Estadística

11-6-2019

Tema 1

Variables
- Discretas
- Continuas

Posición

Cuartiles
- $Q_1 \rightarrow l_{i-1} + \frac{N/4 - N_{i-1}}{n_i} \cdot a_i$
- $Q_2 \rightarrow$ mediana
- $Q_3 \rightarrow l_{i-1} + \frac{3N/4 - N_{i-1}}{n_i} \cdot a_i$

Centralización

$\bar{x}$ = media = $\bar{x} = \frac{\sum x_i n_i}{N}$

$Mo = l_{i-1} + \frac{h_i - h_{i-1}}{(h_i - h_{i-1}) + (h_i - h_{i+1})} \cdot a_i$

$h_i = \frac{n_i}{a_i}$

$Me = l_{i-1} + \frac{N/2 - N_{i-1}}{n_i} a_i$

Deciles $= D_K = l_{i-1} + \frac{\frac{K \cdot N}{10} - N_{i-1}}{n_i} a_i$

Percentiles $= P_K = l_{i-1} + \frac{\frac{K \cdot N}{100} - N_{i-1}}{n_i} \cdot a_i$

Dispersión

Var $\rightarrow \sigma^2 = \frac{\sum x_i^2 n_i}{N} - \bar{x}^2$

Desviación típica $\rightarrow \sigma = \sqrt{var} = \sqrt{\frac{\sum x_i^2 n_i}{N} - \bar{x}^2}$

Rango $\rightarrow$ Valor máximo - valor mínimo

Rango intercuartílico $\rightarrow Q_3 - Q_1$

Coeficiente de variación $C.V. = \frac{\sigma}{\bar{x}}$

$\hookrightarrow$ Expresa el grado de homogeneidad de una muestra. + mas bajo mas homogéneo

Relación 1

$h_i = \frac{n_i}{a_i}$

8.

	salario	x_i	n_i	N_i	h_i	$x_i n_i$	$x_i^2 n_i$
amplitud $a_i = 300$	300-600	450	13	13	0,043̂	5850	2.632.500
$a_i = 400$	600-1000	800	15	28	0,0375	12000	9.600.000
$a_i = 500$	1000-1500	1250	20	48	0,04	25000	31.250.000
$a_i = 300$	1500-1800	1650	8	56	0,026̂	13200	21.780.000
$a_i = 700$	1800-2500	2150	4	60	0,005714	8600	18.490.000
			60			64650	83.752.500

media $\rightarrow \bar{x} = \frac{\sum x_i n_i}{N} = \frac{64650}{60} = 1078$

moda $\rightarrow l_{i-1} + \frac{h_i - h_{i-1}}{(h_i - h_{i-1}) + (h_i - h_{i+1})} \cdot a_i =$

$=$ ——— 8

$$= 1000 + \frac{0{,}04 - 0{,}0375}{(0{,}04 - 0{,}0375) + (0{,}04 - 0{,}02\hat{6})} \cdot 500 = 1078{,}61$$

mediana $\rightarrow$ $Me = l_{i-1} + \frac{N/2 - N_{i-1}}{n_i} a_i = 1000 + \frac{30 - 28}{20} \cdot 500 = 150$

$$Q_1 = 600 + \frac{15 - 13}{15} \cdot 400 =$$

$$P_{63} = 100 + \frac{37{,}8 - 28}{20} \cdot 500 =$$

$$\sigma = \sqrt{\frac{\Sigma x_i^2 n_i}{N} - \bar{x}^2}$$

$$\sigma = \sqrt{\frac{83.757.500}{60} - (1078)^2}$$

$$\sigma = 483{,}52$$

$$CV = \frac{\sigma}{\bar{x}} = \frac{483{,}52}{1078} = 0{,}4485$$

rango = 2800 − 300

8 d) 3) calcular en que percentil está 1700

$$1700 = 1500 + \frac{\frac{k \cdot 60}{100} - 48}{8} \cdot 300$$

$$200 = \frac{\frac{k \cdot 60}{100} - 48}{8} \cdot 300$$

$$\frac{1600}{300} + 48 = \frac{k \cdot 600}{100} \rightarrow k = 88{,}\hat{8} \rightarrow 89\%$$

Tema 2

covarianza $= \sigma_{xy} = \frac{\Sigma\Sigma x_i y_j n_{ij}}{N} - \bar{x} \cdot \bar{y}$

coeficiente de correlación $\rightarrow$ $r = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$ $\rightarrow$ $-1 \leq r \leq +1$

