國立成功大學

110學年度碩士班招生考試試題

編　號：241

系　所：統計學系

科　目：統計學

日　期：0203

節　次：第 3 節

備　註：不可使用計算機

編號： 241

國立成功大學 110 學年度碩士班招生考試試題

系　　所：統計學系

考試科目：統計學

考試日期：0203，節次：3

※ 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。

A. Multiple Choice (17 × 3% = 51%)

1. Let X and Y be two independent standard normal random variables, and U=X/Y, what is the expected value of U ?
   (A) ∞　(B) 0　(C) 1　(D) None of the above

2. Which of the following statements is(are) correct about Poisson distribution?
   i. The median is greater than the mean.
   ii. It is a discrete distribution
   iii. The variance is equal to its mean.
   (A) i　(B) ii　(C) iii　(D) i & ii　(E) i & iii　(F) ii & iii　(G) All　(H) None

3. Two events A and B are disjoint, which of the following statements is(are) correct?
   i. $P(A \cap B) = P(A)P(B)$
   ii. $P(A \cup B) = P(A) + P(B)$
   iii. A and B are independent.

   (A) i　(B) ii　(C) iii　(D) i & ii　(E) i & iii　(F) ii & iii　(G) All　(H) None

4. The data type of room temperature is
   (A) Interval　(B) Ratio　(C) Nominal　(D) Ordinal

5. A household survey is conducted to study the individual income of Tainan city, 1,000 households are randomly selected and each member more than 18 years old in the selected households are asked for their monthly income. The sampling design is
   (A) Simple random sampling　(B) Stratified sampling　(C) Cluster sampling　(D) systematic sampling

6. Two events A and B are independent, and $P(A) = 1/5$, $P(A \cup B)=2/5$, then $P(B) = ?$
   (A) 1/4　(B) 1/5　(C) 1/6　(D) None of the above

7. For a test scores of a midterm exam, 100 students had the mean of 60 and standard deviation of 5, how many students scored between 50 and 70 at least?
   (A) 75%　(B) 95%　(C) 80%　(D) Cannot be determined unless the distribution is provided.

8. David conducted an experiment to practice what he has learned in class. He randomly selected colored chips out of an urn, in which there are 20 chips, 12 of them are red and others are blue.

At first he selected 10 chips with replacement, and then did the same selection again but without replacement. Let X and Y be the number of blue chips of the first and second selection, respectively. Which of the following statements is(are) correct?

i. E(X) = E(Y)
ii. Both of the ranges of X and Y are 0,1,2,.....,9,10.
iii. Var(X) > Var(Y)

(A) i (B) ii (C) iii (D) i & ii (E) i & iii (F) ii & iii (G) All (H) None

9. Which of the following statements is(are) correct?
   i. The mean, median, and mode are the same for the normal distribution.
   ii. The assumption of normality is not necessary to estimate the slope and intercept in a simple regression model by the least squared method.
   iii. The significance level α is the observed type I error rate.

   (A) i (B) ii (C) iii (D) i & ii (E) i & iii (F) ii & iii (G) All (H) None

10. A survey was conducted to study the proportion of high-income (Monthly income > 150,000 NTD) individual in Tainan City, a random sample (without replacement) of 1,000 individuals were selected and the number of sampled high-income individuals was 30. Which of the following statements is(are) correct?
    i. The sampling distribution of the sample mean is approximately a normal distribution.
    ii. The estimated variance of the sample proportion is smaller than 0.03×(1-0.03)/1000.
    iii. A Poisson distribution is appropriate to model the number of high-income individual selected.

    (A) i (B) ii (C) iii (D) i & ii (E) i & iii (F) ii & iii (G) All (H) None

11. Which of the following statements is(are) true?
    i. Under simple random sampling with or without replacement, the average of the observed values is an unbiased estimator of the population mean irrespective of the population distribution.
    ii. The Central Limit Theorem guarantees that the observed data will be normally distributed if the sample size is large enough.
    iii. If three events A, B and C are independent, then $(A \cap B)$ and C are independent as well.

    (A) i (B) ii (C) iii (D) i & ii (E) i & iii (F) ii & iii (G) All (H) None

12. Which of the following states is(are) correct?
    i. The p-value is not affected by the significance level α.
    ii. The z-score, correlation coefficient, and the coefficient of variation will not change with the change of measurement.

iii. We cannot control the type II error rate of a hypothesis test.
(A) i (B) ii (C) iii (D) i & ii (E) i & iii (F) ii & iii (G) All (H) None

13. John did a goodness-of-fit test for independence on a 2x2 contingence table and the value of the test statistic was 3. Which of the following statements is(are) correct?
i. He has a significant evidence to reject the null hypothesis under a 95% confidence level.
ii. He has a significant evidence to reject the null hypothesis under a 90% confidence level.
iii. A table of χ^2 distribution is necessary to make a decision.
(A) i (B) ii (C) iii (D) i & ii (E) i & iii (F) ii & iii (G) All (H) None

