Practicum 1 Analysis

Peter Claussen Ben Derenge Stephanie Liebl

10/5/2021

Data Processing

Our initial step in processing the 'Practicum 1 Data' data set was to remove any rows where the `Snumber` column was "NA". We identified the `Snumber` column as a unique identifier, and thus were only concerned with nonempty rows. Initially, the data set has 392 rows. After removing removing the 'NA' identifiers, we are left with 352 rows.

```
metrics.dat <- read.csv('Practicum 1 Data.csv',header=TRUE)
metrics.dat <- metrics.dat[!is.na(metrics.dat$Snumber),]
```

Next, we verified that the column `Total_Met_Min` had correct computations. We created a separate column `CalcTMM` to check against the values in `Total_Met_Min`. In our observations of the `shift` variable, we discovered that there existed an 'other' category and missing values. We decided to refrain from combining these categories and created a 'missing' category for values of `shift` that were blank.

```
metrics.dat$CalcTMM <- with(metrics.dat, 8*Vig.ex.Time + 4*Mod.ex.time + 3.3*Walk.ex.Time)

#metrics.dat$shift[metrics.dat$shift==''] <- 'missing'
#shift.levels <- c(paste(c(7:11),'am',sep=''),paste(c(12,1:2),'pm',sep=''),'other','missing')

metrics.dat$shift[metrics.dat$shift==''] <- 'other'
shift.levels <- c(paste(c(7:11),'am',sep=''),paste(c(12,1:2),'pm',sep=''),'other')
metrics.dat$shift <- factor(metrics.dat$shift,shift.levels)
summary(metrics.dat$shift)
```

```
##   7am   8am   9am  10am  11am  12pm   1pm   2pm other
##    31   115    56    50    44    14     8    15    19
```

```
metrics.dat$MissingLbs <- is.na(metrics.dat$pounds_gained)
table(metrics.dat$MissingLbs,metrics.dat$weightgain)
```

```
##
##              No Yes
##   FALSE   0   1 231
##   TRUE    4 110   6
```

We consider two subsets for analysis. First we create a data table that has appropriate values for `weightgain`. This will be the larger of the two data sets. Our original data set contained a binary 'yes/no' column titled `weightgain`. This column was missing data for four rows, and thus, we subset the data to obtain a data

table with values for `weightgain` in all rows. This data table left us with 348 rows. It should be noted that the original data set contains a numeric `pounds_gained` column as well. This column had many missing values, but we were able to impute a '0' if we knew that row had a 'No' for `weightgain`. Lastly, we observed a single missing value for the column `CalcTMM`, and so we once more subset the data to obtain a table with values of `CalcTMM` for all rows.

```
gained.dat <- metrics.dat[metrics.dat$weightgain %in% c('Yes','No'),]
gained.dat$WG <- FALSE
gained.dat$WG[gained.dat$weightgain=='Yes'] <- TRUE
gained.dat$pounds_gained[!gained.dat$WG] <- 0
dim(gained.dat)
```

```
## [1] 348  86
```

```
gained.dat$MissingLbs <- is.na(gained.dat$pounds_gained)
table(gained.dat$MissingLbs,gained.dat$weightgain)
```

```
##
##          No Yes
##   FALSE 111 231
##   TRUE    0   6
```

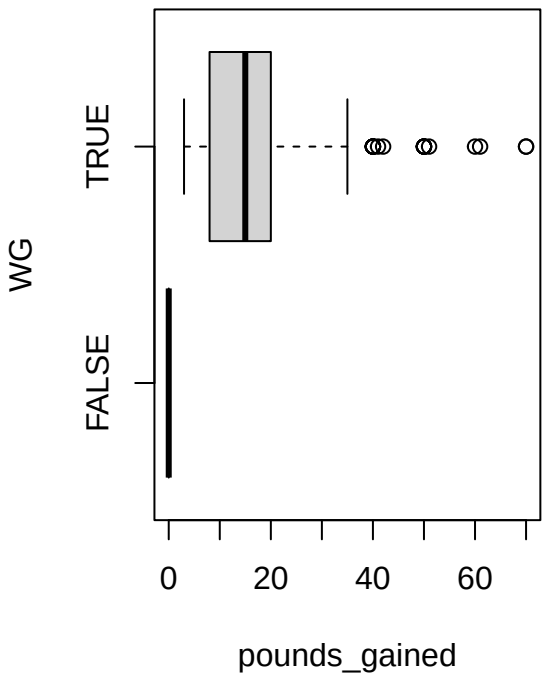
```
gained.dat <- gained.dat[!is.na(gained.dat$CalcTMM),]
dim(gained.dat)
```

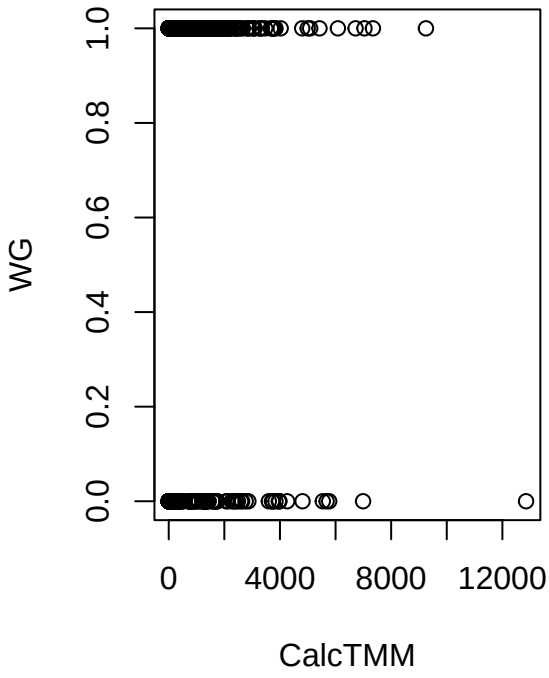
```
## [1] 347  86
```

```
#gained.dat <- gained.dat[!gained.dat$MissingLbs,]
dim(gained.dat)
```

```
## [1] 347  86
```

```
par(mfrow=c(1,2))
boxplot(pounds_gained ~ WG,data=gained.dat,horizontal = TRUE)
plot(WG~CalcTMM,data=gained.dat)
```





Analysis of Binary Response (WG)

(SA1) Does *total metabolic minutes* have an effect on *weight gain*?

Simple logistic regression

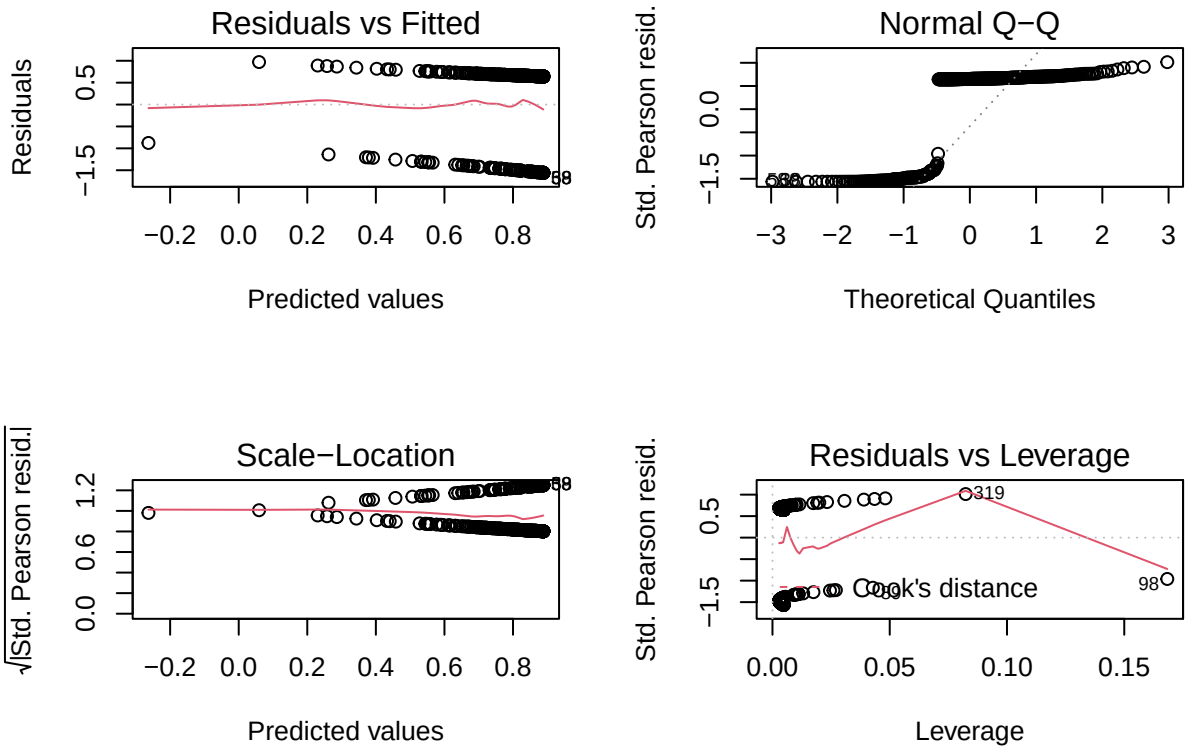
The client provided us two specific aims to address in our analysis. We will begin with the first specific aim, which we will refer to as SA1. SA1 says, "Does total metabolic minutes have an effect on weight gain?". To address this question, we began by creating a simple linear regression model of our binary weight gain versus calculated total metabolic minutes. The summary output of this model suggests that `CalcTMM` has very little effect on weight gain, as the p-value for `CalcTMM` is 0.21, quite large.

```
SA1.model1 <- glm(WG ~ CalcTMM, data=gained.dat,family = binomial)
summary(SA1.model1)
```

```
##
## Call:
## glm(formula = WG ~ CalcTMM, family = binomial, data = gained.dat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.5702  -1.4777   0.8403   0.8697   1.1523
##
```

```
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  8.882e-01  1.517e-01   5.855 4.78e-09 ***
## CalcTMM     -8.959e-05  7.141e-05  -1.255     0.21
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 433.47  on 346  degrees of freedom
## Residual deviance: 431.91  on 345  degrees of freedom
## AIC: 435.91
##
## Number of Fisher Scoring iterations: 4
```

```
par(mfrow=c(2,2))
plot(SA1.model1)
```



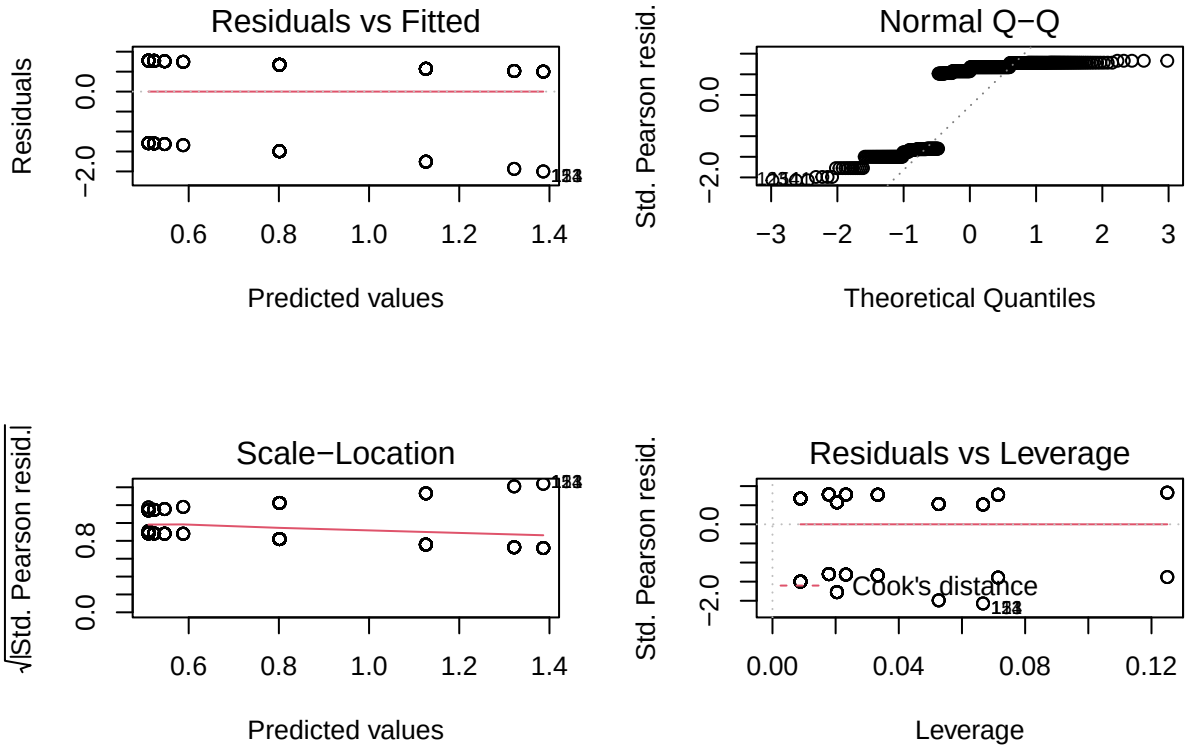
(SA2) Does *shift* have an effect on *weight gain*?

