Capítulo 7 - Testes de Hipóteses

Nos capítulos anteriores considerámos a estimação pontual **(Capítulo 5)** e intervalar **(Capítulo 6)** de um parâmetro populacional desconhecido, digamos θ, partindo de uma amostra aleatória $X_1, \ldots, X_n$ da população respectiva.

No entanto, num estudo estatístico nem sempre o interesse do investigador está em estimar o valor de θ mas sim aferir/**testar** se o mesmo pertence a um determinado subconjunto do espaço de valores possíveis para o parâmetro θ, digamos Θ.

Definição (Hipótese Estatística)

Uma hipótese estatística é uma conjectura acerca de um qualquer parâmetro populacional desconhecido e/ou acerca da própria distribuição da população.

Se a hipótese estatística conjectura um único valor para o parâmetro populacional desconhecido e/ou especifica completamente a distribuição da população então diz-se uma **hipótese simples** *(por exemplo, $\theta \in \{\theta_0\} \subseteq \Theta$).*

Caso contrário é chamada de **hipótese composta** *(por exemplo, $\theta \in \Theta_0 \subseteq \Theta$).*

Testes de Hipóteses

Definição (Hipóteses nula e alternativa)

Para cada hipótese sobre θ que se conjecture, doravante designada por **hipótese nula** *(H_0), há sempre outra hipótese, designada por* **hipótese alternativa** *(H_1). A saber,*

- *H_0 quantifica/conjectura a crença do investigador sobre o valor do parâmetro desconhecido e/ou a distribuição da população*
- *H_1 é qualquer hipótese que contradiga a hipótese nula*

Definição (Teste de hipóteses)

Um teste de hipóteses é um procedimento estatístico que permite testar a hipótese nula H_0 contra a hipótese alternativa H_1 e que envolve a recolha de uma amostra representativa da população em causa (reúne evidência estatística contra ou a favor H_0) bem como uma regra de decisão (quantifica quando temos evidência suficiente para rejeitar H_0).

*Habitualmente H_0 é rejeitada quando se considera ter reunido "***muita***" evidência contra H_0.*

Algumas notas importantes:

- Não rejeitar H_0 não significa que H_0 seja verdade! Significa sim, que não se reuniu, de acordo com a regra de decisão estabelecida, evidência estatística suficiente contra H_0
- Rejeitar H_0 significa que se reuniu evidência estatística suficiente contra H_0, admitindo-se portanto a plausibilidade de H_1

Exemplo

Seja $(X_1, \ldots, X_n)$ *uma amostra aleatória da população* X *dos pesos das formigas Solenopsis onde* $X \sim N(\mu, 2^2)$. *Um biólogo pretende testar*

$$H_0 : \mu \leq 8 \quad vs \quad H_1 : \quad \mu > 8,$$

usando o critério "rejeitar H_0 *se* $\frac{\bar{X}-8}{2/\sqrt{n}} > 1.64$".

Tendo-se observado os pesos seguintes

$$7.88 \; 8.54 \; 12.12 \; 9.14 \; 9.26 \; 12.43 \; 9.92 \; 6.47 \; 7.63 \; 8.11$$

indique se há ou não evidência estatística contra a conjectura do biólogo.

Com base na amostra observada calculamos $\bar{x} = 9.15$ vindo portanto $\frac{9.15-8}{2/\sqrt{10}} \simeq 1.82 > 1.64$ pelo que, segundo a regra de decisão estabelecida, se conclui haver evidência para rejeitar H_0.

Exemplo

No contexto do exemplo anterior, se o biólogo tivesse observado $\bar{x} = 9.15$ *numa amostra de tamanho* $n = 5$ *qual seria a decisão do teste? Comente os dois resultados.*

Neste caso viria $\frac{9.15-8}{2/\sqrt{5}} \simeq 1.29 \not> 1.64$ pelo que, segundo a regra de decisão estabelecida não se reuniu evidência suficiente para rejeitar H_0.

