ЗАДАНИЯ К ПРАКТИЧЕСКИМ ЗАНЯТИЯМ.

Занятие 7.
Данные файлов по индивидуальным анкетам. Регрессионный анализ.

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 проделанную вами работу в виде кода, используя «сохранение» правильных команд в STATA (или функцию “paste” SPSS). В этом случае вы сможете дома повторить все сделанное вами в классе. Кроме того, рекомендуется прикладывать программу к вашим исследованиям.
4) В качестве отчета за семинар нужно предъявить созданные файлы данных, файл аутпута, и файл с кодом.

Исходные файлы.
ind_5_16_S6.dta – основной рабочий файл, который был получен на предыдущем занятии
children_5_29.dta – количество собственных детей индивида, проживающих с ним в одном домохозяйстве, по возрастам (сгенерированный файл на основе семейных данных), с 5 по 29 волну

30. Начало работы.
*Используем директорию на диске

C:\RLMS_work\seminar_7\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_7\data\семинар 7 Рощина.smcl"

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

***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_7\data\ind_5_16_S6.dta", clear

*И сохраните его под другим именем

save "C:\RLMS_work\seminar_7\data\ind_5_16_Sem7.dta"

31. Вспомогательные опции и ресурсы

31.1. Установка полезного приложения, записывающего результаты в удобном виде в текстовый файл - outreg2. (УСТАНАВЛИВАЕМ, кто не сделал – понадобится на занятии)
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
```

31.2. 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

31.3. Регрессионный анализ.
Полезные ресурсы – консультации по статистике в разных программных пакетах:
https://stats.oarc.ucla.edu/other/mult-pkg/seminars/#Stata
https://stats.oarc.ucla.edu/stata/modules/
https://stats.oarc.ucla.edu/stata/webbooks/reg/

Аннотированные аутпуты – детальное пояснение результатов разных регрессий:
https://stats.oarc.ucla.edu/other/annotatedoutput/

Примеры оценки разных регрессионных моделей в разных пакетах:
https://stats.oarc.ucla.edu/other/dae/

Примеры из учебников:
https://stats.oarc.ucla.edu/other/examples/

31.4. Полезные команды для предварительного анализа данных, или пост-оценочные команды (выполняются после оценки конкретной регрессии)

Detecting Unusual and Influential Data
predict -- used to create predicted values, residuals, and measures of influence.
rvpplot --- graphs a residual-versus-predictor plot.
rvfplot -- graphs residual-versus-fitted plot.
lvr2plot -- graphs a **leverage**-versus-squared-residual plot.
dfbeta -- calculates DFBETAs for all the independent variables in the linear model.
avplot -- graphs an added-variable plot, a.k.a. partial regression plot.
Tests for Normality of Residuals
kdensity -- produces kernel density plot with normal distribution overlayed.
pnorm -- graphs a standardized normal probability (P-P) plot.
qnorm --- plots the quantiles of varname against the quantiles of a normal distribution.
iqr -- resistant normality check and outlier identification.
swilk -- performs the Shapiro-Wilk W test for normality.
Tests for Heteroscedasticity
rvfplot -- graphs residual-versus-fitted plot.
hettest -- performs Cook and Weisberg test for heteroscedasticity.
whitetst -- computes the White general test for Heteroscedasticity.
Tests for Multicollinearity
vif -- calculates the variance inflation factor for the independent variables in the linear model.
collin -- calculates the variance inflation factor and other multicollinearity diagnostics
Tests for Non-Linearity
acprplot -- graphs an augmented component-plus-residual plot.
cprplot --- graphs component-plus-residual plot, a.k.a. residual plot.
Tests for Model Specification
linktest -- performs a link test for model specification.
ovtest -- performs regression specification error test (RESET) for omitted variables.

32. Оценим регрессию для модели Минцера.
Обычная множественная регрессия (МНК) для количественной переменной. Команда имеет вид:

regress зависимая переменная объясняющие переменные [if условие] [in диапазон], robust noconst cluster(групповая переменная)

Оценивание линейной регрессии зависимой переменной на объясняющие переменные. Выводятся основные результаты оценивания: количество наблюдений, таблица дисперсионного анализа, статистики F, R2, R2 adj, а также таблица оценок коэффициентов, стандартных отклонений оценок, t-статистик для гипотезы k = 0 и доверительных интервалов. Опция **robust** задает оценку ковариационной матрицы оценок коэффициентов в форме Уайта, учитывающей гетероскедастичность. Опция **cluster** указывает, что ковариационная матрица должна учитывать группировку наблюдений (как в кластерных выборочных обследованиях). Опция **noconst** указывает, что в модель, оценивамую Stata, не следует включать константу (как это делается по умолчанию). После команды **regress** можно получать прогнозные значения, остатки и строить диагностические переменные командой **predict** или проводить диагностику регрессии, не прогоняя регрессию заново.
Команды оценивания статистических моделей в Stata имеют много общего. В частности, после всех таких команд можно отдавать команду **predict**, которая будет строить значения тех или

иных выражений, связанных с результатами оценивания; получать матрицы самих оценок параметров (матрица-столбец e(b)) и их ковариационную матрицу (e(V)); строить тесты на линейные (test) и нелинейные (testnl, с использованием дельта-метода для получения ковариационной матрицы нелинейных функций оценок) комбинации параметров, и т.д.
Stata Annotated Output Regression Analysis OLS – в помощь для интерпретации результатов.
https://stats.oarc.ucla.edu/stata/output/regression-analysis/

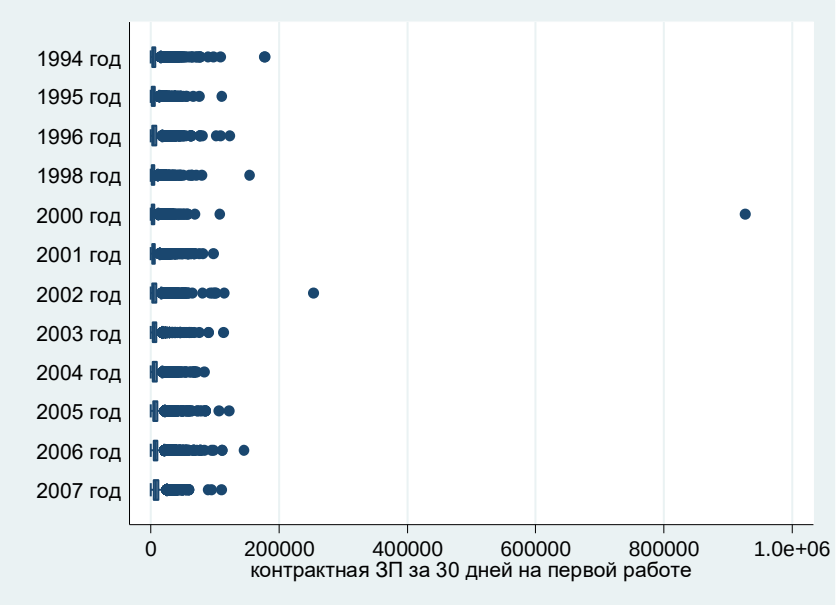
*32.1. Предварительный анализ.
*Так как мы будем оценивать регрессию для занятых, которые получили заработок за последние 30 дней, в будущих командах будем учитывать условие: **if lg_Hwage1 > 0**

*Так как наша зависимая переменная – логарифм ставки заработной платы (дефлированной), посмотрим на ее описательные характеристики и распределение по годам на графике:

sum Wage_1 Hwage1 lg_Hwage1

```
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
      Wage_1 |     56,757    7229.252    8529.321        7.46   926556.7
      Hwage1 |     51,028    50.79457    191.5247   .0264539      24618
   lg_Hwage1 |     51,028    3.440276    .9345589  -3.632352   10.11123
```

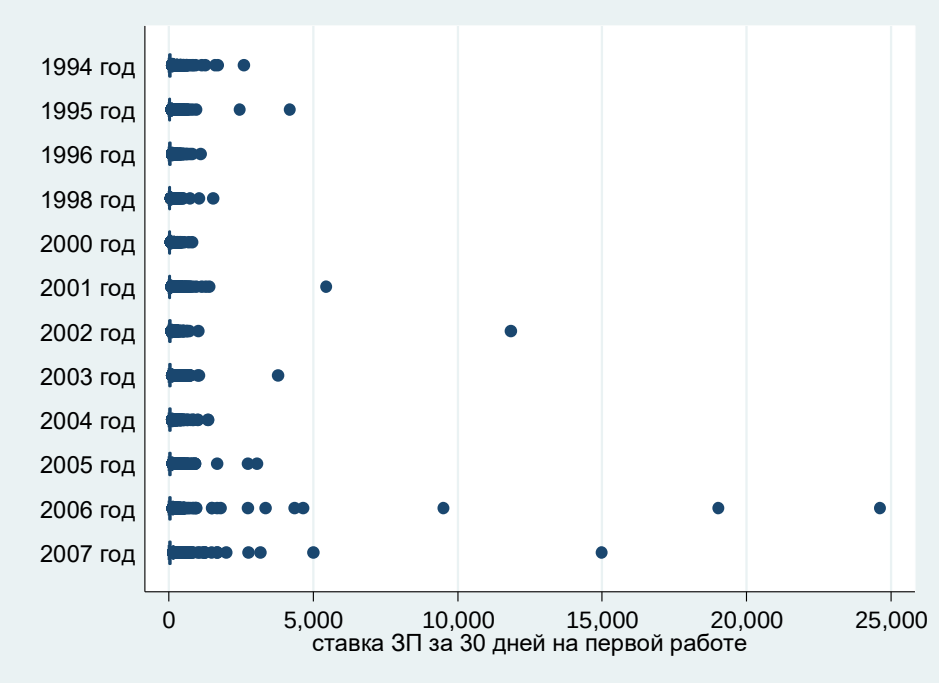
graph hbox Wage_1 , over(id_w)



*Явно выделяются случаи ЗП больше 200 тыс., удалим их.

recode Hwage1 (0 / 30000 = .) if Wage_1 > 200000
recode lg_Hwage1 (-5 / 30000 = .) if Wage_1 > 200000
recode Wage_1 (200000 / 1000000 = .)

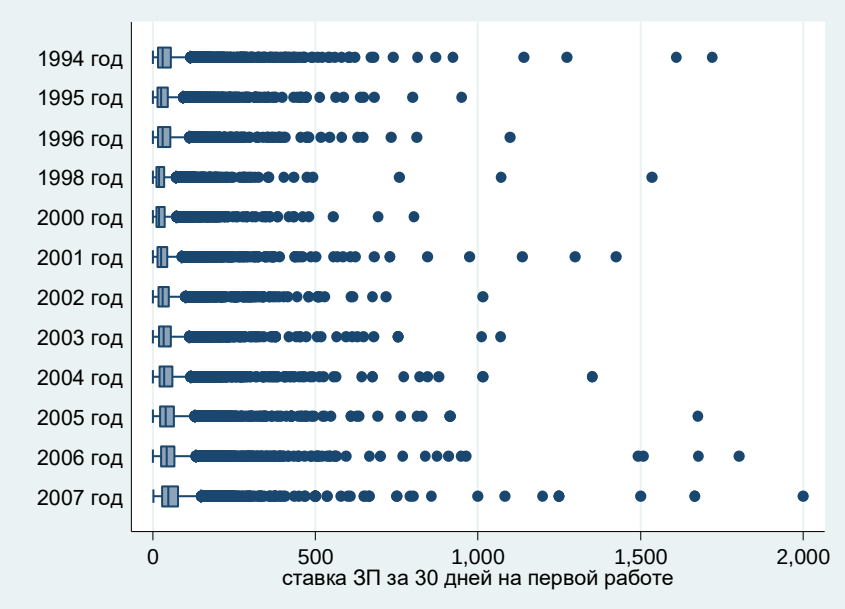
graph hbox Hwage1 , over(id_w)



*Очевидно, что есть сильно выделяющиеся кейсы, где ставка ЗП больше 2 тыс.руб. в час.
*Удалим эти значения

recode lg_Hwage1 (-5 / 30000 = .) if Hwage1 >= 2001
recode Hwage1 (2001 / 25000 = .)

graph hbox Hwage1 , over(id_w)



*Теперь распределение более равномерное. Однако есть еще слишком маленькие значения, где ставка ЗП меньше или равна 1 рублю в час (таких 74 кейса), удалим их тоже:

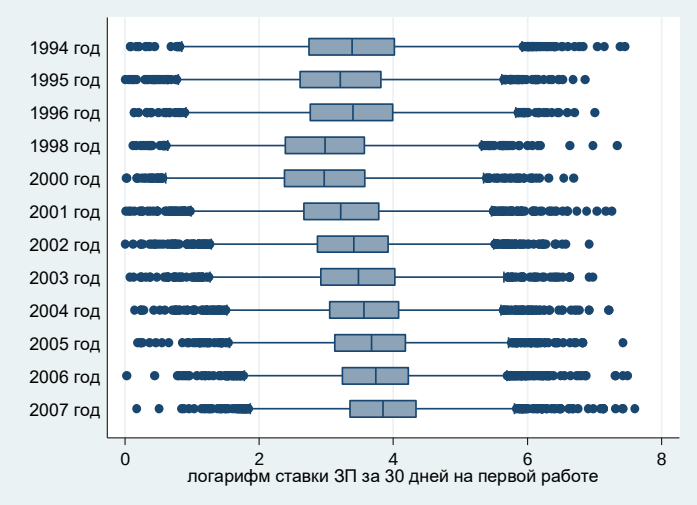
recode lg_Hwage1 (-5 / 30000 = .) if Hwage1 <= 1
recode Hwage1 (0 / 1 = .)

*Посмотрим теперь новые описательные характеристики

sum Wage_1 Hwage1 lg_Hwage1

```
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
      Wage_1 |     56,755    7208.709    7535.796       7.46   178208.9
      Hwage1 |     50,933    48.17861    66.49048   1.005217       2000
   lg_Hwage1 |     50,933    3.443901    .9173143   .0052033   7.600903
```

graph hbox lg_Hwage1 , over(id_w)



Теперь переходим к подготовке к оценке регрессии.
Начинаем ВСЕГДА с описательных статистик всех переменных регрессии.
Для создания дамми-переменных в регрессии, а также в любых других командах, удобно использовать опцию для так называемых «факторных переменных»:
ibN. varname (или **i. varname** если не нужно указать базовую категорию). Она создает на одну меньше дамми, чем есть категорий в переменной; по умолчанию базовой берется первая категория. Если нужно указать другую, можно написать: **ib3.varname**, тогда в этой переменной базовой будет третья категория.
Используем также синтаксис для создания интеракций (взаимодействий), в нашем случае – для переменной возраст\10.

Факторные переменные являются расширениями списков переменных существующих переменных. Когда команда позволяет использовать *факторные переменные*, помимо ввода имен переменных из ваших данных, вы можете вводить факторные переменные, которые могут выглядеть так:

i.varname
i.varname#i.varname
i.varname#i.varname#i.varname
i.varname##i.varname
i.varname##i.varname##i.varname

Факторные переменные создают индикаторные переменные из категориальных переменных, взаимодействий индикаторов категориальных переменных, взаимодействий категориальных и непрерывных переменных и взаимодействий непрерывных переменных (полиномов). Они

разрешены для большинства команд оценки и постоценки, а также для некоторых других команд.

Есть пять операторов:

i. оператор одной переменной для указания индикаторов (unary operator to specify indicators)

c. оператор одной переменной, который следует рассматривать как непрерывный (unary operator to treat as continuous)

o. оператор для пропуска переменной или индикатора (unary operator to omit a variable or indicator)

оператор для двух переменных для нтеракций (binary operator to specify interactions)

оператор интеракций для двух переменных, включающий также сами переменные (binary operator to specify factorial interactions)

Индикаторы и взаимодействия, создаваемые операторами фактор-переменными, называются виртуальными переменными. Они действуют как переменные в списках переменных, но не существуют в наборе данных. Категориальные переменные, к которым применяются операторы переменных-факторов, должны содержать неотрицательные целые числа со значениями в диапазоне от 0 до 32 740 включительно. Факторные переменные можно комбинировать с операторами временных рядов L. и F.

Посмотрим на описательные характеристики всех переменных в будущей регрессии. В первую очередь смотрим на количество кейсов по каждой переменной, а также на минимум и максимум (нет ли ошибок). Если по какой-то переменной намного меньше кейсов, это сигнал проверить, все ли в порядке – в регрессии будет не больше кейсов, чем минимум из количества наблюдений (в реальности меньше из-за пропущенных значений). Главное, чтобы не была упущена какая-то категория респондентов, которая должна быть включена в анализ.
Обратите внимание, что первая категория не включена в описание.

sum lg_Hwage1 i.diplom_k c.age_10##c.age_10 male ln_regwage i.status_1 i.fed_okr i.year

```
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
   lg_Hwage1 |     50,933    3.443901    .9173143   .0052033   7.600903
             |
    diplom_k |
 зак.среднее |    120,311    .3388801     .473331          0          1
среднее проф |    120,311     .227901    .4194802          0          1
     высшее  |    120,311     .168347    .3741757          0          1
-------------+---------------------------------------------------------
             |
      age_10 |    120,424     4.27669      1.8682        1.3       10.2
             |
    c.age_10#|
    c.age_10 |    120,424    21.78022    17.55703       1.69     104.04
             |
        male |    120,436    .4302783    .4951171          0          1
  ln_regwage |    120,436    8.783461    .5138943   7.693017   10.11715
-------------+---------------------------------------------------------
             |
    status_1 |
  обл.центр  |    120,436    .2979508    .4573596          0          1
 другой город|    120,436    .2646219    .4411335          0          1
  село, пгт  |    120,436    .3237653    .4679136          0          1
             |
     fed_okr |
   Северный  |    120,436    .0647896    .2461553          0          1
-------------+---------------------------------------------------------
Центральный  |    120,436    .1792819    .3835898          0          1
Приволжский  |    120,436     .198454    .3988374          0          1
 Юг и С.Кавк.|    120,436    .1644608    .3706947          0          1
 Уральский   |    120,436    .1019463    .3025789          0          1
 'Сибирский  |    120,436    .1275698    .3336116          0          1
-------------+---------------------------------------------------------
```

```
Дальневос~й  |    120,436    .0498356    .2176061          0          1
             |
        year |
        1995 |    120,436    .0698877    .2549588          0          1
        1996 |    120,436    .0692567    .2538912          0          1
        1998 |    120,436    .0722292    .2588682          0          1
-------------+---------------------------------------------------------
        2000 |    120,436    .0753429     .263945          0          1
        2001 |    120,436    .0838454    .2771569          0          1
        2002 |    120,436    .0871749    .2820924          0          1
        2003 |    120,436    .0883125      .28375          0          1
        2004 |    120,436    .0885283    .2840629          0          1
-------------+---------------------------------------------------------
        2005 |    120,436    .0858298    .2801137          0          1
        2006 |    120,436    .1037065    .3048807          0          1
        2007 |    120,436    .1020459    .3027099          0          1
```

Количество кейсов в нашей регрессии будет ограничено информацией о ставке заработной платы, то есть не больше 50933 кейсов. Но в остальных переменных количество кейсов намного больше, поэтому можем оценить модель.

*32.2. Оценим самую простую регрессию по модели Минцера: зависимая переменная – логарифм ЗП, независимые – уровень образования, возраст и возраст в квадрате, мужской пол, контрольные переменные – лог региональной ЗП, год, тип поселения, фед.округ.
*Базовые категории (опущенные) – для образования – школа; для федеральных округов – Москва и СП; для лет – 1994, для типа поселения - города.

reg lg_Hwage1 i.diplom_k c.age_10##c.age_10 male ln_regwage i.status_1 i.fed_okr i.year

```
note: 8.fed_okr omitted because of collinearity

      Source |       SS           df       MS      Number of obs   =    50,917
-------------+----------------------------------   F(27, 50889)    =    885.03
       Model |  13692.1171        27  507.115448   Prob > F        =    0.0000
    Residual |  29158.9764    50,889  .572991736   R-squared       =    0.3195
-------------+----------------------------------   Adj R-squared   =    0.3192
       Total |  42851.0935    50,916  .841603691   Root MSE        =    .75696

------------------------------------------------------------------------------
       lg_Hwage1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----------------+------------------------------------------------------------
         diplom_k |
   законч.среднее|      0.143      0.012   11.925   0.000        0.119       0.166
      средн.проф.|      0.299      0.013   23.779   0.000        0.274       0.323
         высшее  |      0.565      0.013   43.934   0.000        0.540       0.590
                 |
          age_10 |      0.445      0.017   25.857   0.000        0.411       0.479
                 |
        c.age_10#|
        c.age_10 |     -0.058      0.002  -28.007   0.000       -0.063      -0.054
                 |
            male |      0.337      0.007   48.967   0.000        0.324       0.351
      ln_regwage |      0.748      0.014   54.024   0.000        0.720       0.775
                 |
        status_1 |
      обл.центр  |     -0.164      0.019   -8.460   0.000       -0.202      -0.126
   другой город  |     -0.160      0.020   -7.862   0.000       -0.200      -0.120
      село, пгт  |     -0.514      0.019  -26.372   0.000       -0.552      -0.476
                 |
         fed_okr |
       Северный  |      0.230      0.020   11.234   0.000        0.190       0.270
    Центральный  |      0.122      0.019    6.542   0.000        0.085       0.158
    Приволжский  |      0.034      0.019    1.761   0.078       -0.004       0.071
  Юг и С.Кавказ  |      0.062      0.020    3.162   0.002        0.024       0.100
     Уральский   |      0.107      0.019    5.496   0.000        0.069       0.145
     'Сибирский  |     -0.031      0.019   -1.642   0.101       -0.068       0.006
Дальневосточный  |      0.000  (omitted)
                 |
            year |
            1995 |     -0.117      0.017   -6.704   0.000       -0.151      -0.083
            1996 |     -0.184      0.018  -10.165   0.000       -0.219      -0.148
            1998 |     -0.270      0.018  -15.028   0.000       -0.305      -0.235
```

