STAT 4100 Homework 5

Cody Frisby

February 17, 2016

1) Statistical Model

A)

Here, we have 3 treatment groups and 4 blocks. The statistical model is

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} \begin{cases} i = 1, 2, 3 \\ j = 1, 2, 3, 4 \end{cases}$$

where μ is the overall mean of bacteria, τ is the treatment effect from the three different solutions, β is the block effect, here as days, and ε is the error from the model. The model assumes normality of the errors and equal variances. We also need to be aware of possible block-treatment interaction.

B) Hypothesis Test

Our Hypothesis of interest is

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

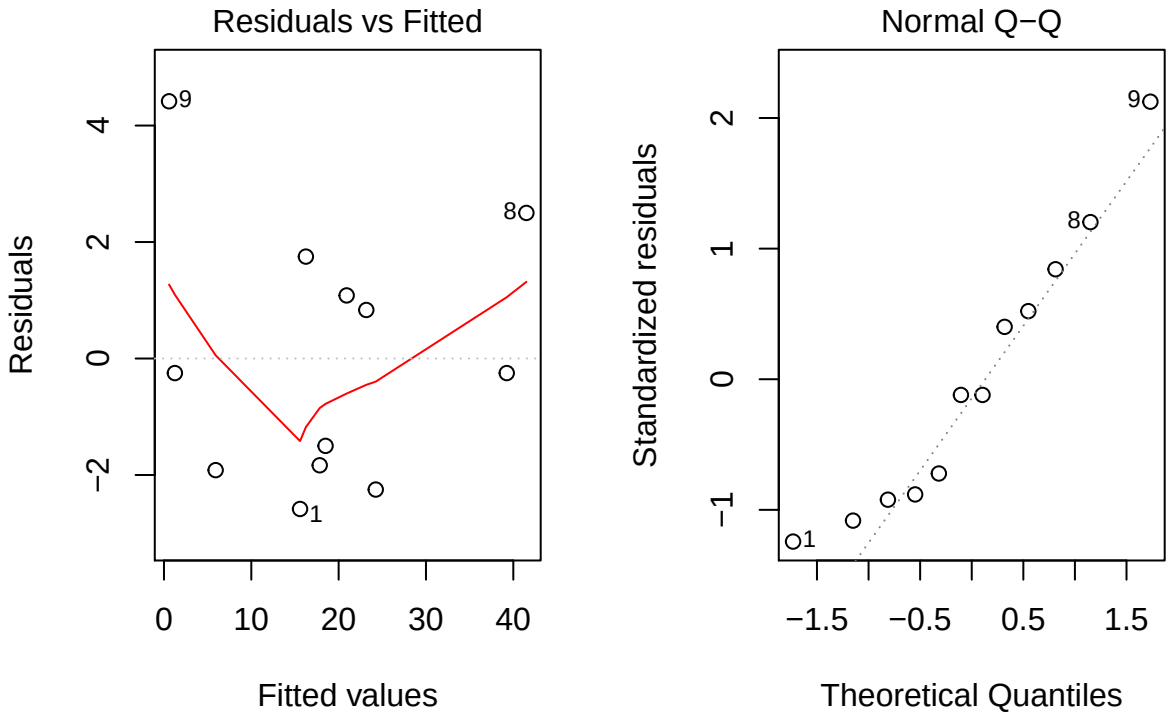
$$H_A : \textit{at least one } \mu_i \neq \mu_j$$

Testing this hypothesis we run an ANOVA test with a treatment group and a blocking group. The result is similar to ANOVA from previous examples only we have added another term to the model. If the F statistic from the test is greater than $F_{\alpha/2:df}$ then we reject the null hypothesis.

