

# Ec 122 - HW1

Ben Juarez

10/4/2021

\*\*\* NOTE: Extra Credit Video was watched. \*\*\*

## Mozart

- 1) The main question/problem is related to the relationship between music cognition and other higher brain functions.
- 2) The main result was that listening to Mozart before taking a particular test resulted in a higher mean SAS (implying higher IQs).
- 3) The method was such that 3 groups had 10 mins of music, relaxation, and silence (respectively) before performing the test which helped to determine IQ results after scoring.
- 4) I do believe these results mostly, but I think it is important to acknowledge that this study only examined one musical sample of one composer, so this effect may look different in varying situations.

## HGL 2.10

```
capm4 = read.table("dat/capm4.dat")
colnames(capm4) = c("date", "dis", "ge", "gm", "ibm", "msft", "xom", "mkt",
                    "riskfree")
```

## Part a

The economics model listed is a simple regression model like those discussed in the chapter because we can clearly see how it resembles the function  $E(y|x) = B_1 + B_2 \cdot x$  (with error) such that  $E(y|x) = (r_j - r_f)$ ,  $B_1 = a_j$ ,  $B_2 = B_j$ , and  $x = (r_m - r_f)$ .

Note that “B” corresponds to beta and “a” corresponds to alpha.

## Part b

We see that  $B_j = (r_j - r_f) / (r_m - r_f)$ .

```
companies = c("Disney", "GE", "GM", "IBM", "Microsoft", "Exxon")

disney = lm((capm4$dis - capm4$riskfree) ~ (capm4$mkt - capm4$riskfree))
disney_est = coef(disney)[2]

general_electric = lm((capm4$ge - capm4$riskfree) ~ (capm4$mkt - capm4$riskfree))
general_electric_est = coef(general_electric)[2]

general_motors = lm((capm4$gm - capm4$riskfree) ~ (capm4$mkt - capm4$riskfree))
general_motors_est = coef(general_motors)[2]
```

```

ibm = lm((capm4$ibm - capm4$riskfree) ~ (capm4$mkt - capm4$riskfree))
ibm_est = coef(ibm)[2]

microsoft = lm((capm4$msft - capm4$riskfree) ~ (capm4$mkt - capm4$riskfree))
microsoft_est = coef(microsoft)[2]

exxon = lm((capm4$xom - capm4$riskfree) ~ (capm4$mkt - capm4$riskfree))
exxon_est = coef(exxon)[2]

estimates = matrix(c(disney_est, general_electric_est, general_motors_est, ibm_est,
                     microsoft_est, exxon_est), dimnames = list(companies, "beta estimate"))

as.table(estimates)

##           beta estimate
## Disney      0.8943871
## GE          0.9012640
## GM          1.2667769
## IBM         1.1871798
## Microsoft   1.3183887
## Exxon       0.4126144

```

With the estimated beta values listed above, we can see that Microsoft appears the most aggressive with a beta estimate of 1.32, while Exxon appears to be the most defensive with a beta estimate of 0.41.

