ACE 564
Spring 2006

Lecture 4

The Multiple Regression Model: Sampling Properties and Interval Estimation

by
Professor Scott H. Irwin

Readings:

Griffiths, Hill and Judge. "Sampling Properties of the Least Squares Estimator," Section 10.1 and "Interval Estimation," Section 10.2 in *Learning and Practicing Econometrics*

Sampling Properties of the Least Squares Estimators in the Multiple Variable Case

In the previous lecture, we specified a linear economic model that led to the following statistical model, known as the multivariate linear regression model

$$y_t = \beta_1 + \beta_2 x_{2,t} + \beta_3 x_{3,t} + e_t$$

We assumed that e_t and y_t are *iid* normal random variables

In compact form, the assumptions of the multiple regression model are

MR1. $y_t = \beta_1 + \beta_2 x_{2,t} + \beta_3 x_{3,t} + e_t, \quad t = 1,\ldots,T$

MR2. $E(e_t) = 0 \Leftrightarrow E(y_t) = \beta_1 + \beta_2 x_{2,t} + \beta_3 x_{3,t}$

MR3. $\text{var}(e_t) = \text{var}(y_t) = \sigma^2$

MR4. $\text{cov}[e_t, e_s] = \text{cov}[y_t, y_s] = 0 \qquad t \neq s$

MR5. The values of $x_{2,t}$ and $x_{3,t}$ are not random or exact linear functions of one another

MR6. $e_t \sim N(0,\sigma^2) \Leftrightarrow y_t \sim N(\beta_1 + \beta_2 x_{2,t} + \beta_3 x_{3,t}, \sigma^2)$

One sample of data was obtained consisting of 52 observations on total revenue, price and advertising for the Bay Area Rapid Food Company

$\Rightarrow$ We assumed that the sample data was generated by the previous statistical model

Given this sample of data, we developed the following rules (estimators) for estimating the intercept and slope parameters of the (true, but unknown) linear statistical model

$$b_1 = \bar{y} - b_2\bar{x}_2 - b_3\bar{x}_3$$

$$b_2 = \frac{\sum_{t=1}^{T} y'_t x'_{2,t} \sum_{t=1}^{T} x'^2_{3,t} - \sum_{t=1}^{T} y'_t x'_{3,t} \sum_{t=1}^{T} x'_{2,t} x'_{3,t}}{\sum_{t=1}^{T} x'^2_{2,t} \sum_{t=1}^{T} x'^2_{3,t} - \left(\sum_{t=1}^{T} x'_{2,t} x'_{3,t} \right)^2}$$

$$b_3 = \frac{\sum_{t=1}^{T} y'_t x'_{3,t} \sum_{t=1}^{T} x'^2_{2,t} - \sum_{t=1}^{T} y'_t x'_{2,t} \sum_{t=1}^{T} x'_{2,t} x'_{3,t}}{\sum_{t=1}^{T} x'^2_{2,t} \sum_{t=1}^{T} x'^2_{3,t} - \left(\sum_{t=1}^{T} x'_{2,t} x'_{3,t} \right)^2}$$

where

$$y'_t = y_t - \bar{y}$$
$$x'_{2,t} = x_{2,t} - \bar{x}_2$$
$$x'_{3,t} = x_{3,t} - \bar{x}_3$$

Using the sample data and least squares estimators, we computed the following least squares estimates of the unknown intercept and slopes for the statistical model

$$b_1 = 104.785 \qquad b_2 = -6.642 \qquad b_3 = 2.984$$

At this point, the estimates are simply computed numbers that have no statistical properties!

We can never know how close these particular numbers are to the true values we are trying to estimate

While we cannot know the accuracy of the least squares estimates, we can examine the properties of the estimator under repeated sampling

$\Rightarrow$ As before, we imagine "hitting" the estimators with many hypothetical samples and examining their performance across the samples

- If we obtained a different sample of 52 weeks, there would be 52 different values for total revenue

- Price and advertising levels would be the same across the samples, so that we just randomly select new values for total revenue for the given levels of price and advertising

- Applying the least squares estimation rules to this new sample would lead to different b_1, b_2 and b_3 estimates

We can imagine drawing many new samples and following the same procedure, with each new sample producing different b_1, b_2 and b_3 estimates

