

Stata, Exemplo Prático I

Heitor Lima

Métodos Quantitativos para Avaliação de Políticas Públicas - MPP - 2023

Professora: Letícia Nunes

Insper Instituto de Ensino e Pesquisa

`heitoraol@al.insper.edu.br`

1. Links úteis

- Exemplo de teste de médias utilizado nestes slides: [Link](#)
- Base de dados utilizada nestes slides: [Link](#) (baixar a base de 2022)
- Documentação oficial do comando ttest do Stata: [Link](#)

Teste de Médias

Contexto e Base de Dados

Um teste de hipótese muito comum é o teste de diferença de média (*mean difference*)

- Dada uma amostra, dois subgrupos são selecionados
- Calcula-se a média de alguma v.a. para cada subgrupo
- Objetivo: Testar se a diferença entre as médias é significativa

Usaremos os dados da *General Social Survey 2022*, uma pesquisa abrangente feita com americanos aleatoriamente selecionados

- Subgrupos: Homens e mulheres
- V.A. selecionada: horas trabalhadas
- Objetivo: Testar se a diferença entre as médias de horas trabalhadas entre homens e mulheres é significativa

Médias de Horas Trabalhadas na Amostra

Pergunta de pesquisa: Homens trabalham mais do que mulheres?

- Usamos o comando `table` para verificar as médias de horas trabalhadas na amostra
- Na amostra, homens trabalham 4.5 horas a mais que mulheres, em média

```
1 . use "GSS2022/GSS2022.dta"
2 . table (sex), statistic(mean hrs1)
3 -----
4                |      Mean
5 -----+-----
6 respondents sex |
7   male          |  42.48269
8   female         |  37.96201
9   dk, na, iap    |  40.19772
10 -----
```

Teste t para Dois Grupos

Podemos dizer que, em toda a **população** dos EUA, homens trabalham 4.5 horas a mais que mulheres?

- Para responder a esta pergunta, usaremos o teste t para dois grupos

A hipótese nula é definida como zero, ou seja, a diferença ($\hat{\mu}$) entre a média de horas trabalhadas por homens (μ_h) e mulheres (μ_m) é zero

- $H_0 : \hat{\mu} = \mu_h - \mu_m = 0$

Intuição: Se H_0 for verdadeira, e tomássemos amostras aleatórias de 1,927 indivíduos, com que frequência veríamos $\hat{\mu} = 4.5$ apenas *por acaso*?

- A medida da frequência com que vemos tal diferença apenas por conta de amostragem aleatória (por acaso) é o **valor-p** (p -value)

→ Bussab e Morettin (2017, cap. 12), Wooldridge (2013, Apêndice C)

Teste t para Dois Grupos

No Stata, usamos o comando `ttest` para calcular o p -value e a estatística t

```
. ttest hrs1, by(sex)
```

Two-sample t test with equal variances

Variable	Obs	Mean	Std. err.	Std. dev.	[95% conf. interval]	
male	953	42.48269	.4627338	14.28491	41.57459	43.39078
female	974	37.96201	.4365433	13.62407	37.10534	38.81869
Combined	1,927	40.19772	.3219576	14.13317	39.56629	40.82914
diff		4.520674	.6358263		3.273693	5.767655

diff = mean(male) - mean(female)

t = 7.1099

H0: diff = 0

Degrees of freedom = 1925

Ha: diff < 0

Ha: diff != 0

Ha: diff > 0

Pr(T < t) = 1.0000

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

Pr(T > t) = 0.0000

Teste t para Dois Grupos

A diferença média é de 4.521. A estatística t é obtida pela fórmula padrão:

$$t = \frac{\hat{\mu} - \mu_0}{SE(\hat{\mu})} = \frac{4.521 - 0}{.636} = 7.11$$

Na parte de baixo da imagem, vemos os testes de significância

- Note que H_0 considera como verdade a hipótese de que $\hat{\mu} = 0$
- O Stata mostra as três possibilidades de hipótese alternativa
- A mais importante é a do meio, que testa se o parâmetro estimado é diferente de zero
- O p -value é igual a 0.0000

Teste t para Dois Grupos: Interpretação

- Se nossa hipótese inicial fosse verdadeira, ou seja, se de fato a diferença entre horas trabalhadas por homens e mulheres fosse igual a zero...
- ... veríamos uma diferença de 4.521 **menos** de 1 vez a cada 10,000 amostragens aleatórias (por acaso)
 - Ou seja, veríamos $\hat{\mu} = 4.521$ (ou algo próximo disso) várias vezes. Isso não é “por acaso”, mas sim uma característica **consistente** da população
- Portanto, a conclusão mais razoável é que, de fato, há uma diferença de 4.521 também na população
- Se o p -value fosse 0.2534, veríamos uma diferença de 4.521 em 25 a cada 100 vezes, apenas devido à amostragem aleatória, mesmo que não houvesse tal diferença na população
- Nesses casos, não podemos rejeitar H_0 , pois pode ser que de fato a diferença na população seja zero

Regressão Linear Simples

Regressão Linear Simples

Com os mesmos dados, vamos regredir salário em educação

- Faremos a análise apenas para mulheres, controlando para número de filhos e idade

```
. global controls childs age
```

```
. reg income educ $controls if sex==2
```

Source	SS	df	MS	Number of obs	=	1,591
				F(3, 1587)	=	39.49
Model	535.067504	3	178.355835	Prob > F	=	0.0000
Residual	7167.43658	1,587	4.51634315	R-squared	=	0.0695
				Adj R-squared	=	0.0677
Total	7702.50409	1,590	4.84434219	Root MSE	=	2.1252

income	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
educ	.1884968	.01854	10.17	0.000	.1521314	.2248622
childs	-.0581465	.034842	-1.67	0.095	-.1264877	.0101947
age	.0031658	.0031323	1.01	0.312	-.0029782	.0093097
_cons	8.366243	.3090597	27.07	0.000	7.760034	8.972451

Regressão Linear Simples: Interpretação

- Um ano adicional de educação está positivamente correlacionado com a renda: $+0.188$ desvio padrão de renda
 - *p-value* menor que 0.05, i.e., o parâmetro é estatisticamente significativo a 5%
- Um(a) filho(a) a mais está negativamente correlacionado(a) com a renda: -0.058 desvio padrão de renda
 - *p-value* maior que 0.05, i.e., o parâmetro não é estatisticamente significativo a 5%
- Um ano adicional na idade está positivamente correlacionado com a renda: $+0.003$ desvio padrão de renda
 - *p-value* maior que 0.05, i.e., o parâmetro não é estatisticamente significativo a 5%
- Podemos dizer que estas relações são causais?
 - Dificilmente! Apesar de ser uma amostra dita aleatória, seria necessário realizar verificações adicionais que garantissem aleatoriedade de características específicas

Referências

Referências

BUSSAB, W., E P. MORETTIN (2017). *Estatística Básica*. 9 ed. São Paulo: Saraiva.

WOOLDRIDGE, J. (2013). *Introductory Econometrics: A Modern Approach*. 5th ed. Andover, UK: Cengage.

Notas de aula da Prof^a. Letícia Nunes.