$0 \leq R \leq 1 \rightarrow r^2 = R$

$(x-\bar{x}) = \frac{\sigma xy}{\sigma y^2} \cdot (y-y^2)$

$(y-y^2) = \frac{\sigma xy}{\sigma x^2} \cdot (x-\bar{x})$

$y = a + bx$

$b = \frac{\sigma xy}{\sigma_x{}^2}$ $\quad a = \bar{y} - b\bar{x}$

Relación 2

8

x\y	25-55	55-65	65-95
5	4	2	2
10	1	4	0
15	5	2	0

X	n_i	N_i	$x_i n_i$	$x_i^2 n_i$
5	8	8	40	200
10	5	13	50	500
15	7	20	105	1575
	20		195	2275

I_i	Y_i	n_i	$x_i n_i$	$x_i^2 n_i$
25-55	40	10	400	16000
55-65	60	8	480	28800
65-95	80	2	160	12800
			1040	57600

$\bar{x} = \frac{195}{20}$ 9,75

$\bar{y} = \frac{1040}{20} = 52$

$\sigma_x = \sqrt{\frac{2275}{20} - 9,75^2} = 4,32$

$\sigma y = \sqrt{\frac{57600}{20} - 52^2} = 13,26$

$CV = \frac{4,32}{9,75} = 0,443$

$CV = \frac{13,26}{52} = 0,255$

covarianza

$800 + 600 + 800 \; 400 + 2400 + 3000 + 1800 = 9800$

$\sigma xy = \frac{9800}{20} - (9,75)\cdot(52) = -17$

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{-17}{(4{,}32)(13{,}26)} = -0{,}28... \simeq -0{,}3$$

x \ y	0-3	3-5	5-8
15-20	5	1	0
20-40	0	5	1
40-50	1	4	7
50-70	0	0	1

x = edad de los clientes

y = horas de utilización

a) Edad más frecuente entre los que usa el software más de 3 h

b) Entre los menores de 40 años que tanto por ciento usan el software menos de 3,5 h semanales

a)

	n_i	h_i
15-20	1	1/5
20-40	6	6/20
40-50	11	11/10
50-70	1	1/20

$$Mo_x = 40 + \frac{\frac{11}{10} - \frac{6}{20}}{\left(\frac{11}{10} - \frac{6}{20}\right) + \left(\frac{11}{10} - \frac{1}{20}\right)} \cdot 10 =$$

$$= 40 + \frac{16}{36} \cdot 10 = 44{,}\hat{4}$$

b)

	n_i	N_i
0-3	5	5
* 3-5	6	11
5-8	1	12
	12	

$$3{,}5 = 3 + \frac{\frac{K \cdot 12}{100} - 5}{6} \cdot 2$$

$$0{,}5 = \frac{\frac{K \cdot 12}{100} - 5}{6} \cdot 2$$

$$1{,}5 = \frac{K \cdot 12}{100} - 5$$

$$6{,}5 = \frac{K \cdot 12}{100} \to \frac{6{,}5 \cdot 100}{100} = K$$

$$K = 54{,}16\%$$

Estadística 13-6-19

Probabilidad

Espacio muestral

6

$P(A) = \dfrac{\text{Casos favorables}}{\text{casos totales}}$ $E = \{1,2,3,4,5,6\}$

sea par $\rightarrow A = \{2,4,6\} = 3/6$

$\geq 5 \rightarrow B = \{5,6\} = 2/6$

$P(A \underset{\text{ó}}{\cup} B) = \{2,4,5,6\} = 4/6$ $P(A \underset{\text{y}}{\cap} B) = \{6\} = 1/6$

$P(A \cap B) = P(A) + P(B) + P(A \cup B) = \frac{3}{6} + \frac{2}{6} - \frac{4}{6} = \frac{1}{6}$

$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{3}{6} + \frac{2}{6} - \frac{1}{6} = \frac{4}{6}$

$P(A/B) = \dfrac{P(A \cap B)}{P(B)}$ → probabilidad de que ocurra A sabiendo que ocurre B

$P(B/A) = \dfrac{P(A \cap B)}{P(A)}$

$P(\bar{A}) = 1 - P(A)$

$P(\bar{A} \cap \bar{B}) = P(\overline{A \cup B}) = 1 - P(A \cup B)$

$P(\bar{A} \cup \bar{B}) = P(\overline{A \cap B}) = 1 - P(A \cap B)$

$P(\bar{A} \cap B) = P(B) - P(A \cap B)$

$P(A \cap \bar{B}) = P(A) - P(A \cap B)$

~~Relación 3~~ otra relación

~~8~~

$P(A) = 0{,}4$

$P(E) = 0{,}6$

$P(A \cap E) = 0{,}15$

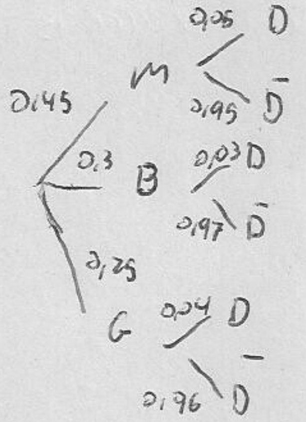
$P(A \cup E)$?