The second specific aim given to us by the client says, "Does shift have an effect on weight gain?". To address this question, we began by creating a simple linear regression model of our binary weight gain versus shift. From the summary output of this model, we see that no value of shift seems to have much significance in the model. The p-value for each of these values is greater than 0.1, suggesting each has little effect on weight gain.

```r
SA2.model1 <- glm(WG ~ shift, data=gained.dat,family = binomial)
summary(SA2.model1)
```

```
##
## Call:
## glm(formula = WG ~ shift, family = binomial, data = gained.dat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.7941  -1.4006   0.7495   0.9400   0.9695
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.54654    0.37887   1.443    0.149
## shift8am     0.25482    0.43004   0.593    0.553
## shift9am    -0.03572    0.46875  -0.076    0.939
## shift10am    0.57947    0.50389   1.150    0.250
## shift11am   -0.02330    0.49303  -0.047    0.962
## shift12pm    0.04124    0.67428   0.061    0.951
## shift1pm    -0.03572    0.82272  -0.043    0.965
## shift2pm     0.83975    0.74847   1.122    0.262
## shiftother   0.77521    0.67839   1.143    0.253
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 433.47  on 346  degrees of freedom
## Residual deviance: 428.11  on 338  degrees of freedom
## AIC: 446.11
##
## Number of Fisher Scoring iterations: 4
```

```r
par(mfrow=c(2,2))
plot(SA2.model1)
```



Model 2 Interactions

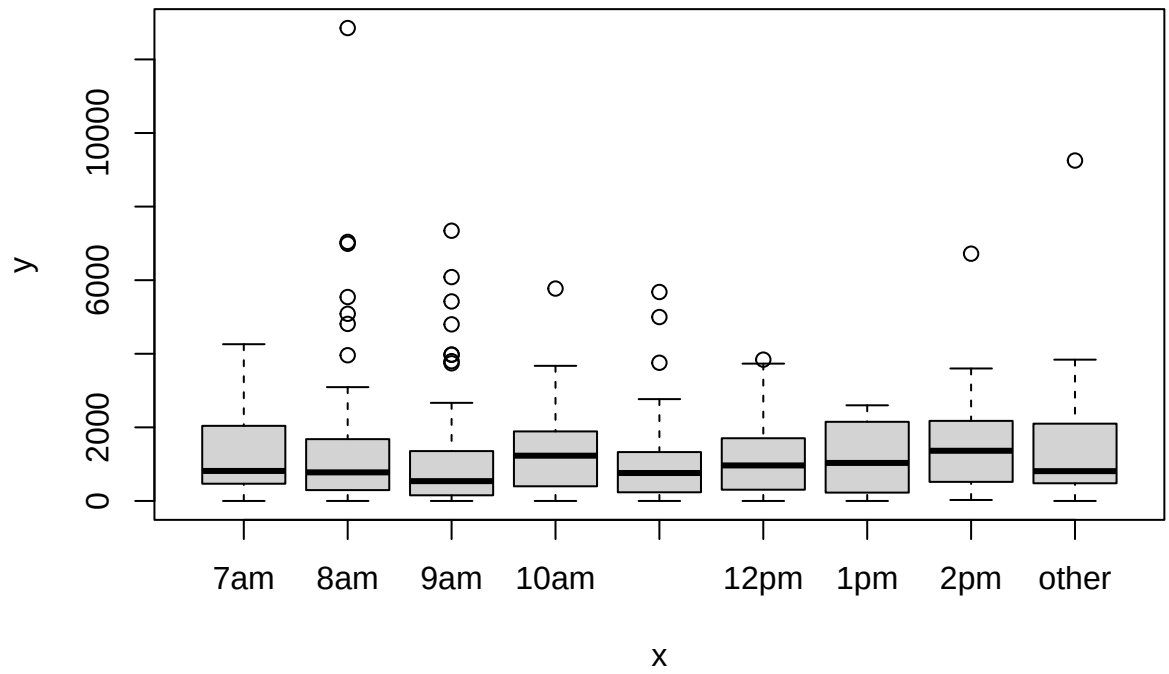
After observing the models representing weight gain versus total met minutes and shift, respectively, we sought to examine whether `shift*CalcTMM` had an effect on weight gain.

```
SA12.model2 <- glm(WG ~ shift*CalcTMM, data=gained.dat,family = binomial)
summary(SA12.model2)
```

```
##
## Call:
## glm(formula = WG ~ shift * CalcTMM, family = binomial, data = gained.dat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9466  -1.3601   0.7570   0.9131   1.4612
##
## Coefficients:
##                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)         1.060e+00  5.867e-01   1.806   0.0709 .
## shift8am            8.686e-02  6.470e-01   0.134   0.8932
## shift9am           -6.779e-01  6.792e-01  -0.998   0.3182
## shift10am           6.941e-01  8.161e-01   0.851   0.3950
## shift11am          -6.229e-01  7.194e-01  -0.866   0.3866
## shift12pm          -2.265e-01  1.000e+00  -0.226   0.8209
```

```
## shift1pm            7.634e-02  1.377e+00   0.055   0.9558
## shift2pm            2.430e-01  1.101e+00   0.221   0.8253
## shiftother         -1.002e+00  1.066e+00  -0.940   0.3473
## CalcTMM            -3.565e-04  2.972e-04  -1.200   0.2303
## shift8am:CalcTMM    1.019e-04  3.249e-04   0.314   0.7538
## shift9am:CalcTMM    4.652e-04  3.455e-04   1.346   0.1782
## shift10am:CalcTMM  -6.172e-05  4.110e-04  -0.150   0.8806
## shift11am:CalcTMM   4.393e-04  3.977e-04   1.105   0.2693
## shift12pm:CalcTMM   1.608e-04  5.436e-04   0.296   0.7674
## shift1pm:CalcTMM   -1.489e-04  8.233e-04  -0.181   0.8565
## shift2pm:CalcTMM    4.062e-04  5.046e-04   0.805   0.4209
## shiftother:CalcTMM  1.696e-03  1.084e-03   1.564   0.1178
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 433.47  on 346  degrees of freedom
## Residual deviance: 415.04  on 329  degrees of freedom
## AIC: 451.04
##
## Number of Fisher Scoring iterations: 6
```

```
plot(metrics.dat$shift, metrics.dat$CalcTMM)
```



Model 3 SA1 and 2 plus anthropometric variables

The original data set included anthropometric variables such as `gender`, `Age`, `height`, and `BMI`. We wanted to inspect whether such variables affect weight gain, and so we created a generalized linear model including them. We began by including just variables for gender and age, and then height. It can be seen that the p-values for these variables in both models are large, suggesting little effect on weight gain.

```
subset3a.dat <- gained.dat[which(complete.cases(gained.dat[, c("gender", "Age", "shift", "CalcTMM")])),]
SA12.model3a <- glm(WG ~ gender + Age + shift + CalcTMM, data=subset3a.dat,family = binomial)
summary(SA12.model3a)
```

```
##
## Call:
## glm(formula = WG ~ gender + Age + shift + CalcTMM, family = binomial,
##     data = subset3a.dat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.0693  -1.3158   0.7675   0.8650   1.4108
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)   1.003e-01  1.626e+00   0.062    0.951
## genderFemale  1.323e+00  1.476e+00   0.896    0.370
## genderMale    8.046e-01  1.489e+00   0.541    0.589
## Age          -6.123e-03  1.265e-02  -0.484    0.628
## shift8am     -1.468e-01  4.819e-01  -0.305    0.761
## shift9am     -3.807e-01  5.238e-01  -0.727    0.467
## shift10am     1.770e-01  5.523e-01   0.320    0.749
## shift11am    -1.926e-01  5.540e-01  -0.348    0.728
## shift12pm    -4.136e-01  7.190e-01  -0.575    0.565
## shift1pm     -4.508e-01  8.569e-01  -0.526    0.599
## shift2pm      1.337e+00  1.150e+00   1.162    0.245
## shiftother    7.151e-01  8.013e-01   0.892    0.372
## CalcTMM      -1.124e-04  7.651e-05  -1.469    0.142
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 393.55  on 318  degrees of freedom
## Residual deviance: 380.39  on 306  degrees of freedom
## AIC: 406.39
##
## Number of Fisher Scoring iterations: 4
```

```
subset3b.dat <- gained.dat[which(complete.cases(gained.dat[, c("gender", "Age", "height","shift", "Calc
SA12.model3b <- glm(WG ~ gender + Age + height +shift + CalcTMM, data=subset3b.dat,family = binomial)
summary(SA12.model3b)
```

```
##
## Call:
## glm(formula = WG ~ gender + Age + height + shift + CalcTMM, family = binomial,
##     data = subset3b.dat)
```

```
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.0225  -1.2973   0.7586   0.8560   1.3909
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -2.136e-01  3.385e+00  -0.063    0.950
## genderFemale  1.355e+00  1.491e+00   0.909    0.364
## genderMale    7.863e-01  1.485e+00   0.529    0.597
## Age          -2.485e-03  1.296e-02  -0.192    0.848
## height        2.457e-03  4.172e-02   0.059    0.953
## shift8am     -1.088e-01  4.832e-01  -0.225    0.822
## shift9am     -3.487e-01  5.275e-01  -0.661    0.509
## shift10am     2.378e-01  5.610e-01   0.424    0.672
## shift11am    -1.791e-01  5.546e-01  -0.323    0.747
## shift12pm    -4.104e-01  7.196e-01  -0.570    0.568
## shift1pm     -4.375e-01  8.575e-01  -0.510    0.610
## shift2pm      1.346e+00  1.153e+00   1.167    0.243
## shiftother    6.555e-01  8.067e-01   0.813    0.416
## CalcTMM      -1.077e-04  7.696e-05  -1.399    0.162
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 384.25  on 312  degrees of freedom
## Residual deviance: 371.25  on 299  degrees of freedom
## AIC: 399.25
##
## Number of Fisher Scoring iterations: 4
```

Model 4 Partition CalcTMM into components

In the background information given to us by our client, it stated that the `Total_Met_Min` column was calculated by a combination of `Vig.ex.Time`, `Mod.ex.Time`, and `Walk.ex.Time`. Although as previously discovered, total met minutes as a whole does not have a large effect on weight gain, we wanted to determine whether its individual components have an effect. We see that none have significant effects ($p > .05$).

```
subset4a.dat <- gained.dat[which(complete.cases(gained.dat[, c("gender", "Age", "shift", "Vig.ex.Time",
SA12.model4a <- glm(WG ~ gender + Age + shift + Vig.ex.Time  + Mod.ex.time + Walk.ex.Time, data=subset4a
summary(SA12.model4a)
```

```
##
## Call:
## glm(formula = WG ~ gender + Age + shift + Vig.ex.Time + Mod.ex.time +
##     Walk.ex.Time, family = binomial, data = subset4a.dat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.0683  -1.3096   0.7628   0.8637   1.4283
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
```

```
## (Intercept)   2.830e-01  1.662e+00   0.170    0.865
## genderFemale  1.133e+00  1.518e+00   0.746    0.456
## genderMale    6.236e-01  1.527e+00   0.408    0.683
## Age          -6.496e-03  1.271e-02  -0.511    0.609
## shift8am     -1.226e-01  4.844e-01  -0.253    0.800
## shift9am     -3.866e-01  5.257e-01  -0.735    0.462
## shift10am     2.055e-01  5.556e-01   0.370    0.712
## shift11am    -1.591e-01  5.581e-01  -0.285    0.776
## shift12pm    -4.121e-01  7.195e-01  -0.573    0.567
## shift1pm     -4.472e-01  8.571e-01  -0.522    0.602
## shift2pm      1.340e+00  1.159e+00   1.156    0.248
## shiftother    7.275e-01  8.023e-01   0.907    0.365
## Vig.ex.Time  -1.443e-03  1.231e-03  -1.172    0.241
## Mod.ex.time  -6.436e-05  1.166e-03  -0.055    0.956
## Walk.ex.Time -2.553e-04  5.873e-04  -0.435    0.664
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 393.55  on 318  degrees of freedom
## Residual deviance: 380.13  on 304  degrees of freedom
## AIC: 410.13
##
## Number of Fisher Scoring iterations: 4
```

```
subset4b.dat <- gained.dat[which(complete.cases(gained.dat[, c("gender", "Age", "height", "shift", "Vig
SA12.model4b <- glm(WG ~ gender + Age + height + shift + Vig.ex.Time  + Mod.ex.time + Walk.ex.Time, data
summary(SA12.model4b)
```

```
##
## Call:
## glm(formula = WG ~ gender + Age + height + shift + Vig.ex.Time +
##     Mod.ex.time + Walk.ex.Time, family = binomial, data = subset4b.dat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.0204  -1.2916   0.7557   0.8589   1.3908
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -0.0537081  3.4176533  -0.016    0.987
## genderFemale  1.1810847  1.5395473   0.767    0.443
## genderMale    0.6147096  1.5279091   0.402    0.687
## Age          -0.0026950  0.0130268  -0.207    0.836
## height        0.0024623  0.0417659   0.059    0.953
## shift8am     -0.0816507  0.4857681  -0.168    0.867
## shift9am     -0.3446730  0.5295749  -0.651    0.515
## shift10am     0.2666934  0.5637140   0.473    0.636
## shift11am    -0.1444094  0.5590046  -0.258    0.796
## shift12pm    -0.4134012  0.7198845  -0.574    0.566
## shift1pm     -0.4307054  0.8575452  -0.502    0.615
## shift2pm      1.3264224  1.1620886   1.141    0.254
## shiftother    0.6683115  0.8076087   0.828    0.408
## Vig.ex.Time  -0.0013616  0.0012588  -1.082    0.279
## Mod.ex.time   0.0000909  0.0011976   0.076    0.939
```

```
## Walk.ex.Time -0.0003440  0.0005891  -0.584    0.559
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 384.25  on 312  degrees of freedom
## Residual deviance: 370.99  on 297  degrees of freedom
## AIC: 402.99
##
## Number of Fisher Scoring iterations: 4
```

Model 5 - Model 4 plus BMI and initial body weight

For these models, we may include BMI, or just the anthropometric variables used to calculate BMI. Using the columns for body weight, height, and pounds gained, we were able to obtain `initial_BMI` and `initial_bweight` columns.

```
gained.dat['initial_bweight'] <- gained.dat$bweight - gained.dat$pounds_gained
gained.dat['initial_BMI'] <- (gained.dat$initial_bweight / (gained.dat$height)^2)*703
```

