Stat 5309 Final Exam

Tom Wilson

May 7th, 2019

Problem 1:

The effect of 4 types of graphite coater types on light box reading are to be studied. The readings might differ from day to day. Here assume we have a fixed effect model. Observations are taken for 3 days on the four types. The results are

```
days<-c('d1','d2','d3')
types<-c('t1','t2','t3','t4')
graphite_data <- expand.grid(day=days,type=types)
graphite_data <- cbind(graphite_data,light=c(4,4.8,4,
                                             4.8,5,4.8,
                                             5,5.2,5.6,
                                             4.6,4.6,5))
graphite_data %>% kable()
```

day	type	light
d1	t1	4.0
d2	t1	4.8
d3	t1	4.0
d1	t2	4.8
d2	t2	5.0
d3	t2	4.8
d1	t3	5.0
d2	t3	5.2
d3	t3	5.6
d1	t4	4.6
d2	t4	4.6
d3	t4	5.0

(a)

Note: Day is used as blocking factor

Write the model equation Let y be the response of light, b be Day (the blocking factor), and tau be the type treatment. i and j are the level of day and type respectively. I and J are the number of levels, n is the number of observations. τ_i, β_j are treatment effects and block effects.

$$y_{ij} = \mu + b_i + \tau_j + \epsilon_{ij}$$

State all the assumptions about residuals

This model assumes that the error term is normally distributed with mean of zero and homogenous variance of σ^2 across levels of the treatment and block.

(b)

Use definitions to estimate the model parameters (14 parameters)

The grand mean

$$\mu_{..} = \frac{1}{n}\sum_{i=1}^{4}\sum_{j=1}^{3} y_{ij} = \frac{57.4}{12} = 4.78$$

$$\mu_1 = \frac{1}{I}\sum_{i=1}^{3} y_{i,1} = \frac{12.8}{3} = 4.267$$

$$\mu_2 = \frac{1}{I}\sum_{i=1}^{3} y_{i,2} = \frac{14.6}{3} = 4.867$$

$$\mu_3 = \frac{1}{I}\sum_{i=1}^{3} y_{i,3} = \frac{15.8}{3} = 5.267$$

$$\mu_4 \frac{1}{I}\sum_{i=1}^{3} y_{i,4} = \frac{14.2}{3} = 4.733$$

$$\tau_1 = \mu_1 - \mu_{..} = 4.267 - 4.78 = -0.513$$

$$\tau_2 = \mu_2 - \mu_{..} = 4.867 - 4.78 = 0.087$$

$$\tau_3 = \mu_3 - \mu_{..} = 5.26 - 4.78 = 0.48$$

$$\tau_4 = \mu_3 - \mu_{..} = 4.733 - 4.78 = -0.047$$

$$\sigma^2 = \frac{1}{(I-1)(J-1)}\sum_{i=1}^{4}\sum_{j=1}^{3} \epsilon_{ij}^2 = \frac{1}{(I-1)(J-1)}\sum_{i=1}^{4}\sum_{j=1}^{3} (y_{ij} - \mu_i)^2 = \frac{0.54}{6} = 0.09$$

$$\beta_1 = \sum_{j=1}^{4} y_{1j} - \mu_{..} = \frac{18.4}{4} - 4.78 = -0.18$$

$$\beta_2 = \sum_{j=1}^{4} y_{1j} - \mu_{..} = \frac{19.6}{4} - 4.78 = 0.12$$

$$\beta_3 = \sum_{j=1}^{4} y_{1j} - \mu_{..} = \frac{19.4}{4} - 4.78 = 0.07$$

(c)

Test the hypothesis that the effects (ie, mean effect) of the 4 graphite coater types are the same.

Hypothesis: Ho:

$$\tau_1 = \tau_2 = \tau_3 = \tau_4 = 0$$

Ha: one of

$$\tau_1, \tau_2, \tau_3, \tau_4 \neq 0$$

Test statistic F=5.67

P-values = 0.03481

Conclusion: at a confidence level of 95% we would reject the joint null hypothesis and conclude that at least one effect is significantly different from zero.

(d)

How would your analysis of variance be different if the experiment had not been blocked?

