FTAP Homework 4

Zachary Fogelson
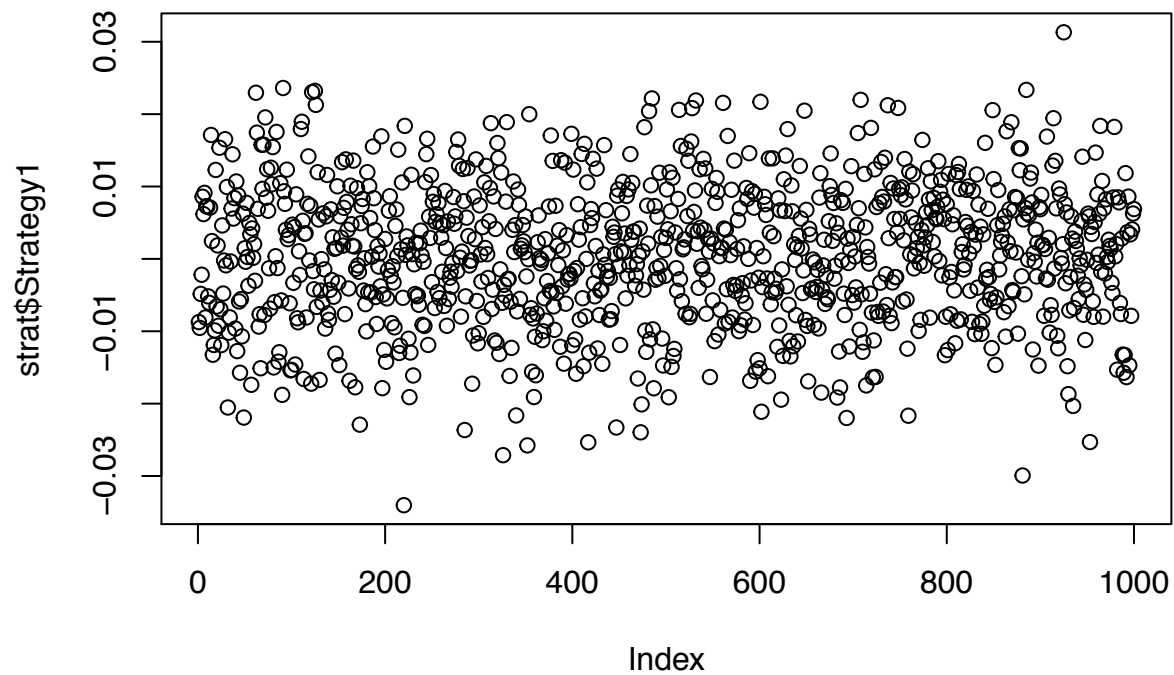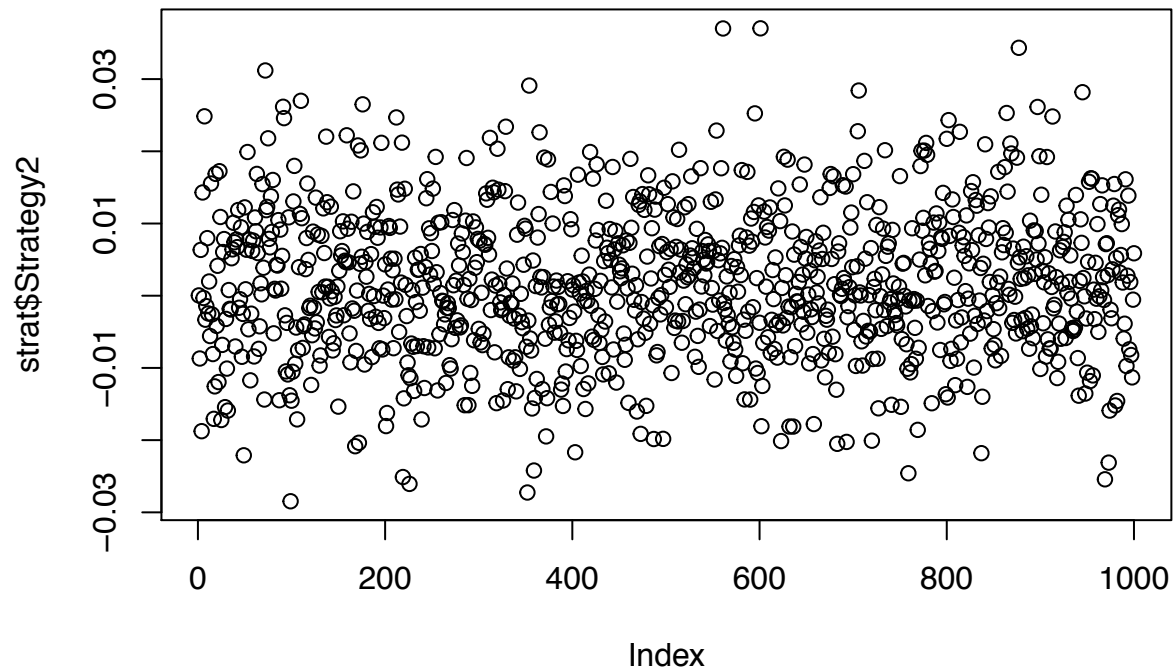
July 16, 2015

