anova postestimation — Postestimation tools for anova

Postestimation commands

The following postestimation commands are of special interest after `anova`:

Command	Description
`dfbeta`	DFBETA influence statistics
`estat hettest`	tests for heteroskedasticity
`estat imtest`	information matrix test
`estat ovtest`	Ramsey regression specification-error test for omitted variables
`estat szroeter`	Szroeter's rank test for heteroskedasticity
`estat vif`	variance inflation factors for the independent variables
`estat esize`	η^2 and ω^2 effect sizes
`rvfplot`	residual-versus-fitted plot
`avplot`	added-variable plot
`avplots`	all added-variables plots in one image
`cprplot`	component-plus-residual plot
`acprplot`	augmented component-plus-residual plot
`rvpplot`	residual-versus-predictor plot
`lvr2plot`	leverage-versus-squared-residual plot

The following standard postestimation commands are also available:

Command	Description
contrast	contrasts and ANOVA-style joint tests of parameters
estat ic	Akaike's, consistent Akaike's, corrected Akaike's, and Schwarz's Bayesian information criteria (AIC, CAIC, AICc, and BIC, respectively)
estat summarize	summary statistics for the estimation sample
estat vce	variance–covariance matrix of the estimators (VCE)
estimates	cataloging estimation results
etable	table of estimation results
forecast	dynamic forecasts and simulations
hausman	Hausman's specification test
lincom	point estimates, standard errors, testing, and inference for linear combinations of parameters
linktest	link test for model specification
lrtest	likelihood-ratio test
margins	marginal means, predictive margins, marginal effects, and average marginal effects
marginsplot	graph the results from margins (profile plots, interaction plots, etc.)
nlcom	point estimates, standard errors, testing, and inference for nonlinear combinations of parameters
predict	predictions and their SEs, leverage statistics, distance statistics, etc.
predictnl	point estimates, standard errors, testing, and inference for generalized predictions
pwcompare	pairwise comparisons of parameters
suest	seemingly unrelated estimation
test	Wald tests of simple and composite linear hypotheses
testnl	Wald tests of nonlinear hypotheses

predict

`predict` after `anova` follows the same syntax as `predict` after `regress` and can provide predictions, residuals, standardized residuals, Studentized residuals, the standard error of the residuals, the standard error of the prediction, the diagonal elements of the projection (hat) matrix, and Cook's D. See [R] **regress postestimation** for details.

margins

`margins` after `anova` follows the same syntax as `margins` after `regress`. See [R] **regress postestimation** for details.

test

Description for test

In addition to the standard syntax of `test` (see [R] **test**), `test` after `anova` has three additionally allowed syntaxes; see below. `test` performs Wald tests of expressions involving the coefficients of the underlying regression model. Simple and composite linear hypotheses are possible.

Menu for test

Statistics > Linear models and related > ANOVA/MANOVA > Test linear hypotheses after anova

Syntax for test

`test, test(`*matname*`) [mtest[(`*opt*`)] matvlc(`*matname*`)]` syntax a

`test, showorder` syntax b

`test [`*term* `[`*term* `...] ] [/` *term* `[`*term* `...] ] [, symbolic]` syntax c

syntax a	test expression involving the coefficients of the underlying regression model; you provide information as a matrix
syntax b	show underlying order of design matrix, which is useful when constructing *matname* argument of the `test()` option
syntax c	test effects and show symbolic forms

Options for test

`test(`*matname*`)` is required with syntax a of `test`. The rows of *matname* specify linear combinations of the underlying design matrix of the ANOVA that are to be jointly tested. The columns correspond to the underlying design matrix (including the constant if it has not been suppressed). The column and row names of *matname* are ignored.

A listing of the constraints imposed by the `test()` option is presented before the table containing the tests. You should examine this table to verify that you have applied the linear combinations you desired. Typing `test, showorder` allows you to examine the ordering of the columns for the design matrix from the ANOVA.

