(本頁中文版如下一頁)

National Chengchi University, 113-2 Academic Year
Midterm Exam of Statistics (II), Bonus Test, R Programming

Department/Grade:_______________ ID:________________ Name:______________

Subject: Statistics (I)

Date: 2025/04/24

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 (`113-2-Stat-R-Midterm.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 http://hmwu.nccu.edu.tw/login.html .
8. Username: `stat113`, Password: (classroom number) `26xxxx`, Folder: "`20250424-MidtermExam`".
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_______________________________

(English version on the previous page)

國立政治大學 113 學年度第二學期

統計學(二) 期中 R 程式加分考

系級:______________ 學號:_______________ 姓名:_____________

考試科目: 統計學(一)

考試日期:2025/04/24

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 (`113-2-Stat-R-Midterm.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**複制「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。
7. 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 http://hmwu.nccu.edu.tw/login.html
8. 帳號: `stat113`，密碼: (上課教室號碼) `26xxxx`，資料夾: 「`20230424-`

MidtermExam」

9. 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)
10. 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)	**Data file: ResidentialWater** **Cost of Residential Water.** On its municipal website, the city of Tulsa states that the rate it charges per 5 CCF of residential water is \$21.62. How do the residential water rates of other U.S. public utilities compare to Tulsa's rate? The file *ResidentialWater* contains the rate per 5 CCF of residential water for 42 randomly selected U.S. cities. a. Formulate hypotheses that can be used to determine whether the population mean rate per 5 CCF of residential water charged by U.S. public utilities differs from the \$21.62 rate charged by Tulsa. b. What is the *p*-value for your hypothesis test in part (a)? c. At $\alpha = .05$, can your null hypothesis be rejected? What is your conclusion? d. Repeat the preceding hypothesis test using the critical value approach.

```
# Q.1
> # Load the data
> library(readxl)
>
>
> #Load data
> data <- read_excel("C:\\Users\\Dnyanaish\\OneDrive\\Attachments\\Desktop\\ResidentialWater.xlsx")
>
> # Now view the data
> View(data)
> # Check columns
> colnames(data)
[1] "Public Utility Company" "Rate (5 CCF)"
>
> # Suppose the column is "Rate per 5 CCF", then:
> rates <- data$`Rate (5 CCF)`
```

```
>
>
> # Part (a) - Hypotheses are defined above.
> #Null: mu = 21.62 and alternative: mu ≠ 21.62
> # Part (b) - Perform t-test and get p-value
> t_test_result <- t.test(rates, mu = 21.62, alternative = "two.sided")
> p_value <- t_test_result$p.value
> cat("P-value:", p_value, "\n")
P-value: 0.2575782
>
> # Part (c) - Decision at alpha = 0.05
> alpha <- 0.05
> if (p_value < alpha) {
+   cat("Reject the null hypothesis at α = 0.05.\nConclusion: The mean rate is sign
ificantly different from $21.62.\n")
+ } else {
+   cat("Fail to reject the null hypothesis at α = 0.05.\nConclusion: No significan
t difference from $21.62.\n")
+ }
Fail to reject the null hypothesis at a = 0.05.
Conclusion: No significant difference from $21.62.
>
> # Part (d) - Critical value approach
> n <- length(rates)
> df <- n - 1
> sample_mean <- mean(rates)
> sample_sd <- sd(rates)
> se <- sample_sd / sqrt(n)
>
> # Compute t-statistic
> t_stat <- (sample_mean - 21.62) / se
> # Critical t-value for two-tailed test
> t_critical <- qt(1 - alpha/2, df)
>
> cat("T-statistic:", t_stat, "\n")
T-statistic: -1.14811
> cat("Critical t-value:", t_critical, "\n")
Critical t-value: 2.019541
>
> if (abs(t_stat) > t_critical) {
+   cat("Reject the null hypothesis using the critical value approach.\n")
+ } else {
+   cat("Fail to reject the null hypothesis using the critical value approach.\n")
+ }
Fail to reject the null hypothesis using the critical value approach.
```

(2) **Data file: BusinessTravel**

Domestic Airfare. The Global Business Travel Association reported the domestic airfare for business travel for the current year and the previous year. Below is a sample of 12 flights with their domestic airfares shown for both years.

