Resumen-Tema-1.pdf

ferluque

ESTADÍSTICA DESCRIPTIVA E INTROD A LA PROBABILIDAD

1º Doble Grado en Ingeniería Informática y Matemáticas

Facultad de Ciencias
Universidad de Granada

Tema 1

Conceptos básicos

- Fenómeno determinístico: mismas condiciones $=$ mismos resultados
- Fenómeno aleatorio: mismas condiciones $\neq$ mismos resultados
- Población: conjunto de unidades con alguna(s) característica(s) en común, sobre la que obtener información
- Muestra: subconjunto de la población
- **Carácter**: propiedad a estudiar (debe observarse en todos los individuos de la **población**):
    - *Modalidad*: "Variantes" del carácter
    - Puede ser *cualitativo* o *cuantitativo*
- **Escalas de medida**:
    - *Nominal*: $x_A = x_b$ ó $x_A \neq x_b$
    - *Ordinal*: Nominal + $x_A < x_b$ ó $x_A > x_B$
    - *De intervalo*: Ordinal + A es $x_A - x_B$ unidades diferente que B
    - *De razón*: Ordinal + A es $x_A - x_B$ veces superior (o inferior) a B

Variables

Pueden ser **discretas** (el paso de un valor a otro representa un salto) o **continuas** (pueden tomar cualquier valor posible entre dos valores)

Frecuencia absoluta del valor x_i (n_i): Número total de individuos de la población que presenta el valor x_i

Frecuencia relativa del valor x_i (f_i): Proporción de individuos que presentan el valor x_i

Frecuencia absoluta acumulada del valor x_i (N_i): Número total de individuos de la población que presentan un valor menor o igual que x_i

Frecuencia relativa acumulada del valor x_i (F_i): Proporción de individuos que presentan un valor menor o igual que x_i

La **distribución de frecuencias de una variable estadística unidimensional** serán los pares (modalidad, frecuencia) $\{(x_i, n_i \dots); i = 1, \dots, k\}$

Tablas estadísticas

Variables discretas y atributos

Modalidades	Frec. Abs.	Frec. Rel.	Frec. Abs. Acum.	Frec. Rel. Acum.
x_1	n_1	f_1	$N_1 = n_1$	$F_1 = f_1$
x_2	n_2	f_2	$N_2 = n_1 + n_2$	$F_2 = f_1 + f_2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_i	n_i	f_i	$N_i = n_1 + n_2 + \cdots + n_i$	$F_i = f_1 + f_2 + \cdots + f_i$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_k	n_k	f_k	$N_k = n_1 + n_2 + \cdots + n_k = n$	$F_k = f_1 + f_2 + \cdots + f_k = 1$

Variables continuas

Intervalos	Marcas	Amplitud	Frec. Abs.	Frec. Rel.	Frec. Acum.
$(e_0, e_1]$	$c_1 = \frac{e_0+e_1}{2}$	$a_1 = e_1 - e_0$	n_1	f_1	$N_1 = n_1$
$(e_1, e_2]$	$c_2 = \frac{e_1+e_2}{2}$	$a_2 = e_2 - e_1$	n_2	f_2	$N_2 = n_1 + n_2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$(e_{i-1}, e_i]$	c_i	a_i	n_i	f_i	$N_i = n_1 + n_2 + \cdots + n_i$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$(e_{k-1}, e_k]$	c_k	a_k	n_k	f_k	$N_k = n_1 + n_2 + \cdots + n_k = n$

Representaciones gráficas

Atributos

- Diagrama de sectores
- Diagrama de rectángulos (o barras)
- Pictograma

Variables discretas

- Diagrama **de barras** (rectángulos para las continuas) (n_i en función de x_i)
- Curva acumulativa o de distribución (o función de distribución) (N_i en función de x_i; en este caso la función será escalonada, y valdrá el valor $N_i \ \forall x \in [x_i, x_{i+1})$). Además, su dominio es $\mathbb{R}$ y vale $0 \ \forall x < x_1$ y $1 \ \forall x \geq x_k$

Variables continuas

- **Histograma**: Rectángulos de base $a_i = x_{i+1} - x_i$, y altura $h_i = \frac{n_i (o\ f_i)}{a_i}$
- **Poligonal de frecuencias**: Unir los puntos centrales de los techos de las columnas del histograma
- **Curva acumulativa o de distribución (o función de distribución)**:

  $F(e_i) = \sum_{j=1}^{i} f_j$

  Para ello suponemos equidistribución de frecuencias en el intervalo. De nuevo su dominio es $\mathbb{R}$ y $F(e) = 0 \ \forall x < x_0$ y $F(e) = 1 \ \forall x \geq x_k$


12
MARVEL STUDIOS
MS MARVEL
Serie Original
8 de junio solo en
Disney+
© 2022 MARVEL. Requiere suscripción. Se aplican términos y condiciones.

