Capítulo 1

Distribuições de freqüências

A Estatística é uma ciência que se ocupa em grande parte do estudo das distribuições de freqüência de variáveis definidas em populações.

Vamos imaginar uma população como um conjunto $\Pi = \{I_1, I_2, \ldots, I_N\}$, em que I_j é o j$^{\underline{\text{O}}}$ indivíduo da população, e $N = \#\Pi$ é o tamanho da população. Uma variável X será uma função definida na população, tomando valores em dado conjunto $\mathcal{V}$. Em símbolos,

$$X : \Pi \to \mathcal{V}.$$

Quando $\mathcal{V}$ for um conjunto de números, diremos que X é uma *variável numérica* ou *quantitativa*. Quando $\mathcal{V}$ for um conjunto de categorias não numéricas, diremos que X é uma *variável categórica* ou *qualitativa*.

Exemplo 1.1 *Suponha que* Π *representa os funcionários de uma dada companhia, numerados de* 1 *a* 36*. Em outras palavras,*

$$\Pi = \{1, \ldots, 36\}, \tag{1.1}$$

e desta forma $I_j = j$, $j = 1, \ldots, 36$*. Sejam as seguintes variáveis.*

$$\begin{aligned} X &= \textit{Sexo} \\ Y &= \textit{Escolaridade} \\ W &= \textit{Número de filhos} \\ Z &= \textit{Salário} \end{aligned}$$

Π	Sexo	Escolaridade	Filhos	Salário	Idade
1	masculino	1º grau	0	4.00	26
2	masculino	1º grau	1	4.56	32
3	feminino	1º grau	2	5.25	36
4	masculino	2º grau	0	5.73	20
5	feminino	1º grau	1	6.26	40
6	masculino	1º grau	1	6.66	28
7	feminino	1º grau	0	6.86	41
8	masculino	1º grau	0	7.39	43
9	feminino	2º grau	1	7.59	34
10	masculino	2º grau	0	7.84	23
11	feminino	2º grau	2	8.12	33
12	masculino	1º grau	0	8.46	27
13	feminino	2º grau	0	8.74	37
14	masculino	1º grau	3	8.95	44
15	feminino	2º grau	0	9.13	30
16	masculino	2º grau	3	9.35	38
17	feminino	2º grau	1	9.77	31
18	masculino	1º grau	2	9.80	39
19	feminino	superior	1	10.53	25
20	masculino	2º grau	0	10.76	37
21	feminino	2º grau	1	11.06	30
22	masculino	1º grau	2	11.59	34
23	feminino	2º grau	4	12.00	41
24	masculino	superior	0	12.79	26
25	feminino	2º grau	2	13.23	32
26	masculino	2º grau	2	13.60	35
27	feminino	1º grau	0	13.85	46
28	masculino	2º grau	2	14.69	29
29	masculino	2º grau	5	14.71	40
30	masculino	2º grau	2	15.99	35
31	feminino	superior	3	16.22	31
32	masculino	2º grau	1	16.61	36
33	feminino	superior	3	17.26	43
34	masculino	superior	4	18.75	33
35	feminino	2º grau	2	19.40	48
36	masculino	superior	3	23.30	42

Tabela 1.1

X e Y são variáveis qualitativas; W e Z são variáveis quantitativas. A tabela da Figura 1.1. representa Π bem como os valores que as variáveis X, Y, W, Z tomam em Π.

Estamos interessados na *distribuição de freqüências* de uma ou mais destas variáveis. Genericamente, dadas uma população Π e uma variável X aí definida, a *distribuição de freqüências (relativas)* de X (em Π) é a função que associa a cada valor possível de X, $v \in \mathcal{V}$, a sua freqüência relativa na população :

$$\begin{aligned} P(X = v) &= N_v/N && (1.2) \\ N_v &= \#\{X = v\} && (1.3) \\ \{X = v\} &= \{I \in \Pi : \; X(I) = v\} && (1.4) \end{aligned}$$

Em outras palavras, $\{X = v\}$ é o subconjunto de indivíduos de Π para os quais X vale v, N_v é o seu tamanho, e $P(X = v)$ a sua proporção em Π.

Definição 1.1 *Diremos que $P(X = v)$ é a* proporção *ou* freqüência *de $\{X = v\}$. Chamaremos também $P(X = \cdot)$ de* função de freqüência *da variável X.*[1] *Definimos também a freqüência de um subconjunto de valores de $\mathcal{V}$: se $U \subset \mathcal{V}$, então*

$$P(X \in U) = \sum_{v \in U} P(X = v) \qquad (1.5)$$

é a freqüência de U (mais precisamente, de $\{X \in U\} = \{I \in \Pi : \; X(I) \in U\}$, o subconjunto de indivíduos de Π para os quais $X \in U$).

No exemplo acima,

$$P(X = f) = 0.44, \; P(X = m) = 0.56$$

$$P(Y = 1) = 0.33, \; P(Y = 2) = 0.50, \; P(Y = s) = 0.17$$

$$P(W = 0) = 0.30, \; P(W = 1) = 0.22, \; P(W = 2) = 0.25,$$
$$P(W = 3) = 0.14, \; P(W = 4) = 0.06, \; P(W = 5) = 0.03$$

A distribuição de Z pode ser obtida de maneira análoga, mas note que neste caso, o resultado não é satisfatório (pois não há muita repetição em

[1] $P(X = \cdot) : \mathcal{V} \to [0,1]; P(X = v)$ dada em (1.2-1.4) para $v \in \mathcal{V}$.

X	freqüência
fem	0.44
masc	0.56

Y	freqüência
1º	0.33
2º	0.50
sup	0.17

Tabela 1.2

W	freqüência
0	0.30
1	0.22
2	0.25
3	0.14
4	0.06
5	0.03

Z	freqüência
$[4.00, 8.00)$	0.28
$[8.00, 12.00)$	0.33
$[12.00, 16.00)$	0.22
$[16.00, 20.00)$	0.14
$[20.00, 24.00)$	0.03

Tabela 1.3

$\mathcal{V}$ neste caso). Uma maneira mais adequada de representar tal distribuição seria considerar *classes* de salário. Por exemplo, os intervalos

$$\begin{aligned}
\mathcal{C}_1 &= [4.00, 8.00),\\
\mathcal{C}_2 &= [8.00, 12.00),\\
\mathcal{C}_3 &= [12.00, 16.00),\\
\mathcal{C}_4 &= [16.00, 20.00),\\
\mathcal{C}_5 &= [20.00, 24.00)
\end{aligned}$$

representam 5 classes de salário, e então

$$P(Z \in \mathcal{C}_1) = 0.28,\ P(Z \in \mathcal{C}_2) = 0.33,\ P(Z \in \mathcal{C}_3) = 0.22,$$
$$P(Z \in \mathcal{C}_4) = 0,14,\ P(Z \in \mathcal{C}_5) = 0.03,$$

onde $P(Z \in \mathcal{C}_i)$ é a proporção da i$^{\underline{a}}$ classe, isto é, $P(Z \in \mathcal{C}_i) = N_i/N$, com

$$N_i = \#\{I \in \Pi :\ X(I) \in \mathcal{C}_i\} \tag{1.6}$$

o número de indivíduos da população com salário na i$^{\underline{a}}$ classe, $i = 1, 2, \ldots$

É conveniente representarmos distribuições de freqüências em tabelas, como nas Tabelas 1.2 e 1.3.

Note que a soma das freqüências da distribuição de freqüências de uma variável naturalmente sempre totaliza 1:

$$\sum_{v \in \mathcal{V}} P(X = v) = 1. \tag{1.7}$$

No caso de divisão em classes:

$$\sum_{i=1}^{k} P(X \in \mathcal{C}_i) = 1, \tag{1.8}$$

onde k é o número de classes.

1.1 Representação gráfica de distribuições de freqüência

Nos gráficos da Figura 1.1, temos representações das distribuições de X, Y e W.

Trata-se de gráficos de barras, onde as alturas das barras representam as freqüências dos valores na escala sob as respectivas barras. (Na Figura 1.1, foram usadas escalas de tamanhos diferentes em cada gráfico, de forma que as alturas das barras são comparáveis apenas dentro de um mesmo gráfico, mas não entre diferentes gráficos.)

Note que no primeiro caso a escala (as distâncias entre valores no eixo das abscissas) é arbitrária e não tem maior significado. A ordem com que dispomos os valores (f e m) também é arbitrária. No segundo caso, a escala é arbitrária, mas a ordem é natural. Por isto dizemos que Y é uma variável *ordinal*, enquanto X é *nominal*. A variável W, e em geral todas as variáveis numéricas também são ordinais, já que para estas, a ordem é natural.

Para representar a distribuição de freqüências de Z, poderíamos fazer também um gráfico de barras. Veja a Figura 1.2.