In the following models, we used the anthropometric variables, the individual total met minute components in addition to the initial BMI and initial body weight variables, respectively. We know that BMI is calculated using height and weight, so we avoided using all three (`initial_BMI`, `height`, `initial_bweight`) in a model to avoid confounding. We see that in the first model, `initial_BMI` has a small p-value and in the second model, `initial_bweight` also has a small p-value. This suggests that the variables effect weight gain in their respective models.

```
subset5a.dat <- gained.dat[which(complete.cases(gained.dat[, c("gender", "Age", "shift", "Vig.ex.Time",
SA12.model5a <- glm(WG ~ gender + Age + shift + Vig.ex.Time  + Mod.ex.time + Walk.ex.Time + initial_BMI
summary(SA12.model5a)
```

```
##
## Call:
## glm(formula = WG ~ gender + Age + shift + Vig.ex.Time + Mod.ex.time +
##     Walk.ex.Time + initial_BMI, family = binomial, data = subset5a.dat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.1210  -1.2011   0.7118   0.8860   1.2865
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)   3.748e-01  1.804e+00   0.208   0.8355
## genderFemale  1.606e+00  1.548e+00   1.038   0.2995
## genderMale    1.126e+00  1.561e+00   0.721   0.4707
## Age           1.232e-02  1.540e-02   0.800   0.4237
## shift8am     -2.743e-01  5.420e-01  -0.506   0.6128
## shift9am      1.975e-01  6.326e-01   0.312   0.7548
## shift10am     5.448e-01  6.418e-01   0.849   0.3960
## shift11am    -2.802e-01  6.240e-01  -0.449   0.6535
## shift12pm     3.528e-01  8.418e-01   0.419   0.6751
## shift1pm      6.855e-02  1.022e+00   0.067   0.9465
```

```
## shift2pm      1.593e+01  9.552e+02   0.017   0.9867
## shiftother    7.029e-01  8.472e-01   0.830   0.4067
## Vig.ex.Time  -4.839e-04  1.427e-03  -0.339   0.7344
## Mod.ex.time  -6.368e-04  1.477e-03  -0.431   0.6663
## Walk.ex.Time -3.513e-04  6.585e-04  -0.533   0.5938
## initial_BMI  -5.113e-02  2.660e-02  -1.922   0.0546 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 301.19  on 241  degrees of freedom
## Residual deviance: 282.10  on 226  degrees of freedom
## AIC: 314.1
##
## Number of Fisher Scoring iterations: 15
```

```
subset5b.dat <- gained.dat[which(complete.cases(gained.dat[, c("gender", "Age", "height", "shift", "Vig
SA12.model5b <- glm(WG ~ gender + Age + height + shift + Vig.ex.Time  + Mod.ex.time + Walk.ex.Time +ini
summary(SA12.model5b)
```

```
##
## Call:
## glm(formula = WG ~ gender + Age + height + shift + Vig.ex.Time +
##     Mod.ex.time + Walk.ex.Time + initial_bweight, family = binomial,
##     data = subset5b.dat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.0781  -1.2123   0.7194   0.8883   1.2876
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)     -2.620e+00  3.892e+00  -0.673   0.5008
## genderFemale     1.647e+00  1.573e+00   1.047   0.2953
## genderMale       1.132e+00  1.561e+00   0.725   0.4683
## Age              1.180e-02  1.542e-02   0.765   0.4441
## height           4.312e-02  5.131e-02   0.840   0.4007
## shift8am        -2.751e-01  5.424e-01  -0.507   0.6120
## shift9am         1.908e-01  6.324e-01   0.302   0.7629
## shift10am        5.317e-01  6.431e-01   0.827   0.4084
## shift11am       -2.783e-01  6.256e-01  -0.445   0.6564
## shift12pm        3.137e-01  8.402e-01   0.373   0.7089
## shift1pm         5.504e-02  1.017e+00   0.054   0.9568
## shift2pm         1.596e+01  9.524e+02   0.017   0.9866
## shiftother       7.036e-01  8.478e-01   0.830   0.4065
## Vig.ex.Time     -4.411e-04  1.425e-03  -0.310   0.7569
## Mod.ex.time     -6.720e-04  1.475e-03  -0.456   0.6487
## Walk.ex.Time    -3.420e-04  6.609e-04  -0.517   0.6049
## initial_bweight -7.408e-03  4.208e-03  -1.760   0.0784 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
```

```
##
##     Null deviance: 301.19  on 241  degrees of freedom
## Residual deviance: 282.60  on 225  degrees of freedom
## AIC: 316.6
##
## Number of Fisher Scoring iterations: 15
```

Lastly, we utilized the function 'stepAIC' to find the simplest model. We called this function twice–once with the model consisting of the anthropometric variables, the total met minutes components, shift, and initial body weight, and once with the model consisting of the same variables but instead of initial weight, we have initial BMI. The stepAIC function suggest that the simplest model that includes initial BMI uses variables `gender` and `initial_BMI` as predictors. The simplest model that includes the initial body weight uses just `initial_bweight` as a predictor.

Thus, we cannot state, to statistical significance, that either specific aim 1 or aim 2 are true. We can, however, suggest that initial BMI and gender are better predictors of weight gain than either Total MET Minutes or shift. See the summary of the recommended model below.

```
summary(best.modela)
```

```
##
## Call:
## glm(formula = WG ~ gender + initial_BMI, family = binomial, data = subset5a.dat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.7865  -1.3332   0.7443   0.8789   1.1867
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   0.88171    1.53086   0.576    0.565
## genderFemale  1.07788    1.42643   0.756    0.450
## genderMale    0.47359    1.43401   0.330    0.741
## initial_BMI  -0.03621    0.02406  -1.505    0.132
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 301.19  on 241  degrees of freedom
## Residual deviance: 293.63  on 238  degrees of freedom
## AIC: 301.63
##
## Number of Fisher Scoring iterations: 4
```

```
summary(best.modelb)
```

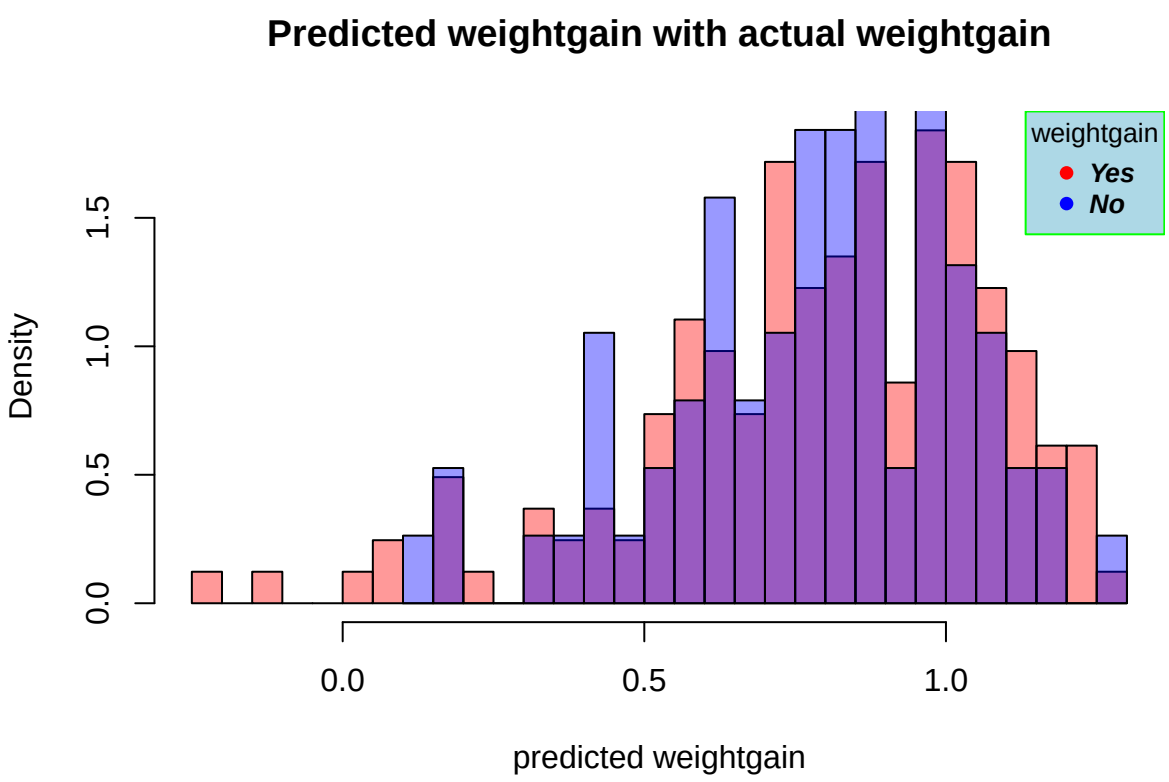
```
##
## Call:
## glm(formula = WG ~ initial_bweight, family = binomial, data = subset5b.dat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.7275  -1.4288   0.7904   0.8762   1.2352
##
```

```
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)      1.874684   0.573019   3.272  0.00107 **
## initial_bweight -0.006504   0.003272  -1.988  0.04684 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 301.19  on 241  degrees of freedom
## Residual deviance: 297.23  on 240  degrees of freedom
## AIC: 301.23
##
## Number of Fisher Scoring iterations: 4
```

```
AIC(best.modela,best.modelb)
```

```
##             df      AIC
## best.modela  4 301.6251
## best.modelb  2 301.2257
```

```
par(mfrow=(c(1,1)))
p = predict.glm(best.modelb)
Histogram_1 <- hist((subset(p, metrics.dat$weightgain == "Yes")), plot = FALSE,breaks=40)
Histogram_2 <- hist((subset(p, metrics.dat$weightgain == "No")), plot = FALSE, breaks =30)
{plot (Histogram_1, col = rgb(1,0,0,0.4),xlab = 'predicted weightgain',freq = FALSE, main = 'Predicted
 plot (Histogram_2, xaxt = 'n', yaxt = 'n',col = rgb(0,0,1,0.4), add = TRUE, freq = FALSE)
legend("topright",c("Yes","No"),cex=.8,col=c("red","blue"),pch=c(19,19),box.col="green", title="weightg
```



We find the best, simplest model for `weightgained` includes only the single predictor, `initial_bodyweight`.

Model 6

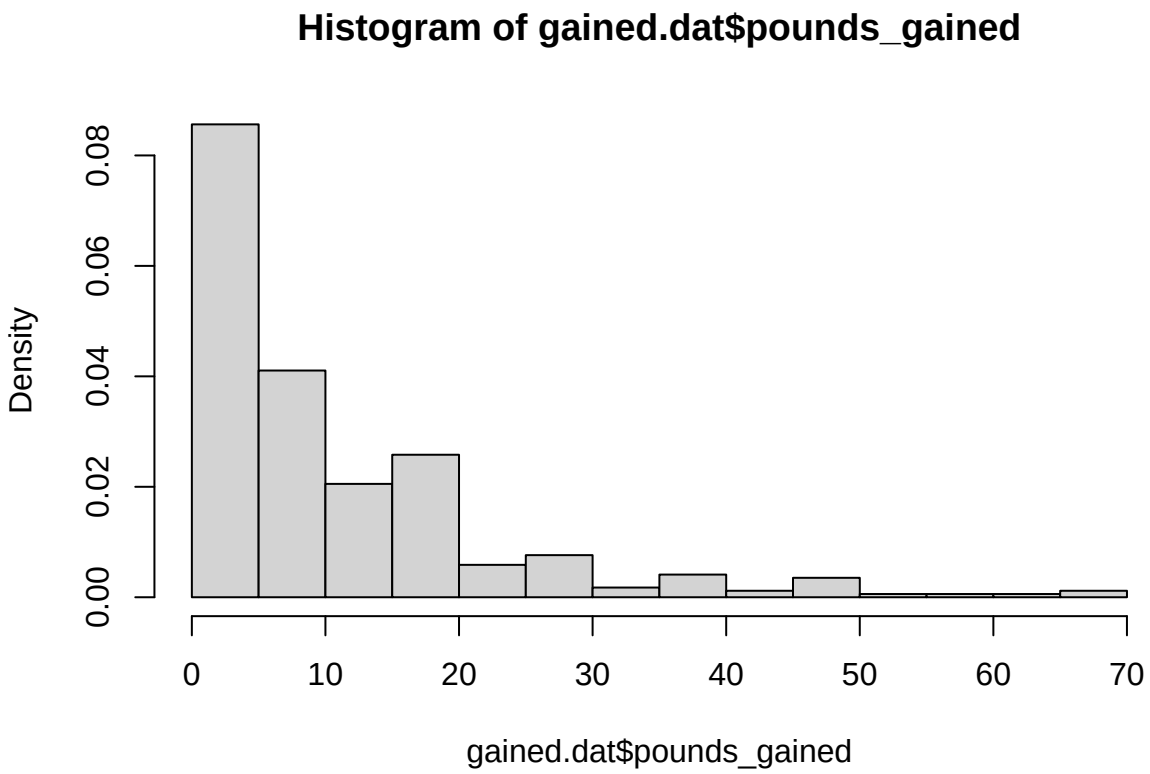
Lastly, we wanted to observe the binomial models created from `initial_BMIxshiftxCalcTMM` as well as `initial_bweightxshiftxCalcTMM`. We then ran the stepAIC() function through both models, and obtained a lower AIC value for the model using `initial_BMIxshiftxCalcTMM`. This suggests this model is simpler than the other. In this model, the following predictors had p-values <0.1: shift10am:CalcTMM, initial_BMI:shift10am:CalcTMM. This suggests that these products are significant in predicting whether an employee has gained weight. See the appendix at the end for model summaries.

