

How can I get ANOVA simple main effects with dummy coding in Stata?

Authored by
stats writer

July 1, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I get ANOVA simple main effects with dummy coding in Stata?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=164200>

ANOVA (Analysis of Variance) is a statistical tool used to compare the means of three or more groups. In Stata, dummy coding is a common method of creating categorical variables for use in ANOVA. It involves creating binary variables that represent the different groups being compared. To obtain ANOVA simple main effects with dummy coding in Stata, you can use the "contrast" command. This command allows you to specify the groups you want to compare and generates the relevant ANOVA results, including the simple main effects. Additionally, the "contrast" command also allows for post hoc tests to be conducted to determine which specific group means differ significantly from each other. This method is useful for examining the effects of categorical variables on a continuous outcome variable in Stata.

How can I get anova simple main effects with dummy coding? | Stata FAQ

Many researchers like to do their anova using regression with dummy coding but wonder how to get tests of simple main effects. This FAQ will show you how to get tests of simple main effects when using dummy coding.

We will begin with a two-factor design using a dataset called crf24.

We will manually compute the various dummy variables and run the regression model.

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

```
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

Since the a*b interaction is statistically significant we will proceed with tests of simple main effects. First, we will do the tests of b at a2 which is the reference level for a.

test b1 b2 b3

(1) b1 = 0

(2) b2 = 0

(3) b3 = 0

F(3, 24) = 68.84

Prob > F = 0.0000

We will follow this with tests of b at a1 which will require that we combine the b dummies with their interaction terms and use the accumulate option to get a test with the correct degrees of freedom.

test b1 + ab1 = 0

(1) b1 + ab1 = 0

F(1, 24) = 46.86

Prob > F = 0.0000

test b2 + ab2 = 0, accumulate

(1) b1 + ab1 = 0

(2) b2 + ab2 = 0

F(2, 24) = 29.51

Prob > F = 0.0000

test b3 + ab3 = 0, accumulate

(1) b1 + ab1 = 0

(2) b2 + ab2 = 0

(3) b3 + ab3 = 0

F(3, 24) = 23.65

Prob > F = 0.0000

The last test command has the F-ratio = 23.65 for b at a1.

We could have done the tests of simple main effects for differences in a for each level of b. Thoses tests would look like this:

/* test of a at b=4 (the reference group) */

test a1

(1) a1 = 0

F(1, 24) = 10.38

Prob > F = 0.0036

/* test of a at b=1 */

```
test a1 + ab1 = 0
```

```
( 1) a1 + ab1 = 0
```

```
F( 1, 24) = 10.38
```

```
Prob > F = 0.0036
```

```
/* test of a at b=2 */
```

```
test a1 + ab2 = 0
```

```
( 1) a1 + ab2 = 0
```

```
F( 1, 24) = 2.59
```

```
Prob > F = 0.1203
```

```
/* test of a at b=3 */
```

```
test a1 + ab3 = 0
```

```
( 1) a1 + ab3 = 0
```

```
F( 1, 24) = 5.84
```

```
Prob > F = 0.0237
```

As you can see, testing an effect, say a, by itself tests the simple

main effect for the reference group. To get the simple main effects for the other levels, you combine the test of a with the appropriate interaction term.

This FAQ only covers the computation of the tests of simple main effects using dummy coding. The FAQ does not cover computing the critical values of these tests. There is a user written ado-program `smecriticalvalue` which can assist in this process (search `smecriticalvalue`).

This method of using dummy coding to compute tests of simple main effects generalizes to more complex designs with multiple factors. Let's try it with a three-factor completely crossed design. Once again we will manually create the dummy variables and run the regression model.

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

```
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

```

In this model a1, b1 and c3 are the reference groups for variables a, b and c, respectively.

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

```
test abc1 abc2 abc2
```

(1) abc1 = 0

(2) abc2 = 0

(3) abc2 = 0

Constraint 3 dropped

F(2, 12) = 6.97

Prob > F = 0.0098

Since the test of the three-way interaction is statistically significant it means that there is one or more two-way interactions that are significant at different levels of a third variable. We have reason to believe that looking at the b*c interaction could be productive, therefore we will test the b*c interactions for each level of a.

/* test b*c at a1 (the reference level) */

test bc1 bc2

(1) bc1 = 0

(2) bc2 = 0

F(2, 12) = 15.25

Prob > F = 0.0005

/* test b*c at a2 */

test bc1+abc1=0

test bc2+abc2=0, accumulate

(1) bc1 + abc1 = 0

(2) bc2 + abc2 = 0

F(2, 12) = 0.1875

Prob > F = 0.8314

We see that b*c at a1 has a pretty large F-ratio and that the

F-ratio for b*c at a2 is less than one.

So we will follow up this analysis by looking for differences in c for each level of b all within a1.

/* test for c at b==1 & a==1 */

test c1 c2

(1) c1 = 0

(2) c2 = 0

F(2, 12) = 24.00

Prob > F = 0.0001

```
/* test for c at b==2 & a==1 */
```

```
test c1+bc1=0
```

```
test c2+bc2=0, accum
```

```
( 1) c1 + bc1 = 0
```

```
( 2) c2 + bc2 = 0
```

```
F( 2, 12) = 0.50
```

```
Prob > F = 0.6186
```

Dummy coding provides a fast and easy way to compute tests of simple main effects.