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

1)

```
~/
> csgpa <- read.table("csgpa.txt", header = TRUE)
> lmfit <- lm(satm ~ gpa, csgpa)
> summary(lmfit)$r.square
[1] 0.06336008
>
> ssm = sum((fitted(lmfit) - mean(csgpa$satm))^2)
> sse = sum((fitted(lmfit) - csgpa$satm)^2)
> sst = ssm + sse
> 1 - (sse/(ssm + sse))
[1] 0.06336008
>
```

```
Values
  sse    1559263.54037206
  ssm    105478.173913651
  sst    1664741.71428571

Files  Plots  Packages  Help  Viewer
Zoom  Export
```

hellooooo

2)

```
> res.aov <- aov(satm ~ gpa, csgpa)
> summary.aov(res.aov)
              Df  Sum Sq Mean Sq F value  Pr(>F)    
gpa            1  105478  105478   15.02 0.00014 ***
Residuals    222 1559264    7024                    
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
```

Under the null hypothesis of beta1 = 0, the ratio F = MSM/MSE follows F-distribution of (1, n - 2) degrees of freedom. Its called the ANOVA and for the SLR it is equivalent to the t-test on beta1. We conclude that we do not reject the null hypothesis

3) $PI_{1-\alpha}(y, x^*) = \left[\hat{y}^* \pm t_{1-\frac{\alpha}{2}} \sqrt{\widehat{Var}[\hat{y}^*]}\right]$ $\alpha = .05$

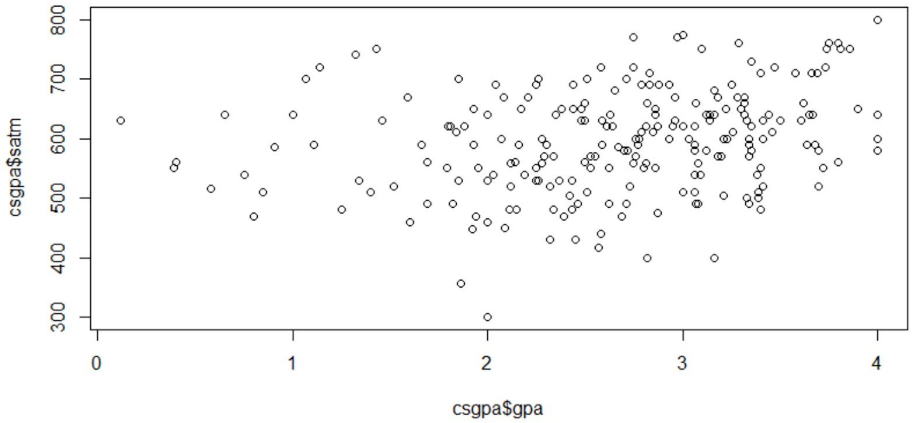
$$\widehat{Var}[\hat{y}^*] = \hat{\sigma}^2 \left(1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}\right)$$

$$\hat{\sigma}^2 = \frac{sse}{(n-2)} = \frac{sse}{222} = 7023.71$$

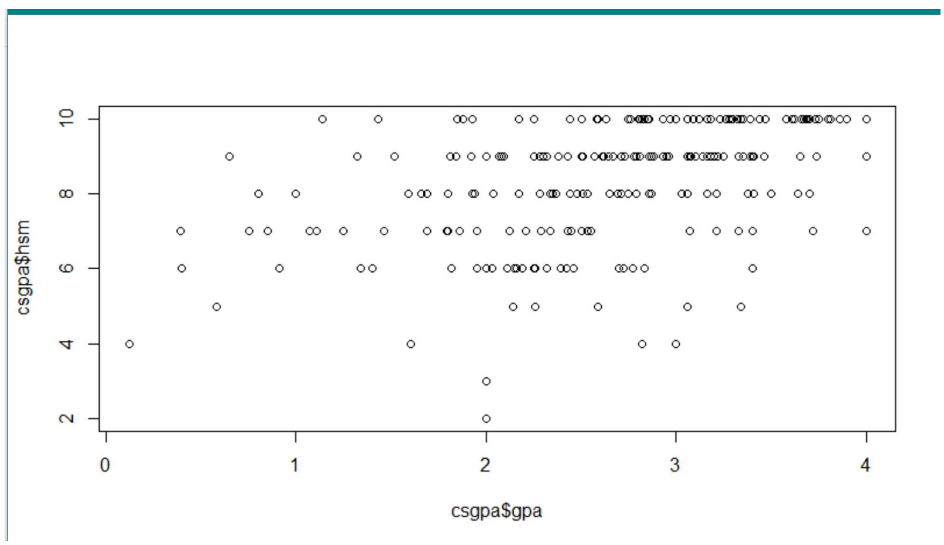
$$7023.71\left(1 + \frac{1}{224} + \frac{(x^* - \bar{x})^2}{135.46}\right)$$

