

How can the MANOVA procedure in SAS be used in statistical analysis?

Authored by
stats writer

June 30, 2024

RECOMMENDED CITATION

stats writer (2024). *How can the MANOVA procedure in SAS be used in statistical analysis?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=160813>

The MANOVA procedure in SAS (Statistical Analysis System) is a powerful tool for conducting statistical analysis. MANOVA stands for Multivariate Analysis of Variance, and it allows for the simultaneous analysis of multiple dependent variables in relation to one or more independent variables. This procedure is useful in situations where there are multiple outcome variables that are related to one another, and when there is a need to examine the overall effect of an independent variable on these variables. The MANOVA procedure in SAS can be used to test for significant differences between groups, to identify relationships between variables, and to determine the strength and direction of these relationships. It is particularly beneficial for analyzing data sets with complex and interrelated variables, making it a valuable tool for researchers and statisticians in various fields. By using the MANOVA procedure in SAS, researchers can gain a deeper understanding of their data and make more informed decisions based on their analysis.

MANOVA | SAS Annotated Output

This page shows an example of multivariate analysis of variance (MANOVA) in SAS with footnotes explaining the output. The data used in this example are from the following experiment.

A researcher randomly assigns 33 subjects to one of three groups. The first group receives technical dietary information interactively from an on-line website. Group 2 receives the same information from a nurse practitioner, while group 3 receives the information from a video tape made by the same nurse practitioner. Each subject then made three ratings:

difficulty, usefulness, and importance of the information in the presentation. The researcher looks at three different ratings of the presentation (difficulty, usefulness and importance) to determine if there is a difference in the modes of presentation. In particular, the researcher is interested in whether the interactive website is superior because that is the most cost-effective way of delivering the information. In the dataset, the ratings are presented in the variables useful, difficulty and importance. The variable group indicates the group to which a subject was assigned.

We are interested in how the variability in the three ratings can be explained by a subject's group. Group is a categorical variable with three possible values: 1, 2 or 3. Because we have multiple dependent variables that cannot be combined, we will choose to use MANOVA. Our null hypothesis in this analysis is that a subject's group has no effect on

any of the three different ratings, and we can test this hypothesis on the dataset, `manova.sas7bdat` .

We can start by examining the three outcome variables.

```
data manova;  
set "C:tempmanova";  
run;
```

```
proc means data = manova;  
var useful difficulty importance;  
run;
```

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
USEFUL	33	16.3303030	3.2924615	11.8999996	24.2999992
DIFFICULTY	33	5.7151515	2.0175978	2.4000001	10.2500000
IMPORTANCE	33	6.4757576	3.9851309	0.2000000	18.7999992

```
proc freq data = manova;
table group;
run;
```

The FREQ Procedure

Cumulative Cumulative

GROUP Frequency Percent Frequency Percent

```
1 11 33.33 11 33.33
2 11 33.33 22 66.67
3 11 33.33 33 100.00
```

```
proc sort data = manova;
by group;
run;
proc means data = manova;
by group;
var useful difficulty importance;
run;
```

GROUP=1

The MEANS Procedure

Variable N Mean Std Dev Minimum Maximum

```
USEFUL 11 18.1181817 3.9037974 13.0000000
```

24.2999992

DIFFICULTY 11 6.1909091 1.8997129 3.7500000

10.2500000

IMPORTANCE 11 8.6818181 4.8630890 3.3000000

18.7999992

GROUP=2

Variable N Mean Std Dev Minimum Maximum

USEFUL 11 15.5272729 2.0756162 12.8000002

19.7000008

DIFFICULTY 11 5.5818183 2.4342631 2.4000001

9.8500004

IMPORTANCE 11 5.1090909 2.5311873 0.2000000

8.5000000

GROUP=3

Variable N Mean Std Dev Minimum Maximum

USEFUL 11 15.3454545 3.1382682 11.8999996

19.7999992

DIFFICULTY 11 5.3727273 1.7590287 2.6500001

8.7500000

IMPORTANCE 11 5.6363637 3.5469065 0.7000000

10.3000002

Next, we can enter our MANOVA command. In SAS, MANOVA is an option within `proc glm`, the generalized linear model procedure. We use the `class` statement to indicate our categorical predictor variable group, then specify our model by listing our outcome variables to the left of the equal sign and our predictor to the right. We are only interested in type III sum of squares, which we indicate with the `SS3` option. In the `manova` statement, we indicate that our hypothesized effect, represented in SAS as `h`, is `group`.

```
proc glm data = manova;  
class group;  
model useful difficulty importance = group / SS3;  
manova h = group;  
run;
```

