pwcompare — Pairwise comparisons

Description Quick start Menu Syntax Options
Remarks and examples Stored results Methods and formulas References Also see

Description

`pwcompare` performs pairwise comparisons across the levels of factor variables from the most recently fit model. `pwcompare` can compare estimated cell means, marginal means, intercepts, marginal intercepts, slopes, or marginal slopes—collectively called margins. `pwcompare` reports the comparisons as contrasts (differences) of margins along with significance tests or confidence intervals for the contrasts. The tests and confidence intervals can be adjusted for multiple comparisons.

`pwcompare` can be used with `svy` estimation results; see [SVY] **svy postestimation**.

See [R] **margins, pwcompare** for performing pairwise comparisons of margins of linear and nonlinear predictions.

Quick start

All pairwise comparisons of the means of `y` across levels of `a` after `regress y i.a`

```
pwcompare a
```

Same as above, and report test statistics and p-values for tests of differences in means

```
pwcompare a, effects
```

Adjust p-values and confidence intervals for multiple comparisons using Tukey's method

```
pwcompare a, effects mcompare(tukey)
```

Same as above, but adjust for multiple comparisons using Bonferroni's method

```
pwcompare a, effects mcompare(bonferroni)
```

Report means for the levels of `a`, and group those that are not significantly different

```
pwcompare a, groups
```

Pairwise comparisons of cell means after `regress y1 a##b`

```
pwcompare a#b
```

Pairwise comparisons of the marginal means of `a`

```
pwcompare a
```

Pairwise comparisons of slopes for continuous `x` after `regress y1 a##c.x`

```
pwcompare a#c.x
```

Pairwise comparisons of log odds after `logit y2 i.a`

```
pwcompare a
```

Pairwise comparisons of the means of `y2` across levels of `a` after `mvreg y1 y2 y3 = i.a`

```
pwcompare a, equation(y2)
```

Same as above, but report pairwise comparisons of `a` for each equation

```
pwcompare a, atequations
```

Pairwise comparisons of overall margins of y1, y2, and y3

```
pwcompare _eqns
```

Menu

Statistics > Postestimation

Syntax

pwcompare *marginlist* [, *options*]

where *marginlist* is a list of factor variables or interactions that appear in the current estimation results or _eqns to reference equations. The variables may be typed with or without the i. prefix, and you may use any factor-variable syntax:

```
. pwcompare i.sex i.group i.sex#i.group

. pwcompare sex group sex#group

. pwcompare sex##group
```

options	Description
Main	
mcompare(*method*)	adjust for multiple comparisons; default is mcompare(noadjust)
asobserved	treat all factor variables as observed
Equations	
equation(*eqspec*)	perform comparisons within equation *eqspec*
atequations	perform comparisons within each equation
Advanced	
emptycells(*empspec*)	treatment of empty cells for balanced factors
noestimcheck	suppress estimability checks
Reporting	
level(#)	confidence level; default is level(95)
cieffects	show effects table with confidence intervals; the default
pveffects	show effects table with *p*-values
effects	show effects table with confidence intervals and *p*-values
cimargins	show table of margins and confidence intervals
groups	show table of margins and group codes
sort	sort the margins or contrasts within each term
post	post margins and their VCEs as estimation results
display_options	control column formats, row spacing, line width, and factor-variable labeling
eform_option	report exponentiated contrasts
df(#)	use *t* distribution with # degrees of freedom for computing *p*-values and confidence intervals

df(#) does not appear in the dialog box.

method	Description
noadjust	do not adjust for multiple comparisons; the default
bonferroni [adjustall]	Bonferroni's method; adjust across all terms
sidak [adjustall]	Šidák's method; adjust across all terms
scheffe	Scheffé's method
* tukey	Tukey's method
* snk	Student–Newman–Keuls's method
* duncan	Duncan's method
* dunnett	Dunnett's method

*tukey, snk, duncan, and dunnett are only allowed with results from anova, manova, regress, and mvreg.
tukey, snk, duncan, and dunnett are not allowed with results from svy.

Time-series operators are allowed if they were used in the estimation.

collect is allowed; see **[U] 11.1.10 Prefix commands**.

Options

Main

mcompare(*method*) specifies the method for computing *p*-values and confidence intervals that account for multiple comparisons within a factor-variable term.

Most methods adjust the comparisonwise error rate, α_c, to achieve a prespecified experimentwise error rate, α_e.

mcompare(noadjust) is the default; it specifies no adjustment.

$$\alpha_c = \alpha_e$$

mcompare(bonferroni) adjusts the comparisonwise error rate based on the upper limit of the Bonferroni inequality:

$$\alpha_e \leq m\alpha_c$$

where *m* is the number of comparisons within the term.

The adjusted comparisonwise error rate is

$$\alpha_c = \alpha_e / m$$

mcompare(sidak) adjusts the comparisonwise error rate based on the upper limit of the probability inequality

$$\alpha_e \leq 1 - (1 - \alpha_c)^m$$

where *m* is the number of comparisons within the term.

The adjusted comparisonwise error rate is

$$\alpha_c = 1 - (1 - \alpha_e)^{1/m}$$

This adjustment is exact when the *m* comparisons are independent.

mcompare(scheffe) controls the experimentwise error rate using the F (or χ^2) distribution with degrees of freedom equal to the rank of the term.

For results from anova, regress, manova, and mvreg (see [R] **anova**, [R] **regress**, [MV] **manova**, and [MV] **mvreg**), pwcompare allows the following additional methods. These methods are not allowed with results that used vce(robust) or vce(cluster *clustvar*).

mcompare(tukey) uses what is commonly referred to as Tukey's honestly significant difference. This method uses the Studentized range distribution instead of the t distribution.

mcompare(snk) is a variation on mcompare(tukey) that counts only the number of margins in the range for a given comparison instead of the full number of margins.

mcompare(duncan) is a variation on mcompare(snk) with additional adjustment to the significance probabilities.

mcompare(dunnett) uses Dunnett's method for making comparisons with a reference category.

mcompare(*method* adjustall) specifies that the multiple-comparison adjustments count all comparisons across all terms rather than performing multiple comparisons term by term. This leads to more conservative adjustments when multiple variables or terms are specified in *marginlist*. This option is compatible only with the bonferroni and sidak methods.

asobserved specifies that factor covariates be evaluated using the cell frequencies observed when the model was fit. The default is to treat all factor covariates as though there were an equal number of observations at each level.

Equations

equation(*eqspec*) specifies the equation from which margins are to be computed. The default is to compute margins from the first equation.

atequations specifies that the margins be computed within each equation.

Advanced

emptycells(*empspec*) specifies how empty cells are handled in interactions involving factor variables that are being treated as balanced.

emptycells(strict) is the default; it specifies that margins involving empty cells be treated as not estimable.

emptycells(reweight) specifies that the effects of the observed cells be increased to accommodate any missing cells. This makes the margins estimable but changes their interpretation.

noestimcheck specifies that pwcompare not check for estimability. By default, the requested margins are checked and those found not estimable are reported as such. Nonestimability is usually caused by empty cells. If noestimcheck is specified, estimates are computed in the usual way and reported even though the resulting estimates are manipulable, which is to say they can differ across equivalent models having different parameterizations.

