# Homework 6 Key

Subrata Paul

10/25/2020

## Problem 3.3

(a) The predictors sex and income are significant at the 0.05 level.

```
lmod = lm(gamble ~., data = faraway::teengamb)
xtable(lmod)
```

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 22.5557 | 17.1968 | 1.31 | 0.1968 |
| sex | -22.1183 | 8.2111 | -2.69 | 0.0101 |
| status | 0.0522 | 0.2811 | 0.19 | 0.8535 |
| income | 4.9620 | 1.0254 | 4.84 | 0.0000 |
| verbal | -2.9595 | 2.1722 | -1.36 | 0.1803 |

(b) We predict that females will spend about 22 lbs less per year on gambling than their male counterparts, assuming that their socioeconomic status, income, and verbal scores are the same.

(c)

$H_0: \ \beta_{\text{sex}} = \beta_{\text{status}} = \beta_{\text{verbal}} = 0 | \beta_{\text{income}} \neq 0$

$H_a: \ \beta_{\text{sex}}$ or $\beta_{\text{status}}$ or $\beta_{\text{verbal}} \neq 0 | \beta_{\text{income}} \neq 0$

```
full_mod = lm(gamble ~ ., data = faraway::teengamb)
reduced_mod = lm(gamble ~ income, data = faraway::teengamb)
(rssf = deviance(full_mod))
```

```
## [1] 21623.77
```

```
(rssr = deviance(reduced_mod))
```

```
## [1] 28008.59
```

```
(f = ((rssr - rssf)/3)/(rssf/42))
```

```
## [1] 4.133761
```

```
1 - pf(f, df1 = 3, df2 = 42)
```

```
## [1] 0.01177211
```

$$F = \frac{(28008.59 - 21623.77)/3}{21623.77/42} = 4.1337611$$

p-value $= P(F_{3,42} \geq 4.13) = 0.012$

There is moderate evidence that at least one of the predictors sex, status, or verbal should be included in the regression model of gamble that already contains the income predictor.

## Problem 3.4

(a)

$H_0: \ \beta_{\text{salary}} = 0|\beta_{\text{ratio}} \neq 0, \beta_{\text{expend}} \neq 0$

$H_0: \ \beta_{\text{salary}} \neq 0|\beta_{\text{ratio}} \neq 0, \beta_{\text{expend}} \neq 0$

```r
fmod = lm(total ~ expend + ratio + salary, data = faraway::sat)
summary(fmod)
```

```
##
## Call:
## lm(formula = total ~ expend + ratio + salary, data = faraway::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 0.00000000000129 ***
## 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
```

$$t = \frac{-8.822632}{4.6967936} = -1.8784372$$

p-value $= 2P(T_{46} \geq 1.8784372) = 0.0666677$

There is weak evidence to conclude the salary predictor should be included in the model of sat scores that already has the ratio and expend predictors.

$H_0: \ \beta_{\text{salary}} = \beta_{\text{ratio}} = \beta_{\text{expend}} = 0$

$H_0: \ \beta_{\text{salary}} \neq 0 \text{ or } \beta_{\text{ratio}} \neq 0 \text{ or } \beta_{\text{expend}} \neq 0$

```r
rmod = lm(total ~ 1, faraway::sat)
(rssf = deviance(fmod))
```

```
## [1] 216811.9
```

```r
(rssr = deviance(rmod))
```

```
## [1] 274307.7
```

```r
(f = ((rssr - rssf)/(rmod$df.residual - fmod$df.residual))/(rssf/fmod$df.residual))
```

```
## [1] 4.066203
```

```r
1 - pf(f, df1 = rmod$df.residual - fmod$df.residual, df2 = fmod$df.residual)
```

```
## [1] 0.01208607
```

$$F = \frac{(274307.68 - 216811.94)/3}{216811.94/46} = 4.0662033$$

p-value $= P(F_{3,42} \geq 4.13) = 0.0120861$.

There is moderate evidence to conclude at least one of the predictors salary, ratio, or expend should be used in modeling the sat predictor.

(b)

F test:

$H_0: \ \beta_{\text{takers}} = 0 | \beta_{\text{ratio}} \neq 0, \beta_{\text{expend}} \neq 0, \beta_{\text{salary}} \neq 0$

$H_a: \ \beta_{\text{takers}} \neq 0 | \beta_{\text{ratio}} \neq 0, \beta_{\text{expend}} \neq 0, \beta_{\text{salary}} \neq 0$

```
options(scipen=5)
rmod = lm(total ~ expend + ratio + salary, faraway::sat)
fmod = lm(total ~ expend + ratio + salary + takers, faraway::sat)
(rssf = deviance(fmod))
```

```
## [1] 48123.9
```

```
(rssr = deviance(rmod))
```

```
## [1] 216811.9
```

```
(f = ((rssr - rssf)/(rmod$df.residual - fmod$df.residual))/(rssf/fmod$df.residual))
```

```
## [1] 157.7379
```

```
1 - pf(f, df1 = rmod$df.residual - fmod$df.residual, df2 = fmod$df.residual)
```

```
## [1] 2.220446e-16
```

$$F = \frac{(216811.94 - 48123.9)/1}{48123.9/45} = 157.7378885$$

p-value $= P(F_{3,42} \geq 4.13) = 2.220446 \times 10^{-16}$.

There is very strong evidence to conclude that the takers predictor should be included in the model of sat scores that already has the ratio, salary, and expend predictors.

t test:

```
library(xtable)
xtable(fmod)
```

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 1045.9715 | 52.8698 | 19.78 | 0.0000 |
| expend | 4.4626 | 10.5465 | 0.42 | 0.6742 |
| ratio | -3.6242 | 3.2154 | -1.13 | 0.2657 |
| salary | 1.6379 | 2.3872 | 0.69 | 0.4962 |
| takers | -2.9045 | 0.2313 | -12.56 | 0.0000 |

$$t = \frac{-2.9044805}{0.23126} = -12.5593745$$

p-value $= 2P(T_{46} \geq 12.5593745) = 2.6065588 \times 10^{-16}$

Notice that $t^2 = F = 157.74$