ESTATÍSTICA BÁSICA

David Blackwell
Departamento de Estatística
Universidade da Califórnia

Tradução de:
CARLOS ALBERTO DE BRAGANÇA PEREIRA
e
WAGNER DE SOUZA BORGES
Professores do Instituto de
Matemática e Estatística da USP

EDITORA McGRAW-HILL DO BRASIL LTDA.
São Paulo — Rio de Janeiro — Belo Horizonte

1973

EDITORA McGRAW-HILL DO BRASIL, LTDA.

Rua Tabapuã, 1105 Itaim-Bibi, S. Paulo Estado de São Paulo	Av. Rio Branco, 156 - s/ 2614 Rio de Janeiro Guanabara	Rua Turmalina, 27 Belo Horizonte M. Gerais

Impresso no Brasil
Printed in Brazil

PREFÁCIO

Este livro contém material para um curso básico de Estatística Elementar, por mim ministrado diversas vezes em Berkeley. Os alunos provinham de todos os departamentos da Universidade e muitos deles já haviam esquecido a Álgebra do ginásio. O nível matemático do curso é modesto: qualquer estudante que saiba lidar com Aritmética, fazendo substituições em fórmulas simples, marcar pontos e passar uma curva suave através de pontos previamente marcados, está pronto para o curso. Mas deve estar preparado para encarar com seriedade os exemplos* frívolos, como bolas em urnas, que serão usados para ilustrar e introduzir praticamente todas as idéias. É um enfoque intuitivo, informal, concreto, do ponto de vista da teoria da decisão e Bayesiana.

DAVID BLACKWELL

---

* Vários exemplos e exercícios deste livro constam da seleção de palavras e letras em frases. Para o objetivo a que se propõem, o significado da frase não tem importância alguma. A estrutura de tais exemplos e exercícios está na quantidade e na disposição das letras e palavras que formam a frase. Com o intuito de não afetar essa estrutura, as frases em apreço não foram traduzidas. (N. dos T.)

PREFÁCIO DA EDIÇÃO BRASILEIRA

Bastante difícil é a tarefa de ministrar um curso introdutório de Probabilidade e Estatística a alunos sem formação matemática sólida. O Prof. Blackwell conseguiu, neste livro, introduzir com grande clareza e o mínimo de formalismo, as idéias básicas do assunto. Desta forma, é inestimável o serviço que a sua publicação, em lingua portuguesa, presta aos professores do assunto e alunos de nossas Universidades.

Edições preliminares da obra circularam, por alguns anos, na Universidade da Califórnia, Berkeley, sendo ela utilizada como texto em cursos ministrados para alunos, na grande maioria, de Ciências Sociais. Tive a oportunidade de constatar, na qualidade de Instrutor, que, salvo raras exceções, os alunos ignoravam a Matemática do nosso curso secundário. Mesmo assim, conseguiram, com bastante sucesso, absorver os conhecimentos ministrados.

O ponto de vista adotado neste livro é o Bayesiano, do qual o autor é um proeminente defensor.

A parte de inferência é uma ótima introdução do enfoque Bayesiano de Estatística, cuja leitura considero interessante, mesmo aos iniciados em Estatística.

Setembro de 1972

DJALMA GALVÃO CARNEIRO PESSOA
Instituto Tecnológico da Aeronáutica
S. José dos Campos

SUMÁRIO

1 PROBABILIDADE

Se uma urna contém cinco bolas e selecionamos uma, de tal forma que cada uma das cinco tenha a mesma chance de ser escolhida, diremos que a bola é *selecionada ao acaso* e que uma bola particular tem 1/5 de *probabilidade* de ser escolhida. Se, das cinco bolas, três são vermelhas, a probabilidade da bola selecionada ser vermelha é de 3/5. De uma maneira geral, quando um elemento é selecionado ao acaso, de um conjunto finito S, de forma que cada elemento de S tenha igual probabilidade de ser escolhido, a probabilidade de que o elemento selecionado pertença a um dado subconjunto A de S é definida pela razão entre o número de elementos de A e o número de elementos de S:

Para seleções ao acaso em um conjunto finito S,

$$P(A) = \frac{\text{número de elementos de } A}{\text{número de elementos de } S}.$$

EXEMPLO 1.1 Um dos cem números seguintes é selecionado ao acaso:

00	01	....	09
10	11	....	19
....			
90	91	....	99

No quadro seguinte, encontram-se alguns eventos, com suas respectivas probabilidades:

Eventos	*Probabilidades*
(*a*) O primeiro dígito é zero	0,1
(*b*) Os dois dígitos são iguais	0,1
(*c*) Os dois dígitos são diferentes	0,9
(*d*) O primeiro dígito é maior que o segundo	0,45
(*e*) O primeiro dígito é maior ou igual ao segundo	0,55
(*f*) O segundo dígito é 1	0,1
(*g*) A soma dos dígitos é 5	0,06
(*h*) A soma dos dígitos é 9	0,1
(*i*) Nenhum dos dígitos é maior que 3	0,16
(*j*) Ambos os dígitos são maiores que 3	0,36
(*k*) Apenas um dos dígitos é maior que 3	0,48
(*l*) O primeiro dígito é maior que 3 e o segundo não	0,24

Em geral, a probabilidade de qualquer evento é um número entre 0 e 1, medindo o grau de incerteza de sua ocorrência. Probabilidades satisfazem as seguintes regras:

1. $P(A) = 0$, se A é *impossível*, isto é, não pode ocorrer.
2. $P(A) = 1$, se A é *certo*, isto é, tem que ocorrer.
3. $P(A \text{ ou } B) = P(A) + P(B)$, se A e B são *mutuamente exclusivos*, isto é, não podem ocorrer juntos.
4. $P(\text{não } A) = 1 - P(A)$

EXEMPLO 1.2 Um ponto é selecionado ao acaso num quadrado S de lado 1, de forma que a probabilidade do ponto pertencer a um subconjunto A de S seja igual à área de A:

$$P(A) = \text{área de } A$$

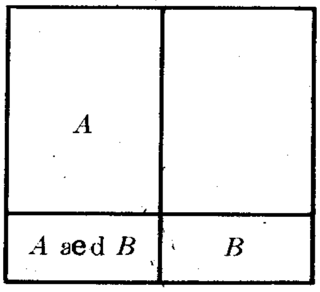


Fig. 1-1

Conforme o diagrama da fig. 1-1, onde A é a metade esquerda de S e B a quarta parte inferior de S,

$$P(A) = \tfrac{1}{2} \qquad P(B) = \tfrac{1}{4} \qquad P(A \text{ e } B) = \tfrac{1}{8}$$

pois os pontos que pertencem a A e B são os que estão no retângulo inferior à esquerda e $P(A \text{ ou } B) = 5/8$, pois todos os pontos que não pertencem ao retângulo superior à direita pertencem a, pelo menos, um entre A e B, isto é, pertencem a A ou B. Fazendo-se $C =$ não (A ou B), retângulo superior à direita,

$$P(C) = \tfrac{3}{8} \qquad P(A \text{ ou } C) = P(A) + P(C) = \tfrac{1}{2} + \tfrac{3}{8} = \tfrac{7}{8}$$

EXEMPLO 1.3 A caixa 1 contém cinco bolas numeradas de 1 a 5 e a caixa 2 contém, igualmente, cinco bolas numeradas de 6 a 10. Uma das caixas é selecionada ao acaso e, em seguida, uma bola é extraída ao acaso da caixa escolhida. Você ganha um prêmio, se o número da bola selecionada for divisível por 3, isto é, se for um dos números 3, 6 ou 9. Como a seleção de qualquer uma das 10 bolas é igualmente provável, P (prêmio) $= 3/10 = 0{,}3$. Suponha que a caixa 1 seja selecionada. Qual é, agora, a probabilidade de que você ganhe o prêmio? Como, agora, a ocorrência de qualquer um dos cinco números, 1, 2, 3, 4, e 5, é igualmente provável e apenas um deles ganha o prêmio, a probabilidade de que você o ganhe, dado que a caixa 1 foi selecionada, é $1/5 = 0{,}2$ e escrevemos

$$P\ (\text{prêmio}\ |\text{caixa 1}) = 0{,}2$$

Analogamente,

$$P\ (\text{prêmio}\ |\text{caixa 2}) = 0{,}4$$

Em geral, para quaisquer dois eventos A e B, a probabilidade de B, dado que A ocorreu, é denotada por $P(B\,|A)$ e satisfaz a fórmula:

$$\boldsymbol{P(B\,|A) = \frac{P(A \text{ e } B)}{P(B)}}$$

Em nosso exemplo, com A = caixa 1 e B = prêmio, (A e B) tem apenas o elemento 3; assim, $P(A$ e $B) = 0{,}1$. Como $P(A) = 0{,}5$, a fórmula nos dá $P(B\,|A) = 0{,}1/0{,}5 = 0{,}2$, coincidindo com o que havíamos encontrado diretamente.

A fórmula para $P(B\,|A)$ pode também ser escrita:

$$P(A \text{ e } B) = P(A)P(B\,|A)$$

e, nesta forma, ela pode ser estendida a mais de dois eventos:

$P(A$ e B e $C\ldots) = P(A)P(B\,|A)P(C\,|A$ e $B)\ldots)$ A probabilidade de que vários eventos ocorram simultâneamente é a probabilidade de que um ocorra, multiplicado pela probabilidade de que um segundo ocorra dado que o primeiro ocorreu, multiplicado pela probabilidade de que um terceiro ocorra dado que os dois primeiros tenham ocorrido juntos e assim por diante.

Por exemplo, para três eventos, temos:

$$\begin{aligned} P(A \text{ e } B \text{ e } C) &= P(A \text{ e } B)P(C\,|A \text{ e } B) = \\ &= P(A)P(B\,|A)P(C\,|A \text{ e } B) \end{aligned}$$

EXEMPLO 1.4 A urna apresentada na fig. 1-2 contém duas bolas brancas (B) e três pretas (W). Duas bolas são retiradas ao acaso da urna, sucessivamente. Calcule a probabilidade de cada um dos quatro resultados, BB, BW, WB e WW.

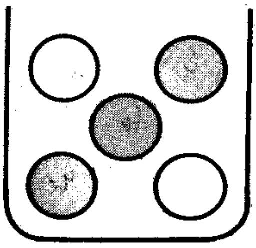


Fig. 1-2

a) se a primeira bola não é recolocada na urna antes da da segunda retirada (retiradas *sem* reposição).

b) se a primeira bola é recolocada (retiradas *com* reposição).

Seja B_1 o evento *a primeira bola é branca*, W_1 o evento *a primeira bola é preta* e assim sucessivamente. No caso (*a*) conforme diagrama da fig. 1-3,

$$\begin{aligned}
P(BB) = P(B_1 \text{ e } B_2) &= P(B_1)P(B_2 \mid B_1) = 2/5(1/4) = 0{,}1 \\
P(BW) &= P(B_1)P(W_2 \mid B_1) = 2/5(3/4) = 0{,}3 \\
P(WB) &= P(W_1)P(B_2 \mid W_1) = 3/5(2/4) = 0{,}3 \\
P(WW) &= P(W_1)P(W_2 \mid W_1) = 3/5(2/4) = 0{,}3
\end{aligned}$$

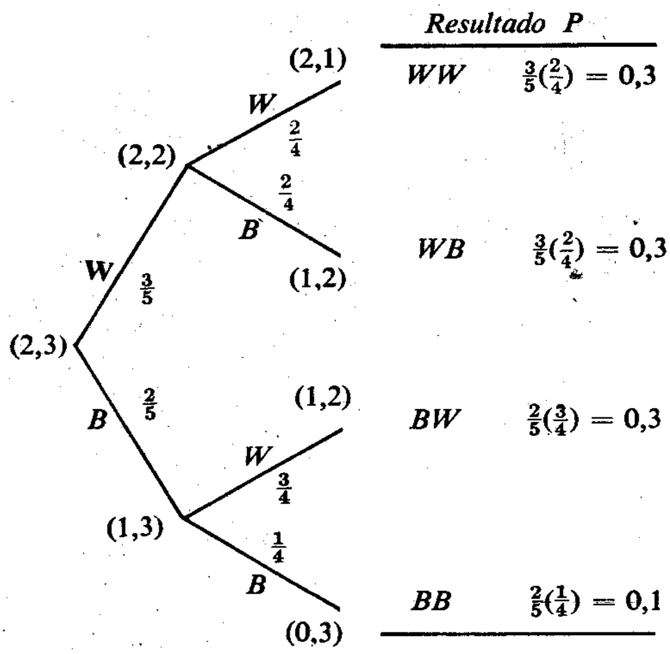
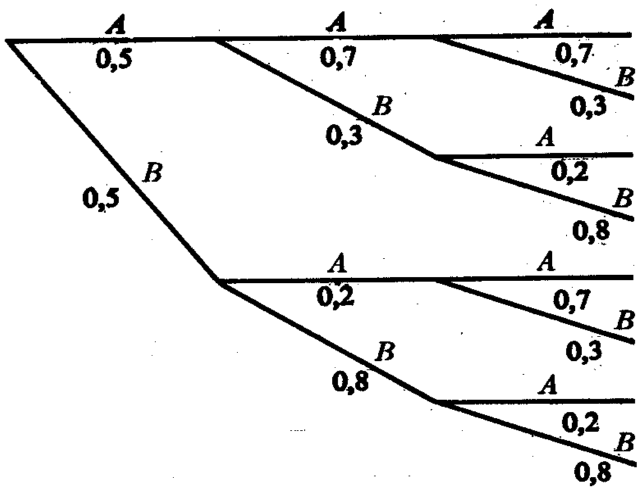


FIG. 1-3

No caso (*b*),

$$
\begin{aligned}
P(BB) &= P(B_1)P(B_2 \mid B_1) &&= 2/5(2/5) = 0{,}16\\
P(BW) &= P(B_1)P(W_2 \mid B_1) &&= 2/5(3/5) = 0{,}24\\
P(WB) &= P(W_1)P(B_2 \mid W_1) &&= 3/5(2/5) = 0{,}24\\
P(WW) &= P(W_1)P(W_2 \mid W_1) &&= 3/5(3/5) = 0{,}36
\end{aligned}
$$

EXEMPLO 1.5 Uma senhora compra determinado produto, às vezes da marca A e, às vezes, da marca B. Se ficou satisfeita com a sua útima aquisição, ela compra novamente a mesma marca; caso contrário, ela muda. Se uma determinada aquisição da marca A tem probabilidade 0,7 de ser satisfatória, enquanto que uma da marca B tem probabilidade 0,8, qual a probabilidade de que sua terceira aquisição seja da marca A, se ela joga uma moeda para decidir que marca deve comprar na primeira vez? O diagrama da fig. 1-4 mostra que P (terceira aquisição é A) $= 0{,}245 + 0{,}030 + 0{,}070 + 0{,}080 = 0{,}425$



Resultado	***P***
AAA	(0,5) (0,7) (0,7) = 0,245
AAB	(0,5) (0,7) (0,3) = 0,105
ABA	(0,5) (0,3) (0,2) = 0,030
ABB	(0,5) $\overset{0,3}{(0,4)}$ (0,8) = 0,120
BAA	(0,5) (0,2) (0,7) = 0,070
BAB	(0,5) (0,2) (0,3) = 0,030
BBA	(0,5) (0,8) (0,2) = 0,080
BBB	(0,5) (0,8) (0,8) = 0,320

FIG. 1-4

EXEMPLO 1.6 Um clube tem dois tipos de sócios: do tipo R e do tipo D. Cada ano, um membro do clube é escolhido ao acaso para proceder à escolha de um novo membro. Um

R sempre escolhe um *R* e um *D* sempre escolhe um *D*. Se, originalmente, o clube foi constituído de um sócio *R* e um sócio *D*, qual a probabilidade de que, após terem sido feitas três escolhas, o clube tenha três *R* e dois *D*? O dagrama da fig. 1-5 mostra que

$$P(3,2) = \tfrac{1}{12} + \tfrac{1}{12} + \tfrac{1}{12} = \tfrac{1}{4}$$

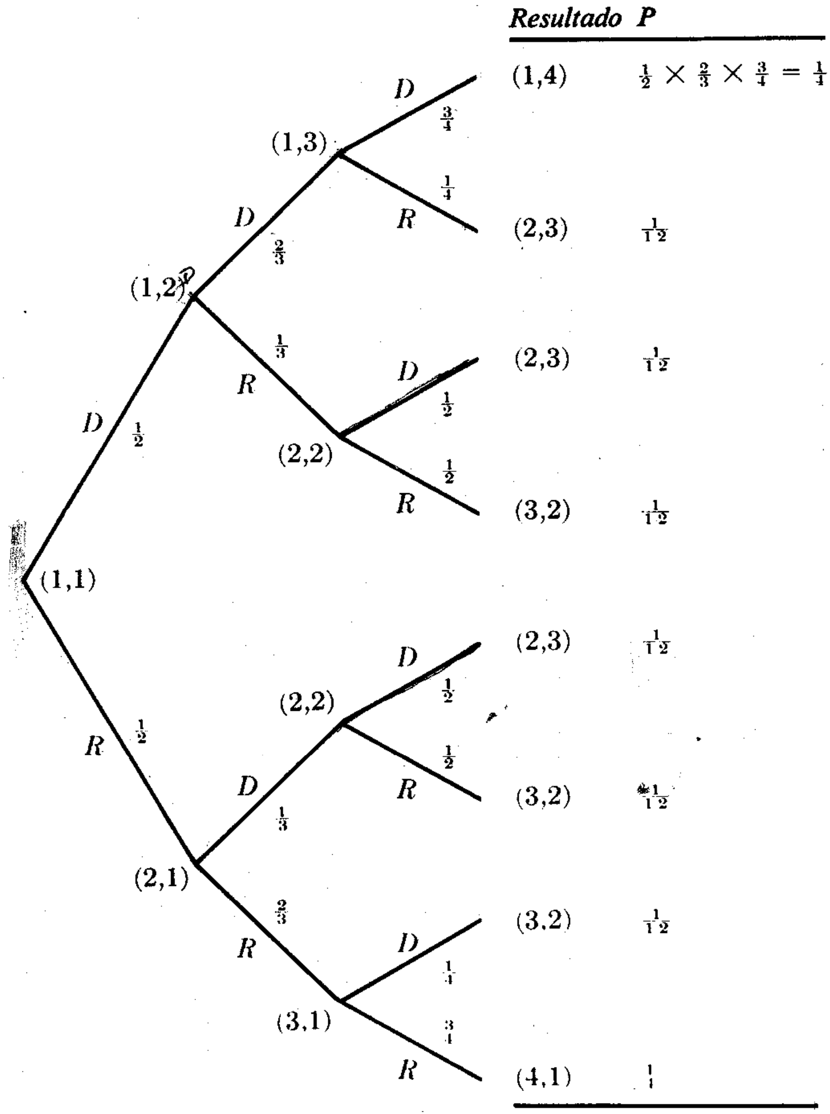
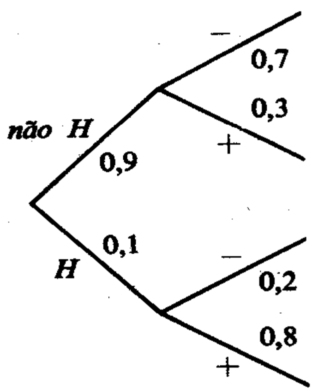


Fig. 1-5

EXEMPLO 1.7 Apenas uma em cada dez pessoas de uma população tem hilose. Das pessoas que têm hilose, 80% reagem positivamente ao teste *X*, enquanto que apenas 30% dos que não têm hilose reagem positivamente. Uma pessoa da população é selecionada ao acaso e o teste *X* é aplicado. Qual a probabilidade de que essa pessoa tenha hilose, se reagiu positivamente ao teste? E negativamente? O diagrama da fig. 1-6 mostra que

$$P(H|+) = \frac{P(H \text{ e } +)}{P(+)} = \frac{0{,}08}{0{,}08 + 0{,}27} = \frac{8}{35} = 0{,}23$$

$$P(H|-) = \frac{P(H \text{ e } -)}{P(-)} = \frac{0{,}02}{0{,}02 + 0{,}63} = \frac{2}{65} = 0{,}031$$



Resultado	*P*
(*não H*, —)	(0,9) (0,7) = 0,63
(*não H*, +)	(0,9) (0,3) = 0,27
(*H*, —)	(0,1) (0,2) = 0,02
(*H*, +)	(0,1) (0,8) = 0,08

Fig. 1-6

EXEMPLO 1.8 Diariamente, uma em cada cinco pessoas é selecionada ao acaso. Qual a probabilidade de que, durante os três primeiros dias,

a) a mesma pessoa seja escolhida diariamente?
b) três pessoas diferentes sejam escolhidas?

Na fig. 1-7, os números dos vértices indicam o número de pessoas diferentes que foram escolhidas. Então,

P (*mesma pessoa todos os dias*) $= 0{,}04$
P (*três pessoas diferentes*) $= 0{,}48$

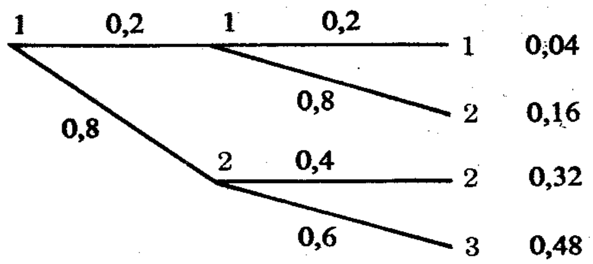


Fig. 1-7

EXERCÍCIOS

Lista 1

Construa diagramas, quando possível.

1. Uma das quatro palavras da frase THE URN IS COLD é selecionada ao acaso. Calcule a probabilidade de que o número de letras da palavra selecionada seja:
a) 2 *b*) 3 *c*) 4.

2. Uma das letras da frase THE URN IS COLD, é selecionada ao acaso. Calcule a probabilidade de que o número de letras da palavra que contém a letra selecionada seja:
a) 2 *b*) 3 *c*) 4.

3. Uma moeda equilibrada é lançada sucessivamente, até que apareça cara ou até que se obtenham 4 coroas. Representa-se coroa por T e cara por H[1]. Calcule a probabilidade de cada um dos cinco resultados: H, TH, TTH, $TTTH$, $TTTT$.

4. Um comitê é formado por quatro homens e duas mulheres. Dois membros são selecionados ao acaso, sucessivamente, sem reposição.
a) Calcule a probabilidade de cada um dos quatro resultados: HH, HM, MH, MM.
b) Qual o resultado mais provável?
c) Qual o número mais provável de homens?

5. Uma urna contém duas bolas com a marca A, uma com a marca B e uma com a marca C. As quatro bolas são retiradas ao acaso, sucessivamente, formando

[1] Representaremos, doravnte, cara por H e coroa por T, para maior simplicidade de notação (N. do T.)

uma *palavra* de quatro letras. Qual a probabilidade de que a *palavra* selecionada comece com:

a) *A* *b*) *B* *c*) *C* *d*) *AA* *e*) *AB* *f*) *ABA*?

6. Uma urna contém duas bolas brancas e duas pretas. As bolas são retiradas ao acaso, sucessivamente e sem reposição.

a) Qual a probabilidade de que a primeira bola preta apareça na primeira retirada? Na segunda retirada? Na terceira retirada? Na quarta retirada?

b) Qual a probabilidade de que a segunda bola preta apareça na segunda retirada? Na terceira retirada? Na quarta?

7. Como você selecionaria, aleatoriamente, uma casa entre cinco? Poderia usar um dado honesto? Uma moeda honesta?

8. No quarto *A*, encontram-se três pessoas e, no quarto *B*, duas. Um dos quartos é selecionado ao acaso; então, uma pessoa é selecionada ao acaso do quarto escolhido e esta pessoa ganha um prêmio. Qual a sua chance de ganhar o prêmio, se você está no quarto *A*? No quarto *B*?

9. Você entrega ao seu amigo uma carta para colocar no correio. A probabilidade de que ele esqueça de colocar é 0,1. A probabilidade de que o correio esqueça de enviá-la, dado que ela foi colocada, é de 0,1 e a probabilidade de que o destinatário não a receba, dado que foi enviada, é de 0,1. Qual a probabilidade de que seu amigo tenha esquecido de colocar a carta no correio, dado que o destinatário não a recebeu?

10. Três pessoas são selecionadas ao acaso, com reposição, de uma população de 100. Qual a probabilidade de que três pessoas diferentes sejam escolhidas?

11. A proporção de pessoas do sexo feminino em uma população é p. Cinco pessoas são selecionadas ao acaso, com reposição. Qual a probabilidade do resultado *MFFFF* (*M* representando as pessoas do sexo masculino e *F* as do feminino), se $p = 0{,}5$? Se $p = 0{,}2$?

12. O colégio 1 tem 50% de moças e o colégio 2 tem 20%. De um colégio selecionado ao acaso, são escolhidos aleatoriamente cinco alunos, com reposição. Qual a probabilidade de que o colégio 1 tenha sido selecionado, se a amostra foi *RMMMM*?

13. Uma parte tem probabilidade 0,9 de funcionar. Um componente consiste em três partes conectadas em série, de tal maneira que o componente funciona somente se as três partes funcionarem. Uma máquina é constituída de três componentes conectados em paralelo, de tal maneira que a máquina funciona se, pelo menos, um de seus componentes funciona. Qual a probabilidade de que a máquina funcione?

14. Cada vez que ocorrer cara no lançamento de uma moeda honesta, o jogador B paga Cr\$ 1,00 ao jogador A e, cada vez que ocorrer coroa, o jogador A paga Cr\$ 1,00 ao jogador B. Os jogadores começam com Cr\$ 2,00 cada e jogam até que um dêles perca todo o seu dinheiro.

a) Qual a probabilidade de que o jogo termine após o segundo lance? Após o quarto lance?

b) Qual a probabilidade de que o jogador A vença, dado que, no primeiro lançamento, ocorreu cara?

15. Um homem se coloca no centro de uma cidade e lança uma moeda honesta. Se ocorrer cara, ele caminha uma quadra para leste; se ocorrer coroa, caminha uma quadra para o norte. Então, êle lança novamente a moeda e usa a mesma regra. Considerando-se o centro da cidade como a origem (0,0), qual a probabilidade de que se encontre em (3,1) após quatro lances?

2 VARIÁVEIS

Qualquer regra que faz corresponder um número a cada resultado de um experimento é chamada *variável*. Ao número que a variável associa a um resultado particular, damos o nome de *valor* da variável para este resultado. Um rol dos valores da variável, com as probabilidades correspondentes, é chamado de *distribuição* da variável.

EXEMPLO 2.1 Dois estudantes são selecionados ao acaso (com reposição) em uma escola na qual 60% dos alunos são do sexo masculino. O número X de alunos do sexo masculino na amostra é uma variável.

Do diagrama apresentado na fig. 2-1, podemos obter a seguinte distribuição de X:

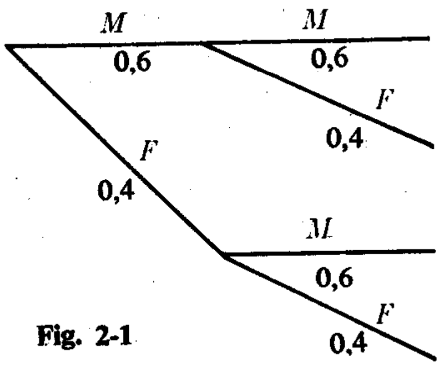


Fig. 2-1

amostra	p	X
MM	0,36	2
MF	0,24	1
FM	0,24	1
FF	0,16	0

v	p
0	0,16
1	0,48
2	0,36

A distribuição de uma variável pode ser representada por um gráfico, chamado *histograma*. A fig. 2-2 mostra o histograma da variável X. O histograma consiste em retângulos disjuntos com bases iguais, centradas em cada

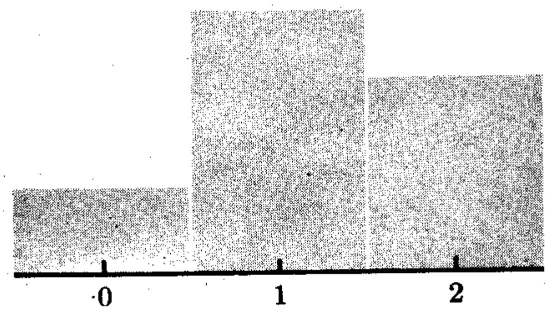


Fig. 2-2

valor da variável e com alturas proporcionais às probabilidades de que a variável assuma aqueles valores. As alturas de nossos retângulos são proporcionais a 16, 48 e 36.

Para quaisquer duas variáveis X e Y, denotamos por $X + Y$ a variável que associa a cada resultado do experimento a soma dos valores de X e Y para o mesmo resultado. As variáveis $X - Y$, XY, X/Y, X^2, $3X + 2Y - 7$ e outras são definidas de maneira análoga.

EXEMPLO 2.2 Uma das cinco palavras da seguinte frase é selecionada ao acaso:

THERE IS NO LARGEST INTEGER

Determine as distribuições de

V, o número de vogais na palavra selecionada
C, o número de consoantes
$V + C$
$C - V - 1$
$(V - 2)^2$

Elemento	V	C	$C + V$	$C - V - 1$	$(V - 2)^2$
THERE	2	3	5	0	0
IS	1	1	2	-1	1
NO	1	1	2	-1	1
LARGEST	2	5	7	2	0
INTEGER	3	4	7	0	1

V:

v	p
1	0,4
2	0,4
3	0,2

C:

v	p
1	0,4
3	0,2
4	0,2
5	0,2

$V + C$:

v	p
2	0,4
5	0,2
7	0,4

$C - V - 1$:

v	p
-1	0,4
0	0,4
2	0,2

$(V - 2)^2$:

v	p
0	0,4
1	0,6

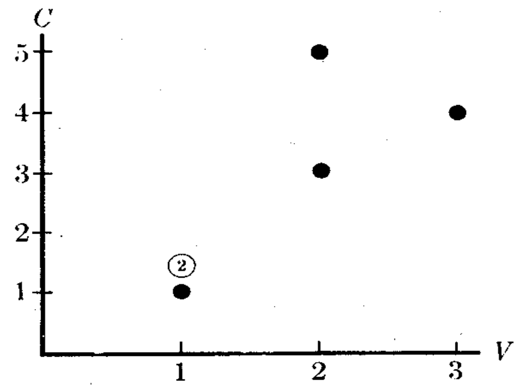


Fig. 2-3

Das distribuições de V e C, separadamente, nada se obtém acerca de seu comportamento conjunto, como, por exemplo, se existe certa tendência em associar grandes valores de V a grandes valores de C. Este tipo de informação é obtida através de um gráfico chamado *diagrama de dispersão*.

A fig. 2-3 apresenta o diagrama de dispersão para V e C. Para sua construção, marcamos todos os pares de valores das variáveis e rotulamos cada um desses pontos com um número proporcional à sua probabilidade; os pontos não rotulados devem ser entendidos como tendo peso 1. Os pontos vão-se tornando mais altos, quando caminhamos para a direita: valores grandes de V tendem a associar-se com valores grandes de C.

PROBLEMA PARA DISCUSSÃO 2.1

Para cada um dos diagramas de dispersão da fig. 2-4, determine as distribuições e construa os histogramas para X, Y, $X + Y$, $X - Y$ e calcule $P(X > Y)$, $P(Y = 2X + 1)$.

