Chapter 1

MANOVA: Multivariate Analysis of Variance

SPSS MANOVA is a generalized multivariate analysis of variance and covariance program which will perform univariate and multivariate linear estimation and tests of hypotheses for any crossed and/or nested design with or without covariates. The user has complete control of the model specification. For example, several effects can be lumped together into a single term. Also, interaction between factors and covariates is allowed.

The sections beginning with 1.2 present univariate analysis of variance models, which include balanced incomplete block designs, confounding designs, nested designs, and split-plot designs. Special features such as collapsing error terms, specifying multiple error terms, partitioning degrees of freedom, contrasts, orthogonal polynomials and analysis of covariance are also discussed.

Tests of significance for a multivariate analysis of variance model include hypotheses and error matrices, four multivariate test criteria, dimension reduction analysis, univariate F tests, and step-down analysis. In addition, principal components analysis and discriminant analysis can be requested. They are documented beginning in Section 1.31.

The sections beginning with 1.38 present multivariate multiple linear regression analysis, which can be considered a special case of multivariate analysis of covariance in which all the independent variables are covariates. Canonical correlation analysis is also discussed.

MANOVA enables the user to analyze a large class of repeated measures designs. The observation can be either single-valued or vector-valued. Covariates, varying or constant across the repeated measures, can also appear in the model. These facilities are described beginning in Section 1.43.

Section 1.51 describes the graphics features available in MANOVA.

MANOVA may require an additional scratch file for which provision must be made in the job setup. See Appendix L for information for the IBM/OS version.

1.1 OVERVIEW

MANOVA specifications are entered via the MANOVA command itself and a number of optional subcommands that fall into the three categories outlined below. For more detail on these, see Section 1.52.

The MANOVA command has the following general format:

```
MANOVA      <dependent variable list> BY <factor list> WITH
            <covariate list>/
```

The MANOVA command, with no subcommands, is the only required specification. A dependent variable list of one variable activates univariate analysis; more than one dependent variable activates multivariate analysis of variance.

Subcommands in the first category specify the factor and data structures of the design. WSFACTOR provides the within-subjects factors for a repeated measures design.

```
WSFACTOR = <factor list>/
```

TRANSFORM requests a linear transformation of the dependent variables and covariates.

```
TRANSFORM (variable list1/variable list2/...) =

    [ORTHONORM]  { DEVIATIONS (refcat)      }
                 { DIFFERENCE               }
    [BASIS]      { HELMERT                  }
    [CONTRAST]   { SIMPLE (refcat)    ,     }   /
                 { REPEATED                 }
                 { POLYNOMIAL [(metric)]    }
                 { SPECIAL (matrix)         }
                 { WSDESIGN <effect list>   }
```

WSDESIGN specifies the model for the within-subjects factors and RENAME can be used to rename the transformed variables.

```
WSDESIGN = < effect list > /

RENAME = newname1, newname2,.../
```

The second category contains subcommands PRINT, PLOT, and PUNCH, which control the amount of optional output produced by MANOVA.

```
PRINT    =  CELLINFO( [MEANS] [SSCP] [COV] [COR] )
 or
NOPRINT     HOMOGENEITY( [BARTLETT] [COCHRAN] [BOXM] )

            DESIGN( [ONEWAY] [OVERALL] [BIAS] [DECOMP]
                    [SOLUTION]                        )

            PRINCOMPS( [COR] [COV] [MINEIGEN(eigcut)]
                       [NCOMP(n)] [ROTATE(rottyp)] )

            ERROR( [SSCP] [COV] [COR] [STDV] )

            SIGNIF( [HYPOTH] [MULTIV] [EIGEN]
                    [DIMENR] [UNIV] [STEPDOWN]
                    [AVERF] [BRIEF] [SINGLEDF] )

            DISCRIM( [RAW] [STAN] [ESTIM] [COR]
                     [ROTATE(rottyp)] [ALPHA(alpha)] )

            PARAMETERS( [ESTIM] [COR] [ORTHO] [NEGSUM] )

            OMEANS[ ( VARIABLES(var list)
                      TABLES( table requests ) ) ]

            PMEANS[ ( VARIABLES(var list)
                      TABLES( table requests )
                      ERROR( errorn ) ) ]

            POBS [ ERROR( errorn ) ]

            FORMAT( [WIDE]     )    /
                    [NARROW]

PLOT     =  [CELLPLOTS] [NORMAL] [BOXPLOTS]
            [STEMLEAF]  [ZCORR]  [PMEANS]
            [POBS]
            [ SIZE( nhor , nvert ) ]          /

PUNCH    =  CELLINFO( [MEAN] [SSCP] [COR] [COV] [STDV] )

            ERROR( [SSCP] [COR] [COV] [STDV] )

            PMEANS [ ( ERROR( errorn ) ) ]

            POBS [ ( ERROR( errorn ) ) ]                /
```

The last category consists of the subcommands that indicate the computational options and model specifications. METHOD provides several options for parameter estimation.

```
METHOD = MODELTYPE(  [MEANS]        )
                     [OBSERVATIONS]

                     [CHOLESKY]
         ESTIMATION( [QR]         [LASTRES]    [CONSTANT]  )
                     [BALANCED]   [NOLASTRES]  [NOCONST]
                     [NOBALANCED]

         SSTYPE( [SEQUENTIAL] )          /
                 [UNIQUE]
```

ANALYSIS subsets and/or reorders the variables.

```
ANALYSIS = <dep var list> WITH <covar list>/
                          - or -
ANALYSIS[(CONDITIONAL)   ]      =    (<dep list 1> WITH <covar list 1>/
        [(UNCONDITIONAL)]                <dep list 2> WITH <covar list 2>/ ...)
                          WITH <covar list> /
```

```
                                        - or -
ANALYSIS( REPEATED  [CONDITIONAL]      )    /
                    [UNCONDITIONAL]
```

PARTITION subdivides the degrees of freedom of a factor.

```
PARTITION(factorname) [= (df1, df2,..)]  /
```

CONTRAST indicates the type of contrast desired for a factor.

```
                                { DEVIATION [(refcat)]   }
                                { DIFFERENCE             }
                                { HELMERT                }
CONTRAST(factorname) =          { SIMPLE [(refcat)]      }  /
                                { REPEATED               }
                                { POLYNOMIAL [(metric)]  }
                                { SPECIAL (matrix)       }
```

ERROR specifies the error term to be used in the model.

```
                { WITHIN or W                }
                { RESIDUAL or R              }
ERROR =         { WITHIN + RESIDUAL or WR    }  /
                { n                          }
```

DESIGN specifies the design model to be analyzed.

```
DESIGN = effect1, effect2,../
```

The DESIGN specification should be the last subcommand of a complete MANOVA run. All the computational and output options are applied to the subsequent DESIGN models unless overridden.

As an example of specifications for MANOVA, consider the following:

```
MANOVA          Y BY A(1,3) B(1,4) WITH X/
                PRINT=CELLINFO(MEANS)/
                METHOD=ESTIMATION(BALANCED)/
                DESIGN=A,B/
                METHOD=ESTIMATION(QR)/
                DESIGN=A,B,A BY B/
```

An analysis of covariance model is specified with Y as the dependent variable, X as the covariate, and A and B as factor variables with three and four levels respectively. The PRINT subcommand requests cell information. The METHOD subcommand indicates that a special balanced processing method be used for parameter estimation. These two options apply to the first DESIGN specification, which requests a main effects model. The second METHOD subcommand requests the (default) QR method for estimating the parameters in the second DESIGN specification (a full model). The PRINT subcommand applied to the first DESIGN will also apply to the second DESIGN.

Note that if the last command is not a DESIGN specification, MANOVA will generate a full model design specification for the problem.

1.2 UNIVARIATE ANALYSIS OF VARIANCE

The basic features of MANOVA useful for univariate analysis of variance are illustrated in the following example taken from Winer (1971, p. 436). An experiment was conducted to evaluate the relative effectiveness of three drugs (Factor DRUG) in bringing about behavioral changes in two categories of patients (Factor CAT). Three patients of each category were assigned at random to one of three drugs, and criterion ratings (Y) were made for each patient. The data are given in Table 1.2.

Table 1.2

		DRUG		
		1	2	3
CAT	1	8 4 0	10 8 6	8 6 4
	2	14 10 6	4 2 0	15 12 9

Figure 1.2 shows SPSS commands to accomplish the analysis of variance of the data. The MANOVA specification defines Y to be the dependent variable and CAT and DRUG the factor variables with two and three levels respectively. Since only one dependent variable (Y) is indicated, a univariate analysis of variance is requested.

Figure 1.2

```
RUN NAME        A UNIVARIATE 2*3 EXAMPLE.
COMMENT         THE DATA ARE TAKEN FROM WINER(1971) PAGE 436.
                Y    : THE DEPENDENT VARIABLE.
                CAT  : FACTOR WITH 2 LEVELS.
                DRUG : FACTOR WITH 3 LEVELS.
VARIABLE LIST   CAT DRUG Y
INPUT FORMAT    FREEFIELD
INPUT MEDIUM    CARD
MANOVA          Y BY CAT(1,2) DRUG(1,3)/
READ INPUT DATA
1 1  8
1 1  4
1 1  0
1 2 10
1 2  8
1 2  6
1 3  8
1 3  6
1 3  4
2 1 14
2 1 10
2 1  6
2 2  4
2 2  2
2 2  0
2 3 15
2 3 12
2 3  9
END INPUT DATA
FINISH
```

The default model generated from the MANOVA specifications is a full factorial. For this example the model is

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$$

where α_i is the main effect of category i, β_j is the main effect of drug j, and $(\alpha\beta)_{ij}$ is the interaction of patient category i and drug j. For the various tests, it is necessary to assume that the error terms, ϵ_{ijk}, are independently identically distributed as normal with mean 0 and variance σ^2.

1.3 Default Output

The default output (without any PRINT subcommand) from a MANOVA run includes

1 An analysis of variance (ANOVA) table. As shown in Figure 1.3a, it gives the sum of squares, degrees of freedom, mean square, *F* value, and the probabilities of each *F* value. The within-cells error term (default error-term if it exists) is used to obtain all the *F* values.

Figure 1.3a

TESTS OF SIGNIFICANCE FOR Y USING SEQUENTIAL SUMS OF SQUARES

SOURCE OF VARIATION	SUM OF SQUARES	DF	MEAN SQUARE	F	SIG. OF F
WITHIN CELLS	106.00000	12	8.83333		
CONSTANT	882.00000	1	882.00000	99.84906	0.0
CAT	18.00000	1	18.00000	2.03774	.179
DRUG	48.00000	2	24.00000	2.71698	.106
CAT BY DRUG	144.00000	2	72.00000	8.15094	.006

2 Statistics for parameter estimation (Figure 1.3b). These consist of estimates of the parameters (COEFF), the standard errors of the estimates (STD. ERR.), the t-value for testing that the parameter is zero, the two-tailed significance of the test, and 95% confidence intervals for the parameters. (Note that the parameters estimated here are not the original α_i, β_j, or $(\alpha\beta)_{ij}$; instead, contrasts of the parameters are estimated. See Section 1.52 for detailed information.)

Figure 1.3b

```
ESTIMATES FOR Y

CONSTANT

 PARAMETER          COEFF.      STD. ERR.       T-VALUE      SIG. OF T    LOWER .95 CL    UPPER .95 CL

       1    7.0000000000          .70053          9.99245          .000         5.47368         8.52632
CAT

 PARAMETER          COEFF.      STD. ERR.       T-VALUE      SIG. OF T    LOWER .95 CL    UPPER .95 CL

       2   -1.0000000000          .70053         -1.42749          .179        -2.52632          .52632
DRUG

 PARAMETER          COEFF.      STD. ERR.       T-VALUE      SIG. OF T    LOWER .95 CL    UPPER .95 CL

       3    0.0                   .99070          0.0             1.000        -2.15854         2.15854
       4   -2.0000000000          .99070         -2.01878          .066        -4.15854          .15854
CAT BY DRUG

 PARAMETER          COEFF.      STD. ERR.       T-VALUE      SIG. OF T    LOWER .95 CL    UPPER .95 CL

       5   -2.0000000000          .99070         -2.01878          .066        -4.15854          .15854
       6    4.0000000000          .99070          4.03756          .002         1.84146         6.15854
```

1.4 Use of the PRINT Subcommand

Additional printed output can be obtained by using the PRINT subcommand. For instance, tests of homogeneity of within-cells variance are produced by specifying

```
MANOVA          Y BY CAT(1,2) DRUG(1,3)/
                PRINT=HOMOGENEITY(BARTLETT,COCHRAN)/
```

The output (Figure 1.4a) includes Bartlett's test and Cochran's test. The significance (P) of both tests is also given.

Figure 1.4a

```
UNIVARIATE HOMOGENEITY OF VARIANCE TESTS

VARIABLE .. Y

     COCHRANS C(2,6) =                    .30189,  P =  .829 (APPROX.)
     BARTLETT-BOX F(5,185) =              .38601,  P =  .858
```

The cell statistics, including the mean, standard deviation, number of observations, and the 95% confidence intervals for the population means can be obtained using

```
MANOVA          Y BY CAT(1,2) DRUG(1,3)/
                PRINT=CELLINFO(MEANS)/
```

The output from the above PRINT subcommand is given in Figure 1.4b.

Figure 1.4b

```
CELL MEANS AND STANDARD DEVIATIONS

VARIABLE .. Y

    FACTOR          CODE                MEAN      STD. DEV.          N   95 PERCENT CONF. INTERVAL

 CAT                   1
  DRUG                 1            4.00000        4.00000          3       -5.93666        13.93666
  DRUG                 2            8.00000        2.00000          3        3.03167        12.96833
  DRUG                 3            6.00000        2.00000          3        1.03167        10.96833

 CAT                   2
  DRUG                 1           10.00000        4.00000          3         .06334        19.93666
  DRUG                 2            2.00000        2.00000          3       -2.96833         6.96833
  DRUG                 3           12.00000        3.00000          3        4.54751        19.45249

FOR ENTIRE SAMPLE                    7.00000        4.31141         18        4.85599         9.14401
```

1.5 Specifying a Model with the DESIGN Subcommand

If the desired model is not the default full factorial, the model must be specified using the DESIGN subcommand. To specify a model that includes only the main effect terms, use

```
MANOVA          Y BY CAT(1,2) DRUG(1,3)/
                DESIGN  CAT,DRUG/
```

If there are three factors, (A, B, and C) with three levels each, the model containing only main effects and the A BY B and B BY C interactions is specified by

```
MANOVA          Y BY A B C (1,3)/
                DESIGN= A, B, C, A BY B, B BY C/
```

The keyword BY in the DESIGN subcommand indicates an interaction term. Thus a three-way interaction is written as A BY B BY C.

1.6 Specifying the ERROR Term

Unless otherwise requested, the within-cells mean square is used as the denominator for all the *F* values. If there is no within-cells error, the residual error is used. The residual mean square is the mean square for all terms not specified in the DESIGN subcommand. For example, if the model containing only main effects for DRUG and CAT is requested using

```
DESIGN= CAT,DRUG/
```

the residual error term is the mean square for the CAT BY DRUG interaction. For the three-factor design specification developed previously, the residual error corresponds to the sum of squares for the pooled A BY C and A BY B BY C interactions since they are not included in the DESIGN specification.

The ERROR subcommand designates the error term to be used for the analysis. See Section 1.91 for rules governing the use of the ERROR subcommand. If different error terms are to be used for the various terms in the design specification, this is indicated in the DESIGN subcommand. See Section 1.92 for further details.

1.7 An Example Using DESIGN and ERROR

The following commands request a main effects model for the data of Figure 1.2. The pooled interaction term (denoted as R for residual) and within-cells error (denoted as W) are used as the error.

```
MANOVA          Y BY CAT(1,2) DRUG(1,3)/
                ERROR=W+R/
                DESIGN=CAT,DRUG/
```

The error subcommand must precede the design specification to which it applies. The analysis of variance table from the preceding commands is given in Figure 1.7.

Figure 1.7

TESTS OF SIGNIFICANCE FOR Y USING SEQUENTIAL SUMS OF SQUARES

SOURCE OF VARIATION	SUM OF SQUARES	DF	MEAN SQUARE	F	SIG. OF F
WITHIN+RESIDUAL	250.00000	14	17.85714		
CONSTANT	882.00000	1	882.00000	49.39200	0.0
CAT	18.00000	1	18.00000	1.00800	.332
DRUG	48.00000	2	24.00000	1.34400	.292

The result in Figure 1.7 can also be obtained by specifying

```
MANOVA          Y BY CAT(1,2) DRUG(1,3)/
                DESIGN = CAT VS W+R, DRUG VS W+R/
```

1.8 Partitioning the Sum of Squares

Often it is desirable to partition the sum of squares associated with the various effects into a number of components that are more relevant to the individual questions of interest. See Cochran and Cox (1957).

In procedure MANOVA partitions are controlled by the keyword PARTITION followed by the name of the factor and the degrees of freedom associated with each component.

To partition the sum of squares for factor DRUG into two components with one degree of freedom each, the following commands can be used.

```
MANOVA          Y BY CAT(1,2) DRUG(1,3)/
                PARTITION(DRUG)=(1,1)/
                DESIGN=CAT,DRUG(1),DRUG(2),CAT BY DRUG/
```

The first component is denoted by DRUG(1), and the second by DRUG(2). The output is given in Figure 1.8.

Figure 1.8

```
TESTS OF SIGNIFICANCE FOR Y USING SEQUENTIAL SUMS OF SQUARES

SOURCE OF VARIATION                SUM OF SQUARES     DF   MEAN SQUARE          F     SIG. OF F

WITHIN CELLS                           106.00000     12       8.83333
CONSTANT                               882.00000      1     882.00000   99.84906           0.0
CAT                                     18.00000      1      18.00000    2.03774          .179
DRUG(1)                                 12.00000      1      12.00000    1.35849          .266
DRUG(2)                                 36.00000      1      36.00000    4.07547          .066
CAT BY DRUG                            144.00000      2      72.00000    8.15094          .006

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

ESTIMATES FOR Y

CONSTANT

 PARAMETER        COEFF.      STD. ERR.      T-VALUE      SIG. OF T   LOWER .95 CL   UPPER .95 CL

         1  7.0000000000         .70053      9.99245           .000        5.47368        8.52632

CAT

 PARAMETER        COEFF.      STD. ERR.      T-VALUE      SIG. OF T   LOWER .95 CL   UPPER .95 CL

         2 -1.0000000000         .70053     -1.42749           .179       -2.52632         .52632

DRUG(1)

 PARAMETER        COEFF.      STD. ERR.      T-VALUE      SIG. OF T   LOWER .95 CL   UPPER .95 CL

         3  0.0                  .99070      0.0              1.000       -2.15854        2.15854

DRUG(2)

 PARAMETER        COEFF.      STD. ERR.      T-VALUE      SIG. OF T   LOWER .95 CL   UPPER .95 CL

         4 -2.0000000000         .99070     -2.01878           .066       -4.15854         .15854

CAT BY DRUG

 PARAMETER        COEFF.      STD. ERR.      T-VALUE      SIG. OF T   LOWER .95 CL   UPPER .95 CL

         5 -2.0000000000         .99070     -2.01878           .066       -4.15854         .15854
         6  4.0000000000         .99070      4.03756           .002        1.84146        6.15854
```

The default contrasts used for partitioning are deviation contrasts (see Section 1.89). The deviation contrasts are not orthogonal, so the two contrasts for DRUG(1) and DRUG(2) are not independent.

1.9 Types of Contrasts

The MANOVA procedure allows specification of six different contrast types: deviation, difference, Helmert, simple, repeated, and polynomial. The user can also input any other contrast matrix via the SPECIAL keyword.

For example, to specify user-supplied orthogonal contrasts for the DRUG factor, the following commands can be used:

```
MANOVA          Y BY CAT(1,2) DRUG(1,3)/
                CONTRAST(DRUG)=SPECIAL(1 1 1 -1 2 -1 1 0 -1)/
                PARTITION(DRUG)=(1,1)/
                DESIGN=CAT,DRUG(1),DRUG(2),CAT BY DRUG(1),
                       CAT BY DRUG(2)/
```

The first set of coefficients (1 1 1) is always the weights for obtaining the constant term. Following the weights vector are the contrasts. The number of contrasts should be equal to the degrees of freedom for the factor. The first contrast (-1 2 -1) defines a contrast between level 2 and the combination of levels 1 and 3 for factor DRUG. The second contrast (1 0 -1) requests a comparison between levels 1 and 3 of DRUG. For most applications, the user shoud be sure that each set of contrast coefficients sum to zero.

Since the inner product of the two contrasts is 0 and the sample sizes in all cells are equal, i.e., (-1)(1) + 2(0) + (-1)(-1) = 0, the two contrasts are independent. In this example, the DRUG(1) partition can be used to test the hypothesis $\beta_2 = (\beta_1 + \beta_3)/2$ while the second contrast tests $\beta_1 = \beta_3$. The ANOVA table is given in Figure 1.9.

