How do I use the anovalator command in Stata?

Authored by

stats writer

July 1, 2024

The anovalator command in Stata is a statistical tool used for conducting analysis of variance (ANOVA) on a dataset. This command allows users to compare the means of three or more groups, and determine if there are significant differences between them. It can be used to analyze both continuous and categorical variables. To use the anovalator command, users must first load their dataset into Stata and then specify the variables to be included in the analysis. The command will then generate output tables that display the results of the ANOVA, including F and p values, as well as graphical representations such as box plots to aid in the interpretation of the data. Overall, the anovalator command is a useful tool for performing ANOVA analysis in Stata, and is commonly used in various fields such as social sciences, economics, and healthcare research.

How do I use the anovalator command? Stata FAQ

The anovalator command refers to a mythical program that displays the results of an estimation command in an anova-like manner, that is, separate multidegree of freedom tests for main effects, two-way and 3-way interactions. In addition, anovalation will do tests of simple main effects, pairwise comparisons and arbitrary linear contrasts.

You may download this program from the UCLA ATS Statistical Consulting Group by typing the following two commands into Stata's command window.

net from https://stats.idre.ucla.edu/stat/stata/ado/analysis

net install anovalator

Internally, anovalator makes extensive use of the margins command to compute the adjusted cell means. Here is the syntax for the anovalator command:

anovalatorvarlist

Well, that's a bit confusing, but it really breaks down to six simpler syntaxes.

anovalatorvarlist /* main effects */
anovalatorvarlist (max 2) /* two-way iteraction */
anovalatorvarlist (max 3) /* 3-way iteraction */
anovalatorvarlist (max 2) /* simple main effects */
anovalatorvarname /* pairwise comparisons */
anovalatorvarname /* linear contrasts */

Now for a bunch of caveats. The anovalator program should be considered experimental in that it has not been tested with every possible estimation command in Stata. In fact, it has been tested a lot with only a handful of estimation procedures. It does not have a "true" help file, just this web page. The program also does not

do as much internal consistency checking as it should. If you enter information incorrectly it may not catch it and issue a warning, it may just crash. You will then most likely have to rerun your estimation command. anovalator will try to do what you request even if it doesn't make any sense. It just doesn't know any better. Do not bother looking for the return results as there aren't any. Further, anovalator does not make any adjustments for multiple tests for pairwise comparisons or linear contrasts. This is left up to the user. To sum it up, the user takes full responsibility for using the command correctly. Please use it carefully.

After all of the above, why would anyone use this program? Well, it can be very useful in certain situations. Everything that anovalator does can, of course, be done manually after running the margins command with the post option. But, if you know what you are doing, anovalator can get you the results you are interested in

more easily and quicker than the manual approach.

Why does anovalator do main effects and two-way interactions? These tests are needed because factor variables in Stata 11 use indicator (dummy) coding for categorical predictor variables.
When using dummy coding in models with two-way or higher interactions the tests of the coefficients are not the tests of the main effects. This is true even if the interactions are not significant.
In models with three-way interactions the tests of the interaction coefficients are not the same as the test of the two-way interaction effect.

We can finally get on with the demonstration of anovalator on four different models.

Two-factor Regression Model

use https://stats.idre.ucla.edu/stat/data/hsbanova, clear

regress write grp##female

Source | SS df MS Number of obs = 200
-------------+------------------------------ F(7, 192) = 11.05

Model | 5135.17494 7 733.59642 Prob > F = 0.0000

Residual | 12743.7001 192 66.3734378 R-squared = 0.2872

-------------+------------------------------ Adj R-squared = 0.2612

Total | 17878.875 199 89.843593 Root MSE = 8.147

--

write | Coef. Std. Err. t P>|t|

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

grp |

2 | 7.31677 2.458951 2.98 0.003 2.466743 12.1668

3 | 10.10248 2.292658 4.41 0.000 5.580454 14.62452

4 | 16.75286 2.525696 6.63 0.000 11.77119 21.73453

|

1.female | 9.136876 2.311726 3.95 0.000 4.577236 13.69652

|

grp#female |

2 1 | -5.029733 3.357123 -1.50 0.136 -11.65131 1.591845

3 1 | -3.721697 3.128694 -1.19 0.236 -9.892723 2.449328

4 1 | -9.831208 3.374943 -2.91 0.004 -16.48793 -3.174482

|

_cons | 41.82609 1.698765 24.62 0.000 38.47545 45.17672

--

Main effects, two-way interaction and tests of simple main effects with F-ratios

F statistics are exact for models in which the disturbances are assumed to be normally distributed, as in the regression above. You can check the main effects and two-way interaction results by running the command: anova write grp##female.

