

ESTADÍSTICA BÁSICA

Soluciones a los Ejercicios para la Evaluación Continua

Curso 2019-2020

Alfonso García Pérez

Universidad Nacional de Educación a Distancia

Copyright ©2020 Alfonso García Pérez

“No está permitida la reproducción total o parcial de este libro, ni su tratamiento informático, ni la transmisión de ninguna forma o por cualquier medio, ya sea electrónico, mecánico, por fotocopia, por registro u otros medios, sin el permiso previo y por escrito de los titulares del Copyright. El contenido de este libro está registrado por el autor en el Registro de la Propiedad Intelectual y protegido por la Ley, que establece penas de prisión además de las correspondientes indemnizaciones para quien lo plagia”

Edita: Universidad Nacional de Educación a Distancia

Capítulo 1

Soluciones a los Ejercicios para la Evaluación Continua

Si se quiere optar por la modalidad de Evaluación Continua, estos ejercicios deberán ser entregados antes del 10 de Enero obligatoriamente en el Curso Virtual. Se ruega entregar en un formato fácil de acceder como por ejemplo pdf. Los Tutores deberán de haber calificado estas pruebas con una nota de 0 a 10 (que ponderadas en la nota final por 0'2 sólo sumarán la calificación de la Prueba Presencial entre 0 y 2) antes del comienzo de la primera semana de las pruebas presenciales. Y, como mucho, al comienzo de la segunda semana de pruebas presenciales los alumnos pueden haber reclamado al Tutor por la nota con la que les calificó, de manera que estas calificaciones serán definitivas al final de la segunda semana de exámenes.

Las calificaciones así obtenidas se sumarán a la de la Prueba Presencial, si en ésta se obtuvo una puntuación de 4 o más puntos, truncando a 10 aquellas notas que superen este valor. Así, el alumno podrá obtener hasta una calificación de 10 puntos. No obstante, para obtener una calificación de Matrícula de Honor deberá haber obtenido un 10 en la Prueba Presencial. Por ejemplo, si un alumno obtiene un 1 en la Evaluación Continua y un 4 en la Prueba Presencial, su calificación final será de 5; si obtiene un 1 en la Evaluación Continua y un 10 en la Prueba Presencial, su calificación final será de 10 (MH); si obtiene un 2 en la Evaluación Continua y un 3'5 en la Prueba Presencial, su calificación final será de 3'5; si obtiene un 1 en la Evaluación Continua y un 9 en la Prueba Presencial, su calificación final será de 10.

Por último recordar que esta prueba es personal e intransferible; cualquier ejercicio cuya solución haya sido copiada será calificada con un 0. Serán los Tutores Intercampus quienes califiquen esta prueba.

Es imprescindible que los ejercicios sean resueltos con el paquete estadístico R teniendo que utilizar en algún caso, alguna librería que tendrá que instalar;

de esta forma veremos su soltura en el manejo de R aunque tan importante como saber usar R es saber obtener las conclusiones correctas con dicha herramienta.

Los tres ejercicios puntúan igual.

Ejercicios para la Evaluación Continua

Problema 1.1

Los datos *volumen.txt* del curso virtual son los volúmenes (en pies cúbicos) de 31 cerezos negros del Parque Nacional de Allegheny en Pensilvania seleccionados al azar (Atkinson, 1982). Se pide determinar con R: La distribución de Frecuencias Absolutas, el Diagrama de Hojas y Ramas, la Media, la Mediana, la Moda, el Primer Decil, el Tercer Cuartil, la Desviación Típica, el Recorrido y el Coeficiente de Asimetría de Pearson.

Primero debemos incorporar los datos a R en formato vector con (1), ya que es una ristra de datos y no una matriz de datos, la cual deberíamos de haber incorporado en formato `data.frame`.

```
> volumen<-scan("e:\\volumen.txt")
```

 (1)

La Distribución de Frecuencias Absolutas (EBR-sección 2.3) corresponderá a la de un carácter cuantitativo sin agrupar y se obtendrá ejecutando (2) como,

```
> table(volumen)
```

 (2)

```
volumen
10.2 10.3 15.6 16.4 18.2 18.8 19.1 19.7 19.9 21 21.3 21.4 22.2 22.6
1 2 1 1 1 1 1 1 1 1 1 1 1 1
24.2 24.9 25.7 27.4 31.7 33.8 34.5 36.3 38.3 42.6 51 51.5 55.4 55.7
1 1 1 1 1 1 1 1 1 1 1 1 1
58.3 77
1 1
```