Figure 1.9

TESTS OF SIGNIFICANCE FOR Y USING SEQUENTIAL SUMS OF SQUARES

SOURCE OF VARIATION	SUM OF SQUARES	DF	MEAN SQUARE	F	SIG. OF F
WITHIN CELLS	106.00000	12	8.83333		
CONSTANT	882.00000	1	882.00000	99.84906	0.0
CAT	18.00000	1	18.00000	2.03774	.179
DRUG(1)	36.00000	1	36.00000	4.07547	.066
DRUG(2)	12.00000	1	12.00000	1.35849	.266
CAT BY DRUG(1)	144.00000	1	144.00000	16.30189	.002
CAT BY DRUG(2)	0.0	1	0.0	0.0	1.000

The above discussion of orthogonal contrasts assumes that the cell frequencies are equal. For the use of the orthogonal contrasts in unbalanced designs, see Section 1.16.

1.10 Designs with Unequal Cell Frequencies

In many experiments, it may not be possible to have equal numbers of observations for each cell. Such designs are termed *unbalanced* or *nonorthogonal.* In nonorthogonal designs the effects are correlated with each other and cannot be estimated independently of one another. That is, the component sum of squares will not add up to the total sum of squares because the main effects will usually not be independent of each other and the interaction effects will not be independent of the main effects. Different ANOVA solutions can be obtained for the same design depending on the "type" of sum of squares calculated. For example, in an unbalanced design with two factors A and B, the sum of squares for main effect A differs depending on whether effect A is the only one in the model or whether it is added to a model already containing effect B.

1.11 Sequential Sums of Squares (Fitting Constants)

Sequential sums of squares are the default type calculated by MANOVA. The sums of squares for each effect are "adjusted" for all effects previously entered into the model. That is, the sum of squares for an effect is adjusted only for all terms to the left of it in the DESIGN subcommand. All terms to the right are ignored. Therefore the order in which terms are specified on the DESIGN subcommand, or the MANOVA command if a DESIGN subcommand is not present, is important. Different orders may produce different results. For the two-factor design specified using

```
DESIGN=A,B/
```

the B main effect is adjusted for A and the overall mean, while A is adjusted only for the mean. If the model is specified as

```
DESIGN=B,A/
```

the A main effect is adjusted for B and the mean, while the B effect is adjusted only for the mean.

Since several DESIGN subcommands can be used in one invocation of the MANOVA procedure, it is possible to obtain easily various sums of squares. For example, in a two-factor model, to obtain the main effect sum of squares adjusted for other main effects and the interaction effect adjusted for main effects, specify both

```
DESIGN=A,B,A BY B/
DESIGN=B,A, A BY B/
```

The first ANOVA table will contain B adjusted for A, and A BY B adjusted for both main effects. The second ANOVA table will contain A adjusted for B and the interaction adjusted for both main effects.

1.12 Regression Model Sum of Squares (Weighted Squares of Means)

It is possible to obtain sums of squares adjusted for all effects listed on the DESIGN subcommand, by specifying

```
METHOD=SSTYPE(UNIQUE)/
```

For the two-factor model this results in main effect A being adjusted for both B and the A BY B interaction. Similarly B is adjusted for A and the interaction, while the interaction is adjusted for main effects A and B.

1.13 Decomposition and Bias Matrices

If the design is unbalanced and the default sequential sums of squares are used, the decomposition and bias matrices may be of interest. They are obtained by specifying

```
PRINT=DESIGN(DECOMP,BIAS)/
```

The elements in the upper triangle of the decomposition matrix are used to obtain the sum of squares for each effect in the model. Consider a 2 × 3 factorial design, where T is the upper triangle of the decomposition matrix.

$$T = \begin{pmatrix} t_{11} & t_{12} & t_{13} & t_{14} & t_{15} & t_{16} \\ 0 & t_{22} & t_{23} & t_{24} & t_{25} & t_{26} \\ 0 & 0 & t_{33} & t_{34} & t_{35} & t_{36} \\ 0 & 0 & 0 & t_{44} & t_{45} & t_{46} \\ 0 & 0 & 0 & 0 & t_{55} & t_{56} \\ 0 & 0 & 0 & 0 & 0 & t_{66} \end{pmatrix}$$

The first row of T represents the CONSTANT effect, the second row represents the effect of A, the third and fourth rows are the effects of B, and the last two rows are the effects of AB. If $\mathbf{h}' = (h_1\ h_2\ h_3\ h_4\ h_5\ h_6)$ is the least-squares estimate of the contrasts of effects, then the sequential sums of squares for the effects are as shown in Table 1.13.

Table 1.13

Source	Sum of Squares
CONSTANT	$(t_{11}h_1+t_{12}h_2+t_{13}h_3+t_{14}h_4+t_{15}h_5+t_{16}h_6)^2$
A	$(t_{22}h_2+t_{23}h_3+t_{24}h_4+t_{25}h_5+t_{26}h_6)^2$
B adjusted A	$(t_{33}h_3+t_{34}h_4+t_{35}h_5+t_{36}h_6)^2+(t_{44}h_4+t_{45}h_5+t_{46}h_6)^2$
AB adjusted A,B	$(t_{55}h_5+t_{56}h_6)^2+(t_{66}h_6)^2$

If the DESIGN specification for this example is

```
DESIGN=A,B,A BY B/
```

then the bias matrix is a 4 × 4 upper triangular matrix, since the order of the bias matrix is the number of effects in the model (in this case, CONSTANT, A, B, and A BY B). The (i,j)th element of this matrix is obtained by summing the squared elements of the T matrix, which are in the rows of effect i and the columns of effect j. The bias matrix for this example is

$$\begin{pmatrix} t_{11}^2 & t_{12}^2 & t_{13}^2 + t_{14}^2 & t_{15}^2 + t_{16}^2 \\ 0 & t_{22}^2 & t_{23}^2 + t_{24}^2 & t_{25}^2 + t_{26}^2 \\ 0 & 0 & t_{33}^2 + t_{34}^2 + t_{44}^2 & t_{35}^2 + t_{36}^2 + t_{45}^2 + t_{46}^2 \\ 0 & 0 & 0 & t_{55}^2 + t_{56}^2 + t_{66}^2 \end{pmatrix}$$

The bias matrix can be used as a measure of the degree of the confounding among effects. For example, the coefficients corresponding to h_3 and h_4 (factor B) in the calculation of sum of squares of A are t_{23} and t_{24}; thus $t_{23}^2 + t_{24}^2$ (squaring is to avoid the sign) can be used as a confounding index between A and B.

1.14 Redundant Effects

If there are empty cells in the design, some effects in the model may not be estimable. MANOVA determines the redundant effects by orthonormalization of the design matrix and prints the information. Figure 1.14 indicates that the interaction effects in columns 10 and 12 in the design matrix are not estimable because of empty cells.

Figure 1.14

```
REDUNDANCIES IN DESIGN MATRIX

   COLUMN      EFFECT

       10    A BY B
       12     (SAME)
```

1.15 Solution Matrices

For any connected design, the hypotheses associated with the sequential sums of squares are weighted functions of the population cell means with weights depending on the cell frequencies (e.g. see Searle(1971), pp. 306-313). For designs with every cell filled, it can be shown that the hypotheses corresponding to the regression model sums of squares are the unweighted hypotheses about the cell means. With empty cells the hypotheses will depend on the pattern of the missingness. In such cases, one can request that the solution matrix, which contains the coefficients of the linear combinations of the cell means being tested, be printed by specifying

```
PRINT=DESIGN(SOLUTION)/
```

For example, in a 2 × 3 (factors A, B) design with one empty cell. The solution matrix P of this design would be

$$P = \begin{pmatrix} p_{11} & p_{21} & p_{31} & p_{41} & p_{51} & 0 \\ p_{12} & p_{22} & p_{32} & p_{42} & p_{52} & 0 \\ p_{13} & p_{23} & p_{33} & p_{43} & p_{53} & 0 \\ p_{14} & p_{24} & p_{34} & p_{44} & p_{54} & 0 \\ p_{15} & p_{25} & p_{35} & p_{45} & p_{55} & 0 \\ p_{16} & p_{26} & p_{36} & p_{46} & p_{56} & 0 \end{pmatrix}$$

The first column of P indicates that the hypothesis corresponding to the sum of squares of CONSTANT is

$$p_{11}\mu_{11}+p_{12}\mu_{12}+p_{13}\mu_{13}+p_{14}\mu_{21}+p_{15}\mu_{22}+p_{16}\mu_{23} = 0$$

where μ_{ij} is the population mean of cell (i,j).

Similarly, column 2 of P represents the coefficients of the linear combinations of cell means being tested for the sum of squares of A, columns 3 and 4 are for the sum of squares of B, and the last two columns are for the sum of squares of AB.

An Example. The following example is taken from Bancroft (1968, p. 20). Quantitative chemical experiments were run to determine the reacting weights of silver (SILVER) and iodine (IODINE) in silver iodine. Five different batches of silver and two different batches of iodine were used in the experiment. These were treated, and then a determination of the reacting weights was made. The coded data are given in Table 1.15. Note that there are two empty cells in the experiment.

Table 1.15

		Silver				
		1	2	3	4	5
Iodine	1	22 25	41 41	29 20 37	49 50	55
	2	-1 40 18	23 13	–	61	–

The MANOVA commands illustrated in Figure 1.15a produce the analysis shown in Figures 1.15b-1.15d.

Figure 1.15a

```
RUN NAME        A 5*2 DESIGN WITH EMPTY CELLS.
COMMENT         DATA ARE TAKEN FROM BANCROFT(1968) PAGE 20.
VARIABLE LIST   SILVER IODINE RESP
INPUT FORMAT    FREEFIELD
INPUT MEDIUM    CARD
MANOVA          RESP BY SILVER(1,5) IODINE(1,2)/
                PRINT=DESIGN(DECOMP,BIAS)/
                DESIGN=SILVER,IODINE,SILVER BY IODINE/
                DESIGN=IODINE,SILVER,SILVER BY IODINE/
READ INPUT DATA
1 1 22
1 1 25
1 2 -1
1 2 40
1 2 18
2 1 41
2 1 41
2 2 23
2 2 13
3 1 29
3 1 20
3 1 37
4 1 49
4 1 50
4 2 61
5 1 55
END INPUT DATA
FINISH
```

The two DESIGN subcommands are used to obtain the sum of squares for IODINE adjusted for SILVER and vice versa. The decomposition and bias matrices are also requested.

The output (Figure 1.15b) indicates that two degrees of freedom for the SILVER BY IODINE interaction effects are lost because of the empty cells. Therefore, instead of four degrees of freedom, it has only two.

Figure 1.15b

```
REDUNDANCIES IN DESIGN MATRIX

   COLUMN      EFFECT

       9    SILVER BY IODINE
      10     (SAME)
```

The decomposition and bias matrices and ANOVA table for the first DESIGN subcommand are given in Figure 1.15c.

Figure 1.15c

TRIANGULAR DECOMPOSITION OF DESIGN

	PARAMETER					
PARAMETER	1	2	3	4	5	6
1	-4.00000	-1.00000	-.75000	-.50000	-.50000	-1.00000
2	1.73205	-2.23607	-.11180	-.22361	-.22361	1.34164
3	1.41421	-.63060	-2.10357	-.28523	-.28523	.76061
4	1.41421	-.63060	1.19303	1.90227	.32521	1.06028
5	1.73205	-.77233	-.27090	1.60122	1.87427	-.17493
6	1.41421	-.63060	-.22119	-.10682	1.33325	3.38625
7	1.00000	-.44590	-.15640	-.07554	.94275	-2.92061
8	1.00000	-1.44590	-1.19649	-1.00287	-.76008	1.11430

	PARAMETER	
PARAMETER	7	8
1	.50000	.25000
2	-.22361	-.55902
3	-.64177	-.53481
4	.53460	.44550
5	.44982	.37485
6	1.41750	1.18125
7	-1.67054	1.00232
8	-1.49674	-1.26491

BIAS COEFFICIENTS FOR SEQUENTIAL ORDERING

	EFFECT			
EFFECT	1	2	3	4
1	16.00000	2.06250	1.00000	.31250
2	0.0	16.93750	3.53333	1.88750
3	0.0	0.0	11.46667	3.40465
4	0.0	0.0	0.0	5.39535

TESTS OF SIGNIFICANCE FOR RESP USING SEQUENTIAL SUMS OF SQUARES

SOURCE OF VARIATION	SUM OF SQUARES	DF	MEAN SQUARE	F	SIG. OF F
WITHIN CELLS	1041.66667	8	130.20833		
CONSTANT	17095.56250	1	17095.56250	131.29392	0.0
SILVER	2572.30417	4	643.07604	4.93882	.027
IODINE	149.95504	1	149.95504	1.15165	.315
SILVER BY IODINE	491.51163	2	245.75581	1.88740	.213

The PRINT subcommand applies to both DESIGN specifications. Figure 1.15d presents only the analysis of variance table for the second design specification.

Figure 1.15d

TESTS OF SIGNIFICANCE FOR RESP USING SEQUENTIAL SUMS OF SQUARES

SOURCE OF VARIATION	SUM OF SQUARES	DF	MEAN SQUARE	F	SIG. OF F
WITHIN CELLS	1041.66667	8	130.20833		
CONSTANT	17095.56250	1	17095.56250	131.29392	0.0
IODINE	473.20417	1	473.20417	3.63421	.093
SILVER	2249.05504	4	562.26376	4.31819	.037
SILVER BY IODINE	491.51163	2	245.75581	1.88740	.213

1.16 Orthogonal Contrasts for Unequal Numbers of Replicates

For balanced designs, two treatment contrasts are orthogonal if the cross products of the contrast coefficients sum to zero. When treatments have unequal numbers of replicates, for contrasts to be orthogonal the weighted sum of cross products, where the weights are the reciprocals of the numbers of replicates, must be zero. For example, suppose the numbers of replicates for five treatments are 4, 2, 1, 5, and 1 respectively; then contrasts (4,2,-6, 0, 0) and (4, 2, 1, 5, -12) are orthogonal, since $4 \times 4/4 + 2 \times 2/2 + (-6)(1)/1 = 0$.

Figure 1.16a illustrates the use of the orthogonal contrasts in a one-way unbalanced design in which the numbers of observations for treatments are 4, 4, 1, and 1, respectively. Note that specification of the PARTITION command without degrees of freedom results in single-degree-of-freedom partitions.

Figure 1.16a

```
RUN NAME        ORTHOGONAL CONTRASTS FOR UNBALANCED DESIGN.
VARIABLE LIST   TREATMNT,Y
N OF CASES      10
INPUT MEDIUM    CARD
INPUT FORMAT    FIXED(F1.0,F2.0)
MANOVA          Y BY TREATMNT(1,4)/
                PRINT=DESIGN(BIAS)/
                CONTRAST(TREATMNT)=SPECIAL(1 1 1 1
                                           1 -1 0 0
                                           4 4 -8 0
                                           4 4 1 -9)/
                PARTITION(TREATMNT)/
                DESIGN=TREATMNT(1),TREATMNT(2),TREATMNT(3)/
READ INPUT DATA
1 8
1 7
2 8
2 9
310
1 6
1 7
2 8
2 6
4 9
FINISH
```

In this example, TREATMNT(1) defines a comparison between treatments 1 and 2; TREATMNT(2) is the contrast between treatment 3 and the combination of treatments 1 and 2; and TREATMNT(3) can be used to test the hypothesis that the average of the first three treatment effects is equal to the last treatment effect. All pairs of contrasts are orthogonal since $(1)(4)/4 + (-1)(4)/4 = 0$, $(1)(4)/4 + (-1)(4)/4 = 0$, and $(4)(4)/4 + (4)(4)/4 + (-8)(1)/1 = 0$. The F tests are therefore independent. The bias matrix and the ANOVA table corresponding to Figure 1.16a are given in Figure 1.16b.

Figure 1.16b

BIAS COEFFICIENTS FOR SEQUENTIAL ORDERING

	EFFECT			
EFFECT	1	2	3	4
1	10.00000	0.0	.00434	.00278
2	0.0	2.00000	0.0	0.0
3	0.0	0.0	.01389	0.0
4	0.0	0.0	0.0	.01111

TESTS OF SIGNIFICANCE FOR Y USING SEQUENTIAL SUMS OF SQUARES

SOURCE OF VARIATION	SUM OF SQUARES	DF	MEAN SQUARE	F	SIG. OF F
WITHIN CELLS	6.75000	6	1.12500		
CONSTANT	608.40000	1	608.40000	540.80000	0.0
TREATMNT(1)	1.12500	1	1.12500	1.00000	.356
TREATMNT(2)	6.12500	1	6.12500	5.44444	.058
TREATMNT(3)	1.60000	1	1.60000	1.42222	.278

The second example is adapted from Cochran and Cox (1957, p. 46). The experiment was conducted to compare the effectiveness of four soil fumigants in keeping down the number of eelworms in the soil. The fumigants were CN, CS, CM, and CK. Each fumigant was tested both in a single and double dose. The control was used as another treatment. The nine treatments are denoted as C00 (control), CN1 (CN with single dose), CS1, CM1, CK1, CN2 (CN with double dose), CS2, CM2, and CK2. There were four replications for each dose of each fumigant and 16 replications of the control. The desired subdivisions of the treatment sum of squares are as follows:

1 If the effect of the fumigants is proportional to the dose, then both CN1 and CN2/2 are the estimate of the effect of CN per unit dose. The pooled estimate of this effect is (CN1+ 2(CN2))/5. The differences in the linear responses to the four fumigants can be measured by the following three contrasts:

```
(0  1 -1  0  0  2 -2  0  0)
(0  1  1 -2  0  2  2 -4  0)
(0  1  1  1 -3  2  2  2 -6)
```

2 The curvature of the treatment CN is measured by C00 − (2CN1) + CN2. The differences in curvature are compared by the quantities CN2 -2(CN1), (the C00 term cancelled out in the comparison) or by the following three contrasts:

```
(0  2 -2  0  0 -1  1  0  0)
(0  2  2 -4  0 -1 -1  2  0)
(0  2  2  2 -6 -1 -1 -1  3)
```

3 The sum of squares between levels (control: 0 level; treatments with single dose: level 1; treatments with double level: level 2) can be partitioned into a component due to the linearity between levels and one representing the curvature between levels. The former is given by the comparison of −1(level 0) +0(level 1) +1(level 2), or the contrast (−4 0 0 0 0 1 1 1 1). The curvature between levels is measured by 1(level 0) −2(level 1) +1(level 2), or the contrast (−4 −2 −2 −2 −2 1 1 1 1).

The above partitions can be summarized by the following MANOVA CONTRAST subcommand:

```
CONTRAST(TREATMNT)=SPECIAL(1  1  1  1  1  1  1  1  1
                           0  1 -1  0  0  2 -2  0  0
                           0  1  1 -2  0  2  2 -4  0
                           0  1  1  1 -3  2  2  2 -6
                           0  2 -2  0  0 -1  1  0  0
                           0  2  2 -4  0 -1 -1  2  0
                           0  2  2  2 -6 -1 -1 -1  3
                          -4  0  0  0  0  1  1  1  1
                           4 -2 -2 -2 -2  1  1  1  1)/
PARTITION(TREATMNT)=(3 3 1 1)/
DESIGN=BLOCK TREATMNT(1) TREATMNT(2)
      TREATMNT(3) TREATMNT(4)/
```

Note that TREATMNT(1), TREATMNT(2), TREATMNT(3), and TREATMNT(4) are the effects of the differences in linear response, in curvature, linear response between levels, and curvature between levels, respectively. Also, it can be verified that the effects are orthogonal.

1.17 Analysis of Covariance

SPSS MANOVA can perform an analysis of covariance in which interval-scaled independent variables (covariates) are used in conjunction with categorical variables (factors). Analysis of covariance is a technique that combines the features of analysis of variance and regression. A two-way analysis of covariance model with two covariates can be described as follows:

$$Y_{ijk} = \mu+\alpha_i+\gamma_j+(\alpha\gamma)_{ij}+\beta_1(X_{1ijk}-\bar{X}_1)+\beta_2(X_{2ijk}-\bar{X}_2)+\epsilon_{ijk}$$

where Y_{ijk} is the dependent variable, α_i, γ_j are the main effects, and $(\alpha\gamma)_{ij}$ is the interaction effect. X_1, X_2 are the covariates, and $\bar{X}_1$, $\bar{X}_2$ are the means for the two covariates.

In the covariance model, Y has a (multiple) linear regression (see Section 1.38) on X_1 and X_2 with regression coefficients β_1 and β_2. The regression procedure is used to remove the variation in the dependent variable due to covariates.

From the standpoint of the analysis of variance model, the covariate model is essentially an analysis of variance model on the corrected scores or

$$Y_{ijk}-\beta_1(X_{1ijk}-\bar{X}_1)-\beta_2(X_{2ijk}-\bar{X}_2) = \mu+\alpha_i+\gamma_j+(\alpha\gamma)_{ij}+\epsilon_{ijk}$$

which is the analysis of variance model for Y adjusted for the two covariates.

