Pairwise Comparisons

Andrew Zieffler

Educational Psychology

University of Minnesota

Driven to Discover℠

Computing the Pairwise Test Results

After fitting the dummy coded models we find...

Contrast	***Unadjusted p-value***
30s. condition vs. 60s. condition	0.0000492
30s. condition vs. 180s. condition	0.00000267
60s. condition vs. 180s. condition	0.218

In general, with j groups there are

$$\frac{j(j-1)}{2}$$

pairwise contrasts

The rationale behind *p*-value adjustment is that the effect of the factor (e.g., teaching method), which is usually tested at the 0.05 significance-level has now been subdivided into many pairwise contrasts.

Is there an effect of condition?

$\alpha = 0.05$

$$\begin{cases} H_{01} : \mu_{30\text{s}} = \mu_{60\text{s}} \\ H_{01} : \mu_{30\text{s}} = \mu_{180\text{s}} \\ H_{01} : \mu_{60\text{s}} = \mu_{180\text{s}} \end{cases}$$

Family of tests

Can we just use the significance-level of 0.05 to test each of the pairwise hypotheses?

Is there an effect of condition?

$\alpha = 0.05$

$$\begin{cases} H_{01} : \mu_{30s} = \mu_{60s} & \alpha = 0.05 \\ H_{01} : \mu_{30s} = \mu_{180s} & \alpha = 0.05 \\ H_{01} : \mu_{60s} = \mu_{180s} & \alpha = 0.05 \end{cases}$$

This significance-level is known as the familywise error rate. It is the probability of making *at least one* type I error **in the family of tests**.

These significance-levels are known as the testwise error rates. Each is the probability of making a type I error **for that particular test**.

If we use a testwise error rate of 0.05 for the three pairwise contrast tests, the familywise error rate is actually higher than 0.05!

$$\alpha_{\mathrm{FW}} = 1 - (1 - \alpha)^k$$

$$\alpha_{\mathrm{FW}} = 1 - (1 - 0.05)^3 = 0.142625$$

The familywise error rate is 0.143.

Big question: What significance-level should we use to test each of the pairwise hypotheses if we want the familywise error rate to be 0.05?

Is there an effect of teaching method?

$\alpha = 0.05$

$$\begin{cases} H_{01} : \mu_{30s} = \mu_{60s} & \alpha = ? \\ H_{01} : \mu_{30s} = \mu_{180s} & \alpha = ? \\ H_{01} : \mu_{60s} = \mu_{180s} & \alpha = ? \end{cases}$$

$$\alpha_{\text{FW}} = 1 - (1 - \alpha)^k$$

$$0.05 = 1 - (1 - \alpha)^3$$

We could use algebra to solve for the testwise rate, but you would have to recall how to solve a polynomial.

Carlo Emilio Bonferroni

Bonferroni solved this type of algebra problem to find the value for alpha that gives an upper-bound for the familywise error rate of 0.05.

Olive Jean Dunn then used Bonferroni's solution in practice.

?

Olive Jean Dunn

This ensemble-adjustment procedure, the most common adjustment method used in the social sciences, is known as the *Dunn-Bonferroni* adjustment.

$$\alpha = \frac{\alpha_{\mathrm{FW}}}{k}$$

This made the family-wise error rate an upper-bound in

$$\alpha_{\mathrm{FW}} = 1 - (1 - \alpha)^k$$

$$\alpha = \frac{0.05}{3} = 0.017$$

The testwise alpha value for each pairwise contrast is 0.017. This means the *p*-value for each pairwise contrast should be compared to 0.017 rather than 0.05.

Ensemble Adjustment Methods

There are many, many, many, many ensemble-adjustment methods.

- Hommel procedure
- Benjamani-Hochberg procedure
- Fisher's Least Significant Difference (LSD) procedure
- Benjamani-Yekutieli procedure
- Tukey's Honestly Significant Difference (HSD) procedure
- Tukey-Kramer procedure
- Scheffé's procedure
- Shaffer's procedure
- Neuman-Keuls procedure
- Holm procedure
- Waller-Duncan procedure
- Hochberg procedure
- Miller-Winer procedure

To make it easier to report and interpret ensemble-adjusted results, it is better to **adjust the *p*-values** rather than the alpha-values.

