SAS/STAT® 14.3 User's Guide

The INBREED Procedure

This document is an individual chapter from *SAS/STAT® 14.3 User's Guide*.

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2017. *SAS/STAT® 14.3 User's Guide*. Cary, NC: SAS Institute Inc.

SAS/STAT® 14.3 User's Guide

Copyright © 2017, 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

September 2017

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 66
The INBREED Procedure

Contents

Overview: INBREED Procedure **4941**
Getting Started: INBREED Procedure **4942**
The Format of the Input Data Set 4942
Performing the Analysis 4943
Syntax: INBREED Procedure **4947**
PROC INBREED Statement 4947
BY Statement 4949
CLASS Statement 4949
GENDER Statement 4949
MATINGS Statement 4950
VAR Statement 4950
Details: INBREED Procedure **4951**
Missing Values 4951
DATA= Data Set 4951
Computational Details 4952
OUTCOV= Data Set 4958
Displayed Output 4960
ODS Table Names 4960
Examples: INBREED Procedure **4961**
Example 66.1: Monoecious Population Analysis 4961
Example 66.2: Pedigree Analysis 4963
Example 66.3: Pedigree Analysis with BY Groups 4964
References **4965**

Overview: INBREED Procedure

The INBREED procedure calculates the covariance or inbreeding coefficients for a pedigree. PROC INBREED is unique in that it handles very large populations.

The INBREED procedure has two modes of operation. One mode carries out analysis on the assumption that all the individuals belong to the same generation. The other mode divides the population into nonoverlapping generations and analyzes each generation separately, assuming that the parents of individuals in the current generation are defined in the previous generation.

PROC INBREED also computes averages of the covariance or inbreeding coefficients within sex categories if the sex of individuals is known.

Getting Started: INBREED Procedure

This section demonstrates how you can use the INBREED procedure to calculate the inbreeding or covariance coefficients for a pedigree, how you can control the analysis mode if the population consists of nonoverlapping generations, and how you can obtain averages within sex categories.

For you to use PROC INBREED effectively, your input data set must have a definite format. The following sections first introduce this format for a fictitious population and then demonstrate how you can analyze this population by using the INBREED procedure.

The Format of the Input Data Set

The SAS data set used as input to the INBREED procedure must contain an observation for each individual. Each observation must include one variable identifying the individual and two variables identifying the individual's parents. Optionally, an observation can contain a known covariance coefficient and a character variable defining the gender of the individual.

For example, consider the following data:

```
data Population;
   input Individual $ Parent1 $ Parent2 $
         Covariance Sex $ Generation;
   datalines;
Mark   George Lisa     .    M  1
Kelly  Scott  Lisa     .    F  1
Mike   George Amy      .    M  1
.      Mark   Kelly  0.50   .  1
David  Mark   Kelly    .    M  2
Merle  Mike   Jane     .    F  2
Jim    Mark   Kelly  0.50   M  2
Mark   Mike   Kelly    .    M  2
;
```

It is important to order the pedigree observations so that individuals are defined before they are used as parents of other individuals. The family relationships between individuals cannot be ascertained correctly unless you observe this ordering. Also, older individuals must precede younger ones. For example, 'Mark' appears as the first parent of 'David' at observation 5; therefore, his observation needs to be defined prior to observation 5. Indeed, this is the case (see observation 1). Also, 'David' is older than 'Jim', whose observation appears after the observation for 'David', as is appropriate.

In populations with distinct, nonoverlapping generations, the older generation (parents) must precede the younger generation. For example, the individuals defined in Generation=1 appear as parents of individuals defined in Generation=2.

PROC INBREED produces warning messages when a parent cannot be found. For example, 'Jane' appears as the second parent of the individual 'Merle' even though there are no previous observations defining her own parents. If the population is treated as an overlapping population, that is, if the generation grouping is ignored, then the procedure inserts an observation for 'Jane' with missing parents just before the sixth observation, which defines 'Merle' as follows:

```
Jane    .       .       .     F  2
Merle   Mike    Jane    .     F  2
```

However, if generation grouping is taken into consideration, then 'Jane' is defined as the last observation in Generation=1, as follows:

```
Mike    George Amy      .     M  1
Jane    .       .       .     F  1
```

In this latter case, however, the observation for 'Jane' is inserted after the computations are reported for the first generation. Therefore, she does not appear in the covariance/inbreeding matrix, even though her observation is used in computations for the second generation (see Figure 66.2).

If the data for an individual are duplicated, only the first occurrence of the data is used by the procedure, and a warning message is displayed to note the duplication. For example, individual 'Mark' is defined twice, at observations 1 and 8. If generation grouping is ignored, then this is an error and observation 8 is skipped. However, if the population is processed with respect to two distinct generations, then 'Mark' refers to two different individuals, one in Generation=1 and the other in Generation=2.

If a covariance is to be assigned between two individuals, then those individuals must be defined prior to the assignment observation. For example, a covariance of 0.50 can be assigned between 'Mark' and 'Kelly' since they are previously defined. Note that assignment statements must have different formats depending on whether the population is processed with respect to generations (see the section "DATA= Data Set" on page 4951 for further information). For example, while observation 4 is valid for nonoverlapping generations, it is invalid for a processing mode that ignores generation grouping. In this latter case, observation 7 indicates a valid assignment, and observation 4 is skipped.

The latest covariance specification between any given two individuals overrides the previous one between the same individuals.

Performing the Analysis

To compute the covariance coefficients for the overlapping generation mode, use the following statements:

```
proc inbreed data=Population covar matrix init=0.25;
run;
```

Here, the DATA= option names the SAS data set to be analyzed, and the COVAR and MATRIX options tell the procedure to output the covariance coefficients matrix. If you omit the COVAR option, the inbreeding coefficients are output instead of the covariance coefficients.

Note that the PROC INBREED statement also contains the INIT= option. This option gives an initial covariance between any individual and unknown individuals. For example, the covariance between any individual and 'Jane' would be 0.25, since 'Jane' is unknown, except when 'Jane' appears as a parent (see Figure 66.4).

