THEORETICAL PORTION

PROBLEM #1.

long-lasting batteries: X ~ Exponential(1000, 1000^2)
short-term batteries: X ~ Gamma(500, 200)

a. For the short-term batteries, the population mean (μ) is 500 while the population variance (σ^2) is 200. Given the variance, we know that the population standard deviation (σ) is 14.1421.
The scale parameter is 200/500 = 0.4 (The variance divided by the mean, giving a rate of 2.5)
The shape parameter is 500/0.4 = 1250 (The mean divided by the scale parameter)

b. Observing a joint probability distribution, the probabilities of the two distributions knowing the ratio is a 4:1 (long-lasting:short-term).
Long-Lasting Batteries, exponential distribution, β = 1000. The probability that these batteries last at least 490 is a right-tailed test. Thus, P(X ≥ 490) = $e^{-490/1000}$ = 0.612626… => 61.26%
Short-Term Batteries, gamma distribution, required a bit more math so I ended up using a gamma calculator with the integral below (with respect to *x*).

$$\int_{490}^{\infty} \frac{1}{\Gamma(a)b}\left(\frac{x}{b}\right)^{a-1} e^{-\frac{x}{b}}$$

The integral for the short-term batteries equates to 0.758795… => 75.88%

To find the probability of that battery lasting at least 490 hours:

$$\frac{4(.6126)}{5} + \frac{.7588}{5} = .49008 + .15176 = .64184$$

The probability that the battery lasts at least 490 hours is ~**64.18%**

c. The probability that the battery has lasted for 490 hours (so at least 490) AND is a short-term battery:

Using the short-term battery probability from *part b* of .7588 and the probability that of the battery selected being a short-term battery is 1 in 5 (.2).
Multiplying these probabilities together gives us .15176 ~ **15.18%**.

Then factoring that the battery is in fact a short-term battery that has lasted at least 490, dividing the above by the probability found in part b of .64184.
.15176 / .64184 = .23645 ~ **23.65%**

The above was approached using Bayes Theorem. The probability that the battery in-hand has lasted for 490 hours and is also a short-term battery is approximately 23.65%.

PROBLEM #2.

a. $$\int_0^1 cx^2\,dx = \left[c\frac{x^3}{3}\right]_0^1 = \frac{c}{3} - 0$$

$$\vdots$$

$$c = 3$$

b. $$E(Y) = E(X^3) = \int_0^1 x^3 * cx^2\,dx = \int_0^1 cx^5 dx = \left[\frac{cx^6}{6}\right]_0^1 = \frac{3(1)^6}{6} - 0 = .5$$

c.
$$\begin{aligned}
&Var(Y) = Var(X^3)\\
&= \int_0^1 \left(x^3 - \frac{1}{2}\right)^2 3x^2\,dx\\
&= \int_0^1 \left(x^6 - x^3 + \frac{1}{4}\right) 3x^2\,dx\\
&= \int_0^1 \left(3x^8 - 3x^5 + \frac{3x^2}{4}\right)\,dx\\
&= \left(\frac{3x^9}{9} - \frac{3x^6}{6} + \frac{3x^3}{12}\right)\Big|_0^1\\
&= \left(\frac{1}{3} - \frac{1}{2} + \frac{1}{4}\right) - 0\\
&= \left(\frac{4}{12} - \frac{6}{12} + \frac{3}{12}\right) - 0 = \frac{1}{12}
\end{aligned}$$

PROBLEM #3.

a. The hypothesis to test is if more than 80% of Americans are right-handed
   H_0: p = 80% H_1: p > 80%
b. Sample size > 30 and using a significant level of .05, z-score for the test statistic:

$$p = 80\% \qquad \hat{p} = \frac{419}{500} = .838 \Rightarrow 83.8\% \qquad n = 500$$

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{.838 - .80}{\sqrt{\frac{.80(1-.80)}{500}}} = \frac{.038}{.01789} = 2.1243 \Rightarrow 2.12$$

   The p-value is then = P(Z > 2.12) = 1 - .9830 = **.017**
   (used the Z-table from http://www.stat.ufl.edu/~athienit/Tables/Ztable.pdf)

c. Since the p-value is ≤ the significant level (α) of .05, we **reject H_0**.
   There is significant evidence to support that more than 80% of Americans are right-handed, with a significant level of .05.

COMPUTATIONAL PORTION

Code and graphs included with each part, not indexed.

