

CH9: ANOVA pairwise comparisons and contrasts

1 Clover Example: Multiple Comparisons for One-Way ANOVA

Nitrogen (N) content of Red Clover inoculated with combinations of *Rhizobium trifolii* and of *Rhizobium meliloti* strains. Five pots of each of six treatments completely randomized in a greenhouse experiment ($t = 6$ trts, $n = 5$ pots/trt, $nT=30$).

We perform Tukey, Bonferoni and Dunnett adjusted pairwise comparisons using the emmeans package.

Notes:

1. Strain is a factor. Check using the `str()` output.
2. For Dunnett's method, the first group is used as "control". For this data, *compos* acts at the control. Hence, we reorder factor levels so that the control group of interest is first. This only needs to be done if Dunnett comparisons are of interest!

```
library(dplyr)
library(emmeans)
clover <- read.csv("C:/hess/STAT511_FA11/RData/CH9_clover.csv")
str(clover)

## 'data.frame': 30 obs. of 2 variables:
## $ Strain: Factor w/ 6 levels "3D0k1","3D0k13",...: 1 1 1 1 1 2 2 2 2 2 ...
## $ N : num 19.4 32.1 32.6 33 27 14.4 14.2 11.8 14.3 11.6 ...

#Reorder levels of strain for convenience with Dunnett's test
levels(clover$Strain)

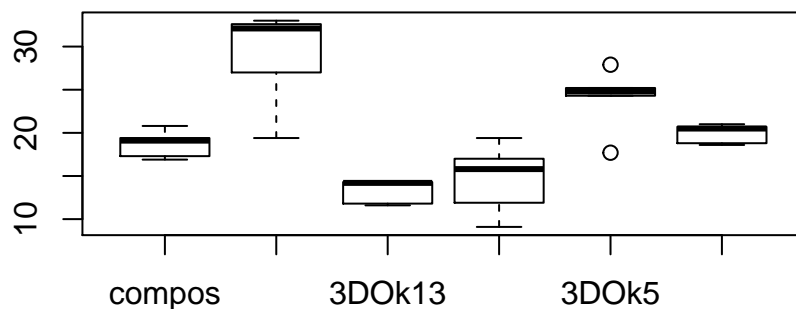
## [1] "3D0k1" "3D0k13" "3D0k4" "3D0k5" "3D0k7" "compos"
clover$Strain <- factor(clover$Strain, levels(clover$Strain)[c(6,1:5)])
levels(clover$Strain)

## [1] "compos" "3D0k1" "3D0k13" "3D0k4" "3D0k5" "3D0k7"
```

1.1 Basic One-way ANOVA analysis

```
boxplot(N ~ Strain, data = clover, main = "Boxplots")
```

Boxplots



```
SumStats <- summarise(group_by(clover, Strain),
  n      = n(),
  mean   = mean(N),
  sd     = sd(N),
  se     = sd / sqrt(n) )
```

```
SumStats
```

```
## # A tibble: 6 x 5
##   Strain      n mean    sd    se
##   <fct> <int> <dbl> <dbl> <dbl>
## 1 compos      5  18.7  1.60  0.716
## 2 3D0k1       5  28.8  5.80  2.59
## 3 3D0k13      5  13.3  1.43  0.638
## 4 3D0k4       5  14.6  4.12  1.84
## 5 3D0k5       5  24.0  3.78  1.69
## 6 3D0k7       5  19.9  1.13  0.505
```

```
OneWayFit <- lm(N ~ Strain, data = clover)
anova(OneWayFit)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Response: N
```

```
##          Df Sum Sq Mean Sq F value    Pr(>F)
## Strain      5  847.05  169.409    14.37 1.485e-06 ***
## Residuals  24  282.93   11.789
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

1.2 Adjusted Pairwise comparisons using emmeans

```
#Tukey Adjusted Pairwise Comparisons
emmout <- emmeans(OneWayFit, pairwise ~ Strain)
emmout
```