`mtest[(`*opt*`)]` specifies that tests are performed for each condition separately. *opt* specifies the method for adjusting p-values for multiple testing. Valid values for *opt* are

`bonferroni`	Bonferroni's method
`holm`	Holm's method
`sidak`	Šidák's method
`noadjust`	no adjustment is to be made

Specifying `mtest` with no argument is equivalent to `mtest(noadjust)`.

`matvlc(`*matname*`)`, a programmer's option, saves the variance–covariance matrix of the linear combinations involved in the suite of tests. For the test $\mathbf{Lb} = \mathbf{c}$, what is returned in *matname* is $\mathbf{L}V\mathbf{L}'$, where V is the estimated variance–covariance matrix of $\mathbf{b}$.

`showorder` causes `test` to list the definition of each column in the design matrix. `showorder` is not allowed with any other option.

`symbolic` requests the symbolic form of the test rather than the test statistic. When this option is specified with no terms (`test, symbolic`), the symbolic form of the estimable functions is displayed.

Remarks and examples

Remarks are presented under the following headings:

See examples 4, 7, 8, 13, 15, 16, and 17 in [R] **anova** for examples that use the `margins` command.

Testing effects

After fitting a model using `anova`, you can test for the significance of effects in the ANOVA table, as well as for effects that are not reported in the ANOVA table, by using the `test` or `contrast` command. You follow `test` or `contrast` by the list of effects that you wish to test. By default, these commands use the residual mean squared error in the denominator of the F ratio. You can specify other error terms by using the slash notation, just as you would with `anova`. See [R] **contrast** for details on this command.

▷ Example 1: Testing effects

Recall our byssinosis example (example 8) in [R] **anova**:

```
. anova prob workplace smokes race workplace#smokes workplace#race smokes#race
> workplace#smokes#race [aweight=pop]
(sum of wgt is 5,419)
                          Number of obs =         65    R-squared     =  0.8300
                          Root MSE      =    .025902    Adj R-squared =  0.7948
                  Source | Partial SS         df         MS        F    Prob>F
-------------------------+----------------------------------------------------
                   Model |  .17364654         11   .01578605     23.53  0.0000
                         |
                workplace |  .09762518          2   .04881259     72.76  0.0000
                   smokes |  .01303081          1   .01303081     19.42  0.0001
                     race |  .00109472          1   .00109472      1.63  0.2070
          workplace#smokes |  .01969034          2   .00984517     14.67  0.0000
            workplace#race |  .00135252          2   .00067626      1.01  0.3718
               smokes#race |  .00166287          1   .00166287      2.48  0.1214
    workplace#smokes#race |  .00095084          2   .00047542      0.71  0.4969
                         |
                 Residual |  .03555777         53    .0006709
-------------------------+----------------------------------------------------
                    Total |   .2092043         64   .00326882
```

We can easily obtain a test on a particular term from the ANOVA table. Here are two examples:

```
. test smokes
```

Source	Partial SS	df	MS	F	Prob>F
smokes	.01303081	1	.01303081	19.42	0.0001
Residual	.03555777	53	.0006709		

```
. test smokes#race
```

Source	Partial SS	df	MS	F	Prob>F
smokes#race	.00166287	1	.00166287	2.48	0.1214
Residual	.03555777	53	.0006709		

Both of these tests use residual error by default and agree with the ANOVA table produced earlier.

We could have performed these same tests with `contrast`:

```
. contrast smokes
Contrasts of marginal linear predictions
Margins: asbalanced
```

	df	F	P>F
smokes	1	19.42	0.0001
Denominator	53		

```
. contrast smokes#race
Contrasts of marginal linear predictions
Margins: asbalanced
```

	df	F	P>F
smokes#race	1	2.48	0.1214
Denominator	53		

◁

❑ Technical note

After `anova`, you can use the '`/`' syntax in `test` or `contrast` to perform tests with a variety of non-$\sigma^2\mathbf{I}$ error structures. However, in most unbalanced models, the mean squares are not independent and do not have equal expectations under the null hypothesis. Also, be warned that you assume responsibility for the validity of the test statistic.

