

※ 考生請注意：本試題不可使用計算機。請於答案卷(卡)作答，於本試題紙上作答者，不予計分。

單選題 (72%): 每題 4 分，答錯者倒扣 1 分

1. If the two populations are normal, but we don't know whether these two populations have equal variance or not. What is the proper probability distribution can be used to test this hypothesis?
 - A. Standard normal distribution
 - B. Student's t distribution
 - C. F distribution
 - D. χ^2 distribution
2. The most common data types of random variables are quantitative variable (continuous and discrete), Ranked (ordinary) variables and categorical data. For the quantitative variable is
 - A. Length, weight, temperature, volume, and height
 - B. Tumor grading and staging, level of customer satisfaction, ...
 - C. Species, gender, genotype, healthy/disease, and material status
 - D. None of above
3. According to previous question (1.), which probability density function can be usually more appropriate for probability determination when the criteria for distribution approximation is not applicable?
 - A. The binomial distribution
 - B. Standard normal distribution
 - C. The Poisson distribution
 - D. The uniform distribution
4. The following option is NOT required for the estimation of sample size?
 - A. The intrinsic variability among observations, σ , standard deviation of population
 - B. Type I error (α)
 - C. The power of the test
 - D. Degree of freedom
5. Following up the question (2.), a simple rule of thumb is that normal approximation to the binomial distribution is acceptable for value n (number of trials) and p (probability of a success) in the binomial density function with the criteria listed below:
 - A. $np > 5$
 - B. $n(1 - p) > 5$
 - C. Both
 - D. None of above
6. If the level of significance of a hypothesis test is decline from 0.05 to 0.01, the type II error of the test
 - A. Will also decrease from 0.05 to 0.01
 - B. Will not change

- C. Will increase
D. Will decrease
7. Before applying two-sample test of hypothesis to your data, which following option is NOT necessary:
A. Check the power of the test
B. Check whether population's distribution is parametric or non-parametric
C. Check if the two populations have equal variances
D. Check the dependence between two samples
8. Suppose that on a field trip to Guatemala (瓜地馬拉) you decide to study handsome fungus beetles (偽瓢蟲科). The population you investigate is composed of 60% females and 40% males. Additionally, it has two color morphs, dull brown (70%) and bronze (30%). Half of all the insects are dull brown females. What is the probability that a randomly selected individual is either dull brown or female?
A. 0.8
B. 0.7
C. 0.2
D. 0.6
9. The fundamental difference between continuous and discrete variables is the probability that a continuous random variable (X) takes any specific value (c), $P(X=c)$,
A. depends on the probability density function
B. is very closed to 1.0
C. is equal to zero
D. is at least 0.5
10. Which following distribution is more properly used to estimate confidence interval for the mean when the population variance is unknown?
A. Binomial distribution
B. Student's t distribution
C. F distribution
D. χ^2 distribution
11. About a confidence interval (C.I.) for the population mean, which following description is NOT correct?
A. The "confidence interval" term directly refer to a probability rather than an interval
B. To determine confidence interval requires a point estimate, the sample mean, and the standard error of the mean
C. The range of 99% C.I. is wider than that of 95% C.I.
D. None of above
12. There are several typical steps in a statistical test of hypothesis. Which typical steps in the following is NOT necessary?
A. Formulate null and alternative hypotheses based on problem statement
B. Determine the critical value and P value based on the level of significance and appropriate test

- statistic
- C. Compare the test statistic to critical value and compare the P value to the chosen level of significance
- D. Estimate sample size based on confidence interval of the sample mean
13. What type of error is the probability of rejecting a true null hypothesis?
- A. Type I (α)
- B. Type II (β)
- C. Either Type I or Type II, depending on the level of significance
- D. Either Type I or Type II, depending on whether the test is one-tailed or two-tailed
14. For formulating null and alternative hypotheses in a statistical test of hypothesis, which following statement is correct?
- A. There are three forms of hypotheses: right tail, left tail, and two-tail formulations. We can test these three forms of hypotheses simultaneously using the same data set.
- B. The alternative hypothesis is always used for statistical testing.
- C. Both null and alternative hypotheses are mutually exclusive and all inclusive.
- D. None of above statements
15. The chi-square test is commonly used for your categorical data, but chi-square distribution is characterized by continuous variable, and originated from normal distribution. When applying this method under small degree of freedom, it turns out testing discrete data using the probability distribution with a continuous random variable. What is the proper way to improve this statistical testing?
- A. Change to Fisher's exact test
- B. Use Bonferroni correction for further analysis
- C. Change to Wilcoxon rank sum test
- D. Use Yales' correction for further analysis
16. In which following condition, Student's t distribution will approximate to the standard normal distribution?
- A. Decrease sample size
- B. Raise confident level
- C. Increase degree of freedom
- D. Lower Type I error
17. Beside sample variance, standard deviation, and range to obtain measures of the spread and variability, "quartiles" divide a distribution in quarters and generate five-number summaries. The visual counterpart to a five-number summary is box plot. What the following statistics is not considered into a box plot?
- A. Median
- B. IQR (interquartile range)
- C. Outliers

D. Mean

18. We define the margin of error for a $(1-\alpha)$ 100% confidence interval for a population proportion to be

$m = z_{1-\alpha/2} SE_{\hat{p}} = z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$. The researcher specifies m and the confidence level, whose value of $z_{1-\alpha/2}$ is determined. Thus, to solve for the sample size n we simply need \hat{p} . However, since researcher

has not yet done the study, is unknown. The way to solve for this equation is to fix the value (n) and maximize the term of $\hat{p}(1-\hat{p})$. By doing this, we will obtain

A. $m \leq z_{1-\alpha/2} \sqrt{\frac{0.50}{n}}$

B. $m \leq z_{1-\alpha/2} \sqrt{\frac{0.90}{n}}$

C. $m \leq z_{1-\alpha/2} \sqrt{\frac{0.99}{n}}$

D. $m \leq z_{1-\alpha/2} \sqrt{\frac{0.25}{n}}$

簡答題 (28%):

Hint for mathematical expression: \bar{X} → sample mean, n → sample size, s → sample standard deviation, population mean → μ , population size → N , and σ → population standard deviation

- (a) What's difference between sample standard (標準差) deviation and standard error/standard error of the mean (SEM, 標準誤) (5%)? (b) In the statement of scientific report or article, there are two ways to express confidence interval for the mean. How? (5%) (c) Based on the same sample, what is the major difference between 95% and 99% confident interval? (5%) (d) What is the general sampling distribution to estimate the confidence interval of population mean and variance respectively? (5%)
- What are the (a) major concerns/assumptions (5%), and (b) proper statistical testing methods (3%) of for continuous data type under non-parametric approach?