Quiz 4 (Sheather 7.3 and 7.4)

ⓘ This is a preview of the published version of the quiz

Started: Sep 28 at 5:11pm

Quiz Instructions

This quiz uses the same "beauty" data set that you are working with for HW04 (I have removed the same variables that I recommended removing in the problem statement for HW04). Below are some analyses. The quiz questions all have to do with these analyses.

```
> library(MASS)
> library(leaps)
> library(car)
Loading required package: carData
>
> ##################
> ## The Data... ##
> ##################
>
> str(beauty.red)
'data.frame':   463 obs. of  24 variables:
 $ tenured          : int  0 1 1 1 0 1 0 1 0 0 ...
 $ minority         : int  1 0 0 0 0 0 0 0 0 0 ...
 $ age              : int  36 59 51 40 31 62 33 51 33 47 ...
 $ beautyf2upper    : int  6 2 5 4 9 5 5 6 5 6 ...
 $ beautyflowerdiv  : int  5 4 5 2 7 6 4 4 3 5 ...
 $ beautyfupperdiv  : int  7 4 2 5 9 6 4 6 7 7 ...
 $ beautym2upper    : int  6 3 3 2 6 6 4 3 5 6 ...
 $ beautymlowerdiv  : int  2 2 2 3 7 5 4 2 5 3 ...
 $ beautymupperdiv  : int  4 3 3 3 6 5 4 3 3 6 ...
 $ btystdave        : num  0.202 -0.826 -0.66 -0.766 1.421 ...
 $ courseevaluation : num  4.3 4.5 3.7 4.3 4.4 4.2 4 3.4 4.5 3.9 ...
 $ didevaluation    : int  24 17 55 40 42 182 33 25 48 16 ...
 $ female           : int  1 0 0 1 1 0 1 1 1 0 ...
 $ formal           : int  0 0 0 0 0 1 0 0 0 0 ...
 $ fulldept         : int  1 1 1 1 1 1 1 1 1 1 ...
 $ lower            : int  0 0 0 0 0 0 0 0 0 0 ...
 $ nonenglish       : int  0 0 0 0 0 0 0 0 0 0 ...
```

```
 $ onecredit        : int  0 0 0 0 0 0 0 0 0 0 ...
 $ percentevaluating: num  55.8 85 100 87 87.5 ...
 $ profevaluation   : num  4.7 4.6 4.1 4.5 4.8 4.4 4.4 3.4 4.8 4 ...
 $ students         : int  43 20 55 46 48 282 41 41 60 19 ...
 $ tenuretrack      : int  1 1 1 1 1 1 1 1 0 ...
 $ blkandwhite      : int  0 0 0 0 0 0 0 0 0 ...
 $ btystdvariance   : num  2.13 1.39 2.54 1.76 1.69 ...
> beauty.red <- beauty.red[,-grep("profevaluation",names(beauty.red))]
> dim(beauty.red)
[1] 463  23
> n <- dim(beauty.red)[1]
> pmax <- dim(beauty.red)[2]
>
> ############################################
> ## Dividing into training and test sets... ##
> ############################################
>
> train <- (1:n) %in% sample(1:n)[1:floor(n/2)]
> test <- !train
> beauty.train <- beauty.red[train,]
> beauty.test <- beauty.red[test,]
>
> ###############################################################################
> ## All Subsets variable selection, with BIC criterion, on the training data ##
> ###############################################################################
>
> all.subsets <- regsubsets(courseevaluation ~ ., data=beauty.train,
+                           nvmax=pmax)
>
> results <- with(summary(all.subsets),data.frame(which,bic))
>
> best.bic.result <- results[results$bic==min(results$bic),]
> best.bic.vars <- paste(names(best.bic.result)[2:pmax]
+                        [unlist(best.bic.result[1,(2:pmax)])],
+                        collapse=" + ")
> best.bic.formula <- paste("courseevaluation ~",best.bic.vars)
> best.bic.model <- lm(best.bic.formula, data=beauty.train)
>
> ###############################################################################
> ## Stepwise variable selection, with BIC criterion, on the training data ##
```

```
> ###############################################################################
>
> best.stepwise.model <- stepAIC(lm(courseevaluation ~ ., data=beauty.train),
+                           direction="both",k=log(n))
>
> ## (omitting output from stepAIC...)
>
> ########################################
> ## Comparison of the fitted models... ##
> ########################################
>
> round(cbind(summary(best.bic.model)$coef,vif=c(NA,vif(best.bic.model))),2)
                  Estimate Std. Error t value Pr(>|t|)  vif
(Intercept)           3.36       0.16   21.26     0.00   NA
beautyf2upper         0.05       0.02    2.89     0.00 1.14
female               -0.24       0.07   -3.40     0.00 1.13
formal                0.27       0.10    2.74     0.01 1.12
nonenglish           -0.38       0.14   -2.72     0.01 1.06
onecredit             0.48       0.15    3.17     0.00 1.06
percentevaluating     0.01       0.00    2.98     0.00 1.16
> round(with(summary(best.bic.model),c(r2=r.squared,r2adj=adj.r.squared)),2)
   r2 r2adj
 0.18  0.16
>
> round(cbind(summary(best.stepwise.model)$coef,vif=c(NA,vif(best.stepwise.model))),2)
                Estimate Std. Error t value Pr(>|t|)     vif
(Intercept)      -32.76       9.92   -3.30        0      NA
beautyf2upper      1.26       0.33    3.84        0  406.11
beautyflowerdiv    1.33       0.37    3.61        0  436.56
beautyfupperdiv    1.32       0.35    3.82        0  413.58
beautym2upper      1.49       0.41    3.61        0  406.42
beautymlowerdiv    1.46       0.41    3.58        0  444.84
beautymupperdiv    1.25       0.32    3.92        0  435.18
btystdave        -15.45       4.16   -3.71        0 9968.47
female            -0.33       0.07   -4.54        0    1.23
formal             0.36       0.11    3.43        0    1.31
nonenglish        -0.47       0.14   -3.27        0    1.17
tenuretrack       -0.28       0.08   -3.40        0    1.17
blkandwhite        0.30       0.11    2.86        0    1.37
> round(with(summary(best.stepwise.model),c(r2=r.squared,r2adj=adj.r.squared)),2)
```

```
   r2 r2adj
 0.24  0.20
>
> BIC(best.bic.model,best.stepwise.model)
                     df      BIC
best.bic.model        8 367.5066
best.stepwise.model  14 383.7878
```

