

I.4

CÁLCULO DAS PROBABILIDADES

ADVERTÊNCIA PRÉVIA

Com o capítulo relativo a populações finitas, inicia-se a publicação das nossas lições de Cálculo das Probabilidades no Instituto Superior de Agronomia. Os capítulos seguintes serão publicados à medida que a experiência nos indicar qual a melhor orientação a seguir no ensino desta matéria, tendo sempre em conta a finalidade prática desse ensino e as condições especiais em que tem de ser realizado.

Na preparação das nossas lições serviram-nos de base, principalmente, as seguintes obras:

G. CASTELNUOVO - *Calcolo delle probabilita*, Zanichelli, Bologna, 1933.

D. J. FINNEY - *An introduction to statistical science in Agriculture*, John Wiley Sons, New York, 1953.

CRAMÉR - *Mathematical methods of Statistics*, Princeton University Press, Princeton, 1951.

G. UDN YULE and M. G. KENDALL - *An introduction to the Theory of Statistics*, Charles Griffin, Londres, 1937.

P. de VARENNES E MENDONÇA - *Noções de Cálculo das Probabilidades*, Instituto Superior de Agronomia, Lisboa, 1950.

Convirá, no entanto, precisar, desde já, que tanto a orientação geral do curso, como muitos dos pormenores didáticos têm carácter pessoal.

Lisboa, Maio de 1955

J. Sebastião e Silva

I.4.2

APONTAMENTOS DE CÁLCULO DAS PROBABILIDADES

A – Distribuições de uma variável contínua (real)

Consideremos na recta real, \mathbf{R} , um intervalo U de extremos a, b , que, para fixar ideias, vamos supor limitado e fechado: $U = [a, b]$. Seja x uma variável casual (comprimento, densidade, percentagem ou qualquer outra espécie de grandeza) que toma todos os seus valores no intervalo $[a, b]$. Trata-se, pois, agora, duma *variável casual contínua* (variável real).

Suponhamos que, para cada intervalo J contido em U , existe uma determinada *probabilidade de que o valor de x esteja em J* . Essa probabilidade representa-se por qualquer dos símbolos

$$\Pr(x \in J) \text{ ou } \Pr(J).$$

Em particular, se for $J = [x_1, x_2]$, poderá, ainda, escrever-se

$$\Pr(J) = \Pr(x_1 \leq x \leq x_2);$$

se for $J = [x_1, x_2[$, poderá escrever-se

$$\Pr(J) = \Pr(x_1 \leq x < x_2), \text{ etc.}$$

É claro que $\Pr(J)$, função numérica do intervalo variável $J \subset U$, deve satisfazer às seguintes condições (ou axiomas):

- I. $\Pr(J) \geq 0$, *qualquer que seja* J .
- II. $\Pr(J) = 1$, se $J = U$.
- III. $\Pr(J_1 + J_2) = \Pr(J_1) + \Pr(J_2)$, se J_1 e J_2 forem intervalos contíguos mas disjuntos.

Note-se que esta última condição pode apresentar-se com vários aspectos. Assim, se forem x_1, x_2, x_3 três pontos de U , tais que $x_1 < x_2 < x_3$, pode ter-se:

$$\begin{aligned} \Pr(x_1 \leq x \leq x_3) &= \Pr(x_1 \leq x < x_2) + \Pr(x_2 \leq x \leq x_3) \\ &= \Pr(x_1 \leq x \leq x_2) + \Pr(x_2 < x \leq x_3) \end{aligned}$$

$$\Pr(x_1 \leq x < x_3) = \Pr(x_1 \leq x < x_2) + \Pr(x_2 \leq x < x_3), \text{ etc.}$$

Dum modo geral, quando $x_1 = x_2$, identificaremos o intervalo $[x_1, x_2]$ com o ponto x_1 e escreveremos

$$\Pr([x_1, x_1]) = \Pr(x = x_1) = \Pr(x_1).$$

Verificadas as referidas condições, diremos que a função $\Pr(J)$ é uma *distribuição de x definida sobre o intervalo U* (universo infinito, visto ser formado por uma infinidade de pontos).

Das condições I, II e III, resulta que é sempre

$$0 \leq \Pr(J) \leq 1.$$

Também é fácil ver que, por exemplo:

$$\Pr(x_1 \leq x \leq x_2) = \Pr(x_1 \leq x < x_2) + \Pr(x_2).$$

Outras consequências poder-se-iam deduzir, ainda, de tais axiomas.

NOTA. Às condições anteriores pode juntar-se, para fins teóricos, uma outra (*axioma de continuidade*), que enunciaremos nos seguintes termos:

IV. Dada uma sucessão infinita (J_n) de intervalos tais que

$$J_1 \subset J_2 \subset \dots \subset J_n \subset \dots,$$

sendo J a reunião de todos estes intervalos, ter-se-á

$$\Pr(J) = \lim_{n \rightarrow \infty} \Pr(J_n).$$

Em estudos de nível mais elevado, os axiomas III e IV são substituídos por um outro, mais geral, relativo à soma de sucessões infinitas de conjuntos C_n disjuntos dois a dois. O axioma traduz-se, simplesmente, pela fórmula

$$\Pr\left(\sum_{n=1}^{\infty} C_n\right) = \sum_{n=1}^{\infty} \Pr(C_n).$$

Os conjuntos C_n podem ser intervalos ou quaisquer outros conjuntos que se possam construir a partir dos intervalos pela aplicação sucessiva ou alternada do símbolo

$$\sum_1^{\infty}$$

e de passagens ao complementar: dá-se-lhes o nome de *conjuntos borelianos* ou *conjuntos de BOREL*.

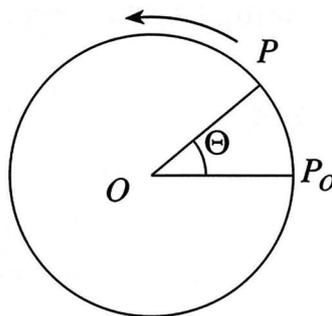


Fig. 1

Exemplo – Numa roleta “perfeita”, cada ponto P da sua circunferência é determinado por um número real Θ , igual ou superior a 0 e

menor que 2π , que é o arco $\widehat{P_0P}$, medido em radianos num sentido prefixo, tomando como origem um ponto P_0 . É claro que a correspondência $P \mapsto \Theta$ assim estabelecida, entre os pontos da circunferência e os pontos do intervalo $[0, 2\pi[$, é biunívoca. As probabilidades correspondentes a dois arcos iguais serão também iguais (na hipótese de a roleta ser perfeita). Por sua vez, a probabilidade correspondente à circunferência será

$$\Pr(0 \leq \Theta \leq 2\pi) = 1.$$

Então, atendendo à condição III, é fácil ver que

$$\begin{aligned} \Pr(\Theta_1 \leq \Theta \leq \Theta_2) &= \Pr(\Theta_1 < \Theta < \Theta_2) \\ &= \frac{\Theta_2 - \Theta_1}{2\pi}. \end{aligned}$$

Em particular, tem-se

$$\Pr(\Theta = \Theta_1) = \frac{\Theta_1 - \Theta_1}{2\pi} = 0,$$

isto é, a probabilidade de cada valor particular Θ_1 de Θ é *nula*; *não quer isto dizer, porém, que qualquer dos acontecimentos individuais $\Theta = \Theta_1$ seja impossível, pois que um deles há-de realizar-se, necessariamente, em cada prova*. O que podemos dizer, ainda aqui, é que se trata de acontecimentos *praticamente impossíveis*.

Dada uma distribuição de probabilidade, $\Pr(J)$, sobre o intervalo $U = [a, b]$, o número

$$\Pr(x \leq u) = \Pr([a, u])$$

é, manifestamente, uma função de u , a que chamaremos *cumulant da distribuição*. Representamo-la por $\Phi(u)$ ou, mesmo, por $\Phi(x)$, tomando x para variável independente, em vez de u . É claro que, aumentando u , $\Phi(u)$ não pode diminuir, em virtude das condições I, II, III: *a função $\Phi(x)$ é, pois, crescente em sentido lato no intervalo U* .

NOTA. Do axioma IV deduz-se que

$$\Pr(x < u) = \lim_{x \rightarrow u^-} \Phi(x) = \Phi(u^-).$$

Com efeito, dada uma sucessão *crescente* de pontos x_n , convergente para u , o intervalo $[a, u[$ será a reunião de todos os intervalos $[a, x_n]$ e, portanto, a probabilidade de x estar em $[a, u[$, ou seja, $\Pr(x < u)$, será o limite, quando $n \rightarrow \infty$, da probabilidade de x estar em $[a, x_n]$. Tem-se, pois, $\lim_{n \rightarrow \infty} \Pr(x < x_n) = \lim_{x \rightarrow u^-} \Pr(x \leq u) = \Phi(u^-)$.

Então, será

$$\Pr(x_1 \leq x \leq x_2) = \Phi(x_2) - \Phi(x_1^-),$$

e, analogamente,

$$\Pr(x_1 < x < x_2) = \Phi(x_2^-) - \Phi(x_1),$$

$$\Pr(x_1 < x \leq x_2) = \Phi(x_2) - \Phi(x_1),$$

$$\Pr(x_1 \leq x < x_2) = \Phi(x_2^-) - \Phi(x_1^-).$$

Assim, *toda a distribuição* $\Pr(J)$ *é determinada pela sua função cumulante, $\Phi(x)$.*

É claro que será

$$\Pr(x) = \Phi(x) - \Phi(x^-), \text{ para todo o } x \in U.$$

$$\Phi(b) = \Pr(U) = 1.$$

A função $\Phi(x)$ será contínua à direita em todos os pontos, isto é: $\Phi(x^+) = \Phi(x)$, qualquer que seja x .

Em particular, pode suceder que, num ponto x_0 , se tenha $\Phi(x_0^-) = \Phi(x_0)$. Então, a função $\Phi(x)$ é contínua em x_0 e a probabilidade deste ponto é nula, visto que $\Phi(x_0) - \Phi(x_0^-) = 0$.

Um caso particular importante é aquele em que a função $\Pr(x)$ do ponto x é nula, excepto num número finito de pontos x_1, x_2, \dots, x_n do intervalo U : recaímos, então, no caso já estudado das variáveis casuais descontínuas, com um número finito de valores. Neste caso, a

probabilidade $\Pr(J)$, correspondente a um dado intervalo J , será a soma $\sum \Pr(x_i)$ das probabilidades dos valores x_i situados em J . A função cumulante $\Phi(x)$ será, pois, neste caso,

$$\Phi(x) = \sum_{x_i \leq x} \Pr(x_i),$$

sendo fácil ver que uma tal função $\Phi(x)$ apresenta uma descontinuidade de 1ª espécie em cada um dos pontos x_i , com um salto positivo igual a $\Pr(x_i)$:

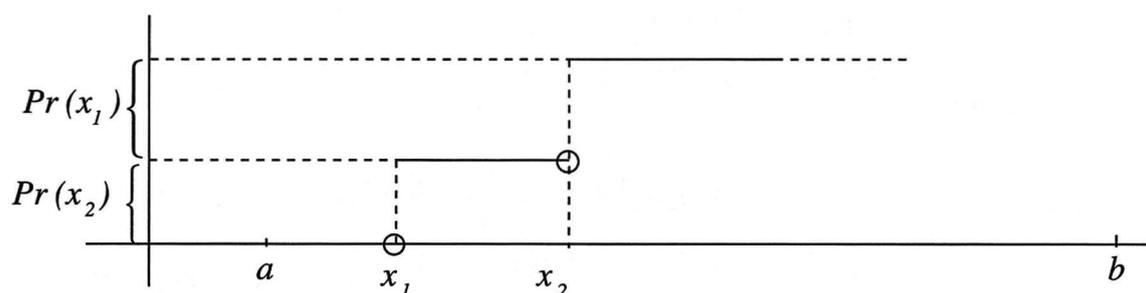


Fig. 2

Fique, pois, assente que o caso das variáveis descontínuas, com um número finito de valores x_1, x_2, \dots, x_r , se pode sempre englobar no caso das variáveis contínuas, desde que se atribua probabilidade nula a todo o valor de x diferente daqueles.

Um outro caso particular importante, mas que já não se reduz ao caso dos universos finitos, é aquele em que a cumulante $\Phi(x)$ admite derivada contínua no intervalo U . Procuremos o significado desta derivada.

Começemos por notar que, sendo neste caso $\Phi(x)$ uma função contínua, não apresenta saltos. Então, a probabilidade de cada valor de $x \neq a$ é nula,⁽¹⁾ tendo-se, pois, sempre:

$$\Pr(x_1 < x < x_2) = \Pr(x_1 \leq x \leq x_2).$$

(1) – Veja-se nota precedente.

Suponhamos que é também $\Pr(a) = 0$. Nestas condições, tem-se, para todo o ponto x_0 de U :

$$\Phi(x_0 + h) - \Phi(x_0) = \begin{cases} \Pr(x_0 \leq x \leq x_0 + h), & \text{para } h > 0 \\ \Pr(x_0 + h \leq x \leq x_0), & \text{para } h < 0 \end{cases}$$

À razão incremental

$$\frac{\Phi(x_0 + h) - \Phi(x_0)}{h}$$

podemos, então, chamar *densidade média de probabilidade no intervalo de extremos $x_0, x_0 + h$* . Por sua vez, à derivada $\Phi'(x_0)$, *limite da razão incremental quando $h \rightarrow 0$* (que existe, por hipótese), será natural chamar *a densidade de probabilidade no ponto x_0* .

Representemos por $\varphi(x)$ a derivada de $\Phi(x)$ no intervalo U . Então, segundo o teorema fundamental do cálculo integral, será

$$\Phi(u) = \Phi(a) + \int_a^u \varphi(x) dx = \int_a^u \varphi(x) dx.$$

É claro que o diferencial $d\Phi = \varphi(x) dx$ (*probabilidade elementar*), representa, a menos de um infinitésimo de ordem superior à de dx , a probabilidade correspondente ao intervalo infinitésimo $[x, x + dx]$.