anovalator grp female, main twoway simple fratio

anovalator main-effect for grp
chi2(3) = 54.866574 p-value = 7.331e-12
scaled as F-ratio = 18.288858

anovalator main-effect for female
chi2(1) = 14.830893 p-value = .00011759
scaled as F-ratio = 14.830893

anovalator two-way interaction for grp#female
chi2(3) = 8.6708394 p-value = .03400302
scaled as F-ratio = 2.8902798

anovalator test of simple main effects for grp at(female=0)
chi2(3) = 46.001924 p-value = 5.666e-10
scaled as F-ratio = 15.333975

anovalator test of simple main effects for grp at(female=1)
chi2(3) = 13.644845 p-value = .00343069
scaled as F-ratio = 4.5482816

Linear contrast and pairwise comparisons

We will start with a contrast amont the levels of grp that is the average of 1 &
2 versus the average of
3 & 4. Please note: It is
up to you to make sure that the weights for the contrast sum to zero. There is no internal
checking.
The contrast will be followed by all pairwise comparisons for grp.

If you don't want to see the table of adjusted group means from the margins command, just use the quiet option.

anovalator grp, wgt(1/2 1/2 -1/2 -1/2) pairwise quiet

anovalator contrast for grp

(1) .5*1bn.grp + .5*2.grp - .5*3.grp - .5*4.grp = 0

| Coef. Std. Err. z P>|z|

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

(1) | -7.638494 1.16622 -6.55 0.000 -9.924243 -5.352746

----

anovalator pairwise comparisons for grp

Comparison Coef. Std. Err. z P>|z|

1 vs 2 -4.8019 1.67856 -2.86 0.004 -8.091884 -1.511923

1 vs 3 -8.24164 1.56435 -5.27 0.000 -11.30776 -5.175516

1 vs 4 -11.8373 1.68747 -7.01 0.000 -15.1447 -8.529812

2 vs 3 -3.43973 1.61019 -2.14 0.033 -6.595705 -.2837594

2 vs 4 -7.03535 1.73006 -4.07 0.000 -10.42626 -3.644445

3 vs 4 -3.59562 1.61948 -2.22 0.026 -6.769794 -.4214472

Linear contrasts

Next we will do one more contrast; group 1 versus the average of 3 &amp 4.

anovalator grp, wgt(1 0 -1/2 -1/2) quiet

anovalator contrast for grp

(1) 1bn.grp - .5*3.grp - .5*4.grp = 0

| Coef. Std. Err. z P>|z|

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

(1) | -10.03945 1.411274 -7.11 0.000 -12.80549 -7.273399

Three-factor Regression Model

This second example involves a 2x2x3 factorial model run using regression.

use https://stats.idre.ucla.edu/stat/data/threeway, clear

regress y a##b##c

Source | SS df MS Number of obs = 24

-------------+------------------------------ F(11, 12) = 33.94

Model | 497.833333 11 45.2575758 Prob > F = 0.0000

Residual | 16 12 1.33333333 R-squared = 0.9689

-------------+------------------------------ Adj R-squared = 0.9403

Total | 513.833333 23 22.3405797 Root MSE = 1.1547

----

y | Coef. Std. Err. t P>|t|

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

2.a | -.5 1.154701 -0.43 0.673 -3.015876 2.015876

2.b | -.5 1.154701 -0.43 0.673 -3.015876 2.015876

|

a#b |

2 2 | 6.5 1.632993 3.98 0.002 2.942014 10.05799

|

c |

2 | 4 1.154701 3.46 0.005 1.484124 6.515876

3 | 8 1.154701 6.93 0.000 5.484124 10.51588

|

a#c |

2 2 | 1 1.632993 0.61 0.552 -2.557986 4.557986

2 3 | -1.10e-14 1.632993 -0.00 1.000 -3.557986 3.557986

|

b#c |

2 2 | -4 1.632993 -2.45 0.031 -7.557986 -.4420135

2 3 | -9 1.632993 -5.51 0.000 -12.55799 -5.442014

|
a#b#c |
2 2 2 | 3 2.309401 1.30 0.218 -2.031753 8.031753
2 2 3 | 8.5 2.309401 3.68 0.003 3.468247 13.53175
|
_cons | 11 .8164966 13.47 0.000 9.221007 12.77899

