32. Distribuição amostral da diferença de duas médias.

Consideremos duas populações X com parâmetros μ_1 e σ_1^2 e Y com parâmetros μ_2 e σ_2^2. Sorteiam-se duas amostras independentes: a da primeira população de tamanho n e a da segunda de tamanho m . Calculam-se as médias amostrais $\bar{X}$ e $\bar{Y}$.

(a) Qual a distribuição amostral de $\bar{X}$? E de $\bar{Y}$?

(b) Defina $D = \bar{X} - \bar{Y}$. O que você entende por distribuição amostral de D?

(c) Calcule $E(D)$ e $Var(D)$.

(d) Como você acha que será a distribuição de D? Por quê?

Solução: Sabemos que

$$E\left(\bar{X}\right) = \mu_1 \quad e \quad Var\left(\bar{X}\right) = \frac{\sigma_1^2}{n}.$$

Se $X \sim N(\mu_1, \sigma_1^2)$ então:

$$\bar{X} \sim N\left(\mu_1, \frac{\sigma_1^2}{n}\right).$$

Se X não tiver distribuição normal podemos usar o teorema central do limite que garante que $\bar{X}$ tenha uma distribuição aproximadamente normal para n grande.

Sabemos que

$$E\left(\bar{Y}\right) = \mu_2 \quad e \quad Var\left(\bar{Y}\right) = \frac{\sigma_2^2}{m}.$$

Se $Y \sim N(\mu_2, \sigma_2^2)$ então:

$$\bar{Y} \sim N\left(\mu_2, \frac{\sigma_2^2}{m}\right).$$

Se Y não tiver distribuição normal podemos usar o teorema central do limite que garante que $\bar{Y}$ tenha uma distribuição aproximadamente normal.

Vamos responder ao item **b**:

Resposta pessoal. Pesquise no livro e responda com suas palavras.

Vamos responder ao item **c**:

$$\mu_D = E(D) = E\left(\bar{X} - \bar{Y}\right) = E\left(\bar{X}\right) - E\left(\bar{Y}\right) = \mu_1 - \mu_2.$$

Temos que $\bar{X}$ e $\bar{Y}$ são independentes. Assim,

$$\sigma_D^2 = Var(D) = Var\left(\bar{X} - \bar{Y}\right) = Var\left(\bar{X}\right) + Var\left(\bar{Y}\right) = \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}.$$

Vamos responder ao item **d**:

Vamos supor normalidade das variáveis X e Y.

Então:

$$D = \bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}\right)$$

Padronizando temos:

$$Z = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \sim N(0,1),$$

aproximadamente.

Mesmo as populações não sendo normais podemos aplicar ao teorema do limite central.

35. Distribuição amostral da diferença de duas proporções. Usando os resultados do Problema 32, qual seria a distribuição de $\hat{p}_1 - \hat{p}_2$, a diferença entre as proporções de amostras independentes retiradas de populações com parâmetros p_1 e p_2?

Solução:

Vamos supor que

$$X \sim Ber(p_1)\ ;\ \mu_1 = p_1 \quad , \sigma_1^2 = p_1\, q_1.$$

$$Y \sim Ber(p_2)\ ;\ \mu_2 = p_2 \quad , \sigma_2^2 = p_2\, q_2.$$

Seja $X_1, X_2, \ldots, X_n$ a amostra aleatória de tamanho n de X:

Então,

$$\hat{p}_1 = \frac{\sum_{i=1}^{n} X_i}{n} = \bar{X}, \qquad \hat{p}_2 = \frac{\sum_{j=1}^{m} Y_j}{m} = \bar{Y}.$$

Note que:

$$S_1 = \sum_{i=1}^{n} X_i \sim Bin(n, p_1) \quad e \quad S_2 = \sum_{j=1}^{m} Y_j \sim Bin(m, p_2).$$

Além disso podemos aproximar a distribuição binomial para Normal.

Assim

$$E\left(\hat{p}_1 - \hat{p}_2\right) = E\left(\hat{p}_1\right) - E\left(\hat{p}_2\right) = p_1 - p_2.$$

$$Var\left(\hat{p}_1 - \hat{p}_2\right) = Var\left(\hat{p}_1\right) + Var\left(\hat{p}_2\right) = \frac{p_1 q_1}{n} + \frac{p_2 q_2}{m}.$$

E usando o teorema do limite central temos que $\hat{p}_1 - \hat{p}_2$ segue uma distribuição normal .
Padronizando temos:

$$Z = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n} + \frac{p_2\, q_2}{m}}} \sim N(0,1),$$

aproximadamente.