

# Burris\_midterm\_knitted.R

*christieburris*

*Thu Jun 7 08:02:16 2018*

```
##### Christie Burris #####
##### Biometry Midterm #####

# Prompt: Dr. Smith happens to like dark chocolate brownies and especially favors
# those that have the right amount of chewiness and flavor. Suppose that chewiness
# can be measured (perhaps by eating some). To do this you decide to run an experiment
# by baking a number of batches of brownies and compare them to come up with the
# best brownies. There were four different brands that were considered.
# In his study he made a batch of each brownies (using 4 8x8 pans, each pan having
# a different brand mix), set the oven temperature to 350 degrees and placed all
# four in the oven at once. This was repeated 10 times on different days.

# Analyze the data - what questions/hypotheses might you ask prior to experimentation?

#####

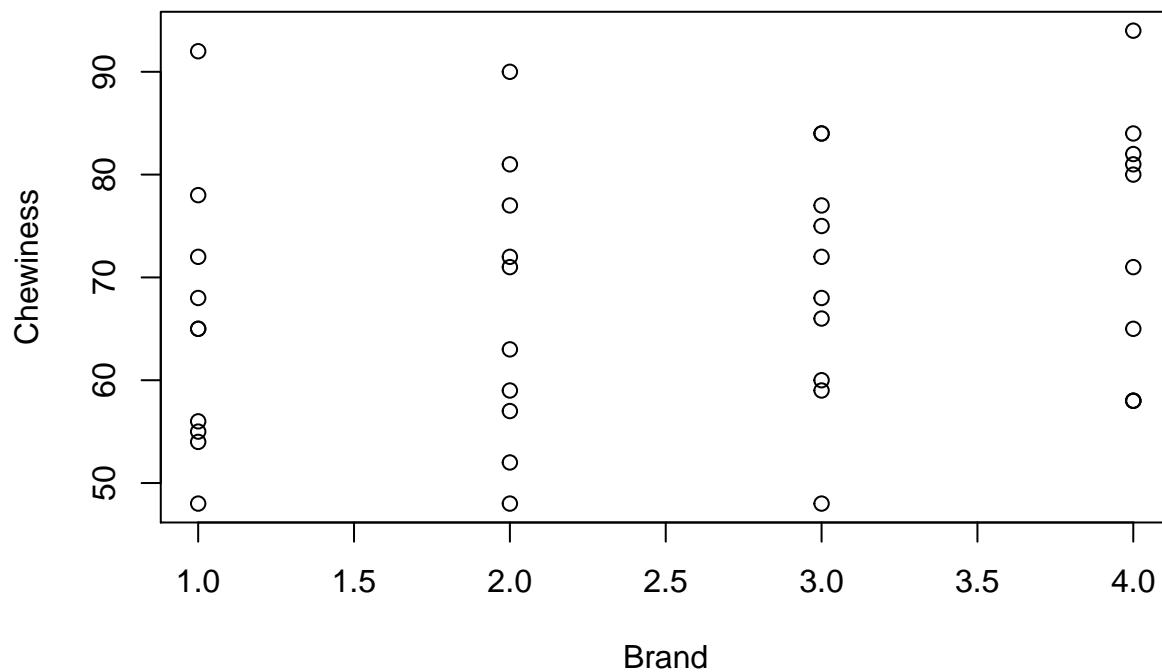
