Intervalo de confiança

Diferença entre duas médias e diferença entre duas proporções

Caio S. R. Laytynher, Thauan N. J. Alves

16 de dezembro de 2021

Universidade Estadual de Santa Cruz - UESC
Ilhéus - BA
DCET - 173: Probabilidade e estatística
2º Semestre

Passos

Passos

1. Determinar o nível de significância α;

Passos

1. Determinar o nível de significância α;
2. Calcular o desvio padrão combinado;

Passos

1. Determinar o nível de significância α;
2. Calcular o desvio padrão combinado;
3. Se a variância for conhecida:
    3.1 Determinar o valor $z_{\frac{\alpha}{2}}$ crítico.

Passos

1. Determinar o nível de significância α;
2. Calcular o desvio padrão combinado;
3. Se a variância for conhecida:
   3.1 Determinar o valor $z_{\frac{\alpha}{2}}$ crítico.
4. Caso contrário
   4.1 Calcular os graus de liberdade (ϕ);
   4.2 Determinar o valor $t_{\frac{\alpha}{2},\phi}$ crítico.

Passos

1. Determinar o nível de significância α;
2. Calcular o desvio padrão combinado;
3. Se a variância for conhecida:
   3.1 Determinar o valor $z_{\frac{\alpha}{2}}$ crítico.
4. Caso contrário
   4.1 Calcular os graus de liberdade (ϕ);
   4.2 Determinar o valor $t_{\frac{\alpha}{2},\phi}$ crítico.
5. Calcular a margem de erro (ME);

Passos

1. Determinar o nível de significância α;
2. Calcular o desvio padrão combinado;
3. Se a variância for conhecida:
   3.1 Determinar o valor $z_{\frac{\alpha}{2}}$ crítico.
4. Caso contrário
   4.1 Calcular os graus de liberdade (ϕ);
   4.2 Determinar o valor $t_{\frac{\alpha}{2},\phi}$ crítico.
5. Calcular a margem de erro (ME);
6. Determinar o intervalo de confiança (IC).

Passos

1. Determinar o nível de significância α;
2. Calcular o desvio padrão combinado;
3. Se a variância for conhecida:
   3.1 Determinar o valor $z_{\frac{\alpha}{2}}$ crítico.
4. Caso contrário
   4.1 Calcular os graus de liberdade (ϕ);
   4.2 Determinar o valor $t_{\frac{\alpha}{2},\phi}$ crítico.
5. Calcular a margem de erro (ME);
6. Determinar o intervalo de confiança (IC).

OBS: Testes para intervalo de confiança são sempre bilaterais, por isso $\frac{\alpha}{2}$.

Passos

Passos

1. Determinar o nível de significância α;

Passos

1. Determinar o nível de significância α;
2. Calcular para ambas amostras:
    2.1 Média: $\bar{x}$;
    2.2 Desvio padrão: s, ou variância: s^2;
    2.3 Quantidade de dados: n.

Passos

1. Determinar o nível de significância α;
2. Calcular para ambas amostras:
    2.1 Média: $\bar{x}$;
    2.2 Desvio padrão: s, ou variância: s^2;
    2.3 Quantidade de dados: n.
3. Verificar se as variâncias podem ser consideradas significativamente iguais a partir do **teste** F, que pode ser feito computacionalmente pela função `var.test()`.

Se forem consideradas iguais

Se forem consideradas iguais

Desvio padrão da diferença

$$s_c = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

$$s_{\bar{x}_1 - \bar{x}_2} = s_c \cdot \sqrt{\frac{n_1 + n_2}{n_1 \cdot n_2}}$$

Se forem consideradas iguais

Desvio padrão da diferença

$$s_c = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

$$s_{\bar{x}_1 - \bar{x}_2} = s_c \cdot \sqrt{\frac{n_1 + n_2}{n_1 \cdot n_2}}$$

Graus de liberdade

$$\phi = n_1 + n_2 - 2$$

Se não forem consideradas iguais

Se não forem consideradas iguais

Desvio padrão da diferença

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Se não forem consideradas iguais

Desvio padrão da diferença

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Graus de liberdade

$$\phi = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2}$$

Se não forem consideradas iguais

Desvio padrão da diferença

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Graus de liberdade

$$\phi = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2}$$

- O $t_{\frac{\alpha}{2},\phi}$ pode ser obtido por meio de tabelas, calculadoras científicas ou a função `qt()`;

