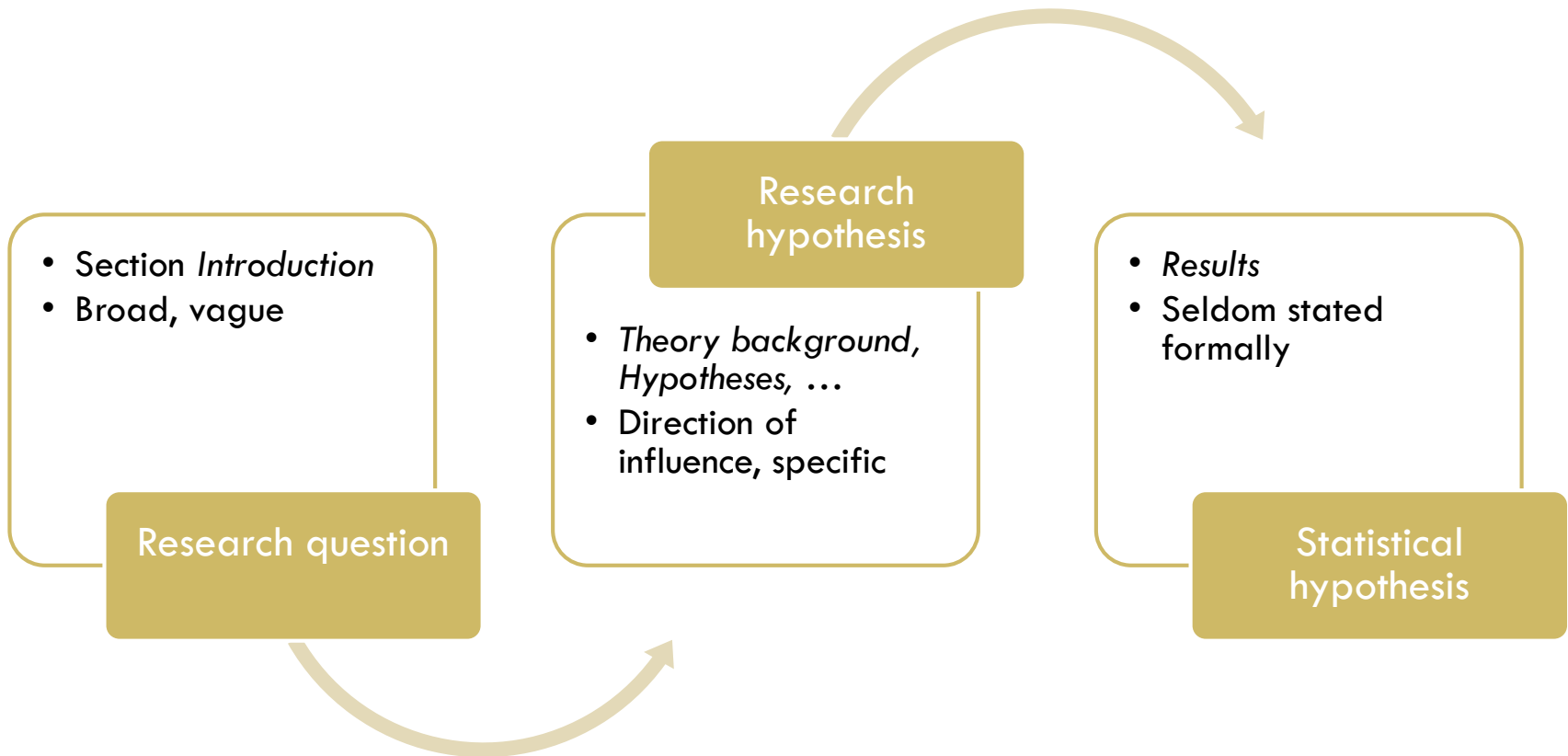


LECTURE 5:  
HYPOTHESIS TESTING

Jan Zouhar

Introductory Econometrics

# Hypothesis testing in empirical research





ELSEVIER

Contents lists available at ScienceDirect

# Journal of Business Venturing



2017 IF: 6.000

2017 AIS: 2.626

## Funders' positive affective reactions to entrepreneurs' crowdfunding pitches: The influence of perceived product creativity and entrepreneurial passion



Blakley C. Davis<sup>a,\*</sup>, Keith M. Hmieleski<sup>b</sup>, Justin W. Webb<sup>c</sup>, Joseph E. Coombs<sup>a</sup>

<sup>a</sup> Department of Management, Virginia Commonwealth University, 301 West Main St., Richmond, VA 23284, United States

<sup>b</sup> Department of Management, Entrepreneurship, and Leadership, Texas Christian University, TCU Box 298530, Fort Worth, TX 76129, United States

<sup>c</sup> Department of Management, University of North Carolina at Charlotte, Friday Building 308A, Charlotte, NC 28223, United States

### ARTICLE INFO

#### Article history:

Received 4 October 2015

Received in revised form 28 October 2016

Accepted 31 October 2016

Available online xxxx

#### Keywords:

Positive affect

Affective events theory

Crowdfunding

Creativity

### ABSTRACT

This study draws upon affective events theory, research regarding funders' perceptions, and research regarding expectation alignment between products and their presenters to develop and test an indirect effects model of crowdfunding resource allocation decisions. To test our hypothesized relationships, we drew upon a sample of 102 participants who each assessed ten different product pitches made by ten different entrepreneurs. Results from the study indicate that perceived product creativity is positively related to crowdfunding performance, both directly and indirectly, via positive affective reactions of prospective funders. Moreover, we find the indirect effect of product creativity is contingent upon the extent to which funders perceive an entrepreneur to be passionate, such that perceived entrepreneurial passion increases the positive nature of the indirect effect. Implications for future theory development, empirical research and implications for practitioners are discussed as well.

## 2. Introduction

Rewards-based crowdfunding represents a relatively new and increasingly common conduit for acquiring financial resources. Rather than providing funds for the general development or growth of a new venture, crowdfunding enables entrepreneurs to garner funds in support of a specific purpose, which often centers on the creation or distribution of a new product (Belleflamme et al., 2014). Since entrepreneurs who utilize crowdfunding are often in the midst of developing products that are unfinished or unproven (Parhankangas and Ehrlich, 2014), prospective funders may be required to rely heavily upon perception-based elements of the pitch (e.g., creativity, passion) when determining whether, and to what extent, they will provide monetary support to the entrepreneur (Maxwell et al., 2011). Research examining perception-based elements of pitches has predominately focused on funders' judgments regarding the merits of business related information that might be used to predict the odds of venture success (e.g., Chen et al., 2009). There is, however, some evidence to suggest that, by engendering positive affective reactions in prospective funders, entrepreneurs might also influence funding through emotional aspects of the product pitch (Baron et al., 2006).

New products often represent the focal point in entrepreneurs' crowdfunding pitches. Although a number of perception-based elements may be considered, one that has been consistently found to differentiate successful new products from failures is product creativity (Cooper and Kleinschmidt, 1987). Creativity is highly valued across cultures (Peterson and Seligman, 2004) and is considered elemental to entrepreneurship (Schumpeter, 1947; Ward, 2004). For these reasons, creative products (i.e., those that are deemed to be both novel and meaningful; Andrews and Smith, 1996) have the potential to elicit positive affective reactions in those who view them (Algoe and Haidt, 2009; Baron, 2008). Moreover, the extent of creative products' affect-inducing potential may be influenced by the entrepreneur delivering the pitch. Indeed, individuals are often unable to partition out the affective reactions they have to new product *pitches* from the reaction they have to the actual *products* (Cialdini, 2016). Despite this relationship, research has generally focused on how funders' judgements and affective reactions are independently influenced by their perceptions of the entrepreneur(s) (e.g., Baron et al., 2006; Chen et al., 2009). Thus, relatively little is known with regard to (1) how funders' perceptions of a product's creativity may influence their resource allocation decisions (at least in part) by engendering positive affective reactions; and (2) how funders' affective reactions to product creativity are influenced by perceptions of the entrepreneur(s) delivering the pitches.