Observamos agora, que apesar de em ambos os casos se ter obtido uma média amostral $\bar{x} = 9.15$, a decisão do teste foi diferente. Isto significa em particular, que **o tamanho amostral é muito importante!!!**

Testes de Hipóteses

Definição (Erros do tipo I e II e potência do teste)

A rejeição de H_0 quando ela é verdadeira é chamado **erro do tipo I** *do teste e a não rejeição de H_0 quando esta é falsa é chamado* **erro do tipo II** *do teste.*

Sejam as probabilidades

$$\alpha = P(\textit{erro de tipo I}) = P(\textit{rejeitar } H_0 | H_0 \textit{ é verdadeira})$$
$$\beta = P(\textit{erro de tipo II}) = P(\textit{não rejeitar } H_0 | H_0 \textit{ é falsa}).$$

Nestas condições, α diz-se o **nível de significância** *do teste e $1-\beta$ diz-se a* **potência do teste**.

Idealmente gostar-se-ia de minimizar os dois erros (tipo I e II) simultaneamente (equivalentemente, minimizar o erro do tipo I e maximizar a potência do teste).

Infelizmente tal é impossível (dado que a redução de um implica o aumento do outro) pelo que, num teste de hipóteses estatístico, o erro que se controla é o erro do tipo I (também chamado **nível do teste**), i.e., o teste é realizado obrigando a que o erro do tipo I não exceda um determinado valor.

Os níveis usuais de um teste de hipóteses estatístico são $\alpha = 0.01,\ 0.05$ **e** 0.1.

Teste de hipóteses para o valor médio populacional, μ

Os testes de hipóteses para a média populacional μ que iremos estudar são

$$H_0 : \mu = \mu_0 \ vs \ H_1 : \mu \neq \mu_0 \qquad \text{(teste bilateral)}$$

$$H_0 : \mu \leq \mu_0 \ vs \ H_1 : \mu > \mu_0 \qquad \text{(teste unilateral direito)}$$

$$H_0 : \mu \geq \mu_0 \ vs \ H_1 : \mu < \mu_0 \qquad \text{(teste unilateral esquerdo)}$$

A ideia para a realização os testes de hipóteses para a média populacional μ passa por:

- observar uma concretização $x_1, \ldots, x_n$ da a.a. $X_1, \ldots, X_n$
- usar o estimador $\bar{X}$ de μ (que sabemos ser centrado e consistente) para estimar μ, i.e, calcular $\bar{x}$
- se $\bar{x}$ estiver "**(muito) longe**" de μ_0 rejeitar H_0, caso contrário indicar que não se reuniu evidência estatística para rejeitar H_0

Teste de hipóteses para o valor médio populacional, μ

No contexto de um teste de hipóteses estatístico, a noção de **"(muito) longe"** está relacionada com $\bar{x}$ cair nas caudas da distribuição de $\bar{X}$ sob H_0, ou equivalentemente, com

$$\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \ (\sigma^2 \text{ conhecido}) \qquad \text{ou} \qquad \frac{\bar{x}-\mu_0}{s/\sqrt{n}} \ (\sigma^2 \text{ desconhecido})$$

caírem nas caudas das distribuições de

$$\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \ (\sigma^2 \text{ conhecido}) \qquad \text{ou} \qquad \frac{\bar{X}-\mu_0}{S/\sqrt{n}} \ (\sigma^2 \text{ desconhecido})$$

sob H_0, respectivamente.

A área atribuída à(s) cauda(s) é igual ao nível de significância α que é pré-especificado pelo investigador. A decisão passa então por rejeitar H_0 ao nível de significância α, se o valor observado da estatística pertencer ao intervalo de valores (designado por **região crítica** do teste) que é definido pela área atribuída à(s) cauda(s).

