Chapter 11
The MIANALYZE Procedure

Chapter Table of Contents

Chapter 11

The MIANALYZE Procedure

Overview

The MIANALYZE procedure is a new, experimental procedure that combines the results of the analyses of imputations and generates valid statistical inferences. It is available in Release 8.1 of the SAS System. Multiple imputation provides a useful strategy for analyzing data sets with missing values. Instead of filling in a single value for each missing value, Rubin's (1976; 1987) multiple imputation strategy replaces each missing value with a set of plausible values that represent the uncertainty about the right value to impute. You can implement the strategy with two SAS procedures: PROC MI, which generates imputed data sets, and PROC MIANALYZE, which combines the results of analyses carried out on the data sets. These two procedures are available in experimental form in Release 8.1 of the SAS System.

These analyses are obtained by using standard SAS procedures (such as PROC REG) for complete data. No matter which complete-data analysis is used, the process of combining results from different imputed data sets is essentially the same. This results in valid statistical inferences that properly reflect the uncertainty due to missing values.

The MIANALYZE procedure reads the parameter estimates and associated covariance matrix which are computed by the standard statistical procedure for each imputed data set. The MIANALYZE procedure then derives valid univariate and multivariate inferences for these parameters.

For some parameters of interest, it is not straightforward to compute estimates and associated covariance matrices with standard statistical SAS procedures. Examples include correlation coefficients between two variables and ratios of variable means. Special cases such as these are described in the "Examples of the Complete-Data Inferences" section on page 14.

Getting Started

The Fitness data set has been altered to contain an arbitrary missing pattern:

```
*----------------- Data on Physical Fitness -----------------*
| These measurements were made on men involved in a physical |
| fitness course at N.C. State University.                   |
| Only selected variables of                                 |
| Oxygen (oxygen intake, ml per kg body weight per minute),  |
| Runtime (time to run 1.5 miles in minutes), and            |
| RunPulse (heart rate while running) are used.              |
| Certain values were changed to missing for the analysis    |
*------------------------------------------------------------*;
```

```
data FitMiss;
   input Oxygen RunTime RunPulse @@;
   datalines;
44.609  11.37  178     45.313  10.07  185
54.297   8.65  156     59.571    .      .
49.874   9.22    .     44.811  11.63  176
  .     11.95  176     49.091  10.85    .
39.442  13.08  174     60.055   8.63  170
50.541    .      .     37.388  14.03  186
44.754  11.12  176     47.273    .      .
51.855  10.33  166     49.156   8.95  180
40.836  10.95  168     46.672  10.00    .
  .     10.25    .     50.388  10.08  168
39.407  12.63  174     46.080  11.17  156
45.441   9.63  164       .      8.92  146
45.118  11.08    .     39.203  12.88  168
45.790  10.47  186     50.545   9.93  148
48.673   9.40  186     47.920  11.50  170
47.467  10.50  170
;
```

Assume that the data are multivariate normally distributed and that the missing data are missing at random (see the “Statistical Assumptions for Using MI” section in “The MI Procedure” chapter for a description of these assumptions). The following statements use the MI procedure to impute missing values for the FitMiss data set.

```
proc mi data=FitMiss noprint out=outmi seed=37851;
   var Oxygen RunTime RunPulse;
run;
```

The MI procedure creates imputed data sets, which are stored in the outmi data set. A variable named $\_Imputation\_$ indicates the imputation numbers. Based on m imputations, m different sets of the point and variance estimates for a parameter can be computed. In this example, $m = 5$ is the default.

The following statements generate regression coefficients for each of the five imputed data sets:

```
proc reg data=miout outest=outreg covout noprint;
   model Oxygen= RunTime RunPulse;
   by _Imputation_;
run;

proc print data=outreg(obs=8);
   var _Imputation_ _Type_ _Name_
      Intercept RunTime RunPulse;
   title 'Parameter Estimates from Imputed Data Sets';
run;
```

Output 11.0.0. Parameter Estimates

Parameter Estimates from Imputed Data Sets

Obs	_Imputation_	_TYPE_	_NAME_	Intercept	RunTime	RunPulse
1	1	PARMS		89.8363	-3.09466	-0.05644
2	1	COV	Intercept	61.5309	-0.61390	-0.31934
3	1	COV	RunTime	-0.6139	0.13780	-0.00494
4	1	COV	RunPulse	-0.3193	-0.00494	0.00217
5	2	PARMS		96.8952	-3.11704	-0.09877
6	2	COV	Intercept	56.3022	-0.27827	-0.31399
7	2	COV	RunTime	-0.2783	0.15900	-0.00826
8	2	COV	RunPulse	-0.3140	-0.00826	0.00237

The following statements combine the five sets of regression coefficients:

```
proc mianalyze data=outreg mult edf=28;
   var Intercept RunTime RunPulse;
run;
```

Output 11.0.1. Multiple-Imputation Variance Information

The MIANALYZE Procedure

Model Information

Data Set	WORK.OUTREG
Number of Imputations	5

Multiple-Imputation Variance Information

Parameter	Variance: Between	Variance: Within	Variance: Total	DF
Intercept	26.871639	57.685471	89.931438	31.112
RunTime	0.015027	0.128275	0.146308	263.31
RunPulse	0.000906	0.002220	0.003308	36.995

Multiple-Imputation Variance Information

Parameter	Relative Increase in Variance	Fraction Missing Information
Intercept	0.558996	0.396169
RunTime	0.140580	0.129837
RunPulse	0.489914	0.362383

The “Model Information” table lists the input data set(s) and the number of imputations. The “Multiple-Imputation Variance Information” table displays the between-imputation, within-imputation, and total variances for combining complete-data inferences. It also displays the degrees of freedom for the total variance, the relative increase in variance due to missing values, and the fraction of missing information for each parameter estimate.

Output 11.0.2. Multiple-Imputation Parameter Estimates

The MIANALYZE Procedure

Multiple-Imputation Parameter Estimates

Parameter	Estimate	Standard Error	95% Confidence Limits		DF
Intercept	93.619640	9.483219	74.28132	112.9580	31.112
RunTime	-2.997349	0.382503	-3.75050	-2.2442	263.31
RunPulse	-0.086246	0.057515	-0.20278	0.0303	36.995

Multiple-Imputation Parameter Estimates

Parameter	Minimum	Maximum	Mu0	t for H0: Parameter=Mu0	Pr > |t|
Intercept	88.272767	100.831579	0	9.87	<.0001
RunTime	-3.117041	-2.808650	0	-7.84	<.0001
RunPulse	-0.127565	-0.056435	0	-1.50	0.1422

The “Multiple-Imputation Parameter Estimates” table displays a combined estimate and standard error for each regression coefficient (parameter). Inferences are based on *t* distributions. The table displays a 95% confidence interval and a *t*-test with the associated *p*-value for the hypothesis that the parameter is equal to the value specified with the MU0= option (in this case, zero by default). The minimum and maximum parameter estimates from the imputed data sets are also displayed.