## Part c

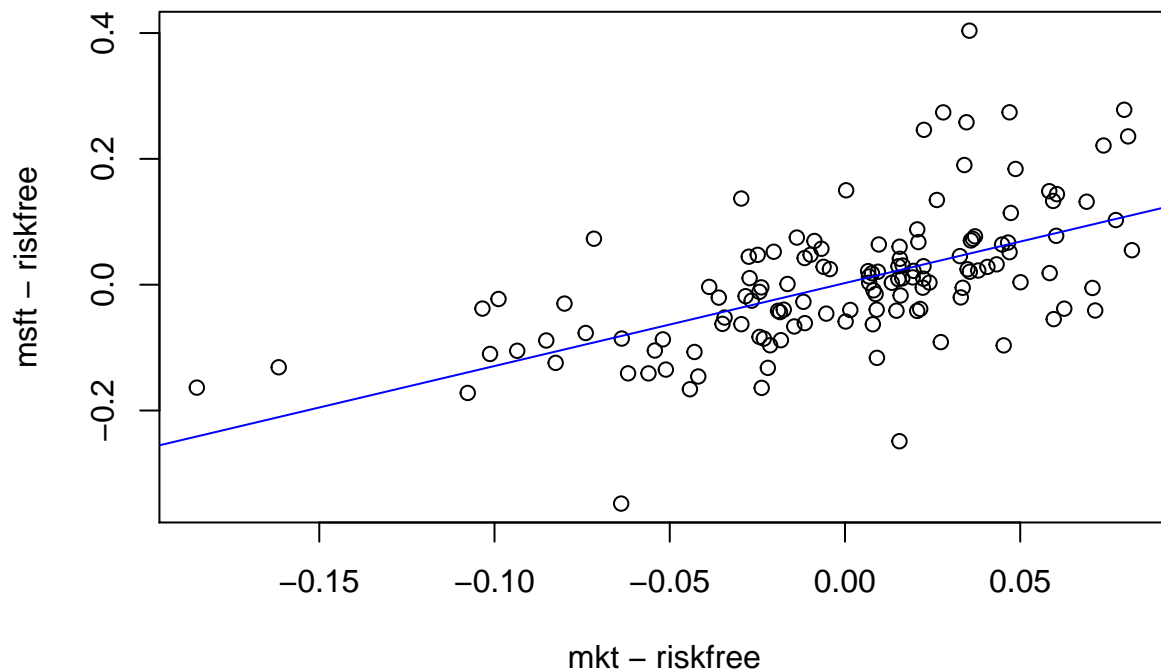
```

intercepts = matrix(c(coef(disney)[1], coef(general_electric)[1], coef(general_motors)[1],
                     coef(ibm)[1], coef(microsoft)[1], coef(exxon)[1]), dimnames =
                     list(companies, "alpha estimate"))
as.table(intercepts)

##           alpha estimate
## Disney      -0.003542738
## GE          -0.003577766
## GM          -0.014938161
## IBM         0.002675016
## Microsoft   0.002570329
## Exxon       0.006776051

plot((capm4$mkt - capm4$riskfree), (capm4$msft - capm4$riskfree), xlab = "mkt - riskfree",
     ylab = "msft - riskfree")
abline(lm((capm4$msft - capm4$riskfree) ~ (capm4$mkt - capm4$riskfree)), col = "blue")

```



This theory does seem correct given the alpha estimates since they are all very close to 0.

## Part d

```
new_estimates = matrix(c(
  coef(lm((capm4$dis - capm4$riskfree) ~ 0 + (capm4$mkt - capm4$riskfree))),
  coef(lm((capm4$ge - capm4$riskfree) ~ 0 + (capm4$mkt - capm4$riskfree))),
  coef(lm((capm4$gm - capm4$riskfree) ~ 0 + (capm4$mkt - capm4$riskfree))),
  coef(lm((capm4$ibm - capm4$riskfree) ~ 0 + (capm4$mkt - capm4$riskfree))),
  coef(lm((capm4$msft - capm4$riskfree) ~ 0 + (capm4$mkt - capm4$riskfree))),
  coef(lm((capm4$xom - capm4$riskfree) ~ 0 + (capm4$mkt - capm4$riskfree))),
  dimnames = list(companies, "beta estimate with alpha = 0"))
as.table(new_estimates)
```

```
##          beta estimate with alpha = 0
## Disney          0.8905834
## GE              0.8974226
## GM              1.2507383
## IBM             1.1900518
## Microsoft       1.3211484
## Exxon           0.4198896
```

