

NOTICE. Unless otherwise indicated, all materials on this page and linked pages at the rbtaylor.net and blue.temple.edu addresses are the sole property of Ralph B. Taylor and © 1999-2001 by Ralph B. Taylor. All users have the right to freely access and copy these pages provided that they: acknowledge the source, do not make changes on any pages, and do not charge more than copying costs for distribution.

NOTES ON HYPOTHESIS TESTING

OBJECTIVES

- To review the underlying common logic of all hypothesis tests
- To review the logic behind one and two tailed hypothesis tests

COMMON LOGIC BEHIND ALL HYPOTHESIS TESTING

In this course we will be discussing a variety of different types of hypothesis tests. All of these tests, however, have the same underlying logic.

Assuming that certain assumptions hold, where does the statistic I observe with a particular sample lie in the larger population of cases from which my sample was drawn? Or, stated differently:

Assuming that certain assumptions hold, does this sample, yielding a particular sample statistic X , represent a random sample drawn from a larger population of cases where the corresponding population parameter is Q , where Q corresponds to a particular hypothesized value that you have articulated beforehand.

In short, what you are trying to do is to understand how a particular sample results relates to a certain hypothesized population value from which the sample was drawn. This is one common feature of all hypothesis tests.

In this course, we will be comparing samples and populations on the following types of features: means, correlations, slopes, partial slopes, explained variance, and so on.

Whenever we test a hypothesis, we are comparing a sample statistic to a population parameter. The latter parameter may be known, or estimated.

Therefore, whenever we conduct a hypothesis, you want to be clear on the following points:

NOT0102BB

- ! what is the attribute in question?
- ! what is the sample that was drawn from the population?
- ! what is the population against which the sample statistic is being compared?
- ! do we know the corresponding population parameter, or are we estimating it to have a certain value?
- ! what IS that value?

Stated differently, whenever we test a hypothesis, or conduct a statistical test, you seek to make inferences from the sample to the population. Social scientists are seldom interested in simply the results they obtain with their sample. They want to know how the results apply to the larger population. For more details on this read p. 159 in Taylor (1994) Research methods in criminal justice.

Some examples:

- ! The National Crime Victimization Survey finds that about 25% of households among the 49,000 households interviewed were "touched by crime" in the past year. A household is "touched by crime" if any type of serious attempted or completed crime was experienced by any household member. What does this tell us about the proportion of US households that were "touched by crime?"
- ! A poll of 800 Philadelphians finds that 60% favor a civilian review board to oversee police operations. What does this say about the proportion of all Philadelphians favoring such a body?

You want to test against the population because your theory says that is where differences should exist. AND the chances are that the policy makers interested in your research don't care about what happened with your dinky sample of X many cases but rather are interested in, for example, all parolees in the state. The nature of what you expect the population results to be, and how you expect that to compare to your sample results, varies somewhat from test to test.

Statistics and Parameters

Sample statistics come in many different forms. Anything you can calculate about a variable from a sample of cases is a sample statistic. These would include: the mean (or average), the median, the standard deviation, the standard error of the mean, the maximum, the minimum, the interquartile range, and so on. When we are talking about more than one variable in a sample we

can calculate additional sample statistics like: correlations, slopes, partial slopes, explained variance, and the like.

In hypothesis testing we compare these sample results to population parameters, features of the population based on all the cases in the population. If we are talking about one attribute or one variable, the parameter might be something like the population mean, the population standard deviation, and so on. If we are talking about two or more variables, the parameter might be something like a difference, a slope, a correlation, explained variance, or a partial slope, or something else. Population parameters are usually represented by Greek symbols.

Some Example Comparisons

Does the mean of my sample differ significantly from the mean of my population (i.e., given that my population mean is what it is, what are the chances that the sample I have obtained, given its mean, came from the same population)? (Z test) If I have two samples that came from different cases, with what surety can I reject the null hypothesis that the two respective population means are identical to one another? (t-test, independent samples) Given the slope of an outcome on a predictor that I have obtained in my sample, what are the chances that in the population the slope is significantly different from 0? [Stated differently: Given the correlation I observe with my sample, what are the chances that in the population I can reject the null hypothesis that the correlation is significantly different from 0?] (t-test for b or beta weight or for correlation) Given the sample variance that I explain in my outcome with my predictors, what are the chances that I can reject the idea that in the population at large the variance I can explain with these predictors is significantly different from 0? (F-test of R squared).

