Overview

The CORR procedure is a statistical procedure for numeric random variables that computes Pearson correlation coefficients, three nonparametric measures of association, and the probabilities associated with these statistics. The correlation statistics include

- Pearson product-moment and weighted product-moment correlation
- Spearman rank-order correlation
- Kendall’s tau-b
Hoeffding's measure of dependence, D

Pearson, Spearman, and Kendall partial correlation.

PROC CORR also computes Cronbach's coefficient alpha for estimating reliability. The default correlation analysis includes descriptive statistics, Pearson correlation statistics, and probabilities for each analysis variable. You can save the correlation statistics in a SAS data set for use with other statistical and reporting procedures.

Output 12.1 on page 272 is the simplest form of PROC CORR output. Pearson correlation statistics are computed for all numeric variables from a study investigating the effect of exercise on physical fitness. The statements that produce the output follow:

options pagesize=60;
proc corr data=fitness;
run;

Output 12.1  Simple Correlation Analysis for a Fitness Study Using PROC CORR

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Sum</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>30</td>
<td>47.56667</td>
<td>5.26330</td>
<td>1427</td>
<td>38.00000</td>
<td>57.00000</td>
</tr>
<tr>
<td>Weight</td>
<td>30</td>
<td>77.70500</td>
<td>8.34152</td>
<td>2331</td>
<td>59.08000</td>
<td>91.63000</td>
</tr>
<tr>
<td>Runtime</td>
<td>29</td>
<td>10.61448</td>
<td>1.41655</td>
<td>307.82000</td>
<td>8.17000</td>
<td>14.03000</td>
</tr>
<tr>
<td>Oxygen</td>
<td>29</td>
<td>47.06445</td>
<td>5.32129</td>
<td>1365</td>
<td>37.38800</td>
<td>60.05500</td>
</tr>
</tbody>
</table>

Pearson Correlation Coefficients
Prob > |r| under H0: Rho=0
Number of Observations

<table>
<thead>
<tr>
<th></th>
<th>Age</th>
<th>Weight</th>
<th>Runtime</th>
<th>Oxygen</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>1.00000</td>
<td>-0.21777</td>
<td>0.19528</td>
<td>-0.32899</td>
</tr>
<tr>
<td></td>
<td>0.2477</td>
<td>0.3100</td>
<td>0.0814</td>
<td></td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>30</td>
<td>29</td>
<td>29</td>
</tr>
<tr>
<td>Weight</td>
<td>-0.21777</td>
<td>1.00000</td>
<td>0.15155</td>
<td>-0.19900</td>
</tr>
<tr>
<td></td>
<td>0.2477</td>
<td>0.4326</td>
<td>0.3007</td>
<td></td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>30</td>
<td>29</td>
<td>29</td>
</tr>
<tr>
<td>Runtime</td>
<td>0.19528</td>
<td>0.15155</td>
<td>1.00000</td>
<td>-0.78346</td>
</tr>
<tr>
<td></td>
<td>0.3100</td>
<td>0.4326</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>29</td>
<td>29</td>
<td>29</td>
<td>29</td>
</tr>
<tr>
<td>Oxygen</td>
<td>-0.32899</td>
<td>-0.19900</td>
<td>-0.78346</td>
<td>1.00000</td>
</tr>
<tr>
<td></td>
<td>0.0814</td>
<td>0.3007</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>29</td>
<td>29</td>
<td>28</td>
<td>29</td>
</tr>
</tbody>
</table>

Output 12.2 on page 274 and Output 12.3 on page 275 illustrate the use of PROC CORR to calculate partial correlation statistics for the fitness study and to store the results in an output data set. The statements that produce the analysis also

 suppress the descriptive statistics
 select and label analysis variables
- exclude all observations with missing values
- calculate the partial covariance matrix
- calculate three types of partial correlation coefficients
- generate an output data set that contains Pearson correlation statistics and print the output data set.

For an explanation of the program that produces the following output, see Example 4 on page 308.
### Output 12.2  Customized Correlation Analysis with Partial Covariances and Correlation Statistics

#### The CORR Procedure

1 Partial Variables: Age  
3 Variables: Weight, Oxygen, Runtime

**Partial Covariance Matrix, DF = 26**

<table>
<thead>
<tr>
<th></th>
<th>Weight</th>
<th>Oxygen</th>
<th>Runtime</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight</td>
<td>72.4374</td>
<td>-12.751</td>
<td>2.067</td>
</tr>
<tr>
<td>Oxygen</td>
<td>-12.751</td>
<td>27.016</td>
<td>-5.593</td>
</tr>
<tr>
<td>Runtime</td>
<td>2.067</td>
<td>-5.593</td>
<td>1.945</td>
</tr>
</tbody>
</table>

**Pearson Partial Correlation Coefficients, N = 28**

<table>
<thead>
<tr>
<th></th>
<th>Weight</th>
<th>Oxygen</th>
<th>Runtime</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight</td>
<td>1.0000</td>
<td>-0.2882</td>
<td>0.1741</td>
</tr>
<tr>
<td>Oxygen</td>
<td>-0.2882</td>
<td>1.0000</td>
<td>-0.7716</td>
</tr>
<tr>
<td>Runtime</td>
<td>0.1741</td>
<td>-0.7716</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

**Spearman Partial Correlation Coefficients, N = 28**

<table>
<thead>
<tr>
<th></th>
<th>Weight</th>
<th>Oxygen</th>
<th>Runtime</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight</td>
<td>1.0000</td>
<td>-0.1640</td>
<td>0.0870</td>
</tr>
<tr>
<td>Oxygen</td>
<td>-0.1640</td>
<td>1.0000</td>
<td>-0.6711</td>
</tr>
<tr>
<td>Runtime</td>
<td>0.0870</td>
<td>-0.6711</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

**Kendall Partial Tau b Correlation Coefficients, N = 28**

<table>
<thead>
<tr>
<th></th>
<th>Weight</th>
<th>Oxygen</th>
<th>Runtime</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight</td>
<td>1.0000</td>
<td>-0.0902</td>
<td>0.0285</td>
</tr>
<tr>
<td>Oxygen</td>
<td>-0.0902</td>
<td>1.0000</td>
<td>-0.5215</td>
</tr>
<tr>
<td>Runtime</td>
<td>0.0285</td>
<td>-0.5215</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
### Procedure Syntax

**Tip:** Supports the Output Delivery System, see “Output Delivery System” on page 18

**Reminder:** You can use the ATTRIB, FORMAT, LABEL, and WHERE statements. See Chapter 3, "Statements with the Same Function in Multiple Procedures," for details. You can also use any global statements as well. See Chapter 2, "Fundamental Concepts for Using Base SAS Procedures," for a list.

```sas
PROC CORR <option(s)>;
   BY <DESCENDING> variable-1
      <DESCENDING> variable-n
      <NOTSORTED>;
   FREQ frequency-variable;
   PARTIAL variable(s);
   VAR variable(s);
   WEIGHT weight-variable;
   WITH variable(s);
```

<table>
<thead>
<tr>
<th>To do this</th>
<th>Use this statement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Produce separate correlation analyses for each BY group</td>
<td>BY</td>
</tr>
<tr>
<td>Identify a variable whose values represent the frequency of each observation</td>
<td>FREQ</td>
</tr>
<tr>
<td>Identify controlling variables to compute Pearson, Spearman, or Kendall partial correlation coefficients</td>
<td>PARTIAL</td>
</tr>
<tr>
<td>Identify variables to correlate and their order in the correlation matrix</td>
<td>VAR</td>
</tr>
</tbody>
</table>

**Output 12.3** Output Data Set with Pearson Partial Correlation Statistics

```
Pearson Correlation Statistics Using the PARTIAL Statement
Output Data Set from PROC CORR

_TYPE_  _NAME_  Weight  Oxygen  Runtime
COV    Weight  72.4374  -12.7511  2.0677
COV    Oxygen  -12.7511  27.0165  -5.5937
COV    Runtime 2.0677  -5.5937  1.9451
MEAN   Weight  0.0000  0.0000  0.0000
STD    Weight  8.5110  5.1977  1.3947
N      Weight  28.0000  28.0000  28.0000
CORR   Weight 1.0000  -0.2882  0.1742
CORR   Oxygen -0.2882  1.0000  -0.7716
CORR   Runtime 0.1742  -0.7716  1.0000
```
To do this | Use this statement
--- | ---
Identify a variable whose values weight each observation to compute Pearson weight product-moment correlation | WEIGHT
Compute correlations for specific combinations of variables | WITH

**PROC CORR Statement**

```plaintext
PROC CORR <option(s)>;
```

<table>
<thead>
<tr>
<th>To do this</th>
<th>Use this option</th>
</tr>
</thead>
</table>
Specify the input data set | DATA= |
Create output data sets |
Specify an output data set to contain Hoeffding's D statistics | OUTH= |
Specify an output data set to contain Kendall correlations | OUTK= |
Specify an output data set to contain Pearson correlations | OUTP= |
Specify an output data set to contain Spearman correlations | OUTS= |
Control statistical analysis |
Request Hoeffding's measure of dependence, D | HOEFFDING |
Request Kendall's tau-b | KENDALL |
Request Pearson product-moment correlation | PEARSON |
Request Spearman rank-order correlation | SPEARMAN |
Control Pearson correlation statistics |
Compute Cronbach's coefficient alpha | ALPHA |
Compute covariances | COV |
Compute corrected sums of squares and crossproducts | CSSCP |
Exclude missing values | NOMISS |
Specify singularity criterion | SINGULAR= |
Compute sums of squares and crossproducts | SSCP |
Specify the divisor for variance calculations | VARDEF= |
Control printed output |
Specify the number and order of correlation coefficients | BEST= |
Suppress Pearson correlations | NOCORR |
Suppress all printed output | NOPRINT |
Suppress significance probabilities | NOPROB |
### Options

**ALPHA**

calculates and prints Cronbach's coefficient alpha. PROC CORR computes separate coefficients using raw and standardized values (scaling the variables to a unit variance of 1). For each VAR statement variable, PROC CORR computes the correlation between the variable and the total of the remaining variables. It also computes Cronbach's coefficient alpha using only the remaining variables.

