

# Dac summary 01/19/2021

Erica Baller

9/29/2021

title: "Dac Summary" output: html\_document: default pdf\_document: default —

## Dac summary

Welcome to the MS project! This is a really exciting project dedicated to understanding psychopathology, particular depression in patients with MS. It is unique in a variety of ways, primarily that it uses clinical images that are research grade.

The team so far is big. PIs- Ted Satterthwaite and Taki Shinohara Senior Scientists - Azeez Adebimpe and Matt Cieslak Research Specialist- Timothy Robert-Fitzgerald Graduate students - Melissa Martin Data Analyst - Sidney Covitz Collaborators - Aaron Alexander-Bloch and Jenna Young DAC Pull done by - Victoria Rautman(2020/2021) and Sunil Thomas (2018)

Date of FINAL Pull: 1/19/2021 Spreadsheet: /Users/eballer/BBL/msdepression/data/dac/investigatingdepressioninmspatient  
Summary of request to Victoria @ DAC: — 9/2/2020

DAC Request Non-funded IRB Approved Research (enter IRB number): 843669 Criteria Display? Description / Exclusions / Limitations / Filters MRN Y Unique identifier requested, but can be masked MRN Visit ID Y

Patient Class(es) Please select only which class(es) you will need.  Inpatient  Outpatient  Emergency Age or DOB ranges Y

Gender Y

Race Y

Department(s) Provide department numbers not just names. N

Provider(s) Provide ID's not just names. N

Date(s) Include in the specific range and date types (eg, admit, order, result) 1/1/2010- 12/31/20 All scans based on date of first treatment with glatiramer acetate Procedure Please include the specific procedure codes. (ICD9/ICD10 is preferred for inpatient) Y 88.91 (Magnetic Resonance Imaging of Brain)

Diagnosis Please include the specific ICD-9/ICD-10 codes including all decimal points. Do not simply include ranges or wildcards. Y Multiple sclerosis (ICD-9: 340; ICD-10: G35), Depressive disorders (ICD-9: 296.99, 296.21, 296.22, 296.23, 296.24, 296.25, 296.26, 296.20, 296.31, 296.32, 296.33, 296.34, 296.35, 296.26, 296.30, 300.4, 293.83, 311, ICD-10: F32.0, F32.1, F32.2, F32.3, F32.4, F32.5, F32.6, F33.0, F33.1, F33.2, F33.41, F33.42, F33.9, F34.1, F06.31, F06.32, F06.34, F32.8, F32.9, major depressive disorder, persistent depressive disorder, depressive disorder due to another medical condition, other specified depressive disorder, unspecified depressive disorder) Orders N

Medication Please list as it is ordered within the UPHS EMR's – medication id's preferred Y Glatiramer Acetate (Copaxone) Lab Result Please list the lab as it is ordered within the UPHS EMR's. Y

Other Y EDSS scores (if present) Other Y All medications patients are receiving Other Y All PACS accession numbers available Other Y PHQ-2 and PHQ-9 if present — 9/16/2020: Per her note to me: We also specified that we wanted to pull images with ORIG\_CPT values of 70551, 70552, 70553, MHDI, IMGMR0128, and we wanted scans with department ID 361, 547, 6004, or 6013. And I may have mentioned this before but the ICD9 and ICD10 codes for MS patients are 340 and G35, respectively

\*For us, we are additionally interested in ALL ICD9/10 codes, ALL medications they are on at the time of scan, specific labs (as indicated), EDSS , PHQ-9 and 2 scores. The particular reason for this is that we are going to be looking at the relationship of brain imaging to depression in the MS group.

Mock columns: DE\_PAT\_ID PAT\_ID EMPI HUP\_MRN PAT\_NM\_WMRN SEX RACE ETHNIC\_GROUP ORDER\_ID ORDERING\_DTTM BEGIN\_EXAM\_DTTM END\_EXAM\_DTTM TECH\_USER\_ID TECHNOLOGIST ACCESSION\_NUM PROC\_ID PROC\_NAME PROC\_CODE ORIG\_CPT PERFORMING\_CSN\_ID COVERAGE\_ID FIN\_CLASS\_NAME VISIT\_EPM\_ID PERFORMING\_DEP\_ID DEPARTMENT\_NAME MODALITY ASSOCIATED\_ICD9 ASSOCIATED\_ICD10 PAT AGE\_AT\_EXAM MRI\_ENC AGE CURRENT AGE All ICD-9 or 10 codes All medication codes wbc rbc hemoglobin csf studies... phq-9 score (and date) phq-2 score (and date) b12 folate TSH RPR Vit D

\_\_\_\_\_  
1/8/2021  
email  
up-  
date  
(me  
to  
Victo-  
ria  
Raut-  
man):  
What  
we re-  
alized  
is that  
the  
best  
way  
to  
define  
inclus-  
ion  
for  
the  
MS  
group  
is to  
focus  
on  
people  
who  
were  
seen  
at  
some  
point  
in  
neu-  
rology  
clinic.

— If you'd be so kind, I'd be grateful for \*hopefully the last pull.

We would use the previous one, but instead of using the ICD 9/10 codes associated with the actual scan, we would use the location of clinic visits, and pull ALL scans, before or after diagnosis.

—  
New  
up-  
dates:  
-Pull  
all the  
scans  
for  
pa-  
tients  
who  
have  
ever  
been  
seen  
by  
“NEU-  
ROL-  
OGY  
SOUTH  
PAVIL-  
LION”  
and  
“NEU-  
ROL-  
OGY  
HUP”  
de-  
part-  
ments  
for  
clinic.

-  
Please  
in-  
clude  
provider  
name  
asso-  
ciated  
with  
these  
de-  
part-  
ment  
visits.

-  
Don't  
worry  
about  
doing  
exclu-  
sions  
based  
on  
ICD9/10  
codes  
for  
the  
scans  
them-  
selves,  
please  
in-  
clude  
all  
and  
we  
will  
sort  
through  
it our-  
selves  
*Please*  
*also*  
*in-*  
*clude*  
*vita-*  
*min*  
*D*  
*levels*  
*in*  
*addi-*  
*tion*  
*to*  
*other*  
*labs*

The rest of the query would be the same as the most recent query you sent back.

Radiology sent images to Taki and group 4/2021

This script goes through data that has been returned from the DAC on 1/19 and summarizes it

```
#preprocessing - because these are sent

#load our df

#from Jan 2021 pull
data <- read.csv("/Users/eballer/BBL/msdepression/data/dac/investigatingdepressioninmspatients_dates_rig
ms_providers <- c("MARKOWITZ, CLYDE E.", "JACOBS, DINA A.", "WILLIAMSON, ERIC MICHAEL-LEE", "BERGER, JO

#keep people with MS providers
data_empi_acc_f_phq <- data[which(data$Provider %in% ms_providers),] #n = 17067

#keep people with ICD10 code G35 (this is the same # as people with IC9 - 340)
data_empi_acc_f_phq <- data_empi_acc_f_phq[grep("G35", data_empi_acc_f_phq$ICD10),] #n = 16830
table(data_empi_acc_f_phq$Provider)

##          BAR-OR, AMIT        BERGER DO, JOSEPH
##                93                      1
##      BERGER, JOSEPH ROBERT       CHAHIN, SALIM
##                  1818                   260
##          JACOBS, DINA A.        KOLSON, DENNIS L.
##                  4765                   643
##      MARKOWITZ, CLYDE E.        NARULA, SONA
##                  6245                   274
##      PRUITT, AMY A. WILLIAMSON, ERIC MICHAEL-LEE
##                  1372                  1359

#####
### some preprocessing #####
#####
```



```

print(paste0("Number of unique modalities (ie. HUP PCAM GR MR4 3T; MR6 3T): ", length(unique(data_empi_f_phq$MODALITY)))

## [1] "Number of unique modalities (ie. HUP PCAM GR MR4 3T; MR6 3T): 11"

print(paste0("Number of unique accession #s n= ", length(unique(data_empi_acc_f_phq$ACCESSION_NUM))) #n = 16830

## [1] "Number of unique accession #s n= 16830"

print(paste0("Number of unique EMPIS n = ", length(unique(data_empi_acc_f_phq$EMPI))) #n = 3737

## [1] "Number of unique EMPIS n = 3737"

##### PHQ-2 and PHQ-9 summaries

#looking at just people with phq-2 and 9
phq2_subset <- data_empi_acc_f_phq[which(!is.na(data_empi_acc_f_phq$PHQ.2)),] #n=4606
phq9_subset <- data_empi_acc_f_phq[which(!is.na(data_empi_acc_f_phq$PHQ.9)),] #n= 261

#phq-2;
print(paste0("N PHQ-2 = ", sum(!is.na(data_empi_acc_f_phq$PHQ.2)), "; n = ", length(which(phq2_subset$PHQ.2 >= 3)))

## [1] "N PHQ-2 = 4606; n = 130 meeting criteria for depression (PHQ2) >= 3"

#phq-9
print(paste0("N PHQ-9 = ", sum(!is.na(data_empi_acc_f_phq$PHQ.9)), "; noDep/mild/mod/mod-severe/severe = ",
            length(which(phq9_subset$PHQ.9 < 5)), "/",
            length(which(phq9_subset$PHQ.9 >= 5 & phq9_subset$PHQ.9 < 10)), "/",
            length(which(phq9_subset$PHQ.9 >= 10 & phq9_subset$PHQ.9 < 15)), "/",
            length(which(phq9_subset$PHQ.9 >= 15 & phq9_subset$PHQ.9 < 20)), "/",
            length(which(phq9_subset$PHQ.9 >= 20)))) #n = 261

## [1] "N PHQ-9 = 261; noDep/mild/mod/mod-severe/severe = 77/73/65/21/25"

#both phq2 and 9

num_both_phq2_and_9 <- length(which(!is.na(data_empi_acc_f_phq$PHQ.2) & !is.na(data_empi_acc_f_phq$PHQ.9)))
print(paste0("num people with both phq2 and 9 = ", num_both_phq2_and_9)) #n = 0

## [1] "num people with both phq2 and 9 = 0"

#Number of unique EMPIs in phq2/phq9 subsets
unique_phq2_empis <- length(unique(phq2_subset$EMPI)) #n=966
print(paste0("N with phq2 (Unique EMPI) n = ", unique_phq2_empis, " out of ", dim(phq2_subset)[1], "; n = ", length(unique(data_empi_f_phq$EMPI))) #n = 966

## [1] "N with phq2 (Unique EMPI) n = 899 out of 4606; n - 32 meeting criteria for depression (PHQ2) >= 3"

```

```

unique_phq9.empis <- length(unique(phq9_subset$EMPI)) #n=76
print(paste0("N with phq9 (Unique EMPI) n = ", unique_phq9.empis, " out of ", dim(phq9_subset)[1], "; n = ", length(unique(phq9_subset$EMPI[phq9_subset$PHQ.9 < 5])), "/", length(unique(phq9_subset$EMPI[(phq9_subset$PHQ.9 >= 5 & phq9_subset$PHQ.9 < 10)])), "/", length(unique(phq9_subset$EMPI[(phq9_subset$PHQ.9 >= 10 & phq9_subset$PHQ.9 < 15)])), "/", length(unique(phq9_subset$EMPI[(phq9_subset$PHQ.9 >= 15 & phq9_subset$PHQ.9 < 20)])), "/", length(unique(phq9_subset$EMPI[phq9_subset$PHQ.9 >= 20]))))

## [1] "N with phq9 (Unique EMPI) n = 68 out of 261; noDep/mild/mod/mod-severe/severe = 18/16/18/8/8"

##### ICD 9/10 for Depression
dep_df <- data_empi_acc_f_phq[grep("F3", data_empi_acc_f_phq$ICD10),] #n = 2123
num_depression <- dim(dep_df)[1]
unique_empi <- which(unique(data_empi_acc_f_phq$EMPI) %in% data_empi_acc_f_phq$EMPI) #index of the location
num_unique_EMPI_and_dep <- length(unique(dep_df$EMPI)) #n = 493

print(paste0("N with ICD 9/10 codes for depression = ", num_depression))

## [1] "N with ICD 9/10 codes for depression = 2123"

print(paste0("N with ICD 9/10 codes for depression and unique EMPI = ", num_unique_EMPI_and_dep, " out of ", dim(data_empi_acc_f_phq)[1]))

## [1] "N with ICD 9/10 codes for depression and unique EMPI = 493 out of 16830"

##### PHQ2 or 9 + ICD 9/10 Depression
num_phq2_and_dep <- length(which(!is.na(dep_df$PHQ.2))) #n = 960
print(paste0("N with ICD 9/10 codes for depression and phq2 (ALL SCANS) = ", num_phq2_and_dep, " out of ", dim(data_empi_acc_f_phq)[1]))

## [1] "N with ICD 9/10 codes for depression and phq2 (ALL SCANS) = 960 out of 2123"

num_phq9_and_dep <- length(which(!is.na(dep_df$PHQ.9))) #n=140
print(paste0("N with ICD 9/10 codes for depression and phq9 (ALL SCANS) = ", num_phq9_and_dep, " out of ", dim(data_empi_acc_f_phq)[1]))

## [1] "N with ICD 9/10 codes for depression and phq9 (ALL SCANS) = 140 out of 2123"

#phq-2;
print(paste0("N dep + PHQ-2 = ", sum(!is.na(dep_df$PHQ.2)), "; n = ", length(which(dep_df$PHQ.2 >= 3)), " out of ", dim(data_empi_acc_f_phq)[1]))

## [1] "N dep + PHQ-2 = 960; n = 83 meeting criteria for depression (PHQ2) >= 3"

#phq-9
print(paste0("N dep + PHQ-9 = ", sum(!is.na(dep_df$PHQ.9)), "; noDep/mild/mod/mod-severe/severe = ", length(which(dep_df$PHQ.9 < 5)), "/", length(which(dep_df$PHQ.9 >= 5 & dep_df$PHQ.9 < 10)), "/", length(which(dep_df$PHQ.9 >= 10 & dep_df$PHQ.9 < 15)), "/", length(which(dep_df$PHQ.9 >= 15 & dep_df$PHQ.9 < 20)), "/", length(which(dep_df$PHQ.9 >= 20)))) #n = 261

## [1] "N dep + PHQ-9 = 140; noDep/mild/mod/mod-severe/severe = 22/39/44/20/15"

```

```

##### # unique people with dep_dx and phq2/9 -> working on this
dep_dx_phq2_unique <- length(unique(dep_df$EMPI[!is.na(dep_df$PHQ.2)])) #n = 201
print(paste0("N with ICD 9/10 codes for depression and phq2 (unique) = ", dep_dx_phq2_unique, " out of 2123"))

## [1] "N with ICD 9/10 codes for depression and phq2 (unique) = 201 out of 2123"

dep_dx_phq9_unique <- length(unique(dep_df$EMPI[!is.na(dep_df$PHQ.9)])) #n = 43
print(paste0("N with ICD 9/10 codes for depression and phq9 (unique) = ", dep_dx_phq9_unique, " out of 2123"))

## [1] "N with ICD 9/10 codes for depression and phq9 (unique) = 43 out of 2123"

#####

##### ICD 9/10 for Healthy Group
healthy_df <- data_empi_acc_f_phq[(-grep("F3", data_empi_acc_f_phq$ICD10)),] #n = 14707
num_healthy <- dim(healthy_df)[1]
unique_empi <- which(unique(healthy_df$EMPI) %in% healthy_df$EMPI) #index of the locations of unique empi
num_unique_EMPI_and_healthy <- length(unique(healthy_df$EMPI)) #n = 3244

print(paste0("N without Depression/ICD 9/10 codes for healthy = ", num_healthy))

## [1] "N without Depression/ICD 9/10 codes for healthy = 14707"

print(paste0("N without Depression/ ICD 9/10 codes for healthy and unique EMPI = ", num_unique_EMPI_and_healthy))

## [1] "N without Depression/ ICD 9/10 codes for healthy and unique EMPI = 3244 out of 14707"

##### PHQ2 or 9 + No Depression ICD9/10
num_phq2_and_healthy <- length(which(!is.na(healthy_df$PHQ.2))) #n = 3646
print(paste0("N without Depression/ ICD 9/10 codes for healthy and phq2 (ALL SCANS) = ", num_phq2_and_healthy))

## [1] "N without Depression/ ICD 9/10 codes for healthy and phq2 (ALL SCANS) = 3646 out of 14707"

#phq-2;
print(paste0("N healthy w/PHQ-2 = ", sum(!is.na(healthy_df$PHQ.2)), "; n = ", length(which(healthy_df$PHQ.2) == 2)))

## [1] "N healthy w/PHQ-2 = 3646; n = 47 meeting criteria for depression (PHQ2) >= 3"

num_phq9_and_healthy <- length(which(!is.na(healthy_df$PHQ.9))) #n=121
print(paste0("N without Depression/ ICD 9/10 codes for healthy and phq9 (ALL SCANS) = ", num_phq9_and_healthy))

## [1] "N without Depression/ ICD 9/10 codes for healthy and phq9 (ALL SCANS) = 121 out of 14707"

#phq-9
print(paste0("N healthy w/PHQ-9 = ", sum(!is.na(healthy_df$PHQ.9)), "; noDep/mild/mod/mod-severe/severe"))
length(which(healthy_df$PHQ.9 < 5)), "/",
length(which(healthy_df$PHQ.9 >= 5 & healthy_df$PHQ.9 < 10)), "/",
length(which(healthy_df$PHQ.9 >= 10 & healthy_df$PHQ.9 < 15)), "/",
length(which(healthy_df$PHQ.9 >= 15 & healthy_df$PHQ.9 < 20)), "/",
length(which(healthy_df$PHQ.9 >= 20))) #n = 55/34/21/1/10

```

```

## [1] "N healthy w/PHQ-9 = 121; noDep/mild/mod/mod-severe/severe = 55/34/21/1/10"

##### # unique people without dep dx and phq2/9
dep_dx_phq2_unique <- length(unique(healthy_df$EMPI[!is.na(healthy_df$PHQ.2)])) #n = 201
print(paste0("N without Depression/ ICD 9/10 codes for healthy and phq2 (unique) = ", dep_dx_phq2_unique))

## [1] "N without Depression/ ICD 9/10 codes for healthy and phq2 (unique) = 698 out of 14707"

dep_dx_phq9_unique <- length(unique(healthy_df$EMPI[!is.na(healthy_df$PHQ.9)])) #n = 43
print(paste0("N without Depression/ ICD 9/10 codes for healthy and phq9 (unique) = ", dep_dx_phq9_unique))

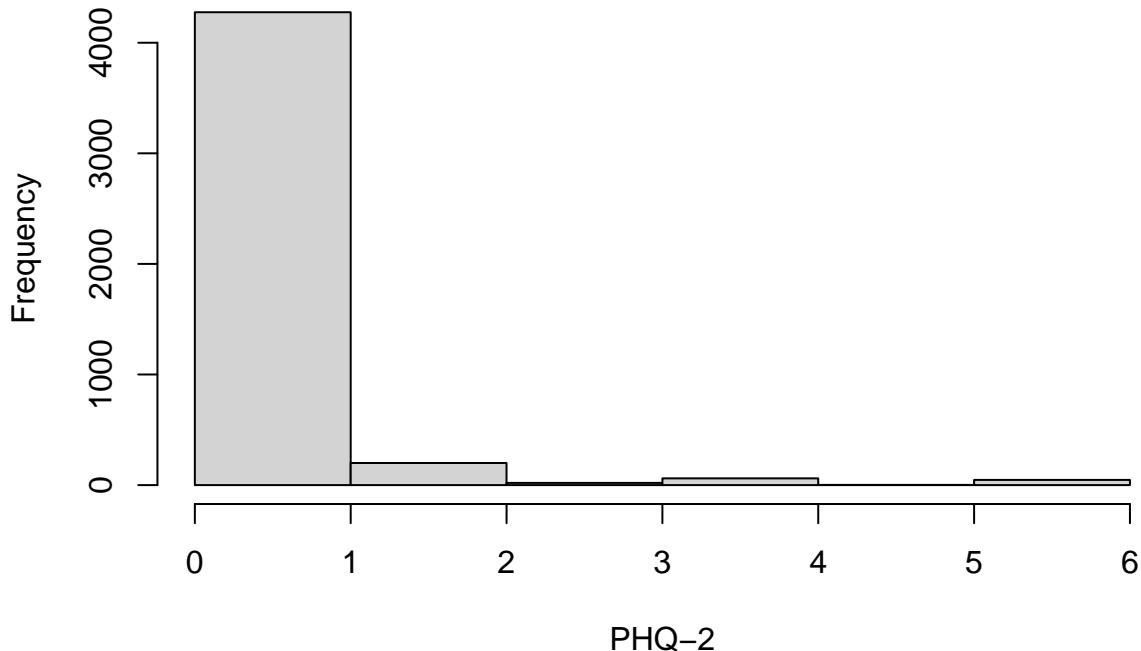
## [1] "N without Depression/ ICD 9/10 codes for healthy and phq9 (unique) = 25 out of 14707"

#####
#demographics and lab histograms
#print histogram of phq-2/9s for people

hist(data_empi_acc_f_phq$PHQ.2[which(!is.na(data_empi_acc_f_phq$PHQ.2))], main = paste0("PHQ-2 Histo : n = ", n))

```

### PHQ-2 Histo : n = 4606

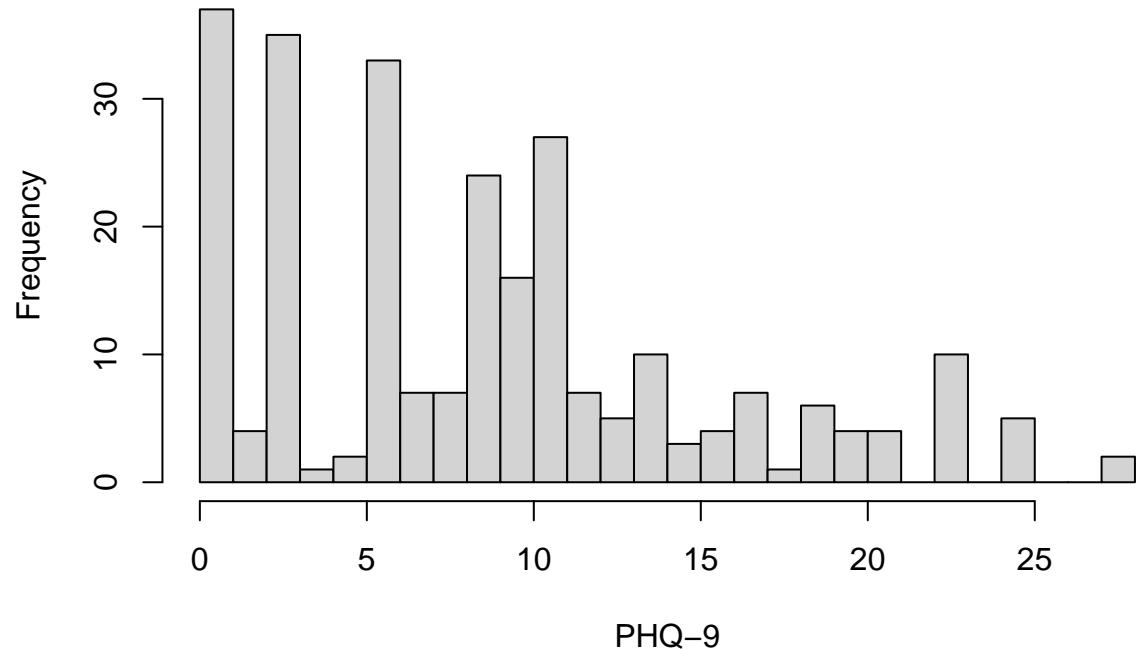


```

hist(data_empi_acc_f_phq$PHQ.9[which(!is.na(data_empi_acc_f_phq$PHQ.9))], main = paste0("PHQ-9 Histo : n = ", n))

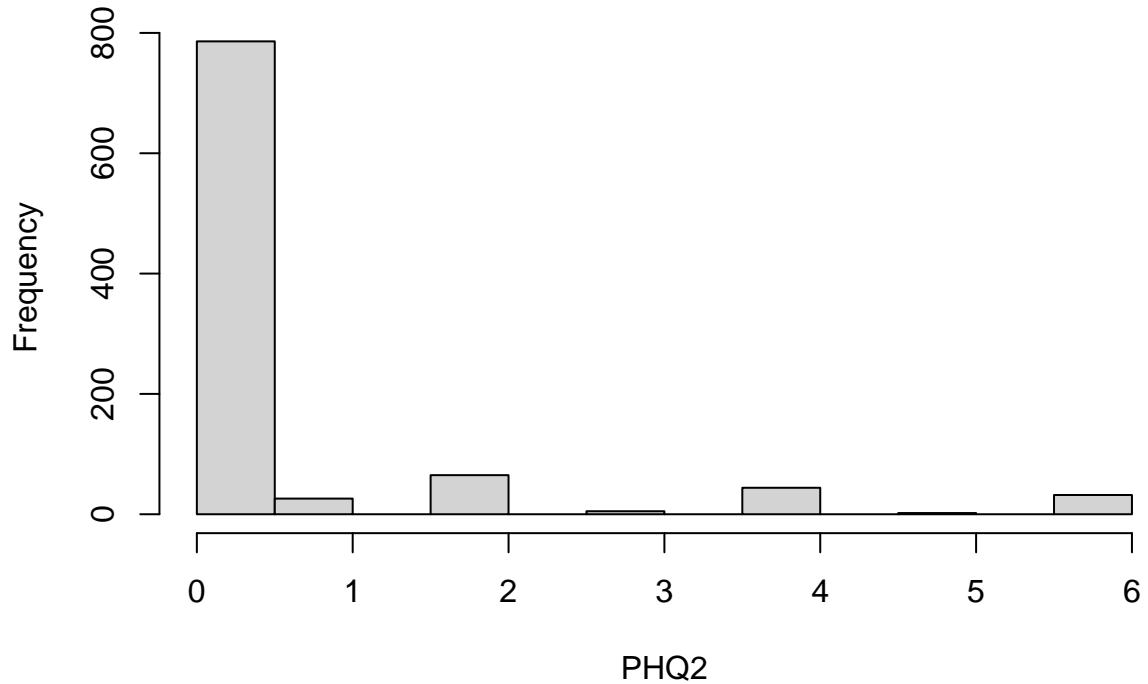
```

**PHQ-9 Histo : n = 261**



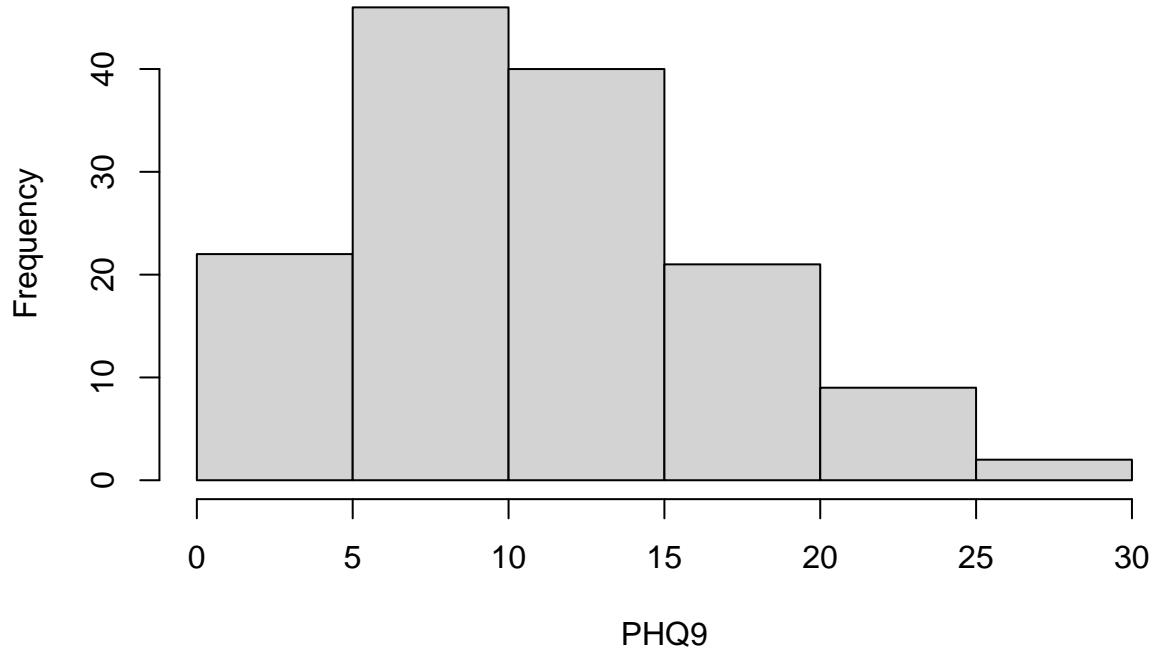
```
#fcode histos
hist(dep_df$PHQ.2, main = paste0("PHQ2 and Dep Histo : n = ", num_phq2_and_dep), xlab = paste0("PHQ2"))
```

**PHQ2 and Dep Histo : n = 960**



```
hist(dep_df$PHQ.9, main = paste0("PHQ9 and Dep Histo : n = ", num_phq9_and_dep), xlab = paste0("PHQ9"))
```

## PHQ9 and Dep Histo : n = 140

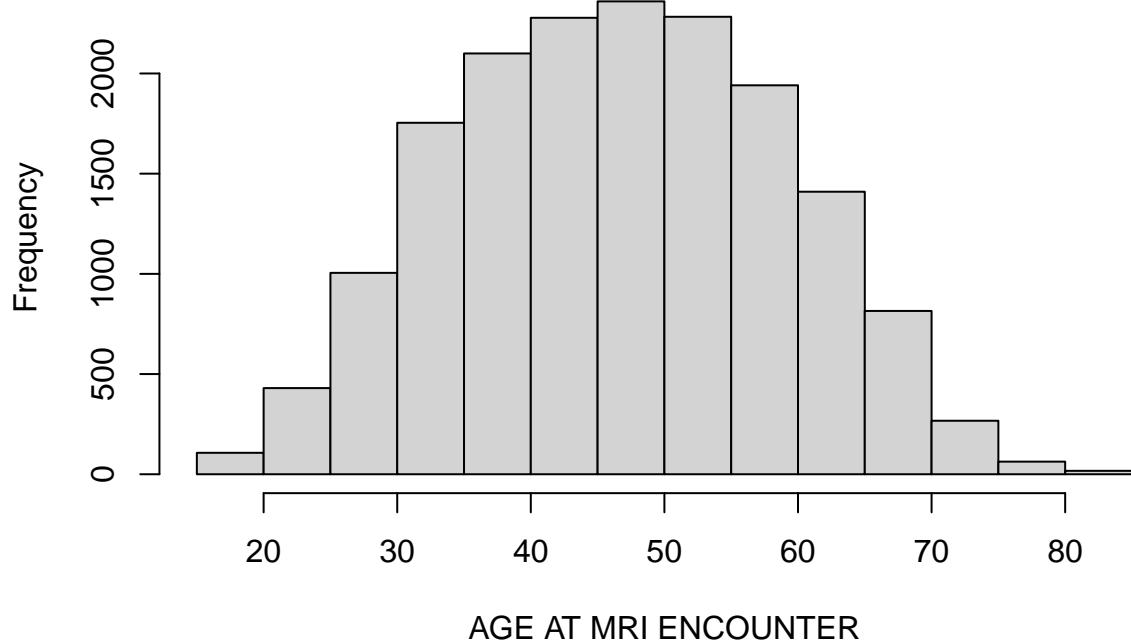


```
###
```

```
#age histogram
```

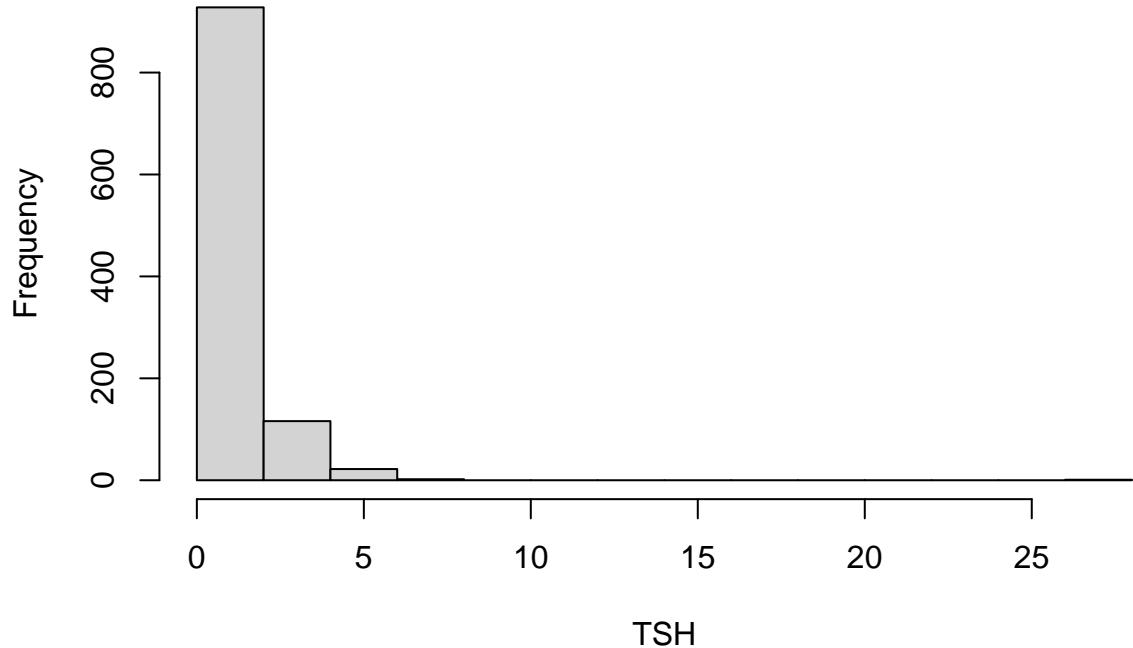
```
hist(data_empi_acc_f_phq$MRI_ENC_AGE[which(!is.na(data_empi_acc_f_phq$MRI_ENC_AGE))], main = paste0("Age
```

**Age Histo : n = 16830**



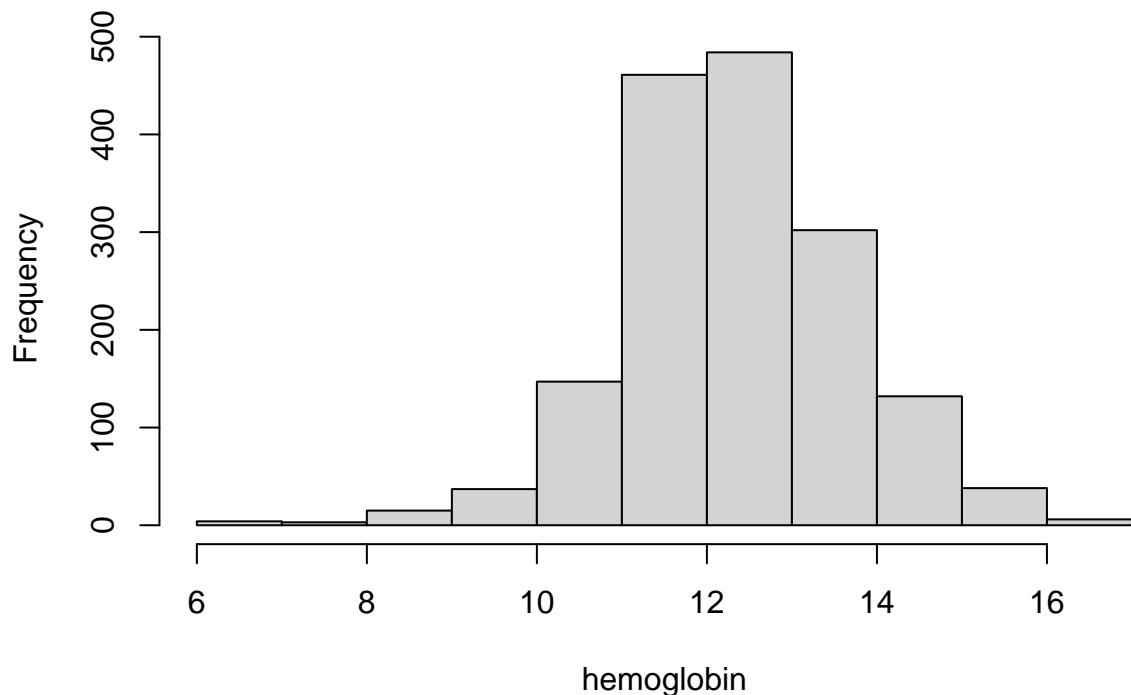
```
#tsh  
hist(data_empi_acc_f_phq$TSH[which(!is.na(data_empi_acc_f_phq$TSH))], main = paste0("TSH Histo : n = ",
```

**TSH Histo : n = 1069**



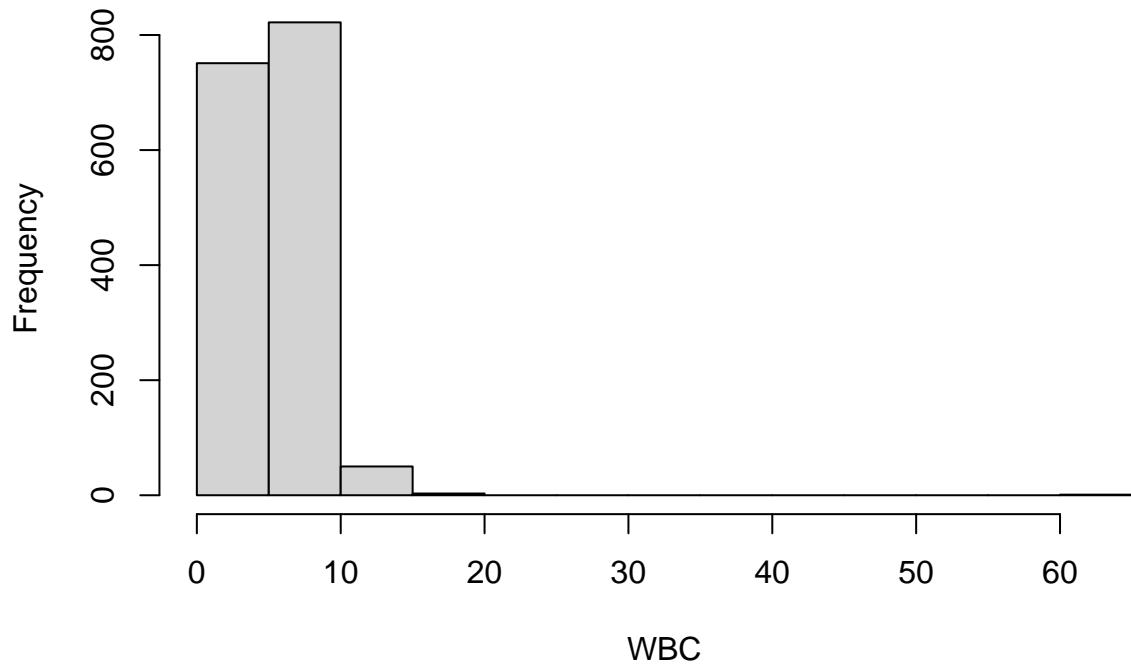
```
#hemoglobin
hist(data_empi_acc_f_phq$hemoglobin[which(!is.na(data_empi_acc_f_phq$hemoglobin))], main = paste0("Hemoglo
```

**Hemoglobin Histo : n = 1629**



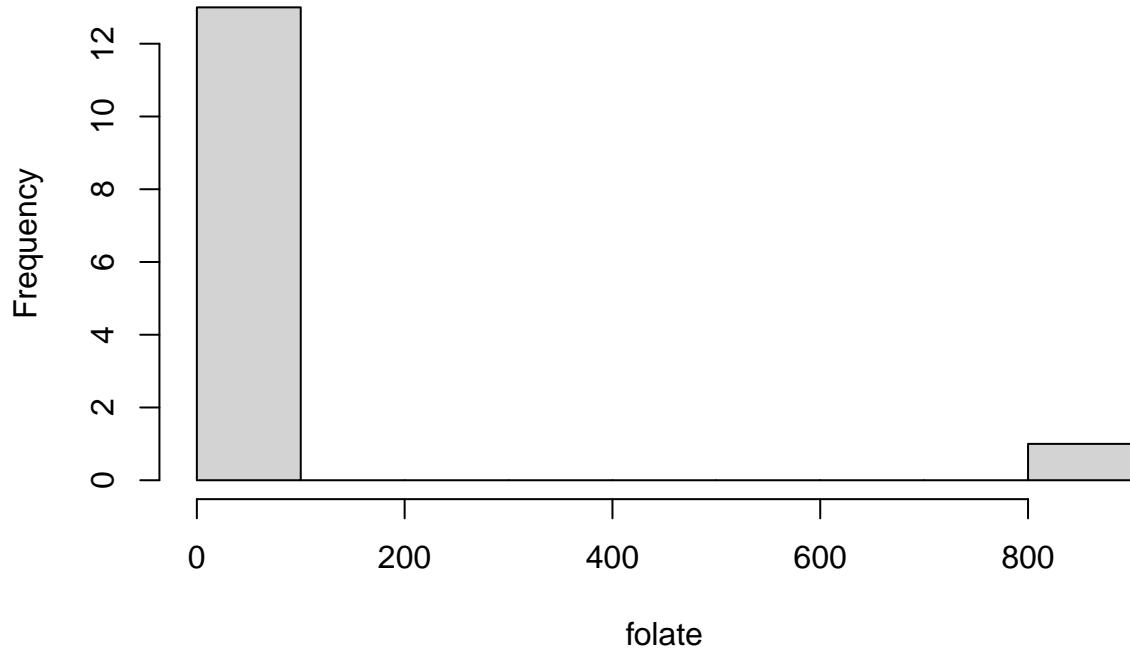
```
#wbc  
hist(data_empi_acc_f_phq$WBC[which(!is.na(data_empi_acc_f_phq$WBC))], main = paste0("WBC Histo : n = ",
```

**WBC Histo : n = 1627**



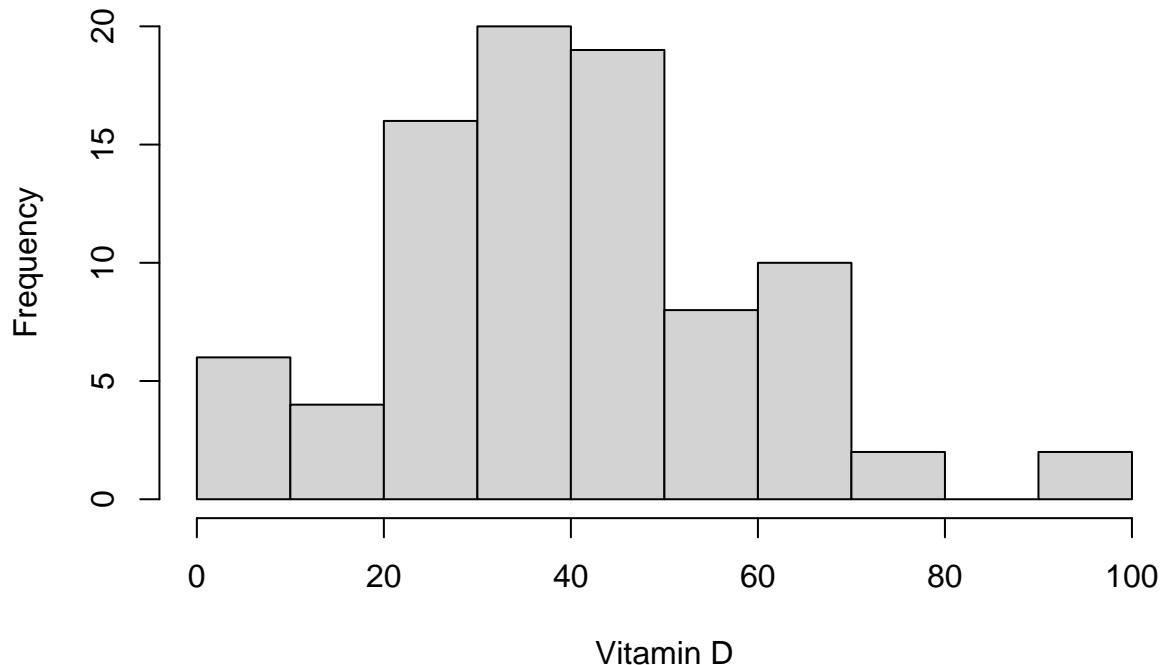
```
#folate
hist(data_empi_acc_f_phq$FOLATE[which(!is.na(data_empi_acc_f_phq$FOLATE))], main = paste0("Folate Histo"))
```

**Folate Histo : n = 14**



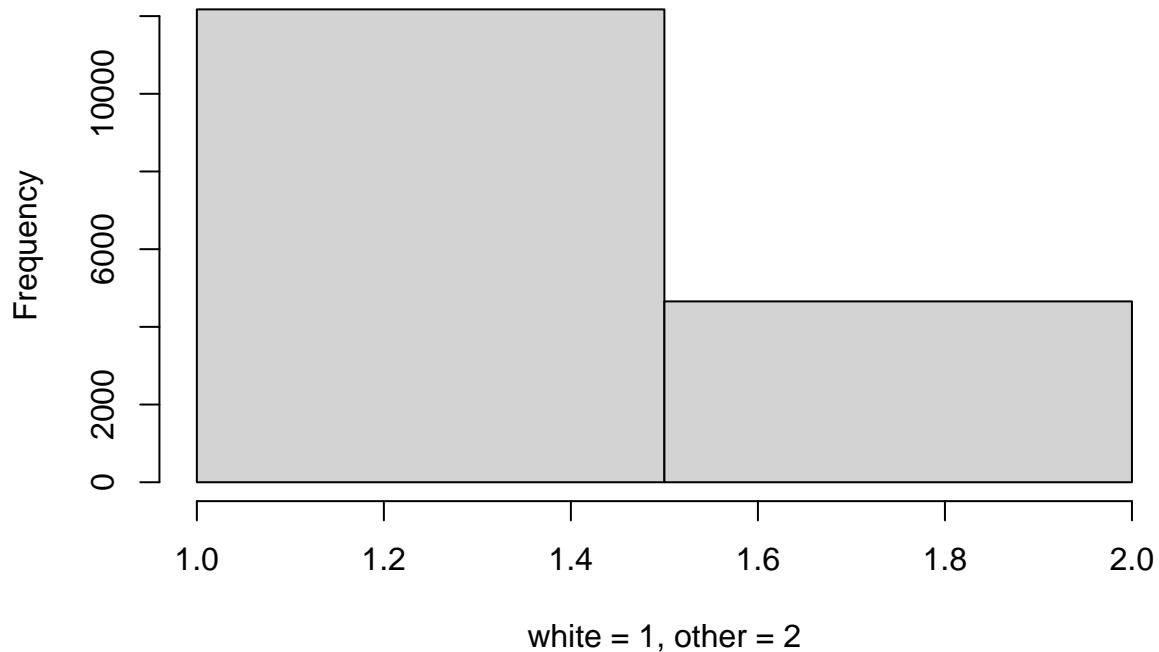
```
#vitamin D
hist(data_empi_acc_f_phq$VitD[which(!is.na(data_empi_acc_f_phq$VitD))], main = paste0("Vitamin D Histo"))
```

**Vitamin D Histo : n = 87**



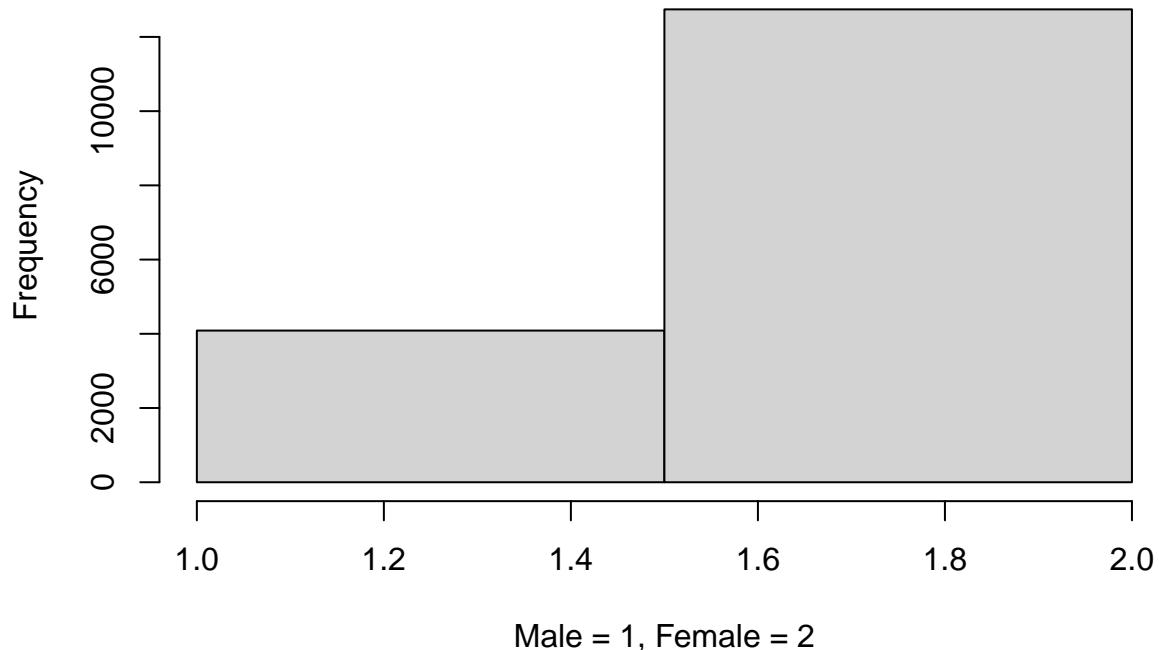
```
#race
hist(data_empi_acc_f_phq$race_binarized[which(!is.na(data_empi_acc_f_phq$race_binarized))],
      main = paste0("Race Histo : n = ", length(which(!is.na(data_empi_acc_f_phq$race))), "; White = ",
      xlab = "white = 1, other = 2",
      breaks = 2)
```

## Race Histo : n = 16830; White = 12174; NW = 4656



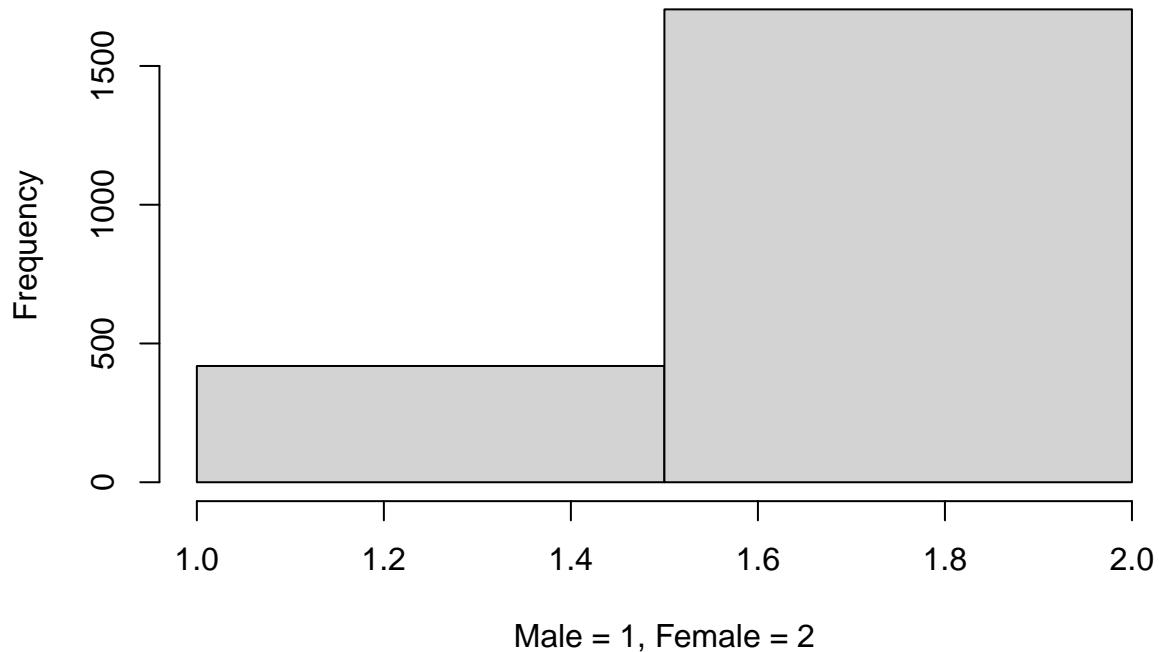
```
#f-codes - people usually have BOTH icd9/10 or neither  
  
#Gender breakdown  
#sex  
hist(data_empi_acc_f_phq$sex_binarized[which(!is.na(data_empi_acc_f_phq$sex_binarized))],  
      main = paste0("Sex Histo : n = ", length(which(!is.na(data_empi_acc_f_phq$sex_binarized))), "; M = ",  
      xlab = "Male = 1, Female = 2",  
      breaks = 2)
```

**Sex Histo : n = 16830; M = 4087; F = 12743**



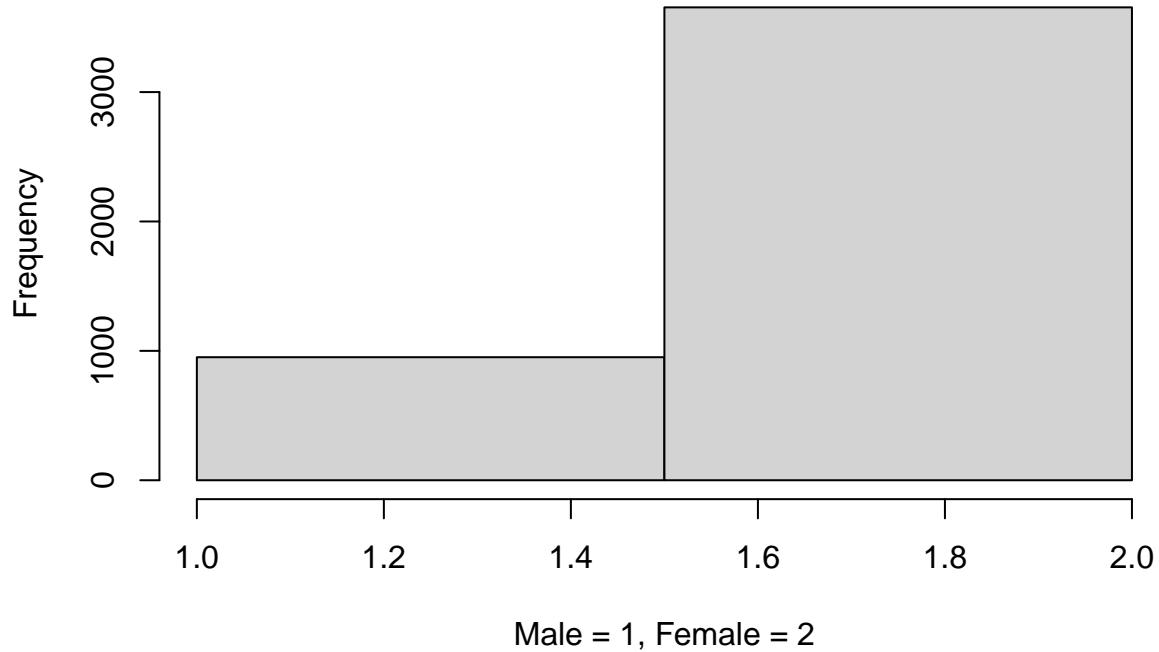
```
#sex with depression diagnosis
hist(dep_df$sex_binarized[which(!is.na(dep_df$sex_binarized))],
     main = paste0("Sex Histo in ICD9/10 depression dx group : n = ", length(which(!is.na(dep_df$sex_bin
     xlab = "Male = 1, Female = 2",
     breaks = 2)
```

## Sex Histo in ICD9/10 depression dx group : n = 2123; M = 419; F = 17



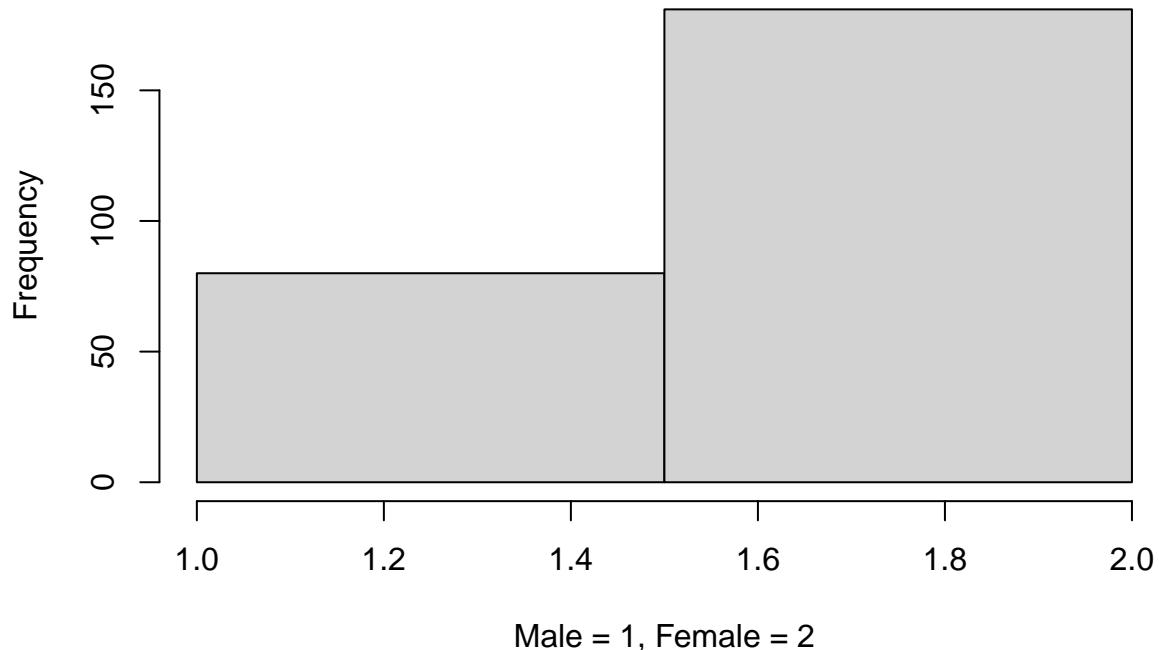
```
#sex w/ phq2
hist(phq2_subset$sex_binarized[which(!is.na(phq2_subset$sex_binarized))],
      main = paste0("Sex Histo w/phq2: n = ", length(which(!is.na(phq2_subset$sex_binarized))), "; M = ",
      xlab = "Male = 1, Female = 2",
      breaks = 2)
```

### Sex Histo w/phq2: n = 4606; M = 951; F = 3655



```
#sex w/phq9
hist(phq9_subset$sex_binarized[which(!is.na(phq9_subset$sex_binarized))],
  main = paste0("Sex Histo w/phq9: n = ", length(which(!is.na(phq9_subset$sex_binarized))), "; M = ",
  xlab = "Male = 1, Female = 2",
  breaks = 2)
```

## Sex Histo w/phq9: n = 261; M = 80; F = 181



```
num_dep_phq2 <- length(which(!is.na(data_empi_acc_f_phq$PHQ.2) & (data_empi_acc_f_phq$PHQ.2 >= 3))) #n
num_dep_phq9 <- length(which(!is.na(data_empi_acc_f_phq$PHQ.9) & (data_empi_acc_f_phq$PHQ.9 >= 5))) #n
num_total_phq2 <- length(which(!is.na(data_empi_acc_f_phq$PHQ.2)))
num_total_phq9 <- length(which(!is.na(data_empi_acc_f_phq$PHQ.9)))
dep_total_phq_table <- data.frame(matrix(data = c(num_dep_phq2, num_dep_phq9, num_total_phq2, num_total_phq9), nrow = 2, byrow = TRUE))
names(dep_total_phq_table) <- c("dep", "healthy")
row.names(dep_total_phq_table) <- c("phq2", "phq9")
chisq_dep_phq2_v_phq9 <- chisq.test(dep_total_phq_table)
chisq_dep_phq2_v_phq9

##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data: dep_total_phq_table
## X-squared = 1058, df = 1, p-value < 2.2e-16

print("observed")

## [1] "observed"

chisq_dep_phq2_v_phq9$observed

##      dep healthy
## phq2 130    4606
## phq9 184    261
```

```

print("expected")

## [1] "expected"

chisq_dep_phq2_v_phq9$expected

##          dep   healthy
## phq2 287.0303 4448.9697
## phq9 26.9697  418.0303

print("residuals")

## [1] "residuals"

chisq_dep_phq2_v_phq9$residuals

##          dep   healthy
## phq2 -9.268716 2.354256
## phq9 30.237469 -7.680325

```

This next step, put together 9/29/2021, aims to define our control and case groups based on the following definitions-

Healthy group - No F33\* code, and PHQ2 and 9 always 0

Lifetime depression - EITHER F33\* diagnosis OR PHQ2  $\geq 3$  OR PHQ9  $\geq 10$  - of note: *PHQ9 0-4 - no depression* 5-9 - mild depression \* 10-14 - moderate depression \* 15-19 - mod/severe \* 20+ - severe

Logic 1) First, separate out people with ICD10 code (F3) *into depressed and non depressed* - *dep\_df* is people with an F3 code - healthy df is people without - look for unique EMPIs in each of these groups

- 2) Within “healthy” group, separate who has EVER had a lifetime history of testing positive on PHQ2
  - Depression is PHQ2  $\geq 3$  at any point
  - Healthy if PHQ2 = 0
  - Look for unique EMPIs in these groups
- 3) Within “healthy” group, separate who has EVER had a lifetime history of testing positive on PHQ9
  - Depression is PHQ9  $\geq 10$  at any point
  - Healthy if PHQ9 = 0
  - Look for unique EMPIs in these groups
- 4) Combine for depressed group and healthy
  - Look for people with PHQ2 or 9 meeting criteria for depression that aren’t in depression group, and add them -> *dep\_num\_icd10\_plus\_phq2\_and\_phq9*
  - remove people from *healthy\_df* if people had EMPIs in *phq2* or 9 depressed group -> *healthy\_num\_no\_icd10\_plus\_phq2\_and\_phq9*
  - make separate group that only includes healthies for whom we have *phq2* or 9 -> *healthy\_num\_no\_icd10\_plus\_phq2\_and\_phq9\_MUST\_HAVE\_PHQ2\_OR\_9*

```

#####
#### Step 1- make F3* depressed and healthy groups* #####
#####

#####
### separate out depressed group
#####
print("separate out depressed group")

## [1] "separate out depressed group"

dep_df <- data_empi_acc_f_phq[grep("F3", data_empi_acc_f_phq$ICD10),] #n = 2123
num_depression <- dim(dep_df)[1]
num_unique_EMPI_and_dep <- length(unique(dep_df$EMPI)) #n = 493
print(paste0("N with ICD 9/10 codes for depression = ", num_depression))

## [1] "N with ICD 9/10 codes for depression = 2123"

print(paste0("N with ICD 9/10 codes for depression and unique EMPI in whole depression group = ", num_u

## [1] "N with ICD 9/10 codes for depression and unique EMPI in whole depression group = 493 out of 2123

#####
### separate out healthy group
#####
print("separate out healthy group")

## [1] "separate out healthy group"

healthy_df <- data_empi_acc_f_phq[-grep("F3", data_empi_acc_f_phq$ICD10),] #n = 14707
num_healthy <- dim(healthy_df)[1]
num_unique_EMPI_and_healthy <- length(unique(healthy_df$EMPI)) #n = 3244
print(paste0("N with WITHOUT depression ICD 9/10 codes (i.e. \"healthy\"): ", num_healthy))

## [1] "N with WITHOUT depression ICD 9/10 codes (i.e. \"healthy\"): 14707"

print(paste0("N with WITHOUT ICD 9/10 depression codes for depression and unique EMPI = ", num_unique_EMPI_and_healthy))

## [1] "N with WITHOUT ICD 9/10 depression codes for depression and unique EMPI = 3244 out of 14707

#####
#### Step 2- Identify PHQ2 >=3 in healthy group #####
#####

print("Get people in healthy group with phq2s for dep/healthy")

## [1] "Get people in healthy group with phq2s for dep/healthy"

```

```

# num in healthy group with PHQ2
phq2_subset <- healthy_df[which(!is.na(healthy_df$PHQ.2)),] #n=3646, unique 698
depressed_phq2_subset <- phq2_subset[which(phq2_subset$PHQ.2 >= 3),] #n = 47
depressed_unique_phq2.empis <- unique(depressed_phq2_subset$EMPI) #n=10
healthy_phq2_subset <- phq2_subset[which(phq2_subset$PHQ.2 == 0),] #n = 3326
healthy_unique_phq2.empis <- unique(healthy_phq2_subset$EMPI) #n = 631

print(paste0("PHQ2 breakdown in healthy group (total n with score = ", dim(phq2_subset)[1], "[unique =
## [1] "PHQ2 breakdown in healthy group (total n with score = 3646[unique = 698]): PHQ2 >= 3 :47[unique

#####
##### Step 3- Identify PHQ9 >=10 in healthy group #####
#####

print("Get people in healthy group with phq9s for dep/healthy")

## [1] "Get people in healthy group with phq9s for dep/healthy"

# num in healthy group with PHQ9
phq9_subset <- healthy_df[which(!is.na(healthy_df$PHQ.9)),] #n=121
depressed_phq9_subset <- phq9_subset[which(phq9_subset$PHQ.9 >= 10),] #n = 32
depressed_unique_phq9.empis <- unique(depressed_phq9_subset$EMPI) #n=9
healthy_phq9_subset <- phq9_subset[which(phq9_subset$PHQ.9 == 0),] #n = 21
healthy_unique_phq9.empis <- unique(healthy_phq9_subset$EMPI) #n =4

print(paste0("PHQ9 breakdown in healthy group (total n with score = ", dim(phq9_subset)[1], "[unique =
## [1] "PHQ9 breakdown in healthy group (total n with score = 121[unique = 25]): PHQ9 >= 10 :32[unique

#####
#Step 4 - Combine for depressed group and healthy groups #
#####

print("Combine groups for new cohort")

## [1] "Combine groups for new cohort"

gain_in_dep_group <- length(depressed_unique_phq2.empis) + length(depressed_unique_phq9.empis)
dep_num_icd10_plus_phq2_and_phq9 <- num_unique_EMPI_and_dep + gain_in_dep_group # we can do this because
print(paste0("number of UNIQUE people with an icd10 code for Depression OR scored positive on PHQ2(>=3) or PHQ9(>=10) = ", dep_num_icd10_plus_phq2_and_phq9))

## [1] "number of UNIQUE people with an icd10 code for Depression OR scored positive on PHQ2(>=3) or PHQ9(>=10) = 153

healthy_num_no_icd10_plus_phq2_and_phq9 <- num_unique_EMPI_and_healthy - gain_in_dep_group #take number

print(paste0("number of UNIQUE people with NO icd10 code for Depression and NEVER had a PHQ2(>=3) or PHQ9(>=10) = ", healthy_num_no_icd10_plus_phq2_and_phq9))

## [1] "number of UNIQUE people with NO icd10 code for Depression and NEVER had a PHQ2(>=3) or PHQ9(>=10) = 147

```

```

healthy_num_no_icd10_plus_phq2_and_phq9_MUST_HAVE_PHQ2_OR_9 <- length(healthy_unique_phq2.empis) + length(healthy_unique_phq9.empis)
print(paste0("number of UNIQUE people with NO icd10 code for Depression and HAS HAD AT LEAST 1 PHQ2 or 9 = ", healthy_num_no_icd10_plus_phq2_and_phq9_MUST_HAVE_PHQ2_OR_9))

## [1] "number of UNIQUE people with NO icd10 code for Depression and HAS HAD AT LEAST 1 PHQ2 or 9 that have never had a dx = ", healthy_num_no_icd10_plus_phq2_and_phq9_never_had_dx)

print(paste0("Cleanest cohort (depressed in icd9/10 or Phq2/9; healthy by at least one phq2/9 and never depressed) = ", final_cleanest_cohort))

## [1] "Cleanest cohort (depressed in icd9/10 or Phq2/9; healthy by at least one phq2/9 and never depressed) = ", final_cleanest_cohort

##### THESE ARE THE DATAFRAMES I WANT TO USE GOING FORWARD #####
#all empirs for depressed group include people with depression dx, and people who got in for phq2 or phq9
empis_for_depressed_group <- append(
  append(unique(dep_df$EMPI), depressed_unique_phq2.empis), depressed_unique_phq9.empis)

final_depressed_group_withICD_AND_depressed_phq2_or_9 <- 
  data.empi.acc.f.phq %>%
  filter(EMPI %in% empirs_for_depressed_group)

#-----
empis_for_healthy_group <- append(healthy.unique_phq2.empis, healthy.unique_phq9.empis)

final_healthy_group_withNOICD_AND_phq2_or_9_0<- 
  data.empi.acc.f.phq %>%
  filter(EMPI %in% empirs_for_healthy_group)

#-----
### add a column to the original dataframe with a 0 if not in any group, -1 if healthy, and 1 if depressed
data.empi.acc.f.phq$depGroupVar = 0
data.empi.acc.f.phq$depGroupVar[which(data.empi.acc.f.phq$EMPI %in% empirs_for_depressed_group)] = 1
data.empi.acc.f.phq$depGroupVar[which(data.empi.acc.f.phq$EMPI %in% empirs_for_healthy_group)] = -1
data.empi.acc.f.phq$depGroupVar = as.factor(data.empi.acc.f.phq$depGroupVar)

#### looking at anyone with phq2s or 9s
males.phq2 <- phq2_subset[which(phq2_subset$SEX == "MALE"),] #n = 951, mean = 0.20
females.phq2 <- phq2_subset[which(phq2_subset$SEX == "FEMALE"),] #n = 3655, mean = 0.26
m_v_f_phq2 <- t.test(males.phq2$PHQ.2, females.phq2$PHQ.2) #p = 0.03 ***

males.phq9 <- phq9_subset[which(phq9_subset$SEX == "MALE"),] #n = 80, mean = 8.63
females.phq9 <- phq9_subset[which(phq9_subset$SEX == "FEMALE"),] #n = 181, mean = 9.13
m_v_f_phq9 <- t.test(males.phq9$PHQ.9, females.phq9$PHQ.9) #p = 0.24

##### Looking at depressed group
## interestingly, not the 2:1 ratio we usually see in general world
## phq2 means suggest people are not depressed
## phq2 means suggest people meet criteria for mild depression

males_dep <- dep_df[(which(dep_df$SEX == "MALE")),] #n = 419/total(4087; 10%) F; mean PHQ2 = 0.61, mean PHQ9 = 8.63
females_dep <- dep_df[(which(dep_df$SEX == "FEMALE")),] #n = 1704/total(12743; 13%)total F; mean = 0.56
m_v_f_dep_phq2 <- t.test(males_dep$PHQ.2, females_dep$PHQ.2) #p = 0.66
m_v_f_dep_phq9 <- t.test(males_dep$PHQ.9, females_dep$PHQ.9) #p = 0.24

#chi sq comparing # of pts dx with depression in males vs females; col 1- male dep n, female dep n; col 2- male dep mean, female dep mean

```

```

### fascinating - only 12% of people even gave a dx of depression in their charts. More likely to have
num_dep_m <- dim(males_dep)[1] #n = 419
num_dep_f <- dim(females_dep)[1] #n = 1704
num_total_m <- length(which(data_empi_acc_f_phq$SEX == "MALE")) #n = 4087
num_total_f <- length(which(data_empi_acc_f_phq$SEX == "FEMALE")) #n =
m_v_f_table <- data.frame(matrix(data = c(num_dep_m, num_dep_f, num_total_m, num_total_f), nrow = 2, nc
names(m_v_f_table) <- c("dep", "total")
row.names(m_v_f_table) <- c("male", "female")
chisq_m_v_f_dist <- chisq.test(m_v_f_table) #p = 4 x 10^-6
chisq_m_v_f_dist

##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data: m_v_f_table
## X-squared = 21.265, df = 1, p-value = 4e-06

print("observed")

## [1] "observed"

chisq_m_v_f_dist$observed

##          dep   total
## male     419   4087
## female  1704  12743

print("expected")

## [1] "expected"

chisq_m_v_f_dist$expected

##          dep   total
## male    504.7348 4001.265
## female 1618.2652 12828.735

print("residuals")

## [1] "residuals"

chisq_m_v_f_dist$residuals

##          dep   total
## male   -3.816149  1.3553713
## female  2.131239 -0.7569463

```

```

#chisq comparing # of

#preprocessing - because these are sent

#load our df

#from Jan 2021 pull
data <- read.csv("/Users/eballer/BBL/msdepression/data/dac/investigatingdepressioninmspatients_dates_rig
ms_providers <- c("MARKOWITZ, CLYDE E.", "JACOBS, DINA A.", "WILLIAMSON, ERIC MICHAEL-LEE", "BERGER, JOSEPH", "CHAHIN, SALIM", "KOLSON, DENNIS L.", "NARULA, SONA", "PRUITT, AMY A.", "WILLIAMSON, ERIC MICHAEL-LEE")

#keep people with MS providers
data_empi_acc_f_phq <- data[which(data$Provider %in% ms_providers),] #n = 17067

#keep people with ICD10 code G35 (this is the same # as people with IC9 - 340)
data_empi_acc_f_phq <- data_empi_acc_f_phq[grep("G35", data_empi_acc_f_phq$ICD10),] #n = 16830
table(data_empi_acc_f_phq$Provider)

## Data frame showing provider names and counts
## BAR-OR, AMIT 93
## BERGER, JOSEPH ROBERT 1818
## JACOBS, DINA A. 4765
## MARKOWITZ, CLYDE E. 6245
## PRUITT, AMY A. WILLIAMSON, ERIC MICHAEL-LEE 1372
## BERGER DO, JOSEPH 1
## CHAHIN, SALIM 260
## KOLSON, DENNIS L. 643
## NARULA, SONA 274
## 1359

#####
### some preprocessing #####
#####

#make accession # and lab findings into integer type, contained in row 28:42, Margin=2 to work on column
data_empi_acc_f_phq$ACCESSION_NUM <- as.integer(data_empi_acc_f_phq$ACCESSION_NUM)
data_empi_acc_f_phq[27:29] <- apply(X = data_empi_acc_f_phq[27:29], FUN = as.integer, MARGIN = 2)
data_empi_acc_f_phq[31:45] <- apply(X = data_empi_acc_f_phq[31:45], FUN = as.integer, MARGIN = 2)

## Warning in apply(X = data_empi_acc_f_phq[31:45], FUN = as.integer, MARGIN = 2):
## NAs introduced by coercion

## Warning in apply(X = data_empi_acc_f_phq[31:45], FUN = as.integer, MARGIN = 2):
## NAs introduced by coercion

## Warning in apply(X = data_empi_acc_f_phq[31:45], FUN = as.integer, MARGIN = 2):
## NAs introduced by coercion

## Warning in apply(X = data_empi_acc_f_phq[31:45], FUN = as.integer, MARGIN = 2):
## NAs introduced by coercion

```



```

## [1] "N PHQ-2 = 4606; n = 130 meeting criteria for depression (PHQ2) >= 3"

#phq-9
print(paste0("N PHQ-9 = ", sum(!is.na(data_empi_acc_f_phq$PHQ.9)), "; noDep/mild/mod/mod-severe/severe =",
            length(which(phq9_subset$PHQ.9 < 5)), "/",
            length(which(phq9_subset$PHQ.9 >= 5 & phq9_subset$PHQ.9 < 10)), "/",
            length(which(phq9_subset$PHQ.9 >= 10 & phq9_subset$PHQ.9 < 15)), "/",
            length(which(phq9_subset$PHQ.9 >= 15 & phq9_subset$PHQ.9 < 20)), "/",
            length(which(phq9_subset$PHQ.9 >= 20)))) #n = 261

## [1] "N PHQ-9 = 261; noDep/mild/mod/mod-severe/severe = 77/73/65/21/25"

#both phq2 and 9

num_both_phq2_and_9 <- length(which(!is.na(data_empi_acc_f_phq$PHQ.2) & !is.na(data_empi_acc_f_phq$PHQ.9)))
print(paste0("num people with both phq2 and 9 = ", num_both_phq2_and_9)) #n = 0

## [1] "num people with both phq2 and 9 = 0"

#Number of unique EMPIs in phq2/phq9 subsets
unique_phq2.empis <- length(unique(phq2_subset$EMPI)) #n=966
print(paste0("N with phq2 (Unique EMPI) n = ", unique_phq2.empis, " out of ", dim(phq2_subset)[1], "; n = ", n))

## [1] "N with phq2 (Unique EMPI) n = 899 out of 4606; n - 32 meeting criteria for depression (PHQ2) >= 3"

unique_phq9.empis <- length(unique(phq9_subset$EMPI)) #n=76
print(paste0("N with phq9 (Unique EMPI) n = ", unique_phq9.empis, " out of ", dim(phq9_subset)[1], "; n = ", n))
length(unique(phq9_subset$EMPI[phq9_subset$PHQ.9 < 5])), "/",
length(unique(phq9_subset$EMPI[(phq9_subset$PHQ.9 >= 5 & phq9_subset$PHQ.9 < 10)])), "/",
length(unique(phq9_subset$EMPI[(phq9_subset$PHQ.9 >= 10 & phq9_subset$PHQ.9 < 15)])), "/",
length(unique(phq9_subset$EMPI[(phq9_subset$PHQ.9 >= 15 & phq9_subset$PHQ.9 < 20)])), "/",
length(unique(phq9_subset$EMPI[phq9_subset$PHQ.9 >= 20]))))

## [1] "N with phq9 (Unique EMPI) n = 68 out of 261; noDep/mild/mod/mod-severe/severe = 18/16/18/8/8"

##### ICD 9/10 for Psychosis
psychosis_df <- data_empi_acc_f_phq[grep("F2", data_empi_acc_f_phq$ICD10),] #n = 2123
num_psychosis <- dim(psychosis_df)[1]
unique_empis <- which(unique(data_empi_acc_f_phq$EMPI) %in% data_empi_acc_f_phq$EMPI) #index of the location
num_unique_EMPI_and_psychosis <- length(unique(psychosis_df$EMPI)) #n = 493

print(paste0("N with ICD 9/10 codes for psychosis = ", num_psychosis))

## [1] "N with ICD 9/10 codes for psychosis = 0"

print(paste0("N with ICD 9/10 codes for psychosis and unique EMPI = ", num_unique_EMPI_and_psychosis, "; n = ", n))

## [1] "N with ICD 9/10 codes for psychosis and unique EMPI = 0 out of 16830"

