POS 3713 Spring 2000
Notes for Healey, Chapter 9
Dr. Mitchell

We will not have time to cover two sample hypothesis testing (Healey, Chapter 9). This handout will serve as a summary of the material covered in that chapter. You will not be tested over this material, but it will be useful should you decide to do all three parts of Lab Assignment #4 (for extra credit).

In Chapter 8 we dealt with hypothesis testing in the one-sample case. In that situation, our concern was with the significance of the difference between a sample value and a population value (such as the difference between people who went through an alcohol treatment program and the community at large with respect to their absentee rates at work). In this chapter, we consider the significance of the difference between two separate populations. For example, do men and women in the United States vary in their support for gun control? We cannot ask every male and female for their opinions on this issue, thus we must draw random samples and then use this information to infer population patterns.

Central Question: Is the difference between the samples large enough to allow us to conclude that the populations represented by the samples are different?

Hypothesis Testing with Sample Means (Large Samples)

- In the one sample case, we assumed that the sample was collected randomly. In the two-sample case, we must assume that the samples are collected randomly and that they are *independent*. Independence means that the chances one case will be selected in one sample is unaffected by the selection of cases in the other sample (our choice of men does not depend on what women we sample). Often we collect a single random sample, and then divide it into two subgroups (such as dividing the National Election Study into male and female sub-samples).
- Step 1: Making Assumptions
  We assume independent random samples, interval level of measurement, and a normal sampling distribution (Z).
- Step 2: Stating the Null hypothesis
  The null hypothesis states that the populations represented by the samples are not different on this variable (such as support for gun control).

  $$H_o: \mu_1 = \mu_2$$
  $$H_1: \mu_1 \neq \mu_2$$

  Note that this example is a two-tailed test. Also, you can see that this is similar to ANOVA, but we are comparing 2 samples (whereas we are comparing more than 2 samples in ANOVA).

- Step 3: Selecting the Sampling Distribution and Establishing the Critical Region
  The sampling distribution is normal, so we should use the Z scores in Appendix A. Assuming $\alpha = .05$, then Z (critical) = +/- 1.96.

- Step 4: Computing the Test Statistic

  $$Z\ (\text{obtained}) = \frac{X_1 - X_2}{\sigma_{X - X}}$$

where $X_1$ = mean for sample 1 (men)
$X_2$ = mean for sample 2 (women)
$\sigma_{X-X}$ = the standard deviation of the sampling distribution of the differences in sample means (see page 205 for formula)

- Step 5: Making a decision
  Because this is a two-tailed test, if our calculated value of Z (obtained) is > 1.96 or < -1.96 (in the reject region), then we can reject the null hypothesis and conclude that there is a significant difference between men and women with respect to gun control attitudes.

<u>Example</u>: Suppose we collect the following data in our support for gun control example. We want to know if there is a significant difference between male and female support for gun control.

| Sample 1 (men) | Sample 2 (women) |
|---|---|
| $X_1 = 6.2$ | $X_2 = 6.5$ |
| $s_1 = 1.3$ | $s_2 = 1.4$ |
| $N_1 = 324$ | $N_2 = 317$ |

$H_o: \mu_1 = \mu_2$
$H_1: \mu_1 \neq \mu_2$

$\sigma_{X-X} = 0.107$ (see page 206 for calculation)

$$Z \text{ (obtained)} = \frac{X_1 - X_2}{\sigma_{X-X}} = \frac{6.2 - 6.5}{0.107} = -0.300/0.107 = -2.80$$

Since -2.80 < -1.96, we can reject $H_o$ and conclude that there is a significant difference between male and female support for gun control. The decision to reject the null hypothesis has only a 0.05 probability (the alpha level) of being incorrect.

<u>Hypothesis Testing with Sample Means (Small Samples)</u>

- As with single-sample means, when the population standard deviation (σ) is unknown and sample size is small (combined *N*s of less than 100), the Z distribution can no longer be used. Instead we must use the t-distribution to find the critical region.
- Df = $N_1 + N_2 - 2$, where $N_1$ = number of cases in sample 1, and $N_2$ = number of cases in sample 2.
- We must assume that the population variances are equal to conduct this test ($\sigma_1^2 = \sigma_2^2$).

<u>Example</u>: A researcher believes that center-city families are significantly larger than suburban families, as measured by number of children. Random samples from both areas are gathered and sample statistics computed.

| Sample 1 (suburban) | Sample 2 (center-city) |
|---|---|
| $X_1 = 2.37$ | $X_2 = 2.78$ |
| $s_1 = 0.63$ | $s_2 = 0.95$ |
| $N_1 = 42$ | $N_2 = 37$ |

In this case, the test is one-tailed because the researcher has specified the direction of the difference (center-city families are larger).

