List 06: Multicollinearity

Nikita V. Artamonov

Contents

sleep equation #1

For the dataset `sleep75` consider a regreaion **sleep ~ 1 + I(totwrk/100) + I(totwrk^2/10000) + age + smsa + , male**.

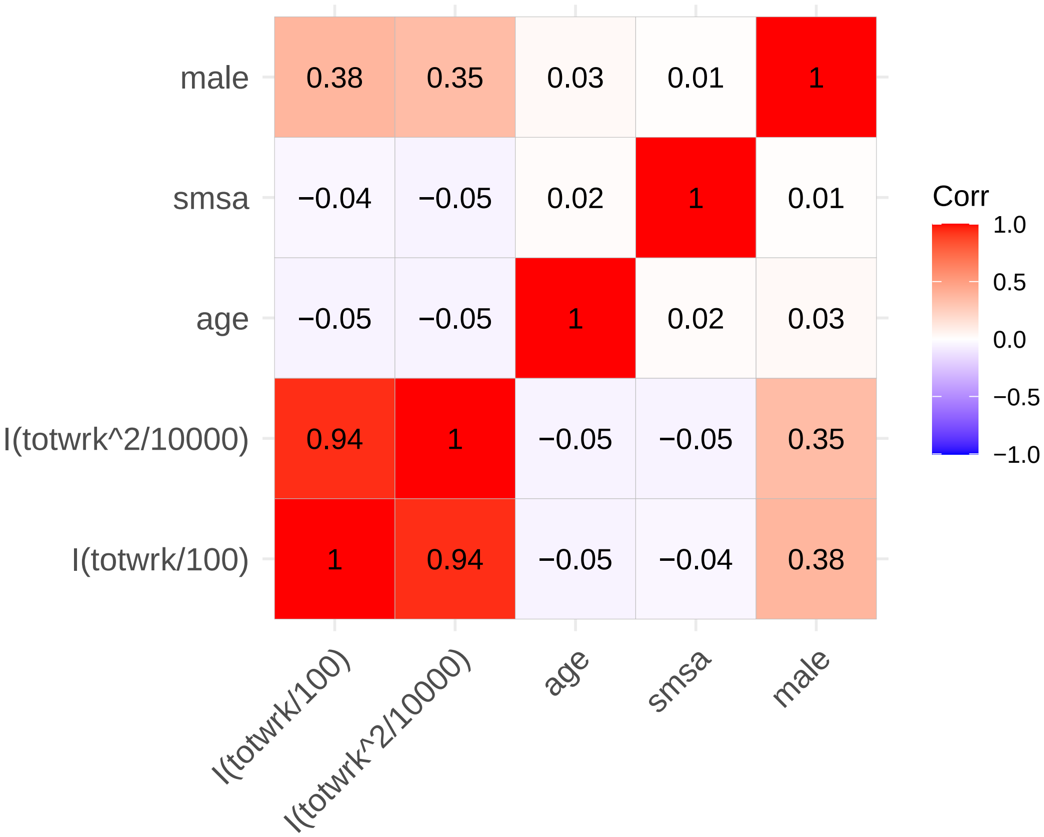
Evaluate VIF for each regressors

```
    I(totwrk/100) I(totwrk^2/10000)               age              smsa
         8.370495          8.199290          1.005968          1.004459
             male
         1.169198
```

Calculate correlation matirx for regressosrs

```
                  I(totwrk/100) I(totwrk^2/10000)    age   smsa  male
I(totwrk/100)             1.000             0.937 -0.050 -0.038 0.376
I(totwrk^2/10000)         0.937             1.000 -0.046 -0.051 0.351
age                      -0.050            -0.046  1.000  0.025 0.032
smsa                     -0.038            -0.051  0.025  1.000 0.007
male                      0.376             0.351  0.032  0.007 1.000
```

and visualize it



Significant level is 5%. Which coefficients are sognoficant (perform non-robust t-test)?

```
[1] "age"   "smsa" "male"
```

We test the significance of working time, i.e. the hypothesis $H_0 : \beta_{totwrk/100} = \beta_{totwrk^2/10000} = 0$. Testing result (Non-robust test):

```
==============
F      Pr(> F)
--------------
45.619    0
--------------
```

Calculate the required critical value. **Round to 2 decimal places.**

```
[1] 3.01
```

Inferences:

```
[1] "We reject the null hypothesis"
```

At first glance we have a contradiction. It is caused by multicollinearity.

sleep equation #2

For the dataset `sleep75` consider a regreaion **sleep ~ totwrk + age + I(age^2) + smsa + male + union**.

