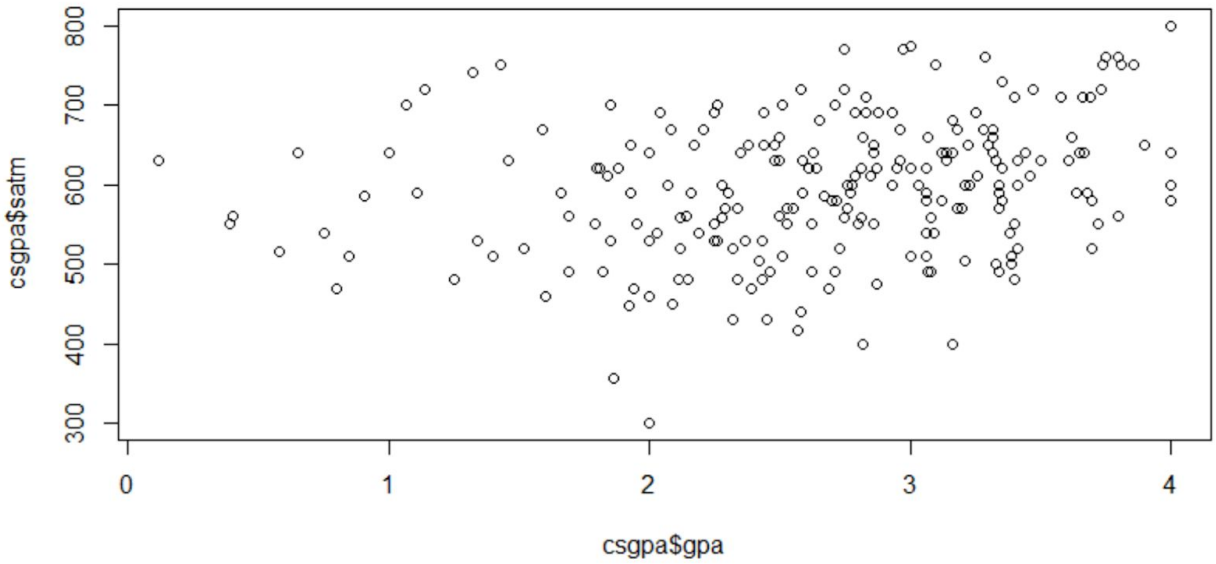
“I pledge my honor that I have abided by the Stevens Honor System” - Himanshu Rana



1) By the scatterplot, as the SAT math scores increase so does the gpa of the first year. I choose to make the x-axis the gpa and the SAT scores the y-axis because it better displays the trend of higher SAT scores translates to higher gpa.

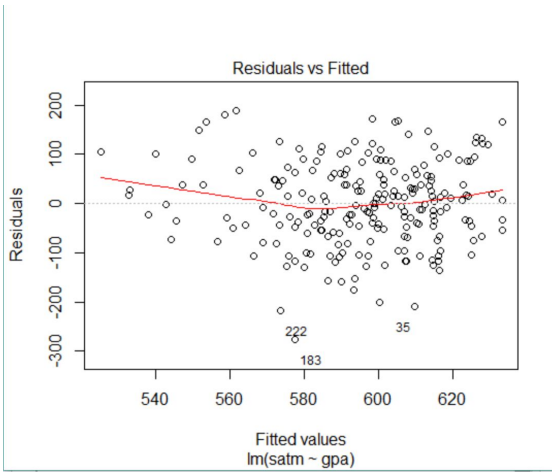
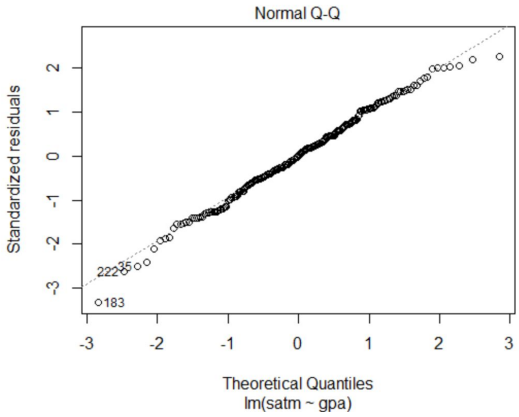
2) $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

