Variáveis bidimensionais

Wagner H. Bonat
Elias T. Krainski
Fernando P. Mayer

Universidade Federal do Paraná
Departamento de Estatística
Laboratório de Estatística e Geoinformação

09/04/2018

UFPR
UNIVERSIDADE FEDERAL DO PARANÁ

DEST/UFPR

Sumário

Introdução

- Interesse no comportamento conjunto de várias variáveis.

Exemplo 5.1

- Uma amostra de 20 alunos do primeiro ano de uma faculdade foi escolhida. Perguntou-se aos alunos se trabalhavam, variável que foi representada por X, e o número de vestibulares prestados, variável representada por Y. Os dados obtidos estão na tabela abaixo.

X	não	sim	não	não	não	sim	sim	não	sim	sim
Y	1	1	2	1	1	2	3	1	1	1

X	não	não	sim	não	sim	não	não	não	sim	não
Y	2	2	1	3	2	2	2	1	3	2

Exemplo 5.1

Distribuição conjunta

(X,Y)	Freq
não,1	5
não,2	6
não,3	1
sim,1	4
sim,2	2
sim,3	2
Sum	20

Exemplo 5.1

Distribuição conjunta (melhor para visualizar)

X/Y	1	2	3	Sum
não	5	6	1	12
sim	4	2	2	8
Sum	9	8	3	20

Distribuição marginal de X

não	sim	Sum
12	8	20

Distribuição marginal de Y

1	2	3	Sum
9	8	3	20

Exemplo 5.2

- Um estudo envolveu 345 pacientes HIV positivos, acompanhados durante um ano, pelo setor de doenças infecciosas de um grande hospital público. Os dados apresentados contêm as ocorrências relacionadas às variáveis número de internações (I) e número de crises com infecções oportunistas (C).

I/C	0	1	2	3	4
0	84	21	8	2	0
1	20	59	35	14	2
2	6	11	43	28	12

- Obtenha as marginais de I e C.
- Exemplo 5.3 tarefa de casa.

Exemplo 5.2

- Marginal de I

0	1	2	Sum
115	130	100	345

- Marginal de C

0	1	2	3	4	Sum
110	91	86	44	14	345

Função de probabilidade conjunta

Sejam X e Y duas VAs discretas originárias do mesmo fenômeno aleatório, com valores atribuídos a partir do mesmo espaço amostral.

A **função de probabilidade conjunta** é definida, para todos os possíveis pares de valores (X, Y), da seguinte forma:

$$p(x, y) = P[(X = x) \cap (Y = y)] = P(X = x, Y = y).$$

Ou seja, $p(x, y)$ representa a probabilidade de (X, Y) ser igual a (x, y).

A função de probabilidade conjunta também pode ser chamada de **distribuição conjunta** ou simplesmente **conjunta** das variáveis.

Exemplo 5.4

Uma empresa atende encomendas de supermercados dividindo os pedidos em duas partes de modo a serem atendidos, de forma independente, pelas suas duas fábricas. Devido à grande demanda, pode haver atraso no cronograma de entrega, sendo que a fábrica I atrasa com probabilidade 0.1 e a II com 0.2. Sejam A_I e A_{II} os eventos correspondentes a ocorrência de atraso nas fábricas I e II, respectivamente.

Para uma entrega, a indústria recebe 200 u.m, mas paga 20 para cada fábrica que atrasar. Considere que o supermercado que recebe a encomenda fez um índice relacionado à pontualidade de entrega. Este índice, atribuiu 10 pontos para cada entrega dentro do cronograma previsto. Denote por X o valor recebido pelo pedido e Y o índice obtido. Obtenha a conjunta de Y e X e as marginais de Y e X.

Exemplo 5.5

Uma região foi dividida em 10 sub-regiões. Em cada uma delas, foram observadas duas variáveis: número de poços artesianos (X) e número de riachos ou rios presentes na sub-região (Y). Os resultados são apresentados na tabela a seguir:

Sub-região	1	2	3	4	5	6	7	8	9	10
X	0	0	0	0	1	2	1	2	2	0
Y	1	2	1	0	1	0	0	1	2	2

- Construa a distribuição conjunta e marginais de X e Y.
- Exemplo 5.6 tarefa de casa.

