MACQUARIE University
SYDNEY·AUSTRALIA

STAT2170 and STAT6180
Applied Statistics

Part 1

Question 1

Six different recipes for commercial pizza dough were evaluated to determine which recipe proofed the fastest. Each of the 6 recipes was replicated four times in a random order. The time taken in minutes for the dough to rise to the top of the bowl was recorded. (modified from the data in O'Neill (1989))

The first seven observations are recorded below. The full dataset of 24 observations is available in the `dough.txt` file available on iLearn (the first six observations are shown in the table below). Load the dough data into `R`.

times	recipe
35.35	F
31.35	E
34.35	C
31.50	B
36.05	A
26.60	D

Carry out the appropriate Analysis of Variance to investigate what differences, if any, there are between the six recipes in proofing time.

a. Before carrying out relevant analysis, check and discuss any assumptions of it.
b. Carry out the analysis of variance procedure in `R`. Comment on the results.
c. If appropriate, carry out multiple pair-wise comparisons of all treatments (6 recipes) at the 5% overall significance level, using both Tukey's and Bonferroni's procedures. Comment on the results. (Note: For the Bonferroni method, you need to first work out the significance level for each individual comparison. Relevant R functions are `pairwise.t.test` and `TukeyHSD`. The function `pairwise.t.test` requires the raw data input. However, `TukeyHSD` requires the ANOVA object as input. Type `? pairwise.t.test` and `? TukeyHSD` for more information)
d. Carry out the Kruskal-Wallis test for the data above in `R`

- interpret the results
- compare the result here with that from the one-way ANOVA in part c, and briefly explain the difference if any
- is this analysis necessary for this data? (may refer to the assumptions for ANOVA)

Part 2

Question 1

Conceptually, `factors` in R or RStudio are variables taking on a limited number of different values. Such variables are also refereed to as categorical variables. Examples of `factors` include gender, educational level, and brand of fast food chains. This question is to explore more about `factors`.

a) Suppose 1, 2, 3 stands for KFC, Subway and McDonald respectively. Use `factor()` to turn the numeric vector `c(2, 1, 2, 1, 3, 3)` into a factor vector named `food`.

b) Now confirm that `food` is indeed a `factor` using `str()`, which reveals the structure of an `R` object.

c) Now please change the first element in `food` from `2` to `3`

   - Hint: `food[6]` will gives the 6th element in `food`.

d) Function `levels()` can produce the labels of the factors. Please correct the all the errors in the code below, which tries to change the labels into `McDonald`, `Subway` and `KFC`.

   ```
   level(food) = c(KFC, Subway, McDonald)
   ```

e) After fixing the codes in the previous sub-question, please use `summary()` or `table()` to check the frequency of `McDonald`, `KFC`, and `Subway` in `food`.

f) Use `plot()` to obtain a bar chart of the variable `food`. Why do you think the bars are sorted on the x-axis this way?

g) There are many ways to re-order the levels in a factor but here we will focus on the following two options here.

   - `relevel()` to specify the reference level, i.e the category that is first on the list. This has a particular implication in a regression setting.
   - Use the `levels` argument in `factor()`

   You should check the help menu to find out how to use these options. Using these information, reorder (and replace) the `food` variable:

   i) Use `McDonald` as the reference level;
   ii) Sort the levels in alphabetical order.

   You can call the variable or use `plot()` or `levels()` to check the factor has been put in the right order.

Question 2

`Lists` in `R` or `RStudio` are objects which contain elements of different types, such as numbers, strings, vectors, and even factors. This question gives out more information about `Lists`.

a) Using function `list()`, create a list named `lunch`, which contains the following elements in order:

   - 4
   - `c("fork","knife")`
   - the factor `food` in the previous question

b) Function `names()` gives names of the element in a list. Now please correct all the errors in the code below, which tries to change the names of `lunch` into

   ```
   name(lunch) <- c("GuestNumber", "Utensil", "Meal")
   ```

c) After correcting the code in the previous sub-question, enter `lunch` in `R` console to see its elements. Now try to change the `"McDonald"` entry in `lunch` into `"KFC"`. Enter `lunch` again in `R` console to see if you have made the change successfully.

   - Hint: `lunch$Utensil` generates output `c("fork","knife")`