

# **SAS/STAT<sup>®</sup> 14.2 User's Guide**

## **The ANOVA Procedure**

This document is an individual chapter from *SAS/STAT® 14.2 User's Guide*.

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2016. *SAS/STAT® 14.2 User's Guide*. Cary, NC: SAS Institute Inc.

### **SAS/STAT® 14.2 User's Guide**

Copyright © 2016, SAS Institute Inc., Cary, NC, USA

All Rights Reserved. Produced in the United States of America.

**For a hard-copy book:** No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, or otherwise, without the prior written permission of the publisher, SAS Institute Inc.

**For a web download or e-book:** Your use of this publication shall be governed by the terms established by the vendor at the time you acquire this publication.

The scanning, uploading, and distribution of this book via the Internet or any other means without the permission of the publisher is illegal and punishable by law. Please purchase only authorized electronic editions and do not participate in or encourage electronic piracy of copyrighted materials. Your support of others' rights is appreciated.

**U.S. Government License Rights; Restricted Rights:** The Software and its documentation is commercial computer software developed at private expense and is provided with RESTRICTED RIGHTS to the United States Government. Use, duplication, or disclosure of the Software by the United States Government is subject to the license terms of this Agreement pursuant to, as applicable, FAR 12.212, DFAR 227.7202-1(a), DFAR 227.7202-3(a), and DFAR 227.7202-4, and, to the extent required under U.S. federal law, the minimum restricted rights as set out in FAR 52.227-19 (DEC 2007). If FAR 52.227-19 is applicable, this provision serves as notice under clause (c) thereof and no other notice is required to be affixed to the Software or documentation. The Government's rights in Software and documentation shall be only those set forth in this Agreement.

SAS Institute Inc., SAS Campus Drive, Cary, NC 27513-2414

November 2016

SAS® and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.

SAS software may be provided with certain third-party software, including but not limited to open-source software, which is licensed under its applicable third-party software license agreement. For license information about third-party software distributed with SAS software, refer to <http://support.sas.com/thirdpartylicenses>.

# Chapter 26

## The ANOVA Procedure

### Contents

---

Overview: ANOVA Procedure . . . . .	<b>972</b>
Getting Started: ANOVA Procedure . . . . .	<b>972</b>
One-Way Layout with Means Comparisons . . . . .	972
Randomized Complete Block with One Factor . . . . .	976
Syntax: ANOVA Procedure . . . . .	<b>980</b>
PROC ANOVA Statement . . . . .	980
ABSORB Statement . . . . .	983
BY Statement . . . . .	983
CLASS Statement . . . . .	984
FREQ Statement . . . . .	985
MANOVA Statement . . . . .	985
MEANS Statement . . . . .	989
MODEL Statement . . . . .	994
REPEATED Statement . . . . .	995
TEST Statement . . . . .	999
Details: ANOVA Procedure . . . . .	<b>1000</b>
Specification of Effects . . . . .	1000
Using PROC ANOVA Interactively . . . . .	1002
Missing Values . . . . .	1003
Output Data Set . . . . .	1003
Computational Method . . . . .	1004
Displayed Output . . . . .	1005
ODS Table Names . . . . .	1006
ODS Graphics . . . . .	1008
Examples: ANOVA Procedure . . . . .	<b>1009</b>
Example 26.1: Factorial Treatments in Complete Blocks . . . . .	1009
Example 26.2: Alternative Multiple Comparison Procedures . . . . .	1011
Example 26.3: Split Plot . . . . .	1015
Example 26.4: Latin Square Split Plot . . . . .	1017
Example 26.5: Strip-Split Plot . . . . .	1020
References . . . . .	<b>1024</b>

---

---

## Overview: ANOVA Procedure

The ANOVA procedure performs *analysis of variance* (ANOVA) for balanced data from a wide variety of experimental designs. In analysis of variance, a continuous response variable, known as a *dependent variable*, is measured under experimental conditions identified by classification variables, known as *independent variables*. The variation in the response is assumed to be due to effects in the classification, with random error accounting for the remaining variation.

The ANOVA procedure is one of several procedures available in SAS/STAT software for analysis of variance. The ANOVA procedure is designed to handle balanced data (that is, data with equal numbers of observations for every combination of the classification factors), whereas the GLM procedure can analyze both balanced and unbalanced data. Because PROC ANOVA takes into account the special structure of a balanced design, it is faster and uses less storage than PROC GLM for balanced data.

Use PROC ANOVA for the analysis of balanced data only, with the following exceptions: one-way analysis of variance, Latin square designs, certain partially balanced incomplete block designs, completely nested (hierarchical) designs, and designs with cell frequencies that are proportional to each other and are also proportional to the background population. These exceptions have designs in which the factors are all orthogonal to each other.

For further discussion, see Searle (1971, p. 138). PROC ANOVA works for designs with block diagonal  $\mathbf{X}'\mathbf{X}$  matrices where the elements of each block all have the same value. The procedure partially tests this requirement by checking for equal cell means. However, this test is imperfect: some designs that cannot be analyzed correctly might pass the test, and designs that can be analyzed correctly might not pass. If your design does not pass the test, PROC ANOVA produces a warning message to tell you that the design is unbalanced and that the ANOVA analyses might not be valid; if your design is not one of the special cases described here, then you should use PROC GLM instead. Complete validation of designs is not performed in PROC ANOVA since this would require the whole  $\mathbf{X}'\mathbf{X}$  matrix; if you are unsure about the validity of PROC ANOVA for your design, you should use PROC GLM.

**CAUTION:** If you use PROC ANOVA for analysis of unbalanced data, you must assume responsibility for the validity of the results.

The ANOVA procedure automatically produces graphics as part of its ODS output. For general information about ODS graphics, see the section “ODS Graphics” on page 1008 and Chapter 21, “Statistical Graphics Using ODS.”

---

## Getting Started: ANOVA Procedure

The following examples demonstrate how you can use the ANOVA procedure to perform analyses of variance for a one-way layout and a randomized complete block design.

---

### One-Way Layout with Means Comparisons

A one-way analysis of variance considers one treatment factor with two or more treatment levels. The goal of the analysis is to test for differences among the means of the levels and to quantify these differences. If there are two treatment levels, this analysis is equivalent to a  $t$  test comparing two group means.

The assumptions of analysis of variance (Steel and Torrie 1980) are that treatment effects are additive and experimental errors are independently random with a normal distribution that has mean zero and constant variance.

The following example studies the effect of bacteria on the nitrogen content of red clover plants. The treatment factor is bacteria strain, and it has six levels. Five of the six levels consist of five different *Rhizobium trifolii* bacteria cultures combined with a composite of five *Rhizobium meliloti* strains. The sixth level is a composite of the five *Rhizobium trifolii* strains with the composite of the *Rhizobium meliloti*. Red clover plants are inoculated with the treatments, and nitrogen content is later measured in milligrams. The data are derived from an experiment by Erdman (1946) and are analyzed in Chapters 7 and 8 of Steel and Torrie (1980). The following DATA step creates the SAS data set Clover:

```

title1 'Nitrogen Content of Red Clover Plants';
data Clover;
    input Strain $ Nitrogen @@;
    datalines;
3DOK1  19.4 3DOK1  32.6 3DOK1  27.0 3DOK1  32.1 3DOK1  33.0
3DOK5  17.7 3DOK5  24.8 3DOK5  27.9 3DOK5  25.2 3DOK5  24.3
3DOK4  17.0 3DOK4  19.4 3DOK4   9.1 3DOK4  11.9 3DOK4  15.8
3DOK7  20.7 3DOK7  21.0 3DOK7  20.5 3DOK7  18.8 3DOK7  18.6
3DOK13 14.3 3DOK13 14.4 3DOK13 11.8 3DOK13 11.6 3DOK13 14.2
COMPOS 17.3 COMPOS 19.4 COMPOS 19.1 COMPOS 16.9 COMPOS 20.8
;

```

The variable Strain contains the treatment levels, and the variable Nitrogen contains the response. The following statements produce the analysis.

```

proc anova data = Clover;
    class strain;
    model Nitrogen = Strain;
run;

```

The classification variable is specified in the **CLASS** statement. Note that, unlike the GLM procedure, PROC ANOVA does not allow continuous variables on the right-hand side of the model. Figure 26.1 and Figure 26.2 display the output produced by these statements.

**Figure 26.1** Class Level Information

### Nitrogen Content of Red Clover Plants

#### The ANOVA Procedure

Class Level Information						
Class	Levels	Values				
Strain	6	3DOK1	3DOK13	3DOK4	3DOK5	3DOK7 COMPOS
Number of Observations Read 30						
Number of Observations Used 30						

The “Class Level Information” table shown in Figure 26.1 lists the variables that appear in the **CLASS** statement, their levels, and the number of observations in the data set.

Figure 26.2 displays the ANOVA table, followed by some simple statistics and tests of effects.

**Figure 26.2** ANOVA Table

### Nitrogen Content of Red Clover Plants

#### The ANOVA Procedure

Dependent Variable: Nitrogen

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	847.046667	169.409333	14.37	<.0001
Error	24	282.928000	11.788667		
Corrected Total	29	1129.974667			

R-Square	Coeff Var	Root MSE	Nitrogen Mean
0.749616	17.26515	3.433463	19.88667

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Strain	5	847.0466667	169.4093333	14.37	<.0001

The degrees of freedom (DF) column should be used to check the analysis results. The model degrees of freedom for a one-way analysis of variance are the number of levels minus 1; in this case,  $6 - 1 = 5$ . The Corrected Total degrees of freedom are always the total number of observations minus one; in this case  $30 - 1 = 29$ . The sum of Model and Error degrees of freedom equals the Corrected Total.

The overall  $F$  test is significant ( $F = 14.37$ ,  $p < 0.0001$ ), indicating that the model as a whole accounts for a significant portion of the variability in the dependent variable. The  $F$  test for Strain is significant, indicating that some contrast between the means for the different strains is different from zero. Notice that the Model and Strain  $F$  tests are identical, since Strain is the only term in the model.

The  $F$  test for Strain ( $F = 14.37$ ,  $p < 0.0001$ ) suggests that there are differences among the bacterial strains, but it does not reveal any information about the nature of the differences. Mean comparison methods can be used to gather further information. The interactivity of PROC ANOVA enables you to do this without re-running the entire analysis. After you specify a model with a **MODEL** statement and execute the ANOVA procedure with a **RUN** statement, you can execute a variety of statements (such as **MEANS**, **MANOVA**, **TEST**, and **REPEATED**) without PROC ANOVA recalculating the model sum of squares.

The following additional statements request means of the Strain levels with Tukey's studentized range procedure.

```
means strain / tukey;
run;
```

Results of Tukey's procedure are shown in Figure 26.3.

**Figure 26.3** Tukey's Multiple Comparisons Procedure  
**Nitrogen Content of Red Clover Plants**

**The ANOVA Procedure**

**Tukey's Studentized Range (HSD) Test for Nitrogen**

Alpha	0.05
Error Degrees of Freedom	24
Error Mean Square	11.78867
Critical Value of Studentized Range	4.37265
Minimum Significant Difference	6.7142

Means with the same letter are not significantly different.			
Tukey Grouping	Mean	N	Strain
A	28.820	5	3DOK1
A			
B A	23.980	5	3DOK5
B			
B C	19.920	5	3DOK7
B C			
B C	18.700	5	COMPOS
C			
C	14.640	5	3DOK4
C			
C	13.260	5	3DOK13

Examples of implications of the multiple comparisons results are as follows:

- Strain 3DOK1 fixes significantly more nitrogen than all but 3DOK5.
- While 3DOK5 is not significantly different from 3DOK1, it is also not significantly better than all the rest, though it is better than the bottom two groups.

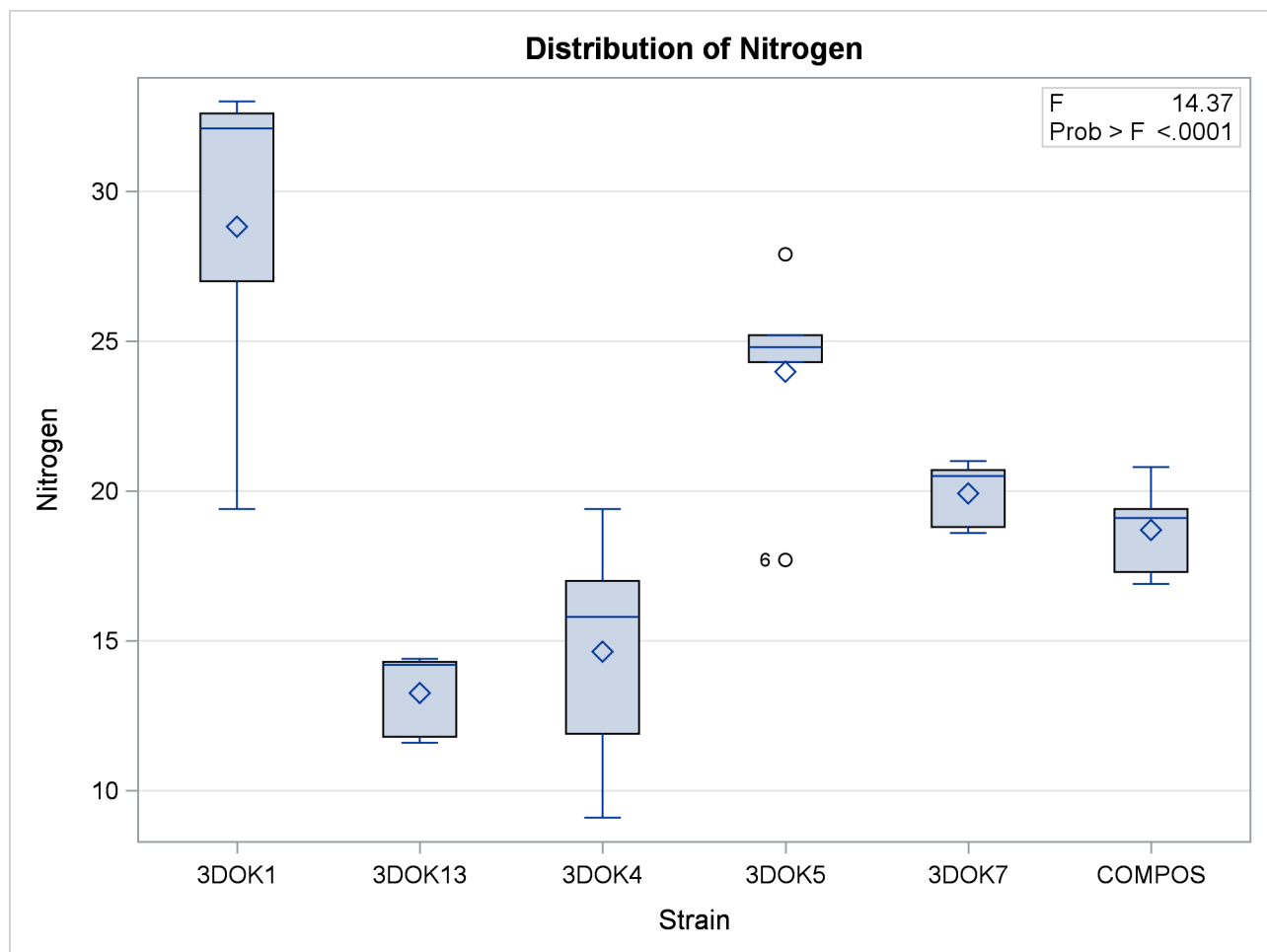
Although the experiment has succeeded in separating the best strains from the worst, more experimentation is required in order to clearly distinguish the very best strain.

If ODS Graphics is enabled, ANOVA also displays by default a plot that enables you to visualize the distribution of nitrogen content for each treatment. The following statements, which are the same as the previous analysis but with ODS graphics enabled, additionally produce [Figure 26.4](#).

```
ods graphics on;
proc anova data = Clover;
  class strain;
  model Nitrogen = Strain;
run;
ods graphics off;
```

When ODS Graphics is enabled and you fit a one-way analysis of variance model, the ANOVA procedure output includes a box plot of the dependent variable values within each classification level of the independent variable. For general information about ODS Graphics, see Chapter 21, “[Statistical Graphics Using ODS](#).” For specific information about the graphics available in the ANOVA procedure, see the section “[ODS Graphics](#)” on page 1008.

**Figure 26.4** Box Plot of Nitrogen Content for each Treatment



## Randomized Complete Block with One Factor

This example illustrates the use of PROC ANOVA in analyzing a randomized complete block design. Researchers are interested in whether three treatments have different effects on the yield and worth of a



particular crop. They believe that the experimental units are not homogeneous. So, a blocking factor is introduced that allows the experimental units to be homogeneous within each block. The three treatments are then randomly assigned within each block.

