

## ЗАДАНИЯ К ПРАКТИЧЕСКИМ ЗАНЯТИЯМ.

### Занятие 6.

**Данные файлов по индивидуальным анкетам. Кластерный и факторный анализ. Матрицы мобильности. Корректировка переменных о доходах.**

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 (лог-файл)**. Первая команда в нем должна быть вида:

**\*Фамилия – номер семинара – номер задания**

3) Сохраняйте сделанную 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.

**ind\_5\_16.dta** – основной рабочий файл

**H\_5\_16\_family.dta** – некоторые переменные из семейного файла

**ind\_relatives\_5\_30\_short.dta** – некоторые переменные из файла родственных связей

**reg\_statistics\_1994\_2021.dta** – некоторые переменные региональной статистики

### 24.0. Преобразования для получения файла на начало занятия (уже сделаны)

\*До занятия, я скачала с сайта лонгитюдный файл по индивидам **RLMS\_IND\_1994\_2021\_2022\_08\_21\_1\_v3\_rus.dta**. Я сохранила его под другим именем:

```
save "C:\RLMS_work\seminar_6\data\ind_5_16.dta"
```

\*оставила в нем только 5-16 волны:

```
drop if id_w >= 17
```

\*отобрала только людей, заполнявших взрослую анкету

```
drop if child ==1
```

\*Оставила сравнительно небольшой набор переменных, относящихся в основном к разделу J

```
keep id_w idind year id_i id_h origsm inwgt region psu status popul int_y adult child marst  
occup08 educ diplom site h3 h4 h4_1 h5 h6 born_m age i4 j1 j2cod08 j4_1 j5a j5b j6 j6_0  
j6_1 j6_1a j6_1b j6_2 j7 j7_1 j8 j8_1 j8_2 j8_3 j9 j10 j11 j11_1 j11_2 j13 j13_1 j13_2 j14 j15  
j16 j17 j18 j18_1 j18_2 j19 j21 j21a j21b j23 j24 j25 j26 j27 j29 j308 j309 j29_0 j29_1 j29a  
j29b j32 j32_1 j33cod08 j35_1 j35_2y j35_2m j36 j36_1 j36_1a j36_1b j36_2 j37 j38 j39 j40  
j41 j41_1 j41_2 j313 j314 j43 j44 j45 j46 j47 j48 j48_1 j48_2 j49 j50 j51 j52 j53 j54 j55 j55_0  
j55_1 j55_2 j56 j57 j58 j59 j59_1 j60 j65 j66 j66_1 j69_9c j72_17 j322 j323m j323y j324 j325m  
j325y j72_18 j72_19 j72_171 j72_172 j72_173 j73 j75 j363 j76 j76_1 j76_2 j133_1 j133_2  
j133_3 j133_4 j133_5 j133_6 j133_7 j133_8 j133_9 j77 j78 j79 j80a j80b j80_1 j80_2c08 j85  
j86 j87 j88 j89 j326 j79a j79_1_1 j79_1_2 j79_1_3 j79_1_4 j79_1_5 j79_1_6 j79_1_7 j79_1_9
```

j791\_10 j791\_11 j791\_12 j791\_13 j161 j161\_1y j161\_1m j161\_2y j161\_2m j161\_3y j161\_3m  
 j81 j81\_1 j81\_2 j82 j83\_1 j83\_2 j83\_3 j83\_4 j83\_5 j83\_6 j83\_7 j83\_8 j84\_1 j90 j193\_1  
 j193\_2 j193\_3 j193\_4 j193\_5 j193\_6 j193\_7 j193\_8 j193\_9 j193\_10 j117\_01 j117\_02  
 j117\_03 j117\_04 j117\_05 j117\_06 j117\_07 j117\_08 j117\_09 j117\_10 j117\_11 j117\_12 j117\_13  
 j117\_14 j117\_15 j117\_16 j117\_17 j207\_1 j207\_2 j207\_3 j207\_5 j207\_6 j207\_7 j207\_8 j207\_9  
 j207\_10 j207\_11 j208\_1 j208\_2 j208\_3 j208\_4 j208\_5 j208\_6 j208\_7 m1 m2 m3

\*Сохранила файл

save "C:\RLMS\_work\seminar\_6\data\ind\_5\_16.dta", replace

tab id\_w

НОМЕР ВОЛНЫ	Freq.	Percent	Cum.
1994 год	8,893	7.38	7.38
1995 год	8,417	6.99	14.37
1996 год	8,341	6.93	21.30
1998 год	8,699	7.22	28.52
2000 год	9,074	7.53	36.06
2001 год	10,098	8.38	44.44
2002 год	10,499	8.72	53.16
2003 год	10,636	8.83	61.99
2004 год	10,662	8.85	70.84
2005 год	10,337	8.58	79.42
2006 год	12,490	10.37	89.80
2007 год	12,290	10.20	100.00
Total	120,436	100.00	

\* Приклеила файл идентификаторов родственников, который мы использовали на третьем занятии

merge 1:1 id\_w id\_i using "C:\RLMS\_work\seminar\_6\data\ind\_relatives\_5\_30\_short.dta"

(note: variable id\_w was byte, now double to accommodate using data's values)

(note: variable idind was long, now double to accommodate using data's values)

(label id\_w already defined)

Result	# of obs.
not matched	298,837
from master	0 (_merge==1)
from using	298,837 (_merge==2)
matched	120,436 (_merge==3)

drop if \_merge == 2

(298,837 observations deleted)

drop \_merge

\*Перекодировала миссинги (возможно, некоторые значения «Затрудняюсь ответить» имеют содержательный смысл, поэтому при необходимости их можно будет перекодировать опять).

```

recode marst occup08 educ diplom h5 h6 born_m age i4 j1 j2cod08 j4_1 j5a j5b j6 j6_0 j6_1
j6_1a j6_1b j6_2 j7 j7_1 j8 j8_1 j8_2 j8_3 j9 j10 j11 j11_1 j11_2 j13 j13_1 j13_2 j14 j15 j16
j17 j18 j18_1 j18_2 j19 j21 j21a j21b j23 j24 j25 j26 j27 j29 j308 j309 j29_0 j29_1 j29a j29b
j32 j32_1 j33cod08 j35_1 j35_2y j35_2m j36 j36_1 j36_1a j36_1b j36_2 j37 j38 j39 j40 j41
j41_1 j41_2 j313 j314 j43 j44 j45 j46 j47 j48 j48_1 j48_2 j49 j50 j51 j52 j53 j54 j55 j55_0
j55_1 j55_2 j56 j57 j58 j59 j59_1 j60 j65 j66 j66_1 j69_9c j72_17 j322 j323m j323y j324 j325m
j325y j72_18 j72_19 j72_171 j72_172 j72_173 j73 j75 j363 j76 j76_1 j76_2 j133_1 j133_2
j133_3 j133_4 j133_5 j133_6 j133_7 j133_8 j133_9 j77 j78 j79 j80a j80b j80_1 j80_2c08 j85
j86 j87 j88 j89 j326 j79a j79_1_1 j79_1_2 j79_1_3 j79_1_4 j79_1_5 j79_1_6 j79_1_7 j79_1_9
j791_10 j791_11 j791_12 j791_13 j161 j161_1y j161_1m j161_2y j161_2m j161_3y j161_3m
j81 j81_1 j81_2 j82 j83_1 j83_2 j83_3 j83_4 j83_5 j83_6 j83_7 j83_8 j84_1 j90 j193_1 j193_2
j193_3 j193_4 j193_5 j193_6 j193_7 j193_8 j193_9 j193_10 j117_01 j117_02 j117_03 j117_04
j117_05 j117_06 j117_07 j117_08 j117_09 j117_10 j117_11 j117_12 j117_13 j117_14 j117_15
j117_16 j117_17 j207_1 j207_2 j207_3 j207_5 j207_6 j207_7 j207_8 j207_9 j207_10 j207_11
j208_1 j208_2 j208_3 j208_4 j208_5 j208_6 j208_7 m1 m2 m3 (99999997 = .a) (99999998 =
.b) (99999999 = .c)

```

\*Перекодировала в «0» значение 2 для дихотомических переменных

```

recode j6 j7 j8_1 j8_3 j9 j11 j11_1 j14 j17 j18_2 j19 j21a j23 j24 j25 j26 j27 j29 j32 j32_1 j36
j37 j39 j41 j41_1 j44 j47 j48_2 j49 j50 j51 j52 j53 j55 j56 j59_1 j72_171 j73 j75 j76_1 j133_1
j133_2 j133_3 j133_4 j133_5 j133_6 j133_7 j133_8 j133_9 j77 j78 j85 j87 j88 j81 j82 j83_1
j83_2 j83_3 j83_4 j83_5 j83_6 j83_7 j83_8 j84_1 (2=0)

```

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

```

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

```

\* И присвоила этот лейбл всему сету переменных:

```

label values j6 j7 j8_1 j8_3 j9 j11 j11_1 j14 j17 j18_2 j19 j21a j23 j24 j25 j26 j27 j29 j32 j32_1
j36 j37 j39 j41 j41_1 j44 j47 j48_2 j49 j50 j51 j52 j53 j55 j56 j59_1 j72_171 j73 j75 j76_1
j133_1 j133_2 j133_3 j133_4 j133_5 j133_6 j133_7 j133_8 j133_9 j77 j78 j85 j87 j88 j81 j82
j83_1 j83_2 j83_3 j83_4 j83_5 j83_6 j83_7 j83_8 j84_1 YES

```

\*расчет дефлятора к уровню 16й волны

```

gen deflat_16 = 1 if id_w == 16
replace deflat_16 = 1.119 if id_w == 15
replace deflat_16 = 1.119 * 1.09 if id_w == 14
replace deflat_16 = 1.119 * 1.09*1.109 if id_w == 13
replace deflat_16 = 1.119 * 1.09*1.109*1.117 if id_w == 12
replace deflat_16 = 1.119 * 1.09*1.109*1.117*1.12 if id_w == 11
replace deflat_16 = 1.119 * 1.09*1.109*1.117*1.12*1.151 if id_w == 10
replace deflat_16 = 1.119 * 1.09*1.109*1.117*1.12*1.151*1.186 if id_w == 9
replace deflat_16 = 1.119 * 1.09*1.109*1.117*1.12*1.151*1.186*1.202*1.365 if id_w == 8
replace
    deflat_16
    =
    1.119
    *
    1.09*1.109*1.117*1.12*1.151*1.186*1.202*1.365*1.844*1.11/1000 if id_w == 7
replace deflat_16 = 1.119 * 1.09 * 1.109 * 1.117 * 1.12 * 1.151 * 1.186 * 1.202 * 1.365 * 1.844
* 1.11 * 1.218 / 1000 if id_w == 6
replace deflat_16 = 1.119 * 1.09 * 1.109 * 1.117 * 1.12 * 1.151 * 1.186 * 1.202 * 1.365 * 1.844
* 1.11 * 1.218 * 2.3/1000 if id_w == 5

```

**label variable deflat\_16 "дефлятор приведения цен к 16 волне по среднерос. ИПЦ (умножать на него)"**

\*расчет рабочего времени на первом месте работы с учетом работы дома (аналогично семинару 2)

```
summarize j8 j8_2 , d  
generate j8_2a = j8_2 if (j8_3 == 0 | j8_3 == .)  
egen j8_total = rowtotal(j8 j8_2a), missing  
summarize j8_total , d  
drop j8_2a  
label variable j8_total "полное рабочее время за 30 дней на первой работе"  
recode j8_total (420 / max = 420)  
sum j8_total, d
```

\*расчет «контрактной» заработной платы за 30 дней: деньги, которые человек получил за 30 дней, + сумму задолженности, деленную на кол-во месяцев задолженности; + сумму, которую можно выручить за проданную продукцию.

```
summarize j10 j15 j16 j18, d  
recode j16 (0 = .d)  
generate j15_16= j15/j16  
generate Wage_1 = .  
egen Wage_sup= rowtotal(j10 j15_16 j18), missing  
replace Wage_1 = Wage_sup  
replace Wage_1 = Wage_1 * deflat_16  
label variable Wage_1 "дефл.контрактная ЗП за 30 дней на первой работе"  
drop j15_16 Wage_sup  
summarize Wage_1 , d
```

\*расчет ставки ЗП, разделив заработок на время, а также логарифм ставки ЗП (понадобится для модели Минцера).

```
gen Hwage1=Wage_1/ j8_total  
replace Hwage1= Hwage1* deflat_16  
label variable Hwage1 "дефл.ставка ЗП за 30 дней на первой работе"  
sum Hwage1, d
```

\*расчет логарифма ставки заработной платы.

```
gen lg_Hwage1= ln(Hwage1)  
label variable lg_Hwage1 "дефл.логарифм ставки ЗП за 30 дней на первой работе"  
replace lg_Hwage1= ln(Hwage1)
```

## 24.1 НАЧАЛО СЕМИНАРА ЗДЕСЬ!!!!

\*Используем директорию на диске

**C:\RLMS\_work\seminar\_6\data**

Это позволит вам использовать готовые коды. Распакуйте архив с данными.

Откройте программу STATA

Начните с открытия файла аутпута, назвав его своей фамилией.

**log using "C:\RLMS\_work\seminar\_6\data\семинар 6 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\_6\data\ind\_5\_16.dta", clear**

\*И сохраните его под другим именем

**save "C:\RLMS\_work\seminar\_6\data\ind\_5\_16\_S6.dta"**

\*25. Факторный и кластерный анализ по переменным для какого-либо года.

\*ПРИМЕР: анализ мотивации труда (15 раунд).

25.1. Вопросы о важности разных мотивов в работе задавались также в 2009 и 2019 (?) годах.

