國立政治大學
113學年度

# 統計學系碩士班甄試考試試題



| 考試科目 | 統計學 | 所(組)別 | 統計學系 | 考試時間 | 112年11月12日 星期六<br>10：00－11：40 |
|---|---|---|---|---|---|

注意：本科目有四份試題，第一份試題答案寫在第一卷答案本上，依此類推。每份試題配分為25分，合計共100分。

## 第一份試題(共25分，答案請寫在第一卷答案本上)

請寫出計算/證明過程，否則不予計分。

1. (10%) A box contains 6 yellow and 6 orange balls. We randomly choose 6 balls (and this is called "1 selection"). If 3 of them are yellow and 3 are orange, we stop. If not, we replace the balls in the box and again random select 6 balls. The experiment continues until exactly 3 of the 6 chosen are yellow. What is the probability that we shall make exactly $n$ selections?

2. (15%) Let $X$ be a continuous random variable that takes on values between $0$ and $c$, where $c > 0$. That is, $P(0 \le X \le c) = 1$. Show that

$$Var(X) \le \frac{c^2}{4}.$$

**Hint:** One of possible solutions is to first show that

$$E(X^2) \le c \cdot E(X) \text{ and } Var(X) \le c^2 \cdot (a - a^2), \text{ where } a = \frac{E(X)}{c},$$

and then complete the proof.

# 統計學系碩士班甄試考試試題



| 考試科目 | 統計學 | 所(組)別 | 統計學系 | 考試時間 | 112 年 11 月 12 日 星期六<br>10：00－11：40 |
|---|---|---|---|---|---|

第二份試題(共 25 分，答案請寫在第二卷答案本上)

The random variable $X$, representing the number of cherries in a cherry puff, has the following probability distribution:

| $x$ | 4 | 5 | 6 | 7 |
|---|---|---|---|---|
| $P(X=x)$ | 0.2 | 0.4 | 0.3 | 0.1 |

(a) Find the probability that the average number of cherries in 36 cheery puffs will be less than 5.5. **(12%)**

(b) If a random sample of 64 cheery puffs has an average of 5.5 cherries, then construct 99% confidence interval for the mean number of cherries in a puff. **(7%)**

(c) How large a sample is needed if we wish to be 95% confident that the sample mean will be within 0.2 cherry of the true mean? **(6%)**

# 統計學系碩士班甄試考試試題



| 考試科目 | 統計學 | 所(組)別 | 統計學系 | 考試時間 | 112年11月12日 星期六<br>10：00－11：40 |
|---|---|---|---|---|---|

第三份試題(共 25 分，答案請寫在第三卷答案本上)

1. Define or state the following terms (2pts for each; Chinese or English are acceptable)

(a) significant level
(b) power
(c) rejection region
(d) type I error

**The following (a) and (b) are NOT multiple-choice questions. You need to explain the reasons.**

2. Let the distribution of a random variable be given by

| outcome | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| Probability | $\theta$ | $2\theta$ | $0.9-2\theta$ | $0.1-\theta$ |

where $\theta$ is unknown and $0<\theta<0.1$. For testing $H_0:\theta=0.05$ vs. $H_1:\theta>0.05$ at level $\alpha=0.05$, a researcher is going to obtain an observation $x$ from this distribution.

(a) 3 methods of region rules are considered as follows. For each decision rule, find the probability of making type I error. (3pts)

I. Reject $H_0$ if $x=0$.
II. Reject $H_0$ if $x=1$ and a fair coin shows a head.
III. Reject $H_0$ if $x=3$.

(b) Determine which of the decision rules has the smallest power if $H_1$ is true. (3pts)

(c) If he is going to collect 100 observations and would like to test $H_0:\theta=0.05$ vs. $H_1:\theta\neq0.05$. Suppose that the counts of (0, 1, 2, 3) are $(y_0,y_1,y_2,y_3)$. Please show a nonparametric testing method. Provide the details such as the name of testing, the test statistic, the decision rule, etc. (5pts)

(d) Since the size of 100 is large enough, please show him how to apply the central limit theorem for testing $H_0:\mu=2$ vs. $H_1:\mu\neq2$, where $\mu$ is mean of the distribution. Please provide the details of the testing procedure. (6pts)

# 統計學系碩士班甄試考試試題



| 考試科目 | 統計學 | 所(組)別 | 統計學系 | 考試時間 | 112 年 11 月 12 日 星期六<br>10：00－11：40 |
|---|---|---|---|---|---|

## 第四份試題(共 25 分，答案請寫在第四卷答案本上)

1. (5 points) Suppose we can reject the null hypothesis that $\beta_2 \geq 0$ at a 5% significance level where $\beta_2$ is the slope coefficient from a bivariate regression. Which of the following is definitely true?

