# MATH405 – HW 19

Nick Huo

2022-11-02

### a.

```
library(MASS)
dd <- read.csv("HW19.csv")
```

### b.

```
empty.mod <- lm(y ~ 1, data=dd)
full.mod <- lm(y ~ ., data=dd)
stepAIC(empty.mod, scope=list(lower=empty.mod, upper=full.mod),
        direction="both")
```

```
## Start:  AIC=7066.3
## y ~ 1
##
##        Df Sum of Sq      RSS    AIC
## + x4    1  12824133  4077740 6073.0
## + x3    1  12811948  4089924 6075.1
## + x1    1   2215968 14685904 6969.9
## + x13   1    877116 16024756 7031.0
## + x11   1     83273 16818599 7064.8
## + x7    1     81818 16820054 7064.9
## + x8    1     54483 16847390 7066.0
## + x12   1     52203 16849669 7066.1
## <none>              16901872 7066.3
## + x10   1     39527 16862345 7066.7
## + x20   1     31159 16870713 7067.0
## + x5    1     24932 16876940 7067.3
## + x15   1     18553 16883319 7067.5
## + x19   1     11554 16890319 7067.8
## + x9    1     10607 16891265 7067.9
## + x16   1      7386 16894486 7068.0
## + x18   1      5318 16896554 7068.1
## + x17   1      2690 16899182 7068.2
## + x2    1      1802 16900070 7068.2
## + x6    1       757 16901115 7068.3
## + x14   1       209 16901663 7068.3
##
```

```
## Step:  AIC=6072.98
## y ~ x4
##
##        Df Sum of Sq      RSS    AIC
## + x1    1   2389347  1688392 5457.7
## + x13   1    425142  3652598 5997.9
## + x11   1    124151  3953588 6053.3
## + x8    1     23549  4054191 6070.9
## + x5    1     18094  4059646 6071.9
## <none>               4077740 6073.0
## + x10   1      8697  4069043 6073.5
## + x2    1      6261  4071478 6073.9
## + x16   1      5509  4072230 6074.0
## + x3    1      4908  4072831 6074.1
## + x6    1      3986  4073754 6074.3
## + x18   1      3973  4073767 6074.3
## + x14   1      3170  4074570 6074.4
## + x17   1      3146  4074593 6074.4
## + x19   1      1867  4075872 6074.7
## + x12   1      1119  4076620 6074.8
## + x15   1       970  4076769 6074.8
## + x9    1       845  4076894 6074.8
## + x20   1       434  4077305 6074.9
## + x7    1       156  4077584 6075.0
## - x4    1  12824133 16901872 7066.3
##
## Step:  AIC=5457.74
## y ~ x4 + x1
##
##        Df Sum of Sq      RSS    AIC
## + x13   1    465148  1223244 5234.2
## + x11   1     70155  1618237 5430.0
## + x5    1     35558  1652834 5444.8
## + x2    1     25390  1663002 5449.1
## + x15   1     19102  1669290 5451.8
## + x8    1     14323  1674069 5453.8
## + x3    1      5799  1682593 5457.3
## <none>               1688392 5457.7
## + x19   1      4272  1684120 5458.0
## + x7    1      4009  1684384 5458.1
## + x20   1      3517  1684876 5458.3
## + x14   1      2226  1686167 5458.8
## + x16   1      1753  1686639 5459.0
## + x12   1      1511  1686881 5459.1
## + x17   1      1137  1687255 5459.3
## + x18   1       580  1687812 5459.5
## + x9    1       168  1688225 5459.7
## + x10   1        94  1688298 5459.7
## + x6    1        30  1688363 5459.7
## - x1    1   2389347  4077740 6073.0
## - x4    1  12997511 14685904 6969.9
##
## Step:  AIC=5234.16
## y ~ x4 + x1 + x13
```

```
##
##        Df Sum of Sq      RSS    AIC
## + x11   1     55823  1167421 5203.5
## + x5    1     36916  1186328 5214.7
## + x2    1     16747  1206497 5226.5
## + x15   1     10577  1212668 5230.1
## + x7    1      8805  1214440 5231.1
## + x14   1      5407  1217838 5233.1
## + x16   1      5345  1217899 5233.1
## + x8    1      3772  1219472 5234.0
## + x20   1      3715  1219529 5234.0
## <none>               1223244 5234.2
## + x18   1      2530  1220715 5234.7
## + x3    1      2202  1221042 5234.9
## + x12   1      1221  1222023 5235.5
## + x19   1       955  1222289 5235.6
## + x9    1       696  1222548 5235.8
## + x17   1       363  1222881 5235.9
## + x6    1       161  1223084 5236.1
## + x10   1        78  1223167 5236.1
## - x13   1    465148  1688392 5457.7
## - x1    1   2429354  3652598 5997.9
## - x4    1  12526981 13750225 6925.8
##
## Step:  AIC=5203.46
## y ~ x4 + x1 + x13 + x11
##