```

```

##### PHQ2 or 9 + ICD 9/10 psychosis
num_phq2_and_psychosis <- length(which(!is.na(psychosis_df$PHQ.2))) #n = 960
print(paste0("N with ICD 9/10 codes for psychosis and phq2 (ALL SCANS) = ", num_phq2_and_psychosis, " out of 960"))

## [1] "N with ICD 9/10 codes for psychosis and phq2 (ALL SCANS) = 0 out of 0"

num_phq9_and_psychosis <- length(which(!is.na(psychosis_df$PHQ.9))) #n=140
print(paste0("N with ICD 9/10 codes for psychosis and phq9 (ALL SCANS) = ", num_phq9_and_psychosis, " out of 140"))

## [1] "N with ICD 9/10 codes for psychosis and phq9 (ALL SCANS) = 0 out of 0"

#phq-2;
print(paste0("N psychosis + PHQ-2 = ", sum(!is.na(psychosis_df$PHQ.2)), "; n = ", length(which(psychosis_df$PHQ.2) == 0)))

## [1] "N psychosis + PHQ-2 = 0; n = 0 meeting criteria for psychosis (PHQ2) >= 3"

#phq-9
print(paste0("N psychosis + PHQ-9 = ", sum(!is.na(psychosis_df$PHQ.9)), "; nopsychosis/mild/mod/mod-severe = ", 
            length(which(psychosis_df$PHQ.9 < 5)), "/",
            length(which(psychosis_df$PHQ.9 >= 5 & psychosis_df$PHQ.9 < 10)), "/",
            length(which(psychosis_df$PHQ.9 >= 10 & psychosis_df$PHQ.9 < 15)), "/",
            length(which(psychosis_df$PHQ.9 >= 15 & psychosis_df$PHQ.9 < 20)), "/",
            length(which(psychosis_df$PHQ.9 >= 20)))) #n = 261

## [1] "N psychosis + PHQ-9 = 0; nopsychosis/mild/mod/mod-severe/severe = 0/0/0/0/0"

##### # unique people with psychosis dx and phq2/9 -> working on this
psychosis_dx_phq2_unique <- length(unique(psychosis_df$EMPI[!is.na(psychosis_df$PHQ.2)])) #n = 201
print(paste0("N with ICD 9/10 codes for psychosis and phq2 (unique) = ", psychosis_dx_phq2_unique, " out of 201"))

## [1] "N with ICD 9/10 codes for psychosis and phq2 (unique) = 0 out of 0"

psychosis_dx_phq9_unique <- length(unique(psychosis_df$EMPI[!is.na(psychosis_df$PHQ.9)])) #n = 43
print(paste0("N with ICD 9/10 codes for psychosis and phq9 (unique) = ", psychosis_dx_phq9_unique, " out of 43"))

## [1] "N with ICD 9/10 codes for psychosis and phq9 (unique) = 0 out of 0"

#####

##### ICD 9/10 for Healthy Group
healthy_df <- data_empi_acc_f_phq[(-grep("F3", data_empi_acc_f_phq$ICD10)),] #n = 14707
num_healthy <- dim(healthy_df)[1]
unique_empi <- which(unique(healthy_df$EMPI) %in% healthy_df$EMPI) #index of the locations of unique empi
num_unique_EMPI_and_healthy <- length(unique(healthy_df$EMPI)) #n = 3244

print(paste0("N without psychosis/ICD 9/10 codes for healthy = ", num_healthy))

## [1] "N without psychosis/ICD 9/10 codes for healthy = 14707"

```

```

print(paste0("N without psychosis/ ICD 9/10 codes for healthy and unique EMPI = ", num_unique_EMPI_and_healthy))

## [1] "N without psychosis/ ICD 9/10 codes for healthy and unique EMPI = 3244 out of 14707"

##### PHQ2 or 9 + No psychosis ICD9/10
num_phq2_and_healthy <- length(which(!is.na(healthy_df$PHQ.2))) #n = 3646
print(paste0("N without psychosis/ ICD 9/10 codes for healthy and phq2 (ALL SCANS) = ", num_phq2_and_healthy))

## [1] "N without psychosis/ ICD 9/10 codes for healthy and phq2 (ALL SCANS) = 3646 out of 14707"

#phq-2;
print(paste0("N healthy w/PHQ-2 = ", sum(!is.na(healthy_df$PHQ.2)), "; n = ", length(which(healthy_df$PHQ.2)))))

## [1] "N healthy w/PHQ-2 = 3646; n = 47 meeting criteria for psychosis (PHQ2) >= 3"

num_phq9_and_healthy <- length(which(!is.na(healthy_df$PHQ.9))) #n=121
print(paste0("N without psychosis/ ICD 9/10 codes for healthy and phq9 (ALL SCANS) = ", num_phq9_and_healthy))

## [1] "N without psychosis/ ICD 9/10 codes for healthy and phq9 (ALL SCANS) = 121 out of 14707"

#phq-9
print(paste0("N healthy w/PHQ-9 = ", sum(!is.na(healthy_df$PHQ.9)), "; nopsychosis/mild/mod/mod-severe/severe = ",
            length(which(healthy_df$PHQ.9 < 5)), "/",
            length(which(healthy_df$PHQ.9 >= 5 & healthy_df$PHQ.9 < 10)), "/",
            length(which(healthy_df$PHQ.9 >= 10 & healthy_df$PHQ.9 < 15)), "/",
            length(which(healthy_df$PHQ.9 >= 15 & healthy_df$PHQ.9 < 20)), "/",
            length(which(healthy_df$PHQ.9 >= 20)))) #n = 55/34/21/1/10

## [1] "N healthy w/PHQ-9 = 121; nopsychosis/mild/mod/mod-severe/severe = 55/34/21/1/10"

##### # unique people without psychosis dx and phq2/9
psychosis_dx_phq2_unique <- length(unique(healthy_df$EMPI[!is.na(healthy_df$PHQ.2)])) #n = 201
print(paste0("N without psychosis/ ICD 9/10 codes for healthy and phq2 (unique) = ", psychosis_dx_phq2_unique))

## [1] "N without psychosis/ ICD 9/10 codes for healthy and phq2 (unique) = 698 out of 14707"

psychosis_dx_phq9_unique <- length(unique(healthy_df$EMPI[!is.na(healthy_df$PHQ.9)])) #n = 43
print(paste0("N without psychosis/ ICD 9/10 codes for healthy and phq9 (unique) = ", psychosis_dx_phq9_unique))

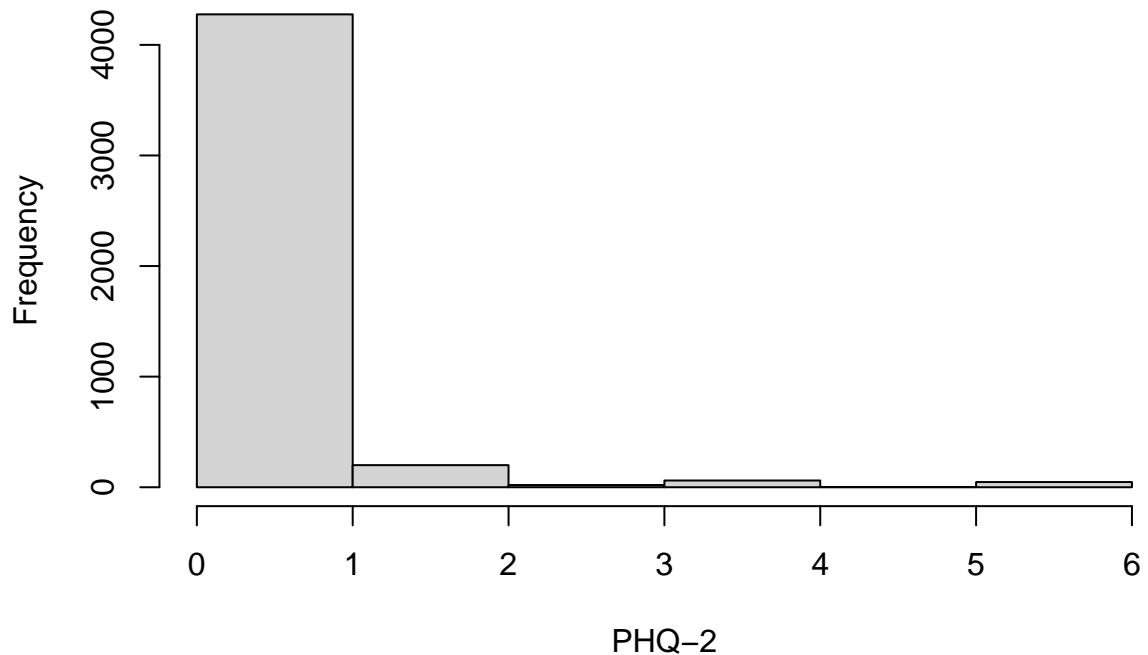
## [1] "N without psychosis/ ICD 9/10 codes for healthy and phq9 (unique) = 25 out of 14707"

#####

#demographics and lab histograms
#print histogram of phq-2/9s for people

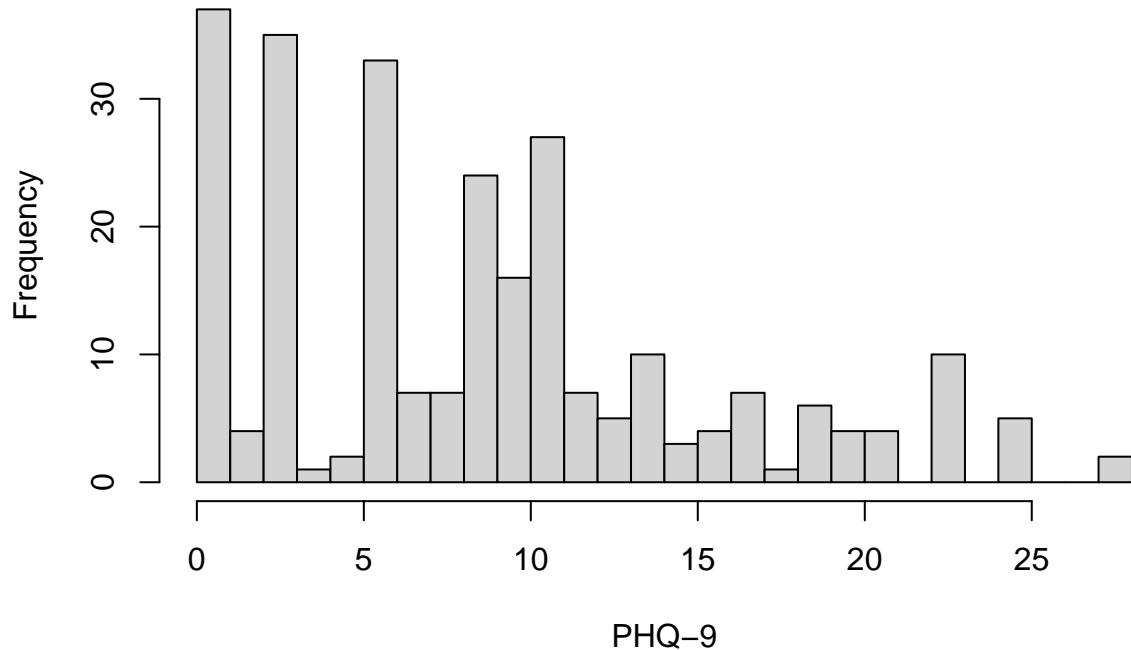
hist(data_empi_acc_f_phq$PHQ.2[which(!is.na(data_empi_acc_f_phq$PHQ.2))], main = paste0("PHQ-2 Histo : ",
```

**PHQ-2 Histo : n = 4606**



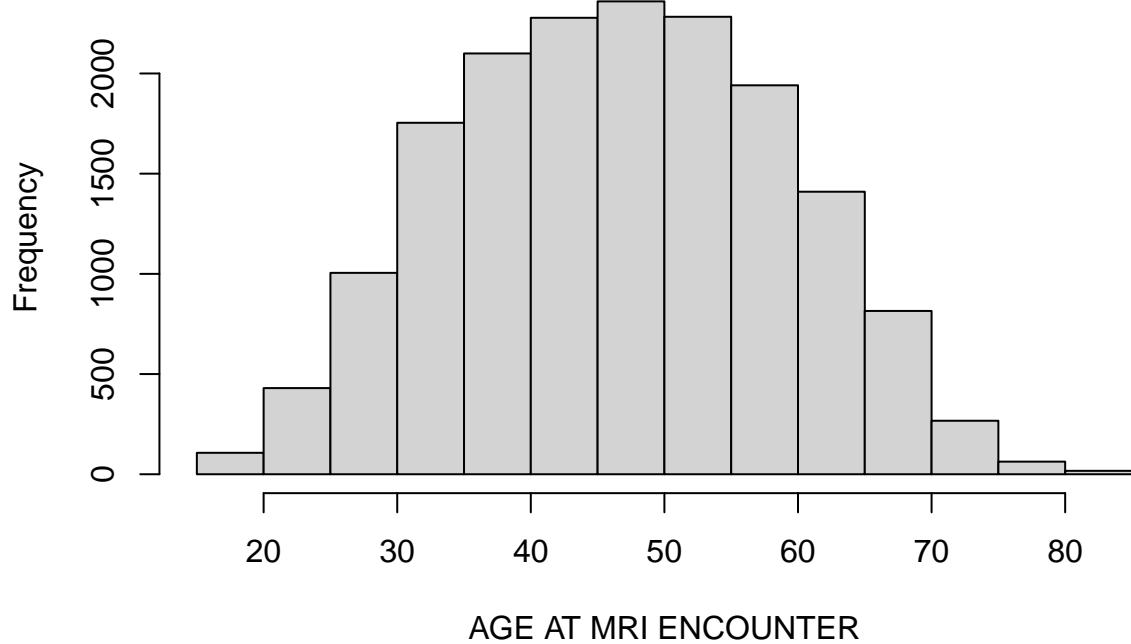
```
hist(data_empi_acc_f_phq$PHQ.9[which(!is.na(data_empi_acc_f_phq$PHQ.9))], main = paste0("PHQ-9 Histo : n = ", nrow(data_empi_acc_f_phq)))
```

**PHQ-9 Histo : n = 261**



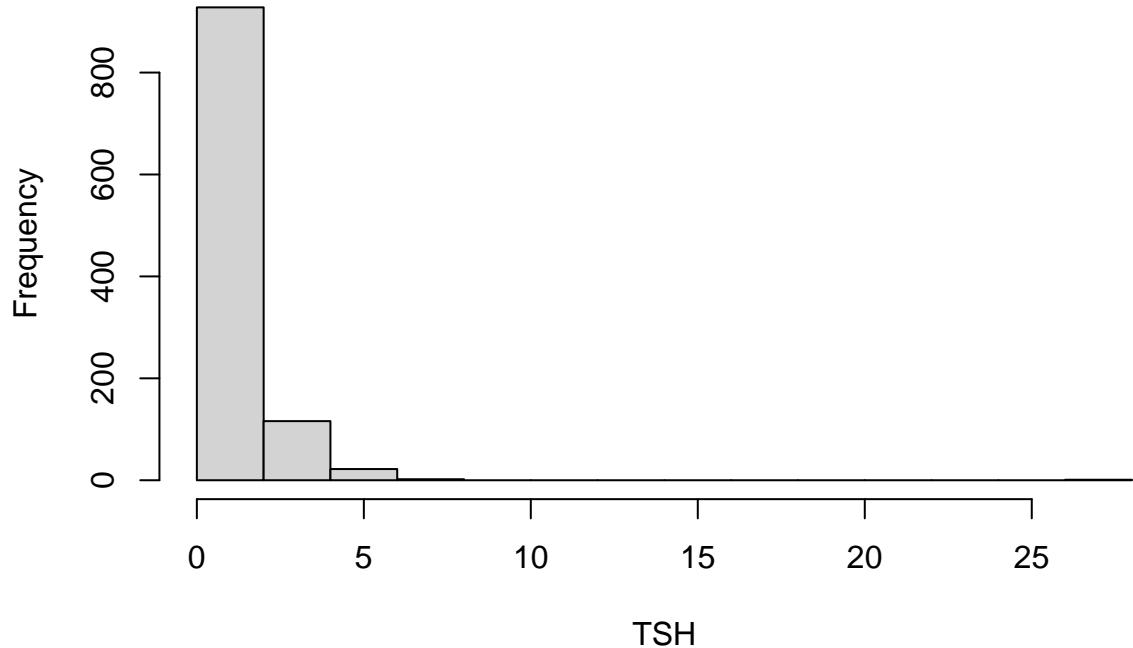
```
#fcode histos - no one with it at this time.  
#hist(psychosis_df$PHQ.2, main = paste0("PHQ2 and psychosis Histo : n = ", num_phq2_and_psychosis), xla  
#hist(psychosis_df$PHQ.9,main = paste0("PHQ9 and psychosis Histo : n = ", num_phq9_and_psychosis), xlab  
###  
#age histogram  
hist(data_empi_acc_f_phq$MRI_ENC_AGE[which(!is.na(data_empi_acc_f_phq$MRI_ENC_AGE))], main = paste0("Age
```

**Age Histo : n = 16830**



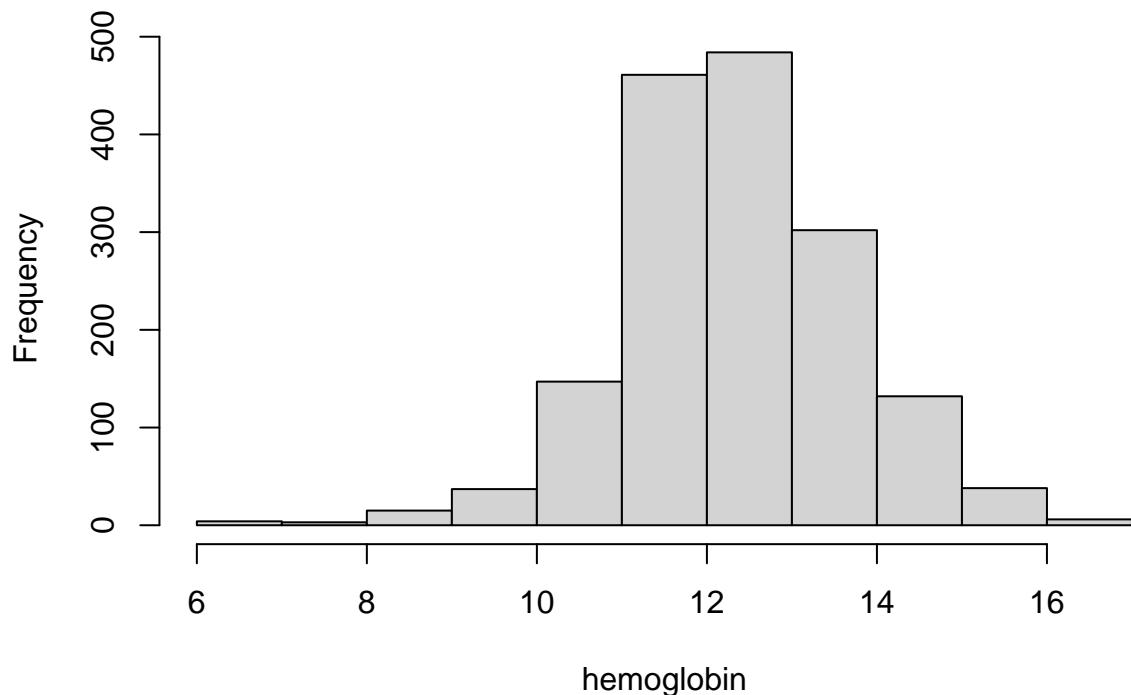
```
#tsh  
hist(data_empi_acc_f_phq$TSH[which(!is.na(data_empi_acc_f_phq$TSH))], main = paste0("TSH Histo : n = ",
```

**TSH Histo : n = 1069**



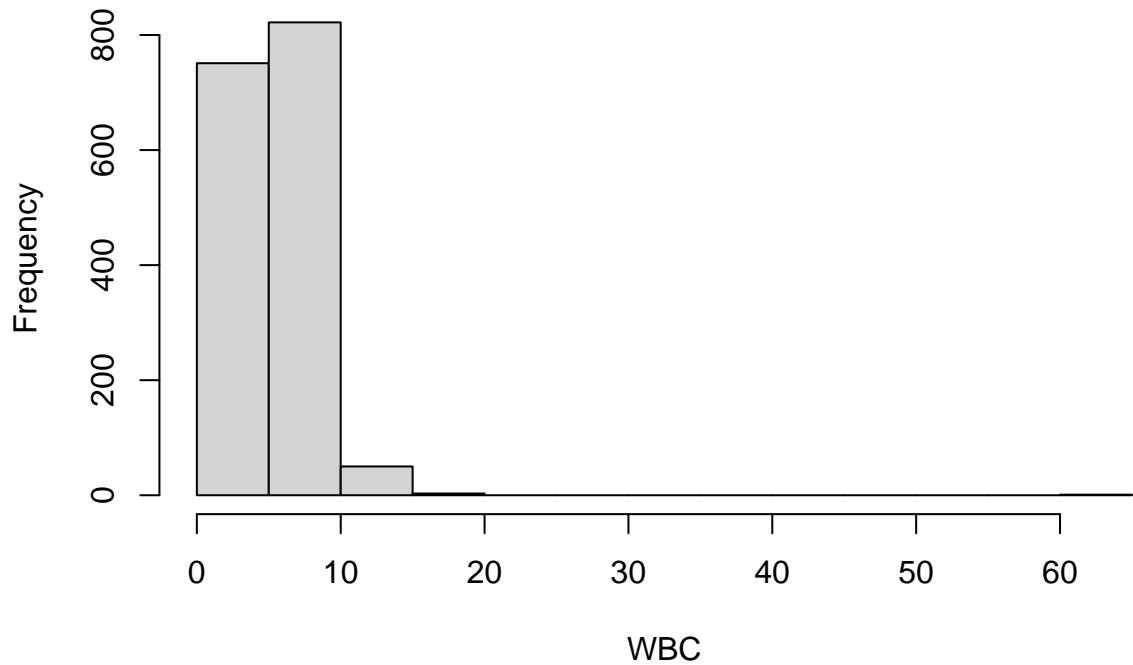
```
#hemoglobin
hist(data_empi_acc_f_phq$hemoglobin[which(!is.na(data_empi_acc_f_phq$hemoglobin))], main = paste0("Hemoglo
```

**Hemoglobin Histo : n = 1629**



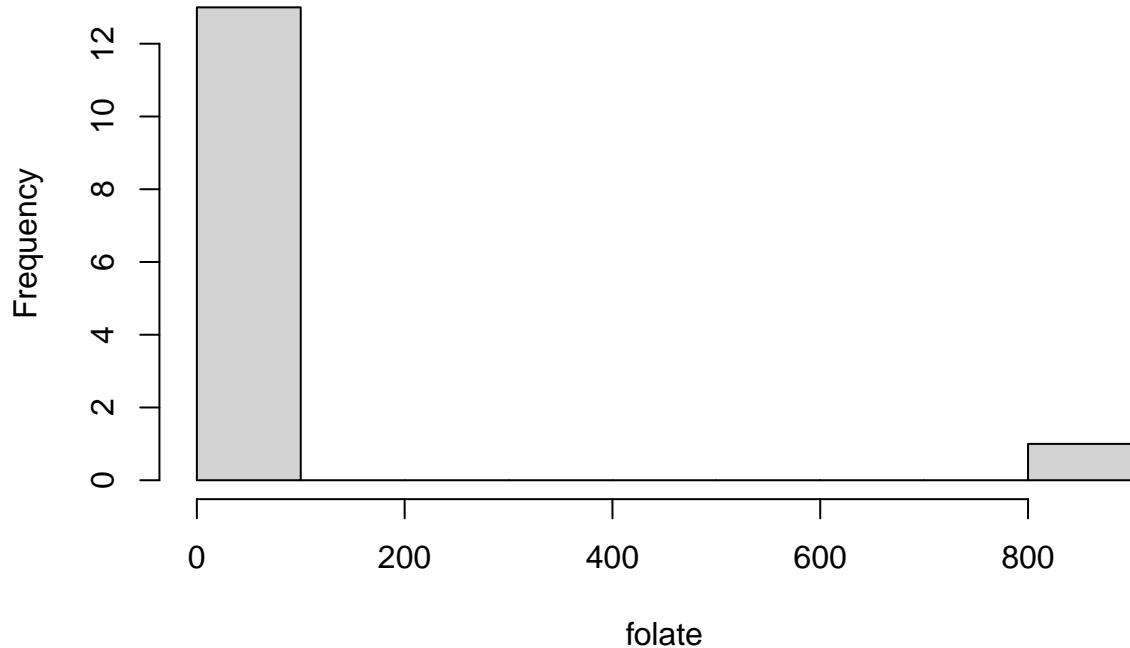
```
#wbc  
hist(data_empi_acc_f_phq$WBC[which(!is.na(data_empi_acc_f_phq$WBC))], main = paste0("WBC Histo : n = ",
```

**WBC Histo : n = 1627**



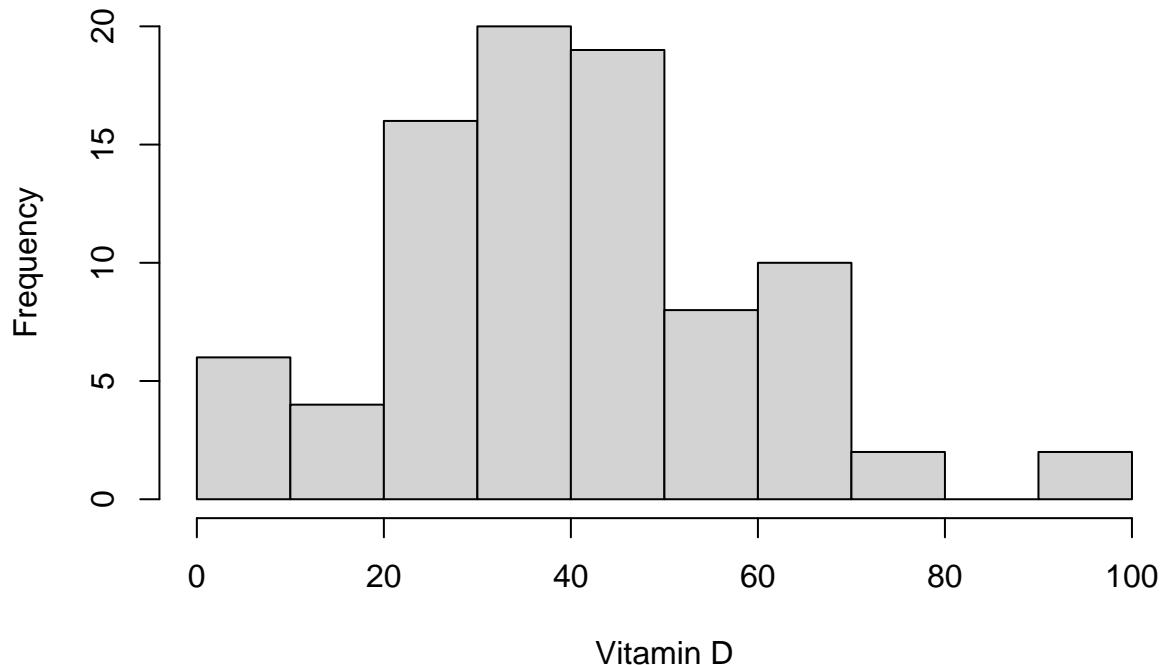
```
#folate  
hist(data_empi_acc_f_phq$FOLATE[which(!is.na(data_empi_acc_f_phq$FOLATE))], main = paste0("Folate Histo"))
```

**Folate Histo : n = 14**



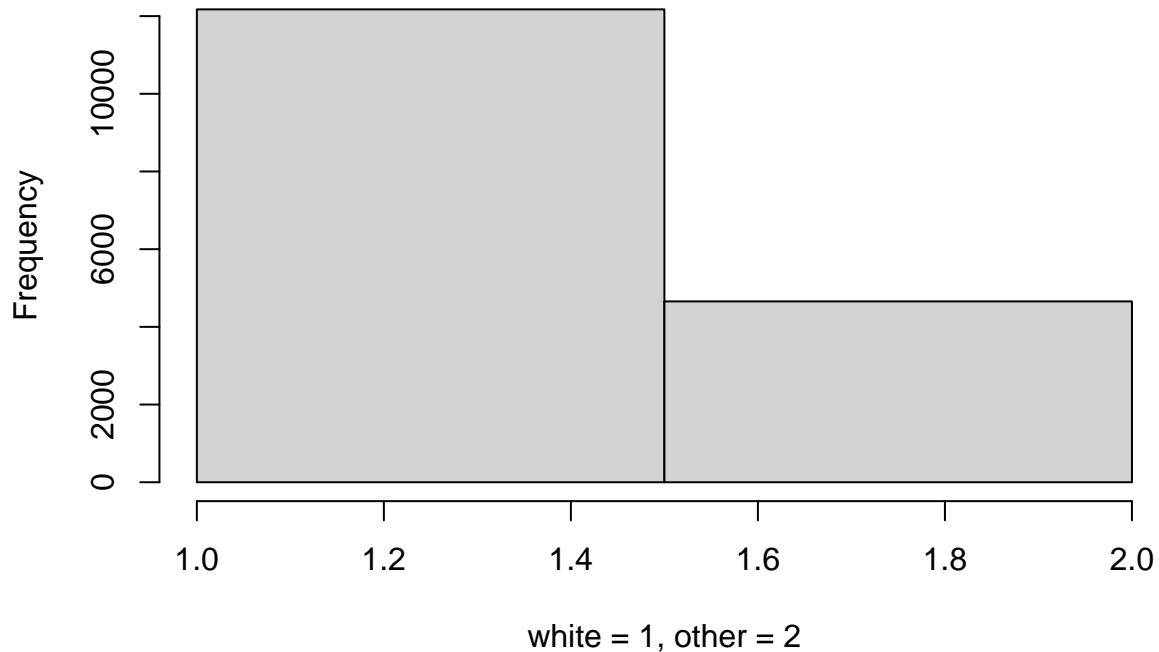
```
#vitamin D
hist(data_empi_acc_f_phq$VitD[which(!is.na(data_empi_acc_f_phq$VitD))], main = paste0("Vitamin D Histo"))
```

**Vitamin D Histo : n = 87**



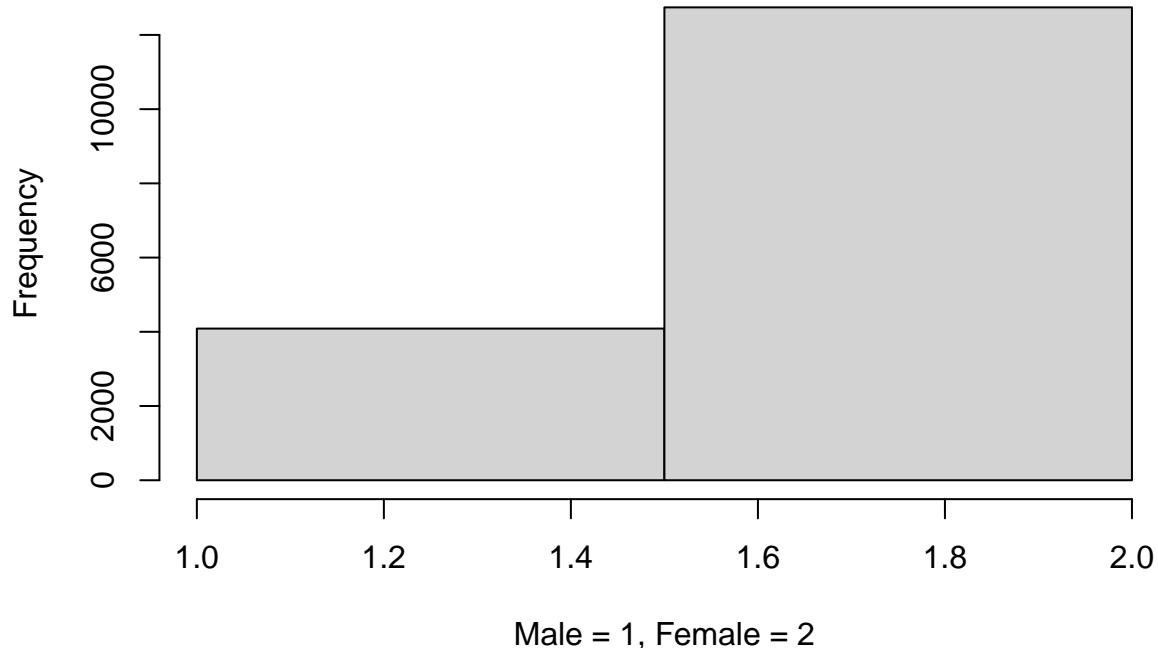
```
#race
hist(data_empi_acc_f_phq$race_binarized[which(!is.na(data_empi_acc_f_phq$race_binarized))],
      main = paste0("Race Histo : n = ", length(which(!is.na(data_empi_acc_f_phq$race))), "; White = ",
      xlab = "white = 1, other = 2",
      breaks = 2)
```

## Race Histo : n = 16830; White = 12174; NW = 4656



```
#f-codes - people usually have BOTH icd9/10 or neither  
  
#Gender breakdown  
#sex  
hist(data_empi_acc_f_phq$sex_binarized[which(!is.na(data_empi_acc_f_phq$sex_binarized))],  
      main = paste0("Sex Histo : n = ", length(which(!is.na(data_empi_acc_f_phq$sex_binarized))), "; M = ",  
      xlab = "Male = 1, Female = 2",  
      breaks = 2)
```

**Sex Histo : n = 16830; M = 4087; F = 12743**



```

##### checking to see whether we have clinical data for patients who have already been segmented as of
### had to do some parsing of tim's spreadsheet: /Users/eballer/BBL/msdepression/data/insurance_images_
### used the following commands to extract the empi and the date -> in tim's they are in the format sub

## more insurance_images_for_depression.csv | cut -f1 -d ',' | perl -pe 's/sub-(.*)/$1/' > tim_empi
## more insurance_images_for_depression.csv | cut -f2 -d ',' | perl -pe 's/ses-(....)(..)(..)/$2/$3/$1'
## more tim_dates | perl -pe 's/(.*)20(..)/$1$2/' | perl -pe 's/(0)(.)\/(.*)(\.)/$2/$3/$4/' | perl

#segmented <- read.csv("/Users/eballer/BBL/msdepression/data/insurance_images_for_depression_with_acces
segmented <- read.csv("/Users/eballer/BBL/msdepression/data/insurance_images_for_depression3.csv", head=1)

names(segmented) <- c("EMPI", "EXAM_DATE", "SES", "ACCESSION_NUM")

#look at the overlap of EMPIs. Won't be able to actually check the matching of scan to person without e
number_of_people_in_tims_group_that_are_also_in_mine <- length(which(segmented$ACCESSION_NUM %in% data_
overlap_group <- data_empi_acc_f_phq[which(data_empi_acc_f_phq$ACCESSION_NUM %in% segmented$ACCESSION_NU

print(paste0("Num people in previously curated group and Erica's group -> ", dim(overlap_group)[1]))

## [1] "Num people in previously curated group and Erica's group -> 1798"

#### PHQ-2 and PHQ-9 summaries

#looking at just people with phq-2 and 9

```

```

phq2_subset <- overlap_group[which(!is.na(overlap_group$PHQ.2)),] #n=4606

phq9_subset <- overlap_group[which(!is.na(overlap_group$PHQ.9)),] #n= 261

#phq-2;
print(paste0("N PHQ-2 = ", sum(!is.na(overlap_group$PHQ.2)), "; n = ", length(which(phq2_subset$PHQ.2 >= 3)))

## [1] "N PHQ-2 = 594; n = 18 meeting criteria for depression (PHQ2) >= 3"

#phq-9
print(paste0("N PHQ-9 = ", sum(!is.na(overlap_group$PHQ.9)), "; noDep/mild/mod/mod-severe/severe = ",
            length(which(phq9_subset$PHQ.9 < 5)), "/",
            length(which(phq9_subset$PHQ.9 >= 5 & phq9_subset$PHQ.9 < 10)), "/",
            length(which(phq9_subset$PHQ.9 >= 10 & phq9_subset$PHQ.9 < 15)), "/",
            length(which(phq9_subset$PHQ.9 >= 15 & phq9_subset$PHQ.9 < 20)), "/",
            length(which(phq9_subset$PHQ.9 >= 20))) #n = 261

## [1] "N PHQ-9 = 22; noDep/mild/mod/mod-severe/severe = 7/6/4/2/3"

#both phq2 and 9

num_both_phq2_and_9 <- length(which(!is.na(overlap_group$PHQ.2) & !is.na(overlap_group$PHQ.9)))
print(paste0("num people with both phq2 and 9 = ", num_both_phq2_and_9)) #n = 0

## [1] "num people with both phq2 and 9 = 0"

#Number of unique EMPIs in phq2/phq9 subsets
unique_phq2.empis <- length(unique(phq2_subset$EMPI)) #n=966
print(paste0("N with phq2 (Unique EMPI) n = ", unique_phq2.empis, " out of ", dim(phq2_subset)[1], "; n = ", dim(phq2_subset)[1]))

## [1] "N with phq2 (Unique EMPI) n = 292 out of 594; n - 9 meeting criteria for depression (PHQ2) >= 3"

unique_phq9.empis <- length(unique(phq9_subset$EMPI)) #n=76
print(paste0("N with phq9 (Unique EMPI) n = ", unique_phq9.empis, " out of ", dim(phq9_subset)[1], "; n = ", dim(phq9_subset)[1]),
      length(unique(phq9_subset$EMPI[phq9_subset$PHQ.9 < 5])), "/",
      length(unique(phq9_subset$EMPI[(phq9_subset$PHQ.9 >= 5 & phq9_subset$PHQ.9 < 10)])), "/",
      length(unique(phq9_subset$EMPI[(phq9_subset$PHQ.9 >= 10 & phq9_subset$PHQ.9 < 15)])), "/",
      length(unique(phq9_subset$EMPI[(phq9_subset$PHQ.9 >= 15 & phq9_subset$PHQ.9 < 20)])), "/",
      length(unique(phq9_subset$EMPI[phq9_subset$PHQ.9 >= 20]))))

## [1] "N with phq9 (Unique EMPI) n = 14 out of 22; noDep/mild/mod/mod-severe/severe = 3/4/3/2/2"

##### ICD 9/10 for Depression
dep_df <- overlap_group[grep("F3", overlap_group$ICD10),] #n = 2123
num_depression <- dim(dep_df)[1]
unique_empi <- which(unique(overlap_group$EMPI) %in% overlap_group$EMPI) #index of the locations of unique empi
num_unique_EMPI_and_dep <- length(unique(dep_df$EMPI)) #n = 493

print(paste0("N with ICD 9/10 codes for depression = ", num_depression))

```

```

## [1] "N with ICD 9/10 codes for depression = 233"

print(paste0("N with ICD 9/10 codes for depression and unique EMPI = ", num_unique_EMPI_and_dep, " out of 123")

## [1] "N with ICD 9/10 codes for depression and unique EMPI = 123 out of total depressed - 233"

##### PHQ2 or 9 + ICD 9/10 Depression
num_phq2_and_dep <- length(which(!is.na(dep_df$PHQ.2))) #n = 960
print(paste0("N with ICD 9/10 codes for depression and phq2 (ALL SCANS) = ", num_phq2_and_dep, " out of 960"))

## [1] "N with ICD 9/10 codes for depression and phq2 (ALL SCANS) = 119 out of 233"

num_phq9_and_dep <- length(which(!is.na(dep_df$PHQ.9))) #n=140
print(paste0("N with ICD 9/10 codes for depression and phq9 (ALL SCANS) = ", num_phq9_and_dep, " out of 140"))

## [1] "N with ICD 9/10 codes for depression and phq9 (ALL SCANS) = 10 out of 233"

#phq-2;
print(paste0("N dep + PHQ-2 = ", sum(!is.na(dep_df$PHQ.2)), "; n = ", length(which(dep_df$PHQ.2) >= 3)), )

## [1] "N dep + PHQ-2 = 119; n = 11 meeting criteria for depression (PHQ2) >= 3"

#phq-9
print(paste0("N dep + PHQ-9 = ", sum(!is.na(dep_df$PHQ.9)), "; noDep/mild/mod/mod-severe/severe = ",
            length(which(dep_df$PHQ.9 < 5)), "/",
            length(which(dep_df$PHQ.9 >= 5 & dep_df$PHQ.9 < 10)), "/",
            length(which(dep_df$PHQ.9 >= 10 & dep_df$PHQ.9 < 15)), "/",
            length(which(dep_df$PHQ.9 >= 15 & dep_df$PHQ.9 < 20)), "/",
            length(which(dep_df$PHQ.9 >= 20)))) #n = 261

## [1] "N dep + PHQ-9 = 10; noDep/mild/mod/mod-severe/severe = 2/3/2/2/1"

##### # unique people with dep dx and phq2/9 -> working on this
dep_dx_phq2_unique <- length(unique(dep_df$EMPI[!is.na(dep_df$PHQ.2)])) #n = 201
print(paste0("N with ICD 9/10 codes for depression and phq2 (unique) = ", dep_dx_phq2_unique, " out of 201"))

## [1] "N with ICD 9/10 codes for depression and phq2 (unique) = 63 out of 233"

dep_dx_phq9_unique <- length(unique(dep_df$EMPI[!is.na(dep_df$PHQ.9)])) #n = 43
print(paste0("N with ICD 9/10 codes for depression and phq9 (unique) = ", dep_dx_phq9_unique, " out of 43"))

## [1] "N with ICD 9/10 codes for depression and phq9 (unique) = 8 out of 233"

#####

##### ICD 9/10 for Healthy Group
healthy_df <- overlap_group[(-grep("F3", overlap_group$ICD10)),] #n = 14707

```

```

num_healthy <- dim(healthy_df)[1]
unique_empi <- which(unique(healthy_df$EMPI) %in% healthy_df$EMPI) #index of the locations of unique empi
num_unique_EMPI_and_healthy <- length(unique(healthy_df$EMPI)) #n = 3244

print(paste0("N without Depression/ICD 9/10 codes for healthy = ", num_healthy))

## [1] "N without Depression/ICD 9/10 codes for healthy = 1565"

print(paste0("N without Depression/ ICD 9/10 codes for healthy and unique EMPI = ", num_unique_EMPI_and_healthy))

## [1] "N without Depression/ ICD 9/10 codes for healthy and unique EMPI = 748 out of 1565"

##### PHQ2 or 9 + No Depression ICD9/10
num_phq2_and_healthy <- length(which(!is.na(healthy_df$PHQ.2))) #n = 3646
print(paste0("N without Depression/ ICD 9/10 codes for healthy and phq2 (ALL SCANS) = ", num_phq2_and_healthy))

## [1] "N without Depression/ ICD 9/10 codes for healthy and phq2 (ALL SCANS) = 475 out of 1565"

#phq-2;
print(paste0("N healthy w/PHQ-2 = ", sum(!is.na(healthy_df$PHQ.2)), "; n = ", length(which(healthy_df$PHQ.2) >= 3)))

## [1] "N healthy w/PHQ-2 = 475; n = 7 meeting criteria for depression (PHQ2) >= 3"

num_phq9_and_healthy <- length(which(!is.na(healthy_df$PHQ.9))) #n=121
print(paste0("N without Depression/ ICD 9/10 codes for healthy and phq9 (ALL SCANS) = ", num_phq9_and_healthy))

## [1] "N without Depression/ ICD 9/10 codes for healthy and phq9 (ALL SCANS) = 12 out of 1565"

#phq-9
print(paste0("N healthy w/PHQ-9 = ", sum(!is.na(healthy_df$PHQ.9)), "; noDep/mild/mod/mod-severe/severe = ",
            length(which(healthy_df$PHQ.9 < 5)), "/",
            length(which(healthy_df$PHQ.9 >= 5 & healthy_df$PHQ.9 < 10)), "/",
            length(which(healthy_df$PHQ.9 >= 10 & healthy_df$PHQ.9 < 15)), "/",
            length(which(healthy_df$PHQ.9 >= 15 & healthy_df$PHQ.9 < 20)), "/",
            length(which(healthy_df$PHQ.9 >= 20)))) #n = 55/34/21/1/10

## [1] "N healthy w/PHQ-9 = 12; noDep/mild/mod/mod-severe/severe = 5/3/2/0/2"

##### # unique people without dep dx and phq2/9
dep_dx_phq2_unique <- length(unique(healthy_df$EMPI[!is.na(healthy_df$PHQ.2)])) #n = 201
print(paste0("N without Depression/ ICD 9/10 codes for healthy and phq2 (unique) = ", dep_dx_phq2_unique))

## [1] "N without Depression/ ICD 9/10 codes for healthy and phq2 (unique) = 229 out of 1565"

dep_dx_phq9_unique <- length(unique(healthy_df$EMPI[!is.na(healthy_df$PHQ.9)])) #n = 43
print(paste0("N without Depression/ ICD 9/10 codes for healthy and phq9 (unique) = ", dep_dx_phq9_unique))

## [1] "N without Depression/ ICD 9/10 codes for healthy and phq9 (unique) = 6 out of 1565"

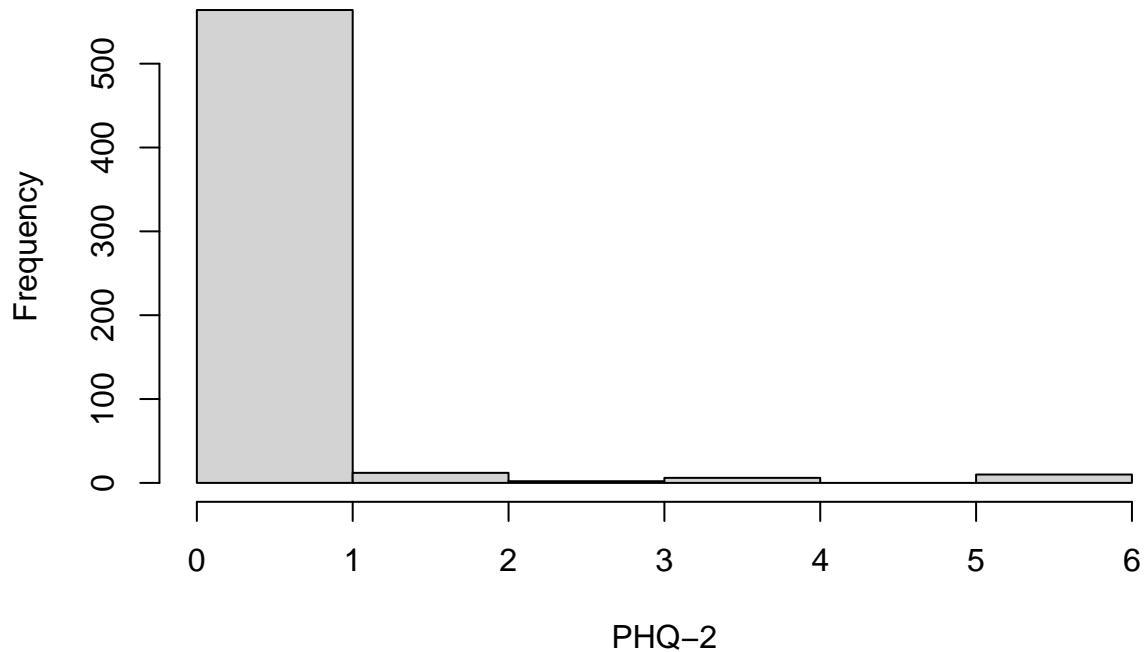
```

```
####
```

```
#demographics and lab histograms  
#print histogram of phq-2/9s for people
```

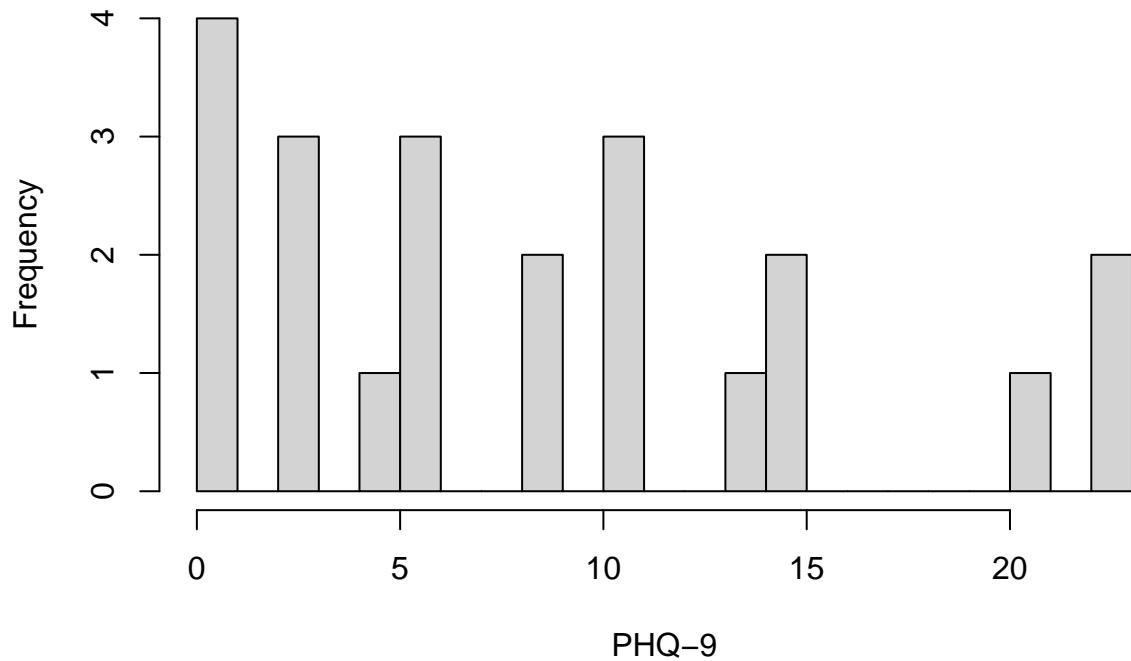
```
hist(overlap_group$PHQ.2[which(!is.na(overlap_group$PHQ.2))], main = paste0("PHQ-2 Histo : n = ", length(
```

**PHQ-2 Histo : n = 594**



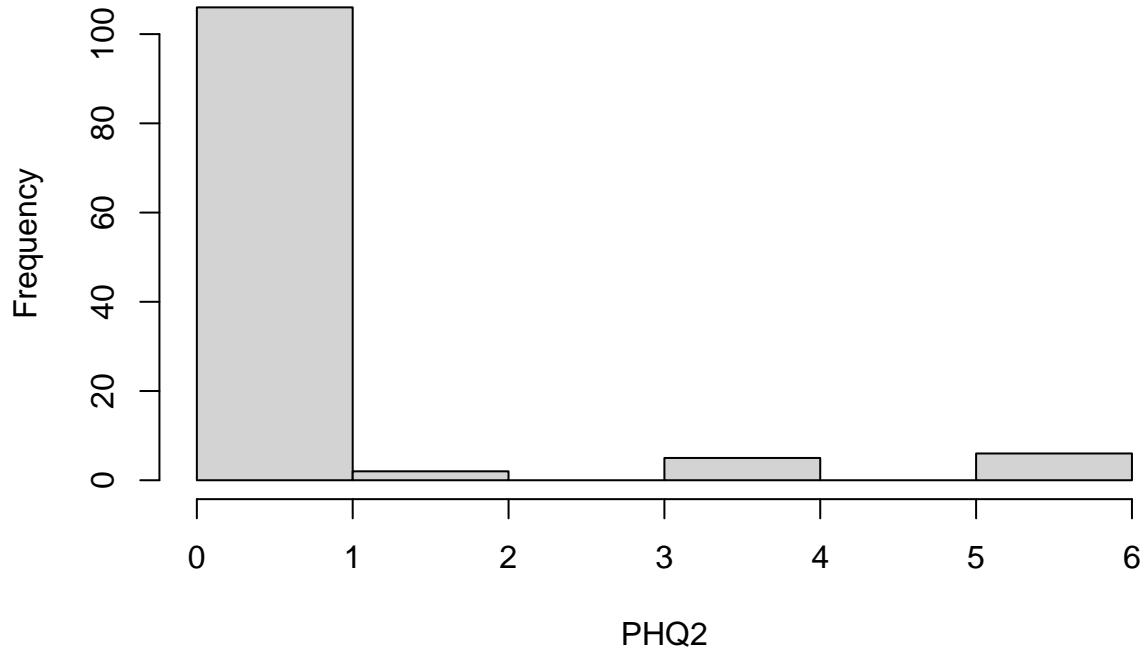
```
hist(overlap_group$PHQ.9[which(!is.na(overlap_group$PHQ.9))], main = paste0("PHQ-9 Histo : n = ", length(
```

**PHQ-9 Histo : n = 22**



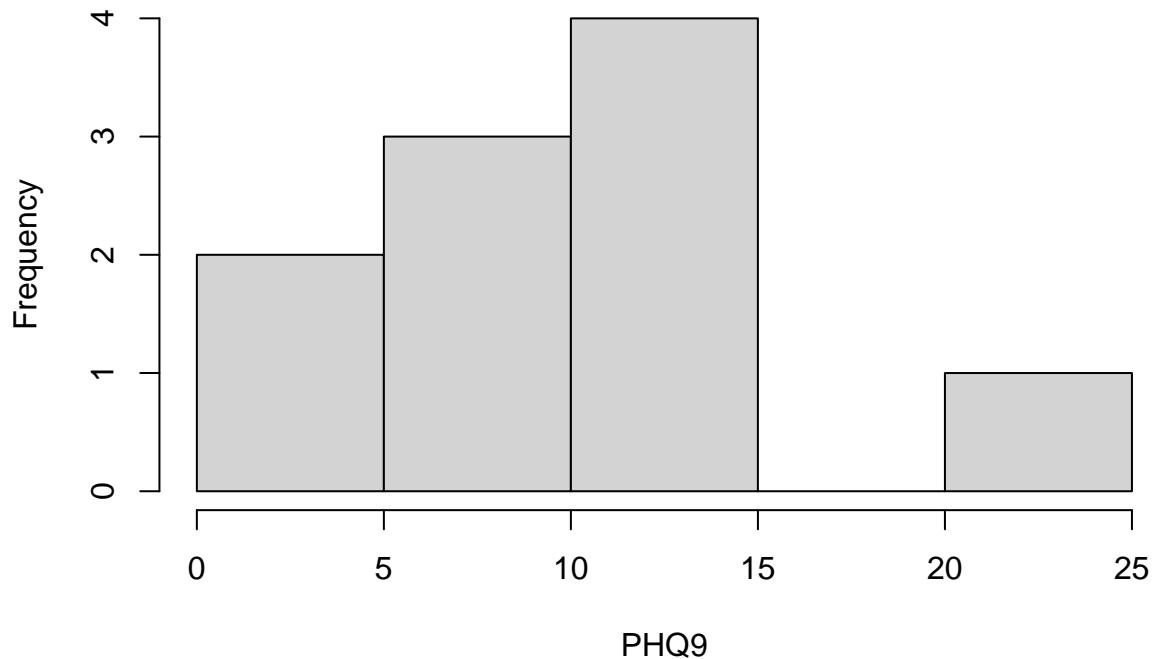
```
#fcode histos
hist(dep_df$PHQ.2, main = paste0("PHQ2 and Dep Histo : n = ", num_phq2_and_dep), xlab = paste0("PHQ2"))
```

## PHQ2 and Dep Histo : n = 119



```
hist(dep_df$PHQ.9, main = paste0("PHQ9 and Dep Histo : n = ", num_phq9_and_dep), xlab = paste0("PHQ9"))
```

### PHQ9 and Dep Histo : n = 10

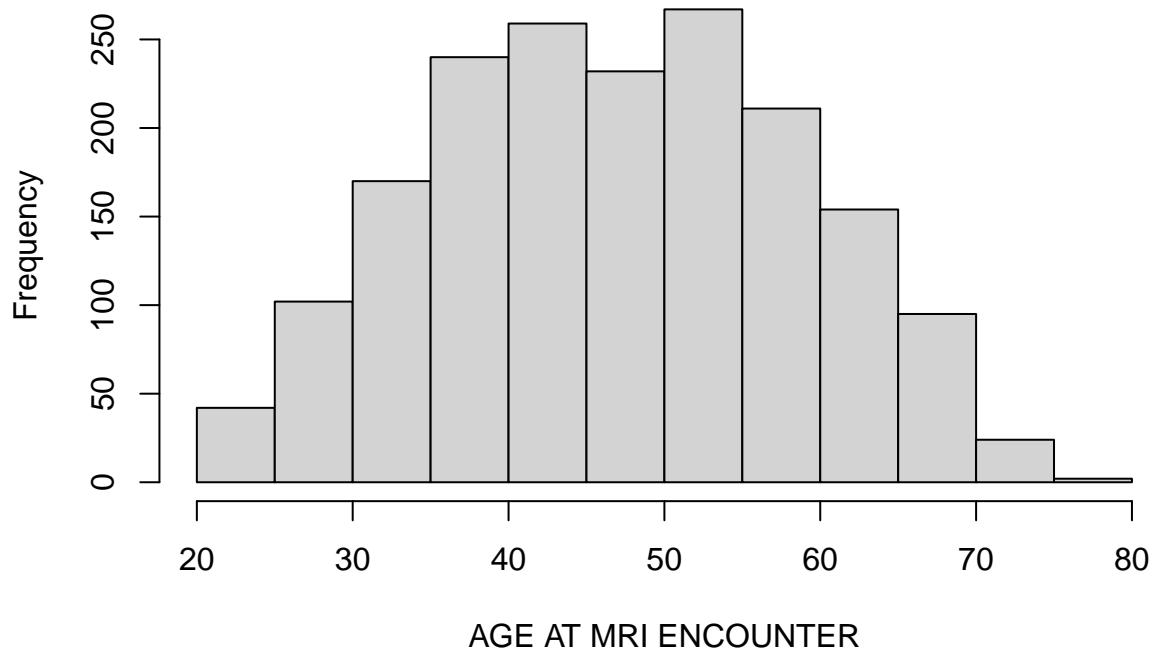


```
###
```

```
#age histogram
```

```
hist(overlap_group$MRI_ENC_AGE[which(!is.na(overlap_group$MRI_ENC_AGE))], main = paste0("Age Histo : n = ",
```

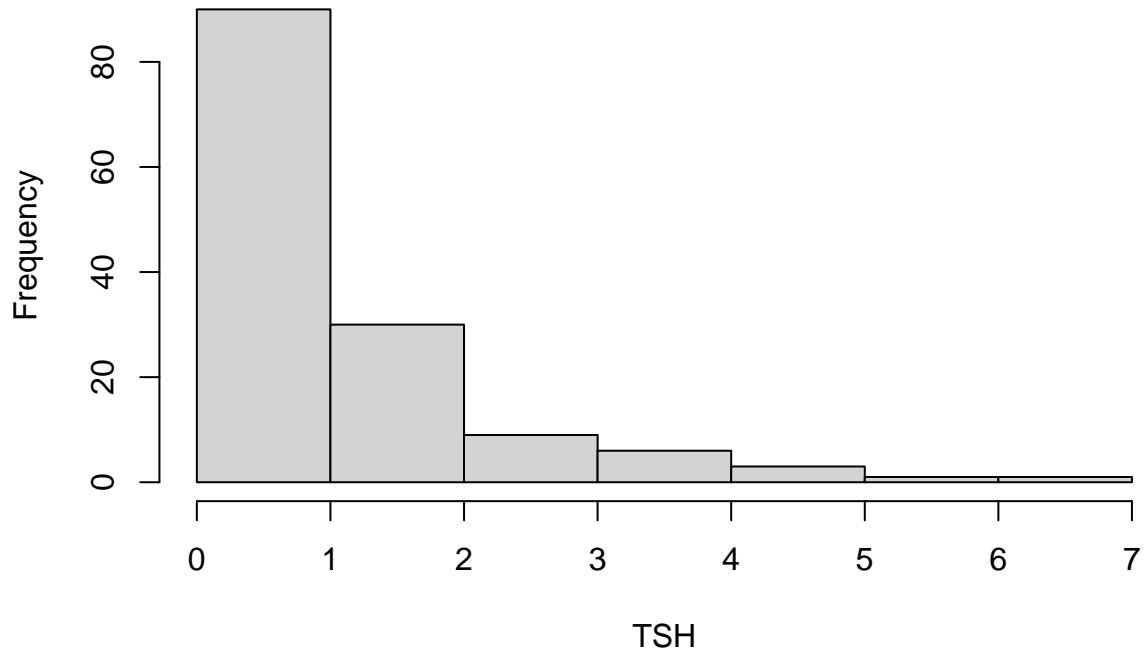
**Age Histo : n = 1798**



```
#tsh
```

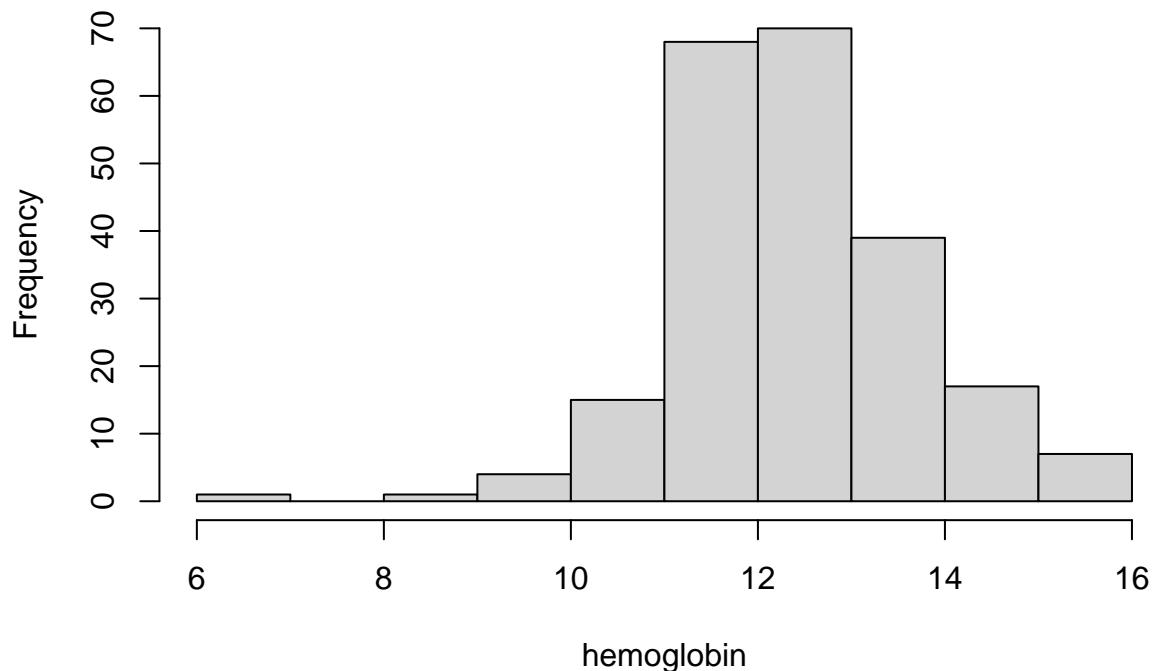
```
hist(overlap_group$TSH[which(!is.na(overlap_group$TSH))], main = paste0("TSH Histo : n = ", length(which(is.na(overlap_group$TSH))))]
```

**TSH Histo : n = 140**



```
#hemoglobin
hist(overlap_group$hemoglobin[which(!is.na(overlap_group$hemoglobin))], main = paste0("Hemoglobin Histo"))
```

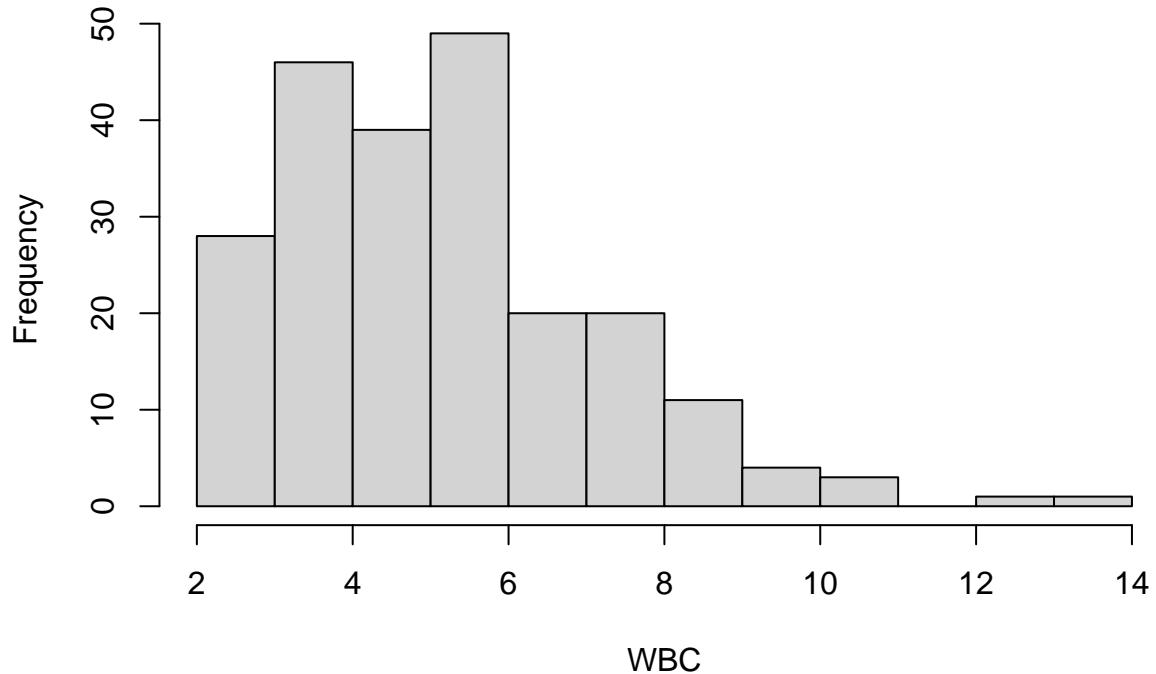
**Hemoglobin Histo : n = 222**



```
#wbc
```

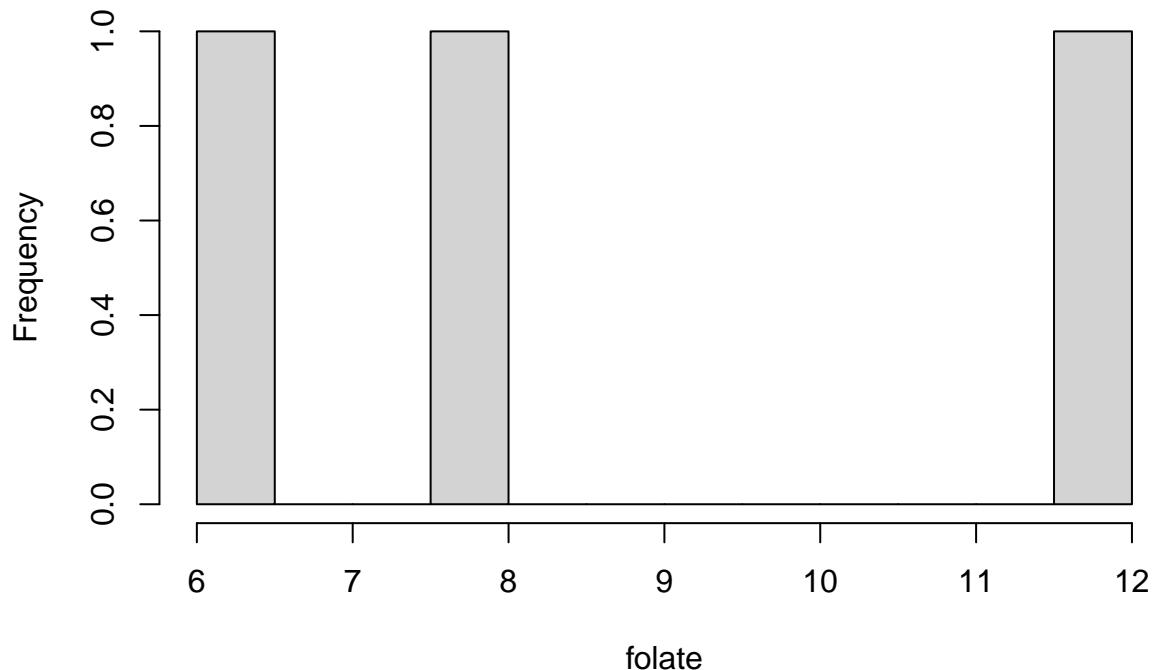
```
hist(overlap_group$WBC[which(!is.na(overlap_group$WBC))], main = paste0("WBC Histo : n = ", length(which(is.na(overlap_group$WBC))))]
```

**WBC Histo : n = 222**



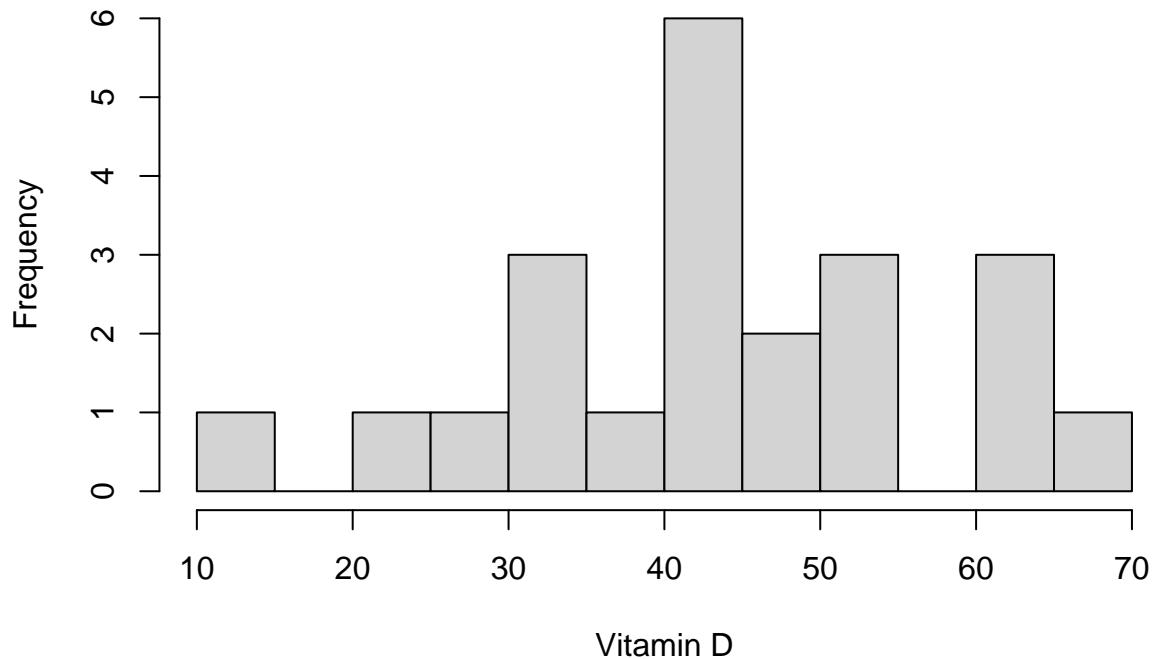
```
#folate
hist(overlap_group$FOLATE[which(!is.na(overlap_group$FOLATE))], main = paste0("Folate Histo : n = ", length(overlap_group$FOLATE)))
```

**Folate Histo : n = 3**



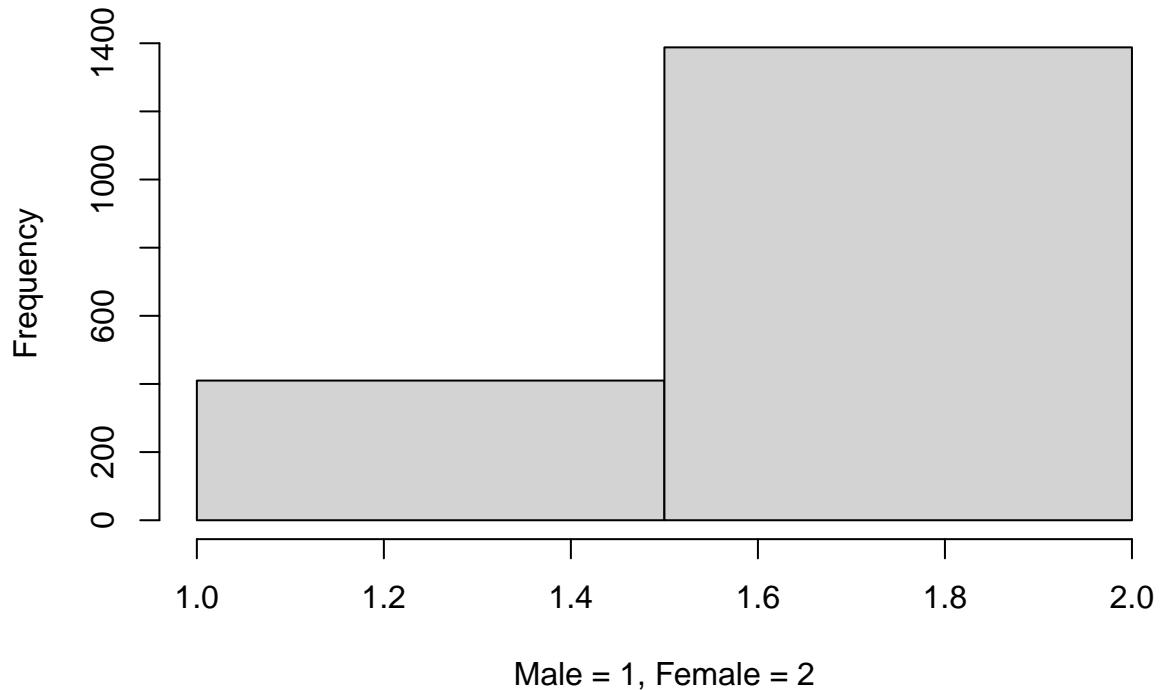
```
#vitamin D
hist(overlap_group$VitD[which(!is.na(overlap_group$VitD))], main = paste0("Vitamin D Histo : n = ", length(overlap_group$VitD)))
```

## Vitamin D Histo : n = 22

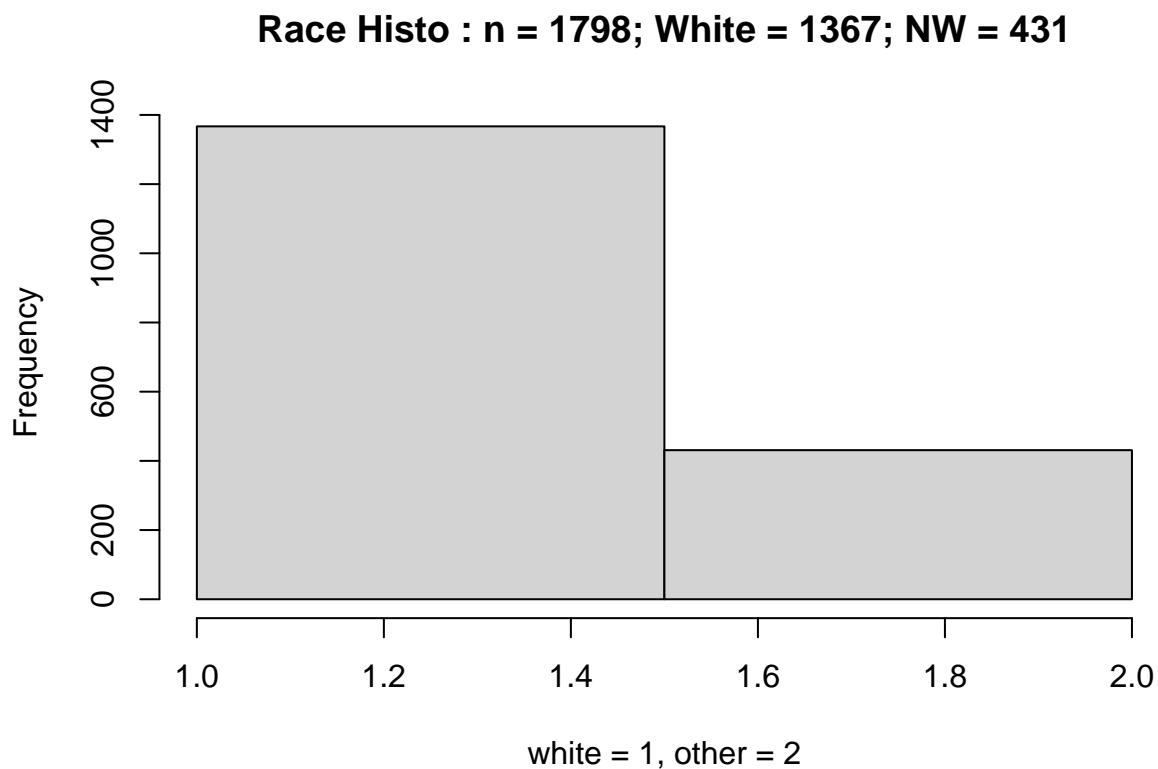


```
#sex
hist(overlap_group$sex_binarized[which(!is.na(overlap_group$sex_binarized))],
      main = paste0("Sex Histo : n = ", length(which(!is.na(overlap_group$sex_binarized))), "; M = ", length(which(is.na(overlap_group$sex_binarized))), " ; SD = ", round(sd(overlap_group$sex_binarized), 2)),
      xlab = "Male = 1, Female = 2",
      breaks = 2)
```

