THE UNIVERSITY OF SYDNEY

STAT5002 Introduction to Statistics

Semester 1 2025

Lecturer: Tiangang Cui

This individual assignment is due by **11:59pm Sunday 25 May 2025**, via Canvas. There are **three** questions in this assignment with a total of 80 points.

Your solution must be submitted in either of the following formats:

- A single combined PDF file that includes your written answers along with screenshots of your code (and relevant outputs, if necessary); or
- A knitted HTML file that includes your code and typed answers to each of the questions. If you choose this option, please submit only the knitted HTML file (not the Rmd source file), as markers will only assess content that is viewable on Canvas.

Regardless of the format, you must include justifications for your calculations and interpretations of your results. You should only submit one combined file with answers to each of the questions clearly labelled.

You may round each step of your calculations to two or three significant digits. Please clearly state which level of rounding you have used.

You may use the output of `t.test()` and `chisq.test()` to check your results. However, you must show your work without relying on these functions unless specified in the question.

Your submitted file must include your SID. To comply with anonymous marking policies, do not include your name anywhere in your assignment. Only your SID should be present.

Please review your submission carefully – what you see is exactly what the marker will see. You may revise and resubmit your work until the due date.

1. **(30 point)** An automotive engineer wants to determine whether a new type of tire provides better vehicle braking deceleration (in metres per second squared, m/s^2) compared to the standard tire. Twenty vehicles are randomly selected from the production line. Each vehicle is tested twice: once with the new tire (`test.A`) and once with the standard tire (`test.B`). The paired results (New Tire vs Standard Tire) are shown below:

```
test.A = c(5.90, 5.26, 2.97, 7.15, 10.06, 11.87, 1.94, 6.27, 6.81, 4.08,
           8.13, 15.18, 8.82, 3.87, 5.23, 11.29, 7.92, 12.82, 7.20, 10.03)
test.B = c(6.07, 4.89, 2.92, 7.00, 9.99, 11.70, 1.94, 5.86, 6.95, 4.03, 7.76,
           15.02, 9.08, 3.73, 4.88, 10.81, 8.05, 12.96, 7.10,  10.07)
```

You may copy the above data into R and use R to assist with the calculations (without using built-in functions for hypothesis tests). Please clearly explain each step of your calculation and its interpretation.

(a) Introduce appropriate parameters and state the null and alternative hypotheses.

(b) What statistical test should be used to test these hypotheses? Justify your choice.

(c) Use appropriate graphical summaries to assess whether the necessary assumptions for applying the chosen test are satisfied.

(d) Compute the observed test statistic and P-value. In your answer, clearly state the distribution of the test statistic and indicate which values of the statistic argue against the null hypothesis.

(e) What is your conclusion based on the calculated P-value? You can either specify your own significance level or use the default 5% level to draw your conclusion.

(f) Perform a bootstrap simulation (with 10,000 repetitions) to simulate the test statistic, and plot the histogram of the simulated statistics. Does the histogram of the simulated test statistics agree with the theoretical test distribution used above?

(g) What is the P-value based on the simulated test statistics?

Solution

(a) Let μ_D be the population mean of the paired differences (test.A − test.B). We want to test whether the new tire improves performance:

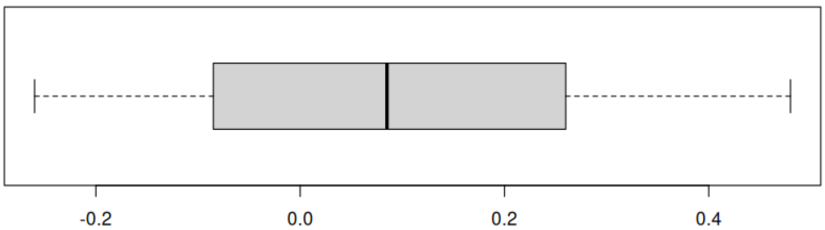
$$H_0 : \mu_D = 0 \quad \text{(no improvement)}$$
$$H_1 : \mu_D > 0 \quad \text{(new tire has higher deceleration)}$$

