

Práctica 12.Comparación de medias (I)

Jesús Martín Fernández

Contenidos

1. Introducción	1
2. Comparación de medias independientes	2
2.1 Comprobación de supuestos de normalidad y homogeneidad de la varianza	3
2.2 T de Student para datos independientes	9
2.3 La U de Mann-Whitney	12
3. Comparación de medias para datos apareados	14
3.1 La T de Student para datos apareados	14
3.2 La prueba de Wilcoxon para datos apareados	15

1. Introducción

Las comparaciones de medias son técnicas estadísticas que buscan determinar si existen diferencias significativas entre los promedios de dos o más grupos. Son fundamentales en áreas como las ciencias sociales, la medicina y la economía, donde se evalúan los efectos de una variable independiente (como un tratamiento) sobre una variable dependiente (como rendimiento o salud). Las pruebas más comunes incluyen la prueba t de Student (para dos grupos) y el ANOVA (para tres o más grupos). Para que estas pruebas sean fiables, deben cumplirse condiciones como la normalidad, homogeneidad de varianzas e independencia de las observaciones.

Cuando no se cumplen las condiciones de normalidad y homogeneidad de varianzas, es necesario utilizar alternativas a las pruebas paramétricas, ya que estas asumen que los datos siguen una distribución normal y tienen varianzas iguales. Para evitar resultados incorrectos o sesgados, se pueden emplear pruebas no paramétricas, que no requieren estas suposiciones. Entre ellas, la prueba de Mann-Whitney compara las medianas de dos grupos independientes, la prueba de Wilcoxon es adecuada para datos pareados, y la prueba de Kruskal-Wallis permite comparar más de dos grupos sin requerir normalidad o homogeneidad de varianzas. Estas pruebas son útiles en situaciones donde los datos no cumplen las condiciones paramétricas.

En esta práctica vamos a aprender a evaluar la normalidad y homocedasticidad (igualdad de varianzas) de las distribuciones y a hacer comparaciones entre dos grupos de variables independientes

En primer lugar, vamos a preparar una base de datos de nuestro directorio de trabajo, `df_iam3_r`, que puedes obtener de la Carpeta de la Práctica 12 en el Aula Virtual:

```
#setwd()
#getwd ()

df_iam3<-read.csv ("df_iam3_r.csv")
```

Esta base recoge una serie de características de 984 sujetos. Se incluyeron características sociodemográficas (`fecha_nac`, `sex` y `clas_soc`), clínicas (`hta` , `DM`, `colesterol` , `salud`) y el hábito tabáquico (`fum`). En un momento en el tiempo (`t`) se recogió qué sujetos habían tenido un evento tipo infarto de miocardio (`iam1`). Posteriormente se siguió a los sujetos hasta otro punto en el tiempo. Tras ese punto `t`, sólo se recogió si el sujeto seguía fumando tras el primer infarto (`fum_p`) , la cifra de colesterol (`colesterol_p`), la percepción de salud posterior al infarto (`salud_p`) y la ocurrencia de un nuevo reinfarto (`iam2`). Se asume, de forma artificial, que los sujetos sin `iam1` no tienen nuevos infartos en el seguimiento.

Comprueba sus variables

```
head (df_iam3)
```

	fecha_nac	sex	alt	peso	imc	hta	fum	DM	colesterol	salud	clas_soc
1	1982-08-06	Mujer	159.9	56.9	22.25437	No	No	1	255	Bien	Media
2	1982-09-23	Varón	165.7	72.7	26.47826	Sí	No	0	192	Mal	Media
3	1939-03-04	Mujer	156.3	51.3	20.99904	No	No	0	188	Bien	Baja
4	1936-01-15	Varón	176.6	91.6	29.37068	Sí	No	0	174	Regular	Baja
5	1940-03-23	Varón	169.2	89.2	31.15761	No	No	0	140	Bien	Baja
6	1962-11-10	Mujer	151.7	45.0	19.55426	Sí	No	0	140	Bien	Media

	iam1	fum_p	colesterol_p	salud_p	iam2	edad	imc_r
1	No	No	255	Bien	No	41	Normal
2	No	No	195	Mal	No	41	Normal
3	No	No	189	Bien	No	85	Normal
4	No	No	182	Regular	No	88	Normal
5	No	No	140	Bien	No	84	Obesidad
6	No	No	144	Bien	No	61	Bajo