Reporting

level(#) specifies the confidence level, as a percentage, for confidence intervals. The default is level(95) or as set by set level; see **[U] 20.8 Specifying the width of confidence intervals**. The significance level used by the groups option is 100 − #, expressed as a percentage.

cieffects specifies that a table of the pairwise comparisons with their standard errors and confidence intervals be reported. This is the default.

pveffects specifies that a table of the pairwise comparisons with their standard errors, test statistics, and p-values be reported.

effects specifies that a table of the pairwise comparisons with their standard errors, test statistics, p-values, and confidence intervals be reported.

cimargins specifies that a table of the margins with their standard errors and confidence intervals be reported.

groups specifies that a table of the margins with their standard errors and group codes be reported. Margins with the same letter in the group code are not significantly different at the specified significance level.

sort specifies that the reported tables be sorted on the margins or differences in each term.

post causes pwcompare to behave like a Stata estimation (e-class) command. pwcompare posts the vector of estimated margins along with the estimated variance–covariance matrix to e(), so you can treat the estimated margins just as you would results from any other estimation command. For example, you could use test to perform simultaneous tests of hypotheses on the margins, or you could use lincom to create linear combinations.

display_options: vsquish, nofvlabel, fvwrap(#), fvwrapon(*style*), cformat(%*fmt*), pformat(%*fmt*), sformat(%*fmt*), and nolstretch.

vsquish specifies that the blank space separating factor-variable terms or time-series–operated variables from other variables in the model be suppressed.

nofvlabel displays factor-variable level values rather than attached value labels. This option overrides the fvlabel setting; see [R] **set showbaselevels**.

fvwrap(#) specifies how many lines to allow when long value labels must be wrapped. Labels requiring more than # lines are truncated. This option overrides the fvwrap setting; see [R] **set showbaselevels**.

fvwrapon(*style*) specifies whether value labels that wrap will break at word boundaries or break based on available space.

fvwrapon(word), the default, specifies that value labels break at word boundaries.

fvwrapon(width) specifies that value labels break based on available space.

This option overrides the fvwrapon setting; see [R] **set showbaselevels**.

cformat(%*fmt*) specifies how to format contrasts or margins, standard errors, and confidence limits in the table of pairwise comparisons.

pformat(%*fmt*) specifies how to format p-values in the table of pairwise comparisons.

sformat(%*fmt*) specifies how to format test statistics in the table of pairwise comparisons.

nolstretch specifies that the width of the table of pairwise comparisons not be automatically widened to accommodate longer variable names. The default, lstretch, is to automatically widen the table of pairwise comparisons up to the width of the Results window. Specifying lstretch or nolstretch overrides the setting given by set lstretch. If set lstretch has not been set, the default is lstretch. nolstretch is not shown in the dialog box.

eform_option specifies that the contrasts table be displayed in exponentiated form. e^{contrast} is displayed rather than contrast. Standard errors and confidence intervals are also transformed. See [R] ***eform_option*** for the list of available options.

The following option is available with pwcompare but is not shown in the dialog box:

df(#) specifies that the t distribution with # degrees of freedom be used for computing p-values and confidence intervals. The default is to use e(df_r) degrees of freedom or the standard normal distribution if e(df_r) is missing.

Remarks and examples

pwcompare performs pairwise comparisons of margins across the levels of factor variables from the most recently fit model. The margins can be estimated cell means, marginal means, intercepts, marginal intercepts, slopes, or marginal slopes. With the exception of slopes, we can also consider these margins to be marginal linear predictions.

The margins are calculated as linear combinations of the coefficients. Let k be the number of levels for a factor term in our model; then there are k margins for that term, and

$$m = \binom{k}{2} = \frac{k(k-1)}{2}$$

unique pairwise comparisons of those margins.

The confidence intervals and p-values for these pairwise comparisons can be adjusted to account for multiple comparisons. Bonferroni's, Šidák's, and Scheffé's adjustments can be made for multiple comparisons after fitting any type of model. In addition, Tukey's, Student–Newman–Keuls's, Duncan's, and Dunnett's adjustments are available when fitting ANOVA, linear regression, MANOVA, or multivariate regression models.

Remarks are presented under the following headings:

Pairwise comparisons of means
- *Marginal means*
- *All pairwise comparisons*

Overview of multiple-comparison methods
- *Fisher's protected least-significant difference (LSD)*
- *Bonferroni's adjustment*
- *Šidák's adjustment*
- *Scheffé's adjustment*
- *Tukey's HSD adjustment*
- *Student–Newman–Keuls's adjustment*
- *Duncan's adjustment*
- *Dunnett's adjustment*

Example adjustments using one-way models
- *Fisher's protected LSD*
- *Tukey's HSD*
- *Dunnett's method for comparisons to a control*

Two-way models
Pairwise comparisons of slopes
Nonlinear models
Multiple-equation models
Unbalanced data
Empty cells

Pairwise comparisons of means

Suppose we are interested in the effects of five different fertilizers on wheat yield. We could estimate the following linear regression model to determine the effect of each type of fertilizer on the yield.

```
. use https://www.stata-press.com/data/r19/yield
(Artificial wheat yield dataset)

. regress yield i.fertilizer

      Source |       SS           df       MS      Number of obs   =       200
-------------+----------------------------------   F(4, 195)       =      5.33
       Model |  1078.84207         4  269.710517   Prob > F        =    0.0004
    Residual |  9859.55334       195   50.561812   R-squared       =    0.0986
-------------+----------------------------------   Adj R-squared   =    0.0801
       Total |  10938.3954       199  54.9668111   Root MSE        =    7.1107

------------------------------------------------------------------------------
       yield | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
  fertilizer |
   10-08-22  |    3.62272   1.589997     2.28   0.024     .4869212    6.758518
   16-04-08  |   .4906299   1.589997     0.31   0.758    -2.645169    3.626428
   18-24-06  |   4.922803   1.589997     3.10   0.002     1.787005    8.058602
   29-03-04  |  -1.238328   1.589997    -0.78   0.437    -4.374127     1.89747
             |
       _cons |   41.36243   1.124298    36.79   0.000     39.14509    43.57977
------------------------------------------------------------------------------
```

In this simple case, the coefficients for fertilizers 10-08-22, 16-04-08, 18-24-06, and 29-03-04 indicate the difference in the mean yield for that fertilizer versus the mean yield for fertilizer 10-10-10. That the standard errors of all four coefficients are identical results from having perfectly balanced data.

Marginal means

We can use `pwcompare` with the `cimargins` option to compute the mean yield for each of the fertilizers.

```
. pwcompare fertilizer, cimargins

Pairwise comparisons of marginal linear predictions

Margins: asbalanced

--------------------------------------------------------------
             |                              Unadjusted
             |     Margin   Std. err.     [95% conf. interval]
-------------+------------------------------------------------
  fertilizer |
   10-10-10  |   41.36243   1.124298      39.14509    43.57977
   10-08-22  |   44.98515   1.124298       42.7678    47.20249
   16-04-08  |   41.85306   1.124298      39.63571     44.0704
   18-24-06  |   46.28523   1.124298      44.06789    48.50258
   29-03-04  |    40.1241   1.124298      37.90676    42.34145
--------------------------------------------------------------
```

Looking at the confidence intervals for fertilizers 10-10-10 and 10-08-22 in the table above, we might be tempted to conclude that these means are not significantly different because the intervals overlap. However, as discussed in *Interaction plots* of [R] **marginsplot**, we cannot draw conclusions about the differences in means by looking at confidence intervals for the means themselves. Instead, we would need to look at confidence intervals for the difference in means.

All pairwise comparisons

By default, pwcompare calculates all pairwise differences of the margins, in this case pairwise differences of the mean yields.

```
. pwcompare fertilizer

Pairwise comparisons of marginal linear predictions

Margins: asbalanced
```

			Unadjusted	
	Contrast	Std. err.	[95% conf.	interval]
fertilizer				
10-08-22 vs 10-10-10	3.62272	1.589997	.4869212	6.758518
16-04-08 vs 10-10-10	.4906299	1.589997	-2.645169	3.626428
18-24-06 vs 10-10-10	4.922803	1.589997	1.787005	8.058602
29-03-04 vs 10-10-10	-1.238328	1.589997	-4.374127	1.89747
16-04-08 vs 10-08-22	-3.13209	1.589997	-6.267889	.0037086
18-24-06 vs 10-08-22	1.300083	1.589997	-1.835715	4.435882
29-03-04 vs 10-08-22	-4.861048	1.589997	-7.996847	-1.725249
18-24-06 vs 16-04-08	4.432173	1.589997	1.296375	7.567972
29-03-04 vs 16-04-08	-1.728958	1.589997	-4.864757	1.406841
29-03-04 vs 18-24-06	-6.161132	1.589997	-9.29693	-3.025333

If a confidence interval does not include zero, the means for the compared fertilizers are significantly different. Therefore, at the 5% significance level, we would reject the hypothesis that the means for fertilizers 10-10-10 and 10-08-22 are equivalent—as we would do for 18-24-06 vs 10-10-10, 29-03-04 vs 10-08-22, 18-24-06 vs 16-04-08, and 29-03-04 vs 18-24-06.

