

PROBABILIDADE & ESTATÍSTICA¹

Rodolfo Hoffmann

3) PROBABILIDADE

Experimento é todo fenômeno ou ação que pode ser repetido e cujo resultado é casual ou aleatório. Espaço amostral é o conjunto de todos os resultados possíveis de um experimento, mutuamente exclusivos e exaustivos. Evento é todo subconjunto do espaço amostral e denominamos evento simples ou elementar ou todo subconjunto unitário do espaço amostral.

→ Conceito Moderno da Probabilidade

Seja E um espaço amostral e seja A qualquer evento em E , isto é, A é um subconjunto de E . Por definição, a probabilidade de ocorrer A é dada pela medida de A , nas seguintes condições:

1) A medida do universo é 1. $P(E) = 1$

2) A medida é não-negativa. $P(A) \geq 0$

3) Se A e B são dois eventos disjuntos, $P(A \cup B) = P(A) + P(B)$

Destes postulados segue-se que $0 \leq P(A) \leq 1$

→ Probabilidade do complemento e o Teorema da Soma

A probabilidade de que A não ocorra é dada por $P(\bar{A}) = 1 - P(A)$.

Se A e B são dois eventos do espaço amostral E , então a probabilidade de que ocorra ou A ou B é

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Se A e B são disjuntos, isto é, $A \cap B = \emptyset$, então

$$P(A \cup B) = P(A) + P(B)$$

→ Probabilidade Condicional e Teorema do Produto

A probabilidade condicional, isto é, a probabilidade de A dada que B ocorreu, é:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Se A e B são dois eventos do espaço amostral E, a probabilidade da ocorrência simultânea de A e B é dada por

$$P(A \cap B) = P(A) \cdot P(B|A) = P(B) \cdot P(A|B)$$

Por definição, dizemos que o evento A é estatisticamente independente do evento B se $P(A|B) = P(A)$. O fato de A ser independente de B implica que B é independente de A; no caso de dois eventos independentes temos $P(A \cap B) = P(A) \cdot P(B)$.

Dados m diferentes elementos, o número de combinações de n elementos que podemos fazer é

$$C_m^n = \binom{m}{n} = \frac{m!}{n!(m-n)!}$$

Dados m diferentes elementos, o n.º de permutações que podemos fazer é $m!$.

Teorema de Bayes

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Em geral, se $i=1,2,\dots,n$ e $j=1,2,\dots,m$,

$$P(e_i|f_j) = \frac{P(e_i \cap f_j)}{\sum_{h=1}^n P(e_h \cap f_j)} = \frac{P(e_i) \cdot P(f_j|e_i)}{\sum_{h=1}^n P(e_h) \cdot P(f_j|e_h)}$$

Permite calcular as probabilidades "a posteriori", isto é, probabilidade dos resultados do primeiro estágio dada o resultado do segundo estágio.

5) MEDIDAS DE TENDÊNCIA CENTRAL

3

A média aritmética de n dados $(x_1, x_2, \dots, x_n)$ é por definição

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i$$

Se os dados estiverem classificados em uma distribuição de frequências com K classes, se X_j ($j=1, 2, \dots, K$) são os valores centrais das classes ou os valores observados (no caso de uma variável discreta) e se f_j são as respectivas frequências, a média aritmética é dada por

$$\bar{x} = \frac{1}{n} \sum_{j=1}^K X_j f_j$$

A mediana de um conjunto de dados é o valor tal que metade dos dados são iguais ou inferiores a esse valor e metade dos dados são iguais ou superiores a esse valor. Se o número de dados é par, por convenção a mediana é a média dos dois valores centrais.

A mediana também pode ser calculada em caso de distribuição de frequências; localiza-se a classe que contém a mediana (classe mediana) e:

$$D = L_h + \frac{1}{f_h} \left(\frac{n}{2} - \sum_{j=1}^{h-1} f_j \right) (L_{h+1} - L_h)$$

Onde se supõe que a classe mediana seja a h -ésima, D é a mediana, L_h é o limite inferior de h , L_{h+1} o limite superior, n o número de dados observados, f_h é a frequência da classe mediana e $\sum_{j=1}^{h-1} f_j$ é a soma das frequências das classes inferiores à classe mediana.

