Ejercicio 1 Estadística Descriptiva (X puntos)

Identifique el o los *outliers* correspondientes a los siguientes datos:

~~23~~	~~34~~	~~27~~	~~7~~	~~30~~	~~26~~	~~28~~	~~31~~	~~34~~	~~8~~

(A) 7

(B) 34

(C) 7 y 34

(D) 31 y 34

(E) 7 y 8

(F) Esta muestra de datos no tiene outliers

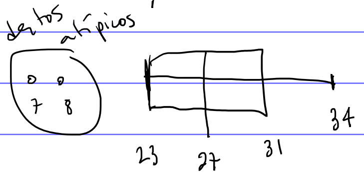
0 | 7 8 (2)
1 |
2 | 3 6 7 8 (6)
3 | 0 1 4 4 (10)

min	q_1	m	q_3	max
7	23	27	31	34

$RIC = 31 - 23 = 8$

$L_{sup} = q_3 + 1.5 \times RIC = 31 + 1.5 \times 8 = 43$

$L_{inf} = q_1 - 1.5 \times RIC = 23 - 1.5 \times 8 = 11$



Ejercicio 2 Regresión (X puntos)

Un grupo de 10 estudiantes realizó dos pruebas. Sea x_i el puntaje obtenido en la primera prueba, y sea y_i el puntaje en la segunda prueba para $i = 1, \ldots, 10$. La siguiente tabla muestra los resultados:

i	1	2	3	4	5	6	7	8	9	10
x_i	55	57	38	69	52	61	55	48	56	44
y_i	67	77	49	79	61	74	64	58	69	52

Los siguientes datos pueden ser de utilidad:

$\sum x_i = 535$	$\sum y_i = 650$	$\sum x_i^2 = 29305$	$\sum y_i^2 = 43182$	$\sum x_i y_i = 35529$

Considere el caso de un estudiante que sacó 63 puntos en la primera prueba pero que no se presentó a la segunda (y por eso no aparece en la tabla). ¿Cuántos puntos estima que podría haber sacado este estudiante si se hubiera presentado a la segunda prueba?

(A) 72 **(B)** 73 **(C)** 74 **(D)** 75 **(E)** 77 **(F)** 78

$(n-1)\,cov(x,y) = \sum (x_i - \bar{x})(y_i - \bar{y})$

$= \sum x_i y_i - n \bar{x} \bar{y}$

$= \sum x_i y_i - \dfrac{(\sum x_i)(\sum y_i)}{n}$

$= \frac{1}{n}\left[n \sum x_i y_i - (\sum x_i)(\sum y_i) \right]$

$(n-1) s_x^2 = \frac{1}{n}\left[n \sum x_i^2 - (\sum x_i)^2 \right]$

$y = \alpha + \beta x$

$\beta = r \dfrac{s_y}{s_x} = \dfrac{cov(x,y)}{s_x \cancel{s_y}} \cdot \dfrac{\cancel{s_y}}{s_x} = \dfrac{cov(x,y)}{s_x^2}$

$\beta = \dfrac{10 \times (35\,529) - (535)(650)}{10 \times (29\,305) - (535)^2} = \dfrac{35\,5290 - 347\,750}{29\,3050 - 286\,225}$

$= \dfrac{7540}{6\,825} = 1.1$

$\alpha = \dfrac{650}{10} - (1.1) \times \dfrac{(535)}{10} = 650 - 588.5 = 61.5 / 10.$

$\boxed{y = 1.1\,x + 6.15}$

$\hat{y} = 1{,}1 \times (63) + 61.5 =$ $\boxed{\hat{y} = 75.45}$

Ejercicio 3 Test de Permutaciones (X puntos)

Ana y Beto trabajan como auxiliar de limpieza en un hotel. Beto le apuesta a Ana que él limpia las habitaciones más rápido que ella. Para comprobarlo asignan las 10 habitaciones del hotel al azar en dos grupos, 5 de ellas las limpia Ana y las otras 5 las limpia Beto. La siguiente tabla muestra el resultado en minutos:

Ana:	22	19	15	19	16
Beto:	17	16	15	15	12

La siguiente figura contiene la distribución de aleatorización (o distribución nula) del estadístico X que consiste en la suma de los minutos totales efectuados por Ana menos la suma de los minutos totales efectuados por Beto.