Here I fill in the ANOVA table:

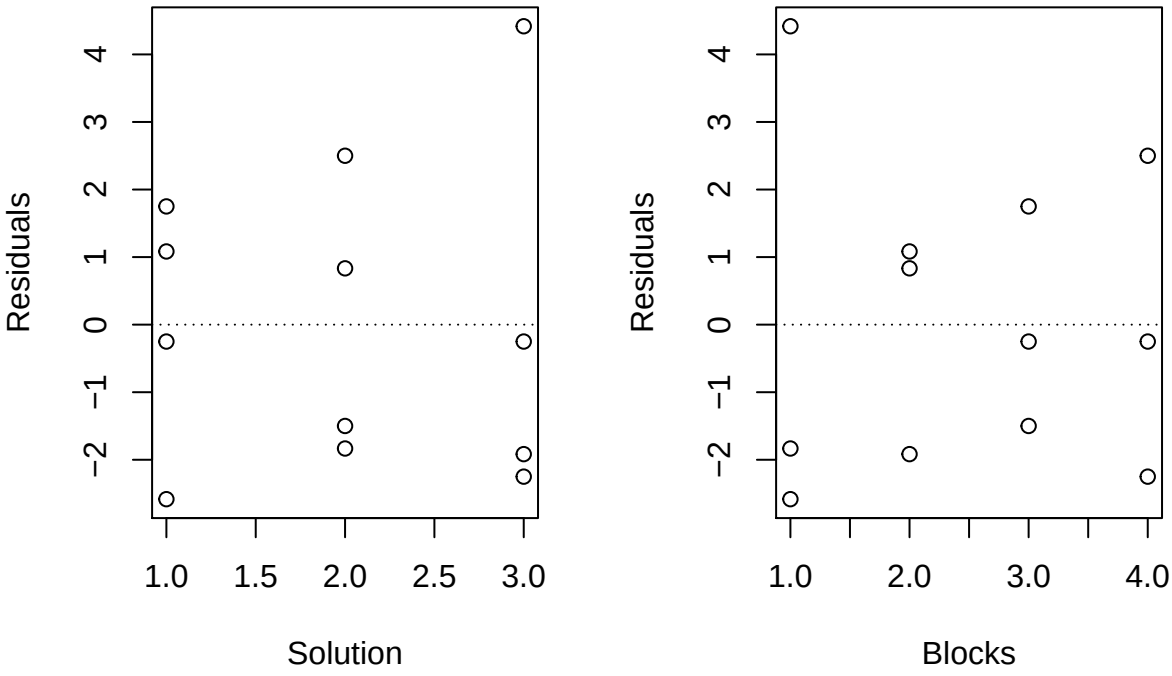
Source	Sum of Squares	Degrees of Freedom	Mean Square	F_0	P-Value
Treatments	703.5	2	351.75	40.7170418	3.2315543×10^{-4}
Blocks	51.8333333	9	8.6388889		
Error	1106.9166667	3	368.9722222		
Total	1862.25	11			

So, we reject the null hypothesis that there is no difference between treatment means. However, we should check our model assumptions to be sure our conclusions are valid, in that we aren't violating any of the above model assumptions.



There doesn't appear to be any major violations of our models' assumptions of normality and equal variance.