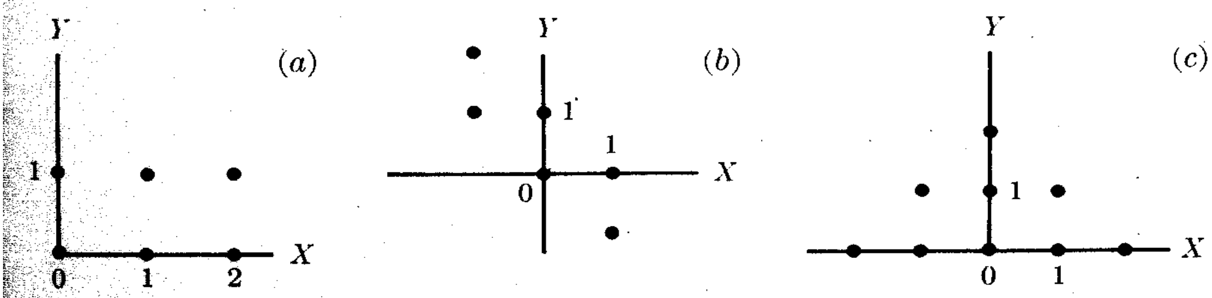


Fig. 2-4

EXERCÍCIOS
Lista 2

1. Uma urna contém três bolas, numeradas de 0 a 2. Duas delas são retiradas ao acaso, sem reposição. O número da primeira bola retirada é X e o número da segunda bola retirada é Y.
   a) Determine as distribuições e construa os histogramas para cada uma das seguintes variáveis: X, Y, $X + Y$, $X - Y$, $|X - Y|$.
   b) Construa o diagrama de dispersão para X e Y.
2. Repita o probl. 1, considerando as retiradas com reposição.
3. Bolas são retiradas sucessivamente com reposição de uma urna na qual a proporção de bolas pretas é 0,8. Você recebe Cr$ 1,00 cada vez que uma bola preta é retirada e não recebe nada, no caso contrário. Represente por X_i seu rendimento proveniente da *i*-ésima retirada e por $S_i = X_1 + \ldots + X_i$ o rendimento total nas primeiras i retiradas.
   a) Determine a distribuição e construa o histograma para cada uma das variáveis: X_1, S_2, S_3, $S_3 - S_2$, $X_1 - X_2$.
   b) Construa o diagrama de dispersão para as variáveis S_2 e S_3.
4. Construa um histograma para a variável S_3 do probl. 3, para cada uma das seguintes proporções de bolas pretas na urna: 0,6; 0,5; 0,1; 0 e 1.

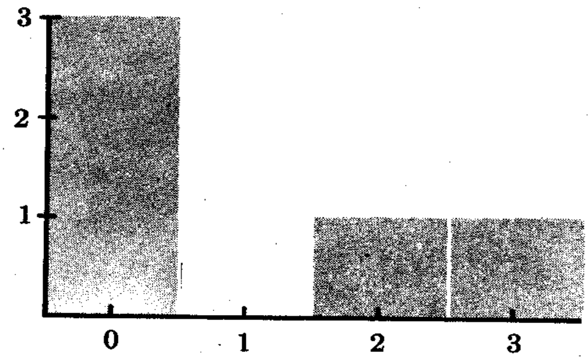


Fig. 2-5

5. A fig. 2-5 representa o histograma de uma variável X. Construa o histograma para as seguintes variáveis: $X + 2$, $2X$, $X - 1$, X^2, $(X - 1)^2$. $X/(X + 1)$.

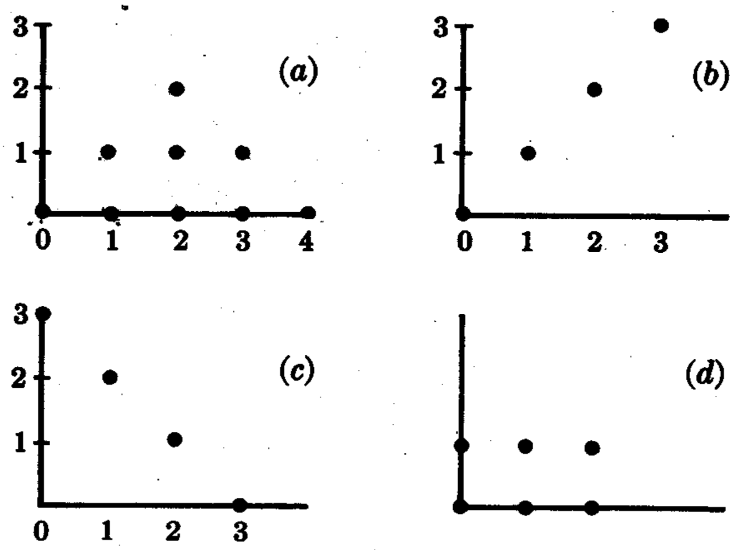


Fig. 2-6

6. Para cada um dos diagramas de dispersão da fig. 2-6, determine as distribuições de X, Y e $X + Y$; construa seus histogramas e faça um diagrama de dispersão para as variáveis X e $Y - X$.

7. Uma urna contém quatro bolas numeradas de 0 a 3. As bolas são retiradas ao acaso, sucessivamente, com reposição, e X_n representa o maior número observado nas primeiras n retiradas. Calcule, então, $P(X_2 = 0)$, $P(X_2 \leq \leq 1)$, $P(X_2 \leq 2)$. Construa um histograma para X_2 e outro para X_3.

8. Uma urna contém cinco bolas numeradas, das quais três com número 0, uma com o número 1 e uma com o número 2. As bolas são retiradas ao acaso, sucessivamente, sem reposição.

   a) Determine os possíveis resultados das duas primeiras retiradas e as respectivas probabilidades.

   b) Determine os possíveis resultados das três primeiras retiradas e as respectivas probabilidades.

   c) Determine a distribuição de X, soma dos números das duas primeiras bolas retiradas, com base no item (*a*) e com base no item (*b*).

d) Determine a distribuição de $(X - 1)^2$, com base no item (*a*), com base no item (*b*) e diretamente da distribuição de X.

9. Um ponto é selecionado ao acaso no quadrado, cujos vértices são (0,0), (1,0), (1,1) e (0,1). Represente por X e Y as coordenadas x, y do ponto selecionado.

   a) Calcule $P(X + Y \leq t)$ para cada um dos seguintes valores de t: 0; 0,5; 1; 1,5; 2.

   b) Faça um gráfico para $P(X + Y \leq t)$ como função de t, $0 \leq t \leq 2$.

10. Um comitê é formado por três pessoas. Cada ano uma delas, selecionada ao acaso, muda de opinião a respeito da pena de morte. De início, todas estão a favor. Se X_n representa o número de pessoas que são a favor da pena de morte no n-ésimo ano (assim, $X_1 = 3$) determine as distribuições e construa os histogramas para X_2, X_3 e X_4. Construa o diagrama de dispersão para X_3 e X_4.

11. Um relojoeiro tem vários relógios para consertar; 40% deles irão requerer uma hora de trabalho, 20% duas horas e 40% três horas. Ele seleciona dois relógios aleatoriamente e conserta-os. Determine a distribuição do tempo X, que ele leva para consertar os dois relógios selecionados.

3 DENSIDADES

Suponha que, numa certa população, a distribuição da variável *H*, altura em polegadas, com arredondamento para o valor inteiro mais próximo, é a seguinte:

H	*Proporção por valor de H*
63	0,01
64	0,03
65	0,07
66	0,12
67	0,17
68	0,20
69	0,17
70	0,12
71	0,07
72	0,03
73	0,01

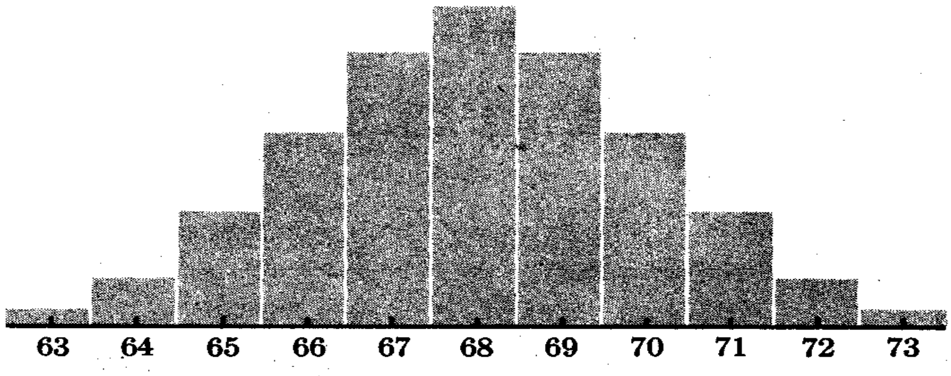


Fig. 3-1

A fig. 3-1 mostra um histograma para H. Se tivéssemos arredondado as alturas até centésimos, teríamos 1.100 valores para H e o histograma seria composto de 1.100 retângulos (mais finos, se mantivéssemos a mesma escala para H) e, se retângulos adjacentes tivessem aproximadamente, a mesma altura, o histograma seria bastante parecido com a curva apresentada na fig. 3-2.

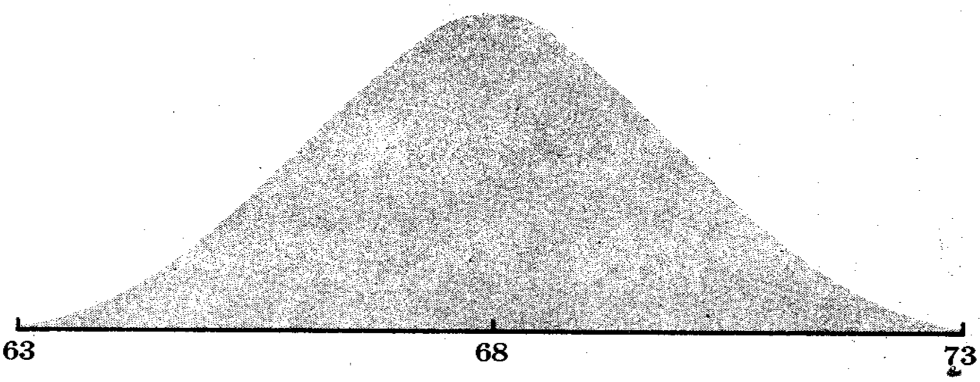


Fig. 3-2

Qualquer curva que represente o gráfico de uma função não-negativa, delimitando entre si e o eixo dos x uma área positiva e finita, pode ser considerada como uma aproximação do histograma de uma variável X, que assume um grande número de valores. A esta curva damos o nome de *densidade*. Para quaisquer dois números a e b, a probabilidade de que o valor de X esteja compreendido entre a e b é proporcional à área sob a densidade entre a e b:

$$P(a \leq X \leq b) = \frac{\text{área sob a densidade, entre } a \text{ e } b}{\text{área total sob a densidade}}$$

EXEMPLO 3.1 A figura 3-3 apresenta a densidade para uma variável X. Calcule $P(11 \leq X \leq 13)$.

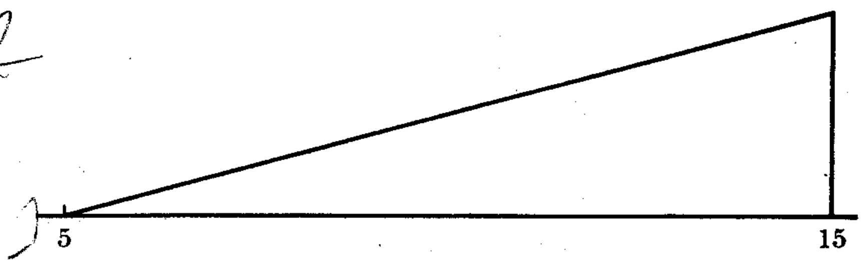


Fig. 3-3

Na fig. 3-4, $P(11 \leq X \leq 13)$ é a razão entre a área sombreada e a área do triângulo ABC. Digamos que a altura do triângulo, no ponto 15, seja 1. Então, a área do trapézio sombreada é $\frac{0,6 + 0,8}{2}$ (2) $= 1,4$ e a área do triângulo é $\frac{1}{2}(10)1 = 5$; logo,

$$P(11 \leq X \leq 13) = \frac{1,4}{5} = 0,28$$

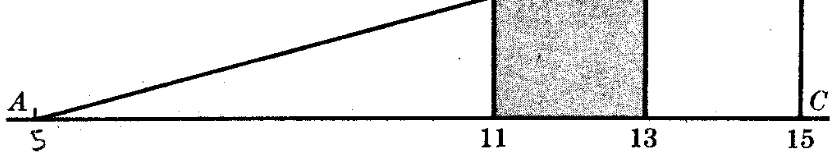


Fig. 3-4

A variável X pode assumir qualquer valor entre 5 e 15. Se decidirmos medir X aproximadamente, por exemplo, arredondando para o número par mais próximo, a variável aproximante X^* assumirá apenas os valores 6, 8, 10, 12 e 14:

Se X está entre 5 e 7, $X^* = 6$
Se X está entre 7 e 9, $X^* = 8$
....
Se X está entre 13 e 15, $X^* = 14$

Dizemos que X^* é a *aproximação* de X com valores 6, 8, 10, 12 e 14.

A distribuição de X^* é a seguinte:

v	p
6	0,04
8	0,12
10	0,20
12	0,28
14	0,36

A fig. 3-5 apresenta o histograma para X^*, juntamente com a densidade de X.

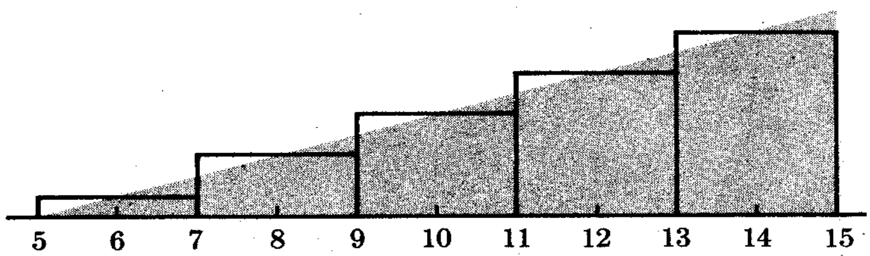


Fig. 3-5

EXEMPLO 3.2 A fig. 3-6 apresenta a densidade de uma variável X. Qual a distribuição da variável aproximante X^* que assume os valores 1, 3, 5, 7 e 9?

Devemos calcular as áreas das cinco regiões separadas pelas linhas verticais. De acordo com a escala vertical adotada (poderíamos ter adotado a escala 1, 2, 3 ou 2, 4, 6, em lugar de 10, 20, 30), as alturas das linhas verticais são

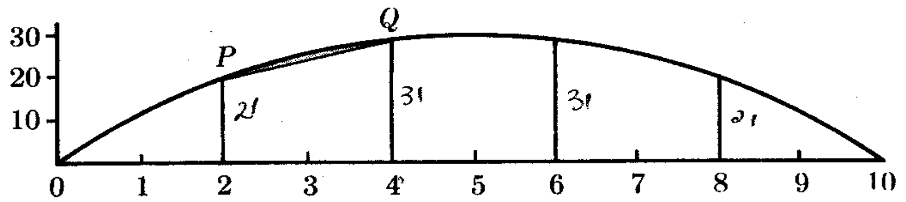


Fig. 3-6

21, 31, 31 e 21. Para determinarmos a área entre 2 e 4, ignoramos a pequena área sombreada entre a linha PQ e a densidade e calculamos a área do trapézio sob PQ, que é $\frac{21+31}{2}(2) = 52$. As outras quatro áreas são calculadas de maneira análoga:

Área sob a densidade entre 0 e 2 =	21
Área sob a densidade entre 2 e 4 =	52
Área sob a densidade entre 4 e 6 =	62
Área sob a densidade entre 6 e 8 =	52
Área sob a densidade entre 8 e 10 =	21
Área total sob a densidade =	208

Assim,

$$P(0 \leq X \leq 2) = P(X^* = 1) = 21/208 = 0{,}10$$
$$P(2 \leq X \leq 4) = P(X^* = 3) = 52/208 = 0{,}25$$
$$P(4 \leq X \leq 6) = P(X^* = 5) = 62/208 = 0{,}30$$
$$P(6 \leq X \leq 8) = P(X^* = 7) = 52/208 = 0{,}25$$
$$P(8 \leq X \leq 10) = P(X^* = 9) = 21/208 = 0{,}10$$

Podemos sempre escolher a escala vertical, de forma que a área total sob a densidade seja 1. Neste caso, a área associada a cada conjunto de valores da variável é a probabilidade de que a variável pertença àquele conjunto.

Por exemplo, se tomarmos a altura CB igual a 0,2 na fig. 3-4, a área do triângulo ABC será 1 e a área sombreada será $\frac{0{,}12+0{,}16}{2}(2) = 0{,}28$, que coincide com o valor de $P(11 \leq X \leq 13)$, calculado anteriormente.

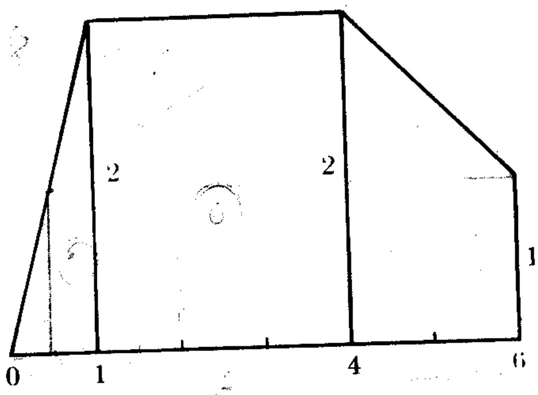


Fig. 3-7

EXERCÍCIOS

Lista 3

1. A fig. 3-7 apresenta a densidade par uma variável X. Verifique que a área total sob a densidade é 10.

Calcule

a) $P(0 \leq X \leq 1)$

b) $P(1 \leq X \leq 4)$

c) $P(4 \leq X \leq 6)$

Qual a distribuição da variável aproximante X^*, cujos valores são 1, 3 e 5? Faça o gráfico de $P(X \leq t)$ como função de t para $0 \leq t \leq 2$. Faça o gráfico de $P(X^* \leq t)$ para $0 \leq t \leq 6$.

2. Construa o gráfico de cada uma das seguintes funções e considere como a densidade de uma variável X:

a) $f(t) = \begin{cases} 1 & 0 \leq t \leq 8 \text{ densidade uniforme ou retangular sobre } 0 \leq t \leq 8 \\ 0 & \text{no complementar} \end{cases}$

b) $f(t) = \begin{cases} t & 0 \leq t \leq 8 \\ 0 & \text{no complementar} \end{cases}$

c) $f(t) = \begin{cases} 8 - t & 0 \leq t \leq 8 \\ 0 & \text{no complementar} \end{cases}$

d) $f(t) = \begin{cases} t(8 - t) & 0 \leq t \leq 8 \\ 0 & \text{no complementar} \end{cases}$

Determine a distribuição da aproximante X^* com valores 1, 3, 5 e 7.

3. Construa um semicírculo com centro no ponto (4,0), raio 4 e diâmetro no eixo dos x. Considere-o como uma densidade e encontre, aproximadamente, a distribuição da variável aproximante cujos valores são 1, 3, 5 e 7.

4. Para uma determinada variável X

$$P(X \leq t) = \begin{cases} 0 & t \leq 0 \\ t/6 & 0 \leq t \leq 6 \\ 1 & t \geq 6 \end{cases}$$

a) Faça o gráfico de $P(X \leq t)$ como função de t.
b) Calcule $P(X \leq 2)$, $P(X \leq 4)$ e $P(2 \leq X \leq 4)$
c) Qual a distribuição de X^* com valores 1, 3 e 5?

Construa um histograma para X^*. Qual a densidade para X?

5. Repita o probl. 4 com

$$P(X \leq t) = \frac{t^2}{36} \quad \text{para} \quad 0 \leq t \leq 6.$$

4 MÉDIA

A *média* ou *valor esperado* de uma variável X é o número, denotado por $E(X)$, obtido de uma das seguintes maneiras:

a) Multiplicando-se o valor que X associa a cada resultado pela probabilidade deste resultado e somando-se tais produtos.

b) Multiplicando-se cada valor da variável X pela probabilidade de sua ocorrência e somando-se os produtos.

a) $E(X) = \Sigma\, X(e)P(e)$
onde e representa os possíveis resultados

b) $E(X) = \Sigma\, vP(X = v)$,
onde v representa os valores da variável X.

EXEMPLO 4.1 As duas fórmulas (*a*) e (*b*) são equivalentes, uma vez que, no lado direito da fórmula (*a*), a soma de todos os termos com $X(e) = v$ é simplesmente $vP(X = v)$.

Duas bolas são selecionadas ao acaso, sem reposição, de uma urna contendo duas bolas pretas e três brancas. Calcule $E(X)$ e $E(X^2)$, onde X é o número de bolas pretas selecionadas. Na fig. 4-1, apresentamos um diagrama.

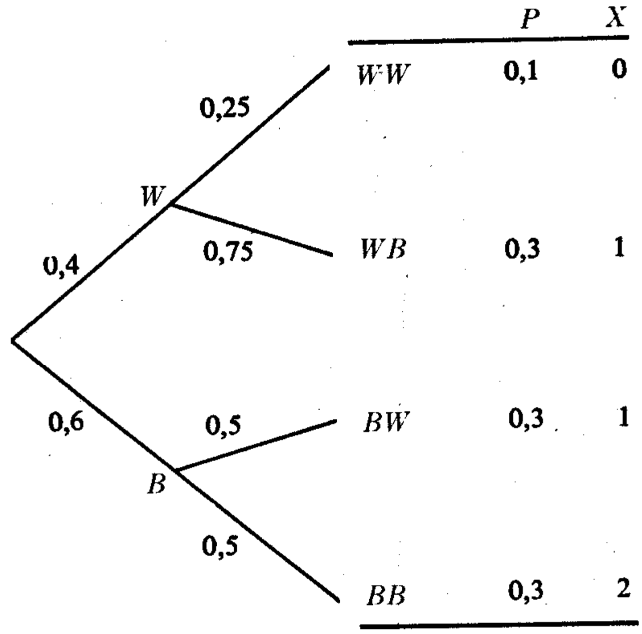


Fig. 4-1

(*Representemos branca por B e preta por W.*)

Usando a fórmula (*a*):

$E(X) \; = 2.(0,3) + 1.(0,3) + 1.(0,3) + 0.(0,1) = 1,2$

$E(X^2) = 4.(0,3) + 1.(0,3) + 1.(0,3) + 0.(0,1) = 1,8$

Para usarmos a fórmula (*b*), determinam-se as distribuições de X e X^2.

X

v	p
0	0,1
1	0,6
2	0,3

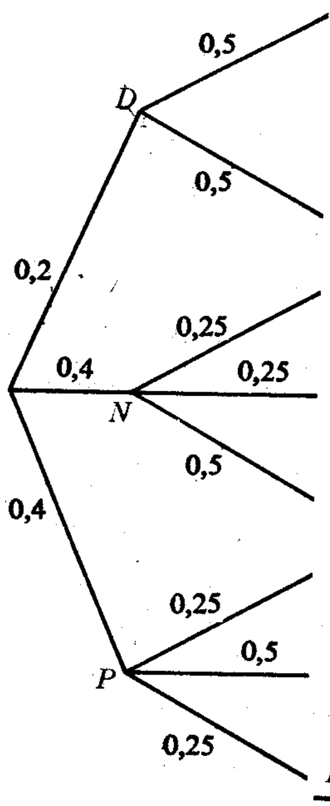
$E(X) = 0(0,1) + 1(0,6) + 2(0,3) = 1,2$

X^2

v	p
0	0,1
1	0,6
4	0,3

$$E(X^2) = 0(0{,}1) + 1(0{,}6) + 4(0{,}3) = 1{,}8$$

EXEMPLO 4.2 Uma urna contém duas moedas de 1 centavo (P), duas de 5 centavos (N) e uma de 10 centavos (D). Duas moedas são selecionadas ao acaso, sem reposição, e dadas a você. Qual a sua renda esperada na primeira retirada (F)? Na segundo retirada (S)? Qual a sua renda total esperada? Veja o diagrama da fig. 4-2.



	P	F	S	$F+S$
DN	0,1	10	5	15
DP	0,1	10	1	11
ND	0,1	5	10	15
NN	0,1	5	5	10
NP	0,2	5	1	6
PD	0,1	1	10	11
PN	0,2	1	5	6
PP	0,1	1	1	2

Fig. 4-2

Pela fórmula (b),

$$E(F) = 1(0,4) + 5(0,4) + 10(0,2) = 4,4$$

$$E(S) = 1(0,4) + 5(0,4) + 10(0,2) = 4,4$$

$$E(F + S) = 2(0,1) + 6(0,4) + 10(0,1) + 11(0,2) + 15(0,2) = \\ = 8,8$$

PROBLEMA PARA DISCUSSÃO 4.1

Suponha que a seleção, no exemplo 4.2, é feita com reposição e você recebe a soma dos valores das moedas selecionadas. A distribuição da sua renda é igual à anterior? Sua renda esperada é igual à anterior?

PROPRIEDADES DA MÉDIA

1. A média de uma soma de variáveis é igual à soma das médias das variáveis:

   $E(X + Y + ...) = E(X) + E(Y) + ...$

2. A média de uma variável multiplicada por uma constante é igual à constante multiplicada pela média da variável:

   $E(cX) = cE(X)$

3. A média de uma constante é igual à constante

   $E(c) = c$

As propriedades acima são conseqüências imediatas da fórmula (a) para a média.

EXEMPLO 4.3

Uma moeda equilibrada será lançada repetidamente. Antes de cada lançamento, você pode apostar qualquer quantia em dinheiro no resultado cara. Apresentamos, a seguir, três maneiras de apostar, usando apenas os três primeiros lançamentos:

1. Aposte Cr$ 1,00 no primeiro lançamento. Se ganhar, abandone o jogo. Se perder, aposte Cr$ 2,00 no segundo lançamente. Se ganhar, abandone o jogo. Se perder novamente, aposte Cr$ 4,00 no terceiro lançamento.
2. Aposte Cr$ 1,00 no primeiro laçamento. Se ganhar, abandone o jogo. Se perder, aposte Cr$ 1,00 em cada um dos lançamentos seguintes.
3. Aposte Cr$ 1,00 em cada lançamento.

Calcule a média de sua renda líquida, segundo cada um dos sistemas de aposta.

Resultado	P	W_1	W_2	W_3
HHH	$\frac{1}{8}$	1	1	3
HHT	$\frac{1}{8}$	1	1	1
HTH	$\frac{1}{8}$	1	1	1
HTT	$\frac{1}{8}$	1	1	-1
THH	$\frac{1}{8}$	1	1	1
THT	$\frac{1}{8}$	1	-1	-1
TTH	$\frac{1}{8}$	1	-1	-1
TTT	$\frac{1}{8}$	-7	-3	-3

Pela fórmula (a):

$$E(W_1) = \frac{1+1+1+1+1+1+1-7}{8} = \frac{0}{8} = 0$$

$$E(W_2) = 0$$

$$E(W_3) = 0$$

Sua renda líquida média será, portanto, zero, segundo qualquer sistema de aposta.

PROBLEMA PARA DISCUTIR 4.2

Imagine um outro sistema que se adapte às circunstâncias descritas no exemplo 4.3 e calcule a renda média.

PROBLEMA PARA DISCUTIR 4.3

No sistema de apostas 3, do exemplo 4.3, qual a distribuição de sua renda líquida no primeiro lançamento da moeda? No segundo? E no terceiro? Qual a sua renda líquida esperada em cada lançamento? Qual sua renda total esperada? Qual seria sua renda líquida esperada, se o sistema 3 fosse utilizado em 10 lançamentos da moeda?

EXEMPLO 4.4

Considere a seguinte densidade de uma variável X:

$$p(t) = \begin{cases} t & 0 < t < 6 \\ 0 & \text{caso contrário} \end{cases}$$

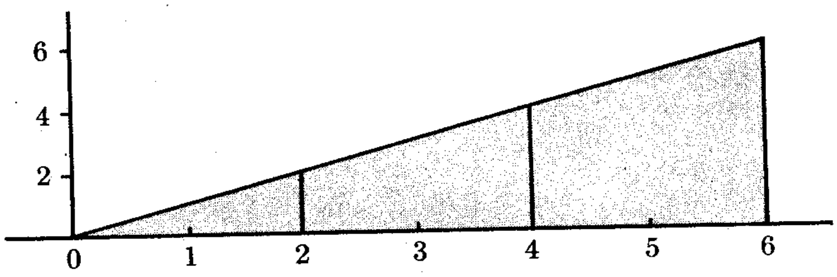


Fig. 4-3

Calcule $E(X)$ e $E(X^2)$, aproximadamente. Veja a fig. 4-3. Para isto, usaremos a variável X^*, cujos valores são 1,3 e 5. Sua distribuição é a seguinte:

v	p
1	1/9
3	3/9
5	5/9

$$E(X^*) = 1.(1/9) + 3.(3/9) + 5.(5/9) = 3{,}89$$
$$E(X^{*2}) = 1.(1/9) + 9.(3/9) + 25.(5/9) = 17$$

Então, $E(X) = 3{,}89$ e $E(X) = 3{,}89$ e $E(X^2) = 17$, aproximadamente. Os valores exatos são $E(X) = 4$ e $E(X^2) = 18$.

Uma aproximante X^* com mais valores dar-nos-ía melhor aproximação, porém com mais trabalho. Por exemplo, se X^* assumisse os valores 0,5, 1,5,...,5,5 teríamos $E(X^*) = 3{,}97$ e $E(X^{*2}) = 17{,}86$.

EXERCÍCIOS

Lista 4

1. Uma urna contém 6 bolas numeradas, das quais três com o número 0, duas com o 2 e uma com o 4 Duas bolas são por você retiradas da urna, sucessivamente, ao acaso e pagam-lhe em cruzeiros a média dos números nelas inscritos. Determine a distribuição e a média da sua renda e construa o histograma para esta (*a*) considerando retiradas com reposição; (*b*) sem reposição.

   Suponha que, após verificar a primeira bola retirada, você possa decidir, dependendo do número nela escrito, se repõe ou não a bola na urna, antes da segunda retirada. Qual seria o seu melhor plano? Qual seria a sua renda esperada, se você usasse este plano?

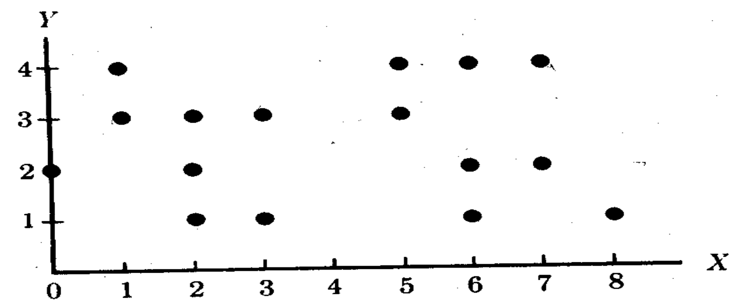


Fig. 4-4

2. Outra propriedade da média é a seguinte: A média de uma variável é o ponto onde seu histograma ou densidade se equilibra. Use este fato para calcular $E(X)$ e $E(Y)$ no diagrama de dispersão da fig. 4-4. Qual o valor de $E(X + Y)$? (verifique, através de cálculos, os resultados).

3. Uma variável X tem a seguinte distribuição:

v	p
0	0,4
1	0,3
2	0,2
3	0,1

Calcule a média da variável $(X - t)^2$ para $t = 0, 1, 3$ e faça o gráfico da média de $(X - t)^2$ como função de t para $0 < t < 3$. Que valor de t torna mínimo $E[(X - t)^2]$? Qual o valor de $E(X)$?

4. Um dos inteiros 1, 2, ..., 1.000 é selecionado ao acaso. Qual a média do inteiro X selecionado (onde estaria o ponto de equilíbrio do histograma)? Como $E(X) = (1 + 2 + ... + 1.000)/1.000$, qual o valor da soma dos primeiros 1.000 inteiros?

5. A distribuição de uma variável X é a seguinte:

v	p
0	$1 - t$
1	t

Calcule: *a*) $E(X)$, *b*) $E(X^2)$, *c*) $[E(X)]^2$, *d*) $E(X^2) - [E(X)]^2$ e faça o gráfico de cada uma delas como função de t, para $0 < t < 1$.

Se uma variável assume apenas os valores 0 e 1, quais os possíveis valores de $E(X)$? E de $E(X^2) - [E(Z)]^2$?