```
## $emmeans
## Strain emmean      SE df lower.CL upper.CL
## compos  18.70 1.535491 24  15.5309  21.8691
## 3D0k1    28.82 1.535491 24  25.6509  31.9891
## 3D0k13   13.26 1.535491 24  10.0909  16.4291
## 3D0k4    14.64 1.535491 24  11.4709  17.8091
## 3D0k5    23.98 1.535491 24  20.8109  27.1491
## 3D0k7    19.92 1.535491 24  16.7509  23.0891
##
## Confidence level used: 0.95
##
## $contrasts
## contrast      estimate      SE df t.ratio p.value
## compos - 3D0k1   -10.12 2.171513 24  -4.660  0.0012
## compos - 3D0k13    5.44 2.171513 24   2.505  0.1622
## compos - 3D0k4     4.06 2.171513 24   1.870  0.4435
## compos - 3D0k5    -5.28 2.171513 24  -2.431  0.1852
## compos - 3D0k7    -1.22 2.171513 24  -0.562  0.9926
## 3D0k1 - 3D0k13   15.56 2.171513 24   7.166 <.0001
## 3D0k1 - 3D0k4   14.18 2.171513 24   6.530 <.0001
## 3D0k1 - 3D0k5    4.84 2.171513 24   2.229  0.2617
## 3D0k1 - 3D0k7    8.90 2.171513 24   4.099  0.0049
## 3D0k13 - 3D0k4   -1.38 2.171513 24  -0.636  0.9871
## 3D0k13 - 3D0k5  -10.72 2.171513 24  -4.937  0.0006
## 3D0k13 - 3D0k7   -6.66 2.171513 24  -3.067  0.0528
## 3D0k4 - 3D0k5   -9.34 2.171513 24  -4.301  0.0030
## 3D0k4 - 3D0k7   -5.28 2.171513 24  -2.431  0.1852
## 3D0k5 - 3D0k7    4.06 2.171513 24   1.870  0.4435
##
## P value adjustment: tukey method for comparing a family of 6 estimates
```

#Bonferroni Adjusted Pairwise Comparisons

```
emmeans(OneWayFit, pairwise ~ Strain, adjust = "bonferroni")
```

```
## $emmeans
## Strain emmean      SE df lower.CL upper.CL
## compos  18.70 1.535491 24  15.5309  21.8691
## 3D0k1    28.82 1.535491 24  25.6509  31.9891
## 3D0k13   13.26 1.535491 24  10.0909  16.4291
## 3D0k4    14.64 1.535491 24  11.4709  17.8091
## 3D0k5    23.98 1.535491 24  20.8109  27.1491
## 3D0k7    19.92 1.535491 24  16.7509  23.0891
##
## Confidence level used: 0.95
##
## $contrasts
## contrast      estimate      SE df t.ratio p.value
## compos - 3D0k1   -10.12 2.171513 24  -4.660  0.0015
## compos - 3D0k13    5.44 2.171513 24   2.505  0.2914
## compos - 3D0k4     4.06 2.171513 24   1.870  1.0000
## compos - 3D0k5    -5.28 2.171513 24  -2.431  0.3431
## compos - 3D0k7    -1.22 2.171513 24  -0.562  1.0000
## 3D0k1 - 3D0k13   15.56 2.171513 24   7.166 <.0001
## 3D0k1 - 3D0k4   14.18 2.171513 24   6.530 <.0001
## 3D0k1 - 3D0k5    4.84 2.171513 24   2.229  0.5317
```

```
## 3D0k1 - 3D0k7      8.90 2.171513 24   4.099 0.0062
## 3D0k13 - 3D0k4     -1.38 2.171513 24   -0.636 1.0000
## 3D0k13 - 3D0k5    -10.72 2.171513 24   -4.937 0.0007
## 3D0k13 - 3D0k7     -6.66 2.171513 24   -3.067 0.0793
## 3D0k4 - 3D0k5      -9.34 2.171513 24   -4.301 0.0037
## 3D0k4 - 3D0k7      -5.28 2.171513 24   -2.431 0.3431
## 3D0k5 - 3D0k7       4.06 2.171513 24    1.870 1.0000
##
## P value adjustment: bonferroni method for 15 tests
#Dunnett Adjusted Comparisons
#Note that the factor levels were reordered above, so that control is listed first!
emmmeans(OneWayFit, dunnett ~ Strain)

