

Testes de hipóteses

Wagner H. Bonat
Fernando P. Mayer
Elias T. Krainski

Universidade Federal do Paraná
Departamento de Estatística
Laboratório de Estatística e Geoinformação

04/06/2018



Sumário

- 1 Introdução
- 2 Teste para a Média Populacional
 - Variância conhecida
 - Variância desconhecida
- 3 Nível descritivo
- 4 Testes Qui-Quadrado
- 5 Exercícios recomendados

Inferência estatística

Na **inferência estatística** os dois principais objetivos são:

- **Estimar** um parâmetro populacional
 - Estimativa pontual
 - Estimativa intervalar
- **Testar** uma hipótese ou afirmativa sobre um parâmetro populacional

Testes de hipótese

Hipótese

É uma afirmativa sobre uma propriedade da população

Teste de hipótese

- É um procedimento para se testar uma afirmativa sobre uma propriedade da população
- Permite tomar **decisões** sobre a população com base em informações de dados amostrais.

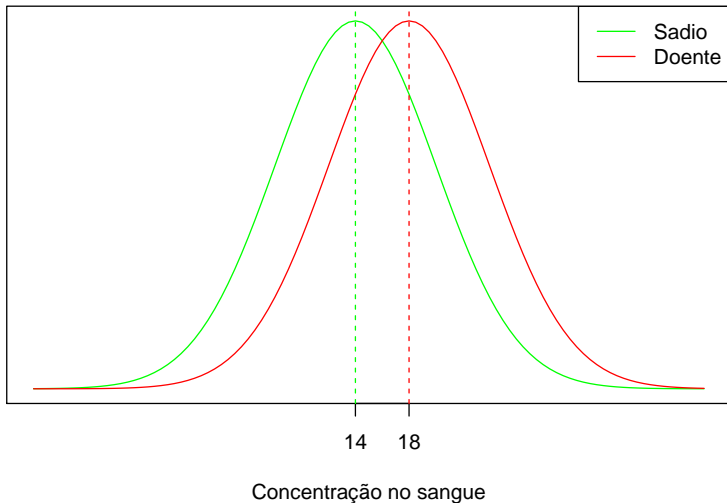
Exemplo 8.1

Suponha que, entre pessoas saudáveis, a concentração de certa substância no sangue se comporta segundo um modelo Normal com média 14 unidades/ml e desvio padrão 6 unidades/ml.

Pessoas sofrendo de uma doença específica tem concentração média da substância alterada para 18 unidades/ml.

Admitimos que o modelo Normal com desvio padrão 6 unidades/ml, continua representado de forma adequada a concentração da substância em pessoas com a doença.

Exemplo 8.1



Exemplo 8.1

Para averiguar se um tratamento é eficaz contra a doença, selecionamos uma amostra de 30 indivíduos submetidos ao tratamento.

Assumimos que todos os elementos da amostra X_1, \dots, X_{30} possuem a mesma distribuição: $X_i \sim N(\mu, 36)$, onde:

- $\mu = 14$ se o tratamento for eficiente
- $\mu = 18$ se o tratamento não for eficiente

Se a média da amostra for próxima de 14, temos **evidências** de que o tratamento é eficaz. Se for mais próxima de 18, as **evidências** são contrárias ao tratamento.

Então a pergunta é: o quão próximo é “próximo”?

Exemplo 8.2

Deseja-se estudar a tolerância de um equipamento eletrônico ao número de impactos termo-elétricos.

Pelas características de fabricação do equipamento, é possível admitir que a probabilidade de falha seja constante, isto é, após cada impacto, existe uma probabilidade p que ele falhe.

Representando por X a variável número de impactos anteriores à falha, pretende-se verificar se o modelo Geométrico com $p = 0.4$ se adequa para caracterizar essa variável.

