labs(title= "ROC curve for Logistic Regression model of Right ATR with log-
  x = "False Positive Rate (1-Specificity)",
  y = "True Positive Rate (Sensitivity)" + geom_abline(intercept = 0, slope=1,linetype = 'dotted') +
  annotate('text',x=0.75, y=0.25,
           label=paste('AUC for Controls ROC curve = ',
           round(mean(multiROC_LogRegLATR_DF$AUC[multiROC_LogRegLATR_DF$Group=='Control']),2)),
           color = '#FF9900',size = 8/.pt) +
  annotate('text',x=0.75, y=0.20,
           label=paste('AUC for PreHD ROC curve = ',
           round(mean(multiROC_LogRegLATR_DF$AUC[multiROC_LogRegLATR_DF$Group=='PreHD']),2)),
           color = 'violet',size = 8/.pt) +
  annotate('text',x=0.75, y=0.15,
           label=paste('AUC for HD ROC curve = ',
           round(mean(multiROC_LogRegLATR_DF$AUC[multiROC_LogRegLATR_DF$Group=='HD']),2)),
           color = '#CCCC33',size = 8/.pt) +
  annotate('text',x=0.75, y=0.10,
           label=paste('AUC for micro-average ROC curve = ',
```

```

    round(mean(multiROC_LogRegLATR_DF$AUC[multiROC_LogRegLATR_DF$Group=='Micro']),2),
    color = '#33CCFF',size = 8/.pt)+
  annotate('text',x=0.75, y=0.05,
    label=paste('AUC for macro-average ROC curve = ',
    round(mean(multiROC_LogRegLATR_DF$AUC[multiROC_LogRegLATR_DF$Group=='Macro']),2)),
    color = '#00CC66',size = 8/.pt)

```

ROC curve for Logistic Regression model of Right ATR with log-transform



### 6.2.2 LATR

Primero, sin transformación logarítmica de AD, RD y MD

```

library(nnet)
# Level of reference in the outcome variable
LATR_avg_demo$Subgroup2<-relevel(LATR_avg_demo$Subgroup, ref = 'Control')
# Fit the model
fitMultinomialLogRegLATR<-multinom(Subgroup2~lhatr_Iron+lhatr_AD+lhatr_RD+lhatr_MD+lhatr_FA,
                                         data=LATR_avg_demo)
# weights: 21 (12 variable)
initial value 60.423676
iter 10 value 56.226073
iter 20 value 55.828933
iter 30 value 55.817687
iter 40 value 55.800403
iter 50 value 55.746919
iter 60 value 55.713543
iter 70 value 55.689973
iter 80 value 55.683016

```

```

iter  90 value 55.665229
iter 100 value 55.633781
final  value 55.633781
stopped after 100 iterations
summary(fitMultinomialLogRegLATR)
Call:
multinom(formula = Subgroup2 ~ lhatr_Iron + lhatr_AD + lhatr_RD +
    lhatr_MD + lhatr_FA, data = LATR_avg_demo)

Coefficients:
(Intercept) lhatr_Iron lhatr_AD lhatr_RD lhatr_MD lhatr_FA
PreHD -16.4061557 40.748348 -0.6126381 -0.4393859 -0.4971262 38.19584
HD 0.4123662 2.289748 3.3075545 1.9843584 2.4254061 -2.05664

Std. Errors:
(Intercept) lhatr_Iron lhatr_AD lhatr_RD lhatr_MD lhatr_FA
PreHD 8.708703 196.1070 1.208109 0.5630415 0.7779546 18.01507
HD 6.558915 160.9015 0.781278 0.3270644 0.4781407 12.76645

Residual Deviance: 111.2676
AIC: 135.2676
# Wald test (z-test)
z_fitMultinomialLogRegLATR<-summary(fitMultinomialLogRegLATR)$coefficients/summary(fitMultinomialLogRegLATR)
round(z_fitMultinomialLogRegLATR,4)
(Intercept) lhatr_Iron lhatr_AD lhatr_RD lhatr_MD lhatr_FA
PreHD -1.8839 0.2078 -0.5071 -0.7804 -0.6390 2.1202
HD 0.0629 0.0142 4.2335 6.0672 5.0726 -0.1611
# p-value
p_fitMultinomialLogRegLATR<-(1-pnorm(abs(z_fitMultinomialLogRegLATR),0,1))*2
round(p_fitMultinomialLogRegLATR,4)
(Intercept) lhatr_Iron lhatr_AD lhatr_RD lhatr_MD lhatr_FA
PreHD 0.0596 0.8354 0.6121 0.4352 0.5228 0.034
HD 0.9499 0.9886 0.0000 0.0000 0.0000 0.872
# Tabla
sum_fitMultinomialLogRegLATR<-summary(fitMultinomialLogRegLATR)
tbl_fitMultinomialLogRegLATR<-rbind(sum_fitMultinomialLogRegLATR$coefficients,
                                         sum_fitMultinomialLogRegLATR$standard.errors,
                                         round(z_fitMultinomialLogRegLATR,4),
                                         round(p_fitMultinomialLogRegLATR,4))
kableExtra::kable(tbl_fitMultinomialLogRegLATR, caption='Multinomial Logistic Regression for Left ATR (controls as a reference)')
  pack_rows(index=c('Coefficients'=2, 'Standard Errors'=2, 'Z-test'=2, 'p-value'=2)) %>%
  kable_styling(position = 'center', font_size = 8, latex_options='hold_position')

```

Cuadro 112: Multinomial Logistic Regression for Left ATR (controls as a reference)

	(Intercept)	lhatr_Iron	lhatr_AD	lhatr_RD	lhatr_MD	lhatr_FA
<b>Coefficients</b>						
PreHD	-16.4061557	40.748348	-0.6126381	-0.4393859	-0.4971262	38.19584
HD	0.4123662	2.289748	3.3075545	1.9843584	2.4254061	-2.05664
<b>Standard Errors</b>						
PreHD	8.7087027	196.107022	1.2081090	0.5630415	0.7779546	18.01507
HD	6.5589146	160.901538	0.7812780	0.3270644	0.4781407	12.76645
<b>Z-test</b>						
PreHD	-1.8839000	0.207800	-0.5071000	-0.7804000	-0.6390000	2.12020
HD	0.0629000	0.014200	4.2335000	6.0672000	5.0726000	-0.16110

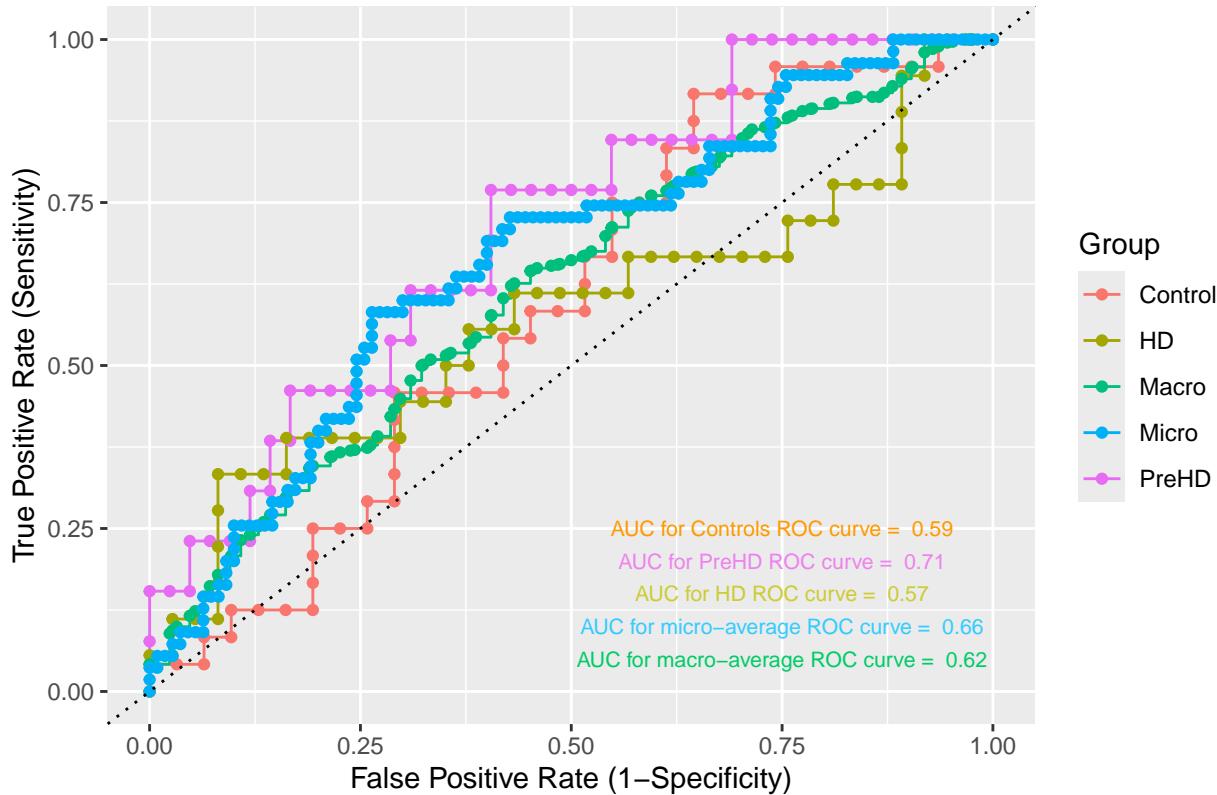
p-value							
PreHD	0.0596000	0.835400	0.6121000	0.4352000	0.5228000	0.03400	
HD	0.9499000	0.988600	0.0000000	0.0000000	0.0000000	0.87200	

