1.1.) a) Fit the model:

>h0 <- lm (total ~ takers + ratio + salary, sat)

All the three predictors have significant coefficients. The model fits well as the R-squared is over 0.8. While the takers and ratio variables are negative correlated to sat total score, indicating that lower percentage of sat takers, lower pupil/teacher ratio and higher average annual salary of teachers might be related to higher sat total score.

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

b)To test hypothesis βsalary=0, fit the model

>h0a<- lm (total ~ takers + ratio, sat)

>anova (h0, h0a)

Analysis of Variance Table

Model 1: total ~ takers + ratio + salary

Model 2: total ~ takers + ratio

Res.Df RSS Df Sum of Sq F Pr(>F)

1 46 48315

2 47 55097 -1 -6781.6 6.4566 0.01449 \*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

c) We found that the p-value < 0.05, so we reject the hypothesis βsalary=0

To test hypothesis βtakers = βratio = βsalary=0, fit the model

>h0b<-lm (total ~ 0, sat)

>anova(h0,h0b)

Analysis of Variance Table

Model 1: total ~ takers + ratio + salary

Model 2: total ~ 0

Res.Df RSS Df Sum of Sq F Pr(>F)

1 46 48315

2 50 46924380 -4 -46876065 11157 < 2.2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

d)We found that the p-value < 0.05, so we reject the hypothesis βtakers = βratio = βsalary=0

2.)

>confint(h0,"salary",level=0.95)

2.5 % 97.5 %

salary 0.5304797 4.574461

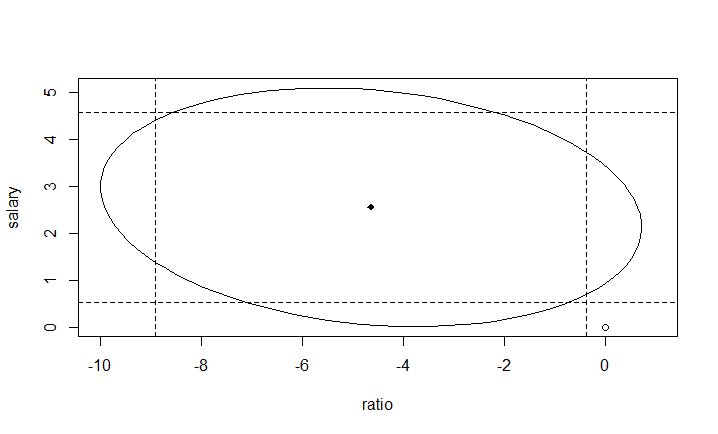
>confint(h0,"salary",level=0.99)

0.5 % 99.5 %

salary -0.146684 5.251624

Because that the 99% CI includes 0, the p-value indicates non-significant , that is >0.01

3.) The joint confidence region for the parameters associated with ratio and salary is shown below:



The origin is located outside of the confidence region, so the hypothesis βratio = βsalary=0 is rejected.

4.) After adding expend to the model, the results are as follows:

>h1<-lm(total~expend+takers+ratio+salary,sat)

Call:

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

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

---

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

The model is still significant, the coefficients of the ratio and salary changed a lot. The coefficients of takers remains significant. The goodness of fit is almost the same compared to the model in question 1.

5.) >h1a<-lm(total~takers,sat)

> anova(h1,h1a)

Analysis of Variance Table

Model 1: total ~ expend + takers + ratio + salary

Model 2: total ~ takers

Res.Df RSS Df Sum of Sq F Pr(>F)

1 45 48124

2 48 58433 -3 -10309 3.2133 0.03165 \*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

The hypothesis βsalary= βexpend = βratio =0 is rejected.

Based on my entire analysis, I feel the salary may have an effect on the response.

2.

TSS-total sum of squares RSS-residual sum of squares p=number of predictors

n-number of observations