Problem 1

a

```
strat <- read.xls(xls = "strategies.xls")
plot(strat$Strategy1, main = "Plot 1")
```

Plot 1


```
plot(strat$Strategy2, main = "Plot 2")
```

Plot 2


```r
runs.test(strat$Strategy1)
```

```
##
##  Runs Test - Two sided
##
## data:  strat$Strategy1
## Standardized Runs Statistic = 1.7085, p-value = 0.08755
```

```r
runs.test(strat$Strategy2)
```

```
##
##  Runs Test - Two sided
##
## data:  strat$Strategy2
## Standardized Runs Statistic = -1.9616, p-value = 0.04981
```

Based on the runs.test, the strategy1 has a p-value of .09 and the p-value of strategy2 is .049. Therefore, at the 95% confidence level I would reject the hypothesis that Strategy2 is iid but I would fail to reject the hypothesis that Strategy1 is iid.

b

```r
jarque.bera.test(strat$Strategy1)
```

```
##
##  Jarque Bera Test
##
## data:  strat$Strategy1
## X-squared = 3.298, df = 2, p-value = 0.1922
```

```
shapiro.test(strat$Strategy1)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  strat$Strategy1
## W = 0.99825, p-value = 0.4055
```

```
jarque.bera.test(strat$Strategy2)
```

```
##
##  Jarque Bera Test
##
## data:  strat$Strategy2
## X-squared = 4.1271, df = 2, p-value = 0.127
```

```
shapiro.test(strat$Strategy2)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  strat$Strategy2
## W = 0.99799, p-value = 0.2774
```

