

Rexample

Lingxiao Zhou

cruise dataset

load data

```
cruise <- read.fwf("https://raw.githubusercontent.com/lingxiaozhou/STA4210Rmaterial/main/data/cruise_sh",
  width = c(20, 20, rep(8, 7)), col.names = c("ship", "cline",
    "age", "tonnage", "passengers", "length", "cabins", "passdens",
    "crew"))

head(cruise)
```

```
##           ship           cline age  tonnage passengers length
## 1 Journey      Azamara         6  30.277         6.94    5.94
## 2 Quest        Azamara         6  30.277         6.94    5.94
## 3 Celebration  Carnival       26  47.262        14.86    7.22
## 4 Conquest     Carnival       11 110.000        29.74    9.53
## 5 Destiny      Carnival       17 101.353        26.42    8.92
## 6 Ecstasy      Carnival       22  70.367        20.52    8.55
##   cabins passdens  crew
## 1   3.55    42.64  3.55
## 2   3.55    42.64  3.55
## 3   7.43    31.80  6.70
## 4  14.88    36.99 19.10
## 5  13.21    38.36 10.00
## 6  10.20    34.29  9.20
```

Fit Full model

```
fit0 <- lm(crew ~ age + tonnage + passengers + length + cabins +
  passdens, data = cruise)
summary(fit0)
```

```
##
## Call:
## lm(formula = crew ~ age + tonnage + passengers + length + cabins +
##     passdens, data = cruise)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -1.7700 -0.4881 -0.0938  0.4454  7.0077
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.5213400  1.0570350  -0.493  0.62258
## age         -0.0125449  0.0141975  -0.884  0.37832
## tonnage      0.0132410  0.0118928   1.113  0.26732
## passengers  -0.1497640  0.0475886  -3.147  0.00199 **
## length       0.4034785  0.1144548   3.525  0.00056 ***
## cabins       0.8016337  0.0892227   8.985 9.84e-16 ***
## passdens    -0.0006577  0.0158098  -0.042  0.96687
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9819 on 151 degrees of freedom
## Multiple R-squared:  0.9245, Adjusted R-squared:  0.9215
## F-statistic: 308 on 6 and 151 DF, p-value: < 2.2e-16
```

```
AIC(fit0)
```

```
## [1] 451.4394
```

```
anova(fit0)
```

```
## Analysis of Variance Table
##
## Response: crew
##           Df Sum Sq Mean Sq  F value    Pr(>F)
## age         1  542.66   542.66  562.8982 < 2.2e-16 ***
## tonnage      1 1118.50  1118.50 1160.2189 < 2.2e-16 ***
## passengers   1   24.17    24.17  25.0735 1.521e-06 ***
## length       1   16.62    16.62  17.2387 5.492e-05 ***
## cabins       1   79.56    79.56  82.5234 5.450e-16 ***
## passdens     1    0.00     0.00   0.0017  0.9669
## Residuals   151  145.57     0.96
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

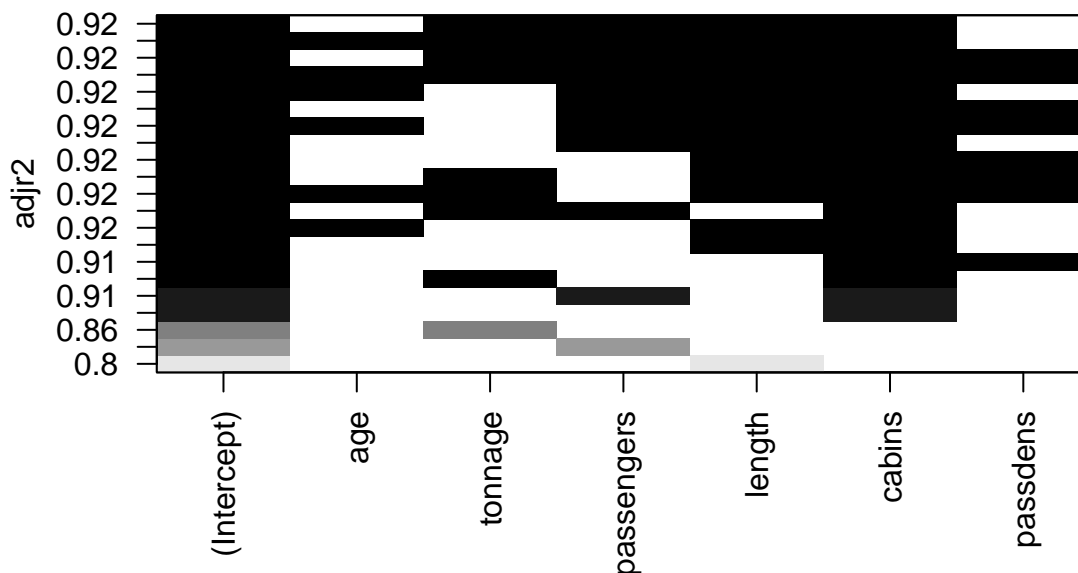
- Fit the linear regression model with all predictors
- According to the summary results and anova table, some predictors are not significant

Perform all possible regressions (aka all subset regressions)