The GLM Procedure

Class Level Information

Class Levels Values

GROUP 3 1 2 3**Number of Observations Read 33****Number of Observations Used 33****Dependent Variable: USEFUL****Sum of****Source DF Squares Mean Square F Value Pr > F****Model 2 52.9242378 26.4621189 2.70 0.0835****Error 30 293.9654425 9.7988481****Corrected Total 32 346.8896803****R-Square Coeff Var Root MSE USEFUL Mean****0.152568 19.16873 3.130311 16.33030****Source DF Type III SS Mean Square F Value Pr > F****GROUP 2 52.92423783 26.46211891 2.70 0.0835****Dependent Variable: DIFFICULTY****Sum of****Source DF Squares Mean Square F Value Pr > F****Model 2 3.9751512 1.9875756 0.47 0.6282****Error 30 126.2872767 4.2095759****Corrected Total 32 130.2624279**

R-Square Coeff Var Root MSE DIFFICULTY Mean
0.030516 35.89975 2.051725 5.715152

Source DF Type III SS Mean Square F Value Pr > F
GROUP 2 3.97515121 1.98757560 0.47 0.6282

Dependent Variable: IMPORTANCE

Sum of

Source DF Squares Mean Square F Value Pr > F
Model 2 81.8296936 40.9148468 2.88 0.0718
Error 30 426.3708962 14.2123632
Corrected Total 32 508.2005898

R-Square Coeff Var Root MSE IMPORTANCE Mean
0.161018 58.21603 3.769929 6.475758

Source DF Type III SS Mean Square F Value Pr > F
GROUP 2 81.82969356 40.91484678 2.88 0.0718

Multivariate Analysis of Variance

Characteristic Roots and Vectors of: E Inverse * H,
where

H = Type III SSCP Matrix for GROUP

E = Error SSCP Matrix

Characteristic Characteristic Vector V'EV=1**Root Percent USEFUL DIFFICULTY IMPORTANCE**

0.89198790 99.42 0.06410227 -0.00186162 0.05375069
 0.00524207 0.58 0.01442655 0.06888878 -0.02620577
 0.00000000 0.00 -0.03149580 0.05943387 0.01270798

MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall GROUP Effect**H = Type III SSCP Matrix for GROUP****E = Error SSCP Matrix****S=2 M=0 N=13****Statistic Value F Value Num DF Den DF Pr > F****Wilks' Lambda 0.52578838 3.54 6 56 0.0049****Pillai's Trace 0.47667013 3.02 6 58 0.0122****Hotelling-Lawley Trace 0.89722998 4.12 6 35.61 0.0031****Roy's Greatest Root 0.89198790 8.62 3 29 0.0003****NOTE: F Statistic for Roy's Greatest Root is an upper bound.****NOTE: F Statistic for Wilks' Lambda is exact.****Class Level Information****The GLM Procedure**

Class Level Information**Classa Levelsb Valuesc****GROUP 3 1 2 3****Number of Observations Read 33****Number of Observations Used 33**

a. Class - This is the categorical predictor variable in the MANOVA.

b. Levels - This is the number of possible values of the specified predictor. Our predictor in this example has three levels (group = 1, group = 2 and group = 3).

c. Values - These are the values of the predictor.