Based on the Jarque Bera tests, I would fail to reject the null that either data set is not normal.
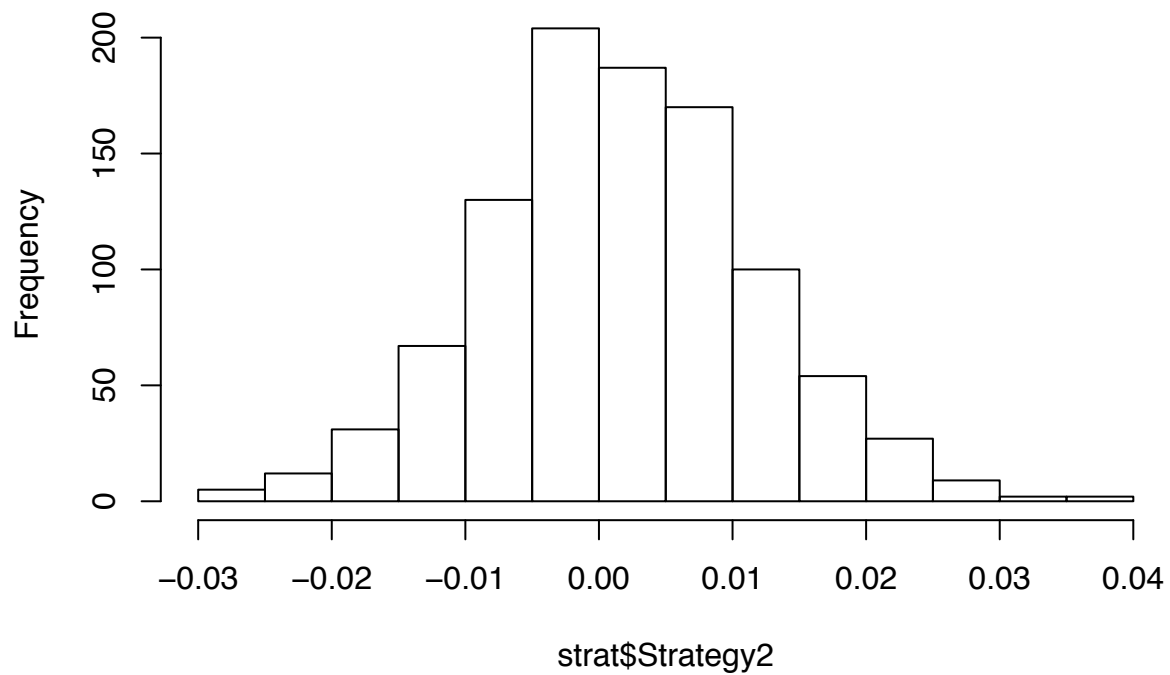
c

```
moms <- data.frame(matrix(all.moments(strat$Strategy1, order.max = 4)[-1], ncol=4, nrow=1))
colnames(moms) <- c("mean", "sd", "skew", "kurt")
moms <- rbind(moms, all.moments(strat$Strategy2, order.max = 4)[-1])
row.names(moms) <- c("Startegy1", "Startegy2")
hist(strat$Strategy1)
```

Histogram of strat$Strategy1
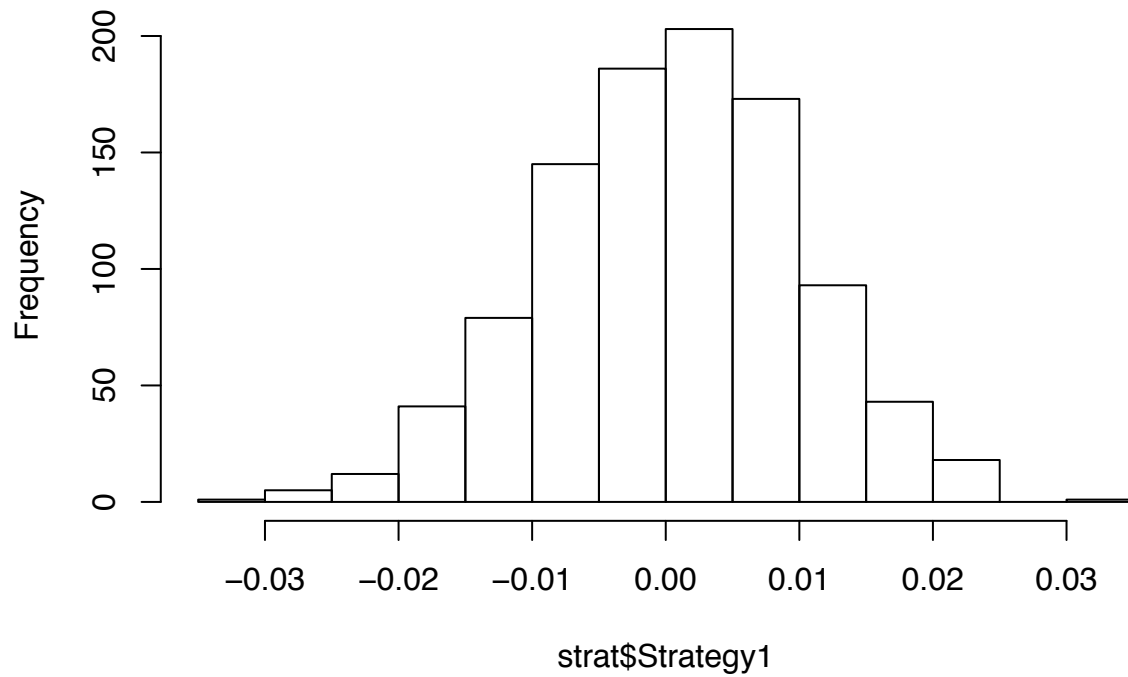
```
hist(strat$Strategy2)
```