Aqui, as barras têm como bases as classes. As alturas das bases representam, como acima, as freqüências relativas das classes. Esta representação gráfica é adequada quando as classes têm todas o mesmo tamanho. Mas quando este não for o caso, haverá distorção. Suponha que em vez da divisão

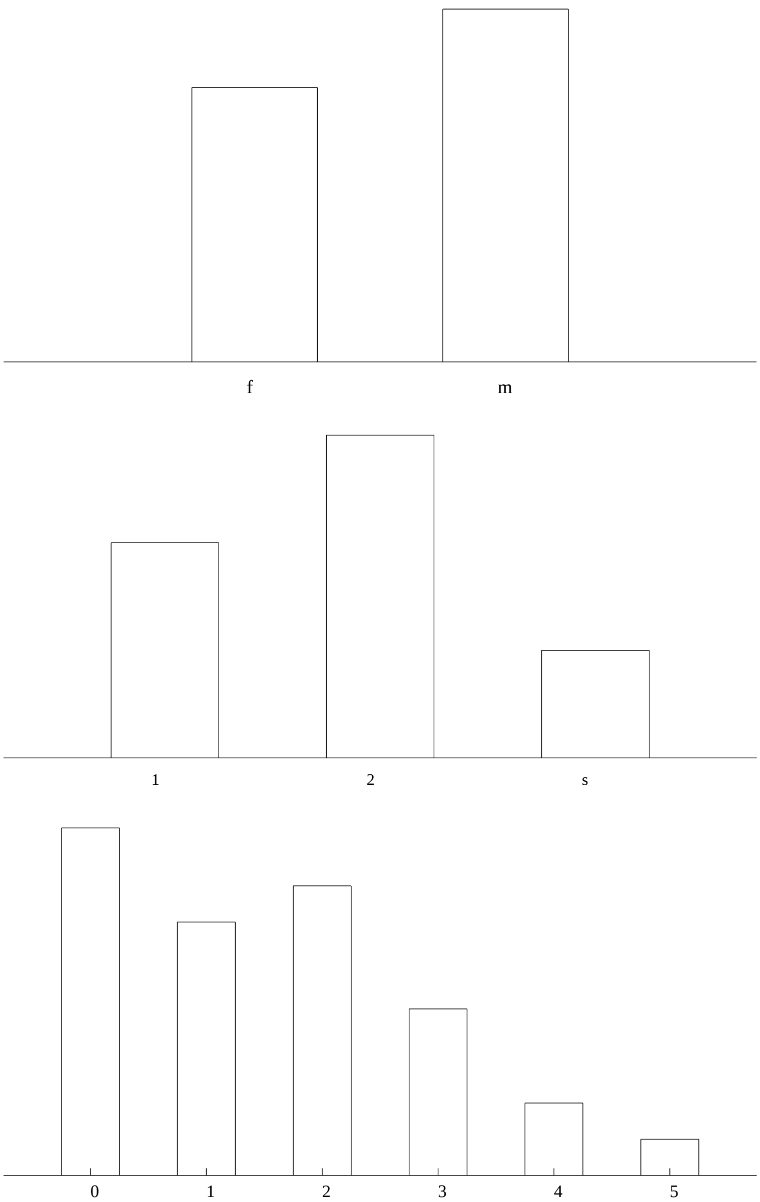

Figura 1.1 Gráficos de freqüências de X, Y, e W, respectivamente

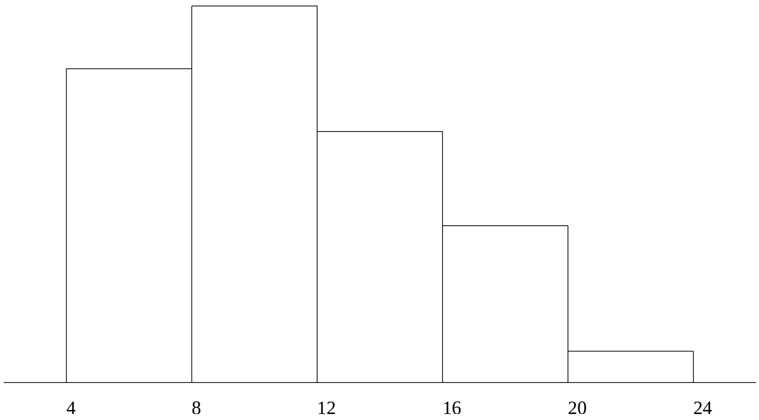

Figura 1.2 Representação gráfica de Z

em classes acima, tomássemos

$$\begin{aligned}
\mathcal{C}'_1 &= [4.00, 8.00),\\
\mathcal{C}'_2 &= [8.00, 12.00),\\
\mathcal{C}'_3 &= [12.00, 16.00),\\
\mathcal{C}'_4 &= [16.00, 24.00),
\end{aligned}$$

Neste caso, a distribuição seria

$$P(Z \in \mathcal{C}'_1) = 0.28,\ P(Z \in \mathcal{C}'_2) = 0.33,\ P(Z \in \mathcal{C}'_3) = 0.22,$$
$$P(Z \in \mathcal{C}'_4) = 0,17$$

e o gráfico de barras como na Figura 1.3.

As classes maiores parecerão mais representativas do que realmente são. Uma saída é tomarmos *densidades de freqüência*.

Para uma variável numérica X tomando valores em $\mathcal{V}$, seja

$$\{\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3, \ldots\} \tag{1.9}$$

uma divisão em classes de $\mathcal{V}$, e

$$P(X \in \mathcal{C}_1),\ P(X \in \mathcal{C}_2),\ P(X \in \mathcal{C}_3), \ldots \tag{1.10}$$

a distribuição de freqüências relativas de X *por classes*, onde $P(X \in \mathcal{C}_i)$ é a freqüência relativa de $\mathcal{C}_i$ para todo i. Seja então para $i = 1, 2, \ldots$

$$f_i = P(X \in \mathcal{C}_i)/\ell_i, \tag{1.11}$$

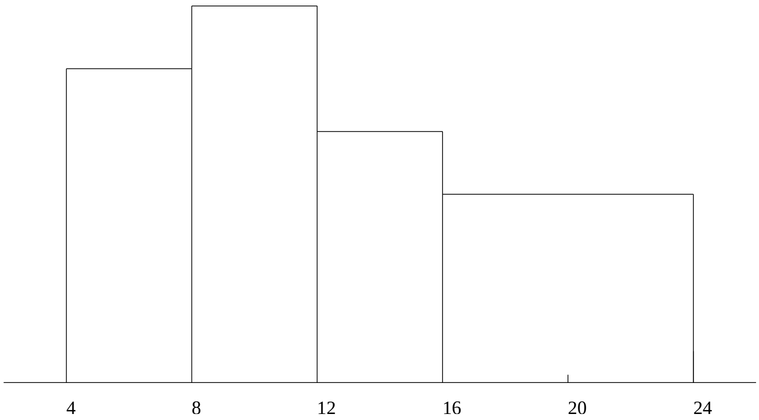

Figura 1.3 Representação gráfica de Z

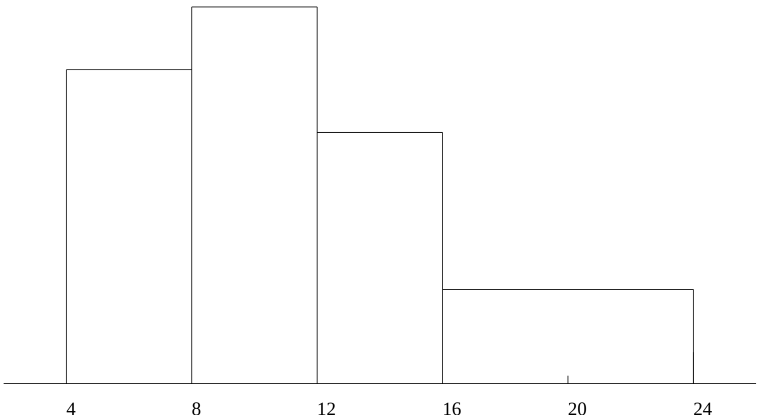

Figura 1.4 Gráfico de densidades de freqüência de Z

onde ℓ_i é o comprimento de $\mathcal{C}_i$, a *densidade de freqüência relativa de* X *em* $\mathcal{C}_i$. A *distribuição de densidades de freqüência relativa de* X é então dada por

$$f_i,\ i = 1, 2, \ldots \tag{1.12}$$

Seja agora o gráfico de barras em que as bases das barras são as classes e suas alturas as densidades de freqüência relativa das classes. Veja na Figura 1.4 o gráfico para Z com a última divisão em classes (em que as alturas representam as densidades respectivas).

Compare com os gráficos das Figuras 1.2 e 1.3 (note que estão em escalas de altura diferentes, mas o importante é notar a maior semelhança de forma,

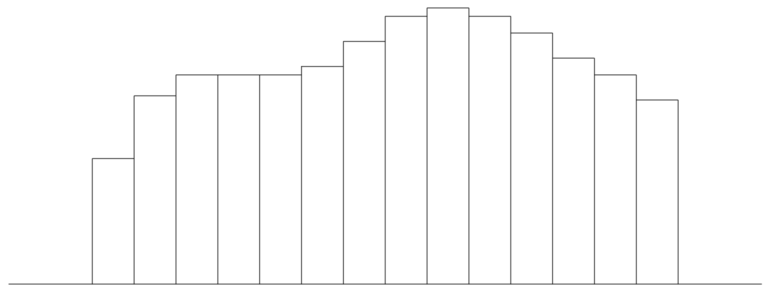

Figura 1.5 Histograma de X.

descontando a escala, entre os gráficos nas Figuras 1.2 e 1.4, do que entre os gráficos nas Figuras 1.2 e 1.3).

Este gráfico de densidades de freqüência relativa é o que chamamos de um *histograma* de Z.

Observação 1.2 *Em vez da altura, como nos gráficos de freqüência, nos histogramas são as* áreas *das barras que representam as freqüências das classes respectivas. De fato,*

$$\text{Área da i-ésima barra} = \textit{base} \times \textit{altura} = \ell_i \times f_i \overset{(1.11)}{=} P(X \in \mathcal{C}_i). \tag{1.13}$$

A área total sob o histograma (soma das áreas de todas as barras) vale portanto 1.

1.2 Função de densidade de freqüência

Suponha que $\mathcal{V}$ seja um intervalo de $\mathbb{R}$ (possivelmente infinito ou $\mathbb{R}$ todo), e tomemos classes bastante pequenas, de tal forma que o histograma de X fique com o aspecto da Figura 1.5.

Poderíamos traçar uma curva pelas extremidades superiores das barras, de forma a obter o desenho da Figura 1.6.

Se retirarmos agora o histograma original, obtendo o gráfico da Figura 1.7, teríamos o que poderíamos chamar de um *histograma pontual*, para o qual o histograma original por classes seria uma aproximação. (Podemos pensar no histograma pontual também como um limite do histograma original, quando os comprimentos das classes tendem a zero.)

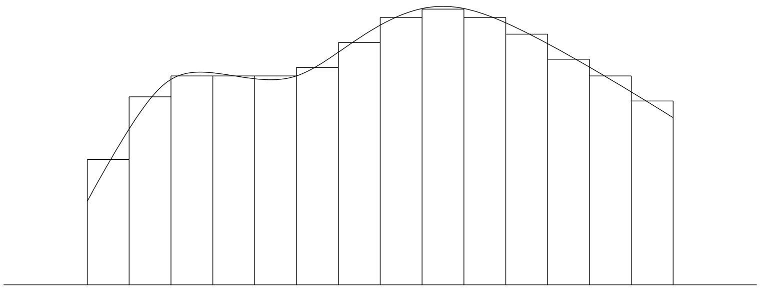

Figura 1.6 Curva superposta ao histograma de X.