**Main discussion:** "Cronbach's Coefficient Alpha" on page 292

**Restriction:** If you use a WITH statement, ALPHA is invalid.

**Interaction:** ALPHA invokes PEARSON.

**Interaction:** If you specify OUTP=, the output data set also contains six observations with Cronbach's coefficient alpha.

**Interaction:** When you use the PARTIAL statement, PROC CORR calculates Cronbach's coefficient alpha for partialled variables.

**See also:** OUTP= option

**Featured in:** Example 3 on page 304

**BEST=n**

prints n correlation coefficients for each variable. Correlations are ordered from highest to lowest in absolute value. Otherwise, PROC CORR prints correlations in a rectangular table using the variable names as row and column labels.

**Interaction:** When you specify HOEFFDING, PROC CORR prints the D statistics in order from highest to lowest.

**Range:** 1 to the maximum number of variables

**COV**

calculates and prints covariances.

**Interaction:** COV invokes PEARSON.

**Interaction:** If you specify OUTP=, the output data set contains the covariance matrix and the _TYPE_ variable value is COV.

**Interaction:** When you use the PARTIAL statement, PROC CORR computes a partial covariance matrix.

**See also:** OUTP= option

**Featured in:** Example 2 on page 302 and Example 4 on page 308

**CSSCP**

prints the corrected sums of squares and crossproducts.

**Interaction:** CSSCP invokes PEARSON.

**Interaction:** If you specify OUTP=, the output data set contains a CSSCP matrix and the _TYPE_ variable value is CSSCP. If you use a PARTIAL statement, the output data set contains a partial CSSCP matrix.

**Interaction:** When you use a PARTIAL statement, PROC CORR prints both an unpartial and a partial CSSCP matrix.
See also: OUTP= option

DATA=SAS-data-set

specifies the input SAS data set.

Main discussion: “Input Data Sets” on page 18

HOEFFDING

calculates and prints Hoeffding’s D statistics. This D statistic is 30 times larger than the usual definition and scales the range between -0.5 and 1 so that only large positive values indicate dependence.

Main discussion: “Hoeffding’s Measure of Dependence, D” on page 289

Restriction: When you use a WEIGHT or PARTIAL statement, HOEFFDING is invalid.

Featured in: Example 1 on page 299

KENDALL

calculates and prints Kendall tau-b coefficients based on the number of concordant and discordant pairs of observations. Kendall’s tau-b ranges from -1 to 1.

Main discussion: “Kendall’s tau-b” on page 288

Restriction: When you use a WEIGHT statement, KENDALL is invalid.

Interactions: When you use a PARTIAL statement, probability values for Kendall’s partial tau-b are not available.

Featured in: Example 4 on page 308

NOCORR

suppresses calculating and printing of Pearson correlations.

Interaction: If you specify OUTP=, the data set type remains CORR. To change the data set type to COV, CSSCP, or SSCP, use the TYPE= data set option.

See also: “Output Data Sets” on page 297

Featured in: Example 3 on page 304

NOMISS

excludes observations with missing values from the analysis. Otherwise, PROC CORR computes correlation statistics using all the nonmissing pairs of variables.

Main discussion: “Missing Values” on page 295

Tip: Using NOMISS is computationally more efficient.

Featured in: Example 3 on page 304

NOPRINT

suppresses all printed output.

Tip: Use NOPRINT when you want to create an output data set only.

NOPROB

suppresses printing the probabilities associated with each correlation coefficient.

NOSIMPLE

suppresses printing simple descriptive statistics for each variable. However, if you request an output data set, the output data set still contains simple descriptive statistics for the variables.

Featured in: Example 2 on page 302
OUTH=output-data-set
creates an output data set containing Hoeffding's D statistics. The contents of the output data set are similar to the OUTP= data set.

Main discussion: “Output Data Sets” on page 297
Interaction: OUTH= invokes HOEFFDING.

OUTK=output-data-set
creates an output data set containing Kendall correlation statistics. The contents of the output data set are similar to the OUTP= data set.

Main discussion: “Output Data Sets” on page 297
Interaction: OUTK= option invokes KENDALL.

OUTP=output-data-set
creates an output data set containing Pearson correlation statistics. This data set also includes means, standard deviations, and the number of observations. The value of the _TYPE_ variable is CORR.

Main discussion: “Output Data Sets” on page 297
Interaction: OUTP= invokes PEARSON.

Interaction: If you specify ALPHA, the output data set also contains six observations with Cronbach's coefficient alpha.

Featured in: Example 4 on page 308

OUTS=SAS-data-set
creates an output data set containing Spearman correlation statistics. The contents of the output data set are similar to the OUTP= data set.

Main discussion: “Output Data Sets” on page 297
Interaction: OUTS= invokes SPEARMAN.

PEARSON calculates and prints Pearson product-moment correlations when you use the HOEFFDING, KENDALL, or SPEARMAN option. If you omit the correlation type, PROC CORR automatically produces Pearson correlations. The correlations range from -1 to 1.

Main discussion: “Pearson Product-Moment Correlation” on page 287
Featured in: Example 1 on page 299

RANK prints the correlation coefficients for each variable. Correlations are ordered from highest to lowest in absolute value. Otherwise, PROC CORR prints correlations in a rectangular table using the variable names as row and column labels.

Interaction: If you use HOEFFDING, PROC CORR prints the D statistics in order from highest to lowest.

SINGULAR=p
specifies the criterion for determining the singularity of a variable when you use a PARTIAL statement. A variable is considered singular if its corresponding diagonal element after Cholesky decomposition has a value less than p times the original unpartialled corrected sum of squares of that variable.

Main discussion: “Partial Correlation” on page 290
Default: 1E-8
Range: between 0 and 1
SPEARMAN

calculates and prints Spearman correlation coefficients based on the ranks of the variables. The correlations range from -1 to 1.

Main discussion: “Spearman Rank-Order Correlation” on page 288

Restriction: When you specify a WEIGHT statement, SPEARMAN is invalid.

Featured in: Example 1 on page 299

SSCP

prints the sums of squares and crossproducts.

Interaction: SSCP invokes PEARSON.

Interaction: When you specify OUTP=, the output data set contains a SSCP matrix and the _TYPE_ variable value is SSCP. If you use a PARTIAL statement, the output data set does not contain an SSCP matrix.

Interaction: When you use a PARTIAL statement, PROC CORR prints the unpartial SSCP matrix.

Featured in: Example 2 on page 302

VARDEF=divisor

specifies the divisor to use in the calculation of variances, standard deviations, and covariances.

Table 12.1 on page 280 shows the possible values for divisor and associated divisors where \( k \) is the number of PARTIAL statement variables.

Table 12.1 Possible Values for VARDEF=

<table>
<thead>
<tr>
<th>Value</th>
<th>Divisor</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>DF</td>
<td>degrees of freedom</td>
<td>( n - k - 1 )</td>
</tr>
<tr>
<td>N</td>
<td>number of observations</td>
<td>( n )</td>
</tr>
<tr>
<td>WDF</td>
<td>sum of weights minus one</td>
<td>( \sum w_i - k - 1 )</td>
</tr>
<tr>
<td>WEIGHT</td>
<td>sum of weights</td>
<td>( \sum w_i )</td>
</tr>
</tbody>
</table>

The procedure computes the variance as \( CSS/d\text{visor} \), where \( CSS \) is the corrected sums of squares and equals \( \sum (x_i - \bar{x})^2 \). When you weight the analysis variables, \( CSS \) equals \( \sum w_i (x_i - \bar{w})^2 \), where \( \bar{w} \) is the weighted mean.

Default: DF

Tip: When you use the WEIGHT statement and VARDEF=DF, the variance is an estimate of \( \sigma^2 \), where the variance of the \( i \)th observation is \( \text{var}(x_i) = \sigma^2/w_i \) and \( w_i \) is the weight for the \( i \)th observation. This yields an estimate of the variance of an observation with unit weight.

Tip: When you use the WEIGHT statement and VARDEF=WGT, the computed variance is asymptotically (for large \( n \)) an estimate of \( \sigma^2/\bar{w} \), where \( \bar{w} \) is the average weight. This yields an asymptotic estimate of the variance of an observation with average weight.

Main discussion: Weighted statistics “Example” on page 68.
**BY Statement**

Calculates separate correlation statistics for each BY group.

Main discussion: “BY” on page 62

```
BY <DESCENDING> variable1 <...<DESCENDING> variablen><NOTSORTED>;
```

**Required Arguments**

*variable*

specifies the variable that the procedure uses to form BY groups. You can specify more than one variable. If you do not use the NOTSORTED option in the BY statement, the observations in the data set must either be sorted by all the variables that you specify, or they must be indexed appropriately. Variables in a BY statement are called BY variables.

**Options**

**DESCENDING**

specifies that the observations are sorted in descending order by the variable that immediately follows the word DESCENDING in the BY statement.

**NOTSORTED**

specifies that observations are not necessarily sorted in alphabetic or numeric order. The observations are grouped in another way, for example, chronological order.

The requirement for ordering or indexing observations according to the values of BY variables is suspended for BY-group processing when you use the NOTSORTED option. In fact, the procedure does not use an index if you specify NOTSORTED. The procedure defines a BY group as a set of contiguous observations that have the same values for all BY variables. If observations with the same values for the BY variables are not contiguous, the procedure treats each contiguous set as a separate BY group.

**FREQ Statement**

Treats observations as if they appear multiple times in the input data set.