Statistical software often implements many of these methods. Software typically adjusts the *p*-value so that you can use the value of 0.05 for comparison.

Put the unadjusted *p*-values in a vector

```
> p.values = c(0.0000492, 0.00000267, 0.218)
```

Use the `p.adjust()` function

```
> p.adjust(p.values, method = "bonferroni")

[1] 0.00014760 0.00000801 0.65400000
```

Contrast	***Unadjusted p-value***	***Bonferroni-adjusted p-values***
30s. condition vs. 60s. condition	0.0000492	0.0001476
30s. condition vs. 180s. condition	0.00000267	0.00000801
60s. condition vs. 180s. condition	0.218	0.654

Yoav Benjamani

Yosef Hochberg

The Benjamini–Hochberg procedure is an ensemble method based on **false discovery rate** (FDR).

FDR is a relatively new approach to the multiple comparisons problem. Instead of controlling the chance of at least one type I error, FDR **controls the expected proportion of type I errors** making these methods less prone to over-adjustment of the *p*-values.

Growing pool of evidence showing that this method may be the best solution to the multiple comparisons problem in many practical situations (Williams, Jones, & Tukey, 1999)

Because of its usefulness, the I*nstitute of Education Sciences* has recommended this procedure for use in its **What Works Clearinghouse** handbook of standards (Institute of Education Sciences, 2008).
http://ies.ed.gov/ncee/wwc/DocumentSum.aspx?sid=19

To understand how the Benjamani–Hochberg procedure makes adjustments, we first rank order the unadjusted *p*-values (from lowest to highest).

Contrast	***Unadjusted p-value***	***Rank***
30s. condition vs. 60s. condition	0.0000492	2
30s. condition vs. 180s. condition	0.00000267	1
60s. condition vs. 180s. condition	0.218	3

Start with the largest *p*-value, and adjust as follows:

$$p_{adj.} = \frac{k \times p_k}{\text{Rank}}$$

Contrast	Unadjusted *p-value*	Rank	Benjamani–Hochberg adjusted *p*-value
30s. condition vs. 60s. condition	0.0000492	2	$\frac{3 \times 0.0000492}{2} = 0.0000738$
30s. condition vs. 180s. condition	0.00000267	1	$\frac{3 \times 0.00000267}{1} = 0.00000801$
60s. condition vs. 180s. condition	0.218	3	$\frac{3 \times 0.218}{3} = 0.218$

```
> p.adjust(p.values, method = "BH")

[1] 0.00007380 0.00000801 0.21800000
```

Contrast	Unadjusted *p-value*	Dunn–Bonferroni adjusted *p*-value	Benjamani–Hochberg adjusted *p*-value
30s. condition vs. 60s. condition	0.0000492	0.0001476	0.0000738
30s. condition vs. 180s. condition	0.00000267	0.00000801	0.00000801
60s. condition vs. 180s. condition	0.218	0.654	0.218

Smaller *p*-values than the Dunn–Bonferroni

In general, some of the ensemble methods are less conservative (produce smaller *p*-values) others are more conservative (produce higher *p*-values)

Another way to Obtain the Pairwise Results

This requires that there is a single column in the data frame that has each subjects' condition. Furthermore, R needs to be treating that column as a factor.

```
> str(wr)

'data.frame':  24 obs. of  5 variables:
 $ condition: Factor w/ 3 levels "30s","60s","180s": 1 1 1 1 1 1 1 1 2 2 ...
 $ words    : num  7 3 6 6 5 8 6 7 7 11 ...
 $ con30    : num  1 1 1 1 1 1 1 1 0 0 ...
 $ con60    : num  0 0 0 0 0 0 0 0 1 1 ...
 $ con180   : num  0 0 0 0 0 0 0 0 0 0 ...
```

The column `condition` meets those requirements.

Use the `pairwise.t.test()` function to test *all* pairwise contrasts. The first argument is the outcome variable, and the second argument is the grouping variable.

```
> pairwise.t.test(wr$words, wr$condition)

    Pairwise comparisons using t tests with pooled SD

data:  wr$words and wr$condition

     30s      60s
60s  0.000098 -
180s 0.000008 0.22

P value adjustment method: holm
```

By default the ensemble adjustment method used is the Holm method.