Вопросы задавали всем, вне зависимости от того, работает человек или нет.

Вопрос: «Для Вас лично при выборе работы насколько важны такие ее стороны, как...».

Варианты:

**j79\_1\_1** возможность получать новые знания, умения

**j79\_1\_2** хорошая репутация предприятия, организации, фирмы

**j79\_1\_3** размер заработной платы

**j79\_1\_4** гарантированная работа, низкая вероятность ее потерять

**j79\_1\_5** удобный график работы

**j79\_1\_6** интересная, творческая работа

**j79\_1\_7** расположение предприятия, организации, фирмы

**j79\_1\_9** соответствие работы Вашим способностям, склонностям

**j791\_10** хорошие отношения в коллективе

**j791\_11** хорошее социальное обеспечение - путевки, дополнительное медицинское обслуживание, питание и т.п.

**j791\_12** возможности должностного роста

**j791\_13** условия труда

Так как фактически в файле обрезаны лейблы, воссоздадим их:

label variable j79\_1\_1 "Важность: возможность получать новые знания, умения"  
 label variable j79\_1\_2 "Важность: хорошая репутация предприятия"  
 label variable j79\_1\_3 "Важность: размер заработной платы"  
 label variable j79\_1\_4 "Важность: гарантированная работа"  
 label variable j79\_1\_5 "Важность: удобный график работы"  
 label variable j79\_1\_6 "Важность: интересная, творческая работа"  
 label variable j79\_1\_7 "Важность: расположение предприятия"  
 label variable j79\_1\_9 "Важность: соответствие работы Вашим способностям"  
 label variable j791\_10 "Важность: хорошие отношения в коллективе"  
 label variable j791\_11 "Важность: хорошее социальное обеспечение"  
 label variable j791\_12 "Важность: возможности должностного роста"  
 label variable j791\_13 "Важность: условия труда "

\*посмотрим распределение нужных переменных в 15й волне:

tab1 j79\_1\_1 j79\_1\_2 j79\_1\_3 j79\_1\_4 j79\_1\_5 j79\_1\_6 j79\_1\_7 j79\_1\_9 j791\_10 j791\_11  
 j791\_12 j791\_13 if id\_w == 15, missing

-> tabulation of j79\_1\_1 if id\_w == 15

Важность:   возможност ь получать   новые   знания,   умения	Freq.	Percent	Cum.
Очень важно	3,289	26.33	26.33
Скорее важно	4,680	37.47	63.80
И да, и нет	1,823	14.60	78.40
Скорее не важно	1,531	12.26	90.66
Совсем не важно	671	5.37	96.03
.a	316	2.53	98.56
.b	87	0.70	99.26
.c	93	0.74	100.00
Total	12,490	100.00	

-> tabulation of j79\_1\_2 if id\_w == 15

Важность:   хорошая   репутация   предприяти я	Freq.	Percent	Cum.
Очень важно	4,168	33.37	33.37
Скорее важно	5,727	45.85	79.22
И да, и нет	1,186	9.50	88.72
Скорее не важно	582	4.66	93.38
Совсем не важно	374	2.99	96.37
.a	265	2.12	98.49
.b	90	0.72	99.22
.c	98	0.78	100.00
Total	12,490	100.00	

-> tabulation of j79\_1\_3 if id\_w == 15

Важность:   размер   заработной   платы	Freq.	Percent	Cum.
Очень важно	8,832	70.71	70.71

Скорее важно	2,812	22.51	93.23
И да, и нет	207	1.66	94.88
Скорее не важно	87	0.70	95.58
Совсем не важно	164	1.31	96.89
.a	198	1.59	98.48
.b	91	0.73	99.21
.c	99	0.79	100.00
-----			
Total	12,490	100.00	

-> tabulation of j79\_1\_4 if id\_w == 15

Важность: гарантированная работа	Freq.	Percent	Cum.
-----			
Очень важно	6,686	53.53	53.53
Скорее важно	4,341	34.76	88.29
И да, и нет	592	4.74	93.03
Скорее не важно	224	1.79	94.82
Совсем не важно	209	1.67	96.49
.a	250	2.00	98.49
.b	92	0.74	99.23
.c	96	0.77	100.00
-----			
Total	12,490	100.00	

-> tabulation of j79\_1\_5 if id\_w == 15

Важность: удобный график работы	Freq.	Percent	Cum.
-----			
Очень важно	5,621	45.00	45.00
Скорее важно	4,736	37.92	82.92
И да, и нет	1,043	8.35	91.27
Скорее не важно	429	3.43	94.71
Совсем не важно	242	1.94	96.65
.a	221	1.77	98.41
.b	92	0.74	99.15
.c	106	0.85	100.00
-----			
Total	12,490	100.00	

-> tabulation of j79\_1\_6 if id\_w == 15

Важность: интересная, творческая работа	Freq.	Percent	Cum.
-----			
Очень важно	3,877	31.04	31.04
Скорее важно	4,489	35.94	66.98
И да, и нет	1,984	15.88	82.87
Скорее не важно	1,160	9.29	92.15
Совсем не важно	508	4.07	96.22
.a	274	2.19	98.41
.b	91	0.73	99.14
.c	107	0.86	100.00
-----			
Total	12,490	100.00	

-> tabulation of j79\_1\_7 if id\_w == 15

Важность: расположение предприятия	Freq.	Percent	Cum.
--	-------	---------	------

	Freq.	Percent	Cum.
Очень важно	3,656	29.27	29.27
Скорее важно	5,009	40.10	69.38
И да, и нет	1,994	15.96	85.34
Скорее не важно	1,047	8.38	93.72
Совсем не важно	357	2.86	96.58
.a	237	1.90	98.48
.b	91	0.73	99.21
.c	99	0.79	100.00
Total	12,490	100.00	

-> tabulation of j79\_1\_9 if id\_w == 15

Важность: соответств ие работы Вашим способност ям	Freq.	Percent	Cum.
Очень важно	4,415	35.35	35.35
Скорее важно	5,621	45.00	80.35
И да, и нет	1,217	9.74	90.10
Скорее не важно	469	3.76	93.85
Совсем не важно	292	2.34	96.19
.a	285	2.28	98.47
.b	92	0.74	99.21
.c	99	0.79	100.00
Total	12,490	100.00	

-> tabulation of j791\_10 if id\_w == 15

Важность: хорошие отношения в коллективе	Freq.	Percent	Cum.
Очень важно	5,697	45.61	45.61
Скорее важно	5,388	43.14	88.75
И да, и нет	628	5.03	93.78
Скорее не важно	155	1.24	95.02
Совсем не важно	201	1.61	96.63
.a	214	1.71	98.34
.b	92	0.74	99.08
.c	115	0.92	100.00
Total	12,490	100.00	

-> tabulation of j791\_11 if id\_w == 15

Важность: хорошее социальное обеспечени е	Freq.	Percent	Cum.
Очень важно	4,262	34.12	34.12
Скорее важно	5,061	40.52	74.64
И да, и нет	1,646	13.18	87.82
Скорее не важно	747	5.98	93.80
Совсем не важно	326	2.61	96.41
.a	258	2.07	98.48
.b	92	0.74	99.22
.c	98	0.78	100.00
Total	12,490	100.00	

-> tabulation of j791\_12 if id\_w == 15

Важность : возможност и должностно го роста	Freq.	Percent	Cum.
Очень важно	3,576	28.63	28.63
Скорее важно	4,170	33.39	62.02
И да, и нет	2,040	16.33	78.35
Скорее не важно	1,467	11.75	90.10
Совсем не важно	721	5.77	95.87
.a	319	2.55	98.42
.b	92	0.74	99.16
.c	105	0.84	100.00
Total	12,490	100.00	

-> tabulation of j791\_13 if id\_w == 15

Важность : условия труда	Freq.	Percent	Cum.
Очень важно	5,578	44.66	44.66
Скорее важно	5,513	44.14	88.80
И да, и нет	612	4.90	93.70
Скорее не важно	145	1.16	94.86
Совсем не важно	205	1.64	96.50
.a	230	1.84	98.34
.b	92	0.74	99.08
.c	115	0.92	100.00
Total	12,490	100.00	

Можно увидеть, что в во всех вопросах количество «затруднившихся» составляет около 2% (градация .a). Можно этим пренебречь, тогда мы потеряем некоторое количество респондентов (причем больше 2%); или предположить, что «затрудняюсь» это близко к ответу «и да, и нет», и сделать соответствующую перекодировку. Или выделить тех, кто затруднился хотя бы с одним из ответов, в отдельную группу – это на усмотрение исследователя.

В этом блоке переменных две «проблемы»:

1) шкала «перевернута» (1 – очень важно, 5 – совсем не важно, а лучше бы наоборот), поэтому ее нужно изменить.

2) У ответов «совсем не важно», «скорее не важно», и отчасти «и да, и нет» очень маленькая доля ответивших – то есть люди плохо различают эти градации, и скорее отвечают в формате «очень важно» или «скорее важно», и лишь немногие выбирают меньшую важность. Если работать с переменными без изменений, результат классификации получается не очень хороший. Поэтому имеет смысл «схлопнуть» ответы 3-4-5 в одну градацию, «не очень важно».

\* перекодирование в другие переменные

```
recode j79_1_1 j79_1_2 j79_1_3 j79_1_4 j79_1_5 j79_1_6 j79_1_7 j79_1_9 j791_10 j791_11
j791_12 j791_13 (2=2)(1=3) (3 / 5=1) , into (j79_1_1a j79_1_2a j79_1_3a j79_1_4a j79_1_5a
j79_1_6a j79_1_7a j79_1_9a j791_10a j791_11a j791_12a j791_13a)
```

```
label variable j79_1_1a "Важность: возможность получать новые знания, умения"
```

```
label variable j79_1_2a "Важность: хорошая репутация предприятия"
```

```
label variable j79_1_3a "Важность: размер заработной платы"
```

```
label variable j79_1_4a "Важность: гарантированная работа"
```

```
label variable j79_1_5a "Важность: удобный график работы"
```

```

label variable j79_1_6a "Важность: интересная, творческая работа"
label variable j79_1_7a "Важность: расположение предприятия"
label variable j79_1_9a "Важность: соответствие работы Вашим способностям"
label variable j791_10a "Важность: хорошие отношения в коллективе"
label variable j791_11a "Важность: хорошее социальное обеспечение"
label variable j791_12a "Важность: возможности должностного роста"
label variable j791_13a "Важность: условия труда "

label define LABOUR_MOTIVE 1 "не очень важно" 2 "важно" 3 "очень важно"
label values j79_1_1a j79_1_2a j79_1_3a j79_1_4a j79_1_5a j79_1_6a j79_1_7a j79_1_9a
j791_10a j791_11a j791_12a j791_13a LABOUR_MOTIVE

```

\*25.2. Факторный анализ мотивации труда для населения в возрасте 18-70 лет.  
 При представлении в эссе, дипломе, статье результатов факторного и кластерного анализа обязательно приводить также распределение ответов на блок исходных вопросов, и проинтерпретировать полученные факторы.

```

tab1 j79_1_1a j79_1_2a j79_1_3a j79_1_4a j79_1_5a j79_1_6a j79_1_7a j79_1_9a j791_10a
j791_11a j791_12a j791_13a if id_w == 15, missing

```

\*Факторный анализ с вращением варимакс. Сохранить факторы в новые переменные.  
 \*Опции: метод извлечения - **pcf** – principal-component factor (метод главных компонент)  
 \***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).

```

factor j79_1_1a j79_1_2a j79_1_3a j79_1_4a j79_1_5a j79_1_6a j79_1_7a j79_1_9a j791_10a
j791_11a j791_12a j791_13a if (id_w == 15 & age >=18 & age <=70), pcf mineigen(1)

```

(obs=10,014)

```

Factor analysis/correlation
Method: principal-component factors
Rotation: (unrotated)
Number of obs = 10,014
Retained factors = 2
Number of params = 23

```

Factor	Eigenvalue	Difference	Proportion	Cumulative
Factor1	5.80274	4.57231	0.4836	0.4836
Factor2	1.23043	0.38421	0.1025	0.5861
Factor3	0.84622	0.19168	0.0705	0.6566
Factor4	0.65455	0.04381	0.0545	0.7112
Factor5	0.61074	0.11299	0.0509	0.7621
Factor6	0.49774	0.03033	0.0415	0.8035
Factor7	0.46742	0.01151	0.0390	0.8425
Factor8	0.45591	0.05605	0.0380	0.8805
Factor9	0.39986	0.03552	0.0333	0.9138
Factor10	0.36434	0.01897	0.0304	0.9442
Factor11	0.34537	0.02069	0.0288	0.9729
Factor12	0.32468	.	0.0271	1.0000

LR test: independent vs. saturated: chi2(66) = 5.5e+04 Prob>chi2 = 0.0000

Factor loadings (pattern matrix) and unique variances

Variable	Factor1	Factor2	Uniqueness
j79_1_1a	0.6687	-0.5028	0.3000
j79_1_2a	0.7234	-0.1931	0.4394
j79_1_3a	0.5443	0.5559	0.3947



j79\_1\_7a "Важность: расположение предприятия" (входит в оба фактора)

j791\_10a "Важность: хорошие отношения в коллективе"

j791\_11a "Важность: хорошее социальное обеспечение"

j791\_13a "Важность: условия труда "

ВАЖНО: факторы представляют из себя **векторы**, количественные переменные, которые принимают некоторое значение для каждого индивида. Так как факторы ортогональны (независимы, корреляции между ними нет), то каждый человек может иметь любую комбинацию значений каждого из них. То есть могут быть люди, у которых оба фактора принимают высокие значения (высоко мотивированные люди); у которых оба фактора имеют низкие значения; или какой-то один из факторов имеет высокое значение, а другой – низкое значение. Таким образом, скорее всего, можно будет выделить 4 группы людей (кластера).

