# A Esperança Matemática e Algumas das suas Aplicações à Teoria da Amostragem

FREDERICO PIMENTEL GOMES

Professor Substituto de Matemática da E. S. A.
"Luiz de Queiroz"

ÍNDICE

## 1 — INTRODUÇÃO

Êste trabalho consta de duas partes bem distintas. Na primeira delas, que abrange os capítulos 2, 3, 4 e 5, examinamos o caso de uma variável casual definida num conjunto numerável. As demonstrações aí feitas são de carater elementar. Na segunda parte, em que discutimos o caso de uma variável contínua, somos forçados a lançar mão de métodos mais elevados, que estão fora do alcance dos que não têm conhecimentos especializados de Matemática.

De todos os métodos usados no estudo das distribuições das pequenas amostras o método das funções características nos parece o mais simples e elegante. Daí o fato de o termos preferido para as demonstrações que serão vistas. Os interessados acharão, porém, numa obra de KENDAL [1, 1° vol., pp. 231-253] ou num livro de USPENSKY [2, pp. 331-346], entre muitos, outros métodos interessantes que permitem resolver problemas análogos.

## 2 — DEFINIÇÃO

Se $x$ é uma variável casual ou aleatória, a *esperança matemática* de $x$, ou o seu *valor médio* é

$$E(x) = x1\ p2 + x2\ p2 + \ldots + xn\ pn\,.$$

Supõe-se aí, implìcitamente, que o conjunto C em que está definida a variável $x$ é numerável.

## 3 — PROPRIEDADES

Antes de tudo, é evidente que, sendo $k$ uma constante, temos

$$E(kx) = kE(x)\,.$$

Com efeito, podemos escrever:

$$\begin{aligned} E(kx) &= k\,x_1p_1 + k\,x_2p_2 + \ldots + k\,x_np_n \\ &= k\,(x_1p_1 + x_2p_2 + \ldots + x_np_n) \\ &= k\,E(x)\,. \end{aligned}$$

Consideremos, agora, duas variáveis casuais $x$ e $y$. Representemos por $p_{ij}$ a probabilidade da ocorrência simultânea de $x_i$ e $y_j$ e obtemos:

$$\begin{aligned} E(x+y) = {} & p_{11}\,(x_1+y_1) + p_{12}\,(x_1+y_2) + \ldots + p_{1n}\,(x_1+y_n) + \\ & + p_{21}\,(x_2+y_1) + p_{22}\,(x_2+y_2) + \ldots + p_{2n}\,(x_2+y_n) + \\ & + \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots + \end{aligned}$$

$$+ p_{n1}(x_n + y_1) + p_{n2}(x_n + y_2) + \dots + p_{nn}(x_n + y_n),$$

$$E(x+y) = x_1(p_{11}+p_{12}+ \dots + p_{1n}) + x_2(p_{21}+p_{22}+ \dots + p_{2n}) +$$
$$+ \dots + x_n(p_{n1}+p_{n2}+ \dots + p_{nn}) + y_1(p_{11}+p_{21}+ \dots + p_{n1}) +$$
$$+ y_2(p_{12}+p_{22}+ \dots + p_{n2}) + \dots + y_n(p_{1n}+p_{2n}+ \dots + p_{nn}).$$

Mas, sendo $P_i$ a probabilidade correspondente a $x_i$, e $Q_i$, a $y_i$, temos por um teorema do Cálculo de Probabilidades (3, pp. 16-17):

$$P_i = p_{i1} + p_{i2} + \dots + p_{in},$$
$$Q_i = p_{1i} + p_{2i} + \dots + p_{ni}.$$

Logo obtemos

$$E(x+y) = x_1 P_1 + x_2 P_2 + \dots + x_n P_n +$$
$$+ y_1 Q_1 + x_2 Q_2 + \dots + x_n Q_n,$$
$$\therefore E(x+y) = E(x) + E(y).$$

Suponhamos, agora, que **x** e **y** sejam variáveis casuais independentes, isto é, que a probabilidade $p_{ij}$ de ocorrerem simultâneamente $x_i$ e $y_j$ seja

$$p_{ij} = P_i \cdot Q_j,$$

onde $P_i$ é a probabilidade de ocorrer $x_i$, e $Q_j$, a de ocorrer $y_j$. Temos, então,

$$E(xy) = p_{11}x_1y_1 + p_{12}x_1y_2 + \dots + p_{1n}x_1y_n +$$
$$+ p_{21}x_2y_1 + p_{22}x_2y_2 + \dots + p_{2n}x_2y_n +$$
$$+ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots +$$
$$+ p_{n1}x_ny_1 + p_{n2}x_ny_2 + \dots + p_{nn}x_ny_n,$$

$$E(xy) = (P_1x_1 + P_2x_2 + \dots + P_nx_n)(Q_1y_1 + Q_2y_2 + \dots + Q_ny_n)$$
$$= E(x) \cdot E(y)$$

