

## ЗАДАНИЯ К ПРАКТИЧЕСКИМ ЗАНЯТИЯМ.

### Занятие 5.

#### Доходы и расходы семьи.

Общая рекомендация ко всем выполняемым вами заданиям:

- 1) Сохраняйте исходные файлы под новым именем, чтобы работать с ними.
- 2) **СРАЗУ открывайте и сохраняйте файл аутпута.** Первая команда в нем должна быть вида:  
**\*Фамилия – номер семинара – номер задания**
- 3) Сохраняйте проделанную вами работу в виде кода, используя «сохранение» правильных команд в STATA (или функцию “paste” SPSS). В этом случае вы сможете дома повторить все сделанное вами в классе. Кроме того, рекомендуется прикладывать программу к вашим исследованиям.
- 4) В качестве отчета за семинар нужно предъявить созданные файлы данных, файл аутпута, и файл с кодом.

#### Исходные файлы.

hh\_5\_16\_N.dta

Это результат нашего прошлого занятия.

#### 19. Корректировка и сравнение средних для душевого дохода.

19.1. Используем директорию на диске

**C:\RLMS\_work\seminar\_5\data**

Это позволит вам использовать готовые коды для переименований файлов 6-7-8 волн. Распакуйте архив данными.

Откройте программу STATA

Начните с открытия файла аутпута, назвав его своей фамилией.

```
log using "C:\RLMS_work\seminar_5\data\семинар 5 Рощина.smcl"
```

Первая команда должна быть такая (тем самым вы подписываете ваш аутпут)

**\*Фамилия - номер семинара**

Если у вас 14я STATA, набираем команду;

```
set more off
```

Желательно также делать комментарии с номером задания (начинающиеся со \*), так как этот файл – ваш главный отчет по работе за семинар.

\*Откройте файл данных

```
use "C:\RLMS_work\seminar_5\data\hh_5_16_N.dta", clear
```

И сохраните его под другим именем

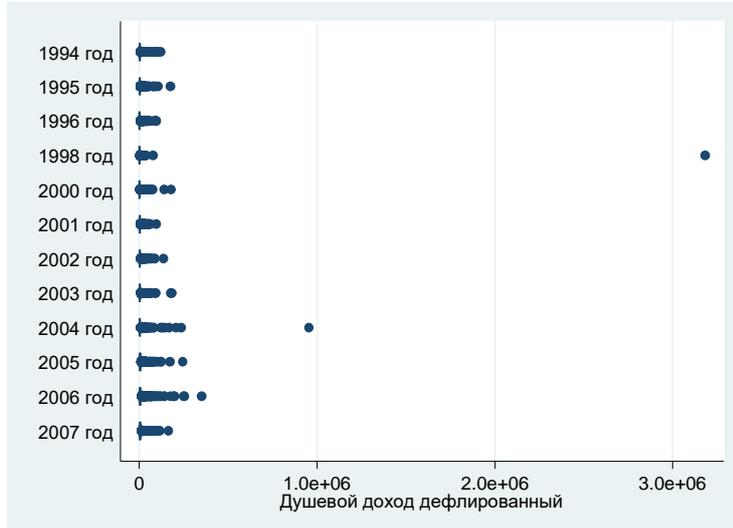
```
save "C:\RLMS_work\seminar_5\data\hh_5_16_S5.dta"
```

\*19.2. Корректировка выбросов в душевом доходе. Тест на равенство среднего душевого

дохода по годам.

\*Построим график распределений душевого дохода по годам, и максимумы

```
graph hbox INCOME_PC , over(id_w)  
tabstat INCOME_PC , statistics( max ) by(id_w) missing
```



Summary for variables: INCOME\_PC  
by categories of: id\_w (Волна (год проведения исследования))

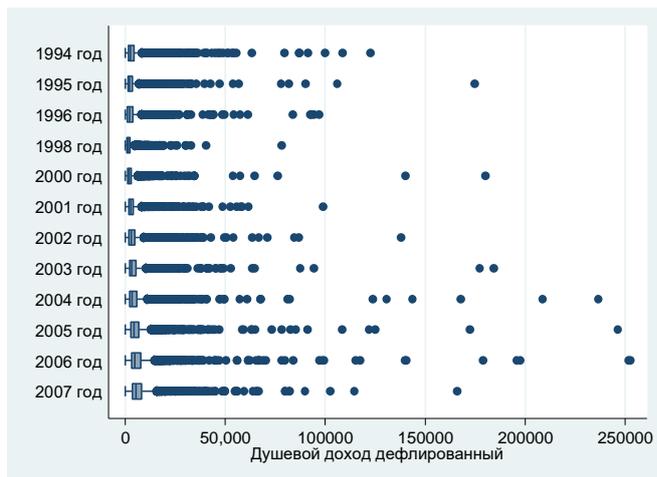
id_w	max
1994 год	122572.9
1995 год	174807.2
1996 год	96972.9
1998 год	3182252
2000 год	180183
2001 год	99010.95
2002 год	137882.9
2003 год	184407.7
2004 год	956280.8
2005 год	246381.4
2006 год	352186.6
2007 год	166000
Total	3182252

\*Удалим эти максимальные значения для 1998, 2004 и 2006 гг.

```
recode INCOME_PC (max = .) if (id_w == 8)  
recode INCOME_PC (max = .) if (id_w == 13)  
recode INCOME_PC (max = .) if (id_w == 15)
```

\*Посмотрим, что получилось

```
graph hbox INCOME_PC , over(id_w)  
tabstat INCOME_PC , statistics( max ) by(id_w) missing
```



Summary for variables: INCOME\_PC  
by categories of: id\_w (Волна (год проведения исследования))

id_w	max
1994 год	122572.9
1995 год	174807.2
1996 год	96972.9
1998 год	78329.78
2000 год	180183
2001 год	99010.95
2002 год	137882.9
2003 год	184407.7
2004 год	236715.2
2005 год	246381.4
2006 год	252586.3
2007 год	166000
Total	252586.3

\*Сохраните файл.

\*Проверим значимость различий среднего душевого дохода по годам (ANOVA).

Statistics > Linear models and related > ANOVA/MANOVA > One way ANOVA

## oneway INCOME\_PC id\_w, tabulate

Волна   (год   прове   дения   иссле   дован   ия)	Summary of Душевой доход дефлированный		
	Mean	Std. Dev.	Freq.
1994 год	4074.3336	6194.5578	3,861
1995 год	3287.9714	5533.7015	3,647
1996 год	3380.8561	5507.9936	3,594
1998 год	2065.7425	2797.7751	3,719
2000 год	2936.7576	5109.4713	3,887
2001 год	3764.1758	4420.3917	4,307
2002 год	4133.0931	5070.4125	4,512
2003 год	4611.9174	6073.6563	4,532
2004 год	5003.345	7865.6021	4,710
2005 год	5931.4216	7823.2995	4,370
2006 год	6771.7471	9688.3761	5,257
2007 год	7073.6661	7005.7264	5,094
Total	4580.523	6660.8527	51,490

Analysis of Variance					
Source	SS	df	MS	F	Prob > F
Between groups	1.1578e+11	11	1.0525e+10	249.84	0.0000
Within groups	2.1686e+12	51478	42127397.9		
Total	2.2844e+12	51489	44366958.2		

Bartlett's test for equal variances:  $\chi^2(11) = 8.5e+03$  Prob> $\chi^2 = 0.000$

F-статистика равна 249,84, а соответствующее значение p равно 0,0000. Так как значение p меньше, чем альфа = 0,05, мы можем отвергнуть нулевую гипотезу о том, что среднее изменение душевого дохода для каждого года одинаково. Другими словами, существует статистически значимая разница в среднем в среднем душевом доходе по крайней мере между двумя годами.

## 20. Получение и оказание материальной помощи другим домохозяйствам.

\*20.1. Посмотрим на вопросы о расходах на помощь другим семьям, которые в разных волнах задавали по-разному.

### codebook e18 e19\_1a e19\_1b e20\_1a e20\_1b e20\_1c e20\_1d

```
-----
e18      Приходилось ли Вашей семье или кому-л из ее членов в течение
последних 30 дней безвозм-но давать деньги или товары -продов-ие,
одежду, др предметы- людям, не из Вашего дх, - детям, родителям,
другим родственникам, просто чужим людям?
-----
```

```

type: numeric (double)
label: e18
range: [1,99999999]          units: 1
unique values: 5             missing .: 20,253/53,511
tabulation: Freq.  Numeric  Label
              8,825      1      Да
              24,344     2      Нет
              27         1.0e+08  ЗАТРУДНЯЮСЬ ОТВЕТИТЬ
              5          1.0e+08  ОТКАЗ ОТ ОТВЕТА
              57         1.0e+08  НЕТ ОТВЕТА
              20,253     .
-----
```

```
-----
e19_1a   Помогала Ваша семья деньгами или товарами в течение последних 30
дней родителям Вашим или супруга(и)?
-----
```

```

type: numeric (double)
label: e19_1a
range: [1,99999999] units: 1
unique values: 3 missing .: 44,686/53,511
tabulation: Freq. Numeric Label
2,018 1 Да
6,501 2 Нет
306 1.0e+08 НЕТ ОТВЕТА
44,686 .

```

-----  
e19\_1b **Сколько составила помощь Вашей семье деньгами или товарами  
в течение последних 30 дней родителям Вашим или супруга? (в рублях)**  
-----

```

type: numeric (double)
label: e19_1b, but 165 nonmissing values are not labeled
range: [10,99999999] units: .1
unique values: 168 missing .: 51,493/53,511
examples: .
.
.
.

```

-----  
e20\_1a **Ваша семья помогает деньгами или товарами родителям Вашим  
или супруга(и), не являющимся членами Вашего домохозяйства?**  
-----

```

type: numeric (double)
label: e20_1a
range: [1,99999999] units: 1
unique values: 3 missing .: 33,258/53,511
tabulation: Freq. Numeric Label
2,157 1 Да
18,073 2 Нет
23 1.0e+08 НЕТ ОТВЕТА
33,258 .

```

-----  
e20\_1b **Как регулярно** Ваша семья помогает деньгами или товарами  
родителям Вашим или супруга(и), не являющимся членами Вашего домохозяйства?  
-----

```

type: numeric (double)
label: e20_1b
range: [1,99999999] units: 1
unique values: 5 missing .: 51,354/53,511
tabulation: Freq. Numeric Label
923 1 Каждый месяц
712 2 Несколько раз в год
106 3 Раз в год
378 4 Когда попросят
38 1.0e+08 НЕТ ОТВЕТА
51,354 .

```

-----  
e20\_1c **В последние 30 дней** Ваша семья помогала деньгами или товарами  
родителям Вашим или супруга(и), не являющимся членами Вашего домохозяйства?  
-----

```

type: numeric (double)
label: e20_1c
range: [1,99999999] units: 1
unique values: 3 missing .: 51,354/53,511
tabulation: Freq. Numeric Label
1,672 1 Да
464 2 Нет
21 1.0e+08 НЕТ ОТВЕТА
51,354 .

```

-----  
e20\_1d **В течение последних 30 дней сколько составляла в рублях** помощь  
родителям Вашим или супруга(и), не являющимся членами Вашего домохозяйства?  
-----

```

type: numeric (double)
label: e20_1d, but 92 nonmissing values are not labeled
range: [15,99999999] units: 1
unique values: 94 missing .: 51,838/53,511
examples: .
.
.
.