Será, ainda, evidentemente:

$$\int_a^b \varphi(x) dx = \Phi(b) = 1.$$

Estas considerações tornam-se mais intuitivas, se imaginarmos a função $\Pr(J)$ como indicando uma distribuição de matéria, de massa total 1, sobre o intervalo $[a, b]$: então, $\varphi(x)$ representará a densidade (ou melhor, a massa específica) no ponto variável x .

Exemplos – 1) No caso duma roleta perfeita, tem-se

$$\Phi(u) = Pr(0 \leq \Theta \leq u) = \frac{u}{2\pi}$$

e, portanto, $\varphi(u) = \Phi'(u) = \frac{1}{2\pi}$. A densidade de probabilidade é, portanto, a mesma em todos os pontos. Mas basta que a roleta não esteja bem centrada ou que não seja homogénea, para que a densidade varie de ponto para ponto.

2) Vejamos, agora, um outro exemplo sugestivo que se apresenta na teoria dos seguros. A probabilidade de que uma criança recém-nascida venha a falecer antes duma certa idade x pode considerar-se como função da variável contínua x , definida num intervalo $[0, L]$, em que L representa um majorante da duração possível da vida humana. Supondo que esta função admite derivada contínua, a probabilidade de que a criança venha a viver até uma idade compreendida entre x_1 e x_2 será

$$\Phi(x_2) - \Phi(x_1) = \int_{x_1}^{x_2} \varphi(x) dx;$$

mas, neste caso, a densidade $\varphi(x)$ é função decrescente de x . Por sua vez, a probabilidade de que uma pessoa de idade a venha a viver até uma idade x entre x_1 e x_2 (com $x_1 > a$) será:

$$\frac{\int_{x_1}^{x_2} \varphi(x) dx}{\int_a^L \varphi(x) dx}$$

probabilidade condicional do acontecimento $x_1 < x < x_2$ a respeito do acontecimento $x \geq a$ (substituição do universo $[0, L]$, pelo universo $[a, L]$).

As precedentes considerações generalizam-se imediatamente ao caso dum intervalo não limitado, por exemplo, o intervalo $]-\infty, +\infty[$. Será, então, $U = \mathbf{R}$. Neste caso, se a cumulante $\Phi(u) = \Pr(x \leq u)$ admite derivada contínua $\varphi(u)$ em todos os pontos, ter-se-á, ainda,

$$\Pr(v \leq x \leq u) = \Phi(u) - \Phi(v) = \int_v^u \varphi(x) dx.$$

Como se deve ter, além disso⁽¹⁾,

$$\Pr(x \leq u) = \lim_{v \rightarrow -\infty} \Pr(v \leq x \leq u),$$

será

$$\Phi(u) = \int_{-\infty}^u \varphi(x) dx.$$

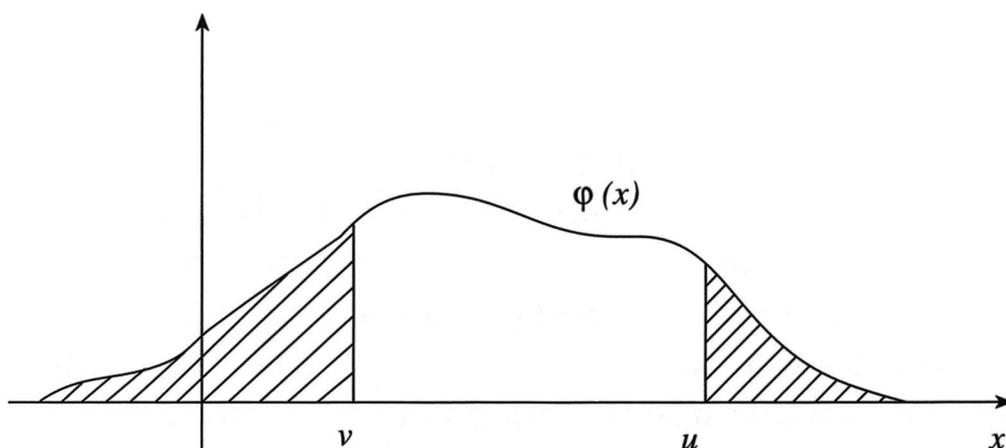


Fig. 3

Se for C a curva representativa da função $\varphi(x)$, o valor de $\Phi(u)$ será a área do domínio limitado por C e pelo eixo dos xx , à esquerda da recta $x = u$. A região a tracejado indica na figura a probabilidade de x estar *fora do intervalo* $[v, u]$, probabilidade esta igual a $1 - \Pr(v \leq x \leq u)$, sendo, por sua vez,

$$\Pr(v \leq x \leq u) = \int_v^u \varphi(x) dx \quad (\text{região a branco}).$$

(1) – Em virtude do axioma IV, enunciado numa nota precedente.

É claro que⁽²⁾

$$\int_{-\infty}^{+\infty} \varphi(x) dx = \Pr(U) = 1 \quad (\text{domínio total}).$$

Note-se, ainda, que toda a distribuição de probabilidade sobre um intervalo $[a, b]$ se pode conceber como distribuição sobre a recta inteira, considerando como nula a probabilidade correspondente a qualquer intervalo contido no complementar de $[a, b]$.

NOTA. Dum modo geral, sendo $\Pr(J)$ uma distribuição sobre um intervalo U , ter-se-á, naturalmente:

$$(1) \quad \Pr(J_1 + J_2 + \dots + J_n) = \Pr(J_1) + \Pr(J_2) + \dots + \Pr(J_n)$$

para todo o sistema finito de intervalos J_1, J_2, \dots, J_n , disjuntos dois a dois, *mesmo que estes não sejam contíguos.*

Se representarmos por \mathcal{H} a totalidade dos conjuntos C contidos em U , que são somas de intervalos, em número finito, disjuntos dois a dois, e se incluirmos em \mathcal{H} o conjunto vazio, é fácil ver que:

1) – *A soma lógica (ou produto lógico) de dois quaisquer conjuntos da família \mathcal{H} ainda é um conjunto desta família.*

2) – *O complementar de qualquer conjunto de \mathcal{H} a respeito de U ainda é um conjunto de \mathcal{H} (por exemplo, o complementar dum intervalo é, geralmente, a soma de dois intervalos disjuntos).*

Exprimiremos este facto dizendo que a família \mathcal{H} é *um corpo de conjuntos* (o que não sucede com a família dos intervalos, pois que a soma de dois intervalos ou o complementar dum intervalo pode não ser um intervalo).

Chamaremos *distribuição em \mathcal{H}* a toda a função real não negativa $\mu(C)$, definida em \mathcal{H} , de modo que se tenha:

$$\mu(C_1 + C_2) = \mu(C_1) + \mu(C_2)$$

quando C_1, C_2 são conjuntos disjuntos da família \mathcal{H} .

(2) – Este facto é, ainda, uma consequência do axioma IV.

Deste modo, a função $\text{Pr}(C)$ será uma distribuição em \mathcal{H} a qual verifica a condição suplementar seguinte: $\text{Pr}(U) = 1$ (*distribuição relativa*).

É claro que, para definir uma distribuição em \mathcal{H} , basta defini-la na família dos intervalos contidos em U , atendendo à fórmula (1). Mas ao contrário do que sucede com os corpos finitos, não é agora suficiente, em geral, conhecer a distribuição nas células do corpo, que são, neste caso, os pontos de U . Como vimos, pode até acontecer que as probabilidades dos pontos sejam todas nulas sem que o sejam a dos intervalos não nulos: é o que sucede nos casos em que existe em cada ponto uma densidade de probabilidade finita e diferente de 0.

Note-se, ainda, que a toda a função não negativa $\Phi(x)$, definida em U , crescente em sentido lato e contínua à direita, corresponde uma distribuição $\mu(C)$ sobre U de que $\Phi(x)$ é a cumulante, isto é, tal que

$$\Phi(u) = \mu(x \leq u).$$

De resto, além do corpo \mathcal{H} , existem outros corpos de conjuntos aos quais se pode prolongar qualquer distribuição definida na família dos intervalos; por exemplo, o corpo dos conjuntos borelianos, citado numa nota precedente.

B – Valores médios para distribuições numa variável real

Começemos por um exemplo: suponhamos que se trata de calcular a média μ das classificações de todos os alunos num exame. A fórmula a usar será, então,

$$\mu = \frac{\sum x v(x)}{N},$$

em que N representa o número total dos alunos, x cada uma das classificações $0, 1, \dots, 20$ e $v(x)$ o número de alunos que tiveram a classificação x . É claro que será, também,

$$\mu = \sum x \frac{v(x)}{N} = \sum x \text{fr}(x),$$

em que $\text{fr}(x)$ designa a frequência relativa de x .

Consideremos, agora, em geral, uma qualquer variável numérica x , susceptível dum número finito de valores x_1, x_2, \dots, x_r , com uma dada distribuição de frequência $\text{fr}(x)$. Chama-se *valor médio* da variável x ao número μ dado pela fórmula

$$\mu = \sum_{i=1}^r x_i \text{fr}(x_i).$$

O valor médio de x também se designa por $M\{x\}$ ou, até, abreviadamente, por \bar{x} . Importa salientar que $M\{x\}$ não é propriamente uma *função da variável x* , mas sim uma *função da distribuição $\text{fr}(x)$* ⁽¹⁾. Por isso mesmo se diz, também, (com mais propriedade) que μ é o *valor médio da distribuição*. Ainda com o mesmo significado se usa a expressão *centro da distribuição*, por analogia com o conceito mecânico de centro da gravidade; com efeito, se assimilarmos cada valor x_i de x a um ponto material de abcissa x_i e massa $\text{fr}(x_i)$, o conjunto de tais valores será um sistema material que tem por centro de gravidade o ponto μ .

O conceito do valor médio traduz-se, naturalmente, em termos de probabilidade. Sendo x uma variável numérica de valores x_1, x_2, \dots, x_r , com uma distribuição da probabilidade $\text{Pr}(x)$, chamaremos valor médio de x ao número

$$\mu = \sum_{i=1}^r x_i \text{Pr}(x_i).$$

Alguns autores usam, neste caso, a expressão *esperança matemática*, em vez de *valor médio*, e a notação $E\{x\}$ em vez de $M\{x\}$.

Por exemplo, se for x a variável casual cujos valores são os números que se obtêm nos lançamentos dum dado, a esperança matemática de x será

$$E\{x\} = 1 \frac{1}{6} + 2 \frac{1}{6} + 3 \frac{1}{6} + 4 \frac{1}{6} + 5 \frac{1}{6} + 6 \frac{1}{6} = \frac{21}{6}.$$

É claro que tudo o que dissermos sobre valores médios para distribuições de frequência se aplica, *mutatis mutandis*, a distribuições

(1) – É aquilo a que, modernamente, se chama *funcional*.

de probabilidade. Também deve, desde já, ficar assente que todas as variáveis a que faremos agora referência são sempre variáveis *numéricas*, com um número finito ou infinito de valores reais.

Seja x uma variável casual contínua definida num intervalo $[a, b]$, com uma função de probabilidade $\varphi(x)$ (densidade). É natural chamar, neste caso, valor médio ou esperança matemática de x ao número μ dado pela fórmula

$$\mu = \int_a^b x \varphi(x) dx,$$

em que o somatório foi substituído por um integral.

O valor deste integral é, ainda, designado por $M\{x\}$ (ou por $E\{x\}$).

Exemplo – Sendo $\varphi(x)dx$ a probabilidade de um recém-nascido viver até uma idade compreendida entre x e $x + dx$, o valor médio de x (que neste caso se chama *esperança de vida média*) será:

$$M\{x\} = \int_0^L x \varphi(x) dx,$$

em que L é um majorante da duração possível da vida. Para uma pessoa de idade a , a esperança de vida média será o integral de $x \varphi(x)$ entre 0 e L dividido pelo integral de $\varphi(x)$ entre a e L .

Na prática usam-se, apenas, valores aproximados destes integrais, obtidos pela regra do trapézio. Por exemplo, a tábua de mortalidade alemã, já citada, dá, para esperança de vida média dum recém-nascido, o valor 44,9 (anos); para um rapaz de 20 anos, o valor 42,6; para um homem de 50 anos, o valor 19,4, etc.

A anterior definição estende-se ao caso dum intervalo infinito, substituindo o integral próprio por um integral impróprio de 2.^a espécie.

Todas as demonstrações que faremos mais adiante acerca de valores médios pressupõem que a variável toma só um número finito de valores. Mas podem facilmente estender-se ao caso das variáveis casuais contínuas, com uma dada densidade de probabilidade $\varphi(x)$. Basta atender às propriedades dos integrais que generalizam as dos somatórios.

Cálculo prático do valor médio – Na prática, quando é muito grande o número de valores possíveis de x , o cálculo dos valores médios

deve subordinar-se a certas normas. Suponhamos que é dada uma tabela de frequências com intervalos-classes de comprimento h . Neste caso, obtém-se um valor aproximado de $M\{x\}$, com erro inferior a h , multiplicando o ponto médio de cada classe pela frequência relativa dessa classe e somando os resultados obtidos.

Para facilitar os cálculos, procede-se do modo seguinte:

1) – Toma-se para a unidade o comprimento h das classes, para que os valores das variáveis sejam inteiros.

2) – Escolhe-se, arbitrariamente, um valor c próximo do centro do intervalo em que varia x e calcula-se o valor médio da variável $x - c$. Como se tem, por definição,

$$M\{x - c\} = \sum (x_i - c) \text{fr}(x_i),$$

será

$$M\{x - c\} = \sum x_i \text{fr}(x_i) - c \sum \text{fr}(x_i) = M\{x\} - c,$$

donde

$$M\{x\} = c + M\{x - c\}.$$

O valor c diz-se *média arbitrária*; o valor $\gamma = M\{x - c\}$ diz-se *correção da média arbitrária*.

A vantagem do processo está em que os números $x_i - c$ são, geralmente, mais pequenos do que os correspondentes valores x_i .

