MICHIGAN STATE UNIVERSITY

AFRE 802
Statistical Methods for Agricultural, Food, & Resource Economists

Linear models & estimation by least squares – Part 3 of 3
(WMS Ch. 11.5 & Wooldridge pp. 113-136)

December 5, 2017

Nicole Mason
Michigan State University
Fall 2017

GAME PLAN

- Housekeeping issues:
  - Office hours this week are Wednesday, 11 AM-1 PM
  - Friday optional review session this week will be 4-5 PM
  - I will hold extra office hours next Tuesday (Dec. 12) from 3-5 PM in the Cook Hall basement
- Return take-home graded exercise (see answer key in 2014 final exam on D2L)
- Collect Thursday's additional practice problem
- Distribute new additional practice problem
- Review

- Linear models & estimation by least squares – Part 3 of 3
  - Classical linear model assumptions
  - Inference
    - Hypothesis testing & p-values
    - Confidence intervals

1

Review: Total, explained, & residual SS, R^2

Total sum of squares: $$\boxed{SST \equiv \sum_{i=1}^{N}(y_i - \bar{y})^2}$$

Explained sum of squares: $$\boxed{SSE \equiv \sum_{i=1}^{N}(\hat{y}_i - \bar{y})^2}$$

Residual sum of squares: $$\boxed{SSR \equiv \sum_{i=1}^{N}\hat{u}_i^2}$$

$$\boxed{SST = SSE + SSR}$$

Coefficient of determination or R^2:

$$\boxed{R^2 = SSE / SST = 1 - (SSR / SST)}$$

Interpretation?

The proportion of the sample variation in y that is explained by x

Review: Simple linear regression assumptions & implications

SLR.1-SLR.4 ➔ OLS estimators are **unbiased**

SLR.1. Linear in parameters:

SLR.2. Random sampling

****SLR.3. Zero conditional mean (exogeneity):**

$$\boxed{E(u \mid x) = E(u) = 0}$$

SLR.4. Sample variation in x

SLR.5. Homoskedasticity (constant variance):

$$\boxed{V(u \mid x) = V(u) = \sigma^2}$$

→ Formulas for variances of OLS estimators are:

$$\boxed{V(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2}} \quad \boxed{V(\hat{\beta}_0) = \frac{\sigma^2 N^{-1}\sum_{i=1}^{N}x_i^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2}}$$

→ SLR.1-SLR.5 ➔ OLS is **BLUE** (Gauss-Markov Theorem)

MICHIGAN STATE UNIVERSITY

Unbiased & consistent estimator of $V(u) = \sigma^2$

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

Use in formulas to estimate variances and obtain standard errors of our OLS estimators:

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

$$\hat{V}(\hat{\beta}_0) = \frac{\hat{\sigma}^2 N^{-1}\sum_{i=1}^{N}x_i^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

$$\hat{\sigma}_{\hat{\beta}_j} = \sqrt{\hat{V}(\hat{\beta}_j)} \text{ for } j = 0,1$$

Review

MICHIGAN STATE UNIVERSITY

Annotated Stata output for simple linear regression so far: N, OLS estimates, SSE, SSR, SST, R^2

```
use "/Users/nicolemason/Documents/AEC802/data/WAGE1_Stata13.dta"

reg wage educ
```

	Source	SS	df	MS	
SSE	Model	1179.73204	1	1179.73204	
SSR	Residual	5980.68225	524	11.4135158	$\hat{\sigma}^2$
SST	Total	7160.41429	525	13.6388844	

Number of obs =	526	
F(1, 524) =	103.36	
Prob > F =	0.0000	
R-squared =	0.1648	
Adj R-squared =	0.1632	
Root MSE =	3.3784	$\hat{\sigma}$

	wage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
$\hat{\beta}_1$	educ	.5413593	.053248	10.17	0.000	.4367534	.6459651
$\hat{\beta}_0$	_cons	-.9048516	.6849678	-1.32	0.187	-2.250472	.4407687

$\hat{\sigma}_{\hat{\beta}_j}$

5

What we know about the sampling distributions of the OLS estimators so far

$$y = \beta_0 + \beta_1 x + u$$

OLS estimators for β_0 and β_1:

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

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Expected values (under SLR.1-SLR.4):

$$E(\hat{\beta}_1) = \beta_1 \text{ and } E(\hat{\beta}_0) = \beta_0$$

Sample variances (under SLR.1-SLR.5):

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

$$\hat{V}(\hat{\beta}_0) = \frac{\hat{\sigma}^2 N^{-1} \sum_{i=1}^{N} x_i^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

$$\text{where } \hat{\sigma}^2 = \frac{1}{N-2}\sum_{i=1}^{N}\hat{u}_i^2 = \frac{SSR}{N-2}$$