Figure 66.1 Analysis for an Overlapping Population

The INBREED Procedure

Covariance Coefficients												
Individual	Parent1	Parent2	George	Lisa	Mark	Scott	Kelly	Amy	Mike	David	Jane	Merle	Jim
George			1.1250	0.2500	0.6875	0.2500	0.2500	0.2500	0.6875	0.4688	0.2500	0.4688	0.4688
Lisa			0.2500	1.1250	0.6875	0.2500	0.6875	0.2500	0.2500	0.6875	0.2500	0.2500	0.6875
Mark	George	Lisa	0.6875	0.6875	1.1250	0.2500	0.5000	0.2500	0.4688	0.8125	0.2500	0.3594	0.8125
Scott			0.2500	0.2500	0.2500	1.1250	0.6875	0.2500	0.2500	0.4688	0.2500	0.2500	0.4688
Kelly	Scott	Lisa	0.2500	0.6875	0.5000	0.6875	1.1250	0.2500	0.2500	0.8125	0.2500	0.2500	0.8125
Amy			0.2500	0.2500	0.2500	0.2500	0.2500	1.1250	0.6875	0.2500	0.2500	0.4688	0.2500
Mike	George	Amy	0.6875	0.2500	0.4688	0.2500	0.2500	0.6875	1.1250	0.3594	0.2500	0.6875	0.3594
David	Mark	Kelly	0.4688	0.6875	0.8125	0.4688	0.8125	0.2500	0.3594	1.2500	0.2500	0.3047	0.8125
Jane			0.2500	0.2500	0.2500	0.2500	0.2500	0.2500	0.2500	0.2500	1.1250	0.6875	0.2500
Merle	Mike	Jane	0.4688	0.2500	0.3594	0.2500	0.2500	0.4688	0.6875	0.3047	0.6875	1.1250	0.3047
Jim	Mark	Kelly	0.4688	0.6875	0.8125	0.4688	0.8125	0.2500	0.3594	0.8125	0.2500	0.3047	1.2500

Number of Individuals	11

In the previous example, PROC INBREED treats the population as a single generation. However, you might want to process the population with respect to distinct, nonoverlapping generations. To accomplish this, you need to identify the generation variable in a CLASS statement, as shown by the following statements:

```
proc inbreed data=Population covar matrix init=0.25;
   class Generation;
run;
```

Note that, in this case, the covariance matrix is displayed separately for each generation (see Figure 66.5).

Figure 66.2 Analysis for a Nonoverlapping Population

The INBREED Procedure

Generation = 1

Covariance Coefficients

Individual	Parent1	Parent2	Mark	Kelly	Mike
Mark	**George**	**Lisa**	1.1250	0.5000	0.4688
Kelly	**Scott**	**Lisa**	0.5000	1.1250	0.2500
Mike	**George**	**Amy**	0.4688	0.2500	1.1250

Number of Individuals	3

The INBREED Procedure

Generation = 2

Covariance Coefficients

Individual	Parent1	Parent2	David	Merle	Jim	Mark
David	**Mark**	**Kelly**	1.2500	0.3047	0.8125	0.5859
Merle	**Mike**	**Jane**	0.3047	1.1250	0.3047	0.4688
Jim	**Mark**	**Kelly**	0.8125	0.3047	1.2500	0.5859
Mark	**Mike**	**Kelly**	0.5859	0.4688	0.5859	1.1250

Number of Individuals	4

You might also want to see covariance coefficient averages within sex categories. This is accomplished by indicating the variable defining the gender of individuals in a GENDER statement and by adding the AVERAGE option to the PROC INBREED statement. For example, the following statements produce the covariance coefficient averages shown in Figure 66.3:

```
proc inbreed data=Population covar average init=0.25;
   class Generation;
   gender Sex;
run;
```

Figure 66.3 Averages within Sex Categories for a Nonoverlapping Generation

The INBREED Procedure

Generation = 1

Averages of Covariance Coefficient Matrix in Generation 1		
	On Diagonal	Below Diagonal
Male X Male	1.1250	0.4688
Male X Female	.	0.3750
Female X Female	1.1250	0.0000
Over Sex	1.1250	0.4063

Number of Males	2
Number of Females	1
Number of Individuals	3

The INBREED Procedure

Generation = 2

Averages of Covariance Coefficient Matrix in Generation 2		
	On Diagonal	Below Diagonal
Male X Male	1.2083	0.6615
Male X Female	.	0.3594
Female X Female	1.1250	0.0000
Over Sex	1.1875	0.5104

Number of Males	3
Number of Females	1
Number of Individuals	4

Syntax: INBREED Procedure

The following statements are available in the INBREED procedure:

PROC INBREED < *options* > **;**
BY *variables* **;**
CLASS *variable* **;**
GENDER *variable* **;**
MATINGS *individual-list1 / mate-list1* <, . . . , *individual-listn / mate-listn*> **;**
VAR *variables* **;**

The PROC INBREED statement is required. Items within angle brackets (< >) are optional. The syntax of each statement is described in the following sections.

PROC INBREED Statement

PROC INBREED < *options* > **;**

The PROC INBREED statement invokes the INBREED procedure. Table 66.1 summarizes the options available in the PROC INBREED statement.

Table 66.1 PROC INBREED Statement Options

Option	Description
Specify Data Sets	
DATA=	Names the SAS data set
OUTCOV=	Names an output data set to contain the inbreeding coefficients
Control Type of Coefficient	
COVAR	Specifies that all coefficients output consist of covariance coefficients
SELFDIAG	Includes an individual's self-mating kinship coefficient
Control Displayed Tables	
AVERAGE	Produces a table of averages of coefficients
IND	Displays the individuals' inbreeding coefficients
MATRIX	Displays the inbreeding coefficient matrix
Specify Default Covariance Value	
INIT=	Specifies the covariance value
Suppress Output	
INDL	Displays individuals' coefficients for only the last generation
MATRIXL	Displays coefficients for only the last generation
NOPRINT	Suppresses the display of all output

AVERAGE

A

produces a table of averages of coefficients for each pedigree of offspring. The AVERAGE option is used together with the GENDER statement to average the inbreeding/covariance coefficients within sex categories.

COVAR

C

specifies that all coefficients output consist of covariance coefficients rather than inbreeding coefficients.

DATA=*SAS-data-set*

names the SAS data set to be used by PROC INBREED. If you omit the DATA= option, the most recently created SAS data set is used.

IND

I

displays the individuals' inbreeding coefficients (diagonal of the inbreeding coefficients matrix) for each pedigree of offspring.