Sumário

- 1 Introdução
- 2 Teste para a Média Populacional
 - Variância conhecida
 - Variância desconhecida
- 3 Nível descritivo
- 4 Testes Qui-Quadrado
- 5 Exercícios recomendados

Exemplo 8.1 continuação

- Interesse geral $\mu = 14$?
- Distribuição da média amostral para $n = 30$: $N(\mu, 36/30)$.
- Critério para decidir sobre o valor de μ .
- **Valor crítico**, digamos x_c tal que se a média amostral (\bar{x}_{obs}) for maior que x_c concluimos que a amostra pertence a população com média 18.
- Como \bar{X} é uma variável aleatória, devem existir erros associados.

Tipos de hipóteses

Hipótese nula H_0

É uma afirmativa de que o valor de um parâmetro populacional é **igual** a algum valor especificado. (O termo *nula* é usado para indicar nenhuma mudança, nenhum efeito). Ex.:

- $\mu = 170$ cm
- $p = 0,5$

Hipótese alternativa H_a

É uma afirmativa de que o parâmetro tem um valor, que, de alguma forma, difere da hipótese nula. Ex.:

- $\mu \neq 170$ $\mu < 170$ $\mu > 170$

Tipos de hipóteses

Quando fazemos um teste de hipótese, chegamos a um dos dois possíveis resultados:

- **Rejeitar** H_0 : em favor da hipótese alternativa H_a
- **Não rejeitar** H_0 : e conclui-se que não existem diferenças

Atenção!

- O termo **aceitar** a hipótese nula é filosoficamente incorreto, pois não se pode aceitar uma hipótese baseada apenas em evidências amostrais (mesmo em um teste de hipótese formal).
- E ainda existe um **erro** associado a todo teste de hipótese

Hipóteses

Hipótese simples:

- H_0 : O tratamento não é eficaz ($\mu = 18$)
- H_a : O tratamento é eficaz ($\mu = 14$)

Hipóteses compostas:

- Hipótese unilateral à esquerda
 - H_0 : O tratamento não é eficaz ($\mu = 18$);
 - H_1 : O tratamento é eficaz ($\mu < 18$).
- Hipótese bilateral:
 - H_0 : O tratamento não é eficaz ($\mu = 18$);
 - H_1 : O tratamento é eficaz ($\mu \neq 18$).

Erros ao realizar um teste de hipótese

- **Erro Tipo I:** rejeitar H_0 , quando H_0 é verdadeira.
- **Erro Tipo II:** não rejeitar H_0 quando H_0 é falsa.

| | H_0 verdadeira | H_0 falsa |
|--------------------|------------------|-----------------|
| Não rejeitar H_0 | Decisão correta | Erro tipo II |
| Rejeitar H_0 | Erro tipo I | Decisão correta |

Erros ao realizar um teste de hipótese

Definimos por α e β as probabilidades de cometer os erros do tipo I e II:

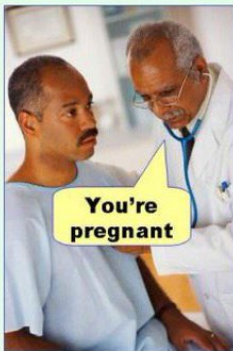
- $\alpha = P(\text{erro tipo I}) = P(\text{rejeitar } H_0 \mid H_0 \text{ verdadeira})$
- $\beta = P(\text{erro tipo II}) = P(\text{não rejeitar } H_0 \mid H_0 \text{ falsa})$

No exemplo 8.1, se $H_0 : \mu = 18$ e $H_a : \mu < 18$, então:

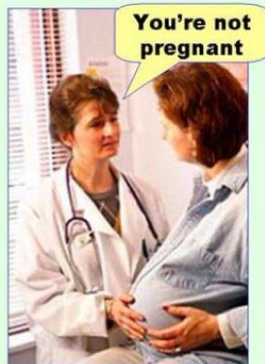
- $\alpha = P(\text{concluir que o tratamento é eficaz quando na verdade não é})$
- $\beta = P(\text{concluir que o tratamento não é eficaz quando na verdade é})$