To answer these research questions, we integrate affective events theory [AET] (Weiss and Cropanzano, 1996) with research findings regarding the importance of funders' perceptions about elements of observed product pitches (Maxwell et al., 2011; Mitteness et al., 2012), and the alignment in expectations of both products and their presenters (Till and Busler, 2000) to develop an indirect effects model. More specifically, our model examines the relationship between a pitched product's perceived level of creativity, funders' positive affective reactions to that perceived creativity, and, ultimately, their decision to provide capital in



## 5. Hypothesis development

### 5.1. Product creativity and crowdfunding performance

When evaluating funding proposals, particularly during early-stage screening, resource providers may place greater importance on the features of the product (or opportunity) as opposed to the qualities of the entrepreneur (c.f., [Mitteneß et al., 2012](#)). Early-stage ventures are typically based on a single or very small number of products. Moreover, such products are often unfinished or unproven ([Parhankangas and Ehrlich, 2014](#)). For this reason, objective evidence of quality or potential for market success is often lacking and resource providers must instead rely more heavily upon their subjective impressions when making investment decisions ([Maxwell et al., 2011](#); [Maxwell and Lévesque, 2014](#)).

Resource providers may look to a variety of different sources for information in forming judgments, yet product creativity is the one perception-based consideration that has been found to consistently differentiate new product successes from failures. This is because new product failures are often 'me too products' and do not provide clear value-added advantages over existing offerings ([Cooper and Kleinschmidt, 1987](#)). Products that are considered creative must be judged as being both *novel* and *meaningful* relative to existing attributes or practices that are associated with similar types of output ([Andrews and Smith, 1996](#); [Gardner, 1993](#); [Hennessey and Amabile, 1988](#)). Novelty refers to the degree to which the offering differs from existing outputs (i.e., the status quo), whereas meaningfulness refers to the extent to which the offering is perceived to generate value that goes above and beyond existing alternatives. Accordingly, we define *product creativity* as the extent to which a product represents a meaningful, value-creating advancement as compared to existing market offerings in the product category ([Andrews and Smith, 1996](#)).

In the context of crowdfunding, entrepreneurs typically seek funds to support a single product that, in turn, serves as the focal point in the funding pitch (e.g., [Mollick, 2014](#)). Moreover, funders are typically provided the option of receiving various tangible rewards in exchange for their capital, with the pitched product often being one of them.<sup>1</sup> The introduction of creative products often represents a key hurdle on the road to broader acceptance and development of competitive advantage ([Ward, 2004](#)). Since failed products often lack creativity ([Cooper and Kleinschmidt, 1987](#)), prospective funders may judge a product's level of creativity as an indicator of potential market performance ([Szymanski et al., 2007](#)), or the entrepreneur's potential for offering other creative products in the future ([Keller, 2012](#)). Taken together, we hypothesize the following:

**Hypothesis 1.** In the crowdfunding context, product creativity will be positively related to crowdfunding performance.

not eliminate the threat of common method variance, it provides evidence that correlations between variables are not driven purely by method bias (Korsgaard and Roberson, 1995).

## 7. Results

Means, correlations, and standard deviations for the variables are presented in Table 1. Table 2a provides multilevel results for the study's two dependent measures: *investment* and *predicted success*, whereas Table 2b presents the multilevel results for

<sup>3</sup> These results mirror those found when preparedness is included in the confirmatory factor analysis.

positive affect (as the dependent variable). The Sobel test for mediation relating to Hypothesis 4 is provided in Table 3. Finally, the conditional indirect and direct effects of perceived product creativity on crowdfunding performance (via funders' positive affect) relating to Hypothesis 5 are found in Table 4, and the plotted interaction between perceived product creativity and perceived entrepreneurial passion is provided in Fig. 2.

Hypothesis 1 predicted that perceived product creativity would be positively related to crowdfunding performance. As can be seen in Table 2a, the coefficient for product creativity is both positive and statistically significant [(Model 3; Investment:  $B = 0.69$ ,  $p < 0.000$ ) (Model 6; Predicted success:  $B = 0.62$ ,  $p < 0.000$ )], thus lending support for Hypothesis 1. Similarly, Hypothesis 2, which predicted a positive relationship between perceived product creativity and funders' positive affect, also received support [(Table 2b; Model 3  $B = 0.34$ ;  $p < 0.000$ )]. Next, in Table 2a, we find a positive and statistically significant relationship between funders' positive affective reactions and crowdfunding performance [(Model 3; Investment:  $B = 0.23$ ,  $p < 0.01$ ) (Model 6; Predicted success:  $B = 0.05$ ,  $p = 0.08$ )]. Thus, Hypothesis 3 is supported.

Hypothesis 4 predicted that the positive relationship between perceived product creativity and crowdfunding performance

# What is hypothesis testing good for?

7

- hypothesis testing uses statistical evidence to answer questions like:
  - ▣ if we hire more policemen, will car theft rate drop significantly in our city?



- ▣ is there a significant gender discrimination in wages?
  - ▣ does the unemployment vs. GDP growth ratio claimed by Okun's Law hold in a particular economy?
- or more subtle questions like:
  - ▣ does my production function exhibit diminishing returns to scale or not?
  - ▣ does the effect of age on wages diminish over time (i.e., with growing age) or not?

# Statistical Hypothesis Testing: A Reminder

8

- **Example:** testing a hypothesis about population mean
  - to remind you about the principle concepts of hypothesis testing, we'll start with an example of a test about the population mean (something you have definitely seen in your statistics classes)
  - imagine you want to find out whether a new diet actually helps people lose weight or whether it is completely useless (most diets are)
  - you've collected data about 100 people, who had been on the diet for 8 weeks
    - let  $d_i$  denote the difference in the weights of  $i$ th person *after* and *before* the diet:
$$d = \text{weight after} - \text{weight before}$$
  - you're testing the *average* effect of a diet, which means you're making a hypothesis about the *population mean* of  $d$  (i.e., about  $E d$ , which we'll denote  $\mu$  for brevity)
  - first, you need to state the **null** and **alternative hypotheses**



# Statistical Hypothesis Testing: A Reminder (cont'd)

9

- in hypothesis testing, the null hypothesis is something you're trying to *disprove* using the evidence in our data
- in order to show the diet works, you'll actually be *disproving it doesn't*
- therefore, the null hypothesis will be:

$$H_0: \mu = 0 \quad (\text{on average, there's no effect})$$

- the alternative hypothesis is a vague definition of what you're trying to show, e.g.:

$$H_1: \mu < 0 \quad (\text{on average, people lose weight})$$

- next, you look at the average effect of diet in your data and find that, say,  $\bar{d} = -1.5$  (in your sample, on average, people lost 1.5 kg)
- is that a reason to reject  $H_0$ ?
  - we don't know yet
  - even if  $H_0$  is actually true, we would *not* expect the sample average to be *exactly* 0
  - the question is whether  $-1.5$  is sufficiently far away from zero so that we can reject  $H_0$

