STAT 445 Lab 8
ASSIGNMENT # 8
Due: Lab 9 (at the end)

Question 1 (Confidence intervals for a correlation coefficient)

The data on a population of $N = 82$ law schools are available on the web ('assign81.dat'). Your objective is to examine different 80% confidence intervals (CIs) for ρ, the Pearson correlation coefficient between LSAT and GPA scores in this population of $N = 82$ law schools, based on a single simple random sample of size $n = 15$.

(a) An advantage of jackknife and the bootstrap methods is that we don't have to make parametric assumptions about the form of the joint distribution of LSAT and GPA scores in the population. Why does it not make sense to assume that the LSAT and GPA scores are bivariate normally distributed in this population of $N = 82$ law schools?

(b) Draw a simple random sample of size $n = 15$ from this population of $N = 82$ law schools. List the ID's, LSAT scores and GPA scores for the $n = 15$ law schools in your random sample and evaluate r, the sample Pearson correlation coefficient.

(c) Pretend for the moment that the LSAT and GPA scores are bivariate normally distributed in this population. In this case,

$$r \approx N\left(\rho, \frac{1}{n}(1-\rho^2)^2\right).$$

Derive the form of an approximate $1-\alpha$ CI for ρ based on this result and evaluate the 80% CI for the correlation between the LSAT and GPA scores in this population based on your random sample of $n = 15$.

(d) Under the bivariate normal assumption, the usual approximation to the distribution of Fisher's transformation of Pearson's correlation coefficient is

$$z_r = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right) \approx N\left(\frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right), \frac{1}{n-3}\right).$$

Derive the form of an approximate $1-\alpha$ CI for ρ based on this result and evaluate the 80% CI for the correlation between the LSAT and GPA scores in this population based on your random sample of $n = 15$.

(e) Use the jackknife procedure to estimate the standard errors of r and z_r for your random sample and use these to evaluate two more approximate CIs for the correlation between the LSAT and GPA scores in this population based on your random sample of $n = 15$.

(f) Generate $B = 1000$ bootstrap samples from your random sample, evaluate the corresponding estimated standard errors for r and z_r for your random sample, and use these to evaluate two more approximate CIs for the correlation between the LSAT and GPA scores in this population based on your random sample of $n = 15$.

(g) Evaluate the 10th and 90th percentiles of the bootstrapped values of r and z_r for your random sample, and use these to evaluate two more approximate CIs for the correlation between the LSAT and GPA scores in this population based on your random sample of $n = 15$.

(h) Comment on your eight approximate CIs for the correlation between the LSAT and GPA scores in this population based on your random sample of $n = 15$. (e.g. Did any fail to cover the true value of the target parameter? Do you have any reason to expect that some of these CIs might behave better than others? Etc.)