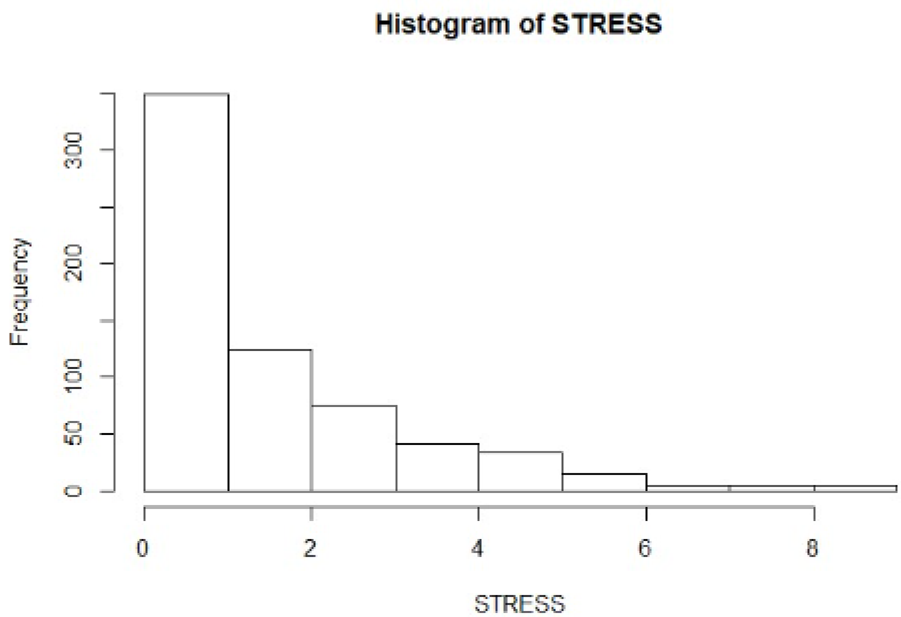
Assignment # 9

May 2021

1)



Min = 0

1st Quartile = 0

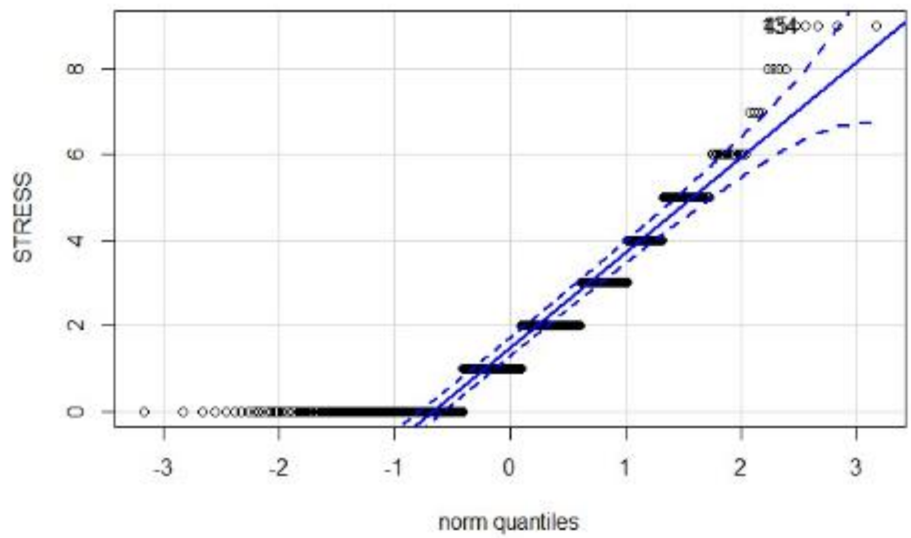
Median = 1

Mean = 1.73

3rd quartile = 3

Max = 9

Variance = 3.419104



After reviewing the results, it seems the stress variable is positively skewed to the right. The count of each stress level decreases as the value increases from 0 to 8. Given the median stress variable is 1, the data looks to be affected by a large number of 0's.

