Chapter 11

Statistical Inference for Multiple Regression

Different samples will lead to different fitted regression lines, due to different random departures in the data from the population conditional mean. Statistical inferential methods control for this randomness.

This chapter presents the statistical properties of the least squares estimates and uses these properties to construct confidence intervals and perform hypothesis tests on the population parameters.

11.1   Properties of the Least Squares Estimator

The assumptions and their consequences for multiple regression are essentially the same as those for bivariate regression detailed in Chapter 7, except the population model now includes additional regressors and conditioning is on these extra regressors. A more abbreviated treatment is given here for multiple regression.

In this chapter we make the strong assumptions that model errors are independent and homoskedastic, in which case inference is based on default standard errors. These assumptions can be relaxed, in which case inference is based on robust standard errors that are presented in Chapter 12.1 and that lead to different t statistics, p-values and confidence intervals.

11.1.1   Data Assumptions

Throughout it is assumed that all the OLS coefficients can be computed. This requires that the sample size exceeds the number of regressors (including the intercept), that there is variation in the sample values taken by each regressor, and that there is no exact linear relationship between the regressors.

If these conditions are not met then regression output from most statistical packages will not report coefficients for all regressors. Instead the computer output will indicate that one or more regressors (or the intercept) are omitted, or will simply have no entry for one or more regressors.

11.1.2   Population Line or Population Conditional Mean

The population relationship between y and $x_2, ..., x_n$ is defined by $E[Y|X_2 = x_2, ..., X_k = x_k]$, the **population conditional mean** of Y given $X_2 = x_2, ..., X_k = x_k$. This is the probability-weighted average of all possible values of Y in the population given $X_2 = x_2, ..., X_k = x_k$. In the econometrics literature it is standard to write the population conditional mean more simply as $E[y|x_2, ..., x_k]$.

For linear regression it is assumed that **conditional mean** is a **linear** function of the regressors. Thus we write that the **conditional mean** or **conditional expectation function** or **population line** is

$$E[y|x_2, ..., x_k] = \beta_1 + \beta_2 x_2 + \cdots + \beta_k x_k.$$

The goal is to estimate the population parameters $\beta_1, \beta_2, ..., \beta_k$ as then we can make predictions, and we can estimate the effect on the conditional mean of changing one or more of the regressors.

Remark 87 *$E[y|x_2, ..., x_k]$ denotes the conditional mean of Y given $X_2 = x_2, ..., X_k = x_k$, the probability-weighted average of all possible values of Y in the population given $X_2 = x_2, ..., X_k = x_k$. For linear regression $E[y|x_2, ..., x_k] = \beta_1 + \beta_2 x_2 + \cdots + \beta_k x_k$.*

11.1.3   Population Assumptions

To obtain the statistical properties of the OLS coefficients we need to make assumptions about the population model and the sampling process that yielded the sample $(x_1, y_1), ..., (x_n, y_n)$.

We introduce the **error term** u and make the standard assumptions are that:

1. The **population model** is

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_k x_{ki} + u_i \text{ for all } i.$$

2. The **error for the ith observation has zero mean conditional on all regressors**:

$$E[u_i|x_{2i}, ..., x_{ki}] = 0 \text{ for all } i.$$

3. The **error for the ith observation has constant variance conditional on the regressors**:

$$\text{Var}[u_i|x_{2i}, ..., x_{ki}] = \sigma_u^2 \text{ for all } i.$$

4. The **errors for different observations are statistically independent**

$$u_i \text{ is independent of } u_j, \text{ for all } i \neq j.$$

Assumptions 1-2 are the crucial assumptions that ensure that the population relationship between y and the regressors is a linear relationship. Specifically the population mean is then

$$E[y_i|x_{2i}, ..., x_{ki}] = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_k x_{ki},$$

so that the data are generated by a linear population relationship. In that case it makes sense to fit the data with a regression line.

Assumptions 3 and 4 are additional assumptions that are used in determining the precision and distribution of the estimates b_1 and b_2. Assumption 3 implies that the error term is homoskedastic, so the conditional variance of y is the same across observations, with

$$\text{Var}[y_i|x_{2i}, ..., x_{ki}] = \sigma_u^2.$$

More detailed discussion of these assumptions was given in Chapter 6.3. See also Chapter 16.

11.1.4 Mean and Variance of a Least Squares Slope Coefficient

Let the typical regressor be the jth regressor x_j. Then inference on its coefficient β_j is based on the OLS estimate b_j. The following results are simply stated here. Proofs in the case of a single regressor are given in Appendix C.1; see also Chapter 16.2.

The **mean of the least squares estimate** b_j is

$$\text{E}[b_j] = \beta_j, \quad j = 1, ...k,$$

given assumptions 1-2. So b_j is unbiased for β_j. If many samples were available, yielding many estimates b_j, the average of the b_j equals β_j.

If we additionally make assumptions 3-4 then the **variance of the least squares slope coefficient** b_j can be shown to be

$$\text{Var}[b_j] = \sigma_{b_j}^2 = \frac{\sigma_u^2}{\sum_{i=1}^n \widetilde{x}_{ji}^2},$$

where $\widetilde{x}_{ji}$ denotes the residual from regressing x_{ji} on an intercept and all regressors other than x_{ji}.

Remark 88 *Under assumptions 1-2 the estimated slope coefficient b_j has mean equal to the population slope parameter β_j. Under assumptions 1-4 the estimated slope coefficient b_j has variance $\sigma_{b_j}^2 = \sigma_u^2 / \sum_{i=1}^n \widetilde{x}_{ji}^2$, where $\widetilde{x}_{ji}$ denotes the residual from regressing x_{ji} on an intercept and all regressors other than x_{ji}.*

The variable $\widetilde{x}_j$ is a residual that measures the net information that the regressor x_j adds beyond that obtained from the other regressors. Smaller values of $\widetilde{x}_j$ mean less net information is conveyed and $\sum_{i=1}^n \widetilde{x}_{ji}^2$ will be smaller. At the same time σ_u^2 will decrease as the other regressors improve model fit. So on balance $\text{Var}[b_j]$ may be bigger or smaller and b_j may be less or more precisely estimated as additional regressors are added to the model.

11.1.5 Standard Error of a Least Squares Slope Coefficient

The variance of b_j depends in part on σ_u^2, the variance of the error term, which is unknown. Under assumptions 1-4, an unbiased estimator of σ_u^2 is the **standard error of the regression**

$$s_e^2 = \frac{1}{n-k} \sum_{i=1}^n (y_i - \widehat{y}_i)^2.$$

The **standard error of** b_j, the estimated standard deviation of b_j, is obtained by replacing σ_u^2 in the formula for $\sigma_{b_j}^2$ by s_e^2 and then taking the square root. Then

$$se(b_j) = \frac{s_e}{\sqrt{\sum_{i=1}^n \widetilde{x}_{ji}{}^2}}.$$

11.1.6   When is a Slope Coefficient Precisely Estimated?

Given the formula for $se(b_2)$ the precision of estimation of β_2 is better the closer the data are to the true regression line (then s_e is small) and the larger is the sample size n (then there are more terms in the sum). This is the same as in bivariate regression.

Additionally $se(b_j)$ is smaller the **less** x_j is explained by the other regressors $x_2, ..., x_{j-1}, x_{j+1}, ..., x_k$. The simple intuition is that the coefficient of x_j will be more precisely estimated the more x_j adds an independent piece of information, rather than duplicating information already contained in the other regressors.