Histogram of strat$Strategy2

```r
kable(head(moms))
```

	mean	sd	skew	kurt
Startegy1	0.0004736	0.0000923	0e+00	0
Startegy2	0.0016572	0.0001049	6e-07	0

For Strategy1 I would model it as the normal distribution $N(.0004, .00009)$ because although the Jacque Bera test said their was a 20% chance that it was normal and the Shapiro test says it had a 40% probability of coming from a normal distribution. Strategy2 has a much lower likelihood of being normal and appears to have some right hand skew. Potentially, I would use a squared normal to model it so that I capture the skew.

d

```r
cor(strat)
```

```
##           Strategy1 Strategy2
## Strategy1  1.000000  0.519215
## Strategy2  0.519215  1.000000
```

These strategies do appear to be correlated.

e

Null Hypothesis: $H_0 : \mu_1 - \mu_2 = 0$

Alternative: $H_1 : \mu_1 - \mu_2 \neq 0$

Two compare whether two samples have the same mean we can use the formula:

$$T = \frac{(\mu_1 - \mu_2) - 0}{\sqrt{(\frac{Var(X_1)}{N_1} + \frac{Var(X_2)}{N_2})}}$$

Because we have >30 observations for both selections assume that the normal distribution will give sufficiently accurate p-values. (This website explains what the degrees of freedom for a t distribution would be).

```r
x1 <- strat$Strategy1
x2 <- strat$Strategy2
t <- ((mean(x1) -  mean(x2)) - 0) / sqrt((var(x1)/length(x1) + var(x2)/length((x2))))
2 * pnorm(t)
```

```
## [1] 0.007262948
```

Because the p-value for is less than 5%, we are 95% confident that we can reject the null and accept the alternative hypothesis that the means are not equal.
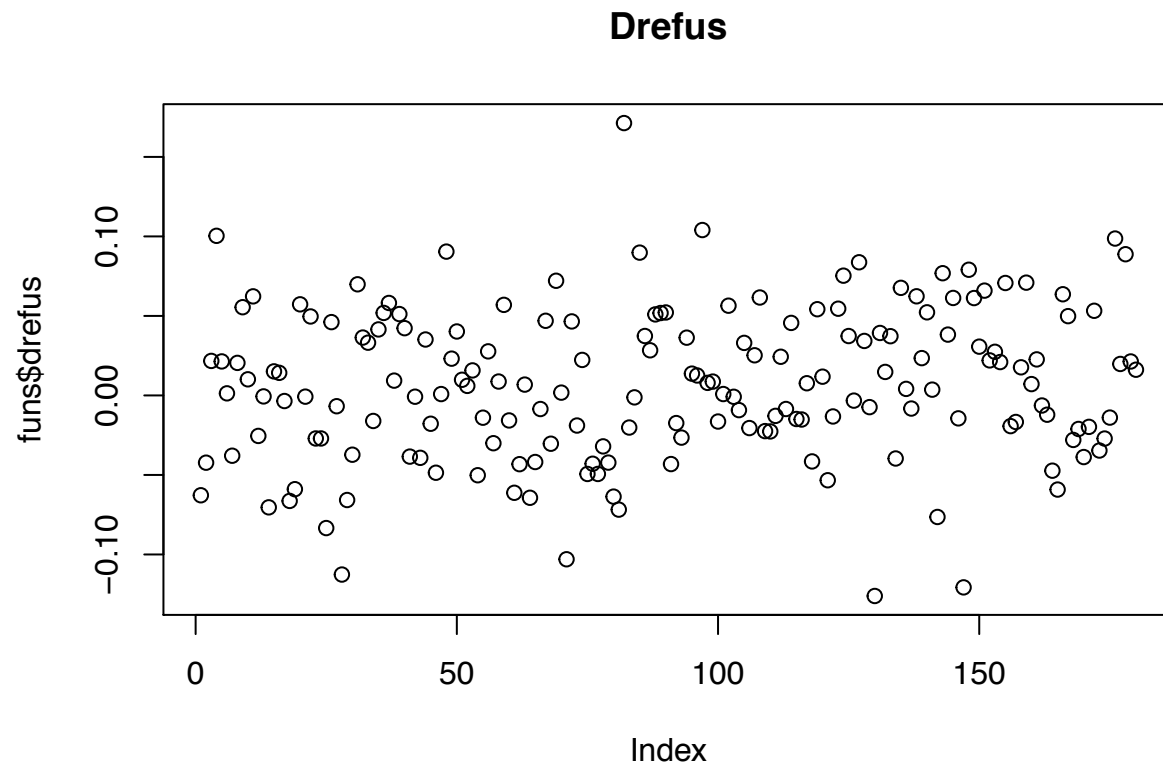
f

