Chapter 3

Measuring association in the multiple regression model

Overview

- The coefficient of multiple determination, R^2.
- Determining if a specific predictor should be included in the model
  - Partial F-test
  - t-test
- The coefficient of partial correlation

3.1 Coefficient of Multiple Determination

- To measure the adequacy of a multiple regression model, we use a measure called the coefficient of multiple determination
- **Definition:** The **coefficient of multiple determination** (sometimes known as $\boldsymbol{R^2}$) is defined as

$$r^2_{y\cdot 1,2,\ldots,p} = R^2 = \frac{SSR}{SST} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

[When $p = 1$, $r^2_{y\cdot 1,2,\ldots,p}$ is simply called the coefficient of determination.]

Coefficient of Multiple Determination (Continued)

- $r^2_{y\cdot 1,2,\ldots,p}$ measures the reduction in the variability of y obtained by using the independent variables x_1, …, x_p.

- Since $0 \le SSR \le SST$, therefore

$$0 \le r^2_{y\cdot 1,2,\ldots,p} \le 1$$

  i.e.

$$0 \le R^2 \le 1$$

Coefficient of Multiple Determination (Continued)

- If a perfect linear relationship exists between the dependent and independent variables, then all the variability in y about $\bar{y}$ can be explained by the variation in the independent variables and

$$r^2_{y\cdot 1,2,\ldots,p} = 1$$

- If no linear relationship exists, then

$$r^2_{y\cdot 1,2,\ldots,p} = 0$$

Coefficient of Multiple Determination (Continued)

- The positive square root of $r^2_{y\cdot 1,2,\ldots,p}$ (R^2) is called the **multiple correlation coefficient** and it is a measure of linear association between y and $x_1, \ldots, x_p$.

Remarks

- A large value of R^2 does not necessarily imply that the regression model is a good one.
- Adding a new independent variable to the existing model will always increase R^2, regardless of whether the additional variable contributes significantly to the model.

Example 1

- Refer to the Example 1 in Chapter 2 on p.2-15.
- From the ANOVA table, we have
- $SST = 888.25$ and $SSR = 692.8226$, therefore

$$R^2 = r^2_{y\cdot 1,2,\ldots,p} = \frac{SSR}{SST} = \frac{692.8226}{888.25} = 0.780$$

- Hence, about 78% of the variation in weight can be explained by the variation in both height and age.
- About 22% of the variation in weight is due to chance and/or the omission of some other independent variables.

3.2 Correlation Coefficient Matrix

- To further study the relationship among the variables, it is often important to examine the correlation between each pair of variables in the model.

- [Recall: The sample correlation coefficient between x and y is defined to be

$$r_{xy} = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

]

Remarks

1. A computational formula for r_{xy} is

$$r_{xy} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sqrt{\sum_{i=1}^{n} y_i^2 - n\bar{y}^2}\sqrt{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}} = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

where $S_{xx} = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$, $S_{yy} = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2$

and $S_{xy} = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}$

2. $-1 \le r_{xy} \le 1$ and $r_{xy} = r_{yx}$

$r_{xy} = \pm 1$ if y is a linear function of x.

(i.e. $y = a + bx$ and $r = 1$ if $b > 0$; $r = -1$ if $b < 0$.)

Correlation Coefficient Matrix of $y, x_1, \ldots, x_p$

$$
\begin{array}{cc}
 & \begin{array}{ccccc} y & x_1 & x_2 & \cdots & x_p \end{array} \\
\begin{array}{c} y \\ x_1 \\ x_2 \\ \vdots \\ x_p \end{array} &
\begin{pmatrix}
1 & r_{y1} & r_{y2} & \cdots & r_{yp} \\
r_{1y} & 1 & r_{12} & \cdots & r_{1p} \\
r_{2y} & r_{21} & 1 & \cdots & r_{2p} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
r_{py} & r_{p1} & r_{p2} & \cdots & 1
\end{pmatrix}
\end{array}
$$

where $r_{y1} = r_{1y} = r_{yx_1}$ and $r_{ij} = r_{ji} = r_{x_i x_j}$

Example 1 (Continued)