❑

Example 2: Testing effects with different error terms

We return to the nested ANOVA example (example 11) in [R] **anova**, where five brands of machinery were compared in an assembly line. We can obtain appropriate tests for the nested terms using `test`, even if we had run the `anova` command without initially indicating the proper error terms.

```
. use https://www.stata-press.com/data/r19/machine
(Machine data)
. anova output machine / operator|machine /
                       Number of obs =         57    R-squared     =  0.8661
                       Root MSE      =    1.47089    Adj R-squared =  0.8077
```

Source	Partial SS	df	MS	F	Prob>F
Model	545.82229	17	32.107193	14.84	0.0000
machine	430.98079	4	107.7452	13.82	0.0001
operator|machine	101.3538	13	7.7964465		
operator|machine	101.3538	13	7.7964465	3.60	0.0009
Residual	84.376658	39	2.1635041		
Total	630.19895	56	11.253553		

In this ANOVA table, `machine` is tested with residual error. With this particular nested design, the appropriate error term for testing `machine` is `operator` nested within `machine`, which is easily obtained from `test`.

```
. test machine / operator|machine
```

Source	Partial SS	df	MS	F	Prob>F
machine	430.98079	4	107.7452	13.82	0.0001
operator|machine	101.3538	13	7.7964465		

This result from `test` matches what we obtained from our `anova` command.

◁

▷ Example 3: Pooling terms when testing effects

The other nested ANOVA example (example 12) in [R] **anova** was based on the sewage data. The ANOVA table is presented here again. As before, we will use abbreviations of variable names in typing the commands.

```
. use https://www.stata-press.com/data/r19/sewage
(Sewage treatment)
. anova particulate s / m|s / f|m|s / w|f|m|s /, dropemptycells
```

```
                    Number of obs =         64    R-squared     =  0.6338
                    Root MSE      =    12.7445    Adj R-squared =  0.5194
```

Source	Partial SS	df	MS	F	Prob>F
Model	13493.609	15	899.57396	5.54	0.0000
solution	7203.7656	1	7203.7656	17.19	0.0536
manager|solution	838.28125	2	419.14063		
manager|solution	838.28125	2	419.14063	0.55	0.6166
facility|manager|solution	3064.9375	4	766.23438		
facility|manager|solution	3064.9375	4	766.23438	2.57	0.1193
worker|facility|manager|solution	2386.625	8	298.32813		
worker|facility|manager|solution	2386.625	8	298.32813	1.84	0.0931
Residual	7796.25	48	162.42188		
Total	21289.859	63	337.93428		

In practice, it is often beneficial to pool nonsignificant nested terms to increase the power of tests on remaining terms. One rule of thumb is to allow the pooling of a term whose p-value is larger than 0.25. In this sewage example, the p-value for the test of `manager` is 0.6166. This value indicates that the manager effect is negligible and might be ignored. Currently, `solution` is tested by `manager|solution`, which has only 2 degrees of freedom. If we pool the `manager` and `facility` terms and use this pooled estimate as the error term for `solution`, we would have a term with 6 degrees of freedom.

Below are two tests: a test of `solution` with the pooled `manager` and `facility` terms and a test of this pooled term by `worker`.

```
. test s / m|s f|m|s
```

Source	Partial SS	df	MS	F	Prob>F
solution	7203.7656	1	7203.7656	11.07	0.0159
manager|solution facility|manager| solution	3903.2188	6	650.53646		

```
. test m|s f|m|s / w|f|m|s
```

Source	Partial SS	df	MS	F	Prob>F
manager|solution facility|manager| solution	3903.2188	6	650.53646	2.18	0.1520
worker|facility|manager| solution	2386.625	8	298.32813		

In the first test, we included two terms after the forward slash (`m|s` and `f|m|s`). `test` after `anova` allows multiple terms both before and after the slash. The terms before the slash are combined and are then tested by the combined terms that follow the slash (or residual error if no slash is present).