Erros ao realizar um teste de hipótese

Type I error
(false positive)



Type II error
(false negative)



Erros ao realizar um teste de hipótese

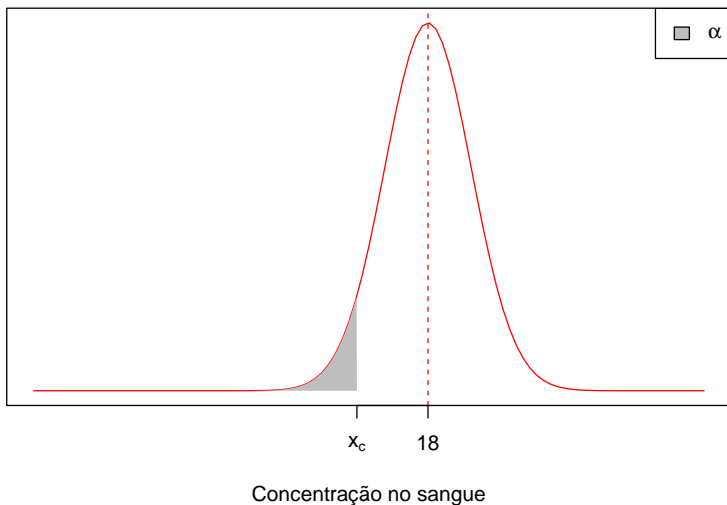
A situação ideal é aquela em que ambas as probabilidades, α e β , são próximas de zero.

No entanto, à medida que diminuimos α , a probabilidade β tende a aumentar.

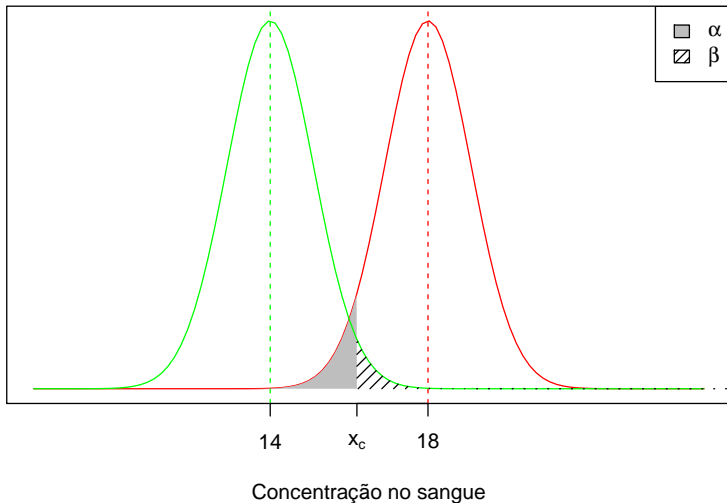
Levando isso em conta, ao formular as hipóteses, **devemos cuidar para que o erro mais importante a ser evitado seja o erro do tipo I.**

Por isso, a probabilidade α recebe o nome de **nível de significância** do teste, e é esse erro que devemos controlar.

Erros ao realizar um teste de hipótese



Erros ao realizar um teste de hipótese



Valor crítico

Supondo α conhecido podemos determinar o valor crítico x_c .

$$\begin{aligned}\alpha &= P(\text{erro tipo I}) = P(\text{rejeitar } H_0 \mid H_0 \text{ verdadeira}) \\ &= P(\bar{X} < x_c \mid \mu = 18) = P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < \frac{x_c - 18}{6/\sqrt{30}}\right) \\ &= P(Z < z_c)\end{aligned}$$

com $Z \sim N(0, 1)$.

Obtendo o valor crítico

Dado α encontramos z_c na tabela normal padrão.