Syntax

The following statements are available in PROC MIANALYZE.

PROC MIANALYZE < *options* > **;**

BY *variables* **;**
VAR *variables* **;**

The rest of this section gives detailed syntax information for each of these statements. The PROC MIANALYZE statement and the VAR statement are required for the MIANALYZE procedure.

PROC MIANALYZE < *options* > **;**

The following table summarizes the options in the PROC MIANALYZE statement.

Table 11.1. Summary of PROC MIANALYZE Options

Tasks	Options
Specify input data sets	
a COV, CORR, or EST type data set	DATA=
parameter estimates and covariance matrices	PARMS=, COVB=
parameter estimates and $(X'X)^{-1}$ matrices	PARMS=, XPXI=
Specify statistical analysis	
parameters under the null hypothesis	MU0=
level for the confidence interval	ALPHA=
complete-data degrees of freedom	EDF=
multivariate inferences	MULT

The following are explanations of the options that can be used in the PROC MIANALYZE statement (in alphabetic order):

ALPHA=*p*

specifies that confidence limits are to be constructed for the parameter estimates with confidence level $100(1-p)\%$, where $0 < p < 1$. The default is p=0.05.

COVB=*SAS-data-set*

names an input SAS data set that contains covariance matrices of the parameter estimates from imputed data sets. If you provide a COVB= data set, you must also provide a PARMS= data set.

DATA=*SAS-data-set*

names a specially structured input SAS data set that contains estimates from imputed data sets. This data set must have a TYPE of EST, COV, or CORR:

- if TYPE=EST, the data set contains the parameter estimates and associated covariance matrices.
- if TYPE=COV, the data set contains the sample means, sample sizes, and covariance matrices. Each covariance matrix for variables is divided by the sample size n to create the covariance matrix for parameter estimates.
- if TYPE=CORR, the data set contains the sample means, sample sizes, standard errors, and correlation matrices. The covariance matrices are computed from the correlation matrices and associated standard errors. Each covariance matrix for variables is divided by the sample size n to create the covariance matrix for parameter estimates.

If you do not specify an input data set with the DATA=, COVB=, or XPXI= option, then the most recently created SAS data set is used as an input DATA= data set.

EDF=*number*
specifies the complete-data degrees of freedom for the parameter estimates. This is used to compute an adjusted degrees of freedom for each parameter estimate. By default, EDF=∞ and the degrees of freedom for each parameter estimate is not adjusted.

MU0=*numbers*
LOCATION=*numbers*
specifies the parameter values μ_0 under the null hypothesis $\mu = \mu_0$ in the *t* tests for location for the variables. If only one number μ_0 is specified, that number is used for all variables. If more than one number is specified, the specified numbers correspond to variables in the VAR statement in the order in which they appear in the VAR statement.

MULT
MULTIVARIATE
requests multivariate inference for the variables.

PARMS=*SAS-data-set*
names an input SAS data set that contains parameter estimates computed from imputed data sets. If you provide a PARMS= data set, you must also provide a COVB= or XPXI= data set.

XPXI=*SAS-data-set*
names an input SAS data set that contains the $(X'X)^{-1}$ matrices associated with the parameter estimates computed from imputed data sets. If you provide an XPXI= data set, you must also provide a PARMS= data set. In this case, PROC MIANALYZE reads the standard errors of the estimates from the PARMS= data. The standard errors and $(X'X)^{-1}$ matrices are used to derive the covariance matrices.

BY Statement

BY *variables* **;**

You can specify a BY statement with PROC MIANALYZE 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 using the SORT procedure with a similar BY statement.
- Specify the BY statement option NOTSORTED or DESCENDING in the BY statement for the MI 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 using the DATASETS procedure.

For more information on the BY statement, refer to the discussion in *SAS Language Reference: Concepts*. For more information on the DATASETS procedure, refer to the discussion in the *SAS Procedures Guide*.

VAR Statement

VAR *variables* **;**

The VAR statement lists the variables to be analyzed. The statement is required, and the variables must be numeric.

Details

Input Data Sets

You can specify input data sets using one of the following option combinations:

- DATA=, which provides both parameter estimates and the associated covariance matrix in a single input data set.
- PARMS= and COVB=, which provide parameter estimates and the associated covariance matrix in separate data sets, respectively.
- PARMS= and XPXI=, which provide parameter estimates and the associated standard errors in a PARMS= data set and the associated $(X'X)^{-1}$ matrix in an XPXI= data set.

The appropriate combination depends on the SAS procedure you used to create the parameter estimates and associated covariance matrix. For instance, if you used PROC REG to create an OUTEST= data set containing the parameter estimates and covariance matrix, you would use the DATA= option to read the OUTEST= data set. Each input data set contains the variable $\_Imputation\_$ to identify the imputation by number.

If you do not specify an input data set with the DATA=, COVB=, or XPXI= option, then the most recently created SAS data set is used as an input DATA= data set.

DATA= data set

The input DATA= data set is a specially structured SAS data set created by statistical procedures available with SAS software. The data set must have a TYPE of EST, COV, or CORR.

With TYPE=EST, the MIANALYZE procedure reads parameter estimates from observations for which the variable _TYPE_ is equal to ’PARM’ or ’PARMS’, and covariance matrices for parameter estimates from observations for which the variable _TYPE_ is equal to ’COV’ or ’COVB’.

With TYPE=COV, the procedure reads sample means from observations for which the variable _TYPE_ is equal to 'MEAN', sample size n from observations for which the variable _TYPE_ is equal to 'N', and covariance matrices for variables from observations for which the variable _TYPE_ is equal to 'COV'.

With TYPE=CORR, the procedure reads sample means from observations for which the variable _TYPE_ is equal to 'MEAN', sample size n from observations for which the variable _TYPE_ is equal to 'N', correlation matrices for variables from observations for which the variable _TYPE_ is equal to 'CORR', and standard errors for variables from observations or which the variable _TYPE_ is equal to ='STD'. The standard errors and correlation matrix are used to generate a covariance matrix for the variables.

Note that with TYPE=COV or CORR, each covariance matrix for the variables is divided by n to create the covariance matrix for the sample means.

PARMS= and COVB= data sets

The input PARMS= data set contains parameter estimates, and the input COVB= data set contains associated covariance matrices computed from imputed data sets. Such data sets are typically created with an ODS OUTPUT statement using procedures such as PROC MIXED and PROC GENMOD.