Teste de hipóteses para o valor médio populacional, μ

População Normal com variância σ^2 conhecida

Considere-se uma amostra aleatória $(X_1, \ldots, X_n)$ de uma população X tal que $X \sim N(\mu, \sigma^2)$, em que σ^2 é conhecido. Queremos testar

$$H_0 : \mu = \mu_0 \ vs \ H_1 : \mu \neq \mu_0$$

(**NOTA:** teste bilateral $\Rightarrow$ rejeição nas duas caudas)

Para isso consideramos o estimador $\bar{X}$ de μ que sabemos nestas condições ser tal que $\bar{X} \sim N(\mu, \sigma^2/n)$. **Sob** H_0 virá então que $\bar{X} \sim N(\mu_0, \sigma^2/n)$, i.e.,

$$\bar{X} \underset{H_0}{\sim} N(\mu_0, \sigma^2/n).$$

Assim a estatística de teste para este teste de hipóteses será então

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \underset{H_0}{\sim} N(0, 1)$$

faltando-nos portanto definir nas caudas da distribuição da estatística de teste sob H_0 a **região crítica** do teste!

Para um **nível de significância** α, pré-especificado, define-se a região de rejeição do teste, como a solução da inequação

$$\left| \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \right| > z_{\alpha/2},$$

(isto equivale a dizer que Z cai na cauda direita ou na cauda esquerda da Normal reduzida)

e escrevemos

$$RC_\alpha =]-\infty, -z_{\alpha/2}[\, \cup \,] + z_{\alpha/2}, +\infty[.$$

Deste modo, observada uma concretização $x_1, \ldots, x_n$ da a.a. $X_1, \ldots, X_n$, rejeitamos H_0 ao nível α sempre que

$$z_{obs} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \in RC_\alpha.$$

Teste de hipóteses para o valor médio populacional, μ

População Normal com variância σ^2 conhecida

Exemplo

Consideremos novamente o exemplo da população X dos pesos (em dg) das formigas Solenopsis tal que $X \sim N(\mu, 2^2)$, da qual observámos a amostra de 4 pesos $(8, 13, 9, 8.5)$. Pretende-se testar, a um nível de significância α de 5%, a hipótese de que o peso médio populacional μ é 10dg.

Informação populacional: $X \sim N(\mu, \sigma^2)$ com $\sigma^2 = 4$ conhecida

Informação amostral: $n = 4$; $\bar{x} = 9.625$

Hipóteses em teste: $H_0 : \mu = 10 \; vs \; H_1 : \mu \neq 10 \quad (\mu_0 = 10)$ (teste bilateral)

Estatística de teste: $Z = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \underset{H_0}{\sim} N(0,1)$ com $X_1, \ldots, X_n$ a.a. da pop. X

Valor observado da estatística de teste: $z_{obs} = \dfrac{\bar{x} - 10}{2/\sqrt{4}} = \dfrac{9.625 - 10}{2/\sqrt{4}} = -0.375$

Região crítica do teste: $\alpha = 0.05 \Leftrightarrow \frac{\alpha}{2} = 0.025$ donde $z_{0.025} \underset{tabela}{\simeq} 1.96$ e assim

$$RC_{0.05} =]-\infty, -1.96[\cup]1.96, +\infty[$$

Decisão do teste: como $z_{obs} = -0.375 \notin RC_{0.05}$ não há evidência para rejeitar H_0 ao nível 5%.

Teste de hipóteses para o valor médio populacional, μ

Distinguimos agora as 4 possíveis estatísticas de teste para o teste de hipóteses bilateral bem como as respectivas regiões críticas do teste

$$H_0 : \mu = \mu_0 \;\; vs \;\; H_1 : \mu \neq \mu_0 \qquad \text{(teste bilateral)}$$

População	Variância σ^2	Estatística de teste	Região crítica RC_α
$N(\mu, \sigma^2)$	conhecida	$\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \underset{H_0}{\sim} N(0,1)$	$]-\infty, -z_{\frac{\alpha}{2}}[\cup]+z_{\frac{\alpha}{2}}, +\infty[$
	desconhecida	$\frac{\overline{X}-\mu_0}{S/\sqrt{n}} \underset{H_0}{\sim} t_{n-1}$	$]-\infty, -t_{n-1,\frac{\alpha}{2}}[\cup]+t_{n-1,\frac{\alpha}{2}}, +\infty[$
Desconhecida $(n \geq 30)$	conhecida	$\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \overset{a}{\underset{H_0}{\sim}} N(0,1)$	$]-\infty, -z_{\frac{\alpha}{2}}[\cup]+z_{\frac{\alpha}{2}}, +\infty[$
	desconhecida	$\frac{\overline{X}-\mu_0}{S/\sqrt{n}} \overset{a}{\underset{H_0}{\sim}} N(0,1)$	$]-\infty, -z_{\frac{\alpha}{2}}[\cup]+z_{\frac{\alpha}{2}}, +\infty[$