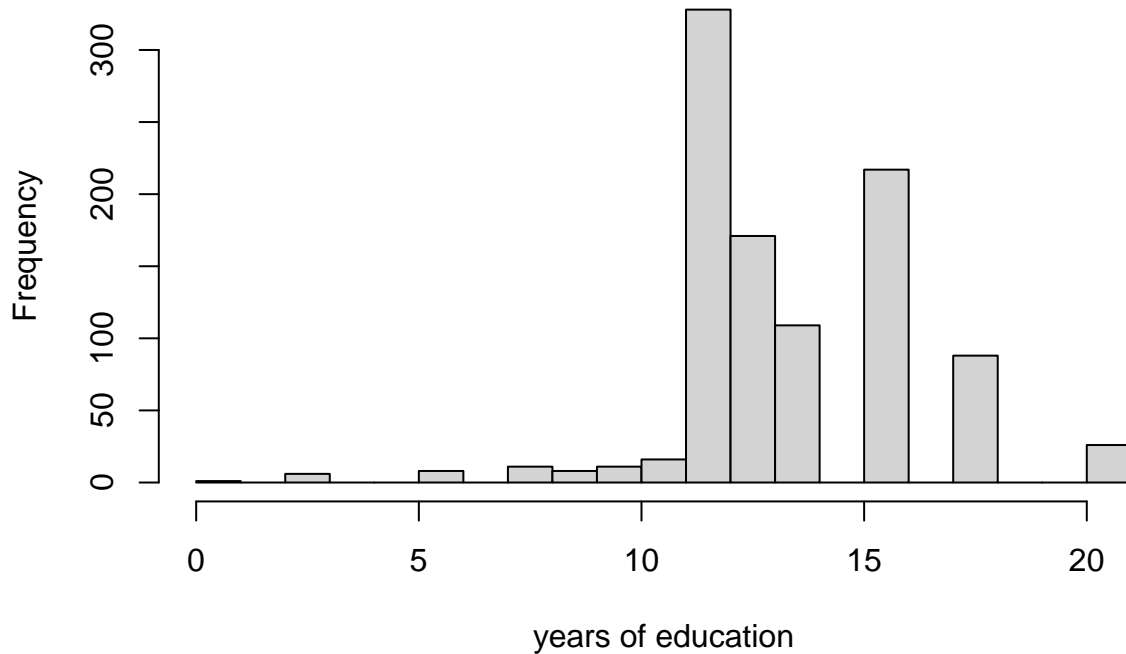
The new beta estimates under the condition that alpha (intercept) = 0 gives us very similar values. We conclude that the beta values do not change much.

## HGL 2.15

```
cps4_s = read.table("dat/cps4_small.dat")
colnames(cps4_s) = c("wage", "educ", "exper", "hrswk", "married", "female", "metro",
  "midwest", "south", "west", "black", "asian")
```

## Part a

```
labels = c("mean", "median", "max", "min", "sd")
educ_hist = hist(cps4_s$educ, breaks = 25, xlab = "years of education", main = "")
```

