Análise Multivariada de Dados - Aula 00

Fundamentos: Tipos de Variáveis, Estatística Descritiva, Principais Variáveis Aleatórias, Amostragem

Kaique Matias de Andrade Roberto

Ciências Atuariais

HECSA - Escola de Negócios

FIAM-FAAM-FMU

Conteúdo

Tipos de Variáveis

Tipos de Variáveis

Definição 1.1

Medidas que descrevem diferenças em tipo ou natureza indicando a presença ou ausência de uma característica ou propriedade, são chamadas de **dados não-métricos ou qualitativos**.

Tipos de Variáveis

As variáveis não métricas ou qualitativas representam características de um indivíduo, objeto ou elemento que não podem ser medidas ou quantificadas; as respostas são dadas em categorias.

Tipos de Variáveis

Um erro comum encontrado em trabalhos que utilizam variáveis qualitativas representadas por números é o cálculo da média da amostra, ou de qualquer outra medida-resumo.

Tipos de Variáveis

O pesquisador calcula, inicialmente, a média dos limites de cada faixa, supondo que esse valor corresponde à média real dos consumidores situados naquela faixa; mas como a distribuição dos dados não é necessariamente linear ou simétrica em torno da média, essa hipótese é muitas vezes violada.

Tipos de Variáveis

Exemplo 1.2

Imagine que um questionário será aplicado para levantar dados da renda familiar de uma amostra de consumidores, com base em determinadas faixas salariais.

Tipos de Variáveis

A Tabela abaixo apresenta as categorias das variáveis.

Classe	Salários Mínimos (SM)	Renda Familiar (em reais)
A	Acima de 20 SM	Acima de 24240 reais
B	10 a 20 SM	De 12120 até 24240 reais
C	4 a 10 SM	De 4848 até 12120 reais
D	2 a 4 SM	De 2424 até 4848 reais
E	Até 2 SM	Até 2424 reais

Tipos de Variáveis

Observe que ambas as variáveis são qualitativas, já que os dados são representados por faixas.

Tipos de Variáveis

Porém, é muito comum a classificação incorreta por parte dos pesquisadores quando a variável apresenta valores numéricos nos dados.

Tipos de Variáveis

Nesse caso, é possível apenas o cálculo de frequências, e não de medidas-resumo, como média e desvio-padrão.

Tipos de Variáveis

Uma possível tabela de frequências obtidas para cada faixa de renda é apresentada abaixo.

Frequências	Renda Familiar (em Reais)
10%	Acima de 24240 reais
18%	De 12120 até 24240 reais
24%	De 4848 até 12120 reais
36%	De 2424 até 4848 reais
12%	Até 2424 reais

Tipos de Variáveis

Para que haja condições de se calcular medidas-resumo, como média e desvio-padrão, a variável em estudo deve ser, necessariamente, quantitativa.

Tipos de Variáveis

Definição 1.3

Os **dados métricos ou quantitativos** são utilizados quando indivíduos diferem em quantia ou grau em relação a um atributo em particular.

Tipos de Variáveis

As variáveis métricas ou quantitativas representam características de um indivíduo, objeto ou elemento resultantes de uma contagem (conjunto finito de valores) ou de uma mensuração (conjunto infinito de valores).

Tipos de Variáveis

Exemplo 1.4

No banco de dados abaixo, as variáveis Idade, Peso e Altura são quantitativas.

Nome	Idade (anos)	Peso (kg)	Altura (m)
Mariana	48	62	1,60
Luiz	54	84	1,76
Roberta	41	56	1,62
Leonardo	30	82	1,90
Melissa	28	54	1,68
Sandro	50	70	1,72

Tipos de Variáveis

Há várias maneiras de representar uma variável métrica, como veremos à seguir.

Tipos de Variáveis

Temos representações Gráficas:

- gráfico de linhas;
- dispersão;
- histograma;
- ramo e-folhas;
- boxplot;

Tipos de Variáveis

medidas de posição ou localização:

- média;
- mediana;
- moda;
- quartis;
- decis;
- percentis;

Tipos de Variáveis

medidas de dispersão ou variabilidade:

- amplitude;
- desvio-médio;
- variância;
- desvio-padrão;
- erro-padrão;
- coeficiente de variação.

