What is multivariate regression analysis and how is it used in SPSS data analysis?

Authored by

stats writer

June 29, 2024

RECOMMENDED CITATION

stats writer (2024). *What is multivariate regression analysis and how is it used in SPSS data analysis?*. PSYCHOLOGICAL SCALES. Retrieved from https://scales.arabpsychology.com/?p=158870

Multivariate regression analysis is a statistical technique used to analyze the relationship between multiple independent variables and a dependent variable. In other words, it helps to understand how changes in one or more variables affect the outcome of another variable. This method is commonly used in data analysis to identify and measure the strength of relationships between variables and to make predictions about the dependent variable based on the independent variables. In SPSS data analysis, multivariate regression can be used to model and analyze complex data sets with multiple variables, allowing researchers to gain insights into the underlying relationships and patterns within the data. It is a powerful tool for understanding and interpreting data in various fields such as social sciences, healthcare, and business.

Multivariate Regression Analysis | SPSS Data Analysis Examples

As the name implies, multivariate regression is a technique that estimates a single regression model with more than one outcome variable. When there is more than one predictor variable in a multivariate regression model, the model is a multivariate multiple regression.

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 and potential follow-up analyses.

Examples of multivariate regression

Example 1. A researcher has collected data on three psychological variables, four academic variables (standardized test scores), and the type of educational program the student is in for 600 high school students. She is interested in how the set of psychological variables is related to the academic variables and the type of program the student is in.

Example 2. A doctor has collected data on cholesterol, blood pressure, and weight. She also collected data on the eating habits of the subjects (e.g., how many ounces of red meat, fish, dairy products, and chocolate consumed per week). She wants to investigate the relationship between the three measures of health and eating habits.

Example 3. A researcher is interested in determining what factors influence the health African Violet plants. She collects data on the average leaf diameter, the mass of the root ball, and the average diameter of the blooms, as well as how long the plant has been in its current container. For predictor variables, she measures several elements in the soil, as well as the amount of light and water each plant receives.

Description of the data

Let's pursue Example 1 from above. We have a hypothetical dataset with 600 observations on seven variables. he psychological variables are locus of control (locus_of_control), self-concept (self_concept), and motivation (motivation). The academic variables are standardized tests scores in reading (read), writing (write), and science (science), as well as a categorical variable (prog) giving the type of program the student is in (general, academic, or vocational).

Let's look at the data mvreg. Note that there are no missing values in this data set.

descriptives variables = locus_of_control self_concept motivation read write science.

Descriptive Statistics
N Minimum Maximum Mean Std. Deviation
locus_of_control 600 -2.00 2.21 0.0965 0.67028
self_concept 600 -2.53 2.09 0.0049 0.70551
motivation 600 -2.75 2.58 0.0039 0.82240
read 600 24.62 80.59 51.9018 10.10298
write 600 20.07 83.93 52.3848 9.72645

science 600 21.99 80.37 51.7633 9.70618
Valid N (listwise) 600

correlations variables = locus_of_control self_concept motivation.
Correlations
locus_of_control self_concept motivation
locus_of_control Pearson Correlation 1 0.171 0.245
Sig. (2-tailed) <0.001 <0.001
N 600 600 600
self_concept Pearson Correlation 0.171 1 0.289
Sig. (2-tailed) <0.001 <0.001
N 600 600 600
motivation Pearson Correlation 0.245 0.289 1
Sig. (2-tailed) <0.001 <0.001
N 600 600 600

correlations variables = read write science.
Correlations
read write science
read Pearson Correlation 1 0.629 0.691
Sig. (2-tailed) <0.001 <0.001
N 600 600 600
write Pearson Correlation 0.629 1 0.569

Sig. (2-tailed) <0.001 <0.001

N 600 600 600

science Pearson Correlation 0.691 0.569 1

Sig. (2-tailed) <0.001 <0.001

N 600 600 600

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.

Multivariate multiple regression, the focus of this page.
Separate OLS Regressions - You could analyze these data using separate OLS regression analyses for each outcome variable. The individual coefficients, as well as their standard errors will be the same as those produced by the multivariate regression. However, the OLS regressions will not produce multivariate results, nor will they allow for testing of coefficients across equations.
Canonical correlation analysis might be feasible if you don't want to consider one set of variables as outcome variables and the other set as predictor variables.

