Inference in Regression Model

Christopher Taber

Department of Economics
University of Wisconsin-Madison

March 7, 2011

Outline

Normality Assumption

We need to add one more assumption to get the final part of the Classical Linear Regression Model

Assumption MLR 6 (Normality)

The population error u is independent of the explanatory variables $x_1, x_2, ..., x_k$ and is normally distributed with zero mean and variance σ^2 : $u \sim N(0, \sigma^2)$

What do you think of this assumption?

It is awfully strong-why should we believe the error term is normally distributed?

Many distributions look like “bell curves” but that does not necessarily mean normality is a good assumption.

When we worry about OLS asymptotics we will see a better way of handling this problem

Why is this assumption useful?

Remember the simple regression model

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{N}(x_i - \bar{x})u_i}{\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

This is a linear combination of normal random variables so it is normally distributed.

Putting the assumption into action gives the following result

Theorem 4.1 (Normal Sampling Distributions)

Under the CLM assumptions MLR.1 through MLR.6, conditional on the sample values of the independent variables

$$\hat{\beta}_j \sim N(\beta_j, Var(\hat{\beta}_j)),$$

where $Var(\hat{\beta}_j)$ was given in Chapter 3 [equation(3.51)]. Therefore,

$$\frac{\left(\hat{\beta}_j - \beta_j\right)}{sd(\hat{\beta}_j)} \sim N(0, 1).$$

We never actually looked at equation 3.51 in the lecture notes, so let me give it now

Theorem 3.2 (Sampling Variance of the OLS Slope Estimators)

$$Var\left(\hat{\beta}_j\right) = \frac{\sigma^2}{SST_j\left(1 - R_j^2\right)}$$

for $1, 2, ..., K$ where $SST_j = \sum_{i=1}^{N}\left(x_{ij} - \bar{x}_j\right)^2$ is the total sample variation in x_j, and R_j^2 is the R-squared from regression x_j on all other independent variables (and including an intercept).

A fly in the Ointment

This is all fine, except for one thing

We don't know σ^2!

We can estimate it in the way you might expect (well close to it)

$$\hat{\sigma}^2 = \frac{1}{N-K-1}\sum_{i=1}^{N}\hat{u}_i^2.$$

This turns out to be an unbiased estimate of σ^2.

We define the standard error of $\hat{\beta}_j$ as

$$se\left(\hat{\beta}_j\right) = \sqrt{\frac{\hat{\sigma}^2}{SST_j\left(1-R_j^2\right)}}$$

Then we have the result that

Theorem 4.2 (t Distribution for the Standardized Estimators)

Under the CLM assumptions MLR.1 through MLR.6 ,

$$\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \sim t_{N-K-1}$$

where $K+1$ is the number of unknown parameters in the population model $y = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + u$ (K slope parameters and the intercept β_0).

Now we don't know β_j, but this won't be particularly important for what we are doing.

Note as well that if N-K-1 is large then the t-distribution is almost identical to the normal distribution

How is this all useful?

Lets forget about regression now and think more generally about inference.

We have talked about how to construct regression estimators and their distribution, but not about statistical inference.

I want to think about this generally and it should be clear that what you learned in your statistics class carries over to the regression model

Lets first think about why we care

Example 1

You want to decide whether to be a doctor or a lawyer.

You have a sample of earnings of people in the population

- For Doctors $\overline{Y} = \$123,452$
- For Lawyers $\overline{Y} = \$112,935$

Should you be a doctor or a lawyer? How confident are you?

Example 2

Basketball Scoring: You go to one basketball game

- Carlos Delfino scores 32 points
- Lebron James scores 17 points

Who is a better?

Example 3

You are a firm thinking about switching to a new mode of production.

It saves about \$1100 on average, but costs \$1000.

Should you buy it?

Example 4

You are trying to predict the stock market.

You think unemployment might be important.

You run a regression of stock gains on unemployment and find $b_2 = 0.01$.

Should you necessarily invest all your money in the stock exchange?

You would like to test formally whether $b_2 > 0$.

Example 5

You want to know whether crime in Chicago is related to the temperature. You run a regression where

A_t : Arrests in Chicago at date t

T_t : Temperature in Chicago at date t

You find that

$$A_t = 143 + 1.3T_t + \varepsilon_t$$

It looks like there are more arrests when the temperature is higher, but are you sure?

Example 6

You are interested in how much cars depreciate. You have information on their age and their price.

P_i : Price of car i in \$1000

A_i : Age of car i

$$P_i = 1.5 - 1.2A_i + \varepsilon_i$$

How do you interpret this? How confident are you?

Outline

Lets forget about regression or sample means or anything else and just assume that we have some estimator $\widehat{\theta}$ that is a function of the data.

This estimator is normally distributed and is an unbiased estimate of θ.

$$\widehat{\theta} \sim N\left(\theta, \sigma_{\theta}^{2}\right)$$

Lets also not worry about T distributions for now, but assume that $N - K - 1$ is large enough that the distribution is approximately normal.

There is a little trick with normal random variables that makes them easy to deal with.

What is the distribution of

$$\frac{\widehat{\theta} - \theta}{\sigma_{\theta}}?$$

First note that θ and σ_θ are nonstochastic while $\widehat{\theta}$ is normally distributed.

Therefore this must be normally distributed.

