lecture 4.3 part 1

preliminaries

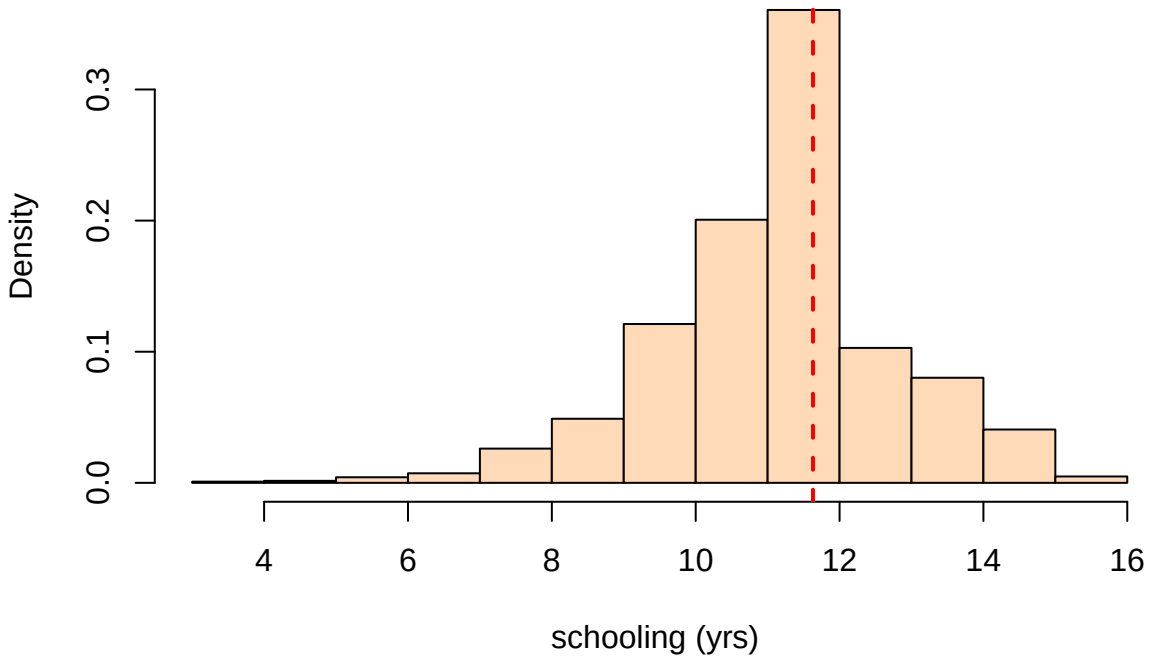
```
#clear workspace
rm(list=ls())

# load data
data = read.csv("~/Desktop/wages1.csv")

# build variables
wage = data$wage
school = data$school
exper = data$exper
```

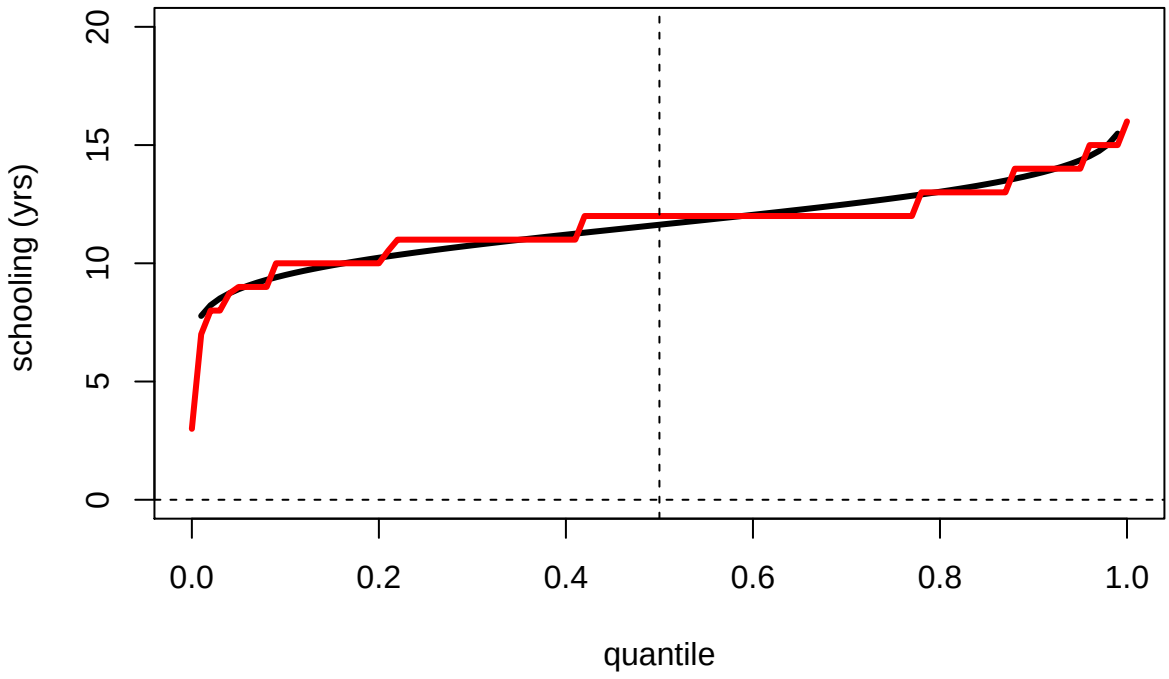
distribution of schooling

```
hist(school,
     breaks = 15,
     col="peachpuff",
     border="black",
     prob = TRUE, # show densities instead of frequencies
     xlab = "schooling (yrs) ",
     main = "")
abline(v=mean(school), lty=2, lwd=2, col="red")
```



