Orthogonal contrasts and multiple comparisons

- Orthogonal contrasts
  - Class comparisons in R
  - Trend analysis in R
- Multiple mean comparisons

Orthogonal contrasts

Planned, single degree-of-freedom orthogonal contrasts are a powerful means of perfectly partitioning the treatment sum of squares (TSS) to gain greater insight into your data; and this method of analysis is available in R via **lm()**, in combination with a new function **contrasts()**.

Whenever you program contrasts, be sure to first use **str()** to inspect your inputted data for how R has ordered the levels within your treatments (usually alphanumerically). This order is important because your contrast coefficients need to correspond accurately to the levels of the indicated classification variable.

Once you know how R has ordered your treatment levels and once you know the contrast coefficients you wish to assign (i.e. you know the questions you want to ask), you will build an appropriate matrix of contrast coefficients. For example, for the gene dosage experiment discussed in lecture:

```
# Contrast 'Linear'      -1,0,1
# Contrast 'Quadratic'   1,-2,1
contrastmatrix<-cbind(c(-1,0,1),c(1,-2,1))
```

After building the contrast matrix, you will assign the matrix to your desired classification variable using the **contrasts()** function within R:

```
contrasts(data.dat$Treatment)<-contrastmatrix
```

This is the general procedure you will follow for contrasts, and it will make sense (don't worry!) in the context of an example.

CLASS COMPARISONS USING CONTRASTS

Example 1 *[Lab4ex1.R]*

This is a CRD in which 18 young *Cleome gynandra* plants were randomly assigned to 6 different treatments (i.e. all combinations of two temperature [High and Low] and three light [8, 12, and 16 hour days] conditions) and their growth measured.

```
#BIOL933 I
#Lab 4, example 1

#This script performs a class comparison analysis using orthogonal contrasts (CRD)

#read in, inspect, and re-classify the data
cleome_dat<-as_data.frame(cleome_dat)
cleome_dat$Trtmt<-as.factor(cleome_dat$Trtmt)
str(cleome_dat, give.attr=F)

#The ANOVA
cleome_mod<-lm(Growth ~ Trtmt, cleome_dat)
anova(cleome_mod)

#Need to assign contrast coefficients
#Notice from str() that R orders the Trtmt levels this way: H08, H12, H16, L08, L12, L16
# Our desired contrasts:
# Contrast 'Temp'                       1,1,1,-1,-1,-1
# Contrast 'Light Linear'               1,0,-1,1,0,-1
# Contrast 'Light Quadratic'            1,-2,1,1,-2,1
# Contrast 'Temp * Light Linear'        1,0,-1,-1,0,1
# Contrast 'Temp * Light Quadratic'     1,-2,1,-1,2,-1

contrastmatrix<-cbind(  c(1,1,1,-1,-1,-1),
                        c(1,0,-1,1,0,-1),
                        c(1,-2,1,1,-2,1),
                        c(1,0,-1,-1,0,1),
                        c(1,-2,1,-1,2,-1)   )
contrastmatrix

contrasts(cleome_dat$Trtmt)<-contrastmatrix
cleome_dat$Trtmt

cleome_contrast_mod<-aov(Growth ~ Trtmt, cleome_dat)
summary(cleome_contrast_mod, split = list(Trtmt = list("Temp" = 1, "Light Lin" = 2, "Light
 Quad" = 3, "T*L Lin" = 4, "T*L Quad" = 5)))


#In general, people do not report contrast SS; so in practice you can simply use lm(), as
 usual:
cleome_cont_mod<-lm(Growth ~ Trtmt, cleome_dat)
summary(cleome_cont_mod)
```

What questions we are asking here, exactly? To answer this, it is helpful to articulate the null hypothesis for each contrast:

Contrast 'Temp' H_0: Mean plant growth under low temperature conditions is the same as under high temperature conditions.
Contrast 'Light Linear' H_0: Mean plant growth under 8 hour days is the same as under 16 hour days (**OR**: The response of growth to light has no linear component).
Contrast 'Light Quadratic' H_0: Mean plant growth under 12 hour days is the same as the average mean growth under 8 and 16 hour days combined (**OR**: The growth response to light is perfectly linear; **OR**: The response of growth to light has no quadratic component).
Contrast 'Temp * Light Linear' H_0: The linear component of the response of growth to light is the same at both temperatures.
Contrast 'Temp * Light Quadratic' H_0: The quadratic component of the response of growth to light is the same at both temperatures.

