STAT425_Homework9

Giang Le

11/12/2021

Problem 1

(a) Is there a difference in conductivity due to coating type?

Yes there is a difference in conductivity due to coating type. I perform manual calculations and an automatic check using aov below.

```
# Manual calculations.
set.seed(425)
coating_1 <- c(143, 141, 150, 146)
coating_2 <- c(152, 149, 137, 143)
coating_3 <- c(134, 136, 132, 127)
coating_4 <- c(129, 127, 132, 129)

mean_1 <- mean(coating_1)
mean_2 <- mean(coating_2)
mean_3 <- mean(coating_3)
mean_4 <- mean(coating_4)

all.data <- matrix(c(coating_1, coating_2, coating_3, coating_4, c(1, 1, 1, 1),
                     c(2, 2, 2, 2), c(3, 3, 3, 3), c(4, 4, 4, 4)), nrow=16, ncol=2)
colnames(all.data) <- c("conductivity", "coating")
all.data <- as.data.frame(all.data)
all.data$coating <- as.factor(all.data$coating)

mean.conductivity <- mean(all.data$conductivity)

# total variation (TSS)
tss <- sum((all.data$conductivity - mean.conductivity)^2)
tss
```

```
## [1] 1080.938
```

```
# between group variation (FSS)
fss <- 4*((mean_1 - mean.conductivity)^2 + (mean_2 - mean.conductivity)^2
+ (mean_3 - mean.conductivity)^2 + (mean_4 - mean.conductivity)^2)
fss
```

```
## [1] 844.6875
```

```
# error (within group variation RSS)
rss <- tss - fss

# F-stat
n <- 16
```

```
r <- 4
fstat <- (fss/(r-1))/(rss/(n-r))
fstat
```

```
## [1] 14.30159
```

```
# p-value
pf(fstat, r-1, n-r, lower.tail = FALSE)
```

```
## [1] 0.0002881237
```

We reject the null hypothesis that the mean conductivity values of coating types are the same. We have evidence that at least one mean conductivity is different. The AOV check produced the same result.

```
# AOV check
one.way <- aov(conductivity ~ coating, data=all.data)
summary(one.way)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)    
## coating      3  844.7  281.56    14.3 0.000288 ***
## Residuals   12  236.3   19.69                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(b) Compute a 95% confidence interval for the mean of coating type 4.

The 95% CI for the mean of coating type 4 is (124.4162, 134.0838), calculated below.

```
#CI for single factor level mean
gl = lm(all.data$conductivity ~ all.data$coating-1)
confint(gl)
```

```
##                       2.5 %   97.5 %
## all.data$coating1 140.1662 149.8338
## all.data$coating2 140.4162 150.0838
## all.data$coating3 127.4162 137.0838
## all.data$coating4 124.4162 134.0838
```

(c) Compute a 99% confidence interval for the mean difference.

The mean difference between coating type 4 and 1 has 99% CI of (-27.95552, -3.5444775)

```
TukeyHSD(aov(all.data$conductivity~all.data$coating), data=all.data, conf.level = 0.99)
```

```
##   Tukey multiple comparisons of means
##     99% family-wise confidence level
## 
## Fit: aov(formula = all.data$conductivity ~ all.data$coating)
## 
## $`all.data$coating`
##       diff       lwr        upr     p adj
## 2-1   0.25 -11.95552 12.4555225 0.9998078
## 3-1 -12.75 -24.95552 -0.5444775 0.0073964
## 4-1 -15.75 -27.95552 -3.5444775 0.0014707
## 3-2 -13.00 -25.20552 -0.7944775 0.0064441
## 4-2 -16.00 -28.20552 -3.7944775 0.0012913
## 4-3  -3.00 -15.20552  9.2055225 0.7759360
```

(d)

```
# Tukey Simultaneous 90% CI for all mean differences
TukeyHSD(aov(all.data$conductivity ~ all.data$coating), conf.level = 0.90)
```

