

Two-Way ANOVA Handout

1 Initialization

```
> library(NCStats)
> library(gdata)
> library(car)
> library(multcomp)
```

2 Bacteria Example

```
> bact <- read.table("Bacteria.txt",head=TRUE)
> str(bact)

'data.frame':      30 obs. of  3 variables:
 $ temp : int  27 27 27 27 27 35 35 35 35 35 ...
 $ conc : num  0.6 0.8 1 1.2 1.4 0.6 0.8 1 1.2 1.4 ...
 $ cells: int  55 120 186 260 151 82 166 179 223 178 ...

> bact$ftemp <- factor(bact$temp)
> bact$fconc <- factor(bact$conc)
> attach(bact)
```

Explore sample size per group, univariate statistics, statistics for each group, and table of group means,

```
> table(ftemp,fconc)

      fconc
ftemp 0.6 0.8 1 1.2 1.4
 27    2  2 2  2  2
 35    2  2 2  2  2
 43    2  2 2  2  2

> Summarize(cells~ftemp)

      n Mean St. Dev. Min. 1st Qu. Median 3rd Qu. Max.
27 10 153.1 69.83544  55  113.2 141.5  177.2 274
35 10 169.3 51.14044  82  157.0 172.0  209.0 236
43 10 161.5 33.95176 115  138.0 149.5  185.2 228

> Summarize(cells~fconc)

      n      Mean St. Dev. Min. 1st Qu. Median 3rd Qu. Max.
0.6  6 108.1667 38.77843  55   85.0 104.5  141.2 153
```

```

0.8 6 144.1667 34.00245 91 126.5 151.0 163.5 186
1 6 154.8333 23.69318 133 135.2 148.0 174.2 186
1.2 6 235.0000 29.85297 189 224.2 232.0 254.0 274
1.4 6 164.3333 37.30773 111 145.8 164.5 181.8 219

```

```
> Summarize(cells~interaction(ftemp,fconc,sep=":"),numdigs=2)
```

```

      n Mean St. Dev. Min. 1st Qu. Median 3rd Qu. Max.
27:0.6 2 102.5    67.18  55  78.75 102.5  126.2 150
35:0.6 2  88.0     8.49  82  85.00  88.0   91.0  94
43:0.6 2 134.0    26.87 115 124.50 134.0  143.5 153
27:0.8 2 105.5    20.51  91  98.25 105.5  112.8 120
35:0.8 2 161.0     7.07 156 158.50 161.0  163.5 166
43:0.8 2 166.0    28.28 146 156.00 166.0  176.0 186
27:1 2 159.5    37.48 133 146.20 159.5  172.8 186
35:1 2 169.5    13.44 160 164.80 169.5  174.2 179
43:1 2 135.5     0.71 135 135.20 135.5  135.8 136
27:1.2 2 267.0     9.90 260 263.50 267.0  270.5 274
35:1.2 2 229.5     9.19 223 226.20 229.5  232.8 236
43:1.2 2 208.5    27.58 189 198.80 208.5  218.2 228
27:1.4 2 131.0    28.28 111 121.00 131.0  141.0 151
35:1.4 2 198.5    28.99 178 188.20 198.5  208.8 219
43:1.4 2 163.5    27.58 144 153.80 163.5  173.2 183

```

```
> aggregate.table(cells,ftemp,fconc,FUN=mean)
```

```

      0.6  0.8  1  1.2  1.4
27 102.5 105.5 159.5 267.0 131.0
35  88.0 161.0 169.5 229.5 198.5
43 134.0 166.0 135.5 208.5 163.5

```

Fit the model and get the ANOVA results

```
> lm1 <- lm(cells~ftemp*fconc)
> anova(lm1)
```