\*Сохраним факторы как новые переменные.

**predict F1\_labmotiv F2\_labmotiv**

**label variable F1\_labmotiv `Scores for factor 1 Мотивация труда: "на себя"``**

**label variable F2\_labmotiv `Scores for factor 2 Мотивация труда: "на окружение"``**

(regression scoring assumed)

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

Variable	Factor1	Factor2
j79_1_1a	0.35864	-0.22720
j79_1_2a	0.19761	-0.03358
j79_1_3a	-0.23185	0.39893
j79_1_4a	-0.16378	0.34326
j79_1_5a	-0.05127	0.23832
j79_1_6a	0.29821	-0.14212
j79_1_7a	0.07818	0.09455
j79_1_9a	0.18511	-0.01488
j791_10a	0.02036	0.16465
j791_11a	0.02430	0.15019
j791_12a	0.27637	-0.12556
j791_13a	-0.01786	0.21416

\*25.3. Кластерный анализ мотивации труда.

Попробуем разделить индивидов на 4 группы в соответствии с предсказанными нами вариантами сочетаний высокого и низкого значений каждого из факторов. НО: кластерный анализ будем делать на исходных переменных, а НЕ на выделенных факторах. Это есть как минимум две причины:

- 1) полученные факторы имеют нормальное распределение, то есть сконцентрированы около значения 0 и «размазаны» по краям. Это не дает хорошей картинки для кластеризации (выделения групп кейсов).
- 2) два фактора объясняют 58% дисперсии, т.е. при использовании факторов мы теряем 42% информации.

Команда из меню:

*Statistics > Multivariate analysis > Cluster analysis > Cluster data > Kmeans*

Опции команды

**cluster kmeans [varlist] [if] [in] , k(#) [ options ]**

**cluster kmeans**

k(#) perform cluster analysis resulting in # groups

measure(measure) similarity or dissimilarity measure; default is L2 (Euclidean)

name(cname)	name of resulting cluster analysis
<b>Options</b>	
start(start_option)	obtain k initial group centers by using start_option
keepcenters	append the k final group means or medians to the data
<b>Advanced</b>	
generate(groupvar)	name of grouping variable

measure	Description
cont_measure	similarity or dissimilarity measure for <b>continuous</b> data
binary_measure	similarity measure for <b>binary</b> data
mixed_measure	dissimilarity measure for <b>a mix</b> of binary and continuous data

cont_measure	Description
L2	<b>Euclidean distance</b> (Minkowski with argument 2)
Euclidean	alias for L2
L(2)	alias for L2
L2squared	<b>squared Euclidean distance</b>
Lpower(2)	alias for L2squared
L1	absolute-value distance (Minkowski with argument 1)
absolute	alias for L1
cityblock	alias for L1
manhattan	alias for L1
L(1)	alias for L1
Lpower(1)	alias for L1
Linfinity	maximum-value distance (Minkowski with infinite argument)
maximum	alias for Linfinity
L(#)	Minkowski distance with # argument
Lpower(#)	Minkowski distance with # argument raised to # power
Canberra	Canberra distance
correlation	correlation coefficient similarity measure
angular	angular separation similarity measure
angle	alias for angular

binary_measure	Description
matching	simple matching similarity coefficient
Jaccard	Jaccard binary similarity coefficient
Russell	Russell and Rao similarity coefficient
Hamann	Hamann similarity coefficient
Dice	Dice similarity coefficient
antiDice	anti-Dice similarity coefficient
Sneath	Sneath and Sokal similarity coefficient
Rogers	Rogers and Tanimoto similarity coefficient
Ochiai	Ochiai similarity coefficient
Yule	Yule similarity coefficient
Anderberg	Anderberg similarity coefficient
Kulczynski	Kulczyński similarity coefficient
Pearson	Pearson's phi similarity coefficient
Gower2	similarity coefficient with same denominator as Pearson

mixed_measure	Description
Gower	Gower's dissimilarity coefficient

**start(start\_option)** indicates how the k initial group centers are to be obtained. The available start\_options are

**krandom(seed#)**, the default, specifies that k unique observations be chosen at random, from among those to be clustered, as starting centers for the k groups. Optionally, a random-number seed may be specified to cause the command set seed seed# (see [R] set seed) to be applied before the k random observations are chosen.

**firstk[, exclude]** specifies that the first k observations from among those to be clustered be used as the starting centers for the k groups. With the exclude option, these first k observations are not included among the observations to be clustered.

`lastk[, exclude]` specifies that the last k observations from among those to be clustered be used as the starting centers for the k groups. With the `exclude` option, these last k observations are then not included among the observations to be clustered.

`random[(seed#)]` specifies that k random initial group centers be generated. The values are randomly chosen from a uniform distribution over the range of the data. Optionally, a random-number seed may be specified to cause the command set seed `seed#` (see [R] set seed) to be applied before the k group centers are generated.

`prandom[(seed#)]` specifies that k partitions be formed randomly among the observations to be clustered. The group means or medians from the k groups defined by this partitioning are to be used as the starting group centers. Optionally, a random-number seed may be specified to cause the command set seed `seed#` (see [R] set seed) to be applied before the k partitions are chosen.

`everykth` specifies that k partitions be formed by assigning observations 1, 1+k, 1+2k, ... to the first group; assigning observations 2, 2+k, 2+2k, ... to the second group; and so on, to form k groups. The group means or medians from these k groups are to be used as the starting group centers.

`segments` specifies that k nearly equal partitions be formed from the data. Approximately the first N/k observations are assigned to the first group, the second N/k observations are assigned to the second group, and so on. The group means or medians from these k groups are to be used as the starting group centers.

`group(varname)` provides an initial grouping variable, `varname`, that defines k groups among the observations to be clustered. The group means or medians from these k groups are to be used as the starting group centers.

`keepcenters` specifies that the group means or medians from the k groups that are produced are to be appended to the data.

```
cluster kmeans j79_1_1a j79_1_2a j79_1_3a j79_1_4a j79_1_5a j79_1_6a j79_1_7a j79_1_9a
j791_10a j791_11a j791_12a j791_13a if (id_w == 15 & age >=18 & age <=70), k(4)
measure(L2) name(Cluster_motive1) start(krandom) generate(Cl_motive1) keepcenters
```

\*КОЛИЧЕСТВО КЕЙСОВ В КЛАСТЕРАХ

**tab Cl\_motive1**

Cl_motive1	Freq.	Percent	Cum.
1	2,648	26.44	26.44
2	2,671	26.67	53.12
3	2,314	23.11	76.22
4	2,381	23.78	100.00
Total	10,014	100.00	

\*кластерные центры (средние значения исходных переменных в кластере).

```
tabstat j79_1_1a j79_1_2a j79_1_3a j79_1_4a j79_1_5a j79_1_6a j79_1_7a j79_1_9a
j791_10a j791_11a j791_12a j791_13a if (id_w == 15 & age >=18 & age <=70),
statistics(mean) by(Cl_motive1)
```

```
Summary statistics: mean
by categories of: Cl_motive1
```

Cl_motive1	j79_1_1a	j79_1_2a	j79_1_3a	j79_1_4a	j79_1_5a	j79_1_6a	j79_1_7a
1	2.143505	2.155967	2.510574	2.207326	1.990937	2.072885	1.702039
2	2.719955	2.82441	2.969674	2.901535	2.900412	2.858854	2.739423
3	1.55013	2.124028	2.940363	2.790838	2.722126	1.812014	2.224287
4	1.178916	1.49895	2.432591	2.037799	1.723226	1.181436	1.391012
Total	1.930797	2.170661	2.713801	2.487018	2.338826	2.010286	2.025464

Cl_motive1	j79_1_9a	j791_10a	j791_11a	j791_12a	j791_13a
1	2.094411	2.13935	1.80287	1.971677	2.102719
2	2.875328	2.921752	2.783227	2.825908	2.938225
3	2.236387	2.638721	2.328436	1.628349	2.656439
4	1.551029	1.863503	1.529189	1.183956	1.847123
Total	2.206311	2.397843	2.120731	1.932894	2.39275

Фактически, у нас получилось, что второй кластер – высокая мотивация по всем переменным, четвертый – низкая (относительно) по всем переменным, хотелось бы 1-й и 3-й кластеры сделать более специфическими, чтобы один был ориентирован на высокие значения переменных из первой группы переменных (вошедших в первый кластер), а другой – на другую группу.

*Первый фактор* (интерпретация: направленность «на себя» - по Ядову):

j79\_1\_1a "Важность: возможность получать новые знания, умения"

j79\_1\_2a "Важность: хорошая репутация предприятия"

j79\_1\_6a "Важность: интересная, творческая работа"

j79\_1\_9a "Важность: соответствие работы Вашим способностям"

j791\_12a "Важность: возможности должностного роста"

*Второй фактор* (интерпретация: направленность «на окружение»)

j79\_1\_3a "Важность: размер заработной платы"

j79\_1\_4a "Важность: гарантированная работа"

j79\_1\_5a "Важность: удобный график работы"

j79\_1\_7a "Важность: расположение предприятия"

j791\_10a "Важность: хорошие отношения в коллективе"

j791\_11a "Важность: хорошее социальное обеспечение"

j791\_13a "Важность: условия труда "

Так как у нас сохранены 4 новый кейса, которые мы можем использовать как новые кластерные центры, мы можем самостоятельно задать их значения. Так как заработная плата фактически важна для всех, оставим ее значение без изменения (мотив 3). В первом кейсе сделаем значения мотивов 1, 2, 6, 9, 12 равными трем (максимальная важность), а остальные, кроме 3го (ЗП), равными 1. В третьем – наоборот, 1, 2, 6, 9, 12 – равными 1 (минимальная важность), остальные, кроме заработной платы, равными 3. Во втором все значения сделаем равными 3, а в четвертом – все 1, кроме заработной платы.

Получим серию команд примерно такого вида (*ее выполнять не нужно*):

**\*replace j79\_1\_1a = 3 in 120437**

\* теперь пересчитаем кластеры, используя последние 4 вспомогательных кейса как начальные кластерные центры

```
cluster kmeans j79_1_1a j79_1_2a j79_1_3a j79_1_4a j79_1_5a j79_1_6a j79_1_7a j79_1_9a
j791_10a j791_11a j791_12a j791_13a if (id_w == 15 & age >=18 & age <=70), k(4)
measure(L2) name(Cluster_motive2) start(lastk, exclude) generate(Cl_motive2)
```

tab **Cl\_motive2**

Cl_motive2	Freq.	Percent	Cum.
1	2,689	26.86	26.86
2	2,659	26.56	53.43
3	2,293	22.91	76.33
4	2,369	23.67	100.00
Total	10,010	100.00	

Распределение по-прежнему равномерное.

tabstat j79\_1\_1a j79\_1\_2a j79\_1\_3a j79\_1\_4a j79\_1\_5a j79\_1\_6a j79\_1\_7a j79\_1\_9a  
j791\_10a j791\_11a j791\_12a j791\_13a if (id\_w == 15 & age >=18 & age <=70),  
statistics(mean) by(**Cl\_motive2**)

Summary statistics: mean  
by categories of: Cl\_motive2

Cl_motive2	j79_1_1a	j79_1_2a	j79_1_3a	j79_1_4a	j79_1_5a	j79_1_6a	j79_1_7a
1	2.714392	2.823354	2.969877	2.901078	2.900335	2.851246	2.736333
2	2.142911	2.156074	2.512599	2.210982	1.993983	2.076344	1.702896
3	1.541212	2.117314	2.939381	2.787614	2.721326	1.802878	2.223288
4	1.179823	1.497678	2.430139	2.03588	1.718024	1.182355	1.389194
Total	1.930669	2.170629	2.713686	2.487013	2.338761	2.01029	2.025475

Cl_motive2	j79_1_9a	j791_10a	j791_11a	j791_12a	j791_13a
1	2.874303	2.921904	2.783563	2.823726	2.938267
2	2.092892	2.13915	1.805942	1.97217	2.10267
3	2.232883	2.636284	2.320105	1.615351	2.656345
4	1.549599	1.862389	1.528071	1.1832	1.844238
Total	2.206294	2.397802	2.120579	1.932468	2.392807

Получилось не совсем так, как ожидалось, теперь первый кластер – самый мотивированный, а четвертый – самый немотивированный. Второй и третий лучше разведены по интересующим нас переменным, но не идеально (например, по мотивам 2 и 9). Можно корректировать дальше начальные центры и искать хорошие решения. Мы пока остановимся на этом. Создадим лейблы. Удалим вспомогательные кейсы: их можно выделить в таблице данных и нажать «delete», появится код:

```
label variable Cl_motive2 "Кластеры по мотивации труда"  
label define CLUSTER_MOTIVE 1 "высоко мотивированные" 2 "ориентация на себя"  
3 "ориентация на среду" 4 "низко мотивированные"  
label values Cl_motive2 CLUSTER_MOTIVE  
drop in 120437/120440
```

Сохраните файл.

\*25.4. Характеристики кластеров.

\*Для того, чтобы понимать, какие именно люди входят в полученные кластеры, мы можем посмотреть на различия между ними.

