Taller 4 AED

Maria Fernanda Palacio, David Santiago Flores Alsina, Estefanía Laverde

26/3/2022

1) Un investigador considera tres índices para medir la severidad de los ataques al corazón. Los valores de esos índices para n = 40 pacientes con ataque al corazón que llegan a las emergencias de un hospital producen las siguientes estadísticas resumidas

```
n = 40
q = 3
xbar = rbind(46.1, 57.3, 50.4)
S = rbind(c(101.3, 63.0, 71.0), c(63.0, 80.2, 55.6), c(71.0, 55.6, 97.4))
xbar
```

```
##      [,1]
## [1,] 46.1
## [2,] 57.3
## [3,] 50.4
```

```
S
```

```
##       [,1] [,2] [,3]
## [1,] 101.3 63.0 71.0
## [2,]  63.0 80.2 55.6
## [3,]  71.0 55.6 97.4
```

a) Los tres índices son evaluados para cada paciente. Realice una prueba para la igualdad de las medias de los índices con $\alpha = 0.05$.

```
C = rbind(c(1,-1,0), c(1,0,-1))
mmuestr = C%*%xbar
covmuestr = C%*%S%*%t(C)

Tcuadrado = n*t(mmuestr)%*%solve(covmuestr)%*%mmuestr

f = ((n-1)*(q-1)/(n-q+1))*qf(p=1-.05, df1=q-1, df2=n-q+1)
f
```

```
## [1] 6.660417
```

```
Tcuadrado
```

```
##          [,1]
## [1,] 90.49458
```

Como el valor de T^2 es mucho mayor que el de F, se rechaza la hipótesis de que las medias sean iguales.

b) Juzgue las diferencias entre pares de las medias de los índices usando intervalos de confianza (T^2) simultaneos del 95%.

```
#Intervalos de confianza mu1-mu2
uint1conf = t(C[1,])%*%xbar+sqrt(f)*sqrt(C[1,]%*%S%*%C[1,]/n)
lint1conf = t(C[1,])%*%xbar-sqrt(f)*sqrt(C[1,]%*%S%*%C[1,]/n)
x1_x2 = xbar[1]-xbar[2]

#Mostrar los intervalos1
print(c(lint1conf,uint1conf))
```

```
## [1] -14.239955  -8.160045
```

```
#Intervalos de confianza mu1-mu3
uint2conf = t(C[2,])%*%xbar+sqrt(f)*sqrt(C[2,]%*%S%*%C[2,]/n)
lint2conf = t(C[2,])%*%xbar-sqrt(f)*sqrt(C[2,]%*%S%*%C[2,]/n)
x1_x3 = xbar[1]-xbar[3]

#MOstrar los intervalos2
print(c(lint2conf,uint2conf))
```

```
## [1] -7.372644 -1.227356
```

```
#Intervalos de confianza mu2-mu3
uint3conf = t(c(0,1,-1))%*%xbar+sqrt(f)*sqrt(c(0,1,-1)%*%S%*%c(0,1,-1)/n)
lint3conf = t(c(0,1,-1))%*%xbar-sqrt(f)*sqrt(c(0,1,-1)%*%S%*%c(0,1,-1)/n)
x2_x3 = xbar[2]-xbar[3]

#MOstrar los intervalos3
print(c(lint3conf,uint3conf))
```

```
## [1]  3.5749 10.2251
```

El 0 no cae en ninguno de los intervalos entonces es evidente que si hay diferencia entre las medias, luego reafirmamos la decisión de rechazar H_0.

###2) Observaciones sobre dos respuestas fueron coleccionadas para dos tratamientos. Las observaciones vectoriales $[x_1, x_2]$ fueron

```
trat21 = rbind(3,3)
trat22 = rbind(1,6)
trat23 = rbind(2,3)

trat31 = rbind(2,3)
trat32 = rbind(5,1)
trat33 = rbind(3,1)
trat34 = rbind(2,3)
```

a) Calcule S_{pooled}

```
trat2 = cbind(trat21,trat22,trat23)
trat3 = cbind(trat31,trat32,trat33,trat34)
n2 = 3
n3 = 4
p = 2

S2 = cov(t(trat2))
S3 = cov(t(trat3))

x2bar = (1/n2)*(trat21+trat22+trat23)
x3bar = (1/n3)*(trat31+trat32+trat33+trat34)

Spooled = ((n2-1)*S2+(n3-1)*S3)/(n2+n3-2)
Spooled
```

```
##      [,1] [,2]
## [1,]  1.6 -1.4
## [2,] -1.4  2.0
```

b) Realice la prueba H_0: $\mu_2 - \mu_3 = 0$ usando un enfoque de dos muestras con $\alpha = .01$.

```
T2 = t(x2bar-x3bar)%*%solve(((1/n2)+(1/n3))*Spooled)%*%(x2bar-x3bar)
f = ((n2+n3-2)*p/(n2+n3-p-1))*qf(p=0.01, df1=p, df2=n2+n3-p-1, lower.tail = FALSE)

f
```

```
## [1] 45
```

```
T2
```

```
##          [,1]
## [1,] 3.870968
```

