Занятие 1. Знакомство с RLMS и STATA. Описательные характеристики и трансформация номинальных переменных.

Общая рекомендация ко всем выполняемым вами заданиям:

1) Сохраняйте исходные файлы под новым именем, чтобы работать с ними.

2) Открывайте и сохраняйте файл аутпута.

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.

Исходный файл.

Данные 24й волны (индивидуальные)

r15iall_32.dta

2. Загрузка файлов в вашу директорию. Знакомство с анкетой и файлами.

2.1. Создать на жестком диске компьютера (директория C: или D:) свою директорию (C:\RLMS_work\). Имейте ввиду, что на директории C: файлы после перезагрузки компьютера (в компьютерном классе) не сохраняются. НИКОГДА не используйте рабочий стол для создания рабочей папки или тем более сохранения туда ваших файлов.

2.2. Переписать туда нужные для работы файлы – Скачать данные **seminar_1.zip** из LMS, и раззиповать. У вас получится директория C:\RLMS_work\seminar_1

В ней находятся папки:

Codebook – codebook для данных РМЭЗ

Data – файлы данных для первого занятия

Quest 15 – вопросники 15й волны

А также файлы с текстом занятия. Из файла в формате pdf переносить команды (при необходимости) будет сложнее (как минимум, мешают номера строк).

2.3. Перед началом преобразований в файлах, ОБЯЗАТЕЛЬНО переименуйте их, чтобы сохранить исходные файлы в неизменном виде. Это поможет в случае ошибок. Запоминайте, какие названия вы дали файлам, старайтесь сделать их разумными (например, использовать букву h в названии семейного файла, а i – в названии индивидуального файла; обязательно указывать волну или волны, если их несколько; возможно добавление вашей фамилии, номера семинара и т.д.).

2.4. Работа проводится ТОЛЬКО С ФАЙЛАМИ, ПЕРЕПИСАННЫМИ В ВАШУ ДИРЕКТОРИЮ и ПЕРЕИМЕНОВАННЫМИ! Созданные вами файлы (например, синтаксиса) в случае необходимости сохраняйте (на флешках, пересылайте себе по почте и т.д.).

2.5. ОБЯЗАТЕЛЬНО сохраняйте преобразования, сделанные вами, а также файлы кода и аутпута, в той же директории (для возможности их сдачи, при необходимости, в конце занятия).

2.6. Есть три 3 анкеты: семейная, анкета для взрослых, анкета для детей, а также «анкета» населенного пункта (см. папку quest 15). Три типа файлов: семейные (домохозяйственные) данные и индивидуальные данные, которые можно скачать с сайта, а также данные с характеристиками населенного пункта, которые можно получить по запросу.

2.7. Названия всех переменных в любом файле: первая буква – номер волны (5- a, 6 – b, 7- c, 8 – d, 9- e, 10 – f, 11 – g, 12 – h, 13 – i, 14 – j, 15 - k и т.д.), вторая буква – номер раздела анкеты (a, b, c, d, e, f, h, i, j, l, m, n, o, k), цифры – номер вопроса в разделе.

2.8. Имейте, пожалуйста, ввиду, что в анкете и в файле данных номер одного и того же вопроса может отличаться, так как в анкете нумерация делается «для респондента», а в файлах одним

и тем же вопросам в разных волнах дается один и тот же номер (для возможности сопоставления и склеивания).
2.9. В дальнейшем имена переменных мы будем «называть» со второй буквы, т.е. имени раздела, т.к. первая буква меняется от волны к волне. Так, одна и та же переменная j1 будет иметь имена aj1 в пятой волне и hj1 в 12 волне. Первая буква м.б. взята в тексте заданий в скобки, это значит, что она меняется в зависимости от волны: (h)j1.

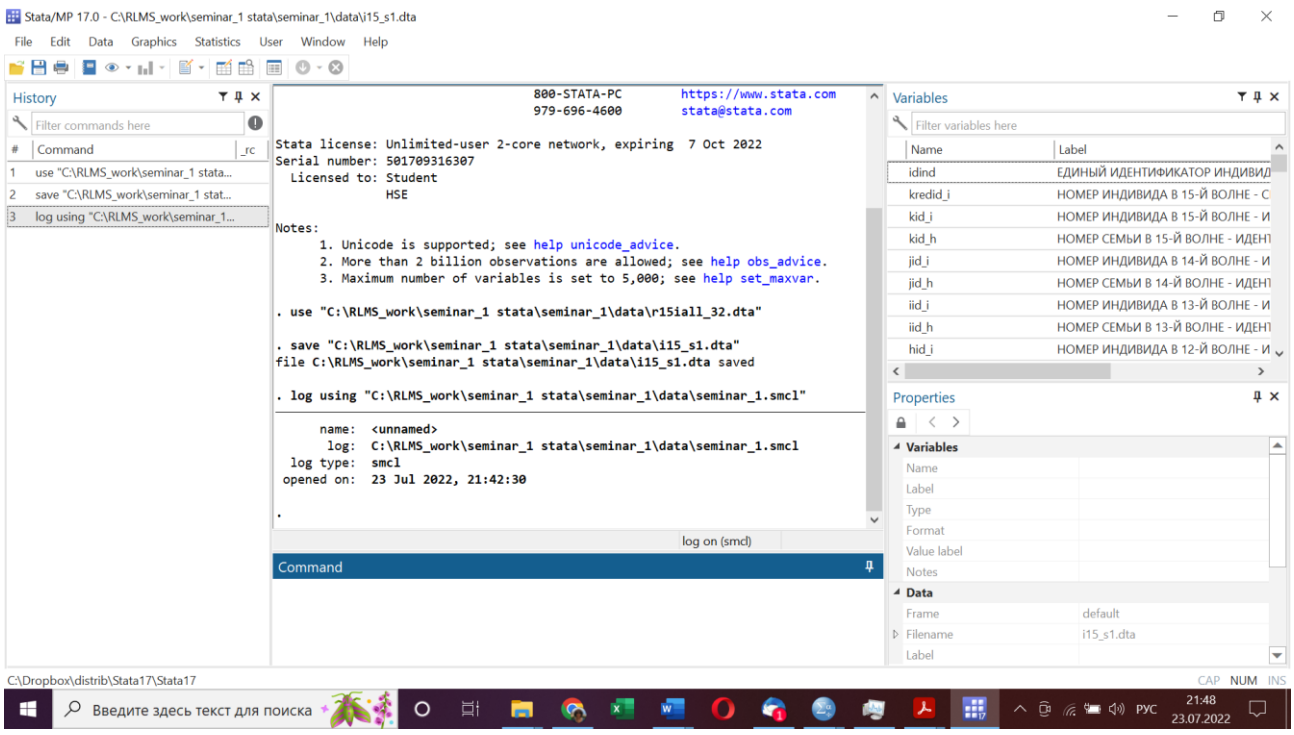
3. Начало работы с программой STATA.

Сайт со ссылками на ресурсы для STATA :
http://www.ats.ucla.edu/stat/stata/

Простой учебник по регрессионному анализу и STATA (6-8 версии)
Коленников С. Прикладной эконометрический анализ в статистическом пакете Stata. РЭШ, 2000-2003 (ест в материалах занятия коленников Stata6Ec.pdf).
Основные приемы работы с пакетом, описанные ниже, приводятся на основе этого учебника.
В материалах занятия также руководство пользователя STATA-17.

3.1. Установка и запуск STATA
Обычно Stata устанавливается в каталог C:/stata, если при установке не было явно указано иное. Подавляющее большинство собственно статистических задач выполняется внешними программами с расширением .ado, находящимися в каталоге C:/stata/ado и его подкаталогах.
При запуске Stata устанавливает ряд внутренних параметров, таких, как объем выделяемой памяти, и некоторые другие (о них можно узнать в [R] limits или в подсказке help limits).

3.2. Интерфейс и команды STATA
Откройте программу STATA.



STATA использует в работе несколько окон: окно ввода команд (Command), окно вывода результатов (Stata Results), окно истории, или предыдущих команд (History), окно переменных

(Variables), окно свойств переменных (Properties), окно поиска и помощи (Help) – открывается из меню, графический экран (Graph) – открывается при создании графика, окно файла-протокола Viewer, или log-файла (smcl) – открывается из меню. Можно также вызвать окна просмотра данных (Stata Browser) или редактирования данных (Stata Editor), а также редактор программ (Stata Do-file Editor).
Переключаться между окнами можно, тыкаясь мышкой в любое место на нужном окне, либо через меню Windows.

В STATA для поиска нужной информации проще всего воспользоваться меню Help, в котором имеются подменю Search (поиск по ключевым словам, например, Durbin Watson statistic) и Stata Command (файл помощи по конкретной команде STATA).
Команды можно генерировать при помощи меню (они автоматически сохраняются в окне COMMAND), набирать вручную или при помощи команды «копировать-вставить». Есть также специальные исполняемые файлы синтаксиса – можно запустить целый файл, а не отдельную команду.
Команды STATA, как правило, имеют следующий вид:
команда [список переменных] [if условие] [in диапазон] [using имя файла] [[веса]], [опции]
Список переменных может состоять из одной переменной (например, если нужно получить сводные статистики или построить гистограмму), из двух (расчет корреляций или построение диаграммы рассеяния) и более (регрессии, графики со многими переменными). Условия if и in выделяют те наблюдения, для которых необходимо провести анализ. Если команда предполагает работу с файлами (чтение, объединение и т.п.), то имя файла, с которым необходимо провести указанные действия, передается в конструкции using. Если разным наблюдениям необходимо придать разные веса, то для этого используется конструкция типа [weight=выражение] (см. help weights; квадратные скобки являются элементами синтаксиса и обязательны).
Многие команды Stata позволяют ограничить свое действие на определенные наблюдения. Делается это с помощью условных модификаторов [if условие] [in диапазон]. Условие, задаемое под if, это логическое выражение, в котором могут использоваться операторы отношений > ("больше"), < ("меньше"), >= ("больше или равно"), <= ("меньше или равно"), = = ("равно", двойной знак использован для того, чтобы не спутать с операцией присвоения), | = или ~ = ("не равно"); логические операции & ("и"), | ("или"), | или ~ ("не"), указание на текущее наблюдение _n и на последнее _N, обычные операции и функции, а также скобки для указания приоритета. Оператор in указывает диапазон наблюдений вида начало/конец, где в качестве конца диапазона может быть использовано последнее наблюдение, обозначаемое латинской "эл"(l) или как ..1.