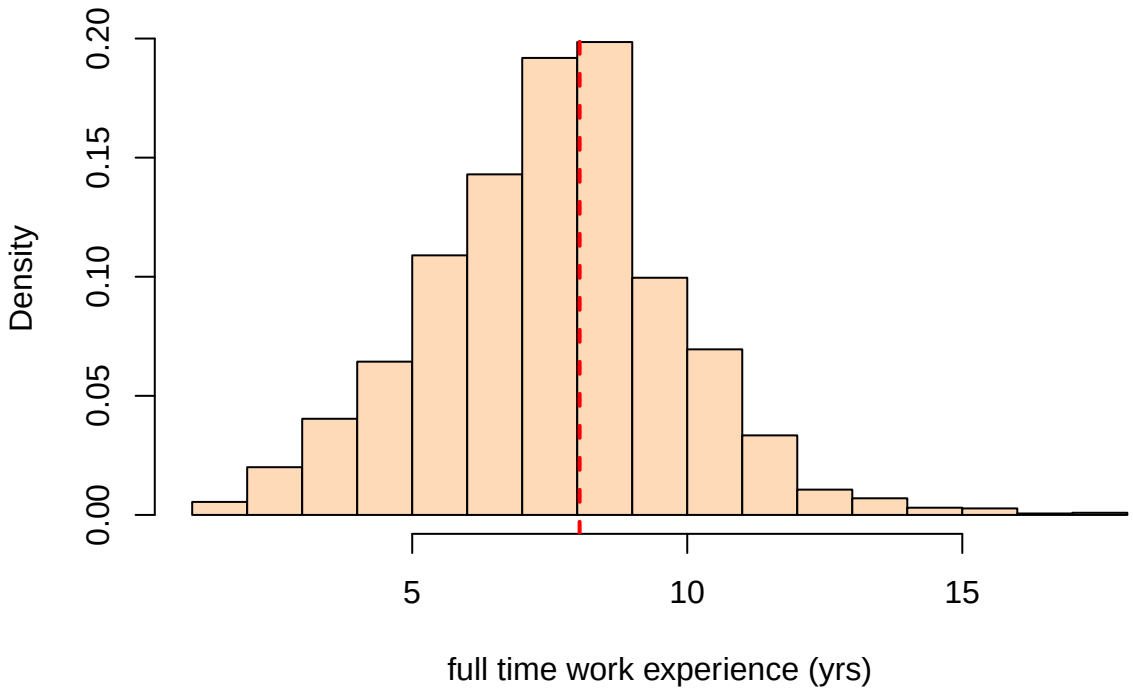

```
x = seq(0,1,by=0.01)
plot(x, qnorm(x, mean(school), sd(school)),
     type="l", ylim=c(-0,20), lwd = 3,
```

```
     xlab = "quantile", ylab = "schooling (yrs) ")
points(x, quantile(school, probs=x), type="l", col="red", lwd=3)
abline(v=0.5, lty=2)
abline(h=0, lty=2)
```

