(本頁中文版如下一頁)

National Chengchi University, 112-2 Academic Year
Final Exam of Statistics (II), Bonus Test, R Programming

Department/Grade:_____ ID:_____ Name:_____

Subject: Statistics (II)

Date: 2024/06/20

This test consists of 5 major questions. (20% each, total score: 100%)

Time period: 15:00~16:00 (total 60 minutes)

_____

Notes:

1. Download the R exam sheet (`112-2-Stat-R-Final.zip`) from the course website and unzip in your laptop. The zip file contain the question sheet, the answer sheet, and the datasets.

2. Answers for this exam should be provided using the R programming language (either Rgui or RStudio). Other programming languages are not permitted.

3. During the exam, you may refer to textbooks, lecture notes (including videos, Please bring your own headphones), or browse the internet. However, the use of communication software/APP such as Messenger, IG, Line, etc., is strictly prohibited.

4. Any form of cheating or suspicious behavior is not allowed.

5. On this answer sheet, please ensure you copy the "executed code and its results (including graphics)" from the **R Console** and paste it here (in `Courier New font`, size 10, black text on a white background). This should include both the code and the output, not just one or the other. Finally, the answers for each sub-question should be highlight by yellow color (not just printing the report; the TA shouldn't have to search for the answers)

6. Please label your answers in sequence, e.g., (1)a, (1)b, (2)a, etc.

7. After completing your answers, save this Word document with the filename "`StudentID-FamilyName-Midterm.docx`" (replace with your actual "`Student ID and FamilyName`") and upload it to https://hmwu.idv.tw/ 【作業考試上傳區】。

8. Username: `stat112`, Password: (classroom number) `26xxxx`, Folder: "`20240620-FinalExam`".

9. If the upload site displays a "blank page", move your cursor to the "address bar" and press "Enter". If that doesn't work, try using a different browser (IE/Edge/Firefox/Chrome).

10. Uploaded files cannot be deleted. If you need to upload a revised file, please add "-2" to the main filename, e.g., "`StudentID-FamilyName-Midterm-2.docx`".

_____Wishing you a successful exam_____

國立政治大學 112 學年度第二學期
統計學(二) 期末 R 程式加分考

系級:_____ 學號:_____ 姓名:_____

考試科目: 統計學(二)

考試日期:2024/06/20

本試題共 5 大題 (各 20%)

考試時間:15:00~16:00 (共 60 分鐘)

_____

注意事項:

1. 從教學網站下載電子考卷 (112-2-Stat-R-Final.zip)，並於自己的筆電解壓縮。 壓縮檔包含題目卷、答案卷和資料集。

2. 本次考題以 R 程式(Rgui 或 RStudio)方式作答，其他程式不允許。

3. 考試過程中可查詢書本、教學講義或上網(含上課影片，請自備耳機)，禁止利用 messenger, IG, Line 等等通訊軟體。

4. 禁止疑似作弊行為。

5. 本答案卷上請務必於 R Console 內複製「執行後的程式碼及結果(含圖形)」，於本答案卷貼上(Courier New, 10 點字，白底黑字)，不是只有程式碼，不是只有報表。最後，將每小題之答案以黃色底高亮起來(不能只印出報表，要助教去找答案)。

6. 請依序註明題號: (1)a, (1)b, (2)a 等等。

11. 作答完請將此 word 檔存檔，檔名為「StudentID-FamilyName-Midterm.docx」(更改成自己「學號」、「姓」)並上傳至 https://hmwu.idv.tw/ 【作業考試上傳區】。

7. 帳號: stat112，密碼: (上課教室號碼) 26xxxx，資料夾:「20240620-MidtermExam」

8. 如果上傳網站出現「空白頁」，請將滑鼠移至「網址列」後，按「Enter」即可。若再不行，請換其它瀏覽器(IE/Edge/Firefox/Chrome)

