DISTRIBUIÇÃO BINOMIAL

Uma variável aleatória Y com distribuição binomial tem a seguinte função de probabilidade :

$$f_y (y) = f_y (y ; n , p) \begin{cases} \binom{n}{y} p^y q^{n-y} & \text{para } y = 0, 1, \dots, n \\ 0 & \text{para outros valores} \end{cases}$$

onde

Y é o número de sucessos em n provas de Bernoulli,

p é a probabilidade de sucesso, $0 \leq p \leq 1$,

q é a probabilidade de insucesso, $q = 1 - p$, $0 \leq q \leq 1$

n são números inteiros positivos,

Se Y tiver uma distribuição binomial então, a seu valor esperado e a variância são respectivamente:

$$E[Y] = np$$

$$V[Y] = npq$$

A distribuição binomial, usa a média amostral para a estimação da média da população np .

A **média amostral** goza das seguintes propriedades: 1) é uma estatística suficiente para a média da população; 2) é de entre os estimadores centrados o de menor variância ou seja é o mais eficiente.

Para deduzir a distribuição binomial, que permite obter a probabilidade de y sucessos em n provas de Bernoulli, começa-se por apresentar um exemplo simples.

CASO 2: Em cinco pacientes escolhidos aleatoriamente entre os que têm enxaquecas, pretende-se calcular a **probabilidade dos três primeiros pacientes sentirem alívio após tomarem a droga A, e os dois restantes não terem alívio**, correspondendo 0,6 à probabilidade de alívio e 0,4 à probabilidade de não ter alívio, $P(A) = 0,6$ e $P(NA) = 0,4$, as quais se mantêm constantes em sucessivas repetições da mesma experiência.

Como as provas são independentes, a probabilidade conjunta dos três primeiros pacientes terem alívio e os dois seguintes não, é igual ao produto das probabilidades marginais, e é um acontecimento pouco provável (perto de zero), como se vê:

$$P(y_1=1, y_2=1, y_3=1, y_4=0, y_5=0) = P(y_1=1)P(y_2=1)P(y_3=1)P(y_4=0)P(y_5=0)$$

$$= p \times p \times p \times q \times q = p^3 q^2 = 0,6^3 0,4^2 = 0,03456$$

Contudo muitas vezes não importa uma ordem específica dos resultados mas simplesmente a probabilidade de um dado número de sucessos em n provas de Bernoulli.

Qual a probabilidade de obter três pacientes com alívio em exactamente 5 pacientes escolhidos aleatoriamente entre os que têm enxaquecas?

Há dez sequências que satisfazem esta condição:

00111; 01101; 01110; 10011; 10101; 10110; 11001; 11010; 11100; 01011

Cada sequência tem a mesma probabilidade de ocorrência: $p^3 q^2 = 0,03456$. Para se obter o número de ocorrências possíveis de 3 sucessos e 2 insucessos, recorre-se às combinações de n objectos tomados de uma só vez, ié, $\binom{5}{3} = \frac{5!}{3!(5-3)!} = 10$.

Multiplicando a probabilidade de cada resultado pelo total de ocorrências, obtém-se a probabilidade do acontecimento “três sucessos em cinco provas”, ser de 0,3456 :

$$\binom{5}{3} p^3 q^2 = 10 \times 0,03456 = 0,3456.$$

Esta probabilidade pode também ser obtida consultando as tabelas da distribuição binomial para $n = 5$, $q = 0,4$ e 2 insucessos.

Generalizando este resultado com vista a obter a probabilidade de exactamente y sucessos em n provas de Bernoulli tem-se: $\binom{n}{y} p^y q^{n-y}$, para $y = 0, 1, 2, \dots, n$; $n = 0, 1, 2, \dots$

Em síntese, nos problemas em que se verificam as suposições do processo de Bernoulli, podem obter-se as probabilidades de $0, 1, 2, 3, \dots, y$ sucessos em n provas recorrendo à distribuição binomial. Esta distribuição tem dois parâmetros n e p .

Repare-se que, a correspondência entre a situação na vida real e o da distribuição binomial deve fazer-se com cuidado, pois em cada caso, a probabilidade p deve ser mantida constante, o que implica que não se alterem as condições subjacentes ao modelo com o decorrer do tempo.

FORMATO DA DISTRIBUIÇÃO

Quanto ao **formato da distribuição** binomial observam-se três situações:

1) Quando n é pequeno e $p > 0,5$, a distribuição diz-se enviesada à direita ou assimétrica negativa.

Por exemplo, para $n = 5$ e $p = 0,6$, uma vez que as tabelas da binomial só têm probabilidades até 0,5, para calcular a probabilidade de $y = 4$ sucessos é o mesmo que calcular a probabilidade de 1 ($= n - y$) insucesso cuja probabilidade é $q = 0,4$.

