QSCI 482 Story 7

Simon Hans Edasi

2023-11-30

1a. Read the data into an R object and use the table() function to check that you have a balanced two-factor ANOVA design.

Balanced design will have same n for each factor.

```
df <- read.csv('7 Guineapigdata.csv')
df$supp <- as.factor(df$supp)
df$dose <- as.factor(df$dose)
table(df$supp)
```

```
##
## OrangeJuice        Pill
##          30          30
```

```
table(df$dose)
```

```
##
## Dose0.5   Dose1   Dose2
##      20      20      20
```

1b. For the rest of the assignment (for grading simplicity), choose the dose to be Factor A and the supplement type to be Factor B. Test the underlying assumptions of the two-factor ANOVA, using Levene's Test (short way around) for homogeneity of variances. Write down your conclusions.

Underlying assumptions:

1. Variances are the same

H_0 : Variances across factors are equal

H_A : Variances across factors are not equal

```
leveneTest(len ~ dose*supp,data = df)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  5  1.7086 0.1484
##       54
```

We fail to reject the null hypothesis that variances are equal.

2. Data in each cell comes from a normally distributed population

```
d1 <- shapiro.test(df$len[df$dose == 'Dose0.5' & df$supp == 'Pill'])
d2 <- shapiro.test(df$len[df$dose == 'Dose0.5' & df$supp == 'OrangeJuice'])
d3 <- shapiro.test(df$len[df$dose == 'Dose1' & df$supp == 'Pill'])
d4 <- shapiro.test(df$len[df$dose == 'Dose1' & df$supp == 'OrangeJuice'])
```

```
d5 <- shapiro.test(df$len[df$dose == 'Dose2' & df$supp == 'Pill'])
d6 <- shapiro.test(df$len[df$dose == 'Dose2' & df$supp == 'OrangeJuice'])

glue('p-values\ncell1: {d1[2]}\ncell2: {d2[2]}\ncell3: {d3[2]}\ncell4: {d4[2]}
cell5: {d5[2]}\ncell6: {d6[2]}')
```

```
## p-values
## cell1: 0.169562726932084
## cell2: 0.182040790059053
## cell3: 0.269785507503429
## cell4: 0.415298280473649
## cell5: 0.919449675002835
## cell6: 0.814790825009568
```

Cells appear normally distributed, we're good here.

1c. Run a two-factor ANOVA on the data the short way around. Copy and paste the output table in R that includes the degrees of freedom, mean square, p-values etc. Interpret the resulting p-values in your own words.

H_0 : No effect of dose on tooth length

H_0 : No effect of supplement on tooth length

H_0 : No interaction effect of dose and supplement on tooth length

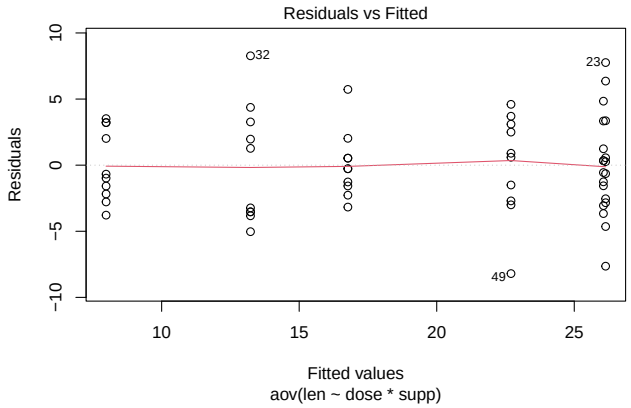
```
aov.res <- aov(len ~ dose*supp, data = df)
summary(aov.res)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## dose         2 2426.4  1213.2  92.000  < 2e-16 ***
## supp         1  205.3   205.3  15.572 0.000231 ***
## dose:supp    2  108.3    54.2   4.107 0.021860 *
## Residuals   54  712.1    13.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

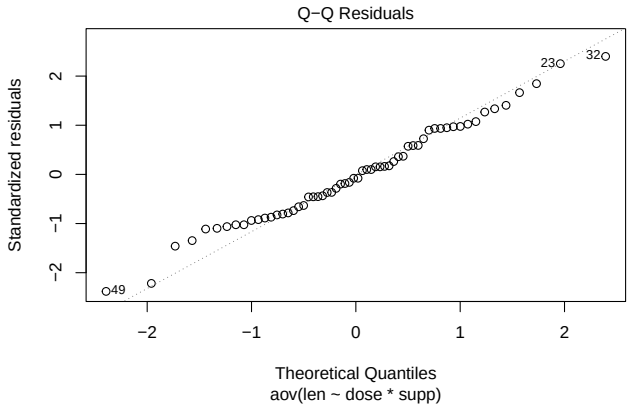
Factor A, B, and AxB all have p-values less than 0.05, which means we should reject all three null hypotheses that both dose and supplement have no effect, and their interactions have no effect. We are left conclude that tooth length is dependent on both the dose size, supplement given, and an interaction between the two factors.

1d. Provide the plot for standardized residuals, the Q-Q plot, and run a Shapiro-Wilks test (short-way around), to check whether the data are really normally distributed. Paste the plots into this document, and explain your conclusions about whether the data are normally distributed.