Exemplo 5.5

Consideramos que cada região tem a mesma probabilidade 1/10 de ser escolhida. Assim a distribuição conjunta é:

(X,Y)	p(x,y)
0,0	0.1
0,1	0.2
0,2	0.2
1,0	0.1
1,1	0.1
2,0	0.1
2,1	0.1
2,2	0.1
Sum	1.0

Exemplo 5.5

Uma forma mais conveniente é

X/Y	0	1	2
0	0.1	0.2	0.2
1	0.1	0.1	0.0
2	0.1	0.1	0.1

Para obter as marginais, efetuamos a soma nas linhas para oobter a marginal de X, e nas colunas para obter a marginal de Y. Por exemplo, $P(X = 0)$ é obtida através de:

$$\begin{aligned} P(X = 0) &= P(X = 0, Y = 0) + P(X = 0, Y = 1) + P(X = 0, Y = 2) \\ &= 0.1 + 0.2 + 0.2 = 0.5 \end{aligned}$$

Exemplo 5.5

Repetindo os cálculos para todos os valores de X e Y, obtemos as marginais:

X/Y	0	1	2	P(X=x)
0	0.1	0.2	0.2	0.5
1	0.1	0.1	0.0	0.2
2	0.1	0.1	0.1	0.3
P(Y=y)	0.3	0.4	0.3	1.0

Marginal de X

0	1	2	Sum
0.5	0.2	0.3	1

Marginal de Y

0	1	2	Sum
0.3	0.4	0.3	1

Funções de probabilidade marginal

Da função de probabilidade conjunta $p(x, y)$, é possível então obter as **funções de probabilidade marginais** de X e Y, através da soma de uma das corrdenadas:

$$P(X = x) = \sum_{y} p(x, y) \quad \text{e} \quad P(Y = y) = \sum_{x} p(x, y)$$

com o somatório percorrendo todos os valores de X ou Y, conforme for o caso.

Associação entre variáveis

Alguma coisa

Exemplo 5.7

Dentre os alunos do 1° ano do ensino médio de uma certa escola, selecionou-se os quinze alunos com melhor desempenho, (nota acima de 7) em inglês. Para esses alunos, foi construída a tabela abaixo com as notas de inglês (I), português (P) e matemática (M):

I	7	7	7	7	8	8	8	8	8	8	8	9	9	9	10
P	8	6	8	9	8	6	9	7	7	6	7	8	9	8	8
M	5	6	7	5	5	5	6	4	7	6	5	5	6	5	5

Exemplo 5.7 - Distribuições conjuntas

Inglês e Português:

I/P	6	7	8	9
7	1	0	2	1
8	2	3	1	1
9	0	0	2	1
10	0	0	1	0

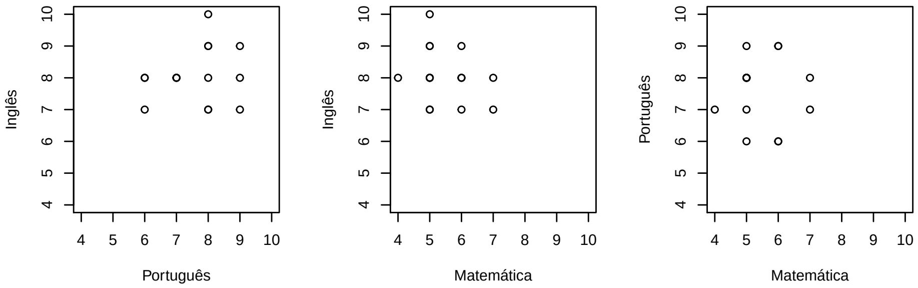
Inglês e Matemática:

I/M	4	5	6	7
7	0	2	1	1
8	1	3	2	1
9	0	2	1	0
10	0	1	0	0

Português e Matemática:

P/M	4	5	6	7
6	0	1	2	0
7	1	1	0	1
8	0	5	0	1
9	0	1	2	0

Exemplo 5.7 - Diagramas de dispersão



Probabilidade condicional para VAs discretas

- A **probabilidade condicional** de $X = x$, dado que $Y = y$ ocorreu, é dada pela expressão:

$$P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}, \quad \text{se} \quad P(Y = y) > 0.$$

- Duas VAs discretas são **independentes**, se a ocorrência de qualquer valor de uma delas não altera a probabilidade de valores da outra. Em termos matemáticos

$$P(X = x|Y = y) = P(X = x).$$

- Definição alternativa

$$P(X = x, Y = y) = P(X = x)P(Y = y), \quad \forall (x, y).$$

Exemplo 5.8

- O Centro Acadêmico de uma faculdade de administração fez um levantamento da remuneração dos estágios dos alunos, em salários mínimos, com relação ao ano que estão cursando. As probabilidades de cada caso são apresentadas na próxima tabela, incluindo as distribuições marginais.