Tipos de Variáveis

Estas variáveis podem ser discretas ou contínuas. As variáveis discretas podem assumir um conjunto finito ou enumerável de valores que são provenientes, frequentemente, de uma contagem, por exemplo, o número de filhos $(0, 1, 2, ...)$.

Tipos de Variáveis

Já as variáveis contínuas assumem valores pertencentes a um intervalo de números reais, por exemplo, peso ou renda de um indivíduo.

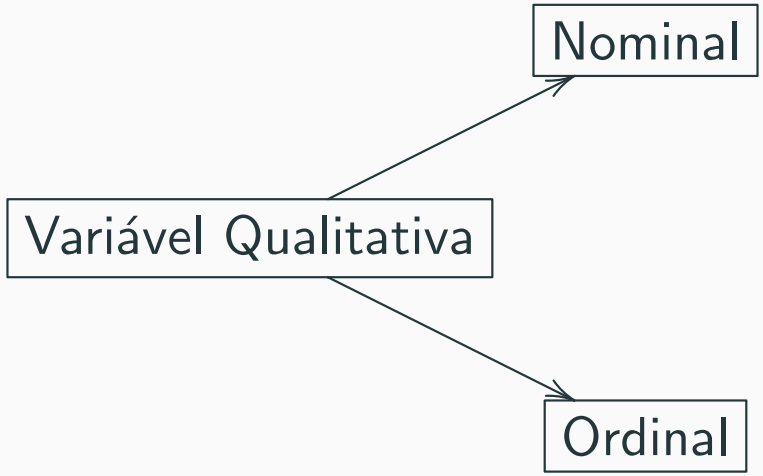
Tipos de Variáveis

As variáveis ainda podem ser classificadas de acordo com o nível ou escala de mensuração.

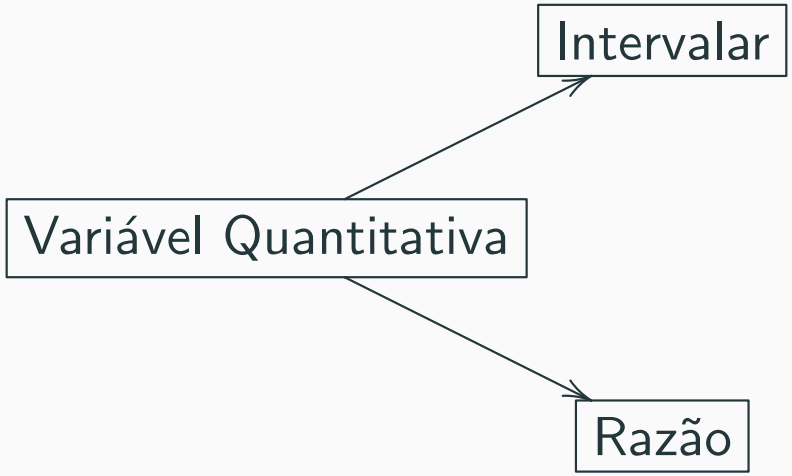
Tipos de Variáveis

Mensuração é o processo de atribuir números ou rótulos a objetos, pessoas, estados ou eventos de acordo com as regras específicas para representar quantidades ou qualidades dos atributos.

Tipos de Variáveis



Tipos de Variáveis



Tipos de Variáveis

Definição 1.5

A **escala nominal** classifica as unidades em classes ou categorias em relação à caraterística representada, não estabelecendo qualquer relação de grandeza ou de ordem. É denominada nominal porque as categorias se diferenciam apenas pelo nome.

Tipos de Variáveis

Podem ser atribuídos rótulos numéricos às categorias das variáveis, porém, operações aritméticas como adição, subtração, multiplicação e divisão sobre esses números não são admissíveis. A escala nominal permite apenas algumas operações aritméticas mais elementares.

