Biometry Midterm.R

christieburris

Sun Feb 25 12:34:29 2018

1a. When conducting an experiment on the chewiness of brownies, the following factors arise. Note that we are assuming the brownies are made from scratch. One can also consider factors 4-8 if the brownies are made from a box mix.

1. Type and amount of sugar (cane, brown sugar, etc.)
2. Type and amount of eggs (egg whites, whipped eggs, etc.)
3. Type and amount of flour (cake flour, bread flour, all purpose, whole grain, gluten free, etc.)
4. Type and amount of fat (assorted vegetable oils, butter, apple sauce, etc.)
5. Type and dimensions of pan (metal, glass, etc.)
6. Temperature of oven
7. Length of cook time
8. Variation amongst ovens

1b. Suppose we want to know the affect of chewiness of different types of fat used in the recipe. Keeping all other ingredients constant, we choose to vary the type of fat in four different ways.

1. Vegetable oil
2. Vegetable oil and butter
3. Applesauce and butter
4. Applesauce

In this study we use one recipe for all brownie batches, varying only the type of fat. We make 10 batches of each of the 4 recipe variations. To ensure control of the other steps in the baking process, one person cooks every batch while observers carefully check every step is done the same as the last. We use 4 8x8 metal pans and cook one of each of the brownie variations at a time. The oven temperature and cook time are constant for each trial. An oven thermometer is used to ensure over temperature is constant throughout the baking process. Once the brownies are cooked, they are cooled for 20 minutes before tested for chewiness from the center of the batch (the rest of the batch is donated).

1c. This method is time consuming and requires judges stay up all day eating brownies. Instead we wish to bake the 10 batches of brownies at the same time in various ovens around town. People are randomly chosen from census data to participate in the study. The 10 ovens are now considered to be blocks in our design as ovens vary in ways that might affect the brownie chewiness. All other elements of the baking and testing process are the same.

2a. I consider Dr. Smith’s design to be a randomized block design since he does not indicate how the baking process varies across days. It is possible that the process changes between days assuming he eats the brownies soon after baking them each time.

2b. I am concerned that the baking process is not entirely constant across the 40 trials. For instance, he does not mention whether the follows the recipes given on the boxes exactly each time, or if there is variation (perhaps he runs out of a particular ingredient and makes a substitute). Does he cook the brownies for the same length of time? Does he cut the piece he is test for chewiness from the same place in the batch each time? Does he let the brownies cool for the same amount of time each day? More information is needed on the baking process in order to test for difference in chewiness across brands.

2c. The data is analyzed first as a completely randomized design and then as a randomized block design. It was originally unclear before plotting the data whether days should be considered as blocks. The hypotheses for this test are the following.

*Null hypothesis:* The mean difference in chewiness across brands is not significant. *Alternative hypothesis:* The differences in mean are significant.

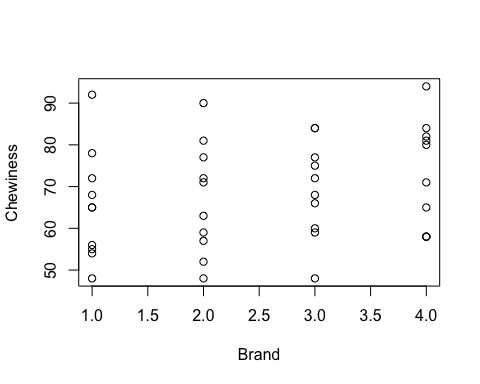
A check of assumptions tells us that our fit residuals are normal. The ANOVA table for the completely randomized design indicates no significant difference in chewiness (p=.5589, df=3). Tukey’s multiple comparisons test confirms the lack of significance (p\_ij>.5).