If we know its expectation and we know its variance we know everything there is to know about it.

$$
\begin{aligned}
E\left(\frac{\widehat{\theta}-\theta}{\sigma_\theta}\right) &= \frac{1}{\sigma_\theta}\left(E\left(\widehat{\theta}\right)-\theta\right)\\
&= 0
\end{aligned}
$$

$$
\begin{aligned}
Var\left(\frac{\widehat{\theta}-\theta}{\sigma_\theta}\right) &= var\left(\frac{1}{\sigma_\theta}\left(\widehat{\theta}-\theta\right)\right)\\
&= \frac{1}{\sigma_\theta^2} var\left(\widehat{\theta}\right)\\
&= \frac{\sigma_\theta^2}{\sigma_\theta^2}\\
&= 1
\end{aligned}
$$

This means that

$$\frac{\widehat{\theta} - \theta}{\sigma_{\theta}} \sim N(0, 1)$$

This is true for any normally distributed random variable.

The distribution $N(0, 1)$ is called a "standard normal"

You can find tables in the back of the book that gives its distribution.

What the tables will show is that for any variable, say Z, that is a standard normal

$$\Pr(-1.96 \leq Z \leq 1.96) = 0.95$$

That is 95% of the time a standard normal random variable will lie between -1.96 and 1.96.

This is true for any standard normal so it has to be true for $\frac{\widehat{\theta}-\theta}{\sigma_\theta}$, thus

$$\Pr\left(-1.96 \leq \frac{\widehat{\theta}-\theta}{\sigma_\theta} \leq 1.96\right) = 0.95$$

Now after doing some algebra, we will get a nice expression

$$\begin{aligned}
.95 &= \Pr(-1.96 \leq \frac{\widehat{\theta}-\theta}{\sigma_\theta} \leq 1.96) \\
&= \Pr(-1.96\sigma_\theta \leq \widehat{\theta}-\theta \leq 1.96\sigma_\theta) \\
&= \Pr(-1.96\sigma_\theta - \widehat{\theta} \leq -\theta \leq 1.96\sigma_\theta - \widehat{\theta}) \\
&= \Pr(-1.96\sigma_\theta + \widehat{\theta} \leq \theta \leq 1.96\sigma_\theta + \widehat{\theta})
\end{aligned}$$

We call this a 95% confidence interval.

If you construct it in this way, it will cover the “true” θ 95% of the time.

It is important to remember that what is random here is $\widehat{\theta}$ and not θ.

Thus it is the interval that is random, not the variable inside it.

Constructing a Confidence Interval for the expected value of a Normal Distribution

Let's start with the case of $Y_i \sim N(\mu_Y, \sigma_Y^2)$ and Y_i and Y_j are independent for $i \neq j$, so $cov(Y_i, Y_j) = 0$.

The sample mean $\overline{Y}$ is

$$\overline{Y} = \frac{1}{N} \sum_{i=1}^{N} Y_i$$

$\overline{Y}$ is a random variable; what is its distribution?

Since it is a linear combination of normal random variables, it's normal. Its mean is

$$E(\overline{Y}) = \mu_Y$$

Its variance is

$$Var(\overline{Y}) = \frac{\sigma_Y^2}{N}$$

The *standard error* of an estimator is its standard deviation (i.e. the square root of its variance), so we also have:

$$se(\overline{Y}) = \frac{\sigma_Y}{\sqrt{N}}$$

Thus

$$\overline{Y} \sim N\left(\mu_Y, \frac{\sigma_Y^2}{N}\right)$$

This is just a special case of what we learned above with $\overline{Y}$ taking the place of $\widehat{\theta}$ and μ_Y taking the place of θ.

Thus

$$\Pr(-1.96\frac{\sigma_Y}{\sqrt{N}} + \overline{Y} \leq \mu_Y \leq 1.96\frac{\sigma_Y}{\sqrt{N}} + \overline{Y})$$

Confidence Intervals for Regression Coefficients

Now we want to construct confidence intervals for regression coefficients.

Once again these are just special cases of $\widehat{\theta}$.

Take β_1 :

$$\hat{\beta}_1 \sim N(\beta_1, Var(\hat{\beta}_1)),$$

Now β_1 takes the place of θ and $\hat{\beta}_1$ takes the place of $\widehat{\theta}$ so that

$$.95 = \Pr(-1.96 se(\hat{\beta}_1) + \hat{\beta}_1 \leq \beta_1 \leq 1.96 se(\hat{\beta}_1) + \hat{\beta}_1)$$

Similarly for any of the other regression coefficients

$$.95 = \Pr(-1.96 se(\hat{\beta}_j) + \hat{\beta}_j \leq \beta_1 \leq 1.96 se(\hat{\beta}_1) + \hat{\beta}_1)$$

Lets look at some examples

Our first example is from the “RDCHEM” data set.

Lets just try the simple regression of profits on Research and Development expenditure

We find that

$$\begin{aligned} \hat{\beta}_1 &= 2.437 \\ se\left(\hat{\beta}_1\right) &= 0.1426 \\ N &= 32 \\ K &= 1 \end{aligned}$$

We want to construct a 95% confidence interval.

N-K-1=30

The critical value for a 2-tailed T test with 30 degrees of freedom is 2.042

The confidence interval is

$$
\begin{aligned}
& \left(-2.042 se\left(\hat{\beta}_1\right) + \hat{\beta}_1, 2.042 se\left(\hat{\beta}_1\right) + \hat{\beta}_1\right) \\
= & \ (-2.042 \times 0.1426 + 2.437, 2.042 \times 0.1426 + 2.437) \\
= & \ (2.146, 2.728)
\end{aligned}
$$

Which is what stata shows.

