

METHODS OF QUANTITATIVE
DATA ANALYSIS
MSR Course, 2011-2012

Harry B.G. Ganzeboom
Lecture 3: More on Regression
April 20-23 2012

AGENDA

- How we should / could continue.
- Assignment 2:
 - REGRESS and UNIANOVA
 - Dummy variable coding / reference
 - High and low R²
 - Interaction models
- How to make tables.
- How does regression control?
- Non-linearities in linear models.

Non-linear relationships

- Linear model ($Y=a+b*X$) generally assume linear relationships, but give plenty of opportunities to model non-linear relationships:
 - Dummy regression (test of linearity)
 - Polynomial regression
 - Spline regression
 - Loess regression

Test of linearity

- Categorize your X-variable into a set of dummy variables.
- Compare continuous linear model with the discrete (dummy variables) model: the F-test (comparing linear correlation with the correlation ration) shows deviations from linearity.
- In SPSS the procedure is standard in Means(ONYWAY ANOVA), but it is easily implemented in Regress and UniAnova (shifting between **with** and **by**).
- This approach tests non-linearity, but does not model it.

Polynomial regression

- You can also fit a polynomial (power) function to the $X \rightarrow Y$ relationship.
- $Y = B_0 + B_1 * X + B_2 * X * X + B_3 * X * X * X$ etc.
- In particular the model with the squared term is fairly popular. It implies a max or a min.
- Although not always necessary, one can reduce collinearity between polynomial terms by centering the X-variable (= subtract the mean).
- Evaluate the functional form by calculating the expected Y-values from the X-terms and then list or plot them.
- Polynomial models are somewhat inflexible, and often imply a max/min when you do not expect these or no min/max if you want it.

Spline regression

- (Linear) spline regression is a trick to fit a broken regression line to a certain trajectory.
- It is also called piecewise linear regression.
- It is rather flexible and particularly useful to model non-linear trends in time series data.
- Splines are available as `mk spline` in Stata, but it easily implemented in any regression program.

Spline recipe

- Choose knots.
- Construct spline vars X_1, X_2, X_3 etc.:
 - $X_1 = X$
 - $X_2 = X - \text{knot}_1$
 - $X_3 = X - \text{knot}_2$
- Regress Y upon the spline vars.
- This is also known as piecewise linear regression.

Evaluation of models

- Specification of non-linear models often requires lots of independent variables in complicated forms.
- For evaluation it is important to balance fit (R^2 , SS-model) with degrees of freedom used. Adj. R^2 and F-test for change are the appropriate tools.
- However, statistical information should also be balanced against substantive considerations: not every spike in your data is interpretable or plausible.