Obtemos x_c

$$z_c = \frac{x_c - 18}{6/\sqrt{30}} \Rightarrow x_c = 18 + z_c \frac{6}{\sqrt{30}}$$

Supondo $\alpha = 0.05$ temos

$$0.05 = P(Z < z_c) \Rightarrow z_c = -1.64$$

logo

$$x_c = 18 - 1.64 \frac{6}{\sqrt{30}} = 16.2$$

Região Crítica

Dada uma amostra, se $\bar{x}_{obs} < 16.2$, **rejeitamos** H_0 , concluindo que o tratamento é eficaz.

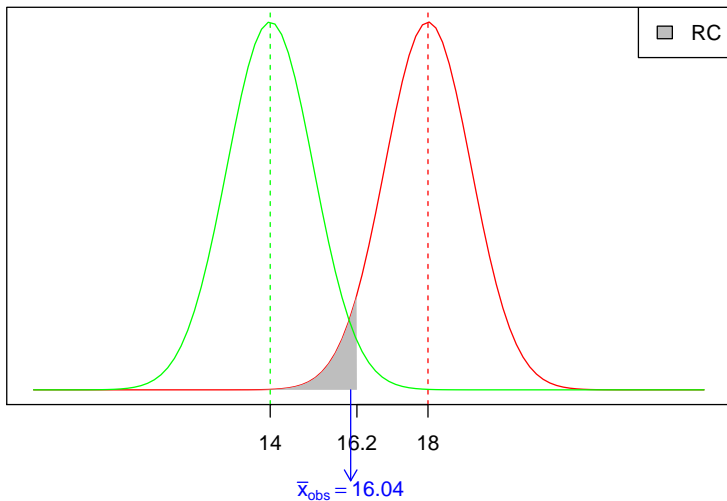
O conjunto dos números reais menores que 16.2 é denominado de **Região de Rejeição** ou **Região Crítica** (RC), isto é:

$$RC = \{x \in \mathbb{R} : x < 16.2\}.$$

No exemplo 8.1, se a média amostral dos 30 indivíduos foi $\bar{x}_{obs} = 16.04$, então **rejeitamos** H_0 , ao nível de significância $\alpha = 0.05$.

Nesse caso, $\bar{x}_{obs} < x_c$ está dentro da RC.

Região Crítica



Teste de hipótese bilateral

Definindo as hipóteses

$$H_0 : \mu = \mu_0 \quad \text{e} \quad H_a : \mu \neq \mu_0$$

A Região Crítica será dada por

$$RC = \{x \in \mathbb{R} \mid x < x_{c1} \quad \text{ou} \quad x > x_{c2}\}$$

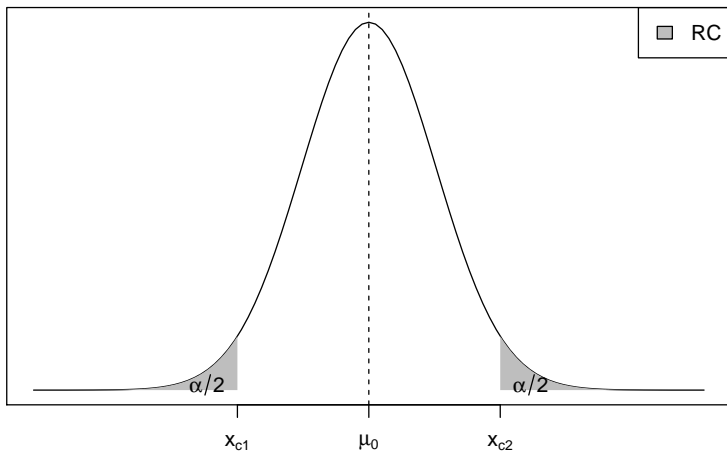
Para um valor de α fixado, determinamos x_{c1} e x_{c2} de modo que

$$P(\bar{X} < x_{c1} \cup \bar{X} > x_{c2}) = \alpha$$

Assim, distribuimos a área α igualmente entre as duas partes da RC

$$P(\bar{X} < x_{c1}) = \frac{\alpha}{2} \quad \text{e} \quad P(\bar{X} > x_{c2}) = \frac{\alpha}{2}$$

Teste de hipótese bilateral



Etapas de um teste de hipótese

- 1 Estabelecer as hipóteses nula e alternativa.
- 2 Definir a forma da região crítica, com base na hipótese alternativa.
- 3 Identificar a distribuição do estimador e obter sua estimativa.
- 4 Fixar α e obter a região crítica.
- 5 Concluir o teste com base na estimativa e na região crítica.

