Practicum 1 Analysis

Peter Claussen Ben Derenge Stephanie Liebl

9/16/2021

Analysis

Data Processing

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

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$shift <- factor(metrics.dat$shift,shift.levels)
summary(metrics.dat$shift)
```

```
##     7am     8am     9am    10am    11am    12pm     1pm     2pm   other missing
##      31     115      56      50      44      14       8      15      15       4
```

```
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.

```
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
```

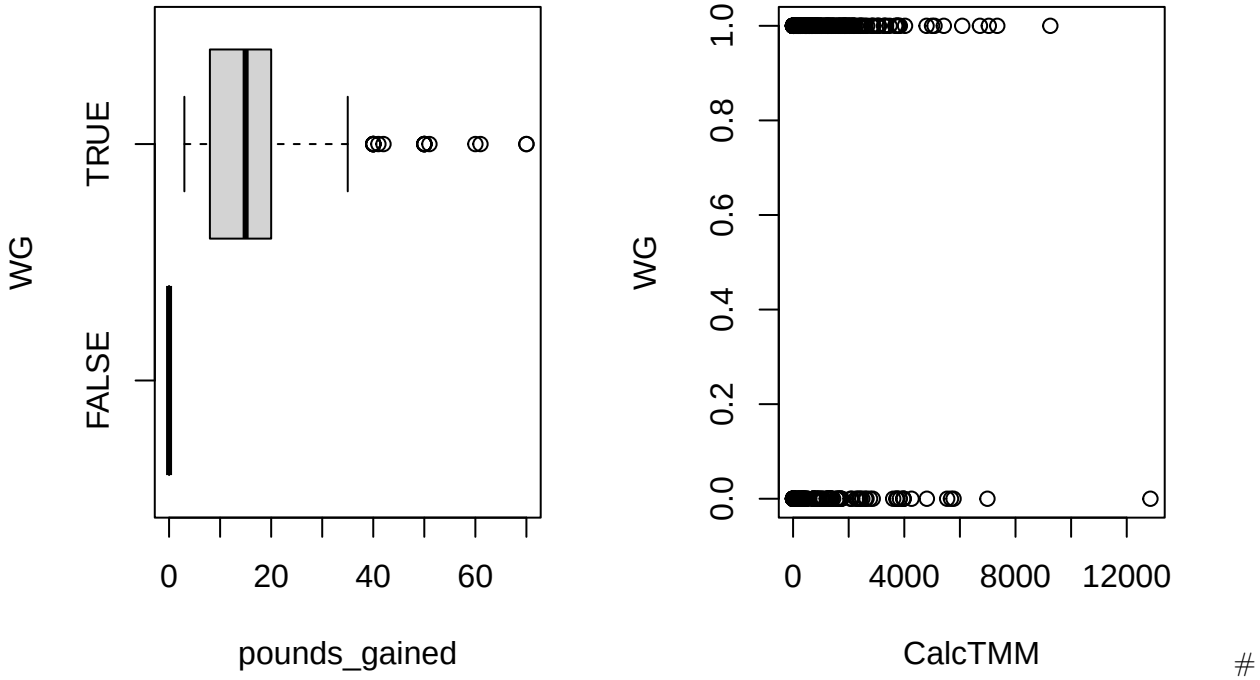
```
gained.dat <- gained.dat[!is.na(gained.dat$CalcTMM),]
dim(gained.dat)
```

```
## [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*?

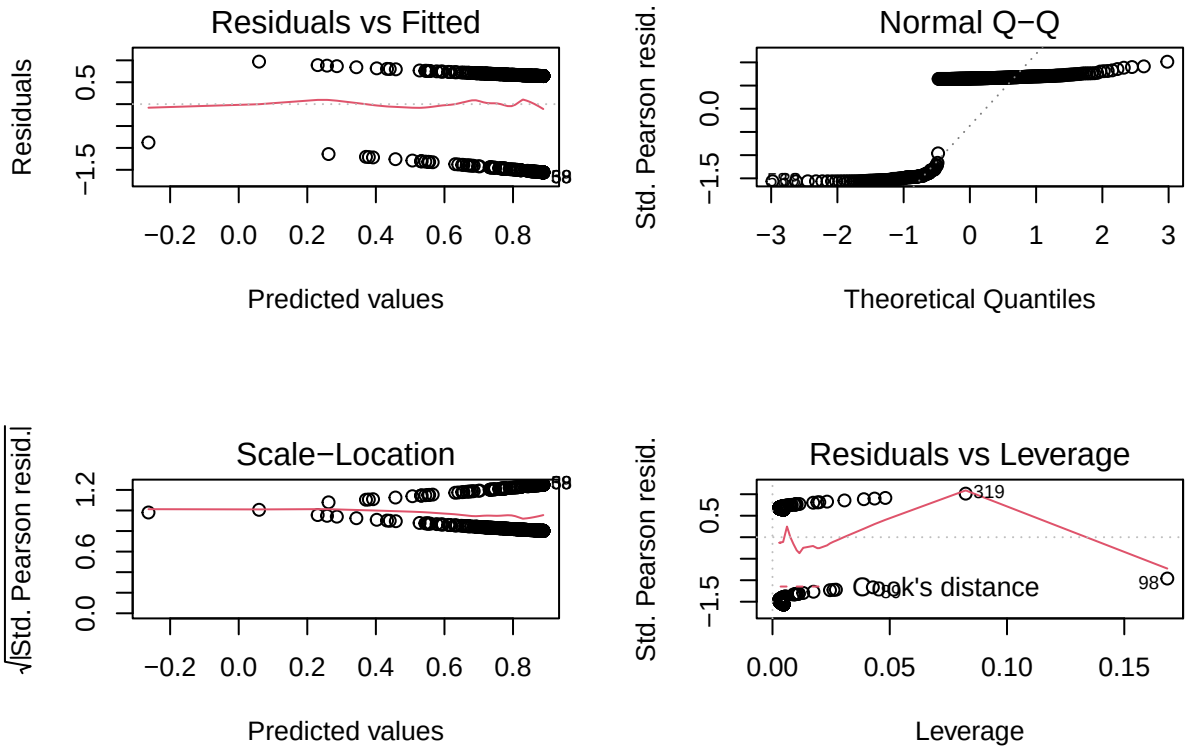
Simple logistic regression

```
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)
```



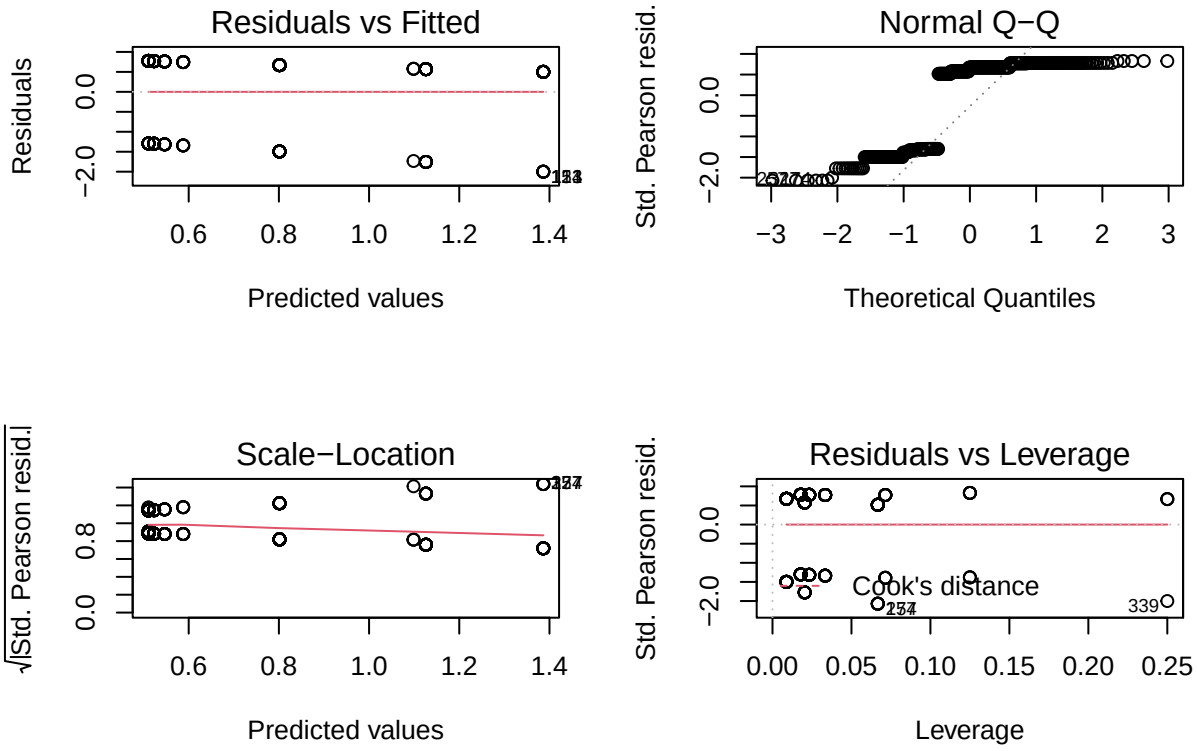
- **(SA2)** Does *shift* have an effect on *weight gain*?