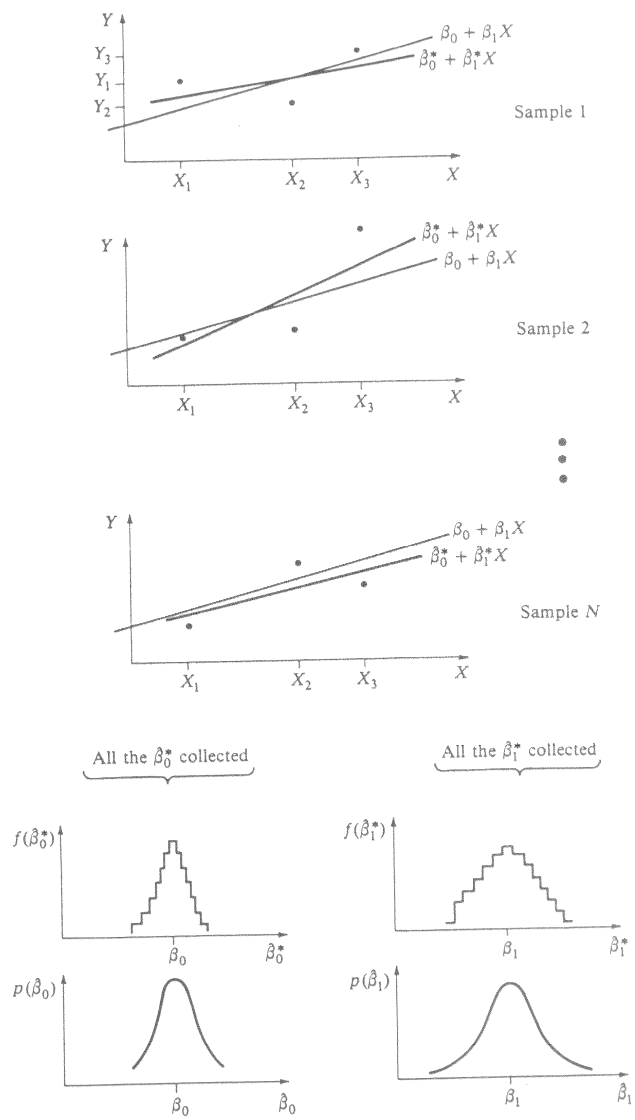

FIGURE 11.3 A thought experiment helps explain the nature of the sampling distributions of the OLS estimators of β_0 and β_1. The same three X values are used in each sample, but different sets of Y values are produced because different values for the disturbances occur in each sample. The $\hat\beta_0^*$ and $\hat\beta_1^*$ values computed in each sample are collected, and their frequency distributions are constructed. The sampling distribution of an estimator is the limiting form of the frequency distribution as N approaches infinity. The actual $\hat\beta_0^*$ and $\hat\beta_1^*$ that we calculate from a set of data are thought of as just one pair of outcomes from these sampling distributions.

Mirer, Thad W. Economic Statistics and Econometrics, Third Edition. Prentice Hall, Englewood Cliffs, NJ. 1995.

Consequently, just as in the simple regression case, the least squares estimation rules b_1, b_2 and b_3 are random variables

- An alternative way of saying that different samples of data will yield different values for b_1, b_2 and b_3

- For this reason, the variability of the estimation rules b_1, b_2 and b_3 is called sampling variability

Viewing b_1, b_2 and b_3 as random variables leads to the following important question:

> $\Rightarrow$ What are the means, variances, covariances, and forms of the repeated sampling pdf's for the random variables b_1, b_2 and b_3?

If someone proposes an estimation rule for a particular statistical model, your next question should be: What are its sampling characteristics and how good is it?

---Griffiths, Hill and Judge, *LPE*, p.209

It may be helpful to review Lecture 5 from ACE 562 when studying the material in this Lecture

Means, Variances and Covariance of Sampling Distributions

Expected values,

$$E(b_1) = \beta_1 \qquad E(b_2) = \beta_2 \qquad E(b_3) = \beta_3$$

$\Rightarrow$ Shows that least squares estimators in the multivariate case are unbiased

Sampling variances,

$$\operatorname{var}(b_1) = \sigma^2 \left[\frac{1}{T} + \frac{\bar{x}_2^2 \sum_{t=1}^{T} x_{3,t}'^2 + \bar{x}_3^2 \sum_{t=1}^{T} x_{2,t}'^2 - 2\bar{x}_2\bar{x}_3 \sum_{t=1}^{T} x_{2,t}' x_{3,t}'}{\sum_{t=1}^{T} x_{2,t}'^2 \sum_{t=1}^{T} x_{3,t}'^2 - \left(\sum_{t=1}^{T} x_{2,t}' x_{3,t}' \right)^2} \right]$$

$$\operatorname{var}(b_2) = \sigma^2 \left[\frac{\sum_{t=1}^{T} x_{3,t}'^2}{\sum_{t=1}^{T} x_{2,t}'^2 \sum_{t=1}^{T} x_{3,t}'^2 - \left(\sum_{t=1}^{T} x_{2,t}' x_{3,t}' \right)^2} \right]$$

$$\operatorname{var}(b_3) = \sigma^2 \left[\frac{\sum_{t=1}^{T} x'^2_{2,t}}{\sum_{t=1}^{T} x'^2_{2,t} \sum_{t=1}^{T} x'^2_{3,t} - \left(\sum_{t=1}^{T} x'_{2,t} x'_{3,t} \right)^2} \right]$$