##        Df Sum of Sq      RSS    AIC
## + x5    1     41678  1125743 5180.0
## + x2    1     15843  1151578 5195.9
## + x7    1     11817  1155604 5198.3
## + x15   1     11007  1156414 5198.8
## + x16   1      5394  1162027 5202.2
## + x20   1      3885  1163536 5203.1
## + x14   1      3773  1163648 5203.2
## <none>               1167421 5203.5
## + x18   1      2325  1165096 5204.1
## + x12   1      2181  1165240 5204.2
## + x8    1      1767  1165654 5204.4
## + x3    1      1697  1165724 5204.4
## + x9    1      1057  1166364 5204.8
## + x6    1       171  1167250 5205.4
## + x17   1       134  1167287 5205.4
## + x10   1        65  1167356 5205.4
## + x19   1        52  1167369 5205.4
## - x11   1     55823  1223244 5234.2
## - x13   1    450816  1618237 5430.0
## - x1    1   2378999  3546420 5979.3
## - x4    1  12555724 13723145 6926.5
##
## Step:  AIC=5180.01
## y ~ x4 + x1 + x13 + x11 + x5
##
##        Df Sum of Sq      RSS    AIC
```

```
## + x2    1     15326  1110418 5172.4
## + x15   1     12347  1113396 5174.3
## + x7    1     11385  1114359 5174.9
## + x16   1      7809  1117935 5177.1
## + x20   1      4135  1121609 5179.4
## <none>               1125743 5180.0
## + x18   1      2993  1122751 5180.1
## + x14   1      2946  1122798 5180.2
## + x3    1      1833  1123911 5180.9
## + x8    1      1277  1124467 5181.2
## + x12   1      1158  1124585 5181.3
## + x9    1       885  1124858 5181.5
## + x6    1       365  1125379 5181.8
## + x10   1       286  1125458 5181.8
## + x17   1       148  1125596 5181.9
## + x19   1        81  1125663 5182.0
## - x5    1     41678  1167421 5203.5
## - x11   1     60585  1186328 5214.7
## - x13   1    451668  1577411 5414.2
## - x1    1   2396797  3522541 5976.5
## - x4    1  12547588 13673331 6925.9
##
## Step:  AIC=5172.42
## y ~ x4 + x1 + x13 + x11 + x5 + x2
##
##        Df Sum of Sq      RSS    AIC
## + x7    1     12951  1097466 5166.2
## + x15   1     12719  1097698 5166.4
## + x16   1      5438  1104980 5171.0
## + x20   1      4300  1106118 5171.7
## + x14   1      3661  1106757 5172.1
## <none>               1110418 5172.4
## + x18   1      2792  1107626 5172.7
## + x3    1      1685  1108733 5173.4
## + x12   1      1479  1108938 5173.5
## + x8    1      1187  1109231 5173.7
## + x9    1      1008  1109410 5173.8
## + x10   1       400  1110018 5174.2
## + x6    1       379  1110039 5174.2
## + x17   1       199  1110219 5174.3
## + x19   1       102  1110315 5174.4
## - x2    1     15326  1125743 5180.0
## - x5    1     41160  1151578 5195.9
## - x11   1     59627  1170044 5207.0
## - x13   1    443624  1554042 5405.7
## - x1    1   2410180  3520598 5978.1
## - x4    1  12562796 13673214 6927.9
##
## Step:  AIC=5166.2
## y ~ x4 + x1 + x13 + x11 + x5 + x2 + x7
##
##        Df Sum of Sq      RSS    AIC
## + x15   1     12059  1085408 5160.5
## + x16   1      4847  1092620 5165.1
```

```
## + x18   1      4099  1093367 5165.6
## + x20   1      3638  1093828 5165.9
## + x14   1      3152  1094314 5166.2
## <none>               1097466 5166.2
## + x3    1      1250  1096216 5167.4
## + x9    1      1100  1096367 5167.5
## + x8    1      1061  1096406 5167.5
## + x12   1      1042  1096424 5167.5
## + x6    1       507  1096960 5167.9
## + x17   1       266  1097201 5168.0
## + x10   1       239  1097227 5168.1
## + x19   1       238  1097229 5168.1
## - x7    1     12951  1110418 5172.4
## - x2    1     16892  1114359 5174.9
## - x5    1     40675  1138141 5189.7
## - x11   1     62824  1160290 5203.2
## - x13   1    448948  1546414 5404.3
## - x1    1   2419169  3516636 5979.4
## - x4    1  12542019 13639486 6928.2
##
## Step:  AIC=5160.47
## y ~ x4 + x1 + x13 + x11 + x5 + x2 + x7 + x15
##
##        Df Sum of Sq      RSS    AIC
## + x18   1      5069  1080339 5159.2
## + x16   1      4124  1081284 5159.8
## + x20   1      3967  1081441 5159.9
## <none>               1085408 5160.5
## + x14   1      2957  1082451 