$\hat{\sigma}$ is the **standard error** of the regression

MICHIGAN STATE UNIVERSITY

The sampling distributions of the OLS estimators

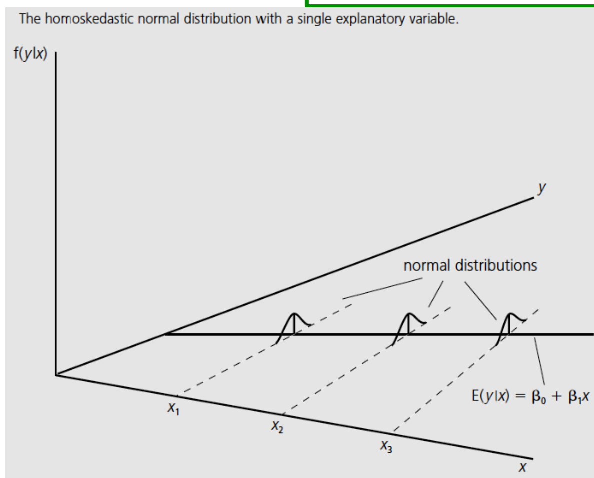
- **By CLT**, Under assumptions SLR.1-SLR.5, the OLS estimators are **asymptotically** (i.e., as $N \to \infty$) **normally distributed** with the means & variances on the previous slides
- If we **add one more assumption**, then we can obtain the sampling distribution of the OLS estimators in **finite samples**

SLR.6. Normality: The population error, ***u*, is independent of *x*** and is **normally distributed** with $E(u)=0$ and $V(u)=\sigma^2$, i.e.: $u \sim Normal(0, \sigma^2)$

7

SLR.1-SLR.6 = "classical linear model assumptions"

- CLM = Gauss-Markov + SLR.6 (normality)
- CLM assumptions imply $\boxed{y \mid x \sim Normal(\beta_0 + \beta_1 x,\ \sigma^2)}$



Source: Wooldridge (2003)

8

MICHIGAN STATE UNIVERSITY

$$y = \beta_0 + \beta_1 x + u$$

The sampling distributions of the OLS estimators under the CLM assumptions (SLR.1-SLR.6):

$\boxed{\hat{\beta}_j \sim Normal\left(\beta_j,\ V(\hat{\beta}_j)\right)}$ where $V(\hat{\beta}_1) = \dfrac{\sigma^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2}$,

$$V(\hat{\beta}_0) = \frac{\sigma^2 N^{-1} \sum_{i=1}^{N} x_i^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2} \qquad \hat{\sigma}^2 = \frac{1}{N-2}\sum_{i=1}^{N}\hat{u}_i^2 = \frac{SSR}{N-2}$$

If we **know σ^2,** then we can standardize beta-hat$_j$ to a **Z-statistic**; otherwise, we can **estimate σ^2** and compute a **T-statistic** – i.e.:

$$\boxed{Z = \frac{\hat{\beta}_j - \beta_j}{\sqrt{V(\hat{\beta}_j)}} \sim Normal(0,\ 1)} \qquad \boxed{T = \frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{V}(\hat{\beta}_j)}} \sim t \text{ with } N-2 \text{ d.f.}}$$

$\sigma_{\hat{\beta}_j}$ $\qquad$ $\hat{\sigma}_{\hat{\beta}_j}$

Testing hypotheses about β_0 or β_1

$$y = \beta_0 + \beta_1 x + u$$

1. State the **null & alternative hypotheses:** e.g., $H_0 : \beta_j = 0, H_1 : \beta_j \neq 0$
2. Define an appropriate **test statistic:** $\hat{\beta}_j$
3. Determine the **distribution of the test statistic under the null** hypothesis $\hat{\beta}_j \sim Normal\left(0,\ V(\hat{\beta}_j)\right)$
4. **Standardize the test statistic** to something with known/tabled probabilities for its sampling distribution (e.g., *Z, t, chi-square, F*)
   $$\boxed{Z = \frac{\hat{\beta}_j - 0}{\sigma_{\hat{\beta}_j}} \sim Normal(0,\ 1)} \qquad \boxed{T = \frac{\hat{\beta}_j - 0}{\hat{\sigma}_{\hat{\beta}_j}} \sim t \text{ with } N-2 \text{ d.f.}}$$
5. Choose a **significance level** (***α***, the *P*(Type I error)=*P*(reject the null when it is true), typically 0.01, 0.05, or 0.10) & a **rejection region** **OR** compute the **p-value** for the test statistic.
6. **Reject the null hypothesis if** the standardized statistic lies **in the rejection region (or if p-value≤*α*)**; fail to reject otherwise

MICHIGAN STATE UNIVERSITY

Example #1: Testing hypotheses about β_0 or β_1

```
. reg bwght cigs
```

Source	SS	df	MS
Model	13060.4194	1	13060.4194
Residual	561551.3	1386	405.159668
Total	574611.72	1387	414.283864

Number of obs	=	1388
F(1, 1386)	=	32.24
Prob > F	=	0.0000
R-squared	=	0.0227
Adj R-squared	=	0.0220
Root MSE	=	20.129

bwght	Coef.	Std. Err.
cigs	-.5137721	.0904909
_cons	119.7719	.5723407

Test the following hypotheses at the a = 0.05 level. Also find the p-values.

