Lecture 4

(Sampling basics and Hypothesis test)

This week we turn from *descriptive* statistics to *inferential* statistics and making decisions about our populations based on the samples we have. For example, our class case research question is really asking if in the entire company population of employees, do males and females receive the same pay for doing equal work. However, we are not analyzing the entire *population*, instead we have a *sample* of 25 males and 25 females to work with.

This brings us to the idea of sampling – taking a small group/sample from a larger population. To paraphrase, not all samples are created equal. For example, if you wanted to study religious feelings in the United States, would you only sample those leaving a fundamentalist church on a Wednesday? While this is a legitimate element of US religions, it does not represent the entire range of religious views – it is representative of only a portion of the US population, and not the entire population.

The key to ensuring that sample descriptive statistics can be used as inferential statistics – sample results that can be used to infer the characteristics (AKA parameters) of a population – is have a random sample of the entire population. A **random sample** is one where, at the start, everyone in the population has the same chance of being selected. There are numerous ways to design a random sampling process, but these are more of a research class concern than a statistical class issue. For now, we just need to make certain that the samples we use are randomly selected rather than selected with an intent of ensuring desired outcomes are achieved.

The issue about using samples that students often new to statistics is that the sample statistic values/outcomes will rarely be exactly the same as the population parameters we are trying to estimate. We will have, for each sample, some *sampling error*, the difference between the actual and the sample result. Researchers feel that this sampling error is generally small enough to use the data to make decisions about the population (Lind, Marchel, & Wathen, 2008).

While we cannot tell for any given sample exactly what this difference is, we can estimate the maximum amount of the error. Later, we will look at doing this; for now, we just need to know that this error is incorporated into the statistical test outcomes that we will be studying.

Once we have our random sample (and we will assume that our class equal pay case sample was selected randomly), we can start with our analysis. After developing the descriptive statistics, we start to ask questions about them. In examining a data set, we need to not only identify if important differences exist or not but also to identify reasons differences might exist. For our equal pay question, it would be legal to pay males and females different salaries if, for example, one gender performed the duties better, or had more required education, or have more seniority, etc. Equal pay for equal work, as we are beginning to see, is more complex than a simple single question about salary equality. As we go thru the class, we will be able to answer

increasingly more complex questions. For this week, we will stay with questions about involving ways to sort our salary results – looking for differences might exist.

Some of these questions for this week with our equal pay case could include:

- Could the means for both males and females be the same, and the observed difference be due to sampling error only?
- Could the variances for the males and female be the same (AKA statistically equal)?
- Could salaries per grade be statistically equal?
- Could salaries per degree (undergraduate and graduate) be the same?
- Etc.

Hypothesis Testing

As we might expect, research and statistics have a set procedure/process on how to go about answering these questions. The hypothesis testing procedure is designed to ensure that data is analyzed in a consistent and recognized fashion so everyone can accept the outcome.

Statistical tests focus on differences – is this difference large enough to be significant, that is not simply a sampling error? If so, we say the difference is statistically significant; if not, the difference is not considered statistically significant. This phrasing is important as it is easy to measure a difference from some point, it is much harder to measure "things are different." It is that pesky sampling error that interferes with assessing differences directly.

Before starting the hypothesis test, we need to have a clear research question. The questions above are good examples, as each clearly asks if some comparison is statistically equal or not. Once we have a clear question – and a randomly drawn sample – we can start the hypothesis testing procedure. The procedure itself has five steps:

- Step 1: State the null and alternate hypothesis
- Step 2: Form the decision rule
- Step 3: Select the appropriate statistical test
- Step 4: Perform the analysis
- Step 5: Make the decision, and translate the outcome into an answer to the initial research question.

Step 1. The null hypothesis is the "testable" claim about the relationship between the variables. It **always** makes the claim of *no* difference exists in the populations. For the question of male and female salary equality, it would be: *Ho: Male mean salary = Female mean salary.* If this claim is found not to be correct, then we would accept the alternate hypothesis claim: *Ha: Male salary mean =/= (not equal) Female salary mean.* (Note, some alternate ways of phrasing these exist, and we will cover them shortly. For now, let's just go with this format.)

Step 2. This step involves selecting the decision rule for rejecting the null hypothesis claim. This will be constant for our class – we will reject the null hypothesis when the *p-value* is equal to or less than 0.05 (this probability is called *alpha*). Other common values are .1, and .01 – the more severe the consequences of being wrong if we reject the null, the smaller the value of

alpha we select. Recall that we defined the p-value last week as the probability of exceeding a value, the value in this case would be the statistical outcome from our test.

Step 3. Selecting the appropriate statistical test is the next step. We start with a question about mean equality, so we will be using the **T-test** – the most appropriate test to determine if two population means are equal based upon sample results.

Step 4. Performing the analysis comes next. Fortunately for us, we can do all the arithmetic involved with Excel. We will go over how to select and run the appropriate T-test below.

