Ejercicios para practicar

del Curso de Estadística y Probabilidad

📊 En este documento encontrarás **ejercicios para practicar los temas de las clases del curso**. Podrás leer la instrucción y dejar tus respuestas en el recuadro de cada ejercicio.

En la **página 26 podrás encontrar las soluciones** a cada ejercicio por si tienes dudas y quieres comprobar tus resultados.

Clase: Conceptos clave de estadística

Ejercicio 1:

Identifica las variables en la siguiente descripción de datos y clasifica las variables como categóricas o cuantitativas. Si la variable es cuantitativa, enumera las unidades.

“La Fórmula 1 es una carrera de autos que se lleva a cabo desde 1950. Un piloto necesita completar entre 48 y 70 vueltas que cubren una distancia de 305 kilómetros. Los resultados de la carrera se informan por número de conductor, nombre del conductor, tipo de automóvil que usa el conductor y el tiempo a la diezmilésima de segundo más cercana. Si un piloto no termina la carrera, en lugar del tiempo para completar la carrera, se registra el número de vueltas completadas”.

Respuesta:

	Variable	Subtipo	Unidad de medida
Número de conductor	CUALITATIVA	Nominal	
Nombre de conductor	CUALITATIVA	Nominal	
Tipo de automóvil	CUALITATIVA	Nominal	
Tiempo	CUANTITATIVA	Continua	Diezmilésima de segundo
Vueltas completadas	CUANTITATIVA	Discreta	Cantidad de vueltas

Clase: Tablas unidimensionales y bidimensionales

Ejercicio 2:

Explica por qué la siguiente es un ejemplo de una tabla de una entrada.

Sabor	Cucharadas vendidas	¿Puede llevar chocolate extra?	Helado o nieve
Mango	67	No	Nieve
Pistache	92	No	Helado
Vainilla	74	Sí	Helado
Nuez	63	Sí	Helado
Fresa	97	No	Nieve
Napolitano	53	No	Helado

Respuesta:

La tabla es **Unidimensional**, ya que en cada columna la pregunta se responde de forma concreta. No se requiere de más preguntas para responder los datos de las variables.

Clase: ¿Qué es la frecuencia estadística y con qué se come?

Ejercicio 3:

Una compañía de paseos de perros lleva registro de cuántas veces se pasean los perros. 40% de los perros paseados por la compañía recibieron entre 25 y 40 paseos y ningún perro recibió más de 40 paseos.

¿Cuántos perros tuvieron entre 0 y 25 paseos si la compañía si la compañía pasea a 400 perros?

Respuesta:

Total = 100% (400)
a= 40% (25, 40)
b= 60% (0, 25)

Total = a+b
400*0.60 = 240

El ejercicio nos pide prácticamente hallar el número de perros restante al 40%, es decir el 60% de 400 perros que es el 100%.

La respuesta es : 240 perros.

Clase: ¿Qué es la frecuencia estadística y con qué se come?

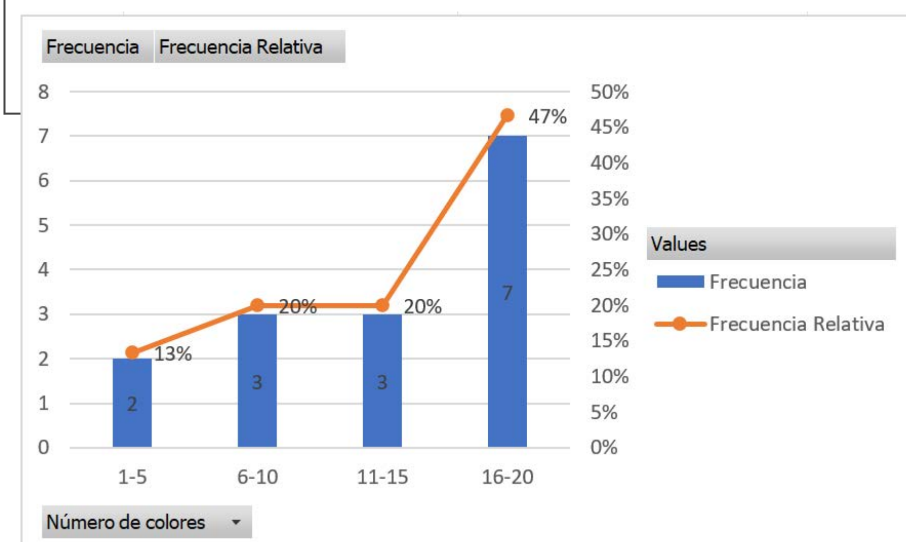
Ejercicio 4:

El número de colores en un estuche de varios estudiantes de kinder es: 5,2,6,6,10,12,16,14,16,15,17,18,17,17,18.

Completa la tabla de frecuencia y frecuencia relativa para los datos. Utilízala para crear un histograma de frecuencias.

Respuesta:

Número de colores	Row Labels	Count of Número de colores	Count of Número de colores2
5	1-5	2	13%
2	6-10	3	20%
6	11-15	3	20%
6	16-20	7	47%
10	**Grand Total**	**15**	**100%**
12			
16			
14			
16			
15			
17			
18			
17			
17			
18			



Clase: ¿Cuál es la mejor visualización para mis datos?

Ejercicio 5:

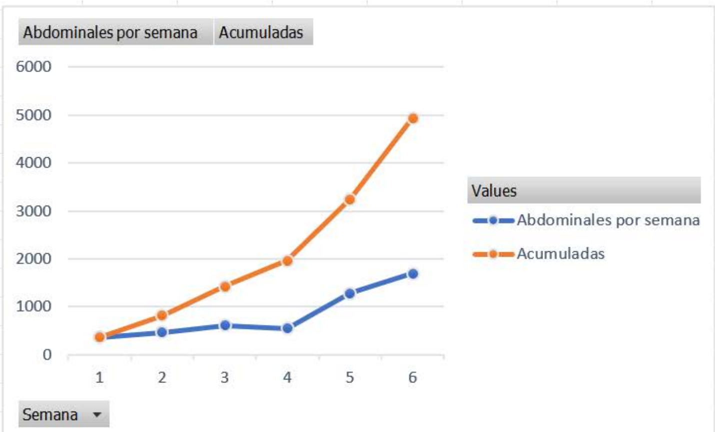
Bertha empezó un programa de ejercicios para poder hacer 2000 abdominales a lo largo de varias semanas. Al final de la semana 6 habrá completado 1685 abdominales.

Crea una gráfica de ojiva de los datos.

Semana	Abdominales
1	355
2	460
3	605
4	545
5	1280
6	1690

Respuesta:

Semana	Abdominales por semana	Acumuladas
1	355	355
2	460	815
3	605	1420
4	545	1965
5	1280	3245
6	1690	4935
Grand Total	**4935**	



Clase: Distribuciones conjuntas

Ejercicio 6:

Una compañía farmacéutica está examinando la acidez como efecto secundario de su nuevo analgésico. ¿Qué conclusiones puedes hacer de la siguiente distribución marginal?

	Analgésico	Placebo	Total
Acidez leve	9	176	185
Acidez intensa	107	31	138
Sin acidez	10573	10473	21046
Total	10689	10680	21369

Respuesta:

Row Labels	Sum of Analgésico
Acidez intensa	1.00%
Acidez leve	0.08%
Sin acidez	98.91%
Grand Total	**100.00%**

Row Labels	Sum of Placebo
Acidez intensa	0.29%
Acidez leve	1.65%
Sin acidez	98.06%
Grand Total	**100.00%**

Row Labels	Analgésico y placebo
Acidez intensa	0.65%
Acidez leve	0.87%
Sin acidez	98.49%
Grand Total	**100.00%**

Conclusiones:

1. Entre el 0.08% y el 1% de las personas tuvieron acidez con el analgésico. El 1% presentó azidez intensa.
2. Entre el 0.29% y 1.65% de las personas tuvieron acidez con el placebo. El 1.65% presentó acidez leve.
3. Menos del 1% de las personas presentaron algún tipo de acidez con el analgésico y el placebo.
4. Teniendo en cuenta los resultados, el efecto secundario de la acidez puede presentarse pero no es muy común que suceda.

