

resumo

Introdução

Elemento Amostral: Unidade sendo observada

Observação (valor): conjunto das características (variáveis) > Cada linha da planilha.

Características: são chamadas de **variáveis** > Colunas da planilha

Variáveis

Natureza Qualitativa: Categoriza, Nomeia e Diferencia

- Nominais: não tem ordem > sexo, religião...
- Ordinais: são ordenáveis > classe social, escolaridade...

Natureza Quantitativa: São numericamente mensuráveis

- Discreta: números inteiros > nº de alunos e filhos...
- Contínua: pode assumir qualquer valor real > peso, altura, salário...

obs: idade é quanti porém em faixa está como categorizada, vira **Quantitativa Categorizada**

Análise gráfica

Escala Nominal

Escalas de medida (Leitura prévia)

- Escala nominal – para variáveis essencialmente qualitativas/categorizadas que não expressam ordem;

- Escala ordinal – para variáveis qualitativas/categorizadas que expressam ordem e quantitativas que foram categorizadas;
- Escala discreta (intervalar) – para variáveis quantitativas que expressam as medidas em termo de unidade fixa;
- Escala contínua (quociente) – para variáveis quantitativas em que qualquer quociente entre dois valores dessa variável for significativo (também chamadas de escala razão).

Escala nominal (Leitura prévia)

- Utilizam-se gráficos que não expressam qualquer tipo de ordem ou importância entre as categorias:
 - exemplo gráfico de setor/pizza (para poucas categorias);
 - barras ou colunas (para muitas categorias).

Exceção: Pareto – a ordem das categorias não importa, e sim a frequência de cada uma.

Exercícios:

- Caracterize as variáveis de 1 a 10 do questionário em nominal, ordinal, ou quantitativa categorizada.
- Elimine os outliers (pessoas que não são da classe A e B e que estão na planilha).
- Construa um gráfico de setor para representar a questão 4 "Como você se locomove em São Paulo?". Você achou o gráfico adequado? Descreva o que acredita que se perdeu com a utilização desse tipo de gráfico.

Escala Ordinal

- Deve-se evitar o gráfico de setor e utilizar gráficos em que a ordem fique clara;

- barras ou colunas.

Exceção: variáveis quantitativas que foram categorizadas. A categorização será estudada mais adiante.

Quando há respostas múltiplas, a tabela dinâmica não poderá ser utilizada porque ela

considera que um conjunto de escolhas qualquer é um item distinto.

- O gráfico dinâmico só pode ser utilizado com uma tabela dinâmica.
- A lógica da tabela dinâmica é sempre colocar os itens em ordem alfabética.
- Com uma escala ordinal, esteja atento à ordem natural da variável.

Escala

Likert: Ao contrário de responder apenas "sim" ou "não", ao dar uma nota em uma escala, o respondente mostra mais especificamente o quanto ele concorda ou discorda de uma atitude ou ação, ou quanto ele acredita que algo é importante ou não.

As escalas Likert mais comuns possuem cinco, sete ou nove itens.

- Concordo plenamente (extremamente importante)
- Concordo parcialmente (muito importante)
- Nem concordo nem discordo (importante)
- Discordo parcialmente (pouco importante)
- Discordo totalmente (nada importante)

Pode também ser utilizada em relação à frequência.

- Sempre
- Frequentemente

- Às vezes
- Raramente
- Nunca

Controvérsia: ordinal x discreta

Exemplo 1 - Escala Likert

Registro Ordinal

No Excel, clique em CTRL+U e substitua

Análise intervalar (Discreta)

5 = A maior parte do tempo
 4 = Uma boa parte do tempo
 3 = Alguma parte do tempo
 2 = Uma pequena parte do tempo
 1 = Nenhuma parte do tempo

Diferencial semântico [Osgood]

- É uma escala que vai de um extremo a outro. O respondente só precisa identificar sua percepção na escala. > Quente - Frio

Escala Stapel

- É uma modificação da escala de Osgood.
- Apresenta valores numéricos e dez pontos.

Fórmulas Excel

Função (utilizada para realizar contagens em respostas múltiplas)

=cont.ses(coluna com os dados; "

"&"parte do texto de busca"&"")

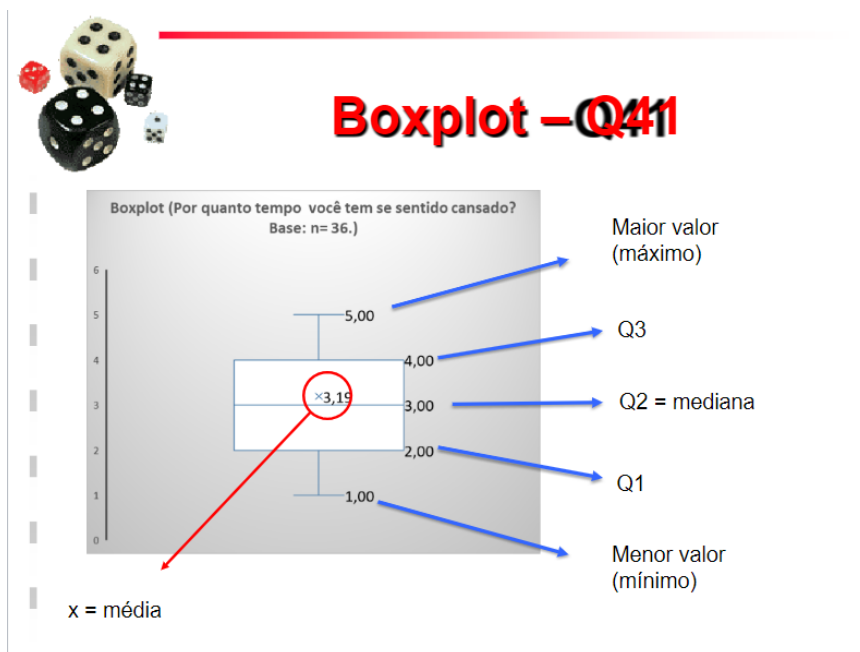
Atenção

& serve para concatenar

* indica que irá buscar em qualquer parte da célula

Construção Boxplot

- Selecione os dados;
- Na aba INSERIR, escolher nas opções de gráfico a opção: Caixa ou caixa estreita.



