

第一部分：單選題，共 20 題，每題 2.5 分，共 50 分，請於試卷上「選擇題作答區」內依序作答。

Without any endorsement, SKII's average daily sales in Taipei are 500 units. Sharon, the brand manager of SKII as well as a graduate of NTU MBA, believes that promoting the product by favorable endorsers will increase the amount of sales. She calculates that the improved sales by additional 75 units daily would pay for the cost of the endorsement. Any further increase in sales would create an extra profit. To test her belief that endorsers would lead to a positive effect on profit, 田中千繪 is invited to endorse SKII. Please answer the following 6 questions (1-6).

1. To examine Sharon's belief, the null hypothesis is
 - A) $H_0: u = 500$
 - B) $H_0: u \leq 500$
 - C) $H_0: u = 575$
 - D) $H_0: u \leq 575$
2. If the cost of endorsement is not considered high in relation to administrative expenses, and Procter & Gamble is eager to raise the profit margin in Asia market, which of the following can best be used as the significance level set by Sharon?
 - A) .01
 - B) .05
 - C) .10
 - D) depending on target customers' buying behaviors
3. If the average daily sales approaches 550 units after 田中千繪 serves as the endorser, then which alternative hypothesis should Sharon use to test her belief?
 - A) $H_a: u > 575$
 - B) $H_a: u \neq 575$
 - C) $H_a: u > 550$
 - D) $H_a: u \neq 550$
4. Continued from the previous questions, let $P(x)$ be the probability of x , and \bar{u} be the sample mean. If $P(\bar{u} > 550) = .75$, the power of the test is
 - A) .25
 - B) .75
 - C) .05
 - D) Unknown

見背面

5. Continued from the previous questions, suppose that when the average daily sales are at least 600 units, Sharon would hate to make the mistake of not using the favorable endorser, such as 田中千繪. And $P(\bar{u} > 550, \text{ given that } u = 600) = .96$. Which of the following is CORRECT?
- A) The probability of rejecting the null hypothesis when it is false is .04.
 - B) The probability of rejecting the null hypothesis when it is true is .04.
 - C) The probability of not rejecting the null hypothesis when it is true is .95.
 - D) None of the above
6. Continued from the previous questions, Sharon's supervisor, Ivory, has a different theory on the endorsement effect. Ivory, referring to her long-term industrial experience, believes that the use of 林志玲 as the endorser would lead the average daily sales of SKII to 590 units. If the average daily sales amount to 600 units after 林志玲 serves as the endorser, and $P(\bar{u} > 600, \text{ given that } u = 590) = .015$. What is the p-value to test Ivory's theory?
- A) .05
 - B) .015
 - C) .03
 - D) None of the above
7. Grant, a senior marketing supervisor in HP, have noticed for a long time that there is a one-to-one correspondence between the number of butterflies observed in his front yard and the number of complaints filed by key accounts on any given day. For example, without exception, whenever a complaint was received, Grant could only find one butterfly in the front yard; whenever two complaints were received, Grant could only find two butterflies in the front yard, and so on. Retrieving from the vague scenes back to 20 years ago, Grant still remembers by what have been learned in the statistics class that he could use $Y = X$ to represent his observations, where Y is the number of butterfly observed and X the number of complaints received. Which of the following is correct?
- A) The number of complaints received is the cause and the number of butterfly observed is the effect.
 - B) The number of complaints received is the effect and the number of butterfly observed is the cause.
 - C) The covariance between the number of complaints received and the number of butterfly observed is positive.
 - D) Grant can also use the regression analysis and the coefficient of the

接次頁

predictor would be 1.

