



Comparando
médias de 2
grupos

Felipe
Figueiredo

t de Student

IC diferença 2
médias

Aprofundamento

Comparando médias de 2 grupos

Intervalos de Confiança da diferença entre as médias

Felipe Figueiredo

Sumário

- 1 A distribuição t de Student
 - A distribuição t de Student
- 2 Intervalo de Confiança da diferença entre duas médias
 - Interpretação
 - Participantes: pareados ou não pareados?
- 3 Aprofundamento
 - Aprofundamento



Comparando
médias de 2
grupos

Felipe
Figueiredo

t de Student

IC diferença 2
médias

Aprofundamento

Discussão da aula passada



Comparando
médias de 2
grupos

Felipe
Figueiredo

t de Student

IC diferença 2
médias

Aprofundamento

Discussão da leitura obrigatória da aula passada

Recapitulando



Comparando
médias de 2
grupos

Felipe
Figueiredo

t de Student

A distribuição t de
Student

IC diferença 2
médias

Aprofundamento

Não vá se perder por aí...

- A distribuição Normal tem **dois** parâmetros
- Seu formato é absolutamente definido por
 - \bar{x} = Média (tendência central)
 - s^2/s = Variância/DP (tendência de dispersão)

⇒ Forma **independe** do n

Nomenclatura

A distribuição *Normal Padrão* também é chamada de **distribuição Z**.

- Vimos que o IC (da média) é composto por 3 componentes
 - a média \bar{x} (centro)
 - o erro padrão da média – SEM (incerteza)
 - *um tal* de t^* , que depende de n
- Quando n era grande, utilizamos $t^* \approx 2$
- Mas de onde vem esse t^* ? Qual seria o valor **correto**?



William Sealy Gosset and the T-Distribution



$$t = \frac{\bar{x}_1 - \bar{x}_2}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$



- Student (pseudônimo de W. S. Gossett [1876-1937]¹)
- Distribuição **t** (baseada na distribuição Normal)
- Melhor se aproxima dos dados de *amostras pequenas*
- 3º parâmetro **graus de liberdade**² vinculado ao tamanho da amostra n .

¹trabalhando para a cervejaria Guinness

²df em inglês

- A distribuição tem forma de sino (simétrica, como a Normal)
- Reflete a maior variabilidade **inerente** às amostras pequenas³
- Formato *depende* do tamanho da amostra (n)

Comparando
médias de 2
grupos

Felipe
Figueiredo

t de Student
A distribuição t de
Student

IC diferença 2
médias

Aprofundamento

Isto é

Quanto mais graus de liberdade, mais a distribuição t se parece com a distribuição Normal padrão (Z)

Pense...

O que deve acontecer com **menos** graus de liberdade?

³graus de liberdade (df) $\approx n$

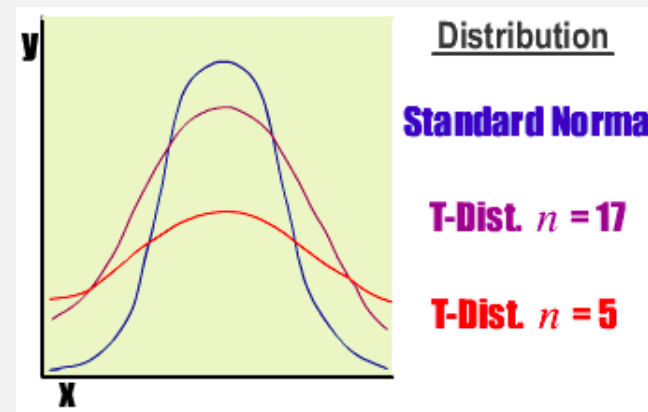


Figura: Duas distribuições t de Student, e a Normal padrão

Comparando
médias de 2
grupos

Felipe
Figueiredo

t de Student
A distribuição t de
Student

IC diferença 2
médias

Aprofundamento

ICs dos exemplos

- IC do ex. 5.1 (PS de 100 alunos): [120.6, 126.2] mmHg
- IC do ex. 5.2 (PS de 5 alunos): [79.2, 118.8] mmHg

Comparando
médias de 2
grupos

Felipe
Figueiredo

t de Student
A distribuição t de
Student

IC diferença 2
médias

Aprofundamento

Pense...

Observe os tamanhos dos ICs.

Lembrete

Para o 5.1, usamos $t^* \approx 2$.

Vimos que esta aproximação **não era apropriada** no 5.2

- $n = 5$ ($df = 4$) $\Rightarrow t^* = 2.776$
- $n = 10$ ($df = 9$) $\Rightarrow t^* = 2.262$
- $n = 15$ ($df = 14$) $\Rightarrow t^* = 2.145$
- $n = 20$ ($df = 19$) $\Rightarrow t^* = 2.093$
- $n = 30$ ($df = 29$) $\Rightarrow t^* = 2.045$

Pense...