Multivariate regression

To conduct a multivariate regression in SPSS, we can use either of two commands, glm or manova. Using the lmatrix subcommand in the glm command, you can test if all of the equations, taken together, are statistically significant. The F-ratios and p-values for three multivariate criterion are given, including Wilks' lambda, Lawley-Hotelling trace, Pillai's trace, and Roy's largest root. We can get the regression coefficients from either the glm or the manova command by including the print subcommand with the keyword parameter.

Below we run the glm command. Notice that we have multiple dependent variables listed before the SPSS keyword with. (The SPSS keyword with indicates that continuous predictor variables will follow.) We use the lmatrix subcommand to request the test of overall model. Semi-colons are required between each predictor.

```
glm locus_of_control self_concept motivation with read write science
/print = parameters
/lmatrix 'multivariate test of entire model' write 1; read 1;
```

science 1.Multivariate Testsa

Effect Value F Hypothesis df Error df Sig.

Intercept Pillai's Trace 0.165 39.239b 3.000 594.000 <0.001

Wilks' Lambda 0.835 39.239b 3.000 594.000 <0.001

Hotelling's Trace 0.198 39.239b 3.000 594.000 <0.001

Roy's Largest Root 0.198 39.239b 3.000 594.000 <0.001

read Pillai's Trace 0.027 5.529b 3.000 594.000 <0.001

Wilks' Lambda 0.973 5.529b 3.000 594.000 <0.001

Hotelling's Trace 0.028 5.529b 3.000 594.000 <0.001

Roy's Largest Root 0.028 5.529b 3.000 594.000 <0.001

write Pillai's Trace 0.056 11.807b 3.000 594.000 <0.001

Wilks' Lambda 0.944 11.807b 3.000 594.000 <0.001

Hotelling's Trace 0.060 11.807b 3.000 594.000 <0.001

Roy's Largest Root 0.060 11.807b 3.000 594.000 <0.001

science Pillai's Trace 0.017 3.397b 3.000 594.000 0.018

Wilks' Lambda 0.983 3.397b 3.000 594.000 0.018

Hotelling's Trace 0.017 3.397b 3.000 594.000 0.018

Roy's Largest Root 0.017 3.397b 3.000 594.000 0.018

a Design: Intercept + read + write + science

b Exact statistic

Tests of Between-Subjects Effects

Source Dependent Variable Type III

Sum of Squares df Mean Square F Sig.

Corrected Model locus_of_control 45.230a 3 15.077 40.135 <0.001

self_concept 1.892b 3 0.631 1.269 0.284

motivation 30.586c 3 10.195 16.224 <0.001

Intercept locus_of_control 37.608 1 37.608 100.116 <0.001

self_concept 0.807 1 0.807 1.624 0.203

motivation 19.477 1 19.477 30.993 <0.001

read locus_of_control 4.802 1 4.802 12.784 <0.001

self_concept 0.257 1 0.257 0.517 0.472

motivation 3.893 1 3.893 6.195 0.013

write locus_of_control 5.391 1 5.391 14.352 <0.001

self_concept 0.351 1 0.351 0.705 0.401

motivation 11.928 1 11.928 18.982 <0.001

science locus_of_control 0.827 1 0.827 2.202 0.138

self_concept 0.623 1 0.623 1.254 0.263

motivation 2.670 1 2.670 4.249 0.040

Error locus_of_control 223.886 596 0.376

self_concept 296.259 596 0.497

motivation 374.542 596 0.628

Total locus_of_control 274.707 600

self_concept 298.165 600

motivation 405.138 600

Corrected Total locus_of_control 269.116 599

self_concept 298.151 599

motivation 405.129 599

a R Squared = 0.168 (Adjusted R Squared = 0.164)

b R Squared = 0.006 (Adjusted R Squared = 0.001)

c R Squared = 0.075 (Adjusted R Squared = 0.071)