```
        2000  |     -0.429      0.018   -24.492   0.000       -0.463      -0.395
        2001  |     -0.369      0.017   -21.338   0.000       -0.403      -0.335
        2002  |     -0.321      0.017   -18.316   0.000       -0.355      -0.286
        2003  |     -0.318      0.018   -17.686   0.000       -0.353      -0.283
        2004  |     -0.295      0.018   -16.109   0.000       -0.331      -0.259
        2005  |     -0.285      0.019   -14.740   0.000       -0.323      -0.247
        2006  |     -0.287      0.020   -14.487   0.000       -0.325      -0.248
        2007  |     -0.280      0.021   -13.355   0.000       -0.321      -0.239
              |
        _cons |     -3.914      0.129   -30.435   0.000       -4.166      -3.662
------------------------------------------------------------------------------
```

Количество кейсов в модели - 50917. Первое, F-test статистически значим, значит, в модели есть ненулевые коэффициенты. R-squared = .3195, значит, почти 32% вариации зависимой переменной объясняется нашей моделью (это довольно неплохой результат!). Для каждой переменной t-test и его значимость показывают, значимо ли переменная отлична от нуля. Это можно также увидеть по доверительным интервалам коэффициента. Незначимы только две переменные (с уровнем 5%). Содержательно нас интересуют коэффициенты для уровней образования и для возраста и возраста в квадрате. Они значимы и положительны, для возраста в квадрате отрицательны, то есть ветви параболы обращены вниз, как и предсказано теоретической моделью. Ставка заработной платы выше у мужчин, и положительно зависит от уровня заработной платы (точнее, логарифма ее) в регионе. Константа показывает, чему равна ставка заработной платы, если все переменные были бы равны 0 (т.е. при базовых категориях для дамми). Коэффициент при каждой переменной показывает, насколько изменится зависимая переменная при изменении независимой переменной на единицу, при том, что все остальные переменные остаются неизменными (и равными своим средним значениям). Это очень важный момент, так как именно это свойство регрессии приводит к необходимости включать в модель все переменные, которые, как мы думаем, могут повлиять на зависимую переменную. Строго говоря, невключение каких-то важных переменных приводит к увеличению ошибки в оценке модели. Серьезная ошибка – сначала оценивать регрессию на одной группе переменных (например, социально-демографических), а потом – на другой группе (например, ценности, мотивы и т.д.). Так как коэффициент показывает изменение зависимой переменной при увеличении независимой переменной на единицу, коэффициент зависит от единицы измерения (поэтому мы возраст разделили на 10), и напрямую сравнивать коэффициенты нельзя (можно только стандартизованные коэффициенты).

Так как коэффициент показывает измерение зависимой переменной при изменении данной независимой переменной на единицу, и средних значениях прочих переменных, принципиально важно в любой статье приводить не только коэффициенты, но и средние значения всех переменных. Чтобы они были посчитаны именно для тех кейсов, которые вошли в регрессию (в нашем случае – это логарифм ставки заработной планы не миссинг), нужно использовать следующую команду (сравните результаты с предыдущей командой sum). Обратите внимание, что общее количество кейсов – такое же, как в регресиии выше. Данная команда относится к последней оцененной регрессии.

estat summarize

```
Estimation sample regress                 Number of obs =     50,917
  ------------------------------------------------------------------
      Variable |         Mean      Std. Dev.         Min          Max
  -------------+----------------------------------------------------
     lg_Hwage1 |      3.44392      .9173896     .0052033     7.600903
               |
      diplom_k |
      зако..   |     .3758273      .4843405            0            1
  среднее п..  |     .2799851       .448996            0            1
       высшее  |     .2385647      .4262102            0            1
               |
        age_10 |     3.924416      1.196277          1.4          8.1
               |
```

```
     c.age_10#|
     c.age_10 |     16.83209      9.886377         1.96     65.61001
              |
         male |     .4734372      .4992988            0            1
   ln_regwage |     8.830315      .5090993     7.693017     10.11715
              |
     status_1 |
   обл.центр  |     .3346427        .47187            0            1
     друг..   |     .2801815      .4490922            0            1
   село, пгт  |     .2588526      .4380088            0            1
              |
      fed_okr |
    Северный  |     .0703891      .2558041            0            1
Центральный   |     .1879922      .3907098            0            1
Приволжский   |     .2160968      .4115851            0            1
   Юг и ..    |     .1353968      .3421503            0            1
 Уральский    |     .0950567      .2932961            0            1
 'Сибирский   |      .123829      .3293896            0            1
Дальневос~й   |   (omitted)
         year |
         1995 |     .0698784      .2549445            0            1
         1996 |     .0658523      .2480261            0            1
         1998 |     .0653809      .2471992            0            1
         2000 |     .0696428      .2545465            0            1
         2001 |     .0788538      .2695131            0            1
         2002 |     .0858456      .2801387            0            1
         2003 |     .0869061      .2817002            0            1
         2004 |     .0916001      .2884634            0            1
         2005 |     .0890862      .2848709            0            1
         2006 |     .1089813      .3116188            0            1
         2007 |     .1093348      .3120619            0            1
-------------------------------------------------------------------------
```

32.3. Проверим модель на мультиколлинеарность.
Для этого после регрессии используем команду **vif** (= *variance inflation factor)*. Как правило, если значение VIF больше 10, присуствует мультиколлинеарность, и надо удалить одну из переменных (обычно ту, для которой это значение самое большое). Tolerance = 1/VIF, это степень коллинеарности, 0,1 соответствует VIF = 10, и означает, что переменная является линейной комбинацией других переменных (некоторые авторы говорят о границе 5 или даже 4, но в руководстве STATA указано 10). Случай мультиколлинеарности не относится к нелинейным моделям, то есть если помимо переменной, включен также ее квадрат (как у нас переменная возраста). В нашей модели одна из дамми, федеральный округ = 8, исключена автоматически из модели в силу коллинеарности (так как в двух переменных – тип поселения и федеральный округ – есть одна и та же градация, Москва и Санкт-Петербург).

vif

```
    Variable |       VIF       1/VIF
-------------+----------------------
    diplom_k |
           2 |      2.99    0.334923
           3 |      2.82    0.354023
           4 |      2.67    0.374485
     age_10  |     37.67    0.026544
    c.age_10#|
    c.age_10 |     37.86    0.026413
        male |      1.05    0.951524
  ln_regwage |      4.41    0.226752
    status_1 |
           1 |      7.41    0.134978
           2 |      7.42    0.134682
           3 |      6.48    0.154421
     fed_okr |
           2 |      2.44    0.409821
           3 |      4.71    0.212510
           4 |      5.54    0.180511
```

```
          5  |      3.98    0.251220
          6  |      2.89    0.345788
          7  |      3.46    0.289331
        year |
        1995 |      1.76    0.567172
        1996 |      1.78    0.560971
        1998 |      1.75    0.570893
        2000 |      1.77    0.566483
        2001 |      1.93    0.518369
        2002 |      2.14    0.468269
        2003 |      2.28    0.438483
        2004 |      2.49    0.402203
        2005 |      2.69    0.371260
        2006 |      3.38    0.295934
        2007 |      3.80    0.263298
-------------+----------------------
    Mean VIF |      5.91
```

Для нашего случая VIF > 10 при возрасте и возрасте в квадрате – это ОК. Но все же в модели довольно высокие значения для переменной «место жительства». Так как для федеральных округов у нас базвая категория – Москва и Санкт-Петербург, возможно, лучше было бы ввести переменную «сельская местность» (сделаем позже).

32.4. Для вывода стандартизованных коэффициентов beta необходима дополнительная опция. Их используют для сравнения силы влияния разных переменных, нивелируя единицы измерения, так как они измеряются в «стандартных ошибках», а не в единицах измерения переменных. Коэффициенты beta показывают те коэффициенты, которые мы бы получили, если бы все независимые переменные преобразовали к стандартному виду, то есть стандартизовали (standard scores, also called z-scores, перевод измерений в стандартную Z-шкалу со средним = 0 и стандартным отклонением =1).
*Рассчитаем ту же модель со стандартизованными коэффициентами

reg lg_Hwage1 i.diplom_k c.age_10##c.age_10 male ln_regwage i.status_1 i.fed_okr i.year, beta

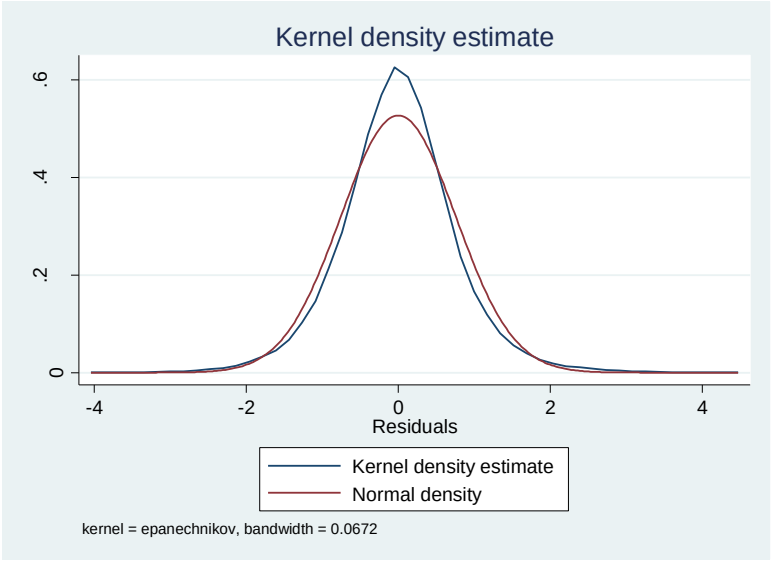
```
note: 8.fed_okr omitted because of collinearity

      Source |       SS           df       MS      Number of obs   =    50,917
-------------+----------------------------------   F(27, 50889)    =    885.03
       Model |  13692.1171        27  507.115448   Prob > F        =    0.0000
    Residual |  29158.9764    50,889  .572991736   R-squared       =    0.3195
-------------+----------------------------------   Adj R-squared   =    0.3192
       Total |  42851.0935    50,916  .841603691   Root MSE        =    .75696

------------------------------------------------------------------------------
       lg_Hwage1 |      Coef.   Std. Err.      t    P>|t|                 Beta
-----------------+------------------------------------------------------------
        diplom_k |
      законч..   |      0.143      0.012   11.925   0.000                0.075
      средне..   |      0.299      0.013   23.779   0.000                0.146
         высшее  |      0.565      0.013   43.934   0.000                0.263
                 |
          age_10 |      0.445      0.017   25.857   0.000                0.580
                 |
        c.age_10#|
        c.age_10 |     -0.058      0.002  -28.007   0.000               -0.630
                 |
            male |      0.337      0.007   48.967   0.000                0.184
      ln_regwage |      0.748      0.014   54.024   0.000                0.415
                 |
        status_1 |
      обл.центр  |     -0.164      0.019   -8.460   0.000               -0.084
   другой город  |     -0.160      0.020   -7.862   0.000               -0.078
      село, пгт  |     -0.514      0.019  -26.372   0.000               -0.245
                 |
         fed_okr |
       Северный  |      0.230      0.020   11.234   0.000                0.064
    Центральный  |      0.122      0.019    6.542   0.000                0.052
    Приволжский  |      0.034      0.019    1.761   0.078                0.015
  Юг и С.Кавказ  |      0.062      0.020    3.162   0.002                0.023
     Уральский   |      0.107      0.019    5.496   0.000                0.034
```

'Сибирский	-0.031	0.019	-1.642	0.101	-0.011
Дальневосточный	0.000	(omitted)			0.000
year					
1995	-0.117	0.017	-6.704	0.000	-0.033
1996	-0.184	0.018	-10.165	0.000	-0.050
1998	-0.270	0.018	-15.028	0.000	-0.073
2000	-0.429	0.018	-24.492	0.000	-0.119
2001	-0.369	0.017	-21.338	0.000	-0.108
2002	-0.321	0.017	-18.316	0.000	-0.098
2003	-0.318	0.018	-17.686	0.000	-0.098
2004	-0.295	0.018	-16.109	0.000	-0.093
2005	-0.285	0.019	-14.740	0.000	-0.088
2006	-0.287	0.020	-14.487	0.000	-0.097
2007	-0.280	0.021	-13.355	0.000	-0.095
_cons	-3.914	0.129	-30.435	0.000	.

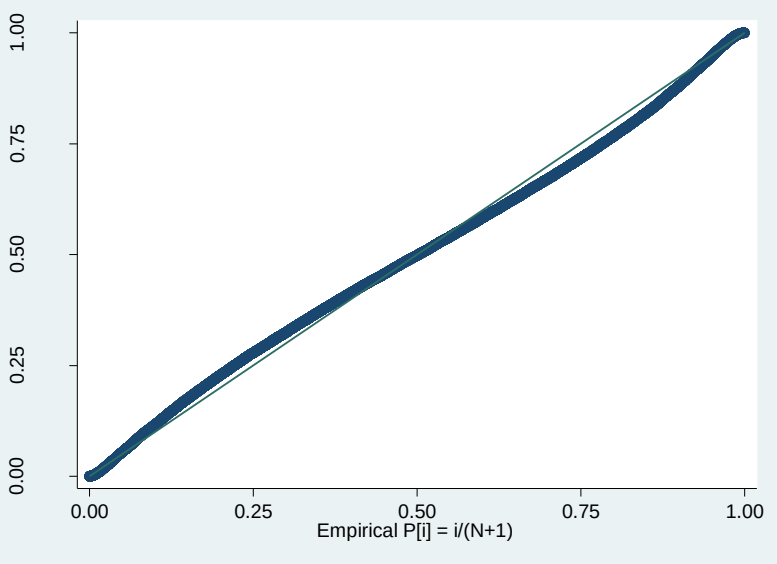
В нашем примере возраст в квадрате имеет максимальный отрицательный Beta coefficient, -0.63 (in absolute value), а возраст – максимальный положительный = **0.58.** **Сибирский** регион имеет минимальный отрицательный Beta coefficient, -.01, но он незначим!. Увеличение возраста в квадрате на «одно стандартное отклонение» уменьшает логарифм ставки заработной платы на -0.62, а возраста – увеличивает на 0,58, при том, что остальные переменные неизменны. То есть для данного периода, самое сильное влияние на ставку заработной платы оказывал возраст (как аппроксимация стажа).
Обратите внимание, что для дихотомических переменных обычный коэффициент показывает изменение зависимой переменной при изменении независимой на единицу (а для набора дамми – по сравнению с базовой категорией), а Beta coefficient – также при изменении этих переменных на стандартное отклонение. В данном случае, по сравнению с отсуствием полного среднего образования, каждый уровень образования дает бОльшую отдачу; но отдача от стажа выше, чем от высшего образования.

*32.5. Тест на нормальность остатков.
*Предскажем переменную, содержащую остатки. Затем используем команду **kdensity** для построения графика (kernel density plot) для сравнения распределения остатков с нормальным. Затем команда **pnorm** command показывает график стандартизованной нормальной вероятности (P-P), а график **qnorm** показывает квинтили переменной по сравнению с квинтилями нормально распределенной переменной. **pnorm** чувствителен к нарушению нормальности в середине распределения, а **qnorm** - на его «хвостах».

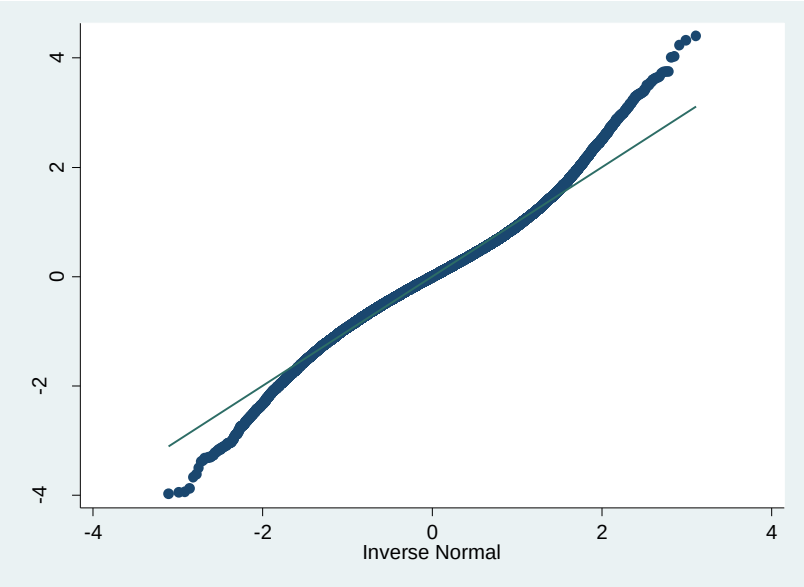
predict r, resid
kdensity r, normal



pnorm r



qnorm r



Как вы видите, результаты **pnorm** не показывают никаких признаков ненормальности, в то время как команда qnorm показывает небольшое отклонение от нормы в верхней части хвоста, как это видно из **kdensity** выше. Тем не менее, это кажется незначительным и тривиальным отклонением от нормы. Можно принять, что остатки близки к нормальному распределению.

*Другим доступным тестом является тест **swilk**, который выполняет W-тест Шапиро-Уилка на нормальность. Значение p основано на предположении, что распределение является нормальным. В нашем примере оно очень мало (0,00 < 0,05), что указывает на то, что мы не можем подтвердить нормальное распределение остатков r.

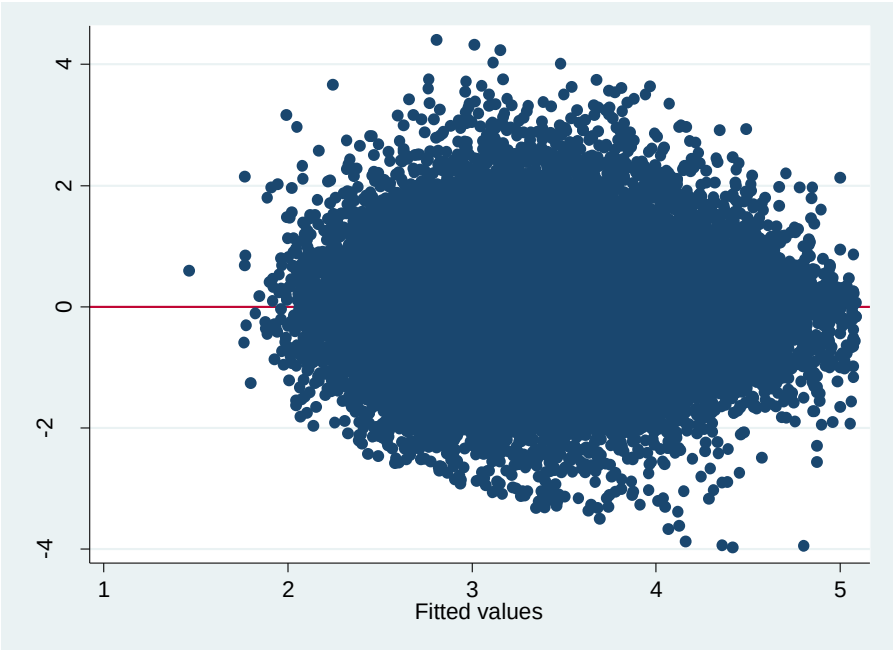
swilk r

```
                    Shapiro-Wilk W test for normal data

    Variable |        Obs       W           V         z       Prob>z
-------------+--------------------------------------------------------
           r |     50,917    0.98561     269.618    15.528     0.00000
```

Тем не менее, это не означает, что наша модель плоха. Многие исследователи считают, что множественная регрессия требует нормальности. Однако нормальность остатков требуется только для проверки достоверности гипотезы, то есть предположение о нормальности гарантирует, что p-значения для t-тестов и F-тестов будут действительными. Нормальность не требуется для получения несмещенных оценок коэффициентов регрессии. Регрессия МНК просто требует, чтобы остатки (ошибки) были одинаково и независимо распределены. Кроме того, нет никаких предположений или требований, чтобы переменные-предикторы были нормально распределены. Если бы это было так, то мы не смогли бы использовать дамми переменные в наших моделях.

*32.6. Тест на гетероскедастичность.
Одним из основных предположений для обычной регрессии наименьших квадратов является однородность дисперсии остатков. Если модель хорошо подобрана, не должно быть никакой закономерности в остатках, нанесенных на график по сравнению с подобранными значениями. Если дисперсия остатков непостоянна, то говорят, что дисперсия остатков является «гетероскедастичной». Существуют графические и неграфические методы обнаружения гетероскедастичности. Обычно используемый графический метод заключается в построении графика остатков по сравнению с подобранными (прогнозируемыми) значениями. Мы делаем это, введя команду **rvfplot**. Ниже мы используем команду **rvfplot** с параметром **yline(0)**, чтобы поместить референсную линию на y=0. Мы видим, что структура точек данных становится немного уже к правому концу, что является признаком гетероскедастичности.

rvfplot, yline(0)



*Первый тест на гетероскедастичность - **imest** - это тест Уайта, а второй тест - **hettest** - это тест Бреуша-Пагана. Оба проверяют нулевую гипотезу о том, что дисперсия остатков однородна. Следовательно, если значение p очень мало, нам придется отклонить гипотезу и принять альтернативную гипотезу о неоднородности дисперсии. Таким образом, в этом случае свидетельство не противоречит нулевой гипотезе о том, что дисперсия является однородной.

*Эти тесты очень чувствительны к предположениям модели, таким как предположение о нормальности. Поэтому общепринятой практикой является объединение тестов с диагностическими графиками, чтобы оценить серьезность гетероскедастичности и решить, нужна ли какая-либо коррекция гетероскедастичности.
*Для 14й STATA еще и нужно увеличить матрицу, иначе расчет будет невозможен.

set matsize 800
estat imtest
(долго считает, надо подождать)

```
Cameron & Trivedi's decomposition of IM-test

---------------------------------------------------
              Source |       chi2     df      p
---------------------+-----------------------------
  Heteroskedasticity |    1397.94    297    0.0000
            Skewness |     162.87     27    0.0000
            Kurtosis |     450.48      1    0.0000
---------------------+-----------------------------
               Total |    2011.29    325    0.0000
---------------------------------------------------
```

estat hettest

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of lg_Hwage1

         chi2(1)      =   331.44
         Prob > chi2  =   0.0000
```

Если значение p очень мало, нам придется отклонить гипотезу и принять альтернативную гипотезу о неоднородности дисперсии. Таким образом, в этом случае доказательство - против нулевой гипотезы о том, что дисперсия является однородной. Это значит, данные гетероскедастичны, и нам нужно корректировать нашу модель – сделать оценки робастными.

