Ejercicio 2

Cedric Prieels

Resumen

Este ejercicio consiste en leer tres ficheros distintos de datos, para analizar y compararlos dos a dos usando diferentes pruebas (parámetricas y no parámetricas) que estudiamos en clase. Estos diferentes tests nos pueden ayudar a determinar si las tres distribuciones tienen la misma media, la misma varianza y son con compatibles con ser la misma distribución o no. El objetivo principal consiste entonces en comparar los resultados obtenidos por cada prueba para compararlos, y para verificar si los resultados devueltos para cada tipo de prueba son compatibles entre si.

Introducción

En este informe, se usan diferentes tipos de tests para comparar la media, la varianza y las distribuciones de tres ficheros de datos. Existen dos categorias principales de tests : los métodos parámetricos (nos ayudan a determinar si los datos tienen la misma varianza o la misma media, pero no son sensibles al hecho de si vienen de la misma distribución) y los métodos no-parámetricos (son sensibles no solamente a diferentes medias y varianzas, pero también a a forma de la distribución).

En particular, los diferentes tests que se usan son los siguientes :

- Distribución t de student (prueba parámetrica de la media)
- Test F de Fischer (prueba parámetrica de la varianza)
- Test U de Wilcoxon-Mann-Whitney (prueba no paramétrica de la media)
- Test Z (prueba parámetrico de la varianza para N grande)
- Test F de Fischer por rangos (prueba de la varianza para N grande)
- Test de Kolmogorov-Smirnov o KS (prueba general de la distribución)

En este ejercicio, vamos a ir comparando dos a dos cada fichero de datos y usando cada test. En cada caso, lo que queremos calcular es la hipotesis nula H0 (suponemos que las dos distribuciones son iguales), y en particular su probabilidad P(H0) que corresponde a la probabilidad de equivocarse si rechazamos esta hipótesis. Por la existencia de errores de medidas, lo único que podemos demonstrar es la diferencia, nunca la igualdad, por el uso de diferentes estadisticos propios a cada test. Un valor de P(H0) muy pequeño nos permite entonces concluir en la diferencia del parámetros estudiado, mientras que un valor de probabilidad cercano a 1 no nos permite concluir nada.

En general, R tiene funciones escritas que nos permiten calcular esta probabilidad directamente. También podemos calcular nosotros el valor del estadístico de cada test, para hacer un “cross-check” y verificar nuestros resultados, porque la probabilidad tiene que ser exactamente igual en los dos casos.

Metodología

Tenemos tres distribuciones de datos a nuestra disposición, en tres ficheros de datos separados. Como siempre, lo primero que hay que hacer es abrirlos para mirar al ojo la pinta que tiene cada fichero. En este caso, vemos que el primer y el tercer fichero tiene números positivos y negativos definidos con unas 10 decimales, mientras el segundo fichero solo tiene números enteros positivos.

La segunda etapa consiste en abrir y leer los ficheros de datos con R, para guardarlos en unos vectores. Se calcula también el número de lineas de cada fichero.

```
rm(list=ls())

setwd('/Users/ced2718/Documents/Universite/Estadistica/Ejercicio_2/')
data1 <- read.table("dat1.dat", header=FALSE)
data2 <- read.table("dat2.dat", header=FALSE)
data3 <- read.table("dat3.dat", header=FALSE)

ndata1 <- nrow(data1)
ndata2 <- nrow(data2)
ndata3 <- nrow(data3)

#Se convierten las listas a data frames
data1 <- data1[,1]
data2 <- data2[,1]
data3 <- data3[,1]

data12 <- c(data1, data2)
data13 <- c(data1, data3)
data23 <- c(data2, data3)

ndata12 <- length(data12)
ndata13 <- length(data13)
ndata23 <- length(data23)
```

Tests

Ya se puede empezar a estudiar dos a dos los ficheros de datos.