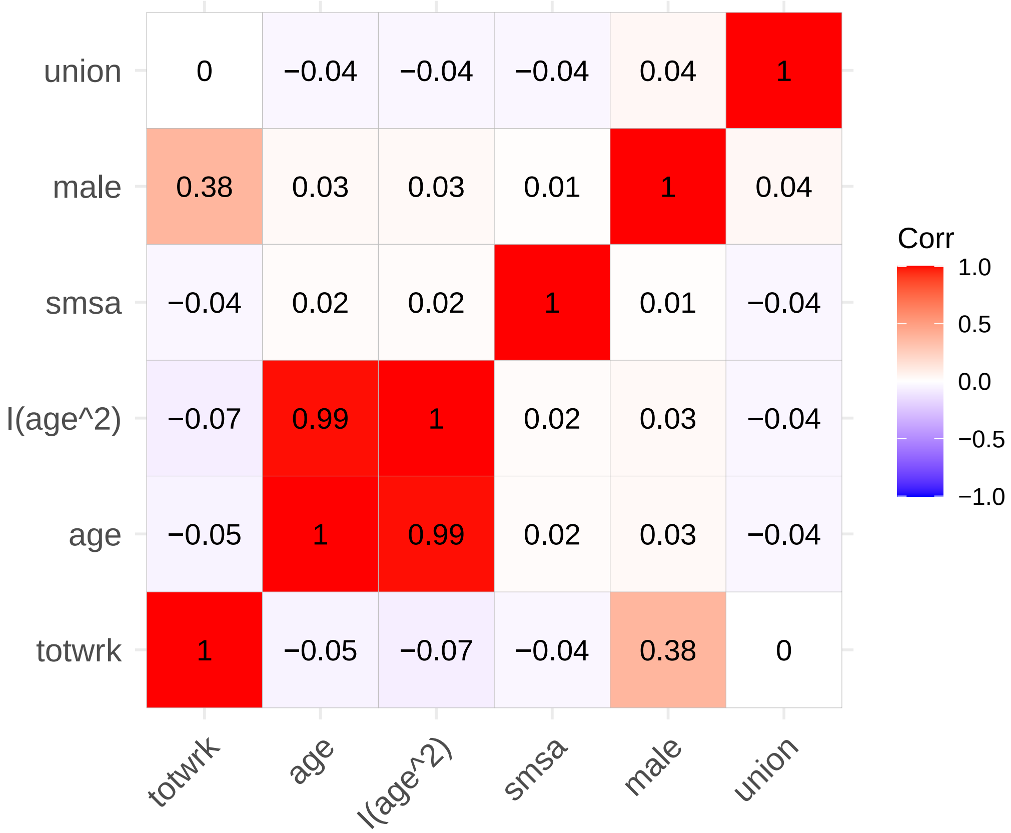
Evaluate VIF for each regressors

```
   totwrk       age  I(age^2)      smsa      male     union 
 1.195469 65.397082 65.561373  1.004278  1.171666  1.007332 
```

Calculate correlation matirx for regressosrs

```
         totwrk    age I(age^2)   smsa  male  union
totwrk    1.000 -0.050   -0.067 -0.038 0.376  0.002
age      -0.050  1.000    0.992  0.025 0.032 -0.037
I(age^2) -0.067  0.992    1.000  0.024 0.026 -0.042
smsa     -0.038  0.025    0.024  1.000 0.007 -0.039
male      0.376  0.032    0.026  0.007 1.000  0.040
union     0.002 -0.037   -0.042 -0.039 0.040  1.000
```

and visualize it



Significant level is 5%. Which coefficients are sognoficant (perform non-robust t-test)?

```
[1] "totwrk" "smsa"   "male"
```

We test the significance of age, i.e. the hypothesis $H_0 : \beta_{age} = \beta_{age^2} = 0$. Testing result (Non-robust test):

```
=============
F     Pr(> F)
-------------
2.497  0.083
-------------
```

Calculate the required critical value. **Round to 2 decimal places.**

```
[1] 3.01
```

Inferences:

```
[1] "We do not reject the null hypothesis"
```

At first glance we have a contradiction. It is caused by multicollinearity.

sleep equation #3

For the dataset `sleep75` consider a regreaion **sleep ~ totwrk + age + smsa + south + I(totwrk * south) + I(age * , south) + I(smsa * south)**.