Now consider a regression of the hours worked in a year

Lets think about a 95% confidence interval for the hourly wage

$$\begin{aligned} \hat{\beta}_1 &= -22.132 \\ se\left(\hat{\beta}_1\right) &= 11.727 \\ N &= 428 \\ K &= 5 \end{aligned}$$

This gives us 422 degrees of freedom which is large enough that this is approximately normal

$$
\begin{aligned}
& \left(-1.96 se\left(\hat{\beta}_1\right)+\hat{\beta}_1, 1.96 se\left(\hat{\beta}_1\right)+\hat{\beta}_1\right) \\
= & \ (-1.96 \times 11.72 - 22.13, 1.96 \times 11.72 - 22.13) \\
= & \ (-45.12, 0.84)
\end{aligned}
$$

Next lets get a 90% confidence interval for the number of kids less than 6.

Again we will be approximately normal, but for a 10% interval the critical value is 1.645.

$$
\begin{aligned}
& \left(-1.645 se\left(\hat{\beta}_4\right) + \hat{\beta}_4, 1.645 se\left(\hat{\beta}_4\right) + \hat{\beta}_4\right) \\
= & \ (-1.645 \times 99.50 - 337.48, 1.645 \times 99.50 - 337.48) \\
= & \ (-501.16, -173.80)
\end{aligned}
$$

Note that this is smaller than the 95% interval

That makes sense

Outline

Hypothesis Testing

First lets think more generally what Hypothesis testing is and then worry about statistical inference

Let H_0 be some hypothesis that you want to test.

Suppose that it is true. We then ask whether the world seems consistent with it.

Specifically: we perform some experiment and see if the results of the experiment are consistent with the hypothesis

Examples

1. There is no gravity
2. I only have one set of keys and I left them in the car
3. I always wear socks

We can easily design experiments and test these hypotheses

In statistics life isn't so easy

We can almost never reject for sure

That is we very rarely know for sure that the null is definitely wrong

For example, consider the hypothesis $H_0 : \mu_x = 0$

Even if $\bar{x} = 1000000$ it is still possible (although unlikely) that the true value is 0

In doing hypothesis testing we formalize these ideas

We must deal with the following:

Type 1 error=$\alpha =$ Probability of rejecting H_0 even when it is true

We would like this to be zero, but it statistics it generally can not be

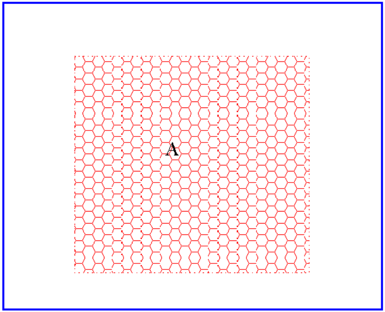
Now we need to construct a test statistic that has to have two features:

1. Depends on the data in a known way
2. We know its distribution when the null hypothesis is true

We then choose an acceptance region A so that when $t \notin A$ we reject the null hypothesis

We choose A in such a way that

$$Pr(\text{reject } H_0 \mid H_0 \text{ true}) = \alpha$$



Lets look at some examples

Example 1: Probability of Getting a Job

Suppose you are out trying to get a job and you think that the probability of getting a job offer is one half.

That is

$$H_0 : Pr(J) = 0.5$$

Now you go out and interview with a bunch of different firms until you get a job.

You want to test the null hypothesis that the probability of getting a job was really 0.5 depending on how long it takes you to find one

Lets define a test statistic t to be the number of jobs for which you interview until you get one

Is this a legitimate test statistic?

- It depends on the data.
- I know its distribution when the null hypothesis is true

To see the last part notice that

Prob. Job at first interview($t = 1$)	0.5
Prob Job at second interview($t = 2$)	0.25
Prob. Job at third interviews($t = 3$)	0.125
Prob. Job at fourth interview($t = 4$)	0.0625
Prob. Job at fiifth interview($t = 5$)	0.03125
Prob. Job after more than 5 interviews($t \geq 6$)	0.03125

How do we do this?

Set $\alpha = 0.0625$ and choose our acceptance region

$$A = \{1, 2, 3, 4\}$$

Thus we reject when we reach our fifth interview

$$\begin{aligned} & \text{Probability of getting to fifth interview} \\ = & \text{ Probability of being rejected after 4 interviews} \\ = & \text{ } 0.0625 \\ = & \text{ } \alpha \end{aligned}$$

Example 2: Sample Mean

Suppose

$$x_i \sim N\left(\mu_x, \sigma_x^2\right)$$

Take the null hypothesis to be that μ_x is some particular value,

$$H_0 : \mu_x = x^*$$

Now set

$$t = \frac{\bar{x} - x^*}{\widehat{se(\bar{x})}}$$

Does this satisfy our criteria?

- It depends on the data. We know x^* and we can estimate $\bar{x}$ and $\widehat{se(\bar{x})}$.
- We know its distribution under the null hypothesis. Under the null

$$x_i \sim N(x^*, \sigma_x^2)$$

So we know that

$$\bar{x} \sim N\left(x^*, \frac{\sigma_x^2}{N}\right)$$

and from statistics class you learned that

$$t \sim T_{N-1}$$

(that is a T distribution with N-1 degrees of freedom)

Example 3: Regression Coefficient

Suppose we want to test something about β_1.

In particular suppose we want to test that

$$H_0 : \beta_1 = \beta_1^*$$

What should we use as our test statistic?

- It needs to depend on the data. That is given a set of X's and Y's, I need to be able to get the test statistic
- I need to know its distribution under the null hypothesis.