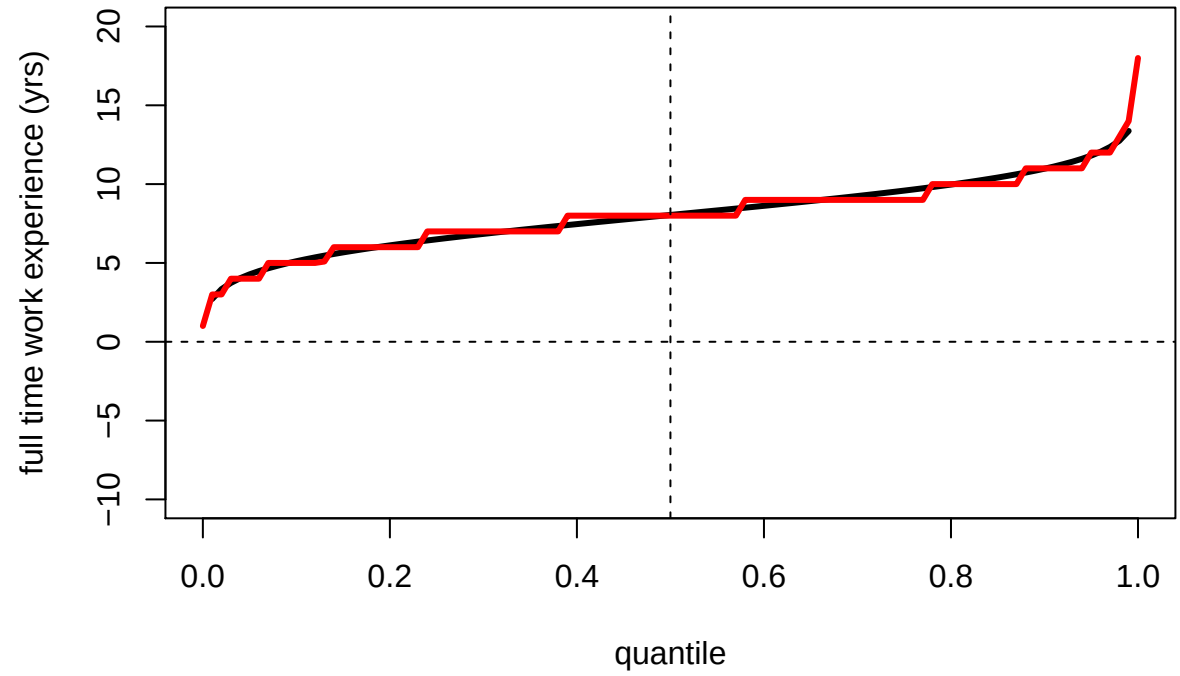


distribution of experience

```
hist(exper,
     breaks = 15,
     col="peachpuff",
     border="black",
     prob = TRUE, # show densities instead of frequencies
     xlab = "full time work experience (yrs)",
     main = "")
abline(v=mean(exper), lty=2, lwd=2, col="red")
```

