

PALESTRA: ROTULAÇÃO DE OUTLIERS

DIA: 27.10.22, das 9h às 12h

INTRODUÇÃO: história vide Barnett e Lewis (1994), motivação, importância, objetivo.

Em Portugal: Rosado (2006).

A) Tukey (1977)

“*outlier labeling*”, traduzido como **rotulação** de *outlier* (SILVA, 2019)

O.I.:= outlier inferior

O.S.:= outlier superior

Q1:= primeiro quartil

Q2:= segundo quartil ou mediana (Me).

Q3:= terceiro quartil

$$\text{O.I.} < Q1 - 1,5*(Q3 - Q1)$$

$$\text{O.S.} > Q3 + 1,5*(Q3 - Q1)$$

B) Adil e Zaman (2020)

P12,5 := décimo segundo e meio percentil.

P37,5 := trigésimo sétimo e meio percentil.

P62,5 := sexagésimo segundo e meio percentil.

P87,5 := octagésimo sétimo e meio percentil.

$$\text{O.I.} < P12,5 - 1,5*(P37,5 - P12,5)$$

$$\text{O.S.} > P87,5 + 1,5*(P87,5 - P62,5)$$

C) Introdução de **assimetria** no trabalho de Hubert e Vandervieren (2008).

D) **Proposição:** coeficiente quartílico de assimetria!

E) Silva et al. (1996, p. 125) e Silva (2011, p. 208): ASSIMETRIA

As:= medida de assimetria (adimensional)

$$As = (Q3 + Q1 - 2Q2) / (Q3 - Q1).$$

Se:

$|As| = 0$; então distribuição simétrica.

$0 < |As| \leq 0,1$, então assimetria fraca.

$0,1 < |As| < 0,3$, então assimetria moderada.

$0,3 \leq |As| \leq 1$, então assimetria forte.

Silva et al. (1996, p. 125) e Silva (2011, p. 208) também informaram que o coeficiente de Bowley (“As”) varia entre -1 a $+1$.

F) Walker et al. (2018)

$$O.I. < Q1 - 1,5*(Q3 - Q1)*[(1 - As)/(1 + As)]$$

$$O.S. > Q3 + 1,5*(Q3 - Q1)*[(1 + As)/(1 - As)]$$

G) Lima et al. (2018)

e := número de Euler; $e \approx 2,718...$ (função exponencial na calculadora)

$$O.I. < Q1 - 1,5*(Q3 - Q1)*e^{-(As)}$$

$$O.S. > Q3 + 1,5*(Q3 - Q1)*e^{+(As)}$$

CONSIDERAÇÕES FINAIS: Síntese dos resultados. Perspectivas.

REFERÊNCIAS

ADIL, Iftikhar Hussain; ZAMAN, Asad. Outliers detection in skewed distributions: split sample skewness based boxplot. **Economic Computation and Economic Cybernetics Studies and Research**, Bucharest, v. 54, n. 3, p. 279-296, 2020.

BARNETT, Vic; LEWIS, Toby. **Outliers in statistical data**. 3. ed. New York: John Wiley & Sons, 1994.

HUBERT, M.; VANDERVIEREN, E. An adjusted boxplot for skewed distributions. **Computational Statistics & Data Analysis**, Amsterdam, v. 52, n. 12, p. 5186-5201, Aug. 2008.

LIMA, Luís Fernando Maia; MAROLDI, Alexandre Masson; SILVA, Dávilla Vieira Odízio da; HAYASHI, Carlos Roberto Massao; HAYASHI, Maria Cristina Piumbato Innocentini. A influência de *outliers* nos estudos métricos da informação: uma análise de dados univariados. **Em Questão**, Porto Alegre, v. 24, Edição Especial 6 EBBC, p. 216-235, 2018.

ROSADO, Fernando. **Outliers em dados estatísticos**. Lisboa: Sociedade Portuguesa de Estatística, 2006.

SILVA, Ermes Medeiros da; SILVA, Elio Medeiros da; GONÇALVES, Valter; MUROLO, Afrânio Carlos. **Estatística para os cursos de: economia, administração e ciências contábeis**. 2 ed. São Paulo: ATLAS, 1996.

SILVA, Domingos J. Lopes da. **Estatística aplicada à investigação científica nas ciências do desporto** – análise exploratória de dados com recursos ao SPSS. Fafe: Instituto de Estudos Superiores de Fafe. 2011.

SILVA, Kelly Cristina Ramos da. Regras robustas para rotular outliers em dados de caudas leves e caudas pesadas. **Tese de Doutorado**. 2019. Disponível em: <<https://teses.usp.br/teses/disponiveis/55/55134/tde-29042019-145141/pt-br.php>>. Acesso em 22 maio 2021.

TUKEY, John Wilder. **Exploratory data analysis**. Reading: Massachusetts, Addison-Wesley, 1977.

WALKER, M. L.; DOVOEDO, Y. H.; CHAKRABORTI, S.; HILTON, C. W. An improved boxplot for univariate data. **The American Statistician**, v. 72, n. 4, p. 348-353, November 2018.