**Tip:** The effects of the FREQ and WEIGHT statements are similar except when calculating degrees of freedom.

**See also:** For an example that uses the FREQ statement, see “FREQ” on page 64

```
FREQ variable;
```
Required Arguments

variable

specifies a numeric variable whose value represents the frequency of the observation. If you use the FREQ statement, the procedure assumes that each observation represents \( n \) observations, where \( n \) is the value of variable. If \( n \) is not an integer, the SAS System truncates it. If \( n \) is less than 1 or is missing, the procedure does not use that observation to calculate statistics.

The sum of the frequency variable represents the total number of observations.

PARTIAL Statement

Computes Pearson partial correlation, Spearman partial rank-order correlation, or Kendall’s partial tau-b.

Restriction: Not valid with the HOEFFDING option.
Interaction: Invokes the NOMISS option to exclude all observations with missing values.
Main discussion: “Partial Correlation” on page 290
Featured in: Example 4 on page 308

PARTIAL variable(s);

Required Arguments

variable(s)

identifies one or more variables to use in the calculation of partial correlation statistics.

Details

- If you use the PEARSON option, PROC CORR also prints the partial variance and standard deviation for each VAR or WITH statement variable.
- If you use the KENDALL option, PROC CORR cannot compute probability values for Kendall’s partial tau-b.

VAR Statement

Specifies the variables to use to calculate correlation statistics.

Default: If you omit this statement, PROC CORR computes correlations for all numeric variables not listed in the other statements.
Featured in: Example 1 on page 299 and Example 2 on page 302
**VAR** variable(s);

**Required Arguments**

variable(s)

identifies one or more variables to use in the calculation of correlation coefficients.

---

**WEIGHT Statement**

Specifies weights for the analysis variables in the calculation of Pearson weighted product-moment correlation.

Restriction: Not valid with the HOEFFDING, KENDALL, or SPEARMAN option.

See also: For information on calculating weighted correlations, see “Pearson Product-Moment Correlation” on page 287.

**WEIGHT** variable;

**Required Arguments**

variable

specifies a numeric variable to use to compute weighted product-moment correlation coefficients. The values of the variable does not have to be integers. If the value of the weight variable is

<table>
<thead>
<tr>
<th>Weight value...</th>
<th>PROC CORR...</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>counts the observation in the total number of observations</td>
</tr>
<tr>
<td>less than 0</td>
<td>converts the value to zero and counts the observation in the total number of observations</td>
</tr>
<tr>
<td>missing</td>
<td>converts the value to zero and counts the observation in the total number of observations</td>
</tr>
</tbody>
</table>

**Tip:** When you use the WEIGHT statement, consider which value of the VARDEF= option is appropriate. See the discussion of the VARDEF= option on page 280 for more information.
WITH Statement

Determines the variables to use in conjunction with the VAR statement variables to calculate limited combinations of correlation coefficients.

Restriction: Not valid with the ALPHA option.

Featured in: Example 2 on page 302

WITH variable(s);

Required Argument

variable(s)

lists one or more variables to obtain correlations for specific combinations of variables. The WITH statement variables appear down the side of the correlation matrix and the VAR statement variables appear across the top of the correlation matrix. PROC CORR computes the following correlations for the VAR statement variables A and B and the WITH statement variables X, Y, and Z:

- X and A
- Y and A
- Z and A
- X and B
- Y and B
- Z and B

Concepts

Interpreting Correlation Coefficients

Correlation coefficients contain information on both the strength and direction of a linear relationship between two numeric random variables. If one variable x is an exact linear function of another variable y, a positive relationship exists when the correlation is 1 and an inverse relationship exists when the correlation is -1. If there is no linear predictability between the two variables, the correlation is 0. If the variables are normal and correlation is 0, the two variables are independent. However, correlation does not imply causality because, in some cases, an underlying causal relationship may exist.

The scatterplots in Figure 12.1 on page 285 depict the relationship between two numeric random variables.
When the relationship between two variables is nonlinear or when outliers are present, the correlation coefficient incorrectly estimates the strength of the relationship. Plotting the data before computing a correlation coefficient enables you to verify the linear relationship and to identify the potential outliers.

**Determining Computer Resources**

The only factor limiting the number of variables that you can analyze is the amount of available memory. The computer resources that PROC CORR requires depend on which statements and options you specify. To determine the computer resources that you need, use

\[
\begin{aligned}
N & \quad \text{number of observations in the data set.} \\
C & \quad \text{number of correlation types (1 to 4).} \\
V & \quad \text{number of VAR statement variables.} \\
W & \quad \text{number of WITH statement variables.} \\
P & \quad \text{number of PARTIAL statement variables.} \\
\end{aligned}
\]

so that

\[
\begin{aligned}
T & \quad V+W+P \\
K & \quad \begin{cases} 
V*W & \text{when } W>0 \\
V*(V+1)/2 & \text{when } W=0 
\end{cases}
\end{aligned}
\]
\[ L = K \quad \text{when } P=0 \]
\[ T*(T+1)/2 \quad \text{when } P>0 \]

For small \( N \) and large \( K \), the CPU time varies as \( K \) for all types of correlations. For large \( N \), the CPU time depends on the type of correlation. To calculate CPU time use

\[ K*N \quad \text{with PEARSON (default)} \]
\[ T*N*\log N \quad \text{with SPEARMAN} \]
\[ K*N*\log N \quad \text{with HOEFFDING or KENDALL} \]

You can reduce CPU time by specifying NOMISS. Without NOMISS, processing is much faster when most observations do not contain missing values.

The options and statements you use in the procedure require different amounts of storage to process the data. For Pearson correlations, the amount of temporary storage in bytes (\( M \)) is

\[ 40T+16L \quad \text{with NOMISS and NOSIMPLE} \]
\[ 40T+16L+56T \quad \text{with NOMISS} \]
\[ 40T+16L+56K \quad \text{with NOSIMPLE} \]
\[ 40T+16L+56K+56T \quad \text{with no options} \]

Using a PARTIAL statement increases the amount of temporary storage by \( 12T \) bytes. Using the ALPHA option increases the amount of temporary storage by \( 32V+16 \) bytes.

The following example uses a PARTIAL statement, which invokes NOMISS.

```sas
proc corr;
  var x1 x2;
  with y1 y2 y3;
  partial z1;
```

Therefore, using \( 40T+16L+56T+12T \), the minimum temporary storage equals 984 bytes \((T=2+3+1 \text{ and } L=T(T+1)/2)\).

Using the SPEARMAN, KENDALL, or HOEFFDING option requires additional temporary storage for each observation. For the most time-efficient processing, the amount of temporary storage in bytes is

\[ 40T+8K+8\times C+12T+N+28N+QS+QP+QK \]

where

\[ QS = \begin{cases} 
0 & \text{with NOSIMPLE} \\
68T & \text{otherwise} 
\end{cases} \]
\[ QP = \begin{cases} 
56K & \text{with PEARSON and without NOMISS} \\
0 & \text{otherwise} 
\end{cases} \]
\[ QK = \begin{cases} 
32N & \text{with KENDALL or HOEFFDING} \\
0 & \text{otherwise.} 
\end{cases} \]

The following example uses KENDALL:

```sas
proc corr kendall;
  var x1 x2 x3;
```
Therefore, the minimum temporary storage in bytes is
\[40 \times 3 + 8 \times 6 + 8 \times 6 \times 1 + 12 \times 3N + 28N + 3 \times 68 + 32N = 420 + 96N\]
where \(N\) is the number of observations.

If \(M\) bytes are not available, PROC CORR must process the data multiple times to compute all the statistics. This reduces the minimum temporary storage you need by \(12(T-2)N\) bytes. When this occurs, PROC CORR prints a note suggesting a larger memory region.

## Statistical Computations

PROC CORR computes several parametric and nonparametric correlation statistics as measures of association. The formulas for computing these measures and the associated probabilities follow.

### Pearson Product-Moment Correlation

The Pearson product-moment correlation is a parametric measure of association for two continuous random variables. The formula for the true Pearson product-moment correlation, denoted \(\rho_{xy}\), is

\[
\rho_{xy} = \frac{\text{cov}(x, y)}{\sqrt{\text{var}(x) \cdot \text{var}(y)}}
\]

\[= \frac{E((x - E(x))(y - E(y)))}{\sqrt{E(x - E(x))^2 \cdot E(y - E(y))^2}}\]

The sample correlation, such as a Pearson product-moment correlation or weighted product-moment correlation, estimates the true correlation. The formula for the Pearson product-moment correlation is

\[
r_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \cdot \sum (y_i - \bar{y})^2}}
\]

where \(\bar{x}\) is the sample mean of \(x\) and \(\bar{y}\) is the sample mean of \(y\).

The formula for a weighted Pearson product-moment correlation is

\[
r_{xy} = \frac{\sum w_i (x_i - \bar{x}_w)(y_i - \bar{y}_w)}{\sqrt{\sum w_i (x_i - \bar{x}_w)^2 \cdot \sum w_i (y_i - \bar{y}_w)^2}}
\]

where
Note that $\bar{x}_w$ is the weighted mean of $x$, $\bar{y}_w$ is the weighted mean of $y$, and $w_i$ is the weight.

When one variable is dichotomous (0,1) and the other variable is continuous, a Pearson correlation is equivalent to a point biserial correlation. When both variables are dichotomous, a Pearson correlation coefficient is equivalent to the phi coefficient.

**Spearman Rank-Order Correlation**

Spearman rank-order correlation is a nonparametric measure of association based on the rank of the data values. The formula is

$$
\theta = \frac{\sum (R_i - \bar{R}) (S_i - \bar{S})}{\sqrt{\sum (R_i - \bar{R})^2 \sum (S_i - \bar{S})^2}}
$$

where $R_i$ is the rank of the $i$th $x$ value, $S_i$ is the rank of the $i$th $y$ value, $\bar{R}$ is the mean of the $R_i$ values, and $\bar{S}$ is the mean of the $S_i$ values.

PROC CORR computes the Spearman’s correlation by ranking the data and using the ranks in the Pearson product-moment correlation formula. In case of ties, the averaged ranks are used.