If you also specify the COVAR option, the individuals' covariance coefficients (diagonal of the covariance coefficients matrix) are displayed.

INDL

displays individuals' coefficients for only the last generation of a multiparous population.

INIT=*cov*

specifies the covariance value *cov* if any of the parents are unknown; a value of 0 is assumed if you do not specify the INIT= option.

MATRIX

M

displays the inbreeding coefficient matrix for each pedigree of offspring.

If you also specify the COVAR option, the covariance matrices are displayed instead of inbreeding coefficients matrices.

MATRIXL

displays coefficients for only the last generation of a multiparous population.

NOPRINT

suppresses the display of all output. Note that this option temporarily disables the Output Delivery System (ODS). For more information on ODS, see Chapter 20, "Using the Output Delivery System."

OUTCOV=*SAS-data-set*

names an output data set to contain the inbreeding coefficients. When the COVAR option is also specified, covariance estimates are output to the OUTCOV= data set instead of inbreeding coefficients.

SELFDIAG

includes an individual's self-mating kinship coefficient instead of the individual's inbreeding coefficient on the diagonal of the matrix in the OUTCOV= data set when the COVAR option is not specified.

BY Statement

BY *variables* **;**

You can specify a BY statement with PROC INBREED 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 INBREED 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).

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 *SAS Visual Data Management and Utility Procedures Guide*.

CLASS Statement

CLASS *variable* **;**

To analyze the population within nonoverlapping generations, you must specify the variable that identifies generations in a CLASS statement. Values of the generation variable, called *generation numbers*, must be integers, but generations are assumed to occur in the order of their input in the input data set rather than in numerical order of the generation numbers. The name of an individual needs to be unique only within its generation.

When the MATRIXL option or the INDL option is specified, each generation requires a unique generation number in order for the specified option to work correctly. If generation numbers are not unique, all the generations with a generation number that is the same as the last generation's are output.

GENDER Statement

GENDER *variable* **;**

The GENDER statement specifies a variable that indicates the sex of the individuals. Values of the sex variable must be character beginning with 'M' or 'F', for male or female. The GENDER statement is needed only when you specify the AVERAGE option to average the inbreeding/covariance coefficients within sex categories or when you want to include a gender variable in the OUTCOV= data set.

PROC INBREED makes the following assumptions regarding the gender of individuals:

- The first parent is always assumed to be the male. See the section "VAR Statement" on page 4950.
- The second parent is always assumed to be the female. See the section "VAR Statement" on page 4950.
- If the gender of an individual is missing or invalid, this individual is assumed to be a female unless the population is overlapping and this individual appears as the first parent in a later observation.

Any contradictions to these rules are reported in the SAS log.

MATINGS Statement

MATINGS *individual-list1 / mate-list1* < **,** *. . .* **,** *individual-listn / mate-listn* > **;**

You can specify the MATINGS statement with PROC INBREED to specify selected matings of individuals. Each individual given in *individual-list* is mated with each individual given in *mate-list*. You can write multiple mating specifications if you separate them by commas or asterisks. The procedure reports the inbreeding coefficients or covariances for each pair of mates. For example, you can use the following statement to specify the mating of an individual named 'David' with an individual named 'Jane':

```
matings david / jane;
```

VAR Statement

VAR *individual parent1 parent2* < *covariance* > **;**

The VAR statement specifies three or four variables: the first variable contains an individual's name, the second variable contains the name of the individual's first parent, and the third variable contains the name of the individual's second parent. An optional fourth variable assigns a known value to the covariance of the individual's first and second parents in the current generation.

The first three variables in the VAR statement can be either numeric or character; however, only the first 12 characters of a character variable are recognized by the procedure. The fourth variable, if specified, must be numeric.

If you omit the VAR statement, then the procedure uses the first three unaddressed variables as the names of the individual and its parents. (Unaddressed variables are those that are not referenced in any other PROC INBREED statement.) If the input data set contains an unaddressed fourth variable, then it becomes the covariance variable.

Details: INBREED Procedure

Missing Values

A missing value for a parent implies that the parent is unknown. Unknown parents are assumed to be unrelated and not inbred unless you specify the INIT= option.

When the value of the variable identifying the individual is missing, the observation is not added to the list of individuals. However, for a multiparous population, an observation with a missing individual is valid and is used for assigning covariances.

Missing covariance values are determined from the INIT=*cov* option, if specified. Observations with missing generation variables are excluded.

If the gender of an individual is missing, it is determined from the order in which it is listed on the first observation defining its progeny for an overlapping population. If it appears as the first parent, it is set to 'M'; otherwise, it is set to 'F'. When the gender of an individual cannot be determined, it is assigned a default value of 'F'.

DATA= Data Set

Each observation in the input data set should contain necessary information such as the identification of an individual and the first and second parents of an individual. In addition, if a CLASS statement is specified, each observation should contain the generation identification; and, if a GENDER statement is specified, each observation should contain the gender of an individual. Optionally, each observation might also contain the covariance between the first and the second parents. Depending on how many statements are specified with the procedure, there should be enough variables in the input data set containing this information.

If you omit the VAR statement, then the procedure uses the first three *unaddressed variables* in the input data set as the names of the individual and his or her parents. Unaddressed variables in the input data set are those variables that are not referenced by the procedure in any other statements, such as CLASS, GENDER, or BY statements. If the input data set contains an unaddressed fourth variable, then the procedure uses it as the covariance variable.

If the individuals given by the variables associated with the first and second parents are not in the population, they are added to the population. However, if they are in the population, they must be defined prior to the observation that gives their progeny.

When there is a CLASS statement, the functions of defining new individuals and assigning covariances must be separated. This is necessary because the parents of any given individual are defined in the previous generation, while covariances are assigned between individuals in the current generation.

Therefore, there could be two types of observations for a multiparous population:

- one to define new individuals in the current generation whose parents have been defined in the previous generation, as in the following, where the missing value is for the covariance variable:

```
Mark   George Lisa    .    M  1
Kelly  Scott  Lisa    .    F  1
```

- one to assign covariances between two individuals in the current generation, as in the following, where the individual's name is missing, 'Mark' and 'Kelly' are in the current generation, and the covariance coefficient between these two individuals is 0.50:

```
.      Mark   Kelly  0.50  .  1
```

Note that the observations defining individuals must precede the observation assigning a covariance value between them. For example, if a covariance is to be assigned between 'Mark' and 'Kelly', then both of them should be defined prior to the assignment observation.