6. Se uma variável X assume apenas os valores 0, 1, 2, é possível atribuir probabilidades a estes números, de tal forma que $E(X) = 1$? De tal forma que $E(X^2) = 5/3$? De tal forma que $E(X) = 1$ e $E(X^2) = 5/3$? Se $E(X) = 1$, qual o maior valor possível de $E(X^2)$? E qual o menor?

7. Uma moeda equilibrada é lançada até que ocorra coroa ou até que se tenham feito três lançamentos de tal forma que os quatro resultados possíveis são: *T*, *HT*, *HHT*, *HHH*. Qual é o número médio de lançamentos? Qual o número médio de lançamentos, se, em vez de, no máximo, 3 lançamentos, tivéssemos 2? Se tivessémos 4? Se tivéssemos 5? Se tivéssemos 1.000 (adivinhe)?

8. Uma urna contém uma bola preta e uma branca. As bolas são retiradas ao acaso, sucessivamente, com reposição, até que a mesma cor tenha ocorrido duas vezes. Desta maneira, os resultados possíveis são: *PP*, *PBP*, *PBB*, *BPP*, *BPB*, *BB*. Calcule o número médio de retiradas.

9. Para cada uma das densidades abaixo, calcule $E(X)$ e $E(X^2)$ aproximadamente:

$$a)\ f(t) = \begin{cases} 1 & 0 \leq t \leq 10 \\ 0 & \text{no complementar} \end{cases}$$

$$b)\ f(t) = \begin{cases} 12 - t & 0 \leq t \leq 12 \\ 0 & \text{no complementar} \end{cases}$$

10. Para uma variável X, a fig. 4-5 mostra o gráfico de $P(X \leq t)$. Calcule $E(X)$ e $E(X^2)$, aproximadamente.

11. Uma moeda, com probabilidade p de ocorrer cara, será lançada duas vezes. Em cada lançamento, você pode

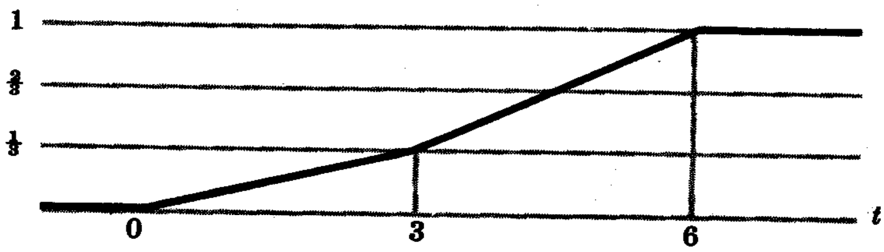


Fig. 4-5

fazer uma aposta de qualquer quantia no resultado cara. Você usa o seguinte esquema: aposta Cr\$ 2,00 no primeiro lançamento. Se ganha, aposta Cr\$ 1,00 no segundo laçamento, mas, se perde, aposta Cr\$ 4,00 no segundo lançamento:

a) Determine a distribuição e calcule a média de seus ganhos líquidos (W) para $p = 0,4$; 0, 5 e 0,6

b) Represente o esquema acima por (2; 1,4). Calcule a média de W para $p = 0,4$ e 0,6 para cada um dos seguintes esquemas: (1; 2,3), (1; 0,2), (1; 0,1).

c) Escolha um esquema qualquer e calcule $E(W)$ para $p = 0,4$ e 0,6. Formule uma conclusão geral sobre o sinal (+) de $E(W)$.

12. Em uma certa população, 60% das pessoas que votam são democratas. Você seleciona cinco eleitores, ao acaso, da população e dá a cada democrata Cr\$ 1,00. Que quantia você espera pagar ao primeiro eleitor sorteado? Ao quarto? Qual o pagamento total esperado? Considere, no problema, o esquema com reposição e o esquema sem reposição.

13. Você tem cinco fregueses, dois em A, dois em B e um em C. Você deve estabelecer-se em qualquer lugar no segmento de reta AC da fig. 4-6. Todos os dias, um dos fregueses é selecionado casualmente e você deve visitá-lo. Onde você deve estabelecer-se para minimizar a distância média percorrida? Suponha que o custo da viagem é o quadrado da distância viajada; então, por exemplo, se você está em 3, uma visita a C custa $(5 + 5)^2 = 100$ cruzeiros. Onde você deve estabelecer-se para minimizar os gastos esperados?

Fig. 4-6

14. Um ponto é sorteado casualmente, dentro de um círculo de raio 1. Qual é a probabilidade da distância X, do ponto sorteado até o centro, ser, pelo menos, 0,5? Calcule $E(X)$, aproximadamente.

15. Uma urna contém três bolas, marcadas com 0, 1 e 2. Duas são sorteadas sucessivamente, ao acaso. Calcule $E(XY)$, onde X e Y são o primeiro e o segundo números sorteados (*a*) com reposição ou (*b*) sem reposição.

16. No exercício 15, encontre o valor esperado para $Z = \max(X, Y)$, isto é, para o maior dos dois números sorteados (se X é igual a Y, então Z é igual a este valor comum).

17. Toda vez que ao jogar uma moeda sai cara, sua fortuna é dobrada; toda vez que sai coroa, sua fortuna é reduzida à metade. Você começa com Cr$ 1,00. Qual é a sua fortuna esperada, após dois lances, (*a*) se a moeda é honesta, e (*b*) se a probabilidade de cara é 1/3?

5 VARIÂNCIA

Suponha que você trabalha com previsões e uma das três palavras da frase GO TO LUNCH será selecionada ao acaso. Você deve prever o número de letras da palavra selecionada e o quadrado do seu erro será descontado do seu salário. Por exemplo, se você prevê 3 e a palavra selecionada é LUNCH, você perde $(5 - 3)^2 = 4$ cruzeiros, ao passo que você perde $(2 - 3)^2 = 1$ cruzeiro, se a palavra selecionada é GO ou TO. Que número você deve prever para tornar a perda esperada tão pequena quanto possível e qual é a menor perda esperada?

Se você escolher o número t como sua previsão, o cálculo da perda esperada é feito da seguinte forma:

1ª
AO
BANCO

Palavra	L (n.º de letras)	$L - t$ (erro)	$(L - t)^2$ (erro quadrático ou perda)
GO	2	$2 - t$	$(2 - t)^2$
TO	2	$2 - t$	$(2 - t)^2$
LUNCH	5	$5 - t$	$(5 - t)^2$

$$
\begin{aligned}
E(\text{perda}) = E(L - t)^2 &= \tfrac{1}{3}.(2 - t)^2 + \tfrac{1}{3}.(2 - t)^2 + \tfrac{1}{3}.(5 - t)^2 \\
&= \tfrac{1}{3}.(33 - 18t + 3t^2) \\
&= t^2 - 6t + 11 \\
&= (t - 3)^2 + 2
\end{aligned}
$$

Sua perda esperada (erro quadrático esperado ou erro médio quadrático, *emq*) será, pelo menos, igual a 2 para todo t e será exatamente 2, se $t = 3$.

Você pode prever 3 para L e o *emq* será 2.

Observe que

$$E(L) = \frac{2 + 2 + 5}{3} = 3$$

igual a sua melhor previsão.

Para qualquer variável X, a constante t que torna menor a média de $(X - t)^2$ é $t = E(X)$. Para esta escolha de t, a média de $(X - t)^2$, que é chamada *variância* de X e simbolizada por $\sigma^2(X)$, é dada pela fórmula:

$$\sigma^2(X) = E(X^2) - [E(X)]^2$$

EXEMPLO 5.2 Considerando a seguinte distribuição de peso (W) num população, calcule $E(W)$ e $\sigma(W)$.

v	p
140	0,5
150	0,3
160	0,2

Tomemos $X = (W - 150)/10$, cuja distribuição é a se guinte:

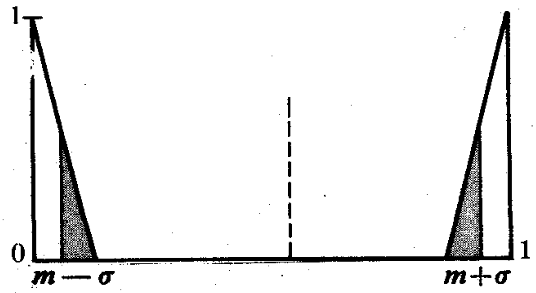
v	p
-1	0,5
0	0,3
1	0,2

$$\begin{aligned} E(X) &= -1(0{,}5) + 0(0{,}3) + 1(0{,}2) = -0{,}3 \\ E(X^2) &= 1(0{,}5) + 0(0{,}3) + 1(0{,}2) = 0{,}7 \\ \sigma^2(X) &= (0{,}7) - (-0{,}3)^2 = 0{,}61 \\ \sigma(X) &= \sqrt{0{,}61} = 0{,}78 \end{aligned}$$

Como

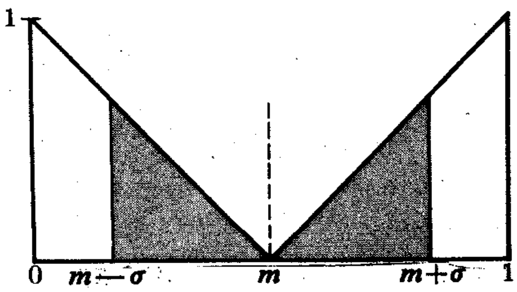
$$\begin{aligned} W &= 10X + 150 \\ E(W) &= 10(-0{,}3) + 150 = 147 \\ \sigma^2(W) &= 10^2(0{,}61) = 61 \\ \sigma(W) &= 10(0{,}78) = 7{,}8 \end{aligned}$$

O desvio padrão é a medida mais comum de dispers (variabilidade) da variável. Em muitos casos, a probabi dade da variável pertencer a um intervalo centrado na m dia e de raio igual a um desvio padrão é de 60% a 70 A fig. 5-1 apresenta 16 densidades com $0 \leq t \leq 1$. Pa cada uma delas, a região entre $m - \sigma$ e $m + \sigma$ está sor breada e a probabilidade p nesta região (isto é, a raz entre a área sombreada e a área total sob a densidade) indicada.



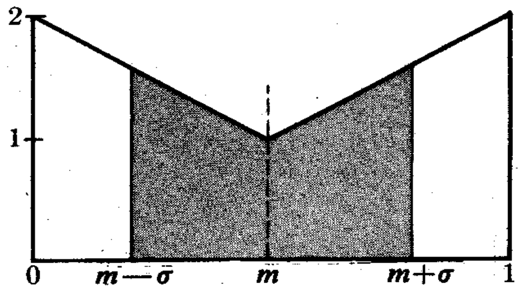
$f(t) = \max(0,|10t-5|-4)$

$\sigma = 0{,}467$ $p = 0{,}452$



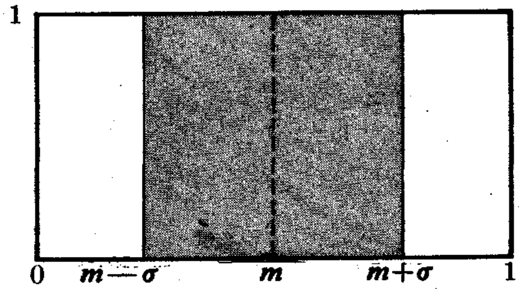
$f(t) = |2t-1|$ $\sigma = 0{,}354$

$m = 0{,}5$ $p = 0{,}500$



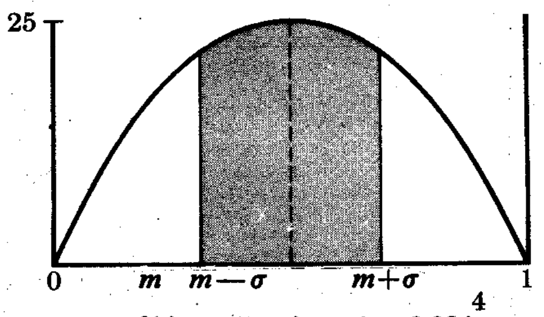
$f(t) = |2t-1|+1$ $\sigma = 0{,}312$

$m = 0{,}5$ $p = 0{,}545$



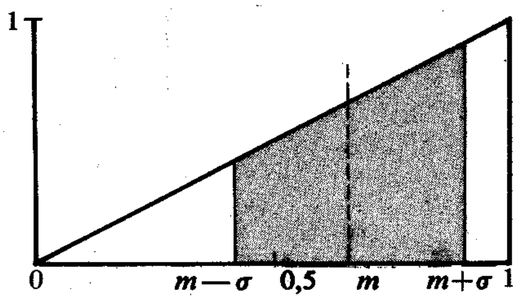
$f(t) = 1$ $\sigma = 0{,}289$

$m = 0{,}5$ $p = 0{,}577$



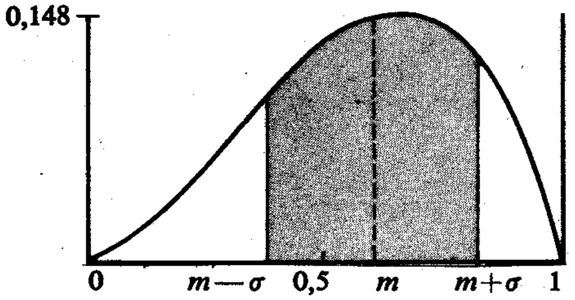
$f(t) = t(1-t)$ $\sigma = 0{,}2246$

$m = 0{,}5$ $p = 0{,}626$



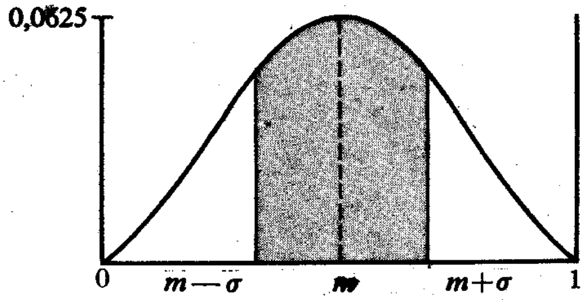
$f(t)\ t$ $\sigma = 0{,}236$

$m = 0{,}667$ $p = 0{,}629$



$f(t) = t^2(1-t)$ $\sigma = 0{,}2$

$m = 0{,}6$ $p = 0{,}640$



$f(t) = t^2(1-t)^2$ $\sigma = 0{,}189$

$m = 0{,}5$ $p = 0{,}644$

Fig. 5-1

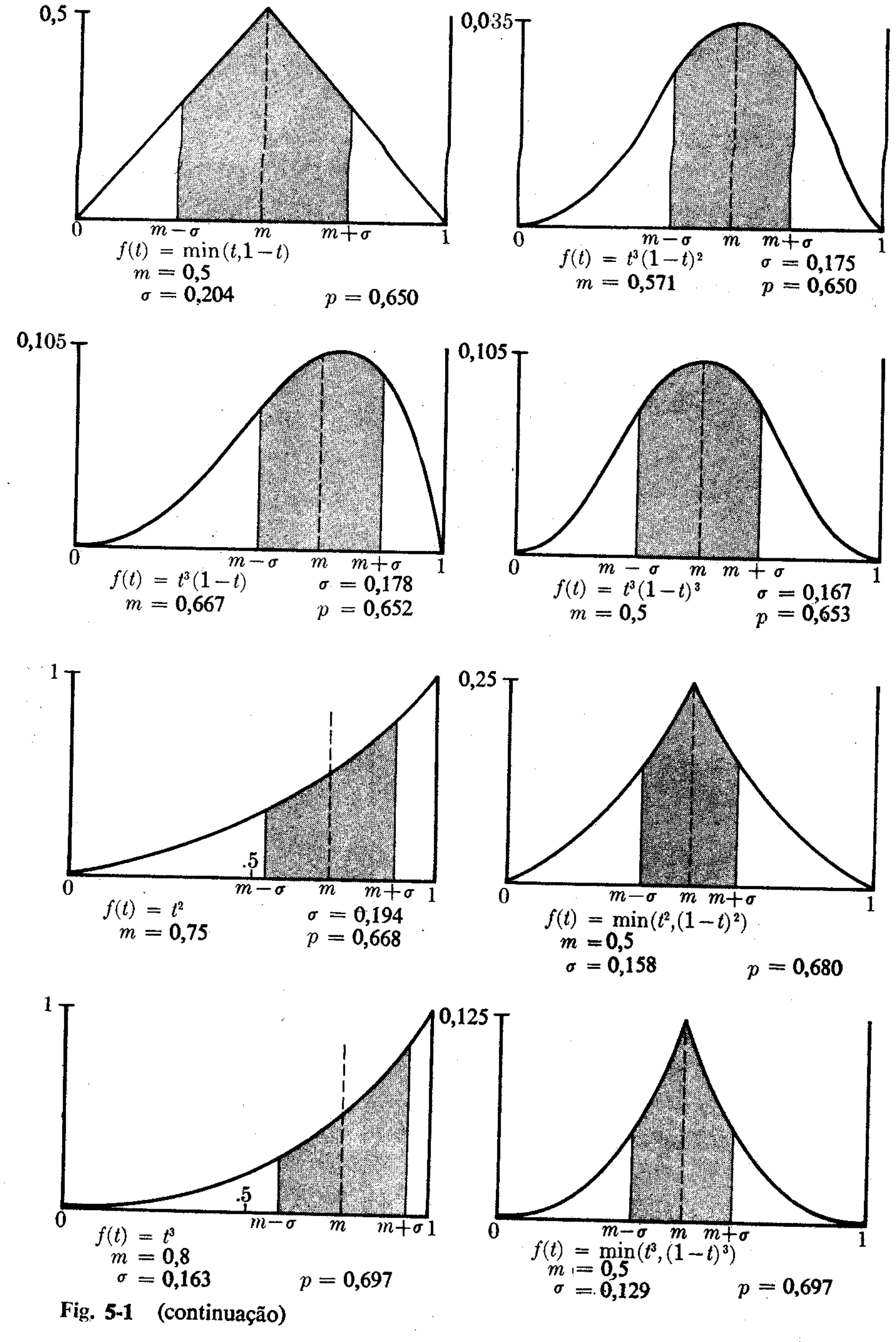


Fig. 5-1 (continuação)

EXERCÍCIOS

Lista 5

1. Em cada diagrama da fig. 5-2, estime a média e o desvio padrão de X e Y. Calcule, então, os valores exatos e marque, em cada diagrama de dispersão, o ponto $(E(X), E(Y))$. Para cada diagrama de dispersão, calcule $\sigma(X + Y)$.

2. Um dos dígitos 0, 1, ..., 9 é selecionado ao acaso. Calcule a média e o desvio padrão do *dígito selecionado*.

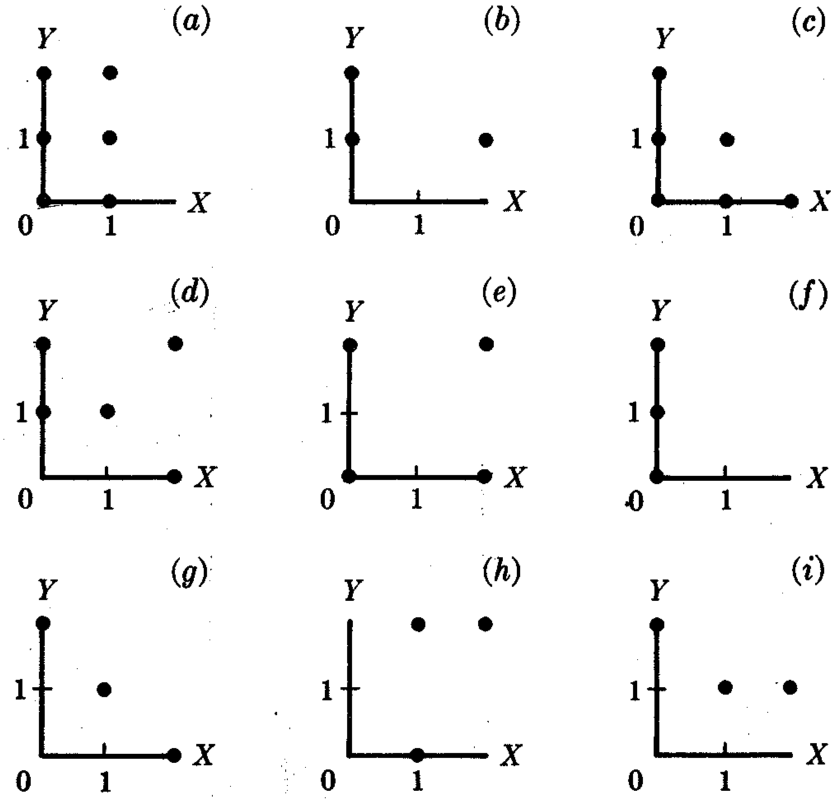


Fig. 5-2

3. Uma urna contém duas bolas brancas e duas pretas. Três bolas são retiradas ao acaso; calcule a média e o desvio padrão do número de bolas brancas retiradas *a*) com reposição; *b*) sem reposição.

4. Na sala 1, estão três mulheres, cujas alturas são 62, 65 e 65 polegadas. Na sala 2, estão dois homens, cujas alturas são 70 e 74. Uma das duas salas é selecionada ao acaso; em seguida, uma pessoa é selecionada ao acaso nesta sala. Você deve prever a altura da pessoa selecionada, perdendo o quadrado do seu erro. Qual a sua melhor previsão e qual o erro médio quadrático? Suponha que, antes de você fazer sua previsão, dizem-lhe qual foi a sala selecionada. Qual a sua melhor previsão se foi a sala 1 e qual o emq? Se foi a sala 2? Suponha que alguém se oferece, por um preço, para dizer a você qual foi a sala selecionada. Quanto você estará disposto a pagar?

5. Bolas são retiradas ao acaso, sucessivamente e com reposição, de uma urna, na qual 60% das bolas são pretas.

a) S_n indica o número de bolas pretas nas n primeiras retiradas. Calcule $E(S_n)$ e $\sigma^2(S_n)$ para $n = 1$, 2, 3 e 4. Tente encontrar uma fórmula geral para $E(S_n)$ e para $\sigma^2(S_n)$.

b) Seja Y_n a proporção de bolas pretas nas n primeiras retiradas, isto é, $Y_n = S_n/n$. Calcule a esperança e a variância de Y_n, para $n = 1$, 2, 3 e 4. Tente encontrar uma fórmula geral para $E(Y_n)$ e $\sigma^2(Y_n)$.

6. Refaça o exercício 5, substituindo 60% por 30%.

7. Substituindo 60% por p no exercício 5, tente encontrar fórmulas gerais para $E(S_n)$, $\sigma^2(S_n)$, $E(Y_n)$ e $\sigma^2(Y_n)$.

8. A seguir, estão algumas distribuições especificadas de forma incompleta. Em cada caso, tente encontrar um valor para t que torne o desvio padrão tão grande quanto possível. Construa os histogramas para os valores de t encontrados.

(*a*)

v	p
0	t
1	$1 - t$

(*b*)

v	p
1	t
6	$1 - t$

(*c*)

v	p
0	t
1	0,4
2	$0,6 - t$

(*d*)

v	p
0	t
1	$0,6 - t$
2	0,4

(*e*)

v	p
0	$\frac{1}{3}$
2	$\frac{1}{3}$
t	$\frac{1}{3}$

(t entre 0 e 2)

(*f*)

v	p
0	$\frac{1}{3}$
2	$\frac{1}{3}$
t	$\frac{1}{3}$

(t irrestrito)

(*g*)

v	p
t	0,5
$t + 1$	0,3
$t + 2$	0,2

(*h*)

v	p
t	0,5
$2t$	0,3
$3t$	0,2

(t entre 1 e 2)

9. Represente graficamente as densidades correspondentes às seguntes funções:

a) $f(t) = t/(t + 1)$ *b*) $f(t) = (2t - 1)^2$

onde $0 \leq t \leq 1$, estime m, σ e $p = P(m - \sigma \leq X \leq \leq m + \sigma)$ a partir do seu gráfico e faça cálculos aproximados para conferir suas estimativas.

10. Um ponto é selecionado ao acaso, num círculo de raio igual a 6 metros. Calcule, aproximadamente, o desvio padrão de X, a distância do ponto sortead ao centro, usando a aproximante X^* com valores 1, 3 e 5.

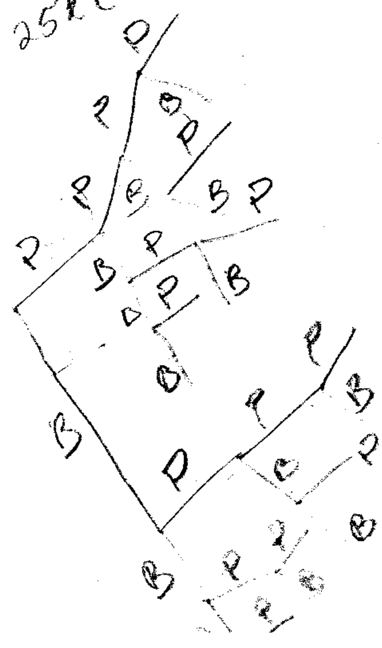
$36 + 60t + 25t^2$
$25t - 25t^2$
$25t(1-t)$

11. Uma moeda honesta é lançada até que ocorra coroa ou até que três lances tenham sido feitos. Calcule o desvio padrão do número N de lances.

P – 0,6
B – 0,4

PP – 0,36
PB – 0,24
BP – 0,24
BB – 0,16

PPP 0,216
PPB 0,144
PBP 0,144
PBB 0,096



E(Y) =

6 VALOR DE UM PREVISOR

Quando você queria prever o comprimento L de uma palavra selecionada ao acaso, da sentença GO TO LUNCH, perdendo o quadrado do seu erro, sua melhor previsão para L era $E(L) = 3$ e a sua perda esperada ou erro médio quadrático dessa melhor previsão era

$$\sigma^2(L) = E(L - 3)^2 = 2$$

Suponha que, antes de prever L, você conheça o número X de letras G da palavra selecionada. Sua melhor previsão para L, usando esta informação, é obviamente 2, quando $X = 1$, pois você sabe que a palavra selecionada foi GO. Quando $X = 0$, você sabe que uma das palavras TO ou LUNCH foi selecionada e que elas são igualmente prováveis, de tal forma que sua melhor estimativa é

$$E(L \mid X = 0) = (2)\tfrac{1}{2} + (5)\tfrac{1}{2} = 3{,}5$$

Sua previsão é, agora, uma variável cujo valor depende de X ou da palavra que foi selecionada. Vamos calcular o emq (erro médio quadrático):

Palavra	L	X	Y	$L - Y$ (erro)	$(L - Y)^2$ (erro quadrático)
Go	2	1	2	0	0
To	2	0	3,5	− 1,5	2,25
Lunch	5	0	3,5	1,5	2,25

emq para Y como previsor de $L = E(L - Y)^2 =$

$$= \frac{0 + 2{,}25 + 2{,}25}{3} = 1{,}5.$$

O uso de Y para prever L provoca uma economia esperada ou uma redução esperada do erro quadrático, de 0,50 em relação ao emq original igual a 2. Sua economia esperada é 25% do custo esperado original. Dizemos que Y tem *valor* 0,25 como previsor de L.

*

Para quaisquer duas variáveis Y e L:

a) O emq de Y como previsor de L é $E(L - Y)^2$

b) O *valor* de Y como previsor de L é

$$W(Y, L) = \frac{\sigma^2(L) - E(L - Y)^2}{\sigma^2(L)}$$

A variável que tem o maior valor para prever L é L, cujo valor é 1:

$$W(L,\ L) = \frac{\sigma^2(L) - E(L - L)^2}{\sigma^2(L)}$$

A média de L tem valor 0 como previsor de L:

$$W(E(L),\ L) = \frac{\sigma^2(L) - E(L - E(L))^2}{\sigma^2(L)} = \frac{\sigma^2(L) - \grave{o}^2(L)}{\sigma^2(L)} = 0$$

Alguns previsores têm valor negativo; por exemplo, qualquer constante diferente de $E(L)$ tem valor negativo, como previsor de L.

EXEMPLO 6.1 A fig. 6-1 representa o diagrama de dispersão para duas variáveis X e Y. Calcule o emq e o valor de $Z = (0{,}1)X + 1$ como previsor de Y.

A reta é o gráfico de Z como uma função de X. Os cálculos são feitos da seguinte maneira:

	X	Y	Z	$(Y - Z)^2$	Y^2
	0	0	1	1	0
	0	1	1	0	1
	1	2	1,1	0,81	4
	3	1	1,3	0,09	1
Σ	4	4	4,4	1,90	6
E	1	1	1,1	0,475	1,5

$$\grave{o}^2(Y) = 1{,}5 - 1^2 = 0{,}5$$

$$W(Z,\ Y) = \frac{0{,}5 - 0{,}475}{0{,}5} = 0{,}05$$

Então, o emq de Z é igual a 0,475 e o valor é 0,05 como previsor de Y.

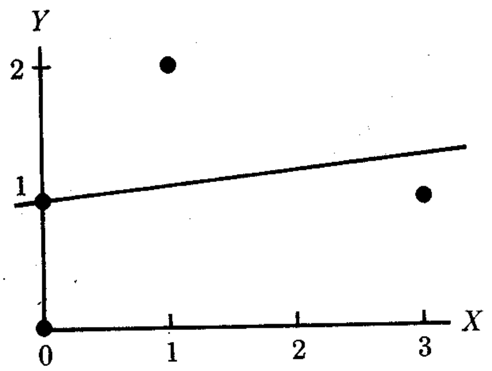


Fig. 6-1

PROBLEMA PARA DISCUSSÃO 6.1

Calcule o valor da constante b que torna menor possível a $E(Y - Z - b)^2$? Para este valor de b, $Z + b$ tem menor emq que Z como previsor de Y? Calcule $W(Z + b, Y)$ para este valor de b.

PROBLEMA PARA DISCUSSÃO 6.2

Qual a melhor função de X para prever Y? Qual o seu valor? Qual a melhor função de Y para prever X? Qual o seu valor?

EXERCÍCIOS Lista 6

1. A fig. 6-2 apresenta o diagrama de dispersão para duas variáveis X e Y.

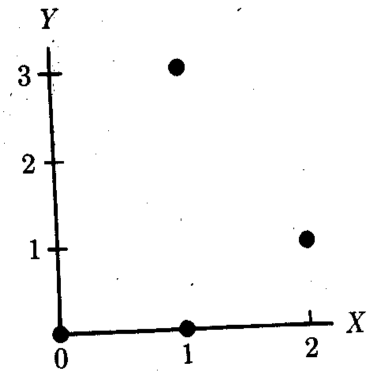


Fig. 6-2

a) Calcule o emq e o valor de cada uma das seguintes constantes, para prever Y: 0; 0,5; 1,5 e 2.

b) Calcule o emq e o valor de cada uma das seguintes funções de X, para prever Y: X, $2X$, $2X - 1$, $(0{,}8)X + 0{,}2$.

c) Determine a melhor função de X para prever Y e calcule seu valor.

d) Cada uma das funções do item b é linear em X, isto é, da forma $aX + b$, onde a e b são constantes.

É possível determinar constantes a e b tais q $aX + b$ seja um melhor previsor de Y (tenha mai valor) que qualquer uma das funções em (*b*)?

e) Encontre a melhor função de Y para prever X determine seu valor. É uma função linear de]

2. Uma moeda equilibrada é lançada três vezes. Repr sentemos por X o número de caras nos dois primeir lançamentos e por Y o número de caras nos três lanç mentos. Construa o diagrama de dispersão para X e (Lembre-se de atribuir a cada ponto um número propo cional à sua probabilidade).

a) Determine a melhor função de X para prever Qual o seu valor? É uma função linear de X?

b) Determine a melhor função de Y para prever . Qual o seu valor? É uma função linear de Y?

3. Refaça o exercício 2, supondo, agora, que a probat lidade de ocorrer cara em cada lançamento é 0,6.

4. Em cada um dos diagramas de dispersão da fig. 6- calcule o valor da melhor função de X para prever e da melhor função de Y para prever X.

