

Computer Project 1

(Due on 10/26/2009, Monday)

In this project, we consider the `faithful` dataset available in R, which record waiting time between eruptions and the duration of the eruption for the Old Faithful geyser in Yellowstone National Park, Wyoming, USA. There are 272 observations on 2 variables

A data frame with 272 observations on 2 variables.

- `eruptions` - Eruption time in mins
- `waiting` - Waiting time to next eruption (in mins)

Concerning the eruption time `eruptions` only, complete the following steps using R (or any other software that you are familiar with). Work by hand is not acceptable.

1. Compute its mean, sample standard deviation, the 10th, 25th, 50th, 75th, 80th, and 90th percentiles.
2. Perform a test to see whether or not the third quartile of the eruption time is below 4.5 minutes. Note that you may use the following command to count number of eruption times below 4.5:

```
count.below <- sum(faithful$ eruptions < 4.5)
```

3. Plot its empirical cumulative distribution function.
4. Apply the runs test to see if the eruption times recorded in the data set seem to be random.
5. Normality Test
 - a. Apply the χ^2 test for normality with grouping of your choice. Make sure that the expected counts are all no less than 5.
 - b. Obtain the Q-Q normal probability plot and perform Shapiro-Wilk's test.
 - c. Apply Komolgorov goodness-of-fit test to check if the data seem to come from $\text{Normal}(\mu = 4, \sigma = 1)$. Then apply Lilliefors test to check normality. Report the p-values.

Remember to interpret every result that you present in the report. Also, provide the R codes in an appendix.