We know that under the assumption of the classical linear regression model *and* under the null hypothesis

$$\hat{\beta}_1 \sim N(\beta_1^*, se(\hat{\beta}_1)^2)$$

so

$$\frac{\hat{\beta}_1 - \beta_1^*}{se(\hat{\beta}_1)} \sim N(0,1)$$

and

$$t = \frac{\hat{\beta}_1 - \beta_1^*}{\widehat{se(\hat{\beta}_1)}} \sim T_{N-K-1}$$

There is still an issue we have not dealt with at all

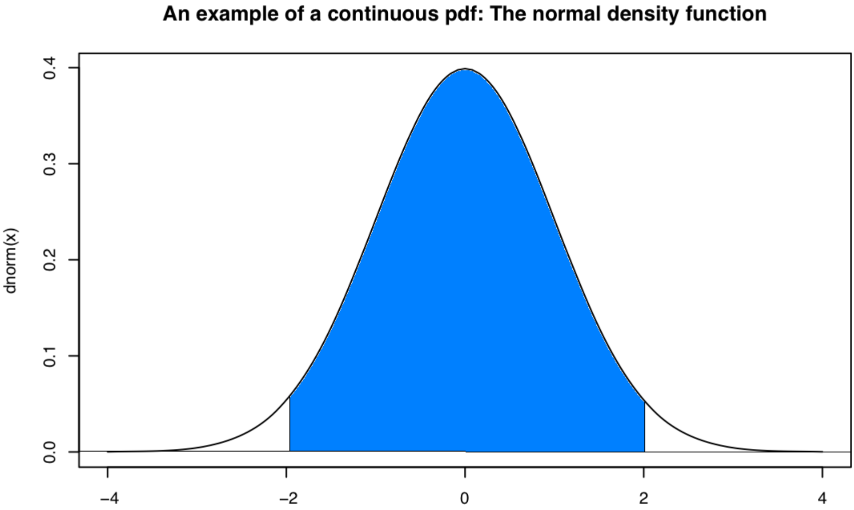
We have talked about how hypothesis testing works and what we need to do it.

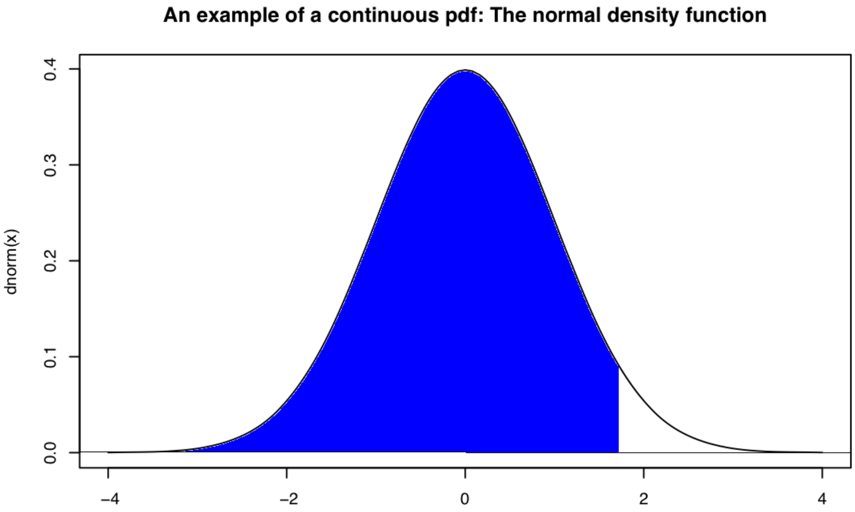
However, we have not talked at all about how we choose the acceptance region

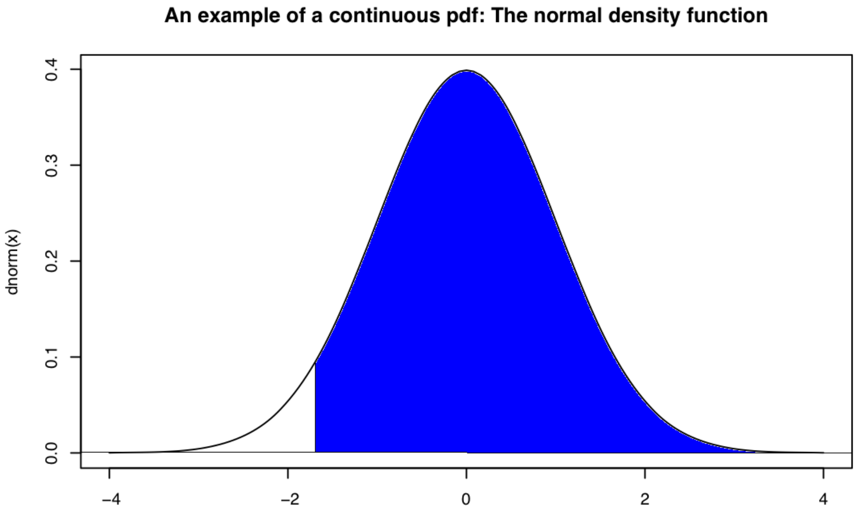
At this point anything will work as long as the probability of rejection is α.

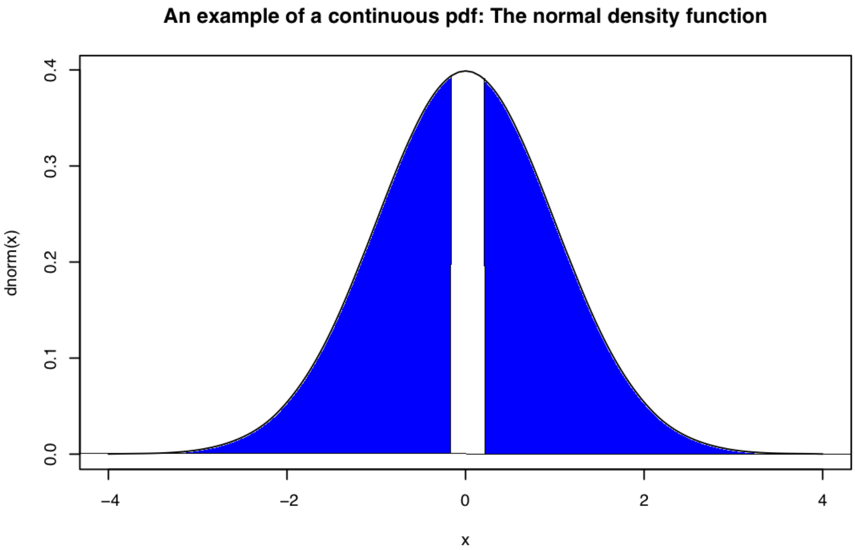
In particular if we know that our distribution is T_{N-K-1}there are many ways to choose the region that will work.