2. Comparación de medias independientes

La comparación de medias independientes para dos grupos se utiliza para determinar si existe una diferencia significativa entre las medias de dos grupos no relacionados. La prueba más

común es la prueba t de Student para muestras independientes, que compara las medias de ambos grupos, bajo los supuestos de normalidad de los datos, homogeneidad de varianzas e independencia entre las observaciones.

Vamos a comparar los niveles de colesterol en aquellos sujetos que tuvieron infarto y los que no

Lo primero será valorar los supuestos de normalidad y homogeneidad de la varianza en ambos grupos

2.1 Comprobación de supuestos de normalidad y homogeneidad de la varianza

La comprobación de los supuestos de normalidad y homogeneidad de varianzas es crucial antes de aplicar pruebas paramétricas como la prueba t de Student o el ANOVA. Para verificar la normalidad, se utilizan pruebas estadísticas como Shapiro-Wilk o Kolmogorov-Smirnov, que contrastan si los datos siguen una distribución normal. También se puede evaluar visualmente mediante histogramas o gráficos Q-Q. Para comprobar la homogeneidad de varianzas, se aplican pruebas como Levene o Bartlett, que verifican si las varianzas entre los grupos son iguales.

Primero vamos a comprobar el supuesto de normalidad del colesterol

```
# Filtrar los datos por grupo
col_infarto <- df_iam3$colesterol[df_iam3$iam1 == "Sí"]
col_no_infarto <- df_iam3$colesterol[df_iam3$iam1 == "No"]

# Prueba de normalidad (Shapiro-Wilk) para el grupo con infarto
shapiro.test(col_infarto)
```

Shapiro-Wilk normality test

```
data:  col_infarto
W = 0.93274, p-value = 9.741e-06
```

```
# Prueba de normalidad (Shapiro-Wilk) para el grupo sin infarto
shapiro.test(col_no_infarto)
```

Shapiro-Wilk normality test

```
data:  col_no_infarto
W = 0.93298, p-value < 2.2e-16
```

Parece que en ninguno de los dos grupos la distribución es muy ajustada a la normal.

Como la muestra es grande, vamos a hacer la misma prueba de normalidad con el test de Kolmogorov-Smirnoff

```
# Prueba de normalidad (Kolmogorov-Smirnov) para el grupo con infarto
ks_test_infarto <- ks.test(col_infarto, "pnorm",
                           mean = mean(col_infarto), sd = sd(col_infarto))
```

```
Warning in ks.test.default(col_infarto, "pnorm", mean = mean(col_infarto), :
ties should not be present for the one-sample Kolmogorov-Smirnov test
```

```
ks_test_infarto
```

Asymptotic one-sample Kolmogorov-Smirnov test

```
data: col_infarto
D = 0.10713, p-value = 0.1135
alternative hypothesis: two-sided
```

```
# Prueba de normalidad (Kolmogorov-Smirnov) para el grupo sin infarto
ks_test_no_infarto <- ks.test(col_no_infarto, "pnorm",
                              mean = mean(col_no_infarto), sd = sd(col_no_infarto))
```

```
Warning in ks.test.default(col_no_infarto, "pnorm", mean =
mean(col_no_infarto), : ties should not be present for the one-sample
Kolmogorov-Smirnov test
```

