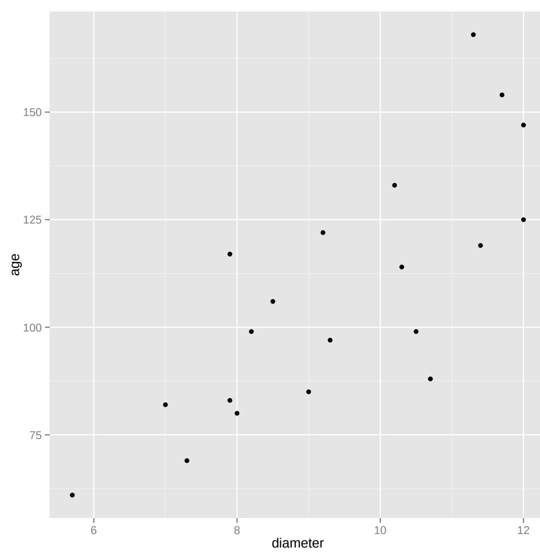
Mike Lederle
lederle@neo.tamu.edu
STAT 607 — HW 4

October 13, 2012

Exercise 4.3

Exercise 4.3 (a)

```
diameter <- c(12, 11.4, 7.9, 9, 10.5, 7.9, 7.3, 10.2, 11.7, 11.3, 5.7, 8, 10.3, 12, 9.2, 8.5,
    7, 10.7, 9.3, 8.2)
age <- c(125, 119, 83, 85, 99, 117, 69, 133, 154, 168, 61, 80, 114, 147, 122, 106, 82, 88,
    97, 99)
trees <- data.frame(diameter, age)
qplot(diameter, age, trees)
```



Exercise 4.3 (b)

```
x.bar.U <- 10.3
N <- 1132
(n <- nrow(trees))

## [1] 20
```

```r
(x.bar <- mean(diameter))

## [1] 9.405

(y.bar <- mean(age))

## [1] 107.4

(B.hat <- y.bar/x.bar)

## [1] 11.42

y.bar.hat.ratio <- B.hat * x.bar.U
paste("ESTIMATOR OF POP MEAN VIA RATIO ====", y.bar.hat.ratio)

## [1] "ESTIMATOR OF POP MEAN VIA RATIO ==== 117.620414673046"

(fpc <- 1 - n/N)

## [1] 0.9823

(term2 <- (x.bar.U/x.bar)^2)

## [1] 1.199

.resid <- function(x, y) {
    y.bar <- mean(y)
    x.bar <- mean(x)
    y - y.bar/x.bar * x
}
variance.resid <- function(r) {
    sum(r^2)/(length(r) - 1)
}
(term3 <- variance.resid(.resid(diameter, age))/n)

## [1] 16.1

SE.y.bar.hat.ratio <- sqrt(fpc * term2 * term3)
paste("STD ERR OF POP MEAN USING RATIO ====", SE.y.bar.hat.ratio)

## [1] "STD ERR OF POP MEAN USING RATIO ==== 4.35487189729141"
```

Exercise 4.3 (c)

```r
simple.reg <- function(x, y) {
    B.1.hat <- cor(x, y) * sd(y)/sd(x)
    B.0.hat <- mean(y) - B.1.hat * mean(x)
    list(B.0.hat, B.1.hat)
}
(trees.coef <- unlist(simple.reg(diameter, age)))

## [1] -7.808 12.250

y.bar.hat.reg <- trees.coef[1] + trees.coef[2] * x.bar.U
paste("ESTIMATOR OF POP MEAN VIA REGRESSION ====", y.bar.hat.reg)
```

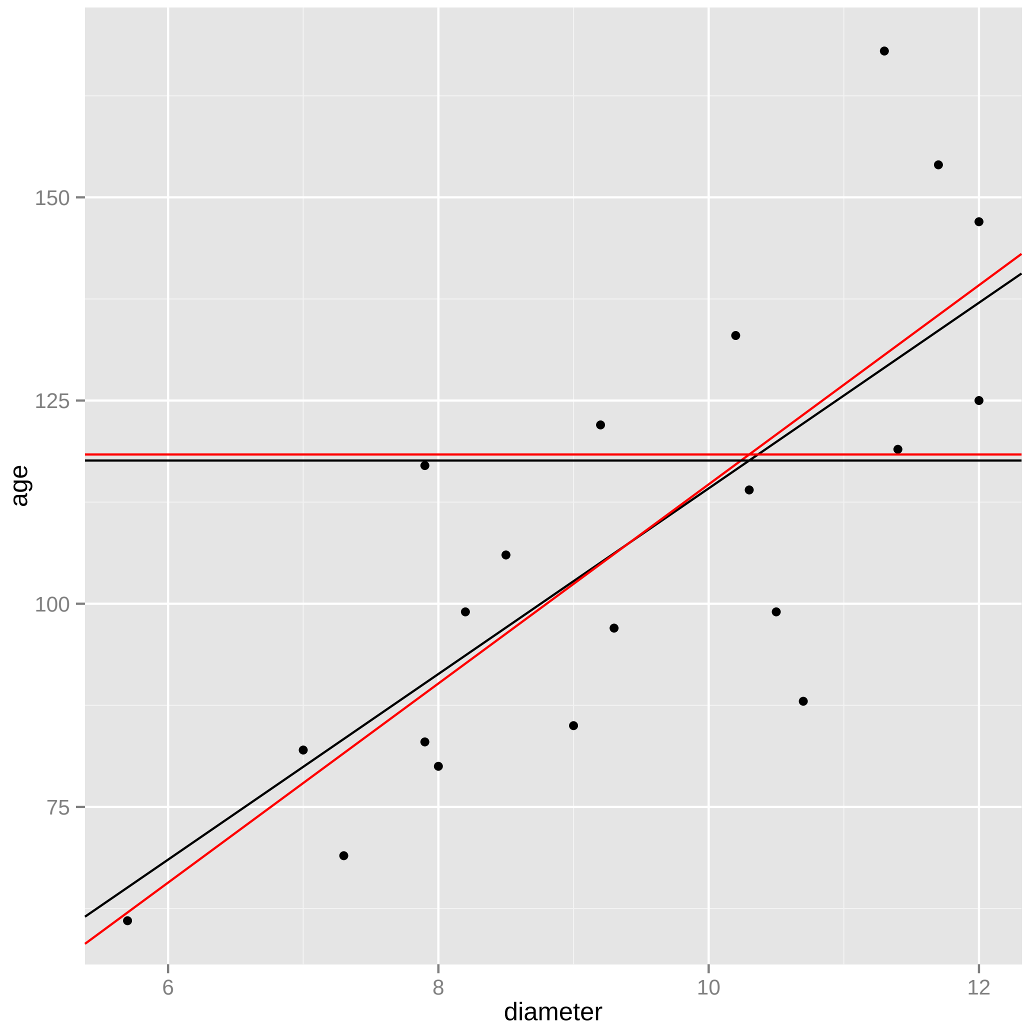
```
## [1] "ESTIMATOR OF POP MEAN VIA REGRESSION ==== 118.36344896498"

se.reg <- function(x, y, N) {
    fpc <- 1 - length(x)/N
    sqrt(fpc * 1/n * var(y) * (1 - cor(x, y)^2))
}
paste("STD ERR OF POP MEAN USING REGRESSION ====", se.reg(diameter, age, N))

## [1] "STD ERR OF POP MEAN USING REGRESSION ==== 3.96220011630061"
```