```
##   Tukey multiple comparisons of means
##     90% family-wise confidence level
##
## Fit: aov(formula = all.data$conductivity ~ all.data$coating)
##
## $`all.data$coating`
##       diff       lwr       upr     p adj
## 2-1   0.25  -7.78265  8.28265 0.9998078
## 3-1 -12.75 -20.78265 -4.71735 0.0073964
## 4-1 -15.75 -23.78265 -7.71735 0.0014707
## 3-2 -13.00 -21.03265 -4.96735 0.0064441
## 4-2 -16.00 -24.03265 -7.96735 0.0012913
## 4-3  -3.00 -11.03265  5.03265 0.7759360
```

```
mean_1
```

```
## [1] 145
```

```
mean_2
```

```
## [1] 145.25
```

```
mean_3
```

```
## [1] 132.25
```

```
mean_4
```

```
## [1] 129.25
```

According to the table above, coating 1 and 2 produce the highest conductivity and the difference between them is not statistically significant at 0.05.

(e)

I would recommend them to keep using coating 4 because coating 4 has the lowest mean of conductivity and this difference is stat. significant compared to the other coating types (except for coating 3).

Problem 2

Consider the butterfat data set in the Faraway library. This data set contains information about the percent of butter fat (more is better) in the milk taken from 100 cows. In the study, there are 5 different breeds of cows and 2 different ages. We are interested in assessing if Age and Breed affect the butterfat content.

(a)

The Factor Effects model for this problem can be written as

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}\epsilon_{ijk}$$

where Y_{ijk} is the butterfat content of breed i, age j, cow k

μ is the mean butterfat content of all cows

α_i is the effect of breed i on the butterfat content

β_i is the effect of age j on the butterfat content

$(\alpha\beta)_{ij}$ is the interaction term

ϵ_{ijk} is the error term, satisfying $\sim \mathcal{N}(0, \sigma^2)$

Sum constraints are also satisfied (such as $\sum \alpha_i = 0$, etc.)

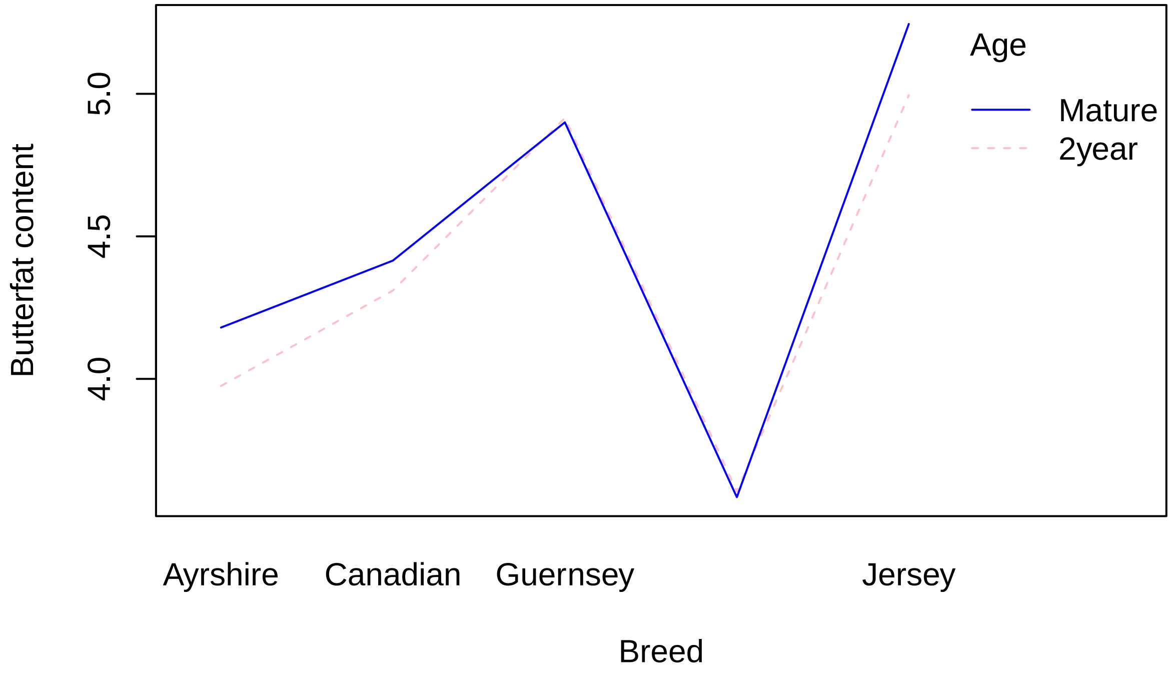
(b) Interaction plot.

