Tutorial 7 HT

Research Methods for Political Science - PO3110

Andrea Salvi

19 March 2019

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

Table of contents

Some things from last time:

Standardized Regression Coefficients

- Standardised beta coefficients are all measured in standard deviations (can range between -1 and +1), instead of the units of the variables: It is possible to compare size of coefficients to one another.

Standardized Regression Coefficients

- Standardised beta coefficients are all measured in standard deviations (can range between -1 and +1), instead of the units of the variables: It is possible to compare size of coefficients to one another.
- Difference between the regular coefficients and the standardized coefficients is the units of measurement.

Standardized Regression Coefficients

- Standardised beta coefficients are all measured in standard deviations (can range between -1 and +1), instead of the units of the variables: It is possible to compare size of coefficients to one another.
- Difference between the regular coefficients and the standardized coefficients is the units of measurement.
- Raw coefficient: A one-unit increase/decrease in X is predicted to increase/decrease Y by xyz units.

Standardized Regression Coefficients

- Standardised beta coefficients are all measured in standard deviations (can range between -1 and +1), instead of the units of the variables: It is possible to compare size of coefficients to one another.
- Difference between the regular coefficients and the standardized coefficients is the units of measurement.
- Raw coefficient: A one-unit increase/decrease in X is predicted to increase/decrease Y by xyz units.
- Standardised coefficient: A one standard deviation increase/decrease is predicted to increase/decrease Y by xyz standard deviations

Standardized Regression Coefficients

- Standardised beta coefficients are all measured in standard deviations (can range between -1 and +1), instead of the units of the variables: It is possible to compare size of coefficients to one another.
- Difference between the regular coefficients and the standardized coefficients is the units of measurement.
- Raw coefficient: A one-unit increase/decrease in X is predicted to increase/decrease Y by xyz units.
- Standardised coefficient: A one standard deviation increase/decrease is predicted to increase/decrease Y by xyz standard deviations

Parametric and Non-parametric Tests

- **Parametric test:** specific assumptions (e.g. normality) are made about the population parameter is known as parametric test. The t-test, for instance, rests on assumption that variable is normally distributed.

Parametric and Non-parametric Tests

- **Parametric test:** specific assumptions (e.g. normality) are made about the population parameter is known as parametric test. The t-test, for instance, rests on assumption that variable is normally distributed.
- **Non-parametric test:** researcher has no idea regarding the population parameter/test does not require the population's distribution to be denoted by specific parameters

Parametric and Non-parametric Tests

- Perform a Shapiro-Wilk Normality Test (commonly used for testing normality assumption) (Analyze … Descriptives … Explore … Plots)

Parametric and Non-parametric Tests

- Perform a Shapiro-Wilk Normality Test (commonly used for testing normality assumption) (Analyze … Descriptives … Explore … Plots)
- What does the Sig. column under Shapiro-Wilk tell us? What is the null hypothesis?

Parametric and Non-parametric Tests

- Perform a Shapiro-Wilk Normality Test (commonly used for testing normality assumption) (Analyze … Descriptives … Explore … Plots)
- What does the Sig. column under Shapiro-Wilk tell us? What is the null hypothesis?
- Null hypothesis: variable not statistically significantly different from a normal distribution

Parametric and Non-parametric Tests

- Perform a Shapiro-Wilk Normality Test (commonly used for testing normality assumption) (Analyze … Descriptives … Explore … Plots)
- What does the Sig. column under Shapiro-Wilk tell us? What is the null hypothesis?
- Null hypothesis: variable not statistically significantly different from a normal distribution
- Rule: p-value < 0.05: reject null hypothesis that data come from a normally distributed population (but be careful with the interpretation (see lecture notes))

Parametric and Non-parametric Tests

- Perform a Shapiro-Wilk Normality Test (commonly used for testing normality assumption) (Analyze … Descriptives … Explore … Plots)
- What does the Sig. column under Shapiro-Wilk tell us? What is the null hypothesis?
- Null hypothesis: variable not statistically significantly different from a normal distribution
- Rule: p-value < 0.05: reject null hypothesis that data come from a normally distributed population (but be careful with the interpretation (see lecture notes))

Parametric and Non-parametric Tests

- Does what a t-test is supposed to test, but when distribution of two samples deviate from a normal distribution

Parametric and Non-parametric Tests

- Does what a t-test is supposed to test, but when distribution of two samples deviate from a normal distribution
- Mann-Whitney test rank orders all scores, determines the rank of each subject, and then computes the average rank for the two groups.