The MIANALYZE procedure uses a PARMS= data set to read parameter names from the variable $Parameter$ or $Effect$, and it reads parameter estimates from the variable $Estimate$.

The MIANALYZE procedure uses a COVB= data set to read parameter names from the variable $Parameter$, $Effect$, or $RowName$, and it reads covariance matrices from the subsequent variables $Col1$, $Col2$, ... or $Prm1$, $Prm2$, ... in the data set.

PARMS= and XPXI= data sets

The input PARMS= data set contains parameter estimates, and the input XPXI= data set contains associated $(X'X)^{-1}$ matrices computed from imputed data sets. Such data sets are typically created with an ODS OUTPUT statement using a procedure such as PROC GLM.

The MIANALYZE procedure uses a PARMS= data set to read parameter names from the variable $Parameter$, it reads parameter estimates from the variable $Estimate$, and it reads standard errors for parameter estimates from the variable $StdErr$.

The MIANALYZE procedure uses an XPXI= data set to read parameter names from the variable $Parameter$ and it reads $(X'X)^{-1}$ matrices from the subsequent variables in the data set.

Combining Inferences from Imputed Data Sets

With m imputations, m different sets of the point and variance estimates for a parameter Q can be computed. Let $\hat{Q}_i$ and $\hat{U}_i$ be the point and variance estimates from the i-th imputed data set, i=1, 2, ..., m. Then the combined point estimate for Q from multiple imputation is the average of the m complete-data estimates:

$$\overline{Q} = \frac{1}{m} \sum_{i=1}^{m} \hat{Q}_i$$

Let $\overline{U}$ be the within-imputation variance, which is the average of the m complete-data estimates:

$$\overline{U} = \frac{1}{m} \sum_{i=1}^{m} \hat{U}_i$$

and B be the between-imputation variance

$$B = \frac{1}{m-1} \sum_{i=1}^{m} (\hat{Q}_i - \overline{Q})^2$$

Then the variance estimate associated with $\overline{Q}$ is the total variance (Rubin 1987)

$$T = \overline{U} + (1 + \frac{1}{m}) B$$

The statistic $(Q - \overline{Q})T^{-(1/2)}$ is approximately distributed as t with v_m degrees of freedom (Rubin 1987), where

$$v_m = (m-1)[1 + \frac{\overline{U}}{(1+m^{-1})B}]^2$$

When the complete-data degrees of freedom v_0 is small, and there is only a modest proportion of missing data, the computed degrees of freedom, v_m, can be much larger than v_0, which is inappropriate. Barnard and Rubin (1999) recommend the use of an adjusted degrees of freedom

$$v_m^* = \left[\frac{1}{v_m} + \frac{1}{\hat{v}_{obs}} \right]^{-1}$$

where $\hat{v}_{obs} = (1-\gamma)\, v_0 (v_0+1)/(v_0+3)$ and $\gamma = (1+m^{-1})B/T$.

If you specify the complete-data degrees of freedom v_0 with the EDF= option, the MIANALYZE procedure uses the adjusted degrees of freedom, v_m^*, in the inferences. Otherwise, the degrees of freedom v_m is used.

The degrees of freedom v_m depends on m and the ratio

$$r = \frac{(1+m^{-1})B}{\overline{U}}$$

The ratio r is called the relative increase in variance due to nonresponse (Rubin 1987). When there is no missing information about Q, the values of r and B are both zero. With a large value of m or a small value of r, the degrees of freedom v will be large and the distribution of $(Q - \overline{Q})T^{-(1/2)}$ will be approximately normal.

Another useful statistic is the fraction of missing information about Q:

$$\hat{\lambda} = \frac{r + 2/(v+3)}{r+1}$$

Both statistics r and λ are helpful diagnostics for assessing how the missing data contribute to the uncertainty about Q.

Multiple Imputation Efficiency

The relative efficiency (RE) of using the finite m imputation estimator, rather than using an infinite number for the fully efficient imputation, in units of variance, is approximately a function of m and λ (Rubin 1987, p. 114).

$$RE = (1 + \frac{\lambda}{m})^{-1}$$

The following table shows relative efficiencies with different values of m and λ. For cases with little missing information, only a small number of imputations are necessary.

Table 11.2. Relative Efficiency

	λ				
m	**10%**	**20%**	**30%**	**50%**	**70%**
3	0.9677	0.9375	0.9091	0.8571	0.8108
5	0.9804	0.9615	0.9434	0.9091	0.8772
10	0.9901	0.9804	0.9709	0.9524	0.9346
20	0.9950	0.9901	0.9852	0.9756	0.9662

Multivariate Inferences

Multivariate inference based on Wald tests can be done with m imputed data sets. The approach is a generalization of the approach taken in the univariate case (Rubin 1987, p. 137; Schafer 1997, p. 113). Suppose that $\hat{\mathbf{Q}}_i$ and $\hat{\mathbf{U}}_i$ are the point and covariance matrix estimates for a vector-valued parameter $\mathbf{Q}$ (such as a multivariate mean) from the ith imputed data set, i=1, 2, ..., m. Then the combined point estimate for $\mathbf{Q}$ from the multiple imputation is the average of the m complete-data estimates:

$$\overline{\mathbf{Q}} = \frac{1}{m} \sum_{i=1}^{m} \hat{\mathbf{Q}}_i$$

Suppose that $\overline{\mathbf{U}}$ is the within-imputation covariance matrix, which is the average of the m complete-data estimates

$$\overline{\mathbf{U}} = \frac{1}{m} \sum_{i=1}^{m} \hat{\mathbf{U}}_i$$

and suppose that $\mathbf{B}$ is the between-imputation covariance matrix

$$\mathbf{B} = \frac{1}{m-1} \sum_{i=1}^{m} (\hat{\mathbf{Q}}_i - \overline{\mathbf{Q}})(\hat{\mathbf{Q}}_i - \overline{\mathbf{Q}})'$$

Then the covariance matrix associated with $\overline{\mathbf{Q}}$ is the total covariance matrix

$$\mathbf{T}_0 = \overline{\mathbf{U}} + (1 + \frac{1}{m})\mathbf{B}$$

The natural multivariate extension of the t statistic used in the univariate case is the F statistic

$$F_0 = (\mathbf{Q} - \overline{\mathbf{Q}})' \mathbf{T}_0^{-1} (\mathbf{Q} - \overline{\mathbf{Q}})$$

with degrees of freedom p and

$$v = (m-1)(1 + 1/r)^2$$

where

$$r = (1 + \frac{1}{m}) trace(\mathbf{B}\overline{\mathbf{U}}^{-1})/p$$

is an average relative increase in variance due to nonresponse (Rubin 1987, p. 137; Schafer 1997, p. 114).