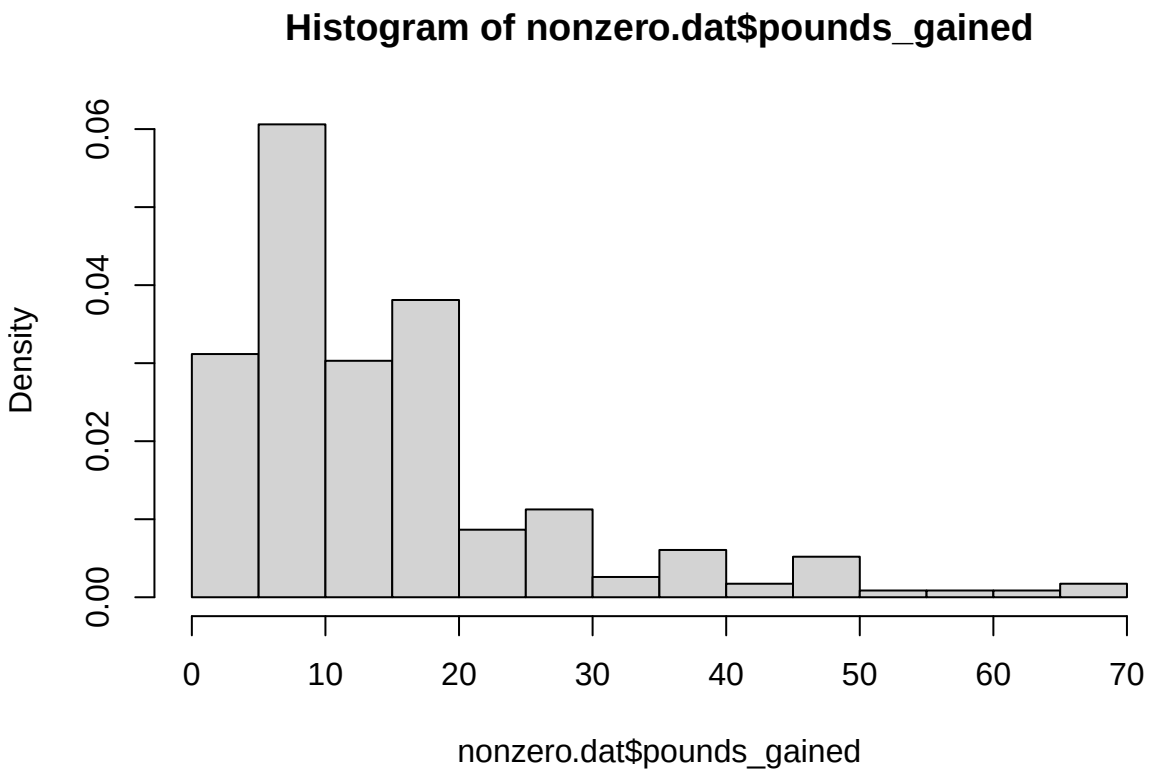
Analysis of Continuous Response (pounds gained)

(SA1) Does *total metabolic minutes* have an effect on *weight gain*?

We have information about net pounds gained. We assume that when `weightgained` is false, we can substitute a value of 0 for `pounds_gained`. This allows us to analyze a full data set; otherwise, we limit our observations. It is worth noting that we may be oversimplifying cases were `pounds_gained` may be negative, thus creating a censored or zero-inflated data set. Therefore, we consider all possible data, and the subset where pounds gained is non-zero.

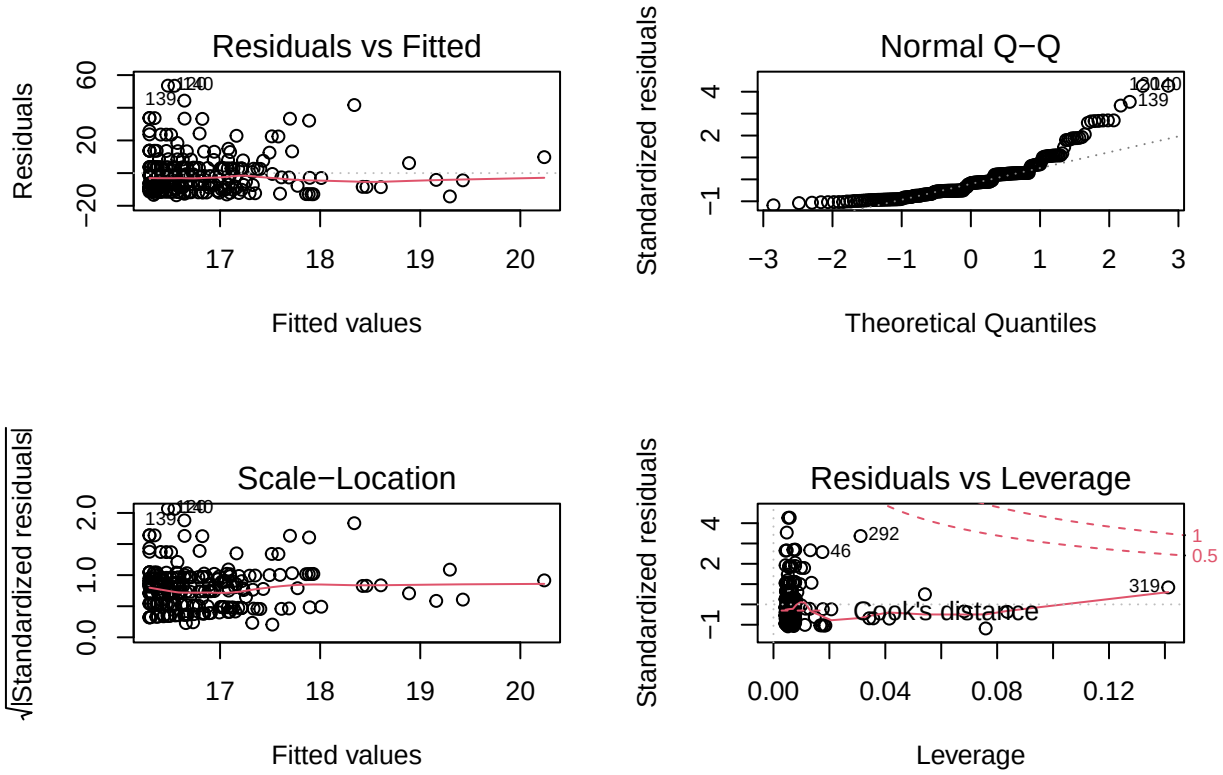
```
## [1] 6
```


Histogram of gained.dat$pounds_gained
Density
0.00
0.02
0.04
0.06
0.08
0
10
20
30
40
50
60
70
gained.dat$pounds_gained



Pounds gained is highly skewed, even when zero observations are excluded. Thus, a linear model, with the assumption of normally distributed errors, may not be appropriate. We include a linear model here for reference only; this is not the recommended analysis. ### Linear model, gaussian errors

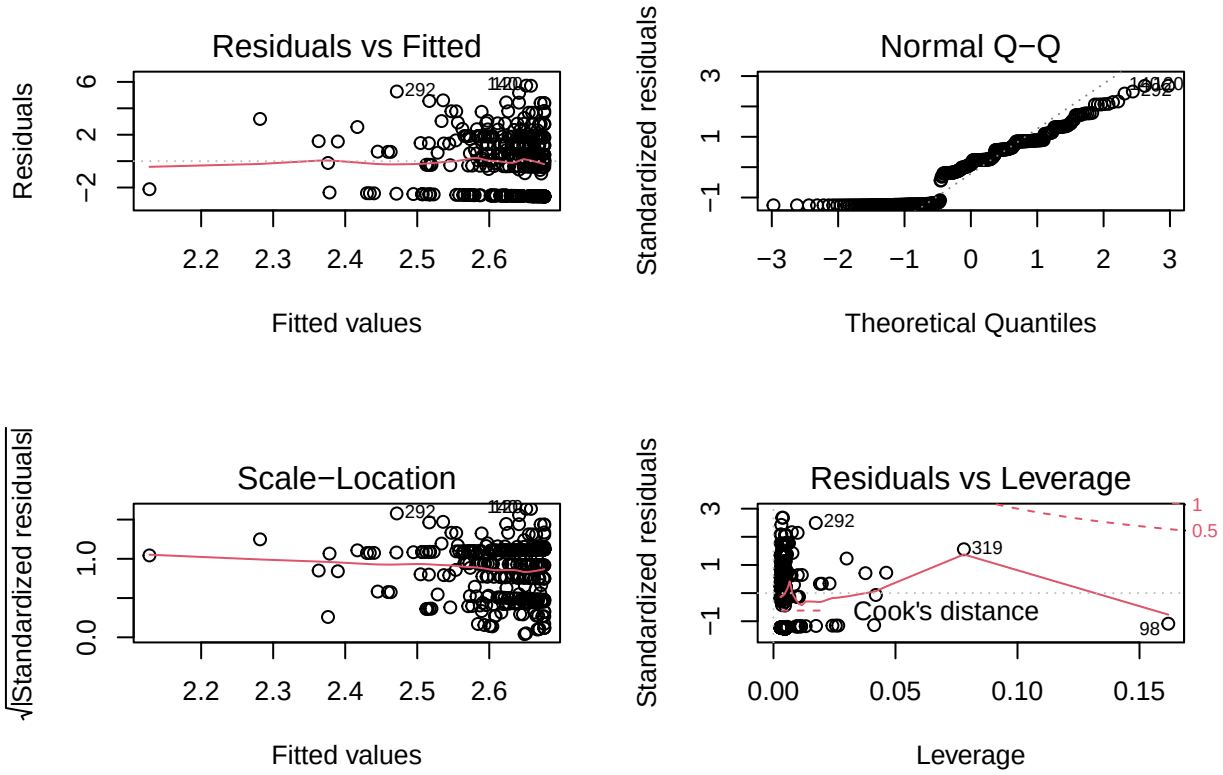
```
SA1.model1.lm <- lm(pounds_gained ~ CalcTMM, data=nonzero.dat)
#summary(SA1.model1.lm)
par(mfrow=c(2,2))
plot(SA1.model1.lm)
```



Square-Root Transform

The distribution of weight gain is highly left-skewed. We may correct this by applying a square-root transformation. We apply this to both the full data (with extra 0s) and the data limited to nonzero weight gain.

```
#error in log(0)
SA1.model1.root <- lm(sqrt(pounds_gained) ~ CalcTMM, data=gained.dat)
#summary(SA1.model1.root)
par(mfrow=c(2,2))
plot(SA1.model1.root)
```

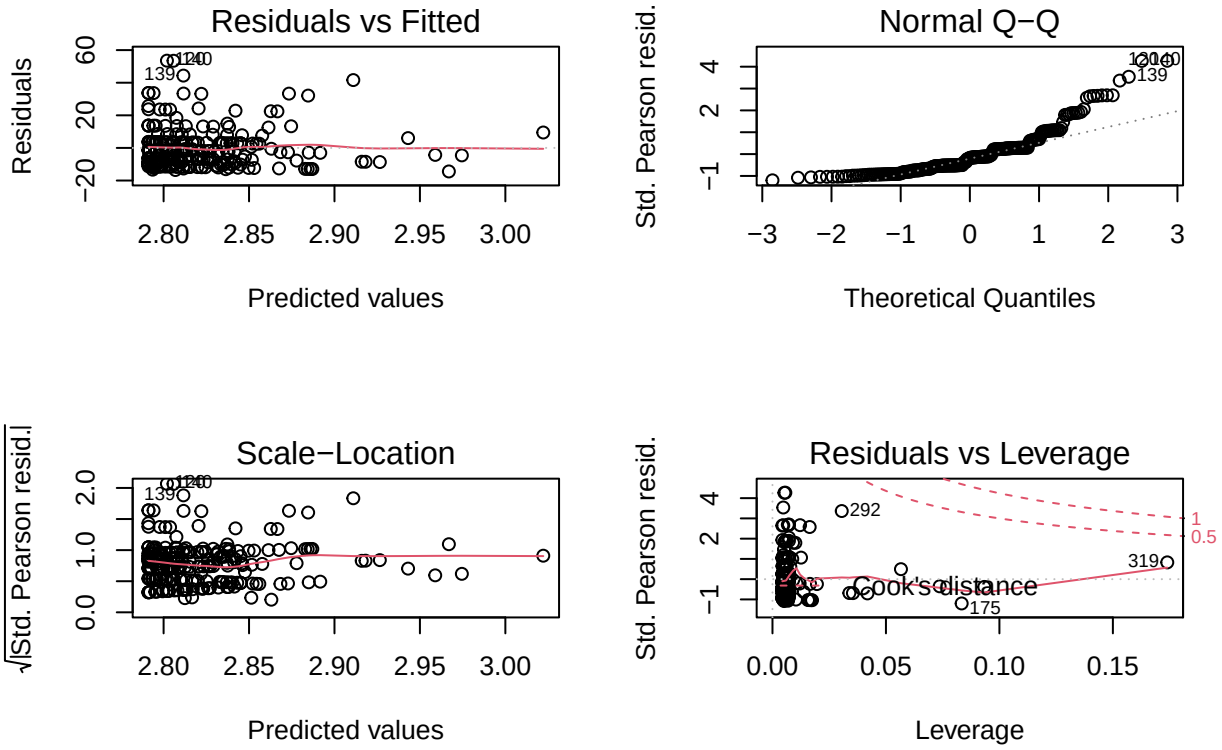


The square-root transformation does appear to improve upon the skewness of the data, although the residuals are not clearly normally distributed.