Covariance of b_2 and b_3,

$$\operatorname{cov}(b_2, b_3) = \sigma^2 \left[\frac{-\frac{\sum_{t=1}^{T} x'_{2,t} x'_{3,t}}{\sum_{t=1}^{T} x'^2_{2,t} \sum_{t=1}^{T} x'^2_{3,t}}}{\sum_{t=1}^{T} x'^2_{2,t} \sum_{t=1}^{T} x'^2_{3,t} - \left(\sum_{t=1}^{T} x'_{2,t} x'_{3,t} \right)^2} \right]$$

where,

$$y'_t = y_t - \bar{y}$$

$$x'_{2,t} = x_{2,t} - \bar{x}_2$$

$$x'_{3,t} = x_{3,t} - \bar{x}_3$$

We can simplify the expressions for the sampling variances and covariance by defining the sample correlation coefficient between $x_{2,t}$ and $x_{3,t}$,

$$r_{23} = \frac{\sum_{t=1}^{T} x'_{2,t} x'_{3,t}}{\sqrt{\sum_{t=1}^{T} x'^2_{2,t} \sum_{t=1}^{T} x'^2_{3,t}}}$$

Then,

$$\text{var}(b_1) = \sigma^2 \left[\frac{1}{T} + \frac{\dfrac{\overline{x}_2^2}{\sum_{t=1}^{T} x'^2_{2,t}} + \dfrac{\overline{x}_3^2}{\sum_{t=1}^{T} x'^2_{3,t}} - 2\overline{x}_2\overline{x}_3 \dfrac{\sum_{t=1}^{T} x'_{2,t} x'_{3,t}}{\sum_{t=1}^{T} x'^2_{2,t} \sum_{t=1}^{T} x'^2_{3,t}}}{1 - r_{23}^2} \right]$$

$$\text{var}(b_2) = \frac{\sigma^2}{\sum_{t=1}^{T} x_{2,t}'^2 \left(1 - r_{23}^2\right)}$$

$$\text{var}(b_3) = \frac{\sigma^2}{\sum_{t=1}^{T} x_{3,t}'^2 \left(1 - r_{23}^2\right)}$$

$$\text{cov}(b_2, b_3) = \frac{-r_{23}^2 \sigma^2}{\left(1 - r_{23}^2\right)\sqrt{\sum_{t=1}^{T} x_{2,t}'^2}\sqrt{\sum_{t=1}^{T} x_{3,t}'^2}}$$

It is worthwhile to examine the special case where r_{23} is equal to zero,

$$\text{var}(b_1) = \sigma^2 \left[\frac{1}{T} + \frac{\bar{x}_2^2}{\sum_{t=1}^{T} x_{2,t}'^2} + \frac{\bar{x}_3^2}{\sum_{t=1}^{T} x_{3,t}'^2} - 2\bar{x}_2\bar{x}_3 \frac{\sum_{t=1}^{T} x_{2,t}' x_{3,t}'}{\sum_{t=1}^{T} x_{2,t}'^2 \sum_{t=1}^{T} x_{3,t}'^2} \right]$$

$$\mathrm{var}(b_2) = \frac{\sigma^2}{\sum_{t=1}^{T} x'^2_{2,t}} \qquad \mathrm{var}(b_3) = \frac{\sigma^2}{\sum_{t=1}^{T} x'^2_{3,t}}$$

$$\mathrm{cov}(b_2, b_3) = 0$$

Important observations:

Just as in the case of simple linear regression, the variance of the error term (σ^2) appears in each formula

- The larger is σ^2 the larger the sampling variances of the least squares estimators; this is to be expected since σ^2 measures the overall uncertainty in the model specification

- If σ^2 is large, then data values may be widely spread about the regression function

- In other words, sample information available to estimate β_1, β_2 and β_3 is less precise the larger is σ^2

The sum of squares for $x_{k,t}$ $(k=2,3)$,

$$\sum_{t=1}^{T} x_{k,t}'^2 = \sum_{t=1}^{T} (x_{k,t} - \bar{x}_k)^2$$

appears in the denominator of the formula for $\text{var}(b_k)$

- Sum of squares for $x_{k,t}$ measures the spread, or variation, of $x_{k,t}$

- The larger the variation in $x_{k,t}$, the smaller the variance of b_k

  $\Rightarrow$ The more information we have about $x_{k,t}$, the more precisely can we estimate β_k

  $\Rightarrow$ In other words, it is easier to measure β_k, the change in y we expect given a change in $x_{k,t}$, the more sample variation (change) in the values of $x_{k,t}$ that we observe

The larger the sample size T, the smaller is $\text{var}(b_k)$

- As T increases, the sum of squares for $x_{k,t}$ increases unambiguously

- More observations means we can estimate the slope parameters more precisely

Correlation of Independent Variables

Recall that the denominator of $\text{var}(b_k)$ contains the term $1-r_{23}^2$

- In general, r_{23}, the correlation between the sample values of $x_{2,t}$ and $x_{3,t}$ will not equal zero