Analysis of Variance Table

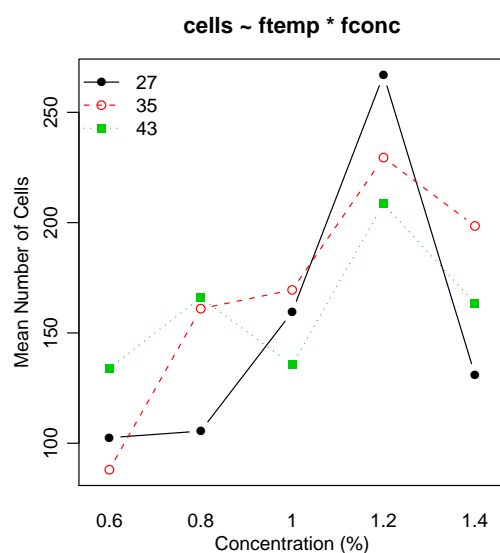
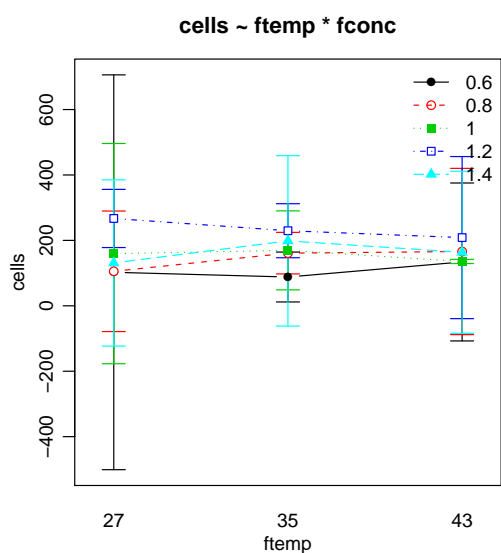
```

Response: cells
      Df Sum Sq Mean Sq F value    Pr(>F)
ftemp    2   1313     656  0.8557  0.44473
fconc    4  51596    12899 16.8154 2.041e-05
ftemp:fconc  8  14703     1838  2.3958  0.06886
Residuals 15  11507     767
Total    29  79118

```

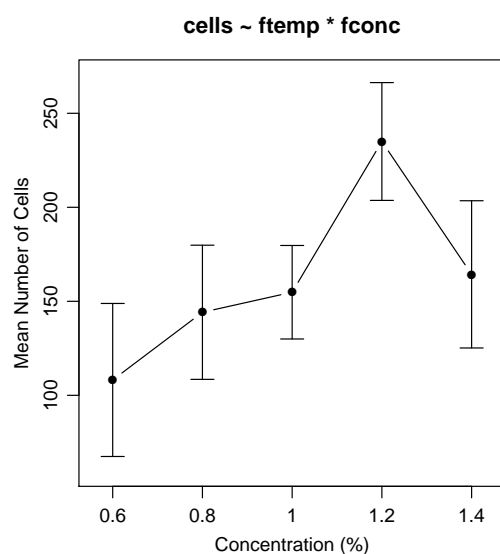
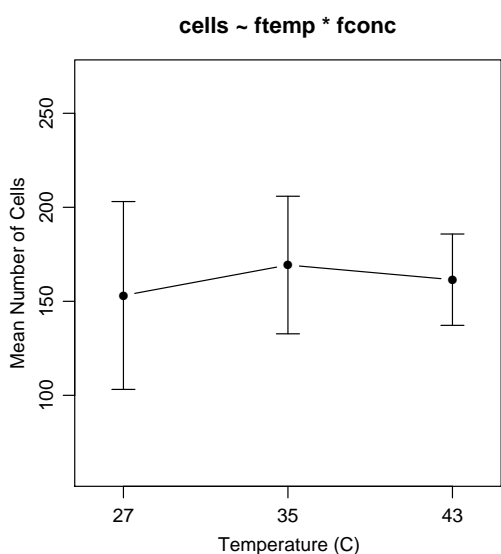
Examine the interaction plots

```
> fit.plot(lm1)
> fit.plot(lm1,interval=FALSE,change.order=TRUE,xlab="Concentration (%)",
+ ylab="Mean Number of Cells",legend="topleft")
```



Examine the main effects plots

```
> fit.plot(lm1,which="ftemp",xlab="Temperature (C)",
+ ylab="Mean Number of Cells",ylim=c(60,270))
> fit.plot(lm1,which="fconc",xlab="Concentration (%)",
+ ylab="Mean Number of Cells",ylim=c(60,270))
```



Examine Tukey's HSD results for concentration main effect and construct a main effects plot with significance letters.

```
> bact.mcl <- glht(lm1, mcp(fconc="Tukey"))  
> summary(bact.mcl)
```

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

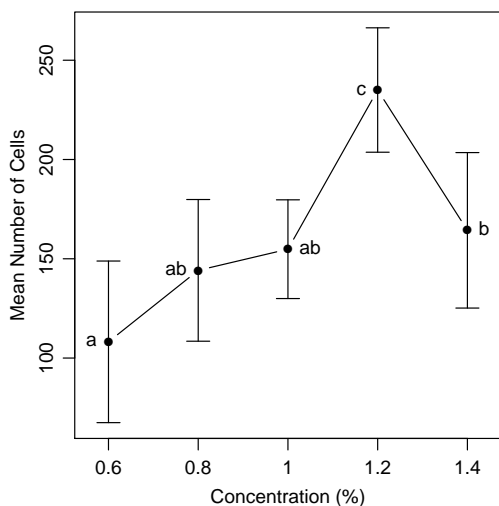
Fit: $\text{lm}(\text{formula} = \text{cells} \sim \text{ftemp} * \text{fconc})$