We may prefer to see the p-values instead of looking at confidence intervals to determine whether the pairwise differences are significantly different from zero. We could use the pveffects option to see the differences with standard errors and p-values, or we could use the effects option to see both p-values and confidence intervals in the same table. Here we specify effects as well as the sort option so that the differences are sorted from smallest to largest.

```
. pwcompare fertilizer, effects sort
Pairwise comparisons of marginal linear predictions
Margins: asbalanced
```

	Contrast	Std. err.	Unadjusted t	Unadjusted P>|t|	Unadjusted [95% conf.	interval]
fertilizer						
29-03-04 vs 18-24-06	-6.161132	1.589997	-3.87	0.000	-9.29693	-3.025333
29-03-04 vs 10-08-22	-4.861048	1.589997	-3.06	0.003	-7.996847	-1.725249
16-04-08 vs 10-08-22	-3.13209	1.589997	-1.97	0.050	-6.267889	.0037086
29-03-04 vs 16-04-08	-1.728958	1.589997	-1.09	0.278	-4.864757	1.406841
29-03-04 vs 10-10-10	-1.238328	1.589997	-0.78	0.437	-4.374127	1.89747
16-04-08 vs 10-10-10	.4906299	1.589997	0.31	0.758	-2.645169	3.626428
18-24-06 vs 10-08-22	1.300083	1.589997	0.82	0.415	-1.835715	4.435882
10-08-22 vs 10-10-10	3.62272	1.589997	2.28	0.024	.4869212	6.758518
18-24-06 vs 16-04-08	4.432173	1.589997	2.79	0.006	1.296375	7.567972
18-24-06 vs 10-10-10	4.922803	1.589997	3.10	0.002	1.787005	8.058602

We find that 5 of the 10 pairs of means are significantly different at the 5% significance level.

We can use the groups option to obtain a table that identifies groups whose means are not significantly different by assigning them the same letter.

```
. pwcompare fertilizer, groups sort
Pairwise comparisons of marginal linear predictions
Margins: asbalanced
```

	Margin	Std. err.	Unadjusted groups
fertilizer			
29-03-04	40.1241	1.124298	A
10-10-10	41.36243	1.124298	A
16-04-08	41.85306	1.124298	AB
10-08-22	44.98515	1.124298	BC
18-24-06	46.28523	1.124298	C

```
Note: Margins sharing a letter in the group label
      are not significantly different at the 5%
      level.
```

The letter A that is assigned to fertilizers 29-03-04, 10-10-10, and 16-04-08 designates that the mean yields for these fertilizers are not different at the 5% level.

Overview of multiple-comparison methods

For a single test, if we choose a 5% significance level, we would have a 5% chance of concluding that two margins are different when the population values are actually equal. This is known as making a type I error. When we perform $m = k(k-1)/2$ pairwise comparisons of the k margins, we have m opportunities to make a type I error.

pwcompare with the mcompare() option allows us to adjust the confidence intervals and p-values for each comparison to account for the increased probability of making a type I error when making multiple comparisons. Bonferroni's adjustment, Šidák's adjustment, and Scheffé's adjustment can be used when making pairwise comparisons of the margins after any estimation command. Tukey's honestly significant difference, Student–Newman–Keuls's method, Duncan's method, and Dunnett's method are only available when fitting linear models after anova, manova, regress, or mvreg.

Fisher's protected least-significant difference (LSD)

pwcompare does not offer an mcompare() option specifically for Fisher's protected least-significant difference (LSD). In this methodology, no adjustment is made to the confidence intervals or p-values. However, it is protected in the sense that no pairwise comparisons are tested unless the joint test for the corresponding term in the model is significant. Therefore, the default mcompare(noadjust) corresponds to Fisher's protected LSD assuming that the corresponding joint test was performed before using pwcompare.

Milliken and Johnson (2009) recommend using this methodology for planned comparisons, assuming the corresponding joint test is significant.

Bonferroni's adjustment

mcompare(bonferroni) adjusts significance levels based on the Bonferroni inequality, which, in the case of multiple testing, tells us that the maximum error rate for all comparisons is the sum of the error rates for the individual comparisons. Assuming that we are using the same significance level for all tests, the experimentwise error rate is the error rate for a single test multiplied by the number of comparisons. Therefore, a p-value for each comparison can be computed by multiplying the unadjusted p-value by the total number of comparisons. If the adjusted p-value is greater than 1, then pwcompare will report a p-value of 1.

Bonferroni's adjustment is popular because it is easy to compute manually and because it can be applied to any set of tests, not only the pairwise comparisons available in pwcompare. In addition, this method does not require equal sample sizes.

Because Bonferroni's adjustment is so general, it is more conservative than many of the other adjustments. It is especially conservative when a large number of tests is being performed.

Šidák's adjustment

mcompare(sidak) performs an adjustment using Šidák's method. This adjustment, like Bonferroni's adjustment, is derived from an inequality. However, in this case, the inequality is based on the probability of not making a type I error. For a single test, the probability that we do not make a type I error is $1-\alpha$. For two independent tests, both using α as a significance level, the probability is $(1-\alpha)(1-\alpha)$. Likewise, for m independent tests, the probability of not making a type I error is $(1-\alpha)^m$. Therefore, the probability of making one or more type I errors is $1-(1-\alpha)^m$. When tests are not independent, the probability of making at least one error is less than $1-(1-\alpha)^m$. Therefore, we can compute an adjusted p-value as $1-(1-{}_u p)^m$, where ${}_u p$ is the unadjusted p-value for a single comparison.

Šidák's method is also conservative although slightly less so than Bonferroni's method. Like Bonferroni's method, this method does not require equal sample sizes.

Scheffé's adjustment

Scheffé's adjustment is used when mcompare(scheffe) is specified. This adjustment is derived from the joint F test and its correspondence to the maximum normalized comparison. To adjust for multiple comparisons, the absolute value of the t statistic for a particular comparison can be compared with a critical value of $\sqrt{(k-1)F_{k-1,\nu}}$, where ν is the residual degrees of freedom. $F_{k-1,\nu}$ is the distribution of the joint F test for the corresponding term in a one-way ANOVA model. Winer, Brown, and Michels (1991, 191–195) discuss this in detail. For estimation commands that report z statistics instead of t statistics for the tests on coefficients, a χ^2 distribution is used instead of an F distribution.

Scheffé's method allows for making all possible comparisons of the k margins, not just the pairwise comparisons. Unlike the methods described above, it does not take into account the number of comparisons that are currently being made. Therefore, this method is even more conservative than the others. Because this method adjusts for all possible comparisons of the levels of the term, Milliken and Johnson (2009) recommend using this procedure when making unplanned contrasts that are suggested by the data. As Winer, Brown, and Michels (1991, 191) put it, this method is often used to adjust for "unfettered data snooping". When using this adjustment, a contrast will never be significant if the joint F or χ^2 test for the term is not also significant.

This is another method that does not require equal sample sizes.

Tukey's HSD adjustment

Tukey's adjustment is also referred to as Tukey's honestly significant difference (HSD) and is used when `mcompare(tukey)` is specified. It is often applied to all pairwise comparisons of means. Tukey's HSD is commonly used as a post hoc test although this is not a requirement.

To adjust for multiple comparisons, Tukey's method compares the absolute value of the t statistic from the individual comparison with a critical value based on a Studentized range distribution with parameter equal to the number of levels in the term. When applied to pairwise comparisons of means,

$$q = \frac{\text{mean}_{\text{max}} - \text{mean}_{\text{min}}}{s}$$

follows a Studentized range distribution with parameter k and ν degrees of freedom. Here mean_{max} and mean_{min} are the largest and smallest marginal means, and s is an estimate of the standard error of the means.

