Follow-Up Comparisons: Sections 39-40

How to Do Pairwise Comparisons of Multiple Means or Proportions

How to Do Pairwise Comparisons of Multiple Medians

Section 39: How to Do Pairwise Comparisons of Multiple Means or Proportions

(Uses data files: Absences.txt)

Suppose you have samples from more than two populations, and have determined (such as by using ANOVA) that at least one population mean is different. The ANOVA does not tell you which one(s) is/are significantly different.

Or perhaps you have proportions based on data from several populations, and you believe that at least one of them is different. Again, you want some justification for deciding which one(s) is/are different from the others.

In these situations, it is helpful to be able to do pairwise comparisons. However, since you want to do multiple comparisons on the same data set, a correction must be made to the p-values to control the overall probability of making a Type 1 error (rejecting a true null hypothesis). There are different correction procedures available, but the most common one is the Bonferroni correction. R software allows you to make this correction automatically.

Both of the data sets used in the following examples are fictitious. The first example uses one called Absences.txt. The second example simply has data entered as part of the R code. While they both deal with the topic of measuring absences from school, they are not related.

EXAMPLE A This example runs an ANOVA to test whether or not the mean number of absences is essentially the same for the months November through February. The ANOVA output leads to a rejection of the hypothesis that the means are the same. Hence you would run a subsequent pairwise t.test to try to identify which mean(s) is/are different. It includes the Bonferroni correction.

The data set contains counts of the numbers of students absent at least three days in the month, for ten classrooms. Different ten-classroom samples are randomly selected each month. The classes are all the same size, so there is no need to weight the counts. The actual data set is as follows.

	Month	Absent		Month	Absent		Month	Absent		Month	Absent
1	Nov	3	11	Dec	6	21	Jan	6	31	Feb	1
2	Nov	1	12	Dec	4	22	Jan	0	32	Feb	0
3	Nov	2	13	Dec	1	23	Jan	4	33	Feb	4
4	Nov	4	14	Dec	7	24	Jan	3	34	Feb	3
5	Nov	3	15	Dec	5	25	Jan	6	35	Feb	2
6	Nov	1	16	Dec	3	26	Jan	4	36	Feb	4
7	Nov	5	17	Dec	6	27	Jan	2	37	Feb	2
8	Nov	2	18	Dec	3	28	Jan	7	38	Feb	3
9	Nov	1	19	Dec	8	29	Jan	6	39	Feb	6
10	Nov	0	20	Dec	8	30	Jan	5	40	Feb	3

As usual, first read in the data set and attach it.

```
> Data = read.table ("E:/Data Files/Absences.txt", header = TRUE)
> attach (Data)
```

It was stated above that the preliminary ANOVA indicates at least one of the monthly means is significantly different. That test is repeated here for reference.

First you create a linear model with the response variable Absent as a function of the variable Month. Run an ANOVA on the resulting model. Assume your level of significance is alpha (α) = .05.

```
> Model = lm (Absent ~ Month)
> anova (Model)
```

The results of the ANOVA are on the left as follows; comments are on the right.

```
Analysis of Variance Table

Response: Absent
           Df   Sum Sq   Mean Sq    F value    Pr(>F)
Month       3    53.4    17.8000    4.6368   0.007667 **  ← p-value is less than α; reject the
Residuals  36   138.2     3.8389                             equal mean hypothesis.
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In this example, the ANOVA leads you to conclude that the mean number of absences is different in at least one month. It does not tell you which one(s) is/are different. Now you probably want to compare the means pairwise to figure that out; this requires a pairwise t-test with the Bonferroni correction. Assume you are still using level of significance α = .05, with a built-in Bonferroni correction.

A few words about the subcommands are needed here.

(1) Note that the subcommand "p.adjust.method" as shown below tells R to make the Bonferroni correction.
(2) Note the subcommand that says: "paired=FALSE"; it does NOT refer to the fact that the means are being compared pairwise. Instead, it is used to specify whether the data does or does not consist of repeated measurements; here it does not. If the same ten classrooms had been used every month, then you would say: "paired=TRUE."

```
> pairwise.t.test (Absent, Month, p.adjust.method = "bonf", paired = FALSE)
```

The resulting output is the following table.

```
Pairwise comparisons using t tests with pooled SD

data:  Absent and Month

      Dec     Feb     Jan
Feb   0.076   -       -
Jan   1.000   0.573   -
Nov   0.013   1.000   0.131

P value adjustment method: bonferroni
```

Only one p-value in the table is less than α. That p-value is 0.013, which came from comparing the means for November and December. Hence these two months are the only pair for which the difference in the mean number of absences is statistically significant.

EXAMPLE B This example compares proportions (absentee rates) among school students, pairwise by month, from November through February. A sample of 150 students in the first grade of a particular school district remains the same each month because no transfers occur. You want to see which months, if any, have significantly different absentee rates (counted as a student being absent more than three days a month). Suppose the counts of students absent more than three days per month are: 13 students in November, 9 in December, 35 in January, and 27 in February.

First enter the data, in order by month: Nov, Dec, Jan, Feb.

```
> Absent = c (13, 9, 35, 27)
> Students = c (150, 150, 150, 150)
```

Calculate the sample proportions.

```
> Absent.Rate = Absent/Students
> Absent.Rate
```

You will get the following proportions; the month labels have been added here for clarity.