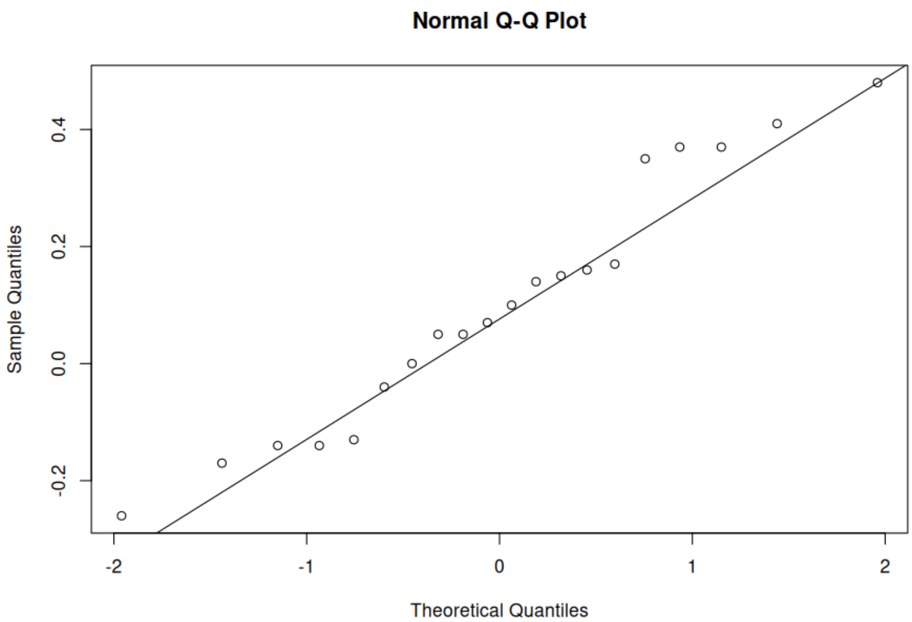
(b) Since measurements are paired for each vehicle, we can use a paired T-test for the mean difference.

(c) Since the testing vehicles are randomly selected from a large population, we may argue that the independence assumption is satisfied.

We can use boxplot and QQ plot to check the normality. You may also use histogram. One can check either the normality of both samples or the normality of the difference.

```
> diff = test.A - test.B
> boxplot(diff, horizontal = T)
> qqnorm(diff)
> qqline(diff)
```





The boxplot appears roughly symmetric and the quantile points of the QQ plot follow closely the QQ line. So we conclude that the normality assumption is satisfied.

(d) The test statistic is the T-statistic

$$T = \frac{\bar{X} - 0}{\widehat{SE}(\bar{X})} = \frac{\bar{X} - 0}{\hat{\sigma}/\sqrt{n}}$$

following Student's t-distribution with $n-1 = 19$ degrees of freedom. Plugging the observed sample mean difference and the sample SD of the difference into the above formula, we obtain the observed T-statistic as follows.

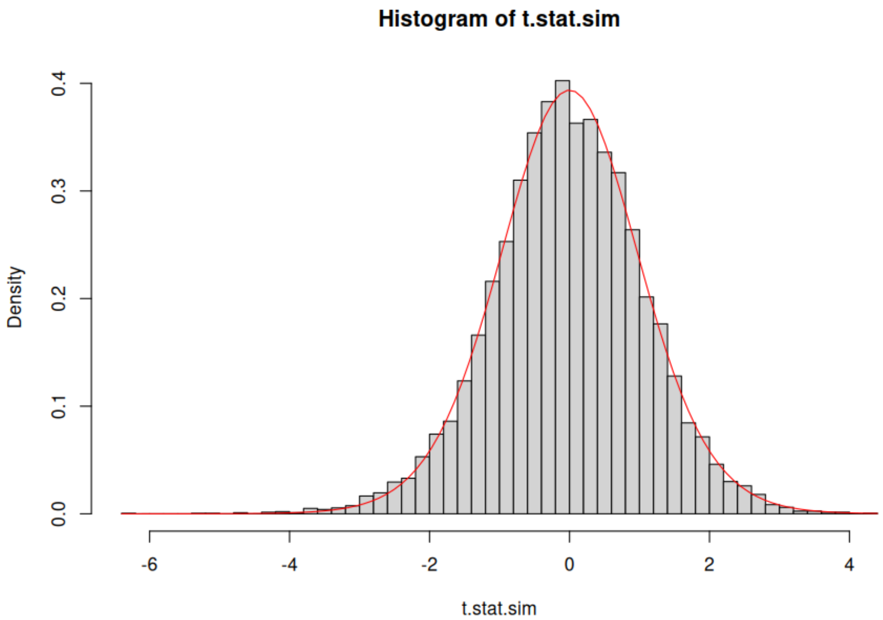
```
> mean.d <- mean(diff)
> sd.d <- sd(diff)
> n <- length(diff)
> t.stat <- mean.d / (sd.d / sqrt(n))
> p.val <- pt(t.stat, df = n - 1, lower.tail=F)
> t.stat
[1] 2.082483
> p.val
[1] 0.0255262
```

