Exam 2

Jyoti Chaudhary

March 30, 2018

```
library(mixlm)
```

```
## Warning: package 'mixlm' was built under R version 3.4.4
```

```
##
## Attaching package: 'mixlm'
```

```
## The following objects are masked from 'package:stats':
##
##     glm, lm
```

```
data1 <- read.csv(paste(getwd(),"/Test2_Data_xr17049.csv",sep = ""),header=T)

full=lm(y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10 , data=data1)

forward(full, alpha = 0.1, full = TRUE)
```

```
## Forward selection, alpha-to-enter: 0.1
##
## Full model: y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10
## <environment: 0x000000000f7f4020>
##
## --= Step 1 =--
## Single term additions
##
## Model:
## y ~ 1
##        Df Sum of Sq     RSS    AIC F value  Pr(>F)
## <none>              10190.5 267.86
## x1      1     56.02 10134.5 269.58  0.2653 0.60884
## x2      1    781.38  9409.1 265.87  3.9862 0.05156 .
## x3      1    654.10  9536.4 266.54  3.2923 0.07586 .
## x4      1     92.78 10097.7 269.40  0.4410 0.50981
## x5      1   1082.84  9107.7 264.24  5.7069 0.02088 *
## x6      1     23.46 10167.0 269.74  0.1108 0.74073
## x7      1     33.99 10156.5 269.69  0.1606 0.69037
## x8      1     23.18 10167.3 269.75  0.1094 0.74223
## x9      1    727.14  9463.4 266.16  3.6882 0.06075 .
## x10     1     22.80 10167.7 269.75  0.1077 0.74426
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## --= Step 2 =--
## Single term additions
##
## Model:
## y ~ x5
##        Df Sum of Sq    RSS    AIC F value  Pr(>F)
## <none>              9107.7 264.24
## x1      1    131.22 8976.4 265.52  0.6871 0.41136
## x2      1    921.06 8186.6 260.91  5.2879 0.02596 *
## x3      1    467.32 8640.3 263.61  2.5420 0.11755
## x4      1    179.96 8927.7 265.25  0.9474 0.33536
## x6      1     50.11 9057.5 265.97  0.2600 0.61249
## x7      1     21.92 9085.7 266.12  0.1134 0.73780
## x8      1     50.10 9057.6 265.97  0.2600 0.61252
## x9      1    436.91 8670.7 263.78  2.3683 0.13053
## x10     1      3.70 9104.0 266.22  0.0191 0.89072
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## --= Step 3 =--
## Single term additions
##
## Model:
## y ~ x5 + x2
##        Df Sum of Sq    RSS    AIC F value  Pr(>F)
## <none>              8186.6 260.91
```

```
## x1      1     94.29 8092.3 262.33  0.5360 0.46781
## x3      1    140.33 8046.3 262.05  0.8022 0.37509
## x4      1    146.06 8040.5 262.01  0.8356 0.36542
## x6      1      1.36 8185.2 262.90  0.0076 0.93077
## x7      1      6.90 8179.7 262.87  0.0388 0.84475
## x8      1     10.42 8176.2 262.85  0.0586 0.80976
## x9      1    698.51 7488.1 258.45  4.2910 0.04395 *
## x10     1     22.47 8164.1 262.77  0.1266 0.72363
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## --= Step 4 =--
## Single term additions
##
## Model:
## y ~ x5 + x2 + x9
##        Df Sum of Sq    RSS    AIC F value Pr(>F)
## <none>              7488.1 258.45
## x1      1   159.865 7328.2 259.37  0.9817 0.3271
## x3      1   109.455 7378.6 259.72  0.6675 0.4182
## x4      1   133.946 7354.1 259.55  0.8196 0.3701
## x6      1     1.255 7486.8 260.44  0.0075 0.9312
## x7      1    61.560 7426.5 260.04  0.3730 0.5444
## x8      1     0.334 7487.8 260.45  0.0020 0.9645
## x10     1    32.153 7455.9 260.24  0.1941 0.6617
```

```
##
## Call:
## lm(formula = y ~ x5 + x2 + x9, data = data1)
##
## Coefficients:
## (Intercept)           x5           x2           x9
##    124.7553      -0.3020      -0.3822      -0.2725
```

```
backward(full, alpha = 0.10, full = TRUE, hierarchy = TRUE)
```