**Kendall’s tau-b**

Kendall’s tau-b is a nonparametric measure of association based on the number of concordances and discordances in paired observations. Concordance occurs when paired observations vary together, and discordance occurs when paired observations vary differently. The formula for Kendall’s tau-b is

$$
\tau = \frac{\sum_{i<j} sgn (x_i - x_j) sgn (y_i - y_j)}{\sqrt{(T_0 - T_1)(T_0 - T_2)}}
$$

where

$$
T_0 = n (n - 1) / 2
$$
$$
T_1 = \sum t_i (t_i - 1) / 2
$$
$$
T_2 = \sum u_i (u_i - 1) / 2
$$
and where \( t_i \) is the number of tied \( x \) values in the \( i \)th group of tied \( x \) values, \( u_i \) is the number of tied \( y \) values in the \( i \)th group of tied \( y \) values, \( f \) is the number of observations, and \( \text{sgn}(z) \) is defined as

\[
\text{sgn}(z) = \begin{cases} 
 1 & \text{if } z > 0 \\
 0 & \text{if } z = 0 \\
 -1 & \text{if } z < 0 
\end{cases}
\]

PROC CORR computes Kendall's correlation by ranking the data and using a method similar to Knight (1966). The data are double sorted by ranking observations according to values of the first variable and reranking the observations according to values of the second variable. PROC CORR computes Kendall's tau-b from the number of interchanges of the first variable and corrects for tied pairs (pairs of observations with equal values of \( X \) or equal values of \( Y \)).

### Hoeffding's Measure of Dependence, \( D \)

Hoeffding's measure of dependence, \( D \), is a nonparametric measure of association that detects more general departures from independence. The statistic approximates a weighted sum over observations of chi-square statistics for two-by-two classification tables (Hoeffding 1948). Each set of \((x, y)\) values are cut points for the classification. The formula for Hoeffding's \( D \) is

\[
D = 30 \frac{(n - 2)(n - 3)D_1 + D_2 - 2(n - 2)D_3}{n(n - 1)(n - 2)(n - 3)(n - 4)}
\]

where

\[
D_1 = \sum_i (Q_i - 1)(Q_i - 2)
\]

\[
D_2 = \sum_i (R_i - 1)(R_i - 2)(S_i - 1)(S_i - 2)
\]

\[
D_3 = \sum_i (R_i - 2)(S_i - 2)(Q_i - 1)
\]

\( R_i \) is the rank of \( x_i \), \( S_i \) is the rank of \( y_i \), and \( Q_i \) (also called the bivariate rank) is 1 plus the number of points with both \( x \) and \( y \) values less than the \( i \)th point. A point that is tied on only the \( x \) value or \( y \) value contributes 1/2 to \( Q_i \) if the other value is less than the corresponding value for the \( i \)th point. A point that is tied on both \( x \) and \( y \) contributes 1/4 to \( Q_i \).

PROC CORR obtains the \( Q_i \) values by first ranking the data. The data are then double sorted by ranking observations according to values of the first variable and reranking the observations according to values of the second variable. Hoeffding's \( D \) statistic is computed using the number of interchanges of the first variable.

When no ties occur among data set observations, the \( D \) statistic values are between -0.5 and 1, with 1 indicating complete dependence. However, when ties occur, the \( D \) statistic may result in a smaller value. That is, for a pair of variables with identical
values, the Hoeffding's D statistic may be less than 1. With a large number of ties in a small data set, the D statistic may be less than -0.5. For more information on Hoeffding's D, see Hollander and Wolfe (1973, p. 228).

---

Partial Correlation

A partial correlation measures the strength of a relationship between two variables, while controlling the effect of one or more additional variables. The Pearson partial correlation for a pair of variables may be defined as the correlation of errors after regression on the controlling variables. Let \( y = (y_1, y_2, \ldots, y_n) \) be the set of variables to correlate. Also let \( \alpha \) and \( \beta \) be sets of regression parameters and \( z \) be the set of controlling variables, where \( \alpha = (\alpha_1, \alpha_2, \ldots, \alpha_c) \), \( \beta \) is the slope, and \( z = (z_1, z_2, \ldots, z_p) \). Suppose

\[
E(y) = \alpha + z\beta
\]

is a regression model for \( y \) given \( z \). The population Pearson partial correlation between the \( i \)th and the \( j \)th variables of \( y \) given \( z \) is defined as the correlation between errors \( (y_i - E(y_i)) \) and \( (y_j - E(y_j)) \).

If the exact values of \( \alpha \) and \( \beta \) are unknown, you can use a sample Pearson partial correlation to estimate the population Pearson partial correlation. For a given sample of observations, you estimate the sets of unknown parameters \( \alpha \) and \( \beta \) using the least-squares estimators \( \hat{\alpha} \) and \( \hat{\beta} \). Then the fitted least-squares regression model is

\[
\hat{y} = \hat{\alpha} + z\hat{\beta}
\]

The partial corrected sums of squares and crossproducts (CSSCP) of \( y \) given \( z \) are the corrected sums of squares and crossproducts of the residuals \( y - \hat{y} \). Using these partial corrected sums of squares and crossproducts, you can calculate the partial variances, partial covariances, and partial correlations.

PROC CORR derives the partial corrected sums of squares and crossproducts matrix by applying the Cholesky decomposition algorithm to the CSSCP matrix. For Pearson partial correlations, let \( S \) be the partitioned CSSCP matrix between two sets of variables, \( z \) and \( y \):

\[
S = \begin{bmatrix}
S_{yx} & S_{xy} \\
S_{yy} & S_{yy}
\end{bmatrix}
\]

PROC CORR calculates \( S_{yxy} \), the partial CSSCP matrix of \( y \) after controlling for \( z \), by applying the Cholesky decomposition algorithm sequentially on the rows associated with \( z \), the variables being partialled out.

After applying the Cholesky decomposition algorithm to each row associated with variables \( z \), PROC CORR checks all higher numbered diagonal elements associated with \( z \) for singularity. After the Cholesky decomposition, a variable is considered singular if the value of the corresponding diagonal element is less than \( p \) times the original unpartialled corrected sum of squares of that variable. You can specify the singularity criterion \( p \) using the SINGULAR= option. For Pearson partial correlations, a controlling variable \( z \) is considered singular if the \( R^2 \) for predicting this variable from
the variables that are already partialled out exceeds $1 - p$. When this happens, PROC CORR excludes the variable from the analysis. Similarly, a variable is considered singular if the $R^2$ for predicting this variable from the controlling variables exceeds $1 - p$. When this happens, its associated diagonal element and all higher numbered elements in this row or column are set to zero.

After the Cholesky decomposition algorithm is performed on all rows associated with $z$, the resulting matrix has the form

$$\begin{bmatrix} T_{zz} & T_{zy} \\ 0 & S_{yy\cdot z} \end{bmatrix}$$

where $T_{zz}$ is an upper triangular matrix with

$$T'_{zz} T_{zz} = S_{zz'}$$
$$T'_{zz} T_{zy} = S_{zy'}$$
$$S_{yy\cdot z} = S_{yy} - T'_{zy} T_{zy}.$$ 

If $S_{zz}$ is positive definite, then the partial CSSCP matrix $S_{yy\cdot z}$ is identical to the matrix derived from the formula

$$S_{yy\cdot z} = S_{yy} - S'_{zy} S^{-1}_{xz} S_{zy}$$

The partial variance-covariance matrix is calculated with the variancedivisor (VARDEF= option). PROC CORR can then use the standard Pearson correlation formula on the partial variance-covariance matrix to calculate the Pearson partial correlation matrix. Another way to calculate Pearson partial correlation is by applying the Cholesky decomposition algorithm directly to the correlation matrix and by using the correlation formula on the resulting matrix.

To derive the corresponding Spearman partial rank-order correlations and Kendall partial tau-b correlations, PROC CORR applies the Cholesky decomposition algorithm to the Spearman rank-order correlation matrix and Kendall tau-b correlation matrix and uses the correlation formula. The singularity criterion for nonparametric partial correlations is identical to Pearson partial correlation except that PROC CORR uses a matrix of nonparametric correlations and sets a singular variable's associated correlations to missing. The partial tau-b correlations range from –1 to 1. However, the sampling distribution of this partial tau-b is unknown; therefore, the probability values are not available.

When a correlation matrix (Pearson, Spearman, or Kendall tau-b correlation matrix) is positive definite, the resulting partial correlation between variables $x$ and $y$ after adjusting for a single variable $z$ is identical to that obtained from the first-order partial correlation formula

$$r_{xy\cdot z} = \frac{r_{xy} - r_{xz} r_{yz}}{\sqrt{(1 - r_{xz}^2)(1 - r_{yz}^2)}}$$

where $r_{xy}, r_{xz},$ and $r_{yz}$ are the appropriate correlations.

The formula for higher-order partial correlations is a straightforward extension of the above first-order formula. For example, when the correlation matrix is positive
definite, the partial correlation between $x$ and $y$, controlling for both $z_1$ and $z_2$ is identical to the second-order partial correlation formula

$$r_{xy|z_1z_2} = \frac{r_{xy|z_1} - r_{xz_1z_1}r_{yz_1z_1}}{\sqrt{(1 - r_{xz_1z_1}^2)(1 - r_{yz_1z_1}^2)}}$$

where $r_{xy|z_1}$, $r_{xz_1z_1}$, and $r_{yz_1z_1}$ are first-order partial correlations among variables $x$, $y$, and $z_2$ given $z_1$.

---

**Cronbach’s Coefficient Alpha**

Analyzing latent constructs such as job satisfaction, motor ability, sensory recognition, or customer satisfaction requires instruments to accurately measure the constructs. Interrelated items may be summed to obtain an overall score for each participant. Cronbach’s coefficient alpha estimates the reliability of this type of scale by determining the internal consistency of the test or the average correlation of items within the test (Cronbach 1951).

