

Getting Started with SHAZAM *to accompany*

R. Carter Hill, William E. Griffiths and Guay C. Lim, *Principles of Econometrics*, Third Edition, Wiley, 2008

Chapters 2 to 4 of the textbook introduce the simple linear regression model with application to the estimation and analysis of the food expenditure function.

The SHAZAM Command File

SHAZAM commands give instructions. Commands are prepared in a SHAZAM command file. The SHAZAM system processes the command file line by line and produces an output file of statistical results.

To edit a SHAZAM command file use a text editor such as Notepad. Do not use a Word processor for editing command files.

SHAZAM commands for work with the food expenditure function are:

```
sample 1 40
read (food.txt) FOOD INCOME
* Descriptive statistics
stat FOOD INCOME / pcor
* OLS estimation
ols FOOD INCOME / list pcov
* Calculate a confidence interval for the slope parameter
confid INCOME / tcrit=2.024
* Test the null hypothesis H0: beta_2 = 15
test INCOME=15
stop
```

The [SAMPLE command](#) sets the sample size.

The [READ command](#) loads the data set and assigns variable names.

The [STAT command](#) gives descriptive statistics for the data set.

The [OLS command](#) is used for least squares estimation.

The [CONFID command](#) and [TEST command](#) are used for interval estimation and hypothesis testing from the previous least squares estimation.

Lines starting with * are [comments](#).

For preparing commands upper case and lower case are interchangeable with the exception of filenames specified on the **READ** command.

SHAZAM Output of Least Squares Estimation Results

For the food expenditure function SHAZAM estimation output shows:

```

|_ols FOOD INCOME / list pcov

OLS ESTIMATION
  40 OBSERVATIONS      DEPENDENT VARIABLE= FOOD
...NOTE...SAMPLE RANGE SET TO:      1,      40

R-SQUARE =      0.3850      R-SQUARE ADJUSTED =      0.3688
VARIANCE OF THE ESTIMATE-SIGMA**2 =      8013.3
STANDARD ERROR OF THE ESTIMATE-SIGMA =      89.517
SUM OF SQUARED ERRORS-SSE=      0.30451E+06
MEAN OF DEPENDENT VARIABLE =      283.57

VARIABLE      ESTIMATED      STANDARD      T-RATIO      PARTIAL STANDARDIZED ELASTICITY
NAME      COEFFICIENT      ERROR      38 DF      P-VALUE CORR. COEFFICIENT      AT MEANS
INCOME      10.210      2.093      4.877      0.000 0.620      0.6205      0.7058
CONSTANT      83.416      43.41      1.922      0.062 0.298      0.0000      0.2942
    
```

The results can be compared with Figure 2.9, page 25 of the textbook.

Statistics reported are:

		Textbook reference
R-SQUARE	R^2	page 83
VARIANCE OF THE ESTIMATE-SIGMA**2	$\hat{\sigma}^2 = \frac{1}{N-2} \sum \hat{e}_i^2$	page 34
STANDARD ERROR OF THE ESTIMATE-SIGMA	$\hat{\sigma}$	
SUM OF SQUARED ERRORS-SSE	$SSE = \sum \hat{e}_i^2$	page 21
MEAN OF DEPENDENT VARIABLE	$\bar{y} = \frac{1}{N} \sum y_i$	page 18

The least squares estimation results listed on the SHAZAM output are interpreted as:

VARIABLE NAME	ESTIMATED COEFFICIENT	STANDARD ERROR	T-RATIO 38 DF	ELASTICITY AT MEANS
INCOME	b_2	$se(b_2)$	$\frac{b_2}{se(b_2)}$	$b_2 \left(\frac{\bar{x}}{\bar{y}} \right)$
CONSTANT	b_1	$se(b_1)$	$\frac{b_1}{se(b_1)}$	

Note: the SHAZAM output reports the intercept estimate as the final parameter estimate in the row with the label `CONSTANT`.

The income elasticity evaluated at the sample means is described on pages 23–4. The estimates of the standard errors of the least squares estimators are given on page 36.

The test statistic given in the `T-RATIO` column is for a test that the coefficient is significantly different from zero. This is the test of significance discussed in Chapter 3. SHAZAM reports an accompanying p-value for a two-tail test. That is, for the slope coefficient the t-statistic is for a test of:

$$H_0: \beta_2 = 0 \quad \text{against} \quad H_1: \beta_2 \neq 0$$

The reported p-value is calculated as: $2 \cdot P\left(t_{(N-2)} \geq \left| \frac{b_2}{\text{se}(b_2)} \right| \right)$

The SHAZAM output reports 3 decimal places of accuracy for the p-value. Therefore, a value of 0.000 means the p-value is less than 0.0005. This leads to the decision that the null hypothesis of a zero coefficient can be rejected at any reasonable significance level.