8. Asus's VP in strategy would like to evaluate the performance of customer relationship management (CRM) in cellular phone division and motherboard division. Three 9-point (-4 to 4) differential semantic scales are used to measure customers' perceptions on Asus's cellular phones and motherboards. The evaluation index is obtained by averaging three scales. By collecting 300 responses from Asus's cellular phone users, the mean evaluation index is 2.8 and the standard error is 3.7. Another 250 responses from Asus's motherboard customers reported a mean of 2.6 and a standard error of 1.3 on the evaluation index. Which of the following statements is CORRECT?
- A) In order to get a more accurate result in analysis, it is better to transform the bipolar scale (e.g., -4 to 4) to the unipolar scale (e.g., 1 to 9).
 - B) When the sample size is getting larger, the sample mean is more likely to get closer to the population mean.
 - C) Asus's VP in strategy should conclude that the performance of CRM in cellular phone and motherboard divisions are doing equivalently well, because according to the analysis of one-way ANOVA, the mean difference is much smaller than the standard deviations.
 - D) Asus's VP in strategy should conclude that the performance of CRM in cellular phone and motherboard divisions are doing equivalently well, because according to the analysis of one-way ANOVA, the treatment mean square is much smaller than the error mean square.

Johnson & Johnson targeting on the young girls attempts to test 4 styles of Hello Kitty package designs for a new line of shampoo. 19 stores with approximately equal sales volumes were selected as the experimental units. The following table is the sales volume of 19 stores in one week.

<i>Style 1</i>	<i>Style 2</i>	<i>Style 3</i>	<i>Style 4</i>
20	16	26	19
23	14	33	12
18	15	27	11
17	17	28	15
	11	22	10

Based on the above information, please answer the next 10 questions (9-18).

9. Laura, the product manager, would like to use the regression analysis to examine

見背面

whether the design style may produce different sales volume. She still remembers that dummy variables should be used in such a regression analysis. How many dummy variables should Laura use in the regression analysis?

- A) 3
- B) 4
- C) 5
- D) None of the above

Continued from the above question, in order to get the regression output, please help Laura fill in the missing values

REGRESSION OUTPUT

<i>Regression Statistics</i>	
Multiple R	"A"
R Square	"B"
Adjusted R Square	"C"
Observations	19

ANOVA

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>
Regression	"D"	"F"	"H"	"I"
Residual	15	158.20	10.55	
Total	"E"	"G"		

10. What is the value of "A"?

- A) .64
- B) .89
- C) .52
- D) .74

11. What is the value of "B"?

- A) .27
- B) .41
- C) .55
- D) .79

12. What is the value of "C"?

接次頁

- A) .61
- B) .69
- C) .77
- D) .75

13. What is the value of "D"?

- A) 3
- B) 4
- C) 5
- D) 6

14. What is the value of "E"?

- A) 18
- B) 19
- C) 20
- D) 21

15. What is the value of "F"?

- A) 327.49
- B) 451.83
- C) 588.22
- D) 658.72

16. What is the value of "G"?

- A) 610.03
- B) 816.92
- C) 485.69
- D) 746.42

17. What is the value of "H"?

- A) 196.07
- B) 150.61
- C) 164.68
- D) 219.57

18. What is the value of "I"?

- A) 14.28
- B) 18.59



見背面

- C) 15.61
- D) 20.81

19. Which of the following statements is CORRECT?

- A) In a two-way ANOVA, if the interaction is significant, either main effect is not supposed to be significant.
- B) In a two-way ANOVA, if the main effect of factor A is not significant, the mean difference between A1 and A2 at the level of B1 can still be significant.
- C) In a two-way ANOVA, if the overall interaction effect is not significant, the interaction of the i th level of factor A and j th level of factor B can still be significantly different from zero.
- D) In a one-way ANOVA, when all factor level means are equal, the expected value of the treatment mean square should be smaller than the expected value of the error mean square such that null hypothesis will not be rejected.

20. Which of the following statement is CORRECT?

- A) In a chi-square test of independence, if marital status is found associated with employment status, then we can predict that the number of married persons who are unemployed is very close to the number of unmarried persons who are unemployed.
- B) When there is not a significant match between the observed and expected values, the chi-square will be significant.
- C) When the test of chi-square is not significant, the chi-square value can sometimes be negative.
- D) The distribution of chi-square is generally symmetric, just like a bell shape. Therefore, statisticians sometimes use normal distributions to approximate chi-square distributions.