$H_0 : \beta_{cigs} = 0$ vs. $H_1 : \beta_{cigs} \neq 0$

and

$H_0 : \beta_{cigs} = 0$ vs. $H_1 : \beta_{cigs} < 0$

$$\boxed{T = \frac{\hat{\beta}_j - 0}{\hat{\sigma}_{\hat{\beta}_j}} \sim t \text{ with } N-2 \text{ d.f.}}$$

11

MICHIGAN STATE UNIVERSITY

T-stats and p-values in Stata output

```
. reg bwght cigs

      Source |       SS       df       MS              Number of obs =    1388
-------------+------------------------------           F(  1,  1386) =   32.24
       Model |  13060.4194     1  13060.4194           Prob > F      =  0.0000
    Residual |    561551.3  1386  405.159668           R-squared     =  0.0227
-------------+------------------------------           Adj R-squared =  0.0220
       Total |   574611.72  1387  414.283864           Root MSE      =  20.129

------------------------------------------------
       bwght |      Coef.   Std. Err.      t    P>|t|
-------------+----------------------------------
        cigs |  -.5137721   .0904909    -5.68   0.000
       _cons |   119.7719   .5723407   209.27   0.000
------------------------------------------------
```

The p-values reported by Stata are for $H_0 : \beta_j = 0$ vs. $H_1 : \beta_j \neq 0$

12

MICHIGAN STATE UNIVERSITY

Example #2: Testing hypotheses about β_0 or β_1

$$\log(crime) = \beta_0 + \beta_1 \log(enroll) + u$$

$$\widehat{\log(crime)} = -6.63 + 1.27 \log(enroll)$$
$$(1.03) \quad (0.11)$$
$$n = 97, R^2 = .585.$$

[Aside on interpreting results in log-log models]

crime is the annual number of crimes on college campuses and *enroll* is student enrollment. The numbers in parentheses are standard errors.

Use the regression output above to test the following hypotheses at the α=0.05 level. Also find the associated p-values.

$H_0 : \beta_1 = 1$ vs. $H_1 : \beta_1 \neq 1$
and
$H_0 : \beta_1 = 1$ vs. $H_1 : \beta_1 > 1$

13

Summary of Functional Forms Involving Logarithms $y = \beta_0 + \beta_1 x + u$

Model	Dependent Variable	Independent Variable	Interpretation of β_1
level-level	y	x	$\boxed{\beta_1 = \frac{\Delta y}{\Delta x}}$ $\Delta y = \beta_1 \Delta x$
level-log	y	$\log(x)$	$\boxed{\frac{\beta_1}{100} = \frac{\Delta y}{\%\Delta x}}$ $\Delta y = (\beta_1/100)\%\Delta x$
log-level	$\log(y)$	x	$\boxed{100\beta_1 = \frac{\%\Delta y}{\Delta x}}$ $\%\Delta y = (100\beta_1)\Delta x$
log-log	$\log(y)$	$\log(x)$	$\boxed{\beta_1 = \frac{\%\Delta y}{\%\Delta x}}$ $\%\Delta y = \beta_1 \%\Delta x$

Source: Wooldridge (2003)

[back]

Confidence intervals for β_0 or β_1

Recall from earlier in the course:

Two-sided, large-sample (1-α)% confidence interval for θ: $\hat{\theta} \pm z_{\alpha/2}\sigma_{\hat{\theta}}$

Two-sided, small-sample (1-α)% confidence interval for μ: $\bar{Y} \pm t_{\alpha/2}\hat{\sigma}_{\bar{Y}}$,

($N - 1$ d.f. for $t_{\alpha/2}$)

Two-sided, finite sample (1-α)% confidence interval for β_j (in the case of simple linear regression):

$$\boxed{\begin{array}{c}\hat{\beta}_j \pm t_{\alpha/2}\hat{\sigma}_{\hat{\beta}_j} \\ (N - 2 \text{ d.f. for } t_{\alpha/2})\end{array}}$$