A moda de um conjunto de dados é o valor que ocorre com maior frequência. Quando os dados estão em distribuição de frequências, pode-se adotar o valor central da classe modal, ou a fórmula

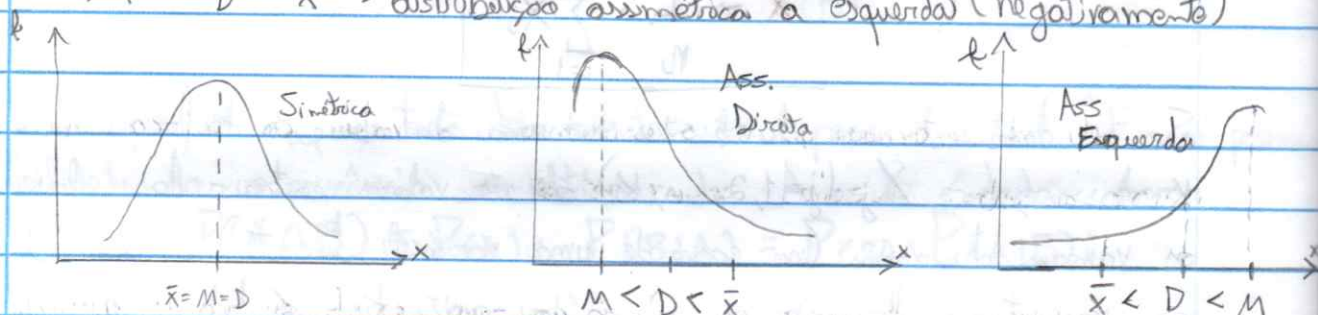
$$M = L_h + \frac{f_h - f_{h-1}}{2f_h - f_{h-1} - f_{h+1}} (L_{h+1} - L_h)$$

→ Posição Relativa

$M = \bar{X} = D \Rightarrow$ distribuição simétrica

$M < D < \bar{X} \Rightarrow$ distribuição assimétrica à direita (positivamente)

$M > D > \bar{X} \Rightarrow$ distribuição assimétrica à esquerda (negativamente)



A média geométrica de n valores não-negativos é

$$G = \sqrt[n]{X_1 X_2 \dots X_n} = \sqrt[n]{\prod_{i=1}^n X_i}$$

$$\log G = \frac{1}{n} \sum_{i=1}^n \log x_i = \bar{X} \text{ (logaritmos)}$$

isto é, o logaritmo da média geométrica é a média aritmética dos logaritmos dos valores observados.

Para qualquer conjunto de n números não-negativos, temos $G \leq \bar{X}$, com $G = \bar{X}$ somente se os n números forem todos iguais. $\triangle$

A média harmônica de n valores diferentes de zero é o recíproco da média aritmética dos recíprocos dos valores, isto é,

$$H = \left(\frac{1}{n} \sum_{i=1}^n X_i^{-1} \right)^{-1}$$

Para qualquer conjunto de n valores positivos temos $H \leq G$, com $H = G$ somente se os n valores forem todos iguais.

Dado um conjunto de n valores observados (X_1, X_n) e conhecidos os respectivos fatores de ponderação (q_1, q_n) a média ponderada de X é:

$$W = \frac{\sum_{i=1}^n x_i q_i}{\sum_{i=1}^n q_i}$$

5

→ Caracterização Adicional

- I) O valor central entre os extremos é o valor em relação ao qual é mínimo o valor absoluto do maior desvio;
- II) A moda é o valor em relação ao qual existe a maior quantidade de desvios nulos;
- III) A média aritmética é o valor em relação ao qual é mínima a soma dos quadrados dos desvios;
- IV) A mediana é o valor em relação ao qual é mínima a soma dos valores absolutos dos desvios.

6) MEDIDAS DE DISPERSÃO

Uma possível medida de dispersão é a amplitude total, que é a diferença entre o maior e o menor valor observado. Essa não é uma boa medida de dispersão pois não leva em consideração todos os dados.

