



Universidade Federal do **ABC**

BIN0406: Introdução à Probabilidade e Estatística

Caderno

Professor: Dr. Antonio Sergio Munhoz

Caio César Carvalho Ortega

RA 21038515

1 Prólogo

As anotações e considerações que se seguem foram realizadas de maneira autônoma e não são fruto de orientação por parte da universidade e/ou de qualquer membro do corpo docente. Foram realizadas para fins de estudo, sendo, portanto, reflexo de um esforço de cunho pessoal para melhor apreensão do conteúdo discutido em sala.

Início da aula de 25/06/2019

2 Introdução à Estatística Descritiva

Conteúdo proposto:

- Conceitos básicos
 - Elemento
 - Dado e variável
 - Valor e conjunto de valores
 - Tipo de dado e tipo de variável
- Tabelas
 - Frequência
- Gráficos
 - Variável qualitativa
 - Variável quantitativa

A coisa mais importante na Estatística são os dados, designados como elementos, que podem ser muita coisa, uma entidade, um sujeito, uma pessoa, uma observação, um experimento. Por exemplo: o professor nasceu em Avaré, eis um dado; dizer que o céu desta noite está estrelado é uma observação, ou seja, um dado que veio de uma observação; já a ideia do experimento é ser um dado fruto de um processo controlado, como uma receita de bolo. Tudo isso são dados.

O dado é um fato, uma característica, de um elemento. Nacionalidade, gentílico, cidade-natal, avaliação do bolo produzido com a receita (se ficou gostoso ou não).

Valor é um agrupamento de dados, ou ainda, valores que os dados poderão ter. Podemos falar também em conjunto de valores.

Tabela 1: Opinião dos funcionários sobre troca de gerente (G = gosta; N = não gosta)

Opinião	Frequência absoluta	Frequência relativa
G	9	45%
N	11	55%

Fonte: questão 8

Quanto ao tipo de dado ou variável, estes podem ser qualitativos (também chamados de categorizados) e quantitativos. Qualitativa: todos aqueles dados e variáveis que não são números (exemplo: local de nascimento), subdivididos em nominal (exemplo: bolo gostoso ou bolo ruim) e ordinal (exemplo: altura de uma pessoa); uma variável qualitativa pode ser uma variável discreta (que só pode assumir valores inteiros) ou contínua (que pode assumir qualquer valor, podendo ser decimal).

Tomando por base o exercício 2 do arquivo “apostila completa.pdf”, Munhoz começa a responder cada uma das questões.

- a. b
- b. 4
- c. 24
- d. Qualitativa: gênero
Quantitativa: idade, questão do aborto, classificação
- e. ?

Feito o exercício, devemos compreender os conceitos de (i) frequência absoluta, que pode ser definido como: quantidade de ocorrência de um valor ou conjunto de valores; (ii) tabela de frequência absoluta, que pode ser definida como a tabela em que as linhas expressam as frequências de cada valor ou conjunto de valores.

Munhoz agora parte para o exercício 8D no mesmo arquivo PDF. Professor inicia a construção de uma tabela de frequências absolutas. Em seguida, complementa com a coluna de frequências relativas. Vide Tabela 1.

Tabelas possuem título, corpo e fonte.

Frequência relativa é o percentual do aparecimento de cada valor.

Professor agora parte para o gráfico, dizendo que este torna a visualização mais fácil. Aponta que o gráfico pode ser de coluna ou de setores (também chamado de “gráfico de pizza”).

Desenha um gráfico de colunas e um gráfico de setores.

Figura 1: Damo-e-folhas para a pontuação

```

2 | 8 6
3 | 5
2 | 6 9 5 4
6 | 4 8 0
9 | 2 5 2 0 9

```

Fonte: questão 10E

Tabela 2: Pontuação dos alunos

Pontuação	Freq. absoluta	Freq. acum.	Freq. acum. (percent.)
50-59	3	3	15%
60-59	2	5	25%
70-79	5	10	50%
80-89	4	14	70%
90-99	6	20	100%

Fonte: questão 10E

Munhoz sugere a utilização do histograma para gráficos de variáveis quantitativas. O histograma também é conhecido como damo-e-folhas (vide Figura 1). Define-o como gráfico de colunas em faixas justapostas.

Para permitir a construção do histograma, agrupa as pontuações em classes (vide ??).

2.1 Teste no Tidia

Questões 2c, 4e, 7 frequência absoluta de A e 11 frequência acumulada percentual da sexta classe da segunda lista, cada uma correspondendo às questões 1 a 4 do teste na plataforma, respectivamente.

