

Multiple Regression Analysis (Inference)

Objectives

1. Students know how to test hypotheses about the parameter (β)
2. Students can test for the validity of the proposed population model.

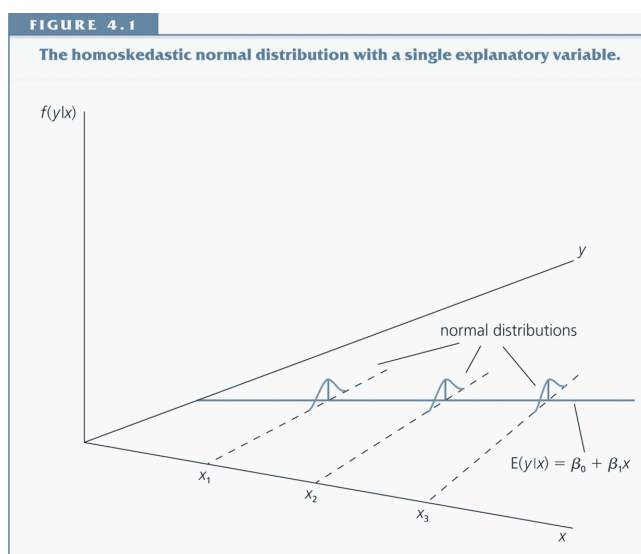
Econometrics Analysis - The Steps

1 Sampling Distribution of the OLS estimators ($\hat{\beta}_{OLS}$)

To be able to test hypotheses about the parameter ($\hat{\beta}$), we need an assumption about the distribution of u .

Assumption MLR 6 - Normality

The population error u is independent of the explanatory variables $X_1, X_2, \dots, X_k$ and is **normally distributed** with zero mean and variable σ^2 .



- By normality of the error term (u), we have normality of $\hat{\beta}$

Or, using the standardization $\frac{Z-\mu}{\sigma} \sim \text{normal}(0,1)$, we have

2 Testing Hypotheses about an individual regression coefficient "the t-test"

- But we do not have $sd.(\hat{\beta}_j)$, we can only calculate the estimator of it :

- for degree of freedom > 30 , $t \approx z$. We can use the $Z - score$ table to find the critical value(s).

2.1 Testing Against Two-Sided Alternatives

Consider the wage equation :

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + u.$$

Suppose we want to test whether experience has a partial effect on wage:

$$\begin{aligned} H_0 &: \beta_2 = 0 \quad (\text{experience has no partial effect}) \\ H_a &: \beta_2 \neq 0. \quad (\text{experience has a partial effect}) \end{aligned}$$

** Note that

Suppose $n = 34$, degree of freedom $= 34 - 3 - 1 = 30$.

From the table t-table, the 5% critical value for a 2-tailed test with 30 d.f. is _____

- If we change the significance level to 10%, then the 2-tailed critical value becomes _____
- If we cannot reject $H_0 : \beta_2 = 0$ (at a given significance level – e.g. 5%), we say that X_2 is statistically significant at the 5% level.
- In other words X_2 has a partial effect on the expected value of Y .

```
. regress log_wage educ exper tenure
```

Source	SS	df	MS	Number of obs = 526		
Model	46.8741776	3	15.6247259	F(3, 522) = 80.39		
Residual	101.455574	522	.194359337	Prob > F = 0.0000		
Total	148.329751	525	.28253286	R-squared = 0.3160		
				Adj R-squared = 0.3121		
				Root MSE = .44086		

log_wage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	.092029	.0073299	12.56	0.000	.0776292	.1064288
exper	.0041211	.0017233	2.39	0.017	.0007357	.0075065
tenure	.0220672	.0030936	7.13	0.000	.0159897	.0281448
_cons	.2843595	.1041904	2.73	0.007	.0796756	.4890435

2.2 Testing Against One-Sided Alternatives

Suppose the rule for rejecting H_0 becomes

$$H_a : \beta_j > 0.$$

We could have the H_0 as

$$\begin{aligned} H_0 & : \beta_j \leq 0 & \text{or} \\ H_0 & : \beta_j = 0 \end{aligned}$$

depending on the context. For $H_0 : \beta_j = 0$, it implies that we are ruling out population values of β_j less than zero. For $H_0 : \beta_j \leq 0$, it implies that we are not ruling out population values of β_j less than zero. To give an example of a test against the one-sided alternative hypothesis, consider the wage equation again :

$$\log(\text{wage}) = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + \beta_3 \text{tenure} + u$$

If we want to test whether experience has a positive partial effect on wage

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

We implicitly rule out the possibility that $\beta_2 < 0$ here.

3 Testing other hypotheses about β_j

- Most common H_0 is $H_0 : \beta_j = 0$.
- However, we can test other types of hypotheses. For example, $H_0 : \beta_j = a_j$ where a_j is a constant number.

If we want to test the H_0 using a 5% significance level, then we reject H_0 if

$$\begin{array}{lcl}
 t_{wife_income} & > & \text{-----} \quad \text{or} \\
 t_{wife_income} & < & \text{-----}
 \end{array}$$

4 Testing Hypotheses about a Single Linear Combination of the Parameter

Consider

$$\log(wage) = \beta_0 + \beta_1 jc + \beta_2 univ + \beta_3 exp\ er + u$$

where jc = number of years attending a two-year college

$univ$ = number of years at a four-year college

$exp\ er$ = months in the workforce.

