Tutorial 07, Michaelmas Term

Research Methods for Political Science (PO3600)

Stefan Müller

28 November 2017

Trinity College Dublin
http://muellerstefan.net/research-methods

Session Outline

1. Questions asked online
2. Gamma and lambda
3. Cross-tables in SPSS
4. Measurement error
5. Research proposal

1. Difference between confidence intervals and t-test? When shall we use which approach?

2. Calculating lambda and gamma by hand?

3. How can I create a dataset with multiple observations easily to check my results?

Feel free to use this service: https://tinyurl.com/questionspo3600.

Calculate Lambda

- Relationship between two nominal variables
- Question: Can we reduce the amount of error by introducing an explanatory variable?

We know distribution of party ID: Lab 50%, Con 30%, Lib Dem 20%

		Father's party identification			
		Lab	Con	Lib Dem	
Resp. party ID	Lab	45	2	3	50
	Con	5	23	2	30
	Lib Dem	10	5	5	20
		60	30	10	100

Father party ID	Our guess	Correct	Wrong
Labour	Labour	45	15
Conservative	Conservative	23	7
Lib Dem	Lib Dem	5	5
Total		73	27

Lambda: Example

- Original error was 50 (when guessing)
- With the additional information (father's party ID) we have 27 errors remaining

General formula: $Association = \frac{Original\ error - Remaining\ error}{Original\ error} = \frac{50-27}{50} = .46$

A value of lambda $= 0.46$ means that the number of errors of prediction can be reduced by 46% if the IV, father's party identification, is know.

Calculate Gamma

- Measure for relationship between two variables (e.g. time spent studying and grade)
- Values between -1 and 1: -1 strong negative relationship; 0 no relationship; $+1$ strong positive relationship

	Time Spent Studying	
Grades	Minimal	Extensive
Bad	20	5
Good	6	10

- Find number of concordant pairs, N_c
- Find number of disconcordant pairs, N_d

Concordant pairs

	Time Spent Studying	
Grades	Minimal	Extensive
Bad	**20**	5
Good	6	**10**

$N_c = 10 \times 20 = 200$

Gamma: Calculate Discordant Pairs

	Time Spent Studying	
Grades	Minimal	Extensive
Bad	20	**5**
Good	**6**	10

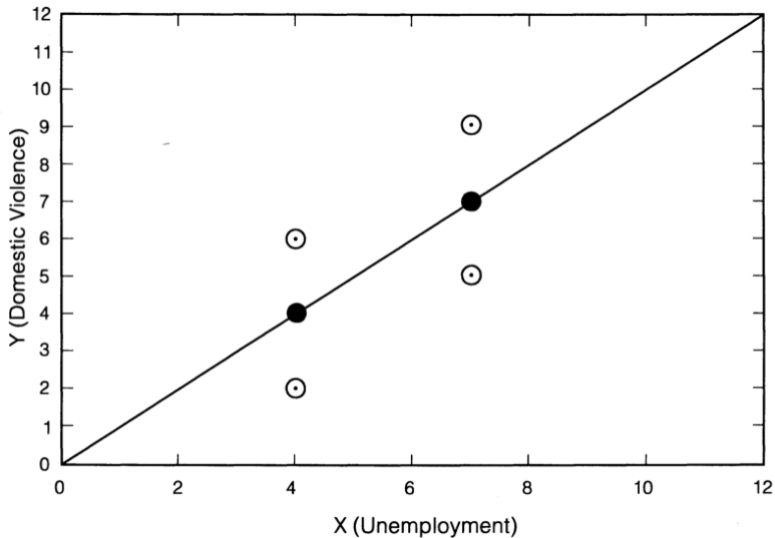$N_c = 10 \times 20 = 200$

$N_d = 6 \times 5 = 30$

Gamma: Insert values into formula

$N_c = 10 \times 20 = 200$

$N_d = 6 \times 5 = 30$
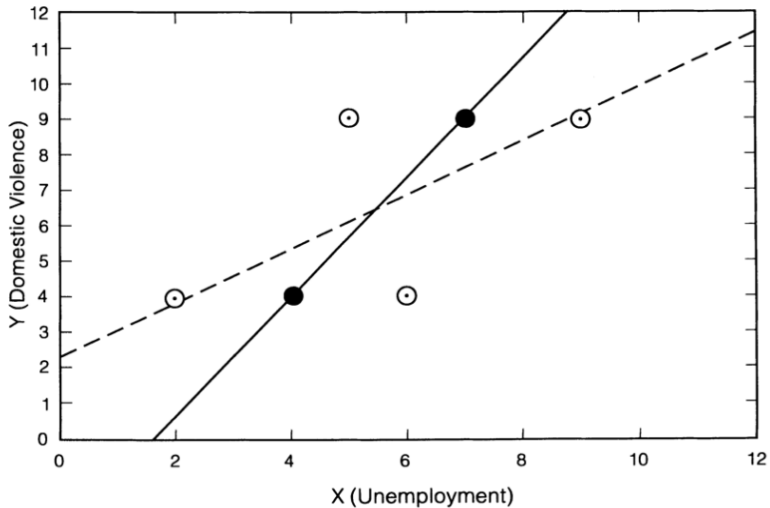
$\gamma = \frac{N_c - N_d}{N_c + N_d}$

$\gamma = \frac{200 - 30}{200 + 30} = 0.73$

Measurement Error

1. Download the two following datasets:
   - `ire_01.sav`: http://tinyurl.com/censusire01
   - `ire_02.sav`: http://tinyurl.com/censusire02
2. Import both datasets into SPSS
3. Sort both datasets ascending by `county_id`
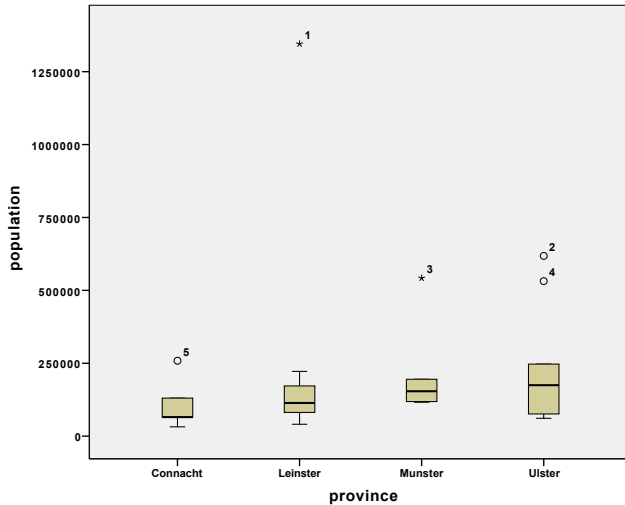4. Merge datasets (Data → Merge Files → Add variables)

Exercise: Summary Statistics

1. Use the merged datasets to create summary statistics of all continuous variables
   (Analyse → Descriptive Statistics → Frequencies)
2. Interpret the variable distributions with your neighbour.
3. Are there information we can leave out?

Exercise: Create a Boxplot

1. Use the merged datasets to create a "simple boxplot"
2. Put `province` on the *Category Axis* and insert `population` as the *Variable*.
3. Interpret the plot. Which observations are outliers?
4. Modify the plot (white background), rename axes manually.
5. Export and save the plot as a pdf file.

SPSS Output: Boxplot (bad)

SPSS Output: Boxplot (better)