Anthony Cunningham

STAT 3200

Due Fri, April 7

**Homework 7**

#1. > barley.data <- read.csv(file="http://www.stat.uiowa.edu/~ernli/ALRdata/HW7barley.csv",

+ head=TRUE,sep=",")

> attach(barley.data)

A. 1)> is.factor(Water) \* R treats both Water and Timing as numerical

[1] FALSE values, not factors.

> is.factor(Timing)

[1] FALSE

> Water = as.factor(Water)

> Timing = as.factor(Timing)

> is.factor(Water) \* R now treats Water and Timing as factors.

[1] TRUE

> is.factor(Timing)

[1] TRUE

2) > unique(Water) \* There are 2(5) = 10 possible combinations of

[1] 4 6 Water and Timing levels.

Levels: 4 6

> unique(Timing)

[1] 1 3 6 9 12

Levels: 1 3 6 9 12

> table(Water, Timing) \* This is a balanced Two-Way ANOVA, as

Timing the number of observations in each possible

Water 1 3 6 9 12 combination of Water and Timing are the same.

4 3 3 3 3 3

6 3 3 3 3 3

> mean(Germination) \* The grand sample mean germination rate is

[1] 14.2 14.2%.

> tapply(Germination, list(Water, Timing), mean)

1 3 6 9 12

4 8.666667 13.333333 21.000000 25.333333 34

6 4.666667 3.666667 7.666667 6.666667 17

3) > par(mfrow = c(2,1))

> interaction.plot(Water, Timing, Germination)

> interaction.plot(Timing, Water, Germination)

\* There does not appear to be significant interaction between Water and Timing groups, as the group means move in the same general direction, implying that the “lines” are parallel to a decent degree.



B. > contrasts(Water)=contr.sum(levels(Water))

> contrasts(Timing)=contr.sum(levels(Timing))

> contrasts(Water)

[,1] \*Dummy variable coding for Water group w/ Sum-to-0 constraint.

4 1

6 -1

> contrasts(Timing)

[,1] [,2] [,3] [,4] \*Dummy variable coding for Timing group w/ Sum-to-0 constraint.

1 1 0 0 0

3 0 1 0 0

6 0 0 1 0

9 0 0 0 1

12 -1 -1 -1 -1

C. 1) > fit = lm(Germination ~ Water + Timing + Water:Timing)

> summary(fit)

Call:

lm(formula = Germination ~ Water + Timing + Water:Timing)

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 14.2000 1.4111 10.063 2.84e-09 \*\*\*

Water1 6.2667 1.4111 4.441 0.000251 \*\*\*

Timing1 -7.5333 2.8221 -2.669 0.014731 \*

Timing2 -5.7000 2.8221 -2.020 0.057011 .

Timing3 0.1333 2.8221 0.047 0.962786

Timing4 1.8000 2.8221 0.638 0.530829

Water1:Timing1 -4.2667 2.8221 -1.512 0.146211

Water1:Timing2 -1.4333 2.8221 -0.508 0.617086

Water1:Timing3 0.4000 2.8221 0.142 0.888706

Water1:Timing4 3.0667 2.8221 1.087 0.290116

Residual standard error: 7.729 on 20 degrees of freedom

Multiple R-squared: 0.6939, Adjusted R-squared: 0.5561

F-statistic: 5.037 on 9 and 20 DF, p-value: 0.001269

2) grand mean = µ-hat = 14.2

µ-hat11 = µ-hat + Water1 + Timing1 + Water1:Timing1 = 14.2 + 6.2667 – 7.5333 – 4.2667 = 8.6667

µ-hat15 = µ-hat + Water1 – Timing1 – Timing2 – Timing3 – Timing4 – Water1:Timing1 – Water1:Timing2 – Water1:Timing3 – Water1:Timing4 = 14.2 + 6.2667 – (-7.5333) – (-5.7) – 0.1333 – 1.8 – (-4.2667) – (-1.4333) – 0.4 – 3.0667 = 34

µ-hat23 = µ-hat – Water1 + Timing3 – Water1:Timing3 = 14.2 – 6.2667 + 0.1333 – 0.4 = 7.6667

µ-hat25 = µ-hat – Water1 – Timing1 – Timing2 – Timing3 – Timing4 + Water1:Timing1 + Water1:Timing2 + Water1:Timing3 + Water1:Timing4 = 14.2 – 6.2667 – (-7.5333) – (-5.7) – 0.1333 – 1.8 – 4.2667 – 1.4333 + 0.4 + 3.0667 = 17

\*Yes, these results are consistent with the grand and individual sample means obtained.

3) \* 69.39% of variability in Germination is explained by Water, Timing, and their interaction.