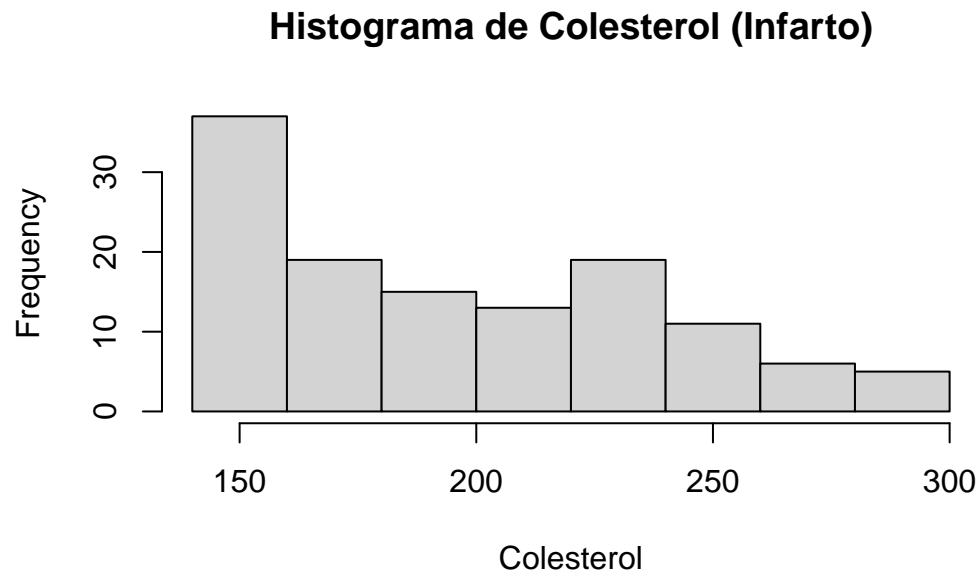
```
ks_test_no_infarto
```

Asymptotic one-sample Kolmogorov-Smirnov test

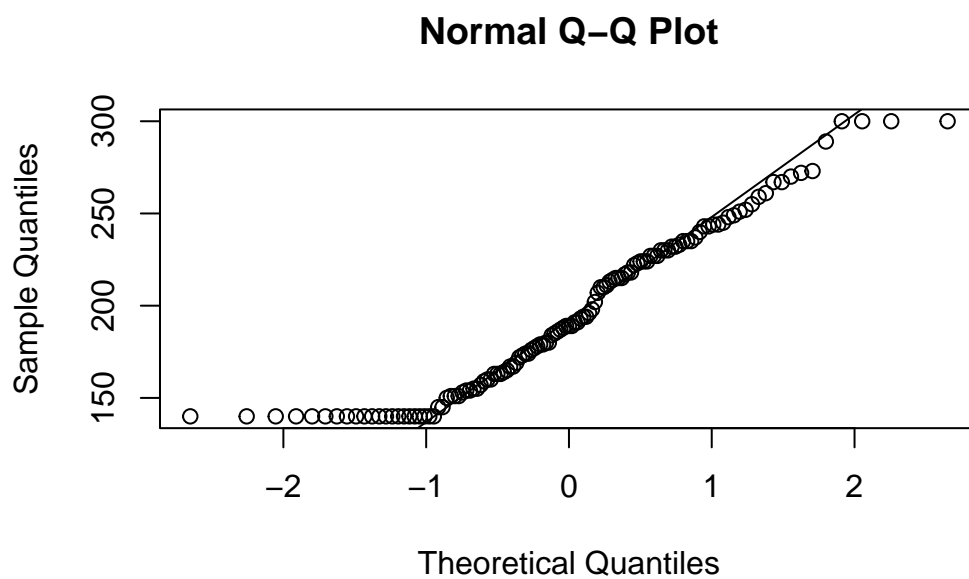
```
data: col_no_infarto
D = 0.11348, p-value = 4.933e-10
alternative hypothesis: two-sided
```