The data from this study are input into the SAS data set RCB:

```

title1 'Randomized Complete Block';
data RCB;
    input Block Treatment $ Yield Worth @@;
    datalines;
1 A 32.6 112    1 B 36.4 130    1 C 29.5 106
2 A 42.7 139    2 B 47.1 143    2 C 32.9 112
3 A 35.3 124    3 B 40.1 134    3 C 33.6 116
;

```

The variables Yield and Worth are continuous response variables, and the variables Block and Treatment are the classification variables. Because the data for the analysis are balanced, you can use PROC ANOVA to run the analysis.

The statements for the analysis are

```

proc anova data=RCB;
    class Block Treatment;
    model Yield Worth=Block Treatment;
run;

```

The Block and Treatment effects appear in the **CLASS** statement. The **MODEL** statement requests an analysis for each of the two dependent variables, Yield and Worth.

Figure 26.5 shows the “Class Level Information” table.

**Figure 26.5** Class Level Information

### Randomized Complete Block

#### The ANOVA Procedure

Class Level Information		
Class	Levels	Values
Block	3	1 2 3
Treatment	3	A B C

Number of Observations Read	9
Number of Observations Used	9

The “Class Level Information” table lists the number of levels and their values for all effects specified in the **CLASS** statement. The number of observations in the data set are also displayed. Use this information to make sure that the data have been read correctly.

The overall ANOVA table for Yield in Figure 26.6 appears first in the output because it is the first response variable listed on the left side in the **MODEL** statement.

**Figure 26.6** Overall ANOVA Table for Yield

**Randomized Complete Block**

**The ANOVA Procedure**

**Dependent Variable: Yield**

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
<b>Model</b>	4	225.2777778	56.3194444	8.94	0.0283
<b>Error</b>	4	25.1911111	6.2977778		
<b>Corrected Total</b>	8	250.4688889			

R-Square	Coeff Var	Root MSE	Yield Mean
0.899424	6.840047	2.509537	36.68889

The overall  $F$  statistic is significant ( $F = 8.94$ ,  $p = 0.0283$ ), indicating that the model as a whole accounts for a significant portion of the variation in Yield and that you can proceed to evaluate the tests of effects.

The degrees of freedom (DF) are used to ensure correctness of the data and model. The Corrected Total degrees of freedom are one less than the total number of observations in the data set; in this case,  $9 - 1 = 8$ . The Model degrees of freedom for a randomized complete block are  $(b - 1) + (t - 1)$ , where  $b$  = number of block levels and  $t$  = number of treatment levels. In this case, this formula leads to  $(3 - 1) + (3 - 1) = 4$  model degrees of freedom.

Several simple statistics follow the ANOVA table. The R-Square indicates that the model accounts for nearly 90% of the variation in the variable Yield. The coefficient of variation (C.V.) is listed along with the Root MSE and the mean of the dependent variable. The Root MSE is an estimate of the standard deviation of the dependent variable. The C.V. is a unitless measure of variability.

The tests of the effects shown in Figure 26.7 are displayed after the simple statistics.

**Figure 26.7** Tests of Effects for Yield

Source	DF	Anova SS	Mean Square	F Value	Pr > F
<b>Block</b>	2	98.1755556	49.0877778	7.79	0.0417
<b>Treatment</b>	2	127.1022222	63.5511111	10.09	0.0274

For Yield, both the Block and Treatment effects are significant ( $F = 7.79$ ,  $p = 0.0417$  and  $F = 10.09$ ,  $p = 0.0274$ , respectively) at the 95% level. From this you can conclude that blocking is useful for this variable and that some contrast between the treatment means is significantly different from zero.

Figure 26.8 shows the ANOVA table, simple statistics, and tests of effects for the variable Worth.

**Figure 26.8** ANOVA Table for Worth

### Randomized Complete Block

#### The ANOVA Procedure

Dependent Variable: Worth

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	1247.333333	311.833333	8.28	0.0323
Error	4	150.666667	37.666667		
Corrected Total	8	1398.000000			

R-Square	Coeff Var	Root MSE	Worth Mean
0.892227	4.949450	6.137318	124.0000

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Block	2	354.6666667	177.3333333	4.71	0.0889
Treatment	2	892.6666667	446.3333333	11.85	0.0209

The overall  $F$  test is significant ( $F = 8.28$ ,  $p = 0.0323$ ) at the 95% level for the variable Worth. The Block effect is not significant at the 0.05 level but is significant at the 0.10 confidence level ( $F = 4.71$ ,  $p = 0.0889$ ). Generally, the usefulness of blocking should be determined before the analysis. However, since there are two dependent variables of interest, and Block is significant for one of them (Yield), blocking appears to be generally useful. For Worth, as with Yield, the effect of Treatment is significant ( $F = 11.85$ ,  $p = 0.0209$ ).

Issuing the following command produces the Treatment means.

```
means Treatment;
run;
```

Figure 26.9 displays the treatment means and their standard deviations for both dependent variables.

**Figure 26.9** Means of Yield and Worth

### Randomized Complete Block

#### The ANOVA Procedure

Level of Treatment	N	Yield		Worth	
		Mean	Std Dev	Mean	Std Dev
A	3	36.8666667	5.22908532	125.000000	13.5277493
B	3	41.2000000	5.43415127	135.666667	6.6583281
C	3	32.0000000	2.19317122	111.333333	5.0332230

## Syntax: ANOVA Procedure

The following statements are available in the ANOVA procedure:

```
PROC ANOVA < options > ;
  CLASS variables < / option > ;
  MODEL dependents = effects < / options > ;
  ABSORB variables ;
  BY variables ;
  FREQ variable ;
  MANOVA < test-options > < / detail-options > ;
  MEANS effects < / options > ;
  REPEATED factor-specification < / options > ;
  TEST < H=effects > E=effect ;
```

The **PROC ANOVA**, **CLASS**, and **MODEL** statements are required, and they must precede the first **RUN** statement. The **CLASS** statement must precede the **MODEL** statement. If you use the **ABSORB**, **FREQ**, or **BY** statement, it must precede the first **RUN** statement. The **MANOVA**, **MEANS**, **REPEATED**, and **TEST** statements must follow the **MODEL** statement, and they can be specified in any order. These four statements can also appear after the first **RUN** statement.

Table 26.1 summarizes the function of each statement (other than the **PROC** statement) in the ANOVA procedure:

**Table 26.1** Statements in the ANOVA Procedure

Statement	Description
<b>ABSORB</b>	Absorbs classification effects in a model
<b>BY</b>	Specifies variables to define subgroups for the analysis
<b>CLASS</b>	Declares classification variables
<b>FREQ</b>	Specifies a frequency variable
<b>MANOVA</b>	Performs a multivariate analysis of variance
<b>MEANS</b>	Computes and compares means
<b>MODEL</b>	Defines the model to be fit
<b>REPEATED</b>	Performs multivariate and univariate repeated measures analysis of variance
<b>TEST</b>	Constructs tests that use the sums of squares for effects and the error term you specify

## PROC ANOVA Statement

```
PROC ANOVA < options > ;
```

The **PROC ANOVA** statement invokes the ANOVA procedure. Table 26.2 summarizes the *options* available in the **PROC ANOVA** statement.

**Table 26.2** PROC ANOVA Statement Options

Option	Description
<b>Specify input and output data sets</b>	
<b>DATA=</b>	Specifies input SAS data set
<b>MANOVA</b>	Requests the multivariate mode of eliminating observations with missing values
<b>MULTIPASS</b>	Requests that the input data set be reread when necessary, instead of using a utility file
<b>NAMELEN=</b>	Specifies the length of effect names
<b>NOPRINT</b>	Suppresses the normal display of results
<b>ORDER=</b>	Specifies the sort order for the levels of the classification variables
<b>OUTSTAT=</b>	Names an output data set for information and statistics on each model effect
<b>PLOTS</b>	Controls the plots produced through ODS Graphics.

You can specify the following *options* in the PROC ANOVA statement:

**DATA=SAS-data-set**

names the SAS data set used by the ANOVA procedure. By default, PROC ANOVA uses the most recently created SAS data set.

**MANOVA**

requests the multivariate mode of eliminating observations with missing values. If any of the dependent variables have missing values, the procedure eliminates that observation from the analysis. The MANOVA option is useful if you use PROC ANOVA in interactive mode and plan to perform a multivariate analysis.

**MULTIPASS**

requests that PROC ANOVA reread the input data set, when necessary, instead of writing the values of dependent variables to a utility file. This option decreases disk space usage at the expense of increased execution times and is useful only in rare situations where disk space is at an absolute premium.

**NAMELEN=*n***

specifies the length of effect names to be *n* characters long, where *n* is a value between 20 and 200 characters. The default length is 20 characters.

**NOPRINT**

suppresses the normal display of results. The NOPRINT option is useful when you want to create only the output data set with the procedure. Note that this option temporarily disables the Output Delivery System (ODS); see Chapter 20, “[Using the Output Delivery System](#),” for more information.

**ORDER=DATA | FORMATTED | FREQ | INTERNAL**

specifies the sort order for the levels of the classification variables (which are specified in the [CLASS](#) statement).

This option applies to the levels for all classification variables, except when you use the (default) ORDER=FORMATTED option with numeric classification variables that have no explicit format. In that case, the levels of such variables are ordered by their internal value.

The ORDER= option can take the following values:

Value of ORDER=	Levels Sorted By
<b>DATA</b>	Order of appearance in the input data set
<b>FORMATTED</b>	External formatted value, except for numeric variables with no explicit format, which are sorted by their unformatted (internal) value
<b>FREQ</b>	Descending frequency count; levels with the most observations come first in the order
<b>INTERNAL</b>	Unformatted value

By default, ORDER=FORMATTED. For ORDER=FORMATTED and ORDER=INTERNAL, the sort order is machine-dependent.

#### **OUTSTAT=SAS-data-set**

names an output data set that contains sums of squares, degrees of freedom,  $F$  statistics, and probability levels for each effect in the model. If you use the CANONICAL option in the MANOVA statement and do not use an M= specification in the MANOVA statement, the data set also contains results of the canonical analysis. See the section “Output Data Set” on page 1003 for more information.

#### **PLOTS <(MAXPOINTS=NONE | number)> <=NONE>**

##### **PLOTS=NONE**

controls the plots produced through ODS Graphics. When ODS Graphics is enabled, the ANOVA procedure can display a grouped box plot of the input data with groups defined by an effect in the model. Such a plot is produced by default if you have a one-way model, with only a single classification variable, or if you use a MEANS statement. Specify the PLOTS=NONE option to prevent these plots from being produced when ODS Graphics is enabled.

ODS Graphics must be enabled before plots can be requested. For example:

```
ods graphics on;
proc anova data = Clover;
  class strain;
  model Nitrogen = Strain;
run;
ods graphics off;
```

For more information about enabling and disabling ODS Graphics, see the section “Enabling and Disabling ODS Graphics” on page 607 in Chapter 21, “Statistical Graphics Using ODS.”

The following option can be specified in parentheses after PLOTS.

##### **MAXPOINTS=NONE | number**

specifies that plots with elements that require processing of more than *number* points be suppressed. The default is MAXPOINTS=5000. This limit is ignored if you specify MAXPOINTS=NONE.

---

## ABSORB Statement

**ABSORB** *variables* ;

Absorption is a computational technique that provides a large reduction in time and memory requirements for certain types of models. The *variables* are one or more variables in the input data set.

For a main effect variable that does not participate in interactions, you can absorb the effect by naming it in an ABSORB statement. This means that the effect can be adjusted out before the construction and solution of the rest of the model. This is particularly useful when the effect has a large number of levels.

Several variables can be specified, in which case each one is assumed to be nested in the preceding variable in the ABSORB statement.

**NOTE:** When you use the ABSORB statement, the data set (or each BY group, if a BY statement appears) must be sorted by the variables in the ABSORB statement. Including an absorbed variable in the CLASS list or in the MODEL statement might produce erroneous sums of squares. If the ABSORB statement is used, it must appear before the first RUN statement or it is ignored.

When you use an ABSORB statement and also use the INT option in the MODEL statement, the procedure ignores the option but produces the uncorrected total sum of squares (SS) instead of the corrected total SS.

See the section “Absorption” on page 3687 in Chapter 47, “The GLM Procedure,” for more information.

---

## BY Statement

**BY** *variables* ;

You can specify a BY statement with PROC ANOVA to obtain separate analyses of observations in groups that are defined by the BY variables. When a BY statement appears, the procedure expects the input data set to be sorted in order of the BY variables. If you specify more than one BY statement, only the last one specified is used.

If your input data set is not sorted in ascending order, use one of the following alternatives:

- Sort the data by using the SORT procedure with a similar BY statement.
- Specify the NOTSORTED or DESCENDING option in the BY statement for the ANOVA procedure. The NOTSORTED option does not mean that the data are unsorted but rather that the data are arranged in groups (according to values of the BY variables) and that these groups are not necessarily in alphabetical or increasing numeric order.
- Create an index on the BY variables by using the DATASETS procedure (in Base SAS software).

Since sorting the data changes the order in which PROC ANOVA reads observations, the sort order for the levels of the classification variables might be affected if you have also specified the ORDER=DATA option in the PROC ANOVA statement.

If the BY statement is used, it must appear before the first RUN statement, or it is ignored. When you use a BY statement, the interactive features of PROC ANOVA are disabled.

When both a **BY** and an **ABSORB** statement are used, observations must be sorted first by the variables in the **BY** statement, and then by the variables in the **ABSORB** statement.

For more information about BY-group processing, see the discussion in *SAS Language Reference: Concepts*. For more information about the DATASETS procedure, see the discussion in the *Base SAS Procedures Guide*.

---

## CLASS Statement

**CLASS** *variable* <(REF= *option*)> ... <*variable* <(REF= *option*)>> </ *global-options*> ;

The CLASS statement names the classification variables to be used in the model. Typical classification variables are Treatment, Sex, Race, Group, and Replication. If you use the CLASS statement, it must appear before the **MODEL** statement.

Classification variables can be either character or numeric. By default, class levels are determined from the entire set of formatted values of the CLASS variables.

**NOTE:** Prior to SAS 9, class levels were determined by using no more than the first 16 characters of the formatted values. To revert to this previous behavior, you can use the TRUNCATE option in the CLASS statement.

In any case, you can use formats to group values into levels. See the discussion of the FORMAT procedure in the *Base SAS Procedures Guide* and the discussions of the FORMAT statement and SAS formats in *SAS Formats and Informats: Reference*. You can adjust the order of CLASS variable levels with the **ORDER=** option in the **PROC ANOVA** statement.

You can specify the following REF= option to indicate how the levels of an individual classification variable are to be ordered by enclosing it in parentheses after the variable name:

### REF='level' | FIRST | LAST

specifies a level of the classification variable to be put at the end of the list of levels. This level thus corresponds to the reference level in the usual interpretation of the estimates with PROC ANOVA's singular parameterization. You can specify the *level* of the variable to use as the reference level; specify a value that corresponds to the formatted value of the variable if a format is assigned. Alternatively, you can specify REF=FIRST to designate that the first ordered level serve as the reference, or REF=LAST to designate that the last ordered level serve as the reference. To specify that REF=FIRST or REF=LAST be used for all classification variables, use the **REF= global-option** after the slash (/) in the CLASS statement.

You can specify the following *global-options* in the CLASS statement after a slash (/):

### REF=FIRST | LAST

specifies a level of all classification variables to be put at the end of the list of levels. This level thus corresponds to the reference level in the usual interpretation of the estimates with PROC ANOVA's singular parameterization. Specify REF=FIRST to designate that the first ordered level for each classification variable serve as the reference. Specify REF=LAST to designate that the last ordered level serve as the reference. This option applies to all the variables specified in the CLASS statement. To specify different reference levels for different classification variables, use **REF=** options for individual variables.



**TRUNCATE**

specifies that class levels be determined by using only up to the first 16 characters of the formatted values of CLASS variables. When formatted values are longer than 16 characters, you can use this option to revert to the levels as determined in releases prior to SAS 9.

---

**FREQ Statement**

**FREQ** *variable* ;

The FREQ statement names a variable that provides frequencies for each observation in the **DATA=** data set. Specifically, if  $n$  is the value of the FREQ variable for a given observation, then that observation is used  $n$  times.

The analysis produced by using a FREQ statement reflects the expanded number of observations. For example, means and total degrees of freedom reflect the expanded number of observations. You can produce the same analysis (without the FREQ statement) by first creating a new data set that contains the expanded number of observations. For example, if the value of the FREQ variable is 5 for the first observation, the first 5 observations in the new data set would be identical. Each observation in the old data set would be replicated  $n_i$  times in the new data set, where  $n_i$  is the value of the FREQ variable for that observation.

If the value of the FREQ variable is missing or is less than 1, the observation is not used in the analysis. If the value is not an integer, only the integer portion is used.

If the FREQ statement is used, it must appear before the first RUN statement or it is ignored.

