1. To make a boxplot:

   - Plot a bar at the median, and at the first and third quartiles.
   - Connect the ends of the bars to make a box with a line in it.
   - Extend whiskers out to a maximum of 1.5*IQR up from the third quartile and down from the first quartile, but only go to the largest or smallest actual data point within that range.
   - Any other data point gets a dot.

2. The Two-sample T-Test (Normal with Equal Variances)

   The data consists of separate simple random samples from two different populations, label them 1 and 2. Let:

   μ_1 = true mean of population 1
   μ_2 = true mean of population 2
   n_1 = sample size taken from population 1
   n_2 = sample size taken from population 2
   σ_1^2 = true variance of population 1
   σ_2^2 = true variance of population 2

   We wish to test:

   $H_0 : \mu_1 - \mu_2 = \delta$
   vs.
   $H_A : \mu_1 - \mu_2 \neq \delta$

   Good numerical and graphical summaries to explore the data might include means, medians, standard deviations, side-by-side boxplots, side-by-side dotplots, stacked histograms, and normal quantile plots, among others.

   If based on our prior knowledge and after exploring the data we are willing to assume:

   - All of the data points are independent, both within and between populations (this will be true if the samples are simple random samples)

   - The two populations each follow normal distributions

   - The variances of the two populations are equal so that $\sigma_1^2 = \sigma_2^2 = \sigma^2$

   Then the test statistic is:

   $$t = \frac{\bar{X}_1 - \bar{X}_2 - \delta}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Where:

$$s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$$

Compare to a t distribution on $\nu = n_1 + n_2 - 2$ degrees of freedom.

In this situation, a $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is given by:

$$\bar{x}_1 - \bar{x}_2 \pm t_{\nu,\alpha/2} * s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}.$$

Where $\nu = n_1 + n_2 - 2$.

3. The Welch T-Test (Normal with Unequal Variances)

The data consists of separate simple random samples from two different populations, label them 1 and 2. Let:

μ_1 = true mean of population 1
μ_2 = true mean of population 2
n_1 = sample size taken from population 1
n_2 = sample size taken from population 2
σ_1^2 = true variance of population 1
σ_2^2 = true variance of population 2

We wish to test:

$H_0 : \mu_1 - \mu_2 = \delta$
vs.
$H_A : \mu_1 - \mu_2 \neq \delta$

Good numerical and graphical summaries to explore the data might include means, medians, standard deviations, side-by-side boxplots, side-by-side dotplots, stacked histograms, and normal quantile plots, among others.

If based on our prior knowledge and after exploring the data we are willing to assume:

- All of the data points are independent, both within and between populations (this will be true if the samples are simple random samples)
- The two populations each follow normal distributions
- The variances of the two populations are not equal

Then the test statistic is:

$$t = \frac{\bar{X}_1 - \bar{X}_2 - \delta}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Compare to a t distribution on ν degrees of freedom, where:

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$$

In this situation, $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is given by:

$$\bar{X}_1 - \bar{X}_2 \pm t_{\nu,\alpha/2} * \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Where ν is given above.

When deciding between a variance equal and variance unequal test, remember that:

- A good guideline when the sample sizes are similar in the two groups is that if $0.5 \leq \frac{s_1}{s_2} \leq 2.0$, we can feel fairly comfortable assuming the sigmas equal. The more our samples sizes differ, the more strict we need to be in having the sigmas similar

- If the variances are truly equal, but are allowed to differ, the test loses some power, but is still a good test.

- If the variances are truly different, but they are assumed equal, the test can make wildly incorrect conclusions.

4. Comparing Two Population Proportions

The data consists of separate samples from two populations, label them 1 and 2. Let:

π_1 = true proportion in population 1
π_2 = true proportion in population 2
n_1 = sample size taken from population 1
n_2 = sample size taken from population 2

We wish to test:

$H_0 : \pi_1 - \pi_2 = 0$
vs.
$H_A : \pi_1 - \pi_2 \neq 0$

When the null is true, $\pi_1 = \pi_2 = \pi$. The unknown π can be estimated using a weighted average of the two individual sample proportions:

$$\hat{\pi} = \frac{\hat{\pi}_1 n_1 + \hat{\pi}_2 n_2}{n_1 + n_2}.$$

where $\hat{\pi}_1$ and $\hat{\pi}_2$ are the sample proportions as computed from the two samples. If based on our prior knowledge we are willing to assume:

- All of the data points are independent, both within and between populations (this will be true if the samples are simple random samples)

- The sample sizes are large enough (πn_1, $(1-\pi)n_1$, πn_2, and $(1-\pi)n_2$ are all greater than 5).

Then the test statistic is:

$$\frac{\hat{\pi}_1 - \hat{\pi}_2}{\sqrt{\hat{\pi}(1-\hat{\pi})(\frac{1}{n_1} + \frac{1}{n_2})}} \dot{\sim} N(0,1).$$

Calculate the p-value and compare to the given significance level α.