

國立政治大學 114 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:2026/06/09

本試題共 6 大題 (共 120%)

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 (**114-2-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，禁止利用 messenger, IG, Line 等等通訊軟體。
4. **禁止使用 AI 模式搜尋。禁止使用 chatGPT 或類似的 AI 平台/工具。** 禁止疑似作弊行為。
1. 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)**。
5. 請依序註明題號: (1)a, (1)b, (2)a 等等。
6. 作答完請將此 word 檔存檔，檔名為「學號-姓名-Stat-R-Final.docx」(更改成自己「學號、姓名」)並上傳至教學網站【作業考試上傳區】或 <http://hmwu.nccu.edu.tw/login.html>
7. 帳號: stat114，密碼: 上課教室號碼，資料夾: 「20260609-Final」
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:「學號-姓名-Stat-R-Final-2.docx」。

祝考試順利

(1) (5 分)	<p>用 R 印出下列字句(姓名改為自己的姓名):</p> <p>"本人(學號)(姓名)恪遵各項考試規則，若如違反，願受校方最嚴厲處罰，謹誓。"</p>
----------------------------	--

(2)
(20 分)

Consider the following data for two variables, x and y .

x_i	135	110	130	145	175	160	120
y_i	145	100	120	120	130	130	110

a. Compute the standardized residuals for these data. Do the data include any outliers? Explain.

b. Plot the standardized residuals against \hat{y} . Does this plot reveal any outliers?

c. Develop a scatter diagram for these data. Does the scatter diagram indicate any outliers in the data?

(3)
(25 分)

Data file: Stroke.xlsx

Risk of a Stroke. A 10-year study conducted by the American Heart Association provided data on how age, blood pressure, and smoking relate to the risk of strokes. Assume that the following data are from a portion of this study. Risk is interpreted as the probability (times 100) that the patient will have a stroke over the next 10-year period. For the smoking variable, define a dummy variable with 1 indicating a smoker and 0 indicating a nonsmoker.

Risk	Age	Pressure	Smoker
12	57	152	No
24	67	163	No
13	58	155	No
56	86	177	Yes
28	59	196	No
51	76	189	Yes
18	56	155	Yes
31	78	120	No
37	80	135	Yes
15	78	98	No
Risk	Age	Pressure	Smoker
22	71	152	No
36	70	173	Yes
15	67	135	Yes
48	77	209	Yes
15	60	199	No
36	82	119	Yes
8	66	166	No
34	80	125	Yes
3	62	117	No
37	59	207	Yes

- Develop an estimated regression equation that relates risk of a stroke to the person's age, blood pressure, and whether the person is a smoker.
- Is smoking a significant factor in the risk of a stroke? Explain. Use $\alpha = .05$.
- What is the probability of a stroke over the next 10 years for Art Speen, a 68-year-old smoker who has blood pressure of 175? What action might the physician recommend for this patient?

(4)
(25 分)

Data file: Lakeland.xlsx

College Retention. Over the past few years the percentage of students who leave Lakeland College at the end of the first year has increased. Last year Lakeland started a voluntary one-week orientation program to help first-year students adjust to campus life. If Lakeland is able to show that the orientation program has a positive effect on retention, they will consider making the program a requirement for all first-year students. Lakeland's administration also suspects that students with lower GPAs have a higher probability of leaving Lakeland at the end of the first year. In order to investigate the relation of these variables to retention, Lakeland selected a random sample of 100 students from last year's entering class. The data are contained in the data set named Lakeland; a portion of the data follows.

Student	GPA	Program	Return
1	3.78	1	1
2	2.38	0	1
3	1.30	0	0
4	2.19	1	0
5	3.22	1	1
6	2.68	1	1
.	.	.	.
.	.	.	.
.	.	.	.
98	2.57	1	1
99	1.70	1	1
100	3.85	1	1

The dependent variable was coded as $y = 1$ if the student returned to Lakeland for the sophomore year and $y = 0$ if not. The two independent variables are:

$$x_1 = \text{GPA at the end of the first semester}$$
$$x_2 = \begin{cases} 0 & \text{if the student did not attend the orientation program} \\ 1 & \text{if the student attended the orientation program} \end{cases}$$

- Write the logistic regression equation relating x_1 and x_2 to y .
- Use both independent variables and statistical software to compute the estimated logit.
- Use $\alpha = .05$ to determine whether each of the independent variables is significant.
- Use the estimated logit computed in part (c) to estimate the probability that students with a 2.5 grade point average who did not attend the orientation program will return to Lakeland for their sophomore year. What is the estimated probability for students with a 2.5 grade point average who attended the orientation program?
- What is the estimated odds ratio for the orientation program?

(5)
(20 分)

Data file: NetflixSubscribers.xlsx

Netflix Subscribers. The following data show the number of Netflix subscribers worldwide for the years 2012 (period 1) to 2017 (period 6) (datawrapper website). The data are in the file *NetflixSubscribers*.

Year	Period	Subscribers (Millions)
2012	1	33.27
2013	2	44.35
2014	3	57.39
2015	4	74.76
2016	5	93.80
2017	6	117.58

- Construct a time-series plot. What type of pattern exists in the data?
- Develop a linear trend equation for this time series.
- Develop a quadratic trend equation for this time series.
- Compare the MSE for each model. Which model appears better according to MSE?
- Use the models in part (b) and (c) to forecast subscribers for 2018.

(6)
(20 分)

Data file: AptExp.xlsx

Lawn Maintenance. Three years of monthly lawn-maintenance expenses (\$) for a six-unit apartment house in southern Florida follow.

Month	Year 1	Year 2	Year 3
January	170	180	195
February	180	205	210
March	205	215	230
April	230	245	280
May	240	265	290
June	315	330	390
July	360	400	420
August	290	335	330
September	240	260	290
October	240	270	295
November	230	255	280
December	195	220	250

- Construct a time series plot. What type of pattern exists in the data?
- Develop an estimated regression equation that can be used to account for any seasonal and linear trend effects in the data. Use the following dummy variables to account for the seasonal effects in the data: Jan = 1 if January, 0 otherwise; Feb = 1 if February, 0 otherwise; Mar = 1 if March, 0 otherwise; . . . Nov = 1 if November, 0 otherwise. Note that using this coding method, when all the 11 dummy variables are 0, the observation corresponds to an expense in December.
- Compute the monthly forecasts for next year based upon both trend and seasonal effects.