CHAPTER 2: Assumptions and Properties of Ordinary Least Squares, and Inference in the Linear Regression Model

Prof. Alan Wan

Table of contents

1. Assumptions in the Linear Regression Model
2. Properties of the O.L.S. Estimator
3. Inference in the Linear Regression Model
4. Analysis of Variance, Goodness of Fit and the F test
5. Inference on Prediction

Assumptions

- The validity and properties of least squares estimation depend very much on the validity of the classical assumptions underlying the regression model. As we shall see, many of these assumptions are rarely appropriate when dealing with data for business. However, they represent a useful starting point dealing with the inferential aspects of the regression and for the development of more advanced techniques.
- The assumptions are as follows:

Assumptions

1. The regression model is linear in the unknown parameters.

Assumptions

1. The regression model is linear in the unknown parameters.
2. The elements in X are non-stochastic, meaning that the values of X are fixed in repeated samples (i.e., when repeating the experiment, choose exactly the same set of X values on each occasion so that they remain unchanged).
    - Notice, however, this does not imply that the values of Y also remain unchanged from sample to sample. The Y values depend also on the uncontrollable values of ϵ, which vary from one sample to another. Y as well as ϵ are therefore stochastic, meaning that their values are determined by some chance mechanism and hence subject to a probability distribution.

Assumptions

1. The regression model is linear in the unknown parameters.
2. The elements in X are non-stochastic, meaning that the values of X are fixed in repeated samples (i.e., when repeating the experiment, choose exactly the same set of X values on each occasion so that they remain unchanged).
   - Notice, however, this does not imply that the values of Y also remain unchanged from sample to sample. The Y values depend also on the uncontrollable values of ϵ, which vary from one sample to another. Y as well as ϵ are therefore stochastic, meaning that their values are determined by some chance mechanism and hence subject to a probability distribution.
   - Essentially this means our regression analysis is conditional on the given values of the regressors.
   - It is possible to weaken the assumption to one of stochastic X distributed independently of the disturbance term.

Assumptions

3. Zero mean value of the disturbance ϵ_i, i.e., $E(\epsilon_i) = 0, \forall \ i$, or in matrix terms,

$$\begin{aligned} E(\epsilon) &= E\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \\ &= 0 \end{aligned}$$

leading to

$$E(Y) = X\beta$$

The zero mean of the disturbances implies that no relevant regressors have been omitted from the model.

Assumptions

4. The variance-covariance matrix of ϵ is a scalar matrix. That is,

$$
\begin{aligned}
E(\epsilon\epsilon') &= E(\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} \begin{bmatrix} \epsilon_1 & \epsilon_2 & \cdots & \epsilon_n \end{bmatrix}) \\
&= \begin{bmatrix} E(\epsilon_1^2) & E(\epsilon_1\epsilon_2) & \cdots & E(\epsilon_1\epsilon_n) \\ E(\epsilon_2\epsilon_1) & E(\epsilon_2^2) & \cdots & E(\epsilon_2\epsilon_n) \\ \vdots & \vdots & \ddots & \vdots \\ E(\epsilon_n\epsilon_1) & E(\epsilon_n\epsilon_2) & \cdots & E(\epsilon_n^2) \end{bmatrix} = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} \\
&= \sigma^2 I.
\end{aligned}
$$

Assumptions

This variance covariance matrix embodies two assumptions:

- $var(\epsilon_i) = \sigma^2 \; \forall \; i$. This assumption is termed *homoscedasticity* (the converse is *heteroscedasticity*).

Assumptions

This variance covariance matrix embodies two assumptions:

- $var(\epsilon_i) = \sigma^2 \ \forall \ i$. This assumption is termed *homoscedasticity* (the converse is *heteroscedasticity*).
- $cov(\epsilon_i \epsilon_j) = 0 \ \forall \ i \neq j$. This assumption is termed *pairwise uncorrelatedness* (the coverse is *serial correlation* or *autocorrelation*).

Assumptions

5. $\rho(X) = rank(X) = k < n$. In other words, the explanatory variables do not form a linear dependent set as X is $n \times k$. We say that X has full column rank. If this conditions fails, then $X'X$ cannot be inverted and O.L.S. estimation becomes infeasible. This problem is known as *perfect multicollinearity*.

Assumptions

5. $\rho(X) = rank(X) = k < n$. In other words, the explanatory variables do not form a linear dependent set as X is $n \times k$. We say that X has full column rank. If this conditions fails, then $X'X$ cannot be inverted and O.L.S. estimation becomes infeasible. This problem is known as *perfect multicollinearity*.
6. As $n \to \infty$, $\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2/n \to Q_j$, where Q_j is finite, $j = 1, \cdots, k$.

Properties of O.L.S.

When some or all of the above assumptions are satisfied, the O.L.S. estimator *b* of β possesses the following properties. Note that not every property requires all of the above assumptions to be fulfilled.

Properties of the O.L.S. estimator:

Properties of O.L.S.

When some or all of the above assumptions are satisfied, the O.L.S. estimator b of β possesses the following properties. Note that not every property requires all of the above assumptions to be fulfilled.

Properties of the O.L.S. estimator:

- b is a linear estimator in the sense that it is a linear combination of the observations of Y:

$$\begin{aligned} b &= (X'X)^{-1}X'Y \\ &= \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{k1} & c_{k2} & \cdots & c_{kn} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \end{aligned}$$

Properties of O.L.S.

- Unbiasedness

$$\begin{aligned} E(b) &= E((X'X)^{-1}X'Y) \\ &= E(\beta + (X'X)^{-1}X\epsilon) \\ &= \beta + (X'X)^{-1}X'E(\epsilon) \\ &= \beta \end{aligned}$$

Properties of O.L.S.

- Unbiasedness

$$
\begin{aligned}
E(b) &= E((X'X)^{-1}X'Y) \\
&= E(\beta + (X'X)^{-1}X\epsilon) \\
&= \beta + (X'X)^{-1}X'E(\epsilon) \\
&= \beta
\end{aligned}
$$

Thus, b is an unbiased estimator of β. That is, in repeated samples, b has an average value identical to β, the parameter b tries to estimate.

Properties of O.L.S.

- Variance-Covariance matrix:

$$
\begin{aligned}
COV(b) &= E((b - E(b))(b - E(b))') \\
&= E((b - \beta)(b - \beta)') \\
&= E((X'X)^{-1}X'\epsilon\epsilon'X(X'X)^{-1}) \\
&= (X'X)^{-1}X'E(\epsilon\epsilon')X(X'X)^{-1} \\
&= \sigma^2(X'X)^{-1}
\end{aligned}
$$

Properties of O.L.S.