3.3. Полезные ресурсы

https://www.stata.com/links/resources-for-learning-stata/ Resources for learning Stata
https://www.youtube.com/user/statacorp - Stata's YouTube Channel

3.4. Загрузка файлов.
Данные для работы в пакете имеют расширение **.dta** ; файлы других форматов (Excel, SAS, SPSS, Statistica и т.п.) необходимо предварительно сохранить в виде текста (с разделением данных запятыми, табуляциями, или в фиксированном формате), либо воспользоваться внешними средствами для конвертации данных. В комплект поставки Professional Stata входит чрезвычайно полезная Windows-утилита StatTransfer (http://www.stattransfer.com), позволяющая преобразовывать данные между двумя десятками различных форматов.
Данные лучше положить в папку, в пути к которой нет РУССКИХ букв – программа плохо их читает в синтаксисе.

Прежде чем начать работу с данными, нам нужно распаковать их. Это индивидуальный и семейный фал 15й волны.

Для версий ниже 16: в окне команд набрать и выполнить команду (команды можно копировать и вставлять в нужное окно):
set more off
Она отключает построчный вывод результатов.

Для загрузки данных нужно воспользоваться командой меню: FILE > OPEN и в браузере найти нужный файл с расширением **.dta.**
r15iall_32.dta (индивидуальный файл данных 15 волны).
Появится примерно такая команда:
use "C:\RLMS_work\seminar_1\data\r15iall_32.dta"

Для перехода в окно данных (просмотр и редактирование) можно использовать соответствующую иконку. В меню DATA есть возможность генерировать команды для склейки файлов (добавление переменных или кейсов).
Для сохранения трансформированных данных можно использовать команду или значок «сохранить», изменив имя файла, например на i15_s1 (инд.данные 15 волны, семинар 1).
При этом генерируется команда (с точностью по пути):

save "C:\RLMS_work\seminar_1\data\i15_s1.dta"

3.5. Окно результатов. Варианты вывода и сохранения результатов. Удобная команда outreg2.

До начала анализа нужно создать (или открыть) файл результатов (аутпут). (Возможно также непосредственно копировать и переносить в Excel, например, результаты из окна RESULTS).
Из меню: FILE ==> LOG ==> BEGIN. При этом получится команды примерно такая:

log using имя файла, [append | replace]
log using "C:\RLMS_work\seminar_1\data\seminar_1.smcl"

ВАЖНО!!!!
Сразу после открытия файла аутпута, наберите команду: звездочка (*) Ваша фамилия – номер семинара, такая команда воспринимается как комментарий – тем самым вы подпишете ваш аутпут и сможете сдать его в качестве отчета за работу на семинаре:

*** Рощина - семинар 1**

После этой команды все, что Stata выводит в окно результатов, будет записано в указанный файл (добавляя либо перезаписывая этот файл, в соответствии с опциями append либо replace, если такой файл существует). log off временно прекращает запись в файл, log on возобновляет запись в файл, log close прекращает запись и закрывает файл.
log on | off |close

Закрыть файл результатов (команда через меню):
log close

Если нужно продолжить уже имеющийся файл результатов (аутпут)
log using " C:\RLMS_work\seminar_1\data\seminar_1.smcl", append

Команды, связанные с log-файлом, продублированы на панели инструментов Stata кнопочкой со светофором. Log-файлы лучше всего печатать непосредственно из Stata, поскольку Stata

умеет автоматически приукрашивать текст (выделяя полужирным шрифтом команды, проставляя даты и т.п.).
Log-файл можно также открыть через меню – FILE > LOG > VIEW.
Или FILE > VIEW
Все, что содержится в Log-файле, после того, как он открыт, можно скопировать и вставить в Word или если скопировать в виде таблицы, то есть в Excel.

Есть еще один вариант сохранения статистических результатов исследований - прекрасная пользовательская команда outreg2, которая записывает результаты регрессий (и не только) в отдельный текстовый файл в соответствии с принятыми в статистической и эконометрической литературе обозначениями: столбцы коэффициентов со стандартными ошибками в скобках, число наблюдений, статистика R2 и прочие статистики. Этот модуль требует отдельной установки, см. stb, help stb – описание ресурсов.

Установка полезного приложения, записывающего результаты в удобном виде в текстовый файл - outreg2. (УСТАНАВЛИВАЕМ**)**
Набрав команду:

findit outreg2
Вы откроете окно помощи, где в разделе

```
outreg2 from http://fmwww.bc.edu/RePEc/bocode/o
    'OUTREG2': module to arrange regression outputs into an illustrative table
    / outreg2 provides a fast and easy way to produce an illustrative / table
    of regression outputs. The regression outputs are produced / piecemeal and
    are difficult to compare without some type of / rearrangement. outreg2
```

Будет ссылка на страницу, с которой можно установить эту опцию. Установив опцию и "help" к ней, вы можете пользоваться командой.

```
INSTALLATION FILES                                  (click here to install)
      outreg2.ado
      outreg2_prf.ado
      outreg2.hlp
      ../s/shellout.ado
      ../s/shellout.hlp
      ../s/seeout.ado
      ../s/seeout.hlp
```

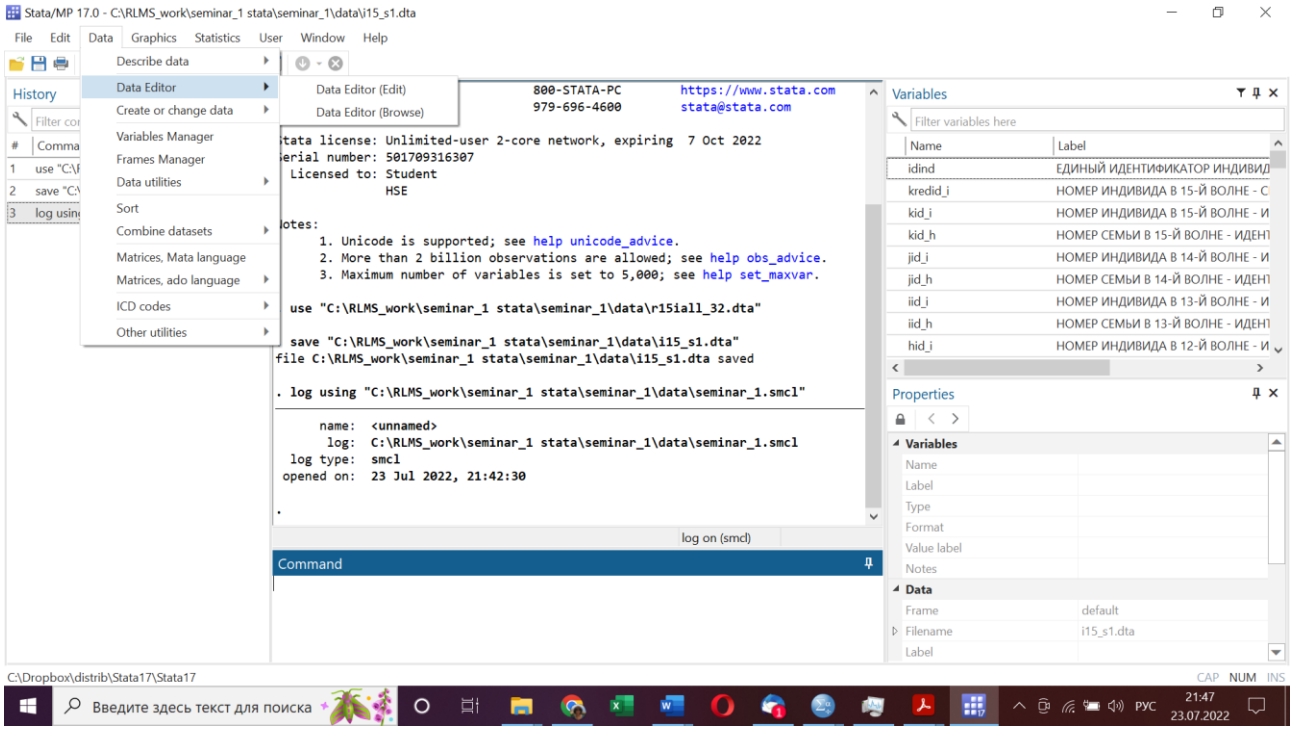
Другой вариант удобного сохранения результатов регрессий. **(**НЕ УСТАНАВЛИВАЕМ, по желанию – самостоятельно дома**)**
Надо скачать модуль - можно сделать в любом месте и на любом компьютере.
Команды для модуля:
ssc install rd
ssc install estout

Далее в Stata вы оцениваете модель и пишете команды сохранить результаты после каждой модели, где
model_x - название вашей модели
xxx.rtf - название word file где будет сохранена табличка со всеми запомненными в эту сессию моделями.

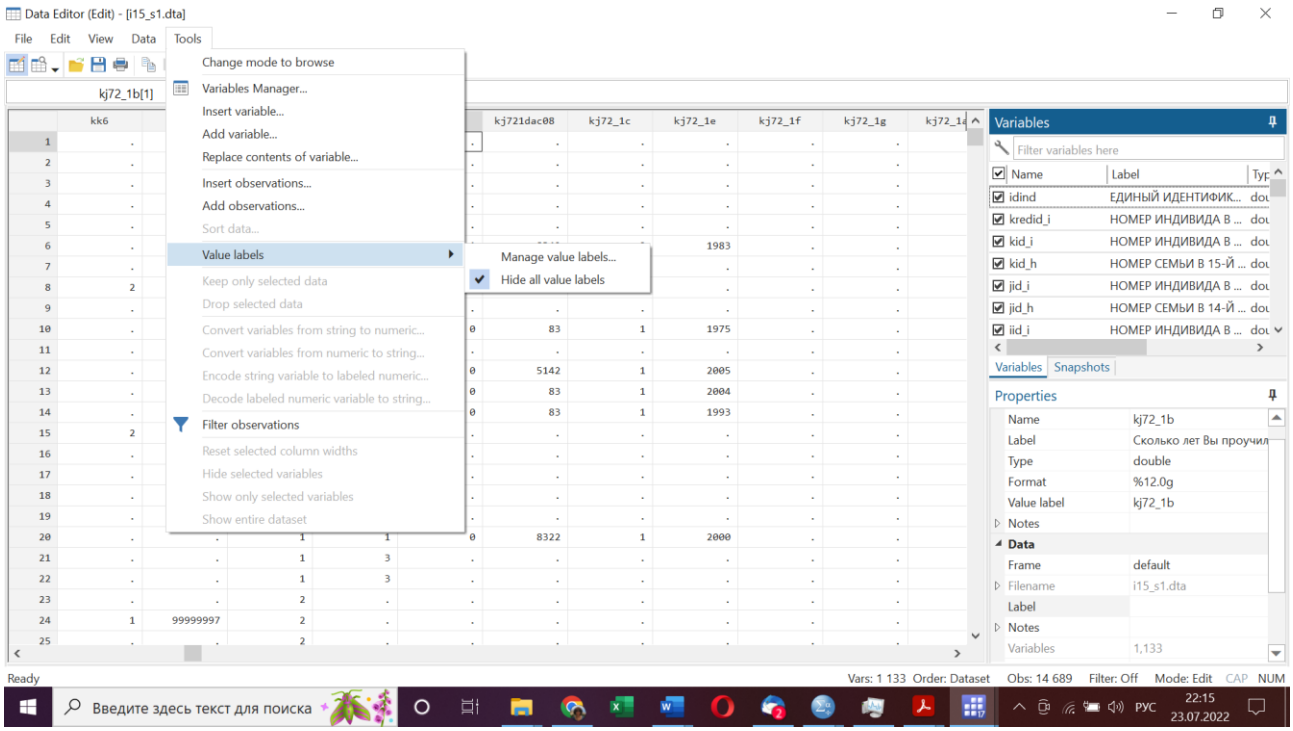
eststo model_x
esttab using xxx.rtf
esttab using "путь\xxx.rtf"
esttab using "путь\xxx.rtf", append

