Print

Practice: Performing Multiple Regression Using PROC REG

Using the **stat1.bodyfat2** table, fit a multiple regression model with multiple predictors, and then modify the model by removing the least significant predictors.

1. Run a regression of **PctBodyFat2** on the variables **Age**, **Weight**, **Height**, **Neck**, **Chest**, **Abdomen**, **Hip**, **Thigh**, **Knee**, **Ankle**, **Biceps**, **Forearm**, and **Wrist**.

   Note: Turn off ODS Graphics.

   Submit the following program:

```
/*st103s02.sas*/  /*Part A*/
ods graphics off;
proc reg data=STAT1.BodyFat2;
    model PctBodyFat2=Age Weight Height
          Neck Chest Abdomen Hip Thigh
          Knee Ankle Biceps Forearm Wrist;
    title 'Regression of PctBodyFat2 on All '
          'Predictors';
run;
quit;
```

   Here are the results.

2. Compare the ANOVA table with this one from the model with only **Weight**. What is different?

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	6593.01614	6593.01614	150.03	<.001
Error	250	10986	43.94389		
Corrected Total	251	17579			

   There are key differences between the ANOVA table for this model and the one for the simple linear regression model. The degrees of freedom for the model are much higher, 13 versus 1. Also, the Mean Square model and the *F* ratio are much smaller.

3. How do the R-Square and the adjusted R-Square compare with these statistics for the **Weight** regression?

Root MSE	6.62902	R-Square	0.3751
Dependent Mean	19.15079	Adj R-Sq	0.3726
Coeff Var	34.61485		

   Both the R-Square and the adjusted R-Square for the full models are larger than the simple linear regression. The multiple regression model explains almost 75% of the variation in the **PctBodtFat2** variable versus approximately 37.5% that is explained by the simple linear regression model.

4. Did the estimate for the intercept change? Did the estimate for the coefficient of **Weight** change?

   Yes, including the other variables in the model changed both the estimate of the intercept and the slope for **Weight**. Also, the *p*-values for both changed dramatically. The slope of **Weight** is now not significantly different from zero.

5. To simplify the model, rerun the model from step 1, but eliminate the variable with the highest *p*-value. Compare the output with the model from step 1.

   Submit the code below. **Knee** was removed because it has the largest *p*-value (0.9552).

```
/*st103s02.sas*/  /*Part B*/
ods graphics off;
proc reg data=STAT1.BodyFat2;
    model PctBodyFat2=Age Weight Height
          Neck Chest Abdomen Hip Thigh
          Ankle Biceps Forearm Wrist;
    title 'Regression of PctBodyFat2 on All '
          'Predictors, Minus Knee';
run;
quit;
```

   Here are the results.

6. Did the *p*-value for the model change?

   The *p*-value for the model did not change to four decimal places.

7. Did the R-Square and the adjusted R-Square values change?

   The R-Square showed essentially no change. The adjusted R-Square increased from .7348 to .7359. When an adjusted R-Square increases by removing a variable from the model, it strongly implies that the removed variable was not necessary.

8. Did the parameter estimates and their *p*-values change?

   Some of the parameter estimates and their *p*-values changed slightly, but none to any large degree.

9. To simplify the model further, rerun the model from step 5, but eliminate the variable with the highest *p*-value. How did the output change from the previous model?

   Submit the following code. **Chest** was removed because it is the variable with the highest *p*-value in the previous model.

```
/*st103s02.sas*/  /*Part C*/
ods graphics off;
proc reg data=STAT1.BodyFat2;
    model PctBodyFat2=Age Weight Height
          Neck Abdomen Hip Thigh
          Ankle Biceps Forearm Wrist;
    title 'Regression of PctBodyFat2 on All '
          'Predictors, Minus Knee, Chest';
```

```
run;
quit;
```

Here are the results.

The ANOVA table did not change significantly. The R-Square remained essentially unchanged. The adjusted R-Square increased again. This confirms that the variable **Chest** did not contribute to explaining the variation in **PctBodyFat2** when the other variables were in the model.

10. Did the number of parameters with *p*-values less than 0.05 change?

The *p*-value for **Weight** changed more than any other and is now slightly more than 0.05. The *p*-values and parameter estimates for other variables changed much less. There are no more variables in this model with *p*-values below 0.05, compared with the previous one.

Hide Solution