- Variance-Covariance matrix:

$$\begin{aligned} COV(b) &= E((b-E(b))(b-E(b))') \\ &= E((b-\beta)(b-\beta)') \\ &= E((X'X)^{-1}X'\epsilon\epsilon'X(X'X)^{-1}) \\ &= (X'X)^{-1}X'E(\epsilon\epsilon')X(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1} \end{aligned}$$

Main diagonal elements are the variances of b_j's, $j = 1, \cdots, k$; off-diagonal elements are covariances. For the special case of a simple linear regression,

$$Cov\left(\begin{bmatrix} b_1 \\ b_2 \end{bmatrix}\right) = \begin{bmatrix} \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i-\bar{x})^2}\right) & -\sigma^2\frac{\bar{x}}{\sum_{i=1}^n (x_i-\bar{x})^2} \\ -\sigma^2\frac{\bar{x}}{\sum_{i=1}^n (x_i-\bar{x})^2} & \frac{\sigma^2}{\sum_{i=1}^n (x_i-\bar{x})^2} \end{bmatrix}$$

Properties of O.L.S.

- b is the best linear unbiased (B.L.U.) estimator of β. Refer to the Gauss-Markov theorem. The B.L.U. properties implies that each b_j, $j = 1, \cdots, k$, has the smallest variance among the class of all linear unbiased estimators of β_j. More discussion in class.

Properties of O.L.S.

- b is the best linear unbiased (B.L.U.) estimator of β. Refer to the Gauss-Markov theorem. The B.L.U. properties implies that each b_j, $j = 1, \cdots, k$, has the smallest variance among the class of all linear unbiased estimators of β_j. More discussion in class.
- b is the minimum variance unbiased (M.V.U.) estimator of β, meaning that b_j has a variance no larger than that of any unbiased estimator of β_j, linear or non-linear. More discussion in class.

Properties of O.L.S.

- b is the best linear unbiased (B.L.U.) estimator of β. Refer to the Gauss-Markov theorem. The B.L.U. properties implies that each b_j, $j = 1, \cdots, k$, has the smallest variance among the class of all linear unbiased estimators of β_j. More discussion in class.
- b is the minimum variance unbiased (M.V.U.) estimator of β, meaning that b_j has a variance no larger than that of any unbiased estimator of β_j, linear or non-linear. More discussion in class.
- b is a consistent estimator of β, meaning that when n becomes sufficiently large, the probability of $b_j = \beta_j$ converges to 1, $j = 1, \cdots, k$. We say that b converges in probability to the true value of β. More discussion in class.

Matters of Inference

- If one assumes additionally that $\epsilon \sim MVN(0, \sigma^2 I)$, then
  - $Y \sim MVN(X\beta, \sigma^2 I)$
  - $b \sim MVN(\beta, \sigma^2 (X'X)^{-1})$
- Using properties of the sampling distribution of b, inference about the population parameters in β can be drawn.

Matters of Inference

- However, we need an estimator of σ^2, the variance around the regression line. This estimator is given by

$$s^2 = \frac{e'e}{n-k} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-k},$$

where $n-k$ is the model's degrees of freedom (d.o.f.) - the number of logically independent pieces of information in the data.

Matters of Inference

- However, we need an estimator of σ^2, the variance around the regression line. This estimator is given by

$$s^2 = \frac{e'e}{n-k} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-k},$$

where $n-k$ is the model's degrees of freedom (d.o.f.) - the number of logically independent pieces of information in the data.

- It can be shown that $e'e/\sigma^2 \sim \chi^2_{(n-k)}$, or $(n-k)s^2/\sigma^2 \sim \chi^2_{(n-k)}$.

Matters of Inference

- However, we need an estimator of σ^2, the variance around the regression line. This estimator is given by

$$s^2 = \frac{e'e}{n-k} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-k},$$

where $n-k$ is the model's degrees of freedom (d.o.f.) - the number of logically independent pieces of information in the data.

- It can be shown that $e'e/\sigma^2 \sim \chi^2_{(n-k)}$, or $(n-k)s^2/\sigma^2 \sim \chi^2_{(n-k)}$.
- Using the properties of the Chi-square distribution, it can be shown that $E(s^2) = \sigma^2$, i.e., s^2 is an unbiased estimator of σ^2.

Matters of Inference

- The quantities b_j, $j = 1, \cdots, k$, are simply point estimates (single numbers). Often it is more desirable to state a range of values in which the parameter is thought to lie rather than a single number. These ranges are called *confidence interval (C.I.) estimates*.

Matters of Inference

- The quantities b_j, $j = 1, \cdots, k$, are simply point estimates (single numbers). Often it is more desirable to state a range of values in which the parameter is thought to lie rather than a single number. These ranges are called *confidence interval (C.I.) estimates*.
- Although the interval estimate is less precise, the confidence that the true population parameters falls between the interval limits is increased. The interval should be precise enough to be practically useful.

Matters of Inference

- Consider a coefficient β_j in β. The interval $L \leq \beta_j \leq U$ is a $100(1-\alpha)$ % confidence interval for β_j in the sense that, prior to sampling,

$$P(L \leq \beta_j \leq U) = 1 - \alpha$$

- This definition states that the C.I. with confidence coefficient $1-\alpha$ is an interval estimate such that the probability is $1-\alpha$ that the calculated limits include β_j for any random trial. That is, in many random samples of size n, $100(1-\alpha)$ percent of the interval estimates will include β_j.

Matters of Inference

- Recall that $b \sim N(\beta, \sigma^2(X'X)^{-1})$. Hence $\frac{b_j - \beta_j}{\sigma c_{jj}} \sim N(0,1)$, where c_{jj}^2 is the jj^{th} element of $(X'X)^{-1}$.

Matters of Inference

- Recall that $b \sim N(\beta, \sigma^2(X'X)^{-1})$. Hence $\frac{b_j - \beta_j}{\sigma c_{jj}} \sim N(0,1)$, where c_{jj}^2 is the jj^{th} element of $(X'X)^{-1}$.
- Hence

$$P(z_{(\alpha/2)} \leq \frac{b_j - \beta_j}{\sigma c_{jj}} \leq z_{(1-\alpha/2)}) = 1 - \alpha$$

- Recognising that $z(\alpha/2) = -z(1-\alpha/2)$ and after some manipulations, we can write

$$P(b_j - z_{(1-\alpha/2)}\sigma c_{jj} \leq \beta_j \leq b_j + z_{(1-\alpha/2)}\sigma c_{jj}) = 1 - \alpha$$

Matters of Inference

- However, σ^2 is typically unknown. Replacing σ^2 by s^2 results in $\frac{b_j - \beta_j}{sc_{jj}} \sim t_{(n-k)}$.