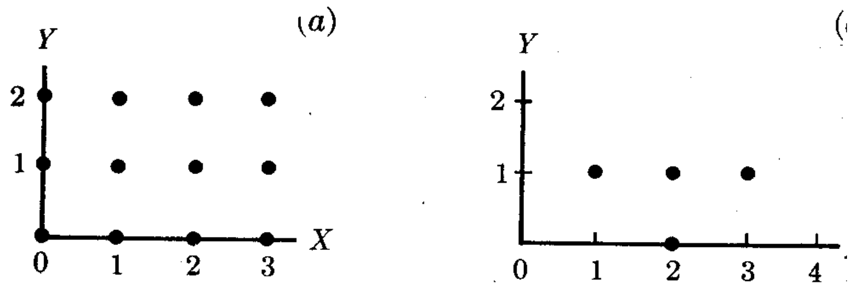


Fig. 6-3

5. Construa um diagrama de dispersão, contendo o men número de pontos, para duas variáveis X e Y, tais qu $E(X) = 1$ e $E(Y) = 3$. Suponha que você vai usar a va riável $2X + c$, onde c é uma constante, para prever Qual seria a melhor escolha para c? Calcule o em para $c = 0, 1, 2,$ e 3.

7 CORRELAÇÃO

Na previsão do comprimento Z de uma palavra, selecionada ao acaso, da frase I SEE THE MOUSE, supondo que você perde o quadrado de seu erro, seu melhor previsor constante é $E(Z) = 3$ e o emq é 2, uma vez que você perde 0 e 4 com probabilidade de $\frac{1}{2}$ cada, $2 = \delta^2(Z)$.

Se, antes de prever Z, você é informado do número X de letras E da palavra selecionda, podemos observar, no diagrama de dispersão para X e Z da fig. 7-1, que as melhores previsões, quando $X = 0$, 1 e 2 são, respectivamente, 1, 4 e 3; os pontos com um círculo em volta constituem o gráfico deste melhor previsor, como uma função de X. Usando este previsor, você perde 0 e 1 com probabilidade de $\frac{1}{2}$ cada, de tal forma que o emq é 0,5 e o valor do previsor é $(2 - 0{,}5)/2 = 0{,}75$. Os três pontos assinalados não estão sobre uma reta e, portanto, a melhor função de X para prever Z não é uma função linear de X, isto é, não é da forma $aX + b$, onde a e b são constantes. Qual é a melhor função linear de X para prever Z e qual o seu valor?

Para quaisquer duas variáveis X e Y, a *covariância* de X e Y é o número

$$\mathrm{Cov}(X, Y) = E(XY) - E(X)E(Y)$$

A melhor função linear de X para prever Y é

$$U = aX + b$$

onde $a = \dfrac{\mathrm{cov}(XY)}{\mathrm{cov}(XX)}$ e $b = E(Y) - aE(X)$

O valor desta melhor função linear de X para prever Y, denotado por $\rho^2(X, Y)$, recebe o nome de *coeficiente quadrático de correlação* entre X e Y e é dado pela fórmula

$$\rho^2(X, Y) = \frac{\mathrm{cov}^2(X, Y)}{\mathrm{cov}(X, X)\,\mathrm{cov}(Y, Y)}$$

A raiz quadrada de ρ^2, afetada do sinal (+) de a, é chamada *coeficiente de correlação* entre X e Y e é denotado por $\rho(X, Y)$.

Note que $\mathrm{cov}(X, Y) = \mathrm{cov}(Y, X)$, de tal forma que $\rho(X, Y) = \rho(Y, X)$ e que $\mathrm{cov}(X, X) = \sigma^2(X)$.

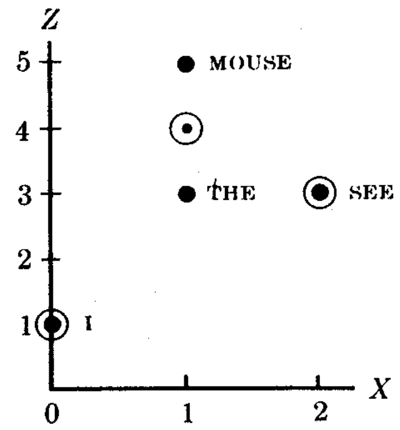


Fig. 7-1

Os cálculos necessarios para a determinação da melhor função linear de X para prever Y e seu valor são os seguintes:

Palavra	X	Z	X^2	XZ	Z^2	$U = X + 2$	$Z - U$ (erro)	$Z - U^2$ (erro quadrático)
I	0	1	0	0	1	2	−1	1
SEE	2	3	4	6	9	4	−1	1
THE	1	3	1	3	9	3	0	0
MOUSE	1	5	1	5	25	3	2	4
Σ	4	12	6	14	44	12	0	6
E	1	3	1,5	3,5	11	3	0	1,5

$$\text{cov}(X, X) = 1{,}5 - (1)^2 = 0{,}5$$

$$\text{cov}(X, Z) = 3{,}5 - 1(3) = 0{,}5$$

$$\text{cov}(Z, Z) = 11 - (3)^2 = 2$$

$$a = \frac{\text{cov}(X, Z)}{\text{cov}(X, X)} = \frac{0{,}5}{0{,}5} = 1$$

$$b = E(Z) - aE(X) = 3 - 1(1) = 2$$

$$\rho^2(X, Z) = \frac{\text{cov}^2(X, Z)}{\text{cov}(X, X)\,\text{cov}(Z, Z)} = \frac{(0{,}5)^2}{(0{,}5)(2)} = 0{,}25$$

A melhor função linear de X para prever Z é, portanto, $U = X + 2$ e seu valor é $\rho^2(X, Z) = 0{,}25$.

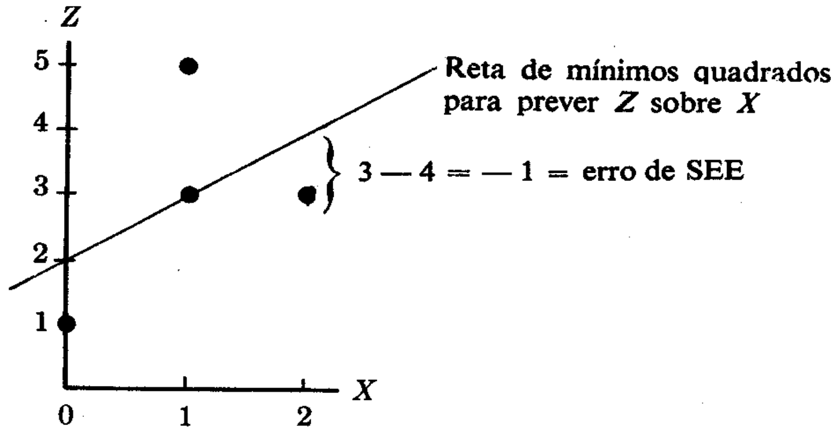


FIG. 7-2

A fig. 7-2 mostra o gráfico de U como uma função de X. Como U é linear, seu gráfico é uma reta, chamada *reta dos mínimos quadrados para prever Z sobre X*.

Observações

1. As três últimas colunas da tabela são obtidas após a e b terem sido calculados.
2. A média de U é igual à média da variável Z (que está sendo prevista); logo, $E(Z - U) = 0$. Isto será sempre verdade no caso do melhor previsor linear, de tal forma que os cálculos verificam nossa afirmação.
3. Calculamos o emq de U como previsor de Z e encontramos 1,5; logo, $W(U, Z) = (2 - 1{,}5)/2 = 0{,}25$, que é igual a $\rho^2(X, Z)$, servindo, novamente, para verificar nosso trabalho.
4. O emq poderia ter sido igual a 0,5 apenas e o valor 0,75 em vez de 1,5 e 0,25, respectivamente, se tivéssemos usado a melhor função de X para prever Z em lugar da melhor função *linear*.
5. A reta de mínimos quadrados pode ser desenhada, calculando-se dois pontos de previsão, isto é, quaisquer dois pares de valores (X, U), marcando-os no plano e passando por eles uma reta, por exemplo, quando $X = 0$, $U = 2$ e quando $X = 2$, $U = 4$, assim a reta de mínimos

quadrados passa pelos pontos (0,2) e (2,4). Outro ponto que pertence, sempre, à reta de mínimos quadrados é $(E(X), E(Z))$. No exemplo, $(E(X), E(Z)) = (1,3)$.

6. O erro associado a cada ponto do diagrama de dispersão é a distância do ponto à reta de mínimos quadrados, tomada verticalmente.

7. O número a mede o quanto nossa previsão de Z cresce, quando X cresce de uma unidade. É chamado *inclinação* da reta de mínimos quadrados. Quando $a = 0$, a melhor reta é horizontal, isto é, o melhor previsor linear é constante e, portanto, tem valor 0: $\rho^2(X, Z) = 0$.

PROBLEMA PARA DISCUSSÃO 7.1

No exemplo que estivemos estudando, qual o maior erro quadrático que poderíamos ter cometido, usando U para prever Z? É possível determinar outra função linear de X para prever Z, cujo maior erro quadrático seja menor que este?

PROBLEMA PARA DISCUSSÃO 7.2

A fig. 7-3 apresenta um diagrama de dispersão. Veja se você pode responder às seguintes questões, sem dispor de nenhum cálculo; depois, faça os cálculos e verifique suas respostas:

a) Qual a melhor reta horizontal para prever Y?
b) Qual o seu emq como previsor de Y, isto é, qual a variância de Y?
c) Qual a melhor função de X para estimar Y?
d) Qual a reta de mínimos quadrados de Y sobre X?
e) Qual seu emq, como previsor de Y?
f) Qual o valor da reta de mínimos quadrados para prever Y através de X, isto é, qual é o valor de $\rho^2(X, Y)$?

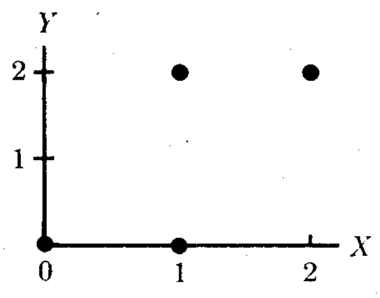


Fig. 7-3

EXERCÍCIOS
Lista 7

1. Uma palavra da frase I SEE THE MOUSE é selecionada ao acaso. X é o número de letras E na palavra selecionada, Y é o número de letras I na palavra selecionada e Z é o tamanho da palavra. Construa o diagrama de dispersão para cada um dos seguintes pares, determine ρ^2 e a melhor função linear do primeiro para prever o segundo e desenhe a reta de mínimos quadrados.

   a) (Z, Y)
   b) (Y, Z)
   c) $(X, Z - Y)$
   d) $(X + 4Y, Z)$
   e) (X^2, Z)

2. Tome qualquer diagrama de dispersão para duas variáveis X e Y, encontre a melhor função linear U de X para prever Y e verifique os seguintes fatos:

   a) $\sigma^2(U) + \sigma^2(Y - U) = \sigma(Y)$
   b) $\rho^2(U, Y - U) = 0$
   c) A melhor função linear de X para prever $2Y + 3$ é $2U + 3$.
   d) $\text{cov}(X - 1, Y + 2) = \text{cov}(X, Y)$ (Somar constantes a variáveis não altera a covariância entre elas).
   e) $\text{cov}(2X, -3Y) = -6\,\text{cov}(X, Y)$ (Multiplicando variáveis por constantes, a covariância fica multiplicada pelo produto dessas constantes).
   f) $\sigma^2(X + Y) = \sigma^2(X) + \sigma^2(Y) + 2\,\text{cov}(X, Y)$
   g) $\rho^2(2X - 1, Y) = \rho^2(X, Y)$ (Enuncie esta propriedade).

3. Para três variáveis quaisquer X, Y e Z, verifique se $\text{cov}(X, Y + Z) = \text{cov}(X, Y) + \text{cov}(X, Z)$.

4. Para cada um dos diagramas de dispersão na fig. 7-4, acrescente um quarto ponto: *a*) para tornar $\rho^2(X, Y) = 0$ ou *b*) para tornar $\rho^2(X, Y)$ maior possível.

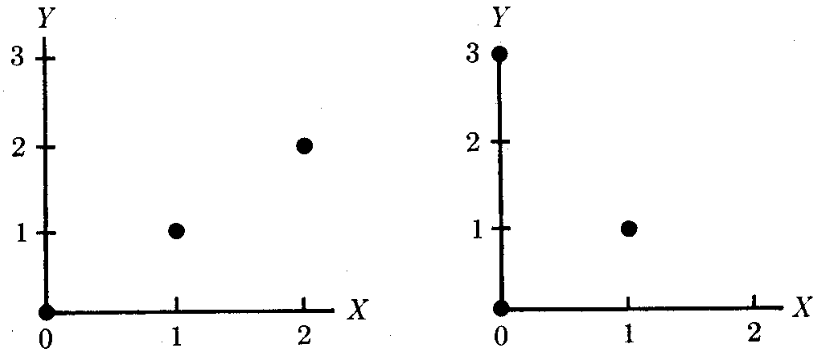


Fig. 7-4

5. Um diagrama de dispersão para X e Y tem três pontos: $(-1{,}0)$, $(1{,}0)$ e um terceiro ponto $Q = (h, k)$. Encontre a melhor função linear $U = aX + b$ de X para prever Y e calcule $\rho^2(X, Y)$, para cada um dos nove pontos Q com h, $k = 1{,}3$ e 10, apresentando o resultado em tabelas:

a

h \ k	1	3	10
1			
3			
10			

b

h \ k	1	3	10
1			
3			
10			

ρ^2

h \ k	1	3	10
1			
3			
10			

Para h fixo, de que modo a depende de k (cresce, decresce, permanece constante)? Para h fixo, de que modo b e ρ^2 dependem de k? Para k fixo, de que modo a, b e ρ^2 dependem de h?

Encontre um ponto Q com $0{,}4 \leq \rho^2 \leq 0{,}6$ e $0 < a < 0{,}5$. Construa o diagrama de dispersão e a reta de mínimos quadrados para este Q.

6. A tabela abaixo mostra a fração do total dos votos que um candidato a governador e um candidato a senador receberam em cada um de quatro distritos eleitorais:

Distrito	*G*	*S*
1	0,5	0,4
2	0,8	0,6
3	0,3	0,3
4	0,4	0,3

Determine a melhor função linear de G para prever S e $\rho^2(G, S)$. (Os cálculos serão simples, se você usar $X = G - 0{,}5$ e $Y = S - 0{,}3$).

7. A urna 1 contém 40% de bolas pretas e a urna 2, por sua vez, 80% de bolas pretas. Uma das duas urnas é selecionada ao acaso e uma bola é extraída também ao acaso, da urna selecionada. Encontre a melhor função linear de X, número de bolas pretas sorteadas, para prever Y, a proporção de bolas pretas na urna selecionada e calcule $\rho^2(X, Y)$.

8. Numa certa população, 50% das pessoas não possuem nenhum carro e 50% possuem um carro. Qualquer pessoa, quando se pergunta quantos carros possui, diz a verdade com probabilidade de 0,8 e mente com porbabilidade de 0,2. Calcule a correlação entre X, a resposta dada, e Y, o número de carros que a pessoa possui.

8 CORRELAÇÃO MÚLTIPLA E PARCIAL

Suponha que você tem de prever o comprimento Z de uma palavra, da frase **I See the Mouse**, selecionada ao acaso, e você perde o quadrado do seu erro. Antes de estimar Z, você terá conhecimento dos valores assumidos pelas duas variáveis:

$X =$ número de letras E da palavra selecionada

$Y =$ número de letras I da palavra selecionada,

porém deverá utilizar uma função linear

$V = cX + dY + e$ $\quad$ c, d e e constantes

de X e Y para prever Z. Qual é a melhor função linear V e qual o seu valor?

A melhor função linear de duas variáveis X e Y, para prever uma terceira variável Z, é

$$V = cX + dY + e,$$

onde

$$c = \frac{\text{cov}(X, Z)\,\text{cov}(Y, Y) - \text{cov}(Y, Z)\,\text{cov}(X, Y)}{D}$$

$$d = \frac{\text{cov}(Y, Z)\,\text{cov}(X, X) - \text{cov}(X, Z)\,\text{cov}(Y, X)}{D}$$

$$e = E(Z) - cE(X) - dE(Y)$$

$$D = \text{cov}(X, X)\,\text{cov}(Y, Y) - \text{cov}^2(X, Y)$$

O valor de V na previsão de Z, chamado *coeficiente quadrático de correlação múltipla entre* Z *e* (X, Y) e simbolizado por $\rho^2(Z, (X, Y))$, é dado por

$$\rho^2(Z, (X, Y)) = \frac{c\,\text{cov}(X, Z) + d\,\text{cov}(Y, Z)}{\text{cov}(Z, Z)}$$

A quantidade que $\rho^2(Z, (X, Y))$ excede $\rho^2(Z, X)$, dividida pelo maior acréscimo possível, $1 - \rho^2(Z, X)$, é chamado *coeficiente quadrático de correlação parcial entre* Y *e* Z *dado* X e denotado por $\rho^2(Y, Z \mid X)$:

$$\rho^2(Y, Z \mid X) = \frac{\rho^2(Z, (X, Y)) - \rho^2(Z, X)}{1 - \rho^2(Z, X)}$$

A raiz quadrada de $\rho^2(Y, Z \mid X)$, afetada do sinal $(\pm)$ de b, é chamado *coeficiente de correlação parcial entre* Y *e* Z *dado* X e representado por $\rho(Y, Z \mid X)$.

Os cálculos para o nosso exemplo são os seguintes:

Palavra	X	Y	Z	X^2	Y^2	Z^2	XY	XZ	YZ	V	$(Z-V)$	$(Z-V)^2$
I	0	1	1	0	1	1	0	0	1	1	0	0
SEE	2	0	3	4	0	9	0	6	0	3	0	0
THE	1	0	3	1	0	9	0	3	0	4	-1	1
MOUSE	1	0	5	1	0	25	0	5	0	4	1	1
Σ	4	1	12	6	1	44	0	14	1	12	0	2
E	1	$\frac{1}{4}$	3	$\frac{3}{2}$	$\frac{1}{4}$	11	0	$\frac{7}{2}$	$\frac{1}{4}$	3	0	$\frac{1}{2}$

Tabela de Covariância

	X	Y	Z
X	$\frac{1}{2}$	$-\frac{1}{4}$	$\frac{1}{2}$
Y	$-\frac{1}{4}$	$\frac{3}{16}$	$-\frac{1}{2}$
Z	$\frac{1}{2}$	$-\frac{1}{2}$	2

$$D = \text{cov}(X, X)\,\text{cov}(Y, Y) - \text{cov}^2(X, Y) = \tfrac{1}{2}(\tfrac{3}{16}) - (-\tfrac{1}{4})^2 = \tfrac{1}{32}$$

$$c = \frac{\text{cov}(X, Z)\,\text{cov}(Y, Y) - \text{cov}(Y, Z)\,\text{cov}(X, Y)}{D} = \frac{\frac{1}{2}(\frac{3}{16}) - (-\frac{1}{2})(-\frac{1}{4})}{\frac{1}{32}} = -1$$

$$d = \frac{\text{cov}(Y, Z)\,\text{cov}(X, X) - \text{cov}(X, Z)\,\text{cov}(Y, X)}{D} = \frac{-\frac{1}{2}(\frac{1}{2}) - \frac{1}{2}(-\frac{1}{4})}{\frac{1}{32}}$$

$$e = E(Z) - cE(X) - dE(Y) = 3 - (-1)(1) - (-4)(\tfrac{1}{4}) = 5$$

$$\rho^2(Z, (X, Y)) = \frac{-1(\frac{1}{2}) + (-4)(-\frac{1}{2})}{2} = \frac{3}{4}$$

$$\rho^2(Z, X) = \frac{(\frac{1}{2})^2}{\frac{1}{2}(2)} = \frac{1}{4}$$

$$\rho^2(Z, Y) = \frac{(-\frac{1}{2})^2}{\frac{3}{16}(2)} = \frac{2}{3}$$

$$\rho^2(Y, Z \mid X) = \frac{\frac{3}{4} - \frac{1}{4}}{1 - \frac{1}{4}} = \frac{2}{3}$$

$$\rho^2(X, Z \mid Y) = \frac{\frac{3}{4} - \frac{2}{3}}{1 - \frac{2}{3}} = \frac{1}{4}$$

A melhor função linear de X e Y para prever Z é $V = -X - 4Y + 5$ e seu valor é $\rho^2(Z, (X, Y)) = \frac{3}{4}$.

Observações

1. $E(Z - V) = E((Z - V)X) = E((Z - V)Y) = 0$. O erro $(Z - V)$ tem média 0 e é não correlacionado com X e Y. Isto será sempre verdade.

2. Podemos calcular o valor de V como previsor de Z diretamente:

$$W(V, Z) = \frac{2 - \frac{1}{2}}{2} = \frac{3}{4}$$

3. O coeficiente a de X é negativo. Se, para um Y fixado, fizermos crescer o valor de X, nossa previsão de Z cresce, em contraposição a $X + 2$, que, como vimos anteriormente, é a melhor função linear de X para estimar Z.

EXERCÍCIOS

Lista 8

1. Damos, a seguir, os valores de três variáveis X, Y e Z, em uma população com 5 elementos

Elemento	X	Y	Z
a	0	1	0
b	0	1	1
c	1	2	2
d	2	2	2
e	2	4	4

Um dos elementos é selecionado ao acaso. Determine a melhor função linear U de X para prever Z, a melhor função linear V de X e Y para prever Z, a melhor função linear W de X para prever Y e calcule $\rho^2(X, Z)$, $\rho^2(Y, Z)$, $\rho^2(Z, (X, Y))$, $\rho^2(Y, Z \mid X)$ e $\rho^2(Z, X \mid Y)$.

2. Para quaisquer três variáveis X, Y, Z, a correlação parcial entre Y e Z, dado X, é igual à correlação entre os erros nas melhores funções lineares de X para prever Y e Z:

$$\rho(Y, Z \mid X) = \rho(Y - W, Z - U)$$

Verifique esta relação para X, Y e Z, no exercício 1.

3. Suponha que X e Y são não-correlacionadas, isto é, $\text{cov}(X, Y) = 0$. Simplifique as fórmulas para c e d neste caso. Você reconhece o resultado?

4. Uma moeda, com probabilidade 0,6 de dar cara, é lançada 3 vezes e X_i representa o número de caras nos primeiros i lançamentos. Determine a melhor função linear de:

a) X_1 e X_2 para prever X_3
b) X_1 e X_2 para prever X_2
c) X_2 e X_3 para prever X_1

5. Para o diagrama de dispersão da fig. 8-1, determine a melhor função linear de X e X^2, para prever Y.

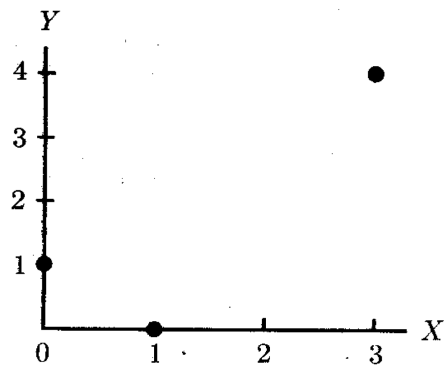
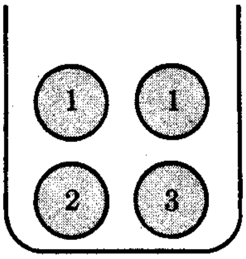
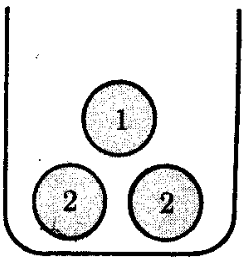


Fig. 8-1

9 INDEPENDÊNCIA



Urna 1



Urna 2

Fig. 9-1

A fig. 9-1 apresenta duas urnas. Retiram-se, ao acaso e com reposição, duas bolas da urna 1 e, em seguida, retira-se, ao acaso, uma bola da urna 2. Denotamos os números das bolas retiradas por X, Y e Z. É claro que qualquer uma das variáveis não tem utilidade alguma na previsão de qualquer das outras duas, bem como quaisquer duas delas não terão também utilidade para a previsão da terceira. De fato, os valores de quaisquer duas não nos dão nenhuma informação a respeito da terceira:

$$P(Y = 2 \mid X = 1,\ Z = 2) = P(Y = 2) = 0{,}25$$

Dizemos que X, Y e Z são *independentes*.

Duas ou mais variáveis X, Y, ... são *independentes* se a distribuição de cada uma, dados os valores de todas as outras, é sempre igual à sua distribuição incondicional, isto é, não depende dos valores das outras variáveis.

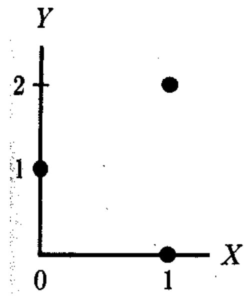


Fig. 9-2

Se duas variáveis são independentes, elas são, certamente, não-correlacionadas. Mas o diagrama de dispersão da fig. 9-2 mostra-nos duas variáveis X e Y que não são correlacionadas, mas que não são independentes.

A variância de uma soma de variáveis independentes é igual à soma de suas variâncias:

$\sigma^2(X + Y + ...) = \sigma^2(X) + \sigma^2(Y) + ...$

se X, Y, ... são independentes.

De uma maneira geral, a variância de uma soma de variáveis é igual à soma das variâncias, quando quaisquer duas variáveis forem não-correlacionadas. Por exemplo, para duas variáveis X e Y,

$$\begin{aligned}\sigma^2(X + Y) &= \text{cov}(X + Y, X + Y)\\ &= \text{cov}(X, X + Y) + \text{cov}(Y, X + Y)\\ &= \text{cov}(X, X) + \text{cov}(X, Y) + \text{cov}(Y, X) +\\ &+ \text{cov}(Y, Y)\\ &= \sigma^2(X) + \sigma^2(Y) + 2\,\text{cov}(X, Y)\end{aligned}$$

Assim, $\sigma^2(X + Y) = \sigma^2(X) + \sigma^2(Y)$, se $\text{cov}(X, Y) = 0$, isto é, se X e Y são não-correlacionadas.

Um caso importante de independência é a amostragem aleatória, com reposição, de uma distribuição conhecida. Por exemplo, se 10 bolas são retiradas da urna 1, ao acaso e com reposição, e $X_1, ..., X_{10}$; representam os números das bolas retiradas, as variáveis $X_1, \ldots, X_{10}$; são independentes, com a mesma distribuição:

v	p
1	0,5
2	0,25
3	0,25

As variáveis constituem uma *amostra aleatória* desta distribuição.

Por outro lado, se selecionássemos uma das duas urnas ao acaso e retirássemos dessa urna 10 bolas ao acaso e com reposição, os números $X_1, \ldots, X_{10}$; das bolas selecionadas não seriam independentes. Se $X_1 = 3$, sabemos que a urna 1 foi selecionada, desta forma $P(X_2 = 1 | X_1 = 3) = 1/2$, enquanto a probabilidade incondicional de ser $X_2 = 1$ é

$$P(X_2 = 1) = P(\text{urna } 1 \text{ e } X_2 = 1) +$$
$$+ P(\text{urna } 2 \text{ e } X_2 = 1) = \tfrac{1}{2}(\tfrac{1}{2}) + \tfrac{1}{2}(\tfrac{1}{3}) = \tfrac{5}{12}$$

EXEMPLO 9.1 Uma amostra aleatória $X_1, \ldots, X_{10}$; será extraída da seguinte distribuição:

v	p
1	0,4
2	0,2
3	0,4

Você deve prever (*a*) a soma $S = X_1 + \ldots + X_{10}$; dos valores amostrais e (*b*) a média $(X_1 + \ldots + X_{10})/10 = S/10$ dos valores amostrais, sabendo que você perde o quadrado do seu erro. Qual a melhor previsão e qual o emq?

A melhor previsão para S é

$$E(S) = E(X_1 + \ldots + E(X_{10})$$
$$= 2 + 2 + \ldots + 2 = 20$$

e o emq é

$$\begin{aligned}\sigma^2(S) &= \sigma^2(X_1) + \ldots + \sigma^2(X_{10})\\ &= 0{,}8 + \ldots + 0{,}8 = 8\end{aligned}$$

Sua melhor previsão para $S/10$ é

$$E\left(\frac{S}{10}\right) = \frac{E(S)}{10} = 20/10 = 2$$

e o emq é

$$\sigma^2\left(\frac{S}{10}\right) = \frac{\sigma^2(S)}{10^2} = 8/100 = 0{,}08$$

Note que

$$\sigma\left(\frac{S}{10}\right) = \sqrt{0{,}08}$$

$$\sigma(X_1) = \sqrt{0{,}8}$$

Desta forma,

$$\sigma\left(\frac{S}{10}\right) = \frac{\sigma(X_1)}{\sqrt{10}}$$

A variável $(X_1 + \ldots + X_{10} / 10 = S/10$ é chamada *média da amostra* e é, usualmente, denotada por $\bar{X}$.

Se $X_1, \ldots, X_n$ é uma amostra aleatória de uma distribuição, a média da média da amostra é a média m da distribuição:

$$\begin{aligned}E(\bar{X}) = E\left(\frac{X_1 + \ldots + X_n}{n}\right) &= E(X_1)\\ &= E(X_n) = m\end{aligned}$$

e o desvio padrão da média da amostra é o desvio padrão σ da distribuição dividido pela raiz quadrada do tamanho da amostra:

$$\sigma^2(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

EXEMPLO 9.2 Numa população, 80% das pessoas são do sexo masculino. Seleciona-se uma amostra aleatória, com reposição, de tamanho 25 e X é a proporção de pessoas do sexo masculino na amostra. Calcule $E(\overline{X})$ e $\sigma(\overline{X})$.

Se escrevermos $X_i = 1$, se o i-ésimo indivíduo é do sexo masculino e $X_i = 0$, se o i-ésimo indivíduo é do sexo feminino, então $X_1 \ldots, X_{25}$ é uma amostra aleatória da seguinte distribuição:

v	p
0	0,2
1	0,8

cuja média é $m = 0.(0{,}2) + 1(0{,}8) = 0{,}8$, o desvio padrão $\sigma = \sqrt{0{,}8 - (0{,}8)^2} = \sqrt{0{,}16} = 0{,}4$ e

$$\overline{X} = (X_1 + \ldots + X_{25})/25$$

Então,

$$E(\overline{X}) = m = 0{,}8$$

$$\sigma(X) = \sigma/\sqrt{25} = \frac{0{,}4}{5} = 0{,}08$$

PROBLEMA PARA DISCUSSÃO 9.1 No exemplo 9.2, que tamanho de amostra seria necessário para termos

$\sigma(\overline{X}) = 0{,}04?$ $0{,}02?$ $0{,}01?$

EXERCÍCIOS Lista 9

1. Existem três urnas, sendo que cada uma contém uma bola preta e a urna i contém i bolas brancas. Uma bola é retirada ao acaso de cada uma das urna se você recebe Cr$ 1,00 por bola preta retirada. Calcule a média, variância e o desvio padrão de sua renda total.

2. Bolas são retiradas, ao acaso e com reposição, de uma urna que contém dez bolas numeradas 0, 1, ..., 9. Você recebe:

Cr\$ 1,00 quando 0, 1, 2 ou 3 é retirada;

Cr\$ 2,00 quando 4, 5, 6, 7 ou 8 é retirada;

Cr\$ 4,00 quando 9 é retirada.

Assim, sua renda Y_i, proveniente da i-ésima extração, é

$$\begin{aligned} Y_i &= 1, \text{ se } X_i \leq 3 \\ &= 2, \text{ se } 4 \leq X_i \leq 8 \\ &= 4, \text{ se } X_i = 9 \end{aligned}$$

onde X_i é o número da i-ésima bola extraída.

Determine a distribuição da renda proveniente das duas primeiras bolas extraídas, sua média e variância. $Y_1, Y_2, \ldots$ é uma amostra de alguma distribuição?