## $emmeans
## Strain emmean      SE df lower.CL upper.CL
## compos  18.70 1.535491 24   15.5309  21.8691
## 3D0k1    28.82 1.535491 24   25.6509  31.9891
## 3D0k13   13.26 1.535491 24   10.0909  16.4291
## 3D0k4    14.64 1.535491 24   11.4709  17.8091
## 3D0k5    23.98 1.535491 24   20.8109  27.1491
## 3D0k7    19.92 1.535491 24   16.7509  23.0891
##
## Confidence level used: 0.95
##
## $contrasts
## contrast      estimate      SE df t.ratio p.value
## 3D0k1 - compos    10.12 2.171513 24    4.660 0.0005
## 3D0k13 - compos   -5.44 2.171513 24   -2.505 0.0776
## 3D0k4 - compos    -4.06 2.171513 24   -1.870 0.2543
## 3D0k5 - compos     5.28 2.171513 24    2.431 0.0902
## 3D0k7 - compos     1.22 2.171513 24    0.562 0.9349
##
## P value adjustment: dunnettx method for 5 tests
```

1.3 CLD, plotting and confidence intervals

The emmout object was created above and contains the Tukey adjusted pairwise comparisons.

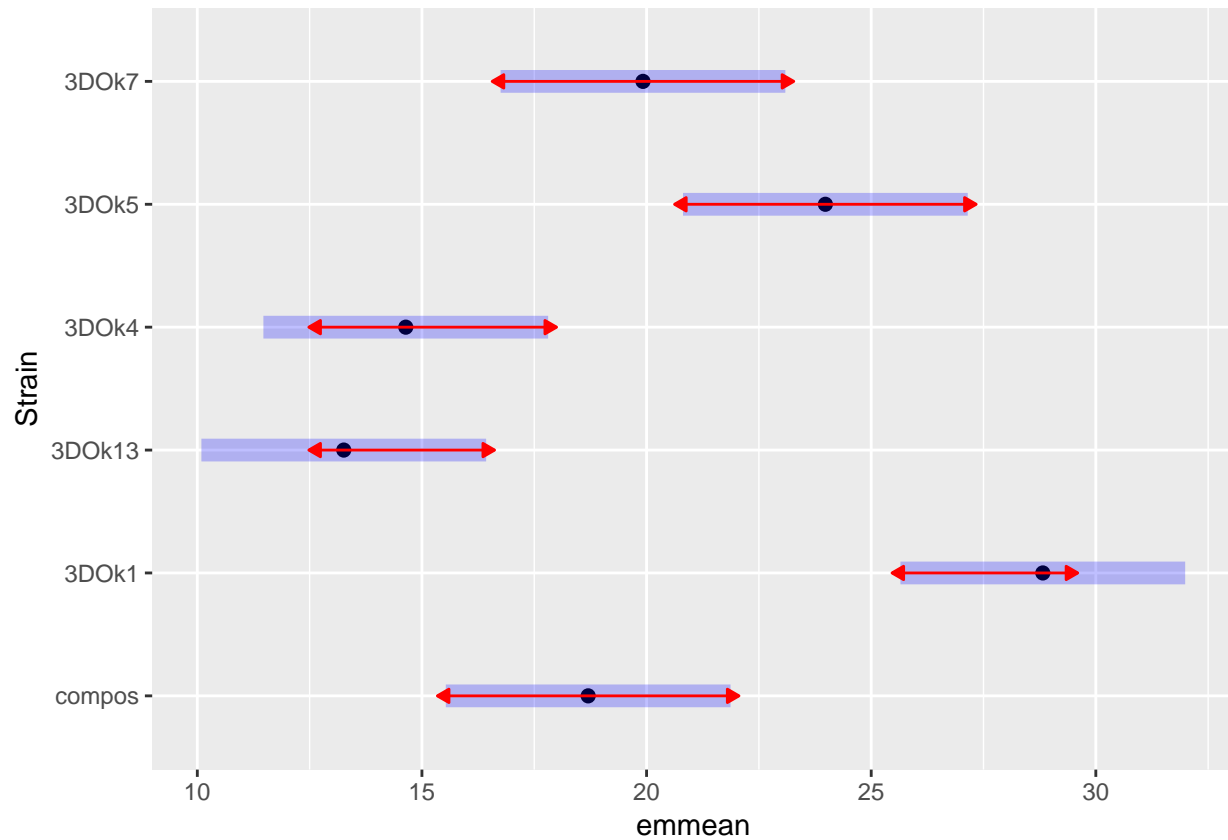
The CLD() function returns the “compact letter display”. Means that share a number grouping are NOT statistically significantly different.