```
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   0.7585   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    0.83975    0.74847   1.122    0.262
## shiftmissing  0.55207    1.21527   0.454    0.650
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 433.47  on 346  degrees of freedom
## Residual deviance: 428.07  on 337  degrees of freedom
## AIC: 448.07
##
## Number of Fisher Scoring iterations: 4
```

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



Model 2 Interactions

```
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
## -1.8851  -1.4015   0.8104   0.9193   1.2372
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)   6.964e-01  3.935e-01   1.770   0.0768 .
## shift8am      2.485e-01  4.312e-01   0.576   0.5645
## shift9am     -5.053e-02  4.702e-01  -0.107   0.9144
## shift10am     5.807e-01  5.049e-01   1.150   0.2501
## shift11am    -5.755e-02  4.947e-01  -0.116   0.9074
## shift12pm     2.385e-02  6.758e-01   0.035   0.9718
## shift1pm     -5.826e-02  8.241e-01  -0.071   0.9436
## shift2pm      8.839e-01  7.514e-01   1.176   0.2395
## shiftother    8.951e-01  7.540e-01   1.187   0.2352
## shiftmissing  5.872e-01  1.218e+00   0.482   0.6297
## CalcTMM      -1.069e-04  7.262e-05  -1.473   0.1408
## ---
## 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: 425.92  on 336  degrees of freedom
## AIC: 447.92
##
## Number of Fisher Scoring iterations: 4
```

Model 3 SA1 and 2 plus anthropometric variables

```
SA12.model3a <- glm(WG ~ gender + Age + shift + CalcTMM, data=gained.dat,family = binomial)
summary(SA12.model3a)
```

```
##
## Call:
## glm(formula = WG ~ gender + Age + shift + CalcTMM, family = binomial,
##     data = gained.dat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9936  -1.3159   0.7661   0.8673   1.4101
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)   1.897e-01  1.623e+00   0.117    0.907
## genderFemale  1.240e+00  1.472e+00   0.842    0.400
## genderMale    7.071e-01  1.486e+00   0.476    0.634
## Age          -6.189e-03  1.263e-02  -0.490    0.624
## shift8am     -1.491e-01  4.821e-01  -0.309    0.757
```

```
## shift9am     -3.842e-01  5.240e-01  -0.733    0.464
## shift10am     1.715e-01  5.524e-01   0.310    0.756
## shift11am    -1.932e-01  5.542e-01  -0.349    0.727
## shift12pm    -4.140e-01  7.192e-01  -0.576    0.565
## shift1pm     -4.538e-01  8.572e-01  -0.529    0.597
## shift2pm      1.331e+00  1.151e+00   1.157    0.247
## shiftother    5.339e-01  8.128e-01   0.657    0.511
## shiftmissing  1.374e+01  6.107e+02   0.022    0.982
## CalcTMM      -1.102e-04  7.636e-05  -1.444    0.149
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 393.55  on 318  degrees of freedom
## Residual deviance: 379.52  on 305  degrees of freedom
##   (28 observations deleted due to missingness)
## AIC: 407.52
##
## Number of Fisher Scoring iterations: 13
```

```
SA12.model3b <- glm(WG ~ gender + Age + height +shift + CalcTMM, data=gained.dat,family = binomial)
summary(SA12.model3b)
```

```
##
## Call:
## glm(formula = WG ~ gender + Age + height + shift + CalcTMM, family = binomial,
##     data = gained.dat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9392  -1.2928   0.7580   0.8566   1.3899
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.657e-01  3.381e+00  -0.049    0.961
## genderFemale  1.268e+00  1.489e+00   0.852    0.394
## genderMale    6.805e-01  1.486e+00   0.458    0.647
## Age          -2.496e-03  1.294e-02  -0.193    0.847
## height        3.124e-03  4.170e-02   0.075    0.940
## shift8am     -1.112e-01  4.834e-01  -0.230    0.818
## shift9am     -3.525e-01  5.278e-01  -0.668    0.504
## shift10am     2.318e-01  5.611e-01   0.413    0.680
## shift11am    -1.797e-01  5.549e-01  -0.324    0.746
## shift12pm    -4.107e-01  7.199e-01  -0.571    0.568
## shift1pm     -4.402e-01  8.578e-01  -0.513    0.608
## shift2pm      1.339e+00  1.153e+00   1.161    0.246
## shiftother    4.589e-01  8.196e-01   0.560    0.576
## shiftmissing  1.375e+01  6.104e+02   0.023    0.982
## CalcTMM      -1.052e-04  7.680e-05  -1.370    0.171
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 384.25  on 312  degrees of freedom
## Residual deviance: 370.32  on 298  degrees of freedom
##   (34 observations deleted due to missingness)
## AIC: 400.32
```

```
##
## Number of Fisher Scoring iterations: 13
```

Model 4 Partition CalcTMM into components

```
SA12.model4a <- glm(WG ~ gender + Age + shift + Vig.ex.Time  + Mod.ex.time + Walk.ex.Time, data=gained.
summary(SA12.model4a)
```

```
##
## Call:
## glm(formula = WG ~ gender + Age + shift + Vig.ex.Time + Mod.ex.time +
##     Walk.ex.Time, family = binomial, data = gained.dat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9919  -1.3054   0.7611   0.8657   1.4278
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)   3.761e-01  1.662e+00   0.226    0.821
## genderFemale  1.046e+00  1.517e+00   0.690    0.490
## genderMale    5.222e-01  1.527e+00   0.342    0.732
## Age          -6.569e-03  1.269e-02  -0.518    0.605
## shift8am     -1.245e-01  4.846e-01  -0.257    0.797
## shift9am     -3.901e-01  5.260e-01  -0.742    0.458
## shift10am     2.003e-01  5.557e-01   0.360    0.719
## shift11am    -1.591e-01  5.583e-01  -0.285    0.776
## shift12pm    -4.125e-01  7.198e-01  -0.573    0.567
## shift1pm     -4.500e-01  8.575e-01  -0.525    0.600
## shift2pm      1.336e+00  1.160e+00   1.152    0.249
## shiftother    5.447e-01  8.137e-01   0.669    0.503
## shiftmissing  1.376e+01  6.103e+02   0.023    0.982
## Vig.ex.Time  -1.434e-03  1.231e-03  -1.165    0.244
## Mod.ex.time  -4.887e-05  1.167e-03  -0.042    0.967
## Walk.ex.Time -2.474e-04  5.875e-04  -0.421    0.674
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 393.55  on 318  degrees of freedom
## Residual deviance: 379.25  on 303  degrees of freedom
##   (28 observations deleted due to missingness)
## AIC: 411.25
##
## Number of Fisher Scoring iterations: 13
```

```
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 = gained.dat)
##
```

```
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9366  -1.2871   0.7556   0.8564   1.3901
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -3.843e-03  3.414e+00  -0.001    0.999
## genderFemale  1.092e+00  1.540e+00   0.709    0.478
## genderMale    5.064e-01  1.531e+00   0.331    0.741
## Age          -2.712e-03  1.301e-02  -0.208    0.835
## height        3.138e-03  4.175e-02   0.075    0.940
## shift8am     -8.385e-02  4.860e-01  -0.173    0.863
## shift9am     -3.485e-01  5.298e-01  -0.658    0.511
## shift10am     2.608e-01  5.638e-01   0.463    0.644
## shift11am    -1.447e-01  5.592e-01  -0.259    0.796
## shift12pm    -4.137e-01  7.202e-01  -0.574    0.566
## shift1pm     -4.334e-01  8.579e-01  -0.505    0.613
## shift2pm      1.321e+00  1.163e+00   1.136    0.256
## shiftother    4.707e-01  8.202e-01   0.574    0.566
## shiftmissing  1.377e+01  6.105e+02   0.023    0.982
## Vig.ex.Time  -1.348e-03  1.259e-03  -1.071    0.284
## Mod.ex.time   1.043e-04  1.198e-03   0.087    0.931
## Walk.ex.Time -3.354e-04  5.892e-04  -0.569    0.569
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 384.25  on 312  degrees of freedom
## Residual deviance: 370.06  on 296  degrees of freedom
##   (34 observations deleted due to missingness)
## AIC: 404.06
##
## Number of Fisher Scoring iterations: 13
```

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

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

```
SA12.model4a <- glm(WG ~ gender + Age + shift + Vig.ex.Time  + Mod.ex.time + Walk.ex.Time + initial_BMI
summary(SA12.model4a)
```

```
##
## Call:
## glm(formula = WG ~ gender + Age + shift + Vig.ex.Time + Mod.ex.time +
##     Walk.ex.Time + initial_BMI, family = binomial, data = gained.dat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.0146  -1.1930   0.7081   0.8825   1.3010
##
## Coefficients:
```

```
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)   5.545e-01  1.816e+00   0.305   0.7602
## genderFemale  1.475e+00  1.555e+00   0.949   0.3428
## genderMale    9.743e-01  1.570e+00   0.620   0.5349
## Age           1.258e-02  1.536e-02   0.819   0.4128
## shift8am     -2.794e-01  5.425e-01  -0.515   0.6065
## shift9am      1.927e-01  6.331e-01   0.304   0.7609
## shift10am     5.366e-01  6.421e-01   0.836   0.4034
## shift11am    -2.815e-01  6.246e-01  -0.451   0.6523
## shift12pm     3.545e-01  8.427e-01   0.421   0.6740
## shift1pm      6.780e-02  1.023e+00   0.066   0.9471
## shift2pm      1.590e+01  9.545e+02   0.017   0.9867
## shiftother    4.309e-01  8.647e-01   0.498   0.6182
## shiftmissing  1.613e+01  1.675e+03   0.010   0.9923
## Vig.ex.Time  -4.905e-04  1.427e-03  -0.344   0.7310
## Mod.ex.time  -5.897e-04  1.473e-03  -0.400   0.6889
## Walk.ex.Time -3.366e-04  6.597e-04  -0.510   0.6098
## initial_BMI  -5.306e-02  2.664e-02  -1.992   0.0464 *
## ---
## 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: 280.85  on 225  degrees of freedom
##   (105 observations deleted due to missingness)
## AIC: 314.85
##
## Number of Fisher Scoring iterations: 15
```

