



Estatística I

Prof. Fernando de Souza Bastos
fernando.bastos@ufv.br

Departamento de Estatística
Universidade Federal de Viçosa
Campus UFV - Viçosa

Introdução

Tipo de Variáveis

Distribuições de Frequências

Introdução

Estatística

É uma ciência que possui uma coleção de métodos para planejar experimentos, obter, organizar, resumir, analisar, interpretar dados e deles extrair conhecimento.

Introdução

Estatística Descritiva

É a parte da Estatística que procura somente descrever e avaliar um certo grupo, seja ele a população ou a amostra. É importante ressaltar que todo trabalho científico pode ser resumido em algumas etapas:

Estatística Descritiva

É a parte da Estatística que procura somente descrever e avaliar um certo grupo, seja ele a população ou a amostra. É importante ressaltar que todo trabalho científico pode ser resumido em algumas etapas:

1. Definição do problema;

Estatística Descritiva

É a parte da Estatística que procura somente descrever e avaliar um certo grupo, seja ele a população ou a amostra. É importante ressaltar que todo trabalho científico pode ser resumido em algumas etapas:

1. Definição do problema;
2. Planejamento (Inclui a revisão de literatura);

Estatística Descritiva

É a parte da Estatística que procura somente descrever e avaliar um certo grupo, seja ele a população ou a amostra. É importante ressaltar que todo trabalho científico pode ser resumido em algumas etapas:

1. Definição do problema;
2. Planejamento (Inclui a revisão de literatura);
3. Coleta, limpeza e organização dos dados;

Estatística Descritiva

É a parte da Estatística que procura somente descrever e avaliar um certo grupo, seja ele a população ou a amostra. É importante ressaltar que todo trabalho científico pode ser resumido em algumas etapas:

1. Definição do problema;
2. Planejamento (Inclui a revisão de literatura);
3. Coleta, limpeza e organização dos dados;
4. Apresentação dos dados;

Estatística Descritiva

É a parte da Estatística que procura somente descrever e avaliar um certo grupo, seja ele a população ou a amostra. É importante ressaltar que todo trabalho científico pode ser resumido em algumas etapas:

1. Definição do problema;
2. Planejamento (Inclui a revisão de literatura);
3. Coleta, limpeza e organização dos dados;
4. Apresentação dos dados;
5. Descrição dos dados;

Estatística Descritiva

É a parte da Estatística que procura somente descrever e avaliar um certo grupo, seja ele a população ou a amostra. É importante ressaltar que todo trabalho científico pode ser resumido em algumas etapas:

1. Definição do problema;
2. Planejamento (Inclui a revisão de literatura);
3. Coleta, limpeza e organização dos dados;
4. Apresentação dos dados;
5. Descrição dos dados;
6. Análise, modelagem e estabelecimento de inferências.

Conceitos básicos

População

Conjunto de dados que têm alguma característica em comum.

Conceitos básicos

População

Conjunto de dados que têm alguma característica em comum.

Amostra

Qualquer conjunto não vazio de elementos retirado da população.

Conceitos básicos

População

Conjunto de dados que têm alguma característica em comum.

Amostra

Qualquer conjunto não vazio de elementos retirado da população.

Parâmetro

Característica numérica qualquer da população.

Conceitos básicos

População

Conjunto de dados que têm alguma característica em comum.

Amostra

Qualquer conjunto não vazio de elementos retirado da população.

Parâmetro

Característica numérica qualquer da população.

Estimador

Qualquer função dos elementos amostrais que seja usada para encontrar uma estimativa de um parâmetro.

Conceitos básicos

População

Conjunto de dados que têm alguma característica em comum.

Amostra

Qualquer conjunto não vazio de elementos retirado da população.

Parâmetro

Característica numérica qualquer da população.

Estimador

Qualquer função dos elementos amostrais que seja usada para encontrar uma estimativa de um parâmetro.

Estimativa

Qualquer valor numérico assumido por um estimador.

Tipo de Variáveis

Tipo de Variáveis

Variável Aleatória

característica da população sujeita a variação.

Tipo de Variáveis

Variável Aleatória

característica da população sujeita a variação.

