RESUMEN DESCRIPTIVA

https://wpd.ugr.es/~bioestad/bioestadistica/tema-1/

TABLA DISTRIBUCION DE FRECUENCIAS VARIABLE CUANTITATIVA

I_i	x_i	n_i	f_i	N_i	F_i	$a_i = e_i - e_{i-1}$	$h_i = n_i / a_i$
$(e_0, e_1]$	x_1	n_1	f_1	N_1	F_1	$a_1 = e_1 - e_0$	h_1
$(e_1, e_2]$	x_2	n_2	f_2	N_2	F_2	$a_2 = e_2 - e_1$	h_2
$(e_2, e_3]$	x_3	n_3	f_3	N_3	F_3	$a_3 = e_3 - e_2$	h_3
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
$(e_{k-1}, e_k]$	x_k	n_k	f_k	$N_k = n$	$F_k = 1$	$a_k = e_k - e_{k-1}$	h_k
		n	1				

Marca de clase (x_i): Punto medio del intervalo $x_i = \dfrac{e_{i-1} + e_i}{2}$

Amplitud del intervalo: Diferencia entre los extremos del intervalo

Densidad de frecuencia: $h_i = \dfrac{n_i}{a_i}$

Medidas de Posición

- MEDIA
- MEDIANA: cuantiles
- MODA

Medidas de Posición

- MEDIA

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

Medidas de Posición: cuantiles: MEDIANA

- MEDIANA (n/2): CUANTILES: Percentiles, Deciles, Cuartiles …

- **Cálculo para variables continuas:** De igual forma que la mediana pero usamos **αn** en vez de **n/2** (equivalentemente α por ½)
- El intervalo que contiene a $C(\alpha)$ será aquel $(I_{i-1}-I_i]$ cuyo $N_i>\alpha n$, o equivalentemente, aquel x_i cuyo $F_i>\alpha$.
- La expresión para el cálculo sería:

$$\boxed{C(\alpha) = I_{i-1} + \frac{\alpha \cdot n - N_{i-1}}{n_i} a_i}$$

Medidas de posición: Cuantil variable discreta

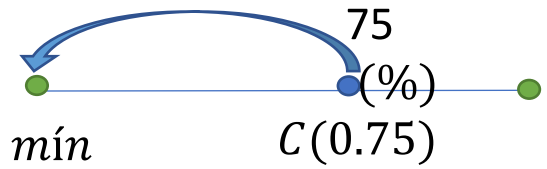
- **Cálculo para variables discretas:** El cálculo se realiza Igual que para la mediana pero cambiando n/2 por **αn** (equivalentemente ½ por α)

- Conociendo el valor de α, el cuantil $C(\alpha)$ será aquel x_i cuyo $\boldsymbol{N_i > \alpha n}$, o equivalentemente, aquel x_i cuyo $\mathbf{F_i > \alpha}$.

- **Ejemplo**: Estamos estudiando el número de reclamaciones en una tienda. Queremos obtener el C(0.75).

x_i	n_i	f_i	N_i
0	4	0,4	4
1	3	0,3	7
2	2	0,2	9
3	1	0,1	10
Sumas	*n=10*	*1*	

$\alpha n = 0.75 \; x \; 10 = 7.5$

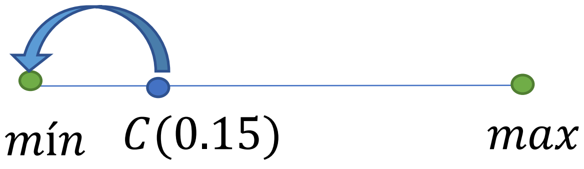


Por lo que el cuantil de orden 0.75 será 2 reclamaciones

Medidas de posición variable cuantitativa continua

- **Ejemplo:** Queremos obtener el peso máximo del 15% de los alumnos que menos pesan, esto es, el C(0.15).
- El intervalo que lo contiene en este ejemplo es (60-75] pues su N_i>0.15n=10.95

I_{i-1} - I_i	n_i	N_i	x_i	a_i
50-60	5	**5**	55	10
60-75	**25**	**30**	**67,5**	**15**
75-90	29	59	82,5	15
90-100	14	73	95	10
Sumas	*n=73*			



$$C(0.15) = I_{i-1} + \frac{\alpha n(\ \text{0.15 x 73 = 10.95}\ ) - N_{i-1}}{n_i} \cdot a_i =$$

$$60 + \frac{10.95 - 5}{25} \cdot 15 =$$

$$60 + 3.57 = 63.57\ kg$$

MODA para variable cuantitative continua

AMPLITUD DEL INTERVALO CONSTANTE
Se elige el intervalo con mayor ni

$$Mo = e_{i-1} + \frac{n_i - n_{i-1}}{(n_i - n_{i+1}) + (n_i - n_{i-1})} \times a_i$$

AMPLITUD DEL INTERVALO VARIA o no es homogénea
Se elige el intervalo con la mayor densidad de frecuencia **hi = ni/ai**

$$M_o = e_{i-1} + \frac{\frac{n_i}{a_i} - \frac{n_{i-1}}{a_{i-1}}}{\left(\frac{n_i}{a_i} + \frac{n_{i+1}}{a_{i+1}}\right) - \left(\frac{n_i}{a_i} - \frac{n_{i-1}}{a_{i-1}}\right)} \times a_i$$

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

Medidas de Dispersión

Medidas de Dispersión Absolutas. Dependen de la unidad de medida de los datos

Rango o Recorrido: $R = x_n - x_1$

Recorrido Intercuartílico: $R_I = Q_3 - Q_1$

Varianza: $$\sigma^2 = \frac{1}{n}\sum_{i=1}^{k} n_i \left(x_i - \bar{x}\right)^2 = \frac{1}{n}\sum_{i=1}^{k} n_i x_i^2 - \bar{x}^2$$

Desviación típica: $\sigma = +\sqrt{\sigma^2}$

Medidas de Dispersión Relativas. No dependen de la unidad de medida de los datos. Permiten la comparación entre distintas poblaciones

Coeficiente de variación de Pearson: $\mathrm{CV} = \frac{\sigma}{\bar{x}}$

Ejemplo

Ejemplo: Un estudio sobre los ingresos de 50 individuos que trabajan en una determinada empresa muestra los siguientes resultados. Se pide calcular la varianza y desviación típica:

Sueldos	x_i	n_i	$n_i x_i$	$n_i x_i^2$
(600 -700]	650	10	6500	4225000
(700- 800]	750	15	11250	8437500
(800- 900]	850	15	12750	10837500
(900- 1000]	950	5	4750	4512500
(1000-1200]	1100	5	5500	6050000
		50	40750	34062500

$$\bar{x} = \frac{\sum_{i=1}^{k} n_i x_i}{n} = \frac{40750}{50} = 815€$$

$$\sigma^2 = \frac{\sum_{i=1}^{k} n_i {x_i}^2}{n} - \overline{x}^2$$

$$= \frac{34062500}{50} - (815)^2 = 17025$$

$$\sigma = +\sqrt{17025} \cong 130.48$$

CV= $\frac{\sigma}{\bar{x}} = \frac{130,48}{815} = 0,1601$