Stat 5309 Lab 4b

Tom Wilson

Feb 21, 2019

1.

An industrial engineer is conducting an experiment on eye focus time. He is interested in the effect of the distance of the object from the eye on the focus time. Four different distances are of interest. He has five subjects available for the experiment. Because there may be differences among individuals, he decides to conduct the experiment in a randomized block design. The data obtained follow. Analyze the data from this experiment (use

$$\alpha = 0.05$$

) and draw appropriate conclusions.

a

Set up the data frame named eye, using subject as a blocking factor, distance as treatment factor and time as response.

```
distances <- c("4","6","8","10")
subjects <- c("1","2","3","4","5")
eye <- expand.grid(subject = subjects, distance = distances)
eye <- cbind(eye,time=c(10,6,6,6,6,
                        7,6,6,1,6,
                        5,3,3,2,5,
                        6,4,4,2,3
                        )
             )
eye %>% kable()
```

subject	distance	time
1	4	10
2	4	6
3	4	6
4	4	6
5	4	6
1	6	7
2	6	6
3	6	6
4	6	1
5	6	6
1	8	5
2	8	3
3	8	3
4	8	2
5	8	5
1	10	6
2	10	4
3	10	4

subject	distance	time
4	10	2
5	10	3

b

Build a linear model named eye.mod.

Are the subject means significantly different?

Are the Distance means significantly different?

```
eye_model <- aov(time ~ subject+distance,data = eye)
eye_anova <- anova(eye_model)
eye_anova
```

```
## Analysis of Variance Table
##
## Response: time
##           Df Sum Sq Mean Sq F value   Pr(>F)
## subject    4  36.30   9.075  7.1176 0.003548 **
## distance   3  32.95  10.983  8.6144 0.002543 **
## Residuals 12  15.30   1.275
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Based on F statistics, there is evidence that both subject and distance have significantly different means.

c

Which distances bring the longest/shortest focus time?

```
dist_means <- eye %>%
              group_by(distance) %>%
              summarise(mean=mean(time))
```

distance of 4 has the highest average time. distance of 10 has the lowest average time.

d

Calculate the sample size (number of treatment replicates) for power > 0.90

```
power.anova.test(groups = 4,
                 between.var = var(dist_means$mean),
                 within.var = eye_anova$`Mean Sq`[3],
                 power=0.9
                 )
```

```
##
##      Balanced one-way analysis of variance power calculation
##
##          groups = 4
##               n = 3.886864
##     between.var = 2.196667
```

```
##      within.var = 1.275
##       sig.level = 0.05
##           power = 0.9
##
## NOTE: n is number in each group
```

need 4 replicates to have a power of 0.9

2

The effect of five different ingredients (A,B,C,D,E) on the reaction time of a chemical process is being studied. Each batch of new material is only large enough to permit five runs to be made. Furthermore, each run requires approximately 1.5 hours, so only five runs can be made in one day. The experimenter decides to run the experiment as a latin square so that day and batch effects may be systematically controlled. She obtains data that follow. Analyze the data from this experiment (use

$$\alpha = 0.05$$

) and draw conclusions.

a

set up a dataframe

```
batches <- c("b1","b2","b3","b4","b5")
days <- c("d1","d2","d3","d4","d5")
chemical <- expand.grid(day = days,
                        batch = batches
                        )
chemical <- cbind(chemical,
                  ingredient = c("A","B","D","C","E",
                                 "C","E","A","D","B",
                                 "B","A","C","E","D",
                                 "D","C","E","B","A",
                                 "E","D","B","A","C"
                                 ),
                  time = c(8,7,1,7,3,
                           11,2,7,3,8,
                           4,9,10,1,5,
                           6,8,6,6,10,
                           4,2,3,8,8)
                  )

