May 12, 2009
Morten Frydenberg
Eva Ørnbøl

# POSTGRADUATE COURSE IN LINEAR AND LOGISTIC REGRESSION
## Day 3

Consider the dataset: "`serumchol.dta`", which is a subset of the dataset `2.20.framingham.dta` used in Dupont. In this exercise, focus is on the dependent variable serum cholesterol (`scl`) and possible explanatory variables systolic blood pressure (`sbp`), diastolic blood pressure (`dbp`), Body Mass Index (`bmi`), and `sex` (`sex`=1 men, `sex`=2 women)..

1. Create a categorical variable from `bmi` according to the WHO definitions by
   `egen bmi_who=cut(bmi), at(10, 18.5, 25, 30, 60) label`
   Is the distribution of `scl` the same in the groups defined by `bmi_who`?

2. Estimate a model (**Model 1**) with `scl` as the dependent variable and `sbp`, `bmi_who` and `sex` as the independent variables; `bmi_who` and `sex` should be entered as a categorical variable in the model with BMI<18.5 kg/m$^2$ and *men* as reference.
   Write down the estimated equation for the expected serum cholesterol.

3. Explain the coefficients for `sbp`, `bmi_who`=3 and `sex`=2.
   Write down the estimated relationship between the expected serum cholesterol and systolic blood pressure for a man with bmi=26.
   Write down the estimated relationship between the expected serum cholesterol and systolic blood pressure for a woman with bmi=26.
   Make a plot of the relationship between the expected serum cholesterol and systolic blood pressure for the eight different combinations of `sex` and `bmi_who`.
   Write down the estimated relationship between serum cholesterol and BMI for a man with a systolic blood pressure of 130 mmHg
   Make a plot of the relationship between the expected serum cholesterol and BMI for a man with a systolic blood pressure of 130 mmHg.

4. Find the expected value with 95% confidence interval for a subject with `sbp`=85, `sex`=2 and `bmi_who`=1.
   (Hint: use the `lincom` command).

5. Create a new variable `sbp2` equal to the square of `sbp`.
   Add `sbp2` to **Model 1** and estimate this model (**Model 2**).

Explain the coefficient of `sbp2`.
Find the expected value for `scl` with 95% confidence interval for a subject with values given in 4. Compare the result with the one you found in 4.

6. Estimate a model (**Model 3**) with `scl` as the dependent variable and `sbp`, `sex` and `bmi` as independent variables (that is, BMI in a non-categorized version).
   Make a plot of the relationship between the expected serum cholesterol and BMI for a man with systolic blood 130 mmHg. Decide from this and the estimates (from **Model 1** and **2**) whether BMI as a continuous variable is preferable to/as good as BMI as a categorized according to WHO.

We stick with **Model 1**.

7. Use the *explanatory* variables in **Model 1** as a basis for an investigation of whether the *dependent* variable would benefit from a transformation. (Hint: plots of the distribution of the residuals, residual versus fitted values, and residual versus independent variable should be made.)

In order to estimate a more realistic model possible interactions should perhaps be included.
Here we focus on two: an interaction between `sex` and `sbp` and an interaction between `sex` and `bmi_who`.

8. Estimate a model (**Model 3**) with ln(`scl`) as the dependent variable and `sbp`, `bmi_who`, `sex` and both interactions as independent variables.

9. Explain the coefficient to the interaction between `sex` and `sbp`.
   Test the hypothesis that the interaction is zero.

10. Explain the coefficient to the interaction between `sex` and `bmi_who`=2.
    Test the hypothesis that all coefficients to the interaction between `sex` and `bmi_who` are zero.