Computational Details

This section describes the rules that the INBREED procedure uses to compute the covariance and inbreeding coefficients. Each computational rule is explained by an example referring to the fictitious population introduced in the section "Getting Started: INBREED Procedure" on page 4942.

Coancestry (or Kinship Coefficient)

To calculate the inbreeding coefficient and the covariance coefficients, use the degree of relationship by descent between the two parents, which is called *coancestry* or *kinship coefficient* (Falconer and Mackay 1996, p.85), or *coefficient of parentage* (Kempthorne 1957, p.73). Denote the coancestry between individuals X and Y by f_{XY}. For information on how to calculate the coancestries among a population, see the section "Calculation of Coancestry" on page 4953.

Covariance Coefficient (or Coefficient of Relationship)

The covariance coefficient between individuals X and Y is defined by

$$\mathrm{Cov}(\mathrm{X},\mathrm{Y}) = 2f_{\mathrm{XY}}$$

where f_{XY} is the coancestry between X and Y. The covariance coefficient is sometimes called the *coefficient of relationship* or the *theoretical correlation* (Falconer and Mackay (1996, p.153); Crow and Kimura (1970, p.134)). If a covariance coefficient cannot be calculated from the individuals in the population, it is assigned to an initial value. The initial value is set to 0 if the INIT= option is not specified or to *cov* if INIT=*cov*. Therefore, the corresponding initial coancestry is set to 0 if the INIT= option is not specified or to $\frac{1}{2}$*cov* if INIT=*cov*.

Inbreeding Coefficients

The inbreeding coefficient of an individual is the probability that the pair of alleles carried by the gametes that produced it are identical by descent (Falconer and Mackay (1996, Chapter 5), Kempthorne (1957, Chapter 5)). For individual X, denote its inbreeding coefficient by F_{X}. The inbreeding coefficient of an individual is equal to the coancestry between its parents. For example, if X has parents A and B, then the inbreeding coefficient of X is

$$F_{\mathrm{X}} = f_{\mathrm{AB}}$$

Calculation of Coancestry

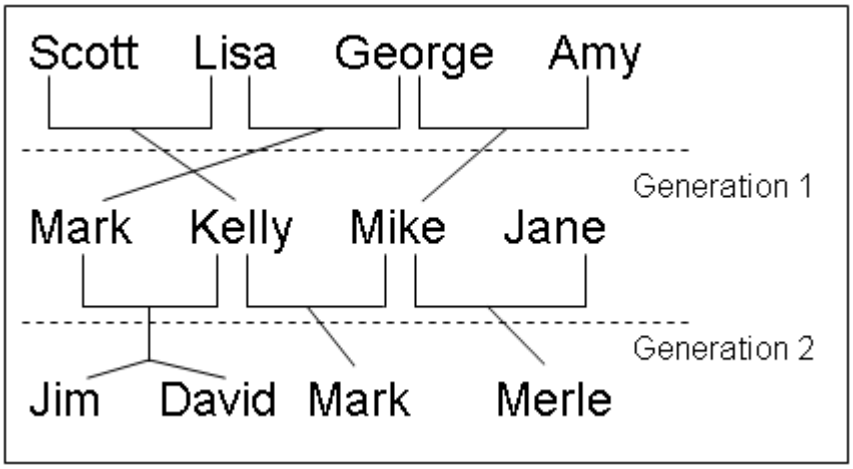
Given individuals X and Y, assume that X has parents A and B and that Y has parents C and D. For nonoverlapping generations, the basic rule to calculate the coancestry between X and Y is given by the following formula (Falconer and Mackay 1996, p.86):

$$f_{\mathrm{XY}} = \frac{1}{4}\left(f_{\mathrm{AC}} + f_{\mathrm{AD}} + f_{\mathrm{BC}} + f_{\mathrm{BD}}\right)$$

And the inbreeding coefficient for an offspring of X and Y, called Z, is the coancestry between X and Y:

$$F_{\mathrm{Z}} = f_{\mathrm{XY}}$$

Figure 66.4 Inbreeding Relationship for Nonoverlapping Population



For example, in Figure 66.4, 'Jim' and 'Mark' from Generation 2 are progenies of 'Mark' and 'Kelly' and of 'Mike' and 'Kelly' from Generation 1, respectively. The coancestry between 'Jim' and 'Mark' is

$$f_{\text{Jim,Mark}} = \frac{1}{\left(f_{\text{Mark,Mike}} + f_{\text{Mark,Kelly}} + f_{\text{Kelly,Mike}} + f_{\text{Kelly,Kelly}}\right)}$$

From the covariance matrix for Generation=1 in Figure 66.4 and the relationship that coancestry is half of the covariance coefficient,

$$f_{\text{Jim,Mark}} = \frac{1}{4}\left(\frac{0.4688}{2} + \frac{0.5}{2} + \frac{0.25}{2} + \frac{1.125}{2}\right) = 0.29298$$

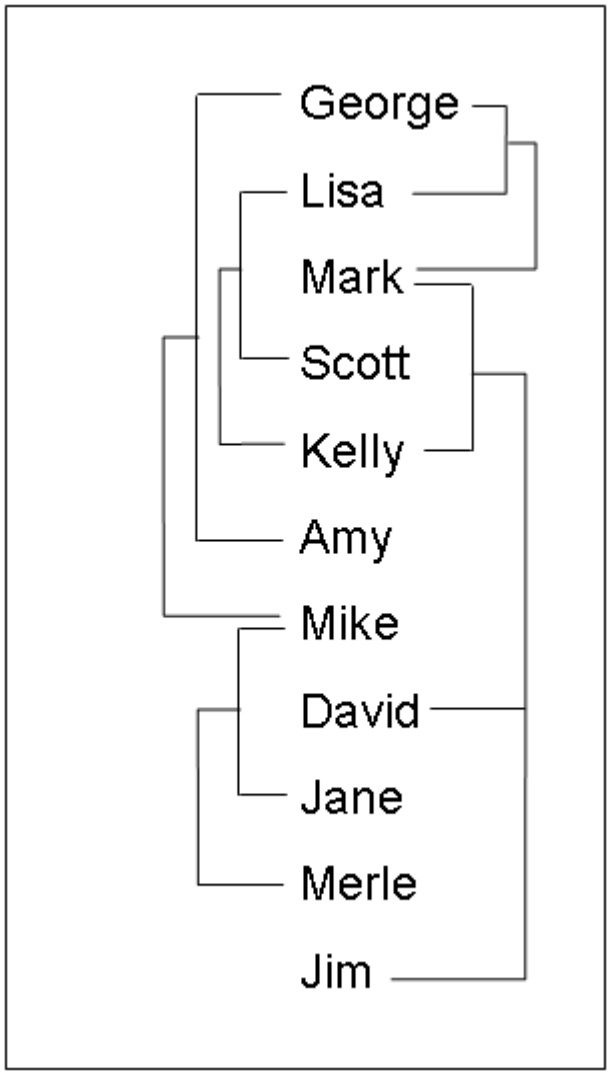
For overlapping generations, if X is older than Y, then the basic rule can be simplified to