Univariate Outputd**Dependent Variablee: USEFUL****Sum of****Sourcef DFg Squaresh Mean Squarei F Valuej Pr > Fk****Model 2 52.9242378 26.4621189 2.70 0.0835****Error 30 293.9654425 9.7988481**

Corrected Total 32 346.8896803

**R-Squarel Coeff Varm Root MSEn USEFUL Meano
0.152568 19.16873 3.130311 16.33030**

**Source DF Type III SSp Mean Square F Value Pr > F
GROUP 2 52.92423783 26.46211891 2.70 0.0835**

d. Univariate Output - Within MANOVA, SAS provides both univariate and multivariate

output. The univariate results are presented separately for each dependent variable.

Here, we see the univariate output for useful (the univariate output for difficulty and importance have been excluded to increase readability).

Within each set of output for a dependent variable, there are two sets of

results. The first set of results matches a one-way ANOVA using the MANOVA predictor and the single dependent variable. The second set of results presents the type III sum of squares results.

e. Dependent Variable - This is one of the dependent variables from the MANOVA.

f. Source - This is the source of the variability in the specified dependent variable.

g. DF - This is the degrees of freedom. Because our predictor, group, has 3 levels, the degrees of freedom associated with the model is 2.

h. Sum of Squares - These are the model, error, and total sum of squares. The model sum of squares is the sum of the squared differences between the predicted values and the mean of the outcome variable. The error sum of squares is the sum of the squared differences between the predicted values and the outcome values. The total sum of squares is the sum of the model and error sums of squares.

i. Mean Square - This is the sum of squares divided by

the degrees of freedom (see g and h).

j. F Value - This is the F statistic associated with the given source.

k. Pr > F - This is the p-value associated with the F statistic of a given source. The null hypothesis that the predictor has no effect on the outcome variable is evaluated with regard to this p-value. For a given alpha level, if the p-value is less than alpha, the null hypothesis is rejected. If not, then we fail to reject the null hypothesis.

l. R-Square - This is the proportion of variability in the dependent variable (useful) that can be explained by the model. It is the ratio of the model sum of squares to the total sum of squares.

m. Coeff Var - This is the coefficient of variation expressed as a percent. The proportion can be calculated as the ratio of

the root mean squared error to the mean of the outcome variable (see n and o), expressed as a percent. It describes the amount of variation in the outcome variable.

n. Root MSE - This is the square root of the Mean Square.

o. USEFUL mean - This is the mean value of the dependent variable.

p. Type III SS - This is a type of sum-of-squares calculation. Here, we are looking at the sum of squares of the predictor, group. Because our model consists of just one predictor, the sum of squares of the predictor is the same as the model sum of squares. Type III sum of squares are calculated for each predictor as if it is the last predictor added to the model. However, in this example, we only have one predictor, and we can see that the Type III sum of squares matches the sum of squares from the ANOVA.

MANOVA Output

Multivariate Analysis of Variance

Characteristic Roots and Vectors of: $E^{-1}H$,
where

H = Type III SSCP Matrix for GROUP

E = Error SSCP Matrix

Characteristic Vector $V'EV=1$

Root	Percent	USEFUL	DIFFICULTY	IMPORTANCE
0.89198790	99.42	0.06410227	-0.00186162	0.05375069
0.00524207	0.58	0.01442655	0.06888878	-0.02620577
0.00000000	0.00	-0.03149580	0.05943387	0.01270798

MANOVA Test Criteria and F Approximations for the
Hypothesis of No Overall GROUP Effect

H = Type III SSCP Matrix for GROUP

E = Error SSCP Matrix

S=2 M=0 N=13s

Statistic	Value	F	Value	Num	DFz	Den	DFaa	Pr > Fab
Wilks' Lambda	0.52578838	3.54	6	56	0.0049			
Pillai's Trace	0.47667013	3.02	6	58	0.0122			
Hotelling-Lawley Trace	0.89722998	4.12	6	35.61	0.0031			

Roy's Greatest Rootx 0.89198790 8.62 3 29 0.0003

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

NOTE: F Statistic for Wilks' Lambda is exact.

q. Characteristic Root -

These are the eigenvalues of the product of the sum-of-squares matrix of the model and the sum-of-squares matrix of the error.