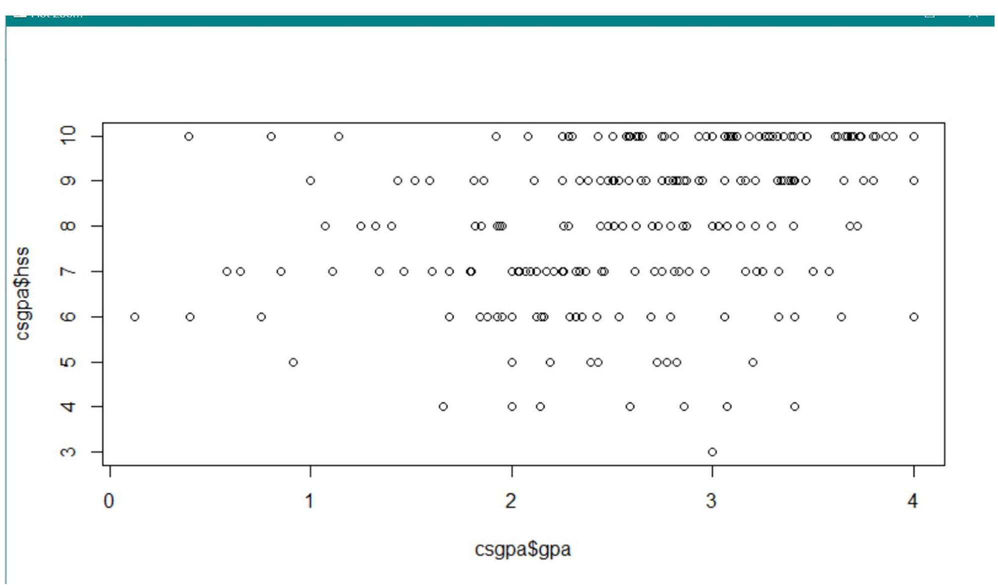
$$\left(\hat{y}^* \pm 1.969\sqrt{7023.7\left(1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{135.46}\right)}\right)$$

3)



4)





One covariate does seem to have a stronger linear effect on the gpa than the rest.

5) $\widehat{y_i} = \widehat{\beta_0} + \sum_{j=1}^{p} \widehat{\beta_j} x_{i,j} + \varepsilon_i$ -> some relevant assumptions are that the x i,j are not random(fixed), beta0 is interpreted as the slope, beta j is slope corresponding to the j-th covariate, and the error term has the same assumption as in SLR.

```
Call:
lm(formula = csgpa$gpa ~ csgpa$satm + csgpa$hsm + csgpa$hss)

Residuals:
     Min       1Q   Median       3Q      Max
-2.10954 -0.37657  0.08842  0.45121  1.68691

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.4843393  0.3601535   1.345    0.180
csgpa$satm  0.0006383  0.0006092   1.048    0.296
csgpa$hsm   0.1597422  0.0381288   4.190 4.05e-05 ***
csgpa$hss   0.0545926  0.0337547   1.617    0.107
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7003 on 220 degrees of freedom
Multiple R-squared:  0.2036,	Adjusted R-squared:  0.1928
F-statistic: 18.75 on 3 and 220 DF,  p-value: 7.222e-11
```

6)

```
 satmr <- lm(satm ~ gpa, csgpa)
 summary(satmr)$r.square
1] 0.06336008
 hsmr <- lm(hsm ~ gpa, csgpa)
 summary(hsmr)$r.square
1] 0.1905312
 hssr <- lm(hss ~ gpa, csgpa)
 summary(hssr)$r.square
1] 0.1085211
```

Each of the covariate's R^2 is less than 0.8, therefore multicollinearity is intact. No, all three covariates cannot be kept in the same model as it is, because it does not properly display the true effects the student's gpa has on each of the categories.