Algebraically, $se(b_j)$ is small if $|\widetilde{x}_{ji}|$ is large since then $\sum_{i=1}^n \widetilde{x}_{ji}{}^2$ is large. Since $\widetilde{x}_{ji} = x_{ji} - \widehat{x}_{ji}$ where $\widehat{x}_{ji}$ is the fitted value from regressing x_{ji} on an intercept and all regressors other than x_{ji}, $|\widetilde{x}_{ji}|$ is large when $\widehat{x}_{ji}$ does a poor job of predicting x_{ji}.

Remark 89 *The standard error of b_j is $se(b_j) = s_e/\sqrt{\sum_{i=1}^n \widetilde{x}_{ji}{}^2}$, where s_e is the standard error of the regression. Under assumptions 1-4 $se(b_j)^2$ has mean equal to the variance of b_j. Bigger samples are better – if the sample is m times larger then $se(b_j)$ is approximately $1/\sqrt{m}$ times as large. And wider dispersion of x_j, after controlling for the other regressors, leads to greater precision.*

11.1.7   The t Statistic

The t-statistic for multiple regression is the exact analog of that for bivariate regression, with the one change that the degrees of freedom are now $n - k$ rather than $n - 2$.

Under assumptions 1-4, $b_j \sim (\beta_j, \sigma_{b_j}^2)$, so the standardized statistic $Z = (b_j - \beta_j)/\sigma_{b_j} \sim (0, 1)$. By the central limit theorem, Z is standard normal distributed as $n \to \infty$. However, σ_{b_j} depends on the unknown parameter σ_u^2. Replacing σ_{b_j} by its estimate $se(b_j)$ leads to the t statistic.

Remark 90 *For OLS regression on k regressors including the intercept, under assumptions 1-4 the t statistic*

$$t = \frac{b_j - \beta_j}{se(b_j)},$$

is a realization of a randomly variable that is approximately $T(n - k)$ distributed, where $T(n - k)$ denotes the t distribution with $(n - k)$ degrees of freedom.

The $T(n - k)$ approximation is exact if additionally the errors are normally distributed or if $n \to \infty$.

11.1.8   How Large Should the Sample Be?

In general errors are not normally distributed and the $T(n-k)$ distribution is an approximation, one that gets more reasonable as the sample size gets large. Unfortunately, there is no simple way to determine whether the sample size for a particular data set and regression is adequate for the $T(n-k)$ approximation to be good, though more advanced analysis using simulations can provide some guidance. Thus there is no hard and fast rule for how large the sample size should be.

For bivariate regression with a single regressor many authors feel that there should be at least thirty observations, though this is by no means sufficient for all types of data. For example, suppose the dependent variable y takes only two values, such as a variable for whether or not an individual is employed. Then a larger sample size will be needed since it is very unlikely that the error term u has a continuous normal distribution when the dependent variable y is discrete and takes only two values.

For multiple regression, as more regressors are added the sample size should be even larger.

If the sample size is too small then the distribution of the t statistic generally has fatter tails than the $T(n-k)$. As a consequence, confidence intervals based on the $T(n-k)$ will be too narrow, and hypothesis tests will tend to reject the null hypothesis too often.

The house price example used in several chapters of this book has only 29 observations. This has the pedagogical advantage of, for example, making it easy to list the complete data. In practice, however, it would be better to have a larger sample.

Aside from considerations of approximating the distribution of the t statistic, another reason for not using samples that are too small is that estimation may be very imprecise, so that statistical analysis is too noisy to be useful.

11.2   Estimators of Model Parameters

The first goal of inference is to obtain estimates of the population parameters $\beta_1, ..., \beta_k$. This is called **point estimation**, to distinguish it from interval estimation using confidence intervals.

11.2.1   Optimal Properties of OLS Estimators

The desirable properties of OLS in bivariate regression, presented in Chapter 6.4, carry over to multiple regression.

First, the estimated OLS coefficients b_j are **unbiased** for the population slope parameters β_j, provided the population model satisfies assumptions 1 and 2 since then $\mathrm{E}[b_j] = \beta_j$.

Second, the estimated OLS coefficients b_j are **consistent** for the population slope parameters β_j, if additionally $\mathrm{Var}[b_j]$ goes to zero as $n \to \infty$. This is the case under the additional assumptions 3 and 4, and under alternative assumptions such as those given in Chapter 7.7.

Third, under assumptions 1-4, the OLS estimates b_j are **best linear unbiased (BLUE)**. They have smallest variance among unbiased estimators that are a weighted average of y_i of the form $b_j = \sum_{i=1}^{n} w_i y_i$ where the weights will depend on the regressors. This result is called the **Gauss-Markov Theorem**.

Fourth, if additionally the errors are normally distributed then the OLS estimates b_j are **best unbiased**, meaning the have minimum variance among unbiased estimators. OLS is also the best estimator among those estimators that are consistent and asymptotically normal, even if the errors are not normally distributed.

Remark 91 *The OLS estimators $b_1, ..., b_k$ are unbiased for $\beta_1, ..., \beta_k$ under assumptions 1-2, best linear unbiased under assumptions 1-4, and best unbiased if additionally the more errors are normally distributed. In large samples the OLS estimators $b_1, ..., b_k$ are consistent for $\beta_1, ..., \beta_k$ under assumptions 1-2, plus assumptions(s) such as assumptions 3-4 that ensure the estimators have finite variance, and are best among consistent and asymptotically normal estimators under assumptions 1-4.*

11.2.2   Relaxing Assumptions 1-4

The slope parameter estimates b_j will usually be **biased** and **inconsistent** if assumptions 1 and/or 2 are not satisfied. In that case alternative analysis is needed.

Often interest lies in the coefficient of a single regressor, say the coefficient of years of schooling in an earnings-schooling regression that includes additional control variables. Alternative analysis that may lead to consistent estimation of the key parameter of interest may entail using a different functional form for the conditional mean, including additional control variables in the regression, using data from a different source, and using an estimation method other than OLS (see Chapter 17).

Assumptions 3-4 generally effect only the estimation of the precision of b_j. These assumptions can be relaxed. First, one can continue to estimate by OLS but use an alternative formula to compute the standard errors of the OLS coefficient estimates; see Chapters 7.7 and 12.1. Second, one can use alternative estimation methods that may lead to coefficient estimates that are more precise than the OLS estimates; see Chapter 12.4 for further discussion.

11.3   Confidence Intervals

11.3.1   Confidence Intervals

As usual, the confidence interval is the estimate plus or minus the relevant critical value times the standard error of the estimate.

Remark 92 *A $100(1 - \alpha)$ **percent confidence interval for the slope parameter** β_j is*

$$b_j \pm t_{n-k,\alpha/2} \times se(b_j),$$

where b_j is the OLS slope coefficient, $t_{n-k,\alpha/2}$ is that value such that a $T(n-k)$ distributed random variable exceeds it in absolute value with probability α, and $se(b_j)$ is the standard error of the slope coefficient estimate b_j.

This confidence interval is exact only if assumptions 1-4 hold and the errors are normally distributed.

The interpretation of the confidence interval is the same as that detailed in Chapter 7.3 for bivariate regression. A 95% confidence interval for β_j is interpreted as an interval that has probability 0.95 of including β_j. That is, if we had many repeated samples leading to many different confidence intervals, 95% of these confidence intervals will include the true unknown value of β_j and 5% will not.

11.3.2 Example: House Price and Size

As an example, consider the house price data in dataset HOUSE. Output from regression of house price on an intercept and six other regressors was presented in Table 10.4. For regressor *Size*, the output lists the 95 percent confidence interval to be $(36.45, 100.29)$.

