2.2 Analysis of Variance in MLR

2.2.1 Sum of Squares and Mean Squares

The Sum of Squares for the analysis of variance in matrix terms are:

$$\begin{aligned}
TSS &= \mathbf{y}^T\mathbf{y} - \frac{1}{n}\mathbf{y}^T\mathbf{J}\mathbf{y} = \mathbf{y}^T\left(\mathbf{I} - \frac{1}{n}\mathbf{J}\right)\mathbf{y} \\
RSS &= \mathbf{r}^T\mathbf{r} = (\mathbf{y} - \mathbf{X}\hat{\beta})^T(\mathbf{y} - \mathbf{X}\hat{\beta}) = \mathbf{y}^T(\mathbf{I} - \mathbf{H})\mathbf{y} \\
FSS &= \hat{\beta}^T\mathbf{X}^T\mathbf{y} - \frac{1}{n}\mathbf{y}^T\mathbf{J}\mathbf{y} = \mathbf{y}^T\left(\mathbf{H} - \frac{1}{n}\mathbf{J}\right)\mathbf{y}
\end{aligned}$$

where $\mathbf{J}$ is an $n \times n$ matrix of $\mathbf{1}$s, and $\mathbf{H}$ is the hat matrix defined above.

TSS has $n-1$ degrees of freedom associated with it. *RSS* has $n-p$ degrees of freedom associated since p parameters need to be estimated in the regression function model. Finally, *FSS* has $p-1$ degrees of freedom associated with it, representing the number of variables $x_2, \dots, x_p$.

The corresponding *Mean Squares* compute as usual by

$$MSR = \frac{FSS}{p-1} \qquad \text{and} \qquad MSE = \frac{RSS}{n-p}$$

The expectation of *MSE* is σ^2 as for the simple linear regression. On the other hand, the expectation of *MSR* is σ^2 plus a nonnegative constant. Indeed, when $p-1=2$, we have

$$\mathbf{E}(MSR) = \sigma^2 + \frac{1}{2}\left(\beta_1^2\sum(x_{i1}-\bar{x}_1)^2 + \beta_2^2\sum(x_{i2}-\bar{x}_2)^2\right) + 2\beta_1\beta_2\sum(x_{i1}-\bar{x}_1)(x_{i2}-\bar{x}_2)$$

Observe that if both β_1 and β_2 are zero, then $\mathbb{E}(MSR) = \sigma^2$. Otherwise, $\mathbb{E}(MSR) > \sigma^2$.

ANOVA Table

Source	df	SS	MS
Regression	$p-1$	FSS	$\mathrm{FSS}/(p-1)$
Error	$n-p$	RSS	$\mathrm{RSS}/(n-p)$
Total	$n-1$	TSS	

Birthweight Example

We can obtain the ANOVA table for the birthweight example in `R` as follows:

```
anova(birthweight.mlr1)
```

```
## Analysis of Variance Table
##
## Response: Birthweight
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Length     1 7.8991  7.8991 69.6588 4.431e-09 ***
## Headcirc   1 1.6594  1.6594 14.6335 0.0006698 ***
## Gestation  1 1.1052  1.1052  9.7461 0.0041461 **
## smoker     1 0.4458  0.4458  3.9313 0.0572841 .
## mage       1 0.0228  0.0228  0.2011 0.6572486
## mnocig     1 0.0224  0.0224  0.1979 0.6598472
## mheight    1 0.0407  0.0407  0.3593 0.5537302
## mppwt      1 0.1958  0.1958  1.7269 0.1994720
## fage       1 0.0849  0.0849  0.7487 0.3942539
## fedyrs     1 0.0164  0.0164  0.1450 0.7062480
## fnocig     1 0.0079  0.0079  0.0696 0.7938981
## fheight    1 0.2135  0.2135  1.8826 0.1809274
## lowbwt     1 0.0632  0.0632  0.5571 0.4616406
## Residuals 28 3.1751  0.1134
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
FSS = sum(anova(birthweight.mlr1)[1:13,2])
FSS
```

```
## [1] 11.77715
```

As you can see the output is slightly different than the table provided above. Indeed, `R` gives us a more detailed picture by computing a more detailed decomposition. To be more specific, instead of giving us *FSS*, we have the *SS* that corresponds to each variable in the model. This means that the *FSS* will be just the **sum** of all the *SS* values corresponding to predictors. Specifically here,

$$FSS = SS_{Length} + SS_{Headcirc} + \ldots + SS_{lowbwt} = 11.77715$$

2.2.2 Goodness of Fit: Multiple R-Square

The coefficient of multiple determination, denoted by R^2, is defined as follows:

$$R^2 = 1 - \frac{\sum_i (\hat{y}_i - y_i)^2}{\sum_i (y_i - \bar{y})^2} = 1 - \frac{RSS}{TSS}$$

It measures *the proportionate reduction of total variation in* y *asociated with the use of the set of* $\mathbf{X}$ *variables* $x_1, \ldots, x_p$. The coefficient of multiple determination reduces to the coefficient of simple determination for SLR when $p - 1 = 1$.

Similarly to the SLR case, we have:

$$0 \leq R^2 \leq 1,$$

where R^2 is zero when all $\beta_k = 0$, and the value 1 when all y observations fall directly on the fitted regression surface, i.e. when $y_i = \hat{y}_i$.

Note that **adding more variables to the regression model can only increase R^2 and never reduce it**, because *RSS* can never become larger with more variables and *TSS* is always the same for a given set of responses.

Remark: A large value of R^2 does not necessarily mean that the fitted model is a useful one. For example, observations may have been taken at only a few levels of the predictor variables. Despite a high R^2 in this case, the fitted model may not be useful if most predictors require extrapolations outside the range of observations. And, again, even thoug R^2 is large, *MSE* may still be too large for inferences to be useful when high precision is required.