STA 4320 CHAP 3.1.2

He Jiang

Aug 2024

Sec 3.1.2

A relationship between the true underlying regression line (f), and the estimated regression line (f_hat) from the sample.

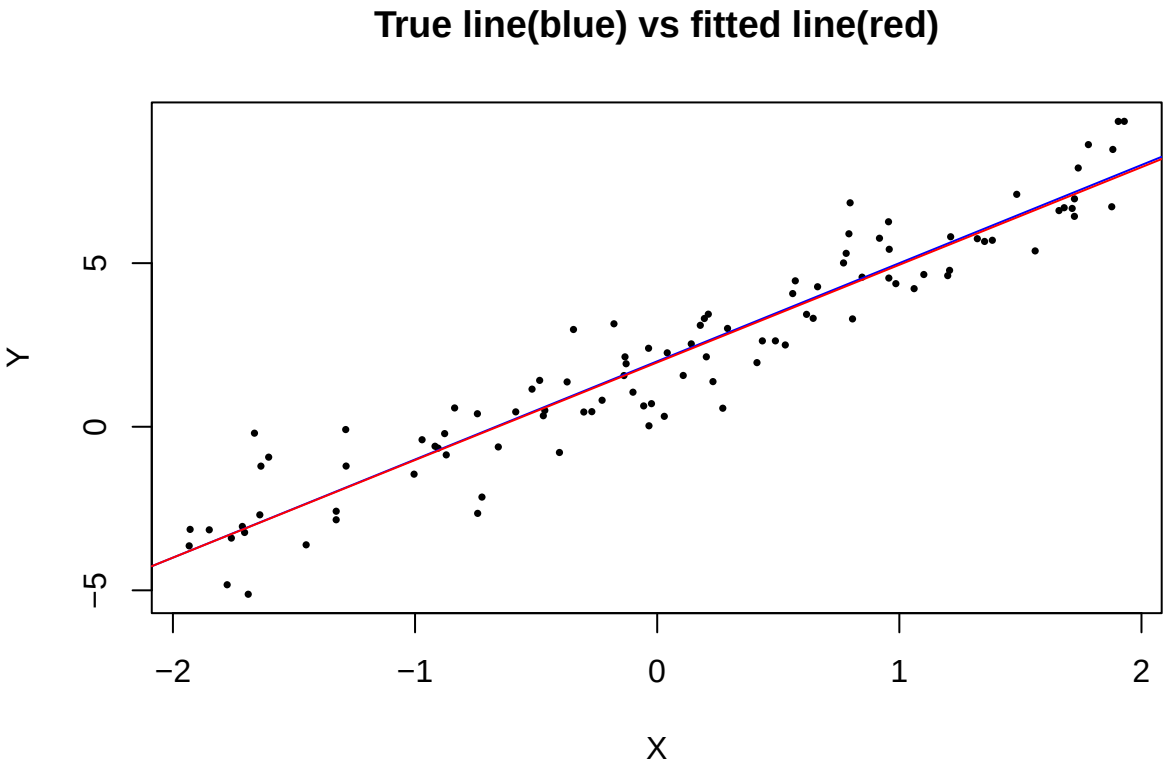
Assume the true line (f) is f(X) = 2+3X.

We conduct a simulation to see how the fitted line (f_hat) actually changes with different samples. This process can be seen by running the below code multiple times.

```
n = 100 # sample size
x = runif(n, -2, 2)
epsilon = rnorm(n, 0, 1)
y = 2 + 3*x + epsilon

res = lm(y ~ x)

plot(x, y,
     main = "True line(blue) vs fitted line(red)",
     pch = 16,
     cex = 0.5,
     xlab = "X",
     ylab = "Y")
abline(a = 2, b = 3, col = "blue")
abline(res, col = "red")
```



Advertising dataset

```
fpath = getwd()
Advertising = read.csv(paste0(fpath, "/Advertising.csv"))
x = Advertising$TV
y = Advertising$sales
```

```
res = summary( lm(y ~ x) )
res
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.3860 -1.9545 -0.1913  2.0671  7.2124
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 7.032594   0.457843   15.36   <2e-16 ***
## x           0.047537   0.002691   17.67   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.259 on 198 degrees of freedom
## Multiple R-squared:  0.6119, Adjusted R-squared:  0.6099
## F-statistic: 312.1 on 1 and 198 DF,  p-value: < 2.2e-16
```

Confidence interval for beta_1

```
# finding the critical value
n = length(x)
t_alpha_2 = qt(0.975, n-2)
t_alpha_2
```

```
## [1] 1.972017
```

```
# confidence interval
c(0.0475 - t_alpha_2 * 0.0027, 0.0475 + t_alpha_2 * 0.0027)
```

```
## [1] 0.04217555 0.05282445
```

Confidence interval for beta_0

```
# confidence interval
c(7.0326 - t_alpha_2 * 0.4578, 7.0326 + t_alpha_2 * 0.4578)
```

```
## [1] 6.12981 7.93539
```