

Problem 3: 6210422036 ธนัท เอี่ยมปรีดี

3.1 จงหา correlation ระหว่าง car age และ distance พร้อมทั้งแปลผล

```
12          28.2      3.8      183013
> cor(d[c(1,2,5)])
      FuelConsumption      CarAge      Distance
FuelConsumption      1.0000000 -0.9324521 -0.9304757
CarAge               -0.9324521  1.0000000  0.9999337
Distance             -0.9304757  0.9999337  1.0000000
```

Car Age & Distance correlation = 0.9999337

หมายความว่า Car Age และ Distance มีความสัมพันธ์กันในทางบวกที่สูงมาก ใกล้ 1 มาก

3.2 พิจารณา สอง models ต่อไปนี้

- Model A ที่มี independent variable เป็น car age
- Model B ที่มี independent variables เป็น car age และ distance

ทำ leave-out cross-validation ใส่ค่า BIC, Adjusted R-squared, R-squared ที่ได้ในตาราง

	BIC	Adjusted R-squared	R-squared
Model A			
Model B			

```
regA <- lm(FuelConsumption ~ CarAge, data=d)
regB <- lm(FuelConsumption ~ CarAge + Distance, data=d)
CV(regA)
CV(regB)
```

```
> CV(regA)
      CV      AIC      AICc      BIC      AdjR2
0.6970580 -4.0031266 -1.0031266 -2.5484066 0.8564135
> CV(regB)
      CV      AIC      AICc      BIC      AdjR2
0.6652746 -4.8595340  0.8547517 -2.9199074 0.8742540
```

```
> summary(regA)$r.squared
[1] 0.8694669
> summary(regB)$r.squared
[1] 0.8971169
```

	BIC	Adj. R-squared	R-squared
Model A	-2.5484066	0.8564135	0.8694669
Model B	-2.9199074	0.8742540	0.8971169

3.3 ควรเลือก model A หรือ B? ให้เหตุผลในการเลือกว่าพิจารณาอะไรบ้าง? (สามารถตอบอย่างอื่นนอกเหนือจากตารางได้, ถ้าจำเป็น)

เลือก Model A

เนื่องจาก car age & distance มี correlation ที่สูงมากจะเกิดปัญหา Multicollinearity ทำให้ไม่ผ่าน diagnostic check ของ multiple regression ไม่ครบ

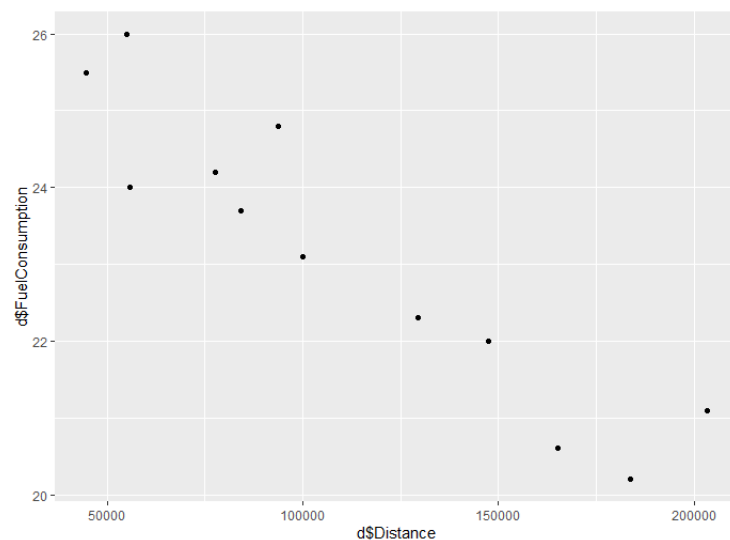
ซึ่งเครื่องหมาย coefficient ของ distance เป็น + ซึ่งควรจะเป็น ลบ นี่ก็ไม่ผ่านอีกข้อ แล้ว coefficient ทั้งสองตัวก็ไม่ significant

```
Call:
lm(formula = FuelConsumption ~ CarAge + Distance, data = d)

Residuals:
    Min       1Q   Median       3Q      Max
-1.05845 -0.36649  0.01306  0.17142  1.31738

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.659e+01  4.932e-01  53.905 1.31e-12 ***
CarAge       -2.021e+01  1.220e+01  -1.656   0.132
Distance      5.163e-04  3.319e-04   1.555   0.154
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6757 on 9 degrees of freedom
Multiple R-squared:  0.8971,    Adjusted R-squared:  0.8743
F-statistic: 39.24 on 2 and 9 DF,  p-value: 3.594e-05
```



ความสัมพันธ์ distance & fuel consumption ควรจะเป็น ลบ