- If the values of $x_{2,t}$ and $x_{3,t}$ are correlated, then $1-r_{23}^2$ is a fraction that is less than 1

- Hence, the larger the correlation between $x_{2,t}$ and $x_{3,t}$ (positive or negative) the larger is the variance of the least squares estimator b_2

Variation in $x_{k,t}$ adds most to the precision of estimation when it is not connected to variation in the other explanatory variables

- "Independent" variables ideally exhibit variation that is "independent" of the variation in other explanatory variables

- When the variation in one explanatory variable is connected to variation in another explanatory variable, it is difficult to disentangle their separate effects

- This is known as the problem of "multicollinearity," which we will devote an entire lecture to later in the course

Summary of Key Factors Affecting Precision of Least Squares Estimators In the Multivariate Case

In the following table, characteristics of the sample data are categorized in terms of impact on precision simply as "More Precision" or "Less Precision"

Sample Characteristic	More Precision	Less Precision
High Variance of y_t		✓
Low Variance of y_t	✓	
High Variation in $x_{k,t}$	✓	
Low Variation in $x_{k,t}$		✓
Large T	✓	
Small T		✓
Low Correlation of $x_{2,t}$ and $x_{3,t}$	✓	
High Correlation of $x_{2,t}$ and $x_{3,t}$		✓
Small Distance from Origin	✓	
Large Distance from Origin		✓

Note that these factors generally work in the same manner for the multiple regression model and the simple linear regression model

Standard Errors of the Sampling Distributions

The standard deviations of the sampling distributions for b_2 and b_3 are found in the usual manner

$$se(b_2) = \sqrt{\mathrm{var}(b_2)} = \sqrt{\frac{\sigma^2}{\sum_{t=1}^{T} x_{2,t}'^2 \left(1 - r_{23}^2\right)}}$$

$$se(b_3) = \sqrt{\mathrm{var}(b_3)} = \sqrt{\frac{\sigma^2}{\sum_{t=1}^{T} x_{3,t}'^2 \left(1 - r_{23}^2\right)}}$$

Notice that the term "standard error" (se) is once again used in place of standard deviation of the sampling distribution

To see why, define the error that arises in estimating one of the true slope parameters as $f = b_2 - \beta_2$

Applying our rules for the mean and variance of the transformation of a random variable we find that

$$E(f) = 0 \text{ and } \operatorname{var}(f) = \operatorname{var}(b_2)$$

Since the expected estimation error is zero, we can say that the typical error (without regard to sign) is given by the standard deviation of f, which equals the standard deviation of b_2

If we replace "typical" with "standard" we can say that $se(b_2)$ measures the standard estimation error for b_2, or in abbreviated form, standard error

Sampling Distributions of the Least Squares Estimators

$$b_1 \sim N\left(\beta_1, \text{var}(b_1)\right)$$

where

$$\text{var}(b_1) = \sigma^2 \left[\frac{1}{T} + \frac{\dfrac{\bar{x}_2^2}{\sum_{t=1}^{T} x_{2,t}'^2} + \dfrac{\bar{x}_3^2}{\sum_{t=1}^{T} x_{3,t}'^2} - 2\bar{x}_2\bar{x}_3 \dfrac{\sum_{t=1}^{T} x_{2,t}' x_{3,t}'}{\sum_{t=1}^{T} x_{2,t}'^2 \sum_{t=1}^{T} x_{3,t}'^2}}{1 - r_{23}^2} \right]$$

$$b_2 \sim N\left(\beta_2, \text{var}(b_2)\right)$$

where

$$\text{var}(b_2) = \frac{\sigma^2}{\sum_{t=1}^{T} x_{2,t}'^2 \left(1 - r_{23}^2\right)}$$

$$b_3 \sim N\left(\beta_3, \text{var}(b_3)\right)$$

where

$$\text{var}(b_3) = \frac{\sigma^2}{\sum_{t=1}^{T} x_{3,t}'^2 \left(1 - r_{23}^2\right)}$$

Gauss-Markov Theorem in the Multiple Variable Case

The least squares estimators b_1, b_2 and b_3 of the population parameters β_1, β_2 and β_3 are,

- Linear
- Unbiased
- Efficient
- Consistent

The first three properties are sufficient to prove that b_1, b_2 and b_3 are the best linear unbiased estimators (BLUE) of β_1, β_2 and β_3

- In this context "best" implies minimum variance sampling distribution
- Known as the Gauss-Markov Theorem

Key Points:

- The estimators b_1, b_2 and b_3 are "best" when compared to similar estimators, those that are linear and unbiased. The Gauss-Markov Theorem does not say that b_1, b_2 and b_3 are the best of all possible estimators.

- The estimators b_1, b_2 and b_3 are best within their class because they have the minimum variance.