Bringing in Sampling Error

In all of these tests a fundamental idea is that we are testing the difference between a sample statistic against a population parameter, controlling for the amount of dispersion or sampling error. Every sample will differ somewhat from its population simply due to sampling error. So I need to control for this. I do so by using as my measure of dispersion a measure of sample standard error. Here is an example of sampling error. Say you have 20 coins in your pocket; 15 of them are pennies, and 5 of them are nickels. In your population the proportion of nickels is 25% of the population. So $p(n) = .25$. Throw the coins into a hat. Grab a handful and throw them on the table. You have drawn a random sample. Count the proportion of nickels. Put them back in the hat. Draw a second sample, trying to make it the same number of coins. Keep on doing this until you get tired. You will see that the proportion of nickels in the samples varies. Sometimes it is below .25, sometimes it is above .25. These differences arise from sampling error: differences around the population mean that arise simply because each sample is not a perfect reflection of the

NOT0102BB

larger population. Differences arise just due to chance. But given the "laws" of probability theory we can estimate how much the sampling error "should" be for certain size samples drawn randomly from populations. We will talk more about sampling error later. But the important point now is simply that we recognize, due to chance, sample results will differ from the corresponding population attribute, even if the sample is a perfect random sample of cases drawn from the population. Every sample will differ somewhat from its population simply due to sampling error. So I need to control for this. I do so by using as my measure of dispersion a measure of sample standard error.

THE COMMON LOGIC IS REFLECTED IN THE STRUCTURE OF THE STATISTICAL TESTS THAT YOU WILL USE

All of the statistical tests that we use in this course have a common underlying structure, reflecting the common underlying logic. Whenever you see a formula, and it will be often, you want to look and see how this common structure is represented in the formula.

$$\frac{[(\text{Sample Statistic}) - (\text{Population Parameter})]}{\text{-----}}$$

Measure of sampling error

WAYS THAT HYPOTHESES DIFFER

We have talked about how all hypotheses are fundamentally similar. But before we get to examples, we must talk about ways hypotheses differ. There are two types of variations.

NULL AND ALTERNATE HYPOTHESES

Your null hypothesis states that you do NOT expect your sample statistics to differ from your population parameter. You expect NO difference between the two. Oftentimes this is the hypothesis you hope to reject. You hope to reject it because you hope that there is a difference. Here are some examples of null hypotheses: * the reported violent crime rate is no different in southern states than it is in the population of all states in the US. * women are not more fearful of crime than are men; the fear levels of the two are not different * the average reported state-level property crime rate in all the United States is the same in 1985 as it was in 1990; there was no

NOT0102BB

difference By contrast your alternate hypothesis DOES expect a difference

ONE EXAMPLE: Z TEST

One type of hypothesis is to see if a sample mean differs "significantly" from a population mean. We are seeing how well a sample statistic compares to a population parameter. In the case of a z test. $Z = \frac{(\text{sample mean} - \text{population mean})}{\text{SEM}}$ where SEM= the standard error of the mean in the sample. The SEM controls for how dispersed the sample is on the variable of interest.

This next example violates a range of assumptions for statistical testing, and we will come back to those assumptions later, for they are important.

In different classes or different readings you may have read about the "subculture of violence" that might help explain the higher rates of violence in southern states.

For some interesting recent work in this area, see the works of R. Nisbett:

- ! R. Nisbett (1993). Violence and U.S. regional culture. American Psychologist 48, 441-449
- ! D. Cohen and R. Nisbett (1994). Self protection and the culture of honor: Explaining Southern violence. Personality and Social Psychology Bulletin 20, 551-567.

Let's look at our ecological data set and see if we can see if the southern states have higher reported violent crime rates, on average, than the population of 50 states. For the population of 50 states, for example, I know the following about the reported violent crime rate per 100k population in 1985 (VIOLRA85):

Mean	406.6
SD	206.77

NOT0102BB

For southern states (n=16; CENREG=3) the figures are as follows:

Mean	458.1
SD	184.9
SEM	46.2

With a sample, we assume that the standard error of the sample mean (SEM) "stands in" for the standard deviation of the sampling distribution of means. (This is an extremely important family of distributions; more later.)

