# Problem 1

1. The rank is .

This has infinitely many solutions because there are four distinct equations, but there are six unknowns. There are too many unknown variables, and not enough distinct equations to find distinct solutions for each unknown variable.

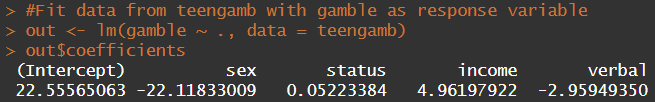
Yes, is estimable.

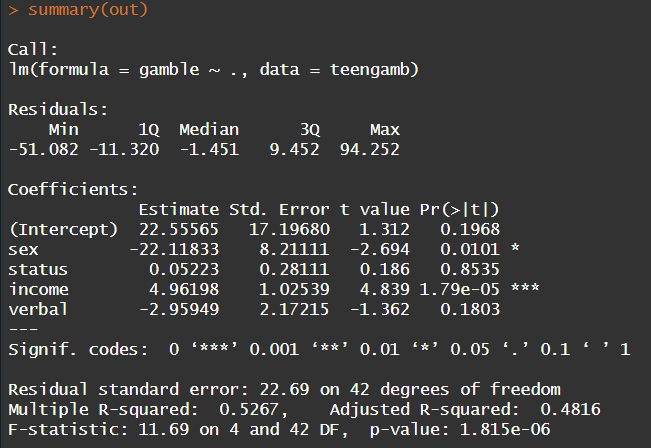
Yes, is estimable.

A screenshot of a computer program

Description automatically generated

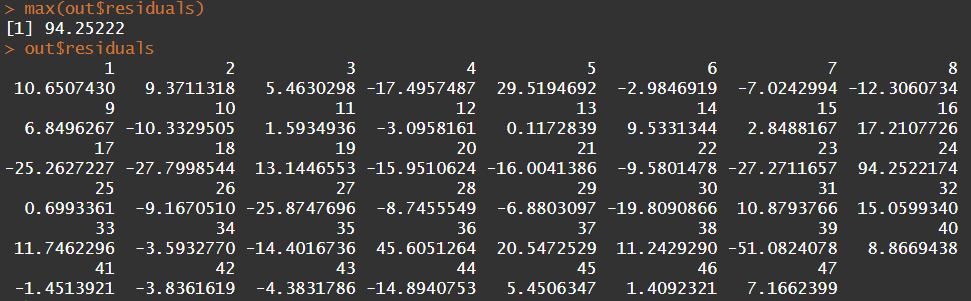
# Problem 2



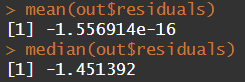


Approximately 52.67% of the variation in gamble is accounted for by the other predictors.

1. Largest residual was 94.2522174, which is case number 24.



1. Mean is approximately 0. Median residual is -1.451392.



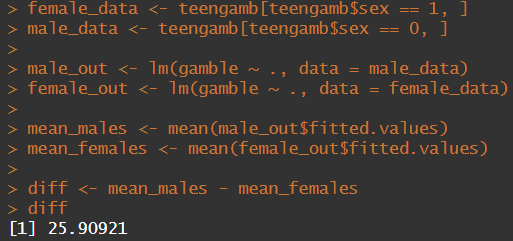
1. Approximately 0.



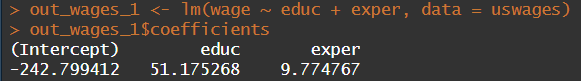
1. Approximately 0.



1. The difference between the predicted expenditure on gambling for a male compared to a female is approximately 25.90921. The mean value for males was approximately 29.775, whereas the mean value for females was 3.866.

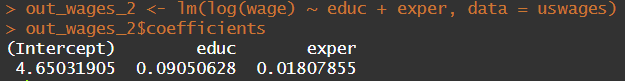


# Problem 3



For the model with wages as the response variable and years of education and experience as the explanatory variables, we found the model below, where is the years of education and is the number of years of experience.

The regression coefficient for years of education is approximately 51.18, which means that the wage would increase on average by 51.18 for every additional year of education.



For the model with log(wages) as the response variable and years of education and experience as the explanatory variables, we found the model below, where is the years of education and is the number of years of experience.

When solved for y becomes:

The regression coefficient for years of education is approximately 0.0905, which becomes a multiplicative factor of 1.0947. This means that the wage will increase on average by 9.47% for each additional year of education.

9.47% growth per year sounds more natural, because you are growing at a percentage based on your previous year of education instead of simply adding a flat rate per each additional year.

# Appendix (R Code)

#####################################

# Author: Evan Whitfield

# Date Last Edit: 1-24-25

# Purpose: To answer problems for ST503 HW2

#####################################

#

# Problem 1

#

#################################

library(estimability)

#Define model matrix

X <- cbind(rep(1,6),

c(rep(1, 3),rep(0,3)),

c(rep(0, 3),rep(1,3)),

c(1,0,0,1,0,0),

c(0,1,0,0,1,0),

c(0,0,1,0,0,1))

#coefficient vectors

cvec1 <- c(0, 1, -1, 0, 0, 0)

cvec2 <- c(0, 0, 0, 1, -2, 1)

nb <- nonest.basis(X)

#checking estimability

is.estble(cvec1,nb)

is.estble(cvec2,nb)

#################################

#

# Problem 2

#

#################################

library(faraway)

#Fit data from teengamb with gamble as response variable

out <- lm(gamble ~ ., data = teengamb)

out$coefficients

summary(out)

#Determining statistics for the residuals

mean(out$residuals)

median(out$residuals)

max(out$residuals)

out$residuals

#Finding correlation between the residuals and the fitted values

cor(out$residuals,out$fitted.values)

#Finding correlation between residuals and income variable

cor(out$residuals,teengamb$income)

#Sub-setting the data based on gender

female\_data <- teengamb[teengamb$sex == 1, ]

male\_data <- teengamb[teengamb$sex == 0, ]

#Determing linear model for each gender

male\_out <- lm(gamble ~ ., data = male\_data)

female\_out <- lm(gamble ~ ., data = female\_data)

#Calculating mean(expected) fitted value

mean\_males <- mean(male\_out$fitted.values)

mean\_females <- mean(female\_out$fitted.values)

#calculating the difference beween the expected values of each gender

diff <- mean\_males - mean\_females

diff

############################################

#

# Problem 3

#

############################################

out\_wages\_1 <- lm(wage ~ educ + exper, data = uswages)

out\_wages\_1$coefficients

out\_wages\_2 <- lm(log(wage) ~ educ + exper, data = uswages)

out\_wages\_2$coefficients