國立清華大學 100 學年度碩士班入學考試試題

系所班組別：工業工程與工程管理學系(甲組、乙組、丙組)

考試科目（代碼）：統計學（1601, 1701, 1801）

　　*請在【答案卡】作答

一、(50%) 答案卡作答

說明: 以下為單選題, 每題 5 分, 答錯倒扣 3 分

1. Let Y_1 and Y_2 be two independent identically distributed random variables with nonzero variance. Which of the following statement is false:

   (a) $Y_1 - Y_2 \neq 0$

   (b) $E(Y_1) - E(Y_2) = 0$

   (c) $Var(Y_1) - Var(Y_2) = 0$

   (d) $Var(Y_1 - Y_2) = 0$

   (e) None of the above

2. Consider the distribution of annual salaries in a small firm. Suppose the mean salary is $14,300 and the standard deviation of the salaries is $1,200. We know from the Chebyshev inequality that the proportion of employees whose salaries fall outside the range of $12,500 to $16,100 is

   (a) at most 1/2.25

   (b) at most 1/2.0

   (c) at most 1/1.5

   (d) at least 1/2.0

   (e) none of the above

3. Suppose that X and Y are two random variables, which may be dependent, and that Var(X)=Var(Y). Assuming that $0<Var(X+Y)<\infty$, and $0<Var(X-Y)<\infty$. Which of the following statement is true:

   (a) E(XY)=E(X)E(Y)

   (b) E(X/Y)=E(X)/E(Y)

   (c) The random variable X+Y and X-Y are uncorrelated.

   (d) The random variable X+Y and X-Y are correlated.

   (e) None of the above

4. The average time that a programmer at an oil service company spends debugging programs is 2.7 hours. The time spent debugging programs is considered to be exponentially distributed. What is the probability that for any randomly selected program, the programmer will spend between 1.5 hours and 2.5 hours debugging the program?

   (a) $e^{-1/25\times1.5}$

   (b) $e^{-2.5/2.7} - e^{-1.5/2.7}$

   (c) $e^{-1.5/2.7} - e^{-2.5/2.7}$

   (d) $e^{-1.5\times2.7} - e^{-1.5\times2.7}$

   (e) $1\text{-}e^{-2.5/2.7}$

5. If a hypothesis test leads to the rejection of the null hypothesis

   (a) A Type I error is always committed

   (b) A Type II error is always committed

   (c) A Type I error may have been committed only

   (d) A Type II error may have been committed only

   (e) Both Type I and II errors may have been committed

6. A pair-difference test is a

   (a) simple example of a chi-square goodness-of-fit test

   (b) simple example of a randomized block design

   (c) simple example of a completely randomized experiment

   (d) simple example of a chi-square homogeneity test

   (e) none of the above

7. A random variable with a F distribution

   (a) can assume only non-negative values

   (b) is the ratio of two chi-square random variables divided by their degrees of freedom

   (c) has an infinite expected value

   (d) has both (a) and (b) true

   (e) has none of the above true

8. Let μ and σ^2 be the mean and variance of population of the random variable X, respectively. What is the variance of the random variable $(X-\mu)/\sigma$?

   (a) 1

   (b) μ/σ^2

   (c) 4

   (d) σ^2

   (e) 0

9. In a simple linear regression model $Y = \beta_0 + \beta_1 X + \varepsilon$, if a confidence interval for β_1 spans the value of 0, one can conclude that

   (a) MSE=0

   (b) $R^2 = 0$

   (c) There is no linear statistical relationship between X and Y

   (d) There is no causal effect of X on Y but there may be a causal effect of Y on X

   (e) The regression function passes through the origin

10. Suppose the true model is $Y = \beta_0 + \beta_1 X + \varepsilon_i$, under what conditions the OLS estimate of β is no longer unbiased.

    (a) $E(\varepsilon_i) \neq 0$

    (b) X is dependent of ε_i

    (c) $Var(\varepsilon_i) \neq \sigma^2$

    (d) ε_i is correlated with ε_j for $i \neq j$

    (e) All of the above

二、(50%) 答案卷作答

1. (10 pts.) Suppose that the joint pdf of X and Y is

$$f_{X,Y}(x,y) = \begin{cases} 1/y, & 0 < x < y < 1 \\ 0, & \text{otherwise.} \end{cases}$$

Prove or disprove that the above $f_{X,Y}(x,y)$ is a joint pdf of X and Y. (No credit is given if you only give the final result without any explanation.)

2. (15 pts.) For each part, name the relevant probability distribution of Y, and state any parameter values. Make any assumptions if needed.

   (a) $Y = \sum_{i=1}^{10} X_i$, where $X_1, \ldots, X_{10}$ are independent geometric random variables, each with parameter p.

   (b) $Y = \sum_{i=1}^{10} X_i$, where $X_1, \ldots, X_{10}$ are independent chi-squared variables, each with parameter $\nu = 1$.

   (c) $Y = \sum_{i=1}^{10} X_i$, where $X_1, \ldots, X_{10}$ are independent Exponential variables, each with expected value 1.

   (d) $Y = \sum_{i=1}^{n} (\frac{X_i - 10}{2})^2$, where $X_i \sim$ normal $(10, \sigma^2 = 4)$, $i = 1, 2, \ldots, n$ are independent.

   (e) $Y = X^2$, where X follows a student t distribution with parameter ν.

3. (25 pts.) Consider a one-factor factorial design for data Y_{ij} listed below:

Treatment	Observations	row average
1	$Y_{11},\ Y_{12},\ \ldots,\ Y_{1,n_1}$	$\overline{Y}_{1\cdot}$
2	$Y_{21},\ Y_{22},\ \ldots,\ Y_{2,n_2}$	$\overline{Y}_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$
a	$Y_{a1},\ Y_{a2},\ \ldots,\ Y_{a,n_a}$	$\overline{Y}_{a\cdot}$

where $\overline{Y}_{i\cdot} = \frac{\sum_{j=1}^{n_i} Y_{ij}}{n_i}$, $\overline{Y}_{\cdot\cdot} = \frac{\sum_{i=1}^{a}\sum_{j=1}^{n_i} Y_{ij}}{N}$, and $N = \sum_{i=1}^{a} n_i$

(a) (15 pts) Suppose $Y_{ij} \sim \mathrm{N}(\mu_i, \sigma^2)$ $i = 1, 2, \ldots, a; j = 1, 2, \ldots, n_i$. Also suppose that Y_{ij} and $Y_{i'j'}$ are statistically independent, where $i \neq i', j \neq j'$. Name the sampling distribution (including the value of parameters) in each part:

i. $\overline{Y}_{i\cdot}$

ii. $\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot}$

iii. $\sum_{i=1}^{a}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{\cdot\cdot})^2$

(b) (5 pts) Based on the table, Write SS_{total}, $SS_{\text{treatment}}$, and SS_{error} as functions of $Y_{ij} =, i = 1, 2, \ldots, a;\ j = 1, 2, \ldots, n_i$, where SS_{total}, $SS_{\text{treatment}}$, and SS_{error} are traditional notation used in ANOVA Table.

(c) (5 pts) Derive the relationship among SS_{total}, $SS_{\text{treatment}}$, and SS_{error}. (No credit is given if you only give the final result without any explanation.)