```

# Evaluación del modelo
## Predicted probabilities
PP_MultinomialLogRegLATR<-fitMultinomialLogRegLATR$fitted.values
PP_MultinomialLogRegLATR_DF<-as.data.frame(PP_MultinomialLogRegLATR)
## Predicted classes
PC_LogRegLATR<-matrix(ncol=ncol(PP_MultinomialLogRegLATR_DF),
                        nrow=nrow(PP_MultinomialLogRegLATR_DF))
for (i in 1:nrow(PP_MultinomialLogRegLATR_DF)){
  PC_LogRegLATR[i,]<-ifelse(PP_MultinomialLogRegLATR_DF[i,]==max(PP_MultinomialLogRegLATR_DF[i,]),1,0)
}
colnames(PC_LogRegLATR)<-colnames(PP_MultinomialLogRegLATR_DF)
## ROC curves one vs the rest by multiROC package
Control_pred_LogRegLATR<- PP_MultinomialLogRegLATR_DF$Control
PreHD_pred_LogRegLATR<- PP_MultinomialLogRegLATR_DF$PreHD
HD_pred_LogRegLATR<- PP_MultinomialLogRegLATR_DF$HD
Control_true<-ifelse(LATR_avg_demo$Subgroup=='Control',1,0)
PreHD_true<-ifelse(LATR_avg_demo$Subgroup=='PreHD',1,0)
HD_true<-ifelse(LATR_avg_demo$Subgroup=='HD',1,0)
DF_ROC_LogRegLATR<-cbind(Control_true,PreHD_true,HD_true,Control_pred_LogRegLATR,
                           PreHD_pred_LogRegLATR,HD_pred_LogRegLATR)
multiROC_LogRegLATR<-multiROC::multi_roc(DF_ROC_LogRegLATR)
multiROC_LogRegLATR_DF<-plot_roc_data(multiROC_LogRegLATR)
ggplot(multiROC_LogRegLATR_DF, aes(x=1-Specificity, y=Sensitivity, color = Group)) +
  geom_step() + geom_point() + labs(title= "ROC curve for Logistic Regression model of Left ATR: One vs One")
  x = "False Positive Rate (1-Specificity)",
  y = "True Positive Rate (Sensitivity)" + geom_abline(intercept = 0, slope=1,linetype = 'dotted') +
  annotate('text',x=0.75, y=0.25,
           label=paste('AUC for Controls ROC curve = ',
                       round(mean(multiROC_LogRegLATR_DF$AUC[multiROC_LogRegLATR_DF$Group=='Control']),2)),
           color = '#FF9900',size = 8/.pt) +
  annotate('text',x=0.75, y=0.20,
           label=paste('AUC for PreHD ROC curve = ',
                       round(mean(multiROC_LogRegLATR_DF$AUC[multiROC_LogRegLATR_DF$Group=='PreHD']),2)),
           color = 'violet',size = 8/.pt) +
  annotate('text',x=0.75, y=0.15,
           label=paste('AUC for HD ROC curve = ',
                       round(mean(multiROC_LogRegLATR_DF$AUC[multiROC_LogRegLATR_DF$Group=='HD']),2)),
           color = '#CCCC33',size = 8/.pt) +
  annotate('text',x=0.75, y=0.10,
           label=paste('AUC for micro-average ROC curve = ',
                       round(mean(multiROC_LogRegLATR_DF$AUC[multiROC_LogRegLATR_DF$Group=='Micro']),2)),
           color = '#33CCFF',size = 8/.pt) +
  annotate('text',x=0.75, y=0.05,
           label=paste('AUC for macro-average ROC curve = ',
                       round(mean(multiROC_LogRegLATR_DF$AUC[multiROC_LogRegLATR_DF$Group=='Macro']),2)),
           color = '#00CC66',size = 8/.pt)

```

## ROC curve for Logistic Regression model of Left ATR: One vs One



Segundo, con transformación logarítmica de AD, RD, MD

```

library(nnet)
# Fit the model
fitMultinomialLogRegLATRLog<-multinom(Subgroup2~lhatr_Iron+log(lhatr_AD)+log(lhatr_RD)+log(lhatr_MD)+lhatr_FA, data=LATR_avg_demo)

# weights:  21 (12 variable)
initial  value 60.423676
iter   10 value 55.864407
iter   20 value 54.606400
iter   30 value 48.922423
iter   40 value 48.215340
iter   50 value 47.407295
iter   60 value 46.942785
iter   70 value 46.906957
iter   80 value 46.661372
iter   90 value 46.632864
iter 100 value 46.594062
final  value 46.594062
stopped after 100 iterations
summary(fitMultinomialLogRegLATRLog)
Call:
multinom(formula = Subgroup2 ~ lhatr_Iron + log(lhatr_AD) + log(lhatr_RD) +
log(lhatr_MD) + lhatr_FA, data = LATR_avg_demo)

Coefficients:
(Intercept) lhatr_Iron log(lhatr_AD) log(lhatr_RD) log(lhatr_MD)
PreHD      7.021926    77.25564     23.65679    -18.38481     -2.253572
HD        240.972934   43.19741    -51.07177     62.93264     22.971767

```

```

        lhatr_FA
PreHD -0.6085196
HD    123.4933564

Std. Errors:
  (Intercept) lhatr_Iron log(lhatr_AD) log(lhatr_RD) log(lhatr_MD) lhatr_FA
PreHD   70.15310   102.9293   41.55374   42.23890   2.505551 77.99149
HD     74.67481   124.9031   43.41982   43.61838   3.125071 82.48408

Residual Deviance: 93.18812
AIC: 117.1881
# Wald test (z-test)
z_fitMultinomialLogRegLATRLog<-summary(fitMultinomialLogRegLATRLog)$coefficients/summary(fitMultinomialLog
round(z_fitMultinomialLogRegLATRLog,4)
  (Intercept) lhatr_Iron log(lhatr_AD) log(lhatr_RD) log(lhatr_MD) lhatr_FA
PreHD   0.1001    0.7506    0.5693    -0.4353   -0.8994 -0.0078
HD     3.2270    0.3458   -1.1762    1.4428    7.3508  1.4972
# p-value
p_fitMultinomialLogRegLATRLog<-(1-pnorm(abs(z_fitMultinomialLogRegLATRLog),0,1))*2
round(p_fitMultinomialLogRegLATRLog,4)
  (Intercept) lhatr_Iron log(lhatr_AD) log(lhatr_RD) log(lhatr_MD) lhatr_FA
PreHD   0.9203    0.4529    0.5691    0.6634    0.3684  0.9938
HD     0.0013    0.7295    0.2395    0.1491    0.0000  0.1343
# Tabla
sum_fitMultinomialLogRegLATRLog<-summary(fitMultinomialLogRegLATRLog)
tbl_fitMultinomialLogRegLATRLog<-rbind(sum_fitMultinomialLogRegLATRLog$coefficients,
                                         sum_fitMultinomialLogRegLATRLog$standard.errors,
                                         round(z_fitMultinomialLogRegLATRLog,4),
                                         round(p_fitMultinomialLogRegLATRLog,4))
kableExtra::kable(tbl_fitMultinomialLogRegLATRLog, caption='Multinomial Logistic Regression for Left ATR (',
  pack_rows(index=c('Coefficients'=2, 'Standard Errors'=2, 'Z-test'=2, 'p-value'=2)) %>%
  kable_styling(position = 'center', font_size = 8, latex_options='hold_position')

```

Cuadro 113: Multinomial Logistic Regression for Left ATR (controls as a reference) with log transform of diffusivities

	(Intercept)	lhatr_Iron	log(lhatr_AD)	log(lhatr_RD)	log(lhatr_MD)	lhatr_FA
<b>Coefficients</b>						
PreHD	7.021926	77.25564	23.65679	-18.38481	-2.253572	-0.6085196
HD	240.972934	43.19741	-51.07177	62.93264	22.971767	123.4933564
<b>Standard Errors</b>						
PreHD	70.153098	102.92928	41.55374	42.23890	2.505551	77.9914927
HD	74.674809	124.90313	43.41982	43.61838	3.125071	82.4840848
<b>Z-test</b>						
PreHD	0.100100	0.75060	0.56930	-0.43530	-0.899400	-0.0078000
HD	3.227000	0.34580	-1.17620	1.44280	7.350800	1.4972000
<b>p-value</b>						
PreHD	0.920300	0.45290	0.56910	0.66340	0.368400	0.9938000
HD	0.001300	0.72950	0.23950	0.14910	0.000000	0.1343000

```

# Evaluación del modelo
## Predicted probabilities
PP_MultinomialLogRegLATR<-fitMultinomialLogRegLATRLog$fitted.values
PP_MultinomialLogRegLATR_DF<-as.data.frame(PP_MultinomialLogRegLATR)
## Predicted classes
PC_LogRegLATR<-matrix(ncol=ncol(PP_MultinomialLogRegLATR_DF),