How do we choose the right one?


An example of a continuous pdf: The normal density function
dnorm(x)
0.0
0.1
0.2
0.3
0.4
−4
−2
0
2
4


An example of a continuous pdf: The normal density function
dnorm(x)
0.0
0.1
0.2
0.3
0.4
−4
−2
0
2
4


An example of a continuous pdf: The normal density function
dnorm(x)
0.0
0.1
0.2
0.3
0.4
−4
−2
0
2
4



An example of a continuous pdf: The normal density function
dnorm(x)
0.0
0.1
0.2
0.3
0.4
−4
−2
0
2
4
x

Power

Our main criterion for picking the right region is

$$\beta = Pr(\text{Reject } H_0 \mid H_0 \text{ false})$$

We want to choose the region so that fixing the size$= \alpha$, the power is as large as possible Lets focus on the null hypothesis that

$$\beta_1^* = 0$$

Suppose in reality that the null is false and that instead

$$\beta_1 = 3$$

What is the distribution of the test statistic under this alternative?

$$
\begin{aligned}
t &= \frac{\hat{\beta}_1}{\widehat{se}(\hat{\beta}_1)} \\
&= \frac{3}{\widehat{se}(\hat{\beta}_1)} + \frac{\hat{\beta}_1 - 3}{\widehat{se}(\hat{\beta}_1)} \\
&\sim \frac{3}{\widehat{se}(\hat{\beta}_1)} + T_{N-K-1}
\end{aligned}
$$

This is going to be a T distribution shifted to the right.

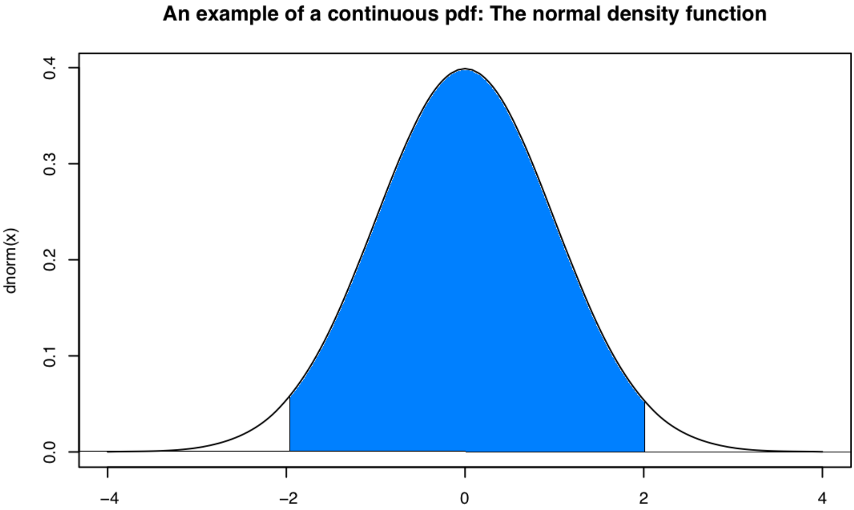
Lets think about the following four 95% acceptance regions

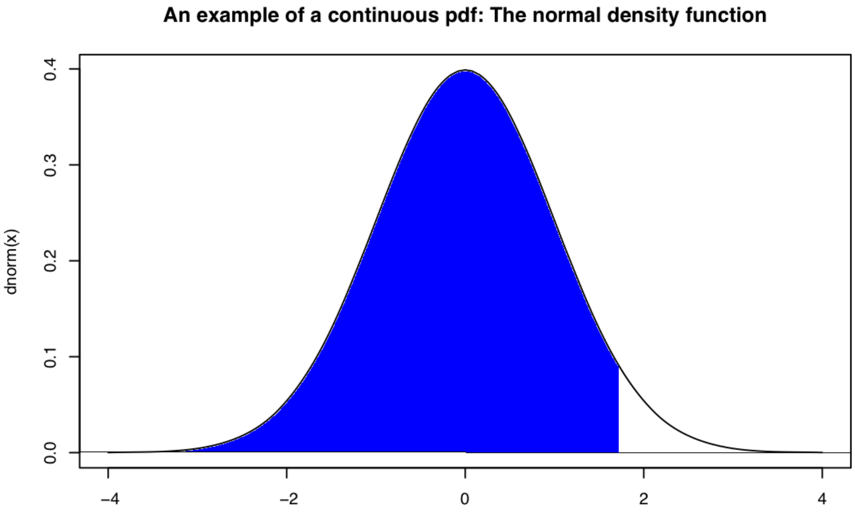
$$\begin{aligned} A &= \{-1.96 \leq t \leq 1.96\} \\ B &= \{t \leq 1.64\} \\ C &= \{-1.64 \leq t\} \\ D &= \{t \leq -0.06, t \geq 0.06\} \end{aligned}$$

Assuming that N-K-1 is large so this is approximately normal, all 4 have the same size $\alpha = 0.05$.