Con estos resultados, no se recahza H_0 con una confianza de $\alpha = 0.01$, luego es posible que haya una igualdad de medias.

c) Construya un intervalo de confianza simultaneo (T^2) del 99% para las diferencias $\mu_{2i} = \mu_{3i}, i = 1, 2$.

```
#Intervalo para i=1
uint1conf = (x2bar[1]-x3bar[1])+sqrt(T2)*sqrt(((1/n2)+(1/n3))*Spooled[1,1])
lint1conf = (x2bar[1]-x3bar[1])-sqrt(T2)*sqrt(((1/n2)+(1/n3))*Spooled[1,1])

print(c(lint1conf,uint1conf))
```

```
## [1] -2.9007639  0.9007639
```

```
#Intervalo para i=2
uint2conf = (x2bar[2]-x3bar[2])+sqrt(T2)*sqrt(((1/n2)+(1/n3))*Spooled[2,2])
lint2conf = (x2bar[2]-x3bar[2])-sqrt(T2)*sqrt(((1/n2)+(1/n3))*Spooled[2,2])

print(c(lint2conf,uint2conf))
```

```
## [1] -0.1251186  4.1251186
```

Como el valor 0 se encuentra en ambos intervalos, si es posible que haya igualdad de medias, como tambien vimos con la prueba de hipotesis.

3) Dados los datos

```
z1 = rbind(10,5,7,19,11,18)
z2 = rbind(2,3,3,6,7,9)
y = rbind(15,9,3,25,7,13)
```

a) ajuste el modelo de regresion lineal.

```
model = lm(as.numeric(y)~as.numeric(z1)+as.numeric(z2))
summary(model)
```

```
##
## Call:
## lm(formula = as.numeric(y) ~ as.numeric(z1) + as.numeric(z2))
##
## Residuals:
##       1       2       3       4       5       6
## -0.5945  4.5054 -5.0592  2.1180  0.5649 -1.5347
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.1480     4.2967   0.500   0.6515
## as.numeric(z1)   1.7823     0.4982   3.578   0.0374 *
## as.numeric(z2)  -2.1883     1.0329  -2.119   0.1243
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.219 on 3 degrees of freedom
## Multiple R-squared:  0.8183, Adjusted R-squared:  0.6972
## F-statistic: 6.757 on 2 and 3 DF,  p-value: 0.07743
```

b) Determine los intervalos de confianza del 95% simultaneos (uno a la vez) para β_1 y β_2.

```
confint.lm(model)
```

```
##                      2.5 %    97.5 %
## (Intercept)    -11.5259154 15.821867
## as.numeric(z1)   0.1968411  3.367791
## as.numeric(z2)  -5.4753736  1.098707
```

Los intervalos de confianza para β_1 son aquellos relacionados a z1 y los de β_2 los de z2.

c) Compruebe la prueba de hipotesis nula de que solo el coeficiente β_1 es cero.

Teniendo en cuenta que el p-valor para β_1 es de 0.0374 y que ademas en el intervalo de confianza no se encuentra el valor 0, se puede rechazar la hipotesis nula y asi decir que β_1 no es cero. Ahora, si tenemos en cuenta el p-valor asociado a z2, que es elevado, y que en su respectivo intervalo de confianza el 0 se encuentra alli, se puede llegar a pensar que el β_2 si sea 0.

d) Determine el valor esperado de la prediccion $(E(Y))$ para $z_1 = 6$ y $z_2 = 4$. Calcule su intervalo de confianza del 95% correspondiente (el del valor esperado).

```
new_data = data.frame(z1=6,z2=4)
predict.lm(model, new_data, level=0.95, interval="confidence")
```

```
##        fit       lwr      upr
## 1 4.088541 -4.705851 12.88293
```

e) Determine el intervalo de confianza del 95% para la prediccion (Y) cuando $z_1 = 6$ y $z_2 = 4$.

```
predict.lm(model, new_data, level=0.95, interval="prediction")
```

```
##        fit       lwr      upr
## 1 4.088541 -11.96283 20.13991
```