---

**MANOVA Statement**

**MANOVA** *< test-options >* *< detail-options >* ;

If the MODEL statement includes more than one dependent variable, you can perform multivariate analysis of variance with the MANOVA statement. The *test-options* define which effects to test, while the *detail-options* specify how to execute the tests and what results to display.

When a MANOVA statement appears before the first RUN statement, PROC ANOVA enters a multivariate mode with respect to the handling of missing values; in addition to observations with missing independent variables, observations with *any* missing dependent variables are excluded from the analysis. If you want to use this mode of handling missing values but do not need any multivariate analyses, specify the MANOVA option in the **PROC ANOVA** statement.

Table 26.3 summarizes the *options* available in the MANOVA statement.

**Table 26.3** MANOVA Statement Options

Option	Description
<b>Test Options</b>	
H=	Specifies hypothesis effects
E=	Specifies the error effect
M=	Specifies a transformation matrix for the dependent variables

**Table 26.3** *continued*

Option	Description
<b>MNAMES=</b>	Provides names for the transformed variables
<b>PREFIX=</b>	Alternatively identifies the transformed variables
<b>Detail Options</b>	
<b>CANONICAL</b>	Displays a canonical analysis of the <b>H</b> and <b>E</b> matrices
<b>MSTAT=</b>	Specifies the method of evaluating the multivariate test statistics
<b>ORTH</b>	Orthogonalizes the rows of the transformation matrix
<b>PRINTE</b>	Displays the error SSCP matrix <b>E</b>
<b>PRINTH</b>	Displays the hypothesis SSCP matrix <b>H</b>
<b>SUMMARY</b>	Produces analysis-of-variance tables for each dependent variable

### Test Options

You can specify the following *options* in the MANOVA statement as *test-options* in order to define which multivariate tests to perform.

**H=effects | INTERCEPT | \_ALL\_**

specifies effects in the preceding model to use as hypothesis matrices. For each SSCP matrix **H** associated with an effect, the H= specification computes an analysis based on the characteristic roots of  $\mathbf{E}^{-1}\mathbf{H}$ , where **E** is the matrix associated with the error effect. The characteristic roots and vectors are displayed, along with the Hotelling-Lawley trace, Pillai's trace, Wilks' lambda, and Roy's greatest root. By default, these statistics are tested with approximations based on the *F* distribution. To test them with exact (but computationally intensive) calculations, use the **MSTAT=EXACT** option.

Use the keyword INTERCEPT to produce tests for the intercept. To produce tests for all effects listed in the **MODEL** statement, use the keyword **\_ALL\_** in place of a list of effects.

For background and further details, see the section “Multivariate Analysis of Variance” on page 3710 in Chapter 47, “The GLM Procedure.”

**E=effect**

specifies the error effect. If you omit the E= specification, the ANOVA procedure uses the error SSCP (residual) matrix from the analysis.

**M=equation,...,equation | (row-of-matrix,...,row-of-matrix)**

specifies a transformation matrix for the dependent variables listed in the **MODEL** statement. The equations in the M= specification are of the form

$$\begin{aligned}
 c_1 \times \text{dependent} - \text{variable} &\pm c_2 \times \text{dependent} - \text{variable} \\
 &\cdots \pm c_n \times \text{dependent} - \text{variable}
 \end{aligned}$$

where the  $c_i$  values are coefficients for the various *dependent-variables*. If the value of a given  $c_i$  is 1, it can be omitted; in other words  $1 \times Y$  is the same as  $Y$ . Equations should involve two or more dependent variables. For sample syntax, see the section “Examples” on page 988.

Alternatively, you can input the transformation matrix directly by entering the elements of the matrix with commas separating the rows, and parentheses surrounding the matrix. When this alternate form of input is used, the number of elements in each row must equal the number of dependent variables. Although these combinations actually represent the columns of the **M** matrix, they are displayed by rows.

When you include an **M=** specification, the analysis requested in the MANOVA statement is carried out for the variables defined by the equations in the specification, not the original dependent variables. If you omit the **M=** option, the analysis is performed for the original dependent variables in the **MODEL** statement.

If an **M=** specification is included without either the **MNAMES=** or the **PREFIX=** option, the variables are labeled MVAR1, MVAR2, and so forth by default.

For further information, see the section “[Multivariate Analysis of Variance](#)” on page 3710 in Chapter 47, “[The GLM Procedure](#).”

**MNAMES=names**

provides names for the variables defined by the equations in the **M=** specification. Names in the list correspond to the **M=** equations or the rows of the **M** matrix (as it is entered).

**PREFIX=name**

is an alternative means of identifying the transformed variables defined by the **M=** specification. For example, if you specify **PREFIX=DIFF**, the transformed variables are labeled DIFF1, DIFF2, and so forth.

## Detail Options

You can specify the following *options* in the MANOVA statement after a slash as *detail-options*:

**CANONICAL**

produces a canonical analysis of the **H** and **E** matrices (transformed by the **M** matrix, if specified) instead of the default display of characteristic roots and vectors.

**MSTAT=FAPPROX | EXACT**

specifies the method of evaluating the multivariate test statistics. The default is **MSTAT=FAPPROX**, which specifies that the multivariate tests are evaluated by using the usual approximations based on the *F* distribution, as discussed in the “Multivariate Tests” section in Chapter 4, “[Introduction to Regression Procedures](#).” Alternatively, you can specify **MSTAT=EXACT** to compute exact *p*-values for three of the four tests (Wilks’ lambda, the Hotelling-Lawley trace, and Roy’s greatest root) and an improved *F*-approximation for the fourth (Pillai’s trace). While **MSTAT=EXACT** provides better control of the significance probability for the tests, especially for Roy’s Greatest Root, computations for the exact *p*-values can be appreciably more demanding, and are in fact infeasible for large problems (many dependent variables). Thus, although **MSTAT=EXACT** is more accurate for most data, it is not the default method. For more information about the results of **MSTAT=EXACT**, see the section “[Multivariate Analysis of Variance](#)” on page 3710 in Chapter 47, “[The GLM Procedure](#).”

**ORTH**

requests that the transformation matrix in the **M=** specification of the MANOVA statement be orthonormalized by rows before the analysis.

**PRINTE**

displays the error SSCP matrix **E**. If the **E** matrix is the error SSCP (residual) matrix from the analysis, the partial correlations of the dependent variables given the independent variables are also produced.

For example, the statement

```
manova / printe;
```

displays the error SSCP matrix and the partial correlation matrix computed from the error SSCP matrix.

**PRINTH**

displays the hypothesis SSCP matrix **H** associated with each effect specified by the **H=** specification.

**SUMMARY**

produces analysis-of-variance tables for each dependent variable. When no **M** matrix is specified, a table is produced for each original dependent variable from the **MODEL** statement; with an **M** matrix other than the identity, a table is produced for each transformed variable defined by the **M** matrix.

**Examples**

The following statements give several examples of using a MANOVA statement.

```
proc anova;
  class A B;
  model Y1-Y5=A B(A);
  manova h=A e=B(A) / printh printe;
  manova h=B(A) / printe;
  manova h=A e=B(A) m=Y1-Y2, Y2-Y3, Y3-Y4, Y4-Y5
    prefix=diff;

  manova h=A e=B(A) m=(1 -1 0 0 0,
                      0 1 -1 0 0,
                      0 0 1 -1 0,
                      0 0 0 1 -1) prefix=diff;

run;
```

The first MANOVA statement specifies **A** as the hypothesis effect and **B(A)** as the error effect. As a result of the **PRINTH** option, the procedure displays the hypothesis SSCP matrix associated with the **A** effect; and, as a result of the **PRINTE** option, the procedure displays the error SSCP matrix associated with the **B(A)** effect.

The second MANOVA statement specifies **B(A)** as the hypothesis effect. Since no error effect is specified, PROC ANOVA uses the error SSCP matrix from the analysis as the **E** matrix. The **PRINTE** option displays this **E** matrix. Since the **E** matrix is the error SSCP matrix from the analysis, the partial correlation matrix computed from this matrix is also produced.

The third MANOVA statement requests the same analysis as the first MANOVA statement, but the analysis is carried out for variables transformed to be successive differences between the original dependent variables. The **PREFIX=DIFF** specification labels the transformed variables as **DIFF1**, **DIFF2**, **DIFF3**, and **DIFF4**.

Finally, the fourth MANOVA statement has the identical effect as the third, but it uses an alternative form of the **M=** specification. Instead of specifying a set of equations, the fourth MANOVA statement specifies rows of a matrix of coefficients for the five dependent variables.

As a second example of the use of the M= specification, consider the following:

```
proc anova;
  class group;
  model dose1-dose4=group / nouni;
  manova h = group
    m = -3*dose1 -   dose2 +   dose3 + 3*dose4,
         dose1 -   dose2 -   dose3 +   dose4,
        -dose1 + 3*dose2 - 3*dose3 +   dose4
    mnames = Linear Quadratic Cubic
    / printe;
run;
```

The M= specification gives a transformation of the dependent variables dose1 through dose4 into orthogonal polynomial components, and the MNames= option labels the transformed variables as LINEAR, QUADRATIC, and CUBIC, respectively. Since the PRINTE option is specified and the default residual matrix is used as an error term, the partial correlation matrix of the orthogonal polynomial components is also produced.

For further information, see the section “[Multivariate Analysis of Variance](#)” on page 3710 in Chapter 47, “[The GLM Procedure](#).”

---

## MEANS Statement

**MEANS** *effects* </ *options* > ;

PROC ANOVA can compute means of the dependent variables for any effect that appears on the right-hand side in the **MODEL** statement.

You can use any number of MEANS statements, provided that they appear after the **MODEL** statement. For example, suppose A and B each have two levels. Then, if you use the following statements

```
proc anova;
  class A B;
  model Y=A B A*B;
  means A B / tukey;
  means A*B;
run;
```

means, standard deviations, and Tukey’s multiple comparison tests are produced for each level of the main effects A and B, and just the means and standard deviations for each of the four combinations of levels for A\*B. Since multiple comparisons options apply only to main effects, the single MEANS statement

```
means A B A*B / tukey;
```

produces the same results.

Options are provided to perform multiple comparison tests for only main effects in the model. PROC ANOVA does not perform multiple comparison tests for interaction terms in the model; for multiple comparisons of interaction terms, see the LSMEANS statement in Chapter 47, “[The GLM Procedure](#).”

Table 26.4 summarizes the *options* available in the MEANS statement.

**Table 26.4** Options Available in the MEANS Statement

Option	Description
<b>Perform multiple comparison tests</b>	
BON	Performs Bonferroni t tests of differences between means for all main effect means
DUNCAN	Performs Duncan's multiple range test on all main effect means
DUNNETT	Performs Dunnett's two-tailed t test
DUNNETTL	Performs Dunnett's one-tailed t test, testing if any treatment is significantly less than the control
DUNNETTU	Performs Dunnett's one-tailed t test, testing if any treatment is significantly greater than the control
GABRIEL	Performs Gabriel's multiple-comparison procedure on all main effect means
REGWQ	Performs the Ryan-Einot-Gabriel-Welsch multiple range test
SCHEFFE	Performs Scheffé's multiple-comparison procedure
SIDAK	Performs pairwise t tests on differences between means with levels adjusted according to Sidak's inequality
SMM or GT2	Performs pairwise comparisons based on the studentized maximum modulus and Sidak's uncorrelated-t inequality
SNK	Performs the Student-Newman-Keuls multiple range test
T or LSD	Performs pairwise t tests
TUKEY	Performs Tukey's studentized range test (HSD)
WALLER	Performs the Waller-Duncan k-ratio t test
<b>Specify additional details for multiple comparison tests</b>	
ALPHA=	Specifies the level of significance for comparisons among the means.
CLDIFF	Presents results from options as confidence intervals
CLM	Options as intervals for the mean of each level of the variables specified
E=	Specifies the error mean square used in the multiple comparisons
KRATIO=	Specifies the Type 1/Type 2 error seriousness ratio for the Waller-Duncan test
LINES	Presents results of options by listing the means in descending order and indicating nonsignificant subsets by line segments
NOSORT	Prevents the means from being sorted into descending order
<b>Test for homogeneity of variances</b>	
HOVTEST	Requests a homogeneity of variance test
<b>Compensate for heterogeneous variances</b>	
WELCH	Requests the Welch (1951) variance-weighted one-way ANOVA

Descriptions of these *options* follow. For a further discussion of these *options*, see the section “Multiple Comparisons” on page 3693 in Chapter 47, “The GLM Procedure.”

**ALPHA=*p***

specifies the level of significance for comparisons among the means. By default, ALPHA=0.05. You can specify any value greater than 0 and less than 1.

**BON**

performs Bonferroni  $t$  tests of differences between means for all main effect means in the MEANS statement. See the **CLDIFF** and **LINES** options, which follow, for a discussion of how the procedure displays results.

**CLDIFF**

presents results of the **BON**, **GABRIEL**, **SCHEFFE**, **SIDAK**, **SMM**, **GT2**, **T**, **LSD**, and **TUKEY** options as confidence intervals for all pairwise differences between means, and the results of the **DUNNETT**, **DUNNETTU**, and **DUNNETTL** options as confidence intervals for differences with the control. The **CLDIFF** option is the default for unequal cell sizes unless the **DUNCAN**, **REGWQ**, **SNK**, or **WALLER** option is specified.

**CLM**

presents results of the **BON**, **GABRIEL**, **SCHEFFE**, **SIDAK**, **SMM**, **T**, and **LSD** options as intervals for the mean of each level of the variables specified in the MEANS statement. For all options except **GABRIEL**, the intervals are confidence intervals for the true means. For the **GABRIEL** option, they are *comparison intervals* for comparing means pairwise: in this case, if the intervals corresponding to two means overlap, the difference between them is insignificant according to Gabriel's method.

**DUNCAN**

performs Duncan's multiple range test on all main effect means given in the MEANS statement. See the **LINES** option for a discussion of how the procedure displays results.

**DUNNETT** <(formatted-control-values)>

performs Dunnett's two-tailed  $t$  test, testing if any treatments are significantly different from a single control for all main effects means in the MEANS statement.

To specify which level of the effect is the control, enclose the formatted value in quotes in parentheses after the *keyword*. If more than one effect is specified in the MEANS statement, you can use a list of control values within the parentheses. By default, the first level of the effect is used as the control. For example,

```
means a / dunnett('CONTROL');
```

where CONTROL is the formatted control value of A. As another example,

```
means a b c / dunnett('CNTLA' 'CNTLB' 'CNTLC');
```

where CNTLA, CNTLB, and CNTLC are the formatted control values for A, B, and C, respectively.

**DUNNETTL** <(formatted-control-value)>

performs Dunnett's one-tailed  $t$  test, testing if any treatment is significantly less than the control. Control level information is specified as described previously for the **DUNNETT** option.

**DUNNETTU** <(formatted-control-value)>

performs Dunnett's one-tailed  $t$  test, testing if any treatment is significantly greater than the control. Control level information is specified as described previously for the **DUNNETT** option.

**E=effect**

specifies the error mean square used in the multiple comparisons. By default, PROC ANOVA uses the residual Mean Square (MS). The effect specified with the E= option must be a term in the model; otherwise, the procedure uses the residual MS.

**GABRIEL**

performs Gabriel's multiple-comparison procedure on all main effect means in the MEANS statement. See the [CLDIFF](#) and [LINES](#) options for discussions of how the procedure displays results.

**GT2**

see the [SMM](#) option.

**HOVTEST****HOVTEST=BARTLETT****HOVTEST=BF****HOVTEST=LEVENE <(TYPE=ABS | SQUARE)>****HOVTEST=OBRIEN <(W=number )>**

requests a homogeneity of variance test for the groups defined by the MEANS effect. You can optionally specify a particular test; if you do not specify a test, Levene's test (Levene 1960) with TYPE=SQUARE is computed. Note that this option is ignored unless your [MODEL](#) statement specifies a simple one-way model.

The HOVTEST=BARTLETT option specifies Bartlett's test (Bartlett 1937), a modification of the normal-theory likelihood ratio test.

The HOVTEST=BF option specifies Brown and Forsythe's variation of Levene's test (Brown and Forsythe 1974).

The HOVTEST=LEVENE option specifies Levene's test (Levene 1960), which is widely considered to be the standard homogeneity of variance test. You can use the TYPE= option in parentheses to specify whether to use the absolute residuals (TYPE=ABS) or the squared residuals (TYPE=SQUARE) in Levene's test. The default is TYPE=SQUARE.

The HOVTEST=OBRIEN option specifies O'Brien's test (O'Brien 1979), which is basically a modification of HOVTEST=LEVENE(TYPE=SQUARE). You can use the W= option in parentheses to tune the variable to match the suspected kurtosis of the underlying distribution. By default, W=0.5, as suggested by O'Brien (1979, 1981).

