

科目：統計學

系所組：統計資訊學系應用統計碩士班

- (15%) 學生完成統計測驗所花的時間是介於 30 至 60 分鐘之間的均勻分配。隨機選取一位學生，求下列事件的機率。(1)學生需要多於 55 分鐘的時間以完成測驗；(2)學生完成測驗的時間介於 30 至 40 分鐘之間；(3)學生完成測驗的時間剛好是 37.23 分鐘。
- (15%) (1)某銀行九點開始營業，每小時平均來 3 人，符合 Poisson 分配，試求某日早上九點到九點半之間沒有客戶上門的機率？(2)若此銀行平均 20 分鐘來 1 人，符合指數分配，試求某日營業後等候 30 分鐘以上才有客戶來的機率？(3)試說明 Poisson 分配與指數分配兩者間的關係？
- (30%)假設 A、B 兩家公司雇用新員工，結果如下表。顯著水準為 0.05。

公司	應徵人數	錄取人數
A	350	64
B	200	22

- (1)請計算 A、B 兩家公司的錄取率的區間估計；(2)以檢定方式比較兩公司之錄取率是否相等；(3)人數與公司是否獨立？( $\chi_1^2 = 3.84$ 、 $\chi_2^2 = 5.99$ 、 $\chi_3^2 = 7.82$ 、 $\chi_4^2 = 9.49$ )
- (40%) 假設台灣北部地區通勤通學民眾的時間平均為 1 小時。洪小姐住新莊，從新莊開車到台北市上班有兩條路：一條經三重到台北（縱貫路），一條由泰山五股走中山高到台北。她記錄 2 星期的行車時間如下(分鐘):  $\sum_{i=1}^5 (x_i - \bar{x})^2 = 34$ ， $\sum_{i=1}^5 (y_i - \bar{y})^2 = 600$ ， $\sum_{i=1}^5 (x_i - \bar{x})(y_i - \bar{y}) = 121$

縱貫路(x)	37	34	39	38	42
中山高(y)	44	23	37	31	55

- (1) 請利用敘述統計方法說明比較走兩條路線差異？
- (2) 在所有(十筆)資料中，請利用柴比雪夫法則與經驗法則指出有多少機率和區間出現在 2 倍標準差的之間？並說明兩方法中需滿足的使用條件。
- (3) 若線性模型為  $y_i = \alpha + \beta x_i + \varepsilon_i$ ，其中  $\varepsilon_i, i = 1, 2, \dots, n$  為相互獨立並服從平均為 0 和變異數為  $\sigma^2$  的常態分佈。請找出參數  $\alpha, \beta$  的估計值？
- (4) 接續(3)假設顯著水準為 0.05， $SSE=169.382$ ，請檢定  $H_0: \beta=1$  是否成立？( $t_{3,0.05} = 2.353, t_{3,0.025} = 3.182, t_{4,0.05} = 2.132, t_{4,0.025} = 2.776, t_{5,0.05} = 2.015, t_{5,0.025} = 2.571$ )(計算到小數第三位)

※ 注意：1.考生須在「彌封答案卷」上作答。

2.本試題紙空白部份可當稿紙使用。

3.考生於作答時可否使用計算機、法典、字典或其他資料或工具，以簡章之規定為準。

科目：計算機概論

系所組：統計資訊學系應用統計碩士班

名詞解釋：

- 1.Primary key (5%)
- 2.Bootstrappings (5%)
- 3.Software as a service (SaaS) (5%)
- 4.Transitive dependency (5%)
- 5.Online to offline (O2O) (5%)
- 6.Data validation (5%)
- 7.Non-repudiation (5%)
- 8.Web crawler (5%)

問答題：

- 1.若想為自己的線上商店架設一個網站，就內容建置與網頁寄存兩方面來說，有哪些方式可以達到目的？試分別說明這些方法的優缺點。(10%)
- 2.試說明資料分類與資料分群的差異，並各舉一個演算法作為例子。(10%)
- 3.已知有一數列{10, 11, 12, 13, 14, 15, 16, 17, 18, 19}。試說明分別以線性搜尋法 (linear search) 與內插搜尋法 (interpolation search) 搜尋 15 的過程，以及各需做幾次比較？(10%)
- 4.請將物件導向程式設計中的繼承(inheritance)與多型(polymorphism)的觀念各寫一小段程式說明其特徵。(10%)
- 5.請舉出兩種你曾使用過的資料分析軟體/工具，並說明其特色。(10%)
- 6.關連式資料庫的完整性規則 (integrity rules) 是什麼？試舉例並說明之。(10%)