$X =$ suma A $-$ suma B $= 91 - 75 = 16$

p-valor $= \dfrac{\#\{X \geq 16\}}{252} = \dfrac{12}{252} = 0.048 < 0.05$

Rechazamos Ho : Ana y Beto son iguales.

Respuesta : A

Ejercicio 4 Estimación (X puntos)

Se tiene la siguiente muestra de $n = 10$ datos que provienen de una variable aleatoria $U(0, \theta)$ con $\theta > 0$ desconocido.

i	1	2	3	4	5	6	7	8	9	10
x_i	3.46	4.96	4.72	4.21	1.06	2.58	2.92	4.85	0.21	0.91

Sea $\hat{\theta}$ el estimador por máxima verosimilitud de θ, entonces:

(A) $\hat{\theta} = 2.988$ **(C)** $\hat{\theta} = 0.21$ **(E)** $\hat{\theta} = 2.585$

(B) $\hat{\theta} = 4.96$ **(D)** $\hat{\theta} = 2.92$ **(F)** $\hat{\theta} = 3.46$

$\hat{\theta} = \max X_i = 4.96$

Ejercicio 5 Estimación (X puntos)

Se tiene una muestra $X_1, \ldots, X_n$ i.i.d. con distribución $U(0, \theta)$ donde $\theta > 0$ es un parámetro desconocido y se propone a $\hat{\theta} = \sum_{i=1}^{n} \alpha_i X_i$ como estimador de θ, siendo $\alpha_i \in \mathbb{R}$ para todo i.

Una *condición necesaria y suficiente* para que $\hat{\theta}$ sea insesgado para θ es:

(A) $\sum_i \alpha_i = 1$.

(B) $\sum_i \alpha_i = 2$.

(C) $\alpha_i = \frac{1}{n}$ para todo i.

(D) $\alpha_i = 1$ para todo i.

(E) El estimador $\hat{\theta}$ presenta sesgo para cualquier elección de los coeficientes $\alpha_1, \ldots, \alpha_n$.

(F) El estimador $\hat{\theta}$ es insesgado para cualquier elección de los coeficientes $\alpha_1, \ldots, \alpha_n$.

$\hat{\theta} = \sum \alpha_i X_i$ $\quad E(\hat{\theta}) = E\left(\sum \alpha_i X_i\right) = \sum \alpha_i \overbrace{E(X_i)}^{\theta/2} = \frac{\theta}{2} \sum \alpha_i$

Ejercicio 6 TCL (X puntos)

Se sabe que en periodo de parciales un estudiante duerme en promedio unas 5.7 horas con una desviación estándar de 1.7 horas. Se selecciona al azar un grupo de 35 estudiantes y se les pregunta cuánto durmieron el día antes del parcial.

¿Cuál es la probabilidad (aproximada) de que este grupo haya dormido en promedio más de 6 horas?

(A) 0.430 **(B)** 0.281 **(C)** 0.148 **(D)** 0.087 **(E)** 0.270 **(F)** 0.852

$n = 35$ $\quad \overline{X} =$ promedio

$$P(\overline{X} > 6) = P\left(\frac{\overline{X} - 5.7}{1.7/\sqrt{35}} > \frac{6 - 5.7}{1.7/\sqrt{35}}\right)$$

$$\simeq P\Big(N(0,1) > \underbrace{\frac{0.3\sqrt{35}}{1.7}}_{1.04}\Big) = P(N(0,1) > 1.04)$$

$$= 1 - 0.8508 = 0.1492$$

Ejercicio 7 Test de Hipótesis (X puntos)

En una fábrica embotelladora de agua mineral se utiliza una máquina para vertir el líquido dentro de las botellas. La máquina vierte el líquido con un error aleatorio cuya distribución es normal de media cero y desvío conocido de 15 ml. En su funcionamiento correcto la máquina debe vertir en promedio $\mu = 1.5$ litros de agua en cada botella.