3.6. Переход в режим данных – просмотр и редактирование (описание)

3.6.1. Перейти в таблицу с непосредственно данными можно при помощи меню (DATA > DATA EDITOR > DATA EDITOR (Edit) \ DATA EDITOR (Browse))
Либо нажав на соответствующие иконки на панели.



По умолчанию выводятся на экран лейблы значений. Для просмотра значений нужно выбрать опцию «скрыть» (Tools > Value labels > Hide all value labels).



В окне «Properties» вы можете редактировать свойства переменных.

3.6.2. Вернемся в основное окно программы, закрыв окно редактирования данных (и переменных).

Вы видите в правом верхнем углу список переменных. Их можно переносить в окно команд снизу, а также редактировать, «открыв» значок замка в окне «Properties».

Давайте посмотрим на основные переменные индивидуального файла.

Для вывода codebook можно использовать меню:
Data > Describe data > Describe data contens (codebook) и выбрать нужные переменные (с **idind** по **kh8b**). Вы получите команду:

codebook idind kredid_i kid_i kid_h jid_i jid_h iid_i iid_h hid_i hid_h gid_i gid_h fid_i fid_h eid_i eid_h did_i did_h cid_i cid_h bid_i bid_h aid_i aid_h k_origsm k_inwgt psu region status popul k_int_y k_born_y k_child k_adult k_marst k_occup08 k_educ k_diplom k_diplom_1 site ssu kh3 kh4 kh4_1 kh5 k_born_m kh6 k_age kh7_1 kh7_2 kh8a kh8b

А также аутпут для каждой переменной такого вида:

```
-------------------------------------------------------------------------------
idind                         ЕДИНЫЙ ИДЕНТИФИКАТОР ИНДИВИДА ДЛЯ ВСЕХ ВОЛН RLMS
-------------------------------------------------------------------------------

                  Type: Numeric (double)

                 Range: [1,35543]                     Units: 1
         Unique values: 14,689                    Missing .: 0/14,689

                  Mean: 16370.2
             Std. dev.: 8912.98

           Percentiles:     10%       25%       50%       75%       90%
                           3093      8242     17659     24843     27170
```

Для каждой переменной мы видим значения:
Type: Numeric (double) – количественная
Range: [1,35543] – диапазон значений (запятая – это не разделитель десятичных знаков. Вообще в STATA для разделителя десятичных знаков используется точка, а для разделителя «тысяч» - запятая, как и в большинстве зарубежных журналов).
Unique values: 14,689 – количество различающихся уникальных значений (в данном случае соотвествует количеству кейсов, так как у каждого индивида свой уникальный идентификатор)
Missing .: 0/14,689 – миссинги отсутствуют
Mean: 16370.2 и Std. dev.: 8912.98 в данном случае не имеют смысла, как и Percentiles

Другой пример:

```
k_origsm                          АДРЕС РЕПРЕЗЕНТАТИВНОЙ ВЫБОРКИ? 15-Я ВОЛНА
-------------------------------------------------------------------------------

                  Type: Numeric (double)
                 Label: k_origsm

                 Range: [0,1]                         Units: 1
         Unique values: 2                         Missing .: 0/14,689

            Tabulation: Freq.   Numeric  Label
```

```
   3,978          0  Нет, не входит в
                     репрезентативную выборку 15-
  10,711          1  Да, адрес репрезентативной
                     выборки 15-ой вол
```

Здесь диапазон значений от 0 до 1, два уникальных значения.
Также есть распределение значений:
0 Нет, не входит в репрезентативную выборку 15- (лейбл оборван) – 3 978 кейсов (запятая разделяет тысячи)
1 - Да, адрес репрезентативной выборки 15-ой вол… - 10 711 кейсов.

Каждая семья и каждый индивид в каждой волне исследования имеет свой уникальный идентификационный номер. В каждом файле есть идентификационные переменные за все предыдущие волны, позволяющие сливать файлы разных раундов (напр., присоединять к файлу одного года данные из файлов других лет для отслеживания динамики) и сопоставлять данные по семьям и индивидам. В семейных файлах есть только идентификаторы семейные, в индивидуальных – и индивидуальные, и семейные.

Раунд		Идентификаторы	
	Год		
5	1994	aid_h	aid_i
6	1995	bid_h	bid_i
7	1996	cid_h	cid_i
8	1998	did_h	did_i
9	2000	eid_h	eid_i
10	2001	fid_h	fid_i
11	2002	gid_h	gid_i
12	2003	hid_h	hid_i
13	2004	iid_h	iid_i
14	2005	jid_h	jid_i
15	2006	kid_h	kid_i
16	2007	lid_h	lid_i
17	2008	mid_h	mid_i
18	2009	nid_h	nid_i
19	2010	oid_h	oid_i
20	2011	pid_h	pid_i
21	2012	qid_h	qid_i
22	2013	rid_h	rid_i
23	2014	sid_h	sid_i
24	2015	tid_h	tid_i
25	2016	uid_h	uid_i
26	2017	vid_h	vid_i
27	2018	wid_h	wid_i
28	2019	xid_h	xid_i
29	2020	yid_h	yid_i
30	2021	zid_h	zid_i
все			idind

Идентификационные переменные (15я волна):

kredid_i - НОМЕР ИНДИВИДА В 15-Й ВОЛНЕ - СПЛОШНАЯ НУМЕРАЦИЯ, эта переменная практически никогда не нужна.

idind - единый уникальный номер индивида для всех волн, не меняется от волны к волне – сплошная нумерация людей, хотя бы раз участвовавших в исследовании, по мере поступления их в панель. Эту переменную ОБЯЗАТЕЛЬНО нужно сохранять во всех файлах, которые вы будете генерировать (с более коротким списком переменных).

kid_h …. fid_h eid_h did_h cid_h bid_h aid_h – идентификаторы домохозяйства во всех волнах, которые были до данной волны (может меняться, если изменялась кодировка, или если человек изменял домохозяйство, т.е. переезжал). Так как домохозяйства имеют свойство разделяться, а также из-за изменений принципа кодировки, НЕ СУЩЕСТВУЕТ «уникального» идентификатора домохозяйства, и в семейных файлах рекомендуется сохранять идентификаторы за ВСЕ предыдущие волны. Идентификатор данной волны (**kid_h**) нужен в любом индивидуальном файле, который вы будете генерировать; остальные – если вы предполагаете использовать данные о семье из предыдущих волн.

kid_i … fid_i eid_i did_i cid_i bid_i aid_i – идентификаторы индивидов во всех волнах, которые были до данной волны, основанные на информации о месте жительства, номере семьи и номере индивида в семье. Редко бывают нужны.

ID_W - № раунда (волны, года) – эту переменную нужно будет создать, если данные за один год. Она всегда присутствует в лонгитюдных файлах (за несколько лет).

В лонгитюдных данных уникальными идентификаторами КЕЙСА являются два идентификатора: **idind + ID_W**. Для домохозяйства – идентификатор в каждой волне, а также условный идентификатор **ID_H + ID_W** (об этом мы поговорим позже).

k_origsm - АДРЕС РЕПРЕЗЕНТАТИВНОЙ ВЫБОРКИ? 15-Я ВОЛНА (переменная-фильтр для отбора репрезентативных данных\кейсов)

k_inwgt - Постстратификационный вес для данного индивида в 15-ой волне (переменная для взвешивания)

psu - ПЕРВИЧНАЯ ЕДИНИЦА ОТБОРА (по сути, регион, т.е. область + возможно насел.пункт в области, если в области опрашивали людей в двух точках)

region – регион (отличается тем, что точно соответствует области, в которой был опрос)

status - ТИП НАСЕЛЕННОГО ПУНКТА (обл.центр, город, село, пгт)

popul - ЧИСЛЕННОСТЬ НАСЕЛЕНИЯ (в населенном пункте)

k_int_y - ГОД ПРОВЕДЕНИЯ ИНТЕРВЬЮ

k_born_y - ГОД РОЖДЕНИЯ РЕСПОНДЕНТА (м.б. скорректированным на основе ответов в других волнах и отличаться от переменной **kh6**)

k_child - Есть детский вопросник 15-ой волны?

k_adult - Есть взрослый вопросник 15-ой волны?

k_marst - СЕМЕЙНОЕ ПОЛОЖЕНИЕ - 15 ВОЛНА (сконструированная переменная)

k_occup08 - ПРОФЕССИОНАЛЬНАЯ ГРУППА - 15 ВОЛНА - по коду kj2cod08 (сконструированная переменная)

k_educ - ОБРАЗОВАНИЕ (ПОДРОБНО): старше 14 лет - 15 ВОЛНА (сконструированная переменная)

k_diplom - ЗАКОНЧЕННОЕ ОБРАЗОВАНИЕ (ГРУППА) - 15 ВОЛНА (сконструированная переменная)

k_diplom_1 - НАИБОЛЕЕ ВЕРОЯТНОЕ ЗАКОНЧЕННОЕ ОБРАЗОВАНИЕ (ГРУППА) - 15 ВОЛНА (скорректированное на основе ответа респондента в других волнах)

site – НОМЕР НАСЕЛЕННОГО ПУНКТА (это не совсем верно)

ssu – Вторичная единица отбора (номер опросного участка в населенном пункте)

kh3 – номер семьи

kh4 - номер члена семьи

kh4_1 - Респондент ранее участвовал(а) в исследовании?

kh5 - Пол респондента

k_born_m - МЕСЯЦ РОЖДЕНИЯ РЕБЕНКА - 15 ВОЛНА (!!!! Это не «месяц рождения ребенка у респондента», а месяц рождения, если респондент – ребенок!!!!)

kh6 - Год рождения респондента 15 ВОЛНА (=kj69.9c)

k_age - Количество полных лет (сконструированная переменная; учитывает месяц

опроса)

ОНИ ДОЛЖНЫ ПРИСУТСТВОВАТЬ В КАЖДОМ ВАШЕМ ФАЙЛЕ!