If the experiement were not blocked, our best estimate of σ^2 would be a pooled variance which is likely to be much larger.

Write the ANOVA table with the value of the test statistic and p-value.

	degrees of freedom	sum of squares	mean square	F stat	P value
day	3-1=2	0.2067	0.1033	1.1481	0.37827
type	4-1=3	1.53	0.51	5.667	0.035
residual	12-2-3-1=6	0.54	0.09		

Problem 2:

The response time in milliseconds was determined for 3 different types of circuits that could be used in an automatic valve shutoff mechanism. The results from a completely randomized experiment are shown in the following table:

```
circuit_types <- c('c1','c2','c3')
circuit_data <- data.frame(circuit_type=rep(circuit_types,5),
                           response_time=c(9,20,6,
                                           12,21,5,
                                           10,23,8,
                                           8,17,16,
                                           15,30,7))
circuit_data
```

```
##    circuit_type response_time
## 1            c1             9
## 2            c2            20
## 3            c3             6
## 4            c1            12
## 5            c2            21
## 6            c3             5
## 7            c1            10
```

```
## 8              c2             23
## 9              c3              8
## 10             c1              8
## 11             c2             17
## 12             c3             16
## 13             c1             15
## 14             c2             30
## 15             c3              7
```

(a)

Write the model equation:

$$y = \mu + \tau_i + \epsilon_i$$

State all the assumptions about residuals

This model assumes that the error term is normally distributed with mean of zero and homogenous variance of σ^2 across levels of the treatment. μ is the grand mean τ_i are treatment effects

(b)

```r
circuit_data %>% group_by(circuit_type) %>% summarise(n=length(response_time),s=sum(response_time),mu=me
```

```
## # A tibble: 3 x 4
##   circuit_type     n     s    mu
##   <fct>        <int> <dbl> <dbl>
## 1 c1               5    54  10.8
## 2 c2               5   111  22.2
## 3 c3               5    42   8.4
```

Use definitions and R to find the estimates for the following parameters:

$$\mu = \frac{1}{n}\sum_{i=1}^{3}\sum_{r=1}^{5} y_{ir} = \frac{207}{15} = 13.8$$

$$\mu_1 = \frac{1}{R}\sum_{r=1}^{5} y_{1,r} = \frac{54}{5} = 10.8$$

$$\mu_2 = \frac{1}{R}\sum_{r=1}^{5} y_{2,r} = \frac{111}{5} = 22.2$$

$$\mu_3 = \frac{1}{R}\sum_{r=1}^{5} y_{1,r} = \frac{42}{5} = 8.4$$

$$\tau_1 = \mu_1 - \mu = 10.8 - 13.8 = -3$$

$$\tau_2 = \mu_2 - \mu = 22.2 - 13.8 = 8.4$$

$$\tau_3 = \mu_3 - \mu = 8.4 - 13.8 = -5.4$$

$$\sigma^2 = \frac{1}{(n-I)}\sum_{i=1}^{3}\sum_{r=1}^{5} \epsilon_{ir}^2 = \frac{1}{(n-I)}\sum_{i=1}^{3}\sum_{r=1}^{5} (y_{ir} - \mu_i)^2 = \frac{202.8}{12} = 16.9$$

```r
circuit_model <- aov(formula = response_time~circuit_type,data=circuit_data)
anova(circuit_model)
```

```
## Analysis of Variance Table
##
## Response: response_time
##              Df Sum Sq Mean Sq F value    Pr(>F)
## circuit_type  2  543.6   271.8  16.083 0.0004023 ***
## Residuals    12  202.8    16.9
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(c)

Perform a Hypothesis Testing that the hypothesis to be tested for response time being equal for different circuit types. Hypothesis.: Ho:

$$\tau_1 = \tau_2 = \tau_3 = 0$$

Ha: one of

$$\tau_1, \tau_2, \tau_3 \neq 0$$

Test statistics F = 16.083

P-value = 0.0004

Decision: At a confidence of 95%, we reject the null hypothesis and conclude that at least one effect is not zero.

(d)

Identify the pairs of treatment means which are different using Tukey's tests with Significant level.

circuit types 2 and 1 are different

circuit types 3 and 2 are different

circuit types 3 and 1 are similar

