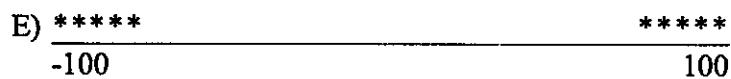
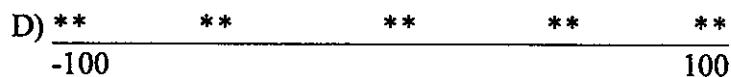
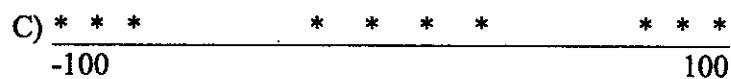
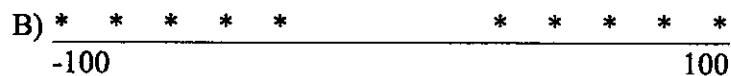


Please choose the most suitable answer among all alternatives for each question and clearly mark it on the answer sheet. Each question is worth 5 points. ※ 注意：請用 2B 鉛筆作答於答案卡，並先詳閱答案卡上之「畫記說明」。

1. For the following dot plots, which dataset has the largest standard deviation?



2. There are three factories: A, B, and C, with 100, 1,000, and 10,000 workers, respectively. The mean wages are the same across all three factories. Which of the following statements is true?

- A) The standard deviations of wages in all three factories are equal.
- B) The ranges of wages in all three factories are equal.
- C) The variance of wages in Factory A is the smallest.
- D) The standard deviation of wages in Factory C is the smallest.
- E) None of the above.

3. Which of the following statements is correct in a negatively skewed distribution?

- A) The arithmetic mean is greater than the mode.
- B) The arithmetic mean is greater than the median.
- C) The difference between the third quartile and the median is equal to the difference between the first quartile and the median.
- D) The difference between the third quartile and the median is less than the difference between the first quartile and the median.
- E) None of the above.

4. What does the intercept in a binary linear regression represent?

- A) The intercept indicates the strength of the relationship between x and y.
- B) The intercept is the expected value of the independent variable when the dependent variable is zero.
- C) The intercept is the expected value of the dependent variable when the independent variable is zero.
- D) The intercept is just a random constant.
- E) None of the above.

5. In a series of n games, the probability of winning each game is 0.5, and each game is independent. A player wins a prize if at least 60% of the games in the series are won. The possible values for n are 5, 20, and 100. Which value of n should the player choose to maximize the probability of winning a prize?
- A) $n=5$
 - B) $n=20$
 - C) $n=100$
 - D) It does not matter since the probabilities are all the same.
 - E) The answer cannot be concluded based on the information provided.
6. The dataset consists of four numbers with a mean of 27. If the mean of the smallest three numbers is 20 and the range of the dataset is 36, what is the mean of the largest three numbers?
- A) 31
 - B) 32
 - C) 33
 - D) 34
 - E) The answer cannot be concluded based on the information provided.
7. If the variance of the first i natural numbers is 14, the variance of the first j even natural numbers is 16, and the variance of the first k odd natural numbers is also 16, which of the following statements is correct?
- A) $i = 11$
 - B) $j = 7$
 - C) $j \neq k$
 - D) all of the above
 - E) none of the above
8. A complex electronic device consists of three components: A, B, and C. The probability of failure for each component in a year is 0.03 for A, 0.05 for B, and 0.1 for C. If any component fails, the entire device will fail. Assuming the components fail independently, what is the closest probability that the device will not fail in a given year?
- A) 0.00015
 - B) 0.82
 - C) 0.83
 - D) 0.95
 - E) 0.97
9. An exam uses the following scoring system to discourage students from guessing or choosing randomly on multiple-choice questions: 10 points for each correct answer, 3 points for each unanswered question, and 0 points for each incorrect answer. If each question has five answer choices, what is the minimum number of choices a student must eliminate before guessing among the remaining choices becomes more advantageous than leaving the question unanswered?
- A) 0
 - B) 1
 - C) 2
 - D) 3
 - E) 4