# Statistical Hypothesis Testing: A Reminder (cont'd)

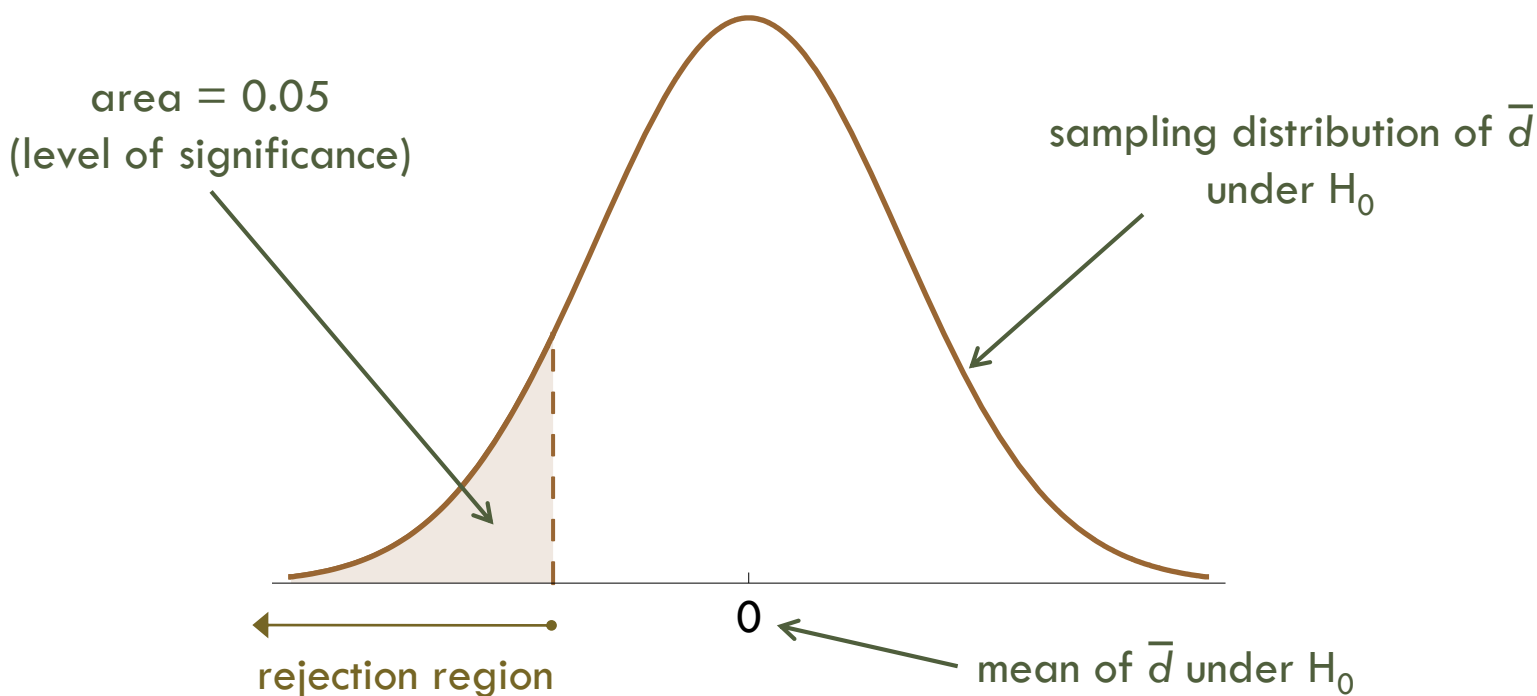
10

- it's not difficult to imagine that even if  $H_0$  is true, you can always end up with an “unlucky” sample with a mean of  $-1.5$
- however, if you use statistics and find out that the probability of obtaining a sample this extreme under  $H_0$  is less than 5% (meaning you get  $-1.5$  or less *in less than 1 in 20 samples*), you'll probably think that you have a strong evidence that  $H_0$  is not true
- the number 5% here is called the **level of significance** of the test
- from here, we can see the following properties of hypothesis testing:
  - we need statistical theory to find the *sampling distribution* of our **test statistic** (in this case, the sample mean)
  - we're working with probabilities all the time: we never know the right answer for sure; it might happen that we reject a null that actually is true (**type I error**)
  - the probability of *type I error* is something we choose ourselves prior to carrying out the test (*level of significance*)
  - once we know these things, we can find the **rejection region**

# Statistical Hypothesis Testing: A Reminder (cont'd)

11

- our test is one-tailed, we reject  $H_0$  only if our statistic is very small:



- Note:* if our test statistic falls out of the rejection region, we use the language “we fail to reject the null at the  $x\%$  level” rather than “the null hypothesis is accepted at the  $x\%$  level”

# Hypotheses about a Single Parameter: *t*-tests

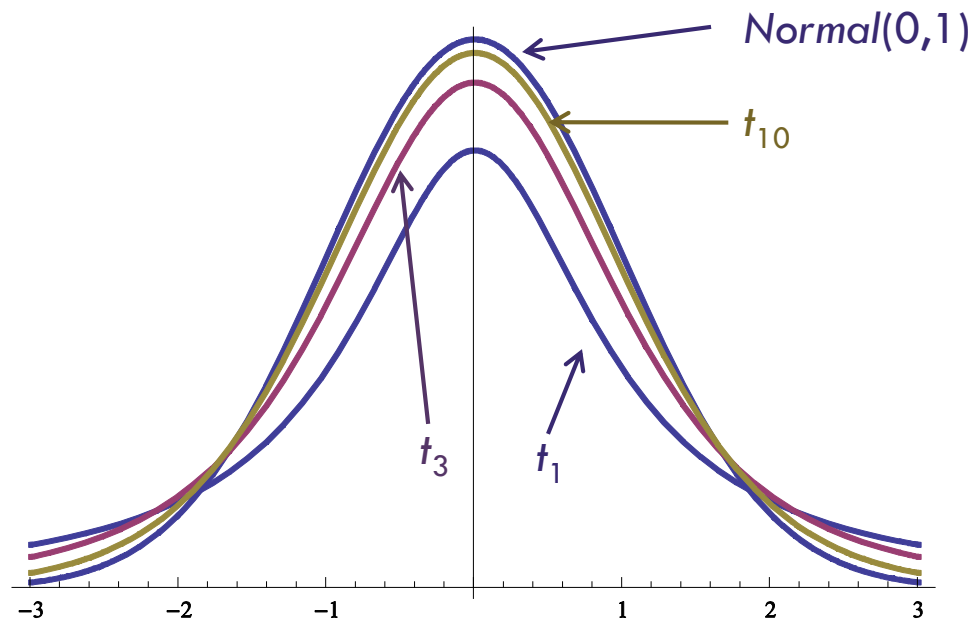
12

- in the previous lecture, we actually developed all the theory needed to carry out the about a single *population* parameter,  $\beta_j$
- we talked about the distribution of the *standardized estimator*:  $\frac{\hat{\beta}_j - \beta_j}{\text{se}(\hat{\beta}_j)}$
- the only thing in the formula that we actually do not know (or, cannot compute from our data) is the true population value of  $\beta_j$ 
  - $\beta_j$  will be supplied by the null hypothesis; i.e., we *hypothesize* about the true population parameter
  - then, we carry out a statistical test to see if we have enough evidence in our data in order to reject this hypothesis
- we had two different results concerning the sampling distributions of standardized estimators:
  1. under MLR.1 through MLR.5, as the sample size increases, the distributions of standardized estimators converge towards the standard normal distribution  $Normal(0,1)$
  2. under MLR.1 through MLR.6, the sampling distribution of standardized estimators is  $t_{n-k-1}$

# Hypotheses about a Single Parameter: $t$ -Tests (cont'd)

13

- this basically means: if we have many observations (high  $n$ ), we don't need the normality assumption (MLR.6), and we can use the Normal distribution instead of Student's  $t$
- actually, we can use Student's  $t$  anyway, because for high  $n$ ,  $t_{n-k-1}$  is *very* close to  $Normal(0,1)$
- hence the name  $t$ -test





## Testing whether the partial effect of $x_j$ on $y$ is significant

- this is the typical test about a population parameter, and the one that *Gretl* shows automatically in every regression output
- as with the effect of a diet, the null hypothesis is the thing we want to *disprove*
- here, we want to disprove there's *no* partial effect of  $x_j$  on  $y$ ; i.e.,

$$H_0: \beta_j = 0.$$

- note that this is the *partial* effect:
  - the null doesn't claim that " $x_j$  has no effect on  $y$ " at all
  - the correct way to put it is: "after  $x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_k$  have been accounted for,  $x_j$  has no effect on the expected value of  $y$ "
- under the null hypothesis, the standardized estimator has the form

$$\frac{\hat{\beta}_j}{\text{se}(\hat{\beta}_j)} = \frac{\text{coefficient}}{\text{standard error}}$$