Clase: Medidas de tendencia central: media, mediana y moda

Ejercicio 7:

¿Cuál es la media del siguiente data set? 107, 252, 360, 424

Hazlo de forma manual y luego comprueba tu resultado con Excel.

Respuesta:

	107
	252
	360
	424
Average	285.75

Para calcular la **media** de un conjunto de números, sumamos todos los números y luego dividimos la suma entre la cantidad de números en el conjunto. En este caso, tenemos los siguientes números: 107, 252, 360 y 424.

La suma de estos números es: 107 + 252 + 360 + 424 = 1143.

Como tenemos 4 números en el conjunto, la media se calcula dividiendo la suma entre 4:

Media = 1143 / 4 = 285.75.

Por lo tanto, la media de los números 107, 252, 360 y 424 es **285.75.**

Clase: Medidas de tendencia central: media, mediana y moda

Ejercicio 8:

¿Cuál es la mediana del siguiente data set? 74, 75, 62, 77, 70, 71, 64, 69, 70, 73

Hazlo de forma manual y luego comprueba tu resultado con Excel.

Respuesta:

74	75	62	77	70	71	64	69	70	73
Average	70.5								
Median	70.5								
Mode	70								

2. Mediana:
   Para calcular la mediana, primero ordenamos los números de menor a mayor y luego encontramos el valor central. Si hay un número impar de elementos, la mediana será el número central. Si hay un número par de elementos, la mediana será el promedio de los dos números centrales.

Ordenando los números de menor a mayor: 62, 64, 69, 70, 70, 71, 73, 74, 75, 77

Hay un número par de elementos (10 en total), por lo que tomaremos los dos números centrales y calcularemos su promedio.

Mediana = (70 + 71) / 2 = 70.5

Por lo tanto, la mediana de los números es 70.5.

Clase: Medidas de dispersión: rango e IQR

Ejercicio 9:

Laura visita centros de investigación para un proyecto y registra cuántos papers le sirven a su proyecto de investigación de cada centro.

Calcula el rango y el IQR de los datos: 6, 20, 21, 0, 3, 7, 7,6, ,21, 3, 3, 1.

Respuesta:

Datos
0
1
3
3
3
6
6
7
7
20
21
21

Rango	21
IQR	7.25
	13.75

Utilizando la fórmula QUARTILE.INC de EXCEL

1. Ordenamos los datos de menor a mayor: 0, 1, 3, 3, 3, 6, 6, 7, 7, 20, 21, 21.
2. Calculamos el primer cuartil (Q1):
Q1 = QUARTILE.INC(data, 1) = QUARTILE.INC([6, 20, 21, 0, 3, 7, 7, 6, 21, 3, 3, 1], 1) = 3.
3. Calculamos el tercer cuartil (Q3):
Q3 = QUARTILE.INC(data, 3) = QUARTILE.INC([6, 20, 21, 0, 3, 7, 7, 6, 21, 3, 3, 1], 3) = 10.25.
4. Calculamos el IQR:
IQR = Q3 - Q1 = 10.25 - 3 = 7.25.

Clase: Medidas de dispersión: rango e IQR

Ejercicio 10:

¿Cuál es el rango de los siguientes datos?

Número de perro	Frecuenci
20	3
25	1
32	1
38	2
39	3
40	2
43	2

Rango Np	23
Rango Fr	2

Respuesta:

Clase: Desplazamiento y escala de valores

Ejercicio 11:

Los estudiantes de una clase de francés terminaron un examen con una calificación promedio de 65 sobre 100. El rango del examen fue de 25 puntos.

Si a cada calificación se le aumenta un 10%, ¿cuáles son los nuevos valores para el promedio y el rango?

Respuesta:

Para determinar los nuevos valores para el promedio y el rango después de aumentar cada calificación en un 10%, podemos seguir los siguientes pasos:

1. Calcular el nuevo promedio:
   - El promedio inicial es de 65 sobre 100.
   - Aumentar cada calificación en un 10% significa sumarle al valor original el 10% de ese valor.
   - Entonces, el nuevo promedio se calcula como:
     nuevo promedio = promedio inicial + (10% del promedio inicial)
     nuevo promedio = 65 + (0.10 * 65)
     nuevo promedio = 65 + 6.5
     nuevo promedio = 71.5

2. Calcular el nuevo rango:
   - El rango inicial es de 25 puntos.
   - Aumentar cada calificación en un 10% no afecta el rango, ya que el rango se define como la diferencia entre el valor máximo y el valor mínimo en un conjunto de datos.
   - Por lo tanto, el nuevo rango se mantiene en 25 puntos.

Entonces, después de aumentar cada calificación en un 10%, el nuevo promedio es de 71.5 sobre 100 y el nuevo rango se mantiene en 25 puntos.

Clase: Box plots y el resumen de 5 números

Ejercicio 12:

¿Cuál es el rango y el IQR del siguiente conjunto de datos?

a. Mediana: 617,594
b. Mínimo: 216,290
c. Máximo: 845,300
d. 1er cuartil: 324,876
e. 3er cuartil: 790,370

Respuesta:

¿Cuál es el rango y el IQR del siguiente conjunto de datos?

Máximo	845,300
Q3	790,370
Mediana (Q2)	617,594
Q1	324,876
Mínimo	216,290

RANGO	**629,010**
IQR	**465,494**

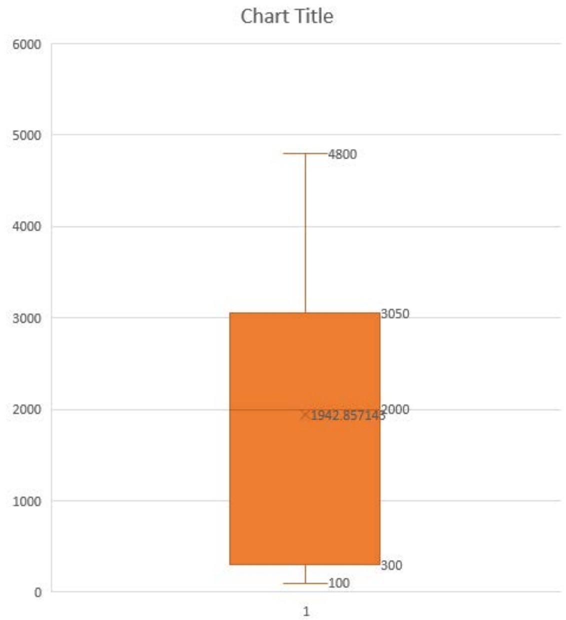
Clase: Box plots y el resumen de 5 números

Ejercicio 13:

Crea el box plot basado en la siguiente información de un conjunto de datos:

a. Moda: 300
b. Mínimo: 100
c. 1er cuartil: 300
d. Mediana: 2000
e. Promedio: 1887.5
f. 3er cuartil: 3050
g. Máximo: 4800

Respuesta:



Nota: el ejercicio tiene un error, ya que el promedio debe ser 1942.837.

No hay forma de que con los datos proporcionados el promedio sea 1887.5

Clase: Media, varianza y desviación estándar

Ejercicio 14:

Los estudiantes de la señora López tomaron un examen el viernes. Ella los califica durante el fin de semana y nota que la calificación promedio del examen es de 68 puntos sobre 100. Para apoyarles decide agregarle a cada uno 10 puntos. ¿Cuáles son los nuevos valores del promedio y desviación estándar?