Now for the comparison of the smallest and largest means, we can say that the probability of not making a type I error is

$$\Pr\left(\frac{\text{mean}_{\text{max}} - \text{mean}_{\text{min}}}{s} \le q_{k,\nu}\right) = 1 - \alpha$$

Then, the following inequality holds for all pairs of means simultaneously:

$$\Pr\left(\frac{|\text{mean}_i - \text{mean}_j|}{s} \le q_{k,\nu}\right) \ge 1 - \alpha$$

Based on this procedure, Tukey's HSD computes the p-value for each of the individual comparisons using the Studentized range distribution. However, because the equality holds only for the difference in the largest and smallest means, this procedure produces conservative tests for the remaining comparisons. Winer, Brown, and Michels (1991, 172–182) discuss this in further detail.

With unequal sample sizes, `mcompare(tukey)` produces the Tukey–Kramer adjustment (Tukey 1953; Kramer 1956).

Student–Newman–Keuls's adjustment

The Student–Newman–Keuls (SNK) method is used when `mcompare(snk)` is specified. It is a modification to Tukey's method and is less conservative. In this procedure, we first order the means. We then test the difference in the smallest and largest means using a critical value from the Studentized range distribution with parameter k, where k is the number of levels in the term. This step uses the same methodology as in Tukey's procedure. However, in the next step, we will then test for differences in the two sets of means that are the endpoints of the two ranges including $k-1$ means. Specifically, we test the difference in the smallest mean and the second-largest mean using a critical value from the Studentized range distribution with parameter $k-1$. We would also test the difference in the second-smallest mean and the largest mean using this critical value. Likewise, the means that are the endpoints of ranges including $k-2$ means when ordered are tested using the Studentized range distribution with parameter $k-2$, and so on.

Equal sample sizes are required for this method.

Duncan's adjustment

When mcompare(duncan) is specified, tests are adjusted for multiple comparisons using Duncan's method, which is sometimes referred to as Duncan's new multiple range method. This adjustment produces tests that are less conservative than both Tukey's HSD and SNK. This procedure is performed in the same manner as SNK except that the p-values for the individual comparisons are adjusted as $1 - (1 - {}_{\text{snk}}p_i)^{1/(r+1)}$, where ${}_{\text{snk}}p$ is the p-value computed using the SNK method and r represents the number of means that, when ordered, fall between the two that are being compared.

Again, equal sample sizes are required for this adjustment.

Dunnett's adjustment

Dunnett's adjustment is obtained by specifying mcompare(dunnett). It is used when one of the levels of a factor can be considered a control or reference level with which each of the other levels is being compared. When Dunnett's adjustment is requested, $k-1$ instead of $k(k-1)/2$ pairwise comparisons are made. Dunnett (1955, 1964) developed tables of critical values for what Miller (1981, 76) refers to as the "many-one t statistic". The t statistics for individual comparisons are compared with these critical values when making many comparisons to a single reference level.

This method also requires equal sample sizes.

Example adjustments using one-way models

Fisher's protected LSD

Fisher's protected LSD requires that we first verify that the joint test for a term in our model is significant before proceeding with pairwise comparisons. Using our previous example, we could have first used the contrast command to obtain a joint test for the effects of fertilizer.

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

	df	F	P>F
fertilizer	4	5.33	0.0004
Denominator	195		

This test for the effects of fertilizer is highly significant. Now we can say we are using Fisher's protected LSD when looking at the unadjusted p-values that were obtained from our previous command,

```
. pwcompare fertilizer, effects sort
```

Tukey's HSD

Because we fit a linear regression model and are interested in all pairwise comparisons of the marginal means, we may instead choose to use Tukey's HSD.

```
. pwcompare fertilizer, effects sort mcompare(tukey)
Pairwise comparisons of marginal linear predictions
Margins: asbalanced
```

	Number of comparisons
fertilizer	10

	Contrast	Std. err.	Tukey t	Tukey P>|t|	Tukey [95% conf.	interval]
fertilizer						
29-03-04 vs 18-24-06	-6.161132	1.589997	-3.87	0.001	-10.53914	-1.78312
29-03-04 vs 10-08-22	-4.861048	1.589997	-3.06	0.021	-9.239059	-.4830368
16-04-08 vs 10-08-22	-3.13209	1.589997	-1.97	0.285	-7.510101	1.245921
29-03-04 vs 16-04-08	-1.728958	1.589997	-1.09	0.813	-6.106969	2.649053
29-03-04 vs 10-10-10	-1.238328	1.589997	-0.78	0.936	-5.616339	3.139683
16-04-08 vs 10-10-10	.4906299	1.589997	0.31	0.998	-3.887381	4.868641
18-24-06 vs 10-08-22	1.300083	1.589997	0.82	0.925	-3.077928	5.678095
10-08-22 vs 10-10-10	3.62272	1.589997	2.28	0.156	-.7552913	8.000731
18-24-06 vs 16-04-08	4.432173	1.589997	2.79	0.046	.0541623	8.810185
18-24-06 vs 10-10-10	4.922803	1.589997	3.10	0.019	.5447922	9.300815

This time, our p-values have been modified, and we find that only four of the pairwise differences are considered significantly different from zero at the 5% level.

If we are interested only in performing pairwise comparisons of a subset of our means, we can use factor-variable operators to select the levels of the factor that we want to compare. Here we exclude all comparisons involving fertilizer 10-10-10.

```
. pwcompare i(2/5).fertilizer, effects sort mcompare(tukey)
Pairwise comparisons of marginal linear predictions
Margins: asbalanced
```

	Number of comparisons
fertilizer	6

	Contrast	Std. err.	t	Tukey P>|t|	Tukey [95% conf.	interval]
fertilizer						
29-03-04 vs 18-24-06	-6.161132	1.589997	-3.87	0.001	-10.28133	-2.040937
29-03-04 vs 10-08-22	-4.861048	1.589997	-3.06	0.013	-8.981242	-.7408538
16-04-08 vs 10-08-22	-3.13209	1.589997	-1.97	0.203	-7.252284	.9881042
29-03-04 vs 16-04-08	-1.728958	1.589997	-1.09	0.698	-5.849152	2.391236
18-24-06 vs 10-08-22	1.300083	1.589997	0.82	0.846	-2.820111	5.420278
18-24-06 vs 16-04-08	4.432173	1.589997	2.79	0.030	.3119792	8.552368

The adjusted p-values and confidence intervals differ from those in the previous output because Tukey's adjustment takes into account the total number of comparisons being made when determining the appropriate degrees of freedom to use for the Studentized range distribution.

Dunnett's method for comparisons to a control

If one of our five fertilizer groups represents fields where no fertilizer was applied, we may want to use Dunnett's method to compare each of the four fertilizers with the control group. In this case, we make only $k - 1$ comparisons for k groups.

```
. pwcompare fertilizer, effects mcompare(dunnett)
Pairwise comparisons of marginal linear predictions
Margins: asbalanced
```

	Number of comparisons
fertilizer	4

	Contrast	Std. err.	Dunnett t	Dunnett P>|t|	Dunnett [95% conf.	interval]
fertilizer						
10-08-22 vs 10-10-10	3.62272	1.589997	2.28	0.079	-.2918331	7.537273
16-04-08 vs 10-10-10	.4906299	1.589997	0.31	0.994	-3.423923	4.405183
18-24-06 vs 10-10-10	4.922803	1.589997	3.10	0.008	1.00825	8.837356
29-03-04 vs 10-10-10	-1.238328	1.589997	-0.78	0.852	-5.152881	2.676225

In our previous `regress` command, fertilizer 10-10-10 was treated as the base. Therefore, by default, it was treated as the control when using Dunnett's adjustment, and the pairwise comparisons are equivalent to the coefficients reported by `regress`. Based on our `regress` output, we would conclude that fertilizers 10-08-22 and 18-24-06 are different from fertilizer 10-10-10 at the 5% level. However, using Dunnett's adjustment, we find only fertilizer 18-24-06 to be different from fertilizer 10-10-10 at this same significance level.

