

Stata 10 Tutorial 7

TOPIC: Estimation and Hypothesis Testing in Multiple Linear Regression Models

DATA: **auto1.dta** (a *Stata*-format data file you created in *Stata Tutorial 1*)

TASKS: *Stata Tutorial 7* introduces you to OLS estimation of **multiple linear regression models** containing two or more regressors, and demonstrates how to perform various common types of hypothesis tests in such models. It is intended to give you some practice in formulating multiple linear regression models and in computing F-tests and t-tests of linear coefficient restrictions on such models. It introduces you to the specification and interpretation of multiple linear regression models that include as regressors **polynomial terms** in explanatory variables and **interaction terms** between continuous explanatory variables. It also introduces the *Stata* **lincom** command, a post-estimation command that computes linear combinations of coefficient estimates, specifically point estimates, standard errors, t-statistics, two-tail p-values, and two-sided confidence intervals for specified linear combinations of coefficient estimates.

- The *Stata commands* that constitute the primary subject of this tutorial are:

regress	Used to perform OLS estimation of multiple linear regression models.
test	Used to compute F-tests of linear coefficient equality restrictions after OLS estimation.
lincom	Used after estimation to compute linear combinations of coefficient estimates and associated statistics.

- The *Stata statistical functions* introduced in this tutorial are:

ttail (<i>df</i> , <i>t</i> ₀)	Computes <i>p-values</i> of <i>calculated t-statistics</i> .
invttail (<i>df</i> , <i>p</i>)	Computes <i>right-tail critical values</i> of <i>t-distributions</i> .
Ftail (<i>df</i> ₁ , <i>df</i> ₂ , <i>F</i> ₀)	Computes <i>p-values</i> of <i>calculated F-statistics</i> .
invFtail (<i>df</i> ₁ , <i>df</i> ₂ , α)	Computes <i>right-tail critical values</i> of <i>F-distributions</i> .

NOTE: *Stata* commands are *case sensitive*. All *Stata* **command names** must be typed in the Command window in **lower case letters**.

□ Overview of Regression Models Used in this Tutorial

The following multiple linear regression models are estimated and tested in this *Stata* tutorial.

$$\text{price}_i = \beta_0 + \beta_1 \text{weight}_i + \beta_2 \text{mpg}_i + u_i \quad (1)$$

$$\ln(\text{price}_i) = \alpha_0 + \alpha_1 \ln(\text{weight}_i) + \alpha_2 \ln(\text{mpg}_i) + u_i \quad (3)$$

$$\text{price}_i = \beta_0 + \beta_1 \text{weight}_i + \beta_2 \text{mpg}_i + \beta_3 \text{weight}_i^2 + u_i \quad (7)$$

$$\text{price}_i = \beta_0 + \beta_1 \text{weight}_i + \beta_2 \text{mpg}_i + \beta_3 \text{weight}_i^2 + \beta_4 \text{mpg}_i^2 + u_i \quad (11)$$

$$\text{price}_i = \beta_0 + \beta_1 \text{weight}_i + \beta_2 \text{mpg}_i + \beta_3 \text{weight}_i^2 + \beta_4 \text{mpg}_i^2 + \beta_5 \text{weight}_i \text{mpg}_i + u_i \quad (13)$$

□ Preparing for Your *Stata* Session

Before beginning your *Stata* session, use Windows Explorer to copy the *Stata*-format dataset **auto1.dta** to the *Stata working directory* on the C:-drive or D:-drive of the computer at which you are working.

- **On the computers in Dunning 350**, the default *Stata* working directory is usually **C:\data**.
- **On the computers in MC B111**, the default *Stata* working directory is usually **D:\courses**.

□ Start Your *Stata* Session

To start your *Stata* session, double-click on the ***Stata 10 icon*** in the Windows desktop.

After you double-click the ***Stata 10 icon***, you will see the now familiar screen of four *Stata* windows.

□ Record Your *Stata* Session and *Stata* Commands -- log using, cmdlog using

To record your *Stata* session, including all the *Stata* commands you enter and the results (output) produced by these commands, make a **.log** file named **351tutorial7.log**. To open (begin) the log file **351tutorial7.log**, enter in the Command window:

```
log using 351tutorial7.log
```

This command opens a file called **351tutorial7.log** in the current *Stata* working directory. Remember that once you have opened the **351tutorial7.log** file, a copy of all the commands you enter during your *Stata* session and of all the results they produce is recorded in that **351tutorial7.log** file.

An alternative way to open the **.log** file **351tutorial7.log** is to click on the **Log** button; click on **Save as type:** and select **Log (*.log)**; click on the **File name:** box and type the file name **351tutorial7**; and click on the **Save** button.

To record only the *Stata* commands you type during your *Stata* session, use the *Stata* **cmdlog using** command. To start (open) the command log file **351tutorial7.txt**, enter in the Command window:

```
cmdlog using 351tutorial7
```

This command opens a plain text-format (ASCII) file called **351tutorial7.txt** in the current *Stata* working directory. All commands you enter during your *Stata* session are recorded in this file.

□ Loading a *Stata*-Format Dataset into *Stata* -- use

Load, or read, into memory the dataset you are using. To load the *Stata*-format data file **auto1.dta** into memory, enter in the Command window:

```
use auto1
```

This command loads into memory the *Stata*-format dataset **auto1.dta**.

❑ Familiarize Yourself with the Current Dataset

To familiarize (or re-familiarize) yourself with the contents of the current dataset, type in the Command window the following commands:

```
describe, short  
describe  
summarize
```

❑ Computing linear combinations of coefficient estimates -- **lincom**

- ***Basic Syntax:***

lincom *exp* [, level(#)]

where *exp* is a user-specified **linear combination of coefficient estimates**. The **lincom** command computes point estimates, standard errors, t-statistics, p-values, and two-sided confidence intervals for a specified linear combination of coefficient estimates from the most recently executed model estimation command, such as the **regress** command. Note that the linear combination specified in *exp* *cannot* contain an equality sign (=).

The **level(#)** *option* on the **lincom** command specifies the **confidence level** to be used in computing the two-sided confidence interval for the specified linear combination of regression coefficients. The value of # specifies the desired confidence level in percentage points; the default confidence level is 95 percent (i.e., $1-\alpha = 0.95$).

- **Examples:** The **lincom** command is most useful in multiple regression models containing two or more explanatory variables. Several examples of its use are given in subsequent sections of this tutorial.

❑ Estimation and Inference in Multiple Linear Regression Models: Introduction

This section introduces you to OLS estimation of **multiple linear regression models** containing two or more regressors. It also demonstrates how to perform various common types of hypothesis tests in such models.

Consider the following multiple linear regression model for price_{*i*}:

$$\text{price}_i = \beta_0 + \beta_1 \text{weight}_i + \beta_2 \text{mpg}_i + u_i \quad (1)$$

where

price_{*i*} = the price of the *i*-th car (in US dollars);

weight_{*i*} = the weight of the *i*-th car (in pounds); and

mpg_{*i*} = the miles per gallon for the *i*-th car (in miles per gallon).

Note: For PRE (1), the number of regression coefficients is $K = 3$.

Regression equation (1) contains **two explanatory variables**, *weight_{*i*}* and *mpg_{*i*}*.

- ◆ The **marginal effect of *weight_{*i*}*** in equation (1) is:

$$\frac{\partial \text{price}_i}{\partial \text{weight}_i} = \beta_1 = \text{a constant}$$

- ◆ The **marginal effect of *mpg_{*i*}*** in equation (1) is:

$$\frac{\partial \text{price}_i}{\partial \text{mpg}_i} = \beta_2 = \text{a constant}$$

- To estimate model (1) by OLS for the full sample of $N = 74$ sample observations in dataset **auto1.dta**, enter in the Command window:

```
regress price weight mpg
```

Testing the individual significance of the slope coefficient estimates

Individual significance tests are **two-tail tests** of the following statistical hypothesis:

$$H_0: \beta_j = 0 \quad \text{versus} \quad H_1: \beta_j \neq 0.$$

Such individual significance tests can be performed using either a *two-tail t-test* or an *F-test*.

