Tutorial 1 HT

Research Methods for Political Science - PO3110

Andrea Salvi

29 January 2019

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

Table of contents

Administrative Stuff

Focus of Tutorials

- Furthering what we did in class;
- Deepening the understanding of Quants developed for PO3110;
- Hands-on SPSS;
- Real-world applications of theories, concepts and quantitative methods;
- Revision of Homeworks;
- Q&A;
- **NB: tutorials do not replace the lectures!**;

Assessment

1. 60% of mark based on end-of-term exam (covers methods and statistics).
2. 1 papers counting 16% (Deadline: 12/04 @ 11:59pm). Work will be done *in groups* submitting joint papers.
3. 4 homework exercises (4 per term), worth 20% of overall mark. Submit online via Blackboard (Turnit in Integration) *on the Monday evening (11:59pm) preceding the tutorial session*.
4. Tutorial participation is worth 4% of your overall mark. They should also attend the presentation sessions (last two weeks of class).
    - Absence policy: **Two unexcused absences in tutorials and 1 in the presentation will be tolerated. Beyond that, the student will receive a zero for participation**.

Late Submissions

- 5 points per day will be taken off your mark on assignments submitted late without a valid excuse (capped at 30 points for the paper).
- As for HMWs, no submissions will be accepted after Tuesdays (midday) following the due date (grade of 0)

Practical Information

Assignments and Blackboard/Turnitin

All the work should be submitted on Blackboard/Turnitin

- LaTeX, Word/Open Office and submitted as **PDFs** (Screen-shots are not sufficients!)
- If you include tables, do not use a screen-shots, but use the “export” function from SPSS. Please save figures appropriately in high resolution.
- Statistical Software: SPSS. You can use alternatives such as R or STATA if you want, but not Excel!
- Please do include the syntax from whichever software you are using!

Important Dates

Available on the Syllabus. Google Calendar on the tutorials' website.

- Week 4: HW 1 (next Monday!)
- Week 6: HW 2 (18 February)
- Week 9: HW 3 (11 March)
- Week 11: HW 4 (25 March)

Paper:

- 12/04 @ 11:59pm

Support

- I am happy to receive your feedback at any time! (content, teaching and tutorial style, too fast/slow?)
    - Short Surveys?
- Online Resources: `http://andrsalvi.github.io/research-methods`
- Questions:
    1. salvia@tcd.ie
    2. Slack Channel
    3. In class!
    4. Office Hours

Review of Uni-variate and Bi-variate Stats

Measures of Central Tendency

- Convenient way to describe a variable through a single number
- Gives us a sense of the where to locate the *"centre"* of the distribution
- Measures:
    1. Mode (at least Nominal variables)
    2. Median (at least Ordinal Variables)
    3. Mean (Interval Ratio)

Practical Calculations of Central Tendency

- Mode:
  - The most frequent value in a distribution.
  - In a more elegant way: ”the value that is most likely to be sampled”.
- Median:
  - The value at the midpoint of a distribution.
  - Data need to be ordered!
  - If the number of observations is *odd*: value at position $\frac{n+1}{2}$
  - If the number of observations is *even*: average between the value at position $\frac{n}{2}$ and the value at position $\frac{n+1}{2}$
- Mean:
  - Estimate mean: $\bar{x} = \frac{\sum x}{n}$
  - We need an interval ratio!
  - Influenced by outliers, while mode and median are not.
  - Always ”internal” with regards to the interval.
  - Remember the difference between μ and $\bar{x}$

Measures of Dispersion

- **Standard deviation:** spread of the sample;
- **Standard error of the mean:** spread of the means of many samples. That is, standard deviation of the sampling distribution;
- **Central limit theorem:** The mean of a large number of random samples will be normally distributed, regardless of the underlying distribution of that variable. That is, the **sampling distribution will take the form of a normal distribution!**

Calculating Dispersion

1. Estimate Mean: $\bar{x} = \frac{\sum x}{n}$
2. Sum of Squared Errors (SS): $\sum(x - \bar{x})^2$
3. Estimate Variance: $\sigma^2 = \frac{SS}{n-1} = \frac{\sum(x-\bar{x})^2}{n-1}$
4. Estimate Standard Deviation: $\sigma = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}} = \sqrt{\sigma^2}$
    - Remember that it is **NEVER** negative.
    - It can be 0 though in case of a uniform distribution.
5. Estimate standard error of the mean: $sd(\bar{X}) = \frac{\sigma}{\sqrt{n}}$

Confidence Intervals

- Confidence Interval: CI = $\bar{x} \pm z * sd(\bar{X})$

Confidence Intervals

- Confidence Interval: CI = $\bar{x} \pm z * sd(\bar{X})$
- Confidence level determined through z. 95% confidence interval has a z-score of 1.96, 90% has one of 1.645 and 99% has a z-score of 2.58.

