Introdução à Probabilidade e à Estatística II

Testes de Hipóteses para a média populacional

Lígia Henriques-Rodrigues

MAE0229 – 1º semestre 2018

1. Teste para a média de uma população com variância conhecida

Seja X uma variável aleatória com distribuição normal de média μ e variância σ^2 conhecida e suponhamos que μ_0 é um valor conhecido.

As hipóteses de interesse são:

Passo 1:

- (i) $H_0 : \mu \leq \mu_0$ contra $H_1 : \mu > \mu_0$;
- (ii) $H_0 : \mu \geq \mu_0$ contra $H_1 : \mu < \mu_0$;
- (iii) $H_0 : \mu = \mu_0$ contra $H_1 : \mu \neq \mu_0$.

Passo 2: A estatística a ser usada é

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

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

$$\overline{X} \underset{\mu=\mu_0}{\sim} N\left(\mu_0, \frac{\sigma^2}{n}\right) \iff Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \underset{\mu=\mu_0}{\sim} N(0,1).$$

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

$$\overline{X} \underset{H_0}{\sim} N\left(\mu_0, \frac{\sigma^2}{n}\right) \iff Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \underset{H_0}{\sim} N(0,1).$$

Passo 3: Determinação da região crítica

(i) $H_0 : \mu \leq \mu_0$ contra $H_1 : \mu > \mu_0$

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

$$RC = \{\overline{X} : \overline{X} \geq x_c\}.$$

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.

$$\alpha_{max} = P_{\mu_0}(\overline{X} \geq x_c) = P\left(Z \geq \frac{x_c - \mu_0}{\sigma/\sqrt{n}}\right)$$

Definindo

$$z_\alpha = \frac{x_c - \mu_0}{\sigma/\sqrt{n}} \iff x_c = \mu_0 + z_\alpha \sigma/\sqrt{n}.$$

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

$$RC = \{\overline{X} : \overline{X} \geq \mu_0 + z_\alpha \sigma/\sqrt{n}\}.$$

(ii) $H_0 : \mu \geq \mu_0$ contra $H_1 : \mu < \mu_0$

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

$$RC = \{\overline{X} : \overline{X} \leq x_c\}.$$

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.

$$\alpha_{max} = P_{\mu_0}(\overline{X} \leq x_c) = P\left(Z \leq \frac{x_c - \mu_0}{\sigma/\sqrt{n}}\right)$$

Definindo

$$-z_\alpha = \frac{x_c - \mu_0}{\sigma/\sqrt{n}} \iff x_c = \mu_0 - z_\alpha \sigma/\sqrt{n}.$$

Logo a região crítica associada ao teste das hipóteses $H_0 : \mu \geq \mu_0$ contra $H_1 : \mu < \mu_0$ é

$$RC = \{\overline{X} : \overline{X} \leq \mu_0 - z_\alpha \sigma/\sqrt{n}\}.$$

(ii) $H_0 : \mu = \mu_0$ contra $H_1 : \mu \neq \mu_0$

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

$$RC = \{\overline{X} : \overline{X} \leq x_{c_1} \vee \overline{X} \geq x_{c_2}\}.$$

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

$$\alpha_{max} = \alpha = P_{\mu_0}(\overline{X} \leq x_{c_1} \vee \overline{X} \geq x_{c_2})$$

em que $\Theta_0 = \{\mu_0\}$.

$$\alpha = P\left(Z \leq \frac{x_{c_1} - \mu_0}{\sigma/\sqrt{n}}\right) + P\left(Z \geq \frac{x_{c_2} - \mu_0}{\sigma/\sqrt{n}}\right)$$

Considerando

$$P\left(Z \leq \frac{x_{c_1} - \mu_0}{\sigma/\sqrt{n}}\right) = P\left(Z \geq \frac{x_{c_2} - \mu_0}{\sigma/\sqrt{n}}\right),$$

então

$$\alpha = 2P\left(Z \geq \frac{x_{c_2} - \mu_0}{\sigma/\sqrt{n}}\right).$$

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

$$x_{c_1} = \mu_0 - z_{\alpha/2}\sigma/\sqrt{n} \quad \text{e} \quad x_{c_2} = \mu_0 + z_{\alpha/2}\sigma/\sqrt{n}.$$

Logo a região crítica associada ao teste das hipóteses $H_0 : \mu = \mu_0$ contra $H_1 : \mu \neq \mu_0$ é

$$RC = \{\overline{X} : \overline{X} \leq \mu_0 - z_{\alpha/2}\sigma/\sqrt{n} \ \vee \ \overline{X} \geq \mu_0 + z_{\alpha/2}\sigma/\sqrt{n}\}.$$

Passo 4: Decisão

Se $\overline{x} \in \mathrm{RC}$, rejeitamos a hipótese H_0 com $\alpha 100\%$ de significância.

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

Assim,

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

Exemplos

1. Um fabricante de uma certa peça afirma que o tempo médio de vida das peças produzidas é de no máximo 1000 horas. Admitindo que o tempo de vida das peças produzidas segue distribuição normal com valor médio μ e desvio padrão $\sigma = 400$. Uma amostra de 50 peças, escolhidas ao acaso, forneceu uma média de 1100 horas. Suponha que os engenheiros de produção têm interesse em verificar se a modificação do processo de fabricação aumenta a duração das peças ($\alpha = 10\%$).

