Vanderbilt University Political Science - Stats I
Fall 2024 - Prof. Jim Bisbee

Lecture 12

12.1 A final topic: small-sample significance tests

- All of our confidence interval building and hypothesis testing has assumed that we have a sample size large enough to be reasonably sure that the CLT applies and our test statistic's sampling distribution approximates the Normal.
- But what happens when our samples are pretty small? Essentially, we need to make some adjustments to the sampling distribution to account for this.
- For the first time in this class, we'll make the simplifying assumption that we are drawing samples from a Normally distributed population (we'll return in a few moments to consider what happens when this assumption is violated). So we assume:
  - $Y_1, Y_2, ...Y_n$ represent a random sample drawn from a Normal population, with $\overline{Y}$ and S_U^2 as the sample mean and our (unbiased) estimator of the population variance, respectively.
  - Goal is to construct a CI for μ (or, equivalently, to conduct hypothesis tests) when $VAR(Y_i) = \sigma^2$ is unknown and the sample size is small.
  - To do this, we need to be able to say something about the sampling distribution of $\overline{Y}$. Again, because we don't have enough n, we can't appeal to the CLT and conclude that it is distributed Normal.
  - What to do instead?
- Well, start with the theorem (proof omitted, see Ch. 6 if you care) that a linear combination of independent, Normally distributed RVs is itself Normally distributed.
- By linear combination, we mean a random variable composed of the sum of the products of a total of J RVs and scalars:

$$\sum_{i=1}^{J} a_i Y_i$$

- $\overline{Y}$ is, of course, one such linear combination, where the a_i 's are each just equal to $\frac{1}{n}$.
- Thus the sampling distribution of $\overline{Y}$ is Normal, as as before we know that $E\left(\overline{Y}\right) = \mu_{\overline{Y}} = \mu$ and $VAR\left(\overline{Y}\right) = \frac{\sigma^2}{n}$.
- Now let's standardize each of the Y_i's, with $Z_i = \frac{Y_i - \mu}{\sigma}$. Now consider the sum of their squares:

$$\sum_{i=1}^{n} Z_i^2 = \sum_i \left(\frac{Y_i - \mu}{\sigma}\right)^2$$

- This sum of squares takes on what's called a **chi-squared $\left(\chi^2\right)$ distribution with n degrees of freedom.**
- More generally, the sum of the squares of any n i.i.d. standard Normal random variables is distributed chi-squared with n degrees of freedom.
  - Just so you know, the chi-squared distribution–like any probablity distribution–has a density function. It happens to look like this:

$$f(y) = \frac{y^{\left(\frac{\nu}{2}-1\right)} e^{-\frac{y}{2}}}{2^{\frac{\nu}{2}} \Gamma\left[\frac{\nu}{2}\right]},$$

    where $\Gamma[\alpha]$ is the gamma function and $\Gamma[\alpha] = \int_0^\infty e^{-u} u^{\alpha-1} du$, and here ν ("nu") is the number of degrees of freedom.
  - This is very complicated. To make more simple, here is what you need to know about the chi-square (draw pdf of chi-square on board):
    - as df increase, chi-square approaches the Normal distribution.
    - The expected value of a chi-square RV is its number of degrees of freedom: $E(X) = E\left(\sum_i Z_i^2\right) = \nu$.
    - For $\nu > 2$, the density function peaks at $\nu - 2$.

12.2 Degrees of freedom

- – By the way, "degrees of freedom" is a characteristic of any statistic that signifies the number of independent pieces of information on which the statistic is based. In gen-

