Introdução à Probabilidade e à Estatística II

Teste de Hipóteses para a variância populacional
Teste de Hipóteses assintótico para a proporção populacional

Lígia Henriques-Rodrigues

MAE0229 – 1º semestre 2018

3. Teste para a variância de uma população

Seja X uma variável aleatória com distribuição normal de média μ e variância σ^2 e suponhamos que σ_0^2 é um valor conhecido.

As hipóteses de interesse são:

Passo 1:

- (i) $H_0 : \sigma^2 \leq \sigma_0^2$ contra $H_1 : \sigma^2 > \sigma_0^2$;
- (ii) $H_0 : \sigma^2 \geq \sigma_0^2$ contra $H_1 : \sigma^2 < \sigma_0^2$;
- (iii) $H_0 : \sigma^2 = \sigma_0^2$ contra $H_1 : \sigma^2 \neq \sigma_0^2$.

Passo 2: A estatística a ser usada é

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1).$$

Nos casos (i) e (ii), a hipótese H_0 é composta, pelo que escrevemos

$$\frac{(n-1)S^2}{\sigma^2} \underset{\sigma^2=\sigma_0^2}{\sim} \chi^2(n-1) \iff \frac{(n-1)S^2}{\sigma_0^2} \underset{\sigma^2=\sigma_0^2}{\sim} \chi^2(n-1).$$

No caso (iii), a hipótese H_0 é simples, pelo que escrevemos

$$\frac{(n-1)S^2}{\sigma^2} \underset{\sigma^2=\sigma_0^2}{\sim} \chi^2(n-1) \iff \frac{(n-1)S^2}{\sigma_0^2} \underset{H_0}{\sim} \chi^2(n-1).$$

Passo 3: Determinação da região crítica

(i) $H_0 : \sigma^2 \leq \sigma_0^2$ contra $H_1 : \sigma^2 > \sigma_0^2$

Neste caso, e sendo x_c o ponto crítico, a RC é

$$RC = \{S^2 : S^2 \geq x_c\}.$$

O ponto crítico x_c será encontrado fixando a probabilidade máxima de cometer o erro tipo I, α_{max}, ou seja, fixando o nível de significância.

$$\alpha_{max} = P_{\sigma_0^2}(S^2 \geq x_c) = P\left(\frac{(n-1)S^2}{\sigma_0^2} \geq \frac{(n-1)x_c}{\sigma_0^2}\right)$$

Definindo

$$\chi_2^2 = \frac{(n-1)x_c}{\sigma_0^2} \iff x_c = \frac{\sigma_0^2\ \chi_2^2}{n-1}.$$

Logo a região crítica associada ao teste das hipóteses $H_0 : \sigma^2 \leq \sigma_0^2$ contra $H_1 : \sigma^2 > \sigma_0^2$ é

$$RC = \left\{S^2 : S^2 \geq \frac{\sigma_0^2\ \chi_2^2}{n-1}\right\}.$$

(ii) $H_0 : \sigma^2 \geq \sigma_0^2$ contra $H_1 : \sigma^2 < \sigma_0^2$

Neste caso, e sendo x_c o ponto crítico, a RC é

$$RC = \{S^2 : S^2 \leq x_c\}.$$

O ponto crítico x_c será encontrado fixando a probabilidade máxima de cometer o erro tipo I, α_{max}, ou seja, fixando o nível de significância.

$$\alpha_{max} = P_{\sigma_0^2}(S^2 \leq x_c) = P\left(\frac{(n-1)S^2}{\sigma_0^2} \leq \frac{(n-1)x_c}{\sigma_0^2}\right)$$

Definindo

$$\chi_1^2 = \frac{(n-1)x_c}{\sigma_0^2} \iff x_c = \frac{\sigma_0^2\ \chi_1^2}{n-1}.$$

Logo a região crítica associada ao teste das hipóteses $H_0 : \sigma^2 \geq \sigma_0^2$ contra $H_1 : \sigma^2 < \sigma_0^2$ é

$$RC = \left\{S^2 : S^2 \leq \frac{\sigma_0^2\ \chi_1^2}{n-1}\right\}.$$

(ii) $H_0 : \sigma^2 = \sigma_0^2$ contra $H_1 : \sigma^2 \neq \sigma_0^2$

Neste caso, e sendo x_c o ponto crítico, a RC é

$$RC = \{S^2 : S^2 \leq x_{c_1} \vee S^2 \geq x_{c_2}\}.$$

Os pontos críticos x_{c_1} e x_{c_2} serão encontrados fixando a probabilidade máxima de cometer o erro tipo I, α_{max}, ou seja, fixando o nível de significância, e considerando que $P_{\sigma_0^2}(S^2 \leq x_{c_1}) = P_{\sigma_0^2}(S^2 \geq x_{c_2})$.