The *feasible test statistics* for such *individual significance tests* are either the *t-statistic* for $\hat{\beta}_j$ given by

$$t(\hat{\beta}_j) = \frac{\hat{\beta}_j - \beta_j}{\widehat{\text{se}}(\hat{\beta}_j)} \sim t[N - K]$$

or the *F-statistic* for $\hat{\beta}_j$ given by

$$F(\hat{\beta}_j) = \frac{(\hat{\beta}_j - \beta_j)^2}{\widehat{\text{Var}}(\hat{\beta}_j)} \sim F[1, N - K].$$

Example 1: In regression equation (1), perform a two-tail test of the hypothesis

$$H_0: \beta_1 = 0 \quad \text{versus} \quad H_1: \beta_1 \neq 0.$$

The null hypothesis $H_0: \beta_1 = 0$ implies that the *marginal effect* of the explanatory variable *weight_i* equals zero in PRE (1):

$$H_0: \beta_1 = 0 \quad \Rightarrow \quad \frac{\partial \text{price}_i}{\partial \text{weight}_i} = \beta_1 = 0$$

This hypothesis can be tested using either a two-tail t-test or an equivalent F-test.

1. A two-tail t-test: Under the null hypothesis $H_0: \beta_1 = 0$, the *sample value of the t-statistic* for $\hat{\beta}_1$ is

$$t_0(\hat{\beta}_1) = \frac{\hat{\beta}_1 - \beta_1}{\widehat{\text{se}}(\hat{\beta}_1)} = \frac{\hat{\beta}_1 - 0}{\widehat{\text{se}}(\hat{\beta}_1)} = \frac{\hat{\beta}_1}{\widehat{\text{se}}(\hat{\beta}_1)}.$$

The *null distribution* of $t_0(\hat{\beta}_1)$ is:

$$t_0(\hat{\beta}_1) \sim t[N - K] = t[74 - 3] = t[71].$$

- Compute the *sample value of the t-statistic*. Enter the commands:

```
scalar t0b1 = _b[weight]/_se[weight]
scalar list t0b1
```

- Compute the *two-tail p-value* of the calculated t-statistic. Enter the command:

```
display 2*ttail(71, abs(t0b1))
```

- Another way to do what the preceding three commands have done is to use the *Stata lincom* command to compute both the *sample value of the t-statistic* and its *two-tail p-value*. Enter the command:

```
lincom _b[weight]
```

What inference would you draw from the outcome of this two-tail t-test?

- Compute the *two-tail 100 α percent critical values* $t_{\alpha/2}[71]$ and $-t_{\alpha/2}[71]$ of the $t[71]$ -distribution. Use the *invttail(df, p)* function, where $p = \alpha/2$.
 - At the **1% significance level** for which $\alpha = 0.01$, $\alpha/2 = 0.01/2 = 0.005$. Enter the commands:

```
display invttail(71, 0.005)      (returns value of  $t_{0.005}[71]$ )
display -invttail(71, 0.005)    (returns value of  $-t_{0.005}[71]$ )
```

- At the **5% significance level** for which $\alpha = 0.05$, $\alpha/2 = 0.05/2 = 0.025$. Enter the commands:

```
display invttail(71, 0.025)      (returns value of  $t_{0.025}[71]$ )
display -invttail(71, 0.025)    (returns value of  $-t_{0.025}[71]$ )
```

- At the **10% significance level** for which $\alpha = 0.10$, $\alpha/2 = 0.10/2 = 0.05$. Enter the commands:

```
display invttail(71, 0.05)      (returns value of  $t_{0.05}[71]$ )
display -invttail(71, 0.05)   (returns value of  $-t_{0.05}[71]$ )
```

2. **An F-test:** Under the null hypothesis $H_0: \beta_1 = 0$, the **sample value of the F-statistic** for $\hat{\beta}_1$ is

$$F_0(\hat{\beta}_1) = \frac{(\hat{\beta}_1 - \beta_1)^2}{\text{Var}(\hat{\beta}_1)} = \frac{(\hat{\beta}_1 - 0)^2}{\text{Var}(\hat{\beta}_1)} = \frac{\hat{\beta}_1^2}{\text{Var}(\hat{\beta}_1)} = \frac{\hat{\beta}_1^2}{\hat{\text{se}}(\hat{\beta}_1)^2}.$$

The **null distribution** of $F_0(\hat{\beta}_1)$ is:

$$F_0(\hat{\beta}_1) \sim F[1, N - K] = F[1, 74 - 3] = F[1, 71].$$

- Compute the **sample value of the F-statistic**. Enter the commands:

```
scalar F0b1 = _b[weight]^2/_se[weight]^2
scalar list F0b1
```

- Compute the **p-value** of the calculated F-statistic. Enter the command:

```
display Ftail(1, 71, F0b1)
```

- Another way to do what the preceding three commands have done is to use the **Stata test** command to compute both the **sample value of the F-statistic** and its **p-value**. Enter either one of the following **test** commands:

```
test weight = 0
test weight
```

What inference would you draw from the outcome of this F-test?

- Compute the **100 α percent critical value** $F_{\alpha}[1, 71]$ of the $F[1, 71]$ -distribution.

Use the **invFtail**(df_1 , df_2 , α) statistical function to compute critical values of the **F**[df_1 , df_2]-distribution, where df_1 is the *numerator* degrees of freedom, df_2 is the *denominator* degrees of freedom, and α is the chosen *significance level*.

In the present case, $df_1 = 1$ (the number of coefficient restrictions specified by the null hypothesis H_0) and $df_2 = 71$ (the degrees of freedom for the RSS of the unrestricted regression equation (1)). The following commands compute and display the *critical values* of the **F**[1,71] distribution for significance levels $\alpha = 0.01, 0.05, 0.10$.

- At the **1% significance level**, $\alpha = 0.01$. Enter the command:

`display invFtail(1,71, 0.01)` (returns value of $F_{0.01}[1, 71]$)

- At the **5% significance level**, $\alpha = 0.05$. Enter the command:

`display invFtail(1,71, 0.05)` (returns value of $F_{0.05}[1, 71]$)

- At the **10% significance level**, $\alpha = 0.10$. Enter the command:

`display invFtail(1,71, 0.10)` (returns value of $F_{0.10}[1, 71]$)

These are the critical values of $F_\alpha[1, 71]$ for testing $H_0: \beta_2 = 0$ against $H_1: \beta_2 \neq 0$. Would you reject the null hypothesis $\beta_2 = 0$ in favor of the alternative hypothesis $\beta_2 \neq 0$ at the 10 percent significance level, at the 5 percent significance level, and/or at the 1 percent significance level? Do the two tests – the two-tail t-test and the F-test – yield the same inference?

Example 2: In regression equation (1), perform a two-tail test of the hypothesis

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

The null hypothesis $H_0: \beta_2 = 0$ implies that the *marginal effect* of the explanatory variable mpg_i equals zero in PRE (1):

$$H_0: \beta_2 = 0 \quad \Rightarrow \quad \frac{\partial \text{price}_i}{\partial \text{mpg}_i} = \beta_2 = 0$$

This hypothesis can be tested using either a two-tail t-test or an equivalent F-test.

1. ***A two-tail t-test:*** Under the null hypothesis $H_0: \beta_2 = 0$, the *sample value of the t-statistic* for $\hat{\beta}_2$ is

$$t_0(\hat{\beta}_2) = \frac{\hat{\beta}_2 - \beta_2}{\widehat{\text{se}}(\hat{\beta}_2)} = \frac{\hat{\beta}_2 - 0}{\widehat{\text{se}}(\hat{\beta}_2)} = \frac{\hat{\beta}_2}{\widehat{\text{se}}(\hat{\beta}_2)}.$$

The *null distribution* of $t_0(\hat{\beta}_2)$ is:

$$t_0(\hat{\beta}_2) \sim t[N - K] = t[74 - 3] = t[71].$$

- Compute the *sample value of the t-statistic*. Enter the commands:

```
scalar t0b2 = _b[mpg]/_se[mpg]
scalar list t0b2
```

- Compute the *two-tail p-value* of the calculated t-statistic. Enter the command:

```
display 2*ttail(71, abs(t0b2))
```

- Another way to do what the preceding three commands have done is to use the *Stata lincom* command to compute both the *sample value of the t-statistic* and its *two-tail p-value*. Enter the command:

```
lincom _b[mpg]
```

What inference would you draw from the outcome of this two-tail t-test?