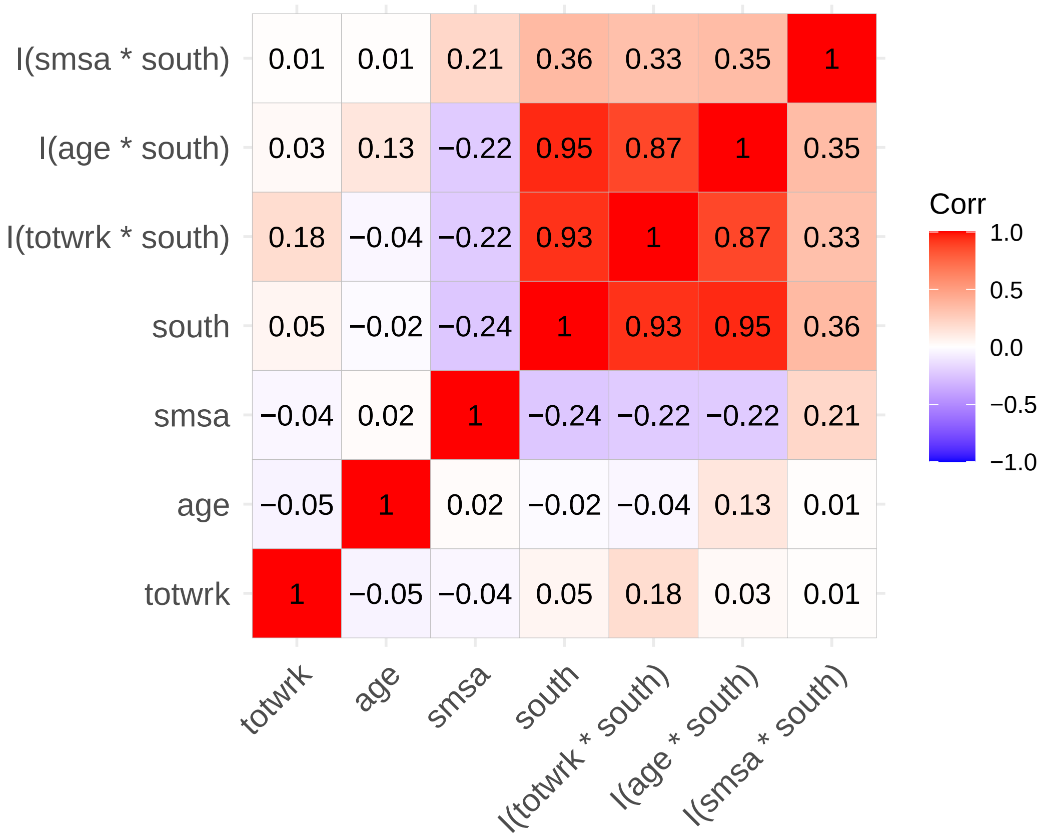
Evaluate VIF for each regressors

```
           totwrk               age              smsa             south 
         1.148498          1.249261          1.187071         22.001994 
I(totwrk * south)    I(age * south)   I(smsa * south) 
         8.851673         12.277039          1.286230 
```

Calculate correlation matirx for regressosrs

```
                  totwrk    age   smsa  south I(totwrk * south) I(age * south)
totwrk             1.000 -0.050 -0.038  0.051             0.175          0.033
age               -0.050  1.000  0.025 -0.018            -0.038          0.126
smsa              -0.038  0.025  1.000 -0.238            -0.224         -0.222
south              0.051 -0.018 -0.238  1.000             0.932          0.947
I(totwrk * south)  0.175 -0.038 -0.224  0.932             1.000          0.868
I(age * south)     0.033  0.126 -0.222  0.947             0.868          1.000
I(smsa * south)    0.012  0.008  0.209  0.359             0.328          0.351
                  I(smsa * south)
totwrk                      0.012
age                         0.008
smsa                        0.209
south                       0.359
I(totwrk * south)           0.328
I(age * south)              0.351
I(smsa * south)             1.000
```

and visualize it



Significant level is 5%. Which coefficients are sognoficant (perform non-robust t-test)?

```
[1] "totwrk"          "south"           "I(age * south)"
```

We test the significance of geographical dummy, i.e. the hypothesis $H_0 : \beta_{south} = \beta_{totwrk*south} = \beta_{age*south} = \beta_{smsa*south} = 0$. Testing result (Non-robust test):

```
=============
F     Pr(> F)
-------------
3.144  0.014
-------------
```

Calculate the required critical value. **Round to 2 decimal places.**

```
[1] 2.38
```

Inferences:

```
[1] "We reject the null hypothesis"
```

At first glance we have a contradiction. It is caused by multicollinearity.

wage equation #1

For the dataset `wage2` consider a regreaion **log(wage)** ~ **age** + **I(age^2)** + **IQ** + **married** + **south** + **urban**.