Respuesta:

La nueva media es 78 y la desviación estándar no cambia.

Cuando añadimos o restamos una constante a cada punto de datos de un conjunto, la media y la mediana aumentarán o disminuirán en la misma cantidad, pero la desviación estándar seguirá siendo la misma. Esto se debe a que la desviación estándar es una medida de la dispersión de los datos, y añadir o restar una constante a cada punto de datos no cambia la dispersión.

En este caso, la media de las puntuaciones de las pruebas era 68. Cuando la Sra. López añadió 10 puntos a cada puntuación, la media aumentó a 78 puntos. La desviación estándar, sin embargo, sigue siendo la misma. Esto se debe a que la desviación estándar es una medida de la dispersión de los datos, y añadir o restar una constante a cada punto de datos no cambia la dispersión.

Clase: Histogramas, polígonos de frecuencia y curvas de densidad

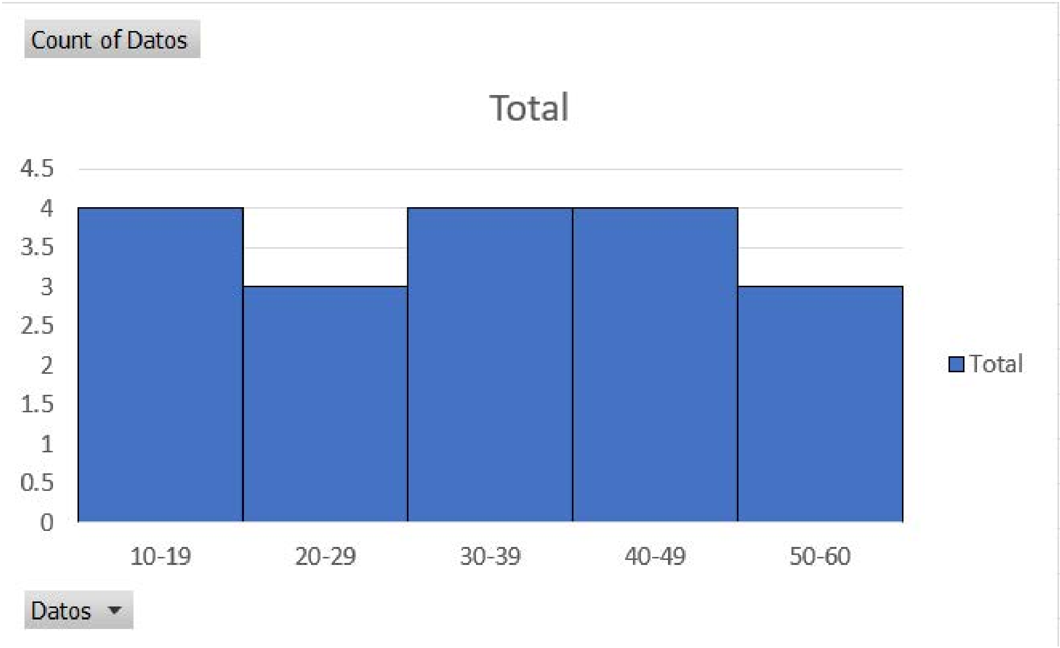
Ejercicio 15:

El siguiente diagrama de tallo y hoja muestra el número de piezas de ropa en cada perchero en una tienda de ropa.

Crea un histograma a partir del diagrama de tallos y usa cubos de tamaño 10. (Recuerda que 1 | = 10)

1	0 1 2 8
2	8 8 9
3	2 6 8 9
4	3 4 4 5
5	2 6
6	0

Respuesta:



Clase: Distribuciones simétricas y asimétricas

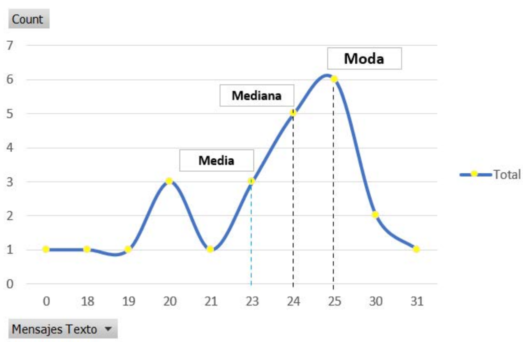
Ejercicio 16:

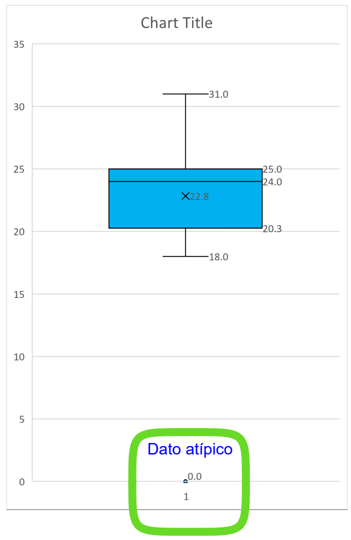
La cantidad de mensajes de texto enviados cada día por la mamá de Lucy son: 0, 18, 19, 20, 20, 20, 21, 23, 23, 23, 24, 24, 24, 24, 24, 25, 25, 25, 25, 25, 25, 30, 30, 31.

¿Hay valores atípicos en el conjunto de datos? Si es así, indica cuáles son. ¿Cuál es la mejor medida de tendencia central para los datos? ¿Cuál es la mejor medida de propagación?

Respuesta:

Media	**22.8**
Mediana (Q2)	**24**
Moda	**25**
Max	**31**
Min	**0**
Q1	**20.8**
Q3	**25**
IQR	**4.3**
Valor Atípico bajo	**14.4**
Valor Atípico alto	**31.4**
Variaza	**33.0**
Desviación Estándar	**5.7**
	5.7





La mejor medida de tendencia central para los datos es la mediana. Ya que no es afectada por los datos atípicos, como si lo es la media.

Los datos tienen un valor atípico bajo de 0 (**que se puede evidenciar en la box plot chart**) porque es inferior a **14.4**. Los datos no tienen valores atípicos altos porque ningún número en el conjunto es mayor que **31.4**.

Dado que los datos **tienen un valor atípico**, la mejor medida de tendencia central es la mediana y la mejor medida de dispersión es el rango intercuartílico.

Para una distribución asimétrica, la mediana siempre será la principal medida de tendencia central, ya que representa el punto de equilibrio.

Clase: Muestreo y sesgo

Ejercicio 17:

El zoológico realizó una encuesta para ver los motivos de sus visitantes para venir al zoológico. Les preguntan a las familias con hijos el por qué les gusta ir al zoológico en la salida.

Da una razón del porqué este método de muestreo puede estar sesgado.

Respuesta:

El método de muestreo utilizado en esta encuesta, que consiste en preguntar solo a las familias con hijos el motivo de su visita al zoológico, puede estar sesgado debido a un sesgo de selección. (Aunque se debe verificar cuál es el objetivo del estudio).

Este sesgo de selección puede ocurrir porque la muestra se limita únicamente a familias con hijos, lo que excluye a otros posibles grupos de visitantes, como parejas sin hijos, adultos solteros o grupos de amigos. Al hacerlo, se ignora una parte de la población que también puede visitar el zoológico por distintas razones.

Como resultado, los motivos de visita que se obtienen a partir de esta encuesta pueden no ser representativos de todos los visitantes del zoológico. Los datos obtenidos estarían sesgados hacia las opiniones y motivaciones de las familias con hijos, mientras que se ignoraría la diversidad de opiniones y razones de otros grupos de visitantes. Esto limitaría la validez y la generalización de los resultados a la población completa de visitantes del zoológico.