Para un control rutinario de calidad se desea tomar una muestra de n botellas, con el objetivo de ralizar el siguiente test de hipótesis:

$$\begin{cases} H_0 : \mu = 1500 \text{ (ml)} \\ H_A : \mu < 1500 \text{ (ml)} \end{cases}$$

Se decide utilizar una región de rechazo de la forma $\{\bar{X} \leq 1500 - \varepsilon\}$, en donde $\bar{X}$ es el promedio de las n botellas, y un nivel de significancia $\alpha = 5\%$.

Suponga que la realidad es que la máquina está descalibrada y vierte en promedio $\mu = 1490$ ml. Hallar el valor de n para que la potencia del test anterior sea de 55 %.

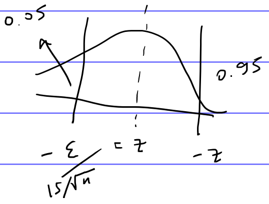
(A) 8 **(B)** 13 **(C)** 17 **(D)** 25 **(E)** 36 **(F)** 15

$\sigma = 15$ $\quad \mu_{\text{correcto}} = 1500$ $\quad \begin{cases} H_0: \mu = 1500 \\ H_A: \mu < 1500 \end{cases}$

$\{\overline{X} < 1500 - \varepsilon\}$ $\quad \alpha = 5\%$ $\quad \mu_{\text{real}} = 1490$

$$P(\bar X < 1500 - \varepsilon \mid \mu = 1500) = 0.05$$

$$P\left(\frac{\bar X - 1500}{15/\sqrt n} < \frac{-\varepsilon}{15/\sqrt n} \;\middle|\; \mu = 1500\right) = 0.05$$



$$z = -1.645$$

$$\boxed{\varepsilon = (1.645) \times \frac{15}{\sqrt n}}$$

$$P(\bar X < 1500 - \varepsilon \mid \mu = 1450) = 0.55$$

$$P\left(\frac{\bar X - 1450}{15/\sqrt n} < \frac{10 - \varepsilon}{15/\sqrt n} \;\middle|\; \mu = 1450\right) = 0.55$$

$$P\left(N(0,1) < \frac{10-\varepsilon}{15/\sqrt n}\right) = 0.55$$

$$\frac{10-\varepsilon}{15/\sqrt n} = 0.125 \qquad 10 - 1.645 \times \frac{15}{\sqrt n} = 0.125 \times \frac{15}{\sqrt n}$$

$$10 = (1.645 + 0.125) \times \frac{15}{\sqrt n}$$

$$n = (1.5)^2 \times (1.645 + 0.125)^2 = 7.05$$

$$n = \lceil 7.05 \rceil = 8$$

Ejercicio 8 Intervalos de Confianza (X puntos)

Un estudio desea determinar cuánto duran en promedio las bolsas de plástico biodegradables. Se asume que la duración medida en meses se distribuye como una variable normal con media y desvío desconocidos.

Para esto se dispone de la siguiente muestra:

17 16 18 19 16 17 17 19 18 15

Calcular el intervalo de confianza para la duración media al nivel de confianza 95 %. Expresar el resultado con 1 cifras decimales.

(A) [16.1, 19.3] (C) [15.8, 18.6] (E) [16.3, 18.1]
(B) [15.7, 21.1] (D) [16.5, 20.3] (F) [16.1, 19.1]

$n = 10$

	X_i	X_i^2
	17	289
	16	256
	18	324
	19	361
	16	256
	17	289
	17	289
	19	361
	18	324
	15	225
Suma	172	2974

$$\bar X \pm t_9^{(0.025)} \frac{S_x}{\sqrt{10}}$$

$$(n-1) S_x^2 = \sum (x_i - \bar x)^2$$

$$= \sum x_i^2 - \frac{(\sum x_i)^2}{n}$$

$$= \frac{1}{n}\left[n \sum x_i^2 - (\sum x_i)^2\right]$$

1,73

''

$$S_x^2 = \frac{1}{90}\left[10(2974) - (172)^2\right] = \frac{1}{90}\left[29740 - 29584\right] = \frac{156}{90}$$

$$\boxed{S_x = 1.32} \qquad \boxed{\bar X = 17.2} \qquad \boxed{t_9(0.025) = 2.26}$$

$$\boxed{\sqrt{10} = 3.16}$$

$17.2 \pm \frac{(2.26)(1.32)}{3.16} = 17.2 \pm 0.94$

$= [16.3 , 18.1]$

Ejercicio 1 Estadística Descriptiva (X puntos)