However, the reference distribution of the statistic F_0 is not easily derived. Especially for small m, the between-imputation covariance matrix $\mathbf{B}$ is unstable and does not have full rank for $m \leq p$ (Schafer 1997, p. 113).

One solution is to make an additional assumption that the population between-imputation and within-imputation covariance matrices are proportional to each other (Schafer 1997, p. 113). This assumption implies that the fractions of missing information for all components of $\mathbf{Q}$ are equal. Under this assumption, a more stable estimate of the total covariance matrix is

$$\mathbf{T} = (1 + r)\overline{\mathbf{U}}$$

With the total covariance matrix $\mathbf{T}$, the F statistic (Rubin 1987, p. 137)

$$F = (\mathbf{Q} - \overline{\mathbf{Q}})' \mathbf{T}^{-1} (\mathbf{Q} - \overline{\mathbf{Q}})/p$$

has an F distribution with degrees of freedom p and v_1, where

$$v_1 = \frac{1}{2}(p+1)(m-1)(1 + \frac{1}{r})^2$$

For $t = p(m-1) \leq 4$, PROC MIANALYZE uses the degrees of freedom v_1 in the analysis. For $t = p(m-1) > 4$, PROC MIANALYZE uses v_2, a better approximation of the degrees of freedom given by Li, Raghunathan, and Rubin (1991)

$$v_2 = 4 + (t-4)(1 + \frac{1}{r}(1 - \frac{2}{t}))^2$$

Examples of the Complete-Data Inferences

For a given parameter of interest, it is not always possible to compute the estimate and associated covariance matrix directly from a SAS procedure. This section gives examples of parameters with their estimates and associated covariance matrices, which provide the input to the MIANALYZE procedure. Some are straightforward, and others require special techniques.

Means

For a population mean vector μ, the usual estimate is the sample mean vector

$$\overline{\mathbf{y}} = \frac{1}{n} \sum \mathbf{y}_i$$

A variance estimate for $\overline{\mathbf{y}}$ is $\frac{1}{n}\mathbf{S}$, where $\mathbf{S}$ is the sample covariance matrix

$$\mathbf{S} = \frac{1}{n-1} \sum (\mathbf{y}_i - \overline{\mathbf{y}})(\mathbf{y}_i - \overline{\mathbf{y}})'$$

These statistics can be computed from a procedure such as CORR. This approach is illustrated in Example 1.

Regression Coefficients

Many SAS procedures are available for regression analysis. Among them, REG provides the most general analysis capabilities, and others like LOGISTIC and MIXED provide more specialized analyses.

Some regression procedures, such as REG and LOGISTIC, create an EST type data set that contains both the parameter estimates for the regression coefficients and their associated covariance matrix. You can read an EST type data set in the MIANALYZE procedure with the DATA= option. This approach is illustrated in Example 2.

Other procedures, such as GLM, MIXED, and GENMOD, do not generate EST type data sets for regression coefficients. For MIXED and GENMOD, you can use ODS OUTPUT statement to save parameter estimates in a data set and the associated covariance matrix in a separate data set. These data sets are then read in the MIANALYZE procedure with the PARMS= and COVB= options, respectively. This approach is illustrated in Example 3 for PROC MIXED and in Example 4 for PROC GENMOD.

PROC GLM does not display tables for covariance matrices. However, you can use ODS OUTPUT statement to save parameter estimates and associated standard errors in a data set and the associated $(X'X)^{-1}$ matrix in a separate data set. These data sets are then read in the MIANALYZE procedure with the PARMS= and XPXI= options, respectively. This approach is illustrated in Example 5.

Correlation Coefficients

For the population correlation coefficient ρ, a point estimate is the sample correlation coefficient r. However, for nonzero ρ, the distribution of r is skewed.

The distribution of r can be normalized through Fisher's Z transformation

$$z(r) = \frac{1}{2} \, log \left(\frac{1+r}{1-r} \right)$$

$z(r)$ is approximately normally distributed with mean $z(\rho)$ and variance $1/(n-3)$.

With a point estimate $\hat{z}$ and an approximate 95% confidence interval (z_1, z_2) for $z(\rho)$, a point estimate $\hat{r}$ and a 95% confidence interval (r_1, r_2) for ρ can be obtained by applying the inverse transformation

$$r = \frac{e^{2z} - 1}{e^{2z} + 1}$$

to $z = \hat{z}, z_1$, and z_2.

This approach is illustrated in Example 6.

Ratios of Variable Means

For the ratio μ_1/μ_2 of means for variables Y_1 and Y_2, the point estimate is $\overline{y}_1/\overline{y}_2$, the ratio of the sample means. The Taylor expansion and delta method can be applied to the function y_1/y_2 to obtain the variance estimate (Schafer 1997, p. 196)

$$\frac{1}{n}\left[\left(\frac{\overline{y}_1}{\overline{y}_2^2}\right)^2 s_{22} - 2\left(\frac{\overline{y}_1}{y_2^2}\right)\left(\frac{1}{\overline{y}_2}\right)s_{12} + \left(\frac{1}{\overline{y}_2}\right)^2 s_{11}\right]$$

where s_{11} and s_{22} are the sample variances of Y_1 and Y_2, respectively, and s_{12} is the sample covariance between Y_1 and Y_2.

A ratio of sample means will be approximately unbiased and normally distributed if the coefficient of variation of the denominator (the standard error for the mean divided by the estimated mean) is 10% or less (Cochran 1977, p. 166; Schafer 1997, p. 196). This approach is illustrated in Example 7.

ODS Table Names

PROC MIANALYZE assigns a name to each table it creates. You must use these names to reference tables when using the Output Delivery System (ODS). These names are listed in the following table. For more information on ODS, see Chapter 16, “Using the Output Delivery System.”

Table 11.3. ODS Tables Produced in PROC MIANALYZE

ODS Table Name	Description	Option
ModelInfo	Model information	
VarianceInfo	Variance information	
ParmEst	Parameter estimates	
WithinCov	Between-imputation covariance matrix	MULT
BetweenCov	Within-imputation covariance matrix	MULT
TotalCov	Total covariance matrix	MULT
MultInf	Multivariate inference	MULT

Examples

The following statements generate five imputed data sets to be used in this section. The data set FitMiss was created in the section “Getting Started”. See “The MI Procedure” chapter for details concerning the MI procedure.

```
proc mi data=FitMiss seed=37851 noprint out=miout;
   var Oxygen RunTime RunPulse;
run;
```

Example 1. Reading Means and Covariance Matrices from a DATA= COV Data Set

This example creates a COV type data set that contains sample means and covariance matrices computed from imputed data sets. These estimates are then combined to generate valid statistical inferences about the population means.