Não se pode utilizar a soma dos desvios como medida de dispersão porque essa soma é sempre zero. Ao tomar os desvios ao quadrado, pode-se considerar os desvios independente do sinal. Dividindo a soma dos desvios pelo n° de observações, obtemos a variância dos dados:

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

Se $X_1, X_2, \dots, X_n$ são os valores observados em uma amostra S^{*2} é o estimador de máxima verossimilhança da variância de X . Enquanto isso, o estimador não-tendencioso da variância de X é

$$s^2 = \frac{\sum (X_i - \bar{X})^2}{n-1}$$

A variância tem o inconveniente de apresentar unidade de medida igual ao quadrado das unidade de medida dos dados. O desvio padrão é, por definição, a raiz quadrada com sinal positivo da variância; a unidade de medida do desvio-padrão é igual à unidade de medida dos dados.

$$\sum (X_i - \bar{X})^2 = \sum X_i^2 - (1/n) (\sum X_i)^2$$

No caso de os dados estarem agrupados em uma distribuição de frequências, cada valor distinto ou valor central da classe deve ser ponderado pela respectiva frequência f_j .

O desvio médio de um conjunto de dados $(X_1, X_2, \dots, X_n)$ é a média aritmética dos valores absolutos dos desvios em relação a $\bar{X}$, isto é:

$$d = (1/n) \sum |X_i - \bar{X}|$$

A diferença média de um conjunto de dados $(X_1, X_2, \dots, X_n)$ é

$$\Delta = \frac{1}{n^2} \sum_i \sum_j |X_i - X_j|$$

7) ESPERANÇA MATEMÁTICA

7

Dizemos que a variável discreta X é uma variável aleatória se a cada um de seus possíveis valores se associa uma probabilidade $P(x)$. O conjunto dos valores da variável e das respectivas probabilidades, isto é, o conjunto dos valores de X_i e $P(X_i)$, com $i=1, \dots, n$, é a distribuição de X . $\sum P(X_i) = 1$.

Se X é uma variável aleatória discreta que assume os valores $X_1 < X_2 < \dots < X_n$, a função de distribuição de X é, por definição:

$$F(x_i) = \sum_{h=1}^i P(X_h)$$

Se a v.a. é contínua, a probabilidade de obtermos exatamente um determinado valor é zero. $P(X = X_0) = 0$.

Para caracterizar a distribuição definimos a função de densidade $f(x)$, de tal maneira que

$$P(a < X < b) = \int_a^b f(x) dx$$

isto é, a probabilidade de X assumir um valor dentro do intervalo (a, b) é numericamente igual à área compreendida entre a função de densidade e o eixo das abscissas, nesse intervalo.

Seja X uma variável discreta que assume os valores $X_1, X_2, \dots, X_n$ com probabilidades $P(X_1), P(X_2), \dots, P(X_n)$, respectivamente. Por definição, a esperança matemática de X é

$$\mu = E(x) = \sum_{i=1}^n X_i P(x_i)$$

Se X é uma variável contínua com função de densidade $f(x)$, temos

$$\mu = E(x) = \int_{-\infty}^{\infty} X f(x) dx$$

→ Distribuição Conjunta e Variáveis Independentes

Sejam X e Y variáveis aleatórias discretas que assumem os valores $X_1, X_2, \dots, X_n$ e $Y_1, Y_2, \dots, Y_k$. Seja $P(X_i, Y_j)$ a probabilidade associada ao par X_i, Y_j . Então o conjunto dos pares de valores X_i, Y_j e das respectivas probabilidades $P(X_i, Y_j)$ constitui a distribuição conjunta de X e Y .

De acordo com o Teorema da soma temos

$$P(X_i) = \sum_j P(X_i, Y_j) \quad \text{e} \quad P(Y_j) = \sum_i P(X_i, Y_j),$$
 que são as probabilidades da distribuição marginal.

Fixando $Y = Y_n$, o conjunto de valores de X_i e de

$$P(X_i | Y = Y_n) = \frac{P(X_i, Y_n)}{P(Y_n)}$$

, o que constitui a distribuição condicional de X dada $Y = Y_n$.

