16 de Febrero 2025

Inteligencia Artificial
Act 7: Laboratorio de Repaso de Probabilidad y Estadística

1 Tipos de datos y Medidas de tendencia central

En una empresa se han recolectado los siguientes datos de 10 empleados:

Nombre	**Edad (años)**	**Área de trabajo**
Ana	25	Ventas
Luis	30	Administración
Marta	40	Producción
Carlos	35	Ventas
Elena	28	Recursos Humanos
Juan	50	Producción
Sofía	45	Administración
Pedro	38	Ventas
Daniel	33	Producción
Laura	27	Recursos Humanos

1. Clasifique las variables en cualitativas o cuantitativas.
   - Variables Cualitativas: nombre, ya que no representa un valor numérico medible y el área de trabajo, ya que representa una categoría
   - Variables Cuantitativas: edad, ya que esta es una variable numérica que se puede medir y se pueden realizar operaciones matemáticas como sumar o restar

2. Determine la media, mediana y moda de la variable "Edad".
   Podemos empezar ordenando los valores para la edad:

   $$25, 27, 28, 30, 33, 35, 38, 40, 45, 50$$

   $$Media = \frac{\sum x_i}{n} = \frac{25 + 27 + 28 + 30 + 33 + 35 + 38 + 40 + 45 + 50}{10} = \frac{351}{10} = 35.1$$

   Por lo tanto, **la media es de 35.1 años.**

   La mediana es el valor central, pero como hay un número par de datos, la mediana es el promedio de los dos valores centrales, es decir 5 y 6.

   $$Mediana = \frac{33 + 35}{2} = 34$$

   La **mediana** de las edades es **34 años**.

   La moda es el número que más se repite, en este caso como no hay ningún valor que se repita, **no hay moda.**

3. Interprete los resultados obtenidos.
   La media nos dice que los empleados de la empresa tienen aproximadamente 35.1 años, la mediana nos indica que la mitad de los empleados están por debajo de los 34 años, mientras que la otra mitad están por arriba de los 34 años, además no hay moda, es decir no hay edades que se repitan entre los empleados, las edades no tienden a un grupo en específico.

2 Medidas de dispersión

Dado el siguiente conjunto de datos correspondiente a las calificaciones de 8 estudiantes en un examen:

$$X = \{70, 85, 90, 95, 88, 92, 75, 80\}$$

1. Calcule la varianza y la desviación estándar de los datos.

   La fórmula para calcular la varianza es la siguiente:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$$

   donde

   n = tamaño de la muestra
   x_i = valores de la muestra
   μ = media

   Primero, calculamos la media:

$$\mu = \frac{\sum x_i}{n} = \frac{70+85+90+95+88+92+75+80}{8} = \frac{675}{8} = 84.375$$

   Ahora la varianza:

$$\sigma^2 = \frac{(70-84.375)^2 + (85-84.375)^2 + (90-84.375)^2 + (95-84.375)^2 + (88-84.375)^2}{8} + ...$$

$$... + \frac{(92-84.375)^2 + (75-84.375)^2 + (80-84.375)^2}{8} = \frac{529.875}{8} = \mathbf{66.23}$$

   La desviación estándar es la raíz cuadrada de la varianza:

$$\sigma = \sqrt{66.23} \approx \mathbf{8.13}$$

2. Interprete la dispersión de los datos.
   La varianza de 66.23 nos dice que las calificaciones están relativamente dispersas con respecto a la media, mientras que la desviación estándar de 8.13 nos dice que aproximadamente las calificaciones estan 8.13 puntos por encima o por debajo del promedio.

3 Probabilidad y Teorema de Bayes

Una empresa de tecnología ha identificado que el 60% de sus empleados son programadores, y el 40% son diseñadores. Se sabe que el 70% de los programadores tienen conocimientos de inteligencia artificial (IA), mientras que solo el 30% de los diseñadores tienen estos conocimientos.

Si se elige un empleado al azar y se sabe que tiene conocimientos de IA, ¿cuál es la probabilidad de que sea programador?