Escala Contínua (razão ou quociente)

Nela, a ideia de continuidade deve ficar clara (evitar barras ou colunas, porque o espaço entre as barras ou colunas dá a ideia de

descontinuidade).

Gráficos apropriados:

- boxplot;
- histograma.

Observação: o histograma é a melhor opção para descrever variáveis em escala contínua, contudo será preciso realizar um processo de categorização da escala.



- Utilizado para representar graficamente distribuições de variáveis contínuas.
- Em estatística, todas as barras precisam ter a mesma amplitude para que o gráfico seja considerado um histograma.
- Sequência de retângulos justapostos com a mesma amplitude nos intervalos de classe.

Amostragem

Quando fazer amostragem?

- diminuir custo
- aumentar velocidade na caracterização (medidas que variam no tempo)

Quando fazer censo?

- população pequena
- precisão completa (não se permite erros)

Tamanho da amostra?

depende:

- da variabilidade original dos dados (maior variabilidade \Rightarrow maior n)
- da precisão requerida no trabalho (maior precisão \Rightarrow maior n)
- do tempo disponível (menor o tempo \Rightarrow menor n)
- do custo da amostragem (maior o custo \Rightarrow menor n)
- E principalmente: da "técnica utilizada"!!!

Como amostrar?

Amostragem probabilística:

- cada elemento da população tem uma probabilidade (não nula) de ser escolhido;

Amostragem não probabilística:

- amostragem restrita aos elementos que se tem acesso (ex: amigos no Facebook);
- amostragem intencional, sem sorteio (ex: compradores de perfume no Shopping Iguatemi);
- voluntários (ex: testes "cegos");
- bola de neve (um passa para o outro);
- quotas.

Amostragem aleatória simples (AAS)

- A população é numerada de 1 a N.
- Por meio de uma tabela de números aleatórios, números compreendidos entre 1 e N são selecionados.
- Esse processo é equivalente a um sorteio no qual se colocam todos os números misturados dentro de uma urna. Os elementos correspondentes aos números escolhidos formarão a amostra.

Amostragem estratificada

-

Quando os elementos da população estão divididos em grupos não superpostos, é mais fácil e mais eficiente escolher, independentemente, uma amostra aleatória simples dentro de cada um destes grupos, os quais são chamados estratos.

- Esta forma de amostragem é uma das mais utilizadas, já que a maioria das populações têm estratos bem definidos: os homens e as mulheres; os alunos das escolas X, Y, Z; os operários pertencentes às classes salariais A, B, C, D etc.

Amostragem Sistemática

-

Se o tamanho da população é desconhecido, não podemos determinar exatamente o valor

de r . Escolheremos intuitivamente um valor razoável para r .

- Às vezes a amostragem sistemática é preferida à amostragem aleatória simples, porque é mais fácil de executar, estando portanto sujeita a erros, e proporciona mais informações com menor custo.

Amostragem por conglomerado

-

Uma amostra por conglomerado é uma amostra aleatória simples na qual cada unidade de amostragem é um grupo, ou conglomerado, de elementos. Exemplo: PNAD.

- Todos os itens dentro dos grupos (conglomerados) selecionados farão parte da amostra.
- A amostragem por conglomerado é recomendada quando: Ou não se tem um sistema de referência listando todos os elementos da população, ou a obtenção dessa listagem é dispendiosa;
- O custo da obtenção de informações cresce com o aumento da distância entre os elementos 20

Medidas

- Desvio padrão: pode ser populacional (dividido por n) ou amostral (dividido por n-1)
- É uma medida de dispersão. Na prática, o desvio-padrão determina um intervalo, centrado na média aritmética, no qual a maior parte dos dados está concentrada. Utilizamos o desvio amostral (despavd.a)
- Desvio elevado ao quadrado é chamado de variância
- Essas medidas são chamadas de medidas de resumo e são muito úteis na comparação de dois grupos, sem necessidade da construção de um gráfico.

Comparação relativa das médias entre os grupos

Como compará-las?

Sexo	Média	Desvio
feminino	2,55	1,07
masculino	1,75	1,01
Com todos os dados	2,03	1,10

a) Calcule a diferença entre as médias [$2,55 - 1,75 = 0,80$]
b) Divida a diferença pelo grupo com a menor média [$0,80/1,75 = 0,4545$]
c) Deixe o resultado em % [45,45%]

Obs. Realizar todos os cálculos, vinculando as células.

Outras medidas de tendência central

- Mediana (=med(...)): mediana: divide as observações em duas partes
- Moda: valor que aparece com mais frequência. Quando os dados estão agrupados, define-se a classe modal (classe com a maior frequência).

Outras medidas de dispersão (variabilidade)

- Intervalo interquartil (Média): diferença entre o valor do terceiro quartil em relação ao primeiro quartil ($IIQ = Q3 - Q1$). Em geral, é utilizado em conjunto com a mediana.
- Amplitude: diferença entre o maior (=máximo(...)) e o menor (=mínimo(...)) valor que uma variável pode assumir. Em geral, é utilizado em conjunto com a moda.