

Problemas Inversos em Python

Sergio Pedro Rodrigues Oliveira

26 December 2024

SUMÁRIO

| | | |
|----------|---|-----------|
| 1 | PRINCIPAIS TÓPICOS | 1 |
| 2 | DADOS | 1 |
| 2.1 | O que são dados? | 1 |
| 2.2 | Informação | 1 |
| 2.3 | Teoria | 1 |
| 2.4 | Tipos de variáveis | 2 |
| 3 | Estatística Básica (Teoria medidas de posição e dispersão) | 3 |
| 3.1 | Preparação dos dados para aplicação de estatística básica | 3 |
| 3.1.1 | Teoria | 3 |
| 3.1.2 | Preparação dos dados (sumariar dados coletados) | 6 |
| 3.1.2.1 | Variável Quantitativa Discreta | 7 |
| 3.1.2.2 | Variável Quantitativa Contínua | 8 |
| 3.1.2.3 | Variáveis Qualitativas | 8 |
| 3.2 | Medidas de posição | 8 |
| 3.3 | Medidas de dispersão | 8 |
| 3.4 | Análise Estatística | 9 |
| 4 | EXCEL | 10 |

LISTA DE FIGURAS

| | | |
|---|---|---|
| 1 | Estatística descritiva. | 3 |
| 2 | Tipos de variáveis. | 5 |
| 3 | Distribuição tabular quantitativo discreta. | 7 |

LISTA DE TABELAS

1 PRINCIPAIS TÓPICOS

2 DADOS

2.1 O que são dados?

Dados são valores brutos atribuídos a algo.

2.2 Informação

- Informação é a ordenação e organização dos dados de forma a transmitir significado e compreensão dentro de um contexto.
- Informação é o significado que a gente obtém a partir dos dados.
- Informação = fazer perguntas para os dados (responder pergunta).
- Nem sempre podemos confiar nos dados, é preciso entender o contexto dos dados.
 - De onde eles vem?
 - Quem são as pessoas que responderam?
 - O que são esses dados?

2.3 Teoria

- Análise descritiva dos dados através da tabulação das variáveis e cálculo de medidas descritivas (média, desvio-padrão, etc).
- Análise descritiva dos dados (Informações preliminares):
 - Contagem dos resultados observados em cada variável do conjunto de dados.
 - Natureza descritiva dos dados, tipo de variáveis (categórica ou numérica).
 - Três objetivos principais:
 - * Verificar erros e anomalias.
 - * Compreender a distribuição de cada uma das variáveis isoladamente.
 - * Compreender a natureza e a força das relações entre as variáveis.
- Após essas etapas, estabelecer um modelo estatístico formal e relatar suas conclusões.

2.4 Tipos de variáveis

- Variável numérica:
 - Continua

Se seus valores pertencer ao conjunto dos números reais.

Ex.: Temperatura corporal, saldo em caixa, peso da carga de um caminhão, etc.
 - Discreta

Se seus valores pertencer ao conjunto dos números inteiros.

Ex.: Número de pessoas com febre, número de empresas, número de caminhões, etc.
- Variável categórica:
 - Ordinal

Se seus valores podem ser ordenados do menor para o maior.

Ex.: Temperatura (baixa, média ou alta), saldo em caixa (negativo, nulo ou positivo), etc.
 - Nominal

Quando não for possível estabelecer ordenamento.

Ex.: Sexo do indivíduo, atividade fim da empresa, marca/modelo do caminhão, etc.

3 Estatística Básica (Teoria medidas de posição e dispersão)

3.1 Preparação dos dados para aplicação de estatística básica

3.1.1 Teoria

- Definição de Estatística:

A Estatística de uma maneira geral compreende aos métodos científicos para **COLETA**, **ORGANIZAÇÃO**, **RESUMO**, **APRESENTAÇÃO** e **ANÁLISE** de Dados de Observação (Estudos ou Experimentos), obtidos em qualquer área de conhecimento. A finalidade é a de obter conclusões válidas para tomada de decisões.

- Estatística Descritiva

Parte responsável basicamente pela **COLETA** e **SÍNTESE** (Descrição) dos Dados em questão.

Disponibiliza de técnicas para o alcance desses objetivos. Tais Dados podem ser provenientes de uma **AMOSTRA** ou **POPULAÇÃO**.

- Estatística Inferencial

É utilizada para tomada de decisões a respeito de uma população, em geral fazendo uso de dados de amostrais.

Essas decisões são tomadas sob condições de **INCERTEZA**, por isso faz-se necessário o uso da **TEORIA DA PROBABILIDADE**.

- O fluxograma da estatística descritiva pode ser esposto da seguinte forma:



Figure 1: Estatística descritiva.