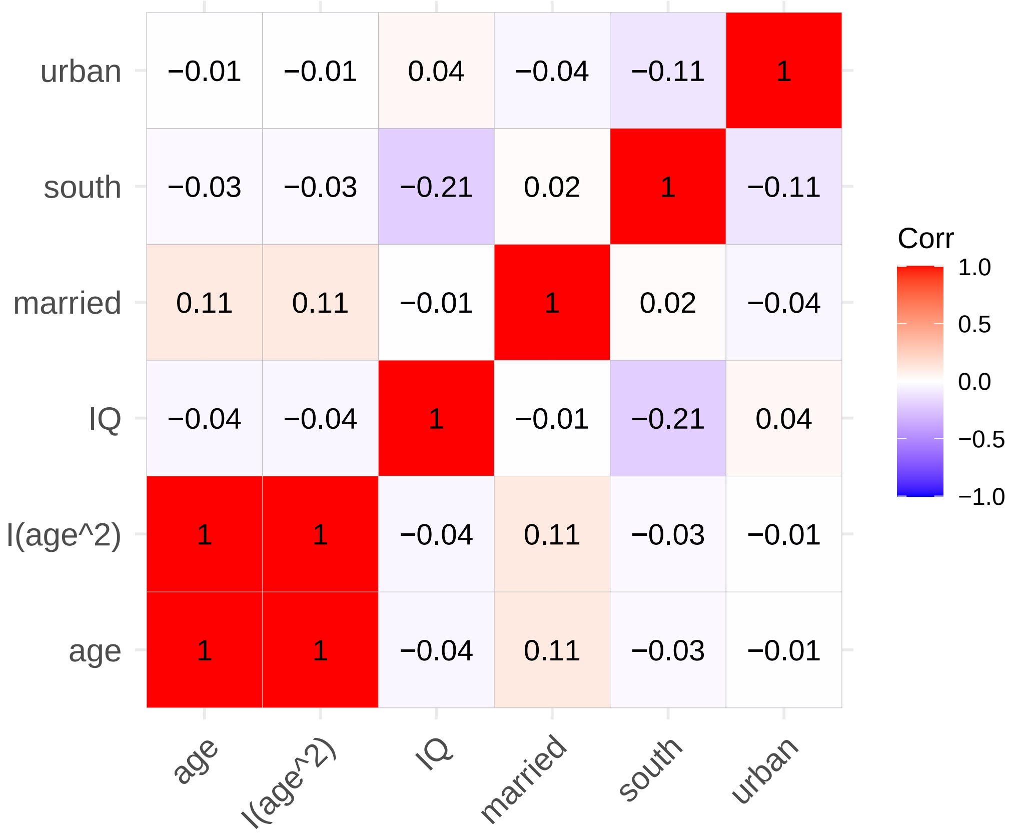
Evaluate VIF for each regressors

```
       age   I(age^2)         IQ    married      south      urban 
632.868517 632.964483   1.049260   1.013807   1.061287   1.016749 
```

Calculate correlation matirx for regressosrs

```
            age I(age^2)     IQ married  south  urban
age       1.000    0.999 -0.044   0.107 -0.029 -0.007
I(age^2)  0.999    1.000 -0.043   0.107 -0.031 -0.009
IQ       -0.044   -0.043  1.000  -0.015 -0.210  0.039
married   0.107    0.107 -0.015   1.000  0.023 -0.040
south    -0.029   -0.031 -0.210   0.023  1.000 -0.110
urban    -0.007   -0.009  0.039  -0.040 -0.110  1.000
```

and visualize it



Significant level is 5%. Which coefficients are sognoficant (perform non-robust t-test)?

```
[1] "IQ"      "married" "south"   "urban"  
```

We test the significance of age, i.e. the hypothesis $_0 : \beta_{age} = \beta_{age^2} = 0$. Testing result (Non-robust test):

```
==============
F      Pr(> F)
--------------
14.833 0.00000
--------------
```

Calculate the required critical value. **Round to 2 decimal places.**

```
[1] 3.01
```

Inferences:

```
[1] "We reject the null hypothesis"
```

At first glance we have a contradiction. It is caused by multicollinearity.

wage equation #2 (structural breaks)

`wage2` **log(wage) ~ age + IQ + south + urban + I(age * urban) + I(IQ * , urban) + I(south * urban)**.

For the dataset `wage2` consider a regreaion **log(wage) ~ age + IQ + south + urban + I(age * urban) + I(IQ * , urban) + I(south * urban)**.