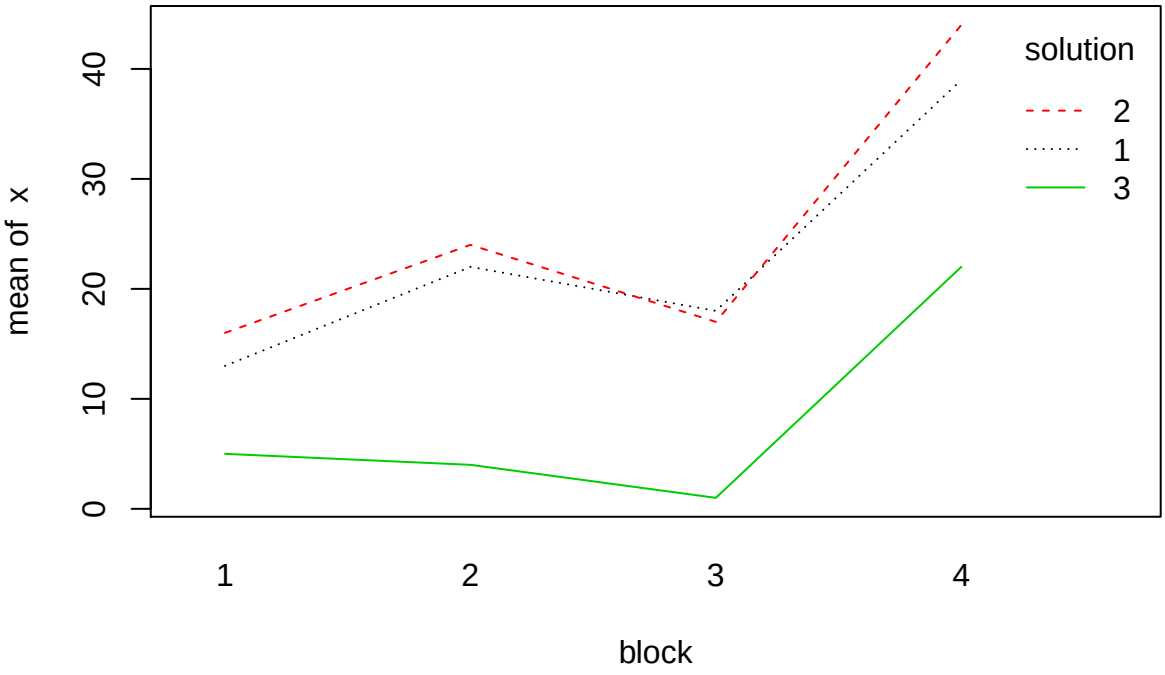
Here, we add two more residual plots. This is to ensure there isn't anything concerning going on between the treatment and blocks and the residuals.



There is a little difference in the variance here where solution = 3 and where block = 1 but nothing we should be too concerned about.

And finally we check for block-treatment interactions. If there is any type of curvilinear shape in the residual

vs $\hat{y}_{ij}$ plot this may be an indication of interaction between blocks and treatments. Additionaly we can create an interactino plot using R, interaction.plot() is the function.



C) Contrast

The coefficients for the contrast will be 1, 0 ,-1 for groups 1, 2, & 3 respectively. Using the formula for calculating the t_0 statistic for the contrasts

$$t_0 = \frac{\sum_{i=1}^{a} c_i \bar{y}_{i.}}{\sqrt{\frac{MS_E}{b} \sum_{i=1}^{a} c_i^2}}$$

