

# Задачи по Эконометрике: Мультиколлинеарность

Н.В. Артамонов (МГИМО МИД России)

## Содержание

1	sleep equation #1	1
2	sleep equation #2	2
3	sleep equation #3	4
4	wage equation #1	5
5	wage equation #2 (структурные сдвиги)	7

## 1 sleep equation #1

Для набора данных sleep75 рассмотрим линейную регрессию  $\text{sleep} \sim 1 + I(\text{totwrk}/100) + I(\text{totwrk}^2/10000) + \text{age} + \text{smsa} + \text{male}$ .

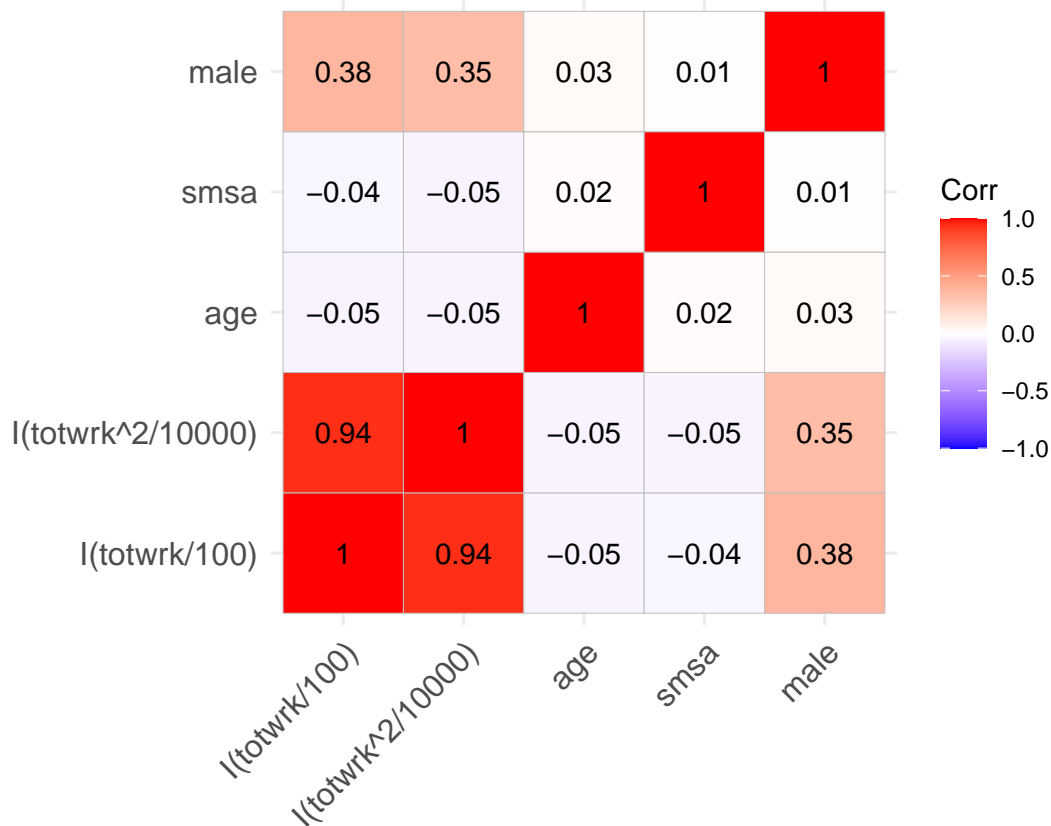
Вычислите на VIF для регрессоров

$I(\text{totwrk}/100)$	$I(\text{totwrk}^2/10000)$	age	smsa
8.370495	8.199290	1.005968	1.004459
male			
1.169198			

Вычислите корреляционную матрицу для регрессоров

	$I(\text{totwrk}/100)$	$I(\text{totwrk}^2/10000)$	age	smsa	male
$I(\text{totwrk}/100)$	1.000	0.937	-0.050	-0.038	0.376
$I(\text{totwrk}^2/10000)$	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

и визуализируйте её



Уровень значимости 5%. Какие коэффициенты значимы (неробастный t-тест)? Ответ

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

Тестируется значимость влияния занятости, т.е. гипотеза  $H_0 : \beta_{totwrk/100} = \beta_{totwrk^2/10000} = 0$ . Результаты тестирования (неробастный тест):

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

Вычислите необходимое критическое значение. Ответ округлите до 2-х десятичных знаков.

```
[1] 3.01
```

Какие результаты тестирования? Ответ

```
[1] "Гипотеза отвергается"
```

**На первый взгляд противоречие.** Оно объясняется мультиколлинеарностью

## 2 sleep equation #2

Для набора данных `sleep75` рассмотрим линейную регрессию `sleep ~ totwrk + age + I(age^2) + smsa + male + union`.