The P-value is 0.0255262.

(e) Using the default 5% level of significance, we reject the null hypothesis, which indirectly suggest that the new tire provides higher braking deceleration.

(f) Here is my code for simulating the test statistic.

```
t.stat.sim = 0
box.g = diff
for (i in 1:10000){
    samp = sample(box.g, size = n, rep = T)
    t.stat.sim[i] = (mean(samp) - mean(box.g))/sd(samp)*sqrt(n)
}
hist(t.stat.sim, freq=F, n=50)
curve(dt(x, df = n - 1), add = T, col = "red")
```



The histogram of the simulated test statistics agree with the theoretical test distribution used above, which also suggests that the assumptions for applying Student's t-distribution are reasonable.

(g) The simulation-based P-value is 0.0251, which is again very close the what we obtained using Student's t-distribution.

```
> mean(t.stat.sim>t.stat)
[1] 0.0251
```

2. (25 point) A school administrator is investigating whether a new online tutoring program leads to a difference in student performance compared to a traditional in-person tutoring method. Students are randomly assigned to two groups: Group A (Online Tutoring) and Group B (In-Person Tutoring). The normalised student performance indicators for these two groups are shown as follows.

```
group.A = c(5.54, 4.41, 6.35, 5.04, 7.33, 6.47, 4.08, 6.00, 7.39, 5.53, 1.54, 6.16,
            4.23, 2.36, 5.09, 5.10, 5.33, 3.75, 6.49, 2.13, 5.44, 7.74, 3.80)
group.B = c(4.31, 6.20, 5.25, 2.14, 3.26, 1.47, 2.24, 4.20, 3.56, 3.68, 7.02, 2.94,
            5.49, 3.37, 4.59, 3.05, 5.24)
```

You may copy the above data into R and use R to assist with the calculations, but do not use built-in functions for hypothesis tests except in Part (e). Please clearly explain each step of your calculation and its interpretation.

(a) State the null and alternative hypotheses to test whether two programs have the same effect on the student performance. In answering, introduce appropriate parameters, as well as a null and alternative hypothesis in terms of these parameters.

(b) Use appropriate graphical and numerical summaries to assess whether the necessary assumptions for applying the classical two-sample t-test are satisfied.

(c) Write down the formula for the test statistic of the classical two-sample t-test, and calculate the observed test statistic. Show your working step by step, rounding each step to three decimal places.

(d) Construct the critical region of rejection at the 5% level of significance. What is your conclusion of the hypothesis test based on the critical region?

(e) Now conduct a Welch test using R, what is the computed P-value, how does it compare with the classical two-sample t-test?

