HW 04

TUAN BUI

9/28/2021

Question 01

```
install.packages('car')
library(carData)
data("Salaries", package = "carData")
Salaries
```

```
##           rank discipline yrs.since.phd yrs.service    sex salary
## 1         Prof          B            19          18   Male 139750
## 2         Prof          B            20          16   Male 173200
## 3     AsstProf          B             4           3   Male  79750
## 4         Prof          B            45          39   Male 115000
## 5         Prof          B            40          41   Male 141500
## 6    AssocProf          B             6           6   Male  97000
## 7         Prof          B            30          23   Male 175000
## 8         Prof          B            45          45   Male 147765
## 9         Prof          B            21          20   Male 119250
## 10        Prof          B            18          18 Female 129000
## 11   AssocProf          B            12           8   Male 119800
## 12    AsstProf          B             7           2   Male  79800
## 13    AsstProf          B             1           1   Male  77700
## 14    AsstProf          B             2           0   Male  78000
## 15        Prof          B            20          18   Male 104800
## 16        Prof          B            12           3   Male 117150
## 17        Prof          B            19          20   Male 101000
## 18        Prof          A            38          34   Male 103450
## 19        Prof          A            37          23   Male 124750
## 20        Prof          A            39          36 Female 137000
## 21        Prof          A            31          26   Male  89565
## 22        Prof          A            36          31   Male 102580
## 23        Prof          A            34          30   Male  93904
## 24        Prof          A            24          19   Male 113068
## 25   AssocProf          A            13           8 Female  74830
## 26        Prof          A            21           8   Male 106294
## 27        Prof          A            35          23   Male 134885
## 28    AsstProf          B             5           3   Male  82379
## 29    AsstProf          B            11           0   Male  77000
## 30        Prof          B            12           8   Male 118223
## 31        Prof          B            20           4   Male 132261
## 32    AsstProf          B             7           2   Male  79916
## 33        Prof          B            13           9   Male 117256
```

```
## 34   AsstProf      B      4     2   Male  80225
## 35   AsstProf      B      4     2 Female  80225
## 36   AsstProf      B      5     0 Female  77000
## 37       Prof      B     22    21   Male 155750
## 38   AsstProf      B      7     4   Male  86373
## 39       Prof      B     41    31   Male 125196
## 40  AssocProf      B      9     9   Male 100938
## 41       Prof      B     23     2   Male 146500
## 42  AssocProf      B     23    23   Male  93418
## 43       Prof      B     40    27   Male 101299
## 44       Prof      B     38    38   Male 231545
## 45       Prof      B     19    19   Male  94384
## 46       Prof      B     25    15   Male 114778
## 47       Prof      B     40    28   Male  98193
## 48       Prof      B     23    19 Female 151768
## 49       Prof      B     25    25 Female 140096
## 50   AsstProf      B      1     1   Male  70768
## 51       Prof      B     28    28   Male 126621
## 52       Prof      B     12    11   Male 108875
## 53   AsstProf      B     11     3 Female  74692
## 54       Prof      B     16     9   Male 106639
## 55  AssocProf      B     12    11   Male 103760
## 56  AssocProf      B     14     5   Male  83900
## 57       Prof      B     23    21   Male 117704
## 58  AssocProf      B      9     8   Male  90215
## 59  AssocProf      B     10     9   Male 100135
## 60   AsstProf      B      8     3   Male  75044
## 61  AssocProf      B      9     8   Male  90304
## 62   AsstProf      B      3     2   Male  75243
## 63       Prof      B     33    31   Male 109785
## 64  AssocProf      B     11    11 Female 103613
## 65   AsstProf      B      4     3   Male  68404
## 66  AssocProf      B      9     8   Male 100522
## 67       Prof      B     22    12   Male 101000
## 68       Prof      B     35    31   Male  99418
## 69       Prof      B     17    17 Female 111512
## 70       Prof      B     28    36   Male  91412
## 71       Prof      B     17     2   Male 126320
## 72       Prof      B     45    45   Male 146856
## 73       Prof      B     29    19   Male 100131
## 74       Prof      B     35    34   Male  92391
## 75       Prof      B     28    23   Male 113398
## 76   AsstProf      B      8     3   Male  73266
## 77       Prof      B     17     3   Male 150480
## 78       Prof      B     26    19   Male 193000
## 79   AsstProf      B      3     1   Male  86100
## 80   AsstProf      B      6     2   Male  84240
## 81       Prof      B     43    28   Male 150743
## 82       Prof      B     17    16   Male 135585
## 83       Prof      B     22    20   Male 144640
```

```