Parametric and Non-parametric Tests

- Does what a t-test is supposed to test, but when distribution of two samples deviate from a normal distribution
- Mann-Whitney test rank orders all scores, determines the rank of each subject, and then computes the average rank for the two groups.
- In SPSS: Analyze … Nonparametric … Legacy Dialogs … 2-Independent Samples Test

Parametric and Non-parametric Tests

- Does what a t-test is supposed to test, but when distribution of two samples deviate from a normal distribution
- Mann-Whitney test rank orders all scores, determines the rank of each subject, and then computes the average rank for the two groups.
- In SPSS: Analyze … Nonparametric … Legacy Dialogs … 2-Independent Samples Test

Probabilities

If we want to understand logistic regression, we need to know the basics about probabilities, odds and the logarithm of odds.

- Probabilities: Simply the likelihood that something will happen.

- Probability of 0.2 of rain = there is a 20% chance of rain

Probabilities

Let's say we have a fair die (6 faces). We are dealing with a discrete probability distribution X. We call our variable a discrete random variable That is a variable that can take on any value from a discrete set of values.

- Finite set of possible values $X = 1, 2, 3, 4, 5, 6$

Probabilities

Let's say we have a fair die (6 faces). We are dealing with a discrete probability distribution X. We call our variable a discrete random variable That is a variable that can take on any value from a discrete set of values.

- Finite set of possible values $X = 1, 2, 3, 4, 5, 6$
- Thus $x_i \in 1, 2, 3, 4, 5, 6$

Probabilities

Let's say we have a fair die (6 faces). We are dealing with a discrete probability distribution X. We call our variable a discrete random variable That is a variable that can take on any value from a discrete set of values.

- Finite set of possible values $X = 1, 2, 3, 4, 5, 6$
- Thus $x_i \in 1, 2, 3, 4, 5, 6$
- What is the probability of obtaining each x_i?

Probabilities

Let's say we have a fair die (6 faces). We are dealing with a discrete probability distribution X. We call our variable a discrete random variable That is a variable that can take on any value from a discrete set of values.

- Finite set of possible values $X = 1, 2, 3, 4, 5, 6$
- Thus $x_i \in 1, 2, 3, 4, 5, 6$
- What is the probability of obtaining each x_i?
- The "mantra" is: $\frac{\text{Favourable Cases}}{\text{Possible Cases}}$

Probabilities

Let's say we have a fair die (6 faces). We are dealing with a discrete probability distribution X. We call our variable a discrete random variable That is a variable that can take on any value from a discrete set of values.

- Finite set of possible values $X = 1, 2, 3, 4, 5, 6$
- Thus $x_i \in 1, 2, 3, 4, 5, 6$
- What is the probability of obtaining each x_i?
- The "mantra" is: $\frac{\text{Favourable Cases}}{\text{Possible Cases}}$
- What is $P(x_i = 7)$?

Probabilities

Let's say we have a fair die (6 faces). We are dealing with a discrete probability distribution X. We call our variable a discrete random variable That is a variable that can take on any value from a discrete set of values.

- Finite set of possible values $X = 1, 2, 3, 4, 5, 6$
- Thus $x_i \in 1, 2, 3, 4, 5, 6$
- What is the probability of obtaining each x_i?
- The "mantra" is: $\frac{\text{Favourable Cases}}{\text{Possible Cases}}$
- What is $P(x_i = 7)$?
- What is $P(1 \leq x_i \leq 1)$?

Odds

Odds: Ratio of the probability that an event will occur divided by the probability that an even will not occur.

Odds

Odds: Ratio of the probability that an event will occur divided by the probability that an even will not occur.

$$\text{Odds} = \frac{P(rain)}{P(\text{not rain})} = \frac{0.2}{0.8} = \frac{1}{4} = 0.25 \tag{1}$$

Odds

Odds: Ratio of the probability that an event will occur divided by the probability that an even will not occur.

$$\text{Odds} = \frac{P(rain)}{P(\text{not rain})} = \frac{0.2}{0.8} = \frac{1}{4} = 0.25 \tag{1}$$

Important: Probabilities always range between 0 and 1, but odds may be ≥ 1.

Odds

Odds: Ratio of the probability that an event will occur divided by the probability that an even will not occur.

$$\text{Odds} = \frac{P(rain)}{P(\text{not rain})} = \frac{0.2}{0.8} = \frac{1}{4} = 0.25 \tag{1}$$

Important: Probabilities always range between 0 and 1, but odds may be ≥ 1. An 80% probability of rain has odds of $\frac{0.8}{0.2} = 4$. A 50% chance of rain has odds of 1.

Logit

Logit: A logit is the natural logarithm (ln) of the odds. If there is a 20chance of rain, then there is a logit of:

Logit

Logit: A logit is the natural logarithm (ln) of the odds. If there is a 20chance of rain, then there is a logit of:

- $ln(0.25) = 1.386$

Logit

Logit: A logit is the natural logarithm (ln) of the odds. If there is a 20chance of rain, then there is a logit of:

- $ln(0.25) = 1.386$

When we run a logistic regression, we estimate the natural logarithm of the odds.