Vamos a representarlas gráficamente

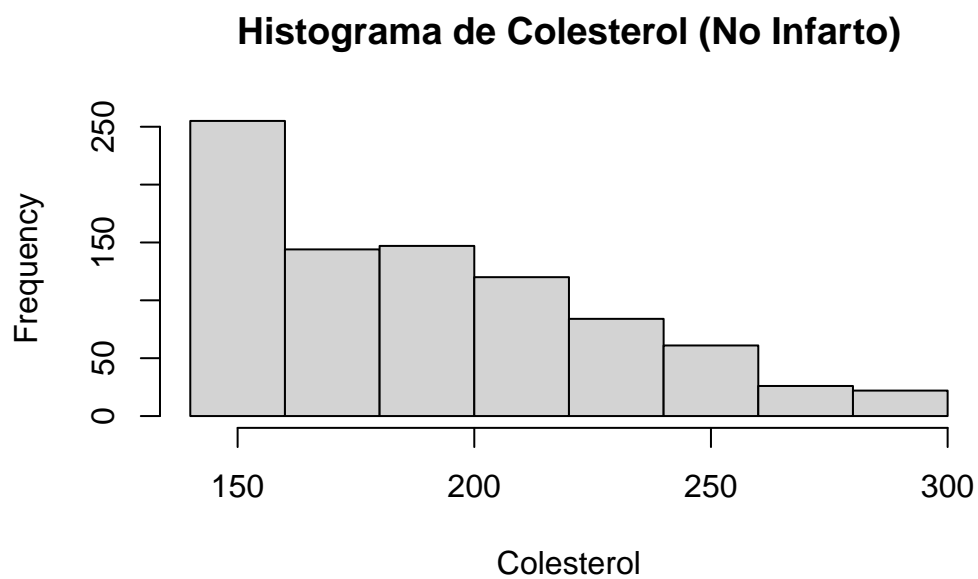
```
# Histograma y gráfico Q-Q para el grupo con infarto
hist(col_infarto,
     main = "Histograma de Colesterol (Infarto)",
     xlab = "Colesterol")
```



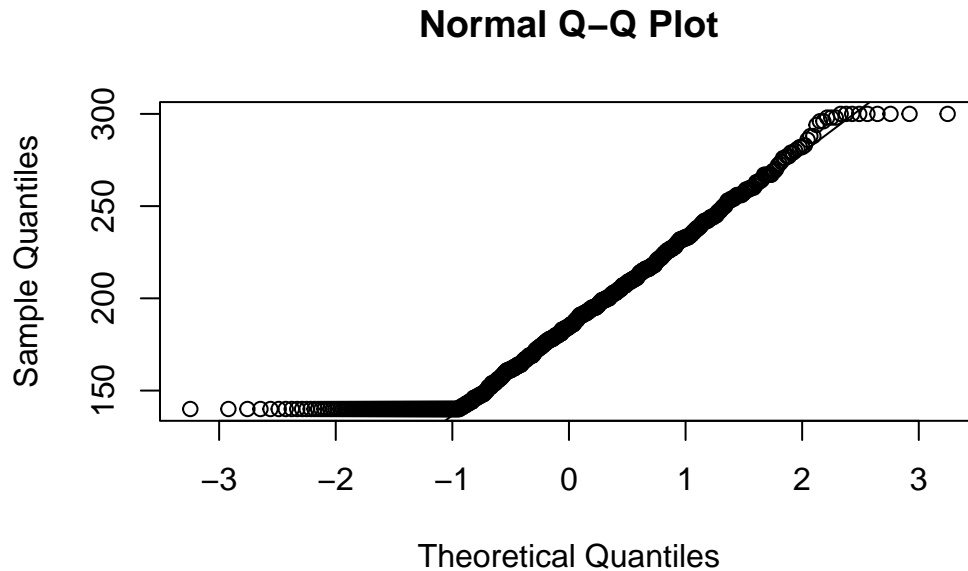
```
qqnorm(col_infarto)
qqline(col_infarto)
```



```
# Histograma y gráfico Q-Q para el grupo sin infarto  
hist(col_no_infarto, main = "Histograma de Colesterol (No Infarto)",  
xlab = "Colesterol")
```



```
qqnorm(col_no_infarto)
qqline(col_no_infarto)
```



Ahora vamos a testar la hipótesis de igualdad de las varianzas.