.0073.

g

Yes, I woudl still reject at the 1 percent level because, the p-value is less than .01.

h

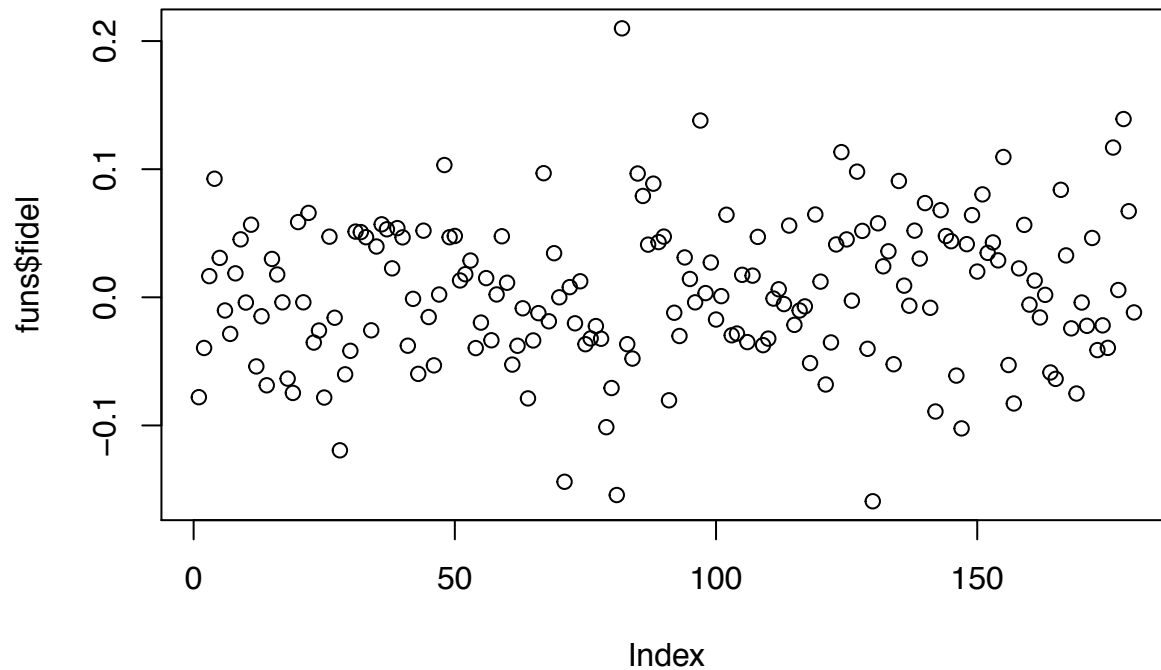The smallest value at which one could reject the null value is the value of the p-value itself.

