

# *Regression #6: Confidence Intervals and Hypothesis Testing*

Econ 671

Purdue University

## Introduction

In the last lecture, we established that, conditional on  $X$  and provided  $\epsilon$  is normally distributed:



How does this result change when we look at *linear combinations* of elements of  $\beta$ ?

To this end, let us consider a  $p \times k$  “selector” matrix  $R$  with  $\text{rank}(R) = p \leq k$ .

## Introduction

For example, setting



leads to  $R\beta$  selecting of the first slope parameter,  $\beta_2$ .

Similarly, setting



leads to  $R\beta$  selecting  $\beta_2 - \beta_3$ .

Finally, choosing



implies that



Lots of other constructions are possible.

## Linear Combinations

We seek to obtain a distribution theory for this arbitrary linear combination of elements of  $\hat{\beta}$ . To this end, we start with



Which implies



or



In our previous lecture, we showed that a quadratic form of the above has a chi-square distribution:



Replacing  $\sigma^2$  with its estimate, we obtain in an analogous manner:



## Linear Combinations

Consider the special case where  $p = 1$ , and  $R$  is a row vector containing a one in the  $i^{\text{th}}$  position and zeros elsewhere. For this special case, we obtain

- 

Or,

- 

Taking the square root of both sides (and letting  $s_{\beta_i}$  denote the standard deviation of  $\hat{\beta}_i$ ;) we obtain the familiar statistic:

- 

a student-t distribution with  $n - k$  degrees of freedom.

## Linear Combinations

$$\frac{\hat{\beta}_i - \beta_i}{s_{\beta_i}} \sim t_{n-k}.$$

Given this result, we can find a value, say  $t_{(\alpha/2), n-k}^*$  such that



(In practice these values are determined by looking up values from a table for various  $\alpha$  and  $n - k$ .) Subbing in for the  $t_{n-k}$  random variable, we can write:



## Linear Combinations

Rearranging, we obtain:



Thus, the interval



is commonly interpreted as a  $100(1 - \alpha)\%$  *confidence interval* for  $\beta_j$ .

## Linear Combinations

Some observations:

- As  $n - k \rightarrow \infty$ ,  $t_{n-k} \rightarrow \mathcal{N}(0, 1)$ . Thus, when our data set is of moderate size (and the number of covariates is modest), we can obtain critical values from the Normal tables rather than the student-t tables.
- Some common choices follow from the Normal tables:

$\alpha = .1$	(90 % Confidence interval)	$t_{.05, \infty}^* = 1.65$
$\alpha = .05$	(95 % Confidence interval)	$t_{.025, \infty}^* = 1.96$
$\alpha = .01$	(99 % Confidence interval)	$t_{.005, \infty}^* = 2.58$



## Joint Confidence Regions

Recall our  $F$ -statistic derived previously:

$$\frac{(R\hat{\beta} - R\beta)' [R(X'X)^{-1}R']^{-1} (R\hat{\beta} - R\beta)}{p\hat{\sigma}^2} \sim F_{p,n-k}.$$

Suppose we are interested in deriving a  $100(1 - \alpha)$  % confidence region for



Naturally, we would construct this region as follows:



where  $F_{p,n-k,\alpha}^*$  is a critical value of the  $F_{p,n-k}$  distribution such that the area to the right of this value equals  $\alpha$ .

In general, these regions are *elliptical* in shape.

## Confidence Regions, Example

To motivate this elliptical shape, consider the specific case where  $k = 2$ ,  $R = I_2$  and  $\sigma^2$  is assumed known and equal to unity.

Further, to fix ideas, suppose



so that  $\rho$  has the interpretation of a correlation parameter.