- this is called the ***t*-ratio**

# Hypotheses about a Single Parameter: $t$ -Tests (cont'd)

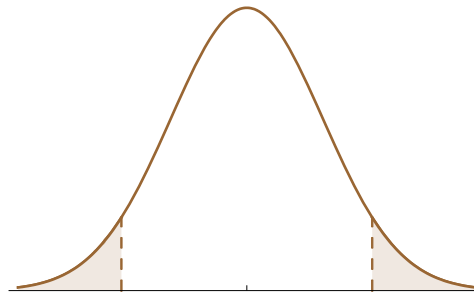
15

- under  $H_0$ ,  $t$ -ratio has the sampling distribution  $t_{n-k-1}$  (approximately or precisely, depending on whether MLR.6 holds)
- *Gretl* automatically carries out the **two-tailed** test:

$$H_0: \beta_j = 0,$$

$$H_1: \beta_j \neq 0.$$

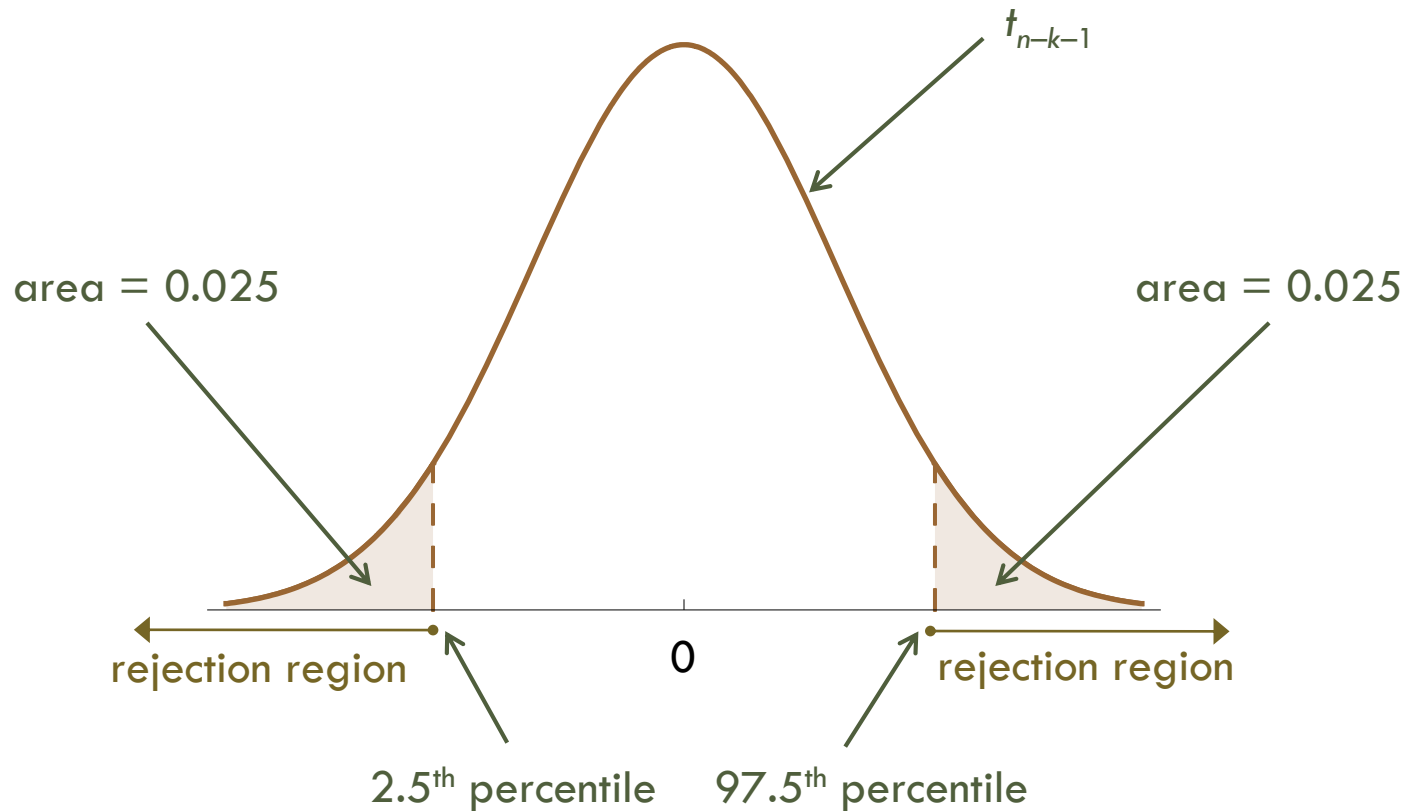
- the *rejection region* is on both tails of the distribution  $t_{n-k-1}$



- if the significance level is 5%, the area below each of the tails is 0.025
- therefore, the bounds of the tails are represented by the 2.5<sup>th</sup> and 97.5<sup>th</sup> percentiles, respectively

# Hypotheses about a Single Parameter: $t$ -Tests (cont'd)

16



Model 1: OLS, using observations 1-328

Dependent variable: l\_price

	coefficient	std. error	t-ratio	p-value	
const	12.5593	0.0427981	293.5	0.0000	***
km1000	-0.00147932	0.000263925	-5.605	4.50e-08	***
age	-0.110341	0.00695126	-15.87	2.95e-042	***
combi	0.0898678	0.0234522	3.832	0.0002	***
diesel	0.164520	0.0240657	6.836	4.12e-011	***
autogas	0.0521381	0.0610311	0.8543	0.3936	
octavia	0.563504	0.0249809	22.56	4.20e-068	***
superb	1.07470	0.0510089	21.07	2.01e-062	***
Mean dependent var	12.18048	S.D. dependent var	0.650827		
Sum squared resid	9.090362	S.E. of regression	0.168545		
R-squared	0.934370	Adjusted R-squared	0.932934		
F(7, 320)	650.8311	P-value(F)	4.6e-185		
Log-likelihood	122.6592	Akaike criterion	-229.3184		
Schwarz criterion	-198.9742	Hannan-Quinn	-217.2119		

97.5<sup>th</sup> percentile of  $t_{320}$  is 1.9674. Which of the explanatory variables **is significant** in the estimated equation? (Note the highlighted jargon.)

# Using p-Values for Hypothesis Testing

18

- as you may have noticed, Gretl reports something called **p-value** next to the *t*-ratio
- what is the *p*-value?
  - *p*-value is the probability of observing a test statistic (in our case, the *t*-ratio) as extreme as we did, if the null hypothesis is true
  - *p*-value is the *smallest* significance level at which the null hypothesis would be rejected for our value of the test statistic
  - **example:** if our *t*-ratio is 1.63, then