The p-value for `solution` using the pooled term is 0.0159. Originally, it was 0.0536. The increase in the power of the test is due to the increase in degrees of freedom for the pooled error term.

We can get identical results if we drop `manager` from the `anova` model. (This dataset has unique numbers for each facility so that there is no confusion of facilities when `manager` is dropped.)

```
. anova particulate s / f|s / w|f|s /, dropemptycells
```

```
                        Number of obs =         64    R-squared     =  0.6338
                        Root MSE      =    12.7445    Adj R-squared =  0.5194
```

Source	Partial SS	df	MS	F	Prob>F
Model	13493.609	15	899.57396	5.54	0.0000
solution	7203.7656	1	7203.7656	11.07	0.0159
facility|solution	3903.2187	6	650.53646		
facility|solution	3903.2187	6	650.53646	2.18	0.1520
worker|facility| solution	2386.625	8	298.32812		
worker|facility| solution	2386.625	8	298.32812	1.84	0.0931
Residual	7796.25	48	162.42188		
Total	21289.859	63	337.93428		

This output agrees with our earlier `test` results.

◁

▷ Example 4: Obtaining overall significance of a term using contrast

In example 10 of [R] **anova**, we fit the model `anova drate region c.mage region#c.mage`. Now, we use the `contrast` command to test for the overall significance of `region`.

```
. contrast region region#c.mage, overall
Contrasts of marginal linear predictions
Margins: asbalanced
```

	df	F	P>F
region	3	7.40	0.0004
region#c.mage	3	0.86	0.4689
Overall	6	5.65	0.0002
Denominator	42		

The overall F statistic associated with the `region` and `region#c.mage` terms is 5.65, and it is significant at the 0.02% level.

In the ANOVA output, the `region` term, by itself, had a sum of squares of 1166.15, which, based on 3 degrees of freedom, yielded an F statistic of 7.40 and a significance level of 0.0004. This is the same test that is reported by `contrast` in the row labeled `region`. Likewise, the test from the ANOVA output for the `region#c.mage` term is reproduced in the second row of the `contrast` output.

◁

Obtaining symbolic forms

`test` can produce the symbolic form of the estimable functions and symbolic forms for particular tests.

▷ Example 5: Symbolic form of the estimable functions

After fitting an ANOVA model, we type `test, symbolic` to obtain the symbolic form of the estimable functions. For instance, returning to our blood pressure data introduced in example 4 of [R] **anova**, let's begin by reestimating `systolic` on `drug`, `disease`, and `drug#disease`:

```
. use https://www.stata-press.com/data/r19/systolic, clear
(Systolic blood pressure data)
. anova systolic drug disease drug#disease
```

```
                   Number of obs =         58    R-squared     =  0.4560
                   Root MSE      =    10.5096    Adj R-squared =  0.3259
```

Source	Partial SS	df	MS	F	Prob>F
Model	4259.3385	11	387.21259	3.51	0.0013
drug	2997.4719	3	999.15729	9.05	0.0001
disease	415.87305	2	207.93652	1.88	0.1637
drug#disease	707.26626	6	117.87771	1.07	0.3958
Residual	5080.8167	46	110.45254		
Total	9340.1552	57	163.86237		

To obtain the symbolic form of the estimable functions, type

```
. test, symbolic
drug
           1  -(r2+r3+r4-r0)
           2   r2
           3   r3
           4   r4
disease
           1  -(r6+r7-r0)
           2   r6
           3   r7
drug#disease
        1  1  -(r2+r3+r4+r6+r7-r12-r13-r15-r16-r18-r19-r0)
        1  2   r6 - (r12+r15+r18)
        1  3   r7 - (r13+r16+r19)
        2  1   r2 - (r12+r13)
        2  2   r12
        2  3   r13
        3  1   r3 - (r15+r16)
        3  2   r15
        3  3   r16
        4  1   r4 - (r18+r19)
        4  2   r18
        4  3   r19
_cons          r0
```