The following illustrative example is taken from Snedecor and Cochran (1967, p. 422). The model is a one-way analysis of covariance with one covariate. The experiment was conducted to evaluate the effect of three drugs on the treatment of leprosy. For each patient, six sites were selected. The variate X, based on laboratory tests, is a score representing the abundance of leprosy

bacilli at these sites before the experiment began. Variate Y is a similar score after several months of treatment. Drugs 1 and 2 are antibiotics, while drug 3 is an inert drug included as a control. Ten patients were selected for each treatment. The MANOVA commands are as follows:

```
MANOVA          Y BY DRUG(1,3) WITH X/
                PRINT= PMEANS/
```

Inclusion of covariates in a model is indicated by the keyword WITH on the MANOVA command. The PRINT = PMEANS (see Section 1.50) specification requests the predicted and adjusted (for covariate) means of treatments.

The output includes the analysis of covariance summary table shown in Figure 1.17a, which gives the sum of squares due to regression (adjusted for the factor DRUG), and the sum of squares due to DRUG adjusted for regression.

Figure 1.17a

TESTS OF SIGNIFICANCE FOR Y USING SEQUENTIAL SUMS OF SQUARES

SOURCE OF VARIATION	SUM OF SQUARES	DF	MEAN SQUARE	F	SIG. OF F
WITHIN CELLS	417.20260	26	16.04625		
REGRESSION	577.89740	1	577.89740	36.01447	0.0
CONSTANT	31.92864	1	31.92864	1.98979	.170
DRUG	68.55371	2	34.27686	2.13613	.138

In addition, the estimated regression coefficient (B), the standardized regression coefficient (BETA), the standard error of the regression coefficient and the t-value of the test that $\beta = 0$ are also given (Figure 1.17b). Note that $(6.00121)^2 = 36.014$, which is the F value for the regression in the ANOVA table.

Figure 1.17b

REGRESSION ANALYSIS FOR WITHIN CELLS ERROR TERM

DEPENDENT VARIABLE ..Y

COVARIATE	B	BETA	STD. ERR.	T-VALUE	SIG. OF T	LOWER .95 CL	UPPER .95 CL
X	.9871838111	.7620649867	.16450	6.00121	.000	.64905	1.32531

The adjusted and predicted means for the factor DRUG are shown in Figure 1.17c.

Figure 1.17c

ADJUSTED AND ESTIMATED MEANS

VARIABLE .. Y

FACTOR	CODE	OBS. MEAN	ADJ. MEAN	EST. MEAN	RAW RESID.	STD. RESID.
DRUG	1	5.30000	6.71496	5.30000	0.0	0.0
DRUG	2	6.10000	6.82393	6.10000	0.0	0.0
DRUG	3	12.30000	10.16110	12.30000	0.0	0.0

Since MANOVA allows the inclusion of interval-scaled variables in the DESIGN specification, the analysis of covariance can also be obtained using the following MANOVA commands:

```
MANOVA          Y,X, BY DRUG(1,3)/
                ANALYSIS = Y/
                DESIGN = X, DRUG/
                DESIGN = DRUG,X/
```

The ANALYSIS subcommand is used to select Y as the dependent variable. The first DESIGN subcommand produces the DRUG effects adjusted for the covariate (X). The output is given in Figure 1.17d.

Figure 1.17d

TESTS OF SIGNIFICANCE FOR Y USING SEQUENTIAL SUMS OF SQUARES

SOURCE OF VARIATION	SUM OF SQUARES	DF	MEAN SQUARE	F	SIG. OF F
WITHIN+RESIDUAL	417.20260	26	16.04625		
CONSTANT	1872.30000	1	1872.30000	116.68144	0.0
X	802.94369	1	802.94369	50.03932	0.0
DRUG	68.55371	2	34.27686	2.13613	.138

The second DESIGN specification requests the regression effect (X) adjusted for the factor DRUG (Figure 1.17e).

Figure 1.17e

TESTS OF SIGNIFICANCE FOR Y USING SEQUENTIAL SUMS OF SQUARES

SOURCE OF VARIATION	SUM OF SQUARES	DF	MEAN SQUARE	F	SIG. OF F
WITHIN+RESIDUAL	417.20260	26	16.04625		
CONSTANT	1872.30000	1	1872.30000	116.68144	0.0
DRUG	293.60000	2	146.80000	9.14855	.001
X	577.89740	1	577.89740	36.01447	0.0

The regression coefficient can be obtained from the estimate of the parameters for factor X (Figure 1.17f).

Figure 1.17f

ESTIMATES FOR Y

CONSTANT

PARAMETER	COEFF.	STD. ERR.	T-VALUE	SIG. OF T	LOWER .95 CL	UPPER .95 CL
1	-2.6957729061	1.91108	-1.41060	.170	-6.62406	1.23252

DRUG

PARAMETER	COEFF.	STD. ERR.	T-VALUE	SIG. OF T	LOWER .95 CL	UPPER .95 CL
2	-1.1850365374	1.06082	-1.11709	.274	-3.36559	.99551
3	-1.0760652052	1.04130	-1.03339	.311	-3.21648	1.06435

X

PARAMETER	COEFF.	STD. ERR.	T-VALUE	SIG. OF T	LOWER .95 CL	UPPER .95 CL
4	.9871838111	.16450	6.00121	.000	.64905	1.32531

From the covariance model given above, it follows that there is a common regression coefficient for the given X. This implies that the within-treatment regression coefficients are homogeneous. The assumption of homogeneity of regression coefficients in the analysis of covariance can be assessed by introducing a treatment by covariate interaction term in the model.

A test for no interaction between DRUG effects and covariate is equivalent to testing the hypothesis that the pooled within-treatment regression coefficient is appropriate. The test for treatment by covariate interaction, which is referred to as the test for regression parallelism, can be obtained in MANOVA as follows:

```
MANOVA      Y, X BY DRUG(1,3)/
            ANALYSIS = Y/
            DESIGN = X, DRUG, X BY DRUG/
```

The analysis of variance table for this DESIGN specification is given in Figure 1.17g.

Since X BY DRUG is not significant, the hypothesis of the homogeneity of the within-treatment regression is not rejected.

Figure 1.17g

TESTS OF SIGNIFICANCE FOR Y USING SEQUENTIAL SUMS OF SQUARES

SOURCE OF VARIATION	SUM OF SQUARES	DF	MEAN SQUARE	F	SIG. OF F
WITHIN+RESIDUAL	397.55795	24	16.56491		
CONSTANT	1872.30000	1	1872.30000	113.02805	0.0
X	802.94369	1	802.94369	48.47255	0.0
DRUG	68.55371	2	34.27686	2.06924	.148
X BY DRUG	19.64465	2	9.82232	.59296	.561

1.18 Analysis of Covariance with Separate Regression Estimates

Consider a 2 × 2 (Factors A, B) design with covariate X. The model (using dummy variables) can be written as

$$Y_{ijk} = \mu + \beta(X_{ijk} - \bar{X}) + \alpha_i Z_{ijk} + \alpha_\varepsilon U_{ijk} + \alpha_s Z_{ijk} U_{ijk} + \epsilon_{ijk}$$

where

Z_{ijk} = 1 if i = 2 (level 2 of A is applied)
0 otherwise

U_{ijk} = 1 if j = 2 (level 2 of B is applied)
0 otherwise

If the interaction terms between the covariate and factor variables are added to the model, then

$$Y_{ijk} = \mu+\beta(X_{ijk}-\bar{X})+\alpha_1 Z_{ijk}+\alpha_2 U_{ijk}+\alpha_3 Z_{ijk}U_{ijk} +(\alpha\beta)_1(X_{ijk}-\bar{X})Z_{ijk}+(\alpha\beta)_2(X_{ijk}-\bar{X})U_{ijk}+(\alpha\beta)_3(X_{ijk}-\bar{X})Z_{ijk}U_{ijk}+\epsilon_{ijk}$$

A test of H_0: $(\alpha\beta)_1 = (\alpha\beta)_2 = (\alpha\beta)_3 = 0$ is equivalent to testing the hypothesis that the regression slopes are the same for all cells. This test can be performed by specifying the following MANOVA commands:

```
MANOVA          Y X BY A(1,2),B(1,2)/
                ANALYSIS=Y/
                DESIGN=X,A,B,A BY B,
                   X BY A + X BY B + X BY A BY B/
```

The effects X BY A, X BY B, and X BY A BY B are lumped together to provide the test of the parallelism hypothesis. If the test is not significant, the usual analysis of covariance model can be used to perform the analysis.

If the assumption of the homogeneity of the slope is violated, one of the following three models might be used:

1 The model of different slopes for each level of factor A. This model can be justified by testing $(\alpha\beta)_2 = (\alpha\beta)_3 = 0$. The MANOVA specification for the test is

```
DESIGN=X,A,B,A BY B,
   X BY B + X BY A BY B/
```

If the test is not significant, the following DESIGN specifications can be used for the analysis of covariance of this model:

```
DESIGN=X WITHIN A, A, B, A BY B/
DESIGN=A, B, A BY B, X WITHIN A/
```

The X WITHIN A term represents the regression effects that are separately estimated within each level of A. The first DESIGN specification requests the main effects and interaction adjusted for the covariate effects. The second DESIGN specification gives the regression effect (last term) adjusted for A, B and AB.

2 The model of different slopes for each level of factor B. The appropriate test for this model is $(\alpha\beta)_1 = (\alpha\beta)_3 = 0$ and is obtained by specifying

```
DESIGN= X, A, B, A BY B, X BY A + X BY A BY B/
```

The analysis of covariance is obtained by using

```
DESIGN=X WITHIN B, A, B, A BY B/
DESIGN=A, B, A BY B, X WITHIN B/
```

3 The model of different slopes for each cell. The MANOVA specifications for this model are

```
DESIGN=X WITHIN A BY B, A, B, A BY B/
DESIGN=A, B, A BY B, X WITHIN A BY B/
```

The X WITHIN A BY B term represents the regression slopes, which are different for each cell.

The same procedure can be simply extended to multiple covariates. For a 2 × 2 design with covariates Z1 and Z2, the X term is replaced by CONTIN(Z1,Z2) throughout the DESIGN specification discussed above. The keyword CONTIN incorporates Z1 and Z2 into a single effect. Thus the following specfications may be used for the analysis of covariance for model 1 with covariates Z1 and Z2.

```
MANOVA          Y Z1 Z2 BY A(1,2) B(1,2)/
                ANALYSIS=Y/
                DESIGN=Z1, Z2, A, B, A BY B,
                   CONTIN(Z1,Z2) BY B + CONTIN(Z1,Z2)BY A BY B/
                DESIGN=CONTIN(Z1,Z2) WITHIN A, A, B, A BY B/
                DESIGN=A, B, A BY B, CONTIN(Z1,Z2) WITHIN A/
```

The first DESIGN specification is used to test the model, while the second and third models are for the analysis of covariance.

An Example.

The following example is taken from Searle (1971, pp. 287,375). An experiment was conducted to compare the effects of three different types of fertilizer and four varieties of grain on the weight of grain (WEIGHT). The milligrams of seed planted (MSEED) for each plot were also recorded and used as the covariate. The SPSS commands and data for model 3 are presented in Figure 1.18a, and the analysis of variance tables in Figure 1.18b.

Figure 1.18a

```
RUN NAME        EMPTY CELLS EXAMPLE FROM SEARLE(1971).
COMMENT         DATA ARE TAKEN FROM P. 287 AND P. 375.
VARIABLE LIST   TREATMNT, VARIETY, WEIGHT, MSEED
N OF CASES      18
INPUT FORMAT    FREEFIELD
INPUT MEDIUM    CARD
MANOVA          WEIGHT MSEED BY TREATMNT(1,3), VARIETY(1,4)/
                ANALYSIS=WEIGHT/
                DESIGN = MSEED WITHIN VARIETY BY TREATMNT ,
                  VARIETY, TREATMNT, VARIETY BY TREATMNT/
                DESIGN = TREATMNT, VARIETY, VARIETY BY TREATMNT,
                  MSEED WITHIN VARIETY BY TREATMNT/
READ INPUT DATA
1 1 8  2
1 1 13 4
1 1 9  3
1 3 12 7
1 4 7  3
1 4 11 5
2 1 6  5
2 1 12 3
2 2 12 6
2 2 14 4
3 2 9  6
3 2 7  2
3 3 14 6
3 3 16 8
3 4 10 4
3 4 14 6
3 4 11 5
3 4 13 7
FINISH
```

Figure 1.18b

TESTS OF SIGNIFICANCE FOR WEIGHT USING SEQUENTIAL SUMS OF SQUARES

SOURCE OF VARIATION	SUM OF SQUARES	DF	MEAN SQUARE	F	SIG. OF F
WITHIN+RESIDUAL	4.30000	3	1.43333		
CONSTANT	2178.00000	1	2178.00000	1519.53488	.000
MSEED WITHIN VARIETY BY TREATMNT	92.11472	8	11.51434	8.03326	.057
VARIETY	5.31810	3	1.77270	1.23677	.433
TREATMNT	36.16611	2	18.08306	12.61609	.035
VARIETY BY TREATMNT	.10107	1	.10107	.07051	.808

TESTS OF SIGNIFICANCE FOR WEIGHT USING SEQUENTIAL SUMS OF SQUARES

SOURCE OF VARIATION	SUM OF SQUARES	DF	MEAN SQUARE	F	SIG. OF F
WITHIN+RESIDUAL	4.30000	3	1.43333		
CONSTANT	2178.00000	1	2178.00000	1519.53488	.000
TREATMNT	10.50000	2	5.25000	3.66279	.157
VARIETY	36.78571	3	12.26190	8.55482	.056
VARIETY BY TREATMNT	34.71429	2	17.35714	12.10963	.037
MSEED WITHIN VARIETY BY TREATMNT	51.70000	7	7.38571	5.15282	.103

1.19 Randomized Block Designs

In this design the experimental unit is divided into groups (blocks). The main object of this is to keep the experimental errors within each group as small as possible. The accuracy of the experiment is increased by making comparisons within the resulting relatively homogeneous experimental units. The model for this design is

$$Y_{ij} = \mu + \beta_i + \tau_j + \epsilon_{ij}$$

where β_i is the block effect and τ_j is the treatment effect.

1.20 Complete Randomized Block Designs

A randomized block design is called complete if each block contains every level of the treatment. Table 1.20 is an example of a complete randomized block design with four treatments, A, B, C, and D, and three blocks.

Table 1.20

	Block		
	1	2	3
	A	D	A
	B	B	C
	C	A	B
	D	C	D

Let Y, TRT, and BLK be the response, treatment, and block variables respectively. The MANOVA commands needed to perform the analysis of this design are

```
MANOVA          Y BY BLK(1,3) TRT(1,4)/
                DESIGN=BLK,TRT/
```

In most applications the significance of the block differences is assumed, and treatment effects are corrected for the block effects. (Although it does not make any difference here since the design is balanced and complete, in general the treatment effects should be adjusted.)

1.21 Balanced Incomplete (Randomized) Block Designs (BIB)

In some randomized block designs it may not be possible to apply all treatments in every block. If the block size is less than the number of treatments, the design is called incomplete. An incomplete block design is called balanced if

- Each block contains exactly k treatments
- Each treatment appears in r blocks
- Any pair of treatments appears together λ times

Thus a BIB can be described in terms of the parameters t (number of treatments), b (number of blocks), k, r, and λ.

Table 1.21 is an example of a BIB design with $t = 4$, $b = 4$, $k = 3$, $r = 3$, and $\lambda = 2$.

Table 1.21

	Block			
	1	2	3	4
	A	D	A	B
	B	B	D	C
	C	A	C	D

The following example is taken from Cochran and Cox (1957, p. 443). It is a BIB design with $t = 6$, $b = 15$, $k = 2$, $r = 5$, and $\lambda = 1$. The blocks are grouped into 5 replications.

The SPSS commands for this analysis are given in Figure 1.21a. The first design model specification requests the blocks within replications adjusted for treatment effects. The second model asks for the treatment effects adjusted for the blocks.

Figure 1.21a

```
RUN NAME        TYPE 1 BALANCED INCOMPLETE BLOCK DESIGN.
COMMENT         DATA ARE TAKEN FROM COCHRAN & COX(1957).
VARIABLE LIST   REPLICS, TREATMNT, BLOCKS, DEP
INPUT MEDIUM    CARD
INPUT FORMAT    FIXED(3F1.0,F2.0)
N OF CASES      UNKNOWN
MANOVA          DEP BY REPLICS(1,5), TREATMNT(1,6), BLOCKS(1,3)/
                DESIGN = REPLICS, TREATMNT, BLOCKS W REPLICS/
                DESIGN = REPLICS, BLOCKS W REPLICS, TREATMNT/
READ INPUT DATA
111 7
12117
13226
14225
 ....
 ....
 ....
54326
55332
56127
END INPUT DATA
FINISH
```

The ANOVA tables from the output for Figure 1.21a are given in Figure 1.21b.

Figure 1.21b

TESTS OF SIGNIFICANCE FOR DEP USING SEQUENTIAL SUMS OF SQUARES

SOURCE OF VARIATION	SUM OF SQUARES	DF	MEAN SQUARE	F	SIG. OF F
RESIDUAL	77.33333	10	7.73333		
CONSTANT	19712.03333	1	19712.03333	2548.96983	0.0
REPLICS	298.46667	4	74.61667	9.64871	.002
TREATMNT	1059.76667	5	211.95333	27.40776	.000
BLOCKS W REPLICS	213.40000	10	21.34000	2.75948	.062

TESTS OF SIGNIFICANCE FOR DEP USING SEQUENTIAL SUMS OF SQUARES

SOURCE OF VARIATION	SUM OF SQUARES	DF	MEAN SQUARE	F	SIG. OF F
RESIDUAL	77.33333	10	7.73333		
CONSTANT	19712.03333	1	19712.03333	2548.96983	0.0
REPLICS	298.46667	4	74.61667	9.64871	.002
BLOCKS W REPLICS	753.00000	10	75.30000	9.73707	.001
TREATMNT	520.16667	5	104.03333	13.45259	.000

1.22 Partially Balanced Incomplete Block Designs (PBIB)

Because balanced incomplete block designs often require a large number of blocks, it may not be possible to find a design that fits the size of the experiment. A general class of BIB designs that do not have the uniform variances for treatment contrasts but still permit the estimation of treatment differences are the partially balanced incomplete block designs.

Consider the design in Table 1.22, with $t = 20$, $k = 4$, $r = 2$ and $b = 10$. Recall that for a BIB design any pair of treatments must appear together λ times. In this design, some treatments occur together in the same blocks and some do not. This is the main difference between BIB and PBIB designs.

Table 1.22

Blocks

1	2	3	4	5	6	7	8	9	10
A	M	E	Q	I	A	B	C	D	E
B	N	F	R	J	K	L	M	N	O
C	O	G	S	K	F	G	H	I	J
D	P	H	T	L	P	Q	R	S	T

PBIB designs represent a large class of designs, many of which can be found in Cochran and Cox (1957). An example with t = 15, b = 15, k = 4, and r = 4 is given on p. 456 of that text. The MANOVA commands and the output ANOVA table are given in Figure 1.22a and Figure 1.22b.

Figure 1.22a

```
RUN NAME        15 X 15 PARTIALLY BAL. INC. BLOCK DESIGN.
COMMENT         DATA ARE TAKEN FROM COCHRAN & COX(1957) P.456.
VARIABLE LIST   BLOCKS, TREATMNT, DEP
INPUT MEDIUM    CARD
N OF CASES      UNKNOWN
INPUT FORMAT    FIXED(2F2.0,8X,F3.0)
MANOVA          DEP BY BLOCKS(1,15), TREATMNT(1,15)/
                DESIGN = BLOCKS, TREATMNT/

READ INPUT DATA
 1 1        2.6
 1 9        2.5
 113        2.0
 115        2.4
 2 1        2.7
  ..         .
  ..         .
  ..         .
15 4        2.5
15 8        3.2
1510        2.4
1511        3.1
END INPUT DATA
FINISH
```

Figure 1.22b

TESTS OF SIGNIFICANCE FOR DEP USING SEQUENTIAL SUMS OF SQUARES

SOURCE OF VARIATION	SUM OF SQUARES	DF	MEAN SQUARE	F	SIG. OF F
RESIDUAL	2.68589	31	.08664		
CONSTANT	448.26654	1	448.26654	5173.80518	0.0
BLOCKS	4.92333	14	.35167	4.05887	.001
TREATMNT	1.56411	14	.11172	1.28948	.268

1.23 Latin and Other Squares

A Latin square is a design in which each treatment appears exactly once in each row and column. The main interest is still on the estimation of treatment differences, but two restrictions are put on the randomization of the treatment assignment. The model of this design is

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \epsilon_{ijk}$$

where α_i, β_j and γ_k are the row, column and treatment effects respectively. An example of a 4 × 4 Latin square is shown in Table 1.23a.