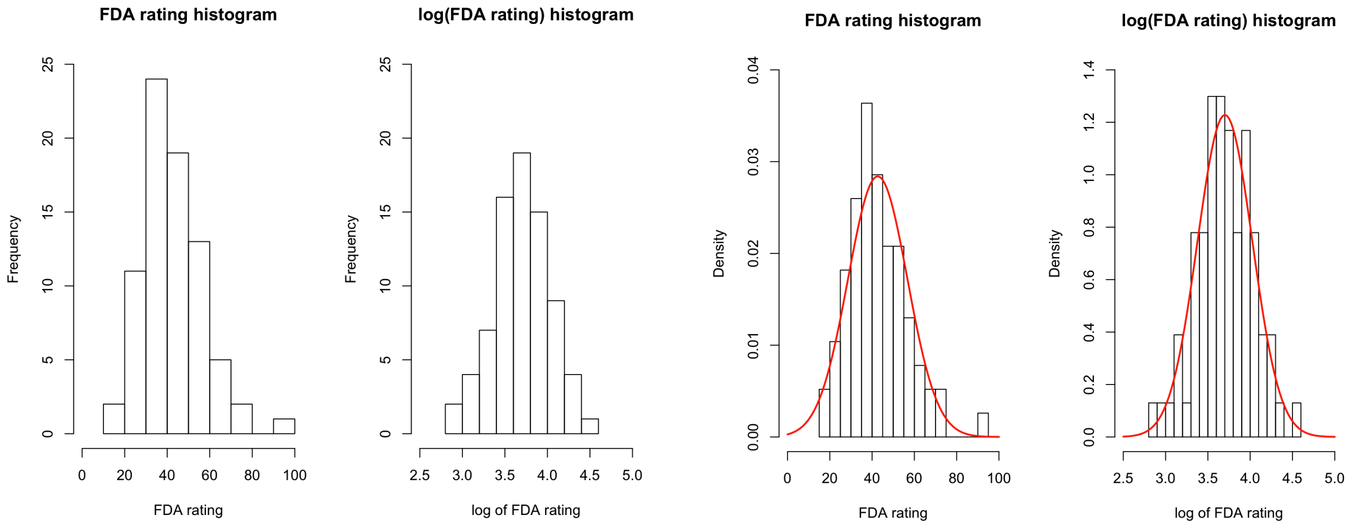
PROBLEM #1 - COMPUTATIONAL

#1, part *a.*

code:

```
cerealData=read.table("cerealdata.txt", header=TRUE)

par(mfrow=c(1,2))

hist(cerealData$rating, main="FDA rating histogram", xlab="FDA rating", ylim=c(0,25),xlim=c(0,100))
hist(log(cerealData$rating), main="log(FDA rating) histogram", xlab="log of FDA rating",
        ylim=c(0,25),xlim=c(2.5,5))

hist(cerealData$rating, main="FDA rating histogram", xlab="FDA rating", ylim=c(0,.04),xlim=c(0,100),
        breaks=20, freq = FALSE)
curve(dnorm(x,mean(cerealData$rating),sd(cerealData$rating)),add=TRUE,col="red",lwd=2)

hist(log(cerealData$rating), main="log(FDA rating) histogram", xlab="log of FDA rating",
        ylim=c(0,1.4),xlim=c(2.5,5), breaks=20, freq=FALSE)
curve(dnorm(x,mean(log(cerealData$rating)),sd(log(cerealData$rating))),add=TRUE,col=“red",lwd=2)
```

output:



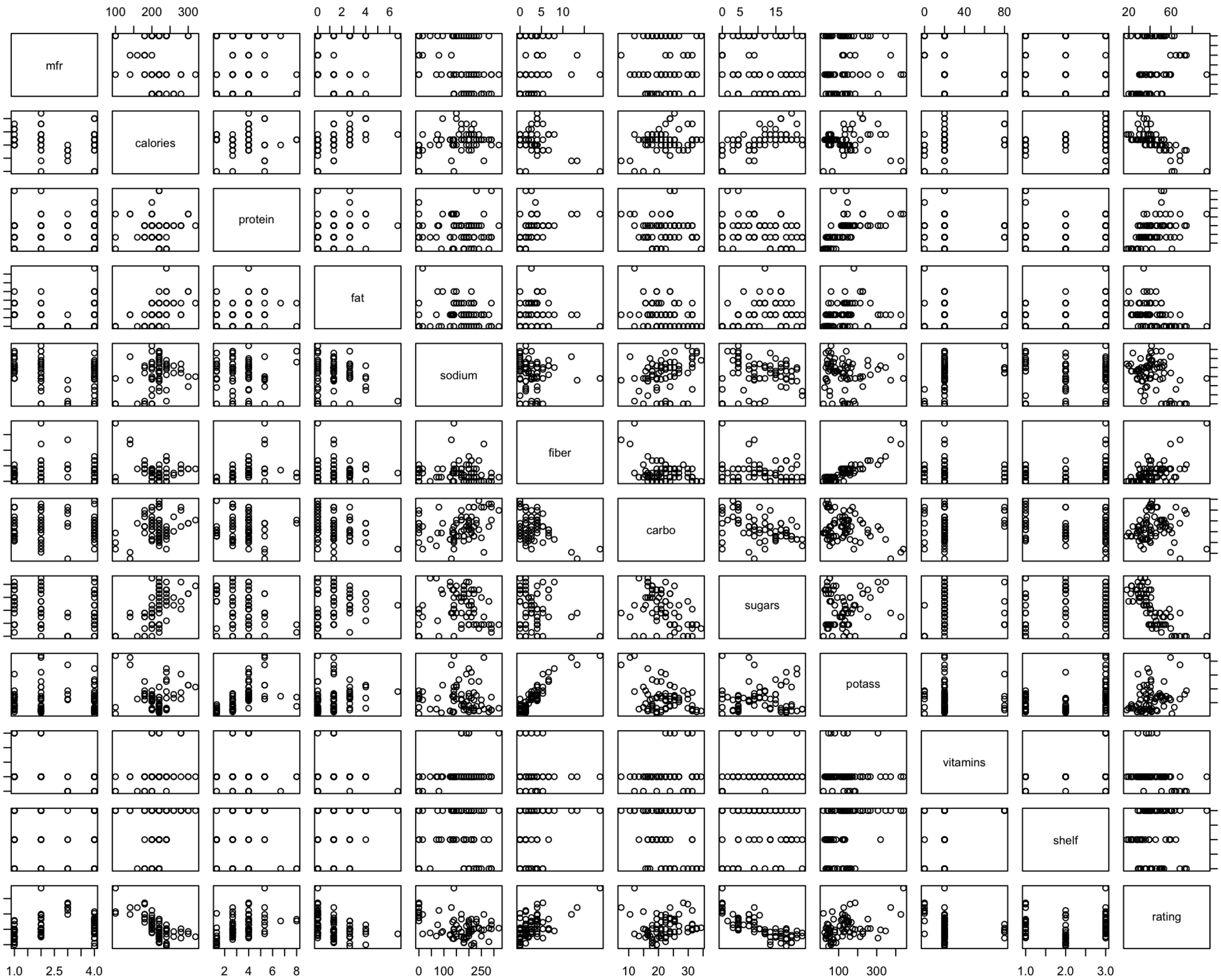
The model which looks more suitable is the “log(rating)” histogram because it appears more normally distributed than the “rating” histogram. The log data looks more ideal for the linear model since it fits a better normal distribution curve,