See the section "[Homogeneity of Variance in One-Way Models](#)" on page 3706 in Chapter 47, "[The GLM Procedure](#)," for more details on these methods. [Example 47.10](#) in the same chapter illustrates the use of the HOVTEST and WELCH options in the MEANS statement in testing for equal group variances.

**KRATIO=value**

specifies the Type 1/Type 2 error seriousness ratio for the Waller-Duncan test. Reasonable values for KRATIO are 50, 100, and 500, which roughly correspond for the two-level case to [ALPHA](#) levels of 0.1, 0.05, and 0.01. By default, the procedure uses the default value of 100.



**LINES**

presents results of the **BON**, **DUNCAN**, **GABRIEL**, **REGWQ**, **SCHEFFE**, **SIDAK**, **SMM**, **GT2**, **SNK**, **T**, **LSD TUKEY**, and **WALLER** options by listing the means in descending order and indicating nonsignificant subsets by line segments beside the corresponding means. The **LINES** option is appropriate for equal cell sizes, for which it is the default. The **LINES** option is also the default if the **DUNCAN**, **REGWQ**, **SNK**, or **WALLER** option is specified, or if there are only two cells of unequal size. If the cell sizes are unequal, the harmonic mean of the cell sizes is used, which might lead to somewhat liberal tests if the cell sizes are highly disparate. The **LINES** option cannot be used in combination with the **DUNNETT**, **DUNNETTL**, or **DUNNETTU** option. In addition, the procedure has a restriction that no more than 24 overlapping groups of means can exist. If a mean belongs to more than 24 groups, the procedure issues an error message. You can either reduce the number of levels of the variable or use a multiple comparison test that allows the **CLDIFF** option rather than the **LINES** option.

**LSD**

see the **T** option.

**NOSORT**

prevents the means from being sorted into descending order when the **CLDIFF** or **CLM** option is specified.

**REGWQ**

performs the Ryan-Einot-Gabriel-Welsch multiple range test on all main effect means in the **MEANS** statement. See the **LINES** option for a discussion of how the procedure displays results.

**SCHEFFE**

performs Scheffé's multiple-comparison procedure on all main effect means in the **MEANS** statement. See the **CLDIFF** and **LINES** options for discussions of how the procedure displays results.

**SIDAK**

performs pairwise  $t$  tests on differences between means with levels adjusted according to Sidak's inequality for all main effect means in the **MEANS** statement. See the **CLDIFF** and **LINES** options for discussions of how the procedure displays results.

**SMM****GT2**

performs pairwise comparisons based on the studentized maximum modulus and Sidak's uncorrelated- $t$  inequality, yielding Hochberg's GT2 method when sample sizes are unequal, for all main effect means in the **MEANS** statement. See the **CLDIFF** and **LINES** options for discussions of how the procedure displays results.

**SNK**

performs the Student-Newman-Keuls multiple range test on all main effect means in the **MEANS** statement. See the **LINES** option for a discussion of how the procedure displays results.

**T****LSD**

performs pairwise  $t$  tests, equivalent to Fisher's least-significant-difference test in the case of equal cell sizes, for all main effect means in the **MEANS** statement. See the **CLDIFF** and **LINES** options for discussions of how the procedure displays results.

**TUKEY**

performs Tukey's studentized range test (HSD) on all main effect means in the MEANS statement. (When the group sizes are different, this is the Tukey-Kramer test.) See the [CLDIFF](#) and [LINES](#) options for discussions of how the procedure displays results.

**WALLER**

performs the Waller-Duncan  $k$ -ratio  $t$  test on all main effect means in the MEANS statement. See the [KRATIO=](#) option for information about controlling details of the test, and see the [LINES](#) option for a discussion of how the procedure displays results.

**WELCH**

requests Welch's (1951) variance-weighted one-way ANOVA. This alternative to the usual analysis of variance for a one-way model is robust to the assumption of equal within-group variances. This option is ignored unless your [MODEL](#) statement specifies a simple one-way model.

Note that using the WELCH option merely produces one additional table consisting of Welch's ANOVA. It does not affect all of the other tests displayed by the ANOVA procedure, which still require the assumption of equal variance for exact validity.

See the section "[Homogeneity of Variance in One-Way Models](#)" on page 3706 in Chapter 47, "[The GLM Procedure](#)," for more details on Welch's ANOVA. [Example 47.10](#) in the same chapter illustrates the use of the [HOVTEST](#) and WELCH options in the MEANS statement in testing for equal group variances.

---

## MODEL Statement

**MODEL** *dependents = effects </ options>* ;

The MODEL statement names the dependent variables and independent effects. The syntax of effects is described in the section "[Specification of Effects](#)" on page 1000. For any model effect involving classification variables (interactions as well as main effects), the number of levels cannot exceed 32,767. If no independent effects are specified, only an intercept term is fit. This tests the hypothesis that the mean of the dependent variable is zero. All variables in effects that you specify in the MODEL statement must appear in the [CLASS](#) statement because PROC ANOVA does not allow for continuous effects.

You can specify the following *options* in the MODEL statement; they must be separated from the list of independent effects by a slash.

**INTERCEPT****INT**

displays the hypothesis tests associated with the intercept as an effect in the model. By default, the procedure includes the intercept in the model but does not display associated tests of hypotheses. Except for producing the uncorrected total SS instead of the corrected total SS, the INT option is ignored when you use an [ABSORB](#) statement.

**NOUNI**

suppresses the display of univariate statistics. You typically use the NOUNI option with a multivariate or repeated measures analysis of variance when you do not need the standard univariate output. The NOUNI option in a MODEL statement does not affect the univariate output produced by the [REPEATED](#) statement.

## REPEATED Statement

**REPEATED** *factor-specification* < / options > ;

When values of the dependent variables in the **MODEL** statement represent repeated measurements on the same experimental unit, the **REPEATED** statement enables you to test hypotheses about the measurement factors (often called *within-subject factors*), as well as the interactions of within-subject factors with independent variables in the **MODEL** statement (often called *between-subject factors*). The **REPEATED** statement provides multivariate and univariate tests as well as hypothesis tests for a variety of single-degree-of-freedom contrasts. There is no limit to the number of within-subject factors that can be specified. For more details, see the section “[Repeated Measures Analysis of Variance](#)” on page 3712 in Chapter 47, “[The GLM Procedure](#).”

The **REPEATED** statement is typically used for handling repeated measures designs with one repeated response variable. Usually, the variables on the left-hand side of the equation in the **MODEL** statement represent one repeated response variable.

This does not mean that only one factor can be listed in the **REPEATED** statement. For example, one repeated response variable (hemoglobin count) might be measured 12 times (implying variables Y1 to Y12 on the left-hand side of the equal sign in the **MODEL** statement), with the associated within-subject factors treatment and time (implying two factors listed in the **REPEATED** statement). See the section “[Examples](#)” on page 998 for an example of how PROC ANOVA handles this case.

Designs with two or more repeated response variables can, however, be handled with the **IDENTITY** transformation; see [Example 47.9](#) in Chapter 47, “[The GLM Procedure](#),” for an example of analyzing a doubly-multivariate repeated measures design.

When a **REPEATED** statement appears, the ANOVA procedure enters a multivariate mode of handling missing values. If any values for variables corresponding to each combination of the within-subject factors are missing, the observation is excluded from the analysis.

The simplest form of the **REPEATED** statement requires only a *factor-name*. With two repeated factors, you must specify the *factor-name* and number of levels (*levels*) for each factor. Optionally, you can specify the actual values for the levels (*level-values*), a *transformation* that defines single-degree-of-freedom contrasts, and *options* for additional analyses and output. When more than one within-subject factor is specified, *factor-names* (and associated level and transformation information) must be separated by a comma in the **REPEATED** statement. These terms are described in the following section, “Syntax Details.”

## Syntax Details

Table 26.5 summarizes the *options* available in the **REPEATED** statement.

**Table 26.5** PROC REPEATED Statement Options

Option	Description
<b>CANONICAL</b>	Performs a canonical analysis of the <b>H</b> and <b>E</b> matrices
<b>MSTAT=FAPPROX</b>	Specifies the method of evaluating the multivariate test statistics
<b>NOM</b>	Displays only the results of the univariate analyses
<b>NOU</b>	Displays only the results of the multivariate analyses
<b>PRINTE</b>	Displays the <b>E</b> matrix
<b>PRINTH</b>	Displays the <b>H</b> (SSCP) matrix

Table 26.5 *continued*

Option	Description
PRINTM	Displays the transformation matrices that define the contrasts
PRINTRV	Produces the characteristic roots and vectors
SUMMARY	Produces analysis-of-variance tables for each contrast
UEPSDEF=	Specifies the univariate $F$ test adjustment

You can specify the following terms in the REPEATED statement.

*factor-specification*

The *factor-specification* for the REPEATED statement can include any number of individual factor specifications, separated by commas, of the following form:

*factor-name levels* <(*level-values*)> <*transformation*>

where

*factor-name* names a factor to be associated with the dependent variables. The name should not be the same as any variable name that already exists in the data set being analyzed and should conform to the usual conventions of SAS variable names.

When specifying more than one factor, list the dependent variables in the **MODEL** statement so that the within-subject factors defined in the REPEATED statement are nested; that is, the first factor defined in the REPEATED statement should be the one with values that change least frequently.

*levels* specifies the number of levels associated with the factor being defined. When there is only one within-subject factor, the number of levels is equal to the number of dependent variables. In this case, *levels* is optional. When more than one within-subject factor is defined, however, *levels* is required, and the product of the number of levels of all the factors must equal the number of dependent variables in the **MODEL** statement.

(*level-values*) specifies values that correspond to levels of a repeated-measures factor. These values are used to label output; they are also used as spacings for constructing orthogonal polynomial contrasts if you specify a **POLYNOMIAL** transformation. The number of level values specified must correspond to the number of levels for that factor in the REPEATED statement. Enclose the *level-values* in parentheses.

The following *transformation* keywords define single-degree-of-freedom contrasts for factors specified in the REPEATED statement. Since the number of contrasts generated is always one less than the number of levels of the factor, you have some control over which contrast is omitted from the analysis by which transformation you select. The only exception is the **IDENTITY** transformation; this transformation is not composed of contrasts, and it has the same degrees of freedom as the factor has levels. By default, the procedure uses the **CONTRAST** transformation.

**CONTRAST**<(ordinal-reference-level)>

generates contrasts between levels of the factor and a reference level. By default, the procedure uses the last level; you can optionally specify a reference level in parentheses after the keyword **CONTRAST**. The reference level corresponds to the ordinal value of the level rather than the level value specified. For example, to generate contrasts between the first level of a factor and the other levels, use

```
contrast (1)
```

**HELMERT**

generates contrasts between each level of the factor and the mean of subsequent levels.

**IDENTITY**

generates an identity transformation corresponding to the associated factor. This transformation is *not* composed of contrasts; it has  $n$  degrees of freedom for an  $n$ -level factor, instead of  $n - 1$ . This can be used for doubly-multivariate repeated measures.

**MEAN**<(ordinal-reference-level)>

generates contrasts between levels of the factor and the mean of all other levels of the factor. Specifying a reference level eliminates the contrast between that level and the mean. Without a reference level, the contrast involving the last level is omitted. See the **CONTRAST** transformation for an example.

**POLYNOMIAL**

generates orthogonal polynomial contrasts. Level values, if provided, are used as spacings in the construction of the polynomials; otherwise, equal spacing is assumed.

**PROFILE**

generates contrasts between adjacent levels of the factor.

For examples of the transformation matrices generated by these contrast transformations, see the section “[Repeated Measures Analysis of Variance](#)” on page 3712 in Chapter 47, “[The GLM Procedure](#).”

You can specify the following *options* in the **REPEATED** statement after a slash:

**CANONICAL**

performs a canonical analysis of the **H** and **E** matrices corresponding to the transformed variables specified in the **REPEATED** statement.

**MSTAT=FAPPROX | EXACT**

specifies the method of evaluating the multivariate test statistics. The default is **MSTAT=FAPPROX**, which specifies that the multivariate tests are evaluated by using the usual approximations based on the  $F$  distribution, as discussed in the “Multivariate Tests” section in Chapter 4, “[Introduction to Regression Procedures](#).” Alternatively, you can specify **MSTAT=EXACT** to compute exact  $p$ -values for three of the four tests (Wilks’ lambda, the Hotelling-Lawley trace, and Roy’s greatest root) and an improved  $F$ -approximation for the fourth (Pillai’s trace). While **MSTAT=EXACT** provides better control of the significance probability for the tests, especially for Roy’s Greatest Root, computations for the exact  $p$ -values can be appreciably more demanding, and are in fact infeasible for large problems

(many dependent variables). Thus, although `MSTAT=EXACT` is more accurate for most data, it is not the default method. For more information about the results of `MSTAT=EXACT`, see the section “[Multivariate Analysis of Variance](#)” on page 3710 in Chapter 47, “[The GLM Procedure](#).”

**NOM**

displays only the results of the univariate analyses.

**NOU**

displays only the results of the multivariate analyses.

**PRINTE**

displays the **E** matrix for each combination of within-subject factors, as well as partial correlation matrices for both the original dependent variables and the variables defined by the transformations specified in the `REPEATED` statement. In addition, the `PRINTE` option provides sphericity tests for each set of transformed variables. If the requested transformations are not orthogonal, the `PRINTE` option also provides a sphericity test for a set of orthogonal contrasts.

**PRINTH**

displays the **H** (SSCP) matrix associated with each multivariate test.

**PRINTM**

displays the transformation matrices that define the contrasts in the analysis. `PROC ANOVA` always displays the **M** matrix so that the transformed variables are defined by the rows, not the columns, of the displayed **M** matrix. In other words, `PROC ANOVA` actually displays **M'**.

**PRINTRV**

produces the characteristic roots and vectors for each multivariate test.

**SUMMARY**

produces analysis-of-variance tables for each contrast defined by the within-subjects factors. Along with tests for the effects of the independent variables specified in the `MODEL` statement, a term labeled **MEAN** tests the hypothesis that the overall mean of the contrast is zero.

**UEPSDEF=unbiased-epsilon-definition**

specifies the type of adjustment for the univariate *F* test that is displayed in addition to the Greenhouse-Geisser adjustment. The default is `UEPSDEF=HFL`, corresponding to the corrected form of the Huynh-Feldt adjustment (Huynh and Feldt 1976; Lecoutre 1991). Other alternatives are `UEPSDEF=HF`, the uncorrected Huynh-Feldt adjustment (the only available method in previous releases of SAS/STAT software), and `UEPSDEF=CM`, the adjustment of Chi et al. (2012). See the section “[Hypothesis Testing in Repeated Measures Analysis](#)” on page 3714 in Chapter 47, “[The GLM Procedure](#),” for details about these adjustments.

**Examples**

When specifying more than one factor, list the dependent variables in the `MODEL` statement so that the within-subject factors defined in the `REPEATED` statement are nested; that is, the first factor defined in the `REPEATED` statement should be the one with values that change least frequently. For example, assume that three treatments are administered at each of four times, for a total of twelve dependent variables on each experimental unit. If the variables are listed in the `MODEL` statement as **Y1** through **Y12**, then the following `REPEATED` statement

```
repeated trt 3, time 4;
```

implies the following structure:

	Dependent Variables											
	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10	Y11	Y12
Value of trt	1	1	1	1	2	2	2	2	3	3	3	3
Value of time	1	2	3	4	1	2	3	4	1	2	3	4

The REPEATED statement always produces a table like the preceding one.

For more information about repeated measures analysis and about using the REPEATED statement, see the section “[Repeated Measures Analysis of Variance](#)” on page 3712 in Chapter 47, “[The GLM Procedure](#).”

---

## TEST Statement

```
TEST <H= effects> E= effect ;
```

Although an  $F$  value is computed for all SS in the analysis by using the residual MS as an error term, you can request additional  $F$  tests that use other effects as error terms. You need a TEST statement when a nonstandard error structure (as in a split plot) exists.

**CAUTION:** The ANOVA procedure does not check any of the assumptions underlying the  $F$  statistic. When you specify a TEST statement, you assume sole responsibility for the validity of the  $F$  statistic produced. To help validate a test, you might want to use the GLM procedure with the RANDOM statement and inspect the expected mean squares. In the GLM procedure, you can also use the TEST option in the RANDOM statement.

You can use as many TEST statements as you want, provided that they appear after the [MODEL](#) statement.

You can specify the following terms in the TEST statement.

**H=effects**

specifies which effects in the preceding model are to be used as hypothesis (numerator) effects.

**E=effect**

specifies one, and only one, effect to use as the error (denominator) term. The **E=** specification is required.

