HW2

Chong Zhang

September 16, 2018

(a) What percentage of variation in the response is explained by these predictors?

```
library(faraway)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
head(teengamb)
```

```
##   sex status income verbal gamble
## 1   1     51   2.00      8    0.0
## 2   1     28   2.50      8    0.0
## 3   1     37   2.00      6    0.0
## 4   1     28   7.00      4    7.3
## 5   1     65   2.00      8   19.6
## 6   1     61   3.47      6    0.1
```

```
model = lm(gamble~., data = teengamb)

summary(model)
```

```
##
## Call:
## lm(formula = gamble ~ ., data = teengamb)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -51.082 -11.320  -1.451   9.452  94.252
##
## 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
##
## Residual standard error: 22.69 on 42 degrees of freedom
## Multiple R-squared:  0.5267, Adjusted R-squared:  0.4816
## F-statistic: 11.69 on 4 and 42 DF,  p-value: 1.815e-06
```

Since **R-squared** is the parameter that is used to percentage of variance explained by these predictors. Thus the answer is **`0.5267234`** . One can also use **Adjusted R-squared**, which is **`0.4816495`**

(b) For all other predictors held constant, what would be the difference in predicted expenditure on gambling for a male compared to a female?

```
male = mutate(teengamb, sex = 0)
female =  mutate(teengamb, sex = 1)

male_exp = mean(predict(object = model, newdata = male))
female_exp = mean(predict(object = model, newdata = female))

difference = male_exp - female_exp
```

The differnce in predicted expenditure on gambling for a male compared to a female is **`22.1183301`** , if hold all other predictors constant. Alternatively, we can also obtain this information from the coefficient of **Sex**, which is also **-22.11833**.

(c) Write down the least squares regression equation using all the variables in the model.

$$Expenditure = 22.55 - 22.12 * Sex + 0.05 * Status + 4.96 * Income - 2.96 * Verba$$

(d) Which variables are statistically significant (at the 5% level of significance)? Interpret.

Both **Sex** and **Income** are statistically significant (at the 5% level of significance). For Sex, it means that if hold all other predictors constant, female (marked as 1) spend 22.11 pounds less than male (marked as 0) on gambling. For income, it means that if hold all other predictors constant, one pound/week increase in income will lead to 4.96 pounds more spent on gambling.

(e) Find a 90% confidence interval for $\hat{\beta}_0$. Interpret how this interval can help in determining the significance of this coefficient in the model.

```
confidence = confint(model, level = 0.90)
confidence[1,]
```

```
##       5 %      95 %
## -6.368553 51.479855
```

The 90% confidence interval for $\hat{\beta}_0$ is (`-6.3685535, 51.4798547`). Since this interval includes **0**, it means that alternative hypothesis that $\hat{\beta}_0$ **is not equal to 0** is rejected. Thus $\hat{\beta}_0$ is not significant in the model.