Analysis1. 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**.
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			

Observation: The ANOVA table with the model with only Weight is only a simple regression model. However, the degrees of freedom for the model are much higher, 13 versus 1. Also, Sum of squares and mean square values are less when we considered all variables.

Analysis2. 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		

Observation: R-square and Adj R-square is higher in case of considering more variables.

Analysis3.

Did the estimate for the intercept change? Did the estimate for the coefficient of **Weight** change?
Yes, Yes

Analysis4.

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

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;
```

From Question1:

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	13	13159	1012.22506	54.50	< 0001
Error	238	4420.06401	18.57170		
Corrected Total	251	17579			

After Eliminating the variable with the highest *p*-value, F value increases, p-value did not change.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	10	13158	1315.76595	71.72	<.0001
Error	241	4421.33035	18.34577		
Corrected Total	251	17579			

Analysis5.

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

From question1 and after eliminating higher value of p-value variables: adj-R-square increases,

Root MSE	4.30949	**R-Square**	0.7486
Dependent Mean	19.15079	**Adj R-Sq**	0.7348
Coeff Var	22.50293		

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.

Analysis6.

Did the parameter estimate and their *p*-values change?
From the results:

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-21.35323	22.18616	-0.96	0.3368
Age	1	0.06457	0.03219	2.01	0.0460
Weight	1	-0.09638	0.06185	-1.56	0.1205
Height	1	-0.04394	0.17870	-0.25	0.8060
Neck	1	-0.47547	0.23557	-2.02	0.0447
Chest	1	-0.01718	0.10322	-0.17	0.8679
Abdomen	1	0.95500	0.09016	10.59	<.0001
Hip	1	-0.18859	0.14479	-1.30	0.1940
Thigh	1	0.24835	0.14617	1.70	0.0906
Knee	1	0.01395	0.24775	0.06	0.9552
Ankle	1	0.17788	0.22262	0.80	0.4251
Biceps	1	0.18230	0.17250	1.06	0.2917
Forearm	1	0.45574	0.19930	2.29	0.0231
Wrist	1	-1.65450	0.53316	-3.10	0.0021

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-25.99962	12.15316	-2.14	0.0334
Age	1	0.06509	0.03092	2.11	0.0363
Weight	1	-0.10740	0.04207	-2.55	0.0113
Neck	1	-0.46749	0.22812	-2.05	0.0415
Abdomen	1	0.95772	0.07276	13.16	<.0001
Hip	1	-0.17912	0.13908	-1.29	0.1990
Thigh	1	0.25926	0.13389	1.94	0.0540
Ankle	1	0.18453	0.21686	0.85	0.3957
Biceps	1	0.18617	0.16858	1.10	0.2705
Forearm	1	0.45303	0.19593	2.31	0.0216
Wrist	1	-1.65666	0.52706	-3.14	0.0019

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

Analysis7.

To simplify the model further, rerun the model from Question 4, but eliminate the variable with the highest *p*-value. How did the output change from 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;
```

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

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.

Analysis8.

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 estimate 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.