

EBB2022_Exercici_Solució

Santi

2023-05-30

1.- Ejercicio Indica que tipo de análisis o que pruebas estadísticas utilizarías y si fuera necesario algún tipo de prueba adicional para llevar a cabo el análisis. Formula la hipótesis a contrastar de acuerdo con las hipótesis seleccionadas

a.- Se efectúa un estudio de seguimiento a 1018 sujetos atendidos en una clínica de obesidad. Se mide el Índice de Masa Corporal(IMC) y el perfil lipídico. Al cabo de 12 meses se evalúa de nuevo el IMC y el colesterol estando interesados en cuantificar la disminución de ambos parámetros

Disponemos de dos variables cuantitativas medidas en dos momentos distintos. Se quiere analizar si los valores pre y post sufren algún cambio, para ello debemos: en primer lugar evaluar la normalidad de las variables. En caso de que se ajusten a una distribución normal, miramos la homogeneidad de las variables y realizamos un test t de Student para muestras apareadas. En el caso de que no sigan una distribución normal, realizaremos el test no paramétrico de Wilcoxon para muestras apareadas.

b.- Se analizan un grupo de variables inmunológica(leucocitos totales, linfocitos B, natural Killer, etc) en una muestra de 102 hombres y 147 mujeres mayores de 65 años. Se está interesado en ver la existencia de diferencias por sexo.

Disponemos de distintas variables cuantitativas (leucocitos, linfocitos, etc.) y una variable cualitativa, sexo. Queremos ver si existen diferencias de las variables cuantitativas según el sexo, para ello debemos: en primer lugar evaluar la normalidad de las variables. En caso de que se ajusten a una distribución normal, miramos la homogeneidad de las variables y realizamos un test t de Student. En el caso de que no sigan una distribución normal, realizaremos el test no paramétrico U de Mann-Whitney.

c.- La supervivencia de los pacientes con cardiopatía isquémica se asocia al valor de la fracción de eyección(FE). Se desea comparar los resultados obtenidos en 125 pacientes mediante la fracción de eyección isotópica en la asignación de sujetos a grupos de alto y bajo riesgo con la asignación obtenida a partir de la FE angiográfica. Plantea el análisis.

Disponemos de dos variables cualitativas, queremos ver si existe relación entre el riesgo y la clasificación según FE angiográfica, para ello debemos: en primer lugar realizamos una tabla de frecuencias, en el caso de obtener en alguna celda un valor inferior a 0.05 realizamos un test exacto de Fisher, en caso contrario un test Chi Cuadrado.

Ejercicio Práctico

Este ejercicio consta de diversas partes en un intento de simular lo que se lleva a cabo en un estudio real. Se ha simplificado para hacerlo más practicable por lo que no hace falta que os agobiéis si algo no os cuadre del todo. De lo que se trata es que veamos como aplicar las distintas técnicas que hemos estudiado, de forma integral, en un problema de análisis de datos.

2.1 Los datos

La demora entre el comienzo de los síntomas y el ingreso hospitalario es un factor que determina la mortalidad del infarto agudo de miocardio (IAM). Se estudian 426 sujetos que acuden al servicio de urgencias de 5 hospitales por dolor torácico , recogiendo el tiempo entre los primeros síntomas y la llegada al hospital y una

serie de variables sociodemográficas. Se está interesado en estimar el retraso prehospitalario y determinar las variables asociadas.

DEMORA	Minutos desde el inicio de los síntomas hasta llegada al hospital
HOSPITAL	
EDAD	
SEXO	
NACIONAL	0 Español 1 Extranjero
MEDIO1	Lugar inicio síntomas. 0 Domicilio 1 Fuera del domicilio
DISTANC	Distancia en isocronas(0,1,2,3,...)
REMITIDO	Remisión por un sanitario 0 Si 1 No
NSE	Nivel socioeconómico 0 Bajo 1 Medio 2 Alto
COHABIT	Convivientes 0 Vive en compañía 1 Vive sólo
CARDIOP	Diagnóstico previo de cardiopatía 0 No 1 Si
DOLOR	Nivel de dolor 0 intenso 1 moderado 2 ligero
NOCHE	Aparición nocturna 0 Dia 1 Noche
AMBULAN	Acudir al hospital en Ambulancia 0 ambulancia 1 Otros medios

Figure 1: demora.png

Los datos los podéis encontrar en los ficheros de Stata demora.dta, de Excel demora.xls y de texto plano separado por comas demora.csv

```
dat <- read.csv("datasets/demora.csv")
dat$noche <- factor(dat$noche, 0:1, c("Dia", "Noche"))
dat$noche <- relevel(dat$noche, ref = "Noche")
```

