Stat 138: Introduction to Sampling Designs
Problem Set 3

Anne Christine Amores

March 28, 2025

Average Age of Trees

Foresters want to estimate the average age of trees in a stand. Determining age is cumbersome, because one needs to count the tree rings on a core taken from the tree. In general, though, the older the tree, the larger the diameter, and diameter is easy to measure. The foresters measure the diameter of all 1132 trees and find that the population mean equals 10.3 They then randomly select 20 trees for age measurement.

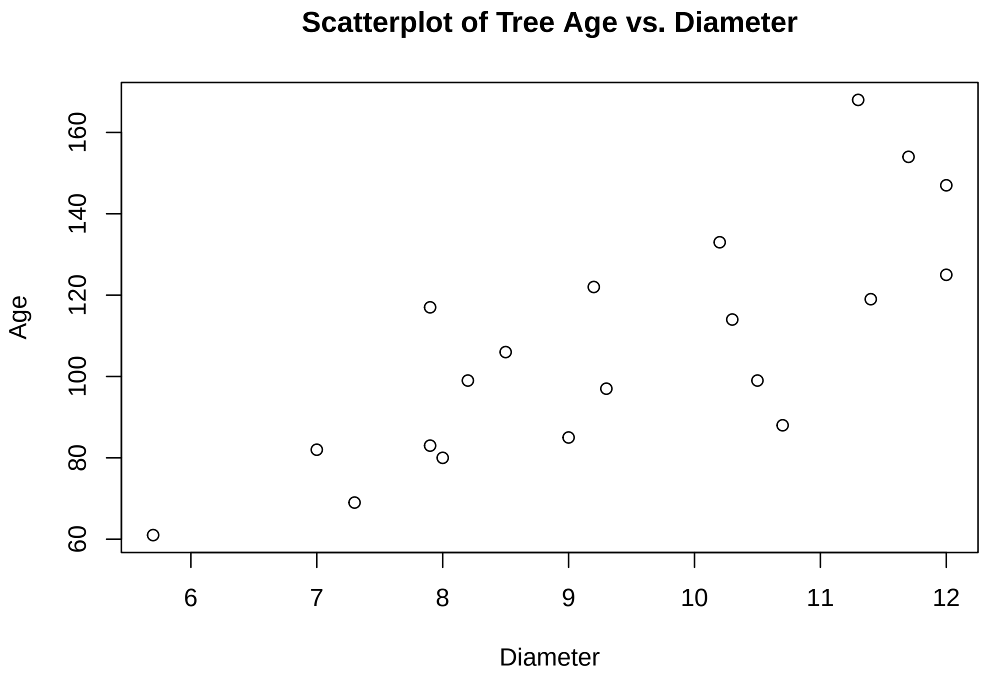
Tree No.	Diameter, x	Age, y	Tree No.	Diameter, x	Age, y
1	12.0	125	11	5.7	61
2	11.4	119	12	8.0	80
3	7.9	83	13	10.3	114
4	9.0	85	14	12.0	147
5	10.5	99	15	9.2	122
6	7.9	117	16	8.5	106
7	7.3	69	17	7.0	82
8	10.2	133	18	10.7	88
9	11.7	154	19	9.3	97
10	11.3	168	20	8.2	99

(a) **Draw a scatterplot of y vs. x.**

```
# Creating the dataset
diameter <- c(12.0, 11.4, 7.9, 9.0, 10.5, 7.9, 7.3, 10.2, 11.7, 11.3,
              5.7, 8.0, 10.3, 12.0, 9.2, 8.5, 7.0, 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)

# Creating scatterplot
plot(diameter, age, pch=1, xlab="Diameter", ylab="Age",
     main="Scatterplot of Tree Age vs. Diameter")
```



(b) **Estimate the population mean age of trees in the stand using ratio estimation and give an approximate standard error for your estimate.**
Given: $N = 1132$, $\bar{x}_u = 10.3$

Estimating $\bar{y}_u$ using ratio estimation,

$$\hat{\bar{y}}_r = \frac{\bar{y}}{\bar{x}} \cdot \bar{x}_u, \tag{1}$$

where $\bar{y}$ is the mean age of trees in the sample, and $\bar{x}$ is the mean diameter of trees in the sample.

