※ 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。

# 1 True or False 2%×15 = 30%

*For each of the following statements, please determine if it is true or false, and answer "T" for true, "F" otherwise on the answer sheet.*

1. Sample mean is the unbiased estimator for the population mean for any population as long as the sampling design is simple random sampling without replacement.
2. If two events are disjoint, then they are independent.
3. If we have outliers in our data, then the range is a better measurement to describe the dispersion of the data.
4. Under a regression model $y = a + bx$, we mean that one unit change of $x$ would lead to the change of $b$ unit on $y$ if all other possible explanatory variables are fixed.
5. The usual F-test for testing if two population variances are equal is very sensitive to the assumption of normality, hence it should be used with caution.
6. The gender is a nominal variable.
7. Type-I error happens when we reject a true null hypothesis.
8. $t_k^{0.95} = -t_k^{0.05} = \sqrt{F_{1,10}^{0.9}} = \sqrt{(F_{10,1}^{0.1})^{-1}}$
9. Chebyshev's theorem can be applied even under a highly skewed population.
10. As log as the number of experiments $n$ is greater than about 30, then it is fine to apply the normal approximation for the binomial distribution.

※ 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。

11. If two events $A$ and $B$ are independent, then $A^c$ and $B^c$ are independent as well.

12. Boxplot is a proper graphical tool to describe qualitative data.

13. The power curve is used to describe the predictive power of a regression model.

14. The assumption of normality is not required when the method of least squares is used to estimate the parameters under a linear regression model.

15. The median tends to be greater than the mean when the population is right-skewed.

## 2 Multiple Choice $2\% \times 10 = 20\%$

1. If a factory would like to examine the time when one of their five independent production lines breaks down, which of the following distribution is necessary for this analysis? Assuming that the occurrence of breakdown follows a Poisson process.

   (A) Poisson distribution (B) Exponential distribution (C)Beta distribution (D) Binomial distribution

2. Cindy would like to study the weekly expenditure of a college student in her school, therefore she at first randomly choose 50 classes out of 380 classes, and then again randomly choose 15 students out of all selected classes to construct her sample of students. What is the name of the sampling design used here?

   (A) Stratifies random design (B) Cluster design (C) Systematic design (D) None of the above is true

※ 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。

3. Tony would like to proceed a one-way ANOVA to example if the products from three production lines are similar, and he collects 10 products randomly from each of the production line. What is the distribution he needs to decide if the test statistic is significant?
   (A) $t_{19}$ (B) $t_{27}$ (C) $F_{2,29}$ (D) $F_{2,27}$

4. Refer to question 3, Tony rejects the null hypothesis that the products are equal, and he would like to further proceed a pairwise comparison to determine why the hull hypothesis is rejected with the Fisher's LSD (Least squared distance) method, what is the distribution he needs to decide if the test statistic(s) is(are) significant?
   (A) $t_{19}$ (B) $t_{27}$ (C) $F_{2,29}$ (D) $F_{2,27}$

5. Refer to question 4, what is the overall Type-I error rate of the pairwise multiple comparisons if the $\alpha$ is set to be 0.05 for each comparison
   (A) 0.00125 (B) 0.15 (C) 0.143 (D) 0.184

6. $X$ and $Y$ are two independent standard normal random variables, and $U = X/Y$, what is the expected value of $U$?
   (A) undetermined (B) 1 (C) 0 (D) Infinity

7. The relationship between the Poisson distribution and Binomial distribution is similar to the relationship between Negative binomial distribution and
   (A) Geometric distribution (B) Beta distribution (C) Gamma distribution (D) Bernoulli distribution

8. Which of the followings does not affect the p-value?

※ 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) The value of the test statistic. (B) The alternative hypothesis. (C) The significant level $\alpha$ (D) None of the above is correct.

9. To examine the efficiency of three different production methods, five workers were selected in this study and each of them experiments all three methods for an hour (the order of methods experimented is randomly assigned). The response is the number of products produced, What distribution is required for the test statistics to examine if

| Worker | Method A | method B | Method C |
|---|---|---|---|
| 1 | 156 | 174 | 128 |
| 2 | 138 | 160 | 145 |
| 3 | 149 | 142 | 120 |
| 4 | 130 | 147 | 139 |
| 5 | 178 | 169 | 142 |

Table 1: Number of products

there is any difference among these three methods

(A) $F_{3,14}$ (B) $F_{2,8}$ (C) $F_{2,14}$ (D) $F_{4,8}$

10. Which of the following statistic does not change with the change of measurement?

(A) Variance (B) Median (C) Estimated slope of the regression model (D) z-score

## 3 Fill in the Blanks $4\% \times 10 = 40\%$

*Exact number is not required, for example, if your answer is* $e^2 \doteq 7.39$ *or* $\sqrt{253} \doteq 15.90$, *it*

※ 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。

*is good enough to give* $e^2$ *or* $\sqrt{253}$

1. To study the effect of declining birth rate, 100 families are selected and the number of children of a family is recorded, the data yields One would like to examine if the number

| Number of children | 0 | 1 | 2 | $\geq 3$ |
|---|---|---|---|---|
| Number of families | 35 | 35 | 25 | 5 |

Table 2: Number of children

of children in a family follows a Poisson distribution, the expected number of families without any children under $H_o$ : Number of children follows a Poisson distribution is __(A)__, and the distribution you need to examine the test statistic is __(B)__. The Pearson's residual of the class of families without any children is __(C)__

2. A study is conducted to investigate the relationship between $x$ and $y$ with the following linear regression model

$$y = a + b \cdot x + \varepsilon,\ \varepsilon \sim N(0, \sigma^2).$$

11 pairs of data $(x_i, y_i)$ are collected and the data yields the following sample statistics listed in Table 3: and the correlation between $x$ and $y$ is 0.6. Based on these sample

| | Mean | Variance |
|---|---|---|
| $x$ | 10 | 900 |
| $y$ | 20 | 2500 |

Table 3: Sample statistics

※ 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。

statistics, the estimated regression equation is

$$y = \hat{a} + \hat{b} \cdot x.$$

where $\hat{a} =$ (D) , $\hat{b} =$ (E) , and the estimated value of $\sigma^2$ is $\hat{\sigma}^2 =$ (F) .

To examine if the effect of $x$ on $y$ is significant, we need to test

$$H_o : \beta = 0$$

The value of the test statistics is (G) , and your conclusion is (H) . In addition, to construct the associated ANOVA table, the value of the total sum of square SSTOT= (I) , and the mean square due to treatment MSTR= (J) .

# 4 Problems 10%

1. (10%) Describe the similarity and/or difference between the significant level $\alpha$ and p-value of a hypothesis test.