Se as variáveis X e Y forem independentes, a probabilidade associada ao par X_i, Y_j é igual ao produto das probabilidades marginais de X_i e de Y_j , isto é,

$$P(X_i, Y_j) = P(X_i) P(Y_j) = \sum_j P(X_i, Y_j) \sum_i P(X_i, Y_j)$$

Conclui-se que, se as variáveis são independentes, as distribuições condicionais são iguais à correspondente distribuição marginal.

→ Propriedades da Esperança Matemática

a) Se K é uma constante, $E(KX) = K \cdot E(X)$

b) Se K é uma constante, $E(X+K) = E(X) + K$

c) X e Y são v.a. quaisquer, $E(X+Y) = E(X) + E(Y)$

d) Se X e Y são v.a.s independentes, $E(XY) = E(X) \cdot E(Y)$ ⁹

→ Variância e Covariância

Por definição a Variância de uma v.a. X de população infinita é:

$$\sigma^2 = V(x) = E(X - E(x))^2 = E(X - \mu)^2$$

Se K é uma constante,

$$V(X + K) = V(x)$$

$$V(KX) = K^2 V(x)$$

Por definição, dadas duas v.a.s X e Y , a covariância entre X e Y é:

$$\text{cov}(X, Y) = E(X - E(x))(Y - E(Y)) = E(X - \mu_x)(Y - \mu_y)$$

$$V(X + Y) = V(x) + V(y) + 2 \text{cov}(X, Y)$$

$$V(X - Y) = V(x) + V(y) - 2 \text{cov}(X, Y)$$

Se X e Y são variáveis independentes, $V(X \pm Y) = V(x) + V(y)$.

OBS.: Embora $\text{cov}(X, Y) = 0$ sempre que X e Y são variáveis aleatórias independentes, nem toda situação de $\text{cov}(X, Y) = 0$ implica que X e Y são independentes.

Seja K uma constante e se X, Y e Z são variáveis aleatórias, a covariância apresenta as seguintes propriedades:

a) $\text{cov}(X + Y, Z) = \text{cov}(X, Z) + \text{cov}(Y, Z)$

b) $\text{cov}(KX, Y) = \text{cov}(X, KY) = K \text{cov}(X, Y)$

c) $\text{cov}(K, X) = \text{cov}(X, K) = 0$

*) $\text{cov}(X, \alpha + \beta X) = \beta V(x)$

→ Momentos

Dada a distribuição de uma variável aleatória X , por definição, o momento de ordem K em relação à origem é:

$$M'_K = E(X^K)$$

o o primeiro momento em relação à origem é a média de X , isto é

$$M'_1 = E(X) = \mu$$

Por definição, o momento de ordem K em relação à média de X é:

$$M_K = E(X - \mu)^K$$

(1) primeiro momento em relação à média é a esperança do desvio de X em relação à sua média, que é zero.

$$M_1 = E(X - \mu) = E(X) - \mu = 0$$

(2) segundo momento em relação à média é a variância de X .

$$M_2 = E(X - \mu)^2 = \sigma^2$$

(3) terceiro momento em relação à média está relacionado com a assimetria da distribuição.

$$\begin{cases} M_3 = 0 \Rightarrow \text{Dist. Simétrica} \\ M_3 > 0 \Rightarrow \text{Dist. Assimétrica à direita} \\ M_3 < 0 \Rightarrow \text{Dist. Assimétrica à esquerda} \end{cases}$$

(4) coeficiente de assimetria $\alpha_3 = M_3 / \sigma^3$ é medida adimensional de assimetria.

(5) quarto momento em relação à média está relacionado com a curtose da distribuição. Considerando $\alpha_4 = M_4 / \sigma^4$,

$$\begin{cases} \alpha_4 = 3 \Rightarrow \text{Mesocúrtica} = \text{Dist Normal} \\ \alpha_4 > 3 \Rightarrow \text{Leptocúrtica} \Rightarrow \text{+alta que a dist. normal} \\ \alpha_4 < 3 \Rightarrow \text{Platicúrtica} \Rightarrow \text{+achatada que a dist. normal} \end{cases}$$

Dada uma v.a. de média μ e variância σ^2 , o coeficiente de variância da distribuição é $\frac{\sigma}{\mu}$.

