Simulation: simple linear regression

Manuel Villarreal

2024-08-29

1. Generate the needed variables for a simple linear regression model.

```
sample_size <- 40

beta_0 <- 4

beta_1 <- 3

x <- rnorm(n = sample_size, mean = 50, sd = 5)

sigma_2 <- 6

epsilon <- rnorm(n = sample_size, mean = 0, sd = sqrt(sigma_2))

y <- beta_0 + beta_1 * x + epsilon

s_xx <- sum((x - mean(x))^2)

x_bar <- mean(x)

hat_beta1 <- sum((y - mean(y)) * (x - mean(x))) / sum((x - mean(x))^2)

hat_beta0 <- mean(y) - hat_beta1 * mean(x)
```

The estimated value of the intercept was: 9.01

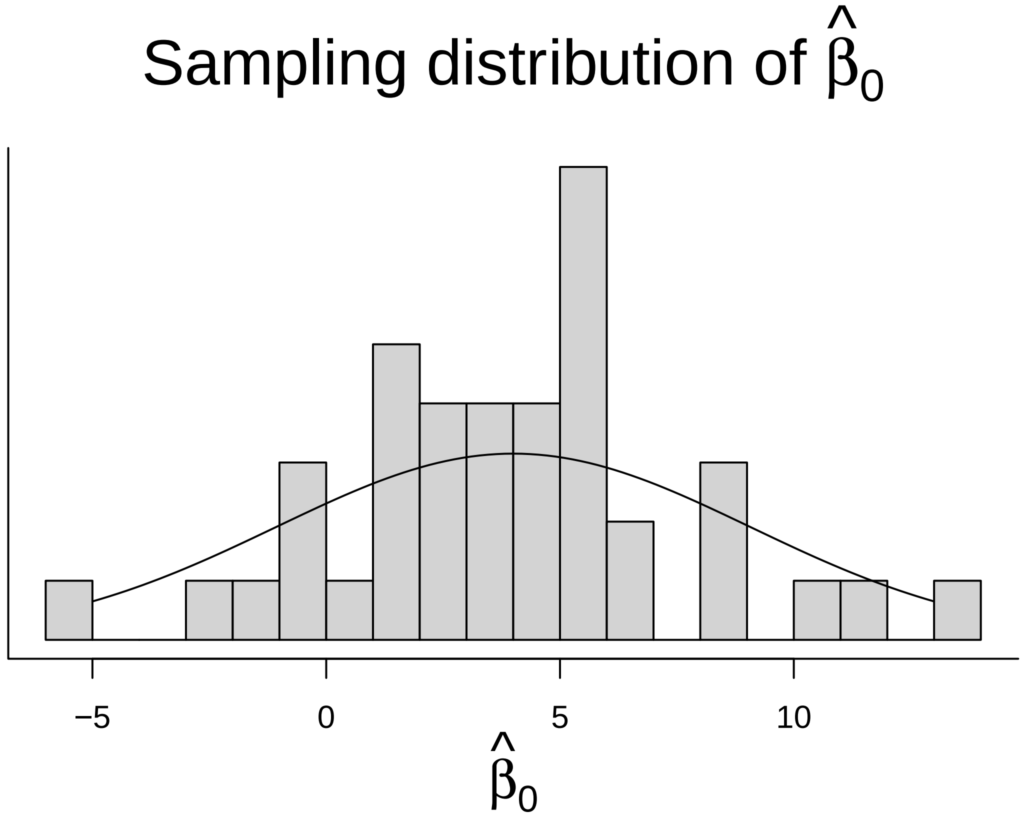
The estimated value of the slope was: 2.89

1. Simulate $k = 20$ "experiments" each with a sample size of 40, and save the parameter estimates of β_0, β_1, and σ^2.

```
n_experiments <- 40

hat_beta0 <- c()
hat_beta1 <- c()

for (i in 1:n_experiments) {
  epsilon <- rnorm(n = sample_size, mean = 0, sd = sqrt(sigma_2))

  y <- beta_0 + beta_1 * x + epsilon

  hat_beta1[i] <- sum((y - mean(y)) * (x - x_bar)) / s_xx
  hat_beta0[i] <- mean(y) - hat_beta1[i] * x_bar
}
```

