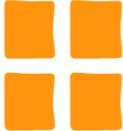


VII ITALIAN STATA USERS MEETING
Bologna, November 11-12, 2010

 **TABULA**

What You See Is What You Tabulate

Giovanni Luca Lo Magno
(lomagno.gl@virgilio.it)
Department of National Accounts and Social Processes Analysis
University of Palermo

Workflow of data analysis



Steps in the workflow

- 1) Cleaning data
- 2) Performing analysis
- 3) Presenting findings
- 4) Saving your work

Tasks within each step

- a) Planning
- b) Organization
- c) Documentation
- d) Execution

Criteria for choosing a workflow



- 1) Accuracy
- 2) Efficiency
- 3) Simplicity
- 4) Standardization
- 5) Automation
- 6) Usability
- 7) Scalability

How do you get it?



What You See Is What You Mean (WYSIWYM)

```
<html >
<body>
<h1>CHAPTER I</h1>
<h2>DOWN THE RABBIT-HOLE</h2>
<p>Alice was beginning to get very
tired of sitting by her sister on
the bank, and of having nothing to
do: once or twice she had peeped
into the book her sister was
reading, but it had no pictures or
conversations in it, "and what is
the use of a book' thought Alice,
'without pictures or
conversatoins?" </p>
</body>
</html >
```

What You See Is What You Get (WYSIWYG)

CHAPTER I

DOWN THE RABBIT-HOLE

Alice was beginning to get very tired of sitting by her sister on the bank, and of having nothing to do: once or twice she had peeped into the book her sister was reading, but it had no pictures or conversations in it, 'and what is the use of a book' thought Alice, 'without pictures or conversatoins?'

Sometimes what you get it's only "abc"!



This is what you see editing an RTF file manually:

```
{ \ rtf1\ ansi\ ansi cpg1252\ uc1 \ deff0\ deflang1033deflangfe1033{ \ fonttbl { \ f0\
froman\ fcharset0\ fprq2{ \ *\ panose\02020603050405020304} Times New Roman; } } { \
colortbl ;\ red0green0\ blue0;\ red0\ green0\ blue255;\ red0\ green255\ blue255;\ red0\
green255\ blue0;\ red255\ green0\ blue255;\ red255\ green0\ blue0;\ red255\ green255\
blue0;\ red255\ green255\ blue255;\ red0\green0\ blue128;\ red0\ green128\ blue128;\
red0\ green128\ blue0;\ red128\ green0\ blue128;\ red128\ green0\ blue0;\ red128\
green128\blue0;\ red128\ green128\ blue128;\ red192\ green192\ blue192; } { \
stylesheet{ \ widctlpar\ adjustright \ fs20\ cgrid \ snext0 Normal; } { \ *\ cs10 \
additive Default Paragraph Font; } } { \ info{ \ title } { \ author Steven Holzner } { \
operator Steven Holzner } { \ creatmyr2000\ mo\ dy\ hr\ min } { \ revtim\ yr2000\ mo\
dy17\ hr13\ min55 } { \ version1 } { \ edmins1 } { \ nofpages1 } { \ nofwords0 } { \
nofchars1 } { \ *\ company SteveCo } { \ nofcharsws1 } { \ vern89 } } \ widowctrl\
ftnbj\ aenddoc\ formshade\ viewkind4\ viewscale100\ pgbdrhead\ pgbdrfoot\ fet0\ sectd
\ psz1\ linex0\ endnhere\ sectdefaultcl { \ *\ psectlvl1\pnucrm\ pstart1\
pni ndent720\ pnhang{ \ pntxta . } } { \ *\ psectlvl2\pnucltr\ pstart1\ pni ndent720\
pnhang{ \ pntxta . } } { \ *\ psectlvl3\pndec\ pstart1\ pni ndent720\ pnhang{ \ pntxta
. } } { \ *\ psectlvl4\pncltr\ pstart1\ pni ndent720\ pnhang{ \ pntxta ) } } { \ *\
psectlvl5\pndec\ pstart1\ pni ndent720\ pnhang{ \ pntxtb ( ) { \ pntxta ) } } { \ *\
psectlvl6\ pncltr\ pstart1\ pni ndent720\ pnhang{ \ pntxtb ( ) { \ pntxta ) } } { \ *\
psectlvl7\ pnclrm\ pstart1\ pni ndent720\ pnhang{ \ pntxtb ( ) { \ pntxta ) } } { \ *\
psectlvl8\ pncltr\ pstart1\pni ndent720\ pnhang{ \ pntxtb ( ) { \ pntxta ) } } { \ *\
psectlvl9\ pnclrm\pstart1\ pni ndent720\ pnhang{ \ pntxtb ( ) { \ pntxta ) } } \ pard\
plain\sl480\ slmul t1\ widctlpar\ adjustright \ fs20\ cgrid { \ b\ fs24\ ul abc } { \ b\
ul \ par } }
```