Conhecer o tipo da variável é importante porque os métodos estatísticos que você pode utilizar em sua análise variam de acordo com o tipo de variável. Existem dois tipos principais para variáveis:

- **Variáveis categóricas** (ou **Variáveis qualitativas**)
apresentam valores que podem somente ser posicionados em categorias tais como sim e não.

Tipo de Variáveis

Variável Aleatória

característica da população sujeita a variação.

Conhecer o tipo da variável é importante porque os métodos estatísticos que você pode utilizar em sua análise variam de acordo com o tipo de variável. Existem dois tipos principais para variáveis:

- **Variáveis categóricas** (ou **Variáveis qualitativas**)
apresentam valores que podem somente ser posicionados em categorias tais como sim e não.
- **Variáveis numéricas** (ou **Variáveis quantitativas**)
apresentam valores que representam quantidades.

Tipo de Variáveis

Dentre as variáveis qualitativas, ainda podemos fazer uma distinção entre dois tipos: **variável qualitativa nominal**, para a qual não existe nenhuma ordenação nas possíveis realizações, e **variável qualitativa ordinal**, para a qual existe uma ordem nos seus resultados.

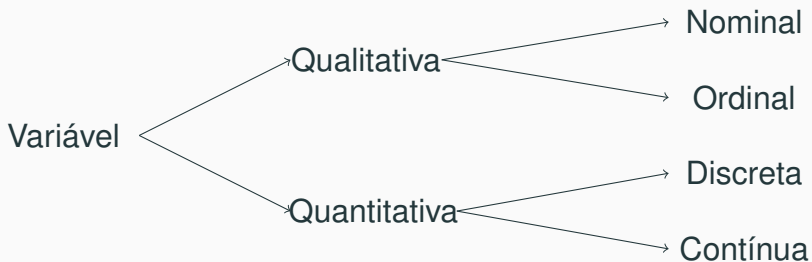
Variáveis numéricas podem ser, ainda, identificadas como **variáveis discretas** ou **variáveis contínuas**.

Variáveis numéricas podem ser, ainda, identificadas como **variáveis discretas** ou **variáveis contínuas**.

Variáveis discretas apresentam valores numéricos que surgem a partir de um processo de contagem. “A quantidade de canais de TV a Cabo Premium que você assina” é um exemplo de uma variável numérica discreta, uma vez que a resposta corresponde a um entre uma quantidade finita de números inteiros.

Variáveis contínuas produzem respostas numéricas que surgem a partir de um processo de medição. O tempo que você espera pelo atendimento de um caixa no banco é um exemplo de variável numérica contínua, uma vez que a resposta pode assumir qualquer valor dentro dos limites de um continuum, ou de um intervalo.

Figura 1: Classificação de uma variável



Em uma primeira análise, identificar o tipo da variável pode parecer fácil, embora algumas variáveis que você poderia desejar estudar possam ser categóricas ou numéricas, dependendo do modo como você as define. Por exemplo, “idade” aparentaria ser uma variável numérica evidente, mas o que acontece se você estiver interessado em comparar os hábitos de compra de crianças, adolescentes, pessoas de meia-idade e pessoas com idade para aposentadoria? Nesse caso, definir “idade” como uma variável categórica faria mais sentido.

Assim, apesar de não indicado na maioria das vezes, podemos transformar variáveis quantitativas em qualitativas. Considere, por exemplo, que a estatura de um homem adulto varie entre 1,40 e 2,20 m. Quando coletarmos dados dos elementos da população, podemos encontrar absolutamente qualquer valor dentro deste intervalo. Caso não queira usar os dados em seu estado bruto, quantitativo, podemos transformar a variável em qualitativa categorizando-a em grupos, por exemplo, baixos (menor que 1,60 m), médios (1,60 – 1,80 m) e altos (maior que 1,80m).

Assim, apesar de não indicado na maioria das vezes, podemos transformar variáveis quantitativas em qualitativas. Considere, por exemplo, que a estatura de um homem adulto varie entre 1,40 e 2,20 m. Quando coletarmos dados dos elementos da população, podemos encontrar absolutamente qualquer valor dentro deste intervalo. Caso não queira usar os dados em seu estado bruto, quantitativo, podemos transformar a variável em qualitativa categorizando-a em grupos, por exemplo, baixos (menor que 1,60 m), médios (1,60 – 1,80 m) e altos (maior que 1,80m).

