SGupta_HW02Question2

```r
# Load necessary library
library(faraway)
# Load the dataset
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
```

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

```r
rsquared <- summary(model)$r.squared
print(paste("Percentage of variation explained:", rsquared * 100, "%"))
```

```
[1] "Percentage of variation explained: 52.6723412826738 %"
```

```r
# Get residuals
residuals <- model$residuals

# (b) Find the index of the largest positive residual
largest_residual_number <- which.max(residuals)
print(paste("Largest positive residual:", max(residuals) ))
```

```
[1] "Largest positive residual: 94.2522174243442"
```

```r
print(paste("Observation with the largest positive residual:", largest_residual_number ))
```

```
[1] "Observation with the largest positive residual: 24"
```

```r
# (c) Mean and median of residuals
mean_residual <- mean(residuals)
median_residual <- median(residuals)

print(paste("Mean of residuals:", mean_residual ))
```

```
[1] "Mean of residuals: 2.64563784591527e-16"
```

```r
print(paste("Median of residuals:", median_residual ))
```

```
[1] "Median of residuals: -1.45139206896952"
```

```
# Get fitted values
values <- model$fitted.values

# (d) Compute correlation
correlation1 <- cor(residuals, values)
print(paste("Correlation of residuals with fitted values:", correlation1 ))
```

[1] "Correlation of residuals with fitted values: 4.76558784202177e-17"

```
# (e) Compute correlation
correlation2 <- cor(residuals, teengamb$income)
print(paste("Correlation of residuals with income:", correlation2 ))
```

[1] "Correlation of residuals with income: -2.00608596581065e-17"

```
# (f) Extract the coefficient for sex
sex_coefficient <- coef(model)["sex"]
print(paste("Difference in predicted expenditure for males vs. females:", sex_coefficient ))
```

[1] "Difference in predicted expenditure for males vs. females: -22.1183300928389"