# We first run ANOVA as if the design was completely randomized.
# An ANOVA table is produced from the linear fit of brand and chewiness
# The results indicate no significant difference in chewiness.
# Visualizing the data tells us that the variance amongst each brand is very high.
# Tukey's multiple comparisons confirms lack of significance.
# We instead decide that a randomized block design better describes the experiment.

# read data
library(readr)
Data <- read.csv("~/Desktop/Biometry/Midterm/brownieCB.csv",header=T)
attach(Data)

blockF <- as.factor(block)
brandF <- as.factor(brand)

##### not taking days to be blocks
# (assuming completely randomized design) #####

# visualize data, lots of variance within brands, less variance between brands
plot(brand,chewy,ylab = "Chewiness", xlab = "Brand")
```



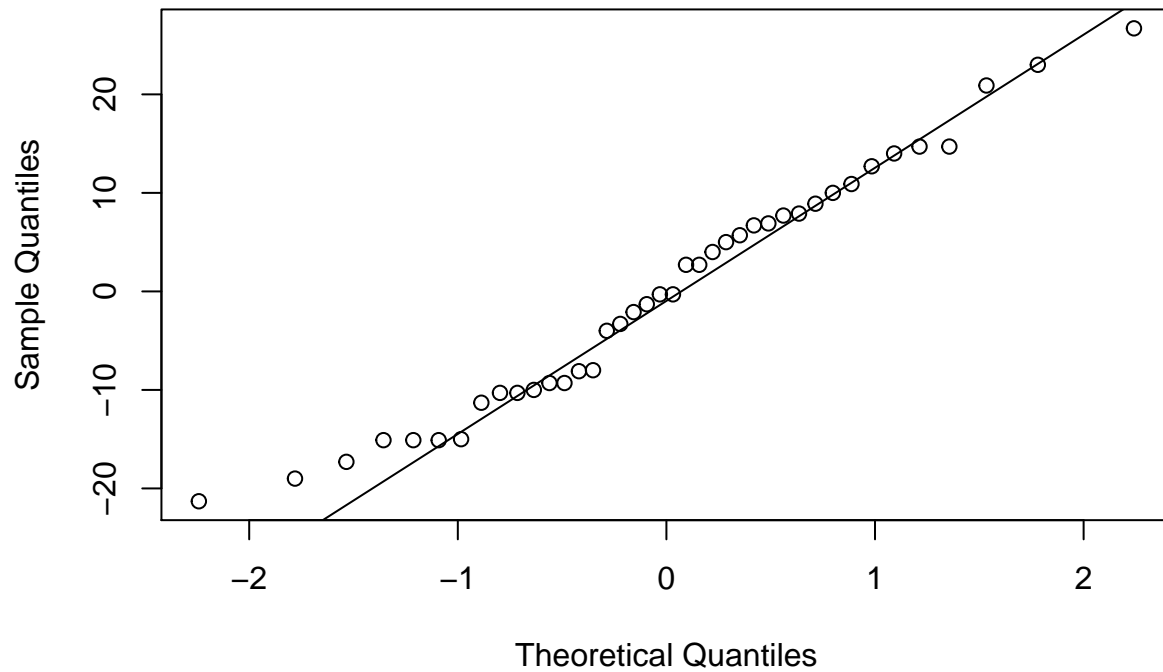
```
# fit linear model
fit <- lm(chewy ~ brandF, data = Data)