```

### tab id\_w e18

Волна (год прове дения иссле дован ия)	Приходилось ли Вашей семье или кому-л из ее членов в течение после					Total
	Да	Нет	ЗАТРУДНЯЮ	ОТКАЗ ОТ	НЕТ ОТВЕТ	
1994 год	1,226	2,731	4	1	13	3,975
1995 год	873	2,898	3	1	8	3,783
1996 год	882	2,835	4	1	28	3,750
1998 год	975	2,849	4	0	2	3,830
2000 год	1,066	2,936	0	1	3	4,006
2001 год	1,271	3,249	6	1	1	4,528
2002 год	1,250	3,412	5	0	1	4,668
2003 год	1,282	3,434	1	0	1	4,718
Total	8,825	24,344	27	5	57	33,258

### tab id\_w e19\_1a

Волна (год прове дения иссле дован ия)	Помогала Ваша семья деньгами или товарами в течение последних 30 д			Total
	Да	Нет	НЕТ ОТВЕТ	
1994 год	291	780	155	1,226
1995 год	183	612	78	873
1996 год	206	613	63	882
1998 год	240	728	7	975
2000 год	237	829	0	1,066
2001 год	291	980	0	1,271
2002 год	286	964	0	1,250
2003 год	284	995	3	1,282
Total	2,018	6,501	306	8,825

### tab id\_w e20\_1a

Волна (год прове дения иссле дован ия)	Ваша семья помогает деньгами или товарами родителям Вашим или суп			Total
	Да	Нет	НЕТ ОТВЕТ	
2004 год	469	4,239	3	4,711
2005 год	408	4,148	16	4,572
2006 год	685	4,858	2	5,545
2007 год	595	4,828	2	5,425
Total	2,157	18,073	23	20,253

Таким образом, в 1994-2003 гг. семьям сначала задавали общий вопрос, помогали ли они кому-то, а затем по каждой категории помощи другим д\х помощи задавали два вопроса: 1) помогали ли за 30 дней; 2) сумма помощи (переменные e19\_1a e19\_1b для помощи родителям, всего 5 разных вариантов, кому семья помогала). При этом если переменная e18 принимает значение 2 («не помогали»), то остальные переменные имеют пропущенные значения.

\*20.2. Сначала восстановим значение «2» (не помогали) для каждой категории помощи для тех д\х, которые вообще не оказывали помощь (e18 принимает значение 2 («не помогали»)).

**recode e19\_1a e19\_2a e19\_3a e19\_4a e19\_5a (. =2) if (e18 ==2 & id\_w <= 12)**

Начиная с 2004 года (13я волна), общий вопрос о помощи уже не задавали, а по каждой помощи другим д\х стали задавать 4 вопроса: 1) помогали ли; 2) как часто помогали; 3) помогали ли за 30 дней; 4) сумма помощи (переменные для помощи родителям: **e20\_1a e20\_1b e20\_1c e20\_1d**). При этом если переменная **e20\_1a** принимает значение 2 («не помогали»), то остальные переменные имеют пропущенные значения. Если мы хотим сравнить, как семьи помогали другим д\х, за весь период, нам нужно привести в соответствие эти переменные за 1994-2003 и за 2004-2007 гг. Проще всего «восстановить» значения переменных **e19\_1a e19\_1b** за 2004-2007 гг.

\*Посмотрите в таблице данных для последней волны (16), как выглядят значения, и что мы хотим получить в результате.

\*С суммами сделать замену для этих лет очень просто, заменив миссинги в переменных **e19\_1b e19\_2b e19\_3b e19\_4b e19\_5b** на значения переменных **e20\_1b e20\_2b e20\_3b e20\_4b e20\_5b** если номер волны больше или равен 13 :

```
replace e19_1b = e20_1d if id_w >= 13
replace e19_2b = e20_2d if id_w >= 13
replace e19_3b = e20_3d if id_w >= 13
replace e19_4b = e20_4d if id_w >= 13
replace e19_5b = e20_5d if id_w >= 13
```

Посмотрите на таблице данных 16й волны: появились значения переменных **e19\_1b e19\_2b e19\_3b e19\_4b e19\_5b**.

Теперь аналогично произведем замены для **e19\_1a e19\_2a e19\_3a e19\_4a e19\_5a**: сначала заменим из значения на значения из переменных **e20\_1c e20\_2c e20\_3c e20\_4c e20\_5c** (это вопросы о наличии помощи за 30 дней) для волн 13-16. Затем если переменная **e20\_1a** равна 2 (то есть д\х вообще не помогали этой категории других д\х), приравняем **e19\_1a** также 2, тем самым в этой переменной значение 2 будет соответствовать и отсутствию помощи вообще, и отсутствию помощи за 30 дней (для волн 13-16). Аналогичные преобразования сделаем для других четырех категорий д\х, которым можно помогать.

```
replace e19_1a = e20_1c if id_w >= 13
replace e19_1a = 2 if id_w >= 13 & e20_1a ==2
replace e19_2a = e20_2c if id_w >= 13
replace e19_2a = 2 if id_w >= 13 & e20_2a ==2
replace e19_3a = e20_3c if id_w >= 13
replace e19_3a = 2 if id_w >= 13 & e20_3a ==2
replace e19_4a = e20_4c if id_w >= 13
replace e19_4a = 2 if id_w >= 13 & e20_4a ==2
replace e19_5a = e20_5c if id_w >= 13
replace e19_5a = 2 if id_w >= 13 & e20_5a ==2
```

\*Обозначим как миссинги значения 99999997- 99999999.

```
recode e18 e19_1a e19_1b e19_2a e19_2b e19_3a e19_3b e19_4a e19_4b e19_5a e19_5b e20_1a
e20_1b e20_1c e20_1d e20_2a e20_2b e20_2c e20_2d e20_3a e20_3b e20_3c e20_3d e20_4a
e20_4b e20_4c e20_4d e20_5a e20_5b e20_5c e20_5d (99999997 = .a) (99999998 = .b) (99999999
= .c)
```

\*Перекодируем значение «2» в значение «0» («нет») для всех используемых нами переменных.

```
recode e18 e19_1a e19_2a e19_3a e19_4a e19_5a e20_1a e20_1c e20_2a e20_2c e20_3a e20_3c
e20_4a e20_4c e20_5a e20_5c (2=0)
```

\*Сделаем лейбл для сета значений 0- нет, 1 – да.

```
label define YES 0 "нет" 1 "да" .a "затр.ответить" .b "отказ" .c "нет ответа"
```

\* И присвоим этот лейбл всему сету переменных:

```
label values e18 e19_1a e19_2a e19_3a e19_4a e19_5a e20_1a e20_1c e20_2a e20_2c e20_3a
e20_3c e20_4a e20_4c e20_5a e20_5c YES
```

Так как переменная **e18** не определена для волн 13-16, сделаем вспомогательную переменную **e18\_sum** как сумму всех видов помощи (на случай пропуска в одной из них используем команду суммирования, в которой миссинг принимается за 0) для этих волн. Все значения этой переменной больше или равные 1 перекодируем в 1; а затем прирачаем значение нужной нам переменной к этой вспомогательной переменной для волн 13-16, и удалим вспомогательную переменную.

```
egen e18_sum= rowtotal(e19_1a e19_2a e19_3a e19_4a e19_5a) if id_w >= 13, missing
recode e18_sum (1 / 5 =1)
replace e18 = e18_sum if id_w >= 13
drop e18_sum
```

\*Посмотрите распределения полученных переменных с учетом пропущенных значений

```
tab1 e18 e19_1a e19_2a e19_3a e19_4a e19_5a e20_1a e20_1c e20_2a e20_2c e20_3a e20_3c
e20_4a e20_4c e20_5a e20_5c, missing
```

\*20.3. Посмотрим средние значения этих дихотомических переменных по годам (они покажут нам долю тех, кто оказывал помощь таким домохозяйствам, так как значения переменных 0 или 1):

```
tabstat e18 e19_1a e19_2a e19_3a e19_4a e19_5a [aweight = hhwt], statistics( mean ) by(id_w)
missing
```

```
Summary statistics: mean
by categories of: id_w (Волна (год проведения исследования))
```

id_w	e18	e19_1a	e19_2a	e19_3a	e19_4a	e19_5a
1994 год	.3112778	.0761971	.1269367	.0096752	.0578852	.1305662
1995 год	.2338628	.0498361	.1008608	.0071707	.03961	.0875247
1996 год	.2374461	.0559082	.1009014	.0039453	.0372547	.0910489
1998 год	.2545294	.0587812	.1191382	.0060523	.0524677	.0832306
2000 год	.2742362	.0581446	.12533	.0067289	.0637033	.0994081
2001 год	.2886018	.059871	.1252101	.0080231	.0671097	.1166066
2002 год	.2723304	.0565922	.1187345	.0072785	.0593705	.1076344
2003 год	.2783162	.0601643	.1221422	.0059447	.0771325	.1004305
2004 год	.3304595	.078729	.1684425	.011252	.1077987	.0686508
2005 год	.307242	.0658416	.1668687	.0055374	.1020715	.0506127
2006 год	.3329638	.0935173	.1561879	.0133411	.1076025	.0618781
2007 год	.2971569	.0752666	.1430935	.0098562	.1023378	.0524705
Total	.2854135	.0662739	.1310994	.008032	.0731246	.0874291

\*Посмотрим, значимы ли различия между долями семей, помогавшим другим домохозяйствам, по годам.

**oneway e18 id\_w, tabulate**

Волна   (год   прове   дения   иссле   дован   ия)	Summary of Приходилось ли Вашей семье или кому-л из ее членов в течение после		
	Mean	Std. Dev.	Freq.
1994 год	.30983068	.46248209	3,957
1995 год	.23150358	.42184934	3,771
1996 год	.23728814	.42547759	3,717
1998 год	.25496862	.43590057	3,824
2000 год	.26636682	.44211353	4,002
2001 год	.28119469	.44963203	4,520
2002 год	.26812527	.44303071	4,662
2003 год	.27184054	.44495533	4,716
2004 год	.31641537	.46512646	4,709
2005 год	.3005045	.45852771	4,559
2006 год	.32371506	.46793494	5,545
2007 год	.29208925	.45476506	5,423
Total	.28207097	.45001192	53,405

Analysis of Variance					
Source	SS	df	MS	F	Prob > F
Between groups	42.6101382	11	3.87364893	19.20	0.0000
Within groups	10772.2728	53393	.201754403		
Total	10814.883	53404	.202510729		

Bartlett's test for equal variances:  $\chi^2(11) = 104.1773$  Prob> $\chi^2 = 0.000$

Так как Prob > F = 0.000, по крайней мере в какие-то два года средние значения различаются.

\*20.4. Посмотрим теперь на переменную **f8** (Получали члены Вашей семьи в течение последних 30 дней безвозмездно как-либо деньги или товары от людей, не являющихся членами Вашего домохозяйства: от детей, родителей, друзей, посторонних лиц или организаций?).