8) A Distribuição Binomial

Consideremos um experimento constituído por n ensaios independentes, onde cada ensaio pode resultar em um de dois eventos mutuamente exclusivos. Suponhamos que estamos interessados em um desses eventos, que denominaremos resultado favorável. Seja p a probabilidade de ocorrer o resultado favorável, em qualquer dos n ensaios; logo a probabilidade de obter um resultado desfavorável será $q = 1 - p$.

Seja X o nº de resultados favoráveis em n ensaios; então X é uma variável aleatória discreta que pode assumir valores inteiros de 0 à n . Dizemos que X tem distribuição binomial com parâmetros n e p .

É conveniente relembrar os termos do desenvolvimento do binómio de Newton:

$$(p + q)^1 = q + p$$

$$(p + q)^2 = q^2 + 2pq + p^2$$

$$(p + q)^3 = q^3 + 3pq^2 + 3p^2q + p^3$$

$$(p + q)^n = q^n + npq^{n-1} + \binom{n}{2} p^2 q^{n-2} + \dots + \binom{n}{k} p^k q^{n-k} + \dots + p^n$$

Generalizando, se X tem distribuição binomial com parâmetros p e n , as probabilidades associadas aos $n+1$ diferentes valores de X são dadas pelos termos do desenvolvimento de $(p + q)^n$.

A probabilidade de que ocorram $X = k$ resultados favoráveis em n ensaios é dada por

$$P(X = k) = \binom{n}{k} p^k q^{n-k}$$

Como X é a soma de n termos cada um com esperança p e variância pq , temos que:

$$\mu = E(X) = np$$

$$\sigma^2 = V(X) = npq$$

9) A Distribuição Normal

Consideremos três distribuições binomiais, que ocupam três diferentes posições e têm diferentes graus de dispersão. Para comparar a forma dessas distribuições, independentemente da suas posições e dispersões, fazemos uma transformação de variáveis definida

$$Z = \frac{X - \mu}{\sigma}$$

Pode-se verificar que $E(Z) = 0$ e $V(Z) = 1$. A nova variável Z , que tem média zero e variância 1, é denominada **variável reduzida**.

A distribuição binomial considerando a variável reduzida tendendo $n \rightarrow \infty$ forma a função de densidade da distribuição normal reduzida, com média zero e variância um.

A função de densidade da distribuição normal com média μ e variância σ^2 é dada por:

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left\{ -\frac{(x - \mu)^2}{2\sigma^2} \right\}, \text{ com } -\infty < x < \infty$$

Pode-se demonstrar que

$$\int_{-\infty}^{\infty} f(x) dx = 1; \quad \int_{-\infty}^{\infty} x f(x) dx = E(X) = \mu$$

$$\int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx = E(X - \mu)^2 = \sigma^2$$

→ Teorema do Limite Central

Se $X_1, X_2, \dots, X_n$ são variáveis aleatórias independentes com média μ , variância σ^2 e se μ e σ^2 são valores finitos, a distribuição de $Y = \sum X_i$ tende a uma distribuição normal com média $E(Y) = n\mu$ e variância $V(Y) = n\sigma^2$, à medida que n cresce.

A distribuição limite de $Z = \frac{\sum X_i - n\mu}{\sqrt{n} \sigma}$,

$n \rightarrow \infty$, é uma distribuição normal reduzida.

A soma de um n° bastante grande de variáveis aleatórias independentes tem distribuição aproximadamente normal, desde que nenhuma delas seja dominante.

A função de densidade da distribuição normal mostra que o valor da função é o mesmo para dois pontos equidistantes de μ ; isto é, definida um acréscimo u , se somarmos e subtrairmos esse acréscimo à média μ ,

$$f(\mu + u) = f(\mu - u)$$

Pode-se então concluir que a distribuição normal é simétrica em torno de μ , logo $\mu = \text{mediana} = \text{moda}$.

A curva normal é assintótica ao eixo das abscissas, já que $f(x) > 0$ e $\lim_{x \rightarrow \pm\infty} f(x) = 0$.

A curva normal tem dois pontos de inflexão, cujas abscissas são $X = \mu \pm \sigma$, que são os pontos em que $f''(X) = 0$.