   Hipóteses: $H_0 : \mu \leq 1000$ contra $H_1 : \mu > 1000$

2. Um comprador de tijolos acha que a qualidade dos tijolos está diminuindo. De experiências anteriores, considera-se que a resistência ao desmoronamento de tais tijolos é normalmente distribuída com valor médio superior ou igual a 200 kg e desvio padrão de 10 kg. Uma amostra de 100 tijolos, escolhidos ao acaso, forneceu uma média de 195 kg. Ao nível de significância de 5%, pode-se afirmar que a resistência média ao desmoronamento diminuiu?

   Hipóteses: $H_0 : \mu \geq 200$ contra $H_1 : \mu < 200$

3. Os registros dos últimos anos de um colégio atestam para calouros admitidos uma nota média 115 (teste vocacional) e um desvio padrão 20. Supondo que as notas dos novos calouros tem distribuição normal com média μ e desvio padrão σ, teste a hipótese de que a média de uma nova turma é a mesma das turmas anteriores. Retirou-se, ao acaso, uma amostra de 20 notas, obtendo-se média 118. Use $\alpha = 0,01$.

   Hipóteses: $H_0 : \mu = 115$ contra $H_1 : \mu \neq 115$

2. Teste para a média de uma população com variância desconhecida e n pequeno

Seja X uma variável aleatória com distribuição normal de média μ e variância σ^2 desconhecida, n pequeno e suponhamos que μ_0 é um valor conhecido.

As hipóteses de interesse são:

Passo 1:

- (i) $H_0 : \mu \leq \mu_0$ contra $H_1 : \mu > \mu_0$;
- (ii) $H_0 : \mu \geq \mu_0$ contra $H_1 : \mu < \mu_0$;
- (iii) $H_0 : \mu = \mu_0$ contra $H_1 : \mu \neq \mu_0$.

Passo 2: A estatística a ser usada é

$$\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1).$$

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

$$\frac{\overline{X}-\mu}{S/\sqrt{n}} \underset{\mu=\mu_0}{\sim} t(n-1) \iff \frac{\overline{X}-\mu_0}{S/\sqrt{n}} \underset{\mu=\mu_0}{\sim} t(n-1).$$

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

$$\frac{\overline{X}-\mu}{S/\sqrt{n}} \underset{H_0}{\sim} t(n-1) \iff \frac{\overline{X}-\mu_0}{S/\sqrt{n}} \underset{H_0}{\sim} t(n-1).$$

Passo 3: Determinação da região crítica

(i) $H_0 : \mu \leq \mu_0$ contra $H_1 : \mu > \mu_0$

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

$$RC = \{\overline{X} : \overline{X} \geq x_c\}.$$

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.

$$\alpha_{max} = P_{\mu_0}(\overline{X} \geq x_c) = P\left(\frac{\overline{X} - \mu}{S/\sqrt{n}} \geq \frac{x_c - \mu_0}{S/\sqrt{n}}\right)$$

Definindo

$$t_\alpha = \frac{x_c - \mu_0}{S/\sqrt{n}} \iff x_c = \mu_0 + t_\alpha S/\sqrt{n}.$$

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

$$RC = \{\overline{X} : \overline{X} \geq \mu_0 + t_\alpha S/\sqrt{n}\}.$$

(ii) $H_0 : \mu \geq \mu_0$ contra $H_1 : \mu < \mu_0$

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

$$RC = \{\overline{X} : \overline{X} \leq x_c\}.$$

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.

$$\alpha_{max} = P_{\mu_0}(\overline{X} \leq x_c) = P\left(\frac{\overline{X} - \mu}{S/\sqrt{n}} \leq \frac{x_c - \mu_0}{S/\sqrt{n}}\right)$$

Definindo

$$-t_\alpha = \frac{x_c - \mu_0}{S/\sqrt{n}} \iff x_c = \mu_0 - t_\alpha S/\sqrt{n}.$$

Logo a região crítica associada ao teste das hipóteses $H_0 : \mu \geq \mu_0$ contra $H_1 : \mu < \mu_0$ é

$$RC = \{\overline{X} : \overline{X} \leq \mu_0 - t_\alpha S/\sqrt{n}\}.$$

(ii) $H_0 : \mu = \mu_0$ contra $H_1 : \mu \neq \mu_0$

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

$$RC = \{\overline{X} : \overline{X} \leq x_{c_1} \vee \overline{X} \geq x_{c_2}\}.$$

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

$$\alpha_{max} = \alpha = P_{\mu_0}(\overline{X} \leq x_{c_1} \vee \overline{X} \geq x_{c_2})$$

em que $\Theta_0 = \{\mu_0\}$.

$$\alpha = P\left(\frac{\overline{X} - \mu}{S/\sqrt{n}} \leq \frac{x_{c_1} - \mu_0}{S/\sqrt{n}}\right) + P\left(\frac{\overline{X} - \mu}{S/\sqrt{n}} \geq \frac{x_{c_2} - \mu_0}{S/\sqrt{n}}\right)$$