- Re-compute the *two-tail 100 α percent critical values* $t_{\alpha/2}[71]$ and $-t_{\alpha/2}[71]$ of the $t[71]$ -distribution at significance levels $\alpha = 0.01, 0.05, 0.10$. Use the *invttail(df, p)* function with $p = \alpha/2$. Enter the commands:

```
display invttail(71, 0.005)      (returns value of  $t_{0.005}[71]$ )
display -invttail(71, 0.005)   (returns value of  $-t_{0.005}[71]$ )
```

```

display invttail(71, 0.025)    (returns value of  $t_{0.025}[71]$ )
display -invttail(71, 0.025)  (returns value of  $-t_{0.025}[71]$ )

display invttail(71, 0.05)   (returns value of  $t_{0.05}[71]$ )
display -invttail(71, 0.05)  (returns value of  $-t_{0.05}[71]$ )

```

2. **An F-test:** Under the null hypothesis $H_0: \beta_2 = 0$, the *sample value of the F-statistic* for $\hat{\beta}_2$ is

$$F_0(\hat{\beta}_2) = \frac{(\hat{\beta}_2 - \beta_2)^2}{\text{Var}(\hat{\beta}_2)} = \frac{(\hat{\beta}_2 - 0)^2}{\text{Var}(\hat{\beta}_2)} = \frac{\hat{\beta}_2^2}{\text{Var}(\hat{\beta}_2)} = \frac{\hat{\beta}_2^2}{\text{s}\hat{\text{e}}(\hat{\beta}_2)^2}.$$

The *null distribution* of $F_0(\hat{\beta}_2)$ is:

$$F_0(\hat{\beta}_2) \sim F[1, N - K] = F[1, 74 - 3] = F[1, 71].$$

- Compute the *sample value of the F-statistic*. Enter the commands:

```

scalar F0b2 = _b[mpg]^2/_se[mpg]^2
scalar list F0b2

```

- Compute the *p-value* of the calculated F-statistic. Enter the command:

```
display Ftail(1, 71, F0b2)
```

- Another way to do what the preceding three commands have done is to use the *Stata test* command to compute both the *sample value of the F-statistic* and its *p-value*. Enter either one of the following *test* commands:

```

test mpg = 0
test mpg

```

What inference would you draw from the outcome of this F-test?

- Re-compute the **100 α percent critical values** $F_\alpha[1, 71]$ of the $F[1, 71]$ -distribution at significance levels $\alpha = 0.01, 0.05, 0.10$.

- At the **1% significance level**, $\alpha = 0.01$. Enter the command:

`display invFtail(1,71, 0.01)` (returns value of $F_{0.01}[1, 71]$)

- At the **5% significance level**, $\alpha = 0.05$. Enter the command:

`display invFtail(1,71, 0.05)` (returns value of $F_{0.05}[1, 71]$)

- At the **10% significance level**, $\alpha = 0.10$. Enter the command:

`display invFtail(1,71, 0.10)` (returns value of $F_{0.10}[1, 71]$)

These are the critical values of $F_\alpha[1, 71]$ for testing $H_0: \beta_2 = 0$ against $H_1: \beta_2 \neq 0$. Would you reject the null hypothesis $\beta_2 = 0$ in favor of the alternative hypothesis $\beta_2 \neq 0$ at the 10 percent significance level, at the 5 percent significance level, and/or at the 1 percent significance level? Do the two tests – the two-tail t-test and the F-test – yield the same inference?

Testing the joint significance of all slope coefficient estimates

A **joint significance test of all slope coefficient estimates** is a test the hypothesis that the two slope coefficients in PRE (1) are jointly equal to zero. The null and alternative hypotheses are:

$$H_0: \beta_j = 0 \quad \forall j=1, 2 \quad \Rightarrow \quad \beta_1 = 0 \text{ and } \beta_2 = 0 \quad (2)$$

$$H_1: \beta_j \neq 0 \quad j=1, 2 \quad \Rightarrow \quad \beta_1 \neq 0 \text{ and/or } \beta_2 \neq 0.$$

- The following sequence of **test** commands with the **accumulate** option will perform a joint F-test of the null hypothesis H_0 in (2). Enter the **test** commands:

```
test weight = 0
test mpg = 0, accum
```

Note that the **accumulate option** can be abbreviated as **accum**. In fact, it can even be abbreviated as **a**. This series of **test** commands computes the sample value of the F-statistic and its p-value.

- Alternatively, a single **test** command can be used to compute this joint F-test when the null hypothesis specifies zero or exclusion restrictions on two or more regression coefficients, as it does in the case of the null and alternative hypotheses in (2) above. Enter the command:

```
test weight mpg
```

This **test** command performs a **joint F-test** of the null hypothesis that the regression coefficient of the regressor weight_i equals zero *and* the regression coefficient of the regressor mpg_i equals zero.

- Use the **invFtail(df_1 , df_2 , α)** statistical function to compute critical values of the **F[df_1 , df_2]-distribution**, where df_1 is the *numerator* degrees of freedom, df_2 is the *denominator* degrees of freedom, and α is the chosen *significance level*.

In the present case, $df_1 = 2$ (the number of coefficient restrictions specified by the null hypothesis H_0) and $df_2 = 71$ (the degrees of freedom for the RSS of the unrestricted regression equation (1)). The following commands compute and display the *critical values* of the **F[2,71] distribution** for significance levels $\alpha = 0.01, 0.05, 0.10$.

- At the **1% significance level**, $\alpha = 0.01$. Enter the command:

```
display invFtail(2,71, 0.01) (returns value of  $F_{0.01}[2, 71]$ )
```

- At the **5% significance level**, $\alpha = 0.05$. Enter the command:

```
display invFtail(2,71, 0.05) (returns value of  $F_{0.05}[2, 71]$ )
```

- At the **10% significance level**, $\alpha = 0.10$. Enter the command:

```
display invFtail(2,71, 0.10) (returns value of  $F_{0.10}[2, 71]$ )
```

These are the critical values appropriate for testing hypothesis (2) above.

Would you reject the null hypothesis in (2) at the 10 percent significance level, at the 5 percent significance level, and/or at the 1 percent significance level?

□ Testing Equality of Slope Coefficients in Multiple Linear Regression Models: Introduction to the *lincom* Command

Recall the multiple linear regression model for price_{*i*} considered in the previous section:

$$\text{price}_i = \beta_0 + \beta_1 \text{weight}_i + \beta_2 \text{mpg}_i + u_i \quad (1)$$

In this regression equation, there is in general no reason to think that the slope coefficients β_1 and β_2 will be equal, because the explanatory variables weight_i and mpg_i will usually have different units of measurement. However, consider the following **log-log regression model** for car prices:

$$\ln(\text{price}_i) = \alpha_0 + \alpha_1 \ln(\text{weight}_i) + \alpha_2 \ln(\text{mpg}_i) + u_i \quad (3)$$

In the log-log regression model (3), the slope coefficients α_1 and α_2 are **partial elasticity coefficients**, and hence do not depend on the units of measurement for price_{*i*}, weight_{*i*} and mpg_{*i*}. In fact,

$$\alpha_1 = \frac{\partial \ln(\text{price}_i)}{\partial \ln(\text{weight}_i)} = \text{the partial elasticity of } \textit{price}_i \text{ wrt } \textit{weight}_i;$$

$$\alpha_2 = \frac{\partial \ln(\text{price}_i)}{\partial \ln(\text{mpg}_i)} = \text{the partial elasticity of } \textit{price}_i \text{ wrt } \textit{mpg}_i.$$

In log-log models such as that given by population regression equation (3), testing the equality of pairs of slope coefficients makes perfect sense, since these slope coefficients are comparable in terms of their magnitude.

- Before estimating regression equation (3), it is necessary to create the variables $\ln(\text{price}_i)$, $\ln(\text{weight}_i)$ and $\ln(\text{mpg}_i)$. Enter the commands:

```
generate lnprice = ln(price)
generate lnweight = ln(weight)
generate lnmpg = ln(mpg)
```

- Estimate by OLS regression equation (3). Enter the command:

```
regress lnprice lnweight lnmpg
```

- **Test 1:** Use the results of OLS estimation of the log-log model (3) to test the proposition that the **partial elasticity of $price_i$ with respect to $weight_i$ is equal to the partial elasticity of $price_i$ with respect to mpg_i .**

- ◆ The **null and alternative hypotheses** are:

$$H_0: \alpha_1 = \alpha_2 \quad \Rightarrow \quad \alpha_1 - \alpha_2 = 0 \quad (4)$$

$$H_1: \alpha_1 \neq \alpha_2 \quad \Rightarrow \quad \alpha_1 - \alpha_2 \neq 0$$

- ◆ Use a *Stata* **lincom** command to compute a **two-tail t-test** of the null hypothesis (4) above. Enter the command:

```
lincom _b[lnweight] - _b[lnmpg]
```

Carefully inspect the results of this **lincom** command. State the inference you would draw from this t-test.