Example #1: Confidence intervals for β_0 or β_1

```
. reg bwght cigs

      Source |       SS       df       MS              Number of obs =    1388
-------------+------------------------------           F(  1,  1386) =   32.24
       Model |  13060.4194     1  13060.4194           Prob > F      =  0.0000
    Residual |    561551.3  1386  405.159668           R-squared     =  0.0227
-------------+------------------------------           Adj R-squared =  0.0220
       Total |   574611.72  1387  414.283864           Root MSE      =  20.129

------------------------------------------------------
       bwght |      Coef.   Std. Err.      t    P>|t|
-------------+----------------------------------------
        cigs |  -.5137721   .0904909    -5.68   0.000
       _cons |   119.7719   .5723407   209.27   0.000
------------------------------------------------------
```

$$\hat{\beta}_j \pm t_{\alpha/2}\hat{\sigma}_{\hat{\beta}_j}$$
$(N-2$ d.f. for $t_{\alpha/2})$

a. Find the 95% (two-sided) confidence interval for β_{cigs}.
Relate this to $H_0: \beta_{cigs} = 0$ vs. $H_1: \beta_{cigs} \neq 0$ at α=0.05.

b. Find the 95% upper confidence interval for β_{cigs}.
Relate this to $H_0: \beta_{cigs} = 0$ vs. $H_1: \beta_{cigs} < 0$ at α=0.05.

MICHIGAN STATE UNIVERSITY

95% confidence intervals in Stata output

```
. reg bwght cigs

      Source |       SS       df       MS              Number of obs =    1388
-------------+------------------------------           F(  1,  1386) =   32.24
       Model |  13060.4194     1  13060.4194           Prob > F      =  0.0000
    Residual |    561551.3  1386  405.159668           R-squared     =  0.0227
-------------+------------------------------           Adj R-squared =  0.0220
       Total |   574611.72  1387  414.283864           Root MSE      =  20.129

------------------------------------------------------------------------------
       bwght |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        cigs |  -.5137721   .0904909    -5.68   0.000    -.6912861   -.3362581
       _cons |   119.7719   .5723407   209.27   0.000     118.6492    120.8946
------------------------------------------------------------------------------
```

18

MICHIGAN STATE UNIVERSITY

Example #2: Confidence intervals for β_0 or β_1

$$\log(crime) = \beta_0 + \beta_1 \log(enroll) + u$$

$$\widehat{\log(crime)} = \underset{(1.03)}{-6.63} + \underset{(0.11)}{1.27} \log(enroll)$$
$$n = 97,\ R^2 = .585.$$

crime is the annual number of crimes on college campuses and *enroll* is student enrollment. The numbers in parentheses are standard errors.

a. Find the 95% (two-sided) confidence interval for β_1.
   Relate this to H_0: $\beta_1 = 1$ vs. H_1: $\beta_1 \sim= 1$ at $\alpha=0.05$.
b. Find the 95% lower confidence interval for β_1.
   Relate this to H_0: $\beta_1 = 1$ vs. H_1: $\beta_1 > 1$ at $\alpha=0.05$.

$$\boxed{\hat{\beta}_j \pm t_{\alpha/2}\hat{\sigma}_{\hat{\beta}_j} \quad (N-2 \text{ d.f. for } t_{\alpha/2})}$$

19

Answers :

$2-sided\ \ 95\%\ CI\ for\ \beta_1:\ \ [1.054,\ 1.486]$

$Lower\ \ 1-sided\ \ 95\%\ CI\ for\ \beta_1:\ \ [1.089,\ \infty)$

Annotated Stata output (see handout)

Homework: Ch. 11 (cont'd)

1. Finish the other parts of Thursday's HW
2. Using the data in WMS 11.3, test H_0: β_1 =0 vs. H_1: β_1 ~= 0, and H_0: β_1 =0 vs. H_1: β_1 < 0, both at the at α=0.05 level. Also find the 95% two-sided and upper CIs, and relate the results to your hypothesis tests above.
3. Using the data in tourism.dta (on D2L) and Stata, regress household tourism expenditure (*tourismexp*) on household income (*income*). Interpret the estimate for β_1, and construct 99% two-sided and lower CIs for β_1. Use the CI results to test H_0: β_1=0.05 vs. H_1: β_1~= 0.05, and H: β_1=0.05 vs. H_1: β_1>0.05 at α=0.01 level.

- Please try to complete all Ch. 11 HW before class on Thursday so that we can go over it then (you won't turn in Ch. 11)

MICHIGAN STATE UNIVERSITY

Game plan for Thursday (last day of class)

- Finish any material on today's slides that we didn't get to
- Go over answers to additional practice problem
- Go over any questions you have on the Ch. 11 HW, past final exams, or other HWs/course material

Final exam details

- Cumulative but with emphasis on Ch. 7-Ch. 11
- Please bring paper, pencil, calculator, and cheat sheets (two 8.5x11" sheets, front and back). Please write last 4 digits of your PID on paper in advance to save time.
- Exam is closed book/notes except for cheat sheets
- Exam is in this room from 12:45-2:45 PM (hard stop) next Thursday, December 14

23