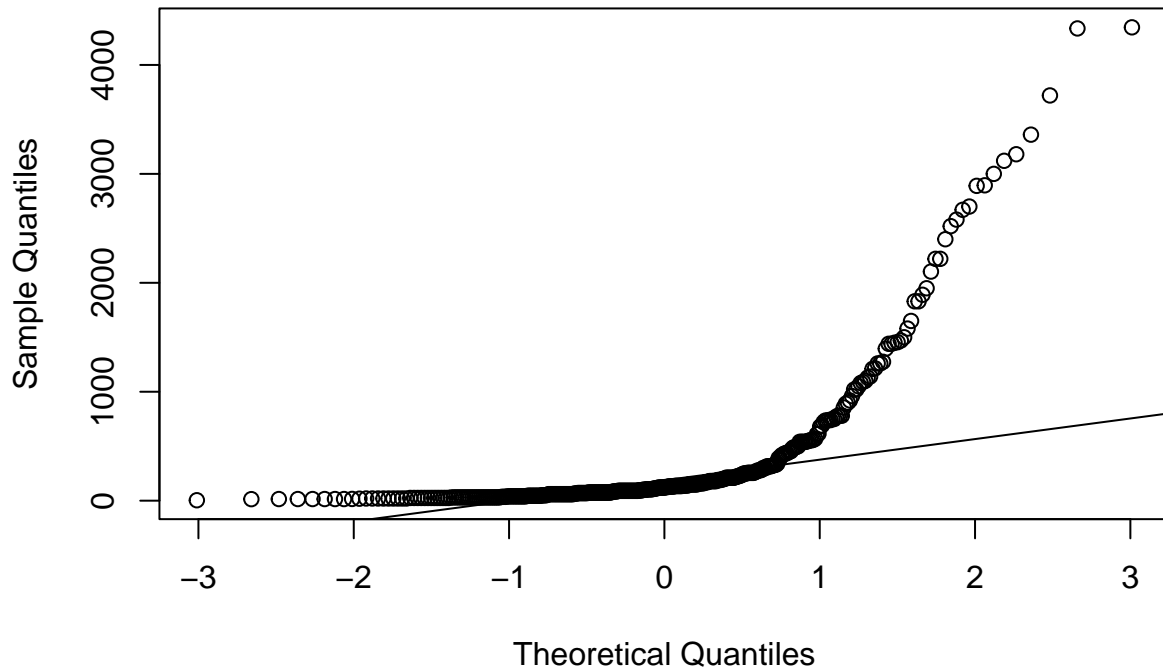
2.2. Los análisis

a) Se está interesado en conocer la relación entre la demora y la aparición nocturna del síntoma. Comprueba la normalidad de la variable

Realizamos gráficos y test de normalidad Shapiro Wilks.

```
qqnorm(dat$demora)
qqline(dat$demora)
```

Normal Q-Q Plot



```
shapiro.test(dat$demora)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data:  dat$demora  
## W = 0.55371, p-value < 2.2e-16
```

Con un p-valor inferior a 0.05 concluimos que los datos no se ajustan a una distribución normal. Indica el procedimiento de análisis. En primer lugar, debemos realizar un análisis gráfico.

Dado que los datos no se ajustan a una distribución normal, las pruebas que usemos deberán ser no paramétricas.

b) Aunque no sea el método más adecuado realiza el contraste paramétrico para contrastar la existencia de relación. Interpreta los resultados

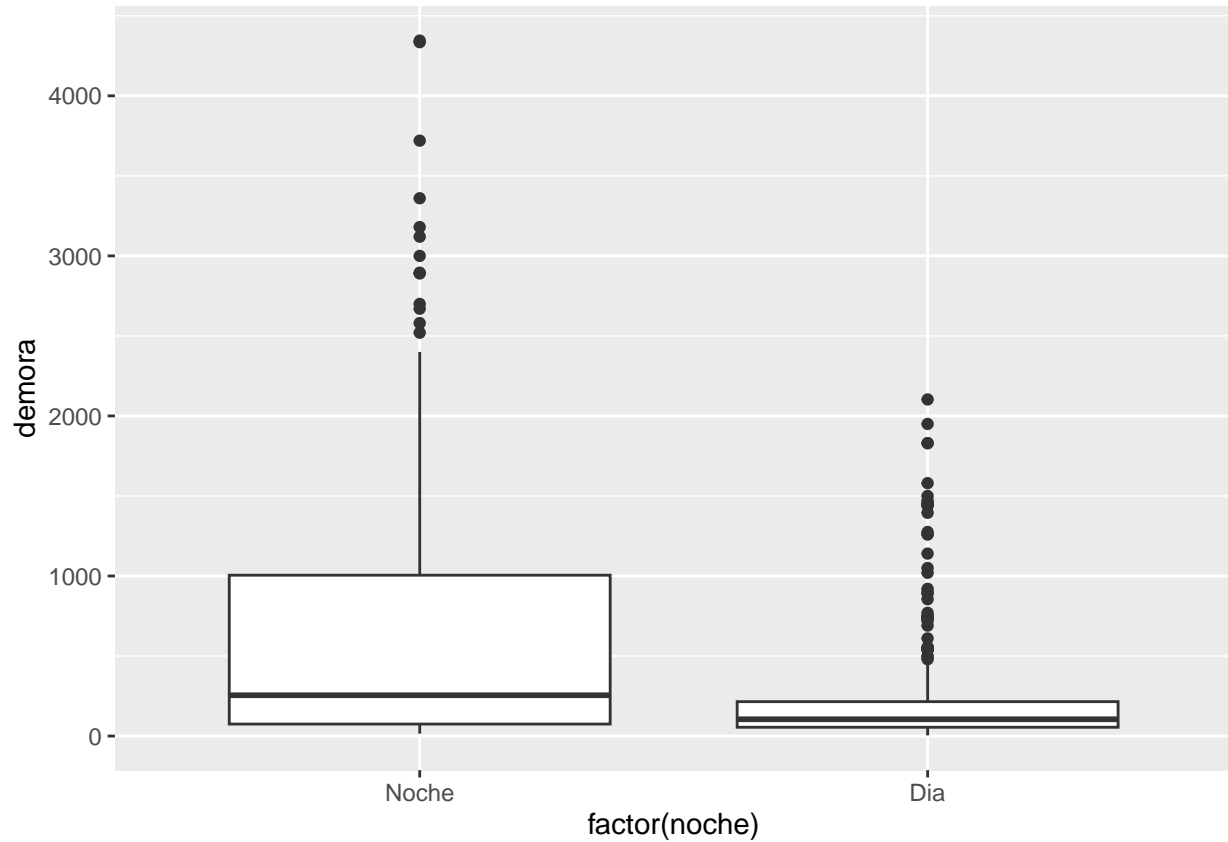
Primero efectuamos un boxplot para comparar los datos antes de efectuar el contraste de hipótesis y elegir el más adecuado