The following statements use the CORR procedure to generate sample means and a covariance matrix for the variables in each imputed data set.

```
proc corr data=outmi cov out=outcov(type=cov) nocorr noprint;
   var Oxygen RunTime RunPulse;
   by _Imputation_;
run;

proc print data=outcov(obs=12);
   title 'CORR Means and Covariance Matrices'
         ' (First Two Imputations)';
run;
```

Output 11.0.3. COV Data Set

CORR Means and Covariance Matrices (First Two Imputations)

Obs	_Imputation_	_TYPE_	_NAME_	Oxygen	RunTime	RunPulse
1	1	COV	Oxygen	28.3399	-6.4698	-21.358
2	1	COV	RunTime	-6.4698	2.0072	4.573
3	1	COV	RunPulse	-21.3575	4.5734	127.656
4	1	MEAN		47.3512	10.6004	171.533
5	1	STD		5.3235	1.4168	11.299
6	1	N		31.0000	31.0000	31.000
7	2	COV	Oxygen	32.3477	-6.9273	-34.993
8	2	COV	RunTime	-6.9273	2.0015	6.972
9	2	COV	RunPulse	-34.9932	6.9721	134.265
10	2	MEAN		47.3489	10.5350	169.172
11	2	STD		5.6875	1.4147	11.587
12	2	N		31.0000	31.0000	31.000

Note that the covariance matrices in the data set outcov are estimated covariance matrices of variables, $V(\mathbf{y})$. The estimated covariance matrix of the sample means is $V(\overline{\mathbf{y}}) = V(\mathbf{y})/n$, where n is the sample size, and is not the same as an estimated covariance matrix for variables.

The following statements combine the results for the imputed data sets, and derive both univariate and multivariate inferences about the means. The EDF= option is specified to request that the adjusted degrees of freedom be used in the analysis. For sample means based on 31 observations, the complete-data error degrees of freedom is 30.

```
proc mianalyze data=outcov edf=30 mult;
   var Oxygen RunTime RunPulse;
run;
```

Output 11.0.4. Multiple-Imputation Variance Information

Multiple-Imputation Variance Information

Parameter	Variance: Between	Within	Total	DF
Oxygen	0.010781	0.971786	0.984724	27.778
RunTime	0.001358	0.069812	0.071443	27.44
RunPulse	1.212934	3.978617	5.434138	15.06

Multiple-Imputation Variance Information

Parameter	Relative Increase in Variance	Fraction Missing Information
Oxygen	0.013313	0.013223
RunTime	0.023351	0.023072
RunPulse	0.365836	0.292770

The "Multiple-Imputation Variance Information" table displays the between-imputation variance, within-imputation variance, and total variance for each univariate inference. It also displays the degrees of freedom for the total variance. The relative increase in variance due to missing values and the fraction of missing information for each variable are also displayed. A detailed description of these statistics is provided in the "Combining Inferences from Imputed Data Sets" section on page 11 and the "Multiple Imputation Efficiency" section on page 12.

Output 11.0.5. Multiple-Imputation Parameter Estimates

Multiple-Imputation Parameter Estimates

Parameter	Estimate	Standard Error	95% Confidence Limits		DF
Oxygen	47.353472	0.992332	45.3200	49.3869	27.778
RunTime	10.540511	0.267287	9.9925	11.0885	27.44
RunPulse	170.107976	2.331124	165.1410	175.0749	15.06

Multiple-Imputation Parameter Estimates

Parameter	Minimum	Maximum	Mu0	t for H0: Parameter=Mu0	Pr > |t|
Oxygen	47.254402	47.524060	0	47.72	<.0001
RunTime	10.509680	10.600357	0	39.44	<.0001
RunPulse	169.171896	171.533410	0	72.97	<.0001

The "Multiple-Imputation Parameter Estimates" table displays the estimated mean and corresponding standard error for each variable. The table also displays a 95% confidence interval for the mean and a t statistic with the associated p-value for testing the hypothesis that the mean is equal to the value specified. You can use the MU0= option to specify the value for the null hypothesis, which is zero by default. The table also displays the minimum and maximum parameter estimates from the imputed data sets.

With the MULT option, the procedure displays the between-imputation covariance matrix, within-imputation covariance matrix, and total covariance matrix for multivariate inference.

Output 11.0.6. Within-Imputation Covariance Matrices

Within-Imputation Covariance Matrix

	Oxygen	RunTime	RunPulse
Oxygen	0.971786448	-0.224990242	-0.888353428
RunTime	-0.224990242	0.069812412	0.181765875
RunPulse	-0.888353428	0.181765875	3.978617483

Between-Imputation Covariance Matrix

	Oxygen	RunTime	RunPulse
Oxygen	0.010781081	-0.000901842	0.068065599
RunTime	-0.000901842	0.001358484	0.021778312
RunPulse	0.068065599	0.021778312	1.212934081

Total Covariance Matrix

	Oxygen	RunTime	RunPulse
Oxygen	1.189781877	-0.275461048	-1.087632794
RunTime	-0.275461048	0.085473041	0.222540400
RunPulse	-1.087632794	0.222540400	4.871118534

Assuming that the between-imputation covariance matrix is proportional to the within-imputation covariance matrix, the procedure also displays a multivariate inference for all the parameters taken jointly.

Output 11.0.7. Multiple-Imputation Multivariate Inference

Multiple-Imputation Multivariate Inference
Assuming Proportionality of Between/Within Covariance Matrices

Avg Relative Increase in Variance	Num DF	Den DF	F for H0: Parameter=Mu0	Pr > F
0.224324	3	181.84	12836.4	<.0001

Assuming that the within-imputation covariance matrix is proportional to the between-imputation covariance matrix, the table shows a significant p-value for the null hypothesis that the population means are all equal to zero.

With the exception of the multivariate inference, the above results could also have been obtained with the MI procedure.

Example 2. Reading Regression Results from a DATA= EST Data Set

This example creates an EST type data set that contains regression coefficients and their corresponding covariance matrices computed from imputed data sets. These estimates are then combined to generate valid statistical inferences about the regression model.