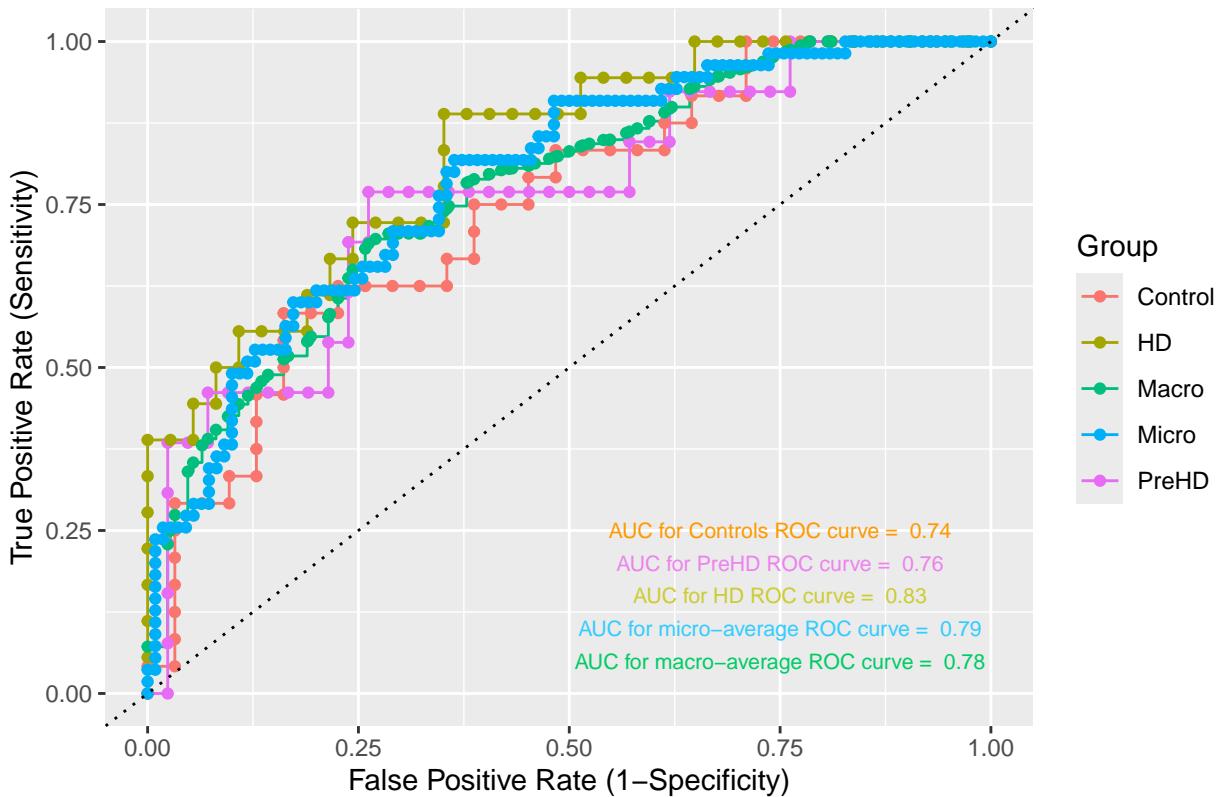
```

```

nrow=nrow(PP_MultinomialLogRegLATR_DF))
for (i in 1:nrow(PP_MultinomialLogRegLATR_DF)){
  PC_LogRegLATR[i,]<-ifelse(PP_MultinomialLogRegLATR_DF[i,]==max(PP_MultinomialLogRegLATR_DF[i,]),1,0)
}
colnames(PC_LogRegLATR)<-colnames(PP_MultinomialLogRegLATR_DF)
## ROC curves one vs the rest by multiROC package
Control_pred_LogRegLATR<- PP_MultinomialLogRegLATR_DF$Control
PreHD_pred_LogRegLATR<- PP_MultinomialLogRegLATR_DF$PreHD
HD_pred_LogRegLATR<- PP_MultinomialLogRegLATR_DF$HD
Control_true<-ifelse(LATR_avg_demo$Subgroup=='Control',1,0)
PreHD_true<-ifelse(LATR_avg_demo$Subgroup=='PreHD',1,0)
HD_true<-ifelse(LATR_avg_demo$Subgroup=='HD',1,0)
DF_ROC_LogRegLATR<-cbind(Control_true,PreHD_true,HD_true,Control_pred_LogRegLATR,
                           PreHD_pred_LogRegLATR,HD_pred_LogRegLATR)
multiROC_LogRegLATR<-multiROC::multi_roc(DF_ROC_LogRegLATR)
multiROC_LogRegLATR_DF<-plot_roc_data(multiROC_LogRegLATR)
ggplot(multiROC_LogRegLATR_DF, aes(x=1-Specificity, y=Sensitivity, color = Group)) +
  geom_step() + geom_point() + labs(title= "ROC curve for Logistic Regression model of Left ATR with log-t")
  x = "False Positive Rate (1-Specificity)",
  y = "True Positive Rate (Sensitivity)" + geom_abline(intercept = 0, slope=1,linetype = 'dotted') +
  annotate('text',x=0.75, y=0.25,
           label=paste('AUC for Controls ROC curve = ',
                       round(mean(multiROC_LogRegLATR_DF$AUC[multiROC_LogRegLATR_DF$Group=='Control']),2)),
           color = '#FF9900',size = 8/.pt) +
  annotate('text',x=0.75, y=0.20,
           label=paste('AUC for PreHD ROC curve = ',
                       round(mean(multiROC_LogRegLATR_DF$AUC[multiROC_LogRegLATR_DF$Group=='PreHD']),2)),
           color = 'violet',size = 8/.pt) +
  annotate('text',x=0.75, y=0.15,
           label=paste('AUC for HD ROC curve = ',
                       round(mean(multiROC_LogRegLATR_DF$AUC[multiROC_LogRegLATR_DF$Group=='HD']),2)),
           color = '#CCCC33',size = 8/.pt) +
  annotate('text',x=0.75, y=0.10,
           label=paste('AUC for micro-average ROC curve = ',
                       round(mean(multiROC_LogRegLATR_DF$AUC[multiROC_LogRegLATR_DF$Group=='Micro']),2)),
           color = '#33CCFF',size = 8/.pt) +
  annotate('text',x=0.75, y=0.05,
           label=paste('AUC for macro-average ROC curve = ',
                       round(mean(multiROC_LogRegLATR_DF$AUC[multiROC_LogRegLATR_DF$Group=='Macro']),2)),
           color = '#00CC66',size = 8/.pt)

```

ROC curve for Logistic Regression model of Left ATR with log-transform o



### 6.2.3 Both ATR

Primero, sin transformación logarítmica de AD, RD y MD

```
library(nnet)
# Level of reference in the outcome variable
ATR_avg$Subgroup2<-relevel(ATR_avg$ATR_subgroup, ref = '1')
# Fit the model
fitMultinomialLogRegATR<-multinom(Subgroup2~ATR_Iron+ATR_AD+ATR_RD+ATR_MD+ATR_FA,
                                     data=ATR_avg)
# weights: 21 (12 variable)
initial value 120.847352
iter  10 value 113.941052
iter  20 value 113.651019
iter  30 value 111.981640
iter  40 value 111.736399
iter  50 value 111.715879
iter  60 value 111.677646
iter  70 value 111.672935
iter  80 value 111.607176
iter  90 value 111.583365
iter 100 value 111.576054
final  value 111.576054
stopped after 100 iterations
summary(fitMultinomialLogRegATR)
Call:
multinom(formula = Subgroup2 ~ ATR_Iron + ATR_AD + ATR_RD + ATR_MD +
ATR_FA, data = ATR_avg)
```

```

Coefficients:
(Intercept) ATR_Iron ATR_AD ATR_RD ATR_MD ATR_FA
2 -14.924088 213.21974 -5.697498 -3.601648 -4.300237 23.717279
3 4.187639 -79.60705 23.244210 13.556697 16.785833 -6.898475

Std. Errors:
(Intercept) ATR_Iron ATR_AD ATR_RD ATR_MD ATR_FA
2 5.944430 146.7788 1.451119 0.7141268 0.9597631 10.798156
3 4.918954 133.2334 1.204128 0.6182137 0.8134861 8.823841

Residual Deviance: 223.1521
AIC: 247.1521
# Wald test (z-test)
z_fitMultinomialLogRegATR<-summary(fitMultinomialLogRegATR)$coefficients/summary(fitMultinomialLogRegATR)$
round(z_fitMultinomialLogRegATR,4)
(Intercept) ATR_Iron ATR_AD ATR_RD ATR_MD ATR_FA
2 -2.5106 1.4527 -3.9263 -5.0434 -4.4805 2.1964
3 0.8513 -0.5975 19.3038 21.9288 20.6344 -0.7818
# p-value
p_fitMultinomialLogRegATR<-(1-pnorm(abs(z_fitMultinomialLogRegATR),0,1))*2
round(p_fitMultinomialLogRegATR,4)
(Intercept) ATR_Iron ATR_AD ATR_RD ATR_MD ATR_FA
2 0.0121 0.1463 1e-04 0 0.0281
3 0.3946 0.5502 0e+00 0 0.4343
# Tabla
sum_fitMultinomialLogRegATR<-summary(fitMultinomialLogRegATR)
tbl_fitMultinomialLogRegATR<-rbind(sum_fitMultinomialLogRegATR$coefficients,
sum_fitMultinomialLogRegATR$standard.errors,
round(z_fitMultinomialLogRegATR,4),
round(p_fitMultinomialLogRegATR,4))
kableExtra::kable(tbl_fitMultinomialLogRegATR, caption='Multinomial Logistic Regression for Both ATR (cont.',
pack_rows(index=c('Coefficients'=2, 'Standard Errors'=2, 'Z-test'=2, 'p-value'=2)) %>%
kable_styling(position = 'center', font_size = 8, latex_options='hold_position')

```

Cuadro 114: Multinomial Logistic Regression for Both ATR (controls as a reference)

	(Intercept)	ATR_Iron	ATR_AD	ATR_RD	ATR_MD	ATR_FA
<b>Coefficients</b>						
2	-14.924088	213.21974	-5.697498	-3.6016483	-4.3002370	23.717279
3	4.187639	-79.60705	23.244210	13.5566971	16.7858328	-6.898475
<b>Standard Errors</b>						
2	5.944430	146.77878	1.451119	0.7141268	0.9597631	10.798156
3	4.918954	133.23338	1.204128	0.6182137	0.8134861	8.823841
<b>Z-test</b>						
2	-2.510600	1.45270	-3.926300	-5.0434000	-4.4805000	2.196400
3	0.851300	-0.59750	19.303800	21.9288000	20.6344000	-0.781800
<b>p-value</b>						
2	0.012100	0.14630	0.000100	0.0000000	0.0000000	0.028100
3	0.394600	0.55020	0.000000	0.0000000	0.0000000	0.434300

```

# Evaluación del modelo
## Predicted probabilities
PP_MultinomialLogRegATR<-fitMultinomialLogRegATR$fitted.values
PP_MultinomialLogRegATR_DF<-as.data.frame(PP_MultinomialLogRegATR)
## Predicted classes

