Appendix

REGRESSION ANALYSIS

Loading the original data:

```
project_data = read.csv("reg_educ_race_sex.csv")
```

Omitting null values:

```
nrow(project_data)
```

```
## [1] 35
```

```
project_data = na.omit(project_data)
nrow(project_data)
```

```
## [1] 35
```

```
project_data
```

```
##                        education_text                 race_text sexs_text
## 1  High School graduates, no college                     White       Men
## 2       Bachelor's degree and higher                     White       Men
## 3             Bachelor's degree only                     White       Men
## 4                    Advanced degree                     White       Men
## 5    Less than a High School diploma Black or African American       Men
## 6    Less than a High School diploma                     White     Women
## 7  High School graduates, no college                     White     Women
## 8       Bachelor's degree and higher                     White     Women
## 9             Bachelor's degree only                     White     Women
## 10                   Advanced degree                     White     Women
## 11 High School graduates, no college Black or African American       Men
## 12      Bachelor's degree and higher Black or African American       Men
## 13            Bachelor's degree only Black or African American       Men
## 14                   Advanced degree Black or African American       Men
## 15   Less than a High School diploma Black or African American     Women
## 16 High School graduates, no college Black or African American     Women
## 17      Bachelor's degree and higher Black or African American     Women
## 18            Bachelor's degree only Black or African American     Women
## 19                   Advanced degree Black or African American     Women
## 20   Less than a High School diploma                     Asian       Men
## 21 High School graduates, no college                     Asian       Men
## 22  Some college or associate degree                     Asian       Men
## 23      Bachelor's degree and higher                     Asian       Men
## 24            Bachelor's degree only                     Asian       Men
## 25                   Advanced degree                     Asian       Men
## 26   Less than a High School diploma                     Asian     Women
## 27 High School graduates, no college                     Asian     Women
## 28  Some college or associate degree                     Asian     Women
## 29      Bachelor's degree and higher                     Asian     Women
## 30            Bachelor's degree only                     Asian     Women
## 31                   Advanced degree                     Asian     Women
```

```
## 32  Some college or associate degree                     White       Men
## 33  Some college or associate degree Black or African American       Men
## 34  Some college or associate degree                     White     Women
## 35  Some college or associate degree Black or African American     Women
##     avg_val
## 1   736.550
## 2  1326.450
## 3  1223.500
## 4  1542.600
## 5   454.750
## 6   382.950
## 7   549.300
## 8   978.250
## 9   901.000
## 10 1120.800
## 11  580.750
## 12  978.100
## 13  904.550
## 14 1179.850
## 15  378.250
## 16  481.900
## 17  880.800
## 18  818.100
## 19 1011.350
## 20  535.375
## 21  687.000
## 22  840.875
## 23 1577.125
## 24 1374.375
## 25 1800.500
## 26  446.625
## 27  568.500
## 28  703.750
## 29 1210.625
## 30 1091.125
## 31 1407.125
## 32  861.300
## 33  669.300
## 34  641.700
## 35  567.750
```

```
attach(project_data)
```

```
high_nocol = as.numeric(education_text == 'High School graduates, no college')
bachandhigh = as.numeric(education_text == "Bachelor's degree and higher")
bachonly = as.numeric(education_text == "Bachelor's degree only")
adv = as.numeric(education_text == 'Advanced degree')
lessthanhigh = as.numeric(education_text == 'Less than a High School diploma')
white =  as.numeric(race_text == 'White')
black = as.numeric(race_text == 'Black or African American')
asian = as.numeric(race_text == 'Asian')
men = as.numeric(sexs_text == 'Men')
```

Creation of Model:

```
model = lm(avg_val ~ + high_nocol + bachandhigh
           + bachonly + adv + lessthanhigh
           + white + black + men)

summary(model)
```

```
##
## Call:
## lm(formula = avg_val ~ +high_nocol + bachandhigh + bachonly +
##     adv + lessthanhigh + white + black + men)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -162.57  -72.79   15.68   43.17  222.10
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    742.45      55.40  13.402 3.48e-13 ***
## high_nocol    -113.45      63.45  -1.788 0.085428 .
## bachandhigh    444.45      63.45   7.005 1.94e-07 ***
## bachonly       338.00      63.45   5.327 1.42e-05 ***
## adv            629.59      63.45   9.923 2.49e-10 ***
## lessthanhigh  -250.87      66.88  -3.751 0.000892 ***
## white         -116.42      46.09  -2.526 0.017973 *
## black         -278.13      44.86  -6.199 1.47e-06 ***
## men            206.37      37.30   5.532 8.30e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 109.9 on 26 degrees of freedom
## Multiple R-squared:  0.9328, Adjusted R-squared:  0.9122
## F-statistic: 45.13 on 8 and 26 DF,  p-value: 2.616e-13
```

```
anova(model)
```

```
## Analysis of Variance Table
##
## Response: avg_val
##              Df  Sum Sq Mean Sq F value    Pr(>F)
## high_nocol    1  638081  638081  52.837 1.016e-07 ***
## bachandhigh   1  301498  301498  24.966 3.391e-05 ***
## bachonly      1  171287  171287  14.184 0.0008578 ***
## adv           1 2209374 2209374 182.951 2.825e-13 ***
## lessthanhigh  1  205534  205534  17.020 0.0003368 ***
## white         1    1002    1002   0.083 0.7755584
## black         1  464135  464135  38.433 1.475e-06 ***
## men           1  369591  369591  30.605 8.299e-06 ***
## Residuals    26  313985   12076
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```