anova(fit)

## Analysis of Variance Table
##
## Response: chewy
##      Df Sum Sq Mean Sq F value Pr(>F)
## brandF    3  341.7   113.89   0.6988 0.5589
## Residuals 36 5867.1   162.97

# check assumptions
qqnorm(fit$residuals)
# compare to normal
qqline(fit$residuals)

# Tukey's multiple comparisons
library(multcompView)
```

## Normal Q-Q Plot



```
library(emmeans)
```

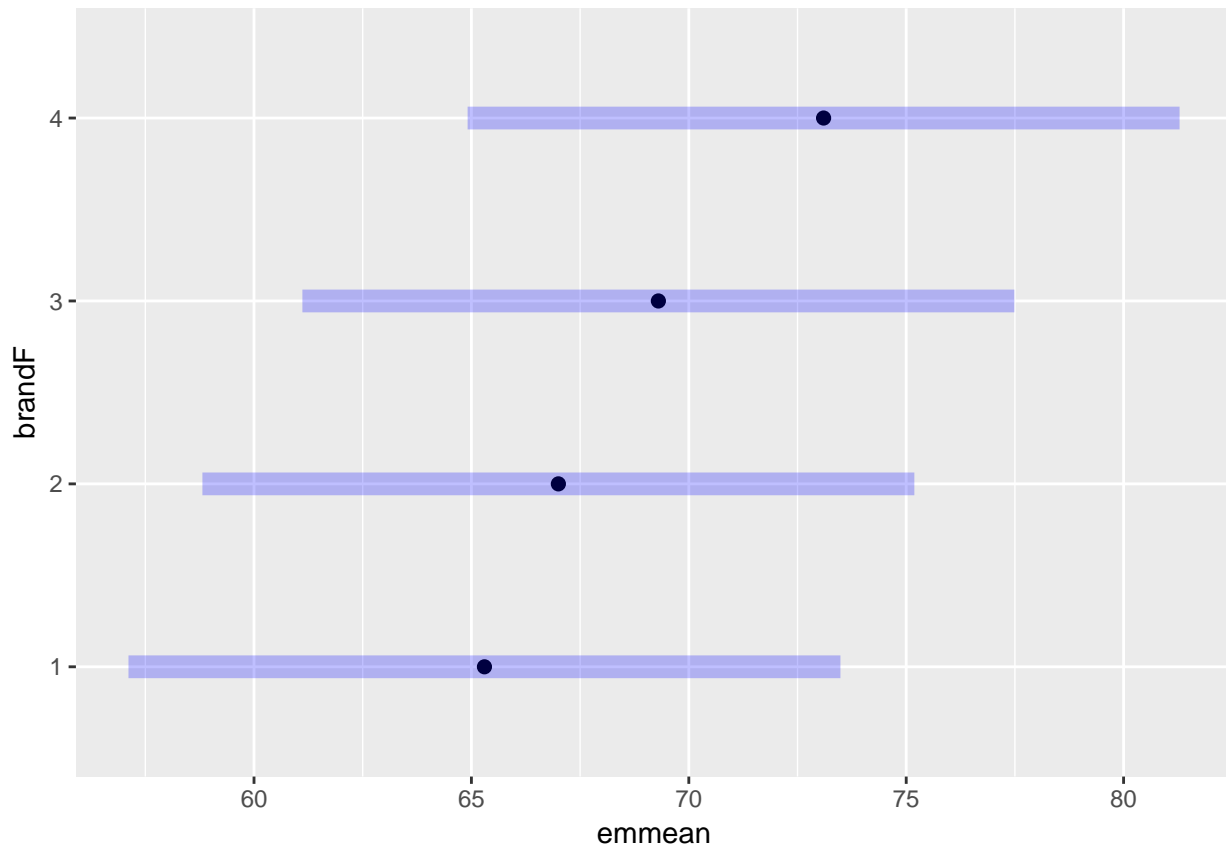
```
## Warning: package 'emmeans' was built under R version 3.4.3
```

```
fit_e <- emmeans(fit, "brandF") #estimate means from fit and factor=brand
pairs(fit_e) # intervals for pairs
```

```
## contrast estimate      SE df t.ratio p.value
## 1 - 2          -1.7 5.709203 36  -0.298  0.9907
## 1 - 3          -4.0 5.709203 36  -0.701  0.8962
## 1 - 4          -7.8 5.709203 36  -1.366  0.5281
## 2 - 3          -2.3 5.709203 36  -0.403  0.9775
## 2 - 4          -6.1 5.709203 36  -1.068  0.7106
## 3 - 4          -3.8 5.709203 36  -0.666  0.9093
##
```