## 84     AsstProf       B        6        2    Male    88825
## 85         Prof       B       17       18  Female  122960
## 86         Prof       B       15       14    Male  132825
## 87         Prof       B       37       37    Male  152708
## 88     AsstProf       B        2        2    Male    88400
## 89         Prof       B       25       25    Male  172272
## 90    AssocProf       B        9        7    Male  107008
## 91     AsstProf       B       10        5  Female   97032
## 92    AssocProf       B       10        7    Male  105128
## 93    AssocProf       B       10        7    Male  105631
## 94         Prof       B       38       38    Male  166024
## 95         Prof       B       21       20    Male  123683
## 96     AsstProf       B        4        0    Male    84000
## 97    AssocProf       B       17       12    Male    95611
## 98         Prof       B       13        7    Male  129676
## 99         Prof       B       30       14    Male  102235
## 100        Prof       B       41       26    Male  106689
## 101        Prof       B       42       25    Male  133217
## 102        Prof       B       28       23    Male  126933
## 103        Prof       B       16        5    Male  153303
## 104        Prof       B       20       14  Female  127512
## 105   AssocProf       A       18       10    Male    83850
## 106        Prof       A       31       28    Male  113543
## 107   AssocProf       A       11        8    Male    82099
## 108   AssocProf       A       10        8    Male    82600
## 109   AssocProf       A       15        8    Male    81500
## 110        Prof       A       40       31    Male  131205
## 111        Prof       A       20       16    Male  112429
## 112   AssocProf       A       19       16    Male    82100
## 113    AsstProf       A        3        1    Male    72500
## 114        Prof       A       37       37    Male  104279
## 115        Prof       A       12        0  Female  105000
## 116        Prof       A       21        9    Male  120806
## 117        Prof       A       30       29    Male  148500
## 118        Prof       A       39       36    Male  117515
## 119    AsstProf       A        4        1    Male    72500
## 120    AsstProf       A        5        3  Female   73500
## 121        Prof       A       14       14    Male  115313
## 122        Prof       A       32       32    Male  124309
## 123        Prof       A       24       22    Male    97262
## 124   AssocProf       A       25       22  Female    62884
## 125        Prof       A       24       22    Male    96614
## 126        Prof       A       54       49    Male    78162
## 127        Prof       A       28       26    Male  155500
## 128    AsstProf       A        2        0  Female    72500
## 129        Prof       A       32       30    Male  113278
## 130    AsstProf       A        4        2    Male    73000
## 131   AssocProf       A       11        9    Male    83001
## 132        Prof       A       56       57    Male    76840
## 133   AssocProf       A       10        8  Female    77500
```

```
## 134   AsstProf       A        3        1 Female   72500
## 135       Prof       A       35       25   Male 168635
## 136       Prof       A       20       18   Male 136000
## 137       Prof       A       16       14   Male 108262
## 138       Prof       A       17       14   Male 105668
## 139  AssocProf       A       10        7   Male   73877
## 140       Prof       A       21       18   Male 152664
## 141  AssocProf       A       14        8   Male 100102
## 142  AssocProf       A       15       10   Male   81500
## 143       Prof       A       19       11   Male 106608
## 144   AsstProf       B        3        3   Male   89942
## 145       Prof       B       27       27   Male 112696
## 146       Prof       B       28       28   Male 119015
## 147   AsstProf       B        4        4   Male   92000
## 148       Prof       B       27       27   Male 156938
## 149       Prof       B       36       26 Female 144651
## 150   AsstProf       B        4        3   Male   95079
## 151       Prof       B       14       12   Male 128148
## 152   AsstProf       B        4        4   Male   92000
## 153       Prof       B       21        9   Male 111168
## 154  AssocProf       B       12       10 Female 103994
## 155   AsstProf       B        4        0   Male   92000
## 156       Prof       B       21       21   Male 118971
## 157  AssocProf       B       12       18   Male 113341
## 158   AsstProf       B        1        0   Male   88000
## 159  AssocProf       B        6        6   Male   95408
## 160       Prof       B       15       16   Male 137167
## 161   AsstProf       B        2        2   Male   89516
## 162       Prof       B       26       19   Male 176500
## 163  AssocProf       B       22        7   Male   98510
## 164   AsstProf       B        3        3   Male   89942
## 165   AsstProf       B        1        0   Male   88795
## 166       Prof       B       21        8   Male 105890
## 167       Prof       B       16       16   Male 167284
## 168       Prof       B       18       19   Male 130664
## 169  AssocProf       B        8        6   Male 101210
## 170       Prof       B       25       18   Male 181257
## 171   AsstProf       B        5        5   Male   91227
## 172       Prof       B       19       19   Male 151575
## 173       Prof       B       37       24   Male   93164
## 174       Prof       B       20       20   Male 134185
## 175  AssocProf       B       17        6   Male 105000
## 176       Prof       B       28       25   Male 111751
## 177  AssocProf       B       10        7   Male   95436
## 178  AssocProf       B       13        9   Male 100944
## 179       Prof       B       27       14   Male 147349
## 180   AsstProf       B        3        3 Female   92000
## 181       Prof       B       11       11   Male 142467
## 182       Prof       B       18        5   Male 141136
## 183  AssocProf       B        8        8   Male 100000
```

```