Example: Voting for Trump

$$ln\left[\frac{P(\text{Vote Trump})}{P(\text{Not Vote Trump})}\right] = \beta_0 + \beta_1 \times partyid + \beta_2 \times education \quad (2)$$

Example: Voting for Trump

$$ln\left[\frac{P(\text{Vote Trump})}{P(\text{Not Vote Trump})}\right] = \beta_0 + \beta_1 \times partyid + \beta_2 \times education \quad (2)$$

To change this from the ln to the odds, we need to take the exponential function of the natural logarithm function:

Example: Voting for Trump

$$ln\left[\frac{P(\text{Vote Trump})}{P(\text{Not Vote Trump})}\right] = \beta_0 + \beta_1 \times partyid + \beta_2 \times education \quad (2)$$

To change this from the ln to the odds, we need to take the exponential function of the natural logarithm function:

$$\frac{P(\text{Vote Trump})}{P(\text{Not Vote Trump})} = e^{\beta_0} + e^{\beta_1 \times partyid} + e^{\beta_2 \times education} \quad (3)$$

Example: Voting for Trump

$$ln\left[\frac{P(\text{Vote Trump})}{P(\text{Not Vote Trump})}\right] = \beta_0 + \beta_1 \times partyid + \beta_2 \times education \quad (2)$$

To change this from the ln to the odds, we need to take the exponential function of the natural logarithm function:

$$\frac{P(\text{Vote Trump})}{P(\text{Not Vote Trump})} = e^{\beta_0} + e^{\beta_1 \times partyid} + e^{\beta_2 \times education} \quad (3)$$

If we want to get the (predicted) probability of voting for trump, we need to rewrite the formula again.

$$\frac{P(\text{Vote Trump})}{P(\text{Not Vote Trump})} = \frac{1}{1 + e^{-\beta_0} + e^{-\beta_1 \times partyid} + e^{-\beta_2 \times education}} \quad (4)$$

Logistic Regression in SPSS: Download the data

Data: `https://tinyurl.com/anes16sav` Codebook: `https://www.electionstudies.org/wp-content/uploads/2016/02/anes_pilot_2016_CodebookUserGuide.pdf`

Logistic Regression in SPSS

Dependent variable: turnout12
Independent variables: birthyr, gender, newsint

Logistic Regression in SPSS

Dependent variable: turnout12
Independent variables: birthyr, gender, newsint

Logistic Regression in SPSS

Dependent variable: turnout12
Independent variables: birthyr, gender, newsint

1. Check how missing values are coded (and recode them to system-missing if necessary)

Logistic Regression in SPSS

Dependent variable: turnout12
Independent variables: birthyr, gender, newsint

1. Check how missing values are coded (and recode them to system-missing if necessary)
2. Recode turnout12 to turnout12 recoded: 0: definitely did not vote (original: 2); 1: definitely voted (original: 1); missing: not completely sure (original: 3)

Logistic Regression in SPSS

Dependent variable: turnout12
Independent variables: birthyr, gender, newsint

1. Check how missing values are coded (and recode them to system-missing if necessary)
2. Recode turnout12 to turnout12 recoded: 0: definitely did not vote (original: 2); 1: definitely voted (original: 1); missing: not completely sure (original: 3)
3. Reverse the coding of newsint and call the new variable newsint recoded

Logistic Regression in SPSS

Dependent variable: turnout12
Independent variables: birthyr, gender, newsint

1. Check how missing values are coded (and recode them to system-missing if necessary)
2. Recode turnout12 to turnout12 recoded: 0: definitely did not vote (original: 2); 1: definitely voted (original: 1); missing: not completely sure (original: 3)
3. Reverse the coding of newsint and call the new variable newsint recoded
4. Run the logit

Logistic Regression in SPSS

Dependent variable: turnout12
Independent variables: birthyr, gender, newsint

1. Check how missing values are coded (and recode them to system-missing if necessary)
2. Recode turnout12 to turnout12 recoded: 0: definitely did not vote (original: 2); 1: definitely voted (original: 1); missing: not completely sure (original: 3)
3. Reverse the coding of newsint and call the new variable newsint recoded
4. Run the logit

Rule of thumbs for coefficients

OR > 1: positive relationship

OR = 1: no relationship

OR < 1: negative relationship

Rule of thumbs for coefficients

OR > 1: positive relationship
OR = 1: no relationship
OR < 1: negative relationship

Interpret Odds Ratio in Logistic Regression:
`https://stats.idre.ucla.edu/other/mult-pkg/faq/general/faq-how-do-i-interpret-odds-ratios-in-logistic-regressi`