*32.7. Для получения робастных оценок регрессии (то есть скорректированных стандартных отклонений с учетом гетероскедастичности) используем специальную опцию. В такой оценке модели коэффициенты не меняются, но меняется их значимость. В нашем случае существенных изменений нет.
Stata Annotated Output Robust Regression
https://stats.oarc.ucla.edu/stata/output/robust-regression/

reg lg_Hwage1 i.diplom_k c.age_10##c.age_10 male ln_regwage i.status_1 i.fed_okr i.year, robust beta

```
Linear regression                               Number of obs     =     50,917
                                                F(27, 50889)      =     892.66
                                                Prob > F          =     0.0000
                                                R-squared         =     0.3195
                                                Root MSE          =     .75696

------------------------------------------------------------------------------
             |               Robust
   lg_Hwage1 |      Coef.   Std. Err.      t    P>|t|                     Beta
-------------+----------------------------------------------------------------
    diplom_k |
   законч  .. |      0.143      0.012   11.459   0.000                   0.075
   средне  .. |      0.299      0.013   23.230   0.000                   0.146
      высшее  |      0.565      0.013   43.305   0.000                   0.263
              |
      age_10  |      0.445      0.017   25.745   0.000                   0.580
              |
    c.age_10#|
```

```
        c.age_10 |     -0.058      0.002  -27.988   0.000                  -0.630
                 |
            male |      0.337      0.007   48.871   0.000                   0.184
       ln_regwage |      0.748      0.014   52.744   0.000                   0.415
                 |
        status_1 |
      обл.центр  |     -0.164      0.021   -7.851   0.000                  -0.084
   другой город  |     -0.160      0.022   -7.325   0.000                  -0.078
      село, пгт  |     -0.514      0.021  -23.955   0.000                  -0.245
                 |
         fed_okr |
       Северный  |      0.230      0.023   10.100   0.000                   0.064
    Центральный  |      0.122      0.020    5.981   0.000                   0.052
    Приволжский  |      0.034      0.021    1.626   0.104                   0.015
  Юг и С.Кавказ  |      0.062      0.022    2.873   0.004                   0.023
     Уральский   |      0.107      0.021    4.987   0.000                   0.034
      'Сибирский |     -0.031      0.021   -1.478   0.139                  -0.011
Дальневосточный  |      0.000  (omitted)                                   0.000
                 |
            year |
            1995 |     -0.117      0.020   -5.983   0.000                  -0.033
            1996 |     -0.184      0.020   -9.317   0.000                  -0.050
            1998 |     -0.270      0.020  -13.721   0.000                  -0.073
            2000 |     -0.429      0.019  -22.260   0.000                  -0.119
            2001 |     -0.369      0.019  -19.471   0.000                  -0.108
            2002 |     -0.321      0.018  -17.385   0.000                  -0.098
            2003 |     -0.318      0.019  -16.757   0.000                  -0.098
            2004 |     -0.295      0.019  -15.493   0.000                  -0.093
            2005 |     -0.285      0.020  -14.297   0.000                  -0.088
            2006 |     -0.287      0.020  -14.133   0.000                  -0.097
            2007 |     -0.280      0.021  -13.091   0.000                  -0.095
                 |
           _cons |     -3.914      0.132  -29.726   0.000                       .
------------------------------------------------------------------------------
```

*32.8. Робастные кластеризованные оценки регрессии.

Регрессия МНК предполагает, что остатки независимы. Наш набор данных, так как они панельные, содержит данные о более чем 50000 человек за 12 лет. Очевидно, что оценки внутри каждого индивидуума не могут быть независимыми, и это может привести к остаткам, которые не являются независимыми внутри индивидуумов. Мы можем использовать параметр кластера, чтобы указать, что наблюдения сгруппированы по индивидуумам (на основе **idind**) и что наблюдения могут быть коррелированы внутри индивидуумов, но будут независимыми между индивидуумами.

Теперь мы можем запустить регрессию с опцией «кластера». Нам не нужно включать робастную оценку, поскольку «кластер» уже подразумевает робастность. Обратите внимание, что стандартные ошибки существенно изменились, гораздо больше, чем изменение, вызванное опцией «робастность».

Робастные кластерные стандартные ошибки нельзя оценить вместе со стандартизованными коэффициентами (увы).

reg lg_Hwage1 i.diplom_k c.age_10##c.age_10 male ln_regwage i.status_1 i.fed_okr i.year, cluster(idind) beta

```
options vce(cluster clustvar) and beta may not be combined
r(184);
```

*Оценим регрессию с калстеризованными робастными стандартными отклонениями

reg lg_Hwage1 i.diplom_k c.age_10##c.age_10 male ln_regwage i.status_1 i.fed_okr i.year, cluster(idind)

```
note: 8.fed_okr omitted because of collinearity

Linear regression                               Number of obs     =     50,917
                                                F(27, 14564)      =     495.38
                                                Prob > F          =     0.0000
                                                R-squared         =     0.3195
                                                Root MSE          =     .75696

                                   (Std. Err. adjusted for 14,565 clusters in idind)
```

	Robust				
lg_Hwage1	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]
diplom_k					
законч ..	0.143	0.017	8.481	0.000	0.110 0.176
средне ..	0.299	0.018	16.417	0.000	0.263 0.334
высшее	0.565	0.019	29.882	0.000	0.528 0.602
age_10	0.445	0.025	18.041	0.000	0.397 0.493
c.age_10#					
c.age_10	-0.058	0.003	-19.281	0.000	-0.064 -0.053
male	0.337	0.011	30.786	0.000	0.316 0.359
ln_regwage	0.748	0.021	36.082	0.000	0.707 0.788
status_1					
обл.центр	-0.164	0.031	-5.216	0.000	-0.225 -0.102
другой город	-0.160	0.033	-4.789	0.000	-0.226 -0.095
село, пгт	-0.514	0.033	-15.441	0.000	-0.579 -0.449
fed_okr					
Северный	0.230	0.035	6.500	0.000	0.161 0.300
Центральный	0.122	0.032	3.805	0.000	0.059 0.185
Приволжский	0.034	0.033	1.037	0.300	-0.030 0.098
Юг и С.Кавказ	0.062	0.034	1.842	0.065	-0.004 0.128
Уральский	0.107	0.034	3.189	0.001	0.041 0.173
'Сибирский	-0.031	0.032	-0.958	0.338	-0.095 0.032
Дальневосточный	0.000	(omitted)			
year					
1995	-0.117	0.018	-6.692	0.000	-0.151 -0.083
1996	-0.184	0.019	-9.872	0.000	-0.220 -0.147
1998	-0.270	0.019	-14.417	0.000	-0.307 -0.233
2000	-0.429	0.019	-23.007	0.000	-0.465 -0.392
2001	-0.369	0.019	-19.808	0.000	-0.405 -0.332
2002	-0.321	0.019	-16.827	0.000	-0.358 -0.283
2003	-0.318	0.020	-15.661	0.000	-0.358 -0.278
2004	-0.295	0.021	-14.066	0.000	-0.337 -0.254
2005	-0.285	0.023	-12.549	0.000	-0.329 -0.240
2006	-0.287	0.024	-11.881	0.000	-0.334 -0.239
2007	-0.280	0.026	-10.719	0.000	-0.331 -0.229
_cons	-3.914	0.191	-20.505	0.000	-4.288 -3.539

Как и в случае робастной оценки, оценка коэффициентов для «кластеров» такая же, как и оценка МНК, но стандартные ошибки учитывают, что наблюдения внутри индивидуумов не являются независимыми. Несмотря на то, что стандартные ошибки в этом случае больше, переменные, которые были значимы в анализе МНК, также значимы в этом анализе. Эти стандартные ошибки рассчитываются на основе агрегированных значений для 14365 человек, поскольку между индивидами наблюдения являются независимыми. Если у вас очень небольшое количество кластеров по сравнению с общим размером выборки, возможно, что стандартные ошибки могут быть значительно больше, чем в результате обычного МНК.

*32.9. Расчет предельных эффектов **dydx(*)**, или производных (Derivatives).
Эта операция занимает довольно много времени. Команда относится к последней выполненной регрессии. Можно указать отдельные переменные или (*) – для всех переменных. Правда, для исходной регрессии предельные эффекты рассчитываются не для всех переменных, поэтому в модели заменим тип поседения на более простую переменную – «сельская местность». При этом в регрессию включаются все федеральные округа (базовая категория – Москва и Санкт-Петербург), и они все значимы (хотя Центральный регион – с уровнем 10%).

reg lg_Hwage1 i.diplom_k c.age_10##c.age_10 male ln_regwage village i.fed_okr i.year, cluster(idind)

```
Linear regression                               Number of obs     =     50,917
                                                F(26, 14564)      =     514.15
                                                Prob > F          =     0.0000
                                                R-squared         =     0.3195
                                                Root MSE          =     .75696

                                  (Std. Err. adjusted for 14,565 clusters in idind)
------------------------------------------------------------------------------
             |               Robust
   lg_Hwage1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    diplom_k |
  законч ..  |      0.143      0.017     8.479   0.000        0.110       0.176
  средне ..  |      0.299      0.018    16.418   0.000        0.263       0.334
     высшее  |      0.565      0.019    29.894   0.000        0.528       0.602
             |
      age_10 |      0.445      0.025    18.054   0.000        0.397       0.494
             |
    c.age_10#|
    c.age_10 |     -0.059      0.003   -19.289   0.000       -0.064      -0.053
             |
        male |      0.337      0.011    30.783   0.000        0.316       0.359
  ln_regwage |      0.748      0.021    36.297   0.000        0.708       0.788
     village |     -0.352      0.015   -23.622   0.000       -0.381      -0.323
             |
     fed_okr |
   Северный  |      0.068      0.026     2.659   0.008        0.018       0.118
Центральный  |     -0.040      0.021    -1.893   0.058       -0.082       0.001
Приволжский  |     -0.128      0.022    -5.887   0.000       -0.170      -0.085
Юг и С.Кавказ|     -0.100      0.025    -4.061   0.000       -0.148      -0.052
 Уральский   |     -0.056      0.023    -2.407   0.016       -0.101      -0.010
  'Сибирский |     -0.193      0.021    -9.124   0.000       -0.234      -0.152
Дальневосточный|     -0.163      0.031    -5.198   0.000       -0.224      -0.101
             |
        year |
        1995 |     -0.117      0.018    -6.692   0.000       -0.151      -0.083
        1996 |     -0.184      0.019    -9.884   0.000       -0.220      -0.147
        1998 |     -0.270      0.019   -14.416   0.000       -0.307      -0.233
        2000 |     -0.429      0.019   -23.029   0.000       -0.466      -0.392
        2001 |     -0.369      0.019   -19.836   0.000       -0.406      -0.333
        2002 |     -0.321      0.019   -16.872   0.000       -0.358      -0.283
        2003 |     -0.318      0.020   -15.711   0.000       -0.358      -0.279
        2004 |     -0.296      0.021   -14.110   0.000       -0.337      -0.255
        2005 |     -0.285      0.023   -12.599   0.000       -0.330      -0.241
        2006 |     -0.287      0.024   -11.934   0.000       -0.334      -0.240
        2007 |     -0.280      0.026   -10.774   0.000       -0.331      -0.229
             |
       _cons |     -3.917      0.190   -20.639   0.000       -4.290      -3.545
```

margins, dydx(*)

(считает довольно долго, надо подождать)

```
Average marginal effects                        Number of obs     =     50,917
Model VCE    : Robust

Expression   : Linear prediction, predict()
dy/dx w.r.t. : 2.diplom_k 3.diplom_k 4.diplom_k age_10 male ln_regwage village
               2.fed_okr 3.fed_okr 4.fed_okr 5.fed_okr 6.fed_okr 7.fed_okr
               8.fed_okr 1995.year 1996.year 1998.year 2000.year 2001.year
               2002.year 2003.year 2004.year 2005.year 2006.year 2007.year

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    diplom_k |
  законч ..  |      0.143      0.017     8.479   0.000        0.110       0.176
  средне ..  |      0.299      0.018    16.418   0.000        0.263       0.334
     высшее  |      0.565      0.019    29.894   0.000        0.528       0.602
             |
      age_10 |     -0.014      0.004    -3.327   0.001       -0.022      -0.006
        male |      0.337      0.011    30.783   0.000        0.316       0.359
  ln_regwage |      0.748      0.021    36.297   0.000        0.708       0.788
     village |     -0.352      0.015   -23.622   0.000       -0.381      -0.323
             |
     fed_okr |
   Северный  |      0.068      0.026     2.659   0.008        0.018       0.118
Центральный  |     -0.040      0.021    -1.893   0.058       -0.082       0.001
Приволжский  |     -0.128      0.022    -5.887   0.000       -0.170      -0.085
```

```
  Юг и С.Кавказ  |     -0.100      0.025   -4.061   0.000       -0.148      -0.052
     Уральский   |     -0.056      0.023   -2.407   0.016       -0.101      -0.010
     'Сибирский  |     -0.193      0.021   -9.124   0.000       -0.234      -0.152
Дальневосточный  |     -0.163      0.031   -5.198   0.000       -0.224      -0.101
                 |
            year |
            1995 |     -0.117      0.018   -6.692   0.000       -0.151      -0.083
            1996 |     -0.184      0.019   -9.884   0.000       -0.220      -0.147
            1998 |     -0.270      0.019  -14.416   0.000       -0.307      -0.233
            2000 |     -0.429      0.019  -23.029   0.000       -0.466      -0.392
            2001 |     -0.369      0.019  -19.836   0.000       -0.406      -0.333
            2002 |     -0.321      0.019  -16.872   0.000       -0.358      -0.283
            2003 |     -0.318      0.020  -15.711   0.000       -0.358      -0.279
            2004 |     -0.296      0.021  -14.110   0.000       -0.337      -0.255
            2005 |     -0.285      0.023  -12.599   0.000       -0.330      -0.241
            2006 |     -0.287      0.024  -11.934   0.000       -0.334      -0.240
            2007 |     -0.280      0.026  -10.774   0.000       -0.331      -0.229
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.
```

Обратите внимание, что нет интеракций (для возраста).
Производная **dydx** показывает, как изменяется зависимая переменная (логарифм ставки ЗП) при малых изменениях каждой из переменных («скорость изменения» в данной точке). Самое сильное влияние по этому параметру – региональной заработной платы.
dydx(varlist), eyex(varlist), dyex(varlist), and eydx(varlist) – эти опции требуют, чтобы **margins** выводила в отчете производные по отношению к переменным, которые указаны в списке varlist.
eyex(), dyex(), and eydx() – эти опции выводят в отчете производные как эластичность (см. в “Help” - Expressing derivatives as elasticities in [R] margins).

Эластичность - мера чувствительности одного из параметров (например, дохода) к изменению другого (например, возраста), показывающая, на сколько процентов изменится первый показатель при изменении второго на 1%. Функция называется эластичной, когда показатель эластичности больше единицы, и неэластичной – если наоборот.

margins, eydx(*)
(считает довольно долго, надо подождать)

```
Average marginal effects                        Number of obs     =     50,917
Model VCE    : Robust

Expression   : Linear prediction, predict()
ey/dx w.r.t. : 2.diplom_k 3.diplom_k 4.diplom_k age_10 male ln_regwage village
               2.fed_okr 3.fed_okr 4.fed_okr 5.fed_okr 6.fed_okr 7.fed_okr
               8.fed_okr 1995.year 1996.year 1998.year 2000.year 2001.year
               2002.year 2003.year 2004.year 2005.year 2006.year 2007.year

------------------------------------------------------------------------------
                 |            Delta-method
                 |      ey/dx   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----------------+------------------------------------------------------------
        diplom_k |
      законч ..  |      0.045      0.005    8.369   0.000        0.034       0.056
      средне ..  |      0.092      0.006   16.074   0.000        0.081       0.103
         высшее  |      0.167      0.006   28.928   0.000        0.156       0.178
                 |
          age_10 |     -0.004      0.001   -3.385   0.001       -0.007      -0.002
            male |      0.100      0.003   30.610   0.000        0.094       0.107
      ln_regwage |      0.222      0.006   36.069   0.000        0.210       0.234
         village |     -0.105      0.004  -23.409   0.000       -0.113      -0.096
                 |
         fed_okr |
        Северный |      0.019      0.007    2.669   0.008        0.005       0.034
    Центральный  |     -0.012      0.006   -1.896   0.058       -0.024       0.000
    Приволжский  |     -0.038      0.006   -5.910   0.000       -0.050      -0.025
  Юг и С.Кавказ  |     -0.029      0.007   -4.062   0.000       -0.043      -0.015
     Уральский   |     -0.016      0.007   -2.404   0.016       -0.030      -0.003
     'Сибирский  |     -0.058      0.006   -9.127   0.000       -0.070      -0.045
Дальневосточный  |     -0.048      0.009   -5.125   0.000       -0.067      -0.030
                 |
            year |
```

```
        1995  |     -0.033      0.005   -6.673   0.000       -0.042      -0.023
        1996  |     -0.052      0.005   -9.901   0.000       -0.062      -0.042
        1998  |     -0.077      0.005  -14.241   0.000       -0.088      -0.067
        2000  |     -0.126      0.005  -22.952   0.000       -0.137      -0.115
        2001  |     -0.107      0.005  -19.964   0.000       -0.118      -0.097
        2002  |     -0.093      0.005  -17.059   0.000       -0.103      -0.082
        2003  |     -0.092      0.006  -15.842   0.000       -0.103      -0.080
        2004  |     -0.085      0.006  -14.211   0.000       -0.097      -0.073
        2005  |     -0.082      0.006  -12.635   0.000       -0.095      -0.069
        2006  |     -0.082      0.007  -11.943   0.000       -0.096      -0.069
        2007  |     -0.080      0.007  -10.739   0.000       -0.095      -0.066
------------------------------------------------------------------------------
Note: ey/dx for factor levels is the discrete change from the base level.
```

Expressing derivatives as elasticities

You specify the `dydx(`*varname*`)` option on the `margins` command to use $dy/d(varname)$ as the response variable. If you want that derivative expressed as an elasticity, you can specify `eyex(`*varname*`)`, `eydx(`*varname*`)`, or `dyex(`*varname*`)`. You substitute `e` for `d` where you want an elasticity. The formulas are

$$\texttt{dydx()} = dy/dx$$
$$\texttt{eyex()} = dy/dx \times (x/y)$$
$$\texttt{eydx()} = dy/dx \times (1/y)$$
$$\texttt{dyex()} = dy/dx \times (x)$$

and the interpretations are

`dydx()`:		change in y for a		change in x
`eyex()`:	proportional	change in y for a	proportional	change in x
`eydx()`:	proportional	change in y for a		change in x
`dyex()`:		change in y for a	proportional	change in x

As `margins` always does with response functions, calculations are made at the observational level and are then averaged. Let's assume that in observation 5, $dy/dx = 0.5$, $y = 15$, and $x = 30$; then

$$\texttt{dydx()} = 0.5$$
$$\texttt{eyex()} = 1.0$$
$$\texttt{eydx()} = 0.03$$
$$\texttt{dyex()} = 15.0$$

Many social scientists would informally explain the meaning of `eyex()` $= 1$ as "y increases 100% when x increases 100%" or as "y doubles when x doubles", although neither statement is literally true. `eyex()`, `eydx()`, and `dyex()` are rates evaluated at a point, just as `dydx()` is a rate, and all such interpretations are valid only for small (infinitesimal) changes in x. It is true that `eyex()` $= 1$ means y increases with x at a rate such that, if the rate were constant, y would double if x doubled. This issue of casual interpretation is no different from casually interpreting `dydx()` as if it represents the response to a unit change. It is not necessarily true that `dydx()` $= 0.5$ means that "y increases by 0.5 if x increases by 1". It is true that "y increases with x at a rate such that, if the rate were constant, y would increase by 0.5 if x increased by 1".

`dydx()`, `eyex()`, `eydx()`, and `dyex()` may be used with continuous x variables. `dydx()` and `eydx()` may also be used with factor variables.