Matters of Inference

- However, σ^2 is typically unknown. Replacing σ^2 by s^2 results in $\frac{b_j - \beta_j}{s c_{jj}} \sim t_{(n-k)}$.
- If $Z \sim N(0,1)$ and $W \sim \chi^2_{(n-k)}$ and Z and W are independently distributed, then $\frac{Z}{\sqrt{W/(n-k)}} \sim t_{(n-k)}$. More discussion on t distribution in class.

Matters of Inference

- However, σ^2 is typically unknown. Replacing σ^2 by s^2 results in $\frac{b_j-\beta_j}{sc_{jj}} \sim t_{(n-k)}$.
- If $Z \sim N(0,1)$ and $W \sim \chi^2_{(n-k)}$ and Z and W are independently distributed, then $\frac{Z}{\sqrt{W/(n-k)}} \sim t_{(n-k)}$. More discussion on t distribution in class.
- Hence the confidence interval becomes

$$P(b_j - t_{(1-\alpha/2,n-k)}sc_{jj} \leq \beta_j \leq b_j + t_{(1-\alpha/2,n-k)}sc_{jj}) = 1-\alpha$$

Matters of Inference

- Hypothesis tests about β_j can also be performed. The most common test about a coefficient in a regression is:

$$H_0 : \beta_j = \beta_j^* \quad \text{vs.} \quad H_1 : \beta_j \neq \beta_j^*$$

at a significance level α, the probability of rejecting H_0 when H_0 is correct, the so-called Type I error.

Matters of Inference

- Hypothesis tests about β_j can also be performed. The most common test about a coefficient in a regression is:

$$H_0 : \beta_j = \beta_j^* \quad \text{vs.} \quad H_1 : \beta_j \neq \beta_j^*$$

  at a significance level α, the probability of rejecting H_0 when H_0 is correct, the so-called Type I error.

- To test this hypothesis, a t statistic is used:

$$t = \frac{b_j - \beta_j^*}{sc_{jj}}$$

  If H_0 is true then t has a t distribution with $n - k$ degrees of freedom.

Matters of Inference

- If H_0 is true, then t is expected to lie not too far from the centre of the distribution.
- The decision rule is:

  Reject H_0 if $t > t_{(1-\alpha/2,n-k)}$ or $t < -t_{(1-\alpha/2,n-k)}$
  Do not reject otherwise.

Matters of Inference

- If H_0 is true, then t is expected to lie not too far from the centre of the distribution.
- The decision rule is:

  Reject H_0 if $t > t_{(1-\alpha/2, n-k)}$ or $t < -t_{(1-\alpha/2, n-k)}$
  Do not reject otherwise.
- Testing this hypothesis is equivalent to asking if β_j^* lies in the $100(1-\alpha)$ percent C.I. of β_j.
- It is a common practice to test the hypothesis of $H_0 : \beta_j = 0$. Failure to reject this hypothesis would imply that β_j is not significantly different from zero, or equivalently, X_j has no significant impact on the behaviour of Y, at level of significance α.

Matters of Inference

- Return to Example 1.3, and consider the estimation of β_2.
- Back in Chapter 1, we already computed $b_2 = 0.332$. The output shows that $e'e = \sum_{i=1}^{25} e_i^2 = 32501.95754$. Note that $d.o.f. = 25 - 3 = 22$. Hence $s^2 = e'e/22 = 1477.362$.

Matters of Inference

- Return to Example 1.3, and consider the estimation of β_2.
- Back in Chapter 1, we already computed $b_2 = 0.332$. The output shows that $e'e = \sum_{i=1}^{25} e_i^2 = 32501.95754$. Note that $d.o.f. = 25 - 3 = 22$. Hence $s^2 = e'e/22 = 1477.362$.
- From $(X'X)^{-1}$, $c_{22}^2 = 0.000020048$. Hence $s.e.(b_2) = \sqrt{1477.362 \times 0.000020048} = 0.1721$.

Matters of Inference

- Return to Example 1.3, and consider the estimation of β_2.
- Back in Chapter 1, we already computed $b_2 = 0.332$. The output shows that $e'e = \sum_{i=1}^{25} e_i^2 = 32501.95754$. Note that $d.o.f. = 25 - 3 = 22$. Hence $s^2 = e'e/22 = 1477.362$.
- From $(X'X)^{-1}$, $c_{22}^2 = 0.000020048$. Hence $s.e.(b_2) = \sqrt{1477.362 \times 0.000020048} = 0.1721$.
- Set $\alpha = 0.05$. From the t distribution table, $t_{(1-0.05/2,22)} = 2.074$. Hence the 95 percent C.I. for β_2 is

$$0.3318 - (2.074)(0.1721) \leq \beta_2 \leq 0.3318 + (2.074)(0.1721)$$

or

$$-0.0251 \leq \beta_2 \leq 0.6887$$

Matters of Inference

- This C.I. contains 0, meaning that if we test $H_0 : \beta_2 = 0$ vs. $H_1 : \beta_2 \neq 0$, we would not be able to reject H_0 at $\alpha = 0.05$. This is indeed the case.

Matters of Inference

- This C.I. contains 0, meaning that if we test $H_0 : \beta_2 = 0$ vs. $H_1 : \beta_2 \neq 0$, we would not be able to reject H_0 at $\alpha = 0.05$. This is indeed the case.
- Note that for testing H_0, $t = \frac{0.33183}{0.1721} = 1.928$, which lies to the left of $t_{(1-0.05/2,22)} = 2.074$. Hence H_0 cannot be rejected at $\alpha = 0.05$.

Matters of Inference

- This C.I. contains 0, meaning that if we test $H_0 : \beta_2 = 0$ vs. $H_1 : \beta_2 \neq 0$, we would not be able to reject H_0 at $\alpha = 0.05$. This is indeed the case.
- Note that for testing H_0, $t = \frac{0.33183}{0.1721} = 1.928$, which lies to the left of $t_{(1-0.05/2,22)} = 2.074$. Hence H_0 cannot be rejected at $\alpha = 0.05$.
- Alternatively, one can base the decision on the p-value, which is the probability of obtaining a value of t at least as extreme as the actual computed value if H_0 is true. In our example, the p-value is 0.0668, meaning that $P(t > 1.928 \text{ or } t < -1.928) = 0.0668$.