- ◆ Now use a *Stata* **test** command to compute an **F-test** of the null hypothesis (4). Enter either of the following commands:

```
test lnweight - lnmpg = 0
test lnweight = lnmpg
```

Carefully inspect the results of this **test** command. How do you know that this **F-test** yields the same inference as the **two-tail t-test** of the same null hypothesis (4)?

- **Test 2:** Use the results of OLS estimation of the log-log model (3) to test the proposition that the **partial elasticity of $price_i$ with respect to $weight_i$ is equal in magnitude but opposite in sign to the partial elasticity of $price_i$ with respect to mpg_i .**

- ◆ The **null and alternative hypotheses** for this test are:

$$H_0: \alpha_1 = -\alpha_2 \quad \Rightarrow \quad \alpha_1 + \alpha_2 = 0 \quad (5)$$

$$H_1: \alpha_1 \neq -\alpha_2 \quad \Rightarrow \quad \alpha_1 + \alpha_2 \neq 0$$

- ◆ Use a *Stata* **lincom** command to compute a **two-tail t-test** of the null hypothesis above. Enter the command:

```
lincom _b[lnweight] + _b[lnmpg]
```

Carefully inspect the results of this **lincom** command. State the inference you would draw from this t-test.

- ◆ Use a *Stata* **test** command to compute an **F-test** of the null hypothesis (5). Enter either of the following commands:

```
test lnweight + lnmpg = 0
test lnweight = -lnmpg
```

Carefully inspect the results of this **test** command. How do you know that this **F-test** yields the same inference as the **two-tail t-test** of the same null hypothesis (5)?

□ Interaction Variables: Polynomial Terms in Explanatory Variables

Nature: Polynomial terms refer to integer powers of explanatory variables. For example, the **second-order polynomial term** in the variable X_{ji} is X_{ji} -squared, X_{ji}^2 ; the **third-order polynomial term** in the variable X_{ji} is X_{ji} -cubed, X_{ji}^3 .

Usage: Polynomial terms in explanatory variables are one way of allowing for **nonlinear relationships** between the dependent variable Y_i and individual explanatory variables such as X_{ji} .

Consider the following regression equation:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{1i}^2 + u_i. \quad (6)$$

Compare the *marginal effects* on Y_i of the two explanatory variables X_{1i} and X_{2i} in equation (6).

- ◆ The **marginal effect of X_{1i}** in equation (6) is:

$$\frac{\partial Y_i}{\partial X_{1i}} = \beta_1 + 2\beta_3 X_{1i} = \text{a linear function of } X_{1i}.$$

- ◆ The **marginal effect of X_{2i}** in equation (6) is:

$$\frac{\partial Y_i}{\partial X_{2i}} = \beta_2 = \text{a constant.}$$

Example 1: a polynomial in one explanatory variable, $weight_i$

Consider the following multiple regression equation:

$$price_i = \beta_0 + \beta_1 weight_i + \beta_2 mpg_i + \beta_3 weight_i^2 + u_i \quad (7)$$

where

$price_i$ = the price of the i -th car (in US dollars);

$weight_i$ = the weight of the i -th car (in pounds);

$weight_i^2$ = the square of $weight_i$; and

mpg_i = the miles per gallon for the i -th car (in miles per gallon).

Note: For PRE (7), the number of regression coefficients is $K = 4$.

Regression equation (7) contains **two explanatory variables**, $weight_i$ and mpg_i , but **three regressors** – $weight_i$, $weight_i^2$, and mpg_i .

Marginal Effects of the Explanatory Variables

- ◆ The **marginal effect of $weight_i$** in equation (7) is:

$$\frac{\partial \text{price}_i}{\partial \text{weight}_i} = \beta_1 + 2\beta_3 \text{weight}_i = \text{a linear function of } \text{weight}_i.$$

- ◆ The **marginal effect of mpg_i** in equation (7) is:

$$\frac{\partial \text{price}_i}{\partial \text{mpg}_i} = \beta_2 = \text{a constant.}$$

- Before estimating regression equation (7), it is necessary to create the variable weight_i^2 . Enter the command:

```
generate weightsq = weight^2
```

- Now estimate by OLS regression equation (7). Enter the command:

```
regress price weight mpg weightsq
```

Testing the marginal effects of individual explanatory variables

- **Test 3:** Test the hypothesis that the *marginal effect of weight_i on price_i is zero* for all cars in regression equation (7).
- ◆ The **marginal effect of weight_i on price_i** is obtained by partially differentiating regression equation (7) with respect to weight_i :

$$\frac{\partial \text{price}_i}{\partial \text{weight}_i} = \beta_1 + 2\beta_3 \text{weight}_i \quad (i = 1, \dots, N)$$

- ◆ Sufficient conditions for $\partial \text{price}_i / \partial \text{weight}_i = 0$ for all i are $\beta_1 = 0$ and $\beta_3 = 0$. We therefore want to test these two coefficient exclusion restrictions.
- ◆ The **null and alternative hypotheses** are:

$$H_0: \beta_j = 0 \quad \forall j=1,3 \quad \beta_1 = 0 \text{ and } \beta_3 = 0 \quad (8)$$

$$H_1: \beta_j \neq 0 \quad j=1,3 \quad \beta_1 \neq 0 \text{ and/or } \beta_3 \neq 0$$

- The following **test** command computes an F-test of H_0 against H_1 . Enter the command:

```
test weight weightsq
```

Inspect the results generated by this command. State the inference you would draw from this test.

Note: The above **test** command is equivalent to the following sequence of **test** commands with the **accumulate** option:

```
test weight = 0
test weightsq = 0, accum
```

- Use the following commands to compute the **critical values** of the **F[2,70] distribution** for significance levels $\alpha = 0.01, 0.05, 0.10$:

```
display invFtail(2,70, 0.01)    (returns value of  $F_{0.01}[2, 70]$ )
display invFtail(2,70, 0.05)    (returns value of  $F_{0.05}[2, 70]$ )
display invFtail(2,70, 0.10)    (returns value of  $F_{0.10}[2, 70]$ )
```

These are the critical values of the $F[2, 70]$ -distribution appropriate for testing hypothesis (8) above. Would you reject the null hypothesis in (8) at the 10 percent significance level, at the 5 percent significance level, and/or at the 1 percent significance level?

- **Test 4:** Test the hypothesis that the **marginal effect of $weight_i$ on $price_i$ is constant** in regression equation (7) – i.e., that it does not depend on $weight_i$.
- ◆ The necessary and sufficient condition for $\partial price_i / \partial weight_i$ in equation (7) to be constant is $\beta_3 = 0$. We therefore want to test this coefficient restriction.
- ◆ The **null and alternative hypotheses** are:

$$H_0: \beta_3 = 0 \tag{9}$$

$$H_1: \beta_3 \neq 0$$

- One way to test this hypothesis is to use the *t-ratio* for $\hat{\beta}_3$ that is printed by the **regress** command used to estimate equation (7). The t-ratio for $\hat{\beta}_3$ is given in general by the expression

$$t_0(\hat{\beta}_3) = \frac{\hat{\beta}_3}{\widehat{\text{se}}(\hat{\beta}_3)} \sim t[N - K] = t[74 - 4] = t[70] \quad \text{under } H_0: \beta_3 = 0.$$

Alternatively, you can calculate the above t-ratio for $\hat{\beta}_3$ and its two-tail p-value by entering the following **lincom** command:

```
lincom _b[weightsq]
```

From the results of OLS estimation of equation (7), state the inference you would draw from this t-test.

- A second way to test this hypothesis is to compute an *F-test* of $H_0: \beta_3 = 0$ against $H_1: \beta_3 \neq 0$. The sample value of the F-statistic for this test is given by the expression

$$F_0(\hat{\beta}_3) = \frac{\hat{\beta}_3^2}{\widehat{\text{Var}}(\hat{\beta}_3)} \sim F[1, N - K] = F[1, 74 - 4] = F[1, 70] \quad \text{under } H_0: \beta_3 = 0.$$

Enter either of the following **test** commands:

```
test weightsq
test weightsq = 0
```

State the inference you would draw from this F-test. Compare the results of this F-test with those of the t-test performed above. Are these two tests equivalent? Why or why not?