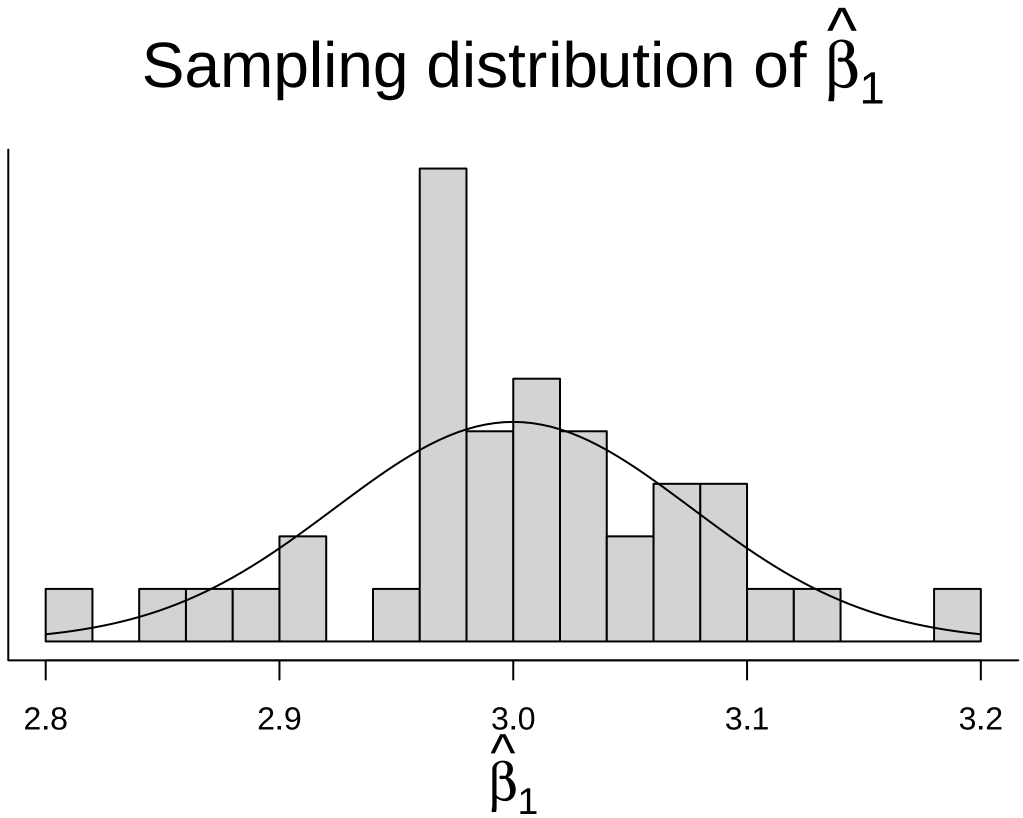
1. Plot the histogram of the estimated values of β_0 and overlay the theoretical normal distribution.

```r
hist(hat_beta0, freq = FALSE, breaks = 20, axes = FALSE, ann = FALSE)
curve(expr = dnorm(x, mean = 4,
                   sd = sqrt((sigma_2 * sum(x^2)) / s_xx)),
      add = TRUE, from = -5, to = 13)
box(bty = "l")
axis(1)
mtext(text = expression(paste("Sampling distribution of ", hat(beta)[0])),
      side = 3, cex = 2, line = 1)
mtext(text = expression(paste(hat(beta)[0])), side = 1, line = 3.2, cex = 1.7)
```



1. Plot the histogram of the estimated values of β_1 and overlay the theoretical normal distribution.

```r
hist(x = hat_beta1, breaks = 20, freq = FALSE, axes = FALSE, ann = FALSE)
box(bty = "l")
axis(1)
mtext(text = expression(paste(hat(beta)[1])), side = 1, line = 3.2, cex = 1.7)
mtext(text = expression(paste("Sampling distribution of ", hat(beta)[1])),
      side = 3, cex = 2, line = 1)
curve(expr = dnorm(x, mean = 3,
                   sd = sqrt(sigma_2 / s_xx)),
      add = TRUE)
```



1. Increase the number of experiments to 200 on step two and repeat steps from 2, 3, and 4.