3. Suponha que você dispõe de uma urna, como a do exercício 2. Como poderia produzir uma amostra aleatória da seguinte distribuição?

v	p
10	0,2
12	0,4
14	0,3
20	0,1

Como poderia produzir uma amostra aleatória de cada uma das distribuições abaixo?

v	p
10	0,18
12	0,43
14	0,28
20	0,11

*v	p
0	1/11
1	1/11
2	.
.	.
.	.
10	1/11

4. Experimentos análogos ao de retirar bolas, ao acaso e com reposição, de uma urna, como a do exercício 2, têm sido constantemente executados e os resultados $X_1, X_2, \ldots$, registrados em tabelas, chamadas *tabelas dos números aleatórios*. Suponha que você dispõe de um par de dados equilibrados, digamos um vermelho e outro verde. Como poderia, usando o par de dados, contruir uma tabela de números aleatórios? Seria possível, usando uma moeda honesta, construir uma tabela de números aleatórios? Construa da maneira que você quiser uma tabela de números aleatórios, com 100 entradas.

5. Suponha que você tem uma moeda, cuja probabilidade de ocorrer cara é de 0,6. Você lança a moeda repetidamente e considere como um par os dois primeiros lançamentos; como outro par, os dois lançamentos seguintes e assim sucessivamente. Agora, você elimina todos os pares de elementos iguais e define $X_i = 1$, se o i-ésimo par restante for HT, ou $X_i = 0$ se o par for TH. Se os 12 primeiros lançamentos são

	$\cancel{HH}$	TH	HT	$\cancel{TT}$	$\cancel{TT}$	HT
você tem		0	1			1
		X_1	X_2			X_3

De que distribuição, se existir, $X_1, X_2, \ldots$ é uma amostra aleatória?

6. Se X e Y são independentes, $\text{cov}(X, Y) = E(XY) - E(X)E(Y) = 0$, isto é, $E(XY) = E(X)E(Y)$. De uma maneira geral, se $X_1, X_2, \ldots, X_n$ são independentes,

$$E(X_1X_2\ldots X_n) = E(X_1)E(X_2)\ldots E(X_n)$$

Usando este fato, calcule $E(X_1X_2 \ldots X_n)$, onde $X_1 \ldots X_n$ é uma amostra aleatória da distribuição:

v	p
0	0,7
1	0,3

Qual a distribuição de $X_1X_2 \ldots X_6$?

7. Extrai-se uma amostra aleatória da distribuição abaixo, até que se obtenha um 1 ou um 3.

v	p
1	0,4
2	0,3
3	0,2
4	0,1

Represente por Y a última variável observada e por n o número de observações. Calcule $P(n = 1 \text{ e } Y = 1)$, $P(Y = 1 \mid n = 1)$, $P(n = 2 \text{ e } Y = 1)$, $P(Y = 1 \mid n = 2)$, $P(Y = 1)$, $P(n > 5)$ e $P(n \leq 5)$.

8. A fig. 9-3 apresenta três páginas de um pequeno livro. Você seleciona uma delas, ao acaso; desta página, seleciona uma linha ao acaso e, finalmente, seleciona ao

1	2	3
Excessive bail shall not be re quired, nor exces-	sive fines imposed, nor cruel and un- usual punish-	ments inflicted.

Fig. 9-3

acaso uma palavra da linha selecionada. (Vamos convencionar que uma palavra pertence à linha em que ela inicia.) Qual a probabilidade de que a palavra *shall* seja selecionada? Que *inflicted* seja selecionada?

Suponha que você deseja selecionar ao acaso uma palavra do livro, sem primeiro contar o número de palavras. Você sabe apenas que (*a*) o livro tem três páginas, (*b*) nenhuma página tem mais de cinco linhas e (*c*) nenhuma linha tem mais de seis palavras.

Propomos o seguinte método: Selecione $u = 1, 2, 3$ ao acaso, $v = 1, 2, 3, 4, 5$ ao acaso e $w = 1, 2, 3, 4, 5, 6$ ao acaso. Então, tome a w-ésima palavra da v-ésima linha da u-ésima folha, se ela existir; caso contrário, repita o experimento. Continue sempre, até ter conseguido a palavra. Assim, (1, 2, 3) nos leva à palavra *required*, mas (2, 3, 3) ou (3, 2, 1) não nos levam a palavra alguma.

Explique por que este processo funciona ou não.

10 DISTRIBUIÇÃO BINOMIAL

Nos últimos cinco anos, foram aplicadas vacinas contra varíola em 30% de uma população. Selecionam-se, ao acaso e com reposição, dez pessoas desta população. Qual a probabilidade de que exatamente quatro pessoas nesta amostra tenham sido vacinadas nos últimos cinco anos? De uma maneira geral, qual a distribuição de X, número de pessoas na amostra que foram vacinadas nestes últimos cinco anos?

Imaginemos que estamos com uma lista contendo os $2^{10} = 1.024$ resultados possíveis do experimento, como vemos no quadro seguinte:

Resultado	*Probabilidade*	X
NNNNNNNNNN	$(0,7)(0,7)\ldots(0,7)(0,7) = (0,7)^{10}$	0
NNNNNNNNNY	$(0,7)(0,7)\ldots(0,7)(0,3) = (0,3)\,(0,7)^9$	1
NNNNNNNNYN	$(0,7)(0,7)\ldots(0,3)(0,7) = (0,3)\,(0,7)^9$	1
. . . . .		
YNNNYNNNYY	$(0,3)(0,7)\ldots(0,3)(0,3) = (0,3)^4(0,7)^6$	4
. . . . .		
YYYYYYYYYY	$(0,3)(0,3)\ldots(0,3)(0,3) = (0,3)^{10}$	10

(Y representa *sim* e N representa *não*)

Ao lado de cada resultado, está sua probabilidade e o valor de X, isto é, o número de letras Y.

Podemos calcular $P(X = 4)$ percorrendo o quadro, marcando os resultados em que $X = 4$ e somando suas probabilidades. Cada resultado com $X = 4$, isto é, com 4 letras Y e 6 letras N, terá probabilidade $(0,3)^4(0,7)^6$, de tal forma que

$$P(X = 4) = C(10, 4)\,(0,3)^4(0,7)^6$$

onde $C(10, 4)$ representa o número de resultados em que aparecem exatamente 4 letras Y.

Em geral, denotamos por $C(n, k)$ o número de palavras com n letras, usando somente as letras N e Y, com exatamente k letras Y. A tabela seguinte mostra-nos valores de $C(n, k)$ para $n = 1, \ldots, 10$.

n \ k	0	1	2	3	4	5	6	7	8	9	10
1	1	1									
2	1	2	1								
3	1	3	3	1							
4	1	4	6	4	1						
5	1	5	10	10	5	1					
6	1	6	15	20	15	6	1				
7	1	7	21	35	35	21	7	1			
8	1	8	28	56	70	56	28	8	1		
9	1	9	36	84	126	126	84	36	9	1	
10	1	10	45	120	210	252	210	120	45	10	1

Você é capaz de explicar a coluna $k = 0$? A coluna $k = 1$? Os números 1 que aparecem na diagonal? O fato de que cada linha é simétrica em relação a seu centro, por exemplo, que $C(8, 6) = C(8, 6)$? Qual a soma dos elementos da primeira linha? Da segunda linha? Da quinta linha? Qual seria a soma dos elementos da décima linha? Explique. Você percebe alguma regra para calcular os números de uma linha, conhecendo os elementos da linha imediatamente acima? Use esta regra para calcular os elementos da décima-primeira linha. Qual a soma dos elementos dessa linha? Você pode explicar porque esta regra funciona?

Em nosso problema original,

$P(X = 4) = 210.(0,3)^4(0,7)^6$

A distribuição de X é a seguinte:

v	p
0	$(0,7)^{10}$
1	$10(0,3)\ (0,7)^9$
2	$45(0,3)^2\ (0,7)^8$
3	$120(0,3)^3\ (0,7)^7$
4	$210(0,3)^4\ (0,7)^6$
5	$252(0,3)^5\ (0,7)^5$
6	$210(0,3)^6\ (0,7)^4$
7	$120(0,3)^7\ (0,7)^3$
8	$45(0,3)^8\ (0,7)^2$
9	$10(0,3)^9\ (0,7)$
10	$(0,3)^{10}$

Como ficaria a distribuição, se 0,3 fosse substituído por 0,31? E se 10 fosse substituído por 8?

O número 10 em nosso problema é chamado *tamanho da amostra* ou *número de ensaios*, o número 0,3 é denominado *probabilidade de sucesso* em um ensaio, a variável X é denominada *número de sucessos em* 10 *ensaios independentes com probabilidade de sucesso* 0,3 e a distribuição de X é denominada *distribuição binomial com* 10 *ensaios e probabilidade de sucesso* 0,3 ou *distribuição binomial com* $n = 10$ *e* $p = 0,3$.

Se retirarmos uma amostra aleatória de n elementos, com reposição, de uma população, na qual a probabilidade de um certo atributo é p, o número X de elementos que, em nossa amostra, são portadores do atributo pode assumir os valores $0, 1, 2, \ldots, n$ e

$$P(X = k) = C(n, k)p^k q^{n-k} \quad \cdot\, q^{n-k}$$

onde $q = 1 - p$. O número X é denominado *variável binomial com parâmetros n e p*, e sua distribuição é denominada *distribuição binomial com parâmetros n e p*.

Se $X_1, X_2, \ldots, Xn$ é uma amostra aleatória da distribuição

v	p
0	q
1	p

a sua soma $S = X_1 + X_2 + \ldots + X_n$ é uma variável binomial com parâmetros n e p, assim:

$$E(S) = p + p + \ldots + p = np$$

$$\sigma^2(S) = pq + pq + \ldots + pq = npq$$

A média e a variância de uma variável binomial S com n ensaios e probabilidade de sucesso p são

$$E(S) = np \text{ e } \sigma^2(S) = npq, \text{ onde } q = 1 - p$$

Vamos, agora, determinar uma fórmula para $C(n, k)$. Para calcular $C(10, 4)$, número de *palavras* que podemos formar com 10 letras, das quais 4 são Y e o restante N, imagine uma urna com 10 bolas, sendo 4 marcadas com Y e 6 marcadas com N. Retiramos ao acaso as bolas da urna, sucessivamente e sem reposição, construindo uma palavra de 10 letras. Um dos possíveis resultados é

YNNYNNNYYN

Qual a probabilidade desse resultado?

$P(YNNYNNNYYN) =$

$$= \frac{4}{10} \times \frac{6}{9} \times \frac{5}{8} \times \frac{3}{7} \times \frac{4}{6} \times \frac{3}{5} \times \frac{2}{4} \times \frac{2}{3} \times \frac{1}{2} \times \frac{1}{1}$$

Note, então, que o denominador é $10 \times 9 \times 8 \times \ldots \times \times 3 \times 2 \times 1$. Denotamos este número por 10! (leia-se 10 *fatorial*). Os numeradores para os Y são 4, 3, 2, 1, de tal forma que seu produto é 4!, e os numeradores para os N são 6, 5, 4, 3, 2, 1, e seu produto é igual a 6! Assim, nossa palavra tem probabilidade

$$\frac{4!(6!)}{10!} = \frac{4 \times 3 \times 2 \times 1}{10 \times 9 \times 8 \times 7} = \frac{1}{210}$$

Para uma outra palavra qualquer, obteríamos, novamente,

$$\frac{4!(6!)}{10!} = \frac{1}{210}$$

para sua probabilidade; desta forma, deverão existir

$$210 = \frac{10!}{4!(6!)} \text{ palavras}$$

$$C(10, 4) = \frac{10!}{4!(6!)}$$

O mesmo método nos dá a fórmula geral

$$C(n, k) = \frac{n!}{k!(n-k)!}$$

A fórmula estará correta para $k = 0$ e $k = n$, se definirmos $0! = 1$.

EXERCÍCIOS
Lista 10

1. Se X é uma variável binomial com parâmetros n e p, calcule $P(X=k)$ para cada conjunto de valores de n, p e k, dados na tabela seguinte:

n	2	2	2	3	4	4	4	4	4	4	4	4
p	0,4	0,4	0,4	0,4	0,4	0,4	0,5	0,6	0,3	0,2	0	1
k	0	1	2	0	0	1	1	1	1	1	1	1
$P(X = k)$												

2. Construa o histograma para uma variável binomial, onde $n = 5$ e p é igual a um dos seguintes valores: 0,4; 0,5; 0,6. Que fração da área do histograma está compreendida num raio de 1 desvio padrão em torno da média?

3. Uma amostra aleatória de 4 elementos será retirada, com reposição, de uma certa população e você ganha um prêmio, se exatamente um dos elementos selecionados tiver um certo atributo. Faça um gráfico de sua probabilidade de ganhar o prêmio como função de p, a probabilidade do atributo na população. Qual o valor de p que lhe dá maior chance? Qual a sua chance para este valor de p?

4. Numa certa população, 40% dos eleitores são democratas. Você pode determinar o tamanho da amostra n e este número de eleitores é selecionado ao acaso. com reposição, dessa população. Você ganha um prêmio se houver exatamente dois democratas na amostra. Que valor de n lhe dá a maior chance? Qual a sua chance para este valor de n?

5. Uma variável X é binomial com $n = 10$. Faça um gráfico de $E(X)$, $\sigma^2(X)$, $E(X/10)$ e $\sigma^2(X/10)$ como função de p, sendo $0 \leq p \leq 1$.

6. Uma variável X é binomial com $p = 0{,}4$. Faça um gráfico de $E(X)$, $\sigma^2(X)$, $E(X/n)$ e $\sigma^2(X/n)$ como função de n; $1 \leq n \leq 10$.

7. Uma urna contém uma bola preta e $n - 1$ bolas brancas. Uma seleção de n bolas será feita ao acaso e você recebe Cr 1,00 cada vez que a bola preta é retirada. Calcule a média e a variância de sua renda total (*a*) com reposição e (*b*) sem reposição, para cada um dos valores seguintes de n: 1, 2, 3, 5, 10. (*c*) Cal-

cule a probabilidade de que sua renda seja 0 para os valores de n acima e extrações com reposição.

8. Se X é binomial com $n = 100$ e $p = 0{,}3$, expresse $P(X = 37)$ em termos de fatoriais. Expresse $P(X = 38)$ e calcule a razão

$$r = \frac{P(X = 38)}{P(X = 37)}$$

Qual é maior: $P(X = 37)$ ou $P(X = 38)$? Calcule a razão

$$\frac{P(X = k + 1)}{P(X = k)}$$

como função de k. Para que valores de k essa razão é maior do que 1? Para que valores de k, $P(X = k + 1)$ é maior do que $P(X = k)$? Qual o valor de X mais provável?

9. Use a fórmula $C(n, k)$, para conferir a tabela dos valores $C(n, k)$ para $n = 2$ e para $n = 7$.

10. Uma moeda honesta é lançada $2k$ vezes de tal forma que X, número de caras obtidas, é uma variável binomial com $n = 2k$ e $p = 0{,}5$ e a probabilidade de que o número de caras iguale o número de coroas é

$$p(k) = \frac{(2k!)}{(k!)\,(k!)}(0{,}5)^k(0{,}5)^k$$

Calcule $p(k)$ para $k = 1, 2, 3$ e 4. Complete a tabela abaixo, dando cada entrada às quatro figuras significantes.

k	$p(k)$	$p^2(k)$	$kp^2(k)$	$(k + \frac{1}{2})p^2(k)$
1				
2				
3				
4				

Tente adivinhar $p(5)$, a partir desta tabela e, a seguir, calcule este valor. Qual a grandeza aproximada de $p(100)$? $p(1.000.000)$? Você acha que os números $kp^2(k)$ crescem quando k cresce? Que os números $(k + 1/2)$ $p^2(k)$ decrescem? O restante deste problema requer álgebra.

Escreva uma fórmula para a razão

$$\frac{(k + 1)\, p^2(k + 1)}{kp^2(k)}$$

simplificando o máximo possível. Chegou à conclusão de que os números $kp^2(k)$ crescem? Escreva uma fórmula para a razão

$$\frac{(k + 1 + 1/2)p^2(k + 1)}{(k + 1/2)p^2(k)}.$$

simplificando o máximo possível. Chegou à conclusão de que os números $(k + 1/2)p^2(k)$ decrescem?

11. Cada um de dois distritos tem 100 eleitores. O número de democratas no distrito A é 20 e, no distrito B, é 60. A seguir, damos três métodos para selecionar dois eleitores:

a) Seleciona-se, ao acaso, um eleitor em cada distrito.

b) Seleciona-se, ao acaso com reposição, dois eleitores entre os 200.

c) Seleciona-se, ao acaso, um distrito e, em seguida, selecionam-se, ao acaso com reposição, dois eleitores do distrito selecionado.

Calcule a média, a variância e determine a distribuição de S, número de democratas selecionados, para cada um dos métodos de seleção.

11 APROXIMAÇÃO NORMAL

A seguir, temos a distribuição de uma variável binominal X com $n = 6$ e $p = 0,4$ e a fig. 11-1 nos mostra seu histograma.

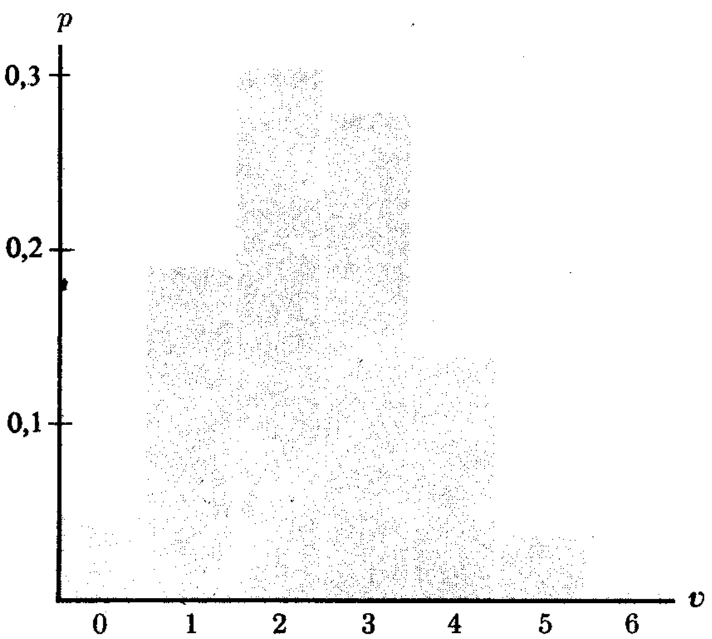


FIG. 11-1

v	p
0	0,047
1	0,187
2	0,311
3	0,276
4	0,13*i*
5	0,037
7	0,004

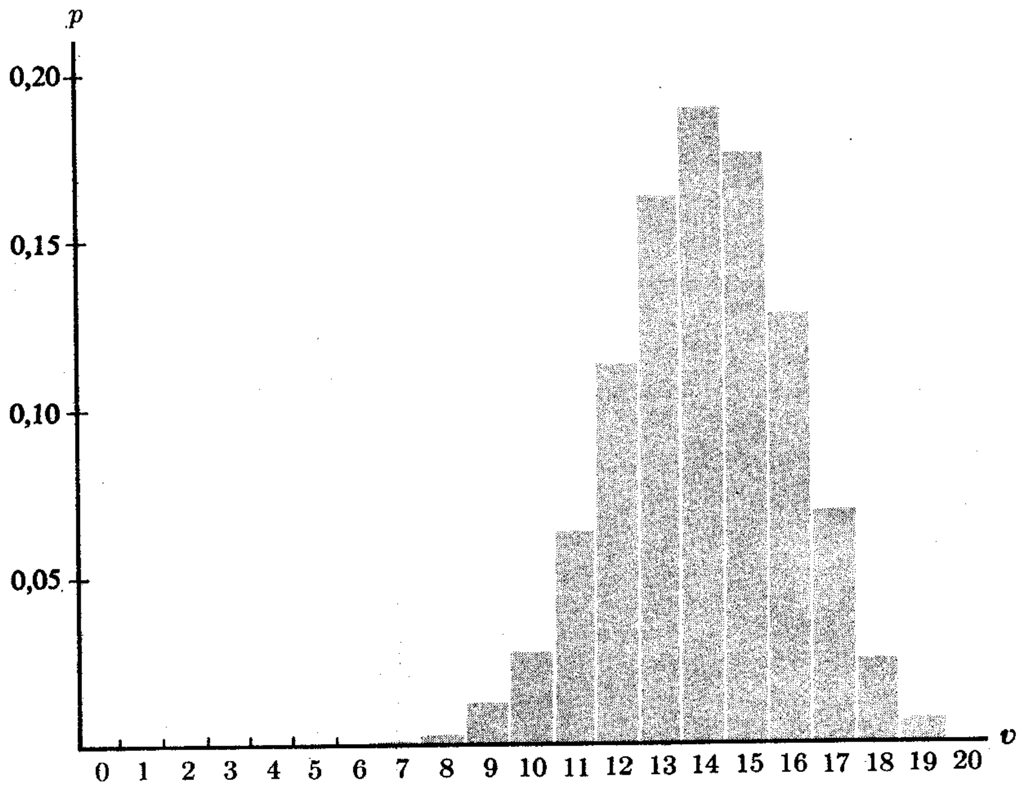


Fig. 11-2

A fig. 11-2 mostra-nos o histograma para uma variável binomial com $n = 20$ e $p = 0,7$. As probabilidades para $n = 20$ e para $n = 0, \ldots, 6$ são, cada uma, menores que 0,001.

Podemos concluir, então, que todos os histogramas para variáveis binomiais em que npq é grande tem, apro-

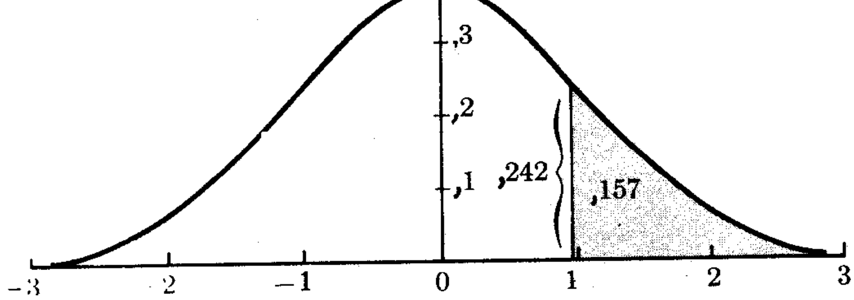


Fig. 11-3

ximadamente, a mesma forma, forma esta de uma curva denominada *curva normal* apresentada na fig. 11-3. A área total sob a curva é 1 e é simétrica em torno do ponto zero.

No final do livro, poderá ser encontrada um atabela da distribuição normal, mostrando as alturas $h(t)$, da curva no ponto t e a área, $H(t)$, sob a curva à direita do ponto t, para vários valores de t.

Uma variável que tenha como densidade a curva normal é denominada *variável normal padrão.* Essa variável tem média 9 e o desvio padrão 1.

Em qualquer histograma, a aproximação normal para a razão

$$\frac{\text{Área do histograma compreendida entre } a \text{ e } b}{\text{Área total sob o histograma}}$$

é a área sob a curva normal compreendida entre t_a e t_b, onde t_x representa o número de desvios padrões pelo qual x excede a média:

$$t_x = \frac{x - m}{\sigma}$$

onde m = média do histograma

σ = desvio padrão do histograma

EXEMPLO 11.1 X é binominal com $n = 6$ e $p = 0{,}4$. Determine a aproximação normal para $P(3 \leq X \leq 5)$.

$P(3 \leq X \leq 5)$ é a parte da área sob o histograma de de X, compreendida entre $a = 2{,}5$ e $b = 5{,}5$. Como $E(X) = 6(0{,}4) = 2{,}4$ e ò(X) = W6(0,4) (0,6) = 1,2,

$$t_a = \frac{2{,}5 - 2{,}4}{1{,}2} = 1/12 = 0{,}083$$

$$t_b = \frac{5{,}5 - 2{,}4}{1{,}2} = 31/12 = 2{,}583$$

Assim, a aproximação normal para $P(3 \leq X \leq 5)$ é a área sob a curva normal compreendida entre 1/12 e 31/12, isto é, $H(1/12)-H(31/12)$ Interpolando-se entre $H(0{,}05) = 0{,}480$ e $H(0{,}10) = 0{,}460$, temos $H(0{,}083) = 0{,}467$. Analogamente, $H(2{,}583) = 0{,}0051$. Então, $P(3 \leq X \leq 5) = 0{,}467 - 0{,}0051 = 0{,}462$

O valor correto de $P|3 \leq X \leq 5|$ é

$0{,}276 + 0{,}138 + 0{,}037 = 0{,}451$

A fig. 11-4 mostra a aproximação normal para $P(3 \leq X \leq 5)$. A escala (x, p) é para o histograma da binomial e a escala (t, h) oara a curva normal. Cono $\sigma(X) = 1{,}2$ e, pois, uma unidade de t corresponde a 1,2 unidade de x, devemos tomar na escala vertical uma unidade p correspondente a 1,2 iguais às áreas unitárias. Note que para cada retângulo, isoladamente, a área sob a parte correspondente da curva é aproximadamente igual a do retângulo.

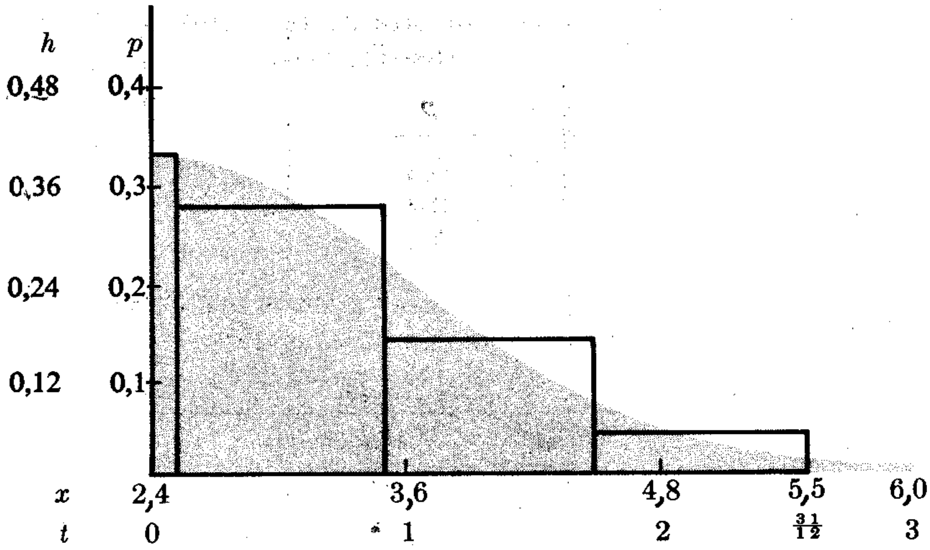


FIG. 11-4

EXEMPLO 11.2 Uma moeda com probabilidade 0,6 de dar cara é lançada 600 vezes. Determine a aproximação normal para a probabilidade de ocorrerem exatamente 372 caras.

Temos:

$$m = 600(0,6) = 360$$

$$\sigma = \sqrt{600(0,6(0,4)} = 12 \qquad a = 371,5 \qquad b = 372,5$$

$$t_a = \frac{371,50 - 360}{12} = \frac{115}{120}$$

$$t_b = \frac{372,50 - 360}{12} = \frac{125}{120}$$

Como t_a e t_b então próximos, podemos aproximar a área $H(t_a) - H(t_b)$ pela área do retângulo cuja base é $t_a - - t_b = 1/12$ e altura $h\left(\frac{t_a + t_b}{2}\right) = h(1) = 0,242$ obtendo:

$$P(372 \text{ caras}) = 1/12(0,242) = 0,0202$$

O valor correto de $P(372 \text{ caras})$ até a quarta casa decimal é 0,0203.

EXEMPLO 11.3 Uma amostra de tamanho 60, $X_1, X_2, \ldots, X_{60}$ é retirada da seguinte distribuição:

v	p
0	0,3
1	0,4
2	0,3

Determine a aproximação normal oara a probabilidade de que a média da amostra

$$X = \frac{X_1 + X^2 + \ldots + X_{60}}{60}$$

seja menor que 0,9.

$$P(X < 0,9) = P(X_1 + \ldots + X_{60}) < 54)$$

$$E(X_1 + \ldots + X_{60}) = 60(X_1) = 60(1) = 60$$

$$\sigma^2(X_1 + \ldots + X_{60}) = \sigma 0 \grave{o}^2(X_1) = 60(0,6) = 36$$

$$\sigma(X_1 + \ldots + X_{60}) = 6$$

Então

$$P(X_1 + \ldots + X_{60} < 54) = 1 - H\left(\frac{53,5 - 60}{6}\right) =$$

$$= 1 - H(\tfrac{13}{12}) = H(\tfrac{13}{12}) = 0,140$$

Existem 14% de chance de que a média da amostra esteja abaixo de 0,9.

EXERCÍCIOS

Lista 11

1. Uma variável X tem média 500, desvio padrão 100 e é, aproximadamente, normal. Calcule a probabilidade de cada um dos seguintes eventos: $X \leq X \leq 600$, $X \leq$ ≤ 800. Determine o número x para o qual $P(X >$ $> x) = 0{,}04$.

2. Desenhe cuidadosamente um gráfico de $H(t)$ pa a $-2{,}5 \leq t \leq 2{,}5$. Poderia você usar este gráfico em vez da tabela de distribuição normal, para responder as questões do problema1?

3. Se 100 bolas são retiradas ao acaso, com reposição, de uma urna, na qual 20% das bolas são pretas, qual a probabilidade de que, pelo menos, 30 das bolas retiradas sejam pretas?

4. Uma urna contém duas bolas vermelhas, uma verde e uma azul. Se 1.000 bolas são retiradas ao acaso, com reposição, qual a probabilidade de que, pelo menos, 300 azuis sejam obtidas? que, pelo menos, 550 vermelhas sejam obtidas?

5. Se X é uma variável normal padronizada, determine a probabilidade de que X pertença a cada um dos intervalos (0; 0,6), (0,6; 1,2), (1,2; 1,8), (1,8; 2,4) e 2,4; 3,0);

   a) usando a tabela da distribuição normal ou *b*) multiplicando o valor de h no ponto médio por 0,6. Use cada uma dessas aproximações para estimar $E(X^2)$, aproximando-se X por uma variável X^*, cujos 10 valores são $+0{,}3$; $+0{,}9$; $+1{,}5$; $+2{,}1$; $+2{,}7$.

6. Um dado equilibrado é arremessado 420 vezes. Determine a aproximação normal para a probabilidade de que a soma dos números produzidos seja, pelo menos, 1.400.

7. Se X é uma binomial com 9 provas e probabilidade de sucesso 0,5, encontre o valor exato de $P(X = 5)$ e sua aproximasão normal.

8. Uma amostra aleatória, com reposição, de $n = 900$ pessoas, é retirada de uma população, na qual a pro-

porção P de pessoas com um certo atributo é 0,7. Determine a aproximação normal para a probabilidade de que a proporção X na nossa amostra esteja compreendida entre 0,68 e 0,72 ou que o número S de pessoas na amostra com o atributo esteja compreendido entre (0,68) (900) = 612 e (0,72) (900) = 648.

9. Repita o problema 8, para cada uma das oito triplas com $n = 225$ ou 900. $p = 0,5$ ou 0,7, $d = 2\%$ ou 5%.

n	p	d	$P(\|X - p\| \leq d)$
225	0,5	0,02	
225	0,5	0,05	
225	0,7	0,02	
225	0,7	0,05	
900	0,5	0,02	
900	0,5	0,05	
900	0,7	0,02	
900	0,7	0,05	

10. Uma moeda honesta é lançada 400 vezes, *a*) calcule a aproximação normal para a probabilidade de exatamente 200 caras; *b*) Determine cotas inferiores e superiores para a probabilidade exata no problema 10 da lista 10.