El Diagrama de Hojas y Ramas, representación adecuada a este tipo de datos, se obtendrá ejecutando (3)

```
> stem(volumen)
```

 (3)

The decimal point is 1 digit(s) to the right of the |

```
1 | 00066899
2 | 00111234567
3 | 24568
4 | 3
5 | 12568
6 |
7 | 7
```

Las medidas de posición y dispersión habituales las obtenemos con R ejecutando (4),

```
> summary(volumen) (4)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
10.20	19.40	24.20	30.17	37.30	77.00

De ahí ya obtenemos que la Media es 30'17, la Mediana es 24'20 y el Tercer Cuartil 37'30 .

Para calcular el Primer Decil ($1/10=0.1$) debemos ejecutar (5)

```
> quantile(volumen,probs=0.1) (5)
```

10%
15.6

obteniendo que es igual a 15'6.

Para calcular la Moda con R es necesario instalar y abrir el paquete `modeest`:

```
> library(modeest)
> mfv(volumen)
[1] 10.3
```

con lo que la Moda es 10'3 como fácilmente se deducía de la distribución de frecuencias absolutas.

Como la función `var` de R nos proporciona la cuasivarianza, la Desviación Típica será 16'17,

```
> sqrt(30*var(volumen)/31)
[1] 16.17055
```

El Rango o Recorrido, diferencia entre el Máximo y el Mínimo será $77 - 10'20 = 66'8$.

```
> range(volumen)[2]-range(volumen)[1]
[1] 66.8
```

Por último, el Coeficiente de Asimetría de Pearson será, EBR-sección 2.3.4, igual a 1'23,

```
> (mean(volumen)-mfv(volumen))/sqrt(30*var(volumen)/31)
[1] 1.228837
```

lo que indica una asimetría a la derecha o positiva.

Problema 1.2

Los datos de la siguiente tabla de contingencia representan las frecuencias absolutas de esqueletos hallados en la necrópolis meroítica de Amir Abdallah (Abri, Nubia sudanesa, siglos IV-I a.C.) en función del Sexo/Edad del esqueleto y del Tipo de tumba (Fernández, 2015):

Sexo/Tipo	Cámara Oeste	Cámara Este	Cámara Norte	Cámara Sur	Fosa Simple
Hombre	96	32	10	4	5
Mujer	95	23	13	0	4
Niño(a)	32	12	5	1	5
Indeterminado	22	6	5	0	0

Determinar con R si puede aceptarse la independencia entre estas dos variables que forman la tabla de contingencia, concluyendo en consecuencia que no enterraban por razón de Sexo o Edad en un tipo u otro de tumba.

Si fuera necesario, utilizar la corrección de Yates.

Se trata de un contraste de independencia de caracteres (EBR-sección 8.2.4), en donde la hipótesis nula es que las variables que forman la tabla son independientes.

Si incorporamos los datos a R

```
> X<-matrix(c(96,32,10,4,5,95,23,13,0,4,32,12,5,1,5,22,6,5,0,0),ncol=5,byrow=T)
> colnames(X)<-c("Cámara Oeste","Cámara Este","Cámara Norte","Cámara Sur","Fosa Simple")
> rownames(X)<-c("Hombre","Mujer","Niño(a)","Indeterminado")
```

y calculamos las frecuencias esperadas,

```
> round(chisq.test(X)$expected,2)
```

	Cámara Oeste	Cámara Este	Cámara Norte	Cámara Sur	Fosa Simple
Hombre	97.34	29.00	13.11	1.99	5.56
Mujer	89.39	26.64	12.04	1.82	5.11
Niño(a)	36.42	10.85	4.91	0.74	2.08
Indeterminado	21.85	6.51	2.94	0.45	1.25

vemos que algunas celdillas tienen frecuencias esperadas menores que 5 por lo que, según el enunciado del ejercicio, procedemos a ejecutar el test de la chi-cuadrado utilizando la corrección de Yates:

```
> chisq.test(X,correct=T)

Pearson's Chi-squared test

data:  X
X-squared = 14.168, df = 12, p-value = 0.2901
```

El p-valor de este test, 0'2901, suficientemente grande, sugiere aceptar la hipótesis nula de independencia de ambas variables y concluir que enterraban a sus muertos en los diferentes tipos de tumbas sin tener en cuenta el Sexo o Edad de los fallecidos.

Problema 1.3

Los datos pulso.txt del curso virtual, representan los valores de X = Estatura (en cm.) e Y = Pulso en reposo (en latidos por minuto) de una muestra de pacientes de un hospital (Hand, et al., 1994, pp. 339).

Determinar con R la recta de regresión lineal simple de Y sobre X , representarla junto a la nube de puntos y, también con R, analizar si esta regresión es significativa para predecir el Pulso en función de la Estatura.

Incorporamos los datos con (1) y obtenemos la recta de regresión solicitada con (2), que resulta ser igual a

$$Y = 46'9069 + 0'2098 X$$

