MSPP PPOL 501–02 & -06: Fall 2014
Course Notes #10: One-Sample Hypothesis Testing for Means; Variations on the theme (*t*-tests; one-sided tests; *p*-values)
Professor Carolyn Hill

I. HYPOTHESIS TESTING: GENERAL POINTS

- Why do hypothesis testing?

  Our goal as social scientists / policy analysts is to make some statements about a *population* of interest, when we only have access to information about particular observations from a *random sample* from that population.

- Why would we only have information about a sample, instead of about the whole population?

  There are many possible reasons, but the biggies are financial cost and time of gathering information about all observations in a population; or logistics that preclude gathering info about all observations – e.g., what if a nonprofit organization that is an element of the population of nonprofits I'm interested in is physically located on the shores of the Bering Sea and has no telephone, e-mail, or other communications devices? (actually, logistical and time problems could be translated into cost metrics – can't everything?!)

- Will a hypothesis test tell me *for certain* an answer to my question of interest?

  Ready for the depressing answer?: NO. Hypothesis testing is all about making statements of probability:

  "how likely is it that I would have drawn a sample that looks like the one that I got, if in fact the null hypothesis were true?"

  What does "looks like the one I got" mean? Usually, we are looking at the *average* or *mean* value of some variable of interest from the sample. This is not always the case, however, and different hypothesis tests are designed to answer different questions of interest.

  Notice that this process does *not* answer the question: "is the null hypothesis true?" or the question "is the null hypothesis false?"

- How do we figure out which type of hypothesis test is appropriate?

  A key determinant is the *level of measurement* of the variables of interest: nominal, ordinal, or interval/ratio (see notes from the first lecture; see me for more info if you are unclear about these definitions)

- Steps in hypothesis testing provide a common language and format:

  1*. Formulate a working hypothesis (i.e., theory-based or informal expectations about the relationships you expect to find)

  2. State a null hypothesis and an alternative hypothesis (this translates your "working hypothesis" developed in #1 into the formal apparatus for statistical inference).

  3. Choose a level of statistical significance (i.e., locate the appropriate critical values for rejecting the null hypothesis): this step identifies your decision rule for determining the areas under the sampling distribution that include unlikely sample outcomes

     (note: Step 3 is associated with *classical hypothesis testing* or the *critical value approach;* other, equivalent ways of testing hypotheses are (a) computing confidence intervals; and (b) interpreting p-values. We'll say more about these alternative forms of hypothesis testing a little bit later)

  4. Compute the test statistic (*Z* score, *t*-statistic, χ^2 statistic, F-statistic, etc.).

  5. Compare the test statistic with the critical value and make a decision.

II. HYPOTHESIS TESTING:
For One Population Mean when the Population Standard Deviation is KNOWN
(used for interval-ratio variables)

Example 1: A brief introduction

[Weiss 9.77]: According to the Bureau of Crime Statistics and Research of Australia…the mean length of imprisonment for motor-vehicle-theft offenders in Australia is 16.7 months. One hundred randomly selected motor-vehicle-theft offenders in Sydney, Australia, had a mean length of imprisonment of 17.8 months. At the 5% significance level, do the data provide sufficient evidence to conclude that the mean length of imprisonment for motor vehicle-theft offenders in Sydney differs from the national mean in Australia? Assume that the population standard deviation of the lengths of imprisonment for motor vehicle-theft offenders in Sydney is 6.0 months.

Go through the steps…

(1*) Informal expectation ……..

(2) State the null and alternative hypotheses:

H_0: $\mu_{SYDNEY} = 16.7$ months

H_1: $\mu_{SYDNEY} \neq 16.7$ months

(3) Choose a level of significance (or a critical / hypothesis rejection region) and identify the critical values of the test statistic. In this case, alpha = α = 0.05 (two-tailed), so Z-critical = ±1.96.

(4) Compute the test statistic: $Z_{calc} = \dfrac{(\overline{X} - \mu)}{\sigma / \sqrt{n}} = \dfrac{17.8 - 16.7}{6 / \sqrt{100}} = \dfrac{1.1}{0.6} = 1.83$

(5) Make a decision. Compare the test statistic to the critical region.

$-1.96 < 1.83 < 1.96$

⇒ **Decision: Fail to Reject the Null Hypothesis.**
Note: we do not *"accept the null hypothesis"!!!*

So we conclude that the sample we drew could have occurred by chance – the average length of imprisonment for motor-vehicle-theft offenders in Sydney, Australia is not statistically different from the average length of imprisonment for all motor-vehicle-theft offenders in Australia.