- **Test 5:** Test the hypothesis that the *marginal effect of mpg_i on price_i is zero* in regression equation (7), i.e., that price_i does not depend on mpg_i.
- ◆ The **marginal effect of mpg_i on price_i** is obtained by partially differentiating regression equation (7) with respect to mpg_i:

$$\frac{\partial \text{price}_i}{\partial \text{mpg}_i} = \beta_2 = \text{a constant} \quad (i = 1, \dots, N)$$

- ◆ The necessary and sufficient condition for $\partial \text{price}_i / \partial \text{mpg}_i$ in equation (7) to equal zero is $\beta_2 = 0$. We therefore want to test this coefficient restriction.
- ◆ The **null and alternative hypotheses** are:

$$H_0: \beta_2 = 0 \quad (10)$$

$$H_1: \beta_2 \neq 0$$

- One way to test hypothesis (10) is to use the **t-ratio** for $\hat{\beta}_2$ that is printed by the **regress** command used to estimate equation (7). The t-ratio for $\hat{\beta}_2$ is given in general by the expression

$$t_0(\hat{\beta}_2) = \frac{\hat{\beta}_2}{\text{s}\hat{\text{e}}(\hat{\beta}_2)} \sim t[N - K] = t[74 - 4] = t[70] \quad \text{under } H_0: \beta_2 = 0.$$

Alternatively, you can calculate the above t-ratio for $\hat{\beta}_2$ by entering the **lincom** command:

```
lincom _b[mpg]
```

From the results of OLS estimation of equation (7), state the inference you would draw from this t-test.

- A second way to test hypothesis (10) is to perform an F-test of the single coefficient restriction specified by the null hypothesis $H_0: \beta_2 = 0$ in (10). Enter either of the following **test** commands:

```
test mpg
test mpg = 0
```

Note: The sample value of the F-statistic computed by the above `test` command equals the square of the sample value of the t-statistic for the null hypothesis H_0 in (10):

$$F_0(\hat{\beta}_2) = \frac{(\hat{\beta}_2 - b_2)^2}{\text{Var}(\hat{\beta}_2)} = \frac{(\hat{\beta}_2 - 0)^2}{\text{Var}(\hat{\beta}_2)} = \frac{(\hat{\beta}_2)^2}{\text{Var}(\hat{\beta}_2)} = \left[\frac{\hat{\beta}_2}{\text{se}(\hat{\beta}_2)} \right]^2 = [t_0(\hat{\beta}_2)]^2.$$

- Use the following commands to verify these relationships between $t_0(\hat{\beta}_2)$ and $F_0(\hat{\beta}_2)$:

```
scalar t0b2 = _b[mpg]/_se[mpg]
scalar varb2 = _se[mpg]^2
scalar F0b2 = _b[mpg]^2/varb2
display t0b2
display sqrt(F0b2)
display t0b2^2
display F0b2
```

- Use the following commands to verify that the *two-tail p-value* of the *calculated t-statistic* $t_0(\hat{\beta}_2)$ equals the *p-value* of the *calculated F-statistic* $F_0(\hat{\beta}_2)$:

```
display 2*ttail(70, abs(t0b2))
display Ftail(1, 70, F0b2)
```

- Finally, calculate and compare the *two-tail 1% critical value* of the **t[70]**-distribution and the *1% critical value* of the **F[1, 70]**-distribution. Enter the **display** commands:

```
display invttail(70, 0.005)^2
display invFtail(1, 70, 0.01)
```

The results of these commands can be used to verify that the **α -level critical value of the F[1, 70]-distribution equals the square of the $\alpha/2$ critical value of the t[70] distribution**, i.e., that

$$F_{0.01}[1, 70] = (t_{0.005}[70])^2 \quad \text{or in general that} \quad F_{\alpha}[1, N - K] = (t_{\alpha/2}[N - K])^2.$$

Example 2: polynomials in two explanatory variables, $weight_i$ and mpg_i

Consider the following multiple regression equation:

$$price_i = \beta_0 + \beta_1 weight_i + \beta_2 mpg_i + \beta_3 weight_i^2 + \beta_4 mpg_i^2 + u_i \quad (11)$$

where

$$mpg_i^2 = \text{the square of } mpg_i.$$

Note: For PRE (11), the number of regression coefficients is $K = 5$.

Regression equation (11) contains **two explanatory variables**, $weight_i$ and mpg_i , but **four regressors** – $weight_i$, $weight_i^2$, mpg_i , and mpg_i^2 .

Marginal Effects of the Explanatory Variables

- ◆ The **marginal effect of $weight_i$** in equation (11) is:

$$\frac{\partial price_i}{\partial weight_i} = \beta_1 + 2\beta_3 weight_i = \text{a linear function of } weight_i.$$

- ◆ The **marginal effect of mpg_i** in equation (11) is:

$$\frac{\partial price_i}{\partial mpg_i} = \beta_2 + 2\beta_4 mpg_i = \text{a linear function of } mpg_i.$$

- Before estimating regression equation (11), it is necessary to create the variable mpg_i^2 . Enter the command:

```
gen mpgsq = mpg^2
```

- Now estimate by OLS regression equation (11). Enter the command:

```
regress price weight mpg weightsq mpgsq
```

- **Test 6:** Test the hypothesis that the *marginal effect of weight_i on price_i is zero* in regression equation (11).
- ◆ The **marginal effect of weight_i on price_i** is obtained by partially differentiating regression equation (11) with respect to **weight_i**:

$$\frac{\partial \text{price}_i}{\partial \text{weight}_i} = \beta_1 + 2\beta_3 \text{weight}_i$$

- ◆ Sufficient conditions for $\partial \text{price}_i / \partial \text{weight}_i = 0$ for all i are $\beta_1 = 0$ and $\beta_3 = 0$. We therefore want to test these two coefficient exclusion restrictions on regression equation (11).
- ◆ The **null and alternative hypotheses** are:

$$H_0: \beta_1 = 0 \text{ and } \beta_3 = 0$$

$$H_1: \beta_1 \neq 0 \text{ and/or } \beta_3 \neq 0.$$

- The following **test** command computes an F-test of H_0 against H_1 . Enter the command:

```
test weight weightsq
```

Inspect the results generated by this command. State the inference you would draw from this test.

- **Test 7:** Use a **lincom** command to compute the estimated value of the marginal effect of weight_i in equation (11) for a car that weighs 3,000 pounds, and to perform a two-tail t-test of the null hypothesis that the marginal effect of weight_i in equation (11) for a car that weighs 3,000 pounds equals zero. Recall that the expression for $\partial \text{price}_i / \partial \text{weight}_i$ in equation (11) is:

$$\frac{\partial \text{price}_i}{\partial \text{weight}_i} = \beta_1 + 2\beta_3 \text{weight}_i.$$

Evaluated at $\text{weight}_i = 3000$, the *estimated* marginal effect of weight_i on price_i in equation (11) is:

$$\text{estimate of } \frac{\partial \text{price}_i}{\partial \text{weight}_i} = \hat{\beta}_1 + 2\hat{\beta}_3(3000).$$

- ◆ The ***null and alternative hypotheses*** are:

$$H_0: \beta_1 + 2\beta_3(3000) = 0$$

$$H_1: \beta_1 + 2\beta_3(3000) \neq 0.$$

- Enter either of the following two **lincom** commands:

```
lincom _b[weight] + 2*_b[weightsq]*3000
```

```
lincom weight + 2*weightsq*3000
```

Note that in the second of the above **lincom** commands *variable names* are used to represent their corresponding *coefficient estimates*. But if you find this notation confusing, just use the first form of the above **lincom** commands.

Inspect the output generated by the above **lincom** commands. Would you infer that the marginal effect of weight_i on price_i is different from zero for a car that weighs 3,000 pounds?