Log Transformation

The log transformation is also commonly used to correct skewed data. The log-transform, however, is not defined for 0 values. We include an model using the log transform for the non-zero pounds gained data. This is included for reference; we do not recommend analyzing these data using a log transformation.

```
#error in log(0)
SA1b.model1.loggauss <- glm(pounds_gained ~ CalcTMM, data=nonzero.dat,family = gaussian(link="log"))
par(mfrow=c(2,2))
plot(SA1b.model1.loggauss)
```

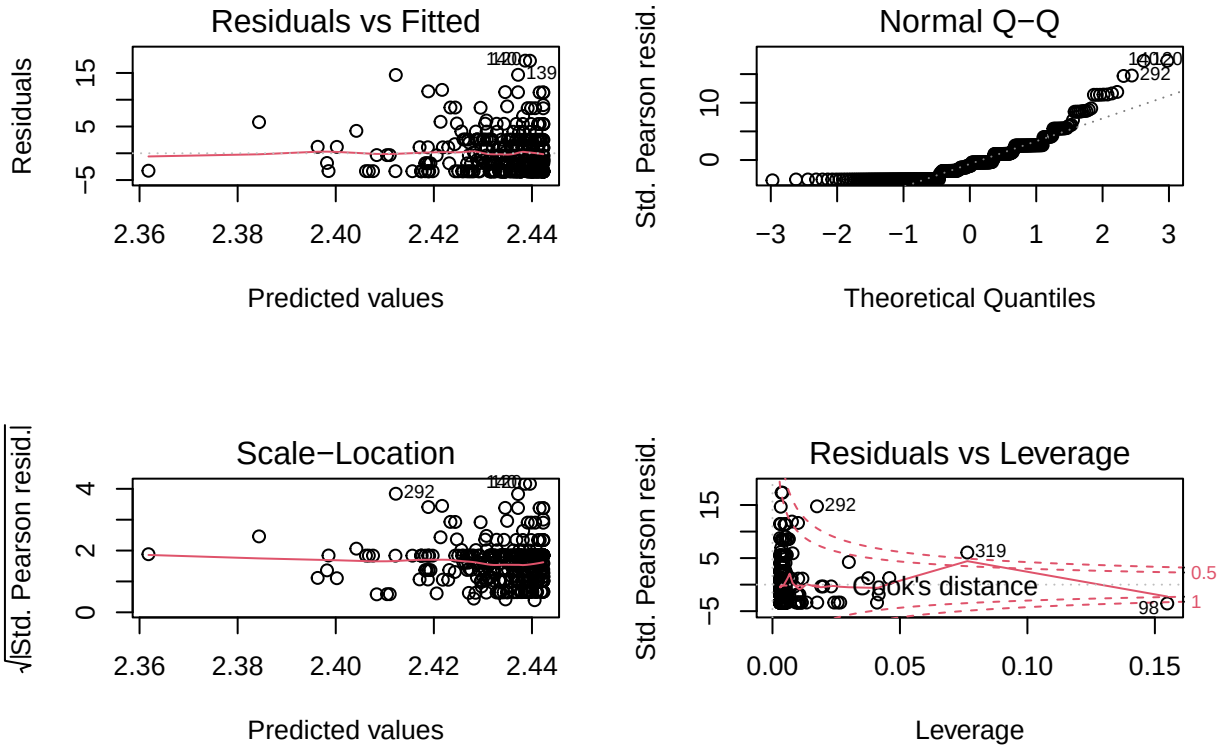


Poisson Regression

The square root transformation is commonly applied to count data. This suggests a possible Poisson model for pounds gained. Poisson regression requires integer values. We'll round pounds gained for this.

```
gained.dat$LBS <- round(gained.dat$pounds_gained)
nonzero.dat$LBS <- round(nonzero.dat$pounds_gained)
```

```
SA1.model1.poisson <- glm(LBS ~ CalcTMM, data=gained.dat,family = poisson)
#SA1b.model1.poisson <- glm(LBS ~ CalcTMM, data=nonzero.dat,family = poisson)
#summary(SA1.model1.poisson)
par(mfrow=c(2,2))
plot(SA1.model1.poisson)
```

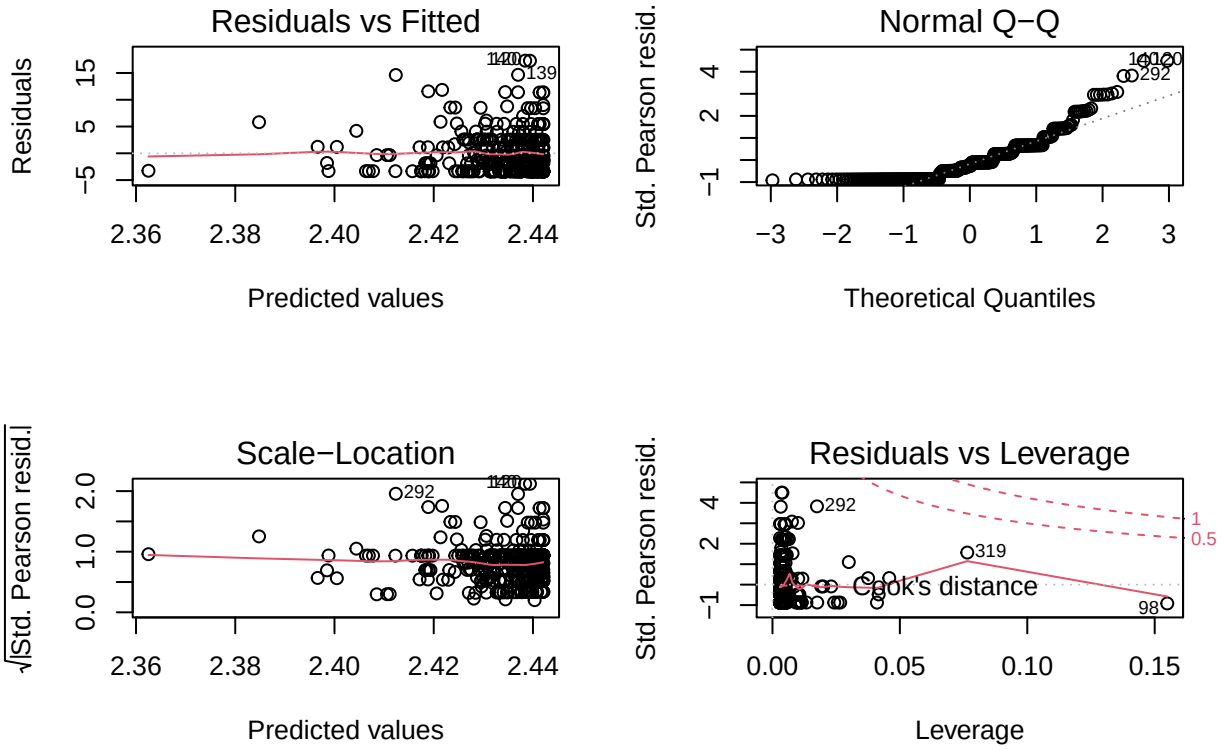


This model provides a similar improvement on residual errors as did the square root transform. Thus, a Poisson model may be recommended for these data.

Quasi-poisson

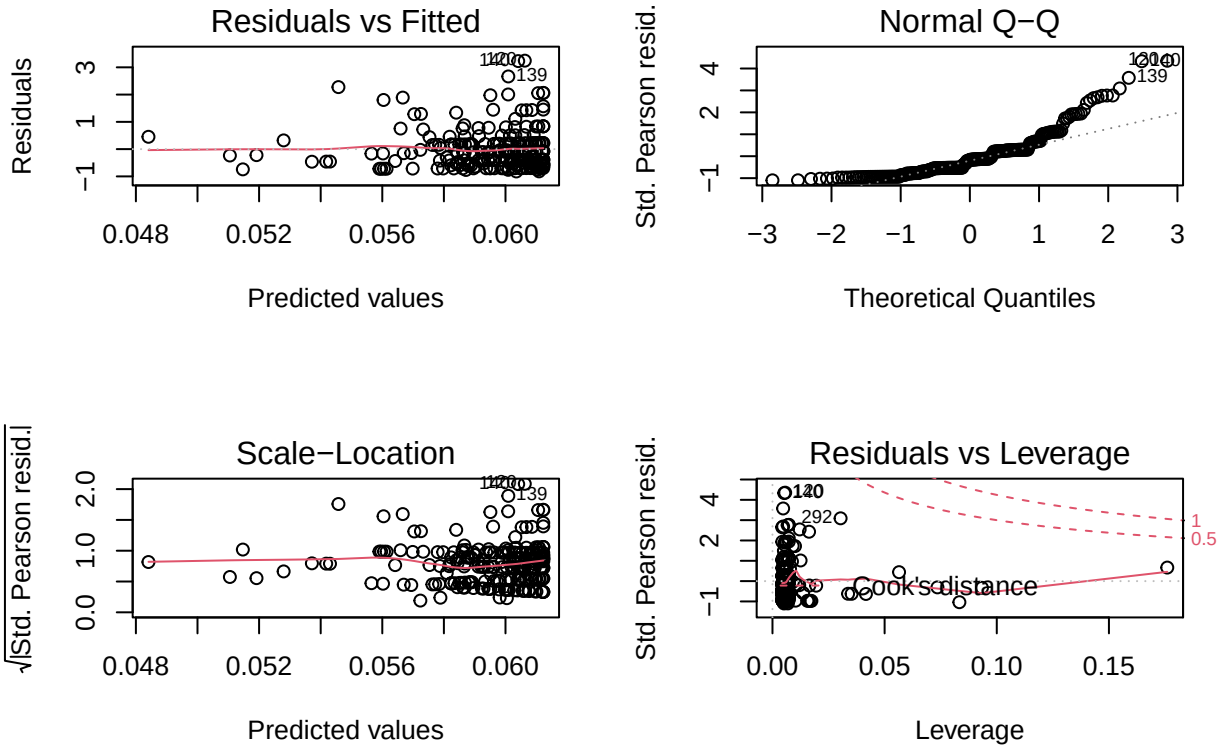
We can, alternatively, fit a quasi-poisson family. This does not require integer values. This also provides a dispersion parameter that may help account for excess 0s.

```
SA1.model1.quasi <- glm(pounds_gained ~ CalcTMM, data=gained.dat,family = quasipoisson)
#SA1b.model1.quasi <- glm(pounds_gained ~ CalcTMM, data=nonzero.dat,family = quasipoisson)
#summary(SA1.model1.quasi)
par(mfrow=c(2,2))
plot(SA1.model1.quasi)
```



The shape of the distribution of weight gain suggests a gamma distribution. However, the gamma distribution is not defined for 0 values. We include a gamma family model for reference, but we do not recommend this model.

```
SA1b.model1.Gamma <- glm(pounds_gained ~ CalcTMM, data=nonzero.dat,family = Gamma)
#summary(SA1b.model1.Gamma)
par(mfrow=c(2,2))
plot(SA1b.model1.Gamma)
```



Zero-inflated Poisson

Of the statistical models considered to this point, the Poisson distribution family seems most suitable for these data. We now consider a zero-inflated Poisson (ZIP) analysis.

Briefly, ZIP defines a conditional probability model. The first stage is modeled as binomial - weight gain is either false (0 pounds gained) or true (a non-zero pounds gained value), with a defined probability. Then, conditional on weight gain being true, the remaining values are fit to a Poisson distribution. This is computed in R using the `pscl` library:

```
library(pscl)
```

```
## Warning: package 'pscl' was built under R version 4.1.1
```

```
## Classes and Methods for R developed in the
## Political Science Computational Laboratory
## Department of Political Science
## Stanford University
## Simon Jackman
## hurdle and zeroinfl functions by Achim Zeileis
```

```
SA1.model1.zero <- zeroinfl(LBS ~ CalcTMM, data = gained.dat)
summary(SA1.model1.zero)
```

```
## Warning in sqrt(diag(object$vcov)): NaNs produced
```

```
##
## Call:
## zeroinfl(formula = LBS ~ CalcTMM, data = gained.dat)
##
## Pearson residuals:
##     Min      1Q  Median      3Q     Max
## -1.3941 -1.2980 -0.4077  0.7914  7.0383
##
## Count model coefficients (poisson with log link):
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) 2.792e+00        NaN     NaN      NaN
## CalcTMM     2.426e-05        NaN     NaN      NaN
##
## Zero-inflation model coefficients (binomial with logit link):
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -8.509e-01  1.851e-01  -4.596 4.32e-06 ***
## CalcTMM      8.009e-05  1.241e-04   0.646    0.519
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of iterations in BFGS optimization: 1
## Log-likelihood: -1631 on 4 Df
```

(SA2) Does *shift* have an effect on *weight gain*?

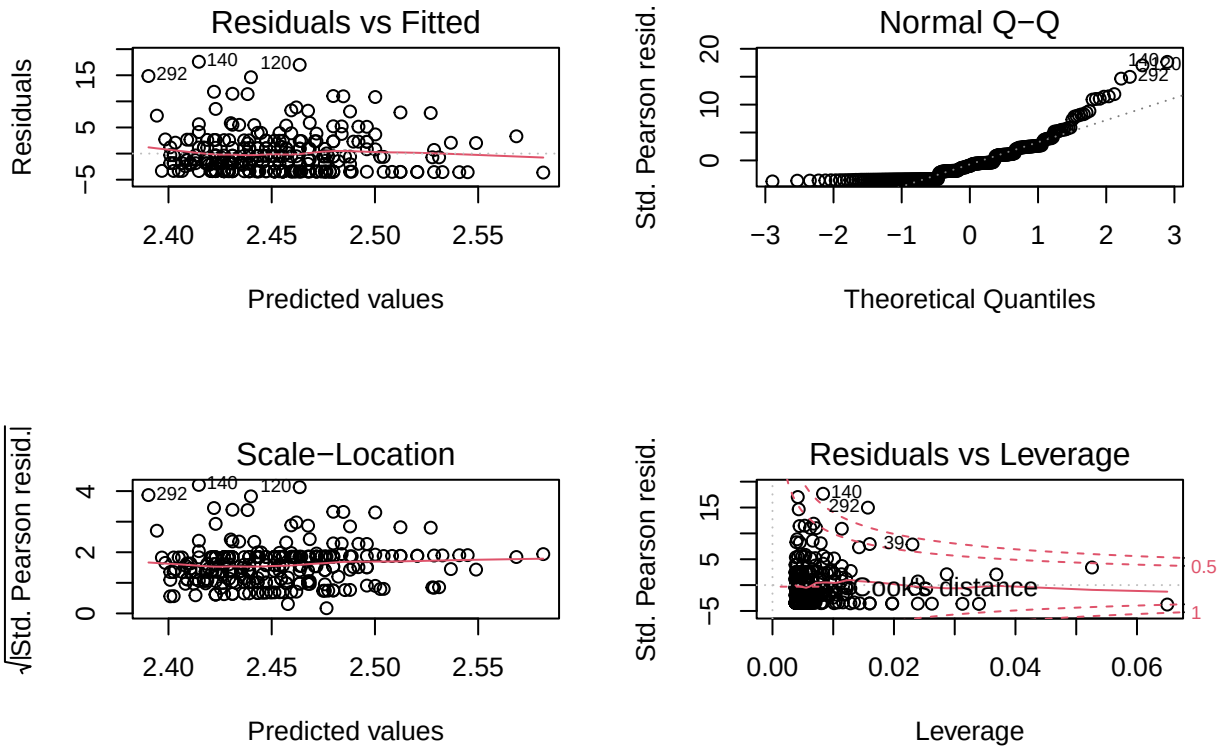
We repeat the analysis of different statistical distributions from above, using *shift* as a predictor variable. See appendix for model summaries.

```
SA2.model1.lm <- lm(pounds_gained ~ shift, data=nonzero.dat)
par(mfrow=c(2,2))
plot(SA2.model1.lm)
SA2.model1.root <- lm(sqrt(pounds_gained) ~ shift, data=gained.dat)
par(mfrow=c(2,2))
plot(SA2.model1.root)
SA2b.model1.loggauss <- glm(pounds_gained ~ shift, data=nonzero.dat,family = gaussian(link="log"))
par(mfrow=c(2,2))
plot(SA2b.model1.loggauss)
SA2.model1.poisson <- glm(LBS ~ shift, data=gained.dat,family = poisson)
par(mfrow=c(2,2))
plot(SA2.model1.poisson)
summary(SA2.model1.poisson)
SA2.model1.quasi <- glm(pounds_gained ~ shift, data=gained.dat,family = quasipoisson)
par(mfrow=c(2,2))
plot(SA2.model1.quasi)
summary(SA2.model1.quasi)
SA2b.model1.Gamma <- glm(pounds_gained ~ shift, data=nonzero.dat,family = Gamma)
par(mfrow=c(2,2))
plot(SA2b.model1.Gamma)
SA2.model1.zero <- zeroinfl(LBS ~ shift, data = gained.dat)
summary(SA2.model1.zero)
```