```
# Load in the data.
library("faraway")
data('butterfat', package='faraway')
# Fit a full model
bf.model <- aov(Butterfat ~ Breed * Age, data=butterfat)
summary(bf.model)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## Breed        4  34.32   8.580  49.565 <2e-16 ***
## Age          1   0.27   0.274   1.580  0.212
## Breed:Age    4   0.51   0.128   0.742  0.566
## Residuals   90  15.58   0.173
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

According to the interaction plot, there might be no interaction between Age and Breed as the lines seem parallel.

```
interaction.plot(x.factor = butterfat$Breed, trace.factor = butterfat$Age,
                 response = butterfat$Butterfat,
                 fun = median,
                 ylab = "Butterfat content",
                 xlab = "Breed",
                 col = c("pink", "blue"),
                 trace.label="Age")
```



(c) Hypothesis testing for the interaction term.

```
# Fit a model with an interaction term and log(Butterfat)
bf.model1 <- aov(log(Butterfat) ~ Breed * Age, data=butterfat)
summary(bf.model1)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)    
## Breed        4 1.7033  0.4258  56.518 <2e-16 ***
## Age          1 0.0137  0.0137   1.814  0.181    
## Breed:Age    4 0.0223  0.0056   0.741  0.567    
## Residuals   90 0.6781  0.0075                   
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The null hypothesis is that the model without the interaction term is sufficient. The alternative is that the model with the interaction term is required. The interaction term's p-value is $0.566 > 0.05$, so it is not statistically significant. We fail to reject the null hypothesis and we conclude that the model without the interaction term is sufficient. This in line with our plot that there is no interaction.

(d) Testing the main effects.

```
# Use partial F-tests to compare the models.
bf.model2 <- aov(log(Butterfat) ~ Breed, data=butterfat)
bf.model3 <- aov(log(Butterfat) ~ Breed + Age, data=butterfat)
anova(bf.model2, bf.model3)
```

```
## Analysis of Variance Table
## 
## Model 1: log(Butterfat) ~ Breed
## Model 2: log(Butterfat) ~ Breed + Age
##   Res.Df     RSS Df Sum of Sq      F Pr(>F)
## 1     95 0.71410                           
## 2     94 0.70043  1  0.013668 1.8343 0.1789
```

The null hypothesis here is that the reduced model with only Breed is adequate. The alternative hypothesis is that the additive model is required. The p-value here is greater than 0.05, so we fail to reject the null hypothesis. We conclude that the reduced model is adequate and Age is not statistically significant.

```
bf.model4 <- aov(log(Butterfat) ~ Age, data=butterfat)
anova(bf.model4, bf.model3)
```

```
## Analysis of Variance Table
## 
## Model 1: log(Butterfat) ~ Age
## Model 2: log(Butterfat) ~ Breed + Age
##   Res.Df     RSS Df Sum of Sq      F    Pr(>F)    
## 1     98 2.40377                                  
## 2     94 0.70043  4    1.7033 57.149 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The null hypothesis here is that the reduced model with only Age is adequate. The alternative hypothesis is that the additive model is required. The p-value here is less than 0.05, so we fail to reject the null hypothesis. We conclude that the reduced model is not adequate and Breed is statistically significant.

(e) Estimate the mean difference in butterfat content between Mature and 2year cows.

