1. Estadística descriptiva

Ejercicio 1

Te dan el siguiente conjunto de siete números: 4, 3, 2, 10, 6, 14, 10.

Vamos a jugar un juego con cinco partes. En cada parte, deberás elegir un solo número (no necesariamente de la lista) para representar el conjunto de siete números.

Empezamos con una suma inicial de 170 puntos, y en cada parte hay una regla diferente para calcular la disparidad entre los valores del conjunto y tu valor sugerido. Esta disparidad es una multa, y la deducirás de tu puntaje total.

Para cada caso, escribe el valor representativo que elijas y calcula tu multa. La cantidad de puntos restantes al final será tu puntaje para el juego. Intenta maximizarlo.

1. La multa es la suma de las diferencias absolutas entre tu valor elegido y los valores del conjunto.
2. Pagas un punto por cada elemento del conjunto que no sea igual a tu número.
3. La multa es igual a la diferencia absoluta máxima.
4. La multa es el valor absoluto de la suma de todas las diferencias de tu valor elegido.
5. La multa es igual a la suma de las diferencias al cuadrado entre tu valor elegido y los valores del conjunto.

Ejercicio 2

En el primer día de clases, un profesor de estadística le pide a cada estudiante que registre en una hoja la cantidad de dinero que tiene en su bolsillo ese día.

Hombres:	81	26	8	10	0	20	14	33	50	10		
	0	12	23	55	28	56	53	55	2			
Mujeres:	4	128	1	0	73	2	8	3	24	94	30	39
	10	146	0	37	10	22	6	8	10	47	33	7

1. Construir un diagrama de tallo y hojas espalda con espalda para la cantidad de dinero que llevan los hombres y las mujeres.
2. Combinar ambos conjuntos de mediciones y construir un diagrama de tallo y hojas para los datos combinados.
3. Comparar el promedio y la mediana de los datos combinados.

Ejercicio 3

Sean $x_1, x_2, \ldots, x_n$ e $y_1, y_2, \ldots, y_n$ dos conjuntos de mediciones. Denotamos por $\bar{x}$ el promedio, y

$$\sigma_x = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

el desvío estándar de las $x_i's$. Análogamente para y. Probar que:

1. $\overline{x+y} = \overline{x} + \overline{y}$.
2. $\overline{ax+b} = a\overline{x} + b \; \forall a, b \in \mathbb{R}$.
3. $\sigma_x^2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2 - \overline{x}^2$.
4. $\sigma_{x+y}^2 = \sigma_x^2 + \sigma_y^2 + 2\left(\frac{1}{n} \sum_{i=1}^{n} x_i y_i - \overline{x}\,\overline{y}\right)$.

Ejercicio 4

Edad	Hombres	Mujeres	Total
0	565	466	1031
1-4	116	87	203
5-9	69	59	128
10-14	78	63	141
15-19	319	181	500
20-24	501	202	703
25-29	633	218	851
30-34	655	328	983
35-39	848	454	1302
40-44	1226	750	1976
45-49	1633	1024	2657
50-54	2459	1569	4028
55-59	3375	2093	5468
60-64	4669	2909	7578
65-69	6152	3745	9897
70-74	7436	4810	12246
75-79	9526	6914	16440
80-84	12619	11290	23909
85-89	12455	15430	27885
90-94	7113	12946	20059
95-99	2104	5637	7741
≥ 100	241	1128	1369
Total	74794	72304	147098

Los datos a la izquierda representan la edad al morir de una gran muestra de personas colectada por la Oficina Australiana de Estadísticas en 2012.

1. Dibujar un histograma (de área 1) de los datos de la tabla para hombres, para mujeres, y para ambos combinados. Para dibujarlos, considerar intervalos de longitud 10 años, por ejemplo, 0-9, 10-19,..., 90-99, y ≥ 100.
2. ¿Es posible calcular el promedio exactamente a partir de los datos dados en la tabla? Dar una cota inferior y cota una superior para el promedio, en los tres casos.
3. Dar una aproximación para la mediana y compararla con el promedio.

Ejercicio 5

Un fabricante está interesado en la vida útil de un nuevo tipo de bombita. Se pone a prueba una muestra aleatoria de 120 bombitas y se registra el número de horas de luz que cada foco proporciona antes de quemarse. A la derecha se muestran los resultados en un diagrama de tallo y hojas (en decenas de horas). Hacer una tabla de frecuencias y dibujar un histograma de la vida útil de estas bombitas.

6	1
6	59
7	000
7	677888
8	11234
8	6667888999
9	000111233333344
9	5679999
10	000122233334
10	5666788889
11	00000011112233334444
11	5567777788889
12	2344
12	57789
13	004
13	8
14	2

Ejercicio 6

Los siguientes datos representan la tasa de ocupación hospitalaria de varios hospitales de Estados Unidos, a la izquierda para el año 1976, y a la derecha para el año 1985. La tasa de ocupación en los hospitales se mide como el número promedio de camas ocupadas por día por cada 100 camas disponibles. Construir un diagrama de caja (boxplot) para la tasa de ocupación para cada año, ambos sobre el mismo eje.

	5	6
	5	88999
0	6	001
	6	22222233
554	6	44445555
77	6	6677777
999998	6	889
111100	7	0011
33333332222	7	223
544	7	5
7777766	7	67
999888	7	999
100	8	00
2	8	
4	8	
7	8	6

Ejercicio 7

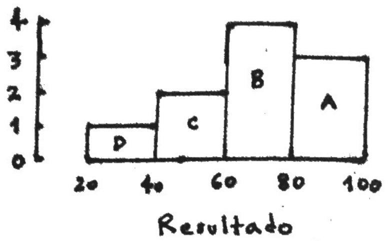
Los siguientes datos corresponden al contenido de azúcar en gramos por porción en cereales para el desayuno que se venden en cierto supermercado. Corresponden a los estantes superior e inferior de las góndolas del supermercado.

Estantes superiores		Estantes inferiores
000	0	
000	1	
000	2	000
000000	3	0
000	4	
00000000	5	
0000000	6	0
0000	7	
0000000	8	
00000000000	9	0000
00000	10	0000000000
000	11	000000
00	12	0000000000000000
000000	13	0000000000000000000
000000000000	14	000000000
0000	15	000000
	16	000000000
0000	17	0
0	18	0
0	19	0000
00	20	

1. Hacer un diagrama de caja (boxplot) para cada grupo, ambos sobre el mismo eje.
2. ¿Qué se puede concluir?

Ejercicio 8

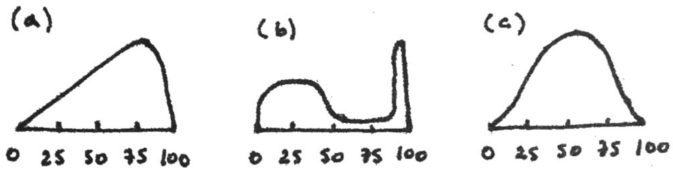
El histograma abajo muestra los resultados finales de un cierto curso.



1. ¿Qué bloque representa los estudiantes que obtuvieron entre 60 y 80?
2. Diez por ciento obtuvo entre 20 y 40. ¿Qué porcentaje obtuvo entre 40 y 60?
3. ¿Queé porcentaje obtuvo más de 60?

Ejercicio 9

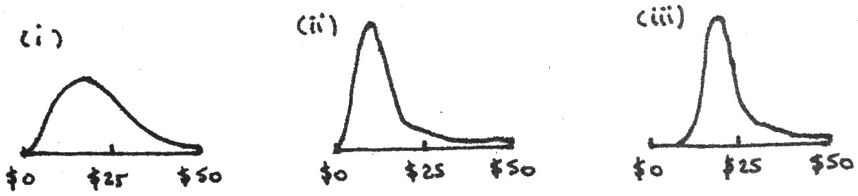
Abajo hay bosquejos de histogramas de los puntajes finales de tres clases diferentes. Los resultados varían de 0 a 100, siendo 50 el puntaje mínimo de aprobación.



1. Para cada clase ¿el porcentaje de los que aprobaron fue: alrededor de 50 %, muy arriba de 50 %, o muy abajo de 50 %?
2. Una clase tuvo marcadamente dos grupos de estudiantes diferentes, uno con muy buenas notas y otro con malas ¿cuál fue?
3. En la clase (b) ¿hubo más gente que obtuvo entre 40-50 o entre 90-100?

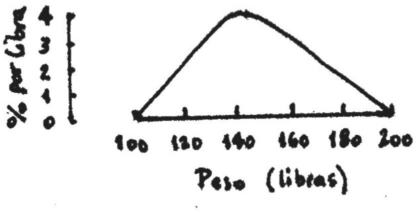
Ejercicio 10

Un investigador colecta datos sobre salarios por hora de tres grupos de personas. Los del grupo B ganan cerca del doble que los del grupo A. Los del grupo C ganan cerca de $10 por hora más que los del grupo A. ¿Qué histograma corresponde a qué grupo?



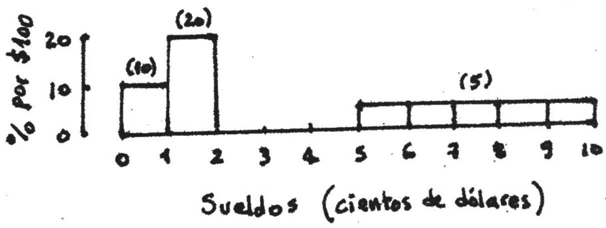
Ejercicio 11

Alguien a hecho un bosquejo de un histograma sobre el peso de un grupo de personas ¿qué está mal?



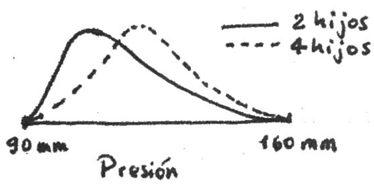
Ejercicio 12

Un histograma de salarios mensuales de un grupo de empleados es mostrado abajo. Tres bloques están faltando. Si se sabe que miden lo mismo, ¿cuál es su altura?