Linear Hypotheses:

	Estimate	Std. Error	t value	p value
0.8 - 0.6 == 0	3.00	27.70	0.108	0.99996
1 - 0.6 == 0	57.00	27.70	2.058	0.28712
1.2 - 0.6 == 0	164.50	27.70	5.939	< 0.001
1.4 - 0.6 == 0	28.50	27.70	1.029	0.83818
1 - 0.8 == 0	54.00	27.70	1.950	0.33500
1.2 - 0.8 == 0	161.50	27.70	5.831	< 0.001
1.4 - 0.8 == 0	25.50	27.70	0.921	0.88452
1.2 - 1 == 0	107.50	27.70	3.881	0.01093
1.4 - 1 == 0	-28.50	27.70	-1.029	0.83817
1.4 - 1.2 == 0	-136.00	27.70	-4.910	0.00156

(Adjusted p values reported -- single-step method)

```
> fit.plot(lm1, which="fconc", xlab="Concentration (%)", ylab="Mean Number of Cells", main="")  
> add.sig.letters(lm1, which="fconc", lets=c("a", "ab", "ab", "c", "b"), pos=c(2,2,4,2,4))
```



```
> detach(bact)
```

3 Soil Phosphorous Example

You must change the directory to where the following file is located.

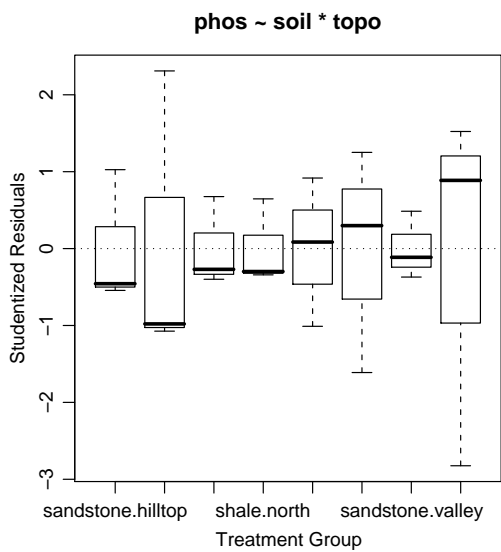
```
> sp <- read.table("SoilPhosphorous.txt",head=TRUE)
> str(sp)

'data.frame':      24 obs. of  3 variables:
 $ soil: Factor w/ 2 levels "sandstone","shale": 2 2 2 2 2 2 2 2 2 ...
 $ topo: Factor w/ 4 levels "hilltop","north",..: 4 4 4 2 2 2 3 3 3 1 ...
 $ phos: int   98 172 185 78 77 100 117 54 96 83 ...

> attach(sp)
> lm1 <- lm(phos~soil*topo)
> levene.test(lm1)
```

```
Levene's Test for Homogeneity of Variance
  Df F value Pr(>F)
group 7  0.3741 0.9043
      16
```

```
> residual.plot(lm1)
```



```
> ad.test(lm1$residuals)
```

Anderson-Darling normality test

```
data: lm1$residuals
A = 0.2126, p-value = 0.8351
```

```
> outlier.test(lm1)
```

```
max|rstudent| = 2.824098, degrees of freedom = 15,
unadjusted p = 0.01282056, Bonferroni p = 0.3076934
```

```
Observation: 1
```

```
> anova(lm1)
```

```
Analysis of Variance Table
```

```
Response: phos
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
soil	1	17876	17876	22.9818	0.0001988
topo	3	9694	3231	4.1542	0.0235128
soil:topo	3	11391	3797	4.8814	0.0134826
Residuals	16	12445	778		
Total	23	51406			

When a two-way ANOVA model has a significant interaction term then multiple comparisons must be computed between each group that can be identified as combinations of the two factors. Unfortunately, this is not a straightforward calculation with the `glht()` function. However, the calculation can be made relatively easily by creating a single factor that consists of the combinations of the two original factors, fitting a one-way ANOVA model to this new single factor, and then submitting this result to the `glht()` function. This process is illustrated below.

