Problem Set 11

Economemtrics

1a.

```r
library(AER)
```

```
## Loading required package: car

## Loading required package: carData

## Loading required package: lmtest

## Loading required package: zoo

##

## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

## Loading required package: sandwich

## Loading required package: survival
```

```r
data("Affairs")


Affairs$regular <- ifelse(Affairs$affairs == 12, 1, 0)
Affairs$female <- ifelse(Affairs$gender == "female", 1, 0)
Affairs$religious <- ifelse(Affairs$religiousness > 2, 1, 0)


head(Affairs)
```

```
##    affairs gender age yearsmarried children religiousness education occupation
## 4        0   male  37        10.00       no             3        18          7
## 5        0 female  27         4.00       no             4        14          6
## 11       0 female  32        15.00      yes             1        12          1
## 16       0   male  57        15.00      yes             5        18          6
## 23       0   male  22         0.75       no             2        17          6
## 29       0 female  32         1.50       no             2        17          5
##    rating regular female religious
## 4       4       0      0         1
## 5       4       0      1         1
## 11      4       0      1         0
## 16      5       0      0         1
## 23      3       0      0         0
## 29      5       0      1         0
```

```r
model <- glm(
    regular ~ education + yearsmarried + age + children + female + religious,
    data = Affairs, family = binomial(link = "logit")
)

model_summary <- summary(model)
print(model_summary)
```

```
##
## Call:
## glm(formula = regular ~ education + yearsmarried + age + children +
##     female + religious, family = binomial(link = "logit"), data = Affairs)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.9898  -0.3928  -0.2806  -0.1961   3.0559
```

```
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)    0.40522    1.52063   0.266 0.789868
## education     -0.15178    0.07286  -2.083 0.037237 *
## yearsmarried   0.22642    0.05863   3.862 0.000113 ***
## age           -0.05121    0.02957  -1.732 0.083261 .
## childrenyes   -0.49653    0.53300  -0.932 0.351554
## female        -0.42646    0.38746  -1.101 0.271057
## religious     -0.94506    0.35903  -2.632 0.008481 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 283.38  on 600  degrees of freedom
## Residual deviance: 255.83  on 594  degrees of freedom
## AIC: 269.83
##
## Number of Fisher Scoring iterations: 6
```

1b.

```r
p_values <- model_summary$coefficients[
    c("female", "religious", "childrenyes"), "Pr(>|z|)"
]

significant <- p_values < 0.05
print(significant)
```

```
##      female   religious childrenyes
##       FALSE        TRUE       FALSE
```

1c.

```
library(margins)

marginal_effects <- margins(model, variables = c("education", "yearsmarried"))
summary(marginal_effects)
```

```
##         factor     AME     SE       z      p   lower   upper
##      education -0.0085 0.0041 -2.0487 0.0405 -0.0166 -0.0004
##   yearsmarried  0.0127 0.0035  3.6236 0.0003  0.0058  0.0195
```

Both education and yearsmarried seem to have a significant effect on the probability of having a regular affair. More educaiton seems to have a negative effect while yearsmarried seems to have a positive effect.

2a.

```
simualte <- function(n) {
    x <- rnorm(n)
    u <- rnorm(n)
    y <- ifelse(1 + x + u > 0, 1, 0)
    return(data.frame(x, y))
}

ame_df <- data.frame(ols = numeric(), probit = numeric())

for (i in 1:1000) {
    data <- simualte(1000)
    ols <- lm(y ~ x, data = data)
    probit <- glm(y ~ x, data = data, family = binomial(link = "probit"))
    ame_df[i, "ols"] <- summary(ols)$coefficients[2, "Estimate"]
    ame_df[i, "probit"] <- summary(margins(probit))$AME
}
```

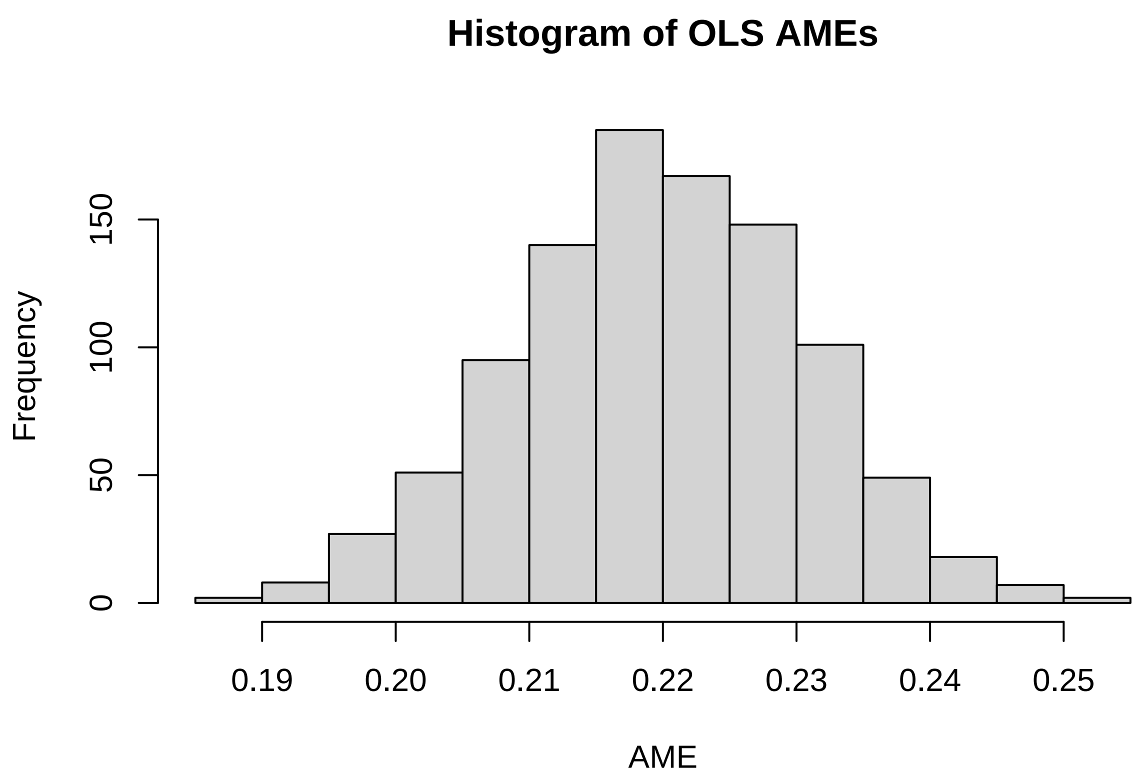
```r
ols_average <- mean(ame_df$ols, na.rm = TRUE)
probit_average <- mean(ame_df$probit, na.rm = TRUE)


print(paste("Average OLS AME: ", ols_average))
```

```
## [1] "Average OLS AME:  0.219811951418305"
```