Table 1.23a

		Column 1	2	3	4
Row	1	A	C	D	B
	2	D	B	C	A
	3	C	A	B	D
	4	B	D	A	C

The following MANOVA specifications may be used to analyze a 4 × 4 Latin square.

```
MANOVA          Y BY ROW(1,4),COL(1,4),TRT(1,4)/
                DESIGN=ROW,COL,TRT/
```

If another restriction on the randomization is placed on a Latin square, we have a Graeco-Latin square. Table 1.23b exhibits a 4 × 4 Graeco-Latin square.

Table 1.23b

		Column			
		1	2	3	4
Row	1	A δ	B α	D β	C γ
	2	B γ	A β	C α	D δ
	3	C β	D γ	B δ	A α
	4	D α	C δ	A γ	B β

In this design the third restriction has levels α, β, γ, δ. Note that α, β, γ and δ not only each appear exactly once within each row and column, but they also appear exactly once with each level of treatments A, B, C, D. The Graeco-Latin square can be constructed by superimposing an orthogonal (same size) Latin square on the original Latin square. In other words, the third restriction factor along with column and row is also a 4 × 4 Latin square. It has treatments α, β, γ, δ and is orthogonal to the original Latin square with treatments A, B, C, and D. Here orthogonality means each letter in one Latin square appears exactly once in the same position as each letter of the other square.

The analysis of variance for a Graeco-Latin square is very similar to that for a Latin square. Let GREEK denote the third restriction factor on a 4 × 4 Graeco-Latin square. The MANOVA specifications would be

```
MANOVA          Y BY ROW(1,4), COL(1,4), GREEK(1,4), TRT(1,4)/
                DESIGN=ROW,COL,GREEK,TRT/
```

Note that a small Graeco-Latin square design may not be very practical, since very few degrees of freedom are left for the residual.

1.24 Factorial Designs

In a factorial design, the effects of several different factors are investigated simultaneously. Suppose we wish to study the effects of two factors on the yield of a chemical. The first factor is temperature at 100°F, 200°F, and 300°F. The other factor is pressure at 20 psi and 40 psi. This experiment is a two-factor factorial design with three levels for the first factor and two levels for the second. The treatments for this experiment are the 6 combinations of the levels of the factors. The model for the 3 × 2 factorial design is

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$$

where α_i is the temperature effect, β_j is the pressure effect, and $(\alpha\beta)_{ij}$ is the temperature-pressure interaction.

A factorial experiment containing one observation per cell (treatment) constitutes one replicate of the design. The design may be replicated k times in two possible ways. If each observation has different experimental conditions for replications within cells (e.g., each replicate is a block), the design is crossed by another factor within k levels (i.e., block effect). If the experimental condition is the same for the replications within cells, the number of factors remains unchanged, and the variation within cells is attributed to the error.

The following example illustrates the use of MANOVA to perform the analysis of a 4 × 4 × 3 factorial in randomized blocks (two blocks) with a covariate. The data are taken from Cochran and Cox (1957, p. 176).

The model contains the main effects (NTREAT, LENPER, CURRENT), all two-way interactions (NTREAT BY LENPER,...,LENPER BY CURRENT), and the three-way interaction (NTREAT BY LENPER BY CURRENT). The SPSS commands for this analysis are shown in Figure 1.24a and the analysis of variance table in Figure 1.24b.

Figure 1.24a

```
RUN NAME        4*4*3 FACTORIAL IN RANDOMIZED BLOCKS.
COMMENT         4*4*3 FACTORIAL IN RANDOMIZED BLOCKS WITH
                COVARIATE. FROM COCHRAN AND COX(1957)  PAGE 176.
VARIABLE LIST   REPLIC,LENPER,CURRENT,NTREAT,Y,X
INPUT MEDIUM    CARD
INPUT FORMAT    FIXED(4F1.0,F2.0,F3.0)
N OF CASES      96
IF              (LENPER EQ 5) LENPER = 4
IF              (NTREAT EQ 3) NTREAT = 2
IF              (NTREAT EQ 6) NTREAT = 3
MANOVA          Y BY REPLIC(1,2), LENPER(1,4), CURRENT(1,4),
                     NTREAT(1,3) WITH X/
                DESIGN = REPLIC,NTREAT,LENPER,CURRENT,NTREAT BY LENPER,
                         NTREAT BY CURRENT, LENPER BY CURRENT,
                         NTREAT BY LENPER BY CURRENT/

READ INPUT DATA
111172152
111374131
111669131
112161130
112361129
112665126
113162141
113365112
113670111
114185147
114376125
114661130
121167136
121352110
121662122
122160111
 . . . .
 . . . .
 . . . .
254159102
254358 98
254688135
FINISH
```

Figure 1.24b

TESTS OF SIGNIFICANCE FOR Y USING SEQUENTIAL SUMS OF SQUARES

SOURCE OF VARIATION	SUM OF SQUARES	DF	MEAN SQUARE	F	SIG. OF F
RESIDUAL	2211.96526	46	48.08620		
REGRESSION	987.52432	1	987.52432	20.53654	.000
CONSTANT	1316.19933	1	1316.19933	27.37166	0.0
REPLIC	.27456	1	.27456	.00571	.940
NTREAT	441.20522	2	220.60261	4.58765	.015
LENPER	180.52285	3	60.17428	1.25138	.302
CURRENT	2111.03300	3	703.67767	14.63367	0.0
NTREAT BY LENPER	211.79056	6	35.29843	.73407	.625
NTREAT BY CURRENT	467.84848	6	77.97475	1.62156	.163
LENPER BY CURRENT	404.37365	9	44.93041	.93437	.505
NTREAT BY LENPER BY CURRENT	1021.61800	18	56.75656	1.18031	.315

1.25 Nested Designs

A nested design arranges the experimental units hierarchically. For example, consider an experiment to compare the yield of wheat per acre for different areas in a given state. Five counties are selected at random, then three townships are randomly selected from each county. From each township two farms are selected and the yield of wheat per acre is obtained. The resulting experiment produces $5 \times 3 \times 2 = 30$ experimental units. The factors of this experiment are county and township, and the township effects are *nested* under the county factor, since a given township appears only under one of the five counties. In other words, the county factor is not *crossed* with township factor and so the interaction between county and township is not estimable.

The model for this two-factor nested design is

$$Y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \epsilon_{ijk}$$

where α_i is the county effect and $\beta_{j(i)}$ is the township effect nested under the county effect.

Since α_i should be tested against variation within α_i, i.e., $\beta_{j(i)}$, the following MANOVA specfications can be used:

```
MANOVA          Y BY COUNTY(1,5),TOWN(1,3)/
                DESIGN=COUNTY VS 1, TOWN WITHIN COUNTY=1 VS WITHIN/
```

Note that the first keyword WITHIN (or just W) indicates nesting. The DESIGN specification requests that COUNTY be tested against the error 1 term, which is the effect of TOWN (nested within COUNTY), and that the within-cells error term (second WITHIN) be used for testing the TOWN effect.

When crossing and nesting are both used in the design, attention must be paid to the choice of appropriate error terms for testing the various effects. Consider a three-factor example, with factors A, B, and C. If

1 C is nested within B, and B is nested within A, the model is

$$Y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \gamma_{k(ij)} + \epsilon_{ijk}$$

The DESIGN specification should be

```
DESIGN=A VS 1, B W A=1 VS 2, C W B W A=2 VS WITHIN/
```

2 C is nested within B, and B is crossed with A, the model is

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_{k(j)} + (\alpha\gamma)_{ik(j)} + \epsilon_{ijk}$$

The rule for writing down the model is that no interaction in which the subscript j appears twice is in the model. For example, interactions $(\beta\gamma)_{jk(j)}$ and $(\alpha\beta\gamma)_{ijk(j)}$ do not exist.

Since C is nested within B, β_j is tested against $\gamma_{k(j)}$. The appropriate error term for α_i and $(\alpha\beta)_{ij}$ is the residual of the A-B two-way table, $(\alpha\gamma)_{ik(j)}$, which is the interaction effect of α_i and $\gamma_{k(j)}$. If the number of observations per cell is greater than one, then $(\alpha\gamma)_{ik(j)}$ and $\gamma_{k(j)}$ can be tested against the within-cells error term. The DESIGN specification for this model is

```
DESIGN=A VS 2, B VS 1, C W B=1 VS WITHIN,
       A BY B VS 2, A BY C W B=2 VS WITHIN/
```

3 C is crossed with B, and B is nested within A. The model and the DESIGN specification are the same as those in (2) except for the names of the effects.

An experiment (Hicks, 1973, p. 195) was conducted to compare a new gun-loading method with the existing one (factor METHOD). Three teams were chosen randomly from each of three groups. Each team used the two methods of gun loading in random order. The data and SPSS commands for the analysis are as given in Figure 1.25a, and the ANOVA table is presented in Figure 1.25b.

Figure 1.25a

```
RUN NAME        NESTED DESIGN.
COMMENT         DATA ARE TAKEN FROM HICKS(1973) P. 194.
COMMENT         METHOD : 2 LEVELS CROSSED WITH GROUP.
COMMENT         GROUP  : 3 LEVELS.
COMMENT         TEAM   : 3 LEVELS NESTED WITHIN GROUP.
COMMENT         NUMBER OF OBSERVATIONS PER CELL = 2.
VARIABLE LIST   GROUP,TEAM,METHOD,RESP
N OF CASES      UNKNOWN
INPUT FORMAT    FREEFIELD
INPUT MEDIUM    CARD
MANOVA          RESP BY METHOD(1,2),GROUP,TEAM(1,3)/
                DESIGN=METHOD VS 1,GROUP VS 2,
                METHOD BY GROUP VS 1,TEAM W GROUP=2,
                METHOD BY TEAM W GROUP=1/
READ INPUT DATA
1 1 1 20.2
1 1 1 24.1
1 1 2 14.2
1 1 2 16.2
1 2 1 26.2
1 2 1 26.9
  . .  .
  . .  .
  . .  .
3 3 1 21.8
3 3 1 23.5
3 3 2 12.7
3 3 2 15.1
END INPUT DATA
FINISH
```

Figure 1.25b

TESTS OF SIGNIFICANCE FOR RESP USING SEQUENTIAL SUMS OF SQUARES

SOURCE OF VARIATION	SUM OF SQUARES	DF	MEAN SQUARE	F	SIG. OF F
WITHIN CELLS	41.58993	18	2.31055		
CONSTANT	13455.99443	1	13455.99443	5823.71557	0.0
ERROR 1	10.72164	6	1.78694		
METHOD	651.95062	1	651.95062	364.84200	0.0
METHOD BY GROUP	1.18721	2	.59361	.33219	.730
ERROR 2	39.25829	6	6.54305		
GROUP	16.05166	2	8.02583	1.22662	.358

1.26 Confounding Designs

In some factorial designs it may not be possible to apply all factor combinations in every block. Two methods can be used to handle this problem. The first one is the BIB designs discussed in Section 1.21. Another method for circumventing this difficulty is to reduce the size of a block by sacrificing the estimation of certain higher-order interactions. Consider a $2 \times 2 \times 2$ factorial experiment, with factors A, B, and C. Let abc denote the experimental unit with all three factors at the high level (since each factor has two levels, one is low and one is high), ab denote the unit where A and B are at the high level and c is at the low level. Thus if a letter appears, that factor is at the high level; otherwise, it is at the low level. When all factors appear at the low level it is designated by (1). Suppose we arrange the $2 \times 2 \times 2$ factorial in two blocks as in Table 1.26a.

Table 1.26a

Block	
1	2
abc	ab
a	ac
b	bc
c	(1)

The effect of A is estimated by comparing the observations receiving high and low levels of A, i.e.,

abc + a + ab + ac − b − c − bc − (1)

and so on.

Note that the ABC interaction is estimated from the comparison

abc + a + b + c − ab − ac − bc − (1)

which is the same as the difference between blocks 1 and 2. Hence we cannot distinguish between the block effects and the ABC interaction. The ABC interaction is said to be *confounded* with the block effect.

If this experiment were replicated four times, the layout might be as shown in Table 1.26b.

Table 1.26b

Replication 1 Block		Replication 2 Block		Replication 3 Block		Replication 4 Block	
1	2	1	2	1	2	1	2
abc	ab	abc	(1)	bc	b	c	ab
a	ac	b	bc	ac	a	b	ac
b	bc	a	ac	ab	c	abc	(1)
c	(1)	c	ab	(1)	abc	a	bc

Since the confounded effect (ABC) is the same for all four replications, ABC is *completely* confounded with blocks. The MANOVA specifications needed for this example are

```
MANOVA      Y BY REPLIC(1,4), BLOCK(1,2), A, B, C(1,2)/
            DESIGN=REPLIC,BLOCK W REPLIC, A, B, C, A BY B, A BY C, B BY C/
```

Note that the model does not include A BY B BY C, which is confounded with BLOCK W REPLIC.

It is possible to test the ABC interaction if some interaction other than ABC is confounded in some of the replications. One possible layout would be that given in Table 1.26c.

Table 1.26c

Replication 1		Replication 2		Replication 3		Replication 4	
Block		Block		Block		Block	
1	2	1	2	1	2	1	2
abc	ab	b	ab	ac	ab	ac	a
a	ac	a	c	(1)	bc	ab	bc
b	bc	ac	(1)	abc	a	b	abc
c	(1)	bc	abc	b	c	c	(1)

In replication 1, ABC is confounded with blocks. In replication 2, the AB interaction is confounded with blocks. For replications 3 and 4, AC and BC are confounded.

For this example, A, B, and C are free of the block effects and three-fourths information for AB, AC, BC, and ABC can be obtained, since the unconfounded interactions can be estimated in three out of four of the replications. Hence we say AB, AC, BC, and ABC are *partially* confounded with blocks. The MANOVA specifications for this 2 × 2 × 2 factorial with partial confounding are

```
MANOVA          Y BY REPLIC(1,4), BLOCK(1,2),A, B, C(1,2)/
                DESIGN=REPLIC,BLOCK W REPLIC, A, B, C, A BY B, A BY C,
                       B BY C, A BY B BY C/
```

More complex confounding designs can be found in Davies (1954) and Cochran and Cox (1957).

Another Example The following example is taken from Cochran and Cox (1957, p. 205). The data are a 3 × 3 × 2 factorial in blocks of six units with three blocks in each of four replications. Interactions AB and ABC are partially confounded with blocks. The SPSS commands for this analysis are given in Figure 1.26a.

Figure 1.26a

```
RUN NAME        CONFOUNDING IN MIXED SERIES.
COMMENT         CONFOUNDING IN MIXED SERIES. 3*3*2 FACTORIAL
COMMENT         FROM COCHRAN AND COX(1957)  P. 205
COMMENT         SECOND ANALYSIS GIVES AB TWO-WAY TABLE ADJUSTED FOR BLOCK
COMMENT         THIRD ANALYSIS GIVES AC TWO-WAY TABLE ADJUSTED FOR BLOCKS
COMMENT
COMMENT         FACTOR A : 8-8-6 FERTILIZER APPLIED IN THE ROW,
COMMENT                    3 LEVELS -- 0 (NONE), 1 (200 LB.), 2 (400 LB.)
COMMENT         FACTOR B : MEALS, 3 LEVELS -- 0 (NONE), 1 (TUNG MEAL),
COMMENT                                       2 (COTTONSEED MEAL).
COMMENT         FACTOR C : 8-8-6 FERTILIZER APPLIED AS SIDE-DRESSING,
COMMENT                    2 LEVELS -- 0 (NONE), 1 (200 LB.).
VARIABLE LIST   REPLICS,BLOCKS,A,B,C,DEP
INPUT MEDIUM    CARD
INPUT FORMAT    FIXED(2X,5F1.0,8X,F3.0)
N OF CASES      72
MANOVA          DEP BY REPLICS(1,4),BLOCKS(1,3),A(0,2),B(0,2),C(0,1)/
                DESIGN = REPLICS,BLOCKS WITHIN REPLICS,A,B,C,
                         A BY B,A BY C, B BY C, A BY B BY C/
                DESIGN = REPLICS,BLOCKS W REPLICS,CONSPLUS A AND B/
                DESIGN = REPLICS,BLOCKS W REPLICS,CONSPLUS A AND C/
READ INPUT DATA
  11011          82
  11020          70
  11100          80
  11121          86
  11201          74
  11210          86
  12001          67
  12010          55
   ....           .
   ....           .
   ....           .
  42210          66
  43001          90
  43010          58
  43100          81
  43121          67
  43211          68
  43220          56
FINISH
```

The first DESIGN specification requests an analysis of variance for this experiment (Figure 1.26b).

Figure 1.26b

TESTS OF SIGNIFICANCE FOR DEP USING SEQUENTIAL SUMS OF SQUARES

SOURCE OF VARIATION	SUM OF SQUARES	DF	MEAN SQUARE	F	SIG. OF F
RESIDUAL	8909.36190	43	207.19446		
CONSTANT	461120.05556	1	461120.05556	2225.54237	0.0
REPLICS	3836.61111	3	1278.87037	6.17232	.001
BLOCKS WITHIN REPLICS	2836.33333	8	354.54167	1.71115	.123
A	1116.02778	2	558.01389	2.69319	.079
B	253.69444	2	126.84722	.61221	.547
C	868.05556	1	868.05556	4.18957	.047
A BY B	1129.34921	4	282.33730	1.36267	.263
A BY C	2995.02778	2	1497.51389	7.22758	.002
B BY C	423.52778	2	211.76389	1.02205	.368
A BY B BY C	1015.95556	4	253.98889	1.22585	.314

The second and third analyses give the AB and AC two-way means adjusted for the block effects (Figure 1.26c). For more information about the use of CONSPLUS to obtain marginal means and summary tables, see Section 1.50.

Figure 1.26c

CONSPLUS A AND B

PARAMETER	COEFF.	STD. ERR.	T-VALUE	SIG. OF T	LOWER .95 CL	UPPER .95 CL
12	72.1964285714	6.02764	11.97757	0.0	60.10109	84.29176
13	73.2261904762	6.02764	12.14841	0.0	61.13086	85.32152
14	79.7023809524	6.02764	13.22283	0.0	67.60705	91.79771
15	86.7738095238	6.02764	14.39600	0.0	74.67848	98.86914
16	87.8035714286	6.02764	14.56684	0.0	75.70824	99.89891
17	79.4226190476	6.02764	13.17641	0.0	67.32729	91.51795
18	89.0297619048	6.02764	14.77026	0.0	76.93443	101.12510
19	74.3452380952	6.02764	12.33406	0.0	62.24990	86.44057
20	77.7500000000	6.02764	12.89892	0.0	65.65467	89.84533

CONSPLUS A AND C

PARAMETER	COEFF.	STD. ERR.	T-VALUE	SIG. OF T	LOWER .95 CL	UPPER .95 CL
12	62.5833333333	4.21611	14.84385	0.0	54.13406	71.03261
13	87.5000000000	4.21611	20.75372	0.0	79.05073	95.94927
14	84.3333333333	4.21611	20.00264	0.0	75.88406	92.78261
15	85.0000000000	4.21611	20.16076	0.0	76.55073	93.44927
16	82.7500000000	4.21611	19.62709	0.0	74.30073	91.19927
17	78.0000000000	4.21611	18.50046	0.0	69.55073	86.44927

1.27 Split-plot Designs

In many factorial designs, it may not be possible to completely randomize the assignment of treatments to the experimental unit. Consider, for example, an experiment to compare three varieties of wheat (factor A) and two different types of fertilizer (factor B). Three locations are selected as blocks. Three levels of A are randomly assigned to plots of equal area within each block. After A is assigned, each plot is "split" into halves (called subplots) to receive the random assignment of B. What is the difference between a complete 3 × 2 factorial and the 3 × 2 split-plot design? In a 3 × 2 factorial, each block is divided into six subplots to receive the random assignment of treatment combinations of A and B. In the split-plot design, two treatment combinations that have the same level of A are always in the same plot. If the subplot is considered the experimental unit, the plot is a "small" block of size 2. The differences among these "small" blocks are the differences between levels of A, since the main effects of A are confounded. A split-plot design is a design in which certain main effects are confounded.

Intuitively, the variation of plots within A should be used as the error term to test for the main effects of A. The effects of plot within A can be partitioned into two parts. One is the block effects and another is the block and A interaction. Thus the model for a split-plot design is

$$Y_{ijk} = \mu + \alpha_i + \beta_k + (\alpha\beta)_{ik} + \gamma_j + (\alpha\gamma)_{ij} + \epsilon_{ijk}$$

where α_i is the A effect, β_k is the block effect, $(\alpha\beta)_{ik}$ is the interaction of A and block and is the error term for testing A, γ_j is the B effect, $(\alpha\gamma)_{ij}$ is the AB interaction, and ϵ_{ijk} is the residual used as the error term for testing B and AB.

Another model is

$$Y_{ijk} = \mu + \alpha_i + \beta_k + (\alpha\beta)_{ik} + \gamma_j + (\alpha\gamma)_{ij} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \epsilon_{ijk}$$

$(\beta\gamma)_{jk}$ is the error term for γ_j, $(\alpha\beta\gamma)_{ijk}$ is the error term for $(\alpha\gamma)_{ij}$, and if the number of observations per cell is greater than 1, then $(\alpha\beta)_{ik}$, $(\beta\gamma)_{jk}$ and $(\alpha\beta\gamma)_{ijk}$ can be tested against the within-cells error.