3) – Exprime-se, finalmente, a média $M\{x\}$ na primitiva unidade. Veremos adiante um exemplo de aplicação.

Valores médios de funções de x – Seja, ainda, x uma variável susceptível dum número finito de valores x_1, x_2, \dots, x_r , com uma dada distribuição de frequência, $\text{fr}(x)$. Qualquer variável y , que seja função unívoca de x , terá também uma distribuição de frequência que se deduz facilmente da primeira: *a frequência dum dado valor de y será, evidentemente, a soma das frequências dos valores de x a que corresponde esse valor de y* . Seja $y = g(x)$; então, o valor médio de y será dado pela fórmula

$$M\{y\} = \sum g(x_i) \text{fr}(x_i).$$

Com efeito, se tivermos, por exemplo, $y_1 = g(x_1) = g(x_2) = \dots = g(x_{n_1})$, a frequência relativa de y_1 será $\text{fr}^*(y_1) = \text{fr}(x_1) + \text{fr}(x_2) + \dots + \text{fr}(x_{n_1})$, donde

$$\begin{aligned} y_1 \text{fr}^*(y_1) &= y_1 [\text{fr}(x_1) + \text{fr}(x_2) + \dots + \text{fr}(x_{n_1})] \\ &= g(x_1) \text{fr}(x_1) + \dots + g(x_{n_1}) \text{fr}(x_{n_1}) \end{aligned}$$

e, analogamente, para os outros valores de y , o que justifica a fórmula precedente.

Desta definição resultam, desde logo, as seguintes proposições:

PROPOSIÇÃO 1. *O valor médio da soma de duas variáveis u , v funções de x é igual à soma dos valores médios dessas variáveis:*

$$M\{u + v\} = M\{u\} + M\{v\}.$$

Com efeito, se for $u = g(x)$, $v = h(x)$, será

$$\begin{aligned} M\{u + v\} &= \sum [g(x_i) + h(x_i)] \text{fr}(x_i) = \\ &= \sum g(x_i) \text{fr}(x_i) + \sum h(x_i) \text{fr}(x_i) = M\{u\} + M\{v\}. \end{aligned}$$

PROPOSIÇÃO 2. *O valor médio dum constante ⁽¹⁾ k é essa mesma constante k .*

Com efeito, tem-se

$$M\{k\} = \sum k \text{fr}(x_i) = k \sum \text{fr}(x_i) = k.$$

PROPOSIÇÃO 3. *O valor médio do produto dum constante k por uma função u de x é igual ao produto da constante pelo valor médio da função.*

(1) – Isto é, dum função $g(x)$ cujos valores $g(x_1), \dots, g(x_r)$ sejam todos iguais a k .

Com efeito, se for $u = g(x)$, será

$$M\{ku\} = \sum k g(x_i) \text{fr}(x_i) = k \sum g(x_i) \text{fr}(x_i) = kM\{u\}.$$

As proposições 1 e 3 exprimem-se dizendo que o símbolo M representa um *operador linear*. As três propriedades anteriores combinadas entre si conduzem à proposição mais geral seguinte:

PROPOSIÇÃO 4. *Dadas n variáveis y_1, \dots, y_n , funções da variável x e $n + 1$ constantes a_0, a_1, \dots, a_n , tem-se:*

$$M\{a_0 + a_1 y_1 + \dots + a_n y_n\} = a_0 + a_1 M\{y_1\} + \dots + a_n M\{y_n\}.$$

Como já se disse previamente, estas considerações generalizam-se, imediatamente, ao caso das distribuições de probabilidade duma variável discreta ou duma variável contínua. Seja, por exemplo, x uma variável casual contínua definida no intervalo $[a, b]$ com uma densidade de probabilidade $\varphi(x)$, e seja $y = f(x)$ uma função de x integrável em $[a, b]$; então, o valor médio de y será dado pela fórmula

$$M\{y\} = \int_a^b f(x) \varphi(x) dx.$$

Como se tem $\int_a^b \varphi(x) dx = 1$, o teorema da média permite-nos afirmar que o valor médio de $f(x)$, ou seja, $M\{y\}$, é um número k compreendido entre os extremos inferior e superior de $f(x)$ em $[a, b]$. As propriedades elementares do integral de RIEMANN habilitam-nos a demonstrar que se tem, ainda,

$$M\{a_0 + a_1 y_1 + \dots + a_n y_n\} = a_0 + a_1 M\{y_1\} + \dots + a_n M\{y_n\},$$

sendo a_0, \dots, a_n constantes, e y_1, \dots, y_n funções de x .

Momentos – Entre as funções da variável x apresentam-se-nos, desde logo, as potências de expoente inteiro ≥ 0 . Chamam-se *momentos* de x os valores médios das funções $x^0, x^1, x^2, \dots, x^n, \dots$. Põe-se, habitualmente:

$$\mu_n = M\{x^n\} \quad (\text{momento de ordem } n).$$

É claro que $\mu_0 = M\{1\} = 1$, $\mu_1 = M\{x\} = \mu$.

Em vez de “momentos da variável x ” também se diz (e até com mais propriedade) “momentos da distribuição de x ”.

Dado um número c qualquer, chama-se *desvio de x a respeito de c* à variável $x - c$. Os desvios a respeito da média dizem-se, simplesmente, *desvios* ou *discrepâncias*, sem qualquer outra referência.

Os momentos da variável $x - c$ (*chamados momentos a respeito de c*) são comparáveis aos momentos dum sistema material a respeito dum ponto ou dum eixo. Interessam, especialmente, os *momentos a respeito do centro* (isto é, a respeito do valor médio). São estes:

$$(1) M\{x - \mu\} = M\{x\} - M\{\mu\} = \mu - \mu = 0,$$

$$(2) M\{(x - \mu)^2\} = M\{x^2 - 2\mu x + \mu^2\} = M\{x^2\} - 2\mu M\{x\} + \mu^2 = \\ = M\{x^2\} - 2\mu^2 + \mu^2 = \mu_2 - \mu^2,$$

$$(3) M\{(x - \mu)^3\} = M\{x^3 - 3x^2\mu + 3x\mu^2 - \mu^3\} = \\ = \mu_3 - 3\mu_2\mu + 3\mu\mu^2 - \mu^3 = \mu_3 + 2\mu^3 - 3\mu\mu_2,$$

etc.

É particularmente importante o segundo momento a respeito do centro, análogo ao momento de inércia dum sistema material: dá-se-lhe o nome de *variância* de x (ou da distribuição considerada) e representa-se por $V\{x\}$. Tem-se, pois, por definição,

$$V\{x\} = M\{(x - M\{x\})^2\}.$$

Como já se viu em (2), é

$$V\{x\} = \mu_2 - \mu^2 = M\{x^2\} - (M\{x\})^2,$$

isto é:

PROPOSIÇÃO 5. *A variância dum variável x é igual à diferença entre o valor médio do seu quadrado e o quadrado do seu valor médio.*

Daqui e das proposições 2 e 3 resultam logo, as seguintes consequências:

PROPOSIÇÃO 6. *A variância dum constante é nula.*

PROPOSIÇÃO 7. *A variância do produto de x por uma constante k é igual ao produto k^2 pela variância de x :*

$$V\{kx\} = k^2 V\{x\}.$$

Uma outra propriedade fundamental da variância é a seguinte:

PROPOSIÇÃO 8. *Qualquer que seja a constante a , a variância de $x + a$ é igual à variância de x :*

$$V\{x + a\} = V\{x\}.$$

Com efeito, $V\{x + a\}$ é, por definição, o valor médio do quadrado de $x + a - M\{x + a\}$; mas, como $M\{x + a\} = M\{x\} + a$, tem-se:

$$x + a - M\{x + a\} = x + a - (M\{x\} + a) = x - M\{x\},$$

donde,

$$V\{x + a\} = M\{(x - M\{x\})^2\} = V\{x\}.$$

Cálculo prático da variância – As proposições 5 e 8 são úteis na prática para o cálculo da variância. Com efeito, uma vez escolhida a “média arbitrária” c , tem-se, em virtude daquelas proposições,

$$V\{x\} = V\{x - c\} = M\{(x - c)^2\} - (M\{x - c\})^2,$$

ou seja:

$$V\{x\} = M\{(x - c)^2\} - \gamma^2,$$

pondo $\gamma = M\{x - c\}$ (correção da média arbitrária).

Basta, portanto, calcular o valor médio de $(x - c)^2$ e subtrair-lhe o quadrado de γ .

Suponhamos, por exemplo, que se trata de achar o centro e a variância da distribuição de frequência dada pela tabela n.º 4 do capítulo anterior. Tomando para média arbitrária c o valor 22,5 (ponto médio do intervalo [22, 23]), podemos dispor os cálculos preliminares no seguinte quadro, em que $v(x)$ designa a *frequência absoluta* de x :

x	$v(x)$	$x - c$	$(x - c) v(x)$	$(x - c)^2 v(x)$
14,5	1	-8	- 8	64
15,5	2	-7	- 14	98
16,5	2	-6	- 12	72
17,5	9	-5	- 45	225
18,5	11	-4	- 44	176
19,5	20	-3	- 60	180
20,5	75	-2	- 150	300
21,5	84	-1	- 84	84
22,5	95	0	0	0
23,5	60	1	60	60
24,5	20	2	40	80
25,5	14	3	42	126
26,5	3	4	12	48
27,5	2	5	10	50
28,5	1	6	6	36
29,5	1	7	7	49
	$\sum v(x) = 400$	—	$\sum (x - c) v(x) = -240$	$\sum (x - c)^2 v(x) = 1648$

Será, então:

$$\gamma = \frac{\sum (x - c) v(x)}{N} = -\frac{240}{400} = -0,60$$

e, portanto, o valor médio de x será

$$\bar{x} = c + \gamma = 22,5 + (-0,60) = 21,90.$$

Por sua vez, o segundo momento de $x - c$ é

$$M\{(x - c)^2\} = \frac{\sum (x - c)^2 v(x)}{N} = \frac{1.648}{400} = 4,12 ,$$

donde,

$$V\{x\} = M\{(x - c)^2\} - \gamma^2 = 4,12 - 0,36 = 3,76.$$

Dum modo geral, no cálculo da variância por este método, o erro proveniente de se agruparem os valores de x por classes pode ser reduzido subtraindo ao valor calculado a quantidade $h^2/12$, sendo h o comprimento das classes. Nisto consiste a chamada *correção de SHEPPARD*.

Uma distribuição diz-se mais ou menos *concentrada*, conforme os valores da variável se acumulam mais ou menos à volta do valor médio. O oposto de concentração é dispersão. Para dar uma ideia do grau de dispersão, poderia utilizar-se a *média dos módulos dos desvios*, isto é, o valor $M\{|x - \mu|\}$ (a média dos desvios não nos diz nada, visto ser nula). Uma medida de dispersão de grande interesse, teórico e prático, é a raiz quadrada da variância: dá-se-lhe o nome de *desvio padrão de x* (“standard deviation”, em inglês), também chamado *desvio quadrático médio*, e representa-se por $\sigma\{x\}$, e por σ_x ou, simplesmente, por σ , quando estiver subentendida a variável de que se trata. Ter-se-á, pois, por definição:

$$\sigma_x = \sqrt{V\{x\}} .$$

É claro que, assim como σ se pode tomar para índice de dispersão, assim, também, o seu inverso $1/\sigma$ se pode tomar para *índice de concentração*.

O valor máximo de $1/\sigma$ só é atingido, evidentemente, quando x assume um único valor, que será, então, a média μ : diz-se, neste caso, que toda a *massa* de distribuição está concentrada em μ . Neste caso será $V\{x\} = 0$ e, portanto, $1/\sigma = \infty$.

Dá-se o nome de *desvio reduzido de x* ao desvio de x dividido pelo desvio padrão. O desvio reduzido de x será, pois, a variável

$$h = \frac{x - \mu}{\sigma},$$

que dá a medida do desvio de x , tomando para unidade o desvio padrão.

É claro que, por sua vez:

$$M\{h\} = \frac{1}{\sigma} M\{x - \mu\} = \frac{1}{\sigma} (M\{x\} - \mu) = 0,$$

$$\sigma\{h\} = \sqrt{V\{h\}} = \sqrt{\frac{1}{\sigma^2} V\{x - \mu\}} = \frac{\sqrt{V\{x\}}}{\sigma} = 1,$$

isto é:

PROPOSIÇÃO 9. *O desvio reduzido dum variável x tem sempre o valor médio igual a 0 e o desvio padrão igual a 1.*

Dum modo geral, dada uma distribuição de frequência ou de probabilidade, podemos sempre substituí-la pela distribuição do desvio reduzido, tomando para nova origem dos eixos o valor médio μ e para unidade dos valores da variável o desvio padrão.

Diremos, então, que a distribuição foi *standardizada ou reduzida*. A standardização consiste, pois, na mudança de variável

$$x \rightarrow h = \frac{x - \mu}{\sigma}.$$

Convém registar a seguinte

PROPOSIÇÃO 10. *Se for $\Phi(x)$ a cumulante da distribuição dada, será*

$$\Psi(h) = \Phi(\mu + \sigma h)$$

a cumulante da distribuição standardizada. Reciprocamente, se for $\Psi(h)$ a cumulante da distribuição standardizada, será

$$\Phi(x) = \Psi\left(\frac{x - \mu}{\sigma}\right)$$

a cumulante da distribuição dada.

Assim, em resumo, o valor médio e o desvio padrão constituem duas características fundamentais duma distribuição: o primeiro indica sumariamente a *posição* ou *localização* da distribuição; o segundo quantifica a *dispersão* dos valores da variável à volta do valor médio.