- In order for the Gauss-Markov Theorem to hold, the assumptions (MR1-MR6) must be true. If any of the assumptions 1-6 are not true, then b_1, b_2 and b_3 are not the best linear unbiased estimators of β_1, β_2 and β_3

- The Gauss-Markov Theorem does not depend on the assumption of normality.

- In the multiple regression model, if we want to use linear and unbiased estimators, then we have to do no more searching.

- The Gauss-Markov theorem applies to the least squares estimators. It does not apply to the least squares estimates from a single sample.

Estimating the Variance of the Error Term

Recall that e_t and y_t are assumed to be *iid* with the following distributions,

$$e_t \sim N(0, \sigma^2) \quad \text{and} \quad y_t \sim N(\beta_1 + \beta_2 x_{2,t} + \beta_3 x_{3,t}, \sigma^2)$$

Unless σ^2 is known, which is highly unlikely, it will have to be estimated as well

We cannot use the least squares principal, as σ^2 does not appear in the sum of squares function,

$$S(\beta_1, \beta_2, \beta_3) = \sum_{t=1}^{T} e_t^{\,2} = \sum_{t=1}^{T} (y_t - \beta_2 x_{2,t} - \beta_3 x_{3,t})^2$$

Just as in the case of simple linear regression, we apply a "heuristic" procedure based on the definition of σ^2

The original definition of σ^2 in the multiple regression model is,

$$\text{var}(y_t) = \text{var}(e_t) = \sigma^2 = E[e_t^2]$$

In other words, the variance is the expected value of the squared errors

Given this definition, it is natural to estimate σ^2 as the average of the squared errors

In order to do this, we must first obtain estimates of the population errors using our sample data as

$$\hat{e}_t = y_t - b_1 - b_2 x_{2,t} - b_3 x_{3,t}$$

We can then develop our sample estimator of σ^2 as,

$$\hat{\sigma}^2 = \frac{\hat{e}_1^2 + \hat{e}_2^2 + \ldots + \hat{e}_T^2}{T-3} = \frac{\sum_{t=1}^{T} \hat{e}_t^2}{T-3}$$

- Notice the squared sample errors are averaged by dividing by $T\text{-}3$

- Accounts for the fact that 3 regression parameters $(\beta_1, \beta_2, \beta_3)$ have to be estimated

Most regression packages report the "standard error of the regression"

This is simply the square root of the estimated variance of the error term,

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2} = \sqrt{\frac{\hat{e}_1^2 + \hat{e}_2^2 + \ldots + \hat{e}_T^2}{T-3}} = \sqrt{\frac{\sum_{t=1}^{T} \hat{e}_t^2}{T-3}}$$

Standard error of the regression is usually reported instead of the variance since the standard error has the same units as y_t

For the Bay Area Rapid Food data,

$$\hat{\sigma}^2 = \frac{\sum_{t=1}^{T} \hat{e}_t^2}{T-3} = \frac{1805.168}{52-3} = 36.84$$

and,

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2} = \sqrt{36.84} = \$6.069 \text{ (thousands)}$$

Estimators of the Variances and Covariances of Least Squares Estimators

Recall that the variances and covariance of the sampling distributions of b_1, b_2 and b_3 are functions of the unknown parameter σ^2

We can generate estimators of the variances and covariances by simply replacing σ^2 with $\hat{\sigma}^2$ in the earlier formulas,

$$\hat{\text{var}}(b_1) = \hat{\sigma}^2 \left[\frac{1}{T} + \frac{\dfrac{\bar{x}_2^2}{\sum_{t=1}^{T} x_{2,t}'^2} + \dfrac{\bar{x}_3^2}{\sum_{t=1}^{T} x_{3,t}'^2} - 2\bar{x}_2\bar{x}_3 \dfrac{\sum_{t=1}^{T} x_{2,t}' x_{3,t}'}{\sum_{t=1}^{T} x_{2,t}'^2 \sum_{t=1}^{T} x_{3,t}'^2}}{1 - r_{23}^2} \right]$$

$$\hat{\text{var}}(b_2) = \frac{\hat{\sigma}^2}{\sum_{t=1}^{T} x_{2,t}'^2 \left(1 - r_{23}^2\right)}$$

$$\hat{\text{var}}(b_3) = \frac{\hat{\sigma}^2}{\sum_{t=1}^{T} x_{3,t}'^2 \left(1 - r_{23}^2\right)}$$

$$\hat{\text{cov}}(b_2, b_3) = \frac{-r_{23}^2 \hat{\sigma}^2}{\left(1 - r_{23}^2\right) \sqrt{\sum_{t=1}^{T} x_{2,t}'^2} \sqrt{\sum_{t=1}^{T} x_{3,t}'^2}}$$