Вам также могут пригодиться переменные:
kh7_1 - Дата проведения интервью: число
kh7_2 - Дата проведения интервью: месяц
kh8a - Интервью продолжалось (часов)
kh8b - Интервью продолжалось (минут)

3.6.3. Можно также использовать удобную команду **describe**:

Она позволяет вывести описание данных и переменных: формат, метки и т. п. Эта команда показывает также количество наблюдений и переменных, изменялись ли данные с момента последнего сохранения, по каким переменным отсортированы наблюдения. Можно указать файл, находящийся на жестком диске.
describe [переменные | using имя файла], [short]

describe idind kredid_i kid_i kid_h jid_i jid_h iid_i iid_h hid_i hid_h gid_i gid_h fid_i fid_h eid_i eid_h did_i did_h cid_i cid_h bid_i bid_h aid_i aid_h k_origsm k_inwgt psu region status popul k_int_y k_born_y k_child k_adult k_marst k_occup08 k_educ k_diplom k_diplom_1 site ssu kh3 kh4 kh4_1 kh5 k_born_m kh6 k_age kh7_1 kh7_2 kh8a kh8b

Получим аутпут (ниже). Обратите внимание, что «value label» выдается в виде «имени», присвоенного определенной переменной (оно может быть одно и то же у разных переменных):

```
Variable      Storage   Display    Value
    name         type    format    label      Variable label
------------------------------------------------------------------------------
idind           double  %12.0g                ЕДИНЫЙ ИДЕНТИФИКАТОР ИНДИВИДА ДЛЯ
                                                ВСЕХ ВОЛН RLMS
kredid_i        double  %12.0g                НОМЕР ИНДИВИДА В 15-Й ВОЛНЕ -
                                                СПЛОШНАЯ НУМЕРАЦИЯ
kid_i           double  %12.0g                НОМЕР ИНДИВИДА В 15-Й ВОЛНЕ -
                                                ИДЕНТИФИКАЦИОННЫЙ
kid_h           double  %12.0g                НОМЕР СЕМЬИ В 15-Й ВОЛНЕ -
                                                ИДЕНТИФИКАЦИОННЫЙ
jid_i           double  %12.0g                НОМЕР ИНДИВИДА В 14-Й ВОЛНЕ -
                                                ИДЕНТИФИКАЦИОННЫЙ
jid_h           double  %12.0g                НОМЕР СЕМЬИ В 14-Й ВОЛНЕ -
                                                ИДЕНТИФИКАЦИОННЫЙ
iid_i           double  %12.0g                НОМЕР ИНДИВИДА В 13-Й ВОЛНЕ -
                                                ИДЕНТИФИКАЦИОННЫЙ
iid_h           double  %12.0g                НОМЕР СЕМЬИ В 13-Й ВОЛНЕ -
                                                ИДЕНТИФИКАЦИОННЫЙ
hid_i           double  %12.0g                НОМЕР ИНДИВИДА В 12-Й ВОЛНЕ -
                                                ИДЕНТИФИКАЦИОННЫЙ
hid_h           double  %12.0g                НОМЕР СЕМЬИ В 12-Й ВОЛНЕ -
                                                ИДЕНТИФИКАЦИОННЫЙ
gid_i           double  %12.0g                НОМЕР ИНДИВИДА В 11-Й ВОЛНЕ -
                                                ИДЕНТИФИКАЦИОННЫЙ
gid_h           double  %12.0g                НОМЕР СЕМЬИ В 11-Й ВОЛНЕ -
                                                ИДЕНТИФИКАЦИОННЫЙ
fid_i           double  %12.0g                НОМЕР ИНДИВИДА В 10-Й ВОЛНЕ -
                                                ИДЕНТИФИКАЦИОННЫЙ
fid_h           double  %12.0g                НОМЕР СЕМЬИ В 10-Й ВОЛНЕ -
                                                ИДЕНТИФИКАЦИОННЫЙ
eid_i           double  %12.0g                НОМЕР ИНДИВИДА В 9-Й ВОЛНЕ -
                                                ИДЕНТИФИКАЦИОННЫЙ
eid_h           double  %12.0g                НОМЕР СЕМЬИ В 9-Й ВОЛНЕ -
                                                ИДЕНТИФИКАЦИОННЫЙ
```

did_i	double	%12.0g		НОМЕР ИНДИВИДА В 8-Й ВОЛНЕ - ИДЕНТИФИКАЦИОННЫЙ
did_h	double	%12.0g		НОМЕР СЕМЬИ В 8-Й ВОЛНЕ - ИДЕНТИФИКАЦИОННЫЙ
cid_i	double	%12.0g		НОМЕР ИНДИВИДА В 7-Й ВОЛНЕ - ИДЕНТИФИКАЦИОННЫЙ
cid_h	double	%12.0g		НОМЕР СЕМЬИ В 7-Й ВОЛНЕ - ИДЕНТИФИКАЦИОННЫЙ
bid_i	double	%12.0g		НОМЕР ИНДИВИДА В 6-Й ВОЛНЕ - ИДЕНТИФИКАЦИОННЫЙ
bid_h	double	%12.0g		НОМЕР СЕМЬИ В 6-Й ВОЛНЕ - ИДЕНТИФИКАЦИОННЫЙ
aid_i	double	%12.0g		НОМЕР ИНДИВИДА В 5-Й ВОЛНЕ - ИДЕНТИФИКАЦИОННЫЙ
aid_h	double	%12.0g		НОМЕР СЕМЬИ В 5-Й ВОЛНЕ - ИДЕНТИФИКАЦИОННЫЙ
k_origsm	double	%12.0g	k_origsm	АДРЕС РЕПРЕЗЕНТАТИВНОЙ ВЫБОРКИ? 15-Я ВОЛНА
k_inwgt	double	%12.0g		Постстратификационный вес для данного индивида в 15-ой волне
psu	double	%12.0g	psu	ПЕРВИЧНАЯ ЕДИНИЦА ОТБОРА
region	double	%12.0g	region	РЕГИОН
status	double	%12.0g	status	ТИП НАСЕЛЕННОГО ПУНКТА
popul	double	%12.0g		ЧИСЛЕННОСТЬ НАСЕЛЕНИЯ
k_int_y	double	%12.0g		ГОД ПРОВЕДЕНИЯ ИНТЕРВЬЮ
k_born_y	double	%12.0g		ГОД РОЖДЕНИЯ РЕСПОНДЕНТА
k_child	double	%12.0g	k_child	Есть детский вопросник 15-ой волны?
k_adult	double	%12.0g	k_adult	Есть взрослый вопросник 15-ой волны?
k_marst	double	%12.0g	k_marst	СЕМЕЙНОЕ ПОЛОЖЕНИЕ - 15 ВОЛНА
k_occup08	double	%80.0g	k_occup08	ПРОФЕССИОНАЛЬНАЯ ГРУППА - 15 ВОЛНА - по коду kj2cod08
k_educ	double	%12.0g	k_educ	ОБРАЗОВАНИЕ (ПОДРОБНО): старше 14 лет - 15 ВОЛНА
k_diplom	double	%12.0g	k_diplom	ЗАКОНЧЕННОЕ ОБРАЗОВАНИЕ (ГРУППА) - 15 ВОЛНА
k_diplom_1	double	%12.0g	k_diplom_1	НАИБОЛЕЕ ВЕРОЯТНОЕ ЗАКОНЧЕННОЕ ОБРАЗОВАНИЕ (ГРУППА) - 15 ВОЛНА
site	double	%12.0g		НОМЕР НАСЕЛЕННОГО ПУНКТА
ssu	double	%12.0g		Вторичная единица отбора
kh3	double	%12.0g		Номер семьи
kh4	double	%12.0g		Номер члена семьи
kh4_1	double	%12.0g	kh4_1	Респондент ранее участвовал(а) в исследовании?
kh5	double	%12.0g	kh5	Пол респондента
k_born_m	double	%12.0g	k_born_m	МЕСЯЦ РОЖДЕНИЯ РЕБЕНКА - 15 ВОЛНА
kh6	double	%12.0g		Год рождения респондента 15 ВОЛНА (=kj69.9c)
k_age	double	%12.0g	k_age	Количество полных лет
kh7_1	double	%12.0g	kh7_1	Дата проведения интервью: число
kh7_2	double	%12.0g	kh7_2	Дата проведения интервью: месяц
kh8a	double	%12.0g	kh8a	Интервью продолжалось (часов)
kh8b	double	%12.0g		Интервью продолжалось (минут)