If the model is fit without a base level for a factor variable, then `pwcompare` will choose the first level as the reference level. If we want to make comparisons with a different level than the one `mcompare(dunnett)` chooses by default, we can use the `b.` operator to override the default. Here we use fertilizer 5 (29-03-04) as the reference level.

```
. pwcompare b5.fertilizer, effects sort mcompare(dunnett)
Pairwise comparisons of marginal linear predictions
Margins: asbalanced
```

	Number of comparisons
fertilizer	4

	Contrast	Std. err.	Dunnett t	P>|t|	Dunnett [95% conf.	interval]
fertilizer						
10-10-10 vs 29-03-04	1.238328	1.589997	0.78	0.852	-2.676225	5.152881
16-04-08 vs 29-03-04	1.728958	1.589997	1.09	0.649	-2.185595	5.643511
10-08-22 vs 29-03-04	4.861048	1.589997	3.06	0.009	.9464951	8.775601
18-24-06 vs 29-03-04	6.161132	1.589997	3.87	0.001	2.246579	10.07568

Two-way models

In the previous examples, we have performed pairwise comparisons after fitting a model with a single factor. Now, we include two factors and their interaction in our model.

```
. regress yield fertilizer##irrigation
```

Source	SS	df	MS			
				Number of obs	=	200
				F(9, 190)	=	27.63
Model	6200.81605	9	688.979561	Prob > F	=	0.0000
Residual	4737.57936	190	24.9346282	R-squared	=	0.5669
				Adj R-squared	=	0.5464
Total	10938.3954	199	54.9668111	Root MSE	=	4.9935

yield	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
fertilizer						
10-08-22	1.882256	1.57907	1.19	0.235	-1.232505	4.997016
16-04-08	-.5687418	1.57907	-0.36	0.719	-3.683502	2.546019
18-24-06	4.904999	1.57907	3.11	0.002	1.790239	8.01976
29-03-04	-1.217496	1.57907	-0.77	0.442	-4.332257	1.897264
1.irrigation	8.899721	1.57907	5.64	0.000	5.784961	12.01448
fertilizer#irrigation						
10-08-22#1	3.480928	2.233143	1.56	0.121	-.9240084	7.885865
16-04-08#1	2.118743	2.233143	0.95	0.344	-2.286193	6.52368
18-24-06#1	.0356082	2.233143	0.02	0.987	-4.369328	4.440545
29-03-04#1	-.0416636	2.233143	-0.02	0.985	-4.4466	4.363273
_cons	36.91257	1.116571	33.06	0.000	34.7101	39.11504

We can perform pairwise comparisons of the cell means defined by the fertilizer and irrigation interaction.

```
. pwcompare fertilizer#irrigation, sort groups mcompare(tukey)
Pairwise comparisons of marginal linear predictions
Margins: asbalanced
```

	Number of comparisons
fertilizer#irrigation	45

	Margin	Std. err.	Tukey groups
fertilizer#irrigation			
29-03-04#0	35.69507	1.116571	A
16-04-08#0	36.34383	1.116571	A
10-10-10#0	36.91257	1.116571	AB
10-08-22#0	38.79482	1.116571	AB
18-24-06#0	41.81757	1.116571	BC
29-03-04#1	44.55313	1.116571	CD
10-10-10#1	45.81229	1.116571	CDE
16-04-08#1	47.36229	1.116571	DEF
18-24-06#1	50.7529	1.116571	EF
10-08-22#1	51.17547	1.116571	F

```
Note: Margins sharing a letter in the group label are
      not significantly different at the 5% level.
```

Based on Tukey's HSD and a 5% significance level, we would conclude that the mean yield for fertilizer 29-03-04 without irrigation is not significantly different from the mean yields for fertilizers 10-10-10, 10-08-22, and 16-04-08 when used without irrigation but is significantly different from the remaining means.

Up to this point, most of the pairwise comparisons that we have performed could have also been obtained with `pwmean` (see [R] **pwmean**) if we had not been interested in examining the results from the estimation command before making pairwise comparisons of the means. For instance, we could reproduce the results from the above `pwcompare` command by typing

```
. pwmean yield, over(fertilizer irrigation) sort group mcompare(tukey)
```

However, `pwcompare` extends the capabilities of `pwmean` in many ways. For instance, `pwmean` only allows for pairwise comparisons of the cell means determined by the highest-level interaction of the variables specified in the `over()` option. However, `pwcompare` allows us to fit a single model, such as the two-way model that we fit above,

```
. regress yield fertilizer##irrigation
```

and compute pairwise comparisons of the marginal means for only one of the variables in the model:

```
. pwcompare fertilizer, sort effects mcompare(tukey)
Pairwise comparisons of marginal linear predictions
Margins: asbalanced
```

```
             Number of
           comparisons

fertilizer          10
```

```
                                                 Tukey                 Tukey
               Contrast   Std. err.         t    P>|t|     [95% conf. interval]

  fertilizer
    29-03-04
          vs
    18-24-06  -6.161132   1.116571     -5.52   0.000     -9.236338   -3.085925
    29-03-04
          vs
    10-08-22  -4.861048   1.116571     -4.35   0.000     -7.936255   -1.785841
    16-04-08
          vs
    10-08-22   -3.13209   1.116571     -2.81   0.044     -6.207297   -.0568832
    29-03-04
          vs
    16-04-08  -1.728958   1.116571     -1.55   0.532     -4.804165    1.346249
    29-03-04
          vs
    10-10-10  -1.238328   1.116571     -1.11   0.802     -4.313535    1.836879
    16-04-08
          vs
    10-10-10   .4906299   1.116571      0.44   0.992     -2.584577    3.565837
    18-24-06
          vs
    10-08-22   1.300083   1.116571      1.16   0.772     -1.775123     4.37529
    10-08-22
          vs
    10-10-10    3.62272   1.116571      3.24   0.012      .5475131    6.697927
    18-24-06
          vs
    16-04-08   4.432173   1.116571      3.97   0.001      1.356967     7.50738
    18-24-06
          vs
    10-10-10   4.922803   1.116571      4.41   0.000      1.847597     7.99801
```

Here the standard errors for the differences in marginal means and the residual degrees of freedom are based on the full model. Therefore, the results will differ from those obtained from `pwcompare` after fitting the one-way model with only `fertilizer` (or equivalently using `pwmean`).

Pairwise comparisons of slopes

If we fit a model with a factor variable that is interacted with a continuous variable, `pwcompare` will even allow us to make pairwise comparisons of the slopes of the continuous variable for the levels of the factor variable.

In this case, we have a continuous variable, `NO3_N`, indicating the amount of nitrate nitrogen already existing in the soil, based on a sample taken from each field.

```
. regress yield fertilizer##c.NO3_N

      Source |       SS           df       MS      Number of obs   =       200
-------------+----------------------------------   F(9, 190)       =     37.61
       Model |  7005.69932         9  778.411035   Prob > F        =    0.0000
    Residual |  3932.69609       190  20.6984005   R-squared       =    0.6405
-------------+----------------------------------   Adj R-squared   =    0.6234
       Total |  10938.3954       199  54.9668111   Root MSE        =    4.5495

------------------------------------------------------------------------------
       yield | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
  fertilizer |
   10-08-22  |   18.65019   8.452061     2.21   0.029      1.97826    35.32212
   16-04-08  |  -13.34076   10.07595    -1.32   0.187    -33.21585    6.534327
   18-24-06  |   24.35061   9.911463     2.46   0.015     4.799973    43.90125
   29-03-04  |   17.58529   8.446736     2.08   0.039     .9238646    34.24671
             |
       NO3_N |   4.915653   .7983509     6.16   0.000     3.340884    6.490423
             |
 fertilizer#|
     c.NO3_N |
   10-08-22  |  -1.282039   .8953419    -1.43   0.154    -3.048126    .4840487
   16-04-08  |   -1.00571   .9025862    -1.11   0.267    -2.786087    .7746662
   18-24-06  |   -2.97627   .9136338    -3.26   0.001    -4.778438   -1.174102
   29-03-04  |  -3.275947   .8247385    -3.97   0.000    -4.902767   -1.649127
             |
       _cons |  -5.459168   7.638241    -0.71   0.476    -20.52581    9.607477
------------------------------------------------------------------------------
```

These are the pairwise differences of the slopes of NO3_N for each pair of fertilizers:

```
. pwcompare fertilizer#c.NO3_N, pveffects sort mcompare(scheffe)
Pairwise comparisons of marginal linear predictions
Margins: asbalanced
```

	Number of comparisons
fertilizer#c.NO3_N	10

	Contrast	Std. err.	Scheffe t	Scheffe P>|t|
fertilizer#c.NO3_N				
29-03-04 vs 10-10-10	-3.275947	.8247385	-3.97	0.004
18-24-06 vs 10-10-10	-2.97627	.9136338	-3.26	0.034
29-03-04 vs 16-04-08	-2.270237	.4691771	-4.84	0.000
29-03-04 vs 10-08-22	-1.993909	.4550851	-4.38	0.001
18-24-06 vs 16-04-08	-1.97056	.612095	-3.22	0.038
18-24-06 vs 10-08-22	-1.694232	.6013615	-2.82	0.099
10-08-22 vs 10-10-10	-1.282039	.8953419	-1.43	0.727
16-04-08 vs 10-10-10	-1.00571	.9025862	-1.11	0.871
29-03-04 vs 18-24-06	-.2996772	.4900939	-0.61	0.984
16-04-08 vs 10-08-22	.276328	.5844405	0.47	0.994

Using Scheffé's adjustment, we find that five of the pairs have significantly different slopes at the 5% level.