Parameter Estimates

Dependent Variable Parameter B Std. Error t Sig. 95% Confidence Interval

Lower Bound Upper Bound

locus_of_control Intercept -1.555 0.155 -10.006 <0.001 -1.861 -1.250

read 0.013 0.004 3.575 <0.001 0.006 0.021

write 0.013 0.003 3.788 <0.001 0.006 0.020

science 0.005 0.004 1.484 0.138 -0.002 0.013

self_concept Intercept -0.228 0.179 -1.274 0.203 -0.579 0.123

read 0.003 0.004 0.719 0.472 -0.005 0.012

write -0.003 0.004 -0.840 0.401 -0.011 0.004

science 0.005 0.004 1.120 0.263 -0.004 0.013

motivation Intercept -1.119 0.201 -5.567 <0.001 -1.514 -0.724

read 0.012 0.005 2.489 0.013 0.003 0.021

write 0.019 0.004 4.357 <0.001 0.011 0.028

science -0.010 0.005 -2.061 0.040 -0.019 0.000

Contrast Results (K Matrix)a

Contrast Dependent Variable

locus_of_control self_concept motivation

L1 Contrast Estimate 0.013 -0.003 0.019

Hypothesized Value 0 0 0

Difference (Estimate - Hypothesized) 0.013 -0.003 0.019

Std. Error 0.003 0.004 0.004

Sig. <0.001 0.401 <0.001

95% Confidence Interval for Difference Lower Bound 0.006 -0.011 0.011

Upper Bound 0.020 0.004 0.028

L2 Contrast Estimate 0.013 0.003 0.012

Hypothesized Value 0 0 0

Difference (Estimate - Hypothesized) 0.013 0.003 0.012

Std. Error 0.004 0.004 0.005

Sig. <0.001 0.472 0.013

95% Confidence Interval for Difference Lower Bound 0.006 -0.005 0.003

Upper Bound 0.021 0.012 0.021

L3 Contrast Estimate 0.005 0.005 -0.010

Hypothesized Value 0 0 0

Difference (Estimate - Hypothesized) 0.005 0.005 -0.010
Std. Error 0.004 0.004 0.005
Sig. 0.138 0.263 0.040
95% Confidence Interval for Difference Lower Bound -0.002 -0.004 -0.019
Upper Bound 0.013 0.013 0.000
a Based on the user-specified contrast coefficients (L') matrix: multivariate test of entire model

Multivariate Test Results
Value F Hypothesis df Error df Sig.
Pillai's trace 0.220 15.745 9.000 1788.000 <0.001
Wilks' lambda 0.784 16.856 9.000 1445.791 <0.001
Hotelling's trace 0.269 17.707 9.000 1778.000 <0.001
Roy's largest root 0.244 48.565a 3.000 596.000 <0.001
a The statistic is an upper bound on F that yields a lower bound on the significance level.

Univariate Test Results
Source Dependent Variable Sum of Squares df Mean Square F Sig.
Contrast locus_of_control 45.230 3 15.077 40.135 <0.001
self_concept 1.892 3 0.631 1.269 0.284
motivation 30.586 3 10.195 16.224 <0.001

Error locus_of_control 223.886 596 0.376
self_concept 296.259 596 0.497
motivation 374.542 596 0.628

The tests for the overall model, shown in the first table above, indicate that the model is statistically significant, regardless of the type of multivariate criteria that is used (i.e., all of the p-values are less than 0.0001). Below the overall model tests are the between-subjects tests. The variables locus_of_control and motivation are statistically significant; the variable self_concept is not. In the third table, we see the parameter estimates for each of the predictor variables for each of the dependent variables. The fifth table gives the multivariate tests, which are all statistically significant, regardless of which test is used. The last table gives the univariate tests, two of which are statistically significant.

The output from the manova command contains most of the same information given by the glm command, but it is organized a little differently.
The first table gives the number of observations, number of parameters, RMSE, R-squared, F-ratio, and p-

value for each of the three models.

manova locus_of_control self_concept motivation with read write science
/print parameters.

- -

The default error term in MANOVA has been changed from WITHIN CELLS to
WITHIN+RESIDUAL. Note that these are the same for all full factorial designs.

*** * * * * * * * * * * * * * * * * * A n a l y s i s o f V a r i a n c e * * * * * * * * * * * * * * * * * ***

600 cases accepted.
0 cases rejected because of out-of-range factor values.
0 cases rejected because of missing data.
1 non-empty cell.

1 design will be processed.

- -

* * * * * * * * * * * * * * * * * * A n a l y s i s o f V a r i a n c e -- Design 1 * * * * * * * * * * * * * * * * *

EFFECT .. WITHIN CELLS Regression
Multivariate Tests of Significance (S = 3, M = -1/2, N = 296)

Test Name Value Approx. F Hypoth. DF Error DF Sig. of F

Pillais 0.22030 15.74505 9.00 1788.00 0.000
Hotellings 0.26889 17.70671 9.00 1778.00 0.000
Wilks 0.78439 16.85640 9.00 1445.79 0.000
Roys 0.19644

- -

EFFECT .. WITHIN CELLS Regression (Cont.)
Univariate F-tests with (3,596) D. F.