◁

▷ Example 6: Symbolic form for a particular test

To obtain the symbolic form for a particular test, we type `test` *term* [*term* ...], `symbolic`. For instance, the symbolic form for the test of the main effect of `drug` is

```
. test drug, symbolic
drug
           1  -(r2+r3+r4)
           2   r2
           3   r3
           4   r4
disease
           1  0
           2  0
           3  0
drug#disease
        1  1  -1/3 (r2+r3+r4)
        1  2  -1/3 (r2+r3+r4)
        1  3  -1/3 (r2+r3+r4)
        2  1   1/3 r2
        2  2   1/3 r2
        2  3   1/3 r2
        3  1   1/3 r3
        3  2   1/3 r3
        3  3   1/3 r3
        4  1   1/3 r4
        4  2   1/3 r4
        4  3   1/3 r4
_cons          0
```

If we omit the symbolic option, we instead see the result of the test:

```
. test drug
```

Source	Partial SS	df	MS	F	Prob>F
drug	2997.4719	3	999.15729	9.05	0.0001
Residual	5080.8167	46	110.45254		

Testing coefficients and contrasts of margins

test allows you to perform tests directly on the coefficients of the underlying regression model. For instance, the coefficient on the third drug and the second disease is referred to as 3.drug#2.disease. This could also be written as i3.drug#i2.disease, or _b[3.drug#2.disease], or even _coef[i3.drug#i2.disease]; see [U] 13.5 Accessing coefficients and standard errors.

Example 7: Testing linear combinations of coefficients

Let's begin by testing whether the coefficient on the third drug is equal to the coefficient on the fourth drug in our blood pressure data. We have already fit the model anova systolic drug##disease (equivalent to anova systolic drug disease drug#disease), and you can see the results of that estimation in example 5. Even though we have performed many tasks since we fit the model, Stata still remembers, and we can perform tests at any time.

```
. test 3.drug = 4.drug
 ( 1)  3.drug - 4.drug = 0
       F(  1,    46) =    0.13
            Prob > F =    0.7234
```

We find that the two coefficients are not significantly different, at least at any significance level smaller than 73%.

For more complex tests, the contrast command often provides a more concise way to specify the tests we are interested in and prevents us from having to write the tests in terms of the regression coefficients. With contrast, we instead specify our tests in terms of differences in the marginal means for the levels of a particular factor. For example, if we want to compare the third and fourth drugs, we can test the difference in the mean impact on systolic blood pressure separately for each disease using the @ operator. We also use the reverse adjacent operator, ar., to compare the fourth level of drug with the previous level.

```
. contrast ar4.drug@disease
Contrasts of marginal linear predictions
Margins: asbalanced
```

	df	F	P>F
drug@disease			
(4 vs 3) 1	1	0.13	0.7234
(4 vs 3) 2	1	1.76	0.1917
(4 vs 3) 3	1	0.65	0.4230
Joint	3	0.85	0.4761
Denominator	46		

	Contrast	Std. err.	[95% conf.	interval]
drug@disease				
(4 vs 3) 1	-2.733333	7.675156	-18.18262	12.71595
(4 vs 3) 2	8.433333	6.363903	-4.376539	21.24321
(4 vs 3) 3	5.7	7.050081	-8.491077	19.89108

None of the individual contrasts shows significant differences between the third drug and the fourth drug. Likewise, the overall F statistic is 0.85, which is hardly significant. We cannot reject the hypothesis that the third drug has the same effect as the fourth drug.