El teorema de Bayes nos dice la probabilidad de un evento dado otro evento, en nuestro caso sería la probabilidad que sea programador dado que sabe IA. La fórmula es la siguiente:

$$P(A/B) = \frac{P(B/A) \cdot P(A)}{P(B)}$$

donde:

$P(A)$ = Probabilidad que sea programador = 60% = 0.6
$P(B)$ = Probabilidad que tenga conocimientos de IA
$P(B/A)$ = Probabilidad que tenga conocimientos de IA dado que es programador = 70% = 0.7

Para obtener $P(B)$ obtenemos el 70% de 60% que sería el porcentaje de que dentro del 60% de los programadores, estos sepan IA, hacemos lo mismo pero con los diseñadores que sería el 30% de 40% y lo sumamos para obtener la probabilidad de que el empleado tenga conocimientos de IA.

$$P(B) = (0.7 \cdot 0.6) + (0.3 \cdot 0.4) = 0.42 + 0.12 = 0.54$$

Sustituyendo los valores:

$$P(A/B) = \frac{(0.7)(0.6)}{0.54} \approx \mathbf{0.77}$$

$\therefore$ La probabilidad de que el empleado elegido al azar sea programador dado que tiene conocimientos de IA es aproximadamente 77%.

4 Distribuciones de probabilidad

Suponga que el número de defectos en un lote de producción sigue una distribución de Poisson con media $\lambda = 3$ defectos por lote.

1. Calcule la probabilidad de que un lote tenga exactamente 2 defectos.

   La distribución de Poisson se define como:

   $$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

   En nuestro caso tenemos $\lambda = 3$ y $k = 2$. Sustituyendo y evaluando:

   $$P(X = 2) = \frac{3^2 \cdot e^{-3}}{2!} = \frac{9e^{-3}}{2} \approx \mathbf{0.2241}$$

   $\therefore$ La probabilidad de que un lote tenga exactamente 2 defectos es aproximadamente 22.41%.

2. Calcule la probabilidad de que un lote tenga al menos 1 defecto.

   Para obtener $P(X \geq 1)$ podemos utilizar la propiedad del complemento:

   $$P(X \geq 1) = 1 - P(X = 0)$$

   Sustituyendo y calculando en la distribución de Poisson para $P(X = 0)$:

   $$P(X = 0) = \frac{3^0 \cdot e^{-3}}{0!} = \frac{1 \cdot e^{-3}}{1} \approx 0.0498$$

   Entonces:

   $$P(X \geq 1) = 1 - 0.0498 = \mathbf{0.9502}$$

   $\therefore$ La probabilidad de que un lote tenga al menos 1 defecto es aproximadamente 95.02%.

5 Funciones de densidad y distribución acumulativa

Sea X una variable aleatoria con distribución normal de media $\mu = 50$ y desviación estándar $\sigma = 10$.

1. Determine la probabilidad de que X tome un valor menor que 45.
   Para determinar $P(X < 45)$ hay que pasar de una distribución normal a una distribución normal estándar (Z), podemos hacer este proceso de estandarización usando la siguiente fórmula:

   $$Z = \frac{X - \mu}{\sigma}$$

   Sustituyendo:

   $$Z = \frac{45 - 50}{10} = \frac{-5}{10} = -0.5$$

   Teniendo la distribución normal estándar, podemos consultar la tabla de distribución normal estándar y obtener el valor para $P(X < 45) = P(Z < -0.5)$
   Entonces,

   $$P(Z < -0.5) \approx \mathbf{0.3085}$$

   $\therefore$ La probabilidad de que X sea menor que 45 es aproximadamente 30.85%.