Se **k** é uma constante, temos ainda

$$E(k) = k\,p_1 + k\,p_2 + \dots + k\,p_n$$
$$= k(p_1 + p_2 + \dots + p_n)$$
$$= k.$$

## 4 — SIGNIFICADO ESTATÍSTICO

Suponhamos que o conjunto de valores da variável casual $x$ constitui uma população de $N$ elementos. A probabilidade de extrair um determinado $x$ qualquer será 1/N. Logo teremos

$$E(x) = \frac{1}{N} x_1 + \frac{1}{N} x_2 + \dots + \frac{1}{N} x_N = \frac{\Sigma x}{N} = \bar{x}$$

**Logo $E(x)$ é a média aritmética $\bar{x}$ da população.**

**Temos ainda**

$$E(x-\bar{x}) = E(x) - E(\bar{x}) = \bar{x}-\bar{x} = 0$$

$$\begin{aligned} E\left[(x-\bar{x})^2\right] &= E\,(x^2 - 2\bar{x}x + \bar{x}^2) \\ &= E(x^2) - 2\bar{x}\,E(x) + E(\bar{x}^2) \\ &= E(x^2) - 2\bar{x}^2 + \bar{x}^2 \\ &= E(x^2) - \bar{x}^2 \end{aligned}$$

$$= \frac{\Sigma x^2}{N} - \left[\frac{\Sigma x}{N}\right]^2$$

$$= \frac{\Sigma x^2 - (\frac{\Sigma x}{N})^2}{N} = \sigma^2 = \frac{\Sigma (x-\bar{x})^2}{N} .$$

**Logo $E(x-\bar{x})^2$ não é mais que a variância $\sigma^2$.**

## 5 — APLICAÇÕES À TEORIA DA AMOSTRAGEM

Suponhamos que da população constituída de todos os valores de $x$ extraímos, ao acaso, uma amostra de $n$ elementos. Podemos, então, com os dados da amostra, tentar estimar os parâmetros $\tilde{x}$ e $\sigma^2$ da população. Somos tentados a tomar

$$\bar{x} = \frac{\Sigma x}{n} \quad \text{e} \quad s^2 = \frac{\Sigma (x-\bar{x})^2}{n}$$

como estimativas, respectivamente, de $\tilde{x}$ e de $\sigma^2$. A conveniência das estimativas, de um modo geral, é um problema complexo, que não abordaremos agora. Uma exigência usual, porém, é que a estimativa deve ser "unbiased", isto é, imparcial ou não tendenciosa. Diz-se que uma estimativa $p$ de um parâmetro P da população é "unbiased" quando temos

$$E(p) = P.$$

Será que as estimativas $\bar{x}$ e $s^2$ são imparciais?

Para $\bar{x}$ temos logo

$$E(\bar{x}) = E\,\frac{x_1 + x_2 + \dots + x_n}{n} = \frac{n\,E(x)}{n} = E(x) = \bar{x}$$

e é, portanto, imparcial.

O caso de $s^2$ é um pouco mais difícil. Temos:

$$s^2 = \frac{\Sigma(x-\bar{x})^2}{n}$$

$$= \frac{\Sigma x^2 - 2\bar{x}\Sigma x + n\,\bar{x}^2}{n} = \frac{\Sigma x^2}{n} - \bar{x}^2 .$$

Consideremos duas variáveis, $x$ e $y$, não independentes.

Temos:

$$E(\bar{x}.\bar{y}) = E\left[\frac{x_1+x_2+\dots+x_n}{n} \cdot \frac{y_1+y_2+\dots+y_n}{n}\right]$$

$$= \frac{1}{n^2}\; E\left[(x_1 + x_2 + \dots + x_n)(y_1 + y_2 + \dots + y_n)\right],$$

$$E(\bar{x}.\bar{y}) = \frac{1}{n^2}\; E\begin{bmatrix} x_1y_1 + x_1y_2 + \dots + x_1y_n + \\ x_2y_1 + x_2y_2 + \dots + x_2y_n + \\ + \dots\dots\dots\dots\dots\dots\dots + \\ + x_ny_1 + x_ny_2 + \dots + x_ny_n \end{bmatrix}$$