So what would it mean to find significant results and to reject each of these null hypotheses?

Reject contrast 'Temp' H_0 = There is a significant response of growth to temperature.
Reject contrast 'Light linear' H_0 = The response of growth to light has a significant linear component.
Reject contrast 'Light quadratic' H_0 = The response of growth to light has a significant quadratic component.
Reject contrast 'Temp * Light Linear' H_0 = The linear component of the response of growth to light depends on temperature.
Reject contrast 'Temp * Light Quadratic' H_0 = The quadratic component of the response of growth to light depends on temperature.

Results

The original ANOVA

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Trtmt	5	718.57	143.714	16.689	4.881e-05	***
Residuals	12	103.33	8.611			

Output of the first summary() statement [Line 35]

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Trtmt	5	718.6	143.7	16.689	4.88e-05	***
Trtmt: Temp	1	606.7	606.7	70.453	**2.29e-06**	***
Trtmt: Light Lin	1	54.2	54.2	6.293	**0.0275**	*
Trtmt: Light Quad	1	35.0	35.0	4.065	**0.0667**	.
Trtmt: T*L Lin	1	11.0	11.0	1.280	**0.2800**	
Trtmt: T*L Quad	1	11.7	11.7	1.356	**0.2669**	
Residuals	12	103.3	8.6			

Things to notice

- Notice the sum of the contrast degrees of freedom. What does it equal? Why?
- Notice the sum of the contrast SS. What does it equal? Why?
- What insight does this analysis give you into your experiment?

For the sake of this demonstration, I used the **aov()** [analysis of variance] function in R instead of **lm()** because for simple experimental designs, **aov()** will report the contrast SS. In general, people report only p-values, not contrast SS; so in practice you should use **lm()**:

Output of the second summary() statement [Line 40]

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  23.1389     0.6917  33.454 3.23e-13 ***
Trtmt1        5.8056     0.6917   8.394 2.29e-06 ***
Trtmt2       -2.1250     0.8471  -2.509   0.0275 *
Trtmt3        0.9861     0.4891   2.016   0.0667 .
Trtmt4        0.9583     0.8471   1.131   0.2800
Trtmt5        0.5694     0.4891   1.164   0.2669
```

Here you see that the reported p-values are the same as above.

TREND ANALYSIS USING CONTRASTS

Example 2 *[Lab4ex2.R]*

This experiment, described in lecture, was conducted to investigate the relationship between plant spacing and yield in soybeans. The researcher randomly assigned five different plant spacings to 30 field plots, planted the soybeans accordingly, and measured the yield of each plot at the end of the season. Since we are interested in the overall relationship between plant spacing and yield (i.e. characterizing the response of yield to plant spacing), it is appropriate to perform a *trend analysis*.

```
#BIOL933
#Soybean spacing example, Lab 4 example 2

#This script performs a trend analysis using orthogonal contrasts (CRD), then using
 regression

#read in, inspect, and re-classify the data
spacing_dat<-as.data.frame(spacing_dat)
spacing_dat$Spacing<-as.factor(spacing_dat$Spacing)
str(spacing_dat, give.attr=F)

#R does not know Spacing is a classification variable.  We need to tell it:
spacing_dat$Spacing<-as.factor(spacing_dat$Spacing)
str(spacing_dat)

#The ANOVA
spacing_mod<-lm(Yield ~ Spacing, spacing_dat)
anova(spacing_mod)

