

Análise descritiva de duas variáveis quantitativas

- Estudam o comportamento conjunto de duas variáveis
- Comportamento:
 - observado: diagrama de dispersão
 - medido: coeficiente de correlação de Pearson

Coeficiente de correlação de Pearson

- Mede a relação entre duas variáveis quantitativas contínuas
- Quantifica o grau em que duas variáveis aleatórias estão relacionadas, desde que a relação seja linear
- No gráfico de dispersão bidimensional, no eixo a variável X e na ordenada a variável Y, o ponto no gráfico representa uma combinação de valores (x,y)

Coeficiente de correlação de Pearson

- A relação entre as variáveis (x,y) é representada por rho (ρ) ou "r"
- Rho é a média dos produtos dos desvios padrão de X e Y
- O coeficiente de correlação é sensível a valores extremos e o resultado pode ser enganoso
- Se X e Y não tem uma relação linear, não fornece uma medida válida de associação

Coeficiente de correlação de Pearson

- O coeficiente de correlação é adimensional, não tem unidade de medida e varia de -1 a +1.
 - Onde: +1 = correlação perfeita positiva
 - 1 = correlação perfeita negativa
 - 0 = não há correlação
- Uma alta correlação entre duas variáveis não significa relação de causa e efeito

$$\rho = r = \frac{\text{Co } x,y}{(sx)(sy)}$$

Coeficiente de correlação de Pearson

$$\text{Co } x,y = \frac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y}) f_i}{n}$$

$(x_i - \bar{x}) \cdot (y_i - \bar{y})$ mostra o quanto a população desviou-se em conjunto da média

- Varia de $-\infty$ $\xrightarrow{0}$ $+\infty$
- O sinal da covariância coincide com o sentido da correlação
- Cada variável tem sua variabilidade e seu desvio padrão
- Lembrar do sobrenome quando houver!

Coeficiente de correlação de Pearson

- Não existe eixo para as frequências das variáveis.
- Uma correlação positiva entre duas variáveis mostram apenas que essas variáveis crescem no mesmo sentido
- as variáveis X e Y são medidas simultaneamente para uma mesma unidade de informação
- Vantagem: previsão

Teste de hipótese para a correlação de Pearson

H_0 : não existe correlação entre as variáveis X e Y na população original

$$t = r \sqrt{\frac{n-2}{1-r^2}}$$

- Pares de observação (xi,yi) foram obtidos aleatoriamente
- X e Y tem distribuição normal
- Tem uma distribuição t com n-2 graus de liberdade somente quando a hipótese nula for verdadeira

Coeficiente de correlação de Spearman

- Método não paramétrico
- Ordenam-se os dois conjuntos de resultados x e y separadamente e calcula-se o coeficiente de correlação de postos
- Símbolo é r_s
- Varia de -1 a +1
- É menos sensível a valores atípicos que o coef. correl Pearson
- Pode ser usado quando uma ou ambas as variáveis são ordinais
- Por se basear em postos e não em observações reais, o método não paramétrico não usa tudo o que é conhecido sobre a distribuição

Coeficiente de correlação de Spearman

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}$$

Teste de hipótese

- Amostra é maior de dez
- os pares de postos são escolhidos aleatoriamente

$$t = r \sqrt{\frac{n - 2}{1 - r^2}}$$
