

Research Methods for Political Science PO3110 (TCD)

HT: Tutorial 6 - Week 7

Letícia Meniconi Barbabela

University College Dublin,
<https://github.com/letmeni/research-methods>

10-11 March 2020

Today's topics

- Categorical predictors (Recoding) (Stats HT05);
- Interaction terms (Stats HT05);
- Bootstrapping (Stats HT05);
- Non-parametric tests (Stats HT06).

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`, `educ`, `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 Recode categorical predictors

Task 1

Question 1 Run a linear regression with `fttrump` as the DV and four IVs (e.g. `compromise`, `educ`, `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 Recode categorical predictors

Task 1

Question 1 Run a linear regression with `fttrump` as the DV and four IVs (e.g. `compromise`, `educ`, `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 Recode categorical predictors

Task 1

Question 1 Run a linear regression with `fttrump` as the DV and four IVs (e.g. `compromise`, `educ`, `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 Recode categorical predictors

Task 1

Question 1 Run a linear regression with `fttrump` as the DV and four IVs (e.g. `compromise`, `educ`, `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 Recode categorical predictors

Task 2

Run and interpret the regression model Let's save both the standardised and unstandardised residuals.

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

Task 2

Run and interpret the regression model Let's save both the standardised and unstandardised residuals.

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

Task 2

Run and interpret the regression model Let's save both the standardised and unstandardised residuals.

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

Task 2

Run and interpret the regression model Let's save both the standardised and unstandardised residuals.

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

Task 2

Run and interpret the regression model Let's save both the standardised and unstandardised residuals.

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

Task 2

- Effect of gender on feelings towards trump might not be just a shift in the intercept;
- Being one category more supportive of gender equality could have a different effect on feelings towards Trump depending on whether the person is a man or a woman.
- How do you account for that in the model?

Task 2

- Effect of gender on feelings towards trump might not be just a shift in the intercept;
- Being one category more supportive of gender equality could have a different effect on feelings towards Trump depending on whether the person is a man or a woman.
- How do you account for that in the model?

Task 2

- Effect of gender on feelings towards trump might not be just a shift in the intercept;
- Being one category more supportive of gender equality could have a different effect on feelings towards Trump depending on whether the person is a man or a woman.
- How do you account for that in the model?

Interaction

- Create an interaction term (e.g. between gender and discselfsex; or between gender and equal pay)
- Include the constitutive elements (why?)
- What can we conclude?
- Are there any issues?

Interaction

- Create an interaction term (e.g. between gender and discselfsex; or between gender and equal pay)
- Include the constitutive elements (why?)
 - What can we conclude?
 - Are there any issues?

Interaction

- Create an interaction term (e.g. between gender and discselfsex; or between gender and equal pay)
- Include the constitutive elements (why?)
- What can we conclude?
- Are there any issues?

Interaction

- Create an interaction term (e.g. between gender and discselfsex; or between gender and equal pay)
- Include the constitutive elements (why?)
- What can we conclude?
- Are there any issues?

Interpreting the Interaction

- What changes in the interpretation of coefficients once we include an interaction?
- How do we interpret the main effects?
- What does the presence of a significant interaction indicates?
- Plot the interaction

Interpreting the Interaction

- What changes in the interpretation of coefficients once we include an interaction?
- How do we interpret the main effects?
- What does the presence of a significant interaction indicates?
- Plot the interaction

Interpreting the Interaction

- What changes in the interpretation of coefficients once we include an interaction?
- How do we interpret the main effects?
- What does the presence of a significant interaction indicates?
- 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`).

- 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.

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.

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 bootstrapping 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.

A note on bootstrapping

- NOTICE: There is no need for bootstrapping 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.
- 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

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

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.
- **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

- **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

Shapiro-Wilk Normality Test: testing normality assumption

- Open the "Pork barrel" data from my Github page;
- Perform a Shapiro-Wilk Normality Test on all scale variables: (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: $p\text{-value} < 0.05$: reject null hypothesis that data come from a normally distributed population (but be careful with the interpretation (see lecture notes))

Shapiro-Wilk Normality Test: testing normality assumption

- Open the "Pork barrel" data from my Github page;
- Perform a Shapiro-Wilk Normality Test on all scale variables: (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: $p\text{-value} < 0.05$: reject null hypothesis that data come from a normally distributed population (but be careful with the interpretation (see lecture notes))

Shapiro-Wilk Normality Test: testing normality assumption

- Open the "Pork barrel" data from my Github page;
- Perform a Shapiro-Wilk Normality Test on all scale variables: (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: $p\text{-value} < 0.05$: reject null hypothesis that data come from a normally distributed population (but be careful with the interpretation (see lecture notes))

Shapiro-Wilk Normality Test: testing normality assumption

- Open the "Pork barrel" data from my Github page;
- Perform a Shapiro-Wilk Normality Test on all scale variables: (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: $p\text{-value} < 0.05$: reject null hypothesis that data come from a normally distributed population (but be careful with the interpretation (see lecture notes))

Shapiro-Wilk Normality Test: testing normality assumption

- Open the "Pork barrel" data from my Github page;
- Perform a Shapiro-Wilk Normality Test on all scale variables: (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: $p\text{-value} < 0.05$: reject null hypothesis that data come from a normally distributed population (but be careful with the interpretation (see lecture notes))

Shapiro-Wilk Normality Test: testing normality assumption

- Open the "Pork barrel" data from my Github page;
- Perform a Shapiro-Wilk Normality Test on all scale variables: (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: $p\text{-value} < 0.05$: reject null hypothesis that data come from a normally distributed population (but be careful with the interpretation (see lecture notes))

Rank-sum test

- 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.
- Let's check if the distribution of "funding after" is the same across parties
- In SPSS: Analyze ... Nonparametric ... Independent samples

Rank-sum test

- 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.
- Let's check if the distribution of "funding after" is the same across parties
- In SPSS: Analyze ... Nonparametric ... Independent samples

Rank-sum test

- 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.
- Let's check if the distribution of "funding after" is the same across parties
- In SPSS: Analyze ... Nonparametric ... Independent samples

Rank-sum test

- 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.
- Let's check if the distribution of "funding after" is the same across parties
- In SPSS: Analyze ... Nonparametric ... Independent samples

Wilcoxon's Signed rank test

- Let's check if "funding before" is different from "funding after"
- In SPSS: Analyze ... Nonparametric ... Related Samples

Wilcoxon's Signed rank test

- Let's check if "funding before" is different from "funding after"
- In SPSS: Analyze ... Nonparametric ... Related Samples