第二部分：計算與簡答題，共 7 題，佔 50 分，請於答案卷上依序作答，並註明題號。

1. (5 points) Suppose that the number of minutes between successive arrivals of the Taipei MRT train at a station is uniformly distributed on $(0,1)$. Passengers arrive according to a Poisson process with a rate λ . Suppose a train has just left the station. Let X denote the number of people who get on the next train. Find
 - (a) $E(X)$
 - (b) $\text{Var}(X)$

接次頁

2. (10 points) Data on the monthly service fee in dollars when a customers' account falls below the minimum required \$5000 balance for a sample of 26 banks yields a sample mean of 71.5 and a sample standard deviation of 30.55.
- Please find a 95% confidence interval for the population mean monthly service fee.
 - Can we say that with probability 0.95 the interval obtained in (a) will include the population mean? If not, then what is the correct interpretation? Explain briefly.
 - Data on 10 more banks were added so that the sample size increase to 36 and we obtained a 95% confidence interval [59.55, 80.45]. What is the mean of the sample of 36?
 - Assume that the estimated population standard deviation in (c) equals the true population s.d. What is the minimum sample size required so that the length of the 95% interval is less than 16?
3. (7.5 points) The output of a fully-functioning power plant has a Gaussian distribution, with a mean of 10MW and a standard deviation of 0.5 MW. Upper and lower control limits are established as 11.0 MW and 9.0 MW respectively. If a certain component fails, the mean output drops to 9.0 MW, without changing the standard deviation.
- What is the probability that the output lies between the control limits, when the power plant is working properly?
 - What is the probability that the output lies between the control limits during one of the partial failures?
 - If the power plant works properly 99% of the time, what is the probability that going outside the control limits means a component has failed?
4. (2.5 points) Suppose $\text{Var}(X) = 49$ and $\text{Var}(Y) = 9$. For what values of the correlation of X and Y is the $\text{Var}(.2X + .8Y)$ less than $\text{Var}(Y)$?
5. (5 points) In a study to investigate the effect of speed on accident severity, 5000 accident reports of fatal automobile accidents were examined, and the vehicle speed at impact was recorded for each one. It was determined that the average speed was 52 mph with a standard deviation of 15 mph. Using Chebyshev's Theorem, write a statement about vehicle speed in fatal accidents.
6. (5 points) Suppose you are a consultant providing professional service to a particular industry. A major firm is interested in your service. There are three options for calculating your fee based on the performance rating. For option 1, the fee equals

見背面

to the performance rating multiply by \$120,000. Under option 2, you will be paid \$300,000 in the beginning of the contract period. If your performance rating is higher than 3, you will receive a bonus of \$150,000 at the end of the contract. According to option 3, you will be paid a fee of \$750,000 up front, but if your performance rating is lower than 5, for each point lower your client will get a rebate of \$200,000 from you. From past experience, you estimated the following probability distribution.

Performance Rating	1	2	3	4	5
Probability	0.1	0.3	0.3	0.2	0.1

If you choose the option with highest expectation, which one should you take? Justify your answer.

7. (15 points) Suppose that the following data represents serum cholesterol levels in mg/dL of a sample of 16 people before and after they adopted a vegetarian diet.

Subject	Before	After	Before - After
1	148	160	- 12
2	200	166	34
3	194	154	40
4	188	180	8
5	236	198	38
6	256	188	68
7	210	180	30
8	198	202	- 4
9	187	172	15
10	175	166	9
11	190	160	30
12	154	150	4
13	165	162	3
14	168	166	2
15	192	178	14
16	200	158	42

- Make a stem-and-leaf plot of the change in cholesterol
- Describe the shape of the data of the change in cholesterol.
- Compute the median of the change in cholesterol.
- Compute the mode of the change in cholesterol.
- Find the interquartile range of the change in cholesterol.
- Name the most appropriate statistical test to adopt if we want to know whether a vegetarian diet has an effect on cholesterol level.