\*перекодируем переменную профессиональных статусов в меньшее количество градаций

(примерно как мы это делали в самостоятельном задании к занятию 3 с учетом, что у нас только взрослые от 14 лет).

```

tab occup08, missing
recode occup08 (1=2) (2=3) (3=4) (4=4) (5=4) (6=5) (7=5) (8=5) (9=6) (0=7) (.a=.a) (.b=.b)
(.c=.c), generate(occup08_k)
label variable occup08_k "ПРОФЕССИОНАЛЬНАЯ ГРУППА - по коду J2COD08
кратко"
recode occup08_k (. = 1) if j1 ==5
label define OCCUP08_k 1 "не работает" 2 "руководители" 3 "специалисты" 4
"служащие" 5 "квалиф.рабочие" 6 "неквал.рабочие" 7 "военные", replace
label values occup08_k OCCUP08_k
tab occup08_k, missing

```

ПРОФЕССИОНАЛЬНАЯ ГРУППА - по коду J2COD08	Freq.	Percent	Cum.
военнослужащие	515	0.43	0.43
законодатели; крупные чиновники; руково	3,901	3.24	3.67
специалисты высшего уровня квалификации	10,773	8.94	12.61
специалисты среднего уровня квалификаци	9,769	8.11	20.72
служащие офисные и по обслуживанию клие	3,389	2.81	23.54
работники сферы торговли и услуг	9,615	7.98	31.52
квалифицированные работники сельского,	348	0.29	31.81
квалифицированные рабочие, занятые ручн	9,555	7.93	39.74
квалифицированные рабочие, использующие	10,112	8.40	48.14
неквалифицированные рабочие всех отрасл	4,996	4.15	52.29
.	57,307	47.58	99.87
.a	31	0.03	99.90
.b	98	0.08	99.98
.c	27	0.02	100.00
Total	120,436	100.00	

ПРОФЕСС ИОНАЛЬН АЯ ГРУППА - по коду J2COD08 кратко	Freq.	Percent	Cum.
не работает	57,235	47.52	47.52
руководители	3,901	3.24	50.76
специалисты	10,773	8.94	59.71
служащие	22,773	18.91	78.62
квалиф.рабочие	20,015	16.62	95.23
неквал.рабочие	4,996	4.15	99.38
военные	515	0.43	99.81
.	72	0.06	99.87
.a	31	0.03	99.90
.b	98	0.08	99.98
.c	27	0.02	100.00
Total	120,436	100.00	

\*Посмотрим кросс-таблицу полученных кластеров и профессионального статуса (есть значимая связь). В группе высоко мотивированных выше, чем в среднем, доля специалистов; в группе ориентированных на себя – выше доля руководителей и специалистов; в группе ориентированных на среду – служащих и рабочих; в группе низкомотивированных – незанятых и рабочих.

```

tabulate occup08_k Cl_motive2 if id_w == 15 & age >=18 & age <=70, chi2 column nofreq

```

ПРОФЕССИОНАЛЬНАЯ ГРУППА - по коду J2COD08 кратко	Кластеры по мотивации труда				Total
	высоко мотивации	ориентации на среду	ориентации на работу	низко мотивации	
не работает	36.00	33.71	29.53	40.50	34.97
руководители	4.88	6.25	2.14	2.96	4.16
специалисты	14.26	14.45	9.23	4.95	10.96
служащие	27.14	25.17	28.26	19.93	25.17
квалиф. рабочие	14.45	16.44	23.93	24.55	19.53
неквал. рабочие	2.79	3.35	6.56	6.77	4.74
военные	0.48	0.64	0.35	0.34	0.46
Total	100.00	100.00	100.00	100.00	100.00

Pearson chi2(18) = 456.1724 Pr = 0.000

\*Посмотрим кросс-таблицу полученных кластеров и пола (есть значимая связь): доля женщин больше, чем в среднем по выборке, среди высокомотивированных, и ориентированных на среду.

**tabulate h5 Cl\_motive2 if id\_w == 15 & age >=18 & age <=70, chi2 column nofreq**

Пол респондента	Кластеры по мотивации труда				Total
	высоко мотивации	ориентации на среду	ориентации на работу	низко мотивации	
МУЖСКОЙ	39.20	46.30	38.12	52.13	43.90
ЖЕНСКИЙ	60.80	53.70	61.88	47.87	56.10
Total	100.00	100.00	100.00	100.00	100.00

Pearson chi2(3) = 126.6796 Pr = 0.000

\*Сравним средний возраст, среднюю ставку заработной платы и среднюю заработную плату за месяц на первой работе (это только для занятых!!!) по кластерам

**oneway age Cl\_motive2 if id\_w == 15 & age >=18 & age <=70, tabulate**

Кластеры по мотивации труда	Summary of Количество полных лет		
	Mean	Std. Dev.	Freq.
высоко мотивации	37.632205	14.698524	2,689
ориентации на среду	38.242948	14.806134	2,659
ориентации на работу	43.195377	13.755444	2,293
низко мотивации	44.366821	14.134702	2,369
Total	40.662637	14.68243	10,010

Analysis of Variance					
Source	SS	df	MS	F	Prob > F
Between groups	87476.7169	3	29158.9056	140.94	0.0000
Within groups	2070201.01	10006	206.895963		
Total	2157677.73	10009	215.573756		

Bartlett's test for equal variances: chi2(3) = 17.3999 Prob>chi2 = 0.001

Высоко мотивированные – самые молодые, и наоборот. Различия значимы.

**oneway Hwage1 CI\_motive2 if id\_w == 15 & age >=18 & age <=70, tabulate**

Класт еры по мотив ации труда	Summary of ставка ЗП за 30 дней на первой работе	Mean	Std. Dev.	Freq.
высоко мо		67.76024	149.48414	1,389
ориентаци		102.94908	841.52075	1,422
ориентаци		56.112374	73.412303	1,385
низко мот		61.189204	281.73067	1,177
Total		72.631287	460.66002	5,373

Source	SS	df	MS	F	Prob > F
Between groups	1872040.57	3	624013.522	2.94	0.0317
Within groups	1.1381e+09	5369	211977.549		
Total	1.1400e+09	5372	212207.651		

Bartlett's test for equal variances:  $\chi^2(3) = 7.6e+03$  Prob> $\chi^2 = 0.000$

Для ставки заработной платы тест показывает значимость различий средних, но значимость чуть хуже (0.03). Самая высокая ставка заработной платы у ориентированных «на себя», самая низкая – у ориентированных «на среду».

Однофакторный дисперсионный анализ (ANOVA) подразумевает, что средние генеральных совокупностей, из которых были извлечены выборки равны, другими словами все они относятся к одной генеральной совокупности и различия носят случайный характер. Для проверки теорий в случае дисперсионного анализа используется F-распределение. F-статистика принимает только положительные или нулевые значения.

Процедура дисперсионного анализа состоит в определении соотношения систематической (межгрупповой) дисперсии к случайной (внутригрупповой) дисперсии в измеряемых данных. В качестве показателя изменчивости используется сумма квадратов отклонения значений параметра от среднего. В случае если верна  $H_0$ , то как внутригрупповая, так и межгрупповая дисперсии служат оценками одной и той же дисперсии и должны быть приблизительно равны. Исходя из этого значение F должно быть близко к 1 в случае, если статистически значимых различий все-таки нет. Критическое значение F определяется уровнем значимости (обычно 0,05 или 0,01) и внутригрупповым и межгрупповым числом степеней свободы ( $\nu$ ).

Критерий Бартлетта (англ. Bartlett's test) — статистический критерий, позволяющий проверять равенство дисперсий нескольких (двух и более) выборок. Нулевая гипотеза предполагает, что рассматриваемые выборки получены из генеральных совокупностей, обладающих одинаковыми дисперсиями.

Однако, при обнаружении статистически значимых отличий мы не сможем сказать лишь об их наличии, но какие именно группы отличаются друг от друга мы определить не сможем, для этого производят так называемые процедуры апостериорных сравнений. Апостериорные сравнения представляют собой попарные сравнения изучаемых групп для обнаружения различий между ними. Апостериорные сравнения могут быть проведены с помощью критерия Стьюдента для независимых выборок.

**oneway Wage\_1 CI\_motive2 if id\_w == 15 & age >=18 & age <=70, tabulate**

Классификация по мотивации труда	Mean	Std. Dev.	Freq.
высоко мотивации	9841.6313	8897.4878	1,545
ориентации	10838.027	9813.1934	1,574
ориентации	8452.0705	7159.8686	1,507
низко мотивации	8700.0986	7936.652	1,289
Total	9503.9852	8601.2036	5,915

Source	Analysis of Variance			F	Prob > F
	SS	df	MS		
Between groups	5.4779e+09	3	1.8260e+09	24.98	0.0000
Within groups	4.3204e+11	5911	73091527.1		
Total	4.3752e+11	5914	73980703.2		

Bartlett's test for equal variances:  $\chi^2(3) = 168.7434$  Prob> $\chi^2 = 0.000$

Для заработной платы за месяц различия средних также значимы.

### 25.5. Самостоятельное задание.

Факторный и кластерный анализ на какой-либо группе других переменных (по выбору) для одного года (по выбору, или для которого есть данные). Для построенных кластеров посмотреть социально-демографические характеристики – 1-2 номинальных переменных, 1-2 количественных переменных).

- 1) j208\_1 j208\_2 j208\_3 j208\_4 j208\_5 j208\_6 j208\_7 – важность демократических институтов
- 2) j207\_1 j207\_2 j207\_3 j207\_5 j207\_6 j207\_7 j207\_8 j207\_9 j207\_10 j207\_11 – степень институционального доверия
- 3) j117\_01 j117\_02 j117\_03 j117\_04 j117\_05 j117\_06 j117\_07 j117\_08 j117\_09 j117\_10 j117\_11 j117\_12 j117\_13 j117\_14 j117\_15 j117\_16 j117\_17 - степень согласия индивида с высказываниями (самооценка, например «Я не могу справиться со своими проблемами») (только 2003, 2004, 2005 годы).
- 4) j193\_1 j193\_2 j193\_3 j193\_4 j193\_5 j193\_6 j193\_7 j193\_8 j193\_9 j193\_10 – вопрос «о ком вы бы могли сказать: «это мы»» (идентификация)
- 5) j83\_1 j83\_2 j83\_3 j83\_4 j83\_5 j83\_6 j83\_7 j83\_8 – каналы поиска работы (только для ищущих работу)
- 6) j133\_1 j133\_2 j133\_3 j133\_4 j133\_5 j133\_6 j133\_7 j133\_8 j133\_9 - ответы на вопрос «на какие средства вы собираетесь жить, когда достигнете пенсионного возраста» (с 2003 г.)

### 26. Матрицы перехода по профессиональному статусу

\*26.1. Создадим переменную «есть работа», при этом учтем, что человек, ответивший, что у него нет работы (переменная j1), может иметь приработки (j56). Нет определенности, считать ли такого человека занятым. Поэтому укажем отдельно эти варианты.

вопросы:

- \* j1 - Ваше основное занятие в настоящее время?
- \* j32 - У Вас есть еще какая-нибудь работа?
- \* j56 - В течение последних 30 дней Вы занимались (еще) какой-нибудь работой, за которую Вам заплатили или должны заплатить?
- \* j77 - Есть ли работа у респондента?

Новая переменная: **labour** - наличие у человека работы (даже если он в отпуске или декрете) или приработков; посмотрим исходные распределения.

**tab id\_w j1**

НОМЕР ВОЛНЫ	Ваше основное занятие в настоящее время. Вы					Total
	Вы сейчас	Вы находи	Вы находи	Вы находи	Или у Вас	
1994 год	4,616	181	55	42	3,993	8,887
1995 год	4,314	173	51	33	3,838	8,409
1996 год	4,156	147	42	37	3,943	8,325
1998 год	4,034	125	51	37	4,442	8,689
2000 год	4,337	124	36	11	4,565	9,073
2001 год	4,871	140	61	17	5,007	10,096
2002 год	5,102	154	55	13	5,173	10,497
2003 год	5,282	157	50	17	5,128	10,634
2004 год	5,339	166	47	12	5,094	10,658
2005 год	5,245	147	25	9	4,906	10,332
2006 год	6,547	184	37	7	5,706	12,481
2007 год	6,589	206	40	8	5,440	12,283
Total	60,432	1,904	550	243	57,235	120,364

**tab id\_w j32**

НОМЕР ВОЛНЫ	Скажите, пожалуйста, у Вас есть еще какая-ни		Total
	нет	да	
1994 год	4,661	224	4,885
1995 год	4,364	194	4,558
1996 год	4,191	183	4,374
1998 год	4,042	194	4,236
2000 год	4,273	224	4,497
2001 год	4,813	269	5,082
2002 год	5,067	252	5,319
2003 год	5,227	272	5,499
2004 год	5,302	257	5,559
2005 год	5,184	232	5,416
2006 год	6,433	336	6,769
2007 год	6,516	316	6,832
Total	60,073	2,953	63,026

**tab id\_w j56**

НОМЕР ВОЛНЫ	В течение последних 30 дней Вы занимались (ещ		Total
	нет	да	
1994 год	8,218	657	8,875
1995 год	7,806	590	8,396
1996 год	7,788	527	8,315
1998 год	8,078	603	8,681
2000 год	8,290	773	9,063
2001 год	9,197	878	10,075
2002 год	9,676	810	10,486
2003 год	9,764	867	10,631
2004 год	9,829	826	10,655
2005 год	9,625	699	10,324
2006 год	11,577	894	12,471
2007 год	11,543	747	12,290
Total	111,391	8,871	120,262

**tab id\_w j77**

НОМЕР ВОЛНЫ	Есть ли работа у респондент а?		Total
	нет	да	
1994 год	3,999	4,894	8,893
1995 год	3,846	4,571	8,417
1996 год	3,959	4,382	8,341
1998 год	4,452	4,247	8,699
2000 год	4,566	4,508	9,074
2001 год	5,009	5,089	10,098
2002 год	5,175	5,324	10,499
2003 год	5,130	5,506	10,636
2004 год	3,974	5,564	9,538
2005 год	3,848	5,426	9,274
2006 год	5,715	6,775	12,490
2007 год	4,255	6,843	11,098
Total	53,928	63,129	117,057

**gen labour = 0 if j77==0**

**replace labour = 0 if j1 == 5 & j56 == 0**

**replace labour = 2 if j1 == 5 & j56 == 1**

**replace labour = 1 if j1 <= 4**

**label define LABOUR 0 "нет работы" 1 "есть постоянная работа" 2 "есть только приработки"**

**label values labour LABOUR**

**tab id\_w labour**

НОМЕР ВОЛНЫ	labour			Total
	нет работ	есть пост	есть толь	
1994 год	3,717	4,894	282	8,893
1995 год	3,523	4,571	323	8,417
1996 год	3,686	4,382	273	8,341
1998 год	4,088	4,247	364	8,699
2000 год	4,064	4,508	502	9,074
2001 год	4,444	5,089	565	10,098
2002 год	4,624	5,324	551	10,499
2003 год	4,555	5,506	575	10,636
2004 год	4,515	5,564	582	10,661
2005 год	4,432	5,426	477	10,335
2006 год	5,067	6,775	648	12,490
2007 год	4,910	6,843	536	12,289
Total	51,625	63,129	5,678	120,432

26.2. Создать переменную «экономическая активность» (работает, не работает и ищет работу - безработный, не ищет работу – экономически не активный).