◁

❑ Technical note

Alternatively, we could have specified these tests based on the coefficients of the underlying regression model using the `test` command. We would have needed to perform tests on the coefficients for `drug` and for the coefficients on `drug` interacted with `disease` to test for differences in the means mentioned above. To do this, we start with our previous `test` command:

```
. test 3.drug = 4.drug
```

Notice that the F statistic for this test is equivalent to the test labeled `(4 vs 3) 1` in the `contrast` output. Let's now add the constraint that the coefficient on the third drug interacted with the third disease is equal to the coefficient on the fourth drug, again interacted with the third disease. We do that by typing the new constraint and adding the `accumulate` option:

```
. test 3.drug#3.disease = 4.drug#3.disease, accumulate
 ( 1)  3.drug - 4.drug = 0
 ( 2)  3.drug#3.disease - 4.drug#3.disease = 0
       F(  2,    46) =    0.39
            Prob > F =    0.6791
```

So far, our test includes the equality of the two drug coefficients, along with the equality of the two drug coefficients when interacted with the third disease. Now, we add two more equations, one for each of the remaining two diseases:

```
. test 3.drug#2.disease = 4.drug#2.disease, accumulate
 ( 1)  3.drug - 4.drug = 0
 ( 2)  3.drug#3.disease - 4.drug#3.disease = 0
 ( 3)  3.drug#2.disease - 4.drug#2.disease = 0
       F(  3,    46) =    0.85
            Prob > F =    0.4761
. test 3.drug#1.disease = 4.drug#1.disease, accumulate
 ( 1)  3.drug - 4.drug = 0
 ( 2)  3.drug#3.disease - 4.drug#3.disease = 0
 ( 3)  3.drug#2.disease - 4.drug#2.disease = 0
 ( 4)  3o.drug#1b.disease - 4o.drug#1b.disease = 0
       Constraint 4 dropped
       F(  3,    46) =    0.85
            Prob > F =    0.4761
```

The overall F statistic reproduces the one from the joint test in the `contrast` output.

You may notice that we also got the message "Constraint 4 dropped". For the technically inclined, this constraint was unnecessary, given the normalization of the model. If we specify all the constraints involved in our test or use `contrast`, we need not worry about the normalization because Stata handles this automatically.

❑

The `test()` option of `test` provides another alternative for testing coefficients. Instead of spelling out each coefficient involved in the test, a matrix representing the test provides the needed information. `test, showorder` shows the order of the terms in the ANOVA corresponding to the order of the columns for the matrix argument of `test()`.

▷ Example 8: Another way to test linear combinations of coefficients

We repeat the last test of example 7 above with the `test()` option. First, we view the definition and order of the columns underlying the ANOVA performed on the systolic data.

```
. test, showorder
 Order of columns in the design matrix
      1: (drug==1)
      2: (drug==2)
      3: (drug==3)
      4: (drug==4)
      5: (disease==1)
      6: (disease==2)
      7: (disease==3)
      8: (drug==1)*(disease==1)
      9: (drug==1)*(disease==2)
     10: (drug==1)*(disease==3)
     11: (drug==2)*(disease==1)
     12: (drug==2)*(disease==2)
     13: (drug==2)*(disease==3)
     14: (drug==3)*(disease==1)
     15: (drug==3)*(disease==2)
     16: (drug==3)*(disease==3)
     17: (drug==4)*(disease==1)
     18: (drug==4)*(disease==2)
     19: (drug==4)*(disease==3)
     20: _cons
```

Columns 1–4 correspond to the four levels of `drug`. Columns 5–7 correspond to the three levels of `disease`. Columns 8–19 correspond to the interaction of `drug` and `disease`. The last column corresponds to `_cons`, the constant in the model.

We construct the matrix `dr3vs4` with the same four constraints as the last test shown in example 7 and then use the `test(dr3vs4)` option to perform the test.

```
. matrix dr3vs4 = (0,0,1,-1,  0,0,0,  0,0,0,0,0,0,0,0,0, 0, 0, 0,  0 \
>                  0,0,0, 0,  0,0,0,  0,0,0,0,0,0,0,1, 0, 0,-1,  0 \
>                  0,0,0, 0,  0,0,0,  0,0,0,0,0,0,0,1,0, 0,-1, 0,  0 \
>                  0,0,0, 0,  0,0,0,  0,0,0,0,0,0,1,0,0,-1, 0, 0,  0)
. test, test(dr3vs4)
 ( 1)  3.drug - 4.drug = 0
 ( 2)  3.drug#3.disease - 4.drug#3.disease = 0
 ( 3)  3.drug#2.disease - 4.drug#2.disease = 0
 ( 4)  3o.drug#1b.disease - 4o.drug#1b.disease = 0
       Constraint 4 dropped
       F(  3,    46) =    0.85
            Prob > F =    0.4761
```