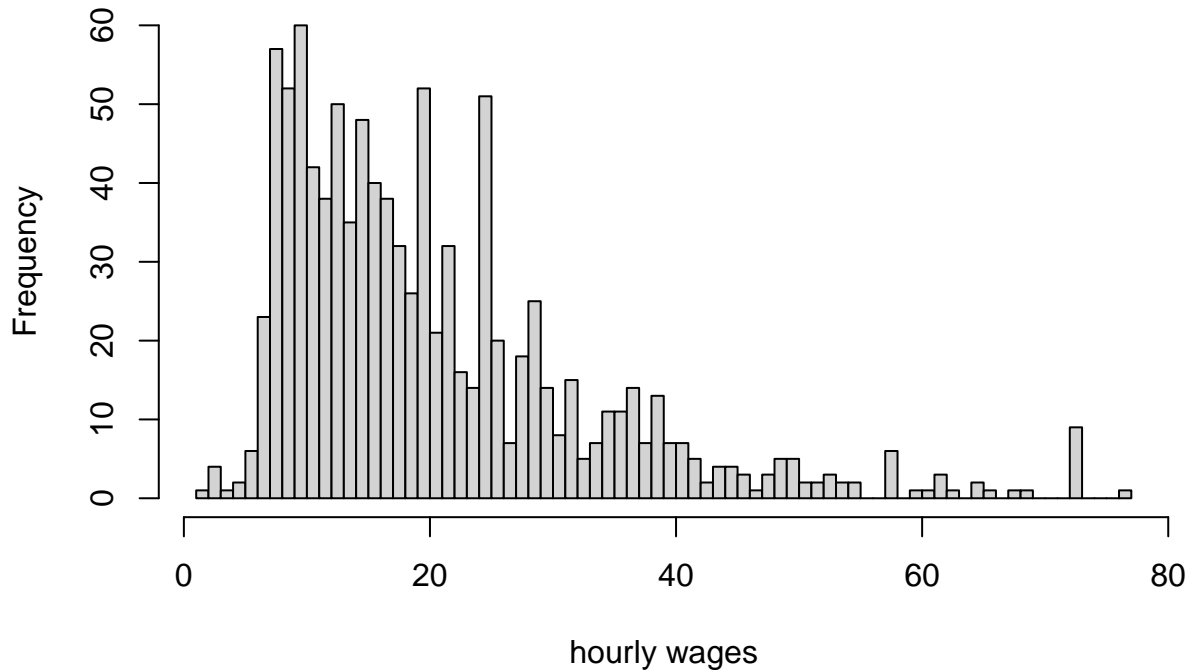


```
educ_stats = matrix(c(mean(cps4_s$educ), median(cps4_s$educ), max(cps4_s$educ), min(cps4_s$educ),
educ_stats
```

```
##
## mean  13.799000
## median 13.000000
## max    21.000000
## min     0.000000
## sd     2.711079
```

We can see that the majority of people had 12 years of education (graduated high school). The next majority is 16 years which implies graduation with college degree after 4 years. The subsequent frequencies at 18 and 21 years could represent master's degrees and PhD, respectively. Overall, these data characteristics makes sense.

```
wage_hist = hist(cps4_s$wage, breaks = 100, xlab = "hourly wages", main = "")
```



```
wage_stats = matrix(c(mean(cps4_s$wage), median(cps4_s$wage), max(cps4_s$wage), min(cps4_s$wage),
as.table(wage_stats)
```

```
##
## mean    20.61566
## median  17.30000
## max     76.39000
## min     1.97000
## sd      12.83472
```

Following the histogram and summary statistics, we can see that half of the people have an hourly wage greater than \$17.30. We can see that the majority of the wages fall between the 6 dollars/hr and 40 dollars/hr. The maximum is \$79.39 per hour and the minimum is \$1.97 per hour. Overall, these wages line up with expectations.