$H_o: \mu_1 = \mu_2$
$H_1: \mu_1 < \mu_2$

Critical value of t: $df = N_1 + N_2 - 2 = 42 + 37 - 2 = 77$ (For $\alpha = .05$, one tailed test); Critical t = -1.671 (Note that the critical region is placed in the lower tail of the sampling distribution in accordance with the direction specified in the alternative hypothesis.)

$$t \text{ (obtained)} = \frac{X_1 - X_2}{\sigma_{X - X}}$$

See page 208 for calculation of $\sigma_{X - X}$. In this example, $\sigma_{X - X} = .19$

$$t \text{ (obtained)} = \frac{2.37 - 2.78}{0.19} = -2.16$$

Decision: because the obtained t < critical t (falling in the reject region), we can reject the null hypothesis and conclude that there is a significant difference in the sizes of suburban and center-city families, with suburban families being smaller on average than center-city families.

Hypothesis Testing with Sample Means (Matched Samples)

The purpose of hypothesis testing with sample means is to compare the mean level of a variable between two groups. In the case of matched samples, we cannot assume that the two groups we have randomly selected are independent. For example, we might want to determine if men and women who are married have different attitudes about gun control. We cannot assume that their attitudes are independent because a husband's views on gun control may influence his wife's views (and vice versa). Suppose that we collect the following data for 5 married couples (assume that we measure each respondent's attitudes about gun control on an 0-100 interval scale).

| Couple # | Husband | Wife | Difference (D = Husband - Wife) |
|---|---|---|---|
| 1 | 75 | 60 | 75 - 60 = 15 |
| 2 | 90 | 85 | 90 - 85 = 5 |
| 3 | 60 | 65 | 60 - 65 = -5 |
| 4 | 30 | 20 | 30 - 20 = 10 |
| 5 | 45 | 42 | 45 - 42 = 3 |

We can see that in most cases, women score lower than their husband's on the support for gun control scale. The hypothesis test will tell us if this difference is large enough to justify the conclusion that it did not occur by random chance alone, but rather reflects an actual difference between husbands and their wives on this issue. The null and alternative hypotheses for this test are (we could set this up as a one-tailed test also if we thought a priori that women would be less supportive):

$$H_o: \mu_D = 0$$
$$H_1: \mu_D \neq 0$$

$\mu_D$ represents the mean for the difference between these matched samples. If there is no difference between the two groups, the mean difference will be zero (under the null). We calculate the difference in the sample as follows.

$X_D = \Sigma D/N = (15 + 5 + -5 + 10 + 3)/5 = 28/5 = 5.6$

$s_D = \sqrt{[(\Sigma D^2 - ((\Sigma D)^2/N))/N]} = \sqrt{[((15^2 + 5^2 + -5^2 + 10^2 + 3^2) - ((28)^2)/5))/5]} = \sqrt{[((384 - (784/5))/5]} = \sqrt{45.44} = 6.74$

We use t because N is small (< 100) and $\sigma_D$ is unknown. The degrees of freedom (df) for the test are N - 1 = 5 - 1 = 4. The critical t for $\alpha$ = .05, two-tailed test = +/- 2.776.

t (obtained) = $X_D/(s_D\sqrt{(N-1)}) = 5.6/(6.74/\sqrt{4}) = 1.66$

Decision: Because our calculated t is less than the critical t (2.776), we must accept the null hypothesis and conclude that there is no significant difference in attitudes about gun control when comparing husbands and their wives.

I am not going to go through the example for sample proportion (pages 213-215), but you should be aware that we can conduct this type of test for variables that are measured nominally.

<u>Summary</u>

For all hypothesis tests, the probability of rejecting the null hypothesis is a function of four independent factors:

1) The size of the observed differences (something the researcher cannot control)
2) The alpha level (the higher the alpha level, the larger the critical region, and hence the easier it becomes to reject the null hypothesis)
3) The use of one or two tailed tests (one tailed tests make rejection of the null hypothesis more likely)
4) The size of the sample (the larger the sample, the higher the probability of rejecting the null hypothesis; we need to think about how important the relationship is substantively when dealing with really large samples)