$$F_{\text{Z}} = f_{\text{XY}} = \frac{1}{2}\left(f_{\text{XC}} + f_{\text{XD}}\right)$$

That is, the coancestry between X and Y is the average of coancestries between older X with younger Y's parents. For example, in Figure 66.5, the coancestry between 'Kelly' and 'David' is

$$f_{\text{Kelly,David}} = \frac{1}{2}\left(f_{\text{Kelly,Mark}} + f_{\text{Kelly,Kelly}}\right)$$

Figure 66.5 Inbreeding Relationship for Overlapping Population



This is so because 'Kelly' is defined before 'David'; therefore, 'Kelly' is not younger than 'David', and the parents of 'David' are 'Mark' and 'Kelly'. The covariance coefficient values Cov(Kelly,Mark) and Cov(Kelly,Kelly) from the matrix in Figure 66.5 yield that the coancestry between 'Kelly' and 'David' is

$$f_{\text{Kelly,David}} = \frac{1}{2}\left(\frac{0.5}{2} + \frac{1.125}{2}\right) = 0.40625$$

The numerical values for some initial coancestries must be known in order to use these rule. Either the parents of the first generation have to be unrelated, with $f = 0$ if the INIT= option is not specified in the PROC INBREED statement, or their coancestries must have an initial value of $\frac{1}{2}cov$, where cov is set by the INIT= option. Then the subsequent coancestries among their progenies and the inbreeding coefficients of their progenies in the rest of the generations are calculated by using these initial values.

Special rules need to be considered in the calculations of coancestries for the following cases.

Self-Mating

The coancestry for an individual X with itself, f_{XX}, is the inbreeding coefficient of a progeny that is produced by self-mating. The relationship between the inbreeding coefficient and the coancestry for self-mating is

$$f_{XX} = \frac{1}{2}(1 + F_X)$$

The inbreeding coefficient F_X can be replaced by the coancestry between X's parents A and B, f_{AB}, if A and B are in the population:

$$f_{XX} = \frac{1}{2}(1 + f_{AB})$$

If X's parents are not in the population, then F_X is replaced by the initial value $\frac{1}{2}cov$ if cov is set by the INIT= option, or F_X is replaced by 0 if the INIT= option is not specified. For example, the coancestry of 'Jim' with himself is

$$f_{\text{Jim,Jim}} = \frac{1}{2}\left(1 + f_{\text{Mark,Kelly}}\right)$$

where 'Mark' and 'Kelly' are the parents of 'Jim'. Since the covariance coefficient Cov(Mark,Kelly) is 0.5 in Figure 66.5 and also in the covariance matrix for GENDER=1 in Figure 66.4, the coancestry of 'Jim' with himself is

$$f_{\text{Jim,Jim}} = \frac{1}{2}\left(1 + \frac{0.5}{2}\right) = 0.625$$

When INIT=0.25, then the coancestry of 'Jane' with herself is

$$f_{\text{Jane,Jane}} = \frac{1}{2}\left(1 + \frac{0.25}{2}\right) = 0.5625$$

because 'Jane' is not an offspring in the population.

Offspring and Parent Mating

Assuming that X's parents are A and B, the coancestry between X and A is

$$f_{XA} = \frac{1}{2}(f_{AB} + f_{AA})$$

The inbreeding coefficient for an offspring of X and A, denoted by Z, is

$$F_Z = f_{XA} = \frac{1}{2}(f_{AB} + f_{AA})$$

For example, 'Mark' is an offspring of 'George' and 'Lisa', so the coancestry between 'Mark' and 'Lisa' is

$$f_{\text{Mark,Lisa}} = \frac{1}{2}\left(f_{\text{Lisa,George}} + f_{\text{Lisa,Lisa}}\right)$$

From the covariance coefficient matrix in Figure 66.5, $f_{\text{Lisa,George}} = 0.25/2 = 0.125$, $f_{\text{Lisa,Lisa}} = 1.125/2 = 0.5625$, so that

$$f_{\text{Mark,Lisa}} = \frac{1}{2}(0.125 + 0.5625) = 0.34375$$

Thus, the inbreeding coefficient for an offspring of 'Mark' and 'Lisa' is 0.34375.

Full Sibs Mating

This is a special case for the basic rule given at the beginning of the section "Calculation of Coancestry" on page 4953. If X and Y are full sibs with same parents A and B, then the coancestry between X and Y is

$$f_{XY} = \frac{1}{4}(2f_{AB} + f_{AA} + f_{BB})$$

and the inbreeding coefficient for an offspring of A and B, denoted by Z, is

$$F_Z = f_{XY} = \frac{1}{4}(2f_{AB} + f_{AA} + f_{BB})$$

For example, 'David' and 'Jim' are full sibs with parents 'Mark' and 'Kelly', so the coancestry between 'David' and 'Jim' is

$$f_{\text{David,Jim}} = \frac{1}{4}\left(2f_{\text{Mark,Kelly}} + f_{\text{Mark,Mark}} + f_{\text{Kelly,Kelly}}\right)$$

Since the coancestry is half of the covariance coefficient, from the covariance matrix in Figure 66.5,

$$f_{\text{David,Jim}} = \frac{1}{4}\left(2 \times \frac{0.5}{2} + \frac{1.125}{2} + \frac{1.125}{2}\right) = 0.40625$$