So my Z test is: $[(\text{sample mean}) - (\text{pop mean})] / \text{SEM} = 51.5/46.2 = 1.11$. To interpret where this lies on the sampling distribution of means, you must convert the Z to an area under the curve (see url below), to see how "far out" it is. So we are testing a mean difference against a measure of "expected dispersion" from sampling error. With a sample, we assume that the standard error of the sample mean (SEM) "stands in" for the standard deviation

We can convert z scores to areas under the curve using a lookup table in the back of a stat book. Or you can do it online. Go to the following url:

<http://www.psychstat.smsu.edu/introbook/normal.htm>

Leave mean at 0 and sd=1. This is what the parameters are for a z distribution anyway. In the box to the right of

X<

type in the z score obtained (e.g., 1.11). Then click on the right pointing arrow. A number should pop up under probability. If you used 1.11 for the z that number should be .867.

WHAT DOES THIS MEAN? It means that, assuming all kinds of things, if the southern states were a random sample from all states, and you took lots of samples of southern states, the chances that you would get a mean score AT OR BELOW THE SAMPLE MEAN YOU HAVE HERE are about 87 times out of 100. Stated differently, if you were to take 100 samples of southern states, and those states were random samples from all US states, you could get a score at or below the mean you have here in about 87 of those samples. The assumption here is repeated random sampling from a larger population.

WHERE DOES THE SEM COME FROM?

We are going to be using the standard error in a lot of statistical tests, so it makes sense for us to get a little more familiar with it. It relates to other concepts with which you are already familiar such as sample variance and sample standard deviation.

VARIANCE is a measure of dispersion and is simply the sum of squared differences around a mean, divided by the n of cases

$$\text{VARIANCE} = s^2 = [\text{Sum} (X_{\text{raw}} - X_{\text{average}})^2] / N$$

Stated differently, it is the average squared difference between a raw score and the mean. It tells you how much, on average, each case contributes to the sum of squared differences around the mean. You might naturally ask, why do we square the differences around the mean, why not just sum the absolute values? Good question.

We square the differences because we want to progressively weight the cases as they become further and further from the mean. In other words, cases closer to the mean are more "typical," and cases farther from the mean are less "typical," but as cases start to get really far from the mean they are really untypical.

There are two technical ideas behind the idea of "far out" cases. First, the mean is a "best guess" on a variable. If you do not know what the score of a case is on a variable, your "best guess" is going to be the mean. For example, if someone says to you, what was the rate of reported violent crime in Oregon in 1985, and you know that in all 50 states the average violent crime rate for 1985 was 406.6 per 100,000 persons, then your best guess for Oregon is 406.6.

Your guess is best in the following sense. If you were asked these questions lots of times, and had no information about the cases, and every time you guessed the mean of all the cases, over lots of questions this strategy would result in the lowest errors, i.e., on average, over a lot of questions, this strategy would come closer than any other strategy to giving you closest to the correct answers.

Second, we often assume that scores on a variable are distributed in a normal fashion, with most scores clustered close to the average. Remember, 66% are within one standard deviation of the mean. Therefore, in a way we are operationalizing the idea of normality by squaring differences between raw scores and the means.

STANDARD DEVIATION. The standard deviation is probably our best measure of dispersion in a sample and can be defined as the square root of the variance.

NOT0102BB

$\sqrt{s^2} = s$ (standard deviation) for sample

$s = \sqrt{\text{arithmetic mean of squared deviations from the mean}}$

$s = \sqrt{\text{average squared difference from mean}}$

$s = \sqrt{[(\text{sum}((x - \text{Mean})^2)) / N]}$

$\sigma =$ standard deviation for population (this is the Greek letter sigma)

Note that for s , like variance: 1. All the cases get used; 2. Extreme cases have large weight. ¹

WHY DO WE SQUARE ROOT? There are two answers (Blalock p. 80). First, we are "compensating" for having squared difference scores. Second, and more importantly, s has important properties that relate to the normal curve. Again, remember that 66% of cases are within one standard deviation of the mean, and almost 95% are within two standard deviations of the mean. When we are working with normal data, or when we are working on the normal distribution, s is our best measure of dispersion.

The more homogeneous the sample on the attribute in question, the smaller the standard deviation will be; heterogeneity => larger standard deviation.