#1, part b.

code:

```
pairs(cerealData[,1:12])
```

output:



The following predictor variables look to be correlated based on the scatterplots above:

- Potassium & fiber linear
- Calories & rating linear
- Calories & sugar somewhat linear
- Carbs & sugar slightly linear
- Sodium & carbs slightly linear
- Sugars & calories slightly linear
- Carbs & potassium slightly linear
- Calories & carbs slightly linear

#1, part c.

$$\hat{Y}_{rating} = \beta_0 + \begin{array}{c} \begin{array}{cccccccccccccc} \beta_{MK} & \beta_{MN} & \beta_{MO} & \beta_{cal} & \beta_{pro} & \beta_{fat} & \beta_{sod} & \beta_{fib} & \beta_{car} & \beta_{sug} & \beta_{pot} & \beta_{vit} & \beta_{sh2} & \beta_{sh3} \end{array} \\ \begin{bmatrix} 0 & 1 & 0 & 140 & 5.33 & 1.33 & 130 & 13.3 & 7.46 & 8.96 & 373 & 20.0 & 0 & 1 \\ 0 & 0 & 1 & 240 & 4.00 & 6.67 & 15 & 2.67 & 11.9 & 11.9 & 180 & 0 & 0 & 1 \\ 1 & 0 & 0 & 140 & 5.33 & 1.33 & 260 & 12.0 & 10.4 & 7.46 & 427 & 20 & 0 & 1 \\ 1 & 0 & 0 & 100 & 5.33 & 0 & 140 & 18.7 & 11.9 & 0 & 440 & 20 & 0 & 1 \\ 0 & 0 & 1 & 220 & 2.67 & 2.67 & 200 & 1.33 & 20.9 & 11.9 & NA & 20 & 0 & 1 \\ 0 & 0 & 0 & 220 & 2.67 & 1.33 & 250 & 0 & 31.3 & 4.48 & 80.0 & 20 & 0 & 1 \\ 0 & 0 & 0 & 220 & 1.33 & 1.33 & 140 & 0 & 19.4 & 17.9 & 33.3 & 20 & 1 & 0 \\ 0 & 0 & 1 & 200 & 4.00 & 1.33 & 230 & 4.00 & 25.4 & 4.48 & 153 & 20 & 0 & 0 \\ 0 & 0 & 0 & 200 & 4.00 & 1.33 & 200 & 4.00 & 25.4 & 4.48 & 146 & 20 & 0 & 0 \\ 0 & 0 & 0 & 220 & 2.67 & 1.33 & 200 & 1.33 & 23.9 & 11.9 & 80.0 & 20 & 0 & 0 \end{bmatrix} \end{array} * \begin{bmatrix} \beta_{MK} \\ \beta_{MN} \\ \beta_{MO} \\ \beta_{cal} \\ \beta_{pro} \\ \beta_{fat} \\ \beta_{sod} \\ \beta_{fib} \\ \beta_{car} \\ \beta_{sug} \\ \beta_{pot} \\ \beta_{vit} \\ \beta_{sh2} \\ \beta_{sh3} \end{bmatrix}$$

#1, d, part i.

$$\hat{Y}_{rating} = 3.885 + .030X_{MK} - .018X_{MN} + .020X_{MO} - .003X_{cal} + .079X_{pro} - .023X_{fat} - .001X_{sod} + .045X_{fib} + .023X_{car} - .005X_{sug} - .0004X_{pot} - .001X_{vit} - .051X_{sh2} + .018X_{sh3}$$

#1, d, part ii.