The following statements use the REG procedure to generate regression coefficients:

```
proc reg data=miout outest=outreg covout noprint;
   model Oxygen= RunTime RunPulse;
   by _Imputation_;
run;

proc print data=outreg(obs=8);
   var _Imputation_ _Type_ _Name_
      Intercept RunTime RunPulse;
   title 'REG Model Coefficients and Covariance matrices'
         ' (First Two Imputations)';
run;
```

Output 11.0.8. EST Type Data Set

REG Model Coefficients and Covariance matrices (First Two Imputations)

Obs	_Imputation_	_TYPE_	_NAME_	Intercept	RunTime	RunPulse
1	1	PARMS		89.8363	-3.09466	-0.05644
2	1	COV	Intercept	61.5309	-0.61390	-0.31934
3	1	COV	RunTime	-0.6139	0.13780	-0.00494
4	1	COV	RunPulse	-0.3193	-0.00494	0.00217
5	2	PARMS		96.8952	-3.11704	-0.09877
6	2	COV	Intercept	56.3022	-0.27827	-0.31399
7	2	COV	RunTime	-0.2783	0.15900	-0.00826
8	2	COV	RunPulse	-0.3140	-0.00826	0.00237

The following statements combine the results for the imputed data sets. The EDF= option is specified to request that the adjusted degrees of freedom be used in the analysis. For a regression model with three independent variables (including the intercept) and 31 observations, the complete-data error degrees of freedom is 28.

```
proc mianalyze data=outreg edf=28;
   var Intercept RunTime RunPulse;
run;
```

Output 11.0.9. Multiple-Imputation Variance Information

Multiple-Imputation Variance Information

Parameter	Variance Between	Variance Within	Variance Total	DF
Intercept	26.871639	57.685471	89.931438	10.91
RunTime	0.015027	0.128275	0.146308	21.123
RunPulse	0.000906	0.002220	0.003308	11.917

Multiple-Imputation Variance Information

Parameter	Relative Increase in Variance	Fraction Missing Information
Intercept	0.558996	0.396169
RunTime	0.140580	0.129837
RunPulse	0.489914	0.362383

The "Multiple-Imputation Variance Information" table displays the between-imputation, within-imputation, and total variances for combining complete-data inferences.

Output 11.0.10. Multiple-Imputation Parameter Estimates

Multiple-Imputation Parameter Estimates

Parameter	Estimate	Standard Error	95% Confidence Limits		DF
Intercept	93.619640	9.483219	72.72616	114.5131	10.91
RunTime	-2.997349	0.382503	-3.79253	-2.2022	21.123
RunPulse	-0.086246	0.057515	-0.21166	0.0392	11.917

Multiple-Imputation Parameter Estimates

Parameter	Minimum	Maximum	Mu0	t for H0: Parameter=Mu0	Pr > |t|
Intercept	88.272767	100.831579	0	9.87	<.0001
RunTime	-3.117041	-2.808650	0	-7.84	<.0001
RunPulse	-0.127565	-0.056435	0	-1.50	0.1597

The "Multiple-Imputation Parameter Estimates" table displays the estimated mean and standard error of the regression coefficients. The inferences are based on the *t* distribution. The table also displays a 95% mean confidence interval and a *t* test with the associated *p*-value for the hypothesis that the regression coefficient is equal to zero. Since the p-value for $RunPulse$ is 0.1597, this variable may be removed from the regression model.

Example 3. Reading Mixed Model Results from PARMS= and COVB= Data Sets

This example creates data sets containing parameter estimates and covariance matrices computed by a mixed model analysis for a set of imputed data sets. These estimates are then combined to generate valid statistical inferences about the parameters.

The following PROC MIXED statements generate the fixed-effect parameter estimates and covariance matrix for each imputed data set:

```
proc mixed data=miout;
   model Oxygen= RunTime RunPulse/solution covb;
   by _Imputation_;
   ods output SolutionF=mixparms CovB=mixcovb;
run;

proc print data=mixparms (obs=6);
   var _Imputation_ Effect Estimate;
   title 'MIXED Model Coefficients (First Two Imputations)';
run;

proc print data=mixcovb (obs=6);
   var _Imputation_ Effect Col1 Col2 Col3;
   title 'MIXED Covariance Matrices (First Two Imputations)';
run;
```

Output 11.0.11. PROC MIXED Model Coefficients

MIXED Model Coefficients (First Two Imputations)

Obs	_Imputation_	Effect	Estimate
1	1	Intercept	89.8363
2	1	RunTime	-3.0947
3	1	RunPulse	-0.05644
4	2	Intercept	96.8952
5	2	RunTime	-3.1170
6	2	RunPulse	-0.09877

Output 11.0.12. PROC MIXED Covariance Matrices

MIXED Covariance Matrices (First Two Imputations)

Obs	_Imputation_	Effect	Col1	Col2	Col3
1	1	Intercept	61.5309	-0.6139	-0.3193
2	1	RunTime	-0.6139	0.1378	-0.00494
3	1	RunPulse	-0.3193	-0.00494	0.002167
4	2	Intercept	56.3022	-0.2783	-0.3140
5	2	RunTime	-0.2783	0.1590	-0.00826
6	2	RunPulse	-0.3140	-0.00826	0.002370

The following statements use the MIANALYZE procedure with PARMS= and COVB= input data sets to produce the same results as in Example 2:

```
proc mianalyze parms=mixparms covb=mixcovb edf=28;
   var Intercept RunTime RunPulse;
run;
```

Example 4. Reading Generalized Linear Model Results from PARMS= and COVB= Data Sets

This example creates data sets containing parameter estimates and corresponding covariance matrices computed by a generalized linear model analysis for a set of imputed data sets. These estimates are then combined to generate valid statistical inferences about the model parameters.

The following statements use PROC GENMOD to generate the parameter estimates and covariance matrix for each imputed data set:

```
proc genmod data=miout;
   model Oxygen= RunTime RunPulse/covb;
   by _Imputation_;
   ods output ParameterEstimates=gmparms
              CovB=gmcovb;
run;

proc print data=gmparms (obs=8);
   var _Imputation_ Parameter Estimate;
   title 'GENMOD Model Coefficients (First Two Imputations)';
run;

proc print data=gmcovb (obs=8);
   var _Imputation_ RowName Prm1 Prm2 Prm3;
   title 'GENMOD Covariance Matrices (First Two Imputations)';
run;
```

Output 11.0.13. PROC GENMOD Model Coefficients

GENMOD Model Coefficients (First Two Imputations)

Obs	_Imputation_	Parameter	Estimate
1	1	Intercept	89.8363
2	1	RunTime	-3.0947
3	1	RunPulse	-0.0564
4	1	Scale	2.6236
5	2	Intercept	96.8952
6	2	RunTime	-3.1170
7	2	RunPulse	-0.0988
8	2	Scale	2.6577

The following table displays the covariance matrices for the first two imputed data sets. Note that the GENMOD procedure computes maximum likelihood estimate for the covariance matrix.

