

# Задачи по Эконометрике: Категориальные переменные

Н.В. Артамонов (МГИМО МИД России)

## Содержание

<b>1 Категориальные признаки</b>	<b>1</b>
1.1 diamond equation #1	1
1.2 diamond equation #2	2
<b>2 Структурные сдвиги</b>	<b>3</b>
2.1 Графики	3
2.2 sleep equation #1	4
2.3 sleep equation #2	5
2.4 wage equation #1	6
2.5 wage equation #2	7

## 1 Категориальные признаки

### 1.1 diamond equation #1

#### 1.1.1 Описательные статистики

Для набора данных Diamond вычислите описательные статистики для количественных и категориальных переменных

carat	colour	clarity	certification	price
Min. :0.1800	D:16	IF :44	GIA:151	Min. : 638
1st Qu.:0.3500	E:44	VS1 :81	HRD: 79	1st Qu.: 1625
Median :0.6200	F:82	VS2 :53	IGI: 78	Median : 4215
Mean :0.6309	G:65	VVS1:52		Mean : 5019
3rd Qu.:0.8500	H:61	VVS2:78		3rd Qu.: 7446
Max. :1.1000	I:40			Max. :16008

#### 1.1.2 Регрессии

Для набора данных Diamond рассмотрим несколько линейных регрессий.

1.  $\log(\text{price}) \sim \text{carat} + I(\text{carat}^2) + \text{colour}$
2.  $\log(\text{price}) \sim \text{carat} + I(\text{carat}^2) + \text{certification}$
3.  $\log(\text{price}) \sim \text{carat} + I(\text{carat}^2) + \text{colour} + \text{certification}$

Оцените регрессии и дайте интерпретацию коэффициентам для категориальных переменных. **Ответ округлите до 2 десятичных знаков**

Результаты оценивания

(Intercept)	carat	I(carat^2)	colourE	colourF	colourG
5.98	5.40	-1.98	-0.07	-0.14	-0.21
colourH	colourI				
-0.30	-0.40				

```

(Intercept)          carat      I(carat^2) certificationHRD
      5.69            5.65          -2.19              0.03
certificationIGI
      0.07

(Intercept)          carat      I(carat^2)      colourE
      5.87            5.67          -2.15          -0.08
      colourF      colourG      colourH      colourI
     -0.15         -0.23         -0.31         -0.41
certificationHRD certificationIGI
      0.04            0.09

```

## 1.2 diamond equation #2

### 1.2.1 Описательные статистики

Для набора данных diamonds вычислите описательные статистики для количественных и категориальных переменных

```

      carat      cut      color      clarity      depth
Min.   :0.2000 Fair      : 1610 D: 6775 SI1      :13065 Min.   :43.00
1st Qu.:0.4000 Good      : 4906 E: 9797 VS2      :12258 1st Qu.:61.00
Median :0.7000 Very Good:12082 F: 9542 SI2      : 9194 Median :61.80
Mean   :0.7979 Premium  :13791 G:11292 VS1      : 8171 Mean   :61.75
3rd Qu.:1.0400 Ideal    :21551 H: 8304 VVS2     : 5066 3rd Qu.:62.50
Max.   :5.0100                I: 5422 VVS1     : 3655 Max.   :79.00
                                J: 2808 (Other): 2531

      table      price      x      y
Min.   :43.00 Min.   : 326 Min.   : 0.000 Min.   : 0.000
1st Qu.:56.00 1st Qu.: 950 1st Qu.: 4.710 1st Qu.: 4.720
Median :57.00 Median : 2401 Median : 5.700 Median : 5.710
Mean   :57.46 Mean   : 3933 Mean   : 5.731 Mean   : 5.735
3rd Qu.:59.00 3rd Qu.: 5324 3rd Qu.: 6.540 3rd Qu.: 6.540
Max.   :95.00 Max.   :18823 Max.   :10.740 Max.   :58.900

      z
Min.   : 0.000
1st Qu.: 2.910
Median : 3.530
Mean   : 3.539
3rd Qu.: 4.040
Max.   :31.800

```

### 1.2.2 Регрессии

Для набора данных diamonds рассмотрим несколько линейных регрессий.