\*Сначала перекодируем в миссинги, а также значение «2» в «0» («нет»):

**recode f8 (99999997 = .a) (99999998 = .b) (99999999 = .c)**

**recode f8 (2=0)**

**label values f8 YES**

\*Сделаем переменную «донор или рецептор» (получало домохозяйство помощь или помогало кому-то) осмотрим распределение:

**gen donor = f8\*10 + e18**

**tab donor**

donor	Freq.	Percent	Cum.
0	27,489	51.53	51.53
1	10,857	20.35	71.88
10	10,807	20.26	92.14
11	4,195	7.86	100.00

Total	53,348	100.00
-------	--------	--------

Так как значение 0 соответствует неучастию в помощи, 1 – только получению помощи, 10 – только оказанию помощи, а 11 – и получению, и оказанию, сделаем соответствующие лейблы.

```
label define donor_label 0 "не участвует" 1 "акцептор" 10 "донор" 11 "донор и акцептор"
label variable donor "донор или акцептор помощи"
label values donor donor_label
```

\*Посмотрим распределение по годам

```
tabulate id_w donor, nofreq row
```

Волна (год прове дения иссле дован ия)	донор или акцептор помощи				Total
	не участв	акцептор	донор	донор и а	
1994 год	54.28	22.12	14.72	8.87	100.00
1995 год	60.89	17.59	15.91	5.61	100.00
1996 год	58.09	17.03	18.11	6.77	100.00
1998 год	55.70	17.83	18.80	7.67	100.00
2000 год	52.41	18.88	20.98	7.73	100.00
2001 год	52.07	19.50	19.81	8.61	100.00
2002 год	53.10	18.58	20.08	8.24	100.00
2003 год	50.99	18.55	21.86	8.60	100.00
2004 год	46.60	23.16	21.76	8.48	100.00
2005 год	46.61	22.09	23.34	7.97	100.00
2006 год	44.83	23.89	22.79	8.50	100.00
2007 год	48.87	22.39	21.91	6.83	100.00
Total	51.53	20.35	20.26	7.86	100.00

\*Посмотрим, различается ли значимо душевой доход в домохозяйствах, по-разному участвующих во взаимопомощи. Можем не учитывать год, так как душевой доход дефлирован.

```
oneway INCOME_PC donor, tabulate
```

донор или акцеп тор помощ и	Summary of Душевой доход дефлированный		
	Mean	Std. Dev.	Freq.
не участв	3974.8758	5657.0667	26,386
акцептор	5938.7541	8649.4876	10,458
донор	4301.8623	6132.6243	10,454
донор и а	5760.9261	7473.3113	4,062
Total	4582.5762	6665.8872	51,360

Source	Analysis of Variance			F	Prob > F
	SS	df	MS		
Between groups	3.5443e+10	3	1.1814e+10	270.06	0.0000
Within groups	2.2466e+12	51356	43746509.1		
Total	2.2821e+12	51359	44434052.7		

Bartlett's test for equal variances: chi2(3) = 3.2e+03 Prob>chi2 = 0.000

\*Сохраните файл.

Save "C:\RLMS\_work\seminar\_5\data\hh\_5\_16\_S5.dta", replace

\*20.5. Самостоятельное задание (в классе или дома).

1) Если вы делаете задание как домашнее, откройте лог-файл (аутпут) и наберите в нем:

\* **Фамилия задание 20.3.**

В качестве отчета за выполненное задание, вы сдаете лог-файл + его копия в формате word, чтобы было возможно сделать комментарии.

Если вы делаете задание в классе или дистанционно, можно его выполнить в вашем общем лог-файле (но укажите номер задания).

2) Аналогично преобразованиям, сделанным для блока переменных о получении помощи (e19\_1a e19\_2a e19\_3a e19\_4a e19\_5a) , сделайте преобразования для блока переменных оказания помощи: f9\_1a f9\_1b f9\_2a f9\_2b f9\_3a f9\_3b f9\_4a f9\_4b f9\_5a f9\_5b f9\_6a f9\_6b f9\_7a f9\_7b f9\_8a f9\_8b f9\_9a f9\_9b f9\_91a f9\_91b f9\_10a f9\_10b

То есть перекодируйте миссинги для всех переменных. Для дихотомических переменных f9\_1a f9\_2a f9\_3a f9\_4a f9\_5a f9\_6a f9\_7a f9\_8a f9\_9a f9\_91a f9\_10a – сначала перекодируйте пропущенное значение «.» в «2» если переменная f8 равна 0 (то есть если домохозяйство не получало помощи: мы ранее уже перекодировали для нее значение 2 в 0). Затем для этого сета переменных перекодируйте «2» в «0» и присвойте уже созданный лейбл **YES**. Посмотрите распределения этих переменных.

3) Рассчитайте средние значения полученных переменных по годам для взвешенных данных при помощи команды **tabstat**.

21. Структура расходов домохозяйства.

21.1. \* Сначала перекодируем миссинги (учитывая, что для переменных e18 e19\_1a e19\_1b e19\_2a e19\_2b e19\_3a e19\_3b e19\_4a e19\_4b e19\_5a e19\_5b мы это уже сделали)

```
recode d5 d6 d17 d18 e2 e3 e4 e5 e6 e6_1 e6_2 e7_1a e7_1b e7_1_0a e7_1_0b e7_1_1a e7_1_1b e7_2a e7_2b e7_3a e7_3b e7_4a e7_4b e7_5a e7_5b e7_6a e7_6b e7_7a e7_7b e7_8a e7_8b e7_9a e7_9b e7_10a e7_10b e8_1a e8_1b e8_2a e8_2b e8_3a e8_3b e9_1a e9_1b e9_2a e9_2b e9_3a e9_3b e9_4a e9_4b e9_4_1a e9_4_1b e9_5a e9_5b e9_6a e9_6b e9_7a e9_7b e9_8a e9_8b e9_9a e9_9b e9_10a e9_10b e10 e11 e12 e12_1 e12_2 e12_3 e13_1a e13_1b e13_2a e13_2b e13_21a e13_21b e13_3a e13_3b e13_31a e13_31b e13_22a e13_22b e13_23a e13_23b e13_24a e13_24b e13_32a e13_32b e13_33a e13_33b e13_34a e13_34b e13_4a e13_4b e13_5a e13_5b e13_6a e13_6b e13_7a e13_7b e13_8a e13_8b e13_9a e13_9b e13_71a e13_71b e13_72a e13_72b e13_10a e13_10b e14 e15 e16 e17 (99999997 = .a) (99999998 = .b) (99999999 = .c)
```

```
describe d5 d6 d17 d18 e2 e3 e4 e5 e6 e6_1 e6_2 e7_1a e7_1b e7_1_0a e7_1_0b e7_1_1a e7_1_1b e7_2a e7_2b e7_3a e7_3b e7_4a e7_4b e7_5a e7_5b e7_6a e7_6b e7_7a e7_7b e7_8a e7_8b e7_9a e7_9b e7_10a e7_10b e8_1a e8_1b e8_2a e8_2b e8_3a e8_3b e9_1a e9_1b e9_2a e9_2b e9_3a e9_3b e9_4a e9_4b e9_4_1a e9_4_1b e9_5a e9_5b e9_6a e9_6b e9_7a e9_7b e9_8a e9_8b e9_9a e9_9b e9_10a e9_10b e10 e11 e12 e12_1 e12_2 e12_3 e13_1a e13_1b e13_2a e13_2b e13_21a e13_21b e13_3a e13_3b e13_31a e13_31b e13_22a e13_22b e13_23a e13_23b e13_24a e13_24b e13_32a e13_32b e13_33a e13_33b e13_34a e13_34b e13_4a e13_4b e13_5a e13_5b e13_6a e13_6b e13_7a e13_7b e13_8a e13_8b e13_9a e13_9b e13_71a e13_71b e13_72a e13_72b e13_10a e13_10b e14 e15 e16 e17 e18 e19_1a e19_1b e19_2a e19_2b e19_3a e19_3b e19_4a e19_4b e19_5a e19_5b
```

variable name	storage type	display format	value label	variable label
d5	double	%12.0g	d5	В последние 12 месяцев Ваша семья платила за пользование землей?

d6	double	%12.0g	d6	Сколько всего рублей Вы заплатили за пользование Вашей землей в последние 12 месяцев?
d17	double	%12.0g	d17	Вы потратили в течение последних 30 дней какие-то деньги на семена,
d18	double	%12.0g	d18	Сколько денег Ваша семья потратила в течение последних 30 дней на семена, удобрения, корма, наем техники, рабочей силы, на что-нибудь другое, необходимое для ведения хозяйства, за исключением расходов на топливо?
e2	double	%12.0g	e2	В течение последних 7 дней Вы или другие члены Вашей семьи питались вне дома?
e3	double	%12.0g	e3	Вспомните, сколько примерно денег все члены Вашей семьи израсходовали на питание вне дома в течение последних 7 дней?
e4	double	%12.0g	e4	Вспомните, сколько примерно денег все члены Вашей семьи израсходовали на питание дома и вне дома в течение последних 30 дней?
e5	double	%12.0g	e5	Покупали Вы или другие члены Вашей семьи в течение последних 3-х месяцев что-либо из одежды или обуви для членов домохозяйства?
e6	double	%12.0g	e6	Сколько Вы и другие члены домохозяйства заплатили за это?
e6_1	double	%12.0g	e6_1	Сколько Вы и другие члены домохозяйства потратили на приобретение одежды и обуви для взрослых, то есть для членов домохозяйства 18-ти лет и старше, в течение последних 3 месяцев?
e6_2	double	%12.0g	e6_2	Сколько Вы и другие члены домохозяйства потратили на приобретение одежды и обуви для детей, то есть для членов домохозяйства младше 18-ти лет, в течение последних 3 месяцев?
e7_1a	double	%12.0g	e7_1a	Ваша семья покупала в течение последних 3-х месяцев культтовары
e7_1b	double	%12.0g	e7_1b	Сколько всего рублей Вы заплатили за культтовары: телевизор, магнитофон...
e7_1_0a	double	%12.0g	e7_1_0a	Ваша семья покупала в течение последних 3-х месяцев культтовары:
e7_1_0b	double	%12.0g	e7_1_0b	Сколько всего рублей Ваша семья заплатила за культтовары: телевизор...
e7_1_1a	double	%12.0g	e7_1_1a	Ваша семья покупала в течение последних 3-х месяцев мобильный телефон
e7_1_1b	double	%12.0g	e7_1_1b	Сколько всего рублей Ваша семья заплатила за мобильный телефон?
e7_2a	double	%12.0g	e7_2a	Ваша семья покупала в течение последних 3-х месяцев предметы домашней обстановки: мебель, ковры и другое?
e7_2b	double	%12.0g	e7_2b	Сколько всего рублей Ваша семья заплатила за предметы домашней обстановки: мебель, ковры и другое?
e7_3a	double	%12.0g	e7_3a	Ваша семья покупала в течение последних 3-х месяцев бытовую технику: холодильник, стиральную машину, пылесос, швейную машину, утюг, кухонный комбайн и тому подобное?
e7_3b	double	%12.0g	e7_3b	Сколько всего рублей Ваша семья заплатила за бытовую технику
e7_4a	double	%12.0g	e7_4a	Ваша семья покупала в течение последних 3-х месяцев автомобиль или микроавтобус?
e7_4b	double	%12.0g	e7_4b	Сколько всего рублей Ваша семья заплатила за автомобиль или микроавтобус?
e7_5a	double	%12.0g	e7_5a	Ваша семья покупала в течение последних 3-х месяцев мотоцикл, мотороллер?
e7_5b	double	%12.0g	e7_5b	Сколько всего рублей Ваша семья заплатила за мотоцикл, мотороллер?
e7_6a	double	%12.0g	e7_6a	Ваша семья покупала в течение последних 3-х месяцев гараж?
e7_6b	double	%12.0g	e7_6b	Сколько всего рублей Ваша семья заплатила за гараж?