Problem 2

```r
funs <- read.xls(xls = "mfunds (2).xls", sheet=2)
plot(funs$drefus, main = "Drefus")
```

Drefus


```r
plot(funs$fidel, main = "Fidel")
```

Fidel

a

```r
runs.test(funs$drefus)
```

```
## 
##  Runs Test - Two sided
## 
## data:  funs$drefus
## Standardized Runs Statistic = -1.9434, p-value = 0.05197
```

```r
runs.test(funs$fidel)
```

```
## 
##  Runs Test - Two sided
## 
## data:  funs$fidel
## Standardized Runs Statistic = -1.7939, p-value = 0.07283
```

b

```r
shapiro.test(funs$drefus)
```

```
## 
##  Shapiro-Wilk normality test
## 
## data:  funs$drefus
## W = 0.99296, p-value = 0.5385
```

```
shapiro.test(funs$fidel)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  funs$fidel
## W = 0.99082, p-value = 0.3051
```

c

Because the Shapiro-Wilk test would fail to reject the hypothesis that either of the funds are non-normal while the at a 90% confidence level we would reject that both of the funds are iid, I would conlude it is unlikely that an iid model would be good for modeling the returns of the two funds.

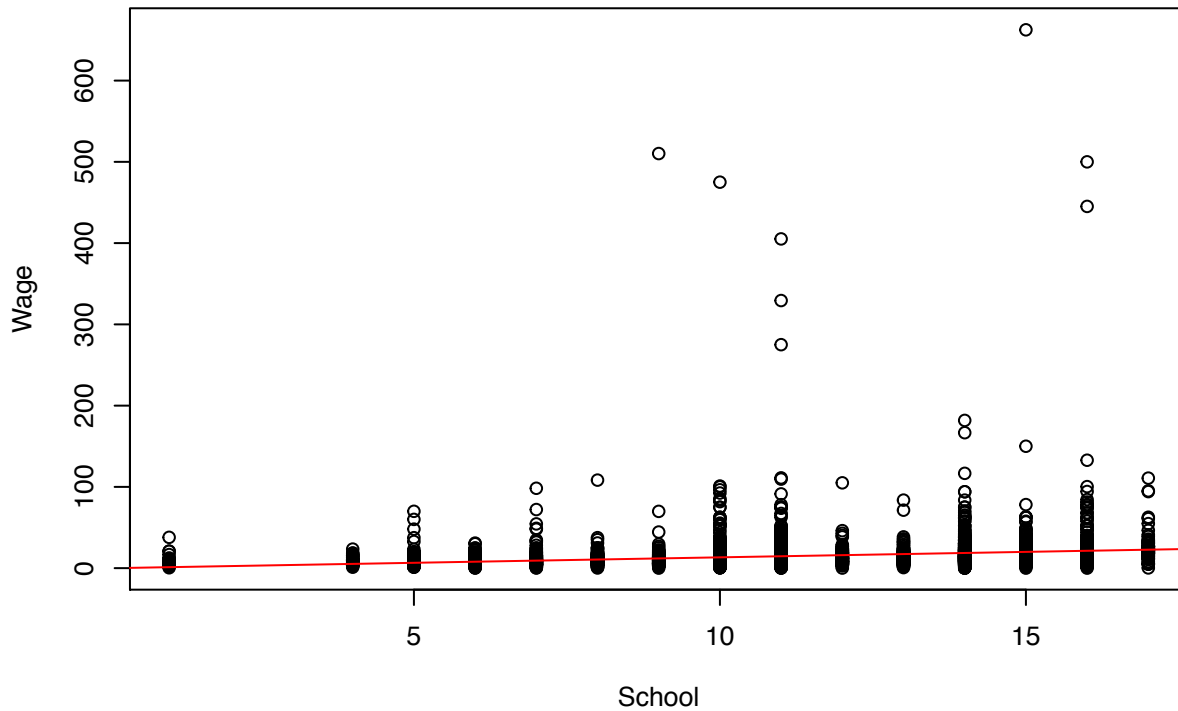**Problem 3**

a

$$E[Y|X=0]=2$$

b

$$E[Y|X=1]=6$$

c

$$Var[Y|X]=4$$

Problem 4

```
cen <- read.xls(xls = "censuswage.xls")
lm.out <- lm(Wage ~ School, cen)
par(cex=.8)
plot(cen, xlab("School"), ylab("Wage"))
abline(lm.out, col="red")
```