Para obtener una imagen más completa y representativa de los motivos de visita al zoológico, sería necesario utilizar un método de muestreo más amplio que incluya a diferentes segmentos de la población de visitantes, como encuestar a todas las familias, parejas sin hijos, adultos solteros y grupos de amigos. Esto permitiría obtener una visión más diversa y precisa de los motivos de visita al zoológico.

Clase: Muestreo y sesgo

Ejercicio 18:

El dueño de un restaurante da una encuesta a cada comensal. Incluye en la misma la pregunta: "¿Has dejado a tu mesero sin propina?"

Da una razón del porqué este método de muestreo puede estar sesgado.

Respuesta:

El método de muestreo utilizado en este caso, en el que el dueño del restaurante da una encuesta a cada comensal y pregunta específicamente si han dejado sin propina al mesero, puede estar sesgado debido a un sesgo de respuesta.

El sesgo de respuesta puede ocurrir porque los comensales pueden sentirse incómodos o avergonzados al admitir que han dejado sin propina al mesero. La presencia del dueño del restaurante al entregar personalmente la encuesta puede aumentar aún más esta presión social y la probabilidad de respuestas falsas o poco sinceras. Como resultado, es probable que los comensales eviten admitir que han dejado sin propina al mesero, lo que sesgaría los resultados de la encuesta.

Además, el sesgo de respuesta también puede influir en las respuestas de los comensales que sí han dejado sin propina al mesero. Al ser conscientes de que el dueño del restaurante está presente y puede identificar a los encuestados, podrían sentir la necesidad de responder de manera más positiva o justificar sus acciones, lo que podría llevar a una subestimación de la frecuencia real de casos en los que se ha dejado sin propina al mesero.

Este sesgo de respuesta podría distorsionar la percepción del dueño del restaurante sobre la frecuencia real de comensales que dejan sin propina al mesero, lo que a su vez podría afectar sus decisiones y acciones en el manejo de su personal o en la implementación de políticas de servicio.

Clase: Regla de la suma, unión e intersección

Ejercicio 19:

Dadas las probabilidades P(A)=0.3 y P(B)=0.6 y P(A∩B)=0.05, ¿cuánto vale P(AUB)? ¿Son A y B mutuamente exclusivos? ¿Por qué? ¿Por qué no?

Respuesta:

P(AUB) = P(A) + P(B) - P(A∩B)

Sustituyendo los valores que se nos dan, tenemos:

P(AUB) = 0.3 + 0.6 - 0.05
P(AUB) = **0.85**

Por lo tanto, la probabilidad de la unión de A y B, P(AUB), es igual a 0.85.

Para determinar si A y B son mutuamente excluyentes, debemos verificar si su intersección, P(A∩B), es igual a cero. En este caso, P(A∩B) es igual a 0.05, lo que significa que hay una superposición entre A y B. Por lo tanto, A y B no son mutuamente excluyentes, ya que tienen elementos en común.

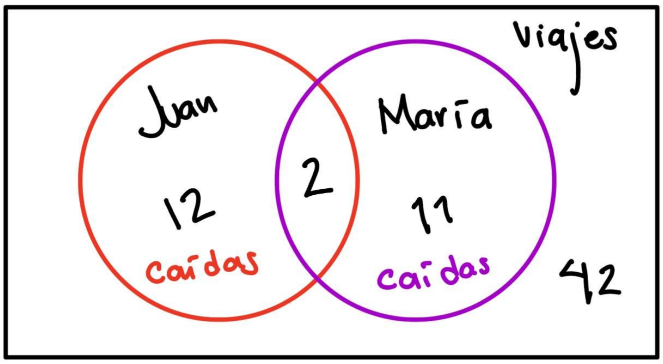
La razón por la que A y B no son mutuamente excluyentes es que la probabilidad de su intersección, P(A∩B), es diferente de cero. Esto indica que hay casos en los que ocurren tanto A como B al mismo tiempo. En otras palabras, hay elementos que pertenecen tanto a A como a B. Si A y B fueran mutuamente excluyentes, su intersección sería vacía y P(A∩B) sería igual a cero.

Clase: Regla de la suma, unión e intersección

Ejercicio 20:

Juan y María hacen varios viajes a una montaña juntos. El siguiente diagrama de Venn muestra el número de veces que se cayeron en sus múltiples viajes.

¿Cuál es la probabilidad de que ambos se hayan caído en un viaje particular y cuál es la probabilidad de que solo Juan o solo María se hayan caído en un viaje particular?



Respuesta:

Primero debemos calcular la cantidad de viajes que Juan y María han hecho juntos para tener el total de los viajes, esto es:

Total viajes: 12 + 2 + 11 + 42 = 67. -- Como Juan y María se cayeron en 2 de los viajes, la probabilidad de caída de los dos sería:

P(caídaJuan ∩ caídaMaría) = 2/67 = **0.02** = **2%**

Por el diagrama de Venn podemos analizar que de los 67 viajes en 12 se cayó solo Juan y en 11 se cayó solo María. Entonces, la probabilidad de que solo Juan se caiga o solo María se caiga es:

P(cayóJuan U cayóMaría) = (12+11 / 67) - P(caídaJuan ∩ caídaMaría) = 0.34 - 0.02 = **0.32** = **32%**

Clase: Probabilidad condicional y eventos dependientes e independientes

Ejercicio 21:

¿Cuál es la probabilidad de obtener 4 cabezas seguidas cuando lanzamos una moneda al aire cuatro veces? Considerando que las 2 opciones de la moneda son cara y cruz.

Respuesta:

La probabilidad sería:

P(Ca1, Ca2, Ca3, Ca4) = 1/2*1/2*1/2*1/2 = 1/16
= 0.0625
= **6.25%**

Clase: Probabilidad condicional y eventos dependientes e independientes

Ejercicio 22:

Camila tiene 12 animales de peluche, 7 de los cuales son elefantes (4 de los elefantes tocan música y se iluminan) y 5 de los cuales son osos (2 de los osos tocan música y se iluminan).

Su madre selecciona al azar un animal para llevarlo de vacaciones. Sea A el evento de que selecciona un elefante y B el evento de que selecciona un animal que toca música y se enciende, encuentra P(A), P(B), P(A|B) y P(B|A).

Indica si los eventos A y B son eventos dependientes o independientes, luego encuentra P(A y B).

Respuesta:

Datos:

12 animales de peluche.

- 7 son elefantes.
  - 4 tocan música y se iluminan.
- 5 son osos.
  - 2 tocan música y se iluminan.

1. P(A) = 7/12 (Selecciona un elefante)
2. P(B) = 6/12 (Selecciona animal que toca música y se enciende)
3. P(A|B) = 4/6 = **2/3** (Hay 6 animales que tocan música y se iluminan, de los cuales 4 son elefantes.)
4. P(B|A) = 4/7 (Sucedido A, probabilidad de que toque música y se ilumine.)
5. A y B: Como P(A) * P(B) ≠ P(A ∩ B), concluimos que A y B son eventos dependientes.

6. Para encontrar P(A y B) (la probabilidad de que se selecciona un elefante que toca música y se enciende), podemos usar la fórmula de la intersección de eventos:

P(A y B) = P(A ∩ B). Según la información proporcionada, hay 4 elefantes que tocan música y se encienden. Entonces:

P(A ∩ B) = 4 / 12
= 1/3 ≈ 0.3333

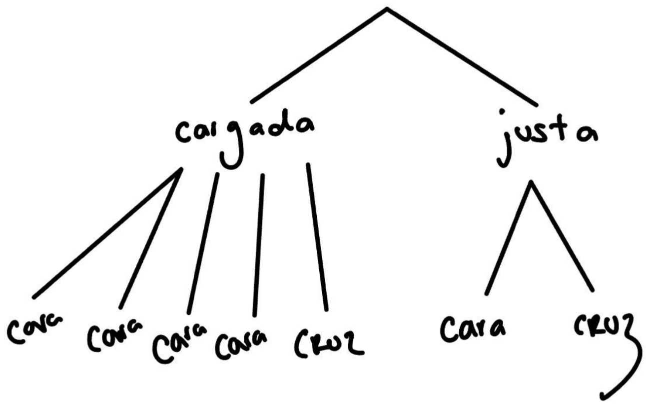
Por lo tanto, la probabilidad de que se seleccione un elefante que toca música y se enciende es de aproximadamente 0.3333 o 1/3.