14. Which of the following statements is(are) true?
i. Range is better than interquartile range to describe the dispersion of our data if extreme value exists
ii. Median is better than mean to describe the location of our data if extreme value exists.
iii. Correlation coefficient is a unit-free measurement.
(A) i (B) ii (C) iii (D) i & ii (E) i & iii (F) ii & iii (G) All (H) None

15. Which of the following statements is(are) true?
i. The Kruskal-Wallis test can be compared to the One-way ANAVA procedure.
ii. The Wilcoxon sighed-rank test is a nonparametric version of one-sample t test.
iii. The R^2 (multiple coefficient determination) increases as the number of explanatory variable increases.
(A) i (B) ii (C) iii (D) i & ii (E) i & iii (F) ii & iii (G) All (H) None

16. To exam $H_0: \mu \leqq 100$ v.s. $H_a: \mu > 100$ based on the one-sample z test, a set of sample was collected. The sample size was 100, and the population variance is given as 400. What is the decision rule to reject H_0 with a level of significance of α=0.05 based on the critical value approach? Let $\bar{x}$ be the sample mean.
(A) $\bar{x} \geq 96.08$ (B) $\bar{x} \leq 96.08$ (C) $\bar{x} \geq 92.16$ (D) $\bar{x} \leq 92.16$ (E) Cannot be determined.

17. Refer to 16, what is range of the type II error rate β when $\bar{x}$ is equal to the critical value, and the true μ is 93?
(A) $0.5 > \beta > 0.1$ (B) $0.1 > \beta > 0.05$ (C) $0.05 > \beta > 0.01$ D) $\beta < 0.01$ (E) Cannot be determined unless a standard normal table is provided.

B. Blank Filling (I-X : 10 × 2% = 20%; XI-XIII : 3 × 3%=9% ; 29%)

1. Jack did a simple linear regression to exam the relationship between two variables x and y for his homework assignment. Unfortunately, the computer output was damaged during a weekend party, and he lost the original data. Please help him to recover the crucial components in the output for his assignment. The damaged output is

```
> round(c(mean(x),mean(y),var(x),var(y),cor(x,y)),2)
[1]    10.00    70.00     4.00  1850.00     0.93
> c(length(x),length(y))
[1] 5 5
> summary(lm(y~x))
Call:
lm(formula = y ~ x)
Coefficients:
             Estimate   Std. Error   t value   Pr(>|t|)
(Intercept)    __I__      46.368      __II__     _____
x             __III__    __IV__      _____      _____

Residual standard error: _____ on ____ degrees of freedom
Multiple R-squared:  __V_____,  Adjusted R-squared:  __VI____
F-statistic:  ___ on __and __ DF,  p-value: 0.02201

> anova(lm(y~x))
Analysis of Variance Table
Response: y
           Df    Sum Sq    Mean Sq    F value    Pr(>F)
x           __   __VII__   _______   __VIII__    0.02201 *
Residuals  __IX__  _____    _______

> round(residuals(lm(y~x)),2)
   1    2    3    4    5
-10   __X__  10  -20   20
```

2. A survey was conducted to study the number of motorcycles of a household. The data is listed as below

Number of Motorcycles (X)	0	1	2	≧3
Number of households	15	30	50	5

One would like to exam if the number of motorcycles of a household, denoted as X, follows the Poisson distribution, that is, he would like to exam the null hypothesis of

$H_o: X \sim Poi(\lambda)$

$H_a: H_o$ is not true.

Under H_o, the estimated value of λ is __XI__, and the expected count of households that owns 1 motorcycle is __XII__. The distribution he needs to evaluate the value of the test statistic is __XIII__.

(For the households having 3 or more than 3 motorcycles, please consider they have 4 motorcycles .)

C. Short Description/Calculation (20%)

1. Suppose that a fast-food chain wants to evaluate the service at three of their chain-restaurants. The customer service director for the chain hires four evaluators with varied experiences in food-service evaluations to act as raters. The four raters evaluate the service at each of the four restaurants in a random order. A rating scale from 0 (low) to 10 (high) is used. The following table summarizes the results:

Restaurant / Rater	A	B	C	Sample Mean	Sample Variance
1	6	8	6	6.67	1.33
2	4	9	7	6.67	6.33
3	4	7	5	5.33	2.33
4	6	8	6	6.67	1.33
Sample Mean	5	8	6	Overall Sample mean = 6.33	
Sample Variance	1.33	0.67	0.67	Overall Sample Variance = 2.42	

Please answer the following questions:

(1) (6%) What is the name of this design, and what is the advantage of this design?

(2) (14%) Please give the proper ANOVA table of this experiment, and describe how to evaluate if the ratings of these restaurants are the same.