Guía de estudio para el examen extraordinario de ESTADÍSTICA

Departamento de Matemáticas
Facultad de Química, UNAM

2024 Facultad de Química, UNAM

https://amyd.quimica.unam.mx

La presente guía es un material de apoyo para los estudiantes de la Facultad de Química que presentan el Examen Extraordinario de Estadística (clave: 1400). Esta guía en ningún momento sustituye el contenido del curso de Estadística ni el contenido de ninguno de los textos mencionados en la bibliografía. Los problemas aquí presentados pueden o no estar incluídos en el Examen Extraordinario.

Primera versión, Noviembre 2024

Índice

1. Estadística descriptiva

Ejercicio 1.1 Identifique cuáles son las unidades experimentales en las que se miden las siguientes variables:

a) Género de un profesor.
b) Número de aciertos en un examen.
c) Edad de un paciente con diabetes.
d) Número de flores en un rosal.
e) Color del auto que sale de un estacionamiento.

Solución:

a) Profesores b) Exámenes c) Pacientes d) Rosales e) Autos

■

Ejercicio 1.2 Identifique cada una de las siguientes variables como cualitativa o cuantitativa:

a) Tiempo que toma armar un rompecabezas.
b) Número de estudiantes en un aula.
c) Calificación de un político en una encuesta (excelente, bueno, regular, malo).

Solución:

a) Cuantitativa b) Cuantitativa c) Cualitativa

■

Ejercicio 1.3 Identifique cuales de las siguientes variables cuantitativas son discretas o continuas:

a) Población en un estado particular de México.
b) Peso de envases recogidos para reciclar en un día.
c) Tiempo para resolver un examen de Estadística.
d) Número de consumidores en una encuesta de 500 que consideran útil aplicar leyendas nutrimentales en productos alimenticios.
e) Número de accidentes en un tramo de 20 Km. de la carretera a Toluca.
f) Tiempo para completar un crucigrama.
g) Costo de una sandía.
h) Número de hermanos de un estudiante.
i) Rendimiento en kilogramos de manzanas en una huerta de una hectárea.

Solución:

a) discreta	b) continua	c) continua	d) discreta	e) discreta
f) continua	g) continua	h) discreta	i) continua	

■

Ejercicio 1.4 Cincuenta estudiantes se inscribieron en cuatro carreras (QFB, Q, IQ, IQM) como se muestra en la tabla:

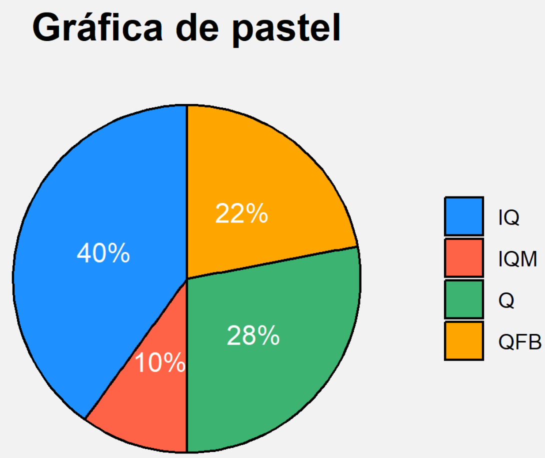
Carrera	QFB	Q	IQ	IQM
Frecuencia	11	14	20	5

a) ¿Cuál es la unidad experimental?
b) ¿Qué variable se mide? ¿Es cualitativa o cuantitativa?
c) Dibuje una gráfica de pastel para describir estos datos.
d) Dibuje una gráfica de barras para describir estos datos.
e) ¿Es importante el orden de presentación en la gráfica anterior?
f) ¿Qué proporción de los estudiantes está en la categoría Q, IQ o IQM?
g) ¿Qué porcentaje de los estudiantes no está en la categoría Q?

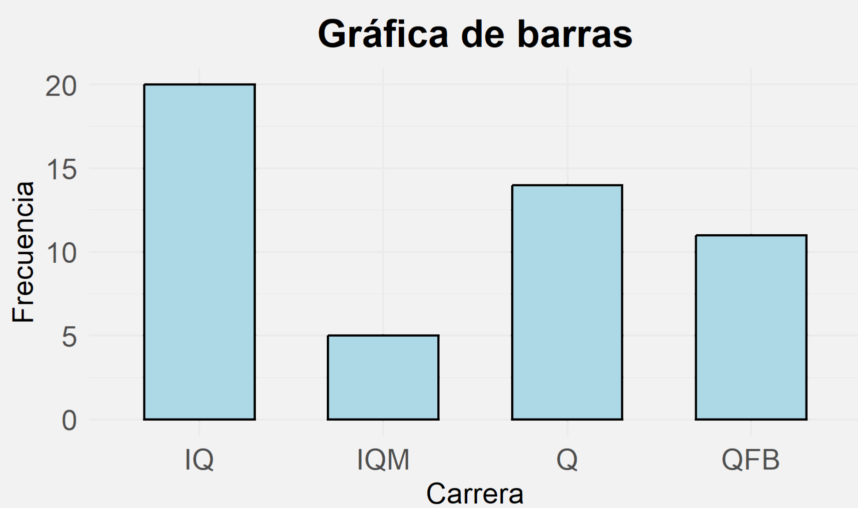
Solución:

a) Las unidades experimentales son los Estudiantes.
b) La variable es Carrera y es de tipo Cualitativa.

c)



d)



e) En este caso no es importante ya que las categorias no tienen un orden natural.
f) Proporción de los estudiantes que está en la categoría Q, IQ o IQM $= 39/50 = 0.78$
g) Porcentaje de los estudiantes que no está en la categoría Q $= 36/50 = 0.72 = 72\,\%$

■

Ejercicio 1.5 Un investigador registró el tiempo (meses) entre el inicio de una enfermedad y su recurrencia para $n = 50$ pacientes:

2.1	4.4	14.7	9.6	4.1	18.4	14.1	1.0	1.6	3.5
2.7	32.3	16.7	7.4	0.2	6.1	2.4	2.4	11.4	18.0
9.9	9.0	2.0	8.2	19.2	6.9	13.5	7.4	0.2	18.0
8.7	24.0	26.7	3.7	12.6	6.6	3.9	1.6	4.3	3.3
1.2	8.3	0.3	1.3	1.4	8.2	5.8	23.1	5.6	0.4

a) Construya un histograma de frecuencia relativa para estos datos y una gráfica de frecuencia relativa acumulada u ojiva.
b) ¿La forma de la distribución es aproximadamente simétrica, sesgada a la izquierda o sesgada a la derecha?
c) Encuentre cuál es la fracción de tiempos de recurrencia menores o iguales a 10 meses.

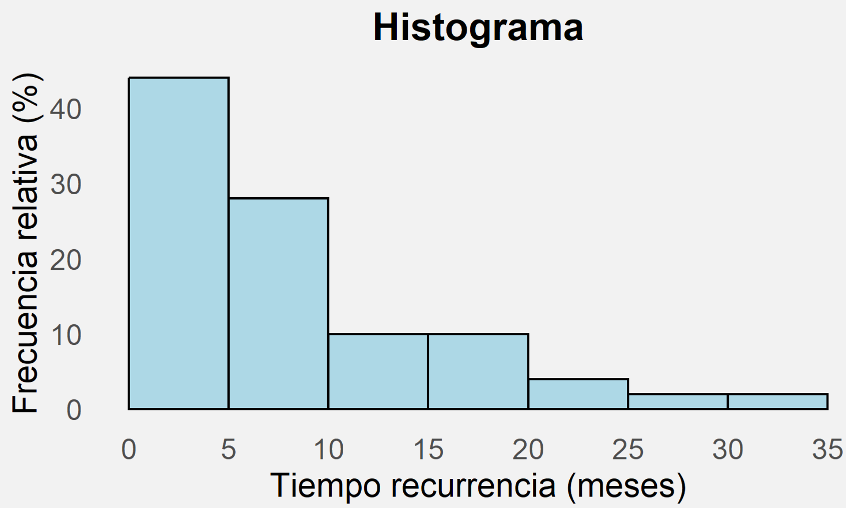
Solución:

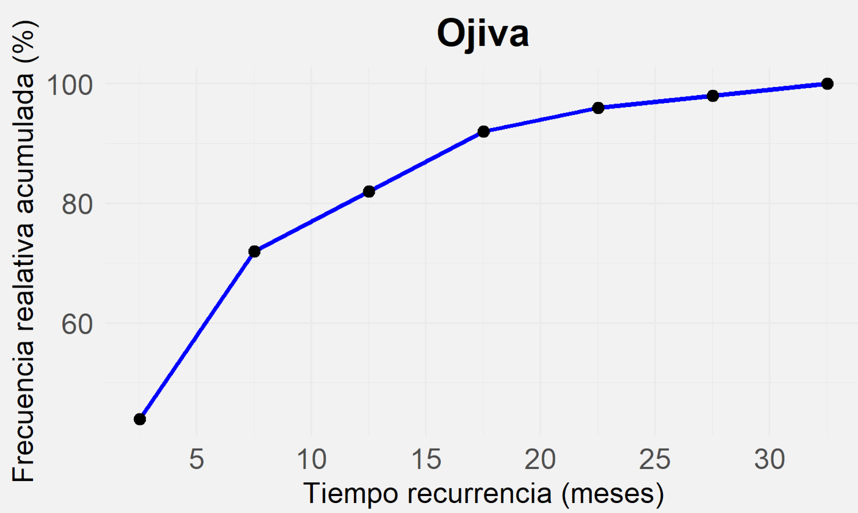
Conviene ordenar los $n = 50$ datos para contarlos más fácilmente:

0.2	0.2	0.3	0.4	1.0	1.2	1.3	1.4	1.6	1.6
2.0	2.1	2.4	2.4	2.7	3.3	3.5	3.7	3.9	4.1
4.3	4.4	5.6	5.8	6.1	6.6	6.9	7.4	7.4	8.2
8.2	8.3	8.7	9.0	9.6	9.9	11.4	12.6	13.5	14.1
14.7	16.7	18.0	18.0	18.4	19.2	23.1	24.0	26.7	32.3

a)

Marca de clase	**Intervalo de clase**	**Frecuencia absoluta**	**Frecuencia Relativa**	**Frecuencia Relativa Acumulada**
2.5	[0-5)	22	0.44	0.44
7.5	[5-10)	14	0.28	0.72
12.5	[10-15)	5	0.10	0.82
17.5	[15-20)	5	0.10	0.92
22.5	[20-25)	2	0.04	0.96
27.5	[25-30)	1	0.02	0.98
32.5	[30-35)	1	0.02	1.00
	Total	50	1.00	





b) El histograma muestra claramente que la distribución tiene cola larga a a la derecha, es decir la distribución tiene “Sesgo positivo”

c) Fracción de tiempos de recurrencia menores o iguales a 10 meses = 0.72

■

Ejercicio 1.6 La siguiente lista contiene las edades (en meses) de 50 niños que se inscribieron en una estancia infantil.

38	47	32	55	42	40	30	35	39	40
48	36	31	36	35	34	43	41	36	41
43	48	40	34	41	30	46	35	40	30
46	37	39	33	32	32	45	42	41	36
50	50	37	39	33	45	38	46	36	31

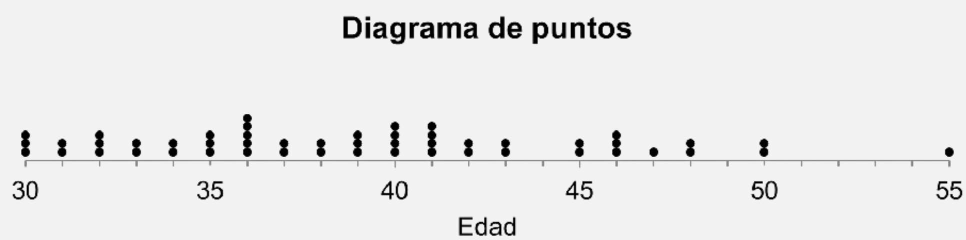
a) Calcule la media, la mediana y la moda para estos datos.
b) Elabore un diagrama de puntos.
c) Elabore un histograma de frecuencia relativa con la frontera inferior de la primera clase en 30 y utilice un ancho de clase de 5 meses.
d) Compare las dos gráficas anteriores. ¿Cuál elegiría como el mejor método para exhibir los datos y por qué?
e) ¿Qué proporción de los niños tenían 35 meses o más, pero menos de 45 meses cuando se inscribieron?
f) Si un niño de este grupo fuera seleccionado al azar, ¿cuál es la probabilidad de que tenga menos de 50 meses cuando se inscribió?

Solución:

Conviene reordenar los datos en orden creciente: 30, 30, 30, 31, 31, 32, 32, 32, 33, **33**, 34, 34, 35, 35, 35, 36, 36, 36, 36, 36, 37, 37, 38, 38, 39, 39, 39, 40, 40, **40**, 40, 41, 41, 41, 41, 42, 42, 43, 43, **45**, 45, 46, 46, 46, 47, 48, 48, 50, 50, **55**

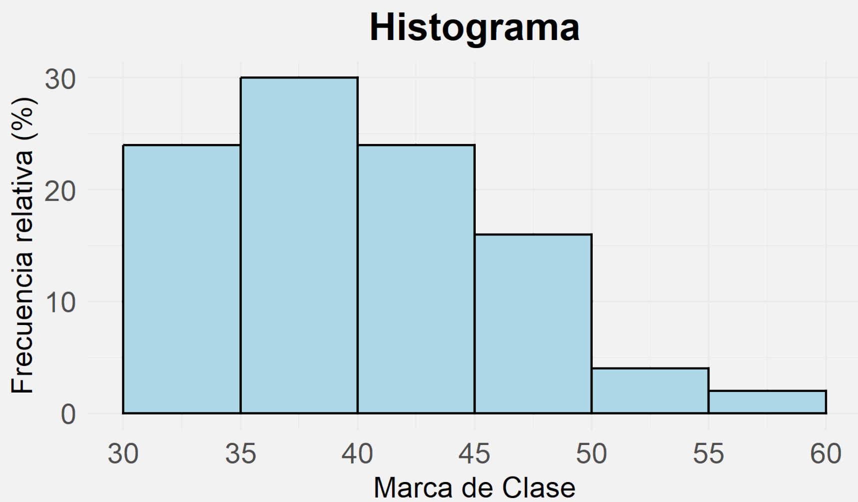
a) Media = $\overline{x} = 39.08$; Mediana = 39 ; Moda = 36

b)



c)

Marca de clase	Intervalo de clase	Frecuencia absoluta	Frecuencia Relativa	Frecuencia Relativa Acumulada
32.5	[30-35)	12	0.24	0.24
37.5	[35-40)	15	0.30	0.54
42.5	[40-45)	12	0.24	0.78
47.5	[45-50)	8	0.16	0.94
52.5	[50-55)	2	0.04	0.98
57.5	[55-60)	1	0.02	1.00
	Total	50	1.00	



d) En este caso el histograma me permite percibir mejor la forma de la distribución.
e) Proporción de niños de 35 meses o más, pero menos de 45 meses = 0.30 + 0.24 = 0.54
f) $P(X < 50) = 1 - P(X \geq 50) = 1 - 0.04 + 0.2 = 0.94$

■

Ejercicio 1.7 Para medir el impacto de la contaminación se midieron las concentraciones de mercurio (microgramos/gramo) en los hígados de 28 delfines, los cuales se listan a continuación: 1.70, 183.00, 1.72, 168.00, 8.80, 218.00, 5.90, 180.00, 101.00, 264.00, 85.40, 81.00, 118.00, 485.00, 221.00, 286.00, 406.00, 315.00, 252.00, 241.00, 329.00, 397.00, 316.00, 209.00, 445.00, 314.00, 278.00, 318.00

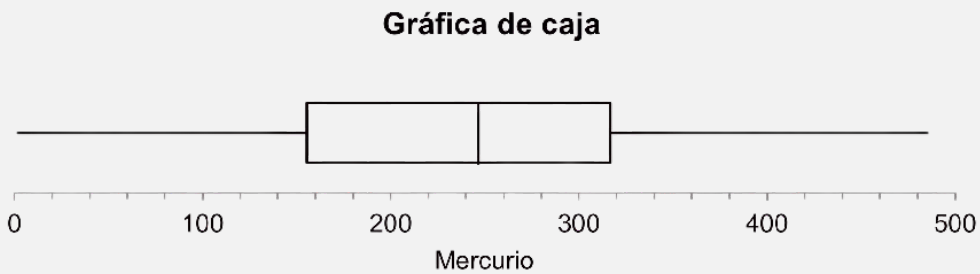
a) Calcule el "resumen de cinco números" para estos datos.
b) Construya una gráfica de caja.
c) ¿Hay algún resultado atípico?
d) Si le revelaran que los cuatro delfines cuyos valores son menores tenían menos de cinco años de edad, en tanto que los otros eran mayores de nueve años, ¿explique si esta información le ayudaría a explicar la gran diferencia entre la magnitud de esos cuatro datos y los demás?.

Solución:

a) Reordenando los $n = 28$ datos: 1.70, 1.72, 5.90, 8.80, 85.40, 101.00, 118.00, 168.00, 180.00, 183.00, 209.00, 218.00, 221.00, 241.00, 252.00, 264.00, 278.00, 286.00, 314.00, 315.00, 316.00, 318.00, 329.00, 397.00, 406.00, 445.00, 481.00, 485.00

Se puede observar que: min = 1.70, $Q_1 = 143$, Med = $Q_2 = 246.5, Q_3 = 317$, max= 485 y entonces el "resumen de cinco números" es: (1.70, 143, 246.5, 317, 485)

b) Gráfica de caja - Esta gráfica se construye a partir del resumen de cinco números:



c) No individualmente, ya que ningún dato sobrepasa los límites $Q_3 + 1.5(Q_3 - Q_1)$, y $Q_3 - 1.5(Q_3 - Q_1)$ pero las 4 datos más pequeños están muy alejados del resto.
d) Sí, ya que han acumulado menos mercurio durante su corta vida.