```
SA12.model4b <- glm(WG ~ gender + Age + height + shift + Vig.ex.Time  + Mod.ex.time + Walk.ex.Time +ini
summary(SA12.model4b)
```

```
##
## Call:
## glm(formula = WG ~ gender + Age + height + shift + Vig.ex.Time +
##     Mod.ex.time + Walk.ex.Time + initial_bweight, family = binomial,
##     data = gained.dat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9710  -1.2121   0.7130   0.8866   1.3026
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)     -2.567e+00  3.891e+00  -0.660   0.5095
## genderFemale     1.520e+00  1.581e+00   0.962   0.3362
## genderMale       9.834e-01  1.572e+00   0.626   0.5316
## Age              1.207e-02  1.539e-02   0.785   0.4326
## height           4.495e-02  5.134e-02   0.875   0.3813
## shift8am        -2.806e-01  5.430e-01  -0.517   0.6053
## shift9am         1.854e-01  6.329e-01   0.293   0.7696
## shift10am        5.229e-01  6.434e-01   0.813   0.4164
## shift11am       -2.799e-01  6.261e-01  -0.447   0.6549
## shift12pm        3.139e-01  8.410e-01   0.373   0.7090
```

```
## shift1pm         5.388e-02  1.018e+00   0.053   0.9578
## shift2pm         1.594e+01  9.521e+02   0.017   0.9866
## shiftother       4.353e-01  8.649e-01   0.503   0.6147
## shiftmissing     1.611e+01  1.670e+03   0.010   0.9923
## Vig.ex.Time     -4.464e-04  1.425e-03  -0.313   0.7541
## Mod.ex.time     -6.272e-04  1.471e-03  -0.426   0.6699
## Walk.ex.Time    -3.272e-04  6.621e-04  -0.494   0.6211
## initial_bweight -7.707e-03  4.215e-03  -1.828   0.0675 .
## ---
## 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: 281.36  on 224  degrees of freedom
##   (105 observations deleted due to missingness)
## AIC: 317.36
##
## Number of Fisher Scoring iterations: 15
```

Pounds Gained analysis

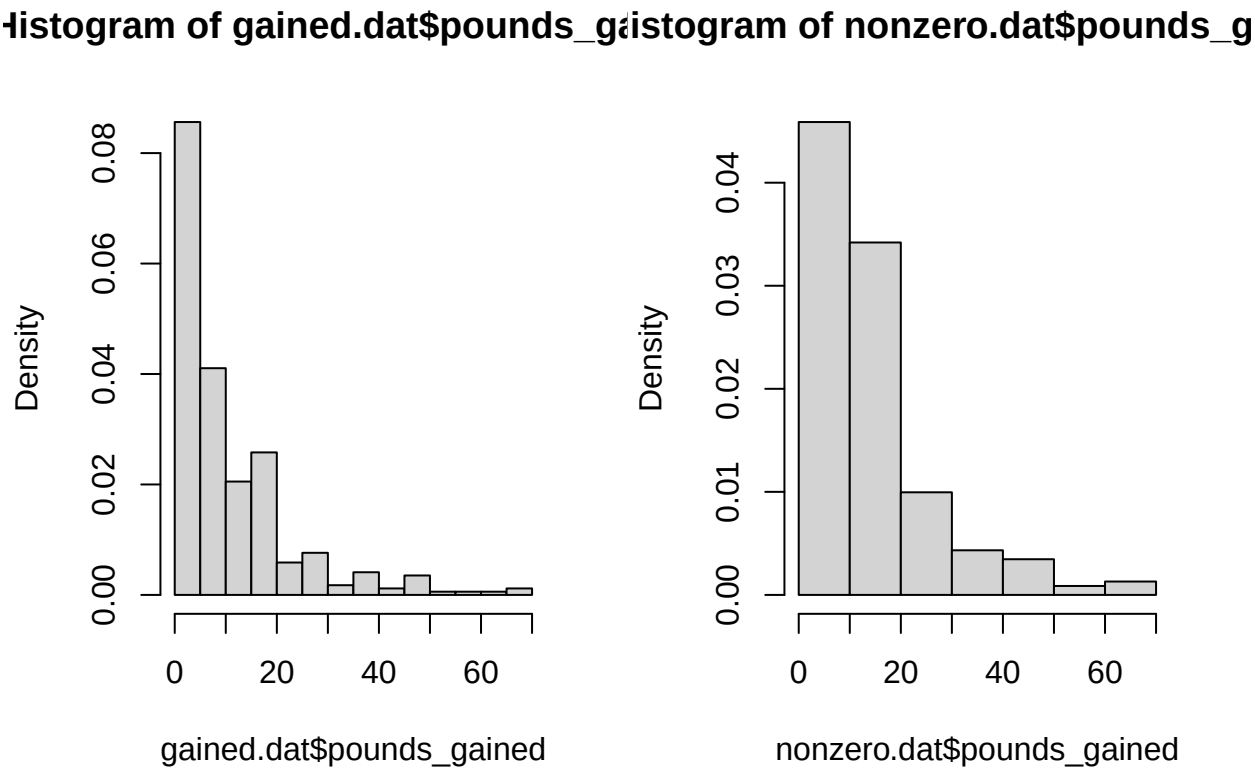
We have information about net pounds gained. We assume that when `weightgained` is false, we can substitute a value of 0 for `pounds_gained`. This allows us to analyze a full data set; otherwise, we limit our observations. It is worth noting that we may be oversimplifying cases were `pounds_gained` may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

```
sum(is.na(gained.dat$pounds_gained))
```

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

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

```
par(mfrow=c(1,2))
hist(gained.dat$pounds_gained, freq = FALSE)
nonzero.dat <- gained.dat[gained.dat$pounds_gained>0,]
hist(nonzero.dat$pounds_gained, freq = FALSE)
```



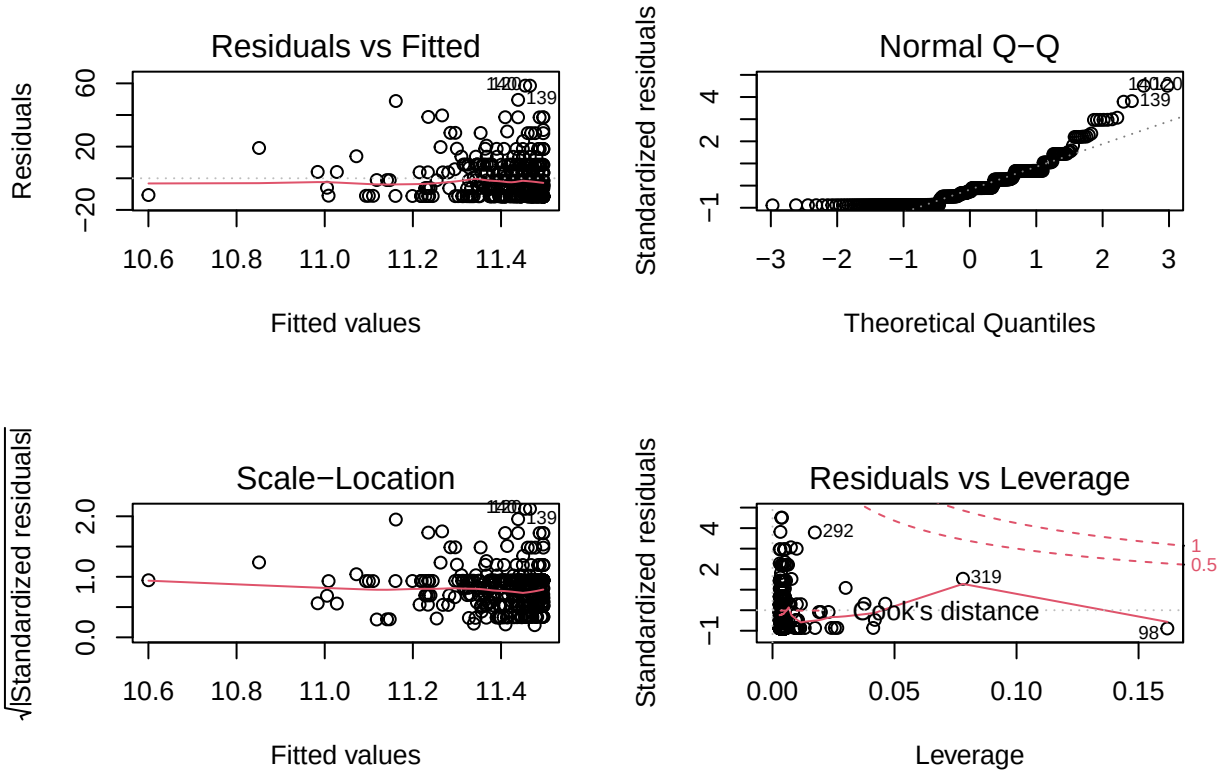
Pounds gained is highly skewed, even when zero observations are excluded. Thus, a linear model, with the assumption of normally distributed errors, may not be appropriate.

Linear model, gaussian errors

```
SA1.model1.lm <- lm(pounds_gained ~ CalcTMM, data=gained.dat)
summary(SA1.model1.lm)
```

