info@tuacademiafacil.com

Estadística

(Grado Ingeniería Informática UOC)

05.568

Contenidos:

- **Estadística Descriptiva**. Introducción al análisis de datos.
- **Probabilidad** variables aleatorias.
- **Inferencia Estadística**.
  - Teorema central del límite
  - Intervalos de confianza
  - Contraste de Hipótesis.
  - Contraste de dos muestras.
- **Regresión Lineal.**

Evaluación continua

Consta de 7 pruebas de evaluación continuada (PEC).

Las **PEC 1-6** constan de dos partes:

- un ejercicio que hay que **resolver con R** y entregar aparte.
- un **cuestionario en Moodle.**
  Una vez abierto el cuestionario, hay 72 horas (3 días) para cerrarlo.
  Se puede realizar dos veces (abrirlos dos veces) y la nota resultante es la mejor de los dos intentos.

En la **PAC7** (o cuestionario final **en Moodle**) sólo habrá

- un intento posible,
- adjuntar la resolución a mano de los ejercicios (paso a paso), y
- tendréis un máximo de 2 horas y 15 minutos para realizarla.

Calificaciones:

- A: Calificación muy buena
- B: Calificación buena
- C+: Calificación suficiente
- C-: Calificación baja
- D: Calificación muy baja
- N: No se emite calificación

Nota final NF = 0,40 · Media Notas PEC 1-6 + 0,60· Nota PEC7

Se os dará un extra de 0.5 a los que hacéis el **cuestionario inicial**.
Esta nota extra se os sumará a Media Notas PEC 1-6.

Para superar la asignatura:
1) **Media Notas PEC 1-6 $\geq$ 4**
2) **Nota PEC7 $\geq$ 3.5**

info@tuacademiafacil.com

Estadística descriptiva

Formulario "oficial" UOC

Media aritmética de datos tabulados	Media aritmética de datos agrupados	Varianza poblacional	Varianza muestral	Desviación típica muestral y poblacional
$\overline{x} = \frac{\sum_{i=1}^{k} n_i x_i}{N} = \sum_{i=1}^{k} f_i x_i$	$\overline{x} = \frac{\sum_{i=1}^{k} n_i m_i}{N} = \sum_{i=1}^{k} f_i m_i$ m_i = marca de clase	$s_x^2 = \frac{\sum_{i=1}^{k} n_i (x_i - \overline{x})^2}{N} = \frac{\sum_{i=1}^{k} n_i x_i^2}{N} - \overline{x}^2$	$s_x^2 = \frac{\sum_{i=1}^{k} n_i (x_i - \overline{x})^2}{N-1}$	$s_x = \sqrt{s_x^2}$

Pregunta 1
Sin responder aún
Puntúa como 1,00
Marcar pregunta

Los siguientes datos corresponden a las notas de una asignatura de 6 estudiantes de la UOC:
{5,4,4,9,6,6}.
Calculad su media, desviación estándar (poblacional) y la mediana y responded en la forma siguiente:
$\overline{x}$ = media (50% de la nota)
σ_x = desviación estándar **poblacional** (25%)
m = mediana (25%)
Usad el punto como separador decimal (por ejemplo 2.45) y redondead a tres decimales

Respuesta:
$\overline{x} =$ 5.667
$\sigma_x =$ 1.7
$m =$ 5.5

Media aritmética: (promedio, mitjana)

$$\overline{x} = \frac{x_1 + x_2 + \cdots + x_n}{N} = \frac{5+4+4+9+6+6}{6} = \frac{34}{6}$$

$\approx 5{,}667$

Nota	Frecuencia		
x_i	n_i	$x_i \cdot n_i$	$x_i^2 \cdot n_i$
4	2		16 · 2
5	1		25 1
6	2		36 · 2
9	1		81 · 1
Total	N=6		210