Here the effort involved with spelling out the coefficients is similar to that of constructing a matrix and using it in the `test()` option. When the test involving coefficients is more complicated, the `test()` option may be more convenient than specifying the coefficients directly in `test`. However, as previously demonstrated, `contrast` may provide an even simpler method for testing the same hypothesis.

◁

After fitting an ANOVA model, various contrasts (1-degree-of-freedom tests comparing different levels of a categorical variable) are often of interest. `contrast` can perform each 1-degree-of-freedom test in addition to the combined test, even in cases in which the contrasts do not correspond to one of the contrast operators.

▷ Example 9: Testing particular contrasts of interest

Rencher and Schaalje (2008) illustrate 1-degree-of-freedom contrasts for an ANOVA comparing the net weight of cans filled by five machines (labeled `A`–`E`). The data were originally obtained from Ostle and Mensing (1975). Rencher and Schaalje use a cell-means ANOVA model approach for this problem. We could do the same by using the `noconstant` option of `anova`; see [R] **anova**. Instead, we obtain the same results by using the standard overparameterized ANOVA approach (that is, we keep the constant in the model).

```
. use https://www.stata-press.com/data/r19/canfill
(Can fill data)
. list, sepby(machine)
```

	machine	weight
1.	A	11.95
2.	A	12.00
3.	A	12.25
4.	A	12.10
5.	B	12.18
6.	B	12.11
7.	C	12.16
8.	C	12.15
9.	C	12.08
10.	D	12.25
11.	D	12.30
12.	D	12.10
13.	E	12.10
14.	E	12.04
15.	E	12.02
16.	E	12.02

```
. anova weight machine
                           Number of obs =         16    R-squared     =  0.4123
                           Root MSE      =   .087758    Adj R-squared =  0.1986

                  Source | Partial SS         df          MS         F    Prob>F
-------------------------+----------------------------------------------------
                   Model |  .05942699          4   .01485675      1.93  0.1757
                         |
                 machine |  .05942699          4   .01485675      1.93  0.1757
                         |
                Residual |   .0847167         11   .00770152
-------------------------+----------------------------------------------------
                   Total |  .14414369         15   .00960958
```

The four 1-degree-of-freedom tests of interest among the five machines are `A` and `D` versus `B`, `C`, and `E`; `B` and `E` versus `C`; `A` versus `D`; and `B` versus `E`. We can specify these tests as user-defined contrasts by placing the corresponding contrast coefficients into positions related to the five levels of machine as described in *User-defined contrasts* of [R] **contrast**.

```
. contrast {machine 3 -2 -2  3 -2}
>          {machine 0  1 -2  0  1}
>          {machine 1  0  0 -1  0}
>          {machine 0  1  0  0 -1}, noeffects
Contrasts of marginal linear predictions
Margins: asbalanced
```

	df	F	P>F
machine			
(1)	1	0.75	0.4055
(2)	1	0.31	0.5916
(3)	1	4.47	0.0582
(4)	1	1.73	0.2150
Joint	4	1.93	0.1757
Denominator	11		

`contrast` produces a 1-degree-of-freedom test for each of the specified contrasts as well as a joint test. We included the `noeffects` option so that the table displaying the values of the individual contrasts with their confidence intervals was suppressed.