```
##
## Call:
## lm(formula = pounds_gained ~ CalcTMM, data = gained.dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -11.496 -11.348  -3.441   6.672  58.545
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 11.4958780  0.9241793  12.439   <2e-16 ***
## CalcTMM     -0.0000697  0.0004500  -0.155    0.877
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.01 on 339 degrees of freedom
## Multiple R-squared:  7.076e-05,  Adjusted R-squared:  -0.002879
## F-statistic: 0.02399 on 1 and 339 DF,  p-value: 0.877
```

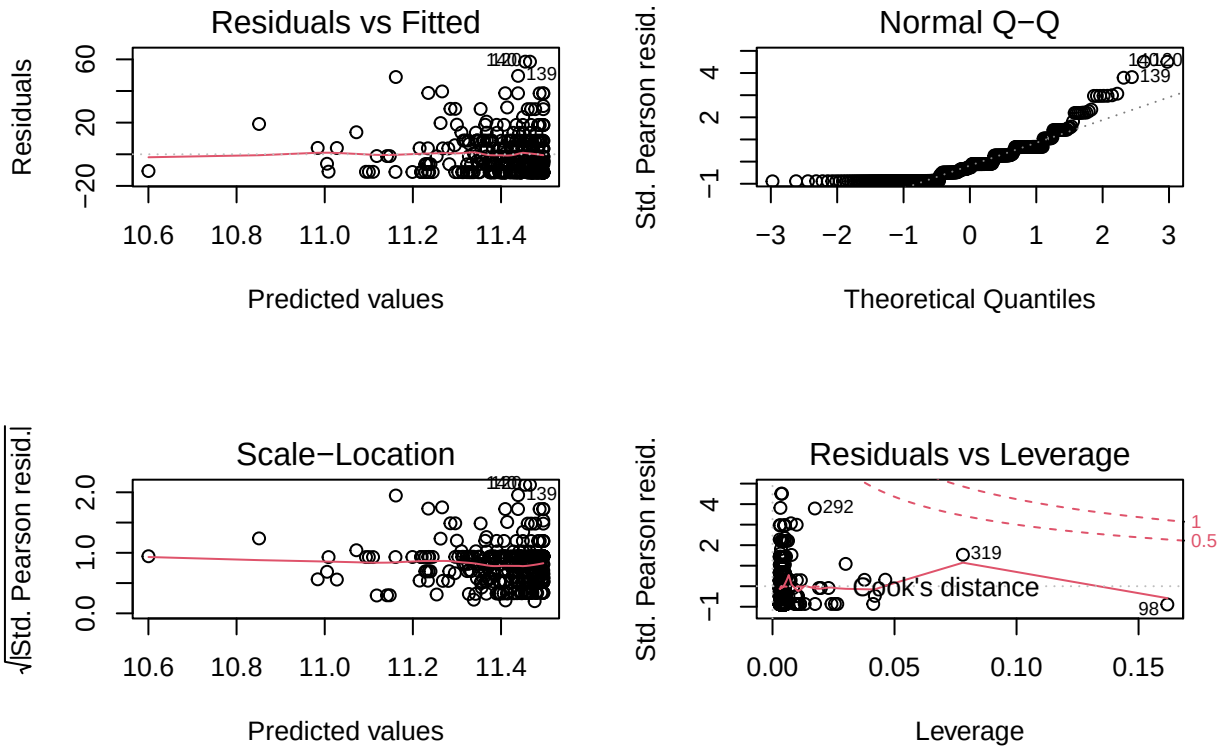
```
par(mfrow=c(2,2))
plot(SA1.model1.lm)
```




```
#equivalent model, just using glm instead of lm
SA1.model1.gauss <- glm(pounds_gained ~ CalcTMM, data=gained.dat,family = gaussian)
SA1b.model1.gauss <- glm(pounds_gained ~ CalcTMM, data=nonzero.dat,family = gaussian)
summary(SA1.model1.gauss)
```

```
##
## Call:
## glm(formula = pounds_gained ~ CalcTMM, family = gaussian, data = gained.dat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -11.496  -11.348   -3.441    6.672   58.545
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 11.4958780  0.9241793  12.439   <2e-16 ***
## CalcTMM     -0.0000697  0.0004500  -0.155    0.877
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 169.2323)
##
##     Null deviance: 57374  on 340  degrees of freedom
## Residual deviance: 57370  on 339  degrees of freedom
## AIC: 2721.5
##
## Number of Fisher Scoring iterations: 2
```

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



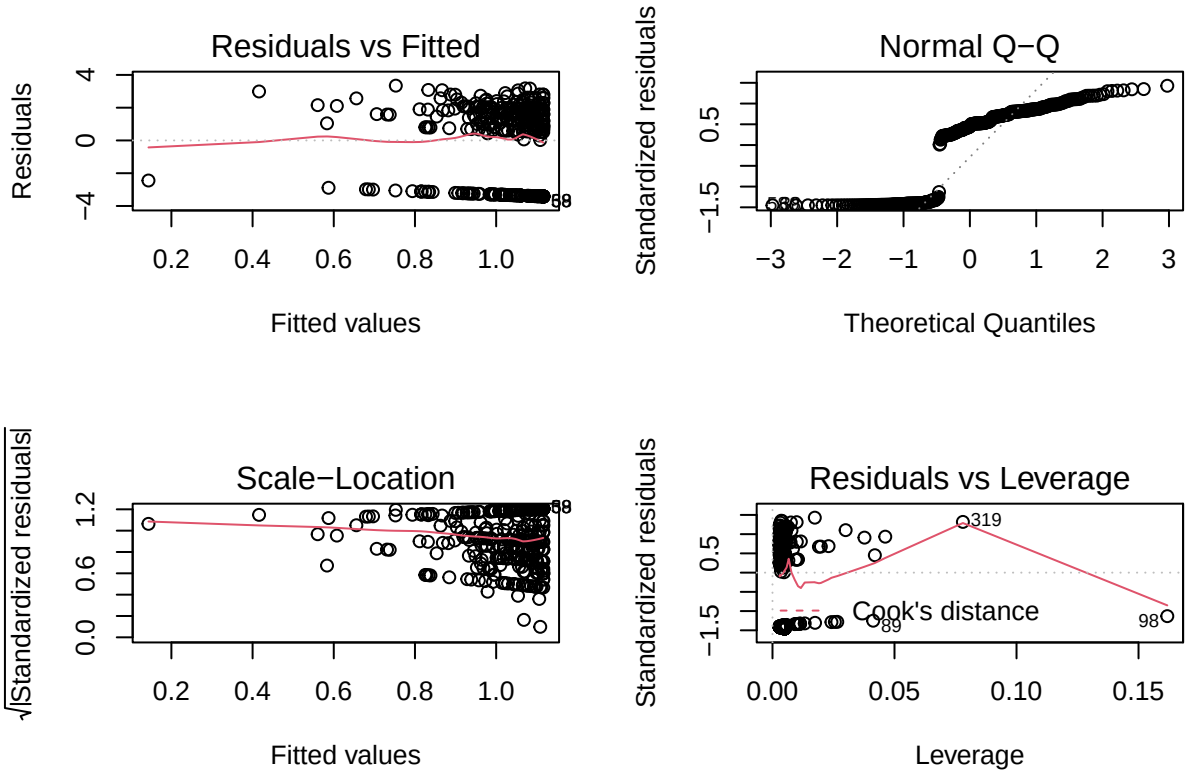
Log-normal

The distribution of weight gain is highly left-skewed. We may correct this by applying a 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.log <- lm(log(pounds_gained+0.1) ~ CalcTMM, data=gained.dat)
summary(SA1.model1.log)
```

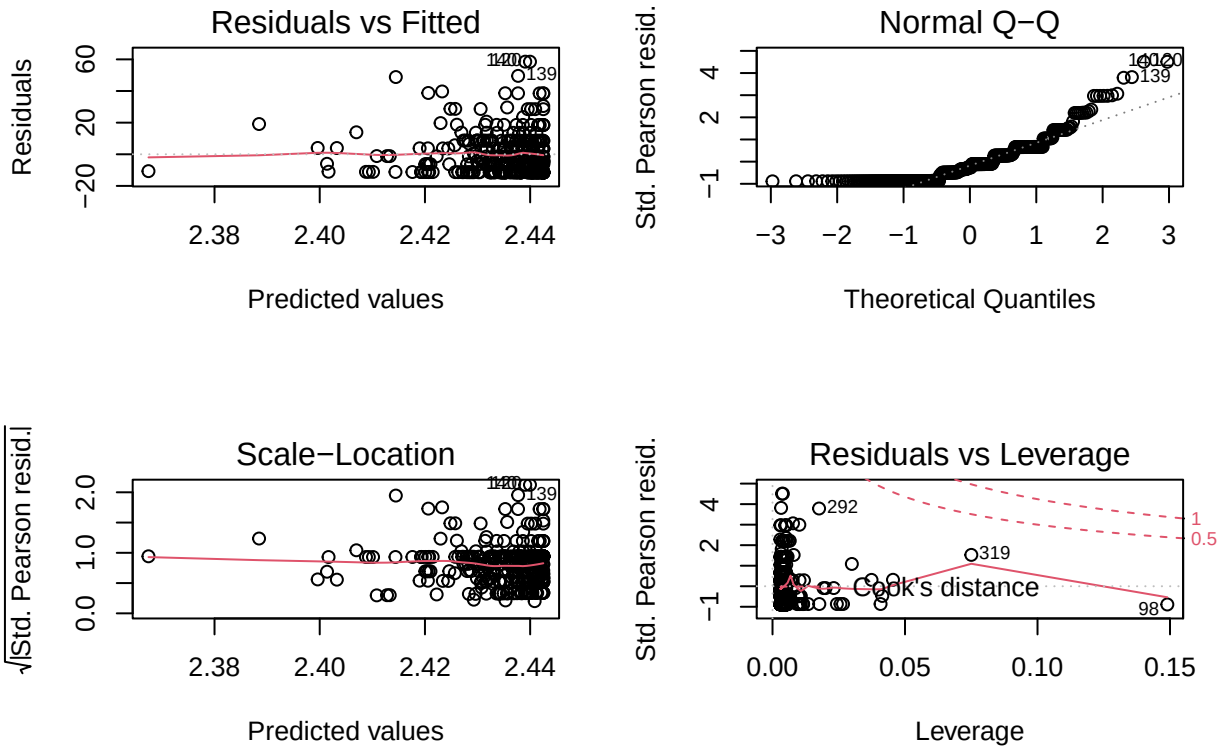
```
##
## Call:
## lm(formula = log(pounds_gained + 0.1) ~ CalcTMM, data = gained.dat)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.419 -3.259  1.046  1.892  3.343
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.116e+00  1.677e-01   6.654 1.15e-10 ***
## CalcTMM     -7.567e-05  8.167e-05  -0.927    0.355
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.361 on 339 degrees of freedom
## Multiple R-squared:  0.002526,   Adjusted R-squared:  -0.0004164
## F-statistic: 0.8585 on 1 and 339 DF,  p-value: 0.3548
```