El test de Levene está implementado en el paquete `car`, y precisa que la variable que divide la muestra sea una variable factor

```
#install.packages("car")
library(car)
```

Warning: package 'car' was built under R version 4.4.1

Cargando paquete requerido: carData

Warning: package 'carData' was built under R version 4.4.1

```
df_iam3$iam1 <- as.factor(df_iam3$iam1)

levene_result <- leveneTest(colesterol ~ iam1, data = df_iam3)
levene_result
```

```
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  1  3.6621 0.05595 .
      982
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

El p-value de 0,05595 es muy próximo al valor umbral, podríamos descartar la hipótesis de homogeneidad de varianzas

Otra opción para comprobar homogeneidad de varianzas en R base , sería la prueba F de homogeneidad de varianzas también conocida prueba F de Snedecor.

```
# Prueba F para comparar las varianzas entre hombres y mujeres
var_test <- var.test(colesterol ~ iam1, data = df_iam3)

# Mostrar resultados
print(var_test)
```

F test to compare two variances

```
data:  colesterol by iam1
F = 0.8289, num df = 858, denom df = 124, p-value = 0.148
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.6259846 1.0683710
sample estimates:
ratio of variances
 0.8289013
```

Este resultado no indica diferencias en las varianzas

Vamos a comprobarlo con el test de Barlett

```
bartlett_result <- bartlett.test(colesterol ~ iam1, data = df_iam3)
bartlett_result
```

Bartlett test of homogeneity of variances

```
data:  colesterol by iam1
Bartlett's K-squared = 1.9932, df = 1, p-value = 0.158
```


Este test, que es más adecuado para distribuciones normales, tampoco detecta diferencias en las varianzas.

2.2 T de Student para datos independientes

Como hemos señalado, la prueba t de Student para datos independientes compara las medias de dos grupos para determinar si hay diferencias significativas. Es adecuada cuando los datos en cada grupo siguen una distribución normal, las varianzas son homogéneas y las observaciones son independientes. Aunque es moderadamente robusta frente a desviaciones de la normalidad en muestras grandes, gracias al teorema central del límite, en muestras pequeñas puede verse afectada por datos no normales. Sin embargo, no es robusta frente a la heterocedasticidad (desigualdad de varianzas).

Hemos visto anteriormente que el colesterol no tiene una distribución normal en pacientes con y sin infarto, pero no se puede descartar la igualdad de varianzas (no era concluyente el test de Levene y no descartaba la hipótesis de igualdad de varianzas el de Barlett), por lo que vamos a comparar las medias del colesterol en pacientes con y sin infarto con una t de student para datos independientes.

```
t_test_result <- t.test(colesterol ~ iam1, data = df_iam3, var.equal = TRUE)
t_test_result
```

Two Sample t-test

```
data:  colesterol by iam1
t = -1.6074, df = 982, p-value = 0.1083
alternative hypothesis: true difference in means between group No and group Sí is
not equal to 0 95 percent confidence interval:
 -14.131574   1.405167
sample estimates:
mean in group No mean in group Sí
      189.3248      195.6880
```

Si no estamos convencidos de la igualdad de las varianzas podemos usar el test de Welch, una variación de la t de Student que no asume varianzas iguales, modificando la orden anterior

```
t_test_result2 <- t.test(colesterol ~ iam1, data = df_iam3, var.equal = FALSE)
t_test_result2
```

Welch Two Sample t-test

```
data:  colesterol by iam1
t = -1.4988, df = 155.39, p-value = 0.136
alternative hypothesis: true difference in means between group No and group Sí is
not equal to 95 percent confidence interval:
 -14.749676   2.023268
sample estimates:
mean in group No mean in group Sí
      189.3248      195.6880
```

Leyendo del final al principio vemos que la media de colesterol en el grupo sin infarto es de 189.32 en el grupo sin infarto y de 195.68 en el grupo con infarto. El p-value de esta diferencia es de 0.1083 (0.136 con el test de Welch), no permite descartar la igualdad de medias.

Para ver el verdadero valor de la p para la toma de decisiones vamos a calcular manualmente los valores de la distribución del colesterol en uno y otro grupo

```
library (psych)
```

Warning: package 'psych' was built under R version 4.4.1

Adjuntando el paquete: 'psych'

The following object is masked from 'package:car':