**Sex Histo : n = 1798; M = 410; F = 1388**



```
#race
hist(overlap_group$race_binarized[which(!is.na(overlap_group$race_binarized))],
      main = paste0("Race Histo : n = ", length(which(!is.na(overlap_group$race))), "; White = ", length(
      xlab = "white = 1, other = 2",
      breaks = 2)
```





```

#####
##### Step 1- make F3* depressed and healthy groups* #####
#####
### This is identical to up above, but repeated so I can run this chunk separately

#####
### separate out depressed group
#####
print("separate out depressed group")

## [1] "separate out depressed group"

dep_df <- data_empi_acc_f_phq[grep("F3", data_empi_acc_f_phq$ICD10),] #n = 2123
num_depression <- dim(dep_df)[1]
num_unique_EMPI_and_dep <- length(unique(dep_df$EMPI)) #n = 493
print(paste0("N with ICD 9/10 codes for depression = ", num_depression))

## [1] "N with ICD 9/10 codes for depression = 2123"

print(paste0("N with ICD 9/10 codes for depression and unique EMPI in whole depression group = ", num_u

## [1] "N with ICD 9/10 codes for depression and unique EMPI in whole depression group = 493 out of 2123

#####
### separate out healthy group
#####
print("separate out healthy group")

## [1] "separate out healthy group"

healthy_df <- data_empi_acc_f_phq[-grep("F3", data_empi_acc_f_phq$ICD10),] #n = 14707
num_healthy <- dim(healthy_df)[1]
num_unique_EMPI_and_healthy <- length(unique(healthy_df$EMPI)) #n = 3244
print(paste0("N with WITHOUT depression ICD 9/10 codes (i.e. \"healthy\"): ", num_healthy))

## [1] "N with WITHOUT depression ICD 9/10 codes (i.e. \"healthy\"): 14707"

print(paste0("N with WITHOUT ICD 9/10 depression codes for depression and unique EMPI = ", num_unique_EMPI

## [1] "N with WITHOUT ICD 9/10 depression codes for depression and unique EMPI = 3244 out of 14707

#####
##### Step 2- Identify PHQ2 >=3 in healthy group #####
#####

print("Get people in healthy group with phq2s for dep/healthy")

## [1] "Get people in healthy group with phq2s for dep/healthy"

```

```

# num in healthy group with PHQ2
phq2_subset <- healthy_df[which(!is.na(healthy_df$PHQ.2)),] #n=3646, unique 698
depressed_phq2_subset <- phq2_subset[which(phq2_subset$PHQ.2 >= 3),] #n = 47
depressed_unique_phq2.empis <- unique(depressed_phq2_subset$EMPI) #n=10
healthy_phq2_subset <- phq2_subset[which(phq2_subset$PHQ.2 == 0),] #n = 3326
healthy_unique_phq2.empis <- unique(healthy_phq2_subset$EMPI) #n = 631

print(paste0("PHQ2 breakdown in healthy group (total n with score = ", dim(phq2_subset)[1], "[unique =
## [1] "PHQ2 breakdown in healthy group (total n with score = 3646[unique = 698]): PHQ2 >= 3 :47[unique

#####
##### Step 3- Identify PHQ9 >=10 in healthy group #####
#####

print("Get people in healthy group with phq9s for dep/healthy")

## [1] "Get people in healthy group with phq9s for dep/healthy"

# num in healthy group with PHQ9
phq9_subset <- healthy_df[which(!is.na(healthy_df$PHQ.9)),] #n=121
depressed_phq9_subset <- phq9_subset[which(phq9_subset$PHQ.9 >= 10),] #n = 32
depressed_unique_phq9.empis <- unique(depressed_phq9_subset$EMPI) #n=9
healthy_phq9_subset <- phq9_subset[which(phq9_subset$PHQ.9 == 0),] #n = 21
healthy_unique_phq9.empis <- unique(healthy_phq9_subset$EMPI) #n =4

print(paste0("PHQ9 breakdown in healthy group (total n with score = ", dim(phq9_subset)[1], "[unique =
## [1] "PHQ9 breakdown in healthy group (total n with score = 121[unique = 25]): PHQ9 >= 10 :32[unique

#####
#Step 4 - Combine for depressed group and healthy groups #
#####

print("Combine groups for new cohort")

## [1] "Combine groups for new cohort"

gain_in_dep_group <- length(depressed_unique_phq2.empis) + length(depressed_unique_phq9.empis)
dep_num_icd10_plus_phq2_and_phq9 <- num_unique_EMPI_and_dep + gain_in_dep_group # we can do this because
print(paste0("number of UNIQUE people with an icd10 code for Depression OR scored positive on PHQ2(>=3) or PHQ9(>=10) = ", gain_in_dep_group))

## [1] "number of UNIQUE people with an icd10 code for Depression OR scored positive on PHQ2(>=3) or PHQ9(>=10) = 47

healthy_num_no_icd10_plus_phq2_and_phq9 <- num_unique_EMPI_and_healthy - gain_in_dep_group #take number

print(paste0("number of UNIQUE people with NO icd10 code for Depression and NEVER had a PHQ2(>=3) or PHQ9(>=10) = ", healthy_num_no_icd10_plus_phq2_and_phq9))

## [1] "number of UNIQUE people with NO icd10 code for Depression and NEVER had a PHQ2(>=3) or PHQ9(>=10) = 32

```

```

healthy_num_no_icd10_plus_phq2_and_phq9_MUST_HAVE_PHQ2_OR_9 <- length(healthy_unique_phq2.empis) + length(healthy_unique_phq9.empis)
print(paste0("number of UNIQUE people with NO icd10 code for Depression and HAS HAD AT LEAST 1 PHQ2 or 9"))

## [1] "number of UNIQUE people with NO icd10 code for Depression and HAS HAD AT LEAST 1 PHQ2 or 9 that

print(paste0("Cleanest cohort (depressed in icd9/10 or Phq2/9; healthy by at least one phq2/9 and never depre

## [1] "Cleanest cohort (depressed in icd9/10 or Phq2/9; healthy by at least one phq2/9 and never depre

##### THESE ARE THE DATAFRAMES I WANT TO USE GOING FORWARD #####
#all empirs for depressed group include people with depression dx, and people who got in for phq2 or phq9
empis_for_depressed_group <- append(
  append(unique(dep_df$EMPI), depressed_unique_phq2.empis), depressed_unique_phq9.empis)

final_depressed_group_withICD_AND_depressed_phq2_or_9 <-
  data_empi_acc_f_phq %>%
filter(EMPI %in% empirs_for_depressed_group)

#-----
empis_for_healthy_group <- append(healthy_unique_phq2.empis, healthy_unique_phq9.empis)

final_healthy_group_withNOICD_AND_phq2_or_9_0<-
  data_empi_acc_f_phq %>%
filter(EMPI %in% empirs_for_healthy_group)

#-----
### add a column to the original dataframe with a 0 if not in any group, -1 if healthy, and 1 if depressed
data_empi_acc_f_phq$depGroupVar = 0
data_empi_acc_f_phq$depGroupVar[which(data_empi_acc_f_phq$EMPI %in% empirs_for_depressed_group)] = 1
data_empi_acc_f_phq$depGroupVar[which(data_empi_acc_f_phq$EMPI %in% empirs_for_healthy_group)] = -1
data_empi_acc_f_phq$depGroupVar = as.factor(data_empi_acc_f_phq$depGroupVar)

#####
#           Step 5 - Read in Fascicle info          #
#####

fascicle_proportions <- read.csv(paste0(homedir, "/results/fascicle_volumes_all_subjects_roi_n2336.csv"))

#fascicle names are contained in all but the first 2 columns of the fascicle_proportions df
fascicle_names <- names(fascicle_proportions[3:dim(fascicle_proportions)[2]])
#####

#           Step 6 - Merge with data_empi          #
#####

df_demo_and_fascicles <- merge(data_empi_acc_f_phq, fascicle_proportions, by = c("EMPI", "EXAM_DATE"))

#####
#   Step 7 - Regressions in dep/non dep          #
#####

#isolate out only the depressed group and healthy group
dep_and_healthy_groups_for_ICD_analysis <- df_demo_and_fascicles[df_demo_and_fascicles$depGroupVar != 0]

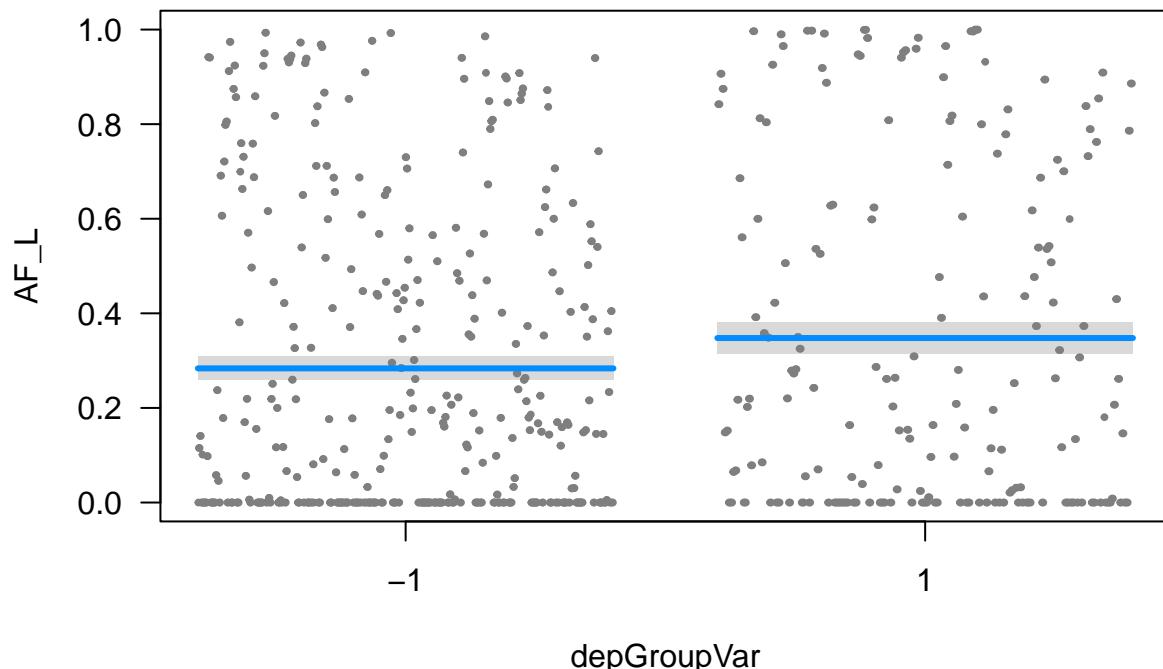
```

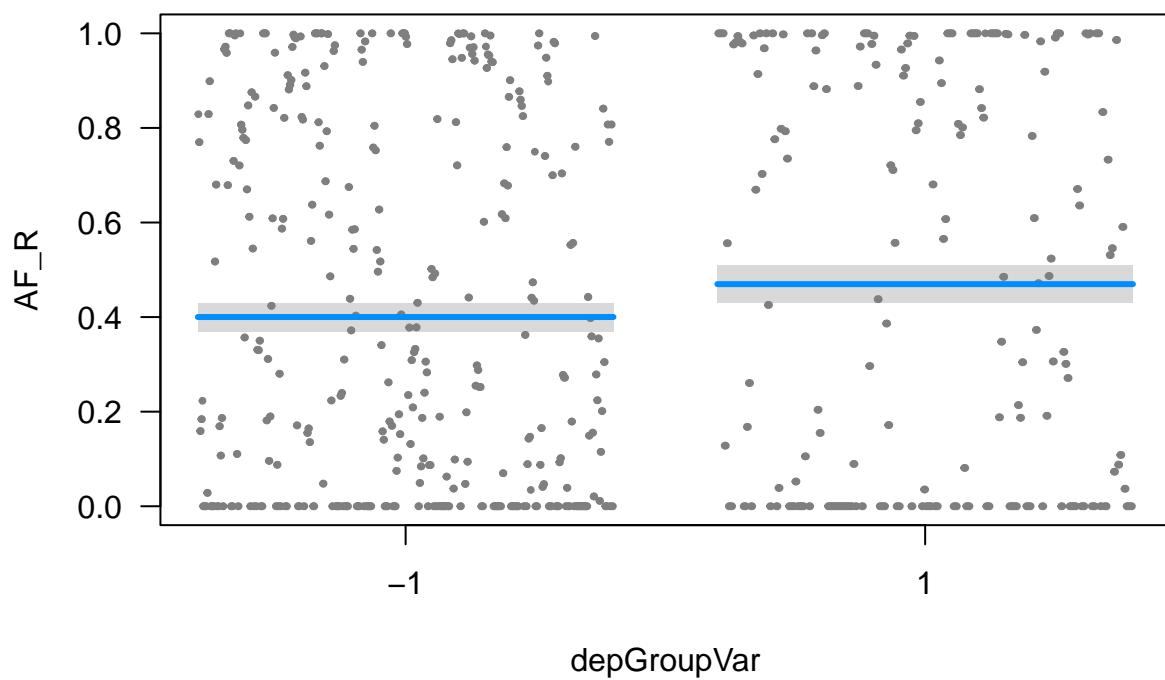
```

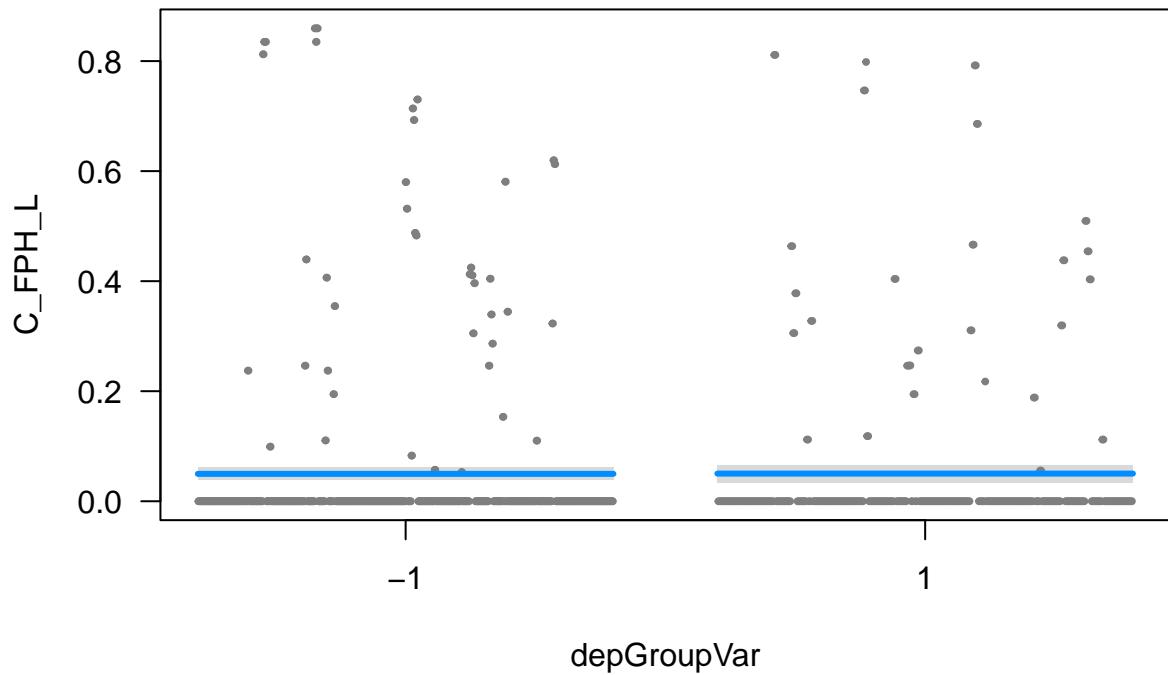
#lm
fascicle_lm <- lapply(fascicle_names, function(x)
{
  lm(substitute(i ~ depGroupVar, list(i = as.name(x))), data = dep_and_healthy_groups_for_ICD_analysis)
})
names(fascicle_lm) <- fascicle_names

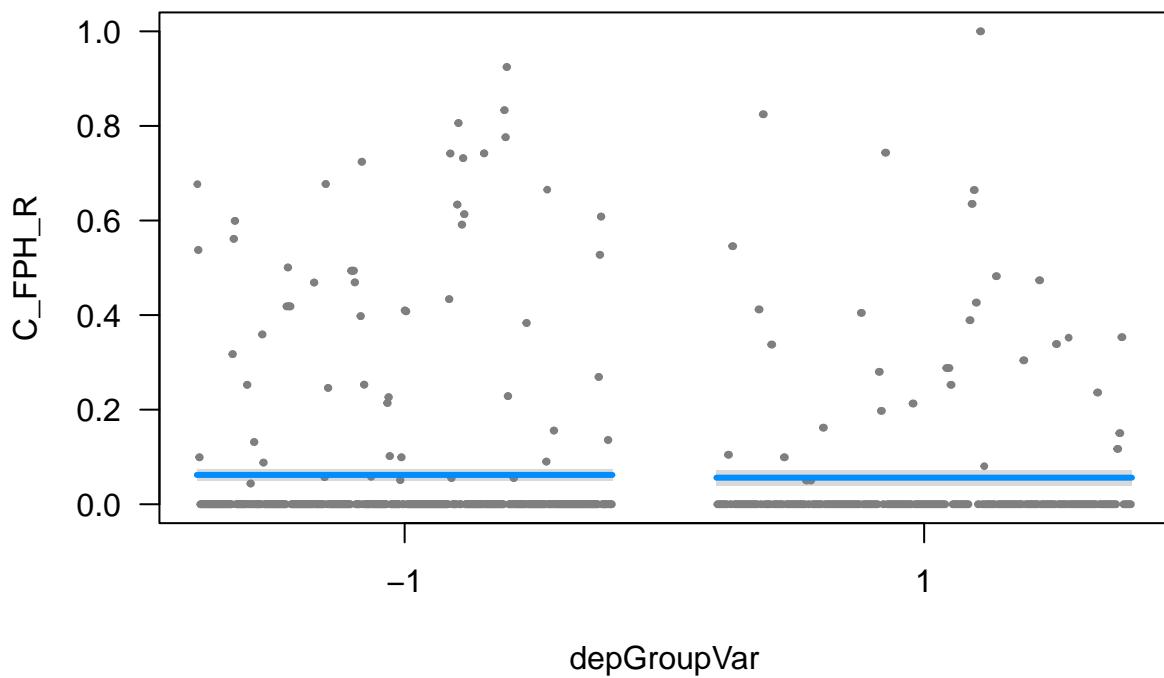
#visreg
sapply(fascicle_lm, visreg)

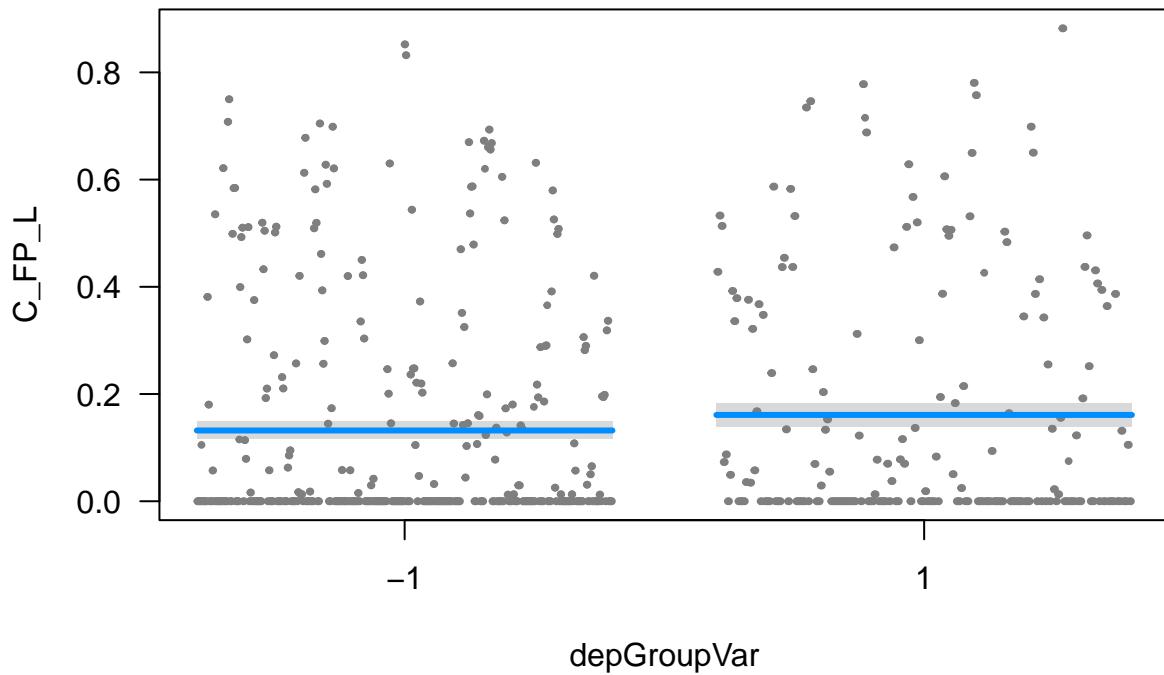
```