Main effects with F-ratios

F statistics are exact for models in which the disturbances are assumed to be normally distributed, as in the regression above.

anovalator a b c, main fratio

anovalator main-effect for a
chi2(1) = 112.5 p-value = 2.777e-26
scaled as F-ratio = 112.5

anovalator main-effect for b
chi2(1) = .5 p-value = .47950012
scaled as F-ratio = .5

anovalator main-effect for c

chi2(2) = 95.6875 p-value = 1.666e-21
scaled as F-ratio = 47.84375

Two-way interactions with F-ratios

You can check the results for this two-way interaction and the main effects above by running the anova command.

anovalator a b, two fratio

anovalator two-way interaction for a#b
chi2(1) = 120.125 p-value = 5.940e-28
scaled as F-ratio = 120.125

anovalator a c, two f

anovalator two-way interaction for a#c
chi2(2) = 13.6875 p-value = .0010661
scaled as F-ratio = 6.84375

anovalator b c, two f

anovalator two-way interaction for b#c
chi2(2) = 16.9375 p-value = .00020993
scaled as F-ratio = 8.46875

3-way interaction with F-ratio

anovalator a b c, 3way fratio

anovalator 3-way interaction for a#b#c
chi2(2) = 13.9375 p-value = .00094083
scaled as F-ratio = 6.96875

Pairwise comparisons

This time we will use the quiet option to omit the table of adjusted group means from the margins command.

anovalator c, pair quiet

anovalator pairwise comparisons for c

Comparison Coef. Std. Err. z P>|z|
1 vs 2 -3.25 .57735 -5.63 0.000 -4.381607 -2.118393
1 vs 3 -5.625 .57735 -9.74 0.000 -6.756607 -4.493393
2 vs 3 -2.375 .57735 -4.11 0.000 -3.506607 -1.243393

Linear Mixed Model

use https://stats.idre.ucla.edu/stat/data/longitudinal,

clear

xtmixed dv x grp##time || sid:, var

Performing EM optimization:
Performing gradient-based optimization:

Iteration 0: log restricted-likelihood = -592.84806
Iteration 1: log restricted-likelihood = -592.84806

Computing standard errors:

Mixed-effects REML regression Number of obs = 205
Group variable: sid Number of groups = 61

Obs per group: min = 1
avg = 3.4
max = 4

Wald chi2(8) = 59.45
Log restricted-likelihood = -592.84806 Prob > chi2 = 0.0000

--

dv | Coef. Std. Err. z P>|z|

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

x | .4739943 .1531749 3.09 0.002 .173777 .7742116

1.grp | -3.336813 1.374692 -2.43 0.015 -6.03116 -.6424663

|

time |

1 | -.5743502 1.079417 -0.53 0.595 -2.689969 1.541268

2 | -2.030516 1.185852 -1.71 0.087 -4.354744 .2937116

3 | -3.884634 1.185852 -3.28 0.001 -6.208861 -1.560406

|

grp#time |

1 1 | -.8623717 1.418394 -0.61 0.543 -3.642372 1.917629

1 2 | -1.758286 1.514289 -1.16 0.246 -4.726238 1.209666

1 3 | -.1996833 1.521116 -0.13 0.896 -3.181015 2.781649

|

_cons | 6.632933 3.343588 1.98 0.047 .0796211 13.18624

--

--

Random-effects Parameters | Estimate Std. Err.

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

sid: Identity |

var(_cons) | 14.9587 3.625251 9.302605 24.05377

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

var(Residual) | 13.40257 1.603409 10.6012 16.9442

--

LR test vs. linear regression: chibar2(01) = 58.81 Prob >= chibar2 = 0.0000

Main effects

F-ratios may not be appropriate for all linear mixed models so we will forego the fratio option for this example.

anovalator grp time, main

anovalator main-effect for grp
chi2(1) = 12.183244 p-value = .00048221

anovalator main-effect for time
chi2(3) = 32.882824 p-value = 3.409e-07

Two-way interaction

anovalator grp time, two

anovalator two-way interaction for grp#time
chi2(3) = 1.602114 p-value = .65891057

Pairwise comparisons

Here are the pairwise comparisons with the quiet option. Once again we will note that no adjustments have been made for multiple testing.

