

東吳大學 115 學年度碩士班招生考試試題

第 1 頁，共 4 頁

系級	經濟學系碩士班 B 組	考試時間	100 分鐘
科目	統計學	本科總分	100 分

※一律作答於答案卷上(題上作答不予計分)；並務必標明題號，依序作答。

1. (8%) A company sends promotional emails to n distinct customers. For each customer $i (i = 1, 2, \dots, n)$, define the indicator random variable

$$I_i = \begin{cases} 1, & \text{if customer } i \text{ opens the email during the observation period,} \\ 0, & \text{otherwise.} \end{cases}$$

Assume that the random variables $\{I_i\}_{i=1}^n$ are independent and identically distributed with

$$P(I_i = 1) = p, \quad 0 < p < 1.$$

Let $X = \sum_{i=1}^n I_i$, denote the total number of emails open. denote the total number of emails opened during the observation period. Suppose that the outcome of the experiment is restricted to the event that the promotional campaign is successful in the sense that at least one customer opens the email. Under this condition, determine the expected number of emails opened, and express your answer in terms of n and p .

2. (10%) A customer service center receives complaints each day. Let the random variable X denote the number of complaints per day, with the following probability distribution:

x	0	1	2	3	4
$P(X = x)$	0.15	0.30	0.25	0.20	0.10

The daily operating cost (in dollars) is modeled as a quadratic function of the number of complaints:

$$C = g(X) = 50 + 20X + 5X^2.$$

If the number of complaints received in a day is at least three, the customer service center incurs an additional fixed penalty cost of \$40 for that day.

Let T denote the total daily operating cost, defined as the sum of the basic operating cost and any penalty cost incurred due to excessive complaints. Compute the $E(T)$ and $\text{Var}(T)$.

3. (12%) Let $X_1, X_2, \dots, X_n$ be a random sample from a population with mean μ and variance σ^2 . Consider the following three estimators of μ :

$$\hat{\theta}_1 = \bar{X}, \quad \hat{\theta}_2 = \frac{X_1 + X_2 + \dots + X_{n-1}}{n-1}, \quad \hat{\theta}_3 = \frac{X_1 + 2X_2 + \dots + nX_n}{\frac{n(n+1)}{2}}.$$

Answer the following questions:

- Determine whether each estimator is unbiased for μ .
- For all estimators that are unbiased, compare their variances and determine which estimator has the smallest variance. Identify the estimator that is therefore the most efficient among them.
- Among the unbiased estimators, determine which one is the minimum variance unbiased estimator (MVUE).

東吳大學 115 學年度碩士班招生考試試題

第 2 頁，共 4 頁

系 級	經濟學系碩士班 B 組	考試 時間	100 分鐘
科 目	統計學	本科 總分	100 分

4. (15%) A company conducts random drug testing among its employees. It is known that 2% of the employees have used an illegal drug. The drug test has the following characteristics:

Sensitivity is 95%: If an employee has used drugs, the probability that the test result is positive is 0.95.
Specificity is 90%: If an employee has not used drugs, the probability that the test result is negative is 0.90.

Equivalently, the probability of a false positive is 10%. Assume that all tests are administered correctly and that different test results are statistically independent unless stated otherwise.

(Give your answer rounded to four decimal places.)

- (a) If a randomly selected employee tests positive, what is the probability that the employee has actually used illegal drugs?

Briefly explain how this probability should be interpreted in light of the low base rate of drug use and the test's false positive rate.

- (b) Suppose an employee is tested twice, and the two tests are independent. If both test results are positive, what is the probability that the employee has actually used illegal drugs?

Briefly explain why this probability differs from the result in part (a).

- (c) Continuing from the previous question, suppose the employee is tested twice, the two tests are independent, and the results consist of one positive and one negative result (in either order). What is the posterior probability that the employee has used illegal drugs, $P(D|\text{one positive and one negative})$?

Briefly explain how the mixed test outcomes affect the posterior probability.

5. (5%) Let $\{(X_i, Y_i)\}_{i=1}^n$ be a set of n paired observations and let r_1 denote the Pearson correlation coefficient between X and Y . Define the transformed variables

$$U_i = aX_i + b, \quad V_i = cY_i + d,$$

where a, b, c, d are constants with $a \neq 0$ and $c \neq 0$. Let r_2 denote the Pearson correlation coefficient between U and V . Show that

$$r_2 = \frac{ac}{|a||c|} r_1.$$

系 級	經濟學系碩士班 B 組	考試 時間	100 分鐘
科 目	統計學	本科 總分	100 分

6. (8%) Consider the simple linear regression model

$$y_i = \beta_1 + \beta_2 x_i + u_i, i = 1, \dots, n, \quad (1)$$

and the regression model based on standardized variables

$$y_i^* = \alpha_1 + \alpha_2 x_i^* + v_i, i = 1, \dots, n, \quad (2)$$

where y_i^* and x_i^* are standardized versions of y_i and x_i , respectively.