This confidence interval can be computed from first principles as follows. The output gives $b_{Size} = 68.37$, from the coefficient column, and $se(b_{Size}) = 15.39$, from the standard error column. And here $t_{n-k,\alpha/2} = t_{22;.025} = 2.074$. Then a 95% confidence interval for β_{Size} is

$$b_{Size} \pm t_{n-k,\alpha/2} \times se(b_{Size}) = 68.37 \pm 2.074 \times 15.39 = 68.37 \pm 31.92 = (36.45, 100.29).$$

11.4 Hypothesis Tests on a Single Parameter

The presentation here is brief. Chapter 12.6 provides a more detailed discussion of the issues involved in hypothesis testing.

11.4.1 Tests on Individual Parameters

Consider tests on the jth regression parameter β_j. A **two-sided test** or **two-tailed test** on the parameter β_j is a test of $H_0 : \beta_j = \beta_j^*$ against $H_a : \beta_j \neq \beta_j^*$, where β_j^* is a specified value for β_j.

The null hypothesis is rejected when b_j is far from β_j^*. Equivalently the null hypothesis is rejected if the t statistic $t = (b_j - \beta_j^*)/se(b_j)$ is large. Under the null hypothesis that $\beta_j = \beta_j^*$ and under assumptions 1-4, the t statistic is approximately a draw from the $T(n-k)$ distribution (and exactly a draw from the $T(n-k)$ distribution if additionally model errors are normally distributed). The null hypothesis is rejected if the observed value $|t|$ is so large that it is very unlikely to have observed this value, if indeed H_0 was true so that the t statistic was indeed a draw from the $T(n-k)$.

Similar to the bivariate case, this leads to the following.

Remark 93 *For a **two-sided test** of $H_0 : \beta_j = \beta_j^*$ against $H_a : \beta_j \neq \beta_j^*$ the t statistic*

$$t = \frac{b_j - \beta_j^*}{se(b_j)}$$

*is a draw from the $T(n - k)$ distribution, approximately, if H_0 is true. The p-**value** is $p = \Pr[|T_{n-k}| \geq |t|\ ]$ and H_0 is rejected if $p < \alpha$, where α is the desired **significance level** of the*

Table 11.1: Hypothesis tests on slope parameter: Summary for multiple regression.

	Two-sided Test	One-sided Upper alternative	One-sided Lower alternative
Null hypothesis	$H_0 : \beta_j = \beta_j^*$	$H_0 : \beta_j \leq \beta_j^*$	$H_0 : \beta_j \geq \beta_j^*$
Alternative Hypothesis	$H_0 : \beta_j \neq \beta_j^*$	$H_0 : \beta_j > \beta_j^*$	$H_0 : \beta_j < \beta_j^*$
t statistic	$t = (b_j - \beta_j^*)/se(b_j)$	$t = (b_j - \beta_j^*)/se(b_j)$	$t = (b_j - \beta_j^*)/se(b_j)$
p-value	$p = \Pr[|T_{n-k}| \geq t]$	$p = \Pr[T_{n-k} \geq t]$	$p = \Pr[T_{n-k} \leq t]$
Rejection rule	$p < \alpha$	$p < \alpha$	$p < \alpha$
Critical value	$c = t_{n-k,\alpha/2}$	$c = t_{n-k,\alpha}$	$c = -t_{n-k,\alpha}$
Rejection region	$|t| > c$	$t > c$	$t < c$

test. The **critical value** c is such that $c = t_{n-k,\alpha/2}$, equivalently $\Pr[|T_{n-k}| \geq c] = \alpha$, and H_0 is rejected at significance level α if $|t| > c$.

It is also possible to perform a **one-sided** or **one-tailed test.** In that case the claim to be tested is set as the alternative hypothesis. Table 11.1 summarizes one-sided and two-sided hypothesis tests on the population parameter slope coefficient.

11.4.2 Relationship between t test and Adjusted R-Squared

It can be shown that the adjusted R-squared, $\bar{R}^2$, increases as one regressor is added to the model if and only if a test for statistical significance of this regressor, one based on default standard errors, yields $|t| > 1$.

By contrast a hypothesis test at the usual significance level of 0.05 has a much higher threshold such as $|t| > 1.96$ for a large sample size. It follows that an increase in $\bar{R}^2$ does not necessarily imply statistical significance of the additional regressor(s) at conventional levels of significance such as $\alpha = 0.05$. While $\bar{R}^2$ provides a penalty for large model size, the penalty is weak.

11.4.3 Tests of Individual Statistical Significance

A special case of tests on an individual parameter is a **test of statistical significance**, in which case the hypothesized value of β_j is $\beta_j^* = 0$.

Regression packages print out the necessary statistics to automatically test this hypothesis, including both the t statistic and the p-value for the test of $H_0 : \beta_j = 0$ against $H_a : \beta_j \neq 0$. An example is given in Table 10.4. H_0 is rejected at statistical significance level α if $p < \alpha$. If instead $p > \alpha$ we do not reject H_0, and conclude that there is no statistically significant relationship or equivalently, that the regressor is statistically insignificant. The most common choice of α is 0.05, followed by 0.01 and 0.10.

It is most common to perform two-sided tests of statistical significance, so computer output reports the p-value for a two-sided test. If we have prior beliefs about the sign of β_j, however, then this prior belief can be employed by performing a one-sided test. The prior belief is setup as the alternative hypothesis, as explained in Chapter 4.6. For example, since we believe that house

price increases with size, a one-tailed test of significance of house size may be a test of $H_0 : \beta_j \leq 0$ against $H_a : \beta_j > 0$. In that case we halve the printed p-value, provided the estimate b_j is positive.

There can be ambiguity in the statement that "the regressor is statistically significant at significance level 0.05", as it is not always made clear whether a one-sided or two-sided test was performed. Most often tests of statistical significance are performed as two-sided tests, even if there are strong prior beliefs on the direction of the relationship.

Statistical significance by itself does not imply automatically that the regressor has **economic significance**. As discussed in Chapter 7.4.3, an economically significant regressor has coefficient b_j that is large enough that changing x_j is associated with meaningful changes in the dependent variable. Statistical significance is determined by the size of the t statistic which is the coefficient b_j scaled by $se(b_j)$. Economic significance is instead determined only by whether the coefficient b_j is "small" or "large". This is a subjective judgement that is context specific.

11.4.4 Example: House Prices

As an example we consider regression of house sale price on several regressors using dataset HOUSE. The example is illustrative. We use default standard errors that presume errors are homoskedastic. For these data it is standard to instead use heteroskedastic-robust standard errors; these are presented in Table 12.2. And a better way to incorporate the variable *Lotsize*, one that takes the three arbitrary values 1, 2 and 3, is to use a set of indicator variables, presented in Chapter 14.4.

Results are given in Table 11.2, which repeats Table 10.4.

The only regressor that is statistically significant at significance level 0.05 is *Size* with $p = 0.000$. The next most statistically significant regressor is *Age* with a p-value of 0.259 that is much higher than 0.05.

Suppose that instead of a test of statistical significance we wish to test whether an increase in house size of one square foot is associated with a \$50 increase in house price. For test of $H_0 : \beta_{Size} = 50$ against $H_0 : \beta_{Size} \neq 50$, $t = (68.37 - 50)/15.39 = 1.194$. Then $p = \Pr[|T_{22}| > |1.194|] = 0.245$ so we do not reject H_0 at significance level 0.05.

Looking at the magnitude of the coefficients, the estimate $b_{Size} = 68.37$ means that a 100 square foot increase in size, equivalent to a small room that is ten feet by ten feet, is associated with a \$6,837 increase in house price, an economically meaningful effect. An additional bathroom is also associated with a substantial increase in house price, but this estimate is so imprecise (the 95% confidence interval is from −\$25,771 to \$39,437) that no attention should be paid to it.