Teste de hipóteses para o valor médio populacional, μ

Analogamente, agora para o teste de hipóteses unilateral direito

$$H_0 : \mu \leq \mu_0 \;\; vs \;\; H_1 : \mu > \mu_0 \qquad \text{(teste unilateral direito)}$$

População	Variância σ^2	Estatística de teste	Região crítica RC_α
$N(\mu, \sigma^2)$	conhecida	$\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \underset{H_0}{\sim} N(0,1)$	$]z_\alpha, +\infty[$
	desconhecida	$\frac{\overline{X}-\mu_0}{S/\sqrt{n}} \underset{H_0}{\sim} t_{n-1}$	$]t_{n-1,\alpha}, +\infty[$
Desconhecida $(n \geq 30)$	conhecida	$\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \underset{H_0}{\overset{a}{\sim}} N(0,1)$	$]z_\alpha, +\infty[$
	desconhecida	$\frac{\overline{X}-\mu_0}{S/\sqrt{n}} \underset{H_0}{\overset{a}{\sim}} N(0,1)$	$]z_\alpha, +\infty[$

Teste de hipóteses para o valor médio populacional, μ

Analogamente, agora para o teste de hipóteses unilateral esquerdo

$$H_0 : \mu \geq \mu_0 \;\; vs \;\; H_1 : \mu < \mu_0 \qquad \text{(teste unilateral esquerdo)}$$

População	Variância σ^2	Estatística de teste	Região crítica RC_α
$N(\mu, \sigma^2)$	conhecida	$\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \underset{H_0}{\sim} N(0,1)$	$]-\infty, -z_\alpha[$
	desconhecida	$\frac{\overline{X}-\mu_0}{S/\sqrt{n}} \underset{H_0}{\sim} t_{n-1}$	$]-\infty, -t_{n-1,\alpha}[$
Desconhecida $(n \geq 30)$	conhecida	$\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \underset{H_0}{\overset{a}{\sim}} N(0,1)$	$]-\infty, -z_\alpha[$
	desconhecida	$\frac{\overline{X}-\mu_0}{S/\sqrt{n}} \underset{H_0}{\overset{a}{\sim}} N(0,1)$	$]-\infty, -z_\alpha[$

Definição ($p - valor$ do teste)

O $p - valor$ de um teste de hipóteses estatístico quantifica a probabilidade de, sob a distribuição postulada em H_0, se observar uma amostra como a amostra que foi observada (equivalentemente, de, sob a distribuição postulada em H_0, se observar um valor da estatística de teste igual ou mais extrema do que a que foi observada). Assim,

- *um $p - valor$ pequeno é desfavorável a H_0.*
- *um $p - valor$ elevado indica que as observações são consistentes com H_0.*

Num teste de hipóteses de nível α, a regra de decisão é

"Rejeitar H_0 se $p - valor < \alpha$"

decisão esta, que é equivalente à decisão baseada na região crítica RC_α.

Regra de cálculo do $p - valor$:

Seja $(x_1, x_2, \ldots, x_n)$ uma concretização da amostra aleatória $X_1, \ldots, X_n$ e $w_{obs} = W(x_1, x_2, \ldots, x_n)$ o valor observado da estatística de teste W.