Por definição, temos

$$\alpha_{max} = \alpha = P_{\sigma_0^2}(S^2 \leq x_{c_1} \vee S^2 \geq x_{c_2})$$

em que $\Theta_0 = \{\sigma_0^2\}$.

$$\alpha = P\left(\frac{(n-1)S^2}{\sigma_0^2} \leq \frac{(n-1)x_{c_1}}{\sigma_0^2}\right) + P\left(\frac{(n-1)S^2}{\sigma_0^2} \geq \frac{(n-1)x_{c_2}}{\sigma_0^2}\right)$$

Considerando

$$P\left(\frac{(n-1)S^2}{\sigma_0^2} \leq \frac{(n-1)x_{c_1}}{\sigma_0^2}\right) = P\left(\frac{(n-1)S^2}{\sigma_0^2} \geq \frac{(n-1)x_{c_2}}{\sigma_0^2}\right) = \frac{\alpha}{2},$$

para um α fixado, e definindo,

$$\chi_1^2 = \frac{(n-1)x_{c_1}}{\sigma_0^2} \quad \text{e} \quad \chi_2^2 = \frac{(n-1)x_{c_2}}{\sigma_0^2},$$

obtemos os pontos críticos

$$x_{c_1} = \frac{\sigma_0^2\ \chi_1^2}{n-1} \quad \text{e} \quad x_{c_2} = \frac{\sigma_0^2\ \chi_2^2}{n-1}.$$

Logo a região crítica associada ao teste das hipóteses $H_0 : \sigma^2 = \sigma_0^2$ contra $H_1 : \sigma^2 \neq \sigma_0^2$ é

$$RC = \left\{S^2 : S^2 \leq \frac{\sigma_0^2\ \chi_1^2}{n-1}\ \vee\ S^2 \geq \frac{\sigma_0^2\ \chi_2^2}{n-1}\right\}.$$

Passo 4: Decisão

Se $s^2 \in$ RC, rejeitamos a hipótese H_0 com $\alpha 100\%$ de significância.

Se $s^2 \notin$ RC, não rejeitamos a hipótese H_0 com $\alpha 100\%$ de significância.

Assim,

- Se $s^2 \in$ RC, dizemos que há evidências para rejeitar H_0, a $\alpha 100\%$ de significância
- Se $s^2 \notin$ RC, dizemos que não há evidências para rejeitar H_0, a $\alpha 100\%$ de significância

Exemplo

1. Com base na amostra aleatória seguinte, teste as hipóteses
   $H_0 : \sigma^2 = 1.3^2$ *versus* $H_1 : \sigma^2 \neq 1.3^2$, a um nível de significância de 5%:
   2.0 3.2 5.0 1.8 3.4 2.6.

4. Teste assintótico para a proporção populacional

Seja X uma variável aleatória com distribuição Binomial com parâmetros n e p, em que $n > 30$ e suponhamos que p_0 é um valor conhecido.

As hipóteses de interesse são:

Passo 1:

- (i) $H_0 : p \leq p_0$ contra $H_1 : p > p_0$;
- (ii) $H_0 : p \geq p_0$ contra $H_1 : p < p_0$;
- (iii) $H_0 : p = p_0$ contra $H_1 : p \neq p_0$.

Passo 2: A estatística a ser usada é

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right).$$

Nos casos (i) e (ii), a hipótese H_0 é composta, pelo que escrevemos

$$\hat{p} \underset{p=p_0}{\sim} N\left(p_0, \frac{p_0(1-p_0)}{n}\right) \iff \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \underset{p=p_0}{\sim} N(0,1).$$

No caso (iii), a hipótese H_0 é simples, pelo que escrevemos

$$\hat{p} \underset{H_0}{\sim} N\left(p_0, \frac{p_0(1-p_0)}{n}\right) \iff \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \underset{H_0}{\sim} N(0,1).$$

Passo 3: Determinação da região crítica

(i) $H_0 : p \leq p_0$ contra $H_1 : p > p_0$

Neste caso, e sendo x_c o ponto crítico, a RC é

$$RC = \{\hat{p} : \hat{p} \geq x_c\}.$$

O ponto crítico x_c será encontrado fixando a probabilidade máxima de cometer o erro tipo I, α_{max}, ou seja, fixando o nível de significância.

$$\alpha_{max} = P_{p_0}(\hat{p} \geq x_c) = P\left(Z \geq \frac{x_c - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}\right)$$