Margem de erro

$$ME = t_{\frac{\alpha}{2},\phi} \cdot s_{\bar{x}_1 - \bar{x}_2}$$

Diferença entre duas médias - Teste t

Margem de erro

$$ME = t_{\frac{\alpha}{2},\phi} \cdot s_{\bar{x}_1 - \bar{x}_2}$$

Intervalo de confiança

$$(\bar{x}_1 - \bar{x}_2) - ME \leq \mu_0 \leq (\bar{x}_1 - \bar{x}_2) + ME$$

Método computacional - Possui acesso aos vetores das amostras

- Determinar se as variâncias são significativamente iguais utilizando `var.test()`;
- A função retorna o p-valor, que pode ser utilizado para rejeitar ou aceitar a hipótese da igualdade, H_0.

```
var.test(x,             # numeric
         y,             # numeric
         conf.level)    # numeric
```

Método computacional - Possui os vetores com os dados

- A função `t.test()`, dentre diversas informações, retorna o intervalo de confiança.

```
t.test(x, y, conf.level=0.95)
```

Método computacional - Não possui acesso aos vetores das amostras

- Calcula-se a razão entre as variâncias para a realização de um teste F. A função `pf()` pode ser utilizada para determinar o p-valor;
- As contas devem ser feitas previamente para determinar ϕ e $s_{\bar{x}_1 - \bar{x}_2}$;
- A partir dos valores de α e ϕ, calcula-se o t crítico utilizando a função `qt()`;
- Os resultados são utilizados para calcular ME e IC.

```
qt(p,              # numeric
   df,             # numeric
   lower.tail)     # boolean
```

Características

Características

- Utilizado para amostras relacionadas entre si;

Características

- Utilizado para amostras relacionadas entre si;
- Geralmente a amostra é a mesma medida em momentos diferentes;

Características

- Utilizado para amostras relacionadas entre si;
- Geralmente a amostra é a mesma medida em momentos diferentes;
- O tamanho das amostras é o mesmo;

Passos

Passos

1. Determinar α;

Passos

1. Determinar α;
2. Tirar a diferença entre cada dado de ambas as amostras;

Passos

1. Determinar α;
2. Tirar a diferença entre cada dado de ambas as amostras;
3. Calcular a média das diferenças;

Passos

1. Determinar α;
2. Tirar a diferença entre cada dado de ambas as amostras;
3. Calcular a média das diferenças;
4. Calcular o desvio padrão das diferenças;

Passos

1. Determinar α;
2. Tirar a diferença entre cada dado de ambas as amostras;
3. Calcular a média das diferenças;
4. Calcular o desvio padrão das diferenças;
5. Calcular os graus de liberdade;

Passos

1. Determinar α;
2. Tirar a diferença entre cada dado de ambas as amostras;
3. Calcular a média das diferenças;
4. Calcular o desvio padrão das diferenças;
5. Calcular os graus de liberdade;
6. O $t_{\frac{\alpha}{2},\phi}$ e o intervalo de confiança são calculados da mesma maneira que a anterior, entretanto a margem de erro tem uma pequena diferença.

Diferença entre duas médias - Teste t pareado

Média das diferenças

$$\bar{d} = \frac{\sum_{i=0}^{n} d_i}{n}$$

Diferença entre duas médias - Teste t pareado

Média das diferenças

$$\bar{d} = \frac{\sum_{i=0}^{n} d_i}{n}$$

Desvio padrão das diferenças

$$s_d = \sqrt{\frac{\sum_{i=0}^{n}(d_i - \bar{d})^2}{n-1}}$$

Diferença entre duas médias - Teste t pareado

Média das diferenças

$$\bar{d} = \frac{\sum_{i=0}^{n} d_i}{n}$$

Desvio padrão das diferenças

$$s_d = \sqrt{\frac{\sum_{i=0}^{n} (d_i - \bar{d})^2}{n-1}}$$

Graus de liberdade

$$\phi = n - 1$$

Diferença entre duas médias - Teste t pareado

Margem de erro

$$ME = t_{\frac{\alpha}{2},\phi} \cdot \frac{s_d}{\sqrt{n}}$$

Diferença entre duas médias - Teste t pareado

Margem de erro

$$ME = t_{\frac{\alpha}{2},\phi} \cdot \frac{s_d}{\sqrt{n}}$$

Intervalo de confiança

$$(\bar{x}_1 - \bar{x}_2) - ME \leq \mu_0 \leq (\bar{x}_1 - \bar{x}_2) + ME$$

Método computacional - Possui acesso aos vetores das amostras

- A função `t.test()` possui um parâmetro chamado `paired`, que pode ser alterado para `TRUE` para indicar a realização de um teste pareado.

```r
t.test(x, y, alternative="greater", paired=TRUE)
```

Método computacional - Não possui acesso aos vetores das amostras

- Os cálculos devem ser feitos previamente para determinar o valor de ϕ e de s_d;
- A partir daí, pode-se encontrar o valor de $t_{\frac{\alpha}{2},\phi}$ pela função `qt()`;
- A partir dos resultados, pode-se calcular ME e determinar IC.

```
alpha = 0.05
phi = 15

qt(alpha / 2, phi, lower.tail=F)
    # [1] 2.13145
```

Passos

Passos

1. Determinar α;

Passos

1. Determinar α;
2. Determinar para ambas amostras:
    2.1 Média: $\bar{x}$;
    2.2 Desvio padrão: σ, ou variância σ^2;
    2.3 Quantidade de dados: n.