<u>Example 1: The extended dance remix</u>

• Now that you've seen the whole process, let's walk through this all again and dig into what's going on.

[Weiss 9.77]: According to the Bureau of Crime Statistics and Research of Australia…the mean length of imprisonment for motor-vehicle-theft offenders in Australia is 16.7 months. So far this year, 100 randomly selected motor-vehicle-theft offenders in Sydney, Australia, had a mean length of imprisonment of 17.8 months. At the 5% significance level, do the data provide sufficient evidence to conclude that the mean length of imprisonment for motor vehicle-theft offenders in Sydney differs from the national mean in Australia? Assume that the population standard deviation of the lengths of imprisonment for motor vehicle-theft offenders in Sydney is 6.0 months.

- This is a hypothesis test about a single population mean: that is, you have information from a one sample of motor-vehicle theft offenders in Sydney, and you want to make some inference from that sample *to the population of interest.*

  In this case, the sample of n=100 is drawn from motor vehicle-theft offenders in Sydney. Remember, we want to make an inference about the *population of motor-vehicle-theft offenders in Sydney,* not just describe the 100 that were sampled.

- The population mean value for all motor-vehicle-theft offenders in all of Australia just provides a point of reference with which to test performance of offenders in Sydney this year in particular.

Think about the information you already know and the information you want to know:

(a) *fact*: the population mean length of imprisonment for all motor-vehicle-theft offenders in Australia is 16.7 months.

(b) *fact*: the sample mean length of imprisonment for n=100 randomly-selected offenders so far this year in Sydney is 17.8 months, and the population standard deviation for this population in Sydney is = 6.0 months.

(c) *unknown*: what is the population mean length of imprisonment for all motor-vehicle theft offenders in Sydney this year?

So the statistical inference you want to make from your sample information is to the <u>population of motor-vehicle-theft offenders in Sydney</u>.

- Go through the steps for hypothesis testing:

  (1*) Informal expectation ……..

  (2) State the null and alternative hypotheses:

  H_0: $\mu_{SYDNEY} = 16.7$ months

  This means that:

  H_0: The true mean length of imprisonment (i.e., the population mean value) of motor-vehicle theft offenders in Sydney is 16.7 months. (the reference value of 16.7 is drawn from all of Australia – it was given to you as a population statistic for that population of interest).
  or

  H_1: $\mu_{SYDNEY} \neq 16.7$ months

  This means that:

  H_1: The true mean length of imprisonment (i.e., the population mean value) of motor-vehicle theft offenders in Sydney is NOT 16.7 months.

Notes: (i) I did not specify whether offenders in Sydney had higher or lower lengths of imprisonment than the reference. This type of test is a "two-tailed test" for reasons that we'll see in a bit. It is a more conservative test than a "one-tailed test" and is typically the one you will use.

(ii) The null and alternative hypotheses <u>do not</u> make statements about the <u>sample mean</u> of the offenders in Sydney, but instead about the ***population mean***

of offenders in Sydney – remember, we're interested in making inferences in this case about the population mean of offenders in Sydney more generally, not the specific sample that we drew. For a point of comparison, we used the population mean of all motor-vehicle-theft offenders in Australia from the official reports.

Here are some examples of null hypotheses that are not correct, with suggestions for how they could be corrected (We do not list the corresponding alternative hypotheses, but if you have questions, please ask)

Ex1: $H_{0:}$ The mean of the population is equal to 16.7.
Corrected: H_0: The mean length of imprisonment for the population of motor-vehicle-theft offenders in Sydney so far this year 16.7 months.

Ex1: $H_{0:}$ The population mean length of imprisonment for motor vehicle theft offenders in Sydney so far this year is not statistically different from 16.7 months.
Corrected: H_0: The mean length of imprisonment for the population of motor-vehicle-theft offenders in Sydney so far this year 16.7 months.
The null and alternative hypothesis do not say anything about "statistically significant" anything: the null and alternative are possible true states of the world and thus statistical significance is not appropriate in their statement. The statement of stat sig comes later in the process when you make a decision based on your hypothesis test.

