



Testes de Hipóteses para Diferença entre duas Médias - Amostras independentes: variância conhecida e desconhecida

Universidade Estadual de Santa Cruz

Ivan Bezerra Allaman

Cronograma

1. Introdução
2. Considerando variância conhecida
 1. Estatística de teste
 2. Exemplos
 3. Aplicações
3. Considerando variância desconhecida
 1. Estatística de teste
 2. Exemplos
 3. Aplicações

Introdução

- Nas aulas anteriores o objetivo foi testar uma hipótese a respeito de um parâmetro de apenas uma população.
- O interesse agora é comparar duas populações independentes quanto a um determinado parâmetro.
- Tais comparações se dá por meio da diferença entre médias ou proporções destas populações com relação a um determinado valor hipotético.
- Tecnicamente poderemos testar as seguintes hipóteses:

Hipótese bilateral

$$\begin{array}{ll} H_0: \mu_1 - \mu_2 = \mu_0 & H_0: \pi_1 - \pi_2 = \pi_0 \\ H_a: \mu_1 - \mu_2 \neq \mu_0 & H_a: \pi_1 - \pi_2 \neq \pi_0 \end{array}$$

Hipótese unilateral à direita

$$\begin{array}{ll} H_0: \mu_1 - \mu_2 = \mu_0 & H_0: \pi_1 - \pi_2 = \pi_0 \\ H_a: \mu_1 - \mu_2 > \mu_0 & H_a: \pi_1 - \pi_2 > \pi_0 \end{array}$$

Hipótese unilateral à esquerda

$$\begin{array}{ll} H_0: \mu_1 - \mu_2 = \mu_0 & H_0: \pi_1 - \pi_2 = \pi_0 \\ H_a: \mu_1 - \mu_2 < \mu_0 & H_a: \pi_1 - \pi_2 < \pi_0 \end{array}$$

- Em muitas aplicações $\mu_0 = 0$ ou $\pi_0 = 0$, ou seja, na maioria dos casos o interesse é comparar se duas populações são iguais ou não.

Considerando variância conhecida

Estatística de teste

- O estimador pontual da diferença entre duas médias populacionais $\mu_1 - \mu_2$ é $\bar{X}_1 - \bar{X}_2$.
- O erro padrão da diferença entre as médias populacionais é:

$$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

- Portanto, a estatística de teste para comparar a diferença entre duas médias populacionais é:

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - \mu_0}{\sigma_{\bar{X}_1 - \bar{X}_2}}$$

Aplicação

1. Considere os seguintes resultados, referentes a duas amostras aleatórias independentes tomadas de duas populações:

Amostra 1	Amostra 2
$n_1 = 50$	$n_2 = 35$
$\bar{X}_1 = 13,6$	$\bar{X}_2 = 11,6$
$\sigma_1 = 2,2$	$\sigma_2 = 3$

Teste a hipótese de que a diferença entre as médias das duas populações é maior do que zero considerando $\alpha = 0,05$.

Elaborando as Hipóteses tem-se:

$$H_0 : \mu_1 - \mu_2 \leq 0$$

$$H_1 : \mu_1 - \mu_2 > 0$$

Calculando as estatísticas tem-se:

$$\begin{aligned}\bar{X}_1 - \bar{X}_2 &= 13,6 - 11,6 \\ &= 2\end{aligned}$$

$$\begin{aligned}\sigma_{\bar{X}_1 - \bar{X}_2} &= \sqrt{\frac{2,2^2}{50} + \frac{3^2}{35}} \\ &= 0,595\end{aligned}$$

$$z_{calc} = \frac{2 - 0}{0,595}$$

$$= 3,362$$

$$p - valor = 0,000387$$

Resp: Como o *p-valor* é menor do que α , rejeita-se H_0 com 95% de confiança.

Considerando variância
desconhecida

Introdução

- Quando a variância for desconhecida, mais uma vez iremos nos recorrer a distribuição *t-Student*.
- Quando as variâncias **são significativamente iguais** (ver teste de hipóteses para duas variâncias), precisamos calcular primeiro uma variância combinada (pooled) da seguinte forma:

$$S_c = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

- Logo, o erro padrão da diferença torna-se:

$$S_{\bar{X}_1 - \bar{X}_2} = S_c \cdot \sqrt{\frac{n_1 + n_2}{n_1 \cdot n_2}}$$

- Quando as variâncias são **significativamente diferentes**, então podemos calcular diretamente o erro padrão da diferença como:

$$S_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

- Logo, a estatística de teste é dada pela seguinte expressão:

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - \mu_0}{S_{\bar{X}_1 - \bar{X}_2}}$$