Test t de student

El test de student es una prueba parámetrica de la media, valido para distribuciones gaussianas. Este test se basa en el estadístico t, y en el número de grados de libertad nu. El estadístico se puede calcular de dos maneras diferentes, en función de si la varianza de un fichero de datos es igual a la otra, o si las varianzas no son iguales. Lo primero que se puede hacer es entonces calcular el valor de la varianza en cada caso, para crear un buleano que nos permitirá después saber en que caso estamos. Después, se puede usar directamente la función de R que nos permite calcular el valor de la probabilidad de la hipótesis nula H0.

```
#Se calculan unos buleanos para saber que caso del test utilizar
varIgual12 <- (var(data1)==var(data2))
varIgual13 <- (var(data1)==var(data3))
varIgual23 <- (var(data2)==var(data3))
```

```
resultadosVarianzas <- matrix(c(var(data1), var(data2), var(data3),
                                varIgual12, varIgual13, varIgual23), ncol=2)
colnames(resultadosVarianzas) <- c("Varianza obtenida", "Valor del buleano")
rownames(resultadosVarianzas) <- c("data1 vs data2","data1 vs data3","data2 vs data3")
rtab <- as.table(resultadosVarianzas)
head(rtab)
```

```
##                Varianza obtenida Valor del buleano
## data1 vs data2          2.001512          0.000000
## data1 vs data3          2.001512          0.000000
## data2 vs data3          2.001512          0.000000
```

En este caso, algo extraño aparece : si imprimos por pantalla el valor de la varianza de cada fichero, R nos da un resultado igual en cada caso. Pero si definimos un buleano de esta manera, R nos dice que las varianzas no son en realidad las mismas! Es probablemente que son casi iguales, pero que hay decimales después de la sexta que cambian (en realidad, las decimales cambian a partir de la decimal número 13). Entonces, tendremos que usar el caso de varianzas diferentes en los pasos siguientes.

```
ttest12 <- t.test(data1, data2, var.equal = varIgual12)
ttest13 <- t.test(data1, data3, var.equal = varIgual13)
ttest23 <- t.test(data2, data3, var.equal = varIgual23)

#Sacamos la probabilidad de la hipotesis nula
ttest12PVal <- ttest12$p.value
ttest13PVal <- ttest13$p.value
ttest23PVal <- ttest23$p.value
```

Ya hemos calculado la probabilidad de la hipotesis H0 en los tres casos. Ahora, queremos volver a calcular esta misma probabilidad pero sin usar las funciones de R, como un "cross-check". Esto se hace usando la función pt de R, que necesita como argumentos el valor del estadístico y el número de grados de libertad. Estos parámetros se puedes calcular de dos maneras distintas : a mano, implementando las formulas vistas en clase, o bien usando directamente el valor de ttest12$statisticydettest$12parameter. La implementación de las formulas vistas en clase se hace por el uso de dos funciones, que devuelven directamente los parámetros buscados.

```
getEstadisticoT<- function(x, y, varIgual) {
  nx = length(x)
  ny = length(y)
  if(varIgual){
    estadisticoT <- (mean(x) - mean(y)) / (sqrt((sd(x)^2/(nx))+(sd(y)^2/(ny))))
  } else {
    denominador <- (sum((x-mean(x))^2) + sum((y-mean(y))^2))/(nx + ny - 2)
    estadisticoT <- (mean(x) - mean(y)) / sqrt(denominador * ((1/nx)+(1/ny)))
  }
  return(estadisticoT)
}
```

```
getNuT <- function(x, y, varIgual) {
  nx=length(x)
  ny=length(y)
  if(varIgual){
    nu <- nx+ny-2
```

```
  } else {
    ratiox <- sd(x)^2/nx
    ratioy <- sd(y)^2/ny
    nu <- (ratiox+ratioy)^2/((ratiox^2/(nx-1))+ratioy^2/(ny-1))
  }
  return(nu)
}
```

```
#Cálculo a mano del estadistico y del número de grados de libertad
estadisticoT12 <- getEstadisticoT(data1, data2, varIgual12)
estadisticoT13 <- getEstadisticoT(data1, data3, varIgual13)
estadisticoT23 <- getEstadisticoT(data2, data3, varIgual23)
nu12 <- getNuT(data1, data2, varIgual12)
nu13 <- getNuT(data1, data3, varIgual13)
nu23 <- getNuT(data2, data3, varIgual23)
```