Região de rejeição	$p - valor$
$]-\infty, -c[\cup]c, +\infty[$ ou $]0, b[\cup]c, +\infty[$	$2 \times \min\left(P(W < w_{obs} \mid H_0), P(W > w_{obs} \mid H_0)\right)$ (teste bilateral)
$]-\infty, c[$ ou $]0, c[$	$P(W < w_{obs} \mid H_0)$ (teste unilateral esquerdo)
$]c, +\infty[$	$P(W > w_{obs} \mid H_0)$ (teste unilateral direito)

Exemplo

Consideremos novamente o exemplo da população X dos pesos (em dg) das formigas Solenopsis tal que $X \sim N(\mu, 2^2)$, da qual observámos a amostra de 4 pesos $(8, 13, 9, 8.5)$. Pretende-se testar, a um nível de significância α de 5%, a hipótese de que o peso médio populacional μ é 10dg. Tome a sua decisão com base no $p - valor$ do teste.

Este exercício é em tudo idêntico ao anterior de modo que passamos imediatamente ao cálculo do $p - valor$

$$
\begin{aligned}
p{-}valor &= 2{\times}\min\{P(Z > z_{obs}), P(Z < z_{obs})\} = 2{\times}\min\{P(Z > -0.375), P(Z < -0.375)\} \\
&= 2 \times P(Z < -0.375) = 2\Phi(-0.375) = 2(1 - \Phi(0.375)) \simeq 2(1 - \Phi(0.38)) \\
&\underset{tabela}{\simeq} 2(1 - 0.6480) = 2 \times 0.352 = 0.704
\end{aligned}
$$

Decisão do teste: como $p - valor = 0.704 \not< 0.05$ não temos evidência para rejeitar H_0 ao nível 5%.

Testes de hipóteses para a variância σ^2 de uma população Normal com média desconhecida

Hipóteses:

1. $H_0 : \sigma^2 = \sigma_0^2 \;\; vs \;\; H_1 : \sigma^2 \neq \sigma_0^2$ (teste bilateral)
2. $H_0 : \sigma^2 \leq \sigma_0^2 \;\; vs \;\; H_1 : \sigma^2 > \sigma_0^2$ (teste unilateral direito)
3. $H_0 : \sigma^2 \geq \sigma_0^2 \;\; vs \;\; H_1 : \sigma^2 < \sigma_0^2$ (teste unilateral esquerdo)

Estatística de teste:

$$X^2 = \frac{(n-1)S^2}{\sigma_0^2} \underset{H_0}{\sim} \chi^2_{n-1}$$

Região de rejeição do teste, para um nível de significância α pré-especificado:

1. $R_\alpha =]0; \chi^2_{n-1,1-\frac{\alpha}{2}}[\, \cup \,]\chi^2_{n-1,\frac{\alpha}{2}}; +\infty[$ (teste bilateral)
2. $R_\alpha =]\chi^2_{n-1,\alpha}; +\infty[$ (teste unilateral direito)
3. $R_\alpha =]0; \chi^2_{n-1,1-\alpha}[$ (teste unilateral esquerdo)

Testes de hipóteses para a variância σ^2 de uma população Normal com média desconhecida

Exemplo

Consideremos novamente o exemplo da população X dos pesos (em dg) das formigas Solenopsis tal que $X \sim N(\mu, \sigma^2)$, da qual observámos a amostra de 4 pesos $(8, 13, 9, 8.5)$. Pretende-se testar, a um nível de significância α de 5%, a hipótese de que o desvio padrão populacional σ é superior a 2.25dg.

Informação populacional: $X \sim N(\mu, \sigma^2)$ com σ^2 desconhecida

Informação amostral: $n = 4$; $\bar{x} = 9.625$; $s^2 \simeq 5.229$

Hipóteses em teste: $H_0 : \sigma^2 \leq 2.25^2 \; vs \; H_1 : \sigma^2 > 2.25^2$ (teste unilateral direito)

Estatística de teste: $X^2 = \dfrac{(n-1)S^2}{\sigma_0^2} \underset{H_0}{\sim} \chi^2_{n-1}$ com $X_1, \ldots, X_n$ a.a. da pop. X

Valor observado da estatística de teste: $x^2_{obs} = 3 \times 5.229/2.25^2 \simeq 3.099$

Região crítica do teste: $\alpha = 0.05$; $\chi^2_{3,0.05} \underset{tabela}{\simeq} 7.815$ e portanto $RC_{0.05} =]7.815, +\infty[$

Decisão do teste: como $x^2_{obs} = 3.099 \notin RC_{0.05}$ não há evidência para rejeitar H_0 ao nível 5%; usando $p-valor = P(X^2 > x^2_{obs}) = P(X^2 > 3.099) \underset{Rstudio}{\simeq} 0.377 \not< 0.05$ $\checkmark$

Testes de hipóteses para a proporção populacional, p

Suponhamos que observamos uma amostra aleatória de dimensão n de uma população, em que determinada proporção desconhecida p dos seus elementos possui certa característica.