**ек\_akt** – экономическая активность, со значениями:

1 'есть работа'

2 'есть только приработки, ищет работу'

3 'есть только приработки, не ищет работу'

4 'нет работы никакой, ищет'

5 'нет работы никакой, не ищет'

Поиск работы: комбинация "ищет" и "обращался в поисках..."

\* **j81** - Хотели бы Вы найти (другую) работу?(5-7 раунды - спрашивали только безработных)

\* **j82** - Вы обращались куда-нибудь или к кому-нибудь в поисках работы в течение последних 30 дней? (5-7 раунды - спрашивали только безработных)

\* **j84\_1** - Если бы на прошлой неделе Вам предложили подходящую работу, Вы смогли бы

приступить к ней сразу? (с 8 раунда)

tab id\_w j81  
 tab id\_w j82  
 tab id\_w j84\_1

НОМЕР ВОЛНЫ	Хотели бы Вы найти работу?		Total
	нет	да	
1994 год	2,553	1,394	3,947
1995 год	2,517	1,260	3,777
1996 год	2,474	1,404	3,878
1998 год	5,383	3,144	8,527
2000 год	5,643	3,273	8,916
2001 год	6,482	3,429	9,911
2002 год	6,757	3,523	10,280
2003 год	6,931	3,468	10,399
2004 год	7,108	3,264	10,372
2005 год	7,274	2,762	10,036
2006 год	8,761	3,304	12,065
2007 год	9,069	2,793	11,862
Total	70,952	33,018	103,970

НОМЕР ВОЛНЫ	Вы обращались куда-нибудь или к кому-нибудь		Total
	нет	да	
1994 год	858	535	1,393
1995 год	717	539	1,256
1996 год	813	590	1,403
1998 год	2,071	1,069	3,140
2000 год	2,272	995	3,267
2001 год	2,390	1,033	3,423
2002 год	2,520	1,002	3,522
2003 год	2,501	967	3,468
2004 год	2,306	956	3,262
2005 год	1,922	831	2,753
2006 год	2,419	879	3,298
2007 год	1,912	878	2,790
Total	22,701	10,274	32,975

НОМЕР ВОЛНЫ	Если бы на прошлой неделе Вам предложили по		Total
	нет	да	
1998 год	493	2,498	2,991
2000 год	116	853	969
2001 год	825	2,436	3,261
2002 год	869	2,508	3,377
2003 год	805	2,482	3,287
2004 год	800	2,302	3,102
2005 год	618	1,962	2,580
2006 год	798	2,344	3,142
2007 год	699	1,931	2,630
Total	6,023	19,316	25,339

gen ek\_akt = 1 if labour ==1

```

replace ek_akt = 5 if (labour == 0 & (j81 == 0 | j81==. | j81 == .a | j81 == .b | j81== .c))
replace ek_akt = 5 if (labour == 0 & j81 == 1 & (j82==. | j82 == .a | j82 == .b | j82== .c))
replace ek_akt = 5 if (labour == 0 & j82 == 0)
replace ek_akt = 4 if (labour == 0 & j82 == 1)
replace ek_akt = 3 if (labour == 2 & (j81 == 0 | j81==. | j81 == .a | j81 == .b | j81== .c))
replace ek_akt = 5 if (labour == 2 & j81 == 1 & (j82==. | j82 == .a | j82 == .b | j82== .c))
replace ek_akt = 3 if (labour == 2 & j82 == 0)
replace ek_akt = 2 if (labour == 2 & j82 == 1)
label variable ek_akt "работает или реально ищет работу"
label define ЕК_АКТ 1 "есть работа" 2 "есть т.прираб., ищет работу" 3 "есть т.прираб.,
не ищет раб. " 4 "нет работы никакой, ищет" 5 "нет работы никакой, не ищет"
label values ek_akt ЕК_АКТ
tab id_w ek_akt

```

НОМЕР ВОЛНЫ	работает или реально ищет работу				Total
	есть т.пр	есть т.пр	нет работ	нет работ	
1994 год	98	184	437	8,174	8,893
1995 год	129	194	410	7,684	8,417
1996 год	118	155	472	7,596	8,341
1998 год	147	216	922	7,414	8,699
2000 год	172	330	823	7,749	9,074
2001 год	180	384	853	8,681	10,098
2002 год	160	391	842	9,106	10,499
2003 год	163	412	804	9,257	10,636
2004 год	184	398	772	9,307	10,661
2005 год	126	347	705	9,157	10,335
2006 год	159	488	720	11,123	12,490
2007 год	108	428	770	10,983	12,289
Total	1,744	3,927	8,530	106,231	120,432

\*26.3. переменная "занятость" - 3 градации

```

recode ek_akt (1=3) (2=3) (3=3) (4=2) (5=1), into (employment)
label variable employment "статус занятости"
label define EMPLOYMENT 1 "неактивный" 2 "безработный" 3 "есть любая работа"
label values employment EMPLOYMENT
tab id_w employment

```

НОМЕР ВОЛНЫ	статус занятости			Total
	неактивны	безработн	есть люба	
1994 год	8,174	437	282	8,893
1995 год	7,684	410	323	8,417
1996 год	7,596	472	273	8,341
1998 год	7,414	922	363	8,699
2000 год	7,749	823	502	9,074
2001 год	8,681	853	564	10,098
2002 год	9,106	842	551	10,499
2003 год	9,257	804	575	10,636
2004 год	9,307	772	582	10,661
2005 год	9,157	705	473	10,335
2006 год	11,123	720	647	12,490
2007 год	10,983	770	536	12,289
Total	106,231	8,530	5,671	120,432

\*26.4. дихотомическая переменная "работает-не работает"

recode employment (3=1) (2=0) (1=0), into (employed)

label variable employed "работает"

label define EMPLOYED 1 "есть любая работа" 0 "не работает или неактивный"

label values employed EMPLOYED

tab id\_w employed

НОМЕР ВОЛНЫ	работает		Total
	не работа	есть люба	
1994 год	8,611	282	8,893
1995 год	8,094	323	8,417
1996 год	8,068	273	8,341
1998 год	8,336	363	8,699
2000 год	8,572	502	9,074
2001 год	9,534	564	10,098
2002 год	9,948	551	10,499
2003 год	10,061	575	10,636
2004 год	10,079	582	10,661
2005 год	9,862	473	10,335
2006 год	11,843	647	12,490
2007 год	11,753	536	12,289
Total	114,761	5,671	120,432

\*26.5. Переменная «состоит в браке, включая неформальный»

НОМЕР ВОЛНЫ	СЕМЕЙНОЕ ПОЛОЖЕНИЕ В ТЕКУЩЕЙ ВОЛНЕ					Total
	Никогда в	Состоите	Живете вм	Разведены	Вдовец (в	
1994 год	1,183	0	0	686	1,017	8,528
1995 год	1,416	0	0	629	1,024	8,384
1996 год	1,456	0	0	631	1,007	8,301
1998 год	1,599	4,868	588	582	1,040	8,677
2000 год	1,781	4,899	691	606	1,087	9,064
2001 год	2,009	5,257	862	725	1,224	10,077
2002 год	2,159	5,270	972	777	1,276	10,491
2003 год	2,167	5,272	1,076	828	1,264	10,607
2004 год	2,251	5,277	1,043	818	1,260	10,649
2005 год	2,205	5,120	980	804	1,214	10,323
2006 год	2,655	6,028	1,221	1,016	1,445	12,464
2007 год	2,575	5,973	1,249	939	1,442	12,271
Total	23,456	47,964	8,682	9,041	14,300	119,836

НОМЕР ВОЛНЫ	СЕМЕЙНОЕ ПОЛОЖЕНИЕ В ТЕКУЩЕЙ ВОЛНЕ		Total
	ОФИЦИАЛЬН	Состоите	
1994 год	0	5,642	8,528
1995 год	0	5,315	8,384
1996 год	0	5,207	8,301
1998 год	0	0	8,677
2000 год	0	0	9,064
2001 год	0	0	10,077
2002 год	37	0	10,491
2003 год	0	0	10,607
2004 год	0	0	10,649
2005 год	0	0	10,323
2006 год	99	0	12,464
2007 год	93	0	12,271
Total	229	16,164	119,836

recode marst (1=0) (2=1) (3=1) (4=0) (5=0) (6=1) (7=1), into (married)

**label variable married " состоит в браке, включая неформальный "**  
**label values married YES**  
**tab id\_w employed**

НОМЕР ВОЛНЫ	работает		Total
	не работа	есть люба	
1994 год	8,611	282	8,893
1995 год	8,094	323	8,417
1996 год	8,068	273	8,341
1998 год	8,336	363	8,699
2000 год	8,572	502	9,074
2001 год	9,534	564	10,098
2002 год	9,948	551	10,499
2003 год	10,061	575	10,636
2004 год	10,079	582	10,661
2005 год	9,862	473	10,335
2006 год	11,843	647	12,490
2007 год	11,753	536	12,289
Total	114,761	5,671	120,432

\* сохранить файл

\*26.6. Объявление данных панельными, то есть указание идентификатора кейсов (переменная) и переменной времени (номера волны)

**tsset idind id\_w**

```
panel variable: idind (unbalanced)
time variable: id_w, 5 to 16, but with gaps
delta: 1 unit
```

### *Time-series varlists*

Video example: <https://www.youtube.com/watch?v=ik8r4WvrPkc>

Before using time-series operators, you must declare the time variable using tsset. Description

Time-series varlists are a variation on varlists of existing variables. When a command allows a time-series varlist, you may include time-series operators. For instance, L.gnp refers to the lagged value of variable gnp. The time-series operators are

Operator	Meaning
L.	lag (x <sub>t</sub> -1)
L2.	2-period lag (x <sub>t</sub> -2)
...	
F.	lead (x <sub>t</sub> +1)
F2.	2-period lead (x <sub>t</sub> +2)
...	
D.	difference (x <sub>t</sub> - x <sub>t</sub> -1)
D2.	difference of difference (x <sub>t</sub> - 2x <sub>t</sub> -1 + x <sub>t</sub> -2)
...	
S.	"seasonal" difference (x <sub>t</sub> - x <sub>t</sub> -1)
S2.	lag-2 (seasonal) difference (x <sub>t</sub> - x <sub>t</sub> -2)

\*Создадим новые переменные, равные значению переменной за прошлый год для данного индивида

**gen marst\_T = L.marst**  
**label variable marst\_T "T-1 семейное положение в прошлой волне"**  
**label values marst\_T marst**  
**gen employment\_T = L.employment**

label variable employment\_T "Т-1 занятость в прошлой волне"  
 label values employment\_T EMPLOYMENT  
 gen occup08\_k\_T = L.occup08\_k  
 label variable occup08\_k\_T "Т-1 проф.статус в прошлой волне"  
 label values occup08\_k\_T OCCUP08\_k

\*26.7. Матрица переходов между профессиональными статусами:

В строках – проф.статус в прошлом году, в колонках – в нынешнем году; горизонтальный %; интерпретация – какая доля людей из данной проф.группы (и незанятых) меняет статус (переходит в другую проф.группу или перестает работать)

В таблице отмечено: желтым – сохранение статуса; голубым – нисходящая мобильность, красным – восходящая мобильность.

by id\_w, sort : tabulate occup08\_k\_T occup08\_k, nofreq row

-----  
 -> id\_w = 1994 г  
 no observations  
 -----

-> id\_w = 1995 г

Т-1 проф.статус в прошлой волне	ПРОФЕССИОНАЛЬНАЯ ГРУППА - по коду J2COD08 кратко				Total
	не работа	руководит	специалис	служащие	
не работает	88.29	0.45	1.39	4.31	100.00
руководители	9.09	76.26	3.54	7.07	100.00
специалисты	5.18	3.60	79.86	9.50	100.00
служащие	12.00	2.55	4.00	76.60	100.00
квалиф.рабочие	10.18	0.86	0.07	4.01	100.00
неквал.рабочие	12.27	0.31	0.00	9.82	100.00
военные	16.67	2.38	2.38	4.76	100.00
Total	44.93	3.38	9.46	17.40	100.00

Т-1 проф.статус в прошлой волне	ПРОФЕССИОНАЛЬНАЯ ГРУППА - по коду J2COD08 кратко			Total
	квалиф.ра	неквал.ра	военные	
не работает	3.99	1.46	0.10	100.00
руководители	3.03	0.51	0.51	100.00
специалисты	1.58	0.14	0.14	100.00
служащие	3.32	1.11	0.43	100.00
квалиф.рабочие	82.44	2.37	0.07	100.00
неквал.рабочие	11.04	66.56	0.00	100.00
военные	7.14	0.00	66.67	100.00
Total	19.78	4.48	0.56	100.00