Likewise, the estimators for the standard errors of b_1 and b_2 are,

$$\hat{se}(b_1) = \sqrt{\hat{\text{var}}(b_1)}$$

$$= \sqrt{\hat{\sigma}^2 \cdot \left[\frac{1}{T} + \frac{\dfrac{\bar{x}_2^2}{\sum_{t=1}^{T} x_{2,t}'^2} + \dfrac{\bar{x}_3^2}{\sum_{t=1}^{T} x_{3,t}'^2} - 2\bar{x}_2 \bar{x}_3 \dfrac{\sum_{t=1}^{T} x_{2,t}' x_{3,t}'}{\sum_{t=1}^{T} x_{2,t}'^2 \sum_{t=1}^{T} x_{3,t}'^2}}{1 - r_{23}^2} \right]}$$

$$\hat{se}(b_2) = \sqrt{\hat{\text{var}}(b_2)} = \sqrt{\frac{\hat{\sigma}^2}{\sum_{t=1}^{T} x_{2,t}'^2 \left(1 - r_{23}^2\right)}}$$

$$\hat{se}(b_3) = \sqrt{\hat{\text{var}}(b_3)} = \sqrt{\frac{\hat{\sigma}^2}{\sum_{t=1}^{T} x_{3,t}'^2 \left(1 - r_{23}^2\right)}}$$

This material is tricky:

- We developed estimators of the variances, covariance and standard errors of the least squares estimators of b_1 , b_2 and b_3!!

- While we will not consider the extra complexity, the estimators of the variances, covariance and standard errors are themselves random variables whose values vary in repeated sampling

Matrix of Estimated Variances and Covariances for Bay Area Rapid Food

It is useful to organize the estimated variances and covariances of the sampling distributions of the least squares estimators into a matrix,

$$\begin{bmatrix} \hat{\text{var}}(b_1) & \hat{\text{cov}}(b_1,b_2) & \hat{\text{cov}}(b_1,b_3) \\ \hat{\text{cov}}(b_2,b_1) & \hat{\text{var}}(b_2) & \hat{\text{cov}}(b_2,b_3) \\ \hat{\text{cov}}(b_3,b_1) & \hat{\text{cov}}(b_3,b_2) & \hat{\text{var}}(b_3) \end{bmatrix}$$

The estimates for the Bay Area Rapid Food problem are,

$$\begin{bmatrix} 42.026 & -19.863 & -0.1611 \\ -19.863 & 10.184 & -0.054 \\ -0.161 & -0.054 & 0.02787 \end{bmatrix}$$

Note that the estimates of the standard errors of the least squares estimators are found by taking the square root of the diagonal elements

Sample Regression Output from Excel

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.93117
R Square	0.86708
Adjusted R Square	0.86166
Standard Error	6.06961
Observations	52

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	11776.1839	5888.0919	159.8280	0.0000
Residual	49	1805.1684	36.8402		
Total	51	13581.3523			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	104.7855	6.4827	16.1638	0.0000	91.7580	117.8130
X Variable 1	-6.6419	3.1912	-2.0813	0.0427	-13.0549	-0.2290
X Variable 2	2.9843	0.1669	17.8769	0.0000	2.6488	3.3198

Summary of Bay Area Rapid Food Estimates

$$T=52 \qquad K=3$$

$$\hat{\sigma}^2 = 36.84$$

$$b_1 = 104.785 \quad \hat{\text{var}}(b_1) = 42.026$$

$$b_2 = -6.642 \quad \hat{\text{var}}(b_2) = 10.184$$

$$b_3 = 2.984 \quad \hat{\text{var}}(b_3) = 0.02787$$

Based on this information and the assumption that the statistical model is correctly specified, we can estimate the distributions of e_t and y_t as,

$$e_t \sim N(0, 36.84)$$

$$y_t \sim N(104.785 - 6.642 x_{1,t} + 2.984 x_{2,t}, 36.84)$$

We also can estimate the sampling distributions of b_1, b_2 and b_3 as,

$$b_1 \sim N(104.785, 42.026) \qquad b_2 \sim N(-6.642, 10.184)$$

$$b_3 \sim N(2.984, 0.02787)$$

Interpretation Guidelines

Regression standard error

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2} = \sqrt{36.84} = \$6.069 \text{ (thousands)}$$

We say, "The typical error of the regression model, without regard to sign, is estimated to be <u>\$6,069</u>."

