COVER PAGE

STAT 608 Homework 06 Summer 2017

Please TYPE your name and email address below, then convert to PDF and attach as the first page of your homework upload.

NAME: RAJAN KAPOOR

EMAIL: r.kapoor@tamu.edu.

12/12

STATISTICS 608
Homework 608 S17 06
Due: *11:59 PM, July 17, 2017*

Question 1 [4+4+4+4+2+2=20]

Work Exercise 3 on page 261 in the textbook.

Question 2 [2+4=6]

Work Exercise 8.3.1, page 294, in the textbook.

Solution 1

(a) All possible subsets

	R^2_{adj}	AIC	AICc	BIC
1	0.25107	−63.126	−63.0018	−56.57053
2	0.48577	−135.85	−135.6493	−126.0244
3	0.53819	−155.31	−154.94498	**−142.1983**
4	0.54278	−156.291	−155.8468	−139.9004
5	**0.54585**	**−156.6416**	**−156.0459**	−136.9729
6	0.54366	−154.7307	−153.9606	−131.7839
7	0.54124	−152.7355	−151.7677	−126.5105

Optimal Model acc to BIC $(x_2 + x_4 + x_6)$ or

$\log(\text{PrizeMoney}) \sim \text{GIR} + \text{BirdieConversion} + \text{Scrambling}$

Optimal model acc to AIC, AICc, R^2_{adj}

$\log(\text{PrizeMoney}) \sim \text{GIR} + \text{BirdieConversion} + \text{Sandsaves} + \text{Scrambling} + \text{PuttsperRound}.$

or $(x_2 + x_4 + x_5 + x_6 + x_7)$

(b) Backward selection AIC

$\log(Y) \sim \text{GIR} + \text{BirdieConversion} + \text{Sandsaves} + \text{Scrambling} + \text{PuttsperRound}$

or $(x_2 + x_4 + x_5 + x_6 + x_7)$

Backward selection BIC

$\log(Y) \sim \text{Scrambling} + \text{GIR} + \text{BirdieConversion}$

or $(x_6 + x_2 + x_4)$

All in (R-notation)

(c) Forward Selection AIC

$\log(Y) \sim$ GIR + PuttsperRound + BirdieConversion + Scrambling + Sandsaves (in R-notation)

or $(x_2 + x_7 + x_4 + x_6 + x_5)$

Forward selection BIC

$\log(Y) \sim$ GIR + PuttsperRound + BirdieConversion + Scrambling (in R-notation)

or $(x_2 + x_7 + x_4 + x_6)$

(d) In Backward selection, all predictors in subset 5 are present in subset 6 and subset 7. So when elimination starts from subset 7 and stops at subset 5 same result as (a) is obtained for AIC. Similar argument can be given for BIC in (b) and (a)

On the other hand, when starting from 1 predictor in forward selection, PuttsperRound is present in subset 2 predictors but not in subset 3 predictors Since in forward selection, once a predictor is added, it cannot be removed in following steps, BIC fails to consider the 3-predictor sub-sets without puttsperRound predictor and different result from (a) is obtained.

Forward selection AIC gives same result because it doesnot stop at 3 predictor set.

(e) The 3-predictor model has all 3 regression coeff as statistically highly significant while 5 predictor model has only 3 of 5 coefficients statistically significant

Final Recommended model is selected using partial F test.

M1: $\log(Y) \sim x_2 + x_4 + x_6$

or $\log(Y) = \beta_0 + \beta_2 x_2 + \beta_4 x_4 + \beta_6 x_6$

M2: $\log(Y) \sim x_2 + x_4 + x_6 + x_7 + x_5$

Using partial F-test in anova,

- the p-value is found to be significant

$\Rightarrow$ cannot reject reduced model M1.

$\therefore$ final Recommended model is

$$\log(Y) = \beta_0 + \beta_2 x_2 + \beta_4 x_4 + \beta_6 x_6$$

(f) The mmp plots show quite a variation when compared with non linear models. Yes, we should be cautious in using the results literally..

2 (a) ① The variable PlayoffAppearances is discrete. It is based on series of yes/no responses which ideally follow a binomial distribution, not Gaussian as assumed by the linear model. The variance in such case is not constant and thus validity of model is in question.

② From histogram, the predictor Population appears to be right skewed. As such, some sort of transformation or inclusion of Log(Population) is expected.

③ Using logistic regression of PlayoffAppearances on Log(Population) results in regression coeff

$$\beta_0 = -1.8525$$

$$\beta_1 = 0.5606$$