## Confidence Regions, Example

Multiplying all this out, it can be shown that an approximate 95% Confidence Region is given as:

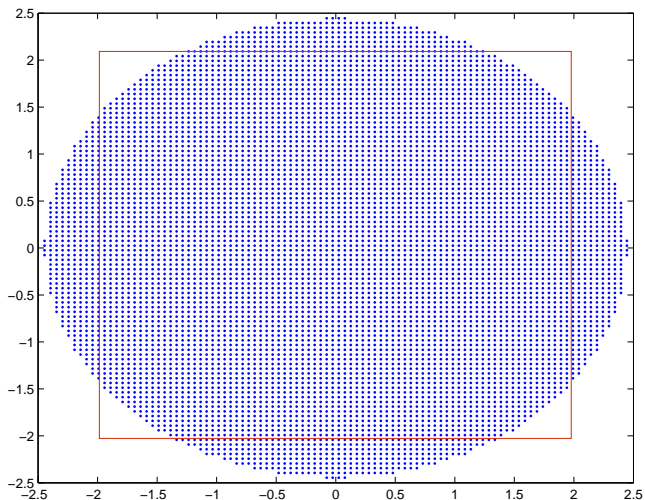


(This makes use of a Chi-square result, which we will demonstrate next lecture). In the following graphs, we provide plots of these regions for two cases: with  $\rho = 0$  and  $\rho = .5$ .

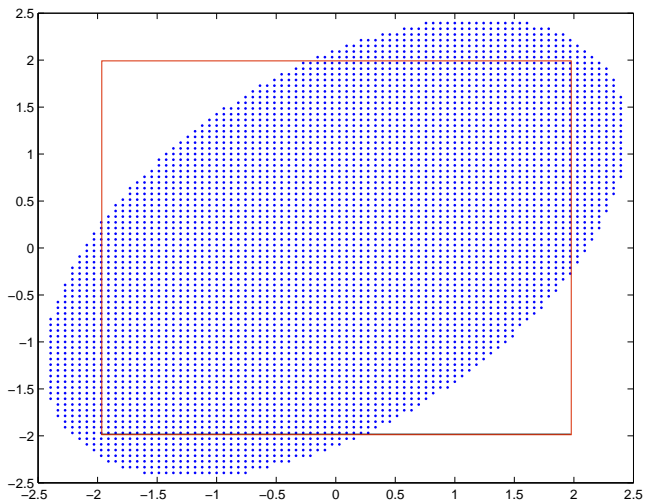
We also provide a plot of a square that would be the naive “joint” region obtained by considering  $\hat{\beta}_1$  and  $\hat{\beta}_2$  separately and piecing together this marginal information.

Finally, we present these plots using  $\beta_j - \hat{\beta}_j$  as the axes so that the plots are centered at zero.

## Joint Confidence Region, $\rho = 0$



## Joint Confidence Region, $\rho = .5$



## Scalar Hypothesis Testing

For a scalar parameter of interest,  $\beta_i$  we derived the result:

$$\Pr \left( \left| \frac{\hat{\beta}_i - \beta_i}{s_{\beta_i}} \right| \leq t_{(\alpha/2), n-k}^* \right) = 1 - \alpha.$$

Thus, we should expect,



100(1 -  $\alpha$ ) percent of the time, in repeated sampling.

## Scalar Hypothesis Testing

So, consider “testing” a null hypothesis of the form:

- 

for some constant  $c$  against the alternative

- 

We would then expect, *if the null were true*:

- 

since this would be true most of the time (specifically,  $100(1 - \alpha)$  percentage of the time) in repeated sampling.

## Scalar Hypothesis Testing

Thus, if



we interpret this as an unusual event, providing evidence against the null hypothesis being true. In this case, we *reject the null* in favor of the alternative  $H_A$ .



## Scalar Hypothesis Testing

A test of particular interest is whether or not a specific covariate  $x_i$  belongs in the model. In this case, it is natural to choose  $c = 0$  and perform the test above.

If we reject this null hypothesis (as we often hope to do), we would reject  $H_0 : \beta_i = 0$  at the  $100\alpha$  % level of significance.

Alternatively, we would say that  $\hat{\beta}_i$  is *statistically significant* at the  $100\alpha$  % level.