Using R to compute $\bar{y}$ and $\bar{x}$,

```
# Computing the sample mean age of trees
ybar = mean(age)
cat("Mean age of trees in the sample:", ybar)
```

```
## Mean age of trees in the sample: 107.4
```

```
# Computing the sample mean diameter of trees
xbar= mean(diameter)
cat("Mean diameter of trees in the sample", xbar)
```

```
## Mean diameter of trees in the sample 9.405
```

Thus, $\bar{y} = 107.4$ and $\bar{x} = 9.405$. Plugging the values computed for $\bar{y}$ and $\bar{x}$ into (1),

$$\hat{\bar{y}}_r = \frac{107.4}{9.405} \cdot 10.3$$

$$\rightarrow \hat{\bar{y}}_r = \boxed{117.6204}$$

Thus, the **estimated population mean age of trees in the stand using ratio** is 117.6204

Now, getting an approximate standard error for our estimate,

$$SE(\hat{\bar{y}}_r) = \sqrt{\hat{Var}(\hat{\bar{y}}_r)}$$

$$\rightarrow SE(\hat{\bar{y}}_r) \approx \sqrt{\left(1 - \frac{n}{N}\right)\left(\frac{\bar{x}_u}{\bar{x}}\right)^2 \frac{s_e^2}{n}}, \tag{2}$$

where $e_i = y_i - \hat{B}x_i$ and $s_e^2 = \frac{1}{n-1}\sum_{i \in S} e_i^2$.

The only value we are missing for the computation of the standard error is s_e^2. Using R to compute for s_e^2:

```
n = 20

# Computing B_hat
B_hat <- ybar / xbar

# Computing residuals
e <- age - B_hat * diameter

# Computing squared residuals
sq_e <- e^2

# Computing sample variance of residuals
var_e <- sum(sq_e) / (n - 1)

cat("Sample variance of residuals:", var_e)
```

```
## Sample variance of residuals: 321.933
```

Thus, $s_e^2 = 321.933$. Plugging this into (2),

$$\rightarrow SE(\hat{\bar{y}}_r) \approx \sqrt{\left(1 - \frac{20}{1132}\right)\left(\frac{10.3}{9.405}\right)^2 \frac{321.933}{20}}$$

$$\rightarrow SE(\hat{\bar{y}}_r) \approx 4.354872$$

$$\rightarrow SE(\hat{\bar{y}}_r) \approx \boxed{4.3549}$$

(c) **Repeat (b) using regression estimation.**

The regression estimator for $\bar{y}_u$ is given by

$$\hat{\bar{y}}_{reg} = \bar{y} + \hat{B}_1(\bar{x}_u - \bar{x}), \tag{3}$$

where

$$\hat{B}_1 = \frac{\sum_{i \in S}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i \in S}(x_i - \bar{x})^2}.$$

Using R to compute for $\hat{B}_1$,

```
# Computing for B_1_hat
xdev <- diameter - xbar
ydev <- age - ybar
numerator <- sum(xdev * ydev)
sq.xdev <- xdev^2
denominator <- sum(sq.xdev)
B_1_hat <- numerator / denominator
cat("B_1_hat =", B_1_hat)
```

```
## B_1_hat = 12.24966
```

Thus, $\hat{B}_1 = 12.24966$. Plugging this into (3),

$$\hat{\bar{y}}_{reg} = 107.4 + 12.24966(10.3 - 9.405)$$

$$\to \hat{\bar{y}}_{reg} = 118.3634457$$

$$\to \hat{\bar{y}}_{reg} \approx \boxed{118.3634}$$

Now, getting an approximate standard error for our estimate,

$$SE(\hat{\bar{y}}_{reg}) = \sqrt{\hat{Var}(\hat{\bar{y}}_{reg})}$$

$$\to SE(\hat{\bar{y}}_{reg}) \approx \sqrt{\left(1 - \frac{n}{N}\right)\frac{s_e^2}{n}}, \tag{4}$$