-----  
 -> id\_w = 1996 г  
 -----

Т-1 проф.статус в прошлой волне	ПРОФЕССИОНАЛЬНАЯ ГРУППА - по коду J2COD08 кратко				Total
	не работа	руководит	специалис	служащие	
не работает	89.16	0.23	1.33	4.49	100.00
руководители	8.26	65.22	8.70	12.61	100.00
специалисты	5.30	2.80	84.74	6.39	100.00
служащие	11.57	2.36	4.14	77.20	100.00
квалиф.рабочие	12.38	0.37	0.22	4.74	100.00
неквал.рабочие	18.18	0.34	0.34	4.04	100.00
военные	7.69	2.56	2.56	12.82	100.00

Total	46.28	3.08	9.67	17.66	100.00
T-1 проф.статус в прошлой волне	ПРОФЕССИОНАЛЬНАЯ ГРУППА - по коду J2COD08 кратко				Total
не работает	2.99	1.73	0.07		100.00
руководители	3.48	0.87	0.87		100.00
специалисты	0.47	0.16	0.16		100.00
служащие	2.28	2.28	0.17		100.00
квалиф.рабочие	78.95	3.26	0.07		100.00
неквал.рабочие	10.44	66.67	0.00		100.00
военные	2.56	0.00	71.79		100.00
Total	18.01	4.77	0.53		100.00

-> id\_w = 1998 г

T-1 проф.статус в прошлой волне	ПРОФЕССИОНАЛЬНАЯ ГРУППА - по коду J2COD08 кратко				Total
	не работа	руководит	специалис	служащие	
не работает	86.86	0.37	1.76	5.52	100.00
руководители	10.99	60.73	8.38	15.71	100.00
специалисты	11.68	4.28	76.97	5.92	100.00
служащие	18.04	1.93	4.38	69.70	100.00
квалиф.рабочие	16.22	0.59	0.93	5.83	100.00
неквал.рабочие	28.06	0.32	0.32	10.00	100.00
военные	13.89	2.78	0.00	16.67	100.00
Total	49.31	2.84	9.25	17.51	100.00

T-1 проф.статус в прошлой волне	ПРОФЕССИОНАЛЬНАЯ ГРУППА - по коду J2COD08 кратко				Total
	квалиф.ра	неквал.ра	военные		
не работает	3.53	1.93	0.03		100.00
руководители	2.62	1.57	0.00		100.00
специалисты	0.49	0.33	0.33		100.00
служащие	4.64	1.23	0.09		100.00
квалиф.рабочие	73.99	2.28	0.17		100.00
неквал.рабочие	8.71	52.58	0.00		100.00
военные	8.33	0.00	58.33		100.00
Total	16.56	4.12	0.42		100.00

-> id\_w = 2000 г

T-1 проф.статус в прошлой волне	ПРОФЕССИОНАЛЬНАЯ ГРУППА - по коду J2COD08 кратко				Total
	не работа	руководит	специалис	служащие	
не работает	84.44	0.63	1.20	6.76	100.00
руководители	8.42	62.63	5.79	17.89	100.00
специалисты	6.75	3.22	80.87	8.20	100.00
служащие	13.84	2.64	4.12	71.66	100.00
квалиф.рабочие	14.72	0.86	1.03	5.42	100.00
неквал.рабочие	27.18	0.70	0.35	9.41	100.00
военные	11.43	2.86	5.71	5.71	100.00
Total	48.94	2.94	8.87	18.33	100.00

T-1 проф.статус в прошлой волне

ПРОФЕССИОНАЛЬНАЯ ГРУППА - по коду J2COD08 кратко

волне	квалиф.ра	неквал.ра	военные	Total
не работает	4.56	2.38	0.03	100.00
руководители	4.74	0.00	0.53	100.00
специалисты	0.64	0.32	0.00	100.00
служащие	5.44	2.22	0.08	100.00
квалиф.рабочие	74.96	2.84	0.17	100.00
неквал.рабочие	12.20	50.17	0.00	100.00
военные	2.86	2.86	68.57	100.00
Total	16.36	4.14	0.41	100.00

-> id\_w = 2001 г

Т-1 проф.статус в прошлой волне	ПРОФЕССИОНАЛЬНАЯ ГРУППА - по коду J2COD08 кратко				Total
	не работа	руководит	специалис	служащие	
не работает	87.69	0.31	1.13	5.07	100.00
руководители	5.45	75.45	5.91	10.00	100.00
специалисты	6.49	3.92	82.05	5.73	100.00
служащие	11.96	3.66	4.71	72.91	100.00
квалиф.рабочие	11.20	0.39	0.47	5.05	100.00
неквал.рабочие	19.24	0.95	0.00	8.20	100.00
военные	6.67	3.33	3.33	13.33	100.00
Total	49.07	3.39	8.63	17.75	100.00

Т-1 проф.статус в прошлой волне	ПРОФЕССИОНАЛЬНАЯ ГРУППА - по коду J2COD08 кратко			Total
	квалиф.ра	неквал.ра	военные	
не работает	3.86	1.91	0.03	100.00
руководители	3.18	0.00	0.00	100.00
специалисты	1.21	0.30	0.30	100.00
служащие	4.15	2.53	0.07	100.00
квалиф.рабочие	80.40	2.41	0.08	100.00
неквал.рабочие	12.30	58.99	0.32	100.00
военные	3.33	0.00	70.00	100.00
Total	16.60	4.22	0.35	100.00

-> id\_w = 2002 г

Т-1 проф.статус в прошлой волне	ПРОФЕССИОНАЛЬНАЯ ГРУППА - по коду J2COD08 кратко				Total
	не работа	руководит	специалис	служащие	
не работает	87.90	0.37	1.21	5.31	100.00
руководители	9.21	68.42	3.29	16.45	100.00
специалисты	6.35	1.65	83.23	7.37	100.00
служащие	10.82	2.96	3.16	77.01	100.00
квалиф.рабочие	10.93	0.72	0.65	4.05	100.00
неквал.рабочие	18.35	0.53	0.53	8.24	100.00
военные	3.57	0.00	10.71	10.71	100.00
Total	48.62	3.38	8.94	18.59	100.00

Т-1 проф.статус в прошлой волне	ПРОФЕССИОНАЛЬНАЯ ГРУППА - по коду J2COD08 кратко			Total
	квалиф.ра	неквал.ра	военные	
не работает	3.24	1.98	0.00	100.00
руководители	0.99	0.99	0.66	100.00
специалисты	1.02	0.25	0.13	100.00
служащие	4.25	1.67	0.13	100.00

квалиф.рабочие	81.77	1.88	0.00	100.00
неквал.рабочие	11.17	61.17	0.00	100.00
военные	10.71	0.00	64.29	100.00
Total	15.94	4.26	0.26	100.00

-> id\_w = 2003 г

Т-1 проф.статус в прошлой волне	ПРОФЕССИОНАЛЬНАЯ ГРУППА - по коду J2COD08 кратко				Total
	не работа	руководит	специалис	служащие	
не работает	87.99	0.50	1.22	5.45	100.00
руководители	6.31	74.09	4.32	11.63	100.00
специалисты	5.31	3.46	84.07	6.30	100.00
служащие	10.31	2.04	3.65	78.85	100.00
квалиф.рабочие	10.39	1.32	0.90	5.33	100.00
неквал.рабочие	14.70	0.26	0.52	8.92	100.00
военные	0.00	3.70	3.70	14.81	100.00
Total	47.83	3.62	9.11	19.42	100.00

Т-1 проф.статус в прошлой волне	ПРОФЕССИОНАЛЬНАЯ ГРУППА - по коду J2COD08 кратко			Total
	квалиф.ра	неквал.ра	военные	
не работает	3.19	1.65	0.00	100.00
руководители	2.99	0.66	0.00	100.00
специалисты	0.74	0.12	0.00	100.00
служащие	3.18	1.86	0.12	100.00
квалиф.рабочие	79.36	2.63	0.07	100.00
неквал.рабочие	12.60	62.99	0.00	100.00
военные	7.41	0.00	70.37	100.00
Total	15.52	4.25	0.24	100.00

-> id\_w = 2004 г

Т-1 проф.статус в прошлой волне	ПРОФЕССИОНАЛЬНАЯ ГРУППА - по коду J2COD08 кратко				Total
	не работа	руководит	специалис	служащие	
не работает	87.97	0.51	1.32	4.85	100.00
руководители	4.04	75.42	5.05	10.10	100.00
специалисты	5.82	2.28	84.45	6.32	100.00
служащие	11.01	2.78	3.23	77.81	100.00
квалиф.рабочие	9.42	0.57	0.43	4.35	100.00
неквал.рабочие	17.57	0.52	0.52	9.82	100.00
военные	4.35	0.00	4.35	0.00	100.00
Total	47.40	3.59	8.96	19.58	100.00

Т-1 проф.статус в прошлой волне	ПРОФЕССИОНАЛЬНАЯ ГРУППА - по коду J2COD08 кратко			Total
	квалиф.ра	неквал.ра	военные	
не работает	3.74	1.55	0.07	100.00
руководители	5.05	0.34	0.00	100.00
специалисты	1.01	0.00	0.13	100.00
служащие	3.52	1.53	0.11	100.00
квалиф.рабочие	83.01	2.21	0.00	100.00
неквал.рабочие	8.79	62.79	0.00	100.00
военные	4.35	0.00	86.96	100.00
Total	16.07	4.10	0.29	100.00

-----  
 -> id\_w = 2005 г

Т-1 проф.статус в прошлой волне	ПРОФЕССИОНАЛЬНАЯ ГРУППА - по коду J2COD08 кратко				Total
	не работа	руководит	специалис	служащие	
не работает	88.71	0.28	0.98	4.85	100.00
руководители	6.56	66.23	4.92	18.36	100.00
специалисты	5.11	2.37	84.56	6.97	100.00
служащие	11.14	1.68	2.97	78.84	100.00
квалиф.рабочие	9.46	0.74	0.47	4.96	100.00
неквал.рабочие	18.60	0.00	0.27	6.20	100.00
военные	4.00	0.00	0.00	12.00	100.00
Total	47.22	3.02	8.77	20.13	100.00

Т-1 проф.статус в прошлой волне	ПРОФЕССИОНАЛЬНАЯ ГРУППА - по коду J2COD08 кратко			Total
	квалиф.ра	неквал.ра	военные	
не работает	3.58	1.60	0.00	100.00
руководители	3.28	0.66	0.00	100.00
специалисты	0.75	0.00	0.25	100.00
служащие	3.58	1.74	0.06	100.00
квалиф.рабочие	81.89	2.41	0.07	100.00
неквал.рабочие	9.43	65.50	0.00	100.00
военные	0.00	0.00	84.00	100.00
Total	16.40	4.19	0.28	100.00

-----  
 -> id\_w = 2006 г

Т-1 проф.статус в прошлой волне	ПРОФЕССИОНАЛЬНАЯ ГРУППА - по коду J2COD08 кратко				Total
	не работа	руководит	специалис	служащие	
не работает	87.25	0.38	1.22	6.09	100.00
руководители	4.01	77.01	4.01	10.58	100.00
специалисты	4.77	2.32	84.20	8.31	100.00
служащие	9.43	2.36	3.65	79.29	100.00
квалиф.рабочие	7.96	0.78	0.70	4.51	100.00
неквал.рабочие	15.79	0.28	0.83	9.70	100.00
военные	0.00	13.04	0.00	4.35	100.00
Total	45.99	3.43	8.63	21.16	100.00

Т-1 проф.статус в прошлой волне	ПРОФЕССИОНАЛЬНАЯ ГРУППА - по коду J2COD08 кратко			Total
	квалиф.ра	неквал.ра	военные	
не работает	3.46	1.60	0.00	100.00
руководители	3.65	0.36	0.36	100.00
специалисты	0.27	0.14	0.00	100.00
служащие	3.54	1.52	0.22	100.00
квалиф.рабочие	83.58	2.47	0.00	100.00
неквал.рабочие	10.25	63.16	0.00	100.00
военные	0.00	0.00	82.61	100.00
Total	16.43	4.09	0.27	100.00

-----  
 -> id\_w = 2007 г

Т-1 проф.статус в прошлой	ПРОФЕССИОНАЛЬНАЯ ГРУППА - по коду J2COD08 кратко		
	квалиф.ра	неквал.ра	военные

волне	не работа	руководит	специалис	служащие	Total
не работает	87.18	0.50	0.96	5.73	100.00
руководители	6.40	74.71	4.07	10.47	100.00
специалисты	4.66	3.07	84.64	6.46	100.00
служащие	8.72	2.38	3.42	80.49	100.00
квалиф.рабочие	7.60	0.76	0.70	5.49	100.00
неквал.рабочие	14.48	0.69	0.69	9.66	100.00
военные	2.94	5.88	2.94	11.76	100.00
Total	44.21	3.63	9.06	21.94	100.00

  

T-1 проф.статус в прошлой волне	ПРОФЕССИОНАЛЬНАЯ ГРУППА - по коду J2COD08 кратко			Total
волне	квалиф.ра	неквал.ра	военные	Total
не работает	3.77	1.83	0.02	100.00
руководители	4.36	0.00	0.00	100.00
специалисты	0.74	0.42	0.00	100.00
служащие	3.28	1.53	0.18	100.00
квалиф.рабочие	82.06	3.39	0.00	100.00
неквал.рабочие	10.11	64.37	0.00	100.00
военные	2.94	0.00	73.53	100.00
Total	16.44	4.42	0.29	100.00

## 26.8. Самостоятельное задание.

Построить аналогичные матрицы переходов для статуса занятости и для брачного статуса.