- ***Test 8:*** Test the hypothesis that the ***marginal effect of mpg_i on price_i is zero*** in regression equation (11).
- ◆ The ***marginal effect of mpg_i on price_i*** is obtained by partially differentiating regression equation (11) with respect to ***mpg_i*** .

$$\frac{\partial \text{price}_i}{\partial \text{mpg}_i} = \beta_2 + 2\beta_4 \text{mpg}_i.$$

- ◆ Sufficient conditions for $\partial \text{price}_i / \partial \text{mpg}_i = 0$ for all i are $\beta_2 = 0$ and $\beta_4 = 0$. We therefore want to test these two coefficient exclusion restrictions.
- ◆ The ***null and alternative hypotheses*** are:

$$H_0: \beta_2 = 0 \text{ and } \beta_4 = 0$$

$$H_1: \beta_2 \neq 0 \text{ and/or } \beta_4 \neq 0.$$

- The following **test** command computes an F-test of H_0 against H_1 . Enter the command:

```
test mpg mpgsq
```

Inspect the results generated by this command. State the inference you would draw from this test.

- **Test 9:** Use a **lincom** command to compute the estimated value of the marginal effect of mpg_i in equation (11) for a car that gets 25 miles to the gallon, and to compute a two-tail t-test of the null hypothesis that the marginal effect of mpg_i in equation (11) for a car that gets 25 miles to the gallon equals zero. Recall that the expression for $\partial \text{price}_i / \partial \text{mpg}_i$ in equation (11) is:

$$\frac{\partial \text{price}_i}{\partial \text{mpg}_i} = \beta_2 + 2\beta_4 \text{mpg}_i.$$

Evaluated at $\text{mpg}_i = 25$, the *estimated* marginal effect of mpg_i on price_i is:

$$\text{estimate of } \frac{\partial \text{price}_i}{\partial \text{mpg}_i} = \hat{\beta}_2 + 2\hat{\beta}_4(25).$$

- ◆ The **null and alternative hypotheses** are:

$$H_0: \beta_2 + 2\beta_4(25) = 0$$

$$H_1: \beta_2 + 2\beta_4(25) \neq 0.$$

- Enter either (or both) of the following **lincom** commands:

```
lincom _b[mpg] + 2*_b[mpgsq]*25
```

```
lincom mpg + 2*mpgsq*25
```

What is the estimated value of the marginal effect of mpg_i on price_i for a car that has fuel efficiency equal to 25 miles per gallon? Would you infer that the marginal effect of mpg_i on price_i is different from zero for a car that has fuel efficiency equal to 25 miles per gallon?

□ **Interaction Variables: Interactions Between Two Continuous Variables**

Nature: Interactions between two *continuous variables* refer to products of explanatory variables. For example, if X_{ji} and X_{ki} are two continuous explanatory variables, the interaction term between them is the product $X_{ji}X_{ki}$. The product $X_{ji}X_{ki}$ is included as a regressor in the population regression function.

Usage: An interaction term between two continuous variables allows the marginal effect of one explanatory variable to be a linear function of the other explanatory variable.

Consider the following regression equation:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{1i}^2 + \beta_4 X_{2i}^2 + \beta_5 X_{1i} X_{2i} + u_i. \quad (12)$$

Compute the marginal effects on Y_i of the two explanatory variables X_{1i} and X_{2i} in equation (12).

- ◆ The **marginal effect of X_{1i}** in equation (12) is:

$$\frac{\partial Y_i}{\partial X_{1i}} = \beta_1 + 2\beta_3 X_{1i} + \beta_5 X_{2i} = \text{a linear function of both } X_{1i} \text{ and } X_{2i}.$$

- ◆ The **marginal effect of X_{2i}** in equation (12) is:

$$\frac{\partial Y_i}{\partial X_{2i}} = \beta_2 + 2\beta_4 X_{2i} + \beta_5 X_{1i} = \text{a linear function of both } X_{1i} \text{ and } X_{2i}.$$

Example 3: polynomials and an interaction term in two explanatory variables

$$\text{price}_i = \beta_0 + \beta_1 \text{weight}_i + \beta_2 \text{mpg}_i + \beta_3 \text{weight}_i^2 + \beta_4 \text{mpg}_i^2 + \beta_5 \text{weight}_i \text{mpg}_i + u_i \quad (13)$$

Note: For PRE (13), the number of regression coefficients is $K = 6$.

Regression equation (13) contains **two explanatory variables**, weight_i and mpg_i , but **five regressors** – weight_i , weight_i^2 , mpg_i , mpg_i^2 , and $\text{weight}_i \text{mpg}_i$.

Marginal Effects of the Explanatory Variables

- ◆ The **marginal effect of weight_i** in equation (13) is:

$$\begin{aligned} \frac{\partial \text{price}_i}{\partial \text{weight}_i} &= \beta_1 + 2\beta_3 \text{weight}_i + \beta_5 \text{mpg}_i \\ &= \text{a linear function of both } \text{weight}_i \text{ and } \text{mpg}_i. \end{aligned}$$

- ◆ The **marginal effect of mpg_i** in equation (13) is:

$$\begin{aligned} \frac{\partial \text{price}_i}{\partial \text{mpg}_i} &= \beta_2 + 2\beta_4 \text{mpg}_i + \beta_5 \text{weight}_i \\ &= \text{a linear function of both } \text{weight}_i \text{ and } \text{mpg}_i. \end{aligned}$$

- Before estimating regression equation (13), it is necessary to create the interaction term ($\text{weight}_i \text{mpg}_i$). Enter the command:

```
generate wgtmpg = weight*mpg
```

- Now estimate by OLS regression equation (13). Enter the command:

```
regress price weight weightsq mpg mpgsq wgtmpg
```

➤ **Test 10:** Test the hypothesis that the *marginal effect of weight_i on price_i is zero* in regression equation (13).

◆ The **marginal effect of weight_i on price_i** in equation (13) is

$$\frac{\partial \text{price}_i}{\partial \text{weight}_i} = \beta_1 + 2\beta_3 \text{weight}_i + \beta_5 \text{mpg}_i.$$

◆ Sufficient conditions for $\partial \text{price}_i / \partial \text{weight}_i = 0$ for all i are $\beta_1 = 0$ and $\beta_3 = 0$ and $\beta_5 = 0$. We therefore want to test these three coefficient exclusion restrictions.

◆ The **null and alternative hypotheses** are:

$$H_0: \beta_1 = 0 \text{ and } \beta_3 = 0 \text{ and } \beta_5 = 0 \quad (14)$$

$$H_1: \beta_1 \neq 0 \text{ and/or } \beta_3 \neq 0 \text{ and/or } \beta_5 \neq 0$$

• The following **test** command computes an F-test of H_0 against H_1 in (14). Enter the command:

```
test weight weightsq wgtmpg
```

Inspect the results generated by this command. State the inference you would draw from this test.

➤ **Test 11:** Use a **lincom** command to compute the estimated value of the marginal effect of weight_i in equation (13) for a car that weighs 3,000 pounds and gets 30 miles to the gallon, and to perform a two-tail t-test of the null hypothesis that this marginal effect of weight_i in equation (13) is equal to zero. Recall that the expression for $\partial \text{price}_i / \partial \text{weight}_i$ in equation (13) is:

$$\frac{\partial \text{price}_i}{\partial \text{weight}_i} = \beta_1 + 2\beta_3 \text{weight}_i + \beta_5 \text{mpg}_i$$

Evaluated at weight_i = 3000 and mpg_i = 30, the *estimated* marginal effect of weight_i on price_i is:

$$\text{estimate of } \frac{\partial \text{price}_i}{\partial \text{weight}_i} = \hat{\beta}_1 + 2\hat{\beta}_3(3000) + \hat{\beta}_5(30).$$

- ◆ The **null and alternative hypotheses** are:

$$H_0: \beta_1 + 2\beta_3(3000) + \beta_5(30) = 0$$

$$H_1: \beta_1 + 2\beta_3(3000) + \beta_5(30) \neq 0.$$

- Enter either (or both) of the following **lincom** commands:

```
lincom _b[weight] + 2*_b[weightsq]*3000 + _b[wgtmpg]*30
```

```
lincom weight + 2*weightsq*3000 + wgtmpg*30
```

Inspect the output generated by the above **lincom** commands. Would you infer that the marginal effect of weight_i on price_i is different from zero for a car that weighs 3,000 pounds and has fuel efficiency equal to 30 miles per gallon?

