單選題 (50%)，請用 2B 鉛筆作答於答案卡；第 1~10 題，每題 3 分；第 11~15 題，每題 4 分
Multiple Choice: Choose the best possible answer out of the choices from the list. (50%)

1. The manager for a grocery store needs to decide how many cabbages to order each day. A cabbage sells for $1.20 and costs the grocery store $0.75. Any unsold cabbage at the end of the day is donated to charity. If the manager purchases 100 cabbages for tomorrow and demand for lettuce that day is 250 heads, what is the profit?
   A) –$75
   B) $0
   C) $45
   D) $112.5
   E) $120

2. The following data represent the number of sweaters sold for each of the past six days for a department store. Using a three-period simple moving average, what is the sales forecast for Day 7?

Day	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6
Sales	6	13	8	4	9	14

   A) 6.5
   B) 7.0
   C) 9.0
   D) 11.5
   E) None of the above

3. When using the chi-square test and the observed frequencies are relatively close to the expected frequencies, which one of the following statements is the most likely to be true?
   A) The scenario will result in a smaller chi-square statistic, which makes it less likely to reject the null hypothesis.
   B) The scenario will result in a smaller chi-square statistic, which makes it more likely to reject the null hypothesis.
   C) The scenario will result in a larger chi-square statistic, makes it less likely to reject the null hypothesis.
   D) The scenario will result in a larger chi-square statistic, makes it more likely to reject the null hypothesis.
   E) Cannot conclude with the information provided

4. One airline company would like to investigate if a difference exists in the proportion of flights that arrive on-time between Alaska Airlines, Delta Airlines, Southwest Airlines, and United Airlines. The following data represent the number of on-time flights from random samples taken from each airline. What is the appropriate statistical test and what is the degree of freedom?

Airline	Alaska	Delta	Southwest	United	Total
Number On-time	90	95	85	60	330
Sample Size	100	120	100	80	400

   A) An F-test with the degree of freedom = 3
   B) An F-test with the degree of freedom = 4
   C) A z-test with the degree of freedom = 3
   D) A chi-square test with the degree of freedom = 3
   E) A chi-square test with the degree of freedom = 4

見背面

5. A Type I error occurs when performing one-way ANOVA when we _____.
   A) conclude that the group means are equal but, in fact, they are not equal.
   B) conclude that the group variances are equal but, in fact, they are not equal.
   C) conclude that the group means are not equal but, in fact, they are equal.
   D) conclude that the group variances are not equal but, in fact, they are equal.
   E) None of the above

6. A professor would like to test the hypothesis that the average number of minutes that a student needs to complete a statistics exam is equal to 60 minutes. A Type II error would occur if the professor concludes that the average exam time is _____.
   A) not equal to 60 minutes when, in reality, the average time is equal to 60 minutes
   B) not equal to 60 minutes when, in reality, the average time is greater than 60 minutes
   C) equal to 60 minutes when, in reality, the average time is less than 60 minutes
   D) equal to 60 minutes when, in reality, the average time is not equal to 60 minutes
   E) None of the above

7. The test statistics for one-way ANOVA follows the _____.
   A) normal distribution
   B) Student's t-distribution
   C) binomial distribution
   D) F-distribution
   E) Poisson distribution

8. Which is present in the regression model when independent variables within the model are highly correlated?
   A) Autocorrelation
   B) Homoscedasticity
   C) Heteroscedasticity
   D) Multicollinearity
   E) A dummy variable

9. What will happen if we increase the sample size when calculating a confidence interval while keeping the confidence level constant?
   A) result in a narrower (less precise) confidence interval
   B) result in a narrower (more precise) confidence interval
   C) result in a wider (less precise) confidence interval
   D) result in a wider (more precise) confidence interval
   E) None of the above

10. A data scientist observed the mean of a sample was below the median. What did this information suggest about the distribution?
   A) The distribution is symmetric.
   B) The distribution is skewed to the right or positively skewed.
   C) The distribution is skewed to the left or negatively skewed.
   D) The distribution is multimodal.
   E) There is insufficient information to determine the shape of the distribution.

接次頁

11. Alan likes to play bowling. He claims that his median bowling score for a round is greater than 190. The following random sample from 11 rounds is collected to test this hypothesis. Using $\alpha = 0.05$, which one of the following statements is true?