Ahora podemos comparar el valor obtenido a mano para los estadísticos con los valores obtenidos por el uso de la función t.test de R (y también los valores del número de grados de libertad en cada caso).

```
#Valor del estadistico
resultadosEstadisticoT <- matrix(c(estadisticoT12, estadisticoT13, estadisticoT23,
                                   ttest12$statistic, ttest13$statistic, ttest23$statistic), ncol=2)
colnames(resultadosEstadisticoT) <- c("Estadistico obtenido a mano", "Obtenido por R")
rownames(resultadosEstadisticoT) <- c("data1 vs data2","data1 vs data3","data2 vs data3")
rtab <- as.table(resultadosEstadisticoT)
head(rtab)
```

```
##                Estadistico obtenido a mano Obtenido por R
## data1 vs data2                1.331764e-13   1.331764e-13
## data1 vs data3                1.997646e-13   1.997646e-13
## data2 vs data3                6.658821e-14   6.658821e-14
```

```
#Numero de grados de libertad
resultadosNuT <- matrix(c(nu12, nu13, nu23, ttest12$parameter, ttest13$parameter, ttest23$parameter), n
colnames(resultadosNuT) <- c("Grados de libertad a mano", "Grados de libertad por R")
rownames(resultadosNuT) <- c("data1 vs data2","data1 vs data3","data2 vs data3")
rtab <- as.table(resultadosNuT)
head(rtab)
```

```
##                Grados de libertad a mano Grados de libertad por R
## data1 vs data2                     19998                    19998
## data1 vs data3                     19998                    19998
## data2 vs data3                     19998                    19998
```

Vemos entonces que el uso de dos métodos diferentes nos devuelve valores muy cercas, lo que tiene buena pinta. Ahora, volvemos a calcular el valor de la probabilidad de la hipotesis nula. Se multiplica por 2 la función pt porque como queremos negar que dos cosas son iguales, hay que calcular dos probabilidades : que sea menor, y que sea mayor. Esto se puede implementar en este caso multiplicando el valor de probabilidad obtenido por dos.

```
ttest12PVal2 <- as.numeric(2*pt(ttest12$statistic, ttest12$parameter))
ttest13PVal2 <- as.numeric(2*pt(ttest13$statistic, ttest13$parameter))
ttest23PVal2 <- as.numeric(2*pt(ttest23$statistic, ttest23$parameter))
```

Ya podemos comparar los valores de probabilidad obtenidos en los dos casos.

```
resultadosT <- matrix(c(ttest12PVal, ttest13PVal, ttest23PVal,
                        ttest12PVal2, ttest13PVal2, ttest23PVal2), ncol=2)

colnames(resultadosT) <- c("Test de R", "Test manual")
rownames(resultadosT) <- c("data1 vs data2","data1 vs data3","data2 vs data3")

rtab <- as.table(resultadosT)
head(rtab)
```

```
##                Test de R Test manual
## data1 vs data2         1           1
## data1 vs data3         1           1
## data2 vs data3         1           1
```

Como podemos ver, en los tres casos de comparación y con los dos métodos de cálculo, obtenemos siempre un valor de probablidad de 1, lo que es rasurante sobre el método implementado. Como tenemos un valor de probabilidad que no es igual a 0 (o muy cerca), no podemos concluir que las medias son diferentes entre todos los ficheros.