Clase: Teorema de Bayes

Ejercicio 23:

Tenemos dos monedas, una justa y una cargada para caer en cara 4/5 de las veces. Sin saber qué moneda agarramos, seleccionamos una al azar, lanzamos la moneda y obtenemos cara.

¿Cuál es la probabilidad de que hayamos lanzado la moneda cargada? Completa el diagrama de árbol para resolver la pregunta.



Respuesta:

Tenemos en el lado cargado 5 posibilidades y del lado justo 2 posibilidades. Entonces:
El MCM = 10

Para que el lado cargado nos de 10, le asignamos un peso de 2 a cada posibilidad y para que el lado justo nos de 10, le asignamos un valor de 5 a cada posibilidad.

Si nos quedamos solo con los valores de "cara", entonces tenemos 2+2+2+2 = 8 del lado cargado.

P(cargado|cara) = (caras cargado) / (total caras) = (2+2+2+2) / (2+2+2+2+5).
= 8 / 13
= 0.61 = **61%**

Clase: Teorema de Bayes

Ejercicio 24:

Carlos sabe que en su escuela:

P(último año = senior)= 0.40

P(jugador futbol)=0.15

P(juega fútbol y es de último año)=0.05

Busca la probabilidad de P(último año | juega fútbol) y resume por qué el teorema de Bayes podría ser o no usado para resolver el problema.

Respuesta:

Queremos resolver P(senior|fútbol). Entonces, la fórmula es:

P(A|B) = P(B|A)*P(A) / P(B)

P(senior|fútbol) = P(futbol|senior)*P(senior) / P(futbol). Se puede decir que:

P(futbol y senior) = P(futbol | senior) * P(senior)

ya que **P(B y A) = P(B | A)*P(A).**

Entonces:

P(senior|futbol) = P(futbol y senior) / P(futbol)

= 0.05 / 0.15

= 1/3 = 33%

Clase: Combinaciones y permutaciones

Ejercicio 25:

¿Qué tan grande es la permutación de 2 en 5 comparado a su combinación?

Respuesta:

Ejercicio:

****n** = 5**

****k** = 2**

Permutación de 2 en 5

P(n, k) = n! / (n-k)!

P = 5! / (5-2)!

= 5! / 3!

= 20

Combinación de 2 en 5

C(n, k) = n! / (k! * (n-k)!)

C = 5! / 2! * (5-2)!

= 5! / 2! * 3!

= 10

Clase: Combinaciones y permutaciones

Ejercicio 26:

El equipo de baloncesto femenino de la escuela secundaria tiene 8 jugadoras, 5 de las cuales son seniors. Necesitan averiguar qué senior será la capitana y qué senior será la co-capitana. Para que sea justo, eligen a dos jugadoras de un sombrero. La primera sorteada será capitana y la segunda será co-capitana.

¿Cuántas parejas diferentes de capitana/co-capitana son posibles?

Respuesta:

$n = 5$

$k = 2$

Como el orden si importa (capitana y co-capitana), entonces:

$P(n, k) = n! / (n-k)!$

$= P(5,2) = 5! / 3!$

$= 20$

Clase: Gráficos de dispersión e introducción a la regresión

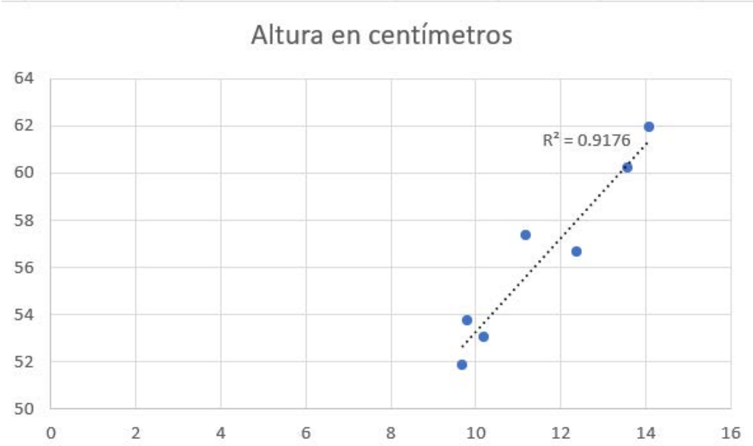
Ejercicio 27:

La siguiente tabla mide el peso en libras y la altura en centímetros a bebés y niñas de 3 meses.

Grafica los puntos de la misma en un gráfico de dispersión y describe la tendencia lo más detallado posible con base en tus observaciones.

Peso en libras	Altura en centímetros
9.7	51.84
10.2	53.04
12.4	56.64
13.6	60.24
9.8	53.76
11.2	57.360
14.1	61.92

Respuesta:



Al observar el gráfico de dispersión, podemos notar que existe una relación entre el peso y la altura de los bebés y niñas de 3 meses. A medida que aumenta el peso, generalmente también aumenta la altura, lo que indica una correlación positiva entre estas dos variables.

Sin embargo, es importante destacar que los puntos no forman una línea perfectamente recta, lo que sugiere que la relación entre el peso y la altura puede ser más compleja. Podemos observar que hay cierta dispersión en los datos, lo que indica que hay otros factores que también influyen en la altura de los bebés además del peso.

A pesar de la dispersión, se puede notar una tendencia general ascendente en el gráfico, lo que implica que en promedio, a medida que aumenta el peso, también aumenta la altura de los bebés y niñas de 3 meses.

Soluciones a los ejercicios

Ejercicio 1:

Variables:

- Número de conductor: categórica.
- Nombre del conductor: categórica.
- Tipo de automóvil que usa el conductor: categórica.
- Número de vueltas completadas: cuantitativa discreta.
- Tiempo a la diezmilésima de segundo más cercana: cuantitativa continua.
- Vueltas completadas: cuantitativa discreta.

Clasificación de las variables:

- Categóricas: Número de conductor, Nombre del conductor y Tipo de automóvil que usa el conductor.
- Cuantitativas:
  - Discretas: Número de vueltas completadas.
  - Continuas: Tiempo a la diezmilésima de segundo más cercana y distancia recorrida.

Unidades de medida:

- Número de vueltas completadas: vueltas.
- Tiempo a la diezmilésima de segundo más cercana: segundos.
- Distancia recorrida: kilómetros.

Ejercicio 2:

Esta tabla es un ejemplo de una tabla de una entrada porque solo presenta información sobre una variable, que en este caso es el sabor del helado o nieve. Cada fila representa un sabor de helado o nieve diferente, y las columnas indican la cantidad de cucharadas vendidas, si puede llevar chocolate extra y si es helado o nieve.

Aunque hay varias variables en la tabla (cantidad de cucharadas vendidas, si puede llevar chocolate extra y si es helado o nieve), todas estas variables están relacionadas con el sabor del helado o nieve y no son variables independientes. Por lo tanto, esta tabla sigue siendo un ejemplo de una tabla de una entrada.

Ejercicio 3:

Si el 40% de los perros recibió entre 25 y 40 paseos y nadie recibió más de 40, entonces el 60% restante recibió menos de 25 paseos. El 40% de 400 = 160. Entonces, 160 perros recibieron entre 25 y 40 paseos. Así que el restante recibió menos de 25 paseos: 400 – 160 = 240.