```
  November     December     January      February
  0.08666667   0.06000000   0.23333333   0.18000000
```

These rates, ranging from 6% to 23.3%, certainly do not appear to be all the same. However, if they are viewed as a sample of a larger population (such as all elementary classrooms in the district), they need to be tested to determine which rates are significantly different over the larger population and which are not. Again, assume your criterion for significance is alpha (α) = .05, with a built-in Bonferroni correction.

The "pairwise.prop.test" command will do this. You have to give it:

-- the counts of the item you are interested in (Absent)
-- the total sample size each time (Students)
-- the instruction to use the Bonferroni correction (p.adjust="bonf").

```
> pairwise.prop.test (Absent, Students, p.adjust = "bonf")
```

Here is the output; explanatory comments are included on the right. The number labels are in order by month: 1 = November, 2 = December, 3 = January, 4 = February. The decimal values in the grid are the p-values for the corresponding comparisons of proportions.

As you can see from the output and the explanatory comments, the proportions of first grade students absent more than three days per month in the period November through February are significantly different:

- between November and January
- between December and January
- between December and February.

The other three pairs show no statistically significant different in proportions.

Pairwise comparisons using Pairwise comparison of proportions

data: Absent out of Students

	1	2	3	
2	1.00000	-	-	← p-value comparing Nov and Dec is 1; no difference
3	0.00565	0.00027	-	← p-value comparing Nov and Jan is .00565; p-value comparing Dec and Jan is .00027. Both p-values are less than α so differences in proportions (Absent.Rate) are both significant.
4	0.16349	0.01515	1.00000	← ---The only p-value that is less than α is .01515; indicates the difference in proportions (Absent.Rate) between Dec and Feb is significant.

P value adjustment method: Bonferroni

Section 40: How to Do Pairwise Comparisons of Multiple Medians

(Uses data files: Anxiety.txt, AnxietyRepeat.txt)

Suppose you have samples from more than two populations, and have determined (by using the Kruskal-Wallis Test or Friedman's ANOVA) that <u>at least</u> one population median is different. Recall that these tests do not tell you which one(s) is/are different, only that at least one is.

In this situation, you need to be able to do pairwise comparisons of medians. However, since you want to do multiple comparisons on the same data set, a correction must be made to the p-values to control the overall probability of making a Type 1 error (rejecting a true null hypothesis). There are different correction procedures available, but the most common one is the Bonferroni correction. R software allows you to make this correction automatically.

Both of the data sets used in the following examples are fictitious. The first example uses one called Anxiety.txt. The second example uses one called AnxietyRepeat.txt.

<u>EXAMPLE A</u> Section 33: "How to Run a Kruskal-Wallis Test" shows the process of using a Kruskal-Wallis test to determine whether or not the median anxiety score of students varies by type of institution. The output there leads to not rejecting the null hypothesis, which suggests that the median score is the same for students from all types of institutions in the study.

However, for purposes of this example, suppose that the test had shown that at least one median was different. You would probably then run a pairwise.wilcox.test to see which one (or ones) differ. Assume the level of significance is alpha (α) = .05, with a built-in Bonferroni correction.

The general syntax for this type of situation is:

```
> pairwise.wilcox.test (Response variable, grouping variable, paired =________,  p.adjust.method="bonf")
```

In particular, in this example, the variable being measured is AnxScore, and the grouping is done by Type. You do not have repeated measures on the same students, so you set: "paired=FALSE." Since you are doing multiple tests using the same data, you need to include the Bonferroni correction factor. Thus, for this example, the command is as follows.

```
> pairwise.wilcox.test (AnxScore, Type, paired=FALSE, p.adjust.method="bonf")
```

Here is the output.

```
Pairwise comparisons using Wilcoxon rank sum test

data:  AnxScore and Type

     COM   LPR   SPR
LPR  1.00  -     -
SPR  1.00  0.69  -
STA  1.00  1.00  1.00
```

This table shows the Bonferroni adjusted p-values for the pairwise comparisons. For example, the adjusted p-value for the comparison of the LPR median and the SPR median is 0.69. It is greater than α, so you can conclude there is no statistically significant difference in the median scores of students from the two types of institutions. In this example, all cases have large p-values; this was expected because the original Kruskal-Wallis test did not indicate any differences.

<u>EXAMPLE B</u> Section 34: "How to Run a Friedman's ANOVA" shows the process of using a Friedman test to determine whether or not the median anxiety score of students varies from one semester to the next. The same students are tested three times each; it is a repeated measurement situation. The output leads to not rejecting the null hypothesis, which suggests that the median score remains the same over time.

However, again suppose the result had been different. That is, suppose that the Friedman test had indicated that at least one median was different. Then you would probably want to run a "pairwise.wilcox.test" on this data. The data now has "paired = TRUE" because you are repeating the anxiety test on the same students several times. Following the general syntax above, you would adapt it for this example as:

```
> pairwise.wilcox.test (Anx.Score, Test.Session, paired=TRUE, p.adjust.method = "bonf")
```

The output is as follows.

```
        Pairwise comparisons using Wilcoxon signed rank test

data:  Anx.Score and Test.Session

        Fall1   Fall2
Fall2   0.54    -
Spr1    0.19    1.00
```

Again, the p-values here are all greater than α, indicating no statistically significant pairwise differences in this example. That was expected, based on the Friedman test results.