Nonlinear models

pwcompare can also perform pairwise comparisons of the marginal linear predictions after fitting a nonlinear model. For instance, we can use the dataset from *Beyond linear models* in [R] **contrast** and fit the following logistic regression model of patient satisfaction on hospital:

```
. use https://www.stata-press.com/data/r19/hospital
(Artificial hospital satisfaction data)
. logit satisfied i.hospital
Iteration 0:  Log likelihood = -393.72216
Iteration 1:  Log likelihood = -387.55736
Iteration 2:  Log likelihood =  -387.4768
Iteration 3:  Log likelihood = -387.47679
Logistic regression                                     Number of obs =    802
                                                        LR chi2(2)    =  12.49
                                                        Prob > chi2   = 0.0019
Log likelihood = -387.47679                             Pseudo R2     = 0.0159
```

satisfied	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
hospital						
2	.5348129	.2136021	2.50	0.012	.1161604	.9534654
3	.7354519	.2221929	3.31	0.001	.2999618	1.170942
_cons	1.034708	.1391469	7.44	0.000	.7619855	1.307431

For this model, the marginal linear predictions are the predicted log odds for each hospital and can be obtained with the `cimargins` option:

```
. pwcompare hospital, cimargins
Pairwise comparisons of marginal linear predictions
Margins: asbalanced
```

	Margin	Std. err.	Unadjusted [95% conf.	interval]
hospital				
1	1.034708	.1391469	.7619855	1.307431
2	1.569521	.1620618	1.251886	1.887157
3	1.77016	.1732277	1.43064	2.10968

The pairwise comparisons are, therefore, differences in the log odds. We can specify `mcompare(bonferroni)` and `effects` to request Bonferroni-adjusted p-values and confidence intervals.

```
. pwcompare hospital, effects mcompare(bonferroni)
Pairwise comparisons of marginal linear predictions
Margins: asbalanced
```

	Number of comparisons
satisfied	
hospital	3

	Contrast	Std. err.	Bonferroni z	Bonferroni P>|z|	Bonferroni [95% conf.	interval]
satisfied						
hospital						
2 vs 1	.5348129	.2136021	2.50	0.037	.0234537	1.046172
3 vs 1	.7354519	.2221929	3.31	0.003	.2035265	1.267377
3 vs 2	.200639	.2372169	0.85	1.000	-.3672535	.7685314

For nonlinear models, only Bonferroni's adjustment, Šidák's adjustment, and Scheffé's adjustment are available.

If we want pairwise comparisons reported as odds ratios, we can specify the or option.

```
. pwcompare hospital, effects mcompare(bonferroni) or
Pairwise comparisons of marginal linear predictions
Margins: asbalanced
```

	Number of comparisons
satisfied	
hospital	3

	Odds ratio	Std. err.	Bonferroni z	Bonferroni P>|z|	Bonferroni [95% conf.	interval]
satisfied						
hospital						
2 vs 1	1.707129	.3646464	2.50	0.037	1.023731	2.846733
3 vs 1	2.086425	.4635888	3.31	0.003	1.225718	3.551525
3 vs 2	1.222183	.2899226	0.85	1.000	.6926341	2.156597

Notice that these tests are still performed on the marginal linear predictions. The odds ratios reported here are the exponentiated versions of the pairwise differences of log odds in the previous output. For further discussion, see [R] **contrast**.

Multiple-equation models

pwcompare works with models containing multiple equations. Commands such as intreg and gnbreg allow their ancillary parameters to be modeled as a function of independent variables, and pwcompare can compare the margins within these equations. The equation() option can be used to specify the equation for which pairwise comparisons of the margins should be made. The atequations option specifies that pairwise comparisons be computed for each equation. In addition, pwcompare allows a special pseudofactor for equation—called _eqns—when working with results from manova, mvreg, mlogit, and mprobit.

Here we use the jaw fracture dataset described in example 4 of [MV] **manova**. We fit a multivariate regression model including one independent factor variable, `fracture`.

```
. use https://www.stata-press.com/data/r19/jaw
(Table 4.6. Two-way unbalanced data for fractures of the jaw, Rencher (1998))
. mvreg y1 y2 y3 = i.fracture

Equation             Obs  Parms        RMSE    "R-sq"          F        P>F
---------------------------------------------------------------------------
y1                    27      3    10.42366    0.2966   5.060804     0.0147
y2                    27      3    6.325398    0.1341   1.858342     0.1777
y3                    27      3    5.976973    0.1024   1.368879     0.2735

------------------------------------------------------------------------------
             | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
y1           |
    fracture |
Two compo..  |  -8.833333   4.957441    -1.78   0.087    -19.06499    1.398322
One simpl..  |          6   5.394759     1.11   0.277    -5.134235    17.13423
             |
       _cons |         37   3.939775     9.39   0.000      28.8687     45.1313
-------------+----------------------------------------------------------------
y2           |
    fracture |
Two compo..  |  -5.761905   3.008327    -1.92   0.067    -11.97079     .446977
One simpl..  |  -3.053571   3.273705    -0.93   0.360    -9.810166    3.703023
             |
       _cons |   38.42857   2.390776    16.07   0.000     33.49425    43.36289
-------------+----------------------------------------------------------------
y3           |
    fracture |
Two compo..  |   4.261905   2.842618     1.50   0.147     -1.60497    10.12878
One simpl..  |   .9285714   3.093377     0.30   0.767    -5.455846    7.312989
             |
       _cons |   58.57143   2.259083    25.93   0.000     53.90891    63.23395
------------------------------------------------------------------------------
```

pwcompare performs pairwise comparisons of the margins using the coefficients from the first equation by default:

```
. pwcompare fracture, mcompare(bonferroni)
Pairwise comparisons of marginal linear predictions
Margins: asbalanced
```

	Number of comparisons
y1	
fracture	3

	Contrast	Std. err.	Bonferroni [95% conf.	interval]
y1				
fracture				
Two compound fractures vs One compound fracture	-8.833333	4.957441	-21.59201	3.925341
One simple fracture vs One compound fracture	6	5.394759	-7.884173	19.88417
One simple fracture vs Two compound fractures	14.83333	4.75773	2.588644	27.07802

We can use the equation() option to get pwcompare to perform comparisons in the y2 equation:

```
. pwcompare fracture, equation(y2) mcompare(bonferroni)
Pairwise comparisons of marginal linear predictions
Margins: asbalanced
```

	Number of comparisons
y2	
fracture	3

	Contrast	Std. err.	Bonferroni [95% conf.	interval]
y2				
fracture				
Two compound fractures vs One compound fracture	-5.761905	3.008327	-13.50426	1.980449
One simple fracture vs One compound fracture	-3.053571	3.273705	-11.47891	5.371769
One simple fracture vs Two compound fractures	2.708333	2.887136	-4.722119	10.13879

Because we are working with mvreg results, we can use the _eqns pseudofactor to compare the margins between the three dependent variables. The levels of _eqns index the equations: 1 for the first equation, 2 for the second, and 3 for the third.

```
. pwcompare _eqns, mcompare(bonferroni)
Pairwise comparisons of marginal linear predictions
Margins: asbalanced
```

	Number of comparisons
_eqns	3

	Contrast	Std. err.	Bonferroni [95% conf.	interval]
_eqns				
2 vs 1	-.5654762	2.545923	-7.117768	5.986815
3 vs 1	24.24603	2.320677	18.27344	30.21862
3 vs 2	24.81151	2.368188	18.71664	30.90637

For the previous command, the only methods available are mcompare(bonferroni), mcompare(sidak), or mcompare(scheffe). Methods that use the Studentized range are not appropriate for making comparisons across equations.

