UFABC

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. 6

b. 4

c. 24

d. Qualitativa: gênero
Quantitativa: idade, questão do aborto, classificação

e. Não

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

Média = $M = \dfrac{X_1 + X_2 + \cdots + X_n}{n}$

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 $X1, X_2, \ldots, X_n$ com média M:

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

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_1X_1	$(X_1 - M)^2$	$f_1(X_1 - M)^2$
X_1	f_1	f_1X_1	$(X_1 - M)^2$	$f_1(X_1 - M)^2$
X_1	f_1	f_1X_1	$(X_1 - M)^2$	$f_1(X_1 - M)^2$
-	-	M	-	$\sigma 2$

Fonte: N/A

4.3 Coeficiente de variação

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

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

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.

5.2 Percentis

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

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

n é a quantidade de dados.

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

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

O ranking é expresso pelos dados ordenados como 30 36 42 42 44 46 50 54

Então $p_{75} = \frac{46+50}{2} = 48$

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

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

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

$p_{50} = Q_2 =$ 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

Início da aula de 05/07/2019

6 Probabilidade: contextos básicos

- Ementa:
  - Probabilidade
  - Probabilidade da ação
  - Probabilidade condicional

6.1 Definições

Evento: conjunto de resultados.

Espaço amostral: conjunto de todos os resultados possíveis

Ponto amostral: é um elemento do espaço amostral

6.2 Exemplo

Dado o lançamento de uma moeda, são:

$A =$ espaço amostral

$A = \{cara, coroa\}$

Pontos amostrais: cara, coroa

6.3 Exemplo

Dado o lançamento de duas moedas, são:

$A =$ espaço amostral

$A = \{(cara, coroa), (coroa, cara), (cara, cara), (coroa, coroa)\}$

Pontos amostrais: (cara,coroa), (coroa,cara), (cara,cara), (coroa,coroa)

$E =$ é o evento, no caso, **sair uma cara somente**

$E = \{(cara, coroa), (coroa, cara)\} \subset A$

6.4 Definição de probabilidade

$P(E) =$ probabilidade do evento E

$E \subset A, A =$espaço amostral

$$P(E) = \frac{n(E)}{n(A)} = \frac{\text{quantidade de elementos de } E}{\text{quantidade de elementos de } A}$$

6.5 Exemplo

Considere o seguinte experimento: lançamento de uma moeda. Temos:

$A = \{cara, coroa\}$

$E =$ sair cara

$E = \{(cara)\}$

$P(E) = P(cara) = \dfrac{n(E)}{n(A)} = \dfrac{1}{2} = 50\%$

6.6 Exemplo

Considere o seguinte experimento: lançamento de duas moedas. Temos:

$E =$ sair só uma cara

$P(E) = \dfrac{2}{4} = \dfrac{1}{2} = 50\%$

6.7 Definição frequentista de probabilidade

Nela, a probabilidade (P) é igual a frequência relativa. Por exemplo, lançar uma moeda e esperar 500 caras. Segundo Bussab e Morettin (2010, p, 121), esta "se baseia na estabilidade das freqüências relativas e no fato de podermos, hipoteticamente, repetir um experimento várias vezes".

6.8 Exemplo

Referência: 3A, apostila 4

a. $\{0, 1, 2, 3, 4\}$
   5

b. 0, cujo $P = 60 \div 200 = 30\%$
   1, cujo $P = 80 \div 200 = 40\%$
   2, cujo $P = 40 \div 200 = 20\%$
   3, cujo $P = 16 \div 200 = 8\%$
   4, cujo $P = 4 \div 200 = 2\%$

c. $P(V = 0) = 30\%$

d. $P(V \geq 2) = 20\% + 8\% + 2\% = 30\%$

e. $P(V = 1$ ou $V = 2) = 40\% + 20\% = 60\%$ ou, alternativamente
$P(V = 1$ ou $V = 2) = \frac{120}{200} = 60\%$

f. $P(V = 2$ ou $V = 1$ ou $V = 0) = \frac{180}{200} = 90\%$ ou, alternativamente
$P(V2$ ou $V = 1$ ou $V = 0) = 20\% + 40\% + 30\% - 90\%$

6.9 Probabilidade da união

6.9.1 Proposição

Proposição: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Prova: $n =$ quantidade total de possibilidades $=$ tamanho do espaço amostral.

Temos:

$n_A \cup B = n_A + n_B - n_{A \cap B}$

$\frac{n_{A \cup B}}{n} = \frac{n_A}{n} + \frac{n_B}{n} - \frac{n_{A \cap B}}{n}$

Logo:

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

6.9.2 Exemplo

Professor utiliza o exemplo $n_{B \cup C} = n_B + n_C - n_{B \cap C}$, vide Figura 2.