Ex3: H_0: $\overline{X} = 16.7$
Corrected: H_0: $\mu_{SYDNEY} = 16.7$ months

Ex4: H_0: $\overline{X} = \mu$
Corrected: H_0: $\mu_{SYDNEY} = 16.7$ months

Ex5: H_0: The length of imprisonment for offenders in Sydney so far this year is equal to the length of imprisonment for those in Australia historically.
Corrected: H_0: The length of imprisonment for all motor-vehicle-theft offenders in Sydney so far this year is equal to the length of imprisonment for all motor-vehicle-theft offenders in Australia historically.
$\mu_{SYDNEY} = 16.7$ months

(3) Choose a level of significance (or a critical / hypothesis rejection region). This step basically sets up a standard for testing the hypothesis. There are 2 possibilities of the "true state of the world," given the sample that we drew:

(i) Perhaps the state of the world is that the true mean length of imprisonment for motor-vehicle-theft offenders in Sydney so far this year is 16.7 months. If this is the true state of the world, then the sample we drew and the corresponding sample mean length of imprisonment we calculated could have just occurred by random chance alone.

(ii) Or, perhaps the state of the world is that the true mean length of imprisonment for motor-vehicle-theft offenders in Sydney so far this year is NOT 16.7 months. If this is the true state of the world, then the sample that we drew and the corresponding sample mean we calculated are unlikely to have occurred by random chance alone.

Key fact: The hypothesis testing process begins with the assumption that (i) is the case.

Choosing the critical region involves selecting a point on the sampling distribution of sample means beyond which the probability of drawing the sample that we did is likely to occur less than 5 times out of 100.

Another way of saying this is that the probability of drawing that sample = 0.05: alpha = $\alpha = 0.05$.

Because we did not specify the direction (higher or lower) in step (2) above, we want to evenly divide this probability into the two "tails" of the normal distribution – so, 0.025 probability in each tail.

Again, the CLT helps us out. Recall that we know the probability associated with areas under the normal curve. Thus, we can figure out a Z-score beyond which the area under the curve is 0.025. Using the appendix, we see that the Z-scores for these points = $\pm$ 1.96.

So, any sample outcome falling in the shaded areas has a probability of less than 0.05 of occurring. We declare such an outcome would be a "rare" event and would cause us to reject explanation (ii).

Draw a picture:

(4) Compute the test statistic: $Z_{calc} = \frac{(\overline{X} - \mu)}{\sigma / \sqrt{n}} = \frac{17.8 - 16.7}{6 / \sqrt{100}} = \frac{1.1}{0.6} = 1.83$

(5) Make a decision. Compare the test statistic to the critical region. 1.83 is not in beyond the critical values of $Z = \pm 1.96$:

-1.96 < 1.83 < 1.96. So we conclude that the sample we drew could have occurred by chance – there is a reasonable chance that the sample mean length of imprisonment for motor vehicle theft offenders in Sydney could have occurred by chance alone, if the true mean length of imprisonment for this population were 16.7 months.

⇒ **Decision: Fail to Reject the Null Hypothesis.**

Note: we do not *"accept the null hypothesis"!!!*

III. VARIATIONS ON THE THEME: *t-tests*

- Up until now, we have used the standard normal distribution when we know the population standard deviation.

- When the standard deviation is estimated from the sample (instead of known from the population), the test statistic actually follows a ***t-* distribution** rather than the standard Normal distribution.

- As noted earlier, the ***t-distribution*** has a symmetric, "bell" shape similar to the Normal distribution, but its precise shape depends on the number of "degrees of freedom," or pieces of information (i.e., observations) in the sample. ***d.f. = n-1***

- Example: A pre-natal program is started at a local health clinic. Before the program started, the average number of visits to the clinic by pregnant mothers was 2.9 visits. After the program started, you took a random sample of 34 mothers enrolled in the pre-natal program and find that on average, they have an average of 3.47 visits, with a standard deviation of 1.40.

  Did the pre-natal program change the number of visits to the clinic by pregnant mothers? Does getting a sample mean of 3.47 from a sample of $n = 34$ mean that behavior has truly changed (relative to what it was before the program), or is it likely that behavior truly has not changed?

  1*: (Informal expectation): ……

  2. State the Null and alternative hypotheses:

     H_0: $\mu_{postprogram} = 2.9$ visits

     H_1: $\mu_{postprogram} \neq 2.9$ visits

  3. Select sampling distribution and establish critical region.

Let's use a 99% confidence level (alpha = $\alpha = 0.01$). This indicates that there is 1% chance that I will reject the null hypothesis even if this sample came from a population with a true mean of 2.9 visits.