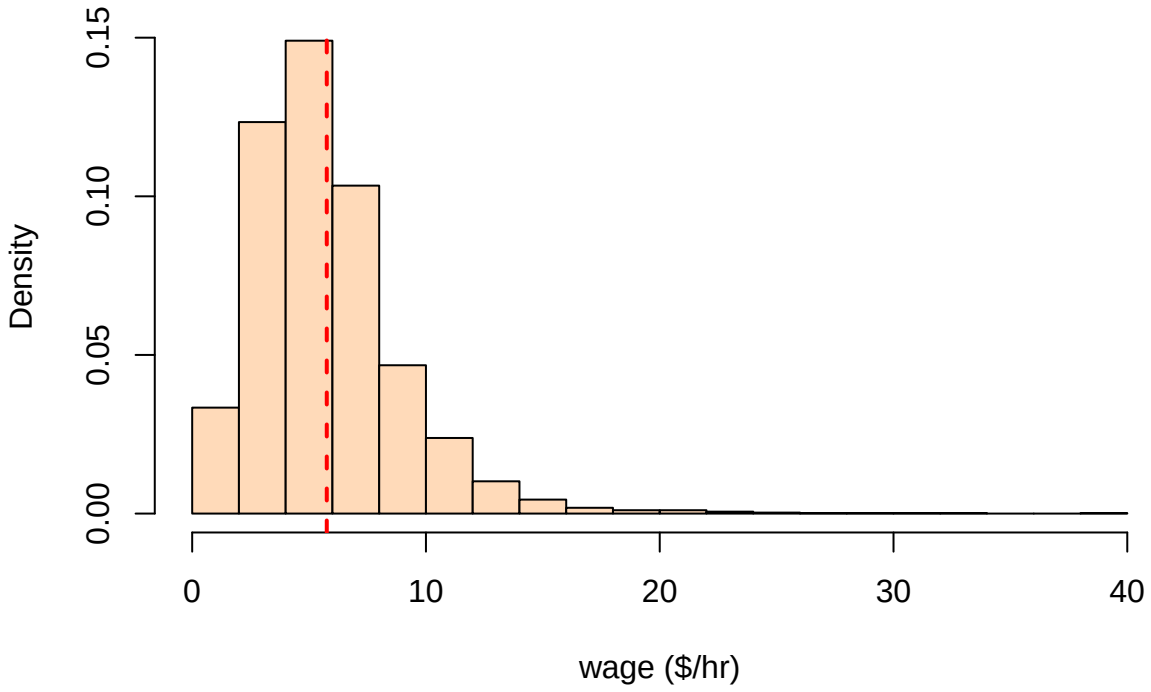



```
x = seq(0,1,by=0.01)
plot(x, qnorm(x, mean(exper), sd(exper)),
     type="l", ylim=c(-10,20), lwd = 3,
     xlab = "quantile", ylab = "full time work experience (yrs)")
points(x, quantile(exper, probs=x), type="l", col="red", lwd=3)
abline(v=0.5, lty=2)
abline(h=0, lty=2)
```

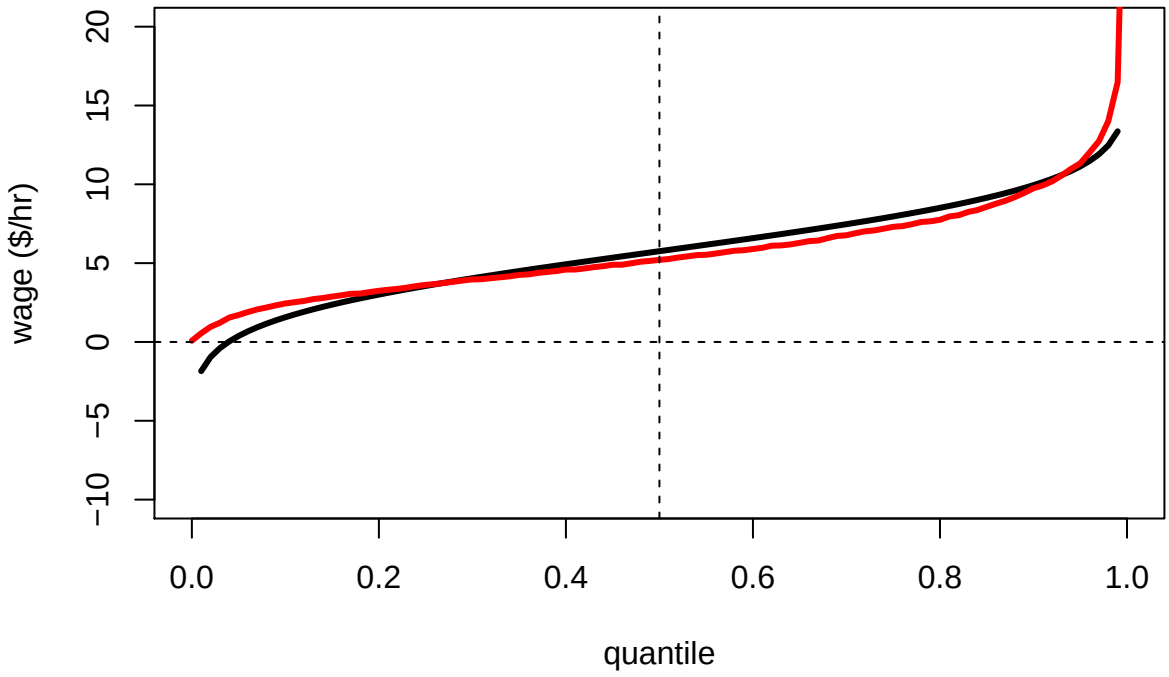


distribution of wages

```
hist(wage,
     breaks = 15,
     col="peachpuff",
     border="black",
     prob = TRUE, # show densities instead of frequencies
     xlab = "wage ($/hr)",
     main = "")
abline(v=mean(wage), lty=2, lwd=2, col="red")
```

