Homework #7 – Due Monday, Nov. 27
COR1-GB.1305 – Statistics and Data Analysis

Problem 1

Here, we consider the two-sample t-test, for the data set `NormTemp.CSV`. The second column (`Gender`) is 1 for male, 2 for female, and the third column (`HeartRate`) is measured in beats per minute.

(a) Make side-by-side boxplots for the temperatures of males and females in the dataset. To do this, use *Graph* ⇒ *Boxplot* then select "One Y, With Groups". Select `Temp` in the "Graph variables" box, and select `Gender` in the "Categorical variables for grouping" box. Do there seem to be any differences?

(b) What are the two samples?

(c) What are the two populations?

(d) What are the null and alternative hypotheses?

(e) Get the descriptive statistics for the two samples. To do this, use *Stat* ⇒ *Basic Statistics* ⇒ *Display Descriptive Statistics*. Select `Temp` in the "Variables" box, and select `Gender` in the "By variables (optional)" box. Find n_1, $\bar{x}_1$, s_1, n_2, $\bar{x}_2$, and s_2.

(f) Compute the test statistic.

(g) Compute an approximate p-value.

(h) Use the p-value to evaluate whether or not there appears to be a significant difference in average temperature between males and females.

(i) Find a 95% confidence for the difference in average temperatures in the populations.

(j) Now, use Minitab to perform the test and construct the confidence interval. Do do so, use *Stat* ⇒ *Basic Statistics* ⇒ *2-Sample t*. Choose the option "Both samples are in one column". Set "Samples" to `Temp` and set "Sample IDs" to `Gender`. The p-value and confidence interval that Minitab computes will be slightly more accurate than the one you compute in part (g), because Minitab uses a t distribution instead of a z distribution to compute the probability.

. . . . . .

Problem 2

We will again use `NormTemp.CSV` data, but now we will investigate `HeartRate`.

(a) Make side-by-side boxplots for the heart rates of males and females in the dataset.

(b) Test whether or not there is a significant difference in average heart rate between all males and females. You can either compute the p-value by hand, or you can compute it using Minitab.

(c) Find a 95% confidence interval for the difference in average heart rates between all males and all females. Again, you can either compute the confidence interval by hand, or you can compute it using Minitab.

(d) What assumptions do you need for the p-value and the confidence interval to be valid?

. . . . . .

Problem 3

(Adapted from Stine and Foster, 17.32) The dataset `retail_sales.csv` gives the sales volume (in dollars per square foot) for 37 retail outlets specializing in women's clothing in 2006 and 2007. Note that these samples are paired: the same stores are measured in both years. Did sales change by a statistically significant amount from 2006 to 2007? To answer the question, use a paired t-test; that is, subtract the sales in the two years, then analyze the store-specific differences. Answer the following:

(a) What is the sample?

(b) What is the population?

(c) If there were no difference in expected sales between the two years, what would be the chance of getting data like that observed?

(d) Find a 95% confidence interval for the difference in expected sales between 2006 and 2007.

. . . . . .