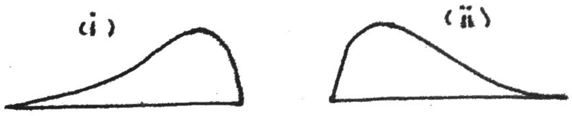
Ejercicio 13

En un estudio médico se compara la presión sanguínea de mujeres con diferentes cantidades de hijos. Abajo hay un bosquejo del histograma para mujeres con 2 y 4 hijos. ¿Qué grupo tiene mayor presión sanguínea? ¿Tener más hijos hace que la presión sangínea de la madre cambie? ¿o puede el cambio deberse a algún otro factor?



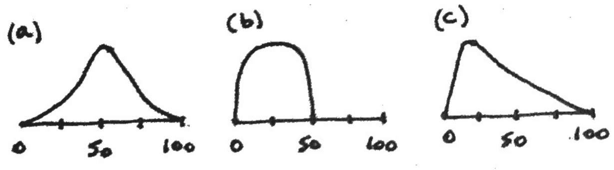
Ejercicio 14

Dos histogramas se bosquejan abajo. Uno muestra la distribución de edad al morir por causas naturales y el otro muestra la distribución de edad al morir por accidentes. ¿Cuál es cual y por qué?



Ejercicio 15

Abajo se bosquejan histogramas de tres tablas de datos. ¿Cuánto vale el promedio? Opciones: 25, 40, 50, 60, 75. ¿Es la mediana igual al promedio? ¿está hacia la izquierda? ¿hacia la derecha?



Ejercicio 16

Abajo se bosquejan histogramas de tres tablas de datos. Unir el bosquejo con la descripción.

(i) $\bar{x} \approx 3.5,\ SD \approx 1$

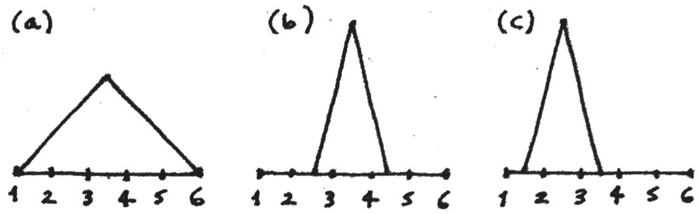
(ii) $\bar{x} \approx 3.5,\ SD \approx 0.5$

(iii) $\bar{x} \approx 3.5,\ SD \approx 2$

(iv) $\bar{x} \approx 2.5,\ SD \approx 1$

(v) $\bar{x} \approx 2.5,\ SD \approx 0.5$

(vi) $\bar{x} \approx 4.5,\ SD \approx 0.5$



2. Test de permutaciones I

Ejercicio 1

Un psicólogo quiere comparar dos terapias para pacientes con depresión. Solo cuatro pacientes tomaron parte; dos fueron asignados al azar al grupo que recibió la terapia 1 y los dos restantes al grupo que recibió la terapia 2. La siguiente tabla muestra las puntuaciones obtenidas en un test de mejoría para estos 4 pacientes.

Paciente:	1	2	3	4
Terapia:	1	1	2	2
Puntuación:	30	60	20	30

1. Enumere las seis formas posibles en que estos cuatro pacientes podrían haber sido asignados a las dos terapias.
2. Usando los puntajes de mejoría observados que se muestran en la tabla, calcular las medias en cada grupo y su diferencia, para cada asignación posible.
3. Indicar cuál es la hipótesis nula para este experimento.

4. Argumentar que si la hipótesis nula es cierta, la distribución muestral de la diferencia en las medias de cada grupo es

$$\mathbf{P}(-20) = 2/6,\ \mathbf{P}(-10) = 1/6,\ \mathbf{P}(10) = 1/6,\ \mathbf{P}(20) = 2/6.$$

5. Calcular el p-valor a una cola, comparando con la alternativa de que la puntuación de mejoría es mayor en la terapia 1. Interpretar el resultado.

Ejercicio 2

En un experimento sobre el tratamiento de la adicción a la cocaína, 48 personas fueron asignadas al azar en dos grupos iguales. Uno de los grupos recibió un tratamiento con desipramina (D), un nuevo fármaco en desarrollo, y el otro con litio (L) un medicamento ya existente.

La variable de interés considerada por los investigadores es la recaída de los pacientes. Los resultados del experimento fueron los siguientes: en el grupo tratado con desipramina hubo 10 recaídas, y en el grupo tratado con litio hubo 18 recaídas.

1. Indicar cuál es la variable de respuesta en este experimento.
2. Escribir la hipótesis siguiente en términos de la variable de respuesta

$$H_0 : \begin{array}{l}\text{Los tratamientos no difieren significativamente}\\ \text{en su efecto sobre las recaídas de los pacientes.}\end{array}$$

3. Construir un modelo de urnas y bolillas en el cual se divide al azar a los pacientes en dos grupos, y tal que la cantidad de recaídas totales es siempre 28.
4. Sean R_D y R_L las variables aleatorias que cuentan el número de recaídas en el grupo D y L respectivamente.
   a) ¿Cuáles son los valores de R_D y R_L observados por los investigadores?
   b) ¿Cuánto es $R_D + R_L$? Determinar la distribución de R_D.
5. Calcular el p-valor $\mathrm{pval}((R_D)_{\mathrm{obs}})$ y concluir si la afirmación H_0 es creíble o no.

Ejercicio 3

Se realizó un estudio de pacientes con infecciones que causan diarrea persistente. Los pacientes fueron asignados al azar en dos tratamientos: algunos pacientes recibieron el antibiótico vancomicina y algunos pacientes recibieron un trasplante fecal. La variable de respuesta de cada paciente es si se curó ó no se curó. Los resultados se muestran en la siguiente tabla:

		Tratamiento		
		Vancomicina	Trasplante	Total
Curado	Si	4	13	17
	No	9	3	12
	Total	13	16	29

1. Se considera la región de rechazo $I_r(c) = [c, \infty)$ y el estadístico X igual al número de curados en el grupo que recibió el transplante. Convenimos en utilizar una probabilidad de error de tipo I de $\alpha = 0.05$. Calcular el valor de c.
2. En base al valor de c calculado en la parte anterior, decidir si rechazar o no H_0.

3. Calcular el p-valor $\mathrm{pval}(X_{\mathrm{obs}})$ y decidir si rechazar o no H_0. ¿Es la misma decisión que en la parte anterior?

Ejercicio 4

Un estudio desea evaluar la relación entre la condición física y las horas de las comidas diarias. Para esto 17 personas son divididas al azar en dos grupos, uno de 11 que realiza las comidas diarias en un horario fijo, y otro con 6 que lo hace en horarios cambiantes. Los resultados del estudio se muestran en la siguiente tabla.

		Horario		
		Fijo	Cambiante	Total
Condición física	Mejor	9	2	11
	Igual	2	4	6
	Total	11	6	17

¿Existe evidencia de que horarios fijos para las comidas es beneficioso para la condición física?

1. Formular la hipótesis nula de interés.
2. Sugerir un estadístico para usar en este test.
3. Responder a la pregunta formulada más arriba.

Ejercicio 5

Los árboles frutales están sujetos a una enfermedad causada por bacterias comúnmente llamada tizón de fuego (porque las ramas muertas resultantes parecen haber sido quemadas). Se realiza un estudio para comparar los resultados de eliminar cuidadosamente las ramas afectadas a medida que van apareciendo. La variable de respuesta es la muerte/vida del árbol después de cuatro años. Un grupo de 24 árboles fue dividido al azar en dos grupos, uno de tratamiento (17) y otro de control (7), y los resultados son resumidos en la siguiente tabla:

		Grupo		
		Control	Tratamiento	Total
Respuesta	Muerto	7	12	19
	Vivo	0	5	5
	Total	7	17	24

1. Formular la hipótesis nula de interés.
2. Enumerar todas las tablas posibles que podrían haber ocurrido en la división aleatoria.
3. Calcular la probabilidad de cada tabla.
4. Calcular el p-valor y decidir si rechazar o no la hipótesis nula.

Ejercicio 6

Una empresa desea cambiar su página web. Luego de un proceso de selección se debe elegir entre dos posibles modelos (A y B). Para decidirlo, la empresa elige 40 clientes y los divide al azar en dos grupos iguales. A uno de ellos se les muestra la página con el modelo A y al otro la página con el modelo B. Luego de un mes se les pregunta si están satisfechos con la página. Los resultados son:

		Modelo		
		A	B	Total
Satisfacción	No	2	9	11
	Si	18	11	29
	Total	20	20	24

1. Formular la hipótesis nula de interés.
2. Enumerar todas las tablas posibles que podrían haber ocurrido en la división aleatoria.
3. Elegir un estadístico de prueba para este test.
4. Calcular el p-valor y decidir si rechazar o no la hipótesis nula.

Ejercicio 7

Se desea estudiar el efecto de un nuevo producto en el tiempo de secado de la pintura. Para esto se divide al azar un lote de 16 paneles pintados en dos grupos de 8, unos reciben el tratamiento con el nuevo producto (Tratamiento) y los otros se usan de control. Los tiempos de secado (en horas) se muestran en el diagrama de tallos (espalda con espalda) que está a la derecha.

Tratamiento		Control
00	21	
0000	22	0
0	23	0
0	24	000
	25	00
	26	0

Considere la hipótesis nula

$$H_0 : \text{El tratamiento no tiene efecto sobre el tiempo de secado}$$

y el estadístico $X =$ (suma de respuestas en tratamiento)-(suma de respuestas en control). Se asume que el tratamiento no aumenta el tiempo de secado de la pintura.

1. Usando la Figura 1 (ver última página), calcular el p-valor $\text{pval}(X_{\text{obs}})$. Decidir si rechazar o no H_0 para el valor umbral $p_u = 0.05$.
2. Se considera la región de rechazo $I_r(c) = (-\infty, c]$. Usando la Figura 1, hallar el mayor valor de c para el cual la probabilidad de error de tipo I es menor o igual que $\alpha = 0.05$. Decidir si rechazar o no H_0.