The following MANOVA specifications may be used to perform an analysis of variance of a 3 × 2 split-plot design:

```
MANOVA          Y BY BLOCK(1,3), A(1,3), B(1,2)/
                DESIGN=BLOCK, A VS 1, A BY BLOCK=1, B, A BY B/
```

In the above DESIGN specification, effect A is tested against the error 1 term which is the interaction of A and BLOCK. Effects B and AB are tested against the residual, since there is no within-cells error in this example.

This type of design can be extended by subdividing each subplot into sub-subplots, etc. The model for a split-split-plot design would be

$$Y_{ijkl} = \mu+\alpha_i+\beta_l+(\alpha\beta)_{il}+\gamma_j+(\alpha\gamma)_{ij}+\pi_{ijl}+\delta_k+(\alpha\delta)_{ik}+(\gamma\delta)_{jk}+(\alpha\gamma\delta)_{ijk}+\epsilon_{ijkl}$$

where δ_k is the effect for the sub-subplot factor, π_{ijl} = subplot residual = $(\gamma\beta)_{jl}$ + $(\alpha\gamma\beta)_{ijl}$, ϵ_{ijkl} is the residual, and $(\alpha\beta)_{il}$, π_{ijl}, and ϵ_{ijkl} are the appropriate error terms for plot, subplot, and sub-subplot factors, respectively.

An example of a split-split-plot design is taken from Hicks (1973, p. 223). The SPSS commands are given in Figure 1.27a.

Figure 1.27a

```
RUN NAME        SPLIT-SPLIT-PLOT DESIGN.
COMMENT         DATA ARE TAKEN FROM HICKS(1973) PAGE 223.
COMMENT         LAB : THREE DIFFERENT LABORATORIES--PLOT FACTOR.
COMMENT         TEM : THREE LEVELS OF TEMPERATURE--SUB-PLOT FACTOR.
COMMENT         MIX : THREE TYPES OF MIX--SUB-SUB-PLOT FACTOR.
COMMENT         FOUR REPLICATES (BLOCK).
VARIABLE LIST   BLOCK LAB TEM MIX RESP
N OF CASES      UNKNOWN
INPUT FORMAT    FIXED(4F1.0,1X,F4.1)
INPUT MEDIUM    CARD
MANOVA          RESP BY BLOCK(1,4),LAB,TEM,MIX(1,3)/
                DESIGN=BLOCK,LAB VS 1,LAB BY BLOCK=1,
                       TEM VS 2,LAB BY TEM VS 2,TEM BY BLOCK+
                       LAB BY TEM BY BLOCK=2,
                       MIX,LAB BY MIX,TEM BY MIX,LAB BY TEM BY MIX/
READ INPUT DATA
1111 18.6
1112 14.5
1113 21.1
1121  9.5
1122  7.8
1123 11.2
1131  5.4
1132  5.2
1133  6.3
1211 20.0
1212 18.4
1213 22.5
   .
   .
   .
4321  9.5
4322  9.0
4323 11.4
4331  4.8
4332  5.4
4333  5.8
END INPUT DATA
FINISH
```

As can be seen from the DESIGN specification, the interaction of LAB and BLOCK is the error term for the plot factor LAB, the interaction of TEM and BLOCK and the interaction of LAB, TEM, and BLOCK are pooled together as the error term for the subplot factors. The sub-subplot factors are to be tested against the residual. The analysis of variance from the output for this run is shown in Figure 1.27b.

Figure 1.27b

TESTS OF SIGNIFICANCE FOR RESP USING SEQUENTIAL SUMS OF SQUARES

SOURCE OF VARIATION	SUM OF SQUARES	DF	MEAN SQUARE	F	SIG. OF F
RESIDUAL	13.40499	54	.24824		
CONSTANT	14697.66168	1	14697.66168	59207.32736	0.0
BLOCK	9.41435	3	3.13812	12.64143	0.0
MIX	145.71785	2	72.85893	293.50127	0.0
LAB BY MIX	.33926	4	.08482	.34167	.849
TEM BY MIX	43.68696	4	10.92174	43.99659	0.0
LAB BY TEM BY MIX	1.07740	8	.13467	.54252	.819
ERROR 1	16.10982	6	2.68497		
LAB	40.66356	2	20.33178	7.57244	.023
ERROR 2	9.88335	18	.54907		
TEM	3119.50650	2	1559.75325	2840.69330	0.0
LAB BY TEM	4.93650	4	1.23412	2.24764	.104

1.28 Analysis of Carry-over Effects

If different treatments are applied in sequence to the same unit, residual or carry-over effects may be present in the experiment. By including dummy factors, MANOVA enables the user to perform an analysis of variance with residual effects.

The following example is taken from Cochran and Cox (1957, p. 133). The experiment compares three feeding methods (A, B, and C) on the milk yield of dairy cows. The experiment consists of two 3 × 3 Latin squares. The rows of the squares represent the successive periods of application, while the columns represent the cows. The data are as follows:

	Square 1			Square 2		
	Cow 1	Cow 2	Cow 3	Cow 4	Cow 5	Cow 6
Period 1	A(38)	B(109)	C(124)	A(86)	B(75)	C(101)
Period 2	B(25)	C(86)	A(72)	C(76)	A(35)	B(63)
Period 3	C(15)	A(39)	B(27)	B(46)	C(34)	A(1)

In addition to the direct (treatment) effects τ_a, τ_b and τ_c, the treatments also contain the residual effects r_a, r_b, and r_c for the period immediately following the one in which they are applied. Thus for cow 2 in the third period, the expected total treatment effect is $\tau_a + r_c$, since A is applied in this period and C in the preceding period. Similarly, the expected total treatment effect is $\tau_a + r_b$ for cow 2 in the second period.

If we let CEFFECT be the (dummy) factor of residual effects and assign

CEFFECT = 1 if no residual effects
2 if r_a is the residual effect
3 if r_b is the residual effect
4 if r_c is the residual effect

then the values of CEFFECT in this example would be

	Square 1			Square 2		
	Cow 1	Cow 2	Cow 3	Cow 4	Cow 5	Cow 6
Period 1	1	1	1	1	1	1
Period 2	2	3	4	2	3	4
Period 3	3	4	2	4	2	3

If the effects of CEFFECT are divided into groups using the following contrasts:

```
( 1   1   1   1 )
( 3  -1  -1  -1 )
( 0   2  -1  -1 )
( 0   0   1  -1 )
```

and the pooled effect of second and third contrasts is CEFFECT(2), then CEFFECT(2) can be used to obtain a test of $r_a = r_b = r_c$. Since the second contrast (0, 2, -1, -1) specifies a test on $r_a = (r_b + r_c)/2$, and the third contrast (0, 0, 1, -1) a test of $r_b = r_c$, jointly they test the hypothesis $r_a = r_b = r_c$.

The above can be done by using the following MANOVA specifications.

```
CONTRAST(CEFFECT)=SPECIAL(1 1 1 1, 3 -1 -1 -1,
                          0 2 -1 -1,0 0 1 -1)/
PARTITION (CEFFECT)=(1,2)/
```

The CONTRAST subcommand indicates the contrast coefficients for factor CEFFECT. The PARTITION subcommand divides the CEFFECT factor into 2 groups for the contrasts. The first group has one degree of freedom with the contrast (3, -1, -1, -1). The second group (CEFFECT(2)) corresponds to the second and third contrasts lumped together and has two degrees of freedom.

The complete MANOVA command file is given in Figure 1.28a.

Figure 1.28a

```
RUN NAME        ANALYSIS OF VARIANCE WITH CARRY-OVER EFFECTS.
COMMENT         DATA IS TAKEN FROM COCHRAN AND COX(1957) PAGE 135.
                CEFFECT REPRESENTS THE CARRY-OVER EFFECTS.
                       CEFFECT=1 IF NO RESIDUAL EFFECTS.
                               2 IF RESIDUAL EFFECT A.
                               3 IF RESIDUAL EFFECT B.
                               4 IF RESIDUAL EFFECT C.
VARIABLE LIST   PERIOD,COW,SQUARE,TREATMNT,CEFFECT,DEP
INPUT MEDIUM    CARD
INPUT FORMAT    FIXED(2X,5F1.0,F10.0)
N OF CASES      18
MANOVA          DEP BY PERIOD(1,3),COW(1,6),SQUARE(1,2),
                       TREATMNT(1,3), CEFFECT(1,4)/

                CONTRAST(CEFFECT) = SPECIAL(1 1 1 1, 3 -1 -1 -1,
                                            0 2 -1 -1, 0 0 1 -1  )/

                PARTITION(CEFFECT) = (1,2)/

                DESIGN = COW, PERIOD WITHIN SQUARE,
                         CEFFECT(2), TREATMNT/
                DESIGN = COW, PERIOD WITHIN SQUARE,
                         TREATMNT, CEFFECT(2)/
READ INPUT DATA
  11111     38.
  12121     109.
  13131     124.
  14211     86.
  15221     75.
  16231     101.
  21122     25.
  22133     86.
  23114     72.
  24232     76.
  25213     35.
  26224     63.
  31133     15.
  32114     39.
  33122     27.
  34224     46.
  35232     34.
  36213     1.
FINISH
```

In the first DESIGN specification, treatment effects are adjusted for the residual effects, and the converse holds in the second DESIGN specification. The ANOVA summary tables for both models are given in Figure 1.28b.

Figure 1.28b

TESTS OF SIGNIFICANCE FOR DEP USING SEQUENTIAL SUMS OF SQUARES

SOURCE OF VARIATION	SUM OF SQUARES	DF	MEAN SQUARE	F	SIG. OF F
RESIDUAL	199.25000	4	49.81250		
CONSTANT	61483.55556	1	61483.55556	1234.29974	0.0
COW	5781.11111	5	1156.22222	23.21149	.005
PERIOD WITHIN SQUARE	11489.11111	4	2872.27778	57.66179	.001
CEFFECT(2)	38.42222	2	19.21111	.38567	.703
TREATMNT	2854.55000	2	1427.27500	28.65295	.004

TESTS OF SIGNIFICANCE FOR DEP USING SEQUENTIAL SUMS OF SQUARES

SOURCE OF VARIATION	SUM OF SQUARES	DF	MEAN SQUARE	F	SIG. OF F
RESIDUAL	199.25000	4	49.81250		
CONSTANT	61483.55556	1	61483.55556	1234.29974	0.0
COW	5781.11111	5	1156.22222	23.21149	.005
PERIOD WITHIN SQUARE	11489.11111	4	2872.27778	57.66179	.001
TREATMNT	2276.77778	2	1138.38889	22.85348	.006
CEFFECT(2)	616.19444	2	308.09722	6.18514	.060

Note that in this example, the number of observations receiving r_a, r_b, and r_c are equal (4). If the design is not balanced with respect to residual effects, contrast coefficients for unequal numbers of replicates must be used to create the desired residual effects.

1.29 Tukey's Test for Nonadditivity

In factorial designs with only one observation per cell there is no within-cell error and thus no direct estimate of the experimental error. Frequently, the highest-order interaction is assumed to be part of the experimental error and its mean square is used to provide a denominator for *F* tests on the remaining model terms. One method of checking the tenability of this no- interaction assumption is provided by Tukey's test for nonadditivity (Tukey(1949)).

SPSS-MANOVA can perform Tukey's test by using the fact that Tukey's sum of squares for nonadditivity is the linear × linear component of interaction in the metric of the estimates of the main effects (see Winer(1971) page 395). Tukey's test requires two separate runs:

1. The first run obtains main effect parameter estimates using an additive main effects model.
2. The second run uses the parameter estimates from the first run as the metric in polynomial contrasts for the factors; the design specifies a linear × linear single-degree-of-freedom interaction term which actually provides the sum of squares for Tukey's test.

To illustrate this procedure consider the data in Table 1.29 taken from Winer(1971), page 474. These data comprise a 3 × 4 factorial with one observation per cell.

Table 1.29

		B			
		1	2	3	4
A	1	8	12	16	20
	2	2	2	14	18
	3	5	4	9	22

First, estimates of main effects are computed by using the following MANOVA specifications.

```
MANOVA          Y BY A(1,3) B(1,4)/
                PRINT=PARAMETERS(NEGSUM)/
                DESIGN= A, B/
```

The PRINT=PARAMETERS(NEGSUM) results in the printing of the estimate of the last main effect as the negative sum of the previous estimates. The default deviation contrast must be used to get these estimates. Figure 1.29a displays the estimates.

Figure 1.29a

ESTIMATES FOR Y

CONSTANT

PARAMETER	COEFF.	STD. ERR.	T-VALUE	SIG. OF T	LOWER .95 CL	UPPER .95 CL
1	11.0000000000	.84984	12.94366	.000	8.92054	13.07946

A

PARAMETER	COEFF.	STD. ERR.	T-VALUE	SIG. OF T	LOWER .95 CL	UPPER .95 CL
2	3.0000000000	1.20185	2.49615	.047	.05919	5.94081
3	-2.0000000000	1.20185	-1.66410	.147	-4.94081	.94081
4	-1.0000000000	.	.	.	.	.

B

PARAMETER	COEFF.	STD. ERR.	T-VALUE	SIG. OF T	LOWER .95 CL	UPPER .95 CL
4	-6.0000000000	1.47196	-4.07620	.007	-9.60174	-2.39826
5	-5.0000000000	1.47196	-3.39683	.015	-8.60174	-1.39826
6	2.0000000000	1.47196	1.35873	.223	-1.60174	5.60174
7	9.0000000000	.	.	.	.	.

In the second run, orthogonal polynomial contrasts for each factor are requested. The metric for each factor consists of the parameter estimates for that factor's categories produced by the initial run:

```
CONTRAST(A)=POLYNOMIAL(3 -2 -1)/
CONTRAST(B)=POLYNOMIAL(-6 -5 2 9)/
```

Each factor is then partitioned so that the first partition contains the linear component of the orthogonal polynomial contrast:

```
PARTITION(A)/
PARTITION(B)/
```

Lastly, the design specifies a main effects model along with the linear × linear component of the interaction:

```
DESIGN=A, B, A(1) BY B(1)/
```

The resulting ANOVA table appears in Figure 1.29b.

Figure 1.29b

TESTS OF SIGNIFICANCE FOR Y USING SEQUENTIAL SUMS OF SQUARES

SOURCE OF VARIATION	SUM OF SQUARES	DF	MEAN SQUARE	F	SIG. OF F
RESIDUAL	34.33855	5	6.86771		
CONSTANT	1452.00000	1	1452.00000	211.42418	.000
A	56.00000	2	28.00000	4.07705	.089
B	438.00000	3	146.00000	21.25890	.003
A(1) BY B(1)	17.66145	1	17.66145	2.57166	.170

The *F* test for the A(1) BY B(1) interaction is Tukey's test for nonadditivity.

Note that Tukey's test for nonadditivity can be extended to higher-order factorial experiments.

1.30 Simple Effects

The presence of a significant interaction in a two-way design precludes the testing of the main effects. Instead, the effect of one factor differs at each level of the other factor. Frequently one may wish to test the significance of these differential effects. Such tests are generally called tests of simple effects.

Simple effects can be tested in SPSS-MANOVA by using the nesting facility of the DESIGN subcommand. As an example, consider the data presented in Figure 1.2 for which the ANOVA table appears in Figure 1.3a. Here the interaction is significant at the 0.006 level. Simple effects tests are desired to examine the category differences for each of the drugs. The following DESIGN subcommand accomplishes this:

```
DESIGN=DRUG, CAT WITHIN DRUG(1), CAT WITHIN DRUG(2),
             CAT WITHIN DRUG(3)/
```

Here CAT WITHIN DRUG(1) tests the difference in means between category 1 and category 2 for the first level of drug. Similarly, the two successive effects test for category differences for the second and third drugs, respectively. Note that DRUG appears first in the design. This eliminates any confounding effects of CAT. Figure 1.30a presents the output of this design.

Figure 1.30a

TESTS OF SIGNIFICANCE FOR Y USING SEQUENTIAL SUMS OF SQUARES

SOURCE OF VARIATION	SUM OF SQUARES	DF	MEAN SQUARE	F	SIG. OF F
WITHIN CELLS	106.00000	12	8.83333		
CONSTANT	882.00000	1	882.00000	99.84906	0.0
DRUG	48.00000	2	24.00000	2.71698	.106
CAT WITHIN DRUG(1)	54.00000	1	54.00000	6.11321	.029
CAT WITHIN DRUG(2)	54.00000	1	54.00000	6.11321	.029
CAT WITHIN DRUG(3)	54.00000	1	54.00000	6.11321	.029

The simple effects of the three drugs within each category of patients can be tested in the same manner.

In higher-order designs one may want tests of simple effects for both interactions and main effects. For example, consider a three-way factorial design with factors A, B, and C, each with two levels. Should the three-way interaction appear significant then an examination of the second-order interaction terms at various levels of the third factor would be in order. To accomplish this, the following DESIGN subcommands would be used:

```
DESIGN=A, B, C, A BY B, A BY C, B BY C,
       B BY C WITHIN A(1), B BY C WITHIN A(2)/

DESIGN=A, B, C, A BY B, A BY C, B BY C,
       A BY B WITHIN C(1), A BY B WITHIN C(2)/

DESIGN=A, B, C, A BY B, A BY C, B BY C,
       A BY C WITHIN B(1), A BY C WITHIN B(2)/
```

Test of simple main effects can be requested as well. To test factor A within the B BY C treatment combinations the following DESIGN subcommand is used:

```
DESIGN=B, C, B BY C, A WITHIN B(1) BY C(1),
       A WITHIN B(1) BY C(2), A WITHIN B(2) BY C(1),
       A WITHIN B(2) BY C(2)/
```

It may also be desirable to compare two or more means at particular levels of another factor or treatment combinations. For example, it may be interesting to compare the effectiveness of drug 1 with drug 2 within each patient category. Such comparisons can be performed by extending the methods used for ordinary simple effects. The procedure is as follows:

1 Define a contrast incorporating the comparisons of interest such as

```
CONTRAST(DRUG)=SPECIAL(1 1 1 1 -1 0 2 -1 -1)/
```

2 Partition the factor into the desired components by specifying

```
PARTITION(DRUG)/
```

In subsequent designs, DRUG(1) will refer to the drug 1 versus drug 2 comparison. DRUG(2) will refer to the drug 1 versus drugs 2 and 3 combined comparison.

3 Request regression-approach sums of squares by using

```
METHOD=SSTYPE(UNIQUE)/
```

This is mandatory even for orthogonal designs, because DRUG(1) and DRUG(2) are not independent.

4 Specify the designs as for ordinary simple effects, but expand the simple effects terms according to the CONTRAST/PARTITION specification:

```
DESIGN=CAT, DRUG(1) WITHIN CAT(1),
            DRUG(1) WITHIN CAT(2),
            DRUG(2) WITHIN CAT(1),
            DRUG(2) WITHIN CAT(2)/
```

Figure 1.30b presents the output for this design.

Figure 1.30b

TESTS OF SIGNIFICANCE FOR Y USING UNIQUE SUMS OF SQUARES

SOURCE OF VARIATION	SUM OF SQUARES	DF	MEAN SQUARE	F	SIG. OF F
WITHIN CELLS	106.00000	12	8.83333		
CONSTANT	882.00000	1	882.00000	99.84906	0.0
CAT	18.00000	1	18.00000	2.03774	.179
DRUG(1) WITHIN CAT(1)	24.00000	1	24.00000	2.71698	.125
DRUG(2) WITHIN CAT(2)	18.00000	1	18.00000	2.03774	.179
DRUG(1) WITHIN CAT(2)	96.00000	1	96.00000	10.86792	.006
DRUG(2) WITHIN CAT(1)	18.00000	1	18.00000	2.03774	.179

1.31 MULTIVARIATE TESTS OF SIGNIFICANCE

1.32 Standard MANOVA Output

In the univariate F test, the F value is a function of the ratio (SSH)/(SSE), where SSH is the sum of squares due to the hypothesis and SSE the sum of squares due to error. Significance tests in multivariate analysis of variance models are based on functions of the eigenvalues of the matrix $S_h S_e^{-1}$, where S_h is the matrix of the sums of squares and cross products (SSCP) for the hypothesis and S_e is the SSCP matrix for the error. The MANOVA procedure computes four statistics used for significance tests: Roy's largest root, Wilks' lambda, Hotelling's trace, and Pillai's criterion. (All of these are functions of the eigenvalues.)

The MANOVA commands for the multivariate analysis are exactly the same as in the univariate case, except that two or more response variables are specified instead of one. Figure 1.32a, given below, illustrates the use of MANOVA to analyze the dental calculus reduction data in Finn (1974). The response variables in this example are RCAN, RLI, and RCI.