Matters of Inference

- The $p-$value can be viewed as the minimum level of significance chosen for the test to result in a rejection of H_0. Thus, a decision rule using p-value may be stated as:

  Reject H_0 if p-value$< \alpha$
  Do not reject H_0 if p-value$\geq \alpha$

Matters of Inference

- The $p-$value can be viewed as the minimum level of significance chosen for the test to result in a rejection of H_0. Thus, a decision rule using p-value may be stated as:

  Reject H_0 if p-value$< \alpha$
  Do not reject H_0 if p-value$\geq \alpha$

- In the last example, the p-value is 0.0668. Hence we cannot reject H_0 at $\alpha = 0.05$. On the other hand, H_0 can be rejected at any significance level at or higher than 0.0668.

Matters of Inference

- The $p-$value can be viewed as the minimum level of significance chosen for the test to result in a rejection of H_0. Thus, a decision rule using p-value may be stated as:

  Reject H_0 if p-value$< \alpha$
  Do not reject H_0 if p-value$\geq \alpha$
- In the last example, the p-value is 0.0668. Hence we cannot reject H_0 at $\alpha = 0.05$. On the other hand, H_0 can be rejected at any significance level at or higher than 0.0668.
- Similarly, we can test $H_0 : \beta_3 = 0$ and conclude that H_0 is rejected at $\alpha = 0.05$.

Matters of Inference

- Altogether, it means allowing for a 5% Type 1 risk, disposable income is not significant for explaining consumption but the total value of assets is significant.
- Note that if we conclude that $\beta_j = 0$, it does not necessarily follow that X_j is unrelated to Y. It simply means that, when the other explanatory variables are included in the model, the marginal contribution of X_j to further improving the model's fit is negligible.

Matters of Inference

- Altogether, it means allowing for a 5% Type 1 risk, disposable income is not significant for explaining consumption but the total value of assets is significant.
- Note that if we conclude that $\beta_j = 0$, it does not necessarily follow that X_j is unrelated to Y. It simply means that, when the other explanatory variables are included in the model, the marginal contribution of X_j to further improving the model's fit is negligible.
- Sometimes it also makes sense to conduct a hypothesis test for the intercept coefficient. This should be done only when there are data that span $X = 0$ or at least near $X = 0$, and the difference between Y equaling zero and not equaling zero when $X = 0$ is scientifically plausible and interesting.

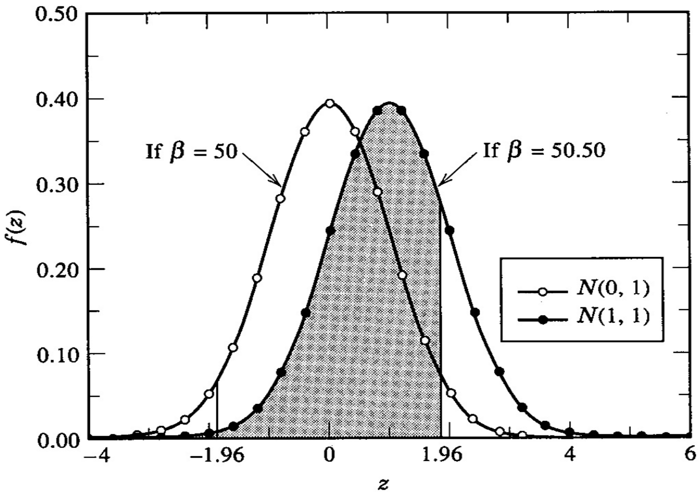
Type 1 and Type 2 errors

- Rejecting H_0 when it is true is called a Type 1 error. Recall that if H_0 is true the probability that it will be (incorrectly) rejected is $P(t > t_{(1-\alpha/2,n-k)}) + P(t < -t_{(1-\alpha/2,n-k)}) = \alpha$. This is the significance level; by choosing α, we effectively determine the probability that the test will incorrectly reject a true hypothesis.

Type 1 and Type 2 errors

- Rejecting H_0 when it is true is called a Type 1 error. Recall that if H_0 is true the probability that it will be (incorrectly) rejected is $P(t > t_{(1-\alpha/2,n-k)}) + P(t < -t_{(1-\alpha/2,n-k)}) = \alpha$. This is the significance level; by choosing α, we effectively determine the probability that the test will incorrectly reject a true hypothesis.
- If H_0 is false and it is not rejected then a Type 2 error has been committed. While we can fix $P(\text{Type 1 error})$, the same control of Type 2 error is not possible. See the following diagram for an illustration for testing $H_0 : \beta_2 = 50$ with σ^2 known and $var(b_2) = 0.25$. Suppose that β_2 is either 50 or 50.5. Note that Type 2 error probability depends on the true value of β_2 which is unknown in practice.

Type 1 and Type 2 errors



Type 1 and Type 2 errors

- Most elementary texts define the "power" of the test as the probability of rejecting a false H_0, i.e., the probability of doing the right thing in the face of an incorrect H_0. By this definition, the power is equal to 1 minus the Type 2 error probability.
- Sometimes the power is simply defined as the probability of rejecting H_0. By this definition, α, the significance level, is a point on the power curve.

Type 1 and Type 2 errors

- Let $H_0 : \beta_2 = \beta_2^*$.
  $P(\text{Type 2 error}|\beta_2 \neq \beta_2^*) = P(\text{Not rejecting } H_0|\beta_2 \neq \beta_2^*)$.
- $\text{Power}(\beta_2) = P(\text{rejecting } H_0|\beta_2)$
- A test is "unbiased" if
  $\text{Power}(\beta_2|\beta_2 \neq \beta_2^*) \geq P(\text{Type 1 error})$.
- For a test where H_0 corresponds to a point in the parameter space (e.g., a two-sided t test), the significance level is a point on the power curve.
- For a test where H_0 corresponds to a region in the parameter space (e.g., a one-sided t test), the significance level is the maximum probability of committing a Type 1 error within the region defined by H_0, and $P(\text{Type 1 error})$ has a range of values with α being the maximum of the range.

