生物統計 HW#2

數據所 0656704 賴品儒

目的: 學習使用 R 語言做 simple regression analysis

題目: How strong is the linear relationship between the age of a driver and the distance the driver can see? If we had to guess, we might think that the relationship is negative — as age increases, the distance decreases. A research firm collected data on a sample of *n* = 30 drivers ("age_distance.txt").

a. Please draw stem-leaf plots for "Age" and "Distance" respectively. Comments on the two plots

(這兩個圖可以讓看個別變數的分布狀況)

stem-leaf plos for Age：

```
The decimal point is 1 digit(s) to the right of the |

1 | 8
2 | 02335789
3 | 27
4 | 169
5 | 35
6 | 35678
7 | 01234579
8 | 2
```

年齡的高峰出現在 20 和 70 幾的族群，出現雙峰

stem-leaf plot for Distance：

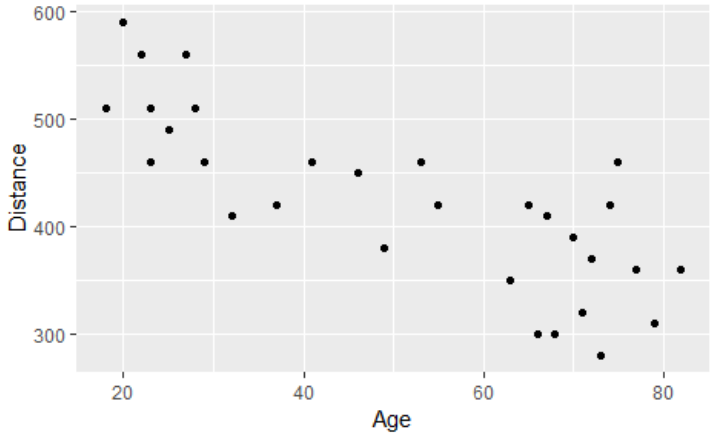
```
The decimal point is 2 digit(s) to the right of the |

2 | 8
3 | 0012
3 | 566789
4 | 112222
4 | 5666669
5 | 111
5 | 669
```

距離大致集中在 30、40 幾

b. Please draw a scatterplot of "Age" (X 軸) and "Distance" (Y 軸). Comments on the plot. Are there any outliers or influential observations?

(這個圖可以讓看兩個變數的關聯性)



Age 和 Distance 大致呈負相關，年齡越高，所看距離有下降趨勢。

沒有特別明顯的 outliers 及 influential observations。

c. For each variable, compute the mean and standard deviation. Then compute the Pearson correlation coefficient.

	mean	Standard deviation
Age	51	21.77629
Distance	423.3333	81.72002

Pearson correlation coefficient：-0.8012447

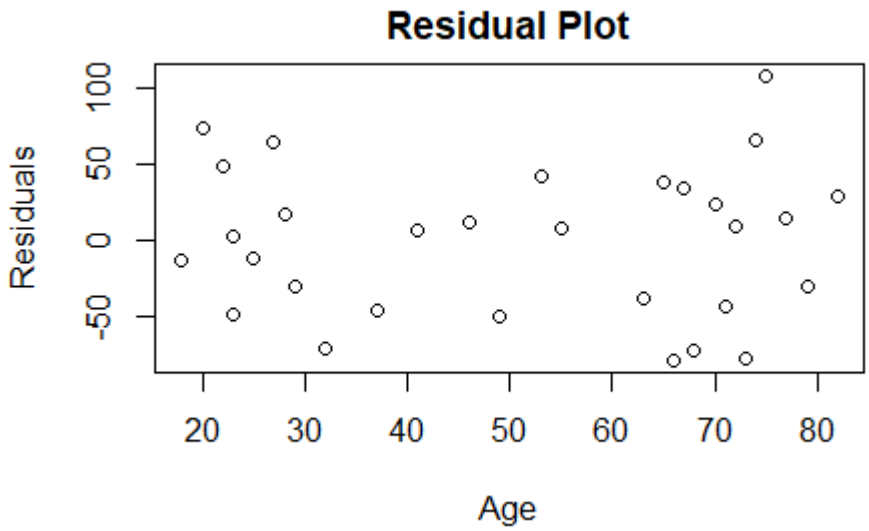
d. Fit the following linear regression:

