

Chapter 5 - OLS Asymptotics

Thalles Quinaglia Liduares

29/03/2022

Exercise 5.5

Upload package

```
library(wooldridge)
library(lmreg)
library(dplyr)
```

Upload database

```
data<-wooldridge::htv

attach(data)
```

Consider the analysis in Computer Exercise C11 in Chapter 4 using the data in HTV.RAW, where `educ` is the dependent variable in a regression.

(i) How many different values are taken on by `educ` in the sample? Does `educ` have a continuous distribution?

```
distinct_educ<- dplyr::n_distinct(educ)

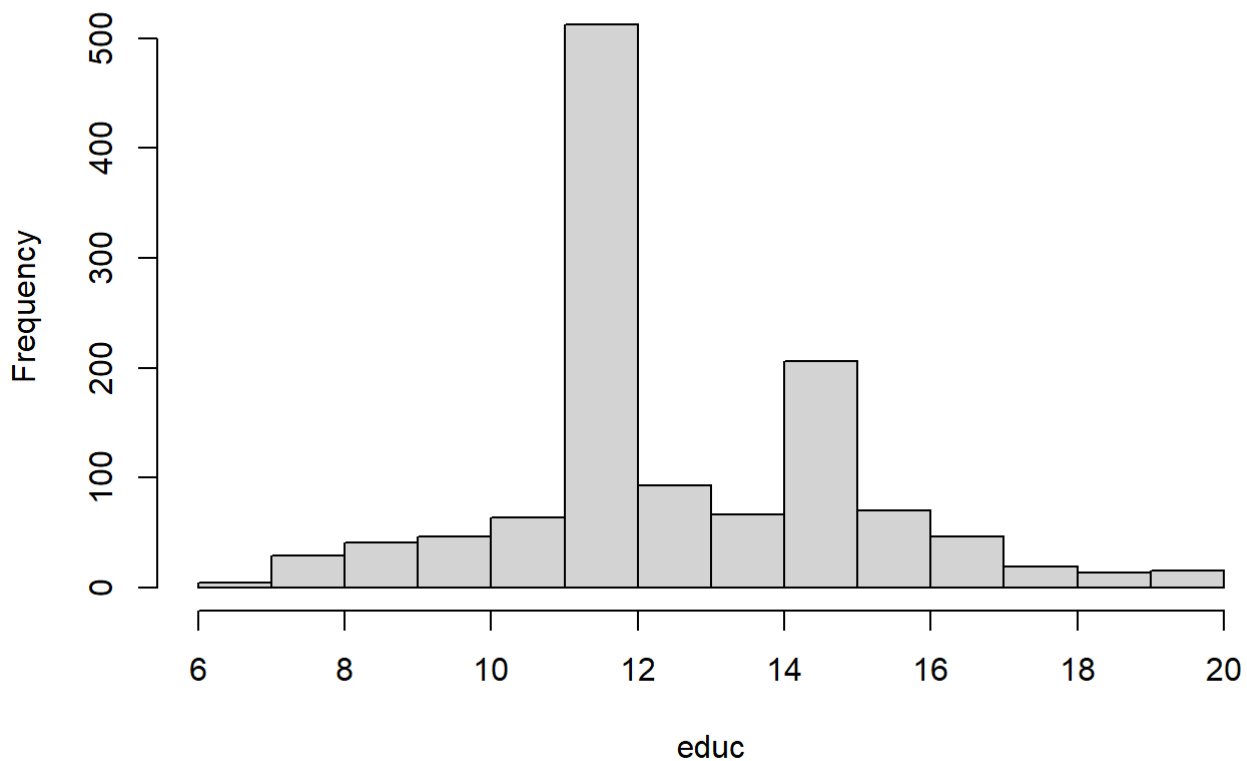
distinct_educ
```

```
## [1] 15
```

Distribution of education

```
hist(educ)
```

Histogram of educ



There are 15 different values from education in the sample. The distribution of `educ` is discrete.