## Part b

```
wage_lm = lm((B_1 = cps4_s$wage) ~ (B_2 = cps4_s$educ))
coef(wage_lm)
```

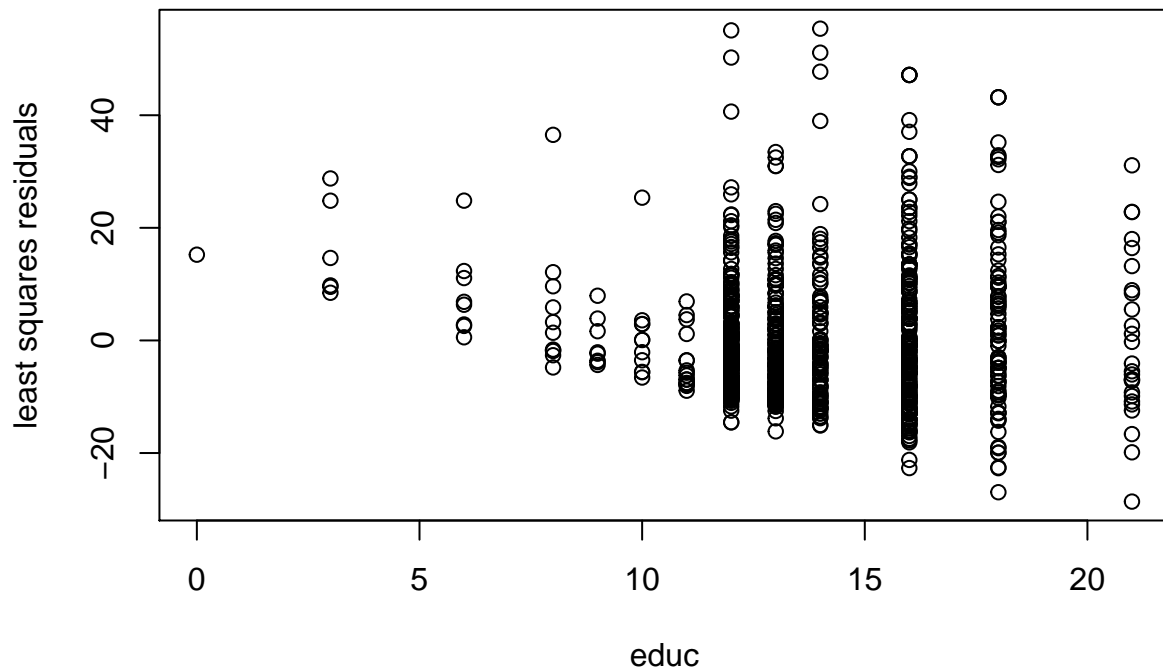
```
##          (Intercept) B_2 = cps4_s$educ
##          -6.710328      1.980288
```

We can see that estimating the linear regression gives us  $WAGE = -6.71 + 1.98(EDUC)$ .

From this, we can infer that each extra year of education implies an increase of \$1.98 in wages. The intercept tells us about the wage of someone with no years of education which is irrelevant in this case since it is negative and due to its logistical meaning.

## Part c

```
resid_plot = plot(cps4_s$educ, resid(wage_lm), xlab = "educ",
ylab = "least squares residuals")
```



We can see a pattern such that as educ increases, the residuals also increase which implies that the variance of the random error grows as educ grows. If SR1-SR5 hold, then we should not be able to notice patterns regarding the residuals.

## Part d

```
female_data = subset(cps4_s, cps4_s$female > 0)
male_data = subset(cps4_s, cps4_s$female == 0)
black_data = subset(cps4_s, cps4_s$black > 0)
white_data = subset(cps4_s, cps4_s$black == 0 & cps4_s$asian == 0)

female_lm = lm(female_data$wage ~ (B_2 = female_data$educ))
male_lm = lm(male_data$wage ~ (B_2 = male_data$educ))
black_lm = lm(black_data$wage ~ (B_2 = black_data$educ))
white_lm = lm(white_data$wage ~ (B_2 = white_data$educ))

regressions = cbind(
  c(coef(female_lm)[1], coef(male_lm)[1], coef(black_lm)[1], coef(white_lm)[1]),
  c(coef(female_lm)[2], coef(male_lm)[2], coef(black_lm)[2], coef(white_lm)[2]))
colnames(regressions) = c("B_1", "B_2")
rownames(regressions) = c("female", "male", "black", "white")

as.table(regressions)
```

	B_1	B_2
female	-14.168046	2.357519
male	-3.054487	1.875300
black	-15.085926	2.449058
white	-6.550655	1.991933

Following the table, we can see that with an extra year of education, a black worker has a higher increase in wages over a white worker while a female worker has a higher increase in wages over a male worker. Intercepts are irrelevant.