Tipos de Variáveis

Por exemplo, pode-se contar o número de elementos de cada classe ou ainda aplicar testes de hipóteses referentes à distribuição das unidades da população nas classes.

Tipos de Variáveis

Desta forma, a maioria das estatísticas usuais, como média e desvio-padrão, não tem sentido para variáveis qualitativas de escala nominal.

Tipos de Variáveis

Exemplo 1.6

Como exemplos de variáveis não métricas em escalas nominais, podemos mencionar profissão, religião, cor, estado civil, localização geográfica ou país de origem.

Tipos de Variáveis

Definição 1.7

Uma variável não métrica em **escala ordinal** classifica as unidades em classes ou categorias em relação à característica representada, estabelecendo uma relação de ordem entre as unidades das diferentes categorias.

Tipos de Variáveis

A escala ordinal é uma escala de ordenação, designando uma posição relativa das classes segundo uma direção. Qualquer conjunto de valores pode ser atribuído às categorias das variáveis, desde que a ordem entre elas seja respeitada.

Tipos de Variáveis

Assim como na escala nominal, operações aritméticas (somas, diferenças, multiplicações e divisões) entre esses valores não fazem sentido. Desse modo, a aplicação das estatísticas descritivas usuais também é limitada para variáveis de natureza nominal.

Tipos de Variáveis

Exemplo 1.8

Exemplos de variáveis ordinais incluem opinião e escalas de preferência de consumidores, grau de escolaridade, classe social, faixa etária, etc.

Tipos de Variáveis

Definição 1.9

A **escala intervalar**, além de ordenar as unidades quanto à característica mensurada, possui uma unidade de medida constante. A origem ou o ponto zero dessa escala de medida é arbitrário e não expressa ausência de quantidade.

Tipos de Variáveis

Exemplo 1.10

Um exemplo clássico de escala intervalar é a temperatura medida em graus Celsius ou Fahrenheit. A escolha do zero é arbitrária e diferenças de temperaturas iguais são determinadas por meio da identificação de volumes guais de expansão no líquido usado no termômetro.

Tipos de Variáveis

A maioria das estatísticas descritivas pode ser aplicada para dados de variável com escala intervalar, com exceção de estatísticas baseadas na escala de razão, como o coeficiente de variação.

Tipos de Variáveis

Definição 1.11

A **escala de razão** ordena as unidades em relação à característica mensurada e possui uma unidade de medida constante. Por outro lado, a origem (ou ponto zero) é única e o valor zero expressa ausência de quantidade. Dessa forma, é possível saber se um valor em um intervalo específico da escala múltiplo de outro.

Tipos de Variáveis

Razões iguais entre valores da escala correspondem a razões iguais entre unidades mensuradas. Assim, escalas de razão são invariantes sob transformações de proporções positivas.

Tipos de Variáveis

Por exemplo, se uma unidade tem 1 metro e outra 3 metros, pode-se dizer que a última tem uma altura três vezes superior à da primeira.

Tipos de Variáveis

Dentre as escalas de medida, a escala de razão é a mais elaborada, pois permite o uso de todas as operações aritméticas. Além disso, todas as estatísticas descritivas podem ser aplicadas para dados de uma variável expressa em escala de razão.

Tipos de Variáveis

Exemplo 1.12

Exemplos de variáveis cujos dados podem estar na escala de razão incluem renda, idade, quantidade produzida de determinado produto e distância percorrida.

Tópicos de Estatística Descritiva

Tópicos de Estatística Descritiva

A **estatística descritiva** descreve e sintetiza as características principais observadas em um conjunto de dados por meio de tabelas, gráficos e medidas-resumo, permitindo ao pesquisador melhor compreensão do comportamento dos dados.

Tópicos de Estatística Descritiva

A análise é baseada no conjunto de dados em estudo (amostra), sem tirar quaisquer conclusões ou inferências acerca da população.