When a value is recorded, the observed value contains some degree of measurement error. Two sets of measurements on the same variable for the same individual may not have identical values. However, repeated measurements for a series of individuals will show some consistency. Reliability measures internal consistency from one set of measurements to another. The observed value $Y$ is divided into two components, a true value $T$ and a measurement error $E$. The measurement error is assumed to be independent of the true value, that is,

$$Y = T + E, \quad \text{cov}(T, E) = 0$$

The reliability coefficient of a measurement test is defined as the squared correlation between the observed value $Y$ and the true value $T$, that is,

$$\rho^2 (Y, T) = \frac{\text{cov}(Y, T)^2}{\text{var}(Y) \text{var}(T)}$$

$$= \frac{\text{var}(T)^2}{\text{var}(Y) \text{var}(T)}$$

$$= \frac{\text{var}(T)}{\text{var}(Y)}$$

which is the proportion of the observed variance due to true differences among individuals in the sample. If $Y$ is the sum of several observed variables measuring the same feature, you can estimate $\text{var}(T)$. Cronbach’s coefficient alpha, based on a lower bound for $\text{var}(T)$, is an estimate of the reliability coefficient.

Suppose $p$ variables are used with $Y_j = T_j + E_j$ for $j = 1, 2, \ldots, p$, where $Y_j$ is the observed value, $T_j$ is the true value, and $E_j$ is the measurement error. The measurement errors ($E_j$) are independent of the true values ($T_j$) and are also independent of each other. Let $Y_0 = \sum Y_j$ be the total observed score and $T_0 = \sum T_j$ be the total true score. Because
The CORR Procedure △ Cronbach’s Coefficient Alpha 293

\[(p - 1) \sum \text{var}(T_j) \geq \sum_{i \neq j} \text{cov}(T_i, T_j),\]

a lower bound for \(\text{var}(T_0)\) is given by

\[\frac{p}{p - 1} \sum_{i \neq j} \text{cov}(T_i, T_j)\]

With \(\text{cov}(Y_i, Y_j) = \text{cov}(T_i, T_j)\) for \(i \neq j\), a lower bound for the reliability coefficient is then given by the Cronbach’s coefficient alpha:

\[
\alpha = \left(\frac{p}{p - 1}\right) \frac{\sum_{i \neq j} \text{cov}(Y_i, Y_j)}{\text{var}(Y_0)}
\]

\[= \left(\frac{p}{p - 1}\right) \left(1 - \frac{\sum_{j} \text{var}(Y_j)}{\text{var}(Y_0)}\right)\]

If the variances of the items vary widely, you can standardize the items to a standard deviation of 1 before computing the coefficient alpha. If the variables are dichotomous (0,1), the coefficient alpha is equivalent to the Kuder-Richardson 20 (KR-20) reliability measure.

When the correlation between each pair of variables is 1, the coefficient alpha has a maximum value of 1. With negative correlations between some variables, the coefficient alpha can have a value less than zero. The larger the overall alpha coefficient, the more likely that items contribute to a reliable scale. Nunnally (1978) suggests .70 as an acceptable reliability coefficient; smaller reliability coefficients are seen as inadequate. However, this varies by discipline.

To determine how each item reflects the reliability of the scale, you calculate a coefficient alpha after deleting each variable independently from the scale. The Cronbach’s coefficient alpha from all variables except the \(k\)th variable is given by

\[
\alpha_k = \left(\frac{p - 1}{p - 2}\right) \left(1 - \frac{\sum_{i \neq k} \text{var}(Y_i)}{\text{var}\left(\sum_{i \neq k} Y_i\right)}\right)
\]

If the reliability coefficient increases after deleting an item from the scale, you can assume that the item is not correlated highly with other items in the scale. Conversely, if the reliability coefficient decreases you can assume that the item is highly correlated with other items in the scale. See SAS Communications, 4th Quarter 1994, for more information on how to interpret Cronbach’s coefficient alpha.

Listwise deletion of observations with missing values is necessary to correctly calculate Cronbach’s coefficient alpha. PROC CORR does not automatically use listwise
deletion when you specify ALPHA. Therefore, use the NOMISS option if the data set contains missing values. Otherwise, PROC FREQ prints a warning message in the SAS log indicating the need to use NOMISS with ALPHA.

---

**Probability Values**

Probability values for the Pearson and Spearman correlations are computed by treating

\[
\frac{(n - 2)^{1/2} r}{(1 - r^2)^{1/2}}
\]

as coming from a t distribution with \( n - 2 \) degrees of freedom, where \( r \) is the appropriate correlation.

Probability values for the Pearson and Spearman partial correlations are computed by treating

\[
\frac{(n - k - 2)^{1/2} r}{(1 - r^2)^{1/2}}
\]

as coming from a t distribution with \( n - k - 2 \) degrees of freedom, where \( r \) is the appropriate partial correlation and \( k \) is the number of variables being partialled out.

Probability values for Kendall correlations are computed by treating

\[
\frac{s}{\sqrt{\text{var}(s)}}
\]

as coming from a normal distribution when

\[
s = \sum_{i<j} \text{sgn}(x_i - x_j) \text{sgn}(y_i - y_j)
\]

and where \( x_i \) are the values of the first variable, \( y_i \) are the values of the second variable, and the function \( \text{sgn}(z) \) is defined as

\[
\text{sgn}(z) = \begin{cases} 
1 & \text{if } z > 0 \\
0 & \text{if } z = 0 \\
-1 & \text{if } z < 0 
\end{cases}
\]

The formula for the variance of \( s \), \( \text{var}(s) \), is computed as

\[
\text{var}(s) = \frac{v_0 - v_1 - v_2}{18} + \frac{v_1}{2n(n-1)} + \frac{v_2}{9n(n-1)(n-2)}
\]
where
\[ v_0 = n(n-1)(2n+1) \]
\[ v_t = \sum t_i (t_i - 1) (2t_i + 5) \]
\[ v_u = \sum u_i (u_i - 1) (2u_i + 5) \]
\[ v_1 = (\sum t_i (t_i - 1)) (\sum u_i (u_i - 1)) \]
\[ v_2 = (\sum t_i (t_i - 1) (t_i - 2)) (\sum u_i (u_i - 1) (u_i - 2)) \]

The sums are over tied groups of values where \( t_i \) is the number of tied \( x \) values and
\( u_i \) is the number of tied \( y \) values (Noether 1967). The sampling distribution of
Kendall's partial tau-b is unknown; therefore, the probability values are not available.

The probability values for Hoeffding's D statistic are computed using the asymptotic
distribution computed by Blum, Kiefer, and Rosenblatt (1961). The formula is
\[
\frac{(n-1)\pi^4}{60} D + \frac{\pi^4}{72}
\]
which comes from the asymptotic distribution. When the sample size is less than 10,
see the tables for the distribution of \( D \) in Hollander and Wolfe (1973).

**Results**

**Missing Values**

By default, PROC CORR uses pairwise deletion when observations contain missing
values. PROC CORR includes all nonmissing pairs of values for each pair of variables
in the statistical computations. Therefore, the correlations statistics may be based on
different numbers of observations.

If you specify the NOMISS option, PROC CORR uses listwise deletion when a value
of the BY, FREQ, VAR, WEIGHT, or WITH statement variable is missing. PROC CORR
excludes all observations with missing values from the analysis. Therefore, the number
of observations for each pair of variables is identical. The PARTIAL statement always
excludes the observations with missing values by automatically invoking NOMISS.
Listwise deletion is needed to correctly calculate Cronbach's coefficient alpha when data
are missing. If a data set contains missing values, when you specify ALPHA use the
NOMISS option.

There are two reasons to specify NOMISS and, thus, to avoid pairwise deletion.
First, NOMISS is computationally more efficient, so you use fewer computer resources.
Second, if you use the correlations as input to regression or other statistical procedures,
a pairwise-missing correlation matrix leads to several statistical difficulties. Pairwise
correlation matrices may not be nonnegative definite, and the pattern of missing values
may bias the results.

**Procedure Output**

By default, PROC CORR prints a report that includes descriptive statistics and
correlation statistics for each variable. The descriptive statistics include the number of
observations with nonmissing values, the mean, the standard deviation, the minimum, and the maximum. PROC CORR reports the following additional descriptive statistics when you request various correlation statistics:

- **sum**
  - for Pearson correlation only

- **median**
  - for nonparametric measures of association

- **partial variance**
  - for Pearson partial correlation

- **partial standard deviation**
  - for Pearson partial correlation.

If variable labels are available, PROC CORR labels the variables. When you specify the CSSCP, SSCP, or COV option, the appropriate sum-of-squares and crossproducts and covariance matrix appears at the top of the correlation report. If the data set contains missing values, PROC CORR prints additional statistics for each pair of variables. These statistics, calculated from the observations with nonmissing row and column variable values, may include:

- **SSCP(W',V')**
  - uncorrected sum-of-squares and crossproducts

- **USS(W')**
  - uncorrected sum-of-squares for the row variable

- **USS(V')**
  - uncorrected sum-of-squares for the column variable

- **CSSCP(W',V')**
  - corrected sum-of-squares and crossproducts

- **CSS(W')**
  - corrected sum-of-squares for the row variable

- **CSS(V')**
  - corrected sum-of-squares for the column variable

- **COV (W',V')**
  - covariance

- **VAR (W')**
  - variance for the row variable

- **VAR (V')**
  - variance for the column variable

- **DF(W',V')**
  - divisor for calculating covariance and variances.

For each pair of variables, PROC CORR always prints the correlation coefficients, the number of observations used to calculate the coefficient, and the significance probability. When you specify the ALPHA option, PROC CORR prints Cronbach's coefficient alpha,
the correlation between the variable and the total of the remaining variables, and Cronbach’s coefficient alpha using the remaining variables for the raw variables and the standardized variables.