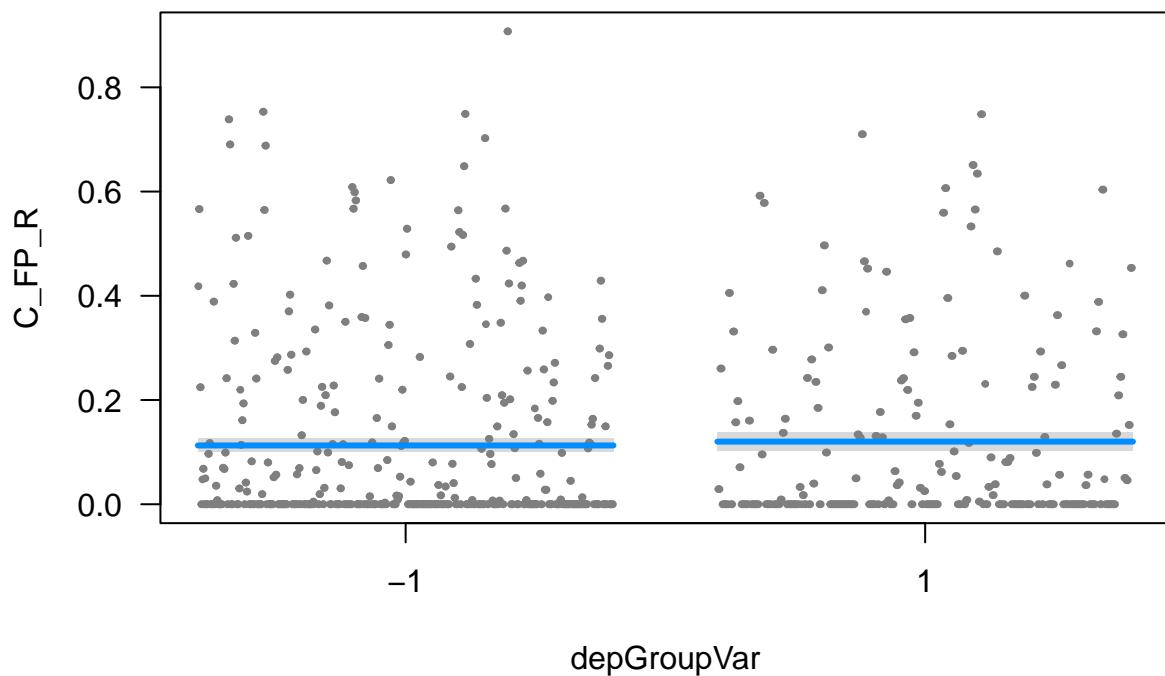


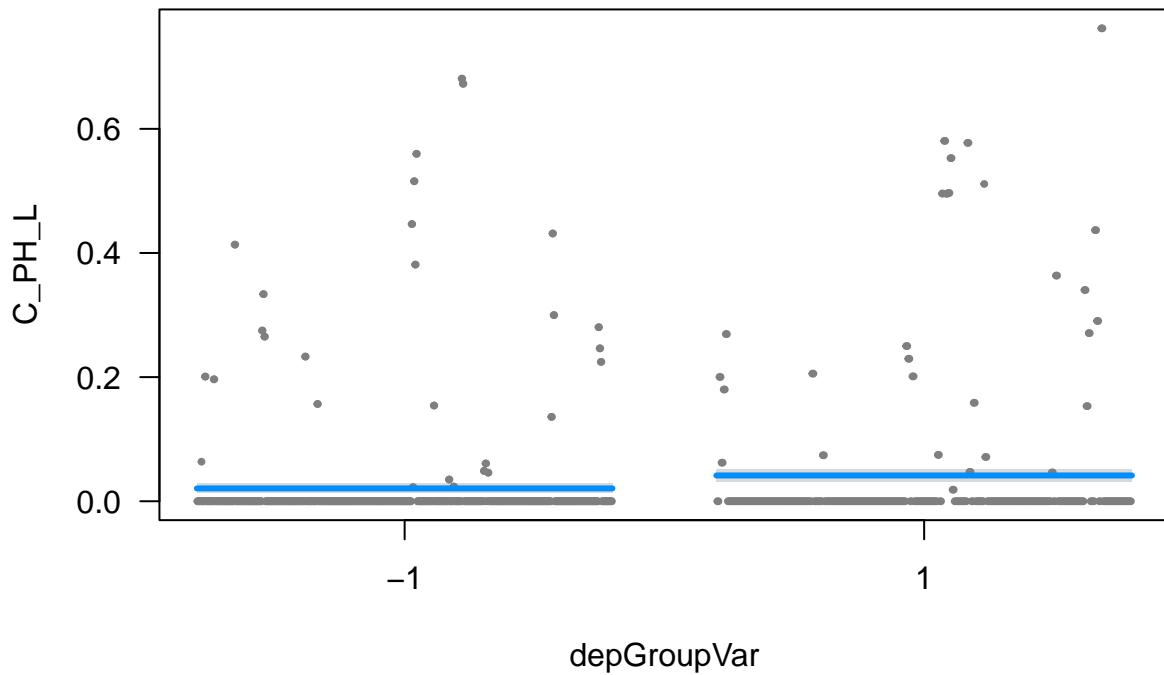


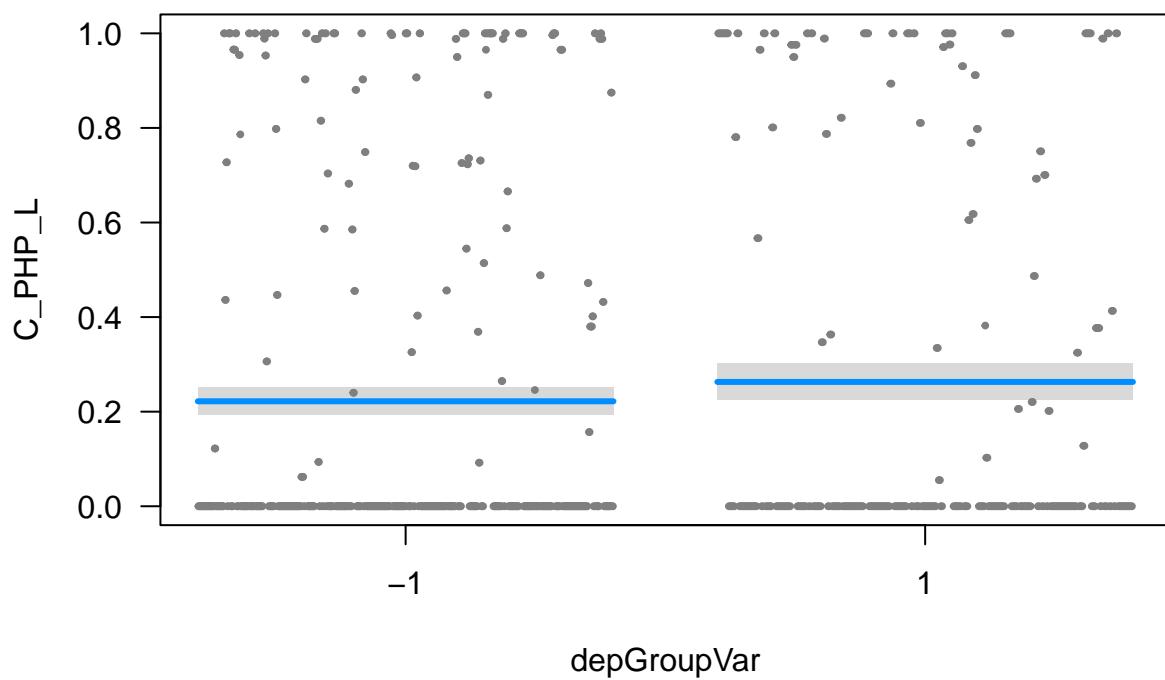


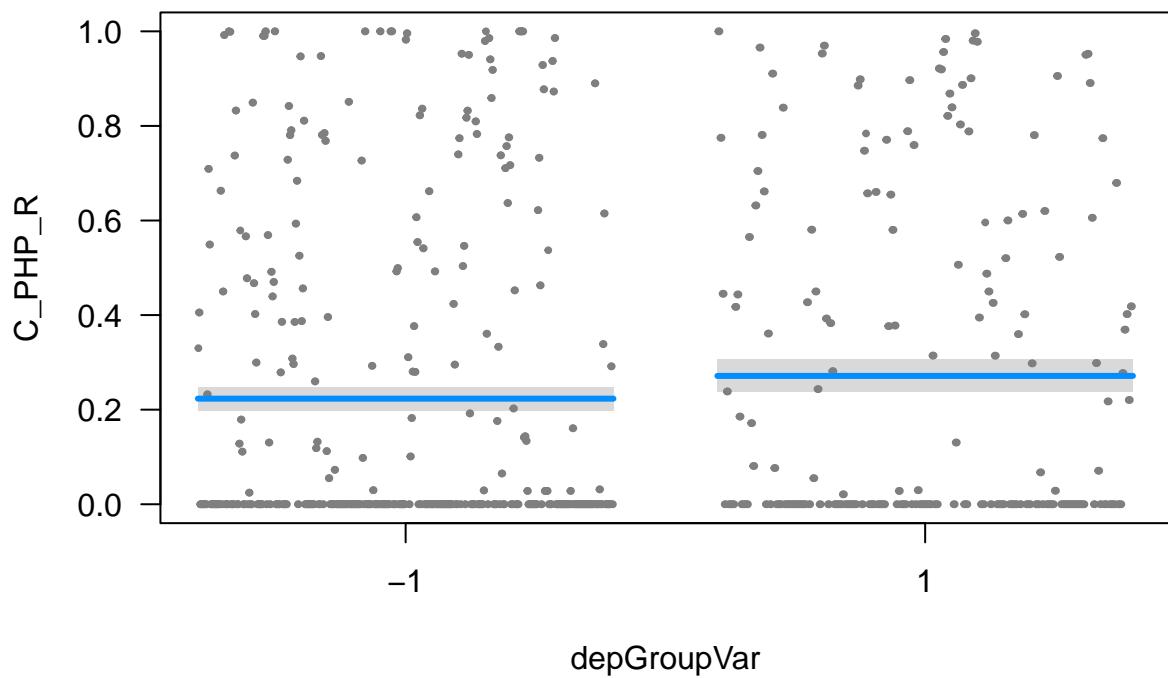


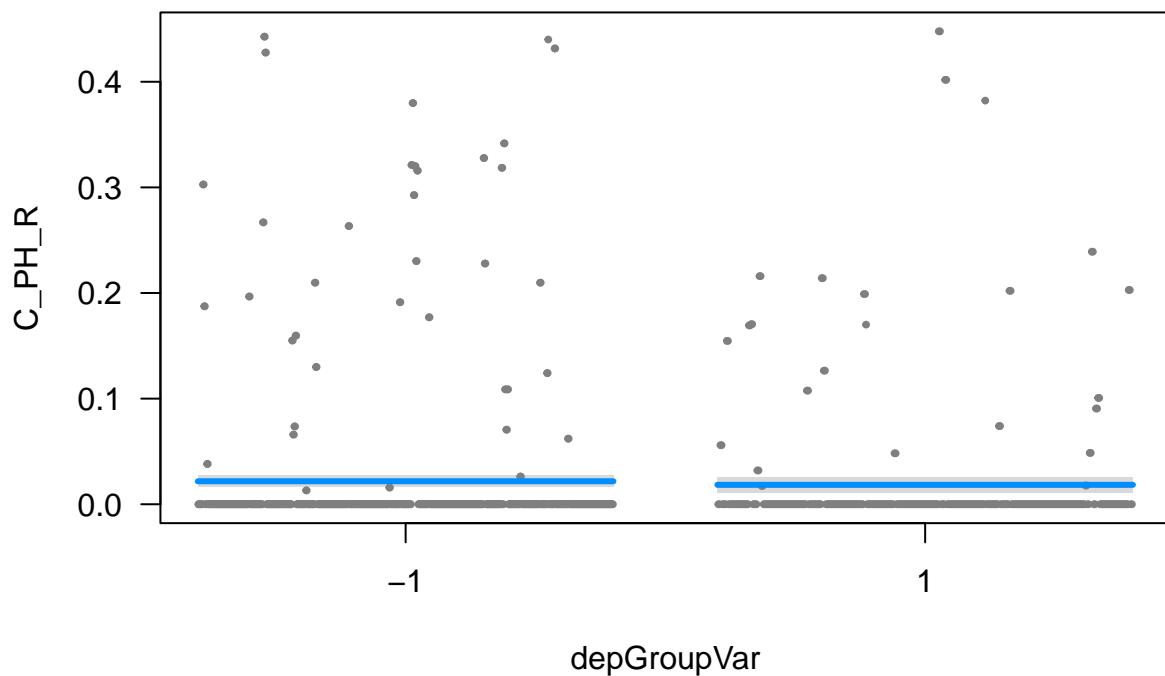


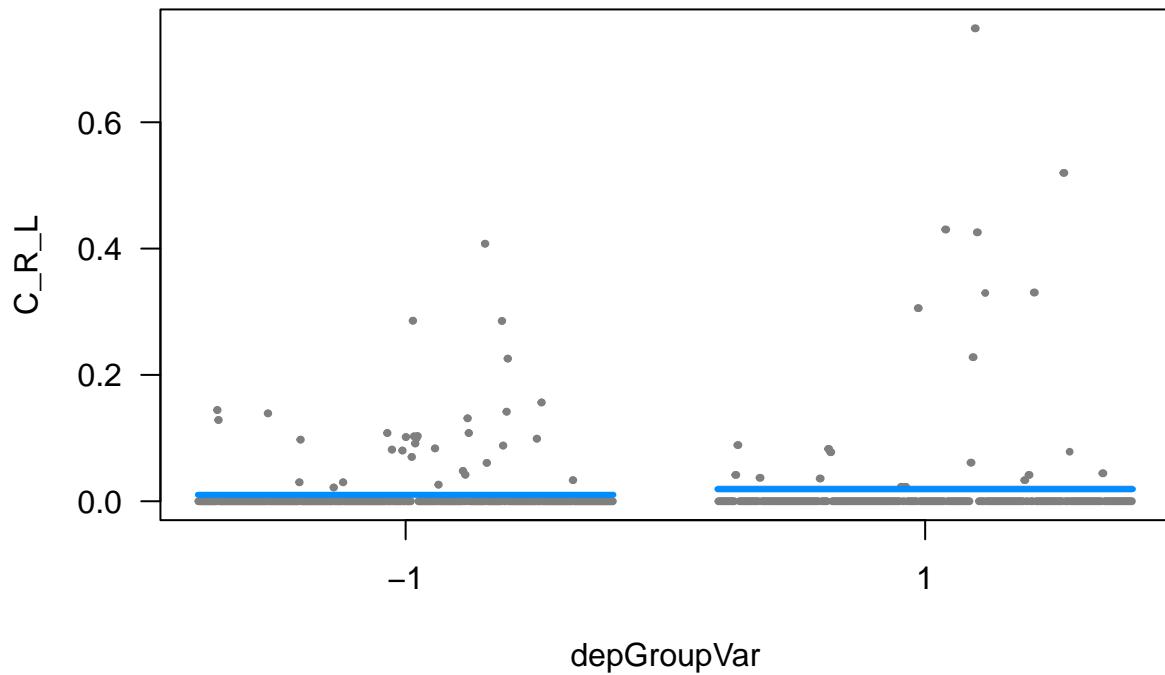


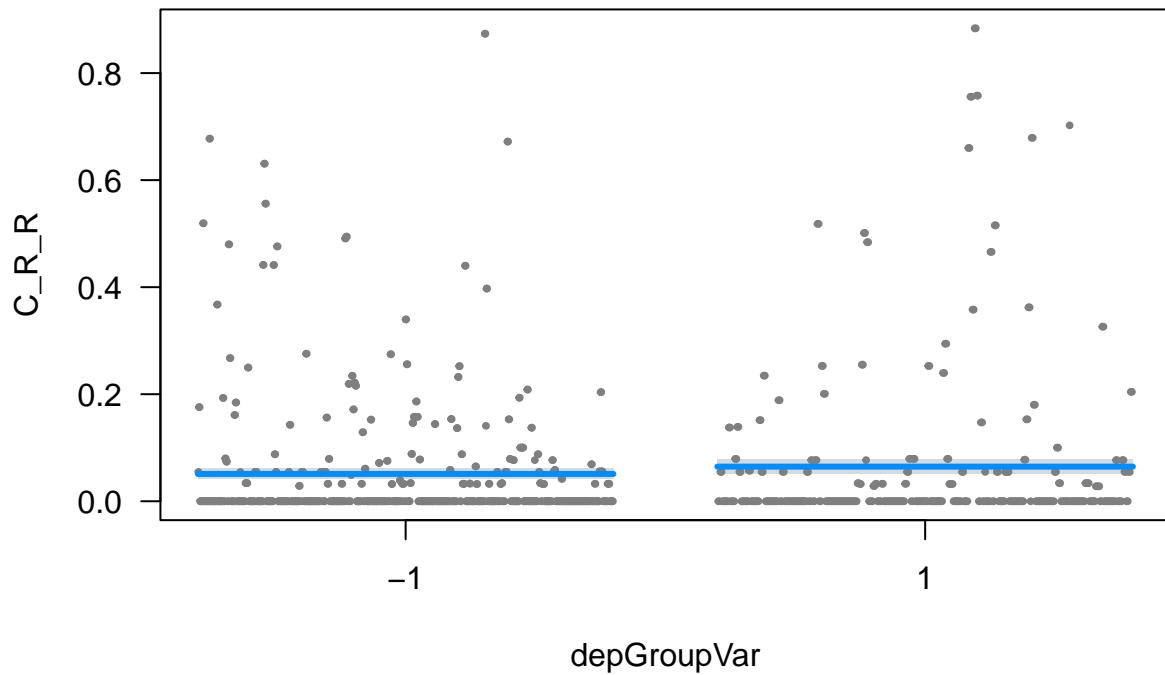


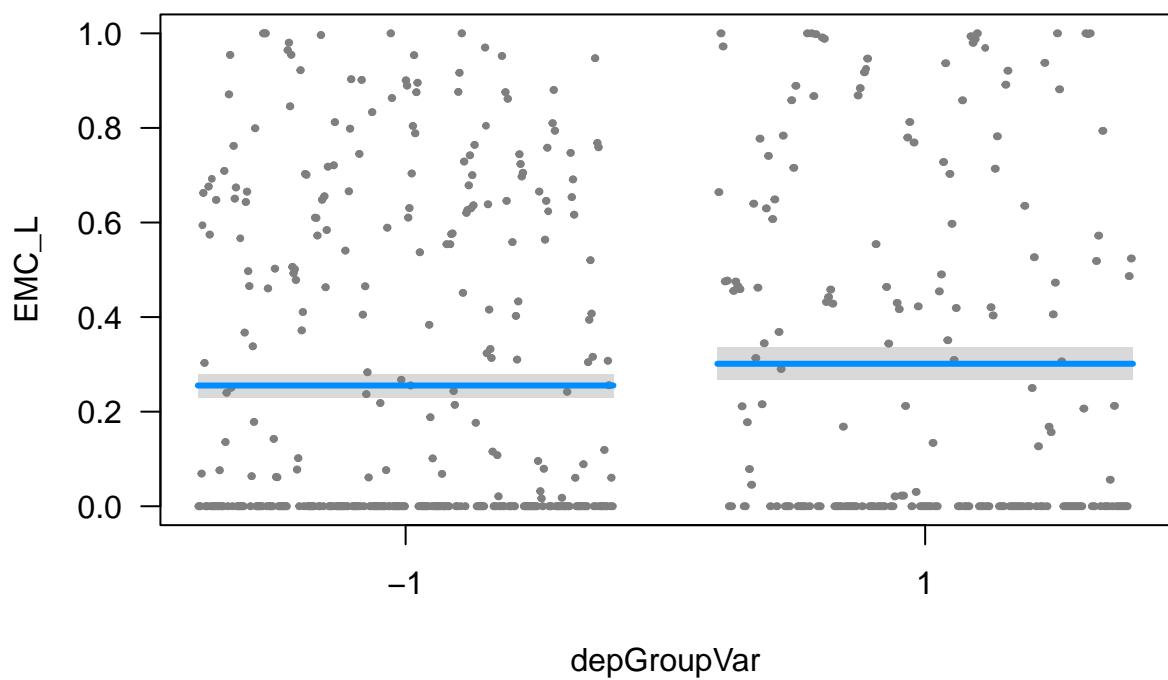


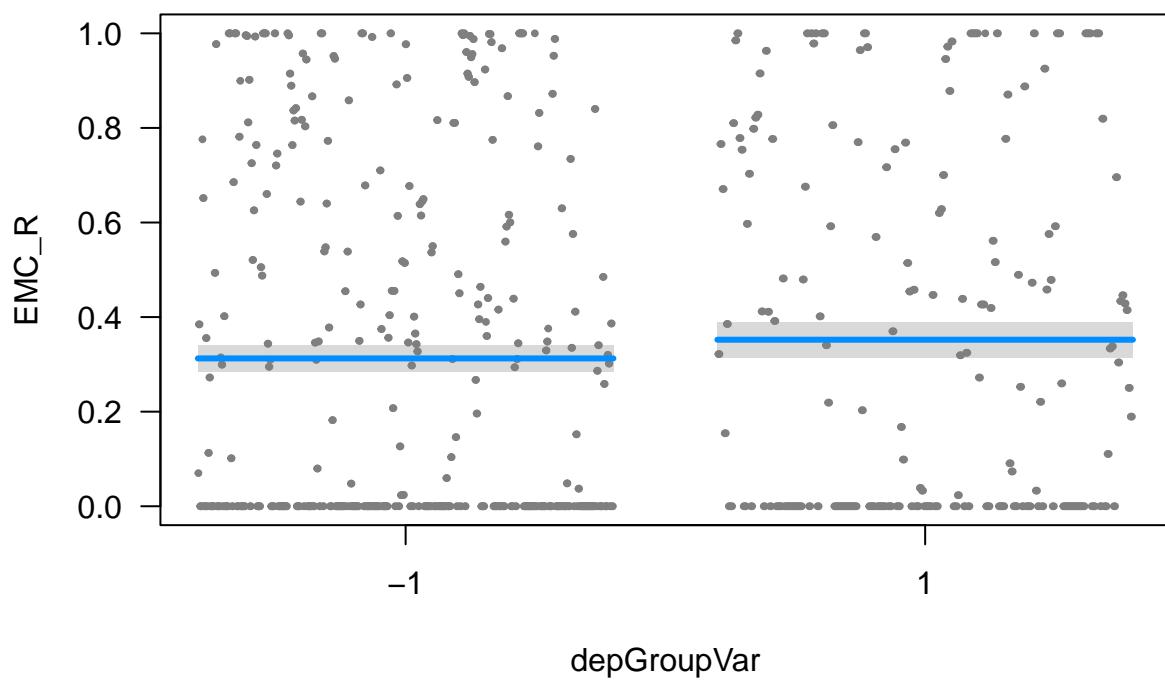


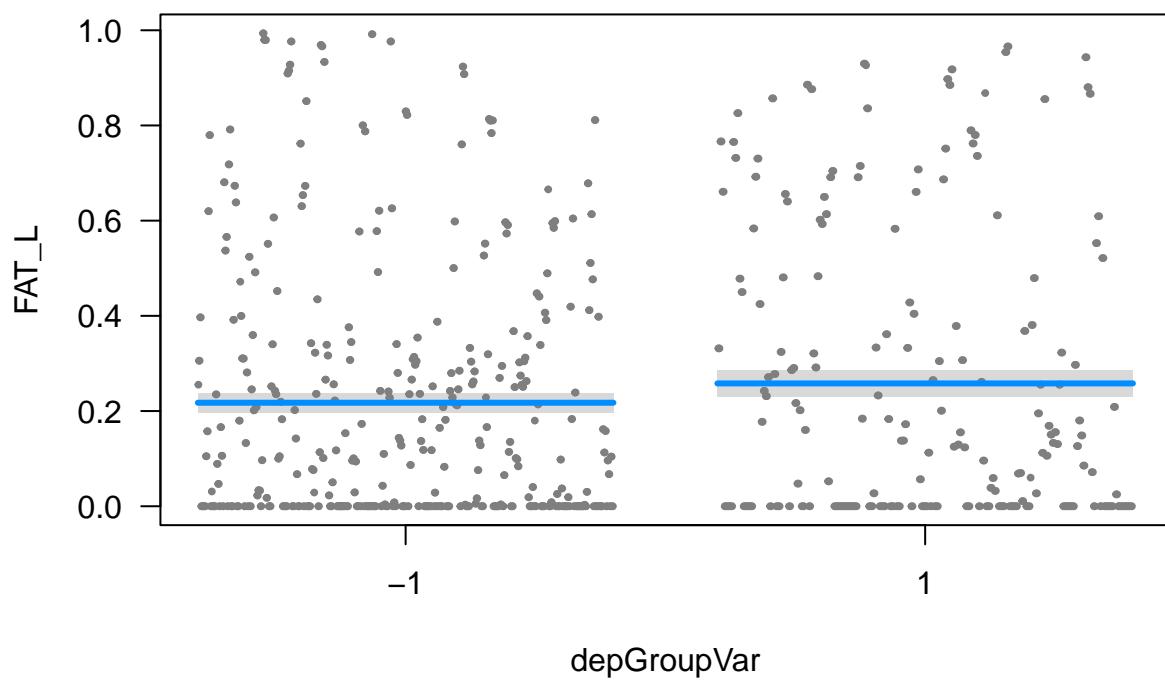


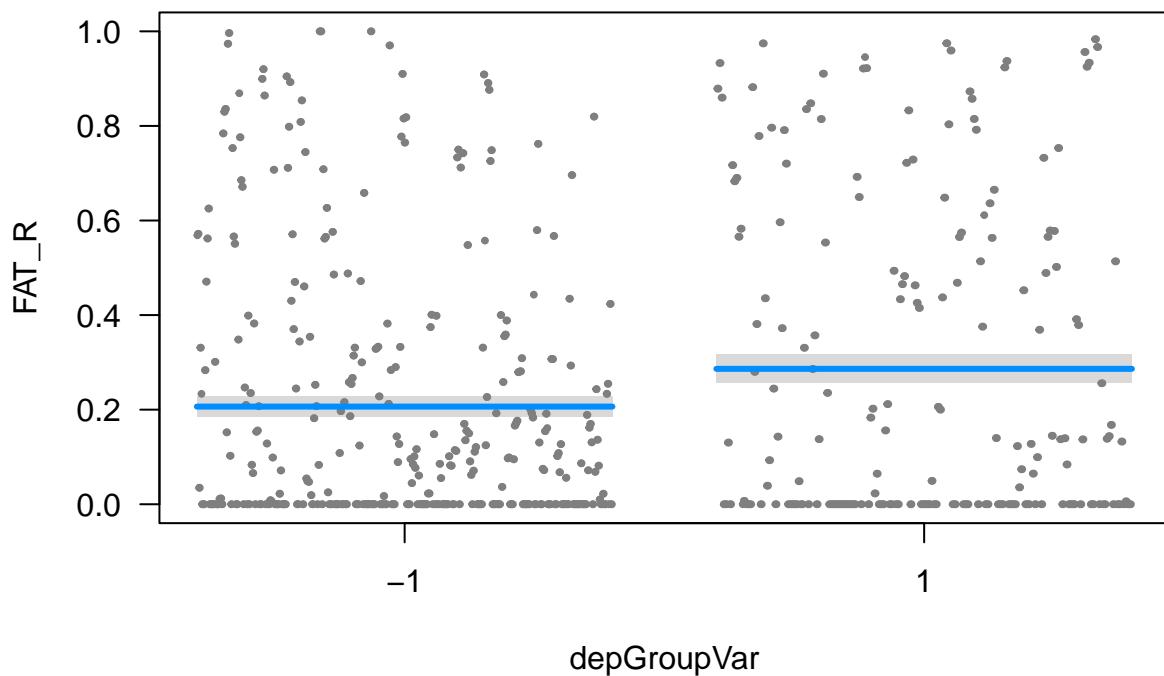


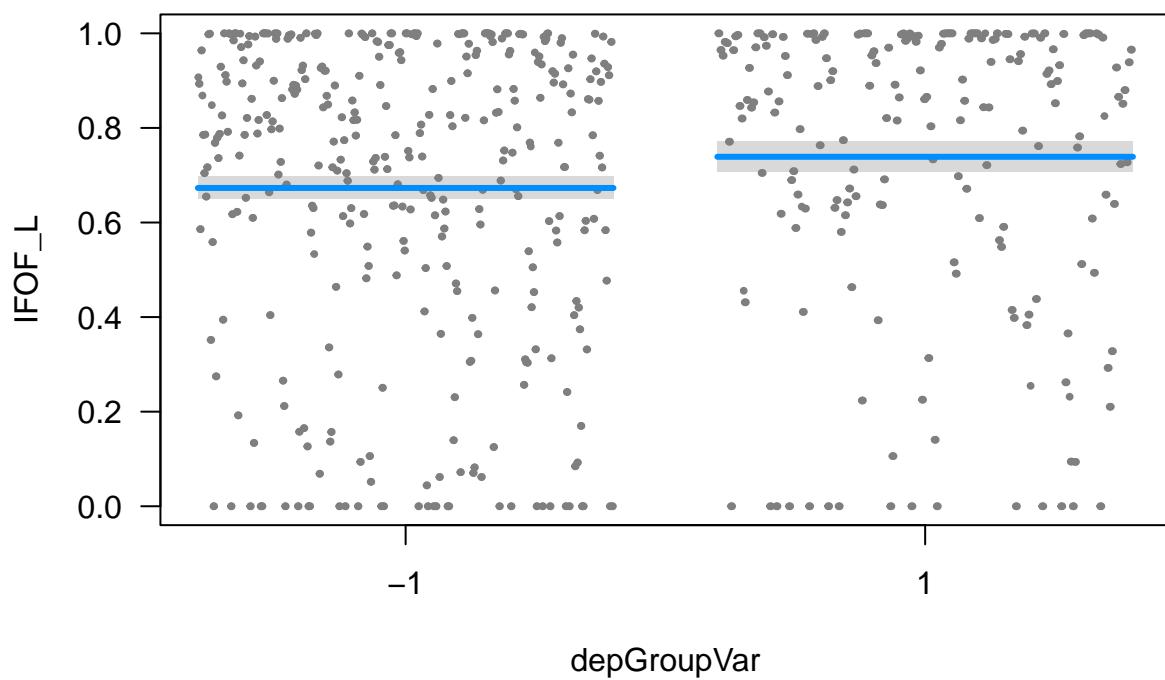


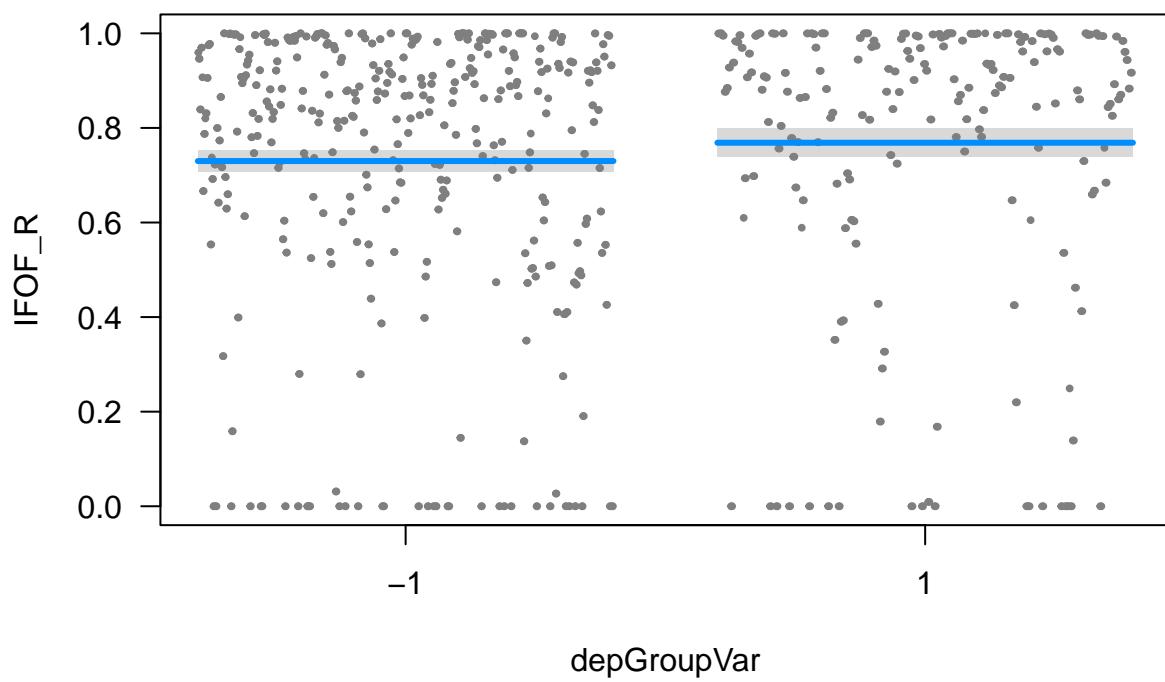


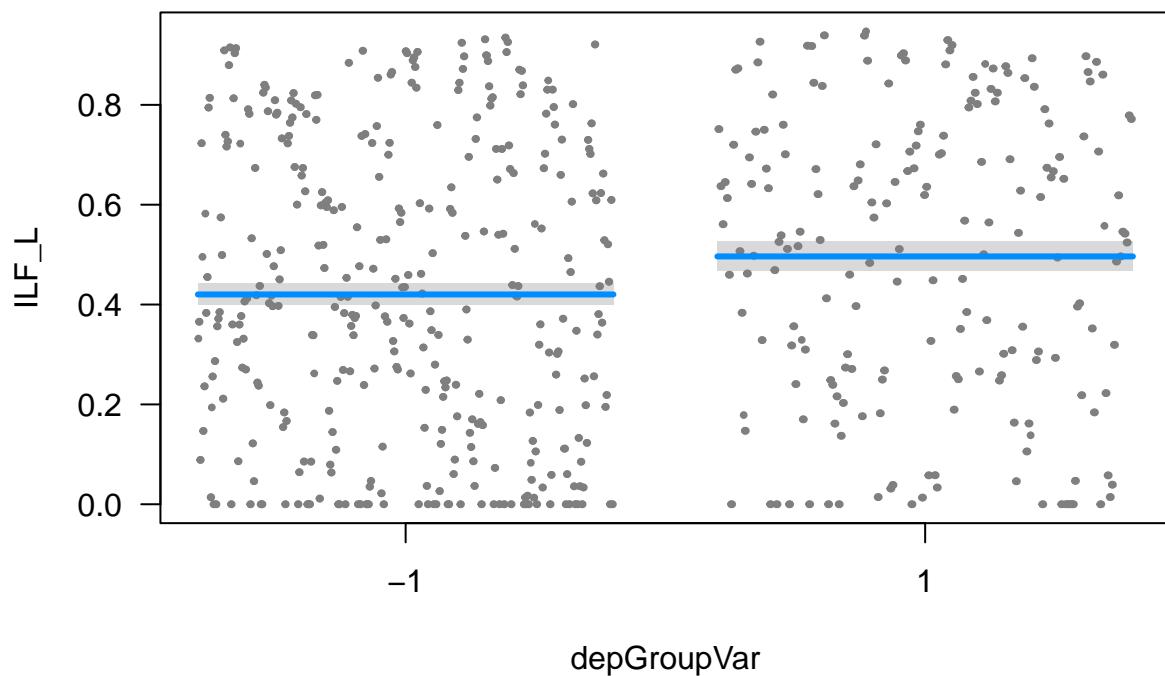


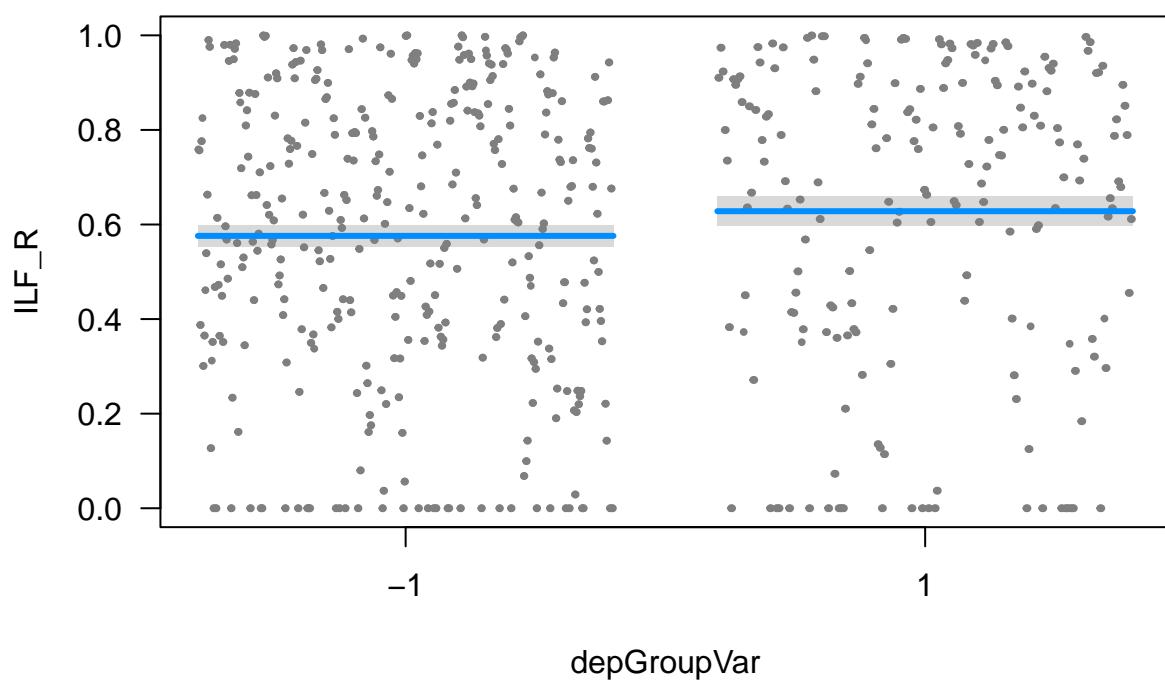


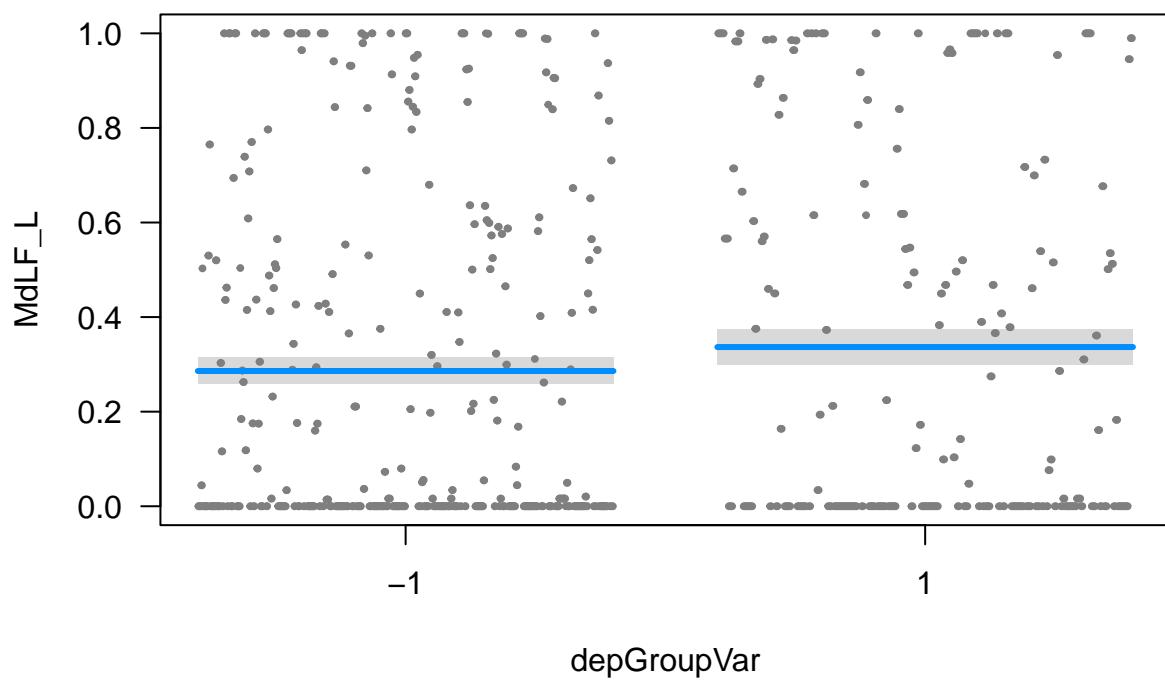


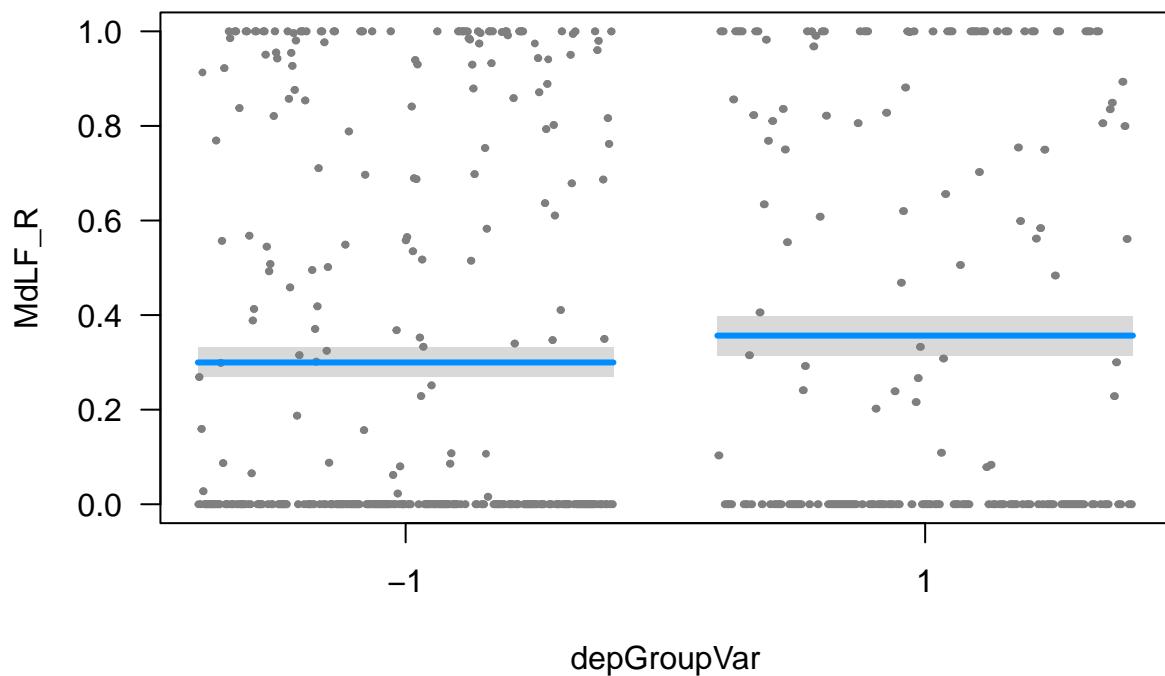


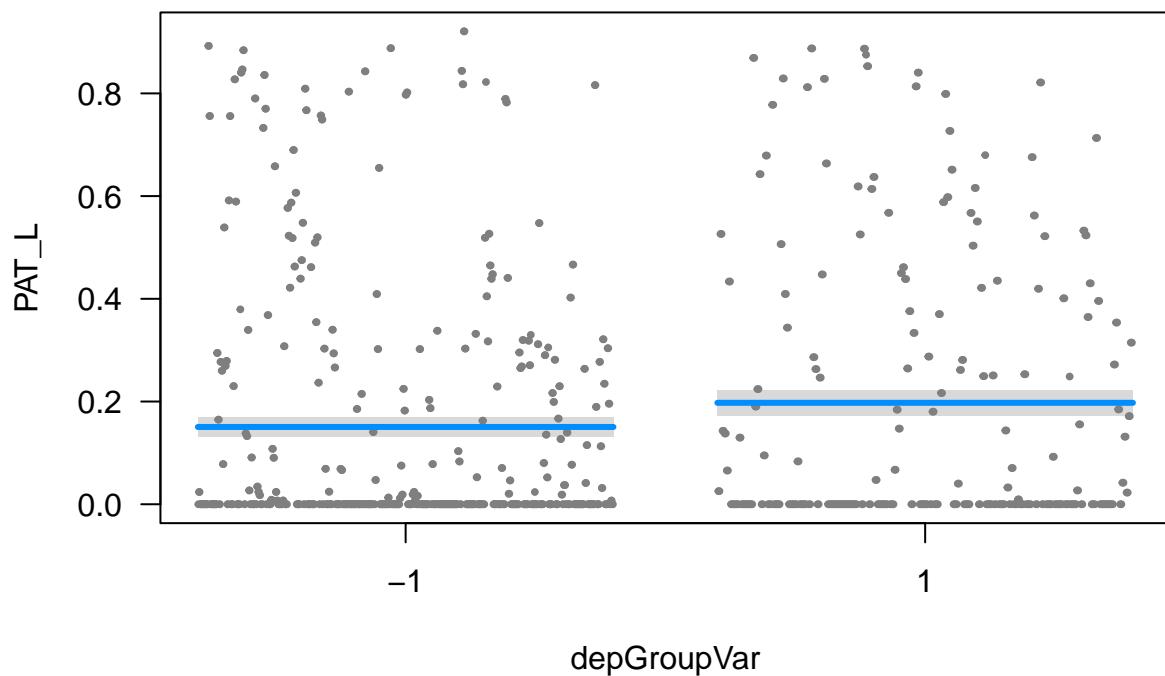


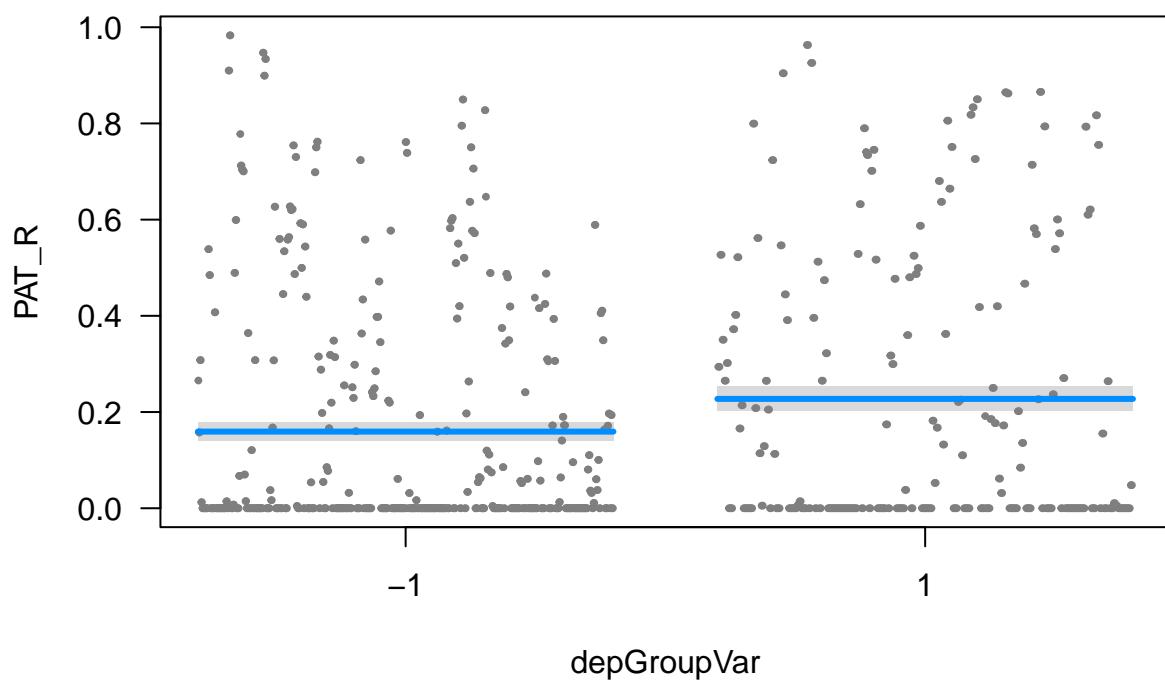


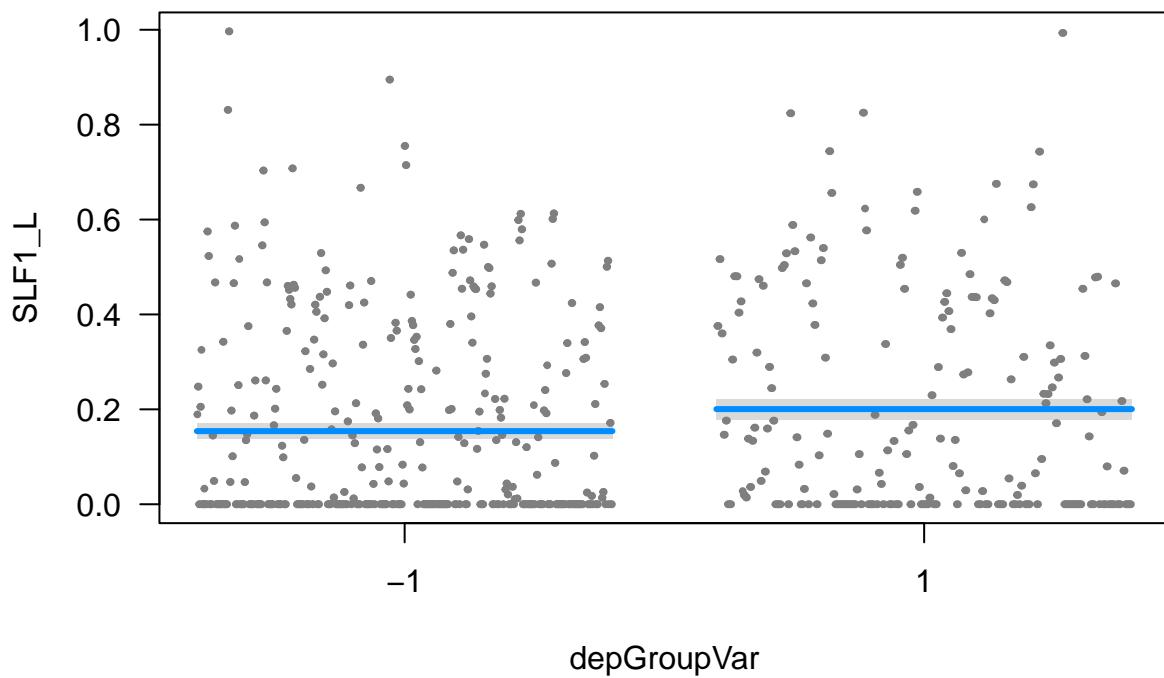


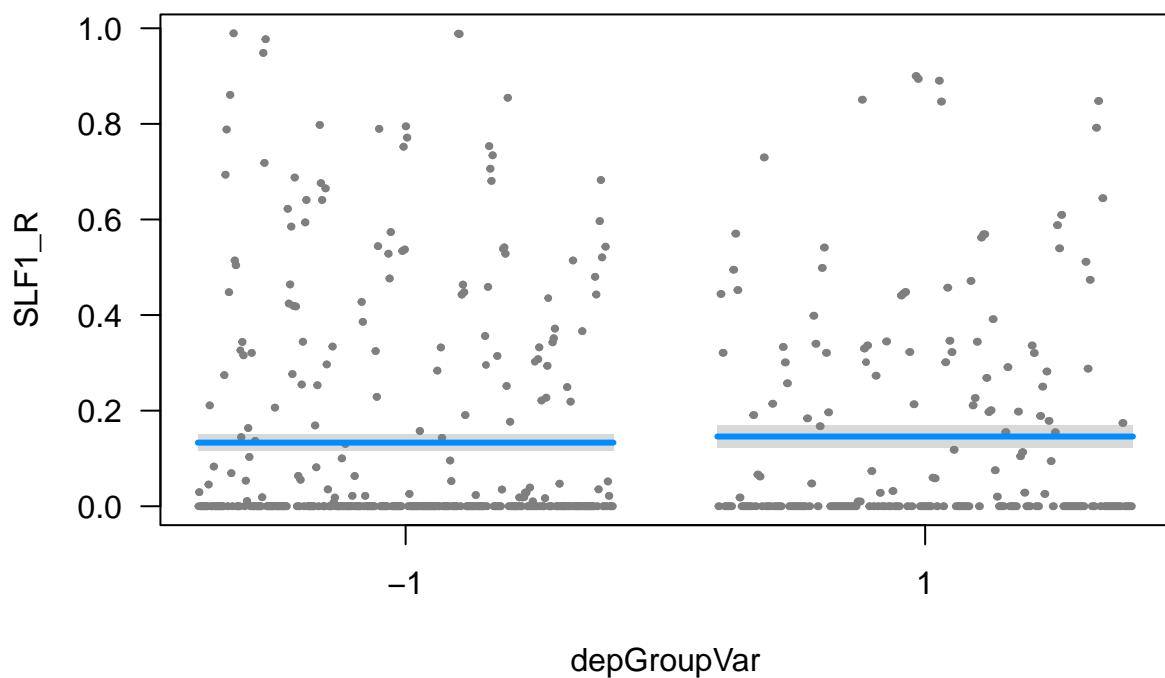


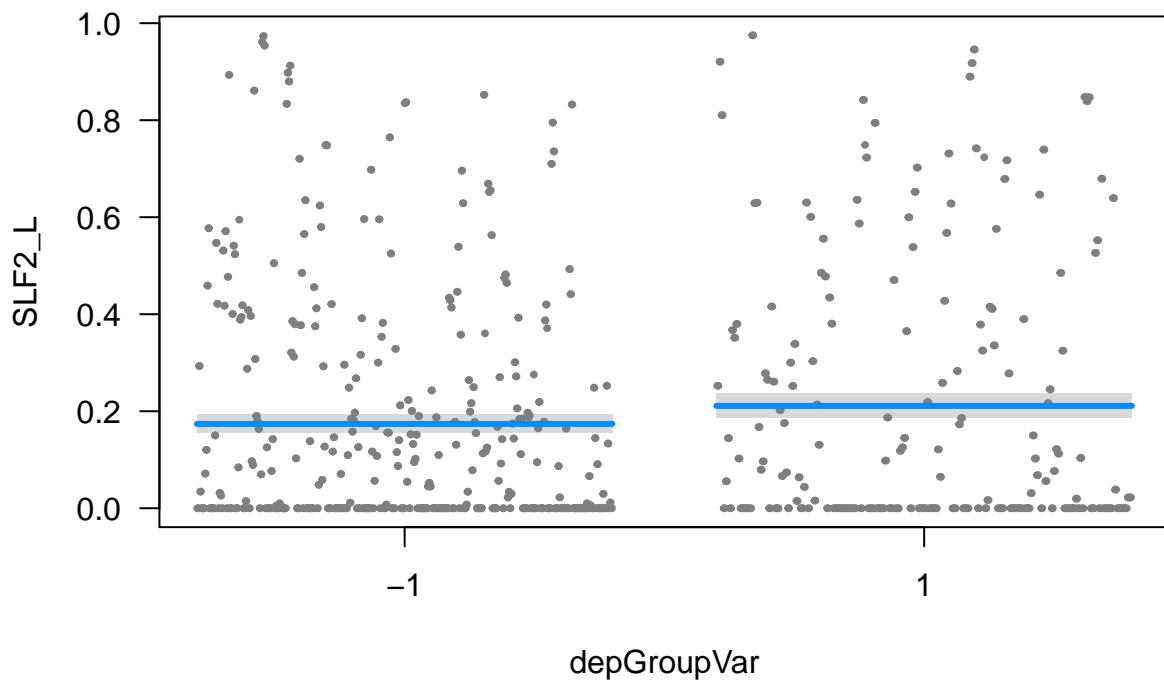


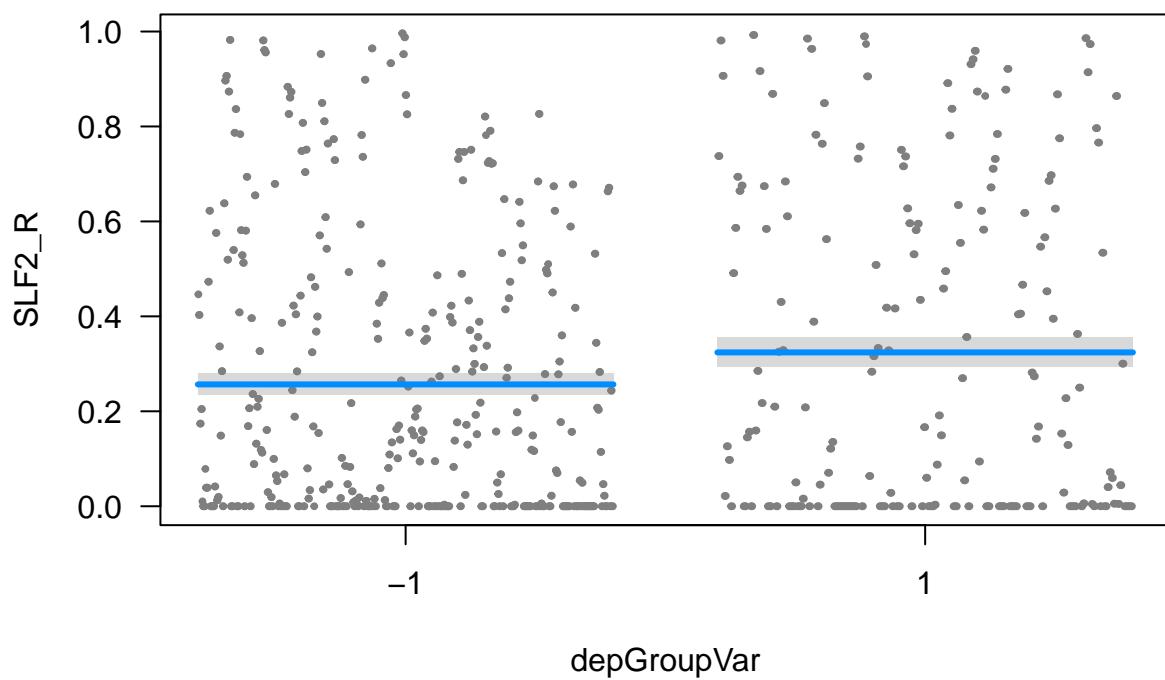


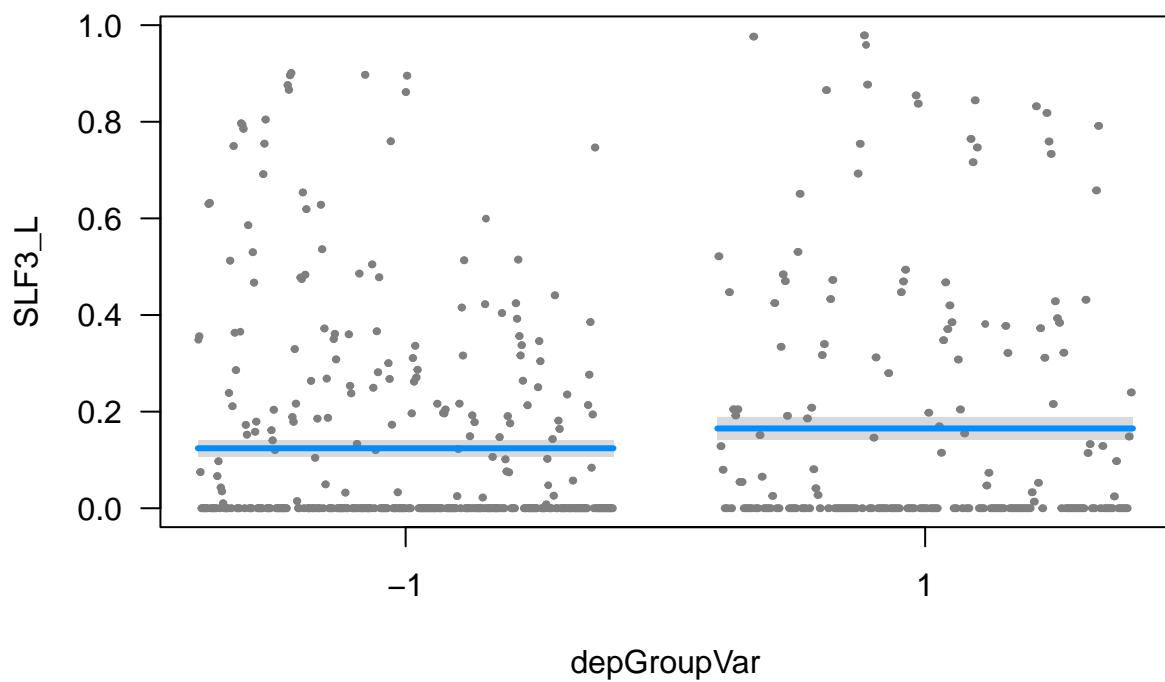


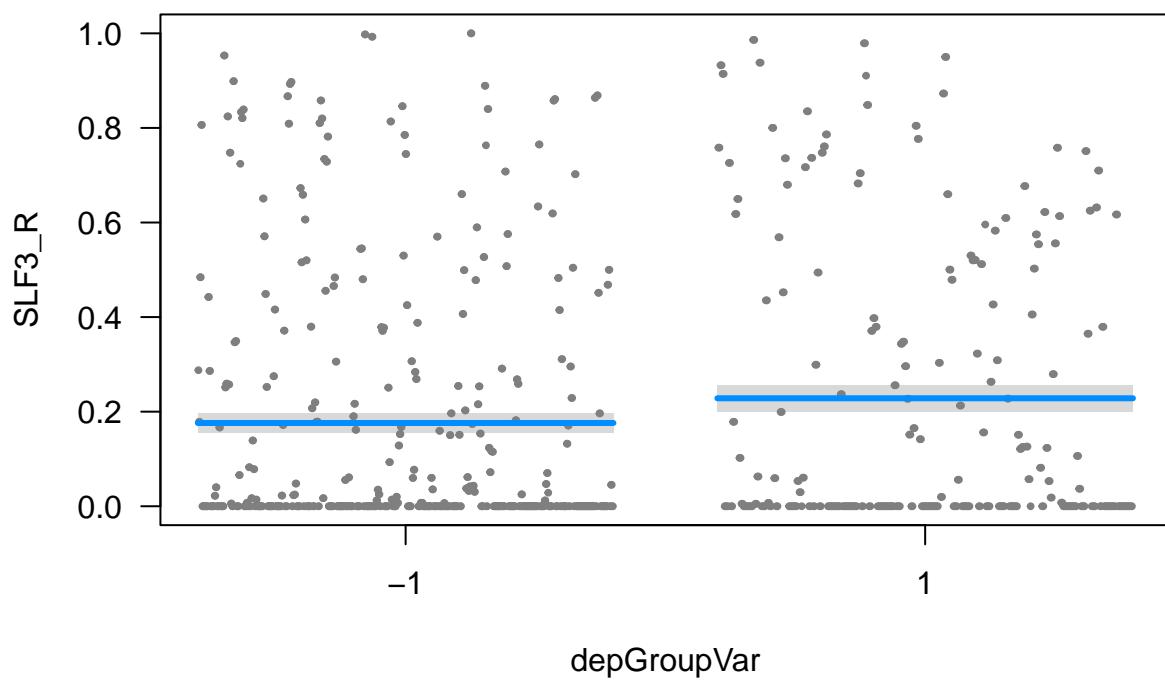


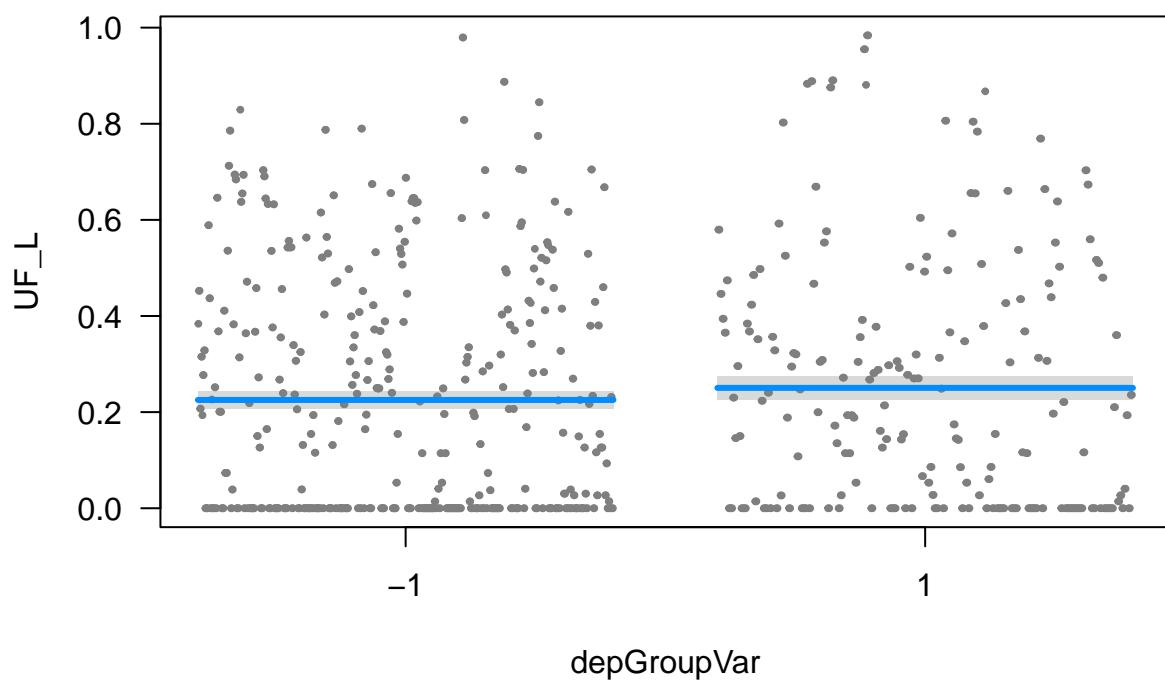


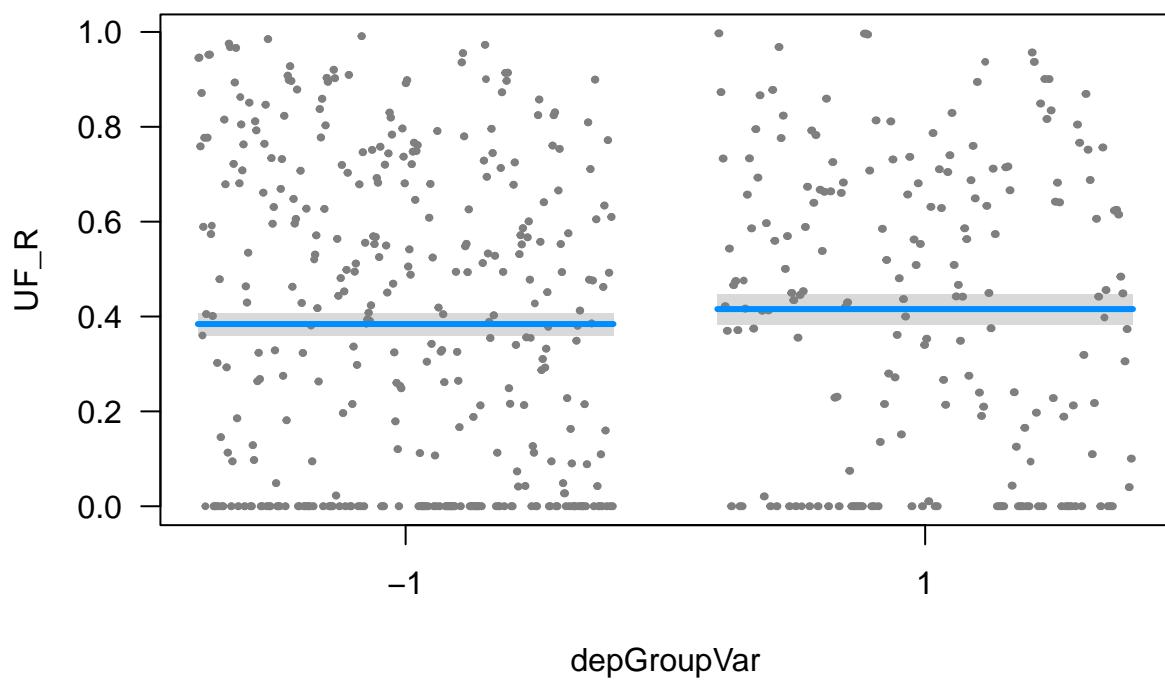


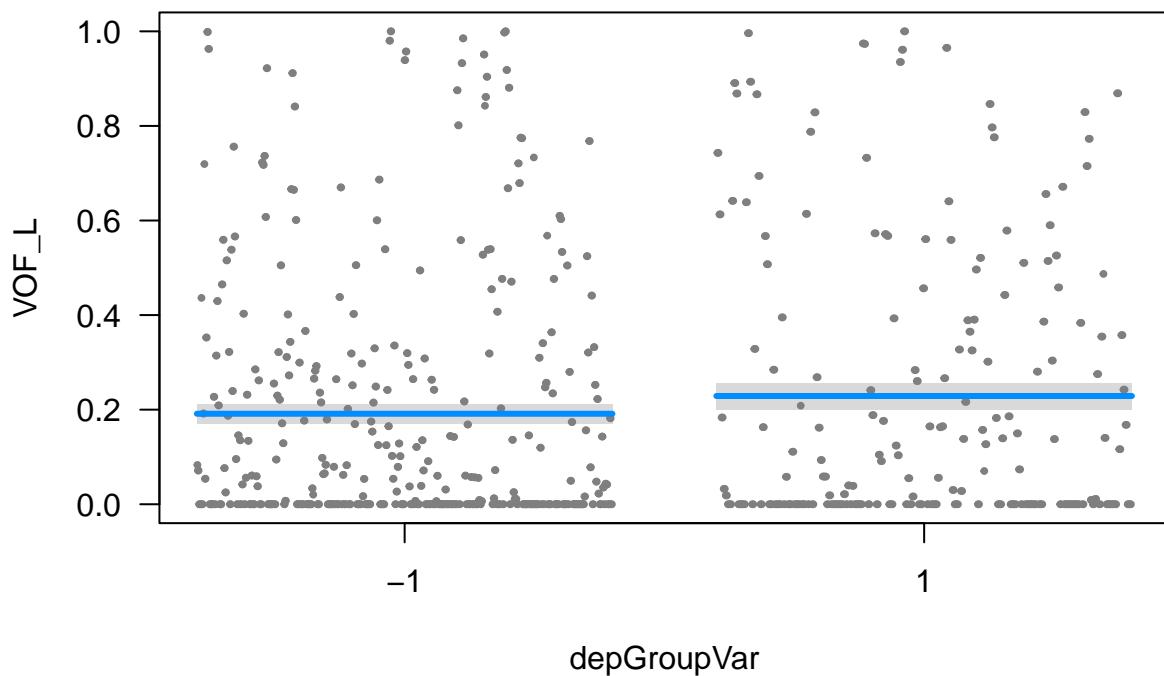


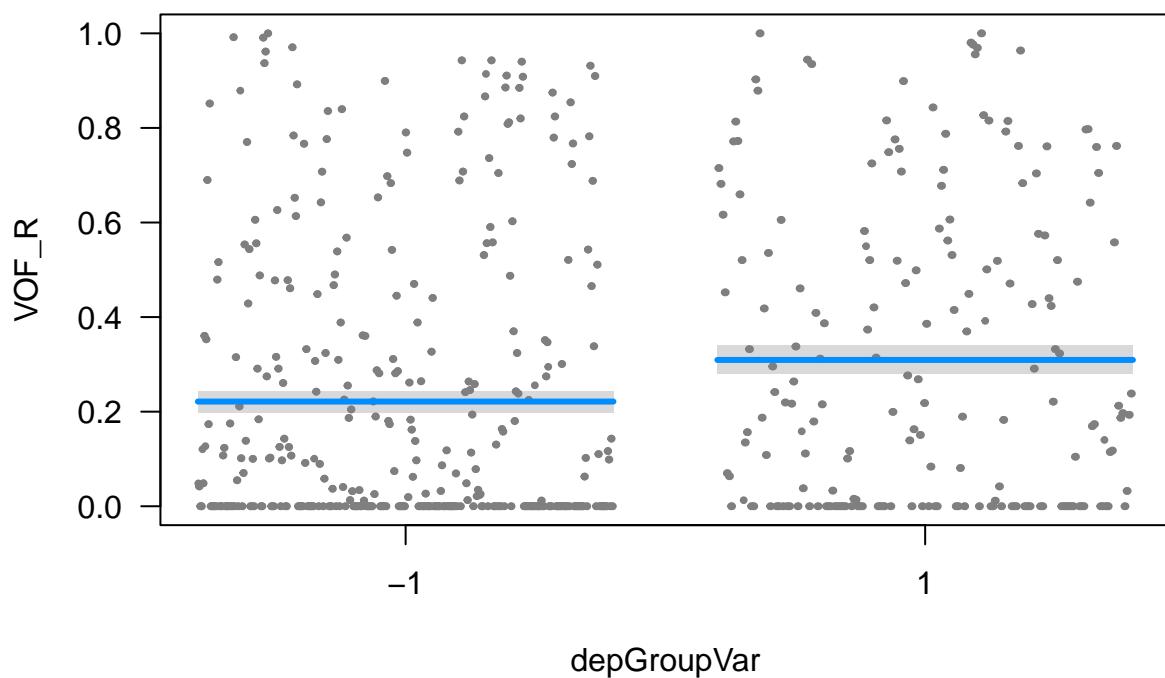


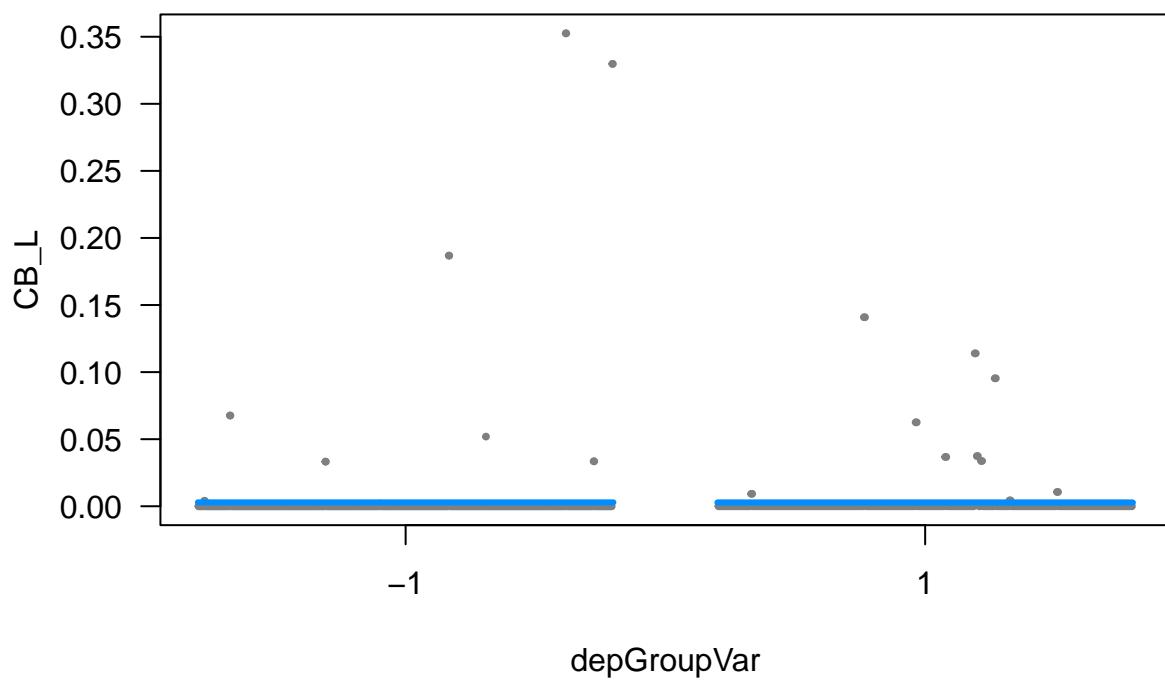


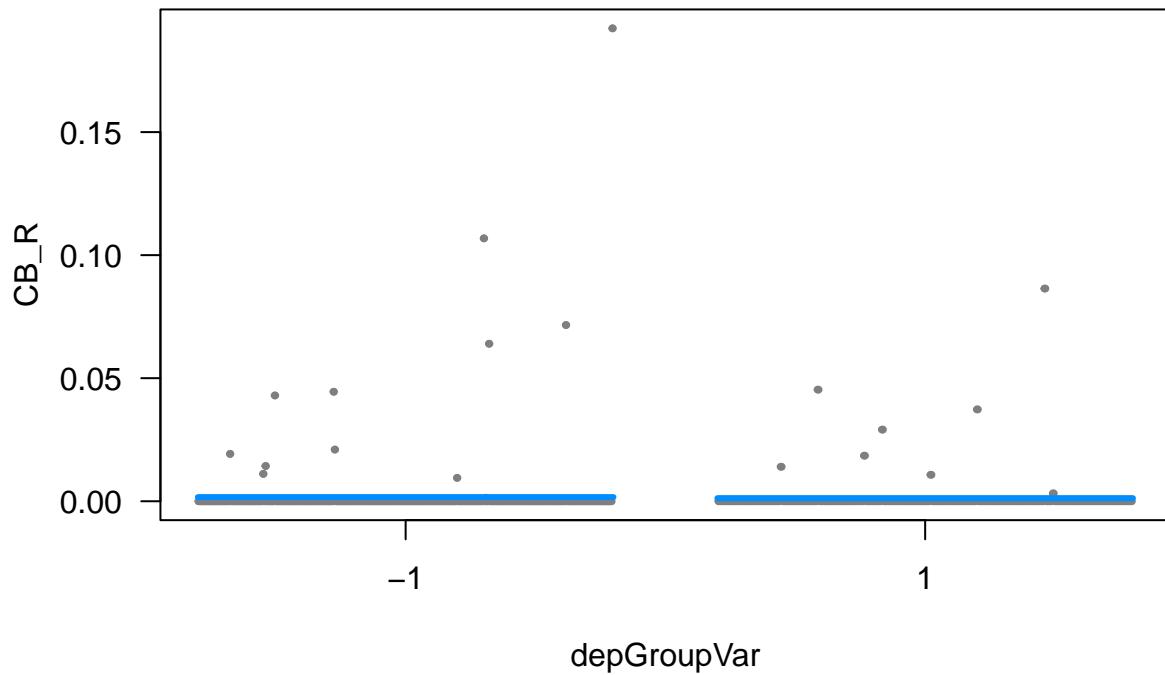


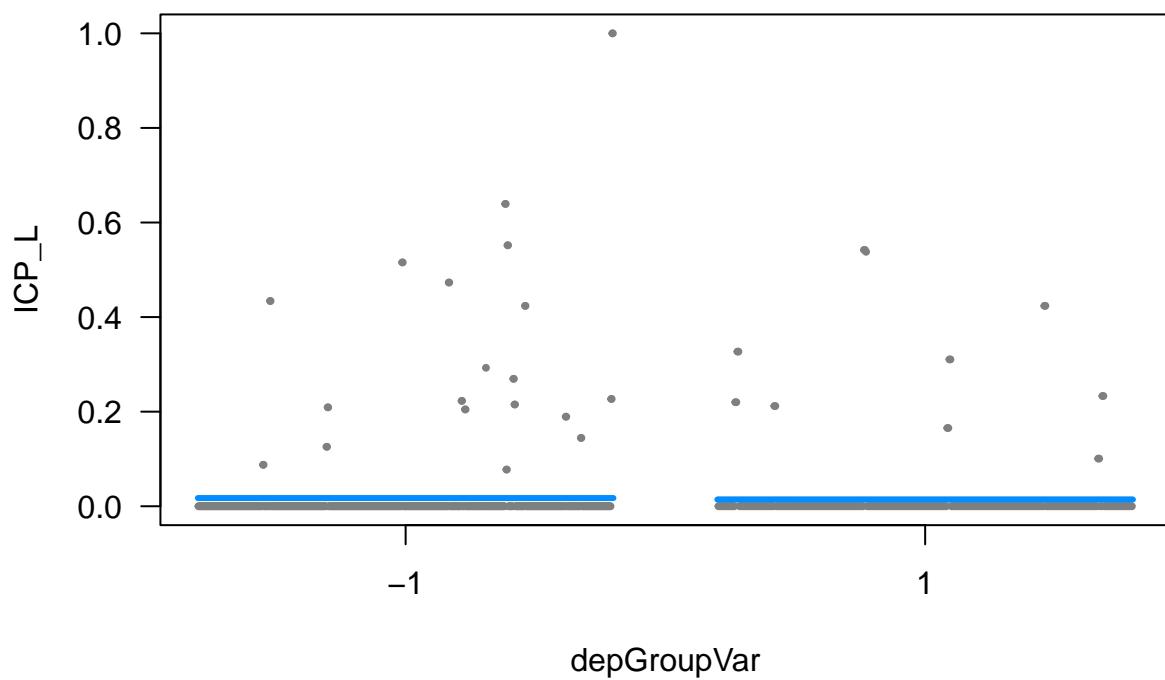


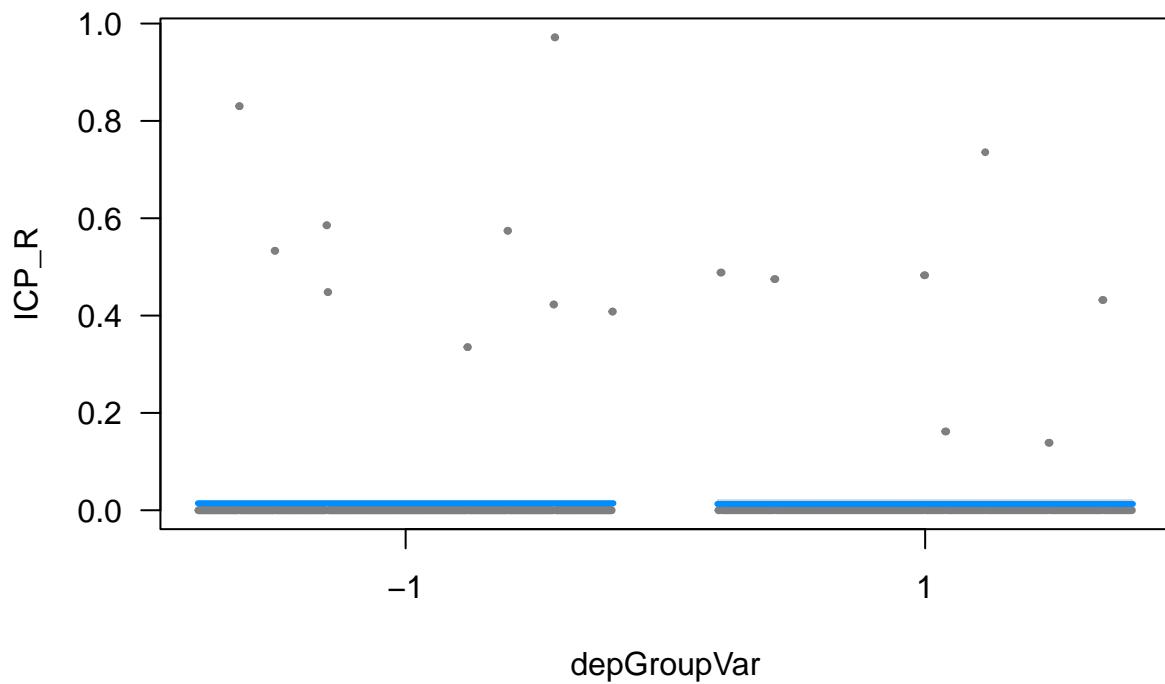


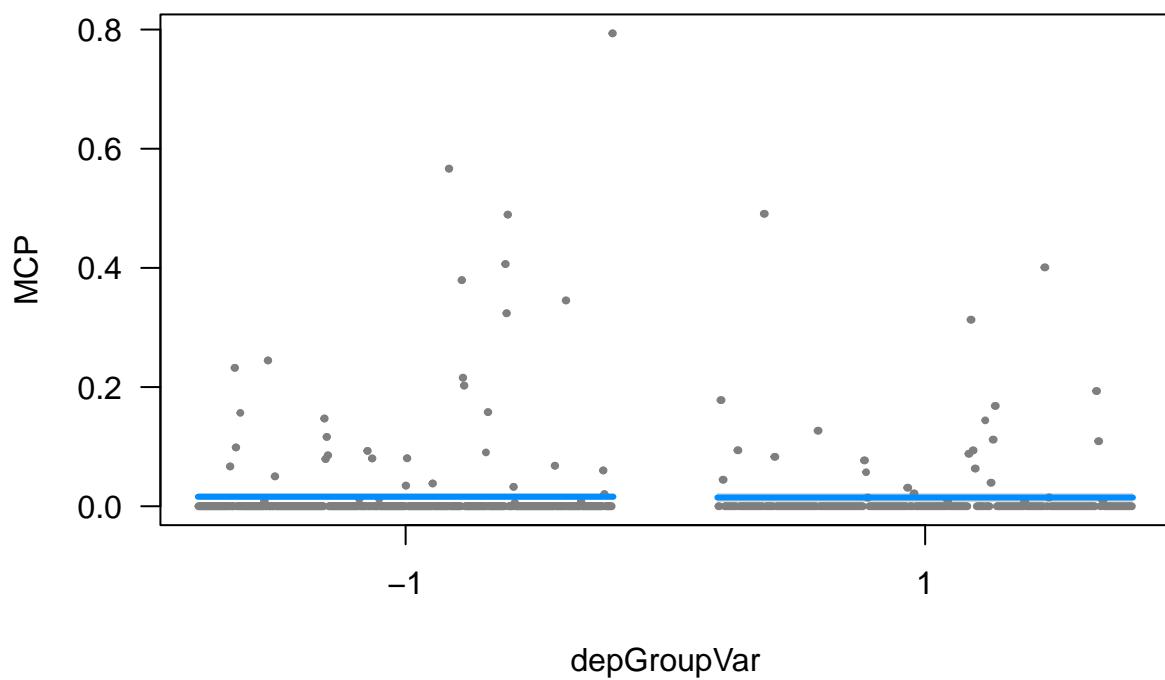


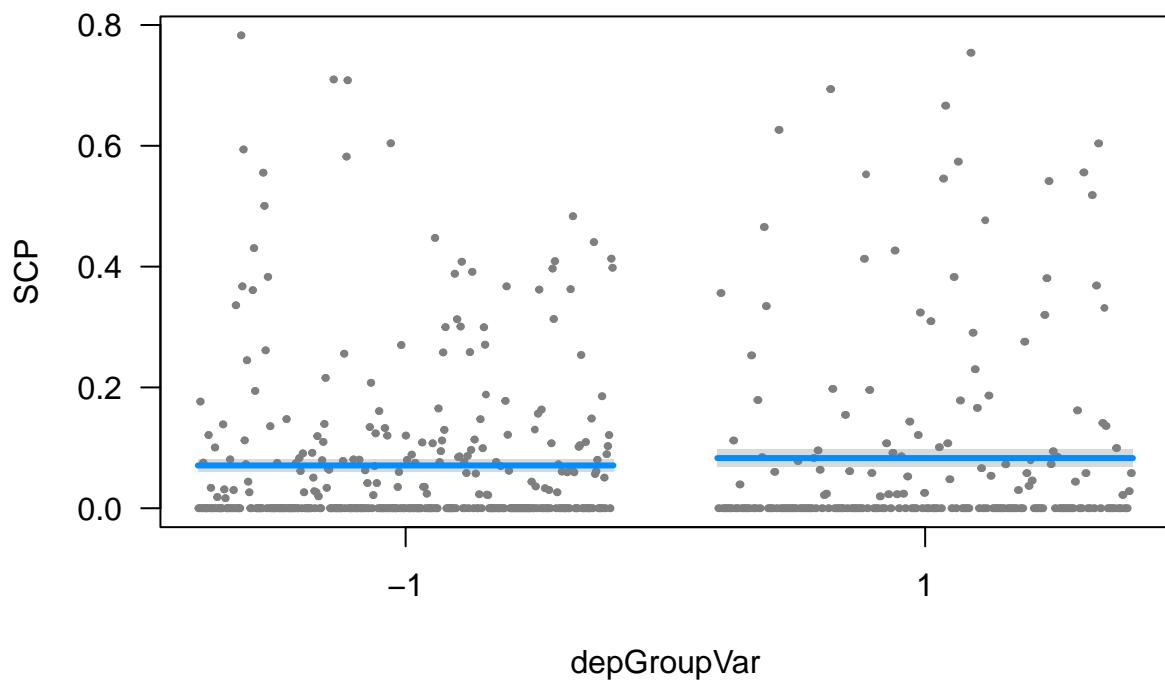


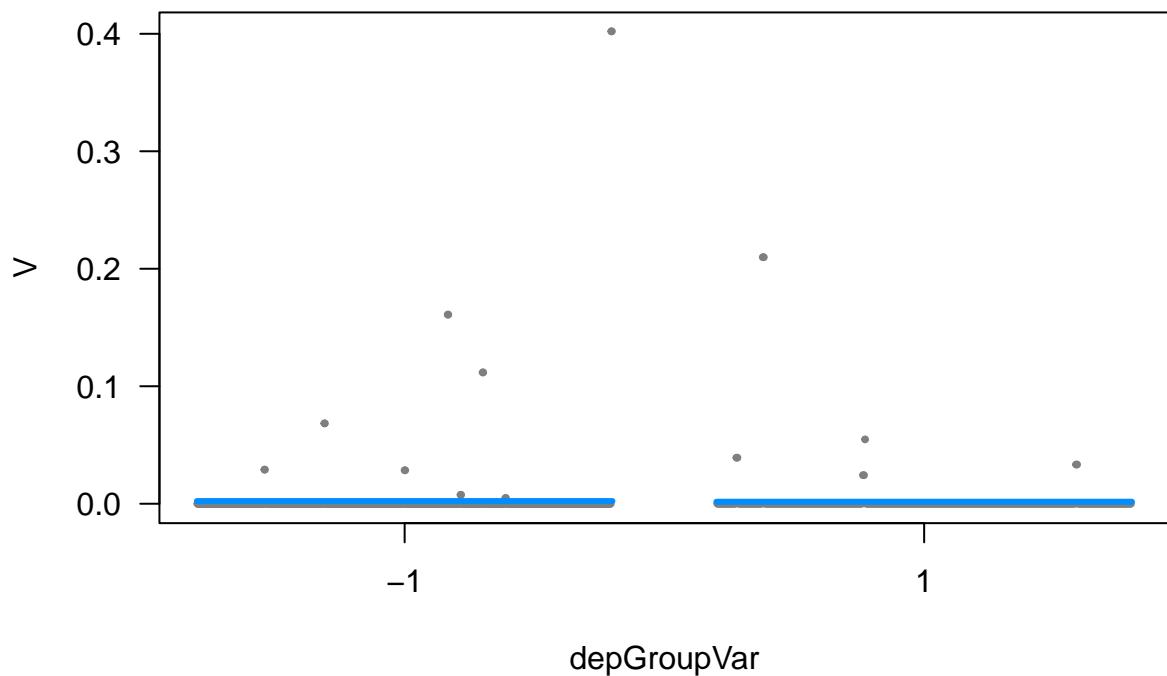


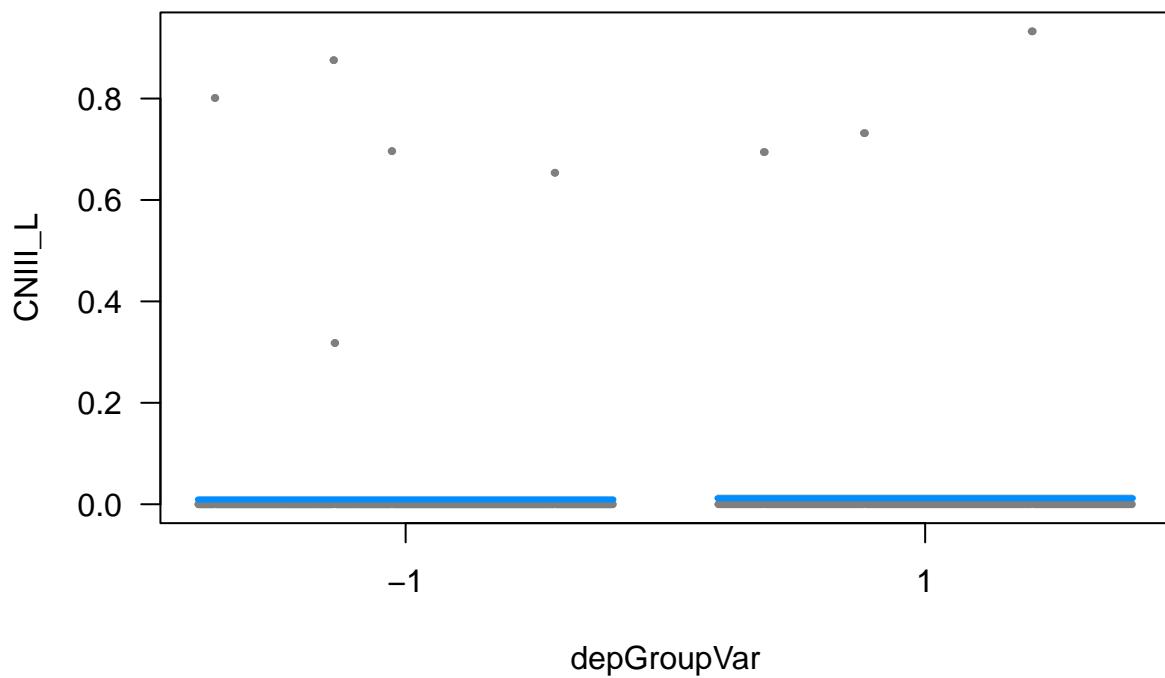


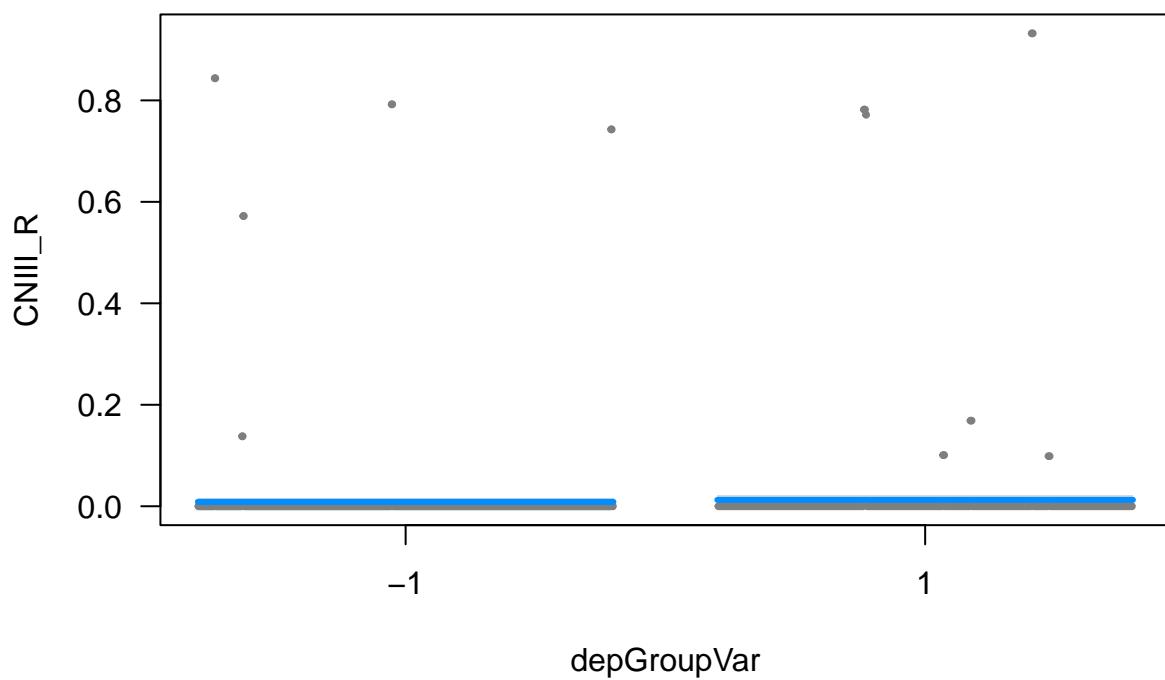


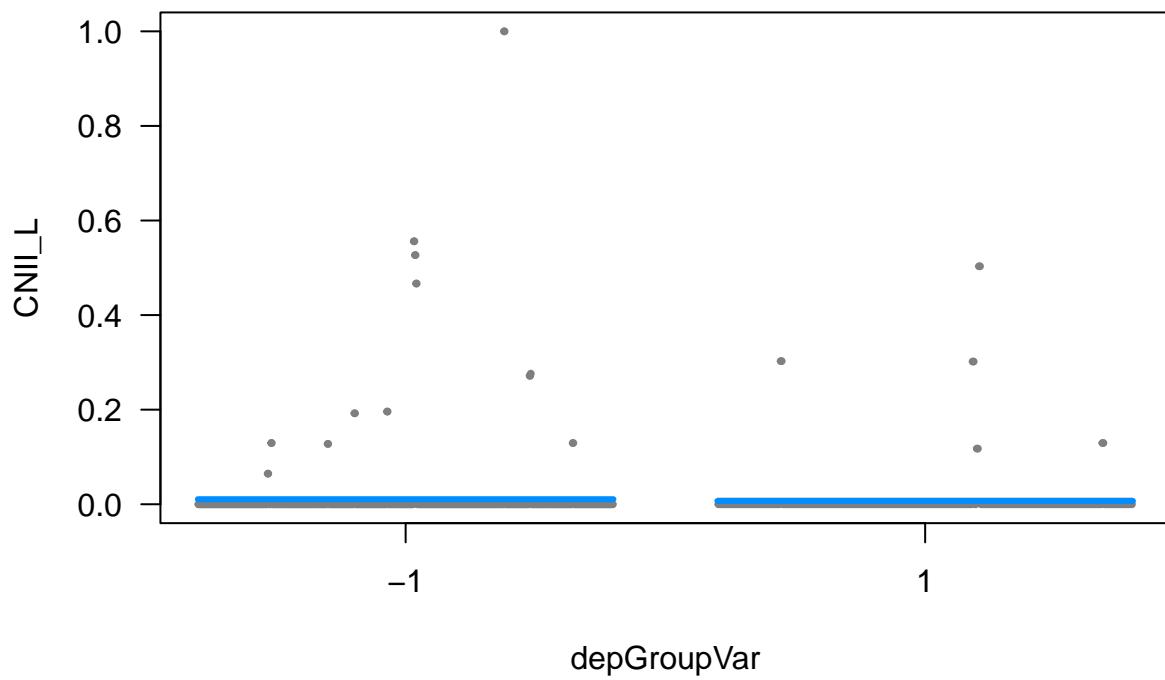


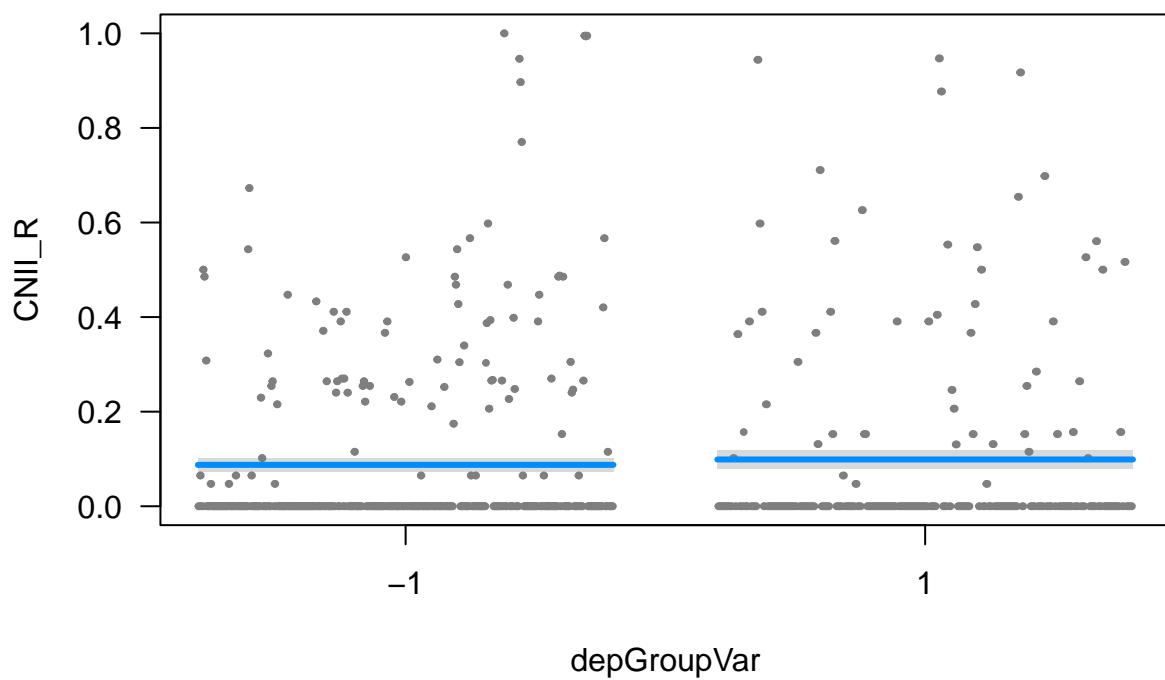


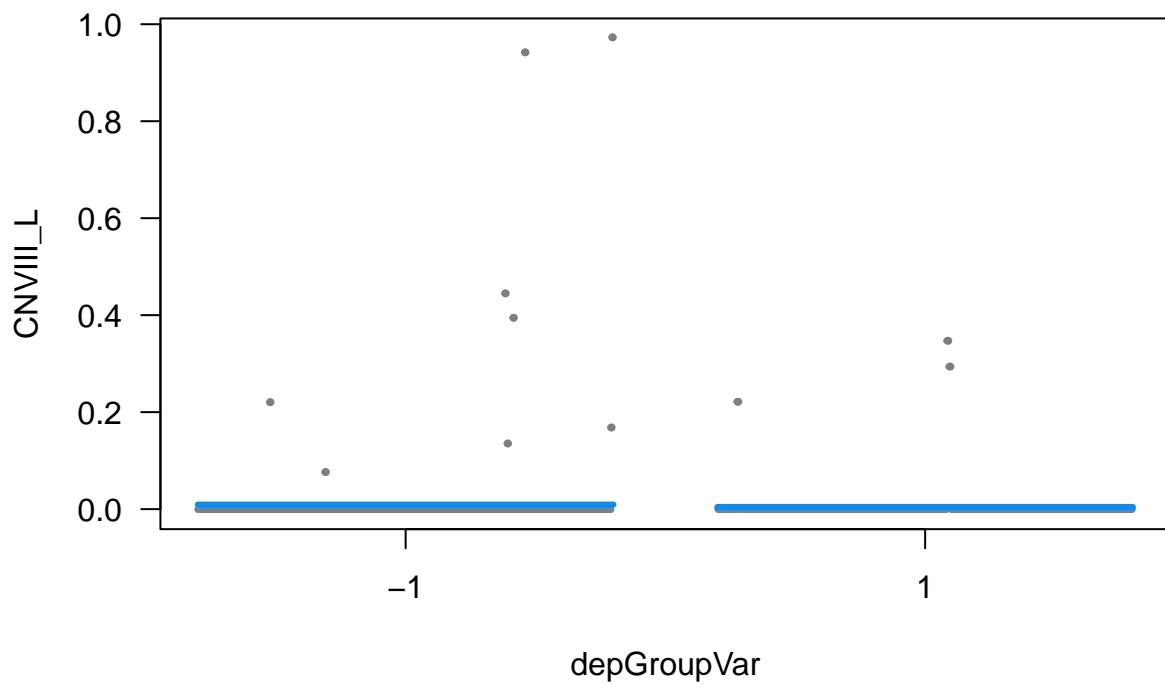


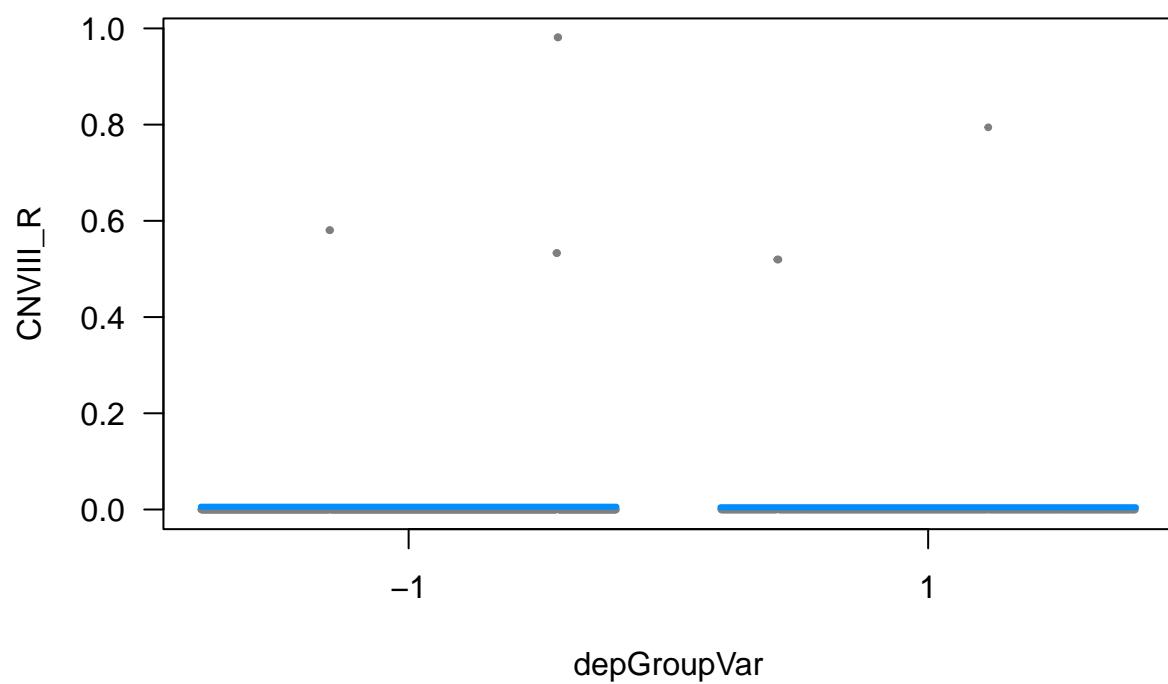


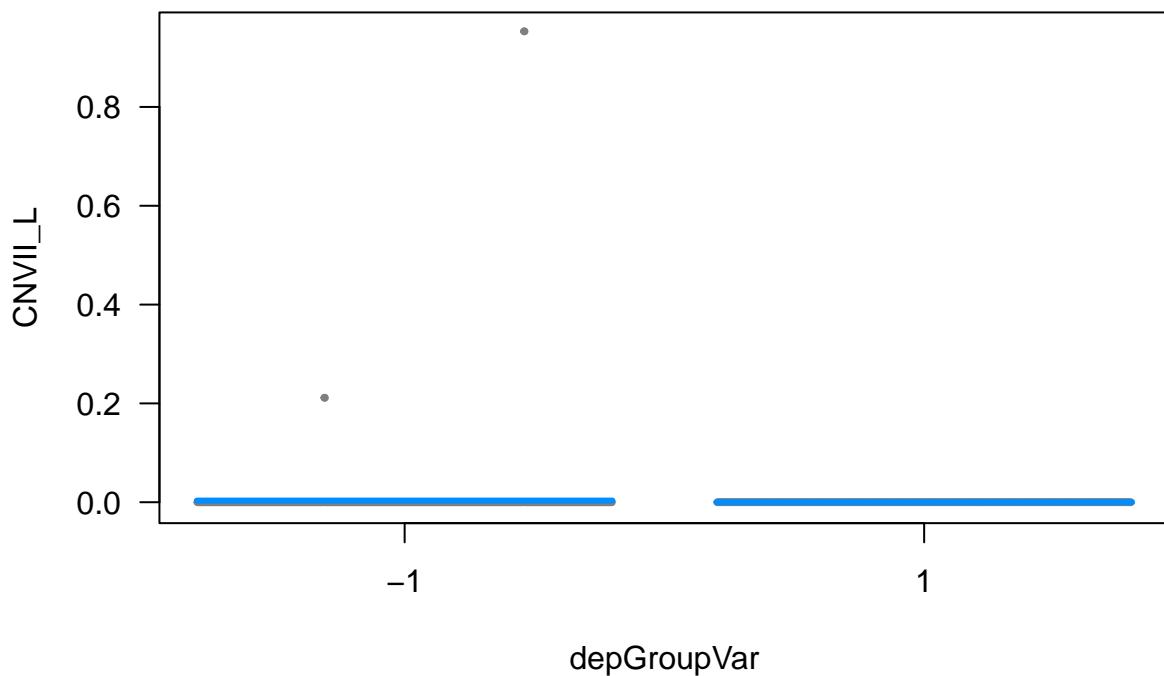


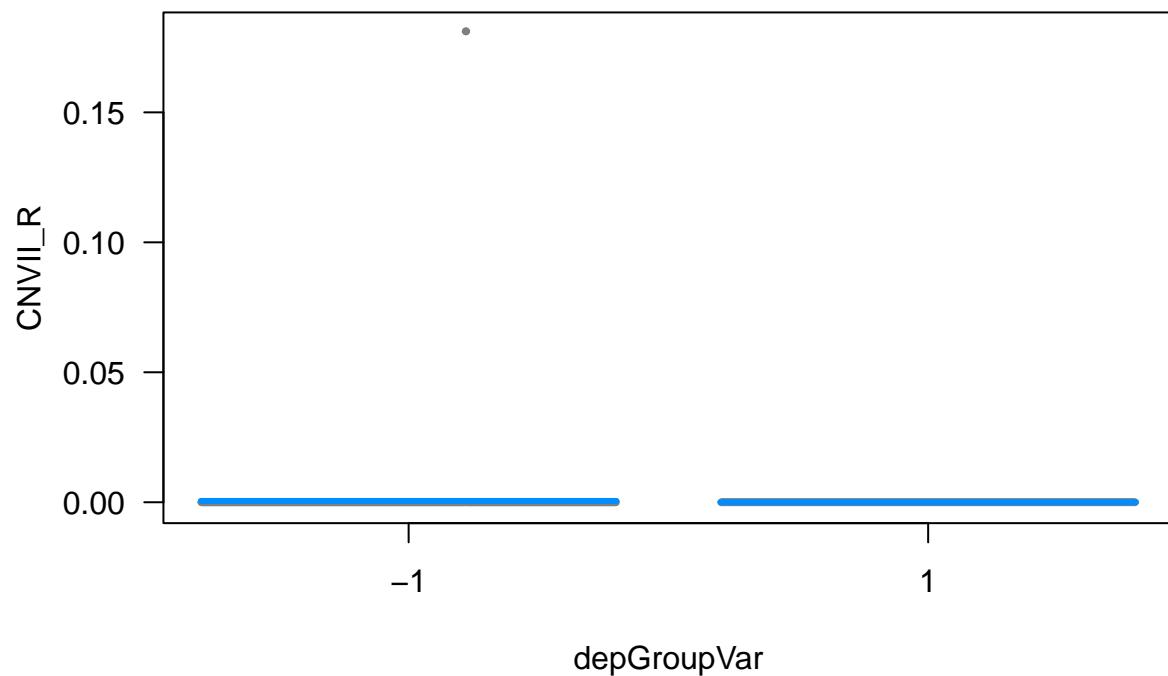


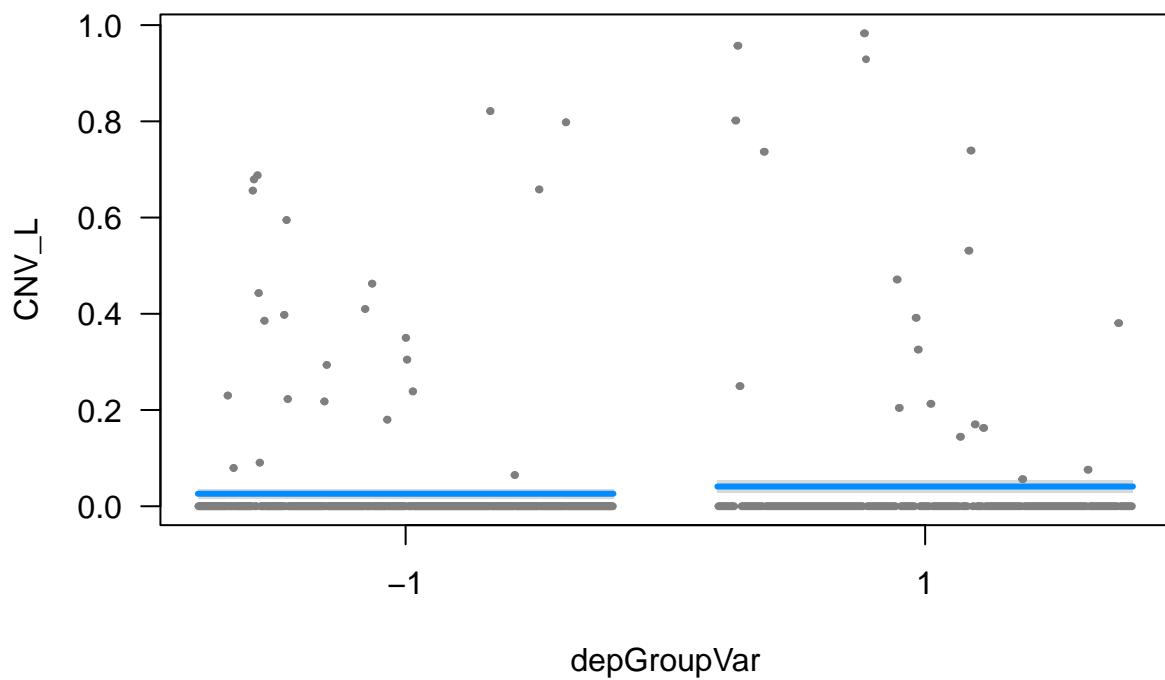


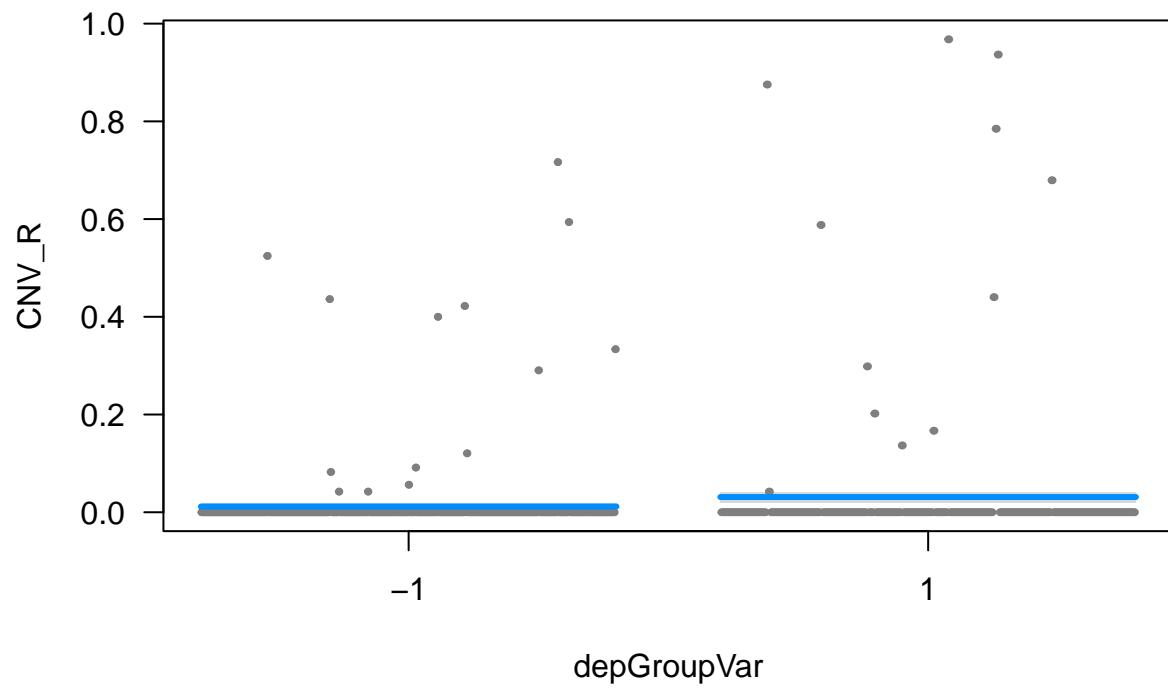


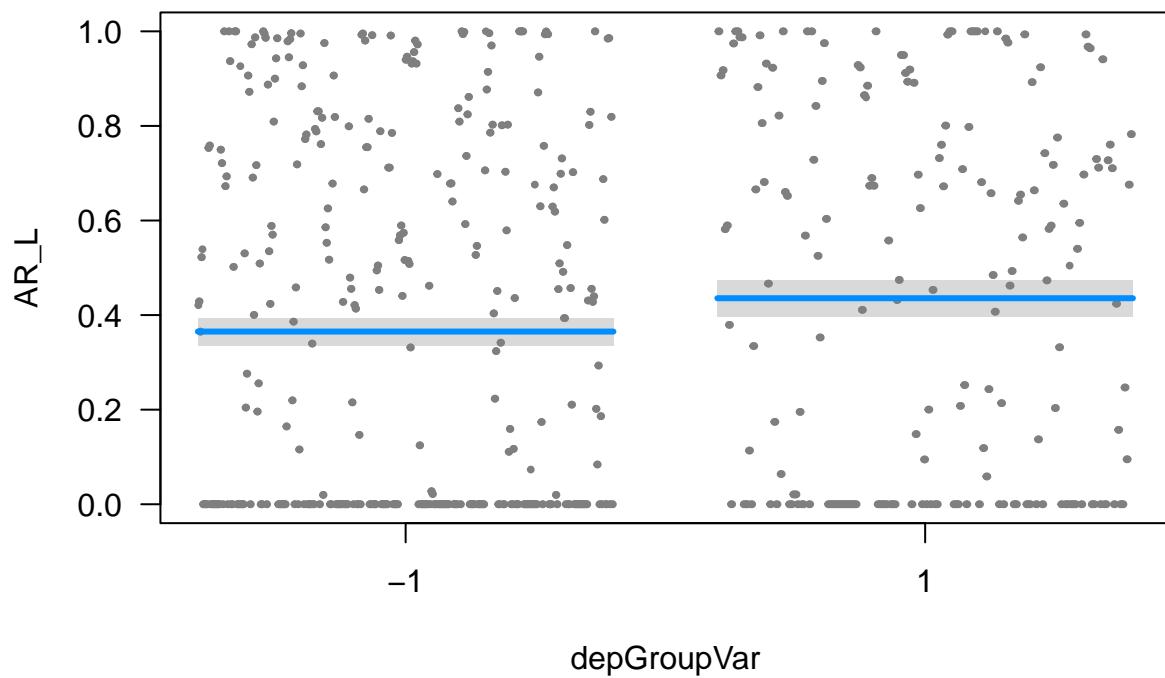


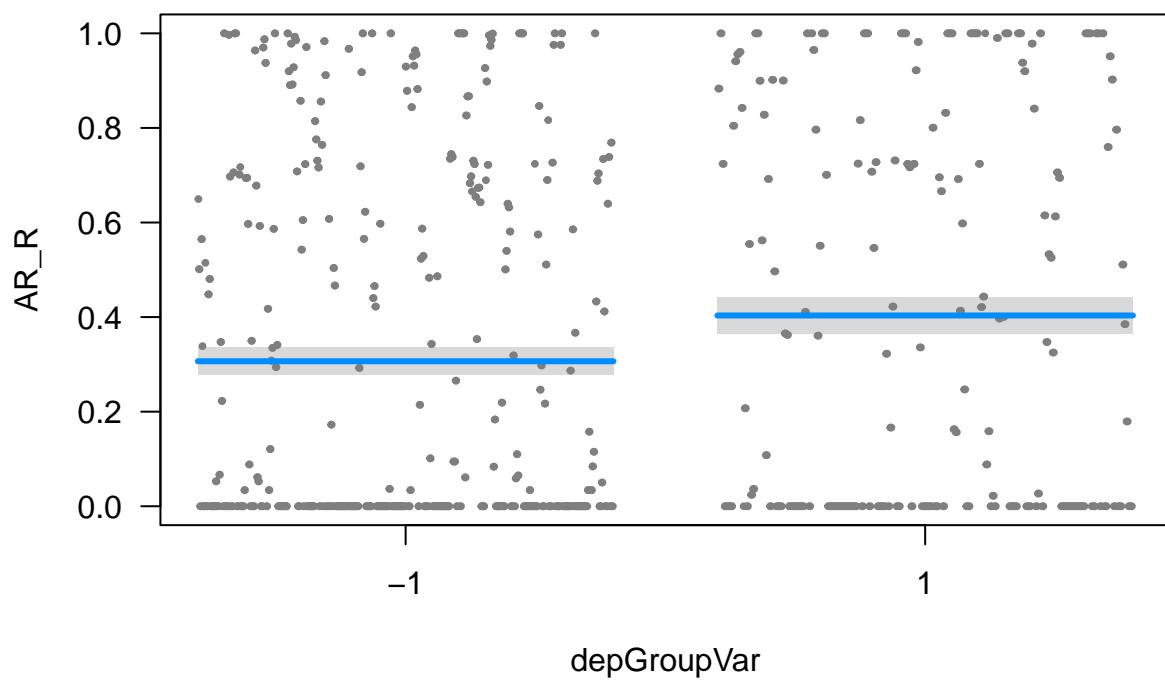


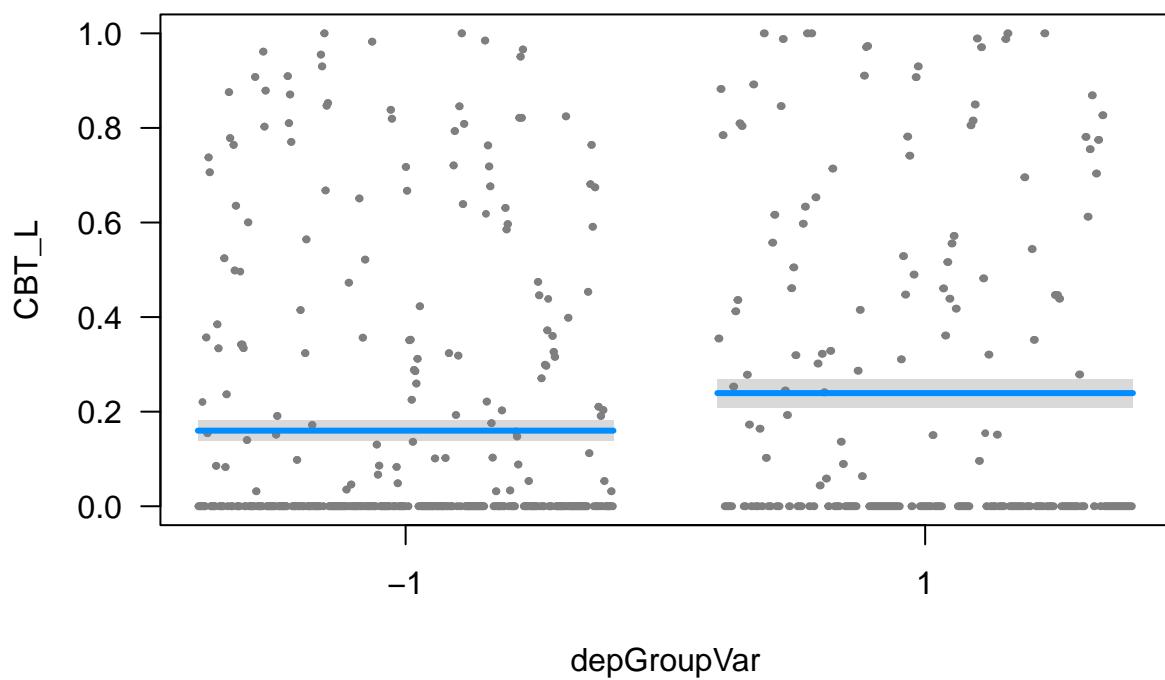


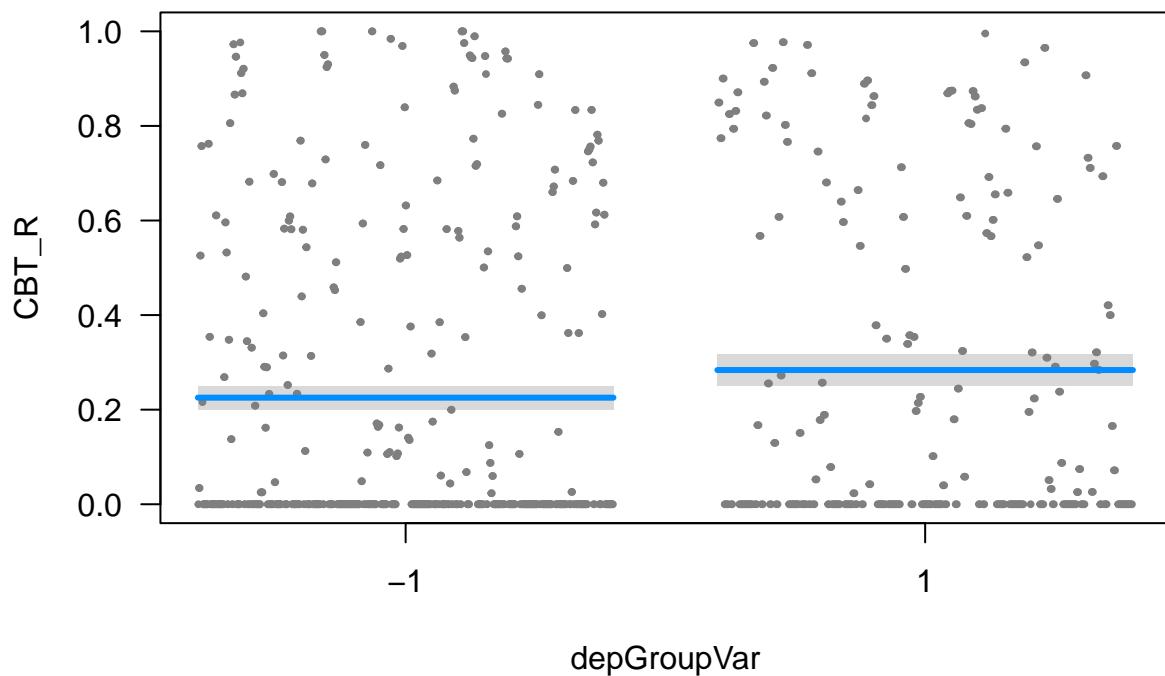


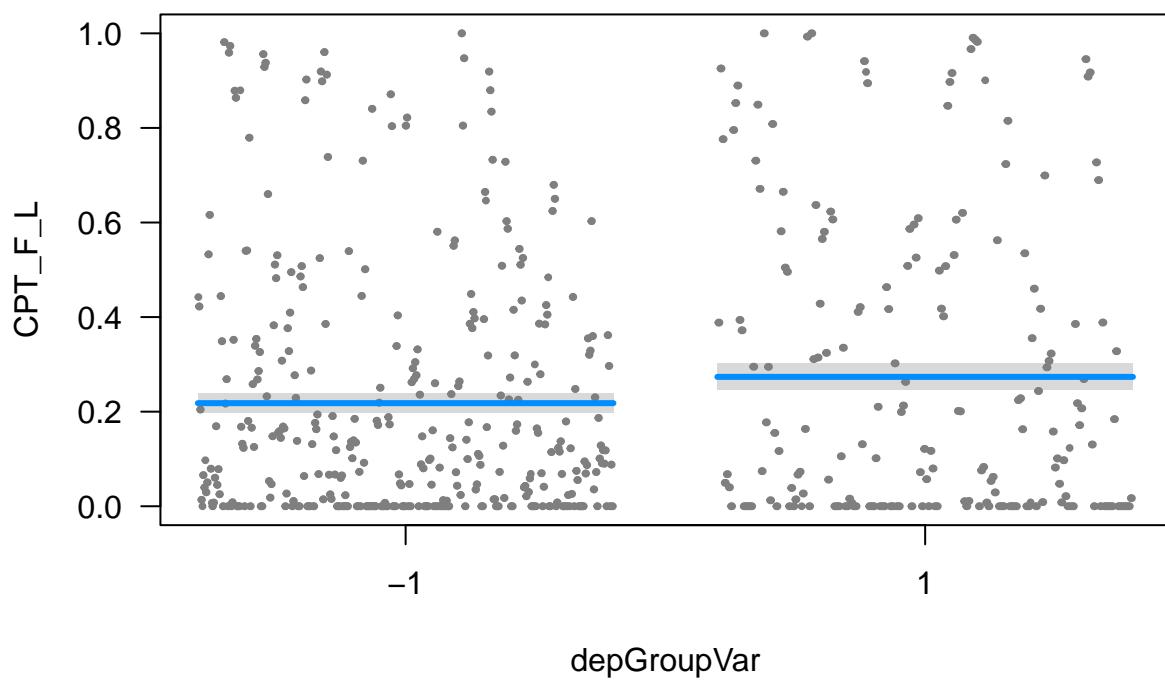


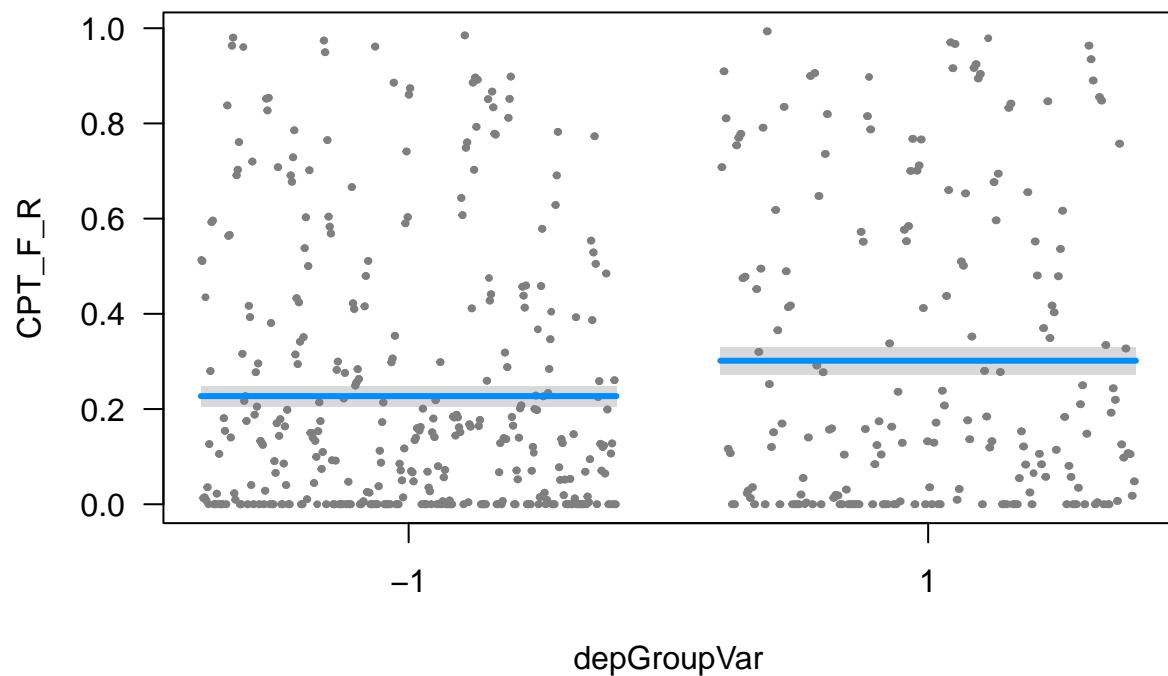


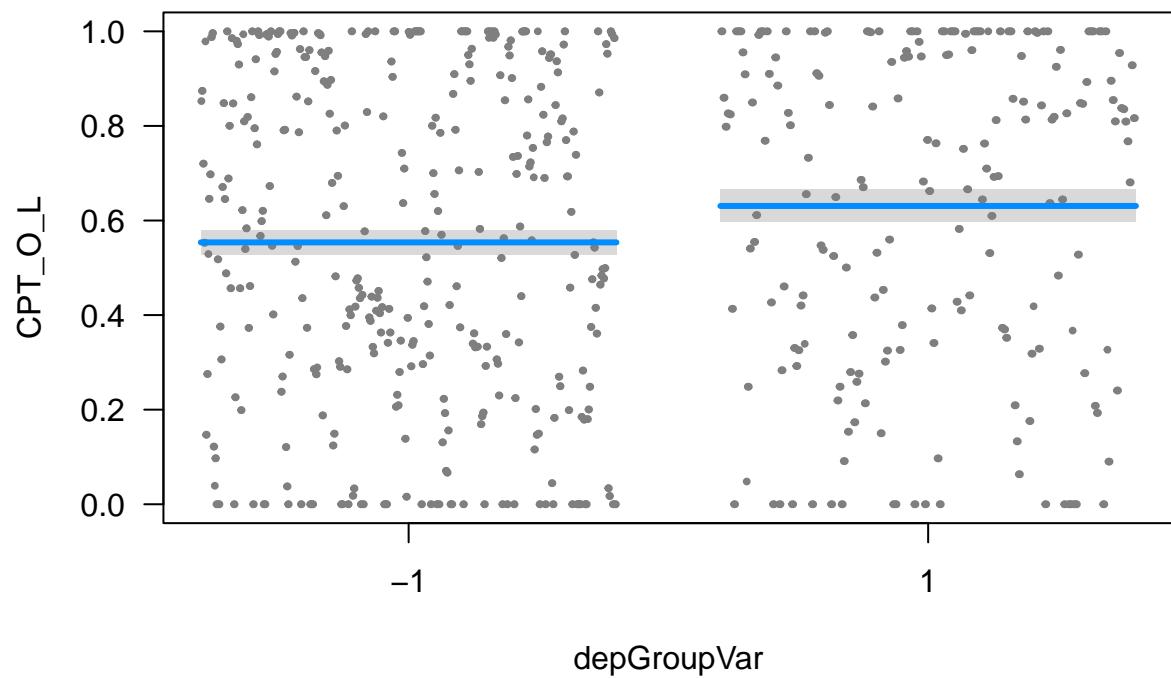


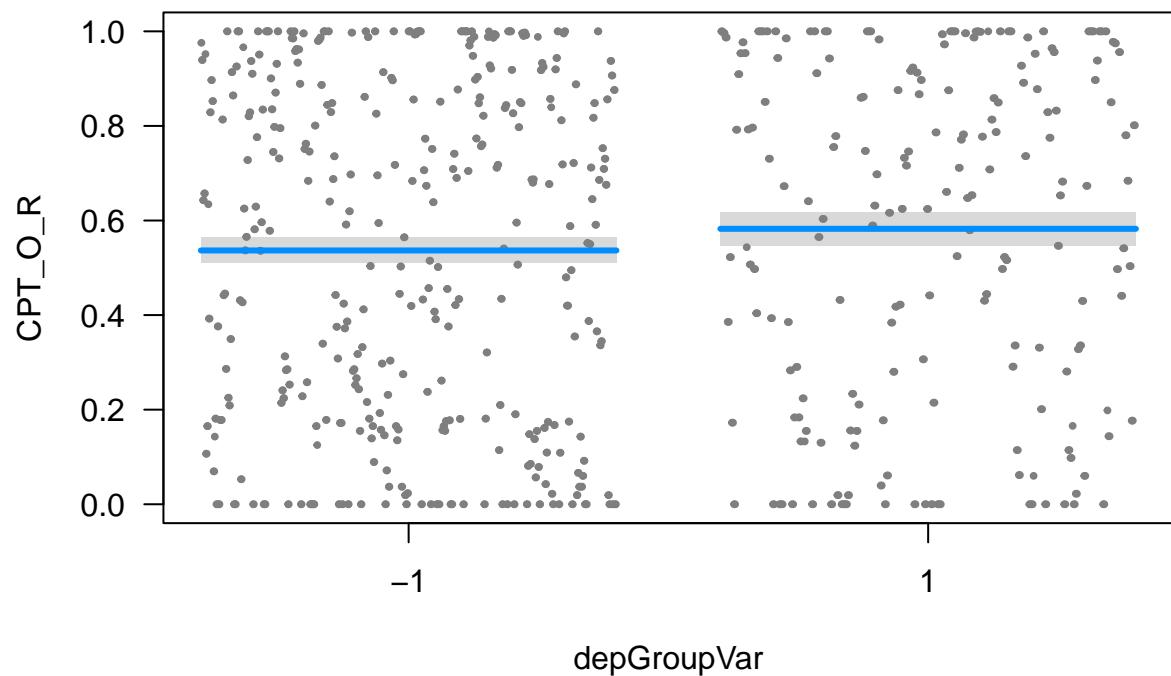


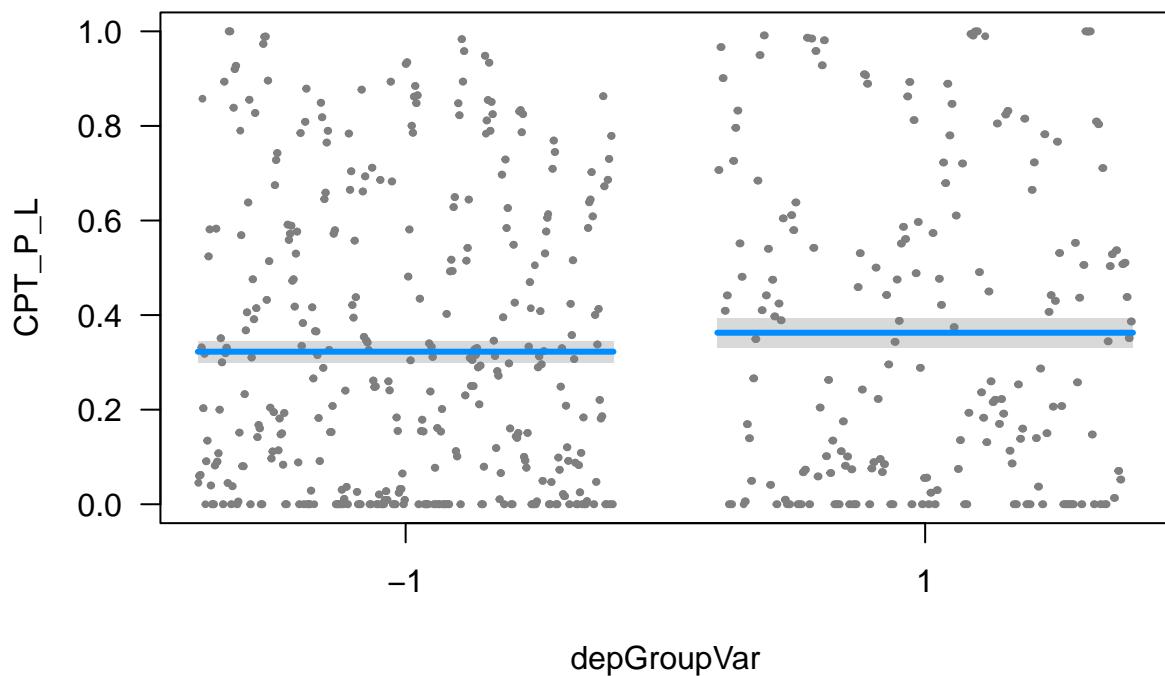


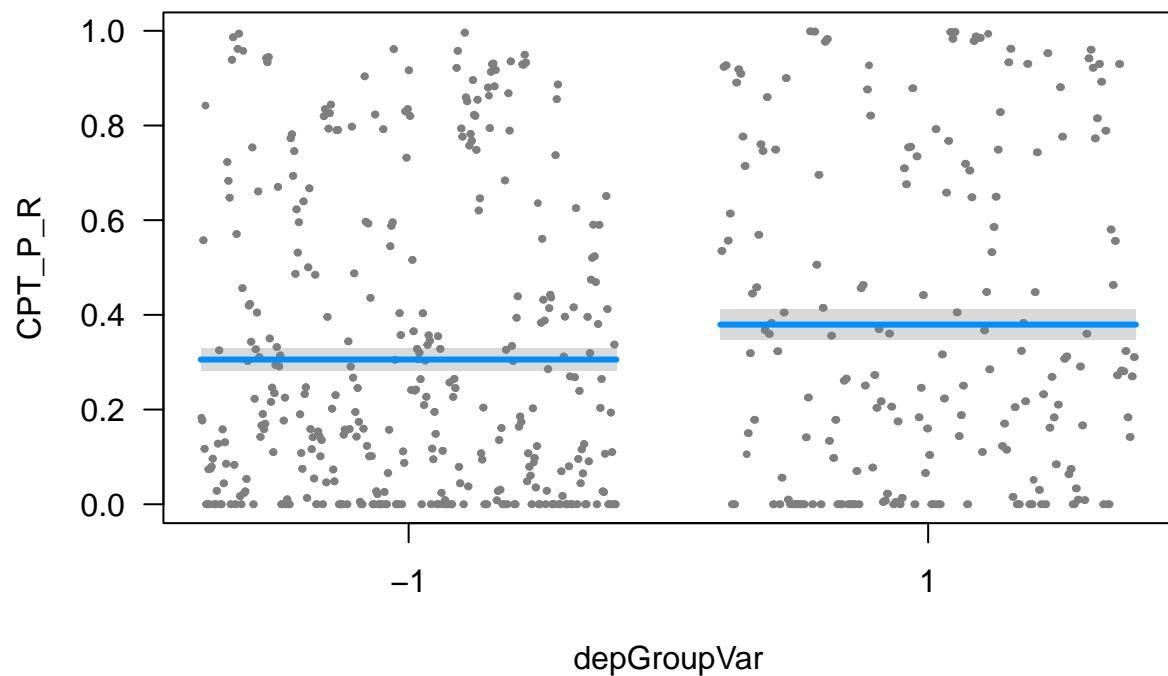


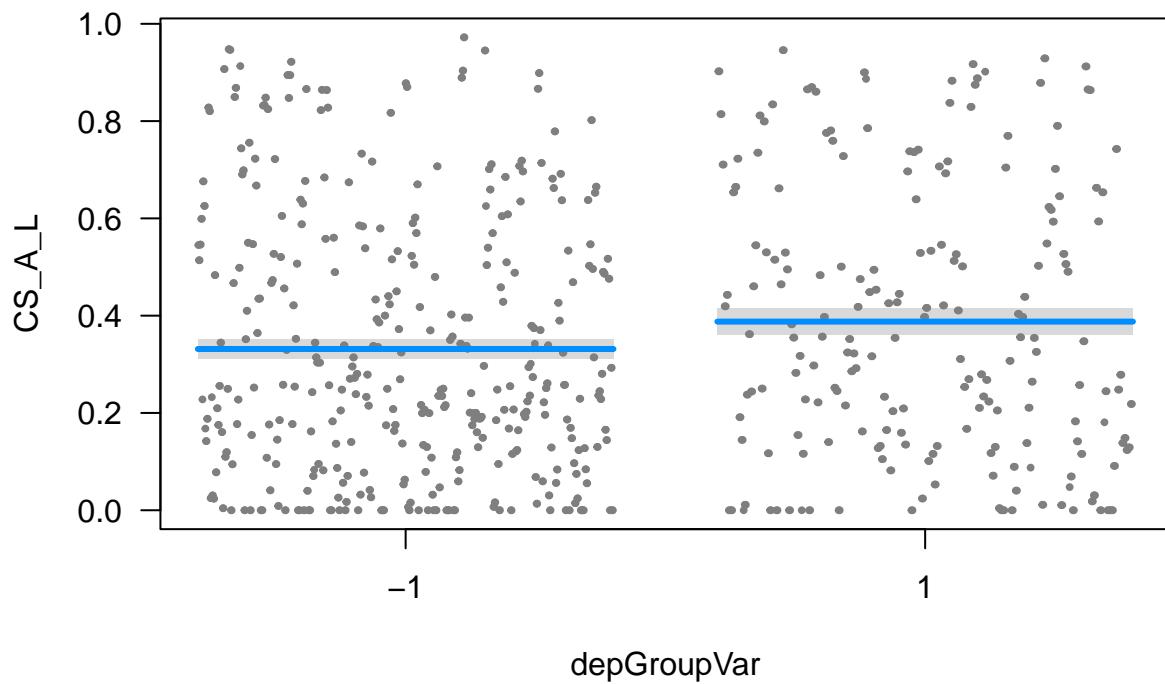


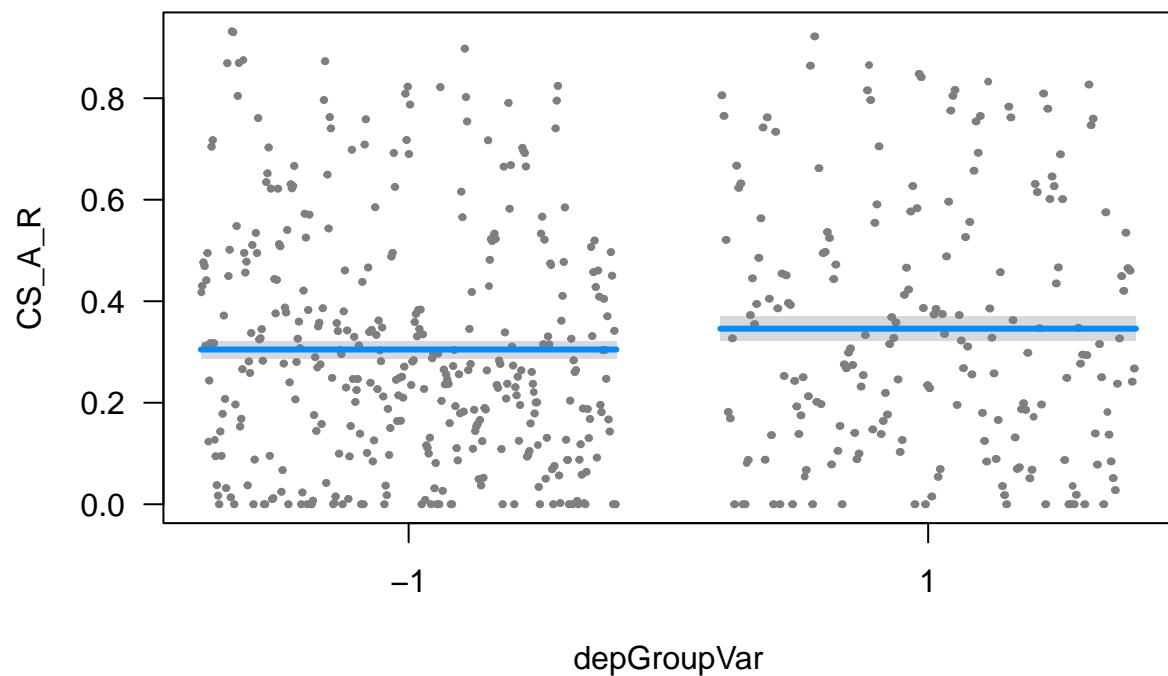


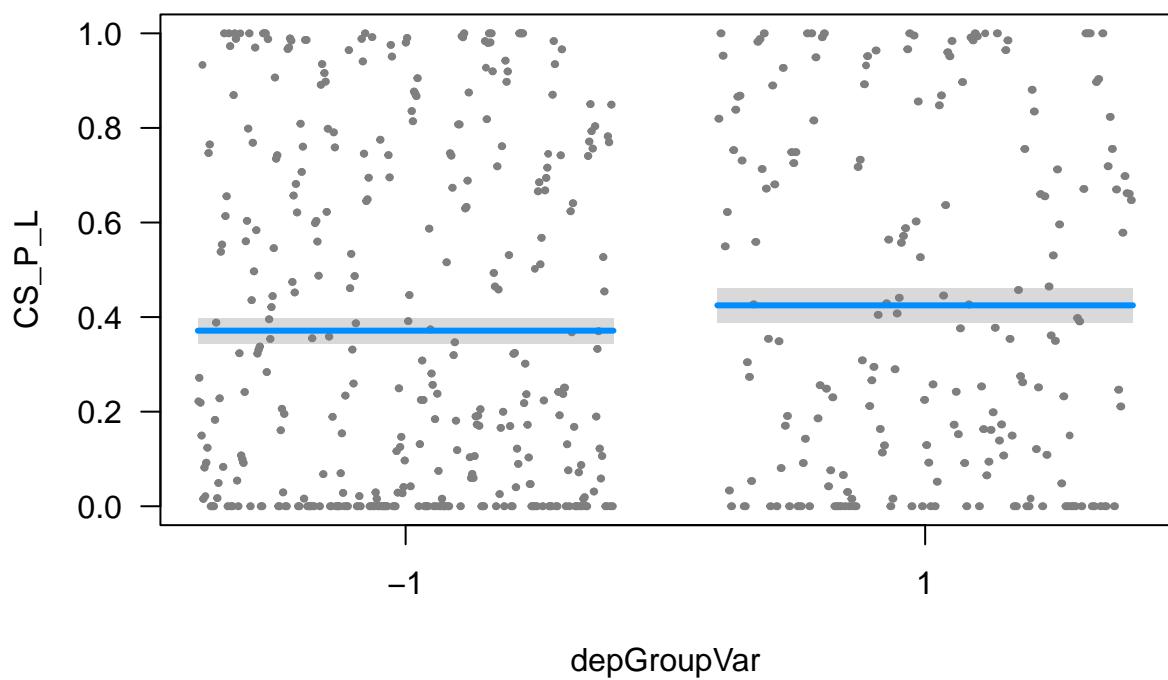


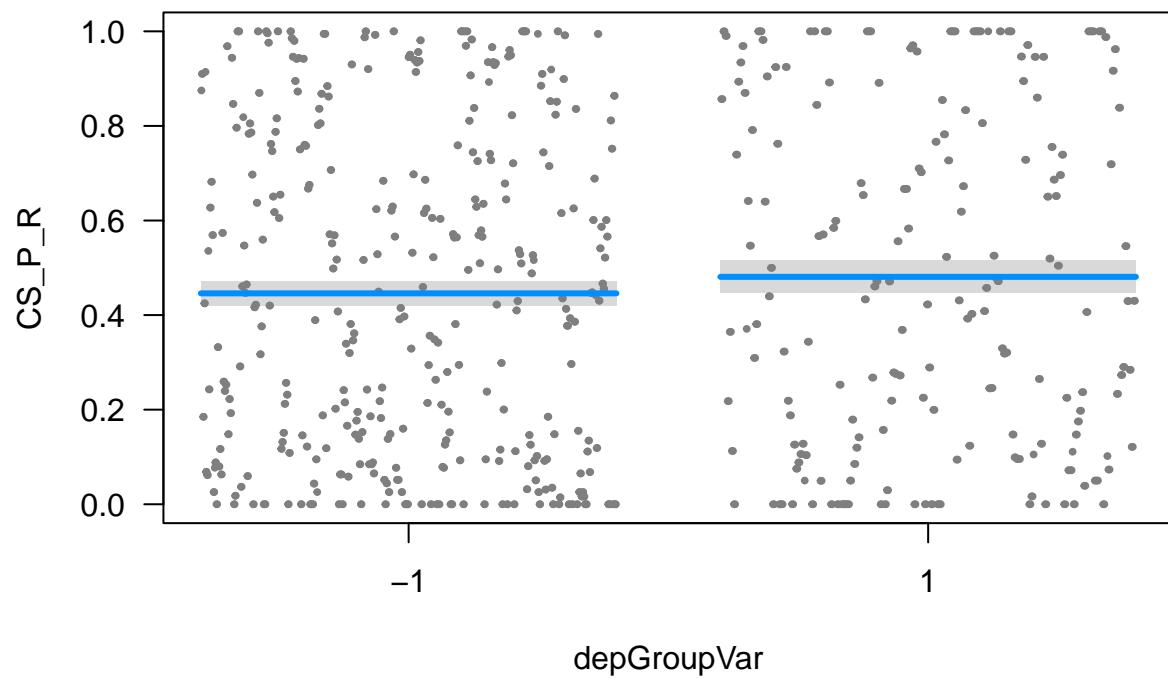


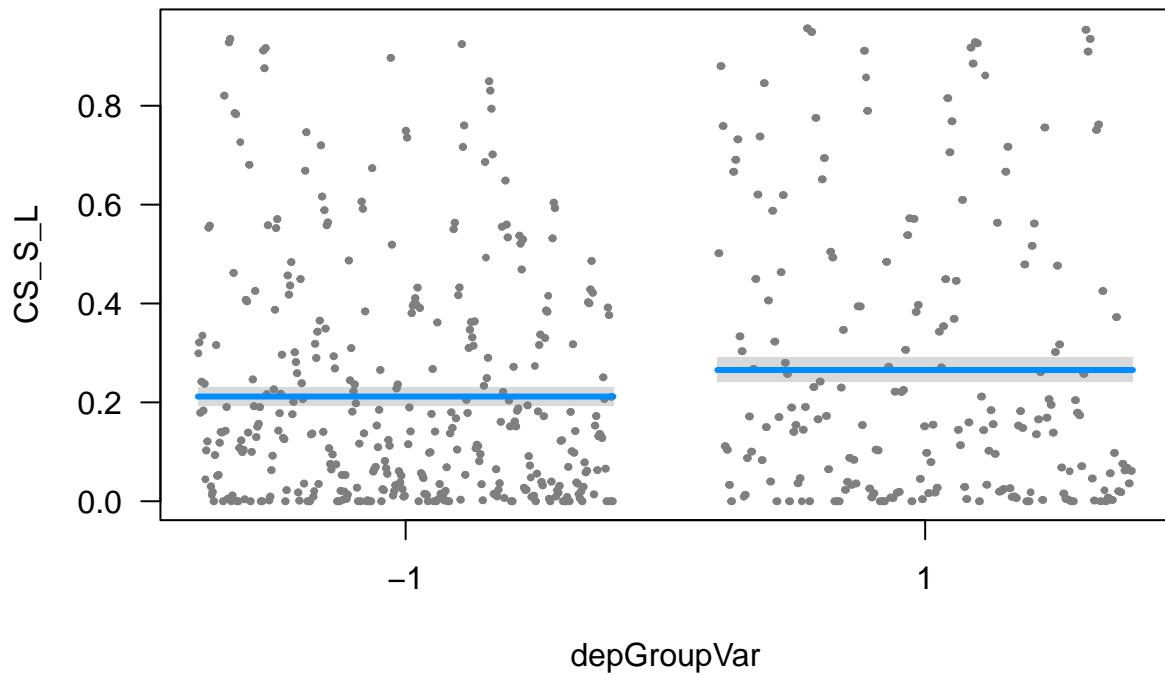


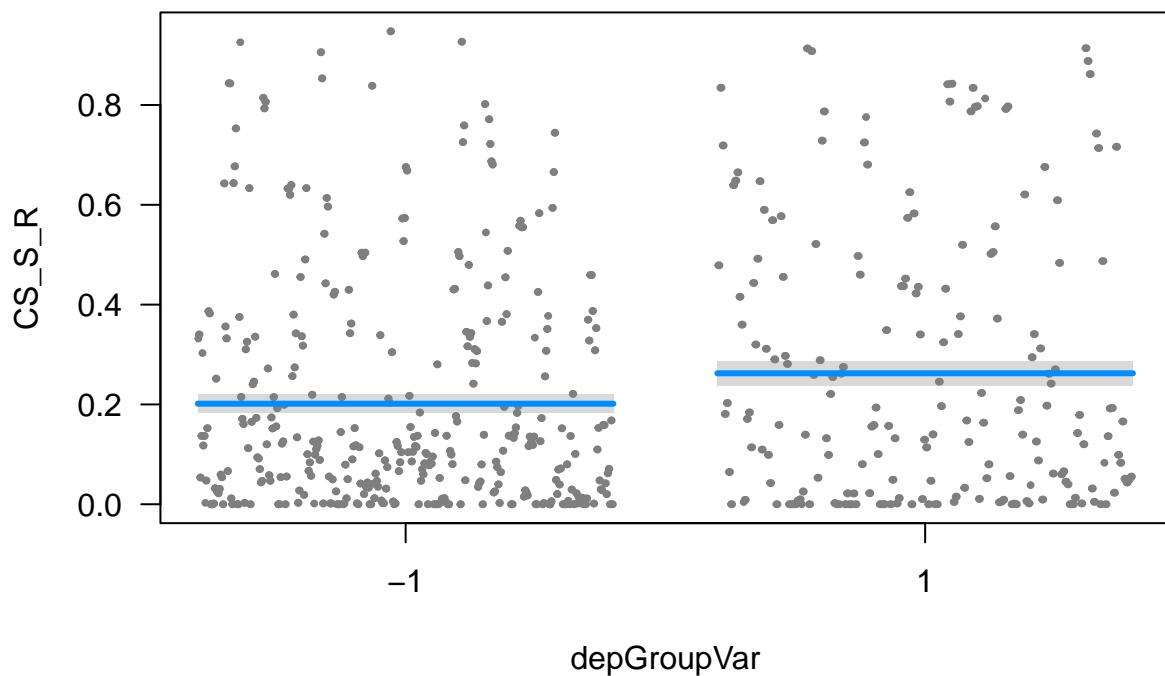


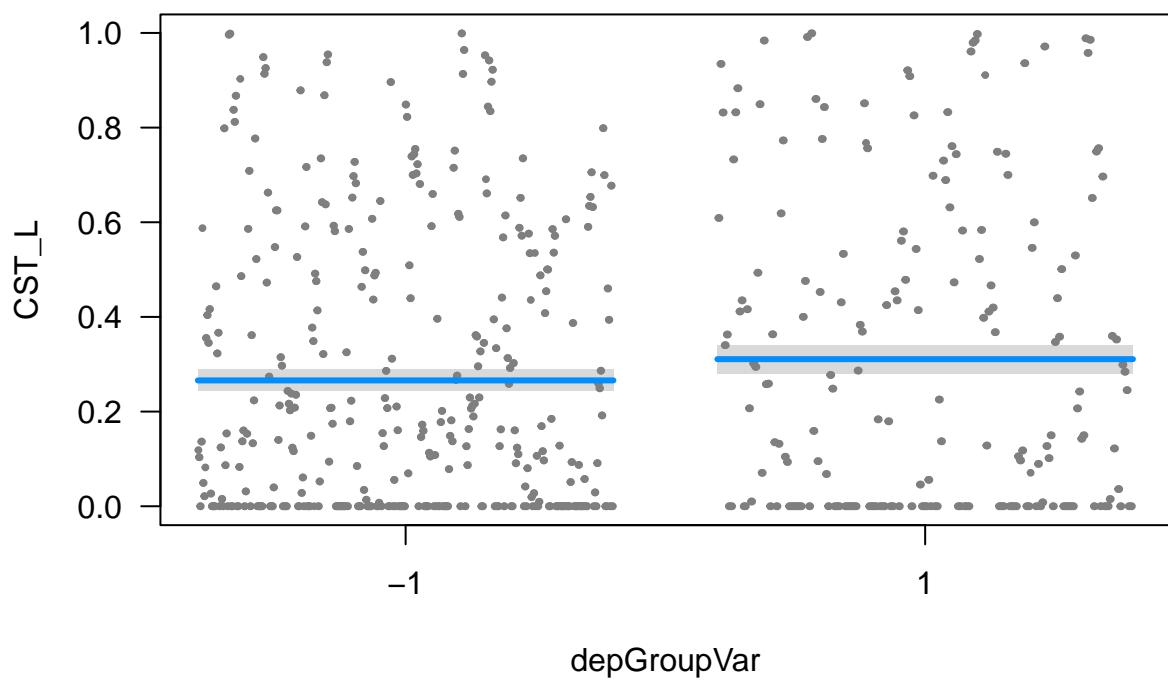


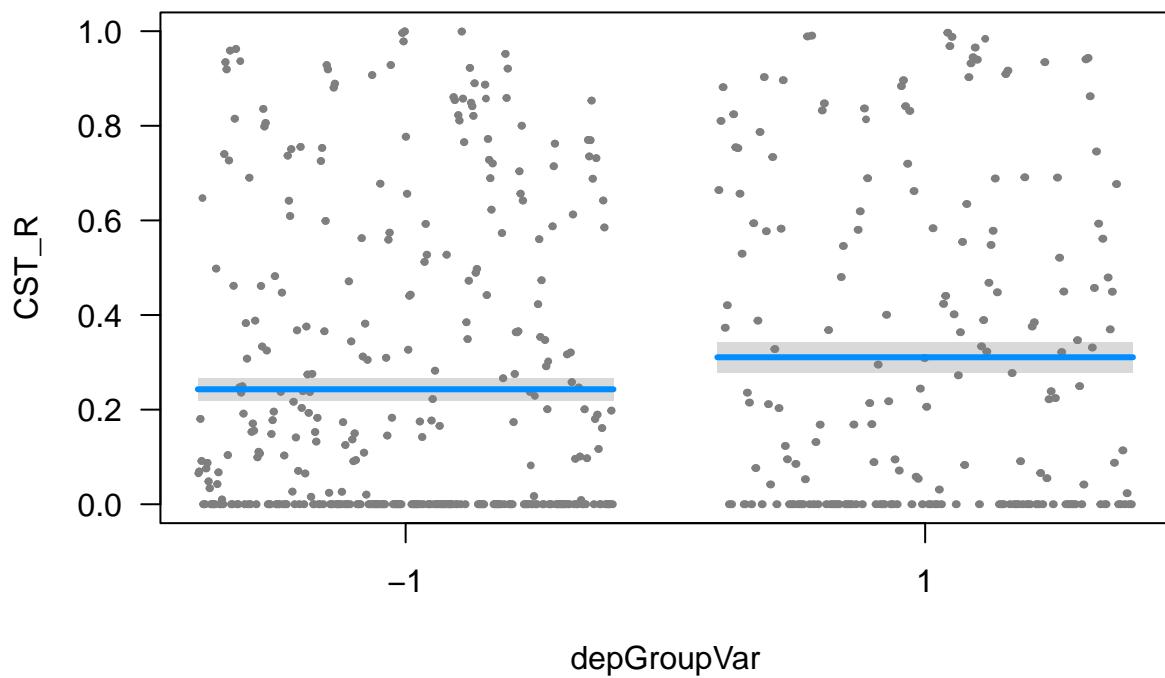


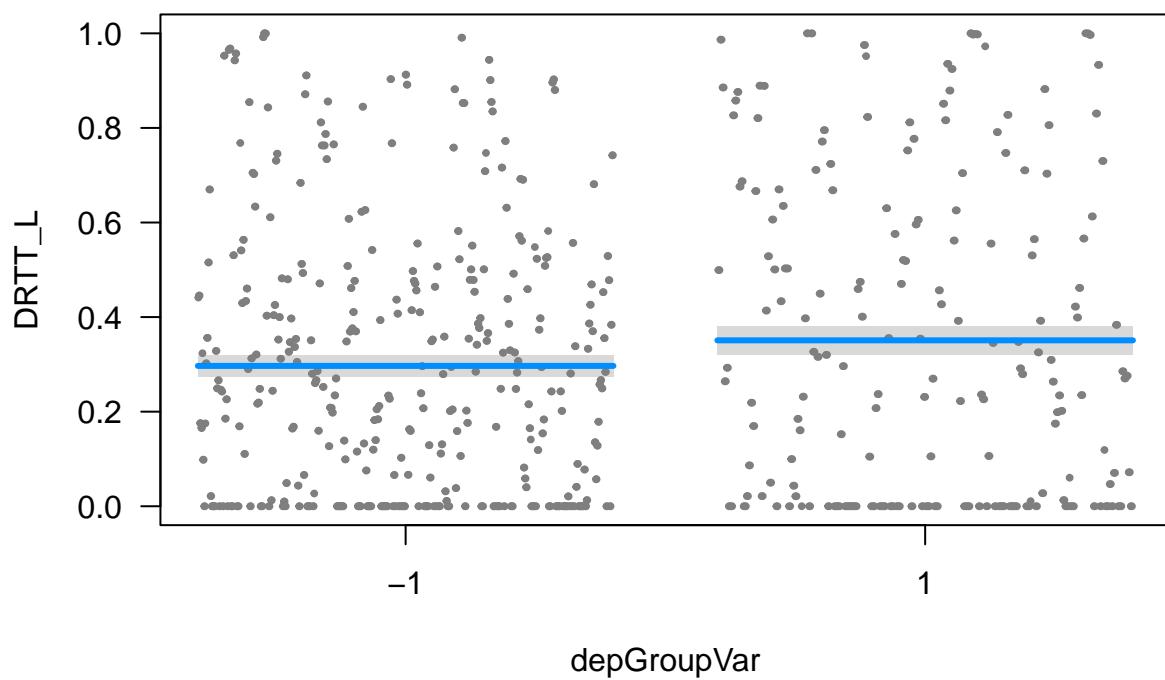


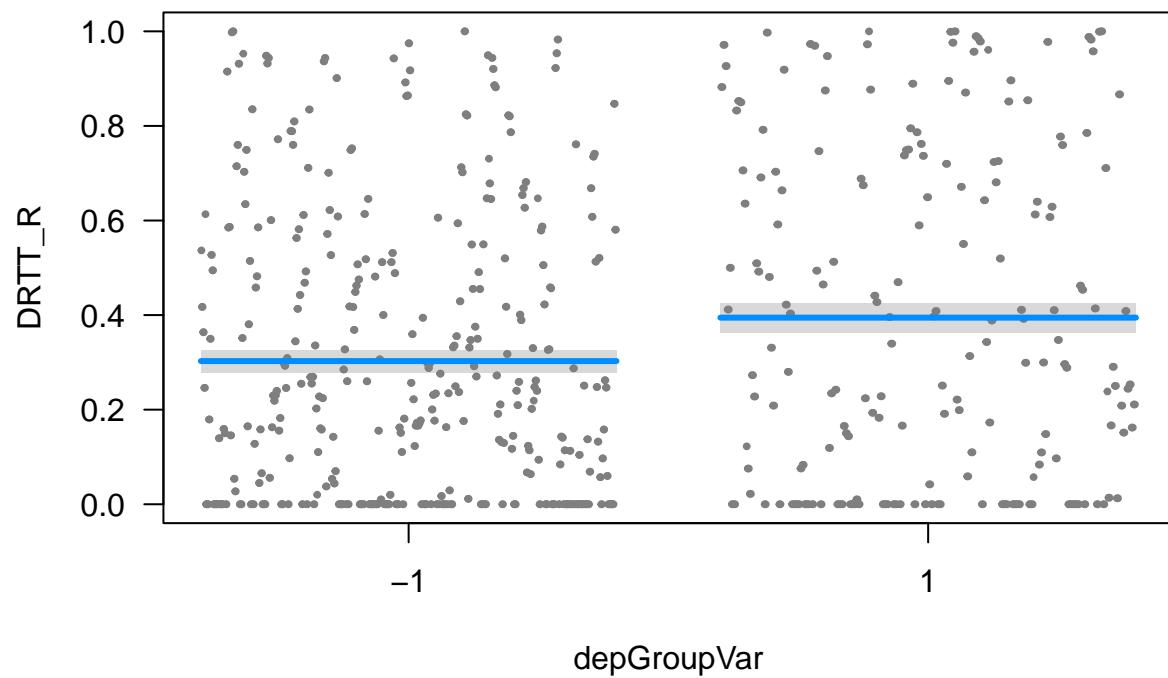


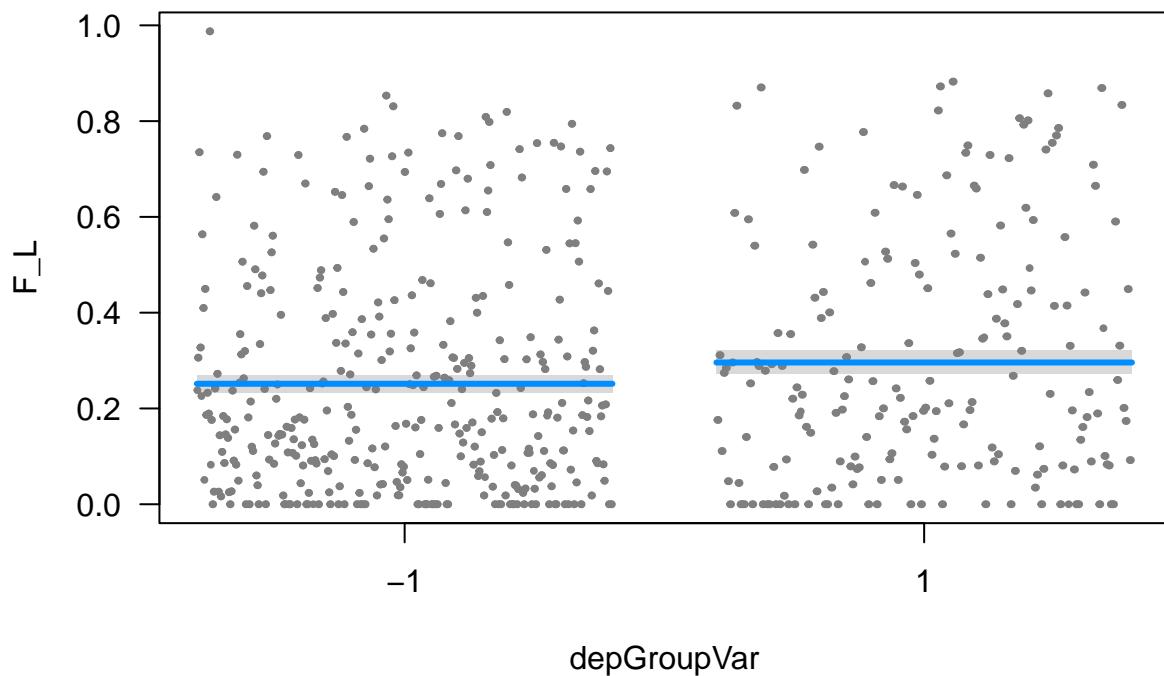


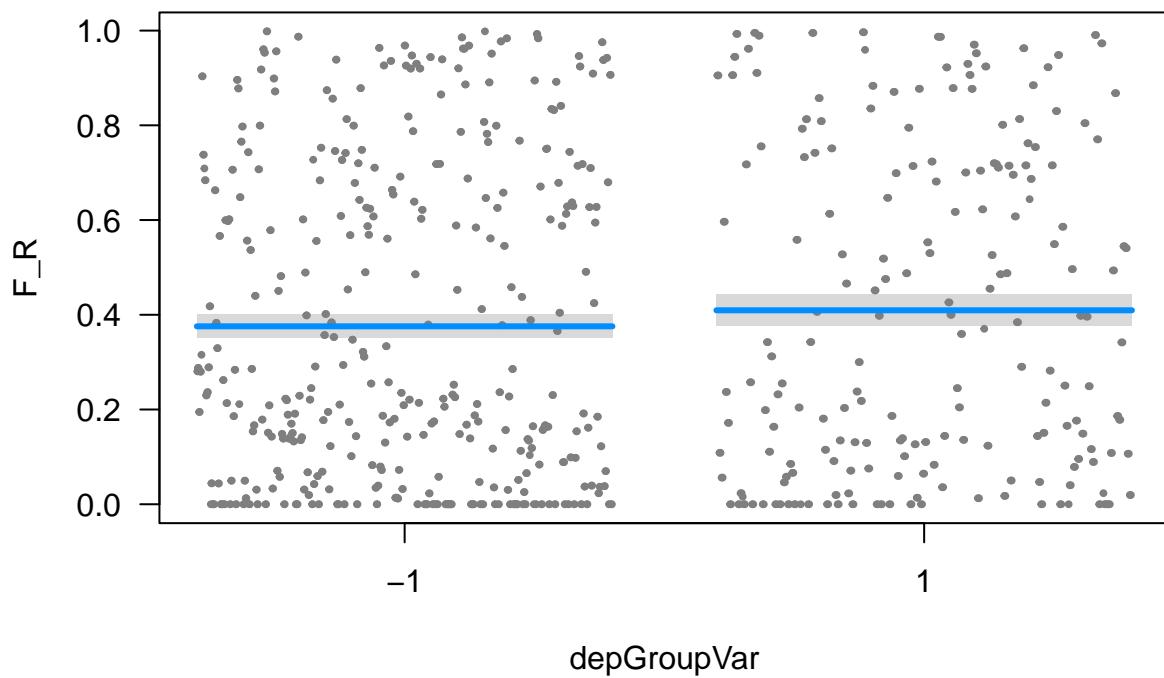


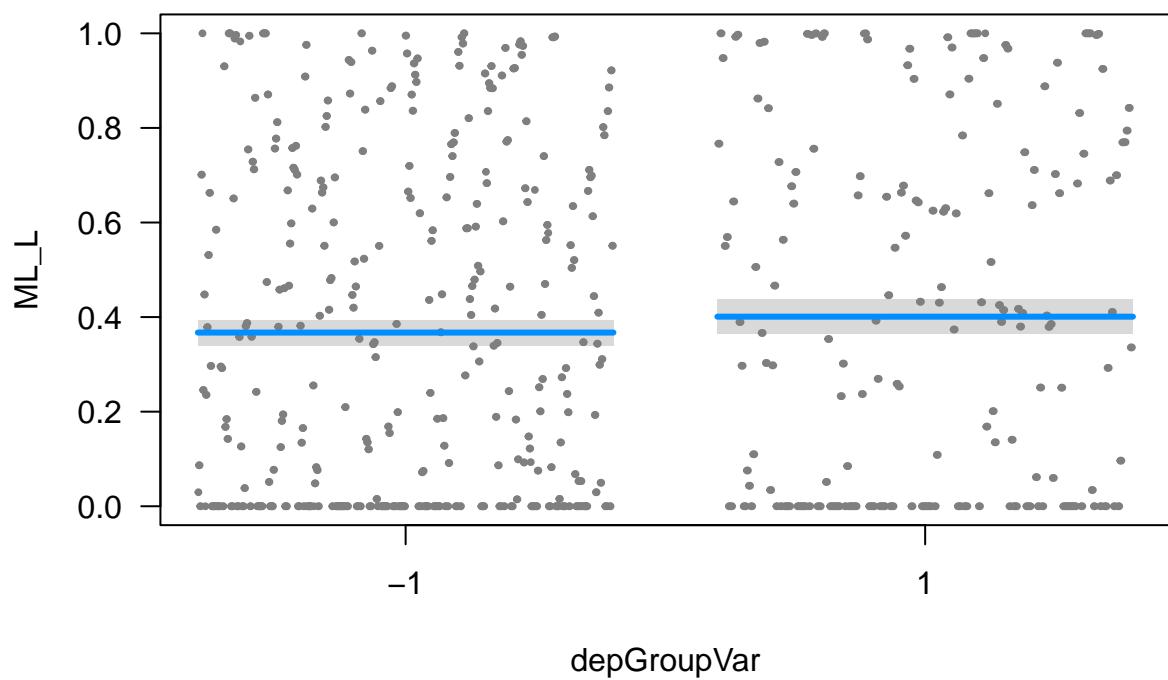


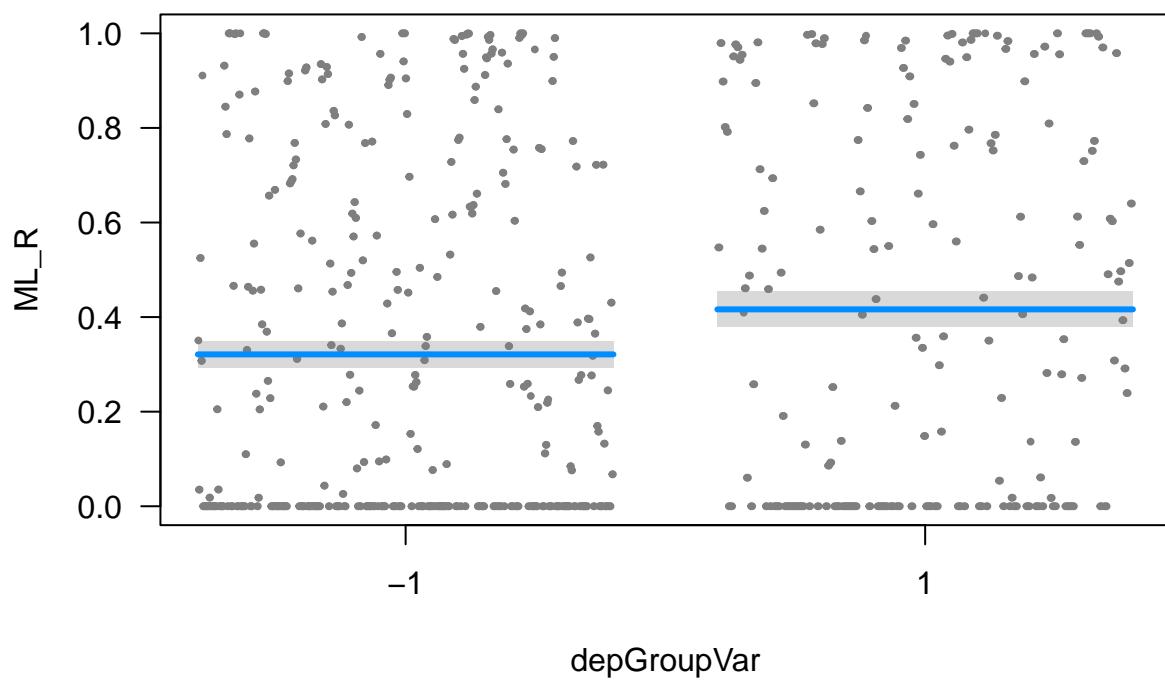


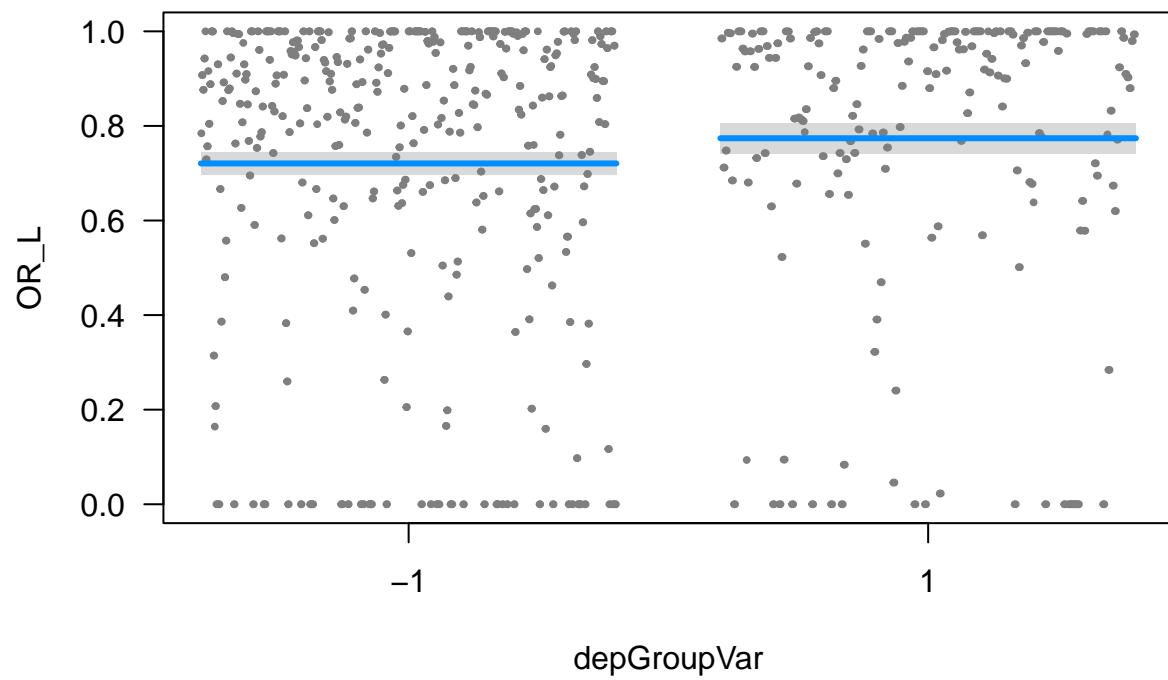


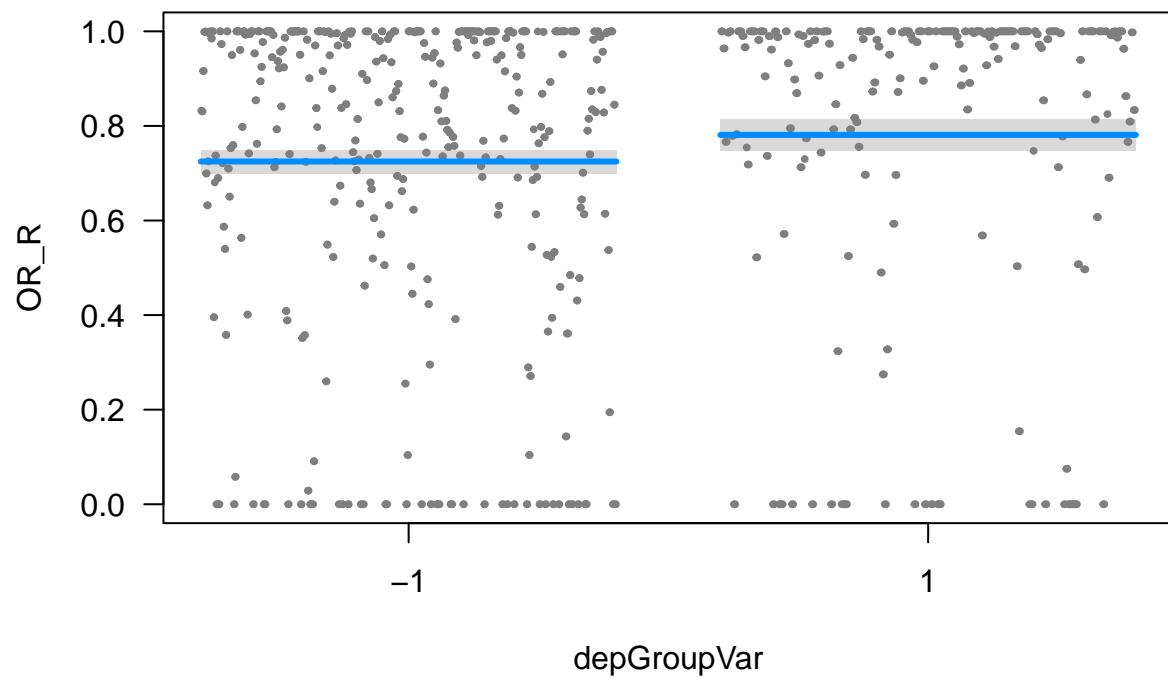


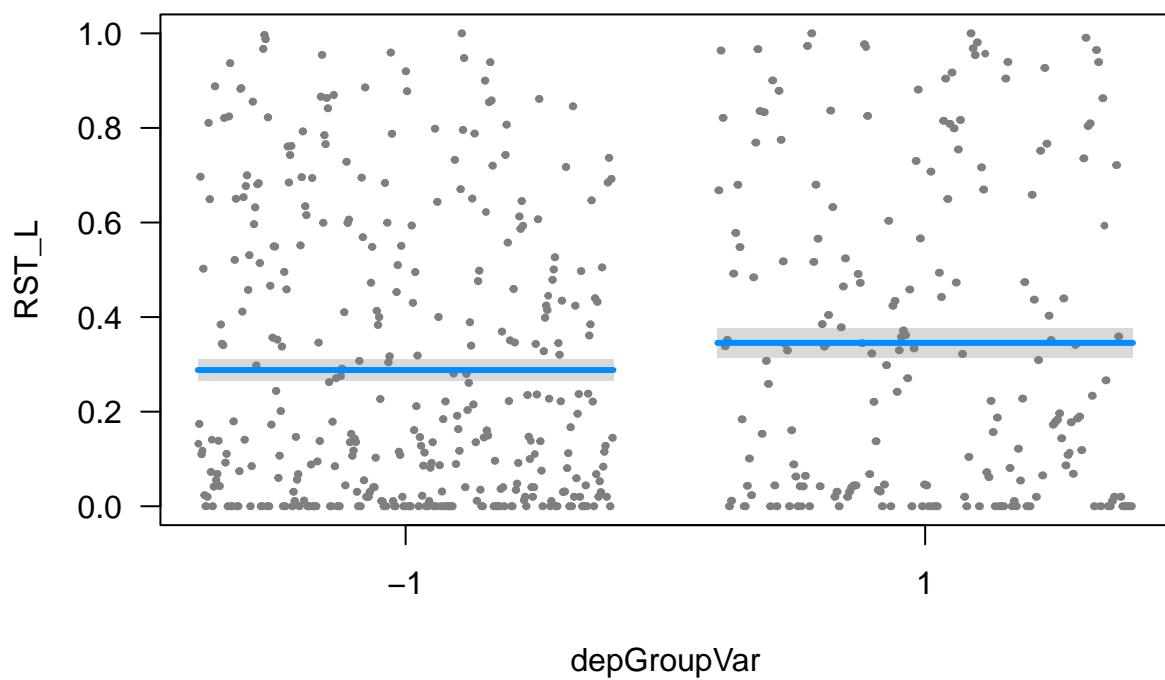


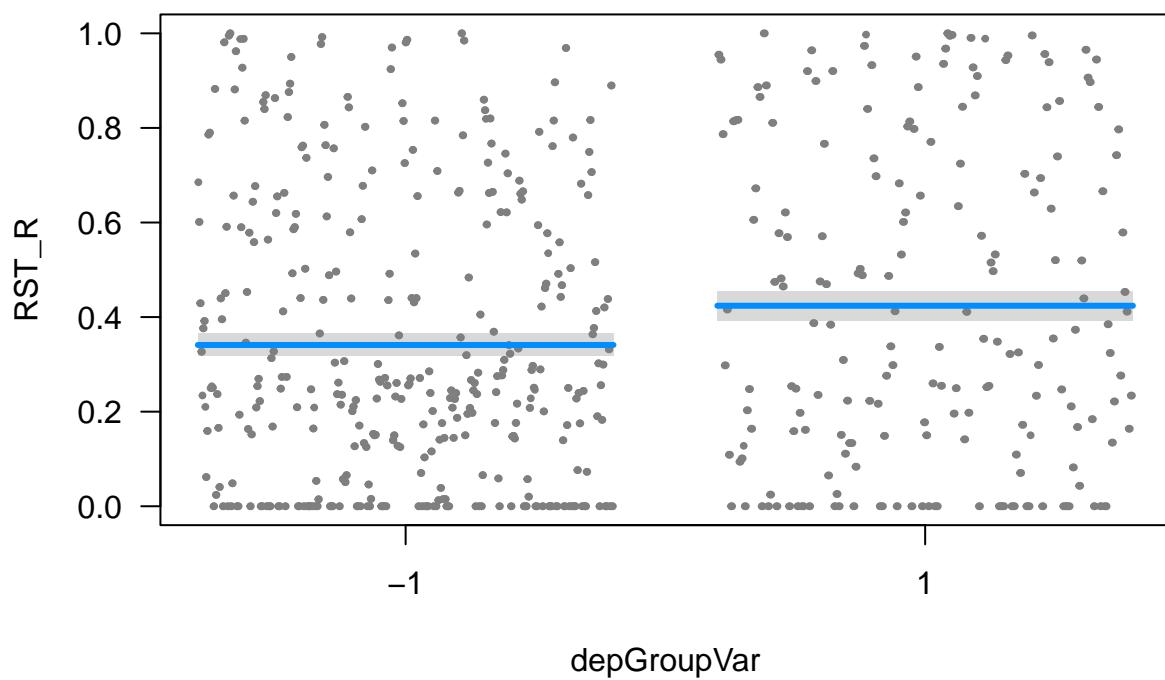


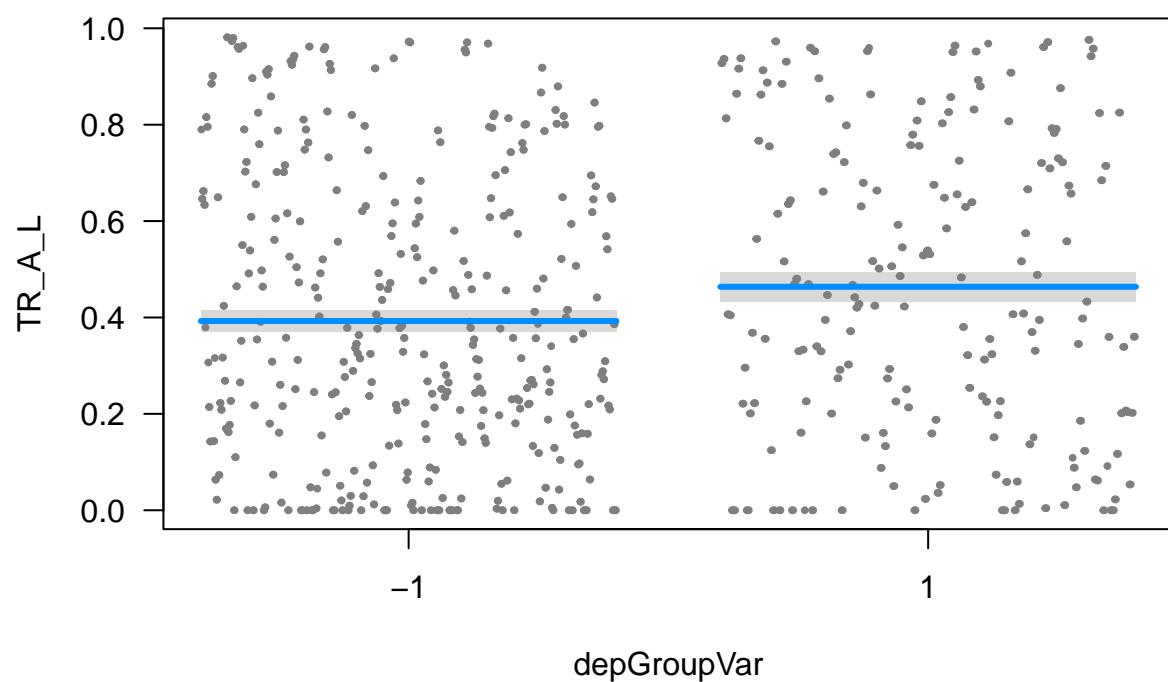


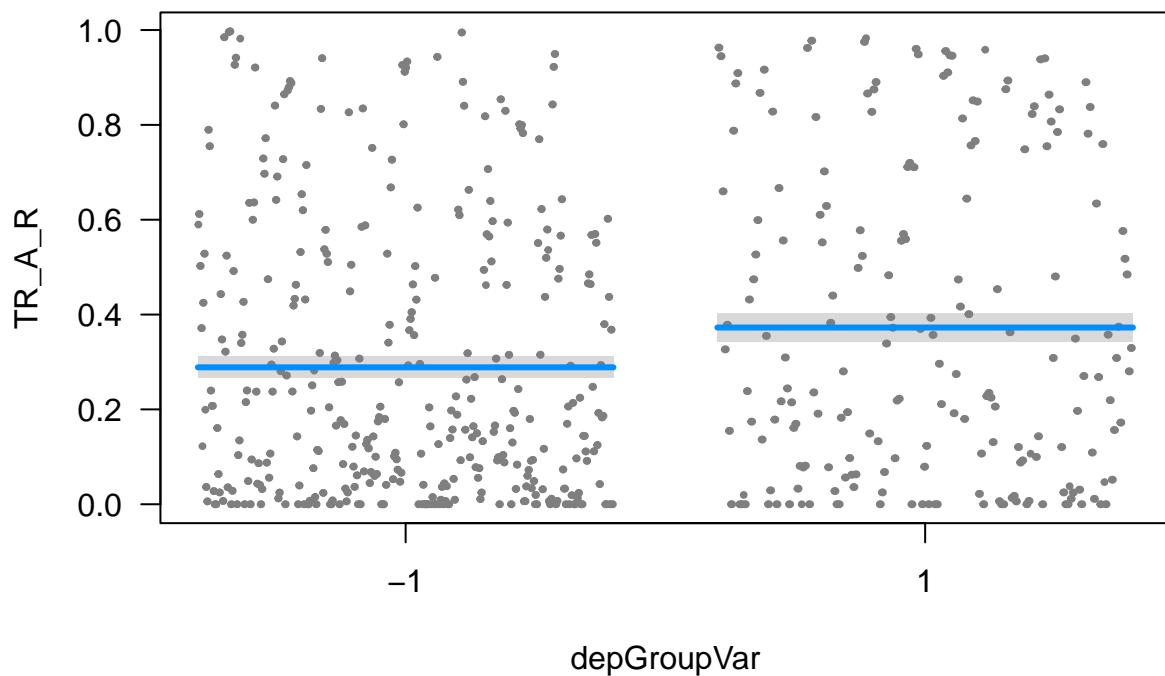


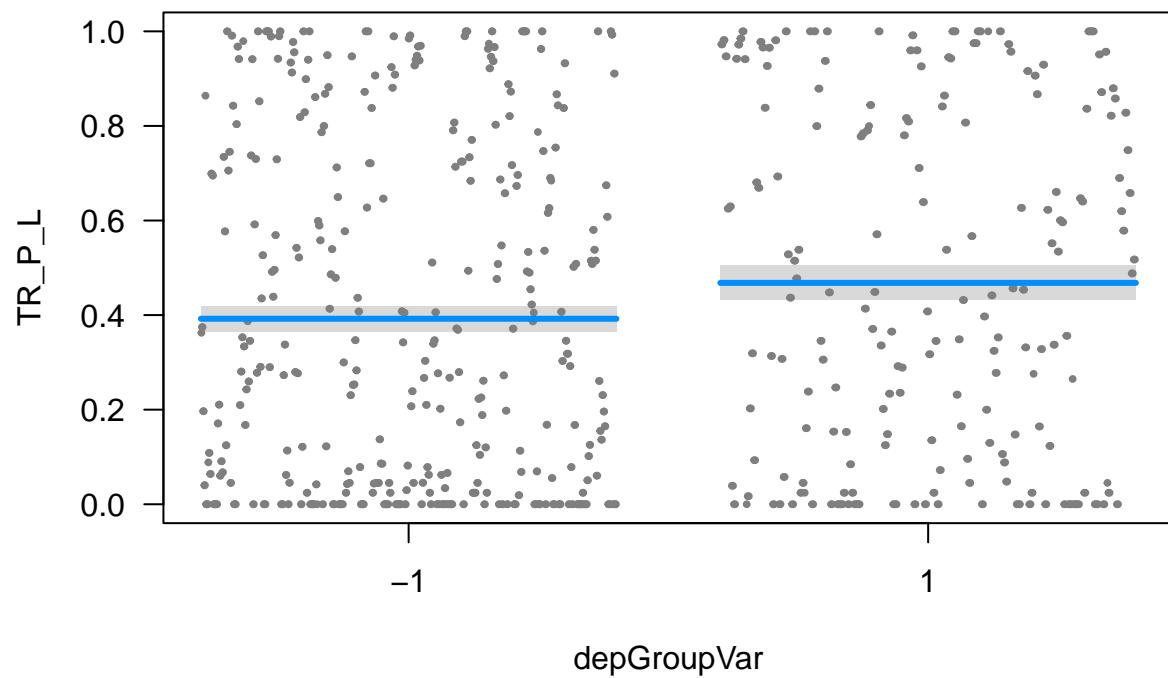


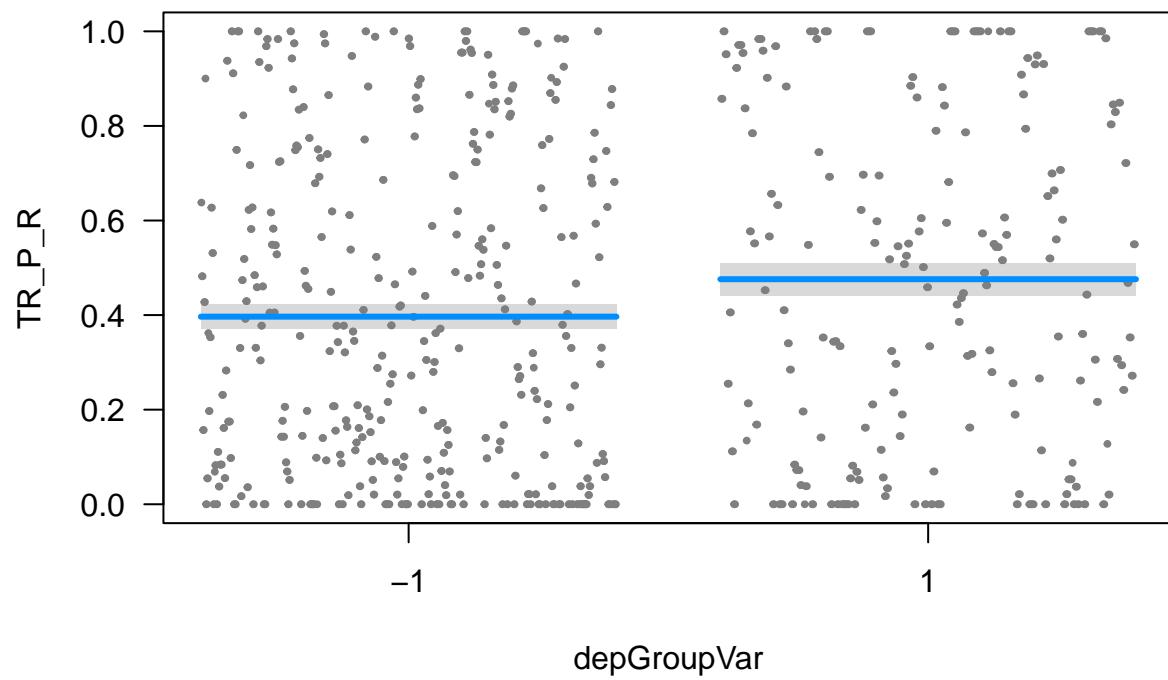


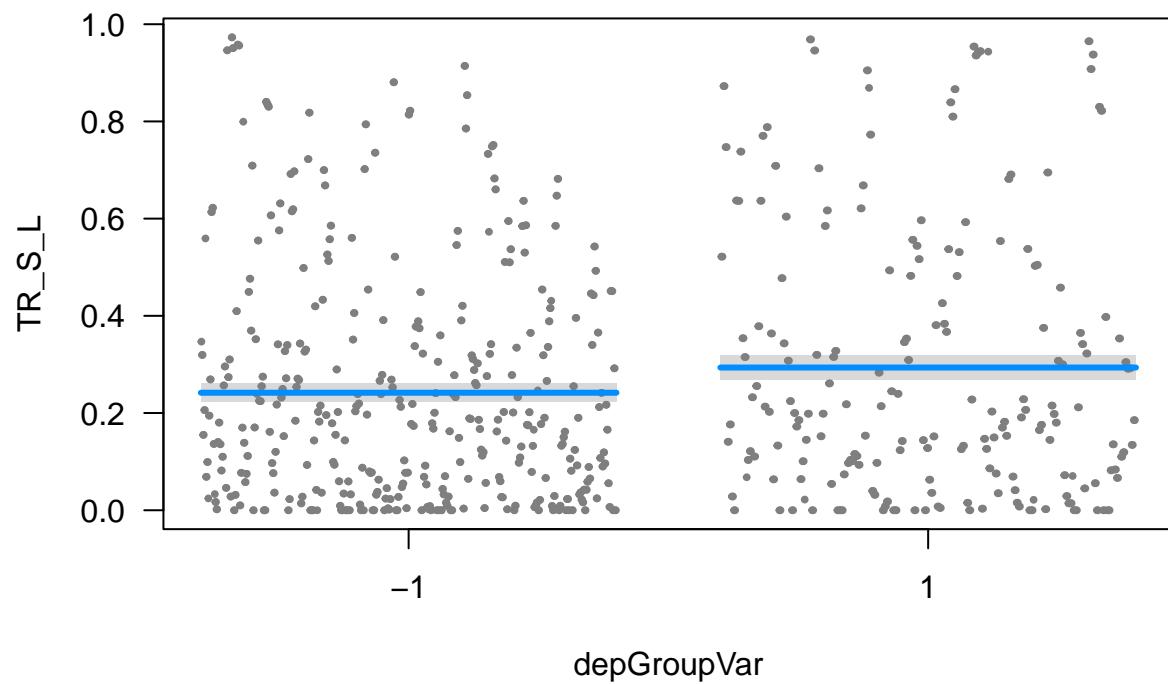


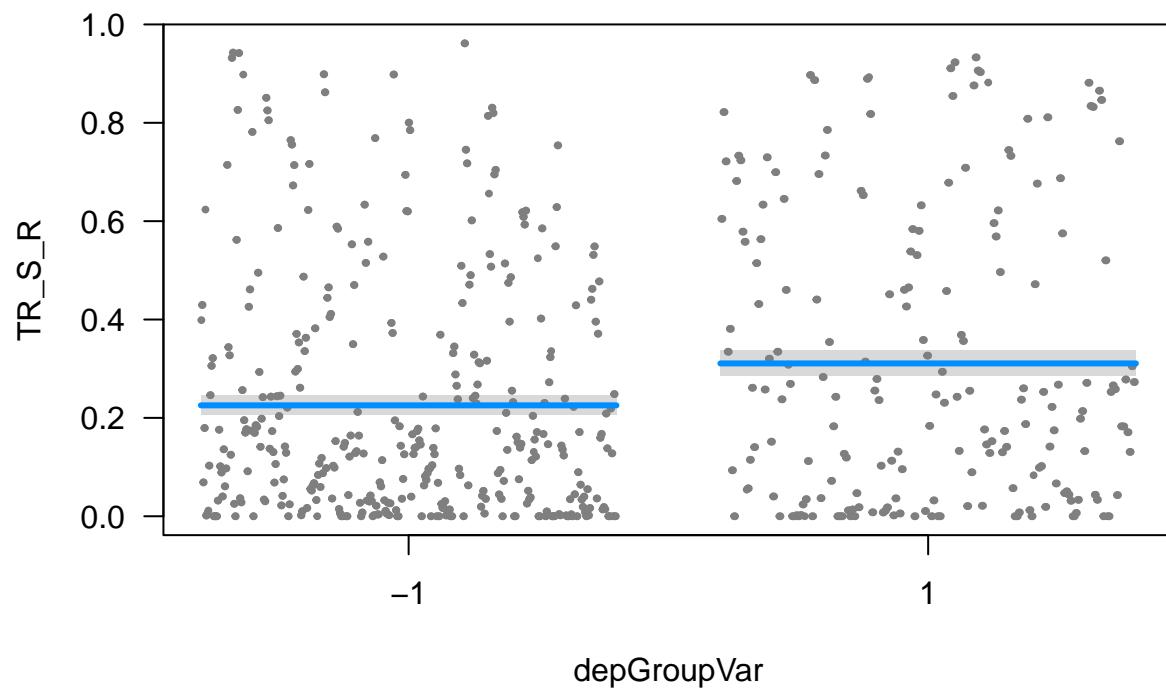


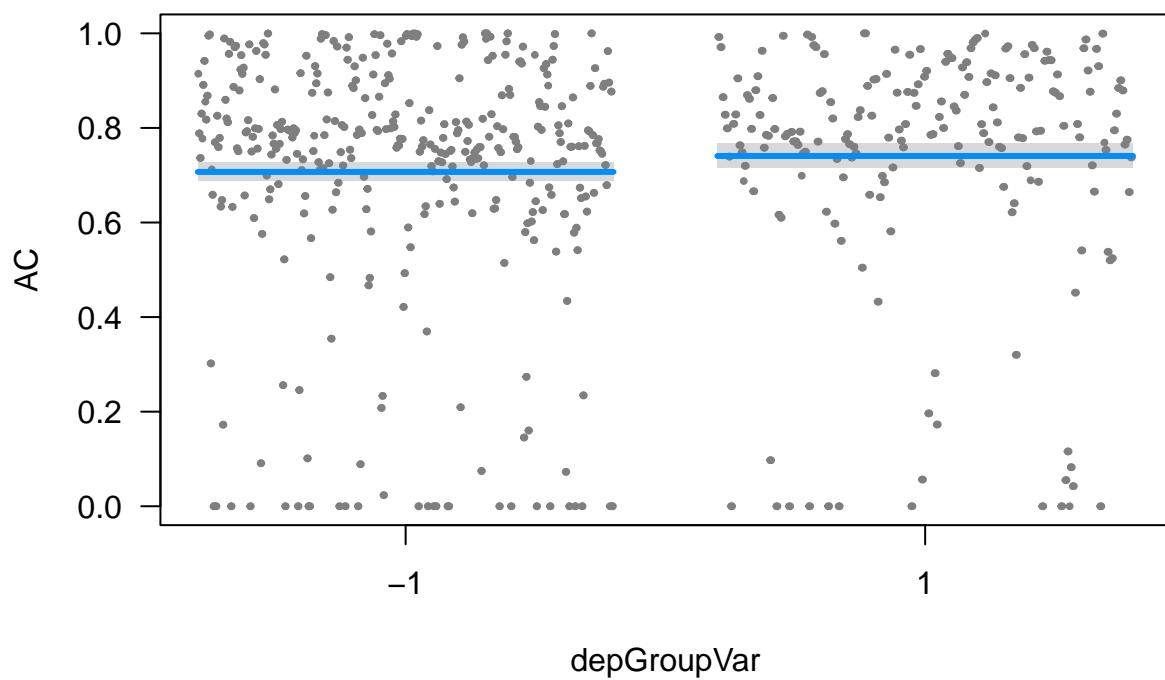


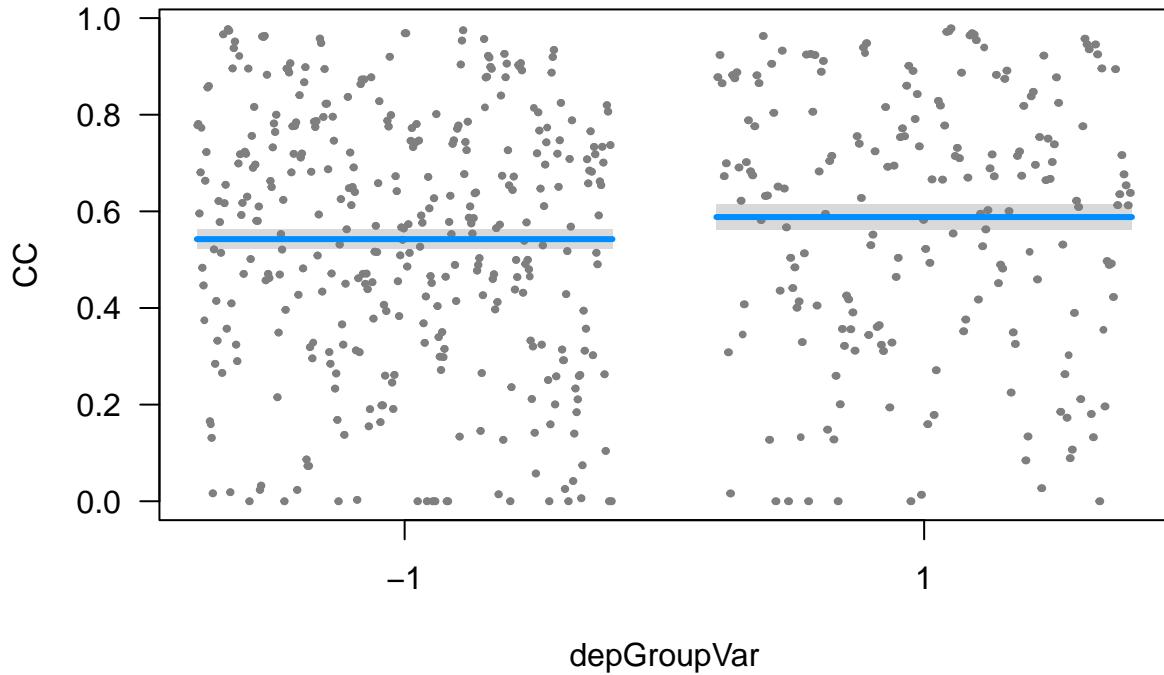












```

##      AF_L          AF_R          C_FPH_L          C_FPH_R          C_FP_L
## fit  data.frame,5 data.frame,5 data.frame,5 data.frame,5 data.frame,5
## res  data.frame,4 data.frame,4 data.frame,4 data.frame,4 data.frame,4
## meta list,6      list,6      list,6      list,6      list,6
##      C_FP_R          C_PH_L          C_PHP_L          C_PHP_R          C_PH_R
## fit  data.frame,5 data.frame,5 data.frame,5 data.frame,5 data.frame,5
## res  data.frame,4 data.frame,4 data.frame,4 data.frame,4 data.frame,4
## meta list,6      list,6      list,6      list,6      list,6
##      C_R_L          C_R_R          EMC_L          EMC_R          FAT_L
## fit  data.frame,5 