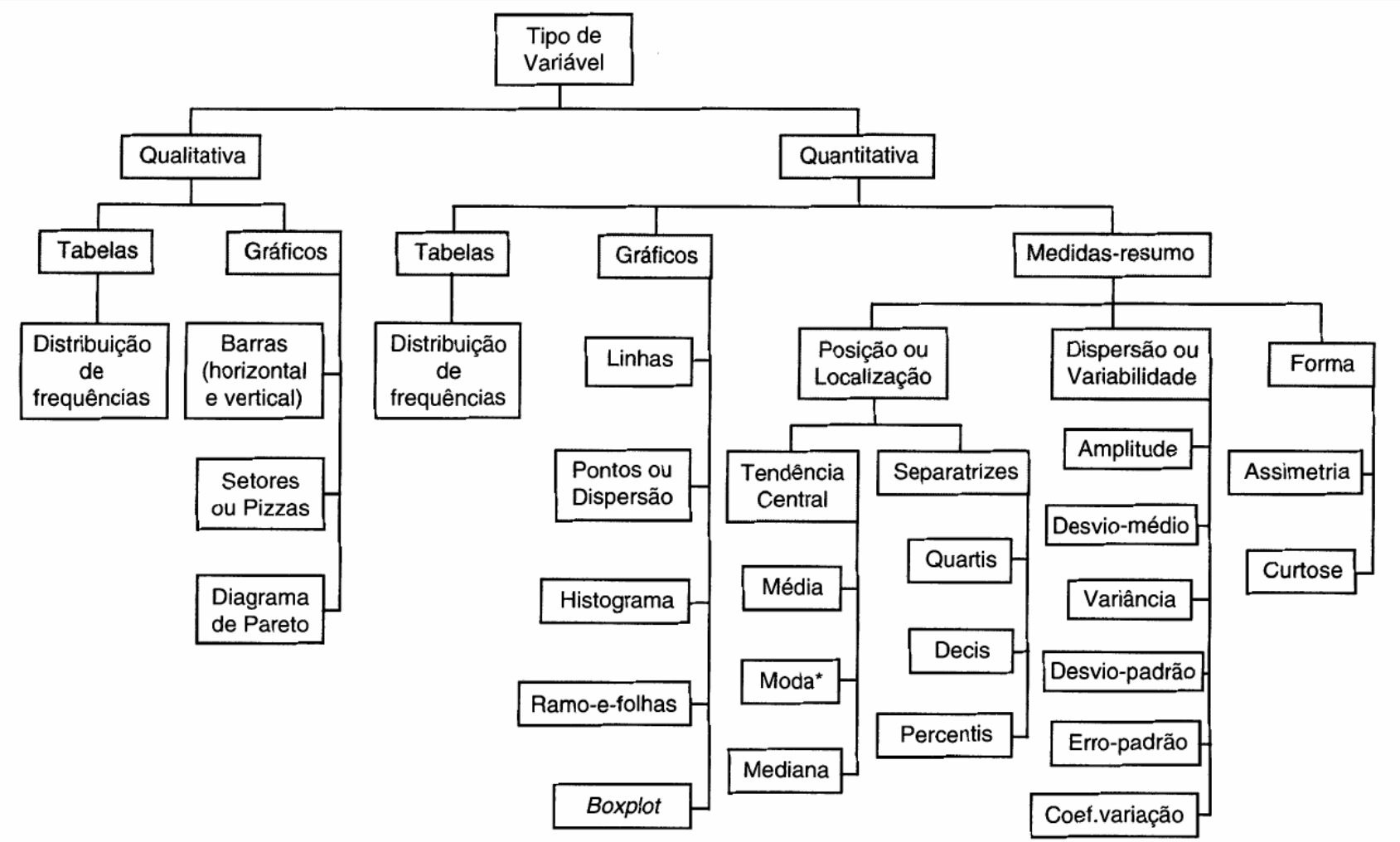
Tópicos de Estatística Descritiva

Antes de iniciarmos o uso da estatística descritiva, é necessário identificarmos o tipo de variável a ser estudada.

