STAT 535 H/W 2

Jason Rich

February 23, 2015

Problem 3.10

a.) model and assuptions

$Yij = \mu + \tau_i + e_{ij}$

e ~ iid N(0, σ^2)

i = 1,2,3,...,a

j = 1,2,3,...,n

$\sum_{i=1}^{a} \tau_i = 0$

a.)

```
## ANOVA Table for Cotton Strength

##             Df Sum Sq Mean Sq F value   Pr(>F)
## weight       4  475.8  118.94   14.76 9.13e-06 ***
## Residuals   20  161.2    8.06
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Test of Hypothosis

H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$

H_1: at least one $\mu_i \neq 0$

```
## grand mean for cotton strength

## [1] 15.04
```

b.) test:

H_0: $\frac{\mu_2+\mu_5}{2} = \frac{\mu_2+\mu_3+\mu_4}{3}$

H_1: $\frac{\mu_2+\mu_5}{2} \neq \frac{\mu_2+\mu_3+\mu_4}{3}$

at the $\alpha = 0.05$ level

$\frac{(9.8+10.8)}{2} = \frac{(15.4+17.6+21.6)}{3}$

with $c_1 = c_5 = \frac{1}{2}$

and $c_2 = c_3 = c_4 = \frac{1}{3}$

$$t_0 = \frac{\sum_{i=1}^{a} c_i \bar{Y}_i}{\sqrt{\frac{MSE}{n} \sum_{i=1}^{a} c_i^2}}$$

```
##   15   20   25   30   35
##  9.8 15.4 17.6 21.6 10.8
```

test statistic

```
## [1] 22.44721
```

t-value

```
## [1] 2.085963
```

our test statistic in absolute value is greater than our t-value. Thus I would reject the null hypothesis

c.)

confidence intervals for the above contrast

```
## [1] 51.70313
```

```
## [1] 62.29687
```

d.)

total number of meaningful contrast

```
## Warning: package 'car' was built under R version 3.1.2
```

```
##   2 3 4 5
## 1 0 0 0 0
## 2 1 0 0 0
## 3 0 1 0 0
## 4 0 0 1 0
## 5 0 0 0 1
```

```
##    15-35 20-35 25-35 30-35
## 15     1     0     0     0
## 20     0     1     0     0
## 25     0     0     1     0
## 30     0     0     0     1
## 35    -1    -1    -1    -1
```

e.) maximum set of mutually orthogonal contrast; max number of orthogonal contrast is *p-1* where *p* is the total number of treatment levels

```
##              .L         .Q            .C         ^4
## [1,] -0.6324555  0.5345225 -3.162278e-01  0.1195229
## [2,] -0.3162278 -0.2672612  6.324555e-01 -0.4780914
## [3,]  0.0000000 -0.5345225 -4.095972e-16  0.7171372
## [4,]  0.3162278 -0.2672612 -6.324555e-01 -0.4780914
## [5,]  0.6324555  0.5345225  3.162278e-01  0.1195229
```

f.)

```
## Warning: package 'agricolae' was built under R version 3.1.2
```

```
##
## Study: aov.2 ~ "weight"
##
## Scheffe Test for strength
##
## Mean Square Error  : 8.06
##
## weight,  means
##
##    strength      std r Min Max
## 15      9.8 3.346640 5   7  15
## 20     15.4 3.130495 5  12  18
## 25     17.6 2.073644 5  14  19
## 30     21.6 2.607681 5  19  25
## 35     10.8 2.863564 5   7  15
##
## alpha: 0.05 ; Df Error: 20
## Critical Value of F: 2.866081
##
## Minimum Significant Difference: 6.079555
##
## Means with the same letter are not significantly different.
##
## Groups, Treatments and means
## a     30      21.6
## ab    25      17.6
## bc    20      15.4
## c     35      10.8
## c     15      9.8
```

g.)

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = strength ~ weight, data = c1)
```

```
##
## $weight
##        diff         lwr        upr     p adj
## 20-15   5.6   0.2270417 10.9729583 0.0385024
## 25-15   7.8   2.4270417 13.1729583 0.0025948
## 30-15  11.8   6.4270417 17.1729583 0.0000190
## 35-15   1.0  -4.3729583  6.3729583 0.9797709
## 25-20   2.2  -3.1729583  7.5729583 0.7372438
## 30-20   6.2   0.8270417 11.5729583 0.0188936
## 35-20  -4.6  -9.9729583  0.7729583 0.1162970
## 30-25   4.0  -1.3729583  9.3729583 0.2101089
## 35-25  -6.8 -12.1729583 -1.4270417 0.0090646
## 35-30 -10.8 -16.1729583 -5.4270417 0.0000624
```