$$7.2173425 = \frac{15}{2.0783273}$$

The standard error is

$$\sqrt{\frac{MS_E}{b} \sum_{i=1}^{a} c_i^2}$$

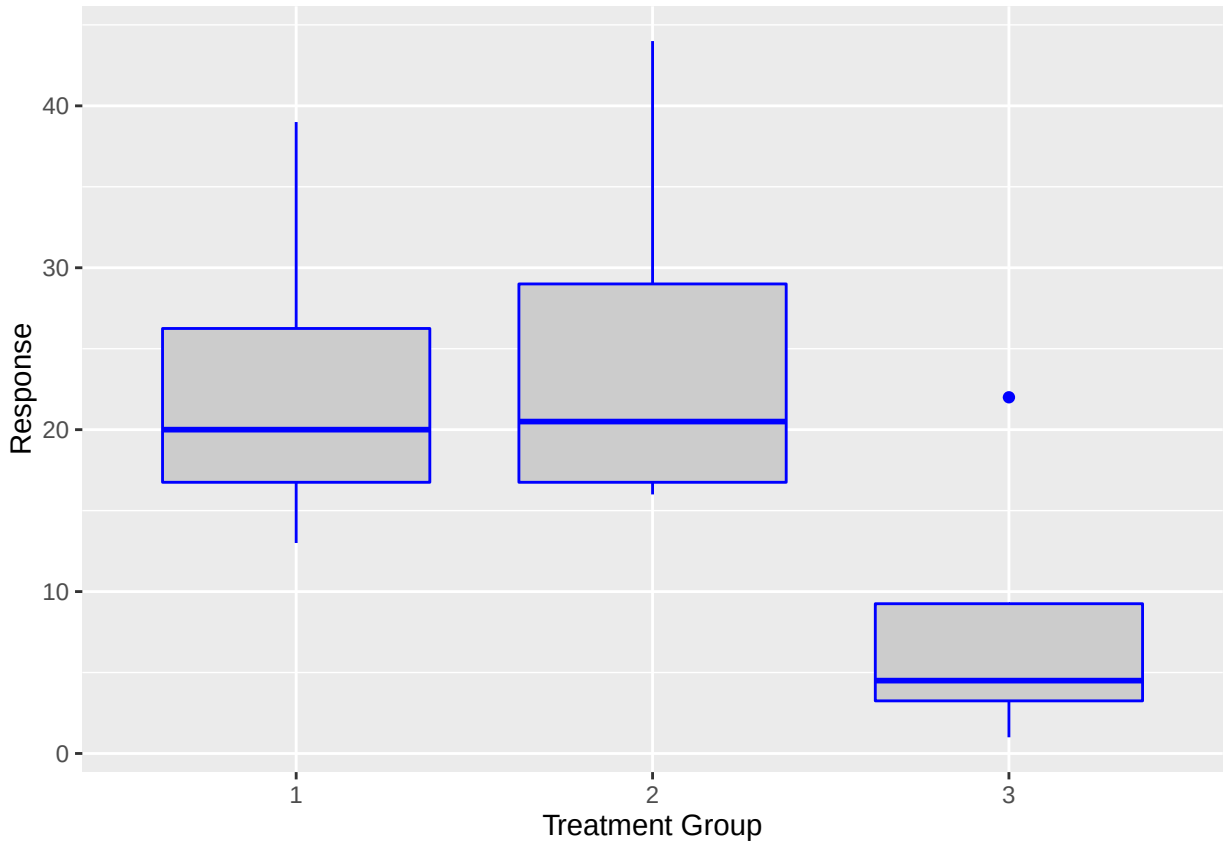
which is the denominator of the t_0 equation, 2.0783273.

Also,

$$7.2173425 > 2.8214379$$

at $\alpha = 0.01$.

Visually, we can see a differnce between groups 1 and 3



2) RCBD

A)

The main advantages of a complete randomized block design when compared to a randomized design is that we can protect ourselves from a lurking variable or some other uncontrollable variable that may have an affect on the response and that is known to us. The above example we blocked for days. This protects us from concluding differences that may or may not be real since the day did have an affect on the response.

B)

Complete randomized designs are superior to RCBD when the experimental units are homogeneous. It is also nesecarry when the lurking variables are unknown and therefore uncontrollable.

3) Latin Squares Design

A) Model

The statistical model fro this experiment is

$$y_{ijk} = \mu + \alpha_i + \tau_i + \beta_k + \varepsilon_{ijk} \begin{cases} i = 1, 2, ..., p \\ j = 1, 2, ..., p \\ k = 1, 2, ..., p \end{cases}$$

where y_{ijk} is the observation in the ith row and kth column for the jth treatment group, μ is the overall mean of our response, α is the effect from the ith strip, τ is the effect from the ith electrode shape, β is the effect from the ith postion, and ε is the overall error, or residuals from our model. This model is completely additive, there is no interaction between rows, columns, and treatments.