There is one eigenvalue for each of the eigenvectors of the product of the model sum of squares matrix and the error sum of squares matrix, a 3×3 matrix. The percents listed next to the characteristic roots indicate the amount of variability in the outcomes a given root and vector account for. In this example, the first root and vector account for 99.42% of the variability in the outcomes and the second for .58% of the variability in the outcomes.

r. Characteristic Vector - These are the eigenvectors of

the product of the sum-of-squares matrix of the model and the sum-of-squares matrix of the error. The three numbers that compose a vector can be read across a row (one under useful, one under difficulty, and one under importance).

s. $S=2$ $M=0$ $N=13$ - These are intermediate results that are used in computing the multivariate test statistics and their associated degrees of freedom. If P is the number of dependent variables, Q is the hypothesis degrees of freedom, and NE is the residual or error degrees of freedom, then $S = \min(P, Q)$, $M = .5(\text{abs}(P-Q)-1)$ and $N = .5(NE-P-1)$.

t. **Statistic** - MANOVA calculates four multivariate test statistics.

All four are based on the characteristic roots (see superscript q). The null hypothesis for each of these tests is the same: the independent variable (group) has no effect on any of the dependent variables (useful,

**difficulty
and importance).**

u. Wilks' Lambda - This can be interpreted as the proportion of the variance in the outcomes that is not explained by an effect. To calculate Wilks' Lambda, for each characteristic root, calculate $1/(1 + \text{the characteristic root})$, then find the product of these ratios. So in this example, you would first calculate $1/(1+0.89198790) = 0.5285446$, $1/(1+0.00524207) = 0.9947853$, and $1/(1+0)=1$. Then multiply $0.5285446 * 0.9947853 * 1 = 0.52578838$.

v. Pillai's Trace - This is another one of the four multivariate test statistics used in MANOVA. To calculate Pillai's trace, divide each characteristic root by $1 + \text{the characteristic root}$, then sum these ratios. So in this example, you would first calculate $0.89198790/(1+0.89198790) = 0.471455394$, $0.00524207/(1+0.00524207) =$

0.005214734, and $0/(1+0)=0$.

When these are added we arrive at Pillai's trace: $(0.471455394 + 0.005214734 + 0) = 0.47667013$.

w. Hotelling-Lawley Trace - This is very similar to Pillai's Trace. It is the sum of the roots of the product of the sum-of-squares matrix of the model and the sum-of-squares matrix of the error for the two linear regression functions and is a direct generalization of the F statistic in ANOVA. We can calculate the Hotelling-Lawley Trace by summing the characteristic roots listed in the output: $0.89198790 + 0.00524207 + 0 = 0.89723$.

x. Roy's Greatest Root - This is the largest of the roots of the product of the sum-of-squares matrix of the model and the sum-of-squares matrix of the error for the two linear regression functions. We can see that the value of Roy's Greatest Root is the largest of the

characteristic roots (see superscript q). Because it is a maximum, it can behave differently from the other three test statistics. In instances where the other three are not significant and Roy's is significant, the effect should be considered non-significant. For further information on the calculations underlying MANOVA results, consult SAS online documentation .

y. F Value - This is the F statistic for the given predictor and test statistic.

z. Num DF - This is the number of degrees of freedom in the model.

aa. Den DF - This is the number of degrees of freedom associated with the model errors. Note that there are instances in MANOVA when the degrees of freedom may be a non-integer (here, the DF associated with Hotelling-Lawley

Trace is a non-integer) because these degrees of freedom are calculated using the mean squared errors, which are often non-integers.

ab. $Pr > F$ - This is the p-value associated with the F statistic of a given effect and test statistic. The null hypothesis that a given predictor has no effect on either of the outcomes is evaluated with regard to this p-value.

For a given alpha level, if the p-value is less than alpha, the null hypothesis is rejected. If not, then we fail to reject the null hypothesis. In this example, we reject the null hypothesis that group has no effect on useful, difficulty or importance scores at alpha level .05 because the p-values are all less than .05.