Comparação de Duas Populações: Amostras dependentes

Duas amostras dependentes, $X_1, X_2, ..., X_n$ e $Y_1, Y_2, ..., Y_n$, as observações são pareadas, podemos considerar que temos na realidade uma amostra de pares $(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)$.

Se definirmos a v.a. $D = X - Y$, teremos a amostra $D_1, D_2, ..., D_n$, resultante das diferenças entre os valores de cada par.

População Normal

Nessa situação, faremos a seguinte suposição: a v.a. D tem distribuição normal $N(\mu_D, \sigma^2_D)$. Podemos deduzir daqui que

$$\bar{D} = \frac{1}{n}\sum_{i=1}^{n} D_i = \frac{1}{n}\sum_{i=1}^{n}(X_i - Y_i) = \bar{X} - \bar{Y} \qquad (13.27)$$

terá distribuição $N(\mu_D, \ \sigma^2_D/n)$.

Considere

$$S^2_D = \frac{1}{n-1}\sum_{i=1}^{n}(D_i - \bar{D})^2. \qquad (13.28)$$

Pelo Teorema 7.1, a estatística

$$T = \frac{\sqrt{n}(\bar{D} - \mu_D)}{S_D} \qquad (13.29)$$

terá distribuição t de Student, com $(n - 1)$ graus de liberdade.

Exemplo 13.10. Cinco operadores de certo tipo de máquina são treinados em máquinas de duas marcas diferentes, A e B. Mediu-se o tempo que cada um deles gasta na realização de uma mesma tarefa, e os resultados estão na Tabela 13.8.

Tabela 13.8: Tempos para realização de tarefa para cinco operadores.

Operador	Marca A	Marca B
1	80	75
2	72	70
3	65	60
4	78	72
5	85	78

Com o nível de significância de 10%, poderíamos afirmar que a tarefa realizada na máquina A demora mais do que na máquina B?

13.5 Comparação de Proporções em Duas Populações

Nosso objetivo agora é a comparação das proporções p_1 e p_2 de indivíduos de duas populações P_1 e P_2, respectivamente, que tenham um mesmo atributo. Para isso, extraímos duas amostras independentes dessas populações, com tamanhos n_1 e n_2, respectivamente, e obtemos os estimadores usuais $\hat{p}_1$ e $\hat{p}_2$. Das seções 10.9 e 12.6 temos que

$$\hat{p}_1 \sim N\left(p_1\,,\frac{p_1(1-p_1)}{n_1}\right),\hat{p}_2 \sim N\left(p_2\,,\frac{p_2(1-p_2)}{n_2}\right).$$

Usando os resultados da seção 13.3.1 e Problema 10.32, obtemos

$$\hat{p}_1-\hat{p}_2 \sim N\left(p_1-p_2,\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}\right),$$

e portanto,

$$z=\frac{(\hat{p}_1-\hat{p}_2)-(p_1-p_2)}{\sqrt{\dfrac{p_1(1-p_1)}{n_1}+\dfrac{p_2(1-p_2)}{n_2}}} \sim N(0\,,1).$$

Pode-se provar que, substituindo p_1 e p_2 por seus estimadores, obtemos que

$$\hat{z}=\frac{(\hat{p}_1-\hat{p}_2)-(p_1-p_2)}{\sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\dfrac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} \sim N(0\,,1). \tag{13.35}$$

Suponha agora que queiramos testar as hipóteses

$$H_0 : p_1 = p_2,$$

$$H_1 : p_1 \neq p_2.$$

Usando os mesmos argumentos apresentados na seção 13.3.1(a), deve-se usar um estimador comum de $p_1 = p_2$, a saber

$$\hat{p}_c = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2},$$

e de (13.35) obtemos, sob H_0,

$$\hat{z} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_c(1 - \hat{p}_c)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0, 1). \tag{13.36}$$

Exemplo 13.12: Para lançamento da nova embalagem do sabonete X a divisão de criação estuda duas propostas, A e B. Em cada um de dois supermercados similares, foram colocados sabonetes com cada tipo de embalagem, e a clientes selecionados aleatoriamente foi perguntado se tinham notado o sabonete e que descrevessem o tipo de embalagem. Abaixo estão os resultados:

Proposta	Notaram?		Total
	Sim	Não	
A	168	232	400
B	180	420	600
Total	348	652	1000