HW6

Anya Conti

March 5, 2017

1. The simple linear model of sat ~ expend is shown below. The coefficient for expenditure is -20.892. This implies that for every increase in education expenditure by $1000, there is a 20.892 decrease in SAT score, which is approximately the 21 point decrease in SAT score the essay claimed.

```
##
## Call:
## lm(formula = sat ~ expend, data = SAT)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -145.074  -46.821    4.087   40.034  128.489
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1089.294     44.390  24.539  < 2e-16 ***
## expend       -20.892      7.328  -2.851  0.00641 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 69.91 on 48 degrees of freedom
## Multiple R-squared:  0.1448, Adjusted R-squared:  0.127
## F-statistic: 8.128 on 1 and 48 DF,  p-value: 0.006408
```

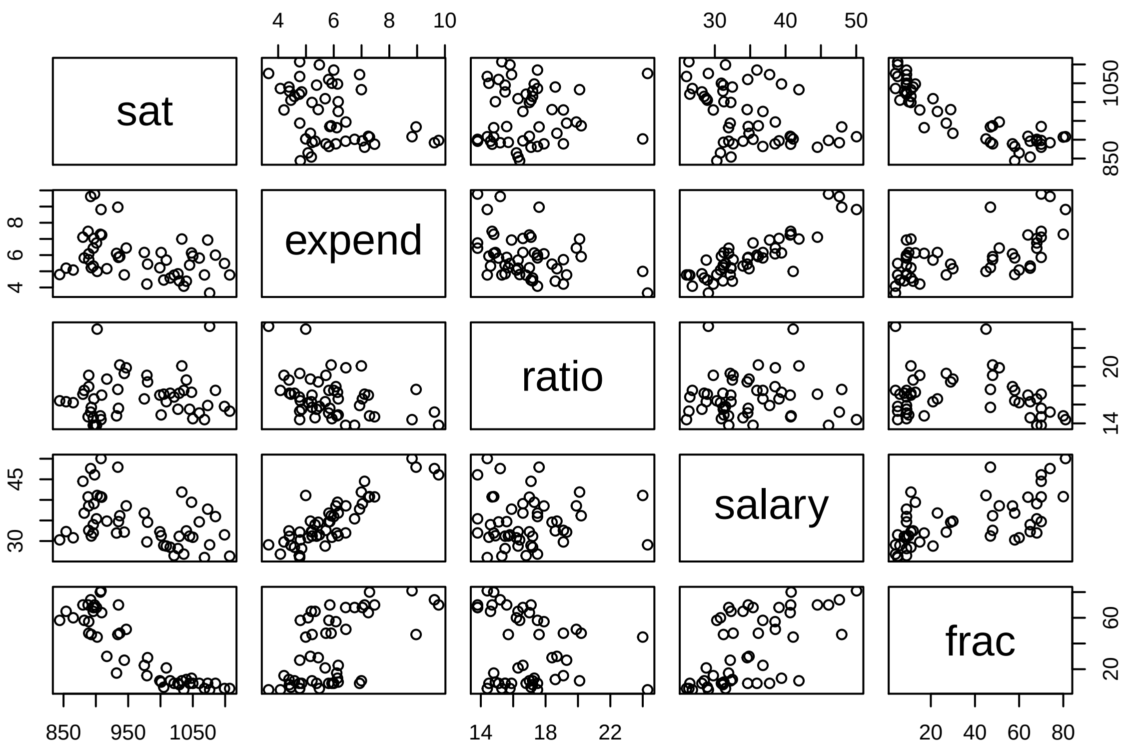
2. The simple linear model of sat ~ salary is shown below. The coefficient for salary is -5.540. This implies that for every increase in average annual salary of public school teachers by $1000, there is a 5.54 decrease in SAT score, which is negative just as the essay claimed.

```
##
## Call:
## lm(formula = sat ~ salary, data = SAT)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -147.125  -45.354    4.073   42.193  125.279
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1158.859     57.659  20.098  < 2e-16 ***
## salary        -5.540      1.632  -3.394  0.00139 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 67.89 on 48 degrees of freedom
## Multiple R-squared:  0.1935, Adjusted R-squared:  0.1767
## F-statistic: 11.52 on 1 and 48 DF,  p-value: 0.001391
```

