Tutorial 8

Research Methods for Political Science - PO3110

Andrea Salvi

19 & 20 November 2019

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

Table of contents

Correlation by hand

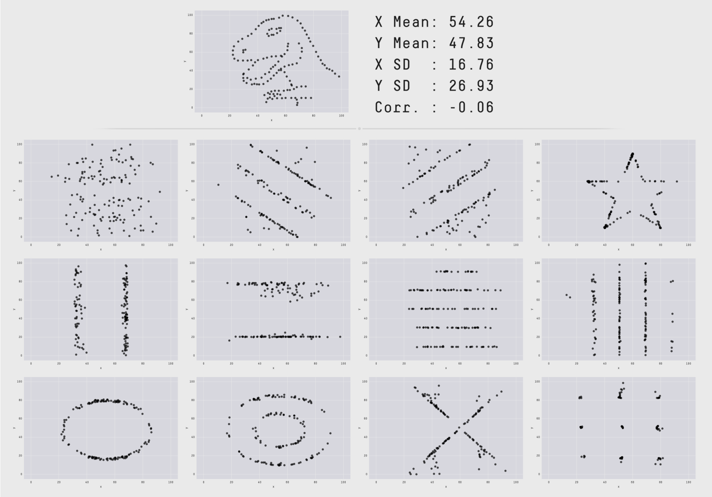
Preliminary Steps

1. Correlation and co-variation measure the association between two interval-ratio variables

Preliminary Steps

1. Correlation and co-variation measure the association between two interval-ratio variables
2. Before you start, always create a scatter-plot!

Behold! The datasaur!



[1]

[1]https://www.autodeskresearch.com/publications/samestats

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

x	y
1	1
2	3
4	5
5	7

Steps:

1. Calculate mean of each variable:

Hands-on

x	y
1	1
2	3
4	5
5	7

Steps:

1. Calculate mean of each variable:
2. Calculate standard deviations of x and y: $\sigma = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$
3. Calculate co-variance: $\sigma_{xy} = \frac{\sum(x-\bar{x})(y-\bar{y})}{n-1}$

Hands-on

x	y
1	1
2	3
4	5
5	7

Steps:

1. Calculate mean of each variable:
2. Calculate standard deviations of x and y: $\sigma = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$
3. Calculate co-variance: $\sigma_{xy} = \frac{\sum(x-\bar{x})(y-\bar{y})}{n-1}$
4. Divide co-variance by the product of standard deviations!

Hands-on

1. Calculate mean of each variable: $\bar{x} = 3$, $\bar{y} = 4$

Hands-on

1. Calculate mean of each variable: $\bar{x} = 3$, $\bar{y} = 4$
2. Calculate standard deviations: σ_x = 1.83 and σ_y = 2.58

Hands-on

1. Calculate mean of each variable: $\bar{x} = 3$, $\bar{y} = 4$
2. Calculate standard deviations: σ_x = 1.83 and σ_y = 2.58
3. Calculate co-variance: σ_{xy} = $((13) \times (14) + (23) \times (34) + (43) \times (54) + (53) \times (74))/(41) = 4.66$

Hands-on

1. Calculate mean of each variable: $\bar{x} = 3$, $\bar{y} = 4$
2. Calculate standard deviations: σ_x = 1.83 and σ_y = 2.58
3. Calculate co-variance: σ_{xy} = $((13) \times (14) + (23) \times (34) + (43) \times (54) + (53) \times (74))/(41) = 4.66$
4. Divide co-variance by the product of standard deviations: $4.66/(1.83 \times 2.58)$

Hands-on

1. Calculate mean of each variable: $\bar{x} = 3$, $\bar{y} = 4$
2. Calculate standard deviations: σ_x = 1.83 and σ_y = 2.58
3. Calculate co-variance: σ_{xy} = $((13) \times (14) + (23) \times (34) + (43) \times (54) + (53) \times (74))/(41) = 4.66$
4. Divide co-variance by the product of standard deviations: $4.66/(1.83 \times 2.58)$

Correlation, $r = 0.99$

Correlation in SPSS

Correlation in SPSS

x	y
1	1
2	3
4	5
5	7

Correlation in SPSS

x	y
1	1
2	3
4	5
5	7

1. Insert dataset into SPSS

Correlation in SPSS

x	y
1	1
2	3
4	5
5	7

1. Insert dataset into SPSS
2. Create scatter-plot of x and y

Correlation in SPSS

x	y
1	1
2	3
4	5
5	7

1. Insert dataset into SPSS
2. Create scatter-plot of x and y
3. Estimate correlation

Significance of r_{xy}

Let's check manually!

- The sampling distribution of r is approximately normal (but bounded at -1.0 and +1.0) when N is large

Significance of r_{xy}

Let's check manually!

- The sampling distribution of r is approximately normal (but bounded at -1.0 and +1.0) when N is large
- It distributes t when N is small!

Significance of r_{xy}

Let's check manually!

- The sampling distribution of r is approximately normal (but bounded at -1.0 and +1.0) when N is large
- It distributes t when N is small!
- $DF = N - 2$

Significance of r_{xy}

Let's check manually!

- The sampling distribution of *r* is approximately normal (but bounded at -1.0 and +1.0) when N is large
- It distributes *t* when N is small!
- $DF = N - 2$
- $t = r\sqrt{\frac{n-2}{1-r^2}}$

Significance of r_{xy}

Let's check manually!

- The sampling distribution of *r* is approximately normal (but bounded at -1.0 and +1.0) when N is large
- It distributes *t* when N is small!
- $DF = N - 2$
- $t = r\sqrt{\frac{n-2}{1-r^2}}$
- $H_0 : r = 0$

Project Work

Which test should I use?

- Ultimately depend on your data!

Which test should I use?

- Ultimately depend on your data!
- When in doubt, have a look at this link: `https://stats.idre.ucla.edu/other/mult-pkg/whatstat/`

Which test should I use?

- Ultimately depend on your data!
- When in doubt, have a look at this link: `https://stats.idre.ucla.edu/other/mult-pkg/whatstat/`
- Further resources have been included the tutorial website!

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!