```
## Backward elimination, alpha-to-remove: 0.1
##
## Full model: y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10
## <environment: 0x0000000016eadde0>
##
## --= Step 1 =--
## Single term deletions
##
## Model:
## y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10
##        Df Sum of Sq    RSS    AIC F value  Pr(>F)
## <none>              7111.6 269.87
## x1      1     14.59 7126.2 267.98  0.0800 0.77880
## x2      1    797.48 7909.1 273.19  4.3734 0.04307 *
## x3      1     31.72 7143.3 268.10  0.1740 0.67890
## x4      1    104.68 7216.3 268.60  0.5741 0.45319
## x5      1    933.79 8045.4 274.04  5.1209 0.02928 *
## x6      1     16.26 7127.9 267.99  0.0891 0.76685
## x7      1     37.16 7148.8 268.13  0.2038 0.65418
## x8      1     14.53 7126.1 267.98  0.0797 0.77925
## x9      1    735.71 7847.3 272.80  4.0346 0.05154 .
## x10     1     82.63 7194.2 268.45  0.4532 0.50481
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## --= Step 2 =--
## Single term deletions
##
## Model:
## y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x9 + x10
##        Df Sum of Sq    RSS    AIC F value  Pr(>F)
## <none>              7126.1 267.98
## x1      1     14.43 7140.6 266.08  0.0810 0.77739
## x2      1    816.15 7942.3 271.40  4.5812 0.03848 *
## x3      1     33.21 7159.3 266.21  0.1864 0.66824
## x4      1    105.95 7232.1 266.71  0.5947 0.44513
## x5      1    921.15 8047.3 272.05  5.1706 0.02841 *
## x6      1     14.52 7140.7 266.08  0.0815 0.77673
## x7      1     31.66 7157.8 266.20  0.1777 0.67559
## x9      1    749.71 7875.8 270.98  4.2083 0.04682 *
## x10     1     68.38 7194.5 266.45  0.3838 0.53909
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## --= Step 3 =--
## Single term deletions
##
## Model:
## y ~ x2 + x3 + x4 + x5 + x6 + x7 + x9 + x10
##        Df Sum of Sq    RSS    AIC F value  Pr(>F)
## <none>              7140.6 266.08
```

