Tutorial 08, Hilary Term

Research Methods for Political Science (PO3600)

Stefan Müller

20 March 2018

Trinity College Dublin
http://muellerstefan.net/research-methods

Outline

1. Standardised regression coefficients
2. Non-parametric tests
3. Logistic regression

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

- 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?

- 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, week 8 HT))

Mann-Whitney 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.
- In SPSS: *Analyze ... Nonparametric ... Legacy Dialogs ... 2-Independent Samples Test*

Standardised Regression Coefficients

- Standardised beta coefficients are all measured in standard deviations (can range between -1 and $+1$), instead of the units of the variables: $\longrightarrow$ 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.

Mathematics of Logistic Regression: Probabilities

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

Probabilities:

Mathematics of Logistic Regression: 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

Mathematics of Logistic Regression: Odds

Odds:

Mathematics of Logistic Regression: Odds

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

$$Odds = \frac{prob(rain)}{prob(no\ rain)} = \frac{0.2}{0.8} = \frac{1}{4} = 0.25$$

Important: Probabilities always range between 0 and 1, but odds may be greater than 1. An 80% probability of rain has odds of $0.8/0.2 = 4.0$.

A 50% change of rain has odds of 1.

Mathematics of Logistic Regression: Logit

Logit:

Mathematics of Logistic Regression: Logit

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

$$ln(0.25) = -1.386\ldots$$

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

Example: Voting for Trump

$$ln\left[\frac{prob(vote\ \ Trump)}{prob(not\ \ vote\ \ Trump)}\right] = b_0 + b_1 \times partyID + b_2 \times education$$

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

$$\frac{prob(vote\ \ Trump)}{prob(not\ \ vote\ \ Trump)} = e^{b_0} + e^{b_1 \times partyID} + e^{b_2 \times education}$$

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

$$prob(vote\ \ Trump) = \frac{1}{1 + e^{-b_0} + e^{-b_1 \times partyID} + e^{-b_2 \times education}}$$

Running a Logistic Regression in SPSS

Dependent variable: `turnout12`

Independent variables: `birthyr`, `gender`, `newsint`

Tasks:

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`

Output and Interpretation

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	newsint_recoded	.852	.088	93.001	1	.000	2.343
	birthyr	-.039	.006	49.537	1	.000	.962
	gender	-.364	.175	4.323	1	.038	.695
	Constant	76.522	10.990	48.482	1	.000	1.710E+33

a. Variable(s) entered on step 1: newsint_recoded, birthyr, gender.

$OR > 1$: 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-regression/