Output 11.0.14. PROC GENMOD Covariance Matrices

GENMOD Covariance Matrices (First Two Imputations)

Obs	_Imputation_	Row Name	Prm1	Prm2	Prm3
1	1	Prm1	55.576282	-0.554494	-0.288436
2	1	Prm2	-0.554494	0.1244661	-0.004459
3	1	Prm3	-0.288436	-0.004459	0.0019571
4	1	Scale	3.221E-15	4.056E-16	-3.99E-17
5	2	Prm1	50.853576	-0.251343	-0.283604
6	2	Prm2	-0.251343	0.1436123	-0.007458
7	2	Prm3	-0.283604	-0.007458	0.0021408
8	2	Scale	1.573E-15	-8.42E-16	4.556E-17

The following statements use the MIANALYZE procedure with PARMS= and COVB= input data sets:

```
proc mianalyze parms=gmparms covb=gmcovb;
   var Intercept RunTime RunPulse;
run;
```

Since the GENMOD procedure computes maximum likelihood estimates for the covariance matrix, the EDF= option is not used. The resulting model coefficients are identical to the estimates from Example 2, but the standard errors are slightly different because in this example, maximum likelihood estimates for the standard errors are combined without the EDF= option, where as in Example 2, unbiased estimates for the standard errors are combined with the EDF= option.

Example 5. Reading GLM Results from PARMS= and XPXI= Data Sets

This example creates data sets containing parameter estimates and corresponding $(X'X)^{-1}$ matrices computed by a general linear model analysis for a set of imputed data sets. These estimates are then combined to generate valid statistical inferences about the model parameters.

The following statements use PROC GLM to generate the parameter estimates and $(X'X)^{-1}$ matrix for each imputed data set:

```
proc glm data=miout;
   model Oxygen= RunTime RunPulse/inverse;
   by _Imputation_;
   ods output ParameterEstimates=glmparms
              InvXPX=glmxpxi;
run;
```

```
proc print data=glmparms (obs=6);
   var _Imputation_ Parameter Estimate;
   title 'GLM Model Coefficients (First Two Imputations)';
run;

proc print data=glmxpxi (obs=8);
   var _Imputation_ Parameter Intercept RunTime RunPulse;
   title 'GLM X'X Inverse Matrices (First Two Imputations)';
run;
```

Output 11.0.15. PROC GLM Model Coefficients

GLM Model Coefficients (First Two Imputations)

Obs	_Imputation_	Parameter	Estimate	StdErr
1	1	Intercept	89.83625110	7.84416236
2	1	RunTime	-3.09465714	0.37121659
3	1	RunPulse	-0.05643530	0.04654854
4	2	Intercept	96.89520358	7.50347743
5	2	RunTime	-3.11704067	0.39874724
6	2	RunPulse	-0.09876620	0.04868479

Output 11.0.16. PROC GLM $(X'X)^{-1}$ Matrices

GLM X'X Inverse Matrices (First Two Imputations)

Obs	_Imputation_	Parameter	Intercept	RunTime	RunPulse
1	1	Intercept	8.0741900546	-0.080557497	-0.041904337
2	1	RunTime	-0.080557497	0.0180825874	-0.000647829
3	1	RunPulse	-0.041904337	-0.000647829	0.0002843269
4	1	Oxygen	89.8362511	-3.094657138	-0.056435305
5	2	Intercept	7.1996847537	-0.035584352	-0.040151747
6	2	RunTime	-0.035584352	0.0203321682	-0.001055814
7	2	RunPulse	-0.040151747	-0.001055814	0.0003030923
8	2	Oxygen	96.895203581	-3.117040668	-0.098766199

The following statements use the MIANALYZE procedure with PARMS= and XPXI= input data sets to produce the same results as in the example 2:

```
proc mianalyze parms=glmparms xpxi=glmxpxi edf=28;
   var Intercept RunTime RunPulse;
run;
```

Example 6. Combining Correlation Coefficients

This example combines sample correlation coefficients and associated variances computed from a set of imputed data sets.

The following statements use the CORR procedure to compute the correlation r between variables Oxygen and RunTime for each imputed data set:

```
proc corr data=outmi out=outcorr;
   var Oxygen RunTime;
   by _Imputation_;
run;

proc print data=outcorr (obs=10);
   title 'Correlations (First Two Imputations)';
run;
```

Output 11.0.17. CORR Type Data Set

Correlations (First Two Imputations)

Obs	_Imputation_	_TYPE_	_NAME_	Oxygen	RunTime
1	1	MEAN		47.3512	10.6004
2	1	STD		5.3235	1.4168
3	1	N		31.0000	31.0000
4	1	CORR	Oxygen	1.0000	-0.8578
5	1	CORR	RunTime	-0.8578	1.0000
6	2	MEAN		47.3489	10.5350
7	2	STD		5.6875	1.4147
8	2	N		31.0000	31.0000
9	2	CORR	Oxygen	1.0000	-0.8609
10	2	CORR	RunTime	-0.8609	1.0000

The following statements compute Fisher's Z transformation of r

$$z = \frac{1}{2} log \left(\frac{1+r}{1-r} \right)$$

and the variance estimate corresponding to z, $1/(n-3)$.

```
data ztrans(type=EST);
   set outcorr (drop= RunTime rename= Oxygen= Z);
   if (_type_ = 'N' or _name_ = 'RunTime');

   if (_type_ = 'CORR')  then do;
      _type_= 'PARMS';
      _name_='';
      Z= 0.5 * log((1+Z)/(1-Z));
      end;
```

```
else if (_type_ = 'N')  then do;
   _type_='COVB';
   _name_='Z';
   Z= 1. / (Z-3);
   end;
run;

proc print data=ztrans;
   title 'EST Type Data Set with Fisher''s Z Transformation';
run;
```

Output 11.0.18. Fisher's Z transformation

```
             EST Type Data Set with Fisher's Z Transformation

           Obs    _Imputation_    _TYPE_    _NAME_        Z

             1          1         COVB        Z        0.03571
             2          1         PARMS               -1.28502
             3          2         COVB        Z        0.03571
             4          2         PARMS               -1.29693
             5          3         COVB        Z        0.03571
             6          3         PARMS               -1.29719
             7          4         COVB        Z        0.03571
             8          4         PARMS               -1.32379
             9          5         COVB        Z        0.03571
            10          5         PARMS               -1.36734
```

The following statements use the MIANALYZE procedure to generate a combined parameter estimate $\hat{z}$ and variance for Fisher's z. The ODS statement is used to save the parameter estimates in an output data set.

```
proc mianalyze data=ztrans;
   ods output ParmEst=parms;
   var z;
run;
```

Output 11.0.19. Inferences Based on Fisher's Z

```
                             The MIANALYZE Procedure

                     Multiple-Imputation Parameter Estimates

                                   Standard
   Parameter      Estimate            Error    95% Confidence Limits        DF

   z             -1.314051       0.192407     -1.69130     -0.93680   3212.4

                     Multiple-Imputation Parameter Estimates

                                                            t for H0:
Parameter       Minimum        Maximum            Mu0   Parameter=Mu0   Pr > |t|

z             -1.367335      -1.285017              0           -6.83     <.0001
```

In addition to the estimate for z, PROC MIANALYZE also generates 95% confidence limits for z, $\hat{z}_{.025}$ and $\hat{z}_{.975}$. An estimate of the correlation coefficient and 95% confidence limits can then be generated from the inverse transformation

$$r = \frac{e^{2z} - 1}{e^{2z} + 1}$$

for $z = \hat{z}$, $\hat{z}_{.025}$, and $\hat{z}_{.975}$.