Teorema de TCHEBICHEFF – O poder representativo do desvio padrão é posto em evidência por um teorema de TCHEBICHEFF, que podemos enunciar do seguinte modo:

Qualquer que seja a distribuição de probabilidade de x , a probabilidade de que o módulo do desvio $x - \mu$ seja igual ou superior a k vezes o desvio padrão (sendo k um número qualquer) é sempre igual ou inferior a $1/k^2$. Isto é, simbolicamente:

$$\Pr(|x - \mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

Faremos a demonstração apenas para o caso em que x toma um número finito de valores x_1, x_2, \dots, x_r , com probabilidades que designaremos, respectivamente, por p_1, p_2, \dots, p_r . Pondo $\varepsilon_i = x_i - \mu$, tem-se, por definição,

$$\sigma^2 = V\{x\} = p_1\varepsilon_1^2 + p_2\varepsilon_2^2 + \dots + p_r\varepsilon_r^2.$$

Sejam $\varepsilon_a, \varepsilon_b, \dots$ os desvios de módulo inferior a $k\sigma$ e sejam $\varepsilon_m, \varepsilon_n, \dots$ os desvios de módulo superior ou igual a $k\sigma$. É claro que, substituindo na soma $\sum p_i\varepsilon_i^2$ os números $\varepsilon_a, \varepsilon_b, \dots$ por 0 e os números $\varepsilon_m, \varepsilon_n, \dots$ por $k\sigma$, se obtém o resultado

$$\sigma^2 k^2 (p_m + p_n + \dots) \leq \sum p_i \varepsilon_i^2 = \sigma^2.$$

Mas a soma $p_m + p_n + \dots$ é a probabilidade dum desvio de módulo igual ou superior a $k\sigma$. Designando essa probabilidade por p , virá

$$\sigma^2 k^2 p \leq \sigma^2,$$

ou seja,

$$p \leq \frac{1}{k^2},$$

como queríamos demonstrar.

Este teorema costuma também ser apresentado com o seguinte aspecto:

A probabilidade dum desvio de módulo inferior a $k\sigma$ é igual ou superior a $1 - 1/k^2$. Simbolicamente:

$$\Pr(|x - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}.$$

Para reconhecer a equivalência dos dois enunciados, basta notar que o acontecimento $|x - \mu| < k\sigma$ é o contrário do acontecimento $|x - \mu| \geq k\sigma$. A sua probabilidade é, pois, complemento para 1 da probabilidade p deste último. Como $p \leq 1/k^2$, será $1 - p \geq 1 - 1/k^2$.

NOTAS IMPORTANTES A RESPEITO DA TERMINOLOGIA E DAS NOTAÇÕES

Dum modo geral, chama-se *parâmetro* dum distribuição de x toda a constante numérica associada a essa distribuição. Assim, serão parâmetros da distribuição os valores médios, não só de x , como de qualquer função de x ; e, ainda, as funções desses valores médios (como, por exemplo, o desvio padrão).

Muitas vezes, para estudar a distribuição dum variável (atributo quantitativo) numa determinada população, é-se obrigado a substituir a população por uma amostra, tão representativa quanto possível da população, e a considerar os parâmetros da distribuição dessa variável *na amostra*, como valores aproximados dos parâmetros da distribuição da mesma variável *na população total*. (É o que se faria, por exemplo, para estudar a distribuição da variável “diâmetro do tronco” numa extensa mata de eucaliptos).

Nesta ordem de ideias, é costume designar os parâmetros da distribuição, na amostra, pelas letras latinas correspondentes às letras gregas com as quais se representam os mesmos parâmetros na população considerada: por exemplo, a média por m , o segundo

momento por m_2 , o desvio padrão por s , etc., etc. É, ainda, nas amostras que, de preferência, se usa a notação \bar{x} para designar o valor médio de x .

Aos parâmetros da distribuição na amostra daremos, ainda, o nome de *constantes estatísticas* da mesma (em inglês, *statistics*).

Importa, finalmente, observar que a distribuição da variável considerada na população tem, muitas vezes, de ser concebida como distribuição de probabilidade. Sucede isto, primeiro que tudo, quando se trata duma população que esteja constantemente a ser acrescida de novos indivíduos, podendo, assim, considerar-se praticamente infinita (por exemplo, uma espécie, uma raça, uma variedade, etc.). Nestes casos, é corrente atribuir à distribuição de probabilidade uma determinada expressão analítica, com base em considerações de carácter teórico-experimental. A distribuição de frequência relativa da variável na amostra deverá, então, *tender* para a distribuição de probabilidade pressuposta na população quando o número de elementos da amostra aumenta indefinidamente.

Este ponto visto é, ainda, aplicado a populações que, embora circunscritas no tempo e no espaço, sejam muito numerosas.

Todas estas considerações se aplicam, *mutatis mutandis*, ao caso das distribuições de duas ou mais variáveis numéricas, que passamos a estudar.

C – Valores médios para distribuições de mais de uma variável real

Começemos por considerar um sistema (x, y) de duas variáveis reais, susceptível dum número finito de valores,

$$(x_i, y_k), \quad i = 1, 2, \dots, r, \quad k = 1, 2, \dots, s,$$

e suponhamos dada uma distribuição de frequências, $fr(x, y)$, sobre o universo destes valores. Nestas condições, qualquer variável z que seja função unívoca de (x, y) terá, também, uma distribuição de frequência que se deduz da primeira do seguinte modo: *a frequência dum dado valor de z será a soma das frequências dos valores de (x, y) a que corresponde esse valor de z* . Seja $z = g(x, y)$; então, é

fácil reconhecer, como para as distribuições duma só variável, que o valor médio de z é dado pela fórmula

$$M\{z\} = \sum_{i, k} g(x_i, y_k) \text{fr}(x_i, y_k).$$

Entre as possíveis funções de (x, y) aparecem-nos, primeiro que tudo, as próprias variáveis x, y . Será, então,

$$\begin{aligned} M\{x\} &= \sum_{i, k} x_i \text{fr}(x_i, y_k) = \sum_i \sum_k x_i \text{fr}(x_i, y_k) \\ &= \sum_i x_i \sum_k \text{fr}(x_i, y_k) = \sum_i x_i \text{fr}(x_i), \end{aligned}$$

visto que $\text{fr}(x_i) = \sum_k \text{fr}(x_i, y_k)$ (*1ª frequência marginal*).

Analogamente,

$$M\{y\} = \sum_k y_k \text{fr}(y_k),$$

com $\text{fr}(y_k) = \sum_i \text{fr}(x_i, y_k)$ (*2ª frequência marginal*)⁽¹⁾.

Uma outra função simples de (x, y) é a sua soma $x + y$. A proposição 1 de B, pode agora generalizar-se do seguinte modo:

PROPOSIÇÃO 1. *O valor médio da soma das duas variáveis x, y é igual à soma dos valores médios de x e de y .*

Com efeito, tem-se, por definição:

$$\begin{aligned} M\{x + y\} &= \sum_{i, k} (x_i + y_k) \text{fr}(x_i, y_k) \\ &= \sum_{i, k} x_i \text{fr}(x_i, y_k) + \sum_{i, k} y_k \text{fr}(x_i, y_k) \\ &= M\{x\} + M\{y\}, \end{aligned}$$

em virtude do que se disse, há pouco, sobre $M\{x\}$ e $M\{y\}$.

(1) – As duas funções $\text{fr}(x)$, $\text{fr}(y)$ tomam, geralmente, valores diferentes para valores iguais das variáveis x, y . Seria, por isso, mais correcto designá-las por símbolos diferentes, por exemplo, $\text{fr}_1(x)$, $\text{fr}_2(y)$. Não o fazemos para não sobrecarregar as notações.

Chamam-se *momentos de distribuição* $fr(x, y)$ os valores médios das funções $x^m y^n$, sendo m, n números inteiros não negativos. Dum modo geral, põe-se

$$\mu_{m,n} = M\{x^m y^n\}.$$

Em particular, $\mu_{1,0} = M\{x\}$, $\mu_{0,1} = M\{y\}$, $\mu_{2,0} = M\{x^2\}$, $\mu_{0,2} = M\{y^2\}$ (primeiros e segundos momentos das variáveis x, y , isoladas). Ao par $(\mu_{1,0}, \mu_{0,1})$ dá-se o nome de *centro da distribuição*, ainda por analogia com o centro da gravidade dum sistema de pontos materiais (x_i, y_k) que tivessem massas iguais a $fr(x_i, y_k)$, respectivamente.

O primeiro momento misto é $\mu_{1,1} = M\{xy\}$. Algumas vezes, para simplificar as notações, representaremos por \bar{x} o valor médio de x e por \bar{y} o valor médio de y (isto é, pomos $\bar{x} = \mu_{1,0}$, $\bar{y} = \mu_{0,1}$, embora as notações \bar{x}, \bar{y} se devam usar, de preferência, para as amostras).

PROPOSIÇÃO 2. *Se as variáveis x, y são independentes (a respeito da distribuição considerada), tem-se*

$$M\{xy\} = M\{x\} M\{y\}, \text{ ou seja, } \mu_{1,1} = \mu_{1,0} \mu_{0,1}.$$

Com efeito, dizer que x, y são independentes equivale a dizer que $fr(xy) = fr(x) \cdot fr(y)$. Então, será

$$\begin{aligned} M\{xy\} &= \sum_{i,k} x_i y_k fr(x_i, y_k) = \sum_{i,k} x_i y_k fr(x_i) fr(y_k) \\ &= \sum_i x_i fr(x_i) \cdot \sum_k y_k fr(y_k) = M\{x\} M\{y\}. \end{aligned}$$

São particularmente importantes os *momentos a respeito do centro*,

$$M\{(x - \bar{x})^m \cdot (y - \bar{y})^n\}, \quad m, n = 0, 1, 2, \dots$$

Entre estes, destacaremos o valor médio do produto dos desvios $x - \bar{x}, y - \bar{y}$. Dá-se-lhe o nome de *covariância* das variáveis x, y e designa-se por $C\{x, y\}$. Portanto:

$$C\{x, y\} = M\{(x - \bar{x})(y - \bar{y})\}.$$

Será, então:

$$\begin{aligned} C\{x, y\} &= M\{xy - x\bar{y} - y\bar{x} + \bar{y}\bar{x}\} \\ &= M\{xy\} - \bar{y}M\{x\} - \bar{x}M\{y\} + \bar{x}\bar{y} = M\{xy\} - \bar{y}\bar{x} - \bar{x}\bar{y} + \bar{x}\bar{y} \\ &= M\{xy\} - M\{x\}M\{y\} = \mu_{1,1} - \mu_{1,0} \cdot \mu_{0,1}, \end{aligned}$$

isto é:

PROPOSIÇÃO 3. *A covariância das variáveis x , y é igual à diferença entre o valor médio do produto dessas variáveis e o produto dos valores médios das mesmas.*

Daqui e da proposição 2 deduz-se, logo, o seguinte

COROLÁRIO. *Se as variáveis x , y são independentes, a sua covariância é nula (a recíproca, porém, não é verdadeira).*

Por sua vez, tem-se

PROPOSIÇÃO 4. *A variância da soma das duas variáveis x , y é igual à soma das respectivas variâncias mais o dobro da sua covariância, isto é:*

$$V\{x + y\} = V\{x\} + V\{y\} + 2C\{x, y\}.$$

Começemos por notar que o desvio de $x + y$ é

$$x + y - M\{x + y\} = x + y - (\bar{x} + \bar{y}) = (x - \bar{x}) + (y - \bar{y}),$$

o que se pode exprimir dizendo que o *desvio da soma é igual à soma dos desvios das parcelas*. O quadrado do desvio de $x + y$ será, pois,

$$(x - \bar{x})^2 + (y - \bar{y})^2 + 2(x - \bar{x}) \cdot (y - \bar{y}),$$

donde, pela definição de variância e pela proposição 1:

$$\begin{aligned} V\{x + y\} &= M\{(x - \bar{x})^2 + (y - \bar{y})^2 + 2(x - \bar{x}) \cdot (y - \bar{y})\} \\ &= M\{(x - \bar{x})^2\} + M\{(y - \bar{y})^2\} + 2M\{(x - \bar{x}) \cdot (y - \bar{y})\} \\ &= V\{x\} + V\{y\} + 2C\{x, y\}. \end{aligned}$$

Daqui e do corolário anterior vem logo este outro

COROLÁRIO. *Se as variáveis x, y são independentes, tem-se*
 $V\{x + y\} = V\{x\} + V\{y\}$.

Registem-se, ainda, as seguintes propriedades, cuja demonstração é imediata:

$$C\{x, x\} = V\{x\}, \quad C\{ax, by\} = ab C\{x, y\}$$

sendo a, b constantes quaisquer.

Será, então, em virtude das propriedades já demonstradas da variância

$$(1) \quad \boxed{V\{ax + by + c\} = a^2V\{x\} + b^2V\{y\} + 2ab C\{x, y\}}$$

em que a, b, c são constantes quaisquer.

Correlação e regressão – Já vimos que se tem $C\{x, y\} = 0$ quando x, y são independentes. Para avaliar o *grau de dependência* ou *associação* das variáveis x, y , usa-se o seguinte índice:

$$\rho = \frac{C\{x, y\}}{\sqrt{V\{x\}V\{y\}}} = \frac{C\{x, y\}}{\sigma_x \sigma_y}$$

chamado *coeficiente de correlação* ou, apenas, *correlação de x e y* .