and this is what you get:

abc

Creating a table of summary statistics



The Stata WYSIWYM approach

```
. sysuse nlsw88  
(NLSW, 1988 extract)
```

```
. table race married, contents(mean wage) format(%3.1f)
```

```
-----  
      race |      married  
           | single  married  
-----+-----  
white |      8.9      7.7  
black |      6.7      7.0  
other |      8.4      8.6  
-----
```

Notes about the `table` command:

- Up to seven-way tables can be created
- Up to five statistics may be displayed in each cell of the table.
- Text inside the table can't be obtained.



What is **TABULA** ?

- It's a software that helps you to create and save complex statistical tables.
- It's a complete software written in C++: it's not a Stata command.
- It's a graphic user interface (GUI) front end for Stata.
- It's not a complete statistical software, but it uses Stata as statistical engine by executing it in batch mode.



- Tabula is written in C++.
- Tabula uses Qt 4.5.3 C++ libraries for the GUI. Qt (pronounced officially as “cute”) is a project maintained by Nokia™ and a developers’ community.
- Qt is cross-platform, but until now Tabula was compiled only for Windows.
- Tabula saves files in XML format.
- The output file is in CSV format.

Disclaimer

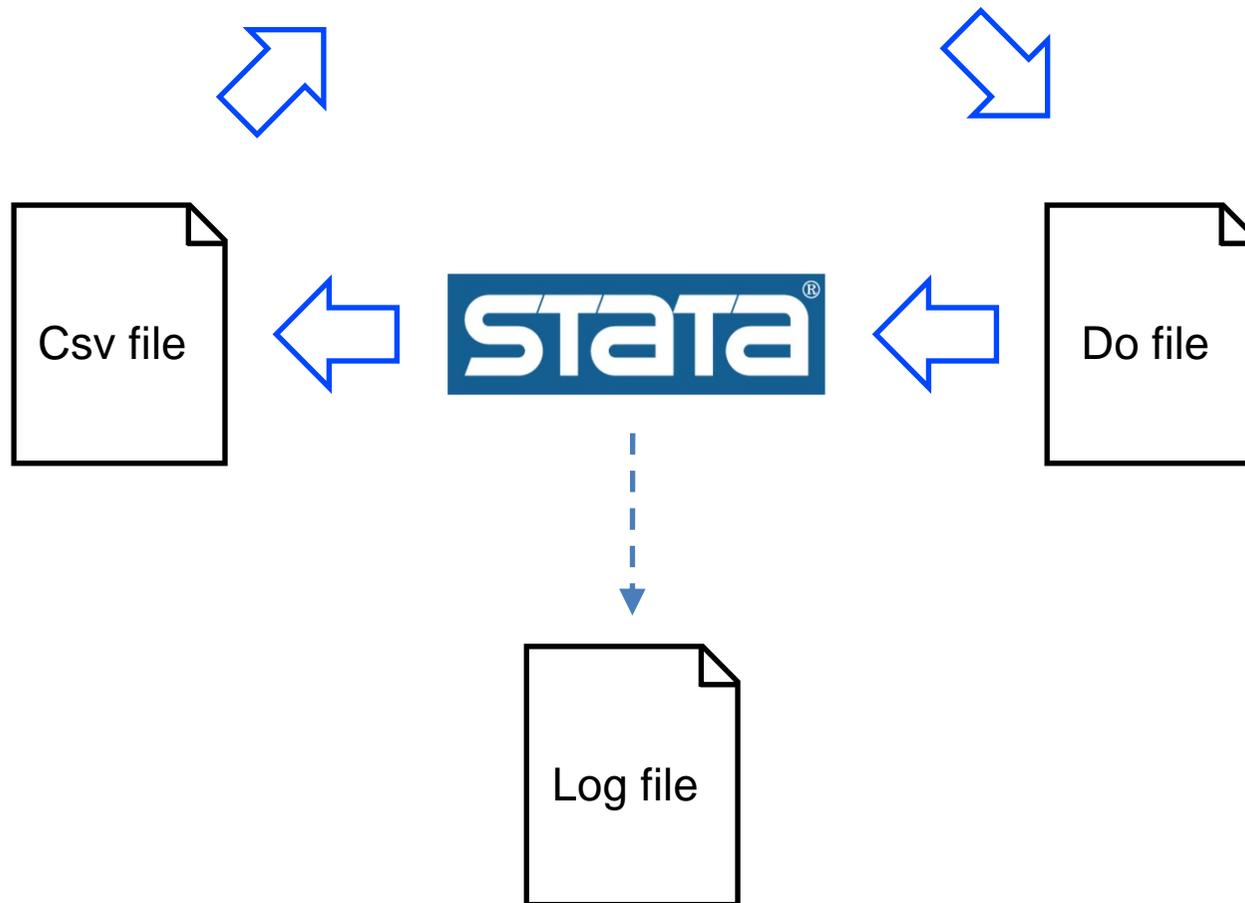


Tabula is released in beta version.