Ejercicio 8

Un ingeniero que trabaja para un fabricante de llantas investiga el efecto de un nuevo compuesto sobre la duración del caucho. Para esto divide al azar un lote de 16 llantas en dos grupos de 8, unas reciben el tratamiento con el nuevo compuesto (Tratamiento) y las otras se usan de control. Los datos, en miles de km, obtenidos son los siguientes:

Tratamiento: 61, 59, 60, 65, 58, 60, 61, 60. Control: 59, 60, 61, 62, 55, 60, 59, 61.

Considere la hipótesis nula

$$H_0 : \text{El tratamiento no tiene efecto sobre la duración del caucho.}$$

y el estadístico $X =$ (suma de respuestas en tratamiento)-(suma de respuestas en control). Se asume que el tratamiento no disminuye la duración del caucho.

a) Usando la Figura 2 (ver página 4), calcular el p-valor $\text{pval}(X_{\text{obs}})$. Decidir si rechazar o no H_0 utilizando como umbral $p_u = 0.05$.

b) Se considera la región de rechazo $I_r(c) = [c, +\infty)$. Usando la Figura 2, hallar el menor valor de c para el cual la probabilidad de error de tipo I es a lo sumo $\alpha = 0.05$. Decidir si rechazar o no H_0.

Ejercicio 9

Se desea estudiar el efecto de un nuevo químico sobre la tensión disruptiva (en kV) de un material. Para esto se divide al azar un lote de 16 muestras del material en dos grupos de 8, unas reciben el tratamiento con el nuevo químico (Tratamiento) y las otras se usan de control. Los resultados se muestran en el diagrama de tallos (espalda con espalda) que está a la derecha.

Tratamiento		Control
00	53	
0000	54	0
0	55	0
0	56	000
	57	00
	58	0

Considere la hipótesis nula

$$H_0 : \text{El tratamiento no tiene efecto sobre la tensión disruptiva del material}$$

y el estadístico $X =$ (suma de respuestas en tratamiento)-(suma de respuestas en control). Se asume que el tratamiento no aumenta la tensión disruptiva del material.

1. Usando la Figura 3, calcular el p-valor $\text{pval}(X_{\text{obs}})$. Decidir si rechazar o no H_0 para el valor umbral $p_u = 0.05$.
2. Se considera la región de rechazo $I_r(c) = (-\infty, c]$. Usando la Figura 3, hallar el mayor valor de c para el cual la probabilidad de error de tipo I es menor o igual que $\alpha = 0.05$. Decidir si rechazar o no H_0.

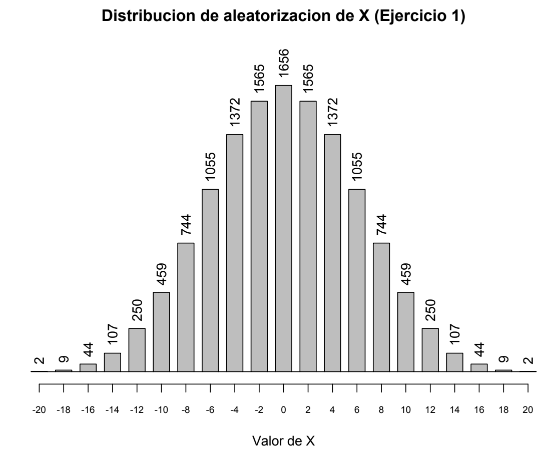


Figura 1: Distribución de aleatorización de X para usar en el Ejercicio 7.

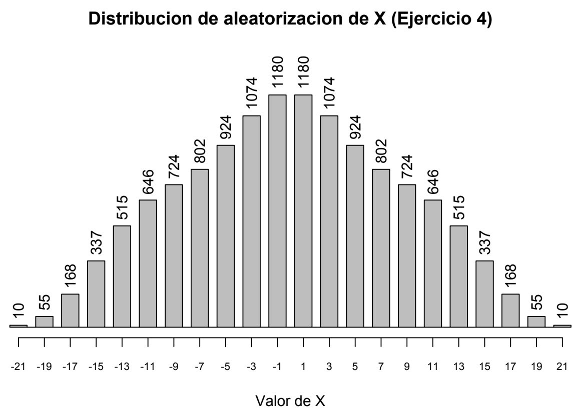


Figura 2: Distribución de aleatorización de X para usar en el Ejercicio 8.

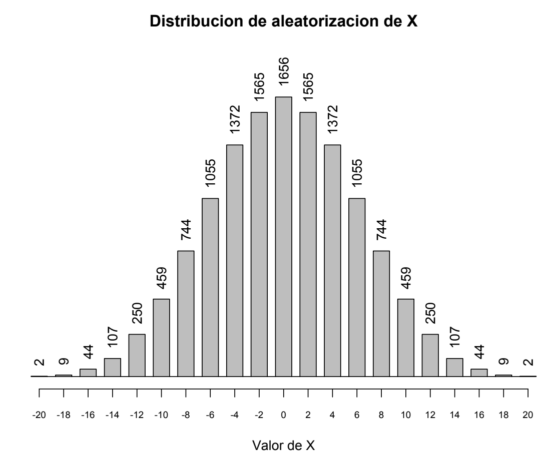


Figura 3: Distribución de aleatorización de X para usar en el Ejercicio 9.

3. Test de permutaciones II

Ejercicio 1

Un psicólogo clínico quiere elegir entre dos terapias para tratar casos graves de depresión. Selecciona seis pacientes que divide al azar en dos grupos, uno que recibe la Terapia 1 y el otro recibe la Terapia 2. Después de un mes de tratamiento, la mejoría en cada paciente se mide por el cambio en una puntuación en un test para medir la gravedad de la depresión. Cuanto mayor sea la puntuación, mejor. Las puntuaciones de mejoría son

Terapia 1:	30	45	45
Terapia 2:	10	20	30

1. Determinar si hay una diferencia entre los efectos de las dos terapias. Discutir si se debe calcular un p-valor a una o a dos colas.

2. Estimar la diferencia de efecto entre las dos terapias. ¿En qué afecta a la estimación elegir los p-valores a una o a dos colas?

Ejercicio 2

Se están analizando dos catalizadores para determinar cómo afectan el rendimiento de un proceso químico. Concretamente, el catalizador A es el que se usa actualmente, pero el catalizador B es más barato, y puede adoptarse en caso de producir un rendimiento similar al actual. Se efectúa una prueba en la planta piloto y da como resultado los datos que se muestran en la siguiente tabla:

	Rendimiento (%)		
Catalizador A:	92	94	92
Catalizador B:	89	91	90

1. Elegir un estadístico de prueba para este test.

2. Decidir si se debe calcular un p-valor a una o a dos colas.

3. Estimar la diferencia de rendimientos entre los dos catalizadores.

Ejercicio 3

La distancia total recorrida por una pelota de golf se prueba golpeando la pelota con Iron Byron, un golfista mecánico con un swing que se dice que emula al legendario campeón Byron Nelson. Se prueban 3 bolas seleccionadas al azar de dos marcas diferentes y se mide la distancia total. Los datos son los siguientes:

Marca 1:	275	285	280
Marca 2:	260	255	260

1. Probar la hipótesis de que ambas marcas de pelota recorren la misma distancia. Usar $\alpha = 0.1$. ¿Cuál es el p-valor?

2. Estimar con una confianza del 90 % la diferencia de distancia recorrida para las dos marcas de pelotas de golf.

Ejercicio 4

Un artículo en Nature (2003, Vol. 48, p. 1013) describe un experimento en el que los individuos consumen diferentes tipos de chocolate para determinar el efecto de comer chocolate en la salud cardiovascular. En el experimento, 12 personas se dividen al azar en dos grupos iguales. En el primer grupo consumen 100 gramos de chocolate negro por día, y en el segundo 200 gramos de chocolate con leche. Luego de un cierto tiempo, se midió la capacidad antioxidante total de su plasma sanguíneo:

Chocolate negro:	119	123	116	114	120	116
Chocolate con leche:	102	106	100	103	99	101

1. ¿Existe evidencia que respalde la afirmación de que el consumo de chocolate negro produce un nivel más alto de la capacidad antioxidante de la sangre que el consumo de chocolate con leche? Usar la Figura 4 para calcular un p-valor a una cola.
2. Estimar la diferencia en el efecto sobre la capacidad antioxidante de los diferentes chocolates. Usar la Figura 5.

Ejercicio 5

Un polímero se fabrica en un proceso químico. Se desea probar un nuevo catalizador más barato. Para poder usarlo, éste no debe modificar demasiado la viscosidad del polímero. Si la diferencia en la viscosidad es de 10 o menos, el fabricante estaría muy interesado en el nuevo catalizador.

Se hace un estudio con los siguientes resultados:

Catalizador actual:	724	718	776	760	745
Catalizador nuevo:	735	775	729	755	783

En base a este estudio, ¿recomendarías el Catalizador nuevo? Usar la Figura 6.