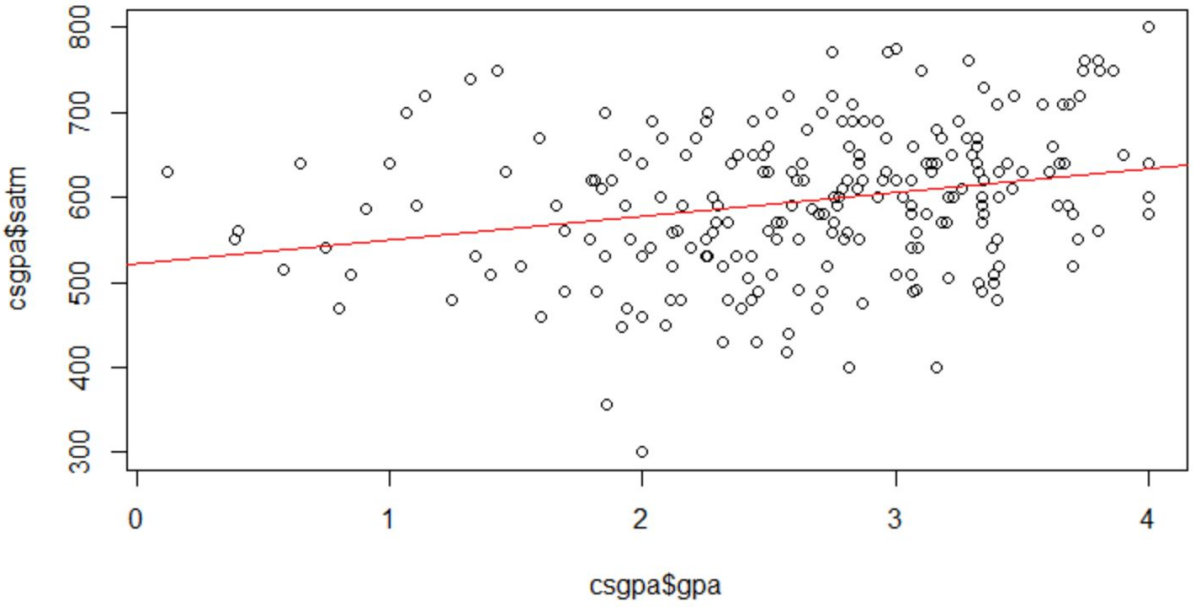
Some assumptions we can make is that the error is i.i.d. ~ N(0, σ^2) which implies that Yi | xi is independent to N(β0 + β1xi , σ^2).

```
> summary(lmfit)

Call:
lm(formula = satm ~ gpa, data = csgpa)

Residuals:
     Min       1Q   Median       3Q      Max
-277.560  -51.003    0.953   59.162  188.345

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  521.752     19.784  26.372  < 2e-16 ***
gpa           27.904      7.201   3.875  0.00014 ***
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 83.81 on 222 degrees of freedom
Multiple R-squared:  0.06336,   Adjusted R-squared:  0.0591
4
F-statistic: 15.02 on 1 and 222 DF,  p-value: 0.0001402
```

3) $\beta_0 = 521.75$ $\beta_1 = 27.9$ The positive sign of the slope indicates that the relationship between the two variables increases as the other increases,

4)

homoscedasticity	normality
	



5)

$R^2 = 0.063$ This indicates a good fit because it is close to zero which means there is less error between the points and the regression line. Also, that the relationship is almost linear.

6) Step 1 $H_0: \beta_1 = 0$ Step 2: $\alpha = .05$
$H_1: \beta_1 \neq 0$

Step 3: $T = \frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)} = \frac{27.1}{2.68} = 10.41$

$t_{obs} = 1.969$

Step 4: $P(10.41 > k_\alpha) = \frac{.05}{2}$

Step 5: $|1.969| > k_\alpha \rightarrow$ not true, do not Reject H_0
GPA and SAT math scores are not linearly related at $\alpha = 5\%$.

7) Test conclusion does not seem consistent. Out of the 244 CS students surveyed only 224 are reported. Its too large of a sample size and that means there is more room for error.

6)
7)