$P(A \cup E) = 0{,}4 + 0{,}6 - 0{,}15 = 0{,}85$

$P(\bar{A}) = 1 - 0{,}4 = 0{,}6$

$P(\bar{A} \cap \bar{E}) = P(\overline{A \cup E}) = 1 - P(A \cup E) =$

$= 1 - 0{,}85 = 0{,}15$



$P(D) = P(M \cap D) + P(B \cap D) + P(G \cap D)$

$P(D) = 0{,}45 \cdot 0{,}05 + 0{,}3 \cdot 0{,}03 + 0{,}25 \cdot 0{,}04 = \cancel{0{,}3395}$ $0{,}0415$

$P(G/D) = \dfrac{P(G \cap D)}{P(D)} = \dfrac{0{,}25 \cdot 0{,}04}{0{,}0415} = 0{,}2409$

12 otra relación

$P(C) = 0{,}05$

$P(D) = 0{,}14$

$P(C \cap D) = 0{,}01$

$P(C \cup D) = 0{,}05 + 0{,}15 - 0{,}015 = 0{,}18$

$P(\bar{D}) = 1 - 0{,}14 = 0{,}86$

$P(\bar{C} \cap \bar{D}) = P(\overline{C \cup D}) = 1 - P(C \cup D) = 1 - 0{,}18 = 0{,}82$

$$P(C/D) = \frac{P(C \cap D)}{P(D)} = \frac{0{,}01}{0{,}14} = 0{,}0714$$

$P(A/B) = P(A)$

$P(A) \cdot P(B) = P(A \cap B) \rightarrow$ si se cumple son independientes

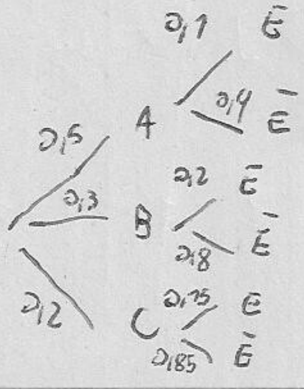
$P(C \cap D) = 0{,}05 \cdot 0{,}14 = 0{,}007 \neq P(C \cap D)$

$P(A) + P(B) = P(A \cup B) \rightarrow$ incompatibles

$P(A \cap B) = \emptyset$

Relación 3

17



$P(E) = P(A \cap E) + P(B \cap E) + P(C \cap E)$

$P(E) = 0{,}5 \cdot 0{,}1 + 0{,}3 \cdot 0{,}2 + 0{,}2 \cdot 0{,}15 = 0{,}14$

$P\left(\frac{A}{\bar{E}}\right) + P\left(\frac{B}{\bar{E}}\right) = P\left(\frac{A \cup B}{\bar{E}}\right) =$

$$= \cancel{0,5} \cdot \frac{P(A \cap \bar{E})}{P(\bar{E})} + \frac{P(B \cap \bar{E})}{P(\bar{E})} = \frac{0,5 \cdot 0,9}{0,86} +$$

$$+ \frac{0,3 \cdot 0,8}{0,86} =$$

Tema 4: Distribuciones de probabilidad

→ Discretas → masa de probabilidad

$P\{x=0\}\ 0,2$

$P\{x=1\} = P\{x=2\} = 0,2$

$P\{x=3\} = 0,2$

$$F(x) = \begin{cases} 0 & x < 0 \\ 0,2 & 0 \le x < 1 \\ 0,5 & 1 \le x < 2 \\ 0,8 & 2 \le x < 3 \\ 1 & x \geq 3 \end{cases}$$

$P[x > 2,2]$?

$P[P > 2,2] = 1 - P[x \le 2,2] = 1 - 0,8 = 0,2$

Esperanza

→ $E[x] = \sum x_i P_i$

$E[x] = 0 \cdot 0,2 + 1 \cdot 0,3 + 2 \cdot 0,3 + 3 \cdot 0,2 = 1,2$

$Var[x] = E[x^2] - (E[x])^2 = 3,3 - (1,5)^2 = 1,05$

$E[x^2] = 0^2 \cdot 0,2 + 1^2 \cdot 0,3 + 2^2 \cdot 0,3 + 3^2 \cdot 0,2 = 3,3$

Distribuciones continuas

Función de densidad

$f(x) =$ $\int_a^b f(x) = 1$

$a \leq x \leq b$

$$f(x) = \begin{cases} kx & 0 \leq x \leq 2 \\ 0 & x < 0 \quad x > 2 \end{cases}$$

averiguar k para que f(x) sea una función de densidad

$$\int_0^2 k\,x\,dx = 1 \rightarrow k \int_0^2 x\,dx \left[k \frac{x^2}{2} \right]_0^2$$

$$\frac{2^2}{2} k - \frac{0^2}{2} = 1 \rightarrow 2k = 1 \rightarrow k = 1/2$$

$$f(x) = \begin{cases} 1/2\, x & 0 \leq x \leq 2 \\ 0 & x < 0 \quad x > 2 \end{cases}$$