5160.6
## + x8    1      1593  1083815 5161.4
## + x3    1      1307  1084100 5161.6
## + x12   1      1079  1084329 5161.8
## + x9    1       659  1084749 5162.0
## + x6    1       493  1084915 5162.2
## + x19   1       408  1085000 5162.2
## + x10   1       220  1085187 5162.3
## + x17   1       124  1085284 5162.4
## - x15   1     12059  1097466 5166.2
## - x7    1     12290  1097698 5166.4
## - x2    1     17230  1102638 5169.5
## - x5    1     41990  1127398 5185.0
## - x11   1     63283  1148691 5198.1
## - x13   1    439764  1525172 5396.6
## - x1    1   2431216  3516624 5981.4
## - x4    1  12513742 13599150 6928.1
##
## Step:  AIC=5159.19
## y ~ x4 + x1 + x13 + x11 + x5 + x2 + x7 + x15 + x18
##
##        Df Sum of Sq      RSS    AIC
## + x16   1      4465  1075874 5158.3
## + x20   1      3901  1076438 5158.7
## <none>               1080339 5159.2
## + x14   1      3015  1077324 5159.2
```

```
## + x8    1      1524  1078816 5160.2
## + x3    1      1351  1078989 5160.3
## - x18   1      5069  1085408 5160.5
## + x12   1       899  1079441 5160.6
## + x9    1       478  1079861 5160.9
## + x6    1       466  1079873 5160.9
## + x19   1       395  1079944 5160.9
## + x10   1       297  1080042 5161.0
## + x17   1       109  1080230 5161.1
## - x15   1     13028  1093367 5165.6
## - x7    1     13707  1094046 5166.0
## - x2    1     17059  1097398 5168.2
## - x5    1     42910  1123249 5184.5
## - x11   1     63262  1143601 5197.0
## - x13   1    442842  1523182 5397.7
## - x1    1   2426971  3507310 5981.5
## - x4    1  12513365 13593704 6929.8
##
## Step:  AIC=5158.29
## y ~ x4 + x1 + x13 + x11 + x5 + x2 + x7 + x15 + x18 + x16
##
##        Df Sum of Sq      RSS    AIC
## + x20   1      4203  1071671 5157.6
## <none>               1075874 5158.3
## + x14   1      2824  1073051 5158.5
## - x16   1      4465  1080339 5159.2
## + x8    1      1541  1074333 5159.3
## + x3    1      1186  1074688 5159.5
## + x12   1       987  1074888 5159.7
## - x18   1      5409  1081284 5159.8
## + x9    1       696  1075178 5159.8
## + x6    1       503  1075371 5160.0
## + x19   1       455  1075420 5160.0
## + x10   1       293  1075581 5160.1
## + x17   1       202  1075672 5160.2
## - x15   1     12284  1088159 5164.2
## - x7    1     13200  1089074 5164.8
## - x2    1     14684  1090559 5165.8
## - x5    1     44740  1120615 5184.8
## - x11   1     63405  1139279 5196.4
## - x13   1    446369  1522243 5399.2
## - x1    1   2418731  3494605 5981.0
## - x4    1  12504156 13580030 6931.1
##
## Step:  AIC=5157.55
## y ~ x4 + x1 + x13 + x11 + x5 + x2 + x7 + x15 + x18 + x16 + x20
##
##        Df Sum of Sq      RSS    AIC
## <none>               1071671 5157.6
## + x14   1      2829  1068842 5157.7
## - x20   1      4203  1075874 5158.3
## + x8    1      1643  1070028 5158.5
## - x16   1      4767  1076438 5158.7
## + x3    1      1187  1070484 5158.8
```

```
## + x12   1       930  1070741 5158.9
## + x9    1       803  1070868 5159.0
## - x18   1      5350  1077021 5159.0
## + x19   1       440  1071231 5159.3
## + x6    1       404  1071267 5159.3
## + x10   1       283  1071388 5159.4
## + x17   1       226  1071445 5159.4
## - x7    1     12449  1084120 5163.6
## - x15   1     12596  1084267 5163.7
## - x2    1     14740  1086411 5165.1
## - x5    1     45092  1116763 5184.4
## - x11   1     63524  1135195 5195.9
## - x13   1    446387  1518058 5399.3
## - x1    1   2422220  3493891 5982.8
## - x4    1  12454037 13525708 6930.3

##
## Call:
## lm(formula = y ~ x4 + x1 + x13 + x11 + x5 + x2 + x7 + x15 + x18 +
##     x16 + x20, data = dd)
##
## Coefficients:
## (Intercept)           x4           x1          x13          x11           x5
##     47.2679       4.4933       4.0507       0.4154      -1.4447       1.5190
##          x2           x7          x15          x18          x16          x20
##      1.6052       0.2110       9.1290       0.2393       2.6715      -8.6540
```