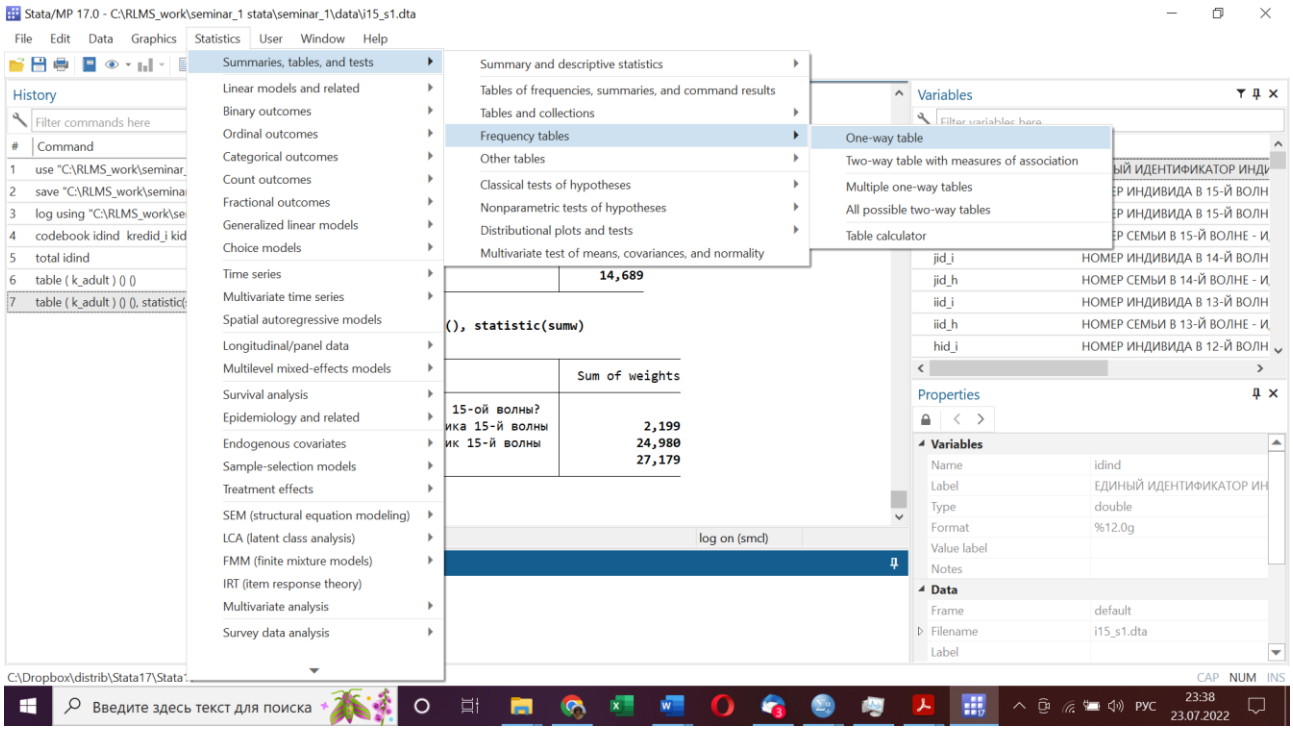
4. Описательные характеристики.

4.1. Распределения переменных.

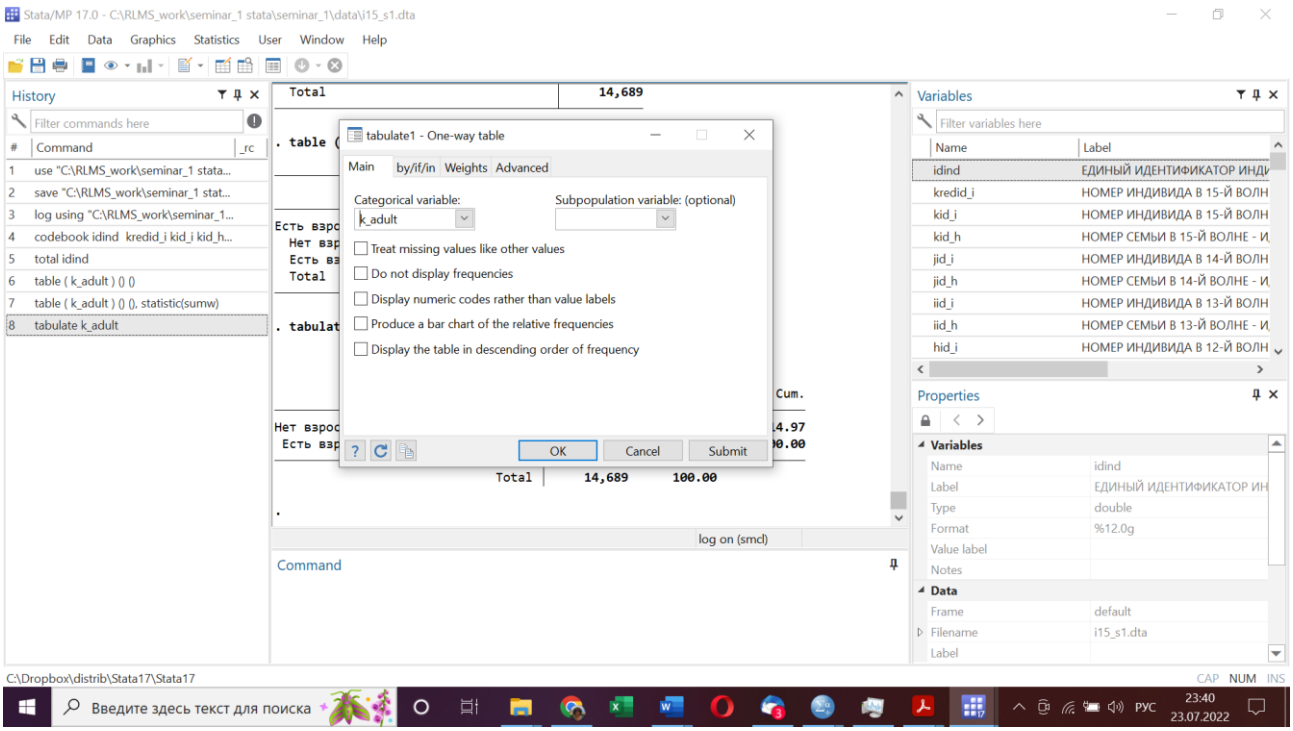
В команде **codebook** мы для каждой переменной видели простые распределения значений (если их не было слишком много).

Посмотрим распределение дихотомической переменной «k_adult» (взрослая или детская анкеты)

Из меню: Statistics > Summaries, tables, and tests > Frequency tables > One-way table



И выберем нужную переменную:



В результате получим следующую команду:

tabulate k_adult

и следующий аутпут:

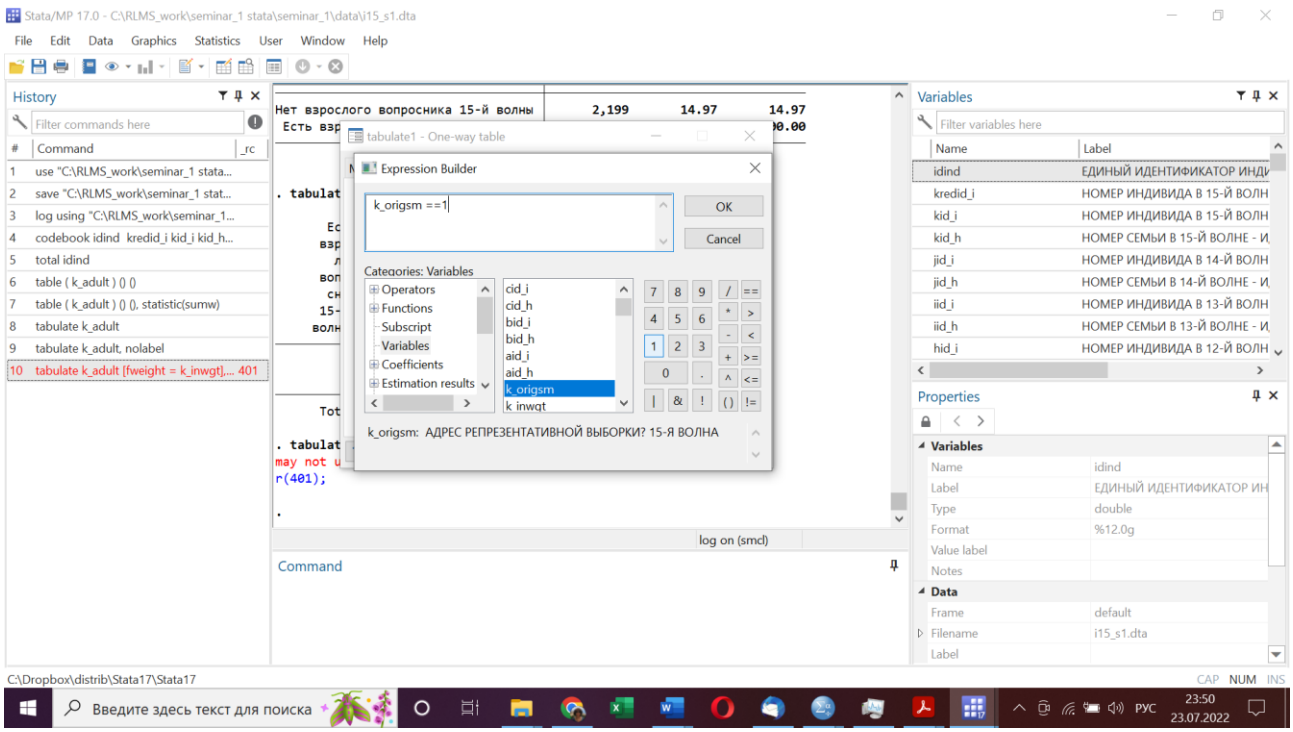
```
                      Есть взрослый |
                     вопросник 15-ой |
                              волны? |      Freq.     Percent        Cum.
------------------------------------+-----------------------------------
Нет взрослого вопросника 15-й волны |      2,199       14.97       14.97
 Есть взрослый вопросник 15-й волны |     12,490       85.03      100.00
------------------------------------+-----------------------------------
                              Total |     14,689      100.00
```

tabulate k_adult, nolabel
После запятой – опции (в данном случае – не использовать метки значений)

```
       Есть |
      взрос |
        лый |
      вопро |
       сник |
      15-ой |
      волны? |      Freq.     Percent        Cum.
-------------+-----------------------------------
           0 |      2,199       14.97       14.97
           1 |     12,490       85.03      100.00
-------------+-----------------------------------
       Total |     14,689      100.00
```

Здесь в базе 14 689 кейсов, из них 85,03% взрослых анкет

4.2. Отберем кейсы из репрезентативной выборки (т.е. где переменная **k_origsm** равна 1)