Unknown or Missing Parents

When individuals or their parents are unknown in the population, their coancestries are assigned by the value $\frac{1}{2}cov$ if cov is set by the INIT= option or by the value 0 if the INIT= option is not specified. That is, if either A or B is unknown, then

$$f_{\mathrm{AB}} = \frac{1}{2}cov$$

For example, 'Jane' is not in the population, and since 'Jane' is assumed to be defined just before the observation at which 'Jane' appears as a parent (that is, between observations 4 and 5), then 'Jane' is not older than 'Scott'. The coancestry between 'Jane' and 'Scott' is then obtained by using the simplified basic rule (see the section "Calculation of Coancestry" on page 4953):

$$f_{\mathrm{Scott,Jane}} = \frac{1}{2}\left(f_{\mathrm{Scott,\cdot}} + f_{\mathrm{Scott,\cdot}}\right)$$

Here, dots (·) indicate Jane's unknown parents. Therefore, $f_{\mathrm{Scott,\cdot}}$ is replaced by $\frac{1}{2}cov$, where cov is set by the INIT= option. If INIT=0.25, then

$$f_{\mathrm{Scott,Jane}} = \frac{1}{2}\left(\frac{0.25}{2} + \frac{0.25}{2}\right) = 0.125$$

For a more detailed discussion on the calculation of coancestries, inbreeding coefficients, and covariance coefficients, see Falconer and Mackay (1996); Kempthorne (1957); Crow and Kimura (1970).

OUTCOV= Data Set

The OUTCOV= data set has the following variables:

- a list of BY variables, if there is a BY statement
- the generation variable, if there is a CLASS statement
- the gender variable, if there is a GENDER statement
- _Type_, a variable indicating the type of observation. The valid values of the _Type_ variable are 'COV' for covariance estimates and 'INBREED' for inbreeding coefficients.
- _Panel_, a variable indicating the panel number used when populations delimited by BY groups contain different numbers of individuals. If there are n individuals in the first BY group and if any subsequent BY group contains a larger population, then its covariance/inbreeding matrix is divided into panels, with each panel containing n columns of data. If you put these panels side by side in increasing _Panel_ number order, then you can reconstruct the covariance or inbreeding matrix.

- _Col_, a variable used to name columns of the inbreeding or covariance matrix. The values of this variable start with 'COL', followed by a number indicating the column number. The names of the individuals corresponding to any given column *i* can be found by reading the individual's name across the row that has a _Col_ value of 'COL*i*'. When the inbreeding or covariance matrix is divided into panels, all the rows repeat for the first *n* columns, all the rows repeat for the next *n* columns, and so on.
- the variable containing the names of the individuals, that is, the first variable listed in the VAR statement
- the variable containing the names of the first parents, that is, the second variable listed in the VAR statement
- the variable containing the names of the second parents, that is, the third variable listed in the VAR statement
- a list of covariance variables Col1–Col*n*, where *n* is the maximum number of individuals in the first population

The functions of the variables _Panel_ and _Col_ can best be demonstrated by an example. Assume that there are three individuals in the first BY group and that, in the current BY group (Byvar=2), there are five individuals with the following covariance matrix.

COV	1	2	3	4	5
1	Cov(1,1)	Cov(1,2)	Cov(1,3)	Cov(1,4)	Cov(1,5)
2	Cov(2,1)	Cov(2,2)	Cov(2,3)	Cov(2,4)	Cov(2,5)
3	Cov(3,1)	Cov(3,2)	Cov(3,3)	Cov(3,4)	Cov(3,5)
4	Cov(4,1)	Cov(4,2)	Cov(4,3)	Cov(4,4)	Cov(4,5)
5	Cov(5,1)	Cov(5,2)	Cov(5,3)	Cov(5,4)	Cov(5,5)
	Panel 1			Panel 2	

Then the OUTCOV= data set appears as follows.

Byvar	_Panel_	_Col_	Individual	Parent	Parent2	Col1	Col2	Col3
2	1	COL1	1			Cov(1,1)	Cov(1,2)	Cov(1,3)
2	1	COL2	2			Cov(2,1)	Cov(2,2)	Cov(2,3)
2	1	COL3	3			Cov(3,1)	Cov(3,2)	Cov(3,3)
2	1		4			Cov(4,1)	Cov(4,2)	Cov(4,3)
2	1		5			Cov(5,1)	Cov(5,2)	Cov(5,3)
2	2		1			Cov(1,4)	Cov(1,5)	.
2	2		2			Cov(2,4)	Cov(2,5)	.
2	2		3			Cov(3,4)	Cov(3,5)	.
2	2	COL1	4			Cov(4,4)	Cov(4,5)	.
2	2	COL2	5			Cov(5,4)	Cov(5,5)	.

Notice that the first three columns go to the first panel (_Panel_=1), and the remaining two go to the second panel (_Panel_=2). Therefore, in the first panel, 'COL1', 'COL2', and 'COL3' correspond to individuals 1, 2, and 3, respectively, while in the second panel, 'COL1' and 'COL2' correspond to individuals 4 and 5, respectively.

Displayed Output

The INBREED procedure can output either covariance coefficients or inbreeding coefficients. Note that the following items can be produced for each generation if generations do not overlap.

The output produced by PROC INBREED can be any or all of the following items:

- a matrix of coefficients
- coefficients of the individuals
- coefficients for selected matings

ODS Table Names

PROC INBREED 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 66.2. For more information on ODS, see Chapter 20, "Using the Output Delivery System."