Notem, no entanto, que perdemos muita informação nesse tipo de transformação!

Podemos também atribuir valores numéricos às várias qualidades ou atributos (ou, ainda, classes) de uma variável qualitativa e depois proceder-se à análise como se esta fosse quantitativa, desde que o procedimento seja passível de interpretação.

Podemos também atribuir valores numéricos às várias qualidades ou atributos (ou, ainda, classes) de uma variável qualitativa e depois proceder-se à análise como se esta fosse quantitativa, desde que o procedimento seja passível de interpretação.

Por exemplo, variáveis dicotômicas para as quais só podem ocorrer duas realizações, usualmente chamadas sucesso e fracasso. Podemos nesse caso associar o fracasso ao valor zero e o sucesso ao valor 1.

Para cada tipo de variável existem técnicas apropriadas para resumir as informações e o maior interesse do pesquisador é conhecer o comportamento da variável, analisando a ocorrência de suas possíveis realizações. Vejamos uma maneira de se dispor um conjunto de realizações, para se ter uma ideia global sobre elas, ou seja, de sua distribuição.

Distribuições de Frequências

Um pesquisador está interessado em fazer um levantamento sobre alguns aspectos socioeconômicos dos empregados da seção de orçamentos da Companhia MB. Usando informações obtidas do departamento pessoal, ele elaborou a Tabela [CompanhiaMB](#).

Distribuições de Frequências

Tabela 1: Frequências e porcentagens dos 36 empregados da seção de orçamentos da Companhia MB segundo o grau de instrução.

Grau de instrução	Frequência n_i	Proporção f_i	Porcentagem $100f_i$
Fundamental	12	0,3333	33,33
Médio	18	0,5000	50,00
Superior	6	0,1667	16,67
Total	36	1,0000	100,00

Fonte: Morettin e Bussab, 2017

Distribuições de Frequências

Tabela 1: Frequências e porcentagens dos 36 empregados da seção de orçamentos da Companhia MB segundo o grau de instrução.

Grau de instrução	Frequência n_i	Proporção f_i	Porcentagem $100f_i$
Fundamental	12	0,3333	33,33
Médio	18	0,5000	50,00
Superior	6	0,1667	16,67
Total	36	1,0000	100,00

Fonte: Morettin e Bussab, 2017

Uma medida bastante útil na interpretação de tabelas de frequências é a proporção de cada realização em relação ao total, pois podem ser utilizadas quando se quer comparar resultados de duas pesquisas distintas.

Uma medida bastante útil na interpretação de tabelas de frequências é a proporção de cada realização em relação ao total, pois podem ser utilizadas quando se quer comparar resultados de duas pesquisas distintas.

Por exemplo, suponhamos que se queira comparar a variável grau de instrução para empregados da seção de orçamentos com a mesma variável para todos os empregados da Companhia MB. Digamos que a empresa tenha 2.000 empregados e que a distribuição de frequências seja a da próxima Tabela.

Tabela 2: frequências e porcentagens dos dos 2.000 empregados da Companhia MB segundo o grau de instrução.

Grau de instrução	Frequência n_i	Proporção f_i	Porcentagem $100f_i$
Fundamental	650	0.325	32.50
Médio	1.020	0.51	51.00
Superior	330	0.165	16.50
Total	2.000	1,0000	100.00

Tabela 2: frequências e porcentagens dos dos 2.000 empregados da Companhia MB segundo o grau de instrução.

Grau de instrução	Frequência n_i	Proporção f_i	Porcentagem $100f_i$
Fundamental	650	0.325	32.50
Médio	1.020	0.51	51.00
Superior	330	0.165	16.50
Total	2.000	1,0000	100.00

Não podemos comparar diretamente as colunas das frequências das Tabelas 1 e 2, pois os totais de empregados são diferentes nos dois casos. Mas as colunas das porcentagens são comparáveis, pois reduzimos as frequências a um mesmo total (no caso 100).