4) Con el dataset Boston,

a) Ajuste de regresion lineal

```
library(MASS)
data(Boston)
```

```
lstat <- Boston$lstat
medv <- Boston$medv

modelo_lineal <- lm(medv~lstat)
summary(modelo_lineal)
```

```
##
## Call:
## lm(formula = medv ~ lstat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -15.168  -3.990  -1.318   2.034  24.500
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 34.55384    0.56263   61.41   <2e-16 ***
## lstat       -0.95005    0.03873  -24.53   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.216 on 504 degrees of freedom
## Multiple R-squared:  0.5441, Adjusted R-squared:  0.5432
## F-statistic: 601.6 on 1 and 504 DF,  p-value: < 2.2e-16
```

La pendiente asociada al valor de lstat es -0.95005. Teniendo en cuenta su p valor asociado, y al de la prueba, podemos asegurar que dicha pendiente no es 0, aunque si es cercana a 0.

b) Intervalo de confianza del 95% para los coeficientes.

```
confint.lm(modelo_lineal)
```

```
##                 2.5 %     97.5 %
## (Intercept) 33.448457 35.6592247
## lstat       -1.026148 -0.8739505
```

c) Predicciones para el valor esperado de medv y los correspondientes intervalos de confianza del 95% para los valores de lstat=c(5,10,15).

Fit es la predicción, lwr and upr son los extremos del intervalo de confianza.

```
new_dat <- data.frame(lstat=5)
predict(modelo_lineal, new_dat, level=0.95, interval="confidence")
```

```
##        fit      lwr      upr
## 1 29.80359 29.00741 30.59978
```

```
predict(modelo_lineal, new_dat, level=0.95, interval="prediction")
```

```
##        fit      lwr      upr
## 1 29.80359 17.56567 42.04151
```

```
new_dat <- data.frame(lstat=10)
predict(modelo_lineal, new_dat, level=0.95, interval="confidence")
```

```
##        fit      lwr      upr
## 1 25.05335 24.47413 25.63256
```

```
predict(modelo_lineal, new_dat, level=0.95, interval="prediction")
```

```
##        fit      lwr      upr
## 1 25.05335 12.82763 37.27907
```

```
new_dat <- data.frame(lstat=15)
predict(modelo_lineal, new_dat, level=0.95, interval="confidence")
```

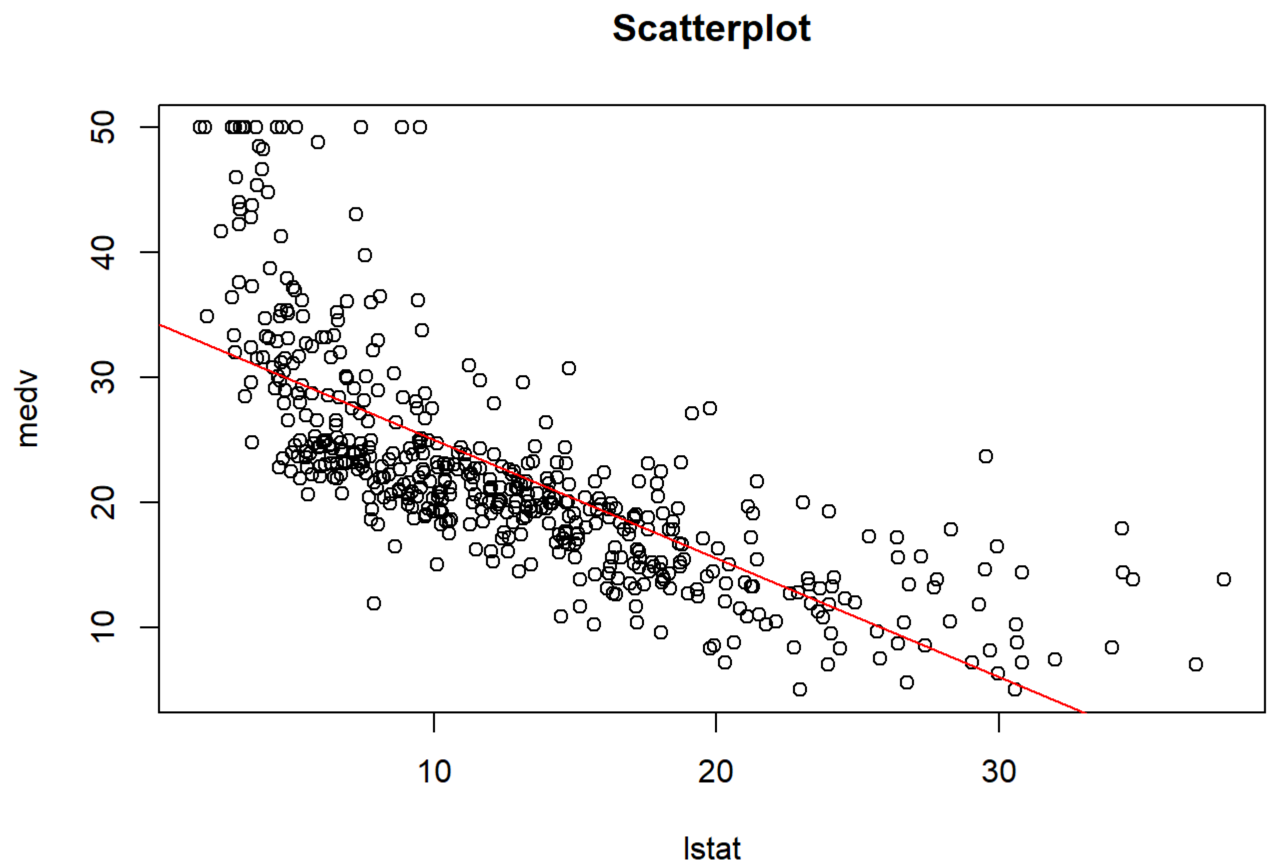
```
##       fit      lwr      upr
## 1 20.3031 19.73159 20.87461
```

```
predict(modelo_lineal, new_dat, level=0.95, interval="prediction")
```

```
##       fit      lwr      upr
## 1 20.3031 8.077742 32.52846
```