Standard error of parameter estimates

$$\hat{se}(b_1) = \sqrt{\hat{\text{var}}(b_1)} = 6.4827$$

We say, "The typical error in estimating β_1***, without regard to sign, is estimated to be <u>6.4827</u>."***

$$\hat{se}(b_2) = \sqrt{\hat{\text{var}}(b_2)} = 3.1912$$

We say, "The typical error in estimating β_2***, without regard to sign, is estimated to be <u>3.1912</u>."***

$$\hat{se}(b_3) = \sqrt{\hat{\text{var}}(b_3)} = 0.1669$$

We say, "The typical error in estimating β_3***, without regard to sign, is estimated to be <u>0.1669</u>."***

Interval Estimation

Interval estimators combine the information from the point estimator with the sampling variance to provide an indication of the reliability of the estimate

When σ^2 is known, and hence, $\text{var}(b_k)$ is known, the standardized random variable,

$$Z_{b_k} = \frac{b_k - \beta_k}{\sqrt{\text{var}(b_k)}}$$

is a function of only one random variable b_k and distributed N(0,1)

When we replace σ^2 by its unbiased estimator $\hat{\sigma}^2$, and hence, replace $\text{var}(b_k)$ with $\hat{\text{var}}(b_k)$, the resulting standardized random variable,

$$t_{b_k} = \frac{b_k - \beta_k}{\sqrt{\hat{\text{var}}(b_k)}}$$

is a function of the ratio of two random variables, b_k and $\hat{\text{var}}(b_k)$, and is distributed as t_{T-K}

In general, we know that we can pick critical values c and $-c$ such that,

$$P[t_{T-K} \geq c] = P[t_{T-K} \leq -c] = \frac{\alpha}{2}$$

or,

$$P[t_{T-K} \geq c] + P[t_{T-K} \leq -c] = \alpha$$

and,

$$P[-c \leq t_{T-K} \leq c] = 1 - \alpha$$

where α is the probability

From a t-table, we know that when T=52, K=3 and α=0.05, then c = 2.008 and $-c$ = -2.008

This can be stated in mathematical form as,

$$P[t_{49} \geq 2.008] = P[t_{49} \leq -2.008] = \frac{0.05}{2} = 0.025$$

or,

$$P[t_{49} \geq 2.008] + P[t_{49} \leq -2.008] = 0.025 + 0.025 = 0.05$$

and,

$$P[-2.008 \leq t_{49} \leq 2.008] = 1 - 0.05 = 0.95$$

Since t_{b_k} is a t-distributed random variable, we can write,

$$P[-2.008 \leq t_{b_k} \leq 2.008] = 0.95$$

Substituting for t_{b_k},

$$P[-2.008 \leq \frac{b_k - \beta_k}{\sqrt{\hat{\text{var}}(b_k)}} \leq 2.008] = 0.95$$

Multiply the inequality in the brackets by $\sqrt{\hat{\text{var}}(b_k)}$,

$$P[-2.008 \cdot \sqrt{\hat{\text{var}}(b_k)} \leq b_k - \beta_k \leq 2.008 \cdot \sqrt{\hat{\text{var}}(b_k)}] = 0.95$$

Now, subtract b_k from each term,

$$P[-b_k - 2.008 \cdot \sqrt{\hat{\text{var}}(b_k)} \leq -\beta_k \leq -b_k + 2.008 \cdot \sqrt{\hat{\text{var}}(b_k)}] = 0.95$$

Next, multiply the inequality by -1 to reverse the direction of the inequality,

$$P[b_k + 2.008 \cdot \sqrt{\hat{\text{var}}(b_k)} \geq \beta_k \geq b_k - 2.008 \cdot \sqrt{\hat{\text{var}}(b_k)}] = 0.95$$

Finally, "flip" the entire inequality as follows,

$$P[b_k - 2.008 \cdot \sqrt{\hat{\text{var}}(b_k)} \leq \beta_k \leq b_k + 2.008 \cdot \sqrt{\hat{\text{var}}(b_k)}] = 0.95$$

This is the interval estimator for β_k at the 95 percent confidence level $[(1\text{-}\alpha)100]$ when σ^2 must be estimated

We can generalize the interval estimator to any desired confidence level as follows,

$$P[b_k - t_{\alpha/2,T-K}\sqrt{\hat{\text{var}}(b_k)} \leq \beta_k \leq b_k + t_{\alpha/2,T-K}\sqrt{\hat{\text{var}}(b_k)}] = 1-\alpha$$

where $t_{\alpha/2,T-K}$ is the appropriate critical value from a t-table for α percent tail probability and T-K degrees of freedom (α is the sum of both lower and upper tail probability)

Interpretation of Interval Estimator when σ^2 is Unknown

Interval is random because b_k and $\hat{\text{var}}(b_k)$ are random variables before sampling

Location and width of interval are variable

- Width of interval is variable because $\hat{\text{var}}(b_k)$ is a random variable

- Location of interval is variable because b_k is a random variable

Correct meaning of interval estimator

- In repeated sampling, we expect 95% of interval estimates to contain β_2

- If we use the interval estimator to compute a "large" number of interval estimates like $b_2 \pm 2.021\sqrt{\hat{\text{var}}(b_2)}$, 95% of these intervals will contain β_2