Но для функции **margins, eyex(*)** мы не можем использовать «factor variables» и интеракции.

```
factor variables not allowed in option eyex()
r(198);
```

*(Если бы мы использовали не интеракции, а **age_10** и переменную «возраст в квадрате» **age_q2,** можно было бы посчитать)

gen age_q2 = age_10 * age_10
reg lg_Hwage1 i.diplom_k age_10 age_q2 male ln_regwage village i.fed_okr i.year, cluster(idind)

margins, eyex(age_10 age_q2)

```
Average marginal effects                        Number of obs     =     50,917
Model VCE    : Robust

Expression   : Linear prediction, predict()
ey/ex w.r.t. : age_10 age_q2

------------------------------------------------------------------------------
             |            Delta-method
             |      ey/ex   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      age_10 |       0.521      0.029   18.108   0.000         0.464       0.577
      age_q2 |      -0.295      0.015  -19.222   0.000        -0.325      -0.265
------------------------------------------------------------------------------
```

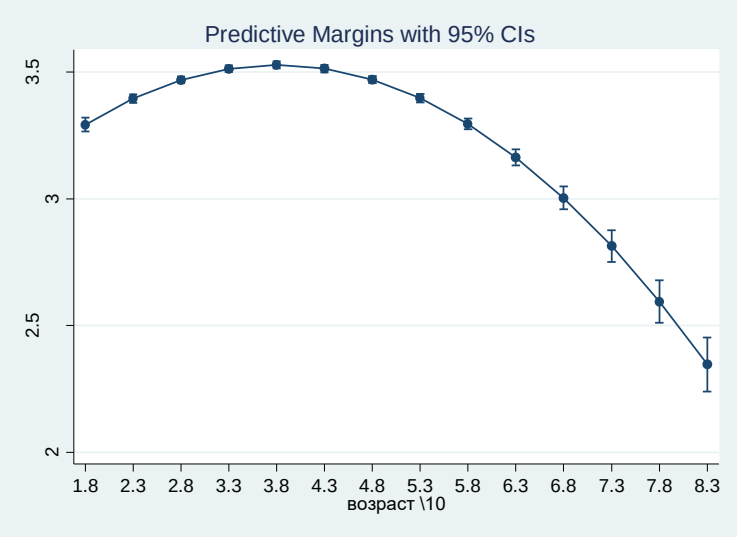
Верно, что eyex () = 1 означает, что если бы темп роста был постоянным, **y** увеличивается с ростом **x**, таким образом, что, **y** удвоился бы, если бы **x** удвоился.

*32.10. Теперь снова вернемся к модели с интеракциями, чтобы можно было построить график предсказанных значений логарифма ставки заработной платы при разных значениях возраста.

reg lg_Hwage1 i.diplom_k c.age_10##c.age_10 male ln_regwage village i.fed_okr i.year, cluster(idind)

*Рассчитаем предсказанные значения y для возраста\10 (возраст от 1,8 до 8,5 с интервалом 0,5, то есть через 5 лет), и построим график. Сохраните график (его можно редактировать).

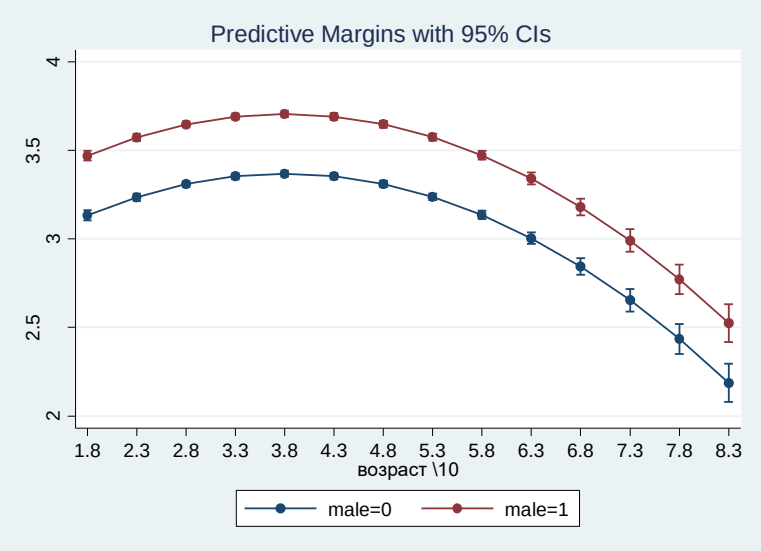
margins, at(age_10 ==(1.8(0.5)8.5))
marginsplot



На графике вы видите предсказанные значения логарифма ставки заработной платы в зависимости от возраста; вы видите, что «пик» достигается примерно в 38 лет; затем ставка заработной платы быстро снижается, к 60ти годам достигая уровня ниже, чем в 18 лет (конечно, это усредненные значения для всех периодов; эффект возраста здесь совмещен с эффектом когорт, то есть старшие поколения быстро теряют в заработке, в том числе из-за обесценения образования, полученного в СССР).

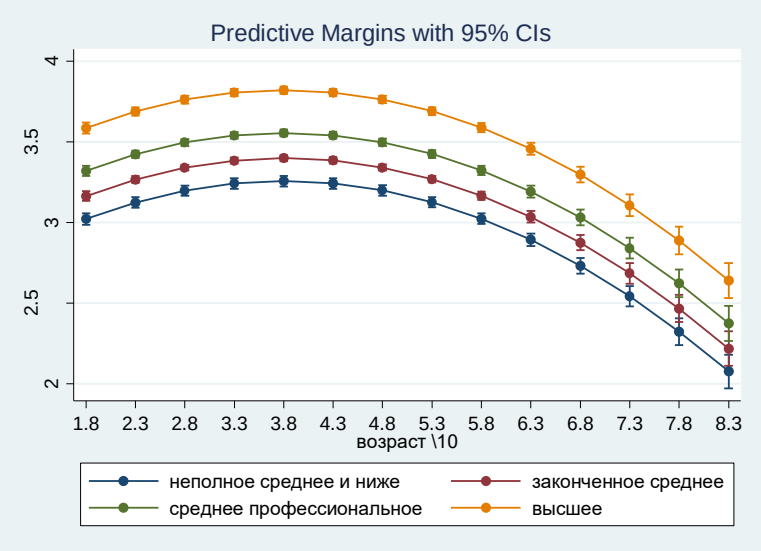
*Теперь учтем также пол

margins, at (age_10 ==(1.8(0.5)8.5) male == (0 1))
marginsplot



*учтем уровень образования

margins, at (age_10 ==(1.8(0.5)8.5) diplom_k == (1 2 3 4))
marginsplot



На втором графике – зависимость от возраста для мужчин и женщин; на третьем – для разных уровней образования.

Video examples

Introduction to margins, part 1: Categorical variables
https://www.youtube.com/watch?v=XAG4CbIbH0k
Introduction to margins, part 2: Continuous variables
https://www.youtube.com/watch?v=L9-PWY79aVA

Introduction to margins, part 3: Interactions
https://www.youtube.com/watch?v=43uX4D_7uaI

*32.11. Взвешенная регрессия (индивидуальные веса). Учтем теперь веса, для репрезентативности. Количество кейсов изменилось, и количество кластеров, то есть индивидов. Есть различия в значимости регионов. Либо, чтобы использовать репрезентативные данные, но не взвешенную регрессию, можно использовать условие **origsm == 1**
pweights, or sampling weights, это веса, обратные вероятности того, что наблюдение включено из-за схемы выборки.

reg lg_Hwage1 i.diplom_k c.age_10##c.age_10 male ln_regwage village i.fed_okr i.year [pweight = inwgt], cluster(idind)

```
(sum of wgt is   3.9175e+04)

Linear regression                               Number of obs     =     39,543
                                                F(26, 12752)      =     389.90
                                                Prob > F          =     0.0000
                                                R-squared         =     0.3112
                                                Root MSE          =     .78597

                                        (Std. Err. adjusted for 12,753 clusters in idind)
------------------------------------------------------------------------------------------
                         |               Robust
               lg_Hwage1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------------------+----------------------------------------------------------------
                diplom_k |
      законченное среднее  |      0.164      0.019     8.533   0.000        0.126       0.202
 среднее профессиональное  |      0.317      0.021    15.344   0.000        0.276       0.357
                  высшее  |      0.569      0.022    26.433   0.000        0.527       0.611
                         |
                  age_10 |      0.464      0.028    16.553   0.000        0.409       0.519
                         |
       c.age_10#c.age_10 |     -0.060      0.003   -17.474   0.000       -0.067      -0.053
                         |
                    male |      0.330      0.012    26.584   0.000        0.306       0.354
              ln_regwage |      0.764      0.022    33.993   0.000        0.720       0.808
                 village |     -0.388      0.017   -22.992   0.000       -0.421      -0.355
                         |
                 fed_okr |
               Северный  |      0.041      0.029     1.396   0.163       -0.017       0.098
            Центральный  |     -0.026      0.024    -1.113   0.266       -0.073       0.020
            Приволжский  |     -0.129      0.024    -5.397   0.000       -0.176      -0.082
          Юг и С.Кавказ  |     -0.081      0.027    -2.971   0.003       -0.135      -0.028
              Уральский  |     -0.043      0.025    -1.719   0.086       -0.093       0.006
              'Сибирский  |     -0.170      0.024    -7.015   0.000       -0.218      -0.123
        Дальневосточный  |     -0.136      0.034    -4.048   0.000       -0.202      -0.070
                         |
                    year |
                   1995  |     -0.119      0.018    -6.535   0.000       -0.155      -0.083
                   1996  |     -0.203      0.020   -10.293   0.000       -0.242      -0.165
                   1998  |     -0.292      0.020   -14.755   0.000       -0.331      -0.253
                   2000  |     -0.455      0.021   -22.111   0.000       -0.495      -0.414
                   2001  |     -0.376      0.021   -18.137   0.000       -0.416      -0.335
                   2002  |     -0.329      0.021   -15.379   0.000       -0.371      -0.287
                   2003  |     -0.331      0.023   -14.557   0.000       -0.376      -0.287
                   2004  |     -0.315      0.023   -13.522   0.000       -0.361      -0.270
                   2005  |     -0.317      0.025   -12.597   0.000       -0.366      -0.268
                   2006  |     -0.308      0.026   -11.651   0.000       -0.360      -0.256
                   2007  |     -0.309      0.028   -10.866   0.000       -0.365      -0.254
                         |
                   _cons |     -4.114      0.208   -19.779   0.000       -4.521      -3.706
------------------------------------------------------------------------------------------
```

*32.12. Предскажем теперь логарифм ставки заработной платы для всех респондентов (как занятых, так и незанятых) на основании значений тех переменных, которые входят в регрессию (это можно сделать, так как мы не включали в регрессию переменных, которые относятся к предприятию или рабочему месту: например, отрасль, тип собственности, и т.д.)

*И посмотрим средние значения для занятых и для незанятых (для репрезентативной выборки; pweight не разрешены).

predict xb
tabstat xb if origsm ==1 , statistics(mean) by(employed)

```
Summary for variables: xb
     by categories of: employed (работает)

        employed |      mean
-----------------+----------
не работает или  |  2.799375
есть любая работ |  3.372102
-----------------+----------
           Total |  3.114137
----------------------------
```

Предсказанная ставка заработной платы для незанятых ниже, чем для занятых; это означает, что они не выходят на работу из-за того, что предлагаемая им ставка заработной платы ниже резервной.

*32.13. Вывод результатов в текстовый файл.
outreg2 обеспечивает быстрый и простой способ создания иллюстративной таблицы выходных данных регрессии. Выходные данные регрессии создаются по частям, и их трудно сравнивать без какой-либо реорганизации. **outreg2** автоматизирует этот процесс, объединяя последовательные выходные данные регрессии в вертикальном формате. Полученная таблица сохраняется на диск в формате, который может быть прочитан другими программами. По умолчанию выводятся коэффициенты и стандартные ошибки в скобках.
Full syntax:
outreg2 [varlist] [estlist] using filename [, options] [: command]
Перед запуском новой команды нужно нажать *«enter»* .

reg lg_Hwage1 i.diplom_k c.age_10##c.age_10 male ln_regwage village i.fed_okr i.year [pweight = inwgt], cluster(idind)
outreg2 using C:\RLMS_work\seminar_7\data\mincer, label seeout
«enter»

Опция **label** позволяет вывести метки переменных, а не их имена. Опция **seeout** позволяет вывести результаты в таблицу (аналог таблицы данных), откуда их легко можно скопировать в excel, например.

Помимо этого, **outreg2** позволяет в туже же таблицу добавить оценку регрессии для разных категорий, например, для мужчин и для женщин.

reg lg_Hwage1 i.diplom_k c.age_10##c.age_10 male ln_regwage village i.fed_okr i.year if male==1 [pweight = inwgt], cluster(idind)
outreg2 using C:\RLMS_work\seminar_7\data\mincer, label seeout
«enter»

После этой команды не нажимайте *«enter»* , а скопируйте результаты в файл excel

reg lg_Hwage1 i.diplom_k c.age_10##c.age_10 male ln_regwage village i.fed_okr i.year if male==0 [pweight = inwgt], cluster(idind)
outreg2 using C:\RLMS_work\seminar_7\data\mincer, label seeout

Пример оформления в исследовании (из файла, полученного при помощи **outreg2)**.

Таблица 1. Коэффициенты оценки регрессии (МНК), зависимая переменная - логарифм ставки ЗП за 30 дней на первой работе, взвешенные данные.

v1	v2	v3	v4
	-1	-2	-3
VARIABLES	ВСЕ	Мужчины	Женщины
законченное образование = 2, законченное среднее (базовая категория: нет полного среднего)	0.164***	0.153***	0.177***
	(0.0192)	(0.0247)	(0.0293)
законченное образование = 3, среднее профессиональное	0.317***	0.297***	0.345***
	(0.0206)	(0.0289)	(0.0297)
законченное образование = 4, высшее	0.569***	0.468***	0.658***
	(0.0215)	(0.0299)	(0.0308)
возраст \10	0.464***	0.472***	0.452***
	(0.0280)	(0.0394)	(0.0397)
c.age_10#c.age_10 (возраст в квадрате)	-0.0601***	-0.0616***	-0.0579***
	(0.00344)	(0.00480)	(0.00493)
мужской пол	0.330***		
	(0.0124)		
логарифм дефлир региональной ЗП	0.764***	0.832***	0.694***
	(0.0225)	(0.0325)	(0.0300)
село	-0.388***	-0.510***	-0.271***
	(0.0169)	(0.0252)	(0.0219)
федеральный округ = 2, Северный (базовая категория – Москва и СП)	0.0409	0.117**	-0.0296
	(0.0293)	(0.0463)	(0.0375)
федеральный округ = 3, Центральный	-0.0263	0.0510	-0.111***
	(0.0236)	(0.0348)	(0.0317)
федеральный округ = 4, Приволжский	-0.129***	-0.0306	-0.231***
	(0.0240)	(0.0356)	(0.0315)
федеральный округ = 5, Юг и С.Кавказ	-0.0812***	0.0355	-0.198***
	(0.0273)	(0.0395)	(0.0369)
федеральный округ = 6, Уральский	-0.0435*	0.0228	-0.114***
	(0.0253)	(0.0359)	(0.0350)
федеральный округ = 7, 'Сибирский	-0.170***	-0.142***	-0.205***
	(0.0243)	(0.0366)	(0.0315)
федеральный округ = 8, Дальневосточный	-0.136***	-0.0759	-0.178***
	(0.0337)	(0.0494)	(0.0446)
1995 (базовая категория – 1994)	-0.119***	-0.135***	-0.102***
	(0.0182)	(0.0275)	(0.0235)
1996	-0.203***	-0.241***	-0.161***
	(0.0198)	(0.0301)	(0.0252)
1998	-0.292***	-0.306***	-0.281***
	(0.0198)	(0.0291)	(0.0266)
2000	-0.455***	-0.461***	-0.452***
	(0.0206)	(0.0307)	(0.0271)
2001	-0.376***	-0.370***	-0.384***
	(0.0207)	(0.0309)	(0.0274)
2002	-0.329***	-0.364***	-0.295***
	(0.0214)	(0.0316)	(0.0283)
2003	-0.331***	-0.351***	-0.313***
	(0.0228)	(0.0339)	(0.0300)
2004	-0.315***	-0.346***	-0.286***
	(0.0233)	(0.0344)	(0.0309)
2005	-0.317***	-0.354***	-0.283***
	(0.0252)	(0.0368)	(0.0336)
2006	-0.308***	-0.392***	-0.228***

	(0.0264)	(0.0387)	(0.0351)
2007	-0.309***	-0.386***	-0.234***
	(0.0285)	(0.0413)	(0.0381)
Constant		-	-
	-4.114***	-4.366***	-3.507***
Observations	(0.208)	(0.300)	(0.278)
R-squared			

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

К сожалению, в файл Word лейблы на русском не выводятся, а файл в формате txt не удобный.

Выделим все команды в левом верхнем окне “Command”, нажмем правую клавишу мыши и выберем опцию «пернести в do-файл». Сохранить получившийся файл. Если перенести куда-то или удалить все файлы результатов, можно этот командный файл «прогнать» целиком. При необходимости можно делать изменения (исправлять ошибки и т.д.)

32.14. Сравнение моделей для подвыборок. Без опции **cluster(idind)** и без робастных оценок. Снова оценим все три модели – для всей выборки, для мужчин и для женщин, и сохраним результаты под оригинальным именем для каждой модели. Веса не будем использовать, так как иначе нельзя выполнить один из тестов. Но используем репрезентативную выборку.

```
reg lg_Hwage1 i.diplom_k c.age_10##c.age_10 male ln_regwage village i.fed_okr i.year  if origsm ==1
estimates store M1_all

reg lg_Hwage1 i.diplom_k c.age_10##c.age_10 male ln_regwage village i.fed_okr i.year if male==1 & origsm ==1
estimates store M2_male

reg lg_Hwage1 i.diplom_k c.age_10##c.age_10 male ln_regwage village i.fed_okr i.year if male==0 & origsm ==1
estimates store M3_female
```

*Команда **estimates table** позволяет сравнить результаты трех моделей. Опции:
b(%7.3f) задает формат вывода коэффициентов
star[(#1 #2 #3)] использует звездочки для обозначения уровней значимости
stats(scalarlist) выводит scalarlist в таблице
Например, stats(N ll chi2 aic) выводит в таблице e(N), e(ll), e(chi2), и AIC. В Stata, e(N) это количество наблюдений; e(ll), - log likelihood; e(chi2) - тест chi-squared - что все коэффициенты в первом уравнении равны нулю.

```
estimates table M1_all M2_male M3_female, b(%7.3f) star stats(N)
```

```
-------------------------------------------------------------
    Variable |    M1_all        M2_male       M3_female
-------------+-----------------------------------------------
    diplom_k |
    зако..   |     0.163***       0.151***       0.178***
среднее п..  |     0.319***       0.299***       0.346***
     высшее  |     0.578***       0.474***       0.661***
             |
      age_10 |     0.459***       0.460***       0.455***
             |
    c.age_10#|
    c.age_10 |    -0.059***      -0.060***      -0.058***
             |
        male |     0.329***   (omitted)      (omitted)
```

```
  ln_regwage |     0.758***       0.829***       0.692***
     village |    -0.381***      -0.507***      -0.273***
             |
     fed_okr |
   Северный  |     0.041*         0.119***      -0.025
Центральный  |    -0.028          0.054*        -0.108***
Приволжский  |    -0.130***      -0.027         -0.224***
   Юг и  ..  |    -0.092***       0.031         -0.202***
 Уральский   |    -0.047**        0.024         -0.114***
 'Сибирский  |    -0.168***      -0.139***      -0.200***
Дальневос~й  |    -0.149***      -0.092**       -0.181***
             |
        year |
       1995  |    -0.118***      -0.134***      -0.101***
       1996  |    -0.198***      -0.235***      -0.162***
       1998  |    -0.293***      -0.306***      -0.284***
       2000  |    -0.455***      -0.459***      -0.453***
       2001  |    -0.378***      -0.371***      -0.385***
       2002  |    -0.326***      -0.360***      -0.296***
       2003  |    -0.331***      -0.350***      -0.312***
       2004  |    -0.311***      -0.344***      -0.280***
       2005  |    -0.311***      -0.348***      -0.279***
       2006  |    -0.303***      -0.388***      -0.230***
       2007  |    -0.305***      -0.385***      -0.234***
             |
       _cons |    -4.052***      -4.321***      -3.501***
-------------+-------------------------------------------
           N |     39544          18464          21080
-------------------------------------------------------------
                  legend: * p<0.05; ** p<0.01; *** p<0.001
```

Сравнение коэффициентов между моделями по значимости не дает существенных различий (кроме регионов). Мы можем формально проверить, что коэффициенты одинаковы для полной модели M1 и моделей на подвыборках M2 и M3, используя команду **hausman**. Команда **hausman** ожидает, что модели будут указаны в порядке «всегда непротиворечивые» (“always consistent”) в первую очередь и «эффективные при H0» (“efficient under H0”) во вторую.

hausman выполняет Hausman's specification test. Чтобы использовать команду **hausman**, нужно выполнить шаги.

(1) получить оценку, которая непротиворечива независимо от того, верна гипотеза или нет;

(2) сохранить результаты оценки в соответствии с именами с помощью **estimates store**;

(3) получить оценку, которая эффективна (и непротиворечива) при гипотезе, которую вы проверяете, но неконсистентна в противном случае ;

(4) сохранить результаты оценки под эффективным именем с помощью **estimates store**;

(5) используйте команду **hausman** чтобы выполнить тест.

Команда выглядит как

hausman name-consistent name-efficient [, options]

alleqs указывает, что все уравнения в моделях должны использоваться для выполнения теста Хаусмана; по умолчанию используется только первое уравнение.

constant указывает, что оценочные точки пересечения (the estimated intercept(s)) должны быть включены в сравнение моделей; по умолчанию они исключены. Поведение по умолчанию подходит для моделей, в которых константа не имеет общей интерпретации в обеих моделях. Предположение, что одна из оценок эффективна (то есть имеет минимальную асимптотическую дисперсию), является обязательным. Это нарушается, например, если ваши наблюдения сгруппированы или взвешены, или если ваша модель каким-то образом неправильно определена. Более того, даже если предположение выполняется, может возникнуть проблема «малой выборки» с тестом Хаусмана. Тест Хаусмана основан на оценке дисперсии var(b-B) разности оценок по разнице var(b)-var(B) дисперсий. При предположениях (1) и (3) var(b)-var(B) является последовательной оценкой var(b-B), но не обязательно положительно определенной «в конечных выборках», то есть в вашем случае. В этом случае тест Хаусмана не определен. К сожалению, это не редкое событие. Stata поддерживает обобщенный тест Хаусмана, который преодолевает обе эти проблемы.