## Scalar Hypothesis Testing

It is not, correct, however, to conclude that  $\beta_i = c$  when

$$\left| \frac{\hat{\beta}_i - c}{s_{\beta_i}} \right| \leq t_{(\alpha/2), n-k}^*$$

Indeed, other nulls of the form  $\beta_i = c + \epsilon$  would yield the same result. Can we conclude that  $\beta_i = c$  and  $\beta_i = c + \epsilon$ ?

For this reason, it is proper to say in this instance that we *fail to reject* the null, as the data simply has not provided sufficient evidence to conclude that the null is (probably) false.

## Scalar Testing and p-values

- Suppose you run a regression and implement a test of the hypothesis that  $\beta_i = 0$ .
- When doing so, you find that you reject  $H_0 : \beta_i = 0$  when setting  $\alpha = .1$ , but fail to reject when setting  $\alpha = .05$ .
- What should you do?
- The fate of your paper hangs in the balance!

## Scalar Testing and $p$ -values

A common practice in this situation is to calculate and report a  $p$ -value or *exact level of significance*.

This (two-sided)  $p$ -value is defined as the (sampling) probability of getting a value of the test statistic that is at least as extreme as the one you obtained with your actual sample of data, given that the null is true (whew!).

Formally, suppose we wish to test  $H_0 : \theta = \theta_0$ . Doing so produces an observed test statistic equal to  $t(y)$ . The  $p$ -value is:



## Scalar Testing and p-values

The “appeal” is that the p-value provides the level of significance, had it been chosen prior to seeing the data, that would lead the researcher to be just indifferent between rejecting and failing to reject. This essentially places the burden on the reader to decide whether or not to “believe” that  $\theta = \theta_0$ .

## Scalar Testing and $p$ -values

- Small  $p$ -values are often considered to provide evidence against  $H_0$ , since we would fail to reject only under very small values of  $\alpha$ .
- Conversely, large  $p$ -values are often considered to provide support for  $H_0$ , since we could not reject at any “reasonable” level of  $\alpha$ .
- In fact, common practice in our profession supports this: when a  $p \in (.05, .10]$ , the associated coefficient estimate gets adorned with a single star. When  $p \in (.01, .05]$ , it gets *two stars*!! And, when  $p \leq .01$ , it gets **THREE STARS!!!** All other coefficients are left undecorated and unwanted.

## Scalar Testing and $p$ -values

Despite this convention, I think it is important to emphasize a few things:

- The  $p$  value *does not* represent the probability of the null hypothesis being true. (And, therefore,  $1 - p$  does not represent the probability that the alternative is true).
- In fact, that statement itself is not well-posed in classical statistics: the null is either true or not true. The  $p$ -value does not represent a degree of belief (or disbelief) about the statement in question.

## Scalar Testing and $p$ -values

Consider the following (adapted from Jacob Cohen “The Earth is Round ( $p < .05$ )” *American Psychologist*, 1994).

The logic of the  $p$ -value goes something like this:

- If the hypothesis is true, I am very unlikely to observe a certain feature/statistic of the data.
- I observed that feature of the data.
- Therefore, the hypothesis is unlikely to be true.



## Scalar Testing and p-values

But does this make sense in the context of the following?

$H_0$  : Person is over 17

- It is very unlikely that a (randomly selected) person over 17 will have attended Purdue University.
- I sample an individual, and observe that she attended Purdue.
- Therefore, the person is unlikely to be over 17.

Wouldn't you make *exactly the opposition conclusion*, upon observing college attendance?

Let's cook up some numbers, just to illustrate the point.

Let  $P = 1$  indicate the event that Purdue is attended. Let  $A = 1$  denote the event that age is over 17. Suppose the following joint distribution describes this pair of outcomes:

	P=1	P=0
A=0	0	.2
A=1	.001	.799

From this table, we see that



(so attending Purdue *IS* unlikely if you are over 17).