■

Ejercicio 1.8 Se contaron el número de cacahuates japoneses en cada paquete de 60gr de una marca genérica y de la marca Kawasaki en 14 paquetes de cada marca. A continuación se presentan los datos:

Marca genérica	25, 26, 26, 26, 26, 25, 28, 28, 28, 27, 27, 24, 25, 26
Kawasaki	25, 29, 24, 24, 28, 24, 28, 22, 25, 28, 30, 27, 28, 24

a) ¿Calcule la media y desviación estándar para la marca genérica?
b) ¿ Calcule la media y desviación estándar para la marca Kawasaki?
c) Compare los centros y variabilidades de las dos marcas usando los resultados de las partes a y b. Explique.

Solución:

1. $\bar{x} = 26.2, s = 1.25$
2. $\bar{x} = 26.1, s = 2.41$
3. Los centros son muy cercanos, pero la variabilidad es más grande para los Kawasaki. Lo cual quiere decir que aunque en promedio traen aproximadamente el mismo número de cacahuates, los paquetes de Kawasaki varían más en su contenido.

■

Ejercicio 1.9 Un fabricante registró las millas por galón (mpg) para cada uno de los 20 autos seleccionados al azar de una línea de producción durante un mes. 23.1, 21.3, 20.2, 24.4, 24.7, 22.7, 25.9, 24.7, 24.9, 22.2 23.6, 23.7, 25.3, 27.0, 26.2, 23.2, 24.4, 24.2, 22.9, 24.6

a) ¿Cuáles son el máximo y el mínimo de millas por galón?
b) ¿Cuál es el rango?
c) Calcule la media y la desviación estándar.
d) Encuentre el valor de los tres cuartiles.
e) Construya una gráfica de caja.

Solución:

Reordenando los datos se encuentran: 20.2, 21.3, 22.2, 22.7, 22.9, 23.1, 23.2, 23.6, 23.7, 24.2, 24.4, 24.4, 24.6, 24.7, 24.7, 24.9, 25.3, 25.9, 26.2, 27.0

a) Max = 27, Min = 20.2, Rango = 6.8
b) $\bar{x} = 23.96, s = 1.64$
c) $Q_1 = 23, Q_2 = 24.3, Q_3 = 24.8$
d) Diagrama de Caja

■

Ejercicio 1.10 En un estudio sobre contaminación ambiental, se determinó el contenido de arsénico de 11 muestras de agua tomadas de un lago. Se obtuvieron los siguientes resultados, en partes por millón (ppm):

170	290	120	150	100	180	210	290	240	120	220

a) Elabore una diagrama de puntos y un diagrama de tallos y hojas.
b) La media, la mediana y moda.
c) Rango, varianza, desviación estándar y coeficiente de variación.

Solución:

a)

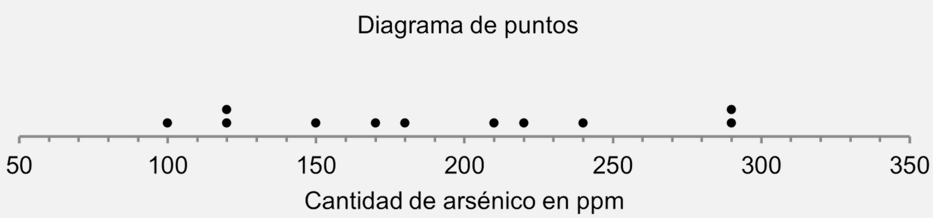


Diagrama de tallos y hojas

Unidad del tallo:100, Unidad de la hoja:10

1A	|	0	2	2
1B	|	5	7	8
2A	|	1	2	4
2B	|	9	9	

b) La media de los datos de una muestra se calcula mediante la siguiente expresión:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum_{i=1}^{11} x_i}{11} = \frac{170+290+\cdots+220}{11}$$

$\bar{x} = 190$ ppm

Para calcular la mediana se ordenan los datos de forma ascendente (menor a mayor):

100	120	120	150	170	180	210	220	240	290	290

Se calcula la posición del dato que está en el centro considerando que el número de datos es impar:

$$\text{Posición de la mediana} = \frac{n+1}{2} = \frac{11+1}{2} = 6$$

entonces: $\tilde{x} = x_6 = 180$ ppm

La moda es el valor que se repite más veces (que tiene la mayor frecuencia), en este caso los valores 120 y 290 se repiten dos veces, siendo este el mayor número de repeticiones de todos los valores (todos los demás valores sólo aparecen una vez), por lo que:

$$\text{Mo} = \{120 \text{ ppm}, 290 \text{ ppm}\}$$

Dado que estos datos tienen 2 modas, se dice que los datos son multimodales (si sólo tienen una moda se les denomina unimodales)

c) Rango:

$$R = x_{\text{máx}} - x_{\text{mín}} = 290 - 100 = 190 \text{ ppm}$$

Varianza muestral:

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^{11} (x_i - 190)^2}{11-1}$$

$$s^2 = \frac{(100-190)^2 + (120-190)^2 + \cdots + (290-190)^2}{10} = \frac{43800}{10}$$

$$s^2 = 4380 \text{ ppm}^2$$

Desviación estándar muestral:

Se define como: $s = \sqrt{s^2}$, por lo que: $s = \sqrt{4380 \text{ ppm}^2} = 66.1816$ ppm

Coeficiente de variación:

Se define como: $CV = \frac{s}{\bar{x}}$, por lo tanto: $CV = \frac{66.1816 \text{ ppm}}{190 \text{ ppm}} = 0.3483$

■

Ejercicio 1.11 Con los datos del ejercicio anterior compruebe que se cumple el teorema de Chebychev para $k = 1.2$.

Solución:

Los datos ordenados del Ejercicio 1.10 son:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}
100	120	120	150	170	180	210	220	240	290	290

El teorema de Chebychev establece que en cualquier población la proporción de datos que se encuentran en el intervalo dado por: $\mu - k\sigma < x < \mu + k\sigma$ es mayor o igual a: $1 - \dfrac{1}{k^2}$. En notación matemática:

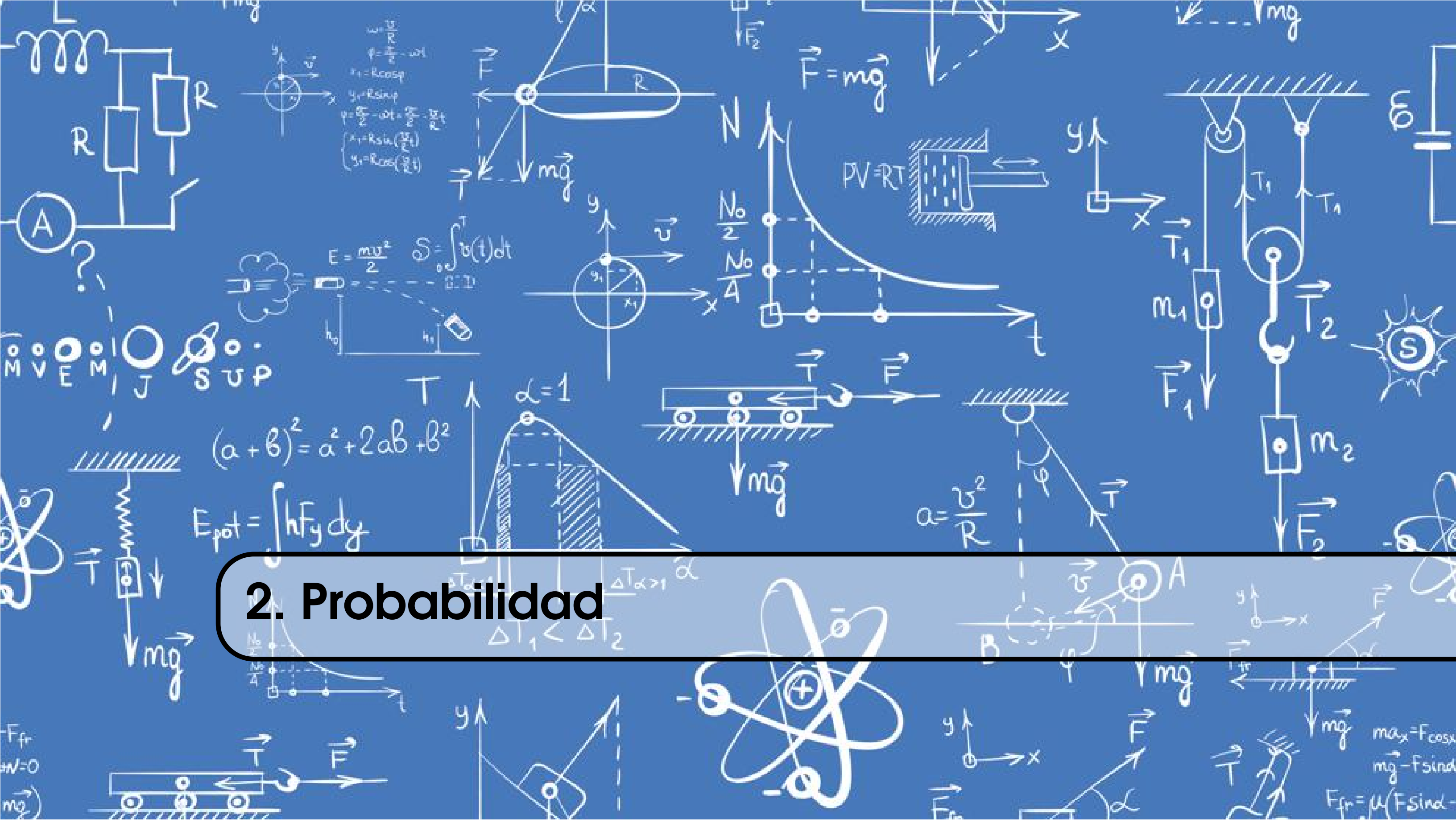
$$P(\mu - k\sigma < x < \mu + k\sigma) \geq 1 - \frac{1}{k^2} \quad ; \quad k > 1$$

Todo conjunto de datos puede ser considerado como una población, por lo que en nuestro caso $\mu = 190$ y $\sigma^2 = 66.1816$. En nuestro ejercicio, para $k = 1.2$, tendremos:

$$P(190 - (1.2)(66.1816) < x < 190 + (1.2)(66.1816)) \geq 1 - \frac{1}{1.2^2} \quad ; \quad k > 1$$

$$P(110.5821 < x < 269.4179) \geq 0.3056$$

Dado que del dato x_2 hasta el dato x_9 pertenecen al intervalo dado por $110.5821 < x < 269.4179$, se tiene que la proporción de datos es $P(110.5821 < x < 269.4179) = \frac{8}{11} = 0.7273$, por lo que se cumple el teorema de Chebychev dado que se cumple que: $0.727 \geq 0.3056$. ■



2. Probabilidad

Ejercicio 2.1 Una caja contiene tres pelotas rojas y dos azules. Se seleccionan dos pelotas al azar y se anotan sus colores. Si consideramos el orden en el que se sacan las pelotas, utilice un diagrama de árbol para hacer la lista de los 20 eventos simples del experimento.¿Cuál es la probabilidad de que las dos sean iguales?

Solución:

Como el orden importa, las 5 pelotas $(A1, A2, R1, R2, R3)$ se pueden extraer de $O_{5,2} = 5!/(5-2)! = 20$ (ordenaciones sin repetición) posibles parejas ordenadas, estas son:

$$
\begin{aligned}
S = \quad & \{(A1A2),(A1R1),(A1R2),(A1R3),(A2R1),(A2R2),(A2R3),(R1R2),(R1R3),(R2R3),\\
& (A2A1),(R1A1),(R2A1),(R3A1),(R1A2),(R2A2),(R3A2),(R2R1),(R3R1),(R3R2)\}
\end{aligned}
$$

Las pelotas son iguales en 8 de las 20 combinaciones . Entonces $P = 8/20 = 0.4$ ■

Ejercicio 2.2 De una bolsa que contiene cuatro monedas: una de cinco, una de 10, una de 25 y una de 50 centavos, se extraen tres monedas al azar.

a) Haga una lista de todos los eventos simples en el espacio muestral S.
b) ¿Cuál es la probabilidad de que la selección contenga la moneda de 10 centavos?
c) ¿Cuál es la probabilidad de que la suma de las monedas extraídas sea igual a 60 centavos o más?

Solución:

a) $S = \{(5,10,25),(5,10,50),(5,25,50),(10,25,50)\}$
b) Ya que 3 de los 4 eventos simples la contienen, la probabilidad de que la selección contenga la moneda de 10 es $P = 3/4$
c) Ya que en 3 de los 4 eventos la suma es mayor o igual a 60, $P = 3/4$

■

Ejercicio 2.3 Tres personas son seleccionadas al azar para ser miembros de un jurado y se toma nota del género de cada persona.

a) Haga una lista de los eventos simples en el espacio muestral S.
b) Asígnele una probabilidad a cada evento simple de S asumiendo que es igualmente probable que cada persona sea hombre o mujer.
c) ¿Cuál es la probabilidad de que sólo una de las tres sea mujer?
d) ¿Cuál es la probabilidad de que las tres personas sean mujeres?

Solución:

a) Considerando el orden de selección y que el género se puede repetir, el número de posibles tercias son $OR_{2,3} = 2^3 = 8$ (ordenaciones con repetición). Estas son:

$$S = \{(H,H,H),(H,H,M),(H,M,H),(M,H,H),(H,M,M),(M,H,M),(M,M,H),(M,M,M)\}$$

donde H =Hombre, M =Mujer
b) La probabilidad de cada evento simple es $P = 1/8$
c) Como 3 de los 8 eventos simples contienen solo 1 mujer, $P = 3/8$
d) Como solo 1 de los 8 eventos simples contienen 3 mujeres, $P = 1/8$

■

Ejercicio 2.4 Una compañía planea efectuar un experimento para comparar su marca de café con la de sus dos principales competidores. Se contrata a un experto para probar y calificar cada una de las tres marcas de café identificadas al azar como A, B y C.

a) Haga una lista de eventos simples en el espacio muestral S.
b) Si el experto no tiene capacidad para distinguir entre las tres marcas, ¿cuál es la probabilidad de que califique el café A como el mejor? ¿Y como el peor?

Solución:

a) Las posibles ordenaciones, sin repetición, de 3 elementos son $O_{3,3} = 3! = 6$.
Estas son: $S = \{(A,B,C),(A,C,B),(B,A,C),(B,C,A),(C,A,B),(C,B,A)\}$
b) Si el experto no distingue, cualquier orden tiene la misma probabilidad de ocurrir y entonces $P(A \text{ mejor}) = 2/6 = 1/3$, y también $P(A \text{ peor}) = 2/6 = 1/3$

■

Ejercicio 2.5 Se lanzan dos dados imparciales.

a) ¿Cuál es la probabilidad de que la suma de los puntos en las caras superiores sea igual a 7? ¿A 11?
b) ¿Cuál es la probabilidad de que los dos dados tengan el mismo número de puntos en la cara superior?
c) ¿Cuál es la probabilidad de que los dos dados muestren un número par?

Solución:

Primero calculemos el número de posibles resultados. Existen $OR_{6,2} = 6^2 = 36$ (ordenaciones con repetición) posibles resultados al lanzar 2 dados:(1,1),(1,2),...(1,6), (2,1)... (6,6)

Sea $P(x,z)$ la probabilidad de que ocurra x en el 1er dado y z en el segundo dado. Entonces:

a) $P(\text{suma de } x+z=7) = P((1,6)\cup(2,5)\cup(3,4)\cup(4,3)\cup(5,2)\cup(6,1))$
$= P(1,6)+P(2,5)+P(3,4)+P(4,3)+P(5,2)+P(6,1) = \frac{6}{36} = \frac{1}{6}$,
y $P(\text{suma de } x+z=11) = P((5,6)\cup(6,5)) = P(5,6)+P(6,5) = \frac{2}{36} = \frac{1}{11}$.
b) $P(x=z) = P((1,1)\cup(2,2)\cup\ldots\cup(6,6)) = \frac{6}{36} = \frac{1}{6}$.
c) $P(\text{ambos pares}) = P(x \text{ es par } \cap z \text{ es par}) = P(x \text{ es par}) * P(z \text{ es par}) = \frac{1}{2} * \frac{1}{2} = \frac{1}{4}$.

■

Ejercicio 2.6 Un detector de humo utiliza dos sensores, A y B. Cuando hay humo, la probabilidad de que el sensor A lo detecte es 0.95, de que el sensor B lo detecte es 0.98 y que los dos sensores lo detecten es 0.94.

a) Si hay humo, encuentre la probabilidad de que sea detectado por el sensor A o el B o por ambos sensores.
b) ¿Cuál es la probabilidad de que ningún sensor detecte el humo?

Solución:

Sea $a =$ el sensor A detecte y $b =$ el sensor B detecte. Entonces:

a) $P(a \cup b) = P(a) + P(b) - P(a \cap b) = 0.95 + 0.98 - 0.94 = 0.99$.
b) $P((a \cup b)^c) = 1 - P(a \cup b) = 1 - 0.99 = 0.01$.

■

Ejercicio 2.7 Un experimento consiste en tirar un dado una sola vez y observar el número de puntos en la cara superior. Los eventos A, B y C están definidos como sigue:

A: resulte un número menor que 4.
B: resulte un número menor o igual a 2.
C: resulte un número mayor que 3.