## 184       Prof       B       26       22   Male 150000
## 185       Prof       B       23       23   Male 101000
## 186       Prof       B       33       30   Male 134000
## 187 AssocProf       B       13       10 Female 103750
## 188       Prof       B       18       10   Male 107500
## 189 AssocProf       B       28       28   Male 106300
## 190       Prof       B       25       19   Male 153750
## 191       Prof       B       22        9   Male 180000
## 192       Prof       B       43       22   Male 133700
## 193       Prof       B       19       18   Male 122100
## 194 AssocProf       B       19       19   Male  86250
## 195 AssocProf       B       48       53   Male  90000
## 196 AssocProf       B        9        7   Male 113600
## 197  AsstProf       B        4        4   Male  92700
## 198  AsstProf       B        4        4   Male  92000
## 199       Prof       B       34       33   Male 189409
## 200       Prof       B       38       22   Male 114500
## 201  AsstProf       B        4        4   Male  92700
## 202       Prof       B       40       40   Male 119700
## 203       Prof       B       28       17   Male 160400
## 204       Prof       B       17       17   Male 152500
## 205       Prof       B       19        5   Male 165000
## 206       Prof       B       21        2   Male  96545
## 207       Prof       B       35       33   Male 162200
## 208       Prof       B       18       18   Male 120000
## 209  AsstProf       B        7        2   Male  91300
## 210       Prof       B       20       20   Male 163200
## 211  AsstProf       B        4        3   Male  91000
## 212       Prof       B       39       39   Male 111350
## 213       Prof       B       15        7   Male 128400
## 214       Prof       B       26       19   Male 126200
## 215 AssocProf       B       11        1   Male 118700
## 216       Prof       B       16       11   Male 145350
## 217       Prof       B       15       11   Male 146000
## 218 AssocProf       B       29       22   Male 105350
## 219 AssocProf       B       14        7 Female 109650
## 220       Prof       B       13       11   Male 119500
## 221       Prof       B       21       21   Male 170000
## 222       Prof       B       23       10   Male 145200
## 223 AssocProf       B       13        6   Male 107150
## 224       Prof       B       34       20   Male 129600
## 225       Prof       A       38       35   Male  87800
## 226       Prof       A       20       20   Male 122400
## 227  AsstProf       A        3        1   Male  63900
## 228 AssocProf       A        9        7   Male  70000
## 229       Prof       A       16       11   Male  88175
## 230       Prof       A       39       38   Male 133900
## 231       Prof       A       29       27 Female  91000
## 232 AssocProf       A       26       24 Female  73300
## 233       Prof       A       38       19   Male 148750
```

```
## 234      Prof        A       36    19 Female 117555
## 235  AsstProf        A        8     3   Male  69700
## 236      Prof        A       28    17   Male  81700
## 237      Prof        A       25    25   Male 114000
## 238  AsstProf        A        7     6 Female  63100
## 239      Prof        A       46    40   Male  77202
## 240      Prof        A       19     6   Male  96200
## 241  AsstProf        A        5     3   Male  69200
## 242      Prof        A       31    30   Male 122875
## 243      Prof        A       38    37   Male 102600
## 244      Prof        A       23    23   Male 108200
## 245      Prof        A       19    23   Male  84273
## 246      Prof        A       17    11 Female  90450
## 247      Prof        A       30    23   Male  91100
## 248      Prof        A       21    18   Male 101100
## 249      Prof        A       28    23   Male 128800
## 250      Prof        A       29     7   Male 204000
## 251      Prof        A       39    39   Male 109000
## 252      Prof        A       20     8   Male 102000
## 253      Prof        A       31    12   Male 132000
## 254  AsstProf        A        4     2 Female  77500
## 255      Prof        A       28     7 Female 116450
## 256 AssocProf        A       12     8   Male  83000
## 257      Prof        A       22    22   Male 140300
## 258 AssocProf        A       30    23   Male  74000
## 259  AsstProf        A        9     3   Male  73800
## 260      Prof        A       32    30   Male  92550
## 261 AssocProf        A       41    33   Male  88600
## 262      Prof        A       45    45   Male 107550
## 263      Prof        A       31    26   Male 121200
## 264      Prof        A       31    31   Male 126000
## 265      Prof        A       37    35   Male  99000
## 266      Prof        A       36    30   Male 134800
## 267      Prof        A       43    43   Male 143940
## 268      Prof        A       14    10   Male 104350
## 269      Prof        A       47    44   Male  89650
## 270      Prof        A       13     7   Male 103700
## 271      Prof        A       42    40   Male 143250
## 272      Prof        A       42    18   Male 194800
## 273  AsstProf        A        4     1   Male  73000
## 274  AsstProf        A        8     4   Male  74000
## 275  AsstProf        A        8     3 Female  78500
## 276      Prof        A       12     6   Male  93000
## 277      Prof        A       52    48   Male 107200
## 278      Prof        A       31    27   Male 163200
## 279      Prof        A       24    18   Male 107100
## 280      Prof        A       46    46   Male 100600
## 281      Prof        A       39    38   Male 136500
## 282      Prof        A       37    27   Male 103600
## 283      Prof        A       51    51   Male  57800
```

```