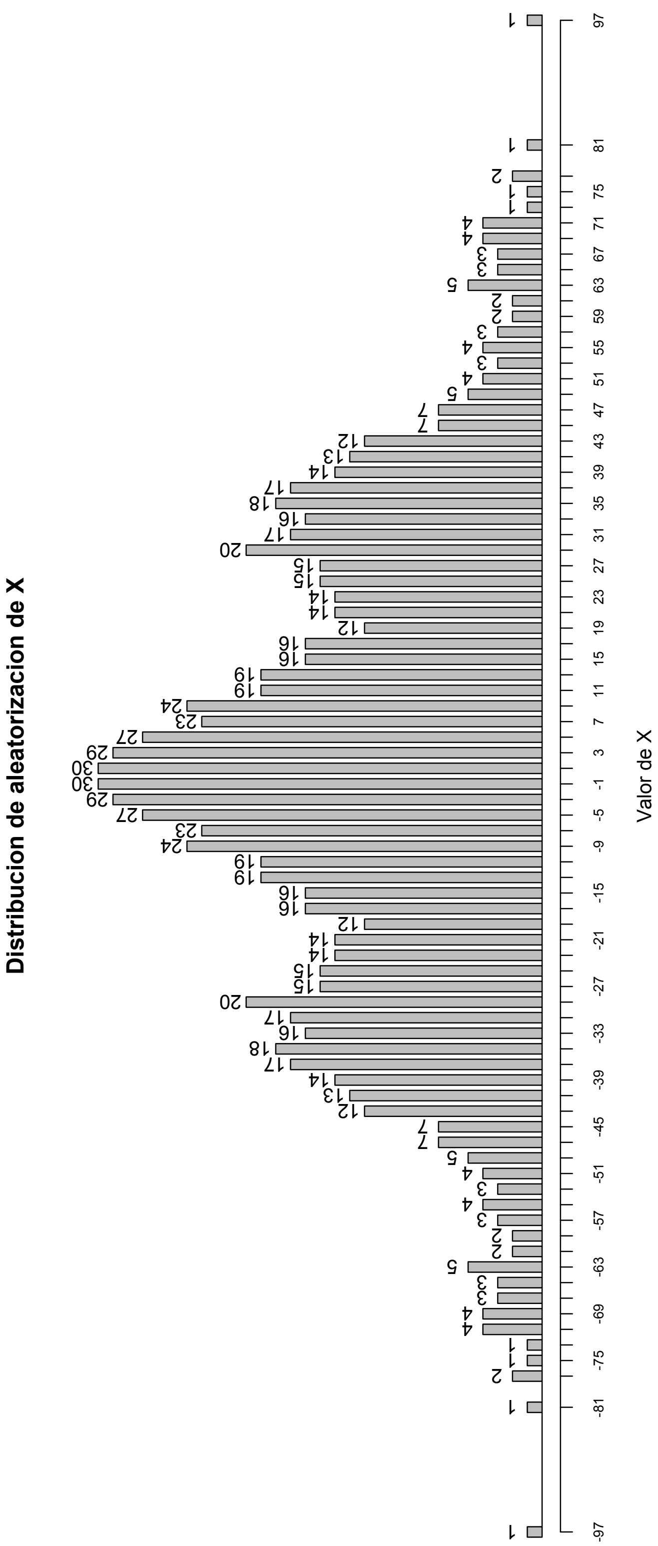


Figura 4: Ejercicio 3. Distribución de aleatorización de X, la diferencia de las sumas en cada grupo. En la gráfica no están indicados, pero el recorrido de X entre -77 y 77 consiste de todos los números impares.

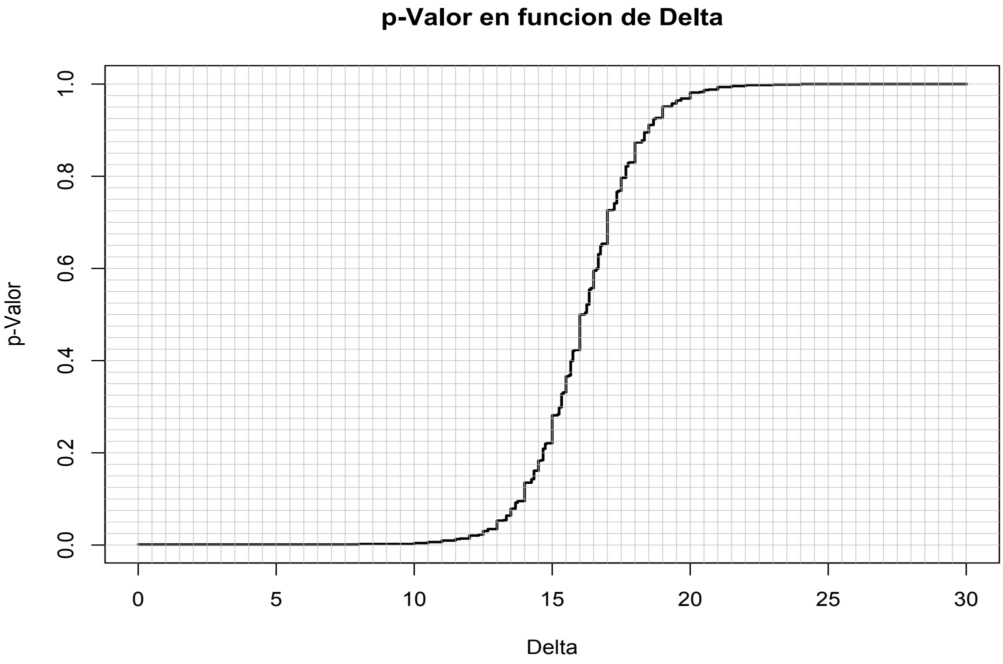


Figura 5: Ejercicio 3. Gráfico del p-valor en función de Delta. Corresponde a restar Δ a las respuestas del grupo de Chocolate negro.

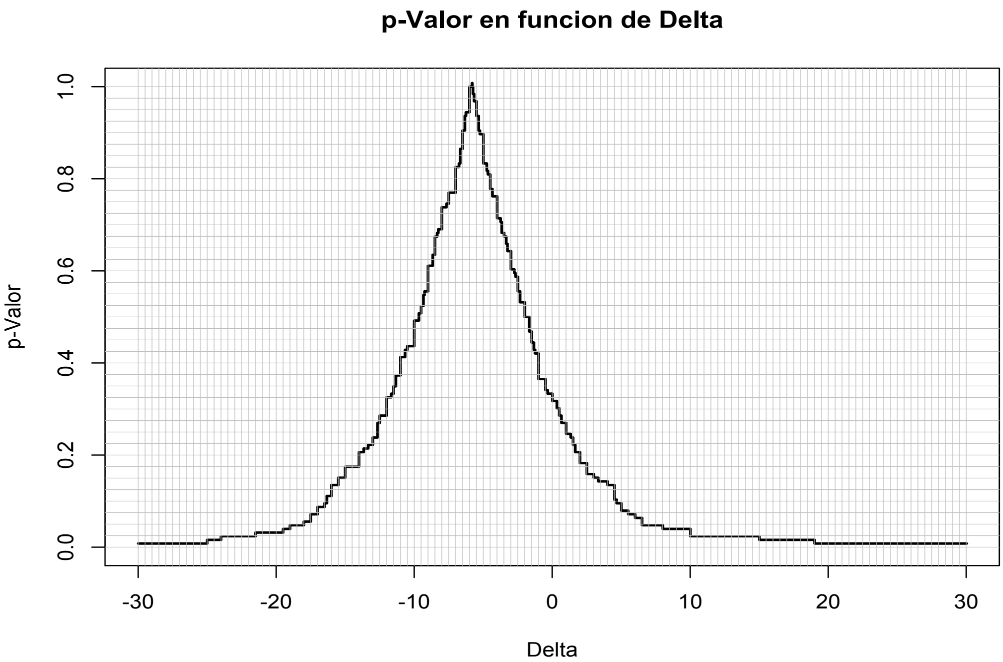


Figura 6: Ejercicio 4. Gráfico del p-valor (a dos colas) en función de Delta. Corresponde a restar Δ a las respuestas del grupo de Catalizador actual.

4. Estimadores I

Ejercicio 1

Sea X una variable discreta con f.p.p. dada por

$$\begin{aligned}
\mathbf{P}(X=0) &= \frac{2}{3}\theta \\
\mathbf{P}(X=1) &= \frac{1}{3}\theta \\
\mathbf{P}(X=2) &= \frac{2}{3}(1-\theta) \\
\mathbf{P}(X=3) &= \frac{1}{3}(1-\theta)
\end{aligned}$$

en donde $0 \leq \theta \leq 1$ es un parámetro. Las siguientes 10 observaciones independientes se tomaron de dicha distribución:

$$3, 0, 2, 1, 3, 2, 1, 0, 2, 1.$$

1. Hallar el estimador de momentos de θ.
2. ¿Cuál es el estimador de máxima verosimilitud de θ?

Ejercicio 2

Supongamos que X es una variable aleatoria discreta con $\mathbf{P}(X=1)=\theta$ y $\mathbf{P}(X=2)=1-\theta$. Se hacen tres observaciones independientes de X:

$$x_1 = 1,\ x_2 = 2,\ x_3 = 2.$$

1. Hallar el estimador de momentos de θ.
2. ¿Cuál es la función de verosimilitud?
3. ¿Cuál es el estimador máxima verosimilitud de θ?

Ejercicio 3

Una empresa ha fabricado ciertos objetos y ha impreso un número de serie en cada objeto fabricado. Los números de serie comienzan en 1 y terminan en N, donde N es el número de objetos que se han fabricado. Uno de estos objetos se selecciona al azar y el número de serie de ese objeto es 888. ¿Cuál es el estimador de momentos de N? ¿Cuál es el estimador de máxima verosimilitud de N?

Ejercicio 4

Beto lanza una moneda tres veces y no observa caras. Luego le da la moneda a Ana. Ella la lanza hasta que se produce la primera cara, esto ocurre con cuatro lanzamientos en total. Sea θ la probabilidad de que la moneda salga cara.

1. ¿Cuál es la función de verosimilitud de θ?
2. ¿Cuál es el estimador de máxima verosimilitud de θ?

Ejercicio 5

Sean $X_1, X_2, \ldots, X_n$ i.i.d. con distribución uniforme en el intervalo de $[0,a]$. Demostrar que el estimador de momento de a es $\hat{a} = 2\overline{X}_n$. ¿Es este un estimador insesgado?

Ejercicio 6

Sea X una variable aleatoria geométrica de parámetro p. Hallar el estimador de máxima verosimilitud de p basado en una muestra aleatoria de tamaño n.

Ejercicio 7

Considere la función de densidad de probabilidad

$$p(x,\theta) = c(1+\theta x), \quad -1 \leq x \leq 1.$$

1. Hallar el valor de la constante c.
2. ¿Cuál es el estimador de momentos de θ?
3. Hallar el estimador de máxima verosimilitud de θ.

Ejercicio 8

Considere la función de densidad de probabilidad

$$p(x;\theta) = \frac{1}{\theta^2} x e^{-x/\theta}, \quad 0 \leq x < \infty, \quad 0 < \theta < \infty$$

Hallar el estimador de máxima verosimilitud de θ.

Ejercicio 9

Una variable aleatoria X tiene función de densidad de probabilidad

$$p(x;\theta) = \frac{1}{2\theta^3} x^2 e^{-x/\theta}, \quad 0 < x < \infty, \quad 0 < \theta < \infty.$$

Calcular el estimador de máxima verosimilitud de θ.

