Chapter 1-6 Quiz

Instructions:

1. **Please answer all questions in English.**
2. **Write your answers on the answer sheet, clearly indicating the question number before each answer.**

I. 20 points

State whether the following statements are true or false. On the answer sheet, mark √ if the statement is TRUE and × if the statement is FALSE.

1. The population regression function (PRF) represents the expected value of the dependent variable for a given value of the explanatory variable.
2. The sum of the deviation of a random variable from its mean value is always equal to zero.
3. The Gauss-Markov theorem establishes that ordinary least squares (OLS) estimators are the best linear unbiased estimators under the classical linear regression assumptions.
4. The OLS estimators b1 and b2 each follow the normal distribution only if ui follows the normal distribution.

5. In a multiple regression model, if the coefficient of determination R^2 equals 0, the F-statistic for overall significance will approach infinity; conversely, if $R^2 = 1$, the F-statistic equals 0.
6. When we say that an estimated regression coefficient is statistically significant, we mean that it is statistically different from 1.
7. The degrees of freedom associated with the total sum of squares (TSS) in a regression model is always n-1, where n is the sample size, irrespective of the number of independent variables included.
8. In a linear-in-variables (LIV) model, the marginal effect of the independent variable on the dependent variable is constant across all levels, while the elasticity varies; in contrast, a log-log model features constant elasticity but varying marginal effects.
9. In the model $Y_i = B_1 + B_2 D_i + u_i$, letting D_i take the values of (0, 2) instead of (0, 1) will *halve* the value of B_2 and will also *halve* the *p* value.
10. The primary objective of the ordinary least squares (OLS) method in linear regression is to minimize the sum of the residuals.

II **34 points**

Based on the data of 64 countries, estimate the following regression model by OLS and obtain the following estimation results:

$$CM_i = B_0 + B_1 FLR_i + B_2 PGNP_i + B_3 TFR_i + u_i \quad \textbf{Model 1}$$

CM: child mortality
FLR: female literacy rate
PGNP: per capita GNP
TFR: total fertility rate

Dependent Variable: CM
Method: Least Squares
Sample: 1 64
Included observations: 64

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	168.307	32.892	5.117	0.000
FLR	-1.768	①	-7.129	0.000
PGNP	-0.006	0.002	②	0.005
TFR	12.869	4.191	3.071	0.003
R-squared	0.747	Mean dependent var		141.5
Adjusted R-squared	③	S.D. dependent var		75.978
S.E. of regression	39.131	Akaike info criterion		10.232
Sum squared resid	91875.383	Schwarz criterion		10.367
Log likelihood	-323.43	Hannan-Quinn criter.		10.285
F-statistic	④	Durbin-Watson stat		2.17
Prob(F-statistic)	0.000			

11. Compute the value of ①, ②, ③ and ④ in the above Eviews output. (4 points)
12. What is the expected relationship between CM and each of the other variables? Are the estimation results consistent with your expectance? (5 points)
13. Interpret the estimated coefficient of PGNP. (5 points)
14. How would you interpret the coefficient of determination (*R-square)* ? (5 points)
15. Test the null hypothesis that $B_1 = 0$ at a 1% significance level. (5 points)
16. Test at 5% significance level, that all partial slope coefficients are equal to zero. (5 points)

17. Regress CM on FLR, and obtain the following results:

Dependent Variable: CM
Method: Least Squares
Sample: 1 64
Included observations: 64

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	263.864	12.225	21.584	0.000
FLR	-2.390	0.213	-11.209	0.000

R-squared	0.670	Mean dependent var	141.500
Adjusted R-squared	0.664	S.D. dependent var	75.978
S.E. of regression	44.024	Akaike info criterion	10.438
Sum squared resid	120163.000	Schwarz criterion	10.506
Log likelihood	-332.019	Hannan-Quinn criter.	10.465
F-statistic	125.646	Durbin-Watson stat	2.315
Prob(F-statistic)	0.000		

How would you decide if it is worth adding the variables PGNP and TFR to the model? Which test would you use? Show the necessary calculations. (5 points)

III 24 points

John estimated Model 1 and Model 2 with Chinese Population Data between 1990-2020.

$Y_t = B_0 + B_1 T + u_t$ **Model 1**

$lnY_t = B_0 + B_1 T + u_t$ **Model 2**

Y_t: Total Population (10000 persons)

T: 1990, T=1, 1991,T=2,… 2020, T=31

The following tables display the estimation output of Model 1 and Model 2.

Model 1

Dependent Variable: Y

Method: Least Squares

Sample: 1990 2020

Included observations: 31

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	115937.900	365.550	317.161	0.000
T	870.007	19.942	43.626	0.000

R-squared	0.985	Mean dependent var	129858.000
Adjusted R-squared	0.984	S.D. dependent var	7970.248
S.E. of regression	993.122	Akaike info criterion	16.702
Sum squared resid	28602423.000	Schwarz criterion	16.794
Log likelihood	-256.880	Hannan-Quinn criterion.	16.732
F-statistic	1903.236	Durbin-Watson stat	0.094
Prob(F-statistic)	0.000		

Model 2

Dependent Variable:LOG(Y)

Method: Least Squares

Sample: 1990 2020

Included observations: 31

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	11.664	0.004	3329.445	0.000
T	0.007	0.000	35.356	0.000

R-squared	0.977	Mean dependent var	11.772
Adjusted R-squared	0.977	S.D. dependent var	0.062
S.E. of regression	0.010	Akaike info criterion	-6.409
Sum squared resid	0.003	Schwarz criterion	-6.316
Log likelihood	101.339	Hannan-Quinn criterion.	-6.379
F-statistic	1250.078	Durbin-Watson stat	0.086
Prob(F-statistic)	0.000		

18. Interpret the estimated coefficient of T (870.007) in **Model 1**. (6 points)

19. Interpret the estimated coefficient of T (0.007) in **Model 2**. (6 points)

20. Can you compare the R squares of the two models? Why or why not? (6 points)

21. How do you decide which is a better model? (6 points)

IV 22 points

Consider the following **Model 1**:

$$\text{Wage}_i = B_0 + B_1\ \text{Experience}_i + B_2\text{College}_i + B_3\text{Master}_i + u_i \quad \text{Model 1}$$

Where Wage= Monthly Salary

Experience = Working experience in years.

College = 1 for those whose highest education level is college，0 if otherwise.

Master = 1 for those whose highest education level is graduate school，0 if otherwise.

The base group for this study are those whose highest education level is high school.

22. What are the expected signs of the various coefficients in **Model1**? (4 points)

23. How would you interpret B_2 and B_3 in **Model1?** (6 points)

Continue with the above variables, but now consider the following **Model 2**:

$$\begin{aligned}\text{Wage}_i = & B_0 + B_1\ \text{Experience}_i + B_2\text{College}_i + B_3\text{Master}_i \\ & + B_4\ \text{Experience}_i * \text{College}_i + B_5\ \text{Experience}_i * \text{Master}_i + u_i \quad \text{Model 2}\end{aligned}$$

24. What is the interpretation of B_4 and B_5? (6 points)

25. If B_4 and B_5 are individually statistically significant, would you choose **Model 2** over **Model 1**? (6 points)