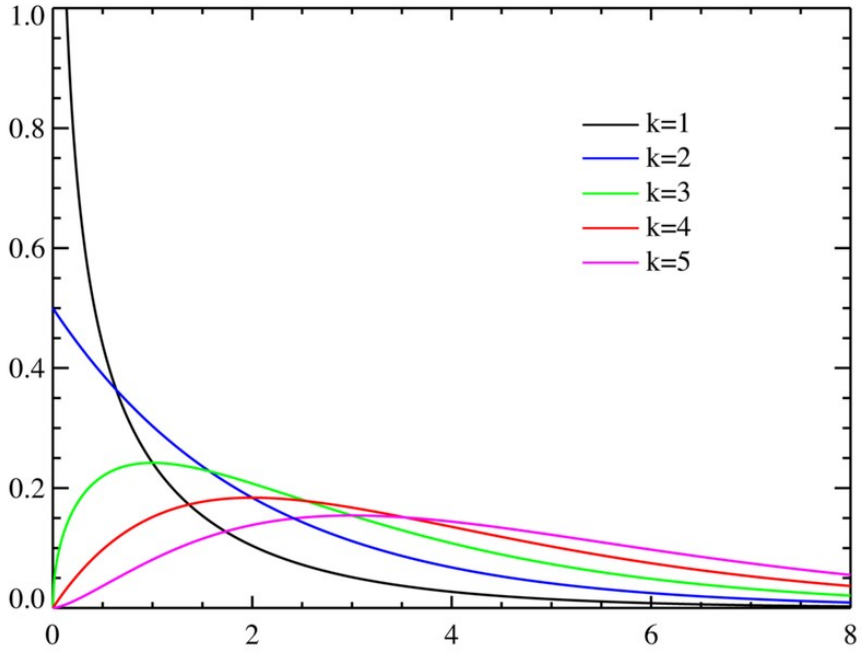

Figure 1:

eral, the degrees of freedom associated with an estimate is equal to the number of pieces of data you have (generally, n) minus the number of parameters needed to generate the estimate.

- Another way to think about it is that a statistic's degrees of freedom is also equal to the number of values in the final calculation of a statistic that are "free to vary."
- For example, if I calculate the mean of three observations, Y_1 and Y_2,and Y_3 as follows:

$$\overline{Y} = \frac{Y_1 + Y_2 + Y_3}{n}$$

this is a statistic with $n = 3$ degrees of freedom. I used three pieces of information to calculate the statistic.

- Compare this to the canonical example is that an estimate of variance using n observations requires that we first estimate the mean and then calculate

$$\begin{aligned} S_U^2 &= \frac{\sum_i \left(\overline{Y} - Y_i\right)^2}{n-1}. \text{ but this is just:} \\ &= \frac{\left(\frac{Y_2+Y_3}{n}\right)^2 + \left(\overline{Y} - Y_2\right)^2 + \left(\overline{Y} - Y_3\right)^2}{n-1}. \end{aligned}$$

  - Thus my calculation of S_U^2 really only uses two pieces of information, because once I've used the three observations to calculate Y-bar, the equation for S_U^2 is fully identified with any two of the observations. So the statistic S_U^2 has d.f. of $n-1$ associated with this estimate.

- Now to return:

- Recall that we are considering the case where $Y_1, Y_2, ...Y_n$ represent a random sample drawn from a Normal population, with $\overline{Y}$ and S_U^2 as the sample mean and our (unbiased) estimator of the population variance, σ^2, respectively. It turns out that the ratio

$$W = \frac{(n-1)\, S_U^2}{\sigma^2} \sim \chi^2_{n-1}$$

  also has a χ^2 distribution with $n-1$ degrees of freedom (proofs omitted).

- Now consider a situation where we divide a standard Normal RV, Z, by the square root of the ratio of a chi-squared RV (W) divided by its degrees of freedom, ν: $\frac{Z}{\sqrt{W/\nu}}$. This is itself a random variable, and it turns out that when Z and W are independent, the quantity

$$T = \frac{Z}{\sqrt{W/\nu}}$$

  has what's called a t distribution with ν degrees of freedom. Why do we care about the t distribution? Well, doing a little math gets us:

$$T = \frac{Z}{\sqrt{W/\nu}} = \frac{\frac{\left(\overline{Y}-\mu\right)}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S_U^2}{\sigma^2}/\,(n-1)}} = \frac{\frac{\sqrt{n}\left(\overline{Y}-\mu\right)}{\sigma}}{\frac{S_U}{\sigma}} = \sqrt{n}\frac{\overline{Y}-\mu}{S_U} = \frac{\overline{Y}-\mu}{S_U/\sqrt{n}}.$$

- Thus Y-bar minus its hypothesized mean divided by Y-bar's estimated standard error, $S_U/\sqrt{n}$, has a "t distribution" with $(n-1)$ d.f.

- That is, we can now fully describe the sampling distribution of Y-bar when we have a small i.i.d. sample of the Normally distributed random variable Y.

- This is the key that allows us to now conduct hypothesis tests with small samples.