Considering the experiment as a randomized block design gives us different results. We first plot the data as chewiness varies by day and notice that the day dramatically affects chewiness (increasing chewiness trend). Our new model is

Chewiness = brand + day + error

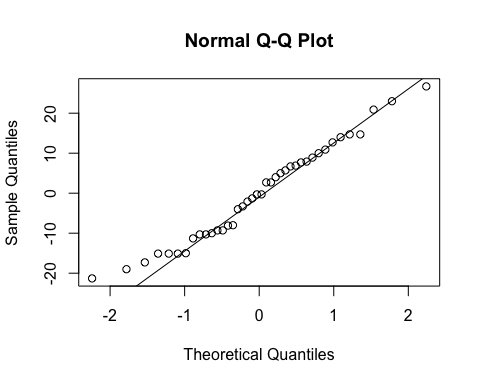
A check of assumptions tells us that our fit residuals are slightly off from normal, but not by much. Maintaining the same hypotheses, we run the ANOVA and Tukey comparisons again and see that both brand and days are significant indicators of chewiness (brand: p=.0008 df=3, day: p=2e-13 df=9*).* The Tukey comparisons test 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.

# 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. However, visualizing the data tells us that the variance amongst each brand is very high.  
# Tukey's multiple comparisons comfirms lack of significance.  
# We instead decide that a randomized block design better describes the experiment.  
  
# read data  
library(readr)  
Data <- read.csv("~/Desktop/Biometry/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 bands  
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, assumptions are met  
qqnorm(fit$residuals)  
# compare to normal  
qqline(fit$residuals)   
  


# Tukey's multiple comparisons   
library(multcompView)

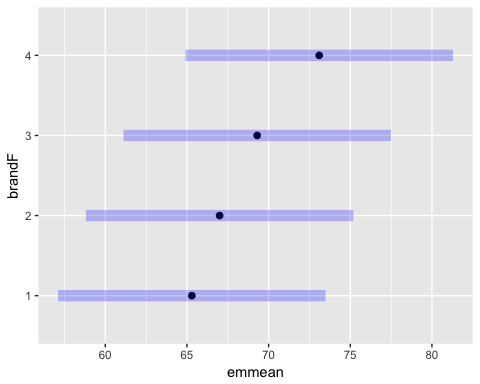
library(emmeans)

## 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

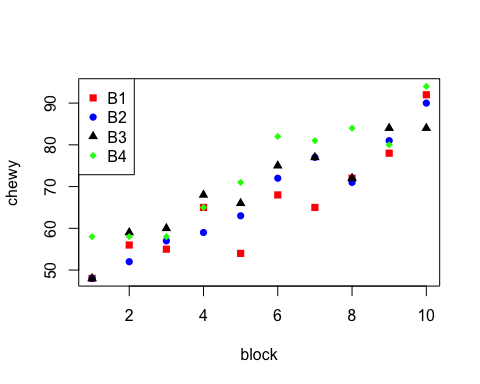
plot(fit\_e) # plot means and std errors



# 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

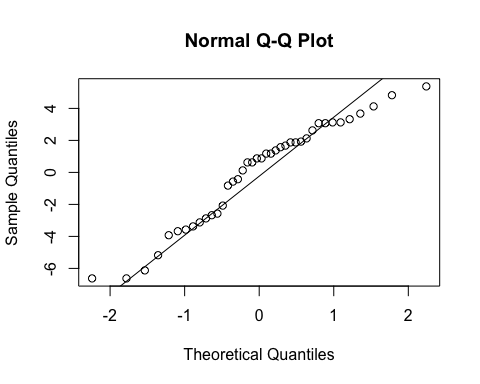
######### taking days to be blocks ##########  
# A plot of the data by days indicates possible significant difference in chewiness across brands.  
# The ANOVA table indicates significance difference in brand means and day means.   
# Tukey comparisons indicates brands 1 and 2 are not significantly 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

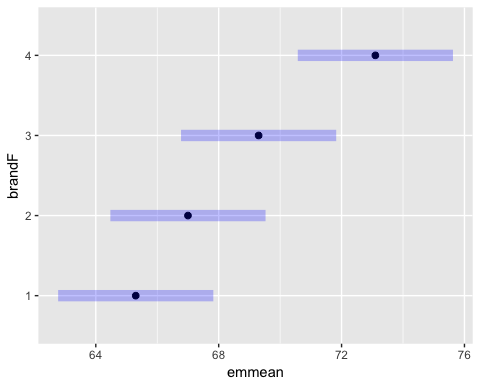
# check assumptions  
qqnorm(fitB$residuals)  
# compare to normal  
qqline(fitB$residuals)



# 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