4.Linear Hypothesis Tests

ECON8011 Microeconometrics

Burcu Duzgun Oncel *

*Marmara University Department of Economics

Spring2023

Table of contents

Testing Linear Hypothesis about β

1. $H_0 = \beta_i = 0$. This sets up the hypothesis that the regressor X_i has no influence on Y. This type of test is very common and is often referred to simply as a significance test.
2. $H_0 = \beta_i = \beta_{i0}$. Here β_{i0}. is some specified value. If, for instance, β_i a price elasticity one might wish to test that the elasticity is -1.
3. $H_0 = \beta_2 + \beta_3 = 1$. If the β indicate labor and capital elasticities in a production function, this formulation hypothesizes constant returns to scale.
4. $H_0 : \beta_3 = \beta_4$. This hypothesizes that X_3 and X_4 have the same coefficient.
5. This sets up the hypothesis that the complete set of regressors has no effect on Y.

$$H_0 : \begin{bmatrix} \beta_2 \\ \beta_3 \\ . \\ . \\ \beta_k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ . \\ . \\ 0 \end{bmatrix}$$

- For the foregoing examples we have:

1. $\mathbf{R} = [0....0 \;\; 1 \;\; 0....0] \quad r = 0 \quad q = 1$ with 1 in the *ith* position.
2. $\mathbf{R} = [0....0 \;\; 1 \;\; 0....0] \quad r = \beta_{i0} \quad q = 1$ with 1 in the *ith* position.
3. $\mathbf{R} = [0 \;\; 1 \;\; 1 \;\; 0....0] \quad r = 1 \quad q = 1$
4. $\mathbf{R} = [0 \;\; 0 \;\; 1 \;\; -1....0] \quad r = 0 \quad q = 1$
5. $\mathbf{R} = [\mathbf{0} \;\; \mathbf{I_{k-1}}] \quad r = 0 \quad q = k-1$ where $\mathbf{0}$ is a vector of $k-1$ zeros

- The efficient way to proceed is to derive a testing procedure for the general linear hypothesis:

  $H_0 : \mathbf{R}\beta - \mathbf{r} = \mathbf{0}$

- Given the OLS estimator, an obvious step is to compute the vector $\mathbf{R}\beta - \mathbf{r} = \mathbf{0}$.

- This vector measures the discrepancy between expectation and observation.

- $\mathbf{R}\hat{\beta} = R\beta$

- $\mathbf{Var}(\mathbf{R}\hat{\beta}) = \sigma^{\mathbf{2}}\mathbf{R}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{R}'$

$H_0 : \beta_i = 0$

- This sets up the hypothesis that the regressor X_i has no influence on Y.
- This type of test is very common and is often refereed to simply as a significance test.

$t = \frac{\hat{\beta}_i}{se(\hat{\beta}_i)}$ with $(n - k)$ degrees of freedom.

$H_0 : \beta_i = \beta_{i,0}$

- Here $\beta_{i,0}$ is some specified value.
- For instance β_i denotes price elasticity one might wish to test that the elasticity is -1.

  $t = \frac{\hat{\beta}_i - \beta_{i,0}}{se(\hat{\beta}_i)}$ with $(n-k)$ degrees of freedom.
- Instead of testing specific hypotheses about β_i one may compute a 95 percent confidence interval for β_i by:

  $\hat{\beta}_i + t_{0.025} se(\hat{\beta}_i)$ and $\hat{\beta}_i - t_{0.025} se(\hat{\beta}_i)$

$H_0 : \beta_2 + \beta_3 = 1$

- If the β's indicate labor and capital elasticities in a production function, this formulation hypothesizes constant returns to scale.

$t = \frac{\hat{\beta}_2 + \hat{\beta}_3 - 1}{\sqrt{var(\hat{\beta}_2 + \hat{\beta}_3)}}$ with $n - k$ degrees of freedom.

- 95 percent confidence interval for the sum $(\beta_2 + \beta_3)$:

$(\hat{\beta}_2 + \hat{\beta}_3) + t_{0.025}\sqrt{var(\hat{\beta}_2 + \hat{\beta}_3)}$ and
$(\hat{\beta}_2 + \hat{\beta}_3) - t_{0.025}\sqrt{var(\hat{\beta}_2 + \hat{\beta}_3)}$

$H_0 = \beta_3 = \beta_4$ or $\beta_3 - \beta_4 = 0$

- This hypothesizes that X_3 and X_4 have the same coefficient.