Вычислите на VIF для регрессоров

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

Вычислите корреляционную матрицу для регрессоров

```
      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
```

и визуализируйте её



Уровень значимости 5%. Какие коэффициенты значимы (неробастный t-тест)? Ответ

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

Тестируется значимость влияния возраста, т.е. гипотеза  $H_0 : \beta_{age} = \beta_{age^2} = 0$ . Результаты тестирования (неробастный тест):

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

Вычислите необходимое критическое значение. Ответ округлите до 2-х десятичных знаков.

```
[1] 3.01
```

Какие результаты тестирования? Ответ

[1] "Гипотеза не отвергается"

На первый взгляд противоречие. Оно объясняется мультиколлинеарностью

### 3 sleep equation #3

Для набора данных `sleep75` рассмотрим линейную регрессию `sleep ~ totwrk + age + smsa + south + I(totwrk * south) + I(age * , south) + I(smsa * south)`.

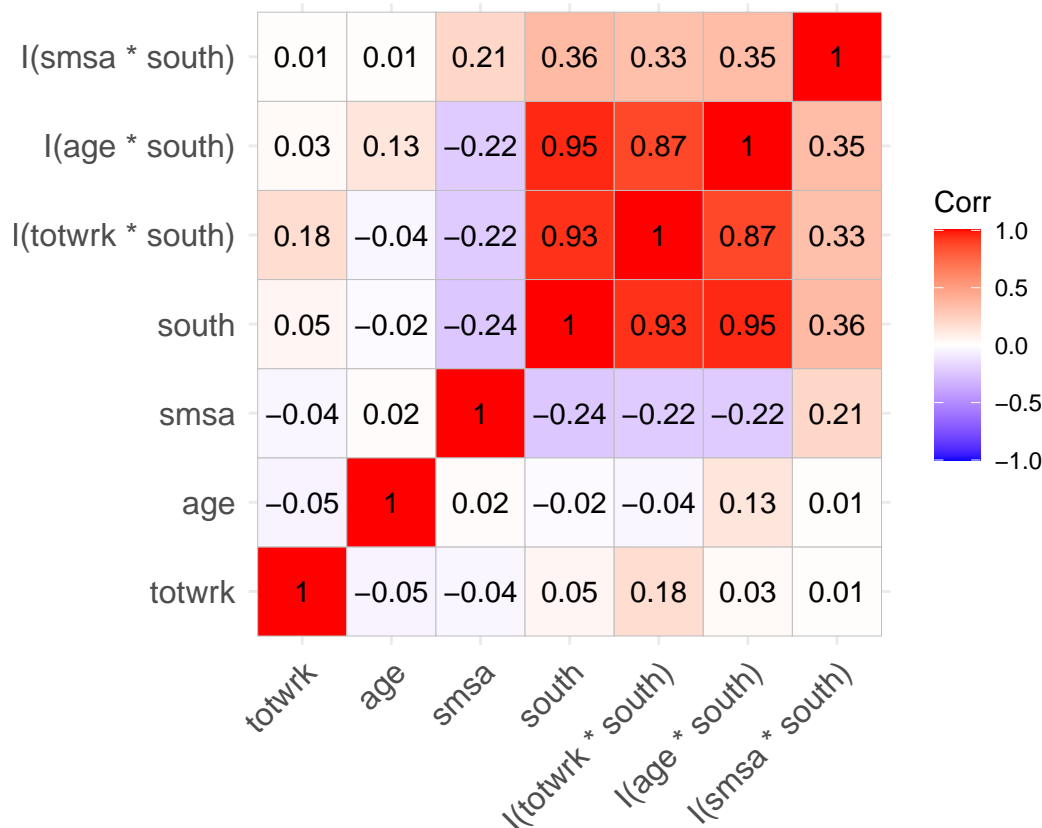
Вычислите на VIF для регрессоров

	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	

Вычислите корреляционную матрицу для регрессоров

	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				

и визуализируйте её



Уровень значимости 5%. Какие коэффициенты значимы (неробастный t-тест)? Ответ

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

Тестируется значимость влияния географической бинарной переменной, т.е. гипотеза  $H_0 : \beta_{south} = \beta_{totwrk*south} = \beta_{age*south} = \beta_{smsa*south} = 0$ . Результаты тестирования (неробастный тест):