```
# install.packages('leaps')
library(leaps)
allcruise <- regsubsets(crew ~ age + tonnage + passengers + length +
  cabins + passdens, nbest = 4, data = cruise)
apROUT <- summary(allcruise)
with(apROUT, round(cbind(which, rsq, adjr2, cp, bic), 3)) ## Prints 'readable' results
```

##	(Intercept)	age	tonnage	passengers	length	cabins	passdens	rsq	adjr2	cp
## 1	1	0	0	0	0	1	0	0.904	0.903	37.772
## 1	1	0	1	0	0	0	0	0.860	0.859	125.086
## 1	1	0	0	1	0	0	0	0.838	0.837	170.523
## 1	1	0	0	0	1	0	0	0.803	0.801	240.675
## 2	1	0	0	0	1	1	0	0.916	0.915	15.952
## 2	1	0	0	0	0	1	1	0.912	0.911	24.261
## 2	1	0	1	0	0	1	0	0.911	0.909	26.792
## 2	1	0	0	1	0	1	0	0.908	0.907	32.443
## 3	1	0	0	1	1	1	0	0.922	0.921	5.857
## 3	1	0	0	0	1	1	1	0.919	0.918	11.341
## 3	1	0	1	1	0	1	0	0.918	0.916	14.023
## 3	1	1	0	0	1	1	0	0.917	0.915	15.909
## 4	1	0	1	1	1	1	0	0.924	0.922	3.847
## 4	1	1	0	1	1	1	0	0.923	0.921	5.084
## 4	1	0	0	1	1	1	1	0.923	0.921	5.197
## 4	1	0	1	0	1	1	1	0.919	0.917	13.056
## 5	1	1	1	1	1	1	0	0.924	0.922	5.002
## 5	1	0	1	1	1	1	1	0.924	0.922	5.781
## 5	1	1	0	1	1	1	1	0.924	0.921	6.240
## 5	1	1	1	0	1	1	1	0.920	0.917	14.904
## 6	1	1	1	1	1	1	1	0.924	0.921	7.000
##	bic									
## 1	-360.238									
## 1	-300.954									
## 1	-277.122									
## 1	-246.201									
## 2	-376.131									
## 2	-368.502									
## 2	-366.249									
## 2	-361.332									
## 3	-382.878									
## 3	-377.413									
## 3	-374.808									
## 3	-373.002									
## 4	-381.933									
## 4	-380.652									
## 4	-380.534									
## 4	-372.631									
## 5	-377.752									
## 5	-376.939									
## 5	-376.462									
## 5	-367.717									
## 6	-372.692									

```
plot(allcruise, scale = "adjr2")
```



- The first column gives the number of predictors ($p-1$), not the number of parameters (p).
- The first row corresponds to the model with an intercept and cabins
- `nbest = 4` tells the program to print out the best 4 models for each possible number of parameters.
- The `with(aprout,round(cbind(which,rsq,adjr2,cp,bic),3))` command prints out the output in readable form, with 3 decimal places.
- Based on adjusted R^2 , the best model is the model with tonnage, passengers, length and cabins

```
fit.new <- update(fit0, . ~ . - age - passdens)
AIC(fit.new)
```

```
## [1] 448.3229
```

```
# install.packages('olsrr')
library(olsrr)

press_value <- ols_press(fit.new)
print(press_value)
```

```
## [1] 154.8479
```