```
## P value adjustment: tukey method for comparing a family of 4 estimates
```

```
plot(fit_e) # plot means and std errors
```



```
# output comparison groups
cld(fit_e,adjust="Tukey")
```

```
## brandF emmean SE df lower.CL upper.CL .group
## 1 65.3 4.037016 36 54.71631 75.88369 1
## 2 67.0 4.037016 36 56.41631 77.58369 1
## 3 69.3 4.037016 36 58.71631 79.88369 1
## 4 73.1 4.037016 36 62.51631 83.68369 1
##
## Confidence level used: 0.95
## Conf-level adjustment: sidak method for 4 estimates
## P value adjustment: tukey method for comparing a family of 4 estimates
## significance level used: alpha = 0.05
```

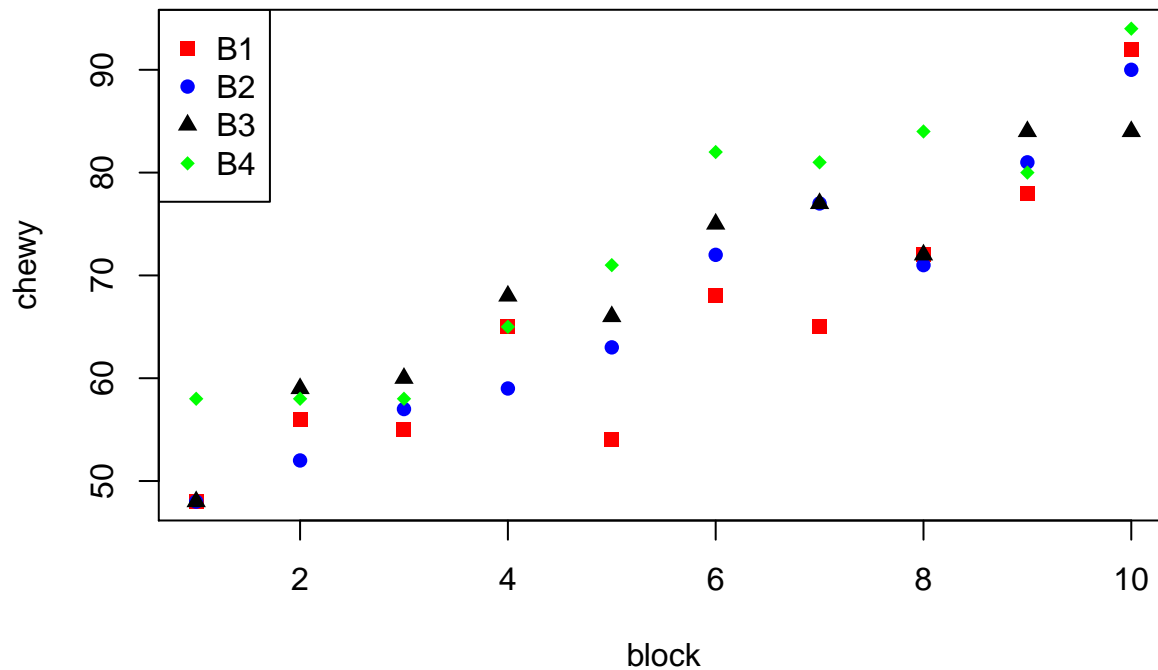
```
##### taking days to be blocks #####
```

```
# A plot of the data by days indicates possible significant difference in chewiness
# across brands.
```