Here I use R and knitr to display the latin squares matrix.

	pos1	pos2	pos3	pos4	pos5
stripI	A64	B61	C62	D62	E62
stripII	B62	C62	D63	E62	A63
stripIII	C61	D62	E63	A63	B62
stripIV	D63	E64	A63	B63	C63
stripV	E62	A61	B63	C63	D62

And here we display the column, row, and treatment means:

Table 3: Column, Row, & Treatment Sums

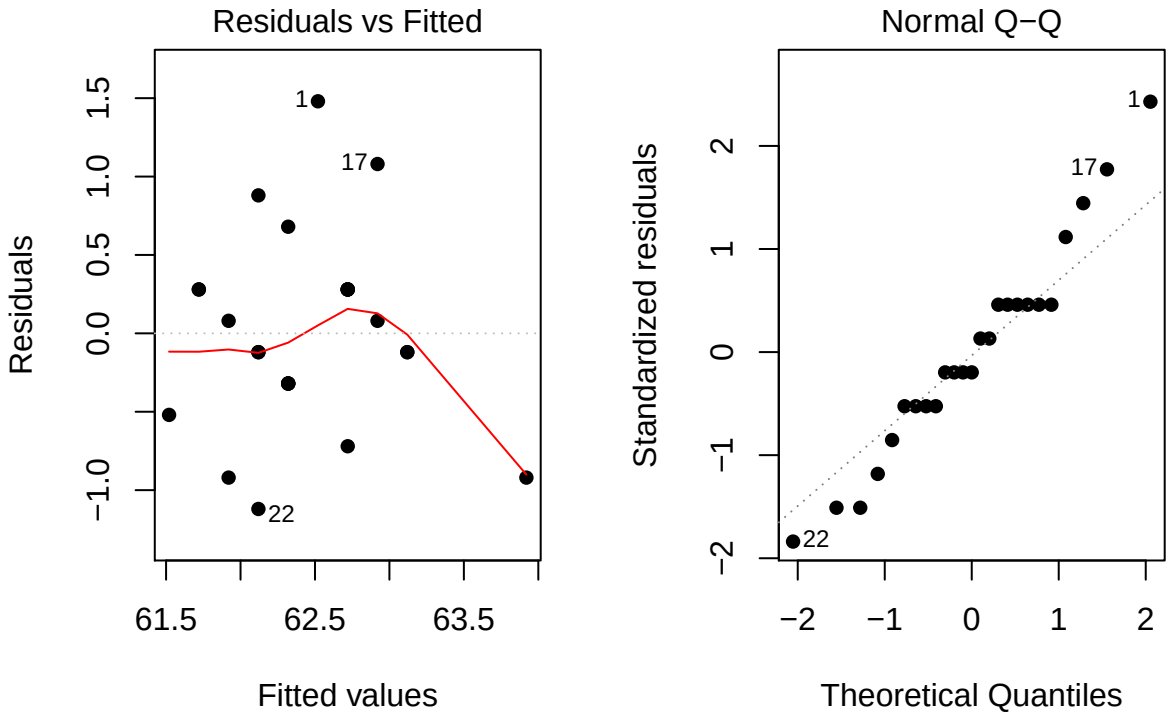
strip	hardness	position	hardness	shape	hardness
stripI	311	pos1	312	A	314
stripII	312	pos2	310	B	311
stripIII	311	pos3	314	C	311
stripIV	316	pos4	313	D	312
stripV	311	pos5	312	E	313

We've actually used R and the library dplyr to compute all the treatment sums, column sums, and row sums. This is quite simple and the syntax is easy to read. Using these computations here I fill in the ANOVA table:

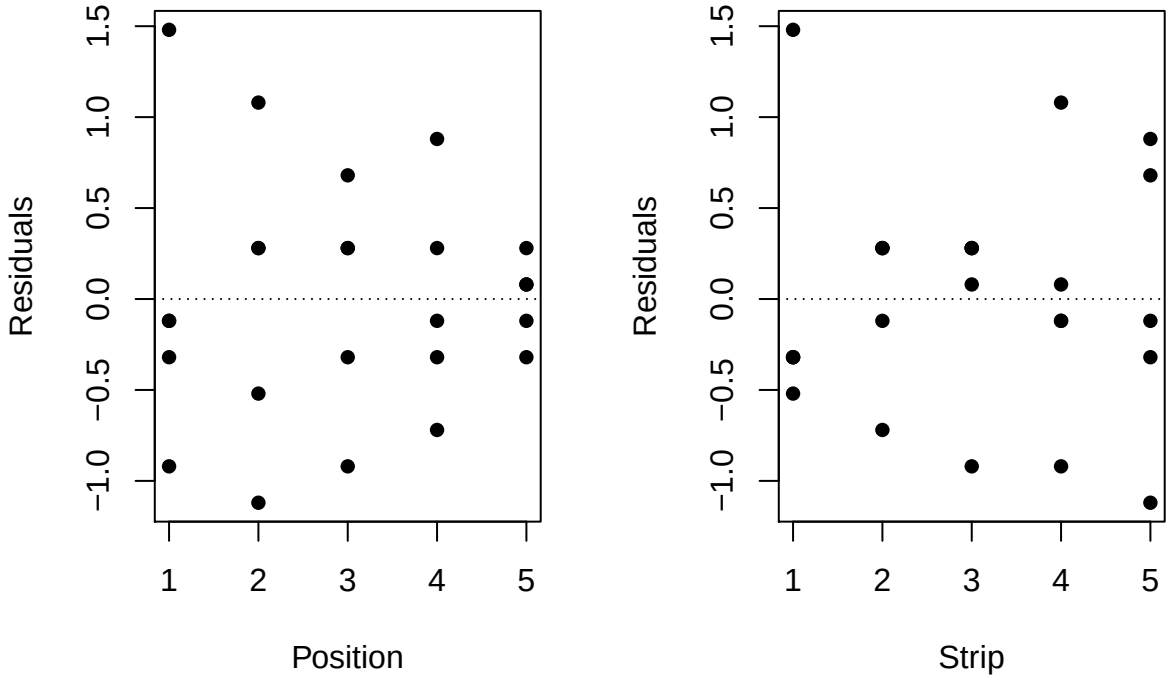
Source	Sum of Squares	Degrees of Freedom	Mean Square	F_0	P-Value
Treatments	1.36	4	0.34	0.4396552	0.7777756
Rows	3.76	4	0.94	1.2155172	0.3546457
Columns	1.76	4	0.44	0.5689655	0.6901732
Error	9.28	12	0.7733333		
Total	16.16	24			

We do not reject the null hypothesis. We cannot conclude that there is a difference between treatments.

Now for some model diagnostics:



Here, the Residual plot looks ok except for the one outlier towards the right side of the plot. The QQ plot is OK as well. We don't wander off the line too far to be concerned about. We can drill down further by looking at more residual plots by the row and column blocks.



Here, there are no violations of our model assumptions. We still conclude no difference between means. We cannot reject the null hypothesis.