Because of the small sample size and because we have an estimate of the population standard deviation from the sample, it's crucial that we get the critical value from the *t*-table.

So, for $d.f. = 34-1 = 33$, a two-sided test with alpha=0.01, the $t_{critical} = \pm 2.733$

4. Calculate the *t*-test statistic:

$$t_{calc} = \frac{(\overline{X} - \mu)}{s/\sqrt{n}} = \frac{3.47 - 2.9}{1.4/\sqrt{34}} = \frac{0.57}{0.240098} = 2.374$$

5. Make a decision: $|2.374| < |2.733| \Rightarrow$ Fail to reject the null hypothesis that the population mean number of visits is unchanged. We declare that it is unlikely (at the 99% confidence level) that pregnant mothers' clinic visits have changed.

Question: What would we conclude if we would have tested this hypothesis using a 95% confidence level?

IV. VARIATIONS ON THE THEME: ONE-TAILED HYPOTHESIS TESTS

- We've been looking at two-tailed tests of null hypotheses—that is, we've been equally concerned with the possibility that the true population value is greater than the hypothesized value a*nd* the possibility that it is less than the hypothesized value. This was reflected in alternative hypotheses that did not specify a direction of difference from the null hypothesis (saying things like "they *are* different"; "there has been *some* change" etc.)

- But maybe the researcher has information about the expected direction of how the true population value relates to the hypothesized value, e.g. when we're testing whether a program had some kind of effect in a particular direction

- Let's take the clinic visits example from earlier. Suppose we wanted to test whether the program *increased* the average number of visits, instead of just testing the fact that visits were different after the program. In this case, the null and alternative hypotheses would be of this form:

  H_0: $\mu_{postprogram} = 2.9$ visits

  H_1: $\mu_{postprogram} > 2.9$ visits

Let's say we still want to set a 99% confidence level. Now, however, instead of splitting the remaining 1% into the two tails of the sampling distribution (because we were agnostic about whether the value we were testing would be higher or lower), we load all 1% into the upper tail:

To get the critical value associated with this, we just go to the t-table, look at the one-tailed test section, α=0.01, d.f.=33, and get a t-critical value of +2.445.

We calculate the test statistic in the very same way as before:

$$t_{calc} = \frac{(\overline{X} - \mu)}{s/\sqrt{n}} = \frac{3.47 - 2.9}{1.4/\sqrt{34}} = \frac{0.57}{0.240098} = 2.374$$

Compare the calculated test statistic to the critical value: 2.374 < 2.445. Fail to reject the null hypothesis and conclude that clinic visits are <u>not statistically significantly greater</u> after the prenatal program began.

Question1: What would have been my decision if were testing this hypothesis at the 95% confidence level?

Question2: What would have been my decision if I had drawn a sample mean of 2.33, and thus calculated the test statistic to be –2.374?

Question3: Do one-tailed tests make it "easier" or "harder" to reject the null hypothesis: Why or why not?

V. VARIATIONS ON THE THEME: *p*-values

- This is a direct extension of topics we have already talked about.

  The ***p-value*** is the exact probability that we would obtain a test statistic as extreme or more extreme than the one we calculated from the sample, assuming that the null hypothesis is true.

  The p-value is also known as the "observed significance level."

- What's the idea behind the p-value? The classical hypothesis testing method sets a predetermined level at which a sample mean will be declared "unlikely" if the hypothesized population mean were true.

- When we test a hypothesis using a p-value, instead of setting up the rejection region in advance, we calculate the exact probability that we would observe the sample mean that we did or one that is greater (in absolute value), given that the null hypothesis is true.

  In other words, it gives us the smallest α at which the null hypothesis would be rejected.

- Using the p-value directly is one of the three types of hypothesis testing

- Different kinds of hypothesis tests require different kinds of test statistics (e.g., the *t*-statistic, the F-statistic, the Chi-square statistic),

  but a *p*-value can be associated with each of these different kinds of test statistics and is thus a universal way of understanding hypothesis testing.

- In interpreting p-values, we still implicitly use the conventional significance levels of 0.10, 0.05, and 0.01 to compare them to. Stating the specific p-value just gives more precise information about the test statistic in relation to the null hypothesis.

- Back to our example: we've already calculated the test statistic, t= 2.374.

