“What is the concept of one-way MANOVA and how can it be applied in Stata for data analysis?”

Authored by

stats writer

June 29, 2024

RECOMMENDED CITATION

stats writer (2024). *“What is the concept of one-way MANOVA and how can it be applied in Stata for data analysis?”*. PSYCHOLOGICAL SCALES. Retrieved from https://scales.arabpsychology.com/?p=158676

One-way MANOVA (Multivariate Analysis of Variance) is a statistical technique used to analyze the relationship between a single dependent variable and multiple independent variables. It allows for the examination of the differences between groups on multiple dependent variables simultaneously. In Stata, one-way MANOVA can be applied by using the "manova" command, which allows for the comparison of means across groups and the detection of significant differences. This can be useful in various fields such as social sciences, healthcare, and business, where researchers may be interested in understanding the impact of several independent variables on a particular outcome. Overall, one-way MANOVA is a powerful tool for exploring complex relationships between variables and can provide valuable insights for data analysis.

One-way MANOVA | Stata Data Analysis Examples

Version info: Code for this page was tested in Stata 12.

MANOVA is used to model two or more dependent variables that are continuous with one or more categorical predictor variables.

Please note: The purpose of this page is to show how to use various data analysis commands. It does not cover all aspects of the research process which researchers are expected to do. In particular, it does not cover data cleaning and checking, verification of assumptions, model diagnostics or

potential follow-up analyses.

Examples of one-way multivariate analysis of variance

Example 1.

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. 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.

Example 2. A clinical psychologist recruits 100 people who suffer from

panic disorder into his study. Each subject receives one of four types of treatment for eight weeks. At the end of treatment, each subject participates in a structured interview, during which the clinical psychologist makes three ratings: physiological, emotional and cognitive. The clinical psychologist wants to know which type of treatment most reduces the symptoms of the panic disorder as measured on the physiological, emotional and cognitive scales. (This example was adapted from Grimm and Yarnold, 1995, page 246.)

Description of the data

Let's pursue Example 1 from above.

We have a data file, manova.dta, with 33 observations on three response variables. The response variables are ratings called useful, difficulty and importance. Level 1 of the group variable is the treatment group, level 2 is control group

1 and
level 3 is control group 2.

Let's look at the data. It is always a good idea to start with descriptive statistics.

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

summarize difficulty useful 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

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 | useful diffic~y import~e

----------+-----------------------------

treatment | 18.11818 6.190909 8.681818
control_1 | 15.52727 5.581818 5.109091
control_2 | 15.34545 5.372727 5.636364

----------+-----------------------------

Total | 16.3303 5.715152 6.475758

correlate useful difficulty importance
(obs=33)

| useful diffic~y import~e

-------------+--------------------------

useful | 1.0000
difficulty | 0.0978 1.0000
importance | -0.3411 0.1978 1.0000

Analysis methods you might consider

Below is a list of some analysis methods you may have encountered. Some of the methods listed are quite reasonable, while others have either fallen out of favor or have limitations.

One-way MANOVA

We will start by running the 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

Stata provides four multivariate tests by default. Each of these tests is statistically significant. For more information on these tests, please see our Stata Annotated Output: MANOVA page.

The overall multivariate test is significant, which means that differences between the levels of the variable group exist. To find where the differences lie, we will follow up with several post-hoc tests. We will begin with the multivariate test of group 1 versus the average of groups 2 and 3. First, we will use the manova, showorder command to determine the order of the elements in the design matrix. Knowing the order of the elements in the design matrix is

necessary to run the post-hoc tests. (Note that the order of the elements in the design matrix changed in Stata 11.)

manovatest, showorder

Order of columns in the design matrix
1: (group==1)
2: (group==2)
3: (group==3)
4: _cons

We will begin by comparing the treatment group (group 1) to an average of the control groups (groups 2 and 3). This tests the hypothesis that the mean control groups equals the treatment group. The output above indicates that the fourth element in the matrix is the constant, so in the matrix command below, we will set it to 0. Once we have created a matrix (which we call c1), we can use the manovatest command to test c1.

matrix c1=(2,-1,-1,0)

manovatest, test(c1)

Test constraint

(1) 2*1.group - 2.group - 3.group = 0

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

-----------+--

manovatest | W 0.5290 1 3.0 28.0 8.31 0.0004 e

| P 0.4710 3.0 28.0 8.31 0.0004 e

| L 0.8904 3.0 28.0 8.31 0.0004 e

| R 0.8904 3.0 28.0 8.31 0.0004 e

|--

Residual | 30

---

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

These results indicate that group 1 is statistically significantly different from the average of groups 2 and 3.

Now we will compare control group 1 (group 2) to

control group 2 (group 3). Again, we need to create a matrix (called c2 in this example) to do this comparison, and then use
that matrix in the manovatest command.

matrix c2=(0,1,-1,0)
manovatest, test(c2)

Test constraint
(1) 2.group - 3.group = 0

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
-----------+---
manovatest | W 0.9932 1 3.0 28.0 0.06 0.9785 e
| P 0.0068 3.0 28.0 0.06 0.9785 e
| L 0.0068 3.0 28.0 0.06 0.9785 e
| R 0.0068 3.0 28.0 0.06 0.9785 e
|---
Residual | 30
---
e = exact, a = approximate, u = upper bound on F

The results indicate that control group 1 is not statistically significantly different from control group 2.