## 284      Prof         A       45      43    Male 155865
## 285 AssocProf         A        8       6    Male  88650
## 286 AssocProf         A       49      49    Male  81800
## 287      Prof         A       28      27    Male 115800
## 288  AsstProf         A        2       0    Male  85000
## 289      Prof         A       29      27    Male 150500
## 290  AsstProf         A        8       5    Male  74000
## 291      Prof         A       33       7    Male 174500
## 292      Prof         A       32      28    Male 168500
## 293      Prof         A       39       9    Male 183800
## 294 AssocProf         A       11       1    Male 104800
## 295      Prof         A       19       7    Male 107300
## 296      Prof         A       40      36    Male  97150
## 297      Prof         A       18      18    Male 126300
## 298      Prof         A       17      11    Male 148800
## 299      Prof         A       49      43    Male  72300
## 300 AssocProf         A       45      39    Male  70700
## 301      Prof         A       39      36    Male  88600
## 302      Prof         A       27      16    Male 127100
## 303      Prof         A       28      13    Male 170500
## 304      Prof         A       14       4    Male 105260
## 305      Prof         A       46      44    Male 144050
## 306      Prof         A       33      31    Male 111350
## 307  AsstProf         A        7       4    Male  74500
## 308      Prof         A       31      28    Male 122500
## 309  AsstProf         A        5       0    Male  74000
## 310      Prof         A       22      15    Male 166800
## 311      Prof         A       20       7    Male  92050
## 312      Prof         A       14       9    Male 108100
## 313      Prof         A       29      19    Male  94350
## 314      Prof         A       35      35    Male 100351
## 315      Prof         A       22       6    Male 146800
## 316  AsstProf         B        6       3    Male  84716
## 317 AssocProf         B       12       9 Female  71065
## 318      Prof         B       46      45    Male  67559
## 319      Prof         B       16      16    Male 134550
## 320      Prof         B       16      15    Male 135027
## 321      Prof         B       24      23    Male 104428
## 322 AssocProf         B        9       9    Male  95642
## 323 AssocProf         B       13      11    Male 126431
## 324      Prof         B       24      15 Female 161101
## 325      Prof         B       30      31    Male 162221
## 326  AsstProf         B        8       4    Male  84500
## 327      Prof         B       23      15    Male 124714
## 328      Prof         B       37      37    Male 151650
## 329 AssocProf         B       10      10    Male  99247
## 330      Prof         B       23      23    Male 134778
## 331      Prof         B       49      60    Male 192253
## 332      Prof         B       20       9    Male 116518
## 333      Prof         B       18      10 Female 105450
```

```
## 334      Prof      B      33      19    Male 145098
## 335 AssocProf      B      19       6  Female 104542
## 336      Prof      B      36      38    Male 151445
## 337      Prof      B      35      23    Male  98053
## 338      Prof      B      13      12    Male 145000
## 339      Prof      B      32      25    Male 128464
## 340      Prof      B      37      15    Male 137317
## 341      Prof      B      13      11    Male 106231
## 342      Prof      B      17      17  Female 124312
## 343      Prof      B      38      38    Male 114596
## 344      Prof      B      31      31    Male 162150
## 345      Prof      B      32      35    Male 150376
## 346      Prof      B      15      10    Male 107986
## 347      Prof      B      41      27    Male 142023
## 348      Prof      B      39      33    Male 128250
## 349  AsstProf      B       4       3    Male  80139
## 350      Prof      B      27      28    Male 144309
## 351      Prof      B      56      49    Male 186960
## 352      Prof      B      38      38    Male  93519
## 353      Prof      B      26      27    Male 142500
## 354      Prof      B      22      20    Male 138000
## 355  AsstProf      B       8       1    Male  83600
## 356      Prof      B      25      21    Male 145028
## 357      Prof      A      49      40    Male  88709
## 358      Prof      A      39      35    Male 107309
## 359      Prof      A      28      14  Female 109954
## 360  AsstProf      A      11       4    Male  78785
## 361      Prof      A      14      11    Male 121946
## 362      Prof      A      23      15  Female 109646
## 363      Prof      A      30      30    Male 138771
## 364 AssocProf      A      20      17    Male  81285
## 365      Prof      A      43      43    Male 205500
## 366      Prof      A      43      40    Male 101036
## 367      Prof      A      15      10    Male 115435
## 368 AssocProf      A      10       1    Male 108413
## 369      Prof      A      35      30    Male 131950
## 370      Prof      A      33      31    Male 134690
## 371 AssocProf      A      13       8    Male  78182
## 372      Prof      A      23      20    Male 110515
## 373      Prof      A      12       7    Male 109707
## 374      Prof      A      30      26    Male 136660
## 375      Prof      A      27      19    Male 103275
## 376      Prof      A      28      26    Male 103649
## 377  AsstProf      A       4       1    Male  74856
## 378  AsstProf      A       6       3    Male  77081
## 379      Prof      A      38      38    Male 150680
## 380 AssocProf      A      11       8    Male 104121
## 381  AsstProf      A       8       3    Male  75996
## 382      Prof      A      27      23    Male 172505
## 383 AssocProf      A       8       5    Male  86895
```

```
## 384      Prof          A          44          44    Male 105000
## 385      Prof          A          27          21    Male 125192
## 386      Prof          A          15           9    Male 114330
## 387      Prof          A          29          27    Male 139219
## 388      Prof          A          29          15    Male 109305
## 389      Prof          A          38          36    Male 119450
## 390      Prof          A          33          18    Male 186023
## 391      Prof          A          40          19    Male 166605
## 392      Prof          A          30          19    Male 151292
## 393      Prof          A          33          30    Male 103106
## 394      Prof          A          31          19    Male 150564
## 395      Prof          A          42          25    Male 101738
## 396      Prof          A          25          15    Male  95329
## 397   AsstProf         A           8           4    Male  81035
```