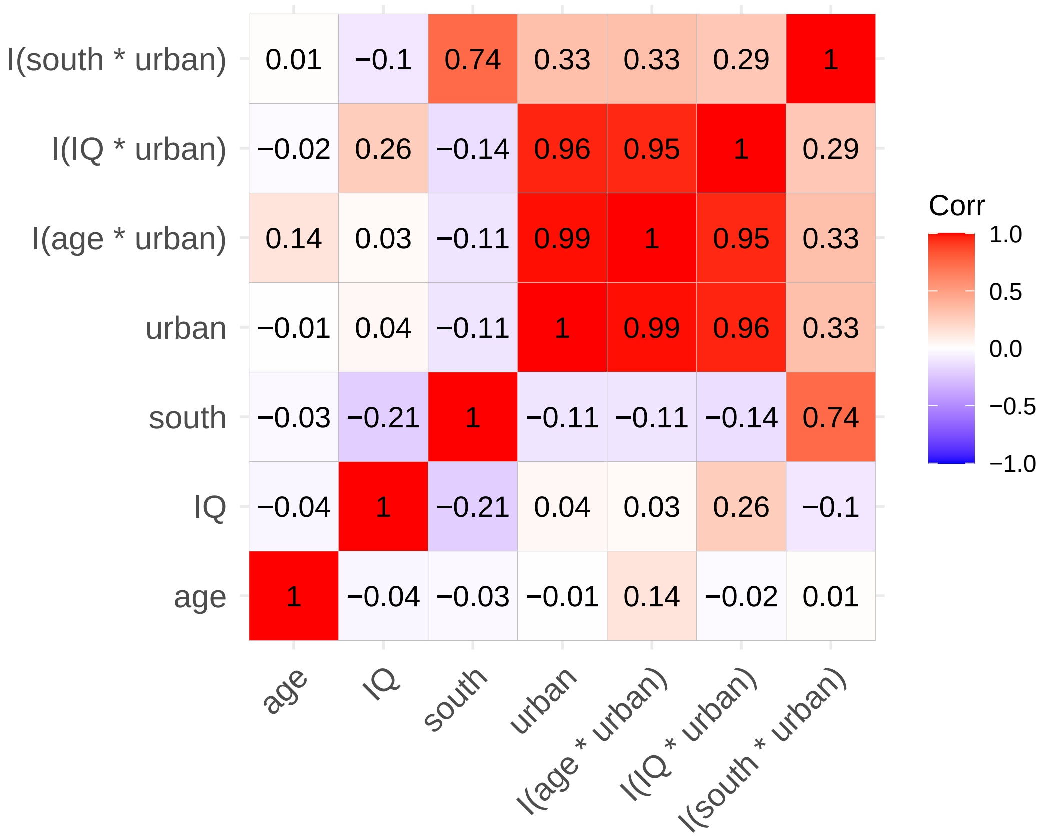
Evaluate VIF for each regressors

```
             age                IQ             south             urban 
        3.394929          3.879824          3.832146        175.876722 
  I(age * urban)     I(IQ * urban) I(south * urban) 
      114.318464         53.135490          4.087171 
```

Calculate correlation matirx for regressosrs

```
                    age     IQ  south  urban I(age * urban) I(IQ * urban)
age               1.000 -0.044 -0.029 -0.007          0.137        -0.020
IQ               -0.044  1.000 -0.210  0.039          0.030         0.260
south            -0.029 -0.210  1.000 -0.110         -0.106        -0.136
urban            -0.007  0.039 -0.110  1.000          0.985         0.964
I(age * urban)    0.137  0.030 -0.106  0.985          1.000         0.947
I(IQ * urban)    -0.020  0.260 -0.136  0.964          0.947         1.000
I(south * urban)  0.010 -0.097  0.741  0.334          0.332         0.288
                 I(south * urban)
age                         0.010
IQ                         -0.097
south                       0.741
urban                       0.334
I(age * urban)              0.332
I(IQ * urban)               0.288
I(south * urban)            1.000
```

and visualize it



Significant level is 1%. Which coefficients are sognoficant (perform non-robust t-test)?

```
[1] "age" "IQ"
```

We test the significance of dwelling dummy, i.e. the hypothesis $H_0 : \beta_{urban} = \beta_{age*urban} = \beta_{IQ*urban} = \beta_{south*urban} = 0$. Testing result (Non-robust test):

```
==============
F      Pr(> F)
--------------
10.250 0.00000
--------------
```

Calculate the required critical value. **Round to 2 decimal places.**

```
[1] 3.34
```

Inferences:

```
[1] "We reject the null hypothesis"
```

At first glance we have a contradiction. It is caused by multicollinearity.