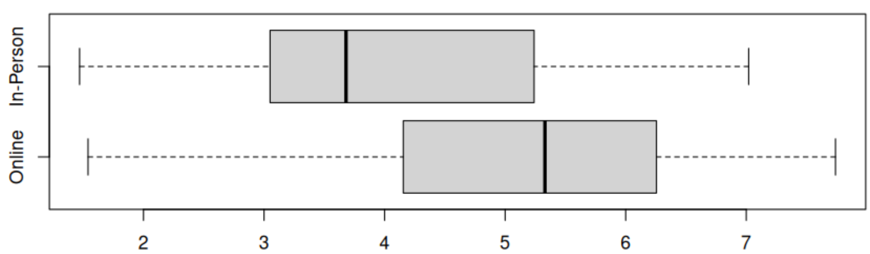
Solution

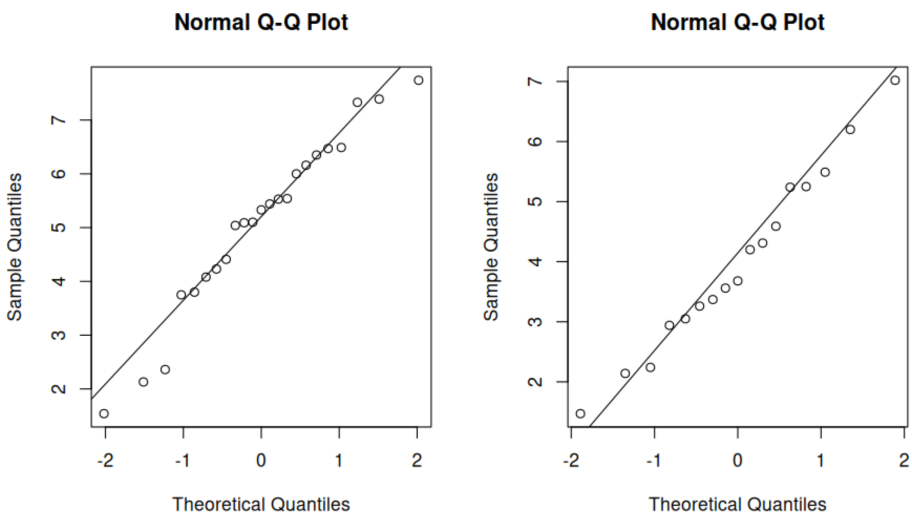
(a) Let μ_A and μ_B be the population means for the online and in-person groups, respectively.

$$H_0 : \mu_A = \mu_B \quad \text{(no difference in performance)}$$
$$H_1 : \mu_A \neq \mu_B \quad \text{(a difference in performance)}$$

(b) We assess normality and variance equality:

```
> boxplot(group.A, group.B, names = c("Online", "In-Person"))
> par(mfrow = c(2,2))
> hist(group.A, main = "Histogram: Online (A)")
> hist(group.B, main = "Histogram: In-Person (B)")
> qqnorm(group.A); qqline(group.A)
> qqnorm(group.B); qqline(group.B)
```





The boxplot appears roughly symmetric and the quantile points of the QQ plots follow closely the QQ lines. So we conclude that the normality assumption is satisfied.

```
> sd.A = sd(group.A)
sd.A
[1] 1.64911
> sd.B = sd(group.B)
sd.B
[1] 1.499781
```

There is a small difference between sample SDs (about 10%), which suggest that the classical two-sample T-test may not be appropriate.

(c) We first calculate the pooled estimate of the common SD:

```
n.A <- length(group.A)
n.B <- length(group.B)
> pooled.SD = sqrt( ( sd.A^2*(n.A-1) + sd.B^2*(n.B-1) ) / (n.A + n.B - 2) )
> pooled.SD
[1] 1.587947
```

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{22 \cdot (1.649)^2 + 16 \cdot (1.50)^2}{38}} \approx 1.588$$

```
> est.SE = pooled.SD*sqrt(1/n.A + 1/n.B)
> est.SE
[1] 0.5078994
> mean.A = mean(group.A)
> mean.B = mean(group.B)
> (mean.A - mean.B)/est.SE
[1] 2.164625
```

This gives the estimated SE of 0.508 and the observed T-statistic

$$t = \frac{\bar{x} - \bar{y}}{\widehat{SE}(\bar{X} - \bar{Y})} = \frac{}{0.5078994} \approx 2.165$$

(d) Since it's a two-sided alternative, the critical value corresponding to the 5% level of significance is

```
> qt(1-0.025, df = n.A+n.B-2)
[1] 2.024394
```

This gives a critical region of $|T| > 2.024$. The observed T-statistic falls inside the critical region, so we reject H_0.