From the data, we have

$$r_{y1} = \frac{\sum_{i=1}^{n} x_{1i} y_i - n\bar{x}_1 \bar{y}}{\sqrt{\sum_{i=1}^{n} y_i^2 - n\bar{y}^2}\sqrt{\sum_{i=1}^{n} x_{1i}^2 - n\bar{x}_1^2}}$$

$$= \frac{40270 - 753(633)/12}{\sqrt{888.25}\sqrt{33903 - 633^2/12}} = 0.8143$$

Example 1 (Continued)

Similarly,

$$r_{y2} = \frac{\sum_{i=1}^{n} x_{2i} y_i - n\bar{x}_2\bar{y}}{\sqrt{\sum_{i=1}^{n} y_i^2 - n\bar{y}^2}\sqrt{\sum_{i=1}^{n} x_{2i}^2 - n\bar{x}_2^2}}$$
$$= \frac{6796 - 753(106)/12}{\sqrt{888.25}\sqrt{976 - 106^2/12}} = 0.7698$$

and $r_{12} = 0.6144$.

Example 1 (Continued)

- Note that all quantities used in the calculations are obtained from the matrices $X'X$ and $X'\underline{y}$ except $\sum y_i^2$

Correlation Coefficient Matrix is given by

$$
\begin{array}{lc}
 & \begin{array}{ccc} y & x_1 & x_2 \end{array} \\
\begin{array}{ll} \text{weight} & y \\ \text{height} & x_1 \\ \text{age} & x_2 \end{array} & \begin{pmatrix} 1 & 0.8143 & 0.7698 \\ 0.8143 & 1 & 0.6144 \\ 0.7698 & 0.6144 & 1 \end{pmatrix}
\end{array}
$$

- The matrix shows that there is a strong positive association between weight and height (y and x_1) and also between weight and age (y and x_2) and a moderate positive association between height and age (x_1 and x_2).

3.3 Evaluating the contribution of each independent variable

- An objective of developing a multiple regression model is to include only those independent variables that are useful in predicting the value of the dependent variable.

- A significant F test does **NOT** necessarily imply **ALL** the independent variables are useful.

3.3 Evaluating the contribution of each independent variable

- 2 ways to determine if a specific independent variable should be included in the regression model.

  i. Partial *F*-test

  ii. *t*-test for slope (β_k)

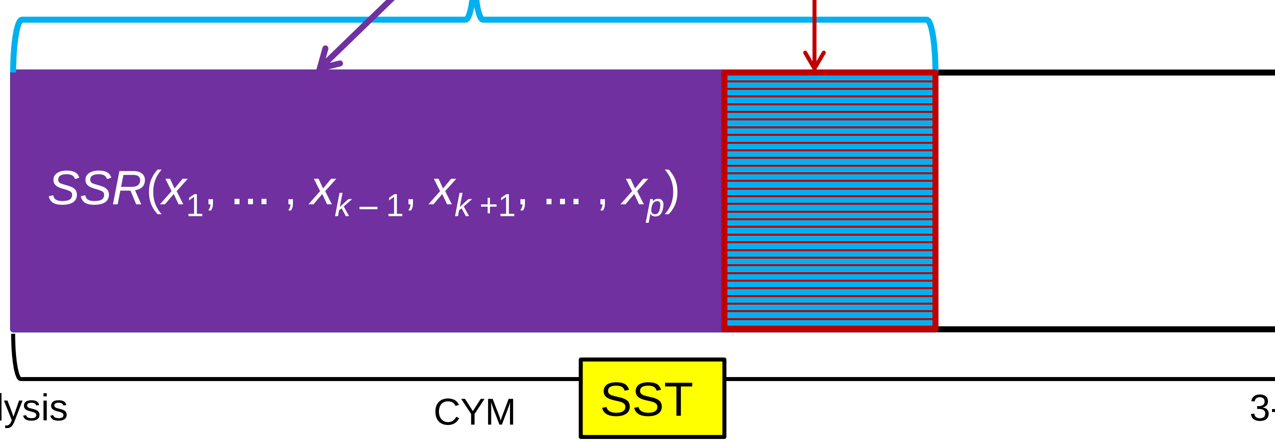
3.3.1 Partial F-test

Let

$$SSR(x_k \mid x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_p)$$
$$= SSR(x_1, \dots, x_p) - SSR(x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_p),$$

where $SSR(x_1, \dots, x_p)$ is the regression sum of squares for a model that includes $x_1, \dots, x_p$.

$SSR(x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_p)$ is the regression sum of squares for a model that includes $x_1, \dots, x_p$ except x_k.



