SGupta_HW03Question4

Question 1

(a) Which variables are statistically significant?

```
library(faraway)
data(teengamb)
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
```

```
# Fit the full model
model <- lm(gamble ~ sex + status + income + verbal, data = teengamb)
summary(model)
```

```
Call:
lm(formula = gamble ~ sex + status + income + verbal, 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
```

Considering = 0.05, sex is statistically significant as the p value 0.01 is less than 0.05.

Status is insignificant is this model preparation as the p value 0.8 is higher than alpha level.

income is highly significant. its p value is only 0.000017

And verbal is also not significant.

(b) What interpretation should be given to the coefficient for sex?

```r
summary(model)$coefficients["sex", c("Estimate", "Std. Error","t value", "Pr(>|t|)")]
```

```
    Estimate   Std. Error      t value     Pr(>|t|)
-22.11833009   8.21111453  -2.69370620   0.01011184
```

A unit increase in sex variable is associated with an average decrease in response variable gamble by about 22.12 units. A female variable tends to gamble less than male by 22.12 units.

(c) Fit a model with just income as a predictor and use an F-test to compare it to the full model.

```r
model_income <- lm(gamble ~ income, data = teengamb)

anova(model_income, model)
```

```
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
```

- The F-statistic is about 4.133 with 3 and 42 degrees of freedom.
- The associated p-value is approximately 0.01177
- The RSS difference between the two models is 6,384.8. This is due to the extra variables (variations) by adding the predictors sex, status, and verbal in the full model.
- Since the p-value (0.01177) is less than the significance level (0.05), we reject the null hypothesis that the extra predictors (sex, status, and verbal) contribute no additional significance over the income-only model.