1.a.

```
fit_1a <- lm(salary ~ sex, data = Salaries)
fit_1a
```

```
##
## Call:
## lm(formula = salary ~ sex, data = Salaries)
##
## Coefficients:
## (Intercept)        sexMale
##      101002          14088
```

```
# The general linear model with the response variable being 'salary' and a
single predictor being 'sex' is: salary = 101002 + 14088 * I(sex = Male)
```

```
summary(fit_1a)
```

```
##
## Call:
## lm(formula = salary ~ sex, data = Salaries)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -57290 -23502  -6828  19710 116455
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   101002       4809  21.001  < 2e-16 ***
## sexMale        14088       5065   2.782  0.00567 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 30030 on 395 degrees of freedom
## Multiple R-squared:  0.01921,    Adjusted R-squared:  0.01673
## F-statistic: 7.738 on 1 and 395 DF,  p-value: 0.005667
```

```
# p-value for sexMale is 0.005667, less than 0.05, reject H0.
# Gender is a significant predictor of salary.
```

1.b.
```
Salaries$sex <- relevel(factor(Salaries$sex), ref = 'Male')
fit_1b <- lm(salary ~ sex, data = Salaries)
fit_1b

##
## Call:
## lm(formula = salary ~ sex, data = Salaries)
##
## Coefficients:
## (Intercept)     sexFemale
##      115090        -14088
```

```
# The general linear model with the response variable being 'salary' and a
# single predictor being 'sex' with 'female' group as baseline is: salary =
# 115090 - 14088 * I(sex = Female)
```

1.c.
```
fit_1c <- lm(salary ~ yrs.service + rank + discipline + sex, data = Salaries)
anova(fit_1c)

## Analysis of Variance Table
##
## Response: salary
##               Df     Sum Sq    Mean Sq   F value      Pr(>F)
## yrs.service    1 4.0709e+10 4.0709e+10  79.3405  < 2.2e-16 ***
## rank           2 1.0358e+11 5.1789e+10 100.9335  < 2.2e-16 ***
## discipline     1 1.7617e+10 1.7617e+10  34.3350 9.861e-09 ***
## sex            1 7.7669e+08 7.7669e+08   1.5137     0.2193
## Residuals    391 2.0062e+11 5.1310e+08
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# p-value for yrs.service is 0.426958, greater than the significance level,
# 0.05, not reject H0. yrs.service is not significant.

# p-value for rankAssocProf is 0.000428, less than the significance level,
# 0.05, reject H0. rankAssocProf is significant.

# p-value for rankProf is < 2e-16, less than the significance level, 0.05,
# reject H0. rankProf is significant.

# p-value for disciplineB is 1.24e-08, greater than the significance level,
# 0.05, reject H0. disciplineB is not significant.

# p-value for sexFemale is 0.219311, greater than the significance level,
# 0.05, not reject H0. sexFemale is not significant.
```

```
summary(fit_1c)

##
## Call:
## lm(formula = salary ~ yrs.service + rank + discipline + sex,
##     data = Salaries)
##
## Residuals:
##     Min      1Q Median      3Q     Max
## -64202 -14255   -1533   10571   99163
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   73122.92    3245.27  22.532  < 2e-16 ***
## yrs.service     -88.78     111.64  -0.795 0.426958
## rankAssocProf 14560.40    4098.32   3.553 0.000428 ***
## rankProf      49159.64    3834.49  12.820  < 2e-16 ***
## disciplineB   13473.38    2315.50   5.819 1.24e-08 ***
## sexFemale     -4771.25    3878.00  -1.230 0.219311
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 22650 on 391 degrees of freedom
## Multiple R-squared:  0.4478, Adjusted R-squared:  0.4407
## F-statistic: 63.41 on 5 and 391 DF,  p-value: < 2.2e-16

# LS of salary = 73122.92 - 88.78*yrs.service + 14560.40 * I(rank =
AssocProf) + 49159.64 * I(rank = Prof) + 13473.38 * I(discipline = B) -
4771.25 * I(sex = Female)
```

1.d.

```
# The coefficient of determination is 0.4478, which is greater than 0.2 and
less than 0.6
# This statistic indicates a good linear model fit
```