Unbalanced data

pwcompare treats all factors as balanced when it computes the marginal means. By "balanced", we mean that the number of observations in each combination of factor levels (in each cell mean) is equal. We can alternatively specify the asobserved option when we have unbalanced data to obtain marginal means that are based on the observed cell frequencies from the model fit. For more details on the difference in these two types of marginal means and a discussion of when each may be appropriate, see [R] **margins** and [R] **contrast**.

In addition, when our data are not balanced, some of the multiple-comparison adjustments are no longer appropriate. Student–Newman–Keuls's method, Duncan's method, and Dunnett's method assume equal numbers of observations per group.

Here we use an unbalanced dataset and fit a two-way ANOVA model for cholesterol levels on race and age group. Then we perform pairwise comparisons of the mean cholesterol levels for each race, requesting Šidák's adjustment as well as marginal means that are computed using the observed cell frequencies.

```
. use https://www.stata-press.com/data/r19/cholesterol3
(Artificial cholesterol data, unbalanced)

. anova chol race##agegrp

                         Number of obs =         67    R-squared     =  0.8179
                         Root MSE      =    8.37496    Adj R-squared =  0.7689

                  Source | Partial SS         df         MS        F    Prob>F
             ------------+----------------------------------------------------
                   Model |  16379.993         14   1169.9995     16.68  0.0000
                         |
                    race |   230.7544          2    115.3772      1.64  0.2029
                  agegrp |  13857.988          4   3464.4969     49.39  0.0000
             race#agegrp |  857.81521          8    107.2269      1.53  0.1701
                         |
                Residual |  3647.2774         52    70.13995
             ------------+----------------------------------------------------
                   Total |   20027.27         66   303.44349

. pwcompare race, asobserved mcompare(sidak)

Pairwise comparisons of marginal linear predictions

Margins: asobserved

---------------------------
             |    Number of
             |  comparisons
-------------+-------------
        race |            3
---------------------------

--------------------------------------------------------------------
                |                                 Sidak
                |   Contrast   Std. err.     [95% conf. interval]
----------------+---------------------------------------------------
           race |
White vs Black  |  -7.232433   2.686089     -13.85924   -.6056277
Other vs Black  |  -5.231198   2.651203     -11.77194    1.309541
Other vs White  |   2.001235   2.414964     -3.956682    7.959152
--------------------------------------------------------------------
```

Empty cells

An empty cell is a combination of the levels of factor variables that is not observed in the estimation sample. When we have empty cells in our data, the marginal means involving those empty cells are not estimable as described in [R] **margins**. In addition, all pairwise comparisons involving a marginal mean that is not estimable are themselves not estimable. Here we use a dataset where we do not have any observations for white individuals in the 20–29 age group. We can use the `emptycells(reweight)` option to reweight the nonempty cells so that we can estimate the marginal mean for whites and compute pairwise comparisons involving that marginal mean.

```
. use https://www.stata-press.com/data/r19/cholesterol2
(Artificial cholesterol data, empty cells)

. tabulate race agegrp

           |                       Age group
      Race |     10-19      20-29      30-39      40-59      60-79 |     Total
-----------+-------------------------------------------------------+----------
     Black |         5          5          5          5          5 |        25
     White |         5          0          5          5          5 |        20
     Other |         5          5          5          5          5 |        25
-----------+-------------------------------------------------------+----------
     Total |        15         10         15         15         15 |        70

. anova chol race##agegrp

                         Number of obs =         70    R-squared     =  0.7582
                         Root MSE      =    9.47055    Adj R-squared =  0.7021

                  Source | Partial SS         df         MS        F    Prob>F
             ------------+----------------------------------------------------
                   Model |  15751.611         13  1211.6624     13.51  0.0000
                         |
                    race |  305.49046          2  152.74523      1.70  0.1914
                  agegrp |  14387.856          4   3596.964     40.10  0.0000
             race#agegrp |  795.80757          7   113.6868      1.27  0.2831
                         |
                Residual |  5022.7156         56   89.69135
             ------------+----------------------------------------------------
                   Total |  20774.327         69   301.0772

. pwcompare race, emptycells(reweight)

Pairwise comparisons of marginal linear predictions

Margins:     asbalanced
Empty cells: reweight

--------------------------------------------------------------
                |                              Unadjusted
                |   Contrast   Std. err.     [95% conf. interval]
----------------+---------------------------------------------
           race |
White vs Black  |   2.922769   2.841166     -2.768769    8.614308
Other vs Black  |   -4.12621   2.678677     -9.492244    1.239824
Other vs White  |  -7.048979   2.841166     -12.74052    -1.35744
--------------------------------------------------------------
```

For further details on the `emptycells(reweight)` option, see [R] **margins** and [R] **contrast**.

Stored results

pwcompare stores the following in r():

Scalars

r(df_r)	variance degrees of freedom
r(k_terms)	number of terms in *marginlist*
r(level)	confidence level of confidence intervals
r(balanced)	1 if fully balanced data, 0 otherwise

Macros

r(cmd)	pwcompare
r(cmdline)	command as typed
r(est_cmd)	e(cmd) from original estimation results
r(est_cmdline)	e(cmdline) from original estimation results
r(title)	title in output
r(emptycells)	*empspec* from emptycells()
r(groups#)	group codes for the #th margin in r(b)
r(mcmethod_vs)	*method* from mcompare()
r(mctitle_vs)	title for *method* from mcompare()
r(mcadjustall_vs)	adjustall or empty
r(margin_method)	asbalanced or asobserved
r(vce)	*vcetype* specified in vce() in original estimation command

Matrices

r(b)	margin estimates
r(V)	variance–covariance matrix of the margin estimates
r(error)	margin estimability codes; 0 means estimable, 8 means not estimable
r(table)	matrix containing the margins with their standard errors, test statistics, *p*-values, and confidence intervals
r(M)	matrix that produces the margins from the model coefficients
r(b_vs)	margin difference estimates
r(V_vs)	variance–covariance matrix of the margin difference estimates
r(error_vs)	margin difference estimability codes; 0 means estimable, 8 means not estimable
r(table_vs)	matrix containing the margin differences with their standard errors, test statistics, *p*-values, and confidence intervals
r(L)	matrix that produces the margin differences from the model coefficients
r(k_groups)	number of significance groups for each term

pwcompare with the post option also stores the following in e():

Scalars

e(df_r)	variance degrees of freedom
e(k_terms)	number of terms in *marginlist*
e(balanced)	1 if fully balanced data, 0 otherwise

Macros

e(cmd)	pwcompare
e(cmdline)	command as typed
e(properties)	b V
e(est_cmd)	e(cmd) from original estimation results
e(est_cmdline)	e(cmdline) from original estimation results
e(title)	title in output
e(emptycells)	*empspec* from emptycells()
e(margin_method)	asbalanced or asobserved
e(asbalanced)	factor variables fvset as asbalanced
e(asobserved)	factor variables fvset as asobserved
e(vce)	*vcetype* specified in vce() in original estimation command

Matrices

e(b)	margin estimates
e(V)	variance–covariance matrix of the margin estimates
e(error)	margin estimability codes; 0 means estimable, 8 means not estimable
e(M)	matrix that produces the margins from the model coefficients
e(b_vs)	margin difference estimates
e(V_vs)	variance–covariance matrix of the margin difference estimates
e(error_vs)	margin difference estimability codes; 0 means estimable, 8 means not estimable
e(L)	matrix that produces the margin differences from the model coefficients

Methods and formulas

Methods and formulas are presented under the following headings:

Notation
Unadjusted comparisons
Bonferroni's method
Šidák's method
Scheffé's method
Tukey's method
Student–Newman–Keuls's method
Duncan's method
Dunnett's method

Notation

pwcompare performs comparisons of margins; see *Methods and formulas* in [R] **contrast**.