Ejercicio 4:

Colores	**Frecuencia**	**Frecuencia relativa**
1 - 5	2	13.33%
6 - 10	3	20%
11 – 15	3	20%
16 - 20	7	46.67%
Totales	15	100%

Ejercicio 5:

Semana	**Abdominales**	**Frecuencia Acumulada**	**% Acumulado**
1	355	355	7.2%
2	460	815	16.5%
3	605	1420	28.8%
4	545	1965	39.8%
5	1280	3245	65.8%
6	1690	4935	100.0%

Ejercicio 6:

La distribución marginal de la columna vertical total muestra el número total de participantes que experimentaron cada nivel de acidez estomacal, independientemente de si recibieron el analgésico o el placebo. Según la tabla, un total de 185 participantes experimentaron acidez leve, 138 experimentaron acidez intensa y 21046 no experimentaron acidez.

A partir de esta información, podemos concluir que la mayoría de los participantes del estudio no experimentaron acidez estomacal, ya que 21046 personas no informaron tener acidez. También podemos concluir que un número significativamente menor de participantes experimentaron acidez intensa en comparación con la acidez leve, ya que solo 138 personas experimentaron acidez intensa en comparación con las 185 personas que experimentaron acidez leve.

Sin embargo, la distribución marginal de la columna vertical total no nos permite sacar conclusiones sobre la eficacia del analgésico o el placebo en el tratamiento de la acidez estomacal. Es necesario examinar la distribución conjunta de las variables para obtener una comprensión completa de los resultados del estudio.

Ejercicio 7:

Para encontrar la media de un conjunto de datos, es necesario sumar todos los valores y luego dividir el resultado por el número total de valores. En este caso, el conjunto de datos es: 107, 252, 360, 424. Para encontrar la media, primero sumamos los valores:

107 + 252 + 360 + 424 = 1143

Luego, dividimos la suma por el número total de valores, que es 4 en este caso:

1143 / 4 = 285.75

Por lo tanto, la media del conjunto de datos es 285.75.

Ejercicio 8:

Para encontrar la mediana de un conjunto de datos es necesario ordenar los valores en orden ascendente o descendente. Luego encontrar el valor central o la media de los dos valores centrales si el conjunto de datos tiene un número par de elementos. En este caso, el conjunto de datos es: 74, 75, 62, 77, 70, 71, 64, 69, 70, 73.

Primero, ordenamos los valores en orden ascendente: 62, 64, 69, 70, 70, 71, 73, 74, 75, 77.

Ejercicio 9:

Primero, es necesario ordenar los datos de menor a mayor: 0, 1, 3, 3, 3, 6, 6, 7, 7, 20, 21, 21.

El rango se calcula como la diferencia entre el valor máximo y el valor mínimo. En este caso, el valor mínimo es 0 y el valor máximo es 21: Rango = 21 - 0 = 21

Para calcular el rango intercuartil (IQR), primero es necesario encontrar los cuartiles Q1 y Q3.

Q1 es el valor que divide al conjunto de datos en dos partes iguales, de modo que el 25% de los datos estén por debajo de él. Para encontrar Q1, podemos utilizar la fórmula: Q1 = (n + 1) * 0.25 para encontrar la posición del valor del Q1

Donde "n" es el número de datos en el conjunto ordenado. En este caso, "n" es igual a 12, por lo que:

Q1 = (12 + 1) * 0.25 = 3.25 como posición.

Como Q1 es un valor entre dos datos, tomamos la media de los dos valores centrales:

Q1 = (3 + 3) / 2 = 3

Q3 es el valor que divide al conjunto de datos en dos partes iguales, de modo que el 75% de los datos estén por debajo de él. Para encontrar Q3, podemos utilizar la fórmula: Q3 = (n + 1) * 0.75

En este caso, Q3 = (12 + 1) * 0.75 = 9.75. Como Q3 es un valor entre dos datos, tomamos la media de los dos valores centrales: Q3 = (7 + 20) / 2 = 13.5

Finalmente, podemos calcular el IQR como la diferencia entre Q3 y Q1:

IQR = Q3 - Q1 = 13.5 - 3 = 10.5. Por lo tanto, el rango de los datos es 21 y el IQR es 10.5.

Ejercicio 10:

Para calcular el rango de los datos, se debe restar el valor mínimo del valor máximo en el conjunto de datos.

En este caso, el valor mínimo es 20 y el valor máximo es 43, por lo tanto:

Rango = Valor máximo - Valor mínimo = 43 - 20 = 23. Por lo tanto, el rango de los datos es 23.

Ejercicio 11:

El decir que se le aumenta un 10% a cada valor significa que cada dato es multiplicado por 1.1, como vimos en las clases, al multiplicar todos los datos por un escalar también las medidas de TC y de dispersión se multiplican. Por lo que el nuevo promedio será de 65*1.1 = 71.5 y el rango de 25 * 1.1 = 27.5

Ejercicio 12:

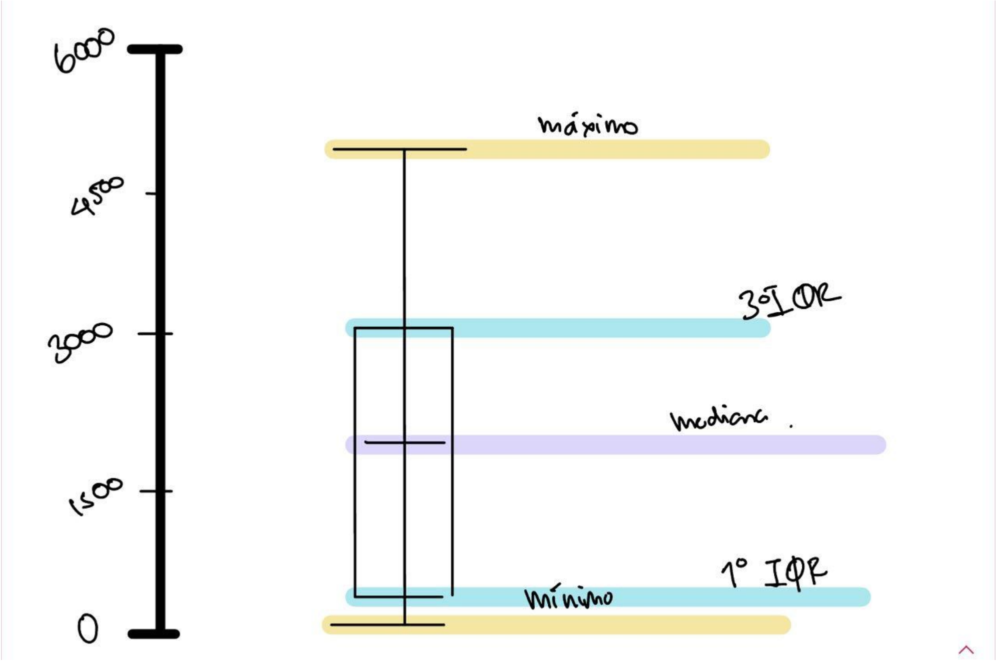
El rango se calcula restando el valor mínimo del valor máximo en el conjunto de datos:

Rango = Máximo - Mínimo = 845,300 - 216,290 = 629,010. El rango es de 629,010.

El IQR (rango intercuartil) es la diferencia entre el tercer cuartil y el primer cuartil:

IQR = Tercer cuartil - Primer cuartil = 790,370 - 324,876 = 465,494. El IQR es de 465,494.