11. Se X_1 e X_2 são uma amostra da distribuição

v	p
0	p_0
1	p_1
2	p_2
.	.
.	.
.	.
n	p_n

então, $S = X_1 + X_2$ tem a distribuição

v	p
0	p_n^2
1	$p_0p_1 + p_1p_0$
2	$p_0p_2 + p_1p_1 + p_2p_0$
3	$p_0p_3 + p_1p_2 + p_2p_1 + p_3p_0$
.	
.	
.	
n	$p_0p_n + p_1p_{n-1} + \ldots + p_np_0$
$n+1$	$p_1p_n + p_2p_{n-1} + \ldots + p_np_1$
$n+2$	$p_2p_n + p_3p_{n-1} + \ldots + p_np_2$
.	
.	
.	
$2n$	p_{n1}

a) Explique $P(S = 3) = p_0p_3 + P_1p_2 + p_2p_1 + p_3p_0$.

b) Determine a distribuição de S, quando a amostra é retirada da distribuição

v	p
0	$\frac{1}{4}$
1	$\frac{1}{4}$
2	$\frac{1}{4}$
3	$\frac{1}{4}$

c) Determine a distribuição de $T = S_1 + S_2$, onde S_1 e S_2 são uma amostra da distribuição de S em (*b*), Por que a distribuição de T é igual à distribuição de $X_1 + X_2 + X_3 + X_4$, onde X_1, X_2, X_3, X_4 são uma amostra da distribuição dada em (*b*)? Construa o histograma para T. Calcule a aproximação normal para $P(T = v)$, para cada valor de v de T.

12 INFERÊNCIA

Em um lago, há um pequeno número, A, de peixes. Você sabe que $A = 1$, 2 ou 3 e atribui a cada um desses valores, probabilidades 0,2, 0,2 e 0,6, respectivamente. Você pega ao acaso um peixe do lago, marca-o de alguma forma, repondo-o no lago, em seguida. No dia seguinte, pega novamente ao acaso, um peixe, verifica se é ou não o marcado e repõe no lago. Agora, você terá de decidir se o número de peixes no lago é ou não igual a 3, ganhando um prêmio se tomar a decisão correta. Qual o melhor método de decisão e em quanto boa é a sua chance de ganhar o prêmio?

Nosso primeiro passo na solução do problema é representar os dados disponíveis em uma tabela que denominaremos *tabela de dados*:

possíveis valores da variável = PARÂMETRO

TABELA DE DADOS

distribuição a priori → probab. antes do experimento

λ	A	X (marcado) T	N (não marcado)
0,2	1	1	0
0,2	2	$\frac{1}{2}$	$\frac{1}{2}$
0,6	3	$\frac{1}{3}$	$\frac{2}{3}$

A coluna A apresenta os possíveis valores da variável A (número de peixes no lago) e a coluna λ as probabilidades atribuídas a cada um desses valores, antes do experimento ser realizado. A variável A, cujo valor é de nosso interesse e da qual se espera reconhecer alguma coisa através do experimento, é chamada o *parâmetro* do problema e a distribuição λ, que representa nossa opinião sobre A antes da realização do experimento, é chamada *distribuição a priori* do parâmetro. A variável X representa o *resultado* do experimento. Este tem 2 resultados possíveis: T, o peixe apanhado no segundo dia foi o peixe marcado, ou N, não foi o peixe marcado. Os números na coluna T mostram a probabilidade do resultado T para cada valor do parâmetro; se $A = 2$, a probabili6ade de pegarmos o peixe marcado é $\frac{1}{2}$ e assim por diante. Os números na coluna N mostram a probabilidade do resultado N para cada valor do parâmetro. A função $p(x\,|a)$, que especifica, para cada resultado x e o valor a do parâmetro, a probabilidade de observarmos o resultado x, dado que o valor do parâmetro é a, é denominado *modelo* para o experimento. Assim, nossa tabela de dados nos mostra duas coisas: (1) a distribuição *a priori* λ e (2) o modelo p.

O segundo passo é construir a *tabela da distribuição conjunta*, que nos dá, para cada valor a do parâmetro e cada resultado x, a probabilidade de ocorrência do par (a, x). Como $P(A = a \text{ e } X = x) = P(A = a) . P(X = x \mid A = a) = \lambda(a)p(x \mid a)$, obtemos a tabela da distribuição conjunta, multiplicando-se cada $p(x \mid a)$ na tabela de dados por $\lambda(a)$:

TABELA DE DISTRIBUIÇÃO CONJUNTA

A	X		Σ
	T	N	
1	0,2	0	0,2
2	0,1	0,1	0,2
3	0,2	0,4	0,6
Σ	0,5	0,5	

Por exemplo: $P(A = 3 \text{ e } X = N) = P(A = 3)P(X = N \mid A = 3) = (0{,}6)(\frac{2}{3}) = 0{,}4$

A linha Σ, que é a soma das linhas do parâmetro, dá-nos a distribuição (*marginal*) de X:

$$P(X = T) = 0{,}5 \quad \text{e} \quad P(X = N) = 0{,}5$$

Analogamente, a coluna Σ, a soma da coluna dos resultados, dá-nos a distribuição de A, que sabíamos ser λ.

O terceiro passo é construir a *tabela da distribuição a posteriori*, que nos dá, para cada resultado x do experimento, a distribuição que atribuímos a A, quando aquele resultado for observado.

Como

$$P(A = a \mid X = x) = \frac{P(A = a \text{ e } X = x)}{P(X = x)},$$

obtemos a tabela da distribuição *a posteriori* dividindo cada entrada na tabela da distribuição conjunta pelo número Σ ao pé da coluna correspondente:

TABELA DA DISTRIBUIÇÃO A POSTERIORI

A	X	
	T	N
1	0,4	0
2	0,2	0,2
3	0,4	0,8

Por exemplo,

$$P(A = 3 \mid X = N) = \frac{P(A = 3 \text{ e } X = N)}{P(X = N)} = \frac{0,4}{0,5} = 0,8$$

Podemos, agora, resolver nosso problema: Deve você decidir que $A = 3$? Como $P(A = 3 \mid T) = 0,4 < 0,5$, se você pegar o peixe marcado, a melhor decisão é $A \neq 3$ e sua chance de estar decidindo corretamente é 0,6. Mas, se pegar o peixe não marcado, dado que $P(A = 3 \mid N) = 0,8 >$ $> 0,5$, você deve decidir que $A = 3$ com chance 0,8 de estar decidindo certo.

Aqui estão dois métodos para você encontrar a chance total de estar decidindo corretamente:

Método 1

Dos seis pares (a, x) que devem ocorrer, aqueles que o fazem ganhar o prêmio (usando seu critério de decisão) são $(1, T)$, $(2, T)$ e $(3, N)$. As probabilidades desses pares, de acordo com a tabela da distribuição conjunta, são 0,2, 0,1 e 0,4 de tal forma que a probabilidade de você estar tomando a decisão correta é

$$0,2 + 0,1 + 0,4 = 0,7$$

Método 2

$$\begin{aligned} P(\text{prêmio}) &= P(T \text{ e prêmio}) + P(N \text{ e prêmio}) \\ &= P(T)P(\text{prêmio} \mid T) + P(N)P(\text{prêmio} \mid N) \\ &= 0,5(0,6) + 0,5(0,8) = 0,7 \end{aligned}$$

Usando seu critério de decisão, qual a probabilidade de que você ganhe o prêmio, se $A = 1$? Se $A = 2$? Se $A = 3$? (Respostas 1, $\frac{1}{2}$ e $\frac{2}{3}$).

Sumário

Se T ocorre, decida $A \neq 3$; P(prêmio) $= 0,6$
Se N ocorre, decida $A = 3$, P(prêmio) $= 0,8$
Probabilidade total do prêmio $= 0,7$.

Suponha que, em vez de ter de decidir se $A = 3$, tenhamos que estimar A, perdendo o quadrado de nosso erro. Qual a melhor estimativa para A e qual é seu emq? Se T ocorre, nossa melhor estimativa para A é

$$E(A \mid T) = (0,4)(1) + (0,2)(2) + (0,4)(3) = 2,0$$

e nosso emq é

$$\begin{aligned}\sigma^2(A \mid T) &= E(A^2 \mid T) - (E(A \mid T))^2 \\ &= (0,4)(1)^2 + (0,2)(2)^2 + (0,4)(3)^2 - (2,0)^2 \\ &= 4,8 - 4,0 = 0,8\end{aligned}$$

Se N ocorre, sua melhor estimativa para A e $E(A \mid N) = 2,8$ e $\sigma^2(A \mid N) = 8 - (2,8)^2 = 0,16$.

Aqui estão dois métodos para calcular o emq total.

Método 1

A tabela a seguir dá-nos o erro quadrático para os seis pares (a, x):

A	X	
	T	N
1	$(1 - 2)^2 = 1$	$(1 - 2,8)^2 = 3,24$
2	$(2 - 2)^2 = 0$	$(2 - 2,8)^2 = 0,64$
3	$(3 - 2)^2 = 1$	$(3 - 2,8)^2 = 0,04$

Multiplique cada erro quadrático pela probabilidade do par correspondente, obtendo:

$$\text{emq total} = 1(0{,}2) + 0(0{,}1) + 1(0{,}2) + (3{,}24)(0) + \\ + (0{,}64)(0{,}1) + (0{,}04)(0{,}4) = 0{,}48$$

Método 2

$$\text{emq total} = P(T)(\text{emq}\,|T) + P(N)(\text{emq}\,|N) = \\ = (0{,}5)(0{,}8) + (0{,}5)(0{,}16) = 0{,}48$$

Seu emq, como função do número de peixes no lago, é:

$$(\text{emq}\,|A = 1) = 1(1-2)^2 + D(1-2{,}8)^2 = 1$$
$$(\text{emq}\,|A = 2) = \tfrac{1}{2}(2-2)^2 + \tfrac{1}{2}(2-2{,}8)^2 = 0{,}32$$
$$(\text{emq}\,|A = 3) = \tfrac{1}{3}(3-2)^2 + \tfrac{2}{3}(3-2{,}8)^2 = 0{,}36$$

Sumário

Se T ocorre, estime $A = 2$; emq $= 0{,}8$
Se N ocorre, estime $A = 2{,}8$; emq $= 0{,}16$
emq total $= 0{,}48$

Observações

1. Se tivermos de decidir se $A = 3$ antes de pescar, nossa melhor decisão é $A = 3$ e temos uma chance de 0,6 de estarmos decidindo certo. O experimento de pescar um peixe aumenta esta probabilidade para 0,7.

2. Se tivermos de estimar A antes de pescar, nossa melhor estimativa é $E(A) = 2{,}4$ e o emq é $\sigma^2(A) = 0{,}64$. O experimento de pescar um peixe reduz o emq para 0,48, de tal forma que a melhor estimativa a partir do experimento tem valor igual a $\dfrac{0{,}64 - 0{,}48}{0{,}64} = 0{,}25$.

O resultado T aumenta realmente o emq para 0,8. Estamos mais incertos acerca de A, após a realização do experimento, do que antes de realizá-lo.

EXERCÍCIOS

Lista 12

1. Nos dados apresentados na tabela que segue, você deve observar X e decidir qual o valor correto de A, ganhando

um prêmio se tomar a decisão correta. Qual o melhor resultado e quais suas chances totais de ganhar o prêmio?

λ	A	X	
		B	W
0,3	a	0,6	0,4
0,7	b	0,9	0,1

2. Descreva um experimento de urna, com os dados tabelados do exercício 1.

3. Resolva o exercício 1, com a coluna 1 substituída por

   a) $\lambda(a) = 0{,}2$ e $\lambda(b) = 0{,}8$

   b) $\lambda(a) = 0{,}1$ e $\lambda(b) = 0{,}9$

4. A proporção p de estudantes graduados, numa certa classe, pode ser 0,1, 0,2 ou 0,5 e essas proporções são igualmente prováveis. Você selecionará uma amostra casual, com reposição, de dois indivíduos da classe e registrará seu *status* (G ou U). Continuará a tabela de dados para este experimento. Determine a melhor estimativa para p, quando o resultado do esperimento é GU, e calcule o emq para este resultado.

5. Suponha, no exercício 4, que alguém seleciona a amostra e lhe informa o número X de G na amostra. Construa a tabela de dados para este esperimento, determine o melhor estimador para p, quando $X = 1$.

6. A urna 1 tem duas bolas pretas e uma bola branca; a urna 2 tem uma bola preta e duas bolas brancas. Uma das urnas é selecionada ao acaso e, em seguida, selecionadas, ao acaso, duas bolas dessa urna. Você será informado do total de bolas pretas retiradas e deverá decidir qual a urna selecionada. Construa a tabela de dados para este problema e determine o melhor método

de decisão e sua chance total de estar decidindo corretamente, (*a*) com reposição e (*b*) sem reposição.

7. Uma das palavras da frase This is it é selecionada ao acaso e, em seguida, desta palavra é selecionada uma letra, que lhe é apresentada. Você deverá, então, adivinhar que palavra foi selecionada, ganhando um prêmio se estiver correto. Construa a tabela de dados, da distribuição conjunta e da distribuição *a posteriori*. Determine o melhor método de decisão, calcule sua chance de ganhar o prêmio para cada possível resultado e sua chance total de receber o prêmio. Usando seu método de decisão, qual sua chance de ganhar esse prêmio, se a palavra selecionada é THIS? IS? IT?

8. Um ponto do diagrama de dispersão da fig. 12-1 é selecionado ao acaso. Você observa X e deve estimar

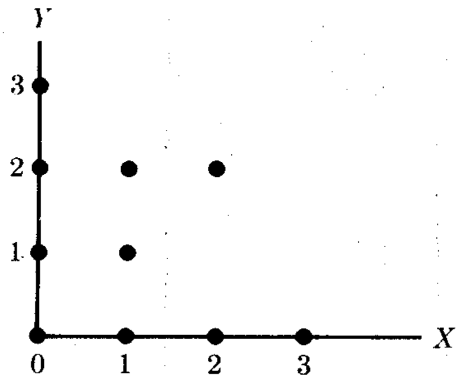


Fig. 12-1

Y, perdendo o quadrado de seu erro. Construa a tabela de dados, da distribuição conjunta, da distribuição *a posteriori*, determine seu melhor método de decisão e calcule seu emq total.

9. Uma urna contém n bolas numeradas 1, 2, ..., n. Você sabe que $n = 3$, 4 ou 5 e considera esses valores igualmente prováveis. Retira, ao acaso, uma bola da urna. Qual a distribuição *a posteriori* de n, se você retira a bola número 1? 2? 3? 4? 5?

13 INFERÊNCIA SOBRE PROPORÇÕES (I)

Suponha que você deve decidir se a proporção de fumantes, num certa população, ultrapassa 25%. Você seleciona uma amostra casual de 100 pessoas e encontra 80 fumantes. Qual deve ser a sua decisão e qual a chance que tem de estar correto?

O parâmetro A de interesse, a proporção de fumantes na população, tem um grande número de possíveis valores: Se a população tem 10.000 elementos, qualquer um dos 10.001 números $0, \frac{1}{10.000}, \frac{2}{10.000}, \ldots, 1$ é um possível valor. Além disso, geralmente não teremos conhecimento acerca do número exato de elementos da população. Imaginaremos que *todos* os números compreendidos entre 0 e 1 são possíveis valores de A e descrevemos nossa opinião *a priori* sobre A, através de uma função de densidade f, definida para $0 \leq x \leq 1$ denominada *densidade a priori* de A.

Qualquer densidade sob (0, 1) é uma possível densidade *a priori*. Tomamos conhecimento de diversas densidades sob (0, 1), no cap. 5.

A relação entre a densidade *a priori* de uma proporção sua densidade *a posteriori*, dada a amostra, é muito simples:

Se f é a densidade *a priori* de uma proporção e uma amostra casual apresenta u indivíduos com a característica e v sem ela, a densidade *a posteriori* da proporção é

$$g(x) = f(x)x^u(1-x)^v$$

Por exemplo, no problema dos fumantes, se nossa opinião *a priori* sobre A é uniforme em $(0, \frac{1}{2})$, isto é, $f(x)=1$ para $0 \leq x \leq \frac{1}{2}$ e $f(x) = 0$ para $x > \frac{1}{2}$, a distribuição *a posteriori* para A tem densidade

$$g(x) = \begin{cases} x^{30}(1-x)^{70} & 0 \leq x \leq \frac{1}{2} \\ 0 & \frac{1}{2} < x \leq 1 \end{cases}$$

Temos de construir o gráfico de g e determinar que proporção da área total sobre g está compreendida entre 0 e 0,25. Isto é bastante complicado, mas podemos trabalhar com amostras bem pequenas. Por exemplo, se selecionarmos um indivíduo e ele não for fumante, teremos $u = 0$ e $v = 1$, de tal forma que a densidade *a posteriori* de A é

$$g(x) = \begin{cases} (1-x) & 0 \leq x \leq \frac{1}{2} \\ 0 & \frac{1}{2} < x \leq 1 \end{cases}$$

Como podemos ver na fig. 13-1, a área total sobre g é

$$\frac{1}{2}\left(\frac{1+\frac{1}{2}}{2}\right) = \frac{3}{8}$$

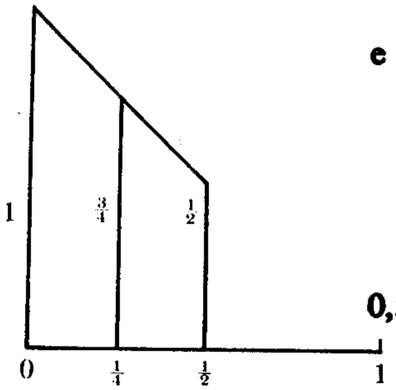


Fig. 13-1

e a área compreendida entre 0 e $\frac{1}{4}$ é

$$\frac{1}{4}\left(\frac{1+\frac{3}{4}}{2}\right)=\frac{7}{32}.$$

Logo, $P(0 \leq A \leq \frac{1}{4}\,|\text{amostra}) = \frac{7}{32}/\frac{3}{8} = \frac{7}{12} = 0{,}57$.

A probabilidade de que $A \leq \frac{1}{4}$ aumentou de 0,5 para 0,57

PROBLEMA PARA DISCUSSÃO 13.1

Calcule $P(0 \leq A \leq \frac{1}{4}\,|\text{amostra})$, se o indivíduo na amostra for um fumante.

(Resposta: $\frac{1}{4}$)

Se tivéssemos selecionado uma amostra de cinco pessoas e encontrado dois fumantes, isto é, $u = 2$ e $v = 3$, teríamos:

$$g(x) = \begin{cases} x^2(1-x)^3 & 0 \leq x \leq \frac{1}{2} \\ 0 & x > \frac{1}{2} \end{cases}$$

Calculamos alguns valores para g:

x	x^2	$(1-x)^3$	g
0	0	1	0
0,1	0,01	0,729	0,00729
0,2	0,04	0,512	0,02048
0,3	0,09	0,343	0,03087
0,4	0,16	0,216	0,03456
0,5	0,25	0,125	0,03125

Construímos o gráfico para g (ou um múltiplo qualquer, que seja conveniente, digamos 1.000 g) e estimamos as áreas sobre g em $(0, \frac{1}{4})$ e $(\frac{1}{4}, \frac{1}{2})$. Como podemos observar na fig. 13-2, a área compreendida entre 0 e $\frac{1}{4}$ parece igual à área sobre a linha pontilhada, a uma altura de 12, e a àrea compreendida entre $\frac{1}{4}$ e $\frac{1}{2}$ parece igual à área sobre à área sobre a linha pontilhada, a uma altura de 32, assim estimamos:

$$P(A \geq 0{,}25\,|\text{amostra}) = \frac{32(0{,}25)}{32(0{,}25) + 12(0{,}25)} = \frac{8}{11} = 0{,}73$$

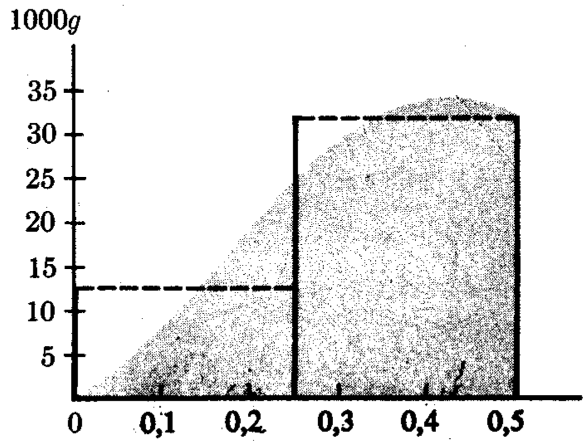


Fig. 13-2

Se todas as pessoas de nossa amostra tivessem sido fumantes, isto é, $u = 5$ e $v = 0$, teríamos, conforme o gráfico da fig. 13-3,

$$g(x) = \begin{cases} x^5 & \text{para } 0 \leq x \leq \frac{1}{2} \\ 0 & \text{para } x > \frac{1}{2} \end{cases}$$

A área compreendida entre $\frac{1}{4}$ e $\frac{1}{2}$ é, aproximadamente, $12(0{,}25) = 3$. A área compreendida entre 0 e $\frac{1}{4}$ é muito pequena para ser estimada pelo gráfico.

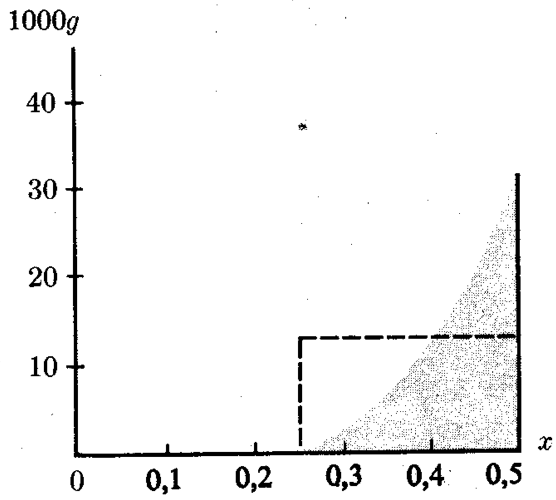


Fig. 13-3

Vamos ignorar a área realmente desprezível, abaixo de 0,2 e calcular

$$1.000g(0{,}2) = 1.000(0{,}2)^5 = 0{,}32$$
$$1.000g(0{,}25) = 1.000(0{,}25)^5 = 0{,}98$$

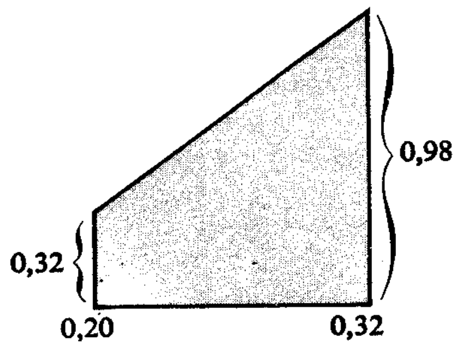


Fig. 13-4

e estimar a área compreendida entre 0,2 e 0,25 como sendo a área do trapézio da fig. 13-4, que é igual a

$$(0{,}05)\left(\frac{0{,}32 + 0{,}98}{2}\right) = 0{,}0325.$$

Assim, estimamos

$$P(A \geq 0{,}25 \,|\text{amostra}) = 3/(3{,}0325) = 0{,}99$$

Mesmo uma amostra de tamanho cinco pode proporcionar uma grande evidência. (O verdadeiro valor de $P(A \geq 0{,}25 \,|\text{amostra})$ é $\frac{63}{64} = 0{,}985$).

EXERCÍCIOS

Lista 13

1. Para a densidade *a priori* $f(x) = x^2(1 - x)$ e uma amostra com $u = 1$ e $v = 2$, faça um gráfico da densidade *a posteriori* de A e estime $\sigma^2(A \,|\text{amostra})$.

2. A área sobre o gráfico de x^n, compreendida entre 0 e t, é igual a $\dfrac{t^{n+1}}{n+1}$. Verifique este fato para os seguintes valores: *a*) $t = \frac{1}{2}$, $n = 0$; *b*) $t = \frac{1}{2}$, $n = 1$; *c*) $t = \frac{1}{2}$, $n = 2$ (aproximadamente).

3. Use o fato apresentado no exercício 2 para mostrar que, se A tem uma distribuição, *a priori*, uniforme em $(0, \frac{1}{2})$ e uma amostra casual de tamanho cinco apresenta cinco sucessos, então,

$$P(A \geq 0{,}25 \mid \text{amostra}) = \tfrac{63}{64}$$

4. Uma amostra de três só tem sucesso. Se a tem distribuição *a priori*, uniforme, use o fato apresentado no exercício 2, para calcular os números

$$P_1 = P(0 \leq A \leq 0{,}4 \mid \text{amostra})$$
$$P_2 = P(0{,}4 \leq A \leq 0{,}8 \mid \text{amostra})$$
$$P_3 = P(0{,}8 \leq A \leq 1 \mid \text{amostra})$$

Calcule a média da variável aproximante A^*, cuja distribuição é

v	p
0,2	P_1
0,6	P_2
0,9	P_2

5. Uma proporção A tem um dos 11 valores: 0; 0,1; 0,2; . . .; 0,9; 1 e estes valores são, *a priori*, igualmente prováveis. Uma amostra casual da população apresenta u sucessos e v falhas. Construa o histograma *a posteriori*, para A, das seguintes amostras:

a) $u = 1, \quad v = 0$
b) $u = 1, \quad v = 1$
c) $u = 2, \quad v = 0$

14 INFERÊNCIA ESTATÍSTICA (II)

Se você inicia com uma densidade, *a priori*, uniforme em (0, 1) para uma proporção A, e uma amostra casual lhe dá u indivíduos com uma certa característica e v sem ela, a densidade, *a posteriori*, resultante para A é

$$x^u(1 - x)^v$$

Chamaremos esta densidade de *densidade* (u, v).

A fig. 14-1 apresenta densidades para alguns (u, v). As três primeiras densidades (primeira linha) têm $u = v$. Elas são simétricas em torno do ponto 0,5, mas tornam-se mais estreitas à medida que u cresce: Quanto maior a amostra com $u = v$, mais é provável que A esteja próximo de 0,5. As três da linha seguinte têm $v = 2u$: duas vezes

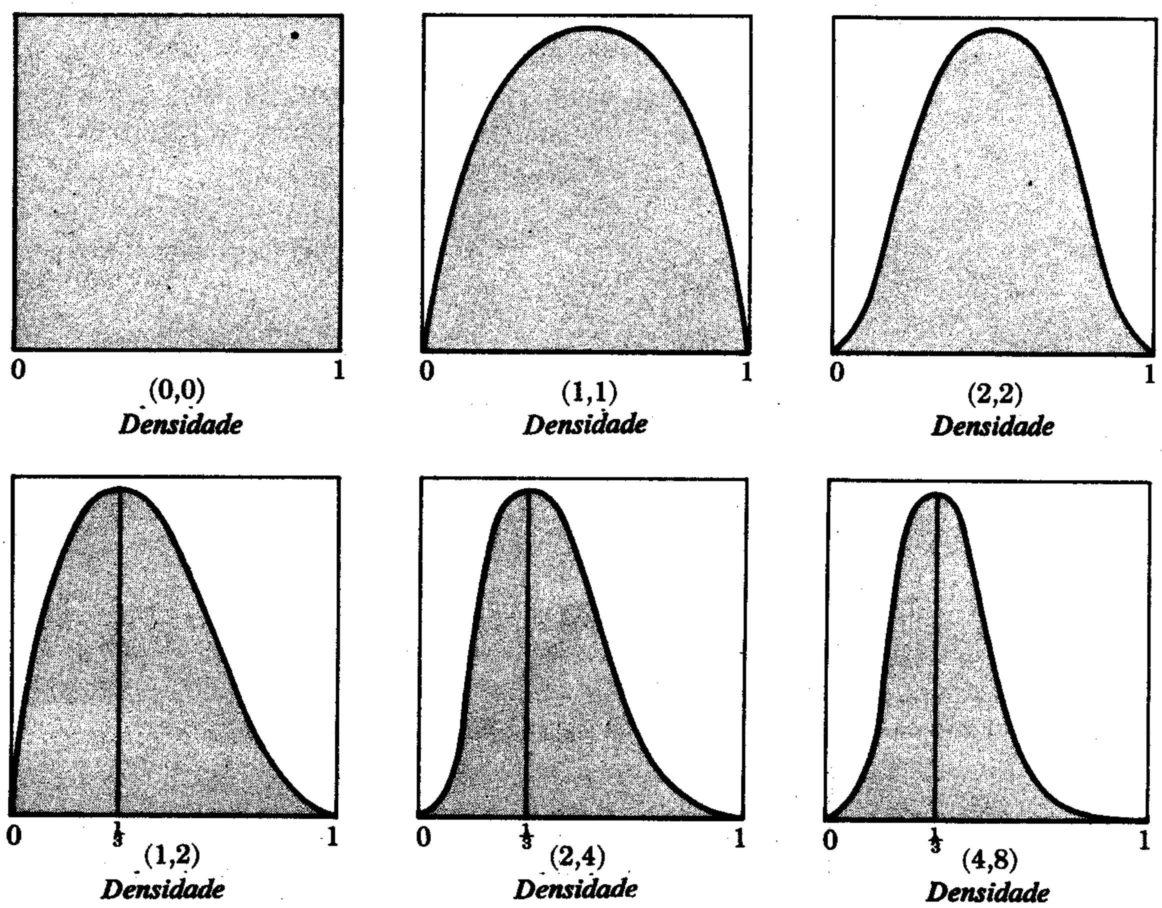


Fig. 14-1

mais falhas que sucessos. Todas elas apresentam um máximo no ponto $\frac{1}{3}$ e se tornam cada vez mais estreitas, à medida que u cresce: quanto maior a amostra com $v = 2u$, é mais provável que A esteja próximo de $\frac{1}{3}$.

EXEMPLO 14.1 Suponha que sua densidade *a priori* para uma proporção A é uniforme, isto é, a densidade é (0, 0). Você seleciona uma amostra de tamanho 100 e obtém 30 sucessos. Qual a sua melhor estimativa para A? Qual a probabilidade P de que o verdadeiro valor de A não difira em mais de 5% de sua estimativa?

Como sua densidade *a posteriori* é a densidade (30, 70), sua melhor estimativa é a média m da densidade (30, 70) e P é a probabilidade de que uma variável com densidade (30, 70) não difira de sua média em mais de 5%.

Para uma densidade (u, v) como esta, a aproximação normal será boa e, assim, tudo o de que precisamos é da média e da variância da densidade.

Se A tem densidade (u, v),

$$E(A) = \frac{u+1}{u+v+2}$$

$$\sigma^2(A) = \frac{(u+1)(v+1)}{(u+v+2)^2(u+v+3)}$$

Para valores grandes de u e v, temos as seguintes fórmulas aproximadas:

$$E(A) = \frac{u}{u+v} = p$$

$$\sigma^2(A) = \frac{u}{(u+v)^2} = \frac{pq}{n}$$

onde $p = u/(u+v)$, $q = 1 - p$, e $n = u + v$

Assim, se A tem densidade (30, 70), aproximadamente,

$$E(A) = \frac{30}{30+70} = 0{,}3$$

$$\sigma^2(A) = \frac{(0{,}3)(0{,}7)}{100} = 0{,}0021$$

$$\sigma(A) = \sqrt{0{,}0021} = 0{,}046.$$

$P(A$ não difira de 0,3 por mais de 5%$) = P(0{,}25 \leq A \leq 0{,}35) =$

$$= H\left(\frac{0{,}25 - 0{,}3}{0{,}046}\right) - H\left(\frac{0{,}35 - 0{,}3}{0{,}046}\right)$$

$$= H(-1{,}09) - H(1{,}09) = 1 - 2H(1{,}09)$$

$$= 1 - 2(0{,}138) = 0{,}724$$

Sua melhor estimativa para A é 0,30, a proporção de sucessos na amostra. Seu emq é $\sigma^2(A) = 0{,}0021$ e a chance que você tem de não se afastar do verdadeiro valor de A por mais de 5% é de cerca de 72%.