  Find the number of d.f. in the sample: n-1 = 34-1 = 33. Now, read across the row for 33 d.f. until you find the range in which 2.374 falls.

  Here, we see that it is between α=0.025 and α=0.01 for a one-tailed test, or thus a corresponding range of α=0.05 and α=0.02 for two-tailed tests. Statistical packages will calculate the p-value for you exactly.

  This just says that if the true population mean were really 2.9 visits, then with repeated samples of size n=34, I would expect to get a sample mean=3.47 visits or greater (in absolute value – i.e., you need to look at both sides for a two-sided test), somewhere between 2 and 5 percent of the time.

Other ways of stating this: Post-program visits are statistically different from pre-program visits a two-tailed significance level between 0.02 and 0.05. (Note: they are also statistically different than the hypothesized mean at the 0.10 level. *Why* can I say this*?)*

- Another example:
  Suppose you are conducting a one-sample t-test and you are interesting in the per capita contributions to nonprofit organizations. A claim has been made by a reputable organization that on average, individuals in the U.S. give $1,000 per year to nonprofits.

  You draw a random sample of 1,600 individuals and ask how much they contributed to nonprofit organizations last year. The sample mean you calculate is $1,009.80 and the sample standard deviation is 200.

  The test statistic = 1.96 (check my math).

  What is the p-value associated with this test statistic? In other words, assuming that the null hypothesis is true, what is the probability that we would obtain a statistic as large or larger than the one we got in this sample (2-sided test)?

- O.K. Now, suppose that I drew another random sample, which had a sample mean of $1,008.23 and a standard deviation of 200.

  The test statistic associated with this sample is t=1.645.
  What is the p-value associated with this sample?

- Take another one: suppose I drew another random sample, which had a sample mean of $1,009 and a standard deviation of 200.
  The test statistic associated with this sample is t = 1.8.
  p-value = ???

- Another sample: sample mean = $1,100 and s=200.
  test statistic = t = 20
  p-value = ???

- Another sample: sample mean = $1,001 and s=200.
  test statistic = t = 0.20.
  p-value = ????

- Although all these examples have dealt with a single-sample mean t-test, the concept of a p-value is the same for other kinds of test statistics too (Chi-Square, F-test, t-tests on regression coefficients, t-tests on correlation coefficients, etc.)

VI. VARIATIONS ON THE THEME: RELATIONSHIP BETWEEN CONFIDENCE INTERVALS AND T-TESTS

- It may have occurred to you that there is some relationship between single-sample hypothesis tests and confidence intervals.

- One specific way to see the relationship is to think about the way we calculate critical values for a *t*-test and then manipulate the formula to see how we get a confidence interval for a sample mean:

$$\frac{(\bar{X}-\mu)}{s/\sqrt{n}} = \pm t \quad \Leftrightarrow \quad (\bar{X}-\mu) = \pm t\left(s/\sqrt{n}\right) \quad \Leftrightarrow \quad \mu = \bar{X} \pm t\left(s/\sqrt{n}\right)$$

- Back to earlier example:

  H_0: $\mu_{postprogram}$ = 2.9 visits

  H_1: $\mu_{postprogram} \neq 2.9$ visits

  Construct a 99% confidence interval around the sample mean:

  $$\text{CI} = \bar{X} \pm 2.733\,\sigma_{\bar{X}}$$

  $$= 3.47 \pm 2.733\left(\frac{1.4}{\sqrt{34}}\right)$$

  $$= 3.47 \pm 0.656188$$

  CI: [2.81, 4.13]

  Since the hypothesized population mean, 2.9 visits, is <u>inside of this interval</u>, we <u>fail to reject</u> the null hypothesis. It is not unlikely that the sample we drew came from a population with a mean of 2.9 visits.