```
# The ANOVA table indicates significance difference in brand means and day means.
# Tukey comparisons indicates brands 1 and 2 are not significantly difference,
# but brand 4 is significantly different from 1 and 2. Brand 3 is not significantly
# different from any of the other brands.
```

```
# plot data, x=days, y=chewiness, color=brand
```

```
plot(block,chewy,col=c("red","blue","black","green")[brand],pch = c(15,16,17,18)[brand])
legend("topleft",legend = c("B1","B2","B3","B4"),col = c("red","blue","black","green"),
      pch = c(15,16,17,18))
```

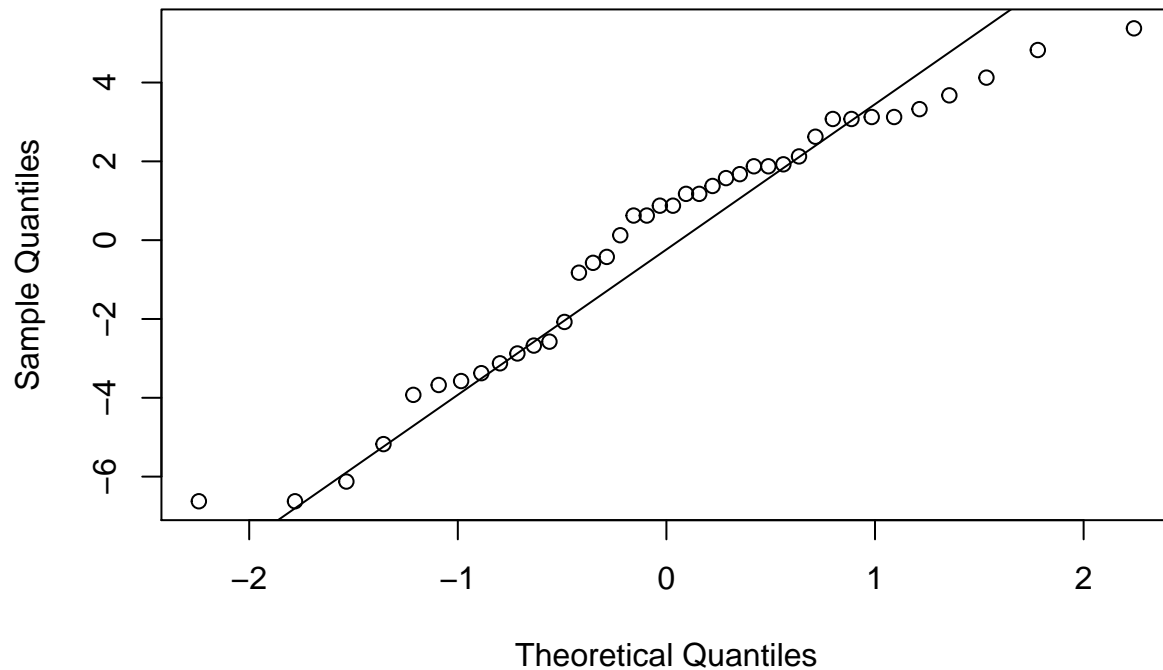


```
# fit block model
fitB <- lm(chewy ~ brandF + blockF, data=Data)
anova(fitB)

## Analysis of Variance Table
##
## Response: chewy
##          Df Sum Sq Mean Sq F value    Pr(>F)
## brandF    3  341.7   113.89   7.5171 0.0008233 ***
## blockF    9 5458.0   606.45  40.0271 2.544e-13 ***
## Residuals 27  409.1    15.15
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

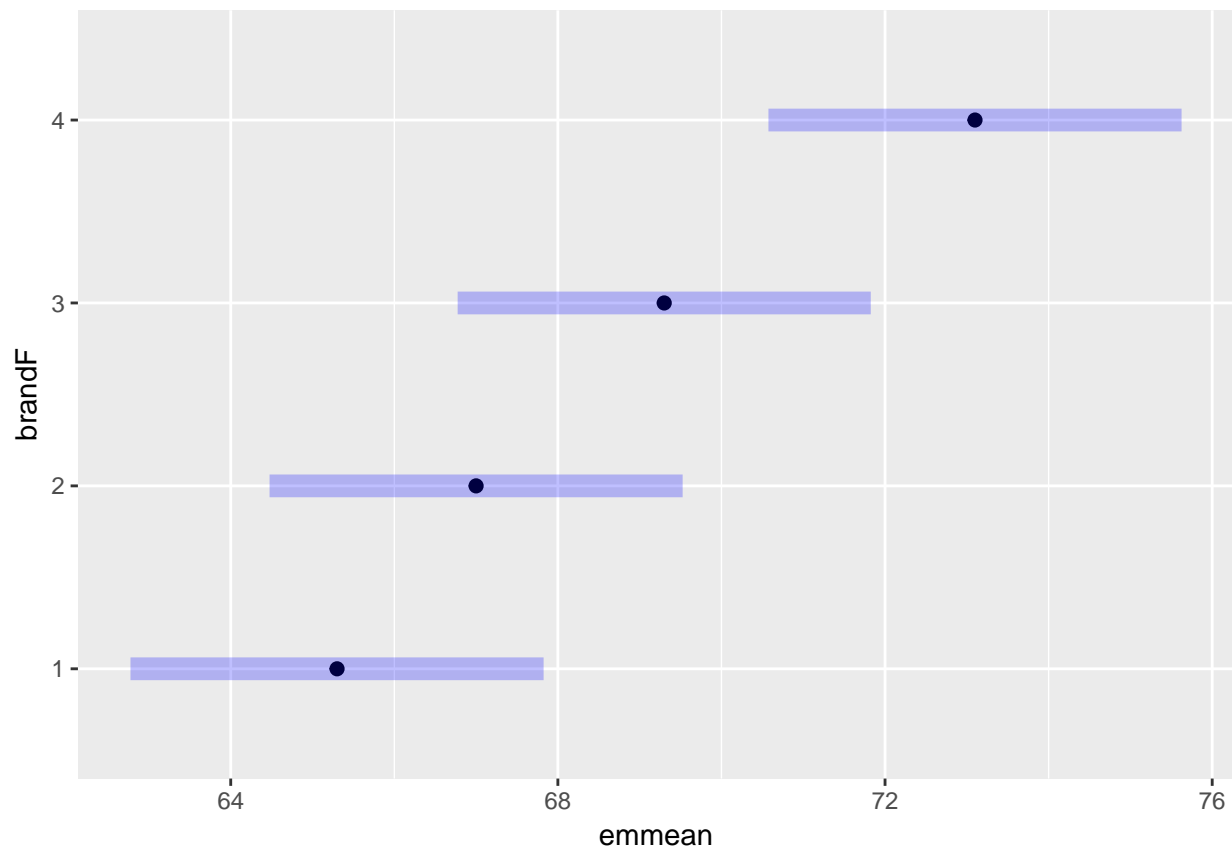
# check assumptions
qqnorm(fitB$residuals)
# compare to normal
qqline(fitB$residuals)
```

## Normal Q-Q Plot



```