SO WHY DON'T WE USE THE STANDARD DEVIATION AS OUR MEASURE OF DISPERSION IN SOMETHING LIKE THE Z TEST? We do, but we are using the standard deviation not of the sample but of the distribution of sample means or the sampling distribution of means.

SAMPLING DISTRIBUTION OF MEANS (SDMs). This is the distribution of means created when you draw repeated random samples, of a certain size, from a population, or large group of scores.

SDMs are theoretical distributions. There exists one SDM for each population mean, and each sample size.

Let's go back to the 1985 reported violent crime rate (VIOLRA85). We know that the mean in all 50 states is 406.6.

NOT0102BB

Suppose that we started drawing random samples. Each sample consists of two states ($n=2$). When we draw each sample we compute the sample average on VIOLRA85.

If we do not allow a state to "go back" into the population after it has been drawn, we will be able to draw 25 samples. If we do allow states to "go back" we can draw an almost infinite number of samples.

Nevertheless, however we do it, we are generating a number of sample means. These sample means themselves will be distributed around the population mean (μ) in a very specific way.

We know the way they will be distributed because we rely on a theorem in probability theory called the central limit theorem. It is stated as follows:

If repeated random samples of size N are drawn from any population (of whatever form) having a mean μ and a variance σ^2 (σ^2), then as N becomes large, the sampling distribution of sample means approaches normality [i.e., a normal distribution] with mean μ and variance [σ^2 / N] (Blalock 1979 p. 180)

You know that your sample means will differ from your population means just due to sampling error. Each sample misses some elements of the population, and thus is not a perfect cross-section of the population. Thus you expect each sample's mean to differ somewhat from the population mean. The above theorem recognizes this but also goes on to say exactly how those sample means will be distributed.

First, it says that the mean of sample means [average (X_{avg})] will be equivalent to the population mean (μ). This assumes that all of the samples were random and were sufficiently large. Therefore for μ we can substitute [average (X_{avg})].

Second, it tells us the variance of these sample means. From that we can get the standard deviation.

If

$$\text{variance SDM} = \sigma^2 / N$$

$$\text{then standard deviation of SDM} = \sqrt{\sigma^2 / N}$$

$$\text{Note that } \sqrt{\sigma^2 / N} \text{ becomes } = \sigma / \sqrt{N}$$

NOT0102BB

For these three types of distributions (Blalock, p. 181) we can say the following about their means and standard deviations [remember N refers to the size of the samples]:

Distribution	Mean	Standard Deviation
Population	$\mu (\mu)$	$\sigma (\sigma)$
Sample	X_{avg}	s
Sampling Distribution	μ	$[\sigma / \sqrt{N}]$

For each sample of a certain size from a population where μ and σ are known, we can know how means of those samples are distributed.

Third, it tells us that the distribution of scores will approach normality under certain assumptions. If things are normal then we know that certain relationships hold between the mean, the standard deviation, and the percentage of cases under certain areas of the normal curve.

If we put the second and third idea together we get an important fourth idea. If I draw 1,000 samples of size N from a population, and each of those is randomly drawn, and unbiased, just due to sampling error I will have about 68.26% of those cases -- REMEMBER, each case is a mean -- or about 683 of them, falling in the range of [mean - (1s)] on the SDM to [mean + (1s)] on the SDM: ²

$[\text{Avg}(X_{avg})] - 1 * ([\sigma / \text{sqrt}(N)])$	\Leftrightarrow	$[\text{Avg}(X_{avg})] + 1 * ([\sigma / \text{sqrt}(N)])$
---	-------------------	---

We let $[\text{Avg}(X_{avg})]$ stand in for μ if the latter is not known.

95.46% of the means, or about 955 of them, will fall in the range of + or - 2s (two standard deviations) around the mean:³

--	--	--

$[\text{Avg}(X_{\text{avg}})] - 2 * ([\text{sigma}/\text{sqrt}(N)])$	\Leftrightarrow	$[\text{Avg}(X_{\text{avg}})] + 2 * ([\text{sigma}/\text{sqrt}(N)])$
--	-------------------	--

This is a crucial idea. It tells us where under certain assumptions various proportions of our sample means should fall.

Fifth, this theorem tells us what happens to SDM if we draw larger vs. smaller samples.

