Lecture Notes 10 Econ 29000, Principles of Statistics Kevin R Foster, CCNY Spring 2011

Examples of hypothesis testing with t-distributions

- 1. For a t Distribution with sample average of 1.98, standard deviation of 1.37, and 8 observations, what is the area in both tails, for a null hypothesis of zero mean ?
- 2. For a t Distribution with sample average of 2.76, standard deviation of 1.16, and 28 observations, what is the area in both tails, for a null hypothesis of zero mean ?
- 3. For a t Distribution with sample average of 0.85, standard deviation of 0.37, and 5 observations, what is the area in both tails, for a null hypothesis of zero mean ?
- 4. For a t Distribution with 22 observations and standard deviation of 2.12, what sample mean leaves 0.10 in the two tails?
- 5. For a t Distribution with 9 observations and standard deviation of 1.19, what sample mean leaves 0.05 in the two tails?
- 6. For a t Distribution with 30 observations and standard deviation of 2.95, what sample mean leaves 0.01 in the two tails?
- 7. Sample A has mean 1.92, standard deviation of 2.24, and 26 observations. Sample B has mean 3.57, standard deviation of 1, and 29 observations. Test the null hypothesis of no difference.
- 8. Sample A has mean 2.16, standard deviation of 1.06, and 17 observations. Sample B has mean 2.69, standard deviation of 0.02, and 16 observations. Test the null hypothesis of no difference.
- 9. Sample A has mean 2.96, standard deviation of 0.89, and 18 observations. Sample B has mean 0.11, standard deviation of 2.89, and 12 observations. Test the null hypothesis of no difference.

Extra to work on:

- 10. For a t Distribution with sample average of 4.41, standard deviation of 0.35, and 7 observations, what is the area in both tails, for a null hypothesis of zero mean ?
- 11. For a t Distribution with sample average of 1.16, standard deviation of 2.7, and 7 observations, what is the area in both tails, for a null hypothesis of zero mean ?
- 12. For a t Distribution with sample average of 0.03, standard deviation of 2.35, and 5 observations, what is the area in both tails, for a null hypothesis of zero mean ?
- 13. For a t Distribution with 24 observations and standard deviation of 1.25, what sample mean leaves 0.56 in the two tails?
- 14. For a t Distribution with 11 observations and standard deviation of 0.05, what sample mean leaves 0.48 in the two tails?

- 15. For a t Distribution with 4 observations and standard deviation of 0.48, what sample mean leaves 0.92 in the two tails?
- 16. For a t Distribution with 8 observations and standard deviation of 1.05, what sample mean leaves 0.69 in the two tails?
- 17. Sample A has mean 4.9, standard deviation of 2.19, and 27 observations. Sample B has mean 4.48, standard deviation of 0.84, and 5 observations. Test the null hypothesis of no difference.
- 18. Sample A has mean 3.24, standard deviation of 0.2, and 9 observations. Sample B has mean 0.49, standard deviation of 1.96, and 22 observations. Test the null hypothesis of no difference.
- 19. Sample A has mean 0.01, standard deviation of 0.01, and 15 observations. Sample B has mean 4.04, standard deviation of 2.6, and 2 observations. Test the null hypothesis of no difference.
- 20. Sample A has mean 0.34, standard deviation of 2.14, and 27 observations. Sample B has mean 4.94, standard deviation of 1.1, and 18 observations. Test the null hypothesis of no difference.
- 21. Sample A has mean 0.54, standard deviation of 0.71, and 20 observations. Sample B has mean 3.09, standard deviation of 0.14, and 27 observations. Test the null hypothesis of no difference.
- 22. Sample A has mean 2.31, standard deviation of 2.98, and 23 observations. Sample B has mean 0.85, standard deviation of 2.65, and 12 observations. Test the null hypothesis of no difference.
- 23. Sample A has mean 3.95, standard deviation of 2.33, and 30 observations. Sample B has mean 4.59, standard deviation of 0.14, and 2 observations. Test the null hypothesis of no difference.
- 24. Sample A has mean 3.99, standard deviation of 1.59, and 26 observations. Sample B has mean 0.1, standard deviation of 1.93, and 4 observations. Test the null hypothesis of no difference.
- 25. Sample A has mean 3.95, standard deviation of 2.56, and 22 observations. Sample B has mean 2.23, standard deviation of 0.82, and 18 observations. Test the null hypothesis of no difference.

You have doubtless noticed by now that much of the basic statistical reasoning comes down to simply putting the numbers into a basic form, $\frac{\overline{X} - \mu}{\sigma_{\overline{x}}}$, where the X-bar is the sample average,

the Greek letter mu, μ , is the value from the null hypothesis, and the Greek letter sigma, σ , is the standard error of the measurement X-bar. This basic equation was used to transform a variable with a Normal distribution into a variable with a Standard Normal distribution (μ =0 and σ =1) by subtracting the mean and dividing by the standard deviation.

It might not be immediately clear that even when we form hypothesis tests on the differences between two samples A and B, so we compare \overline{X}_A with \overline{X}_B , that we're using that same form.