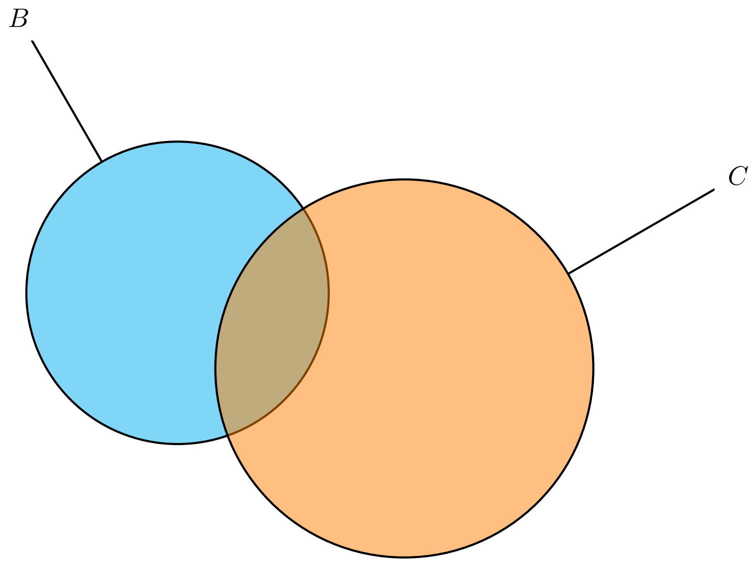
6.9.3 Definição

A e B são eventos excludentes $\leftrightarrow A \cap B = \emptyset$

6.9.4 Proposição

A e B são eventos excludentes $\leftrightarrow P(A \cup B) = P(A) + P(B)$

Figura 2: Observar interseção



6.9.5 Exemplo

Referência: exercício 1A da apostila 4

$P(A) = 23\%$

$P(A) = 19\%$

$P(A \text{ ou } B) = P(A \cup B) = 38\%$

a. $P(A \cap B) =?$

   Temos:

   $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
   $P(A \cap) B) = P(A) + P(B) - P(A \cup B)$
   $P(A \cap) B) = 23\% + 19\% - 36\% = 42\% - 38\%$
   $P(A \cap) B) = 4\%$

b. Não são eventos excludentes, pois $A \cap B \neq \emptyset$ devido a $P(A \cap B) > 0$

6.10 Probabilidade condicional

6.10.1 Definição

$$P(A|B) = \frac{N_{A\cap B}}{N_B}$$

> Se A e B são eventos que podem ocorrer em um dado experimento, a probabilidade condicional de B ter ocorrido, quando se sabe que A ocorreu, é representada por $P(B|A)$. (Leia-se probabilidade de B dado A) (CARVAJAL et al., 2009, p. 75)

6.10.2 Exemplo

Considere a retirada de bolas pretas e brancas numa caixa:

P P P B B

P(sair bola preta na 2a. vez | sair bola branca na 1a. vez) $= \frac{3}{4} = 75\%$

P(sair bola preta na 2a. vez | sair bola branca na 1a. vez) $= \frac{2}{4} = 50\%$

Experimento 1: tira uma bola

Experimento 2: tira uma bola

P P P B B $\rightarrow$ 1a. vez $\rightarrow$ P P B B $\rightarrow$ 2a. vez

6.10.3 Exemplo

$A = \{1, 2, 3, 4, 5\}$

$B = \{1, 2\}$

$C = \{7, 8, 9\}$

6.10.4 Proposição

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

6.10.5 Prova

$$P(A|B) = \frac{n_{A \cap B}}{n_B}$$

$$P(A|B) = \frac{\frac{n_{A \cap B}}{n}}{\frac{n_B}{n}} = \frac{P(A \cap B)}{P(B)}$$

6.10.6 Consequência

$$P(A \cap B) = P(A|B) \cdot P(B)$$

6.10.7 Definição

Eventos A e B são independentes $\leftrightarrow P(A|B) = P(A)$

6.10.8 Proposição

A e B são eventos independentes $\leftrightarrow P(A \cap B) = P(A) \cdot P(B)$

6.10.9 Prova

$$P(A \cap B) = P(A|B)P(B) = P(A)P(B)$$

6.10.10 Exemplo

Referência: exercício 8B da apostila 4

a. $P(\text{cliente mulher}) = \frac{30 + 50}{200} = \frac{80}{200} = 40\%$

b. $P(\text{cliente homem}) = \frac{100}{200} = 50\%$

c. $P(\text{solteira} \mid \text{mulher}) = \frac{30}{80} = 37,5\%$

d. $\frac{12\cancel{0}}{20\cancel{0}} = 60\%$

e. $P(\text{casado} \mid \text{homem}) = \frac{100}{120} = 83,3\%$

f. Não são, pois são iguais

g. $P(\text{solteira} \mid \text{mulher}) = \frac{30}{80} = 37,5\%$
$P(\text{solteiro}) = \frac{50}{200} = 25\%$

6.11 Teste no Tidia

Relação entre questão do Tidia e questão da apostila 4:

1. 7C. Resposta: 50%
2. 9D. Resposta: 49%
3. 11C. Resposta: 50%
4. 10, considerando:
   $x + y$
   $x = P(A \cap B) - P(A) \cdot P(B)$
   $y = P(A \cup B) - P(A) - P(B)$

Início da aula de 12/07/2019

Faltei, mas sei que o Munhoz avançou mais uma apostila.