The following example uses two TEST statements and is appropriate for analyzing a split-plot design.

```
proc anova;
  class a b c;
  model y=a|b(a)|c;
  test h=a e=b(a);
  test h=c a*c e=b*c(a);
run;
```



## Details: ANOVA Procedure

### Specification of Effects

In SAS analysis-of-variance procedures, the variables that identify levels of the classifications are called *classification variables*, and they are declared in the **CLASS** statement. Classification variables are also called *categorical*, *qualitative*, *discrete*, or *nominal variables*. The values of a classification variable are called *levels*. Classification variables can be either numeric or character. This is in contrast to the *response* (or *dependent*) *variables*, which are continuous. Response variables must be numeric.

The analysis-of-variance model specifies *effects*, which are combinations of classification variables used to explain the variability of the dependent variables in the following manner:

- Main effects are specified by writing the variables by themselves in the **CLASS** statement: A B C. Main effects used as independent variables test the hypothesis that the mean of the dependent variable is the same for each level of the factor in question, ignoring the other independent variables in the model.
- Crossed effects (interactions) are specified by joining the **CLASS** variables with asterisks in the **MODEL** statement: A\*B A\*C A\*B\*C. Interaction terms in a model test the hypothesis that the effect of a factor does not depend on the levels of the other factors in the interaction.
- Nested effects are specified by following a main effect or crossed effect with a **CLASS** variable or list of **CLASS** variables enclosed in parentheses in the **MODEL** statement. The main effect or crossed effect is nested within the effects listed in parentheses: B(A) C\*D(A B). Nested effects test hypotheses similar to interactions, but the levels of the nested variables are not the same for every combination within which they are nested.

The general form of an effect can be illustrated by using the **CLASS** variables A, B, C, D, E, and F:

A \* B \* C(D E F)

The crossed list should come first, followed by the nested list in parentheses. Note that no asterisks appear within the nested list or immediately before the left parenthesis.

### Main Effects Models

For a three-factor main effects model with A, B, and C as the factors and Y as the dependent variable, the necessary statements are

```
proc anova;
  class A B C;
  model Y=A B C;
run;
```



## Models with Crossed Factors

To specify interactions in a factorial model, join effects with asterisks as described previously. For example, these statements specify a complete factorial model, which includes all the interactions:

```
proc anova;
  class A B C;
  model Y=A B C A*B A*C B*C A*B*C;
run;
```

## Bar Notation

You can shorten the specifications of a full factorial model by using bar notation. For example, the preceding statements can also be written

```
proc anova;
  class A B C;
  model Y=A|B|C;
run;
```

When the bar (|) is used, the expression on the right side of the equal sign is expanded from left to right by using the equivalents of rules 2–4 given in Searle (1971, p. 390). The variables on the right- and left-hand sides of the bar become effects, and the cross of them becomes an effect. Multiple bars are permitted. For instance,  $A|B|C$  is evaluated as follows:

$$\begin{aligned} A|B|C &\rightarrow \{A|B\}|C \\ &\rightarrow \{A\ B\ A*B\}|C \\ &\rightarrow A\ B\ A*B\ C\ A*C\ B*C\ A*B*C \end{aligned}$$

You can also specify the maximum number of variables involved in any effect that results from bar evaluation by specifying that maximum number, preceded by an @ sign, at the end of the bar effect. For example, the specification  $A|B|C@2$  results in only those effects that contain two or fewer variables; in this case,  $A\ B\ A*B\ C\ A*C$  and  $B*C$ .

The following table gives more examples of using the bar and at operators.

$A C(B)$	is equivalent to	$A\ C(B)\ A*C(B)$
$A(B) C(B)$	is equivalent to	$A(B)\ C(B)\ A*C(B)$
$A(B) B(D\ E)$	is equivalent to	$A(B)\ B(D\ E)$
$A B(A) C$	is equivalent to	$A\ B(A)\ C\ A*C\ B*C(A)$
$A B(A) C@2$	is equivalent to	$A\ B(A)\ C\ A*C$
$A B C D@2$	is equivalent to	$A\ B\ A*B\ C\ A*C\ B*C\ D\ A*D\ B*D\ C*D$

Consult the section “[Specification of Effects](#)” on page 3670 in Chapter 47, “[The GLM Procedure](#),” for further details on bar notation.

## Nested Models

Write the effect that is nested within another effect first, followed by the other effect in parentheses. For example, if A and B are main effects and C is nested within A and B (that is, the levels of C that are observed are not the same for each combination of A and B), the statements for PROC ANOVA are

```
proc anova;
  class A B C;
  model y=A B C(A B);
run;
```

The identity of a level is viewed within the context of the level of the containing effects. For example, if City is nested within State, then the identity of City is viewed within the context of State.

The distinguishing feature of a nested specification is that nested effects never appear as main effects. Another way of viewing nested effects is that they are effects that pool the main effect with the interaction of the nesting variable.

See the “Automatic Pooling” section, which follows.

## Models Involving Nested, Crossed, and Main Effects

Asterisks and parentheses can be combined in the **MODEL** statement for models involving nested and crossed effects:

```
proc anova;
  class A B C;
  model Y=A B(A) C(A) B*C(A);
run;
```

## Automatic Pooling

In line with the general philosophy of the GLM procedure, there is no difference between the statements

```
model Y=A B(A);
```

and

```
model Y=A A*B;
```

The effect B becomes a nested effect by virtue of the fact that it does not occur as a main effect. If B is not written as a main effect in addition to participating in A\*B, then the sum of squares that is associated with B is pooled into A\*B.

This feature allows the automatic pooling of sums of squares. If an effect is omitted from the model, it is automatically pooled with all the higher-level effects containing the **CLASS** variables in the omitted effect (or within-error). This feature is most useful in split-plot designs.

---

## Using PROC ANOVA Interactively

PROC ANOVA can be used interactively. After you specify a model in a **MODEL** statement and run PROC ANOVA with a **RUN** statement, a variety of statements (such as **MEANS**, **MANOVA**, **TEST**, and **REPEATED**) can be executed without PROC ANOVA recalculating the model sum of squares.

The section “[Syntax: ANOVA Procedure](#)” on page 980 describes which statements can be used interactively. You can execute these interactive statements individually or in groups by following the single statement or group of statements with a RUN statement. Note that the [MODEL](#) statement cannot be repeated; the ANOVA procedure allows only one [MODEL](#) statement.

If you use PROC ANOVA interactively, you can end the procedure with a DATA step, another PROC step, an ENDSAS statement, or a QUIT statement. The syntax of the QUIT statement is

```
quit;
```

When you use PROC ANOVA interactively, additional RUN statements do not end the procedure but tell PROC ANOVA to execute additional statements.

When a WHERE statement is used with PROC ANOVA, it should appear before the first RUN statement. The WHERE statement enables you to select only certain observations for analysis without using a subsetting DATA step. For example, the statement **where group ne 5** omits observations with GROUP=5 from the analysis. See *SAS Statements: Reference* for details about this statement.

When a BY statement is used with PROC ANOVA, interactive processing is not possible; that is, once the first RUN statement is encountered, processing proceeds for each BY group in the data set, and no further statements are accepted by the procedure.

Interactivity is also disabled when there are different patterns of missing values among the dependent variables. For details, see the section “Missing Values,” which follows.

---

## Missing Values

For an analysis involving one dependent variable, PROC ANOVA uses an observation if values are nonmissing for that dependent variable and for all the variables used in independent effects.

For an analysis involving multiple dependent variables without the [MANOVA](#) or [REPEATED](#) statement, or without the [MANOVA](#) option in the PROC ANOVA statement, a missing value in one dependent variable does not eliminate the observation from the analysis of other nonmissing dependent variables. For an analysis with the [MANOVA](#) or [REPEATED](#) statement, or with the [MANOVA](#) option in the PROC ANOVA statement, the ANOVA procedure requires values for all dependent variables to be nonmissing for an observation before the observation can be used in the analysis.

During processing, PROC ANOVA groups the dependent variables by their pattern of missing values across observations so that sums and cross products can be collected in the most efficient manner.

If your data have different patterns of missing values among the dependent variables, interactivity is disabled. This could occur when some of the variables in your data set have missing values and either of the following conditions obtain:

- You do not use the [MANOVA](#) option in the PROC ANOVA statement.
- You do not use a [MANOVA](#) or [REPEATED](#) statement before the first RUN statement.

---

## Output Data Set

The OUTSTAT= option in the PROC ANOVA statement produces an output data set that contains the following:

- the BY variables, if any
- `_TYPE_`, a new character variable. This variable has the value 'ANOVA' for observations corresponding to sums of squares; it has the value 'CANCORR', 'STRUCTUR', or 'SCORE' if a canonical analysis is performed through the `MANOVA` statement and no `M=` matrix is specified.
- `_SOURCE_`, a new character variable. For each observation in the data set, `_SOURCE_` contains the name of the model effect from which the corresponding statistics are generated.
- `_NAME_`, a new character variable. The variable `_NAME_` contains the name of one of the dependent variables in the model or, in the case of canonical statistics, the name of one of the canonical variables (CAN1, CAN2, and so on).
- four new numeric variables, SS, DF, F, and PROB, containing sums of squares, degrees of freedom, *F* values, and probabilities, respectively, for each model or contrast sum of squares generated in the analysis. For observations resulting from canonical analyses, these variables have missing values.
- if there is more than one dependent variable, then variables with the same names as the dependent variables represent
  - for `_TYPE_='ANOVA'`, the crossproducts of the hypothesis matrices
  - for `_TYPE_='CANCORR'`, canonical correlations for each variable
  - for `_TYPE_='STRUCTUR'`, coefficients of the total structure matrix
  - for `_TYPE_='SCORE'`, raw canonical score coefficients

The output data set can be used to perform special hypothesis tests (for example, with the IML procedure in SAS/IML software), to reformat output, to produce canonical variates (through the SCORE procedure), or to rotate structure matrices (through the FACTOR procedure).

---

## Computational Method

Let  $\mathbf{X}$  represent the  $n \times p$  design matrix. The columns of  $\mathbf{X}$  contain only 0s and 1s. Let  $\mathbf{Y}$  represent the  $n \times 1$  vector of dependent variables.

In the GLM procedure,  $\mathbf{X}'\mathbf{X}$ ,  $\mathbf{X}'\mathbf{Y}$ , and  $\mathbf{Y}'\mathbf{Y}$  are formed in main storage. However, in the ANOVA procedure, only the diagonals of  $\mathbf{X}'\mathbf{X}$  are computed, along with  $\mathbf{X}'\mathbf{Y}$  and  $\mathbf{Y}'\mathbf{Y}$ . Thus, PROC ANOVA saves a considerable amount of storage as well as time. The memory requirements for PROC ANOVA are asymptotically linear functions of  $n^2$  and  $nr$ , where  $n$  is the number of dependent variables and  $r$  the number of independent parameters.

The elements of  $\mathbf{X}'\mathbf{Y}$  are cell totals, and the diagonal elements of  $\mathbf{X}'\mathbf{X}$  are cell frequencies. Since PROC ANOVA automatically pools omitted effects into the next higher-level effect containing the names of the omitted effect (or within-error), a slight modification to the rules given by Searle (1971, p. 389) is used.

1. PROC ANOVA computes the sum of squares for each effect as if it is a main effect. In other words, for each effect, PROC ANOVA squares each cell total and divides by its cell frequency. The procedure then adds these quantities together and subtracts the correction factor for the mean (total squared over  $N$ ).

2. For each effect involving two **CLASS** variable names, PROC ANOVA subtracts the SS for any main effect with a name that is contained in the two-factor effect.
3. For each effect involving three **CLASS** variable names, PROC ANOVA subtracts the SS for all main effects and two-factor effects with names that are contained in the three-factor effect. If effects involving four or more **CLASS** variable names are present, the procedure continues this process.

---

## Displayed Output

PROC ANOVA first displays a table that includes the following:

- the name of each variable in the **CLASS** statement
- the number of different values or Levels of the **CLASS** variables
- the Values of the **CLASS** variables
- the Number of observations in the data set and the number of observations excluded from the analysis because of missing values, if any

PROC ANOVA then displays an analysis-of-variance table for each dependent variable in the **MODEL** statement. This table breaks down the Total Sum of Squares for the dependent variable into the portion attributed to the Model and the portion attributed to Error. It also breaks down the Mean Square term, which is the Sum of Squares divided by the degrees of freedom (DF). The analysis-of-variance table also lists the following:

- the Mean Square for Error (MSE), which is an estimate of  $\sigma^2$ , the variance of the true errors
- the F Value, which is the ratio produced by dividing the Mean Square for the Model by the Mean Square for Error. It tests how well the model as a whole (adjusted for the mean) accounts for the dependent variable's behavior. This *F* test is a test of the null hypothesis that all parameters except the intercept are zero.
- the significance probability associated with the *F* statistic, labeled "Pr > F"
- R-Square,  $R^2$ , which measures how much variation in the dependent variable can be accounted for by the model. The R square statistic, which can range from 0 to 1, is the ratio of the sum of squares for the model divided by the sum of squares for the corrected total. In general, the larger the R square value, the better the model fits the data.
- C.V., the coefficient of variation, which is often used to describe the amount of variation in the population. The C.V. is 100 times the standard deviation of the dependent variable divided by the Mean. The coefficient of variation is often a preferred measure because it is unitless.
- Root MSE, which estimates the standard deviation of the dependent variable. Root MSE is computed as the square root of Mean Square for Error, the mean square of the error term.
- the Mean of the dependent variable

For each effect (or source of variation) in the model, PROC ANOVA then displays the following:

- DF, degrees of freedom
- Anova SS, the sum of squares, and the associated Mean Square
- the F Value for testing the hypothesis that the group means for that effect are equal
- $\text{Pr} > F$ , the significance probability value associated with the F Value

When you specify a **TEST** statement, PROC ANOVA displays the results of the requested tests. When you specify a **MANOVA** statement and the model includes more than one dependent variable, PROC ANOVA produces these additional statistics:

- the characteristic roots and vectors of  $\mathbf{E}^{-1}\mathbf{H}$  for each **H** matrix
- the Hotelling-Lawley trace
- Pillai's trace
- Wilks' lambda
- Roy's greatest root

See [Example 47.6](#) in Chapter 47, “[The GLM Procedure](#),” for an example of the MANOVA results. These MANOVA tests are discussed in Chapter 4, “[Introduction to Regression Procedures](#).”

---

## ODS Table Names

PROC ANOVA assigns a name to each table it creates. You can use these names to reference the table when using the Output Delivery System (ODS) to select tables and create output data sets. These names are listed in [Table 26.6](#). For more information about ODS, see Chapter 20, “[Using the Output Delivery System](#).”

**Table 26.6** ODS Tables Produced by PROC ANOVA

ODS Table Name	Description	Statement / Option
AltErrTests	Anova tests with error other than MSE	<b>TEST E=</b>
Bartlett	Bartlett's homogeneity of variance test	<b>MEANS / HOVTEST=BARTLETT</b>
CLDiffs	Multiple comparisons of pairwise differences	<b>MEANS / CLDIFF</b> or <b>DUNNETT</b> or (Unequal cells and not <b>LINES</b> )
CLDiffsInfo	Information for multiple comparisons of pairwise differences	<b>MEANS / CLDIFF</b> or <b>DUNNETT</b> or (Unequal cells and not <b>LINES</b> )
CLMeans	Multiple comparisons of means with confidence/comparison interval	<b>MEANS / CLM</b> with ( <b>BON</b> or <b>GABRIEL</b> or <b>SCHEFFE</b> or <b>SIDAK</b> or <b>SMM</b> or <b>T</b> or <b>LSD</b> )

Table 26.6 *continued*

ODS Table Name	Description	Statement / Option
CLMeansInfo	Information for multiple comparisons of means with confidence/comparison interval	MEANS / CLM
CanAnalysis	Canonical analysis	(MANOVA or REPEATED) / CANONICAL
CanCoef	Canonical coefficients	(MANOVA or REPEATED) / CANONICAL
CanStructure	Canonical structure	(MANOVA or REPEATED) / CANONICAL
CharStruct	Characteristic roots and vectors	(MANOVA / not CANONICAL) or (REPEATED / PRINTRV)
ClassLevels	Classification variable levels	CLASS statement
DependentInfo	Simultaneously analyzed dependent variables	default when there are multiple dependent variables with different patterns of missing values
Epsilons	Greenhouse-Geisser and Huynh-Feldt epsilons	REPEATED statement
ErrorSSCP	Error SSCP matrix	(MANOVA or REPEATED) / PRINTE
FitStatistics	R-Square, C.V., Root MSE, and dependent mean	default
HOVFTest	Homogeneity of variance ANOVA	MEANS / HOVTEST
HypothesisSSCP	Hypothesis SSCP matrix	(MANOVA or REPEATED) / PRINTE
MANOVATransform	Multivariate transformation matrix	MANOVA / M=
MCLines	Multiple comparisons LINES output	MEANS / LINES or ((DUNCAN or WALLER or SNK or REGWQ) and not (CLDIFF or CLM)) or (Equal cells and not CLDIFF)
MCLinesInfo	Information for multiple comparison LINES output	MEANS / LINES or ((DUNCAN or WALLER or SNK or REGWQ) and not (CLDIFF or CLM)) or (Equal cells and not CLDIFF)
MCLinesRange	Ranges for multiple range MC tests	MEANS / LINES or ((DUNCAN or WALLER or SNK or REGWQ) and not (CLDIFF or CLM)) or (Equal cells and not CLDIFF)
Means	Group means	MEANS statement
ModelANOVA	ANOVA for model terms	default
MultStat	Multivariate tests	MANOVA statement
NObs	Number of observations	default
OverallANOVA	Over-all ANOVA	default
PartialCorr	Partial correlation matrix	(MANOVA or REPEATED) / PRINTE

**Table 26.6** *continued*

ODS Table Name	Description	Statement / Option
RepeatedTransform	Repeated transformation matrix	<b>REPEATED</b> (CONTRAST or HELMERT or MEAN or POLYNOMIAL or PROFILE)
RepeatedLevelInfo	Correspondence between dependents and repeated measures levels	<b>REPEATED</b> statement
Sphericity Tests	Sphericity tests Summary ANOVA for specified MANOVA H= effects	<b>REPEATED</b> / <b>PRINTE MANOVA</b> / <b>H= SUMMARY</b>
Welch	Welch's ANOVA	<b>MEANS</b> / <b>WELCH</b>

## ODS Graphics

Statistical procedures use ODS Graphics to create graphs as part of their output. ODS Graphics is described in detail in Chapter 21, “[Statistical Graphics Using ODS](#).”