10. The commuting time for a student traveling from home to a college campus follows a normal distribution with a mean of 25 minutes and a standard deviation of 5 minutes. If the student leaves home at 8:40 AM, what is the closest probability that they will arrive at the campus later than 9:10 AM?
- A) 0.05
B) 0.16
C) 0.68
D) 0.84
E) 0.95
11. In a hypothesis test for the equality of two population means with unknown but equal variances, the test statistic is based on:
- A) The pooled variance estimate.
B) The difference between the sample variances.
C) The sum of the sample means.
D) The larger of the two sample variances.
E) The square of the standard deviations.
12. When performing a two-way ANOVA, a significant interaction effect implies:
- A) Neither factor has a significant main effect.
B) The effects of the two factors are independent.
C) The effect of one factor depends on the level of the other factor.
D) Both factors have significant main effects.
E) The dependent variable is not normally distributed.
13. Which of the following is NOT an assumption of ordinary least squares (OLS) regression?
- A) The residuals are normally distributed.
B) The relationship between the dependent and independent variables is linear.
C) The residuals are homoscedastic.
D) The independent variables are uncorrelated.
E) The residuals are independent of each other.
14. In hypothesis testing, what does the power of a test represent?
- A) The probability of rejecting a true null hypothesis.
B) The probability of rejecting a false null hypothesis.
C) The significance level of the test.
D) The probability of failing to reject a true null hypothesis.
E) The probability of obtaining a Type I error.

For Questions 15-17, please read the scenario and output below.

A behavioral scientist wants to investigate whether three different teaching methods (Method A, Method B, and Method C) lead to different academic performance in students. A total of 90 students are randomly assigned to one of the three methods, with 30 students per group. At the end of the semester, their test scores (out of 100) are recorded.

The scientist performs a one-way ANOVA, and the results are summarized below:

題號： 279

國立臺灣大學 114 學年度碩士班招生考試試題

科目： 統計學(H)

節次： 2

題號： 279

共 5 頁之第 4 頁

Source	df	SS	MS	F	p-value
Between Groups	2	450.00	225.00	7.50	0.001
Within Groups	87	2610.00	30.00		
Total	89	3060.00			

15. What is the null hypothesis for the one-way ANOVA conducted in this scenario?

- A) All groups have the same sample size.
- B) The variances within all groups are equal.
- C) The mean test scores are equal across all three teaching methods.
- D) At least one group has a higher variance than the others.
- E) The distribution of scores is normal for all three methods.

16. What conclusion can be drawn from the one-way ANOVA output?

- A) All three teaching methods lead to the same academic performance.
- B) At least one teaching method leads to a significantly different mean test score compared to the others.
- C) The variances of the test scores differ significantly across the three groups.
- D) Teaching method is not a significant factor in academic performance.
- E) Post-hoc tests are unnecessary because the p-value is less than 0.05.

17. Which of the following assumptions is critical for the validity of both the ANOVA and the post-hoc Tukey's HSD tests?

- A) The mean differences among groups must be large.
- B) The sample sizes must be equal in each group.
- C) The population variances must be equal across all groups.
- D) The sample sizes must be greater than 30 in each group.
- E) The F-statistic must be greater than 5.

For Questions 18-20, please read the scenario and output below.

A marketing analyst is investigating the impact of advertising expense and product pricing on sales revenue for a retail chain. Across the chain's 200 stores, the analyst collects data on the following variables and enters in a multiple regression model:

- Y : Sales revenue (in thousands of dollars),
- X_1 : Advertising expense (in thousands of dollars),
- X_2 : Product unit price (in dollars),
- X_3 : The interaction term of X_1 and X_2

接次頁

The regression model output is as follows, with $R^2=0.72$, adjusted $R^2=0.71$, and standard error of the estimate = 8.5.

Predictor	Coefficient (β)	Standard Error	t-statistic	p-value
Constant	20	5	4.00	0.001
X_1	2.5	0.3	8.33	0.001
X_2	-1.2	0.5	-2.40	0.018
X_3	0.04	0.01	4.00	0.001

18. What can be concluded based on the regression coefficients?

- A) Advertising expense has no significant effect on sales revenue.
- B) Advertising expense increases sales revenue at a constant rate regardless of product unit price.
- C) Advertising expense increases sales revenue, but the effect depends on the product unit price.
- D) Advertising expense reduces sales revenue when product unit price is high.
- E) Advertising expense reduces sales revenue regardless of product unit price.

19. Which of the following statements about R^2 is correct?

- A) 72% of the variability in advertising expense is explained by sales revenue.
- B) 72% of the variability in sales revenue is explained by advertising expense, price, and their interaction.
- C) 72% of the variability in sales revenue is explained by advertising expense alone.
- D) $R^2=0.72$ indicates the model overfits the data.
- E) 72% of the model's predictions are accurate.

20. Which of the following assumptions must hold true for the p-values of the predictors to be valid?

- A) X_3 must be statistically significant.
- B) The residuals must have constant variance across all levels of the predictors.
- C) The residuals must be positively correlated with the predictors.
- D) The predictors must be uncorrelated with each other.
- E) The sample size must be greater than 100.