Common Inference Formulas in Statistics

Inference about Means

One Sample Hypothesis Test

$$H_0 : \mu = \mu_0 \qquad H_A : \mu \neq \mu_0 \qquad t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

One Sample Confidence Interval

$$\bar{x} \pm t^* \frac{s}{\sqrt{n}}$$

Two Sample Hypothesis Test

$$H_0 : \mu_A = \mu_B \qquad H_A : \mu_A \neq \mu_B \qquad t = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}}$$

Two Sample Confidence Interval

$$\bar{x}_A - \bar{x}_B \pm t^* \sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}$$

These formulas are robust even if the population is not normal, as long as the sample is "nice" ($n \geq 40$ usually safe; $n \geq 15$ okay for samples w/ little skew & no outliers). Two sample t-distribution methods are usually more robust than one sample methods. If the two samples are similar sizes, then you can add the sizes to evaluate robustness ($n_A + n_B \geq 15$ might be okay, $n_A + n_B \geq 40$ is better).

Inference about Proportions

One Sample Hypothesis Test

$$H_0 : p = p_0 \qquad H_A : p \neq p_0 \qquad z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

This is robust, as long as there are at least 10 expected successes and 10 expected failures in the sample. Otherwise, you might use a binomial model.

One Sample Confidence Interval

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Confidence intervals for proportions are less robust than hypothesis tests (because you are using $\hat{p}$ instead of p_0 to estimate the standard error). On the other hand, plus-4 confidence intervals are actually more robust than hypothesis tests. To make a plus-4 confidence interval, add 2 fake successes and 2 fake failures to the sample.

Two Sample Hypothesis Test

$$H_0 : p_A = p_B \qquad H_A : p_A \neq p_B \qquad z = \frac{\hat{p}_A - \hat{p}_B}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}}$$

where $\hat{p}$ is the pooled proportion. This is fairly robust, as long as their are at least 5 successes and 5 failures in each sample. Otherwise, you can use a permutation test instead.

Two Sample Confidence Interval

$$\hat{p}_A - \hat{p}_B \pm z^* \sqrt{\frac{\hat{p}_A(1-\hat{p}_A)}{n_A} + \frac{\hat{p}_B(1-\hat{p}_B)}{n_B}}$$

To make a plus-4 confidence interval for two sample proportions, add 1 success and 1 failure to each sample.

Inference about Other Statistics

Other common inference techniques include the χ^2 tests for association and goodness of fit, ANOVA F-tests, and various methods related to linear regression. Another tool is bootstrapping, which can be applied to any statistic.