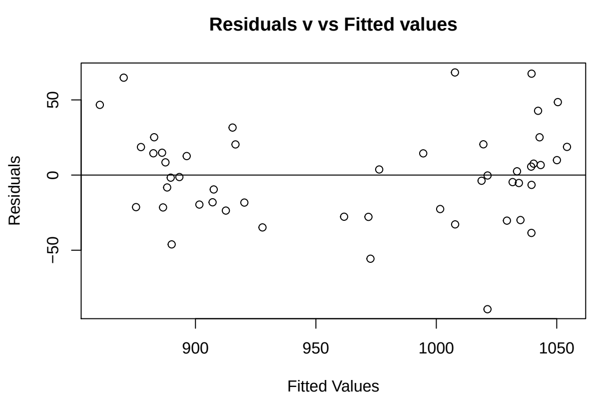
HW 3

1. Fit a model with **total** sat score as the response and **takers**, **ratio** and **salary** as predictors. The *percentage of variation* (R^2) in the response explained by the model is **82.39%**. Also, we are able to see from the summary that all B^ coefficients are significant (i.e non zero & explain variation in response) & that the assumption of constant errors (resdiudals vs fitted) seems to be satisfied. Hence, all these facts together indicate a *good* fit.

```
##
## Call:
## lm(formula = total ~ takers + ratio + salary, data = sat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -89.244 -21.485  -0.798  17.685  68.262
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1057.8982    44.3287  23.865   <2e-16 ***
## takers        -2.9134     0.2282 -12.764   <2e-16 ***
## ratio         -4.6394     2.1215  -2.187   0.0339 *
## salary         2.5525     1.0045   2.541   0.0145 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 32.41 on 46 degrees of freedom
## Multiple R-squared:  0.8239, Adjusted R-squared:  0.8124
## F-statistic: 71.72 on 3 and 46 DF,  p-value: < 2.2e-16
```



2. **H0**: Bsalary <=0 , **HA**: Bsalary > 0. We are going to perform a *t-test*(Bsalary/se(Bsalary)) that follows a distribution of **t(46)** with a value of **2.541**. The *p-value* associated with this test statistics is **0.0145**. Now this is reflective of a two sided hypothesis test (i.e Bsalaray =0, or Bsalary !=0). Hence, the appropriate p-value for a one sieded t test is 0.0145/2 = **0.00725**, which is smaller than our alpha level 0.01. As a result, we *reject* the null hypothesis in which Bsalary<= 0, in favor of the alternative Bsalary>0.

3. **H0**: Bratio = 0, **HA**: Bratio != 0 .We are going to perform an *t-test*(Bratio/se(Bratio)) that follows a distribution of **t(46)** with a value of **-2.187**. The *p-value* associated with this test statistics is **0.0339**. Now this is reflective of a two sided hypothesis test (i.e Bsalaray =0, or Bsalary !=0). As a result, we

fail to reject the null hypothesis,if our alpha level is 0.01. In case our alpha level is 0.05, then we would *reject* the null hypothesis in favour of the alternative.

4. **H0**: Bratio=Bsalary=Btakers=0, **H1**: Btakers!=0 or Bratio!=0 or Bsalary!=0. In this model we are testing the hypthesis that at least one of the coefficients is not zero (i.e Bratio!=0). We are able to use an **F-test** with distribution **F(3, 46)** and an f-value of **71.721** which yields a p-value of ~**0**. Hence, we *reject* the null in favor of the alternative meaning that at least one coefficient will be significant in the model (i.e explain variance in the response).

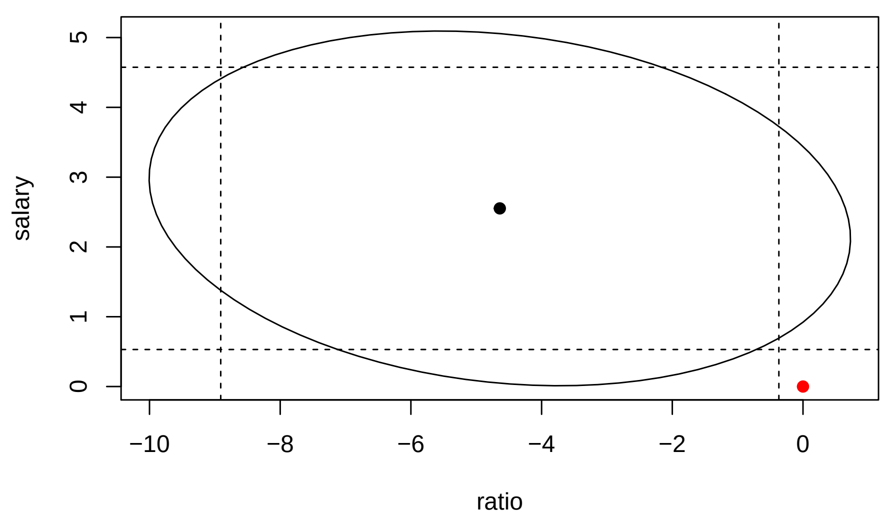
```
## Analysis of Variance Table
##
## Model 1: total ~ 1
## Model 2: total ~ takers + ratio + salary
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     49 274308
## 2     46  48315  3    225992 71.721 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

5. The 95% interval does not include zero, whereas the 99% interval does indeed include zero. As a result, the p-value (probability of getting a result as extreme or more as the observerd) will be contained somewhere between *0.01 and 0.05*, which is indeed within reasonable alpha levels (i.e 5%).

```
## [1] 0.5305464 4.5744536
```

```
## [1] -0.1466051  5.2516051
```

6. Below is the 95% joint confidence region for the parameters associated with **ratio** and salary. The origin point (0,0) is displayed in red. The location of the origin on the plot tells us the outcome of hypothesis **H0:** Bsalary=Bratio = 0, **HA**: Bsalary or Bratio !=0. We *reject* the null hypothesis,at alpha = 0.05, given that the point (0,0) lies outside of the joint confidence interval.



7. For what it pertains the coefficients **takers** hasn't been affected much, and it still retains its significance. On the other side, both **ratio** and **salary** both have drammatically changed coefficients & have lost their significance (p-value greater than any common alpha level). In terms of goodness of fit, we are still able to retain a slightier higher R^2 **82.46%**, but given the addition of a new predictor looking at the adjusted R^2 gives a better view of the data. With the adjusted R^2 decreasing in comparison to the previous model. Overall, I do not see expend as improving the model.

```
##
## Call:
## lm(formula = total ~ takers + ratio + salary + expend, 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 ***
## takers        -2.9045     0.2313 -12.559 2.61e-16 ***
## ratio         -3.6242     3.2154  -1.127    0.266
## salary         1.6379     2.3872   0.686    0.496
## expend         4.4626    10.5465   0.423    0.674
## ---
## 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
```

8. **H0**: Bratio=Bsalary=Bexpend=0 given takers, **HA**: Bratio!= 0 or Bsalary!=0 or Bexpend!=0 given takers. We can run an F-test that follows a distribution F(3, 45) which yields an F-value of **3.2133** and a p-value of **0.03165**. At an alpha level of 5%, we *reject* the null in favor of the alternative.

```
## Analysis of Variance Table
##
## Model 1: total ~ takers
## Model 2: total ~ takers + ratio + salary + expend
##   Res.Df   RSS Df Sum of Sq      F  Pr(>F)
## 1     48 58433
## 2     45 48124  3     10309 3.2133 0.03165 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```