Lab14

David Wiley / Duy Truong

March 27, 2019

Reading in Data

```
dat = read.csv("/home/david/Documents/2019 Spring/Applied Regression/Labs_HW/Data_Sets/Appendices/data-


fit = lm(MORT~., dat)


summary(fit)
```

```
##
## Call:
## lm(formula = MORT ~ ., data = dat)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -91.38 -18.97  -3.56  16.00  91.83
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 995.63646   91.64099  10.865 3.35e-15 ***
## PRECIP        1.40734    0.68914   2.042 0.046032 *
## EDUC        -14.80139    7.02747  -2.106 0.039849 *
## NONWHITE      3.19909    0.62231   5.141 3.89e-06 ***
## NOX          -0.10797    0.13502  -0.800 0.427426
## SO2           0.35518    0.09096   3.905 0.000264 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 37.09 on 54 degrees of freedom
## Multiple R-squared:  0.6746, Adjusted R-squared:  0.6444
## F-statistic: 22.39 on 5 and 54 DF,  p-value: 4.407e-12
```

Finding the influence observations in data:

```
influence.measures(fit)
```

```
## Influence measures of
##   lm(formula = MORT ~ ., data = dat) :
##
##       dfb.1_  dfb.PREC  dfb.EDUC  dfb.NONW   dfb.NOX   dfb.SO2    dffit
## 1  -0.142709  4.67e-01 -0.006572 -8.27e-02  0.169887  0.134297 -0.61684
## 2   0.088512  1.66e-01 -0.180146 -4.60e-02  0.183397  0.052782 -0.43065
## 3  -0.040995  1.01e-01  0.012889 -2.99e-02  0.009176  0.034672 -0.12766
## 4  -0.646098 -8.09e-02  0.714150  4.94e-01 -0.220871  0.294323 -0.99206
## 5   0.016412  3.48e-02 -0.036114  1.28e-02  0.037405 -0.006831 -0.09084
```

```
## 6   0.111584  8.16e-02 -0.168678 -1.60e-01  0.129325  0.088048 -0.38165
## 7   0.920878 -1.28e+00 -0.724788  4.98e-01 -0.410019  0.046590 -1.40319
## 8  -0.021745 -7.53e-02  0.051467  8.10e-02 -0.636539  0.168712 -0.71195
## 9  -0.089134  8.52e-02  0.063880  2.38e-02  0.039750  0.060272 -0.16120
## 10  0.016989 -1.08e-01  0.018958  2.66e-02 -0.075654  0.004407  0.14246
## 11 -0.044053  5.18e-02  0.023194  3.14e-02  0.037015  0.031973 -0.12865
## 12  0.084946 -9.66e-02 -0.081447  6.94e-02 -0.010910  0.014741 -0.15892
## 13 -0.020841  5.79e-02  0.001271 -3.39e-02  0.038741  0.036114 -0.09843
## 14  0.005080 -3.60e-03 -0.005314  3.88e-03 -0.000031 -0.001860 -0.00744
## 15  0.010664 -2.59e-02 -0.007900  2.96e-02 -0.008749  0.004858 -0.03954
## 16 -0.050619  1.88e-02  0.058288 -2.08e-02 -0.016061  0.009640  0.07695
## 17 -0.013710 -6.98e-02  0.025405  8.16e-02 -0.040859  0.052687 -0.14081
## 18  0.019658 -4.08e-02 -0.015333  