

# Comparando médias de 2 grupos

## Intervalos de Confiança da diferença entre as médias

Felipe Figueiredo

- 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?

## 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?

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

# A distribuição T de Student



Comparando  
médias de 2  
grupos

Felipe  
Figueiredo

t de Student

A distribuição t de  
Student

IC diferença 2  
médias

# A distribuição T de Student

Comparando  
médias de 2  
grupos

Felipe  
Figueiredo

t de Student  
A distribuição t de  
Student

IC diferença 2  
médias

## William Sealy Gosset and the T-Distribution



$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{s^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$$



# A distribuição t de Student



Comparando  
médias de 2  
grupos

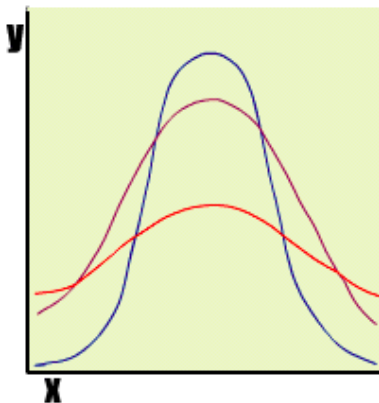
Felipe  
Figueiredo

t de Student  
A distribuição t de  
Student

IC diferença 2  
médias

- Student (pseudônimo de W. S. Gossett [1876-1937], trabalhando para a cervejaria Guinness) criou uma distribuição que melhor se aproxima dos dados de amostras pequenas
- Tem um parâmetro **graus de liberdade** ( $df$  em inglês) vinculado ao tamanho da amostra  $n$ .

# A distribuição t de Student



Distribution

**Standard Normal**

**T-Dist.  $n = 17$**

**T-Dist.  $n = 5$**

Figura: A distribuição t de Student



- A distribuição tem forma de sino (simétrica, assim como a distribuição Normal)
- Reflete a maior variabilidade inerente às amostras pequenas
- O formato da curva depende do tamanho da amostra  $n$
- Quanto mais graus de liberdade ( $df \approx \text{dados}$ ), mais a distribuição  $t$  se parece com a distribuição Normal

## 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

## Pense...

Observe os tamanhos dos ICs.

# Alguns valores de $t$ , para diferentes graus de liberdade



Comparando  
médias de 2  
grupos

Felipe  
Figueiredo

$t$  de Student  
A distribuição  $t$  de  
Student

IC diferença 2  
médias

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

## Pense...

Qual é a relação entre  $N$  e o tamanho do IC?

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

# Alguns valores de $t$ , para diferentes graus de liberdade



Comparando  
médias de 2  
grupos

Felipe  
Figueiredo

$t$  de Student  
A distribuição  $t$  de  
Student

IC diferença 2  
médias

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

Observe que...

- $df = N - 1$
- Para  $N$  grande,  $t \rightarrow 1.960$

Por isso usamos o valor aproximado 2 no primeiro exemplo.

# Exercício 4 (cap 5)

## Exercício 4 do cap 5

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

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

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

- Não-grávidas: [90.0, 96.0]
  - Grávidas: [105.4, 114.6]
- 
- o SEM informa quanto bem você conhece a média de cada grupo
  - Os ICs não tem sobreposição  $\Rightarrow$  2 populações diferentes
  - Como comparar estes dois grupos?

## 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?

# Diferença entre 2 médias



Comparando  
médias de 2  
grupos

Felipe  
Figueiredo

t de Student

IC diferença 2  
médias

Interpretação

Participantes:  
pareados ou  
não-pareados?

- 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.



# Diferença entre 2 médias



Comparando  
médias de 2  
grupos

Felipe  
Figueiredo

t de Student

IC diferença 2  
médias

Interpretação

Participantes:  
pareados ou  
não-pareados?

- Para comparar duas médias  $\bar{x}_1$  e  $\bar{x}_2$ , consideramos a diferença  $\bar{x}_1 - \bar{x}_2$
- Raciocínio: se as médias forem aproximadamente iguais, a diferença será aproximadamente zero
- Além disso, se  $\bar{x}_1$  for maior que  $\bar{x}_2$ , a diferença será positiva
- Analogamente, se  $\bar{x}_1$  for menor que  $\bar{x}_2$ , a diferença será negativa

# Erro padrão da diferença

- Lembre-se que para cada grupo:  $SEM = \frac{DP}{\sqrt{N}}$
- 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 variabilidade)

$$SE = \sqrt{SEM_1^2 + SEM_2^2}$$

- 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

## Exercício 4 (cap 5)

### Exercício 4 do cap 5

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

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

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

## Diferenças: Exercício 5.4 (e 7.1)

- Média grávidas:  $\bar{x}_1 = 110$  unidades/ml
- Média não-grávidas:  $\bar{x}_2 = 93$  unidades/ml
- Diferença entre as médias:  $\bar{x}_d = 17$  unidades/ml
- SEM da diferença: 2.75 unidades/ml
- $N_1 = 100, N_2 = 100 \Rightarrow df = (100 - 1) + 1(100 - 1) = 198$
- $t^* = 1.97$

## IC

[11.6, 22.4] unidades/ml

E o que significa isso?

IC

[11.6, 22.4] unidades/ml

- Estamos 95% certos que a diferença real entre os grupos está entre 11.6 e 22.4
- Conclusão: o fator Y de uma mulher grávida entre 11.6 e 22.4 unidades/ml maior que em uma mulher não grávida

## 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?

# Grupos não-pareados x pareados



Comparando  
médias de 2  
grupos

Felipe  
Figueiredo

t de Student

IC diferença 2  
médias

Interpretação

Participantes:  
pareados ou  
não-pareados?

Aula que vem...



- Até agora assumimos que os grupos e participantes são **independentes**
- Existe um caso importante em que pode-se considerar que eles são dependentes: quando são pareados
- Isto significa que cada participante de um grupo tem um correspondente no outro
-

# Comparação pareada



Comparando  
médias de 2  
grupos

Felipe  
Figueiredo

t de Student

IC diferença 2  
médias

Interpretação

Participantes:  
pareados ou  
não-pareados?