Before you create graphs, ODS Graphics must be enabled (for example, by specifying the ODS GRAPHICS ON statement). For more information about enabling and disabling ODS Graphics, see the section “[Enabling and Disabling ODS Graphics](#)” on page 607 in Chapter 21, “[Statistical Graphics Using ODS](#).”

The overall appearance of graphs is controlled by ODS styles. Styles and other aspects of using ODS Graphics are discussed in the section “[A Primer on ODS Statistical Graphics](#)” on page 606 in Chapter 21, “[Statistical Graphics Using ODS](#).”

When ODS Graphics is enabled, if you specify a one-way analysis of variance model, with just one independent classification variable, or if you use a **MEANS** statement, then the ANOVA procedure will produce a grouped box plot of the response values versus the classification levels. For an example of the box plot, see the section “[One-Way Layout with Means Comparisons](#)” on page 972.

## ODS Graph Names

PROC ANOVA produces a single graph, the name of which you can use for referencing it in ODS. The name is listed in [Table 26.7](#).

**Table 26.7** ODS Graphic Produced by PROC ANOVA

ODS Graph Name	Plot Description
BoxPlot	Box plot of observed response values by classification levels



## Examples: ANOVA Procedure

### Example 26.1: Randomized Complete Block With Factorial Treatment Structure

This example uses statements for the analysis of a randomized block with two treatment factors occurring in a factorial structure. The data, from Neter, Wasserman, and Kutner (1990, p. 941), are from an experiment examining the effects of codeine and acupuncture on post-operative dental pain in male subjects. Both treatment factors have two levels. The codeine levels are a codeine capsule or a sugar capsule. The acupuncture levels are two inactive acupuncture points or two active acupuncture points. There are four distinct treatment combinations due to the factorial treatment structure. The 32 subjects are assigned to eight blocks of four subjects each based on an assessment of pain tolerance.

The data for the analysis are balanced, so PROC ANOVA is used. The data are as follows:

```

title1 'Randomized Complete Block With Two Factors';
data PainRelief;
    input PainLevel Codeine Acupuncture Relief @@;
    datalines;
1 1 1 0.0 1 2 1 0.5 1 1 2 0.6 1 2 2 1.2
2 1 1 0.3 2 2 1 0.6 2 1 2 0.7 2 2 2 1.3
3 1 1 0.4 3 2 1 0.8 3 1 2 0.8 3 2 2 1.6
4 1 1 0.4 4 2 1 0.7 4 1 2 0.9 4 2 2 1.5
5 1 1 0.6 5 2 1 1.0 5 1 2 1.5 5 2 2 1.9
6 1 1 0.9 6 2 1 1.4 6 1 2 1.6 6 2 2 2.3
7 1 1 1.0 7 2 1 1.8 7 1 2 1.7 7 2 2 2.1
8 1 1 1.2 8 2 1 1.7 8 1 2 1.6 8 2 2 2.4
;

```

The variable PainLevel is the blocking variable, and Codeine and Acupuncture represent the levels of the two treatment factors. The variable Relief is the pain relief score (the higher the score, the more relief the patient has).

The following statements invokes PROC ANOVA. The blocking variable and treatment factors appear in the **CLASS** statement. The bar between the treatment factors Codeine and Acupuncture adds their main effects as well as their interaction Codeine\*Acupuncture to the model.

```

proc anova data=PainRelief;
    class PainLevel Codeine Acupuncture;
    model Relief = PainLevel Codeine|Acupuncture;
run;

```

The results from the analysis are shown in [Output 26.1.1](#), [Output 26.1.2](#), and [Output 26.1.3](#).

**Output 26.1.1** Class Level Information**Randomized Complete Block With Two Factors****The ANOVA Procedure**

Class Level Information									
Class	Levels	Values							
PainLevel	8	1 2 3 4 5 6 7 8							
Codeine	2	1 2							
Acupuncture	2	1 2							
Number of Observations Read									32
Number of Observations Used									32

**Output 26.1.2** ANOVA Table**Randomized Complete Block With Two Factors****The ANOVA Procedure****Dependent Variable: Relief**

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	10	11.33500000	1.13350000	78.37	<.0001
Error	21	0.30375000	0.01446429		
Corrected Total	31	11.63875000			

R-Square	Coeff Var	Root MSE	Relief Mean
0.973902	10.40152	0.120268	1.156250

The Class Level Information and ANOVA table are shown in [Output 26.1.1](#) and [Output 26.1.2](#). The classification level information summarizes the structure of the design. It is good to check these consistently in search of errors in the DATA step. The overall  $F$  test is significant, indicating that the model accounts for a significant amount of variation in the dependent variable.

**Output 26.1.3** Tests of Effects

Source	DF	Anova SS	Mean Square	F Value	Pr > F
PainLevel	7	5.59875000	0.79982143	55.30	<.0001
Codeine	1	2.31125000	2.31125000	159.79	<.0001
Acupuncture	1	3.38000000	3.38000000	233.68	<.0001
Codeine*Acupuncture	1	0.04500000	0.04500000	3.11	0.0923

[Output 26.1.3](#) shows tests of the effects. The blocking effect is significant; hence, it is useful. The interaction between codeine and acupuncture is significant at the 90% level but not at the 95% level. The significance level of this test should be determined before the analysis. The main effects of both treatment factors are highly significant.

## Example 26.2: Alternative Multiple Comparison Procedures

The following is a continuation of the first example in the section “One-Way Layout with Means Comparisons” on page 972. You are studying the effect of bacteria on the nitrogen content of red clover plants, and the analysis of variance shows a highly significant effect. The following statements create the data set and compute the analysis of variance as well as Tukey’s multiple comparisons test for pairwise differences between bacteria strains; the results are shown in [Figure 26.1](#), [Figure 26.2](#), and [Figure 26.3](#)

```

title1 'Nitrogen Content of Red Clover Plants';
data Clover;
    input Strain $ Nitrogen @@;
    datalines;
3DOK1  19.4 3DOK1  32.6 3DOK1  27.0 3DOK1  32.1 3DOK1  33.0
3DOK5  17.7 3DOK5  24.8 3DOK5  27.9 3DOK5  25.2 3DOK5  24.3
3DOK4  17.0 3DOK4  19.4 3DOK4   9.1 3DOK4  11.9 3DOK4  15.8
3DOK7  20.7 3DOK7  21.0 3DOK7  20.5 3DOK7  18.8 3DOK7  18.6
3DOK13 14.3 3DOK13 14.4 3DOK13 11.8 3DOK13 11.6 3DOK13 14.2
COMPOS 17.3 COMPOS 19.4 COMPOS 19.1 COMPOS 16.9 COMPOS 20.8
;

proc anova data=Clover;
    class Strain;
    model Nitrogen = Strain;
    means Strain / tukey;
run;

```

The interactivity of PROC ANOVA enables you to submit further **MEANS** statements without re-running the entire analysis. For example, the following command requests means of the Strain levels with Duncan’s multiple range test and the Waller-Duncan  $k$ -ratio  $t$  test.

```

    means Strain / duncan waller;
run;

```

Results of the Waller-Duncan  $k$ -ratio  $t$  test are shown in [Output 26.2.1](#).

### Output 26.2.1 Waller-Duncan $K$ -ratio $t$ Test

#### Nitrogen Content of Red Clover Plants

##### The ANOVA Procedure

##### Waller-Duncan $K$ -ratio $t$ Test for Nitrogen

Kratio	100
Error Degrees of Freedom	24
Error Mean Square	11.78867
F Value	14.37
Critical Value of $t$	1.91873
Minimum Significant Difference	4.1665

**Output 26.2.1** *continued*

Means with the same letter are not significantly different.			
Waller Grouping	Mean	N	Strain
A	28.820	5	3DOK1
B	23.980	5	3DOK5
B			
C	19.920	5	3DOK7
C			
C	18.700	5	COMPOS
E	14.640	5	3DOK4
E			
E	13.260	5	3DOK13

The Waller-Duncan  $k$ -ratio  $t$  test is a multiple range test. Unlike Tukey's test, this test does not operate on the principle of controlling Type I error. Instead, it compares the Type I and Type II error rates based on Bayesian principles (Steel and Torrie 1980).

The Waller Grouping column in [Output 26.2.1](#) shows which means are significantly different. From this test, you can conclude the following:

- The mean nitrogen content for strain 3DOK1 is higher than the means for all other strains.
- The mean nitrogen content for strain 3DOK5 is higher than the means for COMPOS, 3DOK4, and 3DOK13.
- The mean nitrogen content for strain 3DOK7 is higher than the means for 3DOK4 and 3DOK13.
- The mean nitrogen content for strain COMPOS is higher than the mean for 3DOK13.
- Differences between all other means are not significant based on this sample size.

[Output 26.2.2](#) shows the results of Duncan's multiple range test. Duncan's test is a result-guided test that compares the treatment means while controlling the comparison-wise error rate. You should use this test for planned comparisons only (Steel and Torrie 1980). The results and conclusions for this example are the same as for the Waller-Duncan  $k$ -ratio  $t$  test. This is not always the case.

**Output 26.2.2** Duncan's Multiple Range Test

Alpha	0.05				
Error Degrees of Freedom	24				
Error Mean Square	11.78867				
Number of Means	2	3	4	5	6
Critical Range	4.482	4.707	4.852	4.954	5.031

Means with the same letter are not significantly different.				
Duncan Grouping	Mean	N	Strain	
A	28.820	5	3DOK1	
B	23.980	5	3DOK5	
B				
C	19.920	5	3DOK7	
C				
C	18.700	5	COMPOS	
E	14.640	5	3DOK4	
E				
E	13.260	5	3DOK13	

Tukey and Least Significant Difference (LSD) tests are requested with the following **MEANS** statement. The **CLDIFF** option requests confidence intervals for both tests.

```
means Strain/ lsd tukey cldiff ;
run;
```

The **LSD** tests for this example are shown in **Output 26.2.3**, and they give the same results as the previous two multiple comparison tests. Again, this is not always the case.

**Output 26.2.3** T Tests (LSD)**Nitrogen Content of Red Clover Plants****The ANOVA Procedure****t Tests (LSD) for Nitrogen**

Alpha	0.05
Error Degrees of Freedom	24
Error Mean Square	11.78867
Critical Value of t	2.06390
Least Significant Difference	4.4818

Output 26.2.3 continued

Comparisons significant at the 0.05 level are indicated by ***.				
Strain Comparison	Difference Between Means	95% Confidence Limits		
3DOK1 - 3DOK5	4.840	0.358	9.322	***
3DOK1 - 3DOK7	8.900	4.418	13.382	***
3DOK1 - COMPOS	10.120	5.638	14.602	***
3DOK1 - 3DOK4	14.180	9.698	18.662	***
3DOK1 - 3DOK13	15.560	11.078	20.042	***
3DOK5 - 3DOK1	-4.840	-9.322	-0.358	***
3DOK5 - 3DOK7	4.060	-0.422	8.542	
3DOK5 - COMPOS	5.280	0.798	9.762	***
3DOK5 - 3DOK4	9.340	4.858	13.822	***
3DOK5 - 3DOK13	10.720	6.238	15.202	***
3DOK7 - 3DOK1	-8.900	-13.382	-4.418	***
3DOK7 - 3DOK5	-4.060	-8.542	0.422	
3DOK7 - COMPOS	1.220	-3.262	5.702	
3DOK7 - 3DOK4	5.280	0.798	9.762	***
3DOK7 - 3DOK13	6.660	2.178	11.142	***
COMPOS - 3DOK1	-10.120	-14.602	-5.638	***
COMPOS - 3DOK5	-5.280	-9.762	-0.798	***
COMPOS - 3DOK7	-1.220	-5.702	3.262	
COMPOS - 3DOK4	4.060	-0.422	8.542	
COMPOS - 3DOK13	5.440	0.958	9.922	***
3DOK4 - 3DOK1	-14.180	-18.662	-9.698	***
3DOK4 - 3DOK5	-9.340	-13.822	-4.858	***
3DOK4 - 3DOK7	-5.280	-9.762	-0.798	***
3DOK4 - COMPOS	-4.060	-8.542	0.422	
3DOK4 - 3DOK13	1.380	-3.102	5.862	
3DOK13 - 3DOK1	-15.560	-20.042	-11.078	***
3DOK13 - 3DOK5	-10.720	-15.202	-6.238	***
3DOK13 - 3DOK7	-6.660	-11.142	-2.178	***
3DOK13 - COMPOS	-5.440	-9.922	-0.958	***
3DOK13 - 3DOK4	-1.380	-5.862	3.102	

If you only perform the **LSD** tests when the overall model  $F$  test is significant, then this is called Fisher's protected **LSD** test. Note that the **LSD** tests should be used for planned comparisons.

The **TUKEY** tests shown in Output 26.2.4 find fewer significant differences than the other three tests. This is not unexpected, as the **TUKEY** test controls the Type I experimentwise error rate. For a complete discussion of multiple comparison methods, see the section “Multiple Comparisons” on page 3693 in Chapter 47, “The GLM Procedure.”

**Output 26.2.4** Tukey's Studentized Range Test

Alpha	0.05
Error Degrees of Freedom	24
Error Mean Square	11.78867
Critical Value of Studentized Range	4.37265
Minimum Significant Difference	6.7142

---

Comparisons significant at the 0.05 level are indicated by \*\*\*.

Strain Comparison	Difference Between Means	Simultaneous 95% Confidence Limits
3DOK1 - 3DOK5	4.840	-1.874 11.554
3DOK1 - 3DOK7	8.900	2.186 15.614 ***
3DOK1 - COMPOS	10.120	3.406 16.834 ***
3DOK1 - 3DOK4	14.180	7.466 20.894 ***
3DOK1 - 3DOK13	15.560	8.846 22.274 ***
3DOK5 - 3DOK1	-4.840	-11.554 1.874
3DOK5 - 3DOK7	4.060	-2.654 10.774
3DOK5 - COMPOS	5.280	-1.434 11.994
3DOK5 - 3DOK4	9.340	2.626 16.054 ***
3DOK5 - 3DOK13	10.720	4.006 17.434 ***
3DOK7 - 3DOK1	-8.900	-15.614 -2.186 ***
3DOK7 - 3DOK5	-4.060	-10.774 2.654
3DOK7 - COMPOS	1.220	-5.494 7.934
3DOK7 - 3DOK4	5.280	-1.434 11.994
3DOK7 - 3DOK13	6.660	-0.054 13.374
COMPOS - 3DOK1	-10.120	-16.834 -3.406 ***
COMPOS - 3DOK5	-5.280	-11.994 1.434
COMPOS - 3DOK7	-1.220	-7.934 5.494
COMPOS - 3DOK4	4.060	-2.654 10.774
COMPOS - 3DOK13	5.440	-1.274 12.154
3DOK4 - 3DOK1	-14.180	-20.894 -7.466 ***
3DOK4 - 3DOK5	-9.340	-16.054 -2.626 ***
3DOK4 - 3DOK7	-5.280	-11.994 1.434
3DOK4 - COMPOS	-4.060	-10.774 2.654
3DOK4 - 3DOK13	1.380	-5.334 8.094
3DOK13 - 3DOK1	-15.560	-22.274 -8.846 ***
3DOK13 - 3DOK5	-10.720	-17.434 -4.006 ***
3DOK13 - 3DOK7	-6.660	-13.374 0.054
3DOK13 - COMPOS	-5.440	-12.154 1.274
3DOK13 - 3DOK4	-1.380	-8.094 5.334

**Example 26.3: Split Plot**

In some experiments, treatments can be applied only to groups of experimental observations rather than separately to each observation. When there are two nested groupings of the observations on the basis of treatment application, this is known as a *split plot design*. For example, in integrated circuit fabrication it is

of interest to see how different manufacturing methods affect the characteristics of individual chips. However, much of the manufacturing process is applied to a relatively large wafer of material, from which many chips are made. Additionally, a chip's position within a wafer might also affect chip performance. These two groupings of chips—by wafer and by position-within-wafer—might form the *whole plots* and the *subplots*, respectively, of a split plot design for integrated circuits.