4.17e-02 -0.011067  0.002965 -0.05960
## 19 -0.012553 -2.47e-02  0.020639  3.28e-02  0.167857 -0.049387  0.23971
## 20 -0.411192  4.99e-02  0.439630  1.73e-01 -0.080933  0.132680 -0.51490
## 21  0.003684  6.87e-04 -0.003824 -4.28e-03  0.001028 -0.002803  0.00750
## 22 -0.008778  3.83e-03  0.007504  4.73e-03  0.001011  0.006496 -0.01678
## 23 -0.034506 -3.77e-03  0.044625  1.72e-02 -0.021239 -0.009863  0.07315
## 24  0.005700 -1.72e-03 -0.007065  3.41e-03  0.003919 -0.004994 -0.01428
## 25  0.001807 -2.62e-03 -0.001554  1.50e-03 -0.000637  0.000585 -0.00372
## 26  0.000785  8.78e-03 -0.004850  9.10e-03  0.015070 -0.025276 -0.03535
## 27 -0.244862  2.33e-01  0.228372 -1.76e-01  0.058381  0.098394  0.31706
## 28 -0.008238 -1.32e-02  0.017485  1.64e-02 -0.020887 -0.015272  0.06471
## 29  0.077583  3.06e-02 -0.086226 -9.91e-02  0.033367 -0.052564  0.16272
## 30  0.082461 -1.02e-01 -0.060118  5.60e-02 -0.042466 -0.054787  0.13222
## 31 -0.021056 -2.54e-03  0.023202  2.15e-02 -0.002056 -0.004552 -0.03804
## 32 -0.091693  9.82e-02  0.104437 -1.71e-01 -0.022745  0.004676  0.28445
## 33  0.018955 -1.48e-02 -0.017927 -1.92e-02 -0.001927  0.013454 -0.04436
## 34 -0.195126  1.49e-01  0.181011 -9.08e-02  0.037882  0.042368 -0.22768
## 35  0.023614 -1.02e-02 -0.020789  2.27e-03 -0.003369 -0.014752  0.05461
## 36 -0.142679  2.82e-02  0.168165  4.65e-02 -0.050682 -0.017883  0.25037
## 37 -0.016728  3.85e-02  0.008537 -2.74e-02  0.025373 -0.037709 -0.07316
## 38 -0.085901 -2.46e-04  0.103873 -1.14e-01  0.005994 -0.009422 -0.25590
## 39  0.162531  9.10e-02 -0.191217 -2.32e-01  0.084293 -0.075110  0.35192
## 40  0.175850 -5.98e-02 -0.179127 -9.99e-02  0.026473 -0.095847 -0.23266
## 41 -0.047544  4.48e-03  0.057886  4.66e-02 -0.034866 -0.002817  0.12552
## 42 -0.023109 -6.27e-03  0.028610  1.75e-02  0.025143 -0.103569 -0.15012
## 43 -0.012861  8.79e-03  0.012288 -1.93e-02  0.001150  0.015664 -0.02917
## 44  0.229154 -2.33e-01 -0.176342  5.35e-02 -0.096547 -0.112678  0.33411
## 45  0.021260 -1.28e-02 -0.019068 -5.46e-02  0.001491  0.015827 -0.07921
## 46  0.005542 -5.77e-02  0.014959  6.72e-02 -0.049730  0.026435  0.16019
## 47 -0.043681  4.51e-02  0.038189 -8.86e-03  0.030957 -0.061032 -0.10015
## 48  0.039108 -3.26e-02 -0.034090  5.53e-02  0.023217 -0.170739 -0.18798
## 49 -0.033113  6.57e-02  0.022218 -4.83e-02  0.002501  0.089424  0.14740
## 50  0.079272 -9.47e-03 -0.038890 -2.89e-01 -0.062618 -0.006239  0.48968
## 51  0.248649 -1.16e-01 -0.225243 -7.42e-02 -0.018374 -0.109989  0.37832
## 52 -0.218423  2.45e-01  0.195240 -8.17e-02  0.030128  0.043276  0.35029
## 53 -0.024702  5.78e-03  0.046113 -3.48e-01 -0.032288  0.116721 -0.42753
## 54 -0.007939  8.72e-03  0.005535  3.27e-03 -0.007232  0.037014  0.04725
## 55  0.043539  1.98e-02 -0.056713  2.73e-02  0.028128 -0.043161  0.10218
## 56 -0.000285  2.48e-06  0.000297  2.17e-05 -0.000273  0.001082  0.00117
## 57 -0.029507 -1.40e-02  0.031748  1.62e-01 -0.034298 -0.010709  0.19690
## 58  0.005699 -5.91e-02  0.036948 -3.49e-01 -0.074429  0.053907 -0.46983
## 59  0.024740 -9.58e-03 -0.031277  3.62e-02 -0.017781  0.076807  0.12323
```

```
## 60  0.377535  1.29e-01 -0.504803  6.55e-01  0.369854 -0.640507  1.21523
##    cov.r   cook.d    hat inf
## 1  1.034 6.21e-02 0.1493
## 2  0.906 3.01e-02 0.0679
## 3  1.330 2.76e-03 0.1665
## 4  0.559 1.46e-01 0.1134   *
## 5  1.191 1.40e-03 0.0720
## 6  0.882 2.36e-02 0.0524
## 7  0.638 2.93e-01 0.2078   *
## 8  4.845 8.58e-02 0.7730   *
## 9  1.109 4.37e-03 0.0479
## 10 1.306 3.44e-03 0.1541
## 11 1.106 2.79e-03 0.0369
## 12 1.119 4.25e-03 0.0516
## 13 1.151 1.64e-03 0.0487
## 14 1.203 9.40e-06 0.0701
## 15 1.218 2.65e-04 0.0837
## 16 1.188 1.00e-03 0.0672
## 17 1.146 3.35e-03 0.0584
## 18 1.192 6.03e-04 0.0667
## 19 1.383 9.72e-03 0.2107   *
## 20 1.115 4.38e-02 0.1506
## 21 1.196 9.55e-06 0.0644
## 22 1.155 4.78e-05 0.0327
## 23 1.170 9.07e-04 0.0538
## 24 1.148 3.46e-05 0.0267
## 25 1.184 2.35e-06 0.0555
## 26 1.193 2.12e-04 0.0644
## 27 1.201 1.69e-02 0.1342
## 28 1.134 7.09e-04 0.0292
## 29 1.169 4.47e-03 0.0767
## 30 1.167 2.96e-03 0.0670
## 31 1.237 2.46e-04 0.0972
## 32 0.943 1.33e-02 0.0406
## 33 1.192 3.34e-04 0.0647
## 34 1.183 8.73e-03 0.1017
## 35 1.125 5.05e-04 0.0209
## 36 1.016 1.04e-02 0.0464
## 37 1.172 9.07e-04 0.0553
## 38 1.010 1.08e-02 0.0465
## 39 1.108 2.06e-02 0.1041
## 40 1.285 9.14e-03 0.1572
## 41 1.090 2.65e-03 0.0304
## 42 1.147 3.80e-03 0.0619
## 43 1.205 1.44e-04 0.0725
## 44 0.921 1.82e-02 0.0483
## 45 1.208 1.06e-03 0.0813
## 46 1.016 4.27e-03 0.0235
## 47 1.296 1.70e-03 0.1424
## 48 1.445 5.99e-03 0.2356   *
## 49 1.097 3.65e-03 0.0394
## 50 0.547 3.59e-02 0.0329   *
## 51 0.726 2.25e-02 0.0324
## 52 0.861 1.98e-02 0.0425
```

```
## 53 1.199 3.05e-02 0.1614
## 54 1.206 3.79e-04 0.0749
## 55 1.224 1.77e-03 0.0962
## 56 1.503 2.32e-07 0.2557   *
## 57 1.165 6.53e-03 0.0841
## 58 1.210 3.68e-02 0.1765
## 59 1.304 2.57e-03 0.1502
## 60 0.626 2.21e-01 0.1697   *
```