# Calculating the Index of Qualitative Variation (IQV)

| Neighborhood A | | Neighborhood B | | Neighborhood C | |
|---|---|---|---|---|---|
| Race | Frequency | Race | Frequency | Race | Frequency |
| White | 90 | White | 60 | White | 30 |
| Black | 0 | Black | 20 | Black | 30 |
| Other | 0 | Other | 10 | Other | 30 |
| | N=90 | | N=90 | | N=90 |

We calculate the IQV as follows: $IQV = \frac{k\,(N^2 - \Sigma f^2)}{N^2\,(k - 1)}$

Where $k$ = the number of categories (white, black, other)
N = number of cases (90)
$\Sigma f^2$ = the sum of squared frequencies

Finding the sum of the squared frequencies:

| | Neighborhood A | |
|---|---|---|
| Race | Frequency | Squared Frequency |
| White | 90 | 8100 |
| Black | 0 | 0 |
| Other | 0 | 0 |
| | N=90 | $\Sigma f^2 = 8100$ |

Thus for Neighborhood A, IQV = $\frac{3\,(90^2 - 8100)}{90^2\,(3-1)}$ = 3(8100-8100)/16200 = 0

| | Neighborhood B | |
|---|---|---|
| Race | Frequency | Squared Frequency |
| White | 60 | 3600 |
| Black | 20 | 400 |
| Other | 10 | 100 |
| | N=90 | $\Sigma f^2 = 4100$ |

Thus for Neighborhood B, IQV = $\frac{3\,(90^2 - 4100)}{90^2\,(3-1)}$ = 12000/16200 = .74

For Neighborhood C, IQV = 1.

This means that neighborhood C has the most variation in terms of racial makeup, whereas neighborhood A is the least variable.

## POS 3713 Healey Chapter 6, Sampling

The basic problem for social scientists is that we generally do not have enough time and money to collect all of the population data. Thus we collect random samples, and we try to make inferences about the population from the sample. It is important to collect random samples so that the data collected is representative of the larger population.

Collecting sample data produces a dilemma because we know a great deal about the sample distribution (based on the mean, variance, etc.), but we know virtually nothing about the population distribution. And the population distribution is what we are interested in learning about. The sample distribution is interesting only insofar as it allows the researcher to generalize to the population.

Generally, the information necessary to characterize the distribution of a variable would include:

- The shape of the distribution
- Some measure of central tendency
- Some measure of dispersion

Clearly all three can be computed on any variable for the sample. But none of the information can be gathered for the population (except in rare cases where we know things are distributed normally like IQ and height).

In statistics, a device known as the sampling distribution bridges this vast ocean of ignorance. Although we might be ignorant about the population distribution, the characteristics of the sampling distribution are based on the laws of probability, and are very well known.

The general strategy of *inferential statistics* is to move from the sample to the population via the sampling distribution. Three distributions are involved in every application of inferential statistics:

1) The sample distribution: this is empirical and known (what we have collected)
2) The population distribution: which, while empirical (it is exists in reality), is unknown.
3) The sampling distribution: which is non-empirical (theoretical); the laws of probability allow us to use sampling distributions to make inferences about the population from the sample.

Tests of statistical significance

Recall from our discussion of sampling error that there is always the chance that the sample we have drawn is not representative of the population even though it was collected randomly. For example, even though it is unlikely, there is a small chance that in a sample of 100 coin flips, that the coin will turn up heads every time. If the coin is fair, then this sample is not representative of the proportion of heads we would expect (50%).

We talked about various levels of confidence, such as 95% and 99%. 95% confidence means that we are accepting the possibility of error 5% of the time, or in other words, we only expect 95 of 100 samples collected to be representative. These values, .05 and .01 or 5% and 1%, represent the likelihood that any generalization from our sample to the larger population is simply wrong.

Tests of statistical significance perform the same function in evaluating sample statistics, such as measures of association. They tell us just how likely it is that the association we have measured between two variables in a sample might or might not exist in the whole population.

For example, we will be learning about coefficients of association (correlation, lambda, gamma), which are numbers that summarize the amount of improvement in guessing values on one variable for any case based on knowledge of the values of a second. In general, coefficients of association range from 0 to 1 or -1 to 1, with values close to +/- 1 indicating a strong association and those closest to 0 indicating a weak association.

<u>Example</u>: suppose we wanted to examine the relationship between the size of a country's population and the proportion of its adults who are college educated. We might come up with one of three hypotheses:

1) Larger countries generally have a greater proportion of college educated adults than smaller ones. (positive relationship)
2) Smaller countries generally have a greater proportion of college educated adults than larger ones. (negative relationship)
3) There is no systematic difference between the two (null hypothesis). In other words, having knowledge of a country's size will not improve our ability to predict the proportion of adults who are college educated.