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.2.3
```

```
ggplot(dat,aes(x=factor(noche), y=demora))+  
geom_boxplot()
```

```
## Warning: Removed 43 rows containing non-finite values ('stat_boxplot()').
```



```
t.test (dat$demora ~ dat$noche)
```

```
##  
## Welch Two Sample t-test  
##  
## data: dat$demora by dat$noche  
## t = 4.7841, df = 104.89, p-value = 5.641e-06  
## alternative hypothesis: true difference in means between group Noche and group Dia is not equal to 0  
## 95 percent confidence interval:  
## 313.873 758.213  
## sample estimates:  
## mean in group Noche mean in group Dia  
## 780.3061 244.2632
```

```
wilcox.test(dat$demora ~ dat$noche)
```

```
##  
## Wilcoxon rank sum test with continuity correction
```

```
##
## data: dat$demora by dat$noche
## W = 18301, p-value = 4.49e-06
## alternative hypothesis: true location shift is not equal to 0
```

De forma teórica (y basandonos en no normalidad) escogeríamos el test U de Mann-Withney. Usemos el test paramétrico o no paramétrico llegamos a la misma conclusión.

c) Se cree que a un 20% de los pacientes les aparecen los síntomas por la noche. Comprueba dicha hipótesis. Que conclusión se obtiene

Realizamos un test de proporciones para una muestra. Con un p-valor inferior a 0.05 rechazamos que al 20% de los pacientes les aparezcan los síntomas por la noche.

```
prop.test(table(dat$noche),p = .2)
```

```
##
## 1-sample proportions test with continuity correction
##
## data: table(dat$noche), null probability 0.2
## X-squared = 44.866, df = 1, p-value = 2.11e-11
## alternative hypothesis: true p is not equal to 0.2
## 95 percent confidence interval:
## 0.2868619 0.3782098
## sample estimates:
## p
## 0.3309859
```

d) Se esta interesado en estudiar la relación entre el nivel de dolor y la aparición nocturna de los síntomas. Indica el procedimiento de análisis.

Realizamos una tabla de frecuencias para ver como se distribuyen los datos y si en alguna de las celdas tenemos un valor inferior a 5, si es así realizamos un test exacto de Fisher, en caso contrario un test chi cuadrado para evaluar independencia entre variables. Ejecuta dicho análisis e interpreta los resultados.

```
gmodels::CrossTable(dat$dolor, dat$noche, prop.c = F,prop.r = F, chisq = F,prop.chisq = F)
```

```
##
##
## Cell Contents
## |-----|
## | N |
## | N / Table Total |
## |-----|
##
##
## Total Observations in Table: 426
##
##
## | dat$noche
## dat$dolor | Noche | Dia | Row Total |
## -----|-----|-----|-----|
## 0 | 68 | 137 | 205 |
## | 0.160 | 0.322 | |
```

```
## -----|-----|-----|-----|
##          1 |          59 |          124 |          183 |
##          |          0.138 |          0.291 |          |
## -----|-----|-----|-----|
##          2 |          14 |          24 |          38 |
##          |          0.033 |          0.056 |          |
## -----|-----|-----|-----|
## Column Total |          141 |          285 |          426 |
## -----|-----|-----|-----|
##
##
```

Como ningún valor de la tabla es inferior a 5 realizamos un test chi-cuadrado.

```
chisq.test(table(dat$dolor, dat$noche))
```

```
##
## Pearson's Chi-squared test
##
## data:  table(dat$dolor, dat$noche)
## X-squared = 0.30183, df = 2, p-value = 0.8599
```

Con los resultados del test, p.valor superior a 0.05, no podemos rechazar la hipótesis nula, consideramos que noche y dolor son independientes.