– Again, just so you see it, the t-distribution has a density function that looks like this:

$$f(y) = \frac{\Gamma\left[\frac{(\nu+1)}{2}\right]}{\Gamma\left[\frac{\nu}{2}\right]} \cdot \frac{1}{\sqrt{\nu\pi}\left(1+\frac{y^2}{\nu}\right)^{\frac{\nu+1}{2}}},$$

As ν approaches infinity, the t distribution approaches the standard Normal distribution. Draw picture on p. 360.

- These findings allow us to:
  - construct $100(1-\alpha)\%$ CI's around estimates of μ drawn from small samples of a Normally distributed population;
  - perform hypothesis tests with these estimates; and
  - to do the same with estimates of $\mu_1 - \mu_2$.

- For example, for a CI around $\overline{Y}$, an estimate of μ, we proceed as follows:

$$\begin{aligned}
P\left(-t_{\frac{\alpha}{2},\nu} \leq T \leq t_{\frac{\alpha}{2},\nu}\right) &= 1-\alpha, \text{ or} \\
P\left(-t_{\frac{\alpha}{2},\nu} \leq \frac{\overline{Y}-\mu}{\frac{S_U}{\sqrt{n}}} \leq t_{\frac{\alpha}{2},\nu}\right) &= 1-\alpha \\
P\left(-t_{\frac{\alpha}{2},\nu}\frac{S_U}{\sqrt{n}} - \overline{Y} \leq -\mu \leq t_{\frac{\alpha}{2},\nu}\frac{S_U}{\sqrt{n}} - \overline{Y}\right) &= 1-\alpha \\
P\left(t_{\frac{\alpha}{2},\nu}\frac{S_U}{\sqrt{n}} + \overline{Y} \geq \mu \geq -t_{\frac{\alpha}{2},\nu}\frac{S_U}{\sqrt{n}} + \overline{Y}\right) &= 1-\alpha
\end{aligned}$$

- Thus the endpoints of the $100(1-\alpha)\%$ CI are: $\overline{Y} \pm t_{\frac{\alpha}{2},\nu}\frac{S_U}{\sqrt{n}}$. Note how the logic is the same as that which we use to construct the CI used for a large-sample, which is $\overline{Y} \pm z_{\frac{\alpha}{2}}\frac{S_U}{\sqrt{n}}$.

- How about hypothesis tests ? Well, I won't bore you with the details - everything's very

similar to the large-sample test:

$$\begin{aligned} H_0 &: \mu = \mu_0, \\ H_A &: \mu > \mu_0, \mu < \mu_0 \text{ (one tailed)} \\ \mu &\neq \mu_0 \text{ (two-tailed)} \\ \text{Test statistic is:} & \quad T = \frac{\overline{Y} - \mu}{\frac{S_U}{\sqrt{n}}} \\ \text{Reject } H_0 \text{ if } t &> t_{\alpha,\nu} \text{ or } t > -t_{\alpha,\nu} \text{ (one-tailed)} \\ |t| &> t_{\frac{\alpha}{2},\nu} \text{ (two-tailed)} \end{aligned}$$

- Do Examples 10.12 and 10.13 on pp. 521-522.
- And tests for $\mu_1 - \mu_2$?
- Very similar. We assume that our two samples are independent (as we do for large-sample tests). But we typically make an important additional assumption: that the variances of our two populations are the same. That is, $\sigma_1^2 = \sigma_2^2$. Two reasons:
  - a matter of convenience: with small samples it is difficult to get good estimates of small population variances (remember how we need lots of n for consistency property of S_U^2 to kick in)
  - it's a minor sin: since we're already assuming that both populations are Normal; we might as well assume that they have the same variance.
- You'll recall that the test statistic in the large-sample case was

$$Z = \frac{\overline{y}_1 - \overline{y}_2 - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}.$$