Partitioning of Total Sum of Squares

- Analysis of variance (ANOVA) is a useful and flexible way of analysing the fit of the regression. To motivate, consider

$$\begin{aligned} y_i &= \hat{y}_i + e_i \\ y_i - \bar{y} &= \hat{y}_i - \bar{y} + e_i \\ \sum_{i=1}^{n}(y_i - \bar{y})^2 &= \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n} e_i^2 + 2\sum_{i=1}^{n}(\hat{y}_i - \bar{y})e_i \end{aligned}$$

Partitioning of Total Sum of Squares

- Note that

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \text{Total Sum of Squares (TSS)}$$

$$\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 = \text{Regression Sum of Squares (RSS)}$$

$$\sum_{i=1}^{n} e_i^2 = \text{Error Sum of Squares (ESS)}$$

$$\sum_{i=1}^{n}(\hat{y}_i - \bar{y})e_i = 0 \text{ (provided that there is an intercept)}$$

Coefficient of Determination

- Thus,

$$TSS = RSS + ESS$$

or

$$\begin{aligned} R^2 &= \frac{RSS}{TSS} \\ &= 1 - \frac{ESS}{TSS}, \end{aligned}$$

which is the coefficient of determination. It measures the model's "goodness of fit": the proportion of variability of the sample Y values that has been explained by the regression.

- Obviously, $0 \leq R^2 \leq 1$.

Partitioning of Degrees of Freedom

- TSS has $n-1$ d.o.f. because there are n deviations $y_i - \bar{y}$ that enter into TSS, but one constraint on the deviations, namely, $\sum_{i=1}^{n}(y_i - \bar{y}) = 0$. So there are $n-1$ d.o.f. in the deviations.

Partitioning of Degrees of Freedom

- TSS has $n-1$ d.o.f. because there are n deviations $y_i - \bar{y}$ that enter into TSS, but one constraint on the deviations, namely, $\sum_{i=1}^{n}(y_i - \bar{y}) = 0$. So there are $n-1$ d.o.f. in the deviations.
- ESS has $n-k$ d.o.f. because there are n residuals but k d.o.f. are lost due to k constraints on the e_i's associated with estimating the β's.

Partitioning of Degrees of Freedom

- TSS has $n-1$ d.o.f. because there are n deviations $y_i - \bar{y}$ that enter into TSS, but one constraint on the deviations, namely, $\sum_{i=1}^{n}(y_i - \bar{y}) = 0$. So there are $n-1$ d.o.f. in the deviations.
- ESS has $n-k$ d.o.f. because there are n residuals but k d.o.f. are lost due to k constraints on the e_i's associated with estimating the β's.
- RSS has $k-1$ d.o.f. because the regression function contains k parameters but the deviations $\hat{y}_i - \bar{y}$ are subject to the constraint that $\sum_{i=1}^{n}(\hat{y}_i - \bar{y}) = 0$.

Partitioning of Degrees of Freedom

- TSS has $n-1$ d.o.f. because there are n deviations $y_i - \bar{y}$ that enter into TSS, but one constraint on the deviations, namely, $\sum_{i=1}^{n}(y_i - \bar{y}) = 0$. So there are $n-1$ d.o.f. in the deviations.
- ESS has $n-k$ d.o.f. because there are n residuals but k d.o.f. are lost due to k constraints on the e_i's associated with estimating the β's.
- RSS has $k-1$ d.o.f. because the regression function contains k parameters but the deviations $\hat{y}_i - \bar{y}$ are subject to the constraint that $\sum_{i=1}^{n}(\hat{y}_i - \bar{y}) = 0$.
- The d.o.f. add up: $(n-1) = (n-k) + (k-1)$

Mean Squares

- A sum of squares divided by its associated d.o.f. is called a mean square.
- Mean Square Regression (MSR) $= RSS/(k-1)$
- Mean Square Error (MSE) $= ESS/(n-k)$

Mean Squares

- A sum of squares divided by its associated d.o.f. is called a mean square.
- Mean Square Regression (MSR) $= RSS/(k-1)$
- Mean Square Error (MSE) $= ESS/(n-k)$
- In Example 1.3, n=25, k=3, RSS=126186.66, ESS=32501.96, TSS=158688.61. Hence

$$\begin{aligned} R^2 &= 126186.66/158688.61 = 0.7952 \\ MSR &= 126186.66/2 = 63093.33 \\ MSE &= 32501.96/22 = 1477.362 \end{aligned}$$

Overall Significance of the Model

- Frequently, one may wish to test whether or not there is a relationship between Y and the regression model constructed. It is a test of

$$\begin{aligned} H_0 &: \beta_2 = \beta_3 = \cdots = \beta_k = 0 \qquad \text{vs.} \\ H_1 &: \text{at least one of } \beta_j' s, \ (j = 2, \cdots, k), \ \text{is non-zero.} \end{aligned}$$

- The test statistic is

$$F = \frac{MSR}{MSE} = \frac{RSS/(k-1)}{ESS/(n-k)}$$

distributed as $F_{(k-1,n-k)}$ if H_0 is true.

Overall Significance of the Model

- The decision rule is:
  To reject H_0 if $F > F_{1-\alpha,k-1,n-k}$ or p-value$< \alpha$;
  Not to reject H_0 otherwise.

Overall Significance of the Model

- The decision rule is:
  To reject H_0 if $F > F_{1-\alpha,k-1,n-k}$ or p-value$< \alpha$;
  Not to reject H_0 otherwise.
- $F \sim F_{(k-1,n-k)}$ because under H_0, $RSS/\sigma^2 \sim \chi^2_{(k-1)}$, $ESS/\sigma^2 \sim \chi^2_{(n-k)}$, and RSS and ESS are distributed independently.

Overall Significance of the Model

- The decision rule is:
  To reject H_0 if $F > F_{1-\alpha,k-1,n-k}$ or p-value$< \alpha$;
  Not to reject H_0 otherwise.
- $F \sim F_{(k-1,n-k)}$ because under H_0, $RSS/\sigma^2 \sim \chi^2_{(k-1)}$, $ESS/\sigma^2 \sim \chi^2_{(n-k)}$, and RSS and ESS are distributed independently.
- Refer to Example 1.3, $F = \frac{63093.33}{1477.362} = 42.70676$. $F_{(0.95,2,22)} = 3.44$. Hence we reject H_0 convincingly at significance level 0.05. We cannot reject H_0 only if α is set to 2.66073E-08 or lower, as indicated by the test statistic's p-value.

Overall Significance of the Model

Why do we perform an F test in addition to t tests? What can we learn from the F test?

Overall Significance of the Model

Why do we perform an F test in addition to t tests? What can we learn from the F test?

- In the intercept only model, all of the fitted values equal the mean of the response variable. Therefore, if the overall F test is significant, the regression model predicts the response better than than the mean of the response.

Overall Significance of the Model

Why do we perform an F test in addition to t tests? What can we learn from the F test?

