Commented Output: Body Fat Data

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  117.085     99.782   1.173    0.258
Triceps        4.334      3.016   1.437    0.170
Thigh         -2.857      2.582  -1.106    0.285
Midarm        -2.186      1.595  -1.370    0.190
```

COMMENT: This is the analysis of the coefficients and y-intercept. We will ignore interpreting the intercept as keeping it in the model is valuable even if it were zero. The values under the estimate column provide the y-intercept and the slope coefficient for that variable, respectively. The t-value and p-value for each slope is a test of the population slope for that variable being zero (Ho) versus not being zero (Ha), under condition that this variable is added last to the model. That is, the slope test is used to determine if that predictor is a significant linear predictor of the response when added to the model last.

```
Residual standard error: 2.48 on 16 degrees of freedom
Multiple R-squared:  0.8014,      Adjusted R-squared:  0.7641
F-statistic: 21.52 on 3 and 16 DF, p-value: 7.343e-06
```

COMMENT: The residual standard error is the square root of the mean squared error (MSE). It is the MSE that is used in calculating the F-tests in the ANOVA tables. Thus 6.15 would be the MSE. Multiple R-squared is the coefficient of determination which provides the proportion (percentage) of the variation in the response that is explained by the regression model. This can be calculated by taking the SSR/SST or by 1 – SSE/SST. In comments below we will show how to calculate SSR and SST. The adjusted R-squared is a statistic used in model comparison and does not have any useful interpretation. Its value comes from penalizing a model for adding poorly working predictor(s). So, while R-square will increase with addition of predictors to a model, the adjusted R-square could decrease if the variable(s) added were not helpful to improving the overall model. The F-statistic, degrees of freedom and p-value are the test of the overall model (i.e. the null hypothesis that all the model population slopes equal zero versus alternative that at least one population slope does not equal zero.) The F-statistic is calculated by the ratio of the Mean Square Regression (MSR) and the Mean Square Error (MSE). That is, the F-statistic equals MSR/MSE. The MSR is found by taking the SSR and dividing by the DF for the overall regression model (this degree of freedom equals the number of slopes in the model). The degrees of freedom in the output give the numerator (number of model slopes) and denominator (sample size minus number of model estimates)degrees of freedom used in calculating the p-value for the F-test of the overall regression model.

```
Analysis of Variance Table

Response: BodyFat
          Df Sum Sq Mean Sq F value    Pr(>F)
Triceps    1 352.27  352.27 57.2768 1.131e-06 ***
Thigh      1  33.17   33.17  5.3931   0.03373 *
Midarm     1  11.55   11.55  1.8773   0.18956
Residuals 16  98.40    6.15
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

COMMENT: These are the Type I or sequential sum of squares (Seq SS). This information evaluates the significance of a predictor as it comes into the model in the order shown (top predictor is first in model, etc.). The sum of the SS for the predictors provides the Sum of Squares Regression (SSR) for the overall regression model. The sum of squares for the residuals is the Sum of Square Error (SSE). The Sum of Square Total (SST) is the sum of the squares given in the output or by SSR plus SSE. Here, SSR is 352.27 + 33.17 + 11.55 = 396.99; SSE = 98.40; and SST = 396.99 + 98.40 = 495.39 Looking at the previous COMMENT section where were mentioned that R-square is equal to SSR/SST, that would be 396.99/495.39 = 0.8014 as shown in the output as the Multiple R-squared value. The Mean Square values are found by taking each sum of square and dividing by its respective degrees of freedom. The value for the Mean Sq for the Residuals is the MSE of the model (6.15) as stated in previous COMMENT. The Mean Square of the Regression model (MSR) would be the SSR divided by the degrees of freedom for the model (i.e. the number of predictors or slopes). In this example, the MSR would be 396.99/3 which gives MSR = 132.33. Then as stated in previous COMMENT, the F-statistic for the overall regression model is found by MSR/MSE or in this example, 132.33/6.15 = 21.52 as given in the previous output. The F-values are found by taking the Mean Square value and dividing by the MSE. For example, the first F-value for Triceps is found by taking 352.27/6.15 Since each is a test of a single slope (i.e. the slope for that predictor) the degrees of freedom used to calculate the p-value for these tests are 1 and the DF for the error (here that is 16 found by sample size minus number of model estimates). Finally, these F-tests are tests of the significant predictive value of adding that variable to the model containing the preceding predictor(s). Obviously (hopefully!) the first predictor would be the case of a simple linear regression model. The addition to the R-squared that each variable contributes (i.e. the contribution percentage) as it enters the model can be calculated by taking the Type I SS for that variable and dividing by the SST. For example, The contribution of Thigh when added to model with only Triceps would be 11.55/495.39 = 0.067 or 6.7%.

```
Anova Table (Type III tests)

Response: BodyFat
             Sum Sq Df F value Pr(>F)
(Intercept)   8.468  1  1.3769 0.2578
Triceps      12.705  1  2.0657 0.1699
Thigh         7.529  1  1.2242 0.2849
Midarm       11.546  1  1.8773 0.1896
Residuals    98.405 16
```

COMMENT: These are the Type III or variable added last sum of squares. Comparing this ANOVA table to the previous ANOVA table (Type 1 SS) notice that: (1) only the SS for the **last** predictor is the same and (2) that the SS for the Residuals are the same. The reason for (1) is that in both instances the last variable (Midarm) is being evaluated as the last variable to enter the model. Also, notice that the p-values here are the same as those found in the Coefficients section at start of page. Recall from the COMMENT in that section of output that those t-tests were a test of the variable added last, which is what is going on here in the TYPE 3 tests. Therefore, the hypotheses would be the same. The F-values can be found by either squaring the t-statistics from the Coefficient output or by taking the Type III Sum of Squares for the variable and dividing by the MSE of the overall model. For instance, the F-value for Triceps of 2.0657 can be found by squaring the t-statistic (1.437) for Triceps found in the Coefficients table, or by taking the Type III SS for Triceps (12.705) and dividing by the overall model MSE (6.15). Either arrives at an F-value of 2.066, with rounding.