Table 66.2 ODS Tables Produced by PROC INBREED

ODS Table Name	Description	Statement	Option
AvgCovCoef	Averages of covariance coefficient matrix	GENDER	COVAR and AVERAGE
AvgInbreedingCoef	Averages of inbreeding coefficient matrix	GENDER	AVERAGE
CovarianceCoefficient	Covariance coefficient table	PROC	COVAR and MATRIX
InbreedingCoefficient	Inbreeding coefficient table	PROC	MATRIX
IndividualCovCoef	Covariance coefficients of individuals	PROC	IND and COVAR
IndividualInbreedingCoef	Inbreeding coefficients of individuals	PROC	IND
MatingCovCoef	Covariance coefficients of matings	MATINGS	COVAR
MatingInbreedingCoef	Inbreeding coefficients of matings	MATINGS	
NumberOfObservations	Number of observations	PROC	

Examples: INBREED Procedure

Example 66.1: Monoecious Population Analysis

The following example shows a covariance analysis within nonoverlapping generations for a monoecious population. Parents of generation 1 are unknown and therefore assumed to be unrelated. The following statements produce Output 66.1.1 through Output 66.1.3:

```
data Monoecious;
   input Generation Individual Parent1 Parent2 Covariance @@;
   datalines;
1 1 . .  .     1 2 . .  .     1 3 . .  .
2 1 1 1  .     2 2 1 2  .     2 3 2 3  .
3 1 1 2  .     3 2 1 3  .     3 3 2 1  .
3 4 1 3  .     3 . 2 3 0.50   3 . 4 3 1.135
;

title 'Inbreeding within Nonoverlapping Generations';
proc inbreed ind covar matrix data=Monoecious;
   class Generation;
run;
```

Output 66.1.1 Monoecious Population Analysis, Generation 1

Inbreeding within Nonoverlapping Generations

The INBREED Procedure

Generation = 1

Covariance Coefficients

Individual	Parent1	Parent2	1	2	3
1			1.0000	.	.
2			.	1.0000	.
3			.	.	1.0000

Covariance Coefficients of Individuals

Individual	Parent1	Parent2	Coefficient
1			1.0000
2			1.0000
3			1.0000

Number of Individuals	3

Output 66.1.2 Monoecious Population Analysis, Generation 2

Inbreeding within Nonoverlapping Generations

The INBREED Procedure

Generation = 2

Covariance Coefficients					
Individual	Parent1	Parent2	1	2	3
1	1	1	1.5000	0.5000	.
2	1	2	0.5000	1.0000	0.2500
3	2	3	.	0.2500	1.0000

Covariance Coefficients of Individuals			
Individual	Parent1	Parent2	Coefficient
1	1	1	1.5000
2	1	2	1.0000
3	2	3	1.0000

Number of Individuals	3

Output 66.1.3 Monoecious Population Analysis, Generation 3

Inbreeding within Nonoverlapping Generations

The INBREED Procedure

Generation = 3

Covariance Coefficients						
Individual	Parent1	Parent2	1	2	3	4
1	1	2	1.2500	0.5625	0.8750	0.5625
2	1	3	0.5625	1.0000	1.1349	0.6250
3	2	1	0.8750	1.1349	1.2500	1.1349
4	1	3	0.5625	0.6250	1.1349	1.0000

Covariance Coefficients of Individuals			
Individual	Parent1	Parent2	Coefficient
1	1	2	1.2500
2	1	3	1.0000
3	2	1	1.2500
4	1	3	1.0000

Number of Individuals	4

Note that, since the parents of the first generation are unknown, off-diagonal elements of the covariance matrix are all 0s and on-diagonal elements are all 1s. If there is an INIT=*cov* value, then the off-diagonal elements would be equal to *cov*, while on-diagonal elements would be equal to $1 + cov/2$.

In the third generation, individuals 2 and 4 are full siblings, so they belong to the same family. Since PROC INBREED computes covariance coefficients between families, the second and fourth columns of inbreeding coefficients are the same, except that their intersections with the second and fourth rows are reordered. Notice that, even though there is an observation to assign a covariance of 0.50 between individuals 2 and 3 in the third generation, the covariance between 2 and 3 is set to 1.135, the same value assigned between 4 and 3. This is because families get the same covariances, and later specifications override previous ones.

Example 66.2: Pedigree Analysis

In the following example, an inbreeding analysis is performed for a complicated pedigree. This analysis includes computing selective matings of some individuals and inbreeding coefficients of all individuals. Also, inbreeding coefficients are averaged within sex categories. The following statements produce Output 66.2.1:

```
data Swine;
   input Swine_Number $ Sire $ Dam $ Sex $;
   datalines;
3504 2200 2501  M
3514 2521 3112  F
3519 2521 2501  F
2501 2200 3112  M
2789 3504 3514  F
3501 2521 3514  M
3712 3504 3514  F
3121 2200 3501  F
;

title 'Least Related Matings';
proc inbreed data=Swine ind average;
   var Swine_Number Sire Dam;
   matings 2501 / 3501 3504 ,
           3712 / 3121;
   gender Sex;
run;
```

Note the following from Output 66.2.1:

- Observation 4, which defines Swine_Number=2501, should precede the first and third observations where the progeny for 2501 are given. PROC INBREED ignores observation 4 since it is given out of order. As a result, the parents of 2501 are missing or unknown.
- The first column in the "Inbreeding Averages" table corresponds to the averages taken over the on-diagonal elements of the inbreeding coefficients matrix, and the second column gives averages over the off-diagonal elements.

Output 66.2.1 Pedigree Analysis

Least Related Matings

The INBREED Procedure

Inbreeding Coefficients of Individuals

Swine_Number	Sire	Dam	Coefficient
2200			.
2501			.
3504	2200	2501	.
2521			.
3112			.
3514	2521	3112	.
3519	2521	2501	.
2789	3504	3514	.
3501	2521	3514	0.2500
3712	3504	3514	.
3121	2200	3501	.

Inbreeding Coefficients of Matings

Sire	Dam	Coefficient
2501	3501	.
2501	3504	0.2500
3712	3121	0.1563

Averages of Inbreeding Coefficient Matrix

	Inbreeding	Coancestry
Male X Male	0.0625	0.1042
Male X Female	.	0.1362
Female X Female	0.0000	0.1324
Over Sex	0.0227	0.1313

Number of Males	4
Number of Females	7
Number of Individuals	11

Example 66.3: Pedigree Analysis with BY Groups

This example demonstrates the structure of the OUTCOV= data set created by PROC INBREED. Note that the first BY group has three individuals, while the second has five. Therefore, the covariance matrix for the second BY group is broken up into two panels. The following statements produce Output 66.3.1.

```
data Swine;
   input Group Swine_Number $ Sire $ Dam $ Sex $;
   datalines;
1  2789 3504 3514  F
2  2501 2200 3112  .
```

```
2  3504 2501 3782  M
;

proc inbreed data=Swine covar noprint outcov=Covariance
             init=0.4;
   var Swine_Number Sire Dam;
   gender Sex;
   by Group;
run;


title 'Printout of OUTCOV= data set';
proc print data=Covariance;
   format Col1-Col3 4.2;
run;
```