---

**Output Data Sets**

When you specify the OUTP=, OUTS=, OUTK=, or OUTH= option, PROC CORR creates an output data set containing statistics for Pearson correlation, Spearman correlation, Kendall correlation, or Hoeffding’s D, respectively. By default, the output data set is a special data set type (TYPE=CORR) that many SAS/STAT procedures recognize, including PROC REG and PROC FACTOR. When you specify the NOCORR option and the COV, CSSCP, or SSCP option, use the TYPE= data set option to change the data set type to COV, CSSCP, or SSCP. For example, the following statement

```sas
proc corr nocorr cov outp=b(type=cov);
```

specifies the output data set type as COV.

PROC CORR does not print the output data set. Use PROC PRINT, PROC REPORT, or another SAS reporting tool to print the output data set.

The output data set includes the following variables

- **BY variables**
  - identifies the BY group when using a BY statement.

- **_TYPE_ variable**
  - identifies the type of observation.

- **_NAME_ variable**
  - identifies the variable that corresponds to a given row of the correlation matrix.

- **INTERCEP variable**
  - identifies variable sums when specifying the SSCP option.

- **VAR variables**
  - identifies the variables listed in the VAR statement.

You can use a combination of the _TYPE_ and _NAME_ variables to identify the contents of an observation. The _NAME_ variable indicates which row of the correlation matrix the observation corresponds to. The values of the _TYPE_ variable are

- **SSCP**
  - uncorrected sums of squares and crossproducts

- **CSSCP**
  - corrected sums of squares and crossproducts

- **COV**
  - covariances

- **MEAN**
  - mean of each variable

- **STD**
  - standard deviation of each variable
N
number of nonmissing observations for each variable

SUMWGT
sum of the weights for each variable when using a WEIGHT statement

CORR
correlation statistics for each variable.

When you specify the SSCP option, the OUTP= data set includes an additional observation that contains intercept values. When you specify the ALPHA option, the OUTP= data set also includes observations with the following _TYPE_ values:

RAWALPHA
Cronbach’s coefficient alpha for raw variables

STDALPHA
Cronbach’s coefficient alpha for standardized variables

RAWALDEL
Cronbach’s coefficient alpha for raw variables after deleting one variable

STDALDEL
Cronbach’s coefficient alpha for standardized variables after deleting one variable

RAWCTDEL
correlation between a raw variable and the total of the remaining raw variables

STDCTDEL
correlation between a standardized variable and the total of the remaining standardized variables.

When you use a PARTIAL statement, the previous statistics are calculated for the variables after partialling. If PROC CORR computes Pearson correlation statistics, MEAN equals zero and STD equals the partial standard deviation associated with the partial variance for the OUTP=, OUTK=, or OUTS= data set. Otherwise, PROC CORR assigns missing values to MEAN and STD. Output 12.4 on page 298 lists the observations in an OUTP= data set when the COV option and PARTIAL statement are used to compute Pearson partial correlations. The _TYPE_ variable identifies COV, MEAN, STD, N, and CORR as the statistical values for the variables Weight, Oxygen, and Runtime. MEAN always equals 0, while STD is a partial standard deviation.

Output 12.4  OUTP= Data Set with Pearson Partial Correlations

<table>
<thead>
<tr>
<th><em>TYPE</em></th>
<th><em>NAME</em></th>
<th>Weight</th>
<th>Oxygen</th>
<th>Runtime</th>
</tr>
</thead>
<tbody>
<tr>
<td>COV</td>
<td>Weight</td>
<td>72.4374</td>
<td>-12.7511</td>
<td>2.0677</td>
</tr>
<tr>
<td>COV</td>
<td>Oxygen</td>
<td>-12.7511</td>
<td>27.0165</td>
<td>-5.5937</td>
</tr>
<tr>
<td>COV</td>
<td>Runtime</td>
<td>2.0677</td>
<td>-5.5937</td>
<td>1.9451</td>
</tr>
<tr>
<td>MEAN</td>
<td></td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>STD</td>
<td></td>
<td>8.5110</td>
<td>5.1977</td>
<td>1.3947</td>
</tr>
<tr>
<td>N</td>
<td></td>
<td>28.0000</td>
<td>28.0000</td>
<td>28.0000</td>
</tr>
<tr>
<td>CORR</td>
<td>Weight</td>
<td>1.0000</td>
<td>-0.2882</td>
<td>0.1742</td>
</tr>
<tr>
<td>CORR</td>
<td>Oxygen</td>
<td>-0.2882</td>
<td>1.0000</td>
<td>-0.7716</td>
</tr>
<tr>
<td>CORR</td>
<td>Runtime</td>
<td>0.1742</td>
<td>-0.7716</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
Examples

Example 1: Computing Pearson Correlations and Other Measures of Association

Procedure features:

PROC CORR statement options:

HOEFFDING
PEARSON
SPEARMAN
VAR statement

This example

- produces a correlation analysis with descriptive statistics, Pearson product-moment correlation, Spearman rank-order correlation, and Hoeffding's measure of dependence, D
- selects the analysis variables.

Program

```sas
options nodate pageno=1 linesize=80 pagesize=60;
```

The data set FITNESS contains measurements from a study of physical fitness for 30 participants between the ages 38 and 57. Each observation represents one person. Two observations contain missing values.

```sas
data fitness;
  input Age Weight Runtime Oxygen @@;
datalines;
57 73.37 12.63 39.407 54 79.38 11.17 46.080
52 76.32 9.63 45.441 50 70.87 8.92 .
51 67.25 11.08 45.118 54 91.63 12.88 39.203
51 73.71 10.47 45.790 57 59.08 9.93 50.545
49 76.32 . 48.673 48 61.24 11.5 47.920
45 82.78 10.5 47.467 44 73.03 10.13 50.541
45 87.66 14.03 37.388 45 66.45 11.12 44.754
47 79.15 10.6 47.273 54 83.12 10.33 51.855
49 81.42 8.95 40.836 51 77.91 10.00 46.672
48 91.63 10.25 46.774 49 73.37 10.08 50.388
44 89.47 11.37 44.609 40 75.07 10.07 45.313
```
PEARSON, SPEARMAN, and HOEFFDING compute correlation statistics. When you request nonparametric correlations, specify PEARSON to compute Pearson correlations.

```plaintext
proc corr data=fitness pearson spearman hoeffding;
```

The VAR statement specifies the analysis variables and the order to print them.

```plaintext
var weight oxygen runtime;
```

The TITLE statement specifies a title for the report.

```plaintext
title 'Measures of Association for';
title2 'a Physical Fitness Study';
run;
```

Output

The correlation report includes descriptive statistics, Pearson’s rho, Spearman’s rho, and Hoeffding’s D. The report uses the median, instead of the sum, as a descriptive measure when PROC CORR computes nonparametric measures of association. Because missing data are excluded pairwise, the number of observations PROC CORR uses to calculate the correlation coefficients varies.
Measures of Association for a Physical Fitness Study

The CORR Procedure

3 Variables: Weight Oxygen Runtime

Simple Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Median</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight</td>
<td>30</td>
<td>77.70500</td>
<td>8.34152</td>
<td>77.68000</td>
<td>59.08000</td>
<td>91.63000</td>
</tr>
<tr>
<td>Oxygen</td>
<td>29</td>
<td>47.06445</td>
<td>5.32129</td>
<td>46.67200</td>
<td>37.38800</td>
<td>60.05500</td>
</tr>
<tr>
<td>Runtime</td>
<td>29</td>
<td>10.61448</td>
<td>1.41655</td>
<td>10.47000</td>
<td>8.17000</td>
<td>14.03000</td>
</tr>
</tbody>
</table>

Pearson Correlation Coefficients

<table>
<thead>
<tr>
<th></th>
<th>Weight</th>
<th>Oxygen</th>
<th>Runtime</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight</td>
<td>1.00000</td>
<td>-0.19900</td>
<td>0.15155</td>
</tr>
<tr>
<td></td>
<td>0.3007</td>
<td>0.4326</td>
<td></td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>Oxygen</td>
<td>-0.19900</td>
<td>1.00000</td>
<td>-0.78346</td>
</tr>
<tr>
<td></td>
<td>0.3007</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>29</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>Runtime</td>
<td>0.15155</td>
<td>-0.78346</td>
<td>1.00000</td>
</tr>
<tr>
<td></td>
<td>0.4326</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>29</td>
<td>28</td>
<td></td>
</tr>
</tbody>
</table>

Spearman Correlation Coefficients

<table>
<thead>
<tr>
<th></th>
<th>Weight</th>
<th>Oxygen</th>
<th>Runtime</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight</td>
<td>1.00000</td>
<td>-0.13110</td>
<td>0.10546</td>
</tr>
<tr>
<td></td>
<td>0.4979</td>
<td>0.5861</td>
<td></td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>Oxygen</td>
<td>-0.13110</td>
<td>1.00000</td>
<td>-0.68363</td>
</tr>
<tr>
<td></td>
<td>0.4979</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>29</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>Runtime</td>
<td>0.10546</td>
<td>-0.68363</td>
<td>1.00000</td>
</tr>
<tr>
<td></td>
<td>0.5861</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>29</td>
<td>28</td>
<td></td>
</tr>
</tbody>
</table>

Measures of Association for a Physical Fitness Study

The CORR Procedure

Hoeffding Dependence Coefficients

<table>
<thead>
<tr>
<th></th>
<th>Weight</th>
<th>Oxygen</th>
<th>Runtime</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight</td>
<td>0.97559</td>
<td>-0.01789</td>
<td>-0.02418</td>
</tr>
<tr>
<td></td>
<td>&lt;.0001</td>
<td>0.9775</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>Oxygen</td>
<td>-0.01789</td>
<td>1.00000</td>
<td>0.16554</td>
</tr>
<tr>
<td></td>
<td>0.9775</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>29</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>Runtime</td>
<td>-0.02418</td>
<td>0.16554</td>
<td>1.00000</td>
</tr>
<tr>
<td></td>
<td>1.0000</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>29</td>
<td>28</td>
<td></td>
</tr>
</tbody>
</table>
Example 2: Computing Rectangular Correlation Statistics with Missing Data

Procedure features:

PROC CORR statement options:

- COV
- NOSIMPLE
- SSCP

VAR statement

WITH statement

This example:

- suppresses descriptive statistics
- prints uncorrected sum-of-squares and crossproducts
- calculates a rectangular covariance matrix
- calculates a rectangular correlation matrix
- excludes observations with missing values using pairwise deletion (default method).