Sorting through the data, the values that were the smallest disregarding NA. The average rating for this is:

3.885 - .003(100) + .079(1.33)-0-0+0+.023(7.467)-0-.0004(20)-0 = 3.853811 -> which this is the log of it so $e^{3.853811}$ = 47.1725

The predicted average rating of the GM cereal that sits on the bottom shelf is 47.1725. The value could be different due to the amount of rounding from the above data.

Since this is an inaccurate version, I computed it in R to reduce errors due to round:

```
summary(cD_lm_logRating)$coef[1,1]+
            summary(cD_lm_logRating)$coef[5,1]*min(cerealData$calories, na.rm=remove)+
            summary(cD_lm_logRating)$coef[6,1]*min(cerealData$protein, na.rm=remove)+
             summary(cD_lm_logRating)$coef[7,1]*min(cerealData$fat, na.rm=remove)+
             summary(cD_lm_logRating)$coef[8,1]*min(cerealData$sodium, na.rm=remove)+
            summary(cD_lm_logRating)$coef[9,1]*min(cerealData$fiber, na.rm=remove)+
            summary(cD_lm_logRating)$coef[10,1]*min(cerealData$carbo, na.rm=remove)+
            summary(cD_lm_logRating)$coef[11,1]*min(cerealData$sugars, na.rm=remove)+
            summary(cD_lm_logRating)$coef[12,1]*min(cerealData$potass, na.rm=remove)+
            summary(cD_lm_logRating)$coef[13,1]*min(cerealData$vitamins, na.rm=remove)
```

This gives a log value of 3.814114 which $e^{3.814114}$ = **45.3366** for the average rating for GM cereal that sits on the bottom shelf.

#1, d, part iii.

code:

```
cereal_fit2=lm(log(rating)~as.factor(cerealData$mfr), data=cerealData)
summary(cereal_fit2)
```

partial output:

```
                                 Estimate
(Intercept)                      3.50723
as.factor(cerealData$mfr)K       0.23641
as.factor(cerealData$mfr)N       0.70901
as.factor(cerealData$mfr)O       0.20321
```

The manufacturer does appear to have a significant with Nabisco tending to score higher than the other three, based on it's higher estimate value. Though this may appear to be a biased judgement, I do think it is appropriate since brands may tend to produce more unhealthy cereal than other brands.

#1, d, part iv.

code:

```
cereal_fit3=lm(log(rating)~as.factor(cerealData$shelf), data=cerealData)
summary(cereal_fit3)
```

output:

```
                                 Estimate
(Intercept)                      3.79278
as.factor(cerealData$shelf)2     -0.30815
as.factor(cerealData$shelf)3     -0.01504
```

The shelf that the cereal sits on seems to to be most influential for the middle shelf, which tends to decrease in rating. I do not think this is appropriate since which shelf the cereal sits on should not effect it's rating. However, one could argue that the middle shelf tends to be the more unhealthy shelf since it's eye-level to most consumers (as children).

#1, e, part i.

$$\hat{Y}_{rating} = 3.875 - .003X_{cal} + .075X_{pro} - .031X_{fat} - .001X_{sod} + .036X_{fib} + .021X_{car} - .009X_{sug} - .001X_{vit}$$

#1, e, part ii.

The average rating for this was computed as before with:

```
summary(cD_smaller_fitted)$coef[1,1]+
          summary(cD_smaller_fitted)$coef[2,1]*min(cerealData$calories, na.rm=remove)+
          summary(cD_smaller_fitted)$coef[3,1]*min(cerealData$protein, na.rm=remove)+
          summary(cD_smaller_fitted)$coef[4,1]*min(cerealData$fat, na.rm=remove)+
          summary(cD_smaller_fitted)$coef[5,1]*min(cerealData$sodium, na.rm=remove)+
          summary(cD_smaller_fitted)$coef[6,1]*min(cerealData$fiber, na.rm=remove)+
          summary(cD_smaller_fitted)$coef[7,1]*min(cerealData$carbo, na.rm=remove)+
          summary(cD_smaller_fitted)$coef[8,1]*min(cerealData$sugars, na.rm=remove)+
          summary(cD_smaller_fitted)$coef[9,1]*min(cerealData$vitamins, na.rm=remove)
```

This average estimation was somewhat similar as same as the larger model, with an average rating of 3.836277. Since this was logged, $e^{3.836277}$ = **46.3526**. This value makes sense since the potassium value is not being subtracted as well as not factoring the the lesser value of the GM brand (on average).

#1, e, part iii.

The average rating change per 1 gram of fat decreases by a factor of 0.031.

#1, e, part iv.

The percent of the variation in ratings that is explained by this smaller model is 96.29%, given by the lm function in R. To recap from earlier, this entails taking the log of the rating.

#1, part f.

To test the significance of removing predictors from the full model to the smaller model, we observe the rating change when variables are absent vs. present.