Output 66.3.1 Pedigree Analysis with BY Groups

Printout of OUTCOV= data set

Obs	Group	Sex	_TYPE_	_PANEL_	_COL_	Swine_Number	Sire	Dam	COL1	COL2	COL3
1	1	M	COV	1	COL1	3504			1.20	0.40	0.80
2	1	F	COV	1	COL2	3514			0.40	1.20	0.80
3	1	F	COV	1	COL3	2789	3504	3514	0.80	0.80	1.20
4	2	M	COV	1	COL1	2200			1.20	0.40	0.80
5	2	F	COV	1	COL2	3112			0.40	1.20	0.80
6	2	M	COV	1	COL3	2501	2200	3112	0.80	0.80	1.20
7	2	F	COV	1		3782			0.40	0.40	0.40
8	2	M	COV	1		3504	2501	3782	0.60	0.60	0.80
9	2	M	COV	2		2200			0.40	0.60	.
10	2	F	COV	2		3112			0.40	0.60	.
11	2	M	COV	2		2501	2200	3112	0.40	0.80	.
12	2	F	COV	2	COL1	3782			1.20	0.80	.
13	2	M	COV	2	COL2	3504	2501	3782	0.80	1.20	.

References

Crow, J. F., and Kimura, M. (1970). *An Introduction to Population Genetics Theory*. New York: Harper & Row.

Falconer, D. S., and Mackay, T. F. C. (1996). *Introduction to Quantitative Genetics*. 4th ed. Harlow, UK: Longman.

Kempthorne, O. (1957). *An Introduction to Genetic Statistics*. New York: John Wiley & Sons.

Subject Index

coefficient
  of relationship (INBREED), 4952
covariance coefficients, *see* INBREED procedure

full sibs mating
  INBREED procedure, 4957

generation (INBREED)
  nonoverlapping, 4941, 4944
  number, 4949
  overlapping, 4941, 4943
  variable, 4949

INBREED procedure
  coancestry, computing, 4953
  coefficient of relationship, computing, 4952
  covariance coefficients, 4941, 4943, 4945, 4948, 4950, 4952
  covariance coefficients matrix, output, 4948
  first parent, 4950
  full sibs mating, 4957
  generation number, 4949
  generation variable, 4949
  generation, nonoverlapping, 4941, 4944
  generation, overlapping, 4941, 4943
  inbreeding coefficients, 4941, 4943, 4948, 4950, 4953
  inbreeding coefficients matrix, output, 4948
  individuals, outputting coefficients, 4948
  individuals, specifying, 4945, 4949, 4950
  initial covariance value, 4951
  initial covariance value, assigning, 4948
  initial covariance value, specifying, 4943
  kinship coefficient, 4952
  last generation's coefficients, output, 4948
  mating, offspring and parent, 4956, 4957
  mating, self, 4956
  matings, output, 4950
  monoecious population analysis, example, 4961
  offspring, 4948, 4955
  ordering observations, 4942
  OUTCOV= data set, 4948, 4958
  output table names, 4960
  panels, 4958, 4964
  pedigree analysis, 4941, 4942
  pedigree analysis, example, 4963, 4964
  population, monoecious, 4961
  population, multiparous, 4948, 4952
  population, nonoverlapping, 4949
  population, overlapping, 4942, 4943, 4954
  progeny, 4951, 4953, 4956, 4963
  second parent, 4950
  selective matings, output, 4950
  specifying gender, 4945
  theoretical correlation, 4952
  unknown or missing parents, 4958
  variables, unaddressed, 4950, 4951
initial covariance value
  assigning (INBREED), 4948
  INBREED procedure, 4951
  specifying (INBREED), 4943

mating
  offspring and parent (INBREED), 4956, 4957
  self (INBREED), 4956
monoecious population analysis
  example (INBREED), 4961

offspring
  INBREED procedure, 4948, 4955
ordering observations
  INBREED procedure, 4942
output data sets
  OUTCOV= data set (INBREED), 4948, 4958
output table names
  INBREED procedure, 4960

panels
  INBREED procedure, 4958, 4964
pedigree analysis
  example (INBREED), 4963, 4964
  INBREED procedure, 4941, 4942
population (INBREED)
  monoecious, 4961
  multiparous, 4948, 4952
  nonoverlapping, 4949
  overlapping, 4942, 4943, 4954
progeny
  INBREED procedure, 4951, 4953, 4956, 4963

theoretical correlation
  INBREED procedure, 4952

unknown or missing parents
  INBREED procedure, 4958

variables, unaddressed
  INBREED procedure, 4950, 4951

Syntax Index

AVERAGE option
    PROC INBREED statement, 4948

BY statement
    INBREED procedure, 4949

CLASS statement
    INBREED procedure, 4949
COVAR option
    PROC INBREED statement, 4948

DATA= option
    PROC INBREED statement, 4948

GENDER statement, INBREED procedure, 4949

INBREED procedure
    syntax, 4947
INBREED procedure, BY statement, 4949
INBREED procedure, CLASS statement, 4949
INBREED procedure, GENDER statement, 4949
INBREED procedure, MATINGS statement, 4950
INBREED procedure, PROC INBREED statement, 4947
    AVERAGE option, 4948
    COVAR option, 4948
    DATA= option, 4948
    IND option, 4948
    INDL option, 4948
    INIT= option, 4948
    MATRIX option, 4948
    MATRIXL option, 4948
    NOPRINT option, 4948
    OUTCOV= option, 4948
    SELFDIAG option, 4948
INBREED procedure, VAR statement, 4950
IND option
    PROC INBREED statement, 4948
INDL option
    PROC INBREED statement, 4948
INIT= option
    PROC INBREED statement, 4948

MATINGS statement, INBREED procedure, 4950
MATRIX option
    PROC INBREED statement, 4948
MATRIXL option
    PROC INBREED statement, 4948

NOPRINT option
    PROC INBREED statement, 4948

OUTCOV= option
    PROC INBREED statement, 4948

PROC INBREED statement, *see* INBREED procedure

SELFDIAG option
    PROC INBREED statement, 4948

VAR statement
    INBREED procedure, 4950