a

```r
summary(lm.out)
```

```
##
## Call:
## lm(formula = Wage ~ School, data = cen)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -22.44  -6.80  -2.70   2.33 642.48
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.1572     1.1588  -0.136    0.892
## School        1.3453     0.1017  13.231   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.76 on 5334 degrees of freedom
## Multiple R-squared:  0.03178,    Adjusted R-squared:  0.0316
## F-statistic: 175.1 on 1 and 5334 DF,  p-value: < 2.2e-16
```

Based on the summary of the linear model, with 0 years of education you will lose 15 cents an hour and for each additional year of education you will earn an additional ~$1.35.

b

```r
as.numeric(lm.out[[1]][1] + lm.out[[1]][2] *10)
```

```
## [1] 13.2955
```

c

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

$$\frac{SSR}{n-1} = Var(\hat{y})$$

$$\frac{SST}{n-1} = Var(y)$$

```r
yhat <- function(x){
  as.numeric(lm.out[[1]][1] + lm.out[[1]][2] * x)
}

SST <- var(cen$Wage) * (length(cen$Wage) - 1)
SSR <- var(sapply(cen$School, yhat)) * (length(cen$Wage) - 1)
SSE <- SST - SSR
Rsquared = SSR / SST
cat("SST: ", SST, "\nSSR: ", SSR, "\nSSE: ", SSE, "\nR^2:", Rsquared)
```

```
## SST:  2375098
## SSR:  75477.11
## SSE:  2299621
## R^2: 0.03177852
```

d

$$sd(\epsilon) = \sqrt{(\frac{SSE}{n-1})}$$

```r
sqrt(SSE/(length(cen$School)-1))
```

```
## [1] 20.76161
```

Problem 5

a

```r
stocks <- read.xls(xls = "capm.xls")
stocks <- rename(x = stocks, c("Risk.Free.Proxy..30.day.T.Bills"="tbills", "Market.Return.Proxy..Value.U

stocks$rfGE <- stocks$GE - stocks$tbills
stocks$rfMrkt <- stocks$mrkt - stocks$tbills

lm.out <- lm(rfGE ~ rfMrkt, stocks)

summary(lm.out)
```

```
##
## Call:
## lm(formula = rfGE ~ rfMrkt, data = stocks)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.106041 -0.028618 -0.001671  0.029691  0.111813
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.006354   0.002800   2.269   0.0242 *
## rfMrkt      1.088815   0.061666  17.657   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04283 on 238 degrees of freedom
## Multiple R-squared:  0.5671, Adjusted R-squared:  0.5653
## F-statistic: 311.8 on 1 and 238 DF,  p-value: < 2.2e-16
```

$\beta_1 = 1.088815$

b

For each dollar of market excess returns, GE has ~$1.09 of excess returns.

c

```
yhat(.01)
```

```
## [1] 0.0172418
```

d

```
covXY <- cov(stocks$rfGE, stocks$rfMrkt)
varX <- var(stocks$rfMrkt)
covXY / varX
```

```
## [1] 1.088815
```

e

```
mean(stocks$rfGE) - (covXY / varX) * mean(stocks$rfMrkt)
```

```
## [1] 0.006353651
```

Problem 6

a

```
z <- qnorm(.025, 0 , sd = 2)
ypred <- 2 + 4
cat("Lower: ", (ypred + z), "Upper: ", (ypred - z))
```

```
## Lower:  2.080072 Upper:  9.919928
```

b

To find the 95% confidence interval, we need the mean and standard deviation of the sum of two $\hat{Y}s$:

Mean:

Recall that we can substitute $\beta_0 + \beta_1 + \epsilon$ for a given y hat if X = 1.