We conclude that house size is the primary determinant of house price in this market.

One reason for the lack of importance of the other regressors may be that, aside from size, the houses are quite homogeneous as they are of similar vintage and are in a small part of a small city. A real estate maxim is "location, location, location", and here location has already been controlled for by choosing to focus on a fairly homogenous region.

A second reason for statistical insignificance of the other regressors may be the small sample size.

A third reason may be that in fact collectively the house attributes other than size may matter, even if individually each attribute is not statistically significant. This can be tested using the F test presented below.

Table 11.2: House price: Computer output from multiple regression.

Summary Statistics

Observations	29
R-squared	0.651
Adjusted R^2	0.555
St. error of regression	24936

Dependent Variable *Price*

Regressor	Coefficient	Standard Error	t statistic	p value	95% conf. int.	
Size	68.37	15.39	4.44	0.000	36.45	101.29
Bedrooms	2685	9193	0.29	0.773	-16379	21749
Bathrooms	6833	15721	0.43	0.668	-25771	39437
Lot Size	2303	7227	0.32	0.753	-12684	17290
Age	-833	719	-1.16	0.259	-2325	659
Monthsold	-2089	3521	-0.59	0.559	-9390	5213
Intercept	137791	61465	2.24	0.035	10321	265261

ANOVA Table

Source	SS	df	MS	F	p-value
Explained	2.5466×10^{10}	6	4.2444×10^9	6.83	0.0003
Residual	1.3679×10^{10}	22	0.6218×10^9		
Total	3.9146×10^{10}	28	0.1398×10^9		

11.4.5   Tests of a Single Hypothesis on more than One Parameter

The t test can be extended to test a single hypothesis that involves more than one parameter. The following example can be adapted to other cases of a single hypothesis.

Suppose we want to test whether or not $\beta_2 = \beta_3$. Equivalently we want to test $H_0 : \beta_2 - \beta_3 = 0$ against $H_0 : \beta_2 - \beta_3 \neq 0$. This difference can be estimated by $b_2 - b_3$ which has standard error $se(b_2 - b_3)$, leading to the t statistic

$$t = \frac{b_2 - b_3}{se(b_2 - b_3)}.$$

We reject H_0 at level 0.05 if $p = \Pr[|T_{n-k}| < |t|] < 0.05$.

This test requires calculating $se(b_2 - b_3) = \sqrt{se^2(b_2) - 2c_{23} + se(b_3)^2}$, where c_{23} is the estimated covariance between b_2 and b_3. But c_{23} is not automatically provided in standard regression output.

Fortunately, many statistical software packages provide a command to implement such tests, yielding the t statistic (or in some cases the square of the t statistic) and the associated p-value.

If this is not the case then the test can be implemented as follows. Rewrite the model as

$$
\begin{aligned}
y &= \beta_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \cdots + \beta_k x_k + u \\
&= \beta_1 + (\beta_2 - \beta_3)x_2 + \beta_3(x_2 + x_3) + \beta_4 x_4 + \cdots + \beta_k x_k + u.
\end{aligned}
$$

Then regress y on an intercept, x_2, $(x_2 + x_3)$, $x_4, ..., x_k$ and perform a regular t test of whether the coefficient of x_2 equals zero.

As an example, suppose we want to test whether the coefficient of *Bedrooms* equals that of *Bathrooms*. We create a variable *Bedplusbath* equal to the sum of *Bedrooms* and *Bathrooms* and regress *Price* on *Size, Bedrooms, Bedplusbath, Lotsize, Age* and *Monthsold*. The regressor *Bedrooms* has coefficient -4148, standard error 17806, $t = -0.233$ and $p = 0.818$. We do not reject at level 0.05 the null hypothesis that $\beta_{Bedrooms} = \beta_{Bathrooms}$.

11.5 Joint Hypothesis Tests

For multiple regression we often wish to test more than one restriction on the parameters. For example we may wish to test whether or not both $\beta_2 = 0$ and $\beta_3 = 0$. A more complicated example is a test that both $\beta_2 = -\beta_3$ and $2\beta_4 + \beta_6 = 9$.

There are two reasons for performing such tests. One reason is to determine which regressors should be included in the regression model. Models with fewer regressors are often preferred, for reasons of simplicity and because more precise estimates can be obtained if unnecessary regressors are excluded from the model. A second reason is to test restrictions implied by economics theory.

These joint hypothesis tests use test statistics that are most often F distributed, rather than the t distributed. (A variation using the chi-squared distribution is presented later in this section). The test statistics take only positive values so rejection of the null hypothesis occurs if the test statistic takes large positive values.

For joint hypotheses the test statistics are in general a complicated function of coefficient estimates and their associated standard errors and estimated covariances. Many statistical packages provide a command to implement joint tests, and all statistical packages provide as output a joint test of whether or not the regressors are jointly statistically significant, a test of $H_0 : \beta_2 = 0$, $\beta_3 = 0, ..., \beta_k = 0$.

In the special case of independent, homoskedastic errors (assumptions 1-4) it is possible to calculate F test statistics using residual sums of squares. Such calculation does not generalize to inference based on robust standard errors, such as heteroskedastic robust. But it provides useful insights and it is presented in Chapter 11.6.

11.5.1 The F Distribution

The F**distribution** is a continuous right-skewed distribution for a random variable that takes only positive values. The distribution depends on two parameters, called the first and second degrees of freedom, and is denoted $F(v_1, v_2)$ where the degrees of freedom v_1 and v_2 are positive integers.

Note that the order of the degrees of freedom matters, as $F(v_1, v_2) \neq F(v_2, v_1)$. The mean exists if $v_2 > 2$ and equals $v_2/(v_2 - 2)$, so the mean approaches one as v_2 gets large which is the case in regression samples when the sample size gets large. When $v_1 = 1$ the F distribution reduces to the square of the t distribution, i.e. $F_{1,v_2} = (T_{v_2})^2$.

In regression applications v_1 is the number the number of parameter restrictions being tested, so is small, and $v_2 = n - k$, the regression degrees of freedom.

Figure 11.1 presents the probability density function of the F distribution for $(v_1, v_2) = (3, 30)$ and $(v_1, v_2) = (10, 30)$.
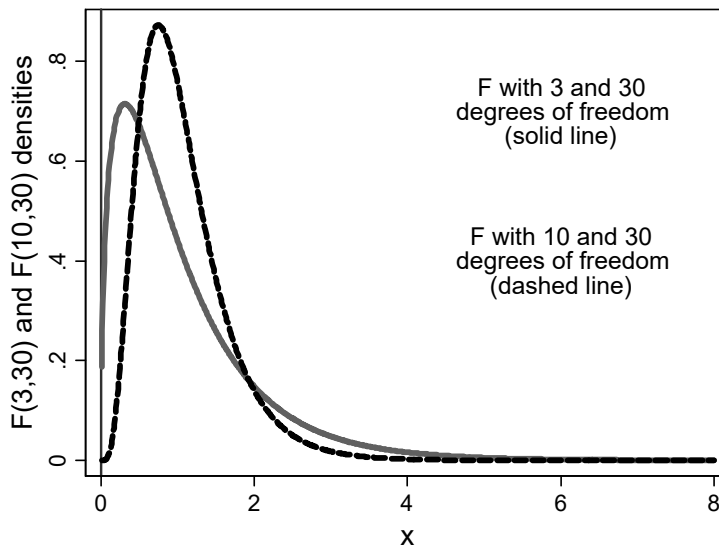
Figure 11.1: F distribution: F(3,30) and F(10,30)

Remark 94 *The $F(v_1, v_2)$ distribution is a right-skewed distribution, with degrees of freedom v_1 and v_2, for a random variable that takes only positive values. $F_{v1,v2,\alpha}$ denotes the value for which the area in the right tail of the distribution equals α.*

11.5.2   Probabilities and Inverse Probabilities for the F Distribution

Probabilities for the F distribution are given by the area under the curve and require use of an appropriate computer command. For example, in Stata $\Pr[F_{10,30} > 2] = $ `Ftail(10,30,2)`.