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

Вычислите необходимое критическое значение. Ответ округлите до 2-х десятичных знаков.

```
[1] 2.38
```

Какие результаты тестирования? Ответ

```
[1] "Гипотеза отвергается"
```

На первый взгляд противоречие. Оно объясняется мультиколлинеарностью

## 4 wage equation #1

Для набора данных wage2 рассмотрим линейную регрессию  $\log(\text{wage}) \sim \text{age} + \text{I}(\text{age}^2) + \text{IQ} + \text{married} + \text{south} + \text{urban}$ .

Вычислите на VIF для регрессоров

```

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

```

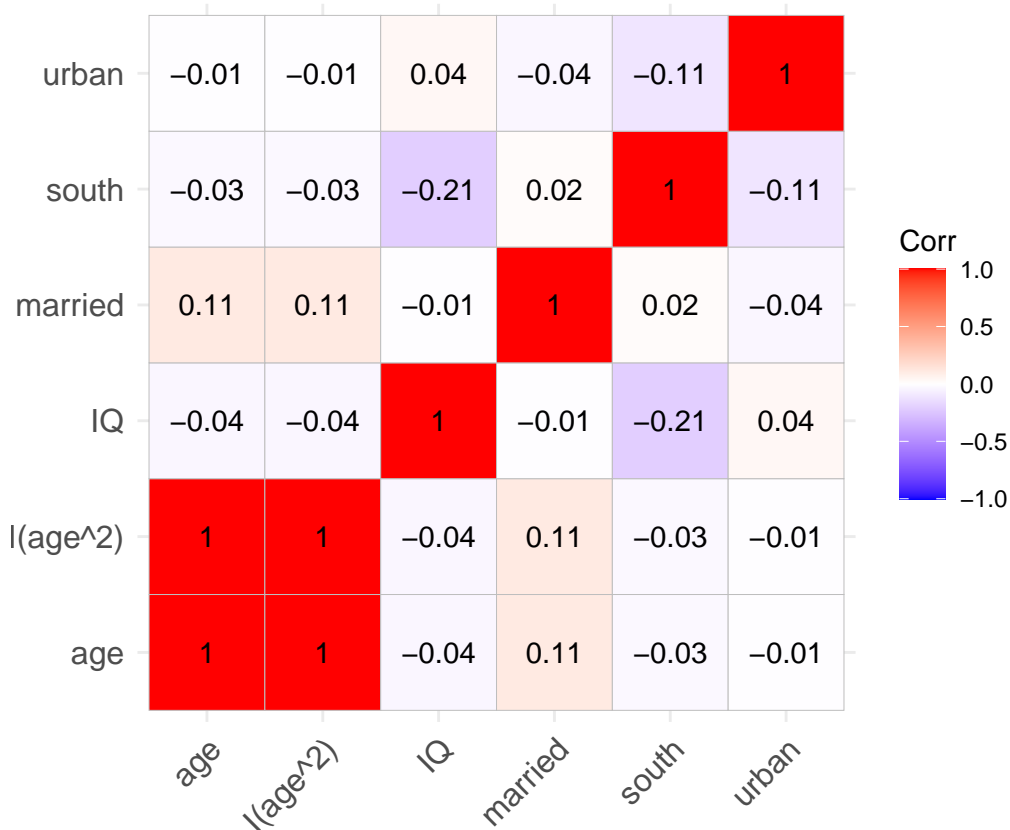
Вычислите корреляционную матрицу для регрессоров

```

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

```

и визуализируйте её



Уровень значимости 5%. Какие коэффициенты значимы (неробастный t-тест)? Ответ

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

Тестируется значимость влияния возраста, т.е. гипотеза  $H_0 : \beta_{age} = \beta_{age^2} = 0$ . Результаты тестирования (неробастный тест):

```

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

```

Вычислите необходимое критическое значение. Ответ округлите до 2-х десятичных знаков.

```
[1] 3.01
```

Какие результаты тестирования? Ответ

[1] "Гипотеза отвергается"

На первый взгляд противоречие. Оно объясняется мультиколлинеарностью

## 5 wage equation #2 (структурные сдвиги)

Для набора данных wage2 рассмотрим линейную регрессию  $\log(\text{wage}) \sim \text{age} + \text{IQ} + \text{south} + \text{urban} + \text{I}(\text{age} * \text{urban}) + \text{I}(\text{IQ} * \text{urban}) + \text{I}(\text{south} * \text{urban})$ .