Best Model (Initial Bodyweight)

We now consider the different statistical families in the analysis of initial body weight and pounds gained, in the context of a zero-inflated poisson model.

```
SA12.best.poisson <- glm(LBS ~ initial_bweight, data=gained.dat,family = poisson)
par(mfrow=c(2,2))
plot(SA12.best.poisson)
```

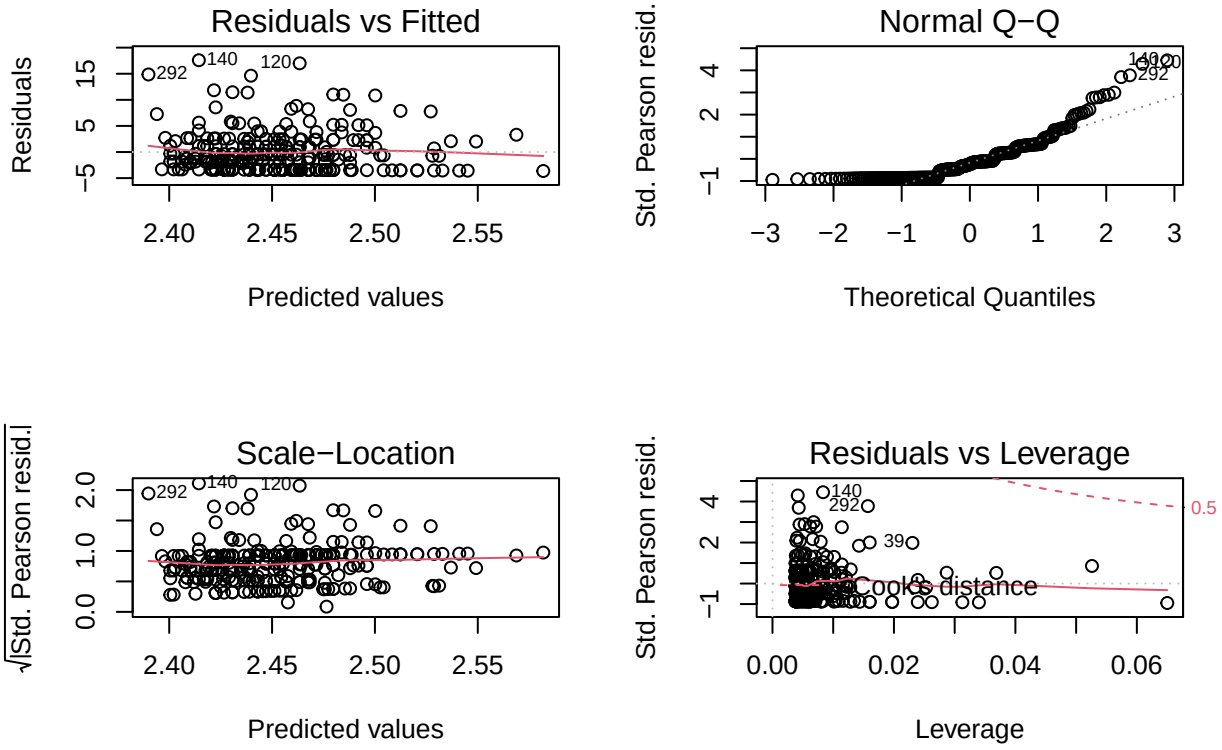



```
summary(SA12.best.poisson)
```

```
##
## Call:
## glm(formula = LBS ~ initial_bweight, family = poisson, data = gained.dat)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -5.141  -4.760  -1.006   1.803  11.794
##
## Coefficients:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)     2.3170820  0.0743207  31.177   <2e-16 ***
## initial_bweight 0.0008137  0.0004303   1.891   0.0586 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 3890.9  on 263  degrees of freedom
## Residual deviance: 3887.3  on 262  degrees of freedom
##   (77 observations deleted due to missingness)
## AIC: 4687.3
##
## Number of Fisher Scoring iterations: 5
```

```
SA12.best.quasi <- glm(pounds_gained ~ initial_bweight, data=gained.dat,family = quasipoisson)
par(mfrow=c(2,2))
plot(SA12.best.quasi)
```




```
summary(SA12.best.quasi)
```

```
##
## Call:
## glm(formula = pounds_gained ~ initial_bweight, family = quasipoisson,
##     data = gained.dat)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -5.142  -4.759  -1.005   1.803  11.796
##
```

```
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     2.3164063  0.2948052   7.857 1.02e-13 ***
## initial_bweight 0.0008167  0.0017068   0.479    0.633
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 15.73252)
##
##     Null deviance: 3891.3  on 263  degrees of freedom
## Residual deviance: 3887.7  on 262  degrees of freedom
##   (77 observations deleted due to missingness)
## AIC: NA
##
## Number of Fisher Scoring iterations: 5
```

```
SA12.best.zero <- zeroinfl(LBS ~ initial_bweight, data = gained.dat)
summary(SA12.best.zero)
```

```
##
## Call:
## zeroinfl(formula = LBS ~ initial_bweight, data = gained.dat)
##
## Pearson residuals:
##     Min      1Q  Median      3Q     Max
## -1.5774 -1.2374 -0.4178  0.6181  7.9549
##
## Count model coefficients (poisson with log link):
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)     2.3636253  0.0739544  31.961  < 2e-16 ***
## initial_bweight 0.0028943  0.0004268   6.781  1.2e-11 ***
##
## Zero-inflation model coefficients (binomial with logit link):
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)     -1.776149   0.552489  -3.215  0.00131 **
## initial_bweight  0.006145   0.003170   1.939  0.05255 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of iterations in BFGS optimization: 1
## Log-likelihood: -1293 on 4 Df
```

Conclusions and Recommendations

- The specific aims stated for this project are partially support with these data. Specifically, we find no significant effect of calculated *Total MET-Minutes* (`CalcTMM`) or *shift* on `weightgain` as a binomial response, using logistic regression models.
- However, we do find a statistical significant effect of `shift` on *pounds gained*. This result is most strongly suggested using a zero-inflated poisson model to account for the individuals reporting 0 pounds gained, and weakly supported using a quasi-poisson model to account for the excess 0 values when no weight gain is reported.

- The logistic regression model suggest that *initial body weight* or *initial BMI* are possible predictors. This was identified from a step-wise model selection algorithm implemented using the `stepAIC` function is R. We considered other combinations of variables, but the best, simplest model included only BMI or initial body weight.

-When interactions among `shift`, `CalcTMM` and `initial_BMI` are included in the model, there is a slight (but not significant at $p<0.05$) suggestion that weight gain may differ among shifts. Thus, initial body weight or initial BMI may be a confounding factor that influences the two variables identified in the specific aims. We note, however, that there were ~100 observations that did not have initial body weight or initial BMI, so this may warrant further investigation, and greater care should be taken when collecting data.

```
boots.zero.model1.tbl$Estimate <- coef(zero.model1)
boots.zero.model1.tbl

confint(zero.model1)
```