If there are k margins for a given factor term, then there are

$$m = \binom{k}{2} = \frac{k(k-1)}{2}$$

unique pairwise comparisons. Let the ith pairwise comparison be denoted by

$$\widehat{\delta}_i = l_i' \mathbf{b}$$

where $\mathbf{b}$ is a column vector of coefficients from the fitted model and l_i is a column vector that forms the corresponding linear combination. If $\widehat{\mathbf{V}}$ denotes the estimated variance matrix for $\mathbf{b}$, then the standard error for $\widehat{\delta}_i$ is given by

$$\widehat{\mathrm{se}}(\widehat{\delta}_i) = \sqrt{l_i' \widehat{\mathbf{V}} l_i}$$

The corresponding test statistic is then

$$t_i = \frac{\widehat{\delta}_i}{\widehat{\mathrm{se}}(\widehat{\delta}_i)}$$

and the limits for a $100(1-\alpha)\%$ confidence interval for the expected value of $\widehat{\delta}_i$ are

$$\widehat{\delta}_i \pm c_i(\alpha)\, \widehat{\mathrm{se}}(\widehat{\delta}_i)$$

where $c_i(\alpha)$ is the critical value corresponding to the chosen multiple-comparison method.

Unadjusted comparisons

pwcompare computes unadjusted p-values and confidence intervals by default. pwcompare uses the t distribution with $\nu =$ e(df_r) degrees of freedom when e(df_r) is posted by the estimation command. The unadjusted two-sided p-value is

$${}_u p_i = 2\,\mathrm{Pr}(t_\nu > |t_i|)$$

and the unadjusted critical value ${}_u c_i(\alpha)$ satisfies the following probability statement:

$$\alpha = 2\,\mathrm{Pr}\left\{t_\nu > {}_u c_i(\alpha)\right\}$$

pwcompare uses the standard normal distribution when e(df_r) is not posted.

Bonferroni's method

For `mcompare(bonferroni)`, the adjusted p-value is

$$_bp_i = \min(1, m\,{_up_i})$$

and the adjusted critical value is

$$_bc_i(\alpha) = {_uc_i}(\alpha/m)$$

Šidák's method

For `mcompare(sidak)`, the adjusted p-value is

$$_{\rm si}p_i = 1 - (1 - {_up_i})^m$$

and the adjusted critical value is

$$_{\rm si}c_i(\alpha) = {_uc_i}\left\{1 - (1 - \alpha)^{1/m}\right\}$$

Scheffé's method

For `mcompare(scheffe)`, the adjusted p-value is

$$_{\rm sc}p_i = \Pr\left(F_{d,\nu} > t_i^2/d\right)$$

where $F_{d,\nu}$ is distributed as an F with d numerator and ν denominator degrees of freedom and d is the rank of the VCE for the term. The adjusted critical value satisfies the following probability statement:

$$\alpha = \Pr\left[F_{d,\nu} > \{{_{\rm sc}c_i}(\alpha)\}^2/d\right]$$

`pwcompare` uses the χ^2 distribution when `e(df_r)` is not posted.

Tukey's method

For `mcompare(tukey)`, the adjusted p-value is

$$_tp_i = \Pr\left(q_{k,\nu} > |t_i|\sqrt{2}\right)$$

where $q_{k,\nu}$ is distributed as the Studentized range statistic for k means and ν residual degrees of freedom (Miller 1981). The adjusted critical value satisfies the following probability statement:

$$\alpha = \Pr\left\{q_{k,\nu} > {_tc_i}(\alpha)\sqrt{2}\right\}$$

Student–Newman–Keuls's method

For `mcompare(snk)`, suppose t_i is comparing two margins that have r other margins between them. Then the adjusted p-value is

$$_{\text{snk}}p_i = \Pr\left(q_{r+2,\nu} > |t_i|\sqrt{2}\right)$$

where r ranges from 0 to $k-2$. The adjusted critical value $_{\text{snk}}c_i(\alpha)$ satisfies the following probability statement:

$$\alpha = \Pr\left\{q_{r+2,\nu} > {_{\text{snk}}c_i(\alpha)}\sqrt{2}\right\}$$

Duncan's method

For `mcompare(duncan)`, the adjusted p-value is

$$_{\text{dunc}}p_i = 1 - (1 - {_{\text{snk}}p_i})^{1/(r+1)}$$

and the adjusted critical value is

$$_{\text{dunc}}c_i(\alpha) = {_{\text{snk}}c_i}\{1 - (1-\alpha)^{r+1}\}$$

Dunnett's method

For `mcompare(dunnett)`, the margins are compared with a reference category, resulting in only $k-1$ pairwise comparisons. The adjusted p-value is

$$_{\text{dunn}}p_i = \Pr(d_{k-1,\nu} > |t_i|)$$

where $d_{k-1,\nu}$ is distributed as the many-one t statistic (Miller 1981, 76). The adjusted critical value $_{\text{dunn}}c_i(\alpha)$ satisfies the following probability statement:

$$\alpha = \Pr\left\{d_{k-1,\nu} > {_{\text{dunn}}c_i(\alpha)}\right\}$$

The multiple-comparison methods for `mcompare(tukey)`, `mcompare(snk)`, `mcompare(duncan)`, and `mcompare(dunnett)` assume the normal distribution with equal variance; thus, these methods are allowed only with results from `anova`, `regress`, `manova`, and `mvreg`. `mcompare(snk)`, `mcompare(duncan)`, and `mcompare(dunnett)` assume equal sample size for each marginal mean. These options will cause `pwcompare` to report a footnote if unbalanced factors are detected.

References

Baldwin, S. 2019. *Psychological Statistics and Psychometrics Using Stata*. College Station, TX: Stata Press.

Dinno, A. 2015. Nonparametric pairwise multiple comparisons in independent groups using Dunn's test. *Stata Journal* 15: 292–300.

Dunnett, C. W. 1955. A multiple comparison for comparing several treatments with a control. *Journal of the American Statistical Association* 50: 1096–1121. https://doi.org/10.2307/2281208.

———. 1964. New tables for multiple comparisons with a control. *Biometrics* 20: 482–491. https://doi.org/10.2307/2528490.

Kramer, C. Y. 1956. Extension of multiple range tests to group means with unequal numbers of replications. *Biometrics* 12: 307–310. https://doi.org/10.2307/3001469.

Miller, R. G., Jr. 1981. *Simultaneous Statistical Inference*. 2nd ed. New York: Springer. https://doi.org/10.1007/978-1-4613-8122-8.

Milliken, G. A., and D. E. Johnson. 2009. *Designed Experiments*. Vol. 1 of *Analysis of Messy Data*, 2nd ed. Boca Raton, FL: CRC Press.

Mitchell, M. N. 2015. *Stata for the Behavioral Sciences*. College Station, TX: Stata Press.

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

Searle, S. R. 1997. *Linear Models for Unbalanced Data*. New York: Wiley.

Taffé, P., M. Peng, V. Stagg, and T. Williamson. 2017. biasplot: A package to effective plots to assess bias and precision in method comparison studies. *Stata Journal* 17: 208–221.

———. 2023. Extended biasplot command to assess bias, precision, and agreement in method comparison studies. *Stata Journal* 23: 97–118.

Tukey, J. W. 1953. The problem of multiple comparisons. Unpublished manuscript, Princeton University.

Winer, B. J., D. R. Brown, and K. M. Michels. 1991. *Statistical Principles in Experimental Design*. 3rd ed. New York: McGraw–Hill.

Also see

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

[R] **contrast** — Contrasts and linear hypothesis tests after estimation

[R] **lincom** — Linear combinations of parameters

[R] **margins** — Marginal means, predictive margins, and marginal effects

[R] **margins, pwcompare** — Pairwise comparisons of margins

[R] **pwmean** — Pairwise comparisons of means

[R] **test** — Test linear hypotheses after estimation

[U] 20 Estimation and postestimation commands

For suggested citations, see the FAQ on citing Stata documentation.