Vamos ver que é sempre

$$-1 \leq \rho \leq 1$$

e que se tem $\rho = 1$ ou $\rho = -1$, se, e só se, a variável y é função linear de x , estando, então, os pontos (x_i, y_k) sobre uma recta. Chega-se a esta conclusão mediante as seguintes considerações:

Se as variáveis x, y , não são independentes, pode presumir-se a existência duma relação funcional entre elas, isto é, duma *lei natural*, que, em primeira aproximação, se procurará exprimir por meio duma função linear,

$$y = a + bx.$$

É claro que, só num caso excepcional, teórico, se poderá ter $y_k = a + bx_i$, ou seja, $y_k - a - bx_i = 0$, para todos os pontos (x_i, y_k) de frequência não nula. Em geral, estes pontos não estão em linha recta. O que se procura, então, é determinar a, b de modo que os desvios $y_k - a - bx_i$ (distâncias verticais dos pontos (x_i, y_k) à recta $y = a + bx$) sejam mínimos; mais precisamente, procura-se tornar mínima a média

$$E = M\{(y - a - bx)^2\} = \sum_{i,k} (y_k - a - bx_i)^2 \text{fr}(x_i, y_k),$$

dos quadrados dos referidos desvios (*método dos mínimos quadrados*). Ora, tem-se, desenvolvendo $[(y - bx) - a]^2$ e atendendo à linearidade do operador M :

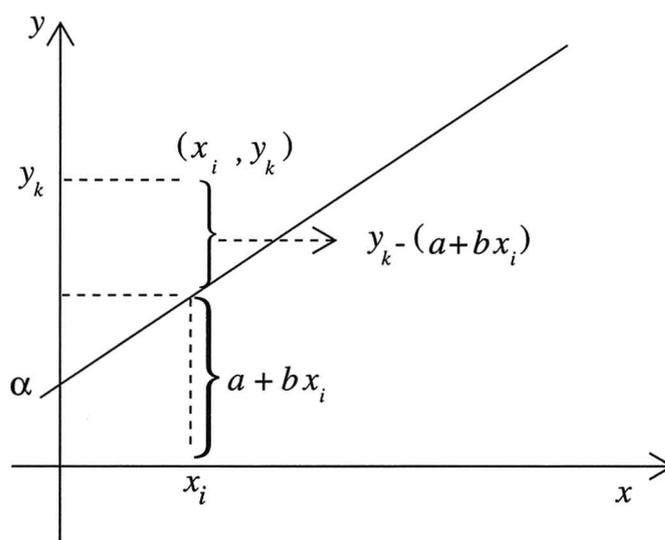


Fig. 4

$$\begin{aligned} E &= M\{(y - bx)^2\} - 2M\{a(y - bx)\} + M\{a^2\} \\ &= M\{y^2\} - 2bM\{xy\} + b^2M\{x^2\} - 2aM\{y\} + 2abM\{x\} + a^2 \\ &= \mu_{0,2} - 2b\mu_{1,1} + b^2\mu_{2,0} - 2a\bar{y} + 2ab\bar{x} + a^2, \end{aligned}$$

onde, para simplificar as notações, pusemos \bar{x} em vez de $\mu_{1,0}$ e \bar{y} em vez de $\mu_{0,1}$. Trata-se, pois, de minimizar a função E de a, b .

Virá, então,

$$\frac{\partial E}{\partial a} = 2a + 2(b\bar{x} - \bar{y}), \quad \frac{\partial E}{\partial b} = 2b\mu_{2,0} + 2(a\bar{x} - \mu_{1,1}),$$

donde, o sistema de equações nas incógnitas a , b :

$$\begin{cases} a + \bar{x}b = \bar{y} \\ \bar{x}a + \mu_{2,0} b = \mu_{1,1} \end{cases}$$

que, resolvido, dá o ponto de estacionaridade (a_1, b_1) definido pelas fórmulas:

$$b_1 = \frac{\mu_{1,1} - \bar{x}\bar{y}}{\mu_{2,0} - \bar{x}^2}, \quad a_1 = \bar{y} - b_1\bar{x}.$$

Notemos, ainda, que

$$\mu_{1,1} - \bar{x}\bar{y} = M\{xy\} - M\{x\}M\{y\} = C\{x, y\}$$

e

$$\mu_{2,0} - \bar{x}^2 = M\{x^2\} - (M\{x\})^2 = V\{x\},$$

o que permite escrever b_1 sob a forma

$$b_1 = \frac{C\{x, y\}}{V\{x\}} = \rho \frac{\sigma_y}{\sigma_x},$$

visto que $C\{x, y\} = \rho \sqrt{V\{x\}V\{y\}} = \rho\sigma_x\sigma_y$ (pondo $\sigma_x = \sqrt{V\{x\}}$, $\sigma_y = \sqrt{V\{y\}}$).

É fácil verificar que se trata, efectivamente, dum mínimo (absoluto). A recta pedida será, pois,

$$y = \bar{y} + b_1(x - \bar{x}), \quad \text{com } b_1 = \frac{C\{x, y\}}{V\{x\}} = \rho \frac{\sigma_y}{\sigma_x}$$

Dá-se-lhe o nome de *recta de regressão de y sobre x*. Analogamente, a recta

$$x = \bar{x} + b_2(y - \bar{y}), \quad \text{com } b_2 = \frac{C\{x, y\}}{V\{y\}} = \rho \frac{\sigma_x}{\sigma_y}$$

é chamada *recta de regressão de x sobre y*. Como se vê, estas duas rectas passam pelo ponto (\bar{x}, \bar{y}) , centro da distribuição. Os coeficientes b_1, b_2 dizem-se *coeficientes de regressão*, ou, apenas, *regressões*.

Note-se que, para $a = a_1, b = b_1$, vem

$$\begin{aligned} E &= M\{[y - \bar{y} - b_1(x - \bar{x})]^2\} \\ &= M\{(y - \bar{y})^2\} - 2b_1 M\{(x - \bar{x})(y - \bar{y})\} + b_1^2 M\{(x - \bar{x})^2\} \\ &= V\{y\} - 2b_1 C\{x, y\} + b_1^2 V\{x\} = V\{y\} - 2\rho^2 V\{y\} + \rho^2 V\{y\} \\ &= V\{y\}(1 - \rho^2), \end{aligned}$$

visto que $C\{x, y\} = \rho\sigma_x\sigma_y$ e $b_1 = \rho\sigma_y/\sigma_x$. Portanto, o valor mínimo de E será $V\{y\}(1 - \rho^2)$. Ora, é evidente que este mínimo (valor médio dos quadrados dos desvios verticais a respeito da 1ª recta) só será nulo, se os pontos (x_i, y_k) de frequência não nula estiverem sobre aquela recta. Vê-se, pois, que, como tínhamos afirmado, os referidos pontos estão em linha recta, se, e só se, $1 - \rho^2 = 0$, ou seja, $\rho^2 = 1$.

Neste caso, as duas rectas de regressão coincidem, pois que será $b_2 = 1/b_1$. As variáveis x, y dizem-se, então, *perfeitamente correlacionadas* (positivamente, se $\rho = +1$, negativamente, se $\rho = -1$). Mas este é um caso ideal, que nunca se verifica exactamente na prática. Vê-se, entretanto, que, se o valor de ρ^2 for bastante próximo de 1, as duas rectas se aproximam bastante uma da outra, ajustando-se ambas, com boa aproximação, ao conjunto dos pontos (x_i, y_k) considerados (regressão linear).

Casos há, todavia, em que a linearidade está longe de traduzir uma possível relação funcional entre as variáveis x, y em questão. Recorre-se, então, a funções mais complicadas – polinómios, exponenciais, etc. – para tentar traduzir a dita relação. Trata-se, portanto, de *ajustar* o mais possível, ao conjunto dos pontos (x_i, y_k) , uma curva de dado tipo, usando, por exemplo, o método dos mínimos quadrados. Assim, a *regressão linear* cede o lugar à *regressão curvilínea*, cuja teoria não podemos expor aqui.

O exemplo que damos a seguir, extraído da citada obra de FINNEY⁽¹⁾, refere-se ao estudo da correlação entre a densidade de produção de trigo (variável x) e o teor do trigo em proteína (variável y), sendo os dados relativos a 10 talhões. A densidade de produção é medida em cwt por acre (o cwt, abreviatura de “hundred weight”, equivale a 50,802 kg, e o acre equivale, aproximadamente, a 0,40467 ha).

TABELA N.º 1

N.º do talhão	x = Produção em cwt por acre	y = Proteína %
1	14,3	10,8
2	12,8	11,4
3	12,7	13,0
4	10,6	14,6
5	10,7	13,8
6	13,0	12,2
7	14,4	10,7
8	12,5	12,8
9	8,7	16,2
10	12,2	11,8

O coeficiente de correlação é, neste caso:

$$\rho = - \frac{26,33}{\sqrt{27,65 \times 27,52}} = - 0,955.$$

Por sua vez, tem-se

$$\bar{x} = 12,19, \bar{y} = 12,73, b_1 = - \frac{26,33}{27,65} = - 0,952,$$

(1) – Importa salientar que, na referida obra, este exemplo é, por sua vez, uma adaptação de resultados expostos pelo Engenheiro Augusto José de Oliveira, num trabalho publicado em “Agronomia Lusitana”, vol. 8 (1946), pp. 147–159 (Estação Agronómica Nacional).

donde, a equação de regressão

$$y = 12,73 - 0,952 (x - 12,19),$$

ou seja,

$$y = 24,3 - 0,95 x.$$

Esta recta está representada na figura junta, em que os pontos marcados indicam os pares (x_i, y_k) observados. *É visível que o teor das sementes em proteínas diminui quando a densidade de produção aumenta, seguindo esta variação uma lei sensivelmente linear.*

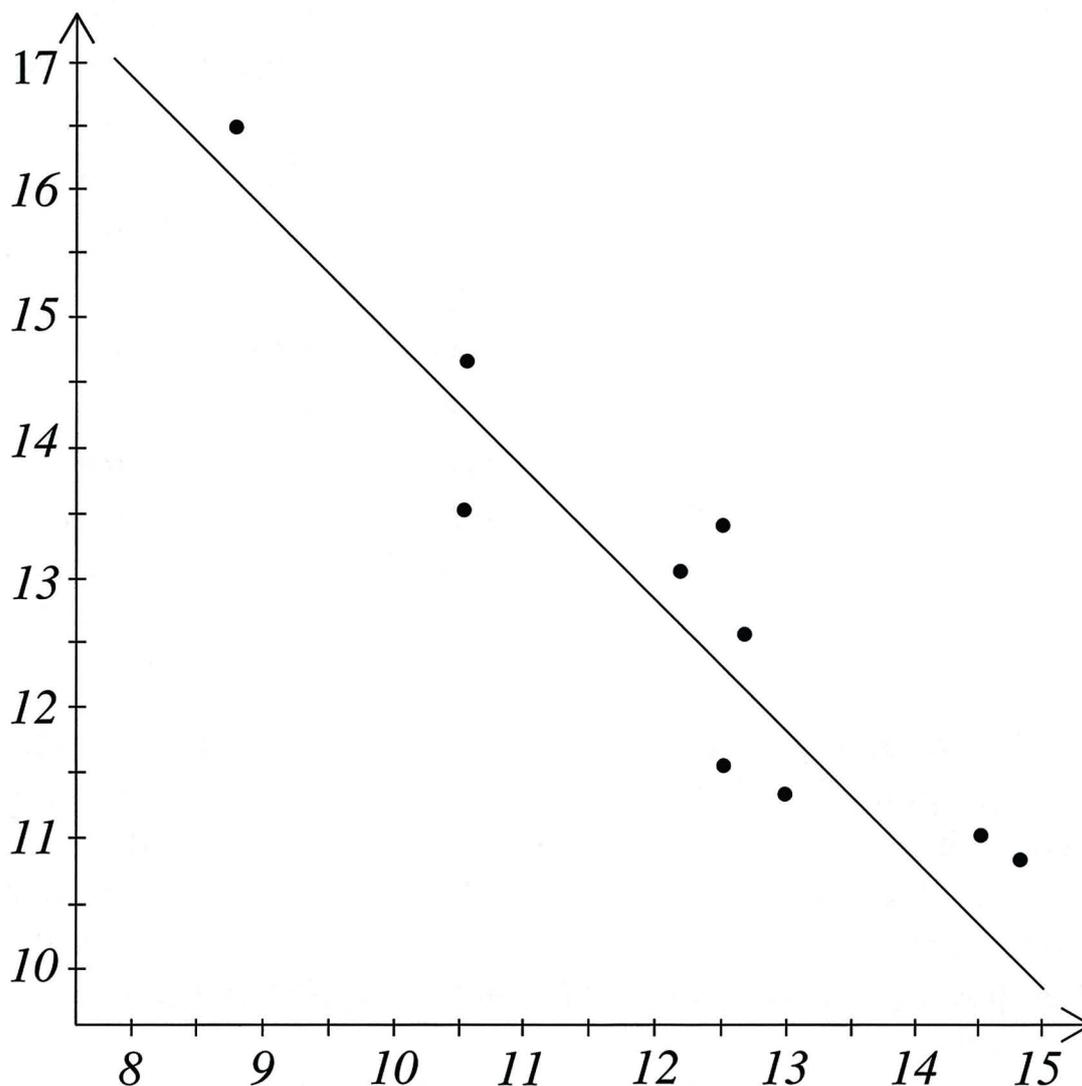


Fig. 5

Para ter uma ideia mais precisa do significado dos valores de ρ , institui-se sobre este índice um teste de significância, que dependerá, naturalmente, do número de pontos (x_i, y_k) observados: é a esse número, diminuído de 2 unidades, que, neste caso, se dá o nome de *número de graus de liberdade*. A *hipótese nula* consiste, agora, em supor $\rho = 0$; para averiguar em que medida o valor de ρ observado é ou não atribuível ao acaso, recorre-se às tábuas que dão valores de ρ correspondentes a diversos níveis de significância. Da mesma obra extraímos a seguinte tabela:

TABELA N.º 2

Nº de graus de liberdade	Nº de pares	Probabilidade		
		0,1	0,05	0,01
1	3	0.988	0.9969	0.9999
2	4	0.90	0.95	0.99
3	5	0.81	0.88	0.96
4	6	0.73	0.81	0.92
6	8	0.62	0.71	0.83
8	10	0.55	0.63	0.76
10	12	0.50	0.58	0.71
15	17	0.41	0.48	0.61
20	22	0.36	0.42	0.54
30	32	0.30	0.35	0.45
60	62	0.21	0.25	0.32

Não esquecer que a tabela dá valores de $|\rho|$. No caso anterior, o número de graus de liberdade é 8, sendo $|\rho| = 0,955$. Ora, este valor excede o limite 0,76 que marca o nível 0,01: quer isto dizer que, sendo válida a hipótese nula, a probabilidade dum ρ igual ou superior a 0,955 (em módulo) é bastante inferior a 1%. O valor de ρ obtido pode, pois, considerar-se *altamente significativa contra a hipótese nula*.