In the case of VIOLRA85, if we draw repeated samples of size 10 (i.e., 10 states in each sample), instead of size 2, we will have a distribution of sampling means with a smaller standard deviation. This necessarily follows from the definition of standard deviation on a sampling distribution. In other words, the sample means will cluster closer to the average sample mean.

BUT HOW DO WE USE THIS INFORMATION WHEN WE DO NOT KNOW THE POPULATION VARIANCE OR POPULATION STANDARD DEVIATION?

We substitute s for sigma and calculate $s / \text{sqrt}(N)$

This is the standard error of the mean (SE or SEM) in one particular sample and it is equivalent to the standard deviation of the sampling distribution of means.

SEM = $s / \text{sqrt}(N)$

BACK TO HYPOTHESIS TESTING

In this section I run through the conceptual and mechanical steps in conducting a hypothesis test, using as the vehicle a z test for comparing a sample mean to a population mean.

You want to become familiar with these steps because the underlying logic is similar no matter what the form of the particular statistical test that you will use.

There are several important steps in hypothesis testing. When we conduct a statistical test we are testing a hypothesis (Blalock 1979, 154 on).

1. STATE ASSUMPTIONS

Every statistical test makes certain assumptions. You need to know them. The assumptions vary somewhat from test to test. For the Z test we assume:

- a. Repeated random sampling. Namely that the cases drawn were a random sample drawn from

NOT0102BB

the population.

This first assumption becomes your null hypothesis. A null hypothesis states that you expect to find no difference exists between two groups. Here your null hypothesis is that the average violent crime rate in southern states is equivalent to the average violent crime rate in the US as a whole.

The null hypothesis stands in contrast to the research or alternate hypothesis. The latter states that you do expect to find a difference.

b. If the N of cases in the sample is not too large, we need to assume that the variable is normally distributed in the population. If N is large (> 50) we need not worry about this.

Since in much of the work we will be doing, we are sticking with the 50 states, and working with non-normal variables, we need to try and make the variables with which we are working as normal as possible. (in this example here, however, we will not worry about it.)

c. You assume that the level of measurement is at least at the interval level. If the data are less than interval the test is not appropriate; you need to do nonparametric tests

2. SELECT CRITICAL REGION

You need to divide possible outcomes of the statistical test into two categories; those outcomes that will allow you to reject the null hypothesis of no difference, and those that will not allow you to reject the null hypothesis of no difference.

What you are doing here is selecting a region of the normal curve (ie., a region on the SDM, which we know approaches normality under certain assumptions) and saying: if my z score is in this region, I will reject the idea that my sample came from a population with $\mu = 406.6$ and $\sigma = 206.77$.

This is called the rejection region

In other words, **if I am in the rejection region with my results** I reject the idea that my sample mean came from a sampling distribution of means based on sample sizes of 16 where $\mu = 406.6$ and $\sigma = [206.77/4]$. ($4 = \sqrt{N}$)

I will reject this idea because, if my z score is in this region, my actual sample mean suggests that it is placed too far out on the SDM that would be generated, with samples of size 16, from a population with the specified mean and standard deviation.

NOT0102BB

In setting the critical region you must decide if your alternate hypothesis is one tailed or two tailed. A one-tailed hypothesis is directional; you are saying that you definitely expect your sample mean to be above or below the population mean. With a two-tailed hypothesis you are saying your sample mean may come out either far above or far below the mean; either outcome leads you to reject the null hypothesis.

The one-tailed vs. two-tailed decision must be made before you examine your results.

In the case here, given the work on the southern subculture of violence that you find convincing, you might decide up front to do a one-tailed test.

In other words, you will reject the null hypothesis if your sample mean is significantly above the population mean. You will not reject the null hypothesis if your sample mean is significantly below the population mean.

You have now defined your critical region. The sum of the critical regions is typically referred to as your significance level (or alpha or p level), and is usually set at $p < .05$ due to convention. It is stating your tolerance for incorrectly rejecting the null hypothesis of "no difference". This is also called your Type I error level.

There can be situations where your statistical test suggests that you reject the null hypothesis of no difference. For example you might find that the violent crime rate is significantly higher in southern states as compared to the US as a whole. BUT YOUR STATISTICAL TEST MAY BE WRONG. In other words, it might really be the case -- out there in the real world -- that the southern violent crime rate is not significantly higher even though your test might say that it is. Your chances of making this kind of mistake are equal to your alpha level.

