

## Problem 5

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
---

```

Sex and income are statistically significant at the 5% level.

```
> confint(lm, level=0.95)
              2.5 %      97.5 %
(Intercept) -12.1489038  57.2602050
sex          -38.6890301 -5.5476301
status       -0.5150722  0.6195399
income        2.8926538  7.0313047
verbal       -7.3430703  1.4240833

```

b) Sex is just a dummy variable representing the categorical data of sex. 1 represents female and 0 represents male.

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

```

Since the p-value is very small the null hypothesis is rejected.

d)

```
Model 1: gamble ~ 1
Model 2: gamble ~ sex + status + income + verbal
  Res.Df  RSS Df Sum of Sq    F    Pr(>F)
1      46 45689
2      42 21624  4    24066 11.686 1.815e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

## Appendix

```
# title: "MSDS596 - HW2"
# author: "Diego Sarachaga"
# date: "10/02/2018"
```

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

```
lm <- lm(gamble ~ sex + status + income + verbal, teengamb)
summary(lm)
```

```
#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
#
# 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
```

```
#Sex and income are statistically significant at the 5% level
confint(lm, level=0.95)
```

```
#Confident intervals
#sex      -38.6890301 -5.5476301
#income    2.8926538  7.0313047
```

```
#b
#Sex is a dummy variable representing the categorical data of sex. A 1
represents female and a 0 represents male.
```

```
#c
lmi <- lm(gamble ~ income, teengamb)
summary(lmi)
```

```
anova(lmi, lm)
# 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
```

#Since the p-value is so small the null hypothesis is rejected.

```
#d
nullmod <- lm(gamble ~ 1, teengamb)
anova(nullmod, lm)
# Analysis of Variance Table
#
# Model 1: gamble ~ 1
# Model 2: gamble ~ sex + status + income + verbal
# Res.Df  RSS Df Sum of Sq    F  Pr(>F)
# 1    46 45689
# 2    42 21624  4   24066 11.686 1.815e-06 ***
# ---
# Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```