Qual é a relação entre n e o tamanho do IC?

$$IC = [\bar{x} - t^* SEM, \bar{x} + t^* SEM]$$

Comparando
médias de 2
grupos

Felipe
Figueiredo

t de Student
A distribuição t de
Student

IC diferença 2
médias

Aprofundamento



Comparando
médias de 2
grupos

Felipe
Figueiredo

t de Student
A distribuição t de
Student

IC diferença 2
médias

Aprofundamento

- $n = 5$ ($df = 4$) $\Rightarrow t^* = 2.776$
- $n = 10$ ($df = 9$) $\Rightarrow t^* = 2.262$
- $n = 15$ ($df = 14$) $\Rightarrow t^* = 2.145$
- $n = 20$ ($df = 19$) $\Rightarrow t^* = 2.093$
- $n = 30$ ($df = 29$) $\Rightarrow t^* = 2.045$

Observe que...

- $df = n - 1$
- Para n grande, $t^* \rightarrow 1.960$

Por isso usamos o valor aproximado 2 no primeiro exemplo.



Comparando
médias de 2
grupos

Felipe
Figueiredo

t de Student
A distribuição t de
Student

IC diferença 2
médias

Aprofundamento

Distribuição Normal - Z

Gostaríamos de poder usar sempre Z como **modelo** para o formato dos nossos dados experimentais.

Distribuição t de Student

- t é uma aproximação da Normal (Z)
- idealizada para n pequeno
- Com n grande ($df \geq 30$) ela *se confunde* com Z.



Comparando
médias de 2
grupos

Felipe
Figueiredo

t de Student
A distribuição t de
Student

IC diferença 2
médias

Aprofundamento

Exercício 4 do cap 5

Os níveis séricos de um hormônio (fator Y) foram medidos em 100 mulheres não grávidas, e em 100 mulheres com até 3 meses de gravidez. Os ICs dos valores dos soros em ambos os grupos são:

- Grávidas: [105.4, 114.6]
- Não grávidas: [90.0, 96.0]

O fator Y médio é diferente em mulheres grávidas e não grávidas?

Requisito

Pelas premissas do IC da média, você tem informações suficientes para calcular/interpretar cada um destes ICs?



Comparando
médias de 2
grupos

Felipe
Figueiredo

t de Student
A distribuição t de
Student

IC diferença 2
médias

Aprofundamento

Exercício 5.4

- Não grávidas: [90.0, 96.0]
- Grávidas: [105.4, 114.6]

Observações:

- O SEM informa *quão bem você estimou a média* de cada grupo
- Os ICs não tem sobreposição \Rightarrow 2 populações diferentes

Pense...

Como comparar estes dois grupos?



Comparando
médias de 2
grupos

Felipe
Figueiredo

t de Student

IC diferença 2
médias

Interpretação
Participantes:
pareados ou não
pareados?

Aprofundamento

- Frequentemente precisamos dividir os dados em dois grupos e comparar as médias.
- Isto pode ser usado para se estudar o efeito de um tratamento em relação a um grupo controle
- ou mesmo para se comparar dois tratamentos diferentes.



Comparando
médias de 2
grupos

Felipe
Figueiredo

t de Student

IC diferença 2
médias

Interpretação
Participantes:
pareados ou não
pareados?

Aprofundamento

Exercício 4 do cap 5

Os níveis séricos de um hormônio (fator Y) foram medidos em 100 mulheres não grávidas, e em 100 mulheres com até 3 meses de gravidez. Os ICs dos valores dos soros em ambos os grupos são:

- Grávidas: [105.4, 114.6]
- Não grávidas: [90.0, 96.0]

O fator Y médio é diferente em mulheres grávidas e não grávidas?

Requisito

Pelas premissas do IC da média, você tem informações suficientes para calcular/interpretar cada um destes ICs?



Comparando
médias de 2
grupos

Felipe
Figueiredo

t de Student

IC diferença 2
médias

Interpretação
Participantes:
pareados ou não
pareados?

Aprofundamento

- x_C Hormônio não grávidas
- x_T Hormônio grávidas (até 3 meses)
- Duas variáveis explícitas

Primeira alternativa

- 1 "Explicar" a "relação" entre o hormônio x_T e o hormônio x_C
- 2 Comparar x_T (grupo de teste) com x_C (referência)

Esta relação pode ser expressa como

$$x_T \sim x_C$$



Comparando
médias de 2
grupos

Felipe
Figueiredo

t de Student

IC diferença 2
médias

Interpretação
Participantes:
pareados ou não
pareados?

Aprofundamento

Suspense dramático...