R code:

```
# read the data
x <- c(13, 22, 18, 39, 16, 24, 17, 44, 5, 4, 1, 22)
trt <- c(1,2,3)
block <- c(1,2,3,4)
# format the data frame coreectly
milk <- data.frame(solution = gl(length(trt),length(block),length(x),
            labels = trt), block=gl(length(block),1,length(x)), x)

milk.aov <- aov(x~block + solution, data = milk)
# anova(milk.aov) # print the model anova table to the console
library(dplyr) # type install.packages("dplyr") is you do not have this.
milk.blocks <- milk %>%
  group_by(block) %>%
  summarize_each(funs(mean, sum), x)
milk.trt <- milk %>%
  group_by(solution) %>%
  summarize_each(funs(mean, sum), x)
y.t <- milk.trt$sum # treatment totals
y.b <- milk.blocks$sum # block totals
y.. <- sum(x)
b <- length(y.b)
a <- length(y.t)
N <- length(x)
n <- N/a
sst <- sum(x^2) - sum(x)^2/N
sstr <- (1/b) * sum(milk.trt$sum^2) - y..^2/N
ssb <- (1/a) * sum(y.b^2) - y..^2/N
sse <- sst - sstr - ssb
mstr <- sstr/(a-1)
msb <- ssb/(b-1)
mse <- sse/((a-1)*(b-1))
F0 <- mstr/mse
p.F <- 1 - pf(F0, a-1, (a-1)*(b-1))
par(mfrow=c(1,2))
plot(milk.aov, which=c(1,2))
par(mfrow=c(1,2))
plot(milk.aov$residuals ~ as.numeric(milk$solution), xlab="Solution",
     ylab="Residuals")
abline(h=0, lty=3)
plot(milk.aov$residuals ~ as.numeric(milk$block), xlab="Blocks",
     ylab="Residuals")
abline(h=0, lty=3)
rmse <- sqrt(mse)
con <- 2
num <- milk.trt$mean[1] - milk.trt$mean[3]
denom <- sqrt((mse/b)*con)
t0 <- num/denom
tcrit <- qt(0.99, N-a)
library(ggplot2)
ggplot(milk, aes(x = solution, y = x)) +
  geom_boxplot(fill = "grey80", colour = "blue") +
```

```
  scale_x_discrete() + xlab("Treatment Group") +
  ylab("Response")
# get the data for our experiment into R
position <- c(rep("pos1",1), rep("pos2",1), rep("pos3",1),
           rep("pos4",1), rep("pos5",1))
strip <- c(rep("stripI",5), rep("stripII",5), rep("stripIII",5),
          rep("stripIV",5), rep("stripV",5))
shape <- c("A", "B", "C", "D", "E", "B", "C", "D", "E", "A", "C", "D", "E",
          "A", "B", "D", "E", "A", "B", "C", "E", "A", "B", "C", "D")
hardness <- c(64, 61, 62, 62, 62, 62, 62, 63, 62, 63, 61, 62, 63, 63, 62, 63,
              64, 63, 63, 63, 62, 61, 63, 63, 62)
mydata <- data.frame(position, strip, shape, hardness)
# 3 plots are created
#anova(myfit) # same model as below
fit <- aov(hardness ~ position+strip+shape, mydata)
A <- matrix(paste(shape, hardness, sep = ""), 5,5, byrow = TRUE)
rownames(A) <- c("stripI","stripII","stripIII","stripIV","stripV")
colnames(A) <- position
knitr::kable(A)
# computing means by all the different grouping factors
library(dplyr)
treat.sums <- mydata %>%
               group_by(shape) %>%
               summarise_each(funs(sum), hardness)
row.sums <- mydata %>%
               group_by(strip) %>%
               summarise_each(funs(sum), hardness)
column.sums <- mydata %>%
               group_by(position) %>%
               summarise_each(funs(sum), hardness)
x <- mydata$hardness
p <- 5
N <- length(x)
y.t <- treat.sums$hardness # treatment totals
y.r <- row.sums$hardness # row totals
y.c <- column.sums$hardness # column totals
y.. <- sum(x)
sst <- sum(x^2) - sum(x)^2/N
sstr <- (1/p) * sum(y.t^2) - y..^2/N
ssr <- (1/p) * sum(y.r^2) - y..^2/N
ssc <- (1/p) * sum(y.c^2) - y..^2/N
sse <- sst - sstr - ssr - ssc
mstr <- sstr/(p-1)
msr <- ssr/(p-1)
msc <- ssc/(p-1)
mse <- sse/((p-2)*(p-1))
F0 <- mstr/mse
p.F <- 1 - pf(F0, p-1, (p-2)*(p-1))
library(knitr)
kable(cbind(row.sums, column.sums, treat.sums), caption = "Column, Row,
      & Treatment Means")
```