Início da aula de 28/06/2019

3 Medidas de posição e variabilidade

3.1 Medida de posição média para dados brutos

$$X_1, X_2, X_3$$

$$\text{Média} = M = \frac{X_1 + X_2 + \dots + X_n}{n}$$

3.2 Medida de posição média para dados agrupados

Tabela 3: Gasto por mês

Mês	Gasto
Janeiro	200
Fevereiro	500
Março	600

Fonte: sem fonte

Tabela 4: Pontuação dos alunos

Chave	X	fa	f	fX
2-6	4	2	20%	0,8
7-11	9	3	30%	2,7
12-16	14	4	40%	5,6
17-21	19	1	10%	1,9
-	-	10	11	

Fonte: questão 6E

3.1.1 Exemplo

Vide gastos na Tabela 3.

O gasto médio então é expresso por:

$$M = \frac{200 + 500 + 600}{3} = \frac{1300}{3} \cong 433$$

3.1.2 Notação

M é a média populacional e \bar{x} é a média de uma amostra; f é a frequência absoluta.

3.2 Medida de posição média para dados agrupados

Passar a limpo da foto.

3.3 Dados arupados por faixas de valores

Troca a faixa pelo ponto central (o ponto médio) e calcula a média. Veja o exemplo da Tabela 4

Logo a média é 11 ($M = 11$).

Tabela 5: Gasto por mês

Valor	$(X_i - M)^2$
200	52900
500	4900
600	28900
-	86700

Fonte: sem fonte

4 Média da variabilidade

4.1 Desvio quadrático

Desvio quadrático é a diferença do quadrado do valor:

$$(X_i - M)^2$$

Considerando novamente os gastos na Tabela 3 e que $M = 430$, temos:

Desvio quadrático do 500.

Não deu tempo de copiar

Calculemos para os gastos dos exemplos anteriores (Tabela 5).

4.2 Variância

Variância é a média dos desvios quadráticos, expressa com o símbolo σ^2 (sigma ao quadrado). Para a variância dos dados X_1, X_2, \dots, X_n com média M :

$$\sigma^2 = \frac{(X_1 - M)^2 + (X_2 - M)^2 + \dots + (X_n - M)^2}{n}$$

Adotemos novamente os gastos dos exemplos anteriores para cálculo da variância:

$$\sigma^2 = \frac{52900 + 4900 + 28900}{3} = \frac{86700}{3} = 28900$$

X	f	fX	$(X_i - M)^2$	$f(X_i - M)^2$
X_1	f_1	$f_1 X_1$	$(X_1 - M)^2$	$f_1 (X_1 - M)^2$
X_1	f_1	$f_1 X_1$	$(X_1 - M)^2$	$f_1 (X_1 - M)^2$
X_1	f_1	$f_1 X_1$	$(X_1 - M)^2$	$f_1 (X_1 - M)^2$
-	-	M	-	σ^2

Fonte: N/A

4.3 Coeficiente de variação

É o CV . A notação é $CV = \frac{\sigma}{M}$. Exemplo:

$$CV = \frac{170}{430} = 39\%$$

5 Variância para dados brutos

Sendo f = frequência relativa, temos a Seção 5.

5.1 Variância populacional e desvio populacional

Passar a limpo da foto.

6 Percentis

É uma medida de posição. Sendo o percentil um ponto entre $p\%$ e $(-1p)\%$, temos:

$$p = \frac{p \times n}{\text{arredondado pra cima}}$$

n é a quantidade de dados.

Exemplo a partir da questão 3A, alínea (d):

$$p_{75} \text{ está na posição } 75\% \times 8 = 6$$

O ranking é expresso pelos dados ordenados como (30) (36) (42) (42) (44)

46 50 54

$$\text{Então } p_{75} = \frac{46 + 50}{2} = 48$$

$$p_{75} = 48 = Q_3 \text{ (3o. quartil)}$$

$$p_{25} = Q_1 \text{ (1o. quartil)}$$

$$p_{25} = \frac{36 + 42}{2} = 39$$

$$p_{50} = Q_2 = \text{mediana} = \frac{42 + 44}{2} = 43$$

Box-plot na foto. Passar a limpo

Exemplo respondendo a questão 9C:

$$CV_{UTC} = \frac{0,84}{2,8} = 30\%$$

$$CV_{UTK} = \frac{0,84}{2,4} = 37\%$$

Logo, a UTK tem notas mais dispersas.

A P1 ficou para dia 19/7!

Responder no Tidia:

1. 3 (a + c + d)
2. 7 (a + b)
3. 10 (cidade e dispersão)
4. 17 resposta: 131

Lista de anotações

<input type="checkbox"/> Início da aula de 25/06/2019	1
<input type="checkbox"/> Início da aula de 28/06/2019	3

Passar a limpo da foto.	4
Não deu tempo de copiar	5
Passar a limpo da foto.	6
Box-plot na foto. Passar a limpo	7
A P1 ficou para dia 19/7!	7

Referências