Partial F-test (Continued)

- Hence $SSR(x_k \mid x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_p)$ measures the contribution of the variable x_k given that all other variables $x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_p$ are already included in the model.

Partial F-test (Continued)

- In particular, for a model with 2 independent variables x_1 and x_2, the contribution of x_1 given that x_2 has been included is

$$SSR(x_1 | x_2) = SSR(x_1, x_2) - SSR(x_2).$$

- Similarly, the contribution of x_2 given that x_1 has been included is

$$SSR(x_2 | x_1) = SSR(x_1, x_2) - SSR(x_1).$$

Partial F-test (Continued)

To test

H$_0$: variable x_k does not significantly improve the model already containing $x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_p$

against

H$_1$: variable x_k significantly improves the model already containing $x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_p$.

The partial F-test criterion is given by

$$F = \frac{SSR\left(x_k \middle| x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_p\right)}{MSE}$$

where $F \sim F(1, n - p - 1)$.

Reject H$_0$ if $F_{\text{obs}} > F_\alpha(1, n - p - 1)$.

Example 1 (Continued)

- Test H_0: height (x_1) does not significantly improve the model containing age (x_2) against

  H_1: H_0 is false.
- From the data on p.2-15, we know that

  SST = 888.25, $SSR(x_1, x_2)$ = 692.8226 and MSE = 21.7142.
- We need to compute $SSR(x_2)$.
- Consider the simple regression model

$$y = \beta_{02} + \beta_{12}x_2 + \epsilon,$$

Example 1 (Continued)

$$X_2{}' = \begin{pmatrix} 1 & \cdots & 1 \\ 8 & \cdots & 9 \end{pmatrix}$$

Hence, $X_2{}'X_2 = \begin{pmatrix} 12 & 106 \\ 106 & 976 \end{pmatrix}$ and $X_2{}'\underline{y} = \begin{pmatrix} 753 \\ 6796 \end{pmatrix}$

Therefore,

$$\underline{\hat{\beta}}_2 = \begin{pmatrix} \hat{\beta}_{02} \\ \hat{\beta}_{12} \end{pmatrix} = \frac{1}{476}\begin{pmatrix} 976 & -106 \\ -106 & 12 \end{pmatrix}\begin{pmatrix} 753 \\ 6796 \end{pmatrix}$$

$$= \begin{pmatrix} 14552/476 \\ 1734/476 \end{pmatrix}$$

Example 1 (Continued)

$$SSR(x_2) = \underline{\hat{\beta}}_2' X_2' \underline{\hat{y}}' - n\bar{y}^2$$
$$= \begin{pmatrix} \frac{14552}{476} & \frac{1734}{476} \end{pmatrix} \begin{pmatrix} 753 \\ 6796 \end{pmatrix} - 12\left(\frac{753}{12}\right)^2$$
$$= 526.3929$$

- Therefore $SSR(x_1 | x_2) = SSR(x_1, x_2) - SSR(x_2)$

$$= 692.8226 - 526.3929$$
$$= 166.4297.$$

Example 1 (Continued)

- Hence the observed F-value is given by

$$F = SSR(x_1 | x_2)/MSE$$
$$= 166.4297/21.7142$$
$$= 7.6646.$$

- Since $F_{obs} = 7.66 > 5.12 = F_{0.05}(1, 9)$, we reject H_0 at the 5% significance level and conclude that the inclusion of x_1 improves significantly the model containing x_2 alone.
- Note: p-value = 0.0218 (i.e. $\Pr(F(1,9) > 7.66) = 0.0218$))

Example 1 (Continued)

- Similarly, for testing
  H_0': age (x_2) does not significantly improve the model containing height (x_1) against
  H_1': H_0' is false.

- To compute $SSR(x_1)$.