```
plot(aov.res,1)
```




```
plot(aov.res,2)
```



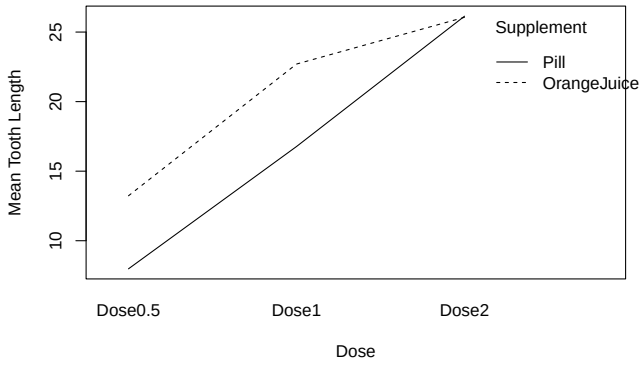

```
shapiro.test(df$len)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  df$len
## W = 0.96743, p-value = 0.1091
```

Both residuals and Q-Q plots look normal, and the Shapiro-Wilk normality test concludes the data are normally distributed.

1e. Plot the interaction plot for these data. What do you conclude from this plot?

```
?interaction.plot()
interaction.plot(x.factor = df$dose, trace.factor = df$supp, response = df$len,
                 xlab = 'Dose', ylab = 'Mean Tooth Length', trace.label = 'Supplement')
```



This interaction plot shows that increasing the dose for either supplement will increase values of mean tooth length. OJ appears to be a better supplement for lower doses but OJ has similar effectiveness as pills at higher dose.

1f. Run the two-factor ANOVA the long way around for these data. Report the intermediate steps involving calculating the means (grand mean, means of the levels of each factor, cell means), sums of squares and DF (total, factor A, factor B, interaction, error), the relevant mean squares, F statistics, and p-values. Report your final answers in a table similar to that presented in the lectures.

```
alpha <- 0.05
# Find grand mean
gm <- mean(df$len)

#find total sum of squares
tss <- sum((df$len - gm)^2)

#total DF
N <- nrow(df)
tdf <- N-1

# find cell SS
doses <- c('Dose0.5','Dose1','Dose2')
supps <- c('Pill','OrangeJuice')
cellss <- 0
a <- length(doses)
b <- length(supps)
for (i in doses){
  for (j in supps){
    cell_mean = mean(df$len[df$supp == j & df$dose == i])
    cellss <- cellss + (cell_mean - gm)^2
  }
}
n <- length(df$len[df$supp == 'Pill' & df$dose == 'Dose1'])
cellss <- n * cellss
celldf <- a*b - 1

# error SS and df
ess <- tss - cellss
```

```r
edf <- tdf - celldf

# factor A (doses) SS and df
fass <- 0
for (i in doses){
  famean <- mean(df$len[df$dose == i])
  fass <- fass + (famean - gm)^2
}
fass <- b*n*fass
fadf <- a-1

# factor B (supp) SS and df
fbss <- 0
for (i in supps){
  fbmean <- mean(df$len[df$supp == i])
  fbss <- fbss + (fbmean - gm)^2
}
fbss <- a*n*fbss
fbdf <- b-1

# Interaction SS and df
intss <- cellss - fass - fbss
intdf <- celldf - fadf - fbdf

# MS
faMS <- fass / fadf
fbMS <- fbss / fbdf
intMS <- intss / intdf
eMS <- ess / edf

# F statistics
fa <- faMS / eMS
fb <- fbMS / eMS
fint <- intMS / eMS

fcrita <- qf(alpha, df1 = fadf, df2 = edf, lower.tail = FALSE)
fcritb <- qf(alpha, df1 = fbdf, df2 = edf, lower.tail = FALSE)
fcritint <- qf(alpha, df1 = intdf, df2 = edf, lower.tail = FALSE)



h1 <- pf(fa, fadf, edf, lower.tail = FALSE)
h2 <- pf(fb, fbdf, edf, lower.tail = FALSE)
h3 <- pf(fint, intdf, edf, lower.tail = FALSE)
print(h1)
```

```
## [1] 4.046291e-18
```

```r
print(h2)
```

```
## [1] 0.0002311828
```

```r
print(h3)
```

```
## [1] 0.02186027
```