d) Grafique el diagrama de dispersion de medv y lstat y la recta de regresion

```
plot(lstat,medv, main = "Scatterplot")
abline(modelo_lineal, col = "red")
```



e) Realice la regresion utilizando todas las variables independientes.

```
#Ver cuales son las variables para guardarlas
summary(Boston)
```

```
##       crim                zn             indus            chas
##  Min.   : 0.00632   Min.   :  0.00   Min.   : 0.46   Min.   :0.00000
##  1st Qu.: 0.08205   1st Qu.:  0.00   1st Qu.: 5.19   1st Qu.:0.00000
##  Median : 0.25651   Median :  0.00   Median : 9.69   Median :0.00000
##  Mean   : 3.61352   Mean   : 11.36   Mean   :11.14   Mean   :0.06917
##  3rd Qu.: 3.67708   3rd Qu.: 12.50   3rd Qu.:18.10   3rd Qu.:0.00000
##  Max.   :88.97620   Max.   :100.00   Max.   :27.74   Max.   :1.00000
##       nox               rm             age              dis
##  Min.   :0.3850   Min.   :3.561   Min.   :  2.90   Min.   : 1.130
##  1st Qu.:0.4490   1st Qu.:5.886   1st Qu.: 45.02   1st Qu.: 2.100
##  Median :0.5380   Median :6.208   Median : 77.50   Median : 3.207
##  Mean   :0.5547   Mean   :6.285   Mean   : 68.57   Mean   : 3.795
##  3rd Qu.:0.6240   3rd Qu.:6.623   3rd Qu.: 94.08   3rd Qu.: 5.188
##  Max.   :0.8710   Max.   :8.780   Max.   :100.00   Max.   :12.127
##       rad              tax           ptratio          black
##  Min.   : 1.000   Min.   :187.0   Min.   :12.60   Min.   :  0.32
##  1st Qu.: 4.000   1st Qu.:279.0   1st Qu.:17.40   1st Qu.:375.38
##  Median : 5.000   Median :330.0   Median :19.05   Median :391.44
##  Mean   : 9.549   Mean   :408.2   Mean   :18.46   Mean   :356.67
##  3rd Qu.:24.000   3rd Qu.:666.0   3rd Qu.:20.20   3rd Qu.:396.23
##  Max.   :24.000   Max.   :711.0   Max.   :22.00   Max.   :396.90
##      lstat            medv
##  Min.   : 1.73   Min.   : 5.00
##  1st Qu.: 6.95   1st Qu.:17.02
##  Median :11.36   Median :21.20
##  Mean   :12.65   Mean   :22.53
##  3rd Qu.:16.95   3rd Qu.:25.00
##  Max.   :37.97   Max.   :50.00
```

```r
medv = Boston$medv
crim = Boston$crim
zn = Boston$zn
indus = Boston$indus
chas = Boston$chas
nox = Boston$nox
rm = Boston$rm
age = Boston$age
dis = Boston$dis
rad = Boston$rad
tax = Boston$tax
ptratio = Boston$ptratio
black = Boston$black
lstat = Boston$lstat

reg <- lm(medv~  crim +  zn +  indus +  chas +  nox +  rm +  age +  dis +  rad +  tax +  ptratio
+  black +  lstat)

summary(reg)
```

```
##
## Call:
## lm(formula = medv ~ crim + zn + indus + chas + nox + rm + age +
##     dis + rad + tax + ptratio + black + lstat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -15.595  -2.730  -0.518   1.777  26.199
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.646e+01  5.103e+00   7.144 3.28e-12 ***
