STAT2008/STAT6038

Examples – Stepwise, All Subsets, Drop Test

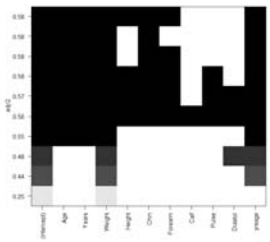
Two basic methods of selecting predictors

Stepwise regression: Enter and remove variables, in a stepwise manner, until no justifiable reason to enter or remove more.

Best subsets regression: Select the subset of variables that do the best at meeting some well-defined objective criterion(we have already seen this method)

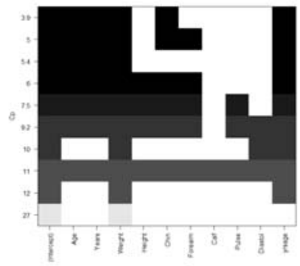
All subsets: adjusted R squared

```
g<-
regsubsets(Systol~Age+Years+Weight+Height+Chin+Forearm+Calf+Pulse+Diastol+yrsage,nvmax=10,nbest=1,data=p
eru)
> plot(g,scale="adjr2")
```



Mallows

```
> plot(g,scale="Cp")
```



Mallows

```
> gsum<-summary(g)
> cbind(gsum$which,gsum$cp)
   (Intercept) Age Years Weight Height Chin Forearm Calf Pulse Diastol yrsage
1            1   0     0      1      0    0       0    0     0       0      0 27.275044
2            1   0     0      1      0    0       0    0     0       0      1 12.060657
3            1   0     0      1      0    0       0    0     0       1      1 10.214927
4            1   1     1      1      0    0       0    0     0       0      1  5.432975
5            1   1     1      1      0    1       0    0     0       0      1  3.904453
6            1   1     1      1      0    1       1    0     0       0      1  5.033319
7            1   1     1      1      1    1       1    0     0       0      1  6.021644
8            1   1     1      1      1    1       1    0     1       0      1  7.533054
9            1   1     1      1      1    1       1    0     1       1      1  9.175292
10           1   1     1      1      1    1       1    1     1       1      1 11.000000
```

Fit the 5 variable model

6

```
> mod1<-lm(Systol~Age+Years+Weight+Chin+yrsage)
> anova(mod1)
Analysis of Variance Table

Response: Systol
          Df  Sum Sq Mean Sq F value    Pr(>F)    
Age        1    0.22    0.22  0.0031  0.955795    
Years      1   82.55   82.55  1.1542  0.290459    
Weight     1 2693.40 2693.40 37.6584 6.457e-07 ***
Chin       1  214.88  214.88  3.0044  0.092373 .  
yrsage     1 1180.14 1180.14 16.5004  0.000282 ***
Residuals 33 2360.23   71.52                      
```