- A representação tabular (Tabelas de Distribuição de Frequências) deve conter:

- Cabeçalho

Deve conter o suficiente para que as seguintes perguntas sejam respondidas “**o que?**” (Relativo ao fato), “**onde?**” (Relativo ao lugar) e “**quando?**” (Correspondente à época).

- Corpo

É o lugar da Tabela onde os dados serão registrados. Apresenta colunas e sub colunas.

- Rodapé

Local destinado à outras informações pertinentes, por exemplo a Fonte dos Dados.

- População e Amostras

- População

É o conjunto de todos os itens, objetos ou pessoas sob consideração, os quais possuem pelo menos uma característica (Variável) em comum. Os elementos pertencentes à uma População são denominados “Unidades Amostrais”.

- Amostras

É qualquer subconjunto (não vazio) da População. É extraída conforme regras pré-estabelecidas, com a finalidade de obter “estimativa” de alguma Característica da População.

- Tipos de variáveis



Figure 2: Tipos de variáveis.

- *Qualitativo nominal:*
Não possuem uma ordem natural de ocorrência.
- *Qualitativo ordinal:*
Possuem uma ordem natural de ocorrência.
- *Quantitativo discreta:*
Só podem assumir valores inteiros, pertencentes a um conjunto finito ou enumerável.
- *Quantitativo contínua:*
Podem assumir qualquer valor em um determinado intervalo da reta dos números reais.

3.1.2 Preparação dos dados (sumariar dados coletados)

- Frequência (conceito):
É a quantidade de vezes que um valor é observado dentro de um conjunto de dado.
- Distribuição em frequências:
 - A distribuição tabular é denominada: “Tabela de Distribuição de Frequências”.
 - Podemos separar em 3 modelos de distribuição tabular:
 - * Variável Quantitativa Discreta.
 - * Variável Quantitativa Contínua.
 - * Variáveis Qualitativas.

3.1.2.1 Variável Quantitativa Discreta

- Passos da preparação dos dados:
 - 1º Passo - **DADOS BRUTOS**:
Obter os dados da maneira que foram coletados.
 - 2º Passo - **ROL**:
Organizar os DADOS BRUTOS em uma determinada ordem (crescente ou decrescente).
 - 3º Passo - **CONSTRUÇÃO TABELA**:
Na primeira coluna são colocados os valores da variável, e nas demais as respectivas frequências.
 - Frequência absoluta simples (Nº de vezes que cada valor da variável se repete).
- Principais campos da **distribuição tabular de variáveis quantitativas discretas**:
 - n é o número total de elementos da amostra.
 - x_i é o número de valores distintos que a variável assume.
 - F_i é a Frequência Absoluta Simples.
 - f_i é a Frequência Relativa Simples.
 - $f_i\%$ é a Frequência Relativa Simples Percentual. $f_i\% = f_i \cdot 100\%$.
 - F_a é a Frequência Absoluta Acumulada.

| <u>xi</u> | <u>Fi</u> | <u>fi</u> | <u>fi%</u> | <u>Fa↓</u> | <u>Fa↑</u> | <u>fa↓</u> | <u>fa↑</u> |
|-----------|-----------|-----------|------------|------------|------------|------------|------------|
| 0 | 6 | 0,2 | 20 | 6 | 30 | 0,2 | 1 |
| 1 | 11 | 0,37 | 37 | 17 | 24 | 0,57 | 0,8 |
| 2 | 8 | 0,27 | 27 | 25 | 13 | 0,84 | 0,43 |
| 3 | 2 | 0,07 | 7 | 27 | 5 | 0,91 | 0,16 |
| 4 | 2 | 0,06 | 6 | 29 | 3 | 0,97 | 0,09 |
| 6 | 1 | 0,03 | 3 | 30 | 1 | 1 | 0,03 |
| Total | 30 | 1 | 100 | - | - | - | - |

Figure 3: Distribuição tabular quantitativo discreta.

Obs.: As setas simbolizam ordem crescente ou decrescente.

3.1.2.2 Variável Quantitativa Contínua

3.1.2.3 Variáveis Qualitativas

3.2 Medidas de posição

3.3 Medidas de dispersão

3.4 Análise Estatística

- Para fazer uma Análise Estatística eficiente de dados, precisamos:

- Limpar os dados:

Remover os *OUTLIER* (valores atípicos, inconsistentes).

- Aplicar Estatística Descritiva aos dados:

As medidas de posição (**Média**, **Mediana** e **moda**) e dispersão (**Amplitude Total**, **Desvio**, **Desvio Médio**, **Variância**, **Desvio-padrão** e **Coefficiente de Variação**) são maneiras de descrever os dados.

- Comparar as medidas dos dados:

Principalmente medidas de dispersão, me especial **Coefficiente de Variação**, são ótimas para comparar dados.

- Previsão de dados:

A principal técnica é de **Regressão**, porém para aplicar, necessita que os dados estejam limpos e com pouca dispersão (quanto menor, melhor).

4 EXCEL