Ejercicio 13:

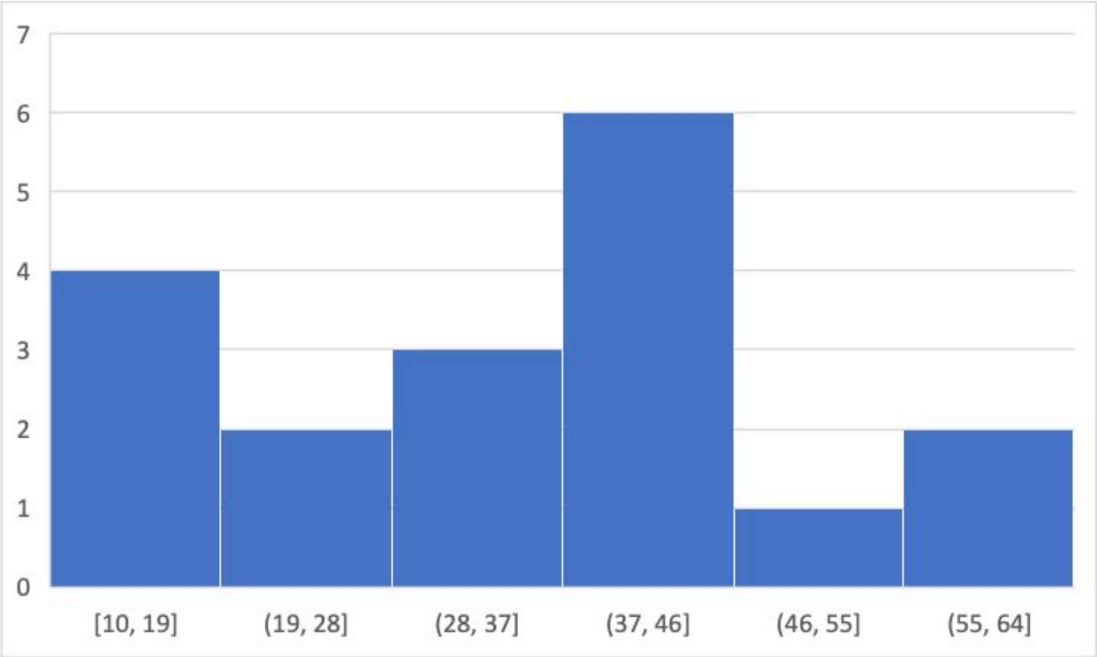


Ejercicio 14:

Si sumamos un valor a todos los datos, solo afecta a las medidas de tendencia central, por lo que el promedio aumentaría en 10 puntos de 68 a 78.

Ejercicio 15:

El diagrama de tallo y hojas cuenta de 10 en 10 a lo largo del lado izquierdo. Luego podemos ver cuántos puntos de datos caen en cada cubo, porque estamos usando cubos de tamaño 10. En otras palabras, los tallos se convierten en cubos y el número de hojas se convierte en las frecuencias graficadas en el histograma.



Ejercicio 16:

Estos datos tienen un valor atípico bajo de 0, por lo que la mejor medida de tendencia central es la mediana y la mejor medida de dispersión es el rango intercuartílico. Para ver si hay valores atípicos en los datos, use la regla 1.5-IQR.

Los valores atípicos bajos están dados por Q1 – 1.5 (IQR)

Los valores atípicos altos están dados por Q3 + 1.5 (IQR)

La mediana del conjunto de datos es 24. El primer y tercer cuartil son

Q1=(20+21)/2 = 20.5

Q3=(25+25)/2 = 25

Entonces el rango intercuartílico es

Q3 – Q1 = 25 – 20.5 = 4.5

Ahora podemos calcular dónde buscar valores atípicos.

Valores atípicos bajos:

Q1 – 1,5 (IQR) -->20.5 – 1.5(4.5) = 13.75

Valores atípicos altos:

Q3 + 1.5 (IQR) --> 25 + 1.5(4.5) = 31.75

Los datos tienen un valor atípico bajo de 0 porque es inferior a 13,75. Los datos no tienen valores atípicos altos porque ningún número en el conjunto es mayor que 31.75. Dado que los datos tienen un valor atípico, la mejor medida de tendencia central es la mediana y la mejor medida de dispersión es el rango intercuartílico.

Ejercicio 17:

El método de muestreo utilizado en la encuesta del zoológico puede estar sesgado debido a la selección de la población objetivo. Al enfocarse únicamente en las familias con hijos, se excluyen a otras posibles poblaciones objetivo, como parejas sin hijos, personas solteras o personas mayores, que podrían tener motivaciones diferentes para visitar el zoológico.

Además, la muestra obtenida puede no ser representativa de la población total de visitantes del zoológico, ya que las familias con hijos podrían ser una minoría en comparación con otros grupos demográficos, y sus motivaciones pueden no ser representativas de los motivos de los visitantes en general. Esto puede sesgar los resultados y llevar a conclusiones erróneas sobre las motivaciones de los visitantes del zoológico en su conjunto.

En resumen, el método de muestreo utilizado puede ser sesgado debido a la selección de una población objetivo limitada, lo que puede no reflejar las motivaciones de la población total de visitantes del zoológico.

Ejercicio 18:

El método de muestreo utilizado en la encuesta del dueño del restaurante puede estar sesgado debido a la naturaleza de la pregunta formulada. La pregunta "¿Has dejado a tu mesero sin propina?" se enfoca en un comportamiento específico, es decir, dejar sin propina al mesero.

Sin embargo, este tipo de pregunta puede estar sesgada debido a que puede hacer que los encuestados se sientan incómodos o avergonzados de admitir que han dejado sin propina a su mesero. Es posible que algunos encuestados no quieran admitir abiertamente este comportamiento negativo por temor a ser juzgados o percibidos como personas irrespetuosas o poco éticas.

Además, el hecho de que se haga la pregunta puede influir en el comportamiento de los comensales. Sabiendo que están siendo encuestados sobre su comportamiento de propina, pueden sentirse obligados a dejar una propina incluso si no hubieran planeado hacerlo originalmente. Esto puede generar respuestas sesgadas y poco confiables.

En resumen, la pregunta formulada en la encuesta del dueño del restaurante puede estar sesgada debido a la naturaleza delicada de la pregunta y a la posibilidad de que los encuestados se sientan incómodos o avergonzados de responder de manera honesta. Además, el hecho de que se haga la pregunta puede influir en el comportamiento de los comensales, lo que puede generar respuestas sesgadas.

Ejercicio 19:

Los eventos A y B no son eventos mutuamente excluyentes porque a veces pueden ocurrir al mismo tiempo. El problema incluso nos dice que $P(A \cap B) = 0.05$, lo que significa que hay un 5 % de posibilidades de que ambos eventos sucedan al mismo tiempo. Para encontrar $P(A \cup B)$, usaremos $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ y reemplazamos $P(A) = 0.3$, $P(B) = 0,6$ y $P(A \cap B) = 0,05$. $P(A \cup B) = 0,3 + 0,6 - 0,05$ $P(A \cup B) = 0,85$

Ejercicio 20:

Del diagrama de Venn, podemos sumar los números de cada una de las cuatro secciones para ver que Juan y Maria hicieron juntos 12 + 2 + 11 + 42 = 67 viajes cuesta arriba. Desde el 2 en el centro del diagrama de Venn donde los círculos se superponen, podemos decir que Juan y María se cayeron en 2 de los viajes cuesta arriba. Entonces, la probabilidad de que Juan se caiga y Maria se caiga es

$P(\text{caidaJuan} \cap \text{caidaMaria}) = 2/67$

Por el diagrama de Venn, sabemos que hicieron 12 viajes en los que solo se cayó Juan y 11 viajes en los que solo se cayó María. Entonces, la probabilidad de que solo Juan se caiga o solo Maria se caiga es

P(JuanCayó U MaríaCayó) = (12 + 11) / 67 = 23

Ejercicio 21:

La probabilidad de obtener una cabeza en un solo lanzamiento de una moneda equilibrada es de 0.5. Como los lanzamientos son independientes entre sí, la probabilidad de obtener 4 cabezas seguidas en 4 lanzamientos es:

P(4 cabezas seguidas) = 0.5 x 0.5 x 0.5 x 0.5 = 0.0625

Por lo tanto, la probabilidad de obtener 4 cabezas seguidas cuando lanzamos una moneda al aire cuatro veces es de 0.0625 o 6.25%.

