Comparação de Duas Populações: Amostras Independentes

duas amostras independentes, $X_1, X_2, ..., X_n$ e $Y_1, Y_2, ..., Y_m$, de duas populações Normais P1 e P2, respectivamente.

Aqui, $P_1 \sim N(\mu_1, \sigma_1^2)$ e $P_2 \sim N(\mu_2, \sigma_2^2)$.

Queremos testar a hipótese (13.1), que aqui fica escrita na forma

$$H_0 : \mu_1 = \mu_2.$$

Na situação da Figura 13.2 (c), a alternativa adequada é

$$H_1 : \mu_2 > \mu_1,$$

mas supondo as variâncias iguais. Se estivermos apenas interessados em verificar se existe diferença entre as médias das duas populações, não importando a direção, então a alternativa adequada será

$$H_1 : \mu_1 \neq \mu_2.$$

Para cada amostra calculamos os estimadores da média e da variância:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i, \quad S_1^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2;$$

$$\bar{Y} = \frac{1}{m} \sum_{i=1}^{m} Y_i, \quad S_2^2 = \frac{1}{m-1} \sum_{i=1}^{m} (Y_i - \bar{Y})^2.$$

Sob a hipótese H_0, isto é, $\mu_1 = \mu_2$,

$$E(\bar{X} - \bar{Y}) = 0, \tag{13.6}$$

$$\mathrm{Var}(\bar{X} - \bar{Y}) = \mathrm{Var}(\bar{X}) + \mathrm{Var}(\bar{Y}) = \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}. \tag{13.7}$$

Como $\bar{X} - \bar{Y}$ tem distribuição normal, se as variâncias fossem conhecidas, a estatística

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\sigma_1^2/n + \sigma_2^2/m}} \tag{13.8}$$

teria distribuição normal padrão, *sob a hipótese nula* H_0, e poderia ser usada para testar H_0 contra H_1. Contudo, nas situações de interesse prático, as variâncias não são conhecidas, devendo ser substituídas por estimativas convenientes.

(a) Mesma Variância, Desconhecida

Suponha que, ao testar a hipótese de igualdade de variâncias, esta não seja rejeitada, isto é, $\sigma_1^2 = \sigma_2^2 = \sigma^2$, porém essa variância comum é desconhecida. Como S_1^2 e S_2^2 são dois estimadores não-viesados de σ^2, podemos combiná-los para obter um estimador comum

$$S_p^2 = \frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2} = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2 + \sum_{i=1}^{m} (Y_i - \bar{Y})^2}{n+m-2}, \tag{13.10}$$

que também é um estimador não-viesado de σ^2.

Pelo Teorema 7.1, a estatística

$$T = \frac{\dfrac{\bar{X} - \bar{Y}}{\sigma\sqrt{1/n + 1/m}}}{S_p/\sigma} = \frac{\bar{X} - \bar{Y}}{S_p\sqrt{1/n + 1/m}} \tag{13.12}$$

terá uma distribuição t de Student, com $(n + m - 2)$ graus de liberdade, *sob a hipótese* H_0, isto é, se $\mu_1 = \mu_2$.

Exemplo 13.4. Duas técnicas de venda são aplicadas por dois grupos de vendedores: a técnica A, por 12 vendedores, e a técnica B, por 15 vendedores. Espera-se que a técnica B produza melhores resultados. No final de um mês, obtiveram-se os resultados da Tabela 13.1.

Tabela 13.1: Dados para duas técnicas de vendas.

Dados	Vendas	
	Técnica A	Técnica B
Média	68	76
Variância	50	52
Vendedores	12	15

Vamos testar, para o nível de significância de 5%, se há diferenças significativas entre as vendas resultantes das duas técnicas. Supondo que as vendas sejam normalmente distribuídas e usando o teste da seção 13.2, vemos que $\sigma^2_A = \sigma^2_B$.

As hipóteses a serem testadas ficam

$$H_0: \mu_A = \mu_B$$

$$H_1: \mu_A < \mu_B.$$

Pelas suposições acima, podemos usar a estatística (13.12), com $n = 12$, $m = 15$ e $S_p^2 = (11S_A^2 + 14S_B^2)/25$. Da Tabela V obtemos RC =]1,708, + ∞[.

Da Tabela 13.1 calculamos

$$s_p^2 = \frac{(11)(50) + (14)(52)}{25} = 51,12,$$

$$t_0 = \frac{76 - 68}{(7,15)\sqrt{\frac{1}{12} + \frac{1}{15}}} = 2,89.$$

Como $t_0 \in$ RC, rejeitamos H_0, ou seja, existe evidência de que a técnica B produz melhores resultados do que a técnica A.

Encontrada diferença entre os métodos, a continuação natural é construir um intervalo de confiança para a diferença $\Delta = \mu_B - \mu_A$. Do resultado (13.12) é fácil verificar que

$$IC(\Delta;\gamma) = (\bar{x}_0 - \bar{y}_0) \pm t_\gamma s_p \sqrt{1/n + 1/m}.$$

Para o nosso exemplo, com $\gamma = 0{,}95$, esse intervalo reduz-se a

$$\begin{aligned} IC(\Delta; 0{,}95) &= 8 \pm (2{,}06)(7{,}15)\sqrt{\frac{1}{12}+\frac{1}{15}} \\ &= 8 \pm 5{,}7 =]2{,}3;\ 13{,}7[. \end{aligned}$$

(b) Variâncias Desiguais, Desconhecidas

Quando a hipótese de igualdade de variâncias for rejeitada, devemos usar a estatística

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{S_1^2/n + S_2^2/m}}. \tag{13.13}$$

Pode-se provar que, sob a veracidade de H_0, a v.a. T aproxima-se de uma distribuição t de Student, com o número de graus de liberdade dado aproximadamente por

$$\nu = \frac{(A+B)^2}{A^2/(n-1) + B^2/(m-1)}, \tag{13.14}$$

na qual

$$A = s_1^2/n, \quad B = s_2^2/m.$$

Como esse valor é geralmente fracionário, arredonde para o inteiro mais próximo para obter o número de graus de liberdade.

Exemplo 13.5. Queremos testar as resistências de dois tipos de vigas de aço, A e B. Tomando-se $n = 15$ vigas do tipo A e $m = 20$ vigas do tipo B, obtemos os valores na Tabela 13.2. Usando um teste F com nível $\alpha = 10\%$ rejeitamos a hipótese de variâncias iguais.

Tabela 13.2: Médias e variâncias para dois tipos de vigas de aço.

Tipo	Média	Variância
A	70,5	81,6
B	84,3	210,8

Consideremos as hipóteses

$$H_0: \mu_A = \mu_B$$
$$H_1: \mu_A \neq \mu_B.$$

A estatística a ser usada é (13.13), com $\nu = (261{,}15)/(2{,}26 + 5{,}85) = 32{,}2$, logo tomamos $\nu = 33$. Com $\alpha = 0{,}05$, obtemos da Tabela V que RA = $]-2{,}0345; 2{,}0345[$. Com os dados da Tabela 13.2, temos $t_0 = (-13{,}8)/(4{,}02) = -3{,}43$.

Como $t_0 \in$ RC, rejeitamos H_0, ou seja, há evidências de que os dois tipos de vigas têm resistências médias diferentes.