e>

Print

Practice: Using the Correlation Analysis Task to Describe the Relationship between Continuous Variables

The percentage of body fat, age, weight, height, and 10 body circumference measurements (for example, abdomen) were recorded for 252 men by Dr. Roger W. Johnson of Calvin College in Minnesota. The data are in the **bodyfat2** data set. Body fat, one measure of health, has been accurately estimated by a water displacement measurement technique.

1. Use PROC CORR to generate scatter plots and correlations for the VAR variables **Age**, **Weight**, and **Height**, and the circumference measures versus the WITH variable, **PctBodyFat2**.

   IMPORTANT: For PROC CORR, ODS Graphics limits you to 10 VAR variables at a time. For this exercise, look at the relationships with **Age**, **Weight**, and **Height** separately from the circumference variables (**Neck**, **Chest**, **Abdomen**, **Hip**, **Thigh**, **Knee**, **Ankle**, **Biceps**, **Forearm**, and **Wrist**).

   Note: This limitation exists only on the graphics that are obtained from ODS. The correlation table displays all variables in the VAR statement by default.

   1. In the Navigation pane, select **Tasks and Utilities**.
   2. Expand **Tasks**.
   3. Expand **Statistics** and open the **Correlation Analysis** task.
   4. Select the **stat1.bodyfat2** data set.
   5. Assign **Age**, **Weight**, **Height**, **Neck**, **Chest**, **Abdomen**, **Hip**, **Thigh**, **Knee**, **Ankle**, **Biceps**, **Forearm**, and **Wrist** to the Analysis variables role.
   6. Assign **PctBodyFat2** to the Correlate with role.
   7. On the OPTIONS tab, under STATISTICS, use the drop-down list for Display statistics and select **Selected statistics**. Then select **Correlations**, **Display p-values**, and **Descriptive statistics**. (Correlations and Display p-values might already be selected.)
   8. Under PLOTS, use the drop-down list for Type of plot and select **Individual scatter plots**. Ensure the check box for **Include inset statistics** is selected, and change the Number of variables to plot value from **5** to **10**.
   9. Run the task.

   Here are the results.

2. Modify the task to generate the scatter plots for the remaining variables, **Biceps**, **Forearm**, and **Wrist**.

   1. On the DATA tab, select all the Analysis variables except **Biceps**, **Forearm**, and **Wrist** and click the **Remove column** icon.
   2. Run the task.

   Here are the results.

3. Examine all of the plots. Can straight lines adequately describe the relationships?

   Yes. **Height** seems to be the only variable that shows no real linear relationship. **Age** and **Ankle** show little linear trend.

4. Are there any outliers that you should investigate?

   One person has outlying values for several measurements. In addition, there are one or two values that seem to be outliers for **Ankle**.

5. Which variable has the highest correlation with **PctBodyFat2**?

   Abdomen, with r=0.81343, is the variable with the highest correlation with **PctBodyFat2**.

6. What is the *p*-value for the coefficient? Is it statistically significant at the 0.05 level?

   The *p*-value is <.0001.

7. Modify the Correlation Analysis task to generate correlations among all the variables previously mentioned (**Age**, **Weight**, **Height**, **Neck**, **Chest**, **Abdomen**, **Hip**, **Thigh**, **Knee**, **Ankle**, **Biceps**, **Forearm**, and **Wrist**) minus **PctBodyFat2**. Don't generate descriptive statistics or plots again. Select only the highest five per variable.

   Note: You'll need to edit the generated code to select the highest five.

   1. On the DATA tab, assign **Age**, **Weight**, **Height**, **Neck**, **Chest**, **Abdomen**, **Hip**, **Thigh**, **Knee** and **Ankle** to the Analysis variables role. **Biceps**, **Forearm**, and **Wrist** should already be assigned to this role.
   2. In the Correlate with box, select **PctBodyFat2** and click the **Remove column** icon.
   3. On the OPTIONS tab, clear the check box for **Descriptive statistics**.
   4. Use the drop-down list for Type of plot and select **None**.
   5. To print the highest five correlated variables for each variable, edit the generated code. Click the **Edit SAS Code** icon in the CODE tab and add **best=5** as follows:

      ```
      ods noproctitle;
      ods graphics / imagemap=on;

      proc corr data=STAT1.BODYFAT2 best=5 pearson nosimple plots=none;
         var Biceps Forearm Wrist Age Weight Height Neck Chest Abdomen Hip Thigh Knee Ankle;
      run;
      ```

   6. Run the code.

   Here are the results. The order of the variables in your results might vary, depending on the order in which you selected the variable names for the Analysis variables role.

8. Are there any notable relationships?

   Several relationships seem to have high correlations (such as those among **Hip**, **Thigh**, and **Knee**). **Weight** seems to correlate highly with all circumference variables.

Hide Solution