#Need to assign contrast coefficients
# Contrast 'Linear'          -2,-1,0,1,2
# Contrast 'Quadratic'       2,-1,-2,-1,2
# Contrast 'Cubic'           -1,2,0,-2,1
# Contrast 'Quartic'         1,-4,6,-4,1
contrastmatrix<-cbind(  c(-2,-1,0,1,2), c(2,-1,-2,-1,2),
                        c(-1,2,0,-2,1), c(1,-4,6,-4,1)    )
contrastmatrix

contrasts(spacing_dat$Spacing)<-contrastmatrix
spacing_dat$Spacing

spacing_contrast_mod<-aov(Yield ~ Spacing, spacing_dat)
summary(spacing_contrast_mod, split = list(Spacing = list("Linear" = 1, "Quadratic" = 2,
 "Cubic" = 3, "Quartic" = 4)))

spacing_con_mod<-lm(Yield ~ Spacing, spacing_dat)
summary(spacing_con_mod)
```

What questions are we asking here exactly? As before, it is helpful to articulate the null hypothesis for each contrast:

Contrast 'Linear' H_0: The response of growth to spacing has no linear component.
Contrast 'Quadratic' H_0: The response of growth to spacing has no quadratic component.
Contrast 'Cubic' H_0: The response of growth to spacing has no cubic component.
Contrast 'Quartic' H_0: The response of growth to spacing has no quartic component.

Can you see, based on the contrast coefficients, why these are the null hypotheses?

Results

The original ANOVA

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Spacing	4	125.661	31.4153	9.9004	6.079e-05	***
Residuals	25	79.328	3.1731			

Output of the first summary() statement [Line 27]

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Spacing	4	125.66	31.42	9.900	6.08e-05	***
Spacing: Linear	1	91.27	91.27	28.762	**1.46e-05**	***
Spacing: Quadratic	1	33.69	33.69	10.618	**0.00322**	**
Spacing: Cubic	1	0.50	0.50	0.159	**0.69357**	
Spacing: Quartic	1	0.20	0.20	0.062	**0.80519**	
Residuals	25	79.33	3.17			

Output of the second summary() statement [Line 30]

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	31.30333	0.32522	96.251	< 2e-16	***
Spacing1	-1.23333	0.22997	-5.363	**1.46e-05**	***
Spacing2	0.63333	0.19436	3.259	**0.00322**	**
Spacing3	-0.09167	0.22997	-0.399	**0.69357**	
Spacing4	0.02167	0.08692	0.249	**0.80519**	

Interpretation

There is a quadratic relationship between row spacing and yield. Why? Because there is a significant quadratic component to the response but no significant cubic or quartic components. Please note that we are only able to carry out trend comparisons in this way because *the treatments are equally spaced*. Now, exactly the same result can be obtained through a regression approach, as shown in the next example.

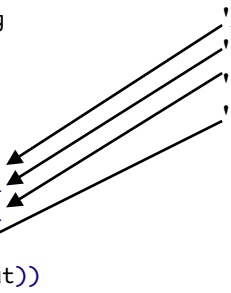
TREND ANALYSIS USING REGRESSION

```
#We can carry out the same analysis using a multiple regression approach
#Read in and inspect the data again
spacing_dat
str(spacing_dat)

#For regression, Spacing is no longer a classification variable, it is a regression variable:
spacing_dat$Spacing<-as.numeric(spacing_dat$Spacing)
str(spacing_dat)
```

