

注意：本試卷共有二頁。請考生答題前，務必閱讀每大題的注意事項說明！

第一大題：計算問答說明題（50分）

注意：(1) 第一大題有三題計算問答說明題。

(2) 請標示清楚，並將所有過程、步驟交代清楚；沒有說明過程者，甚者只給簡單回答如 Yes、No 等，不給分。每大題之下的小題分數，如括號內所示。

Note: You should carefully state the reasons or calculations in the following questions in order to get the points. A short answer, such as "Yes" or "No" will NOT receive any point.

1. (26 points) A researcher conducted a study to examine the effect of college education on earnings in Taiwan. Using data from a large survey, he estimated several regression models. Definitions of the selected variables are provided in Table 1, and the results of the regression analyses are presented in Table 2. Utilize this information to answer to the following questions. Calculate to the third decimal place.

Table 1. Definitions of selected variables.

Variable	Definition	Mean	S. D.
College	A binary variable for whether the respondent holds a college degree.	0.554	0.497
Earnings	Monthly earnings.	-	-
East	A binary variable for whether the respondent resides in eastern Taiwan.	0.033	0.179
North	A binary variable for whether the respondent resides in northern Taiwan.	0.307	0.461
Central	A binary variable for whether the respondent resides in central Taiwan.	0.281	0.450
South	A binary variable for whether the respondent resides in southern Taiwan.	0.379	0.485

Table 2. Regression results

Variable	Model 1		Model 2		Model 3	
	Earnings		Earnings		Log(Earnings)	
	Coef.	S. E.	Coef.	S. E.	Coef.	S. E.
College	5.332 ***	0.097	5.062 ***	0.097	0.322 ***	0.004
East			-3.215 ***	0.117	-0.144 ***	0.005
North			4.029 ***	0.275	0.100 ***	0.012
Central			-0.131	0.120	-0.017 ***	0.005
Constant	30.472 ***	0.072	31.217 ***	0.093	3.113 ***	0.004
$R^2$	0.0121		0.0156		0.0251	
Observations	272,357		272,357		272,357	

Note: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

- What is the estimated correlation coefficient between *College* and *Earnings*? (3 points)
- What is the average earnings among respondents without a college degree? (3 points)
- What are the (i) average and (ii) the standard deviation of earnings for the full sample? (6 points)
- Based on the results in Model 2, what is the expected difference in the average earnings between respondents in Northern Taiwan and respondents in Eastern Taiwan? (3 points)
- Interpret the estimated coefficient of *College* in Model 2 and Model 3, respectively. Be specific. (4 points)
- Can we conclude that Model 3 is superior to Model 2 on the ground of higher  $R^2$ ? Explain. (3 points)
- Based on the regression results in Table 2, can we conclude that college education leads to higher earnings? Explain. (4 points)

見背面

2. (20 points) Answer the following true or false questions with a brief explanation.
- A. A researcher is interested in estimating the mean income of the population. He collected a random sample and found a 98% confidence interval of \$35,000 to \$40,000. This implies that this interval has a 98% chance of containing the mean income in the population. (4 point)
  - B. In the analysis of variance (ANOVA), the larger the variability between group means relative to the within-group variability, the stronger the evidence against the null hypothesis. (4 points)
  - C. It is generally advisable to include a higher number of variables in a regression model for better model performance and accuracy. (4 points)
  - D. In a simple regression model with only one explanatory variable, it is possible that the estimated coefficient is insignificant, even though the global F-test is significant. (4 point)
  - E. In a multiple regression model with more than one explanatory variables, it is possible that all estimated coefficients are insignificant, even though the global F-test is significant. (4 points)
3. (4 points) Given the probability density function of an exponential probability distribution provided below, derive its cumulative density function.

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

第二大題：填充題（每格 5 分，共 50 分）

- (1) 第二大題為填充題，請依空格號 (1)~(10) 之順序作答，每一空格 5 分；此部分 不須 計算過程。
- (2) 如果沒有特別指示，請將答案約分至「最簡分數」表示，否則不予計分。
- A. Let  $Y_1$  denotes the time (in minutes) until the first arrival at a checkout counter,  $Y_2$  denotes the time between the first and second arrival, ..., and  $Y_n$  denotes the time between the  $(n - 1)$ th and  $n$ -th arrival.  $Y_1, Y_2, \dots, Y_n$  are **independent** random variables from the exponential distribution with mean  $\theta$ . The variance of  $Y_1$  is (1) ; the moment generating function (MGF) of  $Y_n$  is (2) . Let  $U = Y_1 + Y_2 + \dots + Y_n$ . The meaning of  $U$  is (3) ; the MGF of  $U$ ,  $m_U(t) =$  (4) ; the variance of  $U$ ,  $\text{Var}(U) =$  (5) .
- B. Let  $Y_1, Y_2, \dots, Y_n$  denote a random sample of size  $n$  from a **Poisson** distribution with unknown mean  $\lambda > 0$  and  $\bar{Y} \equiv (Y_1 + Y_2 + \dots + Y_n)/n$  denote the sample mean. The variance of  $\bar{Y}$ ,  $\text{Var}(\bar{Y}) =$  (6) . A sufficient statistic for  $\lambda$  is (7) ; the minimum-variance unbiased estimator for  $\lambda$  is (8) ; the method-of-moments estimator of  $\lambda$  is (9) ; the maximum likelihood estimator for  $\ln[P(Y = 1)]$  is (10) . (N.B.: Use the notations defined in this question B to answer.)