2)

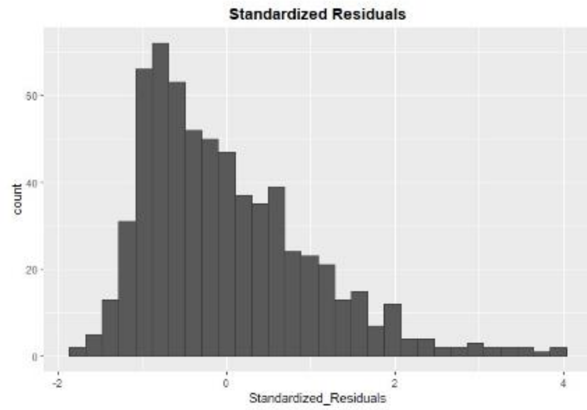
Y = 5.713 -.007X1 -.041X2 -.042X3 - .030X4

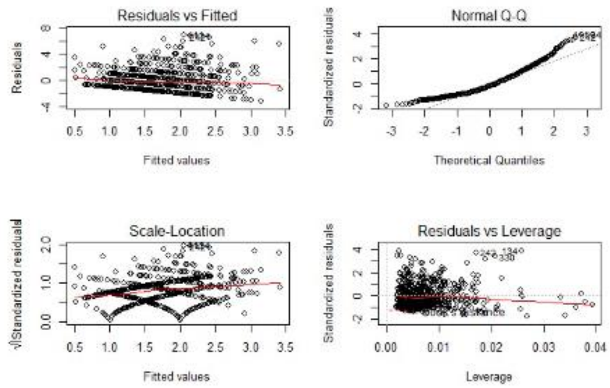
Coefficients Table:

Coefficients:	**Estimate**	**Std. Error**	**T-value**	**Pr(>|t|)**
(Intercept)	5.71281	0.58118	9.830	< 2e-16
COHES	0.00703	9.99793	-3.298	0.00103
ESTEEM	0.04129	0.01933	-2.136	0.03305
GRADES	0.04170	0.02352	-1.773	0.07670
SATTACH	0.03042	0.01412	-2.154	0.03160

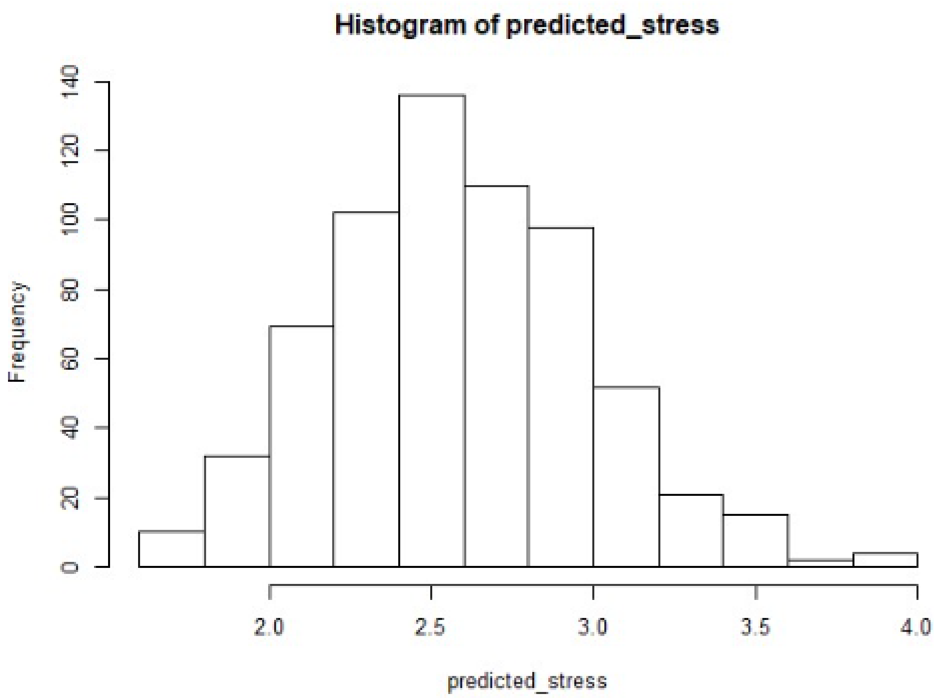
R-squared = .083

F-statistic = 14.65





The results show that this models is not good at predicting stress results. The r-squared value shows that only 8 percent of the variability is explained in the model, a very low percentage. The data in the histograms is skewed to the right and the residuals vs fitted plot shows a high number of outliers.



The results show the log of the dataset produces a normalized distribution of the data. The values in the plot, however, are skewed right which shows that the ols multiple linear regression model is not good for predictive purposes.