Test F de Fischer

El test F es una prueba parámetrica de la varianza. Este test se basa en el estadístico F, y en dos parámetros diferentes, nu1 y nu2, que valen el número de datos en cada fichero menos uno (una ligadura). Como antes, primero se usa directamente la función de R correspondiente para calcular el valor de la probabilidad de la hipotesis nula.

```
Ftest12 <- var.test(data1, data2)
Ftest13 <- var.test(data1, data3)
Ftest23 <- var.test(data2, data3)

#Sacamos la probabilidad de la hipotesis nula
Ftest12PVal <- Ftest12$p.value
Ftest13PVal <- Ftest13$p.value
Ftest23PVal <- Ftest23$p.value
```

También se puede usar el método pf de R, que nos permite volver a calcular estas tres probabilidades para verificar el resultado obtenido. Para poder usar este método, lo primero que hay que hacer es calcular a mano el estadístico F y los grados de libertad. Como la desviación estandár más grande tiene que ir en el numerador, se ponen una serie de condiciones, aunque en este caso como las varianzas son casi iguales (hasta la sexta decimale, como lo vemos), no es estrictamente necesario.

```
if(sd(data1) > sd(data2)) {
  estadisticoF12 <- (sd(data1)^2)/(sd(data2)^2)
} else {
  estadisticoF12 <- (sd(data2)^2)/(sd(data1)^2)
```

```
}

if(sd(data1) > sd(data3)) {
  estadisticoF13 <- (sd(data1)^2)/(sd(data3)^2)
} else {
  estadisticoF13 <- (sd(data3)^2)/(sd(data1)^2)
}

if(sd(data2) > sd(data3)) {
  estadisticoF23 <- (sd(data2)^2)/(sd(data3)^2)
} else {
  estadisticoF23 <- (sd(data3)^2)/(sd(data2)^2)
}

#Cálculo de los grados de libertad
nu1 = ndata1 - 1
nu2 = ndata2 - 1
nu3 = ndata3 - 1
```

Se pueden comparar los resultados obtenidos hasta ahora por los resultados obtenidos por el uso de la función correcta de R.

```
#Comparación desl estadístico obtenido
resultadosEstadisticoF <- matrix(c(estadisticoF12, estadisticoF13, estadisticoF23,
                                   Ftest12$statistic, Ftest13$statistic, Ftest23$statistic), ncol=2)

colnames(resultadosEstadisticoF) <- c("Estadistico obtenido a mano", "Obtenido por R")
rownames(resultadosEstadisticoF) <- c("data1 vs data2","data1 vs data3","data2 vs data3")

rtab <- as.table(resultadosEstadisticoF)
head(rtab)
```

```
##                Estadistico obtenido a mano Obtenido por R
## data1 vs data2                           1              1
## data1 vs data3                           1              1
## data2 vs data3                           1              1
```

```
resultadosNu1F <- matrix(c(nu1, nu2, nu3,
                           Ftest12$parameter[1], Ftest13$parameter[1], Ftest23$parameter[1]), ncol=2)

colnames(resultadosNu1F) <- c("Grado de libertad 1 a mano", "Grado de libertad 1 por R")
rownames(resultadosNu1F) <- c("data1 vs data2","data1 vs data3","data2 vs data3")

rtab <- as.table(resultadosNu1F)
head(rtab)
```

```
##                Grado de libertad 1 a mano Grado de libertad 1 por R
## data1 vs data2                       9999                       9999
## data1 vs data3                       9999                       9999
## data2 vs data3                       9999                       9999
```

```
#Cálculo de la probabilidad
Ftest12PVal2 <- as.numeric(2*pf(Ftest12$statistic, (ndata1-1), (ndata2-1)))
Ftest13PVal2 <- as.numeric(2*pf(Ftest13$statistic, (ndata1-1), (ndata3-1)))
Ftest23PVal2 <- as.numeric(2*pf(Ftest23$statistic, (ndata2-1), (ndata3-1)))
```

Ahora podemos estudiar y comparar en una table los resultados de probabilidad obtenidos por cada uno de los dos métodos.

```
resultadosF <- matrix(c(Ftest12PVal, Ftest12PVal2, Ftest13PVal,
                        Ftest13PVal2, Ftest23PVal, Ftest23PVal2), ncol=2)

colnames(resultadosF) <- c("Test de R", "Test manual")
rownames(resultadosF) <- c("data1 vs data2","data1 vs data3","data2 vs data3")

rtab <- as.table(resultadosF)
head(rtab)
```

```
##                Test de R Test manual
## data1 vs data2         1           1
## data1 vs data3         1           1
## data2 vs data3         1           1
```

Como volvemos a obtener valores de probabilidades diferentes de 0, no podemos concluir que que la varianza sea distinta en todos estos casos.