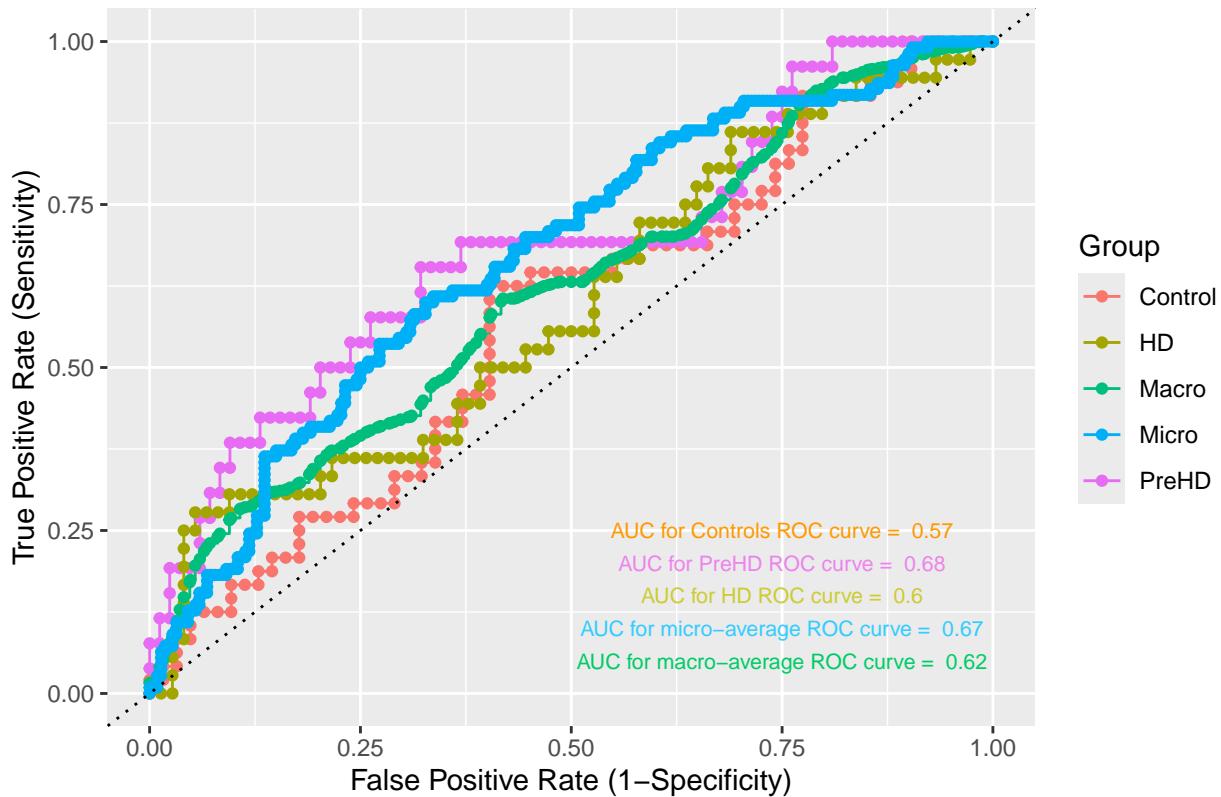
```

```

PC_LogRegATR<-matrix(ncol=ncol(PP_MultinomialLogRegATR_DF),
                      nrow=nrow(PP_MultinomialLogRegATR_DF))
for (i in 1:nrow(PP_MultinomialLogRegATR_DF)){
  PC_LogRegATR[i,]<-ifelse(PP_MultinomialLogRegATR_DF[i,]==max(PP_MultinomialLogRegATR_DF[i,]),1,0)
}
colnames(PC_LogRegATR)<-colnames(PP_MultinomialLogRegATR_DF)
## ROC curves one vs the rest by multiROC package
Control_pred_LogRegATR<- PP_MultinomialLogRegATR_DF$`1`
PreHD_pred_LogRegATR<- PP_MultinomialLogRegATR_DF$`2`
HD_pred_LogRegATR<- PP_MultinomialLogRegATR_DF$`3`
Control_true<-ifelse(ATR_avg$ATR_subgroup=='1',1,0)
PreHD_true<-ifelse(ATR_avg$ATR_subgroup=='2',1,0)
HD_true<-ifelse(ATR_avg$ATR_subgroup=='3',1,0)
DF_ROC_LogRegATR<-cbind(Control_true,PreHD_true,HD_true,Control_pred_LogRegATR,
                           PreHD_pred_LogRegATR,HD_pred_LogRegATR)
multiROC_LogRegATR<-multiROC::multi_roc(DF_ROC_LogRegATR)
multiROC_LogRegATR_DF<-plot_roc_data(multiROC_LogRegATR)
ggplot(multiROC_LogRegATR_DF, aes(x=1-Specificity, y=Sensitivity, color = Group)) +
  geom_step() + geom_point() + labs(title= "ROC curve for Logistic Regression model of Both ATR: One vs One",
    x = "False Positive Rate (1-Specificity)",
    y = "True Positive Rate (Sensitivity)") + geom_abline(intercept = 0, slope=1,linetype = 'dotted') +
  annotate('text',x=0.75, y=0.25,
    label=paste('AUC for Controls ROC curve = ',
                round(mean(multiROC_LogRegATR_DF$AUC[multiROC_LogRegATR_DF$Group=='Control']),2)),
    color = '#FF9900',size = 8/.pt) +
  annotate('text',x=0.75, y=0.20,
    label=paste('AUC for PreHD ROC curve = ',
                round(mean(multiROC_LogRegATR_DF$AUC[multiROC_LogRegATR_DF$Group=='PreHD']),2)),
    color = 'violet',size = 8/.pt) +
  annotate('text',x=0.75, y=0.15,
    label=paste('AUC for HD ROC curve = ',
                round(mean(multiROC_LogRegATR_DF$AUC[multiROC_LogRegATR_DF$Group=='HD']),2)),
    color = '#CCCC33',size = 8/.pt) +
  annotate('text',x=0.75, y=0.10,
    label=paste('AUC for micro-average ROC curve = ',
                round(mean(multiROC_LogRegATR_DF$AUC[multiROC_LogRegATR_DF$Group=='Micro']),2)),
    color = '#33CCFF',size = 8/.pt) +
  annotate('text',x=0.75, y=0.05,
    label=paste('AUC for macro-average ROC curve = ',
                round(mean(multiROC_LogRegATR_DF$AUC[multiROC_LogRegATR_DF$Group=='Macro']),2)),
    color = '#00CC66',size = 8/.pt)

```

## ROC curve for Logistic Regression model of Both ATR: One vs One



Segundo, con transformación logarítmica de AD, RD, MD

```

library(nnet)
# Fit the model
fitMultinomialLogRegATRLog<-multinom(Subgroup2~ATR_Iron+log(ATR_AD)+log(ATR_RD)+log(ATR_MD)+ATR_FA,
                                         data=ATR_avg)

# weights:  21 (12 variable)
initial  value 120.847352
iter   10 value 113.275892
iter   20 value 105.520905
iter   30 value 99.571641
iter   40 value 98.150386
iter   50 value 97.808603
iter   60 value 97.692114
iter   70 value 97.332459
iter   80 value 97.280087
iter   90 value 97.021547
iter  100 value 96.942218
final  value 96.942218
stopped after 100 iterations
summary(fitMultinomialLogRegATRLog)
Call:
multinom(formula = Subgroup2 ~ ATR_Iron + log(ATR_AD) + log(ATR_RD) +
log(ATR_MD) + ATR_FA, data = ATR_avg)

Coefficients:
(Intercept) ATR_Iron log(ATR_AD) log(ATR_RD) log(ATR_MD) ATR_FA
2  0.6388583 92.56587   30.03119   -19.72364   -8.923875 -22.95323
3 203.8244101 30.82737  -45.03236    46.02610   28.234153  96.70087

```

```

Std. Errors:
(Intercept) ATR_Iron log(ATR_AD) log(ATR_RD) log(ATR_MD) ATR_FA
2 46.33636 86.81598 28.40784 28.56422 1.675543 53.96637
3 47.93259 97.15833 30.81562 30.97779 2.002031 59.12645

Residual Deviance: 193.8844
AIC: 217.8844
# Wald test (z-test)
z_fitMultinomialLogRegATRLog<-summary(fitMultinomialLogRegATRLog)$coefficients/summary(fitMultinomialLogRe
round(z_fitMultinomialLogRegATRLog,4)
(Intercept) ATR_Iron log(ATR_AD) log(ATR_RD) log(ATR_MD) ATR_FA
2 0.0138 1.0662 1.0571 -0.6905 -5.3260 -0.4253
3 4.2523 0.3173 -1.4613 1.4858 14.1028 1.6355
# p-value
p_fitMultinomialLogRegATRLog<-(1-pnorm(abs(z_fitMultinomialLogRegATRLog),0,1))*2
round(p_fitMultinomialLogRegATRLog,4)
(Intercept) ATR_Iron log(ATR_AD) log(ATR_RD) log(ATR_MD) ATR_FA
2 0.989 0.2863 0.2904 0.4899 0 0.6706
3 0.000 0.7510 0.1439 0.1373 0 0.1019
# Tabla
sum_fitMultinomialLogRegATRLog<-summary(fitMultinomialLogRegATRLog)
tbl_fitMultinomialLogRegATRLog<-rbind(sum_fitMultinomialLogRegATRLog$coefficients,
                                         sum_fitMultinomialLogRegATRLog$standard.errors,
                                         round(z_fitMultinomialLogRegATRLog,4),
                                         round(p_fitMultinomialLogRegATRLog,4))
kableExtra::kable(tbl_fitMultinomialLogRegATRLog, caption='Multinomial Logistic Regression for Both ATR (c
pack_rows(index=c('Coefficients'=2, 'Standard Errors'=2, 'Z-test'=2, 'p-value'=2)) %>%
kable_styling(position = 'center', font_size = 8, latex_options='hold_position')

```

Cuadro 115: Multinomial Logistic Regression for Both ATR (controls as a reference) with log transform of diffusivities

	(Intercept)	ATR_Iron	log(ATR_AD)	log(ATR_RD)	log(ATR_MD)	ATR_FA
<b>Coefficients</b>						
2	0.6388583	92.56587	30.03119	-19.72364	-8.923875	-22.95323
3	203.8244101	30.82737	-45.03236	46.02610	28.234153	96.70087
<b>Standard Errors</b>						
2	46.3363647	86.81598	28.40784	28.56422	1.675543	53.96637
3	47.9325938	97.15833	30.81562	30.97779	2.002031	59.12645
<b>Z-test</b>						
2	0.0138000	1.06620	1.05710	-0.69050	-5.326000	-0.42530
3	4.2523000	0.31730	-1.46130	1.48580	14.102800	1.63550
<b>p-value</b>						
2	0.9890000	0.28630	0.29040	0.48990	0.000000	0.67060
3	0.0000000	0.75100	0.14390	0.13730	0.000000	0.10190

```

# Evaluación del modelo
## Predicted probabilities
PP_MultinomialLogRegATR<-fitMultinomialLogRegATRLog$fitted.values
PP_MultinomialLogRegATR_DF<-as.data.frame(PP_MultinomialLogRegATR)
## Predicted classes
PC_LogRegATR<-matrix(ncol=ncol(PP_MultinomialLogRegATR_DF),nrow=nrow(PP_MultinomialLogRegATR_DF))
for (i in 1:nrow(PP_MultinomialLogRegATR_DF)){
  PC_LogRegATR[i,]<-ifelse(PP_MultinomialLogRegATR_DF[i,]==max(PP_MultinomialLogRegATR_DF[i,]),1,0)
}
colnames(PC_LogRegATR)<-colnames(PP_MultinomialLogRegATR_DF)