Exemplo 8.3

Um pesquisador deseja estudar o efeito de certa substância no tempo de reação de seres vivos a um certo tipo de estímulo.

Um experimento é desenvolvido com cobaias que são inoculadas com a substância e submetidas a um estímulo elétrico, com seus tempos de reação (em segundos) anotados. Os seguintes valores foram obtidos: 9.1; 9.3; 7.2; 7.5; 13.3; 10.9; 7.2; 9.9; 8.0; 8.6.

Admite-se que o tempo de reação segue o modelo Normal com média 8 e desvio padrão $\sigma = 2$ segundos. O pesquisador desconfia que o tempo médio sofre alteração por influência da substância.

Efetue um teste de hipótese para verificar se o pesquisador tem razão. Use $\alpha = 0.06$, e calcule a probabilidade do erro tipo II supondo que $\mu = 9$.

Exemplo 8.4

Um relatório de uma companhia afirma que 40% de toda a água obtida, através de poços artesianos no nordeste, é salobra.

Há muitas controvérsias sobre essa afirmação, alguns dizem que a proporção é maior, outros que é menor.

Para dirimir as dúvidas, 400 poços foram sorteados e observou-se, em 120 deles água salobra. Qual seria a conclusão, ao nível de 3%?

Teste para a média com variância desconhecida

Já vimos (Cap. 7) que quando não conhecemos σ^2 , usamos como estimativa a variância amostral

$$S^2 = \frac{1}{n-1} \left(\sum_{i=1}^n X_i^2 - n\bar{X}^2 \right),$$

e que a variável padronizada

$$T = \frac{\bar{X} - \mu}{\sqrt{S^2/n}} = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

segue uma distribuição *t-Student* com $n - 1$ graus de liberdade, ou seja,

$$T \sim t_{(n-1)}$$

Exemplo 8.5

Deseja-se investigar se uma certa moléstia que ataca o rim altera o consumo de oxigênio desse órgão. Para indivíduos sadios, admite-se que esse consumo tem distribuição Normal com média $12 \text{ cm}^3/\text{min}$.

Os valores medidos em cinco pacientes com a moléstia foram: 14.4; 12.9; 15.1; 13.7; 13.5.

Qual seria a conclusão, ao nível de 1% de significância?

Sumário

- 1 Introdução
- 2 Teste para a Média Populacional
 - Variância conhecida
 - Variância desconhecida
- 3 Nível descritivo**
- 4 Testes Qui-Quadrado
- 5 Exercícios recomendados

Nível descritivo

Em geral, α é pré-fixado para construir a regra de decisão.

Uma alternativa é deixar em aberto a escolha de α para quem for tomar a decisão.

A ideia é calcular, **supondo que a hipótese nula é verdadeira**, a probabilidade de se obter estimativas mais extremas do que aquela fornecida pela amostra.

Essa probabilidade é chamada de **nível descritivo**, denotada por α^* (ou P -valor).

Valores pequenos de α^* evidenciam que a hipótese nula é falsa.

O conceito de “pequeno” fica para quem decide qual α deve usar para comparar com α^* .

