MANOVA (PROC GLM)

Unlike the exploratory diagnostic procedures covered previously, Multivariate Analysis of Variance (MANOVA) is an inferential procedure. The purpose of MANOVA is to use several responses simultaneously to discern significant differences among controlled experimental treatments. The analysis is very similar to its univariate counterpart, ANOVA, although some of the test statistics are different. Users should be aware that conducting a MANOVA is analogous to carrying out several ANOVAs at once and, therefore, the sample size requirements are multiplied with each multivariate response.

The SAS code required for MANOVA is an extension of the univariate ANOVA procedure, PROC GLM:

```
PROC GLM <optons>;
      CLASS <class vars>;
      MODEL <var1 var2 var3 ... var n> = <class vars>;

      MANOVA H = <options> E = <effect> M = <equation>/<detail options>;
```

The procedure options, **CLASS** statement and **MODEL** statement are all unchanged from the ANOVA GLM case. See the handout on GLM if you are unfamiliar with these statements. The multivariate nature of the analysis is introduced with the **MANOVA** statement. This statement is similar to the **TEST** statement in univariate ANOVA in that a hypothesis effect is defined with the H = syntax, and an error term can be supplied with E = syntax. If the error term is omitted, the model residual is used by default. As with the TEST statement, you can specify as many MANOVA statements as needed.

The options for the hypothesis, H=, are either a specific effect from the model or _ALL_ which indicates all effects should be tested. The third MANOVA item, M=, is not required, but can be used to specify a linear combination of the response variable to test. This option will be demonstrated later. The detail options are mainly related to printing control such as printing the hypothesis and error matrices used in the tests.

Example

The SAS code for a MANOVA testing the experimental factors Flour Concentration, Peak Temperature, and Heating Rate is given below.

```
PROC GLM DATA=EXP_ALL;
      CLASS FLOUR_CONC HEATING_RATE PEAK_TEMP;
      MODEL PEAK_VISC TROUGH_VISC FINAL_VISC BREAKDOWN TOTAL_SETBACK
      TIMEPEAK_VISC = FLOUR_CONC|HEATING_RATE|PEAK_TEMP;
      MANOVA H = _ALL_  ;
```

We are testing all main effects and interactions. The MANOVA statement calls for multivariate tests on all effects and assumes the residual to be the error term. No detail options are requested. The printed output from MANOVA is quite large. The first sections printed are the standard univariate ANOVA results for each response. These are interpreted in the usual manner and, hence, will not be covered here. The following sections give the multivariate results for each effect. A portion of these is shown below.

Characteristic Roots and Vectors of: E Inverse * H, where
H = Type III SSCP Matrix for Flour_Conc
E = Error SSCP Matrix

Characteristic Root	Percent	Characteristic Vector V'EV=1 Peak_Visc	Trough_Visc	Final_Visc	Breakdown	Total_Setback	TimePeak_Visc
134.711152	99.74	**0.00995039**	**-0.00817237**	0.00191232	**-0.00718055**	0.00000000	0.00375331
0.350745	0.26	-0.01206288	0.01075131	0.00339263	0.01046699	0.00000000	-0.00140710
0.000000	0.00	-0.13631959	0.13674894	-0.00046523	0.13641576	0.00000000	-0.00278206
0.000000	0.00	0.00000107	-0.00000107	0.00000000	-0.00000107	0.00000000	0.00000002
0.000000	0.00	0.00266422	-0.00277201	0.00017186	-0.00216898	0.00000000	0.04395392

MANOVA Test Criteria and F Approximations for the Hypothesis of **No Overall Flour_Conc Effect**
H = Type III SSCP Matrix for **Flour_Conc**
E = Error SSCP Matrix

S=2 M=1 N=72

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.00545521	**366.15**	**10**	**292**	**<.0001**
Pillai's Trace	1.25229903	49.24	10	294	<.0001
Hotelling-Lawley Trace	135.06189658	1963.05	10	216.27	<.0001
Roy's Greatest Root	134.71115194	3960.51	5	147	<.0001

NOTE: F Statistic for Roy's Greatest Root is an upper bound.
NOTE: F Statistic for Wilks' Lambda is exact.

This section provides multivariate test results for the Flour Concentration effect. The initial printout produces results similar to the PCA. The characteristic root is the square root of an eigenvalue. The first root accounts for 99.7% of the variability in the effect of Flour Concentration. The rows of this table represent the characteristic vector, i.e. eigenvector, of each root. Similar to the PCA technique, these are interpreted as loadings. They tell the user which variables were influencing Flour Concentration the most. From this we see that Peak Viscosity, Trough Viscosity, and Breakdown have the largest magnitudes. Note that the loadings for Total Setback are always zero. This occurs because that variable is completely redundant in the responses. Total Setback is computed as the difference in Final and Trough Viscosities. In a real analysis, this would be omitted.

The last portion of this printout shows the test results for the Flour Concentration effect. SAS produces four test statistics for each effect. Typically, it does not make any difference which one is used, however, Wilks Lambda is a good general choice. This statistic is an F test and is interpreted in the same manner as the univariate version. Here, the F value is 366.15 with 10 and 292 degrees of freedom. The p-value is <0.0001 which is highly significant. The remaining sections of the complete printout give similar results for each main effect and interaction.

To demonstrate the use of the M= option in the MANOVA statement, consider the response Total Setback again. This variable is specified as Final Viscosity - Trough Viscosity. Thus, a MANOVA statement can be set up to test this as:

```
MANOVA H = FLOUR_CONC*HEATING_RATE M = (0 -1 1 0 0 0);
```

Only the interaction between Flour Concentration and Heating Rate is being tested in this case. The order of the responses in the model statement was:

PEAK_VISC TROUGH_VISC FINAL_VISC BREAKDOWN TOTAL_SETBACK TIMEPEAK_VISC

Therefore, the coefficients in the M= option compute the difference of Final and Trough viscosities or Total Setback. The test results from this statement are:

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.93443809	1.32	8	150	**0.2398**
Pillai's Trace	0.06556191	1.32	8	150	0.2398
Hotelling-Lawley Trace	0.07016185	1.32	8	150	0.2398
Roy's Greatest Root	0.07016185	1.32	8	150	0.2398

This indicates that the interaction for Total Setback is non-significant, which is comparable to the univariate ANOVA for Total Setback below:

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Flour_Conc	2	122445.7320	61222.8660	160.64	<.0001
Heating_Rate	4	9286.0933	2321.5233	6.09	0.0001
Flour_Con*Heating_Ra	**8**	**3890.8029**	**486.3504**	**1.28**	**0.2600**
Peak_Temp	4	354182.2642	88545.5660	232.32	<.0001
Flour_Conc*Peak_Temp	8	62982.3505	7872.7938	20.66	<.0001
Heating_Ra*Peak_Temp	16	28383.0276	1773.9392	4.65	<.0001
Flour_*Heatin*Peak_T	32	14409.7669	450.3052	1.18	0.2502