Middle East Technical University **Spring 2014**
Department of Economics
ECON 301 - 01

Lecturer : Dr. Ozan Eruygur
Research Assistant: Mert Aşçıoğlu

PROBLEM SET I
SIMPLE REGRESSION MODEL

PROBLEM 1

Suppose a researcher wants to estimate the parameters in the simple population equation

$Y = \beta_0 + \beta_1 X + u$

The following data which were calculated from a sample of 10 observations are available.

$\Sigma Y = 580$	$\Sigma xy = -654$
$\Sigma X = 50$	$\Sigma y^2 = 7568$
$\Sigma XY = 2246$	$\Sigma x^2 = 60$
$\Sigma Y^2 = 41208$	
$\Sigma X^2 = 310$	

a. Obtain least square estimates of β_0 and β_1.
b. Compute R^2. Given an interpretation for the coefficients of determination.
c. Compute correlation coefficient between X and Y , between Y and $\hat{Y}$.
d. Test if the regression relation as a whole is significant (take $\alpha = 0.05$).
e. Test if the model must have an intercept term (take $\alpha = 0.05$).
f. Construct the 99 % confidence interval for β_1.
g. Predict the level of Y when X = 15.
h. Construct 95 % confidence intervals for E (Y) and Y when X = 15.
i. Test if $\beta_0 + \beta_1 = 0$ (take $\alpha = 0.05$).
j. Interpret the model you have estimated.
k. Do you think the model should be reformulated? Why or why not?

PROBLEM 2

A sample of 20 observations on price and the quantity supplied of a commodity A are obtained and following supply function is estimated.

$$Q_t^s = b_0 + b_1 P_t + u_t$$

where $E(u_t) = 0$ $\quad t = 1.........T$
$E(u_t^2) = \sigma^2$
$E(u_t u_s) = 0$ $\quad t \neq s$

$\Sigma X_t = 21.9$	$\Sigma (X_t - X)(Y_t - Y) = 106.4$	
$\Sigma (Y_t - Y)^2 = 86.9$	$\Sigma (X_t - X)^2 = 215.4$	$\Sigma Y_t = 186.2$

a. It is claimed that a unit increase in price of A will increase quantity supplied by 0.5 units. Test this hypothesis and based on the results of your test state whether you agree or disagree with the claim.

b. It is also claimed that supply function for commodity A must past through the origin. State whether you agree or disagree with this claim by performing the appropriate tests.
c. Calculate simple correlation coefficients between Q_t^s and $\hat{Q}_t^s$; P_t and Q_t^s .
d. Predict the amount of quantity supplied when $P_t = 30$.
e. Construct 90 % confidence interval for $E(Q_t^s)$

PROBLEM 3

Suppose Y = value of production per worker, X = wage rate, and the subscript t reforms to the t^{th} firm . We are interested in estimating $Y = a + bX + u$ for two industries. The least square results for industry A are ;

$$\underset{}{Y_t = -0.4 + \underset{(0.1)}{1.0}\, X_t} \qquad (T = 50)$$

For industry B the results are

$$Y_t = -0.3 + \underset{(0.1)}{0.8}\, X_t \qquad (T = 50)$$

The figures in the parenthesis are the standard errors. The two samples can be considered to be independent .

a . Show that $R^2_A = 2/3$ and $R^2_B = 4/7$.
b . Test the hypothesis that $b_A = b_B$ (take $\alpha = 0.05$).

PROBLEM 4

From the data concerning the number of accidents and the number of motor vehicles, the following quantities were estimated for the years 1968-1977 .

$\Sigma Y_t = 22$
$\Sigma X_t = 1723$
$\Sigma X_t Y_t = 4180.6$
$\Sigma Y_t^2 = 53.28$
$\Sigma X_t^2 = -328609$

where Y_t = Number of accidents in thousands in year t .
X_t = Number of motor vehicles in thousands in year t .

a. Estimate a linear regression equation (with an intercept term) making use of the information provided above.
b. Test if the regression relation as a whole is significant and interpret the estimated regression.
c. It is claimed that every increase of 100 in the number of motor vehicles will result in 15 extra accidents. Test this hypothesis and based on the results of your test, state whether you agree or disagree with this claim.
d. Given that the number of motor vehicles in 1981 will be 60 000, find the point and interval estimates for the number accidents in 1981.
e. Estimate the elasticity of accidents with respect to motor vehicles at the mean.