- **Test 12:** Test the hypothesis that the **marginal effect of mpg_i on price_i is zero** in regression equation (13).
- ◆ The **marginal effect of mpg_i on price_i** is obtained by partially differentiating regression equation (13) with respect to mpg_i .

$$\frac{\partial \text{price}_i}{\partial \text{mpg}_i} = \beta_2 + 2\beta_4 \text{mpg}_i + \beta_5 \text{weight}_i.$$

- ◆ Sufficient conditions for $\partial \text{price}_i / \partial \text{mpg}_i = 0$ for all i are $\beta_2 = 0$ and $\beta_4 = 0$ and $\beta_5 = 0$. We therefore want to test these three coefficient restrictions.
- ◆ The **null and alternative hypotheses** are:

$$H_0: \beta_2 = 0 \text{ and } \beta_4 = 0 \text{ and } \beta_5 = 0 \tag{15}$$

$$H_1: \beta_2 \neq 0 \text{ and/or } \beta_4 \neq 0 \text{ and/or } \beta_5 \neq 0.$$

- The following **test** command computes an F-test of H_0 against H_1 in (15). Enter the command:

```
test mpg mpgsq wgtmpg
```

Inspect the results generated by this command. State the inference you would draw from this test.

- **Test 13:** Use a **lincom** command to compute the estimated value of the marginal effect of mpg_i in equation (13) for a car that weighs 3,000 pounds and gets 30 miles to the gallon, and to perform a two-tail t-test of the null hypothesis that the marginal effect of mpg_i in equation (13) for a car that weighs 3,000 pounds and gets 30 miles to the gallon equals zero. Recall that the expression for $\partial \text{price}_i / \partial \text{mpg}_i$ in equation (13) is:

$$\frac{\partial \text{price}_i}{\partial \text{mpg}_i} = \beta_2 + 2\beta_4 \text{mpg}_i + \beta_5 \text{weight}_i.$$

Evaluated at $\text{weight}_i = 3000$ and $\text{mpg}_i = 30$, the *estimated* marginal effect of mpg_i on price_i is:

$$\text{estimate of } \frac{\partial \text{price}_i}{\partial \text{mpg}_i} = \hat{\beta}_2 + 2\hat{\beta}_4(30) + \hat{\beta}_5(3000).$$

- ◆ The **null and alternative hypotheses** are:

$$H_0: \beta_2 + 2\beta_4(30) + \beta_5(3000) = 0$$

$$H_1: \beta_2 + 2\beta_4(30) + \beta_5(3000) \neq 0.$$

- Enter either (or both) of the following **lincom** commands:

```
lincom _b[mpg] + 2*_b[mpgsq]*30 + _b[wgtmpg]*3000
```

```
lincom mpg + 2*mpgsq*30 + wgtmpg*3000
```

Would you infer that the marginal effect of mpg_i on price_i is different from zero for a car that weighs 3,000 pounds and has fuel efficiency equal to 30 miles per gallon? What is the estimated value of this marginal effect of mpg_i on price_i ?

□ Testing the Joint Significance of All Slope Coefficient Estimates -- Alternative Methods

A basic test of any estimated sample regression equation is a test of the *joint significance of all its slope coefficient estimates*. The reason for the importance of this particular hypothesis test is that addresses an important question about the empirical usefulness and relevance of the postulated model. That question is: Are the regressors in the regression function jointly relevant in explaining the variation of the dependent variable? Do all the regressors together actually help account for the variation in the dependent variable?

This section demonstrates the equivalence of three alternative methods for computing an *F-test* of the joint significance of all the slope coefficient estimates in a multiple linear regression model. Since a multiple regression model contains at least two non-constant regressors, the null hypothesis for such a joint significance test specifies two or more coefficient exclusion restrictions. An F-test is therefore appropriate.

The Model

To illustrate the alternative methods of performing an F-test of the joint significance of all slope coefficient estimates, use the multiple linear regression model for car prices given by equation (13):

$$\text{price}_i = \beta_0 + \beta_1 \text{weight}_i + \beta_2 \text{mpg}_i + \beta_3 \text{weight}_i^2 + \beta_4 \text{mpg}_i^2 + \beta_5 \text{weight}_i \text{mpg}_i + u_i \quad (13)$$

For regression model (13), the total number of regression coefficients is $K = 6$, and the number of slope coefficients is $k = K - 1 = 6 - 1 = 5$.

- Re-estimate model (13) by entering the following command:

```
regress price weight mpg weightsq mpgsq wgtmpg
```

- Alternatively, you can redisplay the results of the most recently issued **regress** command by simply entering the command:

regress

The Null and Alternative Hypotheses

$$H_0: \beta_j = 0 \text{ for all } j = 1, \dots, 5 \quad \Leftrightarrow \quad \beta_1 = 0 \text{ and } \beta_2 = 0 \dots \text{ and } \beta_5 = 0; \quad (16)$$

$$H_1: \beta_j \neq 0 \text{ for } j = 1, \dots, 5 \quad \Leftrightarrow \quad \beta_1 \neq 0 \text{ and/or } \beta_2 \neq 0 \dots \text{ and/or } \beta_5 \neq 0.$$

Method 1: the ANOVA F-statistic

Method 1 uses the **ANOVA F-statistic** derived from the Analysis-of-Variance table for the OLS sample regression equation obtained by OLS estimation of equation (13).

The **ANOVA F-statistic** can be computed using two alternative but equivalent formulas:

$$F_{\text{ANOVA}} = \frac{\text{ESS}/\text{df}_{\text{ESS}}}{\text{RSS}/\text{df}_{\text{RSS}}} = \frac{\text{ESS}/(K-1)}{\text{RSS}/(N-K)} = \frac{R^2/(K-1)}{(1-R^2)/(N-K)} \quad (17)$$

where:

ESS = the **Explained Sum of Squares** for the **unrestricted OLS-SRE**;

RSS = the **Residual Sum of Squares** for the **unrestricted OLS-SRE**;

$\text{df}_{\text{ESS}} = K - 1$ = the **degrees of freedom for ESS**;

$\text{df}_{\text{RSS}} = N - K$ = the **degrees of freedom for RSS**;

K = the **total number of regression coefficients** in the **unrestricted model**;

N = the **number of sample observations** in the estimation sample.

$$R^2 = \frac{\text{ESS}}{\text{TSS}} = 1 - \frac{\text{RSS}}{\text{TSS}} = \text{the } \mathbf{R\text{-squared}} \text{ for the } \mathbf{unrestricted OLS\text{-}SRE}.$$

The **null distribution of the ANOVA F-statistic** -- that is, the distribution of F_{ANOVA} under the null hypothesis that all $K-1$ slope coefficients jointly equal zero -- is the **$F[K-1, N-K] = F[\text{df}_{\text{ESS}}, \text{df}_{\text{RSS}}]$** distribution:

$$F_{\text{ANOVA}} \sim F[K-1, N-K] = F[\text{df}_{\text{ESS}}, \text{df}_{\text{RSS}}] \text{ under } H_0: \beta_j = 0 \quad \forall j = 1, \dots, k.$$

- Recall from *Stata Tutorial 3* that selected results of the **regress** command are temporarily stored in **e()** functions. To compute the ANOVA F-statistic, you will want to save the values of the terms that enter the formulas for F_{ANOVA} . To do this, enter the following **scalar** commands:

```
scalar N = e(N)
scalar ESS = e(mss)
scalar dfESS = e(df_m)
scalar RSS = e(rss)
scalar dfRSS = e(df_r)
scalar TSS = ESS + RSS
scalar dfTSS = N - 1
scalar list N ESS dfESS RSS dfRSS TSS dfTSS
```

- Calculate the *sample value of the ANOVA F-statistic* using the formula

$$F_{ANOVA1} = \frac{ESS/df_{ESS}}{RSS/df_{RSS}} = \frac{ESS/(K-1)}{RSS/(N-K)} \quad (17a)$$

Enter the commands:

```
scalar Fanova1 = (ESS/dfESS)/(RSS/dfRSS)
scalar list Fanova1 ESS RSS dfESS dfRSS
```

- Calculate and display the *p-value* for the ANOVA F-statistic F_{ANOVA1} . Enter the command:

```
display Ftail(dfESS, dfRSS, Fanova1)
```

- Calculate and display the *5% and 1% critical values* of the $F[K-1, N-K] = F[5, 68]$ distribution. Enter the commands:

```
display invFtail(dfESS, dfRSS, 0.05)
display invFtail(dfESS, dfRSS, 0.01)
```

- Now calculate the *sample value of the ANOVA F-statistic* using the alternative formula

$$F_{ANOVA2} = \frac{R^2/df_{ESS}}{(1-R^2)/df_{RSS}} = \frac{R^2/(K-1)}{(1-R^2)/(N-K)} \quad (17b)$$

You will first want to generate the value of R^2 , the R-squared for the estimated sample regression equation. Enter the commands:

```
scalar Rsq = ESS/TSS
scalar Fanova2 = (Rsq/dfESS)/((1-Rsq)/dfRSS)
scalar list Fanova2 Rsq dfESS dfRSS
```

- Calculate and display the *p-value* for the ANOVA F-statistic F_{ANOVA2} . Enter the command:

```
display Ftail(dfESS, dfRSS, Fanova2)
```

- Finally, compare the sample values of the two ANOVA F-statistics F_{ANOVA1} and F_{ANOVA2} . Enter the command:

```
scalar list Fanova1 Fanova2
```

As expected, $F_{ANOVA1} = F_{ANOVA2}$.