VII. HYPOTHESIS TESTING FOR ONE-SAMPLE MEANS IN STATA

- To show examples of how to do various kinds of hypothesis tests in STATA, we'll use a data set drawn from the National Longitudinal Survey, where the variables are defined as follows:

n= 3010

```
  1. id                        person identifier
  2. nearc2                    =1 if near 2 yr college, 1966
  3. nearc4                    =1 if near 4 yr college, 1966
  4. educ                      years of schooling, 1976
  5. age                       in years
  6. fatheduc                  father's schooling
  7. motheduc                  mother's schooling
  8. weight                    NLS sampling weight, 1976
  9. momdad14                  =1 if live with mom, dad at 14
 10. sinmom14                  =1 if with single mom at 14
 11. step14                    =1 if with step parent at 14
 12. reg661                    =1 for region 1, 1966
 13. reg662                    =1 for region 2, 1966
 14. reg663                    =1 for region 3, 1966
 15. reg664                    =1 for region 4, 1966
 16. reg665                    =1 for region 5, 1966
 17. reg666                    =1 for region 6, 1966
 18. reg667                    =1 for region 7, 1966
 19. reg668                    =1 for region 8, 1966
 20. reg669                    =1 for region 9, 1966
 21. south66                   =1 if in south in 1966
 22. black                     =1 if black
 23. smsa                      =1 in in SMSA, 1976
 24. south                     =1 if in south, 1976
 25. smsa66                    =1 if in SMSA, 1966
 26. wage                      hourly wage in cents, 1976
 27. enroll                    =1 if enrolled in school, 1976
 28. KWW                       knowledge world of work score
 29. IQ                        IQ score
 30. married                   =1 if married, 1976
 31. libcrd14                  =1 if lib. card in home at 14
 32. exper                     age - educ - 6
 33. lwage                     log(wage)
 34. expersq                   exper^2
```

EXAMPLE: Comparing a Sample Mean to a Hypothesized Population Mean
(used for interval-ratio variables)

- Let's start out by looking at the IQ variable: the average IQ in this sample is **102.45**. At least one observation has a minimum IQ of 50; at least one obs has a max IQ of 149.
- The average IQ of all persons in population is equal to 100.
- Does the sample represent a population where the mean is 100?

H_0: The NLS sample comes from a population where the mean IQ is 100.
or
H_0: $\mu_{IQ} = 100$

H_1: The NLS sample comes from a population where the mean IQ is *not* 100.
or
H_1: $\mu_{IQ} \neq 100$

- Choose a level of statistical significance: let's say the 95% level. So, with a sample size *n=2,061* (note: why not 3,010?), the two-sided critical value is $t_{crit} = \pm 1.96$.

- Calculate the test statistic: $t = \frac{(\bar{X} - \mu)}{s/\sqrt{n}} = \frac{102.45 - 100}{15.42/\sqrt{2,061}} = \frac{2.45}{0.3397} = 7.21$

- Compare the absolute value of the calculated t-statistic to the t-critical value: | 7.21 | > |±1.96|

- Conclude that it is highly unlikely that we would have observed a sample mean IQ of 102.45 if in fact the true mean IQ for this population is 100. Thus, we **reject the null hypothesis** that the population mean is equal to 100 for this population.

- Question: Is it possible that this sample was drawn from a population where the population mean is actually equal to 101.5? 102?

- *Alternative ways to test the hypothesis:*

  - **Alternative 1:** We can construct a confidence interval around the sample mean IQ:

$$CI = \bar{X} \pm t\left(s/\sqrt{n}\right) = \bar{X} \pm 1.96\, s_{\bar{X}}$$

$$= 102.45 \pm 1.96 * \left(\frac{15.42}{\sqrt{2061}}\right)$$

$$= 102.45 \pm 1.96 * (0.3397)$$

CI: [101.78, 103.12]

Because we started out with a specific hypothesized population mean IQ (= 100), we can use the confidence interval to make a conclusion from our hypothesis test: the confidence interval we constructed around $\bar{X}$ does *not* include this hypothesized mean value. Thus, we **reject the null hypothesis** and declare it unlikely that our sample came from a population where the mean IQ is actually equal to 100.

Alternative 2: p-values:

- Recall that the p-value is the *exact probability that we would obtain a test statistic as extreme or more extreme (in absolute value for a 2-tail test) than the one we calculated from the sample, assuming that the null hypothesis is true.*

- In interpreting p-values, we still usually use the conventional significance levels of 0.10, 0.05, and 0.01 to compare them to. Stating the specific p-value just gives more precise information about the test statistic in relation to the null hypothesis.

- To find the p-value for the test statistic we calculated above ($t_{calc} = 7.21$), you need to look *inside* the *t* or *Z* table (*Z* is o.k. in this case since the sample size is so large)

- The easiest way to get p-values, though, is to let STATA calculate them for you. It's just important to understand what you're looking at.