Desviación estándar muestral y poblacional: $\sigma = \sqrt{\sigma^2}$

σ^2 : varianza

Desviación poblacional: $\sigma = \sqrt{\frac{\sum_{i=1}^{N} x_i^2 \cdot n_i}{N} - \overline{x}^2}$

$$\sigma = \sqrt{\frac{210}{6} - \left(\frac{34}{6}\right)^2} \Rightarrow \sigma = 1{,}69967\ldots$$

$$\underbrace{4 \quad 4}_{2} \quad 5 \quad 6 \quad \underbrace{6 \quad 9}_{2}$$

$$\frac{5+6}{2} = 5.5$$

1.2 (datos tabulados)

info@tuacademiafacil.com

Pregunta 2

Sin responder aún

Puntúa como 1,00

Marcar pregunta

Dadas las siguientes observaciones sobre el número y la superficie de los pisos de una inmobiliaria,

n_i	x_i
5	71
3	79
6	71
7	116

calculad la media, desviación estándar (poblacional) y la mediana de la superficie de los pisos y responded en forma siguiente:

$\bar{x}$ = media (50%)

σ_x = desviación estándar (poblacional) (25%)

m = mediana (25%)

Usad el punto como separador decimal (ejemplo 2.45) y redondead a tres decimales.

Respuesta:

$\bar{x}$ = 87.143

σ_x = 20.580

m = 71

$$\textit{Media aritmética}: \bar{x} = \frac{\sum_{i=1}^{n} x_i \cdot n_i}{N}$$

$$\bar{x} = \frac{1830}{21} = \frac{610}{7} \approx 87.143$$

M2 Superficie	Frecuencia	Frecuencia Acumulada		
x_i	n_i	N_i	$x_i \cdot n_i$	$x_i^2 \cdot n_i$
71	11	11	781	55451
79	3	11+3 = 14	237	18723
116	7	14+7 = 21	812	94192
Total	N=21		1830	168366

$$\textit{Desviación poblacional}: \sigma = \sqrt{\frac{\sum_{i=1}^{N} x_i^2 \cdot n_i}{N} - \bar{x}^2} = \sqrt{\frac{168366}{21} - \left(\frac{610}{7}\right)^2} = 20.580$$

$\sigma =$

$$\textit{Posición de la Mediana} = \frac{N+1}{2} = \frac{21+1}{2} = \frac{22}{2} = 11^{\circ} \leftarrow \text{posición} \qquad m = 71$$

(mirar en la columna Frecuencia acumulada)

info@tuacademiafacil.com

Pregunta 3
Sin responder aún
Puntúa como 1,00
Marcar pregunta

Los siguientes datos corresponden a la cantidad de libros comprados durante el semestre anterior de 11 estudiantes de la UOC:
{4,2,6,8,17,16,3,20,5,8,14}.
Calculad la mediana y el sus cuartiles, y responded en la siguiente forma:
Q_1= primer cuartil (25% de la nota de la pregunta)
m = mediana (50% de la nota)
Q_3= tercer cuartil (25% de la nota)
Usad el punto como separador decimal (ejemplo 2.45) y redondead a tres decimales.

Respuesta:
Q1 = 4
Q_2 m = 8
Q3 = 16

x	2	3	4	5	6	8	8	14	16	17	20
Posición	1ª	2º	3ª	4ª	5ª	6ª	7º	8º	9º	10º	11º
Cuartiles			Q1			Q2 m			Q3		

$$\textit{Posición de la Mediana}(Q2) = \frac{N+1}{2} = \frac{11+1}{2} = \frac{12}{2} = 6$$

$$\textit{Posición del } Q_1 = \frac{N+1}{4} \cdot 1 = \frac{11+1}{4} \cdot 1 = 3^{a}$$

$$\textit{Posición del } Q_3 = \frac{N+1}{4} \cdot 3 = \frac{11+1}{4} \cdot 3 = 9^{a}$$

info@tuacademiafacil.com

Pregunta 4
Sin responder aún
Puntúa como 1,00
Marcar pregunta

Los siguientes datos corresponden a la cantidad de libros comprados durante el semestre anterior de 10 estudiantes de la UOC:

x_i	n_i
2	1
10	3
15	6

libros	Frecuencia	Frecuencia Acumulada
x_i	n_i	N_i
2	1	1
10	3	1+3=4
15	6	4+6=10
Total	N=	10

Calculad la mediana y los cuartiles, y responded en la siguiente forma:

Q_1= primer cuartil (25% de la nota de la pregunta)

m = mediana (50% de la nota)

Q_3= tercer cuartil (25% de la nota)

Usad el punto como separador decimal (ejemplo 2.45) y redondead a tres decimales.