Current Year	Previous Year	Current Year	Previous Year
345	315	635	585
526	463	710	650
420	462	605	545
216	206	517	547
285	275	570	508
405	432	610	580

a. Formulate the hypotheses and test for a significant increase in the mean domestic airfare for business travel for the one-year period. What is the *p*-value? Using a .05 level of significance, what is your conclusion?
b. What is the sample mean domestic airfare for business travel for each year?
c. What is the percentage change in the airfare for the one-year period?

```
# Q.2
> #Load data
> data <- read_excel("C:\\Users\\Dnyanaish\\OneDrive\\Attachments\\Desktop\\Busines
sTravel.xlsx")
>
> # Part (a) - Hypothesis test
> # H0: mean difference <= 0 (no increase)
> # H1: mean difference > 0 (there is an increase)
>
> # Paired t-test
> t_test_result <- t.test(current_year, previous_year, alternative = "greater", pai
red = TRUE)
>
> # Display the test results
> print(t_test_result)

        Paired t-test

data:  current_year and previous_year
t = 2.0536, df = 11, p-value = 0.03229
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 2.886139      Inf
sample estimates:
mean of the differences 
                     23 

>
```

```
> # Part (b) - Sample means for each year
> mean_current <- mean(current_year)
> mean_previous <- mean(previous_year)
>
> cat("Mean Current Year Airfare:", mean_current, "\n")
Mean Current Year Airfare: 487
> cat("Mean Previous Year Airfare:", mean_previous, "\n")
Mean Previous Year Airfare: 464
>
> # Part (c) - Percentage change
> percentage_change <- ((mean_current - mean_previous) / mean_previous) * 100
> cat("Percentage Change in Airfare:", round(percentage_change, 2), "%\n")
Percentage Change in Airfare: 4.96 %
```

(3)	**Data file: Costco** **Costco Customer Satisfaction.** *Consumer Reports* uses a 100-point customer satisfaction score to rate the nation's major chain stores. Assume that from past experience with the satisfaction rating score, a population standard deviation of $\sigma = 12$ is expected. In 2012, Costco, with its 432 warehouses in 40 states, was the only chain store to earn an outstanding rating for overall quality. A sample of 15 Costco customer satisfaction scores follows. 95 90 83 75 95 98 80 83 82 93 86 80 94 64 62 a. What is the sample mean customer satisfaction score for Costco? b. What is the sample variance? c. What is the sample standard deviation? d. Construct a hypothesis test to determine whether the population standard deviation of $\sigma = 12$ should be rejected for Costco. With a .05 level of significance, what is your conclusion?

```
#Q.3
> # Create the data
> scores <- read_excel("C:\\Users\\Dnyanaish\\OneDrive\\Attachments\\Desktop\\Costco.xlsx"
)
> # Now view the data
> View(scores)
> # Check columns
> colnames(scores)
[1] "Satisfaction Score"
>
> # Suppose the column is "Rate per 5 CCF", then:
> scores <- scores$`Satisfaction Score`
>
>
> # Part (a) - Sample mean
> mean_score <- mean(scores)
> cat("Sample Mean Customer Satisfaction Score:", mean_score, "\n")
Sample Mean Customer Satisfaction Score: 84
>
> # Part (b) - Sample variance
> sample_variance <- var(scores)
> cat("Sample Variance:", sample_variance, "\n")
Sample Variance: 118.7143
>
> # Part (c) - Sample standard deviation
> sample_sd <- sd(scores)
> cat("Sample Standard Deviation:", sample_sd, "\n")
Sample Standard Deviation: 10.89561
>
```

```
> # Part (d) - Hypothesis test for standard deviation
> # H0: sigma = 12
> # H1: sigma != 12
>
> # Chi-square test for variance
> n <- length(scores)
> sigma0 <- 12 # Given population standard deviation under H0
> chi_sq_stat <- (n - 1) * sample_variance / sigma0^2
>
> # Degrees of freedom
> df <- n - 1
>
> # p-value for two-tailed test
> p_value <- 2 * min(pchisq(chi_sq_stat, df), 1 - pchisq(chi_sq_stat, df))
>
> cat("Chi-square Test Statistic:", chi_sq_stat, "\n")
Chi-square Test Statistic: 11.54167
> cat("P-value:", p_value, "\n")
P-value: 0.7138848
>
> # Decision at alpha = 0.05
> alpha <- 0.05
> if (p_value < alpha) {
+   cat("Reject the null hypothesis: The population standard deviation differs from 12.\n"
)
+ } else {
+   cat("Fail to reject the null hypothesis: No significant evidence that the standard dev
iation differs from 12.\n")
+ }
Fail to reject the null hypothesis: No significant evidence that the standard deviation di
ffers from 12.

>
```