e7_7a	double	%12.0g	e7_7a	Ваша семья покупала в течение последних <b>3-х месяцев</b> строительные материалы, материалы для ремонта?
e7_7b	double	%12.0g	e7_7b	Сколько всего рублей Ваша семья заплатила за строительные материалы, материалы для ремонта?
e7_8a	double	%12.0g	e7_8a	Ваша семья покупала в течение последних <b>3-х месяцев</b> дачу, дом, квартиру, землю?
e7_8b	double	%12.0g	e7_8b	Сколько всего рублей Ваша семья заплатила за дом, дачу, квартиру, землю?
e7_9a	double	%12.0g	e7_9a	Ваша семья покупала в течение последних <b>3-х месяцев</b> книги, учебники, учебные пособия, канцтовары?
e7_9b	double	%12.0g	e7_9b	Сколько всего рублей Ваша семья заплатила за книги, учебники, учебные пособия, канцтовары?
e7_10a	double	%12.0g	e7_10a	Ваша семья покупала в течение последних <b>3-х месяцев</b> спортивный инвентарь: велосипед, самокат, коньки?
e7_10b	double	%12.0g	e7_10b	Сколько всего рублей Ваша семья заплатила за спортивный инвентарь: велосипед, самокат, коньки?
e8_1a	double	%12.0g	e8_1a	Ваша семья покупала в течение последних <b>30 дней</b> топливо для заправки автотранспорта, двигателей, генераторов?
e8_1b	double	%12.0g	e8_1b	Сколько всего рублей Ваша семья заплатила за топливо для заправки автотранспорта, двигателей, генераторов?
e8_2a	double	%12.0g	e8_2a	Ваша семья покупала в течение последних <b>30 дней</b> дрова, уголь, торф, керосин?
e8_2b	double	%12.0g	e8_2b	Сколько всего рублей Ваша семья заплатила за дрова, уголь, торф, керосин?
e8_3a	double	%12.0g	e8_3a	Ваша семья покупала в течение последних <b>30 дней</b> баллонный газ?
e8_3b	double	%12.0g	e8_3b	Сколько всего рублей Ваша семья заплатила за баллонный газ?
e9_1a	double	%12.0g	e9_1a	В течение последних <b>30 дней</b> Ваша семья пользовалась транспортом: местного, междугородного?
e9_1b	double	%12.0g	e9_1b	Сколько всего рублей Ваша семья заплатила за услуги транспорта: местного, междугородного?
e9_2a	double	%12.0g	e9_2a	В течение последних <b>30 дней</b> Ваша семья пользовалась услугами по пошиву и ремонту одежды, обуви?
e9_2b	double	%12.0g	e9_2b	Сколько всего рублей Ваша семья заплатила за услуги по пошиву и ремонту одежды, обуви?
e9_3a	double	%12.0g	e9_3a	В течение последних <b>30 дней</b> Ваша семья пользовалась услугами мастерских по ремонту культуртоваров, например телевизора, магнитофона или мебели, бытовой техники, включая покупку запчастей?
e9_3b	double	%12.0g	e9_3b	Сколько всего рублей Ваша семья заплатила за услуги услугами мастерских по ремонту культуртоваров, например телевизора, магнитофона или мебели, бытовой техники, включая покупку запчастей?
e9_4a	double	%12.0g	e9_4a	В течение последних <b>30 дней</b> Ваша семья пользовалась услугами по ремонту и строительству жилья, построек?
e9_4b	double	%12.0g	e9_4b	Сколько всего рублей Ваша семья заплатила за услуги по ремонту и строительству жилья, построек?
e9_4_1a	double	%12.0g	e9_4_1a	В течение последних <b>30 дней</b> Ваша семья пользовалась услугами по ремонту автомобиля, мотоцикла, включая покупку запчастей?
e9_4_1b	double	%12.0g	e9_4_1b	Сколько всего рублей Ваша семья заплатила за услуги по ремонту автомобиля, мотоцикла, включая покупку запчастей?
e9_5a	double	%12.0g	e9_5a	В течение последних <b>30 дней</b> Ваша семья пользовалась прачечной, химчисткой,

				баней, парикмахерской?
e9_5b	double	%12.0g	e9_5b	Сколько всего рублей Ваша семья заплатила за услуги прачечной, химчистки, бани, парикмахерской?
e9_6a	double	%12.0g	e9_6a	В течение последних 30 дней Ваша семья пользовалась почтово-телеграфными услугами, включая междугородные телефонные переговоры (БЕЗ УЧЕТА МОБИЛЬНОЙ СОТОВОЙ СВЯЗИ)?
e9_6b	double	%12.0g	e9_6b	Сколько всего рублей Ваша семья заплатила за почтово-телеграфные услуги?
e9_7a	double	%12.0g	e9_7a	В течение последних 30 дней Ваша семья пользовалась ритуальными услугами: загс, похоронное бюро?
e9_7b	double	%12.0g	e9_7b	Сколько всего рублей Ваша семья заплатила за ритуальные услуги?
e9_8a	double	%12.0g	e9_8a	В течение последних 30 дней Ваша семья пользовалась услугами мобильной сотовой связи?
e9_8b	double	%12.0g	e9_8b	Сколько всего рублей Ваша семья заплатила за услуги мобильной сотовой связи?
e9_9a	double	%12.0g	e9_9a	В течение последних 30 дней Ваша семья пользовалась услугами Интернета?
e9_9b	double	%12.0g	e9_9b	Сколько всего рублей Ваша семья заплатила за услуги Интернета?
e9_10a	double	%12.0g	e9_10a	В течение последних 30 дней Ваша семья пользовалась услугами юристов, нотариусов, риэлторов?
e9_10b	double	%12.0g	e9_10b	Сколько всего рублей Ваша семья заплатила за услуги юристов, нотариусов, риэлторов?
e10	double	%12.0g	e10	Ваша семья в течение последних 30 дней платила за жилье, то есть за квартиру, в том числе и аренду, если Вы снимаете жилье, и коммунальные услуги?
e11	double	%12.0g	e11	Сколько рублей Ваша семья фактически заплатила в течение последних 30 дней за квартиру, включая аренду, и коммунальные услуги?
e12	double	%12.0g	e12	Ваша семья имеет неоплаченные счета за жилье, то есть за квартиру и коммунальные услуги?
e12_1	double	%12.0g	e12_1	Назовите сумму задолженности за квартиру и коммунальные услуги
e12_2	double	%12.0g	e12_2	Если вспомнить последние 3 месяца, сколько рублей в среднем в месяц должна была платить Ваша семья за квартиру, включая аренду, и коммунальные услуги, за вычетом субсидий и льгот, если они у Вас есть?
e12_3	double	%12.0g	e12_3	Сколько киловатт электроэнергии Ваша семья потребила в течение сентября текущего года?
e13_1a	double	%12.0g	e13_1a	Тратила ли Ваша семья в течение последних 30 дней деньги на содержание и оплату занятий детей в дошкольных учреждениях, школах, секциях, кружках, на оплату частных уроков, репетиторов, в том числе и на подарки педагогам?
e13_1b	double	%12.0g	e13_1b	Сколько всего рублей Ваша семья истратила на содержание и оплату занятий детей в дошкольных учреждениях, школах, секциях, кружках, на оплату частных уроков, репетиторов, в том числе и на подарки педагогам?
e13_2a	double	%12.0g	e13_2a	Тратила ли Ваша семья в течение последних 30 дней деньги на путевки в санатории, дома отдыха, детские лагеря, на туристические поездки и другое, исключая расходы на транспорт?
e13_2b	double	%12.0g	e13_2b	Сколько всего рублей Ваша семья истратила на путевки в санатории, дома отдыха, детские лагеря, на туристические поездки и другое, исключая расходы на транспорт?
e13_21a	double	%12.0g	e13_21a	Тратила ли Ваша семья в течение последних 30 дней деньги на билеты: в театр, цирк, кино, на концерты, в парки культуры и другие развлекательные мероприятия?
e13_21b	double	%12.0g	e13_21b	Сколько всего рублей Ваша семья истратила на билеты: в театр, цирк, кино, на концерты, в парки культуры и другие развлекательные мероприятия?

e13_3a	double	%12.0g	e13_3a	Тратила ли Ваша семья в течение последних 30 дней деньги на лечение, обследование, исключая покупку лекарств?
e13_3b	double	%12.0g	e13_3b	Сколько всего рублей Ваша семья истратила на лечение, обследование, исключая покупку лекарств?
e13_31a	double	%12.0g	e13_31a	Тратила ли Ваша семья в течение последних 30 дней деньги на лекарства, включая витамины и другие медикаменты?
e13_31b	double	%12.0g	e13_31b	Сколько всего рублей Ваша семья истратила на лекарства, включая витамины и другие медикаменты?
e13_22a	double	%12.0g	e13_22a	Тратила ли Ваша семья в течение последних 30 дней деньги на лечение, обследование в стационарных лечебных учреждениях: больницах, госпиталях, клиниках, исключая покупку лекарств?
e13_22b	double	%12.0g	e13_22b	Сколько всего рублей Ваша семья истратила на лечение, обследование в стационарных лечебных учреждениях: больницах, госпиталях, клиниках, исключая покупку лекарств?
e13_23a	double	%12.0g	e13_23a	Тратила ли Ваша семья в течение последних 30 дней деньги на лечение, обследование в поликлиниках, амбулаториях, исключая покупку лекарств?
e13_23b	double	%12.0g	e13_23b	Сколько всего рублей Ваша семья истратила на лечение, обследование в поликлиниках, амбулаториях, исключая покупку лекарств?
e13_24a	double	%12.0g	e13_24a	Тратила ли Ваша семья в течение последних 30 дней деньги на лечение, протезирование зубов, исключая покупку лекарств?
e13_24b	double	%12.0g	e13_24b	Сколько всего рублей Ваша семья истратила на лечение, протезирование зубов, исключая покупку лекарств?
e13_32a	double	%12.0g	e13_32a	Тратила ли Ваша семья в течение последних 30 дней деньги на моющие средства, например хозяйственное мыло, стиральный порошок и другие?
e13_32b	double	%12.0g	e13_32b	Сколько всего рублей Ваша семья истратила на моющие средства, например хозяйственное мыло, стиральный порошок и другие?
e13_33a	double	%12.0g	e13_33a	Тратила ли Ваша семья в течение последних 30 дней деньги средства личной гигиены, например, туалетное мыло, шампунь, зубная паста, туалетная бумага, салфетки, памперсы, прокладки?
e13_33b	double	%12.0g	e13_33b	Сколько всего рублей Ваша семья истратила средства личной гигиены, например, туалетное мыло, шампунь, зубная паста, туалетная бумага, салфетки, памперсы, прокладки?
e13_34a	double	%12.0g	e13_34a	Тратила ли Ваша семья в течение последних 30 дней деньги на косметику и парфюмерию?
e13_34b	double	%12.0g	e13_34b	Сколько всего рублей Ваша семья истратила на косметику и парфюмерию?
e13_4a	double	%12.0g	e13_4a	Тратила ли Ваша семья в течение последних 30 дней деньги на оплату занятий взрослых членов семьи в различных секциях, на курсах, у преподавателей?
e13_4b	double	%12.0g	e13_4b	Сколько всего рублей Ваша семья истратила на оплату занятий взрослых членов семьи в различных секциях, на курсах, у преподавателей?
e13_5a	double	%12.0g	e13_5a	Тратила ли Ваша семья в течение последних 30 дней деньги на приобретение акций, облигаций, других ценных бумаг?
e13_5b	double	%12.0g	e13_5b	Сколько всего рублей Ваша семья истратила на приобретение акций, облигаций, других ценных бумаг?
e13_6a	double	%12.0g	e13_6a	Тратила ли Ваша семья в течение последних 30 дней деньги на уплату любых страховых взносов: за страхование жизни, здоровья, машины, жилища и другого?
e13_6b	double	%12.0g	e13_6b	Сколько всего рублей Ваша семья истратила на уплату любых страховых взносов: за страхование жизни, здоровья, машины, жилища и другого?
e13_7a	double	%12.0g	e13_7a	Тратила ли Ваша семья в течение последних 30 дней деньги на погашение кредита, возврат ссуд, уплату долгов?