Tests de rangos

Este test es una prueba no parámetrica de la media, que tiene la vantaja de no suponer que las distribuciones son gaussianas. Además, se usa toda la distribución en este caso, no solamente algunos parámetros como antes (el método es entonces sensible a distintas formas de la distribución).

Este método consiste en sustituir los valores de los ficheros iniciales por los rangos de los elements (en caso de empate, se asigna el rango promedio). La información no se pierde de esta manera, solo está escrite de otra manera. Además, este sistema tiene por ventaja que puntos muy lejados de la distribución quedan cercas al final y ya no influyen tanto, lo que puede ser interesante en algunos casos.

```
#Cálculo de los rangos
data1rangos12 <- rank(data12,ties.method="average") [1:ndata1]
data2rangos12 <- rank(data12,ties.method="average") [(ndata1+1):ndata12]

data1rangos13 <- rank(data13,ties.method="average") [1:ndata1]
data3rangos13 <- rank(data13,ties.method="average") [(ndata1+1):ndata13]

data2rangos23 <- rank(data23,ties.method="average") [1:ndata2]
data3rangos23 <- rank(data23,ties.method="average") [(ndata2+1):ndata23]
```

Se puede calcular ahora la suma de los rangos. Como los ficheros de entrada tienen muchos puntos, consideramos que estas sumas se comportan como gausianas y volvemos entonces a usar el test t, aplicado a los rangos que acabamos de calcular, o bien se puede usar el test U de Wilcoxon-Mann-Whitney directamente sobre los datos.

Test t sobre rangos

Como se usa en este caso el test de Student, hay primero que calcular y comparar las varianzas de cada distribución obtenida como lo hicimos por el test t sobre datos.

```
#Verificación del caso a usar
varIgualRangos12 <- (var(data1rangos12)==var(data2rangos12))
varIgualRangos13 <- (var(data1rangos13)==var(data3rangos13))
varIgualRangos23 <- (var(data2rangos23)==var(data3rangos23))
```

Por fin, se puede aplicar el test a los rangos de los datos que acabamos de calcular.

```
ttestRangos12 <- t.test(data1rangos12, data2rangos12, var.equal=varIgualRangos12)
ttestRangos13 <- t.test(data1rangos13, data3rangos13, var.equal=varIgualRangos13)
ttestRangos23 <- t.test(data2rangos23, data3rangos23, var.equal=varIgualRangos23)

ttestRangos12PVal <- ttestRangos12$p.value
ttestRangos13PVal <- ttestRangos13$p.value
ttestRangos23PVal <- ttestRangos23$p.value
```

Test U de Wilcoxon-Mann-Whitney sobre datos

Esta prueba se puede aplicar directamente a los datos, y no a los rangos.

```
Utest12 <- wilcox.test(data1, data2, paired=FALSE)
Utest13 <- wilcox.test(data1, data3, paired=FALSE)
Utest23 <- wilcox.test(data2, data3, paired=FALSE)

Utest12PVal <- Utest12$p.value
Utest13PVal <- Utest13$p.value
Utest23PVal <- Utest23$p.value
```