3)

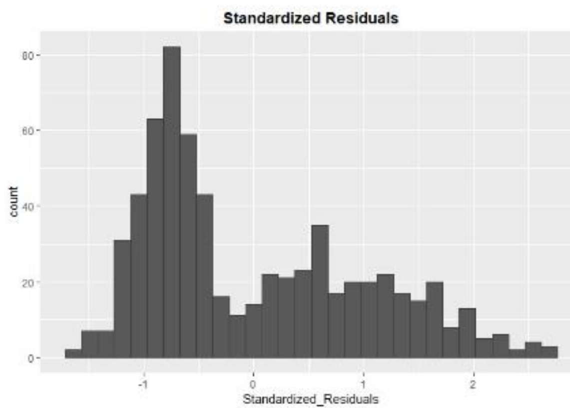
Y = 1.769 - .007X1 - .011X2 -.016X3 - .011X4

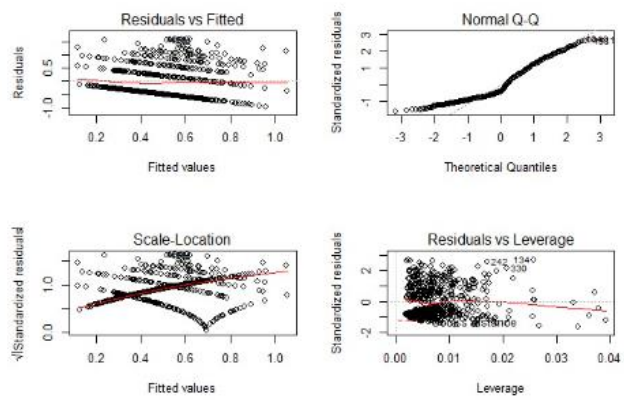
Coefficients Table:

Coefficients:	Estimate	Std. Error	T-value	Pr(>|t|)
(Intercept)	1.769	0.195	9.063	< 2e-16
COHES	-0.007	0.002	-3.037	0.002
ESTEEM	-0.011	0.006	-1.671	0.095
GRADES	-0.016	0.008	-2.004	0.045
SATTACH	-0.011	0.005	-2.325	0.020

R-squared = .077

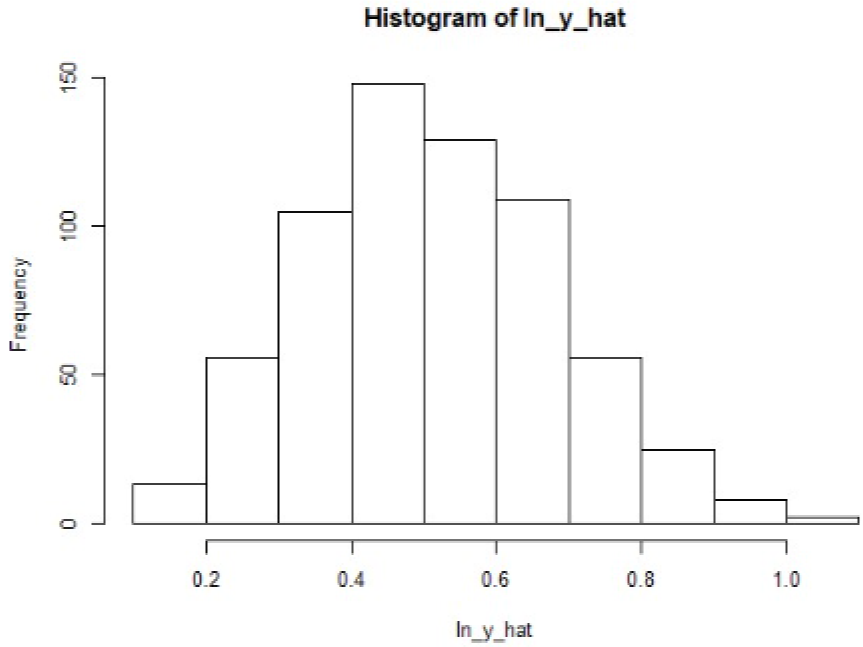
F-statistic = 13.41





The results show this model is also not good a predicting stress levels. The r-squared value shows that only 8 percent of the variability is explained in the model, a very low percentage. The data in the histograms is skewed to the right and the residuals vs fitted plot shows a high number of outliers.

b)



The results show the log of the dataset produces a normalized distribution of the data. The values in the plot, however, are skewed right which shows that the ols multiple linear regression model is not good for predictive purposes.

