Two Sample t-test

Juan Pablo Silvestre y Sebastián López

2024-10-23

Introducción

A continuación vamos a realizar algunos ejercicios de pruebas de hipótesis de comparación de medias.

Realizaremos 6 ejercicios: 2 de two-sample, 2 de Welch t-test y 2 Paired t-test. De cada uno, 1 será con datos y el otro sera solo la informacion muestral.

Recordemos que para ejecutar una prueba de comparación de medias es necesario primero validar si las poblaciones son homocedásticas entre sí, es decir, tienen la misma varianza. Para ello, utilizaremos la prueba de hipótesis F-test par hacerlo; igualmente, recordemos que esta prueba la programamos en la clase anterior.

La prueba F-test es guardarda en un fichero llamado functions.R.

```r
# Cargar las funciones
source("functions.R")
```

Ejercicio 1

Comparación de tiempos de procesamiento en dos máquinas (Two Sample t-Test).

Dos tipos de máquinas producen el mismo componente, pero se sospecha que tienen diferencias en sus tiempos de procesamiento. A continuación se presentan los tiempos (en minutos) de 10 muestras de cada máquina:

```r
m.a = c(12.5, 13.2, 11.8, 14.1, 12.9, 13.5, 11.7, 12.6, 13.3, 12.8)
m.b = c(11.3, 12.7, 11.5, 10.9, 12.2, 11.1, 12.4, 11.6, 10.8, 11.9)
```

¿Existe una diferencia significativa en los tiempos de procesamiento entre las dos máquinas? Utiliza un nivel de significancia del 5 %?

Solución

De acuerdo al enunciado, las hipótesis son:

H0: el tiempo de procesamiento medio real/poblacional de m.a = tiempo procesamiento medio real/poblacional de m.b (Ho: mu.a = mu.b —- mu.a - mu.b = 0).\ Ha: el tiempo de procesamiento medio real/poblacional de m.a != tiempo procesamiento medio real/poblacional de m.b (Ha: mu.a != mu.b —- mu.a - mu.b != 0).

Antes de ejecutar la prueba debemos validar si poblacionalmente las varianzas reales/poblacionales de los tiempos del procesamiento de las máquinas son iguales o no. Para ello debemos ejecutar una prueba F, la cual define las hipótesis de la siguiente forma:

Ho: varianza real/poblacional del tiempo del procesamiento de m.a = varianza real/poblacional del tiempo del procesamiento de m.b (H0: var.a = var.b). \ Ha: varianza real/poblacional del tiempo del procesamiento de m.a != varianza real/poblacional del tiempo del procesamiento de m.b (Ha: var.a != var.b).

Dado lo anterior, es una prueba de dos colas.

```r
alpha = 0.05

var.test(m.a, m.b, ratio =1, alternative ="two.sided", conf.level = 1 - alpha)
```

```
##
##  F test to compare two variances
##
## data:  m.a and m.b
## F = 1.3082, num df = 9, denom df = 9, p-value = 0.6955
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.3249341 5.2667362
## sample estimates:
## ratio of variances
##           1.308183
```

Con una confianza del 95%, NO hay evidencia muestral suficiente para rechazar la hipótesis que las varianzas poblacionales son iguales, esto dado al que el p-vakue = 0.6555 y no es pequeño ni menor que alpha = 0.05. Por lo tanto, se evidencia que a y b tienen la misma varianza.

