POSTGRADUATE COURSE IN
LINEAR AND LOGISTIC REGRESSION
Day 2
Morning exercises:

Part A

Back to the lung function data (`lung.dta.` and `lung.sav`) you looked at yesterday morning.

We will consider the multiple regression of `PEFR` on `height` and `woman`
(generate woman=(sex==1):

$$PEFR = \beta_0 + \beta_1 \cdot height + \beta_2 \cdot woman + E$$

1. What are the assumptions behind the model?
    Which of the assumptions, did you check (informally) yesterday morning?

2. Estimate the model.
    What is the interpretation of the estimates?

3. What is the difference (with CI) in the expected PEFR for a woman and a man both with height 170 cm?
    What is the difference (with CI) in the expected PEFR for a woman and a man both with height 160 cm?

4. Try to check some of the assumption behind the model.

5. Do any of the data points stick out?

Part B

Here we look at the model from the lecture today (`fram200.dta.` and `fram200.sav`).

$$\ln(sbp) = \beta_0 + \beta_1 \cdot age + \beta_2 \cdot woman + \beta_3 \cdot \ln(bmi) + E$$

6. Fit the model using 50 years as reference for `age` and 22 kg/m^2 for `bmi`.
    Give an interpretation of the estimated parameters.

7. Find (*without* confidence interval) the median `sbp` for a man, 55 year old with a bmi of 25 kg/m^2.

8. Check some of the aspects of the model, by plotting the residuals versus `age`, `ln(bmi)` and `sex`.
    What can you see in these plots? What have you checked?

Introduce (`age`-50)^2 to the model and fit it.

9. Comment on the estimates. How have they changed?
    Is the coefficient for (`age`-50)^2 statistical significant?
    Which assumption of the first model have you checked?