Note: For a one-tail test, a p-value can be obtained from the p-value reported for a two-tail test.

The **PCOV** option on the **OLS** command reports the estimated variances and covariances of the least squares estimators as:

VARIANCE-COVARIANCE MATRIX OF COEFFICIENTS		
INCOME	4.3818	
CONSTANT	-85.903	1884.4

The numerical results correspond to the presentation in page 35 of the text and are interpreted as:

VARIANCE-COVARIANCE MATRIX OF COEFFICIENTS		
INCOME	$\hat{\text{var}}(b_2)$	
CONSTANT	$\hat{\text{cov}}(b_1, b_2)$	$\hat{\text{var}}(b_1)$
	INCOME	CONSTANT

The **LIST** option on the **OLS** command gives additional output that can be compared with Table 2.3, page 35.

i	y_i	$\hat{y}_i = b_1 + b_2x_i$	$\hat{e}_i = y_i - \hat{y}_i$	residual plot
OBS.	OBSERVED	PREDICTED	CALCULATED	
NO.	VALUE	VALUE	RESIDUAL	
1	115.22	121.09	-5.8696	*
2	135.98	128.24	7.7437	*
3	119.34	131.91	-12.572	*I
4	114.96	144.98	-30.020	* I
5	187.05	210.73	-23.680	*I
6	243.92	215.94	27.983	I *
7	267.43	228.39	39.037	I *
8	238.71	234.11	4.5997	*
9	295.94	239.83	56.112	I *
10	317.78	250.75	67.028	I *
more output for observations 11 to 40				
RESIDUAL SUM = -0.16058E-11 RESIDUAL VARIANCE = 8013.3				

The final calculations listed above are obtained as:

$$\text{RESIDUAL SUM} = \sum \hat{e}_i \qquad \text{RESIDUAL VARIANCE} = \hat{\sigma}^2$$

The OLS estimated residuals should sum to zero. The output reports the number -0.16058E-11. This shows that computer calculations are subject to rounding error that can vary from machine to machine.

Interval Estimation

The **CONFID** command is used to obtain confidence interval estimates using the results from the previous estimation command. The variable names specified on the **CONFID** command represent coefficients. For the food expenditure example SHAZAM output for a 95% interval estimate for the slope coefficient is:

```
|_confid INCOME / tcrit=2.024
CONFIDENCE INTERVALS BASED ON T-DISTRIBUTION WITH 38 D.F.
- T CRITICAL VALUE = 2.024
NAME          LOWER      COEFFICIENT      UPPER      STD. ERROR
INCOME        5.9729          10.210          14.446          2.0933
```

This matches the results given on page 98. Take careful note of the interpretation of an interval estimate discussed on page 98.

More Hypothesis Testing

The **TEST** command can be used for hypothesis testing. The variable names specified on the **TEST** command represent coefficients from the previous estimation results. Chapter 3.4.2, pages 61–2, studies the left-tail test:

$$H_0 : \beta_2 \geq 15 \quad \text{against} \quad H_1 : \beta_2 < 15$$

SHAZAM output shows the results:

_test INCOME=15				
TEST VALUE =	-4.7904	STD. ERROR OF TEST VALUE	2.0933	
T STATISTIC =	-2.2884634	WITH	38 D.F.	P-VALUE= 0.02776

The above calculations are obtained as:

$$\text{TEST VALUE} = b_2 - 15 \quad \text{STD. ERROR OF TEST VALUE} \quad \text{se}(b_2)$$

$$\text{T STATISTIC} = t = \frac{b_2 - 15}{\text{se}(b_2)} \quad \text{P-VALUE} = 2 \cdot P(t_{(N-2)} \geq |t|)$$

The calculation of a p-value for this test is discussed in Chapter 3.5.2, page 66, and illustrated in Figure 3.6, page 66.

SHAZAM reports p-values for a two-tail test. This can be converted to a p-value for the left-tail test of interest. The SHAZAM results give:

$$2 \cdot P(t_{(38)} \geq 2.29) = 0.02776$$

Therefore,

$$P(t_{(38)} \geq 2.29) = P(t_{(38)} \leq -2.29) = \frac{0.02776}{2} = 0.0139$$