## 27. Перекодирование в новые переменные для регрессионного анализа

\* перед каждой операцией нужно посмотреть распределение исходных переменных, а после - новых

\*27.1. Исправление кода новосибирской области для переменной **psu** для волны от 12 и больше: связано с тем, что в 5-11 волнах код 32 был у Ханты-Мансийского региона, а с 12й волны – у Новосибирской области. Эту переменную будем использовать для приклейки региональных данных.

```
recode psu (32=39) if (id_w >=12 & psu == 32)
label define psu 39 "Новосибирская область", add
```

\*27.2. Создадим переменную номера федерального округа (сначала посмотреть распределение переменной **region**)

\* ОПИСАНИЕ ФЕД.ОКРУГОВ

1- 'М., СП'

138 Москва 140 Новая Москва 141 Санкт-Петербург

2 'Северный'

1 Ленинградская область 89 Коми Республика 105 Коми Республика

3 'Центральный'

14 Калужская область 33 Тамбовская область 67 Тверская область 72 Липецк

135 Смоленск 136 Тула 142 Московская область

4 'Приволжский'

10 Удмуртская Республика 12 Пермский край 45 Казань 47 Оренбургская область 48 Республика Чувашия

70 Саратов 100 Саратовская область 116 Нижний Новгород

5 'Юг и Сев. Кавк.'

9 Краснодар 39 Волгоградская область 52 Ставропольский край 77 Республика Кабардино-Балкария 129 Краснодарский край 137 Ростовская область

6 'Урал'

46 Курган 86 Ханты-Мансийский АО 106 Челябинск 107 Челябинская область 117 Пензенская область

7 'Сибирский'

58 Алтайский край 66 Красноярск 71 Томск 73 Красноярский край

84 Алтайский край 161 Новосибирская область

8 'Дальневосточный'

92 Владивосток 93 Амурская область

**recode region (138=1) (140=1) (141=1) (1=2) (89=2) (105=2) (14=3) (33=3) (67=3) (72=3) (135=3) (136=3) (142=3) (10 =4) (12=4) (45=4) (47=4) (48=4) (70=4) (100=4) (116=4) (9=5) (39=5) (52=5) (77=5) (129=5) (137=5) (46=6) (86=6) (106=6) (107=6) (117=6) (58=7) (66=7) (71=7) (73=7) (84=7) (161=7) (92=8) (93=8) (200=2), into (fed\_окр)**

**label variable fed\_окр "федеральный округ"**

**label define FED\_ОКР 1 "Москва, С.Петербург" 2 "Северный" 3 "Центральный" 4 "Приволжский" 5 "Юг и С.Кавказ" 6 "Уральский " 7 "'Сибирский" 8 "Дальневосточный"**

**label values fed\_окр FED\_ОКР**

**tab id\_w fed\_окр**

федеральный округ	Freq.	Percent	Cum.
Москва, С.Петербург	13,689	11.37	11.37
Северный	7,803	6.48	17.85
Центральный	21,592	17.93	35.77
Приволжский	23,901	19.85	55.62
Юг и С.Кавказ	19,807	16.45	72.06
Уральский	12,278	10.19	82.26
'Сибирский	15,364	12.76	95.02
Дальневосточный	6,002	4.98	100.00
Total	120,436	100.00	

\*27.3. Создадим переменную "тип поселения" (выделение Москвы и СП, объединение ПГТ и села); сначала посмотрим распределение переменной STATUS,

**recode status (4=3) (1=1) (2=2) (3=3) if fed\_окр != 1, into (status\_1)**

**recode status\_1 (1=0) if fed\_окр == 1**

**label define STATUS1 0 "М., СП" 1 "обл.центр" 2 "другой город" 3 "село, пгт"**

**label values status\_1 STATUS1**

**tab id\_w status\_1**

НОМЕР ВОЛНЫ	тип населенного пункта				Total
	М., СП	обл.центр	другой го	село, пгт	
1994 год	962	2,838	2,443	2,650	8,893
1995 год	780	2,671	2,372	2,594	8,417
1996 год	686	2,623	2,337	2,695	8,341
1998 год	629	2,700	2,446	2,924	8,699
2000 год	482	2,981	2,441	3,170	9,074
2001 год	1,388	2,927	2,508	3,275	10,098
2002 год	1,579	2,985	2,557	3,378	10,499
2003 год	1,440	2,974	2,705	3,517	10,636
2004 год	1,497	2,977	2,772	3,416	10,662
2005 год	1,328	2,862	2,704	3,443	10,337
2006 год	1,494	3,675	3,331	3,990	12,490
2007 год	1,424	3,671	3,254	3,941	12,290
Total	13,689	35,884	31,870	38,993	120,436

\*27.4. Создадим возрастные группы.

**recode age (13 / 17 = 1) (18 / 24=2) (25 / 34=3) (35 / 44=4) (45 / 54=5) (55 / 64=6) (65 / 200 =7), into (age\_kod10)**

**label variable age\_kod10 "коды возраста по 5-10 лет"**

**label define age\_kod10 1 "13-17" 2 "18-24" 3 "25-34" 4 "35-44" 5 "45-54" 6 "55-64" 7 ">=65"**

**label values age\_kod10 age\_kod10**

**tab id\_w age\_kod10**

НОМЕР ВОЛНЫ	коды возраста по 5-10 лет				Total
	13-17	18-24	25-34	35-44	
1994 год	564	1,041	1,652	1,737	8,893
1995 год	519	1,008	1,455	1,681	8,417
1996 год	557	1,019	1,432	1,656	8,341
1998 год	702	1,097	1,388	1,721	8,698
2000 год	764	1,209	1,495	1,690	9,074
2001 год	810	1,362	1,736	1,801	10,098
2002 год	871	1,398	1,852	1,792	10,497
2003 год	853	1,432	1,986	1,733	10,633
2004 год	854	1,423	2,051	1,717	10,661
2005 год	787	1,385	2,034	1,626	10,336
2006 год	850	1,760	2,407	1,916	12,488
2007 год	780	1,719	2,359	1,903	12,288
Total	8,911	15,853	21,847	20,973	120,424

НОМЕР ВОЛНЫ	коды возраста по 5-10 лет			Total
	45-54	55-64	>=65	
1994 год	1,175	1,380	1,344	8,893
1995 год	1,141	1,251	1,362	8,417
1996 год	1,081	1,230	1,366	8,341
1998 год	1,232	1,163	1,395	8,698
2000 год	1,428	1,045	1,443	9,074
2001 год	1,637	1,128	1,624	10,098
2002 год	1,701	1,164	1,719	10,497
2003 год	1,773	1,099	1,757	10,633
2004 год	1,758	1,075	1,783	10,661
2005 год	1,735	1,017	1,752	10,336
2006 год	2,163	1,298	2,094	12,488
2007 год	2,150	1,359	2,018	12,288
Total	18,974	14,209	19,657	120,424

\*27.5. Укрупним переменную "национальность".

**recode i4 (1 =1) (50=1) (70=1) (85=1) (99=1) (102=1) (105=1) (106=1) (112=1) (113=1) (114=1) (115=1) (120=1) (121=1) (124=1) (129=1) (130=1) (131=1) (136=1) (137=1) (138=1) (148=1) (151=1) (152=1) (156=1) (159=1) (174=1) (2=2) (4 = 2) (13 = 2) (127=2) (132=2) (141 =2) (18=3) (22=3) (27=3) (28=3) (29=3) (32=3) (33=3) (34=3) (36=3) (48=3) (56=3) (65=3) (76=3) (82=3) (81=3) (98=3) (95=3) (103=3) (123=3) (166=3) (11=4) (12=4) (14=4) (16=4) (35=4) (39=4) (75=4) (17=4) (9=5) (21=5) (158=5) (170=5) (172= 5) (3=6) (5=6) (6=6) (7=6) (8=6) (10=6) (20=6) (24=6) (25=6) (26=6) (31=6) (37=6) (41=6) (47=6) (51=6) (52=6) (67=6) (68=6) (70=6) (75=6) (86=6) (90=6) (96=6) (107=6) (110=6) (111=6) (119=6) (144=6) (145=6) (155=6) (161=6) (167=6) (171=6) (175=6) (176=6) (19=7) (30=7) (38=7) (40=7) (42=7) (43=7) (49=7) (58=7) (60=7) (64=7) (72=7) (77=7) (80=7) (81=7) (83=3) (87=7) (88=7) (93=7) (95=7) (100=7) (118=7) (123=7) (134=7) (142=7) (154=7) (155=7) (158=7) (164=7) (else = 7), into (i4\_k)**

**label variable i4\_k "группы национальностей кратко"**

**label define I4\_K 1 "русские, смешанные русские" 2 "украинцы, белорусы, молдаване"**

3 "народы Сев.Кавказа" 4 "народы Поволжья и Севера" 5 "татары, башкиры" 6  
 "прочие европейские" 7 "прочие не европейские"  
 label values i4\_k I4\_K

tab id\_w i4\_k

НОМЕР ВОЛНЫ	группы национальностей кратко				Total
	русские,	украинцы,	народы Се	народы По	
1994 год	7,447	316	312	341	8,893
1995 год	6,876	302	277	328	8,417
1996 год	6,659	295	292	315	8,341
1998 год	7,175	284	366	348	8,699
2000 год	7,674	219	393	344	9,074
2001 год	8,528	233	408	362	10,098
2002 год	8,561	230	396	373	10,499
2003 год	8,523	230	388	369	10,636
2004 год	8,912	201	508	360	10,662
2005 год	8,011	198	428	345	10,337
2006 год	10,538	234	513	427	12,490
2007 год	10,448	223	501	387	12,290
Total	99,352	2,965	4,782	4,299	120,436

НОМЕР ВОЛНЫ	группы национальностей кратко			Total
	татары, б	прочие ев	прочие не	
1994 год	279	175	23	8,893
1995 год	239	170	225	8,417
1996 год	230	161	389	8,341
1998 год	235	153	138	8,699
2000 год	256	144	44	9,074
2001 год	262	160	145	10,098
2002 год	260	150	529	10,499
2003 год	266	160	700	10,636
2004 год	254	146	281	10,662
2005 год	248	143	964	10,337
2006 год	343	164	271	12,490
2007 год	352	162	217	12,290
Total	3,224	1,888	3,926	120,436

\*27.6. Создадим удобные переменные: "мужской пол=1", возраст/10 (для более компактного коэффициента), село, образование

recode h5 (1=1) (2=0), into (male)

gen age\_10=age/10

recode status\_1 (3=1) (else=0), into (village)

recode diplom (1/3 =1) (4=2) (5=3) (6=4), into (diplom\_k)

label variable male "мужской пол"

label variable age\_10 "возраст \10"

label variable village "село"

label variable diplom\_k "законченное образование кратко"

label define DIPLOM\_K 1 "неполное среднее и ниже" 2 "законченное среднее" 3  
 "среднее профессиональное" 4 "высшее"

label values diplom\_k DIPLOM\_K

tab1 male village diplom\_k

-> tabulation of male

мужск ой пол	Freq.	Percent	Cum.
0	68,615	56.97	56.97
1	51,821	43.03	100.00
Total	120,436	100.00	

-> tabulation of village

село	Freq.	Percent	Cum.
0	81,443	67.62	67.62
1	38,993	32.38	100.00
Total	120,436	100.00	

-> tabulation of diplom\_k

законченное образование кратко	Freq.	Percent	Cum.
неполное среднее и ниже	31,867	26.49	26.49
законченное среднее	40,771	33.89	60.38
среднее профессиональное	27,419	22.79	83.17
высшее	20,254	16.83	100.00
Total	120,311	100.00	

**\*Сохраните файл данных!!!**

## 28. Приклеивание переменных по семье и региональной статистики

### 28.1. Приклеивание к рабочему файлу данных по семье

**merge m:1 id\_w id\_h using "C:\RLMS\_work\seminar\_6\data\H\_5\_16\_family.dta"**

Result	# of obs.	
not matched	30	
from master	0	( _merge==1)
from using	30	( _merge==2)
matched	120,436	( _merge==3)

**drop if \_merge == 2  
drop \_merge**

**tab id\_w nfm**

НОМЕР ВОЛНЫ	Колич ество члено в семьи	13	Total
1994 год	0	0	8,893
1995 год	0	0	8,417
1996 год	0	0	8,341
1998 год	0	0	8,699
2000 год	8	8	9,074

2001 год		0		10,098
2002 год		16		10,499
2003 год		0		10,636
2004 год		0		10,662
2005 год		0		10,337
2006 год		0		12,490
2007 год		8		12,290
-----+-----+-----				
Total		32		120,436

## \*28.2. Приклеивание региональной статистики

```
merge m:1 year psu using "C:\RLMS_work\seminar_6\data\reg_statistics_1994_2021.dta"
drop if _merge == 2
drop _merge
```

\*дефлирование и логарифмирование региональной заработной платы и региональных душевых доходов

```
gen ln_regwage = ln(regwage_N * deflat_16)
gen ln_regincome = ln( regincome_N1 * deflat_16)
label variable ln_regincome "логарифм дефлированных душевых рег доходов"
label variable ln_regwage "логарифм дефлир региональной ЗП"
```

## \*29. Коррекция доходов и расчет новых переменных о доходах.

В этом задании главное – скорректировать разные переменные о доходах из индивидуального и семейного файла, например, в том случае, если человек общую сумму своих доходов назвал меньшей, чем если сложить его доходы из разных источников; или если сумма семейных доходов меньше, чем если сложить доходы всех членов семьи. Если такую корректировку не сделать, то не получится верно рассчитать доходы остальных членов семьи, кроме мужа и жены; то же самое – нетрудовые доходы человека.

