- Chapter 2 will cover a variety of techniques, but none that will set the tone as much as the permutation test which we will discuss immediately.
- Many tests and methods over the next few chapters will be based at least in part on this test.

- **Example:** New versus traditional instructional methods
- Suppose a company is deciding whether to switch how it trains its employees for a certain task

- **Example:** What if we perform a *t*-test?
- We find a test statistic value of $t = 2.08$ and a

WHY CAN'T WE JUST PERFORM A T-TEST?

- What have we assumed here though?

1. Independent observations (reasonable)
2. Populations from which the observations are drawn are normally distributed (Maybe, but we don't know).
3. Variances of the two populations are the same (Maybe, but we don't know).

- If these assumptions do not hold, our *t*-test may suggest a significant difference where there is not one.

The Permutation Test

- Let us assume just however that experimental subjects are assigned randomly to one of the two treatments:

Four subjects to the new method (treatment = assignment of test)
Three subjects to the old method

- How many possible ways to assign seven data points?:

$$\binom{7}{4} = \frac{7!}{4!3!} = 35$$

- If the two methods are equally effective in educating the employees, we would expect both low and high scores in both groups. As such, we would also expect similar means of the scores in each group.

- If however, the new method were more effective than the traditional method, we would expect the larger data values to be in the new method group and the smaller data values.

- **Based on these two thoughts, we can use the difference in the means of the two group scores as a test statistic**

- We can see from Table 2.1.2 that the observed set of data is identified second from the top with an asterisk.
- It has a mean difference of 16.2
- If we think about this in terms of probabilities, the *p*-value for an upper-tail test is the probability of observing a difference of means of 16.2 or greater under the assumption that the treatments do not differ.

To put it another way, under the assumption that all the different permutations* are equally likely to occur. (Independent observations)

- This procedure is called a two-sample **permutation test**.
- The distribution of the 35 differences of means is called the *permutation distribution* for the difference between the two means.

- Instead of the difference between means, we may equivalently use the sum of the observations of one of the two methods, in this case the new method, as our test statistic.
- Example: The observed data as a new method sum of 198, which is the second largest of the 35 sums.

p-value = 2/35 = .0571

A comment about using the sum of treatment 1 observations as opposed to the difference of means. Consider the following:

$$D = \frac{T_1}{m} - \frac{T_2}{n} = \frac{T_1}{m} - \frac{T - T_1}{n} = T_1\left(\frac{1}{m} + \frac{1}{n}\right) - \frac{T}{n}$$

where T_1 is the sum of observations in treatment 1, T_2 is the sum of observations in treatment 2, *T* is the total sum of observations, and *D, m,* and *n* are defined as before.
Note that *D* and T_1 have a one-to-one correspondence (*D* can be determined from T_1 and vice versa).
What does that mean about a test performed based on *D* versus a test based on T_1?

So what? What can we say? About whom or what can we say it?
Don't have enough evidence that they will not be equal, fail to reject the null hypothesis.

1. Randomly assign experimental units (subjects or objects) to one of two treatments with *m* units assigned to treatment 1 and *n* units assigned to treatment 2.
2. Obtain data on the units and compute the difference between the two means, D_obs

In our example, we randomly assign 4 subjects to the new method and 3 subjects to the traditional method.
We then observe their data values and compute the difference between the two means (16.2)

What test should you apply? How? **TEST**

5. If the presumed effect of treatment 1 is to produce observations that are on average larger than those for treatment 2 (upper-tail test), compute the p-value as the proportion of *D*'s greater than or equal to D_obs:

$$P_{\text{upper tail}} = \frac{\text{number of } D\text{'s} \geq D_{obs}}{\binom{m+n}{m}}$$

Lab 2 due Thurs., in class Tues., if not Friday
Lab 3 in class Thurs,