```
## x2      1    803.85 7944.4 269.41  4.6156 0.03764 *
## x3      1     53.98 7194.5 264.45  0.3099 0.58075
## x4      1    146.87 7287.4 265.09  0.8433 0.36383
## x5      1    910.81 8051.4 270.08  5.2298 0.02744 *
## x6      1     26.07 7166.6 264.26  0.1497 0.70083
## x7      1     44.65 7185.2 264.39  0.2564 0.61532
## x9      1    738.80 7879.4 269.00  4.2421 0.04582 *
## x10     1     84.21 7224.8 264.66  0.4835 0.49076
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## --= Step 4 =--
## Single term deletions
##
## Model:
## y ~ x2 + x3 + x4 + x5 + x7 + x9 + x10
##        Df Sum of Sq    RSS    AIC F value  Pr(>F)
## <none>              7166.6 264.26
## x2      1    901.63 8068.3 268.18  5.2840 0.02657 *
## x3      1     63.63 7230.3 262.70  0.3729 0.54473
## x4      1    123.88 7290.5 263.12  0.7260 0.39902
## x5      1    885.31 8051.9 268.08  5.1883 0.02789 *
## x7      1     47.55 7214.2 262.59  0.2787 0.60036
## x9      1    719.26 7885.9 267.04  4.2152 0.04633 *
## x10     1     64.10 7230.7 262.70  0.3757 0.54323
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## --= Step 5 =--
## Single term deletions
##
## Model:
## y ~ x2 + x3 + x4 + x5 + x9 + x10
##        Df Sum of Sq    RSS    AIC F value  Pr(>F)
## <none>              7214.2 262.59
## x2      1    855.84 8070.0 266.19  5.1012 0.02904 *
## x3      1     65.91 7280.1 261.04  0.3928 0.53413
## x4      1    142.75 7356.9 261.57  0.8509 0.36146
## x5      1    876.29 8090.5 266.32  5.2231 0.02728 *
## x9      1    674.18 7888.4 265.06  4.0184 0.05133 .
## x10     1     55.82 7270.0 260.97  0.3327 0.56707
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## --= Step 6 =--
## Single term deletions
##
## Model:
## y ~ x2 + x3 + x4 + x5 + x9
##        Df Sum of Sq    RSS    AIC F value  Pr(>F)
## <none>              7270.0 260.97
```

```
## x2      1    814.03 8084.0 264.28  4.9267 0.03164 *
## x3      1     84.14 7354.1 259.55  0.5092 0.47924
## x4      1    108.63 7378.6 259.72  0.6575 0.42182
## x5      1    820.50 8090.5 264.32  4.9659 0.03101 *
## x9      1    660.35 7930.4 263.32  3.9966 0.05179 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## --= Step 7 =--
## Single term deletions
##
## Model:
## y ~ x2 + x4 + x5 + x9
##        Df Sum of Sq    RSS    AIC F value  Pr(>F)
## <none>              7354.1 259.55
## x2      1   1143.10 8497.2 264.77  6.9946 0.01122 *
## x4      1    133.95 7488.1 258.45  0.8196 0.37011
## x5      1    929.68 8283.8 263.50  5.6887 0.02135 *
## x9      1    686.40 8040.5 262.01  4.2001 0.04628 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## --= Step 8 =--
## Single term deletions
##
## Model:
## y ~ x2 + x5 + x9
##        Df Sum of Sq    RSS    AIC F value   Pr(>F)
## <none>              7488.1 258.45
## x2      1   1182.66 8670.7 263.78  7.2652 0.009786 **
## x5      1    863.71 8351.8 261.91  5.3059 0.025824 *
## x9      1    698.51 8186.6 260.91  4.2910 0.043949 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Call:
## lm(formula = y ~ x2 + x5 + x9, data = data1)
##
## Coefficients:
## (Intercept)           x2           x5           x9
##    124.7553      -0.3822      -0.3020      -0.2725
```

```
stepWise(full, alpha.enter = 0.1, alpha.remove = 0.1, full = TRUE)
```