Considerando

$$P\left(\frac{\overline{X}-\mu}{S/\sqrt{n}} \leq \frac{x_{c_1}-\mu_0}{S/\sqrt{n}}\right) = P\left(\frac{\overline{X}-\mu}{S/\sqrt{n}} \geq \frac{x_{c_2}-\mu_0}{S/\sqrt{n}}\right),$$

então

$$\alpha = 2P\left(\frac{\overline{X}-\mu}{S/\sqrt{n}} \geq \frac{x_{c_2}-\mu_0}{S/\sqrt{n}}\right).$$

Definindo $P\left(\frac{\overline{X}-\mu}{S/\sqrt{n}} \geq t_{\alpha/2}\right) = \alpha/2$, para um α fixado, obtemos os pontos críticos

$$x_{c_1} = \mu_0 - t_{\alpha/2} S/\sqrt{n} \quad \text{e} \quad x_{c_2} = \mu_0 + t_{\alpha/2} S/\sqrt{n}.$$

Logo a região crítica associada ao teste das hipóteses $H_0 : \mu = \mu_0$ contra $H_1 : \mu \neq \mu_0$ é

$$RC = \{\overline{X} : \overline{X} \leq \mu_0 - t_{\alpha/2} S/\sqrt{n} \ \vee \ \overline{X} \geq \mu_0 + t_{\alpha/2} S/\sqrt{n}\}.$$

Passo 4: Decisão

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

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

Assim,

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

Exemplos

1. Um fabricante de uma certa peça afirma que o tempo médio de vida das peças produzidas é de no máximo 1000 horas. Admitindo que o tempo de vida das peças produzidas segue distribuição normal com valor médio μ e desvio padrão σ. Uma amostra de 21 peças, escolhidas ao acaso, forneceu uma média de 1100 horas e um desvio padrão de 350 horas. Suponha que os engenheiros de produção têm interesse em verificar se a modificação do processo de fabricação aumenta a duração das peças ($\alpha = 10\%$).

2. Um comprador de tijolos acha que a qualidade dos tijolos está diminuindo. De experiências anteriores, considera-se que a resistência ao desmoronamento de tais tijolos é normalmente distribuída com valor médio superior ou igual a 200 kg e desvio padrão desconhecido. Uma amostra de 26 tijolos, escolhidos ao acaso, forneceu uma média de 195 kg e um desvio padrão de 9 kg. Ao nível de significância de 5%, pode-se afirmar que a resistência média ao desmoronamento diminuiu?

3. Os registros dos últimos anos de um colégio atestam para calouros admitidos uma nota média 115 (teste vocacional). Supondo que as notas dos novos calouros tem distribuição normal com média μ e desvio padrão σ, teste a hipótese de que a média de uma nova turma é a mesma das turmas anteriores. Retirou-se, ao acaso, uma amostra de 20 notas, obtendo-se média 118 e um desvio padrão de 22. Use $\alpha = 0,01$.

3. Teste para a média de uma população com variância desconhecida e n grande

Seja X uma variável aleatória com média μ, variância σ^2 desconhecida, n grande e suponhamos que μ_0 é um valor conhecido.

As hipóteses de interesse são:

Passo 1:

- (i) $H_0 : \mu \leq \mu_0$ contra $H_1 : \mu > \mu_0$;
- (ii) $H_0 : \mu \geq \mu_0$ contra $H_1 : \mu < \mu_0$;
- (iii) $H_0 : \mu = \mu_0$ contra $H_1 : \mu \neq \mu_0$.

Passo 2: A estatística a ser usada é

$$\overline{X} \approx N\left(\mu, \frac{S^2}{n}\right).$$

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

$$\overline{X} \underset{\mu=\mu_0}{\approx} N\left(\mu_0, \frac{S^2}{n}\right) \iff Z = \frac{\overline{X} - \mu_0}{S/\sqrt{n}} \underset{\mu=\mu_0}{\approx} N(0,1).$$

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

$$\overline{X} \underset{H_0}{\approx} N\left(\mu_0, \frac{S^2}{n}\right) \iff Z = \frac{\overline{X} - \mu_0}{S/\sqrt{n}} \underset{H_0}{\approx} N(0,1).$$

Passo 3: Dado α

(i) $RC = \{\overline{X} : \overline{X} \geq x_c\}$, com $x_c = \mu_0 + z_\alpha S/\sqrt{n}$;

(ii) $RC = \{\overline{X} : \overline{X} \leq x_c\}$, com $x_c = \mu_0 - z_\alpha S/\sqrt{n}$;

(iii) $RC = \{\overline{X} : \overline{X} \leq x_{c_1} \vee \overline{X} \geq x_{c_2}\}$, com $x_{c_1} = \mu_0 - z_{\alpha/2} S/\sqrt{n}$ e
$x_{c_2} = \mu_0 + z_{\alpha/2} S/\sqrt{n}$.

Passo 4: Decisão

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

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

Assim,

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