But it should be clearer if we notate the difference as D, where $D = \overline{X}_A - \overline{X}_B$, so the test statistic general form is $\frac{D-\mu}{\sigma_D}$, where we usually test the null hypothesis of μ =0 so it drops out of the equation. The test statistic, $\frac{\overline{X}_A - \overline{X}_B}{\sigma_D}$, then, is really the usual form, $\frac{(\overline{X}_A - \overline{X}_B) - 0}{\sigma_D}$ but without writing the zero. Then we use the formula already derived to get the standard error of the difference in means, σ_D

The only wrinkle introduced by the t-distribution is that we take the exact same form, $\frac{\overline{X} - \mu}{\sigma_{\overline{X}}}$,

but if there are fewer than 30 observations we look up the value in a different table (for a t distribution); if there are more than 30 or so, we look up the value in the normal table just like always.

Going forward, when we use a different estimator (something more sophisticated than the sample average) we will create other test statistics (sometimes called t-statistics) of the same

basic form, but for different estimators – call it \tilde{X} . Then the test-statistic is $\frac{\tilde{X} - \mu}{\sigma_{\tilde{X}}}$, where we

use the standard error of this other estimator. So it's not a bad idea, in a beginning course in statistics, to reflexively write down that basic formula and start trying to fill in values. Ask what is the estimator's value in this sample? That's \tilde{X} . What is the value of that statistic, if the null hypothesis were true? That's μ . What's the standard error of the estimator? That's $\sigma_{\tilde{X}}$. Put those values into the formula and roll!

Statisticians, who know the whole Greek alphabet, sometimes say \tilde{X} as "X-wiggle" or "X-twiddle" as if they don't know what a tilde is.

Interpretation

In many arguments, it is important show that a certain estimator is statistically significantly different from zero. But that mere fact does not "prove" the argument and you should not be fooled into believing otherwise. It is one link in a logical chain but any chain is only as strong as its weakest link. If there is strong statistical significance then this means one link of the chain is strong, but if the rest of the argument is held together by threads it will not support any weight. As a general rule, you will rarely use a word like "prove" if you want to be precise (unless you're making a mathematical proof). Instead, phrases like "consistent with the hypothesis" or "inconsistent with the hypothesis" are better, since they remind the reader of the linkage: the statistics can strengthen or weaken the argument but they are not a substitute.

For example if you watch a crime drama on TV you'll see court cases where the prosecutor proves that the defendant does not have an alibi for the time the crime was committed. Does that mean that the defendant is guilty? Not necessarily – only that the defendant cannot be proven innocent by proving that they were somewhere else at the time of the crime.

You could find statistics to show that there is a statistically significant link between the time on the clock and the time I start lecture. Does that mean that the clock causes me to start talking? (If the clock stopped, would there be no more lecture?)

There are millions of examples. In the ATUS data that we've been using, we see that people who are not working have a statistically significant increase in time on religious activities. We find a statistically significant negative correlation between the time that people spend on religious activities and their income. Do these mean that religion causes people to be poorer? (We could go on, comparing the income of people who are unusually devout, perhaps finding the average income for quartiles or deciles of time spent on religious activity.) Of course that's a ridiculous argument and no amount of extra statistical tests of increasing sophistication to show that there is that negative correlation, it doesn't change the essential part of the argument. The conclusion is not "proved" by the statistics. The statistics are "consistent with the hypothesis" or "not inconsistent with the hypothesis" that religion makes people poor. If I wanted to argue that religion makes people wealthy, then these statistics would be inconsistent with that hypothesis.

Generally two variables, A and B, can be correlated for various reasons. Perhaps A causes B; maybe B causes A. Maybe both are caused by some other variable. Or they each cause the other (circular causality). Or perhaps they just randomly seem to be correlated. Statistics can cast doubt on the last explanation but it's tough to figure out which of the other explanations is right.

On Sampling

All of these statistical results, which tell us that the sample average will converge to the true expected value, are extremely useful, but they crucially hinge on starting from a random sample -- just picking some observations where the decision on which ones to pick is done completely randomly and in a way that is not correlated with any underlying variable.

For example if I want to find out data about a typical New Yorker, I could stand on the street corner and talk with every tenth person walking by – but my results will differ, depending on whether I stand on Wall Street or Canal Street or 42^{nd} Street or 125^{th} Street or 180^{th} Street! The results will differ depending on whether I'm doing this on Friday or Sunday; morning or afternoon or at lunchtime. The results will differ depending on whether I standing there asking people to stop and answer questions (if the person doing the sample is wearing a formal suit or

sweatpants, if they're white or black or Hispanic or Asian, if the questionnaire is in Spanish or English, etc).

In medical testing the gold standard is "randomized double blind" where, for example, a group of people all get pills but half get a placebo capsule filled with sugar while the other half get the medicine. This is because results differ, depending on what people think they're getting; evaluations differ, depending on whether the examiner thinks the test was done or not. (One study found that people who got pills that they were told were expensive reported better results than people who got pills that were said to be cheap – even though both got placebos.)

Getting a true random sample is tough. Randomly picking telephone numbers doesn't do it since younger people are more likely to have only a mobile number not a land line. Online polls aren't random. Online reviews of a product certainly aren't random. Government surveys such as the ones we've used are pretty good – some smart statisticians worked very hard to ensure that they're a random sample. But even these are not good at estimating, say, the fraction of undocumented immigrants in a population.

There are many cases that are even subtler. This is why most sampling will start by reporting basic demographic information and comparing this to population averages. One of the very first questions to be addressed is, "Are the reported statistics from a representative sample?"