SAS/STAT® 9.22 User's Guide

The INBREED Procedure
(Book Excerpt)

Chapter 44
The INBREED Procedure

Contents

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 44.2).

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 3444 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 44.4).

Figure 44.1 Analysis for an Overlapping Population

```
                        The INBREED Procedure

                      Covariance Coefficients

 Individual   Parent1    Parent2      George       Lisa       Mark      Scott      Kelly

 George                               1.1250     0.2500     0.6875     0.2500     0.2500
 Lisa                                 0.2500     1.1250     0.6875     0.2500     0.6875
 Mark        George     Lisa         0.6875     0.6875     1.1250     0.2500     0.5000
 Scott                               0.2500     0.2500     0.2500     1.1250     0.6875
 Kelly       Scott      Lisa         0.2500     0.6875     0.5000     0.6875     1.1250
 Amy                                 0.2500     0.2500     0.2500     0.2500     0.2500
 Mike        George     Amy          0.6875     0.2500     0.4688     0.2500     0.2500
 David       Mark       Kelly        0.4688     0.6875     0.8125     0.4688     0.8125
 Jane                                0.2500     0.2500     0.2500     0.2500     0.2500
 Merle       Mike       Jane         0.4688     0.2500     0.3594     0.2500     0.2500
 Jim         Mark       Kelly        0.4688     0.6875     0.8125     0.4688     0.8125

                      Covariance Coefficients

 Individual   Parent1    Parent2        Amy       Mike      David       Jane      Merle

 George                               0.2500     0.6875     0.4688     0.2500     0.4688
 Lisa                                 0.2500     0.2500     0.6875     0.2500     0.2500
 Mark        George     Lisa         0.2500     0.4688     0.8125     0.2500     0.3594
 Scott                               0.2500     0.2500     0.4688     0.2500     0.2500
 Kelly       Scott      Lisa         0.2500     0.2500     0.8125     0.2500     0.2500
 Amy                                 1.1250     0.6875     0.2500     0.2500     0.4688
 Mike        George     Amy          0.6875     1.1250     0.3594     0.2500     0.6875
 David       Mark       Kelly        0.2500     0.3594     1.2500     0.2500     0.3047
 Jane                                0.2500     0.2500     0.2500     1.1250     0.6875
 Merle       Mike       Jane         0.4688     0.6875     0.3047     0.6875     1.1250
 Jim         Mark       Kelly        0.2500     0.3594     0.8125     0.2500     0.3047

                      Covariance Coefficients

          Individual   Parent1    Parent2         Jim

          George                               0.4688
          Lisa                                 0.6875
          Mark        George     Lisa         0.8125
          Scott                               0.4688
          Kelly       Scott      Lisa         0.8125
          Amy                                 0.2500
          Mike        George     Amy          0.3594
          David       Mark       Kelly        0.8125
          Jane                                0.2500
          Merle       Mike       Jane         0.3047
          Jim         Mark       Kelly        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 44.5).

Figure 44.2 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 44.3:

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

Figure 44.3 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 PROC INBREED:

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 options listed in Table 44.1 are available in the PROC INBREED statement.

Table 44.1 INBREED Procedure Options

Task	Option
Specify Data Sets	DATA=
	OUTCOV=
Control Type of Coefficient	COVAR
Control Displayed Tables	AVERAGE
	IND
	MATRIX
Specify Default Covariance Value	INIT=
Suppress Output	INDL
	MATRIXL
	NOPRINT

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.

BY Statement

     BY _variables_ **;**

You can specify a BY statement with PROC INBREED to obtain separate analyses on observations in groups 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 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 BY statement option NOTSORTED or DESCENDING in the BY statement for PROC INBREED. 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.

For more information about the BY statement, see _SAS Language Reference: Concepts_. For more information about the DATASETS procedure, see the _Base SAS 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 3443.

- The second parent is always assumed to be the female. See the section "VAR Statement" on page 3443.

- 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 3434.

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 3445.

Covariance Coefficient (or Coefficient of Relationship)

The covariance coefficient between individuals X and Y is defined by

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

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

$$F_X = f_{AB}$$

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):

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

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

$$F_Z = f_{XY}$$

Figure 44.4 Inbreeding Relationship for Nonoverlapping Population
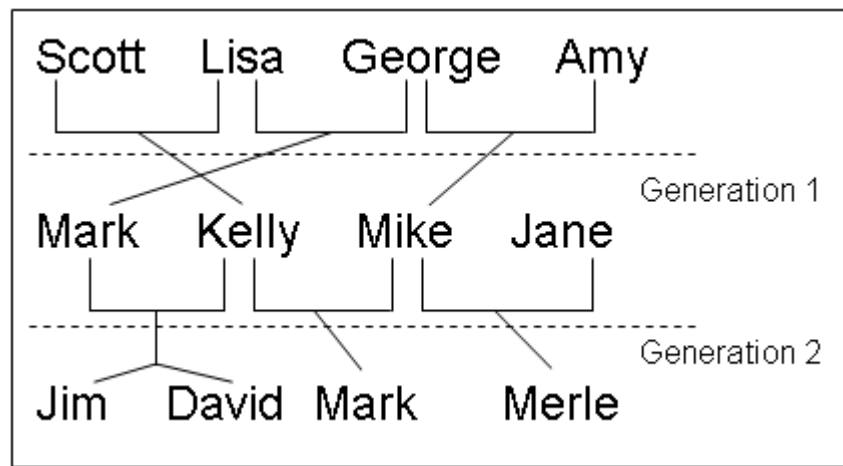
For example, in Figure 44.4, 'Jim' and 'Mark' from Generation 2 are progenies of 'Mark' and 'Kelly' and of 'Mike' and 'Kelly' from Generation 1, respectively. The coancestry between 'Jim' and 'Mark' is

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

From the covariance matrix for Generation=1 in Figure 44.4 and the relationship that coancestry is half of the covariance coefficient,

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

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

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

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 44.5, 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 44.5 Inbreeding Relationship for Overlapping Population
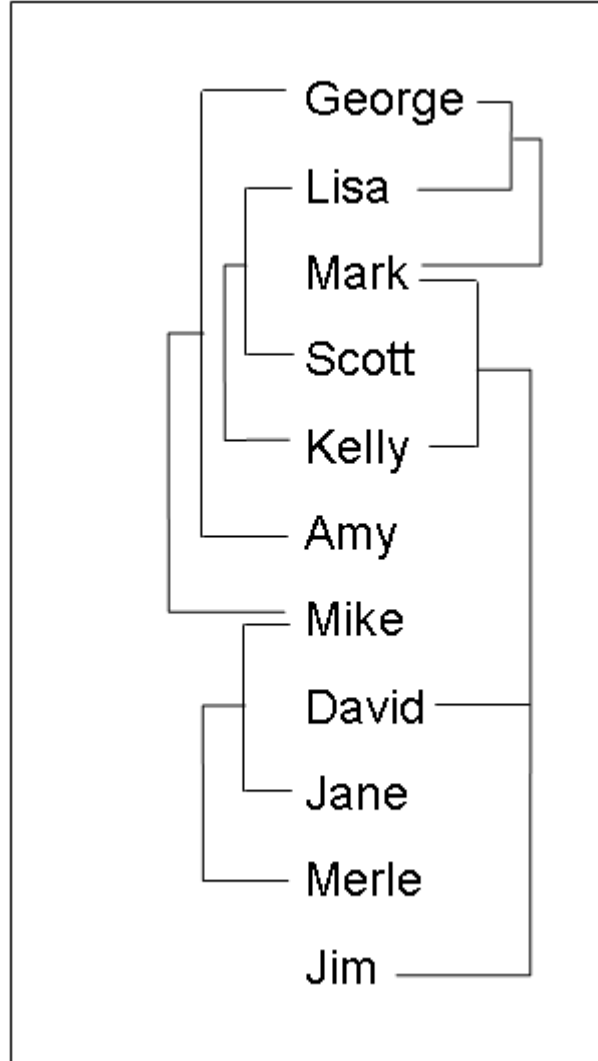
This is so because 'Kelly' is defined before 'David'; therefore, 'Kelly' is not younger than 'David', and the parents of 'David' are 'Mark' and 'Kelly'. The covariance coefficient values Cov(Kelly,Mark) and Cov(Kelly,Kelly) from the matrix in Figure 44.5 yield that the coancestry between 'Kelly' and '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 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 44.5 and also in the covariance matrix for GENDER=1 in Figure 44.4, 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 44.5, $f_{\text{Lisa,George}} = 0.25/2 = 0.125$, $f_{\text{Lisa,Lisa}} = 1.125/2 = 0.5625$, 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 3445. 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

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

Since the coancestry is half of the covariance coefficient, from the covariance matrix in Figure 44.5,

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

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

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

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 3445):

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

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

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

For a more detailed discussion on the calculation of coancestries, inbreeding coefficients, and covariance coefficients, refer to Falconer and Mackay (1996), Kempthorne (1957), and 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 'COLi'. 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 44.2. For more information on ODS, see Chapter 20, "Using the Output Delivery System."

Table 44.2 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 44.1: 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 44.1.1 through Output 44.1.3:

