How can I do post-hoc pairwise comparisons using Stata?

Authored by

stats writer

July 1, 2024

Post-hoc pairwise comparisons refer to a statistical method used to compare individual groups or categories within a larger dataset after conducting a statistical test. In order to perform post-hoc pairwise comparisons using Stata, one must first run the appropriate statistical test, such as ANOVA or t-test, to determine if there is a significant difference between groups. Once this is established, Stata offers several commands, such as "pwcompare" or "postestimation", to conduct post-hoc pairwise comparisons and calculate the p-values for each comparison. These comparisons can provide further insights and understanding of the differences between groups in a dataset.

FAQ:

How can I do post-hoc pairwise comparisons using Stata?

Post-hoc pairwise comparisons are commonly performed after significant effects when there are three or more levels of a factor. Stata has three built-in pairwise methods (sidak, bonferroni and scheffe) in the oneway command. Although these options are easy to use, many researchers consider the methods to be too conservative for pairwise comparisons, especially when the are many levels. The Sidak method is the least conservative of the three followed, in order, by Bonferroni and Scheffe.

We will demonstrate the pairwise options in oneway on

a
dataset looking at write by group which is a four-level predictor.

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

tabstat write, by(group) stat(n mean sd)

Summary for variables: write

by categories of: group

group | N mean sd

---------+---------------------------
1 | 24 46.45833 8.272422
2 | 11 58 7.899367
3 | 20 48.2 9.322299
4 | 145 54.05517 9.172558
---------+---------------------------
Total | 200 52.775 9.478586
----

oneway write group, sidak bonferroni scheffe

Analysis of Variance

Source SS df MS F Prob > F

Between groups 1914.15805 3 638.052682 7.83 0.0001

Within groups 15964.717 196 81.4526375

Total 17878.875 199 89.843593

Bartlett's test for equal variances: chi2(3) = 0.7555 Prob>chi2 = 0.860

Comparison of writing score by group

(Sidak)

Row Mean-|

Col Mean | 1 2 3

---+---

2 | 11.5417

| 0.003

|

3 | 1.74167 -9.8

| 0.988 0.025

|

4 | 7.59684 -3.94483 5.85517

| 0.001 0.658 0.042

Comparison of writing score by group (Bonferroni)

Row Mean-|
Col Mean | 1 2 3
---------+---------------------------------
2 | 11.5417
| 0.003
|
3 | 1.74167 -9.8
| 1.000 0.026
|
4 | 7.59684 -3.94483 5.85517
| 0.001 0.983 0.043

Comparison of writing score by group (Scheffe)

Row Mean-|
Col Mean | 1 2 3
---------+---------------------------------
2 | 11.5417
| 0.007

|
3 | 1.74167 -9.8
| 0.939 0.042
|
4 | 7.59684 -3.94483 5.85517
| 0.003 0.583 0.063

Comparisons 1 versus 2, 1 versus 4, and 2 versus 3 were significant at the 0.05 level or better for all methods while 3vs4 was significant for Sidak and Bonferroni but not Scheffe.

Many researchers prefer pairwise comparisons based upon the Studentized Range distribution.
The IDRE Statistical Consulting Group has developed three programs for the Tukey HSD, the Tukey-Kramer and
the Fisher-Hayter methods. To obtain these programs use the search command (search tukeyhsd, search tkcomp
or search fhcomp). Please note that these programs need the qsturng and
sturng by John R. Gleason which can be found in STB-47/sg101.

The three methods will yield the same test statistic when the cell sizes are equal but will differ when cell sizes are unequal. Computationally, the Tukey-Kramer and the Fisher-Hayter are the same but they use different critical values of the Studentized Range distribution.

The Tukey-Kramer or the Fisher-Hayter are usually preferred when the cell sizes are unequal.

Tukey-Kramer uses degrees of freedom of k and dferror where k is the number of levels and dferror is the degrees of freedom associated with the MSerror in the anova, to obtain the critical value of the Studentized Range statistic.