- In the intercept only model, all of the fitted values equal the mean of the response variable. Therefore, if the overall F test is significant, the regression model predicts the response better than than the mean of the response.
- While R^2 provides an estimate of the strength of the relationship, it does not provide a formal hypothesis test for this relationship. If the overall F test is significant, one can conclude that the R^2 is significantly different from zero. In fact, the F statistic can be written as

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

Overall Significance of the Model

- If the overall F test is significant, but few or none of the t tests are significant then it is an indication that *multicollinearity* might be a problem for the data. More on multicollinearity in Chapter 3.

Overall Significance of the Model

- If the overall F test is significant, but few or none of the t tests are significant then it is an indication that *multicollinearity* might be a problem for the data. More on multicollinearity in Chapter 3.

Notice that for a simple linear regression model, the null hypothesis for the overall F test is simply $\beta_2 = 0$, which is precisely the same null for the t test of $\beta_2 = 0$. In fact, when $k = 1$, $F_{(1,n-k)} = t^2_{(n-k)}$.

- In Example 1.1, for testing $H_0 : \beta_2 = 0$, $F = 94.41 = (9.717)^2 = t^2$, p-values are exactly the same.

Overall Significance of the Model

- If the overall F test is significant, but few or none of the t tests are significant then it is an indication that *multicollinearity* might be a problem for the data. More on multicollinearity in Chapter 3.

Notice that for a simple linear regression model, the null hypothesis for the overall F test is simply $\beta_2 = 0$, which is precisely the same null for the t test of $\beta_2 = 0$. In fact, when $k = 1$, $F_{(1,n-k)} = t^2_{(n-k)}$.

- In Example 1.1, for testing $H_0 : \beta_2 = 0$, $F = 94.41 = (9.717)^2 = t^2$, p-values are exactly the same.
- In Example 1.2, for testing $H_0 : \beta_2 = 0$, $F = 194.252 = (13.937)^2 = t^2$, p-values are exactly the same.

F test for linear restrictions

- In fact, the usefulness of the F test is not limited to testing overall significance. The F test can be used for testing any linear equality restrictions on β.

F test for linear restrictions

- In fact, the usefulness of the F test is not limited to testing overall significance. The F test can be used for testing any linear equality restrictions on β.
- The general formula for the F statistic is

$$\begin{aligned} F &= \frac{(e'e_r - e'e_{ur})/m}{e'e_{ur}/(n-k)} \\ &= \frac{(R^2_{ur} - R^2_r)/m}{(1 - R^2_{ur})/(n-k)} \sim F_{(m,n-k)}|H_0, \end{aligned}$$

where the subscripts ur and r correspond to the unrestricted and restricted models respectively, and m is the number of restrictions under H_0.

F test for linear restrictions

- $e'e_r$ is the ESS associated with the restricted model (i.e., the model that imposes the restrictions implied by H_0; $e'e_{ur}$ is the ESS associated with the unrestricted model (i.e., the model that ignores the restrictions). R_r^2 and R_{ur}^2 are defined analogously.

F test for linear restrictions

- $e'e_r$ is the ESS associated with the restricted model (i.e., the model that imposes the restrictions implied by H_0; $e'e_{ur}$ is the ESS associated with the unrestricted model (i.e., the model that ignores the restrictions). R_r^2 and R_{ur}^2 are defined analogously.
- The F statistic for testing $H_0 : \beta_2 = \beta_3 = \cdots = \beta_k = 0$ is a special case of (1), because under H_0, $m = k - 1$, $e'e_r = TSS$ (the restricted model has no explanatory power) and correspondingly, $R_r^2 = 0$.

F test for linear restrictions

- **Example 2.1** One model of production that is widely used in economics is the Cobb-Douglas production function:

$$y_i = \beta_1^* x_{2i}^{\beta_2} x_{3i}^{\beta_3} exp(\epsilon_i),$$

where y_i=output; x_{2i}=labour input; x_{3i}=capital input.

- Or, in log-transformed terms,

$$\begin{aligned} lny_i &= ln\beta_1^* + \beta_2 lnx_{2i} + \beta_3 lnx_{3i} + \epsilon_i, \\ &= \beta_1 + \beta_2 lnx_{2i} + \beta_3 lnx_{3i} + \epsilon_i, \end{aligned}$$

F test for linear restrictions

- To illustrate, we use annual data for the agricultural sector of Taiwan for 1958-1972.
- Results obtained using SAS:

```
                              The REG Procedure
                                Model: MODEL1
                         Dependent Variable: lny

                 Number of Observations Read          15
                 Number of Observations Used          15


                            Analysis of Variance

                                    Sum of           Mean
Source                   DF        Squares         Square    F Value    Pr > F

Model                     2        0.53804        0.26902      48.07    <.0001
Error                    12        0.06716        0.00560
Corrected Total          14        0.60520


             Root MSE              0.07481    R-Square     0.8890
             Dependent Mean       10.09654    Adj R-Sq     0.8705
             Coeff Var             0.74095


                            Parameter Estimates

                         Parameter       Standard
     Variable     DF      Estimate          Error    t Value    Pr > |t|

     Intercept     1      -3.33846        2.44950      -1.36      0.1979
     lnx2          1       1.49876        0.53980       2.78      0.0168
     lnx3          1       0.48986        0.10204       4.80      0.0004
```

F test for linear restrictions

- β_2 is the elasticity of output with respect to the labour input; it measures the percentage change in output due to a one percent change in labour input; β_3 is interpreted analogously.

F test for linear restrictions

- β_2 is the elasticity of output with respect to the labour input; it measures the percentage change in output due to a one percent change in labour input; β_3 is interpreted analogously.
- The sum $\beta_2 + \beta_3$ gives information on *returns to scale*, that is, the response of output to a proportional change in the inputs. In particular, if this sum is 1, then there are constant returns to scale, that is, doubling the inputs will double the outputs.
- Hence one may be interested in testing $H_0 : \beta_2 + \beta_3 = 1$.

F test for linear restrictions

- The restricted model is one that imposes the restriction $\beta_2 + \beta_3 = 1$ onto the coefficients in the minimisation of the SSE. The least squares estimator (referred to as restricted least squares (R.L.S.)) is obtained by minimising the objective function

$$\phi = (y - Xb_*)'(y - Xb_*) - 2\lambda'(Rb_* - r),$$

where R is a $m \times k$ matrix of constants, r is a $m \times 1$ vector of constants and b_* is the R.L.S. estimator.