However,



That is, *We are certain that the person is over 17, given that she attended Purdue!*

The source of the problem on the previous page is that the  $p$ -value relates to the probability of the data under the null hypothesis  $p(y|H_0)$  while we often (mistakenly) interpret it as the probability of the hypothesis given the data:  $p(H_0|y)$ .

Bayesian statistics and econometrics provides a vehicle for making probability statements about hypotheses - more on that later (if you are sufficiently misguided to seek more).

Consider another critical example associated with the  $p$  – value (think about drug trials to fix ideas).

One set of independent observations produces a point estimate of a parameter of interest  $\theta_1$  as  $\hat{\theta}_1 = 25$ , and  $Std.Err(\hat{\theta}_1) = 10$ , leading the researcher to conclude that  $\theta_1 \neq 0$ . (**SIGNIFICANT!**)

A second set of independent observations for  $\theta_2$  produces  $\hat{\theta}_2 = 10$ , and  $Std.Err(\hat{\theta}_2) = 10$ , leading the researcher to fail to reject  $H_0 : \theta_2 = 0$ . (**INSIGNIFICANT!**)

When comparing  $\theta_1$  and  $\theta_2$ , we obtain:  $\hat{\theta}_1 - \hat{\theta}_2 = 15$ ,  
 $Std.Err(\hat{\theta}_1 - \hat{\theta}_2) = \sqrt{100 + 100} = \sqrt{200} \approx 14$  and thus we cannot reject  $H_0 : \theta_1 = \theta_2$ . (**INSIGNIFICANT!**)

How do we reconcile the following statements? (For more, see Andrew Gelman, “*P* Values and Statistical Practice” *Epidemiology*, 2013).

- (a) Drug one has beneficial health effects, and we reject the position that it is ineffective.
- (b) We can't reject the position that drug 2 is ineffective.
- (c) We can't conclude that drug 1 and drug 2 are different from one another.

A final observation notes that the  $p$ -value is also, at least in part, sample-size dependent. To fix ideas, suppose



(with  $\sigma^2$  known just to make things simple) so that



Suppose we wish to test the null  $H_0 : \mu = c$ , and collect a sample of  $n$  observations yielding the observed sample mean  $\bar{x}_n^o$ . Then, the  $p$ -value associated with this statistic and null hypothesis is:



since  $\bar{x}_n \xrightarrow{P} \mu$ , for  $\mu \neq c$  it is clear that  $p \rightarrow 0$  as the sample size grows.

## Other observations

- The “significance” game is, unfortunately, widespread in economics. Note, among other things that coefficients can be “made” to be significant by simply acquiring more data. Such behind the scene endeavors are surely common, though misplaced.
- *Pretesting* is also an important, though hidden concern. This practice refers to the selection of covariates after a series of initial regressions, where coefficients with low  $t$ -stats are dropped, or models with low  $R^2$  values are discarded. The reported uncertainty measures for the selected model are then incorrect; indeed the probability of obtaining a “significant” coefficient in such a pursuit may be unity!

## Other observations

- Keep in mind the importance of *economic* rather than *statistical* significance.

A coefficient may have a moderate  $t$  statistic of, say, 1.5 and its interval estimate may contain both zero and meaningfully large values.

Conversely, (particularly with large data sets), a coefficient may have huge  $t$ -statistic, though its point and interval estimates cover sufficiently small values that it makes little difference in practice (i.e., no economic or practical significance).



## Joint Testing

To finish this discussion, we quickly discuss how a joint hypothesis of the form:

$$R\beta = r$$

would be implemented. To this end, we go back to our test statistic and note that, if the null is true, then

$$(R\hat{\beta} - r)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r) \leq p\hat{\sigma}^2 F_{p,n-k,\alpha}^*$$

100(1 -  $\alpha$ ) % of the time in repeated sampling. Thus a value of the test statistic

$$\frac{(R\hat{\beta} - r)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r)}{p\hat{\sigma}^2}$$

**LARGER** than the critical value leads to rejection of  $H_0 : R\beta = r$  at the given level of significance, and otherwise we fail to reject.