Looking at the matrix of data:

It looks like there are 8 possible influential observations.

```
summary(fit)
```

```
##
## Call:
## lm(formula = MORT ~ ., data = dat)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -91.38 -18.97  -3.56  16.00  91.83
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 995.63646   91.64099  10.865 3.35e-15 ***
## PRECIP        1.40734    0.68914   2.042 0.046032 *
## EDUC        -14.80139    7.02747  -2.106 0.039849 *
## NONWHITE      3.19909    0.62231   5.141 3.89e-06 ***
## NOX          -0.10797    0.13502  -0.800 0.427426
## SO2           0.35518    0.09096   3.905 0.000264 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 37.09 on 54 degrees of freedom
## Multiple R-squared:  0.6746, Adjusted R-squared:  0.6444
## F-statistic: 22.39 on 5 and 54 DF,  p-value: 4.407e-12
```

```
summary(lm(MORT~., dat[-4,]))#
```

```
##
## Call:
## lm(formula = MORT ~ ., data = dat[-4, ])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -86.694 -15.002  -5.571  14.949  87.933
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1051.48498   88.75357  11.847  < 2e-16 ***
## PRECIP         1.45995    0.65031   2.245 0.028963 *
## EDUC         -19.53521    6.84477  -2.854 0.006147 **
## NONWHITE       2.90929    0.59621   4.880 1.01e-05 ***
## NOX           -0.07984    0.12776  -0.625 0.534716
```

```
## SO2            0.32992    0.08628   3.824 0.000348 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 34.99 on 53 degrees of freedom
## Multiple R-squared:  0.7036, Adjusted R-squared:  0.6756
## F-statistic: 25.16 on 5 and 53 DF,  p-value: 6.73e-13
```

```
summary(lm(MORT~., dat[-7,]))#
```

```
##
## Call:
## lm(formula = MORT ~ ., data = dat[-7, ])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -92.670 -17.131  -2.655  16.797  90.050
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 915.91382   91.33070  10.029 7.61e-14 ***
## PRECIP        2.24260    0.71886   3.120 0.002927 **
## EDUC         -9.98968    6.86711  -1.455 0.151646
## NONWHITE      2.90642    0.59751   4.864 1.07e-05 ***
## NOX          -0.05567    0.12898  -0.432 0.667764
## SO2           0.35117    0.08594   4.086 0.000149 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 35.04 on 53 degrees of freedom
## Multiple R-squared:  0.7068, Adjusted R-squared:  0.6792
## F-statistic: 25.55 on 5 and 53 DF,  p-value: 5.062e-13
```

```
summary(lm(MORT~., dat[-50,]))#
```

```
##
## Call:
## lm(formula = MORT ~ ., data = dat[-50, ])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -87.783 -18.593  -1.122  15.444  84.584
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 988.74747   86.94181  11.373 7.75e-16 ***
## PRECIP        1.41353    0.65352   2.163 0.035075 *
## EDUC        -14.54222    6.66486  -2.182 0.033570 *
## NONWHITE      3.36945    0.59361   5.676 5.90e-07 ***
## NOX          -0.09995    0.12808  -0.780 0.438630
## SO2           0.35571    0.08626   4.124 0.000132 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 35.17 on 53 degrees of freedom
```

```
## Multiple R-squared:  0.7085, Adjusted R-squared:  0.681
## F-statistic: 25.76 on 5 and 53 DF,  p-value: 4.372e-13
```

```
summary(lm(MORT~., dat[-60,]))#
```

```
##
## Call:
## lm(formula = MORT ~ ., data = dat[-60, ])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -87.229 -17.473  -2.911  19.378  90.256
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 962.87594   87.62642  10.988 2.81e-15 ***
## PRECIP        1.32332    0.65330   2.026   0.0479 *
## EDUC        -11.44228    6.77062  -1.690   0.0969 .
## NONWHITE      2.81315    0.60650   4.638 2.34e-05 ***
## NOX          -0.15526    0.12906  -1.203   0.2343
## SO2           0.41034    0.08854   4.634 2.37e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 35.12 on 53 degrees of freedom
## Multiple R-squared:  0.6697, Adjusted R-squared:  0.6386
## F-statistic: 21.49 on 5 and 53 DF,  p-value: 1.101e-11
```

Removing each of the possible influential observations, we can see that the 4th, 7th, 50th, and 60th observations have a high influence on the model. Whereas the other observations show they have little influence on the model. We would have to find out if the data observation is indeed a valid observation. If not, then we can just simply delete the observation. If so, then we cannot delete it but could possibly downweight the observations in proportion to residual magnitude or influence.