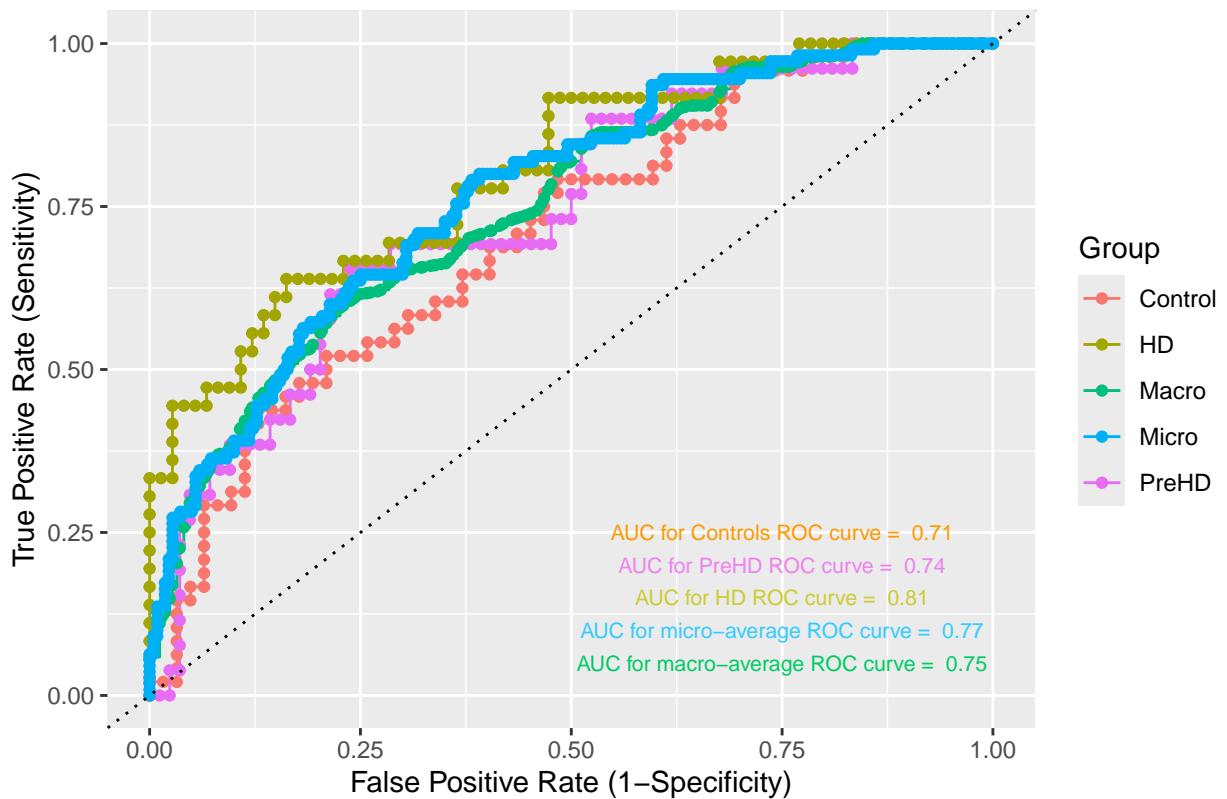
```

```

## ROC curves one vs the rest by multiROC package
Control_pred_LogRegATR<- PP_MultinomialLogRegATR_DF$`1` 
PreHD_pred_LogRegATR<- PP_MultinomialLogRegATR_DF$`2` 
HD_pred_LogRegATR<- PP_MultinomialLogRegATR_DF$`3` 
Control_true<-ifelse(ATR_avg$ATR_subgroup=='1',1,0)
PreHD_true<-ifelse(ATR_avg$ATR_subgroup=='2',1,0)
HD_true<-ifelse(ATR_avg$ATR_subgroup=='3',1,0)
DF_ROC_LogRegATR<-cbind(Control_true,PreHD_true,HD_true,Control_pred_LogRegATR,
                           PreHD_pred_LogRegATR,HD_pred_LogRegATR)
multiROC_LogRegATR<-multiROC::multi_roc(DF_ROC_LogRegATR)
multiROC_LogRegATR_DF<-plot_roc_data(multiROC_LogRegATR)
ggplot(multiROC_LogRegATR_DF, aes(x=1-Specificity, y=Sensitivity, color = Group)) +
  geom_step() + geom_point() + labs(title= "ROC curve for Logistic Regression model of Both ATR with log-likelihood",
    x = "False Positive Rate (1-Specificity)",
    y = "True Positive Rate (Sensitivity)") + geom_abline(intercept = 0, slope=1,linetype = 'dotted') +
  annotate('text',x=0.75, y=0.25,
    label=paste('AUC for Controls ROC curve = ',
    round(mean(multiROC_LogRegATR_DF$AUC[multiROC_LogRegATR_DF$Group=='Control']),2)),
    color = '#FF9900',size = 8/.pt) +
  annotate('text',x=0.75, y=0.20,
    label=paste('AUC for PreHD ROC curve = ',
    round(mean(multiROC_LogRegATR_DF$AUC[multiROC_LogRegATR_DF$Group=='PreHD']),2)),
    color = 'violet',size = 8/.pt) +
  annotate('text',x=0.75, y=0.15,
    label=paste('AUC for HD ROC curve = ',
    round(mean(multiROC_LogRegATR_DF$AUC[multiROC_LogRegATR_DF$Group=='HD']),2)),
    color = '#CCCC33',size = 8/.pt) +
  annotate('text',x=0.75, y=0.10,
    label=paste('AUC for micro-average ROC curve = ',
    round(mean(multiROC_LogRegATR_DF$AUC[multiROC_LogRegATR_DF$Group=='Micro']),2)),
    color = '#33CCFF',size = 8/.pt) +
  annotate('text',x=0.75, y=0.05,
    label=paste('AUC for macro-average ROC curve = ',
    round(mean(multiROC_LogRegATR_DF$AUC[multiROC_LogRegATR_DF$Group=='Macro']),2)),
    color = '#00CC66',size = 8/.pt)

```

ROC curve for Logistic Regression model of Both ATR with log-transform



## 7 Correlaciones clínicas

Generación del dataframe

```
library(ppcor)
RATR_avg_demo_GC<-RATR_avg_demo[RATR_avg_demo$Group=='Carriers',]
LATR_avg_demo_GC<-LATR_avg_demo[LATR_avg_demo$Group=='Carriers',]
GC_Demo_Clin_MRI<-cbind(demo_atr_PHD_HD[1:5], RATR_avg_demo_GC[6:7],
                           demo_atr_PHD_HD[6:11], clinical_vars_gc[6:17],
                           RATR_avg_demo_GC[8:12], LATR_avg_demo_GC[8:12])
```

### 7.1 Correlaciones parciales

#### 7.1.1 CAP controlado por género

```
[,1]
Iron_RATR "rho = -0.254 , p = 0.036"
Iron_LATR "rho = -0.083 , p = 0.301"
AD_RATR   "rho = 0.488 , p = 0.016"
AD_LATR   "rho = 0.434 , p = 0.041"
RD_RATR   "rho = 0.548 , p = 0.005"
RD_LATR   "rho = 0.515 , p = 0.01"
MD_RATR   "rho = 0.538 , p = 0.006"
MD_LATR   "rho = 0.517 , p = 0.009"
FA_RATR   "rho = -0.178 , p = 0.108"
FA_LATR   "rho = -0.2 , p = 0.081"
```

#### 7.1.2 Age\_onset controlado por género

```
[,1]
Iron_RATR "rho = 0.261 , p = 0.33"
Iron_LATR "rho = 0.252 , p = 0.356"
AD_RATR   "rho = -0.121 , p = 0.207"
AD_LATR   "rho = -0.118 , p = 0.214"
RD_RATR   "rho = -0.086 , p = 0.293"
RD_LATR   "rho = -0.109 , p = 0.235"
MD_RATR   "rho = -0.115 , p = 0.221"
MD_LATR   "rho = -0.105 , p = 0.245"
FA_RATR   "rho = -0.034 , p = 0.459"
FA_LATR   "rho = 0.025 , p = 1"
```

#### 7.1.3 UHDRS functional score controlado por género y CAP

```
[,1]
Iron_RATR "rho = 0.147 , p = 1"
Iron_LATR "rho = 0.159 , p = 1"
AD_RATR   "rho = 0.117 , p = 1"
AD_LATR   "rho = 0.295 , p = 0.499"
RD_RATR   "rho = -0.045 , p = 0.141"
RD_LATR   "rho = -0.126 , p = 0.043"
MD_RATR   "rho = 0.015 , p = 1"
MD_LATR   "rho = 0.043 , p = 1"
FA_RATR   "rho = 0.151 , p = 1"
FA_LATR   "rho = 0.248 , p = 0.705"
```

#### 7.1.4 UHDRS functional score controlado por género

```
[,1]
Iron_RATR "rho = 0.233 , p = 0.435"
Iron_LATR "rho = 0.179 , p = 0.638"
AD_RATR "rho = -0.106 , p = 0.242"
AD_LATR "rho = 0.066 , p = 1"
RD_RATR "rho = -0.258 , p = 0.035"
RD_LATR "rho = -0.309 , p = 0.014"
MD_RATR "rho = -0.208 , p = 0.074"
MD_LATR "rho = -0.178 , p = 0.109"
FA_RATR "rho = 0.208 , p = 0.524"
FA_LATR "rho = 0.303 , p = 0.233"
```

#### 7.1.5 UHDRS cognitive score controlado por género y CAP

```
[,1]
Iron_RATR "rho = -0.223 , p = 0.006"
Iron_LATR "rho = -0.115 , p = 0.047"
AD_RATR "rho = 0.115 , p = 1"
AD_LATR "rho = 0.011 , p = 1"
RD_RATR "rho = -0.156 , p = 0.023"
RD_LATR "rho = -0.242 , p = 0.003"
MD_RATR "rho = -0.102 , p = 0.058"
MD_LATR "rho = -0.208 , p = 0.008"
FA_RATR "rho = 0.212 , p = 0.938"
FA_LATR "rho = 0.226 , p = 0.872"
```

#### 7.1.6 UHDRS cognitive score controlado por género

```
[,1]
Iron_RATR "rho = -0.149 , p = 0.154"
Iron_LATR "rho = -0.092 , p = 0.275"
AD_RATR "rho = -0.017 , p = 0.502"
AD_LATR "rho = -0.093 , p = 0.272"
RD_RATR "rho = -0.256 , p = 0.038"
RD_LATR "rho = -0.324 , p = 0.012"
MD_RATR "rho = -0.211 , p = 0.073"
MD_LATR "rho = -0.295 , p = 0.02"
FA_RATR "rho = 0.245 , p = 0.431"
FA_LATR "rho = 0.262 , p = 0.377"
```

#### 7.1.7 UHDRS motor score controlado por género y CAP

```
[,1]
Iron_RATR "rho = -0.027 , p = 0.176"
Iron_LATR "rho = 0.115 , p = 1"
AD_RATR "rho = 0.087 , p = 1"
AD_LATR "rho = -0.063 , p = 0.112"
RD_RATR "rho = 0.189 , p = 0.989"
RD_LATR "rho = 0.229 , p = 0.794"
MD_RATR "rho = 0.169 , p = 1"
MD_LATR "rho = 0.139 , p = 1"
FA_RATR "rho = -0.15 , p = 0.028"
FA_LATR "rho = -0.162 , p = 0.022"
```

### 7.1.8 UHDRS motor score controlado por género

```
[,1]
Iron_RATR "rho = -0.198 , p = 0.084"
Iron_LATR "rho = 0.022 , p = 1"
AD_RATR "rho = 0.399 , p = 0.076"
AD_LATR "rho = 0.267 , p = 0.327"
RD_RATR "rho = 0.5 , p = 0.015"
RD_LATR "rho = 0.503 , p = 0.014"
MD_RATR "rho = 0.481 , p = 0.021"
MD_LATR "rho = 0.45 , p = 0.036"
FA_RATR "rho = -0.23 , p = 0.054"
FA_LATR "rho = -0.254 , p = 0.038"
```

### 7.1.9 Stroop\_interferences controlado por género y CAP

```
[,1]
Iron_RATR "rho = -0.277 , p = 0.001"
Iron_LATR "rho = -0.207 , p = 0.008"
AD_RATR "rho = 0.043 , p = 1"
AD_LATR "rho = -0.103 , p = 0.057"
RD_RATR "rho = -0.086 , p = 0.074"
RD_LATR "rho = -0.133 , p = 0.035"
MD_RATR "rho = -0.083 , p = 0.077"
MD_LATR "rho = -0.161 , p = 0.021"
FA_RATR "rho = 0.037 , p = 1"
FA_LATR "rho = 0.031 , p = 1"
```

### 7.1.10 Stroop\_interferences controlado por género