## crim        -1.080e-01  3.286e-02  -3.287 0.001087 **
## zn           4.642e-02  1.373e-02   3.382 0.000778 ***
## indus        2.056e-02  6.150e-02   0.334 0.738288
## chas         2.687e+00  8.616e-01   3.118 0.001925 **
## nox         -1.777e+01  3.820e+00  -4.651 4.25e-06 ***
## rm           3.810e+00  4.179e-01   9.116  < 2e-16 ***
## age          6.922e-04  1.321e-02   0.052 0.958229
## dis         -1.476e+00  1.995e-01  -7.398 6.01e-13 ***
## rad          3.060e-01  6.635e-02   4.613 5.07e-06 ***
## tax         -1.233e-02  3.760e-03  -3.280 0.001112 **
## ptratio     -9.527e-01  1.308e-01  -7.283 1.31e-12 ***
## black        9.312e-03  2.686e-03   3.467 0.000573 ***
## lstat       -5.248e-01  5.072e-02 -10.347  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.745 on 492 degrees of freedom
## Multiple R-squared:  0.7406, Adjusted R-squared:  0.7338
## F-statistic: 108.1 on 13 and 492 DF,  p-value: < 2.2e-16
```

```
confint.lm(reg)
```

```
##                     2.5 %       97.5 %
## (Intercept)  26.432226009  46.486750761
## crim         -0.172584412  -0.043438304
## zn            0.019448778   0.073392139
## indus        -0.100267941   0.141385193
## chas          0.993904193   4.379563446
## nox         -25.271633564 -10.261588893
## rm            2.988726773   4.631003640
## age          -0.025262320   0.026646769
## dis          -1.867454981  -1.083678710
## rad           0.175692169   0.436406789
## tax          -0.019723286  -0.004945902
## ptratio      -1.209795296  -0.695699168
## black         0.004034306   0.014589060
## lstat        -0.624403622  -0.425113133
```

f) Intervalos de confianza y valor esperado con el promedio de las variables independientes. Intervalo de confianza para la prediccion.

```
new_dat <- data.frame( crim = mean( crim),  zn = mean( zn),  indus = mean( indus),  chas = mean( chas),  nox = mean( nox),  rm = mean( rm),  age = mean( age),  dis = mean( dis),  rad = mean( rad),  tax = mean( tax),  ptratio = mean( ptratio),  black = mean( black),  lstat = mean( lstat))
predict(reg, new_dat, level=0.95, interval="confidence") #Intervalo de conf
```

```
##        fit      lwr      upr
## 1 22.53281 22.11832 22.94729
```

```
predict(reg, new_dat, level=0.95, interval="prediction") #Intervalo de conf para la prediccion
```

```
##        fit      lwr      upr
## 1 22.53281 13.20005 31.86556
```

5) Se realizan observaciones de dos respuestas sobre tres tratamientos. Los vectores de observacion son:

```
Tratamiento1 = cbind(rbind(2,9),rbind(3,2),rbind(7,5),rbind(2,1),rbind(7,5))
Tratamiento2 = cbind(rbind(3,2),rbind(2,4),rbind(9,4))
Tratamiento3 = cbind(rbind(1,4),rbind(7,2),rbind(4,9),rbind(3,2))
```