Que diferença faz o fato de você começar com uma ou com outra opinião *a priori*? No nosso exemplo, se você tivesse iniciado com a densidade (4, 1), sua densidade *a posteriori*, observada uma amostra (30, 70), seria

$$g(x) = x^4(1 - x)x^{30}(1 - x)^{70} = x^{34}(1 - x)^{71}$$

ou, em outras palavras, a densidade (34, 71). Sua estimativa para A seria

$$E(A) = \frac{34}{105} = 0{,}324 \qquad \text{e}$$

$$\sigma^2(A) = \frac{(0{,}324)(0{,}674)}{105} = 0{,}0021 \qquad \sigma(A) = 0{,}046$$

$$P(A \text{ não difira de } 0{,}324 \text{ por mais de } 5\%) = 0{,}724$$

Este exemplo ilustra dois fatos:

1. Se sua opinião *a priori* está razoavelmente bem distribuída e você tem uma amostra razoavelmente grande, sua opinião *a posteriori* será praticamente independente de sua opinião *a priori*.

2. Se sua opinião *a priori* tem densidade (r, s) e você observa uma amostra (u, v), sua opinião *a posteriori* terá densidade $(r + u, v + s)$.

O segundo destes fatos tem uma interpretação natural. Se você inicia com uma opinião *a priori* uniforme sobre A, isto é, com densidade (0, 0) e observa uma amostra (r, s), sua opinião *a posteriori* terá, então, densidade (r, s). Se você, agora, numa segunda amostra, observa (u, v), sua opinião final pode ser calculada tanto como o resultado da observação de uma amostra (u, v) segundo uma opinião *a priori* com densidade (r, s) quanto (agrupando as duas amostras) como o resultado da observação de uma amostra $(r + u, s + v)$, segundo uma opinião *a priori* com densidade (0, 0). O segundo fato diz que os dois métodos de cálculo produzem os mesmos resultados.

EXEMPLO 14.2 Um senhor está fortemente convencido de que a proporção A, nascimentos do sexo masculino numa certa população, é aproximadamente 0,5 e atribui a A uma distribuição

(500, 500). Qual será sua estimativa para A (*a*) após observar uma amostra de 100 nascimentos, dos quais 52 são do sexo masculino? (*b*) após observar uma amostra de 10.000 nascimentos, dos quais 5.200 são do sexo masculino?

Para (*a*):

$$E(A) = \frac{500 + 52}{(500 + 52) + (500 + 48)} = \frac{552}{1.100} = 0{,}502$$

Para (*b*):

$$E(A) = \frac{500 + 5.200}{(500 + 5.200) + (500 + 4.800)}$$

$$= \frac{5.700}{11.000} = 0{,}518$$

Após observar uma amostra do tamanho 100, sua opinião *a priori* ainda predomina, mas uma amostra de tamanho 10.000 a ela se sobrepõe.

EXERCÍCIOS

Lista 14

1. *a*) Determine uma densidade (u, v) que, em sua escola, aproxima sua opinião acerca da proporção A de alunos, cujo sobrenome começa com as letras de de A a F.

b) Da maneira mais simples que tiver ao seu alcance, selecione de sua escola uma amostra (com tamanho entre 20 e 100) de estudantes; registre o número x de alunos em sua amostra cujos nomes iniciam com as letras A a F e o número y de alunos cujos nomes iniciam com as letras de G a Z. Uma forma é pegar um exemplar do jornal da escola e tratar o nome dos alunos, que nele aparecem, como uma amostra casual.

c) Calcule a média m e o desvio padrão σ de sua distribuição *a posteriori* para A.

d) Agora, estime o valor de A com a maior precisão possível, com os dados do Diretório Acadêmico. Verifique a posição deste valor de A em relação à sua distribuição *a posteriori* ou qual o valor de $(A - M)/\sigma$.

2. Repita o exercício 1, mas substitua A a F por A a L, estudantes por moradores de sua cidade, o jornal da escola pelo jornal da cidade (escolha nomes ao acaso em lugar de usar todos) e o diretório pela lista telefônica.

3. Considere a população cujos elementos são seqüência de cinco dígitos: 00000, 00001, ..., 99999. Entre os 100.000 elementos dessa população, uma certa proporção A tem a propriedade de que todos os cinco dígitos são diferentes. Por exemplo, 21387 têm esta propriedade e 10207, não. Comece com uma opinião de que a densidade *a priori* é uniforme sobre A e use uma tabela de números aleatórios para selecionar uma amostra dessa população, proceda continuamente até que sua distribuição *a posteriori* tenha $\sigma \leq 0{,}05$, de forma que a chance de que a sua estimativa não difira do verdadeiro valor de A em mais de 5% seja 68% ($H(-1) - H(1) = 0{,}68$). Calcule, agora, o verdadeiro valor de A.

4. Estime (por amostragem) a proporção A de estudantes canhotos em sua escola. Calcule $P(A \geq 0{,}2 \mid \text{amostra})$.

15 PROPORÇÕES INDEPENDENTES

Se A_1 e A_2 são proporções independentes, isto é, o conhecimento de A_1 não diz nada a respeito de A_2 e c_1 e c_2 constantes quaisquer, então $X = c_1A_1 + c_2A_2$ tem média e variância

$$m = c_1m_1 + c_2m_2$$

$$\sigma^2 = c_1^2\sigma_1^2 + c_2^2\sigma_2^2$$

onde m_i e σ_i^2 são a média e a variância de A_i.

Dois casos interessantes são

a) $c_1 = 1$ e $c_2 = 1$, neste caso $X = A_1 - A_2$ é a *diferença entre as duas proporções*;

b) $c_1 \geq 0$, $c_2 \geq 0$, $c_1 + c_2 = 1$, neste caso X é a *média ponderada entre as duas proporções*.

EXEMPLO 15.1 No tratamento da hilose, os efeitos de duas drogas devem ser testados. Estamos interessados na proporção A_i das pessoas portadoras de hilose, que melhoram quando lhes é aplicada a droga i. Antes de realizarmos o experimento, A_1 e A_2 são consideradas independentes e uniformes no intervalo (0,1). Selecionamos, ao acaso, 100 portadores e aplicamos a droga 1; 20 deles apresentam melhoras. Selecionamos 300 portadores e lhes aplicamos a droga 2; 45 deles apresentam melhoras. Qual a probabilidade de que a droga 1 seja melhor que a 2 ou que $A_1 - A_2 > 0$? Dos dados, segue-se que

$$m_1 = \frac{20}{100} = 0{,}2 \qquad \sigma_1^2 = \frac{(0{,}2)\,(0{,}8)}{100} = 0{,}0016$$

$$m_2 = \frac{45}{300} = 0{,}15 \qquad \sigma_2^2 = \frac{(0{,}15)\,(0{,}85)}{300} = 0{,}000425$$

Então,

$$X = A_1 - A_2$$
$$m = 0{,}2 - 0{,}15 = 0{,}05$$
$$\sigma^2 = 0{,}0016 + 0{,}000425 = 0{,}002025$$
$$\sigma = 0{,}045$$

A aproximação normal dá-nos

$$P(X > 0) = H\left(\frac{0 - 0{,}05}{0{,}045}\right) = 1 - H(1{,}11) = 1 - 0{,}134$$
$$= 0{,}866$$

Existe, aproximadamente, uma chance de 87% de que a droga 1 seja melhor que a droga 2.

EXEMPLO 15.2 Queremos estimar a proporção A de fumantes de uma população, na qual 30% das pessoas são do sexo feminino.

Consideremos as proporções A_1 e A_2 de fumantes do sexo feminino e do sexo masculino, respectivamente, independentes, com densidades *a priori* do tipo (0,0). Calcule a média e o desvio padrão da proporção A de fumantes na população.

Apresentamos, a seguir, em uma tabela de dupla entrada, os resultados de nosso amostra

	Fumantes	Não Fumantes	
Masc.	120	180	300
Femin.	100	100	200
	220	280	

$$m_1 = \frac{100}{200} = 0,5 \qquad \sigma_1^2 = \frac{(0,5)(0,5)}{200} = 0,00125$$

$$m_2 = \frac{120}{300} = 0,4 \qquad \sigma_2^2 = \frac{(0,4)(0,6)}{300} = 0,0008$$

Como $A = 0,3\ A_1 + 0,7\ A_2$ a média e o desvio padrão de A são, respectivamente,

$$m\ (0,3)\ (0,5) + (0,7)\ (0,4) = 0,43$$

$$\sigma^2 = (0,3)^2\ (0,00125) + (0,7)^2 (0,0008) = 0,0005045$$

$$\sigma = \sqrt{0,0005045} = 0,0225$$

Suponha que, no exemplo das drogas, tivéssemos aplicado cada uma delas a 200 pacientes e obtido as mesmas proporções de sucesso: 20% de melhoras para a droga 1 e 15% para a droga 2. Neste caso,

$$m_1 = 0,2 \qquad \sigma_1^2 = \frac{(0.2)(0,5)}{200} = 0,0008$$

$$m_2 = 0,15 \qquad \sigma_2^2 = \frac{(0,15)(0,85)}{200} = 0,000638$$

Logo,

$X = A_1 - A_2$

tem

$m = 0{,}2 - 0{,}15 = 0{,}05$ como antes

$\sigma^2 = 0{,}0008 + 0{,}000638 = 0{,}001438$

$\sigma = \sqrt{0{,}001438} = 0{,}038$ menor do que o resultado anterior

Obviamente, a precisão obtida com o dimensionamento (200, 200) é maior do que a que se obtém com o dimensionamento (100, 300).

No exemplo dos fumantes, se tivéssemos selecionado 250 mulheres e 250 homens em vez de 200 e 300, e obtido novamente $m_1 = 0{,}5$ e $m_2 = 0{,}4$, teríamos encontrado

$m = (0{,}3)\,(0{,}5) + (0{,}7)\,(0{,}4) = 0{,}43$, como antes

$$\sigma^2 = (0{,}3)^2\,\frac{(0{,}5)(0{,}5)}{250} + (0{,}7)^2\,\frac{(0{,}4)(0{,}6)}{250} = 0{,}0056$$

$\sigma = 0{,}024$, maior do que o resultado anterior.

Neste caso, a precisão obtida com o dimensionamento (250, 250) é pior do que a obtida com (200, 300).

Fixado o tamanho total da amostra, que dimensionamento (n_1, n_2) para A_1 e A_2 acarreta maior precisão para $c_1A_1 + c_2A_2$?

Dadas as proporções amostrais m_1 e m_2 e fixado o tamanho total da amostra, a maior precisão para $c_1A_1 + c_2A_2$, isto é, o menor desvio padrão é obtido quando tomamos n_i proporcional a

$|c_i|\sqrt{m_i(1 - m_i)}$

A utilidade desta afirmativa é restrita, pois nossa decisão sobre o dimensionamento é tomada depois de ter sido realizada a amostragem, isto é, depois que m_1 e m_2 são conhecidos. Um processo comum é basear o dimensionamento em estimativas grosseiras de m_1 e m_2 obtidas, por exemplo, a partir de uma amostragem preliminar.

Assim, no exemplo das drogas, se tivéssemos estimado $m_1 = m_2 = 0{,}05$ antes da amostragem, o dimensionamento dos números de pacientes que se submeteriam à droga 1 e à droga 2 estaria na proporção

$$\frac{1\sqrt{(0{,}05)(0{,}95)}}{1\sqrt{(0{,}05)(0{,}95)}} \quad c_1 = 1 \text{ e } c_2 = 1$$

isto é, $n_1 = n_2$. Se tivéssemos tido sorte bastante para estimar $m_1 = 0{,}2$ e $m_2 = 0{,}15$ inicialmente, nosso dimensionamento estaria na razão:

$$\frac{1\sqrt{(0{,}2)(0{,}8)}}{1\sqrt{(0{,}15)(0{,}85)}} = 1{,}12$$

isto é, 112 pacientes com a droga 1, para cada 100 pacientes com a droga 2, portanto:

112/212(400) = 211 pacientes com a droga 1
100/212(400) = 189 pacientes com a droga 2

Analogamente, no exemplo dos fumantes, o melhor dimensionamento para $m_1 = 0{,}5$ e $m_2 = 0{,}4$ é

$$\frac{0{,}3\sqrt{(0{,}5)(0{,}5)}}{0{,}7\sqrt{(0{,}4)(0{,}6)}} = 0{,}44$$

ou 44 mulheres para cada 100 homens. Logo, a amostra deveria ter

100/144(500) = 347 homens
44/144(500) = 153 mulheres

EXERCÍCIOS

Lista 15

1. Diremos que uma palavra é *longa* se contiver mais de seis letras. Tome duas novelas de diferentes autores e denote por A_1 e A_2 a proporção de palavras *longas*. Sendo, inicialmente, que as distribuições a *priori* de A_1 e A_2 são uniformes e independentes, selecione ao acaso 50 palavras de cada novela, encontre a média posterior e o desvio padrão para A_1 e A_2 e calcule $P(A_1 > A_2)$.

Admita que o número de palavras, em cada novela, é proporcional ao número de páginas, e calcule a média posterior e o desvio padrão para A, proporção de palavras longas nas duas novelas juntas.

2. Tome um cilindro sólido, cuja altura é, aproximadamente, 2/3 do seu diâmetro, por exemplo, cortando um pedaço de cabo de vassoura ou empilhando doze moedas de Cr$ 0,05. Qual a sua opinião *a priori* sobre a probabilidade A de que este cilindro repouse sobre seu lado? Decida se $A > 1/2$ lançando o cilindro, até que você tenha 90% de certeza de que sua resposta esteja correta.

3. De posse de uma tachinha, decida se a probabilidade B de que ela repouse sobre seu lado ao ser lançada é $\geq 1/2$, lançando-a até que você tenha 90% de certeza de que sua resposta esteja correta.

4. Decida se $A \geq B$ usando os dados dos probls. 2. e 3 e fazendo lançamentos adicionais do cilindro e/ou tachinha se necessário.

16 QUI QUADRADO

Tomemos uma amostra de 100 indivíduos e classifiquemos cada pessoa segundo o sexo (M ou F) e cor dos cabelos (D para cabelos escuros e L para cabelos claros), obtendo a seguinte tabela de dupla entrada:

	D	L	
M	20	40	60
F	10	30	40
	30	70	

Se não existe, na população, associação alguma entre sexo e cor de cabelos, devemos esperar que 30% dos homens na amostra (isto é, 18) tenham cabelos escuros, pois 30% das pessoas na população têm cabelos escuros. Temos, desta forma, a seguinte tabela de valores esperados:

	D	*L*	
M	18	42	60
F	12	28	40
	30	70	

Uma medida comum da discrepância existente entre as duas tabelas, isto é, o quanto os valores observados diferem dos valores esperados quando se supõe não haver associação, é a estatística χ^2 (*qui-quadrado*).

$$\chi^2 = \text{soma sobre as caselas de } \frac{(\text{observado - esperado})^2}{\text{esperado}}$$

$$\chi^2 = \frac{(20-18)^2}{18} + \frac{(40-42)^2}{42} +$$

$$+ \frac{(10-12)^2}{12} + \frac{(30-28)^2}{28} =$$

$$= 0{,}222 + 0{,}095 + 0{,}333 + 0{,}143 = 0{,}793.$$

A probabilidade de se obter um χ^2 que seja, pelo menos, tão grande quanto o observado, supondo que não existe associação, é chamado *nível* do χ^2 observado e pode ser calculado, aproximadamente, através da seguinte fórmula:

Nível do $\chi^2 = 2H(\sqrt{\chi^2})$

Logo, no nosso caso, o nível do χ^2 é

$2H(\sqrt{0{,}793}) = 2H(0{,}89) = 2(0{,}187) = 0{,}374$

Se não há associação entre sexo e cor de cabelos, na população, obteremos em 37% dos casos uma amostra, onde haverá uma discrepância entre valores observados e esperados, pelo menos, tão grande quanto a obtida, indicada pelo χ^2. Suponha que a tabela de valores observados seja a seguinte:

	D	*L*	
M	30	30	60
F	10	30	40
	40	60	

A tabela de valores esperados, supondo não haver associação, seria a seguinte:

	D	*L*	
M	24	36	60
F	16	24	40
	40	60	

Logo

$$\chi^2 = \frac{(30-24)^2}{24} + \frac{(30-36)^2}{36} +$$

$$+ \frac{(10-16)^2}{16} + \frac{(30-24)^2}{24} =$$

$$= 1{,}5 + 1 + 2{,}25 + 1{,}5 = 6{,}25$$

E o nível do χ^2 seria:

$$2H(\sqrt{6{,}25}) = 2H(2{,}5) = 2(0{,}00621) = 0{,}0124$$

Apenas uma vez em 100 amostras, supondo não haver associação, obteríamos um caso tão extremo quanto este.

O nível do χ^2 *não* mede a probabilidade *a posteriori* de que as duas características não sejam associadas. Mas a metade do nível, ou seja $H(\sqrt{\chi^2})$, mede, grosseiramente, a probabilidade *a posteriori* de que a diferença entre as proporções na população tenha o sinal ($\pm$) contrário àquela na amostra, quando nossa opinião *a priori* é que as duas proporções são independentes e uniformes (de forma que é impossível haver igualdade). Por exemplo, se A_1 e A_2 são proporções de D nas populações (M, F), nossa primeira tabela observada tem

$$m_1 = 1/3 \qquad m_2 = 1/4$$

$$\sigma_1^2 = \frac{1/3(2/3)}{60} = 0{,}00370$$

$$\sigma^2 = \frac{1/4(3/4)}{40} = 0{,}00469$$

Logo, $A_1 - A_2$ tem $m = 1/3 - 1/4 = 1/12 = 0{,}0833$.

$$\sigma^2 = 0{,}00470 + 0{,}00469 = 0{,}00839$$

$$\sigma = \sqrt{0{,}00839} = 0{,}092$$

$$P((A_1 - A_2) < 0) = H\left(\frac{0{,}0833}{0{,}092}\right) = H(0{,}905) = $$
$$= 0{,}183$$

que se compara à metade do nível, 0,187.

Para a segunda tabela observada,

$$m_1 = 0{,}5 \qquad m_2 = 0{,}25$$

$$\sigma_1^2 = \frac{(0{,}5)(0{,}5)}{60} = 0{,}00417$$

$$\sigma_2^2 = 0{,}00469$$

$$m = 0{,}25$$

$$\sigma = \sqrt{0{,}00417 + 0{,}00469} = \sqrt{0{,}00886} = 0{,}094$$

$$P((A_1 - A_2) < 0) = H\left(\frac{0{,}25}{0{,}094}\right) = h(2{,}66) = 0{,}0039$$

A metade do nível do χ^2 é 0,00621.

Se a hipótese de igualdade de duas proporções é encarada com seriedade (e raramente o é), a probabilidade de igualdade, *a posteriori*, é, aproximadamente,

$$Q = \frac{PR}{PR + 1 - P},$$

onde P é a probabilidade de igualdade, *a priori*, e

$$R = \frac{h(m/\sigma)}{\sigma}$$

Se $R = 1$, temos $P = Q$ e a amostra não sugere coisa alguma sobre a igualdade. Se $R > 1$, temos $Q > P$ e a amostra sugere a igualdade. Se $R < 1$, temos $Q < P$ e a amostra sugere a desigualdade. Por exemplo, com nossa primeira tabela,

$$R = \frac{h(0{,}905)}{0{,}902} = \frac{0{,}265}{0{,}092} = 2{,}88$$

A amostra sugere uma ausência de associação entre sexo e cor de cabelos. isto é, sugere a igualdade das proporções de homens e mulheres com cabelos escuros. Se $P = 0{,}5$, por exemplo, isto é, a associação e a ausência de associação são igualmente prováveis *a priori*, então,

$$Q = \frac{(0{,}5)(2{,}88)}{(0{,}5)(2{,}88) + 0{,}5} = 0{,}74$$

Logo, a probabilidade de ausência de associação, *a posteriori*, é de 0,74.

Se $P = 0{,}1$, então,

$$Q = \frac{(0{,}1)(2{,}88)}{(0{,}1)(2{,}88) + 0{,}9} = 0{,}24$$

A associação ainda é menos provável que sua ausência, porém, mais provável do que antes da amostragem.

Com nossa segunda tabela,

$$R = \frac{H(0{,}25/0{,}094)}{0{,}094} = \frac{h(2{,}64)}{0{,}094} = 0{,}13$$

logo, a amostra sugere uma substancial ausência de associação. Por exemplo, se $P = 0{,}5$, então,

$$Q = \frac{(0{,}5)(0{,}13)}{(0{,}5)(0{,}13) + 0{,}5} = 0{,}115$$

EXERCÍCIOS

Lista 16

1. Extraímos uma amostra de tamanho $2n$ de uma população e obtemos a seguinte tabela:

	D	L
M	$0{,}5n$	$0{,}5n$
F	$0{,}4n$	$0{,}6n$

Calcule o valor do χ^2, seu nível e R, para $n = 100$ e para $n = 400$, interpretando, depois, os resultados obtidos.

2. Que palavras são mais prováveis de serem também Palavras-T (que contém, pelo menos, uma letra T): *a*) Palavras-E ou *b*) Palavras que não contenham a letra E? Tome como população as palavras de um livro qualquer e use uma amostra aleatória com tamanho (n), suficiente para que você tenha 90% de certeza na sua resposta.

3. Selecione, ao acaso, um dos valores 0,01, 0,02, ..., 0,99 para uma proporção A_1. (Usar uma tabela de números aleatórios, para esta e outras seleções, no problema). Em seguida, lance uma moeda: se ocorrer cara, tome $A_2 = A_1$; se ocorrer coroa, selecione, ao acaso, um dos 98 valores diferentes de A_1 para A_2. Selecione, agora, amostras casuais de tamanho 100, de populações com proporções A_1 e A_2, calcule R e a probabili-

dade *a posteriori* de $A_1 = A_2$. Você tirará conclusões corretas com a amostra obtida?

4. Cem alunos são selecionados ao acaso, para assistir as aulas de estatística através da televisão. Outros cem alunos semelhantes assistem às mesmas aulas, em classe. Do primeiro grupo, 60 ficam com grau A ou B em um exame, mas apenas 50 do segundo grupo ficam com grau A ou B. Este fato sugere algo contra ou a favor da hipótese de que o ensino pela televisão *não é diferente*?

RESPOSTAS

Respostas - Lista 1

1. (*a*) $\frac{1}{4}$ (*b*) $\frac{1}{2}$ (*c*) $\frac{1}{4}$

2. (*a*) $\frac{1}{6}$ (*b*) $\frac{1}{2}$ (*c*) $\frac{1}{3}$

3. $\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{16}$

4. (*a*) $\frac{12}{30}, \frac{8}{30}, \frac{8}{30}, \frac{2}{30}$ (*b*) $\overset{HH}{MM}$ (*c*) 1

5. (*a*) $\frac{1}{2}$ (*b*) $\frac{1}{4}$ (*c*) $\frac{1}{4}$

(*d*) $\frac{1}{6}$ (*e*) $\frac{1}{6}$ (*f*) $\frac{1}{12}$

6. $\frac{1}{2}, \frac{1}{3}, \frac{1}{6}, \frac{1}{6}, \frac{1}{3}, \frac{1}{2},$

7. (Parcial) sim, sim

8. $\frac{1}{6}, \frac{1}{4}$

9. $\frac{100}{271} = 0{,}37$

10. 0,9702

11. (*a*) $\frac{1}{32} = 0{,}031$ (*b*) 0,00128

12. 0,96

13. 0,93

14. $\frac{1}{2}, \frac{1}{4}, \frac{3}{4}$

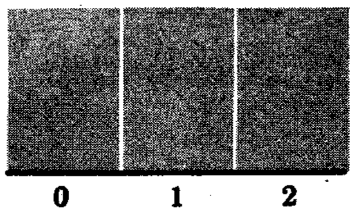
15. $\frac{1}{4}$

Respostas - Lista 2

1. *(Parcial)*

X:

v	p
0	$\frac{1}{3}$
1	$\frac{1}{3}$
2	$\frac{1}{3}$



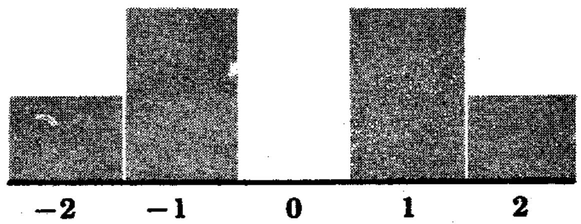
Y: A mesma de X

$X+Y$:

v	p
1	$\frac{1}{3}$
2	$\frac{1}{3}$
3	$\frac{1}{3}$

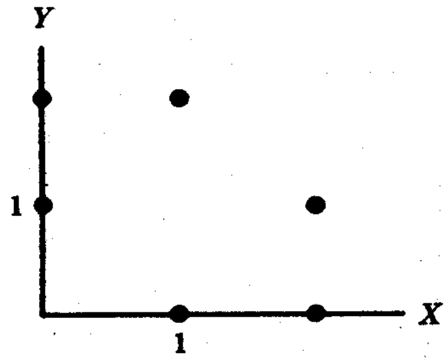
$X-Y$:

v	p
-2	$\frac{1}{6}$
-1	$\frac{1}{3}$
1	$\frac{1}{3}$
2	$\frac{1}{6}$



$|X-Y|$:

v	p
1	$\frac{2}{3}$
2	$\frac{1}{3}$



2. *(Parcial)* X e Y, da mesma forma que no probl. 1.

$X+Y$:

v	p
0	$\frac{1}{9}$
1	$\frac{2}{9}$
2	$\frac{3}{9}$
3	$\frac{2}{9}$
4	$\frac{1}{9}$

$X-Y$:

v	p
-2	$\frac{1}{9}$
-1	$\frac{2}{9}$
0	$\frac{3}{9}$
1	$\frac{2}{9}$
2	$\frac{1}{9}$

5. (*Parcial*)

$X+2$:

v	p
2	0,6
4	0,2
5	0,2

$2X$:

v	p
0	0,6
4	0,2
6	0,2

$X-1$:

v	p
-1	0,6
1	0,2
2	0,2

X^2:

v	p
0	0,6
4	0,2
9	0,2

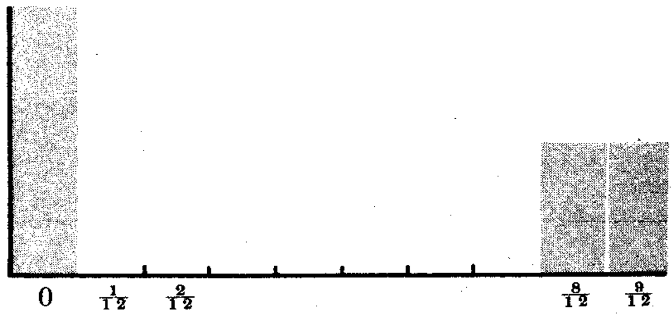
$(X-1)^2$:

v	p
1	0,8
4	0,2

$\dfrac{X}{X+1}$:

v	p
0	0,6
$\frac{2}{3}$	0,2
$\frac{3}{4}$	0,2

Histograma para $\dfrac{X}{X+1}$



6. (*Parcial*)

(*a*) X:

v	p
0	$\frac{1}{9}$
1	$\frac{2}{9}$
2	$\frac{3}{9}$
3	$\frac{2}{9}$
4	$\frac{1}{9}$

Y:

v	p
0	$\frac{5}{9}$
1	$\frac{3}{9}$
2	$\frac{1}{9}$

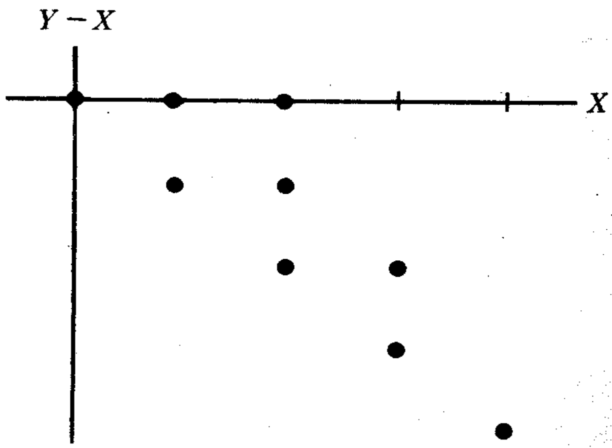
$Y-X$

$X+Y$:

v	p
0	$\frac{1}{9}$
1	$\frac{1}{9}$
2	$\frac{2}{9}$
3	$\frac{2}{9}$
4	$\frac{3}{9}$

(*b*) $Y-X$:

v	p
0	1



7. (*Parcial*) $\frac{1}{16}, \frac{4}{16}, \frac{9}{16}$

X_2:

v	p
0	$\frac{1}{16}$
1	$\frac{3}{16}$
2	$\frac{5}{16}$
3	$\frac{7}{16}$

X_3:

v	p
0	$\frac{1}{64}$
1	$\frac{7}{64}$
2	$\frac{19}{64}$
3	$\frac{37}{64}$

$$P(X_n \geq k) = \left(\frac{k+1}{4}\right)^n \qquad k = 0, 1, 2, 3$$

8. (*a*)

Resultado	p
00	$\frac{6}{20}$
01	$\frac{3}{20}$
02	$\frac{3}{20}$
10	$\frac{3}{20}$
12	$\frac{1}{20}$
20	$\frac{3}{20}$
21	$\frac{1}{20}$

(*b*)

Resultado	p
000	$\frac{2}{20}$
001	$\frac{2}{20}$
002	$\frac{2}{20}$
010	$\frac{2}{20}$
012	$\frac{1}{20}$
020	$\frac{2}{20}$
021	$\frac{1}{20}$
100	$\frac{2}{20}$
102	$\frac{1}{20}$
120	$\frac{1}{20}$
200	$\frac{2}{20}$
201	$\frac{1}{20}$
210	$\frac{1}{20}$

(*c*)

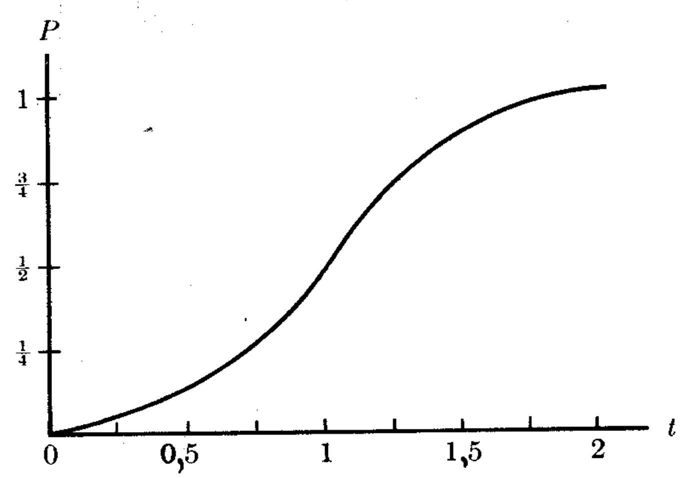
v	p
0	0,3
1	0,3
2	0,3
3	0,1

(*d*)

v	p
0	0,3
1	0,6
4	0,1

9.

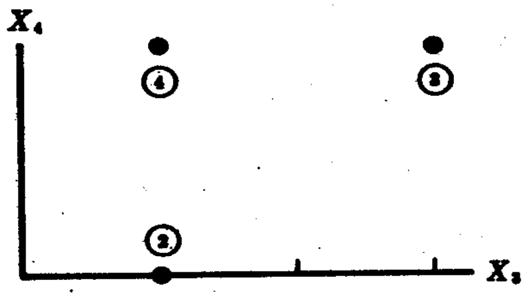
t	$P(X + Y \geq t)$
0	0
0,5	$\frac{1}{8}$
1	$\frac{1}{2}$
1,5	$\frac{7}{8}$
2	1



10. (*Parcial*)

X_4:

v	p
0	$\frac{2}{9}$
2	$\frac{7}{9}$



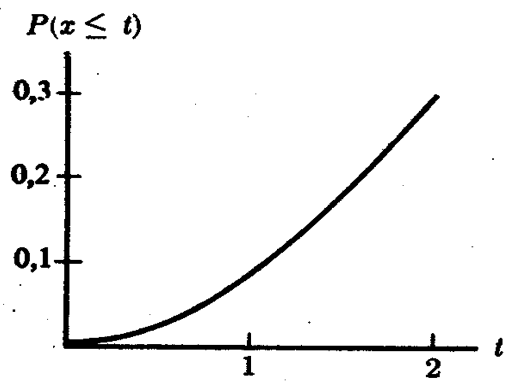
11.