data.frame,5 data.frame,5 data.frame,5 data.frame,5
## res  data.frame,4 data.frame,4 data.frame,4 data.frame,4 data.frame,4
## meta list,6      list,6      list,6      list,6      list,6
##      FAT_R          IFOF_L          IFOF_R          ILF_L          ILF_R
## fit  data.frame,5 data.frame,5 data.frame,5 data.frame,5 data.frame,5
## res  data.frame,4 data.frame,4 data.frame,4 data.frame,4 data.frame,4
## meta list,6      list,6      list,6      list,6      list,6
##      MdLF_L          MdLF_R          PAT_L          PAT_R          SLF1_L
## fit  data.frame,5 data.frame,5 data.frame,5 data.frame,5 data.frame,5
## res  data.frame,4 data.frame,4 data.frame,4 data.frame,4 data.frame,4
## meta list,6      list,6      list,6      list,6      list,6
##      SLF1_R          SLF2_L          SLF2_R          SLF3_L          SLF3_R
## fit  data.frame,5 data.frame,5 data.frame,5 data.frame,5 data.frame,5
## res  data.frame,4 data.frame,4 data.frame,4 data.frame,4 data.frame,4
## meta list,6      list,6      list,6      list,6      list,6
##      UF_L          UF_R          VOF_L          VOF_R          CB_L
## fit  data.frame,5 data.frame,5 data.frame,5 data.frame,5 data.frame,5

```

```

## res data.frame,4 data.frame,4 data.frame,4 data.frame,4 data.frame,4
## meta list,6      list,6      list,6      list,6      list,6
##   CB_R          ICP_L      ICP_R      MCP       SCP
## fit data.frame,5 data.frame,5 data.frame,5 data.frame,5 data.frame,5
## res data.frame,4 data.frame,4 data.frame,4 data.frame,4 data.frame,4
## meta list,6      list,6      list,6      list,6      list,6
##   V            CNIII_L    CNIII_R    CNII_L    CNII_R
## fit data.frame,5 data.frame,5 data.frame,5 data.frame,5 data.frame,5
## res data.frame,4 data.frame,4 data.frame,4 data.frame,4 data.frame,4
## meta list,6      list,6      list,6      list,6      list,6
##   CNVIII_L    CNVIII_R   CNVII_L    CNVII_R   CNV_L
## fit data.frame,5 data.frame,5 data.frame,5 data.frame,5 data.frame,5
## res data.frame,4 data.frame,4 data.frame,4 data.frame,4 data.frame,4
## meta list,6      list,6      list,6      list,6      list,6
##   CNV_R        AR_L      AR_R       CBT_L     CBT_R
## fit data.frame,5 data.frame,5 data.frame,5 data.frame,5 data.frame,5
## res data.frame,4 data.frame,4 data.frame,4 data.frame,4 data.frame,4
## meta list,6      list,6      list,6      list,6      list,6
##   CPT_F_L     CPT_F_R    CPT_O_L    CPT_O_R    CPT_P_L
## fit data.frame,5 data.frame,5 data.frame,5 data.frame,5 data.frame,5
## res data.frame,4 data.frame,4 data.frame,4 data.frame,4 data.frame,4
## meta list,6      list,6      list,6      list,6      list,6
##   CPT_P_R     CS_A_L    CS_A_R     CS_P_L    CS_P_R
## fit data.frame,5 data.frame,5 data.frame,5 data.frame,5 data.frame,5
## res data.frame,4 data.frame,4 data.frame,4 data.frame,4 data.frame,4
## meta list,6      list,6      list,6      list,6      list,6
##   CS_S_L      CS_S_R    CST_L      CST_R     DRTT_L
## fit data.frame,5 data.frame,5 data.frame,5 data.frame,5 data.frame,5
## res data.frame,4 data.frame,4 data.frame,4 data.frame,4 data.frame,4
## meta list,6      list,6      list,6      list,6      list,6
##   DRTT_R      F_L       F_R       ML_L      ML_R
## fit data.frame,5 data.frame,5 data.frame,5 data.frame,5 data.frame,5
## res data.frame,4 data.frame,4 data.frame,4 data.frame,4 data.frame,4
## meta list,6      list,6      list,6      list,6      list,6
##   OR_L        OR_R      RST_L      RST_R     TR_A_L
## fit data.frame,5 data.frame,5 data.frame,5 data.frame,5 data.frame,5
## res data.frame,4 data.frame,4 data.frame,4 data.frame,4 data.frame,4
## meta list,6      list,6      list,6      list,6      list,6
##   TR_A_R      TR_P_L    TR_P_R    TR_S_L    TR_S_R
## fit data.frame,5 data.frame,5 data.frame,5 data.frame,5 data.frame,5
## res data.frame,4 data.frame,4 data.frame,4 data.frame,4 data.frame,4
## meta list,6      list,6      list,6      list,6      list,6
##   AC          CC
## fit data.frame,5 data.frame,5
## res data.frame,4 data.frame,4
## meta list,6      list,6

#anova
fascicle_anova <- lapply(fascicle_lm, anova)

#fdr corrected
fascicle_anova_fdr <- fdr_anova_generic(fascicle_anova, 1)

##           p_anova

