

Model Selection

STAT-UB.0103 – Statistics for Business Control and Regression Models

Model Selection

1. Here are the results from fitting two models for Text. The first model using a single predictor variable, Social:

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	1885519	1885519	3.39	0.072
Error	44	24439192	555436		
Total	45	26324711			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
745.276	7.16%	5.05%	0.27%

Regression Equation

$$\text{Text} = 174 + 0.706 \text{ Social}$$

The second model uses two predictor variables: Social and Audio.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	2563229	1281615	2.32	0.111
Error	43	23761482	552593		
Total	45	26324711			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
743.366	9.74%	5.54%	0.00%

Regression Equation

$$\text{Text} = 77 + 0.712 \text{ Social} + 0.992 \text{ Audio}$$

- (a) Which model has the highest value of R^2 ?
- (b) Compute the value of AIC for the first model.
- (c) Compute the value of AIC for the second model.
- (d) According to AIC, which of these two models is preferable?
- (e) According to R_a^2 , which of these two models is preferable?

Best Subsets Regression

2. Here is the output from using best subsets regression with response Text and predictor variables Video, Audio, Email, Social, and Mail:

Response is Text

Vars	R-Sq	R-Sq (adj)	R-Sq (pred)	Mallows Cp	S	V	A	E	S
1	7.2	5.1	0.3	0.9	745.28				X
2	9.7	5.5	0.0	1.7	743.37	X			X
3	13.0	6.8	0.0	2.2	738.48	X	X		X
4	13.3	4.9	0.0	4.0	745.98	X	X	X	X
5	13.4	2.5	0.0	6.0	755.15	X	X	X	X

Use the output to answer the following questions:

- Of all candidate models with exactly 3 predictors, which fitted model has the smallest value of SSE?
- Of all candidate models with up to 5 predictors, which fitted model has the smallest value of SSE?
- Write an expression for SSE in terms of R^2 and SST.
- Write an expression for AIC in terms of R^2 , SST, n , and k .
- Use the answer from the previous part to find the candidate model with the smallest value of AIC.
- In this situation, does AIC agree with R_a^2 ?