h.) Bonferroni

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  c1$strength and c1$weight
##
##    15      20     25     30
## 20 0.0541  -      -      -
## 25 0.0031  1.0000 -      -
## 30 2.1e-05 0.0251 0.3754 -
## 35 1.0000  0.1859 0.0116 7.0e-05
##
## P value adjustment method: bonferroni
```

Tukey

```
##   Tukey multiple comparisons of means
##     90% family-wise confidence level
##
## Fit: aov(formula = strength ~ weight, data = c1)
##
## $weight
##        diff         lwr        upr     p adj
## 20-15   5.6   0.8560925 10.3439075 0.0385024
## 25-15   7.8   3.0560925 12.5439075 0.0025948
## 30-15  11.8   7.0560925 16.5439075 0.0000190
## 35-15   1.0  -3.7439075  5.7439075 0.9797709
## 25-20   2.2  -2.5439075  6.9439075 0.7372438
## 30-20   6.2   1.4560925 10.9439075 0.0188936
## 35-20  -4.6  -9.3439075  0.1439075 0.1162970
## 30-25   4.0  -0.7439075  8.7439075 0.2101089
## 35-25  -6.8 -11.5439075 -2.0560925 0.0090646
## 35-30 -10.8 -15.5439075 -6.0560925 0.0000624
```

Scheffe

```
##
## Study: aov.2 ~ "weight"
##
## Scheffe Test for strength
##
## Mean Square Error  : 8.06
##
## weight,  means
##
##    strength      std r Min Max
## 15      9.8 3.346640 5   7  15
## 20     15.4 3.130495 5  12  18
## 25     17.6 2.073644 5  14  19
## 30     21.6 2.607681 5  19  25
## 35     10.8 2.863564 5   7  15
##
## alpha: 0.1 ; Df Error: 20
## Critical Value of F: 2.248934
##
## Minimum Significant Difference: 5.385374
##
## Means with the same letter are not significantly different.
##
## Groups, Treatments and means
## a     30      21.6
## ab    25      17.6
## bc    20      15.4
## cd    35      10.8
## d     15      9.8
```

Scheffe's test is the most conservative of the three iterations.

i.)

I would recommend to test all possible contrast, not just those annotated within the problem. Furthermore, I would consider sticking with the Scheffe method of comparing multiple mean contrast, because it offers the smallest, and those the most conservative confidence interval.

Problem 9

```
######################################################################
#Reading the Data
######################################################################
## This data contians three variables:
#1.) site: factor with eight levels
#2.) parcel within each site: factor with 4 levels
#3.) ears: response variable measuring the numbner of ears grown, given the two factors
#source of variation derive from parcel and site
```

```r
#d1 <- read.table("corn.txt", header=TRUE)
#d1
attach(d1)

#ensure the factors are really factors
d1$site <- as.factor(d1$site)
d1$parcel <- as.factor(d1$parcel)

#verifying the data structure
str(d1)
```

```
## 'data.frame':    32 obs. of  3 variables:
##  $ site  : Factor w/ 8 levels "DBAN","LFAN",..: 1 2 3 4 5 6 7 8 1 2 ...
##  $ parcel: Factor w/ 4 levels "I","II","III",..: 1 1 1 1 1 1 1 1 2 2 ...
##  $ ears  : num  43.5 40.5 20 42.5 31.5 32.5 43.5 50 46 46.5 ...
```

a.) model and assumptions

$Y_{ij} = \mu + \tau_i + e_{ij}$

e ~ iid N(0, σ^2)

i = 1,2,3,...,a

j = 1,2,3,...,n

$\sum_{i=1}^{a} \tau_i = 0$