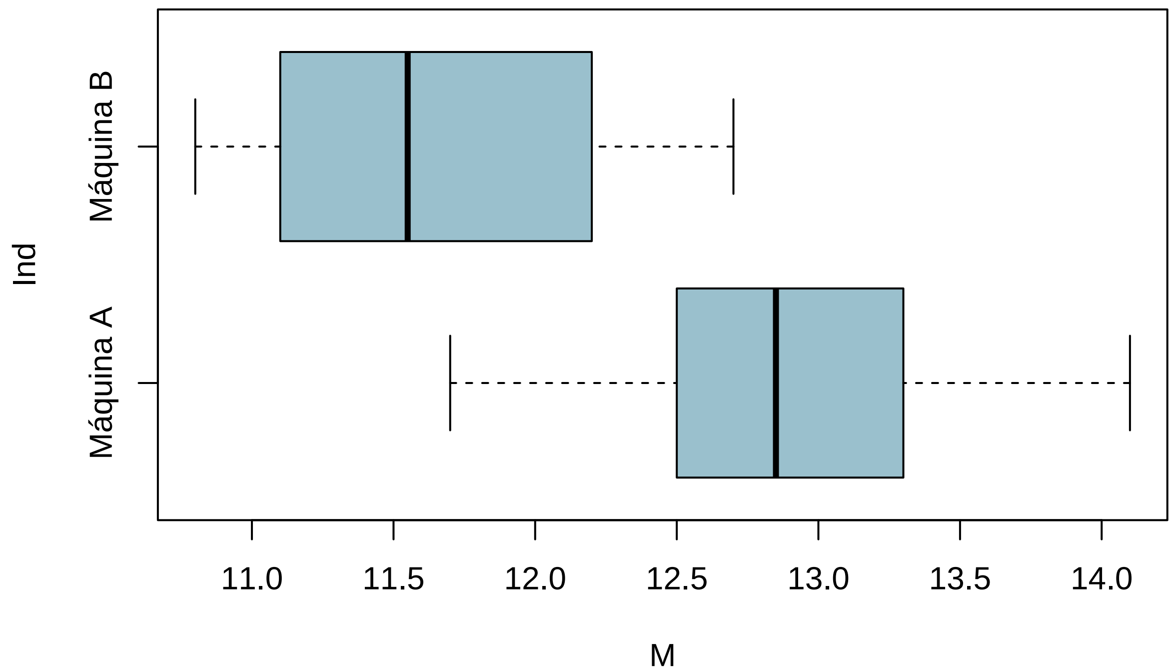
Dado lo anterior se ejecuta el two-sample t-test (no es necesario usar la correcion de welch-satterwaite)

```r
t.test(m.a, m.b, mu = 0, alternative = "two.sided", conf.level = 1 - alpha, var.equal = T)
```

```
##
##  Two Sample t-test
##
## data:  m.a and m.b
## t = 3.8623, df = 18, p-value = 0.001141
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.5472474 1.8527526
## sample estimates:
## mean of x mean of y
##     12.84     11.64
```

De acuerdo con la prueba, con una confianza del 95% HAY evidencia muestral suficiente para rechazar la hipótesis nula de que los tiempos medios reales de procesamiento de las máquinas a y b son iguales. Lo anterior, dado a que el p-value = 0.0011 que es menor y pequeño respecto a alpha = 0.05. Por lo tanto, la evidencia muestral sugiere que los tiempos de procesamiento de las máquinas son diferentes.

```r
M = c(m.a, m.b)
Ind = c(rep("Máquina A", length(m.a)), rep("Máquina B", length(m.b)))
boxplot(M ~ Ind, col = "lightblue3", horizontal = T)
```



Ejercicio 2

Comparación de niveles de glucosa en dos grupos de pacientes (Welch Two Sample t-Test)

En un estudio de diabetes, se compararon los niveles de glucosa en dos grupos de pacientes (20 en cada grupo). Los datos de glucosa (en mg/dL) para los grupos son:

```
g1 = c(105, 110, 115, 107, 112, 108, 115, 117, 109, 114, 106, 113, 118, 116, 119, 120,
108, 109, 110, 112)
g2 = c(128, 130, 133, 127, 135, 131, 136, 134, 129, 130, 132, 133, 135, 137, 129, 128,
130, 131, 132, 134)
```

Determina si existe una diferencia significativa en los niveles de glucosa entre los dos grupos usando un nivel de significancia del 1 %.

Solución

De acuerdo al enunciado, las hipótesis son:

H0: el nivel de glucosa media real/poblacional de g1 = el nivel de glucosa media real/poblacionalde g2 (H0: g1 = g2, g1 - g2 = 0) \

Ha: el nivel de glucosa media real/poblacional de g1 != el nivel de glucosa media real/poblacionalde g2 (Ha: g1 != g2, g1 - g2 != 0)