## Part e

```
educ2 = cps4_s$educ^2
quad = lm(cps4_s$wage ~ educ2)
quad_coeff = as.table(matrix(c(coef(quad)[1], coef(quad)[2])))
colnames(quad_coeff) = c("")
rownames(quad_coeff) = c("a_1", "a_2")
as.table(quad_coeff)
```

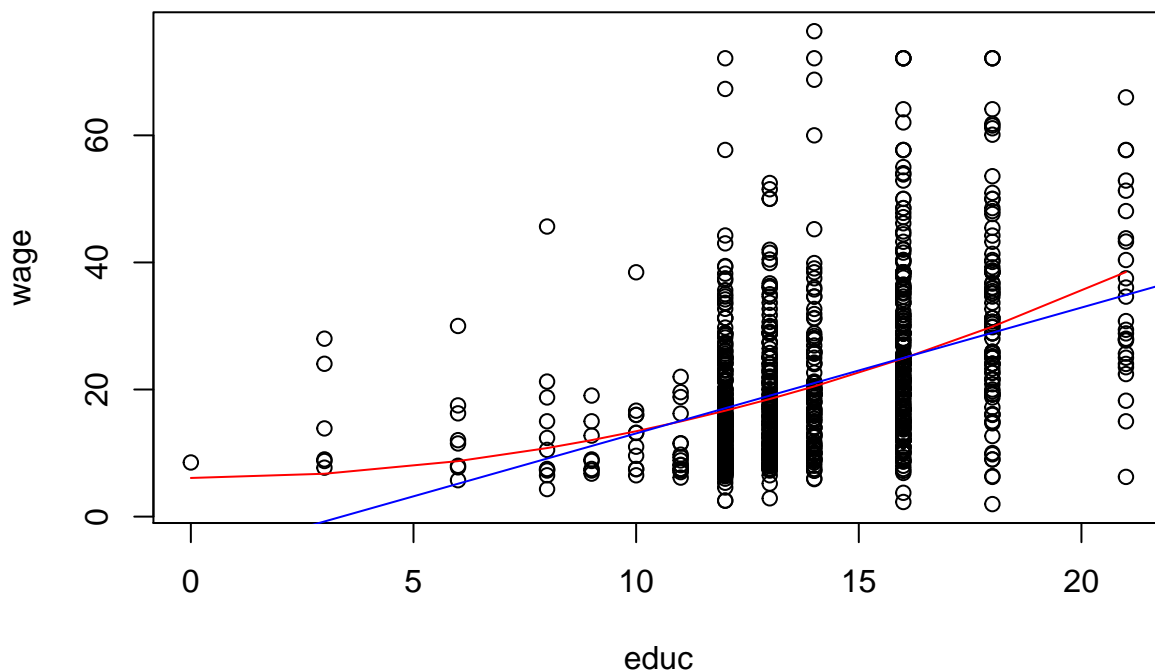
```
##
## a_1 6.08283101
## a_2 0.07348906
```

We can estimate the marginal effect of an additional 12 years by taking the derivative and multiplying by 12 which gives us  $2 * a_2 = 0.073 * 12 = 1.76$ . This means that for someone with 12 years of education, adding another year of education increases the wages by \$1.76.

Applying this process for 14 years gives us 2.06, for comparison. From part (b), we know that an extra year of education implies an increase of wages by \$1.98. This measurement does not take into account the existing years of education beforehand. So, with the quadratic model, we can infer that the marginal effect of an extra year of educ. increases with the years of education already achieved.