- Consider the simple regression model
$$y = \beta_{01} + \beta_{11}x_1 + \epsilon,$$

Example 1 (Continued)

$$X_1{}' = \begin{pmatrix} 1 & \cdots & 1 \\ 57 & \cdots & 59 \end{pmatrix}$$

Hence $X_1{}'X_1 = \begin{pmatrix} 12 & 633 \\ 633 & 33903 \end{pmatrix}$ and $X_1{}'\underline{y} = \begin{pmatrix} 753 \\ 40270 \end{pmatrix}$

Therefore

$$\underline{\hat{\beta}}_1 = \begin{pmatrix} \hat{\beta}_{01} \\ \hat{\beta}_{11} \end{pmatrix} = \frac{1}{6147} \begin{pmatrix} 33903 & -633 \\ -633 & 12 \end{pmatrix} \begin{pmatrix} 753 \\ 40270 \end{pmatrix}$$
$$= \begin{pmatrix} 38049/6147 \\ 6591/6147 \end{pmatrix}$$

Example 1 (Continued)

$$SSR(x_1) = \underline{\hat{\beta}}_1' \, X_1' \, \underline{\hat{y}}' - n\bar{y}^2$$

$$= \begin{pmatrix} \frac{38049}{6147} & \frac{6591}{6147} \end{pmatrix} \begin{pmatrix} 753 \\ 40270 \end{pmatrix} - 12\left(\frac{753}{12}\right)^2$$

$$= 588.9225$$

- Therefore

$$SSR(x_2 | \, x_1) = SSR(x_1, x_2) - SSR(x_1)$$

$$= 692.8226 - 588.9225$$

$$= 103.9001.$$

Example 1 (Continued)

- Hence the observed F-value is given by

$$
\begin{aligned}
F &= SSR(x_2 \mid x_1)/MSE \\
&= 103.9001/21.7142 \\
&= 4.7849.
\end{aligned}
$$

- Since $F_{\text{obs}} = 4.78 < 5.12 = F_{0.05}(1, 9)$ (or p-value = 0.0565), we do not reject H_0 at the 5% significance level.

- Hence based on the partial F-test criterion, age does not improve the prediction of weight significantly in the model containing height.

3.3.2 t-test for the Slope (β_k)

- An equivalent way to perform the partial F-test is to use the t-test.

- Recall :

$$\hat{\beta}_k \sim N\left(\beta_k, \sigma_k^2\right)$$

where $\sigma_k{}^2 = \sigma^2((k+1, k+1)^{\text{th}} \text{ entry of } (X'X)^{-1})$

and

$$SSE/\sigma^2 \sim \chi^2(n-p-1)$$

t-test for the Slope (β_k) (Continued)

- Therefore

$$\frac{\hat{\beta}_k - \beta_k}{s.e.(\hat{\beta}_k)} \sim t(n-p-1)$$

where

$$s.e.(\hat{\beta}_k) = \text{standard error of } \hat{\beta}_k = \sqrt{\widehat{Var}(\hat{\beta}_k)}$$

$$= \sqrt{\hat{\sigma}^2 \times (k+1, k+1)^{\text{th}} \text{ entry of } (X'X)^{-1}}$$

with $\hat{\sigma}^2 = MSE$

t-test for the Slope (β_k) (Continued)

- To test H_0: $\beta_k = 0$ against H_1: $\beta_k \neq 0$, we consider the test statistic

$$t = \frac{\hat{\beta}_k}{s.e.(\hat{\beta}_k)}$$

- Reject H_0 if $|t| > t_{\alpha/2}(n - p - 1)$.

- Note: $\beta_k = 0$ means that the variable x_k does not contribute significantly to the prediction of y in a model already containing $x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_p$.

Example 1 (Continued)

- Test H_0: $\beta_1 = 0$ against H_1: $\beta_1 \neq 0$
- From the data, we have

$$\underline{\hat{\beta}} = \frac{1}{151956}\begin{pmatrix} 995775 \\ 109718 \\ 311529 \end{pmatrix} = \begin{pmatrix} 6.533 \\ 0.722 \\ 2.050 \end{pmatrix}$$

$$\widehat{Var}\left(\underline{\hat{\beta}}\right) = MSE(X'X)^{-1}$$

$$= 21.7142\,\frac{1}{151956}\begin{pmatrix} 838287 & -15834 & 1084 \\ -15834 & 476 & -1050 \\ 1084 & -1050 & 6147 \end{pmatrix}$$

Example 1 (Continued)

- Therefore

$$\widehat{Var}(\hat{\beta}_1) = \frac{21.7142(476)}{151956}$$

- Hence

$$t = \frac{\hat{\beta}_1}{s.e.(\hat{\beta}_1)} = \frac{109718}{151956}\sqrt{\frac{151956}{21.7142(476)}}$$

$$= 2.7685$$

Since $|t| = 2.7685 > 2.262 = t_{0.025}(9)$, we reject H_0 at the 5% level of significance.