*Сравнение полной модели с моделью для мужчин

hausman M1_all M2_male, alleqs constant

```
                ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |     M1_all      M2_male       Difference          S.E.
-------------+------------------------------------------------------------
    diplom_k |
           2 |    .1626427     .1507108        .0119318               .
           3 |      .31916     .2994625        .0196974               .
           4 |    .5777411     .4735398        .1042012               .
      age_10 |    .4586642     .4597992        -.001135               .
    c.age_10#|
    c.age_10 |   -.0593473    -.0599368        .0005894               .
  ln_regwage |    .7578637     .8291796       -.0713159               .
     village |   -.3808714    -.5072342        .1263628               .
     fed_okr |
           2 |    .0410654     .1193437       -.0782783               .
           3 |   -.0283612      .054126       -.0824872               .
           4 |   -.1303683     -.026947       -.1034213               .
           5 |   -.0917446     .0312977       -.1230423               .
           6 |   -.0469315     .0242467       -.0711782               .
           7 |    -.167724    -.1386051       -.0291188               .
           8 |   -.1488885    -.0920129       -.0568756               .
        year |
        1995 |   -.1176535    -.1340331        .0163795               .
        1996 |   -.1984513    -.2349132        .0364618               .
        1998 |   -.2931291    -.3062317        .0131025               .
        2000 |   -.4547292    -.4592371         .004508               .
        2001 |   -.3780346    -.3707137       -.0073209               .
        2002 |   -.3263522    -.3597856        .0334334               .
        2003 |   -.3309666    -.3504915        .0195249               .
        2004 |    -.310711    -.3438228        .0331118               .
        2005 |   -.3108087    -.3482138        .0374051               .
        2006 |   -.3027019    -.3883131        .0856112               .
        2007 |    -.304715    -.3850998        .0803848               .
       _cons |   -4.052189    -4.321114        .2689253               .
------------------------------------------------------------------------------
                        b = consistent under Ho and Ha; obtained from regress
         B = inconsistent under Ha, efficient under Ho; obtained from regress

    Test:  Ho:  difference in coefficients not systematic

                chi2(26) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                         =-1.31e+05    chi2<0 ==> model fitted on these
                                       data fails to meet the asymptotic
                                       assumptions of the Hausman test;
                                       see suest for a generalized test
```

Тест Хаусмана не имеет четкого определения, что происходит довольно часто. Проблема связана с оценкой дисперсии V(b-B) как V(b)-V(B), которая является допустимой оценкой только асимптотически. Здесь это просто неправильная матрица дисперсии, и критерий Хаусмана становится неопределенным. (то же самое для модели для женщин).

Поэтому нужно использовать другой способ.
suest - Seemingly unrelated estimation («Казалось бы, несвязанная оценка»)
Типичными применениями **suest** являются проверки внутримодельных и кросс-модельных гипотез с использованием **test** или **testnl**, например, обобщенного теста спецификации Хаусмана.
suest m1 m2 оценивает одновременную (ко)дисперсию коэффициентов моделей m1 и m2.
Хотя **suest** технически является командой постоценки, она действует как команда оценки, поскольку сохраняет одновременные коэффициенты в **e(b)** и полную матрицу (ко)дисперсии в **e(V)**. Мы могли бы использовать команду **estat vce** для отображения полной матрицы (ко)дисперсии, чтобы показать, что межмодельные ковариации действительно были оценены. Как правило, мы не имеем прямого интереса к e(V).
В данном случае нас интересует проверка теста Чоу после команды **suest.**

Но теперь мы проверим наличие разницы в коэффициентах между моделями для мужчин и для женщин. Можем также добавить в оценку параметр **cluster(idind).**

suest M2_male M3_female, cluster(idind)

```
Simultaneous results for M2_male, M3_female
                                                  Number of obs     =     39,544

                                    (Std. Err. adjusted for 12,753 clusters in idind)
------------------------------------------------------------------------------------
                           |               Robust
                           |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------------------+--------------------------------------------------------------
M2_male_mean               |
                  diplom_k |
     законченное среднее   |      0.151      0.025     6.091   0.000        0.102       0.199
среднее профессиональное   |      0.299      0.029    10.330   0.000        0.243       0.356
                  высшее   |      0.474      0.030    15.781   0.000        0.415       0.532
                           |
                   age_10  |      0.460      0.040    11.635   0.000        0.382       0.537
                           |
        c.age_10#c.age_10  |     -0.060      0.005   -12.473   0.000       -0.069      -0.051
                           |
                     male  |      0.000  (omitted)
               ln_regwage  |      0.829      0.032    25.786   0.000        0.766       0.892
                  village  |     -0.507      0.025   -20.116   0.000       -0.557      -0.458
                           |
                  fed_okr  |
                Северный   |      0.119      0.046     2.617   0.009        0.030       0.209
             Центральный   |      0.054      0.035     1.563   0.118       -0.014       0.122
             Приволжский   |     -0.027      0.036    -0.758   0.448       -0.097       0.043
          Юг и С.Кавказ    |      0.031      0.039     0.797   0.425       -0.046       0.108
             Уральский     |      0.024      0.036     0.674   0.500       -0.046       0.095
              'Сибирский   |     -0.139      0.036    -3.836   0.000       -0.209      -0.068
         Дальневосточный   |     -0.092      0.050    -1.836   0.066       -0.190       0.006
                           |
                     year  |
                     1995  |     -0.134      0.027    -4.933   0.000       -0.187      -0.081
                     1996  |     -0.235      0.029    -7.991   0.000       -0.293      -0.177
                     1998  |     -0.306      0.029   -10.621   0.000       -0.363      -0.250
                     2000  |     -0.459      0.031   -15.051   0.000       -0.519      -0.399
                     2001  |     -0.371      0.031   -12.090   0.000       -0.431      -0.311
                     2002  |     -0.360      0.031   -11.598   0.000       -0.421      -0.299
                     2003  |     -0.350      0.033   -10.549   0.000       -0.416      -0.285
                     2004  |     -0.344      0.034   -10.097   0.000       -0.411      -0.277
                     2005  |     -0.348      0.036    -9.585   0.000       -0.419      -0.277
                     2006  |     -0.388      0.038   -10.193   0.000       -0.463      -0.314
                     2007  |     -0.385      0.041    -9.440   0.000       -0.465      -0.305
                           |
                    _cons  |     -4.321      0.299   -14.465   0.000       -4.907      -3.736
---------------------------+--------------------------------------------------------------
M2_male_lnvar              |
                    _cons  |     -0.410      0.018   -22.378   0.000       -0.446      -0.374
---------------------------+--------------------------------------------------------------
M3_female_mean             |
                  diplom_k |
     законченное среднее   |      0.178      0.030     6.022   0.000        0.120       0.236
среднее профессиональное   |      0.346      0.030    11.512   0.000        0.287       0.405
                  высшее   |      0.661      0.031    21.228   0.000        0.600       0.722
                           |
                   age_10  |      0.455      0.040    11.327   0.000        0.376       0.533
                           |
        c.age_10#c.age_10  |     -0.058      0.005   -11.694   0.000       -0.068      -0.048
                           |
                     male  |      0.000  (omitted)
               ln_regwage  |      0.692      0.030    23.080   0.000        0.633       0.751
                  village  |     -0.273      0.022   -12.389   0.000       -0.316      -0.230
                           |
                  fed_okr  |
                Северный   |     -0.025      0.038    -0.646   0.518       -0.099       0.050
             Центральный   |     -0.108      0.032    -3.416   0.001       -0.170      -0.046
             Приволжский   |     -0.224      0.032    -7.073   0.000       -0.287      -0.162
          Юг и С.Кавказ    |     -0.202      0.037    -5.444   0.000       -0.274      -0.129
             Уральский     |     -0.114      0.035    -3.235   0.001       -0.183      -0.045
              'Сибирский   |     -0.200      0.032    -6.321   0.000       -0.262      -0.138
         Дальневосточный   |     -0.181      0.045    -4.038   0.000       -0.269      -0.093
                           |
```

```
                    year |
                    1995 |     -0.101      0.023   -4.319   0.000       -0.146      -0.055
                    1996 |     -0.162      0.025   -6.444   0.000       -0.211      -0.113
                    1998 |     -0.284      0.027  -10.644   0.000       -0.336      -0.232
                    2000 |     -0.453      0.027  -16.761   0.000       -0.506      -0.400
                    2001 |     -0.385      0.027  -14.093   0.000       -0.439      -0.332
                    2002 |     -0.296      0.028  -10.465   0.000       -0.351      -0.240
                    2003 |     -0.312      0.030  -10.482   0.000       -0.370      -0.254
                    2004 |     -0.280      0.031   -9.079   0.000       -0.341      -0.220
                    2005 |     -0.279      0.033   -8.341   0.000       -0.344      -0.213
                    2006 |     -0.230      0.035   -6.590   0.000       -0.298      -0.161
                    2007 |     -0.234      0.038   -6.166   0.000       -0.309      -0.160
                         |
                   _cons |     -3.501      0.278  -12.616   0.000       -4.045      -2.957
-------------------------+------------------------------------------------------------------
M3_female_lnvar          |
                   _cons |     -0.565      0.017  -33.531   0.000       -0.598      -0.532
------------------------------------------------------------------------------------------
```

Тест на равенство коэффициентов регрессии в двух выборках называют тестом Чоу. Нулевая гипотеза проверяется с помощью E-статистики для гипотезы о том, что коэффициенты при всех добавленных переменных равны нулю. Тест Чоу (Чжоу, англ. Chow test) — применяемая в эконометрике процедура проверки стабильности параметров регрессионной модели, наличия структурных сдвигов в выборке. Фактически тест проверяет неоднородность выборки в контексте регрессионной модели. Истинные значения параметров модели могут теоретически различаться для разных выборок, так как выборки могут быть неоднородны.

test [M2_male_mean = M3_female_mean]

```
 ( 1)  [M2_male_mean]1b.diplom_k - [M3_female_mean]1b.diplom_k = 0
 ( 2)  [M2_male_mean]2.diplom_k - [M3_female_mean]2.diplom_k = 0
 ( 3)  [M2_male_mean]3.diplom_k - [M3_female_mean]3.diplom_k = 0
 ( 4)  [M2_male_mean]4.diplom_k - [M3_female_mean]4.diplom_k = 0
 ( 5)  [M2_male_mean]age_10 - [M3_female_mean]age_10 = 0
 ( 6)  [M2_male_mean]c.age_10#c.age_10 - [M3_female_mean]c.age_10#c.age_10 = 0
 ( 7)  [M2_male_mean]o.male - [M3_female_mean]o.male = 0
 ( 8)  [M2_male_mean]ln_regwage - [M3_female_mean]ln_regwage = 0
 ( 9)  [M2_male_mean]village - [M3_female_mean]village = 0
 (10)  [M2_male_mean]1b.fed_okr - [M3_female_mean]1b.fed_okr = 0
 (11)  [M2_male_mean]2.fed_okr - [M3_female_mean]2.fed_okr = 0
 (12)  [M2_male_mean]3.fed_okr - [M3_female_mean]3.fed_okr = 0
 (13)  [M2_male_mean]4.fed_okr - [M3_female_mean]4.fed_okr = 0
 (14)  [M2_male_mean]5.fed_okr - [M3_female_mean]5.fed_okr = 0
 (15)  [M2_male_mean]6.fed_okr - [M3_female_mean]6.fed_okr = 0
 (16)  [M2_male_mean]7.fed_okr - [M3_female_mean]7.fed_okr = 0
 (17)  [M2_male_mean]8.fed_okr - [M3_female_mean]8.fed_okr = 0
 (18)  [M2_male_mean]1994b.year - [M3_female_mean]1994b.year = 0
 (19)  [M2_male_mean]1995.year - [M3_female_mean]1995.year = 0
 (20)  [M2_male_mean]1996.year - [M3_female_mean]1996.year = 0
 (21)  [M2_male_mean]1998.year - [M3_female_mean]1998.year = 0
 (22)  [M2_male_mean]2000.year - [M3_female_mean]2000.year = 0
 (23)  [M2_male_mean]2001.year - [M3_female_mean]2001.year = 0
 (24)  [M2_male_mean]2002.year - [M3_female_mean]2002.year = 0
 (25)  [M2_male_mean]2003.year - [M3_female_mean]2003.year = 0
 (26)  [M2_male_mean]2004.year - [M3_female_mean]2004.year = 0
 (27)  [M2_male_mean]2005.year - [M3_female_mean]2005.year = 0
 (28)  [M2_male_mean]2006.year - [M3_female_mean]2006.year = 0
 (29)  [M2_male_mean]2007.year - [M3_female_mean]2007.year = 0
       Constraint 1 dropped
       Constraint 7 dropped
       Constraint 10 dropped
       Constraint 18 dropped

           chi2( 25) =  127.2
         Prob > chi2 =    0.0000
```

Мы можем отвергнуть равенство общих коэффициентов при m2 и m3.
Проверим теперь равенство отдельных коэффициентов для мужчин и женщин.

test [M2_male_mean]2.diplom_k = [M3_female_mean]2.diplom_k

```
 ( 1)  [M2_male_mean]2.diplom_k - [M3_female_mean]2.diplom_k = 0
```

```
           chi2(  1) =    0.50
         Prob > chi2 =    0.4798
```

test [M2_male_mean]3.diplom_k = [M3_female_mean]3.diplom_k

```
 ( 1)  [M2_male_mean]3.diplom_k - [M3_female_mean]3.diplom_k = 0
           chi2(  1) =    1.25
         Prob > chi2 =    0.2629
```

test [M2_male_mean]4.diplom_k = [M3_female_mean]4.diplom_k

```
 ( 1)  [M2_male_mean]4.diplom_k - [M3_female_mean]4.diplom_k = 0
           chi2(  1) =   18.75
         Prob > chi2 =    0.0000
```

Тесты показали, что значимо различны только коэффициенты для высшего образования (у женщин отдача от высшего образования выше, чем у мужчин).

32.15. Сохраните рабочий файл; сохраните все команды из «истории» в do-файл, и удалите команды из истории.

33. Модель факторов занятости (Probit).
Stata Annotated Output Probit Regression
https://stats.oarc.ucla.edu/stata/output/probit-regression/

Probit аппроксимирует пробит-модель для бинарной зависимой переменной, предполагая, что вероятность положительного исхода определяется стандартной функцией нормального кумулятивного распределения. Probit может вычислять устойчивые и устойчивые к кластерам стандартные ошибки и корректировать результаты для сложных планов опроса.
Модель, которую мы хотим оценить,
$\Pr(y = 1) = \mathbf{\Phi}(\beta_0 + \beta_1 * x_1 + \beta_2 * x_2 + \ldots..)$
где **Φ** - функция нормального кумулятивного распределения.
Stata интерпретирует значение 0 как отрицательный результат (неудача) и рассматривает все остальные значения (кроме пропущенных) как положительные результаты (успехи). Таким образом, если ваша зависимая переменная принимает значения 0 и 1, то 0 интерпретируется как неудача, а 1 — как успех. Если ваша зависимая переменная принимает значения 0, 1 и 2, то 0 по-прежнему интерпретируется как сбой, но и 1, и 2 рассматриваются как успехи.
Если вы предпочитаете более формальное математическое утверждение, когда вы вводите **probit y x**, Stata соответствует модели

$$\Pr(y_j \neq 0 \mid \mathbf{x}_j) = \Phi(\mathbf{x}_j\beta)$$

где **Φ** - функция нормального кумулятивного распределения.

Для сравнения логистическая модель (logit):
$\Pr(y = 1) = \mathbf{F}(\beta_0 + \beta_1 * x_1 + \beta_2 * x_2 + \ldots..)$
где $F(z) = e^z / (1 + e^z)$ функция логистического кумулятивного распределения

$$\Pr(y_j \neq 0 \mid \mathbf{x}_j) = \frac{\exp(\mathbf{x}_j\beta)}{1 + \exp(\mathbf{x}_j\beta)}$$

Если вы это еще не сделали, очистите все команды из истории, которые вы перенесли в do-файл.

33.1. Сначала приклеим к нашему файлу дополнительные переменные о собственных детях индивида.

merge m:1 id_w idind using "C:\RLMS_work\seminar_7\data\children_5_29.dta"
drop if _merge ==2
drop _merge

```
(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)

    Result                           # of obs.
    -----------------------------------------
    not matched                       280,774
        from master                         0  (_merge==1)
        from using                    280,774  (_merge==2)

    matched                           120,436  (_merge==3)
    -----------------------------------------
```

33.2. Сохраните файл данных. Удалите созданные команды. Мы создадим do-файл и запустим его еще раз в конце занятия.
Для этого создадим команду «открыть файл» (хотя он у нас уже открыт), и повторим команду отмены построчного выведения команд

set more off
use "C:\RLMS_work\seminar_7\data\ind_5_16_Sem7.dta", clear

*33.3. Оценим пробит-регрессию.

probit employed i.diplom_k c.age_10##c.age_10 male married Nind_child1 Nind_child2 Nind_child5 Nind_child17 NUM_adult18 lg_nolab_income lg_S_income lg_Other_income village ln_regwage regunempl i.fed_okr i.id_w if origsm ==1, cluster(idind)

```
Iteration 0:   log pseudolikelihood = -63085.297
Iteration 1:   log pseudolikelihood = -38745.709
Iteration 2:   log pseudolikelihood = -37729.194
Iteration 3:   log pseudolikelihood = -37711.574
Iteration 4:   log pseudolikelihood = -37711.558
Iteration 5:   log pseudolikelihood = -37711.558

Probit regression                                 Number of obs     =     91,538
                                                  Wald chi2(36)     =   10856.82
                                                  Prob > chi2       =     0.0000
Log pseudolikelihood = -37711.558                 Pseudo R2         =     0.4022

                                      (Std. Err. adjusted for 22,436 clusters in idind)
------------------------------------------------------------------------------------------
                         |               Robust
                employed |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------------------+----------------------------------------------------------------
                diplom_k |
     законченное среднее |      0.335      0.022   15.257   0.000        0.292       0.378
среднее профессиональное |      0.510      0.026   19.874   0.000        0.460       0.560
                  высшее |      0.824      0.032   25.761   0.000        0.761       0.886
                         |
                  age_10 |      2.029      0.040   50.315   0.000        1.950       2.108
                         |
       c.age_10#c.age_10 |     -0.242      0.005  -48.371   0.000       -0.252      -0.232
                         |
                    male |      0.255      0.019   13.735   0.000        0.219       0.292
                 married |      0.054      0.027    2.002   0.045        0.001       0.106
             Nind_child1 |      0.085      0.042    2.007   0.045        0.002       0.167
             Nind_child2 |      0.200      0.032    6.237   0.000        0.137       0.263
             Nind_child5 |      0.125      0.027    4.579   0.000        0.072       0.179
            Nind_child17 |      0.054      0.017    3.264   0.001        0.022       0.087
             NUM_adult18 |      0.012      0.008    1.533   0.125       -0.003       0.028
          lg_nolab_income |     -0.108      0.002  -46.767   0.000       -0.113      -0.104
              lg_S_income |     -0.003      0.003   -1.224   0.221       -0.009       0.002
```

```
      lg_Other_income |     -0.019      0.002   -9.131   0.000       -0.024      -0.015
              village |     -0.191      0.022   -8.590   0.000       -0.234      -0.147
            ln_regwage |      0.280      0.038    7.425   0.000        0.206       0.353
             regunempl |     -0.027      0.003   -8.880   0.000       -0.033      -0.021
                       |
               fed_okr |
             Северный  |      0.244      0.048    5.035   0.000        0.149       0.338
          Центральный  |      0.203      0.041    4.976   0.000        0.123       0.283
          Приволжский  |      0.213      0.041    5.255   0.000        0.134       0.292
        Юг и С.Кавказ  |      0.152      0.046    3.279   0.001        0.061       0.242
           Уральский   |      0.139      0.044    3.161   0.002        0.053       0.225
           'Сибирский  |      0.161      0.041    3.936   0.000        0.081       0.241
      Дальневосточный  |      0.188      0.050    3.749   0.000        0.090       0.286
                       |
                 id_w  |
              1995 год |     -0.012      0.020   -0.574   0.566       -0.051       0.028
              1996 год |     -0.280      0.025  -11.303   0.000       -0.328      -0.231
              1998 год |     -0.122      0.029   -4.190   0.000       -0.179      -0.065
              2000 год |     -0.123      0.026   -4.684   0.000       -0.174      -0.072
              2001 год |     -0.171      0.027   -6.235   0.000       -0.225      -0.117
              2002 год |     -0.263      0.029   -8.915   0.000       -0.321      -0.205
              2003 год |     -0.228      0.033   -6.981   0.000       -0.292      -0.164
              2004 год |     -0.252      0.035   -7.223   0.000       -0.320      -0.184
              2005 год |     -0.315      0.038   -8.212   0.000       -0.390      -0.240
              2006 год |     -0.270      0.041   -6.594   0.000       -0.350      -0.190
              2007 год |     -0.318      0.045   -7.129   0.000       -0.406      -0.231
                       |
                 _cons |     -5.516      0.341  -16.164   0.000       -6.185      -4.847
------------------------------------------------------------------------------------------
Note: 117 failures and 0 successes completely determined.
```

Примечание: полностью детерминированы (предсказаны) 117 неудач и 0 успехов.

probit (and logit, logistic, and ivprobit) иногда может не достичь сходимости, тогда появится сообщение типа:

```
Note: 4 failures and 0 successes completely determined.
```

Причина этого сообщения и что делать, если вы его видите, описаны в [R]logit

Есть две причины появления подобного сообщения. Первый — и самый маловероятный — случай возникает, когда непрерывная переменная (или комбинация непрерывной переменной с другими непрерывными или дамми переменными) просто является совершенным предиктором зависимой переменной.