PROBLEM 5

Consider the following models:

Model I : $Y_t = b_0 + b_1 X_t + u_t$
Model II : $Y^*_t = a_0 + a_1 X^*_t + u_t$

where $X^*_t = (X_t - \bar{X}) / S_x$ and $Y^*_t = (Y_t - \bar{Y}) / S_y$
S_x = standard deviation of X in the sample
S_y = standard deviation of Y in the sample

a. Find the least square estimator of a_1.
b. Prove that $a_1 = b_1 (S_x / S_y)$

Note: Variable X* and Y* are known as Standardized variables. A variable is said to be standardized or in standard (deviation) units if it is expressed in terms of deviation from its mean value and divided by its sample standard deviation.

PROBLEM 6

From a sample of 10 families the following quantities were calculated:

$\Sigma C_t = 40$
$\Sigma Y_{d\,t} = 50$
$\Sigma C_t Y_{d\,t} = 236$
$\Sigma Y_{dt}^2 = 300$
$\Sigma C_t^2 = 208.5$

where C_t = Consumption level of the t^{th} family.
$Y_{d\,t}$ = Disposable income of the t^{th} family.

Assume that the consumption function is specified as

$C_t = b_0 + b_1 Y_{d\,t} + u_t$

a. Estimate b_0 , b_1 and the correlation coefficient .
b. Predict the level of consumption for the mean level of disposable income, and construct the 95 % confidence interval for your predicted value.
c. It is claimed that the marginal propensity to consume out of disposable income is 0.75
d. Test this hypothesis and based on the results of your test, state whether you agree or disagree with the claim (take $\alpha = 0.05$).
e. Is the regression relationship between consumption and disposable income significant as a whole? Explain your answer.

PROBLEM 7

A researcher wishes to estimate the parameters of the following model using OLS:

Model 1 $\Rightarrow$ $Y_t = bX_t + u_t$

However, this assistant confuses the dependent and the independent variables and instead estimates the following model using OLS:

Model 2 $\Rightarrow$ $X_t = bY_t + u_t$ and finds

$$\hat{X}_t = 12.17\ \hat{Y}_t \qquad R^2 = 0.993 \qquad SSR = 92.7 \qquad t_a = 35.38 \qquad T = 10$$

a. Using the findings for Model 2, estimate b and its standard error.
b. Find the correlation coefficient for the first model.
c. Test if b is significantly different from zero.

PROBLEM 8

The following data pertaining to the demand for food in the US, 1982-1991 is given;

YEARS	Q_d	P_d
1982	97.6	88.6
1983	97.2	91
1984	97.3	97.9
1985	96	102.3
1986	99.2	102.2
1987	100.3	102.5
1988	100.3	97
1989	104.1	95.8
1990	105.3	96.4
1991	107.6	100.3

Q_d = Food consumption per capita
P_d = Food prices at the retail level/cost of living index

a. Estimate the following linear demand function

   $Q_{dt} = \beta_0 + \beta_1 P_{dt} + u_t$

b. Estimate standard errors of β_0 and β_1.
c. Compute the coefficient of correlation between Q_d and P_d.
d. Compute R^2. Give an interpretation for the coefficients of determination.
e. Estimate the price elasticities of demand for 1989 and 1991.
f. Test if the demand function as a whole significant (take $\alpha = 0.05$).
g. Test if the slope coefficient is different from zero (take $\alpha = 0.05$).
h. Estimate the correlation coefficient between observed demand values and predicted demand values for food.
i. Construct the 99% confidence interval for β_0.
j. Consider the result of hypothesis testing above. Is it necessary to reformulate the model? Explain.

PROBLEM 9

You have the following data for estimating the relationship between Y and X.

Y	X
-2	-2
-1	0
0	1
1	0
2	1

First adapt the model $Y_t = \alpha + \beta X_t + u_t$

a. Compute least squares estimates of α and β.
b. What is an unbiased estimator of $Var(u_t)$?

