

# NOÇÕES DE TESTE DE HIPÓTESES (II)

Teste de Hipóteses sobre  $p$

Nível Descritivo

# Resumo

$X \sim \text{binomial}(n; p)$



(1) Estabelecer as **hipóteses sobre  $p$** :

$$H: p = p_0 \quad \times \quad A: p \neq p_0 ; \quad (\text{ou } A: p > p_0 , \text{ ou } A: p < p_0)$$

(2) Escolher um **nível de significância  $\alpha$** .

(3) Determinar a **região crítica  $RC$**  da forma  
 $\{ X \leq k_1 , X \geq k_2 \}$ ; (ou  $\{ X \geq k \}$  ou  $\{ X \leq k \}$ )

- Fixar  $\alpha$  e obter  $RC$   
ou
- Fixar  $RC$  e obter  $\alpha$

(4) Selecionar uma **amostra**: obter  $X = x$

(5) Regra de **Decisão**:  
 $x \in RC \Rightarrow$  rejeitamos  $H$   
 $x \notin RC \Rightarrow$  não rejeitamos  $H$

**Exemplo 1:** Uma indústria de medicamentos afirma que o processo de fabricação de embalagens produz 90% de itens dentro das especificações. O *IPEM* deseja investigar se este processo de fabricação ainda está sob controle.

Sendo  $p$  a proporção de embalagens dentro das especificações, as hipóteses de interesse são:

$$H: p = 0,9$$

$$A: p < 0,9$$

Seleciona-se uma amostra aleatória de 15 itens e se observa o número de itens satisfatórios ( $X$ ), então  $X \sim b(15, p)$ .

Região crítica:  $RC = \{X \leq k\}$

Logo, para  $\alpha = 6\%$  temos  $k = 11$  e  $RC = \{X \leq 11\}$ .

Para  $\alpha = 1\%$  temos  $k = 9$  e  $RC = \{X \leq 9\}$ .  $\Rightarrow$

Se  $x_{obs} = 10$  itens satisfatórios, então

a)  $\alpha = 6\% \Rightarrow 10 \in RC$

**Rejeitamos  $H$**  ao nível de significância de 6%.

b)  $\alpha = 1\% \Rightarrow 10 \notin RC$

**Não rejeitamos  $H$**  ao nível de significância de 1%.

Crítica: arbitrariedade na escolha da  $RC$  (nível de significância).

Sugestão: determinar um nível de significância associado à evidência experimental.

**NÍVEL DESCRITIVO:**  $P$  (ou "*P-value*")

Nesse exemplo, temos  $RC = \{X \leq k\}$  e  $x_{obs} = 10$ .

*Qual seria o nível de significância se fosse utilizada a  $RC = \{X \leq 10\}$  ?*

O nível de significância seria

$$P(X \leq 10 \mid p = 0,9) = 0,0127.$$

Esse valor é o nível descritivo do teste.

Indicando-o por  $P$ , então  $P = 0,0127$ .

Se o nível de significância  $\alpha$  é maior ou igual ao nível descritivo  $P$ , então a região crítica de nível  $\alpha$  contém o valor observado e, portanto, rejeitamos  $H_0$ .

Portanto, para um  $\alpha$  fixado,

**$P \leq \alpha \Rightarrow$  rejeitamos  $H_0$**

**$P > \alpha \Rightarrow$  não rejeitamos  $H_0$**

No exemplo,  $P = 0,0127$ .

Adotando  $\alpha = 0,05$ , temos  $P < \alpha$  e, portanto, rejeitamos  $H_0$ .

Se tivermos  $x_{obs} = 12$ , qual será o nível descritivo?

$$P = P(X \leq 12 \mid p = 0,9) = 0,1840$$

Conclusão: não rejeitamos  $H_0$ .

**Nível Descritivo** é o menor nível de significância para o qual o resultado observado é significativo, ou seja, conduz à rejeição da hipótese nula  $H_0$ .

**Exemplo 2:** A diretoria de uma escola acredita que neste ano a proporção  $p$  de alunos usuários da Internet é maior que os 70% encontrados no ano anterior. Se uma pesquisa com 30 alunos, escolhidos ao acaso, mostrou que 26 são usuários da Internet, podemos concluir que a afirmação da diretoria é verdadeira?

Hipóteses estatísticas:

$$H: p = 0,7$$

$$A: p > 0,7$$



Seja

$X$ : nº de usuários em 30 entrevistados. Então

$$X \sim b(30, p).$$

Se  $x_{obs} = 26$  usuários da Internet, o *nível descritivo* do teste é dado por:

$$P = P(X \geq 26 \mid p = 0,7) = 0,0301 \quad \Rightarrow$$

**Conclusão:**

Se  $\alpha = 5\%$ ,  $P < \alpha \Rightarrow$  **rejeitamos  $H_0$** , ou seja, os dados indicam que a proporção de usuários aumentou (é maior que 70%).

# Uso da Aproximação Normal

**Exemplo 3:** Proporção de analfabetos em uma cidade  
Pelo Anuário IBGE/1988: 15% *de analfabetos*.

Vamos supor duas pesquisas:

Em 1998, entre 200 entrevistados, 27 eram analfabetos.

Em 2008, entre 200 entrevistados, 19 eram analfabetos.

**Pergunta:** Esses dados suportam a tese de diminuição do analfabetismo na cidade de 1988 para 1998? E de 1988 para 2008?

