Homework 5 Key

Dr. Purna Gamage

Problem 1

Problem 5.10 #12 in Chihara/Hesterberg.

The data set FishMercury contains mercury levels (parts per million) for 30 fish caught in lakes in Minnesota.

(a) Create a histogram or boxplot of the data. What do you observe?

(b) Find the Bootstrap sampling mean and record the bootstrap standard error and the 95% bootstrap percentile interval.

(c) Remove the outlier and find bootstrap sampling mean of the remaining data. Record the bootstrap standard error and the 95% bootstrap percentile interval. Comment on your results.

(d) What effect did removing the outlier have on the bootstrap distribution, in particular, the standard error?

Problem 2

Problem 3.9 #12abc in Chihara/Hesterberg.

Two students went to a local supermarket and collected data on cereals; they classified cereals by their target consumer (children versus adults) and the placement of the cereal on the shelf (bottom, middle, and top). The data are given in *Cereals*.

(a) Create a table (Two-way) to summarize the relationship between age of target consumer and shelf location.

(b) Conduct a chi-square test using R's chisq.test() command.

(c) R returns a warning message. Compute the expected counts for each cell to see why.

Problem 3

Distribution A is a standard normal distribution and distribution B is a $N(1, 2^2)$ distribution. Generate 20 random numbers from distribution A and 30 random numbers form distribution B and record these in a suitable data frame.

Examine the null hypothesis that the means of A and B are the same against the alternative that the mean of B is larger, using a permutation test. Report the p-value and state your conclusion.

Problem 4

The dataset **NCBirths2004.csv** contains data from over 1000 births in the state of North Carolina. One of the columns contains the weight of the newborn baby in grams. Another column tells you whether the mother was a smoker (Yes or No). We want to determine whether the data contain evidence that babies born to mothers who smoke weigh less on average than babies born to non-smoking mothers.

Import the dataset, make side by side boxplots of birth weights for smoking and non-smoking mothers, formulate suitable hypotheses, carry out either a t-test or a permutation test, and state your conclusion.

Problem 5

Write an R function that computes the t-formula confidence interval in (7.8) from sample mean, sample standard deviation, sample size, and confidence level, and use it to do exercise 7.6 #6 in Chihara/Hesterberg.

Q: Julie is interested in the sugar content of vanilla ice cream. She obtains a random sample of $n = 20$ brands and finds an average of $18.05g$ with standard deviation $5g$ (per half cup serving). Assuming that the data come from a normal distribution, find a 90% confidence interval for the mean amount of sugar in a half cup serving of vanilla ice cream.

Problem 6

Exercise 7.6 #12 in Chihara/Hasterberg.

Q: Consider the data set *Girls2004* (see Case Study in Section 1.2).

(a) Create exploratory plots and compare the distribution of weights between babies born to nonsmokers and babies born to smokers.

(b) Find a 95% one-sided lower t confidence bound for the mean difference in weights between babies born to nonsmokers and smokers. Give a sentence interpreting the interval.

(c) What is your conclusion?

BONUS: Submit *ONE* of the Extra Problems:

Ex. Problem 1

Exercise 6.4 #1 in Chihara/Hesterberg.

Let X be a binomial random variable, $X \sim Binom(n, p)$. Show that the MLE of p is $\hat{p} = X/n$.

Ex. Problem 2

Exercise 6.4 #14 in Chihara/Hesterberg.

Let the five numbers 2, 3, 5, 9, 10 come from the uniform distribution on $[\alpha, \beta]$. Find the method of moments estimates of α and β.