- But if we assume $\sigma^2 = \sigma_1^2 = \sigma_2^2$, then we can construct a consistent estimate σ^2 by taking a weighted average of the two sample variances. This weighted average is called "s-squared

pooled" and calculated as

$$s_p^2 = \frac{(n_1 - 1)\, S_{U1}^2 + (n_2 - 1)\, S_{U2}^2}{n_1 + n_2 - 2}.$$

Thus our test statistic is calculated:

$$\begin{aligned} H_0 \quad &: \quad \mu_1 - \mu_2 = 0, \\ H_A \quad &: \quad \mu_1 - \mu_2 > 0,\ \mu_1 - \mu_2 < 0 \ \text{(one tailed)} \\ \mu_1 - \mu_2 \quad &\neq \quad 0 \ \text{(two-tailed)} \end{aligned}$$

$$\text{test statistic is: } T = \frac{\overline{y}_1 - \overline{y}_2 - 0}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},$$

$$\begin{aligned} \text{Reject } H_0 \text{ if } t \quad &> \quad t_{\alpha,\nu} \text{ or } t > -t_{\alpha,\nu} \ \text{(one-tailed)} \\ |t| \quad &> \quad t_{\frac{\alpha}{2},\nu} \ \text{(two-tailed),} \\ &\quad \text{where we have } n_1 + n_2 - 2 \ \text{d.f.} \end{aligned}$$

- If time: do Examples 10.14 and 10.15 on p. 524.
- OK: all of this begs the question: What if the underlying population isn't Normal?
- Statisticians have resorted to empirical studies, where they sample from populations of (known) nonnormal distributions.
  - Moderate departures from normality have little effect on the probability distribution of the test statistic.
  - BUT we are in somewhat treacherous waters here.
- Two more things:
  - because the t is indistinguishable from the Normal at high d.f., a t-test is indisinguishable from a z-test at most levels of n with which we are used to working. This has

led to the ubiquitousness of calling hypothesis tests about μ and $\mu_1 - \mu_2$ "t-tests," even though in most cases they are indistinguishable from z-tests.

– We will revisit t-tests in the context of multivariate regression.

12.3 Summing up inference with one variable

We've spent the past few weeks thinking carefully about the inferences regarding one variable that we can make from a sample to a population. It's helpful to recap the steps we've taken to do this. Let's recap in terms of the goal of making inferences about a population mean μ from a sample statistic Y-bar.

- Assume (big assumption # 1) that we have a random sample, which yields independent, identically distributed observations.
  - *Identicality* assures us that our sample mean, Y-bar, is an unbiased estimator of μ.
- Now we want to know how precise our estimate is. We phrase this question as: how far off, on average, is our estimate typically going to be from the true mean?
- To do this, we need to know the (1) distribution of our estimator and (2) the parameters of the distribution of Y-bar.
- (1) As N becomes large, the Central Limit Theorem tells us that Y-bar is distributed Normal.
- (2) The Normal has two parameters: its mean μ and variance σ^2.
  - Because Y-bar is an unbiased estimator of μ, the mean of Y-bar is μ.
  - If we make the assumption of *independence*, Y-bar's variance is $\frac{\sigma^2}{n}$ and its standard deviation is $\frac{\sigma}{\sqrt{n}}$. We have a consistent, unbiased estimator of σ, which is S_U. A combination of theories allows us to substitute S_U in our estimate of the estimator's standard deviation, that is, $\frac{S_U}{\sqrt{n}}$ for $\frac{\sigma}{\sqrt{n}}$. The resulting quantity $\frac{\overline{Y}-\mu}{\frac{S_U}{\sqrt{n}}}$ converges in probability to the Standard Normal.
- What if we have a small sample? We then make big assumption #2: that Y is distributed Normal. This yields a Y-bar that is distributed T–a distribution that approaches the Normal as N becomes large.

- We now know the pdf and cdf of Y-bar. It is Normal if N is large. And it is T if Y is Normal.
- We can now answer several related questions.
  - *How confident am I about my estimate of Y-bar?* To do this, I identify the values of Y located on either side of Y-bar by an equal distance that between them incorporate (confidence)% of the probability density. I am then "(confidence)% confident" that μ falls in the interval I've created.
  - *How confident that μ is not equal to some hypothesized value μ_0?* To answer this question with "(confidence)% confidence" , I see if μ_0 is found within the confidence interval associated with that level of confidence. If it is, then I am not "(confidence)% confident" that I can reject the null that $\mu = \mu_0$. If it's not, I reject the null at that level of confidence.
  - *At what level of confidence can I be sure that μ is not equal to some hypothesized value μ_0?* To answer this question, I find the smallest level of significance α for which the observed data indicate that the null hypothesis should be rejected. To do this, I find the largest confidence interval around Y-bar that does not contain μ_0. The proportion of the density under the curve contained in this interval is the p-value associated with the hypothesis test.