   (a) Our test statistic was negative.
   (b) We can reject the null hypothesis that $\beta_2 = 0$ at a 5% significance level.
   (c) We can reject the null hypothesis that $\beta_2 \geq 0$ at a 2.5% significance level.
   (d) We can reject the null hypothesis that $\beta_2 < 0$ at a 5% significance level.

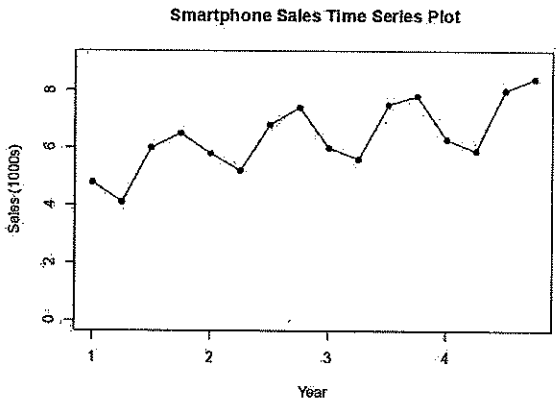
2. (5 points) If the two regression lines are as follows:

   $$Y = a + b\,X \quad \text{and} \quad X = c + d\,Y.$$

   What is the correlation coefficient between variables $X$ and $Y$ ? (If the current information provided doesn't allow for the calculation of the correlation coefficient, please respond with "The information is not sufficient".)

3. (5 points) The table and graph below display the smartphone sales of a specific manufacturer over the last four years. It's evident that there's a rising trend and a recurring seasonal pattern in the sales. When faced with both these factors, we should employ a forecasting method that can address both the trend and seasonality. This can be achieved by merging the dummy variable method for seasonality with the time series regression for linear trends. Please specify the general form of the multiple regression equation that models both the quarterly seasonal pattern and the linear trend in the smartphone sales data.

| Year | Quarter | Sales (1000s) |
|---|---|---|
| 1 | 1 | 4.8 |
| 1 | 2 | 4.1 |
| 1 | 3 | 6.0 |
| 1 | 4 | 6.5 |
| 2 | 1 | 5.8 |
| 2 | 2 | 5.2 |
| 2 | 3 | 6.8 |
| 2 | 4 | 7.4 |
| 3 | 1 | 6.0 |
| 3 | 2 | 5.6 |
| 3 | 3 | 7.5 |
| 3 | 4 | 7.8 |
| 4 | 1 | 6.3 |
| 4 | 2 | 5.9 |
| 4 | 3 | 8.0 |
| 4 | 4 | 8.4 |



# 統計學系碩士班甄試考試試題



| 考試科目 | 統計學 | 所(組)別 | 統計學系 | 考試時間 | 112年11月12日 星期六<br>10：00－11：40 |
|---|---|---|---|---|---|

4. (10 points) We conducted a survey of 46 students to examine the relationship between their primary computer (Mac or PC) and time spent on social media. Our goal was to use regression analysis to determine if one type of computer is associated with more social media usage. The response variable we focused on was Social (amount of time in minutes per week spent on social media). We would like to use "OS" as a predictor variable, which is a categorical (qualitative) variable taking values in the set {Mac, PC}. Suppose we want to explain Social (minutes per week) in terms of OS (PC or Mac) and Email (minutes per week). Below is the output from the regression analysis. (a) Interpret the three estimated regression coefficients in the context of the model. (b) Interpret the *p*-value for the Lack-of-Fit test (assume the level of significance is 0.05).

Analysis of Variance

| Source | DF | Adj SS | Adj MS | F-Value | P-Value |
|---|---|---|---|---|---|
| Regression | 2 | 390597 | 195299 | 2.47 | 0.096 |
| OS_PC | 1 | 293693 | 293693 | 3.72 | 0.060 |
| Email | 1 | 190702 | 190702 | 2.42 | 0.127 |
| Error | 43 | 3394150 | 78934 | | |
| Lack-of-Fit | 29 | 2762459 | 95257 | 2.11 | 0.071 |
| Pure Error | 14 | 631692 | 45121 | | |
| Total | 45 | 3784748 | | | |

Model Summary

| S | R-sq | R-sq(adj) | R-sq(pred) |
|---|---|---|---|
| 280.95i | 10.32% | 6.15% | 0.64% |

Coefficients

| Term | Coef | SE Coef | T-Value | P-Value | VIF |
|---|---|---|---|---|---|
| Constant | 249.0 | 63.6 | 3.92 | 0.000 | |
| OS_PC | -165.7 | 85.9 | -1.93 | 0.060 | 1.07 |
| Email | 0.729 | 0.469 | 1.55 | 0.127 | 1.07 |

Regression Equation

Social = 249.0 - 165.7 OS_PC + 0.729 Email