```
> detach(sp)
> sp$comb <- sp$soil:sp$topo
> levels(sp$comb)
```

```
[1] "sandstone:hilltop" "sandstone:north" "sandstone:south" "sandstone:valley"
[5] "shale:hilltop" "shale:north" "shale:south" "shale:valley"
```

```
> attach(sp)
> lm1a <- lm(phos~comb)
> anova(lm1a)
```

```
Analysis of Variance Table
```

```
Response: phos
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
comb	7	38961	5566	7.1555	0.0005729
Residuals	16	12445	778		
Total	23	51406			

```
> spint.mc <- glht(lm1a, mcp(comb="Tukey"))
> summary(spint.mc)
```

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

```
Fit: lm(formula = phos ~ comb)
```

```
Linear Hypotheses:
```

	Estimate	Std. Error	t value	p value
sandstone:north - sandstone:hilltop == 0	1.667	22.772	0.073	1.00000
sandstone:south - sandstone:hilltop == 0	19.333	22.772	0.849	0.98686
sandstone:valley - sandstone:hilltop == 0	-4.000	22.772	-0.176	1.00000
shale:hilltop - sandstone:hilltop == 0	4.667	22.772	0.205	1.00000
shale:north - sandstone:hilltop == 0	53.333	22.772	2.342	0.33048
shale:south - sandstone:hilltop == 0	57.333	22.772	2.518	0.25567

```

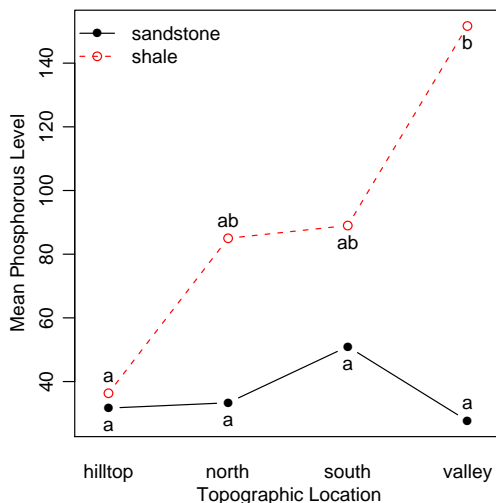
shale:valley - sandstone:hilltop == 0    120.000    22.772    5.270 0.00151
sandstone:south - sandstone:north == 0    17.667    22.772    0.776 0.99219
sandstone:valley - sandstone:north == 0   -5.667    22.772   -0.249 1.00000
shale:hilltop - sandstone:north == 0      3.000    22.772    0.132 1.00000
shale:north - sandstone:north == 0       51.667    22.772    2.269 0.36525
shale:south - sandstone:north == 0       55.667    22.772    2.445 0.28450
shale:valley - sandstone:north == 0     118.333    22.772    5.196 0.00182
sandstone:valley - sandstone:south == 0  -23.333    22.772   -1.025 0.96344
shale:hilltop - sandstone:south == 0    -14.667    22.772   -0.644 0.99746
shale:north - sandstone:south == 0       34.000    22.772    1.493 0.80050
shale:south - sandstone:south == 0       38.000    22.772    1.669 0.70561
shale:valley - sandstone:south == 0     100.667    22.772    4.421 0.00797
shale:hilltop - sandstone:valley == 0     8.667    22.772    0.381 0.99992
shale:north - sandstone:valley == 0      57.333    22.772    2.518 0.25547
shale:south - sandstone:valley == 0      61.333    22.772    2.693 0.19389
shale:valley - sandstone:valley == 0    124.000    22.772    5.445 0.00116
shale:north - shale:hilltop == 0         48.667    22.772    2.137 0.43419
shale:south - shale:hilltop == 0        52.667    22.772    2.313 0.34405
shale:valley - shale:hilltop == 0       115.333    22.772    5.065 0.00229
shale:south - shale:north == 0           4.000    22.772    0.176 1.00000
shale:valley - shale:north == 0         66.667    22.772    2.928 0.13153
shale:valley - shale:south == 0         62.667    22.772    2.752 0.17608
(Adjusted p values reported -- single-step method)

```

```

> fit.plot(lm1,change.order=TRUE,interval=FALSE,main="",ylab="Mean Phosphorous Level",
+ xlab="Topographic Location",legend="topleft")
> add.sig.letters(lm1,change.order=T,lets=c("a","a","a","ab","a","ab","a","b"),
+ pos=c(1,3,1,3,1,1,3,1))

```



```

> detach(sp)

```