HYPOTHESIS TESTS II

Video 1. Watch the introduction to this worksheet.

We start with a **null hypothesis** H_0, which should specify the distribution of a **test statistic**. Exactly what constitutes contrary evidence is determined by our choice of **alternative hypothesis** H_1. It is essential that your hypotheses not come from data that you will use to test them (hypothesis first, data second, always). Collect data, calculate the observed value of the test statistic, then calculate the **p-value** (or observed significance level) of our data. We **reject** H_0 in favor of H_1 if the p-value is small (usually less than 0.05). Otherwise we **fail to reject** H_0.

Method. For tests about the difference between the means of two populations (with unknown variances), the null hypothesis should be $\boxed{H_0 : \mu_1 - \mu_2 = \delta_0}$ and our test statistic is one of the following. We have independent random samples of sizes n_1 and n_2 with means $\overline{x}_1$ and $\overline{x}_2$ and sample variances s_1^2 and s_2^2.

A) $$\boxed{t = \frac{\overline{x}_1 - \overline{x}_2 - \delta_0}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}}$$

Uses a pooled estimator for variance: $s_p^2 = \dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$

Has a t-distribution with $n_1 + n_2 - 2$ degrees of freedom

Appropriate when **populations have the same variance**, all sample sizes if the populations are normal, otherwise large samples only

B) $$\boxed{t = \frac{\overline{x}_1 - \overline{x}_2 - \delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}}$$

Has a t-distribution with ν degrees of freedom where $\nu = \dfrac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1-1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2-1}}$

Appropriate when **populations have different variances**, all sample sizes if the populations are normal, otherwise large samples only

Both test statistics have a t-distribution and you will need to use a calculator or computer to find p-values. An example in **R**: `pt(-2.7, 49)` $= P(T \leq -2.7)$ where T has a t-distribution with 49 degrees of freedom (i.e. comes from a sample of size 50).

1. In a random sample of $n_1 = 16$ games in 2016, the Gonzaga men's basketball team had an average score of $\overline{x}_1 = 81.8750$ with a sample standard deviation of $s_1 = 10.7881$. In a random sample of $n_2 = 9$ games in 2019, the men's basketball team had a mean score of $\overline{x}_2 = 88.4444$ with a sample standard deviation of $s_2 = 8.7050$. Test the hypothesis that the mean number of points per game was the same in both years against the alternative that the mean was different. (Assume that scores for both years are normally distributed with the same variance).

Date: March 30, 2020.

Method. Hypothesis tests involving variances.

A) $\boxed{f = \dfrac{s_1^2}{s_2^2}}$ for tests of $\boxed{H_0 : \sigma_1^2 = \sigma_2^2}$

Has an F-distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom (order matters)
Appropriate only when populations are normally distributed (or samples are very large)
R CDF: `pf(`f`, `$n_1 - 1$`, `$n_2 - 1$`)`

B) $\boxed{x = \dfrac{(n-1)s^2}{\sigma_0^2}}$ for tests of $\boxed{H_0 : \sigma^2 = \sigma_0^2}$

Has a chi-square distribution with $n - 1$ degrees of freedom
Appropriate when populations are normally distributed or the sample is large
R CDF: `pchisq(`x`, `$n - 1$`)`

2. Use the basketball data from the last problem ($n_1 = 16$, $\overline{x}_1 = 81.8750$, $s_1 = 10.7881$; $n_2 = 9$, $\overline{x}_2 = 88.4444$, $s_2 = 8.7050$) to test:

a) $H_0 : \sigma_1^2/\sigma_2^2 = 1$ against $H_1 : \sigma_1^2/\sigma_2^2 \neq 1$
b) $H_0 : \sigma_1^2 = 8$ against $H_1 : \sigma_1^2 > 8$

3. Use some version of R (like RStudio Cloud)to complete R Project 4: Hypothesis tests at `http://web02.gonzaga.edu/faculty/axon/321/r-projects.html`.