```

```

## AF_L      2.578214e-03
## AF_R      6.174158e-03
## C_FPH_L   9.624075e-01
## C_FPH_R   5.844313e-01
## C_FP_L    3.398230e-02
## C_FP_R    5.158020e-01
## C_PH_L    1.357335e-03
## C_PHP_L   9.659307e-02
## C_PHP_R   2.603931e-02
## C_PH_R    4.552169e-01
## C_R_L     1.945615e-02
## C_R_R     1.028564e-01
## EMC_L     3.373818e-02
## EMC_R     9.679307e-02
## FAT_L     2.282320e-02
## FAT_R     2.562681e-05
## IFOF_L    1.121774e-03
## IFOF_R    4.455661e-02
## ILF_L     4.880404e-05
## ILF_R     8.276630e-03
## MdLF_L    3.453679e-02
## MdLF_R    3.176372e-02
## PAT_L     3.443343e-03
## PAT_R     3.748231e-05
## SLF1_L    6.381944e-04
## SLF1_R    3.854758e-01
## SLF2_L    2.085115e-02
## SLF2_R    6.031135e-04
## SLF3_L    5.061243e-03
## SLF3_R    3.607070e-03
## UF_L      1.099838e-01
## UF_R      1.214621e-01
## VOF_L     3.322972e-02
## VOF_R     5.213149e-06
## CB_L      9.309656e-01
## CB_R      5.519092e-01
## ICP_L     5.679475e-01
## ICP_R     8.055311e-01
## MCP       7.730769e-01
## SCP       1.778219e-01
## V         6.658906e-01
## CNIII_L   6.343637e-01
## CNIII_R   4.666709e-01
## CNII_L    3.700836e-01
## CNII_R    3.529199e-01
## CNVIII_L  2.361714e-01
## CNVIII_R  7.577010e-01
## CNVII_L   2.101867e-01
## CNVII_R   2.937322e-01
## CNV_L     6.910983e-02
## CNV_R     3.083527e-03
## AR_L      4.026356e-03
## AR_R      9.457673e-05
## CBT_L     3.048118e-05

```

```

## CBT_R      5.687075e-03
## CPT_F_L   1.886671e-03
## CPT_F_R   5.768941e-05
## CPT_O_L   4.502135e-04
## CPT_O_R   4.413450e-02
## CPT_P_L   4.252117e-02
## CPT_P_R   3.083369e-04
## CS_A_L    1.065540e-03
## CS_A_R    6.444177e-03
## CS_P_L    2.023262e-02
## CS_P_R    1.146695e-01
## CS_S_L    6.937632e-04
## CS_S_R    8.898003e-05
## CST_L     2.073714e-02
## CST_R     8.125128e-04
## DRTT_L    4.802059e-03
## DRTT_R    4.888754e-06
## F_L       3.842894e-03
## F_R       1.112795e-01
## ML_L      1.425907e-01
## ML_R      6.727661e-05
## OR_L      9.572355e-03
## OR_R      8.179265e-03
## RST_L     3.462419e-03
## RST_R     3.550929e-05
## TR_A_L    2.192602e-04
## TR_A_R    1.347576e-05
## TR_P_L    1.056143e-03
## TR_P_R    3.014904e-04
## TR_S_L    1.229346e-03
## TR_S_R    3.440764e-07
## AC        4.324205e-02
## CC        6.941672e-03

```

```
print(fascicle_anova_fdr)
```

```

##      component p_FDR_corr
## 1          AF_L     0.008
## 2          AF_R     0.014
## 3          C_PH_L    0.005
## 4          C_PHP_R    0.046
## 5          C_R_L     0.038
## 6          FAT_L     0.041
## 7          FAT_R      0
## 8          IFOF_L    0.004
## 9          ILF_L      0
## 10         ILF_R     0.017
## 11         PAT_L     0.01
## 12         PAT_R      0
## 13         SLF1_L    0.003
## 14         SLF2_L    0.039
## 15         SLF2_R    0.003
## 16         SLF3_L    0.012
## 17         SLF3_R    0.01

```

```

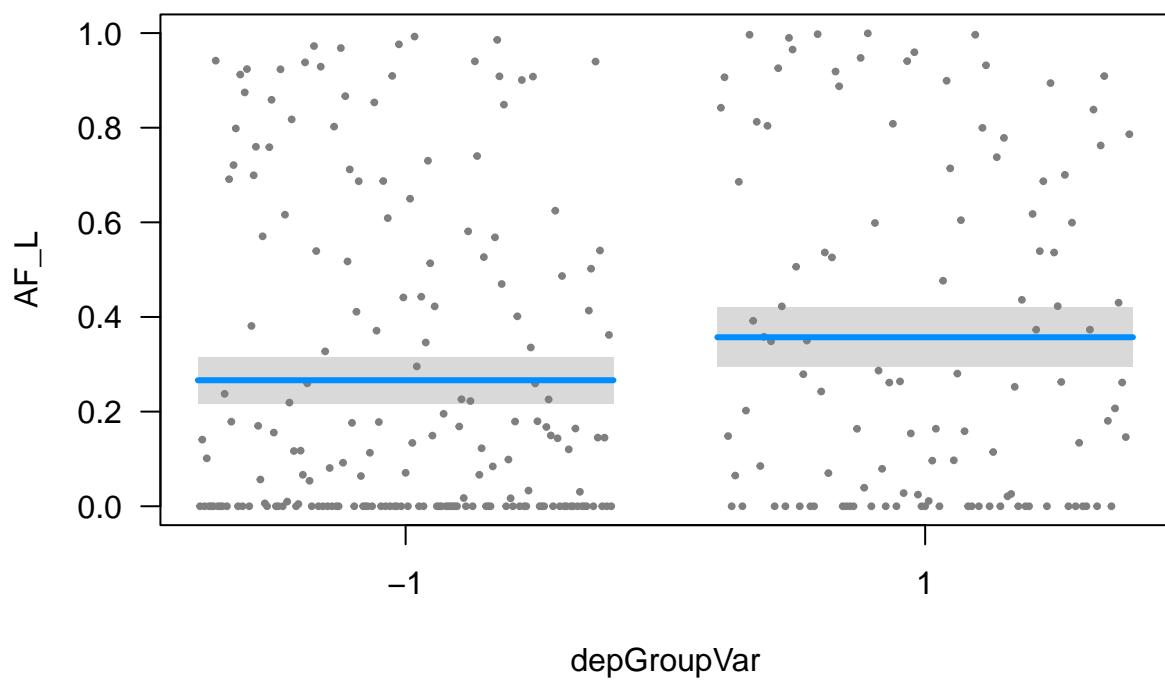
## 18      VOF_R          0
## 19      CNV_R        0.009
## 20      AR_L         0.01
## 21      AR_R        0.001
## 22      CBT_L          0
## 23      CBT_R        0.013
## 24      CPT_F_L       0.006
## 25      CPT_F_R       0.001
## 26      CPT_O_L       0.002
## 27      CPT_P_R       0.002
## 28      CS_A_L        0.004
## 29      CS_A_R        0.014
## 30      CS_P_L        0.039
## 31      CS_S_L        0.003
## 32      CS_S_R        0.001
## 33      CST_L         0.039
## 34      CST_R         0.003
## 35      DRTT_L        0.012
## 36      DRTT_R          0
## 37      F_L           0.01
## 38      ML_R          0.001
## 39      OR_L          0.019
## 40      OR_R          0.017
## 41      RST_L          0.01
## 42      RST_R          0
## 43      TR_A_L        0.001
## 44      TR_A_R          0
## 45      TR_P_L        0.004
## 46      TR_P_R        0.002
## 47      TR_S_L        0.004
## 48      TR_S_R          0
## 49      CC            0.015

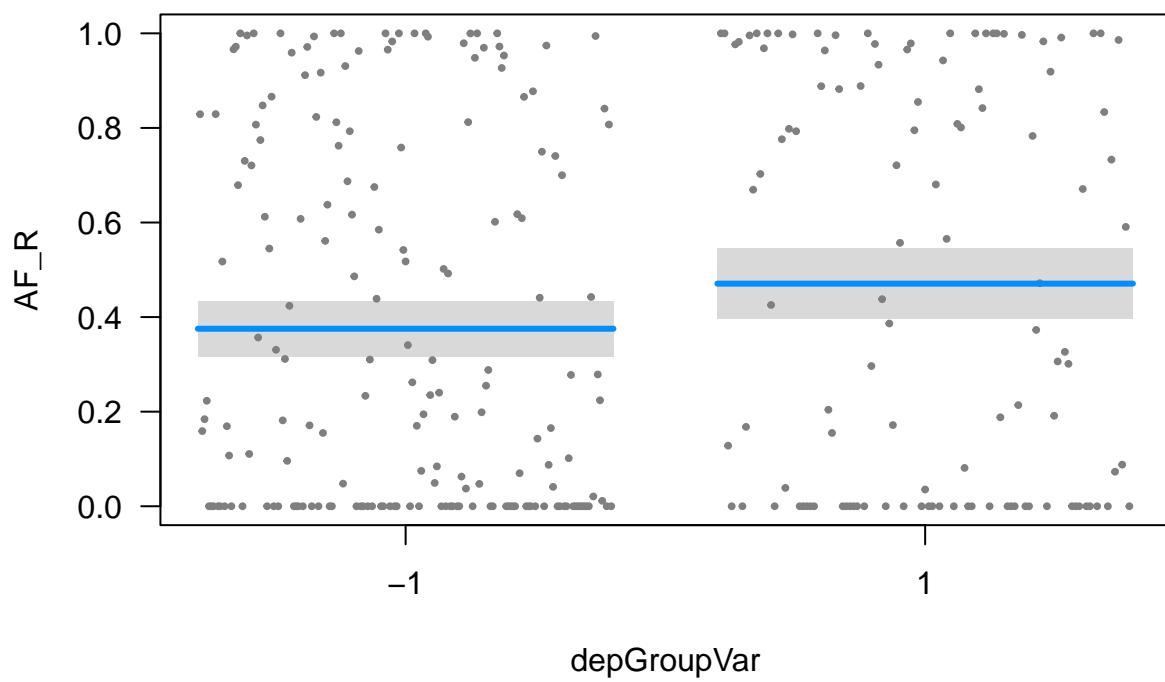
##### repeat above analysis but use only the first instance of each EMPI, sorted by date
df_unique_empi <- dep_and_healthy_groups_for_ICD_analysis %>%
  group_by(EMPI) %>%
  arrange(EXAM_DATE) %>%
  slice(1) %>%
  ungroup() #n = 300, 115 depressed, 185 healthy

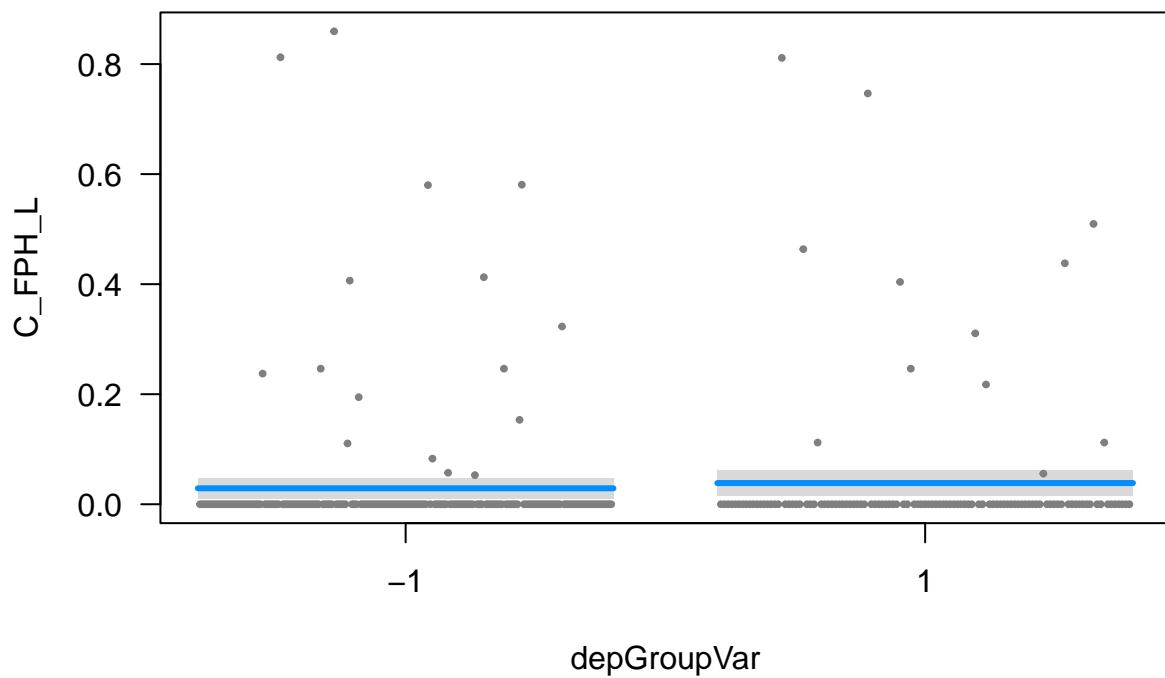
#lm
fascicle_lm_unique <- lapply(fascicle_names, function(x)
{
  lm(substitute(i ~ depGroupVar, list(i = as.name(x))), data = df_unique_empi)
})
names(fascicle_lm_unique) <- fascicle_names

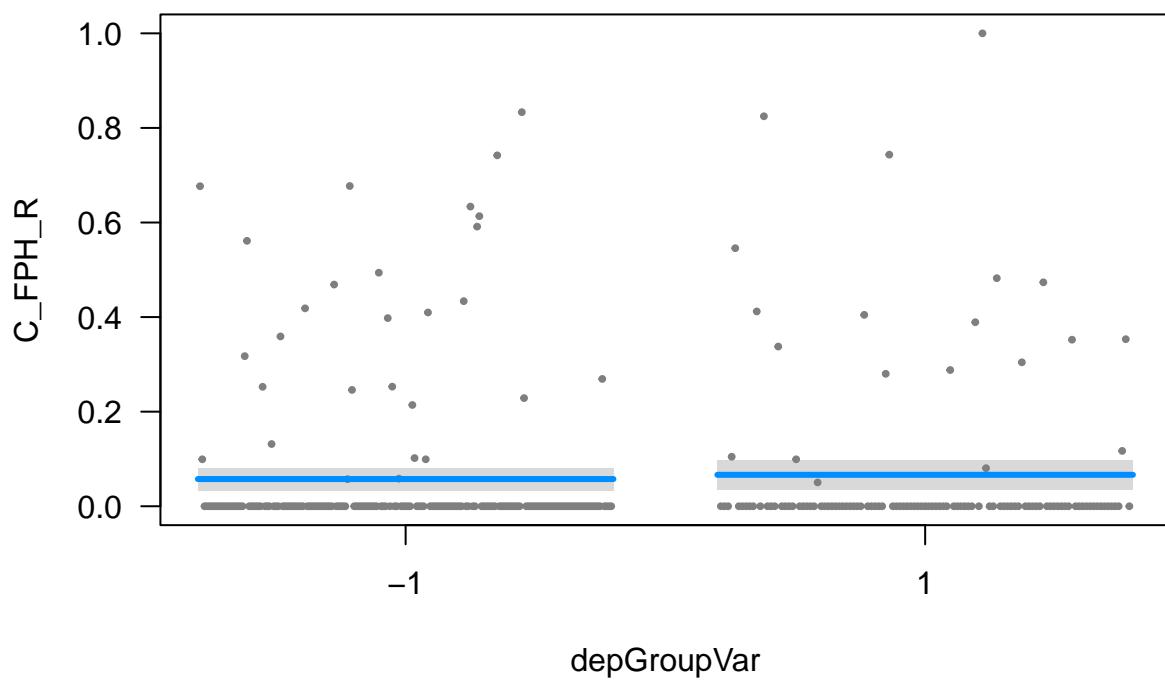
#visreg
sapply(fascicle_lm_unique, visreg)

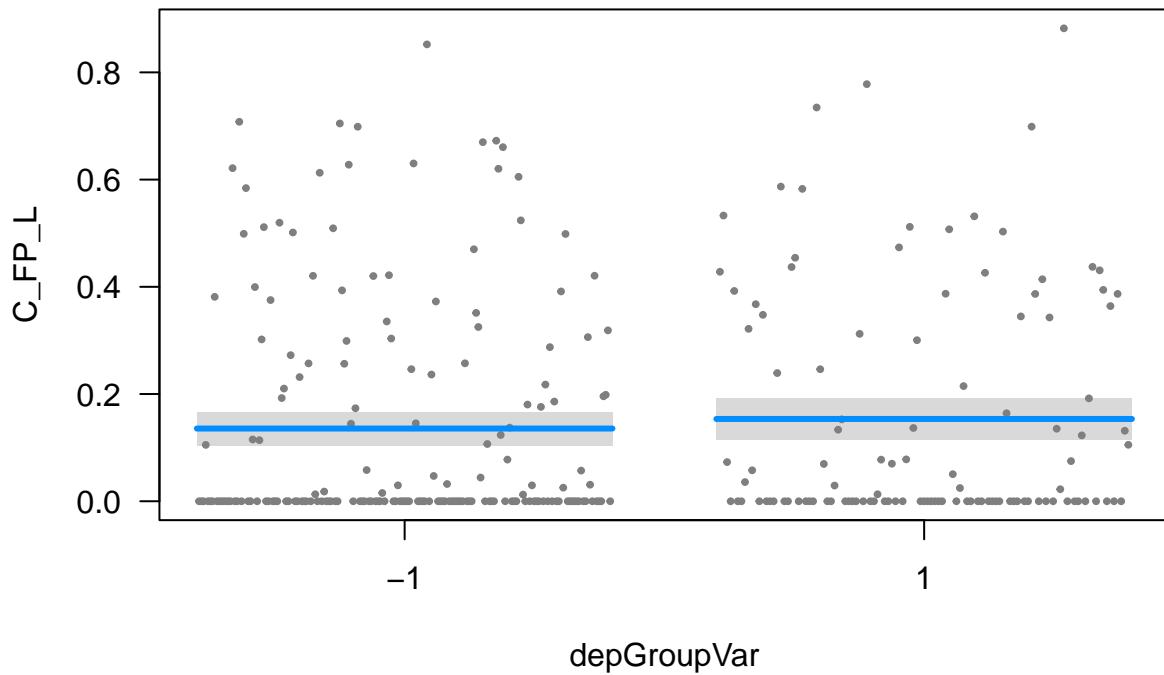
```

