SGupta_HW03Question2

Question 2

Test the hypothesis that Salary = 0. Test the hypothesis that salary= ratio = expend = 0

```
library(faraway)
set.seed(1001)
data(sat)
head(sat)
```

```
           expend ratio salary takers verbal math total
Alabama     4.405  17.2 31.144      8    491  538  1029
Alaska      8.963  17.6 47.951     47    445  489   934
Arizona     4.778  19.3 32.175     27    448  496   944
Arkansas    4.459  17.1 28.934      6    482  523  1005
California  4.992  24.0 41.078     45    417  485   902
Colorado    5.443  18.4 34.571     29    462  518   980
```

```
model1 <- lm(total ~ expend + ratio + salary, data = sat)
summary(model1)
```

```
Call:
lm(formula = total ~ expend + ratio + salary, 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
ratio          6.330      6.542   0.968   0.3383
salary        -8.823      4.697  -1.878   0.0667 .
---
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
```

```r
# Extract t-statistic and p-value for salary
summary(model1)$coefficients["salary", ]
```

```
   Estimate  Std. Error     t value    Pr(>|t|)
-8.82263197  4.69679359 -1.87843723  0.06666771
```

The p value for salary predictor variable is 0.06 which is slightly higher than 0.05 significance level. Therefore we do no have enough evidence to reject null hypothesis. There is evidence that salary has no effect on SAT score.

Testing beta coefficients for expend, ratio and salary all are zero –> We can use the F-test to check if all predictors has any effect on the response variable. The F statistic value is 0.01 which is less than significant level therefore we reject the null hypothesis that all coefficients are zero. Alteast one of the predictor variable has an effect on the SAT score.

- expend: = p = 0.4589 - not significant
- ratio: = p = 0.3383 - not significant
- salary: = p = 0.0667 marginal significance at the 10% level, but not at the 5% level

Now add takers to the model. Test the hypothesis that salary = 0.

```r
data(sat)
model2 <- lm(total ~ expend + ratio + salary + takers, data = sat)
summary(model2)
```

```
Call:
lm(formula = total ~ expend + ratio + salary + 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
ratio         -3.6242     3.2154  -1.127    0.266
salary         1.6379     2.3872   0.686    0.496
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
```

```r
summary(model2)$coefficients["salary", ]
```

```
  Estimate Std. Error    t value   Pr(>|t|)
 1.6379172  2.3872480  0.6861110  0.4961632
```

Since the p-value (0.4962) is much greater than significance level, we fail to reject the null hypothesis After accounting for the takers predictors in model2, salary does not appear to have a significant effect on the response.

```r
anova(model1, model2)
```

```
Analysis of Variance Table

Model 1: total ~ expend + ratio + salary
Model 2: total ~ expend + ratio + salary + takers
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1     46 216812
2     45  48124  1    168688 157.74 2.607e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Model 1: total expend+ratio+salarytotal expend+ratio+salary
- Model 2:total expend+ratio+salary+takerstotal expend+ratio+salary+takers The RSS is reduced in model 2 and this is the extra variation by including the variable takers.The associated p value in model 2 is extremely small which indicates adding takers predictor variable significantly improved model.