Confidence Intervals

- Confidence Interval: CI = $\bar{x} \pm z * sd(\bar{X})$
- Confidence level determined through z. 95% confidence interval has a z-score of 1.96, 90% has one of 1.645 and 99% has a z-score of 2.58.
- **Interpretation:** For a given statistic calculated from a sample, *the confidence interval is a range of values around that statistic that are believed to contain, with a certain probability, the true value of that statistic (population value)*

One Sample T-test

- State the null-hypothesis and the alternative hypothesis.

One Sample T-test

- State the null-hypothesis and the alternative hypothesis.
- Evaluate whether it is a one-tailed t-test of a two-tailed one.
- Choose a level of significance (say 95%IC)

One Sample T-test

- State the null-hypothesis and the alternative hypothesis.
- Evaluate whether it is a one-tailed t-test of a two-tailed one.
- Choose a level of significance (say 95%IC)
- Calculate test statistic

One Sample T-test

- State the null-hypothesis and the alternative hypothesis.
- Evaluate whether it is a one-tailed t-test of a two-tailed one.
- Choose a level of significance (say 95%IC)
- Calculate test statistic
- $t = \frac{\text{observed value - expected value under } H_0}{\text{standard error}}$

One Sample T-test

- State the null-hypothesis and the alternative hypothesis.
- Evaluate whether it is a one-tailed t-test of a two-tailed one.
- Choose a level of significance (say 95%IC)
- Calculate test statistic
- $t = \frac{\text{observed value - expected value under } H_0}{\text{standard error}}$
- $t = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$

One Sample T-test

- State the null-hypothesis and the alternative hypothesis.
- Evaluate whether it is a one-tailed t-test of a two-tailed one.
- Choose a level of significance (say 95%IC)
- Calculate test statistic
- $t = \frac{\text{observed value - expected value under } H_0}{\text{standard error}}$
- $t = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$
- Compare the t with the critical value.

One Sample T-test

- State the null-hypothesis and the alternative hypothesis.
- Evaluate whether it is a one-tailed t-test of a two-tailed one.
- Choose a level of significance (say 95%IC)
- Calculate test statistic
- $t = \frac{\text{observed value - expected value under } H_0}{\text{standard error}}$
- $t = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$
- Compare the t with the critical value.
- If the absolute value of the t-test statistics is greater than the critical value, then the difference is significant. Otherwise it is not!

χ Squared

- H_0: x is independent upon y

χ Squared

- H_0: x is independent upon y
- H_1: x is dependant upon y

χ Squared

- H_0: x is independent upon y
- H_1: x is dependant upon y
- Find critical value:

χ Squared

- H_0: x is independent upon y
- H_1: x is dependant upon y
- Find critical value:
  - In order to do that we resort to a χ^2 table

χ Squared

- H_0: x is independent upon y
- H_1: x is dependant upon y
- Find critical value:
  - In order to do that we resort to a χ^2 table
  - We need to know *the area in the tail* and *our degrees of freedom*

χ Squared

- H_0: x is independent upon y
- H_1: x is dependant upon y
- Find critical value:
    - In order to do that we resort to a χ^2 table
    - We need to know *the area in the tail* and *our degrees of freedom*
    - $df = (\text{rows} - 1) * (\text{columns} - 1)$

χ Squared

- H_0: x is independent upon y
- H_1: x is dependant upon y
- Find critical value:
  - In order to do that we resort to a χ^2 table
  - We need to know *the area in the tail* and *our degrees of freedom*
  - $df = (\text{rows} - 1) * (\text{columns} - 1)$
- Test Statistic: $\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$

χ Squared

- H_0: x is independent upon y
- H_1: x is dependant upon y
- Find critical value:
  - In order to do that we resort to a χ^2 table
  - We need to know *the area in the tail* and *our degrees of freedom*
  - $df = (\text{rows} - 1) * (\text{columns} - 1)$
- Test Statistic: $\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$
  - Calculate the expected frequencies:

