WU #15 - Stepwise

Math 158 - Jo Hardin

Thursday 3/24/2022

Name: _____

1. Calculate the value 161.96. Hints: the penalty is `log(n)` which means the calculation is BIC, not AIC; there are 941 observations; RSS is SSE.

$$BIC_p = n \ln(SSE_p) - n \cdot \ln(n) + \ln(n) \cdot p$$

2. Is `Sum of Sq` the same as SSR or extra sums of squares? (Hint: look at `<none>`)
3. Why aren't there any interaction terms for the first few iterations?
4. Why, toward the end, are the pluses and minuses mixed up?

```
stats::step(lm(weight ~ 1, data = births14),
    weight ~ (mage + mature + weeks + premie + gained  + term + sex + habit + marital + whitemom)^2,
    direction = "both", k = log(941))  # BIC !!
```

```
## Start:  AIC=457.43
## weight ~ 1
##
##
##            Df Sum of Sq    RSS    AIC
## + weeks     1    417.37 1101.6 161.96
## + premie    1    338.28 1180.7 227.20
## + term      2    237.25 1281.7 311.31
## + gained    1     33.86 1485.1 443.06
## + sex       1     33.44 1485.5 443.33
## + habit     1     27.83 1491.1 446.87
## + whitemom  1     24.58 1494.4 448.92
## + marital   1     12.33 1506.6 456.60
## <none>                  1519.0 457.43
## + mage      1      6.31 1512.6 460.36
## + mature    1      4.58 1514.4 461.43
##
## Step:  AIC=161.96
## weight ~ weeks
##
##            Df Sum of Sq    RSS    AIC
## + sex       1     37.70 1063.9 136.03
## + term      2     37.80 1063.8 142.79
## + premie    1     25.02 1076.6 147.18
## + gained    1     24.74 1076.8 147.43
## + whitemom  1     13.72 1087.9 157.01
## + habit     1     13.32 1088.3 157.35
## <none>                  1101.6 161.96
## + marital   1      7.51 1094.1 162.36
## + mage      1      7.47 1094.1 162.40
## + mature    1      7.32 1094.3 162.53
## - weeks     1    417.37 1519.0 457.43
##
## Step:  AIC=136.03
## weight ~ weeks + sex
##
##            Df Sum of Sq    RSS    AIC
## + term      2     41.31 1022.6 112.46
## + premie    1     28.13 1035.8 117.67
```

```
## + gained      1      23.66 1040.2 121.72
## + habit       1      14.34 1049.5 130.11
## + whitemom    1      11.95 1051.9 132.25
## + marital     1       9.30 1054.6 134.62
## + mage        1       8.84 1055.0 135.03
## + mature      1       8.25 1055.6 135.56
## <none>                      1063.9 136.03
## + weeks:sex  1       0.36 1063.5 142.56
## - sex         1      37.70 1101.6 161.96
## - weeks       1     421.63 1485.5 443.33
##
## Step:  AIC=112.46
## weight ~ weeks + sex + term
##
##               Df Sum of Sq    RSS     AIC
## + weeks:term  2     41.756  980.81  86.923
## + gained      1     25.456  997.11  95.586
## + whitemom    1     12.294 1010.28 107.926
## + habit       1     11.829 1010.74 108.359
## + premie      1      9.820 1012.75 110.228
## + mage        1      8.842 1013.73 111.136
## + marital     1      8.156 1014.41 111.772
## + mature      1      7.540 1015.03 112.344
## <none>                     1022.57 112.461
## + weeks:sex  1      0.543 1022.03 118.808
## + term:sex   2      3.545 1019.02 122.887
## - term        2     41.311 1063.88 136.035
## - sex         1     41.207 1063.78 142.790
## - weeks       1    222.908 1245.48 291.179
##
## Step:  AIC=86.92
## weight ~ weeks + sex + term + weeks:term
##
##               Df Sum of Sq    RSS     AIC
## + gained      1     26.662  954.15  67.836
## + habit       1     10.934  969.88  83.221
## + whitemom    1     10.149  970.66  83.982
## + mage        1      8.675  972.14  85.410
## + marital     1      8.651  972.16  85.433
## <none>                      980.81  86.923
## + mature      1      6.993  973.82  87.037
## + premie      1      0.885  979.93  92.921
## + weeks:sex  1      0.170  980.64  93.607
## + term:sex   2      3.526  977.29  97.228
## - weeks:term  2     41.756 1022.57 112.461
## - sex         1     39.579 1020.39 117.303
##
## Step:  AIC=67.84
## weight ~ weeks + sex + term + gained + weeks:term
##
##                Df Sum of Sq    RSS     AIC
## + mage         1     10.062  944.09  64.707
## + habit        1      9.441  944.71  65.325
## + whitemom     1      8.923  945.23  65.841
## + marital      1      8.208  945.94  66.553
## + mature       1      7.864  946.29  66.895
## <none>                       954.15  67.836
## + premie       1      1.147  953.00  73.551
## + weeks:sex   1      0.112  954.04  74.573
## + weeks:gained 1     0.052  954.10  74.632
## + gained:sex  1      0.001  954.15  74.682
## + term:sex    2      2.789  951.36  78.776
## + gained:term  2     0.345  953.81  81.189
## - gained       1     26.662  980.81  86.923
## - weeks:term   2     42.963  997.11  95.586
## - sex          1     38.491  992.64  98.204
##
## Step:  AIC=64.71
```