Como $\binom{n}{y} = \binom{n}{n-y}$

y	$n - y$	$q = 0,4$
0	5	0,0102
1	4	0,0768
2	3	0,2304
3	2	0,3456
4	1	0,2592
5	0	0,0778

2) Quando n e p são pequenos a distribuição é enviesada à esquerda ou assimétrica positiva. Por exemplo $n = 5$, $p = 0,2$

y	$p = 0,2$
5	0,0003
4	0,0064
3	0,0512
2	0,2048
1	0,4096
0	0,3277

3) Quando $p = 0,5$ e/ou n é grande, a distribuição Binomial é simétrica. Por exemplo, $p = 0,5$; $n = 5$

y	$p = 0,5$
0	0,0313
1	0,1563
2	0,3125
3	0,3125
4	0,1563
5	0,0313

Mesmo quando $p \neq 0,5$, a forma da distribuição torna-se cada vez mais simétrica à medida que n assume valores mais elevados.

VALOR ESPERADO E VARIÂNCIA

Corolário:

Se cada variável aleatória X_i , com $i = 1, 2, \dots, n$, são independentes e se cada X_i tiver uma distribuição de Bernoulli com parâmetros $(1, p)$, então a sua soma

$Y = \sum_{i=1}^n X_i$ tem uma distribuição Binomial com parâmetros (n, p) :

$$X_i \cap B(1, p), \text{ então } Y = \sum X_i \cap B(n, p)$$

O valor esperado e variância da distribuição Binomial são respectivamente:

$$1) E[Y] = np$$

$$2) V[Y] = npq$$

Demonstração:

1) $E[Y] = E\left[\sum_{i=1}^n X_i\right]$ Desenvolvendo o somatório tem-se:

$E[Y] = E[X_1 + X_2 + \dots + X_n]$ Como o valor esperado de uma soma é a soma dos valores esperados, vem:

$$E[Y] = E(X_1) + E(X_2) + \dots + E(X_n) =$$

Sabendo que cada X_i tem uma distribuição de Bernoulli, cujo valor esperado é $E(X_i) = p$, então:

$$E[Y] = \underbrace{p + p + \dots + p}_n = np \text{ c.q.d.}$$

2) $V[Y] = V\left[\sum_{i=1}^n X_i\right]$ Desenvolvendo o somatório tem-se:

$V[Y] = V[X_1 + X_2 + \dots + X_n]$, como as variáveis aleatórias são independentes, a variância de uma soma é igual à soma das variâncias, vem:

$$V[Y] = V(X_1) + V(X_2) + \dots + V(X_n) =$$

Sabendo que cada X_i tem uma distribuição de Bernoulli, cuja variância é $V(X_i) = pq$, então:

$$V[Y] = \underbrace{pq + pq + \dots + pq}_n = npq \text{ c.q.d.}$$

Repare-se que a distribuição binomial se reduz à distribuição de Bernoulli quando $n = 1$. Algumas vezes a distribuição de Bernoulli chama-se o ponto binomial.

TEOREMA DA ADITIVIDADE DA BINOMIAL

Se as variáveis aleatórias Y_i , com $i = 1, 2, \dots, k$ são independentes e se cada Y_i tiver distribuição Binomial com parâmetros (n_i, p) , então a sua soma tem distribuição Binomial com parâmetros $(\sum n_i, p)$:

$$Y_i \cap B(n_i, p), \text{ então} \\ \sum_i Y_i \cap B(\sum n_i, p)$$

Repare-se que p é constante.

APROXIMAÇÃO DA BINOMIAL À NORMAL

Quando a dimensão da amostra n é grande e p se aproxima de $1/2$, então a distribuição binomial tende para a distribuição Normal.

Considera-se uma boa aproximação quanto mais perto p estiver de $1/2$, e $(npq) \geq 3$.

Para se aproximar uma distribuição da outra, deverão possuir características semelhantes, nomeadamente as médias e as variâncias serem iguais ou quase iguais: a média da normal (μ) deve ter o mesmo valor que a média da binomial (np), ié, $\mu = np$, e o desvio padrão da normal (σ) deve igualar o desvio padrão da binomial ($\sqrt{npq}$), ié, $\sigma = \sqrt{npq}$.

O cálculo das probabilidades carece da estandardização da variável aleatória Y :

$$Z = \frac{Y - \mu}{\sigma} = \frac{Y - np}{\sqrt{npq}}$$

A mera equivalência entre as médias e os desvios padrões das duas distribuições não assegura que os valores estandardizados Z tenham uma distribuição normal

estandardizada. Porém, pelo teorema do limite central, à medida que n aumenta, o rácio $\frac{Y - np}{\sqrt{npq}}$ tende para a normal estandardizada: $\lim_{n \rightarrow \infty} \frac{Y - np}{\sqrt{npq}} = N(0,1)$.

A distribuição binomial é discreta, envolvendo apenas valores inteiros, pelo que, sendo a um número inteiro, o cálculo das probabilidade até a e depois de a perfaz a unidade, ié, $P[Y \leq a] + P[Y \geq a + 1] = 1$.