Calcule las probabilidades de cada uno de los siguientes eventos, utilizando ya sea la suma de probabilidades de los eventos simples o las reglas y definiciones principales de la probabilidad.

a) S b) A | B c) B d) A y B y C e) A y B f) A y C g) B y C h) A o C i) B o C
j) ¿Los eventos A y B son independientes?
k) ¿Los eventos A y B son mutuamente excluyentes?
l) ¿Los eventos A y C son independientes?
m) ¿Los eventos A y C son mutuamente excluyentes?

Solución:

a) $P(S) = 1$, por definición.
b) $P(A|B) = 1$, porque si B ocurre, es seguro que A también ocurre
c) $P(B) = 2/6 = 1/3$
d) $P(A \cap B \cap C) = 0$, porque A, B y C no pueden ocurrir al mismo tiempo
e) $P(A \cap B) = 2/6 = 1/3$
f) $P(A \cap C) = 0$, porque si C ocurre, A no puede ocurrir.
g) $P(B \cap C) = 0$, porque si C ocurre, B no puede ocurrir.
h) $P(A \cup C) = 1$, porque $A \cup C = S$
i) $P(B \cup C) = 5/6$
j) No, ya que la probabilidad de que A ocurra depende de la ocurrencia o no de B.
k) No, ya que A y B pueden ocurrir al mismo tiempo.
l) No. Al ser eventos complementarios, la probabilidad de que A ocurra depende de la ocurrencia o no de B.
m) Sí, ya que al ser complementarios A y C no pueden ocurrir al mismo tiempo.

■

Ejercicio 2.8 Los proyectos de investigación presentados a una institución fueron evaluados por un comité de expertos para su financiamiento. Los proyectos se enviaron después a un segundo comité de expertos independiente, los cuales revirtieron la decisión del primer comité en 30% de los casos. Si la probabilidad de que un proyecto sea aprobado por el primer grupo es 0.2, ¿cuáles son las probabilidades de los siguientes eventos?

a) Un proyecto es aprobado por los dos grupos.
b) Un proyecto es desaprobado por los dos grupos.
c) Un proyecto es aprobado por un grupo.

Solución:

Sea A_i = Aceptado en comité i, R_i = Rechazado en comité i. Ambos comités son independientes.

a) $P(\text{Un proyecto es aprobado por los dos grupos}) = P(A_1 \cap A_2) = P(A_1) * P(A_2) = 0.2 * (1 - 0.3) = 0.14$.
b) $P(\text{Un proyecto es desaprobado por los dos grupos}) = P(R_1 \cap R_2) = P(R_1) * P(R_2) = 0.8 * (1 - 0.3) = 0.56$.
c) $P(\text{Un proyecto es aprobado por un grupo}) = P(A_1 \cap R_2) + P(R_1 \cap A_2) = P(R_2|A_1)P(A_1) + P(A_2|R_1)P(R_1) = 0.2 * 0.3 + 0.8 * 0.3 = 0.3$.

■

Ejercicio 2.9 Los artículos de una línea de producción son revisados visualmente por dos inspectores consecutivos. La probabilidad de que un artículo defectuoso pase desapercibido por el primer inspector es 0.1. De aquellos que pasan al primer inspector, el segundo inspector dejará pasar a la mitad. ¿Qué porcentaje de artículos defectuosos pasarán desapercibidos por los dos inspectores?

Solución:

Sea A_i = Pase en la etapa i:

$P(\text{Pase desapercibido por los dos inspectores}) = P(A_1 \cap A_2) = P(A_2|A_1)P(A_1) = (0.5)(0.1) = 0.05$. ■

Ejercicio 2.10 Se selecciona al azar un individuo que puede provenir de una de dos poblaciones, S_1 y S_2. La probabilidad de que provenga de S_1 es $P(S_1) = 0.7$ y de S_2 es $P(S_2) = 0.3$. Si proviene de S_1, la probabilidad de observar un evento A es $P(A|S_1) = 0.2$, y si proviene de S_2, la probabilidad de observar A es $P(A|S_2) = 0.3$.

a) Si se extrae una muestra al azar de una de las dos poblaciones, ¿cuál es la probabilidad de que se observe A?
b) Si en la muestra al azar se observa A, ¿cuál es la probabilidad de que la muestra haya sido

seleccionada de la población S_2?

Solución:

a) Usando el Teorema de Probabilidad Total:

$$P(A) = P(A|S_1)P(S_1) + P(A|S_2)P(S_2) = 0.2 * 0.7 + 0.3 * 0.3 = 0.14 + 0.09 = 0.23.$$

b) Usando la definición de probabilidad condicional:

$$P(S_2|A) = \frac{P(S_2 \cap A)}{P(A)} = \frac{P(A|S_2)P(S_2)}{P(A)} = \frac{(0.3)(0.3)}{0.23} = 0.39.$$

■

Ejercicio 2.11 En una población de luciérnagas hay dos subgrupos que se presentan con probabilidades de 40 % y 60 %, respectivamente. Un evento A ocurre en el 50 % de los individuos del primer subgrupo y en 30 % de los del segundo subgrupo. ¿Cuál es la probabilidad incondicional de que ocurra el evento A, sin importar el subgrupo de donde provenga la luciérnaga?

Solución:

Sea $g_i =$ proviene del grupo i.

Usando el Teorema de Probabilidad Total:

$$P(A) = P(A|g_1)P(g_1) + P(A|g_2)P(g_2) = 0.3 * 0.6 + 0.5 * 0.4 = 0.18 + 0.2 = 0.38.$$

■

Ejercicio 2.12 Un operador produce un artículo defectuoso con probabilidad 0.01 si opera la máquina siguiendo el orden recomendado y con probabilidad 0.03 si no lo sigue. Si el operador sigue el orden recomendado 90 % de las veces, ¿qué proporción de artículos defectuosos producirá?

Solución:

Sea $D =$ defectuoso, $S =$ sigue orden, $N =$ no sigue orden. Usando el Teorema de Probabilidad Total:

$$P(D) = P(D|S)P(S) + P(D|N)P(N) = 0.01 * 0.9 + 0.03 * 0.1 = 0.009 + 0.003 = 0.012.$$

■

Ejercicio 2.13 Suponga que un conjunto de síntomas que llamaremos el evento H, ocurre solamente cuando está presente alguna de las enfermedades, A, B o C. Estas enfermedades se excluyen mutuamente y suceden con probabilidades $P(A) = 0.01$, $P(B) = 0.005$ y $P(C) = 0.02$. Además, las probabilidades de desarrollar los síntomas H, para cada una de las enfermedades, son $P(H|A) = 0.90$, $P(H|B) = 0.95$ y $P(H|C) = 0.75$. Suponga que una persona con los síntomas H acude al médico, ¿cuál es la probabilidad de que la persona tenga la enfermedad A?

Solución:

Usando la definición de probabilidad condicional:

$$P(A|H) = \frac{P(H|A)P(A)}{P(H)} = \frac{0.9 * 0.01}{0.9 * 0.01 + 0.95 * 0.005 + 0.75 * 0.02}$$

$$= \frac{0.009}{0.009 + 0.00475 + 0.015} = \frac{0.009}{0.02875} = 0.313.$$

■

Ejercicio 2.14 Suponga que una enfermedad está presente en el 10% de la población, y que la prueba que se utiliza para detectarla no siempre funciona bien. Puede dar negativo cuando la enfermedad está presente o dar positivo cuando está ausente de acuerdo a la siguiente tabla que muestra la proporción de veces que se presenta cada situación.

	Examen Positivo (P)	**Examen Negativo (N)**
Enfermedad Presente (D)	0.08	0.02
Enfermedad Ausente (D^C)	0.05	0.85

a) Encuentre las siguientes probabilidades: $P(D)$, $P(D^C)$, $P(N|D^C)$, $P(N|D)$.
b) Use la regla de Bayes y los resultados anteriores para hallar $P(D|N)$.
c) Use la definición de probabilidad condicional para calcular $P(D|N)$ (Debe ser igual al inciso b).
d) Encuentre la probabilidad de un falso positivo, es decir que el examen dé positivo, dado que la persona no está enferma.
e) Encuentre la probabilidad de un falso negativo, es decir que el examen sea negativo, dado que la persona sí está enferma.

Solución:

a) $P(D) = 0.08 + 0.02 = 0.1$, $\quad P(D^C) = 0.05 + 0.85 = 0.9$,
$P(N|D^C) = 0.85/0.90 = 0.944$, $\quad P(N|D) = 0.02/0.10 = 0.2$.
b) $P(D|N) = \dfrac{0.02}{0.87} = 0.023.$

c) $P(D|N) = \dfrac{P(D \cap N)}{P(N)} = \dfrac{0.02}{0.87} = 0.023.$

d) $P(+|D^C) = \dfrac{P(D^C) * P(+|D^C)}{P(D^C)} = \dfrac{0.05}{0.90} = 0.055.$

e) $P(-|D) = P(D|-) = \dfrac{P(-|D) * P(D)}{P(-)} = \dfrac{0.02}{0.10} = 0.2.$ ∎

Ejercicio 2.15 Clasifique como discretas o continuas las siguientes variables aleatorias:

- X: el número de accidentes automovilísticos que ocurren al mes en Mérida.
- Y: el tiempo para jugar una partida de ajedrez.
- M: la cantidad de leche que una cierta vaca produce al mes.
- N: el número de huevos que pone una cierta gallina a la semana.
- P: el número de licencias para conducir que se emiten cada mes en una ciudad.
- Q: el peso de café producido por hectárea en una finca.

Solución:

X es Discreta, Y es Continua, M es Continua, N es Discreta, P es Discreta y Q es Continua. ∎

Ejercicio 2.16 Sea W la variable aleatoria que da el número de Águilas menos el número de Soles en tres lanzamientos de una moneda. Liste los elementos del espacio muestral S para los tres lanzamientos de la moneda y asigne un valor w de W a cada punto muestral. Por último, asigne una probabilidad de ocurrencia a cada valor de W.

Solución:

$S = \{\text{AAA}, \text{AAS}, \text{ASA}, \text{SAA}, \text{ASS}, \text{SAS}, \text{SSA}, \text{SSS}\}.$
Respectivamente, W toma los valores:

$W = \{3, 1, 1, 1, -1, -1, -1, -3\}.$

Y entonces:

$P(W = -3) = \dfrac{1}{8}, \quad P(W = -1) = \dfrac{3}{8}, \quad P(W = 1) = \dfrac{3}{8}, \quad P(W = 3) = \dfrac{1}{8}.$ ∎

Ejercicio 2.17 El número total de horas, medidas en unidades de 100 horas, que una persona usa la televisión en un año es una variable aleatoria continua X que tiene la siguiente función de densidad:

$$f(x) = \begin{cases} x, & 0 < x < 1 \\ 2 - x, & 1 \leq x \leq 2 \\ 0, & \text{en otro caso.} \end{cases}$$

Calcule la probabilidad de que en un periodo de un año esta persona utilice su televisión

a) menos de 120 horas
b) entre 50 y 100 horas.

Solución:

a)

$$P(X < 120) = \int_0^{1.2} f(x)dx = \int_0^1 xdx - \int_1^{1.2} (2-x)dx = 0.5 + 0.4 - (0.72 - 0.5) = 0.68.$$

b)

$$P(50 < X < 100) = \int_{0.5}^{1} f(x)dx = 0.5 - 0.125 = 0.375.$$

■

Ejercicio 2.18 La distribución de probabilidad de X, el número de baches que se encuentran en cada 100 metros de una carretera, está dada por:

x	0	1	2	3	4
$f(x)$	0.41	0.37	0.16	0.05	0.01

Construya la función de distribución acumulada de X.

Solución:

$$F(x) = \sum_{z=0}^{x} f(z) \Rightarrow F(0) = 0.41, F(1) = 0.78, F(2) = 0.94, F(3) = 0.99, F(4) = 1.0.$$ ■

Ejercicio 2.19 De una urna que contiene 4 bolas rojas, 2 bolas azules y 1 bola verde se sacan 2 bolas sin remplazo. Sea $X =$ número de bolas rojas y $Y =$ número de bolas azules. Se sabe que este experimento sigue la distribución de probabilidad conjunta hipergeomética de dos variables, de modo que:

$$f(x,y)=\begin{cases}\dfrac{\binom{4}{x}\binom{2}{y}\binom{2-x-y}{1}}{\binom{7}{3}} & \text{si} \quad x,y=0,1,2 \quad ; \quad 0\leq 2-x-y\leq 1\\[2ex] 0 & \text{en cualquier otro caso}\end{cases}$$

a) Represente la distribución de probabilidad conjunta en forma de tabla.
b) Determine las distribuciones de probabilidad marginales $g(x)$ y $h(y)$ de esta distribución de dos variables.
c) $P(x=0,y=0)$
d) $P(x=2,y=1)$
e) $P(x=1,y=1)$
f) ¿Cuál es la probabilidad de que al menos una sea roja?
g) ¿Cuál es la probabilidad de que una sea verde?

Solución:

a)

	x			
y	**0**	**1**	**2**	**Total**
0	0	$\frac{4}{21}$	$\frac{6}{21}$	$\frac{10}{21}$
1	$\frac{2}{21}$	$\frac{8}{21}$	0	$\frac{10}{21}$
2	$\frac{1}{21}$	0	0	$\frac{1}{21}$
Total	$\frac{3}{21}$	$\frac{12}{21}$	$\frac{6}{21}$	1

b) Probabilidades marginales:

$$g(x)=\sum_{y=0}^{2} f(x,y)=f(x,0)+f(x,1)+f(x,2)$$

Por ejemplo: $g(1)=\sum_{y=0}^{2} f(1,y)=f(1,0)+f(1,1)+f(1,2)=\frac{4}{21}+\frac{8}{21}+0=\frac{12}{21}$; corresponde con el total de columna con $x=1$ en la tabla anterior.

$$h(y)=\sum_{x=0}^{2} f(x,y)=f(0,y)+f(1,y)+f(2,y)$$

Por ejemplo: $h(2)=\sum_{x=0}^{2} f(x,2)=f(0,2)+f(1,2)+f(2,2)=\frac{1}{21}+0+0=\frac{1}{21}$; corresponde con el total de renglón con $y=2$ en la tabla anterior.

c) $P(x=0,y=0)=f(0,0)=0$ Este resultado es un evento imposible porque el resultado $x=0$, $y=0$ implica que deberían salir 2 bolas verdes, lo cúal no puede ocurrir dado que sólo hay una bola verde en la urna.

d) Es imposible que ocurra algún resultado con 2 bolas rojas y una azul porque implica a 3 bolas y sólo se están sacando 2, por lo que: $P(x=2,y=1)=0$

e) $P(x \geq 1, y) = \sum\limits_{x=1}^{2}\sum\limits_{y=0}^{2} f(x,y) = \sum\limits_{x=1}^{2} = g(1) + g(2)$

Esto se puede obtener sumando el total de las columnas correspondientes a $x = 1$ y $x = 2$, por lo que
$P(x \geq 1, y) = g(1) + g(2) = \frac{12}{21} + \frac{6}{21} = \frac{18}{21}$

f) $P((x,y) : x + y = 1) = f(0,1) + f(1,0) = \frac{2}{21} + \frac{4}{21} = \frac{6}{21}$, la expresión $(x,y) : x + y = 1$ se pueden entender como los casos en que el resultado es una verde y otra de cualquiera de los otros dos colores.

■

Ejercicio 2.20 Un determinado experimento aleatorio tiene la siguiente distribución de probabilidad conjunta de dos variables aleatorias continuas:

$$f(x,y) = \begin{cases} 4xy + 4x^2 & \text{si} \quad 0 \leq y \leq 2x - 1 \;\; ; \;\; 0.5 < x < 1 \\ 0 & \text{en cualquier otro caso} \end{cases}$$

a) Determine las distribuciones de probabilidad marginales $g(x)$ y $h(y)$ de esta distribución de dos variables.
b) $P(0.7 \leq x \leq 1, y \geq 0.5)$

Solución:

a) $g(x) = \displaystyle\int_0^{2x-1} 4xy + 4x^2 \, dy = \; 2xy^2 + 4x^2y\Big|_0^{2x-1} = 2x(2x-1)^2 + 4x^2(2x-1) - 0$

$= 16x^3 - 12x^2 + 2x$

$h(y) = \displaystyle\int_{0.5}^{\frac{y+1}{2}} 4xy + 4x^2 \, dx = \; 2x^2y + \frac{4x^3}{3}\Bigg|_{0.5}^{\frac{y+1}{2}} = \frac{2y^3}{3} + \frac{3y^2}{2} + \frac{y}{2}$

b) $P(0.7 \leq x \leq 1, y \geq 0.5) = \displaystyle\int\limits_{0.7}^{1}\int\limits_{0.5}^{2x-1} xy + x^2 \, dydx = 0.2281$

■

Ejercicio 2.21 En el Ejercicio 2.18 calcule el número promedio de baches que hay en cada 100 metros de esa carretera y la varianza.

Solución:

$\mu = E(X) = \displaystyle\sum_0^4 xf(x) = 0.41*0 + 0.37*1 + \cdots + 4*0.01 = 0.88.$

$\mathrm{Var}(X) = E(X^2) - E(X)^2 = \displaystyle\sum_0^4 x^2 f(x) - \mu^2$

$= 0*0.41 + 1*0.37 + \cdots + 16*0.01 - 0.88^2 = 0.8456$ ■

Ejercicio 2.22 En un juego de azar se le pagan \$3 a una persona si saca un Rey o una Reina, y \$5 si saca un Jota o un As de una baraja normal de 52 cartas. Si saca otra carta, pierde. ¿Cuánto debería pagar por jugar si el juego es justo?

Solución:

$$E(g(x)) = \sum_{\text{todo } x} g(x) * f(x) = 3 * \frac{8}{52} + 5 * \frac{8}{52} = 1.23.$$ ▪

Ejercicio 2.23 Si una persona invierte en las acciones de una empresa, tiene una probabilidad de 0.3 de tener una ganancia de \$4000 y una probabilidad de 0.7 de tener una pérdida de \$1000 en el año. ¿Cuál es su ganancia esperada?