F test for linear restrictions

- The restricted model is one that imposes the restriction $\beta_2 + \beta_3 = 1$ onto the coefficients in the minimisation of the SSE. The least squares estimator (referred to as restricted least squares (R.L.S.)) is obtained by minimising the objective function

$$\phi = (y - Xb_*)'(y - Xb_*) - 2\lambda'(Rb_* - r),$$

where R is a $m \times k$ matrix of constants, r is a $m \times 1$ vector of constants and b_* is the R.L.S. estimator.

- For this example, $R = [0 \;\; 1 \;\; 1]$ and $r = 1$.

F test for linear restrictions

- The SAS results are as follows:

The REG Procedure
Model: MODEL1
Dependent Variable: lny

NOTE: Restrictions have been applied to parameter estimates.

Number of Observations Read	15
Number of Observations Used	15

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.51372	0.51372	73.01	<.0001
Error	13	0.09147	0.00704		
Corrected Total	14	0.60520			

Root MSE	0.08388	R-Square	0.8489
Dependent Mean	10.09654	Adj R-Sq	0.8372
Coeff Var	0.83082		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	1.70856	0.41588	4.11	0.0012
lnx2	1	0.38702	0.09330	4.15	0.0011
lnx3	1	0.61298	0.09330	6.57	<.0001

F test for linear restrictions

- Using the F test procedure, to test
  $H_0 : \beta_2 + \beta_3 = 1 \quad \text{vs.} \quad H_1 : otherwise$

$$
\begin{aligned}
F &= \frac{(R^2_{ur} - R^2_r)/m}{(1 - R^2_{ur})/(n-k)} \\
&= \frac{(0.8890 - 0.8489)/1}{(1 - 0.8890)/12} \\
&= 4.34
\end{aligned}
$$

- At $\alpha = 0.05$, $F_{(0.95,1,12)} = 4.75$. Hence we cannot reject H_0 at 0.05 level of significance and conclude that the returns to scale is constant.

F test for linear restrictions

- SAS can perform the test automatically. The result from the following output concurs with the result based on our calculations. The p-value indicates that H_0 can be rejected only when α is set to at least 0.0592.

The REG Procedure
Model: MODEL1

Test 1 Results for Dependent Variable lny

Source	DF	Mean Square	F Value	Pr > F
Numerator	1	0.02432	4.34	0.0592
Denominator	12	0.00560		

Coefficient of variation

- The coefficient variation is obtained by dividing the standard error of the regression by the mean of y_i values and multiplying by 100.

Coefficient of variation

- The coefficient variation is obtained by dividing the standard error of the regression by the mean of y_i values and multiplying by 100.
- It expresses the standard error of the regression in unit free values. Thus the coefficients of variation for two different regressions can be compared more readily than the standard errors because the influence of the units of the data has been removed.
- The SAS program for Example 2.1 is as follows.

SAS Program for Example 2.1

```
data example21;
input y x2 x3;
ods html close;
ods listing;
lny=log(y);
lnx2=log(x2);
lnx3=log(x3);
cards;
16607.7 275.5 17803.7
17511.3 274.4 18096.8
20171.2 269.7 18271.8
20932.9 267.0 19167.3
20406 267.8 19647.6
20831.6 275 20803.5
24806.3 283 22076.6
26465.8 300.7 23445.2
27403 307.5 24939
28628.7 303.7 26713.7
29904.5 304.7 29957.8
27508.2 298.6 31585.9
29035.8 295.5 33474.5
29281.5 299.0 34821.8
31535.8 288.1 41794.3
;
proc reg data=example21;
model lny=lnx2 lnx3;
test lnx2+lnx3=1;
run;
proc reg data=example21;
model lny=lnx2 lnx3;
restrict lnx2+lnx3=1;
run;
ods html close;
ods html;
run;
```

Adjusted Coefficient of Determination

- Some statisticians have suggested to modify R^2 to recognise the number of independent variables in the model. The reason is that R^2 can generally be made larger if additional explanatory variables are added to the model. A measure that recognises the number of explanatory variables in the model is called the adjusted coefficient of determination:

$$R_a^2 = 1 - \frac{ESS/(n-k)}{TSS/(n-1)}$$

Adjusted Coefficient of Determination

- Some statisticians have suggested to modify R^2 to recognise the number of independent variables in the model. The reason is that R^2 can generally be made larger if additional explanatory variables are added to the model. A measure that recognises the number of explanatory variables in the model is called the adjusted coefficient of determination:

$$R_a^2 = 1 - \frac{ESS/(n-k)}{TSS/(n-1)}$$

- For the unrestricted model of Example 2.1,

$$R_a^2 = 1 - \frac{0.06716/12}{0.6052/14} = 0.8705$$

- Hence the adjustment has only a small effect, as R_a^2 is almost the same as R^2.

Inference on Prediction

- A point prediction is obtained by inserting the given X values into the regression equation, giving

$$\hat{y}_f = b_1 + b_2 x_{f2} + b_3 x_{f3} + \cdots + b_k x_{fk}$$

- Let $g' = (1, x_{f2}, x_{f3}, \cdots, x_{fk})$. Then $\hat{y}_f = g'b$. Note that $var(g'b) = g'var(b)g$. If we assume normality for the disturbance term, it follows that

$$\frac{g'b - g'\beta}{\sqrt{var(g'b)}} \sim N(0,1)$$

Inference on Prediction

- When the unknown σ^2 in $var(b)$ is replaced by s^2, the usual shift to the t distribution occurs, giving

$$\frac{\hat{y}_f - E(y_f)}{s\sqrt{g'(X'X)^{-1}g}} \sim t_{(n-k)},$$

from which a $100(1-\alpha)$ percent confidence (or prediction) interval for $E(y_f)$ is

$$\hat{y}_f \pm t_{(1-\alpha/2, n-k)} s\sqrt{g'(X'X)^{-1}g} \tag{1}$$

Inference on Prediction

- Returning to Example 1.3, the estimated regression equation is:

  $$\hat{y}_i = 36.79 + 0.3318x_{i2} + 0.1258x_{i3}$$

  A family with annual disposable income of $50,000 and liquid assets worth $100,000 is predicted to spend

  $$\begin{aligned} \hat{y}_f &= 36.79 + 0.3318(50) + 0.1258(100) \\ &= 65.96 \end{aligned}$$

  thousand dollars on non-durable goods and services in a year.

