Shannon Keenan
Project 3- Using R to analyze Statistics

Part 1: Confidence Intervals

Background: A confidence interval is a range of values used to estimate the true value of a population parameter. The goal is to construct the confidence intervals for different variables or to find the necessary sample size for given conditions.

Exercise 1: Confidence Interval for means: Large Sample

Open the data: Save the cheese file as a text document and note the file location. Type *Cheese=read.table(file="*YOUR FILE LOCATION*",header=TRUE)* into the command window in R. Type *attach(cheese)* into the command window to access the data.
note: when you type the location name you must place two backslashes between every folder (\\)

Type *taste* to view the variable taste.

Type *qqnorm(taste)* to view a normal quartile plot of the variable taste. Type *t.test(taste,conf.level=.95)* to make a confidence interval for the variable taste.

Repeat and make a confidence interval for the variable H2S

Exercise 2: Confidence interval for means: Small Sample

Open the data: Save the ex07_024 file as a text document and note the file location. Type *mpg=read.table(file="*YOUR FILE LOCATION*",header=TRUE)* Then type *attach(mpg)* to access the data.

Type *stem(MPG)* to make a stem and leaf plot of your variable. Type *hist(MPG) in the command window* to create a histogram of your variable.

Type *qqnorm(MPG)* to make a normal quartile plot of your data.

Type *t.test(MPG,conf.level=.95)* in the command window to construct a 95% confidence interval for the data.

Exercise 3: Confidence Interval for proportion

Open the data: Save the databank.xls file as a text document and note the location of the file. Type *databank= read.table(file="Y*OUR FILE LOCATION*",header=TRUE)* in the command window to open the file. Type *attach(databank)* to access your data.

We are going to do a confidence interval by hand.

Type n=length(GENDER) into the command window to assign a value to n

Type *n* to see that its value is 100.

Type *k=sum(GENDER == "F")* to assign a value to k. Type *k* to see that k is the number of females in our sample.

Type *pbar= k/n* to get a proportion for the number of females in our sample.

Type *SE= sqrt(pbar*(1-pbar)/n)* to calculate margin of error.

Type *E=qnorm(.975)*SE* to normalize the margin of error

Type *pbar + c(-E, E)* to calculate a confidence interval for the proportion of females in our sample.

Alternatively, after assigning values to k and n you can type *prop.test(k, n)* to calculate a confidence interval for the proportion of females in our sample.

Repeat excersize3 and calculate confidence intervals for the proportion of heavy smokers in our sample and for the proportion of subjects who are 25 years or younger.

Part 2: Hypothesis Testing in R

Background: Hypothesis testing refers to answering a Yes / No question about a population. In our project, this question will always be about a parameter (a numerical measure describing a population).

Exercise 1: Two Large Independent Samples

ex_1= read.table(file="C:\\Users\\Shannon\\Documents\\ex07_127.txt",header=TRUE)

Attach(ex_1)

Exercise 2: Two Small, Independent Samples

The data, listed below represent sales (in hundreds of thousands of dollars) for a random sample of fast food outlets located in two cities.

Cooperville	Davenport
6.3	2.8
6.6	6.7
7.6	5.2
3.0	
9.5	
5.9	
6.1	
5.0	
3.6	

Let's try to answer this question: Do fast food outlets do equally well in both cities, or do those in Cooperville do better than those in Davenport?

First, we need to attach the data. Type:

cooperville<- c(6.3, 6.6, 7.6, 3.0, 9.5, 5.9, 6.1, 5.0, 3.6)

davenport<- c(2.8, 6.7, 5.2)

Now, we can perform a t-test on the data to see if those in Cooperville do better than those in Davenport. Type *t.test(davenport, cooperville, var.equal=TRUE, alternative= "less")* And analyze the pvalue to make a decision on the hypothesis.

Exercise 3: Two Sample, Matched Pair

Recent research suggests it is possible for a person to control certain body functions if he or she is trained in a program of *biofeedback* exercises. An experiment is conducted to determine whether blood pressure levels can be consciously reduced in people trained in this program.

Can biofeedback be used to reduce a person's blood pressure?

Six people undergo biofeedback, and their blood pressure is measured both before and after the biofeedback. Results are as follows.

Subject	Before	After
1	136.9	130.2
2	201.4	180.7
3	166.8	149.6
4	150.0	153.2
5	173.2	162.6
6	169.3	160.1

First, we must enter the data: Type

before<- c(136.9, 201.4, 166.8, 150, 173.2, 169.3)

after<- c(130.2, 180.7, 149.6, 153.2, 162.6, 160.1)

Now we can run the test in R. Type
wilcox.test(before, after, paired=T)
in the command window to run a hypothesis test. Analyze the pvalue and make a decision about the hypothesis.

Exercise 4: One Sample

We will use variable SYSTOLIC from DATABANK.XLS.

Do the data present enough evidence to indicate that mean systolic blood pressure is above the population mean of 127? Use 0.1 level of significance.

Open the data: save the databank.xls file as a text document and not the location of the file. Type
databank= read.table(file="YOUR FILE LOCATION",header=TRUE)
to open the data. Type
attach(databank)
in the command window to access the data.

We first need to assign values to the variables used in the calculations. To do this type the following into the command window:

xbar=(mean(SYSTOLIC))
mu0=127
sigma=(sd(SYSTOLIC))
n=100

Now we will calculate the test statistic. Type

z = (xbar−mu0)/(sigma/sqrt(n))

z

Your test statistic should now appear.

We then compute the critical value at .10 significance level. Type:

alpha=.10
z.alpha = qnorm(1−alpha)
- z.alpha

Your critical value is now calculated. Use the critical value and test statistic to analyze the data and make a decision about your hypothesis.