

# Discussion 4

---

Week of 02.09

# Agenda

- Quiz review
- Stata practice & covariates in OLS regressions

# Announcements

- Quiz 3 due Tuesday, Feb. 10 at 9pm.

# Quiz Review

## Question 4

One solution to pandemic-related schooling issues that some people have proposed is after-school programs. You find a study (N=50) showing that a voluntary after-school math tutoring program improved student test scores by 8 percentage points. Why might you be skeptical of this claim?

- 8 percentage points is not an economically or statistically significant increase.
- Students who choose to attend a math tutoring program might be more motivated to study and therefore would perform better on tests even if they had not attended tutoring.
- The sample size is not large enough.
- I'm skeptical because I've been told to be skeptical.

## Question 4

One solution to pandemic-related schooling issues that some people have proposed is after-school programs. You find a study (N=50) showing that a voluntary after-school math tutoring program improved student test scores by 8 percentage points. Why might you be skeptical of this claim?

- 8 percentage points is not an economically or statistically significant increase.
- Students who choose to attend a math tutoring program might be more motivated to study and therefore would perform better on tests even if they had not attended tutoring.
- The sample size is not large enough.
- I'm skeptical because I've been told to be skeptical.

# Creating Tables in STATA

**eststo** allows you to store the results of a regression in STATA.

- Use `ssc install estout, replace` to download the package
- Store regressions with `eststo: reg y x1 x2`
- Use `esttab` to display the results.
- Use `eststo clear` to clear the regressions if you make a mistake.

# Try it yourself!

- Using the auto.dta dataset (sysuse auto.dta), create a table with the following regressions:

$$\mathbf{Price = \beta_0 + \beta_1MPG + \epsilon}$$

$$\mathbf{Price = \alpha_0 + \alpha_1MPG + \alpha_2Weight}$$

$$\mathbf{Price = \gamma_0 + \gamma_1MPG + \gamma_2Weight + \gamma_3Length}$$

```
. esttab
```

---

	(1)	(2)	(3)
	price	price	price
mpg	-238.9*** (-4.50)	-49.51 (-0.57)	-86.79 (-1.03)
weight		1.747** (2.72)	4.365*** (3.74)
length			-104.9* (-2.64)
_cons	11253.1*** (9.61)	1946.1 (0.54)	14542.4* (2.47)

---

N                    74                    74                    74

---

t statistics in parentheses

\* p<0.05, \*\* p<0.01, \*\*\* p<0.001

**P** Using the *nhanes2f* dataset:

1. Find the summary statistics for age, weight, height, and heart attacks
2. Run the following three regressions and create a table with them

$$\mathbf{BloodPressure = \beta_0 + \beta_1 Weight + \varepsilon}$$

$$\mathbf{BloodPressure = \alpha_0 + \alpha_1 Weight + \alpha_2 Age + \varepsilon}$$

$$\mathbf{BloodPressure = \gamma_0 + \gamma_1 Weight + \gamma_2 Age + \gamma_3 Female + \varepsilon}$$

## P Interpret these results.

	(1)	(2)	(3)
	bpsystol	bpsystol	bpsystol
weight	0.434*** (30.32)	0.407*** (32.56)	0.417*** (30.90)
age		0.638*** (57.29)	0.638*** (57.23)
female			0.826* (1.99)
_cons	99.65*** (94.59)	71.28*** (68.35)	70.14*** (59.01)
N	10337	10337	10337

t statistics in parentheses  
\* p<0.05, \*\* p<0.01, \*\*\* p<0.001

**P** Now add the following regression to your table:

$$\mathbf{BloodPressure = \beta_0 + \beta_1 Weight + \beta_2 Age + \beta_3 Diabetic + \epsilon}$$

**P** Which of these regressions do you think is the best?

	(1) bpsystol	(2) bpsystol	(3) bpsystol	(4) bpsystol
weight	0.434*** (30.32)	0.407*** (32.56)	0.417*** (30.90)	0.372*** (27.00)
age		0.638*** (57.29)	0.638*** (57.23)	0.632*** (56.49)
female			0.826* (1.99)	
diabetic				4.338*** (5.90)
_cons	99.65*** (94.59)	71.28*** (68.35)	70.14*** (59.01)	73.69*** (65.91)
N	10337	10337	10337	10335

t statistics in parentheses  
\* p<0.05, \*\* p<0.01, \*\*\* p<0.001

**P** Compare the results of regressions 3 and 4. Why did the coefficient on weight change so much? What might this tell us about our controls?

	(1)	(2)	(3)	(4)
	bpsystol	bpsystol	bpsystol	bpsystol
weight	0.434*** (30.32)	0.407*** (32.56)	0.417*** (30.90)	0.372*** (27.00)
age		0.638*** (57.29)	0.638*** (57.23)	0.632*** (56.49)
female			0.826* (1.99)	
diabetic				4.338*** (5.90)
_cons	99.65*** (94.59)	71.28*** (68.35)	70.14*** (59.01)	73.69*** (65.91)
N	10337	10337	10337	10335

t statistics in parentheses  
\* p<0.05, \*\* p<0.01, \*\*\* p<0.001

# Additional Practice?

 **STATA: A Beginner's Guide to Esttab** (For more esttab practice)

## Question 2

Which of the following would **not** be a useful reference point for evaluating the size of the effect of the pandemic on reading scores?

- The average math score from 2019.
- The average reading score from after the pandemic.
- The annual change in reading test scores over 2014-2019.
- The minimum and maximum scores possible on the reading test.

## Question 2

Which of the following would **not** be a useful reference point for evaluating the size of the effect of the pandemic on reading scores?

- The average math score from 2019.
- The average reading score from after the pandemic.
- The annual change in reading test scores over 2014-2019.
- The minimum and maximum scores possible on the reading test.