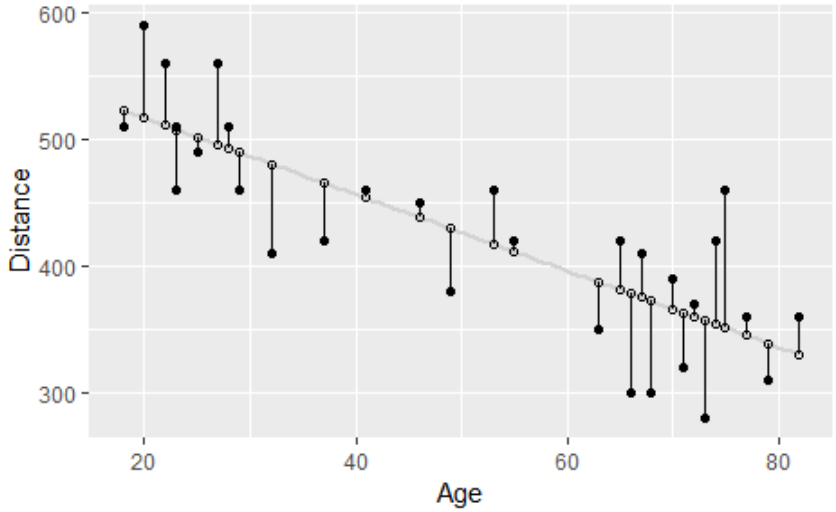
$\text{Distance} = \beta_0 + \beta_1 \text{age} + \varepsilon$.

What are the estimated values of $(\beta_0, \beta_1, Var(\varepsilon))$?

$(\beta_0, \beta_1, Var(\varepsilon))$ = (576.6819, -3.0068, 2476.215)

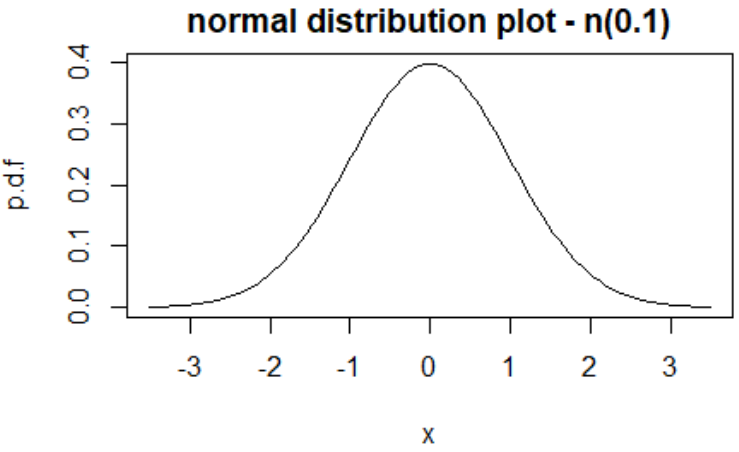
e. Draw the residual plot and comment on the plot.





Residual 在 Age 20~40 和 60~80 的 range 較大，40~60 的 resudual range 較小。

f. Draw the normal probability plot. Commet on the plot.



呈鐘形分布，上圖例子為標準常態分佈，平均數為 0，標準差為 1

g. What is the value of R^2 ? What does this value mean?

R^2 =0.641993

R^2 為衡量迴歸模式解釋力的側度，解釋 response 變數之變異的百分比，R^2 越大越好，介於 0 和 1 之間

R^2 =SSR/TSS

Code：

https://github.com/ayamisea/Applied-Methods-in-Statistics/blob/master/homework/HW02/HW02.Rmd