4) F\* = 5.037 df = 9(num), 20(den) p-value = 0.001269

\* We reject H0 due to significantly small p-value. Therefore, we have evidence that at least one of Water, Timing, or their interaction are significant predictors for barley’s Germination rate.

5) > library(car)

> Anova(fit, type="III")

Anova Table (Type III tests)

Response: Germination

Sum Sq Df F value Pr(>F)

Water:Timing 208.9 4 0.8742 0.496726 \*F\*=(SS/numdf)/(RSS/dendf)

Residuals 1194.7 20 = (208.9/4)/(1194.7/20) = 0.8742

> anova(fit)

Analysis of Variance Table

Response: Germination

Df Sum Sq Mean Sq F value Pr(>F) \*F\*=(SS/numdf)/(RSS/dendf)

Water:Timing 4 208.87 52.22 0.8742 0.496726 = (208.9/4)/(1194.7/20) = 0.8742

Residuals 20 1194.67 59.73

\*Yes, F test stat is the same for both methods. F\* = 0.8742 df = 4(num), 20(den)

p-value = 0.496726

\* We fail to reject H0. Therefore, we do not have evidence that the interaction between groups Water and Timing is a significant predictor for the Germination rate of barley.

D. 1) > fit.me = lm(Germination ~ Water + Timing)

> summary(fit.me)

Call:

lm(formula = Germination ~ Water + Timing)

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 14.2000 1.3962 10.171 3.53e-10 \*\*\*

Water1 6.2667 1.3962 4.488 0.000153 \*\*\*

Timing1 -7.5333 2.7924 -2.698 0.012570 \*

Timing2 -5.7000 2.7924 -2.041 0.052367 .

Timing3 0.1333 2.7924 0.048 0.962311

Timing4 1.8000 2.7924 0.645 0.525294

Residual standard error: 7.647 on 24 degrees of freedom

Multiple R-squared: 0.6404, Adjusted R-squared: 0.5655

F-statistic: 8.547 on 5 and 24 DF, p-value: 9.27e-05

2) µ-hat11 = µ-hat + Water1 + Timing1 = 14.2 + 6.2667 – 7.5333 = 12.9334

µ-hat15 = µ-hat + Water1 – Timing1 – Timing2 – Timing3 – Timing4

= 14.2 + 6.2667 – (-7.5333) – (-5.7) – 0.1333 – 1.8 = 31.7667

µ-hat23 = µ-hat – Water1 + Timing3 = 14.2 – 6.2667 + 0.1333 = 8.0666

µ-hat25 = µ-hat – Water1 –Timing1 – Timing2 – Timing3 – Timing4

= 14.2 – 6.2667 – (-7.5333) – (-5.7) – 0.1333 – 1.8 = 19.2333

\* These results are not consistent with the individual sample means found earlier, as the interaction terms were not included in this model.

3) 64.04% of variability in Germination can be explained by Water and Timing factors.

4) From Overall F-test of Main Effects Model, we reject H0. So, we have evidence that either Water, Timing, or both factors are significant predictors for the Germination rate of Barley.

5) > Anova(fit.me, type="III")

Anova Table (Type III tests)

Response: Germination

Sum Sq Df F value Pr(>F)

(Intercept) 6049.2 1 103.4395 3.529e-10 \*\*\*

Water 1178.1 1 20.1457 0.0001525 \*\*\*

Timing 1321.1 4 5.6477 0.0023801 \*\*

Residuals 1403.5 24

(Water): F\* = 20.1457 df = 1(num), 24(den) p-value = 0.0001525

\* We reject H0. So, we have evidence that, after accounting for the effect of Timing on Germination, Water is a significant predictor of Germination.

(Timing): F\* = 5.6477 df = 4(num), 24(den) p-value = 0.0023801

\* We reject H0. So, we have evidence that, after accounting for the effect of Water on Germination, Timing is a significant predictor of Germination.

6) > par(mfrow=c(2,1))

> plot(fit.me$residuals)

> abline(h=0)

> qqnorm(fit.me$residuals)

> qqline(fit.me$residuals)



\* There does not appear to be any major violations in normality. However, observations appear to get more variant as the group means increase, so the constant variance assumption may be violated.

#2. A. > newspaper <- read.csv(file="http://www.stat.uiowa.edu/~ernli/ALRdata/HW7newspaper.csv",

+ head=TRUE,sep=",")

> attach(newspaper)

> levels(Day)

[1] "Friday" "Monday" "Thursday" "Tuesday" "Wednesday"

> levels(Section)

[1] "Business" "News"

> Day = factor(Day, levels=c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday"))

> levels(Day)

[1] "Monday" "Tuesday" "Wednesday" "Thursday" "Friday"

B. > interaction.plot(Day, Section, Inquiries)



\*There does appear to be significant interaction, since the individual “lines” do not appear to be parallel at all, as they cross at multiple points.