```
logit
```

```
statistics_infarto <- describe (col_infarto, na.rm = TRUE)
statistics_noinfarto <- describe (col_no_infarto, na.rm = TRUE)
```

```
statistics_infarto
```

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
X1	1	125	195.69	44.84	189	192.31	56.34	140	300	160	0.44	-0.8	4.01

```
statistics_noinfarto
```

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
X1	1	859	189.32	40.82	185	185.53	45.96	140	300	160	0.59	-0.39	1.39

Ahora vamos a calcular los valores del t-test manualmente con la función `tsum.test` del paquete BSDA

```
#install.packages("BSDA")
library(BSDA)
```

Warning: package 'BSDA' was built under R version 4.4.1

Cargando paquete requerido: lattice

Adjuntando el paquete: 'BSDA'

The following objects are masked from 'package:carData':

Vocab, Wool

The following object is masked from 'package:datasets':

Orange

```
test_result3 <- tsum.test(mean.x = 189.32,
  s.x = 40.82, n.x = 859, mean.y = 195.69, s.y = 44.84, n.y = 125)
test_result3
```

Welch Modified Two-Sample t-Test

```
data: Summarized x and y
t = -1.5004, df = 155.38, p-value = 0.1355
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -14.756463  2.016463
sample estimates:
mean of x mean of y
 189.32    195.69
```

Si los mismos valores procediesen de una muestra 10 veces más grandes ($n_1=8590$ y $n_2=1250$), observa qué pasaría con el p-value

```
test_result4 <- tsum.test(mean.x = 189.32, s.x = 40.82,  
                          n.x = 8590, mean.y = 195.69, s.y = 44.84, n.y = 1250)  
test_result4
```

Welch Modified Two-Sample t-Test

```
data: Summarized x and y  
t = -4.7447, df = 1565.1, p-value = 2.279e-06  
alternative hypothesis: true difference in means is not equal to 0  
95 percent confidence interval:  
 -9.003414 -3.736586  
sample estimates:  
mean of x mean of y  
 189.32    195.69
```

El valor de la diferencia de medias sería el mismo ($189.32 - 195.69 = -6.37$ mg/dl), con un IC del 95% mucho más estrecho (-9.00 a -3.74), pero el p-value es muy, muy pequeño. Por lo tanto, al aumentar la n , el p-value cambia, mientras que no lo hace la diferencia de medias, que es la que nos da una información útil para la toma de decisiones.

2.3 La U de Mann-Whitney

Ahora vamos a usar la prueba de Mann-Whitney, el equivalente no paramétrico de la prueba t de dos muestras, que se emplea se usa cuando no se asume normalidad en los datos. Vamos a comparar la distribución del `imc` en hombres y mujeres. Utilizaremos la función `wilcox.test()`, que es la función adecuada para esta prueba no paramétrica. Primero comprobaremos la asunción de normalidad para esta variable

```
#Normalidad  
grupo_varon <- df_iam3$imc[df_iam3$sex == "Varón"]  
grupo_mujer <- df_iam3$imc[df_iam3$sex == "Mujer"]  
  
# Prueba de Shapiro-Wilk para normalidad en el grupo de hombres  
shapiro_varon <- shapiro.test(grupo_varon)  
# Prueba de Shapiro-Wilk para normalidad en el grupo de mujeres  
shapiro_mujer <- shapiro.test(grupo_mujer)
```

```
# Mostrar los resultados
print(shapiro_varon)
```

Shapiro-Wilk normality test

```
data: grupo_varon
W = 0.9656, p-value = 2.995e-09
```

```
print(shapiro_mujer)
```

Shapiro-Wilk normality test

```
data: grupo_mujer
W = 0.9594, p-value = 1.757e-10
```

```
# Prueba de Levene para comparar varianzas entre hombres y mujeres
levene_test <- leveneTest(imc ~ sex, data = df_iam3)
```

Warning in leveneTest.default(y = y, group = group, ...): group coerced to factor.