4)

a)

Y = 2.312 - .006X1 - .019X2 - .015X3 - .08X4

Coefficients Table:

Coefficients:	**Estimate**	**Std. Error**	**z-value**	**Pr(>|z|)**
(Intercept)	2.312	0.24	9.633	<2e-16
COHES	-0.006	0.003	-1.124	0.034
ESTEEM	-0.019	0.008	-2.408	0.01
GRADES	-0.015	0.010	-1.513	0.13
SATTACH	-0.008	0.006	-1.371	0.17

AIC = 1552.1

Residual deviance = 380.26

Each individual x variable equates to a unit of change for log(stress), holding all other variables constant. To get the actual predicated value of stress, we take the exponent of our Poisson model. Comparing the coefficients, it seems this model fits very similarly with model 3.

b)

Y = 2.734 - .0013X1 - .024X2 - .023X3 - .016X4

Coefficients Table:

Coefficients Table:

Coefficients:	Estimate	Std. Error	z-value	Pr(>|z|)
(Intercept)	2.734	0.234	11.683	<2e-16
COHES	-0.013	0.003	-4.466	7.98e-06
ESTEEM	-0.024	0.008	-2.947	0.003
GRADES	-0.023	0.010	-2.379	0.017

AIC = 2417.2

Residual deviance = 1245.4

The results show that the model is very similar to the ZIP model, however, the AIC and residual deviance results are much higher in the over-dispersed models because Poisson regression models do not do a good job of predicting low or negative values.

5)

I categorized the data into three groups:

Low Group: Family cohesion less than 1 SD from the mean

Middle Group: Family cohesion = 1 SD from the mean

High Group: Family cohesion greater than 1 SD from the mean

The results show that only the high group is represented while both the middle and low groups produced no results.

6)

Poisson Regression:

AIC = 1552.03

BIC = 1572.372

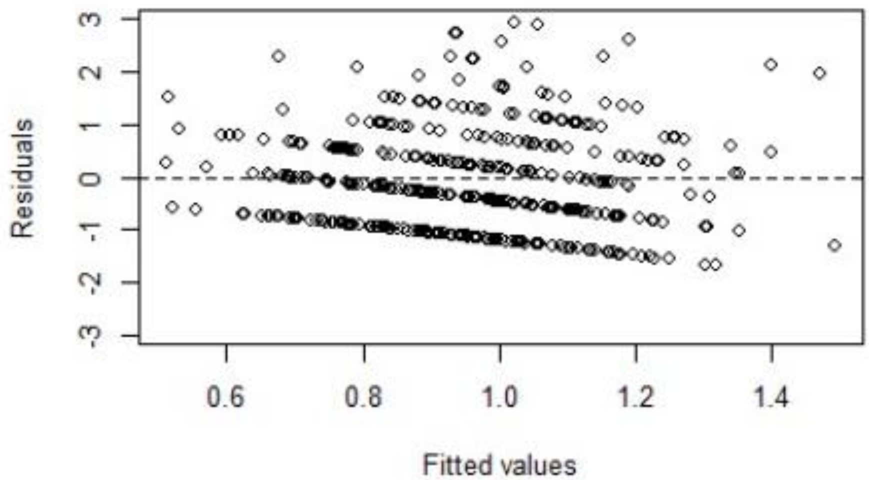
Over-dispersed Poisson Regression:

AIC = 2439.612

BIC = 2417.219

The results show that the Poisson regression model removes zero stress-levels provides a better fit than the over dispersed Poisson Regression model which makes sense as the Poisson model does not do a good job of predicting low values.

7)



The results show there are more positive residual outliers than negative ones, and these negative ones appear to increase as the fitted values increase. There appears to be a greater number of positive residuals and the results are unbalanced when fitted to a line and the model does not seem to eb a great fit for predicting stress levels.

8)

Y = 3.51673 - .021X1 - .019X2 - .025X3 - .028X$

B1 = -.021 change in the log odds ratio for a 1 unit change in X1

For each one unit I change of how an adolescent gets along with their family, stress levels decrease by 2%.

B2 = -.019 change in the log odds ratio for a 1 unit change in X2

For each one unit I change of in level of self-esteem, stress levels decrease by 1.87%.

B3 = -.025 change in the log odds ratio for a 1 unit change in X3

For each one unit I change of in past years school grades, stress levels decrease by 2.52%.