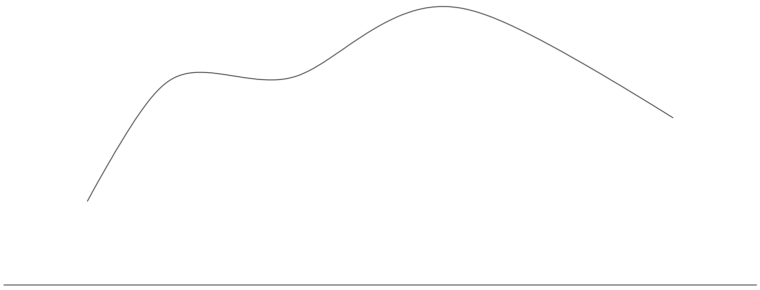

Figura 1.7 Histograma pontual de X

A curva contínua do histograma pontual pode ser vista como o gráfico de uma função $f = f_X$ que denominaremos *função de densidade de freqüência de* X.

A área total sob o histograma pontual, assim como no caso por classes, deve valer 1. Em termos de f, a condição é

$$\int_{-\infty}^{\infty} f(x)\, dx = 1.^2 \tag{1.14}$$

Note que dada f, podemos achar (aproximadamente) a freqüência com que X assume valores num dado intervalo (a, b) (denotada $P(a < X < b)$ como a área sob o gráfico de f (isto é, sob o histograma pontual) entre a e b, como indicado na Figura 1.8.

Em termos de f, teremos

$$P(a < X < b) = \int_a^b f(x)\, dx. \tag{1.15}$$

[2]Veja o Apêndice A para uma discussão breve e informal sobre integração.

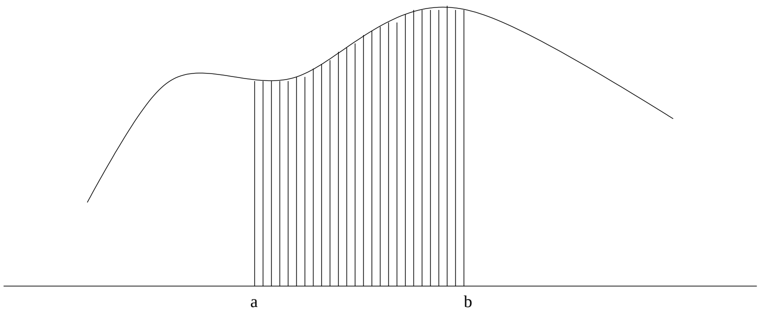

Figura 1.8 Área da região hachurada representa $P(a < X < b)$.

Exemplo 1.3 *Suponha que variável* X *tome valores em* $\mathcal{V} = [0, 1]$ *e que sua distribuição de freqüências em certa população* Π *seja dada (ou bem aproximada) pela função de densidade*

$$f(x) = \begin{cases} 3x^2, & \textit{se } 0 < x < 1; \\ 0, & \textit{caso contrário.} \end{cases}$$

Então,

$$P(1/2 < X < 3/4) = \int_{1/2}^{3/4} 3x^2\,dx = 3\frac{(3/4)^3 - (1/2)^3}{3} = \left(\frac{3}{4}\right)^3 - \left(\frac{1}{2}\right)^3 = \frac{19}{64},$$

o que representa a área na Figura 1.9. (Veja o Apêndice.)

Para variáveis numéricas como Z, podemos usar (e usaremos) a função de densidade de freqüência para representar sua distribuição de freqüências. Neste caso, diremos que Z é *contínua*. Quando a distribuição de Z for dada por uma função de freqüência p_Z, então diremos que Z é *discreta*.

Observação 1.4 *Se* Y *for uma variável contínua, então, para qualquer* y

$$P(Y = y) = \int_y^y f(x)\,dx = \textit{área sob gráfico de } f \textit{ do segmento acima de } y = 0. \tag{1.16}$$

Note que mesmo no caso do histograma por classes, podemos definir uma função de densidade de freqüência de Z como:

$$f(x) = \begin{cases} f_i, & \text{se } x \in \mathcal{C}_i \text{ para algum } i = 1, 2, \ldots, \\ 0, & \text{caso contrário.} \end{cases} \tag{1.17}$$

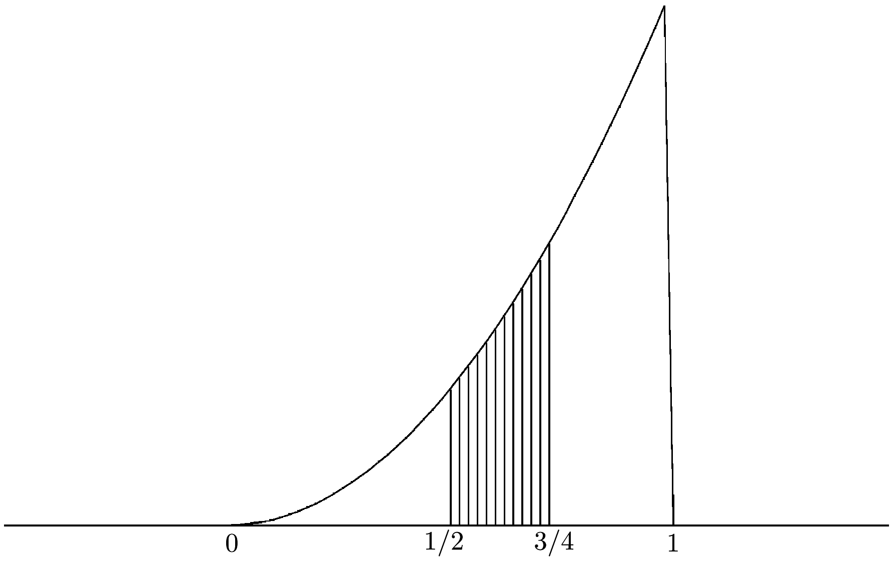

Figura 1.9 $P(1/2 < X < 3/4)$

Z	freqüência	densidade de freqüência
$[4.00, 8.00)$	0.28	0.070
$[8.00, 12.00)$	0.33	0.083
$[12.00, 16.00)$	0.22	0.055
$[16.00, 20.00)$	0.14	0.035
$[20.00, 24.00)$	0.03	0.008

Tabela 1.4

Vamos considerar a variável Z do Exemplo 1.1. Exibimos na Tabela 1.4 as freqüências e densidades de freqüência daquela variável. O gráfico de f_Z, obtido como em (1.17), aparece na Figura 1.10.

Observação 1.5 *Como fizemos no Exemplo 1.3 acima, vamos daqui para frente muitas vezes deixar implícita a população* Π *em que estiver definida uma variável cuja distribuição de freqüências for dada (seja através da função de freqüência, ou da função de densidade de freqüência, seja de outra forma).*

Observação 1.6 (Aditividade da freqüência de uma variável)

Tanto no caso em que é dada por função de freqüência, como quando é dada por função de densidade de freqüência, a freqüência de uma variável X,

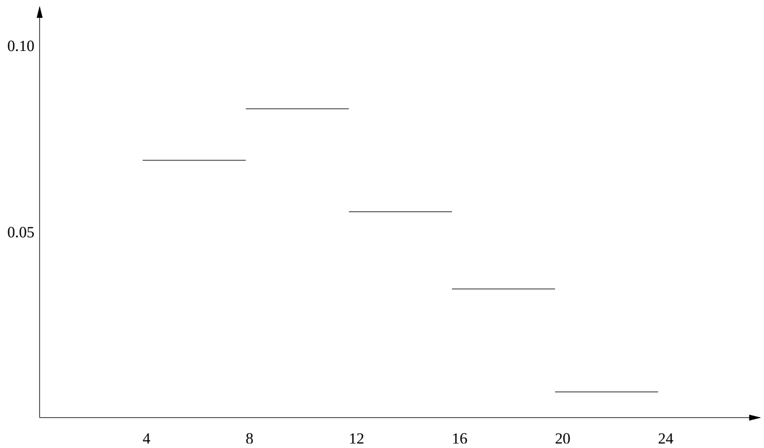

Figura 1.10 Densidade de Z

como na Definição 1.1, é aditiva, isto é, dados U *e* V *subconjuntos* disjuntos *de* $\mathcal{V}$ *(i.e.,* $U, V \subset \mathcal{V}$ *e* $U \cap V = \emptyset$*)*

$$P(X \in U \cup V) = P(X \in U) + P(X \in V). \tag{1.18}$$

Observação 1.7 (Complementaridade da freqüência)

A aditividade da freqüência tem o seguinte corolário. Sejam U *e* V *subconjuntos de* $\mathcal{V}$ *que além de disjuntos, são* exaustivos *(no sentido de que* $U \cup V = \mathcal{V}$*). Neste caso, dizemos que* U *e* V *são complementares, ou que* V *é complementar de* U*, e* U *é complementar de* V*; notação :* $V = U^c$, $U = V^c$. *De (1.18),* $1 = P(X \in \mathcal{V}) = P(X \in U \cup V) = P(X \in U) + P(X \in V)$*, e logo* $P(X \in V) = 1 - P(X \in U)$*. Em outras palavras,*

$$P(X \in U^c) = 1 - P(X \in U). \tag{1.19}$$

1.2.1 A densidade normal

Um caso particular muito importante é o da função de densidade normal dada por

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2},\ x \in \mathbb{R} \tag{1.20}$$

onde $\mu \in \mathbb{R}$ e $\sigma > 0$ são parâmetros.

