Tutorial 6 HT

Research Methods for Political Science - PO3110

Andrea Salvi

12 March 2019

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

Table of contents

1. Hands-on: ANES 2016 Pilot

Hands-on: ANES 2016 Pilot

Download the data

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

Task 1

Question 1 Run a linear regression with fttrump as the DV and four IVs (e.g. compromise, ftsci, equalpay, gender). General question: How do we proceed before running the regression?

Task 1

Question 1 Run a linear regression with fttrump as the DV and four IVs (e.g. compromise, ftsci, equalpay, gender). General question: How do we proceed before running the regression?

1. Check codebook

Task 1

Question 1 Run a linear regression with fttrump as the DV and four IVs (e.g. compromise, ftsci, equalpay, gender). General question: How do we proceed before running the regression?

1. Check codebook
2. Create descriptive statistics

Task 1

Question 1 Run a linear regression with fttrump as the DV and four IVs (e.g. compromise, ftsci, equalpay, gender). General question: How do we proceed before running the regression?

1. Check codebook
2. Create descriptive statistics
3. Recode missing values (into new variables)

Task 1

Question 1 Run a linear regression with fttrump as the DV and four IVs (e.g. compromise, ftsci, equalpay, gender). General question: How do we proceed before running the regression?

1. Check codebook
2. Create descriptive statistics
3. Recode missing values (into new variables)
4. Scatterplot ftsci_recoded over fttrump_recoded

Task 1

Question 1 Run a linear regression with fttrump as the DV and four IVs (e.g. compromise, ftsci, equalpay, gender). General question: How do we proceed before running the regression?

1. Check codebook
2. Create descriptive statistics
3. Recode missing values (into new variables)
4. Scatterplot ftsci_recoded over fttrump_recoded

Task 2

Question 2 Run and interpret the regression model (especially the unstandardised beta coefficients)

Task 2

Question 2 Run and interpret the regression model (especially the unstandardised beta coefficients) Let's save both the standardised and unstandardised residuals.

- Include an interpret the following:
  - SCATTERPLOT of ZRESID and ZPRED

Task 2

Question 2 Run and interpret the regression model (especially the unstandardised beta coefficients) Let’s save both the standardised and unstandardised residuals.

- Include an interpret the following:
  - SCATTERPLOT of ZRESID and ZPRED
  - Include COOK’s D and Durbin Watson Test

Task 2

Question 2 Run and interpret the regression model (especially the unstandardised beta coefficients) Let's save both the standardised and unstandardised residuals.

- Include an interpret the following:
    - SCATTERPLOT of ZRESID and ZPRED
    - Include COOK's D and Durbin Watson Test
    - Include collinearity Diagnostics, Save St and Un Residuals and Predicted Values

Task 2

Question 2 Run and interpret the regression model (especially the unstandardised beta coefficients) Let's save both the standardised and unstandardised residuals.

- Include an interpret the following:
  - SCATTERPLOT of ZRESID and ZPRED
  - Include COOK's D and Durbin Watson Test
  - Include collinearity Diagnostics, Save St and Un Residuals and Predicted Values
- Create an interaction term (e.g. between gender and equalpay)

Task 2

Question 2 Run and interpret the regression model (especially the unstandardised beta coefficients) Let's save both the standardised and unstandardised residuals.

- Include an interpret the following:
  - SCATTERPLOT of ZRESID and ZPRED
  - Include COOK's D and Durbin Watson Test
  - Include collinearity Diagnostics, Save St and Un Residuals and Predicted Values
- Create an interaction term (e.g. between gender and equalpay)
- Plot the interaction

Task 2

Question 2 Run and interpret the regression model (especially the unstandardised beta coefficients) Let's save both the standardised and unstandardised residuals.

- Include an interpret the following:
  - SCATTERPLOT of ZRESID and ZPRED
  - Include COOK's D and Durbin Watson Test
  - Include collinearity Diagnostics, Save St and Un Residuals and Predicted Values
- Create an interaction term (e.g. between gender and equalpay)
- Plot the interaction

Task 2

Question 2 Run and interpret the regression model (especially the unstandardised beta coefficients) Let's save both the standardised and unstandardised residuals.

- Include an interpret the following:
  - SCATTERPLOT of ZRESID and ZPRED
  - Include COOK's D and Durbin Watson Test
  - Include collinearity Diagnostics, Save St and Un Residuals and Predicted Values
- Create an interaction term (e.g. between gender and equalpay)
- Plot the interaction

Task 3

Question 3 Run a bootstrap regression with fttrump as the DV and four IVs (e.g. compromise, ftsci, equalpay, gender).

Task 3

Question 3 Run a bootstrap regression with fttrump as the DV and four IVs (e.g. compromise, ftsci, equalpay, gender).

- If we have problems with the error distribution in terms of heteroskedasticity or non-normality of the error distribution, we can use a robust regression method instead.

Task 3

Question 3 Run a bootstrap regression with fttrump as the DV and four IVs (e.g. compromise, ftsci, equalpay, gender).

- If we have problems with the error distribution in terms of heteroskedasticity or non-normality of the error distribution, we can use a robust regression method instead.
- Bootstrapping: Random sampling with replacement from an observed dataset in order to estimate properties of the sampling distribution
- Remember to tick ”Correct”: This corrects for the fact that our bootstrap samples might be, on average, slightly different from our sample.

Task 3

Question 3 Run a bootstrap regression with fttrump as the DV and four IVs (e.g. compromise, ftsci, equalpay, gender).

- If we have problems with the error distribution in terms of heteroskedasticity or non-normality of the error distribution, we can use a robust regression method instead.
- Bootstrapping: Random sampling with replacement from an observed dataset in order to estimate properties of the sampling distribution
- Remember to tick ”Correct”: This corrects for the fact that our bootstrap samples might be, on average, slightly different from our sample.
- Keep ”simple” for the sample as you have not extracted more complex samples (like stratified samples etc.).

Task 3

Question 3 Run a bootstrap regression with fttrump as the DV and four IVs (e.g. compromise, ftsci, equalpay, gender).

- If we have problems with the error distribution in terms of heteroskedasticity or non-normality of the error distribution, we can use a robust regression method instead.
- Bootstrapping: Random sampling with replacement from an observed dataset in order to estimate properties of the sampling distribution
- Remember to tick ”Correct”: This corrects for the fact that our bootstrap samples might be, on average, slightly different from our sample.
- Keep ”simple” for the sample as you have not extracted more complex samples (like stratified samples etc.).
- We estimate 2000 regressions and their betas. SPSS shows the mean.

Task 3

Question 3 Run a bootstrap regression with fttrump as the DV and four IVs (e.g. compromise, ftsci, equalpay, gender).

- If we have problems with the error distribution in terms of heteroskedasticity or non-normality of the error distribution, we can use a robust regression method instead.
- Bootstrapping: Random sampling with replacement from an observed dataset in order to estimate properties of the sampling distribution
- Remember to tick ”Correct”: This corrects for the fact that our bootstrap samples might be, on average, slightly different from our sample.
- Keep ”simple” for the sample as you have not extracted more complex samples (like stratified samples etc.).
- We estimate 2000 regressions and their betas. SPSS shows the mean.

A note on bootstrapping

- NOTICE: There is no need for bootsrapping in regression analysis if the OLS assumptions are met – In such cases, OLS estimates are the most unbiased and efficient of estimates.

A note on bootstrapping

- NOTICE: There is no need for bootsrapping in regression analysis if the OLS assumptions are met – In such cases, OLS estimates are the most unbiased and efficient of estimates.
- There are situations, however, when we cannot satisfy the assumptions and thus other methods are more helpful.

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