Получим команду
tabulate k_adult if k_origsm ==1

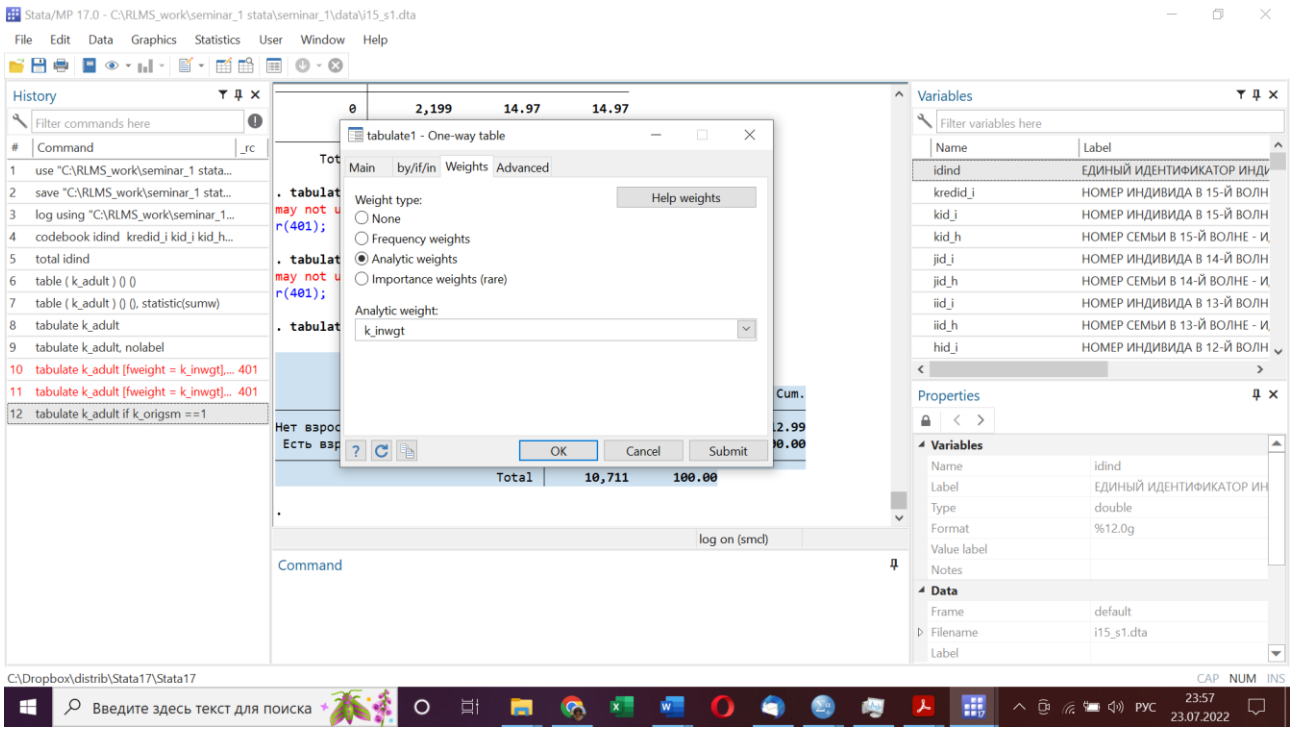
т.е. при условии, что переменная k_origsm равна 1, используется двойное равенство, так как это логическое выражение

Получим аутпут

Есть взрослый вопросник 15-ой волны?	Freq.	Percent	Cum.
Нет взрослого вопросника 15-й волны	1,391	12.99	12.99
Есть взрослый вопросник 15-й волны	9,320	87.01	100.00
Total	10,711	100.00	

Здесь в РЕПРЕЗЕНТАТИВНОЙ базе 10 711 кейсов, из них 87,01% взрослых анкет

4.3. Посмотрим теперь на взвешенные данные.
Переменная **X_INWGT (x_inwgt)** является постстратификационным весом индивида (обратите внимание, что для STATA регистр в имени переменной имеет значение, в отличие от SPSS, и в разных файлах у этой переменной может быть разный регистр). Веса выравнивают выборочную совокупность репрезентативной выборки данной волны, приводя показатели выборки к параметрам генеральной совокупности по полу, возрасту и типу поселения. Использование весов - на усмотрение исследователя.



Команда
tabulate k_adult [aweight = k_inwgt]

Есть взрослый вопросник 15-ой волны?	Freq.	Percent	Cum.
Нет взрослого вопросника 15-й волны	1,592.3132	14.87	14.87
Есть взрослый вопросник 15-й волны	9,117.6868	85.13	100.00
Total	10,710	100.00	

Здесь в также РЕПРЕЗЕНТАТИВНОЙ взвешенной базе 10 710 кейсов, из них 85,13% взрослых анкет. По идее, этот вариант в наилучшей степени соответствует генеральной совокупности. Обратите внимание, что для взвешенных данных частоты нецелые (взрослых – 9117,7 человек).

4.4. Давайте посмотрим на распределение переменной «семейное положение» в полной выборке без взвешивания.

tabulate k_marst

СЕМЕЙНОЕ ПОЛОЖЕНИЕ - 15 ВОЛНА	Freq.	Percent	Cum.
Никогда в браке не состояли	2,655	21.30	21.30
Состоите в зарегистрированном браке	6,028	48.36	69.66
Живете вместе, но не зарегистрированы	1,221	9.80	79.46
Разведены и в браке не состоите	1,016	8.15	87.61
Вдовец (вдова)	1,445	11.59	99.21
ОФИЦИАЛЬНО ЗАРЕГИСТРИРОВАНЫ, НО ВМЕСТЕ	99	0.79	100.00
Total	12,464	100.00	

Вы видите, что здесь 12 464 кейса, хотя в базе 14 689 кейсов. Это происходит потому, что у детей до 13 лет эта переменная принимает пропущенное значение (миссинг). (В общем-то, и у детей 14-15 лет по идее эта переменная не должна принимать какое-то значение, но ОК).
А также есть (или могут быть) пропущенные значения (отсутствие ответа) и для взрослых.

tabulate k_marst, missing

СЕМЕЙНОЕ ПОЛОЖЕНИЕ - 15 ВОЛНА	Freq.	Percent	Cum.
Никогда в браке не состояли	2,655	18.07	18.07
Состоите в зарегистрированном браке	6,028	41.04	59.11
Живете вместе, но не зарегистрированы	1,221	8.31	67.42
Разведены и в браке не состоите	1,016	6.92	74.34
Вдовец (вдова)	1,445	9.84	84.18
ОФИЦИАЛЬНО ЗАРЕГИСТРИРОВАНЫ, НО ВМЕСТЕ	99	0.67	84.85
.	2,225	15.15	100.00
Total	14,689	100.00	

БУДЬТЕ ВНИМАТЕЛЬНЫ!!! Смотрите, для какой совокупности вы смотрите распределение переменных!!!
Если вы хотите получить распределение для взрослых (от 14 лет) в репрезентативной выборке без взвешивания, команда будет такой:

tabulate k_marst if (k_origsm ==1 & k_adult == 1), missing

СЕМЕЙНОЕ ПОЛОЖЕНИЕ - 15 ВОЛНА	Freq.	Percent	Cum.
Никогда в браке не состояли	2,082	22.34	22.34
Состоите в зарегистрированном браке	4,439	47.63	69.97
Живете вместе, но не зарегистрированы	745	7.99	77.96
Разведены и в браке не состоите	780	8.37	86.33
Вдовец (вдова)	1,184	12.70	99.03
ОФИЦИАЛЬНО ЗАРЕГИСТРИРОВАНЫ, НО ВМЕСТЕ	73	0.78	99.82
.	17	0.18	100.00
Total	9,320	100.00	

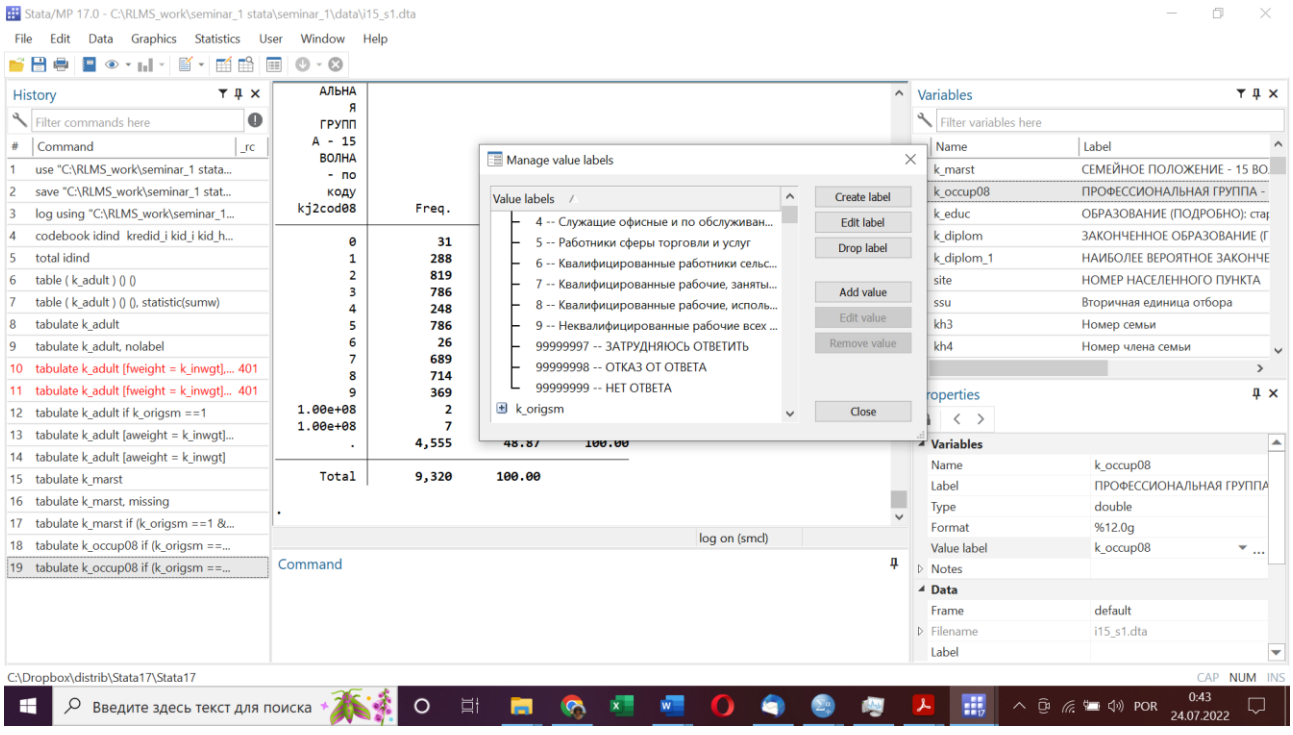
Вы видите, что у 17 взрослых респондентов (0,18%) в репрезентативной выборке у этой переменной значения пропущены.