## Part f

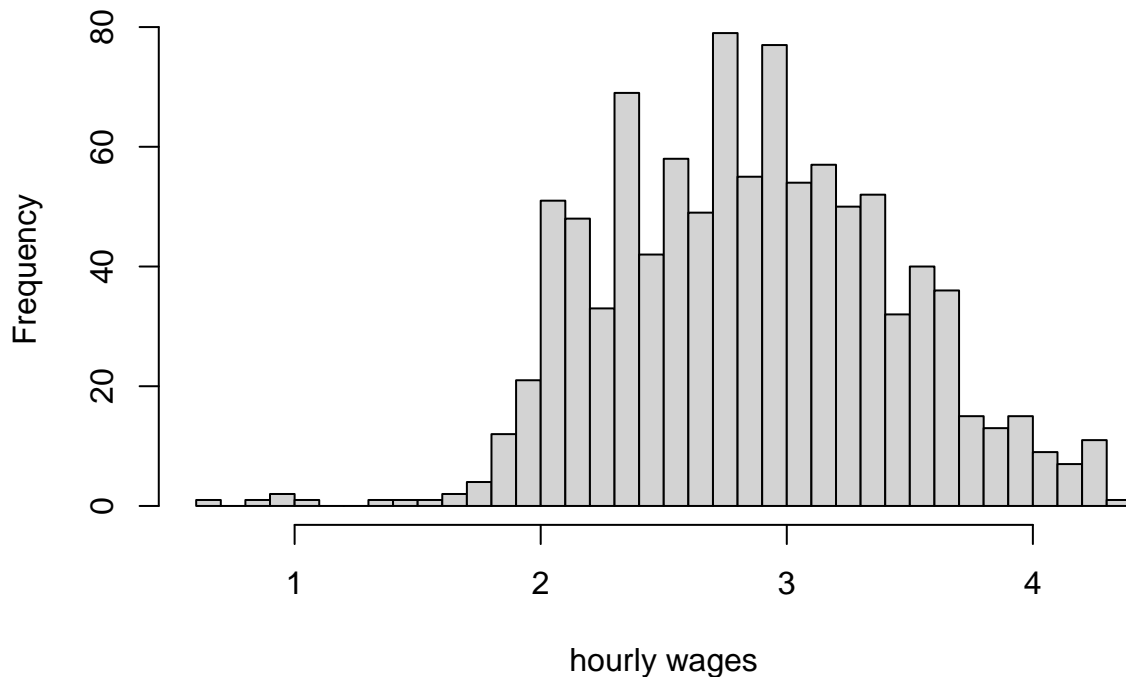
```
quad_function = coef(quad)[1] + coef(quad)[2] * cps4_s$educ^2
educ_v_wage = plot(cps4_s$educ, cps4_s$wage, xlab="educ", ylab="wage")
lines(sort(cps4_s$educ), quad_function[order(cps4_s$educ)], col = "red")
abline(wage_lm, col = "blue")
```



The quadratic model seems to be a better fit for the data.

## Part g

```
hist(log(cps4_s$wage), breaks = 50, main="", xlab="hourly wages")
```



Compared to the histogram from part(a), this distribution is much more bell-shaped and symmetrical.

## Part h

```
llr = lm(log(cps4_s$wage) ~ (y_2 = cps4_s$educ))
llr_coeff = as.table(matrix(c(coef(llr)[1], coef(llr)[2])))
colnames(llr_coeff) = c("")
rownames(llr_coeff) = c("y_1", "y_2")
as.table(llr_coeff)
```

```
##
## y_1 1.60944447
## y_2 0.09040825
```

So, our model is  $\ln(\text{WAGE}) = 1.609 + 0.0904(\text{EDUC})$ . We calculate the marginal effects as follows:

```
wage1 = exp(1.6094 + (0.0904 * 12)) # wage @ 12 years of education
wage2 = exp(1.6094 + (0.0904 * 14)) # wage @ 14 years of education

me1 = wage1 * coef(llr)[2] # marginal effect for 12 years of previous educ.
me2 = wage2 * coef(llr)[2] # marginal effect for 14 years of previous educ.

result = as.table(matrix(c(me1, me2)))
colnames(result) = c("Marginal Effect")
rownames(result) = c("12 yrs. prev. educ.", "14 yrs. prev. educ.")
result
```

```
##
## Marginal Effect
## 12 yrs. prev. educ. 1.337471
```



## 14 yrs. prev. educ.            1.602525

Comparing these values, we see that these marginal effects (log linear model) are lower than the marginal effects for both the linear (1.98) and quadratic models (1.76, 2.06).