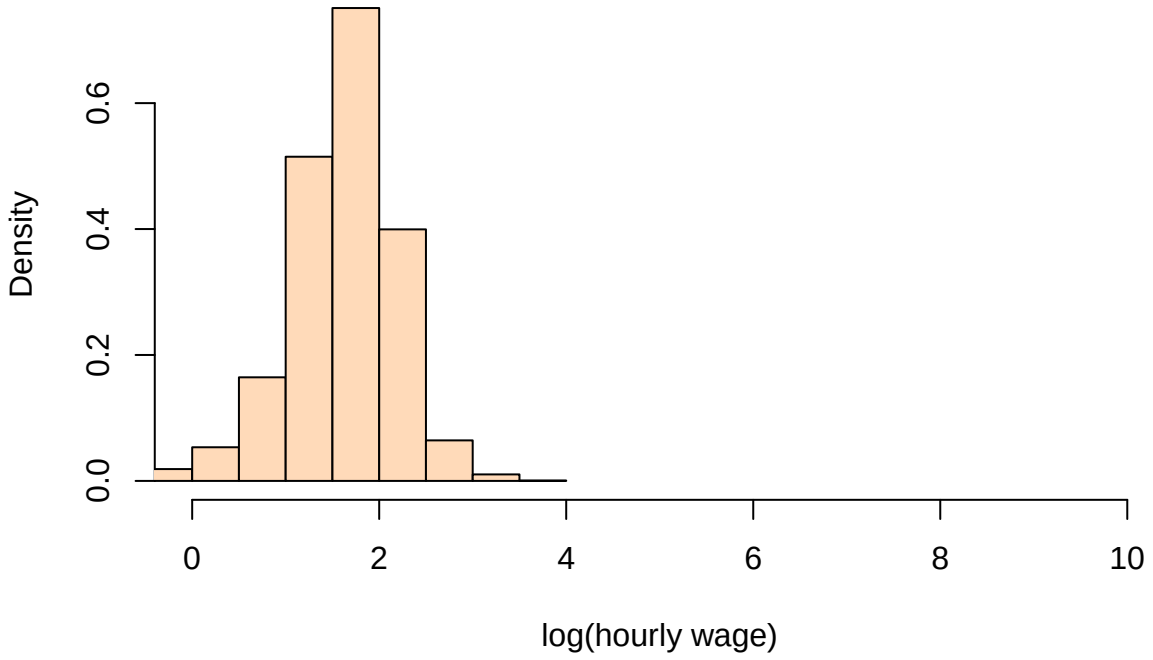



```
x = seq(0,1,by=0.01)
plot(x, qnorm(x, mean(wage), sd(wage)),
     type="l", ylim=c(-10,20), lwd = 3,
     xlab = "quantile", ylab = "wage ($/hr)")
points(x, quantile(wage, probs=x), type="l", col="red", lwd=3)
abline(v=0.5, lty=2)
abline(h=0, lty=2)
```