Now adapt the model $X_t = \gamma + \delta Y_t + v_t$

c. Compute least square estimates of γ and δ .
d. Are the two fitted regression lines the same? Explain this result in terms of the least squares methodology.

PROBLEM 10

Consider the bivariate linear regression model ;

$Y_t = \alpha + \beta X_t + u_t \qquad t = 1,......,N$

in which it is assumed that, for all observations,

i. $E(u_t) = 0$ ii. $E(u_t^2) = \sigma^2$ iii. $E(u_i u_j) = 0 \quad i = j$ iv. X_t is nonstochastic
v. u_t is normally distributed random variables.

Which of the assumptions above are necessary for the least square estimator of β to be

a. Unbiased?
b. Best linear unbiased?
c. Consistent?

PROBLEM 11

Consider the following two alternative bivariate linear regression models;

(I) $Y_t = \beta_0 + \beta_1 X_t + u_t \qquad t = 1,......,N$
(II) $Y_t = \alpha_0 + \alpha_1 (X_t - \overline{X}) + v_t \qquad t = 1,......,N$

where u and v are random variables, β_0 ,β_1 ,α_0 and α_1 are unknown parameters. X_t are nonstochastic and; $\overline{X} = \dfrac{\sum X_i}{N}$.

a. Derive the least square estimator of α_1 in Model II and compare it with the least square estimator of β_1 in Model I.
b. Derive the least square estimator of α_0 in Model II and compare it with the least square estimator of β_0 in Model I.

PROBLEM 12

In the model $Y_t = \alpha + \beta X_t + u_t$, an estimate of β is obtained as follows :

$$\tilde{\beta} = (1/T - 1)\sum \frac{Y_t - Y_{t-1}}{X_t - X_{t-1}}$$

a. Give a geometric interpretation of $\tilde{\beta}$.

b. Show that $\tilde{\beta}$ is unbiased. Be sure to state the assumptions needed to prove this.

c. Without actually deriving the variance of $\tilde{\beta}$, argue why this estimator is inefficient relative to the OLS estimator of β .

PROBLEM 13

In the regression model $Y_t = \beta_0 + \beta_1 X_t + u_t$ let β be the OLS estimator of B.
Then $\hat{u}_t = Y_t - \beta_0 - \beta_1 X_t$ is the residual after removing the effect of X_t on Y_t . Show that X_t and $\hat{u}_t$ are uncorrelated.(i.e: $Cov(X_t , \hat{u}_t) = 0$)

PROBLEM 14

Show that the estimated coefficient can also be written as :

$$\hat{\beta} = \frac{Cov(X,Y)}{Var(X)}$$

PROBLEM 15

Show that the OLS estimator β can also be written as $r(S_y / S_x)$ where r is the sample correlation coefficient between X and Y, S_x^2 is the sample variance of X and S_y^2 is the sample variance of Y.

PROBLEM 16

Suppose; Y: value of production per worker , X: wage rate, and the subscript t refers to the t^{th} firm. We are interested in estimating Y = a + bX + u for two industries. The least square results for industry A are :

$$Y_t = -0.4 + \underset{(0.1)}{1.0}\, X_t \qquad (T = 50)$$

For industry B, the results are :

$$Y_t = -0.3 + \underset{(0.1)}{0.8}\, X_t \qquad (T = 50)$$

The figures in parenthesis are the standard errors. The two samples can be considered to be independent.

a. Show that $R^2_A = 2/3$ and $R^2_B = 4/7$.
b. Test the hypothesis that $b_A = b_B$.

PROBLEM 17

Given ; $X_t = a_0 + a_1 Y_t + u_t$

$$\hat{X} = 5.076 + 0.345\, Y_t$$

s.e. (2.006) (0.069)

and $\Sigma X_t = 12$ $\Sigma X_t^2 = 190$ $T = 25$ $R_r^2 = 0.95$

a. Construct 99% confidence intervals for a_0 and a_1.
b. Test the following hypothesis with $\alpha = 0.05$.

   i. $H_0 : a_0 = 0$
   ii. $H_0 : a_1 = 0$
   iii. $H_0 : a_0 = a_1 = 0$
   iv. $H_0 : a_1 = 0.4$ vs $H_A : a_1 > 0.4$
   v. $H_0 : a_1 = 4.5$ vs $H_A : a_1 < 4.5$

c. Consider the results of the first three hypotheses. Do they necessitate the reformulation of the model? Explain.
d. Assume that $Y_P = 2.5$. Construct 95% confidence interval for $E(X_P)$ and X_P .