Fit the 2 variable model

7

```
> mod2<-lm(Systol~Weight+yrsage)
> anova(mod2)
Analysis of Variance Table

Response: Systol
          Df Sum Sq Mean Sq F value    Pr(>F)
Weight     1 1775.4 1775.38  18.572 0.0001210 ***
yrsage     1 1314.7 1314.69  13.753 0.0006991 ***
Residuals 36 3441.4   95.59
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Carry out a "Drop SSE Test" to compare these two models

8

H_0: Model 2 with SSE_2=3441.4, df_2=36.

H_a: Model 1 with SSE_1=2360.2, df_1=33, MSE_1=71.5.

$$F = \frac{(SSE_2 - SSE_1)/(df_2 - df_1)}{MSE_1}$$
$$= \frac{(3441.4 - 2360.2)/(36 - 33)}{71.5}$$
$$= 5.04 > F_{df_2 - df_1, df_1, \alpha} = F_{3,33,.05} = 2.89$$

So the explanatory capability of Model 1 (on Y) is significantly greater than Model 2, but that doesn't mean it is a "better" model. Need to check for multi-collinearity, consider complexity etc

Stepwise Example: Cement data

Response y: heat evolved in calories during hardening of cement on a per gram basis

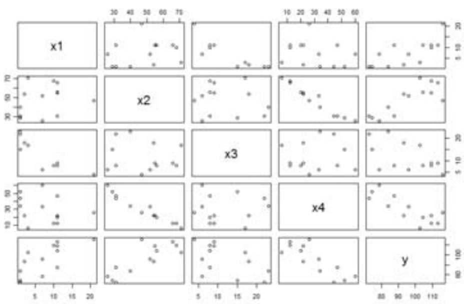
Predictor x_1: percentage of tricalcium aluminate

Predictor x_2: percentage of tricalcium silicate

Predictor x_3: percentage of tetracalcium alumino ferrite

Predictor x_4: percentage of dicalcium silicate

Pairs Cement data



Stepwise regression: the idea

Start with no predictors or all predictors in the model.

At each step, enter or remove a variable based on partial F-tests.

Stop when no more variables can be justifiably entered or removed.

Stepwise regression: the steps

Specify an Alpha-to-Enter (e.g. 0.15) and an Alpha-to-Remove (e.g. 0.15).

Start with no predictors in the model.

Put the predictor with the smallest P-value based on the partial F statistic (a t-statistic) in the model. If P-value > 0.15, then stop. None of the predictors have good predictive ability.

Otherwise

Stepwise regression: the steps

Add the predictor with the smallest P-value (below 0.15) based on the partial F-statistic (a t-statistic) in the model.

If none of the predictors yield P-values < 0.15, stop.

If P-value > 0.15 for any of the partial F statistics, then remove violating predictor.

Continue above two steps, until no more predictors can be entered or removed.

Stepwise regression: the idea (repeated slide- ignore)

Start with no predictors in the "stepwise model."

At each step, enter or remove a predictor based on partial F-tests (that is, the t-tests).

Stop when no more predictors can be justifiably entered or removed from the stepwise model.

Stepwise regression: Preliminary steps

Specify an Alpha-to-Enter ($\alpha_E = 0.15$) significance level.

Specify an Alpha-to-Remove ($\alpha_R = 0.15$) significance level.

Stepwise regression: Step No.1

Fit each of the one-predictor models, that is, regress y on x_1, regress y on x_2, regress y on x_{p-1}.

The first predictor put in the stepwise model is the predictor that has the smallest t-test P-value (below $\alpha_E = 0.15$).

If no P-value < 0.15, stop.

Cement data Step No.1

17

```
> install.packages("MASS")
> library("MASS")
> cement<-read.csv("cement.csv")
> attach(cement)
> nullmodel<-lm(y~1)
> fullmodel<-lm(y~x1+x2+x3+x4)
> addterm(nullmodel,scope=fullmodel,test = "F")
Single term additions
Model:
y ~ 1
       Df Sum of Sq     RSS    AIC F Value     Pr(F)
<none>              2715.76 71.444
x1      1   1450.08 1265.69 63.519 12.6025 0.0045520 **
x2      1   1809.43  906.34 59.178 21.9606 0.0006648 ***
x3      1    776.36 1939.40 69.067  4.4034 0.0597623 .
x4      1   1831.90  883.87 58.852 22.7985 0.0005762 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Try fitting all models that differ from the current model by adding a single term from those supplied, maintaining marginality.

Stepwise regression: Step No.2

Suppose x_4 was the "best" one predictor.

Fit each of the two-predictor models with x4 in the model, that is, regress y on (x_4, x_2), regress y on (x_4, x_3), , and y on (x_4, x_{p-1}).

The second predictor put in stepwise model is the predictor that has the smallest t-test P-value (below $\alpha_E = 0.15$).

If no P-value < 0.15, stop.