(4)	**Data file: Bags** **Bag-Filling Machines at Jelly Belly.** The variance in a production process is an important measure of the quality of the process. A large variance often signals an opportunity for improvement in the process by finding ways to reduce the process variance. Jelly Belly Candy Company is testing two machines that use different technologies to fill three pound bags of jelly beans. The file *Bags* contains a sample of data on the weights of bags (in pounds) filled by each machine. Conduct a statistical test to determine whether there is a significant difference between the variances in the bag weights for two machines. Use a .05 level of significance. What is your conclusion? Which machine, if either, provides the greater opportunity for quality improvements?

```
#Q.4
> # Read the data
> library(readxl)
> data<- read_excel("C:\\Users\\Dnyanaish\\OneDrive\\Attachments\\Desktop\\Bags.xls
x")
> # Separate Machine 1 and Machine 2
> machine1_weights <- data$`Machine 1`
> machine2_weights <- data$`Machine 2`
> # Perform F-test (handle missing values!)
> f_test_result <- var.test(machine1_weights, machine2_weights, alternative = "two.
sided", na.action = na.omit)
> # Print F-test result
> print(f_test_result)

        F test to compare two variances

data:  machine1_weights and machine2_weights
F = 8.3572, num df = 24, denom df = 21, p-value = 6.7e-06
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
  3.529944 19.312889
sample estimates:
ratio of variances 
          8.357233

> # Extract p-value
> p_value <- f_test_result$p.value
> # Set significance level
> alpha <- 0.05
> # Calculate sample variances (ignoring NA)
> var1 <- var(machine1_weights, na.rm = TRUE)
> var2 <- var(machine2_weights, na.rm = TRUE)
> # Display the variances
> cat("Sample variance for Machine 1:", var1, "\n")
Sample variance for Machine 1: 0.01006667
> cat("Sample variance for Machine 2:", var2, "\n")
Sample variance for Machine 2: 0.001204545
> # Decision based on p-value
```

```
> if (p_value < alpha) {
+   cat("The variances of the two machines are significantly different.\n")
+   if (var1 > var2) {
+     cat("Machine 1 has a larger variance, indicating more variability in bag weig
hts.\n")
+     cat("Machine 1 provides a greater opportunity for quality improvements.\n")
+   } else {
+     cat("Machine 2 has a larger variance, indicating more variability in bag weig
hts.\n")
+     cat("Machine 2 provides a greater opportunity for quality improvements.\n")
+   }
+ } else {
+   cat("The variances of the two machines are not significantly different.\n")
+ }
The variances of the two machines are significantly different.
Machine 1 has a larger variance, indicating more variability in bag weights.
Machine 1 provides a greater opportunity for quality improvements.
>
```