В выходных данных отсутствуют пропущенные стандартные ошибки. Если вы получили сообщение «полностью определено» и в выводе отсутствуют одна или несколько стандартных ошибок, см. второй случай, описанный ниже. (См. руководство по STATA, logit)

Второй случай возникает, когда все независимые являются дамми переменными или непрерывными переменными с повторяющимися значениями (например, возраст). Здесь один или несколько оценочных коэффициентов будут иметь недостающие стандартные ошибки.

Это является причиной несхождения, сообщения «полностью определено» и отсутствия стандартных ошибок. Это происходит, когда у вас есть паттерн (или паттерны) ковариации только с одним результатом, и существует коллинеарность, когда наблюдения, соответствующие этому паттерну ковариации, отбрасываются. Если это произойдет с вами, выявите причины. Во-первых, определите модель ковариации только с одним исходом.

Если успехи были полностью определены, это означает, что прогнозируемые вероятности почти равны 1. Если неудачи были полностью определены, это означает, что прогнозируемые вероятности почти равны нулю.

Интерпретация результатов и соответствие теоретическим предсказаниям.

estat summarize

```
  Estimation sample probit
                                  Number of obs        =     91,538

                                  Number of clusters   =     22,436
                                  Obs per cluster: min =          1
                                                   avg =        4.1
                                                   max =         12
```

```
------------------------------------------------------------------------------
    Variable |        Mean      Std. Dev.         Min          Max
-------------+----------------------------------------------------------------
    employed |     .5445607      .4980131           0            1
             |
    diplom_k |
    зако..   |     .3339487      .4716242           0            1
среднее п..  |     .2234373      .4165513           0            1
     высшее  |     .1605672      .3671333           0            1
             |
      age_10 |     4.432312      1.900328         1.3         10.1
             |
    c.age_10#|
    c.age_10 |      23.2566      18.00007        1.69       102.01
             |
        male |     .4281501      .4948134           0            1
     married |     .5827307      .4931108           0            1
 Nind_child1 |      .014606      .1214181           0            2
 Nind_child2 |     .0364985      .1916195           0            2
 Nind_child5 |     .0583146      .2466932           0            3
Nind_child17 |     .3274815      .6608483           0            7
 NUM_adult18 |     2.605322      1.164825           0           12
lg_nolab_i~e |     3.466374      3.849853           0     13.27822
 lg_S_income |      4.23694      4.208042           0     13.73923
lg_Other_i~e |     5.406381      4.246049           0     13.82092
     village |     .3334353      .4714431           0            1
  ln_regwage |     8.730309      .5105221    7.693017     10.11715
   regunempl |     8.878267      4.025184          .8         24.9
             |
     fed_okr |
   Северный  |     .0626625      .2423563           0            1
Центральный  |     .1854749      .3886845           0            1
Приволжский  |      .199338      .3995048           0            1
   Юг и  ..  |     .1687714       .374552           0            1
 Уральский   |     .0970089      .2959715           0            1
 'Сибирский  |     .1308309      .3372172           0            1
Дальневос~й  |     .0526448      .2233246           0            1
             |
        id_w |
   1995 год  |     .0838668      .2771895           0            1
   1996 год  |     .0810265      .2728772           0            1
   1998 год  |     .0820534      .2744475           0            1
   2000 год  |     .0788306      .2694757           0            1
   2001 год  |     .0808407      .2725919           0            1
   2002 год  |      .081955      .2742977           0            1
   2003 год  |     .0804365      .2719693           0            1
   2004 год  |     .0799886      .2712771           0            1
   2005 год  |     .0757718      .2646341           0            1
   2006 год  |      .096266      .2949573           0            1
   2007 год  |     .0923769      .2895589           0            1
------------------------------------------------------------------------------
Std. Dev. not adjusted for clustering
```

*33.4. Предельные оценки (можно сравнивать силу влияния):

margins, dydx(*)

(долго считает, надо подождать)

```
Average marginal effects                        Number of obs     =     91,538
Model VCE    : Robust

Expression   : Pr(employed), predict()
dy/dx w.r.t. : 2.diplom_k 3.diplom_k 4.diplom_k age_10 male married Nind_child1 Nind_child2
               Nind_child5 Nind_child17 NUM_adult18 lg_nolab_income lg_S_income
               lg_Other_income village ln_regwage regunempl 2.fed_okr 3.fed_okr 4.fed_okr
               5.fed_okr 6.fed_okr 7.fed_okr 8.fed_okr 6.id_w 7.id_w 8.id_w 9.id_w 10.id_w
               11.id_w 12.id_w 13.id_w 14.id_w 15.id_w 16.id_w

------------------------------------------------------------------------------
                             |            Delta-method
                             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----------------------------+------------------------------------------------------------------
                   diplom_k |
     законченное среднее     |      0.086      0.006   14.733   0.000        0.074       0.097
среднее профессиональное     |      0.129      0.007   19.249   0.000        0.116       0.142
                    высшее   |      0.202      0.008   25.436   0.000        0.186       0.217
                             |
```

```
                  age_10 |      0.018      0.001    15.865   0.000        0.016       0.020
                    male |      0.059      0.004    13.793   0.000        0.051       0.068
                 married |      0.012      0.006     2.002   0.045        0.000       0.025
             Nind_child1 |      0.020      0.010     2.007   0.045        0.000       0.039
             Nind_child2 |      0.046      0.007     6.225   0.000        0.032       0.061
             Nind_child5 |      0.029      0.006     4.573   0.000        0.017       0.042
            Nind_child17 |      0.013      0.004     3.262   0.001        0.005       0.020
            NUM_adult18 |      0.003      0.002     1.533   0.125       -0.001       0.007
         lg_nolab_income |     -0.025      0.001   -47.438   0.000       -0.026      -0.024
            lg_S_income |     -0.001      0.001    -1.224   0.221       -0.002       0.000
         lg_Other_income |     -0.005      0.000    -9.129   0.000       -0.005      -0.004
                village |     -0.044      0.005    -8.587   0.000       -0.054      -0.034
             ln_regwage |      0.065      0.009     7.421   0.000        0.048       0.082
              regunempl |     -0.006      0.001    -8.922   0.000       -0.008      -0.005
                        |
                fed_okr |
              Северный  |      0.057      0.011     5.070   0.000        0.035       0.079
            Центральный |      0.048      0.010     4.958   0.000        0.029       0.066
            Приволжский |      0.050      0.010     5.236   0.000        0.031       0.069
          Юг и С.Кавказ |      0.036      0.011     3.277   0.001        0.014       0.057
             Уральский  |      0.033      0.010     3.163   0.002        0.012       0.053
              'Сибирский |      0.038      0.010     3.930   0.000        0.019       0.057
        Дальневосточный  |      0.044      0.012     3.768   0.000        0.021       0.067
                        |
                   id_w |
               1995 год |     -0.003      0.005    -0.574   0.566       -0.011       0.006
               1996 год |     -0.064      0.006   -11.424   0.000       -0.075      -0.053
               1998 год |     -0.027      0.007    -4.166   0.000       -0.040      -0.015
               2000 год |     -0.028      0.006    -4.685   0.000       -0.039      -0.016
               2001 год |     -0.039      0.006    -6.258   0.000       -0.051      -0.027
               2002 год |     -0.060      0.007    -8.980   0.000       -0.073      -0.047
               2003 год |     -0.052      0.007    -7.009   0.000       -0.067      -0.037
               2004 год |     -0.058      0.008    -7.249   0.000       -0.073      -0.042
               2005 год |     -0.073      0.009    -8.226   0.000       -0.090      -0.055
               2006 год |     -0.062      0.009    -6.594   0.000       -0.080      -0.043
               2007 год |     -0.073      0.010    -7.114   0.000       -0.093      -0.053
------------------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.
```

33.5. Сравним результаты с логистической регрессией.

logit employed i.diplom_k c.age_10##c.age_10 male married Nind_child1 Nind_child2 Nind_child5 Nind_child17 NUM_adult18 lg_nolab_income lg_S_income lg_Other_income village ln_regwage regunempl i.fed_okr i.id_w if origsm ==1, cluster(idind)

```
Iteration 0:   log pseudolikelihood = -63085.297
Iteration 1:   log pseudolikelihood = -38587.067
Iteration 2:   log pseudolikelihood = -37551.284
Iteration 3:   log pseudolikelihood = -37510.517
Iteration 4:   log pseudolikelihood = -37510.428
Iteration 5:   log pseudolikelihood = -37510.428

Logistic regression                             Number of obs     =     91,538
                                                Wald chi2(36)     =    9522.15
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -37510.428               Pseudo R2         =     0.4054

                                              (Std. Err. adjusted for 22,436 clusters in idind)
------------------------------------------------------------------------------------------
                           |               Robust
                  employed |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------------------+--------------------------------------------------------------
                  diplom_k |
      законченное среднее  |      0.553      0.038    14.558   0.000        0.479       0.628
среднее профессиональное   |      0.864      0.045    19.325   0.000        0.777       0.952
                  высшее   |      1.441      0.058    24.834   0.000        1.327       1.555
                           |
                    age_10 |      3.621      0.071    51.114   0.000        3.482       3.760
                           |
         c.age_10#c.age_10 |     -0.435      0.009   -49.314   0.000       -0.452      -0.418
                           |
                      male |      0.441      0.033    13.378   0.000        0.376       0.506
                   married |      0.090      0.047     1.895   0.058       -0.003       0.183
               Nind_child1 |      0.125      0.074     1.695   0.090       -0.020       0.269
               Nind_child2 |      0.326      0.057     5.672   0.000        0.213       0.439
               Nind_child5 |      0.195      0.049     3.981   0.000        0.099       0.292
              Nind_child17 |      0.071      0.030     2.352   0.019        0.012       0.131
```

```
          NUM_adult18 |      0.022      0.014     1.592   0.111      -0.005       0.050
      lg_nolab_income |     -0.184      0.004   -45.678   0.000      -0.192      -0.176
          lg_S_income |     -0.005      0.005    -0.928   0.353      -0.014       0.005
      lg_Other_income |     -0.034      0.004    -8.972   0.000      -0.041      -0.027
              village |     -0.325      0.039    -8.331   0.000      -0.401      -0.248
           ln_regwage |      0.498      0.067     7.453   0.000       0.367       0.629
            regunempl |     -0.048      0.005    -9.011   0.000      -0.059      -0.038
                      |
              fed_okr |
             Северный |      0.409      0.086     4.735   0.000       0.240       0.578
         Центральный  |      0.338      0.073     4.642   0.000       0.195       0.481
         Приволжский  |      0.356      0.072     4.962   0.000       0.215       0.497
       Юг и С.Кавказ  |      0.245      0.082     3.005   0.003       0.085       0.405
          Уральский   |      0.233      0.079     2.962   0.003       0.079       0.387
          'Сибирский  |      0.266      0.072     3.669   0.000       0.124       0.408
    Дальневосточный   |      0.305      0.089     3.437   0.001       0.131       0.480
                      |
               id_w   |
            1995 год  |     -0.020      0.035    -0.574   0.566      -0.089       0.049
            1996 год  |     -0.492      0.043   -11.331   0.000      -0.577      -0.407
            1998 год  |     -0.212      0.051    -4.163   0.000      -0.312      -0.112
            2000 год  |     -0.216      0.046    -4.691   0.000      -0.306      -0.126
            2001 год  |     -0.305      0.048    -6.354   0.000      -0.399      -0.211
            2002 год  |     -0.465      0.052    -8.957   0.000      -0.567      -0.363
            2003 год  |     -0.406      0.057    -7.066   0.000      -0.519      -0.293
            2004 год  |     -0.454      0.061    -7.374   0.000      -0.574      -0.333
            2005 год  |     -0.564      0.068    -8.332   0.000      -0.697      -0.431
            2006 год  |     -0.485      0.072    -6.722   0.000      -0.627      -0.344
            2007 год  |     -0.571      0.079    -7.265   0.000      -0.726      -0.417
                      |
               _cons  |     -9.774      0.605   -16.155   0.000     -10.960      -8.588
------------------------------------------------------------------------------
```

Как правило, результаты похожи, хотя в данном случае есть некоторые различия по значимости.

33.6. Выведем экспоненту.

logit employed i.diplom_k c.age_10##c.age_10 male married Nind_child1 Nind_child2 Nind_child5 Nind_child17 NUM_adult18 lg_nolab_income lg_S_income lg_Other_income village ln_regwage regunempl i.fed_okr i.id_w if origsm ==1, or cluster(idind)

```
Iteration 0:   log pseudolikelihood = -63085.297
Iteration 1:   log pseudolikelihood = -38587.067
Iteration 2:   log pseudolikelihood = -37551.284
Iteration 3:   log pseudolikelihood = -37510.517
Iteration 4:   log pseudolikelihood = -37510.428
Iteration 5:   log pseudolikelihood = -37510.428

Logistic regression                             Number of obs     =     91,538
                                                Wald chi2(36)     =    9522.15
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -37510.428               Pseudo R2         =     0.4054

                                    (Std. Err. adjusted for 22,436 clusters in idind)
------------------------------------------------------------------------------
                             |               Robust
                    employed | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----------------------------+------------------------------------------------------------------
                    diplom_k |
         законченное среднее |      1.739      0.066    14.558   0.000       1.614       1.873
среднее профессиональное     |      2.373      0.106    19.325   0.000       2.174       2.591
                     высшее  |      4.224      0.245    24.834   0.000       3.770       4.733
                             |
                      age_10 |     37.384      2.649    51.114   0.000      32.537      42.953
                             |
             c.age_10#c.age_10 |      0.647      0.006   -49.314   0.000       0.636       0.659
                             |
                        male |      1.554      0.051    13.378   0.000       1.457       1.658
                     married |      1.094      0.052     1.895   0.058       0.997       1.201
                 Nind_child1 |      1.133      0.083     1.695   0.090       0.981       1.309
                 Nind_child2 |      1.386      0.080     5.672   0.000       1.238       1.551
                 Nind_child5 |      1.216      0.060     3.981   0.000       1.104       1.339
                Nind_child17 |      1.074      0.033     2.352   0.019       1.012       1.140
                 NUM_adult18 |      1.023      0.014     1.592   0.111       0.995       1.051
```

```
------------------------------------------------------------------------------
lg_nolab_income |      0.832      0.003   -45.678   0.000        0.825       0.838
    lg_S_income |      0.995      0.005    -0.928   0.353        0.986       1.005
lg_Other_income |      0.967      0.004    -8.972   0.000        0.960       0.974
        village |      0.723      0.028    -8.331   0.000        0.669       0.780
     ln_regwage |      1.645      0.110     7.453   0.000        1.443       1.875
      regunempl |      0.953      0.005    -9.011   0.000        0.943       0.963
                |
        fed_okr |
       Северный |      1.505      0.130     4.735   0.000        1.271       1.782
  Центральный   |      1.402      0.102     4.642   0.000        1.216       1.618
  Приволжский   |      1.428      0.102     4.962   0.000        1.240       1.643
 Юг и С.Кавказ  |      1.278      0.104     3.005   0.003        1.089       1.500
    Уральский   |      1.262      0.099     2.962   0.003        1.082       1.473
    'Сибирский  |      1.304      0.094     3.669   0.000        1.132       1.503
Дальневосточный |      1.357      0.121     3.437   0.001        1.140       1.615
                |
          id_w  |
     1995 год   |      0.980      0.034    -0.574   0.566        0.915       1.050
     1996 год   |      0.611      0.027   -11.331   0.000        0.562       0.666
     1998 год   |      0.809      0.041    -4.163   0.000        0.732       0.894
     2000 год   |      0.806      0.037    -4.691   0.000        0.736       0.882
     2001 год   |      0.737      0.035    -6.354   0.000        0.671       0.810
     2002 год   |      0.628      0.033    -8.957   0.000        0.567       0.695
     2003 год   |      0.666      0.038    -7.066   0.000        0.595       0.746
     2004 год   |      0.635      0.039    -7.374   0.000        0.563       0.717
     2005 год   |      0.569      0.039    -8.332   0.000        0.498       0.650
     2006 год   |      0.616      0.044    -6.722   0.000        0.534       0.709
     2007 год   |      0.565      0.044    -7.265   0.000        0.484       0.659
                |
          _cons |      0.000      0.000   -16.155   0.000        0.000       0.000
------------------------------------------------------------------------------
```

Интерпретация Odds Ratio (коэффициент шансов, или экспонента).
Шансы – это отношение вероятности того, что событие произойдёт, к вероятности того, что событие не произойдёт (измеряется от 0 до бесконечности); вероятность измеряется от 0 до 1. Отношение, или коэффициент шансов (OR) является мерой связи между воздействием (exposure) и результатом (outcome). OR представляет вероятность того, что исход произойдет при определенном воздействии, по сравнению с шансами того, что исход произойдет в отсутствие этого воздействия.
При расчете логистической регрессии коэффициент регрессии (b1) представляет собой предполагаемое увеличение логарифмических шансов результата, или зависимой переменной (outcome) на единицу увеличения значения воздействия. Экспоненциальная функция коэффициента регрессии (eb1) представляет собой отношение шансов, связанное с увеличением воздействия (exposure), или в данном случае независимой переменной, на одну единицу.
В данной модели – шансы быть занятым для людей с высшим образованием в 4,2 раза выше шансов людей, которые не имеют полного среднего образования, то есть сравнение делается с базовой категорией. В 2007 году шансы быть занятым были в 0.578 выше (то есть в реальности на 42,2% ниже), чем в базовом 1994 году. При увеличении логарифма своих нетрудовых доходов на 1, шансы быть занятым падают на 16,9% (OR = 0.831).

*33.7. Сохраните все команды в do-файл, и сохраните его под именем…
Сохраните все выполненные команды из окна COMMAND в виде DO-файла. Удалите неудачные команды из левого окна историй команд (выделены красным). Выделите все команды (всего 7 команд) , нажмите правую кнопку мыши и выберите опцию "послать в DO-файл", после чего откроется окно редактора файла синтаксиса, и затем сохраните DO-файл с именем "**do_probit.do**". Удалите из этого файла команду **margins, dydx(*)** (так как она очень долго выполняется).

*33.8. Вывод в таблицу odds ratio (эту команду не сохраняем в DO-файл, так как если мы укажем имя уже существующего файла для вывода, будет ошибка).

outreg2 using C:\RLMS_work\seminar_7\data\reg_employed, label eform cti(odds ratio) seeout

v1	v2
VARIABLES	odds ratio
законченное образование = 2, законченное среднее	1.739***
	(0.0661)
законченное образование = 3, среднее профессиональное	2.373***
	(0.106)
законченное образование = 4, высшее	4.224***
	(0.245)
возраст \10	37.38***
	(2.649)
Возраст\10 в квадрате	0.647***
	(0.00571)
мужской пол	1.554***
	(0.0512)
состоит в браке, включая неформальный	1.094*
	(0.0519)
Количество своих детей до 1 года в д\х	1.133*
	(0.0834)
Количество своих детей с 1 до 2-х лет в д\х	1.386***
	(0.0797)
Количество своих детей с 3 до 5-ти лет в д\х	1.216***
	(0.0597)
Количество своих детей с 6 до 17-ти лет в д\х	1.074**
	(0.0326)
кол-во взрослых 18+ в семье	1.023
	(0.0144)
лог дефл нетрудового дохода	0.832***
	(0.00336)
лог дефл.доходов супруга\супруги	0.995
	(0.00488)
лог дефл доходов остальных членов семьи	0.967***
	(0.00365)
село	0.723***
	(0.0282)
логарифм дефлир региональной ЗП	1.645***
	(0.110)
Уровень безработицы населения 15 лет и старше, в среднем за год,% (до	0.953***
	(0.00509)
федеральный округ = 2, Северный	1.505***
	(0.130)
федеральный округ = 3, Центральный	1.402***
	(0.102)
федеральный округ = 4, Приволжский	1.428***
	(0.102)
федеральный округ = 5, Юг и С.Кавказ	1.278***
	(0.104)
федеральный округ = 6, Уральский	1.262***
	(0.0993)
федеральный округ = 7, 'Сибирский	1.304***
	(0.0945)
федеральный округ = 8, Дальневосточный	1.357***
	(0.121)
1995 (базовая категория – 1994)	0.980
	(0.0345)
1996	0.611***

	(0.0265)
1998	0.809***
	(0.0412)
2000	0.806***
	(0.0371)
2001	0.737***
	(0.0354)
2002	0.628***
	(0.0326)
2003	0.666***
	(0.0383)
2004	0.635***
	(0.0391)
2005	0.569***
	(0.0385)
2006	0.616***
	(0.0444)
2007	0.565***
	(0.0444)
Constant	5.69e-05***
	(3.44e-05)
Observations	91,538

34. Модель Хекмана для уравнения Минцера.
34.1. Эта модель применяется, если мы считаем, что есть модель с отбором: сначала человек выбирает, работать или нет (модель отбора), а затем для тех, кто работает, выполняется зависимость заработной платы (мы оцениваем логарифм ставки ЗП) от человеческого капитала, то есть от образования и возраста. Если мы оцениваем только модель для заработной платы, она оказывается смещенной, так как неработающие имеют потенциально более низкую ставку заработной платы.

heckman подходит для оценки моделей регрессии с выбором, используя либо двухэтапную непротиворечивую оценку Хекмана, либо оценку с полным максимальным правдоподобием (используем сначала второй вариант). В модели отбора используется оценка пробит.