The following statements produce an analysis for a split-plot design. The **CLASS** statement includes the variables Block, A, and B, where B defines subplots within BLOCK\*A whole plots. The **MODEL** statement includes the independent effects Block, A, Block\*A, B, and A\*B. The **TEST** statement asks for an *F* test of the A effect that uses the Block\*A effect as the error term. The following statements produce [Output 26.3.1](#) and [Output 26.3.2](#):

```

title1 'Split Plot Design';
data Split;
    input Block 1 A 2 B 3 Response;
    datalines;
142 40.0
141 39.5
112 37.9
111 35.4
121 36.7
122 38.2
132 36.4
131 34.8
221 42.7
222 41.6
212 40.3
211 41.6
241 44.5
242 47.6
231 43.6
232 42.8
;

proc anova data=Split;
    class Block A B;
    model Response = Block A Block*A B A*B;
    test h=A e=Block*A;
run;

```

**Output 26.3.1** Class Level Information and ANOVA Table

### Split Plot Design

#### The ANOVA Procedure

Class Level Information		
Class	Levels	Values
Block	2	1 2
A	4	1 2 3 4
B	2	1 2



**Output 26.3.1** *continued*

Number of Observations Read 16

Number of Observations Used 16

**Split Plot Design****The ANOVA Procedure****Dependent Variable: Response**

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	182.0200000	16.5472727	7.85	0.0306
Error	4	8.4300000	2.1075000		
Corrected Total	15	190.4500000			

R-Square	Coeff Var	Root MSE	Response Mean
0.955736	3.609007	1.451723	40.22500

First, notice that the overall  $F$  test for the model is significant.

**Output 26.3.2** Tests of Effects

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Block	1	131.1025000	131.1025000	62.21	0.0014
A	3	40.1900000	13.3966667	6.36	0.0530
Block*A	3	6.9275000	2.3091667	1.10	0.4476
B	1	2.2500000	2.2500000	1.07	0.3599
A*B	3	1.5500000	0.5166667	0.25	0.8612

**Tests of Hypotheses Using the Anova MS for Block\*A as an Error Term**

Source	DF	Anova SS	Mean Square	F Value	Pr > F
A	3	40.19000000	13.39666667	5.80	0.0914

The effect of Block is significant. The effect of A is not significant: look at the  $F$  test produced by the **TEST** statement, not at the  $F$  test produced by default. Neither the B nor A\*B effects are significant. The test for Block\*A is irrelevant, as this is simply the main-plot error.

**Example 26.4: Latin Square Split Plot**

The data for this example is taken from Smith (1951). A Latin square design is used to evaluate six different sugar beet varieties arranged in a six-row (Rep) by six-column (Column) square. The data are collected over two harvests. The variable Harvest then becomes a split plot on the original Latin square design for whole plots. The following statements produce [Output 26.4.1](#), [Output 26.4.2](#), and [Output 26.4.3](#):

```

title1 'Sugar Beet Varieties';
title3 'Latin Square Split-Plot Design';
data Beets;
  do Harvest=1 to 2;
    do Rep=1 to 6;
      do Column=1 to 6;
        input Variety Y @;
        output;
      end;
    end;
  end;
  datalines;
3 19.1 6 18.3 5 19.6 1 18.6 2 18.2 4 18.5
6 18.1 2 19.5 4 17.6 3 18.7 1 18.7 5 19.9
1 18.1 5 20.2 6 18.5 4 20.1 3 18.6 2 19.2
2 19.1 3 18.8 1 18.7 5 20.2 4 18.6 6 18.5
4 17.5 1 18.1 2 18.7 6 18.2 5 20.4 3 18.5
5 17.7 4 17.8 3 17.4 2 17.0 6 17.6 1 17.6
3 16.2 6 17.0 5 18.1 1 16.6 2 17.7 4 16.3
6 16.0 2 15.3 4 16.0 3 17.1 1 16.5 5 17.6
1 16.5 5 18.1 6 16.7 4 16.2 3 16.7 2 17.3
2 17.5 3 16.0 1 16.4 5 18.0 4 16.6 6 16.1
4 15.7 1 16.1 2 16.7 6 16.3 5 17.8 3 16.2
5 18.3 4 16.6 3 16.4 2 17.6 6 17.1 1 16.5
;

proc anova data=Beets;
  class Column Rep Variety Harvest;
  model Y=Rep Column Variety Rep*Column*Variety
        Harvest Harvest*Rep
        Harvest*Variety;
  test h=Rep Column Variety e=Rep*Column*Variety;
  test h=Harvest          e=Harvest*Rep;
run;

```

**Output 26.4.1** Class Level Information**Sugar Beet Varieties****Latin Square Split-Plot Design****The ANOVA Procedure**

Class Level Information		
Class	Levels	Values
Column	6	1 2 3 4 5 6
Rep	6	1 2 3 4 5 6
Variety	6	1 2 3 4 5 6
Harvest	2	1 2

---

Number of Observations Read 72

---

Number of Observations Used 72

---

**Output 26.4.2** ANOVA Table**Sugar Beet Varieties****Latin Square Split-Plot Design****The ANOVA Procedure****Dependent Variable: Y**

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	46	98.9147222	2.1503200	7.22	<.0001
Error	25	7.4484722	0.2979389		
Corrected Total	71	106.3631944			

R-Square	Coeff Var	Root MSE	Y Mean
0.929971	3.085524	0.545838	17.69028

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Rep	5	4.32069444	0.86413889	2.90	0.0337
Column	5	1.57402778	0.31480556	1.06	0.4075
Variety	5	20.61902778	4.12380556	13.84	<.0001
Column*Rep*Variety	20	3.25444444	0.16272222	0.55	0.9144
Harvest	1	60.68347222	60.68347222	203.68	<.0001
Rep*Harvest	5	7.71736111	1.54347222	5.18	0.0021
Variety*Harvest	5	0.74569444	0.14913889	0.50	0.7729

First, note from [Output 26.4.2](#) that the overall model is significant.

**Output 26.4.3** Tests of Effects**Tests of Hypotheses Using the Anova MS for Column\*Rep\*Variety as an Error Term**

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Rep	5	4.32069444	0.86413889	5.31	0.0029
Column	5	1.57402778	0.31480556	1.93	0.1333
Variety	5	20.61902778	4.12380556	25.34	<.0001

**Tests of Hypotheses Using the Anova MS for Rep\*Harvest as an Error Term**

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Harvest	1	60.68347222	60.68347222	39.32	0.0015

[Output 26.4.3](#) shows that the effects for Rep and Harvest are significant, while the Column effect is not. The average Ys for the six different Varieties are significantly different. For these four tests, look at the output produced by the two **TEST** statements, not at the usual ANOVA procedure output. The Variety\*Harvest interaction is not significant. All other effects in the default output should either be tested by using the results from the **TEST** statements or are irrelevant as they are only error terms for portions of the model.

## Example 26.5: Strip-Split Plot

In this example, four different fertilizer treatments are laid out in vertical strips, which are then split into subplots with different levels of calcium. Soil type is stripped across the split-plot experiment, and the entire experiment is then replicated three times. The dependent variable is the yield of winter barley. The data come from the notes of G. Cox and A. Rotti.

The input data are the 96 values of Y, arranged so that the calcium value (Calcium) changes most rapidly, then the fertilizer value (Fertilizer), then the Soil value, and, finally, the Rep value. Values are shown for Calcium (0 and 1); Fertilizer (0, 1, 2, 3); Soil (1, 2, 3); and Rep (1, 2, 3, 4). The following example produces [Output 26.5.1](#), [Output 26.5.2](#), [Output 26.5.3](#), and [Output 26.5.4](#).

```

title1 'Strip-split Plot';
data Barley;
  do Rep=1 to 4;
    do Soil=1 to 3; /* 1=d 2=h 3=p */
      do Fertilizer=0 to 3;
        do Calcium=0,1;
          input Yield @;
          output;
        end;
      end;
    end;
  end;
datalines;
4.91 4.63 4.76 5.04 5.38 6.21 5.60 5.08
4.94 3.98 4.64 5.26 5.28 5.01 5.45 5.62
5.20 4.45 5.05 5.03 5.01 4.63 5.80 5.90
6.00 5.39 4.95 5.39 6.18 5.94 6.58 6.25
5.86 5.41 5.54 5.41 5.28 6.67 6.65 5.94
5.45 5.12 4.73 4.62 5.06 5.75 6.39 5.62
4.96 5.63 5.47 5.31 6.18 6.31 5.95 6.14
5.71 5.37 6.21 5.83 6.28 6.55 6.39 5.57
4.60 4.90 4.88 4.73 5.89 6.20 5.68 5.72
5.79 5.33 5.13 5.18 5.86 5.98 5.55 4.32
5.61 5.15 4.82 5.06 5.67 5.54 5.19 4.46
5.13 4.90 4.88 5.18 5.45 5.80 5.12 4.42
;

proc anova data=Barley;
  class Rep Soil Calcium Fertilizer;
  model Yield =
    Rep
    Fertilizer Fertilizer*Rep
    Calcium Calcium*Fertilizer Calcium*Rep(Fertilizer)
    Soil Soil*Rep
    Soil*Fertilizer Soil*Rep*Fertilizer
    Soil*Calcium Soil*Fertilizer*Calcium
    Soil*Calcium*Rep(Fertilizer);

```

```

test h=Fertilizer                e=Fertilizer*Rep;
test h=Calcium calcium*fertilizer e=Calcium*Rep(Fertilizer);
test h=Soil                      e=Soil*Rep;
test h=Soil*Fertilizer           e=Soil*Rep*Fertilizer;
test h=Soil*Calcium
      Soil*Fertilizer*Calcium     e=Soil*Calcium*Rep(Fertilizer);
means Fertilizer Calcium Soil Calcium*Fertilizer;
run;

```

### Output 26.5.1 Class Level Information

#### Strip-split Plot

##### The ANOVA Procedure

Class Level Information			
Class	Levels	Values	
Rep	4	1 2 3 4	
Soil	3	1 2 3	
Calcium	2	0 1	
Fertilizer	4	0 1 2 3	

---

Number of Observations Read 96

Number of Observations Used 96

---

### Output 26.5.2 ANOVA Table

#### Strip-split Plot

##### The ANOVA Procedure

##### Dependent Variable: Yield

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	95	31.89149583	0.33569996	.	.
Error	0	0.00000000		.	.
Corrected Total	95	31.89149583			

R-Square	Coeff Var	Root MSE	Yield Mean
1.000000	.	.	5.427292

Output 26.5.2 continued

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Rep	3	6.27974583	2.09324861	.	.
Fertilizer	3	7.22127083	2.40709028	.	.
Rep*Fertilizer	9	6.08211250	0.67579028	.	.
Calcium	1	0.27735000	0.27735000	.	.
Calcium*Fertilizer	3	1.96395833	0.65465278	.	.
Rep*Calcium(Fertili)	12	1.76705833	0.14725486	.	.
Soil	2	1.92658958	0.96329479	.	.
Rep*Soil	6	1.66761042	0.27793507	.	.
Soil*Fertilizer	6	0.68828542	0.11471424	.	.
Rep*Soil*Fertilizer	18	1.58698125	0.08816563	.	.
Soil*Calcium	2	0.04493125	0.02246562	.	.
Soil*Calcium*Fertili	6	0.18936042	0.03156007	.	.
Rep*Soil*Calc(Ferti)	24	2.19624167	0.09151007	.	.

Notice in [Output 26.5.2](#) that the default tests against the residual error rate are all unavailable. This is because the `Soil*Calcium*Rep(Fertilizer)` term in the model takes up all the degrees of freedom, leaving none for estimating the residual error rate. This is appropriate in this case since the `TEST` statements give the specific error terms appropriate for testing each effect. [Output 26.5.3](#) displays the output produced by the various `TEST` statements. The only significant effect is the `Calcium*Fertilizer` interaction.

Output 26.5.3 Tests of Effects

Tests of Hypotheses Using the Anova MS for  
Rep\*Fertilizer as an Error Term

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Fertilizer	3	7.22127083	2.40709028	3.56	0.0604

Tests of Hypotheses Using the Anova MS for  
Rep\*Calcium(Fertili) as an Error Term

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Calcium	1	0.27735000	0.27735000	1.88	0.1950
Calcium*Fertilizer	3	1.96395833	0.65465278	4.45	0.0255

Tests of Hypotheses Using the Anova MS for  
Rep\*Soil as an Error Term

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Soil	2	1.92658958	0.96329479	3.47	0.0999

Tests of Hypotheses Using the Anova MS for  
Rep\*Soil\*Fertilizer as an Error Term

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Soil*Fertilizer	6	0.68828542	0.11471424	1.30	0.3063

Tests of Hypotheses Using the Anova MS for Rep\*Soil\*Calc(Ferti)  
as an Error Term

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Soil*Calcium	2	0.04493125	0.02246562	0.25	0.7843
Soil*Calcium*Fertili	6	0.18936042	0.03156007	0.34	0.9059

**Output 26.5.4** Results of MEANS statement

Yield				
Level of Fertilizer	N	Mean	Std Dev	
0	24	5.18416667	0.48266395	
1	24	5.12916667	0.38337082	
2	24	5.75458333	0.53293265	
3	24	5.64125000	0.63926801	

Yield				
Level of Calcium	N	Mean	Std Dev	
0	48	5.48104167	0.54186141	
1	48	5.37354167	0.61565219	

Yield				
Level of Soil	N	Mean	Std Dev	
1	32	5.54312500	0.55806369	
2	32	5.51093750	0.62176315	
3	32	5.22781250	0.51825224	

Yield				
Level of Calcium	Level of Fertilizer	N	Mean	Std Dev
0	0	12	5.34666667	0.45029956
0	1	12	5.08833333	0.44986530
0	2	12	5.62666667	0.44707806
0	3	12	5.86250000	0.52886027
1	0	12	5.02166667	0.47615569
1	1	12	5.17000000	0.31826233
1	2	12	5.88250000	0.59856077
1	3	12	5.42000000	0.68409197

Output 26.5.4 shows the results of the **MEANS** statement, displaying for various effects and combinations of effects, as requested. You can examine the Calcium\*Fertilizer means to understand the interaction better.

In this example, you could reduce memory requirements by omitting the Soil\*Calcium\*Rep(Fertilizer) effect from the model in the **MODEL** statement. This effect then becomes the **ERROR** effect, and you can omit the last **TEST** statement in the statements shown earlier. The test for the Soil\*Calcium effect is then given in the Analysis of Variance table in the top portion of output. However, for all other tests, you should look at the results from the **TEST** statement. In large models, this method might lead to significant reductions in memory requirements.

## References

- Bartlett, M. S. (1937). "Properties of Sufficiency and Statistical Tests." *Proceedings of the Royal Society of London, Series A* 160:268–282.
- Brown, M. B., and Forsythe, A. B. (1974). "Robust Tests for Equality of Variances." *Journal of the American Statistical Association* 69:364–367.
- Chi, Y.-Y., Gribbin, M. J., Lamers, Y., Gregory, J. F., III, and Muller, K. E. (2012). "Global Hypothesis Testing for High-Dimensional Repeated Measures Outcomes." *Statistics in Medicine* 31:724–742.
- Erdman, L. W. (1946). "Studies to Determine If Antibiosis Occurs among Rhizobia." *Journal of the American Society of Agronomy* 38:251–258.
- Fisher, R. A. (1942). *The Design of Experiments*. 3rd ed. Edinburgh: Oliver & Boyd.
- Freund, R. J., Littell, R. C., and Spector, P. C. (1986). *SAS System for Linear Models*. 1986 ed. Cary, NC: SAS Institute Inc.
- Graybill, F. A. (1961). *An Introduction to Linear Statistical Models*. Vol. 1. New York: McGraw-Hill.
- Henderson, C. R. (1953). "Estimation of Variance and Covariance Components." *Biometrics* 9:226–252.
- Huynh, H., and Feldt, L. S. (1976). "Estimation of the Box Correction for Degrees of Freedom from Sample Data in the Randomized Block and Split Plot Designs." *Journal of Educational Statistics* 1:69–82.
- Lecoutre, B. (1991). "A Correction for the Epsilon Approximate Test with Repeated Measures Design with Two or More Independent Groups." *Journal of Educational Statistics* 16:371–372.
- Levene, H. (1960). "Robust Tests for the Equality of Variance." In *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, edited by I. Olkin, S. G. Ghurye, W. Hoeffding, W. G. Madow, and H. B. Mann, 278–292. Palo Alto, CA: Stanford University Press.
- Neter, J., Wasserman, W., and Kutner, M. H. (1990). *Applied Linear Statistical Models*. 3rd ed. Homewood, IL: Irwin.
- O'Brien, R. G. (1979). "A General ANOVA Method for Robust Tests of Additive Models for Variances." *Journal of the American Statistical Association* 74:877–880.
- O'Brien, R. G. (1981). "A Simple Test for Variance Effects in Experimental Designs." *Psychological Bulletin* 89:570–574.
- Remington, R. D., and Schork, M. A. (1970). *Statistics with Applications to the Biological and Health Sciences*. Englewood Cliffs, NJ: Prentice-Hall.
- Scheffé, H. (1959). *The Analysis of Variance*. New York: John Wiley & Sons.
- Schlotzhauer, S. D., and Littell, R. C. (1987). *SAS System for Elementary Statistical Analysis*. Cary, NC: SAS Institute Inc.
- Searle, S. R. (1971). *Linear Models*. New York: John Wiley & Sons.