Figure 1.32a

```
RUN NAME        DENTAL CALULUS DATA FROM FINN(1974) PAGE C-56
FILE NAME       DATA FOR ANTI-CALCULUS AGENT
VARIABLE LIST   YEAR,TR,RCAN,RLI,RCI,LCI,LLI,LCAN
INPUT FORMAT    FIXED(2F1.0,6F2.0)
N OF CASES      107
MISSING VALUES YEAR TO LCAN(BLANK)
MANOVA          RCAN,RLI,RCI BY YEAR(1,2),TR(1,5)/
READ INPUT DATA
11 2 2 1 2 2 1
11 0 0 0 2 1 0
11 0 0 4 4 0 0
11 2 2 2 3 2 2
 . . . . . .
 . . . . . .
 . . . . . .
23 0 1 3 4 3 0
23 1 0 1 0 1 0
23 0 1 0 0 0 0
23 0 1 6 4 1 0
FINISH
```

Since no DESIGN specifications are given in Figure 1.32a, a full factorial model is assumed. The standard output (without the PRINT subcommand) includes

1 General information about the design. This includes the number of observations, the number of levels of each effect, and the redundant effects (if any) in the model. This output is given in Figure 1.32b for the dental calculus data. (Three degrees of freedom are lost in the interaction effect because of empty cells.)

Figure 1.32b

```
     107 CASES ACCEPTED.
       0 CASES REJECTED BECAUSE OF OUT-OF-RANGE FACTOR VALUES.
       0 CASES REJECTED BECAUSE OF MISSING DATA.
       7 NON-EMPTY CELLS.

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

CORRESPONDENCE BETWEEN EFFECTS AND COLUMNS OF BETWEEN-SUBJECTS DESIGN

 STARTING  ENDING
  COLUMN   COLUMN   EFFECT NAME

     1        1     CONSTANT
     2        2     YEAR
     3        6     TR
     7       10     YEAR BY TR

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

REDUNDANCIES IN DESIGN MATRIX

   COLUMN     EFFECT

      8     YEAR BY TR
      9      (SAME)
     10      (SAME)
```

2 Multivariate tests of the significance of each effect in the model. The four test statistics previously mentioned are given. Each of these statistics is a function of the nonzero eigenvalues λ_i of the matrix $S_h S_e^{-1}$. The number of nonzero eigenvalues, s, is equal to the minimum of the number of dependent variables, q, and the degrees of freedom for the tested effect, n_h. The distributions of these statistics, under the null hypothesis, depend on q, n_h, and n_e (the error degrees of freedom).

Pillai's criterion. This test statistic, sum of $\lambda_i/(1+\lambda_i)$, can be approximated by an *F* variate (see Pillai, 1960). (The degrees of freedom are a function of q, n_h, and n_e.)

Hotelling's trace. This is the statistic T = sum of λ_i, which is equal to the trace of $S_h S_e^{-1}$. The critical points of the distribution of T have been tabulated by Pillai (1960) and depend on S = min(p,q), M = $(|n_h - q| - 1)/2$, and N = $(n_e - q - 1)/2$. (The values of S, M, and N for each effect are printed by MANOVA.) MANOVA also gives an approximate *F* statistic based on T, where the degrees of freedom depend on q, n_h and n_e.

Wilks' lambda. This test statistic, product of $1/(1+\lambda_i)$, can be transformed, using Rao's formula (Rao, 1973), into an approximate *F* statistic with degrees of freedom determined by q, n_h, and n_e.

Roy's largest root criterion. Upper percentage points of the distribution of this test statistic, $\lambda_1/(1+\lambda_1)$, where λ_1 is the largest eigenvalue of $S_h S_e^{-1}$, can be found in Heck (1960), Pillai (1967), and Morrison (1976). This distribution, like that of Hotelling's trace, depends on S, M, and N.

For the dental calculus data, the multivariate tests of the hypothesis that there is no TR effect (adjusted for the YEAR effect) are presented in Figure 1.32c.

Figure 1.32c

EFFECT .. TR

MULTIVARIATE TESTS OF SIGNIFICANCE (S = 3, M = 0, N = 48)

TEST NAME	VALUE	APPROX. F	HYPOTH. DF	ERROR DF	SIG. OF F
PILLAIS	.20122	1.79739	12.00	300.00	.048
HOTELLINGS	.22813	1.83769	12.00	290.00	.042
WILKS	.80733	1.82255	12.00	259.58	.045
ROYS	.14402				

The name of the test statistic is given under TEST NAME and its value listed under VALUE. For Pillai's criterion, Hotelling's trace, and Wilks lambda, approximate *F* statistics are given, with the degrees of freedom under HYPOTH. DF and ERROR DF and the p-values under SIG. OF F. A comparison (with references) of the powers of these four tests can be found in Morrison (1976).

3 Eigenvalues and canonical correlations. The nonzero eigenvalues of $S_h S_e^{-1}$ and the corresponding canonical correlations for each effect in the model are given. For example, the results for the effect TR are shown in Figure 1.32d.

Figure 1.32d

EIGENVALUES AND CANONICAL CORRELATIONS

ROOT NO.	EIGENVALUE	PCT.	CUM. PCT.	CANON. COR.
1	.16825	73.75366	73.75366	.37950
2	.05253	23.02709	96.78075	.22340
3	.00734	3.21925	100.00000	.08538

The canonical correlation coefficients ρ_i are calculated as $\rho_i^2 = \lambda_i/(1 + \lambda_i)$; they are the canonical correlations between the response variables and the effect. ρ_i also measures the correlation between the *i*th canonical variate of the response variables and the tested effect (in certain linear combinations). The canonical correlations in this example can also be obtained by using the following dummy variables to represent the YEAR and TR effects.

X_1 = 1 if YEAR = 2
0 otherwise

Y_1 = 1 if TR = 2
0 otherwise

.
.

Y_4 = 1 if TR = 5
0 otherwise

If X_1 is already in the regression equation (since TR is adjusted for YEAR) and the within-cells SSCP matrix is used as the error matrix, then the ρ_i's above are the canonical correlations between RCAN, RLI and RCI, and Y_1, Y_2, Y_3, and Y_4.

4 Dimension reduction analysis. Dimension reduction analysis, based on Wilks' lambda, is used to assess the dimensionality of a significant relationship between the response variables and the tested effect. The first test is based on all the eigenvalues and is equivalent to the overall Wilks' lambda test; the second test is performed on all the eigenvalues except the largest, and so on. Hence the value of Wilks' lambda for testing roots n_1 to n_2 is found by calculating the product from $i = n_1$ to $i = n_2$ of $1/(1+\lambda_i)$.

MANOVA also prints the approximate *F* statistic for each of these Wilks' lambda statistics. For the effect TR, the output in Figure 1.32e is obtained.

Figure 1.32e

DIMENSION REDUCTION ANALYSIS

ROOTS	WILKS LAMBDA	F	HYPOTH. DF	ERROR DF	SIG. OF F
1 TO 3	.80733	1.82255	12.00	259.58	.045
2 TO 3	.94316	.97402	6.00	240.16	.443
3 TO 3	.99271	.36286	2.00	198.00	.696

Dimension reduction analysis can be interpreted as follows: If the roots from n_o to s are not significant (in other words, if the $s - n_o + 1$ smallest canonical correlations are not significantly different from zero), we may say that the data do not provide evidence of association in more than $n_o - 1$ dimensions (only $n_o - 1$ discriminant functions are significant). In the dental calculus example, only one canonical correlation is significant at the 0.05 level for the TR effect.

5 Univariate analysis of variance results for each of the q response variables. In our example, Figure 1.32f gives the results obtained for the effect TR.

Figure 1.32f

UNIVARIATE F-TESTS WITH (4,100) D. F.

VARIABLE	HYPOTH. SS	ERROR SS	HYPOTH. MS	ERROR MS	F	SIG. OF F
RCAN	6.18306	137.89515	1.54577	1.37895	1.12097	.351
RLI	28.07315	261.87433	7.01829	2.61874	2.68002	.036
RCI	69.55358	423.98046	17.38839	4.23980	4.10123	.004

The sum of squares for the tested effect (HYPOTH. SS) and for the error (ERROR SS) of each response variable are the appropriate diagonal elements of S_h and S_e respectively.

Output for the YEAR effect and the YEAR BY TR interaction is given in Figure 1.32g.

Figure 1.32g

EFFECT .. YEAR BY TR
MULTIVARIATE TESTS OF SIGNIFICANCE (S = 1, M = 1/2, N = 48)

TEST NAME	VALUE	APPROX. F	HYPOTH. DF	ERROR DF	SIG. OF F
PILLAIS	.02445	.81881	3.00	98.00	.487
HOTELLINGS	.02507	.81881	3.00	98.00	.487
WILKS	.97555	.81881	3.00	98.00	.487
ROYS	.02445				

EIGENVALUES AND CANONICAL CORRELATIONS

ROOT NO.	EIGENVALUE	PCT.	CUM. PCT.	CANON. COR.
1	.02507	100.00000	100.00000	.15637

DIMENSION REDUCTION ANALYSIS

ROOTS	WILKS LAMBDA	F	HYPOTH. DF	ERROR DF	SIG. OF F
1 TO 1	.97555	.81881	3.00	98.00	.487

UNIVARIATE F-TESTS WITH (1,100) D. F.

VARIABLE	HYPOTH. SS	ERROR SS	HYPOTH. MS	ERROR MS	F	SIG. OF F
RCAN	.09862	137.89515	.09862	1.37895	.07152	.790
RLI	1.08877	261.87433	1.08877	2.61874	.41576	.521
RCI	9.73563	423.98046	9.73563	4.23980	2.29625	.133

EFFECT .. YEAR

MULTIVARIATE TESTS OF SIGNIFICANCE (S = 1, M = 1/2, N = 48)

TEST NAME	VALUE	APPROX. F	HYPOTH. DF	ERROR DF	SIG. OF F
PILLAIS	.04077	1.38843	3.00	98.00	.251
HOTELLINGS	.04250	1.38843	3.00	98.00	.251
WILKS	.95923	1.38843	3.00	98.00	.251
ROYS	.04077				

EIGENVALUES AND CANONICAL CORRELATIONS

ROOT NO.	EIGENVALUE	PCT.	CUM. PCT.	CANON. COR.
1	.04250	100.00000	100.00000	.20192

DIMENSION REDUCTION ANALYSIS

ROOTS	WILKS LAMBDA	F	HYPOTH. DF	ERROR DF	SIG. OF F
1 TO 1	.95923	1.38843	3.00	98.00	.251

UNIVARIATE F-TESTS WITH (1,100) D. F.

VARIABLE	HYPOTH. SS	ERROR SS	HYPOTH. MS	ERROR MS	F	SIG. OF F
RCAN	3.54279	137.89515	3.54279	1.37895	2.56919	.112
RLI	6.83291	261.87433	6.83291	2.61874	2.60923	.109
RCI	12.97332	423.98046	12.97332	4.23980	3.05989	.083

Remember, the (default) sequential approach is used to obtain the S_h matrix for each effect. Thus YEAR BY TR is adjusted for TR and YEAR, and TR is adjusted for YEAR.

6 Parameter estimates and related statistics for each response variable. These consist of the standard errors of the parameter estimates, t-values and their significance levels (two-tailed), and 95% confidence intervals for the parameters. The parameters estimated depend on the contrasts chosen for the reparameterization. The output shown in Figure 1.32h describes the parameters for the dental calculus example.

Figure 1.32h

ESTIMATES FOR MEAN

CONSTANT

PARAMETER	COEFF.	STD. ERR.	T-VALUE	SIG. OF T	LOWER .95 CL	UPPER .95 CL
1	.7455586081	.15426	4.83313	.000	.43951	1.05161

YEAR

PARAMETER	COEFF.	STD. ERR.	T-VALUE	SIG. OF T	LOWER .95 CL	UPPER .95 CL
2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

TR

PARAMETER	COEFF.	STD. ERR.	T-VALUE	SIG. OF T	LOWER .95 CL	UPPER .95 CL
3	-.0312728938	.26335	-.11875	.906	-.55374	.49120
4	.6443223443	.43965	1.46553	.146	-.22793	1.51658
5	-.2604395604	.24892	-1.04626	.298	-.75430	.23342
6	.3109890110	.51656	.60204	.549	-.71386	1.33583

YEAR BY TR

PARAMETER	COEFF.	STD. ERR.	T-VALUE	SIG. OF T	LOWER .95 CL	UPPER .95 CL
7	.0922619048	.34499	.26744	.790	-.59218	.77671
8	0.0	.	.	.	.	.
9	0.0	.	.	.	.	.
10	0.0	.	.	.	.	.

ESTIMATES FOR RLI

CONSTANT

PARAMETER	COEFF.	STD. ERR.	T-VALUE	SIG. OF T	LOWER .95 CL	UPPER .95 CL
1	1.0793376068	.21258	5.07729	.000	.65758	1.50109

YEAR

PARAMETER	COEFF.	STD. ERR.	T-VALUE	SIG. OF T	LOWER .95 CL	UPPER .95 CL
2	.0327380952	.34757	.09419	.925	-.65683	.72231

TR

PARAMETER	COEFF.	STD. ERR.	T-VALUE	SIG. OF T	LOWER .95 CL	UPPER .95 CL
3	.8313766789	.36291	2.29087	.024	.11138	1.55138
4	.6657020757	.60587	1.09875	.275	-.53633	1.86773
5	-.2549328449	.34304	-.74317	.459	-.93551	.42564
6	-.3120757021	.71186	-.43839	.662	-1.72439	1.10023

YEAR BY TR

PARAMETER	COEFF.	STD. ERR.	T-VALUE	SIG. OF T	LOWER .95 CL	UPPER .95 CL
7	.3065476190	.47542	.64480	.521	-.63667	1.24976
8	0.0	.	.	.	.	.
9	0.0	.	.	.	.	.
10	0.0	.	.	.	.	.

ESTIMATES FOR RCI

CONSTANT

PARAMETER	COEFF.	STD. ERR.	T-VALUE	SIG. OF T	LOWER .95 CL	UPPER .95 CL
1	2.0558302808	.27049	7.60038	0.0	1.51918	2.59248

YEAR

PARAMETER	COEFF.	STD. ERR.	T-VALUE	SIG. OF T	LOWER .95 CL	UPPER .95 CL
2	-.3988095238	.44225	-.90177	.369	-1.27622	.47860

TR

PARAMETER	COEFF.	STD. ERR.	T-VALUE	SIG. OF T	LOWER .95 CL	UPPER .95 CL
3	1.1763125763	.46177	2.54741	.012	.26018	2.09245
4	1.4540903541	.77091	1.88619	.062	-.07538	2.98356
5	-.3713064713	.43648	-.85068	.397	-1.23727	.49466
6	.3429792430	.90578	.37866	.706	-1.45405	2.14001

YEAR BY TR

PARAMETER	COEFF.	STD. ERR.	T-VALUE	SIG. OF T	LOWER .95 CL	UPPER .95 CL
7	.9166666667	.60493	1.51534	.133	-.28349	2.11682
8	0.0	.	.	.	.	.
9	0.0	.	.	.	.	.
10	0.0	.	.	.	.	.

1.33 Optional MANOVA Output

Other output related to multivariate significance tests can be obtained by using the PRINT subcommand. Such optional output includes

1 The error matrix. For every error matrix used in the model,

```
PRINT=ERROR(SSCP)/
```

can be used to obtain the error SSCP matrix, S_e. Although only one error matrix, the within-cells error matrix, was used in Figure 1.32a, more than one error matrix is sometimes used (e.g., in multivariate nested designs). The error matrix for Figure 1.32a is given in Figure 1.33a.

Figure 1.33a

```
WITHIN CELLS SUM-OF-SQUARES AND CROSS-PRODUCTS

                RCAN            RLI            RCI

RCAN        137.89515
RLI         101.90797      261.87433
RCI          81.03938      217.53449      423.98046
```

The error variance-covariance and error correlation matrices can also be obtained, by specifying

```
PRINT=ERROR(COV,COR)/
```

2 The hypothesis SSCP matrix. The matrix S_h for each effect can be obtained by specifying

```
PRINT=SIGNIF(HYPOTH)/
```

This matrix is adjusted for the covariates (if any). The hypothesis SSCP matrix for the TR effect in Figure 1.32a is given in Figure 1.33b.

Figure 1.33b

```
EFFECT .. TR

ADJUSTED HYPOTHESIS SUM-OF-SQUARES AND CROSS-PRODUCTS

                RCAN            RLI            RCI

RCAN         6.18306
RLI          8.26479       28.07315
RCI         15.81805       41.86935       69.55358
```

3 Roy-Bargmann step-down analysis (Roy and Bargmann, 1958). For each effect, step-down tests (which depend on the ordering of the response variables) can be performed by specifying

```
PRINT=SIGNIF(STEPDOWN)/
```

The number of tests for effects in a step-down analysis is equal to the number of response variables in the model. For the first response variable, the test statistic is the same as the univariate F statistic. The test statistic for the second response variable is identical to the univariate test statistic that would result if the first response variable were treated as a covariate. The test statistic for the third response variable is adjusted for the first two variables, and so on. A significant test statistic for the kth response variable indicates that this variable is important for testing the hypothesis that the effect is zero and cannot be accounted for by a linear combination of the preceding $k - 1$ variables. Since testing begins with the last variable and proceeds backwards until a significant result is obtained, the variables assumed to be important in testing an effect should appear early in the step-down ordering. MANOVA uses the ordering of the response variables given in the MANOVA variable list. The step-down analysis for the TR effect in Figure 1.32a is given in Figure 1.33c.

Figure 1.33c

```
ROY-BARGMAN STEPDOWN F - TESTS

VARIABLE    HYPOTH. MS     ERROR MS    STEP-DOWN F    HYPOTH. DF    ERROR DF    SIG. OF F

RCAN          1.54577      1.37895        1.12097             4         100         .351
RLI           4.78488      1.88446        2.53912             4          99         .045
RCI           4.57059      2.48108        1.84218             4          98         .127
```

4 The average F test. If

```
PRINT=SIGNIF(AVERF)/
```

is specified, MANOVA outputs an averaged F test for each tested effect. This is particularly useful for repeated measures designs (see Section 1.44). The sum of squares for the effect and the sum of squares for the error in the averaged F test are obtained by summing over the hypothesis sum of squares and the error sum of squares, respectively, for each variable. The averaged F test for the TR effect in the dental calculus example is given in Figure 1.33d.

Figure 1.33d

AVERAGED F-TEST WITH (12,300) D. F.

	HYPOTH. SS	ERROR SS	HYPOTH. MS	ERROR MS	F	SIG. OF F
(AVER.)	103.80979	823.74994	8.65082	2.74583	3.15052	.000

5 A brief table of multivariate significance tests. A summary table, similar to the univariate ANOVA table, (with Wilks' lambda and the corresponding approximate F statistic replacing the univariate F) can be obtained by specifying

```
PRINT=SIGNIF(BRIEF)/
```

Note that the BRIEF specification overrides requests for the standard multivariate significance tests, the hypothesis SSCP matrix, and step-down analysis. The BRIEF output for Figure 1.32a is given in Figure 1.33e.

Figure 1.33e

TESTS OF SIGNIFICANCE FOR WITHIN CELLS USING SEQUENTIAL SUMS OF SQUARES

SOURCE OF VARIATION	WILKS LAMBDA	APPROX MULT F	SIG. OF F	AVERAGED F	SIG. OF F
CONSTANT	.46843	37.07044	0.0	76.86413	0.0
YEAR	.95923	1.38843	.251	2.83448	.038
TR	.80733	1.82255	.045	3.15052	.000
YEAR BY TR	.97555	.81881	.487	1.32601	.266

1.34 Principal Components Analysis

Principal components analysis (which is performed on each error matrix used in the model) can be requested via the PRINT subcommand. If

```
PRINT=PRINCOMPS(COR)/
```

is specified, the principal components of the error correlation matrix are printed, while

```
PRINT=PRINCOMPS(COV)/
```

produces the principal components of the error-covariance matrix.

The output for a principal components analysis includes

1 A table listing the eigenvalues of the error matrix (COR or COV) and the proportion and cumulative proportion of the total variance accounted for by each component.

2 The principal components of the error matrix.

3 The determinant of the error matrix, the Bartlett test of sphericity, and F max tests. The Bartlett test statistic, which has an approximate chi-square distribution with $q(q-1)/2$ degrees of freedom, is used to test the hypothesis that the population error correlation matrix is an identity matrix (or, equivalently, that the population error variance-covariance matrix is a diagonal matrix). The F max statistic (the ratio of the largest to the smallest diagonal element of the error variance-covariance matrix) is used to test the hypothesis that the variances of the q response variables are equal. The critical points of the distribution of F max under the null hypothesis can be found in Winer (1971) and depend on q and n_e. Both the Bartlett test and F max test can be obtained simply by requesting the error correlation matrix; i.e., by specifying

```
PRINT=ERROR(COR)/
```

in a MANOVA run. It is not necessary to perform a principal components analysis in order to obtain these statistics.

The output from a principal components analysis performed on the dental calculus data is given in Figure 1.34.