H_0: $\beta_{MK} = 0^{and}\ \beta_{MN} = 0^{and}\ \beta_{MO} = 0^{and}\ \beta_{cal} = 0^{and}\ \beta_{pro} = 0^{and}\ \beta_{fat} = 0^{and}\ \beta_{sod} = 0^{and}\ \beta_{fib} = 0^{and}$
$\beta_{car} = 0^{and}\ \beta_{sug} = 0^{and}\ \beta_{pot}= 0^{and}\ \beta_{vit} = 0^{and}\ \beta_{sh2} = 0^{and}\ \beta_{sh3} = 0$
H_1: $\beta_{MK} \neq 0^{or}\ \beta_{MN} \neq 0^{or}\ \beta_{MO} \neq 0^{or}\ \beta_{cal} \neq 0^{or}\ \beta_{pro} \neq 0^{or}\ \beta_{fat} \neq 0^{or}\ \beta_{sod} \neq 0^{or}\ \beta_{fib} \neq 0^{or}$
$\beta_{car} \neq 0^{or}\ \beta_{sug} \neq 0^{or}\ \beta_{pot}\neq 0^{or}\ \beta_{vit} \neq 0^{or}\ \beta_{sh2} \neq 0^{or}\ \beta_{sh3} \neq 0$

Observing the linear model of the full model, we find the F-statistic to be 143.4 with a p-value of $4.125\text{x}10^{-40}$. For the smaller model, the F-statistic is 217.3 with a p-value $8.034\text{x}10^{-45}$ (as well).

#1, part g.

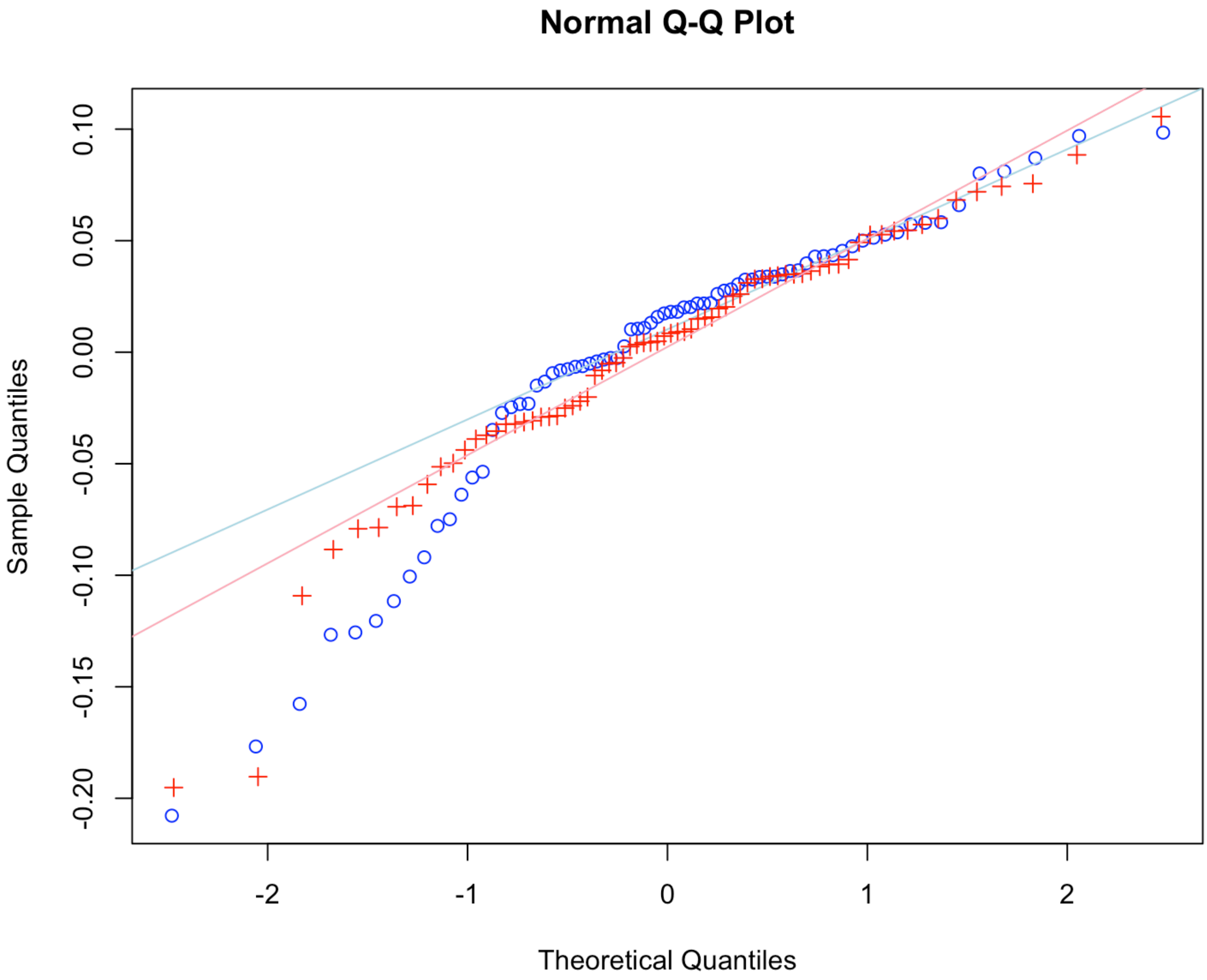
Smaller model variation percentage - 96.29%
Full model variation percentage - 97.15%

Even though the R^2 value is better for the larger model, I believe the smaller model to be the better one because of the potential biased parameters in the larger model. The rating should not be effected by the brand or shelf position, purely on nutritional facts/values.