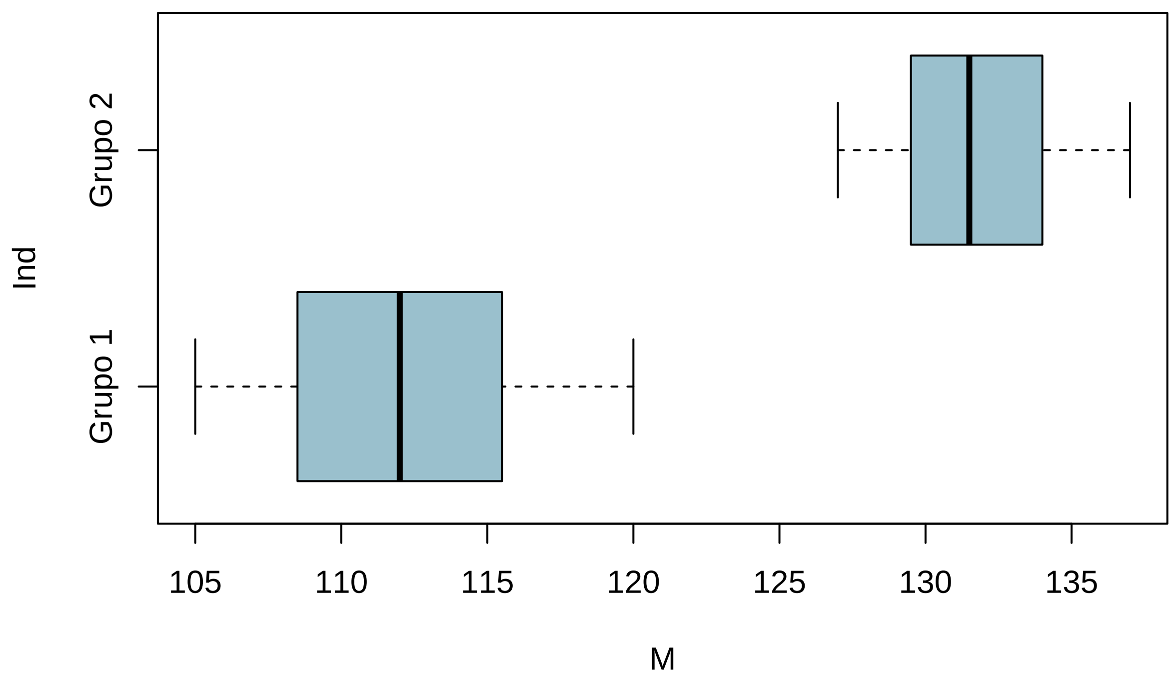
Antes de ejecutar la prueba debemos validar si poblacionalmente las varianzas reales/poblacionales del nivel de glucosa de los dos grupos son iguales o no. Para ello debemos ejecutar una prueba F, la cual define las hipótesis de la siguiente forma:

H0: varianza real/poblacional del nivel de glucosa media real/poblacional de g1 = varianza real/poblacional del varianza real/poblacional del nivel de glucosa media real/poblacional de g1 (H0: var.g1 = var.g2). \

Ha: varianza real/poblacional del nivel de glucosa media real/poblacional de g1 != varianza real/poblacional del varianza real/poblacional del nivel de glucosa media real/poblacional de g1 (Ha: var.g1 != var.g2).

Respecto a lo anterior, Es una prueba de dos colas.

```
M = c(g1, g2)
Ind = c(rep("Grupo 1", length(g1)), rep("Grupo 2", length(g2)))
boxplot(M ~ Ind, col = "lightblue3", horizontal = T)
```




```
alpha = 0.01

var.test(g1, g2, ratio =1, alternative ="two.sided", conf.level = 1 - alpha)
```

```
##
##  F test to compare two variances
##
## data:  g1 and g2
## F = 2.4363, num df = 19, denom df = 19, p-value = 0.05936
## alternative hypothesis: true ratio of variances is not equal to 1
## 99 percent confidence interval:
##  0.7099292 8.3607717
## sample estimates:
## ratio of variances
##            2.4363
```

Dado lo anterior se ejecuta el two-sample t-test (no es necesario usar la correción de Welch-satterwaite)

```r
alpha = 0.01

t.test(g1, g2, mu = 0, alternative = "two.sided", conf.level = 1 - alpha, var.equal = T)
```

```
##
##  Two Sample t-test
##
## data:  g1 and g2
## t = -16.449, df = 38, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99 percent confidence interval:
##  -22.77265 -16.32735
## sample estimates:
## mean of x mean of y
##    112.15    131.70
```

Ejercicio 3

Comparación de concentración de partículas en dos áreas industriales (Welch Two Sample t-Test)

Se midió la concentración de partículas (en ug/m^3?) en dos áreas industriales diferentes:

```r
a1 = c(120, 130, 128, 135, 140, 125, 132, 138, 129, 1331)
a2 = c(150, 158, 160, 155, 162, 157, 159, 163, 161, 154)
```

¿Hay una diferencia significativa en la concentración de partículas entre las dos áreas? Use un nivel de significancia del 5 %.

Solución

```r
alpha = 0.05

var.test(a1, a2, ratio =1, alternative ="two.sided", conf.level = 1 - alpha)
```

```
##
##  F test to compare two variances
##
## data:  a1 and a2
## F = 8949.6, num df = 9, denom df = 9, p-value = 2.22e-16
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##   2222.955 36031.053
## sample estimates:
## ratio of variances
##           8949.604
```

```r
t.test(g1, g2, mu = 0, alternative = "two.sided", conf.level = 1 - alpha, var.equal = T)
```

```
##
##  Two Sample t-test
##
## data:  g1 and g2
## t = -16.449, df = 38, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
```

```
##  -21.95596 -17.14404
## sample estimates:
## mean of x mean of y
##    112.15    131.70
```