(5)	**Academy Awards and Movie Fan Sentiment.** The race for the 2013 Academy Award for Actress in a Leading Role was extremely tight, featuring several worthy performances. The nominees were Jessica Chastain for *Zero Dark Thirty*, Jennifer Lawrence for *Silver Linings Playbook*, Emmanuelle Riva for *Amour*, Quvenzhané Wallis for *Beasts of the Southern Wild*, and Naomi Watts for *The Impossible*. In a survey, movie fans who had seen each of the movies for which these five actresses had been nominated were asked to select the actress who was most deserving of the 2013 Academy Award for Actress in a Leading Role. The responses follow.

	18–30	31–44	45–58	Over 58
Jessica Chastain	51	50	41	42
Jennifer Lawrence	63	55	37	50
Emmanuelle Riva	15	44	56	74
Quvenzhané Wallis	48	25	22	31
Naomi Watts	36	65	62	33

a. How large was the sample in this survey?

b. Jennifer Lawrence received the 2013 Academy Award for Actress in a Leading Role for her performance in *Silver Linings Playbook*. Did the respondents favor Ms. Lawrence?

c. At $\alpha = .05$, conduct a hypothesis test to determine whether people's attitude toward the actress who was most deserving of the 2013 Academy Award for Actress in a Leading Role is independent of respondent age. What is your conclusion?

```
> #Q.5
> # Create the data
> survey_data <- matrix(
+   c(51, 50, 41, 42,   # Jessica Chastain
+     63, 55, 37, 50,   # Jennifer Lawrence
+     15, 44, 56, 74,   # Emmanuelle Riva
+     48, 25, 22, 31,   # Quvenzhané Wallis
+     36, 65, 62, 33), # Naomi Watts
+   nrow = 5,
+   byrow = TRUE
+ )
>
> # Add row and column names
> rownames(survey_data) <- c("Jessica Chastain", "Jennifer Lawrence", "Emmanuelle R
iva", "Quvenzhane Wallis", "Naomi Watts")
> colnames(survey_data) <- c("18-30", "31-44", "45-58", "Over 58")
>
> # View the data
> print(survey_data)
                  18-30 31-44 45-58 Over 58
Jessica Chastain     51    50    41      42
Jennifer Lawrence    63    55    37      50
Emmanuelle Riva      15    44    56      74
Quvenzhane Wallis    48    25    22      31
Naomi Watts          36    65    62      33
>
> # Part (a) - How large was the sample?
> sample_size <- sum(survey_data)
> cat("Sample Size:", sample_size, "\n")
Sample Size: 900
>
> # Part (b) - Did the respondents favor Jennifer Lawrence?
> jennifer_votes <- sum(survey_data["Jennifer Lawrence", ])
> cat("Votes for Jennifer Lawrence:", jennifer_votes, "\n")
```

```
Votes for Jennifer Lawrence: 205
> percentage_jennifer <- (jennifer_votes / sample_size) * 100
> cat("Percentage of votes for Jennifer Lawrence:", round(percentage_jennifer, 2),
"%\n")
Percentage of votes for Jennifer Lawrence: 22.78 %
>
> # Part (c) - Chi-Square Test of Independence
> chi_sq_test <- chisq.test(survey_data)
>
> # Display chi-square test result
> print(chi_sq_test)

        Pearson's Chi-squared test

data:  survey_data
X-squared = 77.736, df = 12, p-value = 1.113e-11

>
> # Conclusion at alpha = 0.05
> alpha <- 0.05
> if (chi_sq_test$p.value < alpha) {
+   cat("Reject the null hypothesis: People's attitude toward the actress depends o
n the respondent's age group.\n")
+ } else {
+   cat("Fail to reject the null hypothesis: People's attitude toward the actress i
s independent of the respondent's age group.\n")
+ }
Reject the null hypothesis: People's attitude toward the actress depends on the res
pondent's age group.
```