MTH5131 Actuarial Statistics

Coursework 4

To be done by 9:00 on Monday 18 March

This coursework counts 20% of your final mark. You should work alone on this coursework. You must upload your single R script containing your solutions by 9:00 Monday 20 March 2023 to get credit for your work. Your R script should be well commented and it should be easy to see from your script what are your answers to the questions. ANY SCRIPTS RECEIVED AFTER THE DEADLINE ABOVE WILL NOT BE MARKED.

Exercise 1 (*40 marks)*.

1. Read the file Computers.csv into a data frame **DF**. Look at its structure. How many of the variables are factors?
2. What are the values that trend takes? How many times does it take the value 4?
3. Create a new data frame called **DF2** consisting of the variables (from left to right) price, hd and ram from **DF**.
4. Look at the scatterplot between the variables in **DF2** and comment on the correlations.
5. Find the Spearman correlations between the variables in **DF2** and comment on them.
6. Using attach, make a new data frame **DF3** consisting of the rows of DF2 for which hd> 500.
7. Making use of ATTACH and DETACH, find the mean of the variable hd for **DF2** and **DF3**
8. Look at the scatterplot of **DF3**. Find the Spearman correlations between the variables in **DF3** and compare them to the correlations you found in step 5.
9. Suppose you know that a vector X has 6 distinct values and that X is ordered from smallest to largest, while vector Y has 6 distinct values and that Y is ordered from largest to smallest. Is it possible to say what is the Spearman correlation from this information? If it is possible, then what is the Spearman correlation?
10. For **DF2**, perform a test of $H_0 : \rho = 0$ versus $H_1 : \rho \neq 0$ for the Spearman correlation of the variables hd and price. Store it in an object called **test**.
11. Extract the p-value from **test** and decide whether or not H_0 should be accepted or rejected at the 5% confidence level.
12. Create a data frame called **DF4** from columns 2 through 6 of DF.
13. Run PRCOMP on **DF4** and use it to estimate how many components are principle.
14. Plot the scree plot for **DF4** and use it to estimate how many components are principle.

Exercise 2 (*15 marks)*.

1. Using a seed of 32, generate 500 values from standard normal random variables and use them to simulate 500 Log-Normal$(12, 16)$ data points. Look at the histogram of the Log-Normal values and comment on it.

2. Using a seed of 23, generate 500 values from Log-Normal$(12, 16)$ random variables (without first generating normal random variables) and use them to simulate 500 standard normal data points. Look at the histogram of the standard normal values and comment on it.

3. Find the empirical probability that the log-normals generated in step 2 are bigger than 10^6. Compare with the true probability.

Exercise 3 (*20 marks)*.

1. Draw the PDFs of the t-distributions for 1,2, and 3 degrees of freedom. Let the x-axis vary from -5 to 5 and make sure all PDF'S are completely visible. Add a legend.

2. Let X be Poisson(3) distributed. Find $P(5 \leq X \leq 17)$.

3. Using a **for** loop and a seed of 53, generate a vector of length 20, whose first value is 0.499 and such that every other value is two times the square of the value previous to it. Display your answer.

Exercise 4 (*15 marks)*. Some data is contained in the file bootstrap.txt It is assumed that the data is N$(0, \sigma^2)$ distributed, where σ^2 is unknown. With seed 43, obtain 200 (non-parametric) bootstrap samples estimating $\hat{\sigma}^2$, where $\hat{\sigma}^2$ is the method of moments estimator of σ. Find the mean and standard deviation of the estimates.