1.e.

```
par(mfrow = c(2,2))
plot(fit_1c)
```

Residuals vs Fitted

Normal Q-Q

Scale-Location

Residuals vs Leverage
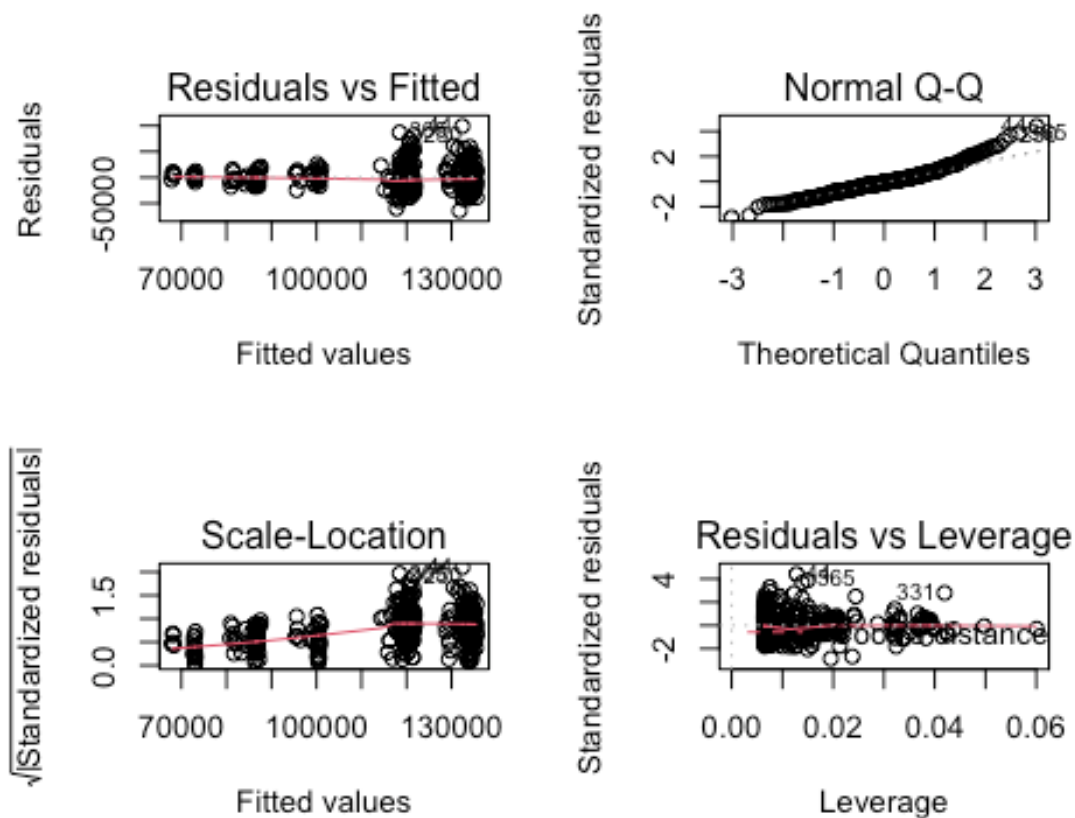
```
# 1. Linearity: it is satisfied because the residuals are symmetrically
distributed around the 0-line in the Residuals vs Fitted plot.

# 2. Homoscedasticity: it is not satisfied because the square root of
standardized residuals is symmetrically distributed around the 1-line in the
Scale-Location plot.

# 3. Independence: assume it is satisfied

# 4. Normality:
shapiro.test(residuals(fit_1c))

##
##  Shapiro-Wilk normality test
##
## data:  residuals(fit_1c)
## W = 0.96073, p-value = 8.202e-09

# p-value is 8.202e-09, less than the significance level 0.05, so residuals
is not normal distributed, normality assumption is not satisfied
```

Question 02

```
install.packages("tidyverse")
install.packages("caret")
install.packages("leaps")
install.packages("MASS")
library(tidyverse)

## — Attaching packages ——————————————————————————— tidyverse
1.3.1 —

## ✓ ggplot2 3.3.5      ✓ purrr    0.3.4
## ✓ tibble  3.1.4      ✓ dplyr    1.0.7
## ✓ tidyr   1.1.3      ✓ stringr 1.4.0
## ✓ readr   2.0.1      ✓ forcats 0.5.1

## — Conflicts ——————————————————————————————
tidyverse_conflicts() —
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

library(caret)

## Loading required package: lattice

##
## Attaching package: 'caret'

## The following object is masked from 'package:purrr':
##
##      lift

library(leaps)
library(MASS)

##
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':
##
##      select

data("swiss")
sample_n(swiss, 3)

##           Fertility Agriculture Examination Education Catholic
Infant.Mortality
## Le Locle       72.7        16.7          22        13    11.22
18.9
## Morges         65.5        59.8          22        10     5.23
18.0
```

```
## Payerne        74.2        58.1        14        8      5.23
23.8
```

