

LAB LECTURE 9

62070501064 อรวิภา คุณเจริญไพศาล

Model 1

Residuals Standard Error: 32.16

```
> mod1 <- lm(Target..Total.orders. ~ Non.urgent.order, data)
> summary(mod1)

Call:
lm(formula = Target..Total.orders. ~ Non.urgent.order, data = data)

Residuals:
    Min       1Q   Median       3Q      Max
-47.301 -21.302  -4.259   12.069  164.290

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   92.99370    11.19444     8.307 1.88e-11 ***
Non.urgent.order  1.20472     0.06025    19.996 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 32.16 on 58 degrees of freedom
Multiple R-squared:  0.8733,    Adjusted R-squared:  0.8711
F-statistic: 399.9 on 1 and 58 DF,  p-value: < 2.2e-16
```

Model 2

Residuals Standard Error: 61.78

```
> mod2 <- lm(Target..Total.orders. ~ Urgent.order, data)
> summary(mod2)

Call:
lm(formula = Target..Total.orders. ~ Urgent.order, data = data)

Residuals:
    Min       1Q   Median       3Q      Max
-117.177  -39.350   -8.785    26.678   179.242

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   14.676    36.098   0.407   0.686
Urgent.order    2.407     0.296   8.129 3.72e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 61.78 on 58 degrees of freedom
Multiple R-squared:  0.5326,    Adjusted R-squared:  0.5245
F-statistic: 66.09 on 1 and 58 DF,  p-value: 3.719e-11
```

Model 3

Residuals Standard Error: 56.7

```
> mod3 <- lm(Target..Total.orders. ~ Urgent.order + Order.type.A, data)
> summary(mod3)

Call:
lm(formula = Target..Total.orders. ~ Urgent.order + Order.type.A,
    data = data)

Residuals:
    Min       1Q   Median       3Q      Max
-153.810  -32.394   -5.891   25.261  168.831

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -12.1445    34.0308  -0.357  0.72251
Urgent.order    1.9825     0.2983   6.646 1.23e-08 ***
Order.type.A    1.4826     0.4304   3.444 0.00108 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 56.7 on 57 degrees of freedom
Multiple R-squared:  0.6131,    Adjusted R-squared:  0.5995
F-statistic: 45.17 on 2 and 57 DF,  p-value: 1.762e-12
```

Model 4

Residuals Standard Error: 27.25

```
> mod4 <- lm(Target..Total.orders. ~ Urgent.order + Order.type.B, data)
> summary(mod4)

Call:
lm(formula = Target..Total.orders. ~ Urgent.order + Order.type.B,
    data = data)

Residuals:
    Min       1Q   Median       3Q      Max
-56.858 -17.686   -1.124   15.623   84.937

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  19.46182    15.92588   1.222   0.227
Urgent.order    1.20787     0.15170   7.962 7.94e-11 ***
Order.type.B    1.26129     0.08123  15.527 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 27.25 on 57 degrees of freedom
Multiple R-squared:  0.9106,    Adjusted R-squared:  0.9075
F-statistic: 290.4 on 2 and 57 DF,  p-value: < 2.2e-16
```

Model 5

Residuals Standard Error: 51.68

```
> mod5 <- lm(Target..Total.orders. ~ Urgent.order + Order.type.C, data)
> summary(mod5)

Call:
lm(formula = Target..Total.orders. ~ Urgent.order + Order.type.C,
    data = data)

Residuals:
    Min       1Q   Median       3Q      Max
-73.965 -29.142  -8.567   14.667  194.840

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   14.1861    30.1939   0.470   0.6403
Urgent.order    0.8995     0.3860   2.330   0.0234 *
Order.type.C    1.2880     0.2531   5.089  4.2e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 51.68 on 57 degrees of freedom
Multiple R-squared:  0.6786,    Adjusted R-squared:  0.6673
F-statistic: 60.18 on 2 and 57 DF,  p-value: 8.918e-15
```

Model 6

Residuals Standard Error: 62.2

```
> mod6 <- lm(Target..Total.orders. ~ Urgent.order + Fiscal.sector.orders, data)
> summary(mod6)

Call:
lm(formula = Target..Total.orders. ~ Urgent.order + Fiscal.sector.orders,
    data = data)

Residuals:
    Min       1Q   Median       3Q      Max
-100.980 -39.728  -9.037   28.768  177.842

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   16.49945    36.53288   0.452   0.653
Urgent.order    2.40497     0.29803   8.069 5.27e-11 ***
Fiscal.sector.orders -0.02103     0.04342  -0.484   0.630
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 62.2 on 57 degrees of freedom
Multiple R-squared:  0.5345,    Adjusted R-squared:  0.5182
F-statistic: 32.72 on 2 and 57 DF,  p-value: 3.433e-10
```

Model 7

Residuals Standard Error: 44.84

```
> mod7 <- lm(Target..Total.orders. ~ Urgent.order + Banking.orders..1., data)
> summary(mod7)

Call:
lm(formula = Target..Total.orders. ~ Urgent.order + Banking.orders..1.,
    data = data)

Residuals:
    Min       1Q   Median       3Q      Max
-72.871 -31.685  -9.092   27.285  149.747

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.371e+01  2.620e+01   0.523   0.603
Urgent.order    2.035e+00  2.208e-01   9.219 6.79e-13 ***
Banking.orders..1. 9.670e-04  1.327e-04   7.289 1.05e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 44.84 on 57 degrees of freedom
Multiple R-squared:  0.7581,    Adjusted R-squared:  0.7496
F-statistic: 89.31 on 2 and 57 DF,  p-value: < 2.2e-16
```