e13_7b	double	%12.0g	e13_7b	Сколько всего рублей Ваша семья истратила на погашение кредита, возврат ссуд, уплату долгов?
e13_8a	double	%12.0g	e13_8a	Тратила ли Ваша семья в течение последних 30 дней деньги на уплату алиментов?
e13_8b	double	%12.0g	e13_8b	Сколько всего рублей Ваша семья истратила на уплату алиментов?
e13_9a	double	%12.0g	e13_9a	Тратила ли Ваша семья в течение последних 30 дней деньги налогов на недвижимость и транспорт, оплату лицензии, пошлин за оформление документов, кроме налогов на доходы и землю?
e13_9b	double	%12.0g	e13_9b	Сколько всего рублей Ваша семья истратила налогов на недвижимость и транспорт, оплату лицензии, пошлин за оформление документов, кроме налогов на доходы и землю?
e13_71a	double	%12.0g	e13_71a	Тратила ли Ваша семья в течение последних 30 дней деньги на уплату долгов частным лицам?
e13_71b	double	%12.0g	e13_71b	Сколько всего рублей Ваша семья истратила на уплату долгов частным лицам?
e13_72a	double	%12.0g	e13_72a	Тратила ли Ваша семья в течение последних 30 дней деньги на погашение кредита, возврат ссуд?
e13_72b	double	%12.0g	e13_72b	Сколько всего рублей Ваша семья истратила на погашение кредита, возврат ссуд?
e13_10a	double	%12.0g	e13_10a	Тратила ли Ваша семья в течение последних 30 дней деньги на покупку валюты с целью сбережений?
e13_10b	double	%12.0g	e13_10b	Сколько всего рублей Ваша семья истратила на покупку валюты с целью сбережений?
e14	double	%12.0g	e14	Ваша семья в течение последних 30 дней давала деньги в долг?
e15	double	%12.0g	e15	Сколько рублей Ваша семья отдала в долг? долг в течение последних 30 дней
e16	double	%12.0g	e16	Ваша семья в течение последних 30 дней откладывала сбережения?
e17	double	%12.0g	e17	Сколько рублей Ваша семья отложила в течение последних 30 дней?
e18	double	%25.0g	YES	Приходилось ли Вашей семье или кому-л из ее членов в течение последних 30 дней безвозм-но давать деньги или товары -продов-ие, одежду, др предметы- людям, не из Вашего дх, - детям, родителям, другим родственникам, просто чужим людям?
e19_1a	double	%25.0g	YES	Помогала Ваша семья деньгами или товарами в течение последних 30 дней родителям Вашим или супруга (и)?
e19_1b	double	%12.0g	e19_1b	Сколько составила помощь Вашей семьи деньгами или товарами в течение последних 30 дней родителям Вашим или супруга (и)?
e19_2a	double	%25.0g	YES	Помогала Ваша семья деньгами или товарами в течение последних 30 дней детям Вашим или супруга (и)?
e19_2b	double	%12.0g	e19_2b	Сколько составила помощь Вашей семьи деньгами или товарами в течение последних 30 дней детям Вашим или супруга (и)?
e19_3a	double	%25.0g	YES	Помогала Ваша семья деньгами или товарами в течение последних 30 дней бабушкам и дедушкам Вашим или супруга (и)?
e19_3b	double	%12.0g	e19_3b	Сколько составила помощь Вашей семьи деньгами или товарами в течение последних 30 дней бабушкам и дедушкам Вашим или супруга (и)?
e19_4a	double	%25.0g	YES	Помогала Ваша семья деньгами или товарами в течение последних 30 дней внукам Вашим или супруга (и)?
e19_4b	double	%12.0g	e19_4b	Сколько составила помощь Вашей семьи деньгами или товарами в течение последних 30 дней внукам Вашим или супруга (и)?
e19_5a	double	%25.0g	YES	Помогала Ваша семья деньгами или товарами в течение последних 30 дней другим людям?

```
e19_5b          double %12.0g          e19_5b          Сколько составила помощь Вашей семьи
                денгами или товарами в течение последних
                30 дней другим людям?
```

\*Перекодируем отсутствие направления расходов код «2» в «0»

```
recode d5 d17 e2 e5 e7_1a e7_1_0a e7_1_1a e7_2a e7_3a e7_4a e7_5a e7_6a e7_7a e7_8a e7_9a
e7_10a e8_1a e8_2a e8_3a e9_1a e9_2a e9_3a e9_4a e9_4_1a e9_5a e9_6a e9_7a e9_8a e9_9a
e9_10a e10 e13_1a e13_2a e13_21a e13_3a e13_31a e13_22a e13_23a e13_24a e13_32a e13_33a
e13_34a e13_4a e13_5a e13_6a e13_7a e13_8a e13_9a e13_71a e13_72a e13_10a e14 e16 (2=0)
```

```
label values d5 d17 e2 e5 e7_1a e7_1_0a e7_1_1a e7_2a e7_3a e7_4a e7_5a e7_6a e7_7a e7_8a
e7_9a e7_10a e8_1a e8_2a e8_3a e9_1a e9_2a e9_3a e9_4a e9_4_1a e9_5a e9_6a e9_7a e9_8a
e9_9a e9_10a e10 e13_1a e13_2a e13_21a e13_3a e13_31a e13_22a e13_23a e13_24a e13_32a
e13_33a e13_34a e13_4a e13_5a e13_6a e13_7a e13_8a e13_9a e13_71a e13_72a e13_10a e14
e16 YES
```

\*Корректировка нереалистичных значений. Посмотрим максимумы для 1998 г. и больше (так как в 1998 была деноминация, до этого большие цифры).

```
tabstat d6 d18 e3 e4 e6 e6_1 e6_2 e7_1b e7_1_0b e7_1_1b e7_2b e7_3b e7_4b e7_5b e7_6b e7_7b
e7_8b e7_9b e7_10b e8_1b e8_2b e8_3b e9_1b e9_2b e9_3b e9_4b e9_4_1b e9_5b e9_6b e9_7b
e9_8b e9_9b e9_10b e11 e12 e12_1 e12_2 e12_3 e13_1b e13_2b e13_21b e13_3b e13_31b e13_22b
e13_23b e13_24b e13_32b e13_33b e13_34b e13_4b e13_5b e13_6b e13_7b e13_8b e13_9b
e13_71b e13_72b e13_10b e15 e17 e19_1b e19_2b e19_3b e19_4b e19_5b if id_w >= 8, statistics(
max) columns(statistics) missing format (% 12.1g)
```

variable	max
d6	69000
d18	250000
e3	50000
e4	100000
e6	.
e6_1	200000
e6_2	53500
e7_1b	150000
e7_1_0b	68000
e7_1_1b	40000
e7_2b	300000
e7_3b	90000
e7_4b	72000000
e7_5b	37000
e7_6b	450000
e7_7b	300000
e7_8b	3500000
e7_9b	15000
e7_10b	24000
e8_1b	50000
e8_2b	20000
e8_3b	7000
e9_1b	30000
e9_2b	11000
e9_3b	15000
e9_4b	180000
e9_4_1b	100000
e9_5b	9400
e9_6b	15000
e9_7b	78000
e9_8b	13000
e9_9b	9000
e9_10b	65000
e11	999998

e12		2
e12_1		100000
e12_2		23000
e12_3		999
e13_1b		30400
e13_2b		120000
e13_21b		15000
e13_3b		5000
e13_31b		32000
e13_22b		60000
e13_23b		25000
e13_24b		70000
e13_32b		4000
e13_33b		5000
e13_34b		15000
e13_4b		130500
e13_5b		280000
e13_6b		70000
e13_7b		160000
e13_8b		20000
e13_9b		40000
e13_71b		100000
e13_72b		100000
e13_10b		50000
e15		620000
e17		500000
e19_1b		150000
e19_2b		490000
e19_3b		15000
e19_4b		48000
e19_5b		70000

В переменных e11 и e12 явно остались коды с прошлых вариантов кодирования пропущенных значений.

Есть явно несколько ошибок: раунд 8, семья с id\_h= 1391412, расходы на автомобиль - e7\_4b - 75 000 000 т.е. более 2 млн долл.; удалить или разделить на 1000.

То же самое по разделу доходов. В переменной f14 очень большое значение, 4198054 руб. для семьи № id\_h =800009 в 8-ом раунде явно не учтена деноминация (мы удалили это значение только в дефлированном душевом доходе); в переменных f13\_1b f14\_3 - 5 000 000 - то же самое.

```
recode e11(999998=.)
```

```
recode e12_3 (999 = .)
```

```
replace f14=f14/1000 if (id_h ==800009 & id_w ==8)
```

```
replace e7_4b=e7_4b/1000 if (id_h ==1391412 & id_w ==8)
```

```
replace f13_1b=f13_1b/1000 if (id_h ==1391412 & id_w ==8)
```

```
replace f14_3=f14_3/1000 if (id_h ==1391412 & id_w ==8)
```

\*21.2. Создадим новые переменные по категориям, суммы расходов с учетом дефлятора

\*Питание 30 дней - e4

```
gen pitanie = e4* deflat_16
```

\*3 месяца: Одежда - e6, Одежда для взрослых - e6\_1, Одежда для детей - e6\_2 (перекодируем 99999996, если в семье нет детей, в 0)

```
recode e6_2 (99999996 = 0)
```

```
egen odejda = rowtotal(e6 e6_1 e6_2), missing
```

```
replace odejda = odejda /3 * deflat_16
```

\*3 месяца: Культтовары + моб.телефон - e7\_1b, Культтовары - e7\_1\_0b, Мобильный телефон

- e7\_1\_1b, Мебель - e7\_2b, Бытовая техника - e7\_3b, Спортивный инвентарь - e7\_10b

```
egen tdp = rowtotal(e7_1b e7_1_0b e7_1_1b e7_2b e7_3b e7_10b), missing  
replace tdp = tdp /3 * deflat_16
```

\*3 месяца: Автомобиль - e7\_4b, Мотоцикл - e7\_5b

```
egen auto = rowtotal(e7_4b e7_5b), missing  
replace auto = auto /3 * deflat_16
```

\*3 месяца: Недвижимость - e7\_8b, Гараж - e7\_6b

```
egen nedv = rowtotal(e7_8b e7_6b), missing  
replace nedv = nedv /3 * deflat_16
```

\*30 дней: Топливо разное - e8\_1b, Дрова и т.д. - e8\_2b, Газ - e8\_3b

```
egen topliv = rowtotal(e8_1b e8_2b e8_3b), missing  
replace topliv = topliv * deflat_16
```

\*30 дней: Лечение (без лекарств) - e13\_3b, Лекарства - e13\_31b, Лечение в больницах - e13\_22b, Лечение в поликлиниках - e13\_23b, Лечение зубов - e13\_24b

```
egen medecine = rowtotal(e13_3b e13_31b e13_22b e13_23b e13_24b), missing  
replace medecine = medecine * deflat_16
```

\*30 дней: Образование детей - e13\_1b, Образование взрослых - e13\_4b

```
egen educ = rowtotal(e13_1b e13_4b), missing  
replace educ = educ * deflat_16
```

\*Отдых и туризм 30 дней - e13\_2b, Культура 30 дней - e13\_21b, Книги 3 месяца - e7\_9b

```
gen var_var = e7_9b/3  
egen cultura = rowtotal(e13_2b e13_21b var_var ), missing  
drop var_var  
replace cultura = cultura * deflat_16
```