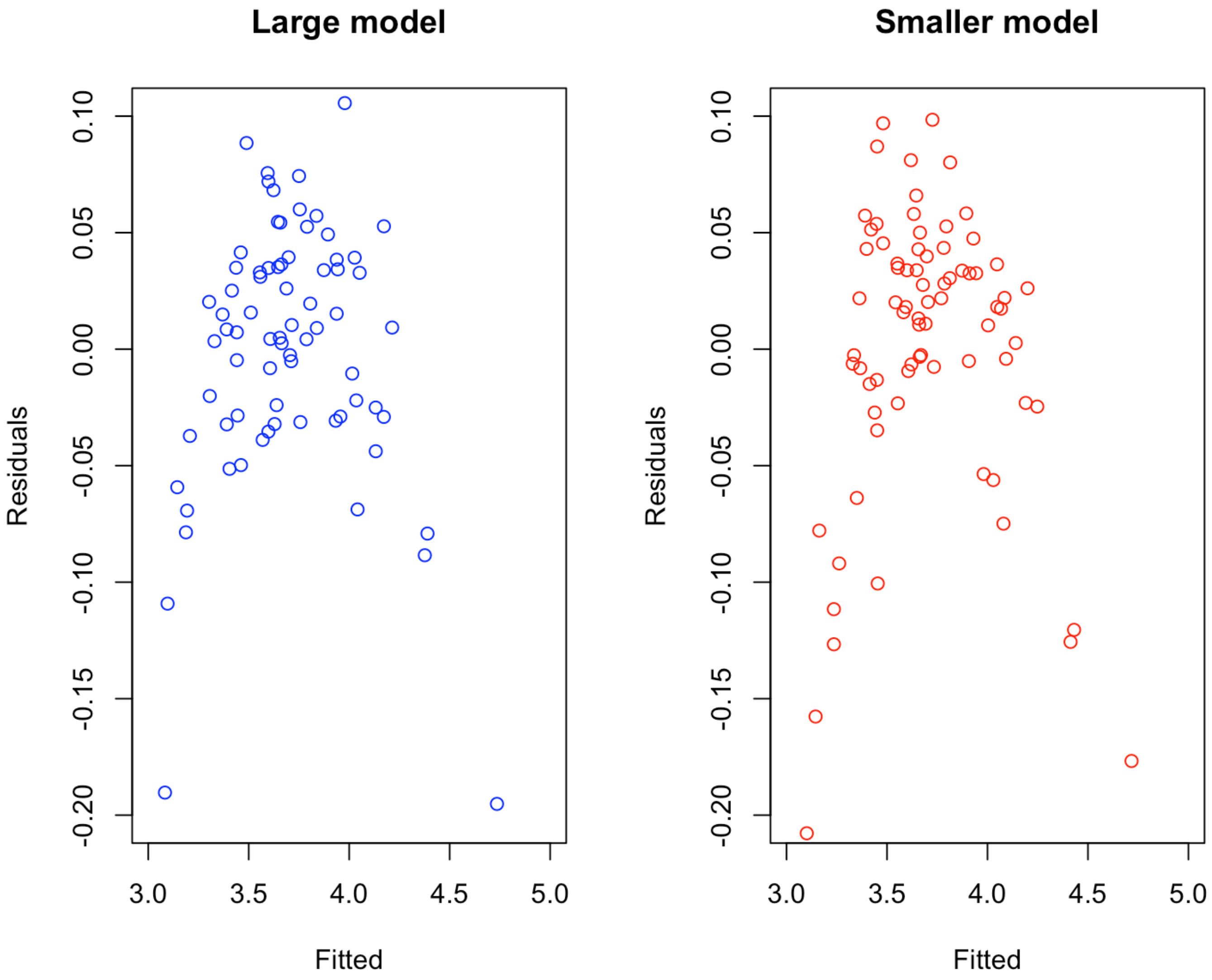
#1, part h.

A couple of ways that the model satisfies the modeling assumptions is to observe a plot of the residuals. We can also check for homoscedasticity, error distribution normality, and independence among the errors. Additionally, checking for the linearity between the variables is another method for checking correct modeling assumptions.

Firstly, lets look a plot of residuals for the full model against the smaller model. We can note that the fitted line fits better for the smaller model vs. the full model.



Secondly, we can look at the fitted values vs. the residuals for both models.