Solución:

$$E(g(x)) = \sum_{\text{todo } x} g(x) * f(x) = 4000 * 0.3 - 1000 * 0.7 = 500.$$ ▪

Ejercicio 2.24 Determine la covarancia de la distribución de probabilidad dada en el Problema 2.19.

Solución:

$$E(XY) = \sum_{y=0}^{2}\sum_{x=0}^{2} xyf(x,y) = f(1,1) + 2f(1,2) + 2f(2,1) + 4f(2,2) = \frac{8}{21}$$
$$\mu_X = \sum_{x=0}^{2} xg(x) = g(1) + 2g(2) = \frac{12}{21} + 2\frac{6}{21} = \frac{24}{21}$$
$$\mu_Y = \sum_{y=0}^{1} yh(y) = h(1) + 2h(2) = \frac{8}{21} + 2\frac{1}{21} = \frac{12}{21}$$
$$\sigma_{XY} = E(XY) - \mu_x\mu_y = \frac{8}{21} - \left(\frac{24}{21}\right)\left(\frac{12}{21}\right) = -\frac{40}{147}$$ ▪

Ejercicio 2.25 Determine la covarianza de la distribución de probabilidad deda en el problema 2.20.

Solución:

$$E(XY) = \int_{\frac{1}{2}}^{1}\int_{0}^{2x-1} xyf(x,y)\, dydx = \int_{\frac{1}{2}}^{1}\int_{0}^{2x-1} xy(4xy + 4x^2)\, dydx = \frac{529}{1440}$$

$$\mu_X = \int_{\frac{1}{2}}^{1} xg(x)\,dx = \int_{\frac{1}{2}}^{1} x(16x^3 - 12x^2 + 2x)\,dx = \frac{209}{240}$$

$$\mu_Y = \int_{0}^{1} yh(y)\,dy = \int_{0}^{1} y\left(\frac{2y^3}{3} + \frac{3y^2}{2} + \frac{y}{2}\right)\,dy = \frac{27}{40}$$

$$\sigma_{XY} = E(XY) - \mu_x\mu_y = \frac{529}{1440} - \left(\frac{209}{240}\right)\left(\frac{27}{40}\right) = -\frac{6349}{28800} \quad \blacksquare$$

3. Distribuciones discretas y continuas

Ejercicio 3.1 Un determinado experimento sigue una distribución uniforme. Se sabe que la variable aleatoria discreta toma los valores $x = 3, 4, 5, \ldots, 12$.

a) Determine la función de masa de la variable aleatoria discreta.
b) Calcule la esperanza y la varianza.
c) Calcule $P(x \leq 6)$

Solución:

a) Dado que es una distribución uniforme, la probabilidad de que $x = 3, 4, 5, \ldots, 12$ esta dada por el valor constante: $\dfrac{1}{b-a+1} = \dfrac{1}{12-3+1} = \dfrac{1}{10}$, entonces la función de masa de esta distribución es:

$$f(x) = \begin{cases} \dfrac{1}{10} & \text{si} \quad x = 3, 4, 5, \ldots, 12 \\ 0 & \text{en cualquier otro caso} \end{cases}$$

b) La esperanza $E(X)$ o μ de la distribución uniforme se calcula con la expresión:

$$E(X) = \mu = \frac{b+a}{2} = \frac{12+3}{2} = \frac{15}{2} = 7.5$$

y la varianza mediante:

$$\text{Var}(X) = \sigma^2 = \frac{(b-a+1)^2-1}{12} = \frac{(12-3+1)^2-1}{12} = \frac{99}{12} = 8.25$$

c) Se pide calcular la función acumulada de masa $F(x)$ hasta $x = 6$, por lo tanto:

$$F(x) = \sum_{x=3}^{6} f(x) = f(3) + f(4) + f(5) + f(6) = \frac{4}{10} = \frac{2}{5}$$

■

Ejercicio 3.2 Un alumno contestará una pregunta de opción múltiple, con 5 opciones y una única opción correcta.

a) Determine la función de masa de la variable aleatoria discreta.
b) Calcule la esperanza y la varianza.

Solución:

Este experimento sigue una distribución Bernoulli, consta de un único ensayo, sólo hay dos posibles resultados, acertar a la respuesta correcta (éxito) y seleccionar una opción incorrecta (fracaso). Consideremos que si se realizaran más de un ensayo, estos serían independiente entre sí y que la probabilidad de éxito p se mantiene constante entre ensayos para que se cumplan las condiciones necesarias para considerar que el ensayo es tipo Bernoulli. La probabilidad de éxito es $p = \frac{1}{5}$, dado que sólo hay una opción correcta de un total de 5 opciones, la probabilidad de fracaso: $q = 1 - p = 1 - \frac{1}{5} = \frac{4}{5}$, entonces:

a)

$$f(x) = \begin{cases} \frac{4}{5} & \text{si} \quad x = 0 \\ \frac{1}{5} & \text{si} \quad x = 1 \\ 0 & \quad \text{en cualquier otro caso} \end{cases}$$

b)

$$E(X) = \mu = p = \frac{1}{5}$$

$$\text{Var} = \sigma^2 = pq = \left(\frac{1}{5}\right)\left(\frac{4}{5}\right) = \frac{4}{25}$$

■

Ejercicio 3.3 En un aeropuerto se han registrado un promedio mensual de cinco accidentes a punto de ocurrir en aterrizajes en los últimos 5 años. Calcule la probabilidad de que:

a) Durante un mes no haya accidentes a punto de ocurrir en aterrizajes en el aeropuerto.
b) Durante un mes haya cinco accidentes a punto de ocurrir.
c) Haya más de cuatro accidentes a punto de ocurrir durante un mes particular.

Solución:

Corresponde a una variable Poisson ($\mu = 5$ accidentes/mes) $\Rightarrow P(X = x) = e^{-5} \cdot 5^x / x!$

a) $P(X = 0) = e^{-5} = 0.0067$
b) $P(X = 5) = 0.0067 \cdot (5^5)/5! = 0.1744$
c) $P(X \geq 5) = 1 - P(X \leq 4) = 1 - 0.0067(1 + 5 + (25/2) + (125/6) + (625/24)) = 0.56$

■

Ejercicio 3.4 De acuerdo con un estudio, los niños en edad escolar que se lesionan dos o más veces tienden a sufrir estas lesiones en un período relativamente corto, por lo general un año o menos. Si el número de lesiones promedio por año para niños en edad escolar es de dos, ¿cuáles son las probabilidades de estos eventos?

a) Un niño en edad escolar sufra dos lesiones durante el año.
b) Un niño en edad escolar sufra dos o más lesiones durante el año.
c) Un niño en edad escolar sufra cuando mucho una lesión durante el año.

Solución:

Corresponde a una variable Poisson ($\mu = 2$ lesiones/año)

a) $P(X = 2) = e^{-2} \cdot 2^2/2! = 0.2706$
b) $P(X \geq 2) = 1 - P(0) - P(1) = 1 - e^{-2} \cdot (1 + 2) = 0.594$
c) $P(X \leq 1) = P(0) + P(1) = 0.406$

■

Ejercicio 3.5 Un médico afirma que el 70 % de las personas con cáncer de pulmón son fumadores. Si su aseveración fuera correcta, calcule la probabilidad de que de 10 pacientes con cáncer de pulmón menos de la mitad sean fumadores.

Solución:

Corresponde a una variable Binomial con $p = 0.7$, $n = 10$,

$$P(X \leq 4) = \sum_{x=0}^{4} \binom{10}{x} p^x (1-p)^{10-x} =$$

$$= P(0)+P(1)+\cdots+P(4) = 1*0.7^0 0.3^{10} + 10*0.7^1 0.3^9 + \cdots + 126*0.7^4*0.3^6 = 0.0474$$

■

Ejercicio 3.6 Un frasco contiene cinco pelotas: tres rojas y dos blancas. Se eligen dos pelotas al azar sin devolverlas al frasco, y se registra el número X de pelotas rojas. Explique por qué X es o no una variable aleatoria binomial. (SUGERENCIA: ¿Cumple con las características de un experimento binomial?). Si el experimento es binomial, dé los valores de n y p.

Solución:

Los valores de n y p cambian de la 1a elección a la 2a elección $\Rightarrow$ **NO ES BINOMIAL** ■

Ejercicio 3.7 En el ejercicio anterior, suponga que el muestreo fue realizado con restitución. Esto es, suponga que la primera pelota se seleccionó del frasco, se observó y después fue devuelta al frasco, y que entonces las pelotas se mezclaron antes de seleccionar la segunda pelota. Explique por qué $\bar{X}$, el número de pelotas rojas observado, es o no una variable aleatoria binomial. Si el experimento es binomial, dé los valores de n y p.

Solución:

Los valores de n y p NO cambian de la 1a elección a la 2a elección $\Rightarrow$ **SÍ ES BINOMIAL** ■

Ejercicio 3.8 Un nuevo sistema de seguridad para casas está diseñado para tener un 99 % de confiabilidad. Suponga que nueve casas que están equipadas con este sistema de seguridad experimentaron un intento de robo. Encuentre las probabilidades de estos eventos:

a) Al menos una alarma se activó;
b) Más de siete alarmas se activaron;
c) Ocho o menos alarmas se activaron.

Solución:

Se trata de una variable Binomial con $n = 9$, $p = 0.99$.
a) $P(X \geq 1) = 1 - P(0) = 1 - 0.01^9 \approx 1$
b) $P(X \geq 8) = P(8) + P(9) = 9 \cdot 0.01 \cdot 0.99^8 + 1 \cdot 0.99^9 = 0.997$
c) $P(X \leq 8) = 1 - P(9) = 1 - 0.99^9 = 0.086$

■

Ejercicio 3.9 En cierta población, el 85 % de las personas tienen sangre tipo Rh positivo. Suponga que dos personas de esta población se casan. ¿Cuál es la probabilidad de que ambos sean Rh negativo, lo cual hace inevitable que sus hijos sean Rh negativo?

Solución:

Corresponde a una variable Binomial con $p = 0.15$, $n = 2$

$$P(X > 0) = 1 - P(X = 0) = 1 - C_{2,0} 0.15^0 (1 - 0.15)^2 = 1 - 0.85^2 = 0.2775$$

■

Ejercicio 3.10 En una gaveta de un laboratorio de química se tienen 20 frascos sin etiquetar con disoluciones contaminadas con bacterias. Se sabe que 8 de los frascos están contaminados con la bacteria A, 7 con la bacteria B y 5 con la bacteria C. Si se toman toman 3 frascos al azar, ¿cuál es la probabilidad de qué...

a) 2 frascos contengan la bacteria A?
b) ninguno de los frascos contenga la bacteria A?
c) al menos 1 tenga la bacteria A?
d) Determine la esperanza y la varianza del experimento.

Solución:

Establezcamos que el éxito es que el frasco seleccionado contenga la bacteria A, de esta forma: $X \sim H(n,N,k) = H(3,20,8)$ con función de masa:

$$f(x) = \begin{cases} \dfrac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} = \dfrac{\binom{3}{x}\binom{17}{3-x}}{\binom{20}{3}} & \text{si} \quad x = 0,1,2,3 \\ 0 & \text{en cualquier otro caso} \end{cases}$$

a) $P(x = 2) = f(2) = \dfrac{\binom{3}{2}\binom{17}{1}}{\binom{20}{3}} = 0.2947$
b) $P(x = 0) = f(0) = 0.193$
c) $P(x \geq 1) = 1 - P(x < 1) = 1 - F(0) = 0.807$
d)

$$E(X) = \mu = n\frac{k}{N} = 3\left(\frac{8}{20}\right) = \frac{24}{20} = \frac{6}{5} = 1.2$$

$$\text{Var}(X)\sigma^2 = n\left(\frac{k}{N}\right)\left(1 - \frac{k}{N}\right)\left(\frac{N-n}{N-1}\right) = 3\left(\frac{8}{20}\right)\left(1 - \frac{8}{20}\right)\left(\frac{20-3}{20-1}\right) = 0.6442$$

■

Ejercicio 3.11 Una empresa compra ciertos componentes en cajas de 40 unidades. Cada caja se inspecciona probando 5 componentes y si ninguno está defectuoso se acepta toda la caja. Supongamos que les llega una caja que, sin saberlo, contiene 3 componentes defectuosos.

a) ¿Cuál es la probabilidad de que se acepte la caja?
b) Calcule los valores de la función de probabilidad para el número de componentes defectuosos en muestras tomadas de esta caja.

Solución:

a) Se trata de un muestreo sin reemplazo que sigue una distribución hipergeométrica, cuya función de masa de probabilidad es

$$f(x) = P(X = x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}},$$

donde N = Número de elementos en la población (40), K = Número de éxitos en la población (3), n = tamaño de la muestra (5) y x = número de éxitos en la muestra. Por lo tanto,

$$P(\text{se acepte la caja}) = P(X = 0) = \frac{\binom{37}{5}\binom{3}{0}}{\binom{40}{5}} = \frac{37! \cdot 35!}{(40! \cdot 32!)} = 0.6624$$

b) Como solo hay tres defectuosos en la caja, el número de defectuosos en la muestra solo puede tomar valores menores o iguales a 3. Entonces, además de $P(X = 0)$, es necesario calcular las siguientes probabilidades:

$$P(X = 1) = 0.3011$$

$$P(X = 2) = 0.0354$$

$$P(X = 3) = 0.0010$$

Resumiendo:

x	0	1	2	3
$f(x)$	0.662	0.301	0.035	0.001

■

Ejercicio 3.12 En un hospital se realizan pruebas para detectar una enfermedad que afecta a 2 personas de cada 10 en la población. Suponga que la eficacia de la prueba para detectar la enfermedad es del 100 %.

a) ¿Cuál es la probabilidad de que la persona 5 a la que se le realiza la prueba sea la tercera en dar positivo?
b) ¿Cuál es la probabilidad de que se realicen al menos 8 pruebas para detectar a 3 personas con la enfermedad?
c) Calcule la varianza y la esperanza de este experimento aleatorio.

Solución:

Establezcamos que el éxito es que la prueba de positivo y que $X =$ el número de fracasos antes de obtener el tercer éxito éxito. De esta forma $X \sim BN(k,p) = BN(3,0.2)$. La función de masa de la distribución es entonces:

$$f(x) = \begin{cases} \binom{x+k-1}{k-1} p^k q^x = \binom{x+2}{2}(0.2)^3(0.8)^x & \text{si} \quad x = 0,1,2,3,\dots \\ 0 & \text{en cualquier otro caso} \end{cases}$$

a) Para tener 3 éxitos en 5 ensayos deben ocurrir 2 fracasos antes de ver el tercer éxito, por lo que:

$$P(x=2) = \binom{4}{2}(0.2)^3(0.8)^2 = 0.0307$$

b) Para tener 3 éxitos en al menos 8 ensayos, se deben tener al menos 5 fracasos antes de ver el tercer éxito, por lo que:

$$P(x \geq 5) = 1 - P(x \leq 4) = 1 - F(4) = 1 - [f(0) + f(1) + f(2) + f(3) + f(4)] = 0.852$$

c)

$$E(X) = \mu = \frac{kq}{p} = \frac{3 * 0.8}{0.2} = 12$$

$$\text{Var} = \sigma^2 = \frac{kq}{p^2} = \frac{3 * 0.8}{0.2^2} = 60$$

■

Ejercicio 3.13 Una persona lanzará un dado y una moneda hasta que el resultado sea sol en la moneda y en el dado caiga un valor mayor que 3.

a) ¿Cuál es la probabilidad de que realice 6 lanzamientos?
b) ¿Cuál es la probabilidad de que realice 6 lanzamientos a lo mucho?
c) Calcule la varianza y la esperanza de este experimento aleatorio.

Solución:

Para determinar la probabilidad de éxito para este experimento, establezcamos que: el evento A = se obtiene sol en la moneda y el evento B = se obtiene un número mayor que 3 en el dado. Entonces $P(A) = \frac{1}{2}$ y $P(B) = \frac{3}{6} = \frac{1}{2}$ (note que el dado debe caer de 4 en adelante para que se cumpla la condición). Los eventos son independientes por lo que: $p = P(A \cap B) = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{4} = 0.25$. Entonces, $X \sim G(0.25)$, donde la x mide el número de fracasos antes de ver el primer éxito. La función de masa es:

$$f(x) = \begin{cases} pq^x = (0.25)(0.75)^x & \text{si} \quad x = 0, 1, 2, 3, \ldots \\ 0 & \text{en cualquier otro caso} \end{cases}$$

a) Para que se realicen 6 lanzamientos antes del primer y único éxito deben ocurrir 5 fracasos, por lo que:

$$P(x = 5) = f(5) = (0.25)(0.75)^5 = 0.0593$$

b) Para que se realicen a lo más 6 lanzamientos antes del primer y único éxito deben ocurrir a lo más 5 fracasos, por lo que:

$$P(x \leq 5) = F(5) = f(0) + f(1) + f(2) + f(3) + f(4) + f(5) = 0.2373$$

c)

$$E(X) = \mu = \frac{q}{p} = \frac{0.75}{0.25} = 3$$

$$\text{Var} = \sigma^2 = \frac{q}{p^2} = \frac{0.75}{0.25^2} = 12$$

■

Ejercicio 3.14 El flujo sanguíneo en el cerebro de personas sanas está normalmente distribuido con una media de 74 y una desviación estándar de 16.

a) ¿Qué proporción de personas sanas tendrán un flujo sanguíneo entre 60 y 80?
b) ¿Qué proporción de personas sanas tendrán un flujo sanguíneo por arriba de 100?
c) Si una persona tiene un flujo sanguíneo por debajo de 40, es considerada como en riesgo de sufrir un accidente cerebrovascular. ¿Cuál es la proporción de personas sanas que se les diagnosticará erróneamente como “en riesgo”?