$$p\text{-value} = \Pr(|T| > 1.63), \quad \text{where } T \sim t_{n-k-1}$$

→ if *p*-value is *less than* our level of significance, we *reject*  $H_0$

→ low *p*-values represent strong evidence for rejecting  $H_0$

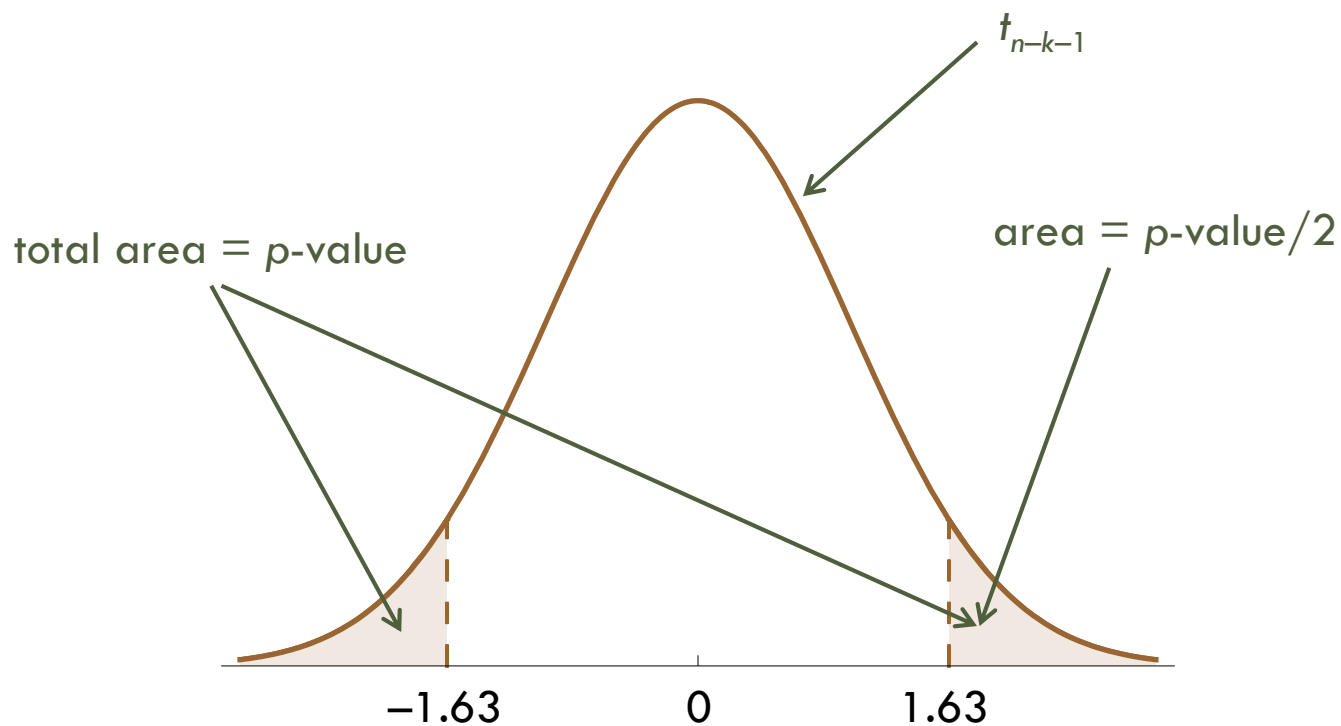
- low levels of *p*-values are highlighted with asterisks in *Gretl*
  - \* ..... *p*-value < **.10**    (→ we can reject  $H_0$  at **10%**)
  - \*\* ..... *p*-value < **.05**    (→ we can reject  $H_0$  at **5%**)
  - \*\*\* ... *p*-value < **.01**    (→ we can reject  $H_0$  at **1%**)



# Using p-Values for Hypothesis Testing

(cont'd)

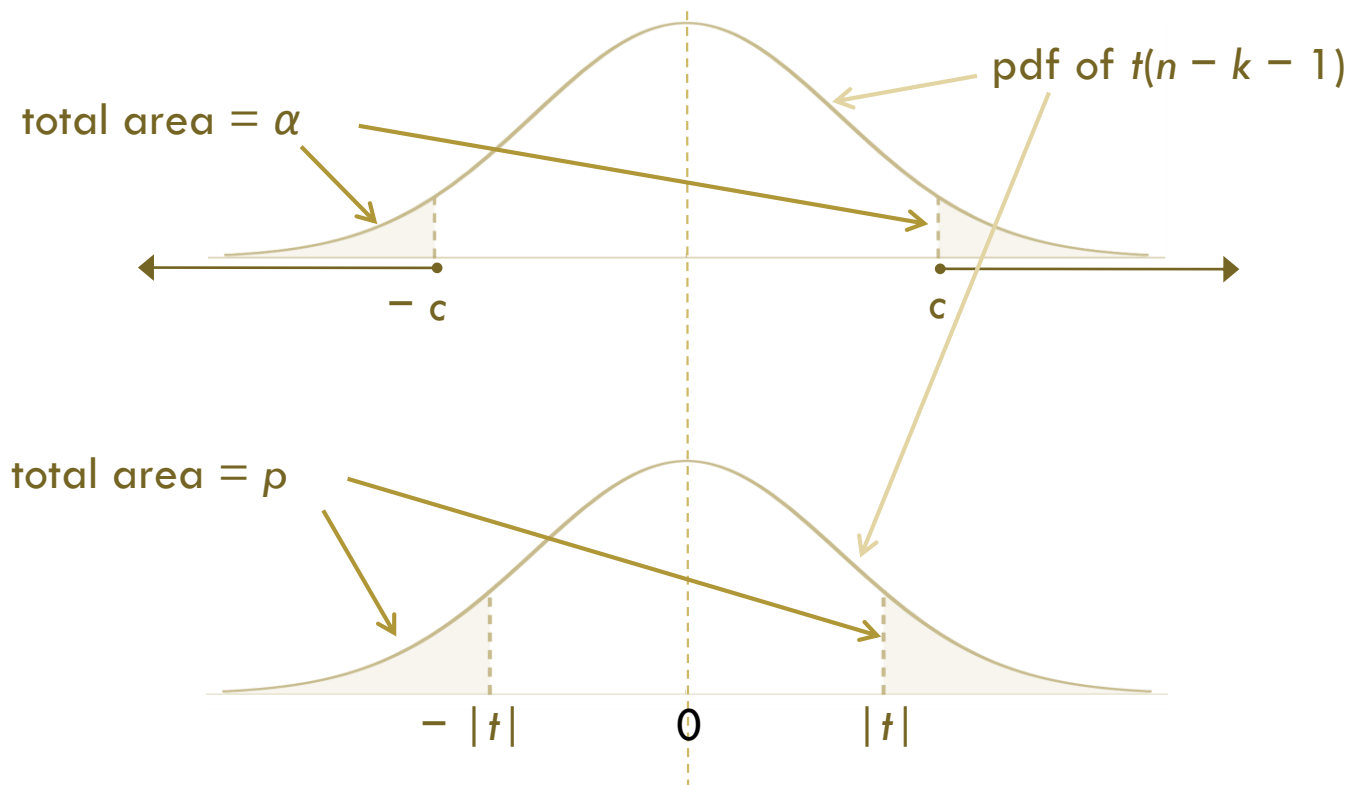
19



# Using p-Values for Hypothesis Testing

(cont'd)

20



Model 1: OLS, using observations 1-328

Dependent variable: l\_price

	coefficient	std. error	t-ratio	p-value	
-----	-----	-----	-----	-----	-----
const	12.5593	0.0427981	293.5	0.0000	***
km1000	-0.00147932	0.000263925	-5.605	4.50e-08	***
age	-0.110341	0.00695126	-15.87	2.95e-042	***
combi	0.0898678	0.0234522	3.832	0.0002	***
diesel	0.164520	0.0240657	6.836	4.12e-011	***
autogas	0.0521381	0.0610311	0.8543	0.3936	
octavia	0.563504	0.0249809	22.56	4.20e-068	***
superb	1.07470	0.0510089	21.07	2.01e-062	***
Mean dependent var	12.18048	S.D. dependent var	0.650827		
Sum squared resid	9.090362	S.E. of regression	0.168545		
R-squared	0.934370	Adjusted R-squared	0.932934		
F(7, 320)	650.8311	P-value(F)	4.6e-185		
Log-likelihood	122.6592	Akaike criterion	-229.3184		
Schwarz criterion	-198.9742	Hannan-Quinn	-217.2119		

Significance stars in Gretl:

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.10$

(in other SW usually different:

\*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ )

# One-Tailed $t$ -Tests

22

- one-tailed  $t$ -tests can be constructed analogously to their two-tailed counterparts
- however, it's important to remember that for the “usual”  $t$ -test with the null hypothesis

$$H_0: \beta_j = 0,$$

all regression packages (including *Gretl*) always calculate the  $p$ -values for the *two-tailed* version of the test