Graus de liberdade

- Quando as variâncias são significativamente iguais, então os graus de liberdade é calculado como:

$$\phi = n_1 + n_2 - 2$$

- Quando as variâncias são significativamente diferentes, então o cálculo do graus de liberdade é dada pela expressão de Satterthwaite:

$$\phi = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \right)^2}{\frac{1}{n_1-1} \left(\frac{S_1^2}{n_1} \right)^2 + \frac{1}{n_2-1} \left(\frac{S_2^2}{n_2} \right)^2}$$

Aplicação

2. O College Board divulgou comparações sobre as pontuações no Scholastic Aptitude Test (SAT) baseando-se no nível educacional mais elevado obtido pelos pais da pessoa que faz os exames. Uma das hipóteses de pesquisa era que os estudantes cujos pais haviam obtido um nível mais elevado de educação obteriam uma pontuação média mais elevada no SAT. Durante 2003, a média global dos exames orais do SAT foi 507. As pontuações nos exames orais do SAT para amostras independentes de estudantes estão nos dados a seguir. A primeira amostra exibe pontuações nos exames orais do SAT correspondentes a estudantes cujos pais têm diplomas universitários com graus de bacharel. A segunda amostra exibe as pontuações nos exames orais do SAT de estudantes cujos pais têm diplomas do segundo grau, mais não tem diplomas universitários. Qua é a conclusão? Os dados estão disponíveis no seguinte link: http://nbcgib.uesc.br/lec/download/R/dados/sat_anderson_etal.txt

Considerando μ_{com} alunos cujos os pais tem diploma e μ_{sem} alunos cujos os pais não tem diploma, tem-se as seguintes hipóteses:

$$H_0 : \mu_{com} - \mu_{sem} \leq 0$$

$$H_1 : \mu_{com} - \mu_{sem} > 0$$

Vamos checar primeiro se as variancias são significativamente iguais. Aplicando o teste F para a razão entre duas variancias conforme já visto temos:

$$F = \frac{S_1^2}{S_2^2} = \frac{3530,8}{2677,8} = 1,318$$

$$p - \text{valor} = 0,3266$$

Portanto, podemos considerar as variâncias iguais. Calculando as estatísticas tem-se:

$$\begin{aligned}\bar{X}_{com} - \bar{X}_{sem} &= 525 - 487 \\ &= 38\end{aligned}$$

$$\begin{aligned}S_c &= \sqrt{\frac{(16 - 1) \cdot 3530,8 + (12 - 1) \cdot 2677,8}{16 + 12 - 2}} \\ &= 56,302\end{aligned}$$

$$\begin{aligned}S_{\bar{X}_{com} - \bar{X}_{sem}} &= 56,302 \cdot \sqrt{\frac{16 + 12}{16 \cdot 12}} \\ &= 21,5\end{aligned}$$

$$gl = 16 + 12 - 2 = 26$$

$$\begin{aligned}t_{calc} &= \frac{38 - 0}{21,5} \\ &= 1,77\end{aligned}$$

$$p - \text{valor} = 0,0444$$

Resp: Como o *p-valor* é menor do que α , rejeita-se H_0 com 95% de confiança.

3. The following data show the liver weights (kg) taken from randomly selected cattle in two farms in southwest England during outbreaks of liver fluke disease. As a preliminary to testing the null hypothesis that the mean liver weights of the cattle in the two farms are the same, check that the variability of the observations in the two groups is similar.

g1	18.0	18.5	18.9	18.2	17.9	15.9	16.8	18.2	17.3	17.5	17.7	17.8	17.1	17	16.3
g2	14.3	13.2	17.3	14.9	16.4	16.0	18.6	17.3	15.5	16.8	15.7	18.0	15.2	NA	NA

Testando primeiro se as variancias são iguais.

$$F = \frac{S_{g2}^2}{S_{g1}^2} = \frac{2,31}{0,671} = 3,45$$

$$p - \text{valor} = 0,019$$

Eu particularmente consideraria um p-valor abaixo de 0,01 como significativo. No entanto, para fins didáticos, vamos rejeitar a hipótese de igualdade de variancias, e utilizar a segunda forma da estatística t de student.

Calculando as estatísticas tem-se:

$$\begin{aligned}\bar{X}_{g1} - \bar{X}_{g2} &= 17,54 - 16,09 \\ &= 1,45\end{aligned}$$

$$\begin{aligned}S_{\bar{X}_{g1} - \bar{X}_{g2}} &= \sqrt{\frac{0,67}{15} + \frac{2,31}{13}} \\ &= 0,47\end{aligned}$$

$$gl_{satterthwaite} = 17,8$$

$$\begin{aligned}t_{calc} &= \frac{1,45 - 0}{0,47} \\ &= 3,08\end{aligned}$$

$$p - valor = 0,0067$$

Portanto, considerando alfa igual a 1% e como o p-valor é menor que alfa, rejeitamos H_0 .