STAT2170 AND STAT6180
APPLIED STATISTICS

SGTA Week 5 : One-Way Analysis of Variance and Multiple Comparisons 2023

Part 1

Question 1

A study of the properties of metal plate connected trusses used for roof support ("Modelling Joints Made with Light Gauge Metal Connector Plates"" J. of Forest Products 1979: p39-44) yielded the following observations on Axial Stiffness Index (ASI) in kips/inch. Each of the 5 different treatments was replicated 7 times:

4in	6in	8in	10in	12in
309.2	402.1	392.4	346.7	407.4
409.5	347.2	366.2	452.9	441.8
311.0	361.0	351.0	461.4	419.9
326.5	404.5	357.1	433.1	410.7
316.8	331.0	409.9	410.6	473.4
349.8	348.9	384.2	384.2	441.2
309.7	381.7	382.0	362.6	465.8

The data are stored in a text file, `axial.txt`, available on iLearn.

a. Check if the data is suitable for an ANOVA to test the mean ASI level across the different plate lengths. That is, check the common variance requirement and that the residual variation in the data follows a normal distribution.
b. Assuming the data is valid, carry out the analysis of variance to determine if there are any differences in ASI for the 5 different plate lengths (4", 6", 8" 10" and 12"). Remember in your answer to
   - Define necessary parameters
   - State the hypotheses
   - State the test statistic
   - State the null distribution
   - Compute the P-value or region
   - Give both the statistical and contextual conclusion.
c. If the 5 treatments are significantly different at the 5% level of significance, you should use the Tukey's procedure to test for significant differences between the 5 different treatments. Clearly state your findings/conclusions here.
d. Compute the Bonferroni adjusted margin of error

Question 2 (to be done by hand)

Compare each of the other four metal plates with the 4inches metal plate (the control treatment here), using the Bonferroni procedure at an overall significance level of 5%.

- Show all calculations, but do use the ANOVA results from Question 1 if needed and get relevant descriptive statistics (eg, means, t-quantile etc) in **R** if necessary.
- Give a brief conclusion based on the results.

Question 3 (by hand)

Suppose a dataset of three populations had the three different treatments applied and a response variable recorded. The numeric summary of the dataset is shown in the following summary table:

	$\overline{y}_{i\cdot}$	s_i	n
Treatment A	71.1	6.86	56
Treatment B	70.8	7.45	54
Treatment C	72.4	6.73	57

Using this information,

a. Show that the Grand (Overall) mean of the data is $\overline{y}_{\cdot\cdot} = 71.44671$.
b. Compute the Treatment Sum of Squares.
c. Compute the Residual Sum of Squares.

Question 4

Give the p-value, as accurately as possible from the F-tables, for the following each of four (4) observed F values from an analysis of variance.

a. $F_{obs} = 4.20$ for an experiment with 4 treatments, each replicated 8 times.
b. $F_{obs} = 4.20$ for an experiment with 10 treatments, each replicated 3 times.
c. $F_{obs} = 3.20$ for an experiment with 4 treatments, each replicated 4 times.
d. $F_{obs} = 3.20$ for an experiment with 2 treatments, each replicated 20 times.

Question 5 (by hand)

For this question, we are going to reuse the same dataset in Questions 1 and 2. Define two general contrasts:

$$C_1 = \frac{\mu_{6in} + \mu_{8in} + \mu_{10in} + \mu_{12in}}{4} - \mu_{4in}, \quad \& \quad C_2 = \mu_{12in} - \mu_{4in}$$

i.e. C_1 compares the other four metal plates as one with the 4inches metal plate (the control treatment here) and C_2 simply compares the 12inches metal plate with the 4inches one.

Test whether these two contrasts are significantly different to 0 simultaneously while controlling the overall significance level at 0.05 with the Bonferroni approach.

Part 2

Data frame is a fundamental data structure in `R` and `RStudio`; for example, the datasets we import into `R` or `RStudio` often exist in the form of data frames. In a data frame, each column contains values of one variable and each row contains one set of values from each column. The exercises below will explore the structure and functions on data frame a bit further.

a) Please import to `R` or `RStudio` the dataset `axial.txt`, available on iLearn, and name it `axial`.

b) Use `is.data.frame()` to confirm `axial` is indeed a data frame.

c) Now let's change some individual value in `axial` — Please first create a duplicate of `axial`, name it as `axial_new`, and then change the value on the third row and first column of `axial_new` from 311.0 to 310.0.

- Hint: `axial_new[1,2]` gives the value on the first row and second column

d) Now let's rescale all the values in `asi` of `axial_new` — Multiply all values on the first column by 10 and then use it to replace the existing column.

- Hint: `axial_new[,2]` and `axial_new$treatment` can both give the values on the second column in `axial_new`

e) To select subsets of data, we may need logical operators. Some logical operators are given in Table 3. Now please correct all the errors in the code below, which tries to select all rows in `axial_new` that are associated with treatment `4in` **or** `6in` and save it to `axial_sub`.

```
axial_sub = subset(axial_new, treatment = "4in" & treatment = "6in")
```

Operator	Description
==	equal to
!=	not equal to
x & y	both events x and y occur
x | y	some of events x or y occurs
!x	event x does not occur

Table 3: Some logical operators

f) After fixing the code in the previous sub-question, please verify that the second column of `axial_sub` contains and only contains `4in` and `6in`, using function `unique()`.

g) The function `write.table` can export a data frame. Please correct all the errors in the code below, which tries to export the `axial_sub` in the sub-question above to a file named `axial_sub.txt` to your working directory.

```
write.table(axial_sub, file = axial_sub.txt)
```

Afterward, using the corrected code to complete this exportation and then try to find this `axial_sub.txt` file in your working directory.