

Máster «IoT»

Complementos de Estadística

19 de diciembre de 2018

Una respuesta se dará por válida sólo si está debidamente justificada. En general, debes incluir el código fuente empleado para llevar a cabo las operaciones. Envía tus respuestas al correo `carleos@uniovi.es` o guárdalas en tu carpeta de `@carleos2.epv.uniovi.es:/home/iot/`.

1. Carga en R los datos del fichero `@carleos2.epv.uniovi.es:/home/iot/temperaturas.one`. Contiene mediciones de los sensores de un ordenador Asus Aspire One a lo largo de una noche, un día y otra noche. La primera variable será la «temperatura interior» y la segunda será la «temperatura exterior».

```
## esta versión funciona en algunas codificaciones:
T <- read.table ("/home/iot/dat/temperaturas.one", sep = "Â")
## esta versión funciona siempre:
T <- read.table ("/home/iot/dat/temperaturas.one", sep = "\xc2")
temp1 <- T[,1]
temp2 <- T[,2]
```

2. ANÁLISIS DESCRIPTIVO UNIVARIANTE de la temperatura exterior:

- a) Calcula su media y su desviación típica.

```
mean (temp2) # 21,362.23
sd (temp2) # 2,769.19
```

- b) Haz los siguientes diagramas:

- diagrama de barras de la distribución de la temperatura exterior
`barplot (table (temp2))`
- diagrama de la secuencia temporal de la temperatura exterior
`plot (temp2, type="l")`

3. ANÁLISIS DESCRIPTIVO BIVARIANTE de las temperaturas:

- a) Representálas mediante una nube de puntos.

```
plot (temp1 ~ temp2)
plot (jitter (temp2), jitter (temp1))
```

- b) Halla el coeficiente de correlación lineal de Pearson entre ellas.

```
cor (temp1, temp2) # 0,952.653.3
```

- c) Halla la recta de regresión lineal de forma que la temperatura exterior sea la variable independiente (X) y la temperatura interior sea la variable dependiente (Y).

```
lm (temp1 ~ temp2) # 29,797.1 + 0,836.3·temp2
```

4. PROBABILIDAD:

- a) ¿Pueden considerarse independientes las observaciones que componen la muestra de temperaturas interiores en los ejercicios de arriba?

Las mediciones se realizan cada minuto, por lo que la temperatura en cierto registro estará muy relacionada con la temperatura del registro precedente y con la del subsiguiente. No son observaciones independientes.

- b) ¿Pueden considerarse independientes las observaciones de la variable «speed» del conjunto de datos «cars» de R? (Véase `?cars`)

La ayuda de R no aclara sobre el muestreo. Tendríamos que adquirir el libro referenciado para tener más detalles. Sólo se menciona que son datos de «coches». Como tampoco tenemos indicios de dependencia, podemos suponer independencia, si avisamos de que es una hipótesis asumida sin más pruebas.

Hay pruebas estadísticas para comprobar independencia entre observaciones (prueba de rachas...) pero no las hemos visto (vimos la prueba χ^2 sobre dependencia de dos variables).

- c) Genera 100 valores pseudoaleatorios con distribución exponencial de media igual a 10.

```
rexp (100, 1/10) # comprobar que su «mean» y su «sd» sean aproximadamente 10
```

- d) Haz un diagrama de barras de 100.000 valores pseudoaleatorios de la variable «suma de dos dados cúbicos».

```
dado1 <- sample (6, 1e5, replace=TRUE)
dado2 <- floor (runif (1e5, 1, 7)) # equivale al "sample" anterior
barplot (table (dado1 + dado2))
```

5. INFERENCIA:

- a) Contrasta si la media de la variable «speed» puede asumirse igual a 20.

```
shapiro.test (cars$speed) # es gaussiana; no importa mucho porque length(cars$speed) > 30
t.test (cars$speed, mu=20) $ p.value # sale 1,367e-07 < 0,05 se rechaza la igualdad a 20
```

- b) Para los datos «sleep» de R, analiza si «group» influye en «extra».

```
## contraste sobre la mediana de la diferencia (muestras pareadas)
?sleep # ¿hay diferencias significativas en sueño extra según somnífero?
boxplot (extra ~ group, sleep)
extra1 <- sleep$extra [sleep$group == '1']
extra2 <- sleep$extra [sleep$group == '2']
dif <- extra1 - extra2
shapiro.test (dif) # 0,033 no es gaussiana «por poco»
wilcox.test (extra1, extra2, paired=TRUE) # asegurarse del orden en ambos vectores
wilcox.test (sleep ~ group, sleep, paired=TRUE) # p-valor = 0,009.091
wilcox.test (dif) # los tres «wilcox.test» son equivalentes
t.test (extra1, extra2, paired=TRUE) # no muy distinto: pval = 0,002.833
```

- c) En un control de calidad de la proporción p de defectuosos, se considera $H_0 : p=0,01$ bajo control; $H_1 : p>0,01$ fuera de control. ¿Qué decisión tomarías si en una muestra de tamaño 100 se observan 10 defectuosos? Considera el nivel de significación $\alpha = 0,001$ (uno por mil).

```
prop.test (10, 100, p=0.01, alternative="greater") # aproximado pval < 2.2e-16
binom.test (10, 100, p=0.01, alternative="greater") # exacto pval = 7,632e-08
```

6. FILTRADO:

- a) Aplica en R una media móvil a los datos de temperatura exterior del ejercicio primero. Realiza un gráfico que incluya la secuencia original y la secuencia alisada.

```
alisada <- filter (temp2, rep(1/61, 61)) # medias por hora
plot (temp2, type="l")
lines (1:length(temp2), alisada, col="2", lwd=5)
```

- b) Aplica en C una mediana móvil a esos mismos datos.

```
#include <stdio.h>
#include <stdlib.h>

#include <gsl/gsl_math.h>
#include <gsl/gsl_movstat.h>
#include <gsl/gsl_vector.h>

int
main(void)
{
    const size_t N = 1938; /* length of time series */
    const size_t K = 61; /* window size */
    gsl_movstat_workspace * w = gsl_movstat_alloc(K);
    gsl_vector *x = gsl_vector_alloc(N);
```

```

gsl_vector *xmediana = gsl_vector_alloc(N);
size_t i;
/* Leeremos desde C un fichero generado en R como sigue: */
/* cat (temp2, sep="\n", file="/tmp/temp2") */
FILE * datos = fopen ("/tmp/temp2", "r");
char * registro;
size_t anchura = 0;
char * cola;

for (i = 0; i < N; ++i)
{
    getline (&registro, &anchura, datos);
    double xi = strtod (registro, &cola);
    gsl_vector_set(x, i, xi);
}

/* compute moving statistics */
gsl_movstat_median(GSL_MOVSTAT_END_PADVALUE, x, xmediana, w);

/* print results */
for (i = 0; i < N; ++i)
{
    printf("%zu %f %f\n",
           i,
           gsl_vector_get(x, i),
           gsl_vector_get(xmediana, i));
}

gsl_vector_free(x);
gsl_vector_free(xmediana);
gsl_movstat_free(w);
fclose (datos);

return 0;
}

```