Example 1 (Continued)

- Test H_0': $\beta_2 = 0$ against H_1': $\beta_2 \neq 0$

- From the data, we have

$$\hat{\beta}_2 = \frac{311529}{151956} = 2.050$$

$$\widehat{Var}\left(\hat{\beta}_2\right) = 21.7142\left(\frac{6147}{151956}\right)$$

Example 1 (Continued)

- Therefore

$$t = \frac{\hat{\beta}_2}{s.e.(\hat{\beta}_2)} = \frac{311529}{151956}\sqrt{\frac{151956}{21.7142(6147)}}$$

$$= 2.1874$$

- Since $|t| = 2.1874 < 2.262 = t_{0.025}(9)$, we do not reject H_0'.

- (Note that the t-test reaches the same conclusions as the partial F-test.)

3.4 Coefficients of partial determination and correlation

- The **coefficient of partial determination** measures the proportion of the variation in the dependent variable that is **explained** by a *particular* independent variable while controlling for, or *holding constant,* the *other* independent variables.

3.4.1 Coefficients of partial determination

- In a model with p independent variables, the coefficient of partial determination of the variable x_k with other variables holding constant is

$$r^2_{yk\cdot\text{all variables except } x_k}$$

$$= \frac{SSR(x_k | x_1, \cdots, x_{k-1}, x_{k+1}, \cdots, x_p)}{SSE(x_1, \cdots, x_{k-1}, x_{k+1}, \cdots, x_p)}$$

This is the variation that unexplained by the model with x_1, …, x_{k-1}, x_{k+1}, …, x_p

$$= \frac{SSR(x_k | x_1, \cdots, x_{k-1}, x_{k+1}, \cdots, x_p)}{SST - SSR(x_1, \cdots, x_{k-1}, x_{k+1}, \cdots, x_p)}$$

$$= \frac{SSR(x_k | x_1, \cdots, x_{k-1}, x_{k+1}, \cdots, x_p)}{SST - SSR(x_1, \cdots, x_p) + SSR(x_k | x_1, \cdots, x_{k-1}, x_{k+1}, \cdots, x_p)}$$

Coefficients of partial determination (Continued)

- In particular, for a model with 2 independent variables,

$$r^2_{y1\cdot 2} = \frac{SSR(x_1|x_2)}{SST - SSR(x_1, x_2) + SSR(x_1|x_2)}$$

$$r^2_{y2\cdot 1} = \frac{SSR(x_2|x_1)}{SST - SSR(x_1, x_2) + SSR(x_2|x_1)}$$

3.4.2 Coefficients of Partial Correlation

- The **coefficient of partial correlation** measures the strength of association between y and an independent variable while controlling for, or holding constant, the other independent variables. It is defined as

$$r_{yk\cdot\text{all variables except } x_k} = \sqrt{r^2_{yk\cdot\text{all variables except } x_k}}$$

- For a model with 2 independent variables

$$r_{y1\cdot 2} = \sqrt{r^2_{y1\cdot 2}} \quad \text{and} \quad r_{y2\cdot 1} = \sqrt{r^2_{y2\cdot 1}}$$

Example 1 (Continued)

- From the data, we have

  $SST = 888.25$, $SSR(x_1, x_2) = 692.8226$,

  $SSR(x_1 | x_2) = 166.4297$

  and $SSR(x_2 | x_1) = 103.9001$.

- Therefore

$$r_{y1\cdot 2}^2 = \frac{166.4297}{888.25 - 692.8226 + 166.4297} = 0.4599$$

$$r_{y2\cdot 1}^2 = \frac{103.9001}{888.25 - 692.8226 + 103.9001} = 0.3471$$

Example 1 (Continued)

- $r^2_{y1\cdot 2}$ can be interpreted to mean that at a fixed or constant age (x_2), about 46% of the variation in weight (y) can be explained by the variation in height (x_1).
- Similar interpretation for $r^2_{y2\cdot 1}$

$r_{y1\cdot 2} = (0.4599)^{0.5} = 0.6782$.

$r_{y2\cdot 1} = (0.3471)^{0.5} = 0.5892$.

- $r_{y2\cdot 1}$ can be interpreted to mean that at a fixed or constant age (x_2), variables y and x_1 are moderately associated.

