Homework 7 answers

1. Validity of t-based conf intervals

a.- d. Included in R file.

e. As sample size n increases, the proportion of confidence intervals containing the true mean of 1 approaches 95%. This is a result of the Central Limit Theorem, which states that means based on large samples have approximately normal sampling distributions, regardless of population distribution. As sample size increases, the normality of the sample distribution increases, and the t-tools are more valid.

a. Which assumption violated? Evidence?
b. Expect robustness?
c. Suggest improvement to diminish violation of assumption?

2. Great Britain Office of Population Census and Surveys – 2-sample t-test

a. Using a two-sample t-test with data collected on husbands and wives violates the assumption of independence between samples, as the data is naturally paired: it is expected that married couples would exhibit a stronger similarity/ difference in body types than they would the greater population.

b. No, I would not expect robustness against the violation of independence.

c. Using a paired t-test would be much more appropriate here, as independence between pairs of observations is maintained by the random sampling.

3. Diet in high and low income households

a. For this study the independence of the observations is violated, as the adults in a household could constitute a sub-group, and exhibit a similarity (or difference) of a higher degree than to the remaining observations in the sample. In other words, knowing one adult's caloric intake in a household gives us some idea of the caloric intake of other adults in the same household.

b. The likelihood of dependencies among observations leads me to suspect that the t-tools would not be particularly robust; the standard error would no longer approximate the standard deviation well.

c. Given the lack of independence, t-tools might not be the best procedure, and other methods might be used. Another approach might be to redesign the study to address the clustering.

4. Oregon and gender pay gap

a. In this case the sample histograms are clearly non-normal, and indicate a high number of outliers. I would expect the standard deviation to differ somewhat between the two, as the male sample exhibits a larger number of outliers.

b. Despite the violations indicated above, I would expect the t-tool to be reasonably robust, as the sample size is quite high.

c. Increasing the sample size would diminish the effects of the violation of assumptions.