Suppose we have a population of 200 nations for which we know for a fact that the coefficient of association between population size and the proportion of adults with a college education is 0. Thus, there is no relationship between these variables in reality. Now suppose we cannot collect information on the entire population, but instead draw a random sample of 30 countries. The question: how can we be sure that the sample we have drawn is representative of this population?

Problem: in the real world we seldom know the underlying population parameter (zero in this example). Thus we are usually forced to sample, and often times we only draw a single sample due to expenses. The question thus becomes one of how confident we can be that a test of association based on a single subgroup of a population accurately reflects an underlying population characteristic. The job of the test of statistical significance is to pin a number on that confidence, that is, to measure the probability or likelihood that we are making an appropriate, or conversely, an inappropriate, generalization.

<u>Sampling Distributions</u>

If we draw several samples from the population, then we would expect to get different sample statistics.

1. Example: Suppose we draw 100 or 1000 separate and independent samples of 30 countries from our population of 200 nations. Because the true coefficient for the entire population is zero, then most of the coefficients will also be at or relatively near zero. But some might yield higher values.
2. Example: The proportion of Republicans in the U.S. (Figure 6-7). In this case, we know that the population proportion is .6, and yet the exact proportion produced in a variety of samples varies from .53 to .64. Even though these are not equal to .6, they are close.
3. What if we drew an infinite number of samples from the population? We would get the *sampling distribution*.

**Sampling Distribution**: A theoretical probability distribution of the possible values of some sample statistic that would occur if we were to draw all possible samples of a fixed size from a given population. A sampling distribution includes all possible values that a statistic, such as a sample mean, can assume for a given sample size.

The sampling distribution is theoretical, which means that it is never obtained in reality by the researcher. The proportion of Republicans example demonstrates how it would be obtained in principle. We would keep collecting different samples of the same size, and then construct a distribution of the various proportions or means. Some values will be higher than the population proportion, some will be lower, but on average, we would expect them to cluster around the true population value.

To illustrate further, suppose that the true mean age of the population is 30. Most of the samples we collect will also be approximately 30, thus the sampling distribution of these sample outcomes (Xbars) should peak at 30. Some of the sample means will miss the mark, but the frequency of such misses should decline as we get further away from 30. We should get something that looks normal if the samples are representative (Figure 6.2).

If repeated random samples of size N are drawn from a normal population with mean $\mu$ and standard deviation $\sigma$, then the sampling distribution of sample means will be normal with a mean $\mu$ and a standard deviation of $\sigma/\sqrt{N}$.

In other words, if we begin with a trait that is normally distributed across a population (like IQ), and take an infinite number of equally sized random samples from that population, then the sampling distribution of sample means will be normal.

This also says that the mean of the sampling distribution will be exactly the same value as the mean of the population. As for dispersion, this states that the standard deviation of the sampling distribution, also called the *standard error*, will be equal to the standard deviation of the population divided by the square root of N. This means that the larger the sample size, the smaller the standard error for the sampling distribution.

Central Limit Theorem: Fortunately, we can generalize this result to any parent population, and this is called the central limit theorem. If repeated random samples of size N are drawn from any population, with mean $\mu$ and standard deviation $\sigma$, then as N becomes large (100 or more), the sampling distribution of sample means will approach normality, with mean $\mu$ and standard deviation $\sigma/\sqrt{N}$.

This is an extremely important result, because it will allow us to use normal sampling distributions to test hypotheses about the sample mean and other statistics.

Why are Sampling distributions important?
1. Because once you are able to describe the sampling distribution of any statistics (like mean, median, s.d), you are in a position to test a wide variety of hypotheses.
2. Whenever we estimate a population parameter from a sample, we ask such questions as "How good an estimate do I have?" Can I conclude that the population parameter is identical with the sample statistic? Or is there likely to be some error? If so, how much?
3. To answer each of these questions, we compare our sample results with the "expected" results. The expected results are in turn given by the appropriate sampling distribution. Fortunately, there are known and predictable relationships between the form of the sampling distributions we consider and the sample size of the statistics on which they are based.
4. As the sample size (N) of a sampling distribution increases:
   - a. the dispersion of the sample statistics becomes less, that is, the sample statistics tend to cluster closer to the parameter of interest (you can see this in Figure 6.3), and
   - b. the form of the distribution becomes increasingly symmetrical and bell-shaped (i.e., looks normal) even when the original population of scores is not distributed normally.