We can use the margins command to obtain adjusted predicted values for each of the groups. In the first example below, we get the predicted means for the dependent variable difficulty. In the next two examples, we get the predicted means for the dependent variables useful and importance. These values can be helpful in seeing where differences between levels of the predictor variable are and describing the model.

margins group, predict(equation(difficulty))

Adjusted predictions Number of obs = 33

Expression : Linear prediction: difficulty, predict(equation(difficulty))

--

| Delta-method

| Margin Std. Err. z P>|z|

-------------+--

group |

1 | 6.190909 .6186184 10.01 0.000 4.978439 7.403379

2 | 5.581818 .6186184 9.02 0.000 4.369349 6.794288

3 | 5.372727 .6186184 8.69 0.000 4.160257 6.585197

----

margins group, predict(equation(useful))

Adjusted predictions Number of obs = 33

Expression : Linear prediction: useful, predict(equation(useful))

| Delta-method

| Margin Std. Err. z P>|z|

-------------+--

group |

1 | 18.11818 .9438243 19.20 0.000 16.26832 19.96804

2 | 15.52727 .9438243 16.45 0.000 13.67741 17.37713

3 | 15.34545 .9438243 16.26 0.000 13.49559 17.19532

----

margins group, predict(equation(importance))

Adjusted predictions Number of obs = 33

Expression : Linear prediction: importance, predict(equation(importance))

--

| Delta-method

| Margin Std. Err. z P>|z|

-------------+--

group |

1 | 8.681818 1.136676 7.64 0.000 6.453973 10.90966

2 | 5.109091 1.136676 4.49 0.000 2.881246 7.336936

3 | 5.636364 1.136676 4.96 0.000 3.408519 7.864208

--

In each of the three outputs above, we see that the predicted means for groups 2 and 3 are very similar; the predicted mean for group 1 is higher than those for groups 2 and 3.

In the examples below, we obtain the differences in the means for each of the

dependent variables for each of the control groups (groups 2 and 3) compared to the treatment group (group1). With respect to the dependent variable difficulty, the difference between the means for control group 1 versus the treatment group is approximately -0.61 (5.58 - 6.19). The difference between the means for control group 2 versus the treatment group is approximately -0.82 (5.37 - 6.19).

margins, dydx(group) predict(equation(difficulty))

Conditional marginal effects Number of obs = 33

Expression : Linear prediction: difficulty, predict(equation(difficulty))
dy/dx w.r.t. : 2.group 3.group

--

| Delta-method
| dy/dx Std. Err. z P>|z|
-------------+--
group |

2 | -.6090908 .8748585 -0.70 0.486 -2.323782 1.1056

3 | -.8181818 .8748585 -0.94 0.350 -2.532873 .8965094

--

Note: dy/dx for factor levels is the discrete change from the base level.

margins, dydx(group) predict(equation(useful))

Conditional marginal effects Number of obs = 33

Expression : Linear prediction: useful, predict(equation(useful))
dy/dx w.r.t. : 2.group 3.group

--

| Delta-method
| dy/dx Std. Err. z P>|z|
-------------+--

group |
2 | -2.590909 1.334769 -1.94 0.052 -5.207008 .0251907
3 | -2.772727 1.334769 -2.08 0.038 -5.388827 -.1566278

--

Note: dy/dx for factor levels is the discrete change from the base level.

margins, dydx(group) predict(equation(importance))

Conditional marginal effects Number of obs = 33

Expression : Linear prediction: importance, predict(equation(importance))
dy/dx w.r.t. : 2.group 3.group

--

| Delta-method
| dy/dx Std. Err. z P>|z|
-------------+--
group |
2 | -3.572727 1.607503 -2.22 0.026 -6.723375 -.4220792
3 | -3.045454 1.607503 -1.89 0.058 -6.196103 .1051936
--
Note: dy/dx for factor levels is the discrete change from the base level.

Finally, let's run separate univariate ANOVAs. We will use a foreach loop to run the ANOVA for each dependent variable.

foreach vname in difficulty useful importance {

anova `vname' group

}

/* useful */

Number of obs = 33 R-squared = 0.1526

Root MSE = 3.13031 Adj R-squared = 0.0961

Source | Partial SS df MS F Prob > F

-----------+--

Model | 52.9242378 2 26.4621189 2.70 0.0835

|

group | 52.9242378 2 26.4621189 2.70 0.0835

|

Residual | 293.965442 30 9.79884808

-----------+--

Total | 346.88968 32 10.8403025

/* difficulty */

Number of obs = 33 R-squared = 0.0305

Root MSE = 2.05173 Adj R-squared = -0.0341

Source | Partial SS df MS F Prob > F

-----------+--

Model | 3.97515121 2 1.9875756 0.47 0.6282

|

group | 3.97515121 2 1.9875756 0.47 0.6282

|

Residual | 126.287277 30 4.20957589

-----------+--

Total | 130.262428 32 4.07070087

/* importance */

Number of obs = 33 R-squared = 0.1610

Root MSE = 3.76993 Adj R-squared = 0.1051

Source | Partial SS df MS F Prob > F

-----------+--

Model | 81.8296936 2 40.9148468 2.88 0.0718

|

group | 81.8296936 2 40.9148468 2.88 0.0718

|

Residual | 426.370896 30 14.2123632

-----------+--

Total | 508.20059 32 15.8812684

While none of the three ANOVAs were statistically significant at the alpha = .05 level,
in particular, the F-ratio for difficulty was less than 1.

Things to consider

See also

References