We want to test whether $\beta_1 = \beta_2$.

another possible hypothesis test (one-tailed alternative)

5 Computing p-Values for t-Tests

- What is the significance level given the computed t-statistics?

- p-value : $P(|T| > |t|)$

Example 1: $H_0 : \beta_j \geq 0$, $H_a : \beta_j < 0$, d.f.= 140.

suppose the calculated $t_{\hat{\beta}_j} = -2.75$

- From the z-table, the value -2.75 corresponds to area = _____.
- Thus, p-value = _____.
- Would we reject H_0 if we use the significance level = 5%?

Example 2: $H_0 : \beta_j = a_j$, $H_a : \beta_j \neq a_j$, d.f.= 18.

suppose the calculated $t_{\hat{\beta}_j} = -2.18$

- From the t-table, the value -2.18 corresponds to area = _____.
- Thus, p-value = _____.
- Would we reject H_0 if we use the significance level = 5%?

6 Confidence Intervals (CI)

- Confidence Intervals for the POPULATION PARAMETER (β_j)
- A 95% CI of β_j is given by

Example 1: **95% CI**

Example 2: **99% CI**

7 Testing Multiple Linear Restrictions: The F-test

Suppose the model is specified by

$$\begin{aligned} y &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u \\ H_0 &: \beta_1 = 0 \text{ and } \beta_2 = 0 \\ H_1 &: H_0 \text{ is not true} \end{aligned}$$

We can use the F-test to test this type of "multiple hypotheses".

1. Our full model is called the "unrestricted" model (ur). Suppose it can be expressed as:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_k x_k + u$$

2. The model which takes out x (which we think its associated $\beta = 0$) is called the restricted model (r).

3. Some useful facts

4. Other ways to calculate the F-statistics:

Example: Suppose we are interested in understanding the determinant of a baseball player's salary.

<i>salary</i>	= season salary
<i>years</i>	= years in major leagues
<i>gamesyr</i>	= games per year in the league
<i>bavg</i>	= career batting average
<i>hrunsyr</i>	= homeruns per year
<i>rbisyr</i>	= runs batted in per year

- the unrestricted model (ur) is defined by

```
. regress log_salary years gamesyr bavg hrunsyr rbisyr
```

Source	SS	df	MS	Number of obs = 353		
Model	308.989208	5	61.7978416	F(5, 347) = 117.06		
Residual	183.186327	347	.527914487	Prob > F = 0.0000		
				R-squared = 0.6278		
				Adj R-squared = 0.6224		
Total	492.175535	352	1.39822595	Root MSE = .72658		

log_salary	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
years	.0688626	.0121145	5.68	0.000	.0450355	.0926898
gamesyr	.0125521	.0026468	4.74	0.000	.0073464	.0177578
bavg	.0009786	.0011035	0.89	0.376	-.0011918	.003149
hrunsyr	.0144295	.016057	0.90	0.369	-.0171518	.0460107
rbisyr	.0107657	.007175	1.50	0.134	-.0033462	.0248776
_cons	11.19242	.2888229	38.75	0.000	10.62435	11.76048

- the restricted model (r) is defined by

```
. regress log_salary years gamesyr
```

Source	SS	df	MS	Number of obs = 353		
Model	293.864058	2	146.932029	F(2, 350) = 259.32		
Residual	198.311477	350	.566604221	Prob > F = 0.0000		
				R-squared = 0.5971		
				Adj R-squared = 0.5948		
Total	492.175535	352	1.39822595	Root MSE = .75273		

log_salary	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
years	.071318	.012505	5.70	0.000	.0467236	.0959124
gamesyr	.0201745	.0013429	15.02	0.000	.0175334	.0228156
_cons	11.2238	.108312	103.62	0.000	11.01078	11.43683

Now, our H_0 and H_a becomes

8 How the Hypothesis Testing is done in Practice

1. Check the values of t – *statistic* reported by the statistical software (i.e. STATA, SPSS, SAS)

⇒ These t – *statistics* are to test $H_0 : \beta_i = 0$

⇒ If the d.f. > 30 , then when $t > 1.96$, we can reject H_0

⇒ **When $t > 1.96$** , we can say that β_i is **statistically significant** at 5% level.
(value of $\beta_i \neq 0$)

⇒ **When $t < 1.96$** we can say that β_i is **not statistically significant** at 5% level.

⇒ If $t < 1.96$ we can drop x_i from the model

⇒ After we drop x_i , we estimate the new regression function and obtain a new set of $\hat{\beta}$.

2. We can also perform other hypothesis testings of interest.

e.g. $H_0 : \beta_i = \beta_j$

or $H_0 : \beta_i = 5$ etc.

or perform an F-test for testing multiple linear restrictions

3. Usually, in economics, the estimation results are reported using this form

Dependent Variable: $\log(\text{salary})$			
Independent Variables	(1)	(2)	(3)
$\log(\text{sales})$	.224 (.027)	.158 (.040)	.188 (.040)
$\log(\text{mktval})$	—	.112 (.050)	.100 (.049)
profmarg	—	–.0023 (.0022)	–.0022 (.0021)
ceoten	—	—	.0171 (.0055)
comten	—	—	–.0092 (.0033)
intercept	4.94 (0.20)	4.62 (0.25)	4.57 (0.25)
Observations	177	177	177
R -squared	.281	.304	.353