NOTES: t-distributions

Hypothesis tests and confidence intervals involving mean $\mu$ with unknown $\sigma$.

Suppose that hypothesis tests and confidence intervals involving $\mu$ are provided with a known standard deviation of the population ($\sigma$). This would enable us to calculate the standard deviation of the sample mean distribution since $SD(\bar{x}) = \frac{\sigma}{\sqrt{n}}$, where $n$ is the sample size. We now assume that $\sigma$ is not known (as is most generally the case in real life), and therefore replace $SD(\bar{x})$ with the Standard Error of the Mean. In other words, we use $SE(\bar{x}) = \frac{S_x}{\sqrt{n}}$, where $S_x$ is the standard deviation of the sample. This, however, causes a problem since the $S_x$ fluctuates with the sample size $n$ in such a way as to create a different (possible not normal) distributions for each sample size $n$. Each of these distributions is called a t-distribution.

Facts about t-distributions:
- bell-shaped curve
- mean = 0
- symmetric about 0
- Each one is different from the standard normal curve. As $n$ gets larger the t-distributions get closer to the standard normal curve. When $n \geq 30$ the t-distributions are almost normal (are very close to the standard normal curve).
- degrees of freedom $= n - 1$. The degrees of freedom (df) indicate which t-distribution you will use.

For hypothesis tests, we will use a “t-test statistic” or “t-score” instead of a z-score. The equation for the t-score is $t = \frac{\bar{x} - \mu_0}{S_x/\sqrt{n}}$, where $\mu_0$ is the assumed population mean from the null hypothesis $H_0$. You will also need to use tcdf instead of normalcdf on your calculator. $2^{\text{nd}}$ VARS tcdf. Inputs are tcdf(LB, UB, df).

For confidence intervals we will use $t^*$ instead of $z^*$. The $t^*$ value depends on the t-distribution you are using, and so depends on the degrees of freedom ($n - 1$). $t^*$ cannot be obtained from your graphing calculator and so it must be looked up on a table in your textbook. The formula for a confidence interval for means (with unknown $\sigma$) is $\bar{x} \pm t^* \frac{S_x}{\sqrt{n}}$.

Requirements for using t-distributions:
1. SRS: The sample mean must be chosen from a random sample.
2. Sufficiently large sample size:
   a. CASE: $n < 15$. The data should be very close to a Normal model. Do not use t-methods if there is strong skewness or outliers.
   b. CASE: $15 \leq n < 40$. t-methods should work as long as the data is unimodal and reasonably symmetric (make a histogram). t-methods should not be used in the presence of outliers or strong skewness.
   c. CASE: $40 \leq n$. t-methods can be used even in the presence of strong skewness or a few outliers. In this case t-methods are called “Robust.”