Appendix

```
library(MASS)
subset6a.dat <- gained.dat[which(complete.cases(gained.dat[, c("initial_BMI", "shift", "CalcTMM")])),]
subset6a.model <- glm(WG~ initial_BMI*shift*CalcTMM, data=subset6a.dat, family =binomial)
summary(subset6a.model)
```

```
##
## Call:
## glm(formula = WG ~ initial_BMI * shift * CalcTMM, family = binomial,
##     data = subset6a.dat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.2295  -0.9711   0.4149   0.8852   1.7377
##
## Coefficients:
##                                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)                   -2.628e-01  6.010e+00  -0.044   0.9651
## initial_BMI                    2.560e-02  2.107e-01   0.122   0.9033
## shift8am                       2.106e+00  6.172e+00   0.341   0.7329
## shift9am                       1.581e+00  6.427e+00   0.246   0.8057
## shift10am                      3.699e+00  7.138e+00   0.518   0.6044
## shift11am                      2.334e+00  6.698e+00   0.348   0.7275
## shift12pm                     -3.378e+01  3.150e+01  -1.072   0.2836
## shift1pm                      -1.231e+02  9.497e+03  -0.013   0.9897
## shift2pm                       9.971e+00  1.507e+01   0.662   0.5082
## shiftother                    -9.214e-01  1.075e+01  -0.086   0.9317
## CalcTMM                        7.085e-03  6.059e-03   1.169   0.2423
## initial_BMI:shift8am          -6.524e-02  2.174e-01  -0.300   0.7641
## initial_BMI:shift9am          -4.026e-02  2.243e-01  -0.180   0.8575
## initial_BMI:shift10am         -4.925e-02  2.552e-01  -0.193   0.8470
## initial_BMI:shift11am         -1.263e-01  2.412e-01  -0.524   0.6005
## initial_BMI:shift12pm          1.555e+00  1.404e+00   1.108   0.2679
## initial_BMI:shift1pm           5.311e+00  3.915e+02   0.014   0.9892
## initial_BMI:shift2pm          -4.352e-01  6.045e-01  -0.720   0.4716
## initial_BMI:shiftother         5.928e-03  3.751e-01   0.016   0.9874
```

```
## initial_BMI:CalcTMM           -2.448e-04  2.083e-04  -1.175   0.2401
## shift8am:CalcTMM              -7.153e-03  6.101e-03  -1.172   0.2411
## shift9am:CalcTMM              -5.916e-03  6.316e-03  -0.937   0.3489
## shift10am:CalcTMM             -1.260e-02  7.094e-03  -1.776   0.0757 .
## shift11am:CalcTMM             -6.444e-04  7.331e-03  -0.088   0.9300
## shift12pm:CalcTMM              3.956e-02  3.951e-02   1.001   0.3167
## shift1pm:CalcTMM               6.347e-02  4.363e+00   0.015   0.9884
## shift2pm:CalcTMM              -1.103e-02  9.804e-03  -1.125   0.2604
## shiftother:CalcTMM             3.146e-04  1.840e-02   0.017   0.9864
## initial_BMI:shift8am:CalcTMM   2.398e-04  2.103e-04   1.140   0.2542
## initial_BMI:shift9am:CalcTMM   1.948e-04  2.196e-04   0.887   0.3750
## initial_BMI:shift10am:CalcTMM  4.277e-04  2.531e-04   1.690   0.0910 .
## initial_BMI:shift11am:CalcTMM  5.416e-05  2.462e-04   0.220   0.8259
## initial_BMI:shift12pm:CalcTMM -1.679e-03  1.610e-03  -1.043   0.2969
## initial_BMI:shift1pm:CalcTMM  -2.472e-03  1.662e-01  -0.015   0.9881
## initial_BMI:shift2pm:CalcTMM   4.828e-04  3.984e-04   1.212   0.2255
## initial_BMI:shiftother:CalcTMM 6.881e-05  6.422e-04   0.107   0.9147
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 311.15  on 248  degrees of freedom
## Residual deviance: 253.08  on 213  degrees of freedom
## AIC: 325.08
##
## Number of Fisher Scoring iterations: 15
```

```
anova(subset6a.model)
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: WG
##
## Terms added sequentially (first to last)
##
##
##                          Df Deviance Resid. Df Resid. Dev
## NULL                                       248     311.15
## initial_BMI               1   2.9539       247     308.19
## shift                     8   5.7818       239     302.41
## CalcTMM                   1   1.9492       238     300.46
## initial_BMI:shift         8  11.4590       230     289.00
## initial_BMI:CalcTMM       1   0.6388       229     288.37
## shift:CalcTMM             8  17.4487       221     270.92
## initial_BMI:shift:CalcTMM 8  17.8323       213     253.08
```

```
summary(stepAIC(subset6a.model, direction="both"))
```

```
## Start:  AIC=325.08
## WG ~ initial_BMI * shift * CalcTMM
```

```
##
##                              Df Deviance    AIC
## <none>                            253.08 325.08
## - initial_BMI:shift:CalcTMM  8   270.92 326.92
```

```
##
## Call:
## glm(formula = WG ~ initial_BMI * shift * CalcTMM, family = binomial,
##     data = subset6a.dat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.2295  -0.9711   0.4149   0.8852   1.7377
##
## Coefficients:
##                                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)                   -2.628e-01  6.010e+00  -0.044   0.9651
## initial_BMI                    2.560e-02  2.107e-01   0.122   0.9033
## shift8am                       2.106e+00  6.172e+00   0.341   0.7329
## shift9am                       1.581e+00  6.427e+00   0.246   0.8057
## shift10am                      3.699e+00  7.138e+00   0.518   0.6044
## shift11am                      2.334e+00  6.698e+00   0.348   0.7275
## shift12pm                     -3.378e+01  3.150e+01  -1.072   0.2836
## shift1pm                      -1.231e+02  9.497e+03  -0.013   0.9897
## shift2pm                       9.971e+00  1.507e+01   0.662   0.5082
## shiftother                    -9.214e-01  1.075e+01  -0.086   0.9317
## CalcTMM                        7.085e-03  6.059e-03   1.169   0.2423
## initial_BMI:shift8am          -6.524e-02  2.174e-01  -0.300   0.7641
## initial_BMI:shift9am          -4.026e-02  2.243e-01  -0.180   0.8575
## initial_BMI:shift10am         -4.925e-02  2.552e-01  -0.193   0.8470
## initial_BMI:shift11am         -1.263e-01  2.412e-01  -0.524   0.6005
## initial_BMI:shift12pm          1.555e+00  1.404e+00   1.108   0.2679
## initial_BMI:shift1pm           5.311e+00  3.915e+02   0.014   0.9892
## initial_BMI:shift2pm          -4.352e-01  6.045e-01  -0.720   0.4716
## initial_BMI:shiftother         5.928e-03  3.751e-01   0.016   0.9874
## initial_BMI:CalcTMM           -2.448e-04  2.083e-04  -1.175   0.2401
## shift8am:CalcTMM              -7.153e-03  6.101e-03  -1.172   0.2411
## shift9am:CalcTMM              -5.916e-03  6.316e-03  -0.937   0.3489
## shift10am:CalcTMM             -1.260e-02  7.094e-03  -1.776   0.0757 .
## shift11am:CalcTMM             -6.444e-04  7.331e-03  -0.088   0.9300
## shift12pm:CalcTMM              3.956e-02  3.951e-02   1.001   0.3167
## shift1pm:CalcTMM               6.347e-02  4.363e+00   0.015   0.9884
## shift2pm:CalcTMM              -1.103e-02  9.804e-03  -1.125   0.2604
## shiftother:CalcTMM             3.146e-04  1.840e-02   0.017   0.9864
## initial_BMI:shift8am:CalcTMM   2.398e-04  2.103e-04   1.140   0.2542
## initial_BMI:shift9am:CalcTMM   1.948e-04  2.196e-04   0.887   0.3750
## initial_BMI:shift10am:CalcTMM  4.277e-04  2.531e-04   1.690   0.0910 .
## initial_BMI:shift11am:CalcTMM  5.416e-05  2.462e-04   0.220   0.8259
## initial_BMI:shift12pm:CalcTMM -1.679e-03  1.610e-03  -1.043   0.2969
## initial_BMI:shift1pm:CalcTMM  -2.472e-03  1.662e-01  -0.015   0.9881
## initial_BMI:shift2pm:CalcTMM   4.828e-04  3.984e-04   1.212   0.2255
## initial_BMI:shiftother:CalcTMM 6.881e-05  6.422e-04   0.107   0.9147
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 311.15  on 248  degrees of freedom
## Residual deviance: 253.08  on 213  degrees of freedom
## AIC: 325.08
##
## Number of Fisher Scoring iterations: 15
```

```
subset6b.dat <- gained.dat[which(complete.cases(gained.dat[,c("initial_bweight", "shift", "CalcTMM")]))
subset6b.model <- glm(WG~ initial_bweight*shift*CalcTMM, data=subset6b.dat, family=binomial)
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

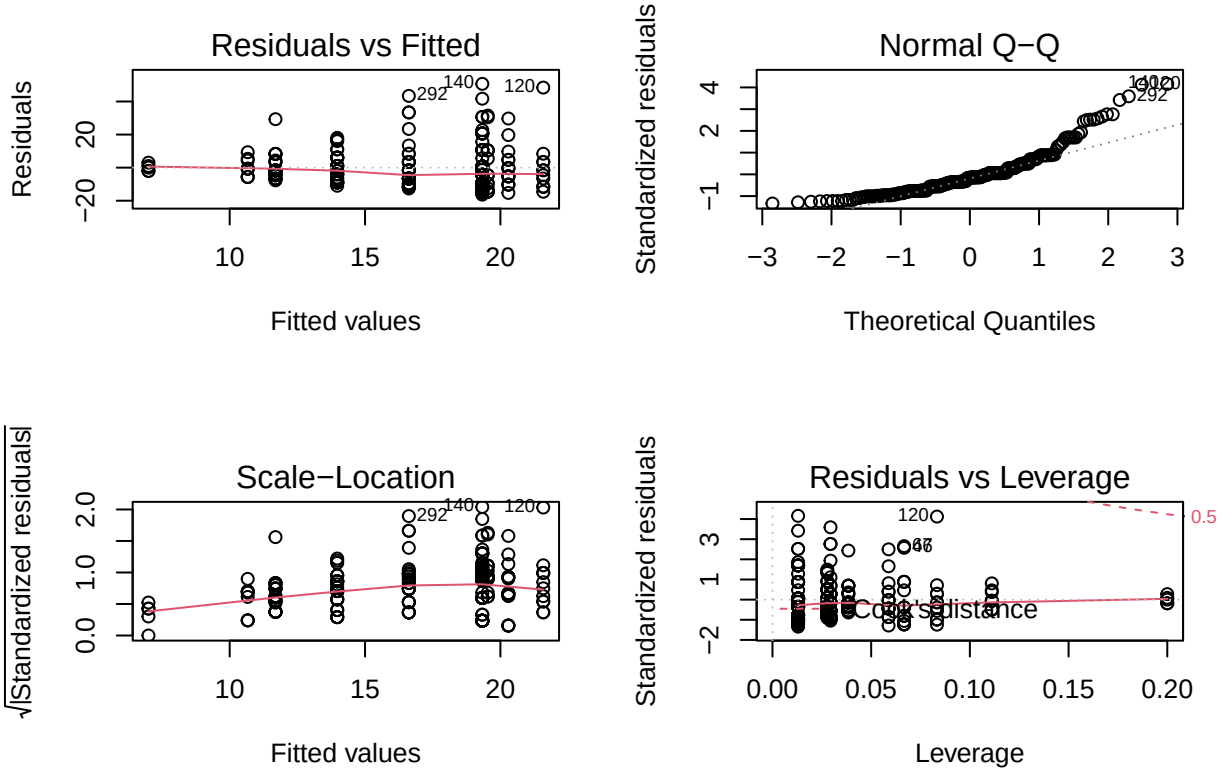
```
stepAIC(subset6b.model, direction="both")
```

```
## Start:  AIC=342.84
## WG ~ initial_bweight * shift * CalcTMM
##
##                                 Df Deviance    AIC
## <none>                               270.84 342.84
## - initial_bweight:shift:CalcTMM  8   295.76 351.76
```

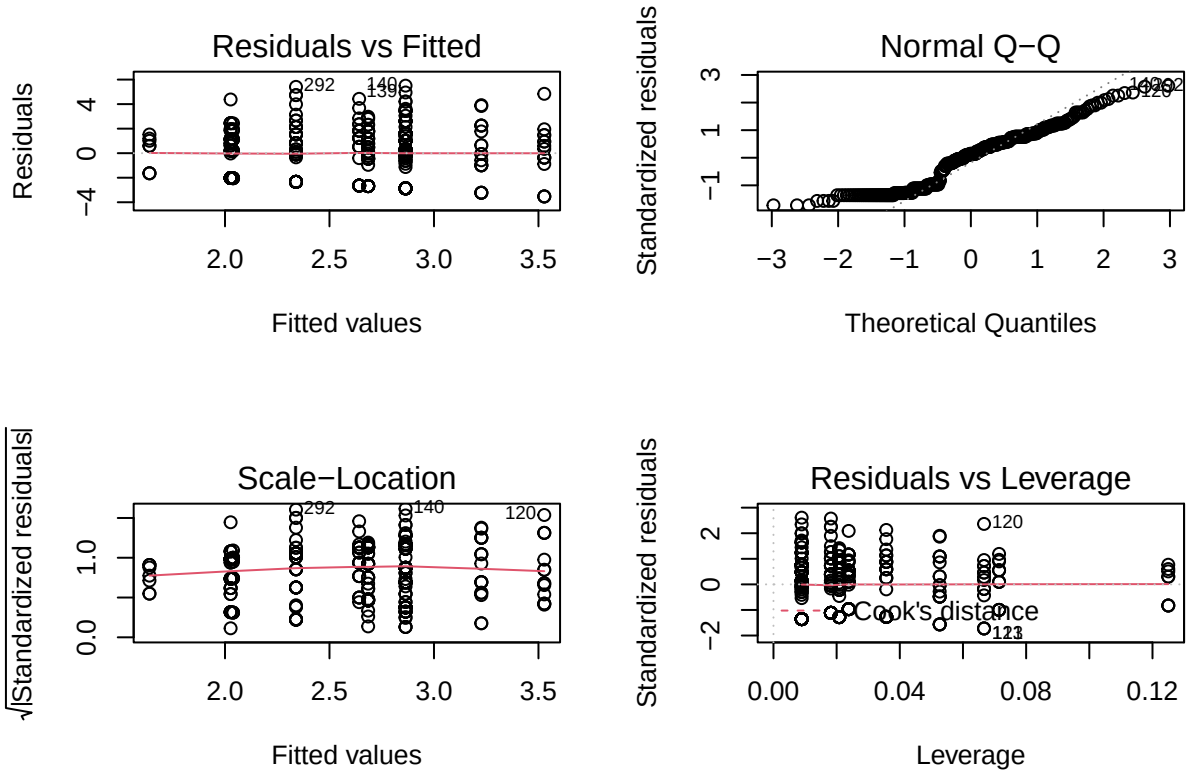
```
##
## Call:  glm(formula = WG ~ initial_bweight * shift * CalcTMM, family = binomial,
##     data = subset6b.dat)
##
## Coefficients:
##                         (Intercept)                      initial_bweight
##                           -1.171e+00                            6.679e-03
##                             shift8am                             shift9am
##                            2.956e+00                            4.596e+00
##                            shift10am                            shift11am
##                            3.806e+00                            2.077e+00
##                            shift12pm                             shift1pm
##                           -1.470e+03                           -1.370e+02
##                             shift2pm                           shiftother
##                            9.155e+00                            1.744e+00
##                              CalcTMM            initial_bweight:shift8am
##                            4.110e-03                           -1.263e-02
##           initial_bweight:shift9am           initial_bweight:shift10am
##                           -2.021e-02                           -8.342e-03
##          initial_bweight:shift11am           initial_bweight:shift12pm
##                           -1.623e-02                            8.984e+00
##           initial_bweight:shift1pm            initial_bweight:shift2pm
##                            1.062e+00                           -5.852e-02
##         initial_bweight:shiftother              initial_bweight:CalcTMM
##                           -1.049e-02                           -1.986e-05
##                    shift8am:CalcTMM                     shift9am:CalcTMM
##                           -4.223e-03                           -4.397e-03
##                   shift10am:CalcTMM                    shift11am:CalcTMM
##                           -6.391e-03                            5.051e-04
```