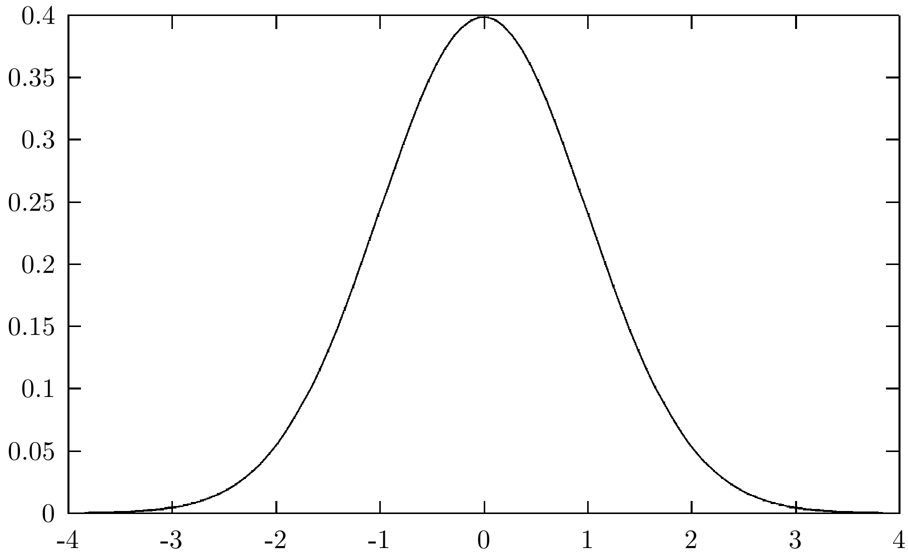

Figura 1.11 Gráfico de f para o caso padrão ($\mu = 0$, $\sigma = 1$).

Uma variável X com distribuição de freqüência dada por f é dita ter distribuição normal com os parâmetros μ e σ dados. Em símbolos,

$$X \sim N(\mu, \sigma^2).^3 \tag{1.21}$$

Veja a Figura 1.11.

μ é um parâmetro de localização da distribuição normal, mais especificamente do eixo de simetria de f. Veja a Figura 1.12.

σ por outro lado é uma medida de largura da distribuição normal. Veja a Figura 1.13.

O caso em que $\mu = 0$ e $\sigma = 1$ é dito *padrão*.

A importância da distribuição normal vem do fato, indicado pelo seu nome, de que ela é bastante comum. Ela também é conhecida como distribuição de Gauss ou Gaussiana. O gráfico da densidade normal será também chamado de *curva* normal.

Para que f seja de fato uma função de densidade de freqüência, é ne-

[3]Ao invés de σ, na notação colocamos o seu quadrado σ^2, seguindo uma convenção que será justificada adiante.

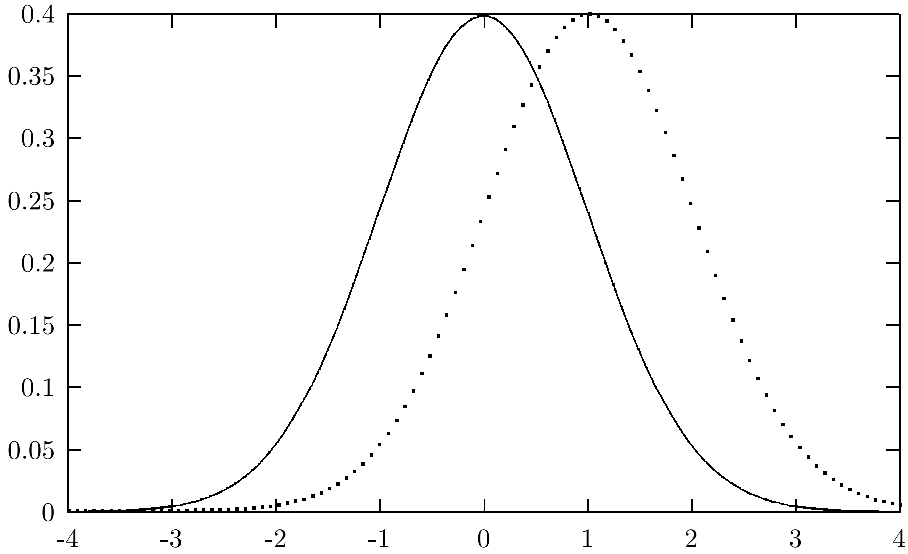

Figura 1.12 Gráficos da densidade normal nos casos padrão (linhas cheias) e $\mu = 1$, $\sigma = 1$ (linhas pontilhadas).

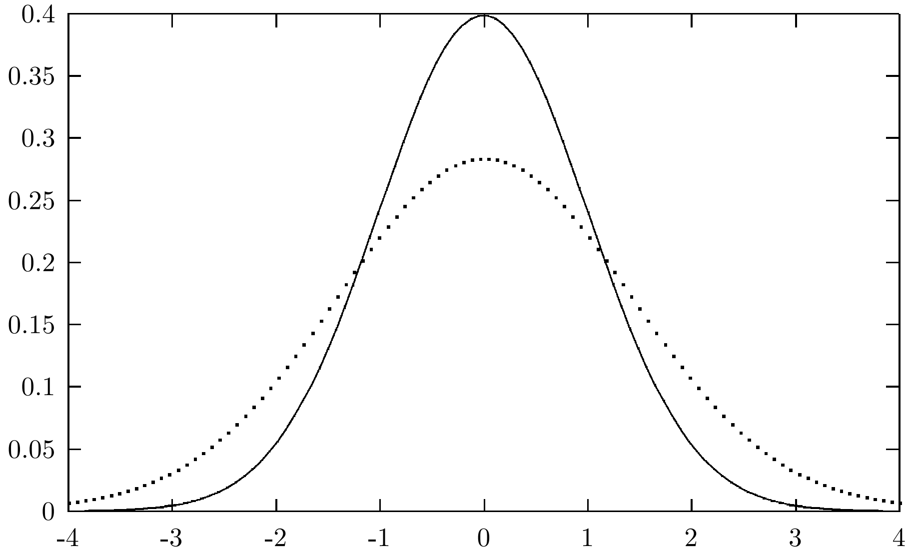

Figura 1.13 Gráficos do caso padrão (cheio) e o caso $\mu = 0$, $\sigma = \sqrt{2}$ (pontilhado).

cessário que

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\,dx = 1, \tag{1.22}$$

o que pode ser demonstrado. Expressões fechadas para

$$P(a < X < b) = \int_{a}^{b} f(x)\,dx = \int_{a}^{b} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\,dx, \tag{1.23}$$

não existem para $a, b \in \mathbb{R}$ genéricos, mas podem ser aproximadas numericamente. Tabelas com os resultados são então construídas e podemos utilizá-las em cálculos específicos, como veremos abaixo.

Padronização da distribuição normal

Uma propriedade útil da distribuição normal é a seguinte. Suponha que a variável X tenha uma distribuição normal com parâmetros μ_0 e $\sigma_0 > 0$ arbitrários. Se Z for a variável obtida de X segundo a transformação

$$Z = \frac{X - \mu_0}{\sigma_0}, \tag{1.24}$$

denominada *padronização* de X, então Z tem uma distribuição normal padrão (com parâmetros $\mu = 0$ e $\sigma = 1$). Em símbolos,

$$X \sim N(\mu_0, \sigma_0^2) \Rightarrow Z \sim N(0, 1). \tag{1.25}$$

Com isto, as freqüências de X podem ser obtidas daquelas de Z da seguinte forma. Para todo $x_1, x_2 \in \mathbb{R}$

$$P(x_1 < X < x_2) = P(z_1 < Z < z_2), \tag{1.26}$$

com $z_i = (x_i - \mu_0)/\sigma_0$, $i = 1, 2$.

Propriedades da distribuição normal padrão

No caso padrão ($\mu = 0$ e $\sigma = 1$), a densidade normal tem a forma

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2},\ x \in \mathbb{R}. \tag{1.27}$$

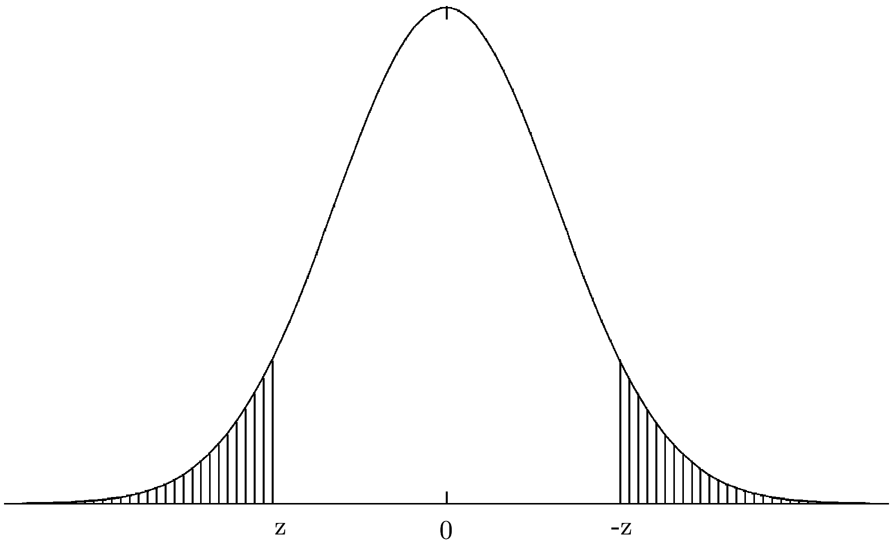

Figura 1.14

A Figura 1.11 representa o histograma pontual neste caso.

Esta densidade é simétrica em torno da origem, como se conclui de (1.27) e se vê na Figura 1.11.

Desta simetria, temos que para $z \in \mathbb{R}$

$$\int_{-\infty}^{z} f(x)\, dx = \int_{-z}^{\infty} f(x)\, dx, \tag{1.28}$$

o que é ilustrado na Figura 1.14 com $z < 0$.

Em termos de freqüências, se $Z \sim N(0,1)$, então

$$P(Z < z) = P(Z > -z). \tag{1.29}$$

Em particular, tomando $z = 0$, temos por um lado

$$P(Z < 0) + P(Z > 0) = \int_{-\infty}^{0} f(x)\, dx + \int_{0}^{\infty} f(x)\, dx = \int_{-\infty}^{\infty} f(x)\, dx = 1$$

por (1.22), e por (1.29)

$$P(Z < 0) + P(Z > 0) = 2P(Z > 0).$$

Concluímos que

$$P(Z < 0) = P(Z > 0) = \frac{1}{2}. \tag{1.30}$$

Seja agora a função $A : \mathbb{R} \to (0,1)$ tal que para $z \in \mathbb{R}$

$$A(z) = P(Z < z). \tag{1.31}$$

Então, se $a < b$

$$P(a < Z < b) = A(b) - A(a). \tag{1.32}$$

Valores de $A(z)$ são tabelados para diversos valores de z. Na tabela disponível no site do curso, os valores de z são positivos apenas, mas isto é suficiente pois, por (1.29), se $z < 0$

$$A(z) = P(Z > -z) = 1 - P(Z < -z) = 1 - A(-z) \tag{1.33}$$

onde a complementaridade da freqüência (veja a Observação 1.7 acima) foi usada na segunda igualdade.