Solución:

X = flujo sanguíneo $\sim$ Normal$(\mu = 74, \sigma = 16)$. Sea $Z = \frac{(X-\mu)}{\sigma}$, entonces:

a) $P(60 < X < 80) = P(X \leq 80) - P(X \leq 60) = P\left(Z \leq \frac{80-74}{16}\right) - P\left(Z \leq \frac{60-74}{16}\right)$
$= P(Z \leq 0.375) - P(Z \leq -0.875)$
y usando las tablas de la Normal(0,1) obtenemos
$P(60 < X < 80) = 0.6461 - 0.1908 = 0.4553$

b) $P(X > 100) = 1 - P(X \leq 100) = 1 - P\left(Z \leq \frac{100-74}{16}\right) = 1 - P(Z \leq 1.625)$
y usando las tablas de la Normal(0,1) obtenemos
$P(X > 100) = 1 - 0.9479 = 0.0521$

c) $P(X < 40) = P\left(Z \leq \frac{40-74}{16}\right) = P(Z \leq -2.125)$
y usando las tablas de la Normal(0,1) obtenemos
$P(X < 40) = 0.0168$

■

Ejercicio 3.15 Un estudio muestra que cuando ocho personas ocupan un elevador, la distribución de probabilidad del peso total tiene una distribución normal con una media igual a 1200 libras y una desviación estándar de 99 libras. ¿Cuál es la probabilidad de que el peso total de ocho personas exceda de 1300 libras? ¿Y de 1500 libras?

Solución:

$X =$ "Peso de ocho" $\sim$ Normal$(\mu = 1200, \sigma = 99)$. Sea $Z = \frac{(X-\mu)}{\sigma}$, entonces:

a) $P(X > 1300) = 1 - P(X \leq 1300) = 1 - P\left(Z \leq \frac{1300-1200}{99}\right) = 1 - P(Z \leq 1.01)$
y usando las tablas de la Normal(0,1) obtenemos
$P(X > 1300) = 1 - 0.8438 = 0.1562$

b) $P(X > 1500) = 1 - P(X \leq 1500) = 1 - P\left(Z \leq \frac{1500-1200}{99}\right) = 1 - P(Z \leq 3.03)$
y usando las tablas de la Normal(0,1) obtenemos
$P(X > 1500) = 1 - 0.9988 = 0.0012$

■

Ejercicio 3.16 Una mina de fosfato descarga sólidos suspendidos todos los días, con una descarga media diaria de 27 miligramos por litro (mg/l) y una desviación estándar de 14 mg/l. ¿Qué proporción de días superará los 50 mg/l la descarga diaria?

Solución:

$X =$ descarga $\sim$ Normal$(\mu = 27, \sigma = 14)$, $Z = \frac{(X-\mu)}{\sigma}$, entonces:

$P(X > 50) = 1 - P(X \leq 50) = 1 - P\left(Z \leq \frac{50-27}{14}\right) = 1 - P(Z \leq 1.64)$

y usando las tablas de la Normal(0,1) obtenemos

$P(X > 50) = 1 - 0.9495 = 0.0505$ ■

Ejercicio 3.17 El número de cierto tipo de bacterias en muestras de 1 mililitro de agua potable tiende a estar normalmente distribuido con media de 85 y desviación estándar de 9. ¿Cuál es la probabilidad de que una muestra determinada de 1 ml de agua potable contenga más de 100 bacterias?

Solución:

$X =$ núm. de bacterias $\sim$ Normal$(\mu = 85, \sigma = 9)$. Sea $Z = \frac{(X-\mu)}{\sigma}$, entonces:

$P(X > 100) = 1 - P(X \leq 100) = 1 - P\left(Z \leq \frac{100-85}{9}\right) = 1 - P(Z \leq 1.66)$
y usando las tablas de la Normal(0,1) obtenemos
$P(X > 100) = 1 - 0.952 = 0.048$ ∎

Ejercicio 3.18 Se estima que la vida media de las barrenas que perforan pozos petroleros es de 75 horas. Suponga que una compañía de exploración petrolera compra barrenas cuya vida útil que está aproximadamente normalmente distribuida, con una media igual a 75 horas y una desviación estándar igual a 12 horas.

a) ¿Qué proporción de barrenas fallarán antes de 60 horas?
b) ¿Qué proporción durará por lo menos 60 horas?
c) ¿Qué proporción fallará después de más de 90 horas de uso?

Solución:

$X =$ vida barrenas $\sim$ Normal$(\mu = 75, \sigma = 12)$. Sea $Z = \frac{(X-\mu)}{\sigma}$, entonces:

a) $P(X < 60) = P\left(Z < \frac{60-75}{12}\right) = P(Z < -1.25)$
y usando las tablas de la Normal(0,1) obtenemos
$P(X < 60) = 0.1056$
b) $P(X \geq 60) = 1 - P(X < 60) = 1 - 0.1056 = 0.8944$
c) $P(X > 90) = 1 - P(X \leq 90) = 1 - P\left(Z < \frac{90-75}{12}\right) = 1 - P(Z < 1.25)$
y usando las tablas de la Normal(0,1) obtenemos
$P(X > 90) = 1 - 0.8944 = 0.1056\ldots$ ver (a)

∎

Ejercicio 3.19 Una empresa utiliza focos cuyas duraciones están exponencialmente distribuidas, con tiempo medio de operación $\mu = 5$ años.

a) ¿Cual es la probabilidad de que un foco siga funcionando después de 8 años?
b) ¿A los cuantos años la mitad de los focos seguirá funcionando?

Solución:

a) Sea X = tiempo de vida con $X \sim$ Exponencial $(\mu = 5)$. Es decir:
$f(x) = e^{-x/\mu}/\mu$, que tiene una $F(x) = P(X < x) = 1 - e^{-x/\mu}$
Por lo tanto
$P(X > 8) = 1 - P(X < 8) = 1 - (1 - e^{-8/5}) = e^{-1.6} = 0.20$

b) La incógnita es el valor c para el cual $P(X < c) = 0.5$
Pero, para una variable exponencial con media 5, $P(X < c) = 1 - e^{-c/5} = 0.5$
Entonces $(-c/5)) = ln(0.5) = 0.6931$ y por lo tanto $c = 3.4657$.
Es decir, aproximadamente a los 3 años, 5 meses y 17 dias la mitad de los focos seguirán funcionando

■

Ejercicio 3.20 El suministro de agua, en millones de litros por día, de una zona industrial sigue una distribución exponencial con parámetro $\lambda = 0.2$

a) La zona industrial requiere un millón de litros por día para funcionar adecuadamente, ¿cuál es la probabilidad de que en un día el abastecimiento de agua sea insuficiente?
b) ¿Cuál es la probabilidad de que en un día cualquiera el suministro de agua supere los 3 millones de litros?
c) Calcule la esperanza y la varianza.

Solución:

$X \sim E(\lambda) = E(0.2)$ con función de densidad:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si} \quad x > 0 \\ 0 & \text{en cualquier otro caso} \end{cases}$$

a) $P(x < 1) = F(1) = \displaystyle\int_0^1 0.2e^{-0.2x}\,dx = 0.1813$

b) $P(x \geq 3) = 1 - P(x < 3) = 1 - \displaystyle\int_0^3 0.2e^{-0.2x}\,dx = 0.5488$

c) $E(X) = \mu = \dfrac{1}{\lambda} = \dfrac{1}{0.2} = 5$

$\text{Var}(X) = \sigma^2 = \dfrac{1}{\lambda^2} = \dfrac{1}{0.2^2} = 25$

■

Ejercicio 3.21 La duración, en meses, de un determinado componente electrónico de un espectrómetro de masa sigue una distribución gamma con media igual a 24 meses y desviación estándar de 12 semanas.

a) Determine los parámetros α y β .
b) ¿Cuál es la probabilidad de que un componente electrónico dure al menos 20 meses?
c) ¿Cuál es la probabilidad de que dure a lo mucho 30 meses?
d) ¿Hasta qué tiempo de duración se encontrará al 90 % de estos componentes?

Solución:

a) Dado que la distribución tiene $E(X) = \mu = \alpha\beta$ y $\text{Var}(X) = \sigma^2 = \alpha\beta^2$ se tiene el siguiente sistema de ecuaciones:

$$\begin{cases} \alpha\beta & = & 24 \\ \alpha\beta^2 & = & 144 \end{cases}$$

Si se divide la segunda ecuación entre la primera: $\dfrac{\alpha\beta^2}{\alpha\beta} = \dfrac{144}{24} \Rightarrow \beta = 6$

Se sustituye este valor en la primera ecuación: $6\alpha = 24 \Rightarrow \alpha = 4$

Con estos valores la función de densidad queda de la siguiente forma:

$$f(x) = \begin{cases} \dfrac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} = \dfrac{1}{6^4 3!} x^3 e^{-\frac{x}{6}} & \text{si} \quad x > 0 \\ 0 & \text{en cualquier otro caso} \end{cases}$$

b) $P(x \geq 20) = 1 - P(x < 20) = 1 - F(20) = 1 - \displaystyle\int_0^{20} \frac{1}{6^5} x^3 e^{-\frac{x}{6}}\, dx = 0.573$

c) $P(x < 30) = F(30) = \displaystyle\int_0^{30} \frac{1}{6^5} x^3 e^{-\frac{x}{6}}\, dx = 0.735$

d) $P(x < a) = F(a) = \displaystyle\int_0^{a} \frac{1}{6^5} x^3 e^{-\frac{x}{6}}\, dx = 0.9$

$-\dfrac{e^{-\frac{a}{6}} a^3}{1296} - \dfrac{1}{72} e^{-\frac{a}{6}} a^2 - \dfrac{1}{6} e^{-\frac{a}{6}} a - e^{-\frac{a}{6}} + 1 = 0.9$ se resuelve para a:

$a = 40.0847$

un método alternativo es usar la función gamma incompleta:

$$P(x < a) = \int_0^{x=a/\beta} \frac{1}{\Gamma(\alpha)} y^{\alpha-1} e^{-y}\, dy$$

En las tablas de gamma incompleta se lee el valor de x en la columna para $\alpha = 4$ en el que la probabilidad sea lo más cercana posible a 0.9. En las tablas para una probabilidad de 0.918 se lee $x = 7$ por lo que:

$$x = \frac{a}{\beta} \Rightarrow 7 = \frac{a}{6} \Rightarrow a = 42$$

∎

4. Muestreo y distribuciones muestrales

Ejercicio 4.1 Se selecciona una muestra aleatoria simple de 25 observaciones de una población distribuida normalmente con media 106 y desviación estándar 12.

a) Calcule la media y la desviación estándar de la distribución muestral de la media muestral $\bar{X}$.
b) Calcule la probabilidad de que $\bar{X}$ exceda de 110.
c) Calcule la probabilidad de que la media muestral se aleje de la media poblacional μ en no más de 4.

Solución:

$X \sim \text{Normal}(\mu = 106, \sigma = 12)$, $n = 25$.

a) $\bar{X} \sim \text{Normal}(\mu = 106, \sigma = \frac{12}{5}) \sim \text{Normal}(\mu = 106, \sigma = 2.4)$
b) Sea $Z = \frac{(\bar{X}-\mu)}{\sigma}$, entonces:
$P(\bar{X} > 110) = 1 - P(\bar{X} < 110) = 1 - P\left(Z < \frac{110-106}{2.4}\right) = 1 - P(Z < 1.666)$
y usando las tablas de la Normal(0,1) obtenemos: $P(\bar{X} > 110) = 1 - 0.9515 = 0.0485$
c) $P(102 < \bar{X} < 110) = P(\bar{X} < 110) - P(\bar{X} < 102) = P\left(Z < \frac{110-106}{2.4}\right) - P\left(Z < \frac{102-106}{2.4}\right) =$
$= P(Z < 1.666) - P(Z < -1.666)$
y usando las tablas de la Normal(0,1) obtenemos:
$P(102 < \bar{X} < 110) = 0.952 - 0.048 = 0.9040$

■

Ejercicio 4.2 Suponga que la distribución de potasio en los plátanos de cierto supermercado está distribuida normalmente, con media igual a 422 mg y desviación estándar de 13 mg por plátano. Si una persona come $n = 3$ plátanos al día y T es el potasio total en miligramos de estos tres plátanos.

a) Encuentre la media y la desviación estándar de T.
b) Encuentre la probabilidad de que la ingesta diaria de potasio de los tres plátanos sea mayor a 1300 mg. (SUGERENCIA: Observe que T es igual a la suma de tres variables aleatorias, X_1, X_2 y X_3, donde X_i es el potasio contenido en el plátano i)

Solución:

$X \sim \text{Normal}(\mu = 422, \sigma = 13)$, $n = 3$. Entonces:

$X_1 + X_2 + X_3 = T \sim \text{Normal}(\mu = 422 \times 3, \sigma = \sqrt{3} \times 13) \sim \text{Normal}(\mu = 1266, \sigma = 22.51)$

Por lo tanto, usando $Z = \frac{(X-\mu)}{\sigma}$, tenemos:

$P(T > 1300) = 1 - P(T \leq 1300) = 1 - P\left(Z < \frac{1300-1266}{22.51}\right) = 1 - P(Z < 1.51)$
y usando las tablas de la Normal(0,1) obtenemos

$P(T > 1300) = 1 - 0.9345 = 0.0655$ ∎

Ejercicio 4.3 Un vendedor afirma que el porcentaje de dulces rojos en un paquete de dulces multicolores es 13%. Suponga que se toma al azar un paquete de estos dulces que contiene 55 piezas y se determina la proporción de dulces rojos del paquete.

a) ¿Cuál será la distribución aproximada de la proporción muestral de dulces rojos en un paquete que contiene 55 dulces?
b) ¿Calcule la probabilidad de que el porcentaje muestral de dulces rojos sea menor que 20%?
c) ¿Calcule la probabilidad de que el porcentaje muestral sea mayor a 35%?

Solución:

a) $n = 55$, $\hat{p} =$ prop. observada $= 0.13$, $\hat{q} = 0.87$. Como $np = 55 \times 0.13 = 7.15 > 5$, entonces $X =$ Número de éxitos en n intentos será aproximadamente $\text{Normal}(\mu = np, \sigma = \sqrt{npq})$ y la proporción muestral $\hat{p}$ será:

   $\hat{p} \sim \text{Normal}(\mu = p, \sigma = \sqrt{pq/n})$, es decir:

   $\hat{p} \sim$ aprox. $\text{Normal}(\mu = 0.13, \sigma = 0.045)$
b) Utilizando $Z = \frac{(X-\mu)}{\sigma}$, tenemos que $P(\hat{p} < 0.2) = P\left(Z < \frac{0.20-0.13}{0.045}\right) = P(Z < 1.55)$ y usando las tablas de la Normal(0,1), encontramos $P(Z < 1.55) = 0.9394$.

Es decir $P(\hat{p} < 0.2) = \mathbf{0.9394}$.

c) De la misma manera, $P(\hat{p} > 0.35) = 1 - P(\hat{p} < 0.35) = 1 - P\left(Z < \frac{0.35-0.13}{0.045}\right) = 1 - P(Z < 4.88) \approx 1 - 1 \approx 0$.

■

5. Estimación y pruebas de hipótesis

Ejercicio 5.1 Los contenidos de una muestra aleatoria de 5 bolsas de cacahuates japoneses de una cierta marca dieron los siguientes pesos netos (en gramos):

$$180 \quad 190 \quad 185 \quad 175 \quad 184$$

a) Calcule un Intervalo de Confianza del 95 % para la media de los contenidos en todas las bolsas de cacahuates japoneses de esa marca.
b) ¿Con qué nivel de confianza estaremos estimando el contenido medio de las bolsas de cacahuates, si los límites de confianza son 177.4 y 188.2? Suponga una distribución Normal.

Solución:

a) Cuando la distribución de la variable aleatoria es Normal, la expresión para calcular el intervalo de confianza para la media es:

$$IC95\,\% \text{ para } \mu = \bar{x} \pm \text{MOE},$$

donde el margen de error

$$\text{MOE} = t_{1-\alpha/2}(\text{gl}) * \frac{s}{\sqrt{n}}$$

En este caso, $n = 5$, $\bar{x} = 182.8$, $s = 5.63$, y gl = 4. En las tablas de la t de Student encontramos el valor crítico $t_{1-\alpha/2}(4) = 2.776$, por lo que:

$$\text{MOE} = 2.776 * \frac{5.63}{\sqrt{5}} = 6.989$$

Entonces, el intervalo de confianza del 95 % es:

$$IC95\,\% \text{ para } \mu = 182.8 \pm 6.99 \quad o \quad (175.8, 189.8)$$

b) Sabemos que el IC(1-α)% para μ es $(177.4, 188.2)$, pero desconocemos α.
Sabemos que el ancho del intervalo de confianza es 2 veces el MOE, por lo que:

$$188.2 - 177.4 = 5.4 = 2 * \text{MOE} \quad \Rightarrow \quad \text{MOE} = 2.7 = t_{1-\alpha/2}(\text{gl}) * \frac{s}{\sqrt{n}}.$$

Despejando $t_{1-\alpha/2}(\text{gl})$ encontramos que:

$$t_{1-\alpha/2}(\text{gl}) = \frac{\sqrt{5} * 5.4}{5.63} = 2.14$$

Buscamos en la tabla de la t de Student con 4 grados de libertad y encontramos que para el valor α en las dos colas de 10 %, el valor crítico correspondiente es 2.13, por lo que concluimos que el intervalo $(177.4, 188.2)$ es de 90 % de confianza.