- Smith, W. G. (1951). Dissertation notes on Canadian Sugar Factories Ltd., Taber, Alberta.
- Snedecor, G. W., and Cochran, W. G. (1967). *Statistical Methods*. 6th ed. Ames: Iowa State University Press.
- Steel, R. G. D., and Torrie, J. H. (1980). *Principles and Procedures of Statistics*. 2nd ed. New York: McGraw-Hill.
- Welch, B. L. (1951). "On the Comparison of Several Mean Values: An Alternative Approach." *Biometrika* 38:330–336.

# Subject Index

- absorption of effects
  - ANOVA procedure, 983
- adjacent-level contrasts, 997
- alpha level
  - ANOVA procedure, 990
- analysis of variance, *see also* ANOVA procedure
  - multivariate (ANOVA), 985
  - one-way layout, example, 972
  - one-way, variance-weighted, 994
  - within-subject factors, repeated measurements, 998
- ANOVA procedure
  - absorption of effects, 983
  - alpha level, 990
  - balanced data, 972
  - Bartlett's test, 992
  - block diagonal matrices, 972
  - Brown and Forsythe's test, 992
  - canonical analysis, 987
  - characteristic roots and vectors, 986
  - complete block design, 976
  - computational methods, 1004
  - confidence intervals, 991
  - contrasts, 997
  - dependent variable, 972
  - disk space, 981
  - effect specification, 1000
  - factor name, 995, 996
  - homogeneity of variance tests, 992
  - hypothesis tests, 999
  - independent variable, 972
  - interactive use, 1002
  - interactivity and missing values, 1003
  - introductory example, 972
  - level values, 995, 996
  - Levene's test for homogeneity of variance, 992
  - means, 989
  - memory requirements, 983, 1004
  - missing values, 981, 1003
  - model specification, 1000
  - multiple comparison procedures, 989
  - multiple comparisons, 991–994
  - multivariate analysis of variance, 981, 985
  - O'Brien's test, 992
  - ODS graph names, 1008
  - ODS table names, 1006
  - ordering of effects, 981
  - orthonormalizing transformation matrix, 987
  - output data sets, 982, 1003
  - pooling, automatic, 1002
  - repeated measures, 995
  - sphericity tests, 998
  - SSCP matrix for multivariate tests, 986
  - transformations, 995, 996
  - transformations for MANOVA, 986
  - unbalanced data, caution, 972
  - Welch's ANOVA, 994
  - WHERE statement, 1003
- at sign (@) operator
  - ANOVA procedure, 1001
- balanced data
  - ANOVA procedure, 972
- bar (|) operator
  - ANOVA procedure, 1001
- Bartlett's test
  - ANOVA procedure, 992
- block diagonal matrices
  - ANOVA procedure, 972
- Bonferroni *t* test, 991
- Brown and Forsythe's test
  - ANOVA procedure, 992
- canonical analysis
  - ANOVA procedure, 987
  - repeated measurements, 997
- canonical variables
  - ANOVA procedure, 987
- categorical variables, *see* classification variables
- characteristic roots and vectors
  - ANOVA procedure, 986
- classification variables
  - ANOVA procedure, 972, 1000
- complete block design
  - example (ANOVA), 976
- completely randomized design
  - examples, 1009
- confidence intervals
  - means (ANOVA), 991
  - pairwise differences (ANOVA), 991
- continuous variables, 1000
- contrasts
  - repeated measurements (ANOVA), 996, 997
- crossed effects
  - specifying (ANOVA), 1000, 1001
- dependent effect, definition, 1000

- difference between means
  - confidence intervals, 991
- discrete variables, *see* classification variables
- Duncan's multiple range test, 991
- Duncan-Waller test, 994
  - error seriousness ratio, 992
  - multiple comparison (ANOVA), 1012
- Dunnett's test
  - one-tailed lower, 991
  - one-tailed upper, 991
  - two-tailed, 991
- effect
  - definition, 1000
  - specification (ANOVA), 1000
- Einot and Gabriel's multiple range test
  - ANOVA procedure, 993
- error seriousness ratio
  - Waller-Duncan test, 992
- Fisher's LSD test, 993
- Gabriel's multiple-comparison procedure
  - ANOVA procedure, 992
- GT2 multiple-comparison method, 993
- Hochberg's GT2 multiple-comparison method, 993
- homogeneity of variance tests, 992
  - Bartlett's test (ANOVA), 992
  - Brown and Forsythe's test (ANOVA), 992
  - Levene's test (ANOVA), 992
  - O'Brien's test (ANOVA), 992
- honestly significant difference test, 994
- Hotelling-Lawley trace, 986
- HSD test, 994
- hypothesis tests
  - custom tests (ANOVA), 999
  - for intercept (ANOVA), 994
- independent variable
  - defined (ANOVA), 972
- interaction effects
  - specifying (ANOVA), 1000, 1001
- intercept
  - hypothesis tests for (ANOVA), 994
- Latin square design
  - ANOVA procedure, 1017
- least-significant-difference test, 993
- Levene's test for homogeneity of variance
  - ANOVA procedure, 992
- main effects
  - specifying (ANOVA), 1000
- means
  - ANOVA procedure, 989
- memory requirements
  - reduction of (ANOVA), 983
- model
  - specification (ANOVA), 1000
- multiple comparisons of means
  - ANOVA procedure, 989
  - Bonferroni *t* test, 991
  - Duncan's multiple range test, 991
  - Dunnett's test, 991
  - error mean square, 992
  - Fisher's LSD test, 993
  - Gabriel's procedure, 992
  - GT2 method, 993
  - Ryan-Einot-Gabriel-Welsch test, 993
  - Scheffé's procedure, 993
  - Sidak's adjustment, 993
  - SMM, 993
  - Student-Newman-Keuls test, 993
  - Tukey's studentized range test, 994
  - Waller-Duncan method, 992
  - Waller-Duncan test, 994
- multivariate analysis of variance, 981, 985
- nested effects
  - specifying (ANOVA), 1000, 1002
- Newman-Keuls' multiple range test, 993
- nominal variables, *see also* classification variables
- O'Brien's test for homogeneity of variance
  - ANOVA procedure, 992
- ODS graph names
  - ANOVA procedure, 1008
- orthogonal polynomial contrasts, 997
- orthonormalizing transformation matrix
  - ANOVA procedure, 987
- Pillai's trace, 986
- qualitative variables, *see* classification variables
- repeated measures
  - ANOVA procedure, 995
  - more than one factor (ANOVA), 995, 998
- response variable, 1000
- Roy's greatest root, 986
- Ryan's multiple range test, 993
- Scheffé's multiple-comparison procedure, 993
- Sidak's inequality, 993
- SMM multiple-comparison method, 993
- sphericity tests, 998
- split-plot design
  - ANOVA procedure, 999, 1015, 1017, 1020
- SSCP matrix

- displaying, for multivariate tests, 988
- for multivariate tests, 986
- strip-split-plot design
  - ANOVA procedure, 1020
- Student's multiple range test, 993
- Studentized maximum modulus
  - pairwise comparisons, 993
- transformation matrix
  - orthonormalizing, 987
- transformations
  - ANOVA procedure, 996
  - for multivariate ANOVA, 986
- Tukey's studentized range test, 994
- Tukey-Kramer test, 994
- unbalanced data
  - caution (ANOVA), 972
- variables, *see also* classification variables
- Waller-Duncan test, 994
  - error seriousness ratio, 992
  - multiple comparison (ANOVA), 1012
- Welch's ANOVA, 994
- Welsch's multiple range test, 993
- Wilks' lambda, 986



# Syntax Index

- ABSORB statement
  - ANOVA procedure, [983](#)
- \_ALL\_ effect
  - MANOVA statement (ANOVA), [986](#)
- ALPHA= option
  - MEANS statement (ANOVA), [990](#)
- ANOVA procedure
  - syntax, [980](#)
- ANOVA procedure, ABSORB statement, [983](#)
- ANOVA procedure, BY statement, [983](#)
- ANOVA procedure, CLASS statement, [984](#)
  - REF= option, [984](#)
  - REF= variable option, [984](#)
  - TRUNCATE option, [985](#)
- ANOVA procedure, FREQ statement, [985](#)
- ANOVA procedure, MANOVA statement, [985](#)
  - \_ALL\_ effect, [986](#)
  - CANONICAL option, [987](#)
  - E= option, [986](#)
  - H= option, [986](#)
  - INTERCEPT effect, [986](#)
  - M= option, [986](#)
  - MNAMES= option, [987](#)
  - MSTAT= option, [987](#)
  - ORTH option, [987](#)
  - PREFIX= option, [987](#)
  - PRINTE option, [988](#)
  - PRINTH option, [988](#)
  - SUMMARY option, [988](#)
- ANOVA procedure, MEANS statement, [989](#)
  - ALPHA= option, [990](#)
  - BON option, [991](#)
  - CLDIFF option, [991](#)
  - CLM option, [991](#)
  - DUNCAN option, [991](#)
  - DUNNETT option, [991](#)
  - DUNNETTL option, [991](#)
  - DUNNETTU option, [991](#)
  - E= option, [992](#)
  - GABRIEL option, [992](#)
  - HOVTEST option, [992](#)
  - KRATIO= option, [992](#)
  - LINES option, [993](#)
  - LSD option, [993](#)
  - NOSORT option, [993](#)
  - REGWQ option, [993](#)
  - SCHEFFE option, [993](#)
  - SIDAK option, [993](#)
  - SMM option, [993](#)
  - SNK option, [993](#)
  - TUKEY option, [994](#)
  - WALLER option, [994](#)
  - WELCH option, [994](#)
- ANOVA procedure, MODEL statement, [994](#)
  - INTERCEPT option, [994](#)
  - NOUNI option, [994](#)
- ANOVA procedure, PROC ANOVA statement, [980](#)
  - DATA= option, [981](#)
  - MANOVA option, [981](#)
  - MULTIPASS option, [981](#)
  - NAMELEN= option, [981](#)
  - NOPRINT option, [981](#)
  - ORDER= option, [981](#)
  - OUTSTAT= option, [982](#)
  - PLOTS= option, [982](#)
- ANOVA procedure, REPEATED statement, [995](#)
  - CANONICAL option, [997](#)
  - CONTRAST keyword, [997](#)
  - factor specification, [996](#)
  - HELMERT keyword, [997](#)
  - IDENTITY keyword, [997](#)
  - MEAN keyword, [997](#)
  - MSTAT= option, [997](#)
  - NOM option, [998](#)
  - NOU option, [998](#)
  - POLYNOMIAL keyword, [997](#)
  - PRINTE option, [998](#)
  - PRINTH option, [998](#)
  - PRINTM option, [998](#)
  - PRINTRV option, [998](#)
  - PROFILE keyword, [997](#)
  - SUMMARY option, [998](#)
  - UEPSDEF option, [998](#)
- ANOVA procedure, TEST statement, [999](#)
  - E= effects, [999](#)
  - H= effects, [999](#)
- BON option
  - MEANS statement (ANOVA), [991](#)
- BY statement
  - ANOVA procedure, [983](#)
- CANONICAL option
  - MANOVA statement (ANOVA), [987](#)
  - REPEATED statement (ANOVA), [997](#)
- CLASS statement
  - ANOVA procedure, [984](#)

- CLDIFF option
  - MEANS statement (ANOVA), 991
- CLM option
  - MEANS statement (ANOVA), 991
- CONTRAST keyword
  - REPEATED statement (ANOVA), 997
- DATA= option
  - PROC ANOVA statement, 981
- DUNCAN option
  - MEANS statement (ANOVA), 991
- DUNNETT option
  - MEANS statement (ANOVA), 991
- DUNNETTL option
  - MEANS statement (ANOVA), 991
- DUNNETTU option
  - MEANS statement (ANOVA), 991
- E= effects
  - TEST statement (ANOVA), 999
- E= option
  - MANOVA statement (ANOVA), 986
  - MEANS statement (ANOVA), 992
- FREQ statement
  - ANOVA procedure, 985
- GABRIEL option
  - MEANS statement (ANOVA), 992
- H= effects
  - TEST statement (ANOVA), 999
- H= option
  - MANOVA statement (ANOVA), 986
- HELMERT keyword
  - REPEATED statement (ANOVA), 997
- HOVTEST option
  - MEANS statement (ANOVA), 992
- IDENTITY keyword
  - REPEATED statement (ANOVA), 997
- INTERCEPT effect
  - MANOVA statement (ANOVA), 986
- INTERCEPT option
  - MODEL statement (ANOVA), 994
- KRATIO= option
  - MEANS statement (ANOVA), 992
- LINES option
  - MEANS statement (ANOVA), 993
- LSD option
  - MEANS statement (ANOVA), 993
- M= option
  - MANOVA statement (ANOVA), 986
- MANOVA option
  - PROC ANOVA statement, 981
- MANOVA statement
  - ANOVA procedure, 985
- MEAN keyword
  - REPEATED statement (ANOVA), 997
- MEANS statement
  - ANOVA procedure, 989
- MNAMES= option
  - MANOVA statement (ANOVA), 987
- MODEL statement
  - ANOVA procedure, 994
- MSTAT= option
  - MANOVA statement (ANOVA), 987
  - REPEATED statement (ANOVA), 997
- MULTIPASS option
  - PROC ANOVA statement, 981
- NAMELEN= option
  - PROC ANOVA statement, 981
- NOM option
  - REPEATED statement (ANOVA), 998
- NOPRINT option
  - PROC ANOVA statement, 981
- NOSORT option
  - MEANS statement (ANOVA), 993
- NOU option
  - REPEATED statement (ANOVA), 998
- NOUNI option
  - MODEL statement (ANOVA), 994
- ORDER= option
  - PROC ANOVA statement, 981
- ORTH option
  - MANOVA statement (ANOVA), 987
- OUTSTAT= option
  - PROC ANOVA statement, 982
- PLOTS= option
  - PROC ANOVA statement, 982
- POLYNOMIAL keyword
  - REPEATED statement (ANOVA), 997
- PREFIX= option
  - MANOVA statement (ANOVA), 987
- PRINTE option
  - MANOVA statement (ANOVA), 988
  - REPEATED statement (ANOVA), 998
- PRINTH option
  - MANOVA statement (ANOVA), 988
  - REPEATED statement (ANOVA), 998
- PRINTM option
  - REPEATED statement (ANOVA), 998
- PRINTRV option
  - REPEATED statement (ANOVA), 998
- PROC ANOVA statement, *see* ANOVA procedure

- PROFILE keyword
  - REPEATED statement (ANOVA), [997](#)
- REF= option
  - CLASS statement (ANOVA), [984](#)
- REGWQ option
  - MEANS statement (ANOVA), [993](#)
- REPEATED statement
  - ANOVA procedure, [995](#)
- SCHEFFE option
  - MEANS statement (ANOVA), [993](#)
- SIDAK option
  - MEANS statement (ANOVA), [993](#)
- SMM option
  - MEANS statement (ANOVA), [993](#)
- SNK option
  - MEANS statement (ANOVA), [993](#)
- SUMMARY option
  - MANOVA statement (ANOVA), [988](#)
  - REPEATED statement (ANOVA), [998](#)
- TEST statement
  - ANOVA procedure, [999](#)
- TRUNCATE option
  - CLASS statement (ANOVA), [985](#)
- TUKEY option
  - MEANS statement (ANOVA), [994](#)
- UEPSDEF option
  - REPEATED statement (ANOVA), [998](#)
- WALLER option
  - MEANS statement (ANOVA), [994](#)
- WELCH option
  - MEANS statement (ANOVA), [994](#)
- WHERE statement
  - ANOVA procedure, [1003](#)