3.5.1 SAS Program

```
*Partial F-test;
data ch3ex1;
  infile "d:/ST3131/Lecture/ch3ex1.txt";
  input weight height age;

proc glm data=ch3ex1;
  model weight = height age;
run;

*Compute a partial correlation coefficient;
proc corr data=ch3ex1 nosimple;
var weight height;
partial age;
proc corr  data=ch3ex1 nosimple;
var weight age;
partial height;
run;
```

3.5.1 Partial SAS Output

```
                        The GLM Procedure
Dependent Variable: weight
                                             Sum of
Source               DF        Squares    Mean Square    F Value    Pr > F
Model                 2    692.8226065    346.4113033      15.95    0.0011
Error                 9    195.4273935     21.7141548
Corrected Total      11    888.2500000

             R-Square     Coeff Var      Root MSE     weight Mean
             0.779986      7.426048      4.659845        62.75000

Source               DF      Type I SS    Mean Square    F Value    Pr > F
height                1    588.9225232    588.9225232      27.12    0.0006
age                   1    103.9000834    103.9000834       4.78    0.0565

Source               DF    Type III SS    Mean Square    F Value    Pr > F
height                1    166.4297494    166.4297494       7.66    0.0218
age                   1    103.9000834    103.9000834       4.78    0.0565

                                         Standard
   Parameter          Estimate                Err     t Value     Pr > |t|
   Intercept       6.553048251        10.94482708        0.60       0.5641
   height          0.722037958         0.26080506        2.77       0.0218
   age             2.050126352         0.93722561        2.19       0.0565
```

Type I SS: height — SSR(height); age — SSR(age | height)

Type III SS: height — SSR(height | age); age — SSR(age | height)

3.5.1 Partial SAS Output (Continued)

```
                  The CORR Procedure
   1 Partial Variables:    age
   2         Variables:    weight   height

Pearson Partial Correlation Coefficients, N = 12
      Prob > |r| under H0: Partial Rho=0

                 weight         height
      weight     1.00000        0.67818
                                 0.0218
      height     0.67818        1.00000
                  0.0218

                  The CORR Procedure
   1 Partial Variables:    height
   2         Variables:    weight   age

Pearson Partial Correlation Coefficients, N = 12
      Prob > |r| under H0: Partial Rho=0
                 weight            age
      weight     1.00000        0.58916
                                 0.0565
         age     0.58916        1.00000
                  0.0565
```

3.5.2 R Program and Output

```
> model1=lm(weight~height+age)
> #SSR(height) and SSR(age|height) are given in the ANOVA table
> anova(model1)
Analysis of Variance Table
Response: weight
          Df Sum Sq Mean Sq F value    Pr(>F)
height     1 588.92  588.92 27.1216 0.0005582 ***
age        1 103.90  103.90  4.7849 0.0564853 .
Residuals  9 195.43   21.71
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #SSR(age) and SSR(height|age) are given in the ANOVA table
> model2=lm(weight~age + height)
> anova(model2)
Analysis of Variance Table
Response: weight
          Df Sum Sq Mean Sq F value    Pr(>F)
age        1 526.39  526.39 24.2419 0.0008205 ***
height     1 166.43  166.43  7.6646 0.0218070 *
Residuals  9 195.43   21.71
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

SSR(height)

SSR(age | height)

SSR(age)

SSR(height | age)

R Program and Output (Continued)

```
> #Compute partial correlation
> model5=anova(model1)
> sst=sum(weight^2)-length(weight)*mean(weight)^2
> p.R2_wt_age.ht = model5$"Sum Sq"[2]/(sst -
   model5$"Sum Sq"[1])
> p.cor_wt_age.ht = sqrt(p.R2_wt_age.ht)
> p.R2_wt_age.ht; p.cor_wt_age.ht
[1] 0.3471117
[1] 0.5891619
>
> model6=anova(model2)
> p.R2_wt_ht.age = model6$"Sum Sq"[2]/(sst-
   model6$"Sum Sq"[1])
> p.cor_wt_ht.age = sqrt(p.R2_wt_ht.age)
> p.R2_wt_ht.age;p.cor_wt_ht.age
[1] 0.4599322
[1] 0.678183
```

The 2nd entry under the column "Sum Sq" in the object "model5". i.e. SSR(age | height) = 103.90