```
Spacing<-spacing_dat$Spacing
Spacing2<-Spacing^2
Spacing3<-Spacing^3
Spacing4<-Spacing^4

anova(lm(Yield ~  Spacing +
                  Spacing2 +
                  Spacing3 +
                  Spacing4,
                  spacing_dat))
```



'Linear' — H_0: There is no linear component
'Quadratic' — H_0: There is no quadratic component
'Cubic' — H_0: There is no cubic component
'Quartic' — H_0: There is no quartic component

Results

Output of the anova() statement [Line 46]

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Spacing	1	91.267	91.267	28.7623	1.461e-05	***
Spacing2	1	33.693	33.693	10.6183	0.003218	**
Spacing3	1	0.504	0.504	0.1589	0.693568	
Spacing4	1	0.197	0.197	0.0621	0.805187	
Residuals	25	79.328	3.173			

Notice in this case that the results match perfectly those from our earlier analysis by contrasts.

For the interested:

To find the best-fit line for this data, you need only one line of code:

```
lm(Yield ~ poly(Spacing,2,raw=TRUE), spacing_dat)
```

With this command, R provides estimates of the polynomial coefficients:

```
Coefficients:
                  (Intercept)  poly(Spacing, 2, raw = TRUE)1  poly(Spacing, 2, raw = TRUE)2
                        52.03667                       -1.26111                        0.01759
```

In this case, the equation of the trend line that best fits our data would be:

$$\text{Yield} = 52.03667 - 1.26111 * \text{Sp} + 0.01759 * \text{Sp}^2$$

Multiple Mean Comparisons

Orthogonal contrasts are planned, *a priori* tests that partition the experimental variance cleanly. They are a powerful tool for analyzing data, but they are not appropriate for all experiments. Less restrictive comparisons among treatment means can be performed using various means separation tests, or multiple comparison tests, accessed in R through the "agricolae" package. To install this package on your computer, use RStudio's "Install Packages" function or simply type in the R console or script window:

```
install.packages("agricolae")
library(agricolae)
```

The first line downloads and installs the "agricolae" library on your computer. The second line loads the library into R's working memory to be used during the current work session. Once a package is loaded on your computer, you don't have to do it again. But each time you start a new work session, if you need a library, you will need to call it up (the 2nd line).

Regardless of the test, the syntax is straightforward and uniform within agricolae:

```
test.function(data.mod, "Factor")
```

This syntax tells R to:

1. Compute the means of the response variable for each level of the specified classification variable(s) ("Factor"), all of which were featured in the original **lm()** statement; then
2. Perform multiple comparisons among these means using the stated **test.function()**.

Some of the available **test.functions** are listed below:

Fixed Range Tests

· LSD.test()	Fisher's least significant difference test
· HSD.test()	Tukey's studentized range test (HSD: Honestly significant difference)
· scheffe.test()	Scheffé's test

Multiple Range Tests

· duncan.test()	Duncan's test
· SNK.test()	Student-Newman-Keuls test

The default significance level for comparisons among means is $\alpha = 0.05$, but this can be changed easily using the option alpha = α, where α is the desired significance level. The important thing to keep in mind is the EER (experimentwise error rate); we want to keep it controlled while keeping the test as sensitive as possible, so our choice of test should reflect that.

Note: Two tests are not supported by "agricolae", but available in other the R packages: "multcomp" contains the Dunnett test; "mutoss" contains the REGWQ test. The syntax for Dunnett is less convenient than the others. Both are demonstrated in the example below.

Example 3 *(One-Way Multiple Comparison)* *[Lab4ex3.R]*

Here's the clover experiment again, a CRD in which 30 different clover plants were randomly inoculated with six different strains of rhizobium and the resulting level of nitrogen fixation measured.

```
#BIOL933
#Lab 4, example 3

#This script illustrates the use of means separations tests

#read in and inspect the data
clover_dat
str(clover_dat)

clover_mod<-lm(NLevel ~ Culture, data = clover_dat)
anova(clover_mod)

# Install the required package
# LSD and other posthoc tests are not in the default packages; the package “agricolae”
contains scripts for LSD, Scheffe, Duncan, and SNK tests, among others. Agricolae was
developed by Felipe de Mendiburu as part of his master thesis "A statistical analysis tool
for agricultural research" – Univ. Nacional de Ingenieria, Lima-Peru (UNI).