Step 5. Interpret the test results, making a decision on rejecting or not rejecting the null hypothesis, and using this outcome to answer the research question is the final step. Excel output tables provide all the information we need to make our decision in this step.

Step 1: Setting up the hypothesis statements

In setting up a hypothesis test for looking at the male and female means, there are actually three questions we could ask and associated hypothesis statements in step 1.

1. Are male and female mean salaries equal?
    a. Ho: Male mean salary = Female mean salary
    b. Ha: Male mean salary =/= Female mean salary
2. Is the male mean salary equal to or greater than the Female mean salary?
    a. Ho: Male mean salary => Female mean salary
    b. Ha: Male mean salary < Female mean salary
3. Is the male salary equal to or less than the female mean salary?
    a. Ho: Male mean salary <= Female mean salary
    b. Ha: Male mean salary > Female mean salary

While they appear similar each answers a different question. We cannot, for example, take the first question, determine the means are not equal and then say that, for example, the male mean is greater than the female mean because the sample results show this. Our statistical test did not test for this condition. If we are interested in a directional difference, we need to use a directional set of hypothesis statements as shown in statements 2 and 3 above.

Rules. There are several rules or guidelines in developing the hypothesis statements for any statistical test.

1. The variables must be listed in the same order in both claims.
2. The null hypothesis must always contain the equal (=) sign.
3. The null can contain an equal (=), equal to or less than (<=) or equal to or greater than (=>) claim.
4. The null and alternate hypothesis statement must, between them, account for all possible actual comparisons outcomes. So, if the null has the equal (=) claim, the alternate must contain the not equal (=/= or ≠) statement. If the null has the equal or less than (<= or ≤) claim, the alternate must contain the greater than (>) claim.

Finally, if the null has the equal to or greater (=> or ≥) claim, the null must contain the less than (<) claim.

Deciding which pair of statements to use depends on the research question being asked – which is why we always start with the question. Look at the research question being asked; does it contain words indicating a simple equality (means are equal, the same, etc.) or inequality (not equal, different, etc.), if so we have the first example Ho: variable 1 mean = variable 2 mean, Ha: variable 1 mean =/= variable 2 mean.

If the research question implies a directional difference (larger, greater, exceeds, increased, etc. or smaller, less than, reduced, etc.) then it is often easier to use the question to frame the alternate hypothesis and back into the null. For example, the question is the male mean salary greater than the female mean salary would lead to an alternate of exactly what was said (Ha: Male salary mean > Female salary mean) and the opposite null (Male salary mean <= Female salary mean).

Step 2: Decision Rule

Once we have our hypothesis statements, we move on to deciding the level of evidence that will cause us to reject the null hypothesis. Note, we always test the *null* hypothesis, since that is where our claim of equality lies. And, our decision is either *reject the null* or *fail to reject the null*. If the latter, we are saying that the alternate hypothesis statement is the more accurate description of the relationship between the two variable population means. We never accept the alternate.

When we perform a statistical test; we are in essence asking if, based on the evidence we have is, the difference we observe be large enough to have been caused by something other than chance or is it due to sampling error?

A statistical test gives us a statistic as a result. We know the shape of the statistical distribution for each type of test, therefore we can easily find the probability of exceeding this test value. Remember we called this the p-value.

Now all we need to decide is what is an acceptable level of chance – that is, when would the outcome be so rare that we would not expect to see it purely by chance sampling error alone? Most researchers agree that if the p-value is 5% (.05) or less than, then chance is not the cause of the observed difference, something else must be responsible. This decision point is called alpha. Other values of alpha frequently used are 10% (often used in marketing tests) and 1% (frequently used in medical studies). The smaller the chosen alpha is, the more serious the error is in rejecting the null when we should not have.

For our analysis, we will use an alpha of .05 for all our tests.

Final Point

You may have noticed that we have two basic types of hypothesis statements – those testing equality and those testing directional differences. This leads to two different types of statistical tests – the two-tail and the one-tail. In the one-tail test, the entire value of alpha is

focused on the distribution tail – either the right or left tail depending upon the phrasing of the alternate hypothesis. A neat hint, the arrow head in the alternate hypothesis shows which tail the result needs to be in to reject the null.

In the case of the two-tail test (equality), we do not care if one variable is bigger or smaller than the other, only that they differ. This means that the rejection statistic could be in either tail, the right or left. Since the reject region is split into two areas, we need to split alpha into these areas – so with a two-tail test, we use alpha/2 as the comparison with our p-value (e.g., $0.05/2 = 0.025$). The example in Lecture 5 will review this in more detail.

References

Lind, D. A., Marchel, W. G., & Wathen, S. A. (2008). *Statistical Techniques in Business & Finance.* (13th Ed.) Boston: McGraw-Hill Irwin.