Analytics Foundations: Lab 2

1. Revisit the *NormTemp* dataset from Lab 1, where we examined the observed mean body temperature (**temperature**) in comparison to the well-known "average" of 98.6.
   a. Perform a statistical test (alpha = 0.05) to determine whether this well-known number is actually the mean body temperature. What is your p-value? Explain in words what this p-value means. What is your conclusion?

We reject the null hypothesis that the mean body temperature is 98.6. The p-value is <0.0001. This means that if we assume the population mean body temperature is 98.6, the probability of obtaining a sample mean body temperature equal to or more extreme than the one we observed is <0.0001. Because that is so unlikely, we reject the assumption that the population mean temperature is 98.6

   b. Give the 95% Confidence Interval for **temperature**. Explain in words what a 95% confidence interval represents. **[98.1220, 98.3765] We can claim with 95% confidence that this interval contains the population mean.**

   c. If we restrict our analysis to only the females in this dataset, would our conclusion change? **No.**

   d. Is there any difference (alpha=0.05) in bodytemp between the two genders recorded in this dataset? **Yes, the 2 sample t-test shows evidence that there is a difference in body temperature between genders.**

2. The *Airline* dataset contains information regarding the number of international airline travelers (variable **AirPassengers**) across different months of the year from 1949-1960. Was there is a significant difference in the airline passengers between the summer months of June, July, and August and the remainder of the year?

Yes, there is a statistically significant difference in the median number of airline passengers from summer months to all other months (using the Wilcoxon test…distribution is NOT normal, but distributions appear similar and the variation appears similar).