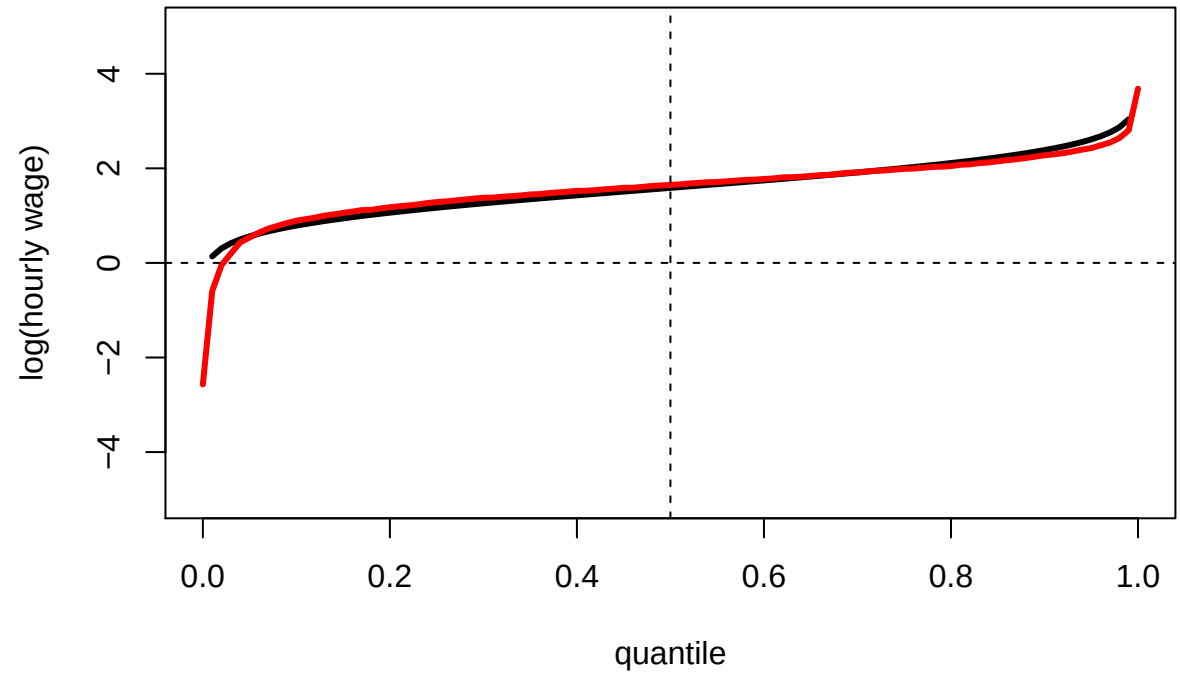


distribution of log(wages)

```
hist(log(wage),
     breaks = 15,
     col="peachpuff",
     border="black",
     prob = TRUE, # show densities instead of frequencies
     xlab = "log(hourly wage)",
     main = "",
     xlim = c(0,10))
```

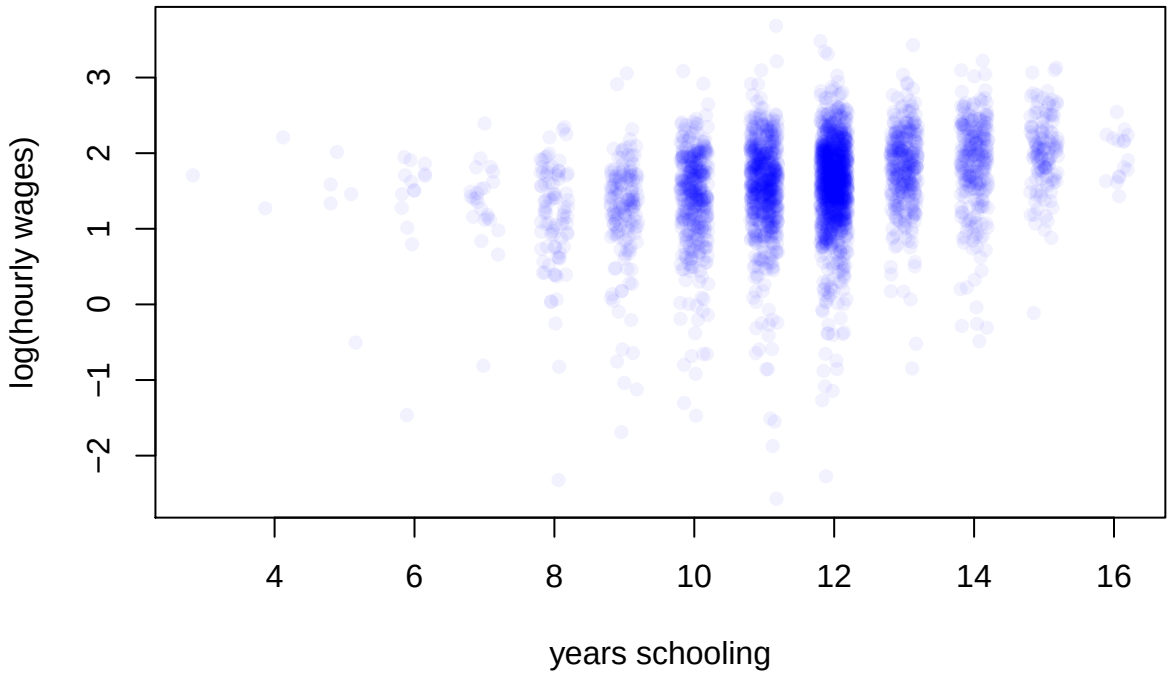



```
x = seq(0,1,by=0.01)
plot(x, qnorm(x, mean(log(wage)), sd(log(wage))),
     type="l", ylim=c(-5,5), lwd = 3,
     xlab = "quantile", ylab = "log(hourly wage)")
points(x, quantile(log(wage), probs=x), type="l", col="red", lwd=3)
abline(v=0.5, lty=2)
abline(h=0, lty=2)
```