```
##                   shift12pm:CalcTMM                     shift1pm:CalcTMM
##                            2.910e+00                            7.789e-02
##                    shift2pm:CalcTMM                   shiftother:CalcTMM
##                           -6.893e-03                           -4.326e-03
##   initial_bweight:shift8am:CalcTMM    initial_bweight:shift9am:CalcTMM
##                            1.946e-05                            2.085e-05
##  initial_bweight:shift10am:CalcTMM   initial_bweight:shift11am:CalcTMM
##                            2.992e-05                            1.521e-06
##  initial_bweight:shift12pm:CalcTMM    initial_bweight:shift1pm:CalcTMM
##                           -1.641e-02                           -5.248e-04
##   initial_bweight:shift2pm:CalcTMM  initial_bweight:shiftother:CalcTMM
##                            5.005e-05                            2.939e-05
##
## Degrees of Freedom: 263 Total (i.e. Null);  228 Residual
## Null Deviance:       331.8
## Residual Deviance: 270.8     AIC: 342.8
```

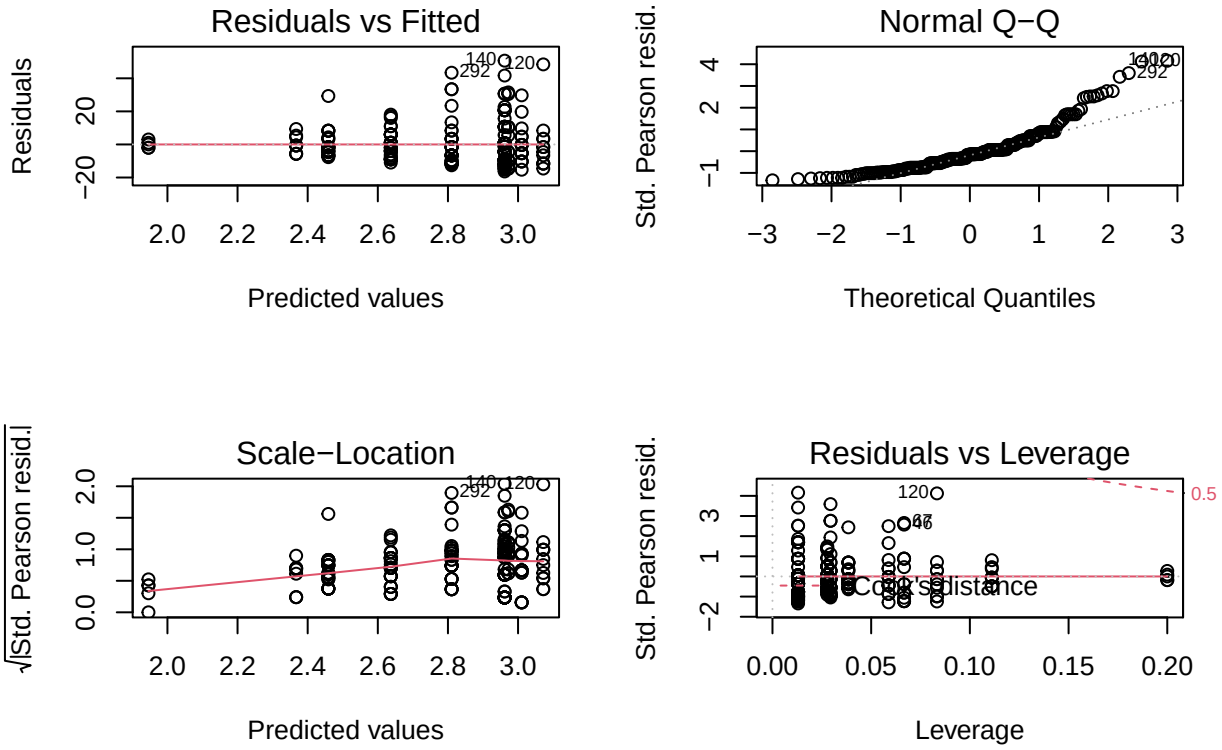
```
SA2.model1.lm <- lm(pounds_gained ~ shift, data=nonzero.dat)
par(mfrow=c(2,2))
plot(SA2.model1.lm)
```



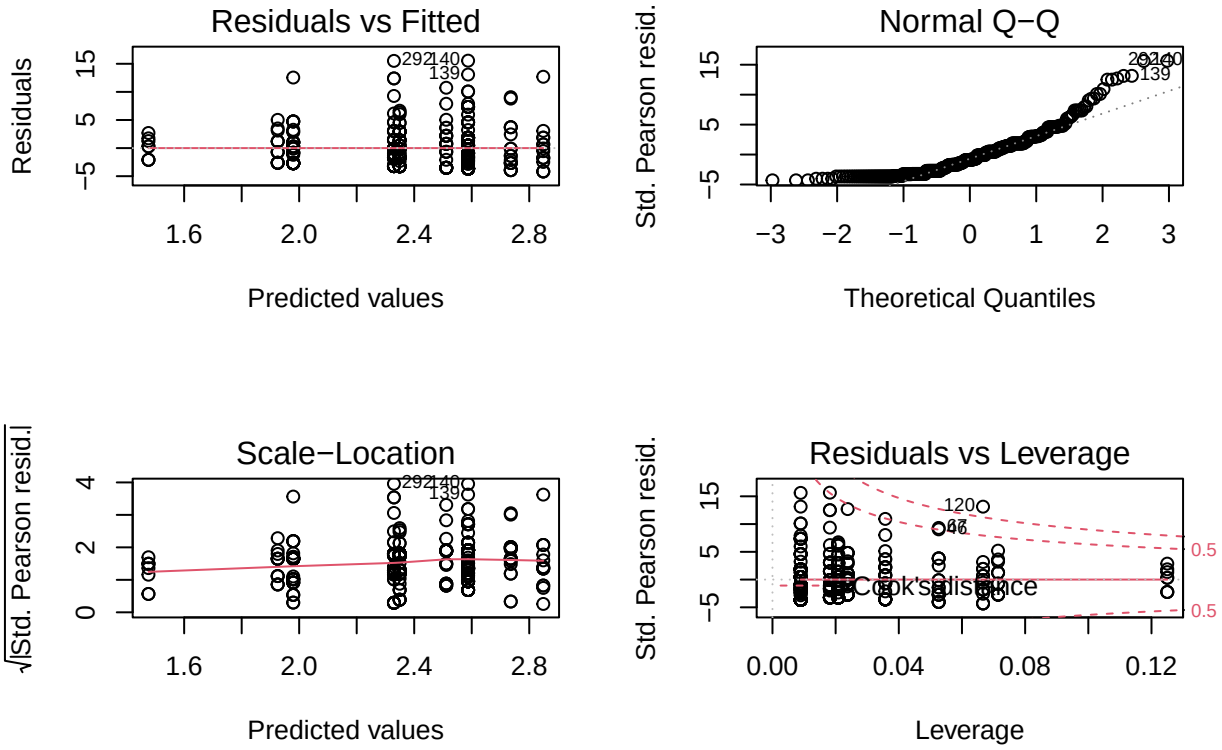

```
SA2.model1.root <- lm(sqrt(pounds_gained) ~ shift, data=gained.dat)
par(mfrow=c(2,2))
plot(SA2.model1.root)
```




```
SA2b.model1.loggauss <- glm(pounds_gained ~ shift, data=nonzero.dat,family = gaussian(link="log"))
par(mfrow=c(2,2))
plot(SA2b.model1.loggauss)
```




```
SA2.model1.poisson <- glm(LBS ~ shift, data=gained.dat,family = poisson)
par(mfrow=c(2,2))
plot(SA2.model1.poisson)
```

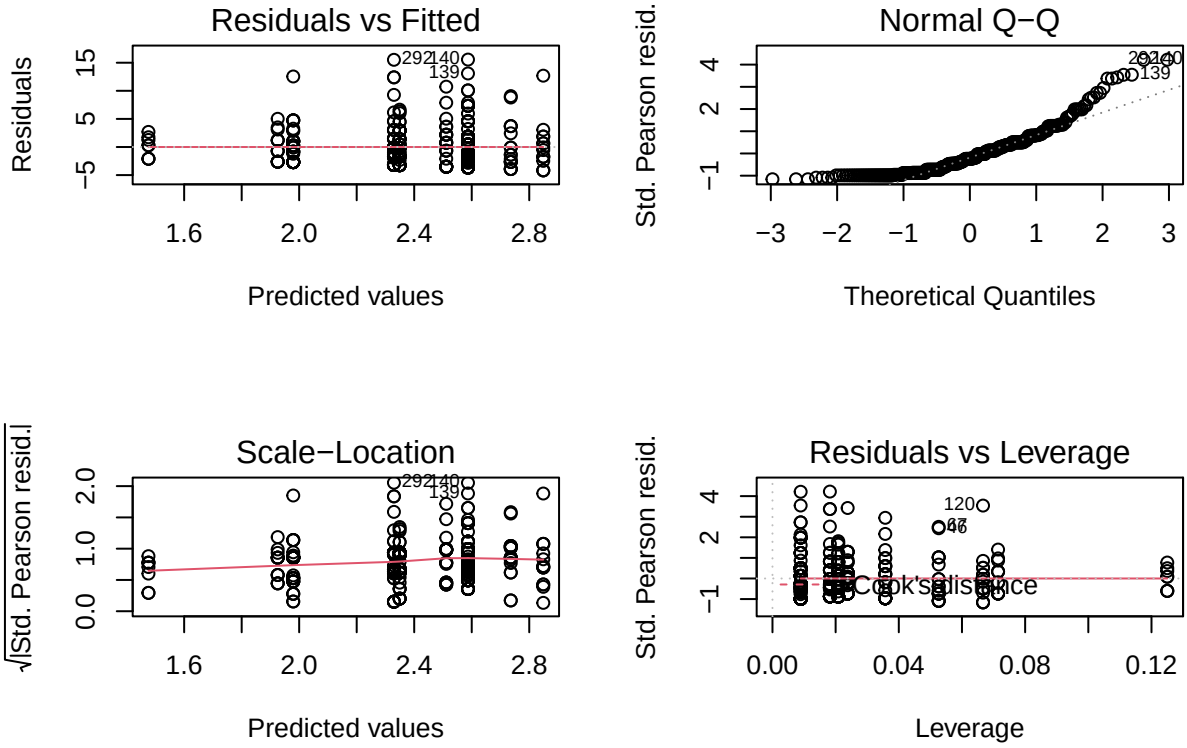



```
summary(SA2.model1.poisson)
```

```
##
## Call:
## glm(formula = LBS ~ shift, family = poisson, data = gained.dat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -5.8765  -4.5327  -0.9454   1.7100  10.9155
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.51134    0.05384  46.648  < 2e-16 ***
## shift8am     0.07602    0.05975   1.272  0.20325
## shift9am    -0.18185    0.06832  -2.662  0.00778 **
## shift10am   -0.16195    0.06990  -2.317  0.02051 *
## shift11am   -0.53198    0.07866  -6.763 1.35e-11 ***
## shift12pm   -0.58605    0.11539  -5.079 3.80e-07 ***
## shift1pm    -1.03543    0.17740  -5.837 5.32e-09 ***
## shift2pm     0.33744    0.08222   4.104 4.06e-05 ***
## shiftother   0.22439    0.07944   2.825  0.00473 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
```

```
##
##     Null deviance: 4843.1  on 340  degrees of freedom
## Residual deviance: 4585.9  on 332  degrees of freedom
## AIC: 5629.6
##
## Number of Fisher Scoring iterations: 5
```

```
SA2.model1.quasi <- glm(pounds_gained ~ shift, data=gained.dat,family = quasipoisson)
par(mfrow=c(2,2))
plot(SA2.model1.quasi)
```

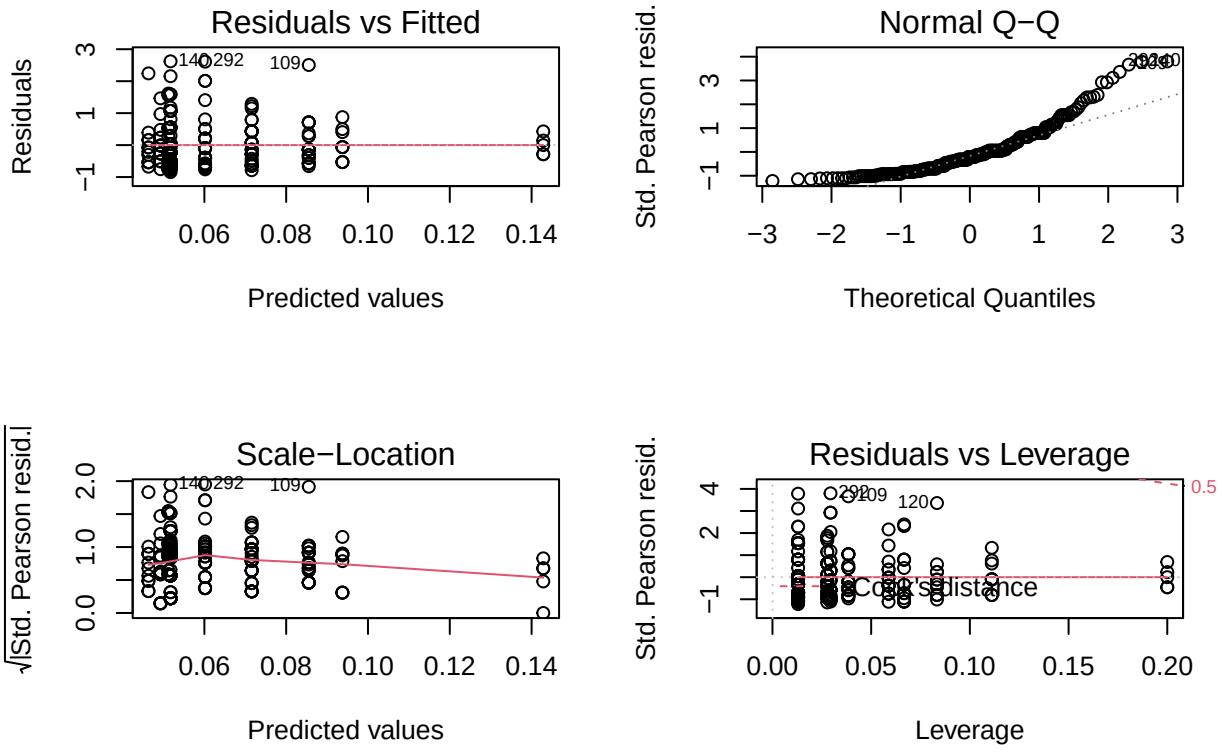



```
summary(SA2.model1.quasi)
```

```
##
## Call:
## glm(formula = pounds_gained ~ shift, family = quasipoisson, data = gained.dat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -5.8765  -4.5327  -0.9442   1.7113  10.9172
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.51134    0.19968  12.577   <2e-16 ***
## shift8am     0.07569    0.22162   0.342   0.7329
```

```
## shift9am    -0.18185    0.25341  -0.718   0.4735
## shift10am   -0.16195    0.25927  -0.625   0.5326
## shift11am   -0.53198    0.29176  -1.823   0.0691 .
## shift12pm   -0.58605    0.42799  -1.369   0.1718
## shift1pm    -1.03543    0.65797  -1.574   0.1165
## shift2pm     0.33744    0.30494   1.107   0.2693
## shiftother   0.22439    0.29466   0.762   0.4469
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 13.75713)
##
##     Null deviance: 4843.5  on 340  degrees of freedom
## Residual deviance: 4586.5  on 332  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 5
```

```
SA2b.model1.Gamma <- glm(pounds_gained ~ shift, data=nonzero.dat,family = Gamma)
par(mfrow=c(2,2))
plot(SA2b.model1.Gamma)
```




```
SA2.model1.zero <- zeroinfl(LBS ~ shift, data = gained.dat)
summary(SA2.model1.zero)
```

```
##
```

```
## Call:
## zeroinfl(formula = LBS ~ shift, data = gained.dat)
##
## Pearson residuals:
##     Min      1Q  Median      3Q     Max
## -1.8022 -1.1719 -0.3405  0.6929  5.8602
##
## Count model coefficients (poisson with log link):
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.01033    0.05384  55.914  < 2e-16 ***
## shift8am    -0.04828    0.05975  -0.808  0.41910
## shift9am    -0.19987    0.06833  -2.925  0.00344 **
## shift10am   -0.37326    0.06990  -5.340 9.32e-08 ***
## shift11am   -0.55141    0.07867  -7.009 2.39e-12 ***
## shift12pm   -0.64323    0.11540  -5.574 2.49e-08 ***
## shift1pm    -1.06534    0.17792  -5.988 2.13e-09 ***
## shift2pm     0.06159    0.08222   0.749  0.45378
## shiftother  -0.03821    0.07945  -0.481  0.63056
##
## Zero-inflation model coefficients (binomial with logit link):
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.43539    0.38696  -1.125    0.261
## shift8am    -0.35314    0.43737  -0.807    0.419
## shift9am    -0.04652    0.47620  -0.098    0.922
## shift10am   -0.66329    0.51074  -1.299    0.194
## shift11am   -0.05019    0.50070  -0.100    0.920
## shift12pm   -0.15247    0.67888  -0.225    0.822
## shift1pm    -0.07558    0.82729  -0.091    0.927
## shift2pm    -0.95098    0.75261  -1.264    0.206
## shiftother  -0.88644    0.68295  -1.298    0.194
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of iterations in BFGS optimization: 2
## Log-likelihood: -1539 on 18 Df
```