\*Помощь разным людям за 30 дней: e19\_1b, e19\_2b, e19\_3b, e19\_4b, e19\_5b

```
egen help = rowtotal(e19_1b e19_2b e19_3b e19_4b e19_5b), missing  
replace help = help * deflat_16
```

\*30 дней: Жилье и коммунальные услуги - e11

```
gen jku = e11 * deflat_16
```

\*Услуги 30 дней: Транспорт - e9\_1b, Пошив одежды - e9\_2b, Ремонт техники - e9\_3b, Ремонт автомобиля - e9\_4\_1b, Чистка, баня и т.д. - e9\_5b, Загс и ритуальные услуги - e9\_7b, Юристы и т.д. - e9\_10b

```
egen uslugi = rowtotal(e9_1b e9_2b e9_3b e9_4_1b e9_5b e9_7b e9_10b), missing  
replace uslugi = uslugi * deflat_16
```

\*Строительство и ремонт жилья 30 дней - e9\_4b, Стройматериалы 3 месяца - e7\_7b

```
gen var_var = e7_7b/3
egen remont = rowtotal(e9_4b var_var), missing
drop var_var
replace remont = remont * deflat_16
```

\*30 дней: Почта - e9\_6b, Мобильная связь - e9\_8b, Интернет - e9\_9b

```
egen sviaz = rowtotal(e9_6b e9_8b e9_9b), missing
replace sviaz = sviaz * deflat_16
```

\*30 дней (с 9й волны): Моющие средства - e13\_32b, Гигиена - e13\_33b, Парфюмерия и косметика - e13\_34b

```
egen hygiene = rowtotal(e13_32b e13_33b e13_34b) if id_w >= 9, missing
replace hygiene = hygiene * deflat_16
```

\*Пользование землей за 12 месяцев - d6, Семена и пр. за 30 дней - d18

```
gen var_var = d6/12
egen agro = rowtotal(var_var d18), missing
drop var_var
replace agro = agro * deflat_16
```

\*30 дней: Страховые взносы - e13\_6b, алименты - e13\_8b, налоги - e13\_9b

```
egen nalog = rowtotal(e13_6b e13_8b e13_9b), missing
replace nalog = nalog * deflat_16
```

\*30 дней: Погашение кредита, возврат долгов - e13\_7b, Уплата долгов - e13\_71b, Погашение кредита - e13\_72b

```
egen vozvr_dolg = rowtotal(e13_7b e13_71b e13_72b), missing
replace vozvr_dolg = vozvr_dolg * deflat_16
```

\*30 дней: Дали в долг - e15, Валюта на сбережения - e13\_10b, Отложили сбережений - e17, Акции, облигации - e13\_5b

```
egen sberejen = rowtotal(e15 e13_10b e17 e13_5b), missing
replace sberejen = sberejen * deflat_16
```

```
label variable pitanie "расходы на питание"
label variable odejda "расходы на одежде"
label variable tdp "расходы на ТДП"
label variable auto "расходы на покупку автотранспорта"
label variable nedv "расходы на недвижимость"
label variable topliv "расходы на топливо"
label variable medecine "расходы на медицину"
label variable educ "расходы на образование"
label variable cultura "расходы на культура"
label variable help "расходы на помощь другим"
```

```

label variable jku "расходы на ЖКУ"
label variable uslugi "расходы на разные услуги"
label variable remont "расходы на ремонт и строительство"
label variable sviaz "расходы на связь"
label variable hygiene "расходы на средства гигиены и т.д."
label variable agro "расходы на землю, семена и т.д."
label variable nalog "налоги, алименты, страховки"
label variable vozvr_dolg "возврат долгов и кредитов"
label variable sberejen "отложено сбережений, акции и т.д."

```

Посмотрите в таблицу данных – в получившихся переменных очень много пропущенных значений. Это те домохозяйства, у которых не было таких расходов, но также и те, которые затруднились назвать либо сумму, либо даже были ли у них такие расходы. Если бы мы работали с данными за один год, лучше было бы по каждой переменной расходов (до объединения в категории) перекодировать миссинг (.) в 0, если данного вида расходов не было (дихотомическая переменная наличия такого вида расходов = 2). Но за разные годы эти переменные могли быть разными (например, `еб еб_1 еб_2`), и тогда мы бы все равно не смогли суммировать эти переменные.

\*Поэтому условно будем считать, что скорее всего, если нет значения – это 0.

```

recode pitanie odejda tdp auto nedv topliv medecine educ cultura help jku uslugi remont sviaz
hygiene agro nalog vozvr_dolg sberejen (. = 0)
recode hygiene (. = 0) if id_w >= 9

```

\*21.3. Рассчитаем сумму расходов домохозяйства за 30 дней (учтем, что переменная `hygiene` определена, начиная с 9го раунда):

```

gen EXPEND = pitanie + odejda + tdp + auto + nedv + topliv + medecine + educ + cultura+
help + jku + uslugi + remont + sviaz + hygiene + agro + nalog + vozvr_dolg +sberejen if id_w
>= 9
replace EXPEND = pitanie + odejda + tdp + auto + nedv + topliv + medecine + educ + cultura+
help + jku + uslugi + remont + sviaz + agro + nalog + vozvr_dolg +sberejen if id_w < 9
label variable EXPEND "суммарные расходы за 30 дней"

```

```
sum EXPEND
```

Variable	Obs	Mean	Std. Dev.	Min	Max
EXPEND	53,511	15649.91	394258.8	0	9.10e+07

\*У 304 домохозяйств суммарные расходы оказались равны 0; удалим значения расходов для них.

```

recode pitanie odejda tdp auto nedv topliv medecine educ cultura help jku uslugi remont sviaz
hygiene agro nalog vozvr_dolg sberejen (0=.) if EXPEND == 0
recode EXPEND(0=.) if EXPEND == 0

```

\*Рассчитаем долю расходов по каждому виду в %

```

gen d_pitanie = pitanie / EXPEND*100 if EXPEND != 0
gen d_odejda = odejda / EXPEND*100 if EXPEND != 0
gen d_tdp = tdp / EXPEND*100 if EXPEND != 0
gen d_auto = auto / EXPEND*100 if EXPEND != 0
gen d_nedv = nedv / EXPEND*100 if EXPEND != 0

```

```

gen d_topliv = topliv / EXPEND*100 if EXPEND != 0
gen d_medecine = medecine / EXPEND*100 if EXPEND != 0
gen d_educ = educ / EXPEND*100 if EXPEND != 0
gen d_cultura = cultura / EXPEND*100 if EXPEND != 0
gen d_help = help / EXPEND*100 if EXPEND != 0
gen d_jku = jku / EXPEND*100 if EXPEND != 0
gen d_uslugi = uslugi / EXPEND*100 if EXPEND != 0
gen d_remont = remont / EXPEND*100 if EXPEND != 0
gen d_sviaz = sviaz / EXPEND*100 if EXPEND != 0
gen d_hygiene = hygiene / EXPEND*100 if EXPEND != 0
gen d_agro = agro / EXPEND*100 if EXPEND != 0
gen d_nalog = nalog / EXPEND*100 if EXPEND != 0
gen d_vozvr_dolg = vozvr_dolg / EXPEND*100 if EXPEND != 0
gen d_sberejen= sberejen / EXPEND*100 if EXPEND != 0

```

```

label variable d_pitanie "% на питание"
label variable d_odejda "% на одежду"
label variable d_tdp "% на ТДП"
label variable d_auto "% на авто"
label variable d_nedv "% на недвижимость"
label variable d_topliv "% на топливо"
label variable d_medecine "% на медицину"
label variable d_educ "% на образование"
label variable d_cultura "% на культура"
label variable d_help "% на помощь другим"
label variable d_jku "% на ЖКУ"
label variable d_uslugi "% на услуги"
label variable d_remont "% на ремонт и строительство"
label variable d_sviaz "% на связь"
label variable d_hygiene "% на стерства гигиены"
label variable d_agro "% на землю, семена и т.д."
label variable d_nalog "% на налоги, алименты, страховки"
label variable d_vozvr_dolg "% на возврат долгов и кредитов"
label variable d_sberejen "% на сбережения и активы"

```

\*Посмотрим описательные характеристики

```

sum pitanie odejda tdp auto nedv topliv medecine educ cultura help jku uslugi remont sviaz
hygiene agro nalog vozvr_dolg sberejen EXPEND d_pitanie d_odejda d_tdp d_auto d_nedv
d_topliv d_medecine d_educ d_cultura d_help d_jku d_uslugi d_remont d_sviaz d_hygiene
d_agro d_nalog d_vozvr_dolg d_sberejen

```

Variable	Obs	Mean	Std. Dev.	Min	Max
pitanie	53,206	3903.088	4083.654	0	176961.6
odejda	53,206	1329.096	2743.5	0	144885.3
tdp	53,206	676.4656	3286.152	0	379599.4
auto	53,206	630.7058	8618.737	0	503639.8
nedv	53,206	610.0068	13928.24	0	1372640
topliv	53,206	482.9152	1483.234	0	75545.97
medecine	53,206	501.8105	2001.055	0	134492.6
educ	53,206	244.278	1546.968	0	184637.9
cultura	53,206	223.1714	2081.682	0	121000
help	53,206	660.3859	5061.088	0	663370.8

jku	53,206	815.0119	1220.128	0	42522
uslugi	53,206	777.5665	2908.169	0	137024.3
remont	53,206	571.3321	5127.041	0	406135.1
sviaz	53,206	226.476	625.3139	0	25383.45
hygiene	53,206	281.4849	530.0398	0	16952.85
agro	53,206	222.752	1742.414	0	281859.3
nalog	53,206	113.2123	1073.836	0	78330
vozvr_dolg	53,206	586.0828	3875.386	0	273289.8
sberejen	53,206	1119.043	9347.423	0	876490.4
EXPEND	53,511	13895.23	27362.35	0	1476549
d_pitanie	53,206	40.72534	26.82412	0	100
d_odejda	53,206	9.754491	13.90465	0	100
d_tdp	53,206	3.0083	9.556902	0	100
d_auto	53,206	.6701687	6.656993	0	97.12038
d_nedv	53,206	.4027315	5.397718	0	99.59258
d_topliv	53,206	3.843087	10.66332	0	100
d_medecine	53,206	4.532935	10.01019	0	100
d_educ	53,206	1.41773	5.373196	0	100
d_cultura	53,206	1.072188	4.32811	0	100
d_help	53,206	4.214453	11.36366	0	100
d_jku	53,206	9.214507	13.11105	0	100
d_uslugi	53,206	5.016827	9.576847	0	100
d_remont	53,206	1.802264	8.170817	0	100
d_sviaz	53,206	1.513177	3.793815	0	100
d_hygiene	53,206	2.659394	5.148286	0	100
d_agro	53,206	2.59055	9.714982	0	100
d_nalog	53,206	.6103498	3.742692	0	100
d_vozvr_dolg	53,206	2.424289	8.219866	0	100
d_sberejen	53,206	4.527213	12.6016	0	100

\*Сохраните файл

save "C:\RLMS\_work\seminar\_5\data\hh\_5\_16\_S5.dta", replace

\*21.4. Посмотрим средние значения доли расходов на разные направления для групп домохозяйств из разных квинтильных групп (лучше для какого-то определенного года). Сравните долю расходов на разные категории в разных доходных группах, а также за 1994 и за 2007 год. Что изменилось? Для корректного сравнения нужно использовать тест ANOVA.