Hipóteses:

1. $H_0 : p = p_0 \;\; vs \;\; H_1 : p \neq p_0$ (teste bilateral)
2. $H_0 : p \leq p_0 \;\; vs \;\; H_1 : p > p_0$ (teste unilateral direito)
3. $H_0 : p \geq p_0 \;\; vs \;\; H_1 : p < p_0$ (teste unilateral esquerdo)

Estatística de teste:

$$Z = \frac{\hat{P} - p_0}{\sqrt{p_0(1-p_0)/n}} \underset{H_0}{\overset{a}{\sim}} N(0,1)$$

Região de rejeição do teste, para um nível de significância α pré-especificado:

1. $R_\alpha =]-\infty; -z_{\frac{\alpha}{2}}[\; \cup \;]z_{\frac{\alpha}{2}}; +\infty[$ (teste bilateral)
2. $R_\alpha =]z_\alpha; +\infty[$ (teste unilateral direito)
3. $R_\alpha =]-\infty; -z_\alpha[$ (teste unilateral esquerdo)

Testes de hipóteses para a proporção populacional, p

Exemplo

Um médico oncologista está convencido de que percentagem de doentes com cancro do cólon que morre 5 *anos após a sua detecção é superior ou igual a* 10%. *De 200 casos de pessoas com cancro do cólon, aleatoriamente detectadas, 12 morreram após 5 anos da detecção. Teste a conjectura do médico aos níveis de significância usuais.*

Informação populacional:

$X =$ nº pessoas que morreram de cancro do cólon em n casos,

sendo p a probabilidade de morrer (sucesso) $\sim Bin(n, p)$

Informação amostral: $n = 200$; $x = 12$; vindo a estimativa $\hat{p} = 0.06$

Hipóteses em teste: $H_0 : \; p \geq 0.1 \; vs \; H_1 : p < 0.1$ (teste unilateral esquerdo)

Estatística de teste: $Z = \dfrac{\hat{P} - p_0}{\sqrt{p_0(1-p_0)/n}} \underset{H_0}{\overset{a}{\sim}} N(0,1)$

Valor observado da estatística de teste: $z_{obs} = \dfrac{0.06 - 0.1}{\sqrt{0.1 \times 0.9/200}} \simeq -1.89$

P-valor do teste: $p - valor \overset{a}{\simeq} P(Z < z_{obs}) = \Phi(-1.89) = 1 - \Phi(1.89) \underset{tabela}{\simeq} 0.0294$

Decisão do teste: sendo $p - valor < 0.05, 0.1$ e tal que $p - valor \not< 0.01$ temos evidência para rejeitar H_0 apenas aos níveis 5 e 10%

Intervalos de confiança e testes de hipóteses

Existe uma relação estreita entre os ICs estudados no Capítulo 6 e os testes de hipóteses bilaterais

Quando construímos um IC a $(1-\alpha)100\%$ para um parâmetro desconhecido θ,

- todos os valores nesse intervalo são valores plausíveis para o parâmetro θ
- todos os valores fora desse intervalo são valores implausíveis para o parâmetro θ

Assim, quando o valor θ_0 especificado pela hipótese nula para θ

- **estiver contido** no $IC_{(1-\alpha)100\%}(\theta)$, **não temos** evidência para rejeitar H_0 ao nível $\alpha 100\%$
- **não estiver contido** no $IC_{(1-\alpha)100\%}(\theta)$, **temos** evidência para rejeitar H_0 ao nível $\alpha 100\%$

Vejam o video

https://www.jbstatistics.com/the-relationship-between-confidence-intervals-and-hypothesis-tests/