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



An alternative method, may make it easier to compare models.

```
#error in log(0)
SA1.model1.loggauss <- glm(pounds_gained+0.01 ~ CalcTMM, data=gained.dat,family = gaussian(link="log"))
SA1b.model1.loggauss <- glm(pounds_gained ~ CalcTMM, data=nonzero.dat,family = gaussian(link="log"))
par(mfrow=c(2,2))
plot(SA1.model1.loggauss)
```



Poisson regression

Poisson regression requires integer values. We'll round pounds gained for this.

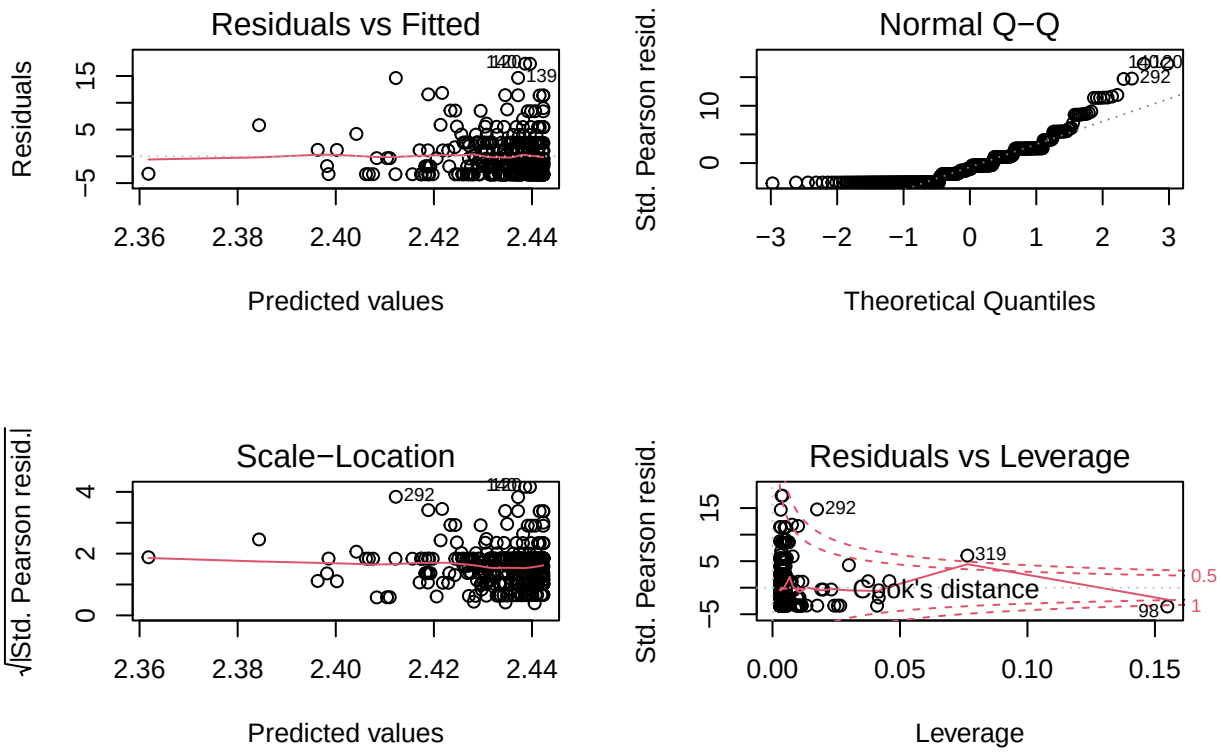
```
gained.dat$LBS <- round(gained.dat$pounds_gained)
nonzero.dat$LBS <- round(nonzero.dat$pounds_gained)
```

```
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)
```

```
##
## Call:
## glm(formula = LBS ~ CalcTMM, family = poisson, data = gained.dat)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -4.796  -4.764  -1.077   1.824  11.675
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.442e+00  2.105e-02 116.024   <2e-16 ***
## CalcTMM     -6.262e-06  1.038e-05  -0.604    0.546
## ---
## 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: 4842.8  on 339  degrees of freedom
```

```
## AIC: 5872.4
##
## Number of Fisher Scoring iterations: 5
```

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



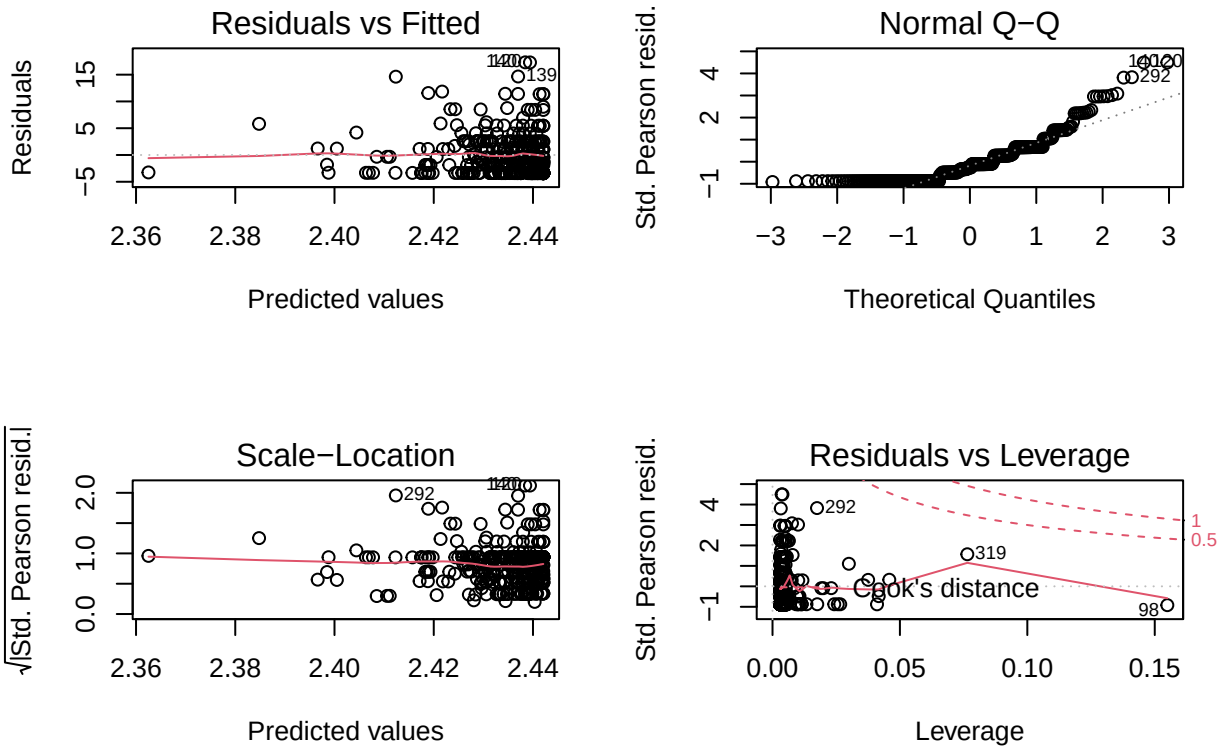
We can, alternatively, fit a quasi-poisson family. This does not require integer values. This also provides a dispersion parameter that may help account for excess 0s.

```
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)
```

```
##
## Call:
## glm(formula = pounds_gained ~ CalcTMM, family = quasipoisson,
##     data = gained.dat)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -4.795  -4.764  -1.076   1.825  11.675
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.442e+00  8.112e-02  30.106   <2e-16 ***
## CalcTMM     -6.192e-06  3.998e-05  -0.155    0.877
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 14.84767)
```

```
##
##     Null deviance: 4843.5  on 340  degrees of freedom
## Residual deviance: 4843.1  on 339  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 5
```