Ejercicio 10

Sea X una variable aleatoria con densidad dada por

$$p(x;\theta) = (1/\theta) x^{(1-\theta)\theta}, \quad 0 < x < 1, \quad 0 < \theta < \infty.$$

Probar que

$$\hat{\theta} = -(1/n) \sum_{i=1}^{n} \ln(X_i)$$

es el estimador de máxima verosimilitud de θ.

5. Estimadores II

Ejercicio 1

Ana y Beto desean estimar el número N de bolitas que hay en una caja. Las bolitas están numeradas del 1 al N. Para esto extraen dos bolitas con reposición de la urna, y obtienen la 3 y la 18. Ana propone que hay $21 = 3+18$ bolitas en la urna, y Beto que hay $18 = \text{máx}\{3,18\}$. ¿Con qué estimación te quedarías, y por qué?

Ejercicio 2

Hemos comprado 25 resistencias del proveedor A y 30 resistencias del vendedor B. Sean $X_1, \ldots X_{25}$ los valores de las resistencias que provienen del proveedor A, que se supone normales independientes con una media μ_A y desviación estándar 1.5 ohms. De manera similar, supongamos que $Y_1, \ldots, Y_{30}$ representan los valores de las resistencias que provienen del proveedor B, que se supone normales independiente con una media μ_B y desviación estándar de 2.0 ohms. Se asume que las resistencias de distintos proveedores son independientes.

¿Cuál es el sesgo y el error cuadrático medio de

$$\hat{\mu}_{AB} = \overline{X} - \overline{Y}$$

estimador de la diferencia $\mu_A - \mu_B$?

Ejercicio 3

Sea $X_1, \ldots, X_n$ un muestreo aleatorio de X, una variable aleatoria con la siguiente densidad de probabilidad:

$$f(x) = \frac{1}{2\sigma} e^{-|x|/\sigma}.$$

1. Hallar el estimador de máxima verosimilitud $\hat{\sigma}$ de σ.
2. Calcular la esperanza y la varianza de $\hat{\sigma}$. ¿Qué pasa cuando n tiende a infinito?
3. Probar que $\hat{\sigma}$ es consistente.
4. Se dispone de la siguiente muestra de X:

   0.14, -7.46, 4.33, -1.59, 6.54, 0.69, 0.56, 8.30, 0.15, -4.27.

   Dar una estimación de σ. Explicar la diferencia entre esta estimación y $\hat{\sigma}$.

Ejercicio 4

Una pieza de una máquina se verifica al final de cada hora de producción y se cambia por una nueva en caso de encontrarse rota. El tiempo de vida T (en horas) de la pieza tiene distribución exponencial de parámetro λ. Por lo tanto, el tiempo en horas que transcurre hasta el recambio de la pieza es $X = \lfloor T \rfloor + 1$, donde $\lfloor T \rfloor$ es la parte entera de T.

1. Hallar la función de probabilidad de X. ¿Es una distribución conocida?
2. Considerar el parámetro $\mu = \mathrm{E}(X)$. Sea $X_1, X_2, \ldots, X_n$ un muestreo aleatorio de los tiempos de recambio de las piezas. Sugerir un estimador $\hat{\mu}$ insesgado, consistente y asintóticamente normal para μ (justificar que cumple las condiciones).

A partir de los tiempos en los que se realiza el recambio de las piezas se desea estimar el parámetro λ del tiempo de vida de dichas piezas.

3. Calcular λ en función de μ. Sugerir un estimador consistente de λ.
4. Se observan 10 tiempos de recambio de piezas y se obtienen lo siguientes datos:

10	21	34	41	50	59	70	77	87	98

   Dar una estimación del valor de λ.

Ejercicio 5

Supongamos que $X_1, X_2, \ldots, X_n$ son i.i.d. con función de densidad

$$p(x;\theta) = e^{-(x-\theta)}, \quad x \geq \theta$$

y $p(x;\theta) = 0$ de lo contrario.

1. Hallar el estimador de momentos $\hat{\theta}_1$ de θ.
2. Calcular el sesgo y el error cuadrático medio de $\hat{\theta}_1$.
3. Hallar el estimador de máxima verosimilitud $\hat{\theta}_2$ de θ.

   Sugerencia: tener cuidado y no derivar antes de pensar. ¿Para qué valores de θ es la verosimilitud positiva?
4. Calcular el sesgo y el error cuadrático medio de $\hat{\theta}_2$.
5. Si el tamaño de la muestra es grande, ¿con qué estimador te quedarías?

Ejercicio 6

Considere el siguiente método para estimar μ para una distribución de Poisson. Observar que

$$p_0 = \mathbf{P}(X = 0) = e^{-\mu}.$$

Sea Y la variable que indica el número de ceros de un muestreo i.i.d. de tamaño n. Entonces μ podría ser estimado por

$$\hat{\mu} = -\ln\left(\frac{Y}{n}\right).$$

1. ¿Es $\hat{\mu}$ un estimador consistente de μ? Tener en cuenta que $Y \sim \mathrm{Bin}(n, p_0)$.
2. Usar el método delta para calcular la varianza asintótica de $\hat{\mu}$.

Ejercicio 7

Las semillas de una planta pueden ser de cuatro tipos:

Tipo	Recuento
A	1997
B	906
C	904
D	32

Según la teoría genética, las probabilidades de cada tipo son

$$p_A = \frac{2+\theta}{4}, \quad p_B = \frac{1-\theta}{4}, \quad p_C = \frac{1-\theta}{4}, \quad p_D = \frac{\theta}{4}$$

donde θ $(0 < \theta < 1)$ es un parámetro relacionado a la vinculación de los tipos.

1. Hallar el estimador de máxima verosimilitud de θ.
2. El número esperado de recuentos de tipo A es $n(2+\theta)/4$. Si este número esperado se iguala al recuento X_A, se obtiene el siguiente estimador de θ:

   $$\hat{\theta}_1 = \frac{4X_A}{n} - 2.$$

   El mismo procedimiento aplicado al tipo D produce el estimador

   $$\hat{\theta}_2 = \frac{4X_D}{n}.$$

   Calcular estas estimaciones. Usando que X_A y X_D son variables aleatorias binomiales, mostrar que los estimadores son insesgados y hallar sus varianzas. Calcular la eficiencia relativa de $\hat{\theta}_1$ y $\hat{\theta}_2$.

Ejercicio 8

Supongamos que en la población de gemelos, los hombres (H) y las mujeres (M) tienen la misma probabilidad de ocurrir y que la probabilidad de que los gemelos sean idénticos es α. Si los gemelos no son idénticos, sus genes son independientes.

1. Probar que
   $$\mathbf{P}(HH) = \mathbf{P}(MM) = \frac{1+\alpha}{4} \quad \mathbf{P}(MH) = \frac{1-\alpha}{2}.$$
2. Supongamos que se muestrean n gemelos. Se encontró que n_1 son HH, n_2 son MM y n_3 son MH, pero no se sabe qué gemelos son idénticos.

   Hallar el estimador de máxima verosimilitud de α y su varianza.

6. Normal, Student, Chi2

Ejercicio 1

El verdadero peso de un objeto es 10g. Se lo pesa dos veces de forma independiente. Llamemos X a la primer medición e Y a la segunda. Asumimos que X e Y son normales con esperanza 10g y desvío 0.2g.

1. Calcular la probabilidad de que la 2da medición esté más cerca de 10g que la primer medición.
2. Calcular la probabilidad de que la 2da medición sea más chica que la primera, pero no en más de 0.2g.

Ejercicio 2

Sean X e Y normales independientes, X estándar e Y con media 1. Supongamos además que $\mathbf{P}(X > Y) = 1/3$. Calcular el desvío estándar de Y.

Ejercicio 3

Sea $\overline{X}$ el promedio de una muestra de 16 variables aleatorias normales independientes con media 0 y varianza 1. Determinar c tal que $\mathbf{P}\left(|\overline{X}| < c\right) = 0.5$.

Ejercicio 4

Si T tiene distribución t de Student con 7 grados de libertad, hallar t_0 tal que

1. $\mathbf{P}(|T| < t_0) = .9$.
2. $\mathbf{P}(T > t_0) = .05$.

Ejercicio 5

Demostrar que la distribución de Cauchy (con $\mu = 0$ y $\sigma = 1$) y la distribución t con 1 grado de libertad son las mismas.

Ejercicio 6

Sean X e Y independientes con distribución exponencial de parámetro $\lambda = 1$. Hallar la distribución de X/Y. Esta distribución se conoce como *distribución F de Fisher*.

Ejercicio 7

Sean $X = (X_1, X_2)$ e $Y = (Y_1, Y_2)$ puntos aleatorios del plano, en donde X_1, X_2, Y_1, Y_2 son normales estándar independientes. Usando la simetría rotacional hallar la distribución de Z la proyección de Y sobre la dirección de X.

Ejercicio 8

Sean $X_1, \dots, X_n$ normales estándar independientes. Sea

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

la varianza muestral. Hallar la varianza de S^2.

7. Test de hipótesis I

Ejercicio 1

Se extraen datos de una distribución binomial $\mathrm{Bin}(5, \theta)$, donde θ es desconocido. A continuación se muestra la tabla de probabilidades $p(x|\theta)$ para 3 valores de θ:

x	0	1	2	3	4	5
$\theta = .5$	.031	.156	.313	.313	.156	.031
$\theta = .6$	.010	.077	.230	.346	.259	.078
$\theta = .8$	.000	.006	.051	.205	.410	.328

Se desea hacer el siguiente test de hipótesis sobre el valor de θ:

$$\begin{cases} H_0 : \theta = 0.5 \\ H_A : \theta > 0.5 \end{cases}$$

al nivel de significancia $\alpha = .1$.

1. Hallar la región de rechazo.
2. Calcular la potencia del test para cada una de las dos hipótesis $\theta = 0.6$ y $\theta = 0.8$.
3. Suponga que hace el experimento y los datos dan $x = 4$. Calcular el p-valor.

Ejercicio 2

Para saber si una moneda está equilibrada se decide tirarla 30 veces y observar la ocurrencia de caras. Al realizarse la experiencia se obtuvieron 11 caras. Sea $p \in (0,1)$ la probabilidad de que salga cara en una tirada.