   Inverse probabilities or **critical values** for the $F(q, n - k)$ distribution can be obtained form an appropriate computer command. For example, in Stata $F_{10,30,.05} = $ `invFtail(10,30,.05)`. Alternatively, tables give these at different degrees of freedom and at key significance levels such as 0.10, 0.05 and 0.01. The critical values decrease as the number of restrictions (q) increases, the degrees of freedom in the unrestricted model ($n - k$) increases, and as test significance level (α) increases. Table 11.3 provides some examples and Appendix E provides more detailed tables.

   Unlike t critical values, F critical values are difficult to interpret directly and it is much easier to rely on p-values computed by a statistical package. If these are unavailable note that if $n - k > 10$ then H_0 is always rejected at level 0.05 if $F > 5$, regardless of the size of q.

   For tests of a single restriction we use the $F(1, n - k)$ distribution which is the square of the $T(n - k)$ distribution. Since the 5% critical value for the $T(\infty)$ distribution is 1.96, it follows that the 5% critical value for the $F(1, \infty)$ distribution is $1.96^2 = 3.84$.

Table 11.3: F distribution: Critical values for various degrees of freedom and confidence levels.

Test size		$q=1$	$q=2$	$q=3$	$q=10$	$q=20$
10%	$n-k=30$	2.88	2.49	2.28	1.82	1.67
	$n-k=\infty$	2.71	2.30	2.08	1.60	1.42
5%	$n-k=30$	4.17	3.32	2.92	2.16	1.93
	$n-k=\infty$	3.84	3.00	2.60	1.83	1.57
1%	$n-k=30$	7.56	5.39	4.51	2.98	2.55
	$n-k=\infty$	6.64	4.61	3.78	2.32	1.88

11.5.3   The F Statistic

The F statistic applies to models that are nested in each other.

The more general model, called the **unrestricted model** or **complete model**, is a model with k regressors, so

$$y = \beta_1 + \beta_2 x + \beta_3 x_3 + \cdots + \beta_k x_k + u.$$

The **restricted model** or **reduced model** is a model that places restrictions on the parameters $\beta_1, \beta_2, ..., \beta_k$. Most often this restricted model is one that omits some of the regressors, so that some of the $\beta's$ are set to zero.

Let q denote the number of restrictions imposed by the restricted models. If just one regressor is dropped then $q = 1$ while if all regressors but the intercept are dropped then $q = k - 1$.

In general the formula for the F statistic is complicated. As an example, suppose we wish to test the two restrictions $\beta_2 = 0$ and $\beta_3 = 0$. So we test $H_0 : \beta_2 = 0$, $\beta_3 = 0$ against H_a : at least one of $\beta_2 \neq 0$, $\beta_3 \neq 0$. Then it can be shown that

$$F = \frac{[b_2 se(b_2)^2 - 2b_2 b_3 c_{23} + b_2 se(b_2)^2]/2}{se(b_2)^2 se(b_3)^2 - c_{23}^2}$$

where c_{23} is the estimated covariance between b_2 and b_3.

For the moment we assume that computer output provides the desired F statistic. In the subsequent section we present formulas in the special case that errors are independent and heteroskedastic so that assumptions 1-4 hold and we can use default standard errors.

11.5.4 F Tests

The main use of the F statistic is in hypothesis testing.

Definition 27 *An F **test** is a **two-sided test of** H_0 : The q parameter restrictions implied by the restricted model are correct, against the alternative hypothesis H_a : At least one of the q parameter restrictions implied by the restricted model is incorrect.*

The F statistic is necessarily positive and large values of the F statistic lead to rejection of H_0.

Remark 95 *The p-**value** for the F test of q restrictions is $p = \Pr[F_{q,n-k} \geq F]$, and H_0 is rejected if $p < \alpha$, where α is the desired **significance level** of the test. The **critical value** c is such that $c = F_{q,n-k,\alpha}$, equivalently $\Pr[|F_{q,n-k}| \geq c] = \alpha$, and H_0 is rejected at significance level α if $F > c$.*

Note that the F test is a **two-sided test**. It is true that the F test p-value or critical value is calculated using only **one tail**, the right tail, of the F distribution. But the test itself is a two-sided test because in the case of testing exclusion of several regressors, for example, the alternative hypothesis can only be that parameters are not equal to zero rather than, for example, greater than zero. A one-sided F test is not possible.

The leading examples of the F test are now presented.

11.5.5   Test of Overall Significance

The first test, called as **test of overall significance**, is a test of whether or not the regressors taken together add substantially to predictive ability compared to an intercept-only model. The null hypothesis is that only an intercept is needed, so all the slope parameters equal zero, while the alternative hypothesis is that at least one of the slope coefficients differs from zero.

Remark 96 *A **test of overall significance** is a **two-sided test of** $H_0 : \beta_2 = 0, ..., \beta_k = 0$, against the alternative hypothesis $H_a :$ At least one of $\beta_2 \neq 0, ..., \beta_k \neq 0$.*

Most statistical packages automatically print out the F statistic for test of overall significance, along with its associated p-value. If $p < 0.05$, for example, then the regressors are said to be jointly statistically significant at significance level 0.05.

For the house price example $k = 7$ and $n = 29$, so a test if overall significance tests $q = 6$ restrictions and $n - k = 22$. The output in Table 11.2 includes $F(6, 22) = 6.83$ with $p = 0.0003$. Since $p < 0.05$ we conclude that the regressors are jointly statistically significant at level 0.05.

Note that the F test is merely a test of whether at least one regressor is statistically significant, not that all regressors are statistically significant. In this example from Table 11.2 only one of the regressors, *Size*, is individually statistically significant at level 0.05.

11.5.6   Test of Subsets of Regressors

Tests of a subset of regressors or of **exclusion restrictions** compare two models with one model a reduced version of the other and test whether the additional regressors in the unrestricted model are jointly statistically significant. This test is often used to test the joint statistical significance of sets of indicator variables, such as for geographic region or for ethnic background, that are presented in Chapter 14.4.

The unrestricted model or complete model with k regressors is specified to be

$$y = \beta_1 + \beta_2 x_2 + \cdots \beta_g x_g + \beta_{g+1} x_{g+1} + \cdots + \beta_k x_k + u.$$

The restricted model or reduced model is specified to include only the first g regressors so

$$y = \beta_1 + \beta_2 x_2 + \cdots \beta_g x_g + u.$$

The restricted model is therefore obtained by setting $k - g$ slope coefficients to zero.

Remark 97 *Let the last $q = k - g$ regressors be omitted in the restricted model. Then a **test of a subset of regressors** is a **two-sided test of** $H_0 : \beta_{g+1} = 0, ..., \beta_k = 0$, against the alternative hypothesis H_a : At least one of $\beta_{g+1} \neq 0, ..., \beta_k \neq 0$.*

Return to the house price example. We have shown that house size is the only one of the six regressors that is individually statistically insignificant. A natural question to ask is whether the five regressors when taken together are statistically significant, or whether a model with just house size as a regressor is adequate. This is a test of five restrictions. Using a specialized test command yields $F = 0.417$ with $p = 0.832 > 0.05$ so we do not reject $H_0 : \beta_3 = 0, ..., \beta_7 = 0$ at significance level 0.05. We conclude that the additional five regressors are jointly statistically insignificant. It is best to just include *Size* as a regressor.