Neste exemplo, o número de pares observados é pequeno. Quando esse número for grande, torna-se necessário repartir os valores de x , y por intervalos de classe, como foi indicado para as

tábuas de frequências, e organizar uma tábua de contingência que, neste caso (atributos quantitativos) se dirá uma *tábua de correlação*. Do próprio exame da tábua se pode já inferir se os dados tendem ou não a acumular-se à volta duma recta.

Caso das distribuições de probabilidade de duas variáveis reais – É claro que todas as considerações precedentes, relativas a distribuições de frequência de duas variáveis x, y , se podem estender, *mutatis mutandis*, ao caso das probabilidades. Poderão, ainda, considerar-se variáveis contínuas em vez de variáveis discretas. Para isso, haverá que estender os conceitos definidos em A) ao caso de duas variáveis contínuas x, y , substituindo os intervalos J por rectângulos Δ de lados paralelos aos eixos coordenados. Assimilando a distribuição de probabilidade a uma distribuição de massa, chega-se, intuitivamente, ao conceito de *densidade (superficial) de probabilidade* $\varphi(x, y)$ ⁽¹⁾. Então, a probabilidade correspondente a um dado rectângulo Δ será dada pelo integral duplo

$$\text{Pr}(\Delta) = \iint_{\Delta} \varphi(x, y) dx dy.$$

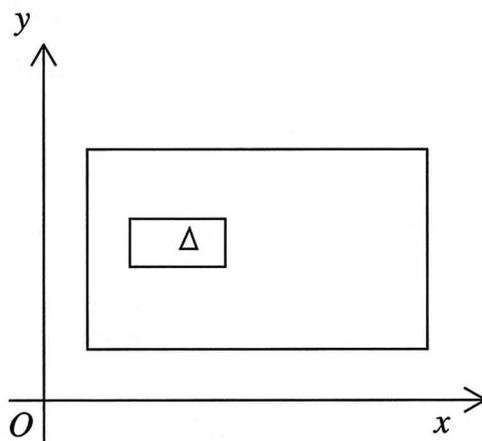


Fig. 6

(1) – Note-se que o conhecido conceito de *densidade de população se refere a uma densidade superficial de frequência absoluta* (número de habitantes por unidade de área). O próprio conceito de massa específica se confunde, na escala atômica, com o de densidade de população (de partículas materiais).

É claro que, se for U o domínio da distribuição, será

$$\Pr(U) = \iint_U \varphi(x, y) dx dy = 1.$$

Por sua vez, o valor médio duma função $z = f(x, y)$, integrável em U , será

$$M\{z\} = \iint_U f(x, y) \varphi(x, y) dx dy,$$

e, pelo teorema da média, tem-se

$$L_1 \leq M\{z\} \leq L_2,$$

sendo L_1 e L_2 os extremos inferior e superior de $f(x, y)$ em U .

Todas as anteriores proposições se podem, então, generalizar a este caso, atendendo às propriedades elementares do integral duplo.

Distribuição de n variáveis reais – Somos, finalmente, conduzidos, na mesma ordem de ideias, a considerar distribuições de n variáveis x_1, x_2, \dots, x_n , discretas ou contínuas. É claro que o caso das variáveis contínuas obriga a considerar domínios do espaço \mathbf{R}^n e a utilizar integrais múltiplos (duplos, triplos, quádruplos, etc., conforme for $n = 2, 3, 4, \dots$).

Quanto a valores médios em \mathbf{R}^n , a sua teoria é perfeitamente análoga à precedente.

Suponhamos, por exemplo, que o sistema de variáveis (x_1, x_2, \dots, x_n) só pode tomar um número finito de valores. Então, dada uma distribuição de frequência, $\text{fr}(x_1, x_2, \dots, x_n)$, destas variáveis, e uma função $y = g(x_1, x_2, \dots, x_n)$ das mesmas, o valor médio de y , será, por definição,

$$M\{y\} = \sum g(x_1, \dots, x_n) \text{fr}(x_1, \dots, x_n).$$

Todas as anteriores proposições se generalizam a este caso. Mas bastará registrar as seguintes fórmulas:

$$M\{x_1 + x_2 + \dots + x_n\} = M\{x_1\} + M\{x_2\} + \dots + M\{x_n\}$$

$$V\{x_1 + x_2 + \dots + x_n\} = \sum_i V\{x_i\} + 2 \sum_{i < k} C\{x_i, x_k\}$$

Em particular, se x_1, \dots, x_n são independentes, será $C\{x_i, x_k\} = 0$ para $i \neq k$ e a anterior fórmula simplifica-se:

$$V\{x_1 + x_2 + \dots + x_n\} = V\{x_1\} + V\{x_2\} + \dots + V\{x_n\}$$

Não esquecer, ainda, que se tem

$$\begin{aligned} V\{x_i\} &= M\{x_i^2\} - (M\{x_i\})^2, \\ C\{x_i, x_k\} &= M\{x_i x_k\} - M\{x_i\} M\{x_k\}, \\ M\{ay\} &= a M\{y\}, \quad V\{ay + b\} = a^2 V\{y\}, \end{aligned}$$

sendo y uma função qualquer de x_1, \dots, x_n , e a, b constantes arbitrárias.

É claro que estas fórmulas se aplicam, igualmente, às distribuições de probabilidade, podendo, nesse caso, substituir-se a notação $M\{x\}$ por $E\{x\}$ (esperança matemática de x).

D – Aplicação à distribuição binomial. Teorema de BERNOULLI

Como se sabe, a distribuição de BERNOULLI

$$\Pr(x) = \binom{n}{x} p^x q^{n-x}$$

dá a probabilidade de que um acontecimento α , de probabilidade p , se realize x vezes em n provas

$$\mathcal{P}_1, \mathcal{P}_2, \dots, \mathcal{P}_n.$$

Para determinar o centro e o desvio padrão desta distribuição, vamos recorrer a um artifício, em que se utilizam as propriedades precedentes, e que consiste no seguinte:

Designemos por $x^{(i)}$ a variável casual assim definida:

$$x^{(i)} \begin{cases} = 1, & \text{se } \alpha \text{ se realiza na prova } \mathcal{P}_i, \\ = 0, & \text{se } \alpha \text{ não se realiza na prova } \mathcal{P}_i, \end{cases}$$

(para $i = 1, 2, \dots, n$). Então, qualquer que seja i , será p a probabilidade de ser $x^{(i)} = 1$ e será $1 - p$ a probabilidade de ser $x^{(i)} = 0$. Pondo $q = 1 - p$, virá, portanto,

$$(1) \quad M\{x^{(i)}\} = 1 \cdot p + 0 \cdot q = p.$$

O desvio de $x^{(i)}$ será, pois, $x^{(i)} - p$, com os valores $1 - p$, $0 - p$ de probabilidades p , q , respectivamente. Portanto:

$$(2) \quad \begin{aligned} V\{x^{(i)}\} &= M\{(x^{(i)} - p)^2\} = (1 - p)^2 p + (-p)^2 q = \\ &= q^2 p + p^2 q = pq(p + q) = pq. \end{aligned}$$

Notemos, agora, que a soma $x^{(1)} + x^{(2)} + \dots + x^{(n)}$ tem tantas parcelas iguais a 1 quantas as vezes que α se realiza, sendo as restantes parcelas nulas; a soma será, pois, igual ao *número de realizações de α nas n provas, ou seja, x* . Tem-se, pois,

$$x = x^{(1)} + x^{(2)} + \dots + x^{(n)}.$$

Além disso, já sabemos que as variáveis casuais $x^{(i)}$ são *independentes*. Logo, atendendo a (1), vem:

$$M\{x\} = M\{x^{(1)}\} + \dots + M\{x^{(n)}\} = np,$$

e atendendo a (2):

$$V\{x\} = V\{x^{(1)}\} + \dots + V\{x^{(n)}\} = npq.$$

Serão, pois,

$$\mu = np, \quad \sigma = \sqrt{npq},$$

o valor médio e o desvio padrão da frequência absoluta x de α .

A frequência relativa de α nas n provas é

$$f = \frac{x}{n}.$$

Aplicando os resultados anteriores, tem-se

$$M\{f\} = \frac{1}{n} M\{x\} = p, \quad V\{f\} = \frac{1}{n^2} V\{x\} = \frac{pq}{n}.$$

O valor médio e o desvio padrão da frequência relativa de α são, pois, respectivamente,

$$p \quad \text{e} \quad \sqrt{\frac{pq}{n}}.$$

Diz-se que uma sucessão

$$u_1, u_2, \dots, u_n, \dots$$

de números reais *converge estocasticamente* (ou *converge em probabilidade*) para um número real a , quando, dado um número $\delta > 0$, por menor que ele seja, a probabilidade de

$$|u_n - a| < \delta$$

tende para 1 quando n tende para ∞ .

Podemos, agora, enunciar o

TEOREMA DE BERNOULLI. *Seja α um acontecimento de probabilidade p . A frequência relativa de α em n provas converge estocasticamente para p , quando $n \rightarrow \infty$.*

Demonstração. Designemos, agora, mais precisamente por f_n a frequência relativa de α nas n provas. Como se viu atrás, tem-se $M\{f_n\} = p$, $\sigma\{f_n\} = \sqrt{pq/n}$. Então, segundo o teorema de Tchebicheff atrás demonstrado, a probabilidade P_n de que se terá

$$|f_n - p| < \delta$$

satisfaz à condição

$$P_n \geq 1 - \frac{1}{k^2}, \text{ em que } k = \frac{\delta}{\sigma} = \delta \sqrt{\frac{n}{pq}}.$$

Ora, quando $n \rightarrow \infty$, também $k \rightarrow \infty$ e, portanto, $1 - 1/k^2$ tende para 1. Como se tem, por outro lado, $P_n \leq 1$, virá pois,

$$\lim_{n \rightarrow \infty} P_n = 1,$$

o que, segundo a definição anterior, é a tese do teorema.

Este teorema vem lançar nova luz sobre as relações entre os conceitos de frequência relativa e de probabilidade. Como vemos, a *frequência relativa, f_n , converge estocasticamente para a probabilidade p , quando $n \rightarrow \infty$* . Daqui resulta que, dado um número positivo δ , tão pequeno quanto quisermos, existe sempre uma ordem a partir da qual é *praticamente certo que $f_n - p < \delta$* . É praticamente certo, mas não certo! *Não será, portanto, lícito escrever*

$$\lim_{n \rightarrow \infty} f_n = p,$$

conforme o conceito de limite da análise matemática, embora, na prática, as coisas se passem, de certo modo, como tal. A variável f_n converge para p de *maneira casual, irregular*, não se podendo, em absoluto, negar a existência de desvios grandes para valores de n elevados.

A probabilidade P_n que intervem na demonstração diz-se de *segunda ordem* a respeito de p . É nas probabilidades de segunda ordem que se baseia a técnica dos chamados *resseguros*, usada entre companhias de seguros, para garantir um maior grau de segurança.

E – Distribuição normal

Uma grande parte das distribuições que se encontram na prática aproxima-se mais ou menos duma distribuição que, em cada ponto x do intervalo $]-\infty, +\infty[$, tem uma densidade $\phi(x)$ tal que

$$\phi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

sendo μ , σ constantes.

A uma tal distribuição dá-se o nome de *distribuição normal*, de GAUSS ou LAPLACE-GAUSS, de parâmetros μ , σ ; ou, abreviadamente, distribuição (N; μ , σ).

Demonstra-se que é:

$$(1) \quad \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1,$$

$$(2) \quad \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \mu,$$

$$(3) \quad \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} (x-\mu)^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \sigma^2.$$

Para a demonstração, veja-se, por exemplo, CASTELNUOVO, obra citada, pp. 253-255 (feita a mudança de variáveis $h = (x-\mu)/\sigma$).

A segunda fórmula diz-nos que o centro da distribuição é μ . A terceira diz-nos que a variância da distribuição é σ^2 , sendo, portanto, σ o desvio padrão. Fica, assim, justificado o uso das letras μ , σ , como parâmetros da distribuição. Convém registar, desde já, este facto: *a distribuição normal é completamente determinada pelo centro e pelo desvio padrão.*

Substituindo x pelo desvio reduzido

$$h = \frac{x - \mu}{\sigma},$$

obtém-se a *distribuição normal estandardizada ou distribuição* $(N; 0,1)$, cuja função de densidade é

$$\varphi(h) = \frac{1}{\sqrt{2\pi}} e^{-\frac{h^2}{2}}.$$

(É claro que podemos passar a usar aqui x no papel de h).

Interpretemos esta distribuição como distribuição de probabilidade. A probabilidade dum desvio reduzido compreendido entre dois limites, λ_1, λ_2 , será dada pelo integral

$$\frac{1}{\sqrt{2\pi}} \int_{\lambda_1}^{\lambda_2} e^{-\frac{x^2}{2}} dx.$$

Se pusermos

$$\Phi(\lambda) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\lambda} e^{-\frac{x^2}{2}} dx,$$

será $\Phi(\lambda)$ a cumulante da distribuição e o valor do anterior integral será $\Phi(\lambda_2) - \Phi(\lambda_1)$. Note-se que existem tabelas com os valores de $\Phi(\lambda)$ para diferentes valores de λ .

Estudemos a função

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Começemos por notar que esta função é definida e positiva em todo o intervalo $]-\infty, +\infty[$ e que se tem $\varphi(-x) = \varphi(x)$ (curva simétrica a respeito do eixo dos yy).

Por outro lado, tem-se

$$\varphi'(x) = -\frac{x}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},$$

$$\varphi''(x) = -\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} + \frac{x^2}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} = (x^2 - 1) \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}.$$

Como se tem sempre $e^{-x^2/2} > 0$, a função $\varphi'(x)$ é nula para $x = 0$, positiva para $x < 0$ e negativa para $x > 0$. Por sua vez, a função $\varphi''(x)$ tem o sinal de $x^2 - 1$, sendo, portanto, negativa para $-1 < x < 1$, positiva para $x < -1$ ou $x > 1$ e nula para $x = \pm 1$.