Model 8

Residuals Standard Error: 23.63

```
> mod8 <- lm(Target..Total.orders. ~ Urgent.order + Banking.orders..1. + Order.type.B, data )
> summary(mod8)

Call:
lm(formula = Target..Total.orders. ~ Urgent.order + Banking.orders..1. +
    Order.type.B, data = data)

Residuals:
    Min       1Q   Median       3Q      Max
-45.85 -13.29  -3.30   10.91   54.77

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.826e+01  1.381e+01   1.322   0.191
Urgent.order    1.268e+00  1.322e-01   9.591 2.03e-13 ***
Banking.orders..1. 3.781e-04  8.493e-05   4.452 4.11e-05 ***
Order.type.B     1.045e+00  8.555e-02  12.216 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 23.63 on 56 degrees of freedom
Multiple R-squared:  0.934,    Adjusted R-squared:  0.9304
F-statistic: 264.1 on 3 and 56 DF,  p-value: < 2.2e-16
```

Model 9

Residuals Standard Error: 26.58

```
> mod9 <- lm(Target..Total.orders. ~ Urgent.order + Banking.orders..2. + Order.type.B, data )
> summary(mod9)

Call:
lm(formula = Target..Total.orders. ~ Urgent.order + Banking.orders..2. +
    Order.type.B, data = data)

Residuals:
    Min       1Q   Median       3Q      Max
-55.528 -18.787   1.505  16.009  86.216

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.953e+01  1.634e+01   1.807   0.0761 .
Urgent.order   1.041e+00  1.701e-01   6.122  9.63e-08 ***
Banking.orders..2. 2.780e-04  1.402e-04   1.982   0.0524 .
Order.type.B   1.148e+00  9.760e-02  11.765 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 26.58 on 56 degrees of freedom
Multiple R-squared:  0.9165,    Adjusted R-squared:  0.912
F-statistic: 204.8 on 3 and 56 DF,  p-value: < 2.2e-16
```

Model 10

Residuals Standard Error: 27.23

```
> mod10 <- lm(Target..Total.orders. ~ Urgent.order + Banking.orders..3. + Order.type.B, data )
> summary(mod10)

Call:
lm(formula = Target..Total.orders. ~ Urgent.order + Banking.orders..3. +
    Order.type.B, data = data)

Residuals:
    Min       1Q   Median       3Q      Max
-53.571 -16.524  -0.617  16.695  87.622

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.341e+01  1.693e+01   0.792   0.432
Urgent.order   1.208e+00  1.516e-01   7.973  8.58e-11 ***
Banking.orders..3. 2.832e-04  2.702e-04   1.048   0.299
Order.type.B   1.256e+00  8.131e-02  15.450 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 27.23 on 56 degrees of freedom
Multiple R-squared:  0.9123,    Adjusted R-squared:  0.9076
F-statistic: 194.3 on 3 and 56 DF,  p-value: < 2.2e-16
```

จากการทดลองและสร้างโมเดลทั้ง 10 ครั้ง เมื่อนำค่า Residual standard error มาเปรียบเทียบกัน จะเห็นได้ว่า Model 8 ที่มีค่า Residual standard error เท่ากับ 23.63 นั้นเป็นค่าที่น้อยที่สุดเมื่อเทียบกับ Model อื่น ๆ จึงถือได้ว่า Model 8 เป็น Best Regression Model for predicting daily

คำถาม: Why is sometime adding predictors do not help prediction?

เมื่อใช้ฟังก์ชัน `cor()` ในการคำนวณค่า correlation ระหว่างแต่ละตัวแปรเพื่อดูค่าความสัมพันธ์นั้น บางตัวแปรที่มีค่า correlation กับตัวแปรเป้าหมายหรือ Target ของเรามาก ๆ ในช่วง $|0.8 \text{ ถึง } 1|$ หมายความว่าตัวแปรนั้นสัมพันธ์กับ Target มาก กลับกันหากมีตัวแปรที่มีค่า correlation กับ Target น้อย ๆ ในช่วง $|0 \text{ ถึง } 0.5|$ หมายความว่าสัมพันธ์กับตัวแปร Target น้อย หรืออาจจะไม่มีความสัมพันธ์กับ Target

ดังนั้นบางครั้งในการเพิ่มตัวแปรในการเอามาทำนาย หากตัวแปรนั้นมีค่า `cor` น้อย ๆ ก็อาจจะไม่ได้ส่งผลหรือช่วยอะไรในการทำนายของเรา หรือหากมีหลายตัวแปร เช่น $\text{Target} \sim a + b$ หาก a มีความสัมพันธ์กับ Target มาก ๆ ในทางกลับกัน b กลับไม่มีความสัมพันธ์กับ Target การนำ b เข้ามาทำนาย ก็อาจจะไม่ได้ช่วยอะไร ยิ่งถ้า a และ b ไม่ได้มีความสัมพันธ์กัน การใช้ในการทำนายร่วมกันก็จะไม่เกิดประโยชน์ใด ๆ เลย