Se você prestou atenção até aqui...

Temos **duas** variáveis.

Portanto temos duas médias (trivial).

Mas também temos **dois SEM!**

Esta relação pode ser expressa como

horm. grávidas ~ horm. não grávidas

Mais precisamente

horm. grávid. = horm. não grávid. + **Erro_C** + **Erro_T**

Comparando
médias de 2
grupos

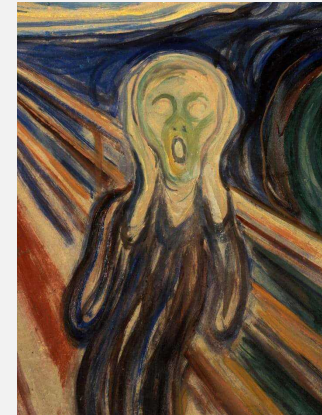
Felipe
Figueiredo

t de Student

IC diferença 2
médias

Interpretação
Participantes:
pareados ou não
pareados?

Aprofundamento



Duas médias, e dois erros?

Comparando
médias de 2
grupos

Felipe
Figueiredo

t de Student

IC diferença 2
médias

Interpretação
Participantes:
pareados ou não
pareados?

Aprofundamento

Exercício 5.4

- Não grávidas: [90.0, 96.0]
- Grávidas: [105.4, 114.6]

Difícil

Calcular os dois ICs (x_C e x_T), e compará-los diretamente

Moleza

Calcular o IC da diferença (x_d) usando o método da aula passada **Calcular o IC da diferença (x_d) usando o método da aula passada**

Comparando
médias de 2
grupos

Felipe
Figueiredo

t de Student

IC diferença 2
médias

Interpretação
Participantes:
pareados ou não
pareados?

Aprofundamento

Neste caso podemos usar um truque para **trocar** um problema de 2 variáveis por outro de 1 variável.



Comparando
médias de 2
grupos

Felipe
Figueiredo

t de Student

IC diferença 2
médias

Interpretação
Participantes:
pareados ou não
pareados?

Aprofundamento

Diferença entre 2 médias

- Comparar duas médias \bar{x}_C e \bar{x}_T , consideramos a diferença média $\bar{x}_d = \bar{x}_T - \bar{x}_C$
- Se \bar{x}_T for maior que $\bar{x}_C \Rightarrow$ diferença média é positiva
- Se \bar{x}_T for menor que $\bar{x}_C \Rightarrow$ a diferença média é negativa

Intuição

Raciocínio: se as médias forem aproximadamente iguais...

... a diferença média (\bar{x}_d) será próxima de zero

Pense em **saldo**

Nesse caso a diferença média (\bar{x}_d) é o **efeito observado**

Efeitos grandes são mais fáceis de ser detectados

- x_C Hormônio não grávidas
- x_T Hormônio grávidas (até 3 meses)
- $d = x_T - x_C$ (uma variável)

Segunda alternativa (método da aula passada)

“Explicar” a “relação” entre a diferença d e a referência (**zero**)

Esta relação pode ser expressa como

$$d \sim 0$$

Estratégia proposta

Temos duas variáveis.

Calculamos a **diferença** entre as médias e aplicamos o método da aula passada – IC de **uma** média.

moleza!

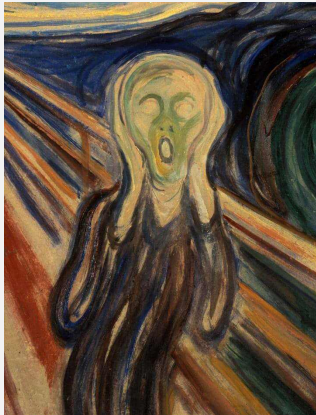
O que falta?

O que falta?

... precisamos do *SEM da diferença*.

Ou seja...

$$d = 0 + \text{Erro}_d$$



SEM da diferença?

Comparando
médias de 2
grupos

Felipe
Figueiredo

t de Student

IC diferença 2
médias

Interpretação

Participantes:
pareados ou não
pareados?

Aprofundamento

- Lembre-se que para cada grupo: $SEM = \frac{s}{\sqrt{n}}$
- Para a diferença entre 2 grupos, “somamos” os SEM
- Mas esta “soma” não é direta!
- É preciso levar em conta o uso do quadrado/raiz quadrada do DP (aula de variabilidade⁴)

$$SEM_d = \sqrt{(SEM_C)^2 + (SEM_T)^2}$$

⁴ não podemos somar DPs, mas podemos somar variâncias

Comparando
médias de 2
grupos

Felipe
Figueiredo

t de Student

IC diferença 2
médias

Interpretação

Participantes:
pareados ou não
pareados?

Aprofundamento



Estratégia proposta

SEM da diferença.