Program

```sas
options nodate pageno=1 linesize=80 pagesize=60;
```

The data set SETOSA contains measurements for four iris parts: sepal length, sepal width, petal length, and petal width based on Fisher's iris data (1936). Fifty iris specimens from the species *Iris setosa* are used. Each observation represents one specimen. Three observations contain missing values. The LABEL statement associates a label with each variable.

data setosa;
  input SepalLength SepalWidth PetalLength PetalWidth @@;
  label sepallength='Sepal Length in mm.';
  sepalwidth='Sepal Width in mm.';
  petallength='Petal Length in mm.';
  petalwidth='Petal Width in mm.';
  datalines;
  50 33 14 02 46 34 14 03 46 36 . 02
  51 33 17 05 55 35 13 02 48 31 16 02
  52 34 14 02 49 36 14 01 44 32 13 02
  50 35 16 06 44 30 13 02 47 32 16 02
  48 30 14 03 51 38 16 02 48 34 19 02
  50 30 16 02 50 32 12 02 43 30 11 .
  58 40 12 02 51 38 19 04 49 30 14 02
  51 35 14 02 50 34 16 04 46 32 14 02
  57 44 15 04 50 36 14 02 54 34 15 04
  52 41 15 . 55 42 14 02 49 31 15 02
  54 39 17 04 50 34 15 02 44 29 14 02
  47 32 13 02 46 31 15 02 51 34 15 02
  50 35 13 03 49 31 15 01 54 37 15 02
```
SSCP displays the uncorrected sum-of-squares and crossproducts matrix and invokes PEARSON. COV calculates the covariance matrix. NOSIMPLE suppresses descriptive statistics.

```plaintext
proc corr data=setosa sscp cov nosimple;
```

The WITH statement together with the VAR statement produces a rectangular correlation matrix. The matrix rows are PetalLength and PetalWidth while the matrix columns are SepalLength and SepalWidth.

```plaintext
var sepallength sepalwidth;
with petallength petalwidth;
```

The TITLE statement specifies a title for the report.

```plaintext
title 'Fisher (1936) Iris Setosa Data';
run;
```

Output

The correlation report includes rectangular sum-of-squares and crossproducts, covariances, and the correlation matrix using the two WITH variables and two VAR variables. The descriptive statistics do not appear. PROC CORR uses variable labels to label matrix rows (WITH variables). PROC CORR calculates sum-of-squares and crossproducts and covariances statistics for each pair of variables by using observations with nonmissing row and column variable values. Because missing data are excluded pairwise, the number of observations PROC CORR uses to calculate the correlation coefficients changes.
Example 3: Computing Cronbach's Coefficient Alpha

Fisher (1936) Iris Setosa Data

The CORR Procedure

With Variables: PetalLength PetalWidth
Variables: SepalLength SepalWidth

Sums of Squares and Crossproducts
SSCP / Row Var SS / Col Var SS

<table>
<thead>
<tr>
<th></th>
<th>SepalLength</th>
<th>SepalWidth</th>
</tr>
</thead>
<tbody>
<tr>
<td>PetalLength</td>
<td>36214.00000</td>
<td>24756.00000</td>
</tr>
<tr>
<td>Petal Length in mm.</td>
<td>10735.00000</td>
<td>10735.00000</td>
</tr>
<tr>
<td></td>
<td>123793.0000</td>
<td>58164.0000</td>
</tr>
<tr>
<td>PetalWidth</td>
<td>6113.00000</td>
<td>4191.00000</td>
</tr>
<tr>
<td>Petal Width in mm.</td>
<td>355.00000</td>
<td>355.00000</td>
</tr>
<tr>
<td></td>
<td>121356.0000</td>
<td>56879.0000</td>
</tr>
</tbody>
</table>

Variances and Covariances
Covariance / Row Var Variance / Col Var Variance / DF

<table>
<thead>
<tr>
<th></th>
<th>SepalLength</th>
<th>SepalWidth</th>
</tr>
</thead>
<tbody>
<tr>
<td>PetalLength</td>
<td>1.270833333</td>
<td>1.363095238</td>
</tr>
<tr>
<td>Petal Length in mm.</td>
<td>2.625000000</td>
<td>2.625000000</td>
</tr>
<tr>
<td></td>
<td>12.33333333</td>
<td>14.60544218</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>48</td>
</tr>
<tr>
<td>PetalWidth</td>
<td>0.911347518</td>
<td>1.048315603</td>
</tr>
<tr>
<td>Petal Width in mm.</td>
<td>1.063386525</td>
<td>1.063386525</td>
</tr>
<tr>
<td></td>
<td>11.80141844</td>
<td>13.62721631</td>
</tr>
<tr>
<td></td>
<td>47</td>
<td>47</td>
</tr>
</tbody>
</table>

Pearson Correlation Coefficients
Prob > |r| under H0: Rho=0
Number of Observations

<table>
<thead>
<tr>
<th></th>
<th>SepalLength</th>
<th>SepalWidth</th>
</tr>
</thead>
<tbody>
<tr>
<td>PetalLength</td>
<td>0.22335</td>
<td>0.22014</td>
</tr>
<tr>
<td>Petal Length in mm.</td>
<td>0.1229</td>
<td>0.1285</td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>49</td>
</tr>
<tr>
<td>PetalWidth</td>
<td>0.25726</td>
<td>0.27539</td>
</tr>
<tr>
<td>Petal Width in mm.</td>
<td>0.0775</td>
<td>0.0582</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>48</td>
</tr>
</tbody>
</table>

Example 3: Computing Cronbach's Coefficient Alpha

Procedure features:
PROC CORR statement options:
ALPHA
NOCORR
NOMISS

This example
computes Cronbach’s coefficient alpha for a multiple-item mixed-rating scale
suppresses Pearson correlation statistics
excludes observations with missing values using listwise deletion.

This example does not examine the correlation matrix but assumes that all items are positively correlated. Normally, you want to examine the correlation and covariance matrices to make sure that all variables are positively correlated. Positive correlation is needed because items measure a common entity. You exclude negatively correlated items from the analysis because they do not measure the same construct.

Program

options nodate pageno=1 linesize=80 pagesize=60;

The data set PYSCHDAT contains responses to a questionnaire assessing the mental stability of 30 randomly selected female psychiatric patients. Each observation represents one patient. The scale includes seven items. The LABEL statement provides a label for each item. Seven observations contain missing values.

data psychdat;
  input Age Anxiety Depression Sleep Sex Life WeightChange @@;
  label age = 'age in years'
    anxiety = 'anxiety level'
    depression = 'depression level'
    sleep = 'normal sleep (1=y 2=n)'
    sex = 'sexual (1=n 2=y)'
    life = 'suicidal (1=n 2=y)'
    weightchange = 'recent weight change';
  datalines;
39 2 2 2 2 2 2 4.9 41 2 2 2 2 2 2 2 2 2 4.9 41 2 2 2 2 2 2 2 2 4.9 41 2 2 2 2 2 2 2 2 4.9 41 2 2 2 2 2 2 2 2 4.9 41 2 2 2 2 2 2 2 2 4.9 41 2 2 2 2 2 2 2 2 4.9 41 2 2 2 2 2 2 2 2 4.9 41 2 2 2 2 2 2 2 2 4.9 41 2 2 2 2 2 2 2 2 4.9 41 2 2 2 2 2 2
42 3 3 . 2 2 4.0 30 2 2 2 2 2 2 -2.6
35 2 1 1 2 1 -0.3 44 . 1 2 1 1 0.9
31 2 2 . 2 2 -1.5 39 3 2 2 2 1 3.5
35 3 2 2 2 2 -1.2 33 2 2 2 2 2 0.8
38 2 1 1 1 1 -1.9 31 2 2 2 . 1 5.5
40 3 2 2 2 2 1 2.7 44 2 2 2 2 2 4.4
43 3 2 2 2 2 3.2 32 1 1 1 2 1 -1.5
32 1 2 2 . 1 -1.9 43 4 3 2 2 2 8.3
46 3 2 2 2 2 3.6 30 2 2 2 2 1 1.4
34 3 3 . 2 2 . 37 3 2 2 2 1 .
35 2 1 2 2 1 -1.0 45 2 2 2 2 2 6.5
35 2 2 2 2 2 -2.1 31 2 2 2 2 1 -0.4
32 2 2 2 2 2 -1.9 44 2 2 2 2 2 3.7
40 3 3 2 2 2 4.5 42 3 3 2 2 2 4.2
;

ALPHA computes Cronbach’s alpha and invokes PEARSON. NOCORR suppresses Pearson correlation statistics. NOMISS excludes observations with missing values. Omitting a VAR statement causes PROC CORR to use all numeric variables.

```sas
proc corr data=psychdat alpha nocorr nomiss;
```

The TITLE statement specifies a title for the report.