Figure 1.34

```
EIGENVALUES OF WITHIN CELLS CORRELATION MATRIX

              EIGENVALUE      PCT OF VAR        CUM PCT

       1         2.02696        67.56528       67.56528
       2          .67398        22.46611       90.03139
       3          .29906         9.96861      100.00000

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

NORMALIZED PRINCIPAL COMPONENTS

          COMPONENTS

VARIABLES              1              2              3

RCAN             -.73816        -.65197        -.17338
RLI              -.90389         .08961         .41827
RCI              -.81551         .49081        -.30667

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

DETERMINANT =                    .40855
BARTLETT TEST OF SPHERICITY =  87.87194 WITH 3 D. F.
SIGNIFICANCE =                   .000

F(MAX) CRITERION =            3.07466 WITH (3,100) D. F.
```

MANOVA also enables the user to rotate the principal components loadings. The keywords for specifying the type of rotation are VARIMAX, QUARTIMAX, and EQUIMAX (see *SPSS*, Second Edition, pp. 484-485, for a description of these three rotations). NOROTATE inhibits rotation. For example, if

```
PRINT=PRINCOMPS(COR,ROTATE(VARIMAX))/
```

is specified, a principal components analysis is performed on each error correlation matrix and the varimax method is used to rotate the component loadings. By default, all components are rotated. Fewer components may be rotated by specifying the number of components to be rotated, in parentheses, after the NCOMP keyword or by specifying a cutoff value for the eigenvalues, in parentheses, after the MINEIGEN keyword. For example, specifying

```
PRINT=PRINCOMPS(COR,ROTATE(VARIMAX),NCOMP(2))/
```

causes only the first two components to be rotated. If

```
PRINT=PRINCOMPS(COR,ROTATE(VARIMAX),MINEIGEN(1.5))/
```

is specified, only those components associated with eigenvalues greater than 1.5 will be rotated.

1.35 Discriminant Analysis

MANOVA can be used to perform discriminant analysis for each effect in the model. The PRINT subcommand requesting discriminant analysis has the format

```
PRINT=DISCRIM(output list)/
```

The output list may include requests for

1. The raw discriminant function coefficients. These are obtained for each tested effect by specifying

   ```
   PRINT=DISCRIM(RAW)/
   ```

2. The standardized discriminant function coefficients. If

   ```
   PRINT=DISCRIM(STAN)/
   ```

   is specified, the standardized discriminant function coefficients (obtained by multiplying each raw coefficient by the corresponding standard deviation of the variable) will be printed.

3. The effect estimates in the discriminant function space. To obtain the estimates of each effect for the canonical variables, specify

   ```
   PRINT=DISCRIM(ESTIM)/
   ```

   The canonical variables are defined here as the canonical variates associated with the response variables.

4. The correlations between response variables and canonical variables. These are obtained by specifying

   ```
   PRINT=DISCRIM(COR)/
   ```

As an indication of how much each response variable contributes to the canonical variate, these correlations aid in the interpretation of the canonical variables.

For the dental calculus data, a discriminant analysis for the effect TR is requested by specifying

```
PRINT=DISCRIM(RAW,STAN,ESTIM,COR)/
```

The resulting output is given in Figure 1.35.

Figure 1.35

```
RAW DISCRIMINANT FUNCTION COEFFICIENTS

          FUNCTION NO.
VARIABLE                1

RCAN                .02507
RLI                -.14814
RCI                -.40728

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

STANDARDIZED DISCRIMINANT FUNCTION COEFFICIENTS

          FUNCTION NO.

VARIABLE                1

RCAN                .02944
RLI                -.23973
RCI                -.83862

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

ESTIMATES OF EFFECTS FOR CANONICAL VARIABLES

          CANONICAL VARIABLE

 PARAMETER              1

       3           -.60303
       4           -.67469
       5            .18246
       6           -.08566

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

CORRELATIONS BETWEEN DEPENDENT AND CANONICAL VARIABLES

          CANONICAL VARIABLE

VARIABLE                1

RCAN               -.38019
RLI                -.77143
RCI                -.98526
```

Discriminant analysis results are reported only for those functions (or corresponding canonical correlations; see Section 1.32) that are significant at level α. The default value of α is 0.15. In Figure 1.32a, the dimension reduction analysis for the TR effect indicates that only the first canonical correlation is significant (the observed significance level is 0.045); hence only one discriminant function is reported in the output displayed above. The value of α can be set by specifying a number between 0 and 1, in parentheses, after the keyword ALPHA. Thus,

```
PRINT=DISCRIM(RAW,COR,ALPHA(0.5))/
```

produces discriminant function coefficients and the correlations between response variables and canonical variables that correspond to discriminant functions with significance levels less than 0.5. If $\alpha = 1.0$ is specified, MANOVA reports all the discriminant functions.

The correlations between the response variables and the canonical variables can be rotated by adding the ROTATE keyword to the PRINT subcommand. (The types of rotation available are described in 1.34.) For example,

```
PRINT=DISCRIM(COR,ROTATE(VARIMAX),ALPHA(1.0))/
```

produces the correlations between the response variables and all the canonical variables and rotates the canonical variables (using the varimax method).

1.36 Box's M Test

The assumption of homogeneous within-cells variance-covariance matrices can be assessed by Box's M test, a multivariate analog of Bartlett's test. If

```
PRINT=HOMOGENEITY(BOXM)/
```

is specified, MANOVA will print Box's M statistic and an approximate *F* statistic with its p- value. The results of Box's M test for the dental calculus data are given in Figure 1.36.

Figure 1.36

```
MULTIVARIATE TEST FOR HOMOGENEITY OF DISPERSION MATRICES

BOXS M =                    114.53559
F WITH (36,2404) DF =         2.67721, P =    .000 (APPROX.)
CHI-SQUARE WITH 36 DF =      98.09416, P =    .000 (APPROX.)
```

1.37 Multivariate Analysis of Covariance

MANOVA will also perform a multivariate analysis of covariance. Figure 1.37a illustrates this use of MANOVA.

Figure 1.37a

```
RUN NAME        DENTAL CALCULUS DATA FROM FINN(1974) PAGE C-56
FILE NAME       DATA FOR ANTI-CALCULUS AGENT
VARIABLE LIST   YEAR,TR,RCAN,RLI,RCI,LCI,LLI,LCAN
INPUT FORMAT    FIXED(2F1.0,6F2.0)
N OF CASES      107
MISSING VALUES YEAR TO LCAN(BLANK)
MANOVA          RCAN,RLI,RCI BY YEAR(1,2),TR(1,5) WITH LCI/
READ INPUT DATA
11 2 2 1 2 2 1
11 0 0 0 2 1 0
11 0 0 4 4 0 0
11 2 2 2 3 2 2
 . . . . . .
 . . . . . .
 . . . . . .
23 0 1 3 4 3 0
23 1 0 1 0 1 0
23 0 1 0 0 0 0
23 0 1 6 4 1 0
FINISH
```

RCAN, RLT, and RCI are the response variables and LCI the covariate in this example. The discussion of univariate analysis of covariance in Section 1.17 can be generalized.

When a covariate is specified, multivariate significance testing and parameter estimation are adjusted for the covariate; i.e., both S_h and S_e are adjusted. For the dental calculus data, the multivariate significance tests for the TR effect, adjusted for the covariate LCI, are given in Figure 1.37b.

Figure 1.37b

```
EFFECT .. TR

MULTIVARIATE TESTS OF SIGNIFICANCE (S = 3, M = 0, N = 47 1/2)

TEST NAME              VALUE       APPROX. F      HYPOTH. DF       ERROR DF       SIG. OF F

PILLAIS                .18149        1.59371           12.00         297.00            .092
HOTELLINGS             .20468        1.63178           12.00         287.00            .082
WILKS                  .82485        1.61631           12.00         256.93            .087
ROYS                   .13641

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

EIGENVALUES AND CANONICAL CORRELATIONS

ROOT NO.      EIGENVALUE          PCT.      CUM. PCT.    CANON. COR.

       1          .15796      77.17431       77.17431         .36934
       2          .04095      20.00838       97.18269         .19835
       3          .00577       2.81731      100.00000         .07572

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

DIMENSION REDUCTION ANALYSIS

ROOTS          WILKS LAMBDA              F      HYPOTH. DF       ERROR DF       SIG. OF F

1 TO 3                .82485        1.61631           12.00         256.93            .087
2 TO 3                .95515         .75412            6.00         237.72            .607
3 TO 3                .99427         .28215            2.00         196.00            .754

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

UNIVARIATE F-TESTS WITH (4,99) D. F.

VARIABLE       HYPOTH. SS       ERROR SS      HYPOTH. MS       ERROR MS              F       SIG. OF F

RCAN             4.20875      119.32324         1.05219         1.20529         .87298            .483
RLI              4.15057      152.39986         1.03764         1.53939         .67406            .612
RCI             18.00210      123.66477         4.50052         1.24914        3.60290            .009
```

MANOVA also prints multivariate significance tests of the hypothesis that the regression coefficients are zero, under the heading EFFECT..WITHIN CELLS REGRESSION. (WITHIN CELLS indicates that the within-cells error matrix was used in the model.) These tests for the dental calculus data of Figure 1.37a are shown in Figure 1.37c. (See Section 1.38 for a more detailed discussion of regression analysis.)

Figure 1.37c

EFFECT .. WITHIN CELLS REGRESSION

MULTIVARIATE TESTS OF SIGNIFICANCE (S = 1, M = 1/2, N = 47 1/2)

TEST NAME	VALUE	APPROX. F	HYPOTH. DF	ERROR DF	SIG. OF F
PILLAIS	.72613	85.72625	3.00	97.00	0.0
HOTELLINGS	2.65133	85.72625	3.00	97.00	0.0
WILKS	.27387	85.72625	3.00	97.00	0.0
ROYS	.72613				

EIGENVALUES AND CANONICAL CORRELATIONS

ROOT NO.	EIGENVALUE	PCT.	CUM. PCT.	CANON. COR.	SQUARED COR.
1	2.65133	100.00000	100.00000	.85213	.72613

DIMENSION REDUCTION ANALYSIS

ROOTS	WILKS LAMBDA	F	HYPOTH. DF	ERROR DF	SIG. OF F
1 TO 1	.27387	85.72625	3.00	97.00	0.0

UNIVARIATE F-TESTS WITH (1,99) D. F.

VARIABLE	SQ. MUL. R	MUL. R	ADJ. R-SQ.	HYPOTH MS	ERROR MS	F	SIG. OF F
RCAN	.13468	.36699	.07350	18.57190	1.20529	15.40872	.000
RLI	.41804	.64656	.37689	109.47447	1.53939	71.11537	.000
RCI	.70832	.84162	.68770	300.31570	1.24914	240.41814	0.0

The estimated parameters for the regression of each response variable on the covariate are also listed, together with standard errors, t-values, and confidence intervals. For Figure 1.37a, the results in Figure 1.37d were obtained.

Figure 1.37d

REGRESSION ANALYSIS FOR WITHIN CELLS ERROR TERM

DEPENDENT VARIABLE ..RCAN

COVARIATE	B	BETA	STD. ERR.	T-VALUE	SIG. OF T	LOWER .95 CL	UPPER .95 CL
LCI	.1731949251	.3669895761	.04412	3.92539	.000	.08565	.26074

DEPENDENT VARIABLE ..RLI

COVARIATE	B	BETA	STD. ERR.	T-VALUE	SIG. OF T	LOWER .95 CL	UPPER .95 CL
LCI	.4204974555	.6465616861	.04986	8.43299	0.0	.32156	.51944

DEPENDENT VARIABLE ..RCI

COVARIATE	B	BETA	STD. ERR.	T-VALUE	SIG. OF T	LOWER .95 CL	UPPER .95 CL
LCI	.6964596479	.8416200881	.04492	15.50542	0.0	.60733	.78559

1.38 MULTIVARIATE MULTIPLE LINEAR REGRESSION

1.39 The Multivariate Linear Regression Model

The univariate regression model

$$Y_i = \beta_0 + \beta_1 X_{i1} + \ldots + \beta_p X_{ip} + \epsilon_i$$

expresses the ith observation of the dependent variable Y as a linear function of p independent variables X_i and the error term ϵ_i.

The ϵ_i are assumed to be independent and normally distributed with mean 0 and variance σ^2, and the β_i's are the unknown parameters to be estimated. The multivariate extension of this model is

$$Y_i = B_0 + B_1X_{i1} + \ldots + B_pX_{ip} + \epsilon_i \quad = B_0 + B'X_i + \epsilon_i$$

where $Y_i = (Y_{i1}\ Y_{i2} \ldots Y_{iq})'$ is a vector of q response variables for observation i, the X_j are independent variables, the B_j are q × 1 vectors containing the regression parameters,and the ϵ_i vectors are the errors (assumed to be independent and to have a q-variate normal distribution with mean **0** and covariance matrix Σ).

1.40 MANOVA Multivariate Regression Analysis

MANOVA provides estimates of B_0, B, and Σ and tests the hypothesis that B = 0. The constant vector B_0 is included in the model unless the subcommand

```
METHOD=ESTIMATION(NOCONSTANT)/
```

is included in the MANOVA run. When NOCONSTANT is specified, the regression line or plane is forced to pass through the origin (i.e., B_0 is assumed to be **0** in the equation). Four test statistics (described in Section 1.32) are given for testing the hypothesis that B = 0: Pillai's criterion, Hotelling's trace, Wilks' lambda, and Roy's largest root. All of these are functions of the nonzero eigenvalues of $S_hS_e^{-1}$, where S_h is the regression SSCP matrix and S_e is the error SSCP matrix.

In Figure 1.40a (taken from Finn, 1974), the dependent variable consists of two divergent measures of achievement, synthesis (SYNTH) and evaluation (EVAL), and the independent variables are a general intelligence index (INTEL) and three measures of creativity (CONOBV, CONRMT, and JOB). Three cross products between the creativity measures and INTEL are formed to represent the interaction terms of the model. Figure 1.40a shows the standard SPSS command file for this problem. COMPUTE statements are used to create the interaction terms.

Figure 1.40a

```
RUN NAME        MULTIVARIATE MULTIPLE REGRESSION
COMMENT         DATA ARE TAKEN FROM FINN(1974)  C-3
VARIABLE LIST   SYNTH EVAL CONOBV CONRMT JOB INTEL
INPUT MEDIUM    CARD
INPUT FORMAT    FREEFIELD
MISSING VALUES  SYNTH TO INTEL(9.9)
N OF CASES      UNKNOWN
COMPUTE         CI1=CONOBV*INTEL
COMPUTE         CI2=CONRMT*INTEL
COMPUTE         CI3=JOB*INTEL
MANOVA          SYNTH EVAL WITH INTEL CONOBV CONRMT JOB CI1 CI2 CI3/
                PRINT=DISCRIM(RAW,STAN,ESTIM,COR)/
READ INPUT DATA
5 1 20.0 5.0 13.0 106.0
0 0 13.0 3.0 10.0  97.0
6 2  9.9 4.0  5.0  90.0
4 2 10.0 3.0 15.0 121.0
1 2 12.0 2.0  4.0  99.0
7 1 25.0 5.0 23.0 120.0
1 2 21.0 3.0 11.0  91.0
  .   .   .    .     .
  .   .   .    .     .
  .   .   .    .     .
3 2 15.0 4.0 12.0 107.0
6 6 22.0 10.0 23.0 143.0
0 0 12.0 2.0 13.0 101.0
4 1 12.0 6.0 10.0 115.0
3 0 10.0 5.0 10.0  97.0
3 1 21.0 3.0 20.0  92.0
END INPUT DATA
```

Note that no factor variables are specified in the MANOVA procedure card and the keyword WITH is used to separate the response and independent variables.

The standard output includes multivariate significance tests and the statistics for parameter estimation described in Section 1.32. The following tests and statistics are of particular interest:

1 Tests of H_0:B=**0** and H_0:B_0=**0**. These are automatically printed, along with the multiple R^2 and adjusted R^2 for each response variable regressed on the independent variables. This portion of the output for Figure 1.40a is given in Figure 1.40b.

Figure 1.40b

EFFECT .. WITHIN CELLS REGRESSION

MULTIVARIATE TESTS OF SIGNIFICANCE (S = 2, M = 2, N = 24 1/2)

TEST NAME	VALUE	APPROX. F	HYPOTH. DF	ERROR DF	SIG. OF F
PILLAIS	.55946	2.88501	14.00	104.00	.001
HOTELLINGS	1.05995	3.78553	14.00	100.00	.000
WILKS	.47077	3.33286	14.00	102.00	.000
ROYS	.49886				

EIGENVALUES AND CANONICAL CORRELATIONS

ROOT NO.	EIGENVALUE	PCT.	CUM. PCT.	CANON. COR.	SQUARED COR.
1	.99544	93.91374	93.91374	.70630	.49886
2	.06451	6.08626	100.00000	.24617	.06060

DIMENSION REDUCTION ANALYSIS

ROOTS	WILKS LAMBDA	F	HYPOTH. DF	ERROR DF	SIG. OF F
1 TO 2	.47077	3.33286	14.00	102.00	.000
2 TO 2	.93940	.55565	6.00	105.00	.765

UNIVARIATE F-TESTS WITH (7,52) D. F.

VARIABLE	SQ. MUL. R	MUL. R	ADJ. R-SQ.	HYPOTH MS	ERROR MS	F	SIG. OF F
SYNTH	.45390	.67372	.38039	11.59727	1.87825	6.17450	.000
EVAL	.36102	.60085	.27500	9.60230	2.28783	4.19712	.001

EFFECT .. CONSTANT

MULTIVARIATE TESTS OF SIGNIFICANCE (S = 1, M = 0, N = 24 1/2)

TEST NAME	VALUE	APPROX. F	HYPOTH. DF	ERROR DF	SIG. OF F
PILLAIS	.01764	.45782	2.00	51.00	.635
HOTELLINGS	.01795	.45782	2.00	51.00	.635
WILKS	.98236	.45782	2.00	51.00	.635
ROYS	.01764				

EIGENVALUES AND CANONICAL CORRELATIONS

ROOT NO.	EIGENVALUE	PCT.	CUM. PCT.	CANON. COR.
1	.01795	100.00000	100.00000	.13281

DIMENSION REDUCTION ANALYSIS

ROOTS	WILKS LAMBDA	F	HYPOTH. DF	ERROR DF	SIG. OF F
1 TO 1	.98236	.45782	2.00	51.00	.635

UNIVARIATE F-TESTS WITH (1,52) D. F.

VARIABLE	HYPOTH. SS	ERROR SS	HYPOTH. MS	ERROR MS	F	SIG. OF F
SYNTH	1.13385	97.66914	1.13385	1.87825	.60367	.441
EVAL	.12773	118.96726	.12773	2.28783	.05583	.814

2 Estimates of the regression coefficients B and B_0 with their standard errors, t values for testing H_0:β_i=0, and 95% confidence intervals for each β_i. The output in Figure 1.40c was obtained for Figure 1.40a.

Figure 1.40c

```
REGRESSION ANALYSIS FOR WITHIN CELLS ERROR TERM

DEPENDENT VARIABLE ..SYNTH

COVARIATE              B          BETA      STD. ERR.      T-VALUE     SIG. OF T  LOWER .95 CL  UPPER .95 CL

INTEL          .0555153073   .4727752433        .05165      1.07475          .287      -.04814        .15917
CONOBV         .2008178054   .7838261472        .24128       .83231          .409      -.28334        .68498
CONRMT         .1410916362   .2648440705        .47795       .29520          .769      -.81799       1.10018
JOB           -.3208770046  -.9994055945        .33236     -.96544           .339      -.98781        .34606
CI1           -.0015680423  -.6955227162        .00234.    -.66986           .506      -.00627        .00313
CI2           -.0009030738  -.2073789045        .00443     -.20380           .839      -.00979        .00799
CI3            .0030798107  1.2548388165        .00314      .98169           .331      -.00322        .00938

DEPENDENT VARIABLE ..EVAL

COVARIATE              B          BETA      STD. ERR.      T-VALUE     SIG. OF T  LOWER .95 CL  UPPER .95 CL

INTEL         -.0094648415  -.0790004191        .05701     -.16602           .869      -.12386        .10493
CONOBV        -.1937798157  -.7413104806        .26629     -.72770           .470      -.72813        .34057
CONRMT         .4295086197   .7901962055        .52750      .81424           .419      -.62900       1.48801
JOB           -.2833034910  -.8648268789        .36682     -.77233           .443     -1.01937        .45277
CI1            .0023782808  1.0339290659        .00258      .92057           .362      -.00281        .00756
CI2           -.0032614042  -.7340404855        .00489     -.66690           .508      -.01307        .00655
CI3            .0025330022  1.0115178798        .00346      .73156           .468      -.00441        .00948

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

ESTIMATES FOR SYNTH ADJUSTED FOR 7 COVARIATES

CONSTANT

 PARAMETER          COEFF.       STD. ERR.        T-VALUE       SIG. OF T    LOWER .95 CL    UPPER .95 CL
         1  -4.0520586339         5.21524         -.77696            .441      -14.51720         6.41309

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

ESTIMATES FOR EVAL ADJUSTED FOR 7 COVARIATES

CONSTANT

 PARAMETER          COEFF.       STD. ERR.        T-VALUE       SIG. OF T    LOWER .95 CL    UPPER .95 CL
         1   1.3600318232         5.75585          .23629            .814      -10.18992        12.90999
```

All of the output related to multivariate significance tests that can be obtained by using the PRINT phrase as described in Section 1.33 is also available in the multivariate regression analysis.