■

Ejercicio 5.2 Del ejemplo anterior, ¿cuántas bolsas se deberán tomar de muestra para que el Intervalo de Confianza de 95 % para la media tenga un margen de error de estimación de 0.8?

Solución:

La expresión para calcular el tamaño de muestra correspondiente a un margen de error y un nivel de confianza, es el mínimo valor entero que cumpla con:

$$n \geq \left(\frac{z_{1-\alpha/2} \cdot \sigma}{\text{MOE}} \right)^2,$$

En este caso no conocemos σ, pero podemos aproximarlo por el valor calculado para $\hat{\sigma} = s = 5.63$. Por lo tanto

$$n \geq \left(\frac{1.96 * 5.63}{0.8} \right)^2 = 190.25,$$

es decir, $n_{\text{mín}} = 191$.

■

Ejercicio 5.3 Una máquina produce tapones de goma. Para estimar la variabilidad de los diámetros, se toma una muestra aleatoria de 10 tapones producidos por la máquina encontrando los siguientes diámetros en milímetros:

10.1 9.7 10.3 10.4 9.9 9.8 9.9 10.1 10.3 9.9

Calcule un intervalo de confianza del 95 % para la desviación estándar de los diámetros de todos los tapones producidos por la máquina. Suponga que los diámetros de las piezas se distribuyen como una Normal.

Solución:

La expresión para calcular el $IC95\,\%$ para σ^2 es:

$$\left(\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}, \frac{(n-1)s^2}{\chi^2_{\alpha/2}}\right),$$

donde $\chi^2_{\alpha/2}$ y $\chi^2_{1-\alpha/2}$ son los percentiles apropiados de la distribución Chi cuadrada con $(n-1)$ grados de libertad.

En este caso, $s = 0.237$, $n = 10$, y para 9 grados de libertad encontramos en las tablas de la distribución Chi cuadrada los valores:

$$\chi^2_{(0.05/2)} = 19.02 \quad \text{y} \quad \chi^2_{(1-0.05/2)} = 2.7$$

Por lo tanto, el intervalo de confianza para σ^2 es:

$$IC95\,\% \text{ para } \sigma^2 = \left(\tfrac{9*0.056}{19.02}, \tfrac{9*0.056}{2.7}\right) = (0.026, 0.187),$$

y para la desviación estándar σ, tendremos el intervalo de confianza:

$$IC95\,\% \text{ para } \sigma = \left(\sqrt{0.026}, \sqrt{0.187}\right) = (0.1627, 0.4319).$$

■

Ejercicio 5.4 Se sabe que la desviación estándar de la vida de cierta cepa de ratones de laboratorio es de 5.8 meses y su distribución es aproximadamente Normal. ¿De qué tamaño debo elegir una muestra de ratones para tener un 99 % de confianza de que la vida media esperada de la muestra esté dentro de dos meses a partir de la vida media de toda la población de esa cepa de ratones?

Solución:

La expresión para calcular el tamaño de muestra correspondiente a un margen de error y un nivel de confianza es el mínimo valor entero que cumpla con:

$$n \geq \left(\frac{z_{1-\alpha/2} \cdot \sigma}{\text{MOE}}\right)^2,$$

por lo que:

$$n \geq \left(\frac{2.575 * 5.8}{2}\right)^2 = 55.7,$$

es decir, $n_{\text{mín}} = 56$ ratones. ▪

Ejercicio 5.5 Se realizó una encuesta de salida a una muestra aleatoria de 600 electores en una votación y se encontró que 240 votaron a favor del candidato A.

a) Utilizando un Intervalo de Confianza del 95 %, estimar el porcentaje de electores a favor del candidato A en toda la población.
b) ¿Qué tan grande tiene que ser el tamaño de muestra si se desea tener una confianza de 94 % de que el error de estimación no supere el 2 %?

Solución:

a) Se trata de una distribución Binomial con $X =$ Núm. de votos a favor de A $=$ Núm. de éxitos $= 240$ y $n =$ Núm. de votantes $= 600$. Siempre que $np > 5$ y $nq > 5$ la expresión para calcular el intervalo es:

$$IC95\,\% \text{ para } p = \left(\hat{p} \pm z_{1-\alpha/2} * \sqrt{\frac{\hat{p}\hat{q}}{n}}\right).$$

En este caso:

$$IC95\,\% \text{ para } p = 0.4 \pm 1.96 * \sqrt{\frac{0.4 * 0.6}{600}} = 0.4 \pm 0.0392 \quad o \quad (0.36, 0.44).$$

b) La expresión para calcular el tamaño de muestra correspondiente a un margen de error y un nivel de confianza para la estimación de p es el mínimo entero que cumpla con:

$$n \geq \left(\frac{z_{1-\alpha/2} * \sigma}{\text{MOE}}\right)^2,$$

En este caso, aproximamos σ por su estimador $s = \sqrt{\hat{p}\hat{q}} = \sqrt{(0.4 * 0.6)} = 0.49$.

Con un nivel de confianza de 94 %, el valor crítico es $z_{0.97} = 1.88$. Entonces:

$$n \geq \left(\frac{1.88 \times 0.49}{0.02}\right)^2 = 2120.6,$$

es decir, $n_{\text{mín}} = 2121$.

■

Ejercicio 5.6 Un estudio acerca de un remedio homeopático en 15 pacientes mostró los siguientes tiempos en aliviar su dolor de cabeza desde la toma del remedio. Construya un Intervalo de Confianza 95 % para el tiempo promedio de respuesta al remedio suponiendo que el tiempo de respuesta sigue una distribución Normal.

8.6, 9.4, 7.9, 6.8, 8.3, 7.3, 9.2, 9.6, 8.7, 11.4, 10.3, 5.4, 8.1, 5.5, 6.9

Solución:

La expresión para calcular el $IC95\,\%$ para μ es $\bar{x} \pm \text{MOE}$, con $\text{MOE} = t_{1-\alpha/2}(\text{gl}) \cdot s/\sqrt{n}$.

En este caso, tenemos que $\bar{x} = 8.226$, $s = 1.672$ y $n = 15$. En las tablas de la t de Student encontramos el valor crítico $t_{0.025}(14) = 2.145$, por lo que:

$$IC95\,\% \text{ para } \mu = 8.226 \pm 2.145 * \frac{1.672}{\sqrt{15}} = 8.226 \pm 0.926 \quad o \quad (7.30, 9.15)$$ ■

Ejercicio 5.7 En un estudio piloto, se midió la cantidad en miligramos de principio activo contenido en un fármaco producido mediante dos procesos químicos distintos. Los datos se muestran en la siguiente tabla:

Proceso A	Proceso B
$n = 50$	$n = 45$
$\bar{x} = 39.7$	$\bar{x} = 41.2$
$s^2 = 1.8$	$s^2 = 2.2$

Calcule un intervalo de confianza para $\mu_1 - \mu_2$ al 95 % de confianza suponiendo que las varianzas poblaciones se pueden aproximar mediante las varianzas muestrales.

Solución:

En este caso, debido a que ambas muestras son mayores a 30, se puede invocar al Teorema Central del Límite y suponer normalidad en los datos. Por la misma razón las varianzas poblacionales se pueden aproximar mediante las varianzas muestrales, y entonces la expresión para calcular el $IC95\,\%$ para $(\mu_1 - \mu_2)$ es $(\overline{x_1} - \overline{x_2}) \pm \text{MOE}$, con $\text{MOE} = z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$.

En las tablas de la normal estándar encontramos el valor crítico $z_{0.025} = 1.96$, por lo que:

$$IC95\,\% \text{ para } (\mu_1 - \mu_2) = (39.7 - 41.2) \pm 1.96\sqrt{\frac{1.8}{50} + \frac{2.2}{45}} = -1.5 \pm 0.5711 \quad o \quad (-2.071, -0.929)$$

■

Ejercicio 5.8 Se realizó un estudio de calidad a dos marcas de vinos sobre el cumplimiento de unos determinados estándares de calidad. Se encontró que de la marca A, 525 de un total de 600 botellas cumplieron con los estándares, mientras que de la marca B, 712 de un total de 800 botellas cumplieron con los estándares. Calcule un intervalo de confianza para la diferencia entre las dos proporciones poblacionales que cumplen con los estándares al 90% de confianza.

Solución:

De acuerdo a los datos del problema:

$$\widehat{p_1} = \frac{525}{600} = 0.875 \qquad \widehat{q_1} = 0.125 \qquad \widehat{p_2} = \frac{712}{800} = 0.89 \qquad \widehat{q_2} = 0.11$$

La expresión para calcular el $IC90\%$ para $(p_1 - p_2)$ es $(\widehat{p_1} - \widehat{p_2}) \pm \text{MOE}$,

con $\text{MOE} = z_{\alpha/2}\sqrt{\dfrac{\widehat{p_1}\widehat{q_1}}{n_1} + \dfrac{\widehat{p_1}\widehat{q_1}}{n_2}}$.

En las tablas de la normal estándar encontramos el valor crítico $z_{0.05} = 1.645$, por lo que:

$$IC90\% \text{ para } (p_1 - p_2) = (0.875 - 0.89) \pm 1.645\sqrt{\frac{(0.875)(0.125)}{600} + \frac{(0.89)(0.11)}{800}}$$

$$= -0.015 \pm 0.0287 \quad \text{o} \quad (-0.0437, 0.0287)$$

■

Ejercicio 5.9 Un supermercado vende paquetes de tortillas que deben contener aproximadamente 1Kg. Una muestra de 35 paquetes al azar dio un promedio de 1.01 Kg. y una desviación estándar de 0.18 Kg.

a) Si desea asegurarse que la cantidad promedio de tortillas es en realidad de 1 Kg. ¿cuáles hipótesis probaría?
b) Encuentre el valor p para la prueba y efectúe con él la prueba de la parte a. Suponga Normalidad de los datos.

Solución:

a) H_0: $\mu = 1$ vs. H_a: $\mu \neq 1$
b) En este caso, $\bar{x} = 1.01$, $s = 0.18$ y $n = 35$, y como desconocemos la varianza poblacional, el estadístico de prueba es:

$$T = \frac{(\bar{x} - \mu)}{(s/\sqrt{n})} = \frac{(1.01 - 1)}{(0.18/\sqrt{35})} = 0.329$$

Como es una prueba de dos colas, el valor p correspondiente a 0.329 es $P(|T_{34}| > 0.329) =$

0.74, el cual es un valor p muy grande y no significativo, por lo que **no se rechaza** H_0. Es decir, no hay evidencia para afirmar que el peso promedio sea diferente de 1 Kg.

Alternativamente, también se puede probar la hipótesis comparando el valor del estadistico de prueba con los valores criticos correspondientes, digamos, a un $\alpha = 0.05$. Estos valores se obtienen de una tabla de la distribución t de Student con 34 grados de libertad, y son $t_{34,.025} = \pm 2.032$. Como $T = 0.329$ cae entre estos dos valores, **no se rechaza** H_0

■

Ejercicio 5.10 Un laboratorio farmacéutico declaró que la potencia media de cierto medicamento es de 80%. Se tomó una muestra aleatoria de 100 tabletas del medicamento y se obtuvo una media muestral de $\bar{x} = 79.7\,\%$ con una desviación estándar de $s = 0.8\,\%$. Utilizando $\alpha = 0.05$, ¿existe suficiente evidencia para rechazar lo dicho por el laboratorio?

a) ¿Cuál es la hipótesis nula a ser probada?
b) ¿Cuál es la hipótesis alternativa?
c) Lleve a cabo la prueba estadística y exprese su conclusión.

Solución:

a) y b) H_0: $\mu = 80$ vs. H_a: $\mu \neq 80$

c) En este caso $n = 100$, por lo que podemos suponer normalidad. Por otro lado, $\bar{x} = 79.7$ y $s = 0.8$, y como desconocemos la varianza poblacional, el estadístico de prueba es:

$$T = \frac{(\bar{x} - \mu)}{(s/\sqrt{n})} = \frac{(79.7 - 80)}{0.8/\sqrt{100}} = -0.3/0.08 = -3.75$$

El valor crítico en tablas de la t de Student para $\alpha = 0.05$ es $t_{99,.025} = \pm 1.99$. Por lo tanto, se **rechaza** H_0.

■

Ejercicio 5.11 Una agencia supervisora del gobierno requiere que las aguas de un río contengan un mínimo de 5 partes por millón de oxígeno disuelto para que el río se considere seguro para la vida. Se tomaron 6 muestras de agua en un lugar específico del río que contenían 4.9, 5.1, 4.9, 5.0, 5.0 y 4.7 de oxígeno disuelto. Utilizando $\alpha = 0.05$ y suponiendo que los contenidos se distribuyen Normalmente ¿los datos nos dan suficiente evidencia para decidir que el contenido de oxígeno es menor al mínimo?

Solución:

H_0: $\mu = 5$ vs. H_a: $\mu < 5$

Para estos datos, $\bar{x} = 4.933$, $s = 0.1366$, $n = 6$, y como desconocemos la varianza poblacional, el estadístico de prueba es:

$$T = \frac{(\bar{x}-\mu)}{(s/\sqrt{n})} = \frac{(4.933-5)}{0.1366/\sqrt{6}} = \frac{-0.067}{0.0558} = -1.194$$

Por otro lado, como se trata de una prueba de una sola cola el valor crítico en tablas de la t de Student es $t_{5,.975} = -2.015$. Y como $-2.015 < -1.194$, **no se rechaza** H_0. ▪

Ejercicio 5.12 Un fabricante de resistencias quiere controlar la variabilidad en la duración de vida útil de sus resistencias de tal manera que σ sea menor que 150 horas. Para probarlo, tomó una muestra de 20 resistencias y midió su vida útil en horas, dando los siguientes resultados. Con un nivel de significancia de $\alpha = 0.01$, ¿existe suficiente evidencia para indicar que el fabricante alcanzó su objetivo? Suponga que los datos siguen una distribución Normal.

2100	2302	1951	2067	2415	1883	2101	2146	2278	2019
1924	2183	2077	2392	2286	2501	1946	2161	2253	1827

Solución:

H_0: $\sigma = 150$ vs. H_a: $\sigma < 150$, es equivalente a:

H_0: $\sigma^2 = 22500$ vs. H_a: $\sigma^2 < 22500$

Se trata de una prueba de 1 cola con $n = 20$, $s^2 = 34854.3$, $\alpha = 0.01$, y el estadístico de prueba apropiado es:

$$\chi^2 = \frac{(n-1)\cdot s^2}{\sigma_o^2} = \frac{19\cdot 34854.3}{22500} = 29.43$$

El valor crítico en tablas de la Chi cuadrada es $\chi^2_{(19,0.01)} = 36.19$, y como $29.43 < 36.19$, **no se rechaza** H_0. ▪

Ejercicio 5.13 Se tomaron al azar 80 mediciones independientes de dos poblaciones 1 y 2. A continuación vemos un resumen de los datos muestrales:

	Muestra 1	Muestra 2
Tamaño muestral	80	80
Media muestral	11.6	9.7
Varianza muestral	27.9	38.4

a) ¿Cuáles serían las hipótesis nula y alternativa si su objetivo es demostrar que μ_1 es mayor que μ_2?
b) ¿Se trata de una prueba de una o de dos colas?
c) Calcule el estadístico de prueba y encuentre el valor-p. ¿Existe una diferencia significativa en las medias poblacionales al nivel de significancia de 1 %?
d) Encuentre ahora el valor crítico y la región de rechazo cuando $\alpha = 0.01$. ¿Existe suficiente

evidencia para pensar que existe una diferencia en las medias poblacionales?

Solución:

a) H_0: $\mu_1 = \mu_2$ vs. H_a: $\mu_1 > \mu_2$, o de manera equivalente:

$$H_0 : \mu_1 - \mu_2 = 0 \quad \text{vs.} \quad H_a : \mu_1 - \mu_2 > 0$$

b) Es una prueba de 1 cola.

c) Se trata de dos poblaciones independientes de cuyas varianzas no sabemos nada, pero el tamaño de muestra es tan grande ($160 \gg 30$) que el estadístico de prueba

$$T = (\bar{x}_1 - \bar{x}_2) / \sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}$$

se puede aproximar por una Normal $(0,1)$.
En nuestro caso:

$$T = \frac{(11.6 - 9.7)}{\sqrt{\left(\frac{27.9}{80} + \frac{38.4}{80}\right)}} = \frac{1.9}{\sqrt{\frac{27.9+38.4}{80}}} = \frac{1.9}{\sqrt{\frac{66.3}{80}}} = \frac{1.9}{0.9111} = 2.08$$

Consultando una tabla de la Normal$(0,1)$, encontramos que:

valor-$p(2.08) = P(T > 2.08) = 0.0188$

El cual no es significativo al nivel de $\alpha = 0.01$. Es decir, **no se rechaza** H_0. No existe una diferencia significativa entre las medias poblacionales al nivel de significancia de 1 %.

d) Alternativamente, la hipótesis también se puede probar comparando el valor del estadístico de prueba con el valor crítico para $\alpha = 0.01$ que se encuentra en tablas de la Normal$(0,1)$, el cual es $z_{0.01} = 2.33$. Como $T = 2.08$ es menor que 2.33, **no rechazamos** H_0. Es decir, no existe suficiente evidencia para afirmar que existe una diferencia en las medias poblacionales entre las dos muestras.

■

Ejercicio 5.14 Se llevó a cabo un experimento para comparar el tiempo medio necesario para recuperarse de un resfriado común en personas adultas que tomaron un suplemento diario de 4 mg de vitamina C, en comparación con aquellas personas adultas que no tomaron dicho suplemento. Se seleccionaron 35 personas adultas al azar para cada uno de los dos grupos. Los tiempos de recuperación promedio y las desviaciones estándar fueron:

	Sin suplemento	Con suplemento
Tamaño muestral	35	35
Media muestral	5.9	5.1
Varianza muestral	2.9	1.2

a) Si se desea demostrar que el uso de vitamina C reduce el tiempo necesario para recuperarse de un resfriado común, escriba las hipótesis nula y alternativa para la prueba.
b) ¿Se trata de una prueba de una o de dos colas?
c) Realice la prueba estadística de la hipótesis nula y exprese su conclusión utilizando $\alpha = 0.05$.

