1.3 Exercises 3: Date 3.5.2013

1. Derive the relation between the F-statistic and the coefficient of determination.

2. Show that $e'e = y'My = u'Mu$ with $M = I - X(X'X)^{-1}X'$, $e = y - X\hat{\beta}$ the vector of residuals of the regression $y = X\beta + u$ with β the unknown parameter vector, u the unknown error vector and $\hat{\beta}$ the least squares estimator.

3. Let $y = X\beta + u$, $\hat{y} = X\hat{\beta}$ and $e = y - X\hat{\beta} = y - \hat{y}$, with $\hat{\beta} = (X'X)^{-1}X'y$ the least squares estimator. Show that the decomposition of the total sum of squares $y'y$ into the explained sum of squares $\hat{y}'\hat{y}$ and the residual sum of squares $e'e$ holds.

4. Use the data file **cps78.xls** Read the information in "Info". Estimate a multiple linear regression model

$$WAGE_i = \beta_0 + \beta_1 AGE_i + \beta_2 ED_i + \beta_3 FE_i + \beta_4 UNION_i + u_i \qquad (6)$$

   where ED stands for years of formal education, FE for female, and $UNION$ for union membership. Which values do you obtain for $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4)$? How do you interpret the estimated coefficients? For each coefficient find out if it is significantly different from zero (95% level).
   Plot the residuals against the years of education (ED). Group the residuals e_i according to the years of education and estimate the variance of the u_i within each group by computing the sample variance of the e_i in each group. Does the assumption of a constant variance σ^2 for all u_i seem justified?

5. Estimate the regression as in (6), but replace the dependent variable $WAGE_i$ by $LNWAGE_i$.

$$LNWAGE_i = \beta_0 + \beta_1 AGE_i + \beta_2 ED_i + \beta_3 FE_i + \beta_4 UNION_i + u_i \qquad (7)$$

   This model explains obviously the natural logarithm of $WAGE$. Which values do you obtain now for $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4)$? How do you now interpret the estimated coefficients in this equation? Repeat the estimation of group variances as in the previous exercise. Do you notice a difference?

6. Test in equation (7) if **all** slope coefficients simultaneously are different from zero against the alternative that at least one coefficient is non-zero.

7. Test if women earn significantly less than men. Formulate the null- and the alternative hypothesis and use a type one error probability of $\alpha = 5\%$.

8. Investigate in (7) how much hourly wage increase somebody would get on average for one more year of formal education. Calculate a 95%-confidence interval and interpret the result.

9. Test (at the 95% level) in the estimated equation (7) if union members receive a higher wage than non-union members.

10. Someone might argue that years of experience should also play a role for the personal wage. Add the variable EX to the model (7) and estimate the revised equation to find out about this issue. Test the coefficients against zero. Do you encounter any problem?