X4 inluded

```
newmod<-lm(y~x4)
> addterm(newmod,scope=fullmodel,test = "F")
Single term additions

Model:
y ~ x4
       Df Sum of Sq    RSS    AIC F Value     Pr(F)
<none>              883.87 58.852
x1      1    809.10  74.76 28.742 108.224 1.105e-06 ***
x2      1     14.99 868.88 60.629   0.172    0.6867
x3      1    708.13 175.74 39.853  40.295 8.375e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Stepwise regression: Step No.2 (continued)

Suppose x_1 was the best second predictor.

Step back and check P-value for $\beta_4 = 0$. If the P-value for $\beta_4 = 0$ has become not significant (above $\alpha_R = 0.15$), remove x4 from the stepwise model.

Check to see if X4 should now be dropped

```
> newmod1<-lm(y~x4+x1)
> dropterm(newmod1,test = "F")
Single term deletions

Model:
y ~ x4 + x1
       Df Sum of Sq     RSS    AIC F Value     Pr(F)
<none>                74.76 28.742
x4      1    1190.9 1265.69 63.519  159.29 1.815e-07 ***
x1      1     809.1  883.87 58.852  108.22 1.105e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The **dropterm** function considers each variable individually and considers what the change in residual sum of squares would be if this variable was excluded from the model. If p-value is large then we should consider dropping it from the model

Stepwise regression: Step No.3

Suppose both x_1 and x_4 made it into the two-predictor stepwise model.

Fit each of the three-predictor models with x_1 and x_4 in the model, that is, regress y on (x_1, x_2, x_4), regress y on (x_1, x_3, x_4),..., and regress y on (x_1, x_4, x_{p-1}).

X1 included

```
> newmod1<-lm(y~x4+x1)
> addterm(newmod1,scope=fullmodel,test = "F")
Single term additions

Model:
y ~ x4 + x1
       Df Sum of Sq    RSS    AIC F Value   Pr(F)
<none>              74.762 28.742
x2      1    26.789 47.973 24.974  5.0259 0.05169 .
x3      1    23.926 50.836 25.728  4.2358 0.06969 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Stepwise regression: Step No.3 (continued)

The third predictor put in stepwise model is the predictor that has the smallest t-test P-value (below $\alpha_E = 0.15$).

If no P-value < 0.15, stop.

Step back and check P-values for $\beta_1 = 0$ and $\beta_4 = 0$. If either P-value has become not significant (above $\alpha_R = 0.15$), remove the predictor from the stepwise model.

After adding X2, check X1 and X4

```
> dropterm(newmod2,test="F")
Single term deletions

Model:
y ~ x4 + x1 + x2
       Df Sum of Sq    RSS    AIC F Value     Pr(F)
<none>              47.97 24.974
x4      1      9.93  57.90 25.420   1.863   0.20540
x1      1    820.91 868.88 60.629 154.008 5.781e-07 ***
x2      1     26.79  74.76 28.742   5.026   0.05169 .
---
```

After dropping X4, see if we should add X3.

```
newmod3<-lm(y~x1+x2)
> addterm(newmod3,scope=fullmodel,test = "F")
Single term additions

Model:
y ~ x1 + x2
       Df Sum of Sq    RSS    AIC F Value  Pr(F)
<none>              57.904 25.420
x3      1    9.7939 48.111 25.011  1.8321 0.2089
x4      1    9.9318 47.973 24.974  1.8633 0.2054
```

Stepwise regression: Stopping the procedure

The procedure is stopped when adding an additional predictor does not yield a t-test P-value below $\alpha_E = 0.15$.

Caution about stepwise regression – Multiple Hypothesis Testing

Many t-tests for $\beta_k = 0$ are conducted in a stepwise regression procedure.

The probability is high:

that we included some unimportant predictors

that we excluded some important predictors

Drawbacks of stepwise regression

The final model is not guaranteed to be optimal in any specified sense.

The procedure yields a single final model, although in practice there are often several equally good models.

It doesnt take into account a researchers knowledge about the predictors