How can I get ANOVA main-effects with dummy coding in Stata version 10 and earlier?

Authored by

stats writer

July 1, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I get ANOVA main-effects with dummy coding in Stata version 10 and earlier?*. PSYCHOLOGICAL SCALES. Retrieved from https://scales.arabpsychology.com/?p=164169

ANOVA (Analysis of Variance) main-effects with dummy coding can be obtained in Stata version 10 and earlier by following a few simple steps. First, create dummy variables for each level of the categorical variable using the "generate" command. Next, use the "tabulate" command to check the frequency of each dummy variable. Then, run the ANOVA model using the "anova" command and specify the dummy variables as independent variables. Finally, use the "test" command to obtain the main-effects for each dummy variable. This method allows for the easy interpretation of the ANOVA results and is applicable in Stata version 10 and earlier versions.

How can get anova main-effects with dummy coding? (Stata version 10 and earlier) | Stata FAQ

Many researchers like to do their anova using regression with dummy coding but find it confusing when they don't get the same main-effects as in anova. This FAQ will show you how to get those main-effects.

Let's begin by showing the normal anova using a dataset called crf24 to use as a comparison.

use https://stats.idre.ucla.edu/stat/stata/faq/crf24, clear

anova y a b a*b

Number of obs = 32 R-squared = 0.9214

Root MSE = .877971 Adj R-squared = 0.8985

```
Source | Partial SS df MS F Prob > F
-----------+--------------------------------------------------
Model | 217 7 31 40.22 0.0000
|
a | 3.125 1 3.125 4.05 0.0554
b | 194.5 3 64.8333333 84.11 0.0000
a*b | 19.375 3 6.45833333 8.38 0.0006
|
Residual | 18.5 24 .770833333
-----------+--------------------------------------------------
Total | 235.5 31 7.59677419
```

Next, we will manually compute the various dummy variables and run the regression model.

```
tab a, gen(a)
tab b, gen(b)
generate ab1 = a1*b1
generate ab2 = a1*b2
generate ab3 = a1*b3

regress y a1 b1 b2 b3 ab1 ab2 ab3

Source | SS df MS Number of obs = 32
-------------+------------------------------ F( 7, 24) = 40.22
```

Model | 217 7 31 Prob > F = 0.0000

Residual | 18.5 24 .770833333 R-squared = 0.9214

-------------+------------------------------ Adj R-squared = 0.8985

Total | 235.5 31 7.59677419 Root MSE = .87797

--

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

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

a1 | -2 .6208194 -3.22 0.004 -3.281308 -.7186918

b1 | -8.25 .6208194 -13.29 0.000 -9.531308 -6.968692

b2 | -7 .6208194 -11.28 0.000 -8.281308 -5.718692

b3 | -4.5 .6208194 -7.25 0.000 -5.781308 -3.218692

ab1 | 4 .8779711 4.56 0.000 2.187957 5.812043

ab2 | 3 .8779711 3.42 0.002 1.187957 4.812043

ab3 | 3.5 .8779711 3.99 0.001 1.687957 5.312043

_cons | 10 .4389856 22.78 0.000 9.093978 10.90602

--

For this model a2 is the reference level for a and b4 is the

reference level for b, i.e., they are the omitted levels.

Here is the test of the a*b interaction.

test ab1 ab2 ab3

(1) ab1 = 0

(2) ab2 = 0

(3) ab3 = 0

F(3, 24) = 8.38

Prob > F = 0.0006

To get the main-effect for a we will use the dummy for a plus the

a*b interaction dummies averaged across the four levels of b.

test a1 + (ab1+ab2+ab3)/4 = 0

(1) a1 + .25 ab1 + .25 ab2 + .25 ab3 = 0

F(1, 24) = 4.05

Prob > F = 0.0554

The main-effect for b is a little bit trickier because it is a 3 degree of

freedom test so we will have to do the test command three times and make use of the

accumulate option.

test b1 + ab1/2 = 0

(1) b1 + .5 ab1 = 0

F(1, 24) = 202.70
Prob > F = 0.0000

test b2 + ab2/2 = 0, accumulate

(1) b1 + .5 ab1 = 0
(2) b2 + .5 ab2 = 0

F(2, 24) = 120.86
Prob > F = 0.0000

test b3 + ab3/2 = 0, accumulate

(1) b1 + .5 ab1 = 0
(2) b2 + .5 ab2 = 0
(3) b3 + .5 ab3 = 0

F(3, 24) = 84.11
Prob > F = 0.0000

The last test command has our main-effect for b

So, what's with all of the division, by 4 in the a main-effect and by 2 in the b main-effect. The dummy variable a1 is actually the simple effect of a. To get the "true" main-effect of a we have to combine the simple effect of a with the average of the interaction effects across the four levels of b. Likewise, for the b main-effect we need to combine the simple main-effects of the levels of b with the average interaction effect across the two levels of a.

