Two samples are **independent** if the sample values from one population are not related to, or naturally paired with, the sample values from the second population.

Samples are **dependent** (or consist of matched pairs) if there is a natural pairing of populations (e.g., before and after, husband and wife, etc.)

NOTATION

Population 1

μ_1 = *population* mean
σ_1 = *population* standard deviation
n_1 = *sample* size
$\overline{x}_1$ = *sample* mean
s_1 = sample standard deviation

Population 2

μ_2 = *population* mean
σ_2 = *population* standard deviation
n_2 = *sample* size
$\overline{x}_2$ = *sample* mean
s_2 = sample standard deviation

REQUIREMENTS

1. The values of σ_1 and σ_2 are unknown and assumed not equal (more on this at the end)
2. The two samples are *independent*.
3. Both samples are simple random samples.
4. Both sample sizes are large ($n_1 > 30$ and $n_2 > 30$) or both come from normal distributions.

TEST STATISTIC FOR TWO MEANS

The null hypothesis is often assumed to be $H_0 : \mu_1 = \mu_2$ but not required.

The sample test statistic is: $t = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$. The "simple" method for calculating the degrees of freedom is the smaller of $df = n_1 - 1$ and $df = n_2 - 1$. This is very conservative estimate. There is a more accurate formula in the text, which is what the TI uses, so answers may differ from the text answers.

EXAMPLE 1 IQ and Lead Exposure Data Set 5 in Appendix B lists IQ scores for a random sample of subjects with low lead levels in their blood and another random sample of subjects with high lead levels in their blood. The statistics are summarized below. Use a 0.05 significance level to test the claim that the mean IQ score of people with low lead levels is higher than the mean IQ score of people with high lead levels.

Low Lead Level: $n_1 = 78$, $\overline{x}_1 = 92.88462$, $s_1 = 15.34451$
High Lead Level: $n_2 = 21$, $\overline{x}_2 = 86.90476$, $s_2 = 8.988352$

Create a 95% confidence interval for the difference in the means from the previous example.

EXAMPLE 2 Heights of Supermodels Listed below are the heights (inches) for the simple random sample of supermodels Lima, Bundchen, Ambrosio, Ebanks, Iman, Rubik, Kurkova, Kerr, Kroes, and Swanepoel. Data Set 1 in Appendix B includes the heights of a simple random sample of 40 women from the general population, and here are the statistics for those heights: n = 40, $\overline{x}$ = 63.7815 in., and s = 2.59665 in. Use a 0.01 significance level to test the claim that the supermodels have heights with a mean that is greater than the mean height of women in the general population.

70 71 69.25 68.5 69 70 71 70 70 69.5

EXAMPLE 3 A company states their test-prep program can increase a math placement score by more than 20 points. A group of 100 new students were randomly sampled into two groups. Group A with 46 students, took the placement test without the test-prep program and had a mean of 214.2 and standard deviation 22.7. Group B, with 54 students, took the two-week test-prep program and scored a mean of 241.5 with a standard deviation of 19.4. Test the companies claim that the program increased the mean score by more than 20 points. Note: the claim is of the form $\mu_1 - \mu_2 > 20$ (assuming the test-prep sample mean is $\overline{x}_1$).

Note: Under certain conditions we might be able to assume the two populations have the same standard deviation (or more precisely, the same variance). In these cases we can *pool* the sample variances which gives a higher degree of freedom of $df = n_1 + n_2 - 2$. The calculation of the test statistic is much more complicated. See the text for the formula.