Statistical packages that provide a post-estimation test command that computes this F test may additionally enable test of more complicated hypotheses, such as a test of $H_0 : \beta_2 + \beta_3 = 1$, $\beta_4 = 2$ against H_0 : At least one of $\beta_2 + \beta_3 \neq 1$, $\beta_4 \neq 2$.

11.5.7 Test of a Single Regressor

The F test for subsets of regressors can be used to test whether a single regressor is statistically significant, in which case $q = 1$. This F test yields exactly the same result as that from a two-sided t test of statistical significance.

First, it can be shown algebraically that for $q = 1$ the F test statistic is the square of the usual t test statistic, so $F = t^2$. Second, a random variable with $T(n - k)$ distribution when squared is $F(1, n - k)$ distributed.

It follows that when $q = 1$, the F-test critical value equals the square of the critical value for a two-sided t test. So a two-sided t-test critical value of 1.96 corresponds to an F-test critical value of $1.96^2 = 3.84$.

Remark 98 *The F test for statistical significance of a single regressor is equivalent to a two-sided t test, since $F = t^2$ and $\Pr[|T_{n-k}| > t] = \Pr[F_{1,n-k} > t^2]$.*

For example, using a specialized test command to test whether *Bedrooms* is statistically significant yields $F = 0.085$ with $p = \Pr[F_{1,n-k} > 0.085] = 0.773$. From Table 11.2 the regressor *Bedrooms* has $t = 0.292$, so $t^2 = 0.292^2 = 0.085 = F$, and the same $p = 0.773$.

11.5.8 Computation of Tests in a Statistical Package

OLS regression output includes tests of statistical significance for each regressor and the test of overall significance. For more complicated tests such as tests of subsets of regressors in Stata one uses the `test` and `testparm` commands; see Appendix A.2 for an example. In R one can use the `linearHypothesis` function in the `car` package; see Appendix A.3 for an example. In Gretl use the `restrict` command with option `wald`. In Eviews use the `wald` view.

11.5.9 Chi-squared Distribution

The t distribution goes to the standard normal distribution as the degrees of freedom go to infinity, that is the $T(\infty)$ distribution is the $N(0,1)$ distribution.

A similar limit result exists for the F distribution. A random variable that has an $F(q,\infty)$ distribution when multiplied by q has the chi-squared distribution with q degrees of freedom, denoted $\chi^2(q)$. The $\chi^2(q)$ distribution has mean q and variance $2q$.

Some statistical commands report a χ^2 statistic that is χ^2 distributed rather than an F statistic that is F distributed. In that case note that a $\chi^2(q)$ random variable divided by its degrees of freedom q has the $F(q,\infty)$ distribution. For example, a $\chi^2(10)$-distributed test statistic of 40 corresponds to an $F(10,\infty)$-distributed test statistic of $40/10 = 4$.

11.5.10 Separate Tests of Many Hypotheses

This section has considered jointly testing several hypotheses, such as whether the regressors are jointly statistically significant. What if instead we perform a series of separate tests?

Suppose we run twenty independent hypothesis tests at significant level 0.05 when in all twenty cases the null hypothesis is correct and there is no relationship. Then there is very high probability that at least one of the twenty tests rejects and obtains a spurious result. To see this, note that for a test of size 0.05 the probability of an individual test not rejecting when there is indeed no relationship is $1 - 0.05 = 0.95$. The probability of none of the twenty tests not rejecting is therefore $0.95^{20} = 0.358$. So the probability of at least one of the twenty tests rejecting is $1 - 0.358 = 0.642$.

Thus if we do many tests of statistical significance there is a high chance of finding relationships even where they do not exist.

One should therefore be sceptical about claims that, for example, eating fish has beneficial effects for the heart if it is likely that the study did a battery of separate tests on models that included a range of foods as regressors, or included a range of health outcomes, and used the usual critical values for each test. Similarly one should be sceptical of claims that a particular investment strategy is a winning strategy because it has historically outperformed the market. Many researchers will have tested many strategies using the same data and there is a bias to reporting only the historically best strategies.

What can be done to avoid erroneously finding results? More advanced methods for **multiple testing** that are beyond the scope of this text can provide an estimate of test size when multiple tests are performed. There are also simpler versions of these methods but unfortunately these are too conservative as they actually over-estimate the size in the common case that the tests are positively correlated rather than independent.

The simplest solution is to minimize the use of separate tests and as much as possible use joint hypothesis tests that simultaneously test several hypotheses.

11.6 F Statistic under Assumptions 1-4

We now specialize to the special case of independent, homoskedastic errors (assumptions 1-4). It is then possible to calculate F test statistics using residual sums of squares.

Intuitively if the fit of the restricted model is very close to that of the unrestricted model we should favor the restricted model since it is a smaller model, whereas if the difference is great then the cost of imposing the restrictions is too high and we should favor the unrestricted model.

Under assumptions 1-4 this intuition can be formalized, using the residual sum of squares, $\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2$, as a measure of model fit. Let $ResSS_r$ denote the residual sum of squares in the restricted model, and $ResSS_u$ denote the residual sum of squares in the unrestricted model. Then $ResSS_r > ResSS_u$ since OLS minimizes the residual sum of squares and this minimum will be lower in the more general unrestricted model. A large value for $ResSS_r - ResSS_u$ indicates that the restricted model has much worse fit, so we should not use the restricted model. The F statistic uses an appropriate rescaling of $ResSS_r - ResSS_u$.

Remark 99 *Under assumptions 1-4 the* **F statistic** *can be computed as*

$$F = \frac{(ResSS_r - ResSS_u)/q}{ResSS_u/(n - k)},$$

where ResSS denotes residual sum of squares, subscripts r and u denote the restricted and unrestricted models, q is the number of parameter restrictions, n is the number of observations, and k is the number of regressors in the unrestricted model.

This statistic is exactly $F(q, n-k)$ distributed if additionally the data are normally distributed.

11.6.1 Test of Overall Significance under Assumptions 1-4

For tests of overall significance we have the following simplifications. First, there are $q = k - 1$ restrictions. Second, the restricted model is regression of y on an intercept that yields $b_1 = \bar{y}$, so $ResSS_r = \sum_{i=1}^{n}(y_i - \bar{y})^2$. But this is just *TSS*, the total sum of squares in the unrestricted model. So $ResSS_r - ResSS_u = TSS - ResSS = ExpSS$, since the total sum of squares in the unrestricted model can be decomposed as $TSS = ExpSS + ResSS$.

Remark 100 *Under assumptions 1-4 A* **test of overall significance** *is a* **two-sided test of** $H_0 : \beta_2 = 0, ..., \beta_k = 0$, *against the alternative hypothesis* $H_a :$ *At least one of* $\beta_2 \neq 0, ..., \beta_k \neq 0$. *Then the general formula simplifies to*

$$F = \frac{(ExpSS)/(k - 1)}{ResSS/(n - k)},$$

where ExpSS and ResSS are, respectively, the explained and residual sum of squares in the unrestricted model.

