ECON2P91: Business Econometrics with Applications
Mid-term 2

NAME:__________SOLUTIONS__

STUDENT NO: ______________________

LECTURE SECTION: Section 1:
Section 2:

LAB SECTION: Please indicate the lab section that you attend as the midterm will be returned to students in the labs.

Lab1: Wed 08:00-10:00 MCJ 202
Lab2: Tue 12:00-14:00 MCJ 202
Lab3: Tue 14:00-16:00 MCJ 202
Lab4: Mon 14:00-16:00 MCJ 202
Lab5: Mon 08:00-10:00 MCJ 202
Lab6: Fri 10:00-12:00 MCJ 202
Lab7: Wed 14:00-16:00 MCJ 202
Lab8: Fri 12:00-14:00 MCJ 202
Lab9: Fri 08:00-10:00 MCJ 202
Lab10: Wed 19:00-21:00 MCJ 202
Lab11: Tue 19:00-21:00 MCJ 202
Lab12: Tue 19:00-21:00 MCJ 201
Lab13: Thu 19:00-21:00 MCJ 202

Instructions: Answer all questions from Sections A and B
The standard normal distribution table is provided on page 12
The F distribution table is provided on page 13

THIS BOX IS FOR MARKERS ONLY:

Section A _____/60

Section B _____/40

Total _____/100

Section A: Multiple Choice **[60 points; 2 points each]**

1. Which of the following statements is correct?

(Note that: ESS=Explained Sum of Squares, SSR=Sum of Squared Residuals; TSS=Total Sum of Squares)

a. ESS = SST + TSS
b. TSS = ESS + SSR
c. ESS > TSS
d. $R^2 = 1 - (ESS/TSS)$

Answer: ___**B**___

2. Consider the regression model $Y_i = \beta_0 + \beta_1 X_i + u_i$ where u_i is the error term. The OLS estimators for β_0 and β_1 are derived by:

a. connecting the Y_i corresponding to the lowest X_i observation with the Y_i corresponding to the highest X_i observation.
b. making sure that the standard error of the regression equals the standard error of the slope estimator.
c. minimizing the sum of absolute residuals.
d. minimizing the sum of squared residuals.

Answer: __**D**____

3. Consider the regression model $Y_i = \beta_0 + \beta_1 X_i + u_i$ where u_i is the error term. The OLS estimator for β_1 is obtained by:

a. dividing the Covariance between X and Y by the variance of Y.
b. dividing the covariance between X and Y by the product of the standard deviation of X and the standard deviation of Y.
c. dividing the covariance between X and Y by the standard deviation of X.
d. dividing the covariance between X and Y by the variance of X.

Answer: ___**D**___

4. The regression R^2 is a measure of:

a. whether or not X causes Y.
b. the goodness of fit of your regression line.
c. whether or not ESS > TSS.
d. the square of the determinant of R.

Answer: __**B**____

5. The OLS residuals:

a. can be calculated using the errors from the regression function.
b. can be calculated by subtracting the fitted values of the dependent variable from the actual values of the dependent variable.
c. are unknown since we do not know the population regression function.
d. should not be used in practice since they indicate that your regression does not run through all your observations.

Answer: __**B**____

6. Which of the following assumptions definitely does not apply to a linear regression with one regressor?

   a. homoskedasticity
   b. zero mean
   c. independence
   d. multicollinearity

Answer: ___**D**___

7. Heteroskedasticity means that:

   a. homogeneity cannot be assumed automatically for the model.
   b. the variance of the error term is not constant.
   c. the observed units have different preferences.
   d. agents are not all rational.

Answer: ___**B**___

8. The confidence interval for the sample regression function slope:

   a. can be used to conduct a test about a hypothesized population regression function slope.
   b. can be used to compare the value of the slope relative to that of the intercept.
   c. adds and subtracts 1.96 from the slope.
   d. allows you to make statements about the economic importance of your estimate.

Answer: ___**A**___

9. Consider the estimated regression line $\hat{y}_i = 698.9 - 2.28X_i$. You are told that the t-statistic on the slope coefficient is 4.38. What is the standard error of the slope coefficient?

   a. 0.52
   b. 1.96
   c. -1.96
   d. 4.38

Answer: ___**A**___

10. A large p-value relative to the level of significance:

   a. indicates evidence in favour of the null hypothesis.
   b. implies that the t-statistic is less than 1.96.
   c. indicates evidence in against the null hypothesis.
   d. will happen roughly one in twenty samples.