2. Determine la probabilidad de que X esté entre 40 y 60.
   Para calcular $P(40 < X < 60)$, hacemos el proceso de estandarización.
   Para $X = 40$
   $$Z = \frac{40 - 50}{10} = \frac{-10}{10} = -1$$
   Para $X = 60$
   $$Z = \frac{60 - 50}{10} = \frac{10}{10} = 1$$
   Ahora tenemos que $P(40 < X < 60) = P(-1 < Z < 1)$, consultando la tabla de distribución normal estándar:
   $$P(Z < 1) \approx 0.8413 \qquad P(Z < -1) \approx 0.1587$$
   Entonces,
   $$P(-1 < Z < 1) = P(Z < 1) - P(Z < -1) = 0.8413 - 0.1587 = \mathbf{0.6826}$$
   $\therefore$ La probabilidad de que X esté entre 40 y 60 es aproximadamente 68.26%.

3. Use la función de distribución acumulativa para verificar sus respuestas.
   La función de distribución acumulativa de una distribución normal es:
   $$F(x) = P(X \leq x)$$
   Entonces para el primero ejercicio tenemos:
   $$F(45) = P(X < 45) = P(Z < -0.5) \approx \mathbf{0.3085}$$
   Para el segundo ejercicio tenemos:
   $$F(60) - F(40) = P(40 < X < 60) = P(Z < 1) - P(Z < -1) = 0.8413 - 0.1587 = \mathbf{0.6826}$$
   $\therefore$ Ambas respuestas coinciden con los resultados anteriores.

6 Probabilidad condicional

Un dado justo de seis caras se lanza dos veces.

1. ¿Cuál es la probabilidad de obtener un número par en el segundo lanzamiento, dado que en el primero salió un número impar?
   Al lanzar un dado justo de 6 caras podemos obtener los siguientes resultados: $\{1,2,3,4,5,6\}$. Identificando los números impares tenemos: $\{1,3,5\}$ y los pares: $\{2,4,6\}$, entonces la probabilidad de obtener un número impar es la siguiente:
   $$P(impar) = \frac{3}{6} = \frac{1}{2}$$
   Aunque sabemos que en el primer lanzamiento salió un número impar, este evento es independiente del segundo evento. Así que la probabilidad de que salga un número par en el segundo lanzamiento es:
   $$P(par) = \frac{3}{6} = \frac{1}{2}$$
   $\therefore$ La probabilidad de obtener un número par en el segundo lanzamiento, dado que en el primero salió un número impar es 50%.

2. Interprete los resultados obtenidos.
   La probabilidad de obtener un número par en el segundo lanzamiento, dado que en el primero salió un número impar es de 50% ya que ambos lanzamientos son independientes uno del otro, es decir el resultado del primer lanzamiento no afecta el segundo resultado del segundo lanzamiento.

7 Distribución binomial

Un examen de opción múltiple tiene 5 preguntas, cada una con 4 posibles respuestas, de las cuales solo una es correcta. Un estudiante responde al azar.

1. ¿Cuál es la probabilidad de que el estudiante acierte exactamente 3 respuestas?
   Se puede utilizar la distribución binomial para el problema, la fórmula es:

   $$P(X = x) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$$

   donde:

   n = número de veces que se repite el experimento = 5
   θ = la probabilidad de éxito = $\frac{1}{4}$ = 0.25
   x = número de éxitos = 3
   $1-\theta$ = probabilidad de fracaso = $\frac{3}{4}$ = 0.75

   Sustituyendo y evaluando, tenemos:

   $$P(X = 3) = \binom{5}{3}(0.25)^3(0.75)^{5-3} = \mathbf{0.0879}$$

   $\therefore$ La probabilidad de que el estudiante acierte exactamente 3 respuestas es de 8.79%.

2. ¿Cuál es la probabilidad de que acierte al menos una respuesta?
   Para obtener $P(X \geq 1)$ podemos usar el complemento de la probabilidad de que no acierte ninguna respuesta :

   $$P(X \geq 1) = 1 - P(X = 0)$$

   Calculando $P(X = 0)$

   $$P(X = 0) = \binom{5}{0}(0.25)^0(0.75)^{5-0} = 0.2373$$

   Sustituyendo tenemos:

   $$P(X \geq 1) = 1 - 0.2373 = \mathbf{0.7627}$$

   $\therefore$ La probabilidad de que acierte al menos una respuesta es de 76.27%.