χ Squared

- H_0: x is independent upon y
- H_1: x is dependant upon y
- Find critical value:
  - In order to do that we resort to a χ^2 table
  - We need to know *the area in the tail* and *our degrees of freedom*
  - $df = (\text{rows} - 1) * (\text{columns} - 1)$
- Test Statistic: $\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$
  - Calculate the expected frequencies:
  - $f_e = \frac{\text{row margin} * \text{column margin}}{\text{total}}$

χ Squared

- H_0: x is independent upon y
- H_1: x is dependant upon y
- Find critical value:
  - In order to do that we resort to a χ^2 table
  - We need to know *the area in the tail* and *our degrees of freedom*
  - $df = (\text{rows} - 1) * (\text{columns} - 1)$
- Test Statistic: $\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$
  - Calculate the expected frequencies:
  - $f_e = \frac{\text{row margin} * \text{column margin}}{\text{total}}$
- Evaluate whether our test-statistic is in the rejection area or not.

χ Squared

- H_0: x is independent upon y
- H_1: x is dependant upon y
- Find critical value:
  - In order to do that we resort to a χ^2 table
  - We need to know *the area in the tail* and *our degrees of freedom*
  - $df = (\text{rows} - 1) * (\text{columns} - 1)$
- Test Statistic: $\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$
  - Calculate the expected frequencies:
  - $f_e = \frac{\text{row margin} * \text{column margin}}{\text{total}}$
- Evaluate whether our test-statistic is in the rejection area or not.

Cramer's V

- Chi squared does not tell you how strong a relationship is, only whether it is statistically significant.

Cramer's V

- Chi squared does not tell you how strong a relationship is, only whether it is statistically significant.
- If N is large, you are likely to find a significant relationship (but it might be a weak one).

Cramer's V

- Chi squared does not tell you how strong a relationship is, only whether it is statistically significant.
- If N is large, you are likely to find a significant relationship (but it might be a weak one).
- We need a measure of association! Let's take Cramers' V.
- $V = \sqrt{\frac{\chi^2}{N * \text{MIN of r-1, c-1}}}$

Cramer's V

- Chi squared does not tell you how strong a relationship is, only whether it is statistically significant.
- If N is large, you are likely to find a significant relationship (but it might be a weak one).
- We need a measure of association! Let's take Cramers' V.
- $V = \sqrt{\frac{\chi^2}{N * \text{MIN of r-1, c-1}}}$

Other measures:

Remember λ and γ? Check them out on the slides from MT7!

From variance to co-variance

- $\sigma^2 = \frac{\sum(x-\bar{x})^2}{n-1} = \frac{\sum(x-\bar{x})(x-\bar{x})}{n-1}$

From variance to co-variance

- $\sigma^2 = \frac{\sum(x-\bar{x})^2}{n-1} = \frac{\sum(x-\bar{x})(x-\bar{x})}{n-1}$
- $\sigma_{xy} = \frac{\sum(x-\bar{x})(y-\bar{y})}{n-1}$

Interpretation:

- $\sigma_{xy} = 0$: variables not related

From variance to co-variance

- $\sigma^2 = \frac{\sum(x-\bar{x})^2}{n-1} = \frac{\sum(x-\bar{x})(x-\bar{x})}{n-1}$
- $\sigma_{xy} = \frac{\sum(x-\bar{x})(y-\bar{y})}{n-1}$

Interpretation:

- $\sigma_{xy} = 0$: variables not related
- $\sigma_{xy} > 0$: variables vary in the same direction

From variance to co-variance

- $\sigma^2 = \frac{\sum(x-\bar{x})^2}{n-1} = \frac{\sum(x-\bar{x})(x-\bar{x})}{n-1}$
- $\sigma_{xy} = \frac{\sum(x-\bar{x})(y-\bar{y})}{n-1}$

Interpretation:

- $\sigma_{xy} = 0$: variables not related
- $\sigma_{xy} > 0$: variables vary in the same direction
- $\sigma_{xy} < 0$: variables vary in the opposite direction

From variance to co-variance

- $\sigma^2 = \frac{\sum(x-\bar{x})^2}{n-1} = \frac{\sum(x-\bar{x})(x-\bar{x})}{n-1}$
- $\sigma_{xy} = \frac{\sum(x-\bar{x})(y-\bar{y})}{n-1}$

Interpretation:

- $\sigma_{xy} = 0$: variables not related
- $\sigma_{xy} > 0$: variables vary in the same direction
- $\sigma_{xy} < 0$: variables vary in the opposite direction

Problem of co-variation? It depends on the scales of the variables of interest!