Method 2: the General Form F-statistic

A second method of computing an F-test of hypothesis (16) employs the following **general form of the F-statistic** (see *Note 17*):

$$F_0 = \frac{(RSS_0 - RSS_1)/(df_0 - df_1)}{RSS_1/df_1} \quad (18)$$

where:

RSS_0 = the residual sum of squares for the **restricted OLS-SRE**;

RSS_1 = the residual sum of squares for the **unrestricted OLS-SRE**;

$df_0 = N - K_0$ = the *degrees of freedom* for RSS_0 , the **restricted RSS**;

$df_1 = N - K$ = the *degrees of freedom* for RSS_1 , the **unrestricted RSS**;

K_0 = the **number of free regression coefficients** in the **restricted model**;

K = the **number of free regression coefficients** in the **unrestricted model**;

$df_0 - df_1 = N - K_0 - (N - K) = N - K_0 - N + K = K - K_0$.

This general form F-statistic involves the residual sum of squares for both the **unrestricted OLS sample regression equation** corresponding to the alternative hypothesis H_1 and the **restricted OLS sample regression equation** corresponding to the null hypothesis H_0 . To compute this general form F-statistic, it is therefore necessary to estimate by OLS both the **unrestricted model** implied the alternative hypothesis H_1 and the **restricted model** implied by the null hypothesis H_0 .

- The **unrestricted model** implied the **alternative hypothesis H_1** in (16) is simply regression equation (13):

$$\text{price}_i = \beta_0 + \beta_1 \text{weight}_i + \beta_2 \text{mpg}_i + \beta_3 \text{weight}_i^2 + \beta_4 \text{mpg}_i^2 + \beta_5 \text{weight}_i \text{mpg}_i + u_i. \quad (13)$$

- The **restricted model** implied the **null hypothesis H_0** in (16) is obtained by setting $\beta_1 = 0$ and $\beta_2 = 0$ and $\beta_3 = 0$ and $\beta_4 = 0$ and $\beta_5 = 0$ in the unrestricted regression equation (13):

$$\text{price}_i = \beta_0 + u_i. \quad (19)$$

- The following series of commands illustrates the computation of the general form F-statistic (18).

- Estimate by OLS the **unrestricted PRE (13)** and retrieve the unrestricted RSS (RSS_1) and its degrees-of-freedom (df_1). Enter the commands:

```
regress price weight mpg weightsq mpgsq wgtmpg
scalar RSS1 = e(rss)
scalar df1 = e(df_r)
scalar list RSS1 df1
```

- Save for later use the sample value of the F-statistic generated by the above **regress** command. Enter the commands:

```
scalar Fstateq10 = e(F)
scalar list Fstateq10 dfESS dfRSS
```

- Estimate by OLS the **restricted PRE (19)** and retrieve the restricted RSS (RSS_0) and its degrees-of-freedom (df_0). Enter the commands:

```
regress price
```

```
scalar RSS0 = e(rss)
scalar df0 = e(df_r)
scalar list RSS0 df0
```

- To see how to interpret the OLS estimate of the intercept coefficient β_0 in the restricted regression equation (19), compute the sample mean of the dependent variable price_i . Enter the command:

```
summarize price
```

Compare the results of this command with the *restricted OLS estimate* $\tilde{\beta}_0$ of the intercept coefficient β_0 in restricted regression equation (19): you will observe that $\tilde{\beta}_0 = \overline{\text{price}}$, the sample mean of the dependent variable price_i .

- Calculate and display the *sample value of the F-statistic*. Enter the commands:

```
scalar F0 = ((RSS0 - RSS1)/(df0 - df1))/(RSS1/df1)
scalar list F0 RSS0 RSS1 df0 df1
```

- Calculate the *p-value* of the calculated F-statistic F_0 . First create scalars representing the *numerator degrees of freedom* $df_{\text{num}} = df_0 - df_1 = K - K_0$ and the *denominator degrees of freedom* $df_{\text{den}} = df_1 = N - K$ for the null distribution of F_0 . Enter the commands:

```
scalar dfnum = df0 - df1
scalar dfden = df1
scalar list dfnum dfden
display Ftail(dfnum, dfden, F0)
```

- Use the following commands to compute the *critical values* of the **F[dfnum, dfden] = F[5, 68]-distribution** for significance levels $\alpha = 0.01, 0.05, 0.10$:

```
display invFtail(5, 68, 0.01)
display invFtail(5, 68, 0.05)
display invFtail(5, 68, 0.10)
```

- Finally, compare the sample values of the two ANOVA F-statistics F_{ANOVA1} and F_{ANOVA2} computed by Method 1 with the sample value of the F-statistic F_0 computed by Method 2 and the sample value of the F-statistic that is outputted by the **regress** command for the unrestricted equation (13). Enter the command:

```
scalar list Fanova1 Fanova2 F0 Fstateq10
```

As expected, these four F-statistics all have the same sample value. What inference would you draw from these F-statistics respecting the joint significance of the slope coefficient estimates in equation (13)?

Method 3: the test command

As you have learned, a single **test** command can be used to perform an F-test of the joint significance of all slope coefficient estimates in a linear regression model such as equation (13). This **test** command produces the same sample value of the F-statistic as do Methods 1 and 2 above, but it uses yet another form of the F-statistic. The following series of commands illustrates the equivalence of the **test** command and the other methods.

- Re-estimate regression equation (13). Enter again the following **regress** command:

```
regress price weight mpg weightsq mpgsq wgtmpg
```

- Now enter the following **test** command:

```
test weight mpg weightsq mpgsq wgtmpg
```

This **test** command performs a joint F-test of the null hypothesis H_0 in (16) that all five slope coefficients in equation (13) are jointly equal to zero against the alternative hypothesis H_1 that at least one of these slope coefficients is not equal to zero.

- Compare the sample value and p-value of the F-statistic computed by this **test** command with those of the F-statistics computed by Methods 1 and 2 above. Enter the commands:

```
scalar list Fanova1 Fanova2 F0 Fstateq10
display Ftail(dfnum, dfden, F0)
display Ftail(dfESS, dfRSS, Fanova1)
display Ftail(dfESS, dfRSS, Fstateq10)
```

□ Preparing to End Your *Stata* Session

Before you end your *Stata* session, you should do three things.

- First, you may want to **save the current dataset** (although you will not need it for future tutorials). Enter the following **save** command to save the current dataset as *Stata*-format dataset **auto7.dta**:

```
save auto7
```

- Second, **close the command log file** you have been recording. Enter the command:

```
cmdlog close
```

- Third, **close the log file** you have been recording. Enter the command:

```
log close
```

□ End Your *Stata* Session -- exit

- **To end your *Stata* session**, use the **exit** command. Enter the command:

```
exit      or      exit, clear
```

□ Cleaning Up and Clearing Out

After returning to Windows, you should copy all the files you have used and created during your *Stata* session to your own diskette. These files will be found in the *Stata working directory*, which is usually **C:\data** on the computers in Dunning 350, and **D:\courses** on the computers in MC B111. There are two files you will want to be sure you have: the *Stata* log file **351tutorial7.log**, and the command log file **351tutorial7.txt**. If you saved the *Stata*-format data set **auto7.dta**, you will probably want to take it with you as well. Use the Windows **copy** command to copy any files you want to keep to your own portable electronic storage device (e.g., flash memory stick) in the E:-drive (or to a diskette in the A:-drive).

Finally, **as a courtesy to other users** of the computing classroom, please delete all the files you have used or created from the *Stata* working directory.