Note que para dado z, $A(z)$ é a área sob a curva normal padrão abaixo de z. As Figuras 1.15 e 1.17 abaixo indicam nas regiões hachuradas tais áreas para dois z's diferentes.

Cálculo de freqüências normais

Combinando (1.26) com (1.32,1.33), podemos calcular freqüências de variáveis normais quaisquer em termos de A.

Exemplo 1.8 *Suponha que* $X \sim N(48, 625)$. *Note que* $\mu = 48$ *e* $\sigma^2 = 625$*,* *logo* $\sigma = 25$. *Então, a variável padronizada é* $Z = \frac{X-48}{25}$ *e*

$$\begin{aligned}
P(X<69) &\overset{(1.26)}{=} P(Z<0.84) \overset{(1.32)}{=} A(0.84) = 0.7995,\\
P(X>51) &\overset{(1.26)}{=} P(Z>0.12) \overset{(1.19)}{=} 1-P(Z<0.12) = 1-A(0.12)\\
&= 1-0.5478 = 0.4522,\\
P(51<X<69) &\overset{(1.26)}{=} P(0.12<Z<0.84) \overset{(1.32)}{=} A(0.84)-A(0.12)\\
&= 0.7995-0.5478 = 0.2517,\\
P(12<X<37) &\overset{(1.26)}{=} P(-1.44<Z<-0.44) \overset{(1.32)}{=} A(-0.44)-A(-1.44),\\
&\overset{(1.33)}{=} (1-A(0.44))-(1-A(1.44)) = A(1.44)-A(0.44)\\
&= 0.9251-0.6700 = 0.2551,\\
P(24<X<78) &\overset{(1.26)}{=} P(-0.96<Z<1.20) \overset{(1.32)}{=} A(1.20)-A(-0.96)\\
&\overset{(1.33)}{=} A(1.20)-(1-A(0.96)) = A(1.20)+A(0.96)-1\\
&= 0.8849+0.8315-1 = 0.0534,
\end{aligned}$$

onde os valores numéricos de A *foram obtidos da tabela disponível no site do curso.*

Invertendo o problema

Vamos inverter o problema e, ao invés de pedir a freqüência com que X assume valores em dado intervalo, vamos perguntar que intervalo corresponde a dada freqüência. Comecemos com o caso padrão.

Observamos primeiramente que a função A (veja (1.31)) é inversível. Isto é, existe uma função $B : (0,1) \to \mathbb{R}$ tal que

$$B(A(z)) = z, \quad \text{para todo } z \in \mathbb{R}. \tag{1.34}$$

Observação 1.9 *Valores de* $B(\alpha)$ *para diversos valores de* α *podem ser obtidos da tabela de* A *consultando-a de forma inversa.*

Exemplo 1.10 *Suponha que* $Z \sim N(0,1)$. *Qual o valor de* z *para o qual*

$$P(Z<z) = 0.90? \tag{1.35}$$

Veja a Figura 1.15.

Queremos z *tal que* $A(z) = 0.90$. *De (1.34),*

$$z = B(0.90) = 1.28.$$

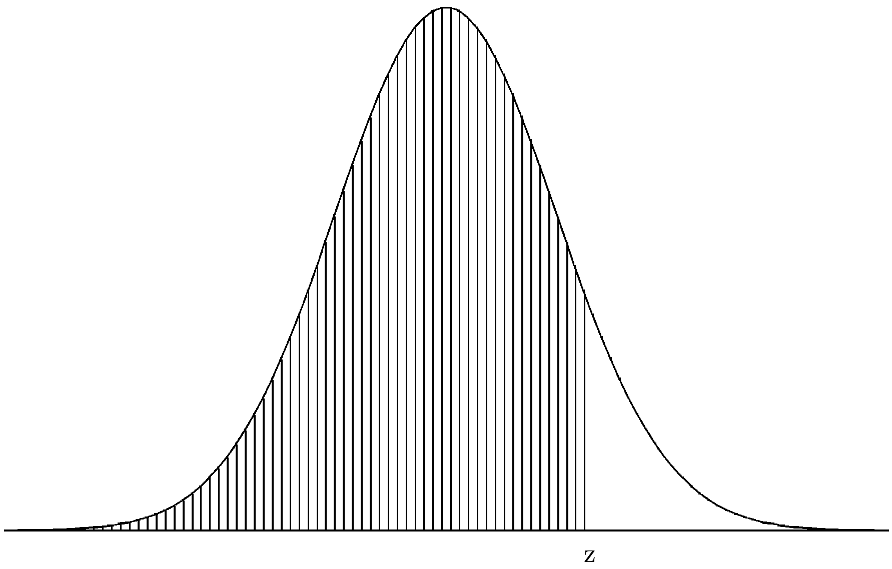

Figura 1.15

Observação 1.11 *Para achar $B(\alpha)$, procuramos α no corpo da tabela de A (aproximando pelo número mais próximo) e lemos o valor correspondente nas margens da tabela.*

Exemplo 1.12 *Se $Z \sim N(0,1)$, qual o valor de z para o qual*

$$P(Z > z) = 0.37? \tag{1.36}$$

Veja a Figura 1.16.

$A(z) = 1 - P(Z > z) = 1 - 0.37 = 0.63$. De (1.34),

$$z = B(0.63) = 0.33.$$

Exemplo 1.13 *Se $Z \sim N(0,1)$, qual o valor de z para o qual*

$$P(Z < z) = 0.22? \tag{1.37}$$

Notemos que z deve ser negativo, do contrário $P(Z < z)$ não poderia ser menor do que 0.50*. Veja a Figura 1.17.*

Neste caso, podemos escrever, pela simetria da distribuição normal padrão (veja (1.29)), $P(Z < z) = P(Z > -z)$ e estamos no caso do Exemplo 1.12.

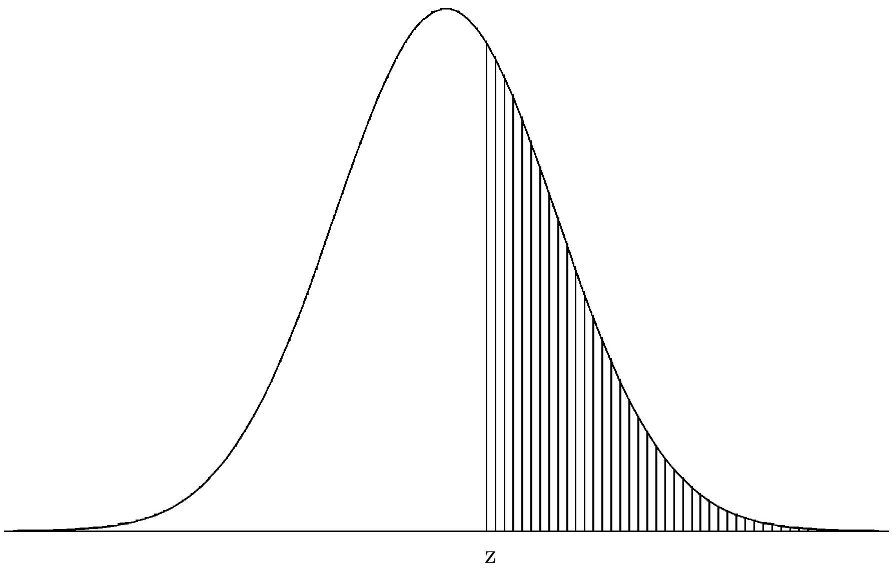

Figura 1.16

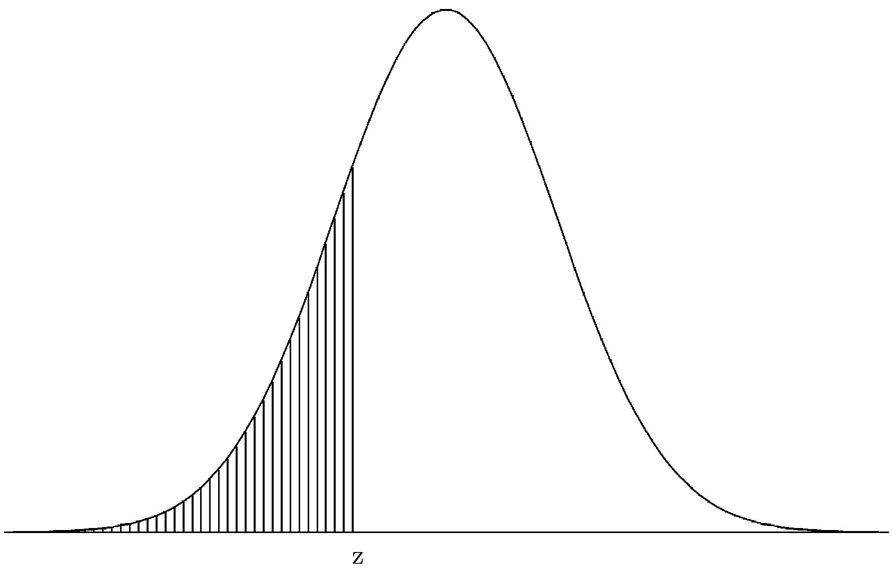

Figura 1.17

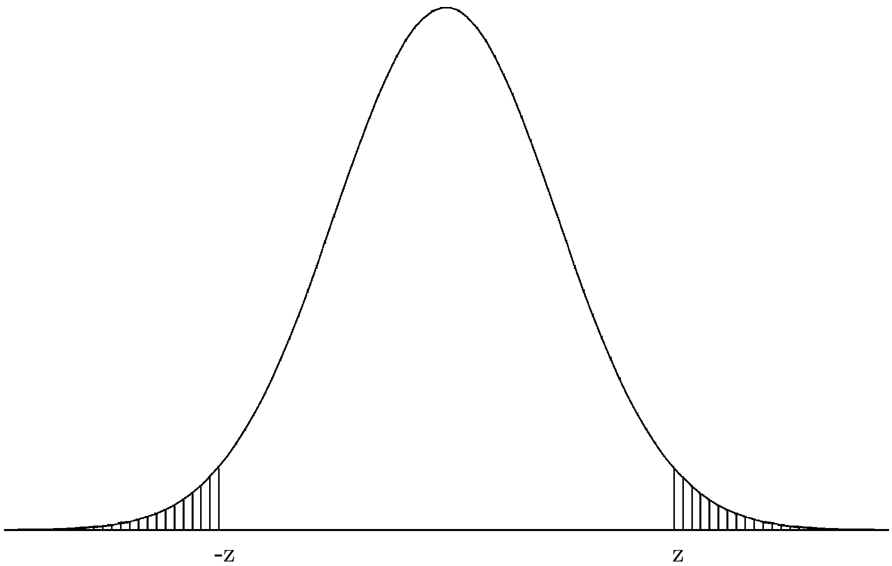

Figura 1.18

Temos que $P(Z > -z) = 1 - A(-z)$ *e logo* $A(-z) = 1 - 0.22 = 0.78$. *De (1.34),*