Example 2

This method generalizes to more complex designs with multiple factors so let's consider a 3-factor completely crossed design.

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

anova y a b c a*b a*c b*c a*b*c

Number of obs = 24 R-squared = 0.9689

Root MSE = 1.1547 Adj R-squared = 0.9403

Source | Partial SS df MS F Prob > F

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

Model | 497.833333 11 45.2575758 33.94 0.0000

|

a | 150 1 150 112.50 0.0000

b | .666666667 1 .666666667 0.50 0.4930

c | 127.583333 2 63.7916667 47.84 0.0000

a*b | 160.166667 1 160.166667 120.13 0.0000

a*c | 18.25 2 9.125 6.84 0.0104

b*c | 22.5833333 2 11.2916667 8.47 0.0051

a*b*c | 18.5833333 2 9.29166667 6.97 0.0098

|

Residual | 16 12 1.33333333

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

Total | 513.833333 23 22.3405797

Once again we will manually create the dummy variables and run the regression model.

recode a (1=0)(2=1)

recode b (1=0)(2=1)

tab c, gen(c)

gen ab=a*b

gen ac1=a*c1

gen ac2=a*c2

gen bc1=b*c1

gen bc2=b*c2

gen abc1=a*b*c1

gen abc2=a*b*c2

regress y a b c1 c2 ab ac1 ac2 bc1 bc2 abc1 abc2

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|

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

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

b | -9.5 1.154701 -8.23 0.000 -12.01588 -6.984124

c1 | -8 1.154701 -6.93 0.000 -10.51588 -5.484124

c2 | -4 1.154701 -3.46 0.005 -6.515876 -1.484124

ab | 15 1.632993 9.19 0.000 11.44201 18.55799
ac1 | 6.39e-14 1.632993 0.00 1.000 -3.557986 3.557986
ac2 | 1 1.632993 0.61 0.552 -2.557986 4.557986
bc1 | 9 1.632993 5.51 0.000 5.442014 12.55799
bc2 | 5 1.632993 3.06 0.010 1.442014 8.557986
abc1 | -8.5 2.309401 -3.68 0.003 -13.53175 -3.468247
abc2 | -5.5 2.309401 -2.38 0.035 -10.53175 -.4682473
_cons | 19 .8164966 23.27 0.000 17.22101 20.77899
--

Here is the test of the three-way a*b*c interaction.

test abc1 abc2

(1) abc1 = 0
(2) abc2 = 0

F(2, 12) = 6.97
Prob > F = 0.0098

Next come the two-way interactions with both a*c and b*c using the
accumulate options.

/* a*b interaction */

```
test ab + (abc1+abc2)/3 = 0

( 1) ab + .3333333 abc1 + .3333333 abc2 = 0

F( 1, 12) = 120.13
Prob > F = 0.0000

/* a*c interaction) */

test ac1 + abc1/2 = 0

( 1) ac1 + .5 abc1 = 0

F( 1, 12) = 13.55
Prob > F = 0.0031

test ac2 + abc2/2 = 0, accumulate

( 1) ac1 + .5 abc1 = 0
( 2) ac2 + .5 abc2 = 0

F( 2, 12) = 6.84
Prob > F = 0.0104

/* b*c interaction */

test bc1 + abc1/2 = 0
```

(1) bc1 + .5 abc1 = 0

F(1, 12) = 16.92

Prob > F = 0.0014

test bc2 + abc2/2 = 0, accumulate

(1) bc1 + .5 abc1 = 0

(2) bc2 + .5 abc2 = 0

F(2, 12) = 8.47

Prob > F = 0.0051

Finally, we get to the main-effects.

/* a main-effect */

test a + ab/2 + (ac1+ac2)/3 + (abc1+abc2)/6 = 0

(1) a + .5 ab + .3333333 ac1 + .3333333 ac2 + .1666667 abc1 + .1666667 abc2 = 0

F(1, 12) = 112.50

Prob > F = 0.0000

/* b main-effect */

test b + ab/2 + (bc1+bc2)/3 + (abc1+abc2)/6 = 0

(1) b + .5 ab + .3333333 bc1 + .3333333 bc2 + .1666667 abc1 + .1666667 abc2 = 0

F(1, 12) = 0.50
Prob > F = 0.4930

/* c main-effect */

test c1 + ac1/2 + bc1/2 + abc1/4 = 0

(1) c1 + .5 ac1 + .5 bc1 + .25 abc1 = 0

F(1, 12) = 94.92
Prob > F = 0.0000

test c2 + ac2/2 + bc2/2 + abc2/4 = 0, accumulate

(1) c1 + .5 ac1 + .5 bc1 + .25 abc1 = 0
(2) c2 + .5 ac2 + .5 bc2 + .25 abc2 = 0

F(2, 12) = 47.84
Prob > F = 0.0000