

Exercise: Using PROC GLMSELECT to Perform Stepwise Selection

Use the **stat1.bodyfat2** data set to identify a set of "best" models. Use significance-level model selection techniques.

1. With the **SELECTION=STEPWISE** option, use **SELECT=SL** in PROC GLMSELECT to identify a set of candidate models that predict **PctBodyFat2** as a function of the variables **Age, Weight, Height, Neck, Chest, Abdomen, Hip, Thigh, Knee, Ankle, Biceps, Forearm, and Wrist**. Use the default values for **SLENTRY=** and **SLSTAY=**. Submit the code and view the results.

```
/*st104s01.sas*/ /*Part A*/
ods graphics on;
proc glmselect data=stat1.bodyfat2 plots=all;
    STEPWISESL: model PctBodyFat2=Age Weight Height
                     Neck Chest Abdomen Hip Thigh
                     Knee Ankle Biceps Forearm Wrist
                     / SELECTION=STEPWISE SELECT=SL;
    title 'SL STEPWISE Selection with PctBodyFat2';
run;
```

Here are the [results](#).

In the results, notice the following:

- Selection stopped because the candidate for entry has $SLE > 0.15$ and the candidate for removal has $SLS < 0.15$.
- The stepwise selection process, using significance level, seems to select an eight-effect model (including the intercept).
- The Coefficient panel shows that the standardized coefficients do not vary greatly when additional effects are added to the model.
- The Fit panel indicates that the best model, according to AIC, AICC, and adjusted R-square, is the final model viewed during the selection process. SBC shows a minimum at step four.
- The parameter estimates from the selected model are presented in the Parameter Estimates table.

2. Modify the code to specify the forward selection process (**FORWARD**). Submit the code and view the results.

```
/*st104s01.sas*/ /*Part B*/
proc glmselect data=stat1.bodyfat2 plots=all;
    FORWARDSL: model PctBodyFat2=Age Weight Height
                     Neck Chest Abdomen Hip Thigh
                     Knee Ankle Biceps Forearm Wrist
                     / SELECTION=FORWARD SELECT=SL;
    title 'SL FORWARD Selection with PctBodyFat2';
run;
```

Here are the [results](#).

In the results, notice the following:

- Selection stopped as the candidate for entry has $SLE > 0.5$.
- The forward selection process, using significance level, seems to select an 11-effect model (including the intercept).
- The Coefficient panel shows that the standardized coefficients do not vary greatly when additional effects are added to the model.

- The Fit panel indicates that the best models, according to AIC, AICC, adjusted R-square, and SBC, are at various steps in the selection progression.
 - The parameter estimates from the selected model are presented in the Parameter Estimates table.
3. How many variables would result from a model using forward selection and a significance-level-for-entry criterion of 0.05, instead of the default SLENTY= value, 0.50? Modify and submit the code, and view the results.

```
/*st104s01.sas*/ /*Part C*/  
proc glmselect data=stat1.bodyfat2 plots=all;  
  FORWARDSL: model PctBodyFat2=Age Weight Height  
                  Neck Chest Abdomen Hip Thigh  
                  Knee Ankle Biceps Forearm Wrist  
                  / SELECTION=FORWARD SELECT=SL  
                  SLENTY=0.05;  
  title 'SL FORWARD (0.05) Selection with PctBodyFat2';  
run;
```

Here are the [results](#).

The results show that, when the value of SLENTY= is changed from the default to 0.05, the number of effects in the selected model is reduced to five (including the intercept).

Hide Solution