hw4 Q6

Lisong he

2024-03-28

a)

```
library(faraway)
data(sat, package = 'faraway')

sat_model <- lm(total ~ expend + salary + ratio, data=sat)
summary(sat_model)
```

```
##
## Call:
## lm(formula = total ~ expend + salary + ratio, data = sat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -140.911  -46.740   -7.535   47.966  123.329
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1069.234    110.925   9.639 1.29e-12 ***
## expend        16.469     22.050   0.747   0.4589
## salary        -8.823      4.697  -1.878   0.0667 .
## ratio          6.330      6.542   0.968   0.3383
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 68.65 on 46 degrees of freedom
## Multiple R-squared:  0.2096, Adjusted R-squared:  0.1581
## F-statistic: 4.066 on 3 and 46 DF,  p-value: 0.01209
```

b) The intercept is estimated to be 1069.234. The coefficient of expend is 16.649, -8.823 for salary, and 6.33 for ratio. But t* for all three predictors are larger than 0.05, which are statistically insignificant. There is a suggestive negative association with teacher salary if alpha is 0.1, meaning that higher teacher salaries could potentially correspond to lower average total SAT scores.

c)

```
sat_model_takers <- lm(total ~ expend + salary + ratio + takers, data=sat)
summary(sat_model_takers)
```

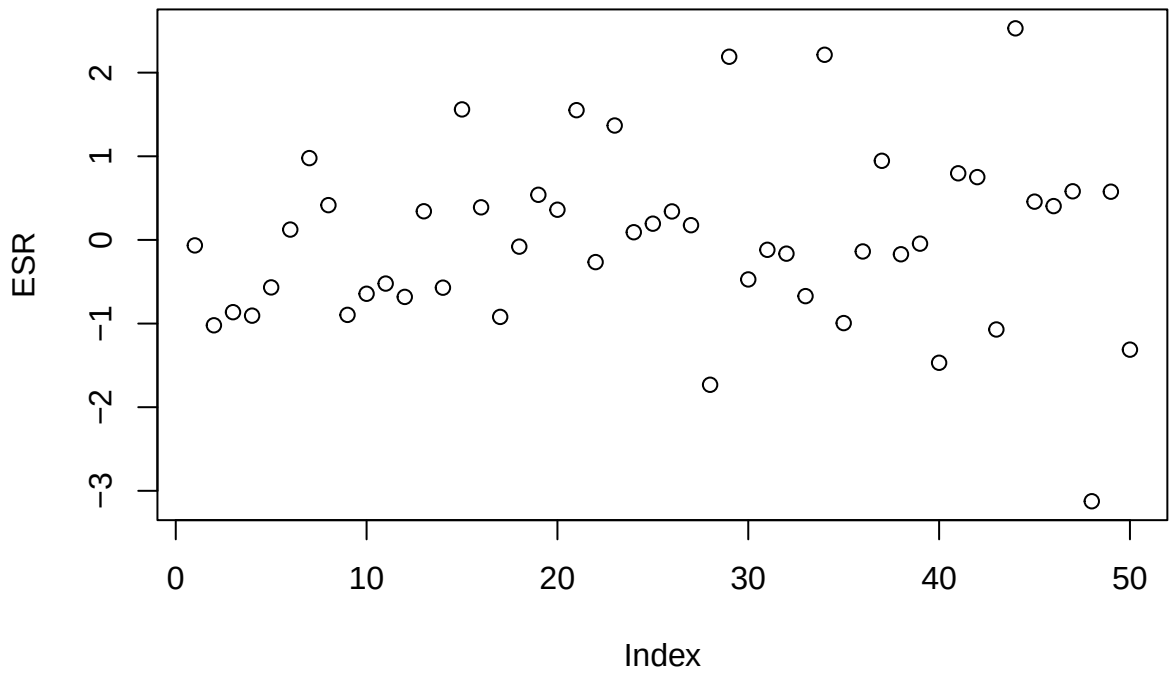
```
##
## Call:
```

```
## lm(formula = total ~ expend + salary + ratio + takers, 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
## takers        -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
```

In this new model, the old three predictors still failed to be statistically significant, whereas the new predictor 'takers' shows a highly significant ($p < 0.001$) negative association with average total SAT scores.

d)

```
ESR <- rstudent(sat_model_takers)
plot(ESR)
```



From the plot, observation number 44 and 48 are likely to be outliers given their large distance from y = 0.

```
t <- qt(1 - 0.1 / (2*50), 50 - 5 - 1, lower.tail = TRUE)
outliers <- which(abs(ESR) > t)
print(outliers)
```

```
## named integer(0)
```

In fact, no outliers are detected under the significance level 0.1.