Ementa:

- Diagrama de árvore
- Probabilidade do caminho
- Probabilidade inversa
- Teorema de Bayes

7 Teorema de Bayes

Uma das relações mais importantes envolvendo probabilidades condicionais Bussab e Morettin (2010, p. 116). Fórmula da versão mais simples (BUSSAB; MORETTIN, 2010, p. 116):

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A) \cdot P(B|A)}{P(B)}$$

Segundo Bussab e Morettin (2010, p. 119), "o Teorema de Bayes fornece um mecanismo formal para atualizar probabilidades", além disso, possui ainda especial importância:

> O Teorema de Bayes, que aparentemente poderia ser encarado como mais um resultado na teoria de probabilidades, tem importância fundamental, pois fornece a base para uma abordagem da inferência estatística conhecida como *inferência bayesiana*. (BUSSAB; MORETTIN, 2010, p. 119)

7.1 Exemplo

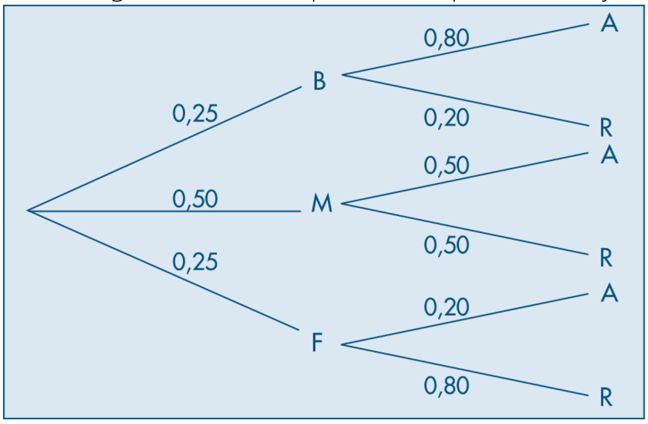
Para selecionar seus funcionários, uma empresa oferece aos candidatos um curso de treinamento durante uma semana. No final do curso, eles são submetidos a uma prova e 25% são classificados como bons (B), 50% como médios (M) e os restantes 25% como fracos (F). Para facilitar a seleção, a empresa pretende substituir o treinamento por um teste contendo questões referentes a conhecimentos gerais e específicos. Para isso, gostaria de conhecer qual a probabilidade de um indivíduo aprovado no teste ser considerado fraco, caso fizesse o curso. Assim, neste ano, antes do início do curso, os candidatos foram submetidos ao teste e receberam o conceito aprovado (A) ou reprovado (R). No final do curso, obtiveram-se as seguintes probabilidades condicionais:

$$P(A|B) = 0,80 \qquad P(A|M) = 0,50 \qquad P(A|F) = 0,20$$

Queremos encontrar $P(F|A)$ e, pelo Teorema de Bayes, essa probabilidade é dada por

$$P(F|A) = \frac{P(A|F)P(F)}{P(A|B)P(B) + P(A|M)P(M) + P(A|F)P(F)}$$

Figura 3: Diagrama em árvore para o exemplo da Subseção 7.1



Fonte: Bussab e Morettin (2010, p. 119)

$$= \frac{(0,20)(0,25)}{(0,80)(0,25) + (0,50)(0,50) + (0,20)(0,25)} = 0,10$$

Então, apenas 10% dos aprovados é que seriam classificados como fracos durante o curso. De modo análogo podemos encontrar $P(B|A) = 0,40$ e $P(M|A) = 0,50$, que poderiam fornecer subsídios para ajudar na decisão de substituir o treinamento pelo teste. (BUSSAB; MORETTIN, 2010, p. 118)

7.2 Diagrama em árvore

Auxiliar para solucionar problemas envolvendo o Teorema de Bayes, como aquele observado na Subseção 7.1.

7.3 Teste no Tidia

Relação entre questão do Tidia e questão da apostila 5.

1. 1A. Resposta: 35%
2. 2F. Resposta: $25,75\%$
3. 3. Resposta: $81,25\%$
4. 6A. Resposta: $56,25\%$

Lista de anotações

Referências

BUSSAB, W. d. O.; MORETTIN, P. A. *Estatística básica*. 6. ed. São Paulo: Saraiva, 2010. ISBN 978-85-02-08177-2.

CARVAJAL, S. et al. *Estatística básica: a arte de trabalhar com dados*. Rio de Janeiro: Elsevier, 2009. ISBN 978-85-352-3030-7.