Identifique el o los *outliers* correspondientes a los siguientes datos:

~~55~~	~~57~~	~~38~~	~~69~~	~~52~~	~~61~~	~~55~~	~~48~~	~~56~~	~~34~~

(A) 34 y 69

(B) 34

(C) 34 y 38

(D) 69

(E) 61 y 69

(F) Esta muestra de datos no tiene outliers

```
3 | 4 8 (2)
4 | 8 (3)
5 | 2 5 5 6 7 (8)
6 | 1 9 (10)
```

min	q_1	m	q_3	max
34	48	55	57	69

$RIC = 57 - 48 = 9$

$L_{sup} = 57 + 1.5 \times 9 = 70.5$

$L_{inf} = 48 - 1.5 \times 9 = 34.5$

Ejercicio 2 Regresión (X puntos)

Un grupo de 10 estudiantes realizó dos pruebas. Sea x_i el puntaje obtenido en la primera prueba, y sea y_i el puntaje en la segunda prueba para $i = 1, \ldots, 10$. La siguiente tabla muestra los resultados:

i	1	2	3	4	5	6	7	8	9	10
x_i	55	57	38	69	52	61	55	48	56	44
y_i	67	77	49	79	61	74	64	58	69	52

Los siguientes datos pueden ser de utilidad:

$\sum x_i = 535$	$\sum y_i = 650$	$\sum x_i^2 = 29305$	$\sum y_i^2 = 43182$	$\sum x_i y_i = 35529$

Considere el caso de un estudiante que sacó 60 puntos en la primera prueba pero que no se presentó a la segunda (y por eso no aparece en la tabla). ¿Cuántos puntos estima que podría haber sacado este estudiante si se hubiera presentado a la segunda prueba?

(A) 70 **(B)** 71 **(C)** 72 **(D)** 73 **(E)** 74 **(F)** 75

$y = 1.1 \times x + 6.15$

$= 1.1 \times 60 + 6.15$

$= 72.15$

Ejercicio 3 Test de Permutaciones

Ana y Beto trabajan como auxiliar de limpieza en un hotel. Beto le apuesta a Ana que él limpia las habitaciones más rápido que ella. Para comprobarlo asignan las 10 habitaciones del hotel al azar en dos grupos, 5 de ellas las limpia Ana y las otras 5 las limpia Beto. La siguiente tabla muestra el resultado en minutos:

Ana:	19	18	17	17	11
Beto:	13	18	14	16	12

La siguiente figura contiene la distribución de aleatorización (o distribución nula) del estadístico X que consiste en la suma de los minutos totales efectuados por Ana menos la suma de los minutos totales efectuados por Beto.

$X_{obs} = 82 - 73 = 9$

$\text{pvalor} = \frac{\#\{X \geq 9\}}{252} = \frac{50}{252} \approx 0.198$

No rechazo H_0

Respuesta E

Ejercicio 4 Estimación (X puntos)

Se tiene la siguiente muestra de $n = 10$ datos que provienen de una variable aleatoria $U(0, \theta)$ con $\theta > 0$ desconocido.

i	1	2	3	4	5	6	7	8	9	10
x_i	0.41	1.15	3.99	2.31	3.39	4.24	4.06	2.31	4.34	2.68

Sea $\hat{\theta}$ el estimador por máxima verosimilitud de θ, entonces:

(A) $\hat{\theta} = 2.888$

(B) $\hat{\theta} = 2.375$

(C) $\hat{\theta} = 0.41$

(D) $\hat{\theta} = 2.68$

(E) $\hat{\theta} = 3.39$

(F) $\hat{\theta} = 4.34$

$\hat{\theta} = \max x_i = 4.34$

Ejercicio 5 Estimación (X puntos)

Se tiene una muestra $X_1,\ldots,X_n$ i.i.d. con distribución $U(0,2\theta)$ donde $\theta > 0$ es un parámetro desconocido y se propone a $\hat{\theta} = \sum_{i=1}^{n} \alpha_i X_i$ como estimador de θ, siendo $\alpha_i \in \mathbb{R}$ para todo i.

Una *condición necesaria y suficiente* para que $\hat{\theta}$ sea insesgado para θ es:

(A) $\sum_i \alpha_i = 1$.