PROBLEM 18

Consider the bivariate linear regression model;

$$Y_t = \beta X_t + u_t \quad t = 1,....,N$$

in which the $u_t \approx N(0,\sigma^2)$, $E(u_i ,u_j) = 0 \quad i \neq j$, and the X_t are nonstochastic.

a . Prove that the least square estimator of β in this model ($\hat{\beta}$) ;

$$\beta = \frac{\sum X_t Y_t}{\sum X_t^2}$$

b . Prove that the estimated variance of $\hat{\beta}$;

$$V\hat{a}r(\hat{\beta}) = \frac{\sum \hat{u}_t^2 / N - 1}{\sum X_t^2}$$

c . Demonstrate that the usual requirement that $\sum \hat{u}_t = 0$ does not hold in this model where $\hat{u}_t$ are the least squares residuals.

PROBLEM 19

For the single regression model :

$$Y_t = \beta X_t + u_t \quad t = 1,....,N$$

with X_t fixed and u_t distributed iid with zero and constant variance

a. Consider the estimator $\overline{\beta} = \frac{\overline{Y}}{\overline{X}}$ where $\overline{Y}$ designates the arithmetic mean. Show that $\overline{\beta}$ is linear and unbiased.
b. Suppose you employ M < N observations and apply OLS. Show this estimator is linear and unbiased but not of minimum variance.
c. Derive the minimum variance, linear estimator without using the condition of unbiasedness and interpret your result.

PROBLEM 20

Think about a model where an economic variable is expressed as its mean plus a random disturbance, u_t, that is normally and independently distributed with mean zero and variance σ^2. Show that the least squares estimator of this model is BLUE.

PROBLEM 21

Based on annual time series for the period 1950-1979 on P= profits and S = sales, the following regression equation was estimated:

$$(1)\quad \hat{P} = \underset{(1.219)}{-2.736} + \underset{(0.002)}{0.054}S, \qquad R^2 = 0.9756 \qquad \hat{\sigma} = 3.6382$$

a . Test if there is a significant relation between profits and sales.
b . The estimated equation in (1) was used to predict the values of P for 1980, 1981 and 1982. The resultant prediction errors and their corresponding standard errors are as follows:

	1980	1981	1982
$\hat{e}$	-8.78	-12.71	-37.38
$\hat{\sigma}_{\hat{e}}$	4.25	4.44	4.35

Asses the predictive performance of (1) .
c . Based on your answers to (a) and (b), what would you conclude about the relationship between profits and sales?

PROBLEM 22

Two variables Y and X are believed to be related by the following stochastic equation:

$$Y = \alpha + \beta X + u$$

where u is the usual random disturbance term with zero mean and constant variance σ^2 . To check this relationship one researcher takes a sample of size 8 and estimates β with OLS. A second researcher takes a different sample size of 8 and also estimates with OLS. The data they used and the results they obtained are as follows:

Researcher 1

Y	**X**
4.0	3
4.5	3
4.5	3
3.5	3
4.5	4
4.5	4
5.5	4
5.0	4

$$Y = \underset{(1.20)}{1.875} + \underset{(0.339)}{0.750}X$$

$R^2 = 0.45$
$\hat{\sigma} = 0.48$

Researcher 2

Y	**X**
2.0	1
2.5	1
2.5	1
1.5	1
11.5	10
10.5	10
10.5	10
11.0	10

$$Y = \underset{(0.27)}{1.5} + \underset{(0.038)}{0.970}X$$

$R^2 = 0.99$
$\hat{\sigma} = 0.48$

Can you explain why the standard error for $\hat{\beta}$ for the first researcher is larger than the standard error of $\hat{\beta}$ for the second researcher?

PROBLEM 23

Consider the following regression equations

$Y_t = \beta_0 + \beta_1 X_t + u_t$ (1)
$Y_t = \alpha_0 + \alpha_1 X_t + v_t$ (2)

where $y_t = Y_t - \overline{Y}$

a. Does multiplying each X_t value by a constant, say 5, change the residuals and fitted values in (1) ?
b. Will OLS estimators of β_0 and α_0 be identical?
c. Will OLS estimators of β_1 and α_1 be identical?