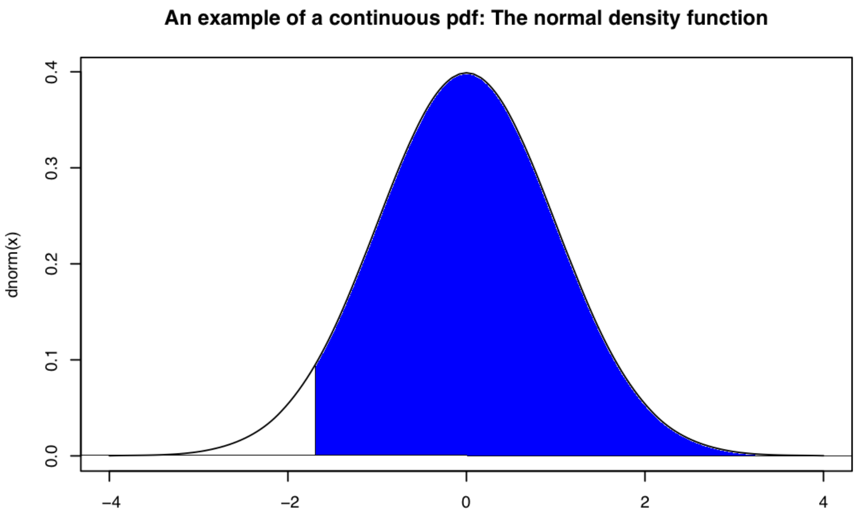
That is under the null hypothesis

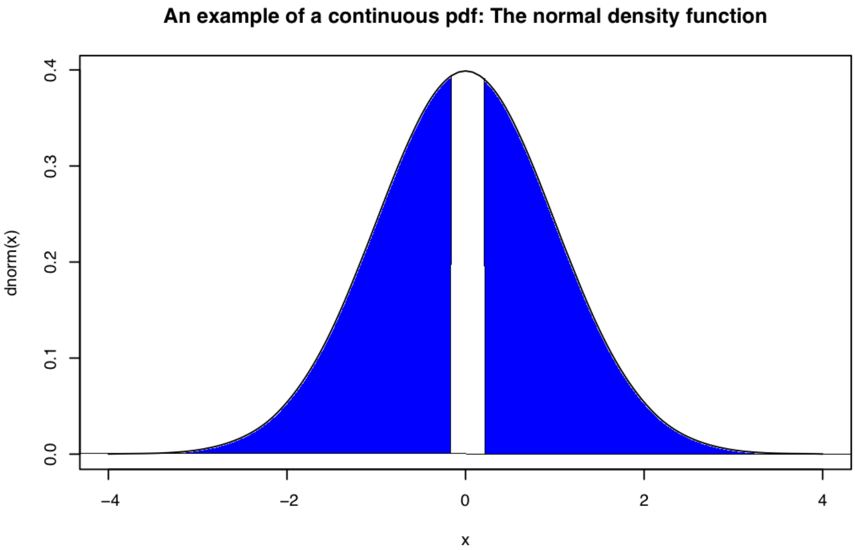
$$
\begin{aligned}
Pr(A \mid H_0) &= Pr\left(-1.96 \leq t \leq 1.96 \mid H_0\right) = 0.95 \\
Pr(B \mid H_0) &= Pr\left(t \leq 1.64 \mid H_0\right) = 0.95 \\
Pr(C \mid H_0) &= Pr\left(-1.64 \leq t \mid H_0\right) = 0.95 \\
Pr(D \mid H_0) &= Pr\left(t \leq -0.06, t \geq 0.06 \mid H_0\right) = 0.95
\end{aligned}
$$


An example of a continuous pdf: The normal density function
dnorm(x)
0.0
0.1
0.2
0.3
0.4
−4
−2
0
2
4



An example of a continuous pdf: The normal density function
dnorm(x)
0.0
0.1
0.2
0.3
0.4
−4
−2
0
2
4


An example of a continuous pdf: The normal density function
dnorm(x)
0.0
0.1
0.2
0.3
0.4
−4
−2
0
2
4


An example of a continuous pdf: The normal density function
dnorm(x)
0.0
0.1
0.2
0.3
0.4
−4
−2
0
2
4
x

For this example which region gives us the most power?

Answer: region B

What about if under the alternative $\beta_1 = -3$?

Region C

What if you are not really sure whether β_1 is positive or negative under the null?

Then you would probably choose region A.

Would you ever choose D?

Not for any reason I can think of

Lets look at a couple of examples.

Cigarette Smoking

Do a higher price of cigarettes lead people to smoke less?

We can write the null hypothesis as

$$H_o : \beta_1 = 0$$

The interesting alternative is $\beta_1 < 0$ so it makes sense to use a one sided test.

N-K-1 is large enough that we will use the normal distribution (degrees of freedom=∞)

Thus the critical value is -1.645. We reject if

$$\begin{aligned} -1.64 >& T \\ =& \frac{-.0317}{0.100} \\ =& -0.32 \end{aligned}$$

We fail to reject the null.

Alternatively we could use a two-sided test. Then we reject if $t < -1.96$ or if $t > 1.96$

We don't reject either of these

We can also try a one-sided test at the 10% level

The critical value here is -1.282

We are still nowhere close

You can see from stata that the p-value is 0.751

You would never use one that large

Wine and Death

Lets look at the relationship between wine consumed and death.

Here you would not really know what to think, so a two sided test makes more sense.

We have 21 observations so the degrees of freedom is 19.

This gives the critical value of a 95% test at 2.093. Calculate

$$
\begin{aligned}
t &= \frac{-16.26}{8.19} \\
&= 1.98
\end{aligned}
$$

We fail to reject the null.

