Lesson 11-2 Comparing Two Means

# Two-sample problems

The goal of inference is to compare the responses to two treatments or to compare two populations. We have separate samples from each treatment or each population.

# Assumptions for comparing two means

We have two SRSs **from two distinct populations.** The samples are **independent**. Matching violates independence. We measure the same variable for both samples.

Both populations are normally distributed. Both means and standard deviations of the populations are unknown.

# Two-sample t statistic



The statistic *t* has the same interpretation as any *z* or *t* statistic. It says how far Ë1 – Ë2 is from its mean in standard deviation units. Unfortunately, a two-sample *t* statistic does not have a *t* distribution. We have two options:

Option 1: Use procedures based on the statistic *t* with degrees of freedom from the data. The degrees of freedom are generally not a whole number. This is a very accurate approximation to the distribution of *t*.

Option 2: Use procedures based on the statistic *t* with critical values from the *t* distribution with degrees of freedom equal to the smaller of n1 – 1 and n2 – 1. These procedures are always conservative for any two normal populations. The true P-values or critical values will always be equal to or less than the value calculated from *t(k).*

# Using the two sample t-procedures

Except in the case of small samples, the assumption that the data are an SRS from the population of interest is more important than the assumption that the population distribution is normal.

* ***Sum of the sample size less than 15***. Use *t* procedures if the data are close to normal. If the data are clearly non-normal or if outliers are present, do not use *t* procedures.
* ***Sum of the sample size at least 15.*** The *t* procedures can be used except in the presence of outliers or strong skewness.
* ***Large samples.*** The *t* procedures can be used even for clearly skewed distribution when the sample is large, roughly n1 + n2 ≥ 40.

When planning a two-sample study, it is best to choose to equal sample sizes whenever possible.

We would like to compare the mean fill of 16 ounce cans of soda from two adjacent filling machines. Past experience has shown that the population standard deviations of fills for the two machines are known to be s1 = 0.043 and s2 = 0.052 respectively.

A sample of 35 cans from machine 1 gave a mean of 16.031 and a sample of 31 cans from machine 2 gave a mean of 16.009. State, perform and interpret an appropriate hypothesis test using the 0.05 level of significance.

In an attempt to determine if two competing brands of cold medicine contain, on the average, the same amount of acetaminophen, twelve different tablets from each of the two competing brands were randomly selected and tested for the amount of acetaminophen each contains. The results (in milligrams) follow. Use a significance level of 0.01.

Brand A Brand B

517, 495, 503, 491 493, 508, 513, 521

503, 493, 505, 495 541, 533, 500, 515

498, 481, 499, 494 536, 498, 515, 515

State and perform an appropriate hypothesis test.