# run tukey comparisons on new model.
fitB_e <- emmeans(fitB, "brandF") #estimate means from fit and factor=brand
pairs(fitB_e) # intervals for pairs
```

```
## contrast estimate      SE df t.ratio p.value
## 1 - 2          -1.7 1.740743 27  -0.977  0.7638
## 1 - 3          -4.0 1.740743 27  -2.298  0.1236
## 1 - 4          -7.8 1.740743 27  -4.481  0.0007
## 2 - 3          -2.3 1.740743 27  -1.321  0.5578
## 2 - 4          -6.1 1.740743 27  -3.504  0.0083
## 3 - 4          -3.8 1.740743 27  -2.183  0.1536
##
## Results are averaged over the levels of: blockF
## P value adjustment: tukey method for comparing a family of 4 estimates
plot(fitB_e) # plot means and std errors
```



```
# output comparison groups
cld(fitB_e,adjust="Tukey")
```

```
## brandF emmean SE df lower.CL upper.CL .group
## 1 65.3 1.230891 27 62.01586 68.58414 1
## 2 67.0 1.230891 27 63.71586 70.28414 1
## 3 69.3 1.230891 27 66.01586 72.58414 12
## 4 73.1 1.230891 27 69.81586 76.38414 2
##
## Results are averaged over the levels of: blockF
## Confidence level used: 0.95
## Conf-level adjustment: sidak method for 4 estimates
## P value adjustment: tukey method for comparing a family of 4 estimates
## significance level used: alpha = 0.05
```