8 Regla de Laplace

Una urna contiene 5 bolas rojas y 7 bolas azules. Se extrae una bola al azar.

1. Determine la probabilidad de que la bola extraída sea roja.
   El número total de eventos posibles es 5+7 = 12, tenemos 12 bolas en total que se pueden extraer, si sabemos que 5 de ellas son rojas, entonces la probabilidad de que la bola extraída sea roja es:

   $$P(roja) = \frac{bolas\ rojas}{total\ de\ bolas} = \frac{5}{12}$$

   $\therefore$La probabilidad de que la bola extraída sea roja es de $\frac{5}{12}$.

2. Si se extraen dos bolas sin reemplazo, ¿cuál es la probabilidad de que ambas sean azules?
   La probabilidad que la primera sea azul es:

   $$P(azul) = \frac{bolas\ azules}{total\ de\ bolas} = \frac{7}{12}$$

   Si ya sacamos una bola, las bolas totales disminuyen en 1, así que ahora nuestro total de bolas es de 11, también el número de bolas azules disminuyen en 1, así que ahora quedan 6. Entonces, la probabilidad de que la segunda sea azul es:

   $$P(azul\_2) = \frac{6}{11}$$

La probabilidad de que ambas sean azules es el producto de ambas probabilidades:

$$P(Ambas\_azules) = \frac{7}{12} \cdot \frac{6}{12} = \frac{7}{22}$$

$\therefore$ La probabilidad de que ambas bolas extraídas sean azules es de $\frac{7}{22}$.

9 Esperanza matemática

Suponga que una persona juega una lotería donde el premio es de 1000 dólares con una probabilidad de 0.01, y el costo del boleto es de 10 dólares.

1. Calcule la esperanza matemática de la ganancia del jugador.
   Si la persona gana el premio, su total sería de 990, considerando el costo del boleto, y la probabilidad de que gane es de 0.01.
   Sin embargo, si la persona no gana el premio su total sería de -10 ya que no recupero lo invertido en el boleto y la probabilidad de que pierda es de 0.99.
   La esperanza matemática de la ganancia del jugador sería sumar las ganancias multiplicadas por la probabilidad de que eso suceda, es decir:

   $$E(X) = (990 \cdot 0.01) + (-10 \cdot 0.99) = 9.9 - 9.9 = \mathbf{0}$$

   $\therefore$ La esperanza matemática de la ganancia del jugador es de 0.

2. Interprete el resultado obtenido.
   La esperanza matemática de la ganancia es de 0, es decir, el jugador por más veces que juegue no tiene ventajas ni desventajas, es decir la cantidad promedio de dinero que ganaría por jugada es de 0, porque por cada vez que el jugador gane, ya habría perdido muchas veces, por lo que su ganancia como tal no existirá.

10 Ley de los grandes números

Un experimento consiste en lanzar una moneda justa 1000 veces y calcular la frecuencia relativa de obtener cara.

1. ¿Cuál es el valor esperado de la frecuencia relativa de obtener cara?
   La probabilidad de obtener cara es de 0.5, ya que estamos hablando de un moneda justa. La frecuencia relativa de obtener cara es el número de veces que obtuvimos cara entre el número total de lanzamientos.

   $$Frecuencia\ relativa = \frac{X}{1000}$$

   Entonces el valor esperado de la frecuencia relativa sería:

   $$E(Frecuencia\ relativa) = \frac{E(X)}{E(1000)} = \frac{1000 \cdot 0.5}{1000} = \mathbf{0.5}$$

   $\therefore$ El valor esperado de la frecuencia relativa de obtener cara es 0.5.

2. ¿Cómo se relaciona esto con la Ley de los Grandes Números?
   La Ley de los Grandes Números nos dice que si repetimos muchas veces un mismo experimento, la frecuencia de que suceda cierto evento tiende a acercarse a la probabilidad de que ocurra este evento, es decir el valor esperado, en nuestro caso tenemos que se repitió muchas veces el experimento de lanzar una moneda, entonces la Ley de los Grandes Números nos dice que la frecuencia relativa se acercará mucho a 0.5.