Econometrics

20 questions, each one on 1 point. 10 questions about a data set, 10 theoretical right/wrong answers. Be aware: this was exceptionally an open book exam, and without Stata. We just got the printed output. We got 3 hours.

Question 1
A data set: budget study in Belgium, >2000 respondents.

- *agerp* = age of the head of the household
- *totexp* = total expenses on non-durable goods in euro
- *npers* = number of persons in the household
- *nchild* = number of children in the household
- *nadult* = number of adults in the household
- *nchi1* = number of children age between 0 and 2
- *nchi2* = number of children age between 2 and 6
- *nchi3* = number of children age between 6 and 15
- *nchi4* = number of children age over 15
- *leis* = total expenses on leisure in euro

Regression model (1):

$$wleis_i = \beta_0 + \beta_1 agerp_i + \beta_2 agerp_i{}^2 + \beta_3 \log(totexp_i) + \beta_4 npers_i + u_i$$

Here, *wleis_i* = *leis_i / totexp_i*. So *wleis* is the ratio of leisure expenditures in the total expenditures.

1. Present the regression results for (1) in the standard way.
2. Interpret the R^2 and the *SER* for this specific regression.
3. If the age of the respondent would change from 20 to 30, how would *wleis* change?
4. Give an accurate interpretation for the coefficient of *log(totexp)*.
5. Based on this information, would you say that leisure is a luxurious good or a necessary good?
6. If *totexp* would be measured in thousands of euros instead of euros, we see that all coefficients, standard errors and the *SER* do not change except for B0. Explain how this comes.
7. Now consider model (2), where we estimate *nchild* and *nadult* instead of *npers*. Perform a *t*-test over whether the household composition (i.e., whether a member is an adult or a child) makes a difference for *wleis*.
8. In model (3), we replace *nchi* by *nchi1*, *nchi2*, *nchi3*, and *nchi4*. Interpret the results. Are they according to your expectations?
   (The results were negative for all nchi's, but less negative for older children.)
9. Perform a test to check whether all *nchi* coefficients are actually equal to zero.

10. Finally, you get the regression results of the residuals squared ($\hat{u}^2$) on all of the regressors. Based on these results, can you say whether there is homoskedasticity or heteroskedasticity?

Question 2

Are the following statements right or wrong? Explain briefly why.

1. Under the OLS assumptions, the standard error of the regression is a consistent estimator for the standard deviation of the error disturbance terms.
2. Consider $Y_i = \beta_0 + \beta_1 X_i + u_i$. If $u \sim N(\gamma, \sigma^2)$ exactly, then we have that $E(u|X) \neq 0$ and thus bias and inconsistency.
3. In regression, the *t* statistic is always exactly *t* distributed as long as the null hypothesis is true, since the error terms are exactly normally distributed.
4. For a joint hypothesis, we can choose between the *F* and *t* statistic. Both have the same results (notably, the same *p*-value).
5. When studying big data, when sample size goes to infinity, the mean square prediction error goes to zero, except if OLS is used as a prediction method.
6. For joint hypothesis testing we may choose between the t-test and the F-test. Their outcomes (in particular their p-values), however, are the same
7. suppose that the disturbance term in the equation

   $$Y_i = \beta_0 + \beta_1 X_i + u_i$$

   satisfies $u_i \sim N(\gamma, \sigma^2)$ where $\gamma > 0$. Then least-squares becomes problematic because one of the least-squares assumptions is that E(u|x)=0. As a result, we cannot estimate β_1
8. in prediction models with many predictions, the mean square prediction error goes to zero as the sample size of the estimation dataset goes to infinity, except when the estimation is carried out with OLS
9. if the number of observations doubles, we can expect R^2 to double and the standard errors of the estimated coefficients to become half as large
10. standard errors of the estimated coefficients are dimensionless: they are expressed as a percentage of the corresponding estimated coefficients