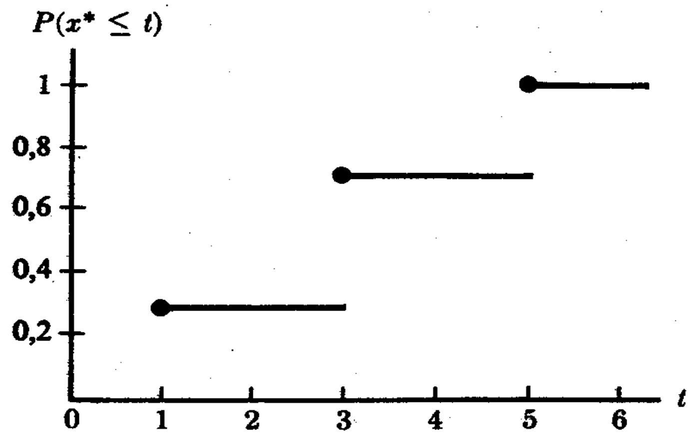
v	p
2	0,16
3	0,16
4	0,36
5	0,16
6	0,16

Respostas — Lista 3

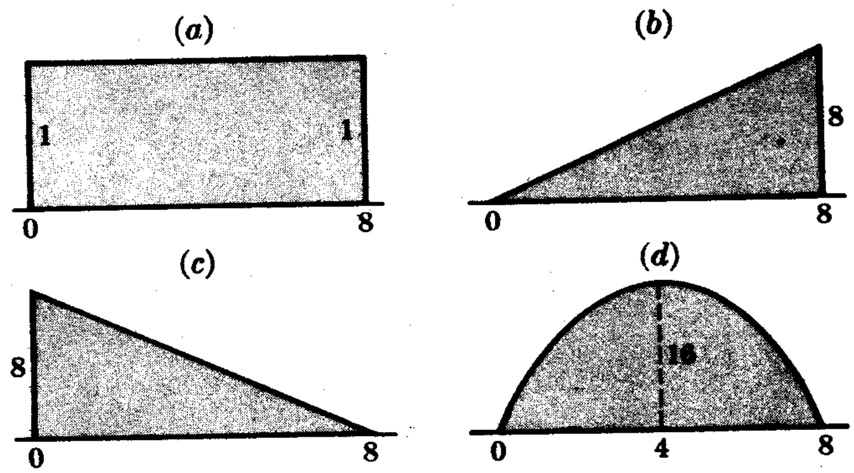
1. (*a*) 0,1 (*b*) 0,6 (*c*) 0,3

v	p
1	0,3
3	0,4
5	0,3





2.



v	$p(a)$	$p(b)$	$p(c)$	$p(d)$
1	0,25	$\frac{1}{16}$	$\frac{7}{16}$	0,15
3	0,25	$\frac{3}{16}$	$\frac{5}{16}$	0,35
5	0,25	$\frac{5}{16}$	$\frac{3}{16}$	0,35
7	0,25	$\frac{7}{16}$	$\frac{1}{16}$	0,15

3. (*Parcial*)

v	p
1	0,16
3	0,34
5	0,34
7	0,16

4. (*Parcial*) $P(2 \leq X \leq 4) = \frac{1}{3}$

v	p
1	$\frac{1}{3}$
3	$\frac{1}{3}$
5	$\frac{1}{3}$

A densidade de X é uniforme em $0 \leq t \leq 6$.

5. (*Parcial*) $P(2 \leq X \leq 4) = \frac{1}{3}$

v	p
1	$\frac{1}{9}$
3	$\frac{3}{9}$
5	$\frac{5}{9}$

densidade de X é $f(t) = \begin{cases} t & 0 \leq t \leq 6 \\ 0 & \text{complementar} \end{cases}$

Respostas - Lista 4

1. (*Parcial*)

v	$p(a)$	$p(b)$
0	$\frac{9}{36}$	$\frac{3}{15}$
1	$\frac{12}{36}$	$\frac{6}{15}$
2	$\frac{10}{36}$	$\frac{4}{15}$
3	$\frac{4}{36}$	$\frac{2}{15}$
4	$\frac{1}{36}$	0

A renda média é $\frac{4}{3}$ em ambos os casos. Substituindo 2, 4, não 0, teríamos a renda média de $\frac{7}{3}$.

2. $E(X) = 4$, $E(Y) = 2{,}5$, $E(X + Y) = 6{,}5$

3. (*Parcial*) $t = 1$ $E(X) = 1$

t	$E(X - t)^2$
0	2
1	1
2	2
3	5

4. $E(X) = 500{,}5$, a soma é 500, 500
5. (*Parcial*) (*a*) t (*b*) t (*c*) t^2 (*d*) $t - t^2$
$0 \leq E(X) \leq 1$, $0 \leq E(X^2) - [E(X)]^2 \leq \frac{1}{4}$
6. (*Parcial*) Pode ser dado. $1 \leq E(X^2) \leq 2$
7. $\frac{8}{4}$, $\frac{3}{2}$, $\frac{15}{8}$, $\frac{31}{16}$, claramente 2
8. 2,5
9. (*Parcial*) Os valores exatos são: (*a*) 5; $\frac{100}{3}$ (*b*) 4; 24
10. (*Parcial*) Os valores exatos são: 3,5 e 15
11. (*a*)

v	p = 0,4	0,5	0,6
3	0,16	0,25	0,36
2	0,24	0,25	0,24
1	0,24	0,25	0,24
−6	0,36	0,25	0,16
média	−0,96	0	0,84

(*b*)

	(1; 2,3)	(1; 0,2)	(1; 0,1)
$p = 0{,}4$	− 0,72	− 0,44	− 0,32
$p = 0{,}6$	0,68	0,36	0,28

(*c*) $E(W)$ tem o mesmo sinal que $p - 0{,}5$.

12. 0,6; 0,6; 3 Com ou sem reposição
13. 1; 2
14. 0,75 O valor exato de $E(X)$ é $\frac{2}{3}$.
15. (*a*) 1 (*b*) $\frac{2}{3}$
16. (*a*) $\frac{13}{9}$ (*b*) $\frac{5}{3}$
17. (*a*) $\frac{25}{16}$ (*b*) 1

Respostas — Lista 5

1. (*Parcial*)

	$E(X)$	$E(Y)$	$\sigma(X)$	$\sigma(Y)$	$\sigma(X+Y)$
(*a*)	$\frac{1}{2}$	1	$\frac{1}{2}$	0,82	0,96
(*b*)	$\frac{2}{3}$	$\frac{4}{3}$	0,94	0,47	0,82
(*c*)	$\frac{2}{3}$	$\frac{2}{3}$	0,75	0,75	0,75
(*d*)	1	$\frac{6}{5}$	0,89	0,75	0,98
(*e*)	1	1	1	1	1,4
(*f*)	0	1	0	0,82	0,82
(*g*)	1	1	0,82	0,82	0
(*h*)	$\frac{4}{3}$	$\frac{4}{3}$	0,47	0,94	1,2
(*i*)	1	$\frac{4}{3}$	0,82	0,47	0,47

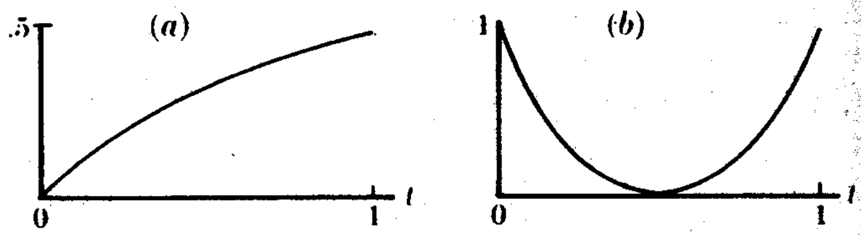
2. 4,5; 2,9 3. (*a*) 1,5; 0,87 (*b*) 1,5; 0,5
4. 68, 19, 64, 2, 72, 4, 16
5. (*Parcial*)

n	$E(S_n)$	$\sigma^2(S_n)$	$E(Y_n)$	$\sigma^2(Y_n)$
1	0,6	0,24	0,6	0,24
2	1,2	0,48	0,6	0,12
3	1,8	0,72	0,6	0,08
4	2,4	0,96	0,6	0,06

8. (*Parcial*)

	Maiores σ	Valor t	Menores σ	Valor t
(*a*)	0,5	0,5	0	0 ou 1
(*b*)	2,5	0,5	0	0 ou 1
(*c*)	0,77	0,3	0,49	0 ou 0,6
(*d*)	0,98	0,6	0,49	0
(*e*)	0,94	perto de 0 ou 2	0,82	1
(*f*)	grande quanto queira	grande	0,82	1
(*g*)	0,78	qualquer	0,78	qualquer
(*h*)	1,6	2	0,78	1

9. (*Parcial*)



Valores exatos:	Valores exatos:
$m = 0{,}63$	$m = 0{,}5$
$\sigma = 0{,}25$	$\sigma = 0{,}39$
$P = 0{,}61$	$P = 0{,}46$

10. 1,4

11. 0,83

Respostas — Lista 6

1. (*Parcial*)

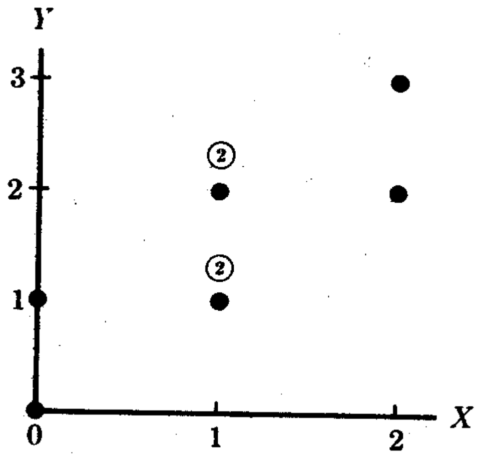
a) emq: 2,5; 1,75; 1,5; 1,75; 2; 2,5
Valor: $-\frac{2}{3}$; $-\frac{1}{6}$; 0; $-\frac{1}{6}$; $-\frac{2}{3}$

b) emq: 1,5; 3,5; 2,5; 1,42
Valor: 0; $-\frac{4}{3}$; $-\frac{2}{3}$; 0,053

c) 0; 1,5; 1 para $X = 0,\ 1,\ 2,$ respectivamente
Valor: 0,25

e) 0,5; 2, 1 para $Y = 0,\ 1,\ 3,$ respectivamente
Valor: 0,75; não.

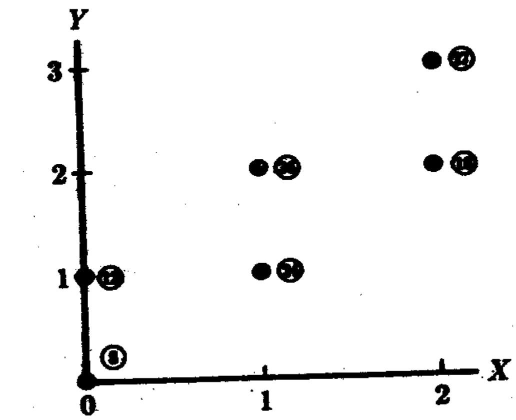
2.



a) 0,5; 1,5; 2,5 para $X = 0,\ 1,\ 2,$ respectivamente
Valor: $\frac{2}{3}$; sim.

b) 0; $\frac{2}{3}$; $\frac{4}{3}$; 2 para $Y = 0,\ 1,\ 2,\ 3,$ respectivamente
Valor: $\frac{2}{3}$; sim.

3.



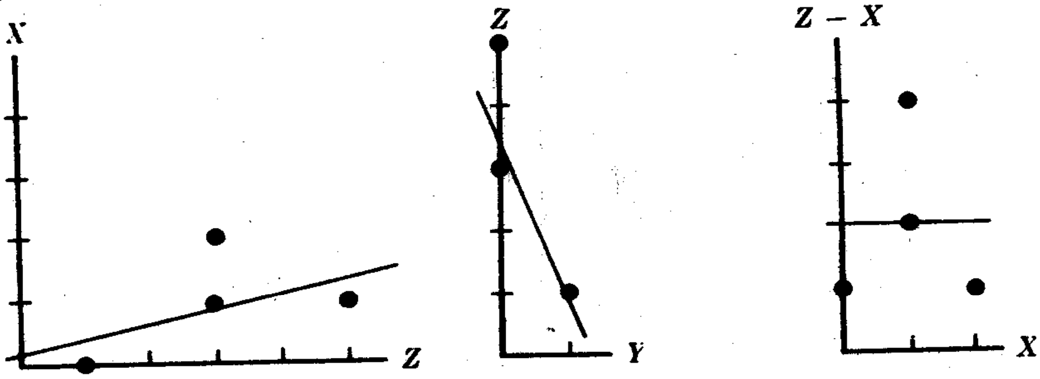
a) 0,6; 1,6; 2,6 para $X = 0, 1, 2$, respectivamente
Valor: $\frac{2}{3}$; sim.

b) O mesmo que em 2(*b*)

4. (*Parcial*) Valores: (*a*) 0; 0 (*b*) $\frac{1}{3}$; 0

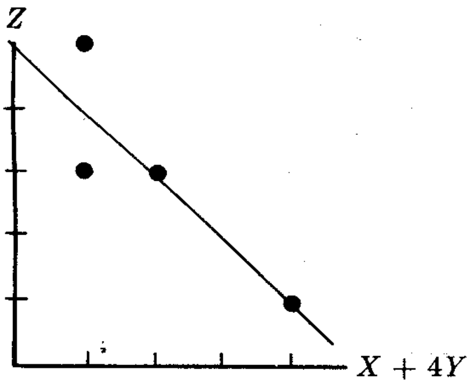
Respostas — Lista 7

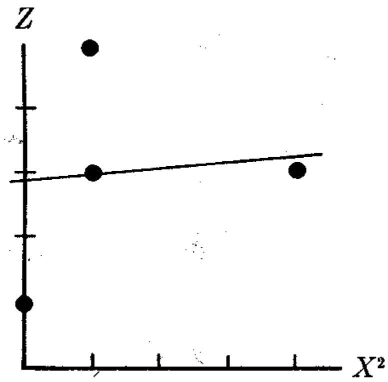
1.



(*a*) $\rho^2 = 0{,}25$ (*b*) $\rho^2 = \frac{2}{3}$ (*c*) $\rho^2 = 0$

$$U = \frac{Z + 1}{4} \qquad U = \frac{11 - 8Y}{3} \qquad U = 2$$





$$U = 5 - (X + 4Y) \qquad\qquad U = \frac{2X^2 + 24}{9}$$

4. Para o segundo diagrama, não é possível tomar $\rho^2 = 1$, mas podemos de diversas maneiras tomar ρ^2 perto de 1; por exemplo (u, u) com u grande. Existem várias soluções para cada um dos outros. Por exemplo: primeiro diagrama (*a*)(0; 11/3) (*b*) (4; 4); segundo diagrama (*a*) (1; 2).

5. (Parcial)

a

$h \backslash k$	1	3	10
1	$\frac{1}{4}$	$\frac{3}{4}$	$\frac{20}{8}$
3	$\frac{1}{4}$	$\frac{3}{4}$	$\frac{20}{8}$
10	$\frac{10}{103}$	$\frac{30}{103}$	$\frac{100}{103}$

b

$h \backslash k$	1	3	10
1	$\frac{1}{4}$	$\frac{3}{4}$	$\frac{5}{2}$
3	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{5}{6}$
10	$\frac{1}{103}$	$\frac{10}{103}$	$\frac{100}{103}$

ρ^2

$h \backslash k$	1	3	10
1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
3	$\frac{3}{4}$	$\frac{3}{4}$	$\frac{3}{4}$
10	$\frac{100}{103}$	$\frac{100}{103}$	$\frac{100}{103}$

6. $(90G + 11)/140$, $\frac{81}{84}$ 7. $(X + 3)/6$, $\frac{1}{6}$ **8.** 0,6

Respostas — Lista 8

1. $(3X + 1)/2$, $(2X + Y)/2$, $X + 1$, 0,9, $\frac{49}{60}$, 0,95, 0,5, $\frac{8}{11}$, 0,71

3. (Parcial $c = \dfrac{\text{cov}(X, Z)}{\text{cov}(X, X)}$ $\quad d = \dfrac{\text{cov}(Y, Z)}{\text{cov}(Y, Y)}$

4. (*a*) $X_2 + 0{,}6$ (*b*) $(X_1 + X_3)/2$ (*c*) $X_2/2$

5. $X^2 - 2X + 1$

Respostas — Lista 9

1. $\frac{13}{12}$, $\frac{95}{144}$, 0,81

2.

v	p
2	0,16
3	0,40
4	0,25
5	0,08
6	0,10
8	0,01

sim para

v	p
1	0,4
2	0,5
4	0,1

5. Para

v	p
0	0,5
1	0,5

6. $(0{,}3)^6 = 0{,}000729$

v	p
0	$1 - (0{,}3)^5$
1	$(0{,}3)^3$

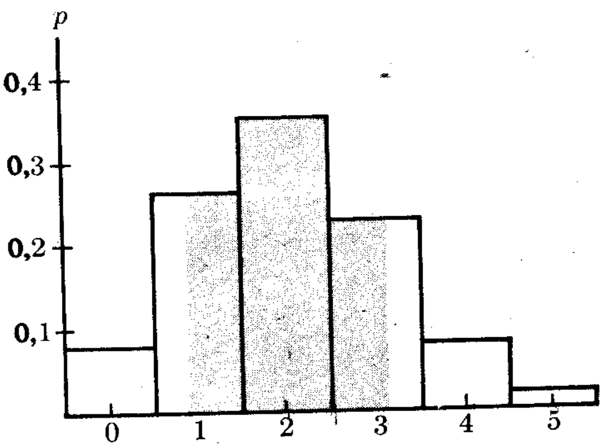
7. 0,4 $\frac{2}{3}$; 0,16; $\frac{2}{3}$; $\frac{2}{3}$; $(0{,}4)^5 = 0{,}01024$; 0,98976.

8. *(Parcial)* $\frac{1}{63}$, $\frac{1}{3}$, funciona

Respostas — Lista 10

1. 0,36; 0,48; 0,16; 0,22; 0,13; 0,38; 0,25; 0,15; 0,41; 0,41; 0; 0

2. *(Parcial)*



$n = 5$; $p = 0{,}4$; $m = 2$; $\sigma = 1{,}1$; $P = 0{,}64$

3. *(Parcial)* Gráfico de $4p(1-p)^2$, 0,25; 0,42
4. 4 ou 5; 0,35
5. *(Parcial)* Gráfico de $10p$, $10p(1-p)$; p; $p(1-p)/10$
6. *(Parcial)* Gráfico de $0,4r$; $0,24r$; 0,4; $0,24/r$.
7. *a*) 1 1; 1; 1; 1; 1; 0; 1/2; 2/3; 4/5; 9/10 *b*) 1; 0 qualquer m
c) 0; 0,25; 0,30; 0,33; $(0,9)^{10} = 0,35$
8. *(Parcial)* $r = \dfrac{(0,3)\,(63)}{(0,7)\,(38)} = 189/266$ 30 é o valor mais provável.

10. *(Parcial)*

k	$p(k)$	$p^2(k)$	$kp^2(k)$	$(k+\frac{1}{2})p^2(k)$
1	0,5000	0,2500	0,2500	0,3750
3	0,3125	0,09766	0,2930	0,3418
5	0,2461	0,06057	0,3028	0,3331

Para grandes k, $kp^2(k)$ está em torno de $1/\pi = 0,3183$, isto é, $p(k)$ está em torno de $1\sqrt{\pi k} = 0,5642/\sqrt{k}$

11.

v	$p(a)$	$p(b)$	$p(c)$
0	0,32	0,16	0.4
1	0,56	0,48	0,4
2	0,12	0,36	0,2
$E(S)$	0,8	0,8	0,8
$\sigma^2(S)$	0,40	0,48	0,56

Respostas — Lista 11

1. 0,159; 0,774; 0,001; 675
3. 0,009 (exato $P = 0,011$)
4. 0,00016; 0,0009
5. *(Parcial)* X^{*2}:

v	p	
	a	b
0,09	0,452	0,458
0,81	0,318	0,319
2,25	0,158	0,156
4,41	0,056	0,053
7,29	0,014	0,012

6. 0,976

7. $\frac{63}{256} = 0{,}246$; 0,247

8.

n	p	d	$P(\lvert X - p \rvert \leq d)$
225	0,5	0,02	0,45
225	0,5	0,05	0,87
225	0,7	0,02	0,49
225	0,7	0,05	0,90
900	0,5	0,02	0,77
900	0,5	0,05	0,997
900	0,7	0,02	0,81
900	0,7	0,05	0,999

9.

10. 0,0399, $\sqrt{0{,}001601}$ (< 0,0410), $\sqrt{0{,}001514}$ (> 0,038)

11. *(Parcial)* (*b*)

v	p
0	$\frac{1}{16}$
1	$\frac{2}{16}$
2	$\frac{3}{16}$
3	$\frac{4}{16}$
4	$\frac{3}{16}$
5	$\frac{2}{16}$
6	$\frac{1}{16}$

(*c*)

v	p	
	Exato	*Aproximadamente*
0,12	$a = 0{,}0039$	0,006
1,11	$4a = 0{,}0156$	0,016
2,10	$10a = 0{,}039$	0,037
3,9	$20a = 0{,}078$	0,073
4,8	$31a = 0{,}121$	0,120
5,7	$40a = 0{,}156$	0,160
6	$44a = 0{,}172$	0,176
	onde $a = \frac{1}{256}$	

Respostas — Lista 12

1. *b* se *B*, *a* se *W*; 0,75

3. *a*) *b* se *B*, qualquer se *W*; 0,8 *b*) *b* sempre, 0,9

4.

λ	*A*	*X*			
		GG	*GU*	*UG*	*UU*
$\frac{1}{3}$	0,1	0,01	0,09	0,09	0,81
$\frac{1}{3}$	0,2	0,04	0,16	0,16	0,64
$\frac{1}{3}$	0,5	0,25	0,25	0,25	0,25

0,332; 0,0294

5. *(Parcial)* 0,332; 0,0294

6. *(Parcial)* *a*) 2 se $X = 0$, qualquer se $X = 1$; 1 se $X = 2$;2/3.

b) O mesmo método de decisão de (*a*); 5/6.

7. *(Parcial)* **Tabela de Dados**

λ	*A*	*X*			
		T	H	I	S
$\frac{1}{3}$	THIS	0,25	0,25	0,25	0,25
$\frac{1}{3}$	IS	0	0	0,5	0,5
$\frac{1}{3}$	IT	0,5	0	0,5	0

Chance total $= \frac{7}{12}$

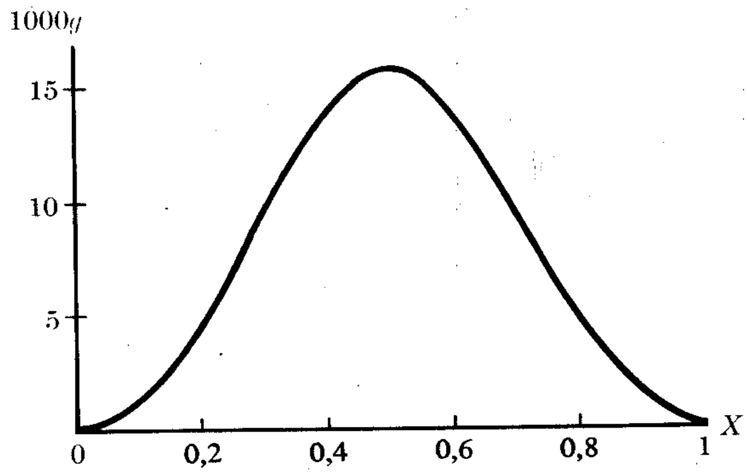
8. *(Parcial)* **Tabela de Dados** em q total = 0,9

λ	Y	X = 0	1	2	3
0,4	0	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
0,2	1	$\frac{1}{2}$	$\frac{1}{2}$	0	0
0,3	2	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{3}$	0
0,1	3	1	0	0	0

9.

n	$p(1)$	$p(2)$	$p(3)$	$p(4)$	$p(5)$
3	$\frac{20}{47}$	$\frac{20}{47}$	$\frac{20}{47}$	0	0
4	$\frac{15}{47}$	$\frac{15}{47}$	$\frac{15}{47}$	$\frac{5}{9}$	0
5	$\frac{12}{47}$	$\frac{12}{47}$	$\frac{12}{47}$	$\frac{4}{9}$	1

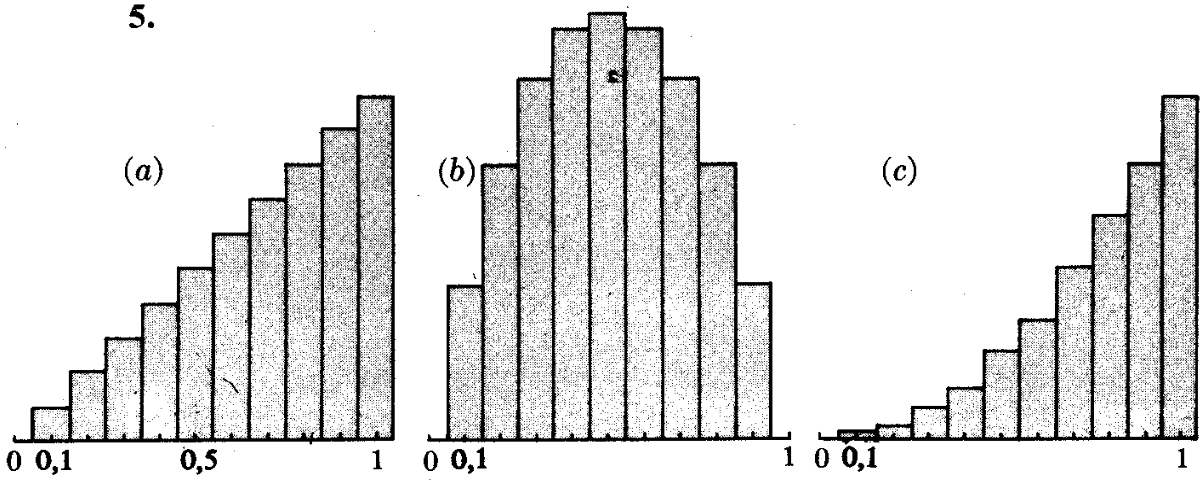
Respostas — Lista 13



σ^2 exato é 1/36.

4. $P_1 = 0{,}0256$ $\quad P_2 = 0{,}3840$ $\quad P_3 = 0{,}5904$ $\quad E(A^*) = 0{,}76688$

5.



Respostas — Lista 14

3. *(Parcial)* 0,3024

Respostas — Lista 16

1. *(Parcial)* 2,02; 0,156; 2,1; 32, perto de 0; 0,2

4. A favor.

APÊNDICE

Tabela de Quadrados, 10-99

	0	1	2	3	4	5	6	7	8	9
1	100	121	144	169	196	225	256	289	324	361
2	400	441	484	529	576	625	676	729	784	841
3	900	961	1024	1089	1156	1225	1296	1369	1444	1521
4	1600	1681	1764	1849	1936	2025	2116	2209	2304	2401
5	2500	2601	2704	2809	2916	3025	3136	3249	3364	3481
6	3600	3721	3844	3969	4096	4225	4356	4489	4624	4761
7	4900	5041	5184	5329	5476	5625	5776	5929	6084	6241
8	6400	6561	6724	6889	7056	7225	7396	7569	7744	7921
9	8100	8281	8464	8649	8836	9025	9216	9409	9604	9801

Tabela da Distribuição Normal

t	$h(t)$	$H(t)$	t	$h(t)$	$H(t)$	t	$h(t)$	$H(t)$
0,00	0,399	0,500	1,50	0,130	0,0668	3,00	0,0^{2}443	0,0^{2}135
0,05	0,398	0,480	1,55	0,120	0,0606	3,05	0,0^{2}381	0,0^{2}115
0,10	0,397	0,460	1,60	0,111	0,0548	3,10	0,0^{2}327	0,0^{3}968
0,15	0,394	0,440	1,65	0,102	0,0495	3,15	0,0^{2}279	0,0^{3}816
0,20	0,391	0,421	1,70	0,0940	0,0446	3,20	0,0^{2}238	0,0^{3}687
0,25	0,387	0,401	1,75	0,0863	0,0401	3,25	0,0^{2}203	0,0^{3}577
0,30	0,381	0,382	1,80	0,0790	0,0359	3,30	0,0^{2}172	0,0^{3}483
0,35	0,375	0,363	1,85	0,0721	0,0322	3,35	0,0^{2}146	0,0^{3}404
0,40	0,368	0,345	1,90	0,0656	0,0287	3,40	0,0^{2}123	0,0^{3}337
0,45	0,361	0,326	1,95	0,0596	0,0256	3,45	0,0^{2}104	0,0^{3}280
0,50	0,352	0,309	2,00	0,0540	0,0228	3,50	0,0^{3}873	0,0^{3}233
0,55	0,343	0,291	2,05	0,0488	0,0202	3,55	0,0^{3}732	0,0^{3}193
0,60	0,333	0,274	2,10	0,0440	0,0179	3,60	0,0^{3}612	0,0^{3}159
0,65	0,323	0,258	2,15	0,0396	0,0158	3,65	0,0^{3}510	0,0^{3}131
0,70	0,312	0,242	2,20	0,0355	0,0139	3,70	0,0^{3}425	0,0^{3}108
0,75	0,301	0,227	2,25	0,0317	0,0122	3,75	0,0^{3}353	0,0 884
0,80	0,290	0,212	2,30	0,0283	0,0107	3,80	0,0^{3}292	0,0^{4}723
0,85	0,278	0,198	2,35	0,0252	0,0^{2}939	3,85	0,0^{3}241	0,0^{4}591
0,90	0,266	0,184	2,40	0,0224	0,0^{2}820	3,90	0,0^{3}199	0,0^{4}481
0,95	0,254	0,171	2,45	0,0198	0,0^{2}714	3,95	0,0^{3}163	0,0^{4}391
1,00	0,242	0,159	2,50	0,0175	0,0^{2}621	4,00	0,0^{3}134	0,0^{4}317
1,05	0,230	0,147	2,55	0,0154	0,0^{2}539	4,05	0,0^{3}109	0,0^{4}256
1,10	0,218	0,136	2,60	0,0136	0,0^{2}466	4,10	0,0^{4}893	0,0^{4}207
1,15	0,206	0,125	2,65	0,0119	0,0^{2}402	4,15	0,0^{4}726	0,0^{4}166
1,20	0,194	0,115	2,70	0,0104	0,0^{2}347	4,20	0,0^{4}589	0,0^{4}133
1,25	0,183	0,106	2,75	0,0^{2}909	0,0^{2}298	4,25	0,0^{4}477	0,0^{4}107
1,30	0,171	0,0968	2,80	0,0^{2}792	0,0^{2}256	4,30	0,0^{4}385	0,0^{5}854
1,35	0,160	0,0885	2,85	0,0^{2}687	0,0^{2}219	4,35	0,0^{4}310	0,0^{5}681
1,40	0,150	0,0808	2,90	0,0^{2}595	0,0^{2}187	4,40	0,0^{4}249	0,0^{5}541
1,45	0,139	0,0735	2,95	0,0^{2}514	0,0^{2}159	4,45	0,0^{4}200	0,0^{5}429

ÍNDICE ANALÍTICO

COMPOSTO E IMPRESSO POR
SEDEGRA SOCIEDADE EDITORA E GRÁFICA LTDA.
RUA MATIPÓ, 101/115 — TEL.: 261-8160 — RIO-GB