```
## grand mean for ears of corn
```

```
## [1] 41.21875
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## site         7 1390.3  198.62   12.54 1.16e-06 ***
## Residuals   24  380.1   15.84
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Call:  glm(formula = ears ~ site)
##
## Coefficients:
## (Intercept)     siteLFAN     siteNSAN     siteORAN     siteOVAN
##      44.375       -1.125      -17.875       -1.125       -4.875
##    siteTEAN     siteWEAN     siteWLAN
##      -6.625        0.250        6.125
##
```

```
## Degrees of Freedom: 31 Total (i.e. Null);  24 Residual
## Null Deviance:       1770
## Residual Deviance: 380.1     AIC: 188
```

b.) teatment effects

```
## Tables of effects
##
##  site
## site
##    DBAN    LFAN    NSAN    ORAN    OVAN    TEAN    WEAN    WLAN
##   3.156   2.031 -14.719   2.031  -1.719  -3.469   3.406   9.281
```

c.) Confidence Intervals for the treatment effects of sites WLAN and WEAN

$\bar{Y}_{wean} - \bar{Y}_{wlan} + / - t_{\alpha/2,N-a}\sqrt{\frac{MSE_p}{n}}$

where MSE_p is the pooled variance for sites WLAN and WEAN

```
##   DBAN   LFAN   NSAN   ORAN   OVAN   TEAN   WEAN   WLAN
## 44.375 43.250 26.500 43.250 39.500 37.750 44.625 50.500
```

```
#vector of response elements per site
wean.1 <- c(43.5,43.5,45.5,46.0)
wlan.1 <- c(50.0,56.0,50.5,45.5)

#site parameters
mean(wean.1)
```

```
## [1] 44.625
```

```
var(wean.1)
```

```
## [1] 1.729167
```

```
length(wean.1)
```

```
## [1] 4
```

```
mean(wlan.1)
```

```
## [1] 50.5
```

```r
var(wlan.1)
```

```
## [1] 18.5
```

```r
length(wlan.1)
```

```
## [1] 4
```

```r
#pooled variance
pooled.var <- ((4-1)*1.73 + (4-1)*18.50)/(6)
pooled.var
```

```
## [1] 10.115
```

```r
#pooled standard deviation
pooled.sd <- sqrt(pooled.var)
pooled.sd
```

```
## [1] 3.180409
```

```r
#t-values: N = 32 a = 8
t.crit <- c(-1,1)*qt(.975,24)
t.crit
```

```
## [1] -2.063899  2.063899
```

```r
conf.inv <- as.numeric(50.50-44.63)+c(-1,1)*qt(.975,24)*sqrt(pooled.var/8)
conf.inv
```

```
## [1] 3.549261 8.190739
```

d.) adding parcel

$Y_{ijk} = \mu + \tau_i + \beta_j + e_{ijk}$

e ~ iid N(0, σ^2)

i = 1,2,3,. . . ,a

j = 1,2,3,. . . ,b

k = 1,2,3,. . . ,n

$\sum_{i=1}^{a} \tau_i = 0,\ \sum_{i=1}^{a} \beta_j = 0$

Where the β_j are he added blocks to reduce the overall variance of the model

```r
#anova model with sources site, parcels,  and total
corn.aov.2 <- aov(ears ~ site + parcel, data= d1)
summary(corn.aov.2)
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)    
## site          7 1390.3  198.62   17.27 2.12e-07 ***
## parcel        3  138.7   46.22    4.02   0.0209 *  
## Residuals    21  241.5   11.50                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#glm summary and coeffieicent
glm(ears ~ site + parcel, data = d1)
```

```
## 
## Call:  glm(formula = ears ~ site + parcel, data = d1)
## 
## Coefficients:
## (Intercept)     siteLFAN     siteNSAN     siteORAN     siteOVAN  
##      41.156       -1.125      -17.875       -1.125       -4.875  
##    siteTEAN     siteWEAN     siteWLAN     parcelII    parcelIII  
##      -6.625        0.250        6.125        5.813        3.375  
##    parcelIV  
##       3.688  
## 
## Degrees of Freedom: 31 Total (i.e. Null);  21 Residual
## Null Deviance:       1770 
## Residual Deviance: 241.5     AIC: 179.5
```

e.)

The variance for the difference of two parcels is 11.50 with the inclusion of the parcel block variable. This is, apposed to before the addition of parcels, which was calculated at 15.84. This is intutive, because adding a block to the equation, spreads the variation across more variables, and those reduces the total variation concentrated in the residuals.