Medidas de posición

Media aritmética

$\bar{x} = \frac{1}{n} \sum_{i=1}^{k} x_i n_i$

Si es continua, $x_i = c_i = \frac{e_i - e_{i-1}}{2}$

$x_1 < \bar{x} < x_k$

Media de las desviaciones respecto de la media es 0

Si $Y = aX + B \Rightarrow \bar{y} = a\bar{x} + b$

Media de los cuadrados de las desviaciones es mínima respecto de la media y es la varianza σ_x^2

Media geométrica

Para variables que tienen efectos acumulativos en el tiempo (como el interés)

$G = \sqrt[n]{\prod_{i=1}^{k} x_i^{n_i}}$

Logaritmo de G es la media aritmética de los logaritmos de la variable

$log\, G = \sum_{i=1}^{k} f_i log\, x_i$

Media armónica

Magnitudes que son cocientes de dos magnitudes (velocidad). **Inversa de la media aritmética de los valores inversos de la variable**

$H = \frac{n}{\sum_{i=1}^{k} \frac{n_i}{x_i}}$

Media cuadrática

Promedios sobre superficies

$Q = \sqrt{\sum_{i=1}^{k} f_i x_i^2}$

Mediana

Valor de la variable que divide a la población en dos partes iguales.

Variables discretas

$x_i \ / \ N_{i-1} < \frac{n}{2} \leq N_i$

- Si $N_i > \frac{n}{2} \Rightarrow Me = x_i$
- Si $N_i = \frac{n}{2} \Rightarrow Me = \frac{x_{i+1} - x_i}{2}$

Variables continuas

$I_i = (e_{i-1},\ e_i] \ / \ N_{i-1} < \frac{n}{2} < N_i$

- Si $N_i > \frac{n}{2} \Rightarrow$ Se interpolará según **la curva de distribución**
- Si $N_i = \frac{n}{2} \Rightarrow \ Me = e_i$

Moda

Valor que más se repite

Variables discretas: El que tenga mayor n_i

Variables continuas: En el intervalo que mayor h_i tenga, interpolamos según el histograma

Percentiles

El percentil de orden r, es un valor de la variable P_r tal que el $r\%$ de los individuos presentan un valor de la variable menor o igual que P_r

Se hace exactamente igual que con la **mediana**, solo que sustituimos $\frac{n}{2}$ por $\frac{nr}{100}$

Medidas de dispersión

Absolutas

- Recorrido: $R = x_k - x_1$
- Recorrido intercuartílico: $RI = Q_3 - Q_1$
- Desviación absoluta media respecto a $\bar{x}$: $D_{\bar{x}} = \frac{\sum_{i=1}^{k} |x_i - \bar{x}| n_i}{n}$
- Desviación absoluta media respecto a Me: $D_{Me} = \frac{\sum_{i=1}^{k} |x_i - Me| n_i}{n}$
- Varianza: $\sigma_x^2 = \frac{\sum_{i=1}^{k} (x_i - \bar{x})^2 n_i}{n}$
  Siempre positiva. Mínima dispersión cuadrática. Si $Y = aX + b \Rightarrow \ \sigma_y^2 = a^2 \sigma_x^2$
- Desviación típica: $\sigma_x = +\sqrt{\sigma_x^2}$

  Si $Y = aX + b \Rightarrow \ \sigma_y = |a| \sigma_x$

  $D_{Me} < D_{\bar{x}} < \sigma_x$

Relativas

- Coeficiente de apertura: $C_A = \frac{x_k}{x_1}$
- Recorrido relativo: $R_R = \frac{x_k - x_1}{\bar{x}}$
- Recorrido semi-intercuartílico: $R_{SI} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$
- Coeficiente de variación de Pearson: $CV(X) = \frac{\sigma_x}{|\bar{x}|}$
- Índice de dispersión respecto a la mediana: $V_{Me} = \frac{D_{Me}}{Me}$

Momentos

Sea $r \in \mathbb{N}$ se llama **momento de orden r respecto al valor "a"** a la cantidad:

${}_a m_r = \sum_{i=1}^{k} f_i (x_i - a)^r$

No centrales: $a = 0$

$m_r = \sum_{i=1}^{k} f_i x_i^r$

Centrales: $a = \bar{x}$

$\mu_r = \sum_{i=1}^{k} f_i (x_i - \bar{x})^r$

Relaciones

$\mu_2 = \sigma_x^2 = m_2 - m_1^2$

$\mu_3 = m_3 - 3m_2m_1 - 2m_1^3$

$\mu_4 = m_4 - 4m_3m_1 + 6m_1^2m_2 - 3m_1^4$

$m_2 = \mu_2 - m_1^2$

...

Medidas de asimetría

Coeficiente de asimetría de Fisher

$\gamma_1(X) = \frac{\mu_3}{\sigma_x^3} = \sum_{i=1}^{k} (\frac{x_i - \bar{x}}{\sigma_x})^3$

- Si $\gamma_1(X) > 0$: Asimétrica por la derecha
- Si $\gamma_1(X) = 0$: Es simétrica
- Si $\gamma_1(X) < 0$: Asimétrica por la izquierda

Coeficiente de asimetría de Pearson

$A_p = \frac{\bar{x} - Mo}{\sigma_x}$; $A_p^* = \frac{3(\bar{x} - Me)}{\sigma_x}$

Con la misma interpretación que $\gamma_1(X)$

Medidas de forma

Coeficiente de curtosis de Fisher

$\gamma_2(X) = \frac{\mu_4}{\sigma_x^4} - 3$

- Platicúrtica si $\gamma_2(X) < 0$
- Mesocúrtica si $\gamma_2(X) = 0$
- Leptocúrtica si $\gamma_2(X) > 0$

Coeficiente de curtosis de Kelley

$K = \frac{1}{2} \frac{Q_3 - Q_1}{D_9 - D_1} - 0,263$

Igual que el coeficiente de Fisher