

Comparação entre duas amostras

PRI5003 - Lab 8

Instituto de Relações Internacionais - Universidade de São Paulo

24 de Maio de 2018

Outline

Onde paramos e para onde vamos

Comparando dois grupos

Amostras dependentes

Amostras independentes

Exemplo

Retomando o programa do curso

Até a prova, cobrimos o seguinte material:

- ▶ Noções de desenho de pesquisa
- ▶ Estatística descritiva
- ▶ Fundamentos de inferência
- ▶ Comparação de uma amostra com um valor hipotético

Esse conteúdo serve como ponto de partida para os testes que costumam ser utilizados em pesquisa. De modo geral, a segunda parte do curso vai se concentrar na comparação de dois ou mais grupos, em vez de apenas um.

Retomando o programa do curso

O caminho das pedras para o resto do curso é o seguinte:

- ▶ Comparação entre duas amostras (hoje)
- ▶ Medidas de associação entre duas variáveis (aula 9)
- ▶ Introdução aos modelos de regressão (aulas 10, 11 e 12)

Comparação de grupos

- ▶ A comparação de grupos é um dos testes mais básicos que podemos fazer com dados estatísticos
 - ▶ O salário dos homens é diferente do das mulheres?
 - ▶ Países democráticos aderem com mais frequência a tratados de direitos humanos do que os autocráticos?
 - ▶ Governos de direita investem mais dinheiro em ajuda externa do que governos de esquerda?
- ▶ Atualmente é muito utilizada em trabalhos com desenho de pesquisa experimental

A pergunta básica que vamos responder é: temos evidência para concluir que a média do grupo 1 é estatisticamente diferente da média do grupo 2?

Amostras dependentes e independentes

Para fazer o teste corretamente, precisamos definir se os 2 grupos que estamos comparando são dependentes ou independentes

- ▶ Amostras dependentes (ou amostra pareada): mesmas unidades são estudadas em períodos de tempo diferentes. Ex: estudos longitudinais de atitudes em relação a políticas comerciais
- ▶ Amostras independentes: as unidades em um grupo são diferentes e independentes em relação ao segundo grupo: Ex: estudos transversais em que apenas um grupo recebe a informação sobre uma política pública

A única coisa que muda entre os dois casos é o cálculo do erro-padrão. Quando a amostra é pareada, esperamos menos variação de características não observadas. Esperamos, portanto, um erro-padrão menor.

Diferença para amostras dependentes

$$\mu_{dif} = \mu_2 - \mu_1$$

- ▶ A boa notícia: é como se tivéssemos apenas uma amostra!
- ▶ A má notícia: muito pouco usado em ciências sociais

Estatística-teste

$$t = \frac{\bar{y} - 0}{ep}, \text{ em que } ep = \frac{s_d}{\sqrt{n}}$$

Diferença para amostras independentes

	Type of Response Variable	
	Categorical	Quantitative
Estimation		
1. Parameter	$\pi_2 - \pi_1$	$\mu_2 - \mu_1$
2. Point estimate	$\hat{\pi}_2 - \hat{\pi}_1$	$\bar{y}_2 - \bar{y}_1$
3. Standard error	$se = \sqrt{\frac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1-\hat{\pi}_2)}{n_2}}$	$se = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
4. Confidence interval	$(\hat{\pi}_2 - \hat{\pi}_1) \pm z(se)$	$(\bar{y}_2 - \bar{y}_1) \pm t(se)$

Diferença para amostras independentes

Significance testing

1. Assumptions

Randomization
 ≥ 10 observations in each
category, for each group

Randomization
Normal population dist.'s
(robust, especially for large n 's)

2. Hypotheses

$$H_0: \pi_1 = \pi_2$$
$$(\pi_2 - \pi_1 = 0)$$
$$H_a: \pi_1 \neq \pi_2$$

$$H_0: \mu_1 = \mu_2$$
$$(\mu_2 - \mu_1 = 0)$$
$$H_a: \mu_1 \neq \mu_2$$

3. Test statistic

$$z = \frac{\hat{\pi}_2 - \hat{\pi}_1}{se_0}$$

$$t = \frac{\bar{y}_2 - \bar{y}_1}{se}$$

4. P -value

Two-tail probability from standard normal or t
(Use one tail for one-sided alternative)

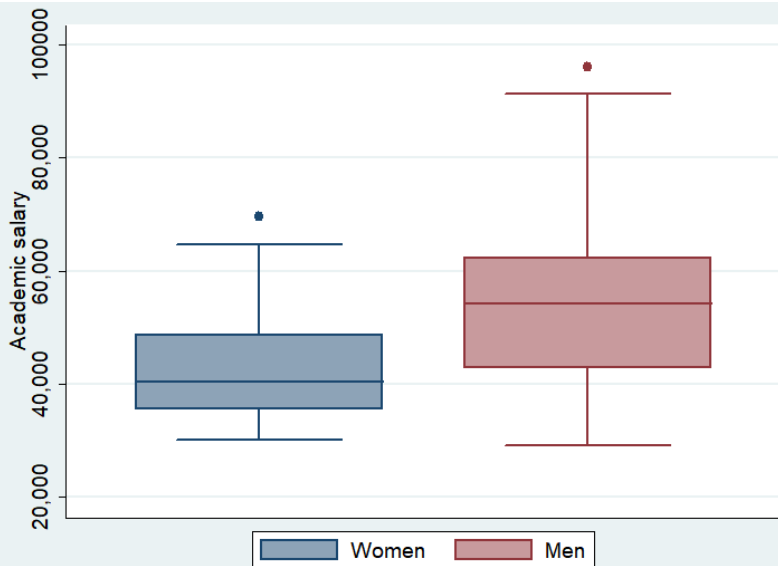
Exemplo

```
. tabstat salary, by(male) statistics(mean sd n)
```

```
Summary for variables: salary  
by categories of: male
```

male	mean	sd	N
Women	42916.6	9161.61	128
Men	53499.24	12583.48	386
Total	50863.87	12672.77	514

Exemplo



Exemplo

```
. ttest salary, by(male)
```

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
Women	128	42916.6	809.7795	9161.61	41314.2	44519.01
Men	386	53499.24	640.4822	12583.48	52239.96	54758.52
combined	514	50863.87	558.972	12672.77	49765.72	51962.03
diff		-10582.63	1206.345		-12952.63	-8212.636

diff = mean(Women) - mean(Men)

t = -8.7725

Ho: diff = 0

degrees of freedom = 512

Ha: diff < 0

Ha: diff != 0

Ha: diff > 0

Pr(T < t) = 0.0000

Pr(|T| > |t|) = 0.0000

Pr(T > t) = 1.0000