**c.**

```
stepAIC(empty.mod, scope=list(lower=empty.mod, upper=full.mod),
        direction="both", k=log(length(dd)))
```

```
## Start:  AIC=7067.34
## y ~ 1
##
##        Df Sum of Sq      RSS    AIC
## + x4    1  12824133  4077740 6075.1
## + x3    1  12811948  4089924 6077.2
## + x1    1   2215968 14685904 6972.0
## + x13   1    877116 16024756 7033.1
## + x11   1     83273 16818599 7066.9
## + x7    1     81818 16820054 7067.0
## <none>              16901872 7067.3
## + x8    1     54483 16847390 7068.1
## + x12   1     52203 16849669 7068.2
## + x10   1     39527 16862345 7068.7
## + x20   1     31159 16870713 7069.1
## + x5    1     24932 16876940 7069.4
## + x15   1     18553 16883319 7069.6
## + x19   1     11554 16890319 7069.9
## + x9    1     10607 16891265 7069.9
```

```
## + x16   1      7386 16894486 7070.1
## + x18   1      5318 16896554 7070.2
## + x17   1      2690 16899182 7070.3
## + x2    1      1802 16900070 7070.3
## + x6    1       757 16901115 7070.4
## + x14   1       209 16901663 7070.4
##
## Step:  AIC=6075.07
## y ~ x4
##
##        Df Sum of Sq      RSS    AIC
## + x1    1   2389347  1688392 5460.9
## + x13   1    425142  3652598 6001.0
## + x11   1    124151  3953588 6056.5
## + x8    1     23549  4054191 6074.1
## + x5    1     18094  4059646 6075.0
## <none>               4077740 6075.1
## + x10   1      8697  4069043 6076.6
## + x2    1      6261  4071478 6077.0
## + x16   1      5509  4072230 6077.2
## + x3    1      4908  4072831 6077.3
## + x6    1      3986  4073754 6077.4
## + x18   1      3973  4073767 6077.4
## + x14   1      3170  4074570 6077.6
## + x17   1      3146  4074593 6077.6
## + x19   1      1867  4075872 6077.8
## + x12   1      1119  4076620 6077.9
## + x15   1       970  4076769 6077.9
## + x9    1       845  4076894 6078.0
## + x20   1       434  4077305 6078.0
## + x7    1       156  4077584 6078.1
## - x4    1  12824133 16901872 7067.3
##
## Step:  AIC=5460.88
## y ~ x4 + x1
##
##        Df Sum of Sq      RSS    AIC
## + x13   1    465148  1223244 5238.3
## + x11   1     70155  1618237 5434.2
## + x5    1     35558  1652834 5449.0
## + x2    1     25390  1663002 5453.3
## + x15   1     19102  1669290 5456.0
## + x8    1     14323  1674069 5458.0
## <none>               1688392 5460.9
## + x3    1      5799  1682593 5461.5
## + x19   1      4272  1684120 5462.1
## + x7    1      4009  1684384 5462.3
## + x20   1      3517  1684876 5462.5
## + x14   1      2226  1686167 5463.0
## + x16   1      1753  1686639 5463.2
## + x12   1      1511  1686881 5463.3
## + x17   1      1137  1687255 5463.5
## + x18   1       580  1687812 5463.7
## + x9    1       168  1688225 5463.9
```

```
## + x10   1        94  1688298 5463.9
## + x6    1        30  1688363 5463.9
## - x1    1   2389347  4077740 6075.1
## - x4    1  12997511 14685904 6972.0
##
## Step:  AIC=5238.33
## y ~ x4 + x1 + x13
##
##        Df Sum of Sq      RSS    AIC
## + x11   1     55823  1167421 5208.7
## + x5    1     36916  1186328 5219.9
## + x2    1     16747  1206497 5231.7
## + x15   1     10577  1212668 5235.3
## + x7    1      8805  1214440 5236.3
## + x14   1      5407  1217838 5238.3
## + x16   1      5345  1217899 5238.3
## <none>               1223244 5238.3