Fisher-Hayter, on the other hand, uses degrees of freedom k-1 and dferror.

anova write group

Number of obs = 200 R-squared = 0.1071
Root MSE = 9.02511 Adj R-squared = 0.0934

Source | Partial SS df MS F Prob > F

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

Model | 1914.15805 3 638.052682 7.83 0.0001

|

group | 1914.15805 3 638.052682 7.83 0.0001

|

Residual | 15964.717 196 81.4526375

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

Total | 17878.875 199 89.843593

tukeyhsd group

Tukey HSD pairwise comparisons for variable group

studentized range critical value(.05, 4, 196) = 3.6647117

uses harmonic mean sample size = 21.111

mean

grp vs grp group means dif HSB-test

--

1 vs 2 46.4583 58.0000 11.5417 5.8759*

1 vs 3 46.4583 48.2000 1.7417 0.8867

1 vs 4 46.4583 54.0552 7.5968 3.8676*

2 vs 3 58.0000 48.2000 9.8000 4.9892*

2 vs 4 58.0000 54.0552 3.9448 2.0083

3 vs 4 48.2000 54.0552 5.8552 2.9809

tkcomp group

Tukey-Krammer pairwise comparisons for variable group
studentized range critical value(.05, 4, 196) = 3.6647117

mean
grp vs grp group means dif TK-test
----
1 vs 2 46.4583 58.0000 11.5417 4.9671*
1 vs 3 46.4583 48.2000 1.7417 0.9014
1 vs 4 46.4583 54.0552 7.5968 5.4018*
2 vs 3 58.0000 48.2000 9.8000 4.0909*
2 vs 4 58.0000 54.0552 3.9448 1.9766
3 vs 4 48.2000 54.0552 5.8552 3.8464*

fhcomp group

Fisher-Hayter pairwise comparisons for variable group
studentized range critical value(.05, 3, 196) = 3.3399493

mean
grp vs grp group means dif FH-test
----
1 vs 2 46.4583 58.0000 11.5417 4.9671*

1 vs 3 46.4583 48.2000 1.7417 0.9014
1 vs 4 46.4583 54.0552 7.5968 5.4018*
2 vs 3 58.0000 48.2000 9.8000 4.0909*
2 vs 4 58.0000 54.0552 3.9448 1.9766
3 vs 4 48.2000 54.0552 5.8552 3.8464*

Groups 1 versus 2, 1 versus 4 and 2 versus 3 were significant using Tukey's HSD while both Tukey-Kramer and Fisher-Hayter also find 3vs4 significant at the 0.05 level.

The three IDRE Statistical Consulting Group programs will also work with factorial designs as shown below.

anova write female group group*female

Number of obs = 200 R-squared = 0.1706
Root MSE = 8.78819 Adj R-squared = 0.1404

Source | Partial SS df MS F Prob > F
----------+----------
Model | 3050.29061 7 435.755802 5.64 0.0000
|
female | 249.988577 1 249.988577 3.24 0.0736
group | 1674.93766 3 558.312554 7.23 0.0001

group*female | 51.0895327 3 17.0298442 0.22 0.8821
|
Residual | 14828.5844 192 77.2322104
------------+--
Total | 17878.875 199 89.843593

tukeyhsd group

Tukey HSD pairwise comparisons for variable group
studentized range critical value(.05, 4, 192) = 3.665369
uses harmonic mean sample size = 21.111

mean
grp vs grp group means dif HSB-test
----
1 vs 2 46.4583 58.0000 11.5417 6.0343*
1 vs 3 46.4583 48.2000 1.7417 0.9106
1 vs 4 46.4583 54.0552 7.5968 3.9718*
2 vs 3 58.0000 48.2000 9.8000 5.1237*
2 vs 4 58.0000 54.0552 3.9448 2.0625
3 vs 4 48.2000 54.0552 5.8552 3.0612

tkcomp group

Tukey-Krammer pairwise comparisons for variable

group
studentized range critical value(.05, 4, 192) = 3.665369

mean
grp vs grp group means dif TK-test
---
1 vs 2 46.4583 58.0000 11.5417 5.1010*
1 vs 3 46.4583 48.2000 1.7417 0.9257
1 vs 4 46.4583 54.0552 7.5968 5.5475*
2 vs 3 58.0000 48.2000 9.8000 4.2012*
2 vs 4 58.0000 54.0552 3.9448 2.0298
3 vs 4 48.2000 54.0552 5.8552 3.9501*

fhcomp group

Fisher-Hayter pairwise comparisons for variable group
studentized range critical value(.05, 3, 192) = 3.3404824

mean
grp vs grp group means dif FH-test
---
1 vs 2 46.4583 58.0000 11.5417 5.1010*
1 vs 3 46.4583 48.2000 1.7417 0.9257
1 vs 4 46.4583 54.0552 7.5968 5.5475*
2 vs 3 58.0000 48.2000 9.8000 4.2012*

2 vs 4 58.0000 54.0552 3.9448 2.0298

3 vs 4 48.2000 54.0552 5.8552 3.9501*

Reference

Kirk, Roger E. (1998) Experimental Design: Procedures for the Behavioral Sciences,
Third Edition. Monterey, California: Brooks/Cole Publishing. ISBN 0-534-25092-0.