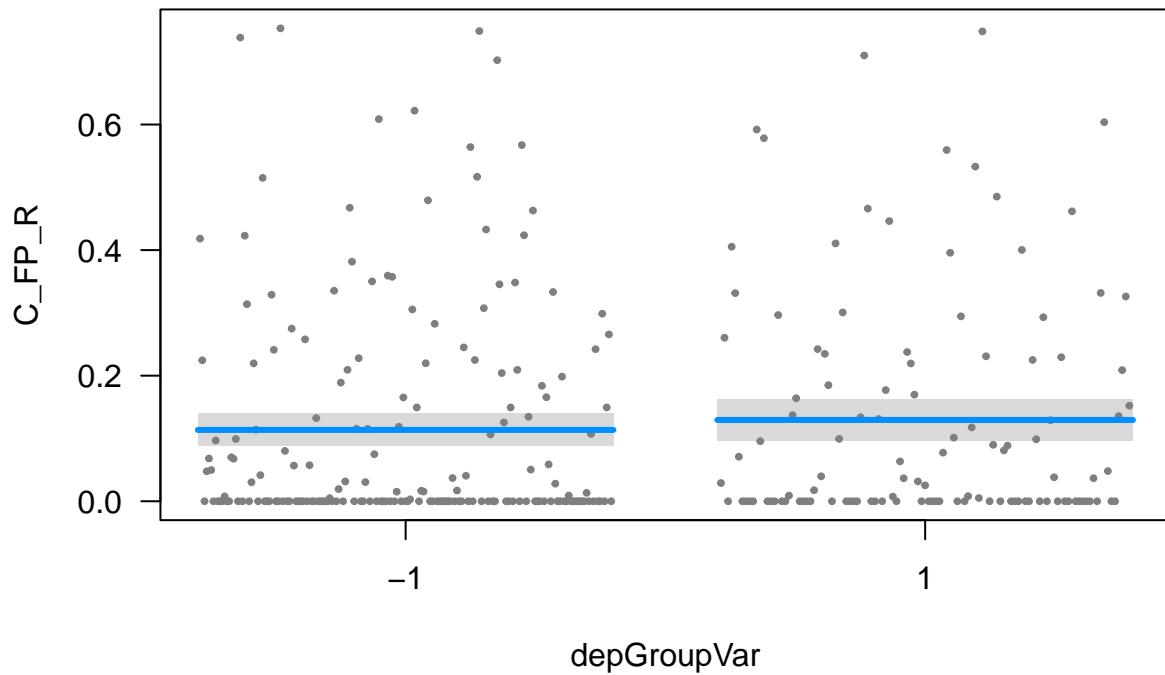


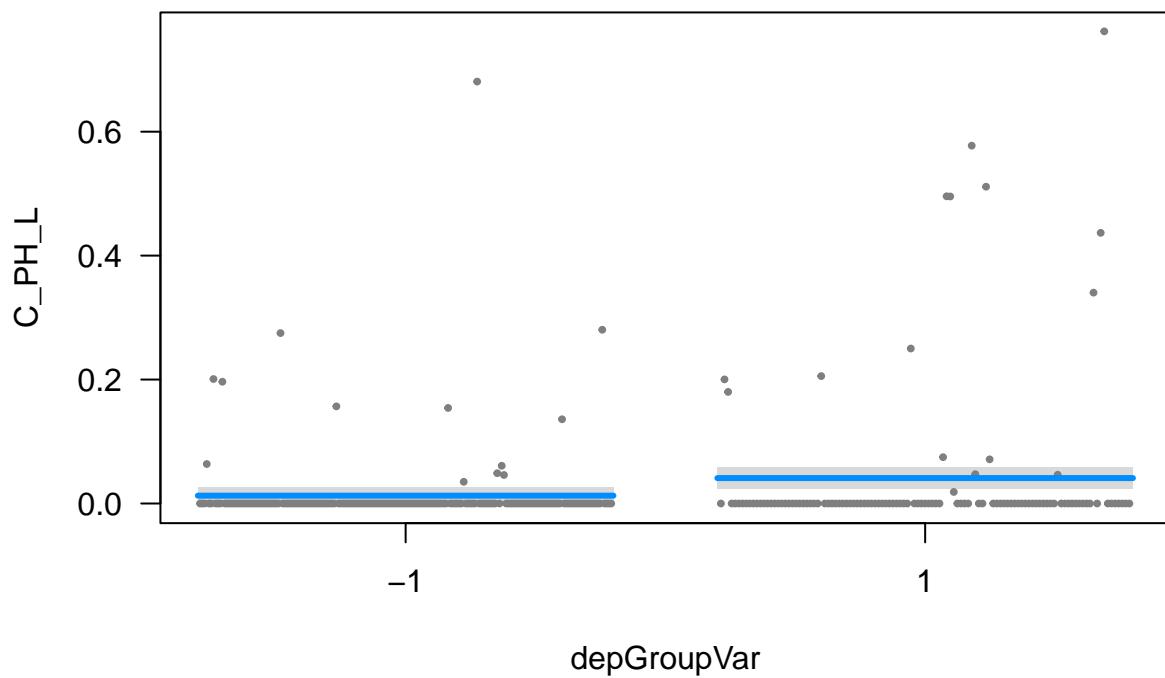


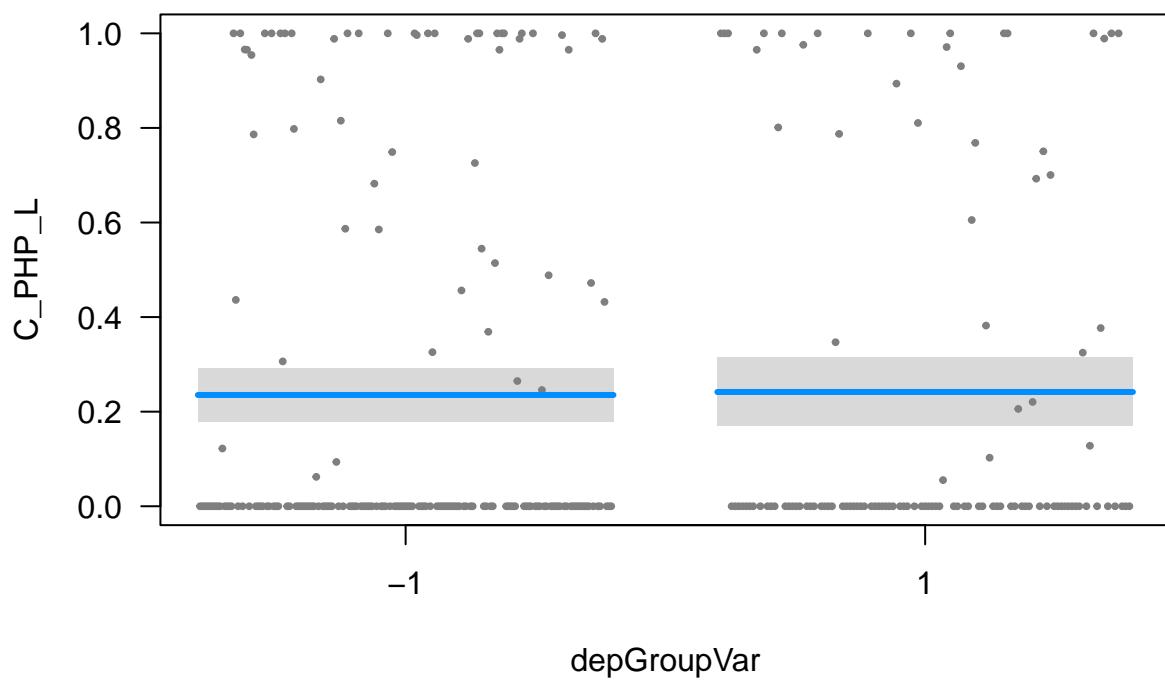


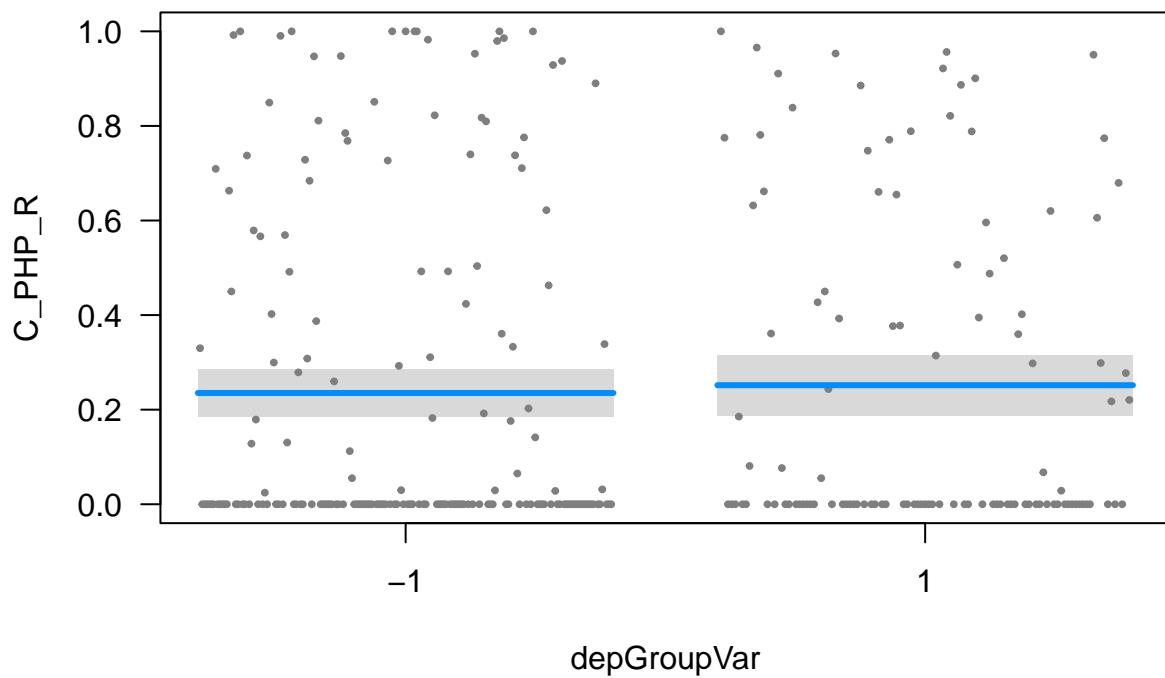


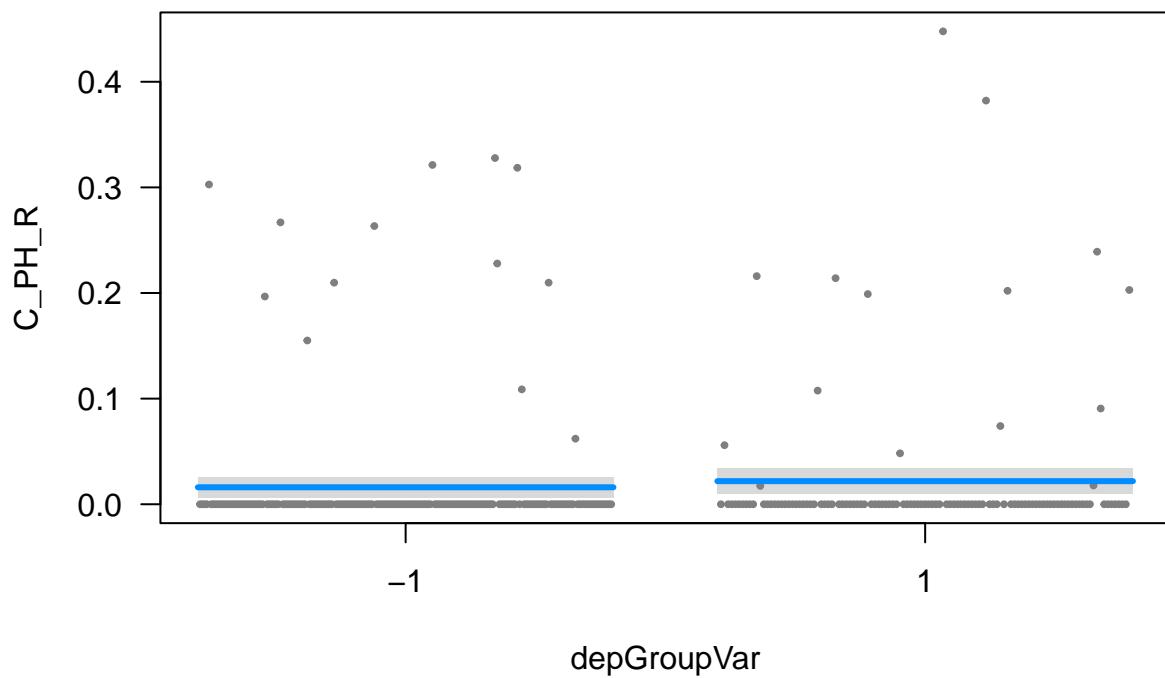


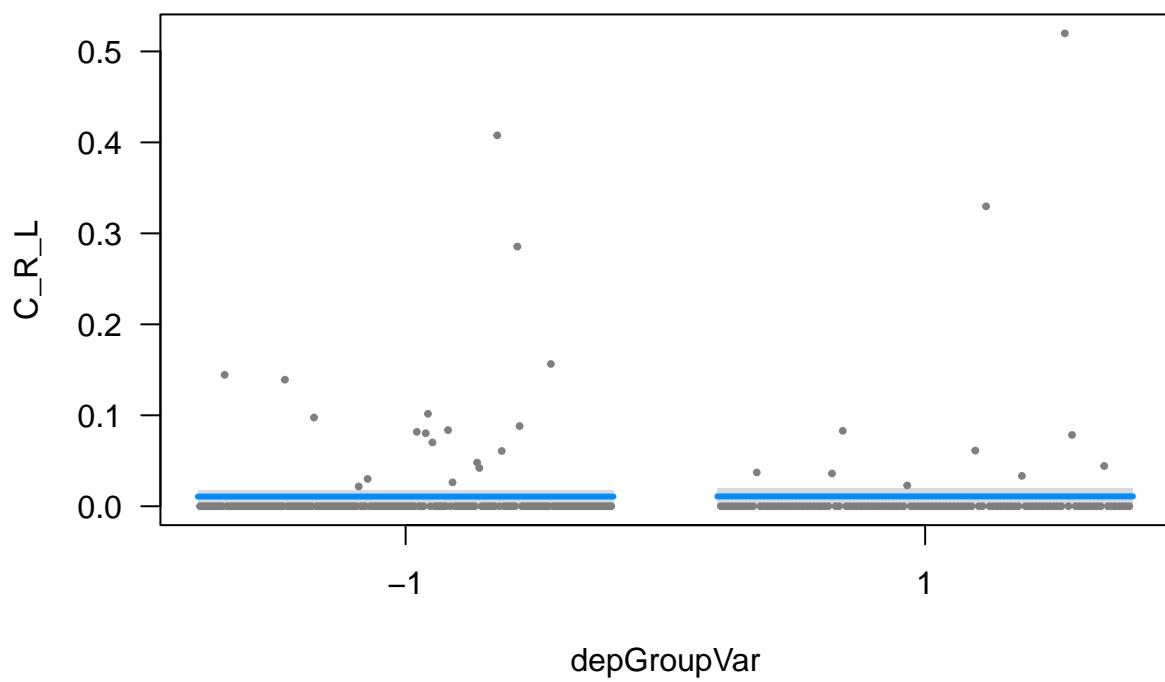


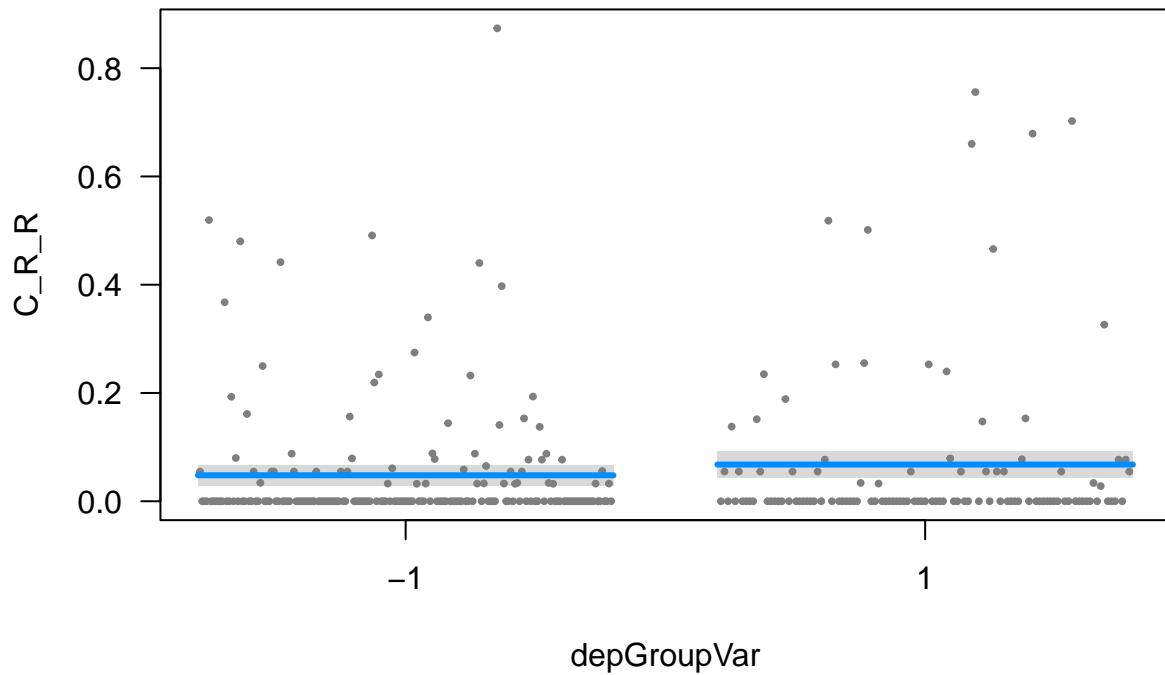


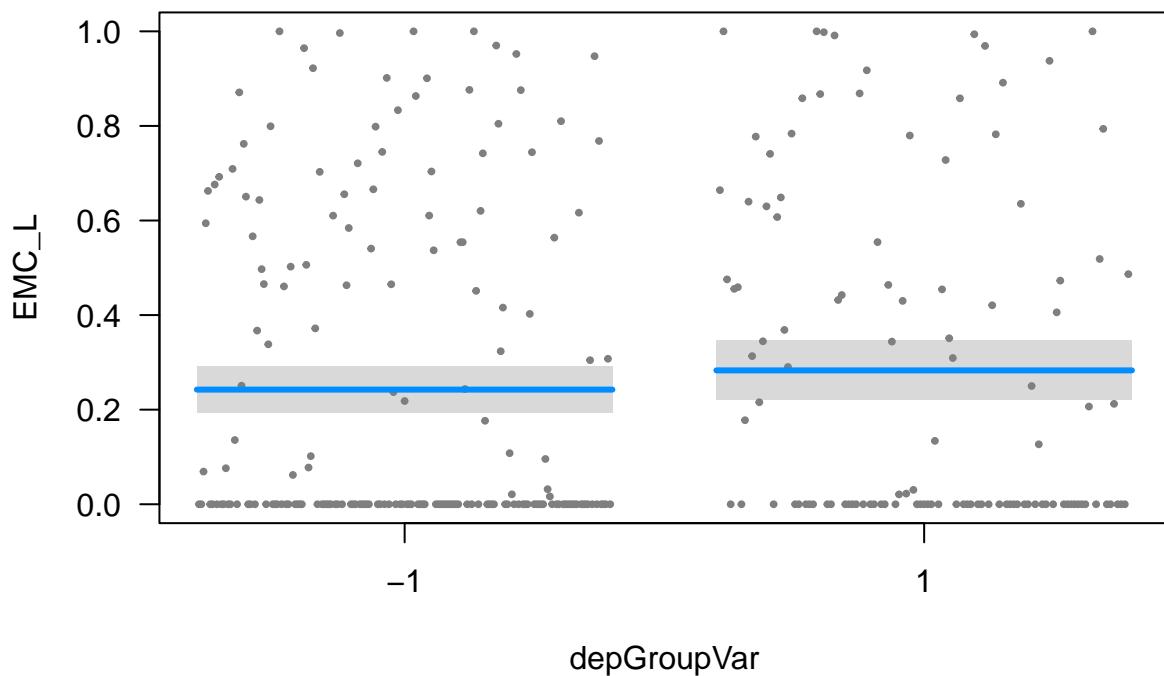


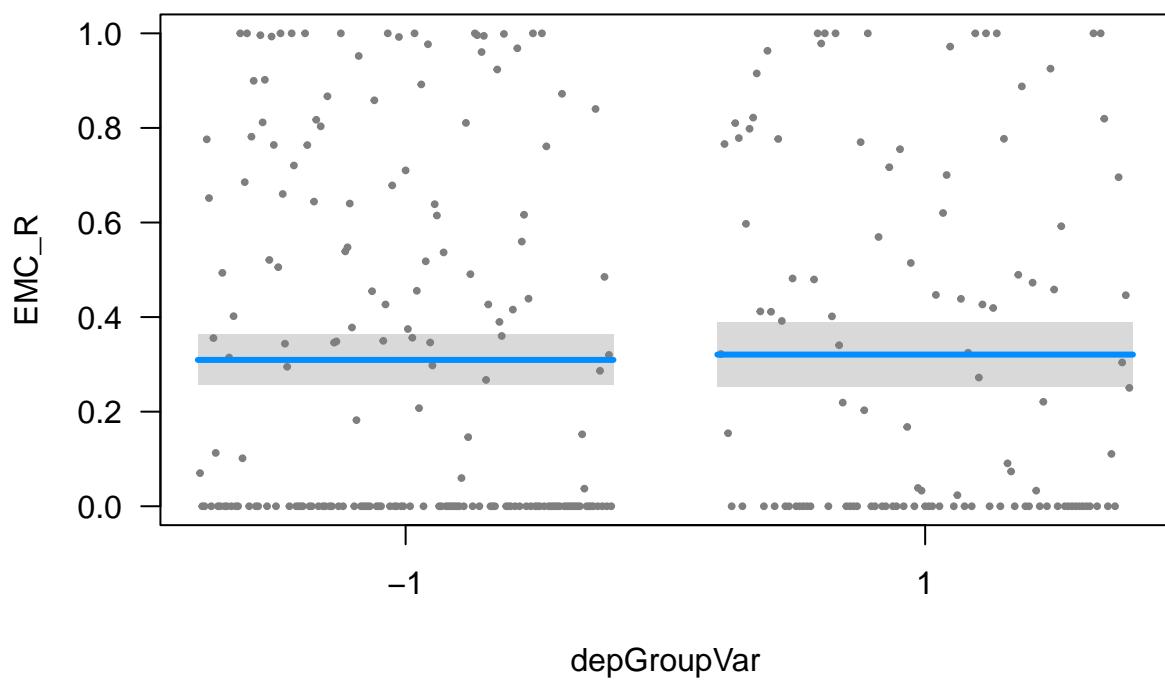


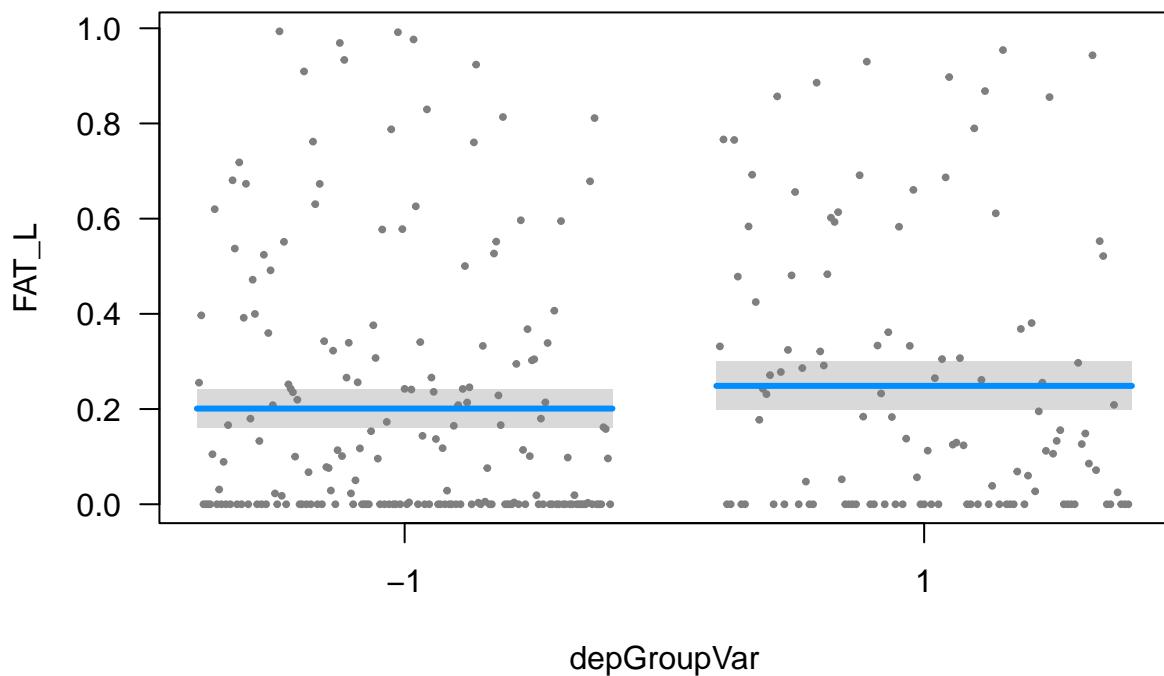


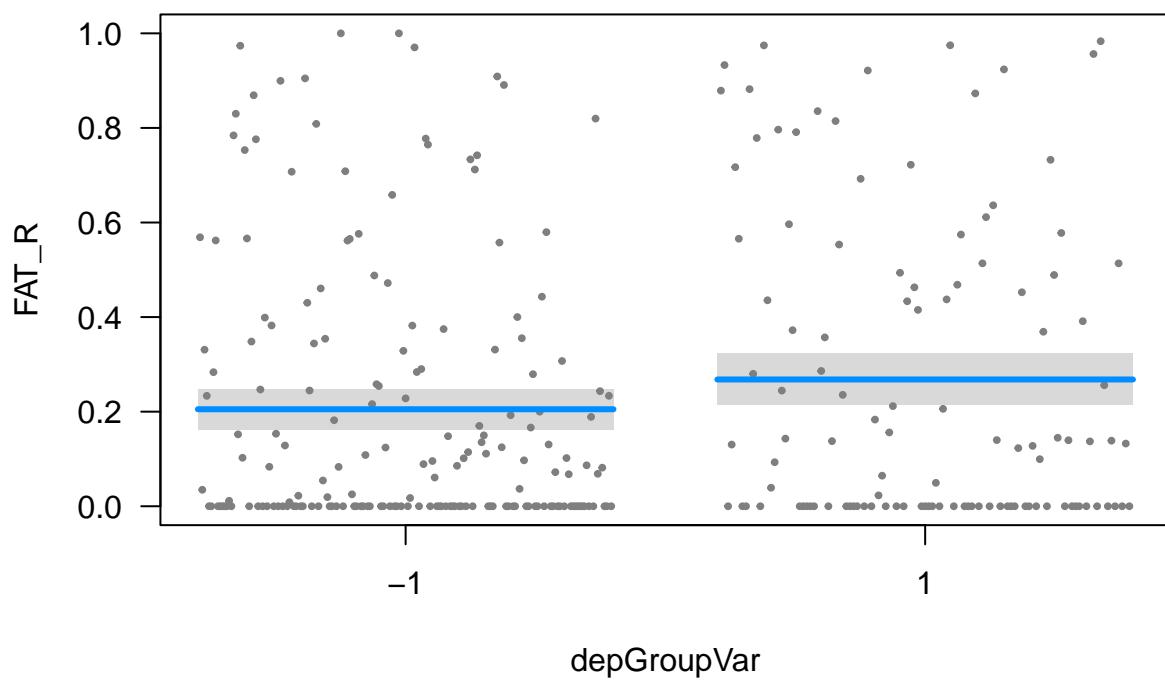


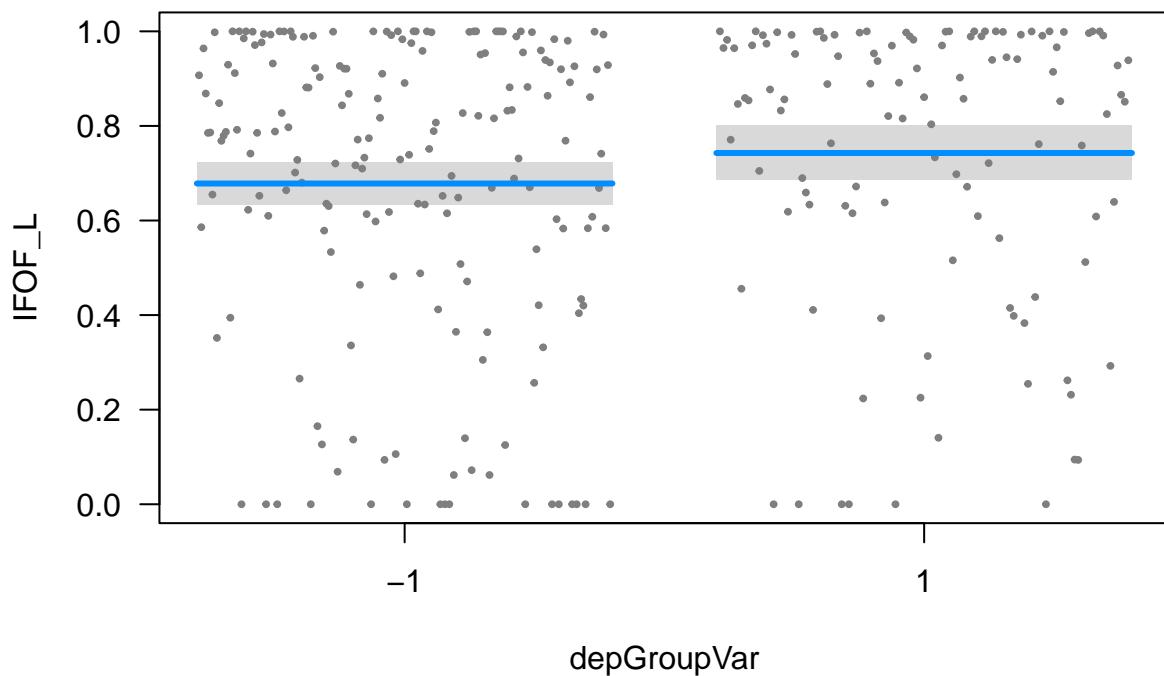


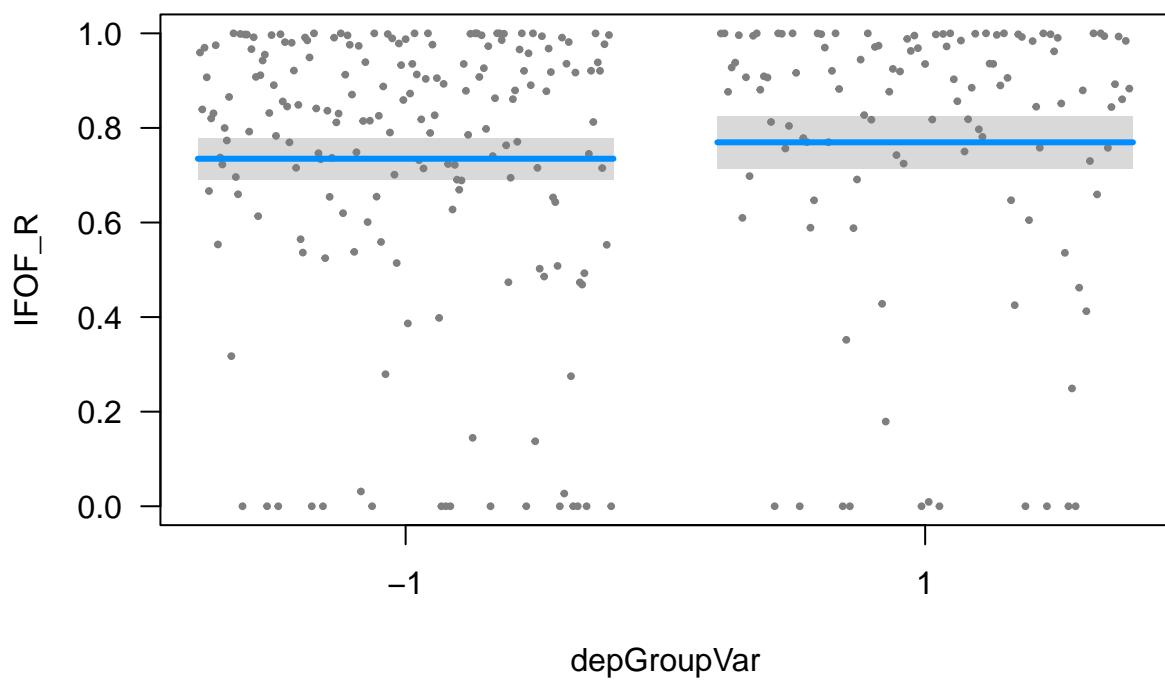


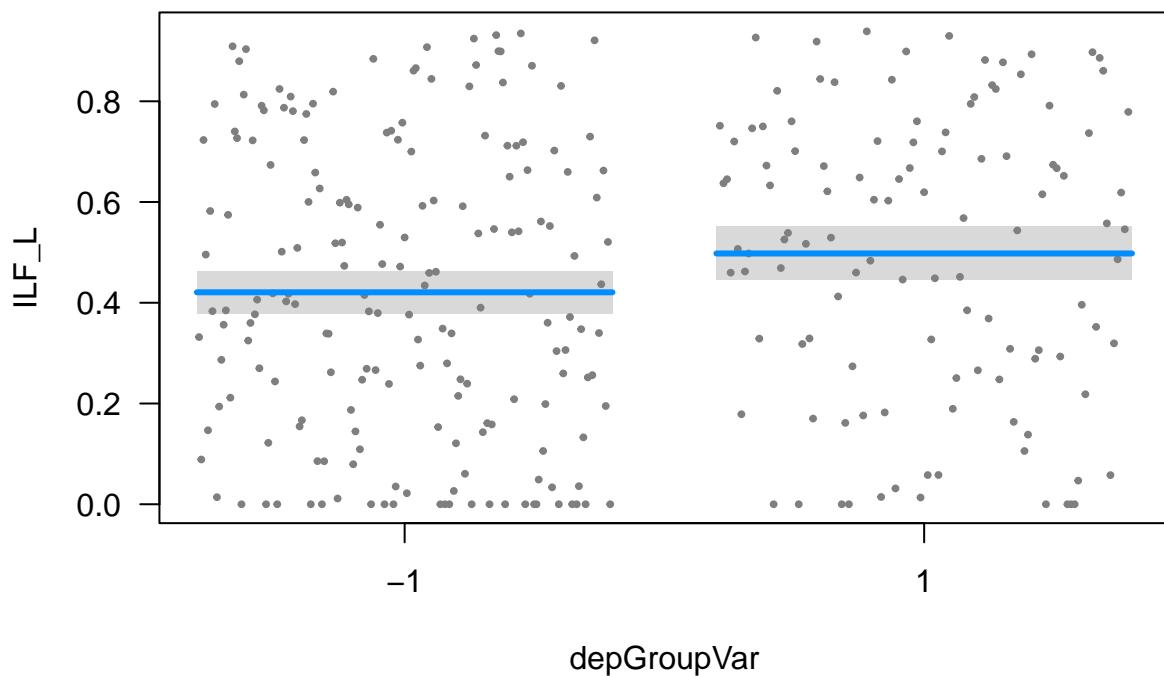


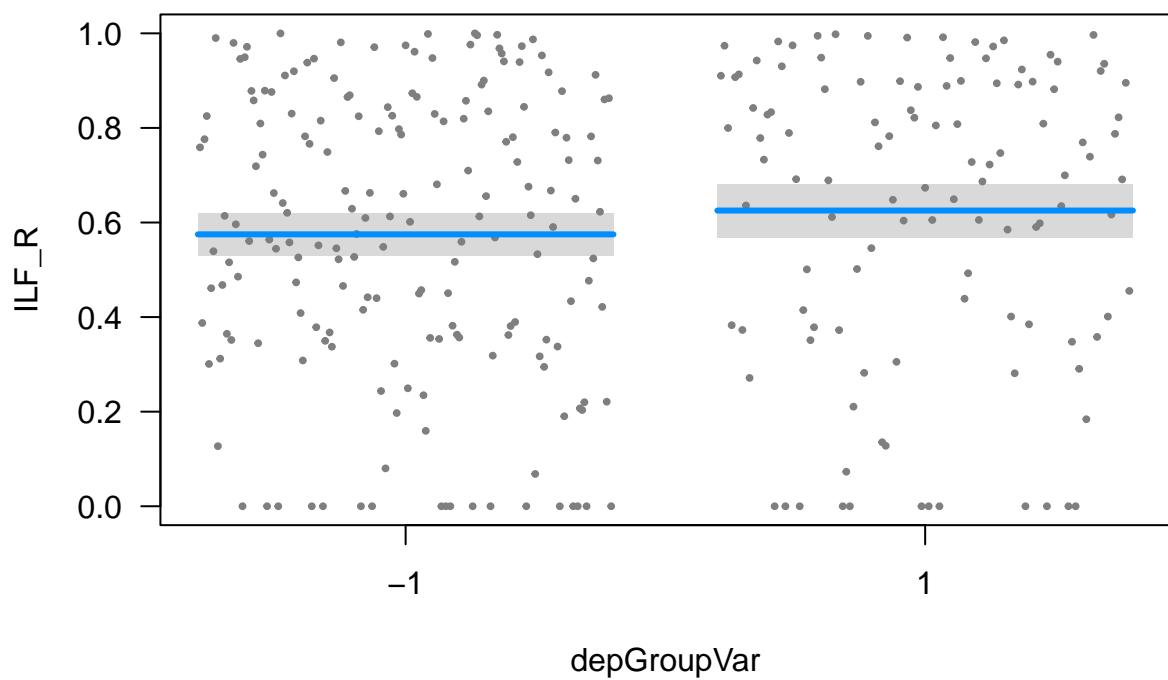


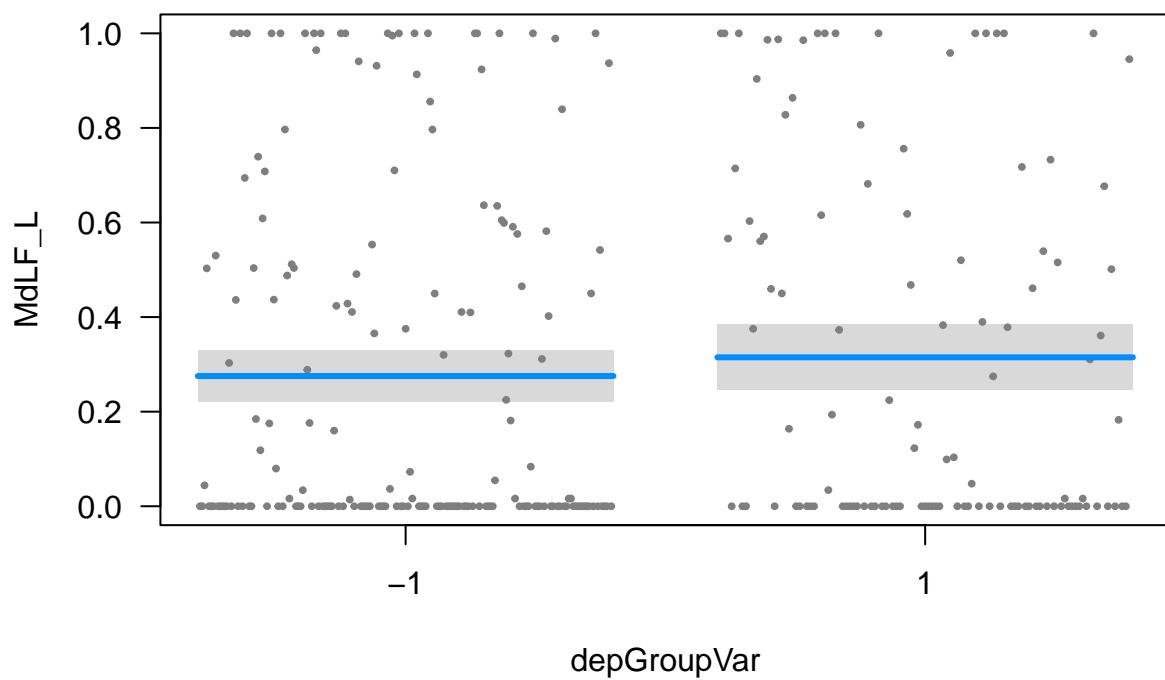


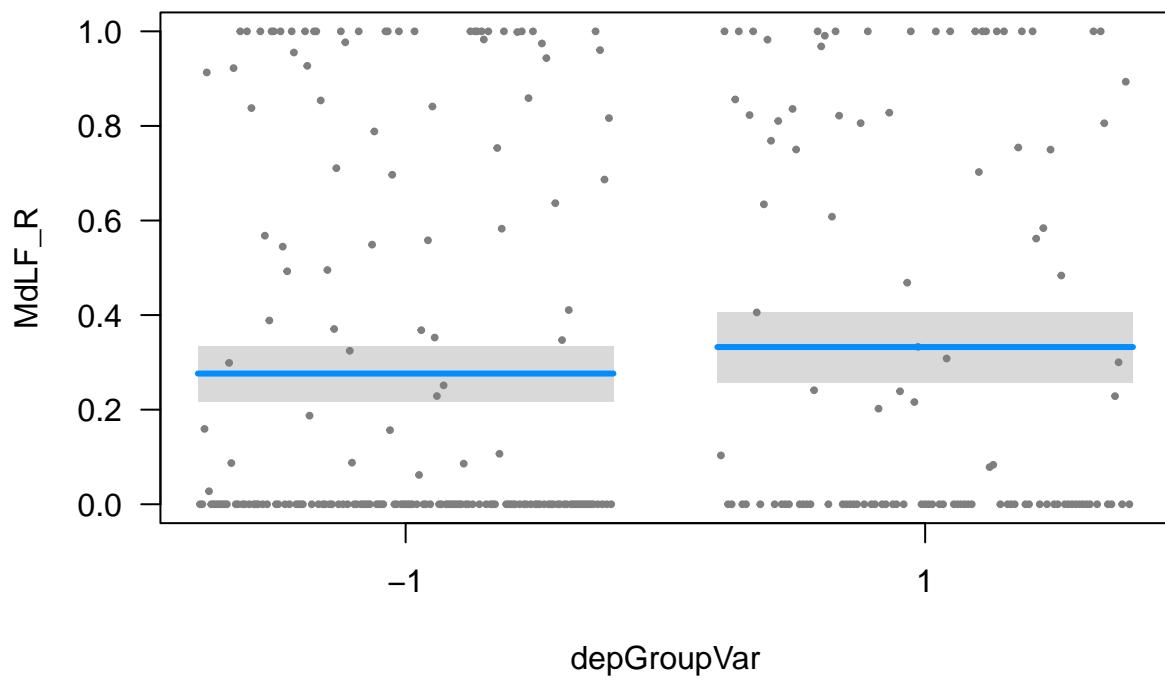


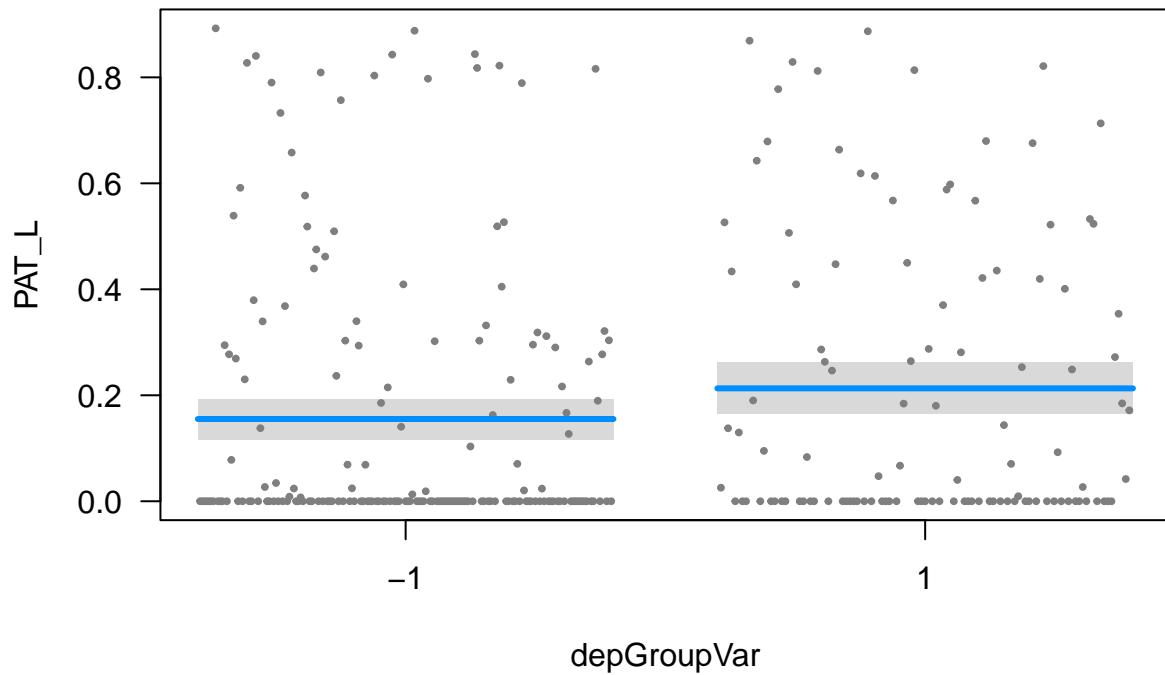


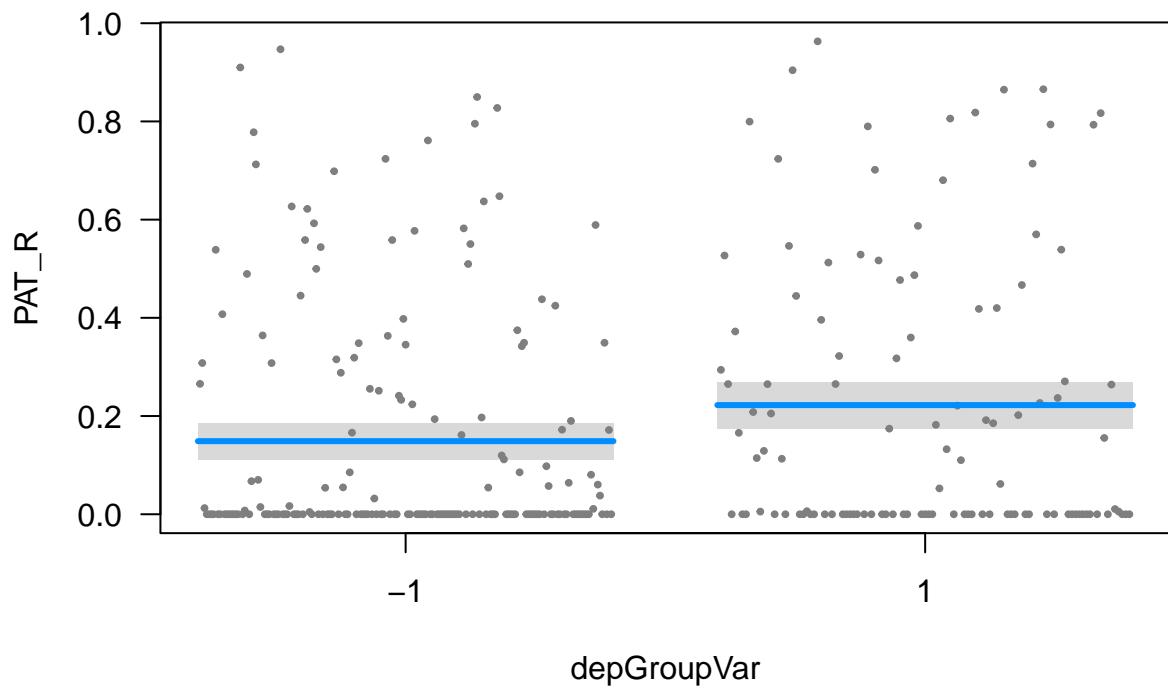


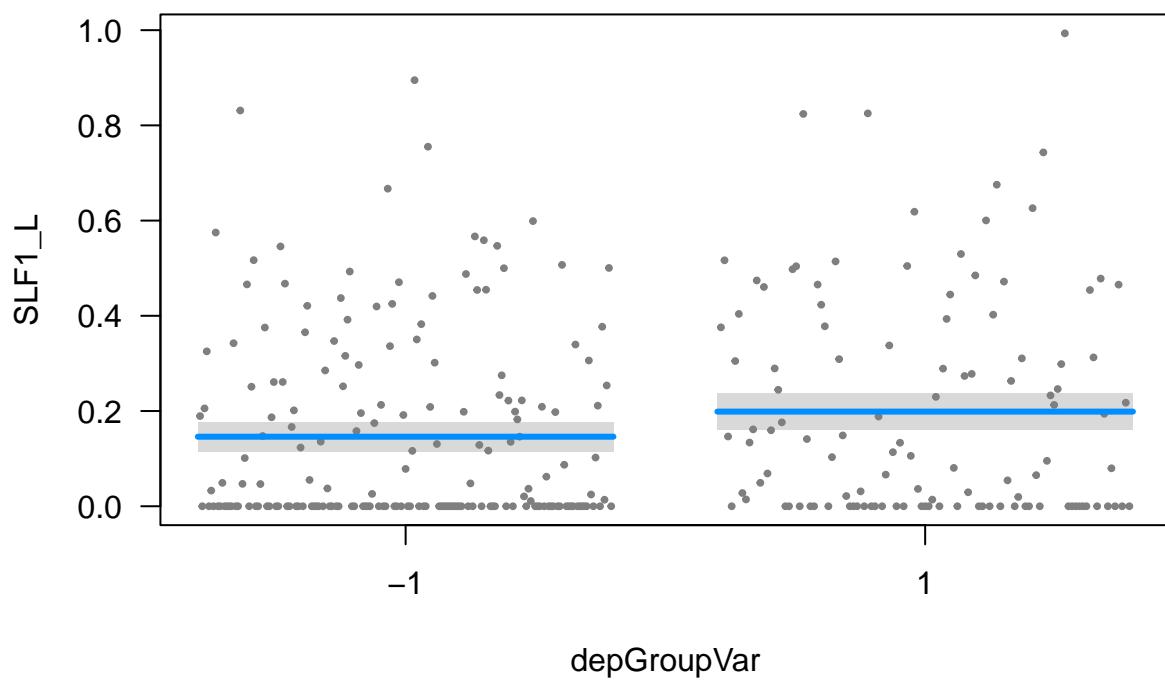


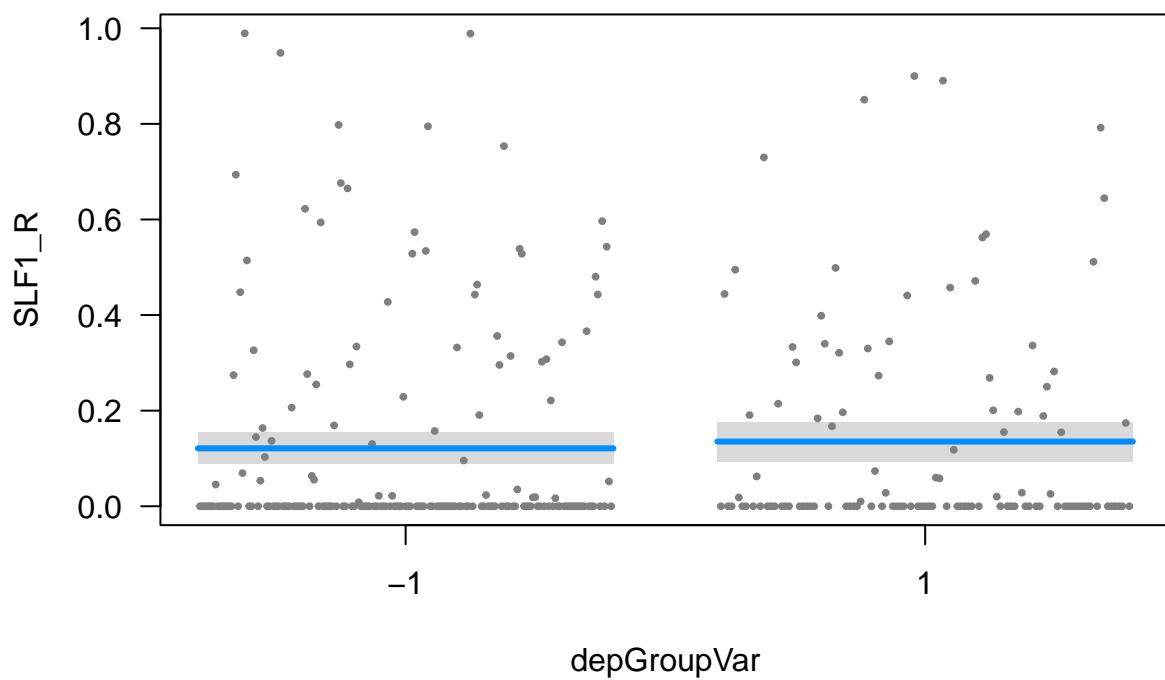


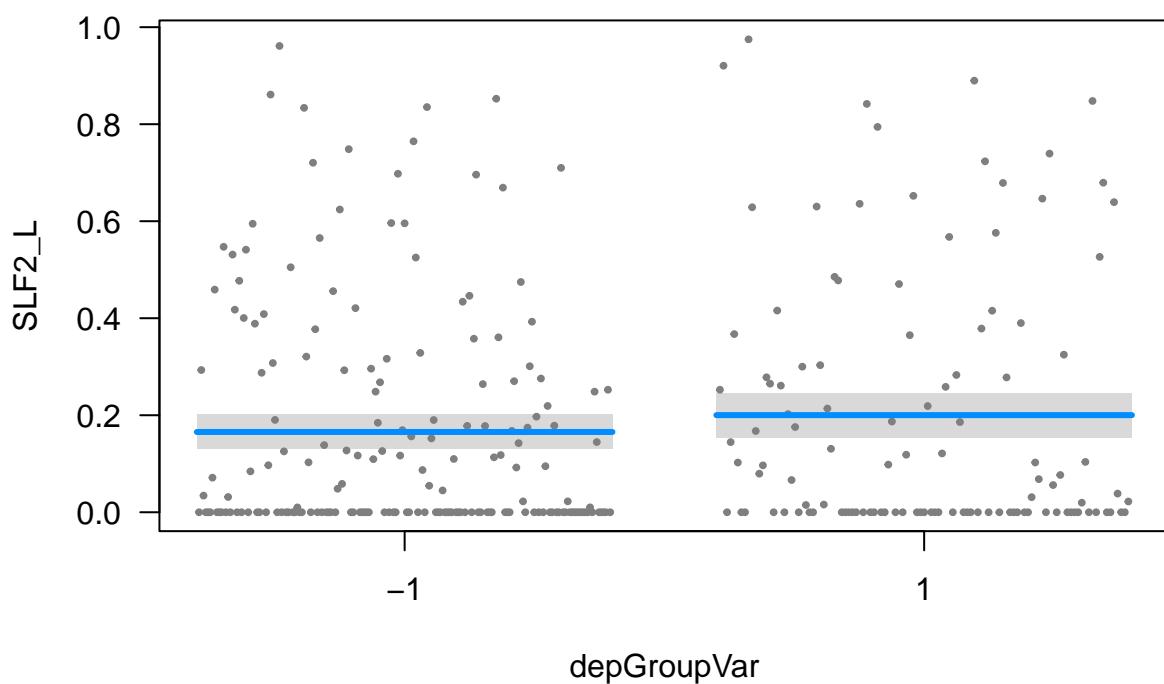


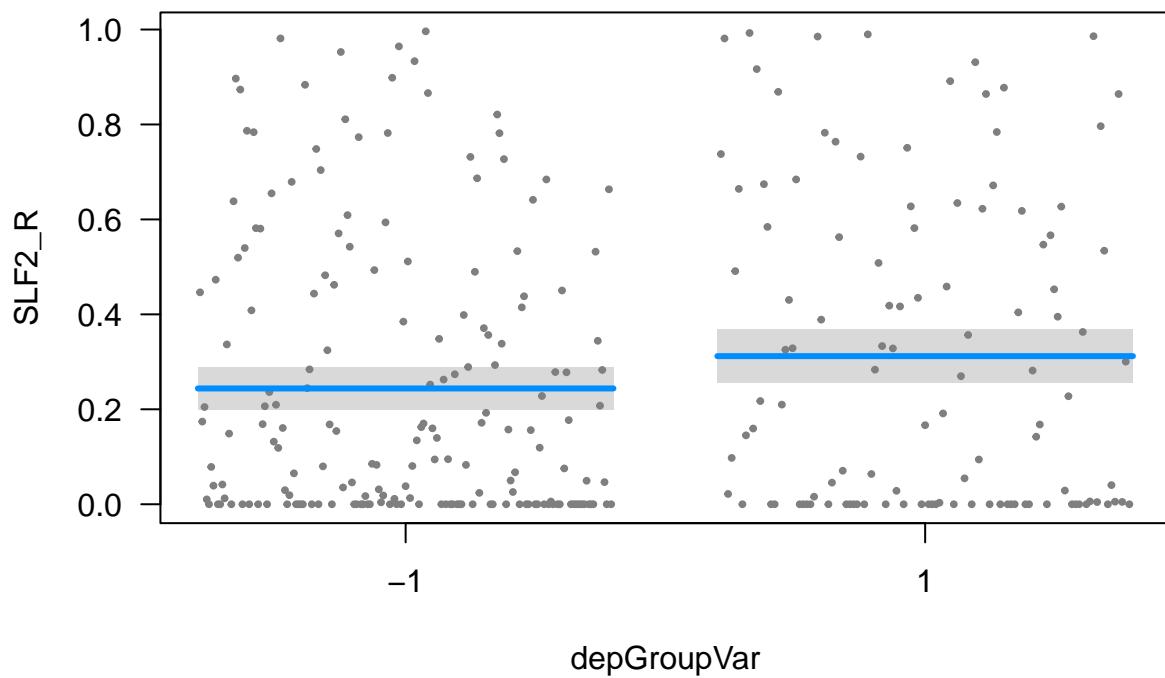


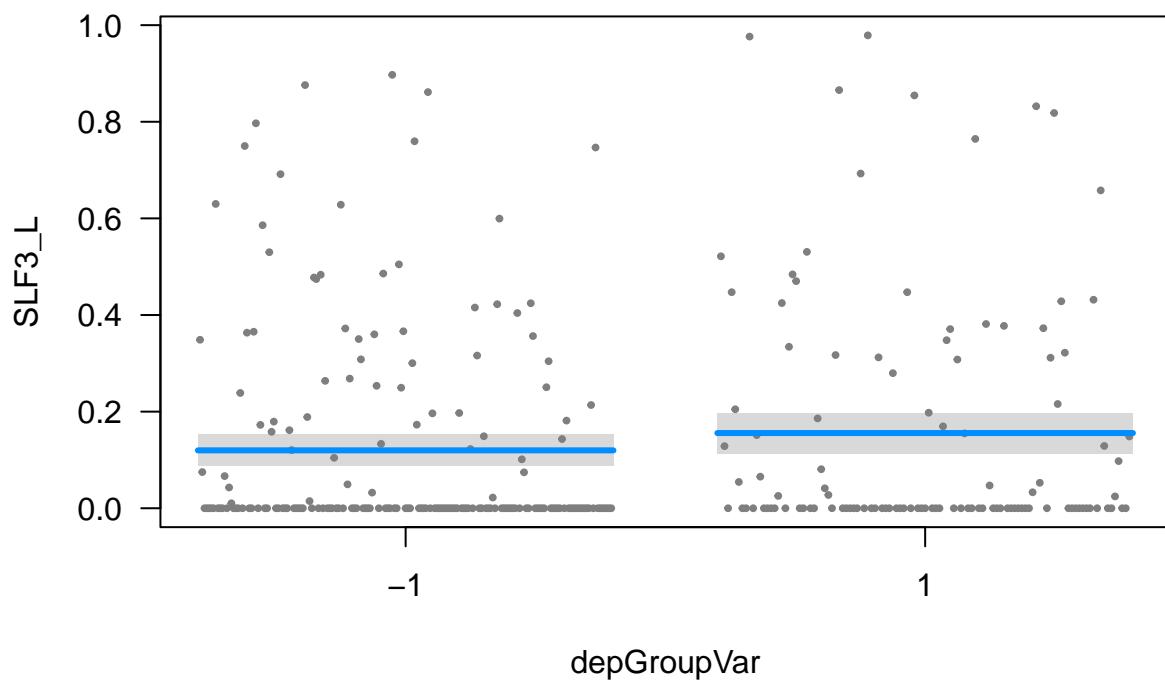


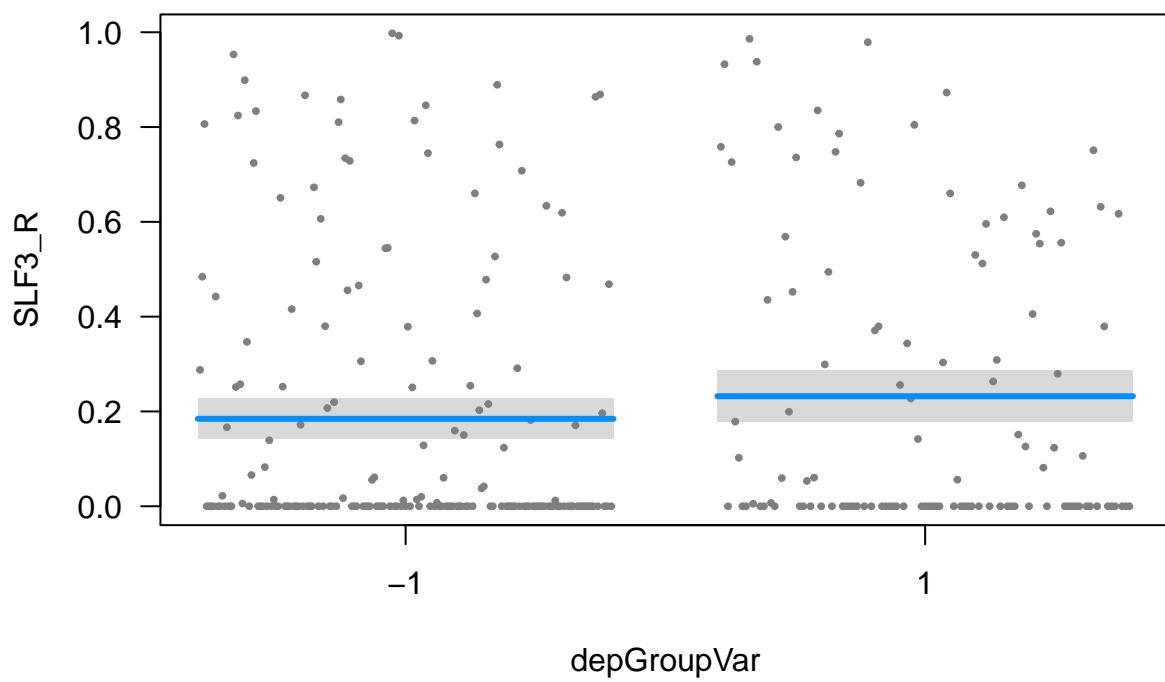


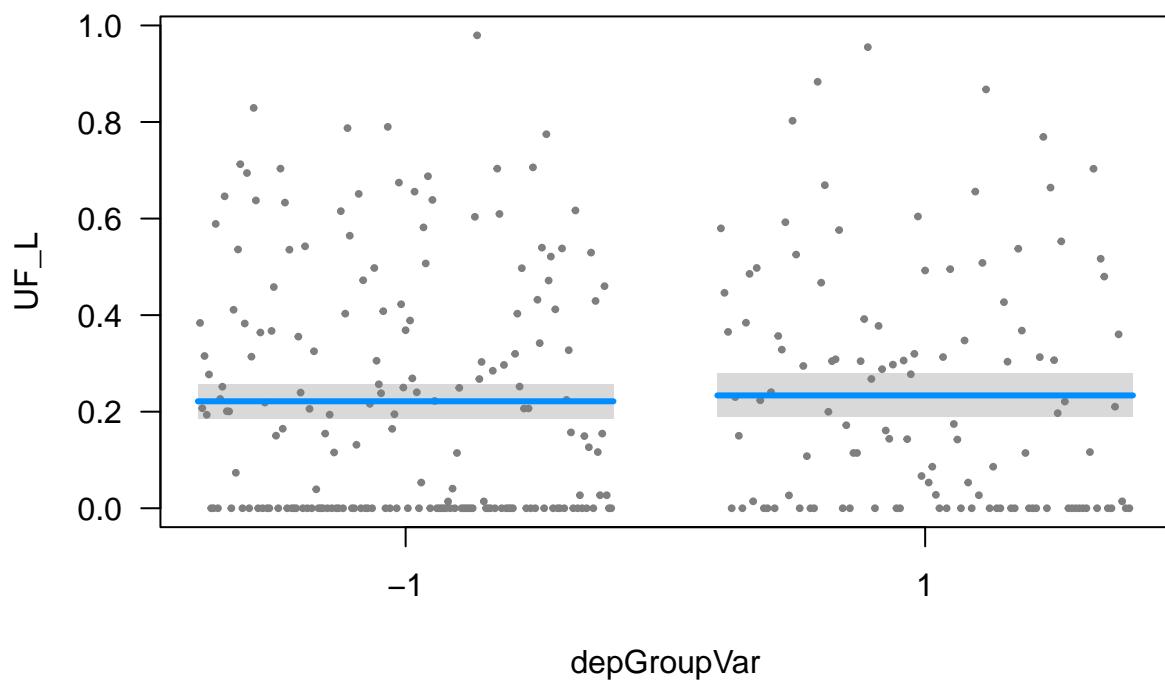


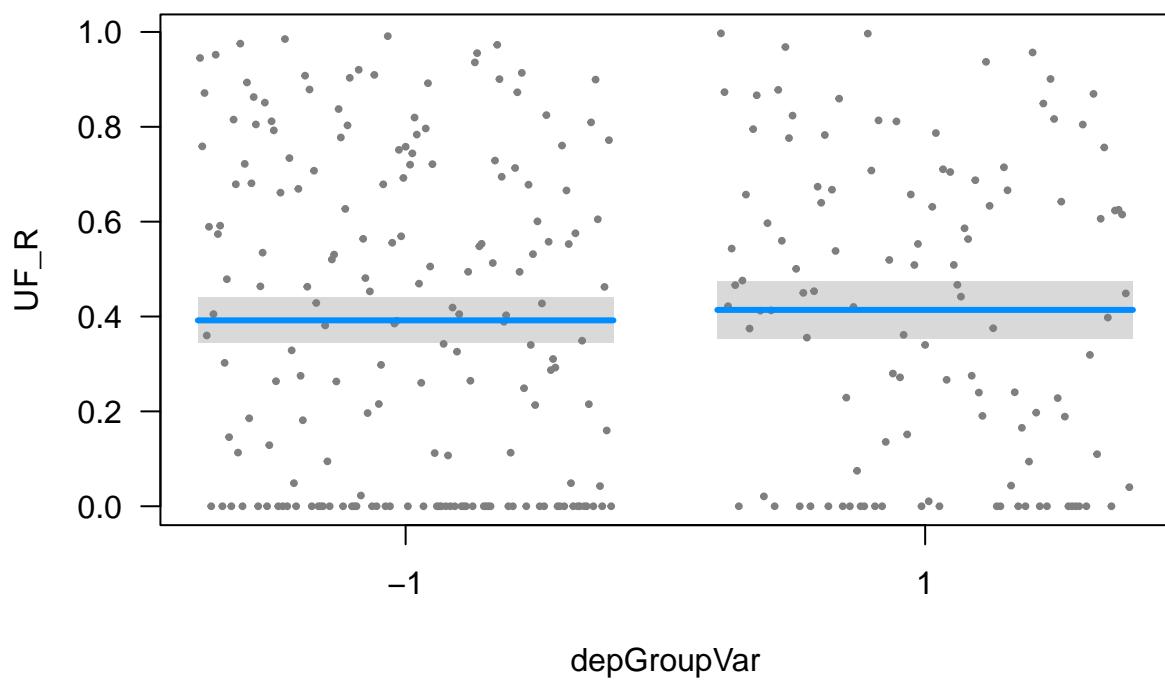


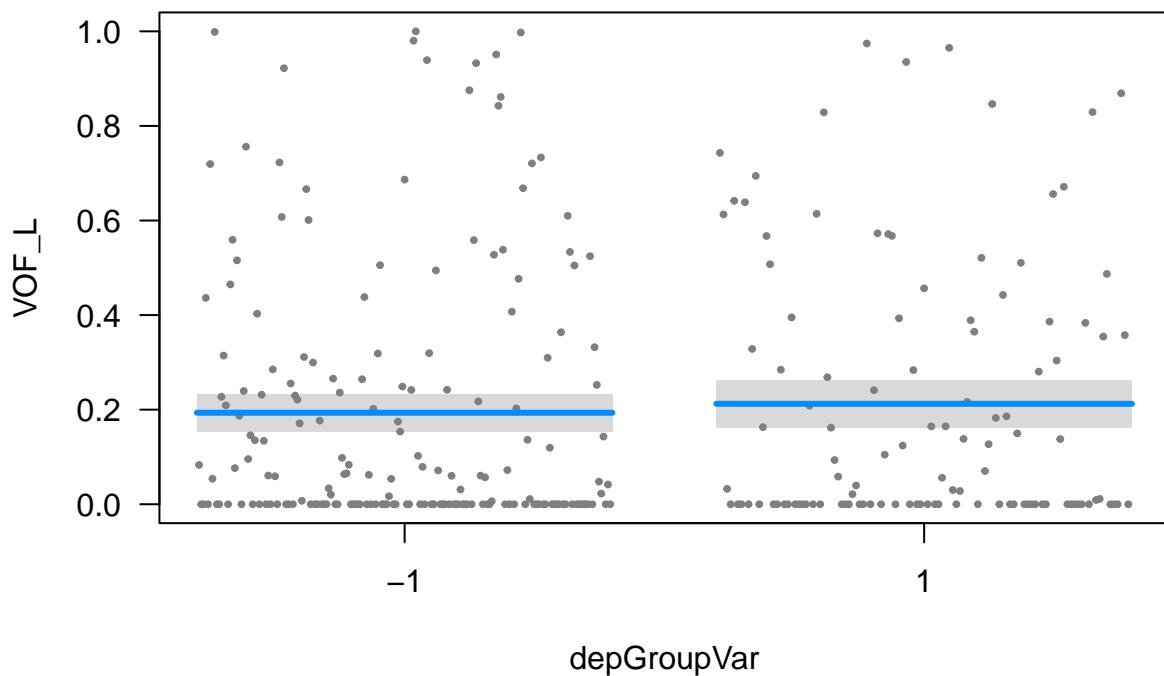


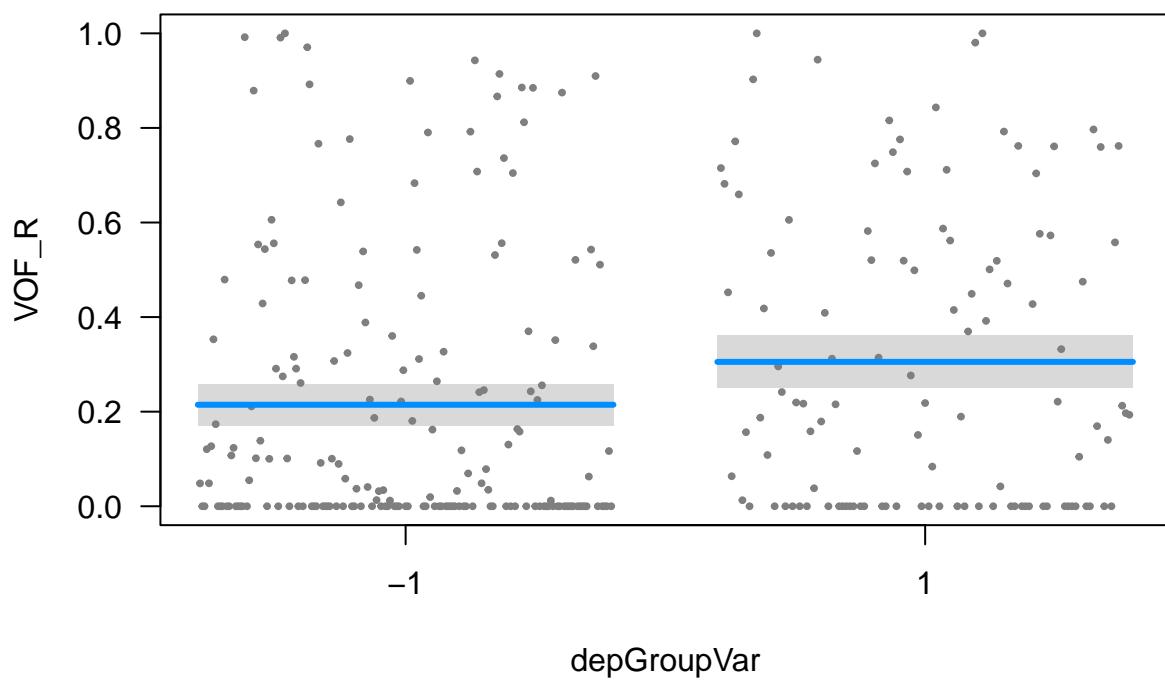


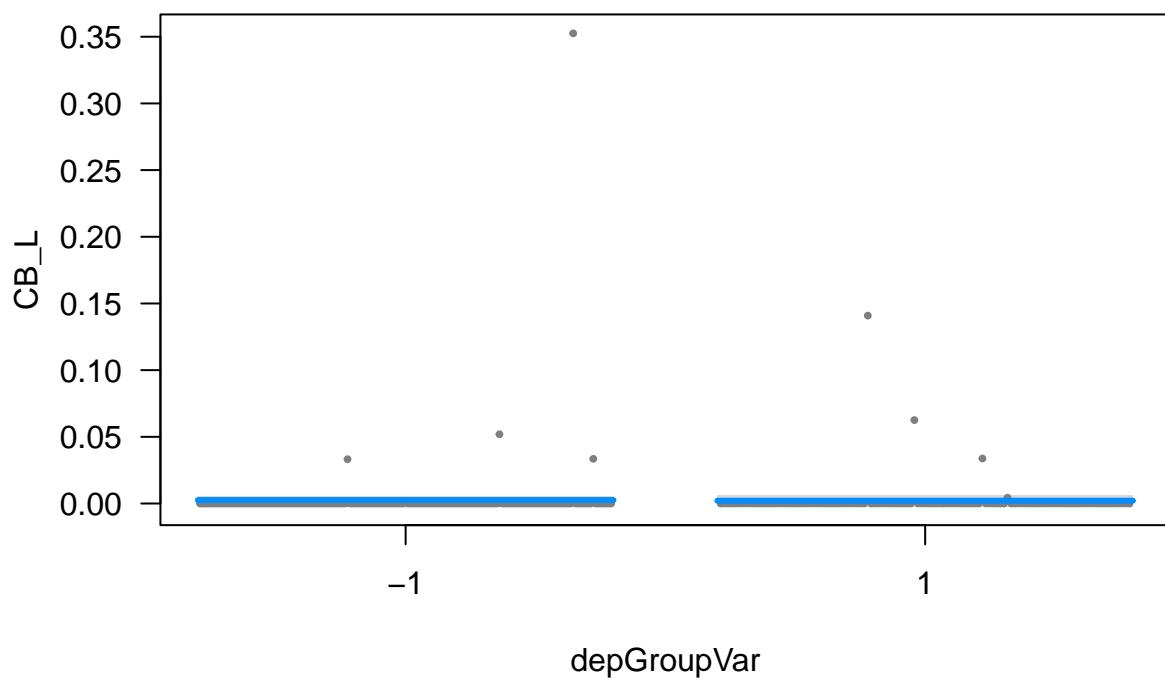


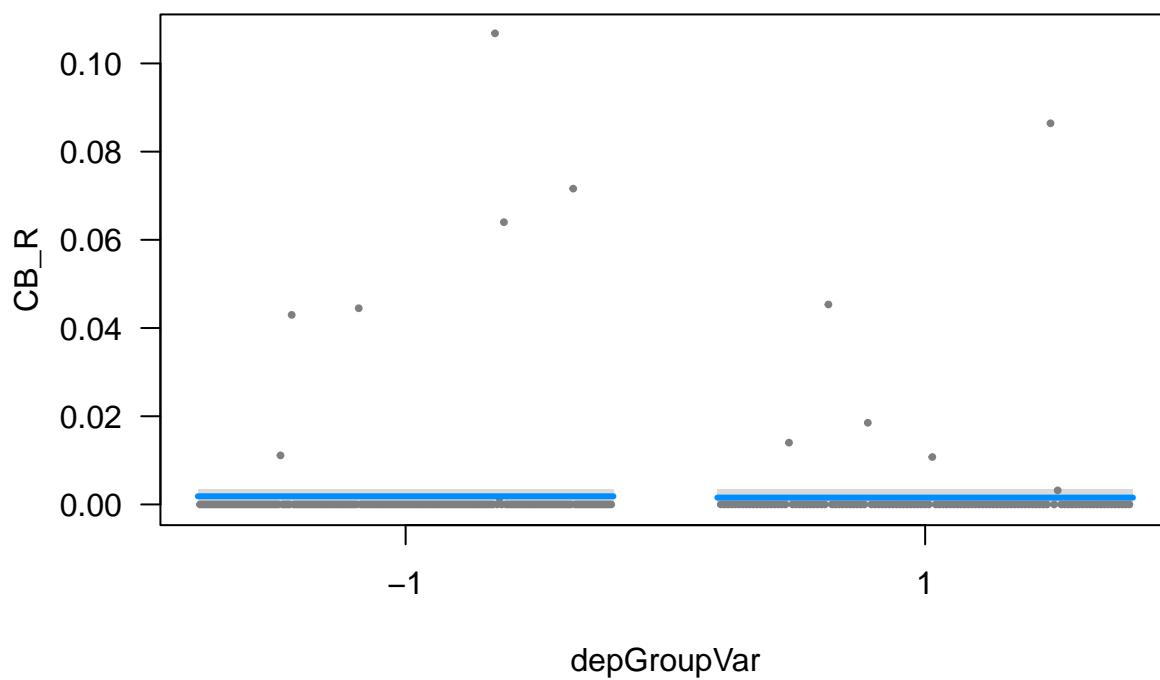


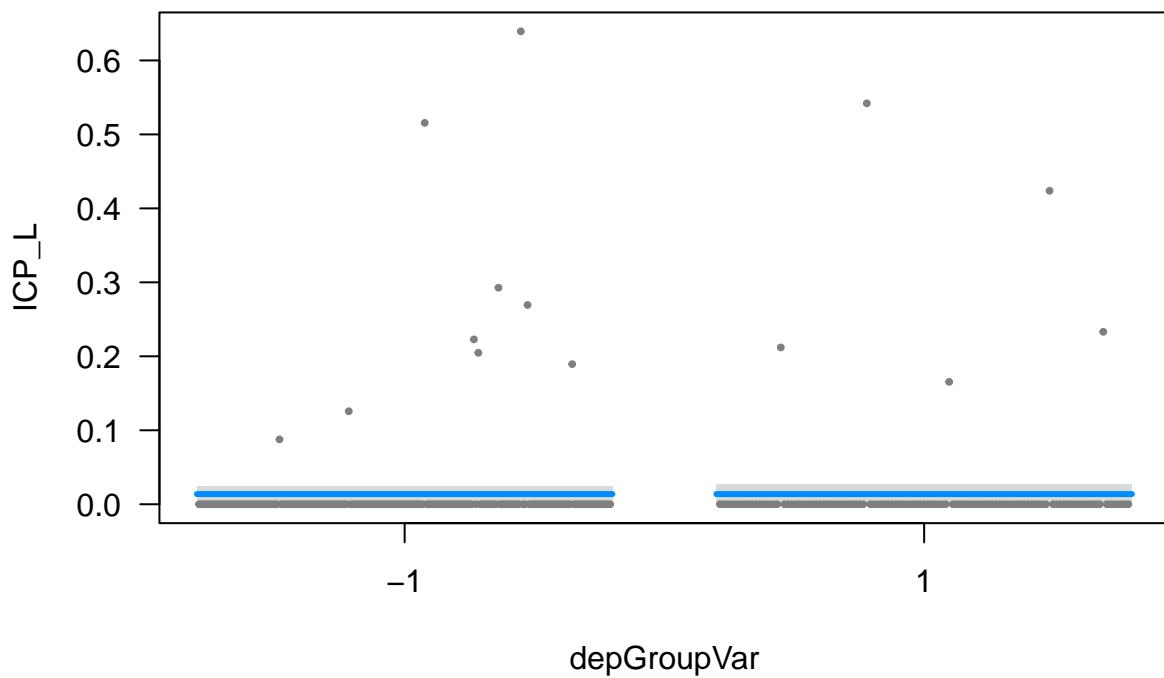


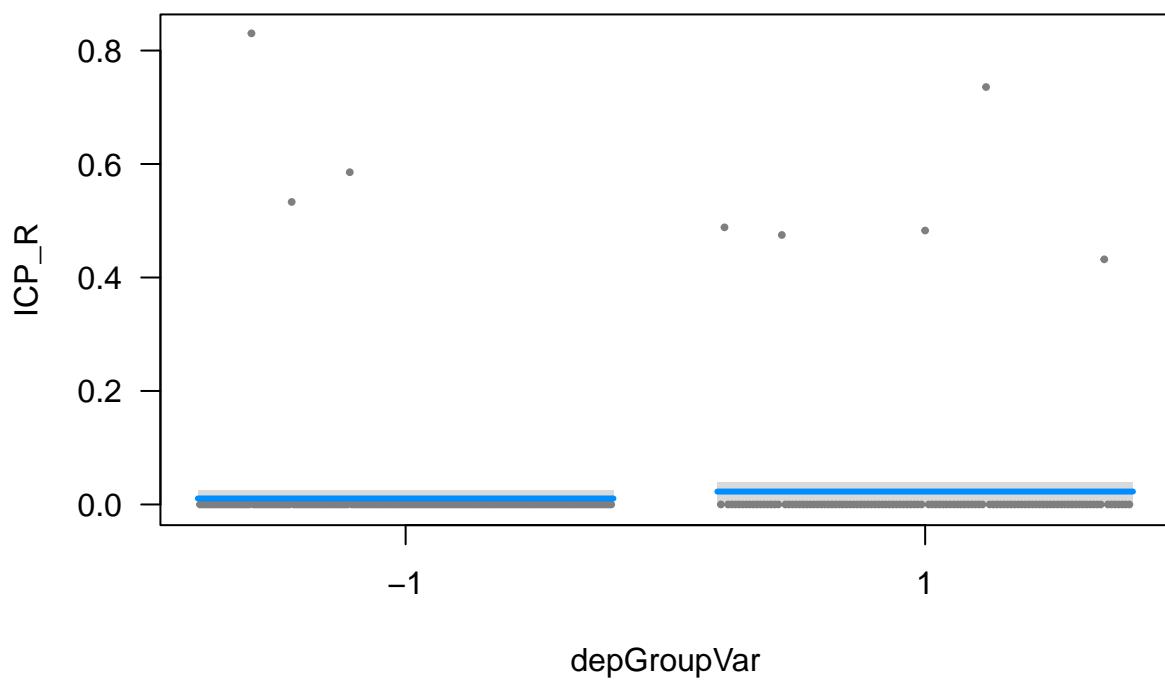


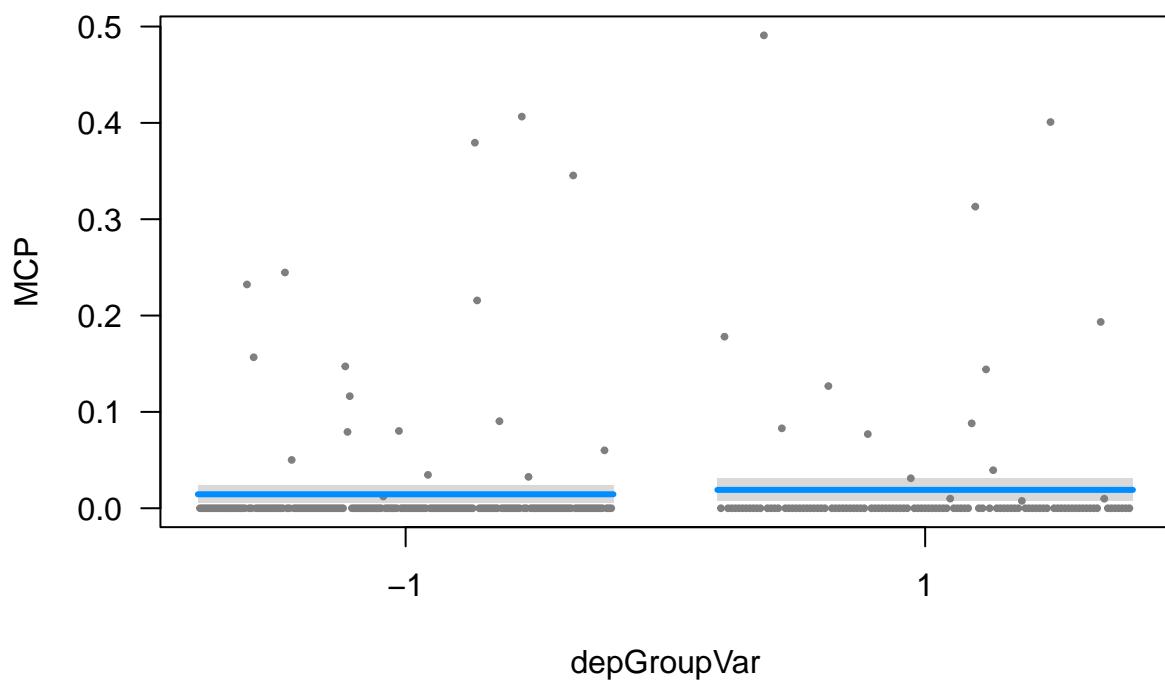


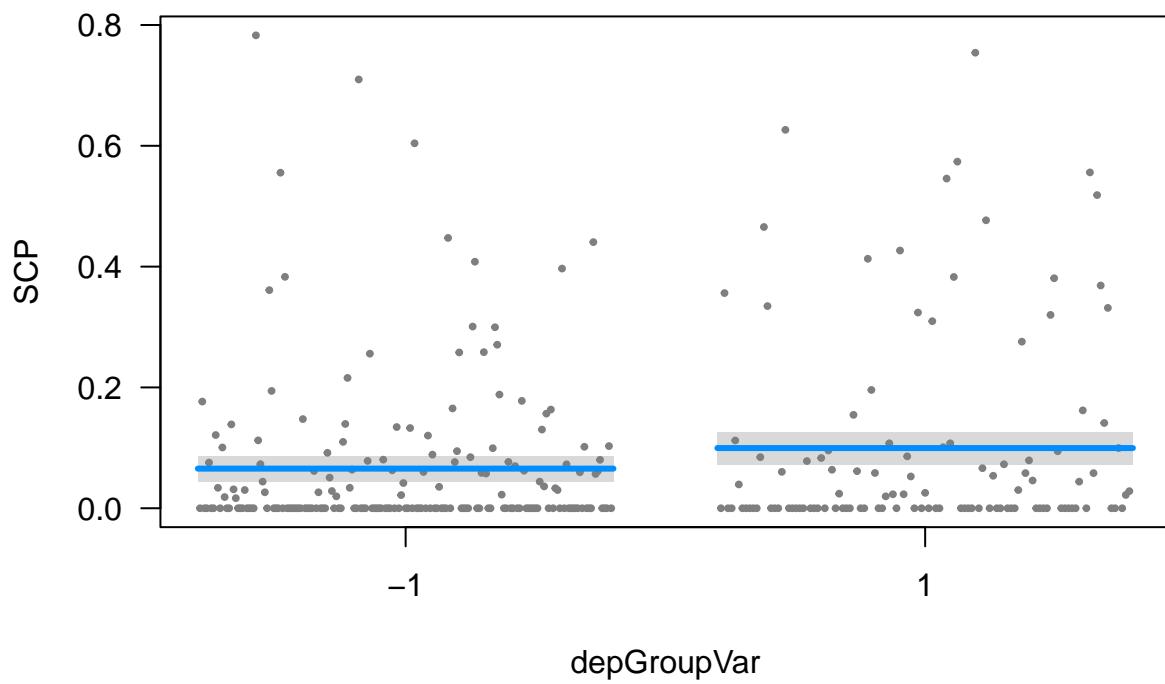


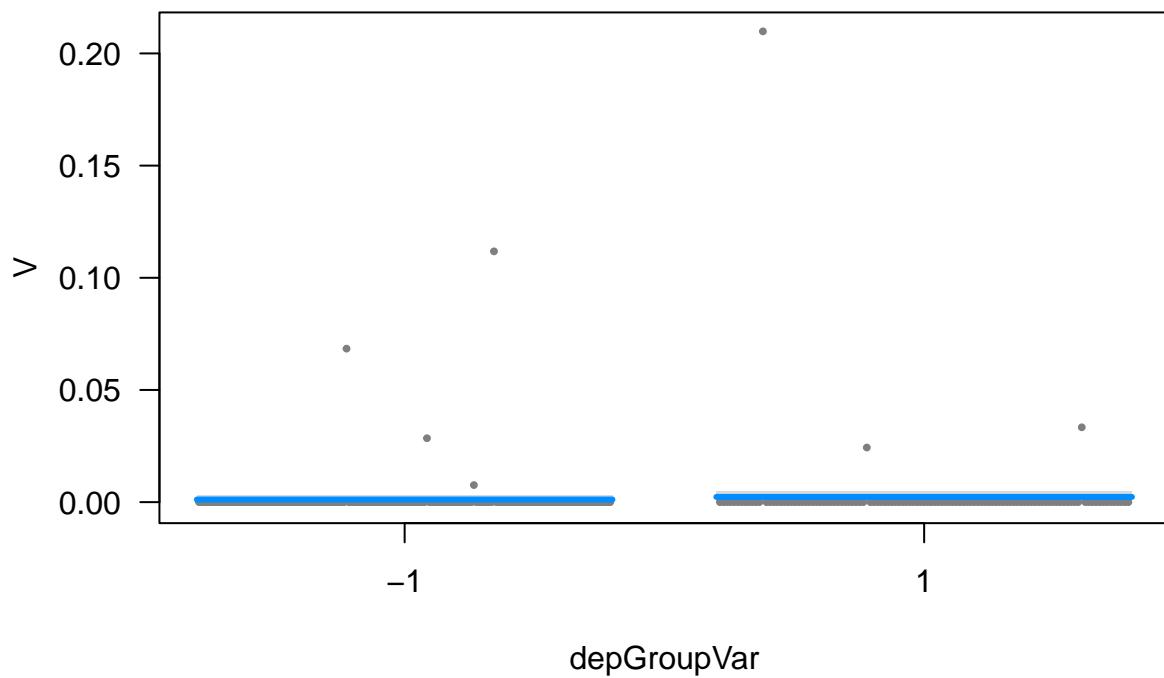


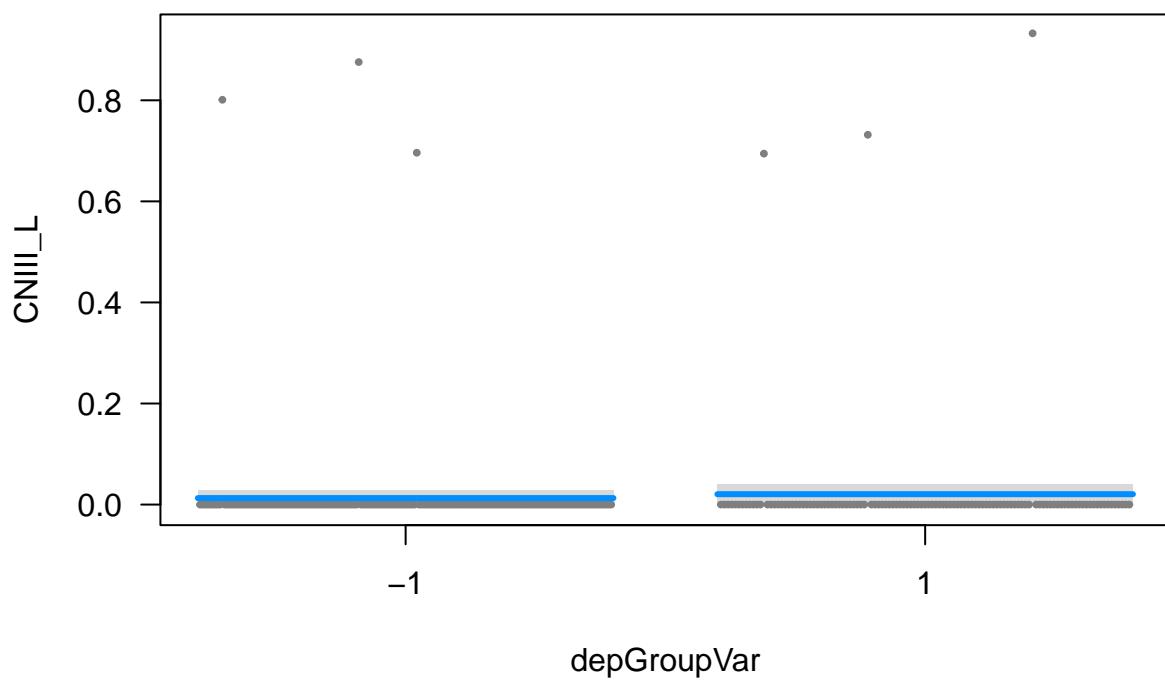


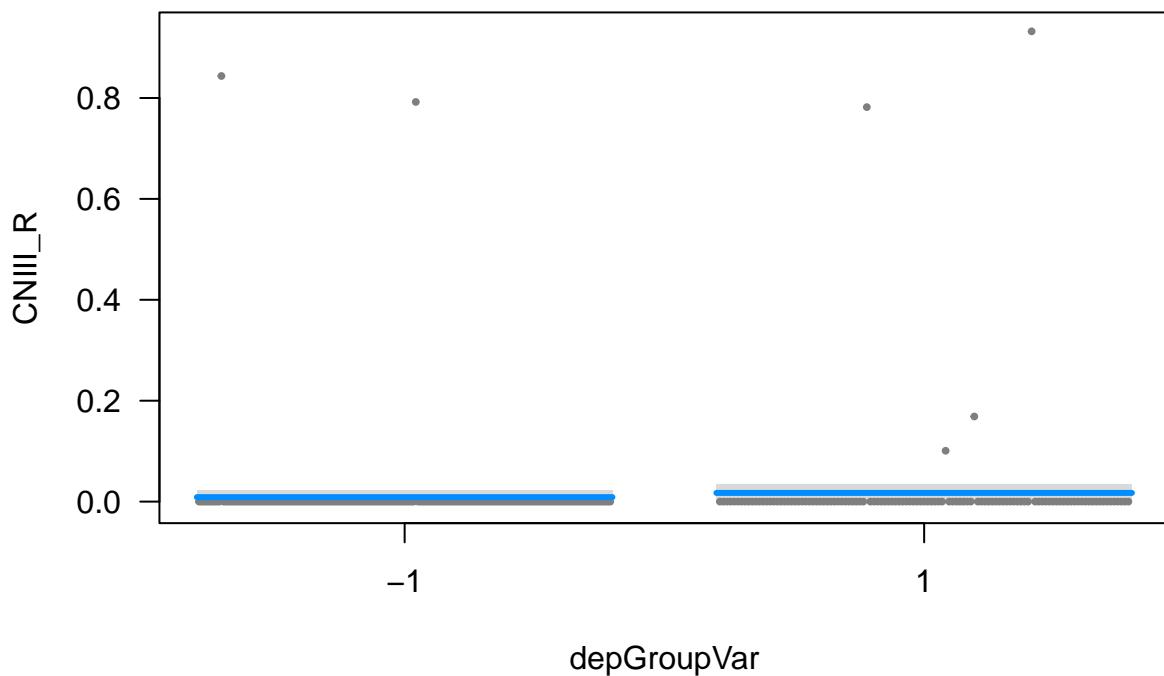


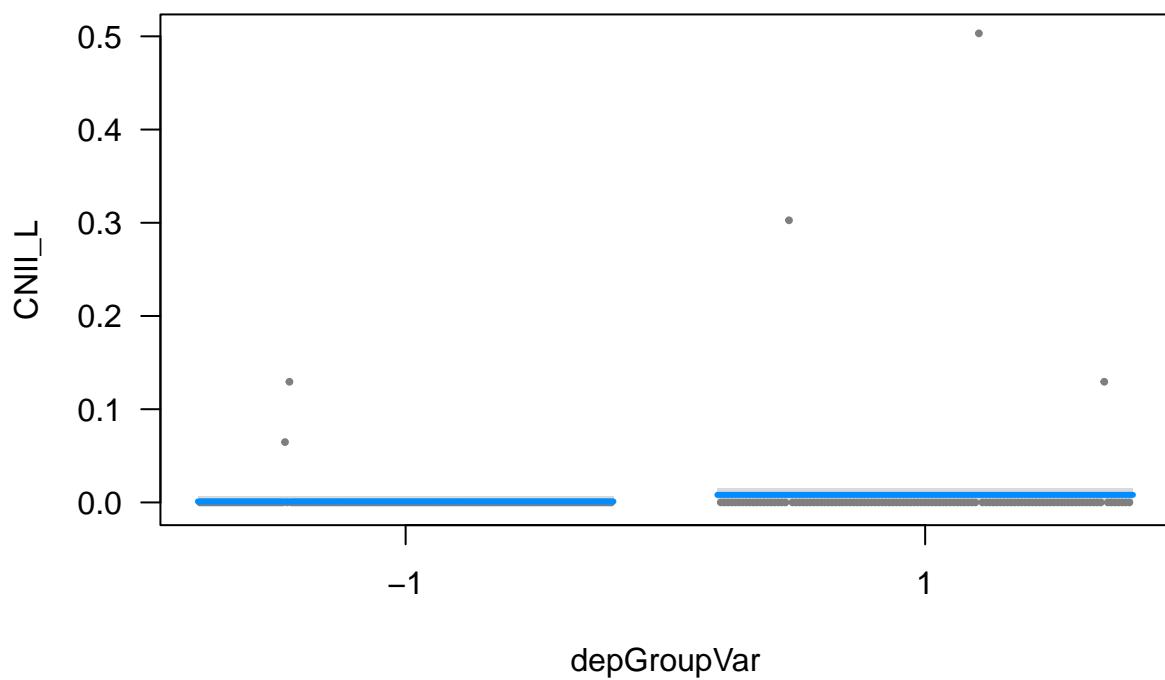


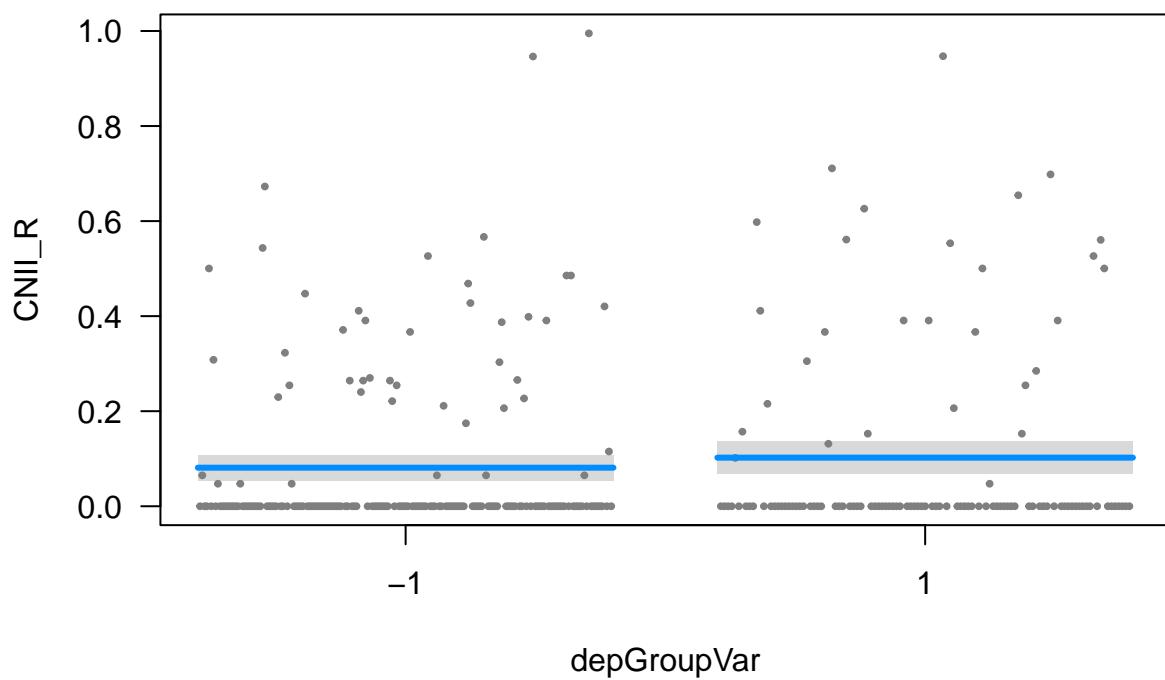


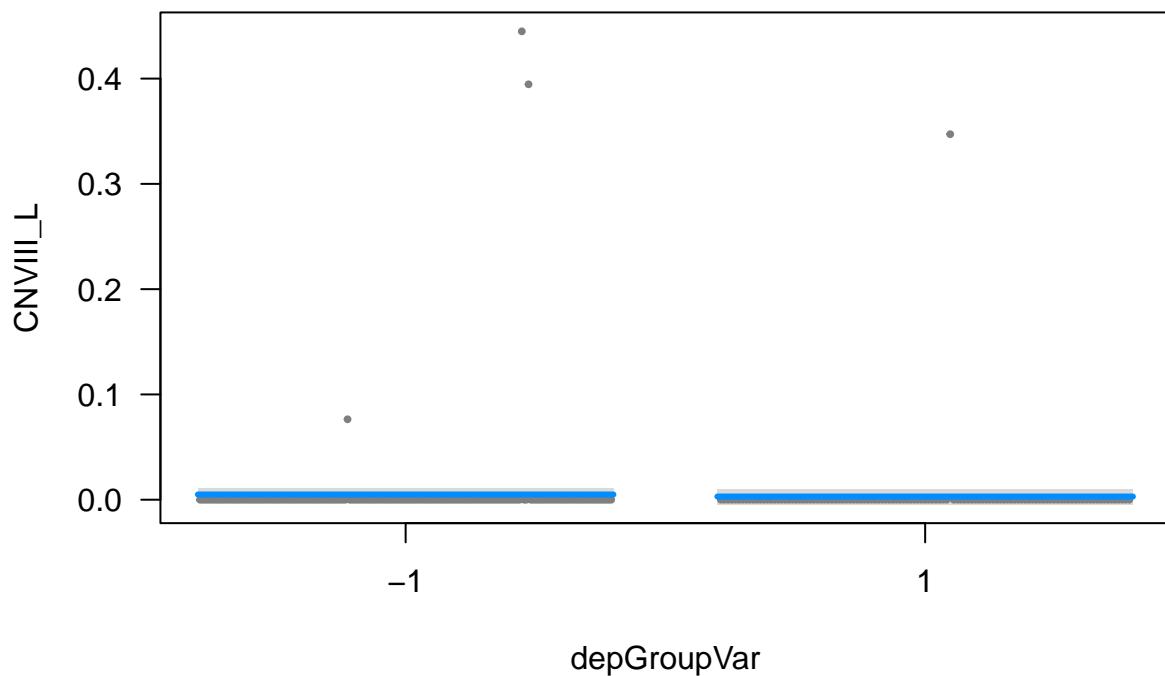


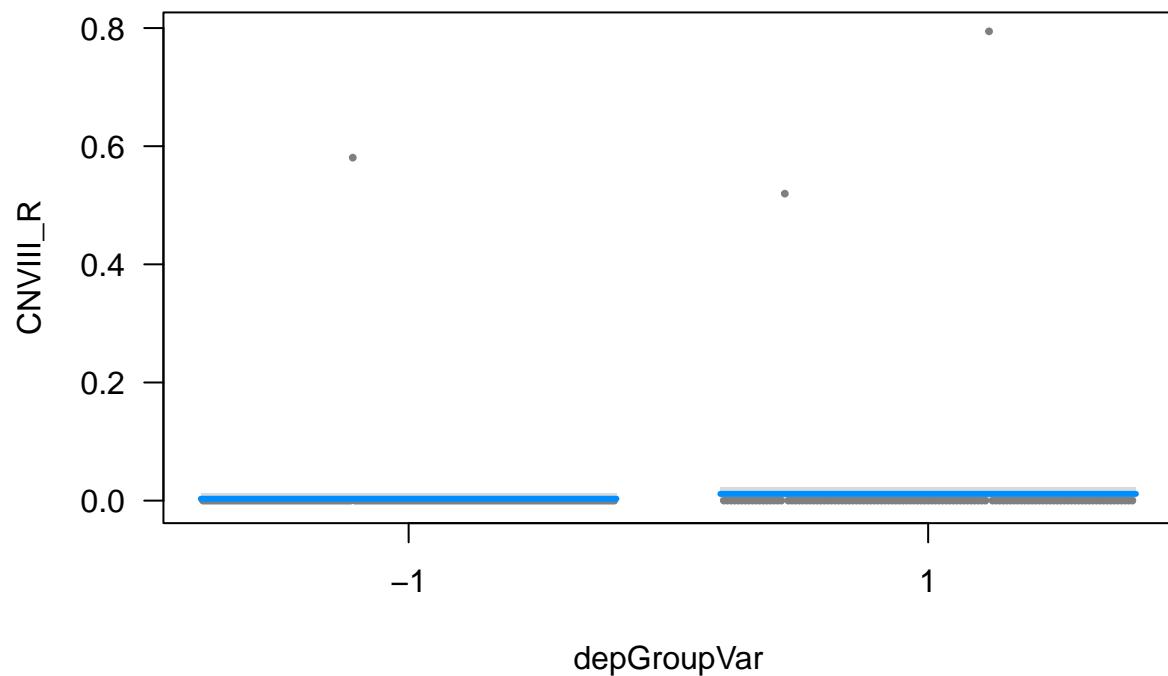


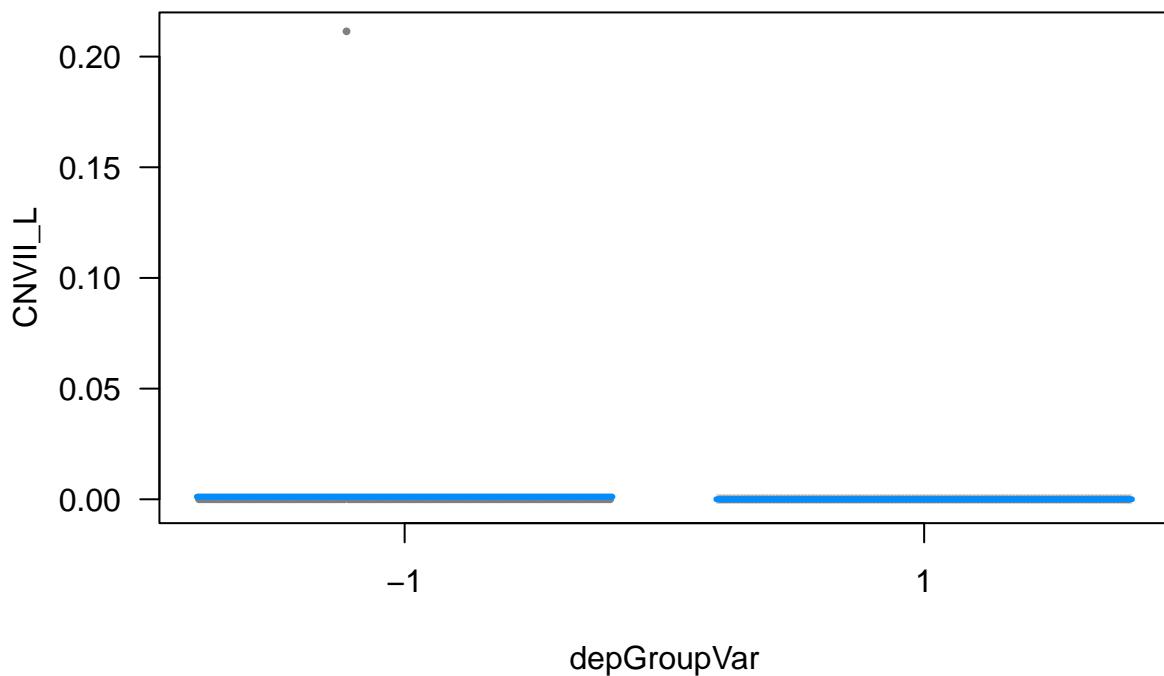


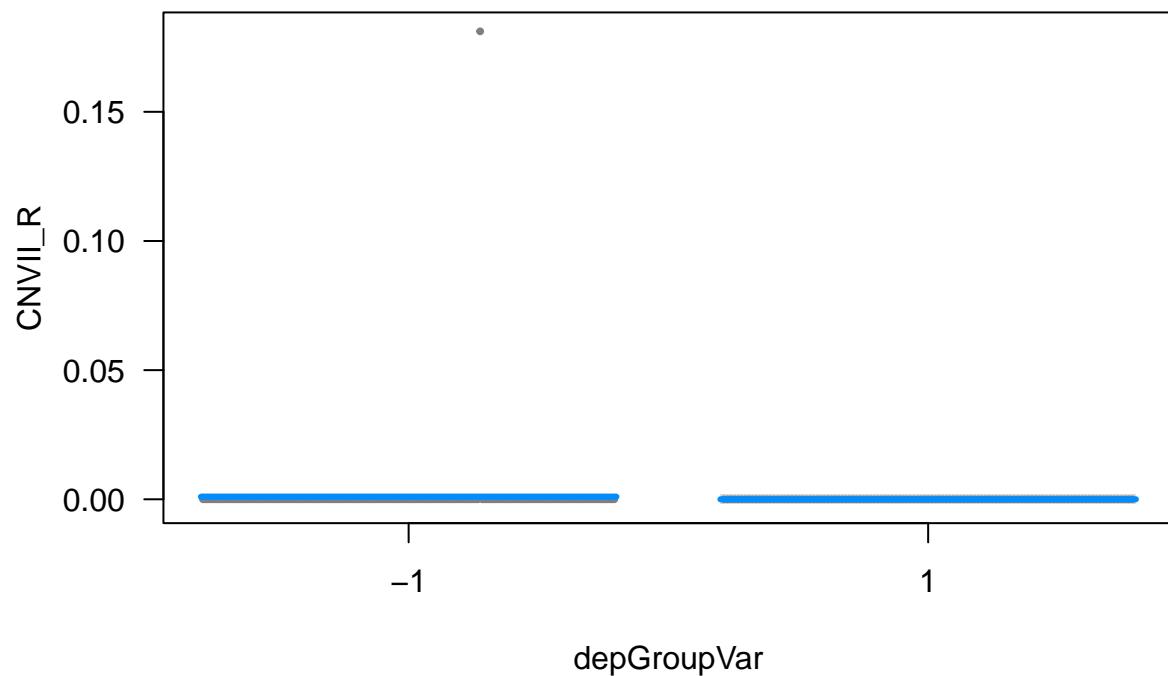


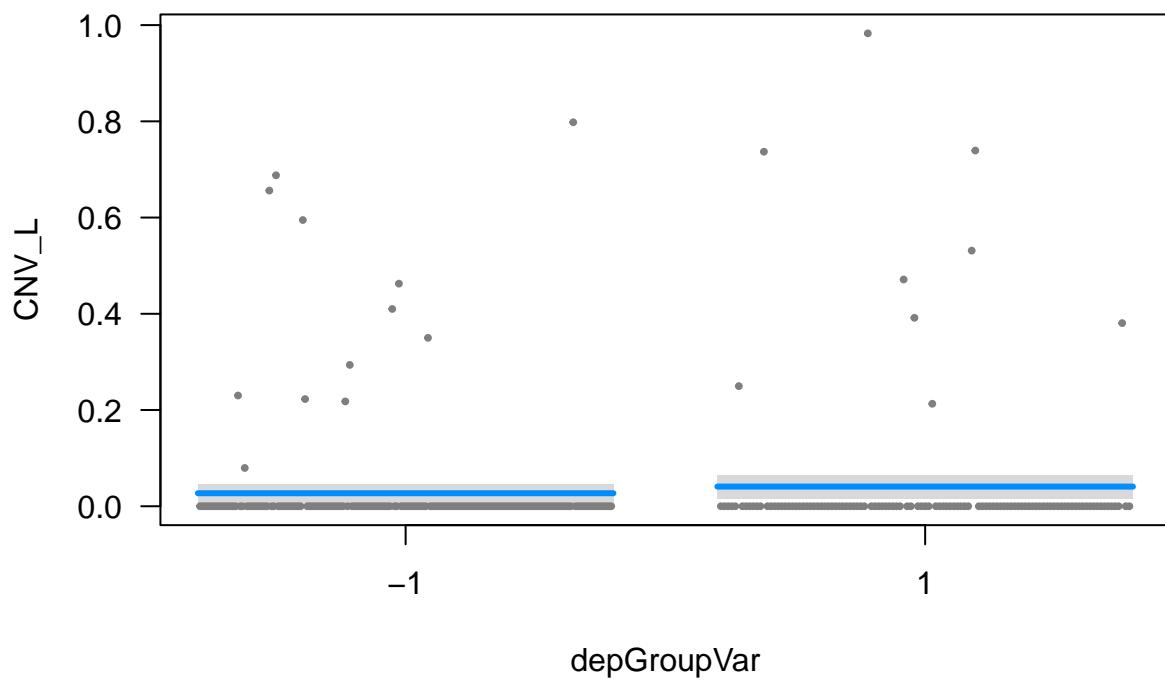


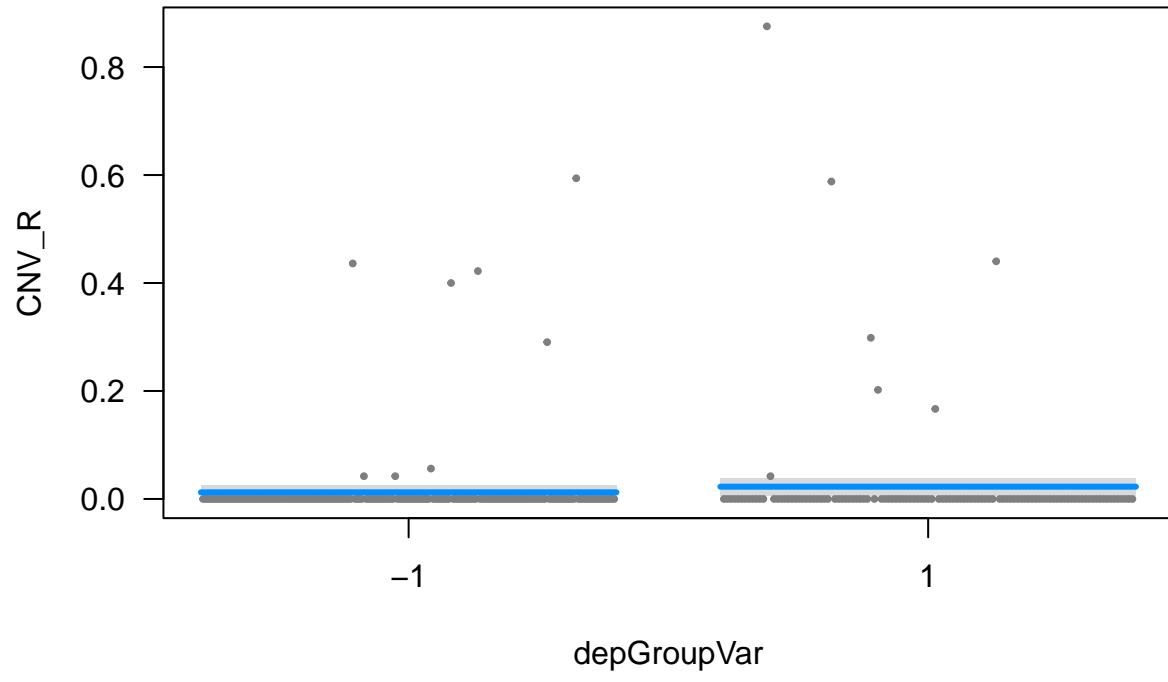


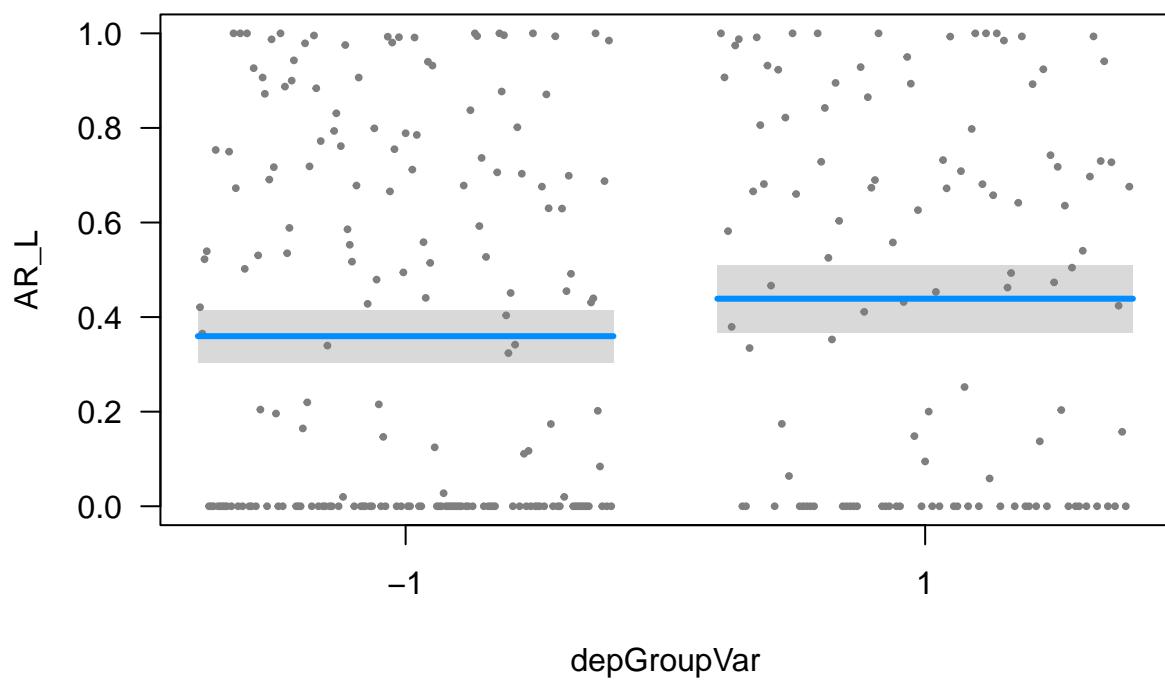


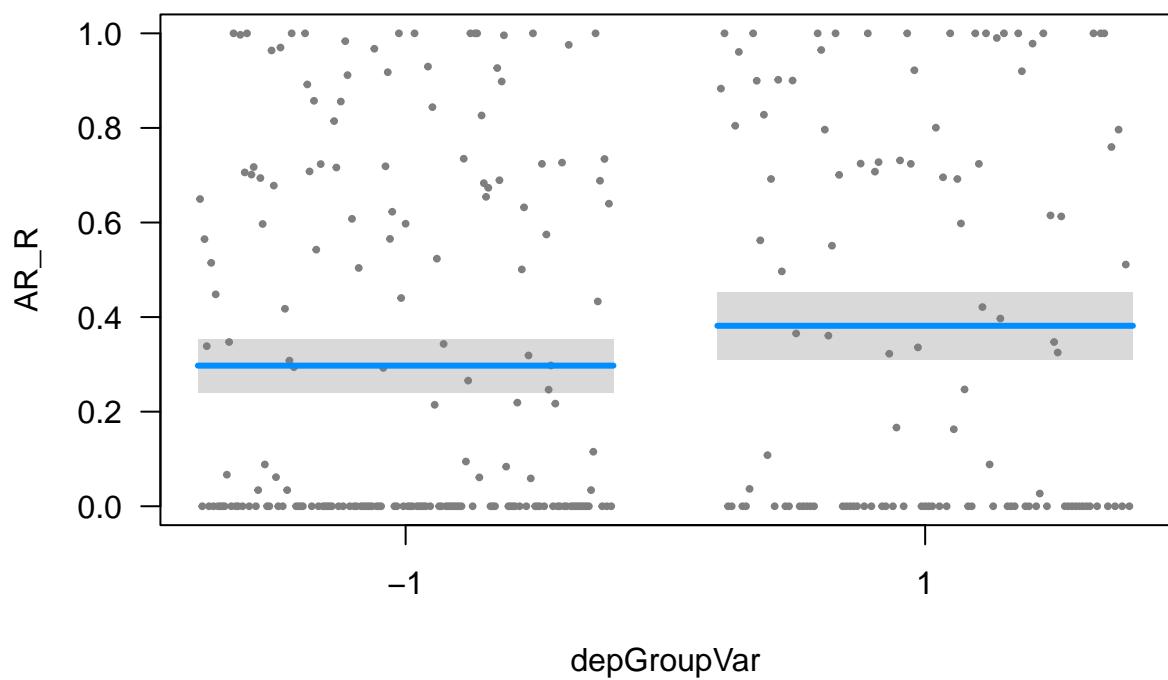


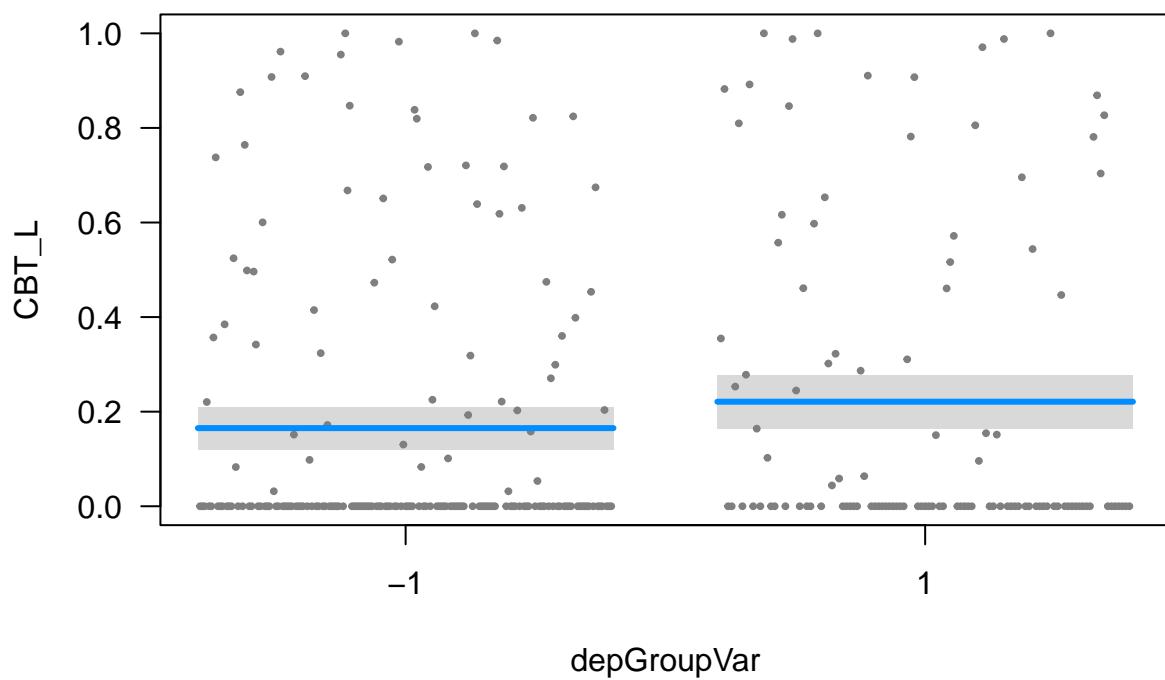


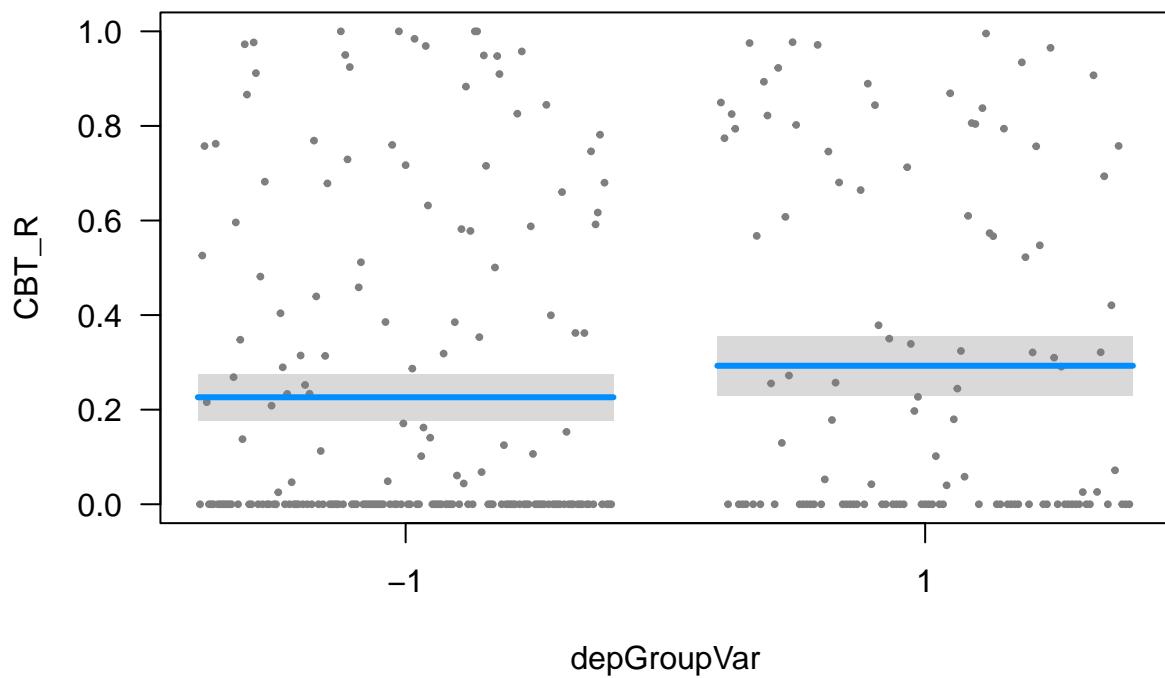


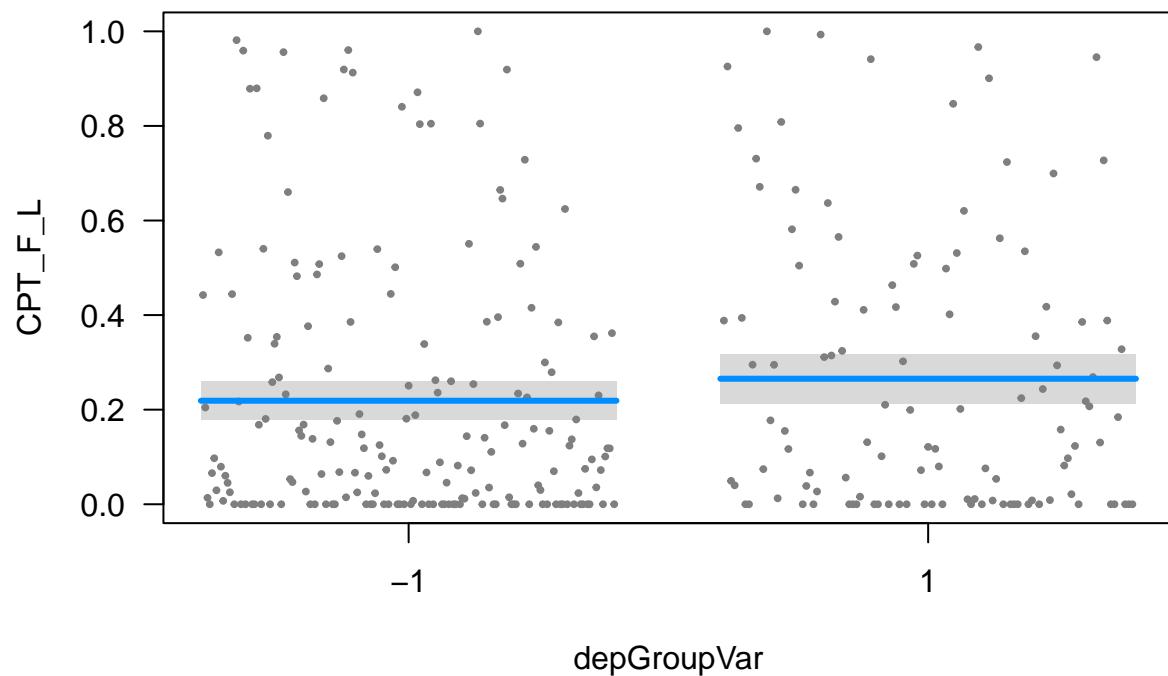


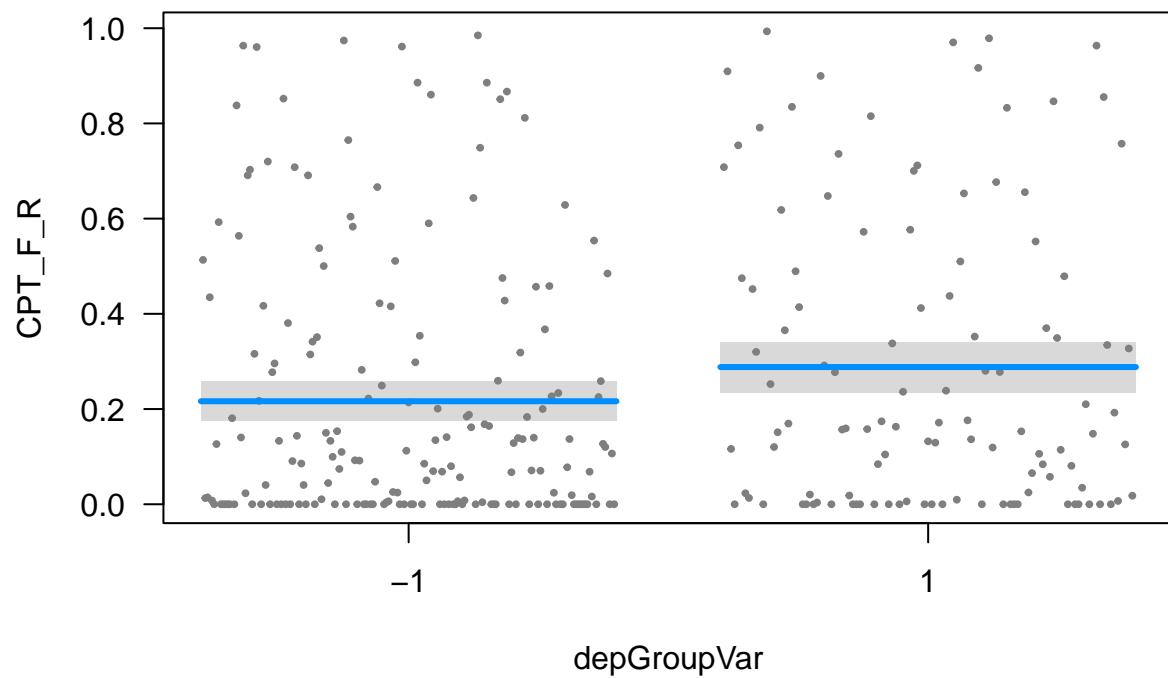


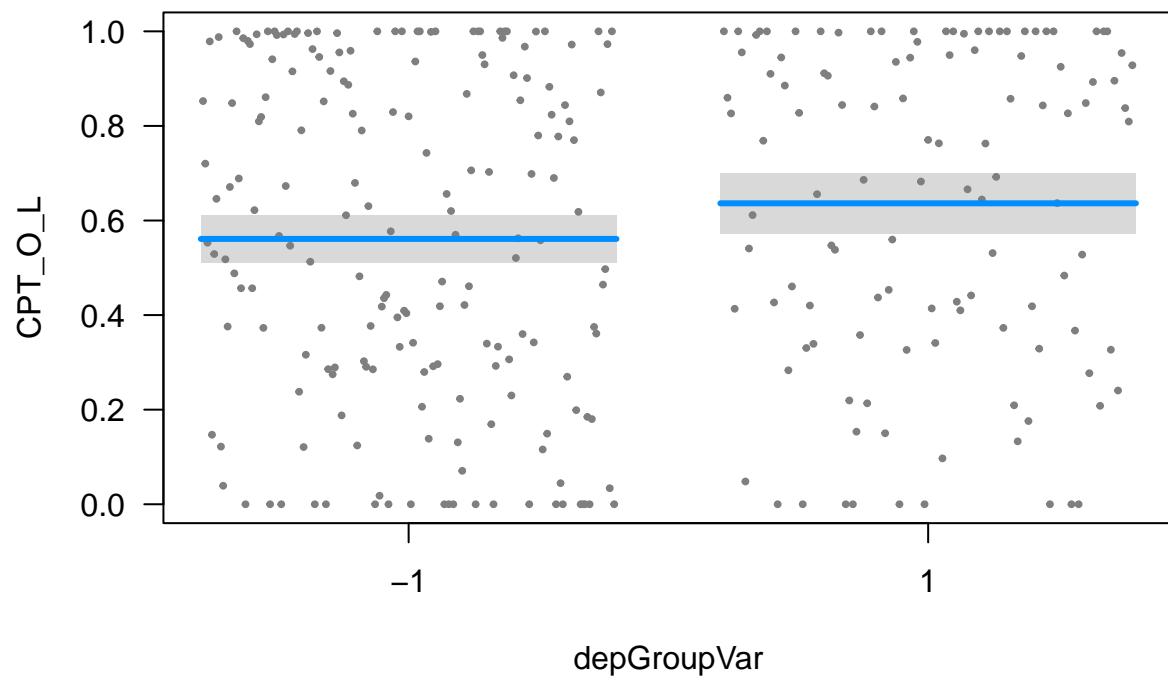


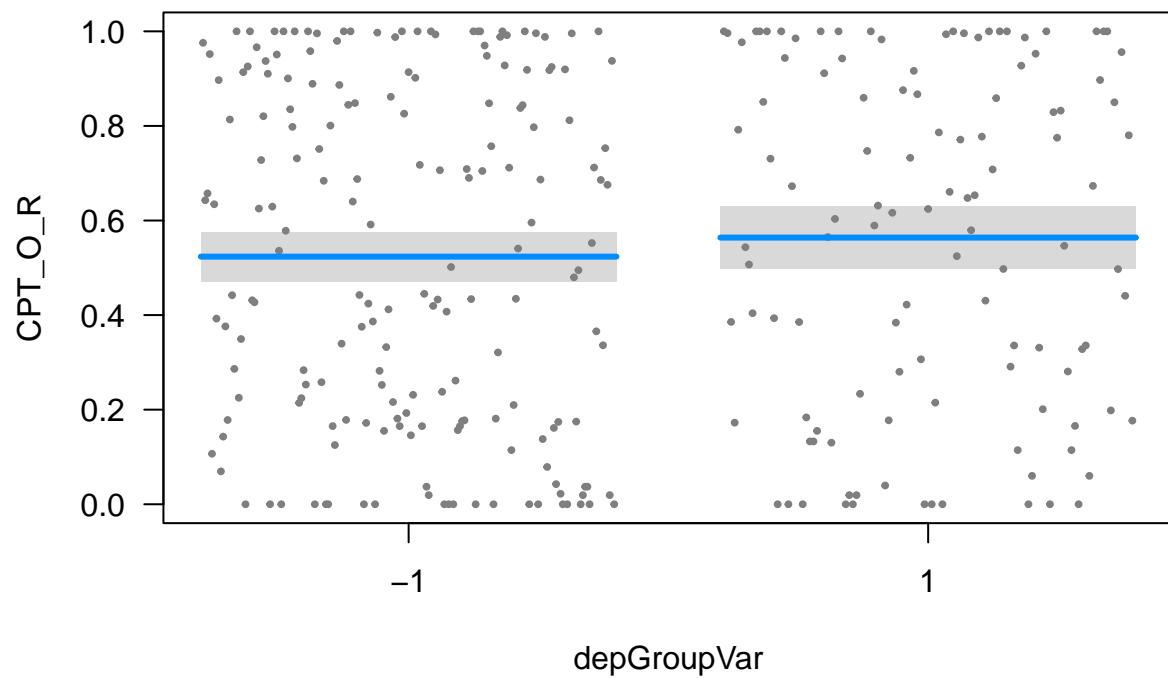


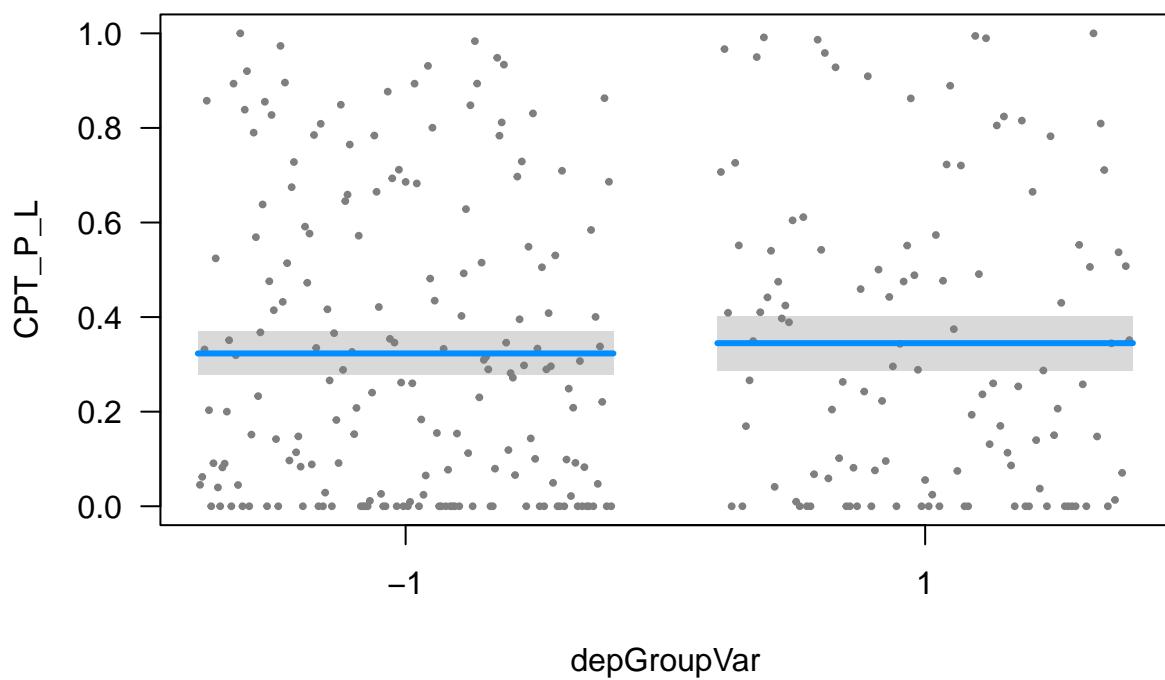


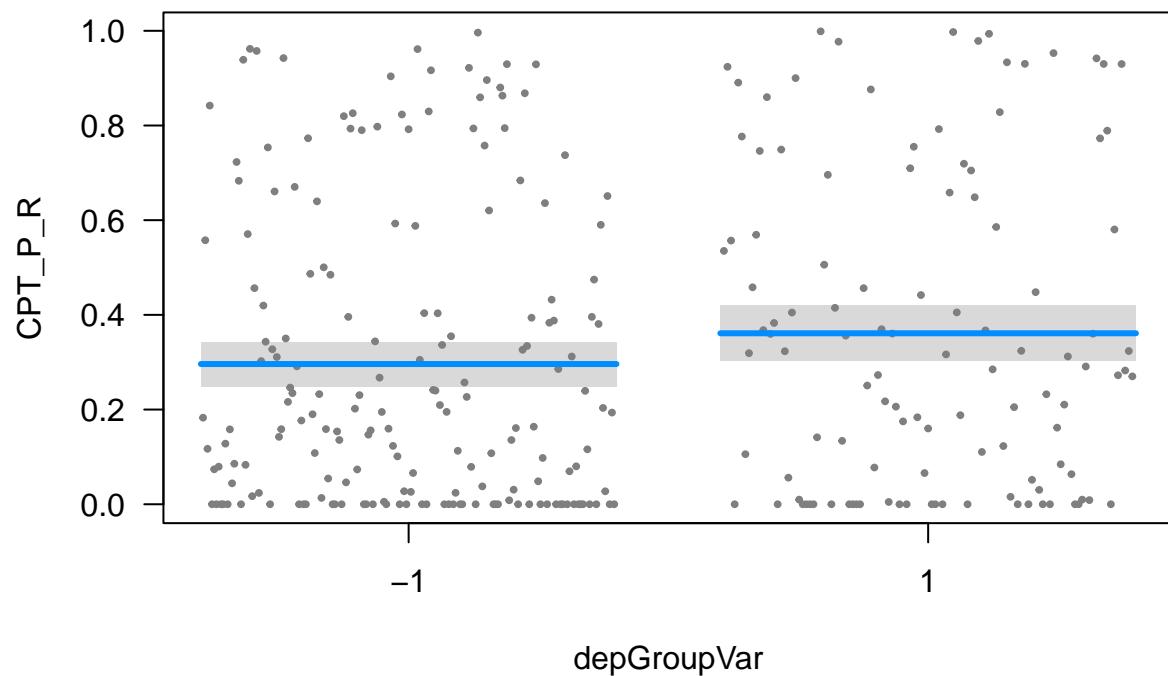


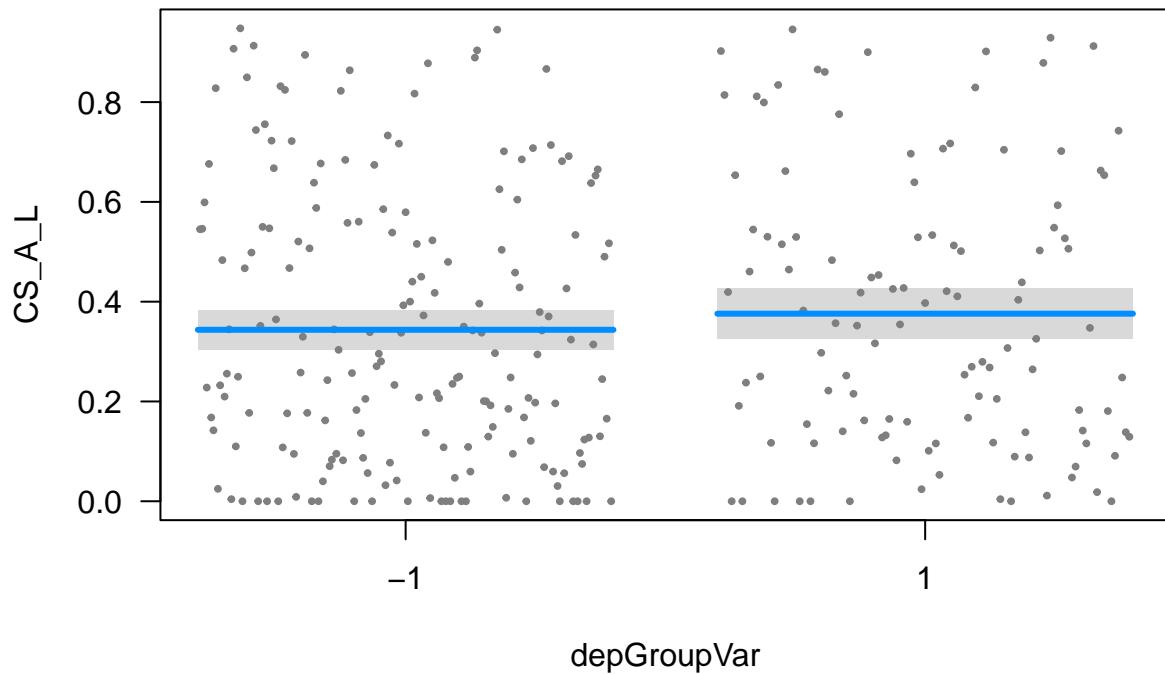


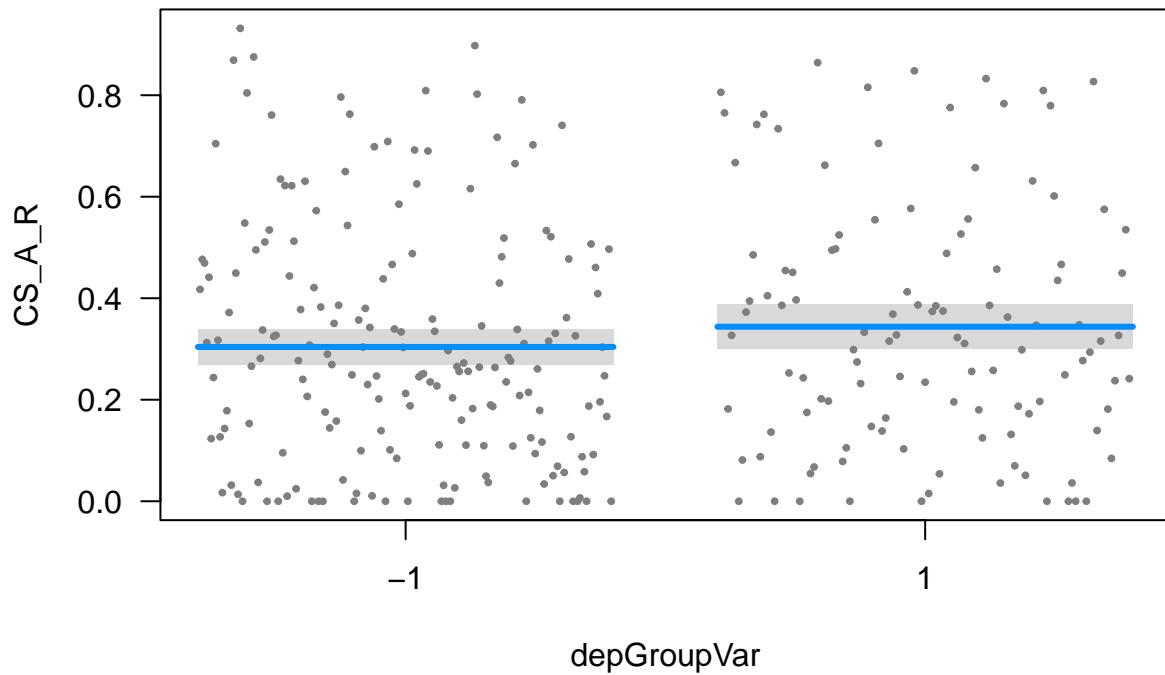


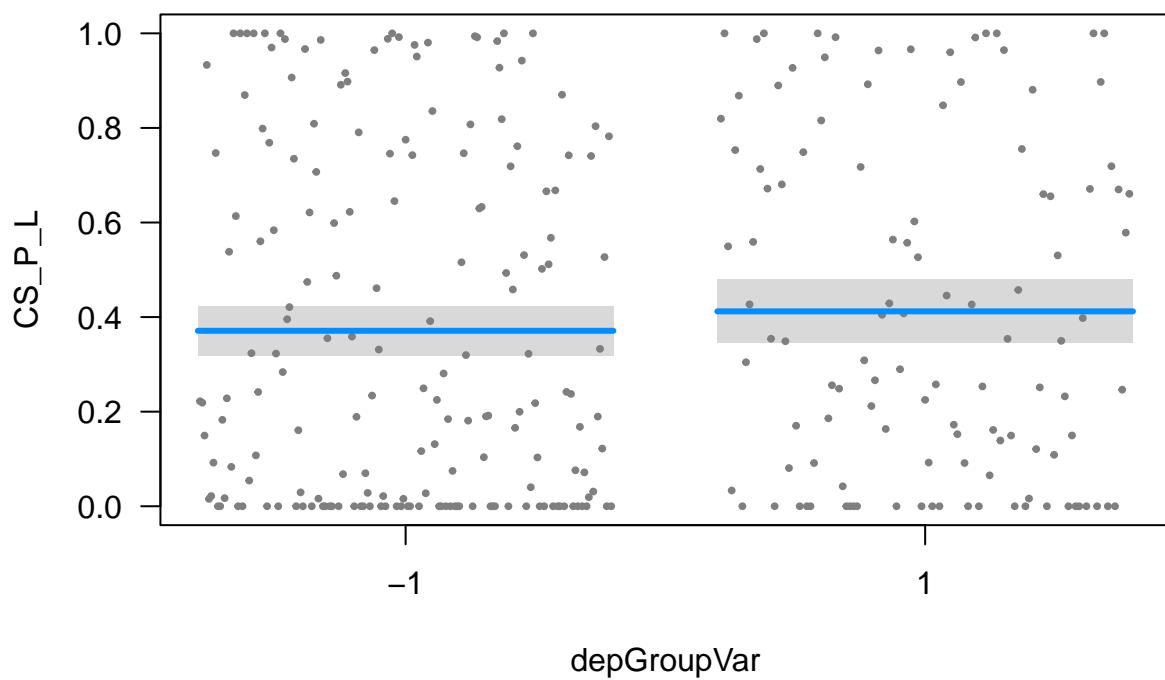


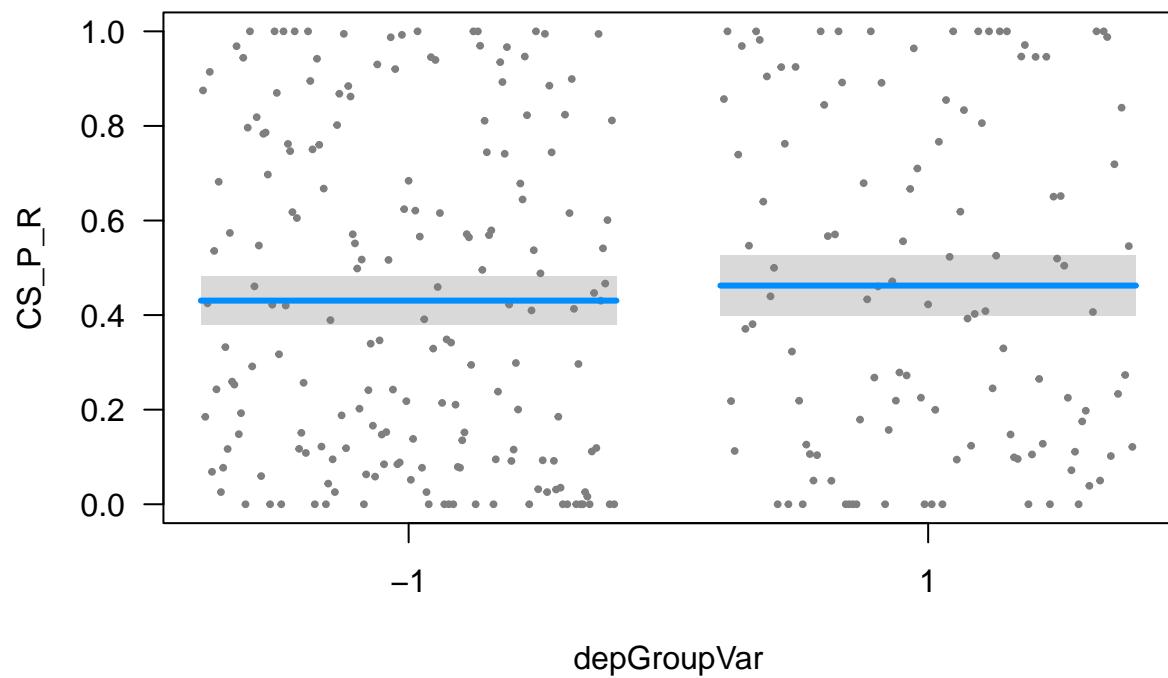


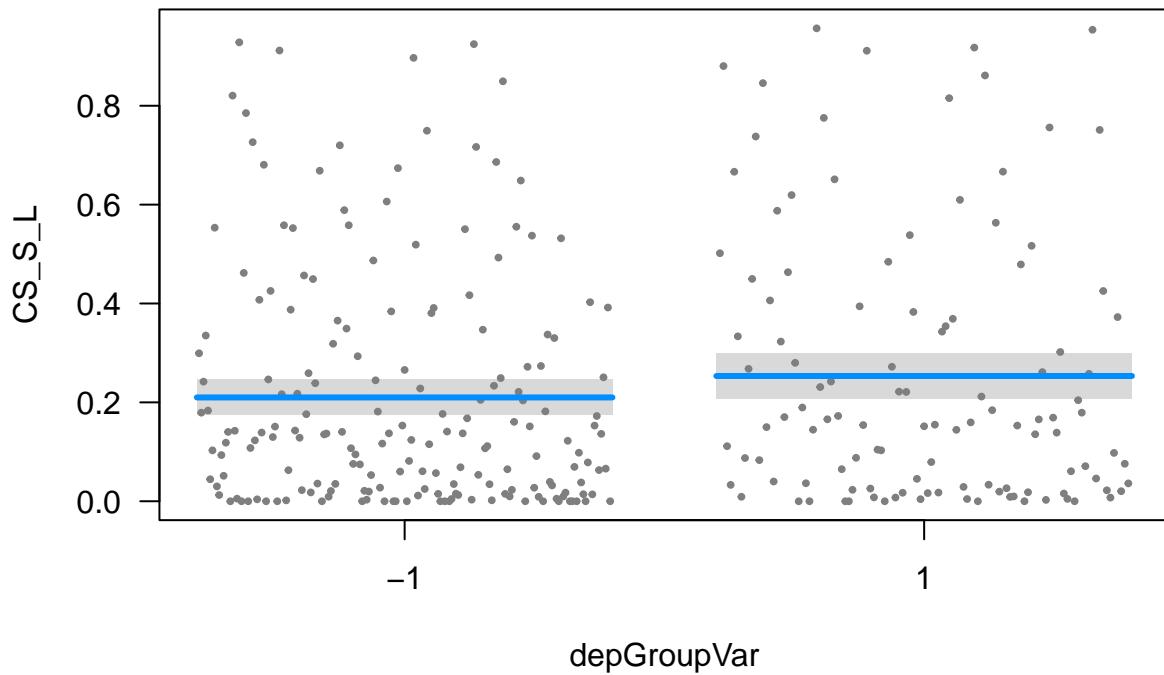


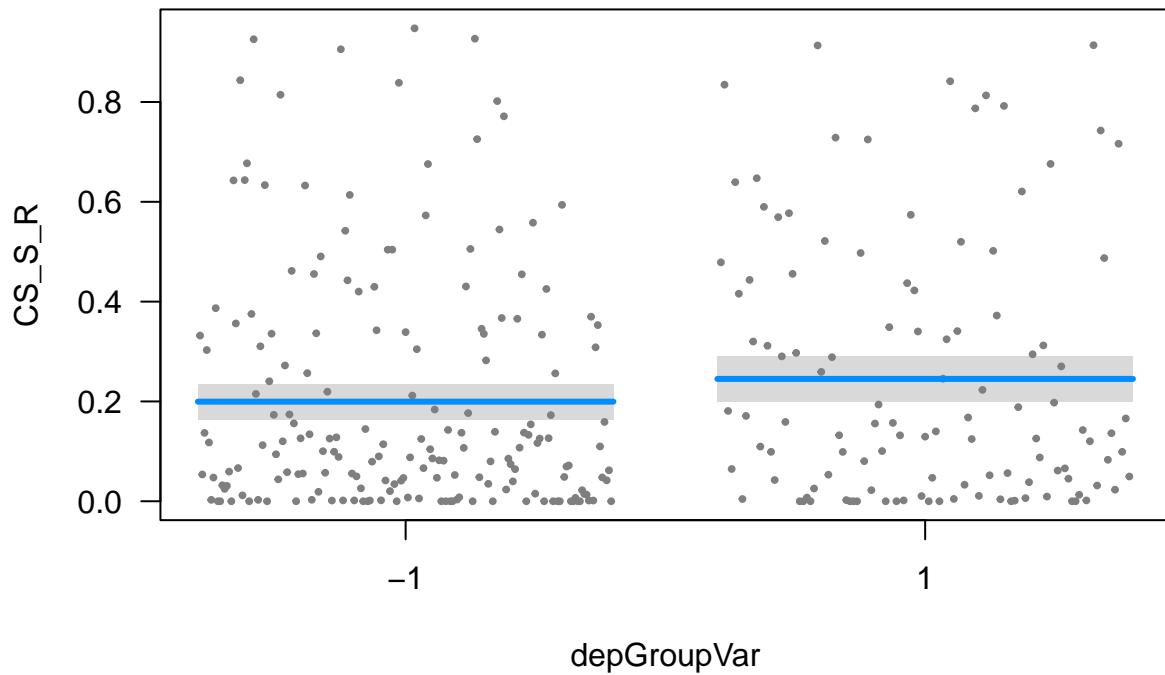


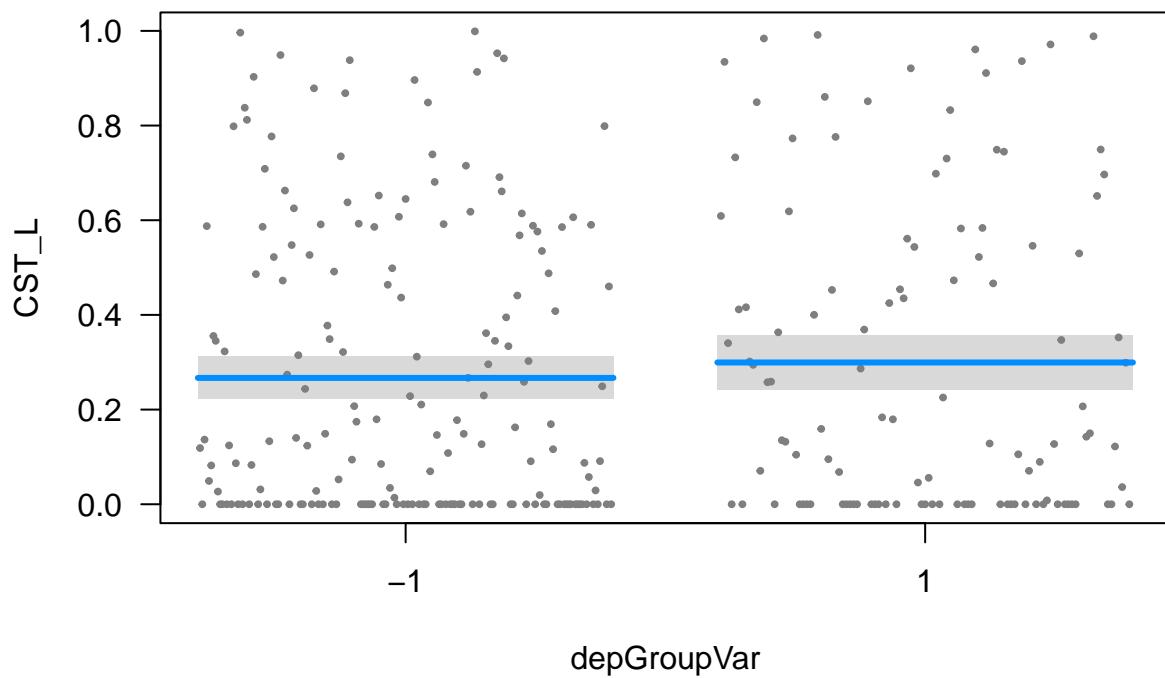


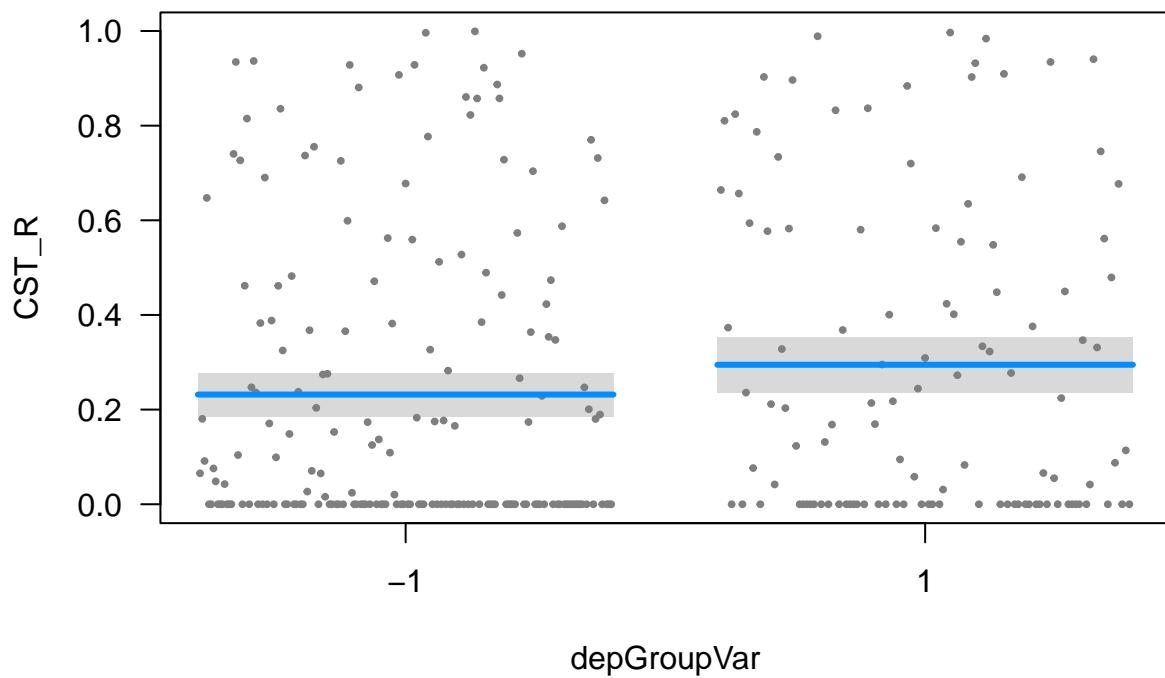


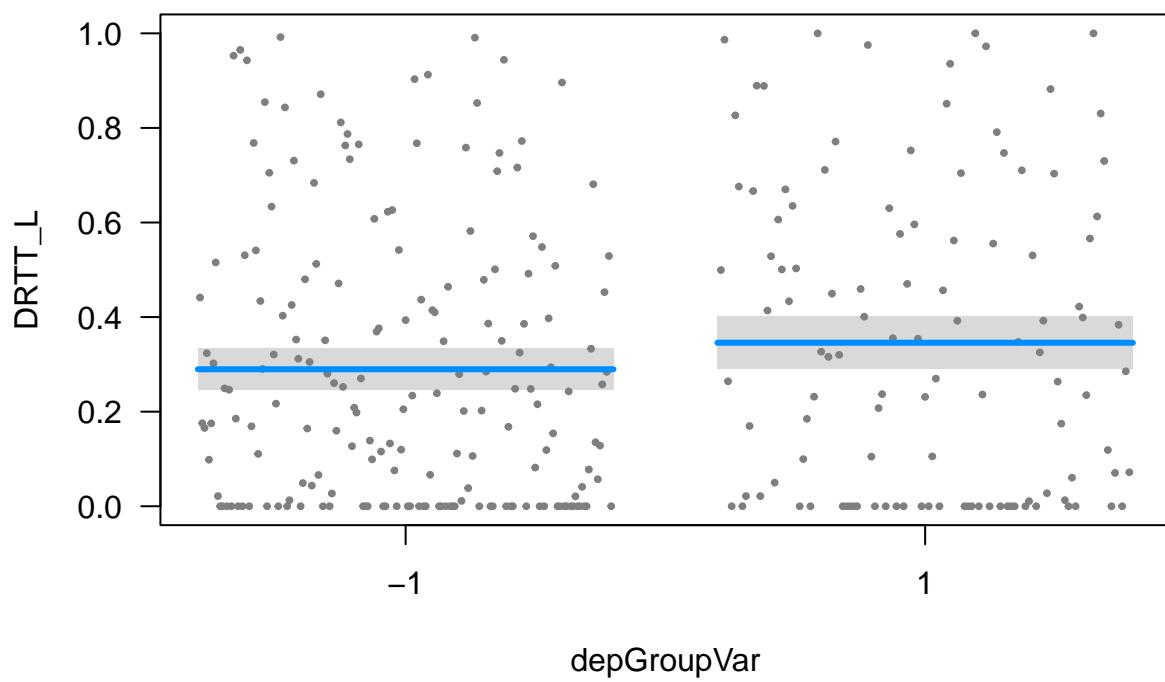


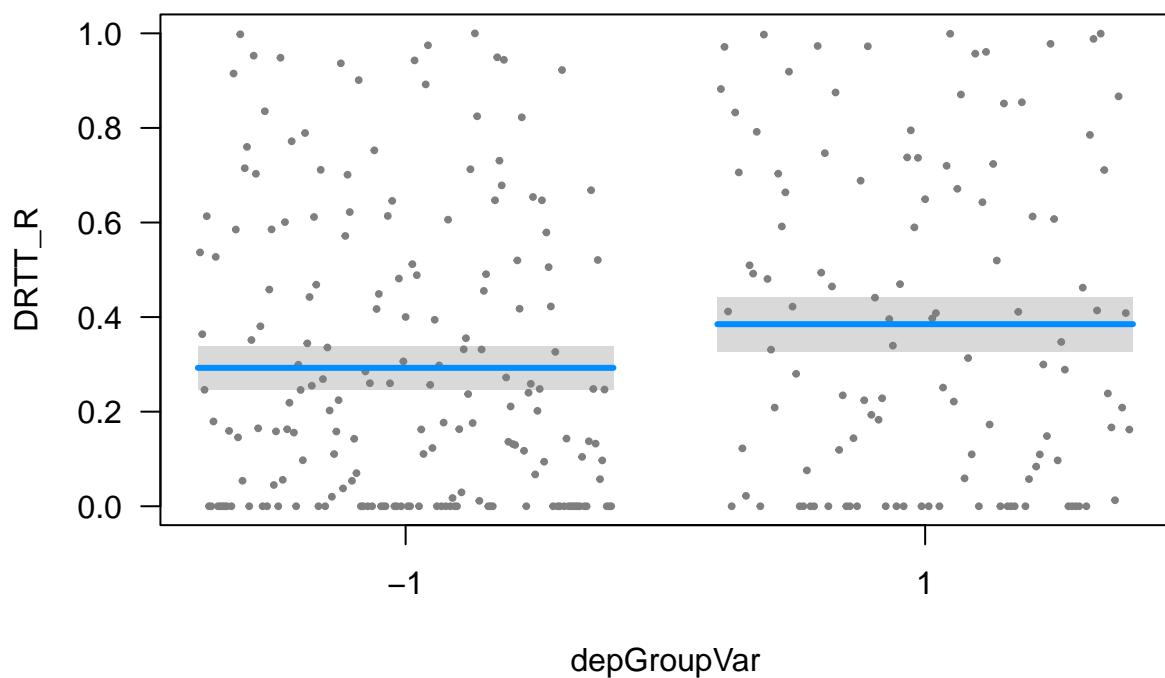


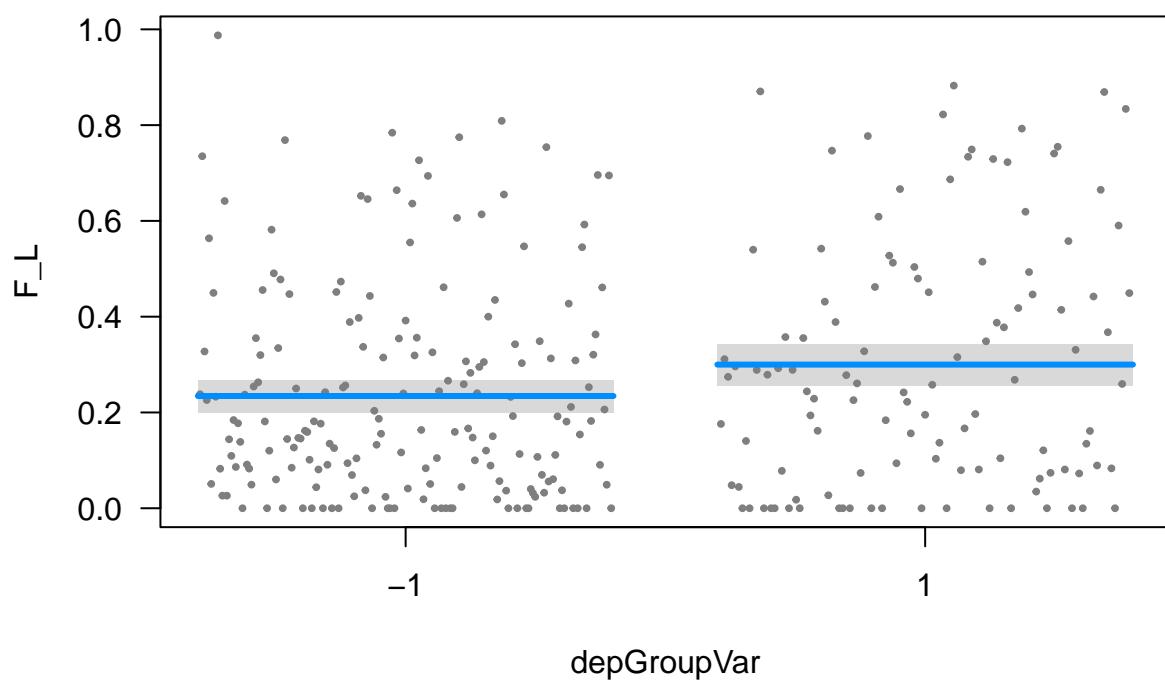


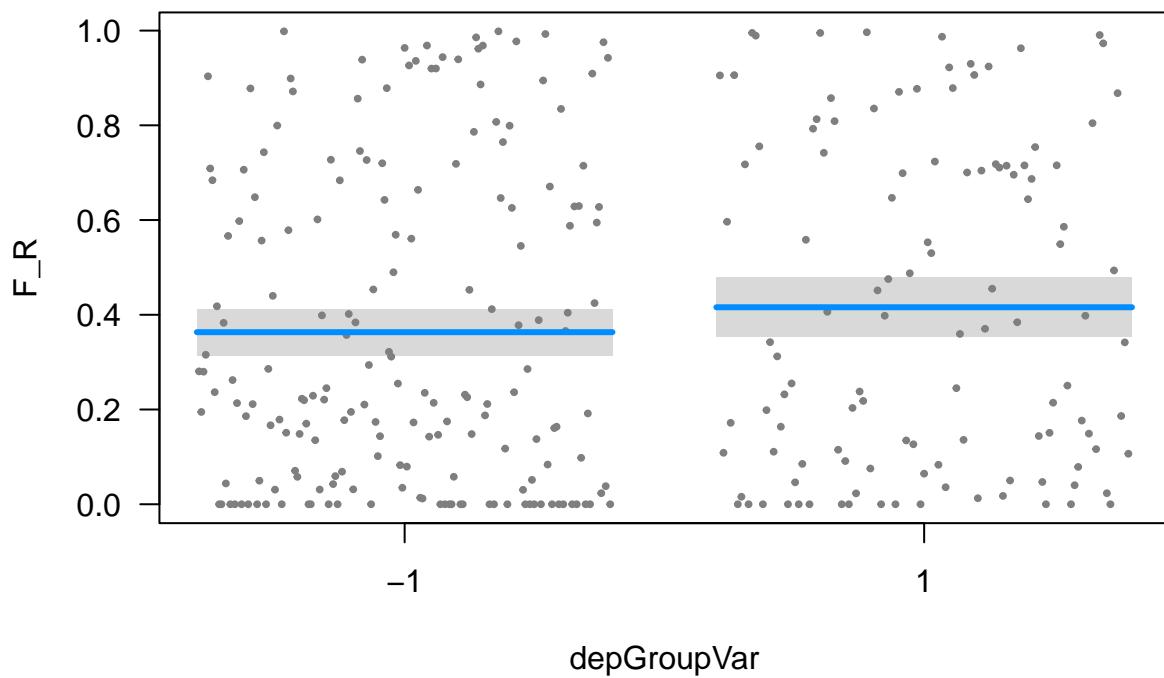


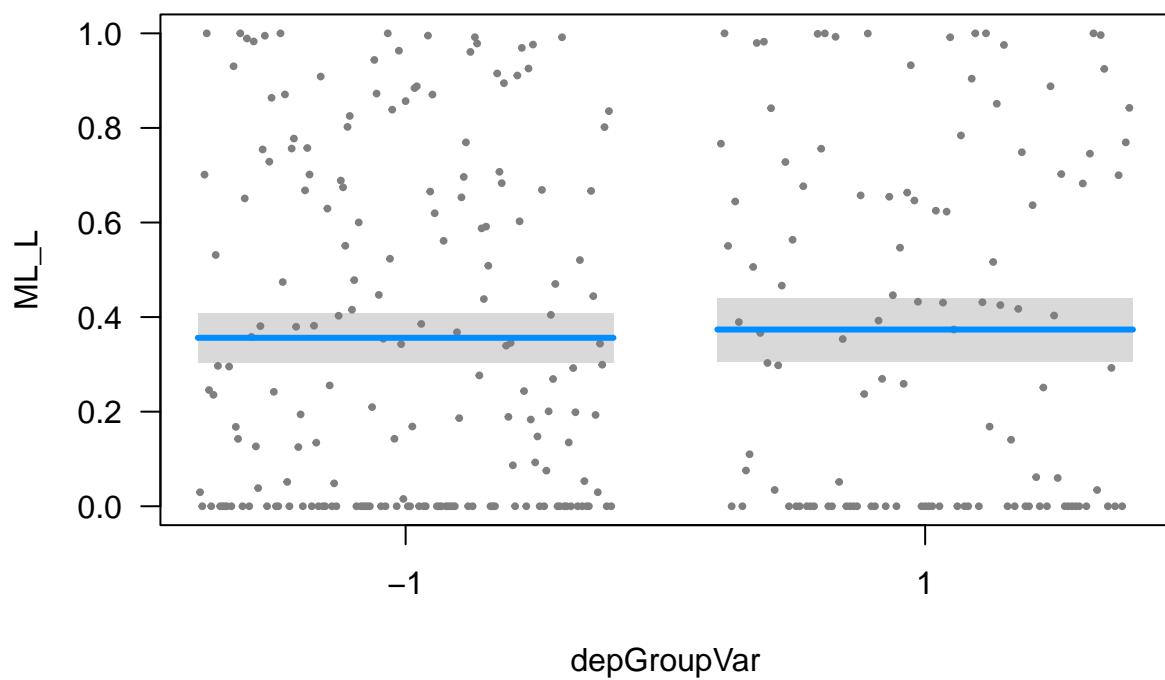


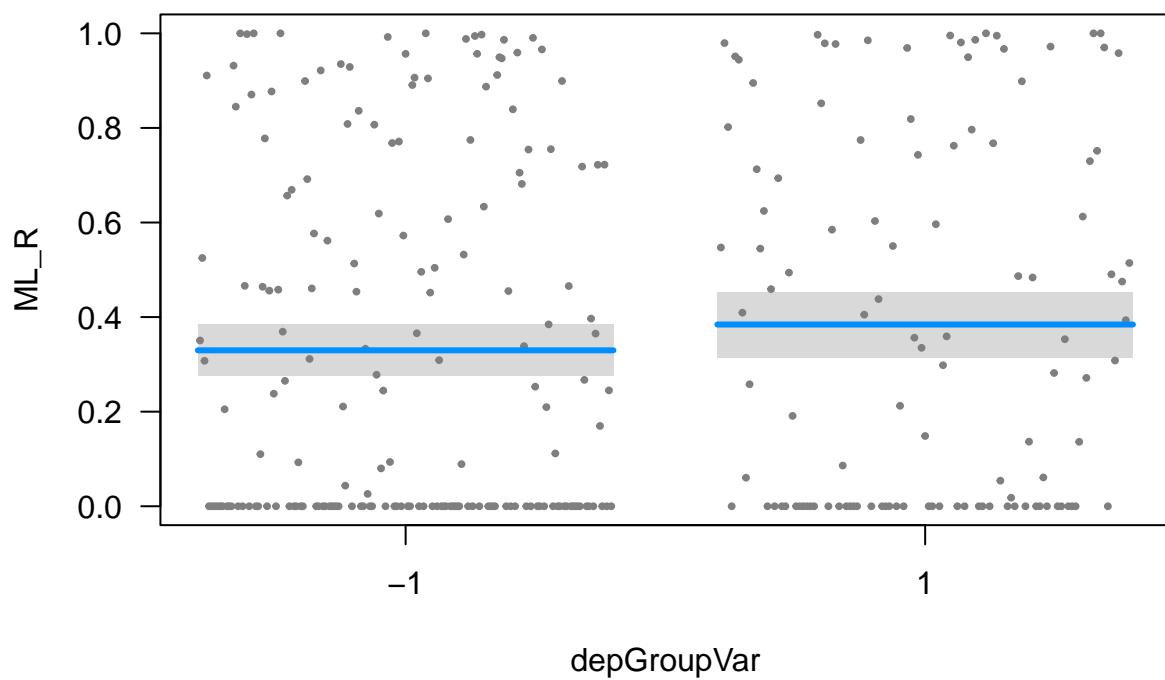


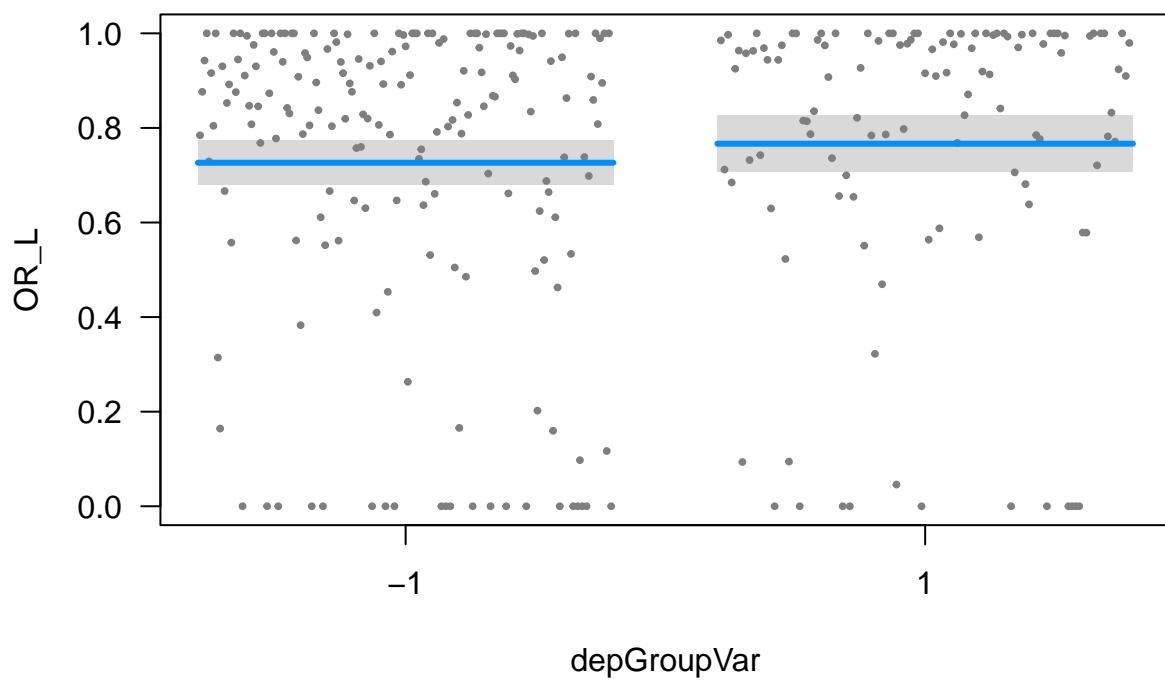


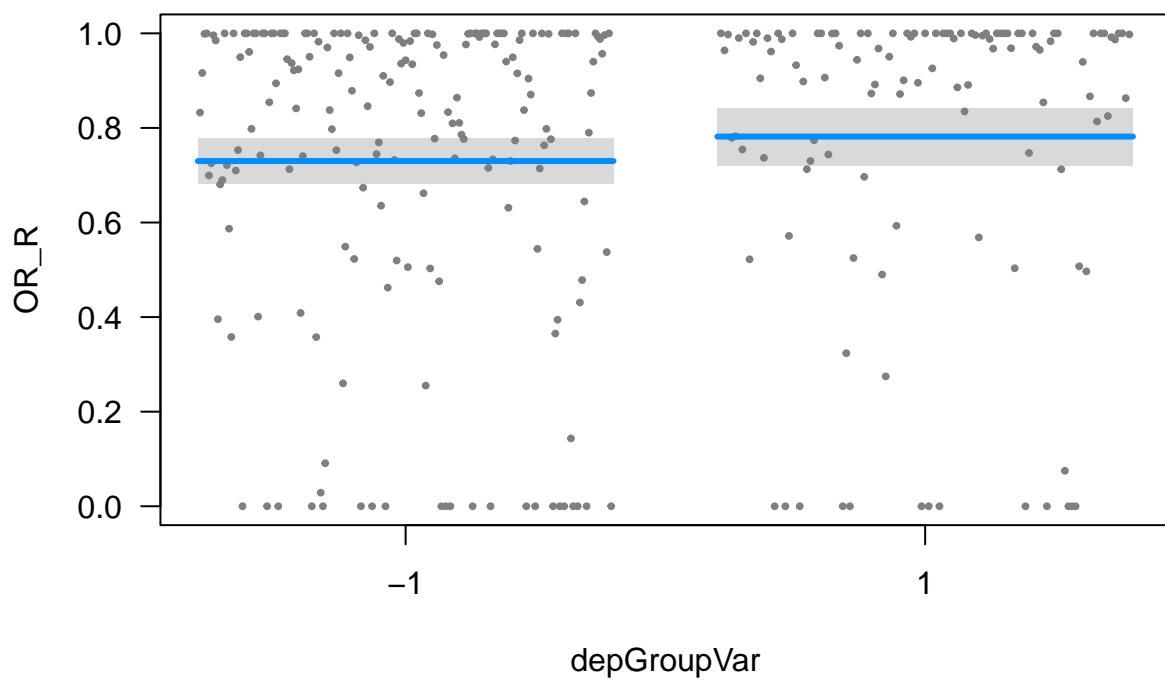


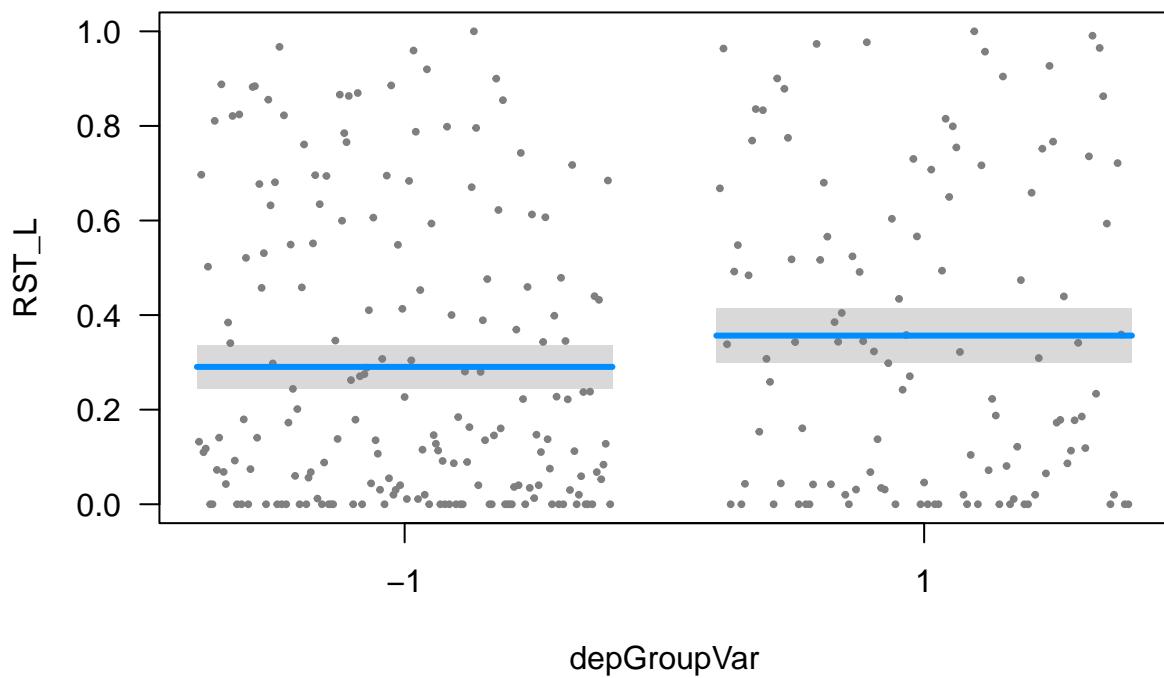


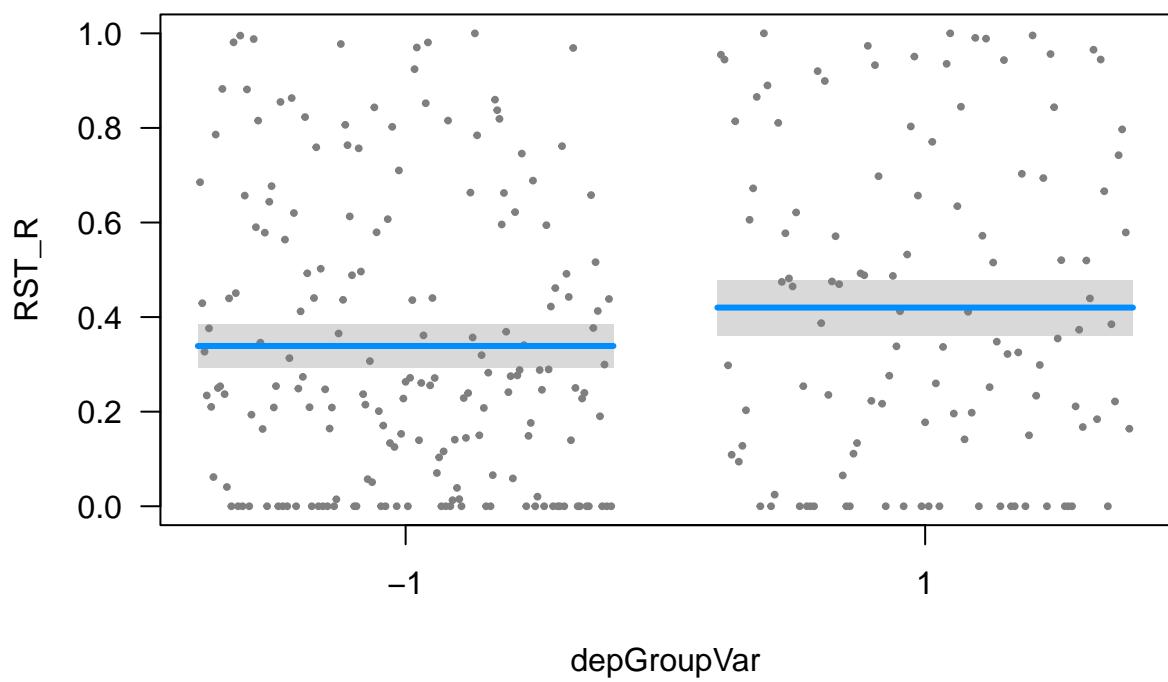


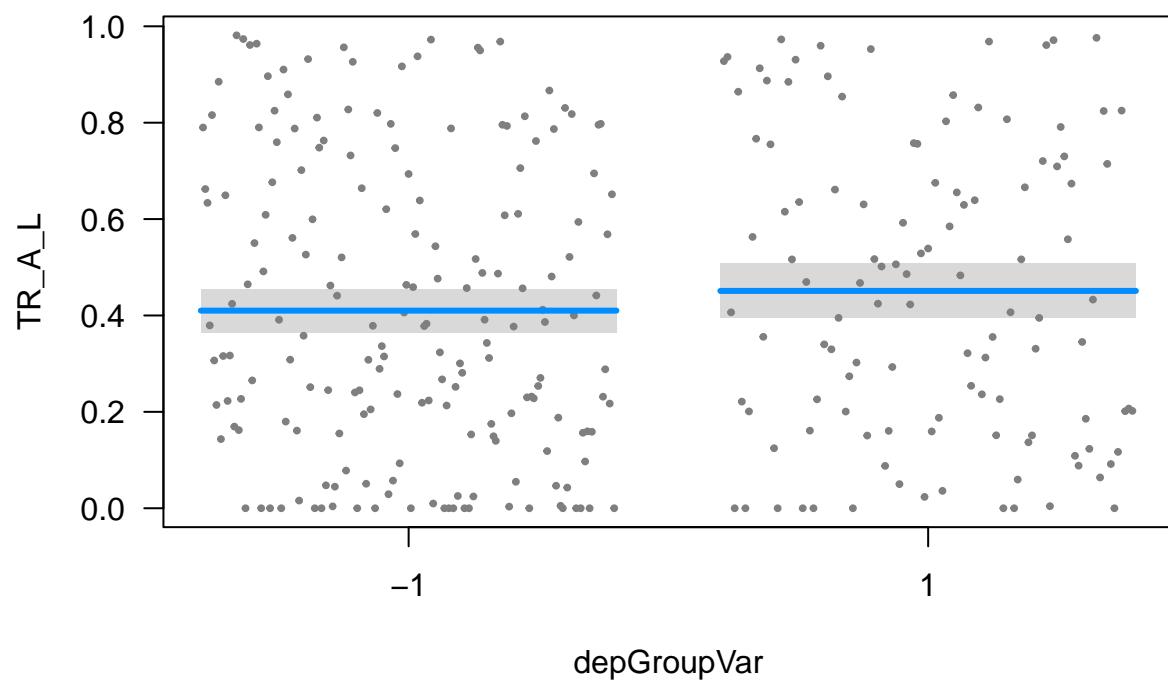


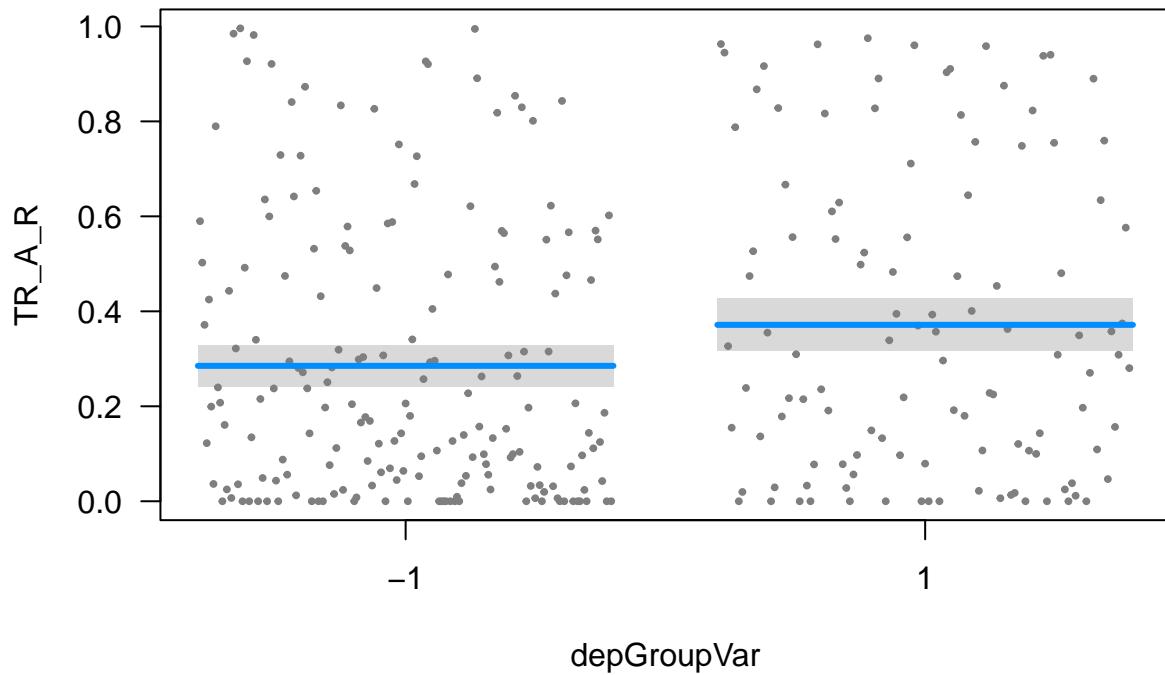


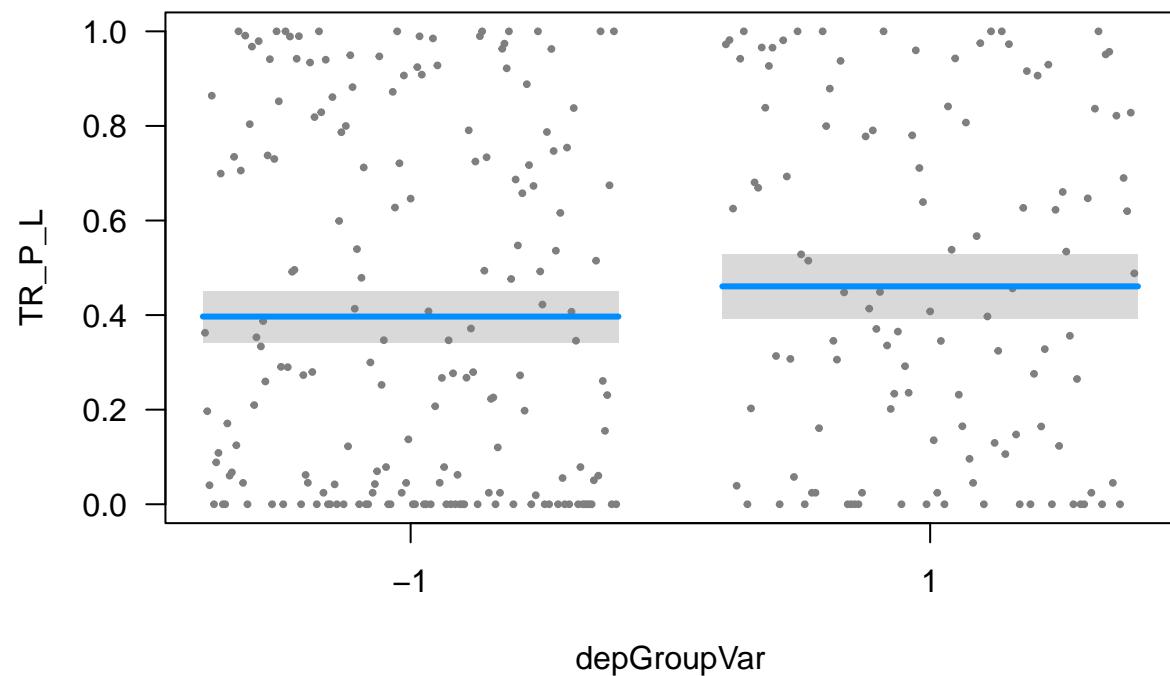


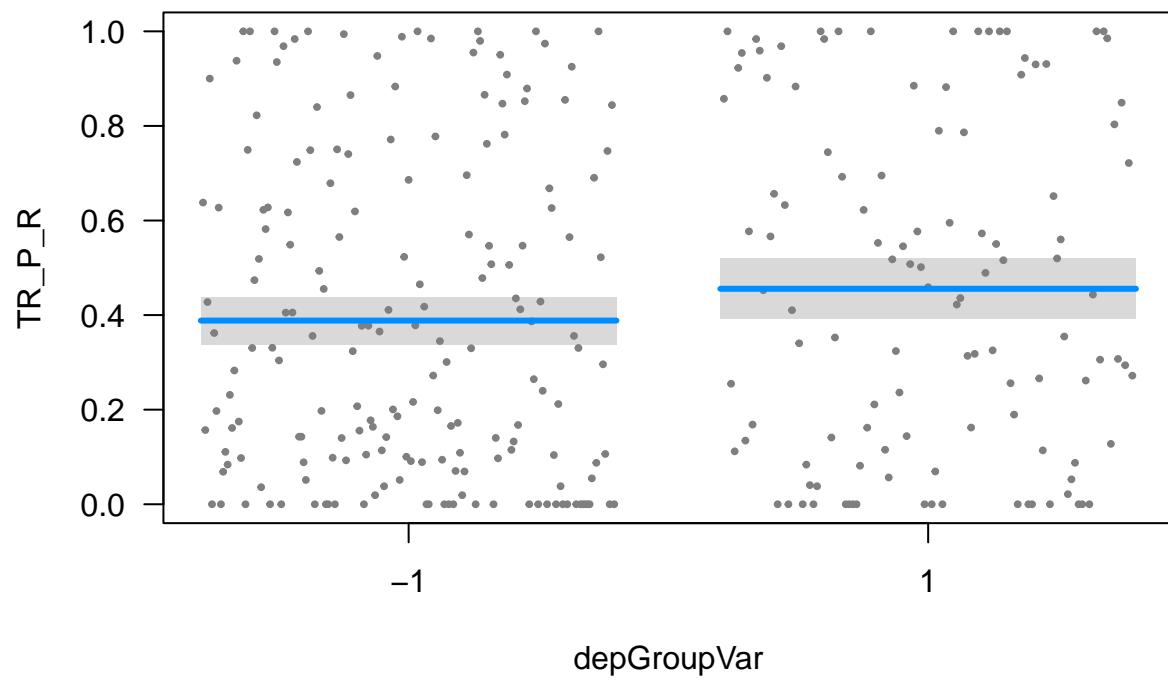


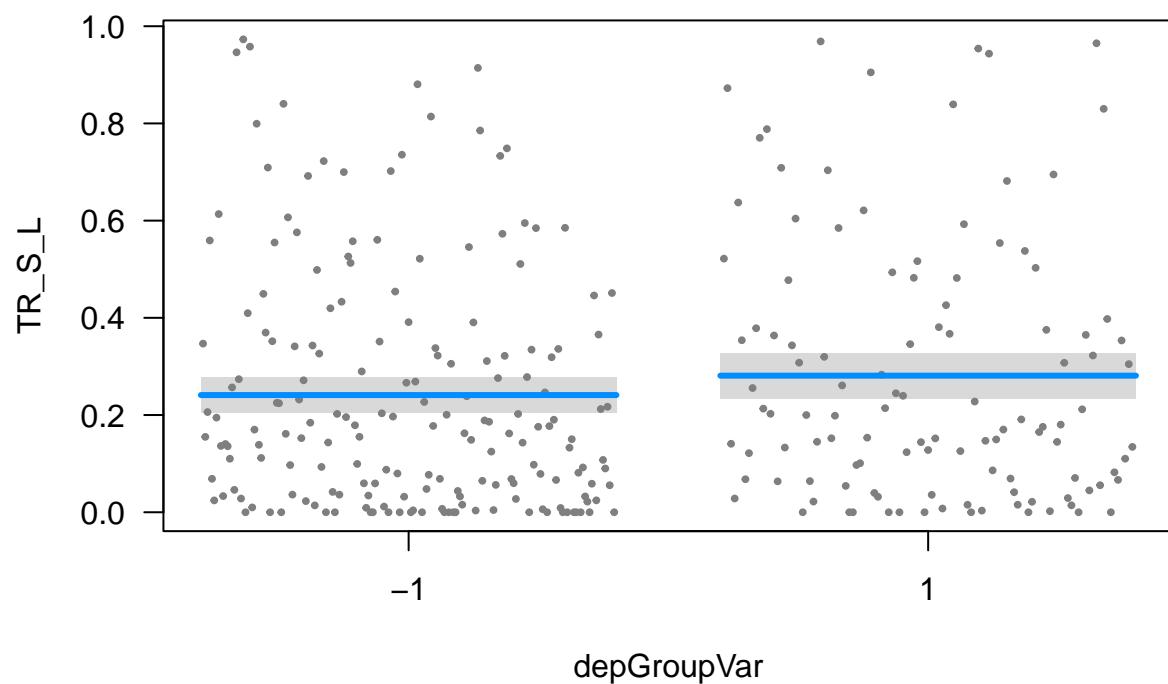


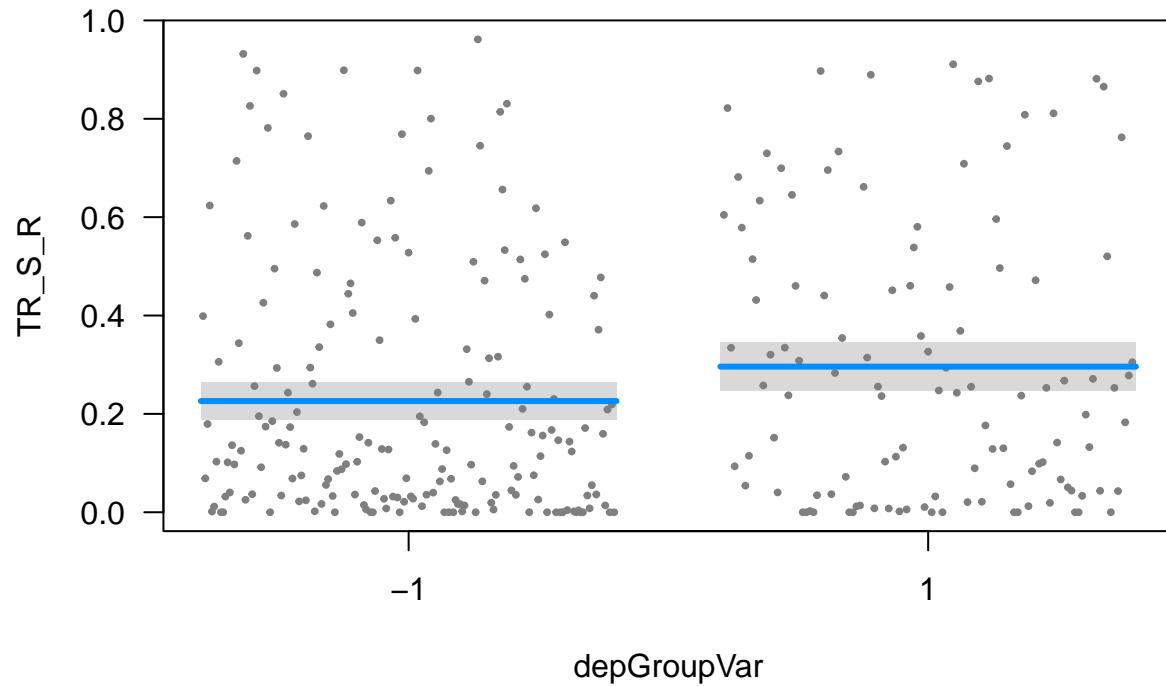


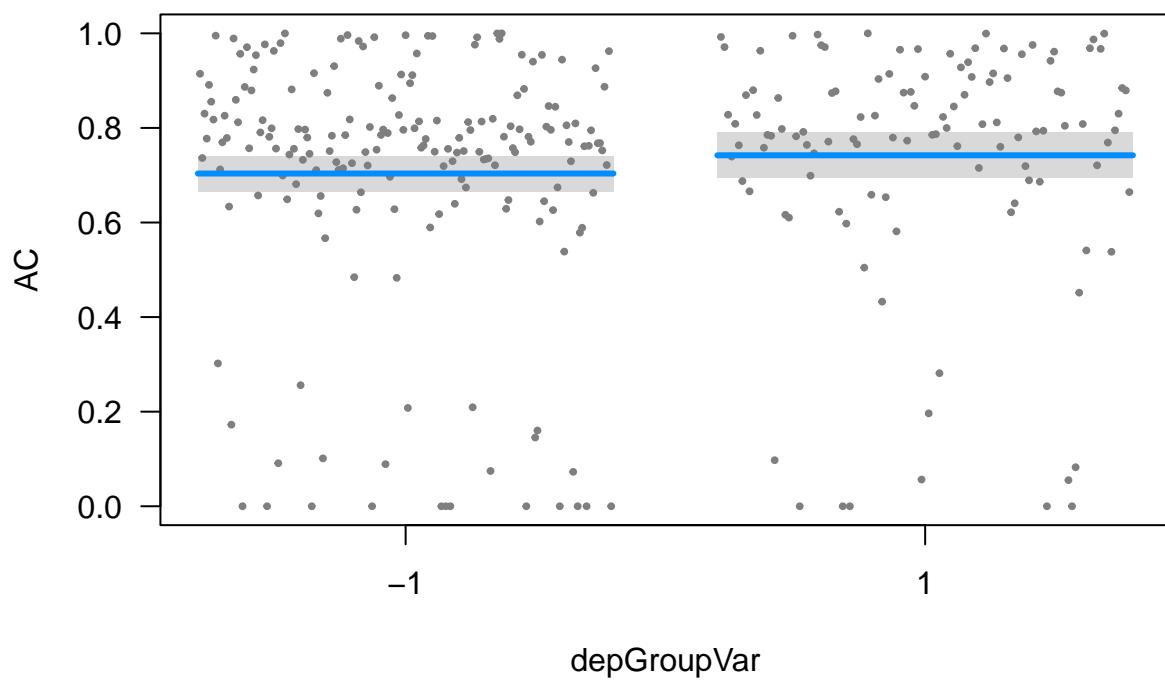


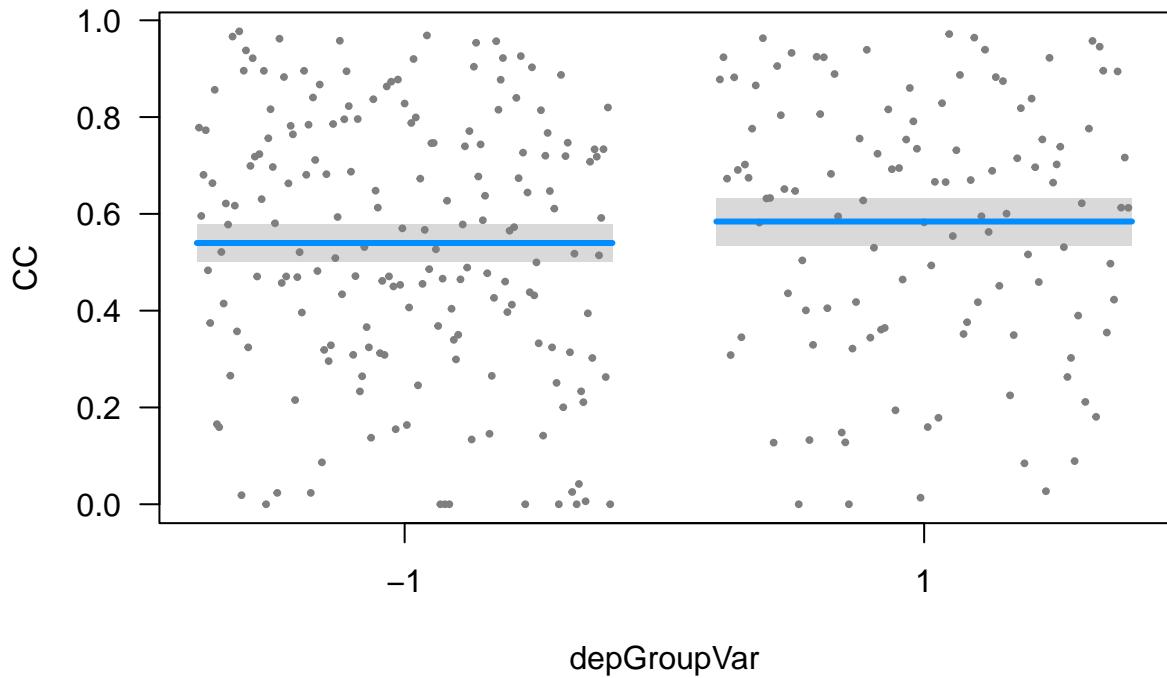












```

##      AF_L          AF_R          C_FPH_L          C_FPH_R          C_FP_L
## fit  data.frame,5 data.frame,5 data.frame,5 data.frame,5 data.frame,5
## res  data.frame,4 data.frame,4 data.frame,4 data.frame,4 data.frame,4
## meta list,6      list,6      list,6      list,6      list,6
##      C_FP_R          C_PH_L          C_PHP_L          C_PHP_R          C_PH_R
## fit  data.frame,5 data.frame,5 data.frame,5 data.frame,5 data.frame,5
## res  data.frame,4 data.frame,4 data.frame,4 data.frame,4 data.frame,4
## meta list,6      list,6      list,6      list,6      list,6
##      C_R_L          C_R_R          EMC_L          EMC_R          FAT_L
## fit  data.frame,5 data.frame,5 data.frame,5 data.frame,5 data.frame,5
## res  data.frame,4 data.frame,4 data.frame,4 data.frame,4 data.frame,4
## meta list,6      list,6      list,6      list,6      list,6
##      FAT_R          IFOF_L          IFOF_R          ILF_L          ILF_R
## fit  data.frame,5 data.frame,5 data.frame,5 data.frame,5 data.frame,5
## res  data.frame,4 data.frame,4 data.frame,4 data.frame,4 data.frame,4
## meta list,6      list,6      list,6      list,6      list,6
##      MdLF_L          MdLF_R          PAT_L          PAT_R          SLF1_L
## fit  data.frame,5 data.frame,5 data.frame,5 data.frame,5 data.frame,5
## res  data.frame,4 data.frame,4 data.frame,4 data.frame,4 data.frame,4
## meta list,6      list,6      list,6      list,6      list,6
##      SLF1_R          SLF2_L          SLF2_R          SLF3_L          SLF3_R
## fit  data.frame,5 data.frame,5 data.frame,5 data.frame,5 data.frame,5
## res  data.frame,4 data.frame,4 data.frame,4 data.frame,4 data.frame,4
## meta list,6      list,6      list,6      list,6      list,6
##      UF_L          UF_R          VOF_L          VOF_R          CB_L
## fit  data.frame,5 data.frame,5 data.frame,5 data.frame,5 data.frame,5

```

