Title

stata.com

pwmean — Pairwise comparisons of means

Syntax

```
pwmean varname, over(varlist) [ options ]
```

options	Description
Main	
* `over(`*varlist*`)`	compare means across each combination of the levels in *varlist*
`mcompare(`*method*`)`	adjust for multiple comparisons; default is `mcompare(noadjust)`
Reporting	
`level(#)`	confidence level; default is `level(95)`
`cieffects`	display a table of mean differences and confidence intervals; the default
`pveffects`	display a table of mean differences and p-values
`effects`	display a table of mean differences with p-values and confidence intervals
`cimeans`	display a table of means and confidence intervals
`groups`	display a table of means with codes that group them with other means that are not significantly different
`sort`	sort results tables by displayed mean or difference
display_options	control column formats, line width, and factor-variable labeling

* `over(`*varlist*`)` is required.

method	Description
`noadjust`	do not adjust for multiple comparisons; the default
`bonferroni`	Bonferroni's method
`sidak`	Šidák's method
`scheffe`	Scheffé's method
`tukey`	Tukey's method
`snk`	Student–Newman–Keuls' method
`duncan`	Duncan's method
`dunnett`	Dunnett's method

Menu

Statistics > Summaries, tables, and tests > Summary and descriptive statistics > Pairwise comparisons of means

Description

pwmean performs pairwise comparisons of means. It computes all pairwise differences of the means of *varname* over the combination of the levels of the variables in *varlist*. The tests and confidence intervals for the pairwise comparisons assume equal variances across groups. pwmean also allows for adjusting the confidence intervals and p-values to account for multiple comparisons using Bonferroni's method, Scheffé's method, Tukey's method, Dunnett's method, and others.

See [R] **pwcompare** for performing pairwise comparisons of means, estimated marginal means, and other types of marginal linear predictions after anova, regress, and most other estimation commands.

See [R] **margins, pwcompare** for performing pairwise comparisons of marginal probabilities and other linear and nonlinear predictions after estimation commands.

Options

Main

over(*varlist*) is required and specifies that means are computed for each combination of the levels of the variables in *varlist*.

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

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 $k - 1$ where k is the number of means being compared.

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 means participating in the range for a given comparison instead of the full number of means.

`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.

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] 20.7 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 of means with their standard errors and confidence intervals be reported. This is the default.

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

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

`cimeans` specifies that a table of the means with their standard errors and confidence intervals be reported.

`groups` specifies that a table of the means with their standard errors and group codes be reported. Means 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 by the mean or difference that is displayed in the table.

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

`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 means, standard errors, and confidence limits in the table of pairwise comparison of means.

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

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

`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. To change the default, use `set lstretch off`. `nolstretch` is not shown in the dialog box.

Remarks and examples

stata.com

`pwmean` performs pairwise comparisons (differences) of means, assuming a common variance among groups. It can easily adjust the p-values and confidence intervals for the differences to account for the elevated type I error rate due to multiple comparisons. Adjustments for multiple comparisons can be made using Bonferroni's method, Scheffé's method, Tukey's method, Dunnett's method, and others.

Remarks are presented under the following headings:

Group means

Suppose we have data on the wheat yield of fields that were each randomly assigned an application of one of five types of fertilizers. Let's first look at the mean yield for each type of fertilizer.

```
. use http://www.stata-press.com/data/r13/yield
(Artificial wheat yield dataset)
. pwmean yield, over(fertilizer) cimeans
Pairwise comparisons of means with equal variances
over         : fertilizer
```

			Unadjusted	
yield	Mean	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

Pairwise differences of means

We can compute all pairwise differences in mean wheat yields for the types of fertilizers.

```
. pwmean yield, over(fertilizer) effects
Pairwise comparisons of means with equal variances
over         : fertilizer
```

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

The contrast in the row labeled `(10-08-22 vs 10-10-10)` is the difference in the mean wheat yield for fertilizer 10-08-22 and fertilizer 10-10-10. At a 5% significance level, we conclude that there is a difference in the means for these two fertilizers. Likewise, the rows labeled `(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)` show differences in these pairs of means. In all, we find that 5 of the 10 mean differences are significantly different from zero at a 5% significance level.