```
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 44.1.1 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 44.1.2 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 44.1.3 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 44.2: 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 44.2.1:

```
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 44.2.1:

- 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 44.2.1 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             .
```

Output 44.2.1 *continued*

```
                    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 44.3: 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 44.3.1.

```
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 44.3.1 Pedigree Analysis with BY Groups

```
                  Printout of OUTCOV= data set

                                    Swine_
   Obs   Group   Sex   _TYPE_   _PANEL_   _COL_   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 and Row.

Falconer, D. S. and Mackay, T. F. C. (1996), *Introduction to Quantitative Genetics*, Fourth Edition, London: Longman.

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

Subject Index

Syntax Index

Your Turn

We welcome your feedback.

- If you have comments about this book, please send them to `yourturn@sas.com`. Include the full title and page numbers (if applicable).

- If you have comments about the software, please send them to `suggest@sas.com`.

SAS® Publishing Delivers!

Whether you are new to the work force or an experienced professional, you need to distinguish yourself in this rapidly changing and competitive job market. SAS® Publishing provides you with a wide range of resources to help you set yourself apart. Visit us online at support.sas.com/bookstore.

SAS® Press

Need to learn the basics? Struggling with a programming problem? You'll find the expert answers that you need in example-rich books from SAS Press. Written by experienced SAS professionals from around the world, SAS Press books deliver real-world insights on a broad range of topics for all skill levels.

support.sas.com/saspress

SAS® Documentation

To successfully implement applications using SAS software, companies in every industry and on every continent all turn to the one source for accurate, timely, and reliable information: SAS documentation. We currently produce the following types of reference documentation to improve your work experience:

- Online help that is built into the software.
- Tutorials that are integrated into the product.
- Reference documentation delivered in HTML and PDF – **free** on the Web.
- Hard-copy books.

support.sas.com/publishing

SAS® Publishing News

Subscribe to SAS Publishing News to receive up-to-date information about all new SAS titles, author podcasts, and new Web site features via e-mail. Complete instructions on how to subscribe, as well as access to past issues, are available at our Web site.

support.sas.com/spn

§sas | THE POWER TO KNOW®