$t = \frac{\hat{\beta}_3 - \hat{\beta}_4}{\sqrt{var(\hat{\beta}_3 - \hat{\beta}_4)}}$ with $n - k$ degrees of freedom

$H_0 : \beta_2 = \beta_3 = ... = \beta_k = 0$

- This sets up the hypothesis that the complete set of regressors has no effect on Y.
- It tests the significance of the overall relation.
- The intercept does not enter into the hypothesis since interest centers on the variation of Y around its mean.

$$F = \frac{ESS/(k-1)}{RSS/(n-k)} \text{ with } F(k-1, n-k)$$

$$F = \frac{R^2/(k-1)}{1-R^2/(n-k)} \text{ with } F(k-1, n-k)$$

- Cross-section data on individuals (from MUS chapter 10)
- Dependent variable docvis is a count. Here do OLS.
- Begin with data description and summary statistics.

```
. use mus10data.dta, clear

. quietly keep if year02==1

. describe docvis private chronic female income

              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
docvis          int    %8.0g                  number of doctor visits
private         byte   %8.0g                  = 1 if private insurance
chronic         byte   %8.0g                  = 1 if a chronic condition
female          byte   %8.0g                  = 1 if female
income          float  %9.0g                  Income in $ / 1000

. summarize docvis private chronic female income

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      docvis |      4412    3.957389    7.947601          0        134
     private |      4412    .7853581    .4106202          0          1
     chronic |      4412    .3263826    .4689423          0          1
      female |      4412    .4718948    .4992661          0          1
      income |      4412    34.34018    29.03987    -49.999    280.777
```

- OLS regression with default standard errors: assumes i.i.d error

```
. * OLS regression with default standard errors
. regress docvis private chronic female income
```

Source	SS	df	MS
Model	35771.7188	4	8942.92971
Residual	242846.27	4407	55.1046676
Total	278617.989	4411	63.1643594

Number of obs =	4412
F(4, 4407) =	162.29
Prob > F =	0.0000
R-squared =	0.1284
Adj R-squared =	0.1276
Root MSE =	7.4233

docvis	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
private	1.916263	.2881911	6.65	0.000	1.351264	2.481263
chronic	4.826799	.2419767	19.95	0.000	4.352404	5.301195
female	1.889675	.2286615	8.26	0.000	1.441384	2.337967
income	.016018	.004071	3.93	0.000	.0080367	.0239993
_cons	-.5647368	.2746696	-2.06	0.040	-1.103227	-.0262465

- Overall fit poor as $R^2 = 0.13$. Often the case for cross-section data.

- Yet all regressors are statistically significant and have large impact.

- OLS regression with robust standard errors for OLS estimator.
- Preferred at this permits model error to be heteroskedastic.

```
. * OLS regression with robust standard errors
. regress docvis private chronic female income, vce(robust)

Linear regression                                      Number of obs =    4412
                                                       F(  4,  4407) =  107.01
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.1284
                                                       Root MSE      =  7.4233
```

docvis	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
private	1.916263	.2347443	8.16	0.000	1.456047	2.37648
chronic	4.826799	.3001866	16.08	0.000	4.238283	5.415316
female	1.889675	.2154463	8.77	0.000	1.467292	2.312058
income	.016018	.005606	2.86	0.004	.0050275	.0270085
_cons	-.5647368	.2069188	-2.73	0.006	-.9704017	-.159072

```
. * Comparison of standard errors
. quietly regress docvis private chronic female income

. estimates store DEFAULT

. quietly regress docvis private chronic female income, vce(robust)

. estimates store ROBUST

. estimates table DEFAULT ROBUST, b(%9.4f) se stats(N r2 F)
```

Variable	DEFAULT	ROBUST
private	1.9163	1.9163
	0.2882	0.2347
chronic	4.8268	4.8268
	0.2420	0.3002
female	1.8897	1.8897
	0.2287	0.2154
income	0.0160	0.0160
	0.0041	0.0056
_cons	-0.5647	-0.5647
	0.2747	0.2069
N	4412.0000	4412.0000
r2	0.1284	0.1284
F	162.2899	107.0104

legend: b/se

- The prefferd heteroskedastic-robust standard errors are within 25% of defalut, sometimes more and sometimes less.