```
[,1]
Iron_RATR "rho = -0.219 , p = 0.065"
Iron_LATR "rho = -0.188 , p = 0.098"
AD_RATR "rho = -0.049 , p = 0.396"
AD_LATR "rho = -0.168 , p = 0.125"
RD_RATR "rho = -0.167 , p = 0.126"
RD_LATR "rho = -0.203 , p = 0.081"
MD_RATR "rho = -0.164 , p = 0.131"
MD_LATR "rho = -0.226 , p = 0.059"
FA_RATR "rho = 0.067 , p = 1"
FA_LATR "rho = 0.065 , p = 1"
```

### 7.1.11 Direct TMT(B-A) controlado por género y CAP

```
[,1]
Iron_RATR "rho = 0.062 , p = 1"
Iron_LATR "rho = 0.144 , p = 1"
AD_RATR "rho = 0.119 , p = 1"
AD_LATR "rho = 0.206 , p = 0.908"
RD_RATR "rho = 0.191 , p = 0.982"
RD_LATR "rho = 0.268 , p = 0.616"
MD_RATR "rho = 0.176 , p = 1"
MD_LATR "rho = 0.274 , p = 0.589"
FA_RATR "rho = -0.137 , p = 0.035"
FA_LATR "rho = -0.081 , p = 0.086"
```

### 7.1.12 Direct TMT(B-A) controlado por género

```
[,1]
Iron_RATR "rho = -0.13 , p = 0.189"
Iron_LATR "rho = 0.047 , p = 1"
AD_RATR "rho = 0.41 , p = 0.065"
AD_LATR "rho = 0.432 , p = 0.047"
RD_RATR "rho = 0.491 , p = 0.018"
RD_LATR "rho = 0.52 , p = 0.01"
MD_RATR "rho = 0.476 , p = 0.023"
MD_LATR "rho = 0.525 , p = 0.009"
FA_RATR "rho = -0.22 , p = 0.062"
FA_LATR "rho = -0.195 , p = 0.088"
```

### 7.1.13 PBA-ENROLL-Depression controlado por género y CAP

```
[,1]
Iron_RATR "rho = -0.247 , p = 0.003"
Iron_LATR "rho = 0.003 , p = 1"
AD_RATR "rho = -0.015 , p = 0.194"
AD_LATR "rho = -0.04 , p = 0.143"
RD_RATR "rho = -0.046 , p = 0.133"
RD_LATR "rho = -0.068 , p = 0.1"
MD_RATR "rho = -0.029 , p = 0.164"
MD_LATR "rho = -0.088 , p = 0.075"
FA_RATR "rho = 0.049 , p = 1"
FA_LATR "rho = 0.014 , p = 1"
```

### 7.1.14 PBA-ENROLL-Depression controlado por género

```
[,1]
Iron_RATR "rho = -0.245 , p = 0.044"
Iron_LATR "rho = 0.001 , p = 1"
AD_RATR "rho = -0.001 , p = 0.57"
AD_LATR "rho = -0.026 , p = 0.478"
RD_RATR "rho = -0.025 , p = 0.48"
RD_LATR "rho = -0.046 , p = 0.41"
MD_RATR "rho = -0.012 , p = 0.53"
MD_LATR "rho = -0.063 , p = 0.357"
FA_RATR "rho = 0.044 , p = 1"
FA_LATR "rho = 0.008 , p = 1"
```

### 7.1.15 PBA-ENROLL-Irritability controlado por género y CAP

```
[,1]
Iron_RATR "rho = -0.277 , p = 0.001"
Iron_LATR "rho = -0.062 , p = 0.109"
AD_RATR "rho = 0 , p = 0.229"
AD_LATR "rho = -0.12 , p = 0.046"
RD_RATR "rho = -0.014 , p = 0.196"
RD_LATR "rho = 0.015 , p = 1"
MD_RATR "rho = -0.019 , p = 0.184"
MD_LATR "rho = -0.089 , p = 0.074"
FA_RATR "rho = -0.105 , p = 0.058"
FA_LATR "rho = -0.062 , p = 0.108"
```

### 7.1.16 PBA-ENROLL-Irritability controlado por género

```
[,1]
Iron_RATR "rho = -0.286 , p = 0.023"
Iron_LATR "rho = -0.068 , p = 0.343"
AD_RATR "rho = 0.037 , p = 1"
AD_LATR "rho = -0.075 , p = 0.322"
RD_RATR "rho = 0.03 , p = 1"
RD_LATR "rho = 0.052 , p = 1"
MD_RATR "rho = 0.024 , p = 1"
MD_LATR "rho = -0.037 , p = 0.439"
FA_RATR "rho = -0.116 , p = 0.219"
FA_LATR "rho = -0.076 , p = 0.32"
```

### 7.1.17 PBA-ENROLL-Psychosis controlado por género y CAP

```
[,1]
Iron_RATR "rho = -0.099 , p = 0.061"
Iron_LATR "rho = 0.166 , p = 1"
AD_RATR "rho = 0.143 , p = 1"
AD_LATR "rho = 0.103 , p = 1"
RD_RATR "rho = -0.024 , p = 0.167"
RD_LATR "rho = -0.082 , p = 0.078"
MD_RATR "rho = 0.043 , p = 1"
MD_LATR "rho = 0 , p = 1"
FA_RATR "rho = 0.131 , p = 1"
FA_LATR "rho = 0.147 , p = 1"
```

### 7.1.18 PBA-ENROLL-Psychosis controlado por género

```
[,1]
Iron_RATR "rho = -0.114 , p = 0.224"
Iron_LATR "rho = 0.159 , p = 0.759"
AD_RATR "rho = 0.159 , p = 0.758"
AD_LATR "rho = 0.124 , p = 0.911"
RD_RATR "rho = 0.019 , p = 1"
RD_LATR "rho = -0.034 , p = 0.446"
MD_RATR "rho = 0.075 , p = 1"
MD_LATR "rho = 0.037 , p = 1"
FA_RATR "rho = 0.116 , p = 0.944"
FA_LATR "rho = 0.13 , p = 0.884"
```

### 7.1.19 PBA-ENROLL-Apathy controlado por género y CAP

```
[,1]
Iron_RATR "rho = -0.338 , p = 0"
Iron_LATR "rho = -0.145 , p = 0.03"
AD_RATR "rho = -0.07 , p = 0.098"
AD_LATR "rho = -0.102 , p = 0.061"
RD_RATR "rho = 0.153 , p = 1"
RD_LATR "rho = 0.204 , p = 0.946"
MD_RATR "rho = 0.084 , p = 1"
MD_LATR "rho = 0.065 , p = 1"
FA_RATR "rho = -0.391 , p = 0"
FA_LATR "rho = -0.357 , p = 0"
```

### 7.1.20 PBA-ENROLL-Apathy controlado por género

```
[,1]
Iron_RATR "rho = -0.337 , p = 0.009"
Iron_LATR "rho = -0.148 , p = 0.157"
AD_RATR "rho = -0.04 , p = 0.431"
AD_LATR "rho = -0.073 , p = 0.327"
RD_RATR "rho = 0.152 , p = 0.772"
RD_LATR "rho = 0.197 , p = 0.585"
MD_RATR "rho = 0.094 , p = 1"
MD_LATR "rho = 0.078 , p = 1"
FA_RATR "rho = -0.392 , p = 0.003"
FA_LATR "rho = -0.358 , p = 0.006"
```

### 7.1.21 PBA-ENROLL-Dysexecutive controlado por género y CAP

```
[,1]
Iron_RATR "rho = -0.133 , p = 0.037"
Iron_LATR "rho = -0.124 , p = 0.043"
AD_RATR "rho = 0.102 , p = 1"
AD_LATR "rho = -0.113 , p = 0.051"
RD_RATR "rho = 0.111 , p = 1"
RD_LATR "rho = 0.099 , p = 1"
MD_RATR "rho = 0.173 , p = 1"
MD_LATR "rho = 0.057 , p = 1"
FA_RATR "rho = -0.219 , p = 0.007"
FA_LATR "rho = -0.248 , p = 0.003"
```

### 7.1.22 PBA-ENROLL-Dysexecutive controlado por género