The F statistic can be directly calculated from an ANOVA table. For the house price example it is the ratio of the first two terms in the last column in Table 11.2. From the `MS` column of that table, $(ExpSS)/(k - 1) = 4.244 \times 10^9$ and $(ResSS)/(n - k) = 0.6218 \times 10^9$. The ratio is then 6.83, which was given as the entry $F(6, 22) = 6.83$.

The F statistic can also be expressed in terms of R^2 as

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

This implies that F necessarily increases as R^2 increases. The F statistic essentially provides a threshold for the increase in R^2 before we can say that the regressors lead to a statistically significant improvement in model fit.

Returning to the house price example, $R^2 = 0.6506$ and the formula for F in terms of R^2 yields $F = (.6506/6)/(.3494/22) = 6.83$.

11.6.2   Test of Subsets of Regressors under Assumptions 1-4

For tests of subsets of regressors we need the restricted sum of squares in restricted and unrestricted models.

For the house price example consider test of whether the five regressors other than house size when taken together are statistically significant, or whether a model with just house size as a regressor is adequate.

From Table 11.2, $ResSS_u = 1.3679 \times 10^{10}$ in the unrestricted model that includes all six regressors. The restricted model is a bivariate regression of house price on house size. From Table 5.5 in Chapter 5.7, $ResSS_r = 1.4975 \times 10^{10}$. Then

$$F = \frac{(1.4975 \times 10^{10} - 1.3679 \times 10^{10})/5}{1.3679 \times 10^{10}/22} = 0.417.$$

Since $p = \Pr[F_{5,22} > 0.417] = 0.832 > 0.05$, we do not reject $H_0 : \beta_3 = 0, ..., \beta_7 = 0$ at significance level 0.05.

11.6.3   Relationship between F test and Adjusted R-Squared

Under assumptions 1-4 it can be shown that the adjusted R-squared, $\bar{R}^2$, increases as one or more regressors are added to the model if and only if the F statistic for these additional regressors exceeds one.

By contrast an F test at the usual significance level of 0.05 has a much higher threshold than $F > 1$. It follows that an increase in $\bar{R}^2$ does not necessarily imply statistical significance of the additional regressors at conventional levels of significance such as $\alpha = 0.05$. While $\bar{R}^2$ provides a penalty for large model size, the penalty is weak.

Remark 101 *As regressors are added the adjusted R-squared increases if and only if $F > 1$ where F is the test statistic for test of joint statistical significance of the additional regressors computed under assumptions 1-4.*

11.7 Presentation of Regression Results

Published articles differ in the method of presentation of regression results due to a desire to economize on space used in reporting results. Key slope coefficients are always reported, though not necessarily all coefficients if there are many regressors. R^2 is usually reported and the F statistic for overall significance is usually reported.

There is great variation in the extent to which combinations of the standard error, t statistic (for test that the population coefficient equals zero), and its associated p-value are reported. Given knowledge of one of these three, and knowledge of the slope coefficient, it is always possible to compute the other two. For example, for the jth regressor given b_j and $se(b_j)$, we can compute $t_j = b_j/se(b_j)$ and $p_j = \Pr[|T_{n-k}| \leq t_j]$ where $T_{n-k} \sim T(n-k)$. Similarly, given b_j and t_j we can compute $se(b_j) = b_j/t_j$. Finally confidence intervals can be computed given b_j and $se(b_j)$.

It is easiest if all four of b_j, $se(b_j)$, t_j and p_j are reported, along with confidence intervals. Indeed these are all given in typical computer output. But for space reasons, especially if there are several different models estimated or if the models have additional regressors, it is quite common for published studies to report only b_j and one of $se(b_j)$, t_j and p_j.

Thus for regression of house price on house size and the number of bedrooms, using default standard errors, we might report the coefficients and standard errors

$$\widehat{Price} = \underset{(27589)}{111691} + \underset{(13.30)}{72.41} \times Size + \underset{(7846)}{1553} \times Bedrooms \qquad R^2 = 0.618.$$

Alternatively we may report the coefficients, along with t statistics for whether the population coefficients equal zero

$$\widehat{Price} = \underset{(4.05)}{111691} + \underset{(5.44)}{72.41} \times Size + \underset{(0.20)}{1553} \times Bedrooms \qquad R^2 = 0.618.$$

Or just the coefficients and p-values (for test of $\beta_j = 0$) may be reported

$$\widehat{Price} = \underset{(0.000)}{111691} + \underset{(0.000)}{72.41} \times Size + \underset{(0.845)}{1553} \times Bedrooms \qquad R^2 = 0.618.$$

A fourth possibility is to report the 95% confidence intervals, possibly also with the coefficients

$$Price = \underset{(54980,168401)}{111691} + \underset{(45.07,99.75)}{72.41} \times Size + \underset{(-14576,17683)}{1553} \times Bedrooms \qquad R^2 = 0.618.$$

And yet another possibility, one that takes less space, is to report just coefficients along with asterisks that indicate the level of statistical significance. Then one, two, or three asterisks are used for statistical significance levels of, respectively, 10%, 5%, and 1%. Thus

$$Price = 111691^{***} + 72.41 \times Size^{***} + 1553 \times Bedrooms \qquad R^2 = 0.618.$$

Table 11.4 presents results reported using these various methods of presentation. Since there are various ways to compute the standard errors it is good practice to additionally include a table footnote stating the method used.

Using any of these alternatives we can verify that the slope coefficient is statistically significant at level 0.05. And while we have focused on the slope coefficient it is clear from this output that the intercept is also statistically significant at level 0.05.

Table 11.4: Regression estimates: Various ways to report the results.

In parentheses:	Results 1 Standard errors	Results 2 t statistics	Results 3 p-values	Results 4 95% Conf. int.	Results 5
Size	72.41	72.41	72.41	72.41	72.41***
	(13.30)	(5.44)	(0.000)	(45.07,99.75)	
Bedrooms	1553	1553	1553	1553	1553
	(7847)	(0.20)	(0.845)	(-14576,17682)	
Intercept	11691	11691	11691	11691	11691***
	(27589)	(4.05)	(0.000)	(54981,168401)	
R^2	0.618	0.618	0.618	0.618	0.618
F(2,26)	21.93	21.93	21.93	21.93	21.93
n	29	29	29	29	29

Note: Default standard errors are used.

11.8   Key Concepts

1. The methods of statistical significance are generally similar to those used in the bivariate case, except the population model includes additional regressors and conditioning is on these additional regressors.

2. It is assumed that there is sufficient variation in the regressor values across observations to enable computing the OLS coefficients.

3. For linear regression $E[y|x] = \beta_1 + \beta_2 x_2 + \cdots + \beta_k x_k$.

4. Population Assumptions 1-4 now condition on $x_2, ..., x_k$ and not just x.

5. Under assumptions 1 and 2, $b_1, ..., b_k$ are unbiased estimators and, with additional assumptions such as assumptions 3 and 4, consistent estimators of $\beta_1, ..., \beta_k$.

6. Under assumptions 1-4, least squares is the best linear unbiased estimator. If additionally the errors are normally distributed then least squares is the best estimator.

7. The slope coefficient standard error $se(b_j)$ is smaller the better the regression line fits the data, the larger the sample, and the greater the variability of the regressors in the sample, after controlling for the other regressors.

8. Statistical inference on β_j is based on $t = (\beta_j - b_j)/se(b_j)$ that is treated as being $T(n-k)$ distributed.

9. Confidence intervals for β_j and hypothesis tests on β_j are similar to the bivariate case except that the $T(n-k)$ distribution is used.

