

## Best mtcars model

### 20160796박보성

R 의 mtcars 데이터:32 종류의 자동차에 대하여, 그 특성을 정리한 자료이다 mtcars 자료에서, mpg 를 종속변수로 하는 가장 좋은 다중 회귀모형을 구성하시오

*가장 좋은 모형이란 무엇을 의미하는 지 먼저 그 의미를 정하고 그 의미에 맞는 가장 좋은 다중 회귀모형을 구성하시오*

좋은 모형이란 가장 예측력 있는 모형이며 AIC, BIC 기준으로 값이 작은 모형이다. Mtcars를 구성하는 변수들의 영향력을 살펴보면서 포함할 수 있는 모형을 정해보겠다.

```
> allmt = lm(data = mtcars, mpg ~ .)
> summary(allmt)
```

Call:  
lm(formula = mpg ~ ., data = mtcars)

Residuals:

Min	1Q	Median	3Q	Max
-3.4506	-1.6044	-0.1196	1.2193	4.6271

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	12.30337	18.71788	0.657	0.5181
cyl	-0.11144	1.04502	-0.107	0.9161
disp	0.01334	0.01786	0.747	0.4635
hp	-0.02148	0.02177	-0.987	0.3350
drat	0.78711	1.63537	0.481	0.6353
wt	-3.71530	1.89441	-1.961	0.0633
qsec	0.82104	0.73084	1.123	0.2739
vs	0.31776	2.10451	0.151	0.8814
am	2.52023	2.05665	1.225	0.2340
gear	0.65541	1.49326	0.439	0.6652
carb	-0.19942	0.82875	-0.241	0.8122

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.65 on 21 degrees of freedom  
Multiple R-squared: 0.869, Adjusted R-squared: 0.8066  
F-statistic: 13.93 on 10 and 21 DF, p-value: 3.793e-07

모든 변수들을 보게 되면 wt가 가장 예측력 있으며 그 다음으로 qsec, am이 영향력 있음을 알 수 있다. 이를 이용하여 다양한 모형을 만들어서 각각 비교해보겠다.

```
nmd1 = lm(data = mtcars, mpg ~ wt + qsec + am)
nmd2 = lm(data = mtcars, mpg ~ wt + qsec)
nmd3 = lm(data = mtcars, mpg ~ wt )
nmd4 = lm(data = mtcars, mpg ~ -1 + wt + qsec + am)
nmd5 = lm(data = mtcars, mpg ~ -1 + wt + qsec)
nmd6 = lm(data = mtcars, mpg ~ -1 + wt )
nmd7 = lm(data = mtcars, mpg ~ wt*qsec)
```

```

nmd8 = lm(data = mtcars, mpg ~ wt*qsec*am)
nmd9 = lm(data = mtcars, mpg ~ -1+ wt*qsec*am)
nmd10 = lm(data = mtcars, mpg ~ -1+ wt*qsec)
nmd11 = lm(data = mtcars, mpg ~ -1+ l(wt^2) + qsec + am)
nmd12 = lm(data = mtcars, mpg ~ -1+ l(wt^2) + qsec)
nmd13 = lm(data = mtcars, mpg ~ -1+ l(wt^2))
nmd14 = lm(data = mtcars, mpg ~ l(wt^2))
nmd15 = lm(data= mtcars, mpg ~ -1+ wt:qsec +am)
nmd16 = lm(data= mtcars, mpg ~ -1+ wt:qsec)

```

이 값들을 AIC, BIC 기준으로 구해봤을 때 다음과 같다.

```

> AIC(nmd1, nmd2, nmd3, nmd4, nmd5, nmd6, nmd7, nmd8, nmd9, nmd10, nmd11, nmd12, nmd13, nmd14, nmd15, nmd16)
      df      AIC
nmd1    5 154.1194
nmd2    4 156.7205
nmd3    3 166.0294
nmd4    4 154.2308
nmd5    3 167.4260
nmd6    2 248.8070
nmd7    5 157.2766
nmd8    9 147.7795
nmd9    8 145.8351
nmd10   4 155.5036
nmd11   4 160.2108
nmd12   3 171.2381
nmd13   2 267.3295
nmd14   3 177.8557
nmd15   3 217.4184
nmd16   2 246.8509
>
> BIC(nmd1, nmd2, nmd3, nmd4, nmd5, nmd6, nmd7, nmd8, nmd9, nmd10, nmd11, nmd12, nmd13, nmd14, nmd15, nmd16)
      df      BIC
nmd1    5 161.4481
nmd2    4 162.5834
nmd3    3 170.4266
nmd4    4 160.0937
nmd5    3 171.8233
nmd6    2 251.7384
nmd7    5 164.6053
nmd8    9 160.9711
nmd9    8 157.5610
nmd10   4 161.3666
nmd11   4 166.0738
nmd12   3 175.6353
nmd13   2 270.2609
nmd14   3 182.2529
nmd15   3 221.8156
nmd16   2 249.7824

```

```

nmd9 = lm(data = mtcars, mpg ~ -1+ wt*qsec*am)

```

경우가 가장 좋은 모형이라고 할 수 있다. 절편이 없는 경우이다.