2

```
## weight ~ weeks + sex + term + gained + mage + weeks:term
##
##               Df Sum of Sq    RSS    AIC
## + whitemom     1     8.780 935.31 62.762
## + mage:weeks   1     8.758 935.33 62.784
## + habit        1     8.525 935.56 63.018
## <none>                      944.09 64.707
## - mage         1    10.062 954.15 67.836
## + marital      1     2.912 941.18 68.647
## + premie       1     1.852 942.24 69.706
## + mature       1     0.864 943.23 70.692
## + weeks:sex    1     0.157 943.93 71.397
## + mage:gained  1     0.114 943.98 71.440
## + weeks:gained 1     0.110 943.98 71.444
## + mage:sex     1     0.029 944.06 71.525
## + gained:sex   1     0.000 944.09 71.554
## + term:sex     2     2.073 942.02 76.333
## + mage:term    2     1.818 942.27 76.587
## + gained:term  2     0.260 943.83 78.141
## - gained       1    28.049 972.14 85.410
## - weeks:term   2    42.780 986.87 92.715
## - sex          1    39.943 984.03 96.853
##
## Step:  AIC=62.76
## weight ~ weeks + sex + term + gained + mage + whitemom + weeks:term
##
##                  Df Sum of Sq    RSS    AIC
## + habit           1     9.751 925.56 59.747
## + mage:weeks      1     8.845 926.46 60.667
## <none>                        935.31 62.762
## - whitemom        1     8.780 944.09 64.707
## - mage            1     9.919 945.23 65.841
## + gained:whitemom 1     2.497 932.81 67.093
## + premie          1     1.788 933.52 67.809
## + marital         1     1.683 933.63 67.914
## + mage:whitemom   1     1.260 934.05 68.340
## + mature          1     1.091 934.22 68.510
## + sex:whitemom    1     1.001 934.31 68.601
## + weeks:gained    1     0.171 935.14 69.437
## + weeks:sex       1     0.123 935.19 69.485
## + mage:gained     1     0.079 935.23 69.530
## + weeks:whitemom  1     0.005 935.30 69.604
## + mage:sex        1     0.001 935.31 69.608
## + gained:sex      1     0.001 935.31 69.608
## + term:sex        2     1.913 933.40 74.529
## + mage:term       2     1.884 933.43 74.558
## + term:whitemom   2     0.440 934.87 76.013
## + gained:term     2     0.224 935.09 76.230
## - gained          1    26.791 962.10 82.490
## - weeks:term      2    40.727 976.04 89.176
## - sex             1    38.435 973.75 93.811
##
## Step:  AIC=59.75
## weight ~ weeks + sex + term + gained + mage + whitemom + habit +
##     weeks:term
##
##                  Df Sum of Sq    RSS    AIC
## + mage:weeks      1     9.206 916.35 57.187
## <none>                        925.56 59.747
## + mage:habit      1     6.419 919.14 60.045
## - mage            1     8.937 934.50 61.942
## - habit           1     9.751 935.31 62.762
## - whitemom        1    10.006 935.56 63.018
## + gained:whitemom 1     2.134 923.42 64.422
## + premie          1     1.500 924.06 65.067
## + mage:whitemom   1     1.307 924.25 65.264
## + mature          1     1.090 924.47 65.484
## + sex:habit       1     0.989 924.57 65.587
```

```
## + sex:whitemom     1      0.782 924.78 65.798
## + marital          1      0.687 924.87 65.894
## + habit:whitemom   1      0.332 925.23 66.256
## + gained:habit     1      0.236 925.32 66.353
## + weeks:habit      1      0.210 925.35 66.380
## + weeks:gained     1      0.138 925.42 66.454
## + weeks:sex        1      0.073 925.49 66.519
## + gained:sex       1      0.053 925.51 66.540
## + mage:gained      1      0.033 925.52 66.560
## + weeks:whitemom   1      0.011 925.55 66.582
## + mage:sex         1      0.004 925.55 66.589
## + mage:term        2      2.493 923.07 70.902
## + term:sex         2      1.656 923.90 71.756
## + term:habit       2      1.087 924.47 72.334
## + term:whitemom    2      0.593 924.97 72.838
## + gained:term      2      0.212 925.35 73.225
## - gained           1     25.108 950.67 78.087
## - weeks:term       2     39.651 965.21 85.525
## - sex              1     39.057 964.62 91.794
##
## Step:  AIC=57.19
## weight ~ weeks + sex + term + gained + mage + whitemom + habit +
##     weeks:term + weeks:mage
##
##                  Df Sum of Sq    RSS    AIC
## <none>                        916.35 57.187
## + mage:habit      1     5.135 911.22 58.746
## - weeks:mage      1     9.206 925.56 59.747
## - habit           1    10.112 926.46 60.667
## - whitemom        1    10.125 926.48 60.681
## + gained:whitemom 1     2.366 913.99 61.601
## + mature          1     1.159 915.19 62.843
## + sex:whitemom    1     0.808 915.54 63.204
## + mage:whitemom   1     0.720 915.63 63.294
## + sex:habit       1     0.664 915.69 63.352
## + marital         1     0.463 915.89 63.558
## + habit:whitemom  1     0.434 915.92 63.588
## + premie          1     0.248 916.10 63.779
## + weeks:gained    1     0.130 916.22 63.900
## + gained:habit    1     0.112 916.24 63.919
## + weeks:sex       1     0.097 916.26 63.934
## + gained:sex      1     0.053 916.30 63.979
## + weeks:habit     1     0.031 916.32 64.001
## + mage:gained     1     0.005 916.35 64.029
## + mage:sex        1     0.001 916.35 64.033
## + weeks:whitemom  1     0.000 916.35 64.034
## + mage:term       2     5.132 911.22 65.596
## + term:sex        2     1.600 914.75 69.236
## + term:habit      2     0.842 915.51 70.016
## + term:whitemom   2     0.700 915.65 70.161
## + gained:term     2     0.124 916.23 70.753
## - gained          1    26.481 942.83 77.147
## - weeks:term      2    42.196 958.55 85.856
## - sex             1    39.999 956.35 90.544
##
## Call:
## lm(formula = weight ~ weeks + sex + term + gained + mage + whitemom +
##     habit + weeks:term + weeks:mage, data = births14)
##
## Coefficients:
##    (Intercept)           weeks          sexmale         termfull
##      -0.308456        0.168305         0.413773         9.435196
##       termlate          gained             mage     whitemomwhite
##      18.519206        0.011030        -0.247747         0.249226
##    habitsmoker   weeks:termfull   weeks:termlate       weeks:mage
##      -0.323727       -0.242098        -0.465941         0.006865
```