From variance to co-variance

- $\sigma^2 = \frac{\sum(x-\bar{x})^2}{n-1} = \frac{\sum(x-\bar{x})(x-\bar{x})}{n-1}$
- $\sigma_{xy} = \frac{\sum(x-\bar{x})(y-\bar{y})}{n-1}$

Interpretation:

- $\sigma_{xy} = 0$: variables not related
- $\sigma_{xy} > 0$: variables vary in the same direction
- $\sigma_{xy} < 0$: variables vary in the opposite direction

Problem of co-variation? It depends on the scales of the variables of interest!

- $\sigma_x\sigma_y \leq \sigma_{xy} \leq \sigma_x\sigma_y$

From variance to co-variance

- $\sigma^2 = \frac{\sum(x-\bar{x})^2}{n-1} = \frac{\sum(x-\bar{x})(x-\bar{x})}{n-1}$
- $\sigma_{xy} = \frac{\sum(x-\bar{x})(y-\bar{y})}{n-1}$

Interpretation:

- $\sigma_{xy} = 0$: variables not related
- $\sigma_{xy} > 0$: variables vary in the same direction
- $\sigma_{xy} < 0$: variables vary in the opposite direction

Problem of co-variation? It depends on the scales of the variables of interest!

- $\sigma_x\sigma_y \leq \sigma_{xy} \leq \sigma_x\sigma_y$
- Solution?

Correlation

Divide the co-variation by the standard deviations of x and y.

Correlation

Divide the co-variation by the standard deviations of x and y.

- correlation $= \frac{\text{covariance}}{\text{sd x} \times \text{sd y}}$

Correlation

Divide the co-variation by the standard deviations of x and y.

- correlation $= \frac{\text{covariance}}{\text{sd x} \times \text{sd y}}$
- $r = \frac{\sigma_{xy}}{\sigma_x \times \sigma_y}$

Correlation

Divide the co-variation by the standard deviations of x and y.

- correlation $= \frac{\text{covariance}}{\text{sd x} \times \text{sd y}}$
- $r = \frac{\sigma_{xy}}{\sigma_x \times \sigma_y}$
- $-1 \leq r \leq 1$

Hands-on

Research Design Challenge!

Since we did not have time back in MT:

Research Design Challenge!

Since we did not have time back in MT: Team up with your colleagues and download the following dataset from James D. Fearon and David D. Laitin, ”Ethnicity, Insurgency, and Civil War,” American Political Science Review 97, 1 (March 2003): 75-90.

- `https://tinyurl.com/method-conflict`

1. Let’s look at the data

Research Design Challenge!

Since we did not have time back in MT: Team up with your colleagues and download the following dataset from James D. Fearon and David D. Laitin, ”Ethnicity, Insurgency, and Civil War,” American Political Science Review 97, 1 (March 2003): 75-90.

- `https://tinyurl.com/method-conflict`

1. Let’s look at the data
2. Come up with a research question

Research Design Challenge!

Since we did not have time back in MT: Team up with your colleagues and download the following dataset from James D. Fearon and David D. Laitin, ”Ethnicity, Insurgency, and Civil War,” American Political Science Review 97, 1 (March 2003): 75-90.

- `https://tinyurl.com/method-conflict`

1. Let’s look at the data
2. Come up with a research question
3. Identify a DV and an IV

Research Design Challenge!

Since we did not have time back in MT: Team up with your colleagues and download the following dataset from James D. Fearon and David D. Laitin, ”Ethnicity, Insurgency, and Civil War,” American Political Science Review 97, 1 (March 2003): 75-90.

- `https://tinyurl.com/method-conflict`

1. Let’s look at the data
2. Come up with a research question
3. Identify a DV and an IV
4. Use plots/tables to describe your data and the relationship you propose

Research Design Challenge!

Since we did not have time back in MT: Team up with your colleagues and download the following dataset from James D. Fearon and David D. Laitin, ”Ethnicity, Insurgency, and Civil War,” American Political Science Review 97, 1 (March 2003): 75-90.

- `https://tinyurl.com/method-conflict`

1. Let's look at the data
2. Come up with a research question
3. Identify a DV and an IV
4. Use plots/tables to describe your data and the relationship you propose
5. Identify a suitable test for your design (among those we reviewed today)!

I am available for further questions/feedback!