Nível descritivo

Para **testes unilaterais**, sendo $H_0 : \mu = \mu_0$, a expressão de α^* depende da hipótese alternativa:

$$\alpha^* = P(\bar{X} < \bar{x}_{obs} \mid H_0 \text{ verdadeira}) \quad \text{para } H_a : \mu < \mu_0$$

$$\alpha^* = P(\bar{X} > \bar{x}_{obs} \mid H_0 \text{ verdadeira}) \quad \text{para } H_a : \mu > \mu_0$$

Para **testes bilaterais**, temos $H_0 : \mu = \mu_0$ contra $H_0 : \mu \neq \mu_0$, a definição do nível descritivo depende da relação entre \bar{x}_{obs} e μ_0 :

$$\alpha^* = 2 \times P(\bar{X} < \bar{x}_{obs} \mid H_0 \text{ verdadeira}) \quad \text{se } \bar{x}_{obs} < \mu_0$$

$$\alpha^* = 2 \times P(\bar{X} > \bar{x}_{obs} \mid H_0 \text{ verdadeira}) \quad \text{se } \bar{x}_{obs} > \mu_0$$

Como estamos calculando a probabilidade para apenas uma das caudas, então esse valor é multiplicado por 2.

Exemplo 8.7

Uma associação de defesa do consumidor desconfia que embalagens de 450 gramas de um certo tipo de biscoito estão abaixo do peso.

Para verificar tal afirmação, foram coletados ao acaso 80 pacotes em vários supermercados, obtendo-se uma média de peso de 447 gramas.

Admitindo-se que o peso dos pacotes segue o modelo Normal com desvio padrão 10 gramas, que conclusão pode ser tirada através do nível descritivo?

Exemplo 8.8

Voltando ao exemplo 8.3, temos:

$$H_0 : \mu = 8 \quad \text{e} \quad H_a : \mu \neq 8$$

Com uma amostra de tamanho 10, observou-se $\bar{x}_{obs} = 9.1$, e sabe-se que $\sigma = 2$. Assim,

$$\begin{aligned}\alpha^* &= 2 \times P(\bar{X} > \bar{x}_{obs} \mid H_0 \text{ verdadeira}) \\ &= 2 \times P(\bar{X} > 9.1 \mid \mu = 8) \\ &= 2 \times P(Z > 1.74) \\ &= 0.0818 \approx 8.18\%\end{aligned}$$

Qual a conclusão ao nível de 5%? E de 10%?

Sumário

- 1 Introdução
- 2 Teste para a Média Populacional
 - Variância conhecida
 - Variância desconhecida
- 3 Nível descritivo
- 4 Testes Qui-Quadrado**
- 5 Exercícios recomendados

Introdução

Os testes que utilizam o modelo **Qui-Quadrado** como estrutura probabilística são denominados genericamente de **Testes Qui-Quadrado**.

Esse mesmo modelo pode ser utilizado para testes com finalidades diferentes, por exemplo:

- **Teste de aderência:** verifica se uma variável segue uma determinada distribuição
- **Teste de independência:** verifica a independência entre duas variáveis
- **Teste de homogeneidade:** verifica se uma variável se comporta de maneira similar em várias subpopulações

Testes Qui-Quadrado

De maneira geral, qualquer teste Qui-Quadrado envolve a quantidade

$$Q^2 = \sum_{i,j} \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

que mede as discrepâncias entre **frequências observadas** (o_{ij}) e as **frequências esperadas** (e_{ij}). Pode-se mostrar que Q^2 possui distribuição Qui-Quadrado com ν **graus de liberdade**, ou seja,

$$Q^2 \sim \chi^2_\nu$$

onde ν depende do teste utilizado.

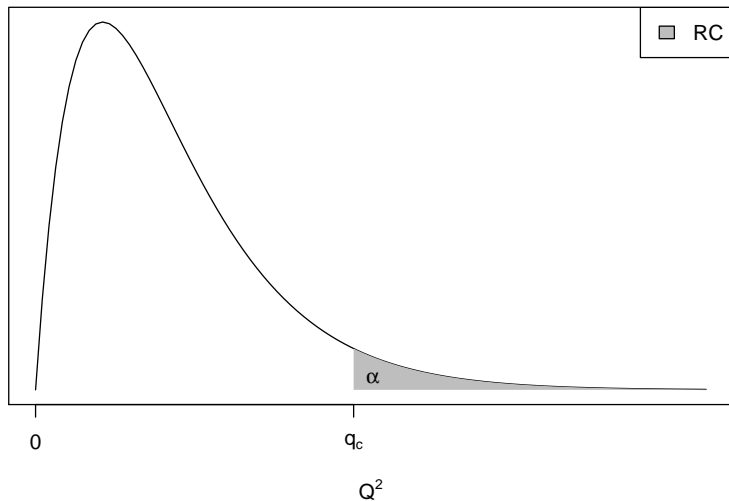
A **Região Crítica** (RC) é constituída de valores grandes de Q^2 ,