B4 = -.028 change in the log odds ratio for a 1 unit change in X4

For each one unit I change of in how the adolescent likes and is attached to their school, stress levels decrease by 2.73%.

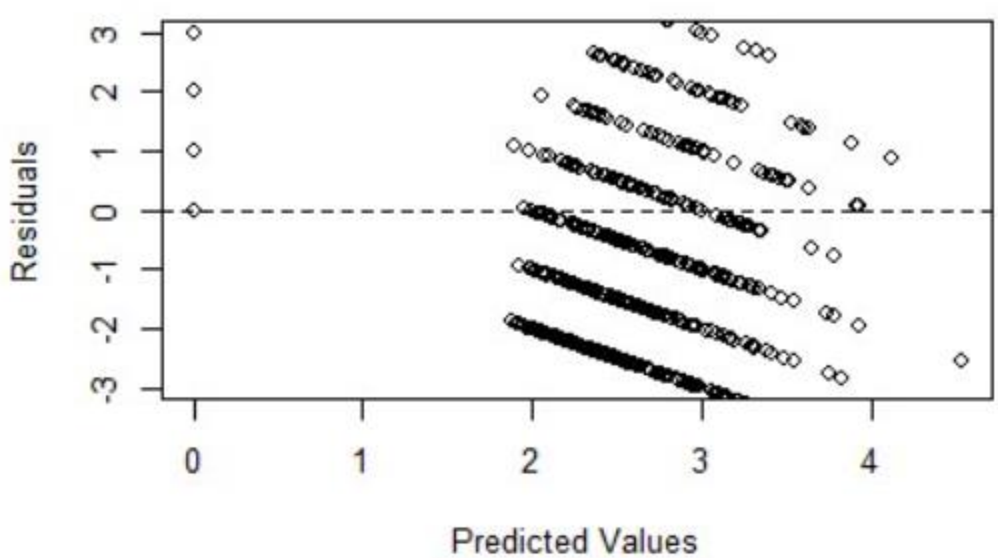
AIC = 821.786

BIC = 844.178

The results based on the AIC and BIC levels show this is the best performing model. It appears the logistical regression model provides a better fit than the ols model, or Poisson distribution model. This is further supported by the residual deviance levels for the poisson distribution models and the r-squared results for these models.

9)

After viewing the results of the models, it is possible to get a better resulting model that combines ensuring that stress levels are present, and determining what stress levels actually are. We fitted a logistical regression model to see if stress is present, we can use a poisson regression model to predict number of stressful events conditioned on the previous regression model. The model appears to be a good fit and could be useful for predictions.



The residuals are more evenly distributed around the predicate values, and the combined fitted model seems to be a better fit than any of the aforementioned models.

10

10)

Model one:

Poisson Coefficients Table:

Coefficients:	Estimate	Std. Error	z-value	Pr(>|z|)
(Intercept)	2.642	0.272	9.7	<2e-16
COHES	-0.008	0.003	-2.418	0.016
ESTEEM	-0.026	0.009	-2.832	0.005
GRADES	-0.019	0.011	-1.792	0.073
SATTACH	-0.01	0.007	-1.571	0.116

Logit Coefficients Table:

Coefficients:	Estimate	Std. Error	z-value	Pr(>|z|)
(Intercept)	-2.835	0.983	-2.884	0.004
COHES	0.019	0.012	1.560	0.119
ESTEEM	-0.004	0.033	-0.132	0.895
GRADES	0.014	0.038	0.380	0.704
SATTACH	0.025	0.024	1.031	0.302

AIC = 2288.802

BIC = 2333.587

Model two:

Coefficients Table:

Coefficients:	Estimate	Std. Error	z-value	Pr(>|z|)
(Intercept)	2.427	0.258	9.421	<2e-16
COHES	-0.013	0.003	-4.101	4.12e-05
ESTEEM	-0.03	0.009	-3.335	0.000852

Logit Table:

Coefficients:	Estimate	Std. Error	z-value	Pr(>|z|)
(Intercept)	-2.152	0.908	-2.371	0.018
ESTEEM	0.033	0.029	1.162	0.245

AIC = 2293.78

BIC = 2316.73

The results show that the combined model provides the best results and is definitely worth exploring more and gathering more summary statistics to determine if it provides a good fit for stress levels. The logistical model with the y indicator has the lowest aic and bic levels and should be explored further. However, since none of the models provided a great fit, additional variables should be tested to see if they provide greater fits.