#install.packages("agricolae")
#library(agricolae)

#FIXED RANGE TESTS
#LSD
LSD <- LSD.test(clover_mod, "Culture")

#Tukey HSD
Tukey <- HSD.test(clover_mod, "Culture")

#Scheffe
Scheffe <- scheffe.test(clover_mod, "Culture")

#MULTIPLE RANGE TESTS
#Duncan
Duncan <- duncan.test(clover_mod, "Culture")

#SNK
SNK <- SNK.test(clover_mod, "Culture")


#---Dunnett Test---
#For this, install package "multcomp"; some less convenient syntax is needed

#install.packages("multcomp")
#library(multcomp)

#The Dunnett Test uses the first treatment (alphanumerically) as the "control" -- hence
renamed 1Comp
#Also note in the output that the "Quantile" (2.6957) is the Critical Value of Dunnett's t
test.dunnett=glht(clover_mod,linfct=mcp(Culture="Dunnett"))
confint(test.dunnett)


#---REGWQ Test---
#For this, package "mutoss" is needed.  Unfortunately, this package has a dependency
#called "multtest" which is no longer available on CRAN.  "multtest" is
#available, however, through bioconductor.  To install, run the following lines:
```

```
source("http://bioconductor.org/biocLite.R")
biocLite("multtest")

#Then install the package "mutoss" as usual.  The REGWQ analysis:
#library(mutoss)
regwq(clover.mod, clover.dat, alpha = 0.05)


#As always, in thinking about the results of various analyses, it is useful to visualize the
data
plot(clover.dat, main = "Boxplot comparing treatment means")
```

In this experiment, there is no obvious structure to the treatment levels and therefore no way to anticipate the relevant questions to ask. We want to know how the different rhizobial strains performed; and to do this, we must systematically make all pair-wise comparisons among them.

In the output below, keep an eye on the least (or minimum) significant difference(s) used for each test.

What is indicated by changes in these values from test to test?

Also notice how the comparisons change significance with the different tests.

t Tests (LSD) for Nlevel

This test controls the Type I comparisonwise error rate, not the experimentwise error rate.

```
Mean Square Error:  11.78867
alpha: 0.05 ; Df Error: 24
Critical Value of t: 2.063899
Least Significant Difference 4.481782

Means with the same letter are not significantly different.

Groups, Treatments and means
a       3DOk1           28.82
b       3DOk5           23.98
bc      3DOk7           19.92
c       1Comp           18.7
d       3DOk4           14.14
d       3DOk13          13.16
```

Dunnett's t Tests for Nlevel

This test controls the Type I experimentwise error for comparisons of all treatments against a control.

```
Multiple Comparisons of Means: Dunnett Contrasts

Quantile = 2.6951 (<- Dunnett critical value; 2.6951*sqrt(2*11.78867/5) = 5.852 (LSD)
95% family-wise confidence level

Linear Hypotheses:
                     Estimate lwr      upr
3DOk1 - 1Comp == 0   10.1200   4.2677  15.9723 ***
3DOk13 - 1Comp == 0  -5.5400 -11.3923   0.3123
3DOk4 - 1Comp == 0   -4.5600 -10.4123   1.2923
3DOk5 - 1Comp == 0    5.2800  -0.5723  11.1323
3DOk7 - 1Comp == 0    1.2200  -4.6323   7.0723

                 Comparisons significant at the 0.05 level are indicated by ***.
```

Tukey's Studentized Range (HSD) Test for Nlevel

This test controls the Type I experimentwise error rate (MEER), but it generally has a higher Type II error rate than REGWQ.

```
Mean Square Error:  11.78867
alpha: 0.05 ; Df Error: 24
Critical Value of Studentized Range: 4.372651
Honestly Significant Difference: 6.714167

Means with the same letter are not significantly different.

Groups, Treatments and means
a       3DOk1           28.82
ab      3DOk5           23.98
bc      3DOk7           19.92
bcd     1Comp           18.7
cd      3DOk4           14.14
c       3DOk13          13.16
```

Scheffe's Test for Nlevel **[*For group comparisons with Scheffe, see lecture notes*]**

This test controls the Type I MEER.