Em ambos os casos, o teste de interesse é das hipóteses:

$$H : p = 0,15$$

$$A : p < 0,15$$

sendo  $p$  a proporção populacional de analfabetos na cidade (no ano da pesquisa).

Seja

$X$ : nº de analfabetos em 200 cidadãos entrevistados (no ano da pesquisa).

Então,

$$X \sim b(200, p).$$

1) Se adotarmos  $RC = \{X \leq 23\}$

Conclusão:

Em 1998,  $x_{obs} = 27 \notin RC \Rightarrow$  Não rejeitamos  $H$

Em 2008,  $x_{obs} = 19 \in RC \Rightarrow$  Rejeitamos  $H$

Cálculo do nível de significância:

$$\begin{aligned}\alpha &= P(\text{rejeitar } H / H \text{ verdadeira}) \\ &= P(X \leq 23 \mid p = 0,15) \\ &= 0,0958\end{aligned}$$



Cálculo do tamanho do erro tipo II:

$$\begin{aligned}\beta(0,10) &= P(\text{não rejeitar } H \mid p = 0,10) \\ &= P(X > 23 \mid p = 0,10) \\ &= 0,2016\end{aligned}$$

## Usando aproximação normal:

$$\alpha = P(X \leq 23 \mid p = 0,15)$$

$$E(X) = np = 200 \times 0,15 = 30$$

$$\text{Var}(X) = np(1-p) = 200 \times 0,15 \times 0,85 = 25,5$$

$$DP(X) = \sqrt{25,5} = 5,05$$

$$\alpha \cong P\left(Z \leq \frac{23 - 30}{5,05}\right) = P(Z \leq -1,39) = 0,0823$$

$$\beta(0,10) = P(X \geq 24 \mid p = 0,10)$$

$$E(X) = np = 200 \times 0,10 = 20$$

$$\text{Var}(X) = np(1-p) = 200 \times 0,10 \times 0,90 = 18$$

$$DP(X) = \sqrt{18} = 4,24$$

$$\beta(0,10) \cong P\left(Z \geq \frac{24 - 20}{4,24}\right) = P(Z \geq 0,94) = 0,1746$$

## 2) Adotando o nível descritivo:

Em 1998,  $x_{obs} = 27 \Rightarrow \mathbf{P} = P(X \leq 27 \mid p = 0,15) = 0,3164$

Em 2008,  $x_{obs} = 19 \Rightarrow \mathbf{P} = P(X \leq 19 \mid p = 0,15) = 0,0148$

**Conclusão:** fixando  $\alpha = 5\%$ ,  
em 1998,  $P > \alpha \Rightarrow$  não rejeitamos  $H$   
em 2008,  $P < \alpha \Rightarrow$  rejeitamos  $H$

Usando aproximação:

Em 1998,  $\mathbf{P} = P(X \leq 27 \mid p = 0,15)$

$$\cong P\left(Z \leq \frac{27 - 30}{5,05}\right) = P(Z \leq -0,59) = 0,2776$$

Em 2008,  $\mathbf{P} = P(X \leq 19 \mid p = 0,15)$

$$\cong P\left(Z \leq \frac{19 - 30}{5,05}\right) = P(Z \leq -2,18) = 0,0146$$

**Exemplo 3:** Uma pesquisa considerou 92 voluntários para fazerem parte do estudo sendo que cada um deles jogava uma moeda. *O* voluntário seria submetido ou não a um *teste*, conforme o resultado fosse cara ou coroa. Os seguintes resultados foram obtidos:

	Voluntários
n° caras	35
n° coroas	57
Total	92

Com base nos resultados dos 92 lançamentos você diria que a moeda usada era honesta? Tome a decisão com base no nível descritivo.

Hipóteses estatísticas:

$$H: p = 0,5$$

$$A: p \neq 0,5$$

sendo  $p$  a probabilidade de *coroa* da moeda.

Seja

$X$ : nº de *coroas* em 92 lançamentos da moeda.

Então,  $X \sim b(92, p)$ .

Sob  $H$ ,  $p=0,5$  e  $E(X)=92 \times 0,5 = 46$ ,  $\text{Var}(X)=92 \times 0,5 \times 0,5 = 23$



Como a hipótese alternativa é bilateral, no cálculo do valor  $P$ , devemos considerar que a região crítica envolve os valores de  $X$  que se distanciam muito (para mais ou para menos) do valor previsto pela hipótese nula  $H$ .

Sob  $H$ , o valor esperado de  $X$  é 46.

Como  $x_{obs} = 57$ , que é maior que 46, calcula-se

$$P(X \geq 57 \mid p = 0,50)$$

Usando a aproximação normal :

$$\cong P\left(Z \geq \frac{57 - 46}{\sqrt{23}}\right) = P(Z \geq 2,29) = 0,011$$

e o nível descritivo será  $P = 2 \times 0,011 = 0,022$

*Para  $\alpha = 5\%$ ,  $P < \alpha \Rightarrow$  Rejeitamos  $H$*

*Para  $\alpha = 1\%$ ,  $P > \alpha \Rightarrow$  Não rejeitamos  $H$*

Se o valor de  $x_{obs}$  fosse menor que 46, por exemplo,  $x_{obs} = 34$ , então

$$P = 2P\left(Z \leq \frac{34 - 46}{\sqrt{23}}\right)$$

$$P = 2P(Z \leq -2,5) = 2 \times [1 - A(2,5)]$$

$$= 2 \times [1 - 0,9938] = 0,0124 .$$