- A picture, and the STATA output….

```
. use "J:\STATA datasets\card.dta", clear

. do "C:\DOCUME~1\gppilab\LOCALS~1\Temp\STD01000000.tmp"

. describe, short

Contains data from J:\STATA datasets\card.dta
  obs:         3,010
 vars:            41
 size:       174,580 (98.3% of memory free)
Sorted by:

. describe, simple
id         fatheduc  step14    reg665    south66   wage      libcrd14  collgrd1  wagenew
nearc2     motheduc  reg661    reg666    black     enroll    exper     hsgrad2
nearc4     weight    reg662    reg667    smsa      kww       lwage     hsgrmore
educ       momdad14  reg663    reg668    south     iq        expersq   collgrd2
age        sinmom14  reg664    reg669    smsa66    married   hsgrad1   collmore

. summarize

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
          id |      3010    2581.749    1500.539          2       5225
      nearc2 |      3010    .4408638    .4965731          0          1
      nearc4 |      3010    .6820598    .4657535          0          1
        educ |      3010    13.26346    2.676913          1         18
         age |      3010     28.1196    3.137004         24         34
-------------+--------------------------------------------------------
    fatheduc |      2320    10.00345    3.720737          0         18
    motheduc |      2657    10.34814    3.179671          0         18
      weight |      3010    321185.3    170645.8      75607    1752340
    momdad14 |      3010    .7893688    .4078247          0          1
    sinmom14 |      3010    .1006645    .3009339          0          1
-------------+--------------------------------------------------------
      step14 |      3010    .0388704    .1933182          0          1
      reg661 |      3010    .0465116    .2106253          0          1
      reg662 |      3010    .1607973     .367405          0          1
      reg663 |      3010    .1956811      .39679          0          1
      reg664 |      3010    .0641196    .2450066          0          1
-------------+--------------------------------------------------------
      reg665 |      3010    .2083056     .406164          0          1
      reg666 |      3010    .0960133    .2946584          0          1
      reg667 |      3010    .1099668    .3129003          0          1
      reg668 |      3010    .0282392     .165683          0          1
      reg669 |      3010    .0903654    .2867522          0          1
-------------+--------------------------------------------------------
     south66 |      3010    .4142857    .4926801          0          1
       black |      3010    .2335548    .4231624          0          1
        smsa |      3010    .7129568    .4524571          0          1
       south |      3010    .4036545    .4907113          0          1
      smsa66 |      3010    .6495017    .4772053          0          1
-------------+--------------------------------------------------------
        wage |      3010    577.2824    262.9583        100       2404
      enroll |      3010    .0923588    .2895799          0          1
         kww |      2963    33.54067    8.611619          4         56
          iq |      2061    102.4498    15.42376         50        149
     married |      3003    2.271395    2.066823          1          6
-------------+--------------------------------------------------------
    libcrd14 |      2997     .674341    .4686987          0          1
       exper |      3010    8.856146    4.141672          0         23
       lwage |      3010    6.261832    .4437976    4.60517   7.784889
     expersq |      3010    95.57907    84.61831          0        529
     hsgrad1 |      3010    .8348837    .3713472          0          1
-------------+--------------------------------------------------------
    collgrd1 |      3010    .2714286    .4447705          0          1
     hsgrad2 |      3010    .3295681    .4701344          0          1
    hsgrmore |      3010     .233887    .4233715          0          1
    collgrd2 |      3010    .1524917    .3595566          0          1
    collmore |      3010    .1189369     .323768          0          1
```

```
-------------+--------------------------------------------------------
     wagenew |      3010    5.772824    2.629583          1      24.04

. summarize wagenew

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
     wagenew |      3010    5.772824    2.629583          1      24.04
```

****** EXAMPLE 1 ******

```
. ttest iq==100

One-sample t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
      iq |    2061    102.4498    .3397435    15.42376    101.7835    103.1161
------------------------------------------------------------------------------
    mean = mean(iq)                                               t =   7.2107
Ho: mean = 100                                   degrees of freedom =     2060

   Ha: mean < 100              Ha: mean != 100               Ha: mean > 100
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

****** EXAMPLE 2 ******

```
. tabulate married

    married |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |      2,144       71.40       71.40
          2 |         14        0.47       71.86
          3 |          3        0.10       71.96
          4 |        155        5.16       77.12
          5 |        102        3.40       80.52
          6 |        585       19.48      100.00
------------+-----------------------------------
      Total |      3,003      100.00
```