Variable Sq. Mul. R Adj. R-sq. Hypoth. MS Error MS F Sig. of F

locus_of 0.16807 0.16388 15.07664 0.37565 40.13509 0.000
self_con 0.00635 0.00135 0.63077 0.49708 1.26896 0.284

motivati 0.07550 0.07084 10.19548 0.62843 16.22382 0.000

- -

Regression analysis for WITHIN CELLS error term
--- Individual Univariate0.9500 confidence intervals
Dependent variable .. locus_of_control

COVARIATE B Beta Std. Err. t-Value Sig. of t Lower -95% CL- Upper

read 0.0133466809 0.2011716274 0.00373 3.57542 0.000 0.00602 0.02068
write 0.0129191741 0.1874705938 0.00341 3.78845 0.000 0.00622 0.01962
science 0.0054541421 0.0789802632 0.00368 1.48403 0.138 -0.00176 0.01267
Dependent variable .. self_concept

COVARIATE B Beta Std. Err. t-Value Sig. of t Lower -95% CL- Upper

read 0.0030876697 0.0442156218 0.00429 0.71906 0.472 -0.00535 0.01152

write -0.0032943533 -0.0454171665 0.00392 -0.83980 0.401 -0.01100 0.00441
science 0.0047345432 0.0651360875 0.00423 1.11988 0.263 -0.00357 0.01304
Dependent variable .. motivation

COVARIATE B Beta Std. Err. t-Value Sig. of t Lower -95% CL- Upper

read 0.0120168331 0.1476238599 0.00483 2.48889 0.013 0.00253 0.02150
write 0.0192165750 0.2272728015 0.00441 4.35678 0.000 0.01055 0.02788
science -0.0097985928 -0.1156455467 0.00475 -2.06130 0.040 -0.01913 -0.00046

- -

*** * * * * * * * * * * * * * * * * A n a l y s i s o f V a r i a n c e -- Design 1 * * * * * * * * * * * * * * * * ***

EFFECT .. CONSTANT
Multivariate Tests of Significance (S = 1, M = 1/2, N = 296)

Test Name Value Exact F Hypoth. DF Error DF Sig. of F

Pillais 0.16540 39.23945 3.00 594.00 0.000
Hotellings 0.19818 39.23945 3.00 594.00 0.000
Wilks 0.83460 39.23945 3.00 594.00 0.000
Roys 0.16540
Note.. F statistics are exact.

- -

EFFECT .. CONSTANT (Cont.)
Univariate F-tests with (1,596) D. F.

Variable Hypoth. SS Error SS Hypoth. MS Error MS F Sig. of F

locus_of 37.60832 223.88586 37.60832 0.37565 100.11601 0.000
self_con 0.80711 296.25868 0.80711 0.49708 1.62370 0.203
motivati 19.47692 374.54227 19.47692 0.62843 30.99314 0.000

- -

Estimates for locus_of_control adjusted for 3 covariates --- Individual univariate 0.9500 confidence intervals

CONSTANT

Parameter Coeff. Std. Err. t-Value Sig. t Lower -95% CL- Upper

1 -1.5552772264 0.15544 -10.00580 0.00000 -1.86055 -1.25001

- -

Estimates for self_concept adjusted for 3 covariates --- Individual univariate 0.9500 confidence intervals

CONSTANT

Parameter Coeff. Std. Err. t-Value Sig. t Lower -95% CL- Upper

1 -0.2278406241 0.17880 -1.27424 0.20307 -0.57900 0.12332

- -

Estimates for motivation adjusted for 3 covariates
--- Individual univariate 0.9500 confidence intervals

CONSTANT

Parameter Coeff. Std. Err. t-Value Sig. t Lower -95% CL- Upper

1 -1.1192470175 0.20104 -5.56715 0.00000 -1.51409 -0.72440

- -

Abbreviated Extended
Name Name

locus_of locus_of_control
motivati motivation
self_con self_concept

If you ran a separate OLS regression for each outcome variable, you would get exactly the same coefficients, standard errors, t- and p-values, and confidence intervals as shown above. So why conduct a multivariate regression? As we mentioned earlier, one

of the advantages of using multivariate regression is that you can conduct tests of the coefficients across the different outcome variables.

Things to consider

The residuals from multivariate regression models are assumed to be multivariate normal. This is analogous to the assumption of normally distributed errors in univariate linear regression (i.e., OLS regression).
Multivariate regression analysis is not recommended for small samples.
The outcome variables should be at least moderately correlated for the multivariate regression analysis to make sense.