1.  $\log(\text{price}) \sim \text{carat} + I(\text{carat}^2) + \text{cut} + x + y + z$
2.  $\log(\text{price}) \sim \text{carat} + I(\text{carat}^2) + \text{color} + x + y + z$
3.  $\log(\text{price}) \sim \text{carat} + I(\text{carat}^2) + \text{color} + \text{cut} + x + y + z$

Оцените регрессии и дайте интерпретацию коэффициентам для категориальных переменных. **Ответ округлите до 2 десятичных знаков**

Результаты оценивания

```

(Intercept)      carat      I(carat^2)      cutGood cutVery Good      cutPremium

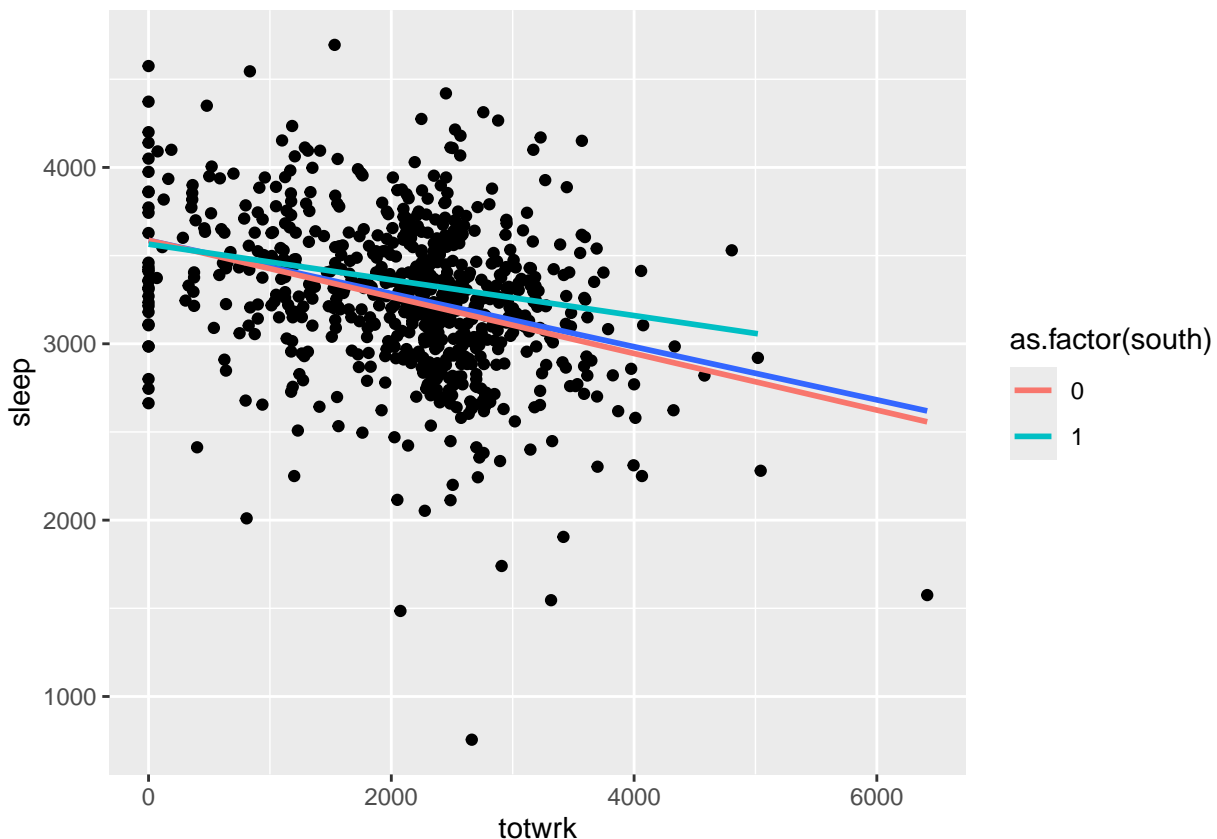
```