PROBLEM 24

Consider the following regression equations

$Y_t = \gamma_0 + \gamma_1 X_t + u_t$ (1)
$X_t = \delta_0 + \delta_1 Y_t + v_t$ (2)

a. Find the relation between $\hat{\gamma}_1$ and $\hat{\delta}_1$
b. Under what conditions will $\hat{\gamma}_1 = \hat{\delta}_1$
c. Show that $R_1^2 = R_2^2 = \hat{\gamma}_1 \hat{\delta}_1$

PROBLEM 25

Create two variables:

$Y_i{*} = Y_i - \overline{Y}$ and $X_i{*} = X_i - \overline{X}$

a. Consider the OLS regression of Y^* on X^*. Show that the intercept term $b_1^* = 0$

b. Show that the slope coefficient b_2^* is exactly equal to the slope coefficient of a regression of Y on X.

PROBLEM 26

Suppose you have the following population regression function, prf:

$Y_i = B_1 + B_2X_{2i} + u_i$

where $\text{var}(u_i) = \sigma^2$

Assume that $b_2 \sim N(B_2, \text{var}(b_2))$ where $\text{var}(b_2)$ is given by the standard formula.

1. Create a new variable $Y_i^* = 2Y_i$. That is, take every Y_i in your sample and double it. Then regress Y_i^* on X_i

$$Y_i^* = b_1^* + b_2^* X_{2i} + e_i$$

a) How do b_1^* and b_2^* compare to b_1 and b_2? (That is, derive the formula for b_1^* and b_2^* in terms of b_1 and b_2)
b) What is the variance of b_2^*? Compare the t-statistics for b_2^* and b_2. Interpret your findings.

2. Now create a new variable $X_i^* = 2X_{2i}$. That is, take every X_{2i} in your sample and double it. Then regress Y_i on X_{2i}^*

$$Y_i = b_1^* + b_2^* X_{2i}^* + e_i$$

a) How do b_1^* and b_2^* compare to b_1 and b_2? (That is, derive the formula for b_1^* and b_2^* in terms of b_1 and b_2)
b) What is the variance of b_2^*? Compare the t-statistics for b_2^* and b_2. Interpret your findings

3. Now create two new variables $Y_i^* = 2Y_i$ and $X_{2i}^* = 2X_{2i}$. Then regress Y_i^* on X_i^*

$$Y_i^* = b_1^* + b_2^* X_{2i} + e_i$$

a) How do b1* and b2* compare to b1 and b2? (That is, derive the formula for b1* and b2* in terms of b1 and b2)
b) What is the variance of b2*? Compare the t-statistics for b2* and b2. Interpret your findings.

PROBLEM 27

Suppose you run a regression of Y on X and also a regression of X on Y. Show that these regressions give you the same R^2 ?

PROBLEM 28

Consider a simple regression model;

a. Prove that the estimated slope of the regression of Y on X will equal the reciprocal of the estimated slope of the regression of X on Y only if $R^2 = 1$.
b. Prove that R^2 for the two-variable regression is unchanged if a linear transformation is made on both variables; that is, Y*=a1 +a2 Y, X*=b1 +b2 X.

PROBLEM 29

Suppose Y_i = log value of production per worker, X_i = log wage rate; and i refers to the i[th] firm. The least squares results for a sample of 52 observations are

$$\hat{Y}_i = -0.4 + 1.0 X_i + e_i$$
$$(0.1)$$

where 0.1 is the estimated standard error. Show that the R^2 for this equation must be 2/3.

PROBLEM 30

Two different researchers have access to the results of a single experiment and they independently estimate the following relationship between yield of a crop and the quantity of fertilizer.

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

How (if at all) would (a) estimated regression coefficient b, (b) the t statistic b/s_b and (c) the value of R^2 computed by the two researchers differ if

a) Researcher A measured both X and Y in kilograms whereas B measured both of these variables in metric tons (1 ton=1000 Kg)?
b) Researcher A measured both X and Y in kilograms whereas B measured X in kilograms and Y in metric tons?

