

## Exercise: Using the Linear Regression Task to Perform Stepwise Selection

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

1. Use the Stepwise Selection Method with Significance Level as the criterion in the Linear Regression task to identify a set of candidate models that predict **PctBodyFat2** as a function of the interval variables **Age, Weight, Height, Neck, Chest, Abdomen, Hip, Thigh, Knee, Ankle, Biceps, Forearm,** and **Wrist**. Use the default values for SLENTY= and SLSTAY=. Submit the task and view the results.
  1. In the Navigation pane, select **Tasks and Utilities**.
  2. Expand **Tasks**.
  3. Expand **Statistics** and open the **Linear Regression** task.
  4. Select the **stat1.bodyfat2** data set.
  5. Assign **PctBodyFat2** as the Dependent variable, and the interval variables **Age, Weight, Height, Neck, Chest, Abdomen, Hip, Thigh, Knee, Ankle, Biceps, Forearm,** and **Wrist** to the Continuous variables role.
  6. On the MODEL tab, click the **Edit this model** icon, select all variables, and click **Add**. Then click **OK**.
  7. On the OPTIONS tab, suppress all diagnostic, residual, and scatter plots.
  8. On the SELECTION tab, use the Selection method drop-down list to choose **Stepwise selection**.
  9. For Add/remove effects with value, choose **Significance level**.
  10. Specify the Significance level to add an effect to the model as 0.15, instead of 0.05.
  11. Specify the Significance level to remove an effect from the model as 0.15 instead of 0.05.
  12. Expand **SELECTION PLOTS** and select both **Criteria plots** and **Coefficient plots**.
  13. Run the task.

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 task to specify the forward selection process (FORWARD). Submit the code and view the results.

1. On the SELECTION tab, choose **Forward selection** as the selection method. Specify the Significance level to add an effect to the model as 0.5.
2. Run the task.

Here are the [results](#).

In the results, note the following:

- Selection stopped as the candidate for entry has SLE > 0.5.

- The forward selection process, using significance level, selects 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.

1. On the SELECTION tab, change the Significance level to add an effect to the model as 0.05.
2. Run the task.

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