What about a 90% test. In that case the critical value is 1.729

Again we can look at the p-value which is 0.062

This means that

- we would reject a test of size 0.061
- we would fail to reject with a size of 0.063

Even if you reject the null does that mean that you should drink a lot of wine?

Suppose (and I do not know the exact numbers but am making them up), you can either drink wine or do yoga

Say that the doctor tells you running will save 30 lives per 100,000

Is drinking wine as effective?

The null hypothesis now is

$$\beta_1 = -30$$

The test statistic is now

$$\begin{aligned} t &= \frac{\hat{\beta}_1 + 30}{\widehat{se(\hat{\beta}_1)}} \\ &= \frac{-16.26 + 30}{8.20} \\ &= 1.68 \end{aligned}$$

The critical values are the same as before so we fail to reject the null both at the 5% and 10% level

Testing linear combination of parameters

Sometimes we want to test more complicated tests about more than parameters at once

One really nice example is work by Kane and Rouse

They want to ask, what is worth more, a year in community college or a year at a four year college.

They look at the regression

$$log(wage)_i = \beta_0 + \beta_1 jc_i + \beta_2 univ_i + \beta_3 exper_i + u_i$$

where jc_i is the number of years of junior college while $univ_i$ is the number of years of university training and $exper_i$ is experience.

It is interesting to test the null hypothesis

$$H_o : \beta_1 = \beta_2$$

How do we do this?

Note that we can write this as

$$H_o : \beta_1 - \beta_2 = 0$$

Now it looks a bit like what we did before

Remember that

$$\hat{\beta}_1 - \hat{\beta}_2$$

is a sum of normal random variables, so it is normal

$$\begin{aligned} E\left(\hat{\beta}_1 - \hat{\beta}_2\right) &= \beta_1 - \beta_2 \\ var\left(\hat{\beta}_1 - \hat{\beta}_2\right) &= var\left(\hat{\beta}_1\right) + var\left(\hat{\beta}_2\right) - 2cov\left(\hat{\beta}_1, \hat{\beta}_2\right) \end{aligned}$$

Thus under the null hypothesis

$$\frac{\hat{\beta}_1 - \hat{\beta}_2}{\sqrt{var\left(\hat{\beta}_1 - \hat{\beta}_2\right)}} \sim N(0,1)$$

and when we deal with the fact that we estimate the standard error

$$\frac{\hat{\beta}_1 - \hat{\beta}_2}{\sqrt{\widehat{var\left(\hat{\beta}_1 - \hat{\beta}_2\right)}}} \sim T_{N-K-1}$$

There is no easy way to do this in stata, but there are two other ways to do it.

The first way is to redefine the model slightly.

What we care about testing is $\beta_1 - \beta_2$ so lets define

$$\theta_1 = \beta_1 - \beta_2.$$

Also define

$$totcol = jc + univ$$

which is just the total number of years of college that a person has.

Then think about the population regression function

$$\begin{aligned} \mathit{log}(\mathit{wage})_i &= \beta_0 + \beta_1 \mathit{jc}_i + \beta_2 \mathit{univ}_i + \beta_3 \mathit{exper} + u_i \\ &= \beta_0 + (\theta_1 + \beta_2)\, \mathit{jc}_i + \beta_2 \mathit{univ}_i + \beta_3 \mathit{exper} + u_i \\ &= \beta_0 + \theta_1 \mathit{jc}_i + \beta_2\, (\mathit{jc} + \mathit{univ}_i) + \beta_3 \mathit{exper} + u_i \\ &= \beta_0 + \theta_1 \mathit{jc}_i + \beta_2 \mathit{totcol} + \beta_3 \mathit{exper} + u_i \end{aligned}$$

This gives us a nice way of testing the restriction

It also makes sense intuitively.

Outline

The other test one might want to do is to test more than one null hypothesis at a time.

For example in the previous example you might be interested in testing the joint null hypothesis

$$
\begin{aligned}
H_0 : \beta_1 &= 0 \\
\beta_2 &= 0.
\end{aligned}
$$

We don't know how to do this yet.

To get some intuition lets think about comparing a restricted regression model with the unrestricted model.

Lets write the unrestricted model as

$$log(wage)_i = \beta_0^u + \beta_1^u jc_i + \beta_2^u univ_i + \beta_3^u exper_i + u_i^u$$

The restricted model is

$$log(wage)_i = \beta_0^r + \beta_3^r exper_i + u_i^r.$$

Under the null hypothesis these models are equivalent.

Intuitively

- If the null hypothesis is true the two models are the same. That means when we include jc_i and $univ_i$ into the model, the sum of squared residuals should not change much.
- However, if the null hypothesis is false that means that at least one of β_1 or β_2 is nonzero and the sum of squared residuals should fall when we include these new variables.

This seems like it could be the basis of a test.

If the sum of squared residuals changes a lot we reject the null hypothesis.

It only makes sense to to this as a one sided test.

It turns out it works the following statistic

$$F = \frac{\left(SSR_r - SSR_u\right)/q}{SSR_u/(N-K-1)}$$

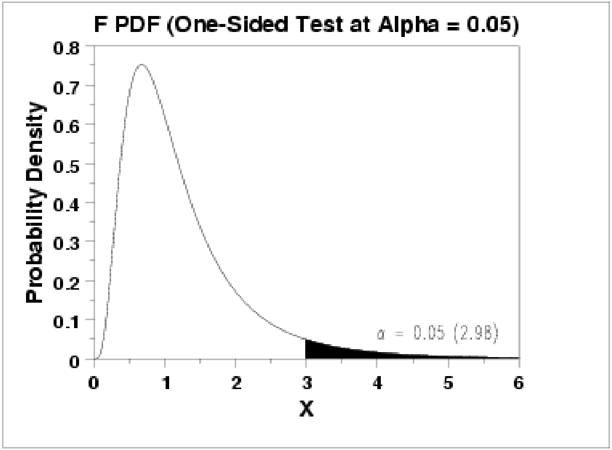
works.