Por fin, se pueden resumir los dos últimos etapas en una table con las probabilidades que acabamos de calcular para compararlas.

```
resultadosU <- matrix(c(ttestRangos12PVal, ttestRangos13PVal, ttestRangos23PVal,
                        Utest12PVal, Utest13PVal, Utest23PVal), ncol=2)

colnames(resultadosU) <- c("t test sobre rangos", "u test sobre datos")
rownames(resultadosU) <- c("data1 vs data2","data1 vs data3","data2 vs data3")

rtab <- as.table(resultadosU)
head(rtab)
```

```
##                t test sobre rangos u test sobre datos
## data1 vs data2        9.023614e-03       9.026857e-03
## data1 vs data3        8.614740e-01       8.614699e-01
## data2 vs data3        2.948848e-05       2.958651e-05
```

Vemos que obtenemos resultados muy similares, utilizando métodos completamente diferentes. Esto da confianza en los valores de probabilidades obtenidos. Además, vemos que en este caso estos valores son

pequeños, que ya no valen 1 como en los últimos casos. Entonces, con estos últimos resultados (que tienen que ser más fiables que los precedientes, porque se basen en un método no parámetrico que compara la forma de toda la distribución, y no solamente unos parámetros como la media o la varianza), podemos concluir que las distribuciones de los ficheros 2 y 3 son muy diferentes. Es muy poco probable que los puntos de estos dos ficheros vengan de una misma distribución. Se ve después que las distribuciones de los ficheros 1 y 2 son probablemente diferentes también, aunque en este caso el valor de la probabilidad de la hipotésis nula esté un poco más alto. Al final, vemos que no podemos concluyir mucho de la comparación de los ficheros 1 y 3.

Test Z

Esta prueba es una prueba parámetrica de la varianza para un número de datos N grande (lo que se justifica perfectamente en nuestro caso). Lo primero que hay que hacer, es calcular el valor del estadístico Z, implementando una función (porque lo vamos a tener que calcular por lo menos 3 veces).

```
getZ <- function (a, b) {
  #Se calculan los parámetros importantes de los datos de entrada a y b
  numa <- length(a)
  numb <- length(b)
  meana <- mean(a)
  meanb <- mean(b)
  sda <- sd(a)
  sdb <- sd(b)
  #Se calcula el estadístico Z por las relaciones vistas en clase
  sumaa <- sum((a-meana)**2)
  sumab <- sum((b-meanb)**2)
  sd <- sqrt((sumaa+sumab)*((1/numa)+(1/numb))/(numa+numb-2))
  Z <- (sqrt(2)*(sda-sdb))/sd
  return(Z)
}
```

Ahora, como lo vimos en clase, sabemos que podemos calcular la probabilidad que buscamos calculando la función de error erf del estadístico Z. Para esto, hay que instalar el paquete VGAM que tien la función de error implementada.

```
#Se calcula Z para las distintas distribuciones de datos usando la función precediente
Z12 <- getZ(data1, data2)
Z13 <- getZ(data1, data3)
Z23 <- getZ(data2, data3)
#Se instala la libreria VGAM y se calculan las probabilidades
library(VGAM)
```

```
## Loading required package: stats4
```

```
## Loading required package: splines
```

```
Ztest12PVal <- erfc(Z12)
Ztest13PVal <- erfc(Z13)
Ztest23PVal <- erfc(Z23)

Ztest12PVal
```

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

```
Ztest13PVal
```

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

```
Ztest23PVal
```

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

Volvemos a obtener una probabilidad de 1 en los tres casos.

Test F con rangos

Este test es una prueba de la varianza para un números de datos N grande. Se usa el método de R var.test sobre los rangos calculados antes para calcular la probabilidad de la hipótesis nula.

```
FtestRangos12 <- var.test(data1rangos12, data2rangos12)
FtestRangos13 <- var.test(data1rangos13, data3rangos13)
FtestRangos23 <- var.test(data2rangos23, data3rangos23)

FtestRangos12PVal <- FtestRangos12$p.value
FtestRangos13PVal <- FtestRangos13$p.value
FtestRangos23PVal <- FtestRangos23$p.value

FtestRangos12PVal
```

```
## [1] 0
```

```
FtestRangos13PVal
```

```
## [1] 1.820766e-14
```

```
FtestRangos23PVal
```

```
## [1] 0.0169599
```

Con este método, no hay manera de verificar nuestros números haciendo un "cross-check".