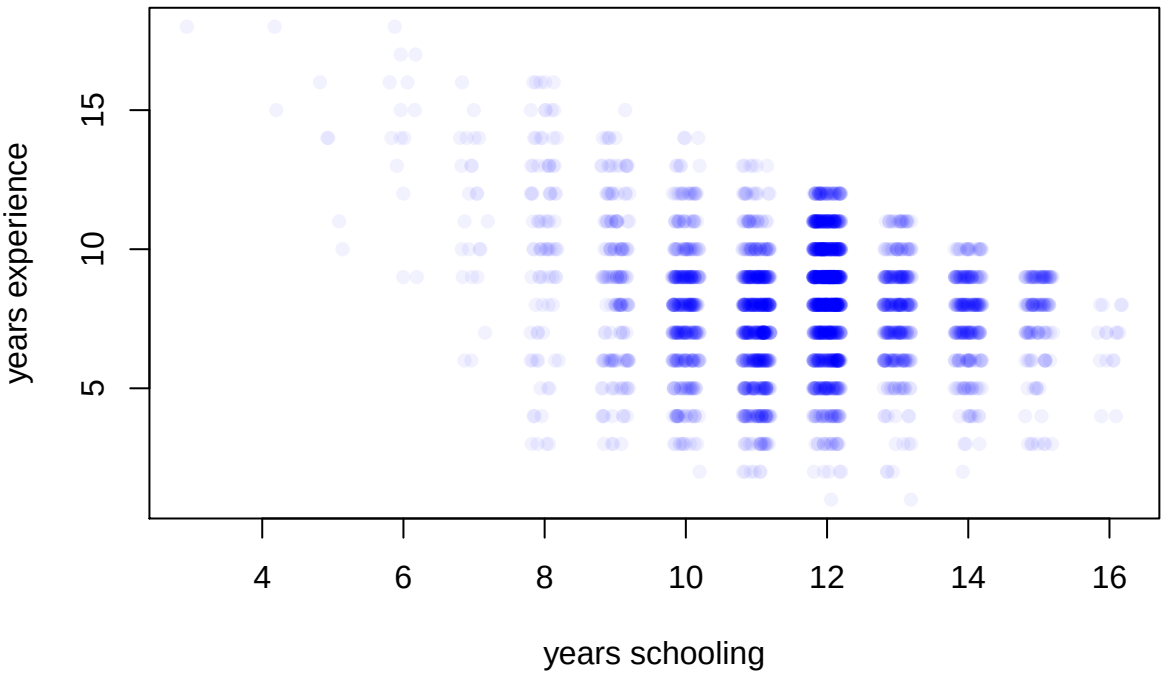


pairwise correlation plots

```
plot(jitter(school,1), log(wage),
     col=rgb(red=0.0, green=0.0, blue=1.0, alpha=0.05),
     pch = 16,
     xlab = "years schooling", ylab = "log(hourly wages)")
```




```
plot(jitter(school,1), exper,
     col=rgb(red=0.0, green=0.0, blue=1.0, alpha=0.05),
     pch = 16,
     xlab = "years schooling",  ylab = "years experience")
```


years experience
5
10
15
4
6
8
10
12
14
16
years schooling