

PEC2 Inferencia estadística 2018-19 semestre 2

Máster Bioinformática y Bioestadística

Junio 2019

Ejercicio 1

A continuación, se presentan diversas situaciones de diferentes estudios basados en datos reales. Se trata de indicar el tipo de análisis o de pruebas estadísticas a utilizar en cada uno de los casos e indicar, si es necesario, si hay que efectuar algún tipo de prueba adicional para llevar a cabo el análisis. Formula las hipótesis oportunas para cada tipo de prueba.

a) La fibrilación auricular es responsable del 20% de todos los ictus. que son más graves, discapacitantes y mortales. Los anticoagulantes orales, tipo antivitamina K, parece que son capaces de prevenir la incidencia de ictus a costa de incrementar el riesgo de una hemorragia intracraneal. Los anticoagulantes de acción directa parece que tienen menos de riesgo de producir una hemorragia intracraneal que los antivitamina K. Se lleva a cabo un ensayo clínico en pacientes con fibrilación auricular en el que se asignan al azar anticoagulantes orales o de acción directa . Se desea contrastar si la incidencia de hemorragia intracraneal es menor en un grupo que en otro

b) En la literatura se indica que la actividad física se asocia con el insomnio En un estudio piloto se dispone de 15 pacientes con insomnio y 20 sin insomnio a los que se les mide el promedio de pasos diario durante una semana. Se está interesado en contrastar la existencia de esta asociación .

c) La procalcitonina (PCT) se asocia con la aparición de la neumonía asociada a ventilación mecánica (NAV) en pacientes de la UCI. Se dispone de una muestra de 500 pacientes divididos en tres grupos: sin NAV, con NAV y con NAV con un shock. Se pretende contrastar si los niveles de procalcitonina son diferente en los tres grupos

d) Se dispone de una muestra de 80 diabéticos recién diagnosticados a los que se les recomienda bajar peso para mejor control de su enfermedad. Se les mide el peso en el momento al diagnóstico y a los 12 meses. Se desea contrastar si efectivamente los pacientes han bajado de peso.

e) Se dispone de un biomarcador construido a partir de una selección de genes asociados con una patología que pretende mejorar la prueba diagnóstica ya existente. Para ello se eligen 30 pacientes con la patología y 30 sin ella según la prueba diagnóstica a los que se les aplica el biomarcador. Se desea conocer si el biomarcador clasifica de la misma manera a los enfermos clasificados por la prueba diagnóstica tradicional.

Ejercicio 2

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 y la base de datos os la proporcionamos en varios formatos: STATA(dta), EXCEL(xls) y Texto plano separado por comas (csv) . De lo que se trata es de ver cómo aplicar diversas técnicas de las aprendidas para responder a algunas de las preguntas.

Los datos

La litiasis biliar es una enfermedad definida por la presencia de calculos. Se dispone de 200 pacientes a los que se les mide una serie de factores asociados a la presencia de cálculos

Los datos los podéis encontrar en los ficheros de Stata litia.dta, de SPSS litia.sav y de texto plano separado por comas litia.csv

Las variables de la base de datos son:

caso Número de caso sexo Sexo (Hombre, Mujer) edad Edad alcohol Bebe alcohol (No,Si) talla Talla peso Peso glucosa Diabetes (No,Si) colestot Colesterol triglice Trigliceridos quetelet BMI. (Índice masa corporal) calculos Calculos(No, Si) edadg Edad (agrupada)

Muestreando la base de datos

Empezaremos tomando una muestra 180 datos de la base de datos original. Esto garantiza que cada uno de vosotros trabaje con un conjunto de datos diferente. Para extraer la muestra debéis adaptar la siguiente sintaxis:

- a) Selecciona una muestra de 180 casos de la base de datos anterior. Recuerda fijar la semilla para que si repites la selección sea la misma y guarda los datos en el objeto litia180.

`set.seed (# número) # Hay que poner un número para que cada vez que ejecutéis seleccionéis la misma muestra`
`litia350 <- litia [sample (1: nrow (litia), 180, replace = FALSE),]`

Las preguntas

a) En numerosas pruebas estadísticas la variable debe de seguir una distribución normal. Comprueba la normalidad de las variables colesterol, ,trigliceridos y bmi(quetelet). Justifica las respuestas.

b) Indica la estrategia de análisis y comprueba si el colesterol, los trigliceridos y el BMI (por separado) se asocian con la litiasis(calculos). Justifica las respuestas.

c) Alguna de estas variables pueden estar asociadas con la edad. Para ella comprueba si hay diferencias entre los trigliceridos y el índice de masa corporal (por separado) con los grupos de edad. Justifica las respuestas

d) Indica si hay diferencias en los calculos en funcion del sexo y la diabetes

Ejercicio 3

- a) Siguiendo las instrucciones debajo vamos a generar 100 veces tres muestra de tamaño 50 bajo la distribución normal con la misma media y varianza. Para cada conjunto de datos comparamos la igualdad de medias entre la muestra 1 y 2, $H_0 : \mu_1 = \mu_2$, La muestra 1 y 3 $H_0 : \mu_1 = \mu_3$ y la muestra 2 y 3 $H_0 : \mu_2 = \mu_3$ con un error tipo I de ($\alpha = 0.05$ cada uno) y guarda si se acepta o no la hipótesis nula.

por ejemplo:

```
set.seed (# número)
m<-0 # sustituye 0 por el valor de media que quieras
s<-1 # sustituye 1 por el valor de media que quieras
comp1.2<-NULL
comp2.3<-NULL
comp1.3<-NULL
for (i in 1:1000) # repite 1000 veces {
  m1<-rnorm(50,m,s) # genera muestra 1
  m2<-rnorm(50,m,s) # genera muestra 2
  m3<-rnorm(50,m,s) # genera muestra 3
  IC1.2<-c(t.test(m1,m2)$conf.int[1],t.test(m1,m2)$conf.int[2])
  comp1.2[i]<-(IC1.2[1]<0 & IC1.2[2]>0)
  IC2.3<-c(t.test(m2,m3)$conf.int[1],t.test(m2,m3)$conf.int[2])
  comp2.3[i]<-(IC2.3[1]<0 & IC2.3[2]>0)# compara 2 y 3
```

```
IC1.3<-c(t.test(m1,m3)$conf.int[1],t.test(m1,m3)$conf.int[2])
comp1.3[i]<-(IC1.3[1]<0 & IC1.3[2]>0) # compara 1 y 3
}
```

vosotros tenéis que cambiar la media y la varianza

a1) ¿ Qué porcentaje de veces no rechazamos cada una de las tres hipótesis Comprobad el resultado en vuestra muestra

a2) ¿Qué % de las veces no rechazamos ninguna de las tres hipótesis de que las tres medias son iguales?

a3) Explicar qué es la corrección de Bonferroni y cuál sería en este caso. Aplícalo a las muestras que habías obtenido y vuelve a calcular el % de no rechazo de las tres hipótesis nula: las tres medias son iguales. Comentar los resultados.

b) Responder a las preguntas de forma breve y razonada

b1) ¿La estimación bootstrap sirve para estimar la distribución de un estadístico o bien para estimar la distribución teórica de la población?

b2) Indica brevemente como procederías para obtener la estimación bootstrap de la mediana de una variable. Genera una muestra de tamaño 100 de una distribución normal con media 180 y desviación típica 20 y calcula el estimador bootstrap de la mediana y su intervalo de confianza al 95%.