Estatística Descritiva

conjunto de técnicas que permitem recolher, organizar, reduzir e apresentar dados estatísticos.

variáveis qualitativas: dados com características não numéricas, identificam uma qualidade ou característica.

escala nominal: dados divididos por categorias sem ordem.
(sexo, distrito residência, cor dos olhos, nacionalidade)

escala ordinal: dados divididos por categorias com sequência.
(opinião, estado civil, grau de escolaridade)

variáveis quantitativas: dados com características numéricas.

discretas: tomam um nº finito ou infinito numerável de valores
(nº filhos, nº pessoas que ..., nota final)

contínuas: tomam um nº infinito não numerável de valores
(idade, altura, peso, temp. do ar, salário)

dados:

variáveis em bruto ⇒ população X

Amostra em bruto ⇒ $(x_1, x_2, x_3, \dots, x_n)$

Amostra ordenada ⇒ $(x_{(1)}, x_{(2)}, x_{(3)}, \dots, x_{(n)})$

dados agrupados - tabela de frequências

x'_i	n_i	f_i	N_i	F_i
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'_k	n_k	f_k	$N_k = n$	$F_k = 1$
	n	1		

n = dimensão da amostra

k = nº categorias

x'_i = categoria

n_i = frequência absoluta simples

$N_i = \sum_{j=1}^{i} n_j$ freq. ~~absolutas~~ ~~simples~~ acumulada

f_i = frequência relativa simples $\left(\frac{n_i}{n}\right)$

$F_i = \sum_{j=1}^{i} f_j$ = frequência relativa

Variáveis Quantitativas Contínuas

construção de classes:

1. nº de classes a construir (Regra de Sturges)

$$k = \left[\frac{\ln n}{\ln 2}\right] + 1 \quad (n = \text{dimensão da amostra})$$

2. amplitude do conjunto de dados (amplitude total)

$$\Delta = X_{(n)} - X_{(1)} = \text{maximo} - \text{minimo}$$

3. amplitude das classes:

$$A = \frac{\Delta}{k}$$

4. classes:

$C_1 = [\text{minimo} ; \text{minimo} + A[$

$C_2 = [\text{minimo} + A ; \text{minimo} + 2A[$

$C_3 = [\text{minimo} + 2A ; \text{minimo} + 3A[$

$\vdots$

$C_k = [\text{minimo} + (k-1)A ; \text{minimo} + kA]$

C_i	x'_i	n_i	f_i	N_i	F_i
C_1	x'_1	n_1	f_1	N_1	F_1
C_2	x'_2	n_2	f_2	N_2	F_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
C_k	x'_k	n_k	f_k	$N_k = n$	$F_k = 1$
		n	1		

$C_i = [l_i ; L_i[$ classes ou intervalos de classes

$x'_i = \frac{(l_i + L_i)}{2}$ pontos médios das classes

Medidas de Assimetria (Skewness)

simétrica: $\bar{X} = Me = Mo$

Assimétrica positiva (enviesado à esquerda): $\bar{X} > Me > Mo$

Assimétrica negativa (enviesado à direita): $\bar{X} < Me < Mo$

- coeficiente de assimetria de Fisher
- grau de assimetria de Pearson
- grau de Assimetria de Bowley

amplitude inter-quartil
$Q = Q_3 - Q_1$

intervalo de variação
$Q' = Q_{0,90} - Q_{0,1} \Leftrightarrow P_{90} - P_{10}$

NO SPSS

$$SK = \frac{\text{skewness}}{\text{std. error Skewness}}$$

simétrica: $|SK| \leq 1,96$

A. positiva: $SK > 1,96$

A. negativa: $SK < -1,96$

Outlier: valor cuja magnitude se afasta de maneiras evidente do centro da distribuição

Moderado: X_i ultrapassa as barreiras moderadas

$X_i < Q_1 - 1.5Q$ ou $X_i > Q_3 + 1.5Q$
$(Q_3 - Q_1)$ $(Q_3 - Q_1)$

Severo: X_i ultrapassa as barreiras severas

$X_i < Q_1 - 3Q$ ou $X_i > Q_3 + 3Q$
$(Q_3 - Q_1)$ $(Q_3 - Q_1)$

Medidas de Achatamento (Kurtosis)

- coeficiente de Achatamento:

$$B_2 = \frac{M_4}{S^4} \begin{cases} > 3 & \text{d. leptocurtica} \\ = 3 & \text{d. mesocurtica} \\ < 3 & \text{d. platicurtica} \end{cases}$$

- coeficiente percentil de achatamento:

$$K = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})} \begin{cases} < 0,263 & \text{D. Lepto} \\ = 0,263 & \text{d. meso} \\ > 0,263 & \text{d. plati} \end{cases}$$

NO SPSS

$$Kt = \frac{\text{Kurtosis}}{\text{std. Error Kurtosis}}$$

distribuição mesocúrtica: $|Kt| \leq 1,96$

distribuição leptocurtica: $Kt > 1,96$

distribuição platicúrtica: $Kt < -1,96$

coeficiente de correlação amostral de spearman:
mede o grau de associação entre duas amostras de dados quantitativos ou qualitativos ordinais.

$$RS = 1 - \frac{6\sum_{i=1}^{n}(d_i^2)}{n(n^2-1)}, \quad -1 \leq R_S \leq 1$$

$d_i \Rightarrow$ diferença entre os valores de ordem de X_i e Y_i

coeficiente de dispersão: $CD = \frac{S}{\bar{X}}$

coeficiente de variação: $CV = CD \times 100\%$

$CV < 50\%$ ⇒ a média será tanto mais representativa quanto menor o valor do coeficiente.

$CV \geq 50\%$ ...