Passos

1. Determinar α;
2. Determinar para ambas amostras:
    2.1 Média: $\bar{x}$;
    2.2 Desvio padrão: σ, ou variância σ^2;
    2.3 Quantidade de dados: n.
3. Calcular desvio padrão combinado;

Passos

1. Determinar α;
2. Determinar para ambas amostras:
    2.1 Média: $\bar{x}$;
    2.2 Desvio padrão: σ, ou variância σ^2;
    2.3 Quantidade de dados: n.
3. Calcular desvio padrão combinado;
4. Achar $z_{\frac{\alpha}{2}}$.

Diferença entre duas médias - Teste z

Desvio padrão combinado

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Desvio padrão combinado

$$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

- O $z_{\frac{\alpha}{2}}$ pode ser obtido pela função `qnorm()`;

Diferença entre duas médias - Teste z

Margem de erro

$$ME = z_{\frac{\alpha}{2}} \cdot \sigma_{\bar{x}_1 - \bar{x}_2}$$

Diferença entre duas médias - Teste z

Margem de erro

$$ME = z_{\frac{\alpha}{2}} \cdot \sigma_{\bar{x}_1 - \bar{x}_2}$$

Intervalo de confiança

$$(\bar{x}_1 - \bar{x}_2) - ME \leq \mu_0 \leq (\bar{x}_1 - \bar{x}_2) + ME$$

Método computacional

- É necessário realizar as contas previamente para determinar o valor de $\sigma_{\bar{x}_1 - \bar{x}_2}$;
- O $z_{\frac{\alpha}{2}}$ é dado pela função `qnorm()`, que funciona da mesma maneira que a `qt()`, porém para a distribuição normal;
- A partir desses valores é possível determinar *ME* e *IC*.

```
alpha = 0.05

qnorm(alpha / 2, lower.tail=F)
     # [1] 1.959964
```

Diferença entre duas médias - Exemplo 1

A tabela a seguir contém dados sobre 10 participantes do 7º exame da coorte de descendência no Framingham Heart Study.

	Homens			Mulheres		
Características	n	$\bar{x}$	s	n	$\bar{x}$	s
Pressão sanguínea sistólica	6	117.5	9.7	4	126.8	12.0
Pressão sanguínea diastólica	6	72.5	7.1	4	69.5	8.1
Colesterol sérico total	6	193.8	30.2	4	215.0	48.8
Peso	6	196.9	26.9	4	146.0	7.2
Altura	6	70.2	1.0	4	62.6	2.3
Índice de massa corporal	6	28.0	3.6	4	26.2	2.0

Diferença entre duas médias - Exemplo 1

Construa um intervalo de confiança de 95% para a diferença entre as médias de **pressão sanguínea sistólica** entre homens e mulheres utilizando estes dados.

Dados:

- $\bar{x}_1 = 117{,}5$
- $\bar{x}_2 = 126{,}8$
- $s_1 = 9{,}7$
- $s_2 = 12{,}0$
- $n_1 = 6$
- $n_2 = 4$
- $\alpha = 0{,}05$

Diferença entre duas médias - Exemplo 1

Teste F

$$F_c = \frac{s_1^2}{s_2^2} = \frac{9{,}7^2}{12{,}0^2} = 0{,}65$$

$$p\text{-valor} = 0{,}61$$

$s_{\bar{x}_1 - \bar{x}_2}$ para igualdade de variâncias

$$s_c = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

$$\implies s_c = \sqrt{\frac{(6-1)9{,}7^2 + (4-1)12{,}0^2}{6+4-2}} = 10{,}62$$

Diferença entre duas médias - Exemplo 1

$s_{\bar{x}_1-\bar{x}_2}$ para igualdade de variâncias

$$s_{\bar{x}_1-\bar{x}_2} = s_c \cdot \sqrt{\frac{n_1 + n_2}{n_1 \cdot n_2}} = 10{,}62 \cdot \sqrt{\frac{6+4}{6 \cdot 4}} = 6{,}86$$

ϕ para igualdade de variâncias

$$\phi = n_1 + n_2 - 2 = 6 + 4 - 2 = 8$$

- $t_{\frac{\alpha}{2},\phi} = t_{0{,}025;8} = 2{,}31$

Margem de erro

$$ME = t_{\frac{\alpha}{2},\phi} \cdot s_{\bar{x}_1-\bar{x}_2} = t_{0{,}025;8} \cdot 6{,}86 = 15{,}81$$