## + x8    1      3772  1219472 5239.2
## + x20   1      3715  1219529 5239.2
## + x18   1      2530  1220715 5239.9
## + x3    1      2202  1221042 5240.1
## + x12   1      1221  1222023 5240.7
## + x19   1       955  1222289 5240.8
## + x9    1       696  1222548 5241.0
## + x17   1       363  1222881 5241.2
## + x6    1       161  1223084 5241.3
## + x10   1        78  1223167 5241.3
## - x13   1    465148  1688392 5460.9
## - x1    1   2429354  3652598 6001.0
## - x4    1  12526981 13750225 6929.0
##
## Step:  AIC=5208.68
## y ~ x4 + x1 + x13 + x11
##
##        Df Sum of Sq      RSS    AIC
## + x5    1     41678  1125743 5186.3
## + x2    1     15843  1151578 5202.2
## + x7    1     11817  1155604 5204.6
## + x15   1     11007  1156414 5205.1
## + x16   1      5394  1162027 5208.5
## <none>               1167421 5208.7
## + x20   1      3885  1163536 5209.4
## + x14   1      3773  1163648 5209.5
## + x18   1      2325  1165096 5210.3
## + x12   1      2181  1165240 5210.4
## + x8    1      1767  1165654 5210.7
## + x3    1      1697  1165724 5210.7
## + x9    1      1057  1166364 5211.1
## + x6    1       171  1167250 5211.6
## + x17   1       134  1167287 5211.6
## + x10   1        65  1167356 5211.7
## + x19   1        52  1167369 5211.7
## - x11   1     55823  1223244 5238.3
## - x13   1    450816  1618237 5434.2
```

```
## - x1    1   2378999  3546420 5983.4
## - x4    1  12555724 13723145 6930.6
##
## Step:  AIC=5186.28
## y ~ x4 + x1 + x13 + x11 + x5
##
##        Df Sum of Sq      RSS    AIC
## + x2    1     15326  1110418 5179.7
## + x15   1     12347  1113396 5181.6
## + x7    1     11385  1114359 5182.2
## + x16   1      7809  1117935 5184.5
## <none>               1125743 5186.3
## + x20   1      4135  1121609 5186.7
## + x18   1      2993  1122751 5187.5
## + x14   1      2946  1122798 5187.5
## + x3    1      1833  1123911 5188.2
## + x8    1      1277  1124467 5188.5
## + x12   1      1158  1124585 5188.6
## + x9    1       885  1124858 5188.8
## + x6    1       365  1125379 5189.1
## + x10   1       286  1125458 5189.1
## + x17   1       148  1125596 5189.2
## + x19   1        81  1125663 5189.3
## - x5    1     41678  1167421 5208.7
## - x11   1     60585  1186328 5219.9
## - x13   1    451668  1577411 5419.4
## - x1    1   2396797  3522541 5981.8
## - x4    1  12547588 13673331 6931.1
##
## Step:  AIC=5179.73
## y ~ x4 + x1 + x13 + x11 + x5 + x2
##
##        Df Sum of Sq      RSS    AIC
## + x7    1     12951  1097466 5174.6
## + x15   1     12719  1097698 5174.7
## + x16   1      5438  1104980 5179.3
## <none>               1110418 5179.7
## + x20   1      4300  1106118 5180.1
## + x14   1      3661  1106757 5180.5
## + x18   1      2792  1107626 5181.0
## + x3    1      1685  1108733 5181.7
## + x12   1      1479  1108938 5181.8
## + x8    1      1187  1109231 5182.0
## + x9    1      1008  1109410 5182.1
## + x10   1       400  1110018 5182.5
## + x6    1       379  1110039 5182.5
## + x17   1       199  1110219 5182.6
## + x19   1       102  1110315 5182.7
## - x2    1     15326  1125743 5186.3
## - x5    1     41160  1151578 5202.2
## - x11   1     59627  1170044 5213.3
## - x13   1    443624  1554042 5412.0
## - x1    1   2410180  3520598 5984.4
## - x4    1  12562796 13673214 6934.2
```

