Recitation: Data analysis

Felipe Balcazar
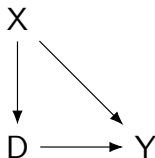
NYU

October, 2021
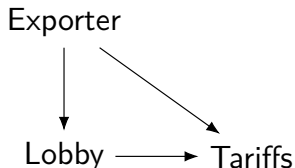
NEW YORK UNIVERSITY

X

D $\longrightarrow$ Y

- We ask whether lobbying increases tariffs on imports.
- Exporting firms have incentives to liberalize.
- Unit of observation (i) is the firm.

Exporter

Lobby $\longrightarrow$ Tariffs

- We ask whether lobbying increases tariffs on imports.
- Exporting firms have incentives to liberalize.
- Unit of observation (i) is the firm.

- How can we begin to analyze Y? We can compute...
  - The mean: $\mu = \frac{\sum_i y}{N}$.
  - The (sample) variance: $\sigma^2 = \frac{\sum_i (y-\mu)^2}{(N-1)}$.
  - The (sample) standard deviation: $\sigma = \sqrt{\frac{\sum_i (y-\mu)^2}{(N-1)}}$.
  - The median or 50th percentile.
  - Maximum or minimum.

Summary statistics

- The mean: $E(tariff) = \frac{\sum_i tariffs}{N_{firms}}$

- The (sample) variance: $Var(tariff) = \frac{\sum_i (tariffs_i - \overline{tariff})^2}{(N_{firms} - 1)}$

- The (sample) standard deviation:

$$Sd(tariff) = \sqrt{Var(tariff)} = \sqrt{\frac{\sum_i (tariff_i - E(tariff))^2}{(N_{firms} - 1)}}$$

- The median or 50th percentile.

- The maximum and the minimum.

How can we can make questions

- Is the average tariff rate 0.07?

  - *Null hypothesis* (H_0): $E(\text{tariff}) = 0.07$.
  - The alternative (H_a): $E(\text{tariff}) \neq 0.07$.

- Is the average tariff rate 0.07?

  - *Null hypothesis* (H_0): $E(tariff) = 0.07$.
  - The alternative (H_a): $E(tariff) \neq 0.07$.

		Null hypothesis	
		TRUE	FALSE
Findings	Reject null	Type I error (α)	Correct decision
	Accept null	Correct decision	Type II error (β)

- α is what is called "significance." Probability of type I error.

- β is what is called "power." Probability of type II error.

- Is the average tariff rate 0.07?

    - *Null hypothesis* (H_0): $E(tariff) = 0.07$.
    - The alternative (H_a): $E(tariff) \neq 0.07$.

		Null hypothesis	
		TRUE	FALSE
Findings	Reject null	False positive (α)	Correct decision
	Accept null	Correct decision	False negative (β)

- α is what is called "significance." Probability of false positive.

- β is what is called "power." Probability of false negative.

- We do not want to reject the null when it is true: *type I error* - i.e., we do not want false positives.

- The probability for this must be small, less than α.
  - α is the statistical significance (usually 0.05 or 5%).

- We do not want to accept the null when it is false: *type II error* - i.e., we do not want false negatives.

  - If the number of unit of observations you have is too small (below 42) you will be underpowered.

  - Important: I talk about the number of unit of observations, not the number of observations!

- To test this hypothesis we compute a t-statistic:

$$\hat{t} = \frac{E(Tariff) - H_0}{SE}.$$

- $SE = \frac{SD(Tariff)}{\sqrt{N_{firms}}}$ - this is the standard error.

- To test this hypothesis we compute a t-statistic:

$$\hat{t} = \frac{E(\text{Tariff}) - 0.07}{SE}.$$

- $SE = \frac{SD(\text{Tariff})}{\sqrt{N_{firms}}}$ - this is the standard error.

- To test this hypothesis we compute a t-statistic:

$$\hat{t} = \frac{E(\textit{Tariff}) - 0.07}{SE}.$$

  - $SE = \frac{SD(\textit{Tariff})}{\sqrt{N_{firms}}}$ - this is the standard error.

- Then we ask $Pr(|t| \geq |\hat{t}|) \equiv$ p-value.
  - p-value is the probability of type I error.

- If p-value $< \alpha$, we reject the null hypothesis.

  - Don't worry, the computer does all of this for you.

  - You probably have seen stars in papers.

- To test this hypothesis we compute a t-statistic:

$$\hat{t} = \frac{E(\textit{Tariff}) - 0.07}{SE}.$$

  - $SE = \frac{SD(\textit{Tariff})}{\sqrt{N_{firms}}}$ - this is the standard error.

- Then we ask $Pr(|t| \geq |\hat{t}|) \equiv$ p-value.
  - p-value is the probability of type I error.

- If p-value $< \alpha$, we reject the null hypothesis.

  - Don't worry, the computer does all of this for you.

  - You probably have seen stars in papers.

Eyeballing the answer to the same question

- Use a confidence interval!

- Focus on the statistical significance (α),

- So $(1 - \alpha) \times 100$ is the confidence that H_0 is true.

    - 90% confidence when $\alpha = 0.1$.

    $$H_0 \in [E(\textit{Tariff}) - 1.65SE; E(\textit{Tariff}) + 1.65SE]$$

    - 95% confidence when $\alpha = 0.05$.

    $$H_0 \in [E(\textit{Tariff}) - 1.96SE; E(\textit{Tariff}) + 1.96SE]$$

    - 99% confidence when $\alpha = 0.01$.

    $$H_0 \in [E(\textit{Tariff}) - 2.56SE; E(\textit{Tariff}) + 2.56SE]$$

Statistical significance

- *Null hypothesis* (H_0): $E(tariff) = 0.0$.
- The alternative (H_a): $E(tariff) \neq 0.0$.
- If we reject the null hypothesis then $E(tariff)$ is statistically significant - that is statistically different from zero.

  - 90% confident that is not zero if

    $$0 \notin [E(Tariff) - 1.65SE; E(Tariff) + 1.65SE]$$

  - 95% confident that is not zero if

    $$0 \notin [E(Tariff) - 1.96SE; E(Tariff) + 1.96SE]$$

  - 99% confident that is not zero if

    $$0 \notin [E(Tariff) - 2.56SE; E(Tariff) + 2.56SE]$$

- *Null hypothesis* (H_0): $E(tariff) = 0.0$.
- The alternative (H_a): $E(tariff) \neq 0.0$.

- Only for this type of hypothesis we can also check the absolute value of the $E(Tariff)$ against the SE:

  - 90% confident that is not zero if

  $$\frac{|E(Tariff)|}{1.65} > SE$$

  - 95% confident that is not zero if

  $$\frac{|E(Tariff)|}{1.96} > SE$$

  - 99% confident that is not zero if

  $$\frac{|E(Tariff)|}{2.56} > SE$$