Q SCI 482 Assignment 3 (100 points)

Q1: Sample size required for desired confidence interval (25 points)

Finches on the various islands of the Galapagos are well known for the variation in beak length among species. Beak length is highly specialized and depends on the diet of each species. In addition, beak length for any particular finch species can shift over time as drought and rainy periods cause shifts in the size of seeds available on a given island. A hybrid population has been growing on the island of Daphne Major that originated in a mating between a vagrant cactus finch and a ground finch, and this new population does not interbreed with the other species on Daphne Major. You have a representative sample of the beak lengths of 30 individuals in the file "`DaphneHybrids.csv`", that can be assumed to be normally distributed, but population standard deviation is not available. To determine if beak length is different for this population, compared to the existing species, the 95% confidence interval for mean beak length must be no wider than 3 mm. What sample size is required from the hybrid population to meet this criterion?

Q2: Sample size desired for given power (35 points)

Ocean acidification is the lowering of seawater pH due to increasing levels of carbon dioxide in the atmosphere. Many shell-forming organisms are affected by ocean acidification, because the associated chemical changes in seawater make it harder to form their shells. At Whiskey Creek oyster hatchery in Oregon, oysters are reared for their first 2 weeks of life and then sold to growers for planting in the wild and eventual commercial harvest. In recent years, they have had growth failures because of ocean acidification in the seawater used, only growing to 310 μm shell length after two weeks instead of 320 μm. A new protocol has been introduced involving adding calcium chloride and sodium carbonate to the seawater to remove the effects of ocean acidification, thus it is expected that shell length will *increase* back to 320 μm (very important note for all following questions: does this imply we need a one-tailed or two-tailed value for $t_{\alpha,v}$?). A sample of 15 shells is measured after two weeks under this new protocol (data in "`WhiskeyCreekAfter.csv`").

(a) If we want to be able to detect *an increase* in shell length from 310 to 320 μm, with a significance level of 0.05, and power of 0.90, what sample size is needed under the new protocol? [20 points]
(b) What sample size is needed for the same requirements as in (a), except with a power of 0.99? [5 points] Tip: reuse your code from part (a) and just change the relevant parameter.
(c) What sample size is needed for the same requirements as in (a), except with a significance level of 0.10? [5 points]
(d) What sample size is needed for the same requirements as in (a), if we need only detect an increase in shell length from 310 to 330 μm? [5 points]

Q3: Minimum detectable difference (20 points)

Using the data in Question 2, and the same scenario, with a significance level of 0.05, what is the minimum detectable difference in shell length, that can be detected 90% of the time, if:

(a) You take a random sample of the shell lengths of 15 baby oysters? [10 points]
(b) You take a random sample of the shell lengths of 500 baby oysters? [10 points]

Q4: Power to detect a difference (20 points)

Consider the data and scenario in Question 2, with a significance level of 0.05. It turns out the new protocol actually increases shell growth more than expected, resulting in mean shell length increasing from <u>310 to 345 µm</u>. What is the power to detect this difference if:

    (a) You take a random sample of the shell lengths of 20 baby oysters? [10 points]

    (b) You take a random sample of the shell lengths of 40 baby oysters? [10 points]

R functions and constants that might be useful for this assignment

```
ceiling()  #rounds a number up to the nearest integer
pt()       #given the tvalue and d.f., returns the area to the left (CDF)
qt()       #given CDF probability and d.f., returns value t from t-distribution
read.csv() #read in a CSV file, e.g. xdata <- read.csv(file="values.csv")
round()    #rounds to the desired number of decimal places
sd()       #returns the sample standard deviation
signif()   #rounds a number to the desired number of significant digits
var()      #returns the sample variance of values in a vector, e.g. var(vector)
```

Other useful operators and constants

```
[] #for extracting parts of a data frame, e.g. X[,10] for the 10th column and all
rows
"" #double quotation marks are needed around text values ("" not "")
```