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{x^2}{4} & 0 \leq x \leq 2 \\ 1 & x > 2 \end{cases}$$

$$\int \frac{1}{2} x\,dx \rightarrow \frac{x^2}{4}$$

$$P[X \leq 1{,}2] = \frac{1{,}2^2}{4} = 0{,}36$$

$$P[X \geq 0{,}8] = 1 - \frac{0{,}8}{4} = 0{,}84$$

$$P[1 < X < 1{,}5] = P[X \leq 1{,}5] - P[X \leq 1] = \frac{1{,}5^2}{4} - \frac{1^2}{4} = 0{,}3125$$

$$E[X] = \int_a^b x\, f(x)\, dx$$

$$E[X] = \int_0^2 x \cdot \tfrac{1}{2} x\, dx \to \int_0^2 \frac{x^2}{2}\, dx \to \frac{1}{2}\int_0^2 x^2 dx \to \left[\frac{x^3}{6}\right]_0^2$$

$$E[X] = \frac{8}{6} - 0 = \frac{4}{3} = 1{,}\bar{3}$$

$$E[X^2] = \int_a^b x^2\, f(x)\, dx$$

$$E[X^2] = \int_0^2 x^2\, \tfrac{1}{2} x\, dx \to var = \tfrac{1}{2}\int_0^2 x^3\, dx$$

$$\frac{1}{2}\int_0^2 x^3\, dx \to \left[\frac{x^4}{8}\right]_0^2 = 2$$

$$var = E[X^2] - (E[X])^2 = 2 - (1{,}33)^2 = 0{,}231$$

otra relación

una variable aleatoria que representa la proporción de accidente automovilísticos fatales en USA tiene la siguiente función de densidad

$$f(x) = \begin{cases} 42x(1-x)^5 & 0 < x \leq 1 \\ 0 & \text{En otro caso} \end{cases}$$

Demostrar que f es una función de densidad

$$\int 42x(1-x)^5 \, dx = 1 - 7x(1-x)^6 - (1-x)^7$$

abandonamos el ejercicio . . . . ..

(5)

$$F(x) \begin{cases} 0 & x \leq 2 \\ \frac{(x-2)^3}{8} & 2 < x < 4 \\ 1 & 4 \leq x \end{cases}$$

función de distribución

sacar función de probabilidad

$$f(x) \begin{cases} \frac{3(x-2)^2}{8} & 2 < x < 4 \\ 0 & \text{otro caso} \end{cases}$$

$P[x \geq 3]$ $1 - P[x < 3] - 1 - \left(\frac{1}{8}\right) = 0{,}875$

$P[x > 4] = 1 - P[x < 4] = 1 - 1 = 0$

$$F(x)\begin{cases} \dfrac{(7+x)}{K} & -7 < x \leq 0 \\ \dfrac{7-x}{K} & 0 < x \leq 7 \\ 0 & \text{en otro caso} \end{cases}$$

$$\int_{-7}^{0} \frac{7+x}{K}\,dx + \int_{0}^{7} \frac{7-x}{k}\,dx = 1 =$$

$$= \frac{1}{k}\left[7x + \frac{x^2}{2}\right]_{-7}^{0} + \frac{1}{k}\left[7x - \frac{x^2}{2}\right]_{0}^{7} =$$

$$\underset{0}{\cancel{F(0)}} - F(-7) + F(7) - \underset{0}{\cancel{F(0)}}$$

$$= \frac{1}{k}\left[\left(49 - \frac{49}{2}\right) - \left(49 + \frac{49}{2}\right)\right] = 1$$

$$\frac{1}{k}\left[98 - 49\right] = 1 \rightarrow \frac{49}{k} = 1 \quad K = 49$$

$$F(x)\begin{cases} 0 & ; x \leq -7 \\ \dfrac{x}{7} + \dfrac{x^2}{98} + \dfrac{C}{49} & ; -7 < x < 0 \\ \dfrac{x}{7} - \dfrac{x^2}{98} + \dfrac{C'}{49} & ; 0 < x \leq 7 \\ 1 & ; x > 7 \end{cases}$$

7