$$-z = B(0.78) = 0.77 \Rightarrow z = -0.77.$$

Exemplo 1.14 *Se* $Z \sim N(0,1)$, *qual o valor de* z *para o qual*

$$P(-z < Z < z) = 0.95? \tag{1.38}$$

Veja a Figura 1.18.

Pela simetria da distribuição normal padrão

$$\begin{aligned} P(-z < Z < z) &= 2P(0 < Z < z) = 2(A(z) - A(0)) \\ &= 2(A(z) - 1/2) = 2A(z) - 1. \end{aligned}$$

Logo, de (1.38)

$$A(z) = (1 + 0.95)/2 = 0.975.$$

De (1.34),

$$z = B(0.975) = 1.96.$$

O caso geral é similar, com a passagem adicional da padronização.

Exemplo 1.15 *Suponha como no Exemplo 1.8 que* $X \sim N(48, 625)$*. Qual o valor de* x *para o qual*

$$P(X > x) = 0.37? \tag{1.39}$$

Padronizando, i.e., fazendo $Z = (X - 48)/25$*, (1.39) é equivalente a*

$$P(Z > z) = 0.37, \tag{1.40}$$

onde

$$z = (x - 48)/25. \tag{1.41}$$

Do Exemplo 1.12 temos que $z = 0.33$*, e de (1.41)*

$$x = 48 + 0.33 \times 25 = 56.25. \tag{1.42}$$

Exemplo 1.16 *Se* $X \sim N(48, 625)$*. Qual o intervalo simétrico em torno do ponto de simetria que concentra* 95% *da freqüência?*

Queremos achar y *tal que*

$$P(48 - y < X < 48 + y) = 0.95? \tag{1.43}$$

Padronizando, (1.43) é equivalente a

$$P(-z < Z < z) = 0.95, \tag{1.44}$$

onde

$$z = y/25. \tag{1.45}$$

Do Exemplo 1.12 temos que $z = 1.96$*, e de (1.45)*

$$y = 1.96 \times 25 = 49. \tag{1.46}$$

Logo, o intervalo procurado é

$$[48 - y, 48 + y] = [-1, 97]. \tag{1.47}$$

1.3 Função de distribuição de freqüência

Para uma variável numérica Z, definimos a *função de distribuição (de freqüência acumulada)* como a função $F = F_X : \mathbb{R} \to [0, 1]$ tal que para todo $x \in \mathbb{R}$, $F(x)$ é a freqüência relativa com que X toma valores menores ou iguais a x. Se a distribuição de freqüências de X for dada pela função de freqüência $P(X=\cdot)$, então

$$F(x)=\sum_{v\leq x} P(X=v);$$

se a distribuição de freqüências de X for dada pela função de densidade de freqüência f_X, então

$$F(x)=\int_{-\infty}^{x} f_X(x)\,dx.$$

Em qualquer caso, podemos escrever

$$F(x)=P(X\leq x).$$

No Exemplo 1.1 acima, vamos achar a função de distribuição de W.

Para $x<0$, W não assume o valor x, logo, neste caso, temos

$$F_W(x)=P(W\leq x)=0,\ \text{se } x<0. \tag{1.48}$$

Para $0\leq x<1$, $\{W\leq x\}=\{W=0\}$. Logo

$$F_W(x)=P(W=0)=0.30,\ \text{se } 0\leq x<1. \tag{1.49}$$

Para $1\leq x<2$, $\{W\leq x\}=\{W\in\{0,1\}\}$. Logo

$$F_W(x)=P(W=0)+P(W=1)=0.52,\ \text{se } 1\leq x<2. \tag{1.50}$$

Para $2\leq x<3$, $\{W\leq x\}=\{W\in\{0,1,2\}\}$. Logo

$$F_W(x)=P(W=0)+P(W=1)+P(W=2)=0.77,\ \text{se } 2\leq x<3. \tag{1.51}$$

Para $3\leq x<4$, $\{W\leq x\}=\{W\in\{0,1,2,3\}\}$. Logo

$$\begin{aligned} F_W(x) &= P(W=0)+P(W=1) \\ &+ P(W=2)+P(W=3)=0.91,\ \text{se } 3\leq x<4. \end{aligned} \tag{1.52}$$

Para $4\leq x<5$, $\{W\leq x\}=\{W\in\{0,1,2,3,4\}\}$. Logo

$$\begin{aligned} F_W(x) &= P(W=0)+P(W=1)+P(W=2) \\ &+ P(W=3)+P(W=4)=0.97,\ \text{se } 4\leq x<5. \end{aligned} \tag{1.53}$$

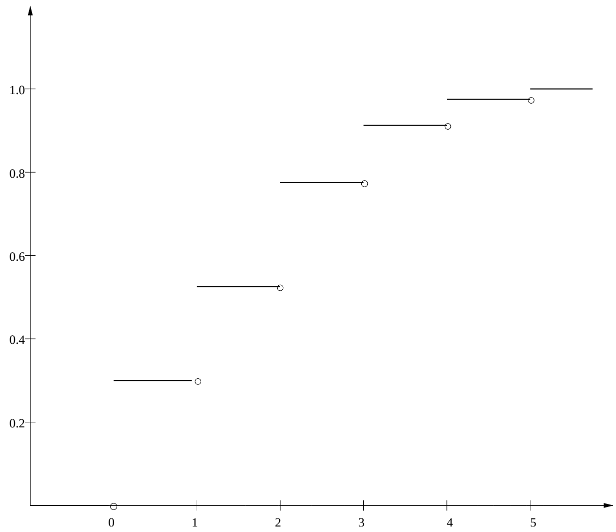

Figura 1.19 Gráfico de F_W

Finalmente, para $x > 5$, todos os valores assumidos por W estão abaixo de x. Logo

$$F_W(x) = 1, \text{ se } x > 5. \tag{1.54}$$

Podemos então resumir (1.48-1.54) da seguinte forma.

$$F_W(x) = \begin{cases} 0, & \text{se } x < 0, \\ 0.30, & \text{se } 0 \leq x < 1, \\ 0.52, & \text{se } 1 \leq x < 2, \\ 0.77, & \text{se } 2 \leq x < 3, \\ 0.91, & \text{se } 3 \leq x < 4, \\ 0.97, & \text{se } 4 \leq x < 5, \\ 1, & \text{se } x > 5. \end{cases} \tag{1.55}$$

Veja o gráfico de W na Figura 1.19.

Observação 1.17 *Note que F_W é uma* função escada*, isto é,* não decrescente *e* constante por intervalos*, com saltos nos pontos em que W toma valores.*

No Exemplo 1.3 acima, se $x < 0$

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f(y)\,dy = \int_{-\infty}^{x} 0\,dy = 0, \tag{1.56}$$

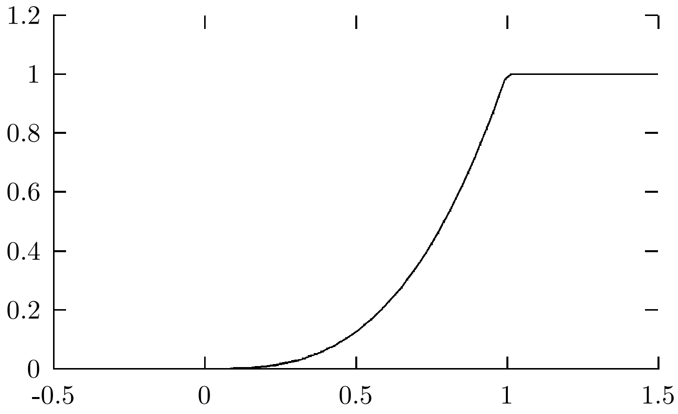

Figura 1.20 Gráfico de F_X

pois $f(y) = 0$, se $y < 0$.

Se $0 \leq x \leq 1$

$$\begin{aligned} F_X(x) = P(X \leq x) &= \int_{-\infty}^{x} f(y)\,dy = \int_{-\infty}^{0} 0\,dy + \int_{0}^{x} 3y^2\,dy \\ &= 0 + 3\frac{x^3 - 0^3}{3} = x^3. \end{aligned} \tag{1.57}$$

Finalmente, se $x > 1$,

$$\begin{aligned} F_X(x) = P(X \leq x) &= \int_{-\infty}^{x} f(y)\,dy \\ &= \int_{-\infty}^{0} 0\,dy + \int_{0}^{1} 3y^2\,dy + \int_{1}^{x} 0\,dy \\ &= 0 + 1 + 0 = 1. \end{aligned} \tag{1.58}$$

E podemos então resumir (1.56-1.58) da seguinte forma.

$$F_W(x) = \begin{cases} 0, & \text{se } x < 0, \\ x^3, & \text{se } 0 \leq x \leq 1, \\ 1, & \text{se } x > 1. \end{cases} \tag{1.59}$$

Veja o gráfico de F_X na Figura 1.20.

