

NATIONAL UNIVERSITY OF SINGAPORE

Department of Statistics and Data Science

**ST2137 Statistical Computing and Programming**

(Semester 2 : AY 2021/2022)

Individual Assignment

**Due Date: 08 April 2022 by 5 pm**

---

**INSTRUCTIONS TO STUDENTS**

1. Students are supposed to submit the answers on time. Any submission after 5 pm of the due date will get penalty: deduct 10% of the awarded mark for every two hours late.
2. **Students are required to complete this assignment individually.**
3. All submission is done online.
4. **Your answer should be an R file (only includes the R code).** Make sure that there is no error when the graders open and run your R code.  
Be sure to lay out systematically the various parts and steps in your working.
5. Your submission file should have the name of the format: A0123456B.R where A0123456B is your matrix number.

---

General idea of this assignment: We have a data of two variables from a random sample. One quantitative variable and one categorical variable which has more than 2 categories. The categorical variable indicates the groups which are labelled either by numeric (1, 2, 3, etc.) or by characters (A, B, C, etc.).

Write a function which helps to conduct the Bonferroni correction for pairwise t-tests, named as `Bonf.PWT`.

I You should have sub-functions as given below before defining `Bonf.PWT`.

1. `normtest`: a function that helps to test the normality of each group given in the dataframe. It should return the p-values for each group separately.
2. `EVT`: a function for equal variance test, to test if variances of all groups are equal. `EVT` should have `df` as its argument. It can use `normtest` in its body.  
It should return a p-value from a Bartlett test or a Levene test depending on the output of `normtest`.

II The main function `Bonf.PWT` should have at least three arguments.

1. `df`: a dataframe which should have two columns: the first column includes all the measurement values (quantitative); the second column is the categorical variable to indicate groups.
2. `norm.alpha`: a significance level for a normality test (Shapiro-Wilk test) for a group. Default value `norm.alpha = 0.05`.

The use of this argument is: if any of the groups is not normal (Shapiro-Wilk test for that group has p-value less than `norm.alpha`), then we would perform a Levene test to check if variances of groups are equal; otherwise a Bartlett test should be used.

3. `var.alpha`: a significance level for equal variance test for all groups. Default value `var.alpha = 0.05`.

The use of this argument is: if the p-value of the equal variance test is larger than `var.alpha` then we would perform all the pairwise t-tests with equal variance, otherwise we should perform all the pairwise t-tests with unequal variances.

III Function `Bonf.PWT` should use both `normtest` and `EVT` in its body. It should return:

1. The information whether all the pairwise t-test were conducted using equal variance or unequal variance
2. Results of all the pairwise t-tests where each test has the information of: two groups tested; the original p-value of the test (must be numeric); and the adjusted p-value (must be numeric).

IV Use the function `Bonf.PWT` that you have defined to run on the three datasets: `tablets1.txt` (mentioned in Topic 8), `locate.txt` (mentioned in Tutorial 8) and `crab.csv` (mentioned in Tutorial 9). For the dataset `crab.csv`, run `Bonf.PWT` to compare the average weight (in kg) of the crabs across three types of spine (1 = both good, 2 = one worn or broken, 3 = both worn or broken).

Report the output when running function `Bonf.PWT` for each dataset as comments in the file of code.

**Note:** for any dataset, the user should transform or change the given data into the format of `df` (with only two columns) by himself before running `Bonf.PWT` on `df`.

END OF ASSESSMENT