

Técnicas de Modelagem para Estudos Ambientais

1º módulo - Estatística básica

1º módulo - Estatística básica

Aula 01 - Tipos de variáveis, apresentação de dados em tabelas, frequência

Aula 02 - Apresentação de dados em gráficos, medidas de tendência central

Aula 03 - A questão da variabilidade, boxplot

Aula 04 - Relação entre duas variáveis

Aula 05 - Probabilidade, população e amostra

Referência: Estatística básica; Vieira, Sonia. CENGAGE Learning, 2012.

Aula 03 - A questão da variabilidade

Para analisar os dados de uma variável, só as medidas de tendência central não são suficientes

Precisamos também de medidas de variabilidade ou de dispersão em torno desse centro

Algumas medidas de variabilidade:

- Máximo, mínimo, amplitude
- Quartis
- Percentis
- Variância e desvio padrão

Aula 03 - A questão da variabilidade

Máximo, mínimo e amplitude

- Dão uma ideia do intervalo de possíveis valores que a variável pode assumir

- Mínimo é o menor valor da variável
- Máximo é o maior valor da variável
- Amplitude é a diferença entre o máximo e o mínimo

A amplitude não é a melhor medida de variabilidade porque usa apenas 2 valores extremos para medí-la

- 2 conjuntos muito diferentes podem ter a mesma amplitude
- É muito sensível aos valores extremos: um *outlier* (valor discrepante) pode alterar muito a amplitude e ela não dará uma boa ideia da variabilidade dos dados
- Muitas vezes é mais útil fornecer o máximo e o mínimo

Aula 03 - A questão da variabilidade

Quartis

- dividem o conjunto de dados em 4 partes iguais
- são 3 quartis:
 - 1º quartil: separa a série entre $\frac{1}{4}$ dos dados com menores valores e o restante
 - 2º quartil: divide a série em 2 partes iguais - **mediana**
 - 3º quartil: separa a série entre $\frac{3}{4}$ dos dados com menores valores e $\frac{1}{4}$ com maiores valores
- para obter os quartis:
 - Organize os dados em ordem crescente
 - Encontre a mediana (2º quartil) e separe a série em 2
 - A mediana da metade com os menores valores da série é o 1º quartil
 - A mediana da metade com os maiores valores da série é o 3º quartil

Aula 03 - A questão da variabilidade

Quartil

- distância interquartílica (ou intervalo interquartil) - IIQ: diferença entre o terceiro e o primeiro quartil

$$\text{IIQ} = 3^{\circ}\text{quartil} - 1^{\circ}\text{quartil}$$

- é uma medida de variabilidade dos dados, pois metade da série fica entre esses intervalos

“A amplitude é muito sensível aos valores discrepantes. Recomenda-se, então, calcular a mediana, a amplitude e a distância interquartílica – principalmente nos casos em que existam dados discrepantes.”

Aula 03 - A questão da variabilidade

Diagrama de caixa - boxplot (candlestick)

- mostra os 3 quartis, o máximo e o mínimo da série, desconsiderando-se os valores discrepantes e pode mostrar também os valores discrepantes
- **Como classificar um valor como discrepante?**
 - Usamos o IIQ: são discrepantes os valores maiores que uma vez e meia o intervalo interquartil para cima do 3º quartil e valores menores que uma vez e meia o intervalo interquartil para baixo do 1º quartil
 - Dessa forma, os limites do boxplot não mostram o valor máximo e o valor mínimo, mas os limites superior e inferior de acordo com a fórmula
 - $L_{inferior} = 1^{\circ}\text{quartil} - 1,5 * IIQ$
 - $L_{superior} = 3^{\circ}\text{quartil} + 1,5 * IIQ$

Aula 03 - A questão da variabilidade

Percentil

É o valor abaixo do qual cai certa porcentagem dos dados

- Muito usado para definição de valores extremos
- Nas planilhas eletrônicas sua-se a função:
=Percentil(dados, percentil)

Aula 03 - A questão da variabilidade

Desvio padrão e variância

- Desvio padrão é uma medida muito utilizada de variabilidade dos dados porque mede de maneira eficaz a dispersão dos dados em torno da média.
- O desvio padrão é calculado a partir da variância
- O desvio é definido como a diferença entre determinado valor e a média

desvio = observação - média

$$desvio = X - \bar{X}$$

Aula 03 - A questão da variabilidade

Desvio padrão e variância

- como há desvios positivos e negativos, não podemos simplesmente usar a média dos desvios como medida da variabilidade dos dados. Portanto, calculamos o quadrado dos desvios

- a variância será então a média dos quadrados dos desvios

- para a população:
$$s^2 = \frac{\sum_{i=1}^n (X - \bar{X})^2}{n}$$

- para a amostra:

$$s^2 = \frac{\sum_{i=1}^n (X - \bar{X})^2}{n-1}$$

Aula 03 - A questão da variabilidade

Desvio padrão e variância

- como o cálculo da variância envolve quadrados, seu valor não representará a mesma unidade de medida dos dados
- o desvio padrão é a raiz quadrada da variância e, portanto, mantém a mesma unidade de medida dos dados
- para a população:

$$s = \sqrt{\frac{\sum_{i=1}^n (X - \bar{X})^2}{n}}$$

- para a amostra:

$$s = \sqrt{\frac{\sum_{i=1}^n (X - \bar{X})^2}{n-1}}$$

Aula 03 - A questão da variabilidade

Exercícios