So if your alpha level is $p < .05$, and all the assumptions we have been talking about hold, and you do this test 100 times, and it really is the case that the null hypothesis is true, then in five out of 100 of those tests, you will mistakenly reject the null hypothesis of no difference. Each time you do the test you have a 5% chance of making this mistake if your alpha level is .05.

You reduce this error level by reducing your significance level and therefore your critical region. For example you might choose $p < .01$.

However, when you decrease your alpha level and your critical region you increase your chances of making a Type II error. Type II errors occur when you fail to reject a null hypothesis, i.e., you conclude "no difference", but in fact in the real world there really is a difference.

There are calculations you can do to specify what the Type II error level is in a particular situation. We will not get into them in this course. The most common source of Type II errors is doing an analysis with too few cases, i.e., samples that are too small and therefore have associated

NOT0102BB

standard errors that are too large. (Can you see why the standard errors would be too large given the formula for SEM?)

SITUATION IN THE REAL WORLD (assumes we have a magic spyglass allowing us to look at the real world situation without using statistics)		SITUATION AS SHOWN BY THE STATISTICAL RESULTS
THERE REALLY IS NOT A SIGNIFICANT DIFFERENCE BETWEEN SAMPLE AND POP; SHOULD ACCEPT NULL HYPOTHESIS	THERE REALLY IS A SIGNIFICANT DIFFERENCE BETWEEN SAMPLE AND POP; SHOULD REJECT NULL HYPOTHESIS	
You have correctly failed to reject the null hypothesis	You have committed a TYPE II ERROR . Most probable reason for doing this is weak statistical power due to too small a sample	STATS SHOW NO SIGNIFICANT DIFFERENCE
You have committed a TYPE I ERROR . The Type I error rate is set simply by the alpha level you use.	You have correctly rejected the null hypothesis ; your chances of doing this are determined by the statistical power of your test	STATS DO SHOW A SIGNIFICANT DIFFERENCE

We will routinely be dealing with 50 cases in statistical analyses of states. You should feel free if you want to set your alpha level at .10 so that your chances of making a Type II error are not too high.

You need to understand the bases of these different kinds of errors.

3. STATE ALL THE ASSUMPTIONS AND THE MODEL WITH WHICH YOU WILL BE WORKING

The next step is to lay out the entire set of assumptions with which you are working. These include the level of measurement, the assumption of repeated random samples from the population

Level of measurement	Interval
Model	Normal population $\mu = 406.6$ $\sigma = 206.77$ $\alpha \leq .10, 1 \text{ tailed}$
Hypothesis: (null)	Random sampling from the population

So you will do a one tailed test. You expect your sample to score higher than the population. You need to look up the z score associated with a value higher than 89.999% of the cases under the normal curve. If the z value you obtain is higher than this value then you are in the critical region on the SDM, and you can reject the null hypothesis that your sample "belongs" on that SDM, and therefore you conclude that your sample did not come from that population.

To do this online, put in .89999 under probability, then click on the left pointing arrow. This will give you your z critical.

4. FIND THE CRITICAL VALUE OF YOUR TEST STATISTICS. HERE: HOW DO I FIND MY CRITICAL Z VALUE?

With any test statistic, you are looking for the value indicating that your result is in the critical region, and therefore allows you to reject the null hypothesis, whatever that may be.

Either the printout provides the critical value needed to reject the null hypothesis, or a lookup table does. Unfortunately, Hamilton does not have a lookup table for z values. Check this online. We see the associated value is 1.281

5. YOU RUN THE TEST.

Here is how the test would run if you use the population standard deviation.

$$z = [X_{\text{avg}} - \mu] / [\sigma / \sqrt{16}]$$

$$z = [458.1 - 406.6] / [206.77 / \sqrt{16}]$$

$$z = 51.5 / 51.59$$

NOT0102BB

$z = .996$

6. INTERPRET

The final and most important part of doing a statistical test is interpreting the results. When you get a result you cannot simply say: "I reject the null hypothesis," or "I fail to reject the null hypothesis." **You have to put the result you obtain in a statistical and substantive context.**