Tópicos de Estatística Descritiva

O tipo de variável é crucial no cálculo de estatísticas descritivas e na representação gráfica de resultados.

Tópicos de Estatística Descritiva



Exercício 2.1

Considerando os dados da Tabela abaixo

Nome	Idade (anos)	Peso (kg)	Altura (m)
Mariana	48	62	1,60
Luiz	54	84	1,76
Roberta	41	56	1,62
Leonardo	30	82	1,90
Melissa	28	54	1,68
Sandro	50	70	1,72

calcule a média, variância e desvio-padrão das variáveis Idade, Peso e Altura.

Tópicos de Estatística Descritiva

Caso você não se sinta seguro com os conceitos da Estatística Descritiva, recomendo que resolva o Exercício 0.4 da Lista de Exercícios (Resumo de Estatística Descritiva).

Introdução à Amostragem

Definição 3.1

A **população** é o conjunto com todos os indivíduos, objetos ou elementos a serem estudados, que apresentam uma ou mais características em comum. O **censo** é o estudo dos dados relativos a todos os elementos da população.

As populações podem ser finitas ou infinitas. As populações finitas são de tamanho limitado, permitindo que seus elementos sejam contados; já as populações infinitas são de tamanho ilimitado, não permitindo a contagem dos elementos.

Exemplo 3.2

Como exemplos de populações finitas, podemos mencionar a quantidade de empregados em determinada empresa, de associados em um clube, de produtos fabricados em determinado período, etc.

Quando o número de elementos da população, embora possa ser contado, for muito grande, assumimos que a população é infinita.

Exemplo 3.3

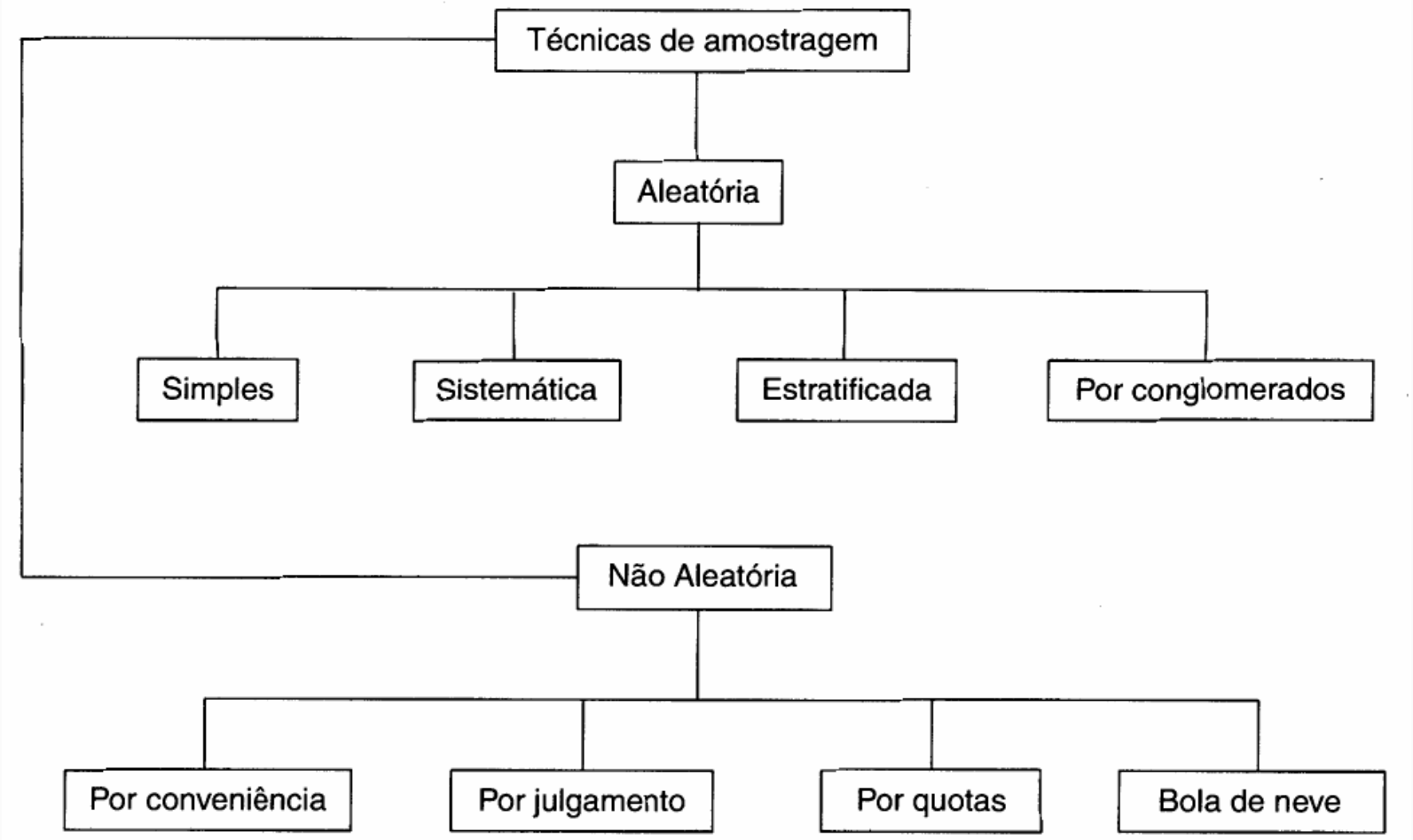
Quando o número de elementos da população, embora possa ser contado, for muito grande, assumimos que a população é infinita. São exemplos de populações consideradas infinitas a quantidade de habitantes no mundo, de residências existentes no Rio de Janeiro, de pontos em uma reta, etc.