1. Se plantea el siguiente test de hipótesis:

$$\begin{cases} H_0 : p = 0.5 \\ H_1 : p < 0.5 \end{cases}$$

   Usando la Figura 7 calcular el p-valor. ¿Cuál sería la decisión para $\alpha = 0.05$ y para $\alpha = 0.15$?

2. En la Figura 8 se grafica en función de p_0 (para $n = 30$ y $\alpha = 0.1$) la región de rechazo para el test de hipótesis

$$\begin{cases} H_0 : p = p_0 \\ H_1 : p \neq p_0 \end{cases}$$

   Indicar la región de rechazo para $p_0 = 0.45$ ¿Qué decisión tomarías en este caso?

3. Usando la Figura 8 determinar un intervalo de confianza al nivel 90% para p.

Ejercicio 3

En un experimento sobre la precisión de las pruebas de mentira de polígrafo, 140 personas recibieron instrucciones de decir la verdad y 140 personas recibieron instrucciones de mentir. Los evaluadores usan un polígrafo para adivinar si cada persona está mintiendo o no. Por analogía, digamos que H_0 corresponde a la persona dice la verdad y H_A corresponde a la persona miente.

1. Describa el significado de los errores de tipo I y de tipo II en este contexto y estime sus probabilidades con base en la siguiente tabla:

	La persona es veraz	La persona miente
El polígrafo cree que la persona es veraz	131	15
El polígrafo cree que la persona está mintiendo	9	125

2. En un TdH, ¿qué relaciones existen entre los términos nivel de significación, potencia, error de tipo 1 y error de tipo 2?

Ejercicio 4

Una moneda se lanza independientemente 10 veces para probar la hipótesis de que la probabilidad de cara es $1/2$, frente a la alternativa de que la probabilidad no es $1/2$. La hipótesis nula se rechaza si se observan 0 o 10 caras.

1. ¿Cuál es el nivel de significación del test?
2. Si de hecho la probabilidad de caras es .1, ¿cuál es la potencia del test?

Ejercicio 5

Verdadero o falso, y explicar por qué:

1. El nivel de significación de un TdH es igual a la probabilidad de que la hipótesis nula sea cierta.
2. Si el nivel de significación de un TdH disminuye, se espera que aumente la potencia.
3. Si se rechaza una prueba en el nivel de significancia α, la probabilidad de que la hipótesis nula sea verdadera es igual a α.
4. La probabilidad de que la hipótesis nula sea rechazada falsamente es igual a la potencia del test.
5. Un error de tipo I ocurre cuando el estadístico cae en la región de rechazo del test.
6. Un error de tipo II es más serio que un error de tipo I.
7. La potencia de un TdH está determinada por la distribución nula del estadístico.

Ejercicio 6

Sean $X_1, \ldots, X_{25}$ una muestra de una distribución normal que tiene varianza 100. Hallar la región de rechazo para un TdH en el nivel $\alpha = .1$ de $H_0 : \mu = 0$ versus $H_A : \mu = 1.5$. ¿Cuál es la potencia del test? Repetir para $\alpha = .01$.

Ejercicio 7

Supongamos que se toma una única observación X de una densidad uniforme en $[0, \theta]$, y considere probar $H_0 : \theta = 1$ versus $H_A : \theta = 2$.

1. Construir un test que tenga un nivel de significancia $\alpha = 0$. ¿Cuál es su potencia?
2. Para $0 < \alpha < 1$, considere el test que rechaza cuando $X \in [0, \alpha]$. ¿Cuál es su nivel de significación y potencia?
3. ¿Cuál es el nivel de significación y la potencia del test que rechaza H_0 cuando $X \in [1-\alpha, 1]$?
4. Encontrar otro test que tenga el mismo nivel de significación y la misma potencia que el anterior.
5. ¿Qué sucede si se intercambian las hipótesis nula y alternativa? $H_0 : \theta = 2$ versus $H_A : \theta = 1$?

Ejercicio 8

Se realizaron ocho pruebas para medir la velocidad de un vehículo submarino, en un modelo a escala de laboratorio. La velocidad promedio observada fue $\overline{X} = 102.2$ metros por segundo. Suponga que la velocidad se distribuye normalmente con una desviación estándar conocida $\sigma = 4$ metros por segundo.

1. Pruebe la hipótesis $H_0 : \mu = 100$ versus $H_A : \mu < 100$ usando $\alpha = 0.05$.
2. ¿Cuál es el p-valor en la parte 1?

3. Calcule la potencia del test si la velocidad media real es 95 metros por segundo.

4. ¿Qué tamaño de muestra se requeriría para detectar una velocidad media real de 95 metros por segundo, si quisiera que la potencia de la prueba fuera al menos de 0.85?

Ejercicio 9

Un fabricante asegura que sus baterías debe recargarse aproximadamente cada 4 horas. Se selecciona una muestra aleatoria de 50 de baterías y se las somete a prueba. La vida media de estas baterías es de 4.05 horas. Suponga que la vida útil de la batería se distribuye normalmente con desviación estándar $\sigma = 0.2$ horas.

1. ¿Hay pruebas que respalden la afirmación de que la vida útil de la batería supera las 4 horas? Utilice $\alpha = 0.05$.

2. ¿Cuál es el p-valor en la parte 1?

3. Calcule la potencia del test si la verdadera duración media de la batería es de 4.5 horas.

4. ¿Qué tamaño de muestra se requeriría para detectar una verdadera duración media de la batería de 4.5 horas, si quisiera que la potencia de la prueba fuera al menos 0.9?

Ejercicio 10

La cepa bacteriana Acinetobacter ha sido probada por sus propiedades de adhesión. Una muestra de cinco mediciones dio lecturas de 2.69, 5.76, 2.67, 1.62 y 4.12 dinas-cm^2. Supongamos que se sabe que la desviación estándar es de 0.66 dinas-cm^2 y que los científicos están interesados en una alta adhesión (al menos 2.5 dinas-cm^2).

1. ¿Debería la hipótesis alternativa ser a una o a dos colas?

2. Probar la hipótesis de que la adhesión media es 2.5 dinas-cm^2.

3. ¿Cuál es el p-valor?

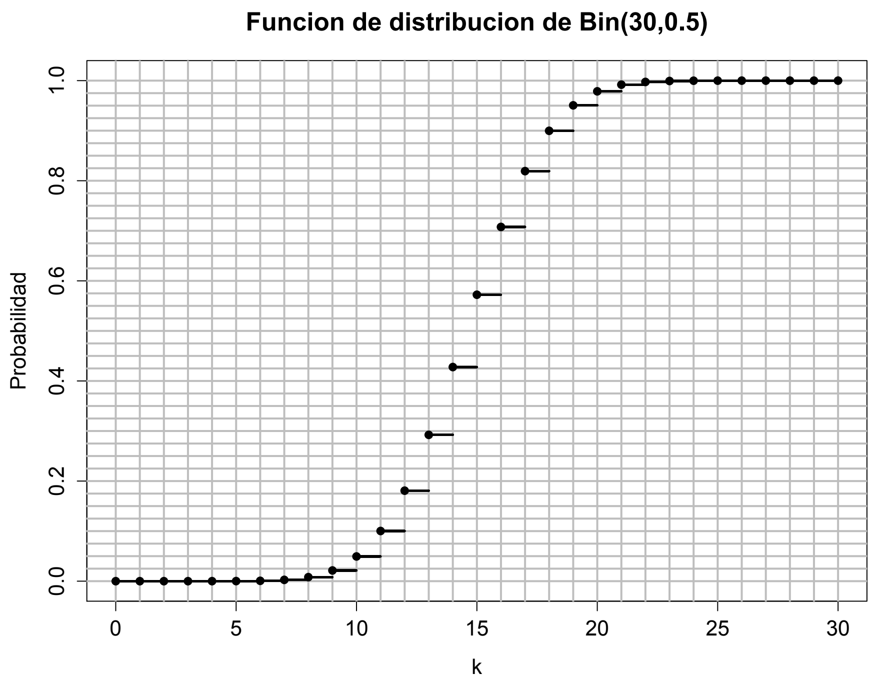


Figura 7: Gráfica de la función de distribución de $\text{Bin}(30, 0.5)$. Las líneas verticales pasan por los enteros, y la distancia entre las líneas horizontales es de 0.025.

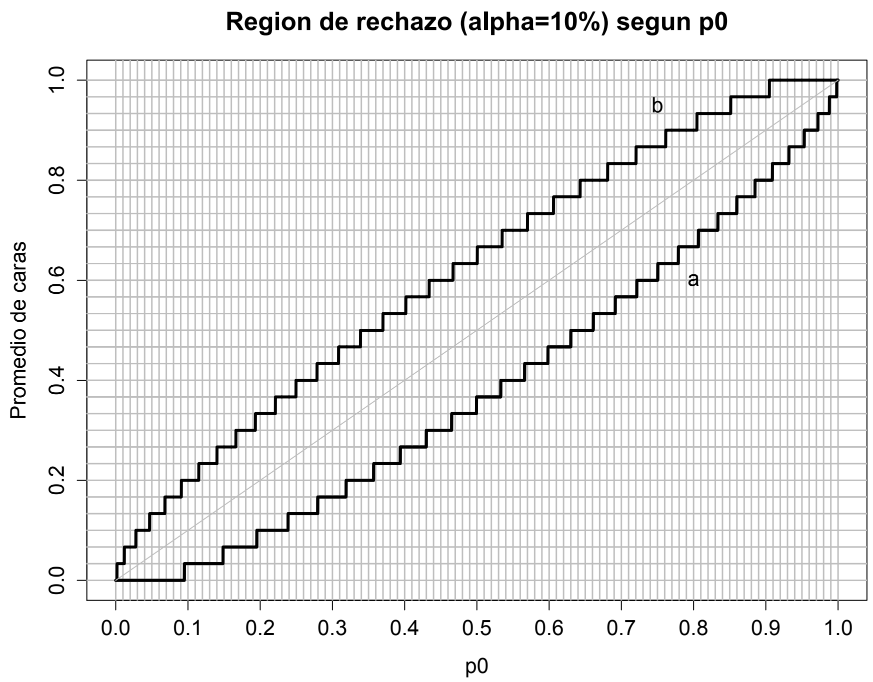


Figura 8: Gráfica de la región de rechazo $\mathscr{R}_\alpha = \left\{ \overline{X}_n \notin [a, b] \right\}$ en función de p_0. Las líneas finas verticales distan 0.01, y las horizontales pasan por los posibles valores del promedio de caras, es decir $\{\frac{k}{30} : k \in \{0, \ldots, 30\}\}$.