※注意：1.考生須在「彌封答案卷」上作答。

2.本試題紙空白部分可當稿紙使用，試題須隨答案卷繳回。

3.考生於作答時可否使用計算機、法典、字典或其他資料或工具，以簡章之規定為準。

科目：微積分

系所組：統計資訊學系應用統計碩士班

令函數  $f(x) = \frac{1}{\sqrt{8\pi}} e^{-\frac{(x-1)^2}{8}}$ ,  $-\infty < x < \infty$ , 為具有平均數為 1 及標準差為 2 之常態分配的機率密度函數

1. 判斷函數  $f(x)$  之增減性(monotonicity)。(10%)
2. 判斷函數  $f(x)$  之凹向性(concavity)。(10%)
3. 求出函數  $f(x)$  之反曲點(inflection points)。(10%)
4. 求出函數  $f(x)$  之相對極大值(relative maximum)和相對極小值(relative minimum)。(10%)
5. 求出函數  $f(x)$  在區間  $[-1, 5]$  之絕對極大值(absolute maximum)和絕對極小值(absolute minimum)。(10%)
6. 求出極限  $\lim_{t \rightarrow 0} \frac{\int_t^2 f(x) dx}{t}$ 。(10%)
7. 令函數  $h(x) = x^{f(x)}$ , 求導數  $h'(1)$ 。(10%)
8. 令函數  $k(x) = \ln(f(x))$ , 求導函數  $k'(x)$ 。(10%)
9. 求出積分  $\int_{-\infty}^{\infty} x \cdot f(x) dx$ 。(須寫計算過程,否則不計分)(10%)
10. 求出積分  $\int_{-\infty}^{\infty} e^{tx} \cdot f(x) dx$ 。(須寫計算過程,否則不計分)(10%)

※注意：1. 考生須在「彌封答案卷」上作答。

2. 本試題紙空白部分可當稿紙使用，試題須隨答案卷繳回。

3. 考生於作答時可否使用計算機、法典、字典或其他資料或工具，以簡章之規定為準。

科目：專業英語

系所組：統計資訊學系應用統計碩士班

**I. Chinese-English Translation (中翻英)**

1. 平均數的計算為觀測值的加總，再除以觀測值的個數。(15%)
2. 假設大學生每晚平均睡眠為 7.1 小時，且標準差為 35 分鐘。如果睡眠時間為常態分配，則大學生睡眠低於 7 小時的比例為何？(15%)

**II. English-Chinese Translation (英翻中)**

(A).

1. MS EXCEL provides statistical functions to compute descriptive statistics for a data set. These functions can be used to compute one statistic at a time (e.g., mean, variance, etc.). (15%)
2. Correlation coefficient is a measure of linear association between two variables that takes on values between  $-1$  and  $+1$ . Values near  $+1$  indicate a strong positive linear relationship; values near  $-1$  indicate a strong negative linear relationship; and values near zero indicate the lack of a linear relationship. (15%)
3. Standard deviation is a measure of variability computed by taking the positive square root of the variance. (15%)

(B). Choose one of the following questions to answer. (任選一題 [註：4(A) 或 4(B)] 作答) (25%)

4(A). Journal article 1. Data learning from big data (Source: *Statistics & Probability Letters*, 2018, 136, pp. 15-19.)

(英翻中)

Technology is generating a huge and growing availability of observations of diverse nature. This big data is placing data learning as a central scientific discipline. It includes collection, storage, preprocessing, visualization and, essentially, statistical analysis of enormous batches of data. In this paper, we discuss the role of statistics regarding some of the issues raised by big data in this new paradigm and also propose the name of data learning to describe all the activities that allow to obtain relevant knowledge from this new source of information.

4(B). Journal article 2. Forecasting daily Covid-19 cases in the world with a hybrid ARIMA and neural network model (Source: *Applied Soft Computing*, 2022, 126, 109315.)

(英翻中)

The use of models to predict disease cases is common in epidemiology and related areas, in the context of Covid-19, both ARIMA and Neural Network models can be applied for purposes of optimized resource management, so the aim of this study is to capture the linear and non-linear structures of daily Covid-19 cases in the world by using a hybrid forecasting model. In summary, the proposed hybrid system methodology consists of two steps. In the first step, an ARIMA model is used to analyze the linear part of the problem. In the second step, a neural network model is developed to model the residuals of the ARIMA model, which would be the non-linear part of it.

※注意：1. 考生須在「彌封答案卷」上作答。

2. 本試題紙空白部分可當稿紙使用，試題須隨答案卷繳回。

3. 考生於作答時可否使用計算機、法典、字典或其他資料或工具，以簡章之規定為準。