The following statements print the estimate for z and 95% confidence limits for z:

```
proc print data=parms;
   title 'Parameter Estimates with 95% Confidence Limits';
   var Estimate LCLMean UCLMean;
run;
```

Output 11.0.20. Parameter Estimates with 95% Confidence Limits

```
            Parameter Estimates with 95% Confidence Limits

             Obs        Estimate     LCLMean     UCLMean

              1        -1.314051    -1.69130    -0.93680
```

The following statements generate an estimate of the correlation coefficient and 95% confidence limits:

```
data corr_ci;
   set parms;
   r=     (exp(2*Estimate)-1)/(exp(2*Estimate)+1);
   r_low= (exp(2*LCLMean)-1) /(exp(2*LCLMean)+1);
   r_upp= (exp(2*UCLMean)-1) /(exp(2*UCLMean)+1);
;

proc print data=corr_ci;
   title 'Estimated Correlation Coefficient'
         ' with 95% Confidence Limits';
   var r r_low r_upp;
run;
```

Output 11.0.21. Estimated Correlation Coefficient

```
          Correlation Coefficient with 95% Confidence Limits

                Obs         r          r_low       r_upp

                 1     -0.86530    -0.93431    -0.73375
```

Example 7. Combining Ratios of Variable Means

This example combines ratios of variable means and associated variances computed from imputed data sets.

The following statements use the CORR procedure to compute the means and covariance matrix of variables Oxygen and RunTime for each imputed data set. Within each imputation, the data set is sorted so that observations with _TYPE_='MEAN' and _TYPE_='N' are read before observations with _TYPE_='COV'.

```
proc corr data=outmi cov nocorr noprint out=outcov;
   var RunTime RunPulse;
   by _Imputation_;
run;

proc sort data=outcov;
   by _Imputation_ _name_;
run;

proc print data=outcov (obs=10);
   title 'Means and Covariance Matrices (First Two Imputations)';
run;
```

Output 11.0.22. Means and Covariance Matrices

Means and Covariance Matrices (First Two Imputations)

Obs	_Imputation_	_TYPE_	_NAME_	RunTime	Run Pulse
1	1	MEAN		10.6004	171.533
2	1	STD		1.4168	11.299
3	1	N		31.0000	31.000
4	1	COV	RunPulse	4.5734	127.656
5	1	COV	RunTime	2.0072	4.573
6	2	MEAN		10.5350	169.172
7	2	STD		1.4147	11.587
8	2	N		31.0000	31.000
9	2	COV	RunPulse	6.9721	134.265
10	2	COV	RunTime	2.0015	6.972

For each imputation, the following statements compute a ratio estimate of the means for the variables RunPulse and RunTime, and a corresponding variance estimate:

```
data vratio (type=EST);
   set outcov;
   keep _Imputation_ _type_ _name_ ratio;
   retain TMean PMean N VarP;

   if (_type_ = 'N')  then N= RunTime;

   if (_type_ = 'MEAN')  then do;
      TMean= RunTime;
      PMean= RunPulse;
      Ratio= RunPulse / RunTime;
      _type_= 'PARMS';
      end;

   if (_type_ = 'COV' and _name_ = 'RunPulse')
      then VarP= RunPulse;

   if (_type_ = 'COV' and _name_ = 'RunTime')  then do;
      Ratio= ( PMean**2/TMean**4 * RunTime
                      - 2 * PMean/TMean**3 * RunPulse
                      + (1./ TMean**2) * VarP ) / N;
      _name_= 'Ratio';
      end;

   if ( _type_= 'STD' or _type_= 'N' or _name_= 'RunPulse')
      then delete;
run;

proc print data=vratio;
  title 'EST Type Data Set with Ratios of Variable Means';
run;
```

Output 11.0.23. Ratio of Variable Means

EST Type Data Set with Ratios of Variable Means

Obs	_Imputation_	_TYPE_	_NAME_	Ratio
1	1	PARMS		16.1819
2	1	COV	Ratio	0.1450
3	2	PARMS		16.0582
4	2	COV	Ratio	0.1240
5	3	PARMS		16.2721
6	3	COV	Ratio	0.1483
7	4	PARMS		16.0773
8	4	COV	Ratio	0.1482
9	5	PARMS		16.1030
10	5	COV	Ratio	0.1681

The following statements use the MIANALYZE procedure to generate a combined point estimate for the ratio of variable means and a variance estimate.

```
proc mianalyze data=vratio;
   var ratio;
run;
```

Output 11.0.24. Inferences for A Ratio of Variable Means

The MIANALYZE Procedure

Multiple-Imputation Parameter Estimates

Parameter	Estimate	Standard Error	95% Confidence Limits		DF
ratio	16.138495	0.395047	15.36337	16.91362	1113

Multiple-Imputation Parameter Estimates

Parameter	Minimum	Maximum	Mu0	t for H0: Parameter=Mu0	Pr > |t|
ratio	16.058154	16.272111	0	40.85	<.0001

The variable RunTime has an estimated mean around 10.6 and a standard error around 1.4 in each imputed data set. An estimated coefficient of variation is $(1.4/\sqrt{31})/10.6 = 2.37\%$. Since the coefficient of variation is less than 10%, the ratio of sample means is approximately unbiased and normally distributed.

References

Barnard, J. and Rubin, D.B. (1999), "Small-Sample Degrees of Freedom with Multiple Imputation," *Biometrika*, 86, 948–955.

Cochran, W.J. (1977), *Sampling Techniques*, 2nd edition, New York: John Wiley & Sons, Inc.

Li, K.H., Raghunathan, T.E., and Rubin, D.B. (1991), "Large-Sample Significance Levels From Multiply Imputed Data Using Moment-Based Statistics and an F Reference Distribution," *Journal of the American Statistical Association*, 86, 1065–1073.

Rubin, D.B. (1976), "Inference and Missing Data," *Biometrika*, 63, 581–592.

Rubin, D.B. (1987), *Multiple Imputation for Nonresponse in Surveys*, New York: John Wiley & Sons, Inc.

Rubin, D.B. (1996), "Multiple Imputation After 18+ Years," *Journal of the American Statistical Association*, 91, 473–489.

Schafer, J.L. (1997), *Analysis of Incomplete Multivariate Data*, New York: Chapman and Hall.