8. Test de hipótesis II

Ejercicio 1

Supongamos que de los 1000 clientes encuestados, 850 están satisfechos o muy satisfechos con los productos y servicios de una empresa. Sea p la proporción de clientes satisfechos. Probar la hipótesis $H_0 : p = 0.9$ contra $H_A : p \neq 0.9$ al nivel $\alpha = 0.05$. Calcular el p-valor.

Ejercicio 2

Se pregunta a una muestra aleatoria de 500 votantes registrados en una ciudad si están a favor del uso de combustibles oxigenados para reducir la contaminación del aire. Si más de 315 votantes responden positivamente, concluiremos que al menos el 60 % de los votantes favorecen el uso de estos combustibles.

1. Calcular la probabilidad de error de tipo I, si exactamente el 60 % de los votantes favorece el uso de estos combustibles.
2. ¿Cuál es la probabilidad β de error de tipo II si el 75 % de los votantes favorece esta acción?

Ejercicio 3

Cuando se preguntó a una muestra aleatoria de 583 trabajadores de una ciudad cuántas horas trabajaron en la semana anterior, la media fue de 37.0 horas, con una desviación estándar de 15.1 horas. ¿Sugiere esto que la media laboral de la población para las mujeres es significativamente diferente a las 40 horas?

Responder de la siguiente manera:

1. Identificación la variable relevante y el parámetro.
2. Planteando las hipótesis nula y alternativa.
3. Calculando e interpretando el p-valor para el valor del estadístico observado.
4. Explicando cómo tomar una decisión al nivel de significancia de 0.01.

Ejercicio 4

Supongamos que tenemos 49 datos con una media muestral de 6.25 y una varianza muestral de 36. Queremos probar las siguientes hipótesis:

$$\begin{cases} H_0: \text{ los datos se extraen de } N(4, \sigma^2), \text{ donde } \sigma \text{ es desconocido.} \\ H_A: \text{ los datos se extraen de } N(\mu, \sigma^2) \text{ donde } \mu \neq 4. \end{cases}$$

1. Hacer el test al nivel de significación $\alpha = 0.05$. Usar la tabla t para calcular el p-valor.
2. Hacer un dibujo que muestre la distribución nula, la región de rechazo y el área utilizada para calcular el p-valor de la parte 1.

Ejercicio 5

Una compañía de agua mineral afirma que sus botellas de agua contienen a lo sumo 1 ppm (partes por millón) de benceno. Una muestra de control proporciona los siguientes datos:

$$n = 25, \quad \overline{X} = 1.16 \text{ ppm}, \quad S = 0.20 \text{ ppm}.$$

Se asume que los datos tienen distribución normal. Hacer un test que ponga a prueba la afirmación, al nivel $\alpha = 0.05$.

Proceder de la siguiente manera:

1. ¿Cuáles son las hipótesis nula y alternativa?
2. Hacer un dibujo de la distribución del estadístico bajo H_0. Incluir valores críticos y la región de rechazo.
3. ¿Cuál es el valor observado del estadístico de prueba?
4. ¿Cuál es tu conclusión?

Ejercicio 6

Sea $X_1, \dots, X_n$ una muestra aleatoria de una distribución exponencial con densidad

$$p(x;\theta) = \theta e^{-\theta x}, \quad (x > 0,\ \theta > 0).$$

Usando el estadístico de la razón de verosimilitud, probar que el test

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_A : \theta \neq \theta_0 \end{cases}$$

la región de rechazo tiene la forma $\{\overline{X}\exp(-\theta_0 \overline{X}) \leq c\}$.

Ejercicio 7

Los siguientes datos provienen de una distribución normal que se sospecha tiene varianza igual a 1. Se desea probar esto con la alternativa de que la varianza es mayor que 1.

1.76, -2.28, -0.56, 1.46, 0.59, 1.26,

-1.94, -0.79, -0.86, -1.41, 2.07, 1.30.

Hay un test de chi-cuadrado para esto. Hacer el test con nivel de significancia 0.05.

Ejercicio 8

La siguiente tabla muestra el número de muertes por caídas accidentales de cada mes del año 1970 en Estados Unidos. ¿Existe alguna evidencia de una desviación de la uniformidad en la tasa a lo largo del tiempo? Es decir, ¿existe un patrón estacional para esta tasa de mortalidad? Si es así, especula sobre las causas.

Mes	Número de muertes
Enero	1668
Febrero	1407
Marzo	1370
Abril	1309
Mayo	1341
Junio	1338
Julio	1406
Agosto	1446
Septiembre	1332
Octubre	1363
Noviembre	1410
Diciembre	1526

Ejercicio 9

En una muestra de 200 individuos, los investigadores están interesados en las proporciones de hombres y mujeres que usan una marca particular de shampoo. ¿Hay alguna diferencia entre hombres y mujeres con respecto a su uso del producto? Hacer un test al nivel $\alpha = .05$.

	Hombres	Mujeres	Total
Usa	20	50	70
No usa	80	50	130
Total	100	100	200

Ejercicio 10

A un importante fabricante de alimentos le preocupa que las ventas de sus papas fritas rústicas hayan disminuido. Como parte de un estudio de factibilidad, la compañía investiga los tipos de papas fritas que se venden en todo el país para determinar si el tipo de papas vendidas es independiente del área del país. Los resultados del estudio están abajo. Realizar una prueba de independencia.

Tipo de papas fritas	Norte	Sur	Centro	Este
Rústicas	70	50	20	25
Comunes	100	60	15	30
Redondas	20	40	10	10

9. Intervalos de confianza

Ejercicio 1

Se sabe que el volumen en un conjunto de botellas de vino sigue una distribución normal $N(\mu, 25)$. Tomas una muestra de las botellas y mides sus volúmenes. ¿Cuántas botellas tienes que muestrear para tener un intervalo de confianza al 95 % para μ con ancho 1?

Ejercicio 2

Supongamos que contra un cierto oponente, la cantidad de puntos que el equipo de basketball de la FIng obtiene se distribuye normalmente con una media desconocida θ y una varianza desconocida σ^2. Supongamos que en el transcurso de los últimos 10 juegos, el equipo obtuvo los siguientes puntos:

$$59, 62, 59, 74, 70, 61, 62, 66, 62, 75.$$

1. Calcular un t-intervalo de confianza al 95 % para θ.
2. Explicar con palabras qué significa la confianza de 95 % del intervalo que acabas de calcular.
3. Ahora supongamos que se sabe que $\sigma^2 = 25$. Calcular un z-intervalo de confianza al 95 % para θ. ¿Cómo se compara este con el intervalo en 1?

Ejercicio 3

Usted hace una encuesta para ver qué fracción p de la población apoya al candidato A sobre el candidato B. Responder usando la aproximación normal:

1. ¿Cuántas personas necesita encuestar para conocer p con un error del 1 % y con un 95 % de confianza?
2. Si encuesta a 400 personas, ¿cuántos tienen que preferir al candidato A para que el intervalo de confianza del 90 % esté completamente por encima de $p = 0.5$?

Ejercicio 4

Supongamos que hiciste 40 intervalos de confianza con un nivel de confianza del 95 %. ¿Cuántos de ellos esperarías que estén "equivocados"? Es decir, ¿cuántos no contendrían realmente el parámetro estimado? ¿Deberías sorprenderte si 10 de ellos están equivocados?

Ejercicio 5

Un estadístico elige 27 fechas seleccionadas al azar, y al examinar los registros de ocupación de un motel en particular para esas fechas, encuentra una desviación estándar de 5.86 habitaciones alquiladas. Si la cantidad de habitaciones alquiladas tiene distribución normal, hallar un intervalo de confianza del 95 % para la desviación estándar de la cantidad de habitaciones alquiladas.

Ejercicio 6

Datos se recopilan sobre el tiempo entre llegadas de taxis consecutivos a un hotel del centro de la ciudad. Recopilamos un conjunto de datos de tamaño 45 con una media muestral $\overline{X} = 5.0$ y una desviación estándar muestral $S = 4.0$. Suponemos que los datos siguen una distribución normal.

1. Calcular un intervalo de confianza del 80 % para la media μ.
2. Calcular un intervalo de confianza del 80 % para la varianza σ^2.

Ejercicio 7

Un investigador está interesado en medir el puntaje promedio de satisfacción laboral (0 = más bajo y 100 = más alto) entre los docentes de tiempo completo en una universidad. Muestrea aleatoriamente a 25 profesores y encuentra que la satisfacción laboral promedio es de 48.4 con una desviación estándar de 6.1. Supongamos que la satisfacción laboral sigue una distribución normal. Usando los datos anteriores, construir un IdC del 95 % para la satisfacción laboral en esta universidad.

Ejercicio 8

Un político afirma que el 70 % de las personas en su distrito son de su partido. Un investigador toma una muestra de 100 personas y encuentra que solo 50 son de ese partido. ¿Es el político un mentiroso? Hacer un IdC al 95 %.

Ejercicio 9

Se desea estimar la media de una población normal. Supongamos que $\sigma = 20$ basado en estudios previos. Nos gustaría estimar la media de la población dentro de ± 10 de su valor verdadero, con $\alpha = .05$ (es decir, 95 % de confianza). ¿Qué tamaño de muestra debemos tomar?