Desta forma, existem situações em que o estudo com todos os elementos da população é impossível ou indesejável, de modo que a alternativa seja extrair um subconjunto da população em análise, denominado **amostra**.

Introdução à Amostragem

A amostra deve ser representativa da população em estudo, daí a importância deste capítulo. A partir das informações colhidas na amostra e utilizando procedimentos estatísticos apropriados, os resultados obtidos podem ser utilizados para generalizar, inferir ou tirar conclusões acerca da população (inferência estatística).

Introdução à Amostragem



Caso os conceitos de Amostragem sejam uma completa novidade para você, recomendo que resolva o Exercício 0.7 da Lista de Exercícios (Resumo de Amostragem).

Variáveis Aleatórias e as Principais Distribuições

Variáveis Aleatórias e as Principais Distribuições

Vamos relembrar alguns conceitos e terminologias relacionados à Teoria das Probabilidades.

Definição 4.1

Um **experimento** consiste em qualquer processo de observação ou medida. Um **experimento aleatório** é aquele que gera resultados imprevisíveis, de modo que, se o processo for repetido inúmeras vezes, torna-se impossível prever seu resultado.

Exemplo 4.2

O lançamento de uma moeda ou de um dado são exemplos de experimentos aleatórios.

Definição 4.3

O espaço amostral S consiste em todos os possíveis resultados de um experimento.

Exemplo 4.4

Por exemplo, no lançamento de uma moeda, podemos obter cara (k) ou coroa (c). Logo,

$$S = \{k, c\}.$$

Já no lançamento de um dado, o espaço amostral é representado por

$$S = \{1, 2, 3, 4, 5, 6\}.$$

Definição 4.5

Um **evento** é qualquer subconjunto de um espaço amostral.

Exemplo 4.6

Por exemplo, o evento A contém apenas as ocorrências pares do lançamento de um dado. Logo,

$$A = \{2, 4, 6\}.$$

Definição 4.7 (Variável Aleatória)

Consideremos $\mathcal{E}$ um experimento aleatório e S o espaço amostral associado ao experimento. A função X que associa a cada elemento $s \in S$ um número real $X(s)$ é denominada **variável aleatória**.