We can specify the `sort` option to order the differences from smallest to largest in the table.

```
. pwmean yield, over(fertilizer) effects sort
Pairwise comparisons of means with equal variances
over         : fertilizer
```

yield	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

Ordering the pairwise differences is particularly convenient when we are comparing means for a large number of groups.

Group output

We can use the `group` option to see the mean of each group and a visual representation of the tests for differences.

```
. pwmean yield, over(fertilizer) group sort
Pairwise comparisons of means with equal variances
over         : fertilizer
```

yield	Mean	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: Means sharing a letter in the group label
      are not significantly different at the 5%
      level.
```

Fertilizers 29-03-04, 10-10-10, and 16-04-08 are all in group A. This means that at our 5% level of significance, we have insufficient information to distinguish their means. Likewise, fertilizers 16-04-08 and 10-08-22 are in group B and cannot be distinguished at the 5% level. The same is true for fertilizers 10-08-22 and 18-24-06 in group C.

Fertilizer 29-03-04 and fertilizer 10-08-22 have no letters in common, indicating that the mean yields of these two groups are significantly different at the 5% level. We can conclude that any other fertilizers without a letter in common have significantly different means as well.

Adjusting for multiple comparisons

The statistics in the examples above take no account that we are performing 10 comparisons. With our 5% significance level and assuming the comparisons are independent, we expect 1 in 20 tests of comparisons to be significant, even if all the population means are truly the same. If we are performing many comparisons, then we should account for the fact that some tests will be found significant by chance alone. More formally, the test for each pairwise comparison is made without adjusting for the elevated type I experimentwise error rate that is introduced when performing multiple tests. We can use the `mcompare()` option to adjust the confidence intervals and p-values for multiple comparisons.

Tukey's method

Of the available adjustments for multiple comparisons, Tukey's honestly significant difference, Student–Newman–Keuls' method, and Duncan's method are most often used when performing all pairwise comparisons of means. Of these, Tukey's method is the most conservative and Duncan's method is the least conservative. For further discussion of each of the multiple-comparison adjustments, see [R] **pwcompare**.

Here we use Tukey's adjustment to compute p-values and confidence intervals for the pairwise differences.

```
. pwmean yield, over(fertilizer) effects sort mcompare(tukey)
Pairwise comparisons of means with equal variances
over         : fertilizer
```

	Number of Comparisons
fertilizer	10

yield	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

When using a 5% significance level, Tukey's adjustment indicates that four pairs of means are different. With the adjustment, we no longer conclude that the difference in the mean yields for fertilizers 10-08-22 and 10-10-10 is significantly different from zero.

Dunnett's method

Now let's suppose that fertilizer 10-10-10 actually represents fields on which no fertilizer was applied. In this case, we can use Dunnett's method for comparing each of the fertilizers to the control.

```
. pwmean yield, over(fertilizer) effects mcompare(dunnett)
Pairwise comparisons of means with equal variances
over          : fertilizer
```

	Number of Comparisons
fertilizer	4

yield	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

Using Dunnett's adjustment, we conclude that only fertilizer 4 (18-24-06) produces a mean yield that is significantly different from the mean yield of the field with no fertilizer applied.

By default, `pwmean` treats the lowest level of the group variable as the control. If, for instance, fertilizer 3 (16-04-08) was our control group, we could type

```
. pwmean yield, over(b3.fertilizer) effects mcompare(dunnett)
```

using the `b3.` factor-variable operator to specify this level as the reference level.

Multiple over() variables

When we specify more than one variable in the `over()` option, pairwise comparisons are performed for the means defined by each combination of levels of these variables.

