MBA em Ciência de Dados
Aula 4 – Exemplos
Prof. Francisco Rodrigues

1 - O tempo para desenvolver um servidor web em uma empresa é descrito por uma variável aleatória X, medida em dias, com distribuição normal de média $\mu = 45$ e variância $\sigma = 400$. Calcule a probabilidade de que um novo servidor web será finalizado entre 40 e 45 dias.

RESOLUÇÃO

Temos que $X \sim N(\mu = 45, \sigma^2 = 400)$. Podemos usar a transformação

$$Z = \frac{X - \mu}{\sigma} \sim N(0,1)$$

Ou seja, Z tem distribuição normal com média igual a zero variância igual a 1.
Assim, temos:

$$P(30 \leq X \leq 40) = P\left(\frac{40-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{45-\mu}{\sigma}\right) = P\left(\frac{40-45}{20} \leq Z \leq \frac{45-45}{20}\right)$$

$$P(-0{,}75 \leq Z \leq 0) = 0{,}273.$$

Portanto, há 27,3% de chance de que o servidor seja concluído entre 40 e 45 dias.

2 - Uma população é descrita pela seguinte distribuição de probabilidades:

$$P(X = 2) = 0{,}2; P(X = 4) = 0{,}4; P(X = 6) = 0{,}4$$

Uma amostra com 50 observações é sorteada. Calcule a probabilidade de que a média dessa amostra seja maior do que 5.

RESOLUÇÃO

Precisamos calcular o desvio padrão da amostra. Primeiramente, calculamos o valor esperado:

$$E[X] = 2 \times 0{,}2 + 4 \times 0{,}4 + 6 \times 0{,}4 = \ 4{,}4$$

E o desvio padrão:

$$\sigma^2 = E[X^2] - E[X]^2 = 2^2 \times 0{,}4 + 4^2 \times 0{,}6 + 6^2 0{,}4 - 4{,}4^2 = 2{,}24$$

$$\sigma = \sqrt{2{,}24} = 1{,}49$$

Usando o Teorema Central do Limite, vamos calcular a probabilidade pedida:

$$P(\bar{X} > 5) = P\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} > \frac{5 - \mu}{\frac{\sigma}{\sqrt{n}}}\right) = P\left(Z > \frac{5 - 4{,}4}{\frac{1{,}49}{\sqrt{50}}}\right) = P(Z > \ 0.31) = 0.378$$

3 - Em uma empresa de venda de softwares, a duração de conversas telefônicas (em minutos) segue o modelo exponencial com parâmetro $\lambda = 1/5$. Observando-se uma

amostra aleatória de 50 dessas chamadas, qual será a probabilidade de que tais amostras em média não ultrapassem 6 minutos?

RESOLUÇÃO

Seja X a variável aleatória que representa o tempo das chamadas. Temos que X segue o modelo exponencial:

$$f(x) = \begin{cases} \lambda e^{-\lambda} \; se \; x \geq 0 \\ 0 \; se \; x < 0 \end{cases}$$

Sendo a esperança e variância:

$$E[X] = \frac{1}{\lambda} = \frac{1}{\frac{1}{5}} = 5$$

$$\sigma^2 = \frac{1}{\lambda^2} = \frac{1}{\frac{1}{25}} = 25$$

Queremos calcular:

$$P(\bar{X} < 6)$$

Usando o Teorema Central do Limite:

$$P(\bar{X} < 6) = P\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < \frac{6 - \mu}{\frac{\sigma}{\sqrt{n}}}\right) = P\left(Z < \frac{6-5}{\frac{5}{\sqrt{50}}}\right) = P(Z < 1{,}41) = \; 0{,}92$$

Portanto, há uma probabilidade de 92% de que as chamadas não ultrapassem 6 minutos.

4 - Qual deve ser o tamanho de uma amostra a ser retirada de uma população $X \sim N(\mu = 100, \sigma = 50)$ para que $P(-90 < \bar{X} < 110) = 0{,}95$?