Variáveis Aleatórias e as Principais Distribuições

As variáveis aleatórias podem ser discretas ou contínuas. Em ambos os casos, estamos interessados em associar uma distribuição de probabilidades para a variável aleatória, bem como calcular a sua esperança, variância, distribuição acumulada e etc.

Dentre as principais distribuições discretas, ressaltamos:

- Distribuição Uniforme;
- Distribuição Bernoulli e Binomial;
- Distribuição Poisson;
- Distribuição Geométrica e Hipergeométrica.

Variáveis Aleatórias e as Principais Distribuições

Dentre as principais distribuições contínuas, ressaltamos:

- Distribuição Uniforme;
- Distribuição Exponencial;
- Distribuição Normal;
- Distribuição t de Student;
- Distribuição Qui-quadrado;
- Distribuição Gamma.

Caso você não se sinta seguro com os conceitos desta Seção, recomendo que resolva os Exercícios 0.5 e 0.11 da Lista de Exercícios (Resumo de Probabilidade e Resumo de Variáveis Aleatórias).

Comentários Finais

Comentários Finais

Em resumo, na aula de hoje nós:

- recapitulamos os tipos de variáveis;
- aprendemos como classificar variáveis;
- recapitulamos alguns tópicos de Estatística Descritiva;
- tivemos um primeiro contato com os conceitos de Amostragem;
- relembramos o conceito de Variável Aleatória;
- listamos as principais distribuições discretas e contínuas.

Comentários Finais

Na próxima aula nós vamos focar em:

- descrever os objetivos de uma análise multivariada;
- entender os conjuntos multivariados;
- entender como os conjuntos multivariados são representados.

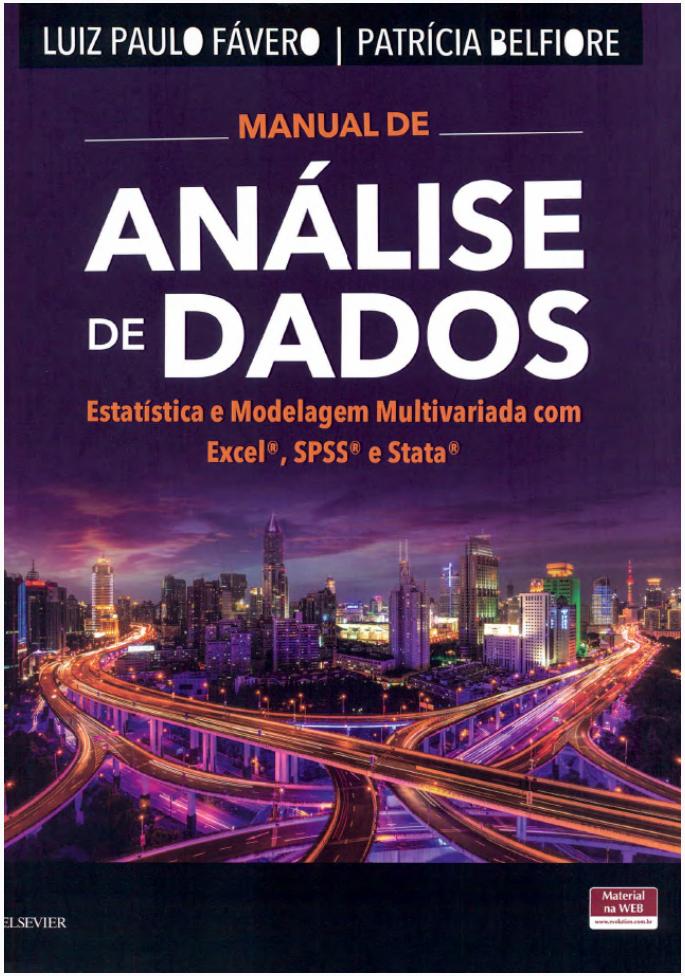
Comentários Finais

ATIVIDADE PARA ENTREGAR (E COMPOR A NOTA N1)

Resolva em grupos de até 4 integrantes os Exercícios 0.6, 0.9, 0.12 e 0.13.

Referências

Referências



Bons Estudos!