Plugging in the s_e^2 that was computed in the previous item into (4),

$$\to SE(\hat{\bar{y}}_{reg}) \approx \sqrt{\left(1 - \frac{20}{1132}\right)\frac{321.933}{20}}$$

$$\to SE(\hat{\bar{y}}_{reg}) \approx 3.976462862$$

$$\to SE(\hat{\bar{y}}_{reg}) \approx \boxed{3.9765}$$

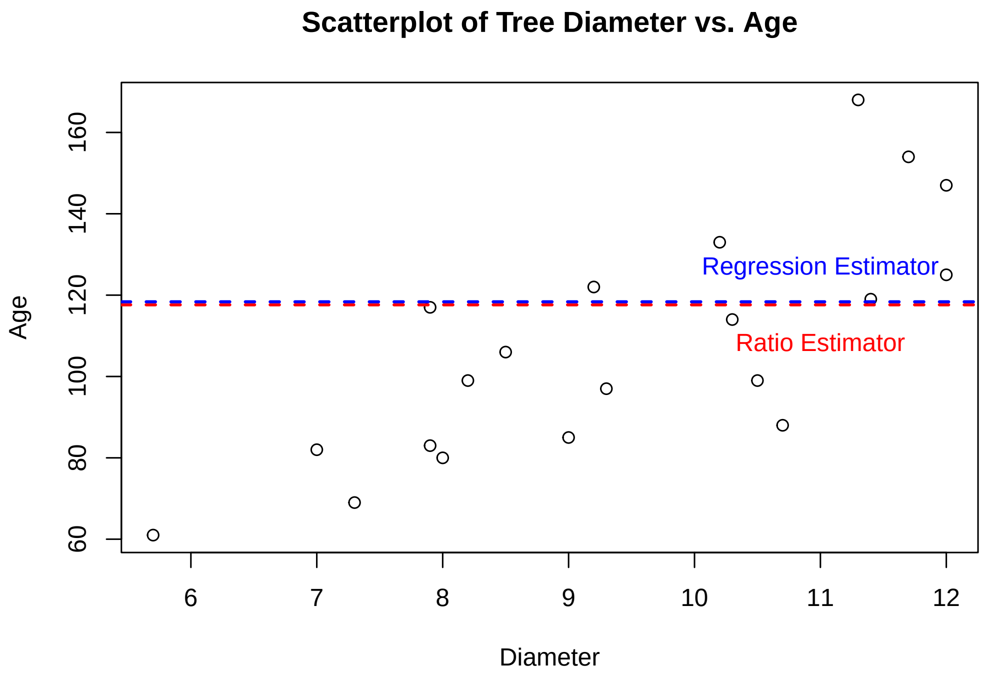
(d) **Label your estimates on your graph. How do they compare?**

```r
# Estimates
y_bar_hat_r <- 117.6204 # Ratio estimator
y_bar_hat_reg <- 118.3634 # Regression estimator

# Creating revised scatterplot
plot(diameter, age,
     main = "Scatterplot of Tree Diameter vs. Age",
     xlab = "Diameter",
     ylab = "Age",
     pch = 1)

# Adding horizontal lines for the estimates
abline(h = y_bar_hat_r, col = "red", lty = 2, lwd = 2) # Ratio estimator
abline(h = y_bar_hat_reg, col = "blue", lty = 2,
       lwd = 2) # Regression estimator

# Adding a legend for the estimates
text(x = max(diameter) - 1, y = y_bar_hat_r - 15,
    labels = "Ratio Estimator", pos = 3, col = "red")
text(x = max(diameter) - 1, y = y_bar_hat_reg + 3,
    labels = "Regression Estimator", pos = 3, col = "blue")
```



The ratio estimate and regression estimate are close to each other, but the regression estimate is slightly larger.