Ejercicio 22:

Hay 7 + 5 = 12 animales de peluche en total. P(A) es la probabilidad de seleccionar un elefante, y hay 7 elefantes.

$P(A)= 7/12$

P(B) es la probabilidad de seleccionar un animal que toque música y se ilumine. Hay 4 + 2 = 6 animales que tocan música y se iluminan.

$P(B) = 6/12 = 1/ 2$

P(A|B) es la probabilidad de seleccionar un elefante, dado que el animal toca música y se enciende. Hay 4 elefantes que tocan música y se iluminan de 4 + 2 = 6 animales en total que tocan música y se iluminan.

$P(A \mid B) = 4/6 = 2/3$

P(B|A) es la probabilidad de elegir un juguete que reproduce música y se ilumina dado que el juguete es un elefante. Hay 4 elefantes que tocan música y se iluminan de un total de 7 elefantes.

$P(B \mid A) = 4/7$

Porque $P(A) \neq P(A|B)$ y $P(B) \neq P(B|A)$, ya que A y B son eventos dependientes.

P(A y B) es la probabilidad de elegir un elefante que toca música y se ilumina. Sabemos que los eventos son eventos dependientes, entonces

$P(A \text{ y } B) = P(A) \cdot P(B|A)$

$P(AyB)= 7/12 * 4/7$

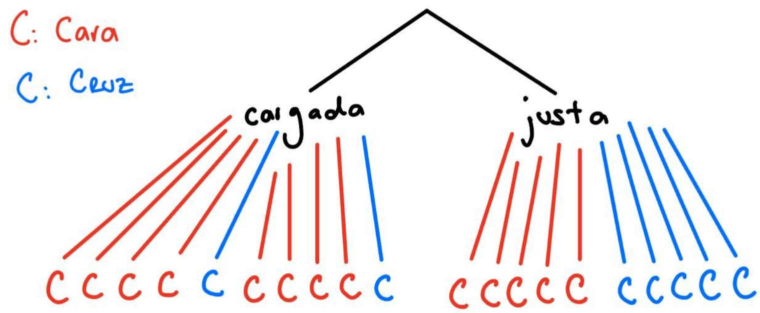
$P(AyB)= 28/84 = 4/12$

$P(A \text{ y } B) = 1/3$

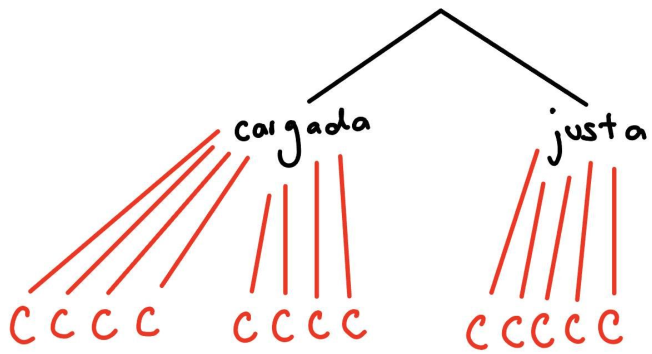
Ejercicio 23:

Estamos buscando la probabilidad de que la moneda está cargada dado que ya lanzamos cara, por lo que estamos buscando P (cargada | cara).

El siguiente paso para el diagrama de árbol es asegurarse de que las ramas estén equilibradas. Usamos fracciones equivalentes para hacer esto. Para el lado cargado, sabemos que obtenemos cara 4 de 5 veces. Esto es lo mismo que 8 de cada 10 veces. Para la moneda imparcial, obtenemos cara 1 de 2 veces, que es lo mismo que 5 de 10 veces.



Solo nos interesan las caras, así que ahora debemos podar el árbol.



Ahora buscamos la probabilidad de que hayamos lanzado la moneda cargada. 8 de las caras provinieron de la moneda cargada y 5 no.

P(cargada) = 8/(8+5) = 8/13

La probabilidad de que lancemos la moneda sesgada, sabiendo que salió cara, es 8/13.

Ejercicio 24:

Veamos si podemos usar el teorema de Bayes para encontrar la probabilidad. Primero, tomemos el teorema de Bayes y escríbalo en términos de nuestro problema. Queremos resolver la probabilidad P(senior|futbol), entonces

P(A|B) =(P(B|A)P(A)) / P(B)

P(senior|futbol) =(P(futbol|senior)P(senior)) / P(futbol)

Recuerda que la regla de la multiplicación dice que

P(B y A) = P(B | A) · P(A).

Entonces también podemos decir que

P(futbol y senior) = P(futbol | senior) · P(senior).

Entonces podemos usar el Teorema de Bayes.

P(senior|soccer) = P(soccer y senior) / P(soccer)

Ahora podemos usar la información que nos han dado para resolver el problema.

P(fútbol y sénior) = 0,05

P(jugar al fútbol) = 0,15

P(senior|fútbol) = 0.05/0.15 = 1/3

También podríamos haber usado un diagrama de Venn, en lugar del teorema de Bayes, para resolver este problema.

Ejercicio 25:

Calcularemos ambos valores, luego encontraremos la diferencia.

$$_5P_2 = \frac{5!}{(5-2)!} = \frac{5!}{3!} = \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{3 \cdot 2 \cdot 1} = 5 \cdot 4 = 20$$

$$_5C_2 = \frac{5!}{2!\,(5-2)!} = \frac{5!}{2! \cdot 3!} = \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{2 \cdot 1 \cdot 3 \cdot 2 \cdot 1} = \frac{5 \cdot 4}{2} = 10$$

$_5P_2$ es el doble de $_5C_2$

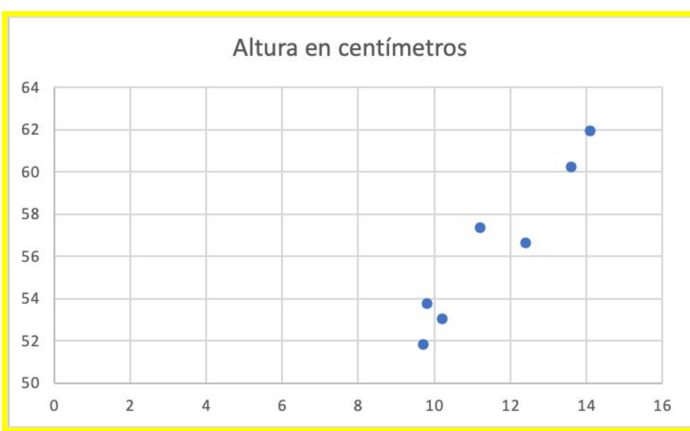
Ejercicio 26:

Como el orden importa, tenemos que calcular las permutaciones. Hay 5 mujeres entre las que podemos elegir y 2 lugares para ubicarlas.

$$_nP_k = \frac{n!}{(n-k)!} = \frac{5!}{(5-2)!} = \frac{5!}{3!} = \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{3 \cdot 2 \cdot 1} = 5 \cdot 4 = 20$$

Hay 20 parejas posibles de capitana/co-capitana.

Ejercicio 27:



Los puntos se elevan de izquierda a derecha y son bastante lineales. Podemos decir que existe una fuerte correlación lineal positiva entre los puntos. No parece haber valores atípicos en los datos.