Diferença entre duas médias - Exemplo 1

Intervalo de confiança

$$(\bar{x}_1 - \bar{x}_2) - ME \leq \mu_0 \leq (\bar{x}_1 - \bar{x}_2) + ME$$

$$(117{,}5 - 126{,}8) - 15{,}81 \leq \mu_0 \leq (117{,}5 - 126{,}8) + 15{,}81$$

$$-25{,}11 \leq \mu_0 \leq 6{,}51$$

Passos

Passos

1. Determinar α;

Passos

1. Determinar α;
2. Calcular ambas proporções;

Passos

1. Determinar α;
2. Calcular ambas proporções;
3. Achar $z_{\frac{\alpha}{2}}$, ME e IC pelo mesmo método já descrito anteriormente.

Diferença entre duas proporções - Teste z

Desvio padrão da diferença

$$\sigma_{p_1-p_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

Diferença entre duas proporções - Teste z

Desvio padrão da diferença

$$\sigma_{p_1 - p_2} = \sqrt{\frac{p_1(1 - p_1)}{n_1} + \frac{p_2(1 - p_2)}{n_2}}$$

Margem de erro

$$ME = z_{\frac{\alpha}{2}} \cdot \sigma_{p_1 - p_2}$$

Diferença entre duas proporções - Teste z

Desvio padrão da diferença

$$\sigma_{p_1-p_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

Margem de erro

$$ME = z_{\frac{\alpha}{2}} \cdot \sigma_{p_1-p_2}$$

Intervalo de confiança

$$(p_1 - p_2) - ME \leq \pi_0 \leq (p_1 - p_2) + ME$$

Método computacional

- O R possui a função `prop.test()`;
- São fornecidos a frequência de elementos pertencentes a cada grupo analisado bem como o tamanho total de ambas amostras;
- Para os propósitos deste curso, `correct` será sempre definido como `FALSE`.

```
prop.test(x,                # numeric vector
          n,                # numeric vector
          conf.level,       # numeric
          correct=FALSE)
```

Método computacional

- Os valores são passados na forma de vetor;
- Dentre diversas informações, a função retorna o intervalo de confiança.

```r
freq_vec = c(13, 17)
n_vec = c(20, 23)

prop.test(x=freq_vec,
          n=n_vec,
          correct=FALSE)
```

Diferença entre duas proporções - Exemplo 2

Um pesquisador da área médica conjectura que fumar pode resultar em rugas ao redor dos olhos. O pesquisador recrutou 150 fumantes e 250 não fumantes para participar em um estudo observacional e constatou que 95 dos fumantes e 105 dos não fumantes vieram a ter rugas proeminentes ao redor dos olhos (baseado em uma pontuação de rugas padronizada administrada por uma pessoa que não sabia quem era ou não fumante.). Alguns resultados do estudo se encontram na tabela a seguir.

	Fumantes	**Não fumantes**
Tamanho da amostra	150	250
Proporção amostral de rugas proeminentes	$\frac{95}{150} = 0.63$	$\frac{105}{250} = 0.42$
Desvio padrão	0.0394	0.0312

Diferença entre duas proporções - Exemplo 2

Como os fumantes se comparam com os não fumantes? Faça um intervalo de confiança de 95%.

Dados:

- $\alpha = 0{,}05$
- $p_1 = 0{,}63$
- $p_2 = 0{,}42$
- $n_1 = 150$
- $n_2 = 250$

Diferença entre duas proporções - Exemplo 2

$\sigma_{p_1-p_2}$

$$\sigma_{p_1-p_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

$$\Longrightarrow \sigma_{p_1-p_2} = \sqrt{\frac{0{,}63(1-0{,}63)}{150} + \frac{0{,}42(1-0{,}42)}{250}} = 0{,}050$$

- $z_{\frac{\alpha}{2}} = z_{0{,}025} = 1{,}96$

Margem de erro

$$ME = z_{\frac{\alpha}{2}} \cdot \sigma_{p_1-p_2} = z_{0{,}025} \cdot 0{,}050 = 0{,}098$$

Diferença entre duas proporções - Exemplo 2

Intervalo de confiança

$$(p_1 - p_2) - ME \leq \pi_0 \leq (p_1 - p_2) + ME$$

$$(0{,}63 - 0{,}42) - 0{,}098 \leq \pi_0 \leq (0{,}63 - 0{,}42) + 0{,}098$$

$$0{,}11 \leq \pi_0 \leq 0{,}31$$

Obrigado pela atenção!