Respuesta:

Q1 = 10

m = 15

Q3 = 15

La Mediana (Q2) de un conjunto con **N par** es el valor medio de las dos observaciones que ocupan el lugar más próximo a la Posición de la Mediana.

$$\text{posic. } Q_1 = \frac{N+1}{4} \cdot 1 = \frac{10+1}{4} \cdot 1 = \frac{11}{4} = 2.75^{\circ}$$

$$\text{posc. } Q_2 = \text{mediana} = \frac{10+1}{4} \cdot 2 = \frac{11}{2} = 5.5^{\circ}$$

$$\text{posc. } Q_3 = \frac{10+1}{4} \cdot 3 = \frac{11 \cdot 3}{4} = 8.25^{\circ}$$

$$\boldsymbol{Posición\ de\ la\ Mediana = \frac{N+1}{2}}$$

Posición del $Q_1 = \frac{N+1}{4} . 1 =$

Posición del $Q_3 = \frac{N+1}{4} . 3 =$

2 10 10 10 15

1.5 Transformaciones

info@tuacademiafacil.com

Pregunta 5
Sin responder aún
Puntúa como 1,00
Marcar pregunta

Supongamos que tenemos una muestra de tamaño 101 con media 44.644, mediana 45, moda 18 y desviación estándar 21.025

1. Si sumamos 30 a todas las observaciones, entonces la nueva media será 74.644, la nueva mediana 75, la nueva moda 48 y la nueva desviación estándar 21.025
2. Si multiplicamos por 7 todas las observaciones, entonces la nueva media será 312.508, la nueva mediana 315, la nueva moda 126 y la nueva desviación estándar 147.175
3. Si multiplicamos por −6 todas las observaciones, entonces la nueva media será −267.864, la nueva mediana −270, la nueva moda −108 y la nueva desviación estándar 126,15
4. Si sumamos 35 a las 37 observaciones más grandes de la muestra, la nueva mediana será 45.

1. **Sumamos** una constante (c = 30) a todas las observaciones.
Cambio de origen. Esto equivale a desplazar todas las observaciones c unidades. Creamos una **nueva variable: y = x + c**.

Medida de centro	*x*	*y = x + c*	Cambio de origen
Media	$\bar{x} = 44.644$	$\bar{y} = \bar{x} + c$	$\bar{y} = 44.644 + 30 = 74.644$
Mediana	$m_x = 45$	$m_y = m_x + c$	$m_y = 45 + 30 = 75$
Moda	$Mo_x = 18$	$Mo_y = Mo_x + c$	$Mo_y = 18 + 30 = 48$
Desviación estándar	$\sigma_x = 21.025$	$\sigma_y = \sigma_x$	$\sigma_y = 21.025$

2. **Multiplicamos** por una constante (c = 7) todas las observaciones.
Cambio de escala. Creamos una **nueva variable: y = c · x**.

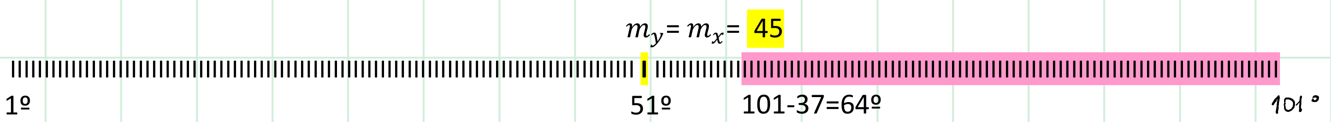
Medida	*x*	*y = c · x*	Cambio de escala
Media	$\bar{x} = 44.644$	$\bar{y} = c \cdot \bar{x}$	$\bar{y} = 44.644 \cdot 7 = 312.508$
Mediana	$m_x = 45$	$m_y = c \cdot m_x$	$m_y = 45 \cdot 7 = 315$
Moda	$Mo_x = 18$	$Mo_y = c \cdot Mo_x$	$Mo_y = 18 \cdot 7 = 126$
Desviación estándar	$\sigma_x = 21.025$	$\sigma_y = \lvert c \rvert \cdot \sigma_x$	$\sigma_y = \lvert 7 \rvert \cdot 21.025 = 147,175$
Coeficiente de variación		$CV_y = \dfrac{\sigma_x}{\bar{x} + C}$	

3. **Multiplicamos** por una constante (c = -6) todas las observaciones.