C. > contrasts(Day) = contr.sum(levels(Day))

> contrasts(Section) = contr.sum(levels(Section))

> fit.interact = lm(Inquiries~Day + Section + Day:Section)

> summary(fit.interact)

Call:

lm(formula = Inquiries ~ Day + Section + Day:Section)

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 9.0000 0.2151 41.849 < 2e-16 \*\*\*

Day1 0.8750 0.4301 2.034 0.050841 .

Day2 0.5000 0.4301 1.162 0.254206

Day3 -0.1250 0.4301 -0.291 0.773342

Day4 -3.0000 0.4301 -6.975 9.5e-08 \*\*\*

Section1 0.1000 0.2151 0.465 0.645296

Day1:Section1 1.5250 0.4301 3.546 0.001308 \*\*

Day2:Section1 -0.8500 0.4301 -1.976 0.057389 .

Day3:Section1 -0.4750 0.4301 -1.104 0.278219

Day4:Section1 1.6500 0.4301 3.836 0.000598 \*\*\*

Residual standard error: 1.36 on 30 degrees of freedom

Multiple R-squared: 0.7648, Adjusted R-squared: 0.6943

F-statistic: 10.84 on 9 and 30 DF, p-value: 2.841e-07

D. > Anova(fit.interact, type="III")

Anova Table (Type III tests)

Response: Inquiries

Sum Sq Df F value Pr(>F)

Day:Section 75.4 4 10.1824 2.493e-05 \*\*\*

Residuals 55.5 30

\* Yes, the interaction term is significant after the main effects are accounted for due to significantly small p-value.

E. > levels(Day:Section)

[1] "Monday:Business" "Monday:News" "Tuesday:Business"

[4] "Tuesday:News" "Wednesday:Business" "Wednesday:News"

[7] "Thursday:Business" "Thursday:News" "Friday:Business"

[10] "Friday:News"

> grouping = Day:Section

> contrasts(grouping) = contr.sum(levels(grouping))

> contrasts(grouping)

[,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9]

Monday:Business 1 0 0 0 0 0 0 0 0

Monday:News 0 1 0 0 0 0 0 0 0

Tuesday:Business 0 0 1 0 0 0 0 0 0

Tuesday:News 0 0 0 1 0 0 0 0 0

Wednesday:Business 0 0 0 0 1 0 0 0 0

Wednesday:News 0 0 0 0 0 1 0 0 0

Thursday:Business 0 0 0 0 0 0 1 0 0

Thursday:News 0 0 0 0 0 0 0 1 0

Friday:Business 0 0 0 0 0 0 0 0 1

Friday:News -1 -1 -1 -1 -1 -1 -1 -1 -1

> lm1 = lm(Inquiries~grouping)

> lm1$coefficients

(Intercept) grouping1 grouping2 grouping3 grouping4

9.000000e+00 2.500000e+00 -7.500000e-01 -2.500000e-01 1.250000e+00

grouping5 grouping6 grouping7 grouping8 grouping9

-5.000000e-01 2.500000e-01 -1.250000e+00 -4.750000e+00 -4.739625e-16

1) H0: Grouping5 – Grouping6 = 0 Ha: Grouping5 – Grouping6 ≠ 0

> linearHypothesis(lm1, c(0,0,0,0,1,-1,0,0,0,0), rhs=0)

Hypothesis:

grouping4 - grouping5 = 0

Model 1: restricted model

Model 2: Inquiries ~ grouping

Res.Df RSS Df Sum of Sq F Pr(>F) We fail to reject H0. So, there is no evidence

1 31 61.625 that the mean # of inquiries significantly

2 30 55.500 1 6.125 3.3108 0.07882 . differs between the two sections on the

Wednesday paper.

2) H0: 2xGrouping9 + Grouping1 + …+ Grouping8 = 0

Ha: 2xGrouping9 + Grouping1 + … + Grouping8 ≠ 0

> linearHypothesis(lm1, c(0,1,1,1,1,1,1,1,1,2), rhs=0)

Hypothesis:

grouping1 + grouping2 + grouping3 + grouping4 + grouping5 + grouping6 + grouping7 + grouping8 + 2 grouping9 = 0

Model 1: restricted model

Model 2: Inquiries ~ grouping

\*We reject H0. So, we have evidence that

Res.Df RSS Df Sum of Sq F Pr(>F) the mean # of inquiries significantly differs

1 31 80.0 between Sections on the Friday newspaper.

2 30 55.5 1 24.5 13.243 0.001019 \*\*