Homework 10: Due on Apr. 28

Guideline

- Homework should be submitted via Gradescope by Friday midnight (11:59 pm. CDT).
- Homework answers to Simulations and data analysis should be written in R Markdown. Please use `\newpage` in R Markdown to separate answers to different questions. For simulations, please use `set.seed(4211)` to make your result reproducible.
- Please also upload your executable "xxx.Rmd" file to the corresponding homework assignment on Canvas.
- The total score for the homework is 50 points.

1. Nonparametric bootstrap. Recall the following question from HW9.

   In a genetic inheritance study discussed by Margolin [1988], samples of individuals from several ethnic groups were taken. Blood samples were collected from each individual and several variables were measured. We shall compare the groups labeled "Native American" and "Caucasian" with respect to the variable mean sister chromatid exchange (MSCE). The data is as follows:

   Native American: 8.50 9.48 8.65 8.16 8.83 7.76 8.63 Caucasian: 8.27 8.20 8.25 8.14 9.00 8.10 7.20 8.32 7.70

   (a) Please use nonparametric bootstrap to find the variance of the sample median in each group. Please overlay the histograms of bootstrap medians for the two groups.

   (b) Please use boot::boot.ci function to report different bootstrap confidence intervals of the sample median in each group.

2. Parametric Bootstrap. Consider a sustainable fishery in the previous homework.

   The total population will only stabilize if $R = S$. If fewer recruits are produced than the number of spawners who died producing them, the total population will decline. If too many recruits are produced, the population will not be sustainable due to insufficient food. We are interested in estimating the stable point such that $R = S$.

   Based on the BH model, one can find the stable point satisfies

   $$\frac{1}{R} = \beta_0 + \beta_1 \frac{1}{R}.$$

   That is

   $$R = \frac{1 - \beta_1}{\beta_0}.$$

   Plugin the least square estimates of β_0 and β_1 to estimate the stable point.

   (a)Please use the nonparametric bootstrap to obtain a corresponding 95% confidence interval and a standard error for your estimate.

Hint: Resample pairs of $\frac{1}{R}$ and $\frac{1}{S}$ together, with replacement.

(b) Please use the parametric bootstrap to obtain a corresponding 95% confidence interval and a standard error for your estimate.

Hint: Resample residuals of linear regression, with replacement.

(c) Compare the histogram of bootstrap estimates in (a) and (b), and comment on the difference in your results.

3. Nonparametric Bootstrap.

   Efron and Tibshirani discussed the test score data on 88 students who took exams in 5 subjects: mechanics, vectors, algebra, analysis, and statistics. The first two were closed book and the last three were open book exams. The data are in "scor" from the "bootstrap" library.

   (a) Please use a panel display to display the scatter plots for each pair of test scores. You may use `PerformanceAnalytics::chart.Correlation` function.

   (b) Obtain nonparametric bootstrap estimates of the standard errors for each of the following estimates of Pearson's correlation: $\hat{\rho}_{\text{mec,vec}}$, $\hat{\rho}_{\text{alg,ana}}$, $\hat{\rho}_{\text{alg,sta}}$ $\hat{\rho}_{\text{ana,sta}}$.

   (c) Please put the histograms of the four bootstrap correlation estimates in part (b) side-by-side. Test each of the four correlations. Report the bootstrap standard error for each correlation estimator, and construct a normal bootstrap confidence interval for each correlation.

4. Continue on Bootstrap.

   The 5-dim scores data have a 5x5 covariance matrix Σ, with positive eigenvalues $\lambda_1 > \lambda_2 > \cdots > \lambda_5$. In principal components analysis (PCA),

$$\theta = \frac{\lambda_1}{\sum_{i=1}^{5} \lambda_j}$$

   measures the proportion of variance explained by the first leading principal component.

   Hint: In R, use `eigen(samplecov)$eigenvalues` to get the eigen values of the sample covariance.

   (a) Use bootstrap to estimate the standard error of $\hat{\theta}$.

   (b) Report a 95% percentile bootstrap confidence interval and BCa confidence interval for θ.