heckman lg_Hwage1 i.diplom_k c.age_10##c.age_10 male ln_regwage village i.fed_okr i.year if origsm ==1, select (employed = i.diplom_k c.age_10##c.age_10 male married Nind_child1 Nind_child2 Nind_child5 Nind_child17 NUM_adult18 lg_nolab_income lg_S_income lg_Other_income village ln_regwage regunempl i.fed_okr i.id_w) cluster(idind)
estat summarize

```
Iteration 0:   log pseudolikelihood = -75081.334
Iteration 1:   log pseudolikelihood = -75077.805
Iteration 2:   log pseudolikelihood = -75077.788
Iteration 3:   log pseudolikelihood = -75077.788

Heckman selection model                         Number of obs     =     79,508
(regression model with sample selection)        Censored obs      =     41,690
                                                Uncensored obs    =     37,818

                                                Wald chi2(26)     =    7957.28
Log pseudolikelihood = -75077.79                Prob > chi2       =     0.0000

                                       (Std. Err. adjusted for 21,205 clusters in idind)
------------------------------------------------------------------------------------------
                         |               Robust
                         |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------------------+----------------------------------------------------------------
lg_Hwage1                |
                diplom_k |
     законченное среднее |      0.133      0.020     6.657   0.000        0.094       0.172
```

среднее профессиональное	0.280	0.022	12.761	0.000	0.237	0.322
высшее	0.525	0.023	22.442	0.000	0.479	0.570
age_10	0.260	0.041	6.310	0.000	0.179	0.341
c.age_10#c.age_10	-0.035	0.005	-6.759	0.000	-0.045	-0.025
male	0.309	0.013	23.504	0.000	0.284	0.335
ln_regwage	0.730	0.026	28.449	0.000	0.680	0.781
village	-0.369	0.017	-21.672	0.000	-0.402	-0.335
fed_okr						
Северный	0.041	0.030	1.375	0.169	-0.017	0.100
Центральный	-0.042	0.025	-1.678	0.093	-0.091	0.007
Приволжский	-0.144	0.025	-5.662	0.000	-0.194	-0.094
Юг и С.Кавказ	-0.096	0.029	-3.360	0.001	-0.152	-0.040
Уральский	-0.058	0.028	-2.064	0.039	-0.114	-0.003
'Сибирский	-0.169	0.025	-6.770	0.000	-0.217	-0.120
Дальневосточный	-0.154	0.035	-4.452	0.000	-0.222	-0.086
year						
1995	-0.121	0.019	-6.442	0.000	-0.158	-0.084
1996	-0.188	0.021	-9.163	0.000	-0.229	-0.148
1998	-0.283	0.020	-13.949	0.000	-0.323	-0.243
2000	-0.446	0.021	-21.031	0.000	-0.487	-0.404
2001	-0.359	0.022	-16.514	0.000	-0.401	-0.316
2002	-0.312	0.023	-13.821	0.000	-0.357	-0.268
2003	-0.316	0.024	-12.996	0.000	-0.364	-0.269
2004	-0.292	0.025	-11.478	0.000	-0.341	-0.242
2005	-0.289	0.028	-10.445	0.000	-0.343	-0.235
2006	-0.280	0.029	-9.557	0.000	-0.338	-0.223
2007	-0.283	0.032	-8.864	0.000	-0.346	-0.220
_cons	-3.359	0.255	-13.163	0.000	-3.860	-2.859
employed						
diplom_k						
законченное среднее	0.356	0.026	13.894	0.000	0.306	0.406
среднее профессиональное	0.587	0.030	19.768	0.000	0.529	0.646
высшее	0.925	0.036	25.556	0.000	0.854	0.996
age_10	2.239	0.045	49.949	0.000	2.151	2.326
c.age_10#c.age_10	-0.264	0.006	-47.380	0.000	-0.274	-0.253
male	0.243	0.021	11.423	0.000	0.202	0.285
married	0.098	0.030	3.224	0.001	0.038	0.158
Nind_child1	-0.331	0.050	-6.601	0.000	-0.429	-0.233
Nind_child2	-0.006	0.036	-0.164	0.870	-0.077	0.065
Nind_child5	0.131	0.030	4.330	0.000	0.072	0.190
Nind_child17	0.039	0.018	2.150	0.032	0.003	0.074
NUM_adult18	0.001	0.009	0.145	0.885	-0.017	0.020
lg_nolab_income	-0.128	0.003	-48.236	0.000	-0.133	-0.123
lg_S_income	-0.005	0.003	-1.500	0.134	-0.011	0.001
lg_Other_income	-0.020	0.002	-8.491	0.000	-0.025	-0.016
village	-0.248	0.025	-9.728	0.000	-0.297	-0.198
ln_regwage	0.296	0.043	6.922	0.000	0.212	0.379
regunempl	-0.046	0.004	-12.383	0.000	-0.053	-0.039
fed_okr						
Северный	0.366	0.055	6.659	0.000	0.258	0.473
Центральный	0.279	0.046	6.031	0.000	0.188	0.369
Приволжский	0.327	0.046	7.085	0.000	0.237	0.417
Юг и С.Кавказ	0.223	0.053	4.209	0.000	0.119	0.326
Уральский	0.197	0.051	3.895	0.000	0.098	0.296
'Сибирский	0.244	0.047	5.194	0.000	0.152	0.336
Дальневосточный	0.235	0.058	4.073	0.000	0.122	0.348
id_w						
1995 год	-0.026	0.022	-1.166	0.244	-0.070	0.018
1996 год	-0.317	0.027	-11.544	0.000	-0.371	-0.263
1998 год	-0.090	0.033	-2.758	0.006	-0.154	-0.026
2000 год	-0.177	0.029	-6.108	0.000	-0.233	-0.120
2001 год	-0.256	0.030	-8.519	0.000	-0.314	-0.197
2002 год	-0.326	0.032	-10.038	0.000	-0.390	-0.262
2003 год	-0.270	0.036	-7.407	0.000	-0.341	-0.198
2004 год	-0.302	0.039	-7.818	0.000	-0.378	-0.226
2005 год	-0.344	0.043	-8.076	0.000	-0.428	-0.261
2006 год	-0.320	0.046	-6.983	0.000	-0.410	-0.230
2007 год	-0.383	0.050	-7.635	0.000	-0.481	-0.284

```
                       _cons |     -6.099      0.387  -15.759   0.000       -6.857      -5.340
-------------------------+------------------------------------------------------------------
                    /athrho |     -0.159      0.025   -6.401   0.000       -0.208      -0.110
                   /lnsigma |     -0.237      0.007  -35.752   0.000       -0.250      -0.224
-------------------------+------------------------------------------------------------------
                        rho |     -0.158      0.024                       -0.205      -0.110
                      sigma |      0.789      0.005                        0.779       0.800
                     lambda |     -0.124      0.019                       -0.162      -0.087
------------------------------------------------------------------------------------------
Wald test of indep. eqns. (rho = 0): chi2(1) =    40.98   Prob > chi2 = 0.0000
```

Тест отношения правдоподобия, указанный в нижней части выходных данных, является эквивалентным тестом для ρ = 0 (rho = 0) и в вычислительном отношении представляет собой сравнение совместной вероятности независимой пробит-модели для уравнения выбора и регрессионной модели на основе наблюдаемых данных. данные о заработной плате против вероятности модели Хекмана. Поскольку $\chi^2 = 45{,}78$, это явно оправдывает уравнение отбора Хекмана с этими данными.

`heckman` assumes that `wage` is the dependent variable and that the first variable list (`educ` and `age`) are the determinants of `wage`. The variables specified in the `select()` option (`married`, `children`, `educ`, and `age`) are assumed to determine whether the dependent variable is observed (the selection equation). Thus, we fit the model

$$\texttt{wage} = \beta_0 + \beta_1\texttt{educ} + \beta_2\texttt{age} + u_1$$

and we assumed that `wage` is observed if

$$\gamma_0 + \gamma_1\texttt{married} + \gamma_2\texttt{children} + \gamma_3\texttt{educ} + \gamma_4\texttt{age} + u_2 > 0$$

where u_1 and u_2 have correlation ρ.

The reported results for the wage equation are interpreted exactly as though we observed wage data for all women in the sample; the coefficients on age and education level represent the estimated marginal effects of the regressors in the underlying regression equation. The results for the two ancillary parameters require some explanation. `heckman` does not directly estimate ρ; to constrain ρ within its valid limits, and for numerical stability during optimization, it estimates the inverse hyperbolic tangent of ρ:

$$\operatorname{atanh}\rho = \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right)$$

This estimate is reported as `/athrho`. In the bottom panel of the output, `heckman` undoes this transformation for you: the estimated value of ρ is 0.7035061. The standard error for ρ is computed using the delta method, and its confidence intervals are the transformed intervals of `/athrho`.

Similarly, σ, the standard error of the residual in the wage equation, is not directly estimated; for numerical stability, `heckman` instead estimates $\ln\sigma$. The untransformed `sigma` is reported at the end of the output: 6.004797.

Finally, some researchers—especially economists—are used to the selectivity effect summarized not by ρ but by $\lambda = \rho\sigma$. `heckman` reports this, too, along with an estimate of the standard error and confidence interval.

34.2. Для двухшаговой процедуры Хекмана (классической, с sample selection) нельзя использовать робастные оценки и взвешивание. Результаты несколько отличаются.

heckman lg_Hwage1 i.diplom_k c.age_10##c.age_10 male ln_regwage village i.fed_okr i.year if origsm ==1, select (employed = i.diplom_k c.age_10##c.age_10 male married Nind_child1 Nind_child2 Nind_child5 Nind_child17 NUM_adult18 lg_nolab_income lg_S_income lg_Other_income village ln_regwage regunempl i.fed_okr i.id_w) twostep
estat summarize

Heckman selection model -- two-step estimates
(regression model with sample selection)

Number of obs	79,508
Censored obs	41,690
Uncensored obs	37,818
Wald chi2(26)	13322.81
Prob > chi2	0.0000

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
lg_Hwage1						
diplom_k						
законченное среднее	0.125	0.015	8.427	0.000	0.096	0.154
среднее профессиональное	0.268	0.016	16.591	0.000	0.236	0.299
высшее	0.508	0.017	29.210	0.000	0.474	0.542
age_10	0.195	0.038	5.147	0.000	0.121	0.270
c.age_10#c.age_10	-0.027	0.005	-5.660	0.000	-0.036	-0.017
male	0.303	0.009	33.775	0.000	0.285	0.320
ln_regwage	0.723	0.018	40.227	0.000	0.688	0.758
village	-0.365	0.011	-34.387	0.000	-0.386	-0.345
fed_okr						
Северный	0.040	0.020	2.023	0.043	0.001	0.079
Центральный	-0.046	0.018	-2.588	0.010	-0.081	-0.011
Приволжский	-0.149	0.018	-8.110	0.000	-0.184	-0.113
Юг и С.Кавказ	-0.097	0.020	-4.758	0.000	-0.137	-0.057
Уральский	-0.060	0.020	-3.033	0.002	-0.099	-0.021
'Сибирский	-0.170	0.018	-9.566	0.000	-0.205	-0.135
Дальневосточный	-0.154	0.023	-6.832	0.000	-0.199	-0.110
year						
1995	-0.120	0.019	-6.380	0.000	-0.158	-0.083
1996	-0.183	0.020	-9.151	0.000	-0.222	-0.144
1998	-0.280	0.020	-14.065	0.000	-0.318	-0.241
2000	-0.441	0.020	-21.950	0.000	-0.481	-0.402
2001	-0.353	0.021	-17.228	0.000	-0.394	-0.313
2002	-0.307	0.021	-14.554	0.000	-0.348	-0.265
2003	-0.310	0.022	-14.142	0.000	-0.353	-0.267
2004	-0.286	0.023	-12.675	0.000	-0.330	-0.241
2005	-0.282	0.024	-11.710	0.000	-0.329	-0.235
2006	-0.273	0.024	-11.187	0.000	-0.321	-0.225
2007	-0.276	0.026	-10.577	0.000	-0.327	-0.225
_cons	-3.147	0.193	-16.271	0.000	-3.526	-2.768
employed						
diplom_k						
законченное среднее	0.357	0.016	21.672	0.000	0.324	0.389
среднее профессиональное	0.588	0.018	32.909	0.000	0.553	0.623
высшее	0.926	0.020	45.841	0.000	0.887	0.966
age_10	2.244	0.024	92.314	0.000	2.196	2.292
c.age_10#c.age_10	-0.264	0.003	-91.261	0.000	-0.270	-0.259
male	0.244	0.012	20.255	0.000	0.221	0.268
married	0.098	0.022	4.524	0.000	0.056	0.141
Nind_child1	-0.333	0.051	-6.533	0.000	-0.433	-0.233
Nind_child2	-0.013	0.032	-0.422	0.673	-0.075	0.049
Nind_child5	0.130	0.024	5.493	0.000	0.084	0.176
Nind_child17	0.039	0.010	3.891	0.000	0.020	0.059
NUM_adult18	0.001	0.007	0.080	0.936	-0.012	0.013
lg_nolab_income	-0.127	0.002	-67.708	0.000	-0.131	-0.124
lg_S_income	-0.005	0.002	-2.096	0.036	-0.010	-0.000
lg_Other_income	-0.020	0.002	-10.684	0.000	-0.023	-0.016
village	-0.250	0.015	-17.227	0.000	-0.279	-0.222
ln_regwage	0.296	0.026	11.248	0.000	0.244	0.347
regunempl	-0.046	0.002	-19.254	0.000	-0.051	-0.042
fed_okr						
Северный	0.375	0.034	11.158	0.000	0.309	0.440
Центральный	0.284	0.026	10.783	0.000	0.232	0.336
Приволжский	0.333	0.028	12.037	0.000	0.279	0.387
Юг и С.Кавказ	0.230	0.032	7.280	0.000	0.168	0.292
Уральский	0.204	0.030	6.850	0.000	0.146	0.263
'Сибирский	0.249	0.028	8.772	0.000	0.193	0.305
Дальневосточный	0.245	0.035	7.081	0.000	0.177	0.313

```
                id_w |
        1995 год  |     -0.024      0.028   -0.874   0.382       -0.079       0.030
        1996 год  |     -0.315      0.029  -10.850   0.000       -0.372      -0.258
        1998 год  |     -0.086      0.031   -2.799   0.005       -0.146      -0.026
        2000 год  |     -0.174      0.029   -5.959   0.000       -0.231      -0.117
        2001 год  |     -0.254      0.030   -8.608   0.000       -0.312      -0.196
        2002 год  |     -0.327      0.030  -10.815   0.000       -0.387      -0.268
        2003 год  |     -0.269      0.032   -8.412   0.000       -0.331      -0.206
        2004 год  |     -0.303      0.033   -9.232   0.000       -0.367      -0.239
        2005 год  |     -0.344      0.035   -9.898   0.000       -0.413      -0.276
        2006 год  |     -0.320      0.035   -9.034   0.000       -0.390      -0.251
        2007 год  |     -0.384      0.038  -10.202   0.000       -0.458      -0.310
                  |
            _cons |     -6.108      0.241  -25.364   0.000       -6.580      -5.636
------------------+------------------------------------------------------------------
mills             |
           lambda |     -0.167      0.021   -7.930   0.000       -0.208      -0.126
------------------+------------------------------------------------------------------
             rho |   -0.21082
           sigma |   .79225256
------------------------------------------------------------------------------------
```

В двухступенчатой модели сначала рассчитывается mills lambda на основании модели отбора (работает – не работает), и затем эта новая переменная входит в модель для заработной платы. В этой модели она значима, значит, коррекцию смещенности выборки имело смысл делать.

*34.3. сравним предсказанные значения логарифма ставки заработной платы в модели Хекмана для занятых и незанятых

predict xb1
tabstat xb1 if origsm ==1, statistics(mean) by(employed)

```
Summary for variables: xb1
     by categories of: employed (работает)

        employed |      mean
-----------------+----------
не работает или  |  3.062849
есть любая работ |  3.446951
-----------------+----------
           Total |  3.273945
-----------------------------
```

* И теперь повторим модель Минцера без корректировки, без робастности и без взвешивания для сравнения результатов

reg lg_Hwage1 i.diplom_k c.age_10##c.age_10 male ln_regwage village i.fed_okr i.year if origsm ==1
predict xb2
tabstat xb2 if origsm ==1, statistics(mean) by(employed)

```
Summary for variables: xb2
     by categories of: employed (работает)

        employed |      mean
-----------------+----------
не работает или  |  2.802537
есть любая работ |   3.37106
-----------------+----------
           Total |  3.114988
-----------------------------
```

Как видим, с корректировкой средние значения для незанятых выше (для занятых тоже, но отличия невелики и могту быть незначимы, нужно проверять).
Сохраните файл данных.

35. Мультиномиальная регрессия (Multinomial Logistic Regression): вероятность изменить статус занятости в будущем году, для неактивных в этом году на рынке труда **(employment ==1)**
Stata Annotated Output Multinomial Logistic Regression
https://stats.oarc.ucla.edu/stata/output/multinomial-logistic-regression/

*35.1. Создадим переменные «будущего» (это возможно только для индивидов!). Объявим данные панельными, то есть указание идентификатора кейсов (переменная **idind**) и переменной времени (номера волны **id_w**)

tsset idind id_w

*переменные «статус занятости в году T+1» и «состоит ли в браке (включая неформальный) в году T+1»
gen married_T1 = F.married
label variable married_T1 "T+1 состоит ли в браке (включая неформальный)"
gen employment_T1 = F.employment
label variable employment_T1 "T+1 занятость в прошлой волне"
label values employment_T1 EMPLOYMENT

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_t-1)
        L2.       2-period lag (x_t-2)
        ...
        F.        lead (x_t+1)
        F2.       2-period lead (x_t+2)
```

Сохраните рабочий файл.

*35.2. Оценим мультиномиальную регрессию для тех, чей статус занятости в году T = 1 (неактивные, то есть не работающие и не ищущие работу), и посмотрим, какие факторы влияют на то, что они в будущем году останутся неактивными, или перейдут в статус безработного, или в статус занятого. Базовую категорию зависимой переменной берем такую же, как в году T, то есть первую (**baseoutcome(1)**), что означает, что человек остается неактивным на рынке труда. С этой категорией будем сравнивать остальные исходы.

mlogit employment_T1 i.diplom_k c.age_10##c.age_10 male married Nind_child1 Nind_child2 Nind_child5 Nind_child17 NUM_adult18 lg_nolab_income lg_S_income lg_Other_income village ln_regwage regunempl i.fed_okr i.id_w if employment ==1 & origsm ==1, cluster(idind) baseoutcome(1)

```
Iteration 0:   log pseudolikelihood = -14185.642
Iteration 1:   log pseudolikelihood = -12059.362
Iteration 2:   log pseudolikelihood = -11379.079
Iteration 3:   log pseudolikelihood = -11272.862
Iteration 4:   log pseudolikelihood =   -11269.4
Iteration 5:   log pseudolikelihood = -11269.384
Iteration 6:   log pseudolikelihood = -11269.384
```

```
Multinomial logistic regression                 Number of obs     =     28,675
                                                Wald chi2(70)     =    2236.89
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -11269.384               Pseudo R2         =     0.2056

                                          (Std. Err. adjusted for 8,857 clusters in idind)
------------------------------------------------------------------------------------------
                             |               Robust
           employment_T1     |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----------------------------+------------------------------------------------------------
неактивный                   |  (base outcome)
-----------------------------+------------------------------------------------------------
безработный                  |
                    diplom_k |
     законченное среднее     |      0.655      0.094     6.940   0.000        0.470       0.840
среднее профессиональное     |      0.743      0.110     6.779   0.000        0.528       0.958
                    высшее   |      1.254      0.158     7.956   0.000        0.945       1.562
                             |
                     age_10  |      1.827      0.178    10.262   0.000        1.478       2.175
                             |
        c.age_10#c.age_10    |     -0.288      0.023   -12.634   0.000       -0.333      -0.243
                             |
                      male   |      0.207      0.075     2.746   0.006        0.059       0.355
                   married   |     -0.082      0.144    -0.570   0.568       -0.365       0.200
               Nind_child1   |     -0.295      0.243    -1.214   0.225       -0.772       0.181
               Nind_child2   |      0.538      0.149     3.617   0.000        0.246       0.830
               Nind_child5   |     -0.051      0.149    -0.343   0.732       -0.342       0.240
              Nind_child17   |      0.044      0.063     0.691   0.490       -0.081       0.168
               NUM_adult18   |      0.041      0.036     1.125   0.261       -0.030       0.111
          lg_nolab_income    |     -0.029      0.011    -2.667   0.008       -0.050      -0.008
              lg_S_income    |     -0.032      0.016    -2.045   0.041       -0.062      -0.001
          lg_Other_income    |     -0.030      0.012    -2.460   0.014       -0.054      -0.006
                   village   |     -0.409      0.094    -4.337   0.000       -0.593      -0.224
                ln_regwage   |      0.087      0.160     0.545   0.586       -0.226       0.400
                 regunempl   |      0.025      0.012     2.136   0.033        0.002       0.048
                             |
                    fed_okr  |
                  Северный   |      0.310      0.207     1.502   0.133       -0.095       0.715
                Центральный  |      0.012      0.184     0.066   0.948       -0.348       0.372
                Приволжский  |     -0.072      0.184    -0.394   0.693       -0.432       0.287
              Юг и С.Кавказ  |      0.020      0.205     0.096   0.924       -0.381       0.420
                 Уральский   |      0.133      0.194     0.684   0.494       -0.247       0.513
                 'Сибирский  |      0.149      0.185     0.806   0.420       -0.213       0.511
             Дальневосточный |     -0.179      0.235    -0.764   0.445       -0.640       0.281
                             |
                      id_w   |
                  1995 год   |     -0.129      0.153    -0.841   0.400       -0.429       0.171
                  1996 год   |     -0.275      0.163    -1.692   0.091       -0.593       0.044
                  1998 год   |     -0.489      0.169    -2.897   0.004       -0.820      -0.158
                  2000 год   |     -0.624      0.160    -3.899   0.000       -0.938      -0.310
                  2001 год   |     -0.743      0.164    -4.544   0.000       -1.064      -0.423
                  2002 год   |     -0.712      0.169    -4.200   0.000       -1.044      -0.380
                  2003 год   |     -0.766      0.184    -4.169   0.000       -1.126      -0.406
                  2004 год   |     -0.734      0.185    -3.968   0.000       -1.096      -0.371
                  2005 год   |     -1.067      0.209    -5.106   0.000       -1.476      -0.657
                  2006 год   |     -0.959      0.212    -4.533   0.000       -1.374      -0.545
                             |
                      _cons  |     -5.059      1.462    -3.461   0.001       -7.924      -2.194
-----------------------------+------------------------------------------------------------
есть_любая_работа            |
                    diplom_k |
     законченное среднее     |      0.646      0.059    11.040   0.000        0.531       0.761
среднее профессиональное     |      0.698      0.068    10.237   0.000        0.564       0.832
                    высшее   |      0.865      0.103     8.377   0.000        0.662       1.067
                             |
                     age_10  |      1.720      0.114    15.139   0.000        1.498       1.943
                             |
        c.age_10#c.age_10    |     -0.254      0.014   -17.939   0.000       -0.281      -0.226
                             |
                      male   |      0.412      0.049     8.457   0.000        0.317       0.508
                   married   |      0.122      0.093     1.307   0.191       -0.061       0.304
               Nind_child1   |      0.154      0.153     1.010   0.312       -0.145       0.454
               Nind_child2   |      0.424      0.114     3.709   0.000        0.200       0.649
               Nind_child5   |      0.150      0.096     1.556   0.120       -0.039       0.338
              Nind_child17   |      0.062      0.048     1.284   0.199       -0.033       0.157
               NUM_adult18   |      0.095      0.023     4.116   0.000        0.050       0.140
          lg_nolab_income    |     -0.026      0.007    -3.489   0.000       -0.040      -0.011
              lg_S_income    |     -0.032      0.010    -3.258   0.001       -0.051      -0.013
          lg_Other_income    |     -0.020      0.008    -2.642   0.008       -0.035      -0.005
                   village   |     -0.179      0.060    -2.983   0.003       -0.296      -0.061
```