Definindo

$$z_\alpha = \frac{x_c - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \iff x_c = p_0 + z_\alpha\sqrt{\frac{p_0(1-p_0)}{n}}.$$

Logo a região crítica associada ao teste das hipóteses $H_0 : \mu \leq \mu_0$ contra $H_1 : p > p_0$ é

$$RC = \left\{\hat{p} : \hat{p} \geq p_0 + z_\alpha\sqrt{\frac{p_0(1-p_0)}{n}}\right\}.$$

(ii) $H_0 : p \geq p_0$ contra $H_1 : p < p_0$

Neste caso, e sendo x_c o ponto crítico, a RC é

$$RC = \{\hat{p} : \hat{p} \leq x_c\}.$$

O ponto crítico x_c será encontrado fixando a probabilidade máxima de cometer o erro tipo I, α_{max}, ou seja, fixando o nível de significância.

$$\alpha_{max} = P_{p_0}(\hat{p} \leq x_c) = P\left(Z \leq \frac{x_c - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}\right)$$

Definindo

$$-z_\alpha = \frac{x_c - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \iff x_c = p_0 - z_\alpha\sqrt{\frac{p_0(1-p_0)}{n}}.$$

Logo a região crítica associada ao teste das hipóteses $H_0 : p \geq p_0$ contra $H_1 : p < p_0$ é

$$RC = \left\{\hat{p} : \hat{p} \leq p_0 - z_\alpha\sqrt{\frac{p_0(1-p_0)}{n}}\right\}.$$

(ii) $H_0 : p = p_0$ contra $H_1 : p \neq p_0$

Neste caso, e sendo x_c o ponto crítico, a RC é

$$RC = \{\hat{p} : \hat{p} \leq x_{c_1} \vee \hat{p} \geq x_{c_2}\}.$$

Os pontos críticos x_{c_1} e x_{c_2} serão encontrados fixando a probabilidade máxima de cometer o erro tipo I, α_{max}, ou seja, fixando o nível de significância, e considerando a simetria da distribuição.

Por definição, temos

$$\alpha_{max} = \alpha = P_{p_0}(\hat{p} \leq x_{c_1} \vee \hat{p} \geq x_{c_2})$$

em que $\Theta_0 = \{p_0\}$.

$$\alpha = P\left(Z \leq \frac{x_{c_1} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}\right) + P\left(Z \geq \frac{x_{c_2} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}\right)$$

Considerando

$$P\left(Z \leq \frac{x_{c_1} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}\right) = P\left(Z \geq \frac{x_{c_2} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}\right),$$

então

$$\alpha = 2P\left(Z \geq \frac{x_{c_2} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}\right).$$

Definindo $P\left(Z \geq z_{\alpha/2}\right) = \alpha/2$, para um α fixado, obtemos os pontos críticos

$$x_{c_1} = p_0 - z_{\alpha/2}\sqrt{\frac{p_0(1-p_0)}{n}} \quad \text{e} \quad x_{c_2} = p_0 + z_{\alpha/2}\sqrt{\frac{p_0(1-p_0)}{n}}.$$

Logo a região crítica associada ao teste das hipóteses $H_0 : p \geq p_0$ contra $H_1 : p < p_0$ é

$$RC = \left\{\hat{p} : \hat{p} \leq p_0 - z_{\alpha/2}\sqrt{\frac{p_0(1-p_0)}{n}} \ \vee \ \hat{p} \geq p_0 + z_{\alpha/2}\sqrt{\frac{p_0(1-p_0)}{n}}\right\}.$$

Passo 4: Decisão

Se $\hat{p} \in RC$, rejeitamos a hipótese H_0 com $\alpha 100\%$ de significância.

Se $\hat{p} \notin RC$, não rejeitamos a hipótese H_0 com $\alpha 100\%$ de significância.

Assim,

- Se $\hat{p} \in RC$, dizemos que há evidências para rejeitar H_0, a $\alpha 100\%$ de significância
- Se $\hat{p} \notin RC$, dizemos que não há evidências para rejeitar H_0, a $\alpha 100\%$ de significância

Exemplo

1. Um estudo é realizado para determinar a relação entre uma certa droga e certa anomalia em embriões de frango. Injetou-se 50 ovos fertilizados com a droga no quarto dia de incubação. No vigésimo dia de incubação, os embriões foram examinados e 7 apresentaram a anomalia. Suponha que deseja-se averiguar se a proporção verdadeira é inferior a 25% com um nível de significância de 0,05.
   Hipóteses: $H_0 : p \geq 0.25$ contra $H_1 : p < 0.25$