(e) The P-value of the classical two-sample T-test is 0.044. Using the function `t.test()`, the calculated P-value of the Welch test is 0.035, which is smaller than that of classical two-sample T-test.

```
> 2*pt(t.stat, df = n.A+n.B-2, lower.tail = F)
[1] 0.04408162
> t.test(group.A, group.B)

        Welch Two Sample t-test

data:  group.A and group.B
t = 2.1964, df = 36.294, p-value = 0.03453
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.0845232 2.1143003
sample estimates:
mean of x mean of y
 5.100000  4.000588
```

3. (25 point) Consider the table below, which shows the number of individuals in each of four age groups (under 18, 18–29, 30–49, 50+) who prefer one of three device types: laptop, desktop, or tablet. The data comes from a survey on technology usage preferences:

	Laptop	Desktop	Tablet	Total
Under 18	12	6	12	30
18–29	14	10	6	30
30–49	16	12	12	40
50+	8	16	6	30
Total	50	44	36	130

We are interested in testing whether device preference (laptop, desktop, or tablet) is independent of age group. Follow the steps below to perform a chi-squared test to investigate this without using built-in functions for hypothesis tests such as `chisq.test()`.

(a) State the null and alternative hypotheses.

(b) Set up the table of expected frequencies.

(c) Discuss whether the necessary assumptions for applying the chi-squared test are satisfied.

(d) Compute the observed test statistic.

(e) Construct the critical region of rejection at the 5% level of significance. What is your conclusion of the hypothesis test? Justify your answer.

Solution

(a) H_0: Device preference (laptop, desktop, tablet) is independent of age group.
H_1: Device preference is not independent of age group. E.g., device preference varies by age group.

(b)
```
> device_data <- matrix(c(
+     12, 6, 12,  # Under 18
+     14, 10, 6,  # 18{29
+     16, 12, 12, # 30{49
+     8, 16, 6    # 50+
+ ),
+ nrow = 4, byrow = TRUE)
> device_data
     [,1] [,2] [,3]
[1,]   12    6   12
[2,]   14   10    6
[3,]   16   12   12
[4,]    8   16    6
> rs = rowSums(device_data)
> cs = colSums(device_data)
> E = round(outer(rs, cs, FUN="*")/sum(device_data), 2)
> E
      [,1]  [,2]  [,3]
[1,] 11.54 10.15  8.31
[2,] 11.54 10.15  8.31
[3,] 15.38 13.54 11.08
[4,] 11.54 10.15  8.31
```

We compute the expected count in each cell using:

$$E_{ij} = \frac{(\text{Row total}) \times (\text{Column total})}{\text{Grand total}}$$

This gives the table of expected counts:

Age Group	Laptop	Desktop	Tablet
Under 18	11.54	10.15	8.31
18–29	11.54	10.15	8.31
30–49	15.38	13.54	11.08
50+	11.54	10.15	8.31

(c) Since observations are independent, and all expected counts are above 5 (as shown above), and the sample size is 130, so it is appropriate to apply the chi-squared test.

(d) The chi-squared test statistic is calculated as:

$$\chi^2 = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

The observed test statistics is approximately 9.90.

(e) The degrees of freedom (df) for the chi-squared test are calculated as:

$$df = (\text{number of rows} - 1) \times (\text{number of columns} - 1) = (4 - 1) \times (3 - 1) = 6$$

The chi-squared test statistic argues against the null hypothesis when it is large.
For a significance level of $\alpha = 0.05$, the critical value for $df = 6$ from the chi-squared distribution table is 12.59, i.e., $P(\chi^2 \geq 12.59) = 0.05$. Therefore, any test statistic larger than 12.59 would provide evidence to reject the null hypothesis, which gives a critical region $(12.59, \infty)$. Since our computed chi-squared test statistic is 9.90, which is not in the critical region, we do not have sufficient evidence to reject the null hypothesis.