Solution:

1. Calculate the value 161.96. Hints: the penalty is `log(n)` which means the calculation is BIC, not AIC; there are 941 observations; RSS is SSE.

$$BIC_p = n\ln(SSE_p) - n \cdot \ln(n) + \ln(n) \cdot p$$

Note that the last column is **BIC** (not AIC) because of the `k = log(941)` penalization in the argument of the function.

$$BIC = 941 \cdot \ln(1101.6) - 941 \cdot \ln(941) + \ln(941) \cdot 2 = 161.96$$

2. Is `Sum of Sq` the same as SSR or extra sums of squares? (Hint: look at `<none>`)

Extra sum of squares. Note that each value in the column `Sum of Sq` is a difference in SSE for two nested model (= SSE(reduced) - SSE(full)). Sometimes the full model is the starting place (if there is a `-` indicating subtraction of a variable); sometimes the full model is the new possible model (if there is a `+` indicating addition of a variable).

3. Why aren't there any interaction terms for the first few iterations?

The software doesn't let you add interaction terms unless you have already included the main effects – good modeling practice. It would almost never make sense to have an interaction term if the main effects weren't in the model.

4. Why, toward the end, are the pluses and minuses mixed up?

The choices for the variable (in or out) are ranked by BIC. So sometimes removing a variable will actually produce a better (smaller) BIC than adding a variable. Remember, adding a variable will always create a smaller SSE, but the change in SSE might not be enough to counter balance the penalization of adding extra coefficients to estimate.

Note:

Some other ways to run the function include the following arguments:

- `k=2` gives AIC
- `test = "F"` (without `k`) gives nested F tests (sorted by AIC)
- `C_p` can be calculated using the `scale` argument