3. The simple linear model of sat ~ ratio is shown below. The coefficient for ratio is 2.682. This implies that for every increase in the average student to teacher ratio by 1, there is a 2.682 increase in SAT score, which is positive just as the essay claimed. Thus, the more students the average teacher has in their class, the better the test scores.

```
##
## Call:
## lm(formula = sat ~ ratio, data = SAT)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -120.69  -63.33  -27.31   64.08  145.26
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  920.699     80.770  11.399 2.94e-15 ***
## ratio          2.682      4.749   0.565    0.575
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 75.35 on 48 degrees of freedom
## Multiple R-squared:  0.006602,   Adjusted R-squared:  -0.01409
## F-statistic: 0.319 on 1 and 48 DF,  p-value: 0.5748
```

4. The scatterplot matrix is shown below. sat and frac do appear to have a medium negative linear correlation, so as a higher fraction of students in the state take the SATs, the state has lower test scores. In addition, expend and frac appear to have a medium positive linear relationship. So as a higher fraction of students in the state take the SATs, the state also spends more on education.



5. The three models sat ~ expend + frac, sat ~ salary + frac, and sat ~ ratio + frac are all shown below in that order. In the first model, the coefficient for expend is 12.2865 which is positive where it was negative previously. In the second model, the coefficient for salary is 2.1804 which is positive where it was negative previously. In the third model, the coefficient for ratio is -2.7787 which is negative where it was positive previously. All 3 coefficients are different from what was in the previous models. In every single model, the coefficient for frac is negative and somewhere between -2.5 and -3. The p-value beling less than 2eˆ-16 shows that it is extremely statistically significant.

```
##
## Call:
## lm(formula = sat ~ expend + frac, data = SAT)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -88.400 -22.884   1.968  19.142  68.755
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 993.8317    21.8332  45.519  < 2e-16 ***
## expend       12.2865     4.2243   2.909  0.00553 **
## frac         -2.8509     0.2151 -13.253  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 32.46 on 47 degrees of freedom
## Multiple R-squared:  0.8195, Adjusted R-squared:  0.8118
## F-statistic: 106.7 on 2 and 47 DF,  p-value: < 2.2e-16
```

```
##
## Call:
## lm(formula = sat ~ salary + frac, data = SAT)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -78.313 -26.731   3.168  18.951  75.590
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 987.9005    31.8775  30.991   <2e-16 ***
## salary        2.1804     1.0291   2.119   0.0394 *
## frac         -2.7787     0.2285 -12.163    4e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 33.69 on 47 degrees of freedom
## Multiple R-squared:  0.8056, Adjusted R-squared:  0.7973
## F-statistic: 97.36 on 2 and 47 DF,  p-value: < 2.2e-16
```

```
##
## Call:
## lm(formula = sat ~ ratio + frac, data = SAT)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -88.053 -23.427   3.057  18.332  58.242
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1118.5087    39.4733  28.336   <2e-16 ***
## ratio         -3.7264     2.2089  -1.687   0.0982 .
```

```
## frac          -2.5474     0.1871 -13.618   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 34.24 on 47 degrees of freedom
## Multiple R-squared:  0.7991, Adjusted R-squared:  0.7906
## F-statistic:  93.5 on 2 and 47 DF,  p-value: < 2.2e-16
```

6. The model sat ~ expend + salary + ratio + frac is shown below. Expend has a coefficient of 4.4626 so as education expenditure increases by $1000, then SAT scores increase by 4.4626 points, holding salary, ratio, and frac constant. Salary has a coefficient of 1.6379 so as average teacher salary increases by $1000, then SAT scores increase by 1.6379 points, holding expend, ratio, and frac constant. Ratio has a coefficient of -3.6242 so as the average student to teacher ratio increases by 1, then SAT scores decrease by 3.6242 points, holding expend, salary, and frac constant. Frac has a coefficient of -2.9045 so as the fraction of students taking the SATs in the state increases by 1, then SAT scores decrease by 3.6242 points, holding expend, salary, and ratio constant.

```
##
## Call:
## lm(formula = sat ~ expend + salary + ratio + frac, data = SAT)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -90.531 -20.855  -1.746  15.979  66.571
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1045.9715    52.8698  19.784  < 2e-16 ***
## expend         4.4626    10.5465   0.423    0.674
## salary         1.6379     2.3872   0.686    0.496
## ratio         -3.6242     3.2154  -1.127    0.266
## frac          -2.9045     0.2313 -12.559 2.61e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 32.7 on 45 degrees of freedom
## Multiple R-squared:  0.8246, Adjusted R-squared:  0.809
## F-statistic: 52.88 on 4 and 45 DF,  p-value: < 2.2e-16
```