The estimated CI is (-0.003788486, 0.212988486). See calculation below.

```
mean_mature <- mean(butterfat[butterfat$Age =="Mature",c(1)])
mean_twoyears <- mean(butterfat[butterfat$Age =="2year",c(1)])
diff <- mean_mature - mean_twoyears
# estimator
diff
```

```
## [1] 0.1046
```

```
# find  (= 0.00745)
anova(bf.model3)
```

```
## Analysis of Variance Table
##
## Response: log(Butterfat)
##           Df  Sum Sq Mean Sq F value Pr(>F)
## Breed      4 1.70334 0.42584 57.1486 <2e-16 ***
## Age        1 0.01367 0.01367  1.8343 0.1789
## Residuals 94 0.70043 0.00745
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# variance
n <- 1
a <- 5
var <- 2*0.00745/(n*a)

# find t-value
alpha = 0.05
dof = 94
t.score = qt(p=alpha/2, df=dof, lower.tail=F)
margin = t.score * sqrt(var)
lb <- diff - margin
ub <- diff + margin
print(c(lb, ub))
```

```
## [1] -0.003788486  0.212988486
```

```
butterfat$Breed
```

```
##   [1] Ayrshire          Ayrshire          Ayrshire          Ayrshire
##   [5] Ayrshire          Ayrshire          Ayrshire          Ayrshire
##   [9] Ayrshire          Ayrshire          Ayrshire          Ayrshire
##  [13] Ayrshire          Ayrshire          Ayrshire          Ayrshire
##  [17] Ayrshire          Ayrshire          Ayrshire          Ayrshire
##  [21] Canadian          Canadian          Canadian          Canadian
##  [25] Canadian          Canadian          Canadian          Canadian
##  [29] Canadian          Canadian          Canadian          Canadian
##  [33] Canadian          Canadian          Canadian          Canadian
##  [37] Canadian          Canadian          Canadian          Canadian
##  [41] Guernsey          Guernsey          Guernsey          Guernsey
##  [45] Guernsey          Guernsey          Guernsey          Guernsey
##  [49] Guernsey          Guernsey          Guernsey          Guernsey
##  [53] Guernsey          Guernsey          Guernsey          Guernsey
##  [57] Guernsey          Guernsey          Guernsey          Guernsey
##  [61] Holstein-Fresian Holstein-Fresian Holstein-Fresian Holstein-Fresian
##  [65] Holstein-Fresian Holstein-Fresian Holstein-Fresian Holstein-Fresian
##  [69] Holstein-Fresian Holstein-Fresian Holstein-Fresian Holstein-Fresian
##  [73] Holstein-Fresian Holstein-Fresian Holstein-Fresian Holstein-Fresian
##  [77] Holstein-Fresian Holstein-Fresian Holstein-Fresian Holstein-Fresian
##  [81] Jersey           Jersey           Jersey           Jersey
##  [85] Jersey           Jersey           Jersey           Jersey
##  [89] Jersey           Jersey           Jersey           Jersey
##  [93] Jersey           Jersey           Jersey           Jersey
##  [97] Jersey           Jersey           Jersey           Jersey
## Levels: Ayrshire Canadian Guernsey Holstein-Fresian Jersey
```

(f) Estimate the following contrast with a 95% confidence interval.

The 95% CI is (4.249123, 4.591877)

```
mean_a <- mean(butterfat[butterfat$Breed =="Ayrshire",c(1)])
mean_c <- mean(butterfat[butterfat$Breed =="Canadian",c(1)])
mean_g <- mean(butterfat[butterfat$Breed =="Guernsey",c(1)])
mean_j <- mean(butterfat[butterfat$Breed =="Jersey",c(1)])
estimator <- (1/2)*(mean_a + mean_c - mean_g + mean_j)
estimator
```

```
## [1] 4.4205
```

```
# find variance
b <- 2
var2 <- 2*0.00745/(n*b) * 4 * (1/2)^2
# find margin
margin2 = t.score * sqrt(var2)
lb <- estimator - margin2
ub <- estimator + margin2
print(c(lb, ub))
```

```
## [1] 4.249123 4.591877
```