\* Доли расходов за 5ю волну

tabstat d\_pitanie d\_odejda d\_tdp d\_auto d\_nedv d\_topliv d\_medecine d\_educ d\_cultura d\_help d\_jku d\_uslugi d\_remont d\_sviaz d\_agro d\_nalog d\_vozvr\_dolg d\_sberejen [aweight = hhwgt] if id\_w == 5, statistics( mean ) by(NI\_income)

NI_income	d_pitanie	d_odejda	d_tdp	d_auto	d_nedv	d_topliv	d_medecine
1	48.61075	12.48162	2.316894	.4883207	.0404417	6.205762	1.218169
2	53.23593	8.282963	2.55414	.1917025	.2337438	5.299951	.8217503
3	55.66783	7.579706	2.234796	.6053459	.5131386	3.495605	.6463496
4	53.2439	10.13869	3.03699	.8493641	.368583	2.957246	.9127732
5	38.84535	12.50925	5.853929	1.40761	.5132573	1.972797	1.759213
Total	49.9694	10.06508	3.259943	.7265123	.3538012	3.836434	1.064465

NI_income	d_educ	d_cultura	d_help	d_jku	d_uslugi	d_remont	d_sviaz
1	1.946283	.0821069	4.039102	2.780749	5.312863	1.218149	.9162539
2	1.357613	.0029036	4.63566	3.332097	5.493336	.8816421	.5701749
3	.9817115	.2022565	5.513378	3.046633	4.715193	1.103788	1.119691

4		1.378108	.1421131	5.722715	3.140424	4.906574	1.032012	1.000446
5		2.042017	.4384045	7.103931	2.06508	5.776	1.610137	1.154262
-----								
Total		1.516499	.1813202	5.495809	2.873328	5.23403	1.169454	.9584581
-----								

NI_income		d_agro	d_nalog	d_vozv~g	d_sber~n
1		7.307993	1.139419	1.213264	2.681859
2		5.545638	.8111424	2.119429	4.630176
3		4.04056	.9461072	1.380211	6.207698
4		2.243149	.8891176	1.613022	6.424772
5		1.590825	1.138233	2.402909	11.81679
-----					
Total		3.940957	.9771175	1.775577	6.601815
-----					

\* Доли расходов за 16ю волну

**tabstat d\_pitanie d\_odejda d\_tdp d\_auto d\_nedv d\_topliv d\_medecine d\_educ d\_cultura d\_help d\_jku d\_uslugi d\_remont d\_sviaz d\_hygiene d\_agro d\_nalog d\_vozvr\_dolg d\_sberejen [aweight = hhwt] if id\_w == 16, statistics( mean ) by(NI\_income)**

NI_income		d_pita~e	d_odejda	d_tdp	d_auto	d_nedv	d_topliv	d_mede~e
1		39.53655	8.644181	2.542063	.2655312	.0126247	4.691339	4.493511
2		39.54413	6.214581	2.551301	.3220335	.1110487	3.474632	6.25714
3		37.50583	7.188204	3.013018	.716281	.1264128	3.001545	5.240868
4		34.06307	8.32197	3.927389	1.45108	.2607781	2.985453	4.828076
5		28.49198	8.228396	4.788032	1.297777	.8516975	3.187885	4.230292
-----								
Total		35.9485	7.686181	3.338499	.7971527	.2653708	3.466035	5.035613
-----								

NI_income		d_educ	d_cult~a	d_help	d_jku	d_uslugi	d_remont	d_sviaz
1		.9086035	1.327956	2.789513	13.88781	3.74344	1.14466	2.860423
2		.6049886	.8313011	3.868919	15.27449	4.004901	1.74098	2.826655
3		.8068759	1.037291	4.152584	13.38775	4.465088	2.264569	2.999546
4		1.214515	1.42485	4.128077	11.05108	4.744813	2.964365	3.666961
5		1.36163	1.863975	5.236688	9.251411	5.750619	3.685375	3.679628
-----								
Total		.9688389	1.283667	4.02742	12.64324	4.525851	2.340395	3.194249
-----								

NI_income		d_hygi~e	d_agro	d_nalog	d_vozv~g	d_sber~n
1		4.828498	2.865216	.4622369	3.531785	1.46406
2		3.399071	1.686114	.3783009	3.484436	3.42498
3		3.315465	1.42242	.5624087	4.569181	4.224664
4		3.218424	.991738	.6744585	4.61087	5.472024
5		3.095662	.4833462	.831437	5.762874	7.92129
-----						
Total		3.569802	1.499167	.5768992	4.372758	4.460365
-----						

## 22. Факторный анализ % расходов для 16й волны.

*Statistics > Multivariate analysis > Factor and principal component analysis > Factor analysis*

Команда ФА имеет вид:

**factor varlist [if] [in] [weight] [, method options]**

Метод извлечения: **pf** - principal factor, по умолчанию; **pcf** - principal-component factor (метод главных компонент); **ipf** - iterated principal factor; **ml** - maximum-likelihood factor, метод максимального правдоподобия, для дихотомических переменных.

### Maximization

**protect(#)** - perform # optimizations and report the best solution (ml only)

**random** - use random starting values (ml only); seldom used.

**mineigen(#)** – какое минимальное собственное значение устанавливается для факторов; определяет максимальное количество факторов, которые будут сохранены косвенно, сохраняя все факторы с собственными значениями больше, чем указанное значение. Многие авторы рекомендуют mineigen(1), хотя доказательство сложно и сомнительно (из «помощи» Stata).

В STATA довольно много пост-оценок (пост-эстимация), в частности, вращение и сохранение факторов как новых переменных.

**rotate** - Perform varimax rotation, the default

**factor d\_pitanie d\_odejda d\_tdp d\_auto d\_nedv d\_topliv d\_medecine d\_educ d\_cultura d\_help d\_jku d\_uslugi d\_remont d\_sviaz d\_hygiene d\_agro d\_nalog d\_vozvr\_dolg d\_sberejen if id\_w == 16, pcf mineigen(1)**

(obs=5,413)

```
Factor analysis/correlation          Number of obs   =      5,413
Method: principal-component factors  Retained factors =         9
Rotation: (unrotated)                Number of params =     135
```

Factor	Eigenvalue	Difference	Proportion	Cumulative
Factor1	1.64974	0.15064	0.0868	0.0868
Factor2	1.49910	0.30488	0.0789	0.1657
Factor3	1.19422	0.04930	0.0629	0.2286
Factor4	1.14492	0.03599	0.0603	0.2888
Factor5	1.10894	0.05030	0.0584	0.3472
Factor6	1.05864	0.04939	0.0557	0.4029
Factor7	1.00924	0.00216	0.0531	0.4560
Factor8	1.00708	0.00192	0.0530	0.5090
Factor9	1.00516	0.01753	0.0529	0.5619
Factor10	0.98764	0.01107	0.0520	0.6139
Factor11	0.97657	0.01500	0.0514	0.6653
Factor12	0.96157	0.00306	0.0506	0.7159
Factor13	0.95851	0.00932	0.0504	0.7664
Factor14	0.94919	0.02896	0.0500	0.8163
Factor15	0.92024	0.02029	0.0484	0.8648
Factor16	0.89995	0.01993	0.0474	0.9121
Factor17	0.88002	0.09073	0.0463	0.9585
Factor18	0.78929	0.78929	0.0415	1.0000
Factor19	0.00000	.	0.0000	1.0000

LR test: independent vs. saturated: chi2(171) = 1.7e+05 Prob>chi2 = 0.0000

Factor loadings (pattern matrix) and unique variances

Variable	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
d_pitanie	-0.7438	-0.1281	-0.5697	-0.1506	-0.0633	0.0671
d_odejda	0.3061	0.4211	-0.2001	0.1341	-0.2946	-0.0593
d_tdp	0.4109	-0.0294	-0.0338	-0.3528	0.0003	-0.3756
d_auto	0.1922	-0.1804	0.0348	-0.0354	0.3435	0.4437
d_nedv	0.1026	-0.1260	0.0945	-0.1026	-0.0305	-0.0619
d_topliv	0.2357	-0.1635	-0.3088	0.5248	0.1812	0.0380
d_medecine	-0.2109	-0.0841	0.4719	0.0576	0.3447	-0.0599
d_educ	0.1967	0.2426	-0.1027	-0.2082	-0.0383	0.1676
d_cultura	0.1812	0.2283	-0.1497	-0.1175	-0.2636	0.2541
d_help	0.0462	-0.3231	0.2858	0.3257	-0.1878	0.0556
d_jku	-0.4851	0.2233	0.3842	-0.1928	0.1599	-0.0895
d_uslugi	0.1222	0.3864	0.2136	-0.0334	0.0231	0.2496
d_remont	0.2445	-0.2317	0.0444	-0.1747	-0.0738	-0.4780
d_sviaz	0.0105	0.5760	0.0457	0.2247	-0.0131	-0.0026
d_hygiene	-0.1015	0.5296	0.1675	0.3057	0.0367	-0.2153



d_hygiene		0.1455	0.6755	0.0169	-0.1114	-0.0232	0.0526
d_agro		0.0793	0.2068	-0.0338	0.3516	-0.5127	-0.0479
d_nalog		0.1348	-0.1003	-0.0954	0.2517	0.0760	0.0569
d_vozvr_dolg		0.3406	-0.3399	0.0761	0.0354	0.1680	-0.0009
d_sberejen		0.2280	-0.1888	0.0356	-0.0884	-0.0070	-0.1463

Variable		Factor7	Factor8	Factor9		Uniqueness
d_pitanie		0.0845	-0.0824	-0.0419		0.0486
d_odejda		-0.1128	-0.1117	-0.1663		0.4919
d_tdp		0.2680	-0.1607	-0.1088		0.4887
d_auto		-0.0191	0.8671	0.0212		0.2266
d_nedv		-0.0531	0.0173	0.8223		0.2995
d_topliv		0.0890	-0.0155	0.0394		0.3933
d_medecine		0.0011	-0.1326	-0.0104		0.2970
d_educ		0.0110	-0.1834	-0.1249		0.7040
d_cultura		-0.2932	-0.1306	-0.2019		0.5830
d_help		-0.1443	-0.0649	-0.1348		0.2643
d_jku		0.0000	0.0477	-0.0990		0.4122
d_uslugi		0.0556	-0.0134	0.2312		0.5147
d_remont		0.1099	-0.1167	0.3667		0.5569
d_sviaz		0.1739	-0.1319	-0.0288		0.5426
d_hygiene		0.1430	-0.0487	-0.0757		0.4780
d_agro		-0.0429	-0.0304	-0.0146		0.5580
d_nalog		0.1218	0.4042	-0.1699		0.6833
d_vozvr_dolg		0.5013	0.0466	-0.0730		0.4745
d_sberejen		-0.7577	0.0263	0.0360		0.3058

Factor rotation matrix

		Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
Factor1		0.8045	-0.2916	0.4050	0.2811	0.0794	0.0521	0.0697
Factor2		0.1575	0.6176	0.1365	-0.1754	0.4921	0.4090	0.2289
Factor3		0.5145	0.1907	-0.5354	-0.4273	0.0818	-0.3361	-0.2961
Factor4		0.0617	0.5865	-0.0680	0.6874	-0.1974	-0.3319	-0.1139
Factor5		0.0549	-0.0511	-0.4795	0.1320	0.0317	0.0603	0.7262
Factor6		-0.1686	-0.1190	-0.0379	0.1885	0.6368	0.0196	-0.4359
Factor7		-0.0028	-0.0666	-0.3657	0.3686	0.1165	0.4858	-0.1704
Factor8		-0.1025	0.3319	0.3964	-0.1854	-0.0503	-0.0885	0.0629
Factor9		0.1324	0.1478	-0.0618	-0.1093	-0.5322	0.5992	-0.3028

		Factor8	Factor9
Factor1		0.0933	0.0462
Factor2		-0.2170	-0.1884
Factor3		0.0392	0.1443
Factor4		0.0194	-0.1119
Factor5		0.4628	-0.0291
Factor6		0.5404	-0.1802
Factor7		-0.2560	0.6179
Factor8		0.4710	0.6721
Factor9		0.3857	-0.2493

**predict** creates new variables containing predictions such as factors scored by the regression method or by the Bartlett method.