188	185	193	186	186	190	194	188	187	187	184

A) Because the p-value is greater than α, we reject the null hypothesis and conclude that the median bowling score is greater than 190.
B) Because the p-value is greater than α, we fail to reject the null hypothesis and cannot conclude that the median bowling score is greater than 190.
C) Because the p-value is less than α, we fail to reject the null hypothesis and cannot conclude that the median bowling score is greater than 190.
D) Because the p-value is less than α, we reject the null hypothesis and conclude that the median bowling score is greater than 190.
E) None of the above

12. A student would like to test the hypothesis that the average roundtrip airfare between New York and Paris is higher for a flight originating in New York when compared to a flight originating in Paris. The following data summarizes the sample statistics for roundtrip flights originating in both cities. Assume that the population variances are not equal. Using $\alpha = 0.05$, what is the conclusion for this hypothesis test?

	Originating City	
	New York	Paris
Sample mean	$1,315	$1,035
Sample size	16	16
Sample standard deviation	$270	$240

A) The average roundtrip airfare between New York and Paris is higher for a flight originating in Paris when compared to a flight originating in New York.
B) The average roundtrip airfare between New York and Paris is higher for a flight originating in New York when compared to a flight originating in Paris.
C) The average roundtrip airfare between New York and Paris is similar for a flight originating in New York and a flight originating in Paris.
D) The average roundtrip airfare between New York and Paris is not the same for a flight originating in New York and a flight originating in Paris, but we cannot tell which one is higher.
E) Information is not enough to reach a conclusion.

13. The table below shows the number of cars sold last month by a sample of seven employees at a car dealer and their number of years of sales experience. What is the correlation coefficient for this data?

Employee	No. 1	No. 2	No. 3	No. 4	No. 5	No. 6	No. 7
Experience	2	2	1	4	5	6	8
Sales	6	7	8	14	9	13	10

A) -0.441
B) -0.112
C) 0.553
D) 0.884
E) None of the above

見背面

14. To investigate the effectiveness of a diet, a random sample of 9 patients is drawn from a population of adult using the diet. The weight of each individual in the sample is taken at the start and at the end of the diet. Assume that the population of differences in weight before and after the diet follows a normal distribution. Suppose the mean decrease in weights over all 9 subjects in the study is 1.0 kilograms with the standard deviation computed as 4.5 kilograms. What conclusion can be made about the effectiveness of the diet, based on the test, at the 0.05 level of significance?
A) The diet is not effective.
B) The diet is effective.
C) The sample size is relatively small to make a conclusion.
D) The variance is relatively large to make a conclusion.
E) None of the above.

15. A hotel chain has 5 resorts in different small islands. The following table shows the hotel occupancies on 4 randomly selected days in the 5 locations. What is the conclusion of the test about whether the mean of the hotel occupancies is all the same for the 5 locations, at the 0.1 level of significance?

Occupancy	Location A	Location B	Location C	Location D	Location E
Observation 1	28	40	11	37	27
Observation 2	33	35	11	47	29
Observation 3	41	36	10	45	25
Observation 4	36	37	12	39	28

A) We can reject the null hypothesis and conclude that there is a difference in hotel occupancies in the 5 locations.
B) We can reject the null hypothesis and conclude that there is no difference in hotel occupancies in the 5 locations.
C) We cannot reject the null hypothesis and conclude that there is a difference in hotel occupancies in the 5 locations.
D) We cannot reject the null hypothesis and conclude that there is no difference in hotel occupancies in the 5 locations.
E) Information is not enough to reach a conclusion.

接次頁

簡答題 (50%)，請依題號順序作答，每小題作答字數不得超過 5 行，可以用中文作答。

Question I (18 points, 3 points each)

Suppose that a researcher collects data on houses that have sold in a particular neighborhood over the past year and obtains the regression results in the table shown below. Standard errors appear in parentheses below the estimated coefficients. For the purposes of these questions let

Price = sale price ($)
Size = house size (in square feet)
Bedrooms = number of bedrooms
Pool = binary variable (1 if house has a swimming pool, 0 otherwise)
View = binary variable (1 if house has a nice view, 0 otherwise)
Condition = binary variable (1 if realtor reports house is in excellent condition, 0 otherwise).

