ECON 301

Optional HW

1) Suppose that classical least squares assumptions apply and the true value of intercept term α is 0 in following model:

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

Now estimate this model with and without an intercept term and compare the variance of the estimator of the slope term.

2) Suppose that regression model is $y_t = \mu + \varepsilon_t$ where $E[\varepsilon_t \mid x_t] = 0$,

$\text{cov}(\varepsilon_t, \varepsilon_s \mid x_t, x_s) = 0 \ \forall t \neq s$ but $\text{var}(\varepsilon_t \mid x_t) = \sigma^2 x_t^2,\ x_t > 0 \ \forall t$

a) Given a sample of observations on y_t and x_t what is the most efficient estimator of μ?,

b) Is this estimator unbiased? What is its variance?

c) What is the OLS estimator of μ?, and what is its variance?.

d) Prove that the estimator in part (a) is at least as efficient as the estimator in part (c).(bonus) (Hint: use Cauchy-Schwarz inequality)

3) Consider following linear regression model:

$$y_i = \beta_1 d_{i1} + \beta_2 d_{i2} + \varepsilon_i \ \forall i = 1, ..., n$$

Where $d_{i,1} = \begin{cases} 1 & if\ 1 \le \text{i} \le \text{N} \\ 0 & if\ \text{N+1} \le \text{i} \le \text{n} \end{cases}$ for some $1 < N < n$ and $d_{i2} = 1 - d_{i1} \ \forall i = 1, ..., n$

a) Derive OLS estimates $\hat{\beta}_1$ *and* $\hat{\beta}_2$ (this should be a function of *N* and *n*

b) Could $\hat{\beta}_1$ *and* $\hat{\beta}_2$ be estimated separately? (i.e regress y_i on d_{i1} to obtain $\hat{\beta}_1$ and regress y_i on d_{i2} to obtain $\hat{\beta}_2$)

4) Show that if regression equations contains a constant term, then $\frac{1}{N}\sum_{i=1}^{N} \hat{y}_i = \bar{y}$ thus,

sample mean of predicted y is equal to actual sample mean of y. where $\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i$.

5) A researcher investigating the determinants of juvenile delinquency has the following data for 2007 for a sample of 100 cities in a certain country: *A*, the number of arrests per 1,000 juveniles, defined as persons aged 14–18, in the city, *P*, the number of households per 1,000 in the city with incomes below the poverty line, and *S*, the number of single-parent households per 1,000 in the city. He is considering fitting the model

$$A = \beta_1 + \beta_2 P + \beta_3 S + u$$

where *u* is a disturbance term that may be assumed to satisfy the usual regression model assumptions. The correlation between *P* and *S* is 0.96. State what is correct, mistaken, confused or incomplete in the following statements, giving an explanation when the statement is not correct.

a) Multicollinearity does not cause the estimates of the coefficients to be biased but it does cause them to be inconsistent.
b) The standard errors will be biased, probably downwards.
c) The problem will be even worse if there is a high correlation between *A* and *P* or between *A* and *S*.
d) Mention and explain two ways of addressing the problem described above.

Empirical Question

An econometric analysis of wage determination is conducted using a sample of 114 workers from a town in India, in 1990. The statistical model is such that the dependent variable is weekly wage income in rupees, (W_i). The explanatory variables are age (AGE_i) of the worker defined in years and some qualitative factors such as gender, education and the type of the job with a distinction between permanent and temporary workers. The dummy variables used to explain these qualitative factors are as follows:

$D_{SEX\,i} = 1$ for male workers and 0 for female workers (otherwise).

DE_{2i} a dummy variable taking the value of 1 for workers who has the primary school diploma as the highest degree, and 0 otherwise.

DE_{3i} a dummy variable taking the value of 1 for workers who has the high school diploma as the highest degree, and 0 otherwise..

DE_{4i} a dummy variable taking the value of 1 for workers with an education level above high school diploma, 0 otherwise.

DPT_i A dummy variable taking a value of 1 for workers with permanent jobs and a value of for temporary workers (otherwise).

$$W_i = \beta_1 + \beta_2 AGE_i + \beta_3 D_{SEXi} + \beta_4 DE_{2i} + \beta_5 DE_{3i} + \beta_6 DE_{4i} + \beta_6 DPT_i + u_i$$

a) What is the base category and its estimated equation?
b) Is there a difference in the wages of female and male workers?
c) What is the intercept term of female worker who has primary school education?
d) Does education make any difference in wages?
e) How will you test the hypothesis that primary and high school education affect wages differently?

f) How will you test the hypothesis that having above high school education will bring less wages than having university education?
g) How will you test the hypothesis that temporary workers receive lower wages than permanent workers?
h) If you think that the effect of age on wages differ across male and female workers, how should you modify your statistical equation?

[Conduct formal hypothesis testing for each test. This should include null and alternative hypothesis, test statistics, comparison with critical values and decision and finally interpretation of the result]