Many bugs are present and many hours of testing,
debugging and more development are still needed.
Users are the only people responsible for the
correctness of computations.

How Tabula works



TABULA



How humans read a table

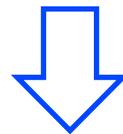


TAB. 7.5 MEAN HOURLY WAGE

Race	Marital status	
	<i>Single</i>	<i>Married</i>
<i>White</i>	8.9	7.7
<i>Black</i>	6.7	7.0
<i>Other</i>	8.4	8.6

Annotations: A blue arrow labeled "column" points to the "Marital status" header. A blue arrow labeled "row" points to the "Black" row. A dashed blue box highlights the value 7.0 in the "Black" row and "Married" column.

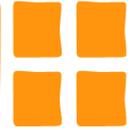
The figure 7.0 is the mean hourly wage of individuals whose race is "black" and marital status is "married"



The implicit condition:

IF RACE=*BLACK* AND MARITAL STATUS=*MARRIED*

Tabula: the basic idea



```
. label list race1 bl  
race1 bl :
```

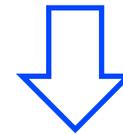
```
1 whi te  
2 bl ack  
3 othe r
```

```
. label list mar1 bl  
mar1 bl :
```

```
0 si ngl e  
1 marri ed
```

TAB. 7.5 MEAN HOURLY

WAGE	Marital status	
	<i>Single</i>	<i>Married</i>
<i>White</i>	8.9	7.7
<i>Black</i>	6.7	7.0
<i>Other</i>	8.4	8.6



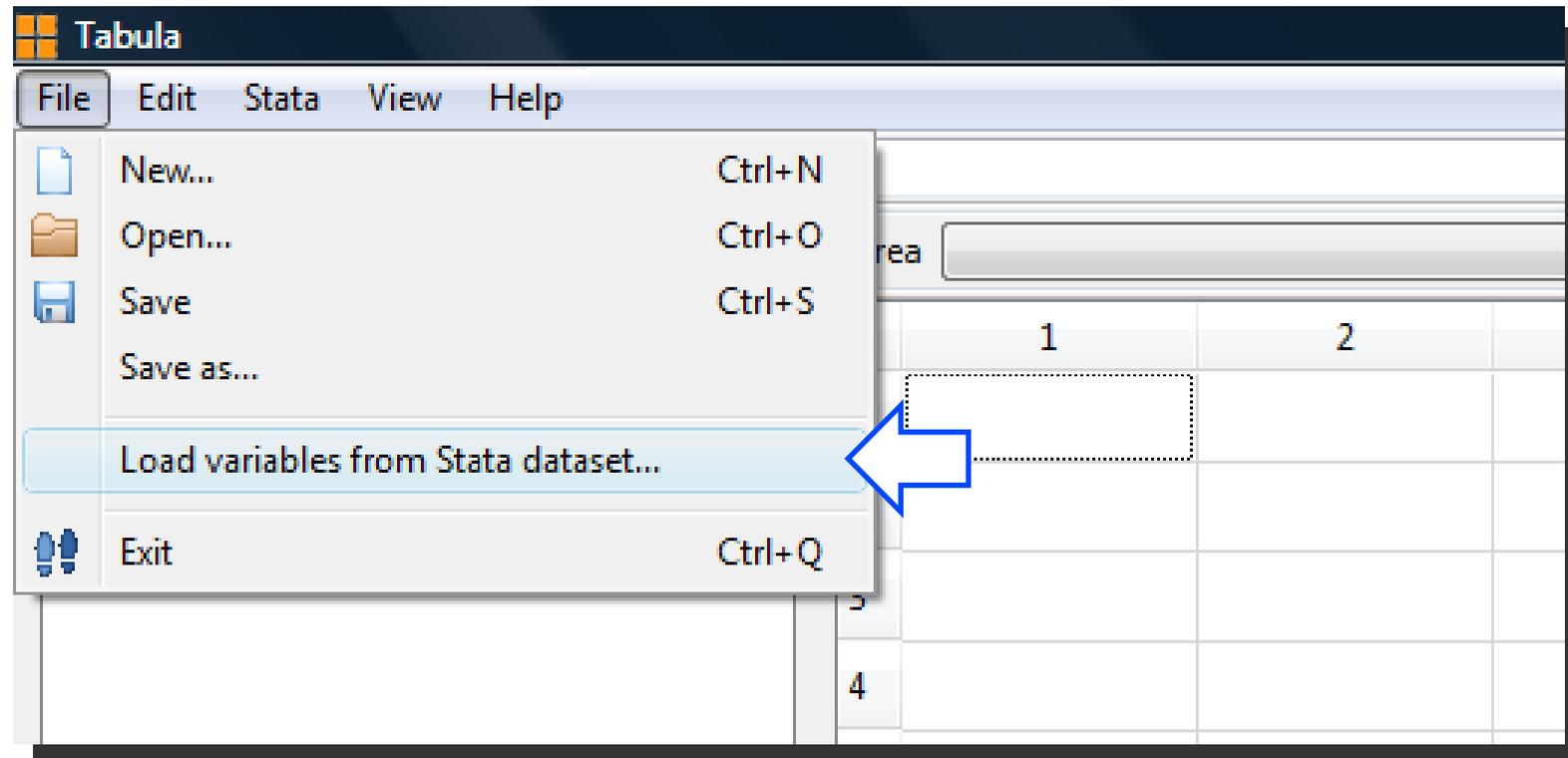
The implicit command:

```
summarize wage if race==1 and married==1
```