	3.80	2.37	-0.64	0.12	0.17	0.16
cutIdeal		x	y	z		
	0.24	0.38	0.02	0.06		
(Intercept)	carat	I(carat^2)	colorE	colorF	colorG	
	4.08	2.35	-0.61	-0.03	-0.04	-0.07
colorH	colorI	colorJ	x	y	z	
	-0.21	-0.30	-0.43	0.38	0.03	0.05
(Intercept)	carat	I(carat^2)	colorE	colorF	colorG	
	3.99	2.54	-0.65	-0.03	-0.04	-0.07
colorH	colorI	colorJ	cutGood	cutVery Good	cutPremium	
	-0.21	-0.31	-0.43	0.12	0.18	0.16
cutIdeal		x	y	z		
	0.25	0.34	0.01	0.06		

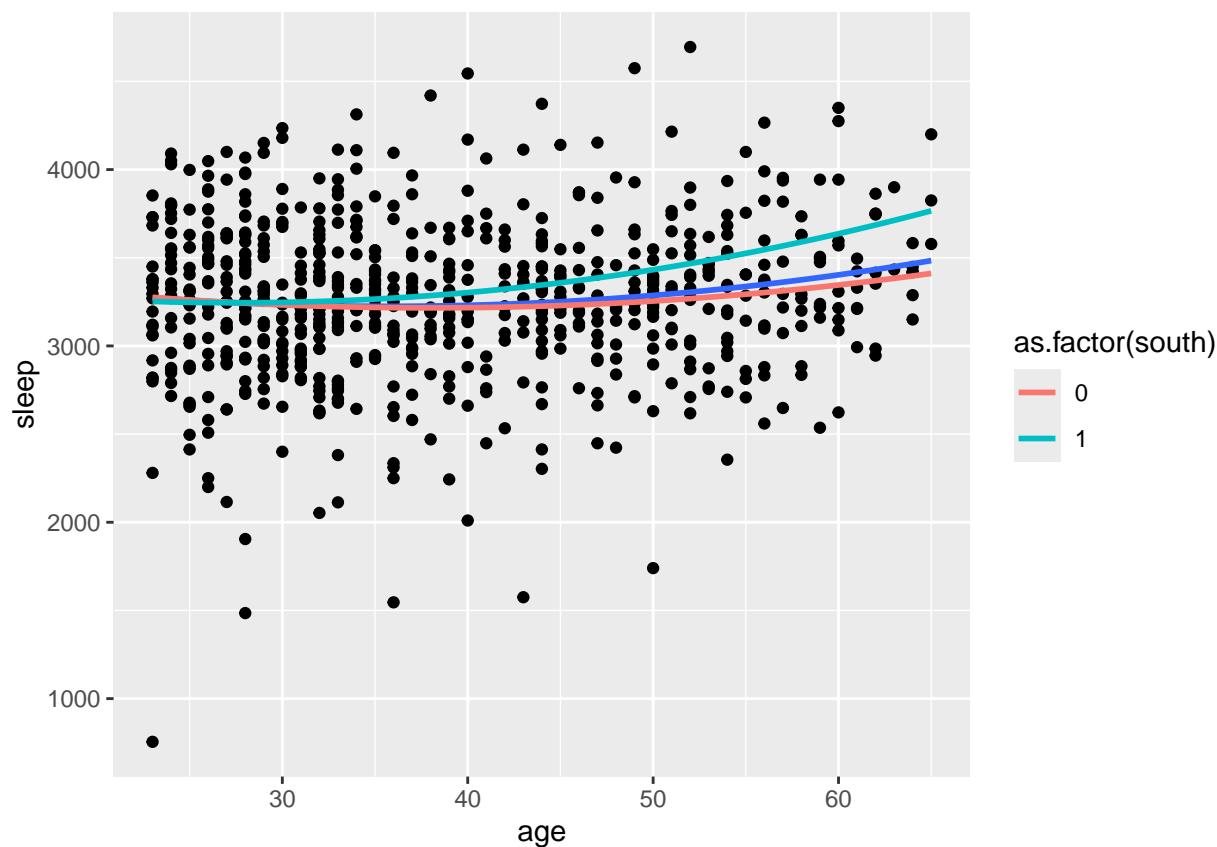
## 2 Структурные сдвиги

### 2.1 Графики

Для набора данных `sleep75` рассмотрим линейную регрессию **sleep** на **totwrk** и подгоним отдельно для `south=0`, отдельно для `south=1` и по полному датасету



Для набора данных `sleep75` рассмотрим линейную регрессию **sleep** на **age**, **age^2** и подгоним отдельно для `south=0`, отдельно для `south=1` и по полному датасету



## 2.2 sleep equation #1

Для набора данных `sleep75` рассмотрим линейную регрессию **sleep** на **totwrk**, **age**, **smsa**: исходную и со структурными сдвигами относительно гендерной переменной.