(B) $\sum_i \alpha_i = 2$.

(C) $\alpha_i = \frac{1}{n}$ para todo i.

(D) $\alpha_i = 1$ para todo i.

(E) El estimador $\hat{\theta}$ presenta sesgo para cualquier elección de los coeficientes $\alpha_1,\ldots,\alpha_n$.

(F) El estimador $\hat{\theta}$ es insesgado para cualquier elección de los coeficientes $\alpha_1,\ldots,\alpha_n$.

$E(X_i) = \theta \qquad E(\hat{\theta}) = \theta \underbrace{\sum \alpha_i}_{=1}$

Ejercicio 6 TCL (X puntos)

Se sabe que en periodo de parciales un estudiante duerme en promedio unas 5.7 horas con una desviación estándar de 1.7 horas. Se selecciona al azar un grupo de 35 estudiantes y se les pregunta cuánto durmieron el día antes del parcial.

¿Cuál es la probabilidad (aproximada) de que este grupo haya dormido en promedio menos de 5 horas?

(A) 0.007 **(B)** 0.148 **(C)** 0.993 **(D)** 0.852 **(E)** 0.076 **(F)** 0.924

$P(\bar{X} < 5) = P\left(\frac{\bar{X} - 5.7}{1.7/\sqrt{35}} < \frac{5 - 5.7}{1.7/\sqrt{35}}\right)$

$\simeq P\left(N(0,1) < \frac{-0.7 \times \sqrt{35}}{1.7}\right) = P(N(0,1) < -2.44)$

$= 1 - 0.9927 = 0.0073$

Ejercicio 7 Test de Hipótesis (X puntos)

En una fábrica embotelladora de agua mineral se utiliza una máquina para vertir el líquido dentro de las botellas. La máquina vierte el líquido con un error aleatorio cuya distribución es normal de media cero y desvío conocido de 15 ml. En su funcionamiento correcto la máquina debe vertir en promedio $\mu = 1.5$ litros de agua en cada botella.

Para un control rutinario de calidad se desea tomar una muestra de n botellas, con el objetivo de ralizar el siguiente test de hipótesis:

$$\begin{cases} H_0 : \mu = 1500 \text{ (ml)} \\ H_A : \mu < 1500 \text{ (ml)} \end{cases}$$

Se decide utilizar una región de rechazo de la forma $\{\bar{X} \leq 1500 - \varepsilon\}$, en donde $\bar{X}$ es el promedio de las n botellas, y un nivel de significancia $\alpha = 5\%$.

Suponga que la realidad es que la máquina está descalibrada y vierte en promedio $\mu = 1490$ ml. Hallar el valor de n para que la potencia del test anterior sea de 75%.

(A) 8 **(C)** 17 **(E)** 36

(B) 13 **(D)** 25 **(F)** 15

$n = [1.5\,(1.645 + 0.675)]^2 = 12.11 \qquad n = \lceil 12.11 \rceil = 13$

Ejercicio 8 Intervalos de Confianza (X puntos)

Un estudio desea determinar cuánto duran en promedio las bolsas de plástico biodegradables. Se asume que la duración medida en meses se distribuye como una variable normal con media y desvío desconocidos.

Para esto se dispone de la siguiente muestra:

16 17 20 21 20 16 14 22 18 20

Calcular el intervalo de confianza para la duración media al nivel de confianza 95%. Expresar el resultado con 1 cifras decimales.

(A) [16.1, 19.3] **(C)** [15.8, 18.6] **(E)** [16.3, 18.1]

(B) [15.7, 21.1] **(D)** [16.5, 20.3] **(F)** [16.1, 19.1]

16	256
17	289
20	400
21	441
20	400
16	256
14	196
22	484
18	324
20	400
184	3446

$\bar{X} = 18.4 \qquad s_X^2 = \frac{1}{90}\left[10(3446) - (184)^2\right]$

$= \frac{1}{90}\left[34460 - 33856\right] = 6.71 \qquad S_X = 2.59$

$t = 2.26 \qquad \sqrt{10} = 3.16$

$18.4 \pm \frac{(2.26)(2.59)}{3.16} = 18.4 \pm 1.85$

$[16.6\ ,\ 20.3]$