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



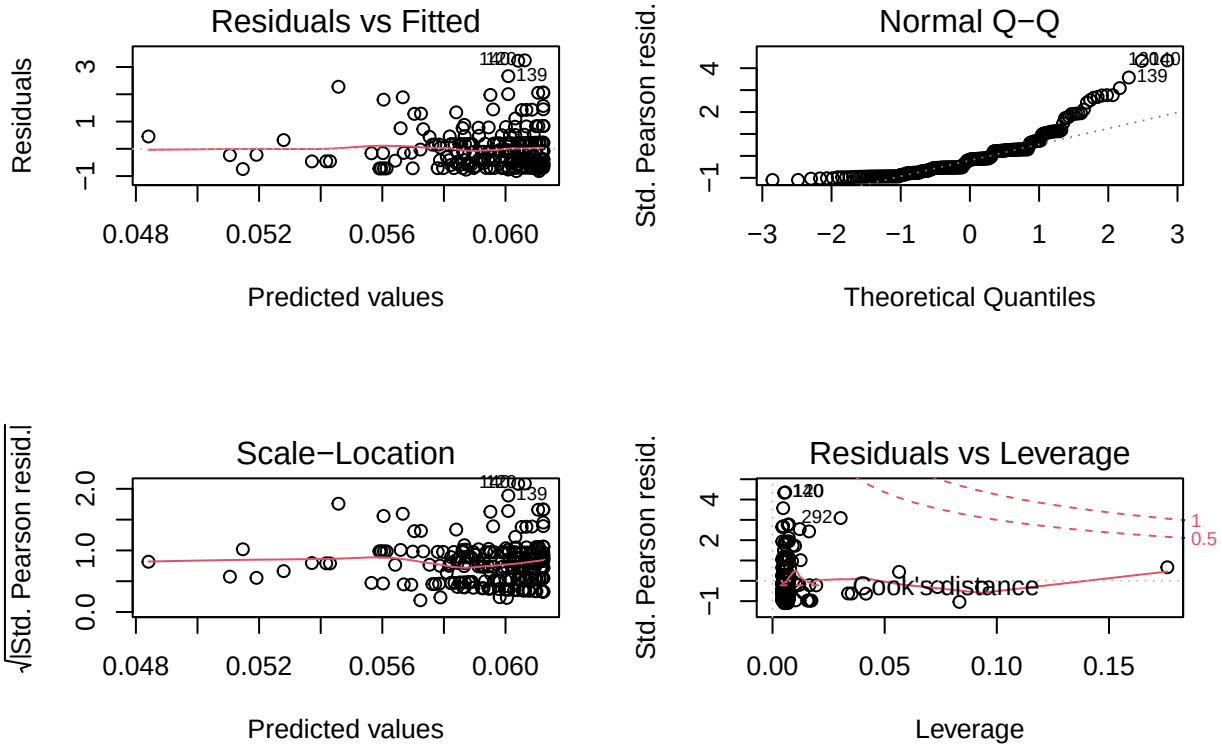
The shape of the distribution of weight gain suggests a gamma distribution. However,

```
SA1b.model1.Gamma <- glm(pounds_gained ~ CalcTMM, data=nonzero.dat,family = Gamma)
summary(SA1b.model1.Gamma)
```

```
##
## Call:
## glm(formula = pounds_gained ~ CalcTMM, family = Gamma, data = nonzero.dat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.3342  -0.6515  -0.1289   0.1995   1.8965
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.123e-02  3.902e-03  15.690   <2e-16 ***
## CalcTMM     -1.385e-06  1.890e-06  -0.733    0.464
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Gamma family taken to be 0.5586371)
```

```
##
##     Null deviance: 109.58  on 230  degrees of freedom
## Residual deviance: 109.30  on 229  degrees of freedom
## AIC: 1701.6
##
## Number of Fisher Scoring iterations: 6
```

```
par(mfrow=c(2,2))
plot(SA1b.model1.Gamma)
```



Model family comparison

```
anova(SA1.model1.gauss,
      SA1.model1.loggauss,
      SA1.model1.poisson,
      SA1.model1.quasi)
```

```
## Warning in anova.glmlist(c(list(object), dotargs), dispersion = dispersion, :
## models with response 'c("pounds_gained + 0.01", "LBS")' removed because response
## differs from model 1
```

```
## Analysis of Deviance Table
##
## Model 1: pounds_gained ~ CalcTMM
## Model 2: pounds_gained ~ CalcTMM
##   Resid. Df Resid. Dev Df Deviance
## 1       339      57370
## 2       339       4843  0    52527
```

```
anova(SA1b.model1.gauss,
      SA1b.model1.loggauss,
      SA1b.model1.poisson,
      SA1b.model1.quasi,
      SA1b.model1.Gamma)
```

```
## Warning in anova.glmlist(c(list(object), dotargs), dispersion = dispersion, :
## models with response '"LBS"' removed because response differs from model 1
```

```
## Analysis of Deviance Table
##
## Model 1: pounds_gained ~ CalcTMM
## Model 2: pounds_gained ~ CalcTMM
## Model 3: pounds_gained ~ CalcTMM
## Model 4: pounds_gained ~ CalcTMM
##   Resid. Df Resid. Dev Df Deviance
## 1       229      36174
## 2       229      36172  0        2
## 3       229       1810  0    34362
## 4       229        109  0     1700
```

```
library(pscl)
```

```
## 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
```

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

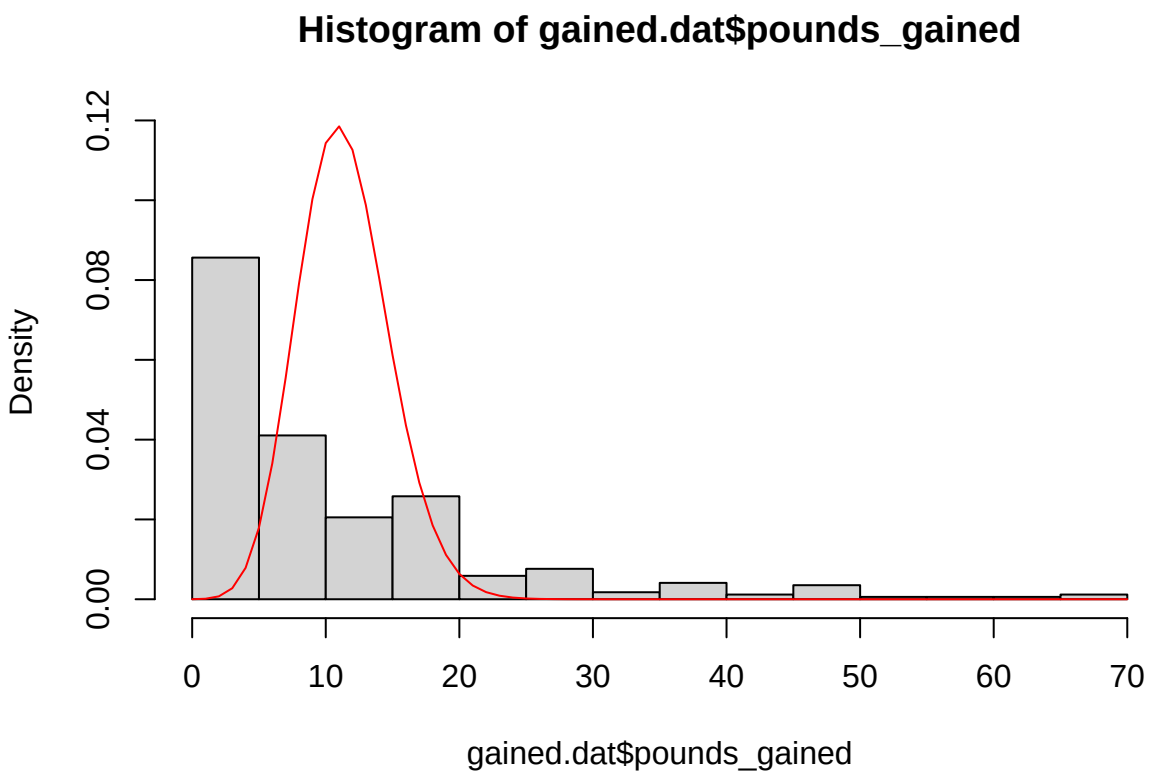
```
## Warning in sqrt(diag(object$vcov)): NaNs produced
```

```
##
## 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
```

```
#par(mfrow=c(1,1))
hist(gained.dat$pounds_gained, freq = FALSE,ylim=c(0,max(dpois(0:70, mean(gained.dat$pounds_gained)))))
100*abs(sd(gained.dat$pounds_gained)-mean(gained.dat$pounds_gained))/mean(gained.dat$pounds_gained)
```

```
## [1] 13.91729
```

```
lines(0:70,dpois(0:70,mean(gained.dat$pounds_gained)),col='red')
```