```
          ln_regwage |      0.392      0.106     3.712   0.000        0.185       0.599
           regunempl |      0.002      0.008     0.200   0.841       -0.014       0.017
                     |
             fed_okr |
           Северный  |     -0.086      0.146    -0.594   0.553       -0.372       0.199
        Центральный  |     -0.033      0.116    -0.280   0.779       -0.260       0.195
        Приволжский  |      0.003      0.118     0.022   0.983       -0.229       0.234
      Юг и С.Кавказ  |      0.042      0.139     0.299   0.765       -0.231       0.314
         Уральский   |      0.063      0.129     0.487   0.627       -0.190       0.315
         'Сибирский  |      0.067      0.121     0.551   0.582       -0.170       0.303
    Дальневосточный  |     -0.034      0.147    -0.231   0.818       -0.322       0.254
                     |
               id_w  |
          1995 год   |     -0.301      0.110    -2.735   0.006       -0.516      -0.085
          1996 год   |     -0.169      0.112    -1.514   0.130       -0.389       0.050
          1998 год   |      0.184      0.108     1.700   0.089       -0.028       0.396
          2000 год   |     -0.251      0.106    -2.363   0.018       -0.459      -0.043
          2001 год   |     -0.348      0.108    -3.238   0.001       -0.559      -0.138
          2002 год   |     -0.321      0.110    -2.914   0.004       -0.537      -0.105
          2003 год   |     -0.432      0.120    -3.608   0.000       -0.666      -0.197
          2004 год   |     -0.690      0.128    -5.409   0.000       -0.940      -0.440
          2005 год   |     -0.438      0.131    -3.353   0.001       -0.694      -0.182
          2006 год   |     -0.523      0.136    -3.841   0.000       -0.790      -0.256
                     |
               _cons |     -7.189      0.961    -7.477   0.000       -9.074      -5.305
------------------------------------------------------------------------------
```

estat summarize – не работает!!!

```
conformability error
```

Интерпретация коэфициентов – смотрим на значимость. Абсолютные значения сравнивать между собой нельзя.

*35.3. Мультиноминальная регрессия с опцией **rrr**: report relative-risk ratios

Относительный риск определяется как отношение вероятностей наступления событий в одной группе к аналогичной вероятности в другой. Понятно, что если данное отношение больше единицы, это означает, что вероятность события (например, смерти) выше в одной группе, нежели в другой, иными словами, мы имеем дело с фактором риска. Если же величина меньше единицы, то мы имеем дело с протективным фактором (например, если группы определялись применением метода лечения).

Relative Risk Ratio – Это коэффициенты относительного риска для мультиномиальной логит-модели. Их можно получить, возведя в степень полиномиальные логит-коэффициенты, или указав опцию **rrr**. Напомним, что мультиномиальная логит-модель оценивает модели k-1, где k-е уравнение относится к референтной группе. Если модель должна быть записана в экспоненциальной форме, где интересующий предиктор оценивается при x + δ и при x для результата m относительно референтной группы, где δ — изменение интересующего нас предиктора (δ традиционно устанавливается равным единице), в то время как другие переменные в модели остаются постоянными. Если мы затем возьмем их отношение, то отношение уменьшится до отношения двух вероятностей, относительного риска. В этом смысле возведенный в степень полиномиальный логит-коэффициент обеспечивает оценку относительного риска. Однако возведенный в степень коэффициент обычно интерпретируется как отношение шансов. Стандартная интерпретация соотношений относительного риска заключается в том, что для единичного изменения предикторной переменной ожидается, что отношение относительного риска исхода m по отношению к референтной группе изменится на коэффициент соответствующей оценки параметра при условии, что переменные в модели остаются постоянными.

(These are the relative risk ratios for the multinomial logit model shown earlier. They can be obtained by exponentiating the multinomial logit coefficients, $e^{coef.}$, or by specifying the **rrr** option. Recall that the multinomial logit model estimates k-1 models, where the k^{th} equation is relative to the referent group. If the model was to be written out in an exponentiated form where the predictor of interest is evaluated at x + δ and at x for outcome m relative to referent group, where δ is the change in the predictor we are interested in (δ is traditionally is set to one) while the other variables in the model

are held constant. If we then take their ratio, the ratio would reduce to the ratio of two probabilities, the relative risk. In this sense, the exponentiated multinomial logit coefficient provides an estimate of relative risk. However, the exponentiated coefficient are commonly interpreted as odds ratios. Standard interpretation of the relative risk ratios is for a unit change in the predictor variable, the relative risk ratio of outcome m relative to the referent group is expected to change by a factor of the respective parameter estimate given the variables in the model are held constant.)

mlogit employment_T1 i.diplom_k c.age_10##c.age_10 male married Nind_child1 Nind_child2 Nind_child5 Nind_child17 NUM_adult18 lg_nolab_income lg_S_income lg_Other_income village ln_regwage regunempl i.fed_okr i.id_w if employment ==1 & origsm ==1, cluster(idind) baseoutcome(1) rrr

```
Iteration 0:   log pseudolikelihood = -14185.642
Iteration 1:   log pseudolikelihood = -12059.362
Iteration 2:   log pseudolikelihood = -11379.079
Iteration 3:   log pseudolikelihood = -11272.862
Iteration 4:   log pseudolikelihood =   -11269.4
Iteration 5:   log pseudolikelihood = -11269.384
Iteration 6:   log pseudolikelihood = -11269.384

Multinomial logistic regression                 Number of obs     =     28,675
                                                Wald chi2(70)     =    2236.89
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -11269.384               Pseudo R2         =     0.2056

                                          (Std. Err. adjusted for 8,857 clusters in idind)
------------------------------------------------------------------------------------------
                           |               Robust
            employment_T1 |        RRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------------------+--------------------------------------------------------------
неактивный                 |  (base outcome)
---------------------------+--------------------------------------------------------------
безработный                |
                  diplom_k |
     законченное среднее   |      1.925      0.182     6.940   0.000        1.600       2.316
среднее профессиональное   |      2.102      0.230     6.779   0.000        1.696       2.605
                   высшее  |      3.503      0.552     7.956   0.000        2.572       4.770
                           |
                    age_10 |      6.213      1.106    10.262   0.000        4.383       8.806
                           |
          c.age_10#c.age_10 |      0.750      0.017   -12.634   0.000        0.717       0.784
                           |
                      male |      1.230      0.093     2.746   0.006        1.061       1.426
                   married |      0.921      0.133    -0.570   0.568        0.694       1.222
               Nind_child1 |      0.744      0.181    -1.214   0.225        0.462       1.199
               Nind_child2 |      1.713      0.255     3.617   0.000        1.280       2.293
               Nind_child5 |      0.950      0.141    -0.343   0.732        0.710       1.271
              Nind_child17 |      1.045      0.066     0.691   0.490        0.923       1.183
              NUM_adult18  |      1.041      0.038     1.125   0.261        0.970       1.118
            lg_nolab_income |      0.971      0.011    -2.667   0.008        0.951       0.992
                lg_S_income |      0.969      0.015    -2.045   0.041        0.940       0.999
            lg_Other_income |      0.971      0.012    -2.460   0.014        0.948       0.994
                    village |      0.664      0.063    -4.337   0.000        0.552       0.799
                 ln_regwage |      1.091      0.174     0.545   0.586        0.798       1.492
                  regunempl |      1.025      0.012     2.136   0.033        1.002       1.049
                           |
                    fed_okr |
                   Северный  |      1.364      0.282     1.502   0.133        0.910       2.045
                Центральный  |      1.012      0.186     0.066   0.948        0.706       1.450
                Приволжский  |      0.930      0.171    -0.394   0.693        0.649       1.333
             Юг и С.Кавказ   |      1.020      0.209     0.096   0.924        0.683       1.523
                Уральский    |      1.142      0.222     0.684   0.494        0.781       1.670
                'Сибирский   |      1.160      0.214     0.806   0.420        0.808       1.666
           Дальневосточный   |      0.836      0.196    -0.764   0.445        0.527       1.324
                           |
                      id_w |
                 1995 год  |      0.879      0.135    -0.841   0.400        0.651       1.187
                 1996 год  |      0.760      0.123    -1.692   0.091        0.552       1.044
                 1998 год  |      0.613      0.104    -2.897   0.004        0.440       0.854
                 2000 год  |      0.536      0.086    -3.899   0.000        0.392       0.733
                 2001 год  |      0.476      0.078    -4.544   0.000        0.345       0.655
                 2002 год  |      0.491      0.083    -4.200   0.000        0.352       0.684
                 2003 год  |      0.465      0.085    -4.169   0.000        0.324       0.666
                 2004 год  |      0.480      0.089    -3.968   0.000        0.334       0.690
```

```
               2005 год  |      0.344      0.072   -5.106   0.000        0.229       0.518
               2006 год  |      0.383      0.081   -4.533   0.000        0.253       0.580
                         |
                   _cons |      0.006      0.009   -3.461   0.001        0.000       0.111
-------------------------+------------------------------------------------------------------
есть_любая_работа        |
                diplom_k |
     законченное среднее  |      1.908      0.112   11.040   0.000        1.701       2.140
среднее профессиональное  |      2.010      0.137   10.237   0.000        1.758       2.297
                  высшее  |      2.374      0.245    8.377   0.000        1.939       2.906
                         |
                   age_10 |      5.586      0.635   15.139   0.000        4.471       6.980
                         |
        c.age_10#c.age_10 |      0.776      0.011  -17.939   0.000        0.755       0.798
                         |
                     male |      1.510      0.074    8.457   0.000        1.373       1.662
                  married |      1.129      0.105    1.307   0.191        0.941       1.355
              Nind_child1 |      1.167      0.178    1.010   0.312        0.865       1.574
              Nind_child2 |      1.529      0.175    3.709   0.000        1.222       1.913
              Nind_child5 |      1.162      0.112    1.556   0.120        0.962       1.403
             Nind_child17 |      1.064      0.052    1.284   0.199        0.968       1.170
              NUM_adult18 |      1.100      0.025    4.116   0.000        1.051       1.150
          lg_nolab_income |      0.975      0.007   -3.489   0.000        0.961       0.989
              lg_S_income |      0.968      0.010   -3.258   0.001        0.950       0.987
          lg_Other_income |      0.980      0.008   -2.642   0.008        0.965       0.995
                 village  |      0.836      0.050   -2.983   0.003        0.744       0.941
               ln_regwage |      1.480      0.156    3.712   0.000        1.203       1.821
                regunempl |      1.002      0.008    0.200   0.841        0.986       1.018
                         |
                  fed_okr |
                Северный  |      0.917      0.133   -0.594   0.553        0.690       1.220
             Центральный  |      0.968      0.112   -0.280   0.779        0.771       1.216
             Приволжский  |      1.003      0.118    0.022   0.983        0.796       1.263
           Юг и С.Кавказ  |      1.042      0.145    0.299   0.765        0.794       1.369
              Уральский   |      1.065      0.137    0.487   0.627        0.827       1.370
              'Сибирский  |      1.069      0.129    0.551   0.582        0.844       1.354
           Дальневосточный  |      0.967      0.142   -0.231   0.818        0.725       1.289
                         |
                    id_w |
               1995 год  |      0.740      0.081   -2.735   0.006        0.597       0.918
               1996 год  |      0.844      0.094   -1.514   0.130        0.678       1.051
               1998 год  |      1.202      0.130    1.700   0.089        0.972       1.486
               2000 год  |      0.778      0.083   -2.363   0.018        0.632       0.958
               2001 год  |      0.706      0.076   -3.238   0.001        0.572       0.872
               2002 год  |      0.726      0.080   -2.914   0.004        0.585       0.900
               2003 год  |      0.649      0.078   -3.608   0.000        0.514       0.821
               2004 год  |      0.501      0.064   -5.409   0.000        0.391       0.644
               2005 год  |      0.645      0.084   -3.353   0.001        0.500       0.834
               2006 год  |      0.593      0.081   -3.841   0.000        0.454       0.774
                         |
                   _cons |      0.001      0.001   -7.477   0.000        0.000       0.005
-------------------------------------------------------------------------------------------
```

36. Модель вероятности вступить в брак для женщин (формальный или неформальный) в будущем году, для волн 9-16 – по ограниченному набору детерминант.

logit married_T1 c.age_10##c.age_10 employed lg_I_income Other_income nfm ln_regincome i.diplom_k i.i4_k i.status_1 i.id_w if (married == 0 & id_w >= 9 & male == 0), cluster(idind) or
estat summarize

```
Iteration 0:   log pseudolikelihood = -3609.1314
Iteration 1:   log pseudolikelihood = -3306.2498
Iteration 2:   log pseudolikelihood = -3220.7014
Iteration 3:   log pseudolikelihood = -3215.4688
Iteration 4:   log pseudolikelihood = -3215.4253
Iteration 5:   log pseudolikelihood = -3215.4253

Logistic regression                             Number of obs     =     16,177
                                                Wald chi2(25)     =     326.11
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -3215.4253               Pseudo R2         =     0.1091

                                      (Std. Err. adjusted for 4,829 clusters in idind)
-------------------------------------------------------------------------------------------
                         |               Robust
             married_T1 | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
```

age_10	**2.317**	**0.417**	**4.668**	**0.000**	1.628	3.298
c.age_10#c.age_10	**0.867**	**0.020**	**-6.341**	**0.000**	0.829	0.906
employed	1.253	0.140	2.017	0.044	1.006	1.559
lg_I_income	1.016	0.017	0.948	0.343	0.983	1.049
Other_income	1.000	0.000	-1.426	0.154	1.000	1.000
nfm	**1.090**	**0.030**	**3.094**	**0.002**	1.032	1.151
ln_regincome	**1.452**	**0.216**	**2.504**	**0.012**	1.084	1.944
diplom_k						
законченное среднее	**1.573**	**0.189**	**3.761**	**0.000**	1.242	1.991
среднее профессиональное	**1.425**	**0.190**	**2.653**	**0.008**	1.097	1.851
высшее	1.168	0.190	0.952	0.341	0.848	1.608
i4_k						
украинцы, бе..	0.621	0.210	-1.408	0.159	0.320	1.205
народы Сев.Кавказа	0.832	0.176	-0.868	0.386	0.549	1.261
народы Поволжья и Севера	0.822	0.182	-0.882	0.378	0.532	1.270
татары, башкиры	**0.542**	**0.145**	**-2.289**	**0.022**	0.320	0.916
прочие европейские	1.186	0.414	0.490	0.624	0.599	2.349
прочие не европейские	0.355	0.079	-4.658	0.000	0.230	0.549
status_1						
обл.центр	1.447	0.310	1.723	0.085	0.950	2.203
другой город	1.414	0.314	1.559	0.119	0.915	2.185
село, пгт	**2.019**	**0.467**	**3.038**	**0.002**	1.283	3.176
id_w						
2001 год	0.814	0.109	-1.542	0.123	0.626	1.057
2002 год	0.863	0.113	-1.126	0.260	0.667	1.115
2003 год	0.831	0.117	-1.314	0.189	0.630	1.096
2004 год	0.730	0.106	-2.167	0.030	0.549	0.970
2005 год	0.851	0.136	-1.009	0.313	0.622	1.164
2006 год	**0.622**	**0.107**	**-2.769**	**0.006**	0.444	0.870
_cons	0.001	0.001	-5.193	0.000	0.000	0.010

```
estat summarize
```

Estimation sample logit

Number of obs = 16,177

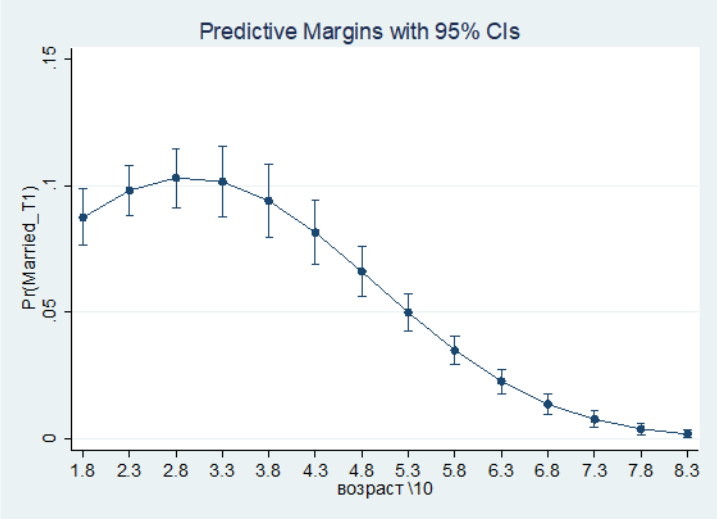
Number of clusters = 4,829
Obs per cluster: min = 1
avg = 3.3
max = 7

Variable	Mean	Std. Dev.	Min	Max
married_T1	.0586017	.234885	0	1
age_10	4.647982	2.317927	1.3	10.1
c.age_10#				
c.age_10	26.97619	22.09319	1.69	102.01
employed	.3823948	.4859872	0	1
lg_I_income	6.732072	3.098706	0	13.77081
Other_income	8748.741	24010.89	0	1007100
nfm	2.906287	1.692187	1	13
ln_regincome	8.784061	.5379421	7.800455	10.4148
diplom_k				
зако..	.260926	.4391532	0	1
среднее п..	.2282253	.4197015	0	1
высшее	.1450207	.3521326	0	1
i4_k				
украинцы,..	.0226865	.1489068	0	1
наро..	.0352352	.1843795	0	1
народы По..	.0371515	.1891388	0	1
тата..	.0263337	.1601306	0	1
проч..	.0116214	.1071778	0	1
прочие не..	.0586635	.2350011	0	1
status_1				
обл.центр	.2980157	.4574006	0	1
друг ..	.2590097	.4381045	0	1

```
   село, пгт  |      .322804      .4675629            0            1
              |
        id_w  |
    2001 год  |     .1359955      .3427944            0            1
    2002 год  |     .1434135      .3505049            0            1
    2003 год  |     .1421153      .3491792            0            1
    2004 год  |     .1463807      .3534984            0            1
    2005 год  |      .141559       .348608            0            1
    2006 год  |     .1702417      .3758566            0            1
------------------------------------------------------------------------------
Std. Dev. not adjusted for clustering
```

Предскажем, как изменяется вероятность вступить в брак с возрастом для женщин.

margins, at(age_10 ==(1.8(0.5)8.5))
marginsplot



37. Выполнение DO-файлов.
Сохраните рабочий файл. Закройте программу.
Выполните ранее сохраненный DO-файл **do_probit.do**

Файлы для отчета за работу на семинаре (дистант):
1. Файл с данными **ind_5_16_s7**
2. Два do-файла (код)
3. Файл аутпута STATA
4. График для возраста
5. Файл excel или word с оценкой модели Минцера (все, мужчины, женщины) из команды outreg2.

38. Самостоятельное задание.
38.1. Оцените модель факторов перехода из статуса занятого в году T в другие статусы занятости.
38.2. Включите в модель Минцера (без корректировки по Хекману) дамми с самооценкой здоровья. Сначала перекодируйте переменную **m3** в другую переменную **m3a**, так, чтобы значение 1 было «очень плохое здоровье», 2 «плохое здоровье», 3 – «среднее здоровье», 4 –

«хорошее здоровье», 5 – «очень хорошее здоровье». За базовую категорию этого набора дамми возьмите 3 – «среднее здоровье».
39.3. Оцените модель вероятности вступить в брак для мужчин.