

# What is the Stata output for a Multivariate Analysis of Variance (MANOVA)?

Authored by  
**stats writer**

June 30, 2024

## RECOMMENDED CITATION

stats writer (2024). *What is the Stata output for a Multivariate Analysis of Variance (MANOVA)?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=160801>

The Stata output for a Multivariate Analysis of Variance (MANOVA) is a statistical summary of the results from conducting a MANOVA in Stata software. It includes various tables and graphs that display the key findings and statistical measures, such as the F-statistic, p-value, and effect sizes, for all dependent variables and independent variables included in the analysis. This output provides important information on the overall significance of the MANOVA, as well as the specific effects of each independent variable on the dependent variables. It is used to interpret and communicate the results of the MANOVA analysis, allowing researchers to gain insights into the relationships between multiple dependent and independent variables.

## MANOVA | Stata Annotated Output

**This page shows an example of multivariate analysis of variance (MANOVA) in Stata 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 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.

use <https://stats.idre.ucla.edu/stat/stata/dae/manova>,  
clear

We can start by examining the three outcome variables.

Note that Stata labels

group 1 as the treatment group, group 2 as control\_1,  
and group 3 as  
control\_2.

summarize useful difficulty importance

Variable | Obs Mean Std. Dev. Min Max

```
-----+-----
useful | 33 16.3303 3.292461 11.9 24.3
difficulty | 33 5.715152 2.017598 2.4 10.25
importance | 33 6.475758 3.985131 .2 18.8
```

tabulate group, nolabel

group | Freq. Percent Cum.

```
-----+-----
treatment | 11 33.33 33.33
control_1 | 11 33.33 66.67
control_2 | 11 33.33 100.00
-----+-----
```

**Total | 33 100.00**

**tabstat difficulty useful importance, by(group)**

**Summary statistics: mean  
by categories of: group**

**group | diffic~y useful import~e**

group	diffic~y	useful	import~e
treatment	6.190909	18.11818	8.681818
control_1	5.581818	15.52727	5.109091
control_2	5.372727	15.34545	5.636364
<b>Total</b>	<b>5.715152</b>	<b>16.3303</b>	<b>6.475758</b>

Next, we can enter our MANOVA command.

**manova difficulty useful importance = group**  
Number of obs = 33

**W = Wilks' lambda L = Lawley-Hotelling trace**  
**P = Pillai's trace R = Roy's largest root**

**Source | Statistic df F(df1, df2) = F Prob>F**

-----+-----

```

group | W 0.5258 2 6.0 56.0 3.54 0.0049 e
| P 0.4767 6.0 58.0 3.02 0.0122 a
| L 0.8972 6.0 54.0 4.04 0.0021 a
| R 0.8920 3.0 29.0 8.62 0.0003 u

```

```

|-----

```

```

Residual | 30

```

```

-----+-----

```

```

Total | 32

```

```

-----

```

**e = exact, a = approximate, u = upper bound on F**

As we look at our results, we will want to refer to the eigenvalues of the sum-of-squares matrix of the model and the sum-of-squares matrix of the error.

These values will be informative in understanding the MANOVA output. To display the values, we ask Stata to list the matrix of eigenvalues from the model.

```

matrix list e(eigvals_m)

```

```

c1 c2

```

```

r1 .8919879 .00524207

```

**Eigenvaluesa****c1 c2****r1 .8919879 .00524207****MANOVA Outputb****Number of obs = 33****W = Wilks' lambdac L = Lawley-Hotelling tracee****P = Pillai's traced R = Roy's largest rootf****Sourceg | Statistich dfi F(df1, df2) = Fj Prob>Fk**

```

-----+-----
group | W 0.5258 2 6.0 56.0 3.54 0.0049 e
| P 0.4767 6.0 58.0 3.02 0.0122 a
| L 0.8972 6.0 54.0 4.04 0.0021 a
| R 0.8920 3.0 29.0 8.62 0.0003 u
|-----
Residual | 30

```

**Total | 32****e = exact, a = approximate, u = upper bound on Fj**

**a. Eigenvalues -** 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 three eigenvectors of the product of the model sum of squares matrix and the error sum of squares matrix, a 3×3 matrix.

Because only two are listed here, we can assume the third eigenvalue is zero.

These eigenvalues are among the saved results of our manova in Stata.

They are used in the calculation of the multivariate test statistics and are therefore useful to consider when looking at MANOVA output.

**b. MANOVA Output -** In Stata, MANOVA output includes four multivariate test statistics for each predictor variable. The four tests are listed above the output table. For each of the four test statistics, an F statistic and associated p-value are also displayed.

**c. 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 eigenvalue, calculate  $1/(1 + \text{the eigenvalue})$ , then find the product of these ratios. So in this example, you would first calculate  $1/(1+0.8919879) = 0.5285446$ ,  $1/(1+0.00524207) = 0.9947853$ , and  $1/(1+0)=1$ . Then multiply  $0.5285446 * 0.9947853 * 1 = 0.5258$ .

**d. Pillai's trace** - This is another multivariate test statistic. To calculate Pillai's trace, divide each eigenvalue by  $1 + \text{the characteristic root}$ , then sum these ratios. So in this example, you would first calculate  $0.8919879/(1+0.8919879) = 0.471455394$ ,  $0.00524207/(1+0.00524207) = 0.005214734$ , and  $0/(1+0)=0$ . When these are added we arrive at Pillai's trace:  $(0.471455394 + 0.005214734 + 0) = 0.4767$ .

**e. Lawley-Hotelling 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:  $0.8919879 + 0.00524207 + 0 = 0.8972$ .**

**f. Roy's largest 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. 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 insignificant.**

**g. Source** - This indicates the predictor variable in question. In our model, we are looking at group as a source of variability in the ratings.

**h. Statistic** - This is the test statistic for the given source listed in the prior column and the multivariate statistic indicated with the letter (W, P, L or R). For each independent variable, there are four multivariate test statistics calculated. See superscripts c, d, e and f.

**i. df** - This is the number degrees of freedom. Here, our predictor has three categories and our dataset has 33 observations, so we have 2 degrees of freedom for the hypothesis, 30 residual degrees of freedom, and 32 total degrees of freedom.

**j. F(df1, df2), F** - The first two columns (df1 and df2) list the degrees of freedom used in determining the F statistics.

### **The third column**

**lists the F statistic for the given source and multivariate test.**

**k. Prob > 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 the three different ratings at alpha level .05 because the p-values are all less than .05.**

**l. e = exact, a = approximate, u = upper bound on F - This indicates how the F statistic was calculated (whether it was an exact calculation, an approximation, or**

**an upper bound) for each of the multivariate tests.**

ARABPSYCHOLOGY.COM