Inference on Prediction

- For this example,

$$(X'X)^{-1} = \begin{bmatrix} 0.202454971 & -0.001159287 & 0.000046500 \\ -0.001159287 & 0.000020048 & -0.000003673 \\ 0.000046500 & -0.000003673 & 0.000000961 \end{bmatrix}$$

- Hence $g'(X'X)^{-1}g =$

$$\begin{bmatrix} 1 & 50 & 100 \end{bmatrix} \begin{bmatrix} 0.202454971 & -0.001159287 & 0.000046500 \\ -0.001159287 & 0.000020048 & -0.000003673 \\ 0.000046500 & -0.000003673 & 0.000000961 \end{bmatrix} \begin{bmatrix} 1 \\ 50 \\ 100 \end{bmatrix}$$

$$= 0.1188$$

Inference on Prediction

- $s = 38.436$ and $t_{(0.975,22)} = 2.074$. Thus, the 95% prediction interval for $E(y_f)$ is

$$65.96 \pm 2.074(38.436)\sqrt{0.1188}$$

or 38.484 to 93.436

Inference on Prediction

- $s = 38.436$ and $t_{(0.975,22)} = 2.074$. Thus, the 95% prediction interval for $E(y_f)$ is

$$65.96 \pm 2.074(38.436)\sqrt{0.1188}$$

$$\text{or} \qquad 38.484 \text{ to } 93.436$$

- Sometimes one may wish to obtain a prediction interval for y_f rather than $E(y_f)$. The two differ only by the disturbance term ϵ_f, which is unpredictable with a mean of 0, so the point prediction remains the same.

Inference on Prediction

- $s = 38.436$ and $t_{(0.975,22)} = 2.074$. Thus, the 95% prediction interval for $E(y_f)$ is

$$65.96 \pm 2.074(38.436)\sqrt{0.1188}$$

$$\text{or} \qquad 38.484 \text{ to } 93.436$$

- Sometimes one may wish to obtain a prediction interval for y_f rather than $E(y_f)$. The two differ only by the disturbance term ϵ_f, which is unpredictable with a mean of 0, so the point prediction remains the same.
- However, the uncertainty of the prediction increases due to the presence of ϵ_f. Now, $y_f = g'\beta + \epsilon_f$. Therefore,

Inference on Prediction

- $e_f = y_f - \hat{y}_f = \epsilon_f - g'(b - \beta)$.
- Squaring both sides and taking expectations gives

$$\begin{aligned} var(e_f) &= \sigma^2 + g' var(b) g \\ &= \sigma^2(1 + g'(X'X)^{-1}g) \end{aligned}$$

from which we can derive the following t statistic:

$$\frac{\hat{y}_f - y_f}{s\sqrt{1 + g'(X'X)^{-1}g}} \sim t_{(n-k)}$$

Inference on Prediction

- which leads to the $100(1-\alpha)$ percent confidence interval for y_f:

$$\hat{y}_f \pm t_{(1-\alpha/2, n-k)} s \sqrt{1 + g'(X'X)^{-1} g}$$

Inference on Prediction

- which leads to the $100(1-\alpha)$ percent confidence interval for y_f:

$$\hat{y}_f \pm t_{(1-\alpha/2,n-k)} s\sqrt{1+g'(X'X)^{-1}g}$$

- Comparison with (1) shows that the only difference is an increase of 1 inside the square root term. Thus, for the data in Example 1.3, the prediction interval for y_f is:

$$65.96 \pm 2.074(38.436)\sqrt{1+0.1188}$$

or -18.359 to 150.279

- One can obtain these outputs directly using SAS by adding the following options to PROC REG:
/p CLM CLI;

Inference on Prediction

The REG Procedure
Model: MODEL1
Dependent Variable: y

Output Statistics

Obs	Dependent Variable	Predicted Value	Std Error Mean Predict	95% CL Mean		95% CL Predict		Residual
1	52.3000	62.0385	14.2215	32.5448	91.5322	-22.9553	147.0322	-9.7385
2	78.4400	55.5902	13.9639	26.6309	84.5495	-29.2196	140.4000	22.8498
3	88.7600	86.9782	12.9449	60.1321	113.8243	2.8666	171.0899	1.7818
4	54.0800	84.5425	11.8498	59.9675	109.1175	1.1279	167.9571	-30.4625
5	111.4400	79.5205	11.8869	54.8686	104.1724	-3.9168	162.9578	31.9195
6	105.2000	123.6442	12.5496	97.6180	149.6704	39.7906	207.4978	-18.4442
7	45.7300	97.8786	10.4611	76.1835	119.5736	15.2666	180.4905	-52.1486
8	122.3500	118.8644	9.6670	98.8162	138.9126	36.6696	201.0592	3.4856
9	142.2400	97.9041	11.8335	73.3629	122.4453	14.4995	181.3087	44.3359
10	86.2200	123.5498	8.8086	105.2818	141.8177	41.7709	205.3286	-37.3298
11	174.5000	158.9706	9.8584	138.5256	179.4156	76.6781	241.2631	15.5294
12	185.2000	143.1395	8.0145	126.5185	159.7606	61.7128	224.5663	42.0605
13	111.8000	144.3668	8.6545	126.4184	162.3152	62.6588	226.0748	-32.5668
14	214.6000	168.0892	7.8309	151.8489	184.3295	86.7393	249.4391	46.5108
15	144.6000	174.1641	7.9982	157.5769	190.7514	92.7443	255.5840	-29.5641
16	174.3600	222.2363	12.2390	196.8542	247.6183	138.5804	305.8922	-47.8763
17	215.4000	179.6848	10.6226	157.6550	201.7147	96.9843	262.3853	35.7152
18	286.2400	239.1630	13.2191	211.7482	266.5778	154.8681	323.4579	47.0770
19	188.5600	184.3890	12.6369	158.1818	210.5962	100.4791	268.2989	4.1710
20	237.2000	232.0104	10.7190	209.7806	254.2402	149.2564	314.7644	5.1896
21	181.8000	225.9031	11.5935	201.8597	249.9465	142.6436	309.1626	-44.1031
22	373.0000	316.3485	25.2885	263.9033	368.7936	220.9307	411.7662	56.6515
23	191.6000	230.0371	17.7840	193.1553	266.9189	142.2059	317.8683	-38.4371
24	247.1200	304.4020	17.7175	267.6581	341.1459	216.6286	392.1754	-57.2820
25	269.6000	228.9247	23.0191	181.1861	276.6633	136.0106	321.8388	40.6753
26	.	65.9602	13.2497	38.4820	93.4384	-18.3553	150.2757	.

Sum of Residuals	0
Sum of Squared Residuals	32502
Predicted Residual SS (PRESS)	48738