- the  $p$ -value of the one-tailed test can be easily calculated from the two-tailed version:

$$p\text{-value one-tailed} = p\text{-value two-tailed} / 2$$

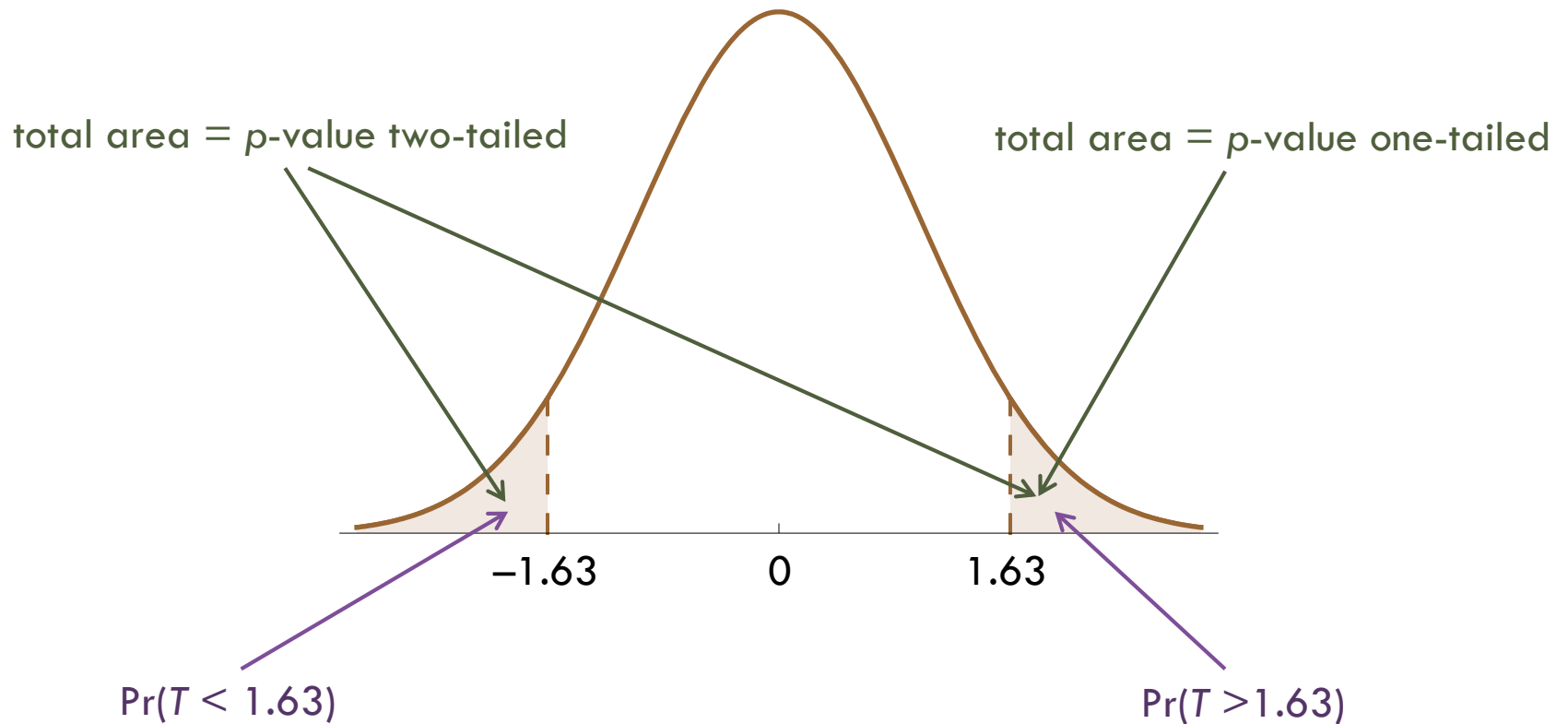
- **example:**
  - let's again consider the test of the significance of  $x_j$  (i.e., the above  $H_0$ ) with a  $t$ -ratio of 1.63
  - this time, the alternative will be:

$$H_1: \beta_j > 0$$

# One-Tailed $t$ -Tests

(cont'd)

23





# *t*-Tests: The General Case

- obviously, we do not have to state the null in the form

$$H_0: \beta_j = 0$$

- the theory we developed enables us to hypothesize about any value of the population parameter, e.g.:

$$H_0: \beta_j = 1$$

- with this  $H_0$ , we can't say anything about the sampling distribution of

$$\frac{\hat{\beta}_j}{\text{se}(\hat{\beta}_j)} = \frac{\text{coefficient}}{\text{standard error}}$$

under  $H_0$

- instead, we'll use the more general version of the *t*-statistic

$$\frac{\hat{\beta}_j - \beta_j}{\text{se}(\hat{\beta}_j)} = \frac{\text{coefficient} - \text{hypothesized value}}{\text{standard error}},$$

which again has the  $t_{n-k-1}$  distribution under  $H_0$

# Confidence Intervals

25

- a **95% confidence interval** (or **interval estimate**) for  $\beta_j$  is the interval given by

$$\hat{\beta}_j \pm c \cdot \text{se}(\hat{\beta}_j)$$

where  $c$  is the 97.5<sup>th</sup> percentile of  $t_{n-k-1}$

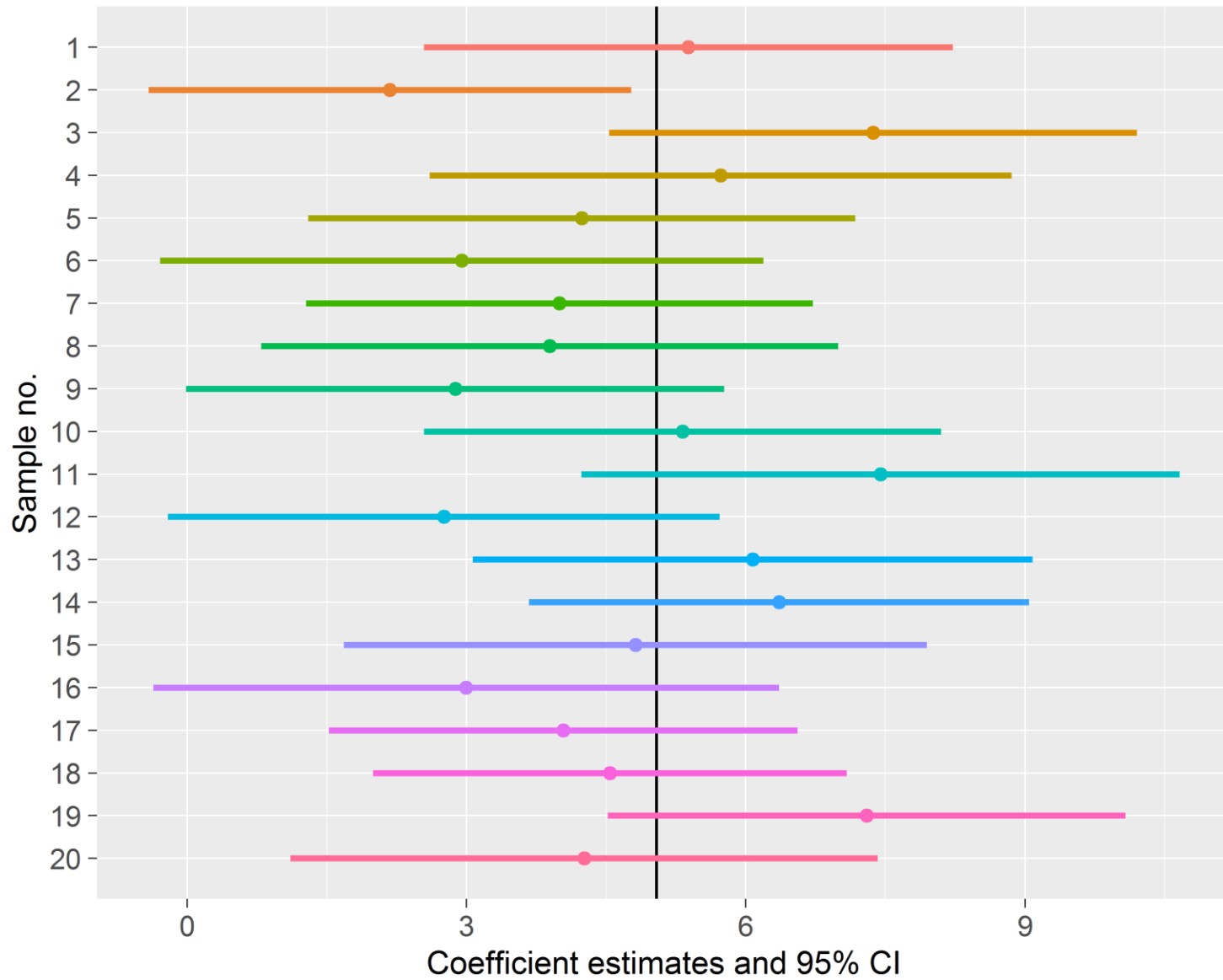
- *interpretation*: roughly speaking, it's an interval that covers the true population parameter  $\beta_j$  in 19 out of 20 samples (i.e., 95% of all samples)
- *basic idea*:
  - the standardized estimator has (either precisely or asymptotically) the  $t_{n-k-1}$  distribution