Medida	*x*	*y = c · x*	Cambio de escala
Media	$\bar{x} = 44.644$	$\bar{y} = c \cdot \bar{x}$	$\bar{y} = -6 \cdot 44.644 = -267.864$
Mediana	$m_x = 45$	$m_y = c \cdot m_x$	$m_y = -6 \cdot 45 = -270$
Moda	$Mo_x = 18$	$Mo_y = c \cdot Mo_x$	$Mo_y = -6 \cdot 18 = -108$
Desviación estándar	$\sigma_x = 21.025$	$\sigma_y = \lvert c \rvert \cdot \sigma_x$	$\sigma_y = \lvert -6 \rvert \cdot 21.025 = 126,15$

4. Sumamos 35 a las 37 observaciones más grandes de la muestra, la nueva mediana será…
Calculamos la $\textit{Posición de la Mediana} = \frac{N+1}{2} = \frac{101+1}{2} = \frac{102}{2} = 51$

Si sumamos 35 a las 37 observaciones más grandes de la muestra, esto no afectará al valor de la mediana.



info@tuacademiafacil.com

Pregunta 6
Sin responder aún
Puntúa como 0,50
Marcar pregunta

Imaginemos que queremos estudiar cual es la bebida favorita de los nadadores federados. Completad las siguientes frases

Si tenemos una lista de los nadadores federados y enviamos un correo a los 30 primeros de la lista	Elegir... No estamos haciendo un m. aleatorio
Un ejemplo de muestro por cuotas sería	Elegir... 100 ♂ y 100 mujeres
Si seleccionamos 5 comarcas al azar y después seleccionamos 53 nadadores federados de estas comarcas	Elegir... M. por conglomerados
Si asignamos un número a cada uno de los N nadadores federados, escogemos un número al azar A y a partir de éste vamos escogiendo los de la forma A+x[N/k] donde k es el tamaño de la muestra.	Elegir... M. Sistemático

Elegir...

- Elegir...
- no estamos haciendo un muestreo aleatorio
- estamos haciendo un muestreo sistemático
- estaremos haciendo un muestreo por conglomerados
- enviar entrevistadores hasta que encontremos 100 mujeres y 200 hombres entre los nadadores federados
- estamos haciendo un muestreo aleatorio simple no sistemático

Muestreo	Propiedades
Aleatorio simple	• Todos los individuos tienen la misma probabilidad de formar parte de la muestra. • Selección de los individuos uno a uno y con reposición (siempre sobre el total de la población).
Aleatorio simple usando tablas de dígitos aleatorios	• Se numeran los individuos de la población. • El primero individuo de la muestra se selecciona al azar, los siguientes, a partir de la tabla.
Sistemático	• Se numeran los N individuos de la población. • El primero individuo de la muestra de tamaño k se selecciona a partir del cálculo $\left[\frac{N}{k}\right] = \text{Parte entera del cociente.}$ • Los siguientes, a intervalos fijos.
Estratificado	• La población es dividida en Estratos (grupos disjuntos homogéneos con características distintivas). • Se hace un muestreo aleatorio simple en cada estrato. • En la muestra se conserva la proporción que cada estrato representa en la población.
Por Conglomerados	• Los Conglomerados son unidades(físicas o geográficas) cada uno de los cuales representa la heterogeneidad de la población. • Se selecciona una muestra de conglomerados y a su vez se escogen, al azar, individuos de cada uno de los conglomerados.
Polietápico	Se aplica la estratificación a los conglomerados.
Por cuotas	• Se categorizan los individuos de la población • La proporción de individuos de cada categoría en la muestra es equivalente a la proporción de esa categoría en la población. El entrevistador selecciona individuos hasta cubrir las cuotas.