```
Mean Square Error  : 11.78867
alpha: 0.05 ; Df Error: 24
Critical Value of F: 2.620654
Minimum Significant Difference: 7.860537

Means with the same letter are not significantly different.

Groups, Treatments and means
a        3DOk1           28.82
ab       3DOk5           23.98
bc       3DOk7           19.92
bc       1Comp           18.7
c        3DOk4           14.14
c        3DOk13          13.16
```

Duncan's Multiple Range Test for Nlevel

This test controls the Type I comparisonwise error rate, not the MEER.

```
Mean Square Error:  11.78867
alpha: 0.05 ; Df Error: 24

Critical Range
       2        3        4        5        6
4.481781 4.707218 4.851958 4.954192 5.030507

Means with the same letter are not significantly different.

Groups, Treatments and means
a        3DOk1           28.82
b        3DOk5           23.98
bc       3DOk7           19.92
c        1Comp           18.7
d        3DOk4           14.14
d        3DOk13          13.16
```

Student-Newman-Keuls (SNK) Test for Nlevel

This test controls the Type I experimentwise error rate under the complete null hypothesis (EERC) but not under partial null hypotheses (EERP).

```
Mean Square Error:  11.78867
alpha: 0.05 ; Df Error: 24

Critical Range
       2        3        4        5        6
4.481781 5.422890 5.990354 6.397338 6.714167

Means with the same letter are not significantly different.

Groups, Treatments and means
a        3DOk1           28.82
b        3DOk5           23.98
b        3DOk7           19.92
b        1Comp           18.7
c        3DOk4           14.14
c        3DOk13          13.16
```

Ryan-Einot-Gabriel-Welsch (REGWQ) Multiple Range Test for Nlevel

This test controls the Type I MEER.

The output here is very inconvenient:

```
Number of hyp.:         15
Number of rej.:         7
  rejected pValues adjPValues
1        1       0          0
2        3       0          0
3       10  0.0012     0.0012
4        4  0.0229     0.0229
5        2   4e-04      4e-04
6        6   5e-04      5e-04
7        5   7e-04      7e-04