Ejercicio 10

Considere el intervalo de confianza para μ de una población normal con desviación estándar conocida σ:

$$\overline{X} - z_{\alpha_1}\sigma/\sqrt{n} \leq \mu \leq \overline{X} + z_{\alpha_2}\sigma/\sqrt{n}$$

donde $\alpha_1 + \alpha_2 = \alpha$. Sea $\alpha = 0.05$ y calcule el intervalo para $\alpha_1 = \alpha_2 = \alpha/2 = 0.025$. Ahora calcule el intervalo para el caso $\alpha_1 = 0.01$ y $\alpha_2 = 0.04$. ¿Qué intervalo es más corto? ¿Hay alguna ventaja en usar un intervalo de confianza simétrico?

10. Comparación de dos muestras

Ejercicio 1

La tabla contiene datos sobre los ciclos hasta la falla de las juntas de soldadura a dos temperaturas diferentes (20 y 60 C) para diferentes tipos de placas de circuito impreso:

(20°C) 218, 265, 279, 282, 336, 469, 496, 507, 685, 685

(60°C) 185, 242, 254, 280, 305, 353, 381, 504, 556, 697

1. Pruebe la hipótesis nula con $\alpha = 0.05$ de que los ciclos de falla son los mismos a ambas temperaturas. ¿La alternativa es a una o dos colas?
2. Calcular un intervalo de confianza del 95 % para la diferencia entre los ciclos medios hasta la falla para las dos temperaturas.
3. ¿Está el valor cero contenido en el intervalo de confianza del 95 %? Explique la conexión con la conclusión a la que llegó en la parte 1.
4. ¿Qué supuestos has hecho sobre la distribución de los datos?

Ejercicio 2

Una empresa está interesada en determinar si existe una diferencia en la satisfacción laboral entre los empleados de la mañana y los de la tarde. La satisfacción laboral se mide en una escala de 0 (nada satisfecho) a 100 (extremadamente satisfecho).

Mañana	91	87	67	99	83	87	90	45	77	81	92	79	83	99	99	75	75
Tarde	74	66	81	39	49	55	67	71	62	50	58	60	73	70			

Haga un test al nivel de significación de .05. Suponga datos normales con varianzas iguales.

Ejercicio 3

El gerente de una flota de automóviles está probando dos marcas de llantas y asigna una llanta de cada marca al azar a las dos ruedas traseras de ocho autos y los hace funcionar hasta que se gastan. Los datos (en kilómetros) se muestran en la tabla. Calcular un intervalo de confianza del 99 % para la diferencia en la vida media de las llantas. ¿Qué marca preferirías en base a este cálculo?

Auto	Marca 1	Marca 2
1	36925	34318
2	45300	42280
3	36240	35500
4	32100	31950
5	37210	38015
6	48360	47800
7	38200	37810
8	33500	33215

Ejercicio 4

Se pueden usar dos pruebas analíticas diferentes para determinar el nivel de impureza en las aleaciones de acero. Se analizan ocho muestras utilizando ambos procedimientos, y los resultados se muestran en la siguiente tabla.

Ejemplar	Prueba 1	Prueba 2
1	1.2	1.4
2	1.3	1.7
3	1.5	1.5
4	1.4	1.3
5	1.7	2.0
6	1.8	2.1
7	1.4	1.7
8	1.3	1.6

1. ¿Hay evidencia suficiente para concluir que las pruebas difieren en el nivel de impureza promedio, usando $\alpha = 0.01$?
2. ¿Hay pruebas que apoyen la afirmación de que la prueba 1 genera una diferencia de medias de 0.1 unidades por debajo de la prueba 2? Utilice $\alpha = 0.05$.
3. Si la media de la prueba 1 está 0.1 por debajo de la media de la prueba 2, es importante detectar esto con probabilidad de al menos 0.90. ¿El uso de ocho aleaciones fue un tamaño de muestra adecuado? De no ser así, ¿cuántas aleaciones deberían haberse usado?

11. Dependencia

Ejercicio 1

Sea X uniforme en $[0,1]$ y A un evento tal que $\mathbf{P}(A|X=x) = x^2$. ¿Cuánto vale $\mathbf{P}(A)$?

Ejercicio 2

Sea (X,Y) un punto al azar en el interior del triángulo de vértices $(1,0)$, $(0,1)$, $(-1,0)$. Para $x \in [-1,1]$, hallar:

a) $\mathbf{P}\left(Y \geq \frac{1}{2} \middle| X = x\right)$; b) $\mathbf{P}\left(Y < \frac{1}{2} \middle| X = x\right)$; c) $\mathrm{E}(Y|X=x)$; d) $\mathrm{Var}(Y|X=x)$.

Ejercicio 3

La distribución de probabilidad condicional de Y dado $\{X = x\}$ es

$$p_{Y|x}(y) = xe^{-xy}, \quad \text{para } y > 0,$$

y la distribución de probabilidad marginal de X es uniforme (continua) de 0 a 10.

1. Graficar $p_{Y|x}(y)$ para varios valores de x.
2. Determinar:

$$\mathbf{P}(Y < 2|X = 2), \quad \mathrm{E}(Y|X = 2), \quad \mathrm{E}(Y|X = x), \quad p_{XY}(x,y), \quad p_Y(y).$$

Ejercicio 4

Las longitudes de los ejes menor y mayor se utilizan para describir las partículas de polvo que tienen una forma aproximadamente elíptica. Sean X e Y las longitudes de los ejes mayor y menor (micrómetros), respectivamente. Supongamos que

$$p_X(x) = e^{-x}, \quad (x > 0)$$

y que la distribución condicional

$$p_{Y|x}(y) = e^{-(y-x)}, \quad (x < y).$$

1. Demostrar que $p_{Y|x}(y)$ es una función de densidad de probabilidad para cualquier valor de x.
2. Calcular $\mathbf{P}(X < Y)$.
3. Hallar la densidad de probabilidad conjunta $p_{XY}(x,y)$.
4. Hallar la densidad de probabilidad condicional de X dado $\{Y = y\}$.
5. Calcular $\mathbf{P}(Y < 2|X = 1)$ y $\mathrm{E}(Y|X = 1)$.
6. Calcular $\mathbf{P}(X < 1, Y < 1)$ y $\mathbf{P}(Y < 2)$.
7. Hallar c tal que $\mathbf{P}(Y < c) = 0.9$.
8. ¿Son X e Y independientes?

12. Normal bi-variada y regresión lineal

Ejercicio 1

Los valores de presión arterial sistólica y diastólica (en mmHg) son las presiones cuando el músculo cardíaco se contrae y relaja (las denotados por Y y X respectivamente). La distribución de la presión diastólica es normal con media 73 y desviación estándar 8. La presión sistólica tiene distribución condicional normal con media $1.6x$ cuando $\{X = x\}$ y desviación estándar de 10.

1. Determinar la densidad condicional de Y dado $\{X = 73\}$.
2. Calcular $\mathbf{P}(Y < 115|X = 73)$.
3. Calcular $\mathrm{E}(Y|X = 73)$.

4. Reconocer la distribución $f_{XY}(x,y)$ e identificar la media y la varianza de Y.

Ejercicio 2

Un estudio indica que la elasticidad (en MPa) y la densidad (en g cm^3) de un tipo de goma se ajustan al modelo normal bi-variado. La siguiente tabla indica la estimación de sus parámetros:

Densidad	media: 1	desvío: 0.2
Elasticidad	media: 100	desvío: 8
correlación: -0.8		

1. Se dispone de un ejemplar de goma cuya densidad es 0.75. ¿En cuánto estimarías su elasticidad?
2. Para el ejemplar de la parte anterior, dar un intervalo centrado en el pronóstico que contenga su elasticidad con probabilidad 0.95.
3. Se dispone de un ejemplar de goma cuya densidad se sabe es mayor que 0.8. Estimar su elasticidad.

Ejercicio 3

En un experimento para investigar la relación entre el alargamiento (Y en%) y la temperatura (X en grados) de la mozzarella se observa que dichas variables se ajustan al modelo normal bi-variado. La siguiente tabla indica la estimación de sus parámetros:

Temperatura	media: 70	desvío: 5
Alargamiento	media: 175	desvío: 20
correlación: 0.75		

a) Hallar y tal que $\mathbf{P}(Y \geq y|X = 60) = 0.05$.

b) Se dispone de un ejemplar de mozzarella cuya temperatura es de 60 grados. Dar un intervalo que contenga su alargamiento con probabilidad 0.9.

Ejercicio 4

La siguiente tabla resume los puntajes de un grupo de estudiantes en dos pruebas correspondientes a asignaturas distintas en el mismo período de exámenes:

1era prueba	media: 65	desvío: 10
2da prueba	media: 75	desvío: 9
correlación: 0.6		

Asumimos que los datos se distribuyen como una normal bi-variada.

1. Un estudiante obtuvo 60 puntos en la primera prueba ¿Qué puntaje pronosticarías para la segunda prueba?
2. ¿Cuál es la probabilidad de que tu pronóstico anterior sea erróneo por más de 5 puntos?
3. De los estudiantes que sacaron más de 60 en la primera prueba, ¿qué porcentaje sacó más que la media en la segunda prueba?

13. Regresión lineal en la práctica

Ejercicio 1

La siguiente tabla muestra los datos de un estudio sobre los efectos de reducir el consumo de corriente por medios electrónicos.

Tensión	Corriente (mA)
0.66	7.32
1.32	12.22
1.98	16.34
2.64	23.66
3.3	28.06
3.96	33.39
4.62	34.12
3.28	39.21
5.94	44.21
6.6	47.48

Se asume que los datos siguen un modelo normal bi-variado.

1. Bosquejar un diagrama de dispersión de los datos y ajustar una línea de regresión para predecir la corriente en función de la tensión.
2. Estimar el coeficiente de correlación.
3. Probar la hipótesis de que $\rho = 0$ contra la alternativa $\rho \neq 0$ al nivel $\alpha = 0.05$. ¿Cuál es el p-valor?

Ejercicio 2

Se estudiaron las propiedades físicas de seis muestras de telas ignífugas. Utilice los datos adjuntos y un nivel de significancia de .05 para determinar si existe una correlación significativa entre la rigidez X (mg-cm) y el espesor Y (mm). ¿Es sorprendente el resultado de la prueba a la luz del valor de r?

X	7.98	24.52	12.47	6.92	24.11	35.71
Y	.28	.65	.32	.27	.81	.57