Ejercicio 5

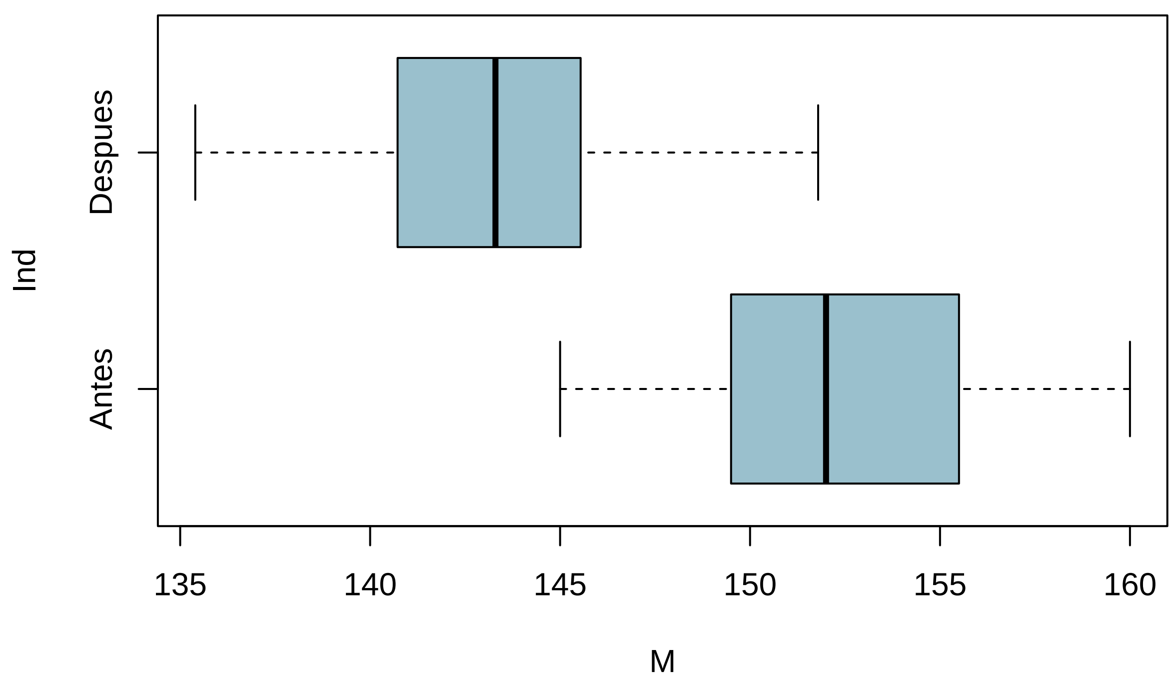
Comparación de presión arterial sistólica en pacientes pre y post tratamiento (Paired t-Test)

Se mide la presión arterial sistólica de 15 pacientes antes y después de un tratamiento para la hipertensión:

```
t1 = c(150, 148, 155, 160, 145, 158, 152, 149, 156, 147, 150, 154, 159, 151, 153)
t2 = c(140, 138, 145, 150, 135, 148, 142, 139, 146, 137, 140, 144, 149, 141, 143) + rnorm(15)
```

Realice un *paired t-test* para determinar si el tratamiento ha reducido significativamente la presión arterial.

```
M = c(t1, t2)
Ind = c(rep("Antes", length(t1)), rep("Despues", length(t2)))
boxplot(M ~ Ind, col = "lightblue3", horizontal = T)
```




```
alpha = 0.01

var.test(g1, g2, ratio =1, alternative ="two.sided", conf.level = 1 - alpha)
```

```
##
##  F test to compare two variances
##
## data:  g1 and g2
## F = 2.4363, num df = 19, denom df = 19, p-value = 0.05936
```

```
## alternative hypothesis: true ratio of variances is not equal to 1
## 99 percent confidence interval:
##  0.7099292 8.3607717
## sample estimates:
## ratio of variances
##            2.4363
```

De la anterior gráfica podeos observar que hay evidencia muestral que indica un efecto de tratamiento. Lo anterior dado a que la presión sistólica disminuyó luego que los pacientes se sometieron al tratamiento. Vamos ahora a validar poblacionalmente esto.

H0 : La presión sistolica media real antes del tratamiento = presión sistólica media real despues del tratamiento (H0: mu.antes <= mu.después) \

Ha : La presión sistolica media real antes del tratamiento != presión sistólica media real despues del tratamiento (Ha: mu.antes > mu.después)

Luego, es una prueba de cola derecha.

```r
alpha = 0.05
t.test(t1, t2, mu = 0, conf.level = 1 - alpha, paired = T, alternative = "greater", var.equal = F)
```

```
##
##  Paired t-test
##
## data:  t1 and t2
## t = 29.484, df = 14, p-value = 2.65e-14
## alternative hypothesis: true mean difference is greater than 0
## 95 percent confidence interval:
##  8.974132      Inf
## sample estimates:
## mean difference
##        9.544288
```