Dependent variable: *ln(Price)*					
Regressor	(1)	(2)	(3)	(4)	(5)
Size	0.00042 (0.000038)				
ln(Size)		0.69 (0.054)	0.68 (0.087)	0.57 (2.03)	0.70 (0.055)
$ln(Size)^2$				0.0078 (0.14)	
Bedrooms			0.0036 (0.037)		
Pool	0.082 (0.032)	0.071 (0.034)	0.072 (0.034)	0.073 (0.036)	0.074 (0.035)
View	0.037 (0.029)	0.027 (0.028)	0.026 (0.026)	0.025 (0.029)	0.024 (0.030)
*Pool*View*					0.0022 (0.10)
Condition	0.13 (0.045)	0.12 (0.035)	6.63 (0.53)	7.02 (7.50)	6.60 (0.40)
Intercept	10.97 (0.069)	6.60 (0.39)	6.63 (0.53)	7.02 (7.50)	6.60 (0.40)
$\bar{R}^2$	0.72	0.74	0.73	0.73	0.73

16. What is the expected change in price of building a 500-square foot addition to a house?
17. Is it better to use *Size* or *ln(Size)* to explain house prices? Explain.
18. Is the quadratic term $ln(Size)^2$ important? Explain.
19. Construct a 95% confidence interval for the estimated effect of an additional bedroom.
20. Using the regression in column (5) to compute the expected change in price when a pool is added to a house with a view.
21. Explain why the coefficients on *Condition* in columns (1) and (2) are far different from those in columns (3), (4), and (5).

見背面

Question II (32 points; 8 points each)

Gauss-Markov Theorem: If assumptions A1—A5 hold, the least squares estimators $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ are the best linear unbiased estimators of the parameters in the multiple regression model $\beta_0, \beta_1, \dots, \beta_k$, respectively.

A1: Observations on $(y_i, \mathbf{x}_i) = (y_i, x_{i1}, x_{i2}, \dots, x_{ik})$ satisfy the population relationship $y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + u_i$.

A2: Given all the explanatory variable observations $\mathbf{X} = \{x_i, i = 1,2, \dots, N\}$, the conditional expectation of the error term is $\mathrm{E}(u_i|\mathbf{X}) = 0$.

A3: Conditional on $\mathbf{X}$, the variance of the error term is $\mathrm{var}(u_i|\mathbf{X}) = \sigma^2$.

A4: Conditional on $\mathbf{X}$, the covariance between different error terms is $\mathrm{cov}(u_i, u_j|\mathbf{X}) = 0$ for $i \neq j$.

A5: The only values of $c_1, c_2, \dots, c_k$ for which $c_1 x_{i1} + c_2 x_{i2} + \cdots + c_k x_{ik} = 0$ for all observations $i = 1,2, \dots, N$ are the values $c_1 = c_2 = \cdots = c_k = 0$.

22. Does the Gauss-Markov theorem say that the estimators $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ have the minimum variance of all possible estimators? Explain.
23. Suppose that u is independent of the explanatory variables, and it takes on the values -2, -1, 0, 1, and 2 with equal probability of 1/5. Which Gauss-Markov assumption(s) does this violate. Explain.
24. Consider the regression model $y_i = \beta_0 + \beta_1 x_{1i} + u_i$ where the pairs (y_i, x_{1i}), $i = 1,2, \dots, N$, are random independent draws from a population, $x_{1i} \sim N(0,1)$, and $\mathrm{E}(u_i|x_{1i}) = x_{1i}^2 - 1$. Note that $\mathrm{E}(u_i) = \mathrm{E}_x[\mathrm{E}(u_i|x_{1i})]$ and $\mathrm{E}_x(x_{1i}^2) = 1$. Will the least squares estimator for β_1 be unbiased? Explain.
25. Is the assumption A6: $u \sim N(0, \sigma^2)$ essential to carry out hypotheses tests? Explain.

接次頁

t Table

Degrees of freedom	Significance level					
	20% (0.20)	10% (0.10)	5% (0.05)	2% (0.02)	1% (0.01)	0.1% (0.001)
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.598
3	1.638	2.353	3.182	4.541	5.841	12.941
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.859
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.405
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073
16	1.337	1.746	2.120	2.583	2.921	4.015
17	1.333	1.740	2.110	2.567	2.898	3.965
18	1.330	1.734	2.101	2.552	2.878	3.922
19	1.328	1.729	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850
21	1.323	1.721	2.080	2.518	2.831	3.819
22	1.321	1.717	2.074	2.508	2.819	3.792
23	1.319	1.714	2.069	2.500	2.807	3.767
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	2.485	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
27	1.314	1.703	2.052	2.473	2.771	3.690
28	1.313	1.701	2.048	2.467	2.763	3.674
29	1.311	1.699	2.043	2.462	2.756	3.659
30	1.310	1.697	2.042	2.457	2.750	3.646
40	1.303	1.684	2.021	2.423	2.704	3.551
60	1.296	1.671	2.000	2.390	2.660	3.460
120	1.289	1.658	1.980	2.158	2.617	3.373
∝	1.282	1.645	1.960	2.326	2.576	3.291

Image source: https://jimgrange.wordpress.com/2015/12/05/statistics-tables-where-do-the-numbers-come-from/

試題隨卷繳回