Let $\hat{\beta}_1, \hat{\beta}_2$ and $\hat{\alpha}_1, \hat{\alpha}_2$ denote the ordinary least squares estimators from models (1) and (2).

(a) Derive the exact relationship between $\hat{\alpha}_2$ and $\hat{\beta}_2$.

(b) For model (1), prove that the sample mean of the fitted values equals the sample mean of the dependent variable; that is,

$$\frac{1}{n} \sum_{i=1}^n \hat{y}_i = \bar{y}.$$

7. (10%) Let the random variable follow a normal distribution $N(\mu, \sigma^2)$. It is given that :

$$P(X > 100) = 0.1587, \quad P(X < 80) = 0.0668.$$

(Hint: Standard normal distribution function values: $\Phi(1) = 0.8413$, $\Phi(1.5) = 0.9332$, $\Phi(0.25) = 0.5987$) (Give your answer rounded to four decimal places.)

(a) Find μ and σ .

(b) If $Y = 2X + 10$, find $P(190 < Y < 210)$.

8. (5%) In a certain region, the monthly household electricity consumption (in kWh) is normally distributed with mean $\mu = 350$ and standard deviation $\sigma = 80$. A random sample of size n is taken and let $\bar{X}$ denote the sample mean consumption. Determine the minimum sample size n such that

$$P(|\bar{X} - \mu| \leq 10) \geq 0.95.$$

9. (7%) Two production lines, A and B, manufacture the same type of battery. To compare their mean lifetimes, random samples are taken and the observed battery lifetimes (in hours) are as follows:

Production Line A: 49.1, 49.7, 46.8, 50.3, 52.3, 50.8, 51.6, 48.1

Production Line B: 46.7, 50.2, 53.7, 57.1, 48.6, 53.9, 53.5, 53.3, 51.9

Assume that the battery lifetimes from both production lines are normally distributed. At the significance level $\alpha = 0.1$, test whether there is a significant difference between the mean lifetimes of the two production lines. You may use the following critical values:

$$F_{0.05}(7,8) = 3.5005, \quad F_{0.95}(7,8) = 0.2684, \quad t_{0.025}(15) = 2.131, \quad t_{0.05}(15) = 1.753.$$

(Note: Not all the above critical values may be necessary.)

(Give your answer rounded to four decimal places.)

東吳大學 115 學年度碩士班招生考試試題

第 4 頁，共 4 頁

系級	經濟學系碩士班 B 組	考試時間	100 分鐘
科目	統計學	本科總分	100 分

10. (10%) The measurement error of a precision instrument is assumed to follow a normal distribution

$$N(\mu, \sigma^2).$$

An engineer claims that the population standard deviation of the instrument is $\sigma_0 = 2.5$. To assess this claim, a random sample of $n = 7$ measurements is collected:

$$2.1, -1.4, 3.8, 0.6, -2.7, 4.5, 1.9$$

- (a) Computing the sample mean $\bar{x}$ and the sample variance s^2 using the unbiased estimator (denominator $n - 1$).
- (b) Construct a 90% confidence interval for the population variance σ^2 .
- (c) At significance level $\alpha = 0.10$, test $H_0: \sigma^2 = 6.25$ versus $H_a: \sigma^2 \neq 6.25$.
Clearly state the rejection region and your conclusion using “reject H_0 ” or “do not reject H_0 ”.

For a chi-square distribution with $df = 6$, the right-tail critical values are:

$$\chi_{0.995}^2 = 0.486, \chi_{0.99}^2 = 0.6767, \chi_{0.975}^2 = 0.8721, \chi_{0.95}^2 = 1.635, \chi_{0.90}^2 = 2.204, \\ \chi_{0.10}^2 = 10.645, \chi_{0.05}^2 = 12.592, \chi_{0.025}^2 = 14.449, \chi_{0.01}^2 = 16.812$$

(Note: χ_{α}^2 denotes the value such that $P(X^2 > \chi_{\alpha}^2) = \alpha$. Not all the above critical values may be necessary.)

(Give your answer rounded to four decimal places.)

11. (10%) A researcher studies the linear relationship between an explanatory variable x and a response variable y . A random sample of 10 paired observations is collected as follows:

x	2	4	5	7	9	11	13	14	16	18
y	15	14	13	11	10	9	7	6	5	3

Assume the simple linear regression model

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad \varepsilon_i \sim iid \quad N(0, \sigma^2).$$

All numerical answers should be rounded to four decimal places.

- (a) Using the least squares method, find the fitted regression line.
- (b) Compute the sample correlation coefficient r between x and y . Briefly explain the result.
- (c) Compute the residual sum of squares, SSE.