Answer: __**C**____

11. The Gauss-Markov theorem states the OLS estimator is BLUE provided that all the standard assumptions are met. The U in BLUE means that:

   a. OLS estimator has the smallest variance
   b. the sum of OLS residuals is zero
   c. OLS estimator yields a U-shaped relationship
   d. the mean of the sampling distribution of the OLS estimator of the regression slope is equal to the value of the corresponding population parameter

Answer: __**D**____

12. The standard error of a regression (SER):

a. is unit free.
b. measures the fraction of the variation of the dependent variable explained by the independent variable.
c. should be low as an indication of good fit.
d. is always less than 1.

Answer: __**C**____

13. The value of R-squared for a regression of earnings on experience is reported as 0.8. It can therefore be deduced that

a. 80 percent of the variance in earnings is explained by variation in experience
b. 80 percent of the variance in experience is explained by variation in earnings
c. the standard error of the regression is 0.8
d. the correlation coefficient between earnings and experience is 0.8

Answer: ___**A**___

14. The sum of all OLS residuals is equal to

a. the mean of the dependent variable
b. zero
c. the Total Sum of Squares (TSS)
d. none of the above

Answer: ___**B**___

15. The statistical software package has included the following quantities in its output for the regression equation $y_i = \beta_0 + \beta_1 x_i + u_i$, where y and x denote sales-revenue and advertisement-expenditure, respectively. Total Sum of Squares (TSS)=50; Explained Sum of Squares (ESS)=15; and Sum of Squared Residuals (SSR))=35. How much of the variation in sales-revenue is explained by variation in advertisement-expenditure?

a. 30%
b. 70%
c. 35%
d. 15%

Answer: ___**A**___

16. All the following statements refer to the Gauss-Markov theorem except:

a. OLS estimator has the smallest variance
b. the sum of OLS residuals is zero
c. OLS estimator is linear in the dependent variable
d. the mean of the sampling distribution of the OLS estimator of the regression slope is equal to the value of the corresponding population parameter

Answer: ___**A**___

17. Suppose that a researcher, using data on advertisement expenditure (x) and sales revenue (y), estimates the OLS regression $\hat{y}_i = 520.4 - 5.82 x_i$. The standard error of the intercept coefficient is 100 and that of the slope coefficient is 10. The 95% confidence interval for the regression intercept coefficient is

a. $-5.82 \pm (1.96 x 10)$
b. $-5.82 \pm (1.96 x 100)$
c. $520.4 \pm (1.96 x 100)$
d. $520.4 \pm (1.96 x 10)$

Answer: ___C___

18. The confidence interval for a single coefficient in a multiple regression

a. makes little sense because the population parameter is unknown.
b. should not be computed because there are other coefficients present in the model.
c. contains information from a large number of hypothesis tests.
d. should only be calculated if the regression R^2 is identical to the adjusted R^2.

Answer: ___C___

19. Binary variables

a. are generally used to control for outliers in your sample.
b. can take on more than two values.
c. exclude certain individuals from your sample.
d. can take on only two values.

.

Answer: ___D___

20. You are asked to study the relationship between the number of deaths due to vehicle accidents (y) and the number of registered vehicles (x). For this purpose, you specified the model $y_i = \beta_0 + \beta_1 x_i + u_i$, where u is the error term. This regression presupposes that:

a. the direction of causation is from y to x
b. the direction of causation is from y to u
c. the direction of causation is from x to u
d. the direction of causation is from x to y

Answer: ___D___

21. For the regression equation $y_i = \beta_0 + \beta_1 x_i + u_i$, where u is the error term. The parameter β_1 is interpreted as

a. the effect of a 1 unit change in y on x
b. the effect of a 1 unit change in x on y
c. the value of y when x is 0
d. the value of x when y is 0

Answer: ___B___

22. For the regression (visits)=b0+b1(health concerns)+u, a researcher obtained b1 = -0.153. The 95 percent confidence interval for b1 is -0.265 to -0.041. Using this confidence interval the null hypothesis that b1=0 is

a. rejected
b. not rejected
c. there is not enough information to reject or not reject the null hypothesis
d. none of the above

Answer: ___**A**___

23. For the regression (visits)=b0+b1(health concerns)+u, a researcher obtained b1=-0.153. The 95 percent confidence interval for b1 is -0.265 to -0.041. Using this confidence interval the null hypothesis that b1= -0.1 is

a. rejected
b. accepted
c. neither accepted nor rejected
d. there is not enough information to accept or reject the null hypothesis

Answer: ___**B**___

24. In the multiple regression model, the adjusted R^2, $\bar{R}^2$

a. cannot be negative.
b. will never be greater than the regression R^2.
c. equals the square of the correlation coefficient *r*.
d. cannot decrease when an additional explanatory variable is added.

Answer: __**B**____

25. Under perfect multicollinearity

a. the OLS estimator cannot be computed.
b. two or more of the regressors are highly correlated.
c. the OLS estimator is biased even in samples of $n > 100$.
d. the error terms are highly, but not perfectly, correlated.