Результаты оценивания

=====		
	Зависимая переменная	
	-----	
	sleep	
	(1)	(2)
-----	-----	-----
totwrk	-0.150*** (0.017)	-0.143*** (0.026)
age	2.997** (1.393)	1.704 (2.085)
smsa	-73.501** (32.219)	-56.649 (48.902)
male		135.200 (143.832)
totwrk:male		-0.045 (0.036)

age: male		1.500	
		(2.825)	
smsa: male		-34.735	
		(64.938)	
Constant	3498.670***	3486.973***	
	(69.133)	(92.418)	

-----

Observations	706	706
R2	0.115	0.126
Adjusted R2	0.112	0.117
Residual Std. Error	418.877	417.530
F Statistic	30.527***	14.387***

=====

Note:                    \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Для второй регрессии дайте интерпретацию коэффициентов.

## 2.3 sleep equation #2

Для набора данных sleep75 рассмотрим линейную регрессию **sleep** на **totwrk**, **age**, **smsa**: исходную и со структурными сдвигами относительно географической переменной.

Результаты оценивания

=====

	Зависимая переменная	
	sleep	
	(1)	(2)

-----

totwrk	-0.150***	-0.161***
	(0.017)	(0.018)
age	2.997**	1.485
	(1.393)	(1.545)
smsa	-73.501**	-61.554*
	(32.219)	(34.856)
south		-407.206**
		(189.627)
totwrk:south		0.078
		(0.050)
age:south		8.267**
		(3.495)
smsa:south		49.429
		(107.111)

Constant	3498.670*** (69.133)	3557.513*** (75.504)
----------	-------------------------	-------------------------

Observations	706	706
R2	0.115	0.131
Adjusted R2	0.112	0.122
Residual Std. Error	418.877	416.341
F Statistic	30.527***	15.039***

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Для второй регрессии дайте интерпретацию коэффициентов.

## 2.4 wage equation #1

Для набора данных wage2 рассмотрим линейную регрессию **log(wage)** на **age, IQ, urban, married**: исходную и со структурными сдвигами относительно географической переменной.

Результаты оценивания

	Зависимая переменная	
	log (wage)	
	(1)	(2)
age	0.022*** (0.004)	0.024*** (0.005)
IQ	0.009*** (0.001)	0.008*** (0.001)
urban	0.186*** (0.028)	0.217*** (0.035)
married	0.198*** (0.041)	0.201*** (0.049)
south		-0.011 (0.348)
age:south		-0.007 (0.009)
IQ:south		0.002 (0.002)
urban:south		-0.114** (0.057)
married:south		-0.003 (0.087)
Constant	4.849***	4.914***

	(0.163)	(0.204)
-----		
Observations	935	935
R2	0.188	0.205
Adjusted R2	0.184	0.197
Residual Std. Error	0.380	0.377
F Statistic	53.799***	26.481***
=====		
Note:	*p<0.1; **p<0.05; ***p<0.01	

Для второй регрессии дайте интерпретацию коэффициентов.

## 2.5 wage equation #2

Для набора данных wage2 рассмотрим линейную регрессию **log(wage)** на **age, IQ, south, married**: исходную и со структурными сдвигами относительно места жительства.

Результаты оценивания

	Зависимая переменная	
	log (wage)	
	(1)	(2)
-----		
age	0.023*** (0.004)	0.024*** (0.007)
IQ	0.008*** (0.001)	0.008*** (0.002)
south	-0.113*** (0.028)	-0.014 (0.052)
urban		0.237 (0.368)
age:urban		-0.001 (0.009)
IQ:urban		-0.00001 (0.002)
south:urban		-0.117* (0.061)
Constant	5.214*** (0.168)	4.999*** (0.312)
-----		
Observations	935	935
R2	0.146	0.182
Adjusted R2	0.143	0.176
Residual Std. Error	0.390	0.382

F Statistic	52.872***	29.417***
=====		
Note:	*p<0.1; **p<0.05; ***p<0.01	

Для второй регрессии дайте интерпретацию коэффициентов.