Solución:

a) H_0: $\mu_s = \mu_c$ vs. H_a: $\mu_s > \mu_c$ (Prueba de 1 cola).
b) Es una prueba de 1 cola.
c) No sabemos nada acerca de las varianzas por lo que se debe usar la prueba de Welch. El estadístico apropiado es:

$$T = (\bar{x}_1 - \bar{x}_2) / \sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}$$

que se comporta, cuando H_0 es cierta, como una t de Student con:

$$v = \left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2 / \left(\frac{\left(s_1^2/n_1)\right)^2}{n_1 - 1} + \frac{\left(s_2^2/n_2\right)^2}{n_2 - 1}\right) \quad \text{grados de libertad.}$$

En nuestro caso:

$$T = \frac{(5.9 - 5.1)}{\sqrt{\left(\frac{2.9}{35} + \frac{1.2}{35}\right)}} = \frac{0.8}{\sqrt{\left(\frac{2.9+1.2}{35}\right)}} = 2.074 \quad \text{y} \quad v = 58\,\text{gl}.$$

En tablas de la t de Student encontramos con $\alpha = 5\,\%$ que el valor crítico para $v = 60$ es 1.671, y para $v = 50$ es 1.676, por lo que **se rechaza** H_0 ya que 2.074 es mayor que cualquiera de esos valores críticos.

Alternativamente, como los tamaños de muestra son mayores que 30, podemos aproximar la distribución de T por una Normal $(0,1)$ y, utilizando estas tablas, obtenemos que:

$$p = P(Z > 2.074) = (1 - 0.98088) = 0.0096,$$

que es significativo al nivel $5\,\%$ y por lo tanto **se rechaza** H_0.

Es decir, **sí** existe suficiente evidencia para pensar que existe una diferencia en las medias poblacionales del tiempo de recuperación entre quienes toman suplemento y quienes no lo toman.

∎

Ejercicio 5.15 Un geólogo recolectó 20 muestras de mineral de una mina. Todas las muestras tenían el mismo peso. Al azar las dividió en dos grupos y midió los contenidos de zirconio de las muestras utilizando dos métodos diferentes. Los resultados fueron:

Método 1				
0.011	0.013	0.013	0.015	0.014
0.013	0.010	0.013	0.011	0.012

Método 2				
0.011	0.016	0.013	0.012	0.015
0.012	0.017	0.013	0.014	0.015

Pruebe estadísticamente si existe una diferencia significativa en el promedio de los contenidos de zirconio medidos por los dos métodos diferentes y asumiendo que siguen una distribución normal y que las varianzas poblacionales no difieren.

Solución:

H_0: $\mu_1 = \mu_2$ vs. H_a: $\mu_1 \neq \mu_2$ (Prueba de 2 colas).

Como nos informan que $\sigma_1^2 = \sigma_2^2$, el estadístico apropiado es:

$$T = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{s_c^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

que se comporta, cuando H_0 es cierta, como una t de Student con $(n_1 + n_2 - 2)$ grados de libertad, y donde:

$$s_c^2 = \frac{s_1^2 \cdot (n_1 - 1) + s_2^2 \cdot (n_2 - 1)}{n_1 + n_2 - 2}$$

es la estimación de la varianza común.

En nuestro caso, $\bar{x}_1 = 0.0125$, $s_1 = 0.0015$, $\bar{x}_2 = 0.0138$, $s_2 = 0.0019$, $n_1 = n_2 = 10$, y el valor del estadístico es $T = -1.6982$.

El valor crítico para $v = 10 + 10 - 2 = 18$ y $\alpha = 5\,\%$ es $t_{18,0.025} = \pm 2.101$, por lo que **no se rechaza** H_0 ya que $-2.101 < -1.698 < 2.101$. ∎

Ejercicio 5.16 Un inspector sospecha que en cierta población están vertiendo aguas negras a un río. Para comprobarlo, obtuvo cinco muestras de agua del río seleccionadas al azar en un lugar aguas arriba de la población, y otras cinco muestras de aguas abajo. Las mediciones de oxígeno disuelto (ppm) fueron:

Aguas arriba	4.8	5.2	5.0	4.9	5.1
Aguas abajo	5.0	4.7	4.9	4.8	4.9

Utilizando $\alpha = 0.05$ y suponiendo que las mediciones siguen una distribución Normal, ¿existe suficiente evidencia para indicar que la media del contenido de oxígeno aguas abajo es menor que la media del contenido de oxígeno aguas arriba?

Solución:

H_0: $\mu_1 = \mu_2$ vs. H_a: $\mu_1 > \mu_2$ (Prueba de 1 cola).

Ahora bien, no sabemos nada acerca de las varianzas, por lo que se debe usar la prueba de Welch. El estadístico apropiado es:

$$T = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

que se comporta, cuando H_0 es cierta, aproximadamente como una t de Student con:

$$v = \left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2 \Big/ \left(\frac{\left(s_1^2/n_1)\right)^2}{n_1 - 1} + \frac{\left(s_2^2/n_2\right)^2}{n_2 - 1}\right) \quad \text{grados de libertad.}$$

En nuestro caso, $\bar{x}_1 = 5$, $s_1 = 0.158$, $\bar{x}_2 = 4.86$, $s_2 = 0.114$, por lo que:

$T = 1.6068 \quad \text{y} \quad v = 7.3.$

En tablas de la t de Student encontramos con $\alpha = 5\,\%$ que el valor crítico para $v = 7$ es 1.895, por lo que **no se rechaza** H_0 ya que $1.607 < 1.895$. ■

Ejercicio 5.17 La prima anual del seguro de un auto en 2010 variaba dependiendo de la ciudad de residencia

Ciudad	Inbursa ($)	Seguros ASA ($)
Acapulco	2780	2352
Culiacán	2411	2464
Guadalajara	2261	2284
Cd. México	2263	2520

a) ¿Por qué es razonable pensar que estos pares de observaciones fueran dependientes?
b) Utilizando un nivel de significancia de $\alpha = 0.01$, ¿los datos proveen suficiente evidencia para indicar que sí existe una diferencia en el promedio de primas anuales entre los seguros Inbursa y ASA? Utilice $\alpha = 0.05$.
c) Calcule un intervalo de confianza de 99 % para la diferencia en el promedio de primas anuales para seguros Inbursa y ASA.

Solución:

a) Es razonable esperar que el costo de los seguros sean proporcionales a la peligrosidad de cada ciudad.
b) Las hipótesis a contrastar son: H_0: $\mu_1 = \mu_2$ vs. H_a: $\mu_1 \neq \mu_2$

Se trata de una prueba de 2 colas de comparación entre las medias de dos poblaciones pareadas, equivalente a:

$H_0 : \mu_{\text{diferencias}} = 0 \quad \text{vs.} \quad H_a : \mu_{\text{diferencias}} \neq 0$

El estadístico apropiado es:

$$T = \bar{d}/\sqrt{\frac{s_d^2}{n}} = -23.75/\sqrt{\frac{288.85^2}{4}} = -0.1644$$

Como $n = 4$, el valor crítico para $v = 3$ y $\alpha = 0.01$ se encuentra en tablas de la t de Student y es $t_{3,0.005} = \pm 5.84$, y como $-5.84 < -0.1644 < 5.84$, entonces **no se rechaza** H_0.

c) El intervalo de confianza para $\mu_{\text{diferencias}}$ al 99 % es $\bar{d} \pm \text{MOE}$, donde:

$$\text{MOE} = t_{3,0.005} * \frac{s_d}{\sqrt{n}}$$

Es decir:

$$IC99\,\% \text{ para } \mu_{\text{dif}} = 23.75 \pm 5.84 * 288.85/\sqrt{4} = 23.75 \pm 843.59, \text{ o sea } (-819.8, 867.3)$$

■

Ejercicio 5.18 Para comparar el color de dos tintes verdes en la línea de producción de una fábrica, se tomaron nueve muestras de tela y cada una se dividió en dos piezas. A una de las piezas elegida al azar se le aplicó el tinte verde 1 y a la otra pieza se le aplicó el tinte verde 2. A continuación se presenta la "calificación de color" para cada pieza. Suponiendo Normalidad de los datos y un $\alpha = 0.05$, ¿considera que existe suficiente evidencia para indicar una diferencia en la calificación media para los dos tintes?

Muestra	1	2	3	4	5	6	7	8	9
Tinte 1	10	12	9	8	15	12	9	10	15
Tinte 2	8	11	10	6	12	13	9	8	13

Solución:
Las hipótesis a contrastar son: H_0: $\mu_1 = \mu_2$ vs. H_a: $\mu_1 \neq \mu_2$ que son equivalentes a

$H_0 : \mu_{\text{diferencias}} = 0 \quad \text{vs.} \quad H_a : \mu_{\text{diferencias}} \neq 0$

En nuestro caso, $\bar{d} = -1.11$, $s_d = 1.453$, $n = 9$, y el estadístico apropiado es:

$$T = \bar{d}/\sqrt{s_d^2/n} = -1.11/\sqrt{1.453^2/9} = -2.29$$

El valor crítico para $v = 8$ y $\alpha = 5\,\%$ es $t_{8,0.025} = \pm 2.306$, por lo que **no se rechaza** H_0 ya que $-2.306 < -2.29 < 2.306$. ■

Ejercicio 5.19 En un estudio sobre la eficacia de dos fármacos para tratar un padecimiento poco común, los investigadores tienen como hipótesis que el fármaco A es más eficaz que el fármaco B para tratar la enfermedad. Para comprobar dicha hipótesis se administró el fármaco A a 375 pacientes y el fármaco B a 425 pacientes, 155 pacientes que usaron el fármaco A se recuperaron de la enfermedad, mientras que 175 de los que usaron el fármaco B se recuperaron. Realice una prueba de hipótesis adecuada para comprobar la hipótesis de los investigadores con un nivel de significancia del 5 %.

Solución:

Es una prueba de cola superior:

$H_0 : p_1 - p_2 = 0$ vs. $H_a : p_1 - p_2 > 0$

Según los datos del problema:

$$\widehat{p}_1 = \frac{155}{375} = 0.4133 \qquad \widehat{p}_2 = \frac{175}{425} = 0.4118 \qquad \widehat{p} = \frac{155+275}{375+425} = 0.4125$$

El estadístico de prueba:

$$Z = \frac{(\widehat{p}_1 - \widehat{p}_2) - d_0}{\sqrt{\widehat{p}\widehat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{(0.4133 - 0.4118) - 0}{\sqrt{(0.4125)(0.5875)\left(\frac{1}{375} + \frac{1}{425}\right)}} = 0.043$$

El valor crítico z_α para cola superior es $z_{0.05} = 1.645$. Dado que $Z < z_{0.05}$ **no se rechaza** H_0. Por lo que no hay evidencia que apoye la hipótesis de los investigadores. ■

Ejercicio 5.20 Un fabricante de baterías recargables sospecha que las baterías producidas en una de sus líneas de producción tienen una duración que varía demasiado. Para demostrarlo, seleccionó 50 baterías al azar de la línea sospechosa y 50 baterías al azar de una línea con variación controlada y midió su duración en horas hasta su reducción a un cierto voltaje, obteniéndose los siguientes datos:

Línea sospechosa	Línea regulada
$\bar{x}_1 = 9.40$	$\bar{x}_2 = 9.25$
$s_1 = 0.25$	$s_2 = 0.12$

¿Existe suficiente evidencia para indicar que las baterías producidas por la línea sospechosa tienen una varianza más grande que la de las producidas por la línea controlada con un $\alpha = 0.05$? Suponga Normalidad de los datos.

Solución:

H_0: $\sigma_S^2 = \sigma_C^2$ vs. H_a: $\sigma_S^2 > \sigma_C^2$ (Es una prueba de 1 sola cola).

El estadístico de prueba es:

$$F = \frac{s_S^2}{s_C^2} = \frac{(0.25)^2}{(0.12)^2} = 4.34$$

El valor crítico para $(v_1 = 49, v_2 = 49)$ y $\alpha = 5\,\%$ es $F_{0.05,(49,49)} = 1.607$, por lo que **se rechaza** H_0 ya que $1.607 < 4.34$. ■

6. Análisis de varianza, de regresión y correlación

Ejercicio 6.1 Se quiere comparar 3 máquinas diferentes para fabricar tapones de látex, en cuanto a la resistencia a la tensión del tapón que producen. Se utiliza una muestra aleatoria de cuatro tapones fabricados con cada máquina para determinar si la media de la resistencia a la tensión varía de una máquina a otra. A continuación se presentan las medidas de la resistencia a la tensión en hectogramos por centímetro cuadrado:

Máquina		
1	2	3
17	18	16
18	20	15
19	19	18
18	20	17

Realice un análisis de varianza con un nivel de significancia de 0.05 y argumente si la media de la resistencia a la tensión de las tres máquinas difiere o no de manera significativa.

Solución:

Recordemos que en el Diseño Completamente al Azar, la Suma de Cuadrados Total:

$$SC_{\text{tot}} = \sum_{i=1}^{k}\sum_{j=1}^{n} y_{ij}^2 - \frac{Y_{**}^2}{N}$$

se puede descomponer en una Suma de Cuadrados debida a los tratamientos y una Suma de

Cuadrados Residual o debida al error. Es decir,

$$SC_{\text{trat}} = n\sum_{i=1}^{k}\frac{Y_{i*}^2}{n} - \frac{Y_{**}^2}{N}, \qquad SC_{\text{err}} = SC_{\text{tot}} - SC_{\text{trat}}$$

donde $Y_{i*} = \sum_{j=1}^{n} y_{ij}$, $Y_{**} = \sum_{i=1}^{k}\sum_{j=1}^{n} y_{ij}$ y $N = n \times k$.

En nuestro caso $n = 4$, $k = 3$ y $N = 12$, por lo que tenemos que $SC_{\text{tot}} = 24.92$, $SC_{\text{trat}} = 15.17$ y $SC_{\text{err}} = 9.75$, y la tabla de ANOVA queda:

Fuente	SC	G.L.	CM	F	p value
Tratamientos	15.17	2	7.58	7	0.015
Error	9.75	9	1.08		
Total	24.92	11			

El valor-p = 0.015 es menor que $\alpha = 0.05$, por lo que concluimos que sí existe una diferencia significativa entre la resistencia promedio a la tensión de las tres máquinas.
También se puede probar la hipótesis comparando el valor calculado del estadístico de prueba con el valor crítico de cola derecha de la distribución F en tablas, el cuál es $F_{2,9,0.05} = 4.2565$. Dado que este valor es menor al valor F calculado $F_0 = 7$, se llega a la misma conclusión.

Una prueba de comparaciones múltiples post - hoc nos podría ayudar a definir si alguna máquina es significativamente mejor que otra. ▪

Ejercicio 6.2 Los siguientes datos son los porcentajes de aditivos, medidos por 5 analistas, de 3 marcas distintas de mermelada de fresa, A, B y C:

Analista	A	B	C
1	3.8	2.7	3.6
2	1.6	5.2	7.5
3	2.7	2.8	6.4
4	1.7	1.9	2.6
5	2.0	4.8	8.1

A un nivel de significancia de 0.05, realice un análisis de varianza de bloques completos aleatorizados para probar la hipótesis de que el porcentaje de aditivos es el mismo para las tres marcas de mermelada. ¿Cuál de ellas parece tener menos aditivos?

Solución:

Recordemos que en este caso de Diseño de Bloques Completos al Azar, la Suma de Cuadrados Total

$$SC_{\text{tot}} = \sum_{i=1}^{k}\sum_{j=1}^{b} y_{ij}^2 - \frac{Y_{**}^2}{N},$$

se descompone en

$$SC_{\text{trat}} = b\sum_{i=1}^{k}\frac{Y_{i*}^2}{n} - \frac{Y_{**}^2}{N}, \quad SC_{\text{bloq}} = k\sum_{j=1}^{b}\frac{Y_{*j}^2}{n} - \frac{Y_{**}^2}{N} \quad y \quad SC_{\text{err}} = SC_{\text{tot}} - SC_{\text{trat}} - SC_{\text{bloq}}$$

donde $Y_{*j} = \sum_{i=1}^{k} y_{ij}$, $Y_{i*} = \sum_{j=1}^{b} y_{ij}$, $Y_{**} = \sum_{i=1}^{k}\sum_{j=1}^{n} y_{ij}$ y $N = b \times k$.

En nuestro caso $b = 5$, $k = 3$ y $N = 15$, por lo que tenemos que $SC_{\text{tot}} = 62.89$, $SC_{\text{trat}} = 27.80$, $SC_{\text{bloq}} = 16.54$, y $SC_{\text{err}} = 18.56$. La tabla de ANOVA queda:

Fuente	SC	G.L.	CM	F	valor-p
Tratamientos	27.80	2	13.9	6	0.0257
Bloques	16.54	4	4.1	1.8	
Error	18.56	8	2.3		
Total	62.89	14			

El valor-p = 0.0257 es menor que $\alpha = 0.05$, por lo que concluimos que sí existe una diferencia significativa entre el porcentaje de aditivos en las tres mermeladas. De forma alternativa se puede probar la hipótesis buscando el valor crítico de cola derecha de la distribución F en tablas, el cuál es $F_{2,8,0.05} = 4.4590$, dado que este valor es menor al valor F calculado $F_o = 6$, se llega a la misma conclusión.