```
[,1]
Iron_RATR "rho = -0.166 , p = 0.126"
Iron_LATR "rho = -0.135 , p = 0.18"
AD_RATR "rho = 0.164 , p = 0.72"
AD_LATR "rho = -0.033 , p = 0.453"
RD_RATR "rho = 0.177 , p = 0.664"
RD_LATR "rho = 0.164 , p = 0.719"
MD_RATR "rho = 0.228 , p = 0.472"
MD_LATR "rho = 0.129 , p = 0.873"
FA_RATR "rho = -0.24 , p = 0.047"
FA_LATR "rho = -0.271 , p = 0.029"
```

Cuadro 116: Correlations between clinical measurements and Iron

	Iron	
	RATR	LATR
CAP controlled by Gender	$\rho = -0.254, p = 0.036$	$\rho = -0.083, p = 0.301$
Age_Onset controlled by Gender	$\rho = 0.261, p = 0.33$	$\rho = 0.252, p = 0.356$
UHDRS-Functional controlled by Gender and CAP	$\rho = 0.147, p = 1$	$\rho = 0.159, p = 1$
UHDRS-Functional controlled by Gender	$\rho = 0.233, p = 0.435$	$\rho = 0.179, p = 0.638$
UHDRS-Cognitive controlled by Gender and CAP	$\rho = -0.223, p = 0.006$	$\rho = -0.115, p = 0.047$
UHDRS-Cognitive controlled by Gender	$\rho = -0.149, p = 0.154$	$\rho = -0.092, p = 0.275$
UHDRS-Motor controlled by Gender and CAP	$\rho = -0.027, p = 0.176$	$\rho = 0.115, p = 1$
UHDRS-Motor controlled by Gender	$\rho = -0.198, p = 0.084$	$\rho = 0.022, p = 1$
Stroop-Interferences controlled by Gender and CAP	$\rho = -0.277, p = 0.001$	$\rho = -0.207, p = 0.008$
Stroop-Interferences controlled by Gender	$\rho = -0.219, p = 0.065$	$\rho = -0.188, p = 0.098$
Direct TMT controlled by Gender and CAP	$\rho = 0.062, p = 1$	$\rho = 0.144, p = 1$
Direct TMT controlled by Gender	$\rho = -0.13, p = 0.189$	$\rho = 0.047, p = 1$
PBA-Depression controlled by Gender and CAP	$\rho = -0.247, p = 0.003$	$\rho = 0.003, p = 1$
PBA-Depression controlled by Gender	$\rho = -0.245, p = 0.044$	$\rho = 0.001, p = 1$
PBA-Irritability controlled by Gender and CAP	$\rho = -0.277, p = 0.001$	$\rho = -0.062, p = 0.109$
PBA-Irritability controlled by Gender	$\rho = -0.286, p = 0.023$	$\rho = -0.068, p = 0.343$
PBA-Psychosis controlled by Gender and CAP	$\rho = -0.099, p = 0.061$	$\rho = 0.166, p = 1$
PBA-Psychosis controlled by Gender	$\rho = -0.114, p = 0.224$	$\rho = 0.159, p = 0.759$
PBA-Apathy controlled by Gender and CAP	$\rho = -0.338, p = 0$	$\rho = -0.145, p = 0.03$
PBA-Apathy controlled by Gender	$\rho = -0.337, p = 0.009$	$\rho = -0.148, p = 0.157$
PBA-Dysexecutive controlled by Gender and CAP	$\rho = -0.133, p = 0.037$	$\rho = -0.124, p = 0.043$
PBA-Dysexecutive controlled by Gender	$\rho = -0.166, p = 0.126$	$\rho = -0.135, p = 0.18$

Cuadro 117: Correlations between clinical measurements and AD

	AD	
	RATR	LATR
CAP controlled by Gender	$\rho = 0.488, p = 0.016$	$\rho = 0.434, p = 0.041$
Age_Onset controlled by Gender	$\rho = -0.121, p = 0.207$	$\rho = -0.118, p = 0.214$
UHDRS-Functional controlled by Gender and CAP	$\rho = 0.117, p = 1$	$\rho = 0.295, p = 0.499$
UHDRS-Functional controlled by Gender	$\rho = -0.106, p = 0.242$	$\rho = 0.066, p = 1$
UHDRS-Cognitive controlled by Gender and CAP	$\rho = 0.115, p = 1$	$\rho = 0.011, p = 1$
UHDRS-Cognitive controlled by Gender	$\rho = -0.017, p = 0.502$	$\rho = -0.093, p = 0.272$
UHDRS-Motor controlled by Gender and CAP	$\rho = 0.087, p = 1$	$\rho = -0.063, p = 0.112$
UHDRS-Motor controlled by Gender	$\rho = 0.399, p = 0.076$	$\rho = 0.267, p = 0.327$
Stroop-Interferences controlled by Gender and CAP	$\rho = 0.043, p = 1$	$\rho = -0.103, p = 0.057$
Stroop-Interferences controlled by Gender	$\rho = -0.049, p = 0.396$	$\rho = -0.168, p = 0.125$
Direct TMT controlled by Gender and CAP	$\rho = 0.119, p = 1$	$\rho = 0.206, p = 0.908$
Direct TMT controlled by Gender	$\rho = 0.41, p = 0.065$	$\rho = 0.432, p = 0.047$
PBA-Depression controlled by Gender and CAP	$\rho = -0.015, p = 0.194$	$\rho = -0.04, p = 0.143$
PBA-Depression controlled by Gender	$\rho = -0.001, p = 0.57$	$\rho = -0.026, p = 0.478$
PBA-Irritability controlled by Gender and CAP	$\rho = 0, p = 0.229$	$\rho = -0.12, p = 0.046$
PBA-Irritability controlled by Gender	$\rho = 0.037, p = 1$	$\rho = -0.075, p = 0.322$
PBA-Psychosis controlled by Gender and CAP	$\rho = 0.143, p = 1$	$\rho = 0.103, p = 1$
PBA-Psychosis controlled by Gender	$\rho = 0.159, p = 0.758$	$\rho = 0.124, p = 0.911$
PBA-Apathy controlled by Gender and CAP	$\rho = -0.07, p = 0.098$	$\rho = -0.102, p = 0.061$
PBA-Apathy controlled by Gender	$\rho = -0.04, p = 0.431$	$\rho = -0.073, p = 0.327$
PBA-Dysexecutive controlled by Gender and CAP	$\rho = 0.102, p = 1$	$\rho = -0.113, p = 0.051$
PBA-Dysexecutive controlled by Gender	$\rho = 0.164, p = 0.72$	$\rho = -0.033, p = 0.453$

Cuadro 118: Correlations between clinical measurements and RD

	RD	
	RATR	LATR
CAP controlled by Gender	$\rho = 0.548, p = 0.005$	$\rho = 0.515, p = 0.01$
Age_Onset controlled by Gender	$\rho = -0.086, p = 0.293$	$\rho = -0.109, p = 0.235$
UHDRS-Functional controlled by Gender and CAP	$\rho = -0.045, p = 0.141$	$\rho = -0.126, p = 0.043$
UHDRS-Functional controlled by Gender	$\rho = -0.258, p = 0.035$	$\rho = -0.309, p = 0.014$
UHDRS-Cognitive controlled by Gender and CAP	$\rho = -0.156, p = 0.023$	$\rho = -0.242, p = 0.003$
UHDRS-Cognitive controlled by Gender	$\rho = -0.256, p = 0.038$	$\rho = -0.324, p = 0.012$
UHDRS-Motor controlled by Gender and CAP	$\rho = 0.189, p = 0.989$	$\rho = 0.229, p = 0.794$
UHDRS-Motor controlled by Gender	$\rho = 0.5, p = 0.015$	$\rho = 0.503, p = 0.014$
Stroop-Interferences controlled by Gender and CAP	$\rho = -0.086, p = 0.074$	$\rho = -0.133, p = 0.035$
Stroop-Interferences controlled by Gender	$\rho = -0.167, p = 0.126$	$\rho = -0.203, p = 0.081$
Direct TMT controlled by Gender and CAP	$\rho = 0.191, p = 0.982$	$\rho = 0.268, p = 0.616$
Direct TMT controlled by Gender	$\rho = 0.491, p = 0.018$	$\rho = 0.52, p = 0.01$
PBA-Depression controlled by Gender and CAP	$\rho = -0.046, p = 0.133$	$\rho = -0.068, p = 0.1$
PBA-Depression controlled by Gender	$\rho = -0.025, p = 0.48$	$\rho = -0.046, p = 0.41$
PBA-Irritability controlled by Gender and CAP	$\rho = -0.014, p = 0.196$	$\rho = 0.015, p = 1$
PBA-Irritability controlled by Gender	$\rho = 0.03, p = 1$	$\rho = 0.052, p = 1$
PBA-Psychosis controlled by Gender and CAP	$\rho = -0.024, p = 0.167$	$\rho = -0.082, p = 0.078$
PBA-Psychosis controlled by Gender	$\rho = 0.019, p = 1$	$\rho = -0.034, p = 0.446$
PBA-Apathy controlled by Gender and CAP	$\rho = 0.153, p = 1$	$\rho = 0.204, p = 0.946$
PBA-Apathy controlled by Gender	$\rho = 0.152, p = 0.772$	$\rho = 0.197, p = 0.585$
PBA-Dysexecutive controlled by Gender and CAP	$\rho = 0.111, p = 1$	$\rho = 0.099, p = 1$
PBA-Dysexecutive controlled by Gender	$\rho = 0.177, p = 0.664$	$\rho = 0.164, p = 0.719$

Cuadro 119: Correlations between clinical measurements and MD

	MD	
	RATR	LATR
CAP controlled by Gender	$\rho = 0.538, p = 0.006$	$\rho = 0.517, p = 0.009$
Age_Onset controlled by Gender	$\rho = -0.115, p = 0.221$	$\rho = -0.105, p = 0.245$
UHDRS-Functional controlled by Gender and CAP	$\rho = 0.015, p = 1$	$\rho = 0.043, p = 1$
UHDRS-Functional controlled by Gender	$\rho = -0.208, p = 0.074$	$\rho = -0.178, p = 0.109$
UHDRS-Cognitive controlled by Gender and CAP	$\rho = -0.102, p = 0.058$	$\rho = -0.208, p = 0.008$
UHDRS-Cognitive controlled by Gender	$\rho = -0.211, p = 0.073$	$\rho = -0.295, p = 0.02$
UHDRS-Motor controlled by Gender and CAP	$\rho = 0.169, p = 1$	$\rho = 0.139, p = 1$
UHDRS-Motor controlled by Gender	$\rho = 0.481, p = 0.021$	$\rho = 0.45, p = 0.036$
Stroop-Interferences controlled by Gender and CAP	$\rho = -0.083, p = 0.077$	$\rho = -0.161, p = 0.021$
Stroop-Interferences controlled by Gender	$\rho = -0.164, p = 0.131$	$\rho = -0.226, p = 0.059$
Direct TMT controlled by Gender and CAP	$\rho = 0.176, p = 1$	$\rho = 0.274, p = 0.589$
Direct TMT controlled by Gender	$\rho = 0.476, p = 0.023$	$\rho = 0.525, p = 0.009$
PBA-Depression controlled by Gender and CAP	$\rho = -0.029, p = 0.164$	$\rho = -0.088, p = 0.075$
PBA-Depression controlled by Gender	$\rho = -0.012, p = 0.53$	$\rho = -0.063, p = 0.357$
PBA-Irritability controlled by Gender and CAP	$\rho = -0.019, p = 0.184$	$\rho = -0.089, p = 0.074$
PBA-Irritability controlled by Gender	$\rho = 0.024, p = 1$	$\rho = -0.037, p = 0.439$