Here q is the difference in degrees of freedom in the restricted model versus the unrestricted model

Equivalently it is the number of restrictions that you are testing in the null hypothesis.

In the case we are looking at, $q = 2$

Note that it increases with SSR_r so we reject when it is a big number.


F PDF (One-Sided Test at Alpha = 0.05)
Probability Density
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
α = 0.05 (2.98)
0
1
2
3
4
5
6
X

We call this an F distribution with q degrees of freedom in the numerator and $N-K-1$ degrees of freedom in the denominator.

We can also write it in terms of R^2,

$$
\begin{aligned}
F &= \frac{(SSR_r - SSR_u)/q}{SSR_u/(N-K-1)} \\
&= \frac{\left(\frac{SSR_r}{SST} - \frac{SSR_u}{SST}\right)/q}{\frac{SSR_u}{SST}/(N-K-1)} \\
&= \frac{\left(R_u^2 - R_r^2\right)/q}{\left(1 - R_u^2\right)/(N-K-1)}
\end{aligned}
$$

This gives us a really easy way to test the joint hypothesis that all of the coefficients in the regression are zero other than the intercept.

That is the null hypothesis

$$H_0 : \beta_1 = \beta_2 = ... = \beta_K = 0.$$

What is the R^2 in the restricted model?

It has to be zero.

Thus the test statistic for this regression is just

$$F = \frac{R^2/q}{\left(1 - R^2\right)/(N - K - 1)}$$

Now back to the two year college example.

I said there was another way to do it.

Well there is nothing that says you can't use an F-test when you only have one restriction ($q = 1$).

For the community college model the restricted model would have $\beta_1 = \beta_2$ so we can write this as

$$\begin{aligned}
log(wage)_i &= \beta_0 + \beta_1 jc_i + \beta_2 univ_i + \beta_3 exper + u_i \\
&= \beta_0 + \beta_1 jc_i + \beta_1 univ_i + \beta_3 exper + u_i \\
&= \beta_0 + \beta_1 (jc_i + univ_i) + \beta_3 exper + u_i \\
&= \beta_0 + \beta_1 totcol + \beta_3 exper + u_i
\end{aligned}$$

We have

$$
\begin{aligned}
SSR_u &= 1250.54352 \\
SSR_r &= 1250.94205 \\
q &= 1 \\
N - K - 1 &= 6760
\end{aligned}
$$

Thus we get

$$
\begin{aligned}
F &= \frac{(SSR_r - SSR_u)/q}{SSR_u/(N-K-1)} \\
&= \frac{(1250.94205 - 1250.54352)/1}{1250.54352/6760} \\
&= 2.15
\end{aligned}
$$

The critical value for an F-test with 1 degree of freedom in the numerator and ∞ in the denominator is 3.84 so we fail to reject the null hypothesis.

Relationship between F and T tests

It turns out that this F-test is closely related to the t-test

In fact the the t-statistic squared has an F distribution with 1 degree of freedom in the numerator.

Note that $1.96^2 = 3.84$ the critical value of the F-test.

Testing in stata

Stata does things slightly different than the way I presented it.

Lets look at a number of examples

Outline

Example 1: Baseball Salaries

Use the data mlb1 from Wooldridge

First test the overall significance of the regression

$$\begin{aligned} F &= \frac{R^2/q}{\left(1-R^2\right)/(N-K-1)} \\ &= \frac{0.6278/5}{(0.3722)/(347)} \\ &= 117.06 \end{aligned}$$

You can see that stata did it for you already

Test whether the performance variables are jointly zero That is

$$H_0 : \beta_3 = \beta_4 = \beta_5 = 0$$

Run the restricted model which gives an $R^2 = 0.5791$ Then you get

$$\begin{aligned} F &= \frac{\left(R_u^2 - R_r^2\right)/q}{\left(1 - R_u^2\right)/(N - K - 1)} \\ &= \frac{(0.6278 - 0.5971)/3}{(0.3722)/(347)} \\ &= 9.54 \end{aligned}$$

The critical value is 2.6 so we reject

Example 2: Firm size and participation rate

Lets look at the relationship between various things and participation rates in 401K plans

Does type of plan matter?

Does size of firm matter?

Even though firm size is statistically significant it is not "economically significant."

Example 3: The Death Penalty

The death penalty is an extremely controversial policy

Proponents claim it deters crime

Opponents claim there is no evidence of this

Lets look at the evidence using death.do and murder.raw

Key variables:

mrdrte	murders per 100,000 population
exec	total executions, past 3 years

One conclusion from this is that there is no evidence that the death penalty deters crime

In fact point estimates go other way, murder rises with executions

Therefore there is no argument for it

This is exactly the type of argument about which one has to be very careful

Just because a coefficient is insignificant doesn't mean it is zero

Confidence interval is (-.218, .0.548)

Lower bound is about -0.2

Interpreting this number as causal would mean that one execution every three years would save 0.2 lives for every 100,000 people each year

Lets try to figure out whether this is a big number or not

The state of Wisconsin is about 5,600,000

That -0.5 would mean that every execution saves about

$$\frac{5,600,000}{100,000} \times 0.2 \times 3 = 33.6$$

lives

This seems like a big number to me

- Can’t reject that deterrent effect is zero
- Also can’t reject that it is really powerful
- Also can’t reject that is is really powerful, but goes in other direction
- In other words this analysis is completely uninformative

I suspect gun control works in exactly the same way (but opposite politically)