a) Construya la tabla de one-way MANOVA

```
t1Bar = rowMeans(Tratamiento1)
t2Bar = rowMeans(Tratamiento2)
t3Bar = rowMeans(Tratamiento3)

tratamientos = cbind(Tratamiento1,Tratamiento2,Tratamiento3)
tbar = rowMeans(tratamientos)
totalObservations = 12
```

```
row1=0
row2=0
for (i in tratamientos[1,]) {
  row1 = row1+i^2
}
for (i in tratamientos[2,]) {
  row2 = row2+i^2
}

SSobs = cbind(row1,row2)
SSmean = totalObservations*cbind(tbar[1]^2,tbar[2]^2)
```

```r
row1=0
row2=0
for (i in tratamientos[1,]) {
  row1 = row1+i^2
}
for (i in tratamientos[2,]) {
  row2 = row2+i^2
}

Rrow1 =0
Rrow2 =0

for (k in tratamientos[1,1:5]) {
  Rrow1 = Rrow1+(k-t1Bar[1])^2
}
for (k in tratamientos[1,6:8]) {
  Rrow1 = Rrow1+(k-t2Bar[1])^2
}
for (k in tratamientos[1,9:12]) {
  Rrow1 = Rrow1+(k-t3Bar[1])^2
}
for (k in tratamientos[2,1:5]) {
  Rrow2 = Rrow2+(k-t1Bar[2])^2
}
for (k in tratamientos[2,6:8]) {
  Rrow2 = Rrow2+(k-t2Bar[2])^2
}
for (k in tratamientos[2,9:12]) {
  Rrow2 = Rrow2+(k-t3Bar[2])^2
}

SSobs = cbind(row1,row2)
SSmean = totalObservations*cbind(tbar[1]^2,tbar[2]^2)
SStrt = (5*cbind((t1Bar[1]-tbar[1])^2,(t1Bar[2]-tbar[2])^2))+(3*cbind((t2Bar[1]-tbar[1])^2,(t2Ba
r[2]-tbar[2])^2))+(4*cbind((t3Bar[1]-tbar[1])^2,(t3Bar[2]-tbar[2])^2))
SSres = cbind(Rrow1,Rrow2)
SStotal = SSobs-SSmean
row1t=0
```

```r
SSobst =(2*9)+(3*2)+(7*5)+(2*1)+(7*5)+(3*2)+(2*4)+(9*4)+(1*4)+(7*2)+(4*9)+(3*2)
SSmeant = totalObservations*tbar[1]*tbar[2]
SStrtt = (5*(t1Bar[1]-tbar[1])*(t1Bar[2]-tbar[2]))+(3*(t2Bar[1]-tbar[1])*(t2Bar[2]-tbar[2]))+(4*
(t3Bar[1]-tbar[1])*(t3Bar[2]-tbar[2]))
SSrest = 3.2
SStotalt = SSobst-SSmeant
```

```
B <- matrix(c(SStrt[1], SStrtt,
              SStrtt, SStrt[2]), ncol = 2, byrow = TRUE)
W <- matrix(c(SSres[1],SSrest,
              SSrest, SSres[2]), ncol = 2, byrow = TRUE)
BW = B+W
B
```

```
##       [,1]  [,2]
## [1,]  1.45 -1.35
## [2,] -1.35  2.30
```

```
W
```

```
##          [,1]     [,2]
## [1,] 74.21667  3.20000
## [2,]  3.20000 74.61667
```

```
BW
```

```
##          [,1]     [,2]
## [1,] 75.66667  1.85000
## [2,]  1.85000 76.91667
```

Source of variation	Sum of square matrix	Degres
Treatment	B	2
Residual	W	9
Total	B+W	11

b) Evalue el Lambda de Wilk, Λ^* y realice una prueba de hipótesis de los efectos de tratamiento.Set $\alpha = 0.05$

$H_0 = T_1 = T_2 = T_3$ $H_a = T_l = 0 \quad l = 1, 2, 3$

```
g=3
p=2
Lam = det(W)/(det(BW))
prueba = ((totalObservations-g-1)/(g-1))*((1-sqrt(Lam))/(sqrt(Lam)))
prueba
```

```
## [1] 0.1032506
```

```
f1 = qf(1-0.05,df1=4,df2=16)
f1
```

```
## [1] 3.006917
```

Podemos observas que $0.103 < 3.006$ es decir que no vamos a rechazar la prueba H_0

c) Repita la prueba considerando que la prueba es grande.Set $\alpha = 0.05$

```
prueba2 = -(totalObservations-1-((p+g)/2))*log(Lam)
prueba2
```

```
## [1] 0.4332473
```

```
chi = qchisq(1-0.05,df=4)
chi
```

```
## [1] 9.487729
```

Debido a que $0.43 < 9.48$ decimos que no vamos a rechazar la prueba H_0, en esta prueba debido a que era una muestra grande utilizamos la distribución chi cuadrado.