The argument `p.adjust=` can be used to change the ensemble adjustment method.

```
> pairwise.t.test(wr$words, wr$condition, p.adjust = "BH")

    Pairwise comparisons using t tests with pooled SD
    Pairwise comparisons using t tests with pooled SD

data:  wr$words and wr$condition

     30s      60s
60s  0.000074 -
180s 0.000008 0.22

P value adjustment method: BH
```

Yet Another way to Obtain the Pairwise Results

To test the pairwise contrasts, we first set up a vector in which each of the levels from the grouping variable is specified as a difference (in a string). Each of these differences represents a pairwise contrasts we want to test.

```
> levels(wr$condition)
[1] "30s"  "60s"  "180s"

# Set up the contrasts to test
> contr = c("`30s` - `60s` = 0",
            "`30s` - `180s` = 0",
            "`60s` - `180s` = 0")
```

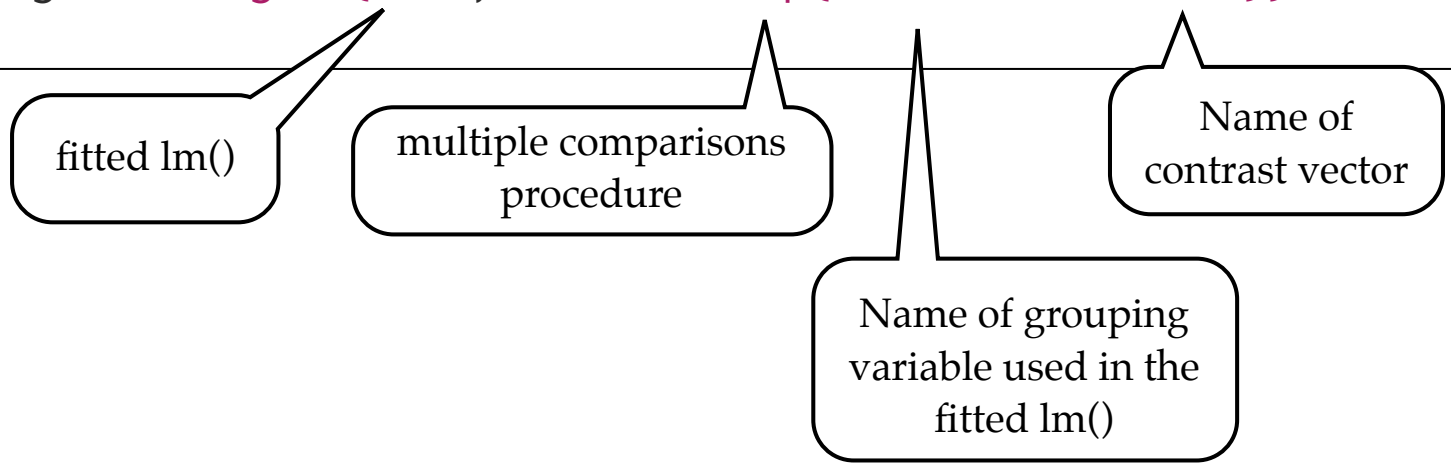
Because the factor levels start with a number we have to put their names in backticks in the contrast vector.

To obtain the test results we will use the `glht()` function from the **multcomp** library

```
# Fit a model using the grouping variable
> lm.1 = lm(words ~ condition, data = wr)

# Load multcomp library (you may need to install it first)
> library(multcomp)

# Use the glht() function to test pairwise contrasts
> glht.1 = glht(lm.1, linfct = mcp(condition = contr))
```



Use `summary()` on the fitted `glht()` model

```
# Get the Benjamani-Hochberg adjusted p-values
> summary(glht.1, test = adjusted("BH"))


Linear Hypotheses:
                    Estimate Std. Error t value  Pr(>|t|)
`30s` - `60s` == 0   -4.0000     0.7868  -5.084 0.0000738 ***
`30s` - `180s` == 0  -5.0000     0.7868  -6.355 0.0000080 ***
`60s` - `180s` == 0  -1.0000     0.7868  -1.271     0.218
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Adjusted p values reported -- BH method)
```

- Subjects in the 60 second condition recall four more words, on average, than subjects in the 60 second condition ($p < 0.001$).
- Subjects in the 180 second condition recall four more words, on average, than subjects in the 60 second condition ($p < 0.001$).
- Subjects in the 180 second condition recall one more word, on average, than subjects in the 60 second condition ($p = 0.218$).

