Sorting through the different situations for tests of significance and confidence intervals

Source: *Modified from the work of Mr. Chuck Tiberio of Wellesley High School on AP Statistics list serve*

The following assumptions and rules-of thumb should be assessed before doing any hypothesis test or constructing any confidence interval.

1. Investigating a mean, μ, with **one sample**

   Use normal distribution if
   - It is a SRS
   - sigma (population SD) is known (this almost never occurs)
   - population is known to be normal

   Use t-distribution with d.f. = n - 1 if
   - It is a SRS
   - sigma (population SD) is unknown
   - at least one of the following is true:
     - population is normal
     - sample size is large (n > 40)
     - sample size is medium (15 to 40) and there is little skewness in the sample data and no extreme outliers
     - sample size is small (n < 15) and there is <u>no</u> skewness and <u>no</u> outliers

2. Investigating a difference of means, $\mu_1 - \mu_2$, with **two dependent samples**

   Use t-distribution with df = n - 1 if
   - the samples are "matched"
   - the list of differences satisfies at least one of the following:
     - the list is large (n > 40)
     - the list is medium (15 to 40) and there is little skewness in the sample data and no extreme outliers
     - the list is small (n < 15) and there is <u>no</u> skewness and <u>no</u> outliers

3. Investigating a difference of means, $\mu_1 - \mu_2$, with **two independent samples**

Use normal distribution if

- BOTH samples satisfy all 3 conditions in part 1 above for investigating the sample mean for one sample

Use t-distribution with d.f. = Satterthwaite's formula
or df = $\min(n_1 - 1, n_2 - 1)$ if

- BOTH samples satisfy all 3 conditions in part 1 above under using a t-distribution

4. Investigating a proportion, P, with one sample

Use normal distribution if

- It is a SRS
- np > 10 and nq > 10
- n < 0.1*N

5. Investigating a difference of proportions, $P_1 - P_2$, with two samples

Use normal distribution if

- both samples are SRS
- np > 10 and nq > 10 for BOTH samples
- n < 0.1*N for BOTH samples

6. Invesigating Goodness of Fit (1 sample, 1 variable) OR Independence (1 population, 2 variables) OR Homogeneity (2 populations, 1 variable)

Use a chi-squared distribution if

- you are working with <u>counts</u>
- at least one of the following is true:
  - all expected counts are at least 5
  - all of the possible outcomes fall into a category

7. Investigating the slope, b_1, of a linear regression model

Use t-distribution with df = n - 2 (you lose 2 d.f. if you are estimating both the intercept and the slope) if

- residuals are normally distributed
- residual plot does not show a curved of fan shaped pattern (homoscedastic)
- residual plot shows little skewness and no extreme outliers