Homework 4

3.3

a)

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 22.55565 17.19680 1.312 0.1968

sex -22.11833 8.21111 -2.694 0.0101 \*

status 0.05223 0.28111 0.186 0.8535

income 4.96198 1.02539 4.839 1.79e-05 \*\*\*

verbal -2.95949 2.17215 -1.362 0.1803

---

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

The sex and income variables are statistically significant at the 5% level .

b)

There is moderate evidence that the sex variable should be included in the model, when all other predictors – status, income, and verbal – are included in the model.

c)

Analysis of Variance Table

Model 1: gamble ~ income

Model 2: gamble ~ sex + status + income + verbal

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

1 45 28009

2 42 21624 3 6384.8 4.1338 0.01177 \*

---

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

Based on the analysis, we have moderate evidence to reject the null hypothesis, and conclude that the coefficients for sex and status are not equal to 0.

4.

a)

Analysis of Variance Table

Model 1: total ~ expend + ratio

Model 2: total ~ expend + ratio + salary

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

1 47 233443

2 46 216812 1 16631 3.5285 0.06667 .

---

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

The test concludes with a p-value of .06667, so we have to fail to reject the null hypothesis that the coefficient for the salary predictor is 0, given that the coefficient for ratio and expend are not 0. is in our model.

Analysis of Variance Table

Model 1: total ~ 1

Model 2: total ~ expend + ratio + salary

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

1 49 274308

2 46 216812 3 57496 4.0662 0.01209 \*

---

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

The test concludes with a p-value of .01209, so we have moderate evidence that we should reject null hypothesis and say there is reason to include the predictors in our model.

b)

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

The test concludes with a p-value of near 0, so we have very strong evidence that we should reject the null, and claim that the takers regressor should be included in our model, along with ratio, expend, and salary.

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

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

Using the summary output from our model that explains total using predictors expend, ratio, salary, and takers, we want to show that the t-test and the F-test are equivalent. We know from the chapter that:

And that:

We know from the ANOVA output that our F-statistic is 157.74. Additionally, we know from our model summary output that , and . Thus, we have

Which rounds to 157.74. Thus, I have shown that the square of our t-statistic is equal to the F-statistic that would be.