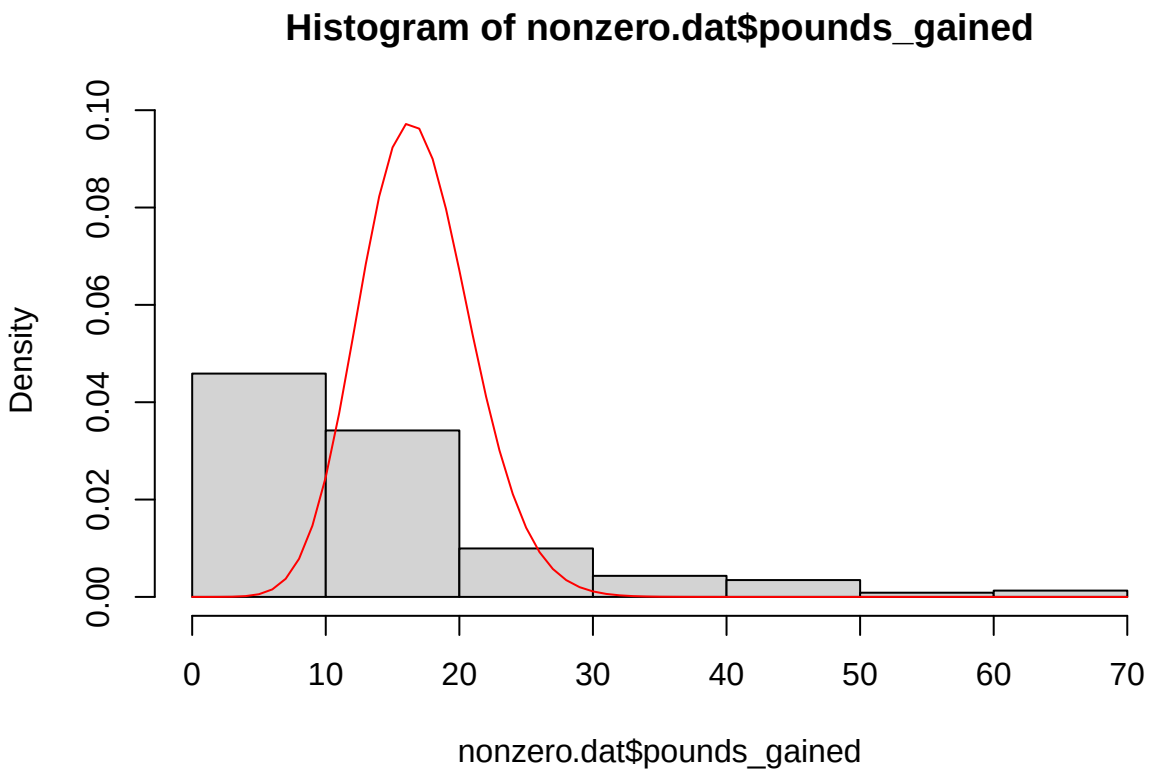

```
mean(gained.dat$pounds_gained)
```

```
## [1] 11.40323
```

```
sd(gained.dat$pounds_gained)
```

```
## [1] 12.99025
```

```
hist(nonzero.dat$pounds_gained, freq = FALSE,ylim=c(0,max(dpois(0:70, mean(nonzero.dat$pounds_gained)))))
lines(0:70,dpois(0:70,mean(nonzero.dat$pounds_gained)),col='red')
```




```
mean(nonzero.dat$pounds_gained)
```

```
## [1] 16.83333
```

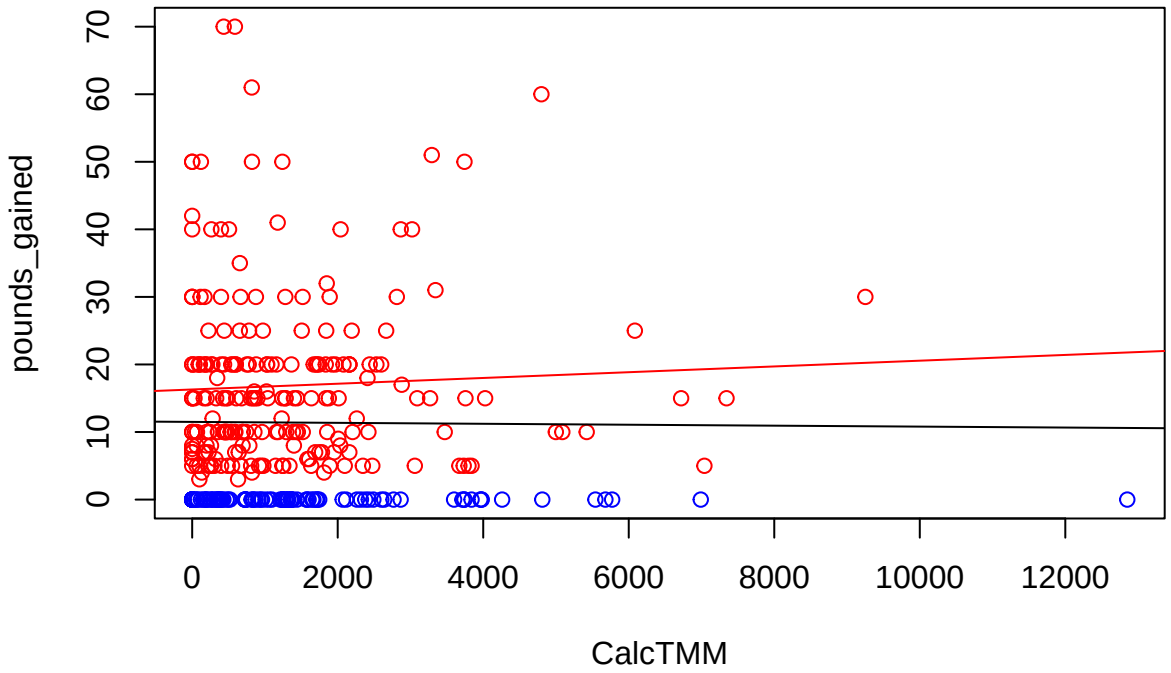
```
sd(nonzero.dat$pounds_gained)
```

```
## [1] 12.55576
```

```
100*abs(sd(nonzero.dat$pounds_gained)-mean(nonzero.dat$pounds_gained))/mean(nonzero.dat$pounds_gained)
```

```
## [1] 25.41133
```

```
colors <- rep('red',dim(gained.dat)[1])
colors[gained.dat$pounds_gained==0] <- 'blue'
plot(pounds_gained~CalcTMM,gained.dat,col=colors)
abline(lm(pounds_gained~CalcTMM,gained.dat))
abline(lm(pounds_gained~CalcTMM,nonzero.dat),col='red')
```




```
#missing values for gender
sum(gained.dat$gender=='')
```

```
## [1] 5
```

```
gained.dat <- gained.dat[!gained.dat$gender=='',]

sum(is.na(gained.dat$Age))
```

```
## [1] 25
```

```
sum(is.na(gained.dat$height))
```

```
## [1] 17
```

```
sum(is.na(gained.dat$shift))
```

```
## [1] 0
```

```
sum(is.na(gained.dat$initial_BMI))
```

```
## [1] 89
```

```
SA12.model4.quasi <- glm(pounds_gained ~ gender + Age + height + shift + Vig.ex.Time  + Mod.ex.time + Wa
SA12b.model4.quasi <- glm(pounds_gained ~ gender + Age + height + shift + Vig.ex.Time  + Mod.ex.time + W
summary(SA12.model4.quasi)
```

```
##
## Call:
## glm(formula = pounds_gained ~ gender + Age + height + shift +
##     Vig.ex.Time + Mod.ex.time + Walk.ex.Time, family = quasipoisson,
##     data = gained.dat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -5.8262  -3.9908  -0.7968   1.5235  10.2409
```

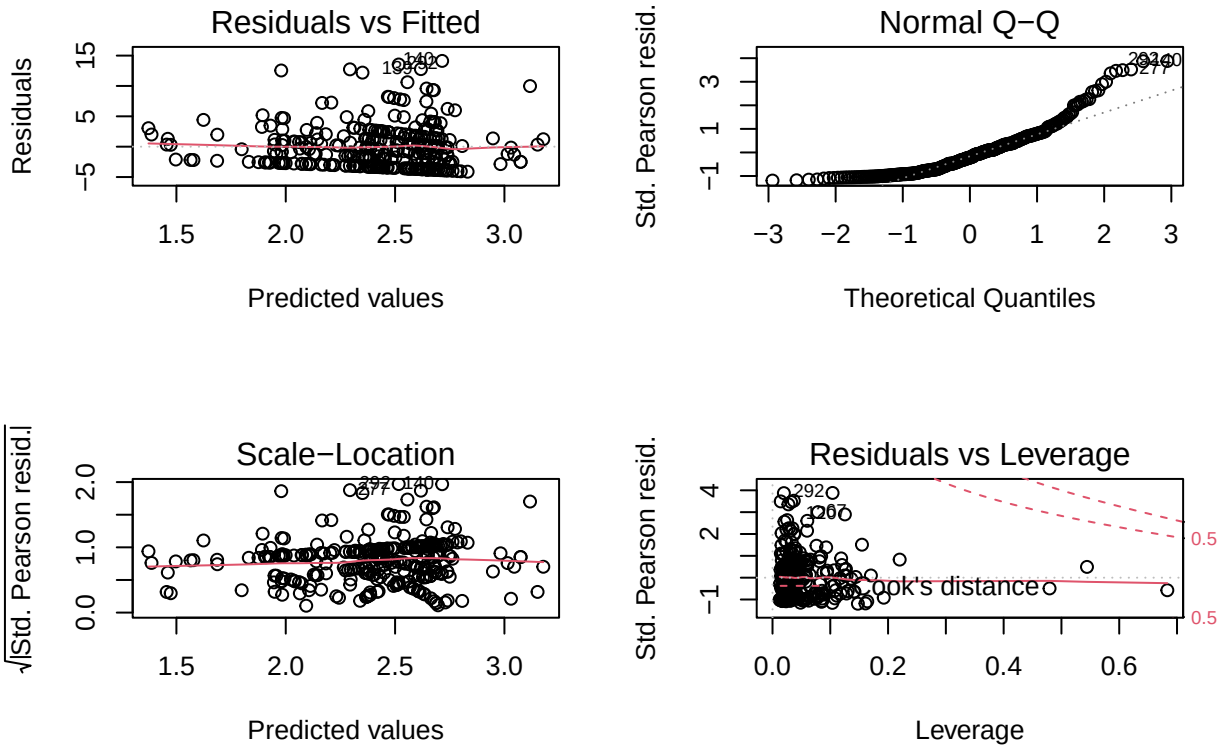
```
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.997e+00  1.423e+00   1.403   0.1617
## genderMale   -1.361e-01  1.926e-01  -0.707   0.4804
## Age          -4.010e-03  6.488e-03  -0.618   0.5370
## height        1.219e-02  2.106e-02   0.579   0.5630
## shift8am      2.228e-02  2.320e-01   0.096   0.9235
## shift9am     -2.681e-01  2.656e-01  -1.009   0.3136
## shift10am    -2.353e-01  2.731e-01  -0.862   0.3896
## shift11am    -5.791e-01  3.045e-01  -1.902   0.0582 .
## shift12pm    -6.464e-01  4.392e-01  -1.472   0.1421
## shift1pm     -1.140e+00  6.580e-01  -1.733   0.0841 .
## shift2pm      3.968e-01  3.503e-01   1.133   0.2582
## shiftother    7.590e-02  3.435e-01   0.221   0.8253
## shiftmissing  5.668e-01  5.897e-01   0.961   0.3373
## Vig.ex.Time  -7.877e-04  6.256e-04  -1.259   0.2090
## Mod.ex.time  -3.765e-05  4.212e-04  -0.089   0.9288
## Walk.ex.Time  1.799e-04  2.860e-04   0.629   0.5300
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 13.56852)
##
##     Null deviance: 4266.8  on 304  degrees of freedom
## Residual deviance: 3966.5  on 289  degrees of freedom
##   (31 observations deleted due to missingness)
## AIC: NA
##
## Number of Fisher Scoring iterations: 5
```

```
summary(SA12b.model4.quasi)
```

```
##
## Call:
## glm(formula = pounds_gained ~ gender + Age + height + shift +
##     Vig.ex.Time + Mod.ex.time + Walk.ex.Time, family = quasipoisson,
##     data = nonzero.dat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -4.5213  -2.2259  -0.5795   0.9458   8.7157
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.365e+00  1.784e+00   0.765   0.4450
## genderFemale  1.167e+00  1.337e+00   0.873   0.3838
## genderMale    1.257e+00  1.345e+00   0.934   0.3513
## Age          -2.322e-03  4.957e-03  -0.468   0.6400
## height        7.164e-03  1.775e-02   0.404   0.6869
## shift8am      3.490e-02  1.852e-01   0.188   0.8508
## shift9am     -1.810e-01  2.167e-01  -0.835   0.4046
## shift10am    -3.269e-01  2.158e-01  -1.515   0.1314
## shift11am    -5.329e-01  2.405e-01  -2.215   0.0279 *
## shift12pm    -5.345e-01  3.484e-01  -1.534   0.1266
```