A construção de tabelas de frequências para variáveis contínuas necessita de certo cuidado. Por exemplo, a construção da tabela de frequências para a variável salário, usando o mesmo procedimento acima, **não resumirá as 36 observações** num grupo menor, pois não existem observações iguais. A solução empregada é agrupar os dados por faixas de salário.

Tabela 3: Frequências e porcentagens dos dos 2.000 empregados da seção de orçamentos da Companhia MB por faixa de salário.

Classe de Salários	Frequência n_i	Porcentagem $100f_i$
4,00 -8,00	10	27,78
8,00 -12,00	12	33,33
12,00 -16,00	8	22,22
16,00 -20,00	5	13,89
20,00 -24,00	1	2,78
Total	36	100,00

Procedendo-se desse modo, ao resumir os dados referentes a uma variável contínua, perde-se alguma informação. Por exemplo, não sabemos quais são os oito salários da classe de 12 a 16, a não ser que investiguemos a tabela original. Sem perda de muita precisão, poderíamos supor que todos os oito salários daquela classe fossem iguais ao ponto médio da referida classe, isto é, 14.

Note que estamos usando a notação $a| - b$ para o intervalo de números contendo o extremo a mas não contendo o extremo b . Podemos também usar a notação $[a, b)$ para designar o mesmo intervalo $a| - b$.

A escolha dos intervalos é arbitrária e a familiaridade do pesquisador com os dados é que lhe indicará quantas e quais classes (intervalos) devem ser usadas.

- Número pequeno de classes \rightarrow perda de informação;

A escolha dos intervalos é arbitrária e a familiaridade do pesquisador com os dados é que lhe indicará quantas e quais classes (intervalos) devem ser usadas.

- Número pequeno de classes → perda de informação;
- Número grande de classes → perda da visão geral dos dados como um conjunto;

A escolha dos intervalos é arbitrária e a familiaridade do pesquisador com os dados é que lhe indicará quantas e quais classes (intervalos) devem ser usadas.

- Número pequeno de classes → perda de informação;
- Número grande de classes → perda da visão geral dos dados como um conjunto;
- A sugestão é usar de 5 a 15 classes com a mesma amplitude;

Para construir uma distribuição de frequências separando por classes uma determinada variável podemos utilizar:

- número de classes(n_c) $\approx \sqrt{n}$ ou usamos a regra de Sturges $n_c = \ln(n)$;

Para construir uma distribuição de frequências separando por classes uma determinada variável podemos utilizar:

- número de classes(n_c) $\approx \sqrt{n}$ ou usamos a regra de Sturges $n_c = \ln(n)$;
- Amplitude da classe (c):

$$c = \frac{AT}{n_c}$$

em que $AT = \text{Maior valor} - \text{Menor valor}$.

Considere

- $x_{(1)}$ como o menor valor do conjunto de dados;
- $x_{(n)}$ como o maior valor do conjunto de dados;

Considere

- $x_{(1)}$ como o menor valor do conjunto de dados;
- $x_{(n)}$ como o maior valor do conjunto de dados;
- Defina $Li_1 = x_{(1)}$, $Ls_k = Li_k + c$, $1 \leq k \leq n_c$ e $Li_k = Ls_{(k-1)}$, $2 \leq k \leq n_c$. Em que Li_k representa o limite inferior da classe k e Ls_k representa o limite superior da classe k .

Considere

- $x_{(1)}$ como o menor valor do conjunto de dados;
- $x_{(n)}$ como o maior valor do conjunto de dados;
- Defina $Li_1 = x_{(1)}$, $Ls_k = Li_k + c$, $1 \leq k \leq n_c$ e $Li_k = Ls_{(k-1)}$, $2 \leq k \leq n_c$. Em que Li_k representa o limite inferior da classe k e Ls_k representa o limite superior da classe k .
- A frequência absoluta (f_a) de uma classe k é encontrada contabilizando os valores x_1, x_2, \dots, x_n que pertencem ao intervalo $[Li_k, Ls_k)$.

Referências



Morettin, Pedro Alberto e Wilton Oliveira Bussab (2017).
Estatística básica. Saraiva Educação.