```
# Mostrar los resultados
print(levene_test)
```

Levene's Test for Homogeneity of Variance (center = median)

	Df	F value	Pr(>F)
group	1	14.699	0.0001342 ***
	982		

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

Vemos que la distribución no es normal , ni las varianzas homogéneas. Está justificado usar la prueba de Mann-Whitney

```
MW_test<- wilcox.test(imc ~ sex, data = df_iam3)
MW_test
```

Wilcoxon rank sum test with continuity correction

```
data: imc by sex
W = 110972, p-value = 0.02427
alternative hypothesis: true location shift is not equal to 0
```

La salida de esta función nos dice que las distribuciones son diferentes, está basada en los rangos. Esto puede implicar que una mediana es significativamente mayor o menor que la otra. Veamos las medianas

```
medianas_imc <- by(df_iam3$imc, df_iam3$sex, median, na.rm = TRUE)

# Mostrar resultados
print(medianas_imc)
```

```
df_iam3$sex: Mujer
[1] 26.18833
```

```
-----
df_iam3$sex: Varón
[1] 27.05867
```

3. Comparación de medias para datos apareados

Las pruebas de comparación de medias para datos apareados se utilizan para analizar las diferencias entre dos mediciones realizadas sobre el mismo grupo de individuos, como en un diseño antes-después. En este enfoque, se asume que la diferencia entre las dos observaciones sigue una distribución normal. El objetivo es determinar si existe un cambio significativo entre ambas mediciones, contrastando la hipótesis nula de que no hay diferencia (media de las diferencias igual a cero). Si las diferencias siguen una distribución normal, se aplica la prueba t de Student para datos apareados, que es la más adecuada en este contexto, ya que el estadístico de contraste sigue esta distribución.

3.1 La T de Student para datos apareados

Vamos a realizar una prueba T de Student para datos apareados comparando el colesterol medido antes y después del punto t (cuando se evidenció si había habido o no un infarto, `colesterol` y `colesterol_p`)

```
#Evaluamos normalidad de las distribución de las diferencias

diferencias <- df_iam3$colesterol - df_iam3$colesterol_p

# Test de normalidad (Shapiro-Wilk)
shapiro.test(diferencias)
```

Shapiro-Wilk normality test

```
data:  diferencias
W = 0.74283, p-value < 2.2e-16
```

Las distribuciones no son normales , pero sabemos que cuando la n es suficientemente grande, la t de student admite desviaciones de la normalidad , por lo que nos permitimos utilizar la t de student para datos apareados

```
t.test(df_iam3$colesterol, df_iam3$colesterol_p, paired = TRUE)
```

Paired t-test

```
data:  df_iam3$colesterol and df_iam3$colesterol_p
t = 7.1395, df = 983, p-value = 1.819e-12
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
 1.808424 3.179381
sample estimates:
mean difference
 2.493902
```

Nos dice que el valor de la p es muy pequeño (1.819e-12), pero además nos ofrece cuál es la diferencia de medias y su IC del 95% (2.493902, IC95%: 1.808424 a 3.179381 mg/dl, colesterol es mayor que colesterol_p), que nos ofrece una perspectiva más útil para la toma de decisiones.

3.2 La prueba de Wilcoxon para datos apareados

La prueba de Wilcoxon para datos apareados es una prueba no paramétrica utilizada cuando no se puede asumir la normalidad de las diferencias entre dos mediciones relacionadas, como

ocurría en el caso anterior. La prueba de Wilcoxon evalúa si las diferencias entre las dos muestras tienden a ser simétricamente distribuidas alrededor de la mediana.

Vamos a utilizar esta prueba para determinar si las medias del colesterol son iguales cuando se midieron antes y después del punto t

```
wilcox.test(df_iam3$colesterol, df_iam3$colesterol_p, paired = TRUE)
```

Wilcoxon signed rank test with continuity correction

```
data: df_iam3$colesterol and df_iam3$colesterol_p  
V = 193229, p-value = 0.002578  
alternative hypothesis: true location shift is not equal to 0
```

Nos dice que las diferencias existen y que es muy improbable que se deba al azar, pero no analiza cómo de grande es esa diferencia.