PROBLEM #2 - COMPUTATIONAL

#2, part a*.*

The students who didn't take pre-calculus in high-school are the baseline for the model. This occurs since x_i is zero and the combination with β_1 of 20 becomes NULL. Thus, $Y_i = 60 + 0 + \varepsilon_i$ is the model for students who didn't take pre-cal in HS.

#2, part b*.*

β_o - the mean test score (y-intercept) for the students who didn't take pre-calculus in high school
β_1 - the increase in test score if the student took pre-cal in high school
ε_i - factors in the error amount from the regression line for the *ith* student

#2, part c*.*

75% have had pre-cal before
25% [$x_{HS\ pre\text{-}cal}$] and 25% [$x_{no\ HS\ pre\text{-}cal}$]

As interpretation, I setup the coding so that it would generate an n_x3 matrix. The *n* value represented the number of students while the 1st column represented the β_o variable. The 2nd column represented the β_1 variable with a random sample generated with 75% probability that the student had taken pre-cal in high school. The additional variable added to the original regression model was the third column, β_2, which represented a weighted group based on an even number of x_i vs non-x_i students. That is, students who took pre-cal in HS and those that had not. It created an equal number of each to be in the group, not to exceed half of the total number of students. So for example, if 75% of the students (out of 100) had taken pre-cal in HS -> 25 non-pre-cal HS and 25 pre-cal HS students were in the group. If it happened to be greater than 75% of the group had prior pre-cal exposure, the intervention group size would limit itself to only double that of the number of students without pre-cal in HS. The regression model now functions as such:

$$\hat{y} = 60 + 20x_{HP} + 10x_{AG}$$

The variable x_{HP} is the indicator for if the student took pre-cal in HS and x_{AG} is the indicator variable for if the student is a part of the active group work.

code:

```
n=100
studentMatrix=matrix(1, nrow=n, ncol=3)

sumPreCal=0;sumNoPreCal=0;sumGroup=0
for(i in 1:n){
        studentMatrix[i,2]=sample(0:1, size=1, prob=c(.25,.75))

        if(sumGroup<(n/2)){
                if(sumNoPreCal<(n/4) && studentMatrix[i,2]==0){ #0 is didn't have pre-cal in HS
                        studentMatrix[i,3]=1  #1 is representing no HS Pre-cal class and select for active group work
                        sumNoPreCal=sumNoPreCal+1
                        sumGroup=sumGroup+1
                }else if(sumPreCal<sumNoPreCal && studentMatrix[i,2]==1){ #1 is had pre-cal in HS
                         studentMatrix[i,3]=2 #2 is representing HS Pre-cal class and select for active group work
                        sumPreCal=sumPreCal+1
                        sumGroup=sumGroup+1
                }else
                        studentMatrix[i,3]=0 #0 - don't select this student for active group work
        }else if(sumNoPreCal!=sumPreCal && sumGroup<(n/2)){   #to make sure each student-base even percentage-wise
```

```
                if(sumPreCal>=sumNoPreCal && studentMatrix[i,2]==0){ #0 is didn't have pre-cal in HS
                        studentMatrix[i,3]=1  #1 is representing no HS Pre-cal class and select for active group work
                        sumNoPreCal=sumNoPreCal+1
                        sumGroup=sumGroup+1
                }else if(sumPreCal<=sumNoPreCal && studentMatrix[i,2]==1){ #1 is had pre-cal in HS
                        studentMatrix[i,3]=2 #2 is representing HS Pre-cal class and select for active group work
                        sumPreCal=sumPreCal+1
                        sumGroup=sumGroup+1
                }
        }else
                 studentMatrix[i,3]=0 #0 - don't select this student for active group work
}

#data verification
summary(studentMatrix[,2])
as.data.frame(table(studentMatrix[,2]))
as.data.frame(table(studentMatrix[,3]))
sumNoPreCal;sumPreCal;sumGroup

#convert column 3 to ones and zeroes
for(i in 1:n){
        if(studentMatrix[i,3]!=0)studentMatrix[i,3]=1
}
```

#2, part d.

To perform the power for calculating the power for n students, I used:

```
pwr.2p.test(mean(studentMatrix[,3]),sig.level=.05,n=n,alternative = “two.sided")
```

which resulted in a power of 99.96%. As the number of students for the test decreased, so did the power and thus the ability to correctly reject the null hypothesis also decreased. The student size for this test was at 100, as listed above. The random sample also generated 75% of students which had taken pre-calculus in high school.

#2, part e.

I had initially calculated a critical value and type II error using a .05 significant level until discovering the power function above. It also made the approach for the next part easier as well.

The initial code:

```
x_critical_.05=matrix(0,nrow=n,ncol=1)
type_II_.05=matrix(0,nrow=n,ncol=1)
for(i in 1:n){
        x_critical_.05[i]=(qnorm(.05)*(20/sqrt(n)))+(60+((studentMatrix[i,2])*20)+(studentMatrix[i,3]*10))
        type_II_.05[i]=(x_critical_.05-(60+((studentMatrix[i,2])*20)+(studentMatrix[i,3]*10)))/(20/sqrt(n))
}
type_II_.05
x_critical_.05

1-mean(pnorm(type_II_.05))
```

#2, part f.