2.a.

```
models_2a <- regsubsets(Fertility ~ ., data = swiss, nvmax = 5)
summary(models_2a)

## Subset selection object
## Call: regsubsets.formula(Fertility ~ ., data = swiss, nvmax = 5)
## 5 Variables  (and intercept)
##                 Forced in Forced out
## Agriculture        FALSE      FALSE
## Examination        FALSE      FALSE
## Education          FALSE      FALSE
## Catholic           FALSE      FALSE
## Infant.Mortality   FALSE      FALSE
## 1 subsets of each size up to 5
## Selection Algorithm: exhaustive
##          Agriculture Examination Education Catholic Infant.Mortality
## 1  ( 1 ) " "         " "         "*"       " "      " "
## 2  ( 1 ) " "         " "         "*"       "*"      " "
## 3  ( 1 ) " "         " "         "*"       "*"      "*"
## 4  ( 1 ) "*"         " "         "*"       "*"      "*"
## 5  ( 1 ) "*"         "*"         "*"       "*"      "*"
```

```
# Best model with 1 variable: Fertility ~ Education
# Best model with 2 variables: Fertility ~ Education + Catholic
# Best model with 3 variables: Fertility ~ Education + Catholic +
Infant.Mortality
# Best model with 4 variables: Fertility ~ Agriculture + Education + Catholic
+ Infant.Mortality
# Best model with 5 variable: Fertility ~ Agriculture + Examination +
Education + Catholic + Infant.Mortality
```

2.b.

```
get_model_formula <- function(id, object, outcome){
  models <- summary(object)$which[id,-1]
  predictors <- names(which(models == TRUE))
  predictors <- paste(predictors, collapse = "+")
  as.formula(paste0(outcome, "~", predictors))
}

get_cv_error <- function(model.formula, data){
  set.seed(1)
  train.control <- trainControl(method = "cv", number = 5)
  cv <- train(model.formula, data = data, method = "lm",
            trControl = train.control)
  cv$results$RMSE
}
```

```
model.ids <- 1:5
cv.errors <-  map(model.ids, get_model_formula, models_2a, "Fertility") %>%
  map(get_cv_error, data = swiss) %>%
  unlist()
cv.errors

## [1] 9.464156 8.517433 7.855267 7.601072 7.736328

which.min(cv.errors)

## [1] 4

coef(models_2a, 4)

##       (Intercept)        Agriculture        Education        Catholic
##        62.1013116        -0.1546175        -0.9802638        0.1246664
## Infant.Mortality
##         1.0784422

# The equation of the best overall model: Fertility = 62.1013116 -
0.1546175*Agriculture - 0.9802638*Education + 0.1246664*Cartholic +
1.0784422*Infant.Mortality
```

2.c.

```
stepAIC(lm(Fertility ~ Agriculture + Examination + Education + Catholic +
Infant.Mortality, data = swiss), direction = 'both', k = log(nrow(swiss)))

## Start:  AIC=201.79
## Fertility ~ Agriculture + Examination + Education + Catholic +
##     Infant.Mortality
##
##                    Df Sum of Sq    RSS    AIC
## - Examination       1     53.03 2158.1 199.11
## <none>                          2105.0 201.79
## - Agriculture       1    307.72 2412.8 204.35
## - Infant.Mortality  1    408.75 2513.8 206.28
## - Catholic          1    447.71 2552.8 207.00
## - Education         1   1162.56 3267.6 218.61
##
## Step:  AIC=199.11
## Fertility ~ Agriculture + Education + Catholic + Infant.Mortality
##
##                    Df Sum of Sq    RSS    AIC
## <none>                          2158.1 199.11
## - Agriculture       1    264.18 2422.2 200.69
## + Examination       1     53.03 2105.0 201.79
## - Infant.Mortality  1    409.81 2567.9 203.43
## - Catholic          1    956.57 3114.6 212.50
## - Education         1   2249.97 4408.0 228.83
```

```
## 
## Call:
## lm(formula = Fertility ~ Agriculture + Education + Catholic + 
##     Infant.Mortality, data = swiss)
## 
## Coefficients:
##     (Intercept)      Agriculture        Education         Catholic
##         62.1013          -0.1546          -0.9803           0.1247
## Infant.Mortality
##          1.0784

# The equation of the best overall model: Fertility = 62.1013 -
0.1546*Agriculture - 0.9803*Education + 0.1247*Cartholic +
1.0784*Infant.Mortality
```
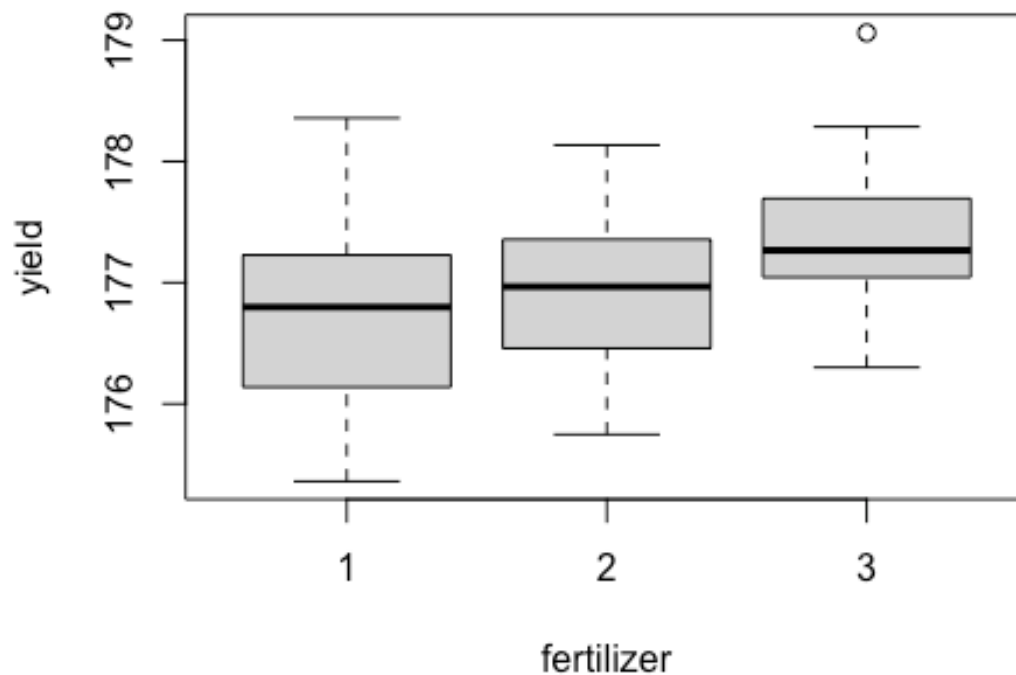
Question 03

```
crop_data <- read.csv('~/OneDrive - Stony Brook University/SBU/MAT + AMS/Fall
2021/AMS 380/hw/04/crop.data.csv', header = T)
```