Nos $n$ produtos $x_1y_1, x_2y_2, \ldots, x_ny_n$, as variáveis são correlacionadas. Mas nos $n^2 - n$ restantes não pode haver correlação. Logo temos

$$E(\bar{x}.\bar{y}) = \frac{1}{n^2}\left[ n\,E\,(x\,y) + (n^2 - n)\;E\,(x)\;E\,(y)\right]$$

$$= \frac{E(x\,y) + (n-1)\,E\,(x)\,E\,(y)}{n}$$

Para $x = y$ temos, então,

$$E(\bar{x}^2) = \frac{E(x^2) + (n-1)\,\bar{x}^2}{n}$$

Logo, vamos obter

$$E\,(s^2) = E\left[\frac{\Sigma x^2}{n} - \bar{x}^2\right]$$

$$= \frac{E(\Sigma x^2)}{n} - E\,(\bar{x}^2)$$

$$= \frac{n\,E\,(x^2)}{n} - \frac{E(x^2) + (n-1)\,\bar{x}^2}{n}$$

$$= \frac{n-1}{n}\left[E\,(x^2) - \bar{x}^2\right] = \frac{n-1}{n}\,\sigma^2.$$

Logo $s^2$ é uma estimativa tendenciosa de $\sigma^2$. Mas

$$s_1^2 = \frac{\Sigma(x-\bar{x})^2}{n-1} = \frac{n}{n-1}\,s^2$$

é imparcial, pois temos então

$$E(s_1^2) = \frac{n}{n-1}\,E(s^2) = \frac{n}{n-1}\cdot\frac{n-1}{n}\,\sigma^2 = \sigma^2.$$

E fica demonstrado elementarmente porque, ao estimar a variância, se divide a soma dos quadrados dos desvios por *n — 1*, e não por *n*.

## 6 — VARIÁVEL CASUAL CONTÍNUA

Quando o conjunto C, campo de definição da variável casual $x$, é do tipo do contínuo, somos obrigados a substituir as probabilidades por diferenciais e as somatórias por integrais.

Sendo $y = f(x)$ a equação de uma curva de frequências, a probabilidade de obter um valor entre $x - \frac{1}{2} dx$ e $x + \frac{1}{2} dx$ será

$$dF = f(x)\ dx.$$

A esperança matemática de uma função $g(x)$ será, por definição,

$$E\left[g(x)\right] \int_{-\infty}^{\infty} g(x)\ f(x)\ dx\ .$$

No caso da <u>curva normal</u> temos

$$dF = \frac{1}{\sigma\sqrt{2\pi}}\ e^{-\frac{(x-\bar{x})}{2\sigma^2}}\ dx\ .$$

A média será

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

e a variância é

$$E\left[(x-\bar{x})^2\right] = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-\bar{x})^2\ e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}\ dx = \sigma^2.$$

É importante a <u>função caraterística</u>

$$\phi\ (t) = E(e^{itx}) = \int_{-\infty}^{\infty} e^{itx}\ f(x)\ dx\ ,$$

na qual $i = \sqrt{-1}$ .

Pode-se demonstrar (1, 1º vol., pp. 91 - 94 ) que

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi\ (t)\ .\ e^{-itx}\ dx.$$

No caso de n variáveis independentes $x_1$ , $x_2$ ,..., $x_n$ , prova-se fàcilmente (1, 1º vol., pp. 104) que

$$\phi(t) = \phi_1(t) \cdot \phi_2(t) \dots \phi_n(t),$$

onde

$$\phi_j(t) = \int_{-\infty}^{\infty} e^{itx_j} f(x_j)\, dx_j ,$$

para j = 1, 2, ... , n.

No caso de uma distribuição normal de média zero e desvio "standard" $\sigma$ a função caraterística é

$$\phi(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx}\, e^{-\frac{x^2}{2\sigma^2}}\, dx = e^{-\frac{1}{2} t^2 \sigma^2}$$

## 7 — DISTRIBUIÇÃO DA MÉDIA

Consideremos a média $\bar{x}$ de uma amostra extraída de uma população normalmente distribuída com média zero e desvio "standard" $\sigma$ . Qual será a distribuição dessa média x ?

Temos
$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$
$$= \frac{x_1}{n} + \frac{x_2}{n} + \dots + \frac{x_n}{n} .$$

Para a primeira variável $x_1$ vamos ter

$$\phi_1(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{it\frac{x}{n}}\, e^{-\frac{x^2}{2\sigma^2}}\, dx$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\left(\frac{x^2}{2\sigma^2} - \frac{it}{n} x\right)}\, dx.$$

Mas

$$\int_{-\infty}^{\infty} e^{-(ax^2+2bx)}\, dx = \int_{-\infty}^{\infty} e^{-a(x+\frac{b}{a})^2 + \frac{b^2}{a}}\, dx$$

$$= \frac{e^{\frac{b^2}{a}}}{\sqrt{a}} \int_{-\infty}^{\infty} e^{-u^2}\, du = \sqrt{\frac{\pi}{a}}\; e^{\frac{b^2}{a}} ,$$

pois, como se pode demonstrar ( 4, pp. 5-7) , a última integral vale $\sqrt{\pi}$ .