$$RC = \{\omega : \omega \geq q_c\}$$

com q_c sendo determinado pelo nível de significância do teste

$$\alpha = P(Q^2 \geq q_c \mid H_0 \text{ verdadeira})$$

Distribuição Qui-Quadrado



Exemplo 8.9 (teste de aderência)

No exemplo 8.2, definimos X como sendo o número de impactos anteriores à falha em um equipamento eletrônico.

Uma amostra de 80 ensaios foi obtida, cada ensaio representando os testes feitos até a interrupção por falha no equipamento, resultado 80 observações da variável de interesse.

Pretende-se verificar se o modelo Geométrico com $p = 0.4$ é adequado.

Exemplo 8.10 (teste de independência)

A tabela abaixo contém os resultados obtidos por estudantes do ensino médio, em um exame com questões nas disciplinas de física e matemática. Para efeito de apresentação na tabela e análise de comportamento, as notas foram classificadas nas categorias alta, média e baixa.

| F/M | Alta | Média | Baixa | Sum |
|-------|------|-------|-------|-----|
| Alta | 56 | 71 | 12 | 139 |
| Média | 47 | 163 | 38 | 248 |
| Baixa | 14 | 42 | 85 | 141 |
| Sum | 117 | 276 | 135 | 528 |

Deseja-se testar se existe dependência entre as notas dessas duas disciplinas.

Exemplo 8.10 (teste de independência)

Vamos testar as hipóteses:

H_0 : As notas de física e matemática são independentes

H_a : As notas não são independentes

Para isso, construímos uma tabela de valores esperados para a casela i, j :

$$e_{i,j} = \frac{\text{Total da linha } i \times \text{Total da coluna } j}{\text{Total geral}}$$

E assim, obtemos

| | Alta | Média | Baixa |
|-------|-------|--------|-------|
| Alta | 30.80 | 72.66 | 35.54 |
| Média | 54.95 | 129.64 | 63.41 |
| Baixa | 31.24 | 73.70 | 36.05 |

Exemplo 8.10 (teste de independência)

Calculamos

$$q_{obs}^2 = \frac{(56 - 30.8)^2}{30.8} + \dots + \frac{(85 - 36.05)^2}{36.05} = 145.78$$

Nesse caso, Q^2 possui distribuição Qui-Quadrado com $(r - 1) \times (s - 1)$ graus de liberdade (r linhas e s colunas). Para $\alpha = 0.01$

$$P(Q^2 \geq q_c \mid H_0) = \alpha \Rightarrow P(Q^2 \geq q_c \mid H_0) = 0.01$$

obtemos $q_c = 13.28$ para χ_4^2 , e portanto,

$$RC = \{\omega : \omega \geq 13.28\}.$$

Como $q_{obs}^2 \in RC$, concluímos por rejeitar a hipótese nula (de independência entre as notas).

Sumário

- 1 Introdução
- 2 Teste para a Média Populacional
 - Variância conhecida
 - Variância desconhecida
- 3 Nível descritivo
- 4 Testes Qui-Quadrado
- 5 Exercícios recomendados

Exercícios recomendados

- Seção 8.2 - 1 a 6.
- Seção 8.3 - 1 a 6.
- Seção 8.4 - 1 a 4.
- Seção 8.5 - 1 a 5.