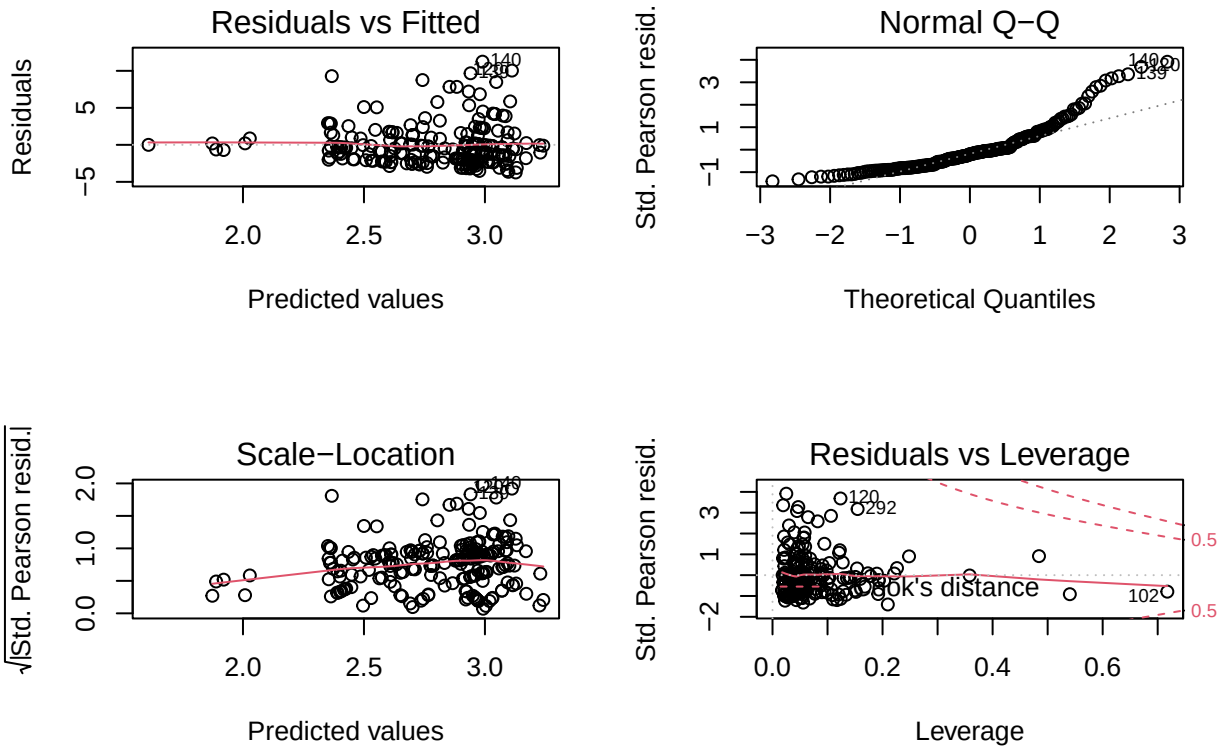
```
## shift1pm     -1.015e+00  5.190e-01  -1.956   0.0519 .
## shift2pm      1.534e-01  2.803e-01   0.547   0.5848
## shiftother   -1.222e-02  2.731e-01  -0.045   0.9644
## shiftmissing  1.540e-01  4.661e-01   0.330   0.7414
## Vig.ex.Time  -2.909e-04  5.325e-04  -0.546   0.5855
## Mod.ex.time  -3.893e-05  3.479e-04  -0.112   0.9110
## Walk.ex.Time  2.865e-04  2.303e-04   1.244   0.2149
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 8.460038)
##
##     Null deviance: 1644.3  on 211  degrees of freedom
## Residual deviance: 1419.4  on 195  degrees of freedom
##   (19 observations deleted due to missingness)
## AIC: NA
##
## Number of Fisher Scoring iterations: 5
```

```
par(mfrow=c(2,2))
plot(SA12.model4.quasi)
```




```
par(mfrow=c(2,2))
plot(SA12b.model4.quasi)
```

```
## Warning: not plotting observations with leverage one:
##   211
```




```
zero.model4 <- zeroinfl(LBS ~ gender + Age + height + shift + Vig.ex.Time + Mod.ex.time + Walk.ex.Time +
summary(zero.model4)
```

```
##
## Call:
## zeroinfl(formula = LBS ~ gender + Age + height + shift + Vig.ex.Time +
##     Mod.ex.time + Walk.ex.Time + initial_BMI, data = gained.dat)
##
## Pearson residuals:
##     Min      1Q  Median      3Q     Max
## -3.1316 -1.1042 -0.3269  0.6359  8.7833
##
## Count model coefficients (poisson with log link):
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)    2.310e+00  4.613e-01   5.007 5.53e-07 ***
## genderMale     1.250e-01  5.941e-02   2.103 0.035433 *
## Age           -6.671e-03  1.996e-03  -3.342 0.000833 ***
## height         5.499e-03  6.793e-03   0.810 0.418180
## shift8am       7.081e-02  7.100e-02   0.997 0.318546
## shift9am      -7.748e-02  8.296e-02  -0.934 0.350334
## shift10am     -3.156e-01  8.144e-02  -3.876 0.000106 ***
## shift11am     -6.429e-01  9.783e-02  -6.572 4.97e-11 ***
## shift12pm     -5.664e-01  1.238e-01  -4.577 4.73e-06 ***
## shift1pm      -9.416e-01  1.939e-01  -4.857 1.19e-06 ***
## shift2pm       2.807e-01  1.135e-01   2.474 0.013373 *
## shiftother     1.088e-01  1.001e-01   1.086 0.277324
## shiftmissing   5.939e-02  1.637e-01   0.363 0.716857
## Vig.ex.Time   -3.060e-04  1.940e-04  -1.578 0.114679
## Mod.ex.time   -7.765e-05  1.100e-04  -0.706 0.480137
## Walk.ex.Time   2.890e-04  7.583e-05   3.812 0.000138 ***
```

```
## initial_BMI   1.769e-02  3.648e-03   4.849 1.24e-06 ***
##
## Zero-inflation model coefficients (binomial with logit link):
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -7.770e-01  3.433e+00  -0.226   0.8209
## genderMale    6.207e-01  4.201e-01   1.478   0.1395
## Age          -1.485e-02  1.558e-02  -0.953   0.3406
## height       -1.887e-02  4.969e-02  -0.380   0.7041
## shift8am      2.712e-01  5.439e-01   0.499   0.6180
## shift9am     -1.910e-01  6.342e-01  -0.301   0.7632
## shift10am    -6.465e-01  6.564e-01  -0.985   0.3246
## shift11am     2.726e-01  6.273e-01   0.435   0.6638
## shift12pm    -3.460e-01  8.457e-01  -0.409   0.6825
## shift1pm     -7.094e-02  1.023e+00  -0.069   0.9447
## shift2pm     -1.576e+01  1.582e+03  -0.010   0.9921
## shiftother   -1.496e-01  8.437e-01  -0.177   0.8593
## shiftmissing -1.618e+01  2.767e+03  -0.006   0.9953
## Vig.ex.Time   8.323e-04  1.457e-03   0.571   0.5678
## Mod.ex.time   3.164e-04  1.502e-03   0.211   0.8332
## Walk.ex.Time  3.375e-04  6.878e-04   0.491   0.6236
## initial_BMI   5.452e-02  2.689e-02   2.027   0.0427 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of iterations in BFGS optimization: 1
## Log-likelihood: -1069 on 34 Df
```

Bootstrap the coefficients

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

confint(zero.model1)
```