The test below used the same data above, except without a set student sample size.

```
pwr.2p.test(mean(studentMatrix[,3]),sig.level=.05,power=.8,alternative = ″two.sided″)
```

This test concluded an *n* value of 27.91 or 28 students to detect their hypothesized difference of 10 at an 80% power.

APPENDIX HOSTMATH

T#1 part B:

```
\int_{490}^{\infty} \frac{1}{\Gamma(a)b}\Big(\frac{x}{b}\Big)^{a-1}e^{-\frac{x}{b}}
```

T#2 part:

```
E(Y) = E(X^3) = \int_{0}^{1}x^3*cx^2s\space dx =  \int_{0}^{1}cx^5dx = \large\begin{bmatrix}\frac{cx^6}{6}\end{bmatrix}\scriptsize\begin{matrix}
1\\0\end{matrix}\small=\frac{3(1)^6}{6} - 0 = .5
```

T#2 part C:

```
Var(Y) = Var(X^3)\\=
\int_{0}^{1}\left(x^3-\frac{1}{2}\right)^2 3x^2\ dx\\
=\int_{0}^{1}\left(x^6-x^3+\frac{1}{4}\right) 3x^2\ dx\\
=\int_{0}^{1}\left(3x^8-3x^5+\frac{ 3x^2}{4}\right)\ dx\\
=\left(\frac{3x^9}{9} - \frac{3x^6}{6} + \frac{3x^3}{12}\right) \  \Bigl_{0}^1 \\
= \left(\frac{1}{3} - \frac{1}{2} + \frac{1}{4}\right) - 0 \\
= \left(\frac{4}{12} - \frac{6}{12} + \frac{3}{12}\right) - 0 = \frac{1}{12}
```

T#3 part B:

```
p = 80\%
\qquad\hat p = \frac{419}{500}=.838\Rightarrow 83.8\%
\qquad n=500
\\\ \\
z=\frac{\hat p - p_{0}}{\sqrt{\frac{p_0(1-p_0)}{n}}} =

\frac{ .838 - .80}{\sqrt{\frac{.80(1-.80)}{500}}} = \frac{.038}{.01789} = 2.1243 \Rightarrow 2.12
```

C#1 part C:

```
\begin{matrix}\scriptsize
\beta_{MK}\space\space
\beta_{MN}\space\space
\beta_{MO}\space\space\space\space\space
\beta_{cal}\quad\space\space
\beta_{pro}\quad\space\space
\beta_{fat}\quad\space\space
\beta_{sod}\quad\space\space
\beta_{fib}\quad\space\space
\beta_{car}\quad\space\space
\beta_{sug}\quad\space\space
\beta_{pot}\quad\space\space
\beta_{vit}\quad\space\space
\beta_{sh2}\space\space
\beta_{sh3}\space\space\ \ \ \ \ \
\end{matrix}

\small
\\\hat{Y}_{rating}=\beta_0 + \begin{bmatrix}
0&1&0&140&5.33&1.33&130&13.3&7.46&8.96&373&20.0&0&1\\
0&0&1&240&4.00&6.67&15  &2.67&11.9&11.9&180&0&0&1\\
1&0&0&140&5.33&1.33&260&12.0&10.4&7.46&427&20&0&1\\
1&0&0&100&5.33&0&140    &18.7&11.9&0&440&20&0&1\\
0&0&1&220&2.67&2.67&200&1.33&20.9&11.9&NA&20&0&1\\\\\

0&0&0& 220&2.67&1.33 &250&0&31.3     &4.48&80.0&20 &0&1\\
0&0&0& 220&1.33&1.33 &140&0&19.4     &17.9&33.3&20 &1&0\\
0&0&1& 200&4.00&1.33 &230&4.00&25.4 &4.48&153&20 &0&0\\
0&0&0& 200&4.00&1.33 &200&4.00&25.4 &4.48&146&20 &0&0\\
0&0&0& 220&2.67&1.33 &200&1.33&23.9 &11.9&80.0&20 &0&0\\
\end{bmatrix}

*

\begin{bmatrix}
```

```
\beta_{MK}\\
\beta_{MN}\\
\beta_{MO}\\
\beta_{cal}\\
\beta_{pro}\\
\beta_{fat}\\
\beta_{sod}\\
\beta_{fib}\\
\beta_{car}\\
\beta_{sug}\\
\beta_{pot}\\
\beta_{vit}\\
\beta_{sh2}\\
\beta_{sh3}\\
\end{bmatrix}

+\begin{bmatrix}
\epsilon_1\\
\epsilon_2\\
\epsilon_3\\
\epsilon_4\\
\epsilon_5\\
\epsilon_{73}\\
\epsilon_{74}\\
\epsilon_{75}\\
\epsilon_{76}\\
\epsilon_{77}\\


\end{bmatrix}
```

C#1 part e.i:

```
\hat {Y}_{rating} = 3.875 -.003X_{cal}+.075X_{pro}-.031X_{fat}-.001X_{sod}+.036X_{fib}+.021X_{car}-.009X_{sug}-.001X_{vit}
```