RESOLUÇÃO

Podemos usar o Teorema Central do Limite:

$$P(-90 < \bar{X} < 110) = 0{,}95$$

$$P\left(\frac{-90 - 100}{\frac{50}{\sqrt{n}}} < Z < \frac{110 - 100}{\frac{50}{\sqrt{n}}}\right) = 0{,}95$$

Pela tabela normal padronizada, para $P(-z_\alpha < Z < z_\alpha) = 0{,}95$, temos que $z_\alpha = 1{,}96$. Assim,

$$\frac{110 - 100}{\frac{50}{\sqrt{n}}} = 1{,}96$$

Resolvendo:

$$n = (1{,}96 \times 5)^2 = 96{,}0$$

Portanto, o tamanho da amostra deve ser igual a 96.

5 - Seja $[X_1, X_2, \dots, X_n]$ uma amostra aleatória de uma população com média μ e variância σ^2. Mostre que para a estatística $Z = (\bar{X} - \mu)/(\sigma/\sqrt{n})$, temos $E[Z] = 0$ e $V(Z) = 1$.

RESOLUÇÃO

Temos:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

Aplicando a esperança em ambos os lados:

$$E[Z] = E\left[\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}\right]$$

Mas, a esperança é uma função linear, onde $E[aX + b] = aE[X] + b$, para a e b constantes.

$$E[Z] = \frac{1}{\frac{\sigma}{\sqrt{n}}}(E[\bar{X}] - \mu)$$

Mas, como $\bar{X}$ é um estimador não-viesado, temos $E[\bar{X}] = \mu$ e, portanto,

$$E[Z] = 0.$$

No caso da variância, temos a seguinte propriedade:

$$V(aX + Y + c) = a^2V(X) + V(Y)$$

Se Xe Y são independentes e a e c são constantes.

Então:

$$V(Z) = V\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}\right) = \frac{n}{\sigma^2}V(\bar{X}).$$

Mas, $V(\bar{X}) = \ \sigma^2/n$. Portanto,

$$V(Z) = V\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}\right) = \frac{n}{\sigma^2}\frac{\sigma^2}{n} = 1$$

Portanto, $Z \sim N(\mu = 0, \sigma = 1)$.

6 - Mostre que se $(X_1, X_2, \dots, X_n)$ (todas independentes) é uma variável aleatória de uma população com média μ e variância σ^2, então $E[\bar{X}] = \mu$ e $V(\bar{X}) = \sigma^2/n$.

RESOLUÇÃO

Temos:

$$E[\bar{X}] = E\left[\sum_{i=1}^{n} \frac{X_i}{n}\right] = \frac{1}{n}\left[E[X_1] + E[X_2] + \dots + E[X_n]\right]$$

Mas, temos que $E[X_i] = \mu$ para $i = 1,2, \dots, n$.

Então

$$E[\bar{X}] = E\left[\sum_{i=1}^{n} \frac{X_i}{n}\right] = \frac{1}{n}\big[E[X_1] + E[X_2] + \cdots + E[X_n]\big] = \frac{n\mu}{n} = \mu.$$

Portanto,

$$E[\bar{X}] = \mu.$$

Para a variância:

$$V(\bar{X}) = V\left(\sum_{i=1}^{n} \frac{X_i}{n}\right) = \frac{1}{n^2}\big(V(X_1) + V(X_2) + \cdots + V(X_n)\big)$$

Mas, temos que $V(X_i) = \sigma^2$, pois cada elemento da amostra tem a mesma distribuição que a população. Assim,

$$V(\bar{X}) = \frac{1}{n^2}\big(V(X_1) + V(X_2) + \cdots + V(X_n)\big) = \frac{1}{n^2}(\sigma + \sigma + \cdots + \sigma) = \frac{1}{n^2}(n\sigma) = \frac{\sigma}{n}$$

Portanto,

$$V(\bar{X}) = \frac{\sigma}{n}.$$