La mermelada que presenta menor cantidad de aditivos es la A, pero habría que hacer una prueba de comparaciones múltiples post-hoc para ver cuáles son las diferencias o bien otro experimento para ver si realmente existe diferencia entre B y A. ■

Ejercicio 6.3 Varias soluciones acuosas de fluoresceína fueron examinadas con un espectrómetro y se obtuvieron las siguientes lecturas de intensidad de fluorescencia:

Intensidad	2	4	10	12	8	21	25
Concentración, g/ml	0	2	4	6	8	10	12

a) Calcule la recta de regresión de mínimos cuadrados.
b) Grafique los puntos y la recta ajustada. ¿Le parece razonable la suposición de una relación lineal?
c) Use la recta de regresión para predecir la intensidad de fluorescencia para una concentración de fluoresceína de 9 g/ml.
d) Construya la tabla ANOVA para la regresión lineal.
e) Calcule R^2. ¿Qué le dice este valor acerca de la efectividad del análisis de regresión lineal?

Solución:

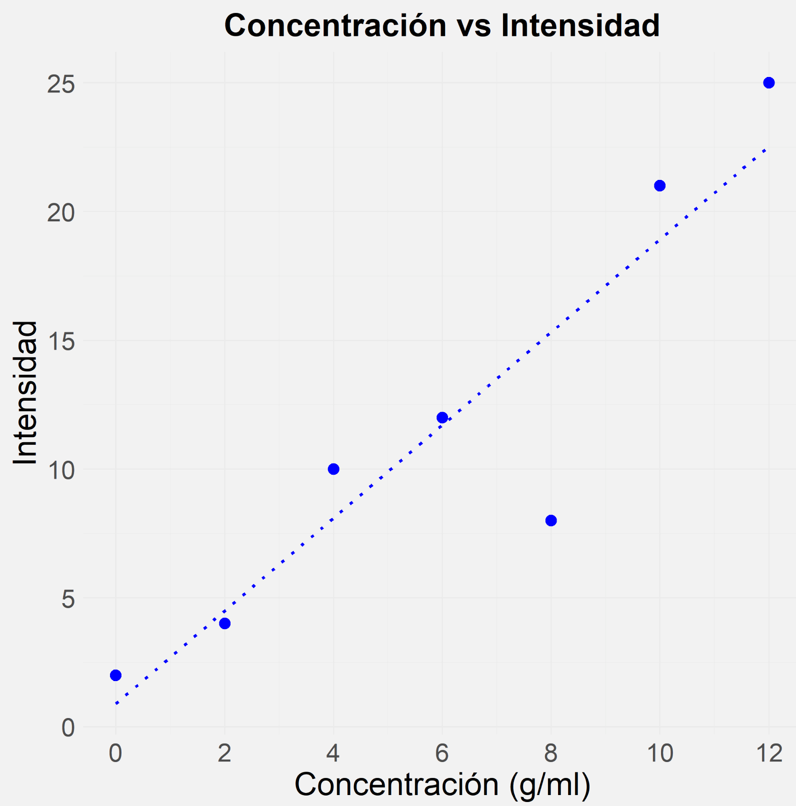
a) Nos ayuda construir la siguiente tabla:

x_i	y_i	$x_i-\bar{x}$	$(x_i-\bar{x})^2$	$y_i-\bar{y}$	$(y_i-\bar{y})^2$	$(y_i-\bar{y})(x_i-\bar{x})$
0	2	-6	36	-11.14	124.16	66.86
2	4	-4	16	-9.14	83.59	36.57
4	10	-2	4	-3.14	9.88	6.29
6	12	0	0	-1.14	1.31	0
8	8	2	4	4.86	23.59	9.71
10	21	4	16	7.86	61.73	31.43
12	25	6	36	11.86	140.59	71.14

Entonces $S_{xx}=\sum(x_i-\bar{x})^2=112$, $S_{yy}=\sum(y_i-\bar{y})^2=444.85$ y $S_{xy}=\sum(x_i-\bar{x})(y_i-\bar{y})=222$. Por lo tanto $b=S_{xy}/S_{xx}=1.982$ y $a=\bar{y}-b\bar{x}=(92/7)-1.982\times(42/7)=1.25$.

La recta de regresión es $y_i=a+bx_i=1.25+1.982x_i$.

b)



Sí es razonable la suposición de una relación lineal.

c) $y(9) = a + b \times 9 = 1.25 + 1.982 \times 9 = 19.089$

d) También podemos calcular $SC_{\text{tot}} = \sum(y_i - \bar{y})^2 = 444.86$, $SC_{\text{mod}} = \sum(\hat{y}_i - \bar{y})^2 = 417.3432$, y $SC_{\text{err}} = \sum(y_i - \hat{y}_i)^2 = 0.9368$ para construir la tabla de ANOVA:

Fuente	SC	$g.l.$	CM	F	Valor-p
Modelo	440.04	1	440.04	456	< 0.001
Error	4.82	5	0.964		
Total	444.86	6			

e) El coeficiente de determinación $R^2 = \frac{S_{xy}^2}{S_{xx} \times S_{yy}} = 0.9892$. Es muy buen ajuste, pues el modelo explica el 98.92 % de la variabilidad.

■

Ejercicio 6.4 Se determinó la concentración de plata en una muestra de material por espectrometría de absorción atómica con el método de adiciones estándar y se obtuvieron los siguientes resultados:

Plata añadida, $\mu g/ml$	0	5	10	15	20	25	30
Absorbancia	0.30	0.45	0.50	0.65	0.70	0.85	0.90

a) Calcule la recta de regresión de mínimos cuadrados.
b) Grafique los puntos y la recta ajustada. ¿Le parece razonable la suposición de una relación lineal?
c) Use la recta de regresión para predecir la absorbancia que se obtendría para 13 $\mu g/ml$ de plata añadida.
d) Construya la tabla ANOVA para la regresión lineal.
e) Calcule R^2. ¿Qué le dice este valor acerca de la efectividad del análisis de regresión lineal?

Solución:

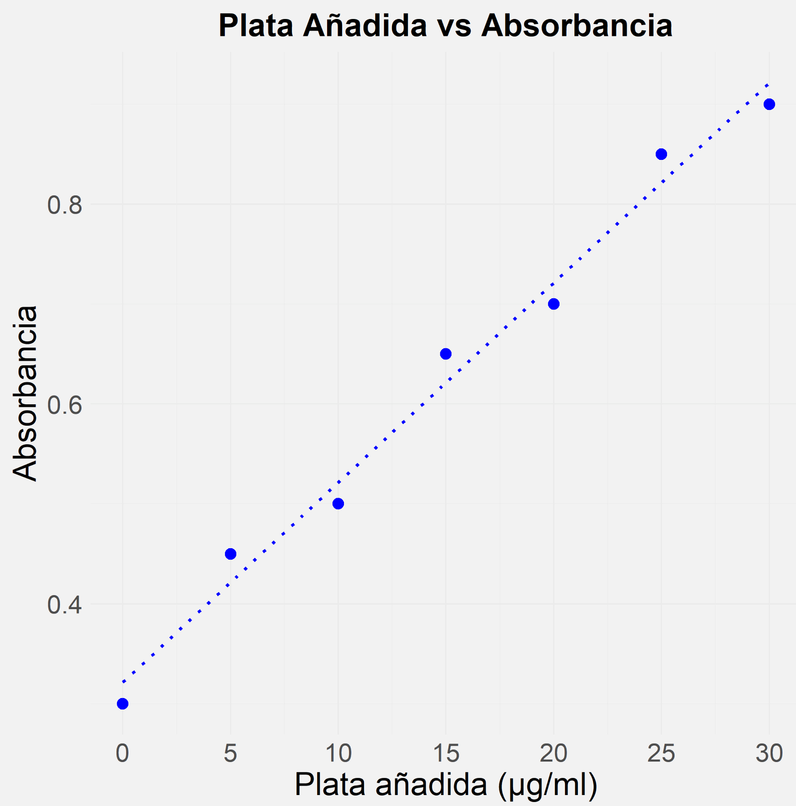
a) Nos ayuda construir la siguiente tabla:

x_i	y_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$y_i - \bar{y}$	$(y_i - \bar{y})(x_i - \bar{x})$
0	0.30	-15	225	-0.321	4.821
5	0.45	-10	100	-0.171	1.714
10	0.50	-5	25	-0.121	0.607
15	0.65	0	0	0.029	0.000
20	0.70	5	25	0.079	0.393
25	0.85	10	100	0.229	2.286
30	0.90	15	225	0.279	4.179

Entonces, $S_{xx} = \sum(x_i - \bar{x})^2 = 700$, $S_{yy} = \sum(y_i - \bar{y})^2 = 0.284$ y $S_{xy} = \sum(x_i - \bar{x})(y_i - \bar{y}) = 14$. Por lo que $b = S_{xy}/S_{xx} = 0.020$ y $a = \bar{y} - b\bar{x} = 0.321$.

La recta de regresión es $y_i = a + bx_i = 0.321 + 0.020x_i$.

b) a



Sí es razonable la suposición de una relación lineal.

c) Use la recta de regresión para predecir la absorbancia que se obtendría para 13 $\mu g/ml$ de plata añadida:

$y(13) = 0.321 + 0.020 \times 13 = 0.581$

d) También podemos calcular $SC_{\text{tot}} = \sum(y_i - \bar{y})^2 = 0.2843$, $SC_{\text{mod}} = \sum(\hat{y}_i - \bar{y})^2 = 0.2800$ y $SC_{\text{err}} = \sum(y_i - \hat{y}_i)^2 = 0.0043$ para construir la tabla de ANOVA:

Fuente	SC	$g.l.$	CM	F	Valor-p
Modelo	0.2800	1	0.2433	327	< 0.001
Error	0.0043	5	0.00012		
Total	0.2843	6			

e) El coeficiente de determinación $R^2 = \frac{S_{xy}^2}{S_{xx} \times S_{yy}} = 0.9849$. Muy buen ajuste. El modelo explica el 98.49 % de la variabilidad total.

■

7. Estadística no paramétrica

Ejercicio 7.1 Los siguientes datos representan los años de vida de una muestra de ratones de laboratorio pertenecientes a cierta cepa.

$$1.7, 1.5, 2.2, 0.9, 1.3, 2.0, 1.6, 1.8, 1.5, 2.0, 1.2$$

Lleve a cabo una **prueba de signo** para probar la hipótesis de que la mediana de la población es de 1.8 años. Utilice $\alpha = 0.05$.

Solución:

Las hipótesis a contrastar son:

$H_0 : \tilde{\mu} = 1.8 \quad \text{vs} \quad H_1 : \tilde{\mu} \neq 1.8.$

En este caso, el estadístico de prueba es X = Número de "MASSES" en 11 intentos, que es Binomial con $p = 1/2$. Donde MÁS (o éxito) es cuando un valor de la muestra excede el valor a probar $\tilde{\mu}_0 = 1.8$, y MENOS (o fracaso) es cuando un valor de la muestra es menor a $\tilde{\mu}_0 = 1.8$; si un valor es igual a $\tilde{\mu}_0$ se descarta, reduciendo el tamaño de la muestra. En nuestra muestra esto último sucede con la octava observación, por lo cual obtenemos, en el orden de los datos, la secuencia $- - + - - + - / - + -$, en la cual $n = 10$, $x = 3$.

Por lo tanto, el valor-p correspondiente a una cola se obtiene de la distribución binomial

$$P(X \leq 3; n = 10, p = 1/2) = \sum_{x=0}^{3} C_{10,x} \cdot (0.5)^{10} = 0.1719$$

y como la prueba es de dos colas, valor-p $= 2P(X \leq 3) = 0.3438$ que es mayor a $\alpha = 0.05$, por lo que No se rechaza H_0.

■

Ejercicio 7.2 Una compañía de transportes desea probar si las llantas de la Marca A ahorrarían más gasolina que la marca B que utilizan actualmente. Para decidirlo, primero se equipan con llantas A a 16 vehículos seleccionados al azar de su flotilla, los cuales se hacen circular por una carretera previamente elegida anotando sus consumos en Km/l. Después, los mismos 16 vehículos se preparan con llantas B y se vuelven a probar en la misma carretera anotando al final su consumo. Los datos se presentan abajo.

A un nivel de significancia de 0.05 ¿Podemos afirmar que los vehículos con llantas marca A ahorran más combustible que los vehículos con llantas marca B?

AUTOMOVIL	1	2	3	4	5	6	7	8
Marca A	4.2	4.7	6.6	7	6.7	4.5	5.7	6
Marca B	4.1	4.9	6.2	6.9	6.8	4.4	5.7	5.8

AUTOMOVIL	9	10	11	12	13	14	15	16
Marca A	7.4	4.9	6.1	5.2	5.7	6.9	6.8	4.9
Marca B	6.9	4.9	6	4.9	5.3	6.5	7.1	4.8

Solución:
Sean $\tilde{\mu}_A$ y $\tilde{\mu}_B$ las medianas de los kilómetros por litro para los automóviles equipados con la Marca A y con la Marca B, respectivamente. Las hipótesis a contrastar son:

$H_0 : \tilde{\mu}_A - \tilde{\mu}_B = 0 \quad \text{vs} \quad H_1 : \tilde{\mu}_A - \tilde{\mu}_B > 0$

El estadístico de prueba es X = Número de "MASSES" en n intentos, que bajo la hipótesis nula es Binomial con $p = 1/2$, donde MÁS o éxito es cuando un valor de la muestra A menos el valor correspondiente de la marca B excede el valor a probar (en este caso 0), y MENOS o fracaso es cuando la diferencia es menor a 0; si un valor es igual a 0 se descarta, reduciendo el tamaño de la muestra. En nuestro caso, la secuencia sería: $+-++-+/++/+++++-+$, por lo cual $n = 14$ y $x = 11$.

Utilizando la aproximación del estadístico con la distribución normal (con corrección por continuidad), encontramos que

$$Z = \frac{(X - np_0 - 0.5)}{\sqrt{np_0q_0}} = \frac{(11 - 14(0.5) - 0.5)}{\sqrt{14(0.5)(0.5)}} = 1.87$$

y entonces valor-$p = P(X \geq 11) \approx P(Z > 1.87) = 0.0307$.
Por lo tanto, se rechaza H_0 y se concluye que, en promedio, las llantas A ahorran más combustible. ■

Ejercicio 7.3 Resuelva el problema anterior sobre la vida de ratones de laboratorio, suponiendo que los valores observados provienen de una población simétrica y usando la prueba de rango con signo (de Wilcoxon).

Solución:

Las hipótesis a contrastar son: $H_0 : \tilde{\mu} = 1.8$ vs $H_1 : \tilde{\mu} \neq 1.8$.

Para calcular la región crítica del estadístico de prueba primero descartaremos el valor que es igual a 1.8, lo que reduce el tamaño de muestra a $n = 10$. De una tabla de valores críticos para la prueba de rangos con signo obtenemos que la región crítica para $\alpha = 0.05$, es $w \leq 8$.

Para calcular el estadístico de prueba debemos restar 1.8 a cada valor y ordenar las diferencias asignándoles rangos sin tomar en cuenta su signo, es decir:

d_i	-0.3	0.4	-0.9	-0.5	0.2	-0.2	-0.3	0.2	-0.6	-0.1
Rangos	5.5	7	10	8	3	3	5.5	3	9	1

Al sumar los rangos de aquellos correspondientes al mismo signo obtendremos $w_+ = 7 + 3 + 3 = 13$ y $w_- = 42$. El menor de los dos es $w = 13$.

Como este valor no cae en la región crítica ($w \leq 8$), NO se rechaza H_0. Es decir, la mediana del tiempo de operación no difiere significativamente de 1.8 horas. ■

Ejercicio 7.4 Se hizo un estudio para comparar el contenido de alquitrán en miligramos en dos marcas líder de cigarros. Los resultados son los siguientes:

Marca A	2.1	4.0	6.3	5.4	4.8	3.7	6.1	3.3		
Marca B	4.1	0.6	3.1	2.5	4	6.2	1.6	2.2	1.9	5.4

Utilice la Prueba de suma de rangos de Wilcoxon (también llamada Prueba de Mann-Whitney) con un nivel de significancia $\alpha = 0.05$ para probar la hipótesis de que las medianas del contenido de alquitrán de ambas marcas son iguales, contra la hipótesis alternativa de que no lo son.

Solución:

Las hipótesis a contrastar son: $H_0 : \tilde{\mu}_A - \tilde{\mu}_B = 0$ vs $H_1 : \tilde{\mu}_A - \tilde{\mu}_B \neq 0$.

La región crítica para el estadístico de prueba la encontramos en una tabla de valores críticos para la prueba de suma de rangos de Wilcoxon. Esta es $u \leq 17$.

Para calcular el estadístico de prueba, primero debemos ordenar las observaciones y asignarles rangos (los marcados con * corresponden a Marca A):

Observ.	0.6	1.6	1.9	2.1	2.2	2.5	3.1	3.3	3.7
Rango	1	2	3	4^*	5	6	7	8^*	9^*

Observ.	4.0	4.0	4.1	4.8	5.4	5.4	6.1	6.2	6.3
Rango	10.5^*	10.5	12	13^*	14.5^*	14.5	16^*	17	18^*

La suma de los rangos que corresponden a la muestra más pequeña es $w_1 = 4 + 8 + \cdots + 18 = 93$ y los correspondientes a la más grande es $w_2 = 78$. Entonces

$$U_1 = w_1 - n_1(n_1+1)/2 = 93 - 72/2 = 57 \quad y \quad U_2 = w_2 - n_2(n_2+1)/2 = 78 - 110/2 = 23$$

Por lo tanto, NO se rechaza la hipótesis nula. Es decir, no hay diferencia significativa entre las medianas del contenido de alquitrán de la marca A y la de la marca B. ∎