```
CLD(emmout$emmeans)

## Strain emmean      SE df lower.CL upper.CL .group
## 3D0k13  13.26 1.535491 24   10.0909  16.4291    1
## 3D0k4   14.64 1.535491 24   11.4709  17.8091    1
## compos  18.70 1.535491 24   15.5309  21.8691   12
## 3D0k7   19.92 1.535491 24   16.7509  23.0891   12
## 3D0k5   23.98 1.535491 24   20.8109  27.1491   23
## 3D0k1   28.82 1.535491 24   25.6509  31.9891    3
##
## Confidence level used: 0.95
## P value adjustment: tukey method for comparing a family of 6 estimates
## significance level used: alpha = 0.05
```

The `plot()` function provides a graphical summary. The blue bars are confidence intervals for the EMMs, and the red arrows are for the comparisons among them. If an arrow from one mean overlaps an arrow from another group, the difference is not significant, based on the `adjust` setting (which defaults to “tukey”). (Note: Don’t ever use confidence intervals for EMMs to perform comparisons; they can be very misleading.)

```
plot(emmout$emmeans, comparisons = TRUE)
```



The `confint()` function returns the adjusted confidence intervals.

```
confint(emmout$contrasts)
```

```
## contrast      estimate      SE df  lower.CL  upper.CL
## compos - 3D0k1    -10.12 2.171513 24 -16.834167 -3.40583296
## compos - 3D0k13     5.44 2.171513 24  -1.274167 12.15416704
## compos - 3D0k4      4.06 2.171513 24  -2.654167 10.77416704
## compos - 3D0k5     -5.28 2.171513 24 -11.994167  1.43416704
## compos - 3D0k7     -1.22 2.171513 24  -7.934167  5.49416704
## 3D0k1 - 3D0k13    15.56 2.171513 24   8.845833 22.27416704
## 3D0k1 - 3D0k4    14.18 2.171513 24   7.465833 20.89416704
## 3D0k1 - 3D0k5      4.84 2.171513 24  -1.874167 11.55416704
## 3D0k1 - 3D0k7      8.90 2.171513 24   2.185833 15.61416704
## 3D0k13 - 3D0k4    -1.38 2.171513 24  -8.094167  5.33416704
## 3D0k13 - 3D0k5   -10.72 2.171513 24 -17.434167 -4.00583296
## 3D0k13 - 3D0k7    -6.66 2.171513 24 -13.374167  0.05416704
## 3D0k4 - 3D0k5     -9.34 2.171513 24 -16.054167 -2.62583296
## 3D0k4 - 3D0k7     -5.28 2.171513 24 -11.994167  1.43416704
## 3D0k5 - 3D0k7      4.06 2.171513 24  -2.654167 10.77416704
##
```