- therefore,

$$\Pr \left\{ -c \leq \frac{\hat{\beta}_j - \beta_j}{\text{se}(\hat{\beta}_j)} \leq c \right\} = 0.95,$$

||

$$\Pr \left\{ \hat{\beta}_j - c \cdot \text{se}(\hat{\beta}_j) \leq \beta_j \leq \hat{\beta}_j + c \cdot \text{se}(\hat{\beta}_j) \right\}$$



- for quick reference, it's good to know the values of  $c$  (i.e., the 97.5<sup>th</sup> percentiles of  $t$ ) for different degrees of freedom:

$df$	2	5	10	20	50	100	1000
$c$	4.30	2.57	2.23	2.09	2.01	1.98	1.96

- a simple rule of thumb for a 95% confidence interval:  
estimated coefficient  $\pm$  two of its standard errors
- similarly, for a 99% confidence interval:  
estimated coefficient  $\pm$  three of its standard errors
- the 99.5<sup>th</sup> percentiles of  $t$ :

$df$	2	5	10	20	50	100	1000
$c$	9.92	4.03	3.17	2.84	2.67	2.62	2.58

# Using Confidence Intervals for Hypothesis Testing

28

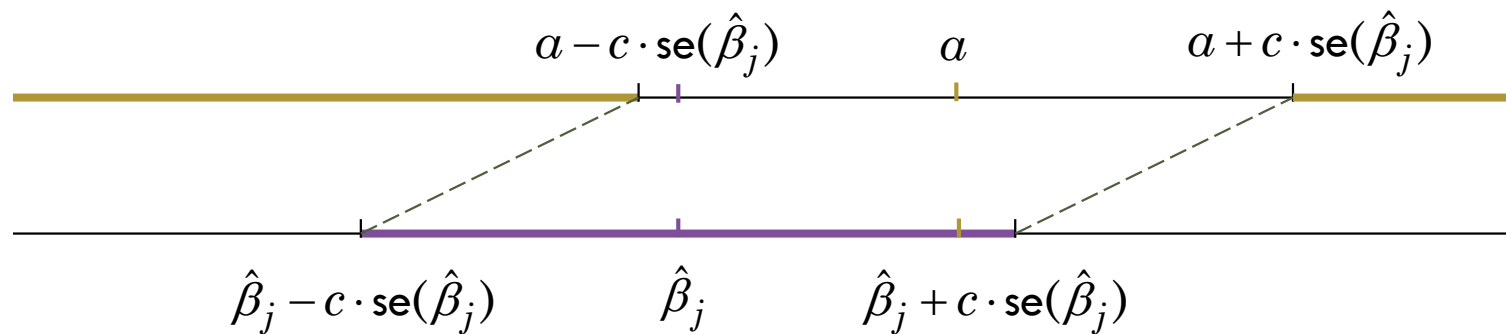
- confidence intervals can be used to easily carry out the two-tailed test

$$H_0: \beta_j = a$$

$$H_1: \beta_j \neq a$$

for any value  $a$ .

- the rule is as follows:
  - $H_0$  is rejected at the 5% significance level if, and only if,  $a$  is *not* in the 95% confidence interval for  $\beta_j$



# Using Confidence Intervals for Hypothesis Testing (cont'd)

29

$$\beta_j \in 95\% \text{ CI} \Leftrightarrow \hat{\beta}_j - c \cdot \text{se}(\hat{\beta}_j) \leq \beta_j \leq \hat{\beta}_j + c \cdot \text{se}(\hat{\beta}_j)$$

$$\Leftrightarrow -\frac{\hat{\beta}_j - \beta_j}{\text{se}(\hat{\beta}_j)} \leq c \leq \frac{\hat{\beta}_j - \beta_j}{\text{se}(\hat{\beta}_j)}$$

$$\Leftrightarrow \left| \frac{\hat{\beta}_j - \beta_j}{\text{se}(\hat{\beta}_j)} \right| \geq c$$

$\Leftrightarrow$   $t$ -statistic is in the rejection region

t(320, 0.025) = 1.967

VARIABLE	COEFFICIENT	95% CONFIDENCE INTERVAL	
const	12.5593	12.4751	12.6435
km1000	-0.00147932	-0.00199856	-0.000960068
age	-0.110341	-0.124017	-0.0966651
combi	0.0898678	0.0437279	0.136008
diesel	0.164520	0.117173	0.211867
autogas	0.0521381	-0.0679347	0.172211
octavia	0.563504	0.514356	0.612652
superb	1.07470	0.974346	1.17506

# Sum-up: Three ways to conclude about the $t$ -test

31

- **Rejection region**
  - No need to know the test statistic in order to determine the rejection region
  - Critical value around two at the usual 5% level
- **Confidence interval**
  - Interesting in its own right
  - No need to specify the hypothesized value first
  - Problem with one-tailed tests
- **p-value**
  - No need to specify the significance level in advance, or: results immediately seen for varying significance levels



# Testing Multiple Linear Restrictions: $F$ -tests

32

□ what do I mean by “linear restrictions”?

■ one linear restriction:

$$H_0: \beta_1 = \beta_2$$

- testing whether two variables have the same effect on the dependent one
- Wooldridge, p. 136: are the returns to education at junior colleges and four-year colleges identical?

■ two linear restrictions:

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

- these are called *exclusion restrictions* (by setting  $\beta_3 = \beta_4 = 0$ , we effectively exclude  $x_3$  and  $x_4$  from the equation)
- we’re testing a *joint significance* of  $x_3$  and  $x_4$
- if  $H_0$  is true,  $x_3$  and  $x_4$  have no effect on  $y$ , once the effect of the remaining explanatory variables have been accounted for

# Testing Multiple Linear Restrictions: $F$ -tests (cont'd)

33

- we won't cover all the theory behind the  $F$ -test here
- *basic idea*:
  - we estimate two models: one with the restrictions imposed in  $H_0$  (the **restricted model**,  $R$ ) and the one without these restrictions (the **unrestricted model**,  $U$ )
  - by similar argumentation as the one we used for the discussion of  $R^2$  in nested models, we can observe that  $SSR$  can only increase as we impose new restrictions, i.e.

$$SSR_R \geq SSR_U$$

- if the increase in  $SSR$  is huge, we have a strong evidence that the restrictions are not fulfilled in our data (the unrestricted model fits our data *much* better), and we reject  $H_0$
- the  $F$ -statistic used for hypothesis testing is:

$$\frac{(SSR_R - SSR_U)/q}{SSR_U/n - k - 1}$$

the number of  
linear restrictions

# Testing Multiple Linear Restrictions: $F$ -tests (cont'd)

34

- ▣ it's obvious that large  $F$ -statistics make us reject  $H_0$
- ▣ but, which values are “large enough”?
- ▣ it can be shown that under  $H_0$ , under the MLR.1 through MLR.5, the  $F$ -statistic has (asymptotically or precisely) the  $F$ -distribution with  $q$  and  $n - k - 1$  degrees of freedom, i.e.

$$\frac{(SSR_R - SSR_U)/q}{SSR_U/n - k - 1} \sim F_{q, n-k-1}$$

- ▣ therefore, we reject  $H_0$  if the  $F$ -statistic is greater than the 95<sup>th</sup> percentile of the  $F$  distribution (this can be found in statistical tables)
- ▣ fortunately, *Gretl* calculates all these things for us, and reports the  $p$ -value for the  $F$ -test
- ▣ so, it's enough to remember what the null hypothesis tells us and how we can use  $p$ -values to evaluate a hypothesis test ☺

