Tutorial 7

Research Methods for Political Science - PO3110

Andrea Salvi

12 November 2018

Trinity College Dublin,
https://andrsalvi.github.io/research-methods/

Table of contents

About HMW3

Systematic and Random Measurement Errors

Systematic:

- Consistently overestimate values for certain types of units (e.g. students always over-reporting their income)
- Biased inferences
- Examples: quantitative text analysis; misreporting/consistent misunderstanding; sample selection bias

Systematic and Random Measurement Errors

Systematic:

- Consistently overestimate values for certain types of units (e.g. students always over-reporting their income)
- Biased inferences
- Examples: quantitative text analysis; misreporting/consistent misunderstanding; sample selection bias

Random:

- Random fluctuations without a clear pattern
- More uncertainty in results, only sometimes biased inferences
- Examples: human coding (if not depending on codebook)

Other measures of association: λ and γ

Lambda - λ

- Relationship between two nominal variables

Lambda - λ

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

Lambda - λ

- Relationship between two nominal variables
- Can we reduce the amount of error by introducing an explanatory variable?
- Lambda is a PRE measure and its value has fairly direct interpretation.

Lambda - λ

- Relationship between two nominal variables
- Can we reduce the amount of error by introducing an explanatory variable?
- Lambda is a PRE measure and its value has fairly direct interpretation.
- **PRE** stands for *Proportional Reduction in Error*

Lambda - λ

- Relationship between two nominal variables
- Can we reduce the amount of error by introducing an explanatory variable?
- Lambda is a PRE measure and its value has fairly direct interpretation.
- **PRE** stands for *Proportional Reduction in Error*
- Lambda tells us the improvement in predicting Y while taking X into account.

Lambda - λ

- Relationship between two nominal variables
- Can we reduce the amount of error by introducing an explanatory variable?
- Lambda is a PRE measure and its value has fairly direct interpretation.
- **PRE** stands for *Proportional Reduction in Error*
- Lambda tells us the improvement in predicting Y while taking X into account.
- $0 \leq \lambda \leq 1$
- Useful to compare the strength of bi-variate relationships.

Lambda- λ

- Association $= \frac{\text{Original Error} - \text{Remaining Error}}{\textit{OriginalError}}$

Lambda- λ

- Association $= \frac{\text{Original Error} - \text{Remaining Error}}{OriginalError}$
- $\lambda = \frac{E_1 - E_2}{E_1}$

Lambda- λ

- Association $= \frac{\text{Original Error} - \text{Remaining Error}}{OriginalError}$
- $\lambda = \frac{E_1 - E_2}{E_1}$
- E_1 = (N – the largest row total)
- E_2 = (For each column, subtract the largest cell frequency from its column total and then add the differences together)

Practical Example [1]

Efficiency/Authoritarianism	Low	High	Total
Low	10	12	22
High	17	5	22
Total	27	17	44

[1] From Professor Patricia Pakvis material

Practical Example [1]

Efficiency/Authoritarianism	Low	High	Total
Low	10	12	22
High	17	5	22
Total	27	17	44

- E_1 = N – largest row total = 44 – 22 = 22

[1] From Professor Patricia Pakvis material

Practical Example [1]

Efficiency/Authoritarianism	Low	High	Total
Low	10	12	22
High	17	5	22
Total	27	17	44

- E_1 = N – largest row total = 44 – 22 = 22
- E_2 = For each column, subtract the largest cell frequency from the col. total and add together = = (27 – 17) + (17 – 12) = 10 + 5 = 15

[1] From Professor Patricia Pakvis material

Practical Example [1]

Efficiency/Authoritarianism	Low	High	Total
Low	10	12	22
High	17	5	22
Total	27	17	44

- E_1 = N – largest row total = 44 – 22 = 22
- E_2 = For each column, subtract the largest cell frequency from the col. total and add together = = (27 – 17) + (17 – 12) = 10 + 5 = 15
- $\lambda = \frac{22-15}{22} = 0.32$

[1] From Professor Patricia Pakvis material

Practical Example [1]

Efficiency/Authoritarianism	Low	High	Total
Low	10	12	22
High	17	5	22
Total	27	17	44

- E_1 = N – largest row total = 44 – 22 = 22
- E_2 = For each column, subtract the largest cell frequency from the col. total and add together = = (27 – 17) + (17 – 12) = 10 + 5 = 15
- $\lambda = \frac{22-15}{22} = 0.32$
- Error reduced by 32%

[1] From Professor Patricia Pakvis material

Practice in SPSS

- Download data with party ID:
  `https://tinyurl.com/datapartyid`

Practice in SPSS

- Download data with party ID: `https://tinyurl.com/datapartyid`
- Create a cross-table

Practice in SPSS

- Download data with party ID: `https://tinyurl.com/datapartyid`
- Create a cross-table
- Calculate Lambda

- Measure for relationship between two ordinal variables (e.g. time spent studying and grade)

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

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

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

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

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

- Find number of concordant pairs, N_c

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

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

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

Concordant pairs

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

$N_c = 10 \times 20 = 200$

Discordant pairs

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

$N_c = 10 \times 20 = 200$ $N_d = 6 = 30$

- $N_c = 10 \times 20 = 200$
- $N_d = 6 = 30$

- $N_c = 10 \times 20 = 200$
- $N_d = 6 = 30$
- $\gamma = \frac{N_c - N_d}{N_c + N_d}$

- $N_c = 10 \times 20 = 200$
- $N_d = 6 = 30$
- $\gamma = \frac{N_c - N_d}{N_c + N_d}$
- $\gamma = \frac{200 - 30}{200 + 30} = 0.73$

Furthering your project

Team up and discuss (some of) the following aspects:

1. Research question + relevance
2. Theoretical argument + hypothesis
3. Type of data + operationalisation of variables
4. Ways of analysing your data

I am available for further questions/feedback!