**regression** produces factors scored by the regression method. This is the default.

**bartlett** produces factors scored by the method suggested by Bartlett (1937, 1938). This method produces unbiased factors, but they may be less accurate than those produced by the default regression method suggested by Thomson (1951). Regression-scored factors have the smallest mean squared error from the true factors but may be biased.

**norotated** specifies that unrotated factors be scored even when you have previously issued a rotate command. The default is to use rotated factors if they are available and unrotated factors otherwise.

\*Сохраним факторы как новые переменные.

### predict Fc1 Fc2 Fc3 Fc4 Fc5 Fc6 Fc7 Fc8 Fc9

(regression scoring assumed)

Scoring coefficients (method = regression; based on varimax rotated factors)

Variable	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
d_pitanie	-0.95879	0.10918	0.08604	0.16186	-0.02470	0.18272
d_odejda	0.00436	0.32636	0.29147	0.14933	-0.10903	0.36737
d_tdp	0.11332	-0.16501	0.16368	-0.18929	-0.20031	0.07172
d_auto	-0.07277	0.10691	0.07230	-0.00670	-0.02311	0.10330
d_nedv	-0.05723	0.05075	-0.02241	0.09856	0.07048	0.14658
d_topliv	-0.12432	0.10658	-0.11326	0.72897	-0.05962	0.08430
d_medecine	0.04812	0.02399	-0.73851	0.13964	-0.01367	0.12360
d_educ	-0.00329	-0.15524	-0.03103	0.09307	0.29849	0.22623
d_cultura	-0.01883	-0.07097	0.03701	0.13816	0.15801	0.40923
d_help	-0.07718	0.10757	0.04622	0.10865	0.06525	-0.67588
d_jku	-0.16310	0.29525	-0.37135	-0.30455	-0.11871	0.23931
d_uslugi	-0.01784	0.19159	-0.01124	0.03583	0.50444	-0.07529
d_remont	0.03688	0.00011	0.11111	-0.09142	-0.36608	0.08609
d_sviaz	-0.00203	0.41341	0.10847	0.05624	0.24743	-0.07283
d_hygiene	0.05978	0.57582	0.01365	-0.04131	-0.10387	0.04228
d_agro	-0.00257	0.27129	-0.04781	0.33581	-0.46885	0.03596
d_nalog	0.03008	-0.03295	-0.12141	0.22735	0.06605	0.08279
d_vozvr_dolg	0.05198	-0.17121	-0.01788	0.07993	0.12048	0.04312
d_sberejen	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000

Variable	Factor7	Factor8	Factor9
d_pitanie	1.28563	-0.08000	-0.03519
d_odejda	0.44633	-0.08717	-0.14499
d_tdp	0.75360	-0.17464	-0.12854
d_auto	0.42937	0.80740	0.00626
d_nedv	0.21191	0.00296	0.79989
d_topliv	0.58235	-0.05207	0.03489
d_medecine	0.50256	-0.15429	-0.02589
d_educ	0.21121	-0.17756	-0.11383
d_cultura	-0.04309	-0.11857	-0.18899
d_help	0.46888	-0.09514	-0.16062
d_jku	0.69086	0.06257	-0.09351
d_uslugi	0.43605	-0.00622	0.24472
d_remont	0.61937	-0.13395	0.33331
d_sviaz	0.36818	-0.09498	0.00674
d_hygiene	0.40380	-0.00987	-0.04470
d_agro	0.29306	-0.04376	-0.02482
d_nalog	0.28136	0.36255	-0.17175
d_vozvr_dolg	1.04481	0.00907	-0.08359
d_sberejen	0.00000	0.00000	0.00000

## 23. Кластерный анализ методом К-средних на переменных % расходов на разные направления.

### 23.1. Виды кластерного анализа в STATA.

#### Partition-clustering methods for observations

(see [MV] cluster kmeans and kmedians)

Kmeans

cluster kmeans

Kmedians

cluster kmedians

#### Hierarchical clustering methods for observations

(see [MV] cluster linkage)

Single linkage

cluster singlelinkage

```

Average linkage                cluster averagelinkage
Complete linkage               cluster completelinkage
Weighted-average linkage      cluster waveragelinkage
Median linkage                 cluster medianlinkage
Centroid linkage              cluster centroidlinkage
Ward's linkage                 cluster wardslinkage
Postclustering commands
Stopping rules                 cluster stop
Dendrograms (cluster trees)   cluster dendrogram
                               (synonym: cluster tree)
Generate summary variables    cluster generate
User utilities
Cluster notes                  cluster notes
Other user utilities           (see [MV] cluster utility)
                               cluster dir
                               cluster list
                               cluster drop
                               cluster use
                               cluster rename
                               cluster renamevar

```

#### cluster kmeans

```

k(#)                perform cluster analysis resulting in # groups
  measure(measure)  similarity or dissimilarity measure; default is L2 (Euclidean)
  name(clname)      name of resulting cluster analysis
Options
  start(start_option)  obtain k initial group centers by using start_option
  keepcenters          append the k final group means or medians to the data
Advanced
  generate(groupvar)  name of grouping variable

```

#### Statistics > Multivariate analysis > Cluster analysis > Cluster data > Kmeans

Команда создает только переменную с именем номера кластера, кластерные центры нужно рассчитать на основе средних значений.

```

cluster kmeans d_pitanie d_odejda d_tdp d_auto d_nedv d_topliv d_medecine d_educ
d_cultura d_help d_jku d_uslugi d_remont d_sviaz d_hygiene d_agro d_nalog d_vozvr_dolg
d_sberejen if id_w ==16, k(4) measure(L2) name(Cluster_expend) start(krandom)
generate(Cl_exp)

```

```
tab Cl_exp if id_w ==16
```

Cl_exp	Freq.	Percent	Cum.
1	2,220	41.01	41.01
2	414	7.65	48.66
3	2,360	43.60	92.26
4	419	7.74	100.00
Total	5,413	100.00	

Посмотрим средние значения исходных переменных по выделенным кластерам

```

tabstat d_pitanie d_odejda d_tdp d_auto d_nedv d_topliv d_medecine d_educ d_cultura d_help
d_jku d_uslugi d_remont d_sviaz d_hygiene d_agro d_nalog d_vozvr_dolg d_sberejen if id_w
== 16, statistics( mean ) by(Cl_exp)

```

```

Summary statistics: mean
  by categories of: Cl_exp

```

Cl_exp	d_pita~e	d_odejda	d_tdp	d_auto	d_nedv	d_topliv	d_mede~e
1	55.80597	5.551153	1.123573	.0032127	.003464	1.952153	4.264661
2	18.72914	5.91508	2.686111	.2041921	.5371419	1.969187	3.231456
3	19.74107	11.9761	6.638187	2.361219	.6100847	5.401847	5.67756
4	18.95755	4.489499	2.747963	.1003057	.0829333	2.664516	3.108093
Total	34.3941	8.298006	3.77312	1.054161	.3149114	3.512621	4.712119

Cl_exp	d_educ	d_cult~a	d_help	d_jku	d_uslugi	d_remont	d_sviaz
1	.6509585	.9116667	1.867844	13.57751	3.643774	.3566809	3.248408
2	.6474159	.7737179	3.942019	7.499984	3.638675	1.049491	2.233113
3	1.794445	1.996141	1.873162	12.53404	5.981962	1.053421	3.892479
4	.555258	.5220772	25.19919	6.51399	2.843168	20.75684	2.299448
Total	1.141826	1.343777	3.834793	12.11099	4.600833	2.292539	3.378107

Cl_exp	d_hygi~e	d_agro	d_nalog	d_vozv~g	d_sber~n
1	3.595944	.9635959	.2571909	1.111505	1.110736
2	1.938246	1.059425	.4178389	1.756212	41.77155
3	4.403974	1.718222	1.074743	9.768057	1.503278
4	2.288634	1.556164	.6786149	2.506169	2.12958
Total	3.720256	1.345801	.6585411	5.042918	4.470587

\*Посмотрим на средние значения по кластерам (нужен тест на разницу средних) – душевого дохода, количества человек в семье и количества детей.

**tabstat INCOME\_PC nfm num\_children17 if id\_w == 16, statistics( mean ) by(Cl\_exp)**

Summary statistics: mean  
by categories of: Cl\_exp

Cl_exp	INCOME~C	nfm	num_c~17
1	5662.343	2.584685	.4477477
2	10223.62	2.490338	.468599
3	7640.572	3.027119	.7050847
4	8463.239	2.458234	.3627685
Total	7078.751	2.760576	.5549603

\*Немного преобразуем переменную «тип населенного пункта»

```
gen status1 = status
recode status1 (4=3)
recode status1 (1=0) if psu <= 2
label variable status1 "тип населенного пункта"
label define status1_lab 0 "М., СП" 1 "обл.центр" 2 "другой город" 3 "село, пгт"
label values status1 status1_lab
```

\*Посмотрим, как распределены домохозяйства из полученных кластеров по типам населенных пунктов (посмотрите на разницу между вертикальным и горизонтальным %: как их интерпретировать).

**tabulate Cl\_exp status1, column nofreq**

	тип населенного пункта
--	------------------------

Cl_exp	М., СП	обл.центр	другой го	село, пгт	Total
1	54.57	40.14	40.44	36.94	41.01
2	8.69	7.49	7.01	7.97	7.65
3	31.55	45.72	44.38	45.58	43.60
4	5.18	6.65	8.17	9.51	7.74
Total	100.00	100.00	100.00	100.00	100.00

### tabulate Cl\_exp status1, nofreq row

Cl_exp	М., СП	тип населенного пункта			Total
		обл.центр	другой го	село, пгт	
1	16.13	30.18	26.76	26.94	100.00
2	13.77	30.19	24.88	31.16	100.00
3	8.77	32.33	27.63	31.27	100.00
4	8.11	26.49	28.64	36.75	100.00
Total	12.12	30.83	27.14	29.91	100.00

1я группа – самые бедные, 56% тратят на еду и 14% на ЖКУ, 2я группа самые богатые, 41% тратят на сбережения; 3я группа – невысокие доходы, но на еду тратят не слишком много (20%), но выше доля затрат на культуру, образование и т.д., самая высокая доля д\х с детьми; 4я группа отличается высокой долей затрат на ремонт или строительство, и на помощь другим людям.

\*23.2. Сохраните файл с данными; сохраните do-файл.

### 23.3. Самостоятельное задание.

Для дихотомических переменных о наличии тех или иных видов получения или оказания материальной помощи (f9\_1a f9\_2a f9\_3a f9\_4a f9\_5a f9\_6a f9\_7a f9\_8a f9\_9a f9\_91a f9\_10a e19\_1a e19\_2a e19\_3a e19\_4a e19\_5a) для 16й волны оцените модель кластерного анализа методом К-средних, с вращением, мера связи для бинарных данных **measure(matching)**. Количество кластеров 4-6 (выберите наилучший вариант). Сохраните переменную номера кластера. Посмотрите ее распределение. Рассчитайте кластерные центры. Рассчитайте средний душевой доход в полученных кластерах.