9. 上傳檔案無法刪除，若要上傳更新檔，請於主檔名後加「-2」，例如:「StudentID-FamilyName-Midterm-2.docx」。

_____祝考試順利_____

(1)	**America's Favorite Sports.** *The Harris Poll* tracks the favorite sport of Americans who follow at least one sport. Results of the poll show that professional football is the favorite sport of 33% of Americans who follow at least one sport, followed by baseball at 15%, men's college football at 10%, auto racing at 6%, men's professional basketball at 5%, and ice hockey at 5%, with other sport at 26%. Consider a survey in which 344 college undergraduates who follow at least one sport were asked to identify their favorite sport produced the following results:

Professional Football	Baseball	Men's College Football	Auto Racing	Men's Professional Basketball	Ice Hockey	Other Sports
111	39	46	14	6	20	108

Do college undergraduate students differ from the general public with regard to their favorite sports? Use $\alpha = .05$.

(請直接印出 p 值或從報表中擷取出 p 值，而不是只有印出報表。)
(Please directly print the p-value or extract it from the report, rather than just printing the report.)

(2) **Rental Car Revenue and Fleet Size.** The following data were used to investigate the relationship between the number of cars in service (1000s) and the annual revenue ($millions) for six smaller car rental companies (*Auto Rental News* website).

Company	Cars (1000s)	Revenue ($ millions)
U-Save Auto Rental System, Inc.	11.5	118
Payless Car Rental System, Inc.	10.0	135
ACE Rent A Car	9.0	100
Rent-A-Wreck of America	5.5	37
Triangle Rent-A-Car	4.2	40
Affordable/Sensible	3.3	32

With x = cars in service (1000s) and y = annual revenue ($ millions),

a. Use the data to develop an estimated regression equation

b. Compute the coefficient of determination r^2.

c. Did the estimated regression equation provide a good fit? Explain.

d. What is the value of the sample correlation coefficient?

e. Predict the annual revenue ($ millions) that the number of cars in service (1000s) = 6

(3)	**Data file: RentMortgage**
	Buy Versus Rent. Occasionally, it has been the case that home prices and mortgage rates dropped so low that in a number of cities the monthly cost of owning a home was less expensive than renting. The following data show the average asking rent for 10 markets and the monthly mortgage on the median priced home (including taxes and insurance) for 10 cities where the average monthly mortgage payment was less than the average asking rent (*The Wall Street Journal*). a. Develop the estimated regression equation that can be used to predict the monthly mortgage given the average asking rent. b. Construct a residual plot against the independent variable. c. Do the assumptions about the error term and model form seem reasonable in light of the residual plot?

(4)	**Data file: Checkout**
	Supermarket Checkout Lines. Retail chain Kroger has more than 2700 locations and is the largest supermarket in the United States based on revenue. Kroger has invested heavily in data, technology, and analytics. Feeding predictive models with data from an infrared sensor system called QueVision to anticipate when shoppers will reach the checkout counters, Kroger is able to alert workers to open more checkout lines as needed. This has allowed Kroger to lower its average checkout time from four minutes to less than 30 seconds (*Retail Touchpoints*).
	Consider the data in the file *Checkout*. The file contains 32 observations. Each observation gives the arrival time (measured in minutes before 6 P.M.) and the shopping time (measured in minutes).
	a. Develop a scatter diagram for arrival time as the independent variable.
	b. What does the scatter diagram developed in part (a) indicate about the relationship between the two variables? Do there appear to be any outliers or influential observations? Explain.
	c. Using the entire data set, develop the estimated regression equation that can be used to predict the shopping time given the arrival time.
	d. Use residual analysis to determine whether any outliers or influential observations are present.

| (5) | **Data file: NetflixSubscribers**

Netflix Subscribers. The following data show the number of Netflix subscribers worldwide for the years 2012 (period 1) to 2018 (period 7) (datawrapper website). The data are in the file *NetflixSubscribers*.

  a. Construct a time-series plot. What type of pattern exists in the data?
  b. Develop a linear trend equation for this time series.
  c. Develop a quadratic trend equation for this time series.

  d. Compare the MSE for each model. Which model appears better according to MSE?
  e. Use the models in part (b) and (c) to forecast subscribers for 2018.
  f. Which of the two forecasts in part e would you use? Explain. |