This is slightly less than one standard deviation above the population mean and corresponds to an alpha level of .84. In other words, 84% of the random sample means based on a sample size of 16, from this population ($\mu = 406.6$; $\sigma = 206.77$) fall at or below this value, and 16% fall above. Therefore you are unable to reject the null hypothesis that this southern sample of 16 states, on the attribute of 1985 reported violent crime rate, represents a random sample from the larger population of 50 states. The result you got (a discrepancy this extreme from the population mean, or less extreme) could happen about 16 times out of 100 and your null hypothesis of no difference could still be true. You have said, by selecting $\alpha = .10$ and a one tailed test, that you needed a z obtained that would equal or exceed your z critical (1.282) - i.e., you needed a result that would happen under the assumption of the null hypothesis being true only 10 times out of a 100 (remember we are assuming repeated random samples) or less often.

So do we find support that for the idea that the southern subculture is more violent than the rest of the nation?

THE ONE SAMPLE T TEST

If you do not know the population standard deviation, the appropriate test is a one sample t-test or "Student's" t-test. The formula for this is:

$$t = [X_{\text{avg}} - \mu] / [s / \sqrt{n}]$$

where n = sample size and s = sample standard deviation.⁴

Of course, $[s / \sqrt{n}]$ is actually the standard error of the sample mean (SEM)

Go through steps 1-4 as you did above for the z test. Decide your assumptions and your model. For example:

NOT0102BB

Level of measurement	Interval
Model: ⁵	Normal population $\mu = 406.6$ alpha < or = .10, 1 tailed
Hypothesis: (null)	Random sampling from the population ($x_{\text{avg}} = \mu$)

You find your critical value. Consult p. 350 in Hamilton.

You will need to examine your degrees of freedom. The table shows you the sampling distribution of t for different degrees of freedom. In this test the

$$df = n-1 = 15$$

Conceptually, what do degrees of freedom refer to? They refer to:

the number of quantities that are unknown minus the number of independent equations linking these unknowns. (Blalock, p. 205).

In the case here df refer specifically to computations for s. Recall the definition of s. In a sample, in computing s, there are as many unknown quantities as there are cases, minus 1. We take one out because we have to calculate the average. That becomes a fixed point around which we calculate the difference scores, and squared difference scores.

Although it will get more complicated when we get to multiple regression, with one-sample t-tests df always = (n-1).

So what is our critical t value for df=15 and p<.10, one-tailed?

Once we have found this we know that the t value we obtain must exceed this value for us to be sure our results are in the rejection (or critical) region, and therefore for us to reject the null hypothesis.

So, for the un-transformed 1985 reported violent crime rate we have:

$$t(15) = 458.1 - 406.6 / 46.2 = 1.11$$

Have we obtained or exceeded our critical t-value? Are the results within the rejection region?

NOT0102BB

GLOSSARY

Based on the above notes you should have a sound understanding of the following terms:

alpha level

central limit theorem

critical region (rejection region)

degrees of freedom

failing to reject the null hypothesis

null hypothesis

one-tailed vs. two-tailed test

rejecting the null hypothesis

research or alternate hypothesis

sampling distribution of means

sampling error

standard deviation

Student's t (one sample t-test)

Type I error

Type II error

variance

z test (formula)

REFERENCES

NOT0102BB

Blalock, H. (1979) Social statistics. New York: McGraw Hill.

Taylor, R. B. (1994) Chapter 10 ("Sampling") in Research methods in criminal justice. New York: McGraw Hill.

FOOTNOTES

1. Sometimes you will see formulas for the standard deviation that use (N-1) in the denominator rather than N. When N-1 is used this is called the sample standard deviation. Sample standard deviation should be used for samples smaller than 30 cases although, most of the time, it does not make much practical difference. What the sample s does -- intuitively -- is counter for fact that with smaller (random) sample (of population) we are less likely to include extreme scores in the sample. For the real story, see Blalock, pp. 205 - 206.

2. The n within the specified range will not come out exactly right each time, but if you were to do this over a large number of distributions, it would work out to this number on average.

3. Same comment as in Fn. above.

4. An alternate and perhaps more widely used formula is

$$t = [X_{\text{avg}} - \mu] / [s / \text{sqrt}(n-1)]$$

We are simply not going to worry about such niceties at this point.

5. No assumption is necessary about since we will be using $s / \text{sqrt}(N)$ instead, and we know these quantities.