anovalator time, pair quiet

anovalator pairwise comparisons for time

Comparison Coef. Std. Err. z P>|z|
0 vs 1 1.00554 .709184 1.42 0.156 -.3844646 2.395537
0 vs 2 2.90966 .757128 3.84 0.000 1.425689 4.393629
0 vs 3 3.98448 .760546 5.24 0.000 2.493804 5.475146
1 vs 2 1.90412 .768825 2.48 0.013 .3972266 3.411019
1 vs 3 2.97894 .772193 3.86 0.000 1.465442 4.492437
2 vs 3 1.07482 .794014 1.35 0.176 -.4814507 2.631083

Linear contrast

Example of a linear contrast without the quiet option.

anovalator time, wgt(1 -1/3 -1/3 -1/3)

Predictive margins Number of obs = 205

Expression : Linear prediction, fixed portion, predict()

at : grp (asbalanced)

time (asbalanced)

| Delta-method

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

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

time |

0 | 14.93653 .6864093 21.76 0.000 13.5912 16.28187

1 | 13.931 .7271827 19.16 0.000 12.50575 15.35625

2 | 12.02688 .7740375 15.54 0.000 10.50979 13.54396

3 | 10.95206 .7773841 14.09 0.000 9.428414 12.4757

----

anovalator contrast

(1) 0bn.time - .3333333*1.time - .3333333*2.time - .3333333*3.time = 0

| Coef. Std. Err. z P>|z|

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

(1) | 2.633224 .5912336 4.45 0.000 1.474427 3.79202

Logit Model

Users need to exercise great care in using anovalator with nonlinear models to ensure that they are testing what they really want to test. anovalator gives users the choice of testing effects in the probability metric (the default) or in terms of log-odds using the linear predictor xb. When working in the probability metric the effect of all covariates need to be taken into account when estimating effects. After the logit model below we run a series of anovalator commands first using predicted probabilites and then using the linear prediction.

```
use https://stats.idre.ucla.edu/stat/data/hsbdemo, clear

logit honors prog##female read

Iteration 0: log likelihood = -115.64441
Iteration 1: log likelihood = -86.328713
Iteration 2: log likelihood = -83.40009
Iteration 3: log likelihood = -83.325781
Iteration 4: log likelihood = -83.325724
```

Iteration 5: log likelihood = -83.325724

Logistic regression Number of obs = 200

LR chi2(6) = 64.64

Prob > chi2 = 0.0000

Log likelihood = -83.325724 Pseudo R2 = 0.2795

----

honors | Coef. Std. Err. z P>|z|

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

prog |

2 | 1.432693 .8684243 1.65 0.099 -.2693873 3.134773

3 | .1211212 1.332814 0.09 0.928 -2.491146 2.733388

|

1.female | 2.154428 .9916074 2.17 0.030 .2109131 4.097943

|

prog#female |

2 1 | -1.47558 1.089442 -1.35 0.176 -3.610847 .6596859

3 1 | -.3118623 1.545874 -0.20 0.840 -3.341721 2.717996

|

read | .1409589 .0251746 5.60 0.000 .0916176 .1903003

_cons | -10.38834 1.733946 -5.99 0.000 -13.78681 -6.989865

Main Effects -- Using Predicted Probabilities

Notice that the chi-square and p-value change depending on the value that read is fixed at in the at option. Note: In anovalator each covariate can only be set at one value.

anovalator prog female, main at((mean) read)

anovalator main-effect for prog at((mean) read)
chi2(2) = 2.6100879 p-value = .27116062

anovalator main-effect for female at((mean) read)
chi2(1) = 9.6879421 p-value = .00185481

anovalator prog female, main at(read=70)

anovalator main-effect for prog at(read=70)
chi2(2) = 1.960464 p-value = .37522404

anovalator main-effect for female at(read=70)
chi2(1) = 7.1977024 p-value = .0072997

Two-way Interaction -- Using Predicted Probabilities

anovalator prog female, two at((mean) read)

anovalator two-way interaction for prog#female at((mean) read)
chi2(2) = .93142487 p-value = .62768776

anovalator prog female, two at(read=70)

anovalator two-way interaction for prog#female at(read=70)
chi2(2) = 3.2364218 p-value = .19825308

Pairwise Comparisons -- Using Predicted Probabilities

By now you know there is no adjustment for multiplicity.