```r
TukeyHSD(circuit_model)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = response_time ~ circuit_type, data = circuit_data)
##
## $circuit_type
##        diff        lwr       upr     p adj
## c2-c1  11.4   4.463555 18.336445 0.0023656
## c3-c1  -2.4  -9.336445  4.536445 0.6367043
## c3-c2 -13.8 -20.736445 -6.863555 0.0005042
```

(e)

Suppose we want to test for the significant difference of

$\frac{\mu_1+\mu_3}{2} = \mu_2$

Perform the appropriate test for the contrast.

Write Contrast C: 1,-2,1

Find the estimate : c = -25.2 se(c) = 4.5

```
fit.contrast(circuit_model,'circuit_type',coeff = c(1,-2,1))
```

```
##                              Estimate Std. Error   t value     Pr(>|t|)
## circuit_type c=( 1 -2 1 )     -25.2   4.503332 -5.595856 0.0001169069
## attr(,"class")
## [1] "fit_contrast"
```

Hypothesis : Ho: μ_2 is the average of μ_2 and μ_3; Ha: μ_2 <> average of μ_2 and μ_3

Test statistics -5.6 P-value 0.00012

Conclusion: Reject the null hypothesis and conclude that μ_2 is not the average of μ_1 and μ_3

Problem 3

A study is conducted to compare 4 menus in terms of numbers of calories. The 4 menus are: A- No calories B- Calories C-Rank-ordered Calories D-Color-ordered Calories

Suppose n=20 per each menu. The sample means and estimated variance are

$y_1 = 17.6$ $y_2 = 16.8$ $y_3 = 16$ $y_4 = 14.4$ $y_{\cdot\cdot} = 16.2$ \$sˆ2 = MSE = 196 \$

(a)

Complete the following table to test the $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$

```
df <- 79 - 3
mse <- 196
sserror <- mse * df
sstotal <- (17.6)^2*20 + (16.8)^2*20 + (16)^2*20 + (14.4)^2*20 - 16.2^2/80
sstrt <- sstotal - sserror
```

```
1 - pf(10.55,3,76)
```

```
## [1] 7.046401e-06
```

source	SS	df	MS	F	P-value
trt	6207.919	3	2069.3	10.55	7.05e-6
error	14896	76	196		
total	21103.9	79			

Reject the null

(b)

Difference between 2 means: $y_1 - y_2$

$$SE = \sqrt{\frac{\sigma^2}{n}} = \sqrt{\frac{196}{80}} = 1.565$$

Compute the standard error : 1.565

```
t <- qt(1-0.025,40)
t
```

```
## [1] 2.021075
```

```
2.02*1.565*0.3162
```

```
## [1] 0.9996031
```

$$LSD = t * SE * \sqrt{\frac{1}{20} + \frac{1}{20}} = 2.02 * 1.565 * 0.3162 = 0.9996$$

Compute the Fisher LSD 0.9996

Write a 95%-CI for the difference, using Fisher LSD

$ -0.9996 < y_1 - y_2 < 0.9996 $

(c)

Contrast C: Write the contrast C which compares Menu 1 vs Menus {1,2,3} combined.

$$\mu_1 = \frac{\mu_1 + \mu_2 + \mu_3}{3}$$

$$3\mu_1 = \mu_1 + \mu_2 + \mu_3$$

$$0 = \mu_1 - 3\mu_1 + \mu_2 + \mu_3$$

(-2,1,1,0)

Calculate an estimate of C : c=-2.4

```
-2*17.6+16.8+16.0
```

```
## [1] -2.4
```

Calculate the standard error of the contrast: se(c) =3.13

```
sqrt(4*1.565^2)
```

```
## [1] 3.13
```

Write a 95%-CI for C, using t-distribution :(-6.26,6.26)

```
qt(1-0.025,60)*3.13
```

```
## [1] 6.260932
```

Problem 4

A router is used to cut locating notches on a printed circuit board. The vibration level at the surface of the board as it is cut is considered to be a major source of dimensional variation in the notches. Two factors are thought to influence vibration: bit size (A) and cutting speed (B). Two sizes (1/16 and 1/8 inch) two speeds (40 and 90 rpm) are selected and four boards are cut at each set of conditions shown below. The response variable is vibration measuerd as the resultant vecor of three accelerometers (x,y,z) on each test circuit board.