```sas
title1 'Mental Stability Scale for Female Psychiatric Patients';
run;
```

**Output**

The correlation report includes descriptive statistics and Cronbach’s coefficient alpha, the correlation between the variable and the total of the remaining variables, and Cronbach’s coefficient alpha using the remaining variables for both the raw variables and the standardized variables. These calculations use the 23 observations without missing values.

Because the variances of some variables vary widely, you use the standardized scores to estimate reliability. The overall standardized alpha of .85 is an acceptable reliability coefficient. This is greater than Nunnally’s suggested value of .70.

The standardized alpha provides information on how each item reflects the reliability of the scale. Notice that the standardized alpha decreases after removing Depression from the construct. Therefore, this variable appears strongly correlated with other items in the scale. The standardized alpha increases slightly after removing Sex from the construct. Thus, removing this variable from the scale makes the construct more reliable.
### Mental Stability Scale for Female Psychiatric Patients

#### The CORR Procedure

7 Variables: Age Anxiety Depression Sleep Sex Life WeightChange

#### Simple Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>23</td>
<td>37.91304</td>
<td>5.13378</td>
<td>872.00000</td>
</tr>
<tr>
<td>Anxiety</td>
<td>23</td>
<td>2.34783</td>
<td>0.64728</td>
<td>54.00000</td>
</tr>
<tr>
<td>Depression</td>
<td>23</td>
<td>1.95652</td>
<td>0.56232</td>
<td>45.00000</td>
</tr>
<tr>
<td>Sleep</td>
<td>23</td>
<td>1.86957</td>
<td>0.34435</td>
<td>43.00000</td>
</tr>
<tr>
<td>Sex</td>
<td>23</td>
<td>1.95652</td>
<td>0.20851</td>
<td>45.00000</td>
</tr>
<tr>
<td>Life</td>
<td>23</td>
<td>1.56522</td>
<td>0.50687</td>
<td>36.00000</td>
</tr>
<tr>
<td>WeightChange</td>
<td>23</td>
<td>1.78261</td>
<td>3.06381</td>
<td>41.00000</td>
</tr>
</tbody>
</table>

#### Simple Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>30.00000</td>
<td>46.00000</td>
<td>age in years</td>
</tr>
<tr>
<td>Anxiety</td>
<td>1.00000</td>
<td>4.00000</td>
<td>anxiety level</td>
</tr>
<tr>
<td>Depression</td>
<td>1.00000</td>
<td>3.00000</td>
<td>depression level</td>
</tr>
<tr>
<td>Sleep</td>
<td>1.00000</td>
<td>2.00000</td>
<td>normal sleep (1=y 2=n)</td>
</tr>
<tr>
<td>Sex</td>
<td>1.00000</td>
<td>2.00000</td>
<td>sexual (1=n 2=y)</td>
</tr>
<tr>
<td>Life</td>
<td>1.00000</td>
<td>2.00000</td>
<td>suicidal (1=n 2=y)</td>
</tr>
<tr>
<td>WeightChange</td>
<td>-2.60000</td>
<td>8.30000</td>
<td>recent weight change</td>
</tr>
</tbody>
</table>

#### Cronbach Coefficient Alpha

<table>
<thead>
<tr>
<th>Variables</th>
<th>Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw</td>
<td>0.627754</td>
</tr>
<tr>
<td>Standardized</td>
<td>0.845339</td>
</tr>
</tbody>
</table>

#### Cronbach Coefficient Alpha with Deleted Variable

<table>
<thead>
<tr>
<th>Deleted Variable</th>
<th>Correlation with Total Alpha</th>
<th>Correlation with Total Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>0.742614</td>
<td>0.557515</td>
</tr>
</tbody>
</table>

#### Cronbach Coefficient Alpha with Deleted Variable

<table>
<thead>
<tr>
<th>Deleted Variable</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>age in years</td>
</tr>
</tbody>
</table>
Example 4: Storing Partial Correlations in an Output Data Set

### The CORR Procedure

#### Cronbach Coefficient Alpha with Deleted Variable

<table>
<thead>
<tr>
<th>Deleted Variable</th>
<th>Correlation with Total Alpha</th>
<th>Correlation with Total Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anxiety</td>
<td>0.577129</td>
<td>0.590851</td>
</tr>
<tr>
<td>Depression</td>
<td>0.554983</td>
<td>0.770956</td>
</tr>
<tr>
<td>Sleep</td>
<td>0.378930</td>
<td>0.618367</td>
</tr>
<tr>
<td>Sex</td>
<td>0.155115</td>
<td>0.333368</td>
</tr>
<tr>
<td>Life</td>
<td>0.622207</td>
<td>0.625338</td>
</tr>
<tr>
<td>WeightChange</td>
<td>0.843939</td>
<td>0.749261</td>
</tr>
</tbody>
</table>

#### Cronbach Coefficient Alpha with Deleted Variable

<table>
<thead>
<tr>
<th>Deleted Variable</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anxiety</td>
<td>anxiety level</td>
</tr>
<tr>
<td>Depression</td>
<td>depression level</td>
</tr>
<tr>
<td>Sleep</td>
<td>normal sleep (1=y 2=n)</td>
</tr>
<tr>
<td>Sex</td>
<td>sexual (1=n 2=y)</td>
</tr>
<tr>
<td>Life</td>
<td>suicidal (1=n 2=y)</td>
</tr>
<tr>
<td>WeightChange</td>
<td>recent weight change</td>
</tr>
</tbody>
</table>

### Procedure features:

- PROC CORR statement options:
  - COV
  - KENDALL
  - NOSIMPLE
  - OUTP=
  - SPEARMAN
- PARTIAL statement
- VAR statement

### Data set: FITNESS on page 299

This example:
- suppresses descriptive statistics
- calculates three types of partial correlation coefficients
- calculates a partial covariance matrix
- excludes observations with missing values using listwise deletion
- selects the analysis variables
- creates an output data set with Pearson correlation statistics.

See Output 12.4 on page 298 for a listing of the output data set.
Program

options nodatepageno=1 linesize=120 pagesize=60;

SPEARMAN and KENDALL request correlation statistics. COV calculates the covariance matrix and invokes PEARSON. NOSIMPLE suppresses descriptive statistics. OUT= creates the FITCORR data set that contains Pearson correlation statistics.

proc corr data=fitness spearman kendall
   cov nosimple
   outp=fitcorr;

The VAR statement specifies the analysis variables and the order to print them.

   var weight oxygen runtime;

The PARTIAL statement calculates partial correlations using Age as the controlling variable.

   partial age;

The LABEL statement associates a label with each variable for the duration of the PROC step.

   label age = 'Age of subject'
     weight = 'Wt in kg'
     runtime = '1.5 mi in minutes'
     oxygen = 'O2 use';

The TITLE statement specifies a title for the report.

   title1 'Partial Correlations for a Fitness and Exercise Study';
   run;

Output

The report includes a partial covariance matrix and partial correlations for Pearson's rho, Spearman's rho, and Kendall's tau-b. The p-values for Kendall's tau-b are not available. Because observations with missing data are excluded, PROC CORR uses 28 observations to calculate correlation coefficients.
Partial Correlations for a Fitness and Exercise Study

The CORR Procedure

1 Partial Variables: Age
3 Variables: Weight Oxygen Runtime

Partial Covariance Matrix, DF = 26

<table>
<thead>
<tr>
<th></th>
<th>Weight</th>
<th>Oxygen</th>
<th>Runtime</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight</td>
<td>72.43742055</td>
<td>-12.75113194</td>
<td>2.06766763</td>
</tr>
<tr>
<td>Oxygen</td>
<td>-12.75113194</td>
<td>27.01654904</td>
<td>-5.59370556</td>
</tr>
<tr>
<td>Runtime</td>
<td>2.06766763</td>
<td>-5.59370556</td>
<td>1.94512451</td>
</tr>
</tbody>
</table>

Pearson Partial Correlation Coefficients, N = 28
Prob > |r| under H0: Partial Rho=0

<table>
<thead>
<tr>
<th></th>
<th>Weight</th>
<th>Oxygen</th>
<th>Runtime</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight</td>
<td>1.00000</td>
<td>-0.28824</td>
<td>0.17419</td>
</tr>
<tr>
<td>Wt in kg</td>
<td>0.1448</td>
<td>0.3849</td>
<td></td>
</tr>
<tr>
<td>Oxygen</td>
<td>-0.28824</td>
<td>1.00000</td>
<td>-0.77163</td>
</tr>
<tr>
<td>O2 use</td>
<td>0.1448</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>Runtime</td>
<td>0.17419</td>
<td>-0.77163</td>
<td>1.00000</td>
</tr>
<tr>
<td>1.5 mi in minutes</td>
<td>0.3849</td>
<td>&lt;.0001</td>
<td></td>
</tr>
</tbody>
</table>

Spearman Partial Correlation Coefficients, N = 28
Prob > |r| under H0: Partial Rho=0

<table>
<thead>
<tr>
<th></th>
<th>Weight</th>
<th>Oxygen</th>
<th>Runtime</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight</td>
<td>1.00000</td>
<td>-0.16407</td>
<td>0.08708</td>
</tr>
<tr>
<td>Wt in kg</td>
<td>0.4135</td>
<td>0.6658</td>
<td></td>
</tr>
<tr>
<td>Oxygen</td>
<td>-0.16407</td>
<td>1.00000</td>
<td>-0.67112</td>
</tr>
<tr>
<td>O2 use</td>
<td>0.4135</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td>Runtime</td>
<td>0.08708</td>
<td>-0.67112</td>
<td>1.00000</td>
</tr>
<tr>
<td>1.5 mi in minutes</td>
<td>0.6658</td>
<td>0.0001</td>
<td></td>
</tr>
</tbody>
</table>

Kendall Partial Tau b Correlation Coefficients, N = 28

<table>
<thead>
<tr>
<th></th>
<th>Weight</th>
<th>Oxygen</th>
<th>Runtime</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight</td>
<td>1.00000</td>
<td>-0.09021</td>
<td>0.02854</td>
</tr>
<tr>
<td>Wt in kg</td>
<td></td>
<td>0.02854</td>
<td></td>
</tr>
<tr>
<td>Oxygen</td>
<td>-0.09021</td>
<td>1.00000</td>
<td>-0.52158</td>
</tr>
<tr>
<td>O2 use</td>
<td></td>
<td>-0.52158</td>
<td>1.00000</td>
</tr>
<tr>
<td>Runtime</td>
<td>0.02854</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.5 mi in minutes</td>
<td></td>
<td>1.00000</td>
<td></td>
</tr>
</tbody>
</table>