Comparando
médias de 2
grupos

Felipe
Figueiredo

t de Student

IC diferença 2
médias

Interpretação

Participantes:
pareados ou não
pareados?

Aprofundamento

- As amostras foram selecionadas aleatoriamente das respectivas populações
- As populações são Normais (Gaussianas)
- As duas populações possuem DP idênticos
- Todos os indivíduos de cada grupo vêm da mesma população
- Cada indivíduo é independente de todos os outros

Comparando
médias de 2
grupos

Felipe
Figueiredo

t de Student

IC diferença 2
médias

Interpretação

Participantes:
pareados ou não
pareados?

Aprofundamento

Exercício 4 do cap 5

Os níveis séricos de um hormônio (fator Y) foram medidos em 100 mulheres não grávidas, e em 100 mulheres com até 3 meses de gravidez. Os ICs dos valores dos soros em ambos os grupos são:

- Grávidas: [105.4, 114.6]
- Não grávidas: [90.0, 96.0]

O fator Y médio é diferente em mulheres grávidas e não grávidas?

Requisito

Pelas premissas do IC da média, você tem informações suficientes para calcular/interpretar cada um destes ICs?

Diferenças: Exercício 5.4 (e 7.1)

- Média grávidas: $\bar{x}_C = 110$ unidades/ml
- Média não grávidas: $\bar{x}_T = 93$ unidades/ml
- Diferença entre as médias: $\bar{x}_d = 17$ unidades/ml

- SEM da diferença: **2.75** unidades/ml

- $n_C = 100, n_T = 100$
- $df = (100 - 1) + (100 - 1) = 198$

- $t^* = 1.97$ (valor crítico tabelado)

Bastidores: Exercício 5.4 (e 7.1)

- Média grávidas: $\bar{x}_C = 110$ unidades/ml
- Média não grávidas: $\bar{x}_T = 93$ unidades/ml
- Diferença entre as médias: $\bar{x}_d = 17$ unidades/ml
- SEM da diferença: **2.75** unidades/ml
- $n_C = 100, n_T = 100$
- $df = (100 - 1) + (100 - 1) = 198$
- $t^* = 1.97$ (valor crítico tabelado)

Resultado: IC da diferença

[11.6, 22.4] unidades/ml

E o que isso significa?

Interpretação

Estamos 95% *confiantes* que a diferença **real** entre os grupos está entre 11.6 e 22.4.

Conclusão (“*nossos dados indicam que...*”)

o (...) fator Y de uma (...) grávida é (...) 17 unidades/ml maior que uma (...) não grávida (variando entre 11,6 e 22,4 unidades/ml).

Pense...

Preencha as lacunas acima.

Grupos não pareados

- Até agora assumimos que os grupos e participantes são **independentes**
- A única coisa que podemos fazer: **comparação global**
- ... a média do grupo A \times a média do grupo B

Grupos pareados

- Existe um caso importante em que pode-se considerar que eles são dependentes: quando são pareados
- Isto é: cada participante de um grupo tem um correspondente no outro
- ... diferença entre cada par \Rightarrow média das diferenças

Quando faz sentido parear indivíduos de dois grupos?

- Mensurar o **mesmo** indivíduo antes e depois do procedimento (*baseline* x intervenção)
- Recrutamento aos pares, quando o par tem a(o) mesma(o)
 - idade/faixas etária
 - região demográfica
 - diagnóstico
- irmãos, pai/filho
- lateralidade (tratamento = lado E, controle = lado D)

Exemplo 7.2

Ye e Grantham (1993) estudaram o mecanismo de absorção de fluido em cistos renais removidos de pacientes com doença renal policística. Incubaram os cistos em meio de cultura celular e mediram a diferença de peso em cada cisto (antes e depois da incubação).

Não pareado

- peso médio (todos, antes) = 6.51g (SEM 2.26g)
- peso médio (todos, depois) = 7.02g (SEM 2.40g)
- IC 95% da diferença [-6.48, 7.50]

Pareado

- ganho em **cada** cisto \Rightarrow depois - antes
- ganho médio dos cistos = 0.50g (SEM 0.23g).
- IC 95% da diferença [-0.03, 1.04]

A escolha entre grupos pareados e grupos não pareados é estratégica (planejamento do estudo), e não uma questão de “preferência”.

Leitura obrigatória

- Capítulo 5. Seção: A distribuição t
- Capítulo 7: Pular as seções
 - Cálculo do IC de grupos independentes
 - Cálculo do IC de grupos pareados

Leitura recomendada

- **ICH - E10** Choice of Control Group in Clinical Trials
 - Seção 2.1 (*Placebo Control*)
 - Cap. 3 (*CHOOSING THE CONCURRENT CONTROL GROUP*)

<http://www.ich.org> (este link é clicável)