```
> pulso<-read.table("e:\\pulso.txt",header=T) (1)
> pulso
      X   Y
1  160  68
2  167  80
.....
48 145  64
49 170  84
50 175  72

> recta<-lm(pulso$Y~pulso$X) (2)
> recta

Call:
lm(formula = pulso$Y ~ pulso$X)
```


Coefficients:

(Intercept)	pulso\$X
46.9069	0.2098

La representación gráfica de la nube de puntos y la recta de regresión la podemos obtener con, por ejemplo, la siguiente secuencia aunque se pueden utilizar diferentes alternativas,

```
> plot(pulso$X,pulso$Y,pch=16,col=2)
> abline(recta,col=4,lwd=2)
```

obteniendo la Figura 1.1.

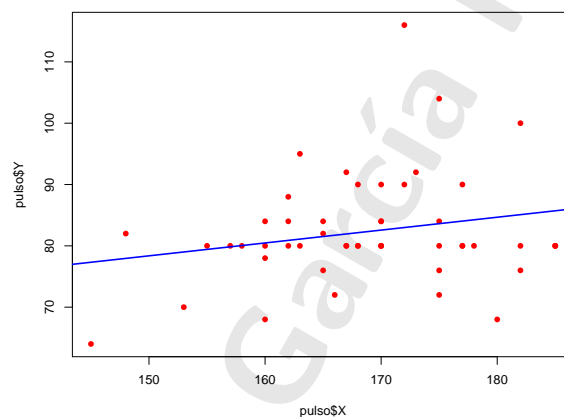


Figura 1.1 : Nube de puntos y recta de regresión

De esta figura ya se desprende que la regresión no va a ser muy buena. Para confirmarlo ejecutamos un test de hipótesis nula de ser cero el coeficiente de regresión obteniendo como p-valor el dado en (3), 0'1279, suficientemente alto como para aceptar la hipótesis nula y concluir que la recta de regresión no es significativa para predecir la variable X a la variable Y .

```
> summary(recta)
```

Call:

```
lm(formula = pulso$Y ~ pulso$X)
```

Residuals:

Min	1Q	Median	3Q	Max
-16.666	-4.876	-1.520	3.424	33.012

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	46.9069	22.8793	2.050	0.0458 *
pulso\$X	0.2098	0.1354	1.549	0.1279

(3)

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.811 on 48 degrees of freedom

Multiple R-squared: 0.04762, Adjusted R-squared: 0.02778

F-statistic: 2.4 on 1 and 48 DF, p-value: 0.1279

Referencias

- Atkinson, A.C. (1982). Regression diagnostics, transformations and constructed variables (with discussion). *Journal of the Royal Statistical Society, Series B*, **44**, 1-36.
- Cabrero Ortega, Y. y García Pérez, A. (2015). *Análisis Estadístico de Datos Espaciales con QGIS y R*. Editorial UNED. (código 0105010CT01A01).
- Fernández Martínez, V.M. (2015). *Arqueo-Estadística. Métodos Cuantitativos en Arqueología*. Alianza Editorial.
- García Pérez, A. (2008). (EAR): *Estadística Aplicada con R*. Editorial UNED. (código 0137352PB01A01).
- García Pérez, A. (2008). (CB): *Estadística Aplicada: Conceptos Básicos*, segunda edición. Editorial UNED. (código 0184011EP01A02).
- García Pérez, A. (2010). (EBR): *Estadística Básica con R*. Editorial UNED. (código 6102104GR01A01).
- García Pérez, A. (2014). (ID): *La Interpretación de los Datos. Una Introducción a la Estadística Aplicada*. Editorial UNED. (código 0105008CT01A01).
- García Pérez, A. (1998). *Problemas Resueltos de Estadística Básica*. Editorial UNED. (código 0184011EP01A02).
- García Pérez, A. (2008). *Ejercicios de Estadística Aplicada*. Editorial UNED. (código 0135284CU01A01).
- Hand, D.J., Daly, F., Lunn, K.J., McConway y Ostrowski, E. (1994). *Small Data Sets*. Editorial Chapman and Hall.