$$E[\frac{\beta_0 + \beta_1 + \epsilon_1 + \beta_0 + \beta_1 + \epsilon_2}{2}]$$
$$E[\frac{2(\beta_0 + \beta_1)}{2} + \frac{\epsilon_1 + \epsilon_2}{2}]$$
$$E[\beta_0 + \beta_1] + 1/2 * (E[\epsilon_1] + E[\epsilon_2])]$$
$$6 + 1/2(0 + 0) = 6$$

Variance:

$$Var[\frac{\beta_0 + \beta_1 + \epsilon_1 + \beta_0 + \beta_1 + \epsilon_2}{2}]$$
$$Var[\frac{2(\beta_0 + \beta_1)}{2} + \frac{\epsilon_1 + \epsilon_2}{2}]$$

We know from class that the standard errors have no covariance.

$$Var[\beta_0 + \beta_1] + (1/2)^2 * (Var[\epsilon_1] + Var[\epsilon_2])$$

$$0 + 0 + 1/4 * 2 * 4 = 2$$

```r
ypred <- 2 + 4
cat("Lower: ", ypred - 1.96*sqrt(2), "Upper: ", ypred + 1.96*sqrt(2))
```

```
## Lower:  3.228141 Upper:  8.771859
```

c

```r
ypred <- 2 + 4*2
cat("Mean: ", ypred, "Var: ", 4)
```

```
## Mean:  10 Var:  4
```

d

Assuming that we have a large number of observations where X = 2:

```r
pnorm(8, mean = 10, sd = sqrt(4))
```

```
## [1] 0.1586553
```

e

The distrobution is $N(12, 4)$

f

$$Pr(Y_2 > Y_1) = Pr(\epsilon_2 - \epsilon_1 > 10 - 12)$$

$$Pr(\epsilon_2 - \epsilon_1 > -2)$$

$$E[\epsilon_2 - \epsilon_1] = 0$$

$$Var[\epsilon_2 - \epsilon_1] = 2Var[\epsilon] = 8$$

```
1 - pnorm(-2, 0, sqrt(8))
```

```
## [1] 0.7602499
```

Problem 7

```
cen <- read.xls(xls = "censuswage.xls")
lm.out <- lm(Wage ~ School, cen)
```

a

$$cov(\beta_0, \beta_1) = \frac{\sigma^2 \bar{x}}{(n-1)s_x^2}$$

```
sigmasquared = sum((lm.out$residuals^2))/length(lm.out$residuals)
sigmasquared * mean(cen$School) / ((length(cen$School)-1)*var(cen$School))
```

```
## [1] 0.1141807
```

b

Null Hypothesis: $H_0 : \beta_1 = 0$

Alternative: $H_1 : \beta_1 \neq 0$

$$\frac{b_j - B_j}{s_{b_j}}$$

```
sBj <- sqrt((1/(length(cen$School)-2)) * sum((lm.out$residuals^2))) / (sqrt(var(cen$School))*sqrt(lengt
t <- (as.numeric(lm.out$coefficients["School"]) - 0) / sBj
2*(1 - pt(t, length(cen$School)-2))
```

```
## [1] 0
```

We are 95% confident we can reject the null hypothesis. This tells us we are confident in the fact that there is a relationship between number of years in school and your wage.

c

Null Hypothesis: $H_0 : \beta_1 = 1$

Alternative: $H_1 : \beta_1 \neq 1$

```
t <- (as.numeric(lm.out$coefficients["School"]) - 1) / sBj
print(t)
```

```
## [1] 3.395897
```

```
2*(1 - pt(t, length(cen$School)-2))
```

```
## [1] 0.0006890356
```

We reject the null hypothesis that each year of school translates into exactly one more dollar of income per hour. However, the slope of the best fit line is close to 1.

d

The p-value for this test is .00068.

e

```
t <- qt(.025, length(cen$School) - 2)
cat("Lower: ", as.numeric(lm.out$coefficients["School"]) + t*sBj, "Upper: ", as.numeric(lm.out$coefficie
```

```
## Lower:  1.14595 Upper:  1.544589
```