3.a.

```
boxplot(yield ~ fertilizer, data = crop_data)
```

3.b.

```r
fit_3b <- lm(yield ~ as.factor(fertilizer), data = crop_data)
anova(fit_3b)

## Analysis of Variance Table
##
## Response: yield
##                       Df Sum Sq Mean Sq F value    Pr(>F)
## as.factor(fertilizer)  2  6.068 3.03402  7.8628 0.0006999 ***
## Residuals             93 35.886 0.38587
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# H0: mu1 = mu2 = mu3     Ha: at least one mu is different.

# p-value for the F-test is 0.0006999, less than the significance level 0.05,
reject H0.
# The effect of fertilizer is significant and the mean of different groups is
different.
```

3.c.

```r
res.aov <- aov(yield ~ factor(fertilizer), data = crop_data)
TukeyHSD(res.aov)

##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = yield ~ factor(fertilizer), data = crop_data)
##
## $`factor(fertilizer)`
##          diff         lwr       upr     p adj
## 2-1 0.1761687 -0.19371896 0.5460564 0.4954705
## 3-1 0.5991256  0.22923789 0.9690133 0.0006125
## 3-2 0.4229569  0.05306916 0.7928445 0.0208735

# p-value of comparison of fertilizer 1 and 2 is 0.4954705, greater than the
significance level 0.05, not reject H0. The mean of fertilizer 1 and 2 are
the same.

# p-value of comparison of fertilizer 1 and 3 is 0.0006125, less than the
significance level 0.05, reject H0. The mean of fertilizer 1 and 3 are
different.

# p-value of comparison of fertilizer 2 and 3 is 0.0208735, less than the
significance level 0.05, reject H0. The mean of fertilizer 2 and 3 are
different.
```

3.d.

```r
# H0: mu2 = mu3     Ha: mu2 != mu3
```

```
yield_2 <- crop_data$yield[crop_data$fertilizer == 2]
yield_3 <- crop_data$yield[crop_data$fertilizer == 3]

shapiro.test(yield_2)

##
##  Shapiro-Wilk normality test
##
## data:  yield_2
## W = 0.98329, p-value = 0.8875

shapiro.test(yield_3)

##
##  Shapiro-Wilk normality test
##
## data:  yield_3
## W = 0.95878, p-value = 0.2542
```

 # p-value of the shapiro test of yield_2 and yield_3 are 0.8875 and 0.2542,
both greater than the significance level 0.05, not reject H0, the samples are
both normal.

```
var.test(yield_2,yield_3)

##
##  F test to compare two variances
##
## data:  yield_2 and yield_3
## F = 0.91811, num df = 31, denom df = 31, p-value = 0.8135
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.448169 1.880826
## sample estimates:
## ratio of variances
##           0.918111
```

 # p-value of variance test is 0.8135, greater than the significance level
0.05, the variances of yield_2 and yield_3 are assumed to be the same.

```
t.test(yield_2, yield_3, mu = 0, var.equal = T)

##
##  Two Sample t-test
##
## data:  yield_2 and yield_3
## t = -2.8835, df = 62, p-value = 0.0054
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.7161699 -0.1297438
## sample estimates:
```

```
## mean of x mean of y
##  176.9332  177.3562
```

```
# p-value of the t-test is 0.0054, less than the significance level 0.05,
reject H0. The mean of yield of fertilizer 2 and 3 are significantly
different.
```