Answer: __**A**____

26. When there are omitted variables in the regression, which are determinants of the dependent variable, then

a. you cannot measure the effect of the omitted variable, but the estimator of your included variable(s) is (are) unaffected.
b. this has no effect on the estimator of your included variable because the other variable is not included.
c. this will always bias the OLS estimator of the included variable.
d. the OLS estimator is biased if the omitted variable is correlated with the included variable.

Answer: __**D**____

27. Consider the model $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_k X_{ki} + u_i, i = 1,..., n$. For this model, the adjusted R^2, or $\overline{R}^2$, is given by

a. $1 - \frac{n-2}{n-k-1}\frac{SSR}{TSS}$.
b. $1 - \frac{n-1}{n-k-1}\frac{ESS}{TSS}$.
c. $1 - \frac{n-1}{n-k-1}\frac{SSR}{TSS}$.
d. $\frac{ESS}{TSS}$.

Answer: __C___

28. When testing joint hypothesis, you should

a. use *t*-statistics for each hypothesis and reject the null hypothesis is all of the restrictions fail.
b. use the *F*-statistic and reject all the hypothesis if the statistic exceeds the critical value.
c. use *t*-statistics for each hypothesis and reject the null hypothesis once the statistic exceeds the critical value for a single hypothesis.
d. use the *F*-statistics and reject at least one of the hypothesis if the statistic exceeds the critical value.

Answer: __D____

29. Consider the regression model $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + u_i$, where u_i denotes the usual error term. To test the hypothesis that β_1 and β_2 are jointly significant, the appropriate restricted regression is:

a. $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + u_i$
b. $y_i = \beta_0 + u_i$
c. $y_i = \beta_0 + \beta_3 x_{3i} + \beta_4 x_{4i} + u_i$
d. $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$

Answer: ___C___

30. The overall regression *F*-statistic tests the null hypothesis that

a. all slope coefficients are zero.
a. all slope coefficients and the intercept are zero.
b. the intercept in the regression and at least one, but not all, of the slope coefficients is zero.
c. the slope coefficient of the variable of interest is zero, but that the other slope coefficients are not.

Answer: ___A___

Section B: Longer Questions and Answers [40 points]

Question 1

What is the relationship between the average hourly earnings of full-time full-year workers (AHE) and age (AGE)? Are there gender differences in average hourly earnings?

To answer this question, a researcher specified the model

$AHE_i = \beta_0 + \beta_1 AGE_i + \beta_2 FEMALE_i + u_i$ *i=1,2,...,7986.*

The units of measurement of the variables are as follows: AHE is in dollars, AGE is in years; and FEMALE is a binary variable for gender (i.e., FEMALE=1 if the person is female and FEMALE=0 is the person is male).

The GRETL output presented on page 11 summarizes the regression results. Use these reported results to answer the following questions:

(a) Report the regression results using the same format as on the assignments (i.e., including R^2 and SER). **[3 points]**

$\hat{AHE}_i = 4.60686 + 0.441542 AGE_i - 2.34676 FEMALE_i$ *R2=0.0396 SER=8.5843*

SE (0.9990293) (0.0332389) (0.195032)

(b) Interpret the reported value of the coefficient of determination. Be specific in your interpretation. **[3 points]**

Coefficient of determination (R-squared) = 0.039671
It means that only 3.96% of the variation in dependent variable (AHE) can be explained by the variation in all independent variables (AGE & FEMALE)

(c) Interpret the estimated coefficient of AGE. **[3 points]**

Coefficient of AGE=0.441542
It means that one unit change in AGE (years), would lead to 0.441542 change in AHE (by units of AHE), HOLDING FEMALE CONSTANT.

(d) Interpret the estimated coefficient of FEMALE. **[3 points]**

Coefficient of FEMALE=-2.34676
It means that changing FEMALE to MALE (binary variable, changes between 0 and 1 only), would lead to lower AHE by 2.34676 units of AHE (dollars in this case), HOLDING AGE CONSTANT.

Or:

Females in this study receive an average of 2.34676 dollars less salary than males, HOLDING AGE CONSTANT.

.

(e) Test the hypothesis that coefficient of FEMALE is 0 against the one-sided alternative that it is less than 0. Use the p-value approach and 5% level of significance and interpret the results of this hypothesis test.
[3 points]

H0: $\beta_2 = 0$
Ha: $\beta_2 < 0$ *(one-sided test)*

T-ratio = (-2.34676-0)/0.195032 = -12.0326

P-value approach;
p-value= $\Phi(-12.0326) = 0$
Since the p-value (0) is less than the level of significance (0.05), we reject the null hypothesis and conclude that the coefficient of FEMALE is significant, which is indicative of discrimination.