PBA-Psychosis controlled by Gender and CAP	$\rho = 0.043, p = 1$	$\rho = 0, p = 1$
PBA-Psychosis controlled by Gender	$\rho = 0.075, p = 1$	$\rho = 0.037, p = 1$
PBA-Apathy controlled by Gender and CAP	$\rho = 0.084, p = 1$	$\rho = 0.065, p = 1$
PBA-Apathy controlled by Gender	$\rho = 0.094, p = 1$	$\rho = 0.078, p = 1$
PBA-Dysexecutive controlled by Gender and CAP	$\rho = 0.173, p = 1$	$\rho = 0.057, p = 1$
PBA-Dysexecutive controlled by Gender	$\rho = 0.228, p = 0.472$	$\rho = 0.129, p = 0.873$

Cuadro 120: Correlations between clinical measurements and FA

	FA	
	RATR	LATR
CAP controlled by Gender	$\rho = -0.178, p = 0.108$	$\rho = -0.2, p = 0.081$
Age_Onset controlled by Gender	$\rho = -0.034, p = 0.459$	$\rho = 0.025, p = 1$
UHDRS-Functional controlled by Gender and CAP	$\rho = 0.151, p = 1$	$\rho = 0.248, p = 0.705$
UHDRS-Functional controlled by Gender	$\rho = 0.208, p = 0.524$	$\rho = 0.303, p = 0.233$
UHDRS-Cognitive controlled by Gender and CAP	$\rho = 0.212, p = 0.938$	$\rho = 0.226, p = 0.872$
UHDRS-Cognitive controlled by Gender	$\rho = 0.245, p = 0.431$	$\rho = 0.262, p = 0.377$
UHDRS-Motor controlled by Gender and CAP	$\rho = -0.15, p = 0.028$	$\rho = -0.162, p = 0.022$
UHDRS-Motor controlled by Gender	$\rho = -0.23, p = 0.054$	$\rho = -0.254, p = 0.038$
Stroop-Interferences controlled by Gender and CAP	$\rho = 0.037, p = 1$	$\rho = 0.031, p = 1$
Stroop-Interferences controlled by Gender	$\rho = 0.067, p = 1$	$\rho = 0.065, p = 1$
Direct TMT controlled by Gender and CAP	$\rho = -0.137, p = 0.035$	$\rho = -0.081, p = 0.086$
Direct TMT controlled by Gender	$\rho = -0.22, p = 0.062$	$\rho = -0.195, p = 0.088$
PBA-Depression controlled by Gender and CAP	$\rho = 0.049, p = 1$	$\rho = 0.014, p = 1$
PBA-Depression controlled by Gender	$\rho = 0.044, p = 1$	$\rho = 0.008, p = 1$
PBA-Irritability controlled by Gender and CAP	$\rho = -0.105, p = 0.058$	$\rho = -0.062, p = 0.108$
PBA-Irritability controlled by Gender	$\rho = -0.116, p = 0.219$	$\rho = -0.076, p = 0.32$
PBA-Psychosis controlled by Gender and CAP	$\rho = 0.131, p = 1$	$\rho = 0.147, p = 1$
PBA-Psychosis controlled by Gender	$\rho = 0.116, p = 0.944$	$\rho = 0.13, p = 0.884$
PBA-Apathy controlled by Gender and CAP	$\rho = -0.391, p = 0$	$\rho = -0.357, p = 0$
PBA-Apathy controlled by Gender	$\rho = -0.392, p = 0.003$	$\rho = -0.358, p = 0.006$
PBA-Dysexecutive controlled by Gender and CAP	$\rho = -0.219, p = 0.007$	$\rho = -0.248, p = 0.003$
PBA-Dysexecutive controlled by Gender	$\rho = -0.24, p = 0.047$	$\rho = -0.271, p = 0.029$

Cuadro 121: Correlations between clinical measurements and MRI metrics  
(Iron, FA)

	RATR - IRON			LATR - IRON			RATR - FA			LATR - FA		
	rho	p	p-adj	rho	p	p-adj	rho	p	p-adj	rho	p	p-adj
CAP (Gender)	-0.254	0.036	0.090	-0.083	0.301	0.653	-0.178	0.108	0.238	-0.200	0.081	0.215
Age_Onset (Gender)	0.261	0.330	0.382	0.252	0.356	0.653	-0.034	0.459	0.777	0.025	1.000	1.000
UHDRS-Functional (Gender,CAP)	0.147	1.000	1.000	0.159	1.000	1.000	0.151	1.000	1.000	0.248	0.705	1.000
UHDRS-Functional (Gender)	0.233	0.435	0.478	0.179	0.638	1.000	0.208	0.524	0.823	0.303	0.233	0.466
UHDRS-Cognitive (Gender,CAP)	-0.223	0.006	0.025	-0.115	0.047	0.257	0.212	0.938	1.000	0.226	0.872	1.000
UHDRS-Cognitive (Gender)	-0.149	0.154	0.226	-0.092	0.275	0.653	0.245	0.431	0.777	0.262	0.377	0.638
UHDRS-Motor (Gender,CAP)	-0.027	0.176	0.242	0.115	1.000	1.000	-0.150	0.028	0.152	-0.162	0.022	0.124
UHDRS-Motor (Gender)	-0.198	0.084	0.142	0.022	1.000	1.000	-0.230	0.054	0.152	-0.254	0.038	0.138
Stroop-Interferences (Gender,CAP)	-0.277	0.001	0.011	-0.207	0.008	0.180	0.037	1.000	1.000	0.031	1.000	1.000
Stroop-Interferences (Gender)	-0.219	0.065	0.120	-0.188	0.098	0.398	0.067	1.000	1.000	0.065	1.000	1.000
Direct TMT (Gender,CAP)	0.062	1.000	1.000	0.144	1.000	1.000	-0.137	0.035	0.152	-0.081	0.086	0.215
Direct TMT (Gender)	-0.130	0.189	0.244	0.047	1.000	1.000	-0.220	0.062	0.152	-0.195	0.088	0.215
PBA-Depression (Gender,CAP)	-0.247	0.003	0.018	0.003	1.000	1.000	0.049	1.000	1.000	0.014	1.000	1.000
PBA-Depression (Gender)	-0.245	0.044	0.097	0.001	1.000	1.000	0.044	1.000	1.000	0.008	1.000	1.000
PBA-Irritability (Gender,CAP)	-0.277	0.001	0.011	-0.062	0.109	0.398	-0.105	0.058	0.152	-0.062	0.108	0.238
PBA-Irritability (Gender)	-0.286	0.023	0.072	-0.068	0.343	0.653	-0.116	0.219	0.437	-0.076	0.320	0.586
PBA-Psychosis (Gender,CAP)	-0.099	0.061	0.120	0.166	1.000	1.000	0.131	1.000	1.000	0.147	1.000	1.000
PBA-Psychosis (Gender)	-0.114	0.224	0.273	0.159	0.759	1.000	0.116	0.944	1.000	0.130	0.884	1.000
PBA-Apathy (Gender,CAP)	-0.338	0.000	0.004	-0.145	0.030	0.257	-0.391	0.000	0.000	-0.357	0.000	0.002
PBA-Apathy (Gender)	-0.337	0.009	0.032	-0.148	0.157	0.492	-0.392	0.003	0.029	-0.358	0.006	0.041
PBA-Dysexecutive (Gender,CAP)	-0.133	0.037	0.090	-0.124	0.043	0.257	-0.219	0.007	0.048	-0.248	0.003	0.035
PBA-Dysexecutive (Gender)	-0.166	0.126	0.198	-0.135	0.180	0.495	-0.240	0.047	0.152	-0.271	0.029	0.128

Cuadro 122: Correlations between clinical measurements and MRI metrics  
 (Diffusivities: AD, RD, MD)

	RATR - AD			LATR - AD			RATR - RD			LATR - RD			RATR - MD			LATR - MD		
	rho	p	p-adj															
CAP (Gender)	0.488	0.016	0.357	0.434	0.041	0.222	0.548	0.005	0.105	0.515	0.010	0.053	0.538	0.006	0.132	0.517	0.009	0.068
Age_Onset (Gender)	-0.121	0.207	0.667	-0.118	0.214	0.471	-0.086	0.293	0.496	-0.109	0.235	0.431	-0.115	0.221	0.443	-0.105	0.245	0.490
UHDRS-Functional (Gender,CAP)	0.117	1.000	1.000	0.295	0.499	0.646	-0.045	0.141	0.310	-0.126	0.043	0.117	0.015	1.000	1.000	0.043	1.000	1.000
UHDRS-Functional (Gender)	-0.106	0.242	0.667	0.066	1.000	1.000	-0.258	0.035	0.140	-0.309	0.014	0.053	-0.208	0.074	0.241	-0.178	0.109	0.240
UHDRS-Cognitive (Gender,CAP)	0.115	1.000	1.000	0.011	1.000	1.000	-0.156	0.023	0.127	-0.242	0.003	0.053	-0.102	0.058	0.241	-0.208	0.008	0.068
UHDRS-Cognitive (Gender)	-0.017	0.502	1.000	-0.093	0.272	0.514	-0.256	0.038	0.140	-0.324	0.012	0.053	-0.211	0.073	0.241	-0.295	0.020	0.093
UHDRS-Motor (Gender,CAP)	0.087	1.000	1.000	-0.063	0.112	0.342	0.189	0.989	1.000	0.229	0.794	0.971	0.169	1.000	1.000	0.139	1.000	1.000
UHDRS-Motor (Gender)	0.399	0.076	0.536	0.267	0.327	0.514	0.500	0.015	0.127	0.503	0.014	0.053	0.481	0.021	0.170	0.450	0.036	0.132
Stroop-Interferences (Gender,CAP)	0.043	1.000	1.000	-0.103	0.057	0.222	-0.086	0.074	0.231	-0.133	0.035	0.109	-0.083	0.077	0.241	-0.161	0.021	0.093
Stroop-Interferences (Gender)	-0.049	0.396	0.947	-0.168	0.125	0.342	-0.167	0.126	0.310	-0.203	0.081	0.178	-0.164	0.131	0.360	-0.226	0.059	0.183
Direct TMT (Gender,CAP)	0.119	1.000	1.000	0.206	0.908	1.000	0.191	0.982	1.000	0.268	0.616	0.847	0.176	1.000	1.000	0.274	0.589	0.925
Direct TMT (Gender)	0.410	0.065	0.536	0.432	0.047	0.222	0.491	0.018	0.127	0.520	0.010	0.053	0.476	0.023	0.170	0.525	0.009	0.068
PBA-Depression (Gender,CAP)	-0.015	0.194	0.667	-0.040	0.143	0.350	-0.046	0.133	0.310	-0.068	0.100	0.199	-0.029	0.164	0.401	-0.088	0.075	0.183
PBA-Depression (Gender)	-0.001	0.570	1.000	-0.026	0.478	0.646	-0.025	0.480	0.755	-0.046	0.410	0.695	-0.012	0.530	0.896	-0.063	0.357	0.655
PBA-Irritability (Gender,CAP)	0.000	0.229	0.667	-0.120	0.046	0.222	-0.014	0.196	0.360	0.015	1.000	1.000	-0.019	0.184	0.406	-0.089	0.074	0.183
PBA-Irritability (Gender)	0.037	1.000	1.000	-0.075	0.322	0.514	0.030	1.000	1.000	0.052	1.000	1.000	0.024	1.000	1.000	-0.037	0.439	0.743
PBA-Psychosis (Gender,CAP)	0.143	1.000	1.000	0.103	1.000	1.000	-0.024	0.167	0.335	-0.082	0.078	0.178	0.043	1.000	1.000	0.000	1.000	1.000
PBA-Psychosis (Gender)	0.159	0.758	1.000	0.124	0.911	1.000	0.019	1.000	1.000	-0.034	0.446	0.701	0.075	1.000	1.000	0.037	1.000	1.000
PBA-Apathy (Gender,CAP)	-0.070	0.098	0.536	-0.102	0.061	0.222	0.153	1.000	1.000	0.204	0.946	1.000	0.084	1.000	1.000	0.065	1.000	1.000
PBA-Apathy (Gender)	-0.040	0.431	0.947	-0.073	0.327	0.514	0.152	0.772	1.000	0.197	0.585	0.847	0.094	1.000	1.000	0.078	1.000	1.000
PBA-Dysexecutive (Gender,CAP)	0.102	1.000	1.000	-0.113	0.051	0.222	0.111	1.000	1.000	0.099	1.000	1.000	0.173	1.000	1.000	0.057	1.000	1.000
PBA-Dysexecutive (Gender)	0.164	0.720	1.000	-0.033	0.453	0.646	0.177	0.664	0.974	0.164	0.719	0.931	0.228	0.472	0.865	0.129	0.873	1.000