PROBLEM 31

Suppose the sample covariance between u and x is defined as

$$\hat{\text{cov}}(u_t, x_t) = \frac{1}{T}\sum (u_t - \bar{u})(x_t - \bar{x})$$

where $\bar{u}$ and $\bar{x}$ are sample means. Show that

$$\hat{\text{cov}}(u_t, x_t) = \frac{1}{T}\sum u_t x_t - \bar{u}\bar{x})$$

PROBLEM 32

Suppose we have the Simple Linear Model

$$y_t = \beta_1 + \beta_2 x_t + e_t, \quad t=1,...,T$$

$E(\varepsilon_t) = 0$
$E(\varepsilon_t^2) = \sigma^2$
$E(\varepsilon_t \varepsilon_s) = 0$ for all $t \neq s$

a. Show that

$$\sum \hat{e}_t = 0.$$

b. Show that

$$\text{cov}(x_{t,} \hat{e}_t) = 0.$$

c. Show that

$R^2 = corr(y_{t,} \hat{y}_t)^2$ and interpret.

where $\hat{y}_t = b_1 + b_2 x_t$

Interpret.

d. Show that: $R^2 = c\hat{o}rr(y_{t,} x_t)^2$ and interpret.

PROBLEM 33

You are given the model:

$$Y_t = \beta X_t + e_t$$

Derive the variance of $\hat{\beta}$.

PROBLEM 34

Suppose we have the Simple Linear Model

$y_t = \beta_1 + \beta_2 x_t + \varepsilon_t$, $t=1,...,T$

$E(\varepsilon_t) = 0$
$E(\varepsilon_t^2) = \sigma^2$
$E(\varepsilon_t \varepsilon_s) = 0$ for all $t \neq s$

a. Derive the LS estimators of. β_1 and β_2.
b. Derive the variance of the LS estimator of β_2.
c. Show that the LS residuals average to zero. Interpret.

PROBLEM 35

True or False? Explain.

a. Variance is a measure of dispersion, while mean is a measure of central tendancy.
b. Covariance cannot be negative.
c. The correlation coefficient between two variables is not changed if either variable is multiplied by a constant.
d. A bigger *t*-statistic (in absolute value, say bigger than 2.0) leads to rejection of the null hypothesis. Other things being equal, using a smaller sample size results in a bigger *t*-statistic. .
e. A bigger *t*-statistic (bigger than 2.0) is a sign of the existence of a relationship between the dependent variable and the independent variables. To have a bigger *t*-statistic, the sum of squared residuals (SSR) should be smaller.
f. Other things being equal, the confidence interval becomes wider as the sample size increases.
g. A simple regression line passes through the origin, if it does not include a constant term, while it passes through the point of means of the dependent variable and the independent variable if it includes a constant term.
h. A simple regression line passes through the origin, if it does not include a constant term, while it passes through the point of means of the dependent variable and the independent variable if it includes a constant term.
i. The signs of the regression coefficient in a simple regression, corresponding t-statistic, and the correlation coefficient between the dependent variable and the independent variable are identical.
j. An R^2 from the regression model without a constant term is bigger than the R^2 from the same regression model including a constant term. Note that a constant term is also considered as one of independent variables.
k. A bigger R^2 is always good.
l. Marginal effect is measured as the coefficient of the independent variable in a regression.
m. The coefficient of the dummy independent variable is interpreted as the difference of two groups that the dummy variable implies.

n. R^2 will always increase as we add *MORE OBSERVATIONS*.
o. SSR (sum of squared residuals or called RSS) will always decrease by adding more variables.

PROBLEM 36

For observations on investment y and profits x of each of 100 firms, it is known that Y=a+bX+e and it is proposed to estimate a and b by OLS. Suppose every firm in the sample had the same profits. What, if any, problem would this create?

PROBLEM 37

Suppose the classical linear regression model applies to Y=bX+e. The slope coefficient in the regression of X on Y is just the inverse of the slope from the regression of Y on X. True, false, or uncertain? Explain.

PROBLEM 38

Suppose your data produce the regression result Y=5+2X. Consider scaling the data to express them in a different base year dollar, by multiplying observations by 0.8.

