Tutorial 7

ST2137-2420

Material

This tutorial offers practice with SAS. All the questions are centered around reproducing previous R/Python analysis in SAS. Please revise the material from chapter 6 of the course textbook. There are a few questions where you will need to read the SAS documentation. In particular, the `proc univariate` and `proc freq` will be useful.

Dataset: Student Performance

If necessary, the following code snippet can be used to convert `G1` from character to numeric, in a new dataset in SAS.

```
data st2137.stud_perf2;
  set st2137.stud_perf;
  G1_num = input(G1, 8.);
run;
```

1. Generate summary statistics for `G1` scores, conditioned on `Medu`.
2. Generate the following boxplots of `G1` scores, by `Medu`. Compare the distribution of `G1` scores with those of `G3` scores (we had used this variable throughout chapter 3 of the textbook).
3. Conduct a χ^2 test of independence of the variables `famrel` and `goout` at 5% significance level.
4. Obtain and interpret the 90% confidence interval for the odds ratio between variables `nursery` and `higher`.

Working with contingency tables

5. Reproduce the χ^2 test that we performed on the political association data in Example 4.9 of the text.

Robust statistics

6. Upload the `mass_chem` dataset from chapter 5 (robust statistics) to SAS. Use `PROC UNIVARIATE` to obtain the following robust estimates of location and scale:
   - Trimmed mean ($\gamma = 0.1$)
   - Winsorised mean ($\gamma = 0.1$)
   - MAD

   Can you explain the differences with the estimates we obtained from R/Python?