Tutorial 8

ST2137-2420

Material

This tutorial covers the topics and concepts from chapter 7, which covers hypothesis tests for comparing means between two groups.

As always, work out each question with R, Python and SAS unless otherwise stated.

Question 1

The number of pages in magazines devoted to advertisements varies widely from magazine-to-magazine and from issue-to-issue within the same magazine. Advertising expenditures, and therefore the number of advertising pages in magazines, tend to be the highest during periods of economic growth. The data in the file `weeklies.txt` gives the number of advertising pages in the current issues of 19 weekly magazines and the number of advertising pages in the same issue of the previous calendar year.

1. Conduct an appropriate t-test at 10% level to assess if the mean difference in advertising expenditure is significantly different from 0. Print out the 90% CI for the mean difference.
2. Compute the difference (as `diff_vec` = `current` - `lastyear`) and assess if this variable is Normally distributed using:
   (i) the histogram
   (ii) qq-plot
   (iii) using a hypothesis test of your choice.
3. Write R and Python functions to compute estimates of skewness and kurtosis, as defined in the lecture notes. Apply the functions to `diff`. Ensure that your values agree with these:

```
library(DescTools)
Skew(diff_vec, method=1)
## [1] 0.1283815
# [1] 0.1283815
Kurt(diff_vec, method=1)
## [1] -0.4401952
```

4. The paper Kim and White (2004) contains a robust estimate of skewness h_1:

$$h_1 = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$$

Implement this in R and Python, and apply it to the `diff_vec` data. Is there a large change from part (3)? Can you explain why or why not?

Question 2

The purchasing director for an industrial parts factory is investigating the possibility of purchasing a new type of milling machine. He determines that the new machine will be bought if there is evidence that the parts produced have a higher average breaking strength than those from the old machine. The

data file `machine.txt` represents the breaking strength of samples of 50 parts from the old and the new machines.

5. Is there evidence that the purchasing director should buy the new machine? Ensure that you check all assumptions as required.

Question 3

A flexible working hour program permits employees to design their own 42-hour work week to meet their personal needs. The management of a large manufacturing firm is considering adopting a flextime program for its administrators and professional employees, depending on the success or failure of a pilot program. Ten employees were randomly selected and given a questionnaire designed to measure their attitudes toward their jobs. Each was then permitted to design and follow a flextime workday. After six months, attitudes toward their jobs were again measured. The resulting attitude scores are given in the data file `flextime.txt`. The higher the score, the more favorable the employee's attitude toward his or her work.

6. Use a nonparametric test procedure in SAS to evaluate the success of the pilot flextime program at 5% significance level.
7. Repeat the question above using R and Python.
8. Convert the dataset to long form (use `stack` in R, `pd.DataFrame.stack` in Python). This means that the dataframe should now look like this:

```
   values    ind
1      54 before
2      25 before
3      82 before
4      76 before
5      63 before
6      82 before
7      94 before
8      72 before
9      33 before
10     90 before
11     68  after
12     42  after
13     80  after
14     91  after
15     70  after
16     88  after
17     90  after
18     81  after
19     38  after
20     93  after
```

References

Kim, Tae-Hwan, and Halbert White. 2004. "On More Robust Estimation of Skewness and Kurtosis." *Finance Research Letters* 1 (1): 56–73.