Observação 1.18 *Note que F_X é uma* função contínua.

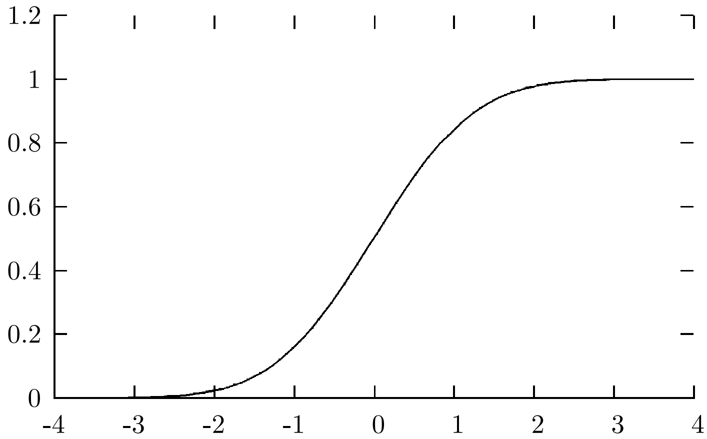

Figura 1.21 Função de distribuição da normal padrão

Na Figura 1.21, exibimos a função de distribuição de uma variável normal padrão (veja (1.27)). Note que também se trata de uma função contínua.

Em geral, a função de distribuição F de uma variável numérica X tem seguintes propriedades.

- É não decrescente;
- $\lim_{x\to-\infty} F(x) = 0$;
- $\lim_{x\to+\infty} F(x) = 1$;
- Se X for discreta, então F é uma função escada;
- Se X for contínua, então F é uma função contínua.

Observação 1.19 *F determina a distribuição de freqüências de X. Se X for discreta, então*

$$P(X = v) = P(X \leq v) - P(X < v) = F(v) - F(v-),$$

onde $F(x-) = \lim_{y\to x, y<x} F(y)$. Isto é, os pontos em que X toma valores (com freqüência positiva) são os pontos de salto de F (isto é, v's tais que $F(v) > F(v-)$), e as freqüências respectivas são os tamanhos dos saltos (isto é, $F(v) - F(v-)$).

E se X for contínua, então

$$f_X(x) = \frac{d}{dx}\int_{-\infty}^{x} f_X(x)\,dx = \frac{d}{dx}F(x).^4$$

1.4 Distribuição de freqüências de mais de uma variável

Suponha que Π é a população de funcionários do Exemplo 1.1 do início do capítulo. Consideremos as variáveis

$$\begin{aligned} X &= \text{Sexo} \\ W &= \text{Número de filhos} \end{aligned}$$

Queremos descrever a distribuição de freqüências da dupla (X, W), isto é, para os diversos valores possíveis (x, w) da dupla (X, W), queremos indicar a freqüência relativa respectiva. Por exemplo, para $(x, w) = (f, 2)$,

$$\begin{aligned} P(X = f, W = 2) &= \frac{\#\{I \in \Pi : X(I) = f, W(I) = 2\}}{36} \\ &= \frac{\#\{X = f, W = 2\}}{36} = \frac{4}{36} = 0.11. \end{aligned} \tag{1.60}$$

Fazendo o mesmo para todos os possíveis valores de (x, w), obtemos a distribuição de freqüência *conjunta* de (X, W):

$$P(X = x, W = w),\ \ (x, w) \text{ os valores possíveis de } (X, W). \tag{1.61}$$

Definição 1.2 *Como no caso com apenas uma variável, consideramos a função $P(X = \cdot, W = \cdot)$: valores possíveis de $(X, W) \to [0, 1]$, tal que $P(X = x, W = w)$ é a freqüência relativa com que (X, W) assume o valor (x, w) na população . Desta forma, chamaremos $P(X = \cdot, W = \cdot)$ também de* função de freqüência conjunta de (X, W).

De forma geral, dadas $X_1, \ldots, X_n$, n variáveis definidas em certa população Π, tomando valores em $\mathcal{V}_1, \ldots, \mathcal{V}_n$, respectivamente, a função de freqüência conjunta de $(X_1, \ldots, X_n)$ é dada pela função

$$P(X_1 = \cdot, \ldots, X_n = \cdot) : \mathcal{V}_1 \times \ldots \times \mathcal{V}_n \to [0, 1],$$

[4] A noção de derivada de uma função, indicada pelo símbolo $\frac{d}{dx}$, em particular sua relação com a integral (em muitos casos, como aqui, ela é a inversa da integral), é objeto do curso de Cálculo. Não entraremos em maiores detalhes nestas notas.

X \ W	0	1	2	3	4	5
f	0.11	0.14	0.11	0.06	0.03	0
m	0.19	0.08	0.14	0.08	0.03	0.03

Tabela 1.5

onde

$$P(X_1 = x_1, \dots, X_n = x_n) \\ = \frac{\#\{I \in \Pi : X_1(I) = x_1, \dots, X_n(I) = x_n\}}{N}.$$

Chamaremos $P(X_1 = \cdot, \dots, X_n = \cdot)$ *de* função de freqüência conjunta de $(X_1, \dots, X_n)$. *(Note que* $\mathcal{V}_1 \times \ldots \times \mathcal{V}_n$ *é o conjunto de valores possíveis de* $(X_1, \dots, X_n)$.*)*

Uma maneira prática de representar $P(X = \cdot, W = \cdot)$ é através de uma tabela de dupla entrada, como na Tabela 1.5.[5]

A distribuição conjunta de várias variáveis contém informação das distribuições de subconjuntos destas variáveis, em particular da distribuição de cada variável individualmente.

Proposição 1.20 *Sejam* $X_1, \dots, X_n$, $n \geq 2$, *variáveis com função de freqüência conjunta* $P(X_1 = \cdot, \dots, X_n = \cdot)$ *e seja* $1 \leq m < n$. *Então a função de freqüência conjunta* $P(X_1 = \cdot, \dots, X_m = \cdot)$ *de* $X_1, \dots, X_m$ *pode ser obtida de* $P(X_1 = \cdot, \dots, X_n = \cdot)$ *da seguinte forma. Dados* $x_1 \in \mathcal{V}_1, \dots, x_m \in \mathcal{V}_m$*,*

$$P(X_1 = x_1, \dots, X_m = x_m) = \\ \sum_{y_{m+1} \in \mathcal{V}_{m+1}, \dots, y_n \in \mathcal{V}_n} P(X_1 = x_1, \dots, X_m = x_m,\ X_{m+1} = y_{m+1}, \dots, X_n = y_n). \tag{1.62}$$

Note que a soma é sobre todos os valores possíveis $\mathcal{V}_{m+1} \times \ldots \times \mathcal{V}_n$ de $(X_{m+1}, \dots, X_n)$.

No exemplo acima,

$$P(W = 2) = P(X = f, W = 2) + P(X = m, W = 2) = 0.11 + 0.14 = 0.25.$$

[5]Nesta tabela, algumas aproximações numéricas foram feitas de forma não usual.

X \ W	0	1	2	3	4	5	total
f	0.11	0.14	0.11	0.06	0.03	0	0.44
m	0.19	0.08	0.14	0.08	0.03	0.03	0.56
total	0.30	0.22	0.25	0.14	0.06	0.03	

Tabela 1.6

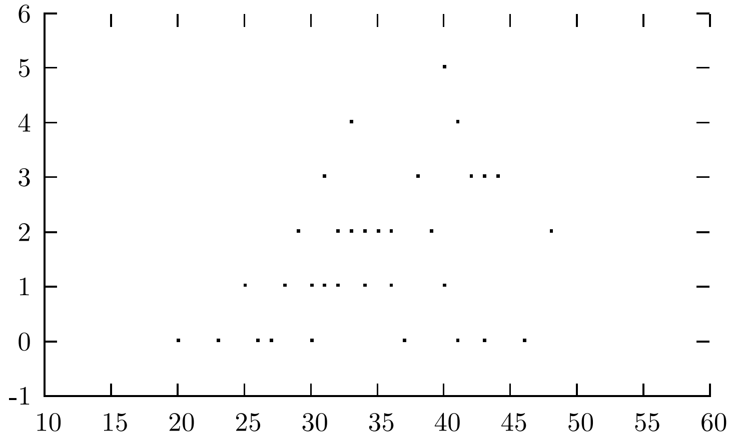

Figura 1.22 Diagrama de dispersão de $V \times W$

Na Tabela 1.6, isto é ilustrado ao tomarmos os totais das linhas e colunas da Tabela 1.5, obtendo desta forma nas margens da tabela as distribuições individuais de X e W. Disto vem que a distribuição individual de cada variável de um conjunto de variáveis é chamada de *distribuição marginal* da variável.

Diagrama de dispersão

Quando há pouca ou nenhuma repetição de valores de uma dupla de variáveis numéricas (X, Y), a tabela de dupla entrada não será muito informativa. Nestes casos, é muitas vezes conveniente representar os valores possíveis de (X, Y) no plano coordenado, num *diagrama de dispersão* de $X \times Y$. Nas Figuras 1.22 e 1.23, exemplificamos com $V \times W$ e $V \times Z$.

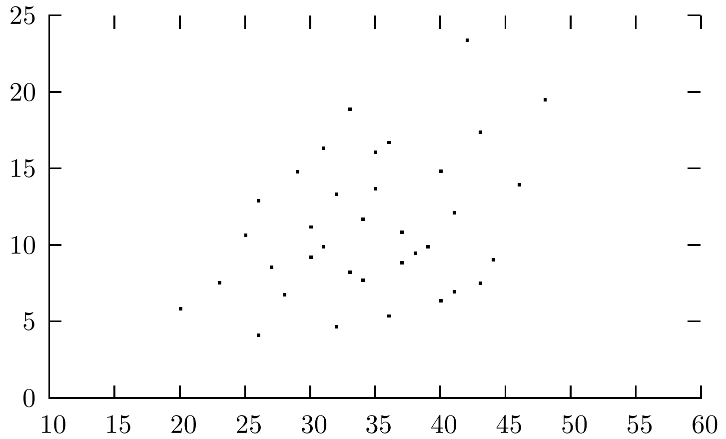

Figura 1.23 Diagrama de dispersão de $V \times Z$

1.4.1 Distribuição condicional e independência

A distribuição conjunta de várias variáveis dá informação sobre associações entre as variáveis. Vamos considerar as variáveis X e W do início da seção. Vamos por exemplo tomar as duas subpopulações de cada sexo. Isto é, vamos dividir a população Π segundo os valores de X. Vamos considerar agora a distribuição de W dentro destas duas subpopulações . Temos portanto duas distribuições, uma para cada subpopulação. Se estas duas distribuições forem iguais, então diremos que X e W são *não associadas* ou *independentes*; se as duas distribuições forem diferentes, então diremos que X e W são *associadas* ou *dependentes*.