1.41 Canonical Analysis

MANOVA can also be used to obtain the canonical correlation between the dependent and independent variables entered into the multivariate regression model. Canonical correlation analysis obtains the linear combinations $u_i=\mathbf{a}_i'\mathbf{Y}$ and $v_i=\mathbf{b}_i'\mathbf{X}$ (i=1,2,..min(p,q)) such that the sample correlation between u_1 and v_1, is maximized. The sample correlation between u_2 and v_2 is greatest among all linear combinations uncorrelated with u_1 and v_1, and so on. The $\mathbf{a}_i$ and $\mathbf{b}_i$ are the canonical coefficients for the dependent and independent variables, respectively, and the pairs of linear combinations u_i and v_i are called the canonical variates.

The format of the PRINT subcommand requesting canonical analysis is

```
PRINT=DISCRIM(output list)/
```

The output list may include requests for

1 The raw canonical coefficients. If

```
PRINT=DISCRIM(RAW)/
```

is specified, the raw canonical coefficients for the dependent variables and the independent variables are produced. For Figure 1.40a, the output in Figure 1.41a is obtained.

Figure 1.41a

```
RAW CANONICAL COEFFICIENTS FOR DEPENDENT VARIABLES

          FUNCTION NO.

VARIABLE                1

SYNTH                .40444
EVAL                 .22637

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

```
RAW CANONICAL COEFFICIENTS FOR COVARIATES
          FUNCTION NO.
COVARIATE                1
INTEL                .02876
CONOBV               .05288
CONRMT               .21845
JOB                 -.27454
CI1                 -.00014
CI2                 -.00156
CI3                  .00258
```

2 The standardized canonical coefficients. If

```
PRINT=DISCRIM(STAN)/
```

is specified, the standardized canonical coefficients (obtained by multiplying each raw coefficient by the corresponding standard deviation of the variable) are printed. The standardized canonical coefficients for Figure 1.40a are given in Figure 1.41b.

Figure 1.41b

```
STANDARDIZED CANONICAL COEFFICIENTS FOR DEPENDENT VARIABLES
          FUNCTION NO.
VARIABLE                 1
SYNTH                .70415
EVAL                 .40212
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
STANDARDIZED CANONICAL COEFFICIENTS FOR COVARIATES
          CAN. VAR.
COVARIATE                1
INTEL                .42636
CONOBV               .35939
CONRMT               .71393
JOB                 -1.48875
CI1                  -.10475
CI2                  -.62467
CI3                  1.82693
```

3 The correlations between the variables and each canonical variate. These correlations are obtained by specifying

```
PRINT=DISCRIM(COR)/
```

and indicate the contribution of each variable to the canonical variate. The percentage and cumulative percentage of the total variation accounted for by each canonical variate are printed as well. The percentage of variation in the dependent variable accounted for by the *i*th canonical variate is calculated as (the sum of squares of correlations between dependent variable and the *i*th canonical variable) × 100/ (number of response variables). The percentage of variation in the independent variable accounted for by the *i*th canonical variate is obtained similarly. Finally, MANOVA prints the redundancy of the dependent variable given the availability of the independent variables (Cooley and Lohnes, 1971), under the heading PCT VAR COV. This is calculated as the proportion of variance accounted for by the *i*th canonical variate multiplied by the corresponding squared canonical coefficient. The redundancy of the independent variables given the availability of the dependent variable appears in the printed output under PCT VAR DEP and is obtained in a similar way. For Figure 1.40a, the output in Figure 1.41c was obtained.

Figure 1.41c

```
CORRELATIONS BETWEEN DEPENDENT AND CANONICAL VARIABLES
          FUNCTION NO.
VARIABLE                 1
SYNTH                .94733
EVAL                 .82794
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

```
VARIANCE EXPLAINED BY CANONICAL VARIABLES OF DEPENDENT VARIABLES

CAN. VAR.      PCT VAR DEP     CUM PCT DEP     PCT VAR COV     CUM PCT COV

      1         79.14597        79.14597        39.48249        39.48249

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

CORRELATIONS BETWEEN COVARIATES AND CANONICAL VARIABLES

         CAN. VAR.

COVARIATE               1

INTEL                .94646
CONOBV               .30260
CONRMT               .56188
JOB                  .57787
CI1                  .62672
CI2                  .69288
CI3                  .79114

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

VARIANCE EXPLAINED BY CANONICAL VARIABLES OF THE COVARIATES

CAN. VAR.      PCT VAR DEP     CUM PCT DEP     PCT VAR COV     CUM PCT COV

      1         22.34706        22.34706        44.79657        44.79657
```

Note that although the number of canonical variates is equal to s = min(p,q), MANOVA prints only those variates that have a significant canonical correlation. The default significance level is 0.15 and can be changed by using the ALPHA specification, as described in Section 1.35.

1.42 Residuals

MANOVA will calculate and print predicted values and residuals for each response variable if

```
PRINT=POBS/
```

is specified in a MANOVA run (POBS stands for predicted observation). The output also includes the case numbers, observed values, and standardized residuals (obtained by dividing the residuals by the error standard deviation).

If multiple error terms are specified in an analysis of covariance model and the residuals for each case are needed, the ERROR subphrase should be used to designate which error term's regression coefficients are to be used in calculating the predicted values. Any error term defined in the design can be used. Consider, for example, a 3 × 2 factorial design with repeated measures on factor B, a SUBJECT factor nested within factor A, and a covariate X. (See Section 1.44 for a discussion of the repeated measures design.) The following MANOVA cards may be used to obtain residuals and significance tests for the model.

```
MANOVA          Y BY A(1,3) SUBJECT(1,3) B(1,2) WITH X/
                PRINT=POBS(ERROR(2))/
                DESIGN=A VS 1, B VS 2, A BY B VS 2,
                    SUBJECT W A = 1, B BY SUBJECT W A = 2/
```

ERROR(2) within the POBS phrase indicates that the regression coefficients associated with error term 2 are to be used to calculate the predicted values for the model (error term 2 is defined in the DESIGN specification as the interaction between B and SUBJECT (within A)).

Various residual plots (observed versus predicted values, observed values versus standardized residuals, predicted values versus standardized residuals, and case number versus standardized residuals) are also available. For a discussion of the graphic features of MANOVA see Section 1.51.

1.43 SPECIAL TOPICS

1.44 Repeated Measures Designs

1.45 Introduction

Designs in which multiple observations are made on a single experimental unit are called repeated measures designs. For example, if a patient's blood pressure is recorded daily for five days after administration of antihypertensive are medication, five repeated observations are obtained for the same case. If only one variable is being measured, say systolic blood pressure, the design is termed singly multivariate. If several variables, such as standing and recumbent systolic and diastolic blood pressures are recorded, the design is doubly multivariate. Since multiple observations are made on the same experimental unit, they are not independent. Special procedures must therefore be used for analysis of repeated measures data.

There are several possible strategies for analysis of repeated measures designs. Both univariate and multivariate solutions can be obtained. Selection of a strategy should be based on the appropriateness of the necessary assumptions as well as power considerations.

1.46 An Example

Data from a repeated measures design found in Winer (1971, p. 546) are shown in Table 1.46. They consist of accuracy scores obtained by adjusting three dials (DIAL) under two levels of background noise (NOISE) during three consecutive ten-minute periods (PERIOD). Each subject is observed nine times, once at each combination of period and dial type. PERIOD and DIAL are called within- subjects factors, while NOISE is called a between-subjects factor. If subject is considered a factor, then the subject factor is crossed with PERIOD and DIAL, but nested under NOISE level.

Table 1.46

Noise	Subject	Periods:	1			2			3		
		Dials:	1	2	3	1	2	3	1	2	3
1	1		45	53	60	40	52	57	28	37	46
	2		35	41	50	30	37	47	25	32	41
	3		60	65	75	58	54	70	40	47	50
2	4		50	48	61	25	34	51	16	23	35
	5		42	45	55	30	37	43	22	27	37
	6		36	60	77	10	19	37	11	19	16

1.47 Obtaining a Univariate Analysis the Hard Way

The univariate analysis of the repeated measures design displayed in Table 1.46 is obtained by treating subject as a random effect nested under the NOISE factor. The model is called a mixed-effects model, and the resulting analysis is a mixed-model analysis of the repeated measures design.

The technique described in Section 1.25 can be used to determine the appropriate error terms for testing the various effects. Table 1.47 summarizes the effects and corresponding error terms for this example.

Table 1.47

Effect	Error Term
NOISE	Subject within NOISE
PERIOD NOISE × PERIOD	PERIOD × Subject within NOISE
DIAL NOISE × DIAL	DIAL × Subject within NOISE
PERIOD × DIAL NOISE × PERIOD × DIAL	PERIOD × DIAL × SUBJECT within NOISE

Figure 1.47a shows an SPSS command file that can be used to perform a univariate analysis of the repeated measures design for the data in Table 1.46. The resulting ANOVA table is presented in Figure 1.47b. A somewhat complicated DESIGN specification is needed because of the multiple error terms in the model. In the next section, a much easier approach to the same problem is given.

Figure 1.47a

```
RUN NAME        UNIVARIATE ANALYSIS OF REPEATED MEASURES DESIGN.
COMMENT         DATA ARE TAKEN FROM WINER(1971) PAGE 546.
VARIABLE LIST   NOISE SUBJECT PERIOD DIAL Y
INPUT FORMAT    FIXED(4F1.0,1X,F2.0)
N OF CASES      54
INPUT MEDIUM    CARD
MANOVA          Y BY NOISE(1,2) SUBJECT(1,3) PERIOD DIAL(1,3)/
                DESIGN=NOISE VS 1, SUBJECT W NOISE=1, PERIOD VS 2,
                       DIAL VS 3, PERIOD BY SUBJECT W NOISE=2,
                       DIAL BY SUBJECT W NOISE=3, NOISE BY PERIOD VS 2,
                       NOISE BY DIAL VS 3, PERIOD BY DIAL VS 4,
                       PERIOD BY DIAL BY SUBJECT W NOISE=4,
                       NOISE BY PERIOD BY DIAL VS 4/
READ INPUT DATA
1111 45
1112 53
1113 60
1121 40
1122 52
 ...  .
 ...  .
 ...  .
2322 39
2323 57
2331 31
2332 29
2333 46
FINISH
```

Figure 1.47b

TESTS OF SIGNIFICANCE FOR Y USING SEQUENTIAL SUMS OF SQUARES

SOURCE OF VARIATION	SUM OF SQUARES	DF	MEAN SQUARE	F	SIG. OF F
RESIDUAL	0.0	0			
CONSTANT	105868.16667	1	105868.16667	.	.
ERROR 1	2491.11111	4	622.77778		
NOISE	468.16667	1	468.16667	.75174	.435
ERROR 2	234.88889	8	29.36111		
PERIOD	3722.33333	2	1861.16667	63.38884	.000
NOISE BY PERIOD	333.00000	2	166.50000	5.67077	.029
ERROR 3	105.55556	8	13.19444		
DIAL	2370.33333	2	1185.16667	89.82316	0.0
NOISE BY DIAL	50.33333	2	25.16667	1.90737	.210
ERROR 4	127.11111	16	7.94444		
PERIOD BY DIAL	10.66667	4	2.66667	.33566	.850
NOISE BY PERIOD BY DIAL	11.33333	4	2.83333	.35664	.836

The mixed-model analysis requires that the variances of the dependent variable be equal for all factor combinations, and that the correlations of the dependent variable at different combinations of within-subjects factors be equal. The MANOVA procedure provides a test, discussed in the next section, for this assumption of compound symmetry.

If compound symmetry appears to be violated, the multivariate approach can be used. In general, the univariate approach is somewhat more powerful, especially for small sample sizes. Note that in the MANOVA procedure, the univariate results can be obtained from the multivariate analysis output. This is important since the multivariate specifications are much simpler than the univariate mixed- model approach just outlined.

1.48 Trend Analysis

Since both PERIOD and DIAL are statistically significant, one may wish to investigate the growth trends for PERIOD and DIAL. If a trend analysis for PERIOD is desired, this effect can be partitioned into a linear effect, PERIOD(1), and a quadratic effect, PERIOD(2), by using the following specifications.

```
CONTRAST(PERIOD)=POLYNOMIAL/
PARTITION(PERIOD)/
```

Equally spaced PERIOD levels are assumed here; for the use of CONTRAST and PARTITION subcommands when levels are unequally spaced, see Sections 1.88 and 1.89.

As shown in Table 1.47, the test for a PERIOD effect used the PERIOD × (subject within NOISE) error term. For the orthogonal polynomial components of PERIOD, we can either use this error term to test for PERIOD(1) and PERIOD(2) effects or decompose PERIOD × (subject

within NOISE) into PERIOD(1) × (subject within NOISE) and PERIOD(2) × (subject within NOISE) and use these as the error terms for PERIOD(1) and PERIOD(2), respectively. The choice of procedure depends in part on the assumptions of the model (see Bock, 1975, p. 460). Unless PERIOD(1) × (subject within NOISE) and PERIOD(2) × (subject within NOISE) both have a fairly large number of degrees of freedom, the single error term PERIOD × (subject within NOISE) is generally used because this test is more powerful.

All interaction terms containing PERIOD can also be partitioned; for example, NOISE × PERIOD has two components, NOISE × PERIOD(1) and NOISE × PERIOD(2), and the pooled and separated error terms described above may be used to test for these two effects. The MANOVA specifications for trend analyses of PERIOD and DIAL are presented in Figure 1.48a, and the resulting ANOVA table is displayed in Figure 1.48b.

Figure 1.48a

```
MANOVA          Y BY NOISE(1,2) SUBJECT(1,3) PERIOD DIAL(1,3)/
                CONTRAST(PERIOD)=POLYNOMIAL/
                CONTRAST(DIAL)=POLYNOMIAL/
                PARTITION(PERIOD)/
                PARTITION(DIAL)/
                DESIGN=NOISE VS 1, SUBJECT W NOISE=1, PERIOD(1) VS 2,
                       PERIOD(2) VS 2, DIAL(1) VS 3,
                       DIAL(2) VS 3, PERIOD BY SUBJECT W NOISE=2,
                       DIAL BY SUBJECT W NOISE=3, NOISE BY PERIOD VS 2,
                       NOISE BY DIAL VS 3, PERIOD BY DIAL VS 4,
                       PERIOD BY DIAL BY SUBJECT W NOISE=4,
                       NOISE BY PERIOD BY DIAL VS 4/
```

Figure 1.48b

TESTS OF SIGNIFICANCE FOR Y USING SEQUENTIAL SUMS OF SQUARES

SOURCE OF VARIATION	SUM OF SQUARES	DF	MEAN SQUARE	F	SIG. OF F
RESIDUAL	0.0	0	.		
CONSTANT	105868.16667	1	105868.16667	.	.
ERROR 1	2491.11111	4	622.77778		
NOISE	468.16667	1	468.16667	.75174	.435
ERROR 2	234.88889	8	29.36111		
PERIOD(1)	3721.00000	1	3721.00000	126.73226	0.0
PERIOD(2)	1.33333	1	1.33333	.04541	.837
NOISE BY PERIOD	333.00000	2	166.50000	5.67077	.029
ERROR 3	105.55556	8	13.19444		
DIAL(1)	2256.25000	1	2256.25000	171.00000	0.0
DIAL(2)	114.08333	1	114.08333	8.64632	.019
NOISE BY DIAL	50.33333	2	25.16667	1.90737	.210
ERROR 4	127.11111	16	7.94444		
PERIOD BY DIAL	10.66667	4	2.66667	.33566	.850
NOISE BY PERIOD BY DIAL	11.33333	4	2.83333	.35664	.836

1.49 The Multivariate Approach

In the multivariate analysis of repeated measures designs, the responses of a case are treated as an h-dimensional response vector. In the current example each subject responds to nine variables, each variable representing a unique DIAL and PERIOD combination. Thus the design for Table 1.46 can be treated as a multivariate one-way design with NOISE as the grouping variable. The model can be written as

$$\mathbf{Y}_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

where $\mathbf{Y}_{ij} = (Y_{ij1} \ldots Y_{ijh})'$, α_i is the treatment effect and the ϵ_{ij} are the errors (assumed to be independent with an h-variate normal distribution having mean $\mathbf{0}$ and a covariance matrix Σ). As long as Σ is positive definite, the covariance structure of the Y_{ijk} can have any pattern. This assumption is of course much less restrictive than the mixed-model assumption of compound symmetry.

The following SPSS MANOVA commands can be used to perform a multivariate analysis of the repeated measures data in Table 1.46.

```
MANOVA          Y1 TO Y9 BY NOISE(1,2)/
                WSFACTORS = PERIOD(3), DIAL(3)/
                WSDESIGN = PERIOD DIAL PERIOD BY DIAL/
                PRINT = SIGNIF(BRIEF)/
                ANALYSIS(REPEATED)/
                DESIGN = NOISE
```

Variables Y1 to Y9 are the nine response variables. The WSFACTORS subcommand indicates that there are two within-subjects factors, each having three levels. The order in which the variables are specified in the WSFACTORS list is very important since it indicates the levels of PERIOD and DIAL corresponding to Y1 to Y9. The following table gives the correspondence between the variables:

Variable	PERIOD	DIAL
Y_1	1	1
Y_2	1	2
Y_3	1	3
Y_4	2	1
Y_5	2	2
Y_6	2	3
Y_7	3	1
Y_8	3	2
Y_9	3	3

If the order of the two within-subjects factors is reversed in the WSFACTORS subcommand, the PERIOD and DIAL headings must be interchanged in the above table. For example, Y_7 would correspond to DIAL level 3 and PERIOD 1.

The WSDESIGN subcommand specifies the model for the within-subjects factors. The model fit need not be saturated. To specify an additive model, use

```
WSDESIGN = PERIOD DIAL/
```

The subcommand ANALYSIS(REPEATED) indicates that a repeated measures analysis is desired. The model for the between-subjects factors is specified, as always, by the DESIGN subcommand. Since there is only one between-subjects factor in this experiment, the command is DESIGN = NOISE.

The subcommand PRINT = SIGNIF(BRIEF) requests printing of brief multivariate output. Excerpts from this output are shown in Figure 1.49.

Figure 1.49

```
TESTS OF SIGNIFICANCE FOR Y1 USING SEQUENTIAL SUMS OF SQUARES

SOURCE OF VARIATION                    SUM OF SQUARES     DF   MEAN SQUARE          F     SIG. OF F

WITHIN CELLS                             2491.11111      4     622.77778
CONSTANT                               105868.16667      1  105868.16667  169.99349          .000
NOISE                                     468.16667      1     468.16667     .75174          .435

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

TESTS OF SIGNIFICANCE FOR WITHIN CELLS USING SEQUENTIAL SUMS OF SQUARES

SOURCE OF VARIATION                  WILKS LAMBDA  APPROX MULT F    SIG. OF F   AVERAGED F    SIG. OF F

PERIOD                                    .05060       28.14526          .011     63.38884          .000
NOISE AND PERIOD                          .15607        8.11102          .062      5.67077          .029

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

TESTS OF SIGNIFICANCE FOR WITHIN CELLS USING SEQUENTIAL SUMS OF SQUARES

SOURCE OF VARIATION                  WILKS LAMBDA  APPROX MULT F    SIG. OF F   AVERAGED F    SIG. OF F

DIAL                                      .01614       91.45623          .002     89.82316          0.0
NOISE AND DIAL                            .56498        1.15495          .425      1.90737          .210

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

TESTS OF SIGNIFICANCE FOR WITHIN CELLS USING SEQUENTIAL SUMS OF SQUARES

SOURCE OF VARIATION                  WILKS LAMBDA  APPROX MULT F    SIG. OF F   AVERAGED F    SIG. OF F

PERIOD BY DIAL                            .00075      331.44500          .041      .33566          .850
NOISE AND PERIOD BY DIAL                  .00043      581.87500          .031      .35664          .836
```

Wilks' lambda (with the corresponding approximate F) can be used to test for within- subjects factor effects, if the compound symmetry assumption appears to be violated. The averaged F statistics in the output are identical to the univariate mixed-model results displayed in Figure 1.47b.

Testing the hypothesis of compound symmetry is equivalent to testing the hypothesis that the covariance matrix of the transformed variables is a diagonal matrix (Bock, 1975, p. 459). Thus, the Bartlett test for sphericity can be used. MANOVA performs this Bartlett test for the transformed variables if the TRANSFORM or WSDESIGN subcommand is present.

MANOVA also performs the analysis of covariance on repeated measures data. If the covariates are constant over the repeated measures, only between-subject factors are adjusted; and if the covariates vary across the repeated measures, both between- and within-subjects factors are adjusted.