There are two schools of thought about performing multiple *post hoc* tests on data:

- Use the unadjusted *p*-values
- Adjust the *p*-values for the number of post hoc tests carried out

Methods of adjustment invariably adjust the *p*-values upwards, making them **larger** than unadjusted *p*-values.

When an unadjusted *p*-value is very small or very large, rarely will an adjustment change the judgment regarding statistical reliability (significance).

For unadjusted *p*-values that hover around 0.05, adjustment might lead to a different judgment about the group differences.

Table 1.
Results from Testing Each of the Pairwise Contrasts in the Rutherford Data.

Contrast	Estimate	*SE*	*t*	*p*
30s. –. 60s. = 0	–4.00	0.79	–5.08	<0.001
30s. – 180s. = 0	–5.00	0.79	–6.36	<0.001
60s. – 180s. = 0	–1.00	0.79	–1.27	0.218

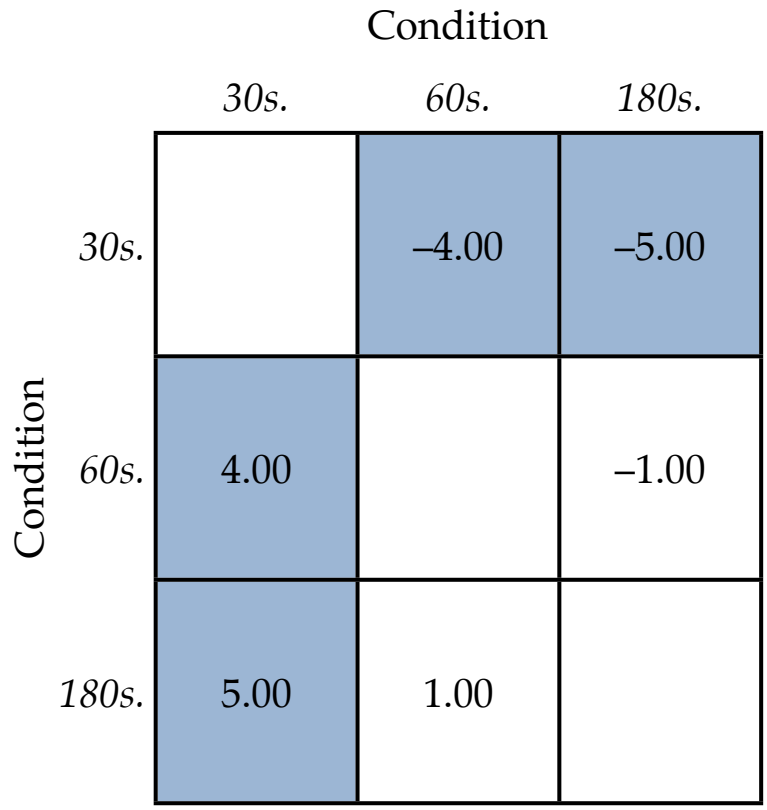
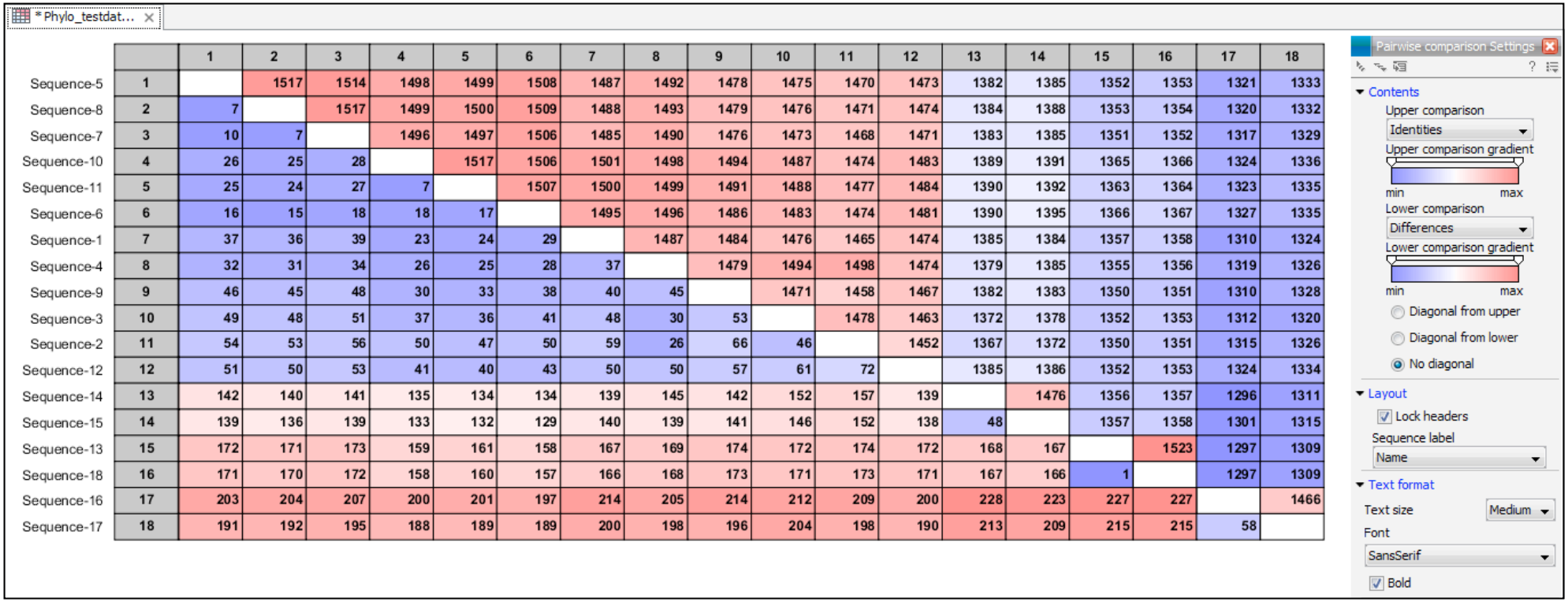