```
bit_sizes <- c(0,1)
speeds <- c(0,1)
vibration_data <- expand.grid(bit_size = bit_sizes,
                              speed = rep(speeds,4))
interactions =vibration_data$bit_size * vibration_data$speed
vibration_data <- cbind(vibration_data,
                        interaction = interactions,
                        vibation=c(18.2,27.2,15.9,41.0,
                                   18.9,24.0,14.5,43.9,
                                   12.9,22.4,15.1,36.3,
                                   14.4,22.5,14.2,39.9
                                   ))
vibration_data %>% kable()
```

bit_size	speed	interaction	vibation
0	0	0	18.2
1	0	0	27.2
0	1	0	15.9
1	1	1	41.0
0	0	0	18.9
1	0	0	24.0
0	1	0	14.5
1	1	1	43.9
0	0	0	12.9
1	0	0	22.4
0	1	0	15.1
1	1	1	36.3
0	0	0	14.4
1	0	0	22.5
0	1	0	14.2
1	1	1	39.9

(a)

Calculate, from scratch, effects of A, B, AB[Hint: Chapter 6]

```
n <- 4
one <- (18.2+18.9+12.9+14.4)
a <- (27.2+24.0+22.4+22.5)
b <- (15.9+14.5+15.1+14.2)
ab <- (41.0+43.9+36.3+39.9)
A <- 1/(2*n) * ((ab-b)+(a-one))
B <- 1/(2*n) * ((ab-a)+(b-one))
AB <- 1/(2*n) * ((ab-b)+(a-one))
A
```

```
## [1] 16.6375
```

```
B
```

```
## [1] 7.5375
```

```
AB
```

```
## [1] 16.6375
```

A = 16.6375 B = 7.5375 AB = 16.6

(b)

Calculate the Sum Squares: SS(A), SS(B), SS(AB).

```
SSA  <- (ab+a-b-one)^2/(4*n)
SSB  <- (ab+b-a-one)^2/(4*n)
SSAB <- (ab+one-a-b)^2/(4*n)
SSA
```

```
## [1] 1107.226
```

```
SSB
```

```
## [1] 227.2556
```

```
SSAB
```

```
## [1] 303.6306
```

SSA = 1107.226 SSB = 227.2556 SSAB = 303.6306

(c)

Complete the ANOVA table: df, MS, F-stat, P-values of A, B, AB

```
SST <- sum(vibration_data$vibation^2) - mean(vibration_data$vibation)^2/(4*n)
SSE <- SST - SSA - SSB - SSAB
SST
```

```
## [1] 10761.19
```

```
SSE
```

```
## [1] 9123.083
```

SST = 10761.19 SSE = 9123.083

source	SS	df	MS	F	P-value
A	1107.23	1	1107.23	1.456	<0.01
B	227.26	1	227.26	0.298	0.85
AB	303.63	1	303.63	0.399	0.711
error	9123.083	12	760.26		
total	10761.19	15			

(d)

Use R, find the model matrix X.

```r
model_matrix <- cbind(1,vibration_data[,c('bit_size','speed','interaction')])
X <- model_matrix %>% as.matrix
y <- vibration_data$vibation
X %>% kable()
```

1	bit_size	speed	interaction
1	0	0	0
1	1	0	0
1	0	1	0
1	1	1	1
1	0	0	0
1	1	0	0
1	0	1	0
1	1	1	1
1	0	0	0
1	1	0	0
1	0	1	0
1	1	1	1
1	0	0	0
1	1	0	0
1	0	1	0
1	1	1	1

Use X to calculate the vector of coefficients , for the regression model.

```r
solve(t(X)%*%X) %*% t(X) %*% y
```

```
##               [,1]
## 1           16.100
## bit_size     7.925
## speed       -1.175
## interaction 17.425
```

$B_0 = 16.1$ $B_A = 7.925$ $B_B = -1.175$ $B_AB = 17.425$