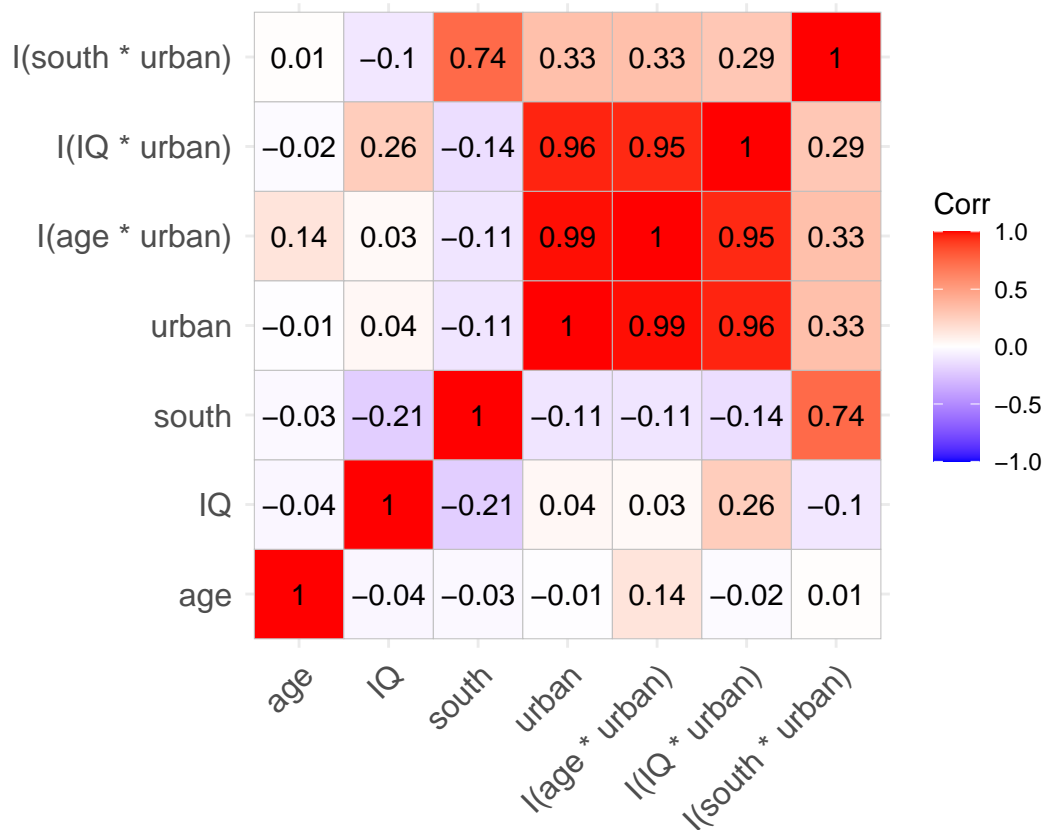
Вычислите на VIF для регрессоров

	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	

Вычислите корреляционную матрицу для регрессоров

	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				

и визуализируйте её



Уровень значимости 1%. Какие коэффициенты значимы (неробастный t-тест)? Ответ

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

Тестируется значимость влияния места жительства, т.е. гипотеза  $H_0 : \beta_{urban} = \beta_{age*urban} = \beta_{IQ*urban} = \beta_{south*urban} = 0$  Результаты тестирования (неробастный тест):

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

Вычислите необходимое критическое значение. Ответ округлите до 2-х десятичных знаков.

```
[1] 3.34
```

Какие результаты тестирования? Ответ

```
[1] "Гипотеза отвергается"
```

**На первый взгляд противоречие.** Оно объясняется мультиколлинеарностью