

Análise exploratória II

Resumos numéricos - medidas de posição, dispersão e forma.

Prof. Me. Lineu Alberto Cavazani de Freitas

CE003 – Estatística II

Departamento de Estatística
Laboratório de Estatística e Geoinformação



Resumos numéricos

- ▶ Uma forma de resumir a informação contida em um conjunto de dados é por meio dos resumos numéricos.
- ▶ Resumos numéricos são basicamente números que resumem números.
- ▶ Os dois principais grupos são as medidas de posição (central e relativa) e dispersão.
- ▶ Existem outros conjuntos de medidas, como as medidas de forma e também as de relação.

Medidas de posição central

- ▶ As medidas de posição central buscam expressar o centro de uma variável por meio de ideias como:
 - ▶ Centro de massa.
 - ▶ Valor que divide a amostra em partes iguais.
 - ▶ Valores de maior frequência ou densidade.
- ▶ Algumas possibilidades são
 - ▶ Média.
 - ▶ Mediana.
 - ▶ Moda.
 - ▶ Média geométrica.
 - ▶ Média harmônica.
 - ▶ Média aparada.

Medidas de posição central

- ▶ As medidas de posição relativa buscam representar pontos do domínio em que a variável apresenta porções com frequências conhecidas.
- ▶ Algumas possibilidades são
 - ▶ Quartis.
 - ▶ Decis.
 - ▶ Percentis.
 - ▶ Máximo.
 - ▶ Mínimo.

Medidas de dispersão

- ▶ As medidas de dispersão são utilizadas para expressar informações como o domínio da variável, grau de dispersão ao redor do centro (variabilidade), e também distanciamento dos valores com relação ao centro.
- ▶ Algumas medidas possíveis são
 - ▶ Amplitude.
 - ▶ Variância.
 - ▶ Desvio padrão.
 - ▶ Desvio absoluto (médio ou mediano).
 - ▶ Coeficiente de variação.

Medidas de forma e associação

- ▶ As medidas de forma expressam informações como assimetria e curtose da distribuição de valores por meio de coeficientes.
- ▶ As medidas de relação visam verificar se existe algum tipo de relação entre pares de variáveis.



Medidas de posição central

Média aritmética

- ▶ Soma de todos os valores dividida pela quantidade de elementos.
- ▶ Interpretação física de centro de gravidade.
- ▶ Medida influenciada por valores extremos.

EXPRESSÃO

Média aritmética

EXEMPLO



Média aritmética ponderada

- ▶ Indicada para dados agrupados ou situações em que existe motivo para unidades receberem um peso maior.
- ▶ Obtêm-se os produtos entre frequências relativas ou os pesos e os valores que a variável assume.
- ▶ Somam-se os produtos e divide-se pela soma das frequências (quantidade de elementos).
- ▶ No caso de faixas de valores, usa-se o centro da faixa.

EXPRESSÃO

Média aritmética ponderada

EXEMPLO



Outros tipos de média

- ▶ Média aritmética e ponderada são os tipos de média mais comuns.
- ▶ Contudo existem outras possibilidades como
 - ▶ Média geométrica.
 - ▶ Média harmônica.
 - ▶ Média aparada.

Mediana

- ▶ Valor que ocupa a posição intermediária dos valores ordenados.
- ▶ Divide o vetor de valores em 2 partes de mesmo tamanho.
- ▶ Metade dos valores é menor que a mediana e a outra metade maior que a mediana.
- ▶ Basta ordenar o conjunto de valores e verificar qual é o valor central.
- ▶ Se o número de observações for ímpar, a mediana é o valor central.
- ▶ Se o número de observações for par, a mediana é a média dos dois valores centrais.

EXPRESSÃO

Mediana

EXEMPLO



Moda

- ▶ Valor ou classe que apresenta maior frequência ou densidade.
- ▶ Valor mais típico, aquele que mais se repete.
- ▶ Quando todos os valores são distintos, não existe moda.
- ▶ Quando a maior frequência está associada a mais de um valor, existe mais de uma moda.

EXPRESSÃO

EXEMPLO



Média, mediana e moda

- ▶ Na prática, estas medidas possuem vantagens e desvantagens.
- ▶ Caso haja valores discrepantes a média é uma medida altamente influenciada, o que não acontece com a moda e a mediana.
- ▶ Já a mediana é difícil de ser obtida quando existem muitos dados, dado que o processo de ordenação é custoso.
- ▶ A dificuldade com a moda surge quando trabalha-se com distribuições multimodais, isto é diversos valores tem a mesma frequência de ocorrência.

Média, mediana e moda

- ▶ As três medidas perdem significado em distribuições multimodais.
- ▶ A média tende a ser uma boa alternativa quando a distribuição é unimodal, simétrica e sem valores extremos.
- ▶ A mediana tende a ser uma boa alternativa para distribuições assimétricas ou com presença de valores extremos.
- ▶ A moda tende a ser uma boa alternativa quando valores se repetem, estão agrupados em classe ou trata-se de uma variável qualitativa.
- ▶ Média, moda e mediana aproximam-se em distribuições unimodais simétricas.



Medidas de dispersão

Medidas de dispersão

- ▶ Medidas de posição por si só não são de grande valia para tirarmos conclusões preliminares a respeito de um conjunto de dados.
- ▶ Há a necessidade de aliar as medidas de posição a quantidades que mensurem a variabilidade.
- ▶ Em geral usamos uma medida de posição central, que nos dá uma ideia de centro dos dados.
- ▶ Aliamos esta informação com uma medida de dispersão, que nos dá uma noção de quão dispersos são os dados.

Medidas de dispersão

- ▶ Lembre-se que conjuntos de dados com diferentes valores podem gerar as mesmas medidas de posição.
- ▶ Contudo um pode ser mais disperso que o outro.
- ▶ Outra utilidade das medidas de dispersão é expressar o domínio da variável.
- ▶ Já as medidas de forma expressam informações como assimetria e curtose da distribuição de valores.

Amplitude total - Diferença entre o maior e o menor valor da variável. - Sensível a valores extremos. - Usa apenas duas medidas.

Desvio

- ▶ Um desvio é uma medida de distância da observação para uma medida de posição central.
- ▶ Podemos usar como referência a média ou a mediana.
- ▶ Tomamos todos os desvios absolutos.
- ▶ Calculamos a média.

EXPRESSÃO

Desvio

EXEMPLO



Variância

- ▶ Uma alternativa melhor do que os desvios é usar a soma dos quadrados dos desvios.
- ▶ Esta ideia dá origem à variância de um conjunto de dados.
- ▶ O problema da variância é a unidade de medida (ao quadrado).

EXPRESSÃO

Variância

EXEMPLO



Desvio-padrão

- Para ter uma medida de dispersão com a mesma unidade de medida dos dados originais definiu-se o desvio-padrão como a raiz quadrada da variância.

EXPRESSÃO

Desvio-padrão

EXEMPLO



Coeficiente de variação

- ▶ Medida de variabilidade relativa à média.
- ▶ Quociente do desvio-padrão pela média.
- ▶ Medida adimensional, geralmente apresentada na forma de porcentagem.
- ▶ Permite comparar a variabilidade de variáveis de diferentes naturezas

EXPRESSÃO

Coeficiente de variação

EXEMPLO



Desvio, variância, desvio padrão, coeficiente de variação]

DISCUSSÃO MEDIDAS

Para variáveis qualitativas existem medidas que avaliam a dispersão são funções das frequências das classes.