4.5. Посмотрим теперь на распределение переменной k_occup08 (ПРОФЕССИОНАЛЬНАЯ ГРУППА – 15 ВОЛНА – по коду kj2cod08) для взрослых респондентов
tabulate k_occup08 if (k_origsm ==1 & k_adult == 1), missing

```
                    ПРОФЕССИОНАЛЬНАЯ |
                ГРУППА - 15 ВОЛНА - по |
                            коду kj2cod08 |      Freq.     Percent        Cum.
------------------------------------------+-----------------------------------
                       Военнослужащие |         31        0.33        0.33
Законодатели; крупные чиновники; руково |        288        3.09        3.42
Специалисты высшего уровня квалификации |        819        8.79       12.21
Специалисты среднего уровня квалификаци |        786        8.43       20.64
Служащие офисные и по обслуживанию клие |        248        2.66       23.30
      Работники сферы торговли и услуг |        786        8.43       31.74
Квалифицированные работники сельского,  |         26        0.28       32.02
Квалифицированные рабочие, занятые ручн |        689        7.39       39.41
Квалифицированные рабочие, использующие |        714        7.66       47.07
Неквалифицированные рабочие всех отрасл |        369        3.96       51.03
                  ЗАТРУДНЯЮСЬ ОТВЕТИТЬ |          2        0.02       51.05
                       ОТКАЗ ОТ ОТВЕТА |          7        0.08       51.13
                                     . |      4,555       48.87      100.00
------------------------------------------+-----------------------------------
                                 Total |      9,320      100.00
```

Посмотрите, в этой переменной есть значения «затрудняюсь ответить» (2 кейса), «отказ о ответа» (7 кейсов), которые по сути являются также отсутствием значений, а также 4 555 пропущенных значений. Кто эти люди??? – те, кто не имеет работы!!!
Это очень важно, понимать, для какой совокупности респондентов задан вопрос.
Не очень удобно также, что мы не видим значения, а только их лейблы. Даже команда ниже не сильно спасает, так как в аутпуте эти значения выдаются в виде 1.00e+08.

tabulate k_occup08 if (k_origsm ==1 & k_adult == 1), missing nolabel



Поэтому посмотрим на лейблы в разделе «Manage value labels»: мы видим, что значения «затрудняюсь ответить» - 99999997, «отказ от ответа» - 99999998, «нет ответа» - 99999999, то есть восьмизначные коды. Они удобны для SPSS, но очень не удобны для STATA.

Чтобы «объявить» эти значения миссингами, нужно их перекодировать в специальные значения. Для «пользовательских» миссингов зарезервированы специальные значения:
.a, .b, .c, …, .z
Что, вообще говоря, не очень-то удобно.

5. Редактирование переменных - основы

В меню DATA есть возможность редактировать метки и имена переменных, удалять и сохранять переменные, удалять и выбирать кейсы, а также генерировать новые переменные (например, на основе старых – создавать логарифмы и т.д.) и перекодировать переменные.

Некоторые наиболее часто используемые команды:

- Изменить значения переменной. Актуально для перекодировки значений категорийной переменной или для соединения нескольких категорий в одну.
recode

- Заменить значения уже существующей переменной.
replace имя переменной =выражение [if условие] [in диапазон]

- Переименовать переменную.
rename имя переменной новое имя

- Удалить наблюдения, удовлетворяющие указанным условиям.
drop if условие | in диапазон

- Удалить указанные переменные.
drop переменные

- Отсортировать данные по указанным переменным.
sort переменные

- Приписать метки к данным или переменным.
label

- Создать метку переменной, которая выводится командой describe и видна в окне переменных. Можно также задать метку для файла данных label data (информация о файле данных хранится в сопровождающем его объекте _dta). Эта метка будет выводиться при исполнении use и describe. Можно также задать метки для отдельных значений дискретной переменной через label define и label values. Признаком хорошего стиля работы с данными является придание меток создаваемым переменным: после любой команды generate или egen должно идти label variable .
label variable имя переменной "текст"

6. Редактирование переменных

6.1. Перекодирование миссингов для профессионального статуса
Перекодируем значения миссингов

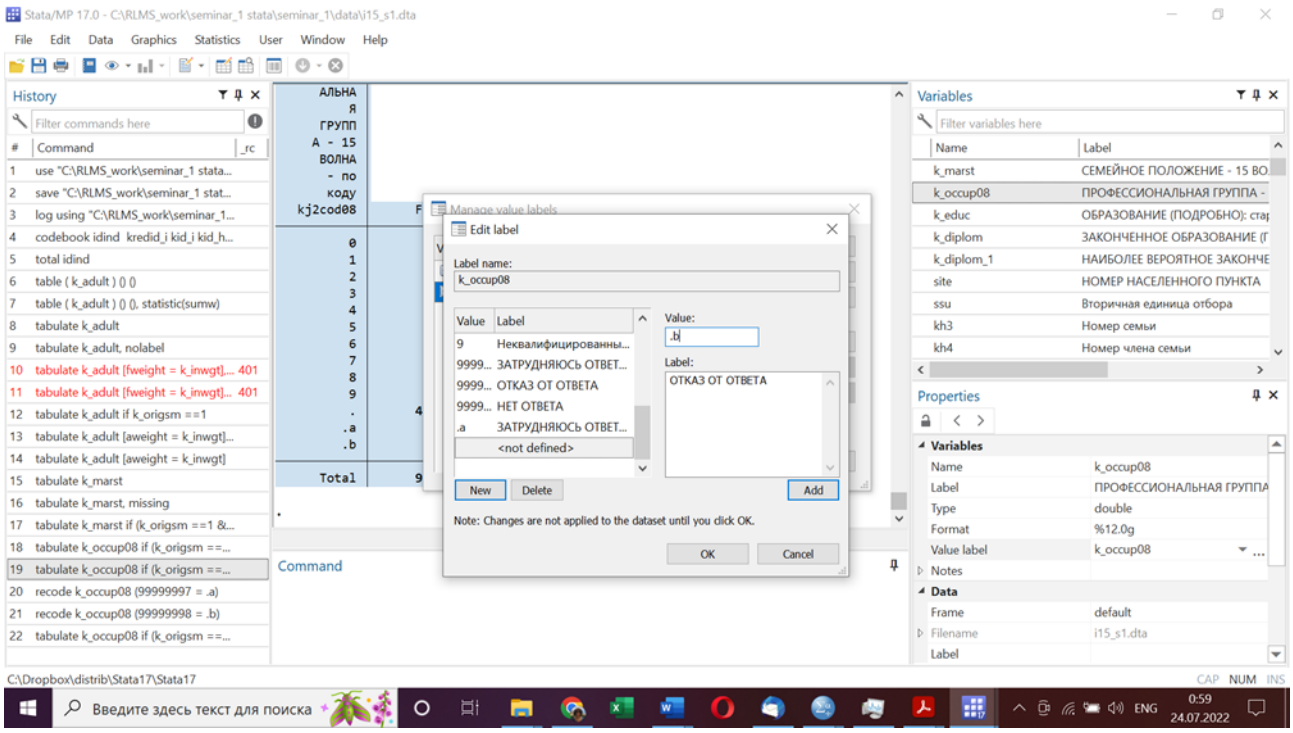
recode k_occup08 (99999997 = .a) (99999998 = .b) (99999999 = .c)

(третий код добавим, так как он может встречаться в панельной подвыборке)

Если мы теперь выполним снова команду
tabulate k_occup08 if (k_origsm ==1 & k_adult == 1), missing nolabel
то получим

```
      ПРОФЕ |
      ГРУПП |
    A - 15  |
      ВОЛНА |
       - по |
       коду |
  kj2cod08  |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |         31        0.33        0.33
          1 |        288        3.09        3.42
          2 |        819        8.79       12.21
          3 |        786        8.43       20.64
          4 |        248        2.66       23.30
          5 |        786        8.43       31.74
          6 |         26        0.28       32.02
          7 |        689        7.39       39.41
          8 |        714        7.66       47.07
          9 |        369        3.96       51.03
          . |      4,555       48.87       99.90
         .a |          2        0.02       99.92
         .b |          7        0.08      100.00
------------+-----------------------------------
      Total |      9,320      100.00
```

6.2. Отредактируем лейблы (можно в меню)



Получим команду

label define k_occup08 0 "Военнослужащие" 1 "Законодатели; крупные чиновники; руководит" 2 "Специалисты высшего уровня квалификации" 3 "Специалисты среднего уровня квалификации;" 4 "Служащие офисные и по обслуживанию клиент." 5 "Работники сферы торговли и услуг" 6 "Квалифицированные работники сельского, лес" 7 "Квалифицированные рабочие, занятые ручным" 8 "Квалифицир. рабочие, использующие машины" 9 "Неквалифицированные рабочие всех отраслей" 99999997 "ЗАТРУДНЯЮСЬ ОТВЕТИТЬ" 99999998 "ОТКАЗ ОТ ОТВЕТА" 99999999 "НЕТ ОТВЕТА" .a "ЗАТРУДНЯЮСЬ ОТВЕТИТЬ" .b "ОТКАЗ ОТ ОТВЕТА" .c "НЕТ ОТВЕТА", replace

6.3. Теперь если мы выполним следующую команду, то увидим ВСЕ значения (для репрезентативной выборки и взрослых респондентов)

tabulate k_occup08 if (k_origsm ==1 & k_adult == 1), missing

ПРОФЕССИОНАЛЬНАЯ ГРУППА - 15 ВОЛНА - по коду kj2cod08	Freq.	Percent	Cum.
Военнослужащие	31	0.33	0.33
Законодатели; крупные чиновники; руково	288	3.09	3.42
Специалисты высшего уровня квалификации	819	8.79	12.21
Специалисты среднего уровня квалификаци	786	8.43	20.64
Служащие офисные и по обслуживанию клие	248	2.66	23.30
Работники сферы торговли и услуг	786	8.43	31.74
Квалифицированные работники сельского,	26	0.28	32.02
Квалифицированные рабочие, занятые ручн	689	7.39	39.41
Квалифицированные рабочие, использующие	714	7.66	47.07
Неквалифицированные рабочие всех отрасл	369	3.96	51.03
.	4,555	48.87	99.90
ЗАТРУДНЯЮСЬ ОТВЕТИТЬ	2	0.02	99.92
ОТКАЗ ОТ ОТВЕТА	7	0.08	100.00
Total	9,320	100.00	

А если выполним эту команду (без опции: **, missing**), то получим распределение без пропущенных значений.

tabulate k_occup08 if (k_origsm ==1 & k_adult == 1)

ПРОФЕССИОНАЛЬНАЯ ГРУППА - 15 ВОЛНА - по коду kj2cod08	Freq.	Percent	Cum.
Военнослужащие	31	0.65	0.65
Законодатели; крупные чиновники; руково	288	6.06	6.71
Специалисты высшего уровня квалификации	819	17.22	23.93
Специалисты среднего уровня квалификаци	786	16.53	40.45
Служащие офисные и по обслуживанию клие	248	5.21	45.67
Работники сферы торговли и услуг	786	16.53	62.20
Квалифицированные работники сельского,	26	0.55	62.74
Квалифицированные рабочие, занятые ручн	689	14.49	77.23
Квалифицированные рабочие, использующие	714	15.01	92.24
Неквалифицированные рабочие всех отрасл	369	7.76	100.00
Total	4,756	100.00	

Однако это в некотором роде «опасно», если забыть, что эта переменная определена только для занятых.