```
## Confidence level used: 0.95
## Conf-level adjustment: tukey method for comparing a family of 6 estimates
rm(clover, SumStats, OneWayFit, emmout)
```

2 Wheat Example: Contrasts for One-Way ANOVA

For many analyses, the ANOVA table and pairwise comparisons of means address all research questions. However, in some cases contrasts are required to address additional comparisons of interest.

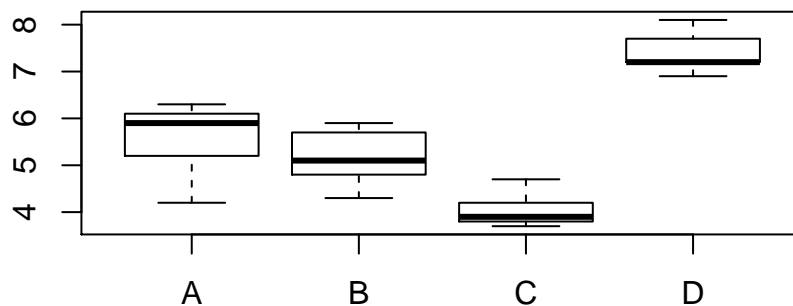
In this example, we analyze a one-way completely randomized design with 4 treatments (A, B, C, D) and 5 observations per treatment.

```
library(dplyr)
library(emmeans)
wheat <- read.csv("C:/hess/STAT511_FA11/RData/CH8_Wheat.csv")
str(wheat)
```

```
## 'data.frame': 20 obs. of 2 variables:
## $ variety: Factor w/ 4 levels "A","B","C","D": 1 1 1 1 1 2 2 2 2 2 ...
## $ yield : num 6.3 5.9 4.2 5.2 6.1 4.3 4.8 5.7 5.9 5.1 ...
```

2.1 Basic One-way ANOVA analysis

```
boxplot(yield ~ variety, data = wheat)
```



```
SumStats <- summarise(group_by(wheat, variety),
  n = n(),
  mean = mean(yield),
  sd = sd(yield),
  se = sd / sqrt(n) )
```

```
SumStats
```

```
## # A tibble: 4 x 5
##   variety     n mean    sd    se
##   <fct>   <int> <dbl> <dbl> <dbl>
## 1 A         5  5.54 0.856 0.383
## 2 B         5  5.16 0.654 0.293
## 3 C         5  4.06 0.404 0.181
## 4 D         5  7.42 0.476 0.213
```

```

OneWayFit <- lm(yield ~ variety, data = wheat)
anova(OneWayFit)

## Analysis of Variance Table
##
## Response: yield
##           Df Sum Sq Mean Sq F value    Pr(>F)
## variety    3 29.346   9.7818  25.227 2.64e-06 ***
## Residuals 16   6.204   0.3878
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

emmout <- emmeans(OneWayFit, "variety")
emmout

##   variety emmean      SE df lower.CL upper.CL
##   A             5.54 0.278478 16  4.949653  6.130347
##   B             5.16 0.278478 16  4.569653  5.750347
##   C             4.06 0.278478 16  3.469653  4.650347
##   D             7.42 0.278478 16  6.829653  8.010347
##
## Confidence level used: 0.95

pairs(emmout, adjust = "none")

##   contrast estimate      SE df t.ratio p.value
##   A - B           0.38 0.3938274 16   0.965  0.3490
##   A - C           1.48 0.3938274 16   3.758  0.0017
##   A - D          -1.88 0.3938274 16  -4.774  0.0002
##   B - C           1.10 0.3938274 16   2.793  0.0130
##   B - D          -2.26 0.3938274 16  -5.739 <.0001
##   C - D          -3.36 0.3938274 16  -8.532 <.0001

```

2.2 Contrasts

The order of coefficients is the order of sample means. Can check the contrast estimates using the sample means or emmeans.

```

contrast(emmout, list(
  #Contrast1: A vs B
  AvB = c(1, -1, 0, 0),
  #Contrast2: A+B vs C+D
  ABvCD = c(0.5, 0.5, -0.5, -0.5),
  #Contrast3: D vs others
  DvOthers = c(-1/3, -1/3, -1/3, 1),
  #Contrast4: C vs (A+B)
  CvAB = c(-0.5, -0.5, 1, 0)
))

##   contrast estimate      SE df t.ratio p.value
##   AvB           0.38 0.3938274 16   0.965  0.3490
##   ABvCD         -0.39 0.2784780 16  -1.400  0.1805
##   DvOthers       2.50 0.3215587 16   7.775 <.0001
##   CvAB          -1.29 0.3410645 16  -3.782  0.0016

```