10. Joint hypothesis tests on several parameters can be implemented using an F test.

11. A test of overall significance is a test of whether $\beta_2 = 0, ..., \beta_k = 0$.

12. A test of subset of regressors is a test of whether a subset of $\beta_{g+1}, ..., \beta_k$ is equal to zero.

13. F is approximately $F(q, n - k)$ distributed under the null hypothesis that the q parameter restrictions are correct.

14. In general the formula for F is complicated.

15. In the special case of inference under assumptions 1-4, $F = [(ResSS_r - ResSS_u)/q]/ResSS_u/(n - k)]$ where $ResSS$ is the residual sum of squares and subscripts r and u denote, respectively, the restricted and unrestricted models.

16. Regression results given in published articles usually give the estimated coefficients along with just one of the following: standard error, t statistic, p-value and 95% confidence interval.

17. Key terms: population model; error term; disturbance term; assumptions 1-4; standard error of the regression coefficient; standard error of the regression; t statistic; t distribution; degrees of freedom; parameter; unbiased; best unbiased; best linear unbiased; confidence interval; confidence region; null hypothesis; alternative hypothesis; one-sided test; two-sided test; rejection; p-value; critical value; critical region; F distribution; F test; test of overall significance; test of subsets of regressors.

11.9 Exercises

1. Suppose the population model is $y = 1 + 2x_2 + 3x_3 + u$ and $E[u|x] = 0$. We obtain a fitted regression line $\widehat{y} = 1.7 + 2.5x_2 + 1.7x_3$. For an observation with $(x_2, x_3, y) = (2, 4, 17.8)$ give each of the following.

    (a) The conditional mean of y given x.   (b) The error term.   (c) The fitted value of y.

    (d) The residual term.

2. Repeat exercise 1 with $y = 2 + 2x_2 + 3x_3 + u$ and $\widehat{y} = 2.5 + 1.2x_2 + 3.7x_3$ and observation $(x_2, x_3, y) = (3, 2, 12)$.

3. Suppose $y = 3 + 5x_2 + 14x_3 + u$ with $u \sim (0, 9)$.

    (a) Give the conditional mean of y given x.   (b) Give the conditional variance of y given x.

4. For each of the following state when the slope coefficient is likely to be more precisely estimated, holding other things constant.

    (a) The sample size is 100 or the sample size is 400.

    (b) The regressor of interest ranges from 30 to 40 or the regressor ranges from 0 to 70.

    (c) The error term has variance 2 or the error term has variance 10.

5. The population model is $y_i = 1 + 2x_{2i} + 4x_{3i} + u_i$ where the error term has conditional mean 0 and conditional variance $4x_i^2$.

   (a) Will the OLS estimator of β_2 be unbiased? Explain.

   (b) Will the OLS estimator of β_3 be unbiased? Explain.

   (c) Will the OLS estimator of β_2 have the variance given in Chapter 11.1? Explain.

6. For each of the following situations state whether the OLS estimator of a slope coefficient is unbiased and whether its variance is that given in Chapter 11.1.

   (a) Assumptions 1-4 hold.

   (b) Assumptions 1-2 hold but assumptions 3-4 do not hold.

   (c) Assumptions 1-2 do not hold but assumptions 3-4 do hold.

7. A detailed description of dataset SALARYSAT is given in exercise 3 of Chapter 7. We consider the relationship between salary (*salary*) and *satmath, satverb, age* and *highgrade*. Use heteroskedastic-robust standard errors.

   (a) Obtain the correlations between the variables. Which variable is most highly correlated with *salary*?

   (b) Regress *salary* on the other variables. Which variables are statistically significant at 5%?

   (c) Do any of the regressors have coefficients with unexpected sign? Explain.

   (d) Provide an interpretation of the coefficient of *satmath*.

   (e) Are the regressors jointly statistically significant at 5%? Explain.

   (f) Which model has better fit: this model or the model with just *satmath* as a regressor? Explain.

   (g) Are the variables other than *satmath* jointly statistically significant at 5%? Explain.

8. Repeat the previous exercise using dataset HOUSE2015 described in exercise 7 of Chapter 5. Regress *price* on an intercept, *size, bedrooms, bathrooms* and *daysonmarket*. Compare models with just *size* as a regressor with the models with all the regressors.

9. A detailed description of dataset SALARYSAT is given in exercise 3 of Chapter 7. We consider the best fitting relationship between salary (*salary*) and most of the other variables. Use heteroskedastic-robust standard errors.

   (a) Regress *salary* on *satmath, satverb, highgrade, age, sex, minority, height, weight* and *genhealth*.

   (b) Sequentially drop the least statistically significant variable until you have a model where all included regressors are statistically significant at 5%. List the variables in the order dropped, and list the retained variables.

   (c) Now separately add back into your preferred model each of the dropped variables. Do any of them become statistically significant at 5%?

10. Repeat the previous exercise using dataset HOUSE2015. Regress *price* on an intercept, *size, bedrooms, bathrooms* and *daysonmarket*.

11. Consider regression with $n = 50$ observations, $k = 7$ regressors including the intercept and test at 5% of the overall statistical significance of the regressors.

   (a) If $F = 3.0$ do you reject the null hypothesis? Explain.

   (b) If $p = 0.05$ what was the value of F?

   (c) Now suppose a chi-squared test is used. If $\chi^2 = 12.0$ do you reject the null hypothesis? Explain. (Hint: See Figure E.6 or Table A.2 or A.3).

   (d) If $p = 0.05$ what was the value of the χ^2 statistic.

12. Repeat the previous exercise with $n = 70$, $k = 8$, $F = 2.7$, $p = 0.14$, and $\chi^2 = 13.0$.

13. Suppose we add a regressor. For each of the following state whether the regressor is definitely statistically significant at level 0.05 using default standard errors.

   (a) A t test of statistical significance of the variable has $p = 0.04$.

   (b) The inclusion of the regressor increased R^2.

   (c) The inclusion of the regressor increased adjusted R^2.

   (d) The inclusion of the regressor decreased s_e.

   (e) The F statistic comparing the model without this additional regressor and with this additional regressor had $p = 0.04$.

14. Consider tests of overall statistical significance under assumptions 1-4. Then $F = [ExpSS/(k-1)]/[ResSS/(n-k)]$.

   (a) Show that this implies $F = [R^2/(k-1)]/[(1-R^2)/(n-k)]$. Hint: Divide numerator and denominator by TSS and use the definitions of R^2.

   (b) Will F necessarily increase as R^2 increases, holding n and k fixed?

   (c) Will F necessarily increase as n increases, holding R^2 and k fixed?

15. You are given the estimates $\widehat{y} = \underset{(6)}{10} + \underset{(40)}{72}\, x_2 + \underset{(5)}{16} x_3$ where standard errors are given in parentheses. The sample size is large so you can use standard normal critical values.

   (a) Obtain the t statistic for b_2, the coefficient of x_2.

   (b) Is x_2 statistically significant at 5%?

   (c) Provide a 95% confidence interval for β_2.

16. Repeat the previous exercise if $\widehat{y} = \underset{(28)}{110} + \underset{(8)}{18}x_2 + \underset{(10)}{24}\, x_3$.

17. You are given the estimates $\widehat{y} = \underset{(2.5)}{10} + \underset{(3)}{72}x_2 + \underset{(5)}{16}x_3$ where t **statistics** are given in parentheses. The sample size is large so you can use standard normal critical values.

   (a) Is x_2 statistically significant at 5%?

   (b) Obtain the standard error for b_2, the coefficient of x_2.

   (c) Provide a 95% confidence interval for β_2.

18. Repeat the previous exercise if $\widehat{y} = \underset{(3.5)}{110} + \underset{(12)}{18}\, x_2 + \underset{(10)}{24}\, x_2$.