Temos, então, os seguintes esquemas:

	$-\infty$	0	$+\infty$
$\varphi'(x)$	+	-	
$\varphi(x)$		↘	
		Máximo	

	$-\infty$	-1	$+1$	$+\infty$
$\varphi''(x)$	+	-	+	
$\varphi(x)$	∪	∩	∪	
		inflexão	inflexão	

Note-se, finalmente, que

$$\lim_{x \rightarrow \infty} \varphi(x) = 0,$$

o que significa que o eixo dos xx é assíntota do gráfico de φ . É fácil ver que não há outras assíntotas.

A curva representativa da função $\varphi(x)$ considerada, terá, pois, o aspecto indicado na figura.

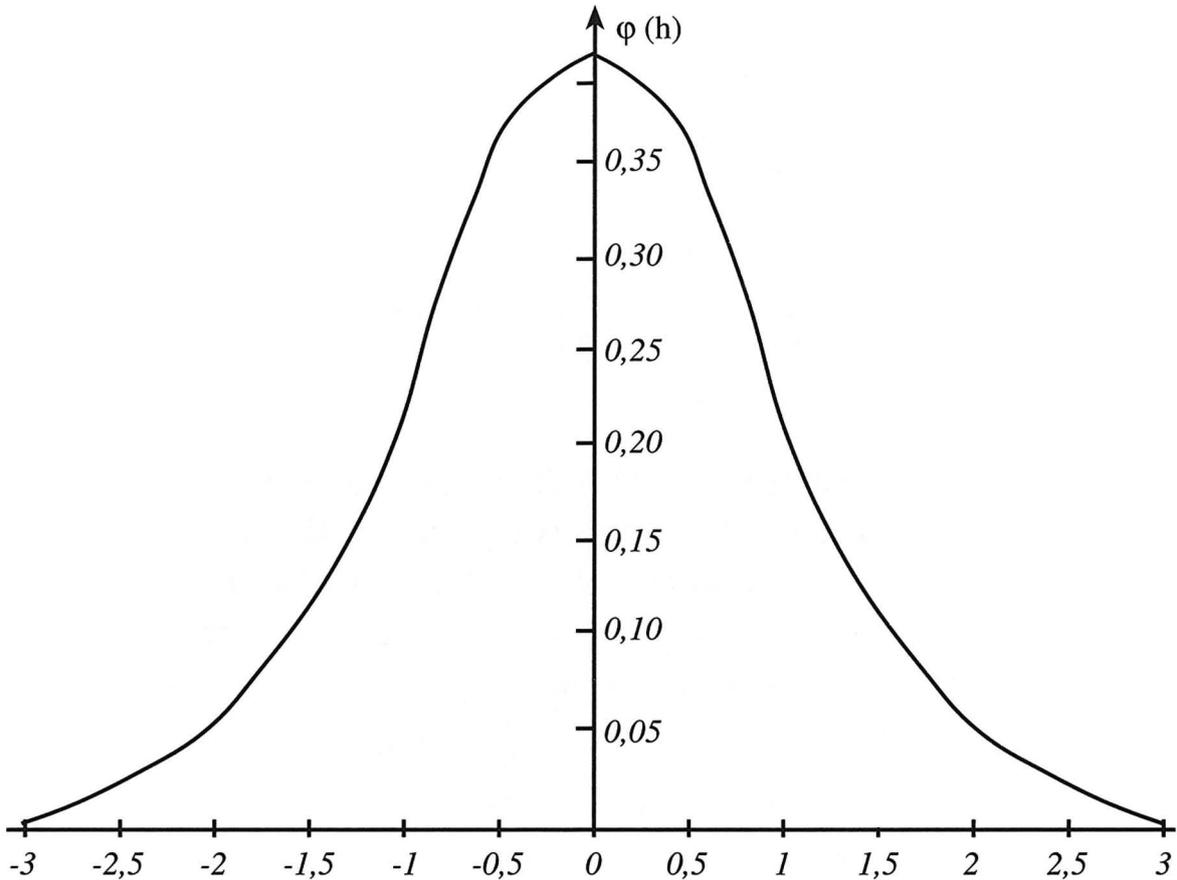


Fig. 7

Dá-se-lhe o nome de *curva de GAUSS* ou *curva em sino*.

Note-se que os pontos de inflexão têm as abcissas -1 , 1 correspondentes ao desvio padrão (unitário neste caso).

Daqui se deduz, logo, que o gráfico da função mais geral

$$\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

acusará um máximo no ponto da abcissa μ , será simétrico a respeito da recta $x = \mu$, terá inflexões nos pontos $\mu - \sigma$, $\mu + \sigma$, e admitirá o eixo dos xx como assíntota.

Por sua vez, a função $\Phi(x)$, cumulante da distribuição normal estandardizada, será crescente em todo o intervalo $]-\infty, +\infty[$, o seu

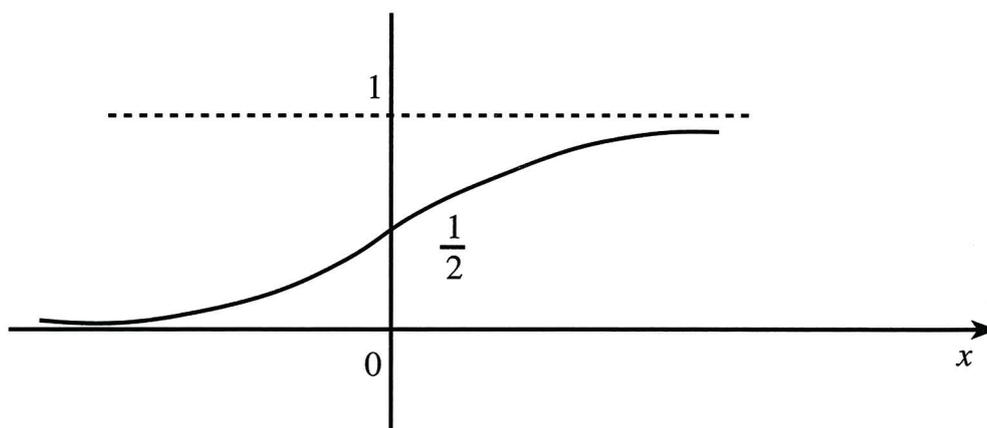


Fig. 8

gráfico terá uma inflexão no ponto de abcissa 0, a respeito do qual é simétrico, e admitirá como assintotas as rectas $y = 0$, $y = 1$, visto que $\lim_{x \rightarrow -\infty} \Phi(x) = 0$, $\lim_{x \rightarrow +\infty} \Phi(x) = 1$ ⁽¹⁾.

Costuma, ainda, escrever-se

$$\Theta(u) = \frac{1}{\sqrt{2\pi}} \int_0^u e^{-\frac{x^2}{2}} dx.$$

A função $\Theta(u)$ também se encontra tabelada. É claro que para cada valor γ de u , $\Theta(\gamma)$ é a área do trapezóide determinado pela curva de GAUSS no intervalo $[0, \lambda]$, igual à área do trapezóide correspondente ao intervalo $[0, -\lambda]$, visto a curva ser simétrica a respeito do eixo dos yy . Deste modo, a probabilidade dum desvio reduzido h compreendido entre $-\lambda$ e λ será

$$\Pr(-\lambda < h < \lambda) = 2\Theta(\lambda)$$

e a probabilidade dum desvio reduzido superior a λ , em valor absoluto, será $1 - 2\Theta(\lambda)$.

Por exemplo, a tabela n.º 3 reproduzida em YULE and KENDALL (loc. cit., pág. 533), dá os valores de $1 - 2\Theta(\lambda)$ a menos de 0,00001,

(1) – É claro que o facto de as rectas $y = 0$, $y = 1$ serem assintotas do gráfico se verifica para a cumulante de *qualquer* distribuição definida no intervalo $]-\infty, +\infty[$.

com primeiras e segundas diferenças. Para $\lambda = 3$, o valor registado é 0,00270; quer isto dizer que:

Em qualquer distribuição normal, a probabilidade dum desvio de módulo superior ao triplo do desvio padrão é um pouco inferior a 3%.

(O teorema de TCHEBICHEFF, que, como vimos, é aplicável a qualquer distribuição, dá para um tal desvio uma probabilidade inferior ou igual a $1/3^2 \approx 11\%$).

Para $\lambda = 4$, a mesma tabela dá o valor 0,00006 e, para $\lambda = 4,5$, o valor de 0,00001.

Exemplos – Numerosos são os exemplos concretos de variáveis casuais que seguem aproximadamente a lei normal. Assim, é-se geralmente induzido a considerar como *normalmente distribuídas* as variáveis:

altura, numa espécie, raça ou variedade de animais ou plantas (dentro de certos limites de idade);

diâmetro do tronco, numa conveniente população de árvores;

comprimento duma espiga, teor em proteínas, produtividade, etc., numa dada variedade de trigo;

teores em gordura, em proteínas e em hidratos de carbono do leite produzido por vacas duma determinada raça;

etc., etc.

Note-se que os agrupamentos biológicos, tais como as espécies, as raças e as variedades, constituem populações praticamente infinitas, que vêm do passado e se prolongam no futuro, com uma certa constância de caracteres. (Muitas vezes, em vez de populações praticamente infinitas como estas, consideram-se população que, embora numerosas e homogéneas, estão bem delimitadas no tempo e no espaço: por exemplo, uma mata constituída por indivíduos duma mesma variedade). Entre as constantes biométricas duma tal população, figuram, precisamente, os valores médios e os desvios padrões de atributos quantitativos como os precedentes. À semelhança do que sucede para as grandezas físicas, idealizam-se para esses parâmetros valores exactos, *dos quais se calculam valores aproximados em amostras casuais extraídas da população: a aproximação*

obtida considera-se tanto melhor quanto mais representativa, e portanto, mais numerosa for a amostra⁽¹⁾. (Convém recordar aqui as considerações que fizemos a propósito do conceito de probabilidade).

Nessas constantes biométricas e nos seus valores aproximados baseiam-se novos testes de significância, relativos, por exemplo, à comparação de duas variedades de trigo do ponto de vista de produtividade, ao efeito dum adubo ou dum insecticida, etc., etc. Trata-se, como se pode imaginar, de assuntos de maior interesse para o agrónomo e para o silvicultor, mas não é possível desenvolver neste curso.

Erros de observação – Invocámos há pouco o exemplo das grandezas físicas. Como se sabe, efectuando várias medições duma mesma grandeza, os resultados não concordam geralmente entre si. Admitida a existência duma medida exacta da grandeza, as diferenças $x - x_0$ entre os valores obtidos, x , e o valor exacto, x_0 , chamam-se *erros de observação*. Estes classificam-se em *sistemáticos* (devidos a uma causa bem determinada, que pode ser um defeito do instrumento de medida ou do observador) e *acidentais, fortuitos* ou *casuais* (dependentes do “acaso”). Suponhamos eliminados os erros sistemáticos. Então, admitidos certos axiomas, demonstra-se que os resultados de medição duma grandeza se distribuem segundo a lei normal, com centro no valor exacto da grandeza. Esta conclusão é confirmada pela experiência, de maneira satisfatória.

É claro que, sendo assim, os erros de observação também se distribuem normalmente, sendo 0 o seu valor médio. O desvio padrão σ chama-se, agora, *erro padrão* ou *erro quadrático médio*. Dá-se o nome de *parâmetro de precisão* ao número $1/(\sigma\sqrt{2})$, que permite avaliar o grau de concentração da distribuição considerada.

É claro que, na prática, se trabalha apenas com um número finito de valores aproximados,

$$x_1, x_2, \dots, x_N,$$

da grandeza medida, alguns dos quais podem aparecer repetidos.

(1) – A teoria da avaliação dos parâmetros duma distribuição, a partir de amostra da população, constitui um dos mais importantes capítulos da Estatística.

Estes valores constituem, por assim dizer, uma amostra dos possíveis resultados x da medição. Os parâmetros da distribuição de x na amostra, nomeadamente, o valor médio,

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_N}{N},$$

e o desvio padrão,

$$s = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_N - \bar{x})^2}{N}},$$

constituem valores aproximados, respectivamente, do valor exacto x_0 da grandeza, e do desvio padrão σ da distribuição normal considerada.

Propriedade reprodutiva da distribuição normal – Uma importante propriedade das distribuições normais é a seguinte, chamada PROPRIEDADE REPRODUTIVA ou TEOREMA DA ESTABILIDADE, que não demonstraremos:

Dadas n variáveis casuais x_1, x_2, \dots, x_n , independentes e normalmente distribuídas, toda a função linear

$$y = a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_n x_n$$

daquelas variáveis segue ainda a lei normal.

As proposições estabelecidas em C permitem-nos determinar o valor médio e o desvio padrão da variável y a que se refere este enunciado, em função dos valores médios e dos desvios padrões de x_1, x_2, \dots, x_n . Deverá ter-se, com efeito,

$$M\{y\} = a_0 + a_1 M\{x_1\} + \cdots + a_n M\{x_n\},$$

$$V\{y\} = a_1^2 V\{x_1\} + \cdots + a_n^2 V\{x_n\},$$

donde, designando por σ o desvio padrão de y e por $\sigma_1, \sigma_2, \dots, \sigma_n$, os desvios padrões de x_1, x_2, \dots, x_n :

$$\sigma = \sqrt{a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2 + \dots + a_n^2 \sigma_n^2}.$$

Como exemplo de aplicação, consideremos o caso duma amostra casual constituída por N valores independentes, x_1, x_2, \dots, x_N , duma mesma variável casual x , normalmente distribuída, com o valor médio μ e o desvio padrão σ . É claro que a média daqueles valores

$$\bar{x} = \frac{\sum x_i}{N} = \frac{1}{N} x_1 + \frac{1}{N} x_2 + \dots + \frac{1}{N} x_N$$

é uma função linear das variáveis casuais x_1, x_2, \dots, x_N , *independentes e normalmente distribuídas com os parâmetros μ, σ* . Então, segundo o teorema anterior, a média \bar{x} será, também, uma variável normalmente distribuída, tendo por valor médio

$$M\{\bar{x}\} = \frac{1}{N} \sum M\{x_i\} = \frac{N\mu}{N} = \mu,$$

isto é,

$$M\{\bar{x}\} = M\{x\},$$

e por desvio padrão

$$\sigma\{\bar{x}\} = \sqrt{\sum \frac{1}{N^2} \sigma^2} = \sqrt{\frac{N\sigma^2}{N^2}} = \sqrt{\frac{\sigma^2}{N}} = \frac{\sigma}{\sqrt{N}},$$

isto é,

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{N}}.$$

O desvio padrão da média \bar{x} obtém-se, pois, a partir do desvio padrão de x , dividindo este por \sqrt{N} .