```

## res data.frame,4 data.frame,4 data.frame,4 data.frame,4 data.frame,4
## meta list,6      list,6      list,6      list,6      list,6
##   CB_R          ICP_L      ICP_R      MCP       SCP
## fit data.frame,5 data.frame,5 data.frame,5 data.frame,5 data.frame,5
## res data.frame,4 data.frame,4 data.frame,4 data.frame,4 data.frame,4
## meta list,6      list,6      list,6      list,6      list,6
##   V            CNIII_L    CNIII_R    CNII_L    CNII_R
## fit data.frame,5 data.frame,5 data.frame,5 data.frame,5 data.frame,5
## res data.frame,4 data.frame,4 data.frame,4 data.frame,4 data.frame,4
## meta list,6      list,6      list,6      list,6      list,6
##   CNVIII_L    CNVIII_R   CNVII_L    CNVII_R   CNV_L
## fit data.frame,5 data.frame,5 data.frame,5 data.frame,5 data.frame,5
## res data.frame,4 data.frame,4 data.frame,4 data.frame,4 data.frame,4
## meta list,6      list,6      list,6      list,6      list,6
##   CNV_R        AR_L      AR_R       CBT_L     CBT_R
## fit data.frame,5 data.frame,5 data.frame,5 data.frame,5 data.frame,5
## res data.frame,4 data.frame,4 data.frame,4 data.frame,4 data.frame,4
## meta list,6      list,6      list,6      list,6      list,6
##   CPT_F_L     CPT_F_R    CPT_O_L    CPT_O_R    CPT_P_L
## fit data.frame,5 data.frame,5 data.frame,5 data.frame,5 data.frame,5
## res data.frame,4 data.frame,4 data.frame,4 data.frame,4 data.frame,4
## meta list,6      list,6      list,6      list,6      list,6
##   CPT_P_R     CS_A_L    CS_A_R     CS_P_L    CS_P_R
## fit data.frame,5 data.frame,5 data.frame,5 data.frame,5 data.frame,5
## res data.frame,4 data.frame,4 data.frame,4 data.frame,4 data.frame,4
## meta list,6      list,6      list,6      list,6      list,6
##   CS_S_L      CS_S_R    CST_L      CST_R     DRTT_L
## fit data.frame,5 data.frame,5 data.frame,5 data.frame,5 data.frame,5
## res data.frame,4 data.frame,4 data.frame,4 data.frame,4 data.frame,4
## meta list,6      list,6      list,6      list,6      list,6
##   DRTT_R      F_L       F_R       ML_L      ML_R
## fit data.frame,5 data.frame,5 data.frame,5 data.frame,5 data.frame,5
## res data.frame,4 data.frame,4 data.frame,4 data.frame,4 data.frame,4
## meta list,6      list,6      list,6      list,6      list,6
##   OR_L        OR_R      RST_L      RST_R     TR_A_L
## fit data.frame,5 data.frame,5 data.frame,5 data.frame,5 data.frame,5
## res data.frame,4 data.frame,4 data.frame,4 data.frame,4 data.frame,4
## meta list,6      list,6      list,6      list,6      list,6
##   TR_A_R      TR_P_L    TR_P_R    TR_S_L    TR_S_R
## fit data.frame,5 data.frame,5 data.frame,5 data.frame,5 data.frame,5
## res data.frame,4 data.frame,4 data.frame,4 data.frame,4 data.frame,4
## meta list,6      list,6      list,6      list,6      list,6
##   AC          CC
## fit data.frame,5 data.frame,5
## res data.frame,4 data.frame,4
## meta list,6      list,6

#anova
fascicle_anova_unique <- lapply(fascicle_lm_unique, anova)

#fdr corrected
fascicle_anova_unique_fdr <- fdr_anova_generic(fascicle_anova_unique, 1)

##           p_anova

```

```

## AF_L      0.02567250
## AF_R      0.04770448
## C_FPH_L   0.52391294
## C_FPH_R   0.65399543
## C_FP_L    0.48257224
## C_FP_R    0.45804086
## C_PH_L    0.01296162
## C_PHP_L   0.89250446
## C_PHP_R   0.69018950
## C_PH_R    0.46121161
## C_R_L     0.97178852
## C_R_R     0.21083969
## EMC_L     0.31552969
## EMC_R     0.79906299
## FAT_L     0.14705546
## FAT_R     0.07693380
## IFOF_L    0.08459977
## IFOF_R    0.33245803
## ILF_L     0.02549913
## ILF_R     0.16956127
## MdLF_L    0.37358728
## MdLF_R    0.24894845
## PAT_L     0.06478979
## PAT_R     0.01735310
## SLF1_L    0.03644751
## SLF1_R    0.59627268
## SLF2_L    0.23908639
## SLF2_R    0.06233292
## SLF3_L    0.18713491
## SLF3_R    0.17736128
## UF_L      0.67930840
## UF_R      0.57479717
## VOF_L     0.56658390
## VOF_R     0.01189975
## CB_L      0.86501258
## CB_R      0.81213034
## ICP_L     0.99405802
## ICP_R     0.27810364
## MCP       0.54939427
## SCP       0.04967510
## V         0.50289423
## CNIII_L   0.56195527
## CNIII_R   0.46653398
## CNII_L    0.09357141
## CNII_R    0.34501658
## CNVIII_L  0.68381315
## CNVIII_R  0.27648526
## CNVII_L   0.43136407
## CNVII_R   0.43136407
## CNV_L     0.38363115
## CNV_R     0.32577122
## AR_L      0.08576154
## AR_R      0.06881451
## CBT_L     0.13009678

```

```
## CBT_R      0.10177583
## CPT_F_L    0.16822470
## CPT_F_R    0.03750849
## CPT_O_L    0.06952521
## CPT_O_R    0.34891233
## CPT_P_L    0.56189636
## CPT_P_R    0.09021897
## CS_A_L     0.32434188
## CS_A_R     0.16230369
## CS_P_L     0.34417997
## CS_P_R     0.44656372
## CS_S_L     0.14708780
## CS_S_R     0.12134611
## CST_L      0.38231082
## CST_R      0.09567975
## DRTT_L     0.12146445
## DRTT_R     0.01457703
## F_L        0.02068663
## F_R        0.19322091
## ML_L       0.68628017
## ML_R       0.22670168
## OR_L       0.29395359
## OR_R       0.19239239
## RST_L      0.07766356
## RST_R      0.03199621
## TR_A_L     0.26505645
## TR_A_R     0.01705549
## TR_P_L     0.15036940
## TR_P_R     0.10361940
## TR_S_L     0.18944704
## TR_S_R     0.02805414
## AC         0.21650960
## CC         0.16499416

print(fascicle_anova_unique_fdr)

## [1] component p_FDR_corr
## <0 rows> (or 0-length row.names)
```