Assignment 2

PSY5113, Spring 2015

This assignment is due by Feb. 17 in class. Please attach only the relevant syntax and output from using statistical packages.

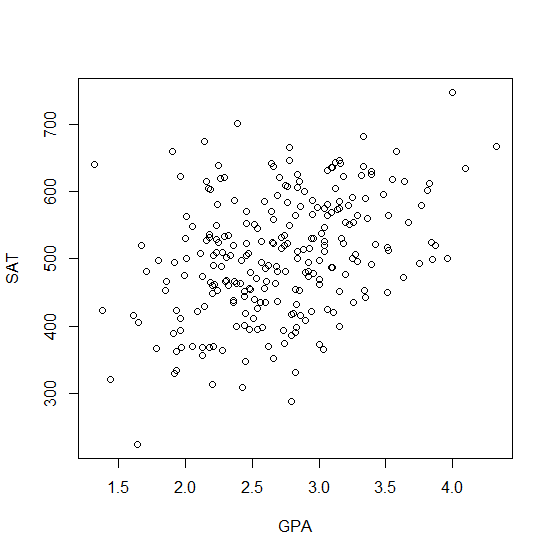
1. Why are sampling distributions so important in the process of making statistical inferences? Also give a concrete example of a statistical test to demonstrate the role of sampling distribution in hypothesis tests.

Sampling distributions are important because they inform researchers about the amount of error they can reasonably expect in their estimates of parameters. Moreover, they allow researchers to make hypothesis tests with narrow confidence intervals.

For example, we might assume that the distribution of GPA scores for students with higher than average “previous accomplishment” scores was different from the distribution of GPA scores for students as a whole. To test this, we might perform take a single sample, compute the average and sample standard deviation, and use this data to test the hypothesis that the higher-achieving students represent a different population than lower-achieving students. The average GPA for students scoring above the mean “prevach” score is 2.89 with a standard deviation of 0.53. We can compare this to the expected mean of 2.52, and as this value is easily within a standard deviation of the mean, we may be tempted to conclude that we cannot reject the null hypothesis that these two groups follow the same GPA distribution. However, we can approximate a sampling distribution by dividing the standard deviation by the square root of the sample size and conduct a two-sample t-test. This will yield an estimate of the true population average for High-Prevach students with a much smaller amount of variance. A t-test yields a t-score of -5.8, p<0.001, a highly significant result.

1. The dataset “cgpa.txt” includes three variables: SAT, GPA (College GPA), and PrevAch (previous achievement test score). Your goal is to study the relationship between SAT and college GPA for now. Note that this data set is the modified version of the one used in your first assignment.
2. Check possible outliers when examining the relationship between SAT and GPA. Be sure to detect outliers by (1) using a bivariate scatter plot and (2) looking at Leverage, Cook’s D, and Studentized residuals.

In your answers, (1) show the scatter plot and circle or highlight the potential outliers in the plot, (2) report cases with above- threshold Leverage, Cook’s D, and Studentized residuals, and (3) identify outliers and list the values of SAT and GPA for the outlier cases.



1. Remove the cases that were identified as outliers from the data set. Then conduct a regression analysis predicting GPA from SAT: Write up the results as if you are working on the Results section in a manuscript. Be sure to include the results on overall significant test of the model, the significance tests of individual model parameters, and interpret all these results (APA format is required).
2. Conduct diagnostics on model assumptions, including normality of errors, homogeneity of residuals, and independence of errors. Be sure to report both graphs and statistical tests for your diagnostics on each assumption.
3. Now you are predicting college GPA from SAT and PrevAch using the dataset of “gpa.txt” with outliers removed.
4. Descriptive statistics: report (1) the mean and standard deviation for each of the three variables, and (2) the correlation matrix for the three variables.
5. Compute the estimates for the intercept, slope of SAT, and slope of PrevAch using the formulas provided in the lecture. (show your steps)
6. Compute the standardized coefficients using the formulas provided in the lecture. (show your steps) State which variable is more important in predicting GPA and why?
7. Use a computer program to run the regression of GPA from SAT and PrevAch. Report both raw and standardized coefficients and their significance. Interpret the respective effect of SAT and PrevAch on GPA.