No caso acima temos $a = \frac{1}{2\sigma^2}$ , $b = -\frac{it}{2n}$ . Logo, vamos ter

$$\phi_1(t) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \sqrt{2\pi\sigma^2}\; e^{-\frac{t^2\sigma^2}{2n^2}} = e^{-\frac{t^2\sigma^2}{2n^2}} .$$

Logo

$$\phi(t) = \phi_1(t) \cdot \phi_2(t) \ldots \phi_n(t) = e^{-\frac{t^2\sigma^2}{2n}}$$

Por comparação com a função característica da distribuição normal, concluimos logo que a distribuição de x se dá segundo uma nova curva normal com a mesma média zero e com variância igual a $\frac{\sigma^2}{n}$ , isto é, a distribuição é dada pela diferencial

$$dF = \frac{\sqrt{n}}{\sigma\sqrt{2\pi}}\; e^{-\frac{x^2 n}{2\sigma^2}}\, dx.$$

8. A ESPERANÇA DE $s^2$ - Já sabemos que

$$s^2 = \frac{\Sigma(x-\bar{x})^2}{n} = \frac{\Sigma x^2}{n} - \bar{x}^2 .$$

Logo temos

$$E(s^2) = \frac{E(\Sigma x^2)}{n} - E(\bar{x}^2) = E(x^2) - E(\bar{x}^2)$$

Mas, pelo que vimos,

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

$$= \frac{e}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2\sigma^2}} dx .$$

Tomemos $u = \frac{x^2}{2\sigma^2}$ e obteremos

$$E(x^2) = \frac{2\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{\infty} u^{\frac{1}{2}} e^{-u} du = \frac{2\sigma^2}{\sqrt{\pi}} \Gamma\left(\frac{3}{2}\right) = \sigma^2 .$$

Do mesmo modo provamos que

$$E(\bar{x}^2) = \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-\frac{x^2 n}{2\sigma^2}} dx = \frac{\sigma^2}{n} .$$

Logo temos

$$E(s^2) = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n} \sigma^2 ,$$

de onde se segue que uma estimativa imparcial da variância será

$$s_1^2 = \frac{n}{n-1} s^2 = \frac{\Sigma(x-\bar{x})^2}{n-1} .$$

## 9 — UM CASO MAIS GERAL

Vamos supor que a amostra de *n* elementos encerra *m* amostras parciais, cada uma com *k* elementos. E' claro que temos

$$n = m.k.$$

Em cada amostra parcial poderemos obter uma estimativa da variância.

$$s_j^2 = \frac{\sum\limits_{i=1}^{k} (x_{ij}-\bar{x}_j)^2}{n-1} \quad ,$$

onde representamos por $x_{ij}$ $(i=1,2, \ldots , k)$ os elementos da amostra parcial $j$.

Vamos admitir a hipótese de que

$$E(s_j^2) = \sigma^2 ,$$

para $j=1,2,\ldots, m$, isto é, que tôdas as amostras parciais nos dão estimativas de uma mesma variância $\sigma^2$. Vamos mostrar que

$$s^2 = \frac{\Sigma s_j^2}{m}$$

é uma estimativa imparcial de $\sigma^2$. Com efeito teremos

$$E(s^2) = E\left[\frac{\Sigma s_j^2}{m}\right] = \frac{m\,\sigma^2}{m} = \sigma^2$$

Mas então teremos ainda

$$s^2 = \frac{\sum\limits_{j=1}^{m} \frac{\sum\limits_{i=1}^{k}(x_{ij}-\bar{x}_j)^2}{k-1}}{m}$$

$$= \frac{\sum\limits_{j=1}^{m}\sum\limits_{i=1}^{k} (x_{ij}-\bar{x}_j)^2}{mk-m}$$

$$= \frac{\sum\limits_{j=1}^{m}\sum\limits_{i=1}^{k} (x_{ij}-\bar{x}_j)^2}{n-m}$$

Logo, quando se consideram $n$ elementos distribuídos em $m$ amostras parciais e se calculam desvios em relação às suas médias, a estimativa da variância será dada pela soma dos quadrados dos desvios dividida por $n - m$, que é o *grau de liberdade* da estimativa.

## 10 — BIBLIOGRAFIA CITADA

1 — KENDALL, Maurice G. — The advanced Theory of Statistics. 3a. edição, 1947. Londres.

2 — USPENSKY, J. V. — Introduction to Mathematical Probability. 1a. edição, 1937. Nova York.

3 — CASTELNUOVO, Guido — Calcolo delle Probabilitá. 3a. edição, 1947. Bolonha.

4 — PIMENTEL GOMES, Frederico — Introdução ao Estudo dos Derigrais. 1948. Piracicaba.