anovalator prog, pair at((mean) read)

Adjusted predictions Number of obs = 200
Model VCE : OIM

Expression : Pr(honors), predict()
at : prog (asbalanced)

female (asbalanced)

read = 52.23 (mean)

--

| Delta-method

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

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

prog |

1 | .1246801 .053032 2.35 0.019 .0207394 .2286209

2 | .2220197 .0493918 4.50 0.000 .1252136 .3188258

3 | .1209305 .0642166 1.88 0.060 -.0049318 .2467928

--

anovalator pairwise comparisons for prog at((mean) read)

Comparison Coef. Std. Err. z P>|z|

1 vs 2 -.0973396 .0685826 -1.42 0.156 -.2317615 .03708235

1 vs 3 .00374962 .0829856 .0452 0.964 -.1589022 .1664014

2 vs 3 .101089 .0802136 1.26 0.208 -.05612955 .2583079

Linear Contrast -- Using Predicted Probabilities

Test the average of 1 & 3 versus 2.

anovalator prog, wgt(1/2 -1 1/2) quiet at((mean) read)

anovalator contrast at((mean) read)

(1) .5*1bn.prog - 2.prog + .5*3.prog = 0

--

| Coef. Std. Err. z P>|z|

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

(1) | -.0992144 .0620262 -1.60 0.110 -.2207835 .0223547

--

anovalator prog, wgt(1/2 -1 1/2) quiet at(read=70)

anovalator contrast at(read=70)

(1) .5*1bn.prog - 2.prog + .5*3.prog = 0

--

| Coef. Std. Err. z P>|z|

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

(1) | -.1459797 .1065456 -1.37 0.171 -.3548052 .0628458

--

Now we will run some of the same anovalator command using the predict(xb) option to get linear predictor in the log-odds metric.

Main Effects -- Using the Linear Predictor

Because we are using the linear predictor xb it will not matter where we hold the value of the covariate constant so the results for at((mean) read) will be the same as at(read=70).

anovalator prog female, main at((mean) read) predict(xb)

anovalator main-effect for prog at((mean) read) predict(xb)
chi2(2) = 2.440604 p-value = .29514102

anovalator main-effect for female at((mean) read) predict(xb)
chi2(1) = 7.8079316 p-value = .00520174

anovalator prog female, main at(read=70) pr(xb)

anovalator main-effect for prog at(read=70) predict(xb)
chi2(2) = 2.440604 p-value = .29514102

anovalator main-effect for female at(read=70) predict(xb)
chi2(1) = 7.8079316 p-value = .00520174

Two-way Interaction -- Using the Linear Predictor

From this point on we won't bother holding the covariate constant at any particular value.

anovalator prog female, two pr(xb)

anovalator two-way interaction for prog#female predict(xb)
chi2(2) = 2.2838769 p-value = .31919968

Pairwise Comparisons -- Using the Linear Predictor

By now you know there is no adjustment for multiplicity.

anovalator prog, pair pr(xb)

Predictive margins Number of obs = 200
Model VCE : OIM

Expression : Linear prediction, predict(xb)

at : prog (asbalanced)
female (asbalanced)

| Delta-method
| Margin Std. Err. z P>|z|
-------------+--
prog |
1 | -1.948838 .4859299 -4.01 0.000 -2.901243 -.9964327
2 | -1.253935 .2859528 -4.39 0.000 -1.814392 -.6934777
3 | -1.983648 .6040716 -3.28 0.001 -3.167606 -.7996893
----

anovalator pairwise comparisons for prog predict(xb)

Comparison Coef. Std. Err. z P>|z|
1 vs 2 -.694903 .537408 -1.29 0.196 -1.748222 .3584163
1 vs 3 .03481 .772502 .0451 0.964 -1.479294 1.548914
2 vs 3 .729713 .663054 1.1 0.271 -.5698737 2.029299

Linear Contrast -- Using the Linear Predictor

Test the average of 1 & 3 versus 2.

anovalator prog, wgt(1/2 -1 1/2) quiet pr(xb)

anovalator contrast for prog predict(xb)

(1) .5*1bn.prog - 2.prog + .5*3.prog = 0

--

| Coef. Std. Err. z P>|z|

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

(1) | -.7123078 .4637178 -1.54 0.125 -1.621178 .1965624

--

This concludes the demonstration of the anovalator command. Please use anovalator responsibly.

Date revised: 02/05/10