```
##
## Step:  AIC=5174.56
## y ~ x4 + x1 + x13 + x11 + x5 + x2 + x7
##
##        Df Sum of Sq      RSS    AIC
## + x15   1     12059  1085408 5169.9
## + x16   1      4847  1092620 5174.5
## <none>               1097466 5174.6
## + x18   1      4099  1093367 5175.0
## + x20   1      3638  1093828 5175.3
## + x14   1      3152  1094314 5175.6
## + x3    1      1250  1096216 5176.8
## + x9    1      1100  1096367 5176.9
## + x8    1      1061  1096406 5176.9
## + x12   1      1042  1096424 5176.9
## + x6    1       507  1096960 5177.3
## + x17   1       266  1097201 5177.4
## + x10   1       239  1097227 5177.5
## + x19   1       238  1097229 5177.5
## - x7    1     12951  1110418 5179.7
## - x2    1     16892  1114359 5182.2
## - x5    1     40675  1138141 5197.0
## - x11   1     62824  1160290 5210.5
## - x13   1    448948  1546414 5411.6
## - x1    1   2419169  3516636 5986.7
## - x4    1  12542019 13639486 6935.5
##
## Step:  AIC=5169.87
## y ~ x4 + x1 + x13 + x11 + x5 + x2 + x7 + x15
##
##        Df Sum of Sq      RSS    AIC
## + x18   1      5069  1080339 5169.6
## <none>               1085408 5169.9
## + x16   1      4124  1081284 5170.3
## + x20   1      3967  1081441 5170.4
## + x14   1      2957  1082451 5171.0
## + x8    1      1593  1083815 5171.9
## + x3    1      1307  1084100 5172.1
## + x12   1      1079  1084329 5172.2
## + x9    1       659  1084749 5172.5
## + x6    1       493  1084915 5172.6
## + x19   1       408  1085000 5172.7
## + x10   1       220  1085187 5172.8
## + x17   1       124  1085284 5172.8
## - x15   1     12059  1097466 5174.6
## - x7    1     12290  1097698 5174.7
## - x2    1     17230  1102638 5177.9
## - x5    1     41990  1127398 5193.4
## - x11   1     63283  1148691 5206.5
## - x13   1    439764  1525172 5404.9
## - x1    1   2431216  3516624 5989.7
## - x4    1  12513742 13599150 6936.5
##
## Step:  AIC=5169.64
```

```
## y ~ x4 + x1 + x13 + x11 + x5 + x2 + x7 + x15 + x18
##
##        Df Sum of Sq      RSS    AIC
## <none>               1080339 5169.6
## + x16   1      4465  1075874 5169.8
## - x18   1      5069  1085408 5169.9
## + x20   1      3901  1076438 5170.2
## + x14   1      3015  1077324 5170.7
## + x8    1      1524  1078816 5171.7
## + x3    1      1351  1078989 5171.8
## + x12   1       899  1079441 5172.1
## + x9    1       478  1079861 5172.4
## + x6    1       466  1079873 5172.4
## + x19   1       395  1079944 5172.4
## + x10   1       297  1080042 5172.5
## + x17   1       109  1080230 5172.6
## - x15   1     13028  1093367 5175.0
## - x7    1     13707  1094046 5175.4
## - x2    1     17059  1097398 5177.6
## - x5    1     42910  1123249 5193.9
## - x11   1     63262  1143601 5206.4
## - x13   1    442842  1523182 5407.1
## - x1    1   2426971  3507310 5990.9
## - x4    1  12513365 13593704 6939.2

##
## Call:
## lm(formula = y ~ x4 + x1 + x13 + x11 + x5 + x2 + x7 + x15 + x18,
##     data = dd)
##
## Coefficients:
## (Intercept)           x4           x1          x13          x11           x5
##     43.3801       4.4985       4.0517       0.4131      -1.4417       1.4776
##          x2           x7          x15          x18
##      1.7132       0.2210       9.2717       0.2328
```