(f) Test for significance of the overall regression using 5% level of significance and interpret the results. **[4 points]**

H0: $\beta_1 = \beta_2 = 0$
Ha: at least one of β_1, β_2 *are not equl to zero*

F-ratio = F(2, 7983) = 12150.5 / 73.6899 = 164.887
F-critical value (2,∞)=3

Since F-ratio is bigger than F-critical, H0 is rejected, and we conclude that β_1 *or* β_2 *are significantly different from 0.*

(g) Use the estimated regression to predict the average hourly earnings for a 30-year old male. **[3 points]**

$\widehat{AHE} = 4.60686 + 0.441542\ (30) - 2.34676\ (0) = 17.85312$ *($)*

Average hourly earnings of a 30 years old male in this database is 17.85312 $

Question 2
Suppose that a researcher, using data on class size (CS) and average test scores from 100 third grade classes, estimates the OLS regression

$$\widehat{TestScore} = 520.4 - 5.82xCS, R^2 = 0.08, SER = 11.5$$
$$St.Error \quad (20.4) \quad (2.21)$$

(a) Construct a 95% confidence interval for β_1, the regression slope coefficient. **[3 points]**

95% confidence interval for $\beta_1 : \hat{\beta}_1 \pm 1.96\ SE\ (\hat{\beta}_1)$

= -5.82 $\pm 1.96 * (2.21)$ *= (-10.1516 , -1.4884)*

(b) Calculate the p-value for the two-sided test of the null hypothesis $H_0 : \beta_1 = 0$, where β_1 is the regression slope coefficient. Do you reject the null hypothesis at the 5% level? **[3 points]**

H0: $\beta_1 = 0$
Ha: $\beta_1 \neq 0$ *(2-sided test)*

t-ratio = -5.82/2.21 = -2.633

P-value approach;
p-value=$2\phi\ (-2.633) = 0.0043$
Since the p-value (0.0043) is less than the level of significance (0.05), we reject the null hypothesis and conclude that the slope coefficient is significantly different from zero.

(c) Explain the reason for reporting both R^2 and SER. **[3 points]**

$R^2 = 0.08$, *which indicates 8.00% of variation in test score can be explained by variation in CS.* R^2 *shows the goodness of fit of the regression line.* $R^2 = 0.08$ *is a poor fit.*
$SER = 11.5$ *It means that the model has a high value of error. SER also indicates the level of goodness of fit. In this example, SER of 11.5 is a large number, indicating a poor fit also.*

(d) What are the limitations of R^2? **[3 points]**
R^2 *indicates the percentage of variation in dependent variable that can be explained by the variation in independent variable.*

1-It only lies between 0 and 1. Higher values indicate better fit.
2- $R^2 = 0$ *indicates horizontal regression line (the worst fit). (ESS = 0)*
3- $R^2 = 1$ *indicates all points on a scatter line are on a straight line, the regression line (the best fit). (ESS = TSS)*
4- R^2 *is the square of the correlation coefficient between the dependent variable (Y) and independent variable (X).*
5- $R^2 = 0.08$ *indicates that 8.00% of the variation in the dependent variable is explained by variation in the regressor (independent variable) included in the regression. The other 92.0% is attributed to other factors or random error.*

(e) What are the units of SER? **[3 points]**

SER has the same unit of dependent variable (in this example, Test Score).

(f) Should multicollinearity be a concern in this model? Explain. **[3 points]**

Multicollinearity is not a concern in this model, as we are dealing with only one regressor.

GRETL OUTPUT
++

(a) OLS estimates

Model 1: OLS, using observations 1-7986
Dependent variable: AHE

	coefficient	std. error	t-ratio	p-value	
Const	4.60686	0.999029	4.611	4.06e-06	***
AGE	0.441542	0.0332389	13.28	7.59e-040	***
FEMALE	-2.34676	0.195032	-12.03	4.62e-033	***

Mean dependent var	16.77115	S.D. dependent var	8.758696
Sum squared resid	588266.3	S.E. of regression	8.584281
R-squared	0.039671	Adjusted R-squared	0.039430
F(2, 7983)	164.8868	P-value(F)	6.76e-71
Log-likelihood	-28499.51	Akaike criterion	57005.01
Schwarz criterion	57025.97	Hannan-Quinn	57012.18

(b) ANOVA table

Analysis of Variance:

	Sum of squares	df	Mean square
Regression	24301	2	12150.5
Residual	588266	7983	73.6899
Total	612567	7985	76.7147

R^2 = 24301 / 612567 = 0.039671
F(2, 7983) = 12150.5 / 73.6899 = 164.887 [p-value 6.76e-071]