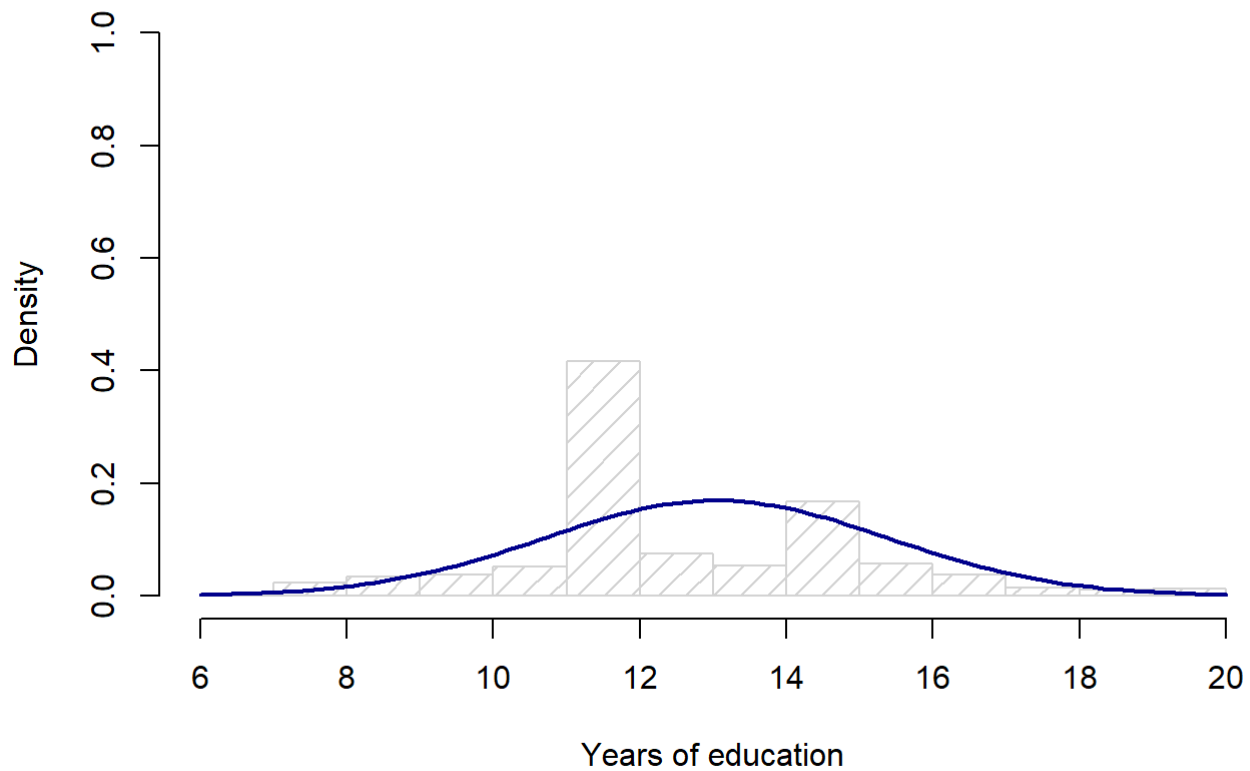
(ii) Plot a histogram of `educ` with a normal distribution overlay. Does the distribution of `educ` appear anything close to normal?

```
educ_mean<-round(mean(educ),2)

educ_sd<-round(sd(educ),2)

hist(educ, density=10, breaks=10, prob=TRUE,
     xlab="Years of education", ylim=c(0,1),
     main="normal curve over histogram")
curve(dnorm(x, mean=educ_mean, sd=educ_sd),
      col="darkblue", lwd=2, add=TRUE, yaxt="n")
```

normal curve over histogram



(iii) Which of the CLM assumptions seems clearly violated in the model

$$educ = \beta_0 + \beta_1 motheduc + \beta_2 fatheduc + \beta_3 abil + \beta_4 abil^2 + u$$

How does this violation change the statistical inference procedures carried out in Computer Exercise C11 in Chapter 4?

```
summary(lm1<-lm(educ~motheduc+fatheduc+abil+I(abil^2)))
```

```
##
## Call:
## lm(formula = educ ~ motheduc + fatheduc + abil + I(abil^2))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.2506 -1.1274 -0.1355  1.0223  7.0482
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.240226   0.287410  28.671 < 2e-16 ***
## motheduc     0.190126   0.028096   6.767 2.03e-11 ***
## fatheduc     0.108939   0.019601   5.558 3.35e-08 ***
## abil         0.401462   0.030288  13.255 < 2e-16 ***
## I(abil^2)    0.050599   0.008304   6.093 1.48e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.758 on 1225 degrees of freedom
## Multiple R-squared:  0.4444, Adjusted R-squared:  0.4425
## F-statistic: 244.9 on 4 and 1225 DF,  p-value: < 2.2e-16
```

In progress...