Figure 1. *Cells indicate the sample contrast estimates (row condition – column condition) for the difference in average number of words recalled. Color indicates statistical significance.*



		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Sequence-5	1		1517	1514	1498	1499	1508	1487	1492	1478	1475	1470	1473	1382	1385	1352	1353	1321	1333
Sequence-8	2	7		1517	1499	1500	1509	1488	1493	1479	1476	1471	1474	1384	1388	1353	1354	1320	1332
Sequence-7	3	10	7		1496	1497	1506	1485	1490	1476	1473	1468	1471	1383	1385	1351	1352	1317	1329
Sequence-10	4	26	25	28		1517	1506	1501	1498	1494	1487	1474	1483	1389	1391	1365	1366	1324	1336
Sequence-11	5	25	24	27	7		1507	1500	1499	1491	1488	1477	1484	1390	1392	1363	1364	1323	1335
Sequence-6	6	16	15	18	18	17		1495	1496	1486	1483	1474	1481	1390	1395	1366	1367	1327	1335
Sequence-1	7	37	36	39	23	24	29		1487	1484	1476	1465	1474	1385	1384	1357	1358	1310	1324
Sequence-4	8	32	31	34	26	25	28	37		1479	1494	1498	1474	1379	1385	1355	1356	1319	1326
Sequence-9	9	46	45	48	30	33	38	40	45		1471	1458	1467	1382	1383	1350	1351	1310	1328
Sequence-3	10	49	48	51	37	36	41	48	30	53		1478	1463	1372	1378	1352	1353	1312	1320
Sequence-2	11	54	53	56	50	47	50	59	26	66	46		1452	1367	1372	1350	1351	1315	1326
Sequence-12	12	51	50	53	41	40	43	50	50	57	61	72		1385	1386	1352	1353	1324	1334
Sequence-14	13	142	140	141	135	134	134	139	145	142	152	157	139		1476	1356	1357	1296	1311
Sequence-15	14	139	136	139	133	132	129	140	139	141	146	152	138	48		1357	1358	1301	1315
Sequence-13	15	172	171	173	159	161	158	167	169	174	172	174	172	168	167		1523	1297	1309
Sequence-18	16	171	170	172	158	160	157	166	168	173	171	173	171	167	166	1		1297	1309
Sequence-16	17	203	204	207	200	201	197	214	205	214	212	209	200	228	223	227	227		1466
Sequence-17	18	191	192	195	188	189	189	200	198	196	204	198	190	213	209	215	215	58	