A distribuição normal é contínua, podendo assumir qualquer valor real, pelo que a área contida abaixo da sua curva correspondente a $P[Y \leq a] + P[Y \geq a + 1]$ não soma um, pois falta-lhe a área entre a e $a + 1$.

Deste modo, antes de se estandardizar a variável discreta Y , deve fazer-se a sua correcção de continuidade. Esta correcção para $P[Y \leq a]$ é $P[Y \leq a + 0,5]$, e para $P[Y \geq (a + 1)]$ é $P[Y \geq (a + 1) - 0,5]$.

DISTRIBUIÇÃO DA MÉDIA AMOSTRAL

$\bar{X} = \frac{\sum X_i}{n}$ é a proporção de sucessos numa amostra

Quando $n > 30$, tem-se.

$\bar{X} = \frac{\sum X_i}{n} \hat{\cap} N\left(p; \sqrt{\frac{pq}{n}}\right)$, na forma estandardizada:

$$Z = \frac{\bar{X} - p}{\sqrt{\frac{pq}{n}}} \hat{\cap} N(0;1)$$

Demonstração do valor esperado e da variância:

$$E(\bar{X}) = E\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \frac{1}{n} E\left(\underbrace{\sum_{i=1}^n X_i}_{\text{Binomial}}\right) = \frac{1}{n} np = p, \text{ c.q.d}$$

$$V(\bar{X}) = V\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \frac{1}{n^2} V\left(\underbrace{\sum_{i=1}^n X_i}_{\text{Binomial}}\right) = \frac{1}{n^2} npq = \frac{pq}{n}, \text{ c.q.d}$$

Caso 3:

Sejam as variáveis aleatórias independentes X_i e Y_i , em que:

$X_1, X_2, \dots, X_9$ com distribuição Binomial $\cap b(n_i = i; p = 0,5)$

e

Y_1, Y_2 com distribuição Normal $\cap N(2;1)$

a) Deduza a distribuição amostral de $T = \sum_{i=1}^5 X_i$. Identifique os seus parâmetros e determine o valor esperado e variância.

Resolução:

$T = \sum_{i=1}^5 X_i$ Aplicando probabilidades e desenvolvendo o somatório tem-se:

$$P[T] = P[X_1 + X_2 + X_3 + X_4 + X_5],$$

Como cada X_i tem distribuição binomial dada por:

$$X_1 \cap b(n_1 = 1; 0,5)$$

$$X_2 \cap b(n_2 = 2; 0,5)$$

$$X_3 \cap b(n_3 = 3; 0,5)$$

$$X_4 \cap b(n_4 = 4; 0,5)$$

$$X_5 \cap b(n_5 = 5; 0,5)$$

Pelo teorema da aditividade da Binomial tem-se:

$$P[T] = P\left[\sum_{i=1}^5 X_i\right] \cap b(1+2+3+4+5; 0,5), \text{ ié}$$

$$P[T] = P\left[\sum_{i=1}^5 X_i\right] \cap b(15; 0,5)$$

Os parâmetros são $n = 15$ e $p = 0,5$

Numa distribuição Binomial o valor esperado é dado por:

$$E[T] = E\left[\sum_{i=1}^5 X_i\right] = np = 15 \times 0,5 = 7,5, \text{ e a variância é:}$$

$$V[T] = V\left[\sum_{i=1}^5 X_i\right] = npq = 15 \times 0,5 \times 0,5 = 3,75$$

b) Deduza a distribuição amostral de $R = \sum_{i=1}^9 X_i - \sum_{j=1}^2 Y_j$.

Resolução:

Pelo teorema da aditividade da Binomial tem-se:

$$P\left[\sum_{i=1}^9 X_i\right] \cap b(1+2+3+4+5+6+7+8+9; 0,5), \text{ ié}$$

$$P\left[\sum_{i=1}^9 X_i\right] \cap b(45; 0,5), \text{ cujos parâmetros são } n = 45 \text{ e } p = 0,5$$

Como $n > 30$ e $p = 0,5$, a distribuição Binomial tende para a normal. Assim,

$$\sum_{i=1}^{45} X_i \dot{\cap} N(np; \sqrt{npq}), \text{ concretizando:}$$

$$\sum_{i=1}^{45} X_i \dot{\cap} N(45 \times 0,5; \sqrt{45 \times 0,5 \times 0,5}), \text{ ié}$$

$$\sum_{i=1}^{45} X_i \dot{\cap} N(22,5; \sqrt{11,5})$$

Pelo teorema da aditividade da Normal, a distribuição de R é dada por:

$$R = \sum_{i=1}^9 X_i - \sum_{j=1}^2 Y_j \cap N(22,5 - 4; \sqrt{11,25 + 2}), \text{ ié}$$

$$R = \sum_{i=1}^9 X_i - \sum_{j=1}^2 Y_j \cap N(18,5; \sqrt{13,25})$$