The significance values above are not adjusted for multiple comparisons. We could have produced the Bonferroni-adjusted significance values by using the `mcompare(bonferroni)` option.

```
. contrast {machine 3 -2 -2  3 -2}
>          {machine 0  1 -2  0  1}
>          {machine 1  0  0 -1  0}
>          {machine 0  1  0  0 -1}, mcompare(bonferroni) noeffects
Contrasts of marginal linear predictions
Margins: asbalanced
```

	df	F	P>F	Bonferroni P>F
machine				
(1)	1	0.75	0.4055	1.0000
(2)	1	0.31	0.5916	1.0000
(3)	1	4.47	0.0582	0.2329
(4)	1	1.73	0.2150	0.8601
Joint	4	1.93	0.1757	
Denominator	11			

```
Note: Bonferroni-adjusted p-values are reported for tests
      on individual contrasts only.
```

Example 10: Linear and quadratic contrasts

Here there are two factors, `A` and `B`, each with three levels. The levels are quantitative so that linear and quadratic contrasts are of interest.

```
. use https://www.stata-press.com/data/r19/twowaytrend
. anova Y A B A#B
```

Number of obs =	36	R-squared =	0.9304		
Root MSE =	2.6736	Adj R-squared =	0.9097		
Source	Partial SS	df	MS	F	Prob>F
Model	2578.5556	8	322.31944	45.09	0.0000
A	2026.7222	2	1013.3611	141.77	0.0000
B	383.72222	2	191.86111	26.84	0.0000
A#B	168.11111	4	42.027778	5.88	0.0015
Residual	193	27	7.1481481		
Total	2771.5556	35	79.187302		

We can use the p. contrast operator to obtain the 1-degree-of-freedom tests for the linear and quadratic effects of A and B.

```
. contrast p.A p.B, noeffects
Contrasts of marginal linear predictions
Margins: asbalanced
```

	df	F	P>F
A			
(linear)	1	212.65	0.0000
(quadratic)	1	70.88	0.0000
Joint	2	141.77	0.0000
B			
(linear)	1	26.17	0.0000
(quadratic)	1	27.51	0.0000
Joint	2	26.84	0.0000
Denominator	27		

All the above tests appear to be significant. In addition to presenting the 1-degree-of-freedom tests, the combined tests for A and B are produced and agree with the original ANOVA results.

Now, we explore the interaction between A and B.

```
. contrast p.A#p1.B, noeffects
Contrasts of marginal linear predictions
Margins: asbalanced
```

	df	F	P>F
A#B			
(linear) (linear)	1	17.71	0.0003
(quadratic) (linear)	1	0.07	0.7893
Joint	2	8.89	0.0011
Denominator	27		

The 2-degrees-of-freedom test of the interaction of A with the linear components of B is significant at the 0.0011 level. But, when we examine the two 1-degree-of-freedom tests that compose this result, the significance is due to the linear A by linear B contrast (significance level of 0.0003). A significance value of 0.7893 for the quadratic A by linear B indicates that this factor is not significant for these data.

```
. contrast p.A#p2.B, noeffects
Contrasts of marginal linear predictions
Margins: asbalanced
```

	df	F	P>F
A#B			
(linear) (quadratic)	1	2.80	0.1058
(quadratic) (quadratic)	1	2.94	0.0979
Joint	2	2.87	0.0741
Denominator	27		

The test of A with the quadratic components of B does not fall below the 0.05 significance level.

◁

Video example

Introduction to contrasts in Stata: One-way ANOVA

References

Mitchell, M. N. 2021. *Interpreting and Visualizing Regression Models Using Stata*. 2nd ed. College Station, TX: Stata Press.

Ostle, B., and R. W. Mensing. 1975. *Statistics in Research*. 3rd ed. Ames, IA: Iowa State University Press.

Rencher, A. C., and G. B. Schaalje. 2008. *Linear Models in Statistics*. 2nd ed. New York: Wiley.

Also see

[R] **anova** — Analysis of variance and covariance

[R] **regress postestimation** — Postestimation tools for regress

[R] **regress postestimation diagnostic plots** — Postestimation plots for regress

[U] 20 Estimation and postestimation commands