```
## Stepwise regression (forward-backward), alpha-to-enter: 0.1, alpha-to-remove: 0.1
##
## Full model: y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10
## <environment: 0x000000000e395480>
##
## --= Step (forward) 1 =--
## Single term additions
##
## Model:
## y ~ 1
##        Df Sum of Sq     RSS    AIC F value  Pr(>F)
## <none>              10190.5 267.86
## x1      1     56.02 10134.5 269.58  0.2653 0.60884
## x2      1    781.38  9409.1 265.87  3.9862 0.05156 .
## x3      1    654.10  9536.4 266.54  3.2923 0.07586 .
## x4      1     92.78 10097.7 269.40  0.4410 0.50981
## x5      1   1082.84  9107.7 264.24  5.7069 0.02088 *
## x6      1     23.46 10167.0 269.74  0.1108 0.74073
## x7      1     33.99 10156.5 269.69  0.1606 0.69037
## x8      1     23.18 10167.3 269.75  0.1094 0.74223
## x9      1    727.14  9463.4 266.16  3.6882 0.06075 .
## x10     1     22.80 10167.7 269.75  0.1077 0.74426
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## --= Step (forward) 2 =--
## Single term additions
##
## Model:
## y ~ x5
##        Df Sum of Sq    RSS    AIC F value  Pr(>F)
## <none>              9107.7 264.24
## x1      1    131.22 8976.4 265.52  0.6871 0.41136
## x2      1    921.06 8186.6 260.91  5.2879 0.02596 *
## x3      1    467.32 8640.3 263.61  2.5420 0.11755
## x4      1    179.96 8927.7 265.25  0.9474 0.33536
## x6      1     50.11 9057.5 265.97  0.2600 0.61249
## x7      1     21.92 9085.7 266.12  0.1134 0.73780
## x8      1     50.10 9057.6 265.97  0.2600 0.61252
## x9      1    436.91 8670.7 263.78  2.3683 0.13053
## x10     1      3.70 9104.0 266.22  0.0191 0.89072
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## --= Step (forward) 3 =--
## Single term additions
##
## Model:
## y ~ x5 + x2
##        Df Sum of Sq    RSS    AIC F value  Pr(>F)
## <none>              8186.6 260.91
```

```
## x1      1     94.29 8092.3 262.33  0.5360 0.46781
## x3      1    140.33 8046.3 262.05  0.8022 0.37509
## x4      1    146.06 8040.5 262.01  0.8356 0.36542
## x6      1      1.36 8185.2 262.90  0.0076 0.93077
## x7      1      6.90 8179.7 262.87  0.0388 0.84475
## x8      1     10.42 8176.2 262.85  0.0586 0.80976
## x9      1    698.51 7488.1 258.45  4.2910 0.04395 *
## x10     1     22.47 8164.1 262.77  0.1266 0.72363
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## --= Step (forward) 4 =--
## Single term additions
##
## Model:
## y ~ x5 + x2 + x9
##        Df Sum of Sq    RSS    AIC F value Pr(>F)
## <none>              7488.1 258.45
## x1      1   159.865 7328.2 259.37  0.9817 0.3271
## x3      1   109.455 7378.6 259.72  0.6675 0.4182
## x4      1   133.946 7354.1 259.55  0.8196 0.3701
## x6      1     1.255 7486.8 260.44  0.0075 0.9312
## x7      1    61.560 7426.5 260.04  0.3730 0.5444
## x8      1     0.334 7487.8 260.45  0.0020 0.9645
## x10     1    32.153 7455.9 260.24  0.1941 0.6617
```

```
##
## Call:
## lm(formula = y ~ x5 + x2 + x9, data = data1)
##
## Coefficients:
## (Intercept)           x5           x2           x9
##    124.7553      -0.3020      -0.3822      -0.2725
```

The final model is y ~ x5 + x2 + x9.

a. If alpha is changed to 0.05 then x9 will still be added to the model as p-value for x9 = 0.04395 which is less than 0.05.

```
final_fit <- lm(y ~ x5 + x2 + x9, data=data1)

#PRESS Residuals
pr <- resid(final_fit)/(1 - lm.influence(final_fit)$hat)
pr_st<-sum(pr^2)
an<-anova(final_fit)
SST<-sum(an$`Sum Sq`)

R2_pred=1-(pr_st/SST)
R2_pred
```

```
## [1] 0.1384949
```

```
AIC(final_fit)
```

```
## [1] 402.3461
```

b. R2_pred is 0.1384949. AIC = 402.3461

```
library(leaps)

tmp =regsubsets(y ~ x5 + x2 + x9, data=data1 ,nbest=10,really.big=T, intercept=T)
names(summary(tmp))
```

```
## [1] "which"  "rsq"    "rss"    "adjr2"  "cp"     "bic"    "outmat" "obj"
```

```
almdl=summary(tmp)[[1]]
SSE=summary(tmp)[[3]]
adjR2=summary(tmp)[[4]]
Cp=summary(tmp)[[5]]
BIC=summary(tmp)[[6]]

fnl=cbind(almdl,SSE,adjR2,Cp,BIC)
fnl
```

```
##   (Intercept) x5 x2 x9      SSE      adjR2        Cp       BIC
## 1           1  1  0  0 9107.660 0.08764016  9.949191 2.2070414
## 1           1  0  1  0 9409.119 0.05744152 11.801081 3.8352161
## 1           1  0  0  1 9463.359 0.05200805 12.134281 4.1226185
## 2           1  1  1  0 8186.599 0.16245863  6.291031 0.7882030
## 2           1  0  1  1 8351.800 0.14555755  7.305874 1.7871280
## 2           1  1  0  1 8670.746 0.11292734  9.265188 3.6610149
## 3           1  1  1  1 7488.086 0.21726710  4.000000 0.2409585
```