Em outras palavras, sejam

$$\Pi_f = \{X = f\} = \{I \in \Pi : X(I) = f\} \tag{1.63}$$
$$\Pi_m = \{X = m\} = \{I \in \Pi : X(I) = m\} \tag{1.64}$$

as subpopulações dos sexos feminino e masculino respectivamente.

Então as distribuições de W em $P_f(W = \cdot)$ e $P_m(W = \cdot)$ são dadas por

$$\begin{aligned} P_v(W = i) &= \frac{\#\{I \in \Pi_v : W(I) = i\}}{\#\Pi_v} = \frac{\#\{I \in \Pi : X(I) = v, W(I) = i\}}{\#\Pi_v} \\ &= \frac{\#\{X = v, W = i\}}{\#\{X = v\}} \end{aligned} \tag{1.65}$$

para $v = f, m$ e $i = 0, \ldots, 5$.

Temos então

$$P_f(W=0)=0.25,\ P_f(W=1)=0.31,\ P_f(W=2)=0.25,$$
$$P_f(W=3)=0.13,\ P_f(W=4)=0.06,\ P_f(W=5)=0$$

$$P_m(W=0)=0.35,\ P_m(W=1)=0.15,\ P_m(W=2)=0.25,$$
$$P_m(W=3)=0.15,\ P_m(W=4)=0.05,\ P_m(W=5)=0.05,$$

o que mostra que X e W são associadas ou dependentes em Π.

Observação 1.21 *Se dividirmos o quociente à direita em (1.65) por $\#\Pi$, teremos*

$$P_v(W=i)=\frac{P(X=v,W=i)}{P(X=v)}. \tag{1.66}$$

As distribuições $P_f(W=\cdot)$ e $P_m(W=\cdot)$ são chamadas distribuições condicionais de W dado respectivamente $X=f$ e $X=m$. Usaremos daqui para frente a seguinte notação unificadora:

$$P(W=i|X=v)=\frac{P(X=v,W=i)}{P(X=v)},\ i=0,\ldots,5, \tag{1.67}$$

que será então a função de *distribuição condicional* de W dado $X=v$, $v=f,m$. Também diremos que

$$P(W=i|X=v),\ i=0,\ldots,5,\ v=f,m \tag{1.68}$$

é a função de distribuição condicional de W dado X.

Observação 1.22 *Da Proposição 1.20, temos que $P(X\in\cdot)$ pode ser expressa em termos da distribuição conjunta de (X,W), temos então que o lado direito de (1.66) pode ser escrito como*

$$\frac{P(X=v,W=i)}{\sum_{j=0}^{5}P(X=v,W=j)}, \tag{1.69}$$

o que mostra que a distribuição condicional é uma função da distribuição conjunta, isto é, pode ser obtida da distribuição conjunta.

Generalizando o exemplo, dada uma dupla de variáveis (X, Y) com (função de) distribuição conjunta $P(X = \cdot, Y = \cdot)$, a (função de) distribuição condicional de X dado Y é dada por

$$P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{P(X = x, Y = y)}{\sum_z P(X = x, Y = z)}; \ x, y, \quad (1.70)$$

onde x, y variam entre os valores possíveis de (X, Y).

Observação 1.23 *Note que $P(X = \cdot|Y = \cdot)$ é uma função de freqüência no primeiro argumento, mas* não *no segundo. Isto é*

$$\sum_x P(X = x|Y = y) = 1 \quad \textit{para todo } y, \qquad (1.71)$$

(veja (1.7)), mas pode ocorrer (e tipicamente ocorrerá)

$$\sum_y P(X = x|Y = y) \neq 1 \quad \textit{para algum } x. \qquad (1.72)$$

Observação 1.24 *Invertendo a definição em (1.70), temos*

$$P(X = x, Y = y) = P(X = x|Y = y)\, P(Y = y) \qquad (1.73)$$

para todo x, y. Temos então que

$$P(X = x) = \sum_y P(X = x, Y = y) = \sum_y P(X = x|Y = y)\, P(Y = y), \qquad (1.74)$$

onde a primeira igualdade vem da Proposição 1.20 e a segunda de (1.73). A Equação (1.74) nos diz que podemos obter a distribuição marginal de X a partir da distribuição condicional de X dado Y e da distribuição marginal de Y usando o lado direito em (1.74).

Independência

Como indicado na discussão acima, a condição de independência entre uma dupla de variáveis (X, Y) é que a distribuição condicional de X dado Y não dependa de Y, isto é, para todo x, $P(X = x|Y = y)$ não depende de

y, ou seja, é constante em y, digamos $\pi(x)$. Se (X, Y) forem independentes, então de (1.74), obtemos

$$\begin{aligned} P(X = x) &= \sum_y P(X = x|Y = y)\, P(Y = y) = \sum_y \pi(x)\, P(Y = y), \\ &= \pi(x) \sum_y P(Y = y) = \pi(x), \end{aligned} \tag{1.75}$$

já que $\sum_y P(Y = y) = 1$. Concluímos que se (X, Y) forem independentes, então

$$P(X = x|Y = y) = P(X = x) \quad \text{para todo } x, y. \tag{1.76}$$

De (1.73), temos que (1.76) é equivalente a

$$P(X = x, Y = y) = P(X = x)\, P(Y = y) \quad \text{para todo } x, y. \tag{1.77}$$

Podemos então tomar (1.77) como condição para independência. Em palavras, (X, Y) são variáveis independentes se sua função de distribuição conjunta *fatorar* nas funções marginais respectivas.

Observação 1.25 *Note que a condição (1.77) indica claramente a simetria da noção de independência entre X e Y; esta simetria não é tão clara, pelo menos a princípio, com a condição equivalente (1.76).*

Exemplo 1.26 *Suponha que as variáveis (X, Y) tenham distribuição conjunta dada na Tabela 1.7, em que as distribuições marginais de X e Y aparecem nas margens (como no caso das variáveis X e W do Exemplo 1.1 na Tabela 1.6). Note que Y assume valores no conjunto de categorias $\{a, b, c\}$ (e logo é categórica), enquanto X é numérica.*

Na Tabela 1.8 representamos em cada linha, correspondendo a cada valor y de Y, a distribuição de X dado $Y = y$. Como as distribuições condicionais apresentam diferenças entre si, concluimos que X e Y são dependentes.

A Tabela 1.9 representa a distribuição conjunta das variáveis categóricas (X', Y'). Elas são independentes. Uma forma de atestar esta afirmação é verificar a fatoração (1.77), e isto pode ser feito da seguinte maneira: para cada entrada do corpo da tabela, faça o produto dos respectivos valores nas margens, e verifique que o resultado equivale ao valor da entrada.

Duas ou mais variáveis

Y \ X	1	2	3	4	total
a	0.06	0.12	0.09	0.03	0.30
b	0.10	0.22	0.09	0.09	0.50
c	0.04	0.06	0.07	0.03	0.20
total	0.20	0.40	0.25	0.15	

Tabela 1.7

Y \ X	1	2	3	4
a	0.20	0.40	0.30	0.10
b	0.20	0.44	0.18	0.18
c	0.20	0.30	0.35	0.15

Tabela 1.8

Y' \ X'	x	y	z	total
a	0.16	0.20	0.04	0.40
b	0.24	0.30	0.06	0.60
total	0.40	0.50	0.10	

Tabela 1.9

Podemos estender a noção de condicionamento e independência para mais do que duas variáveis seguindo as idéias acima. Fazemos uma discussão breve a seguir.

Dadas $X_1, \ldots, X_n$, $n \geq 2$, variáveis com função de freqüência conjunta $P(X_1 = \cdot, \ldots, X_n = \cdot)$ e $1 \leq m < n$, a *distribuição condicional* de $(X_1, \ldots, X_m)$ dado $(X_{m+1}, \ldots, X_n)$ é dada pela função de distribuição condicional de $(X_1, \ldots, X_m)$ dado $(X_{m+1}, \ldots, X_n)$

$$\begin{aligned} &P(X_1 = x_1, \ldots, X_m = x_m | X_{m+1} = x_{m+1}, \ldots, X_n = x_n) \\ &\qquad = \frac{P(X_1 = x_1, \ldots, X_n = x_n)}{P(X_{m+1} = x_{m+1}, \ldots, X_n = x_n)}, \end{aligned} \tag{1.78}$$

onde $x_1, \ldots, x_n$ variam entre os valores possíveis de $(X_1, \ldots, X_n)$.

$P(X_1 = x_1, \ldots, X_m = x_m | X_{m+1} = x_{m+1}, \ldots, X_n = x_n)$ é uma distribuição conjunta de freqüência em $(x_1, \ldots, x_m)$ mas *não necessariamente* em $(x_{m+1}, \ldots, x_n)$.

Dizemos que $X_1, \ldots, X_n$ são independentes se a função de distribuição conjunta *fatorar*, isto é,

$$P(X_1 = x_1, \ldots, X_n = x_n) = P(X_1 = x_1) \ldots P(X_n = x_n) \tag{1.79}$$

para todo $x_1, \ldots, x_n$. Esta condição é equivalente a pedir que para qualquer par de subconjuntos disjuntos de variáveis $\{X'_1, \ldots, X'_j\}$, $\{X''_1, \ldots, X''_k\}$ de $X_1, \ldots, X_n$ (note que $j + k \leq n$), a distribuição condicional de $(X'_1, \ldots, X'_j)$ dado $(X''_1, \ldots, X''_k)$ não dependa dos valores das variáveis no último subconjunto.