```
. pwmean yield, over(fertilizer irrigation) group
Pairwise comparisons of means with equal variances
over         : fertilizer irrigation
```

yield	Mean	Std. Err.	Unadjusted Groups
fertilizer#irrigation			
10-10-10#0	36.91257	1.116571	A
10-10-10#1	45.81229	1.116571	B
10-08-22#0	38.79482	1.116571	A C
10-08-22#1	51.17547	1.116571	E
16-04-08#0	36.34383	1.116571	A
16-04-08#1	47.36229	1.116571	B
18-24-06#0	41.81757	1.116571	CD
18-24-06#1	50.7529	1.116571	E
29-03-04#0	35.69507	1.116571	A
29-03-04#1	44.55313	1.116571	B D

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

Here the row labeled `10-10-10#0` is the mean for the fields treated with fertilizer 10-10-10 and without irrigation. This mean is significantly different from the mean of all fertilizer/irrigation pairings that do not have an A in the "Unadjusted Groups" column. These include all pairings where the fields were irrigated as well as the fields treated with fertilizer 18-24-06 but without irrigation.

Equal variance assumption

`pwmean` performs multiple comparisons assuming that there is a common variance for all groups. In the case of two groups, this is equivalent to performing the familiar two-sample t test when equal variances are assumed.

```
. ttest yield, by(irrigation)
Two-sample t test with equal variances
```

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
0	100	37.91277	.5300607	5.300607	36.86102	38.96453
1	100	47.93122	.5630353	5.630353	46.81403	49.0484
combined	200	42.92199	.5242462	7.413961	41.8882	43.95579
diff		-10.01844	.7732872		-11.54338	-8.493509

```
    diff = mean(0) - mean(1)                                      t = -12.9557
Ho: diff = 0                                     degrees of freedom =      198

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 1.0000

. pwmean yield, over(irrigation) effects
Pairwise comparisons of means with equal variances
over         : irrigation
```

yield	Contrast	Std. Err.	Unadjusted t	Unadjusted P>|t|	Unadjusted [95% Conf.	Interval]
irrigation						
1 vs 0	10.01844	.7732872	12.96	0.000	8.493509	11.54338

The signs for the difference, the test statistic, and the confidence intervals are reversed because the difference is taken in the opposite direction. The p-value from `pwmean` is equivalent to the one for the two-sided test in the `ttest` output.

`pwmean` extends the capabilities of `ttest` to allow for simultaneously comparing all pairs of means and to allow for using one common variance estimate for all the tests instead of computing a separate pooled variance for each pair of means when using multiple `ttest` commands. In addition, `pwmean` allows adjustments for multiple comparisons, many of which rely on an assumption of equal variances among groups.

Stored results

pwmean stores the following in e():

Scalars

e(df_r)	variance degrees of freedom
e(balanced)	1 if fully balanced data; 0 otherwise

Macros

e(cmd)	pwmean
e(cmdline)	command as typed
e(title)	title in output
e(depvar)	name of variable from which the means are computed
e(over)	*varlist* from over()
e(properties)	b V

Matrices

e(b)	mean estimates
e(V)	variance–covariance matrix of the mean estimates
e(error)	mean estimability codes; 0 means estimable, 8 means not estimable
e(b_vs)	mean difference estimates
e(V_vs)	variance–covariance matrix of the mean difference estimates
e(error_vs)	mean difference estimability codes; 0 means estimable, 8 means not estimable
e(k_groups)	number of significance groups for each term

Methods and formulas

pwmean is a convenience command that uses pwcompare after fitting a fully factorial linear model. See *Methods and formulas* described in [R] **pwcompare**.

Reference

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

Also see

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

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

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

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

[R] **pwcompare** — Pairwise comparisons

[R] **ttest** — t tests (mean-comparison tests)

[U] **20 Estimation and postestimation commands**