Test de de Kolmogorov-Smirnov

Este test es una prueba general de las distribuciones. Se usa solamente mediante las funciones implementadas en R.

```
KStest12 <- ks.test(data1, data2)
```

```
## Warning in ks.test(data1, data2): p-value will be approximate in the
## presence of ties
```

```
KStest13 <- ks.test(data1, data3)
```

```
## Warning in ks.test(data1, data3): p-value will be approximate in the
## presence of ties
```

```
KStest23 <- ks.test(data2, data3)
```

```
## Warning in ks.test(data2, data3): p-value will be approximate in the
## presence of ties
```

```
KS12.pval <- KStest12$p.value
KS13.pval <- KStest13$p.value
KS23.pval <- KStest23$p.value
KS12.pval
```

```
## [1] 0
```

```
KS13.pval
```

```
## [1] 1.110223e-16
```

```
KS13.pval
```

```
## [1] 1.110223e-16
```

Conclusión

Para concluir este estudio, volvemos a hacer una table con todos los resultados de probabilidad de la hipotésis nula obtenidos en este informe.

Comparacion	t.test	pt	var.test	pf	Test Z
Data 1 vs data 2	1	1	1	1	1
Data 1 vs data 3	1	1	1	1	1
Data 2 vs data 3	1	1	1	1	1

Comparación	t.test (rangos)	wilcox.test	var.test (rangos)	ks.test
Data 1 vs data 2	0.01	0.01	0	0
Data 1 vs data 3	0.86	0.86	0	0
Data 2 vs data 3	0	0	0.02	0

Como ya explicado a lo largo del informe, se han verificado estos resultados cada vez que era posible, haciendo a veces un segundo test completamente diferente que tiene que dar lo mismo, y calculando a mano el valor del estadìstico cada vez que se podía.

A la vista de estos resultados, podemos sacar diferentes conclusiones. Primero, se puede ver que las pruebas parámetricas (test t, test F y test Z) siempre devuelven un resultado de probabilidad de la hipotésis nula igual a 1, lo que significa que tenemos distribuciones que tienen la misma media y la misma varianza. Es importante no sacar más conclusiones de estos resultados (en particular, no podemos concluir que las distribuciones

son iguales, porque sabemos después del ejercicio 1 que no lo son), especialmente porque sabemos que estas pruebas reponen en la hipotésis de que las distribuciones sean gaussianas, y sabemos después del ejercicio 1 que esta hipotésis no se justifica en este caso.

Del otro lado lado, el test U devuelve un valor de P(H0) muy cerca a 0 (data1 comparado con data2, y data2 comparado con data3) y un valor de 0.86 al comparar las distribuciones data1 con data3. Finalmente, las últimas pruebas no parámetricas que son el Test F por rangos y el Test KS devuelven una probabilidad muy pequeña en todos los casos, lo que parece indicar que todas las distribuciones son distintas.

Cabe indicar que estas pruebas nos dan más informaciones sobre nuestras distribuciones puesto que comparan directamente toda la forma de las distribuciones, mientras que las pruebas parámetricas solo comparan un parámetro (media o varianza en este caso). Como lo vimos en el ejercicio 1, comparar solamente estos parámetros es en general no suficiente : sabemos que las tres distribuciones tienen la misma media y la misma varianza, pero sabemos también que vienen de distribuciones muy diferentes (uniforme, de Poisson y de Gauss). Esto explica los resultados obtenidos a lo largo del informe : las pruebas parámetricas devuelven siempre un resultado de probabilidad igual a 1 mientras que las pruebas no parámetricas devuelven un valor de probabilidad de hipotésis nula igual a 0, puesto que las distribuciones son diferentes.

Bibliografía

R Markdown, *Markdown basics*, http://rmarkdown.rstudio.com/authoring_basics.html. Consultado por última vez el 29 de octubre 2016.

R Development Core Team (2008). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.