Salario/Ano	2	3	4	5	$P(Sal = x)$
2	2/25	2/25	1/25	0	5/25
3	2/25	5/25	2/25	2/25	11/25
4	1/25	2/25	2/25	4/25	9/25
$P(Ano = y)$	5/25	9/25	5/25	6/25	1

- X e Y são independentes?

Exemplo 5.9

- Em uma clínica médica foram coletados dados em 150 pacientes, referentes ao último ano. Observou-se a ocorrência de infecções urinárias (U) e o número de parceiros sexuais (N). Deseja-se verificar se essas variáveis estão associadas.

U/N	0	1	2 +	Total
Sim	12	21	47	80
Não	45	18	7	70
Total	57	39	54	150

- Estude a associação entre U e N.
- Exemplo 5.10 tarefa de casa.

Exemplo 5.9

Tabela com porcentagens em relação ao total de coluna.

U/N	0	1	2 +	Total
Sim	21,1%	53,8%	87,0%	53,3%
Não	78,9%	46,2%	13,0%	46,7%
Total	100%	100%	100%	100%

Caso não haja associação de U com N, devemos esperar porcentagens similares em cada valor de N.

Correlação entre variáveis num conjunto de dados

- Considere um conjunto de dados com n pares de valores para as variáveis X e Y. O coeficiente de correlação mede a dependência linear entre as variáveis e é calculado por

$$\rho_{XY} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{[\sum_{j=1}^{n}(x_j - \bar{x})^2][\sum_{j=1}^{n}(y_j - \bar{y})^2]}}.$$

- Formula mais conveniente para cálculos

$$\rho_{XY} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sqrt{[\sum_{j=1}^{n} x_j^2 - n\bar{x}^2][\sum_{j=1}^{n} y_j^2 - n\bar{y}^2]}}.$$

Correlação entre variáveis num conjunto de dados

- A quantidade de chuva é um fator importante na produtividade agrícola. Para medir esse efeito foram anotados, para 8 diferentes regiões produtoras de soja, o índice pluviométrico em milímetros (X) e a produção do último ano em toneladas (Y). Determine o coeficiente de correlação.

```
# [1] 1085
```

```
# [1] 151533
```

```
# [1] 310
```

```
# [1] 12640
```

```
# [1] 43245
```

```
# [1] 0.7245956
```

- Exemplo 5.12 tarefa de casa.

Covariância de duas v.a

- Dependência linear entre X e Y é a covariância:

$$Cov(X, Y) = \sigma_{XY} = E[(X - \mu_X)(Y - \mu_y)].$$

- Exemplo 5.14 - As variáveis X e Y têm a seguinte distribuição conjunta.

X - Y & (2,2) & (3,4) & (3,8) & (4,6) & (5,4) & (5,8) & (6,10) \ p(x,y) & 0,1 & 0,2 & 0,1 & 0,2 & 0,1 & 0,2 & 0,1 \

- Calcule a covariância entre X e Y.

Correlação de duas v.a

- Exemplo 5.15 - Para os dados do exemplo 5.5 calcule a covariância e a correlação.

X - Y & 0 & 1 & 2 \ 0 & 1/10 & 2/10 & 2/10 \ 1 & 1/10 & 1/10 & 0 \ 2 & 1/10 & 1/10 & 1/10 \

Resultados importantes

- $ E(X + Y) = E(X) + E(Y).$
- $ E(XY) = E(X)E(Y)$ se e somente se X e Y são independentes.
- $Var(X + Y) = Var(X) + Var(Y) - 2Cov(X, Y)$.

Exercícios recomendados

- Seção 5.1 - 1, 2, 3, 4 e 6.
- Seção 5.2 - 1, 2, 3, 4 e 5.