Estes resultados são de grande importância, não só em estatística agronómica como, ainda, na teoria dos erros.

F – Convergência de distribuições. Relação entre as distribuições normal e binomial

Consideremos uma sucessão infinita de distribuições sobre a recta, cujas funções cumulantes sejam

$$\Phi_1(x), \Phi_2(x), \dots, \Phi_n(x), \dots$$

Diz-se que esta sucessão de distribuições *converge para* uma determinada distribuição, cuja cumulante seja $\psi(x)$, se a sucessão de funções $\Phi_n(x)$ converge uniformemente para $\psi(x)$ em todo o intervalo limitado J da recta.

Consideremos, em particular, o caso da distribuição binomial de x , que dá a probabilidade de que um acontecimento de probabilidade p se verifique x vezes em n provas. Supondo a probabilidade p fixa, e atribuindo a n os valores 1, 2, 3, ..., obtém-se uma sucessão de distribuições cujo termo geral é

$$\text{Pr}_n(x) = \binom{n}{x} p^x q^{n-x}, \quad \text{com } q = 1 - p.$$

É claro que cada uma destas distribuições é definida apenas para os valores inteiros de x tais que $0 \leq x \leq n$; mas podemos supô-la prolongada a toda a recta, atribuindo a probabilidade 0 a cada valor de x que não seja inteiro ou não verifique a condição $0 \leq x \leq n$. A cumulante da distribuição $\text{Pr}_n(x)$ acusará, então, saltos positivos nos pontos

$$0, 1, 2, \dots, n-1, n,$$

sendo constante no interior dos intervalos determinados por estes pontos: em particular, será nula no intervalo $]-\infty, 0[$ e igual a 1 no intervalo $[n, +\infty[$.

O valor médio e o desvio padrão de $\text{Pr}_n(x)$ são, respectivamente, como vimos,

$$\mu_n = np, \quad \sigma_n = \sqrt{npq}.$$

Efectuando em $\text{Pr}_n(x)$ a mudança de variável

$$x \rightarrow h = \frac{x - \mu_n}{\sigma_n} = \frac{x - np}{\sqrt{npq}},$$

obtém-se a *distribuição binomial estandardizada*. Pois bem, demonstra-se o seguinte teorema, de importância fundamental em Cálculo das Probabilidades e Estatística:

TEOREMA. *A distribuição binomial estandardizada converge para a distribuição normal estandardizada quando $n \rightarrow \infty$.*

Para a demonstração, deveras delicada, pode ver-se, por exemplo, a obra de CRAMÉR já citada.

Na prática, o anterior resultado interpreta-se do seguinte modo:

Para valores de n elevados, a distribuição binomial aproxima-se da normal. A aproximação aumenta quando n cresce, mas diminui quando o desvio reduzido aumenta.

G – A distribuição de χ^2 de PEARSON

Vimos que toda a função linear de variáveis independentes normalmente distribuídas também é normalmente distribuída. Porém, uma função *não linear* de variáveis independentes normalmente distribuídas não tem, geralmente, distribuição normal. Com efeito, demonstra-se o seguinte teorema:

Dadas n variáveis casuais x_1, x_2, \dots, x_n , independentes e com distribuição normal estandardizada, a raiz quadrada da soma dos seus quadrados, que costuma ser designada por χ :

$$\chi = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2}$$

tem uma distribuição cuja densidade é uma função de χ da forma

$$\varphi(\chi) = C e^{-\frac{\chi^2}{2}} \chi^{n-1},$$

sendo C uma constante.

Para a demonstração, veja-se, por exemplo, P. de VARENNES E MENDONÇA, obra citada.

Note-se que, sendo a raiz aritmética de $\sum x_i^2$, o valor de χ será sempre não negativo: o domínio de distribuição é, pois, o intervalo $[0, +\infty[$. Deverá ter-se, então,

$$\int_0^{+\infty} \varphi(\chi) d\chi = 1,$$

o que permite determinar a constante C :

$$C = 1 : \int_0^{+\infty} e^{-\frac{\chi^2}{2}} \chi^{n-1} d\chi.$$

Deste modo, a probabilidade de que χ^2 seja superior a um dado valor χ_0^2 (igual à probabilidade de que seja $\chi \geq \chi_0$), será dada pela fórmula

$$\Pr(\chi^2 \geq \chi_0^2) = C \int_{\chi_0}^{+\infty} e^{-\frac{\chi^2}{2}} \chi^{n-1} d\chi,$$

com o valor de C dado pela fórmula anterior.

O que interessa, na prática, não é propriamente a distribuição de χ , mas sim a distribuição de χ^2 , que, evidentemente, se reduz imediatamente à anterior como mostra a última fórmula.

Como se vê, a anterior distribuição de χ^2 (chamada distribuição de PEARSON) tem como único parâmetro a variável n , que recebe aqui o nome de *número de graus de liberdade* da distribuição. Para indicar a distribuição de χ^2 com n graus de liberdade, usa-se a notação $(\chi^2; n)$.

O teorema atrás enunciado admite a seguinte generalização:

TEOREMA. *Dadas M variáveis x_1, x_2, \dots, x_M , todas com distribuição normal standardizada, sendo n dessas variáveis independentes (estocasticamente) e as restantes funções lineares homogêneas das primeiras, a variável*

$$\chi^2 = x_1^2 + x_2^2 + \dots + x_M^2$$

tem a distribuição de PEARSON com n graus de liberdade.

Para a demonstração, veja-se o trabalho do Prof. P. de VARENNES E MENDONÇA, “Das distribuições estatísticas mais usadas em provas de significação”, publicado na Revista Agronômica XXVIII (1940).

Este importante resultado permite ver como a distribuição considerada pode ser usada em testes de significância (ou *provas de significação*) aplicáveis a tábuas de contingência, ajustamento de curvas ou superfícies, etc.

Consideremos, por exemplo, uma partição em M atributos $\alpha_1, \alpha_2, \dots, \alpha_M$, num universo infinito U , e seja p_i a probabilidade de que um indivíduo escolhido ao acaso tenha o atributo α_i ($i = 1, 2, \dots, M$). Consideremos, por outro lado, uma amostra casual constituída por N indivíduos. Será, então, sensivelmente,

$$p_i N$$

o valor *esperado* da frequência absoluta do atributo α_i na amostra (supondo N bastante grande). Designemos por m_i este valor e por o_i o valor observado (da referida frequência absoluta). A discrepância

$$\delta_i = o_i - m_i$$

terá, então, o valor médio $M\{\delta_i\} = 0$. Além disso, *supondo N bastante grande e nenhum dos p_i demasiado pequeno*, demonstra-se que *cada uma das variáveis δ_i se distribui normalmente em torno de 0, com um desvio padrão σ_i tal que*

$$\sigma_i^2 = m_i.$$

Então, supondo que n dos desvios δ_i são independentes estocasticamente, sendo os restantes função linear homogénea dos primeiros (como sucede nas tábuas de contingência), o anterior teorema habilita-nos a afirmar que a variável

$$\chi^2 = \sum \frac{\delta_i^2}{m_i} = \sum \left(\frac{\delta_i}{\sigma_i} \right)^2$$

tem aproximadamente a distribuição de PEARSON para n graus de liberdade, visto que cada uma das variáveis δ_i/σ_i tem valor médio nulo e desvio padrão unitário.

E assim ficam esboçados os fundamentos teóricos do teste do χ^2 . A falta de tempo impede-nos de aprofundar este assunto e de abordar o estudo de outras distribuições de grande interesse em Estatística agronómica. Para complementos, lembramos as obras de Estatística já citadas, bem como o referido trabalho do Prof. Varennes e Mendonça.

A obra de R. A. FISHER e F. YATES, “*Statistical Tables for Biological, Agricultural and Medical Research*” (Oliver and Boyd, Edinburgh and London) contém tábuas referentes a várias distribuições importantes; a tábua IV desta obra fornece, para diferentes valores de n e de p , o valor de χ_0^2 tal que $\Pr(\chi^2 \geq \chi_0^2) = p$, na distribuição de χ^2 de PEARSON com n graus de liberdade.

NOTA SOBRE A AVALIAÇÃO DA VARIÂNCIA

Embora não possamos aqui abordar o estudo da teoria da avaliação, convém, desde já, esclarecer um ponto.

É fácil ver que o valor esperado da variância, S^2 , numa amostra com N elementos, está relacionada com a variância, σ^2 , na população, por meio da fórmula

$$M\{S^2\} = \frac{N-1}{N} \sigma^2.$$

Daqui, o considerar o valor $(N/N-1)S^2$ como a *melhor avaliação da variância* σ^2 , a partir da amostra. Este valor é representado por $\hat{\sigma}^2$. Será, portanto,

$$\hat{\sigma} = S \sqrt{\frac{N}{N-1}}$$

a melhor avaliação do desvio padrão, σ .

Dum modo geral, o acento circunflexo sobre um símbolo é usado para indicar a melhor avaliação do parâmetro designado por esse símbolo.

Na fórmula anterior, o coeficiente $N/N-1$ é chamado *correção de BESSEL*. É claro que, para valores de N elevados, aquele coeficiente é sensivelmente igual a 1; a correção só é, portanto, necessária para pequenas amostras.

ÍNDICE

CÁLCULO DAS PROBABILIDADES

ADVERTÊNCIA PRÉVIA	317
I.4.1 INTRODUÇÃO AO CÁLCULO DAS PROBABILIDADES: POPULAÇÕES FINITAS	319
A – Frequências	319
1. Primeiros exemplos	319
2. Populações. Álgebra dos atributos	322
3. Álgebra dos acontecimentos	325
4. Acontecimentos expressos em forma proposicional	326
5. Frequência dum atributo numa população	328
6. Frequência de um acontecimento numa série de provas	330
7. Partições	331
8. Corpos de conjuntos, corpos de atributos, corpos de acontecimentos	333
9. Distribuição em universos finitos	337
10. Soma de conjuntos não disjuntos (atributos ou acontecimentos compatíveis)	339
11. Atributos quantitativos	341
12. Representação gráfica das distribuições: histogramas e polígonos de frequência	345
13. Independência e associação de atributos. Distribuições de duas ou mais variáveis	348

14. Associação e independência de partições múltiplas. Tábuas de contingência	355
15. Associações parciais de atributos. Independência de atributos no caso em que o seu número é superior a dois	359
16. Interpretação de uma tábua de contingência. Testes de significância	363
B – Probabilidades	373
1. Lógica indutiva	373
2. Lógica dedutiva	376
3. Conceito natural de probabilidade	378
4. Axiomatização do conceito de probabilidade.	382
5. Alguns exemplos de cálculo de prodabilidades <i>a priori</i>	385
6. Independência e associação de acontecimentos	396
7. Sistema de duas experiências	398
8. Sistema de várias experiências	403
9. Distribuição binomial ou de Bernoulli	404
10. Conceito de moda. Caso da distribuição normal	409
11. Distribuição polinomial. Amostras casuais	411
BIBLIOGRAFIA	414

I.4.2 APONTAMENTOS DE CÁLCULO DAS PROBABILIDADES	415
A – Distribuições de uma variável contínua real	415
B – Valores médios para distribuições de uma variável real	425
C – Valores médios para distribuições de mais de uma variável real	438
D – Aplicação à distribuição binomial. Teorema de BERNOULLI	451
E – Distribuição normal	455
F – Convergência de distribuições. Relação entre as distribuições normal e binomial.	464
G – A distribuição de χ^2 de PEARSON	465
NOTA SOBRE A AVALIAÇÃO DA VARIÂNCIA	469

I.4.3 ADITAMENTO ÀS LIÇÕES DE CÁLCULO DE PROBABILIDADES	471
A – Regressões. Ajustamentos. Correlação	471
1. Formulação geral do problema	471
2. Ajustamentos pelo método dos mínimos quadrados	476
3. Outra forma das equações normais	479
4. Regressão polinomial	480
5. Regressão linear. Correlação	481
6. Segunda recta de regressão	485
7. Organização prática dos cálculos	487
8. Ajustamentos com mudanças não lineares de variáveis	490
9. Regressão múltipla. Índice de correlação em geral	493
10. Nota sobre as notações	497
B – Distribuições de STUDENT e de FISHER.	
Suas aplicações	497
1. A melhor estimativa do desvio padrão deduzida duma amostra	497
2. Distribuição t de STUDENT	499
3. A melhor estimativa de σ deduzida a partir de várias amostras	501
4. Distribuição da diferença entre duas médias	502
5. Prova do t (de significação)	503
6. Intervalos de tolerância e intervalos de confiança	506
7. Intervalos de confiança quando não é conhecido o σ da população	509
8. Aplicações agronómicas	510
9. Distribuição de F e z de FISHER	511
TÁBUA DA DISTRIBUIÇÃO NORMAL	512
TÁBUA DA DISTRIBUIÇÃO DE t DE STUDENT	513
TÁBUA DA DISTRIBUIÇÃO DE χ^2 DE PEARSON	514
INDICAÇÕES BIBLIOGRÁFICAS	515