Interval Estimates for Bay Area Rapid Food Data

Previously, we reported the following results,

$$b_1 = 104.785 \quad \hat{\text{var}}(b_1) = 42.026$$

$$b_2 = -6.642 \quad \hat{\text{var}}(b_2) = 10.184$$

$$b_3 = 2.984 \quad \hat{\text{var}}(b_3) = 0.02787$$

We will use this information to construct interval estimates for

$\beta_2 =$ the response of revenue to a price change

$\beta_3 =$ the response of revenue to a change in advertising expenditure

Note that

- Degrees of freedom are given by $(T\text{-}K) = (52\text{-}3) = 49$

- Critical value for $\alpha = 0.05$ is $t_c = 2.008$

The interval estimates can be formed as,

$$b_1 \pm t_{0.05/2,\,52-3}\sqrt{\hat{\text{var}}(b_1)} = 104.785 \pm 2.008 \cdot 6.483$$

$$b_2 \pm t_{0.05/2,\,52-3}\sqrt{\hat{\text{var}}(b_2)} = -6.642 \pm 2.008 \cdot 3.191$$

$$b_3 \pm t_{0.05/2,\,52-3}\sqrt{\hat{\text{var}}(b_3)} = 2.984 \pm 2.008 \cdot 0.1669$$

or,

$$(91.767 \leq \beta_1 \leq 117.803)$$

$$(-13.049 \leq \beta_2 \leq -0.234)$$

$$(2.649 \leq \beta_3 \leq 3.319)$$

Sample Regression Output from Excel

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.93117
R Square	0.86708
Adjusted R Square	0.86166
Standard Error	6.06961
Observations	52

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	11776.1839	5888.0919	159.8280	0.0000
Residual	49	1805.1684	36.8402		
Total	51	13581.3523			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	104.7855	6.4827	16.1638	0.0000	91.7580	117.8130
X Variable 1	-6.6419	3.1912	-2.0813	0.0427	-13.0549	-0.2290
X Variable 2	2.9843	0.1669	17.8769	0.0000	2.6488	3.3198

Interpretation Guidelines

- In repeated sampling, we expect 95% of interval estimates to contain β_k

- For example, if we use the interval estimator to compute a "large" number of interval estimates like $-6.642 \pm 2.008 \cdot 3.191$, 95% of these intervals will contain β_2

It is incorrect to state,

"There is a 0.95 probability that β_2 falls in the interval (-13.049, -0.234)."

- This implies β_2 is random and the interval is not, when in reality just the opposite is true!

It is also incorrect to state,

"There is a 0.95 probability that the interval (-13.049, -0.234) contains β_2."

- We will never know whether β_2 falls in the estimated interval (-13.049, -0.234)

- Remember that our confidence is in the estimator not the particular estimate

Compromise language,

> ***"We are 95% confident that the interval (-13.049, -0.234) contains β_2."***

where "confident" is understood to apply to the interval estimator in repeated sampling, not the (-13.049, -0.234) interval estimate

HGJ take a more conservative approach and argue it would only be appropriate to state,

> ***Based on our one sample of data, given the reliability of the interval estimate procedure, we would be "surprised" if β_2 did not fall in the interval (-13.049, -0.234)***

HGJ argue that confidence interval estimates are best understood as giving a general notion of reliability

- A "wide" interval suggests there is not much information in the sample about β_2

- A "narrow" interval suggests we have learned more about the true value of β_2

Whether an interval estimate should be considered "wide" or "narrow" depends on

- Research problem under study

- Economic magnitude of interval

Specific Interpretation

The 95% interval estimate for β_2, the response of revenue to price, is given by

$$(-13.049, -0.234)$$

- Indicates we are 95% confident that decreasing price by \$1 will lead to an increase in revenue between \$234 and \$13,049

- From an economic perspective, this is a wide interval and it is not very informative

- Another way of describing this situation is to say that the point estimate of $b_2 = -6.642$ is not very reliable

- A narrower interval can only be obtained by reducing the variance of the estimator

  ⇒ One way is to obtain more or "better" data

  ⇒ Alternatively, we might introduce some kind of non-sample information on the parameters

The 95% interval estimate for β_3, the response of revenue to advertising, is given by

$$(2.649, 3.319)$$

- Indicates we are 95% confident that increasing advertising expenditure by $1000 leads to an increase in total revenue between $2,649 and $3,319

- From an economic perspective, this is a relatively narrow interval and it is informative

- Another way of describing this situation is to say that the point estimate of $b_3 = 2.984$ is reliable

Generalizing to More than Two Independent Variables

Nearly all of the results discussed with respect to specification, estimation, sampling distributions, and interval estimation for the three-variable regression model generalize to multiple regression models with more than three independent variables

The general format for a multiple linear regression model with *K-1* independent variables is,

$$y_t = \beta_1 + \beta_2 x_{2,t} + \beta_3 x_{3,t} + \ldots + \beta_k x_{k,t} + e_t$$

This more general model is much easier to examine using matrix algebra

- Our good friend scalar algebra begins to fail us when we move beyond three variables!
- Excellent treatment of the matrix formulation of the multiple regression model can be found in *Learning and Practicing Econometrics* by Griffiths, Hill, and Judge (begin with Section 5.4)