## d.

The model we found using BIC for Step-wise Selection is more parsimonious because it has 9 predictors, fewer than the model found using AIC, which has 11 predictors.

## e.

### AIC Model – 2-fold cross-validation

```
set.seed(893745)
gp1 <- sample(1:700, size=350, replace=FALSE)
gp2 <- (1:700)[-gp1]
```

### Randomly split the data into train and test groups

```
aic_model1 <- lm(y ~ x4 + x1 + x13 + x11 + x5 + x2 + x7 + x15 +
       x18 + x16 + x20, data = dd[gp1,])
aic_gp2_pred <- predict(aic_model1, dd[gp2,])
aic_gp2_y <- dd[gp2,1]
mspe_aic_mod1 <- mean((aic_gp2_y - aic_gp2_pred)^2)
```

**Calculate MSPE: model trained on group 1, test on group 2**

```
aic_model2 <- lm(y ~ x4 + x1 + x13 + x11 + x5 + x2 + x7 + x15 +
       x18 + x16 + x20, data = dd[gp2,])
aic_gp1_pred <- predict(aic_model1, dd[gp1,])
aic_gp1_y <- dd[gp1,1]
mspe_aic_mod2 <- mean((aic_gp1_y - aic_gp1_pred)^2)
```

**Calculate MSPE: model trained on group 2, test on group 1**

```
mean(mspe_aic_mod1, mspe_aic_mod2)
```

**AIC Model Overall MSPE**

```
## [1] 1615.653
```

**BIC Model – 2-fold cross-validation**

```
bic_model1 <- lm(formula = y ~ x4 + x1 + x13 + x11 + x5 + x2 + x7 + x15 + x18,
                 data=dd[gp1,])

bic_gp2_pred <- predict(bic_model1, dd[gp2,])
bic_gp2_y <- dd[gp2,1]
mspe_bic_mod1 <- mean((bic_gp2_y - bic_gp2_pred)^2)
```

**Calculate MSPE: model trained on group 1, test on group 2**

```
bic_model2 <- lm(formula = y ~ x4 + x1 + x13 + x11 + x5 + x2 + x7 + x15 + x18,
                 data=dd[gp2,])
bic_gp1_pred <- predict(bic_model1, dd[gp1,])
bic_gp1_y <- dd[gp1,1]
mspe_bic_mod2 <- mean((bic_gp1_y - bic_gp1_pred)^2)
```

**Calculate MSPE: model trained on group 2, test on group 1**

```
mean(mspe_bic_mod1, mspe_bic_mod2)
```

**AIC Model Overall MSPE**

```
## [1] 1625.888
```

## f.

```
tab_df <- data.frame(MSPE=c(1615.653, 1625.888),
                     row.names=c("AIC Model","BIC Model"))
knitr::kable(tab_df)
```

| | MSPE |
|---|---|
| AIC Model | 1615.653 |
| BIC Model | 1625.888 |

The AIC model has a slightly lower MSPE. So the AIC model (with more predictors) seem to be better at prediction. Also noticing that