APLM hw01

san teng

2018 年 10 月 11 日

(Programming work) The director of admission of a small college select 120 students at random from the new freshman class in a study to determine whether a student's grade point average (GPA) at the end of freshman year (Y) can be predicted from ACT score (X). The data information are shown in "CH01PR19.txt". Assume X and Y follows the simple linear regression model with $cov(\varepsilon_i, \varepsilon_j) = 0$ if $i \neq j$, and $cov(\varepsilon_i, \varepsilon_j) = \sigma^2$ if $i = j$.

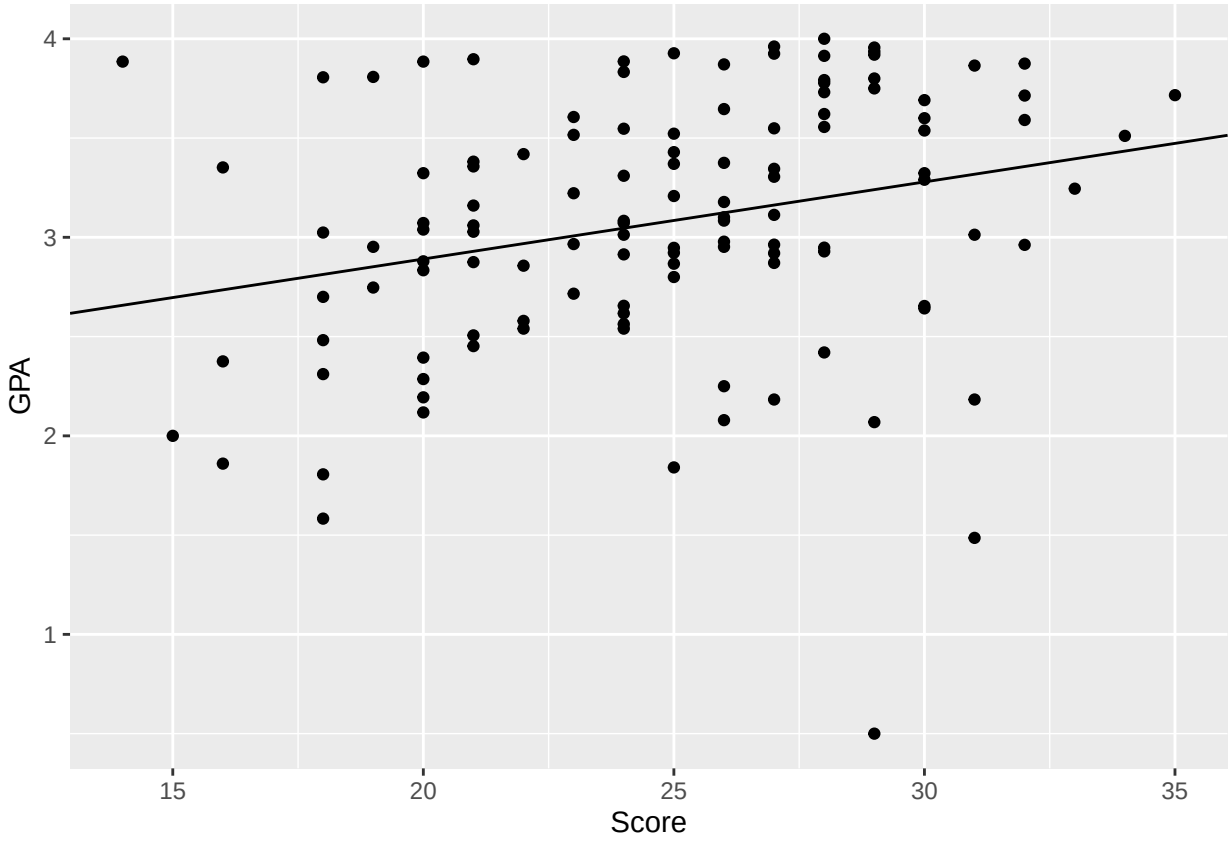
(a) the least squares estimators $\hat{\beta} = (\beta_0, \beta_1)$ is:

```
beta_hat <- as.vector( solve( t(X) %*% X) %*% t(X) %*% Y )
beta_hat
```

```
## [1] 2.11404929 0.03882713
```

(b) Plot the fitted regression with data:

```
g <- ggplot(data, aes(x = Score, y = GPA)) + geom_point()
g + geom_abline(intercept = beta_hat[1] ,slope = beta_hat[2])
```



我認為這條線 **(fitted line)** 並沒有對 **Score** 和 **GPA** 作出明顯的趨勢呈現。

Score 和 **GPA** 並沒有到很明顯的正比關係。在 **Score** 介於 **25~30** 間，有一些點是 **GPA** 低的點，使得如果不看這些資料點，則有著 **Score** 愈高，**GPA** 愈高的趨勢。

(c) the residuals and sum of residuals is:

```
residual <- Y - X %*% beta_hat
sum_of_residual <- sum(residual^2)
head(residual)
```

```
##             [,1]
## [1,]  0.96758105
## [2,]  1.22737094
## [3,]  0.57679116
## [4,] -0.42824608
## [5,]  0.09858105
## [6,]  0.54730978
```

```
sum_of_residual
```

```
## [1] 45.81761
```

(d) Estimate σ^2 and σ by MSE and $\sqrt{MSE}$:

```
MSE <- sum_of_residual / (dim(data)[1] - 2)
MSE
```

```
## [1] 0.3882848
```

```
sqrt(MSE)
```

```
## [1] 0.623125
```