Question 1

1 pts

Suppose you have fitted these models in order to offer advice to a college dean about what (besides teaching quality) might affect course evaluations. Write ONE advantage of discussing the model `best.bic.model` with the dean, and ONE advantage of discussing the model `best.stepwise.model` with the dean.

(write two sentences, one for each model)

p

0 words

</>

Question 2

1 pts

Which model will you choose to discuss with the dean, `best.bic.model` or `best.stepwise.model`, and why?

(Write one sentence for your answer)

p

0 words

</>

Question 3

1 pts

Now we use the test data to compare the models (see below). Does the result cause you to change your answer to question 2? Write one sentence explaining why or why not.

```
> ##############################
> ## test set cross validation ##
```

```
> ##############################
>
> best.bic.pred <- predict(best.bic.model,newdata=beauty.test)
> best.bic.pred.RMSE <-
+     sqrt(mean((beauty.test$courseevaluation - best.bic.pred)^2))
>
> best.stepwise.pred <- predict(best.stepwise.model,newdata=beauty.test)
> best.stepwise.pred.RMSE <-
+     sqrt(mean((beauty.test$courseevaluation - best.stepwise.pred)^2))
>
> best.bic.pred.RMSE
[1] 0.5210218
> best.stepwise.pred.RMSE
[1] 0.5423852
```

Edit View Insert Format Tools Table

12pt Paragraph

p 0 words

Question 4 1 pts

For the model you chose in question 2:

Not including transformations, what modification would you try on this model? Write one sentence describing your modification, or just write "no modifications" if you don't think any are needed.

Edit View Insert Format Tools Table

12pt Paragraph

p 0 words

Not saved Submit Quiz