chemical %>% kable()
```

day	batch	ingredient	time
d1	b1	A	8
d2	b1	B	7
d3	b1	D	1
d4	b1	C	7
d5	b1	E	3
d1	b2	C	11
d2	b2	E	2

day	batch	ingredient	time
d3	b2	A	7
d4	b2	D	3
d5	b2	B	8
d1	b3	B	4
d2	b3	A	9
d3	b3	C	10
d4	b3	E	1
d5	b3	D	5
d1	b4	D	6
d2	b4	C	8
d3	b4	E	6
d4	b4	B	6
d5	b4	A	10
d1	b5	E	4
d2	b5	D	2
d3	b5	B	3
d4	b5	A	8
d5	b5	C	8

b

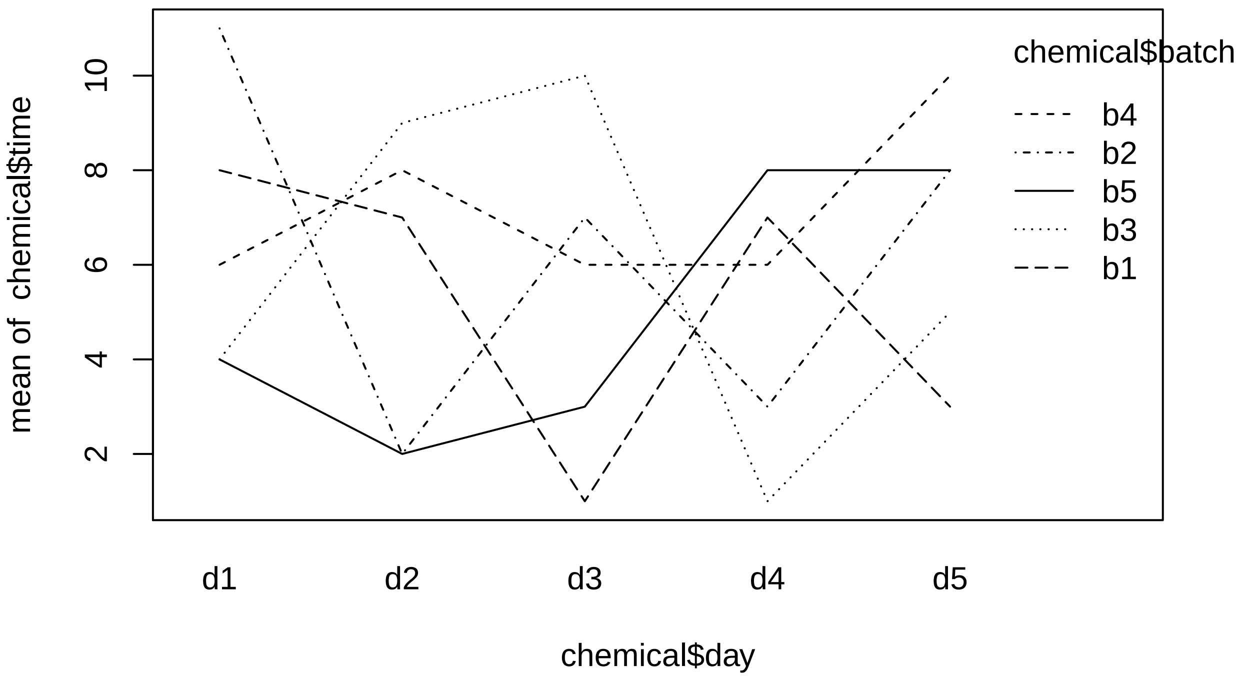
build a linear model using aov. Do the ingredients affect the reaction time? Day means, Batch means, Ingredient means, are significantly different? Check interaction between day and batch.

```
chem_model <- aov(time ~ ingredient+batch+day, data = chemical)
chem_anova <- anova(chem_model)
chem_anova
```

```
## Analysis of Variance Table
##
## Response: time
##            Df Sum Sq Mean Sq F value    Pr(>F)
## ingredient  4 141.44  35.360 11.3092 0.0004877 ***
## batch       4  15.44   3.860  1.2345 0.3476182
## day         4  12.24   3.060  0.9787 0.4550143
## Residuals  12  37.52   3.127
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Ingredient is significantly different, but day and bath are not.

```
interaction.plot(x.factor = chemical$day,
                 trace.factor = chemical$batch,
                 response = chemical$time)
```



There are many potential interactions between day and batch. It can be difficult to judge because the number of replicates is 1.

c

Find the lowest reaction time.

```
chemical %>% group_by() %>%
              summarise(mean=mean(time))
```

```
## # A tibble: 1 x 1
##    mean
##   <dbl>
## 1  5.88
```

3

An industrial engineer is investigating the effect of four assembly methods (A,B,C,D) on the assembly time for a color television component. Four operators are selected for the study. Furthermore, the engineer knows that each assembly method produces such factigue that the time required for the last assembly may be greater than the time required for the first, regardless of the method. That is, a trend develops in the required assembly time. To account for this source of variability, the engineer uses the Latin square design shown below. Analyze the data from this experiment and draw appropriate conclusions.

a

setup the dataframe.

```
operators <- c("1","2","3","4")
assembly_order <- c("1","2","3","4")
tv_component <- expand.grid(operator = operators,
                            assembly = assembly_order
                            )
tv_component <- cbind(tv_component,
                      treatment = c("C","D","A","B",
                                    "B","C","D","A",
                                    "A","B","C","D",
                                    "D","A","B","C"
                                    ),
                      time = c(10,14,7,8,
                               7,18,11,8,
                               5,10,11,9,
                               10,10,12,14)
                      )
tv_component %>% kable()
```

operator	assembly	treatment	time
1	1	C	10
2	1	D	14
3	1	A	7
4	1	B	8
1	2	B	7
2	2	C	18
3	2	D	11
4	2	A	8
1	3	A	5
2	3	B	10
3	3	C	11
4	3	D	9
1	4	D	10
2	4	A	10
3	4	B	12
4	4	C	14

b

build a linear model using aov. Do the treatments affect the assembly time? Operator means, assembly means, treatment means, are they significantly different?

```
tv_model <- aov(time ~ treatment+assembly+operator,
                data=tv_component)
summary(tv_model)
```

```
##               Df Sum Sq Mean Sq F value  Pr(>F)
## treatment      3   72.5  24.167  13.810 0.00421 **
## assembly       3   18.5   6.167   3.524 0.08852 .
## operator       3   51.5  17.167   9.810 0.00993 **
```

```
## Residuals    6   10.5   1.750
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Treatment and operator are significant, while assembly order is not.

c

Find the lowest assembly time.

```r
tv_component %>%
  group_by(treatment) %>%
  summarise(time_mean = mean(time),
            time_std = sd(time)) %>%
  kable()
```

treatment	time_mean	time_std
A	7.50	2.081666
B	9.25	2.217356
C	13.25	3.593976
D	11.00	2.160247

Treatement A is associated with the lowest assembly time.