\*29.1. Посчитаем суммарные заработки на всех 3х работах, и суммарные доходы индивиде из всех источников (заработки, пенсия, пособие по безработице). Для дефлирования создадим несколько вспомогательных переменных, так как у нас рассчитана дефлированная заработная плата.

```
gen double j45_46_d = j45/j46 * deflat_16 if j46 !=0
gen double j40_d = j40 * deflat_16
gen double j48_d = j48 * deflat_16
gen double j57_d = j57 * deflat_16
gen double j76_d = j76 * deflat_16
gen double j89_d = j89 * deflat_16
gen double j60_d = j60 * deflat_16

egen double wage_123 = rowtotal(Wage_1 j40_d j45_46_d j48_d j57_d), missing
egen double ind_income_sum = rowtotal(Wage_1 j40_d j45_46_d j48_d j57_d j76_d j89_d),
missing
drop j76_d j89_d j45_46_d j40_d j48_d j57_d
label variable wage_123 "дефлир. ЗП на 3х работах за 30 дней"
format %9.2f wage_123
label variable ind_income_sum "дефлир сумма доходов индивида из всех источников"

tabstat wage_123, statistics( mean count min max ) by(id_w) columns(statistics)
format(%9.0f)
```

Summary for variables: wage\_123  
by categories of: id\_w (НОМЕР ВОЛНЫ)

id_w	mean	N	min	max
1994 год	6647	4747	33	178209
1995 год	5713	4366	0	111499
1996 год	6976	4116	39	124125
1998 год	4311	4148	0	154259
2000 год	4849	4545	0	926557
2001 год	5935	5090	0	214253
2002 год	6655	5364	0	253834
2003 год	7159	5515	0	151092
2004 год	7632	5684	0	87923
2005 год	8425	5457	0	146365
2006 год	9346	6789	0	179040
2007 год	10018	6671	0	110000
Total	7203	62492	0	926557

**tabstat ind\_income\_sum, statistics( mean count min max ) by(id\_w) columns(statistics)  
format(%9.0f)**

Summary for variables: ind\_income\_sum  
by categories of: id\_w (НОМЕР ВОЛНЫ)

id_w	mean	N	min	max
1994 год	5534	7016	33	178209
1995 год	4716	6437	0	111499
1996 год	6001	5609	39	124125
1998 год	3570	6126	0	154259
2000 год	3985	6828	0	926557
2001 год	4938	7593	0	214253
2002 год	5549	7991	0	253834
2003 год	6004	8113	0	151092
2004 год	6417	8218	0	87923
2005 год	7177	7862	0	146365
2006 год	7939	9683	0	179040
2007 год	8451	9497	0	110000
Total	6032	90973	0	926557

\*У нас есть переменная j60 – это сумма, которую индивид назвал как сумму своих всех доходов. Построим переменную, которая будет максимумом из этих двух переменных.

```
egen double I_income = rowmax(j60_d ind_income_sum)
format %9.2f I_income
label variable I_income "дефлир.инд.доходы (максимум из разных оценок)"
```

\*Рассчитаем нетрудовой доход = свои доходы минус заработная плата (за 30 дней)  
\*Для этого сделаем вспомогательную переменную: заработки умножим на минус 1.

```
gen double wage_123_m = wage_123 * -1
egen double nolab_income = rowtotal(I_income wage_123_m) if (I_income != . & I_income
>= wage_123_m)
replace nolab_income = I_income if wage_123_m == .
recode nolab_income (-1 / 1 = 0)
label variable nolab_income "дефлир.нетрудовые доходы"
drop wage_123_m
```

\*29.2. Приклеивание данных по супругу\супруге (аналогично семинару 3)

\*Сохраним наш файл под новым именем – это будут данные по супругу, и сохраним только несколько переменных

```
save "C:\RLMS_work\seminar_6\data\spouse_5_16.dta"  
keep id_w idind id_i I_income wage_123 nolab_income employment employed occup08_k  
i4_k  
save "C:\RLMS_work\seminar_6\data\spouse_5_16.dta", replace
```

\*Переименовать переменные, так как это будут данные о супруге(супругу), их нужно отличать от данных респондента. Переименовать переменную **id\_i** в **idi\_spouse**, остальным переменным добавить индекс (например, S).

```
rename (idind id_i occup08_k employment employed i4_k wage_123 I_income  
nolab_income) (s_idind idi_spouse s_occup08_k s_employment s_employed s_i4_k  
s_wage_123 S_income S_nolab_income)  
label variable s_idind "S Единый идентификационный номер индивида"  
label variable idi_spouse "S НОМЕР ИНДИВИДА - ИДЕНТИФИКАЦИОННЫЙ"  
label variable s_j60 "S Сколько всего денег в течение последних 30 дней вы получили"  
label variable s_occup08_k "S ПРОФЕССИОНАЛЬНАЯ ГРУППА - по коду J2COD08  
кратко"  
label variable s_employment "S статус занятости"  
label variable s_employed "S работает"  
label variable s_i4_k "S группы национальностей кратко"  
label variable s_wage_123 "S дефлир. ЗП на 3х работах за 30 дней"  
label variable S_income "S дефлир.инд.доходы (максимум из разных оценок) "  
label variable S_nolab_income "S дефлир нетруд доходы"  
save "C:\RLMS_work\seminar_6\data\spouse_5_16.dta", replace
```

\*29.3. Снова открыть рабочий файл и приклеить к нему данные по супругу\супруге по ключевым переменным **id\_w** **idi\_spouse**

```
use "C:\RLMS_work\seminar_6\data\ind_5_16_S6.dta", clear
```

```
merge m:1 id_w idi_spouse using "C:\RLMS_work\seminar_6\data\spouse_5_16.dta"
```

Result	# of obs.	
not matched	101,060	
from master	50,530	( _merge==1)
from using	50,530	( _merge==2)
matched	69,906	( _merge==3)

```
drop if _merge == 2  
drop _merge
```

```
tab id_w s_employment
```

	S статус	
НОМЕР	занятости	
ВОЛНЫ	неактивны безработн	есть люба   Total

1994 год	5,228	236	132	5,596
1995 год	4,841	217	186	5,244
1996 год	4,684	258	164	5,106
1998 год	4,471	566	195	5,232
2000 год	4,567	492	273	5,332
2001 год	4,973	475	298	5,746
2002 год	5,198	464	288	5,950
2003 год	5,306	424	306	6,036
2004 год	5,254	420	310	5,984
2005 год	5,139	395	262	5,796
2006 год	6,223	367	358	6,948
2007 год	6,237	416	283	6,936
Total	62,121	4,730	3,055	69,906

\*29.4. Дефлируем суммарные доходы семьи

```
gen double f14_d = f14 * deflat_16
label variable f14_d "дефл. суммарные доходы семьи"
format %9.2f f14_d
```

```
tabstat f14_d , statistics( mean count min max ) by(id_w) columns(statistics) format(%9.0f)
```

```
Summary for variables: f14_d
by categories of: id_w (НОМЕР ВОЛНЫ)
```

id_w	mean	N	min	max
1994 год	12888	8630	0	478121
1995 год	9909	8127	0	349614
1996 год	10081	7973	0	290919
1998 год	6220	8432	0	234989
2000 год	8606	8748	0	360366
2001 год	11659	9575	0	594066
2002 год	12557	10098	0	689414
2003 год	14172	10163	0	737631
2004 год	15757	10662	0	956281
2005 год	17881	9868	0	517157
2006 год	22169	11785	0	1056560
2007 год	21668	11451	0	320000
Total	14164	115512	0	1056560

\*Если в домохозяйстве 1 человек, индивидуальные доходы равны семейным. Если они меньше, заменим:

```
replace I_income = f14_d if (nfm ==1 & f14_d > I_income)
```

\*Если у респондента нет супруга\супруги (тогда идентификатор супруги\га – миссинг), то доходы супруга\супруга равны 0:

```
recode S_income (. = 0) if idi_spouse == .
label variable S_income "дефл. доходы супруга\супруги"
```

\*Сохраните рабочий файл

```
save "C:\RLMS_work\seminar_6\data\ind_5_16_S6.dta", replace
```

\*Сохраним файл под новым именем. Создадим агрегированную переменную «сумма доходов всех членов семьи», при этом получится датасет с тремя переменными, переименуем ее и сохраним в отдельный файл, а затем откроем снова наш рабочий файл.

```
save "C:\RLMS_work\seminar_6\data\sum_income.dta"
collapse (sum) I_income, by(id_w id_h)
rename I_income sum_income_all
save "C:\RLMS_work\seminar_6\data\sum_income.dta", replace
use "C:\RLMS_work\seminar_6\data\ind_5_16_S6.dta", clear
```

\*Приклеим эту переменную суммарных доходов к рабочему файлу

```
merge m:1 id_w id_h using "C:\RLMS_work\seminar_6\data\sum_income.dta"
```

```
Result                                     # of obs.
-----
not matched                               0
matched                                  120,436  (_merge==3)
-----
```

```
drop _merge
```

\*Создадим переменную, = максимуму из f14\_d sum\_income\_all

```
egen double fam_income = rowmax(f14_d sum_income_all)
label variable fam_income "дефл.суммарный доход, как максимум суммы дох.всех
членов или 14"
format %9.2f sum_income_all fam_income
```

\*Создадим вспомогательные переменные дохода респондента и супруга, умноженные на минус 1

```
gen double I_income_m = I_income * -1
gen double S_income_m = S_income * -1
```

```
egen double Other_income = rowtotal(fam_income I_income_m S_income_m ) if
fam_income!=. , missing
format %9.2f Other_income
label variable Other_income "дефл.доходы ост.членов семьи"
drop I_income_m S_income_m
replace Other_income = 0 if nfm ==1
recode Other_income (-1/1 =0)
```

\*Сохраните файл

\*Посмотрим описательные статистики. Обратите внимание, что средний доход «супруга\супруги» намного меньше, чем у респондента. А минимумы и максимумы – такие же. Попробуйте объяснить, почему.

```
tabstat I_income, statistics( mean count min max ) by(id_w) columns(statistics)
format(%9.0f)
tabstat S_income, statistics( mean count min max ) by(id_w) columns(statistics)
format(%9.0f)
tabstat Other_income, statistics( mean count min max ) by(id_w) columns(statistics)
format(%9.0f)
```

```
Summary for variables: I_income
```

by categories of: id\_w (HOMEР ВОЛНЫ)

id_w	mean	N	min	max
1994 год	4981	8660	0	178209
1995 год	4180	8126	0	111499
1996 год	4616	8112	0	155157
1998 год	2801	8513	0	154259
2000 год	3430	8908	0	926557
2001 год	4417	9829	0	584327
2002 год	4782	10283	0	253834
2003 год	5122	10407	0	151092
2004 год	5621	10472	0	956281
2005 год	6268	10138	0	487884
2006 год	7012	12184	0	341295
2007 год	7559	11874	0	166000
Total	5212	117506	0	956281

Summary for variables: S\_income  
by categories of: id\_w (HOMEР ВОЛНЫ)

id_w	mean	N	min	max
1994 год	3615	8661	0	178209
1995 год	3020	8176	0	111499
1996 год	3329	8122	0	155157
1998 год	2045	8461	0	154259
2000 год	2453	8823	0	926557
2001 год	2970	9727	0	280707
2002 год	3185	10175	0	253834
2003 год	3471	10304	0	151092
2004 год	3712	10305	0	204928
2005 год	4111	10049	0	243942
2006 год	4588	12068	0	341295
2007 год	4958	11841	0	110000
Total	3536	116712	0	926557

Summary for variables: Other\_income  
by categories of: id\_w (HOMEР ВОЛНЫ)

id_w	mean	N	min	max
1994 год	6875	8893	0	478121
1995 год	5050	8417	0	349614
1996 год	5453	8341	0	290919
1998 год	3388	8699	0	214144
2000 год	4604	9074	0	935085
2001 год	6226	10098	0	594066
2002 год	6506	10499	0	689414
2003 год	7186	10636	0	737631
2004 год	8652	10662	0	910339
2005 год	8783	10337	0	517157
2006 год	11728	12490	0	1007100
2007 год	10214	12290	0	320000
Total	7332	120436	0	1007100

Summary for variables: nolab\_income

by categories of: id\_w (НОМЕР ВОЛНЫ)

id_w	mean	N	min	max
1994 год	1293	8667	0	120617
1995 год	1086	8146	0	94490
1996 год	1046	8123	0	155157
1998 год	693	8520	0	62196
2000 год	920	8911	0	115502
2001 год	1326	9851	0	584327
2002 год	1290	10291	0	101534
2003 год	1312	10421	0	151092
2004 год	1352	10472	0	204928
2005 год	1701	10157	0	487884
2006 год	1772	12202	0	313320
2007 год	1888	11897	0	85000
Total	1344	117658	0	584327

Прологарифмируем доходы респондента, доходы супруга\супруги, и остальных членов семьи (+ 1, т.к. логарифм от 0 - не существует, логарифм от 1 = 0)

```
gen double lg_I_income = ln(I_income +1)
label variable lg_I_income "лог дефлир.своих доходов"
gen double lg_S_income = ln(S_income +1)
label variable lg_S_income "лог дефл.доходов супруга\супруги"
gen double lg_Other_income = ln(Other_income +1)
label variable lg_Other_income "лог дефл доходов остальных членов семьи"
gen double lg_nolab_income = ln(nolab_income +1)
label variable lg_nolab_income "лог дефл нетрудового дохода"
```

\*Сохраните ду-файл с кодом.