$confIntervals
               [,1] [,2] [,3]
3DOk1-3DOk13 15.66   NA   NA
3DOk5-3DOk13 10.82   NA   NA
3DOk1-3DOk4  14.68   NA   NA
3DOk7-3DOk13  6.76   NA   NA
3DOk5-3DOk4   9.84   NA   NA
3DOk1-1Comp  10.12   NA   NA
1Comp-3DOk13  5.54   NA   NA
3DOk7-3DOk4   5.78   NA   NA
3DOk5-1Comp   5.28   NA   NA
3DOk1-3DOk7   8.90   NA   NA
3DOk4-3DOk13  0.98   NA   NA
1Comp-3DOk4   4.56   NA   NA
3DOk7-1Comp   1.22   NA   NA
3DOk5-3DOk7   4.06   NA   NA
3DOk1-3DOk5   4.84   NA   NA
```

Wrestling with all the above, one can manually pull together the following table:

Groups,	Treatments	and means
a	3DOk1	28.82
ab	3DOk5	23.98
bc	3DOk7	19.92
bcd	1Comp	18.7
cd	3DOk4	14.14
d	3DOk13	13.16

To make the relationships among the tests easier to see here is a summary table of the above results:

	Significance Groupings						
Culture	**LSD**	**Dunnett**	**Tukey**	**Scheffe**	**Duncan**	**SNK**	**REGWQ**
3DOk1	A	***	A	A	A	A	A
3DOk5	B		AB	AB	B	B	AB
3DOk7	BC		BC	BC	BC	B	BC
Comp	C		BCD	BC	C	B	BCD
3DOk4	D		CD	C	D	C	CD
3DOk13	D		D	C	D	C	D

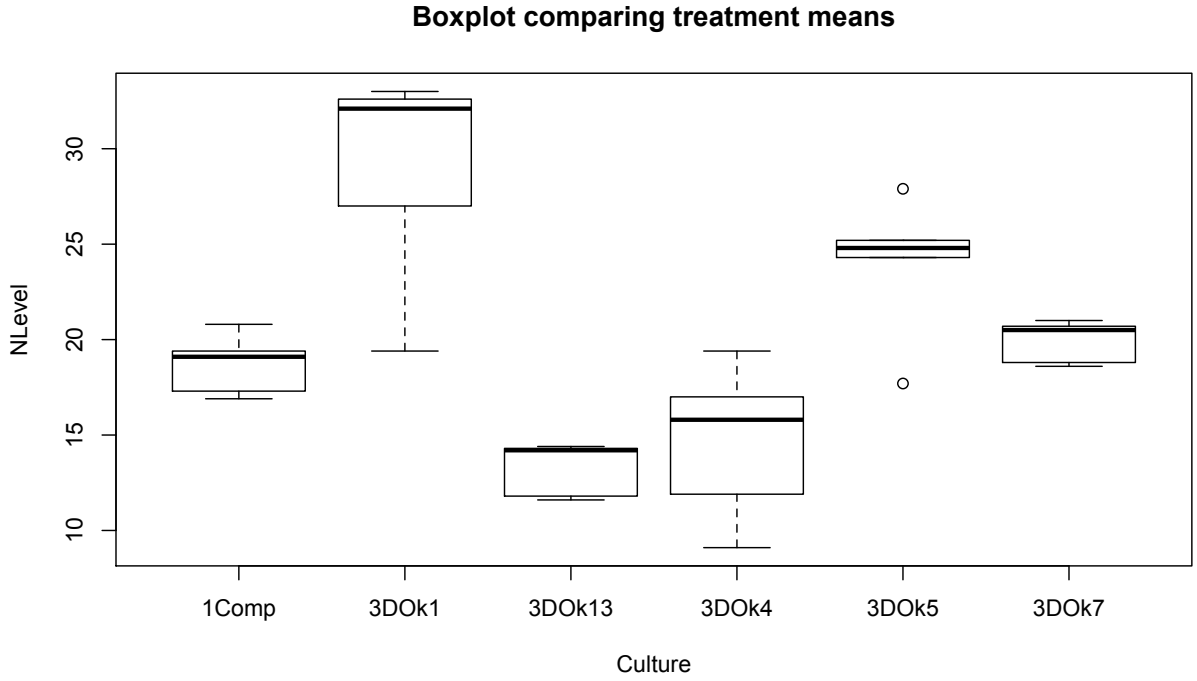
Least Sig't Difference	4.482 fixed	5.852 fixed	6.714 fixed	7.861 fixed	4.482 5.031	4.482 6.714	>4.482 6.714
EER Control	no	yes	yes	yes	no	EERC only	yes

Notice where the non-EER-controlling tests get you into potential Type I trouble, namely by their readiness to declare significant difference between 3DOk5 and Comp, as well as between 3Dok4 and Comp.

On the other hand, regarding potential Type II trouble, notice where the large LSD (7.861) of the relatively insensitive Scheffe's test (insensitive due to its ability to make unlimited pair-wise and group comparisons) failed to pick up a difference detected by other EER-controlling tests (e.g. between 3DOk7 and 3DOk13). In this particular case, the more elaborate REGWQ method returned the same results as Tukey.

Remember, while you should steer clear of tests that do not control for EER, there's no "right" test or "wrong" test. There's only knowing the characteristics of each and choosing the most appropriate one for your experiment (and the culture of your discipline).

As always, it is instructive (nay, encouraged) to consider the above table of comparisons with the boxplot below in hand:



Something to think about:

Does the boxplot above raise any red flags for you about your data?
How would go about investigating such concerns?