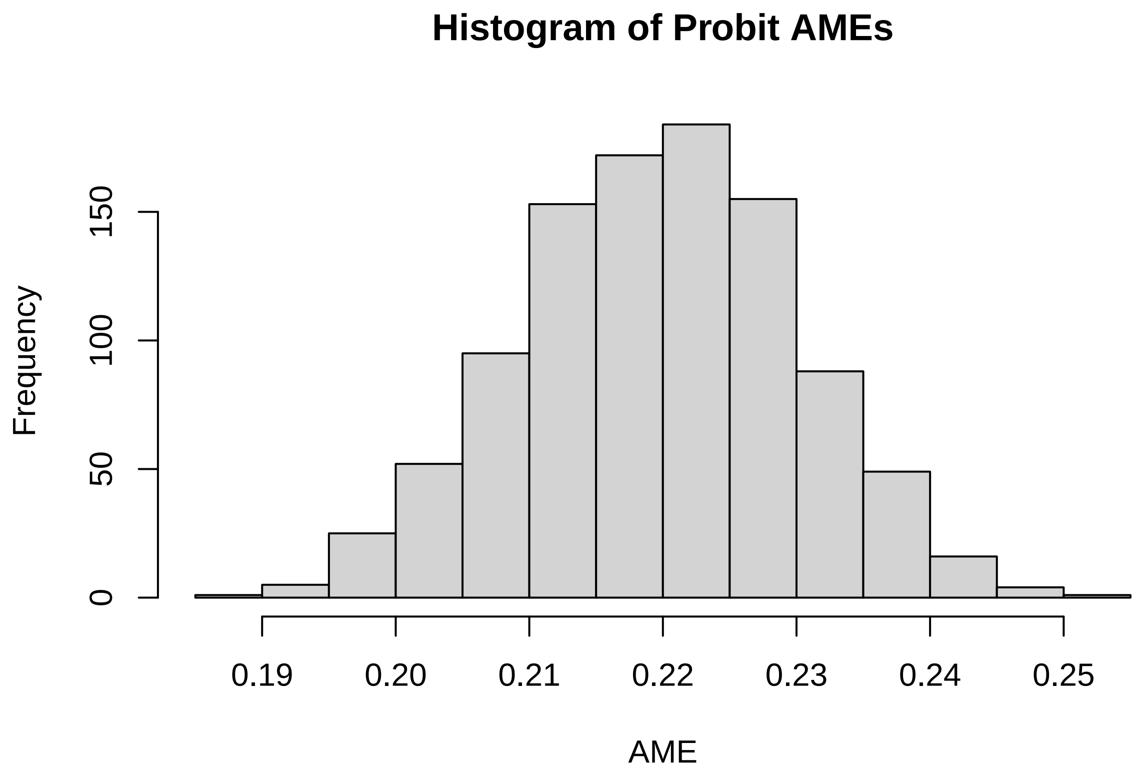
```r
print(paste("Average Probit AME: ", probit_average))
```

```
## [1] "Average Probit AME:  0.219733143855364"
```

```r
hist(ame_df$ols, main = "Histogram of OLS AMEs", xlab = "AME")
```




```r
hist(ame_df$probit, main = "Histogram of Probit AMEs", xlab = "AME")
```



LPM and AMEs look to be similar. In terms of their distributions, they both look similar as well.

2b.

```
simualte <- function(n) {
    x <- rchisq(n, df = 1)
    u <- rnorm(n)
    y <- ifelse(1 + x + u > 0, 1, 0)
    return(data.frame(x, y))
}

ame_df <- data.frame(ols = numeric(), probit = numeric())

options(warn = -1)
```

```r
for (i in 1:1000) {
    data <- simualte(1000)
    ols <- lm(y ~ x, data = data)
    probit <- glm(y ~ x, data = data, family = binomial(link = "probit"))
    ame_df[i, "ols"] <- summary(ols)$coefficients[2, "Estimate"]
    ame_df[i, "probit"] <- summary(margins(probit))$AME
}

ols_average <- mean(ame_df$ols, na.rm = TRUE)
probit_average <- mean(ame_df$probit, na.rm = TRUE)

print(paste("Average OLS AME: ", ols_average))
```

```
## [1] "Average OLS AME:  0.0298830165933"
```

```r
print(paste("Average Probit AME: ", probit_average))
```

```
## [1] "Average Probit AME:  0.128152611166883"
```

The estimated AMEs from probit and LPM are very different in this case.

2c.

The LPM looks to be able to reliably approximate the true AMEs when data is generated from the probit model. This is likely because the LPM is a linear approximation of the probit model.