Extensions to the multiple linear regression model
(STATS544.2: Applied multiple linear regression)

Gunnar Stefansson

August 7, 2013

Dummy variables

Suppose a single change affects the response so that it increases by a constant from then on.

To describe such a simple change we can add a dummy variable to a regression.

This is done by defining a column of 0/1-values.

Examples: A pollution incident; different intercepts depending on sex. A regression with only a dummy variable is equivalent to a t-test. Dummy variables can be used to test whether slopes or intercepts or both are different for groups of subjects or time periods.

These dummy variables can be included in very many different ways.

Typical example: Data on Icelandic summer spawning herring, see http://tgax14.rhi.hi.is/html/data/biol/iceher.txt or other fish examples at http://tgax14.rhi.hi.is/html/data/biol/fish-samples/

It is important to understand the meaning of estimates depending on how the variables are included.

Factors

A factor is an independent variable which can only take on (few) distinct levels.

Two values: Like a dummy variable

Multiple: Can use many dummy variables, but packages do this automagically.

Example: Yield from a field will depend on location (farm).

Factors can also be used to test very generally whether a straight line is appropriate using a lack-of-fit test.

A model with only a single factor will give parameter estimates which are the sample mean for each group. If only these point estimates are needed, then tapply() can be used to compute means for each level.

Factors can also be used to test very generally whether a straight line is appropriate using a lack-of-fit test.

ANOVA

ANOVA is used to analyse a linear model with only factors.

Example...

ANCOVA

ANCOVA: Analysis of models with both continuous and discrete independent variables.

Example: two lines

Poisson regression, via case studies

Logistic regression, via case studies

Survival analysis (Coxph), via case studies