Model 1: OLS, using observations 1-328

Dependent variable: l\_price

	coefficient	std. error	t-ratio	p-value	
-----					
const	12.5593	0.0427981	293.5	0.0000	***
km1000	-0.00147932	0.000263925	-5.605	4.50e-08	***
age	-0.110341	0.00695126	-15.87	2.95e-042	***
combi	0.0898678	0.0234522	3.832	0.0002	***
diesel	0.164520	0.0240657	6.836	4.12e-011	***
autogas	0.0521381	0.0610311	0.8543	0.3936	
octavia	0.563504	0.0249809	22.56	4.20e-068	***
superb	1.07470	0.0510089	21.07	2.01e-062	***
Mean dependent var	12.18048	S.D. dependent var	0.650827		
Sum squared resid	9.090362	S.E. of regression	0.168545		
R-squared	0.934370	Adjusted R-squared	0.932934		
F(7, 320)	650.8311	P-value(F)	4.6e-185		
Log-likelihood	122.6592	Akaike criterion	-229.3184		
Schwarz criterion	-198.9742	Hannan-Quinn	-217.2119		

Does fuel type matter? In other words, are *diesel* and *autogas* **jointly significant**?

Null hypothesis: the regression parameters are zero for the variables diesel, autogas

Test statistic:  $F(2, 320) = 23.4104$ , p-value  $3.24608e-010$

Omitting variables improved 0 of 3 information criteria.

Model 2: OLS, using observations 1-328

Dependent variable: l\_price

	coefficient	std. error	t-ratio	p-value	
const	12.6247	0.0444711	283.9	0.0000	***
km1000	-0.00122462	0.000278073	-4.404	1.45e-05	***
age	-0.121677	0.00720144	-16.90	2.67e-046	***
combi	0.115870	0.0246988	4.691	4.02e-06	***
octavia	0.584548	0.0264572	22.09	1.75e-066	***
superb	1.11115	0.0541394	20.52	1.92e-060	***

Mean dependent var	12.18048	S.D. dependent var	0.650827
Sum squared resid	10.42042	S.E. of regression	0.179893
R-squared	0.924767	Adjusted R-squared	0.923599
F(5, 322)	791.6110	P-value(F)	1.7e-178
Log-likelihood	100.2646	Akaike criterion	-188.5291
Schwarz criterion	-165.7710	Hannan-Quinn	-179.4493

Log-likelihood for price = -3894.93

Model 1: OLS, using observations 1-328  
Dependent variable: l\_price

	coefficient	std. error	t-ratio	p-value	
const	12.5593	0.0427981	293.5	0.0000	***
km1000	-0.00147932	0.000263925	-5.605	4.50e-08	***
age	-0.110341	0.00695126	-15.87	2.95e-042	***
combi	0.0898678	0.0234522	3.832	0.0002	***
diesel	0.164520	0.0240657	6.836	4.12e-011	***
autogas	0.0521381	0.0610311	0.8543	0.3936	
octavia	0.563504	0.0249809	22.56	4.20e-068	***
superb	1.07470	0.0510089	21.07	2.01e-062	***
Mean dependent var	12.18048	S.D. dependent var	0.650827		
Sum squared resid	9.090362	S.E. of regression	0.168545		
R-squared	0.934370	Adjusted R-squared	0.932934		
F(7, 320)	650.8311	P-value(F)	4.6e-185		
Log-likelihood	122.6592	Akaike criterion	-229.3184		
Schwarz criterion	-198.9742	Hannan-Quinn	-217.2119		

**Overall F-test:** null hypothesis is (simple way to write all  $k$  restrictions):

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$$

# Relationship Between $F$ and $t$ Tests

38

- Consider three null hypotheses:
  - a)  $H_0: \beta_1 = 0, \beta_2 = 0$  ( $\rightarrow F$ -test of joint significance of  $x_1$  and  $x_2$ )
  - b)  $H_0: \beta_1 = 0$  ( $\rightarrow t$ -test of individual significance of  $x_1$ )
  - c)  $H_0: \beta_2 = 0$  ( $\rightarrow t$ -test of individual significance of  $x_2$ )
  
- Which of the following can happen?
  - 1) Reject all.
  - 2) Reject none.
  - 3) Reject (a) & (b), do not reject (c).
  - 4) Reject (a), do not reject (b) & (c).
  - 5) Reject (b), do not reject (a) & (c).
  - 6) Reject (b) & (c), do not reject (a).

	(1)	(2)	(3)
const	12.6** (0.0428)	12.6** (0.0753)	12.7** (0.0441)
km1000	-0.00148** (0.000264)	-0.00148** (0.000264)	-0.00148** (0.000264)
age	-0.110** (0.00695)	-0.110** (0.00695)	-0.110** (0.00695)
combi	0.0899** (0.0235)	0.0899** (0.0235)	0.0899** (0.0235)
octavia	0.564** (0.0250)	0.564** (0.0250)	0.564** (0.0250)
superb	1.07** (0.0510)	1.07** (0.0510)	1.07** (0.0510)
diesel	0.165** (0.0241)	0.112* (0.0637)	
autogas	0.0521 (0.0610)		-0.112* (0.0637)
petrol		-0.0521 (0.0610)	-0.165** (0.0241)



Intercepts differ.  
Why?



No change here.



Stars light up and  
fade away...



Test on Model 1:

Null hypothesis: the regression parameters are zero for the variables  
diesel, autogas

Test statistic:  $F(2, 320) = 23.4104$ , p-value  $3.24608e-010$

Omitting variables improved 0 of 3 information criteria.

Test on Model 2:

Null hypothesis: the regression parameters are zero for the variables  
diesel, petrol

Test statistic:  $F(2, 320) = 23.4104$ , p-value  $3.24608e-010$

Omitting variables improved 0 of 3 information criteria.

Test on Model 3:

Null hypothesis: the regression parameters are zero for the variables  
petrol, autogas

Test statistic:  $F(2, 320) = 23.4104$ , p-value  $3.24608e-010$

Omitting variables improved 0 of 3 information criteria.

# Relationship Between $F$ and $t$ Tests

41

- you might have noticed that we can also use the  $F$ -test for

$$H_0: \beta_j = 0,$$

$$H_1: \beta_j \neq 0.$$

- this is what we used  $t$ -tests for
- then you might ask: which one is better?
- the answer is that it doesn't really matter: the results will be the same
- in fact the  $t$ -statistic squared has an  $F$ -distribution with 1 degree of freedom in the numerator
  - ▣ formally,  $T \sim t_{n-k-1} \Rightarrow T^2 \sim F_{1,n-k-1}$
  - ▣ note that, for large degrees of freedom, the critical values ( $c$ ) are
    - 1,96 for the  $t$ -test (using  $t_{df}$ )
    - 3.84 for the  $F$  test (using  $F_{1,df}$ )
    - and  $1.96^2 = 3.84$

# Relationship Between $F$ and $t$ Tests

42

	coefficient	std. error	t-ratio	p-value	
const	12.5593	0.0427981	293.5	0.0000	***
km1000	-0.00147932	0.000263925	-5.605	4.50e-08	***
age	-0.110341	0.00695126	-15.87	2.95e-042	***
combi	0.0898678	0.0234522	3.832	0.0002	***
diesel	0.164520	0.0240657	6.836	4.12e-011	***
autogas	0.0521381	0.0610311	0.8543	0.3936	
octavia	0.563504	0.0249809	22.56	4.20e-068	***
superb	1.07470	0.0510089	21.07	2.01e-062	***

Test on Model 1:

Null hypothesis: the regression parameter is zero for autogas

Test statistic:  $F(1, 320) = 0.729807$ , p-value 0.393585

*Note:*  $0.8543^2 = 0.7298$

LECTURE 5:  
HYPOTHESIS TESTING

Jan Zouhar

Introductory Econometrics