Exercise 4.3 (d)

```
p <- qplot(diameter, age, trees)
p + geom_abline(intercept = 0, slope = B.hat) + geom_abline(intercept = trees.coef[1], slope = trees.coe
    color = "red") + geom_hline(yintercept = y.bar.hat.ratio) + geom_hline(yintercept = y.bar.hat.reg,
    color = "red")
```



Seem pretty close to me. Regression estimate has lower SE.

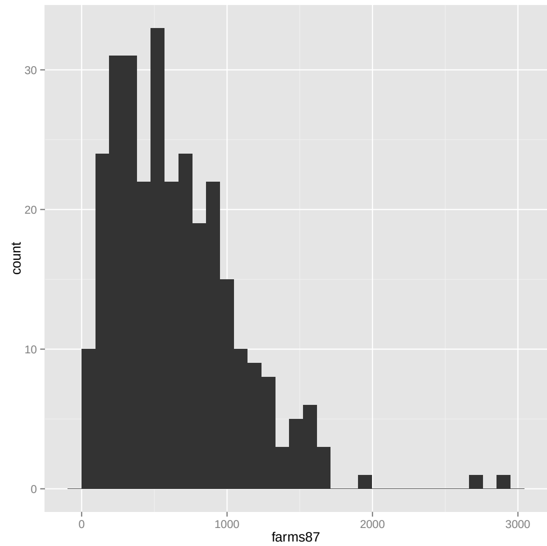
Exercise 4.8

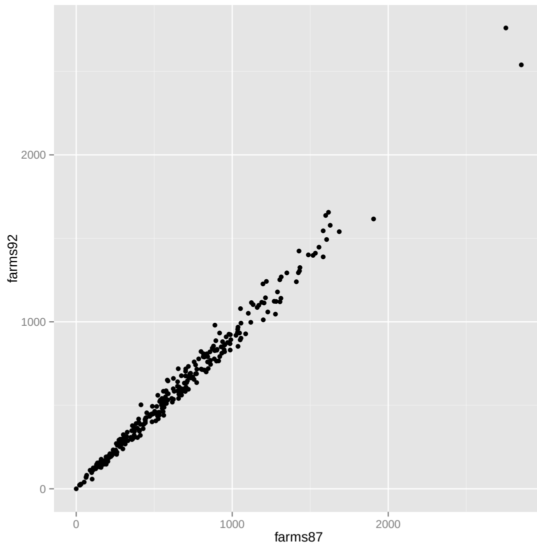
Exercise 4.8 (a)

```
agsrs <- read.csv("~/Courses/STAT 607/STAT-607/data/Dataset/agsrs.csv")
qplot(farms87, data = agsrs)

## stat_bin:  binwidth defaulted to range/30.  Use 'binwidth = x' to adjust this.

qplot(farms87, farms92, data = agsrs)
```





Exercise 4.8 (b)

```
t.x <- 2087759
n <- 300
N <- 3078
(x.bar <- mean(agsrs$farms87))

## [1] 647.7

(y.bar <- mean(agsrs$farms92))

## [1] 599.1

(B.hat <- y.bar/x.bar)

## [1] 0.9248

t.hat.y.ratio <- B.hat * t.x
paste("ESTIMATOR OF POP MEAN VIA RATIO ====", t.hat.y.ratio)

## [1] "ESTIMATOR OF POP MEAN VIA RATIO ==== 1930836.49967065"
```

Exercise 4.8 (c)

```r
(farms.coef <- with(agsrs, simple.reg(farms87, farms92)))

## [[1]]
## [1] -1.045
##
## [[2]]
## [1] 0.9265

(y.bar.hat.reg <- farms.coef[[1]] + farms.coef[[2]] * t.x/N)

## [1] 627.4

(t.hat.y.regression <- y.bar.hat.reg * N)

## [1] 1930988
```