(a) If both X and Y are scaled, what regression results would you obtain?
(b) If Y is scaled, but X is not, what regression results would you obtain?
(c) If X is scaled, but Y is not, what regression results would you obtain?

PROBLEM 39

Prove that the estimated regression line passes through the means of X and Y.
Hint: Show that the means satisfy the equation Y=a+bX, where a and b are the OLS estimators.

PROBLEM 40

The most common measure of fit in a regression model is given by

a. the sum of squared residuals.
b. The explained sum of squares over the residual sum of squares.
c. The explained sum of squares over the total sum of squares.
d. The total sum of squares.

PROBLEM 41

OLS is the Best Linear Unbiased Estimator because

a. the residuals always sum to zero.
b. It gives the minimum variance of all linear unbiased estimators.
c. It produces an estimator with zero variance.
d. It minimizes the sum of squared residuals.

PROBLEM 42

When using OLS to estimate a linear regression model, a sample of 100 data points will

a. always produces statistically significant coefficient estimates.

b. always produce a better goodness of fit than an alternative sample of 75 data points.
c. always give an R squared higher than 0.5.
d. produce an average of residual equal to zero.

PROBLEM 43

The regression model includes a random error or disturbance term for a variety of reasons. Which of the following is NOT one of them?

a. measurement errors in the observed variables
b. omitted influences on Y (other than X)
c. linear functional form is only an approximation
d. the observable variables do not exactly correspond with their theoretical counterparts
e. there may be approximation errors in the calculation of the least squares estimates

PROBLEM 44

Which of the following assumptions about the error term is not part of the so called "classical assumptions"?

a. it has a mean of zero
b. it has a constant variance
c. its value for any observation is independent of its value for any other observation
d. it is independent of the value of X
e. it has a normal distribution

PROBLEM 45

Which of the following is NOT true?

a. the point Xbar, Ybar always lies on the regression line
b. the sum of the residuals is always zero
c. the mean of the fitted values of Y is the same as the observed values of Y
d. there are always as many points above the fitted line as there are below it
e. the regression line minimizes the sum of the squared residuals

PROBLEM 46

In a simple linear regression model the slope coefficient measures

a. the elasticity of Y with respect to X
b. the change in Y which the model predicts for a unit change in X
c. the change in X which the model predicts for a unit change in Y
d. the ratio Y/X
e. the value of Y for any given value of X

PROBLEM 47

Changing the units of measurement of the Y variable will affect all but which one of the following?

a. the estimated intercept parameter
b. the estimated slope parameter
c. the Total Sum of Squares for the regression
d. R squared for the regression

e. the estimated standard errors

PROBLEM 48

A fitted regression equation is given by Yhat = 20 + 0.75X. What is the value of the residual at the point X=100, Y=90?

a. 5
b. -5
c. 0
d. 15
e. -5

PROBLEM 49

a. Under the Gauss-Markov conditions, OLS can be shown to be BLUE. The phrase "linear" in this acronym refers to the fact that we are estimating a linear model. true or false?
b. In order to apply a t-test, the Gauss-Markov conditions are strictly required. true or false?
c. A regression of the OLS residual upon the regressors included in the model by definition yields an R-squared of zero. true or false?
d. The hypothesis that the OLS estimator is equal to zero can be tested by means of a t-test. true or false?
e. From asymptotic theory we learn that - under appropriate conditions - the error terms in a regression model will be approximately normally distributed if the sample size is sufficiently large. true or false?
f. If the absolute t-value of a coefficient is smaller than 1.96, we can accept the null hypothesis that the coefficient is zero, with 95% confidence. true or false?
g. Because OLS provides the BEST linear approximation of a variable y from a set of regressors, OLS also gives BEST linear unbiased estimators for the coefficients of these regressors. true or false?
h. If a variable in a model is significant at the 10% level, it is also significant at the 5% level. true or false?
i. Consider two alternative models that are nested. If we compare the models on the basis of the adjusted R squared and an F-test, the F-test will prefer the extended model more often than the adjusted R-squared. true or false?
j. By including an irrelevant variable in a regression model, you can reduce the coefficient of determination (R squared). True or false?
k. If the intercept is omitted from the regression equation, the estimated regression line continues to pass through the means of X and Y.