6.4. Создание новой переменной

- Команда позволяет создать новую переменную, возможно, указанного типа, и присвоить ей значение выражения. В выражение могут входить числа, переменные, фигурировать арифметические операции, функции (математические, статистические, строковые и пр.), логические условия (которые вычисляются как 1 - истина и 0 - ложь) и пропущенные значения.
generate [тип] имя переменной = выражение [if условие] [in диапазон]

generate k_occup08_k = k_occup08

Присвоим лейблы

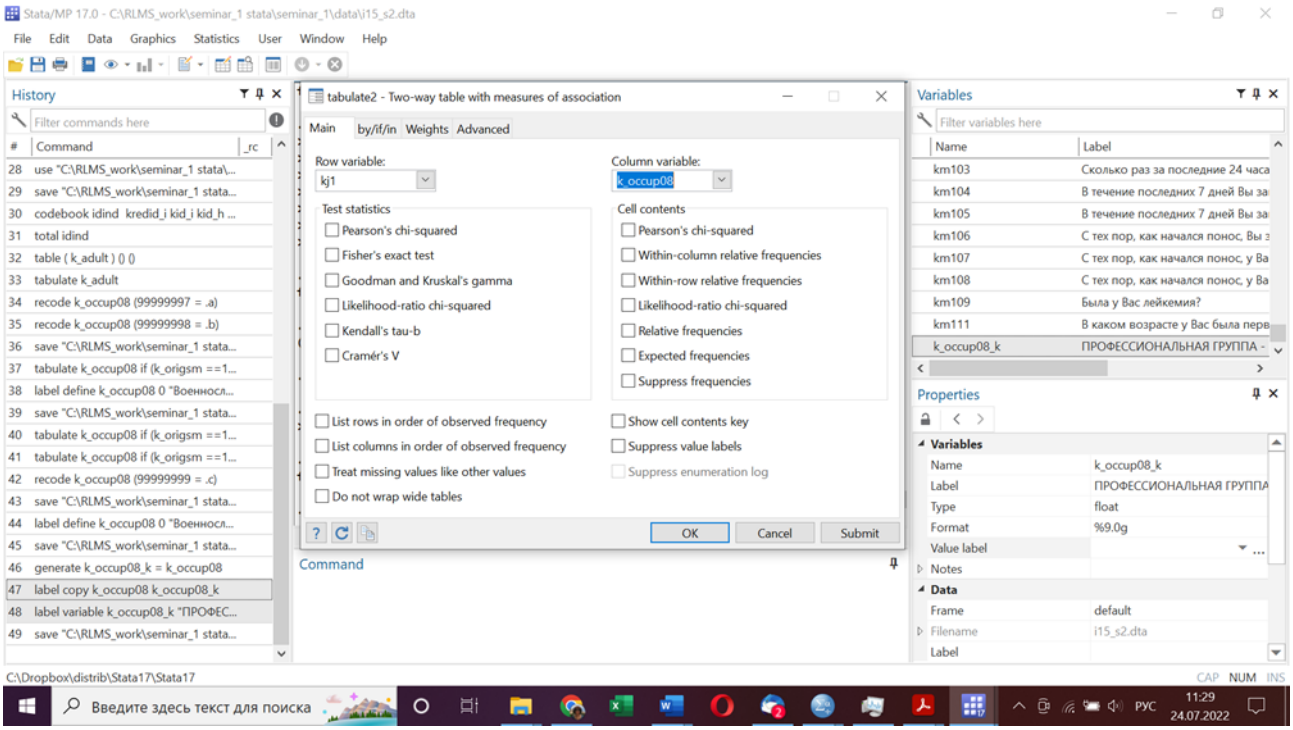
label variable k_occup08_k "ПРОФЕССИОНАЛЬНАЯ ГРУППА - 15 ВОЛНА - по коду kj2cod08 + незанятые"
label copy k_occup08 k_occup08_k

6.5. Рассмотрим переменную **kj1**

tabulate kj1 if k_adult == 1, missing

Ваше основное занятие в настоящее время?	Freq.	Percent	Cum.
Вы сейчас работаете	6,547	52.42	52.42
Вы находитесь в отпуске - декретном или	184	1.47	53.89
Вы находитесь в любом другом оплачиваем	37	0.30	54.19
Вы находитесь в неоплачиваемом отпуске	7	0.06	54.24
Или у Вас сейчас нет работы	5,706	45.68	99.93
ЗАТРУДНЯЮСЬ ОТВЕТИТЬ	5	0.04	99.97
ОТКАЗ ОТ ОТВЕТА	4	0.03	100.00
Total	12,490	100.00	

И рассчитаем кросс-таблицу



tabulate k_occup08 kj1 if k_adult == 1, missing

ПРОФЕССИОНАЛЬНАЯ ГРУППА - 15 ВОЛНА - по коду kj2cod08	Ваше основное занятие в настоящее время? Вы сейчас	Вы находи	Вы находи	Вы находи	Total
Военнослужащие	45	0	1	0	46
Законодатели; крупные	418	3	3	0	424
Специалисты высшего у	1,074	47	7	1	1,129
Специалисты среднего	1,058	43	9	1	1,111
Служащие офисные и по	347	13	0	0	360
Работники сферы торго	1,105	40	3	1	1,149
Квалифицированные раб	31	0	0	1	32
Квалифицированные раб	967	11	8	2	988
Квалифицированные раб	993	11	6	1	1,011
Неквалифицированные р	496	15	0	0	511
.	0	0	0	0	5,715
ЗАТРУДНЯЮСЬ ОТВЕТИТЬ	2	0	0	0	2
ОТКАЗ ОТ ОТВЕТА	10	1	0	0	11
НЕТ ОТВЕТА	1	0	0	0	1
Total	6,547	184	37	7	12,490

ПРОФЕССИОНАЛЬНАЯ ГРУППА - 15 ВОЛНА - по коду kj2cod08	Ваше основное занятие в настоящее время? Или у Вас	ЗАТРУДНЯЮ	ОТКАЗ ОТ	Total
Военнослужащие	0	0	0	46
Законодатели; крупные	0	0	0	424
Специалисты высшего у	0	0	0	1,129
Специалисты среднего	0	0	0	1,111
Служащие офисные и по	0	0	0	360
Работники сферы торго	0	0	0	1,149
Квалифицированные раб	0	0	0	32
Квалифицированные раб	0	0	0	988
Квалифицированные раб	0	0	0	1,011
Неквалифицированные р	0	0	0	511
.	5,706	5	4	5,715
ЗАТРУДНЯЮСЬ ОТВЕТИТЬ	0	0	0	2
ОТКАЗ ОТ ОТВЕТА	0	0	0	11
НЕТ ОТВЕТА	0	0	0	1
Total	5,706	5	4	12,490

Таким образом, если у человека «нет работы» (значение 5 вопроса **kj1**), в вопросе **k_occup08** будет пропущенное значение.

6.6. Перекодируем вновь созданную переменную

recode k_occup08_k (. = 10) if (kj1 ==5 & k_adult == 1)

label define k_occup08_k 10 "не работает", add

label values k_occup08_k k_occup08_k

tabulate k_occup08_k if (k_origsm ==1 & k_adult == 1), missing

ПРОФЕССИОНАЛЬНАЯ ГРУППА - 15 ВОЛНА - по коду kj2cod08 + незанятые	Freq.	Percent	Cum.
Военнослужащие	31	0.33	0.33
Законодатели; крупные чиновники; руково	288	3.09	3.42
Специалисты высшего уровня квалификации	819	8.79	12.21
Специалисты среднего уровня квалификаци	786	8.43	20.64
Служащие офисные и по обслуживанию клие	248	2.66	23.30
Работники сферы торговли и услуг	786	8.43	31.74
Квалифицированные работники сельского,	26	0.28	32.02
Квалифицированные рабочие, занятые ручн	689	7.39	39.41
Квалифицированные рабочие, использующие	714	7.66	47.07
Неквалифицированные рабочие всех отрасл	369	3.96	51.03
не работает	4,549	48.81	99.84
.	6	0.06	99.90
ЗАТРУДНЯЮСЬ ОТВЕТИТЬ	2	0.02	99.92
ОТКАЗ ОТ ОТВЕТА	7	0.08	100.00
Total	9,320	100.00	

6.7. Самостоятельное задание.

A) Для взрослых респондентов от 14 лет построить распределения переменных **kj72_171 kj72_172 kj72_173** (учитывая миссинги; для полной выборки)
B) Для всех трех переменных перекодировать 99999997 99999998 99999999 в пропущенные значения и изменить соответствующие лейблы
C) для переменной kj72_171 перекодировать значение «2» в «0» и изменить соотвествующий лейбл
D) построить новую переменную **kj72_172a**, в которой количество детей =0, если **kj72_171**=0 (то есть если у респондента нет детей)
E) построить новую переменную **kj72_173a,** в которой количество детей до 18 лет =0 если у респондента нет детей, или если у него нет детей до 18 лет.
F) построить распределения этих новых переменных, с учетом миссингов, для взрослых респондентов от 14 лет для репрезентативной выборки.
- Скопируйте результаты и команды из окна аутпута, и вставьте их в текстовый файл с вашими ответами на задания этого семинара. Один текстовый файл для всех выполненных заданий. Назовите файл вашей ФИО и номер группы, укажите номер семинара.