Problem Set 5 ECON387 Queens College

Adriana Sham

October 16, 2019

```
my_ols = function(y, x){
    b_1 = cor(x, y) * sd(y) / sd(x)
    b_0 = mean(y) - b_1 * mean(x)
    mylist = as.numeric( list( b_1, b_0))
   return (mylist)
}# this function fits a linear regression line to the simple linear
# regression model using the least squares methodology

set.seed(37)
n = 1000
x = rnorm(n)
beta_0 = 0.5
beta_1 = 1.8

y = beta_0 + beta_1 * x
b = round(my_ols(y,x), 3)
paste("The intercept estimate is",b[2],
      "and the slope estimate is", b[1], ";
      when sample noise is not taken accountable")
```

```
## [1] "The intercept estimate is 0.5 and the slope estimate is 1.8 ; \n      when sample noise is not
```

```
epsilons = rnorm(n, mean = 0, sd = 1)
y1 = beta_0 + beta_1 * x + epsilons
b1 = round(my_ols(y1,x), 3)
paste("The intercept estimate is",b1[2],
      "and the slope estimate is", b1[1], ";
      the estimates are different from the true
      values due to the sampling noise")
```

```
## [1] "The intercept estimate is 0.534 and the slope estimate is 1.754 ; \n      the estimates are dif
```

```
epsilons2 = rnorm(n, mean = 0, sd = 1)
y2 = beta_0 + beta_1 * x + epsilons2
b2 = round(my_ols(y2,x), 3)
paste("The intercept estimate is",b2[2],
      "and the slope estimate is", b2[1],
      "; the estimates are different from the
      previous ones again due to the sampling noise")
```

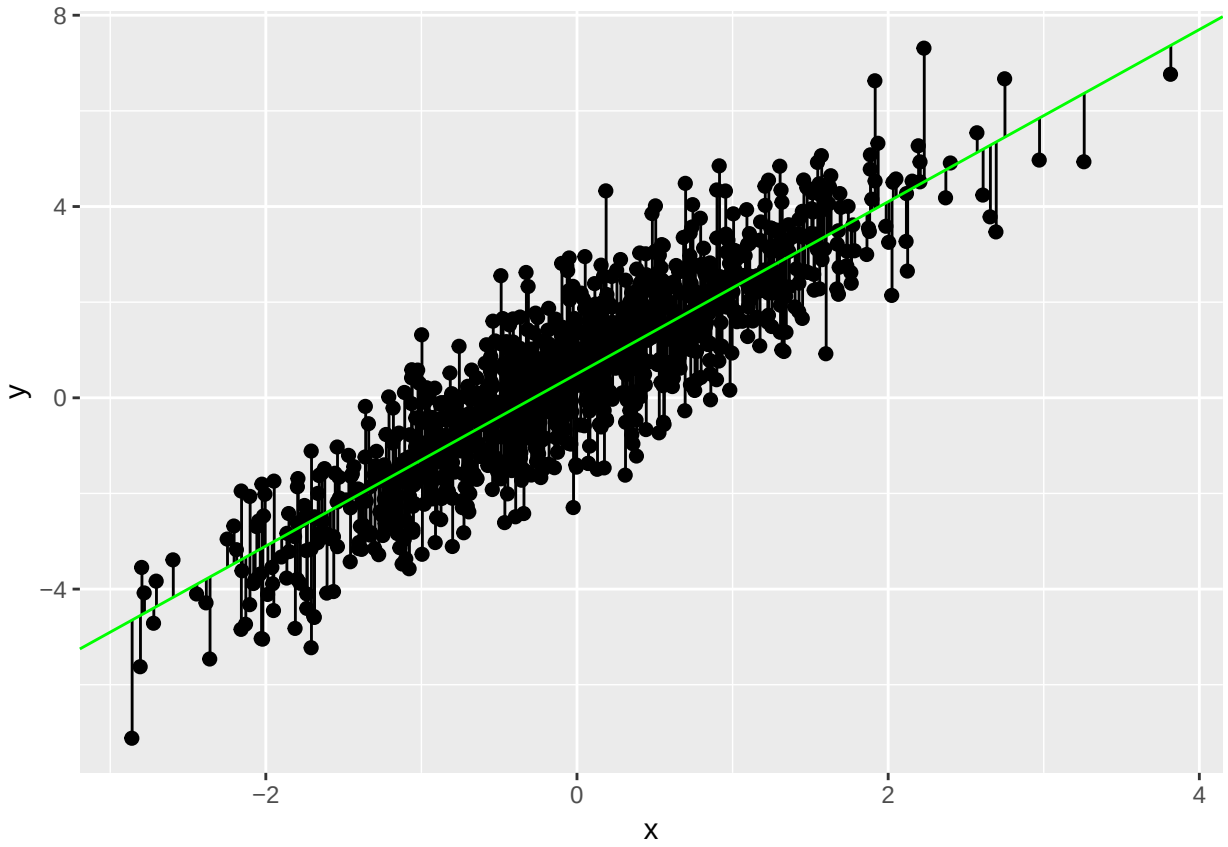
```
## [1] "The intercept estimate is 0.514 and the slope estimate is 1.843 ; the estimates are different f
```

```
pacman::p_load(ggplot2)
simple_df = data.frame(x = x, y = y2)

true_hstar_line = geom_abline(intercept = beta_0, slope = beta_1, color = "green")

simple_df$hstar = beta_0 + beta_1 * simple_df$x
simple_viz_obj = ggplot(simple_df, aes(x, y)) +
```

```
  geom_point(size = 2)
epsilon_line_segments = geom_segment(aes(xend = x, yend = hstar), position = position_nudge(x = 0.002))
simple_viz_obj + epsilon_line_segments + true_hstar_line
```




```
# the noise (epsilons) are the difference between the data
# and the best fit line
```