Loading variables from Stata dataset



Load nlsw88.dta (example dataset installed in Stata)



Tabula 0.3 beta loads variables only from format-114 datasets
(not Stata 9 datasets)

Variables and modalities loaded in Tabula



The screenshot shows the Tabula software interface. The main window is titled "Tabula" and has a menu bar with "File", "Edit", "Stata", "View", and "Help". Below the menu bar, there are fields for "Result", "Format" (set to 0/0), and "Area".

The "Variables" panel is open, showing a list of variables with their names and labels. A blue bracket on the left side of the panel is labeled "Variable list". The variables listed are:

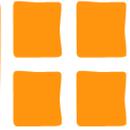
Name	Label
idcode	NLS id
age	age in current year
race	race
married	married
never_marri...	never married
grade	current grade completed
collgrad	college graduate
south	lives in south
smsa	lives in SMSA
c_city	lives in central city
industry	industry
occupation	occupation
union	union worker

The "Modality" panel is also open, showing the modalities for the selected "race" variable. A blue bracket on the left side of the panel is labeled "Modalities of 'race' variable". The modalities listed are:

Value	Label
1	white
2	black
3	other

The main data grid on the right side of the interface shows a table with 14 rows and 4 columns. The columns are labeled "1", "2", and "3". The rows are numbered 1 through 14. The data grid is currently empty.

Dragging and dropping a modality on a cell



Step 1: select modality

The screenshot shows the Tabula software interface. The 'Variables' panel on the left lists various variables, with 'married' selected. The 'Modality' panel below it shows the modalities for the selected variable: 0 (single) and 1 (married). The main data grid on the right is empty, with columns labeled 1, 2, 3, and 4, and rows numbered 1 through 14.

	1	2	3	4
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				

Dragging and dropping a modality on a cell



Step 2: drag modality

The screenshot shows the Tabula software interface. On the left, the 'Variables' panel lists variables with their labels. The 'married' variable is selected. Below it, the 'Modality' panel shows two modalities: '0 single' and '1 married'. A dashed blue arrow originates from the '0 single' modality and points to the cell at row 1, column 2 of the data grid. The data grid has columns labeled 1, 2, and 3, and rows labeled 1, 2, 3, and 4. The cell at row 1, column 2 is highlighted with a dashed border and a mouse cursor is positioned over it.

	1	2	3
1			
2			
3			
4			

Dragging and dropping a modality on a cell



Step 3: drop modality

The screenshot shows the Tabula software interface. The main window displays a data table with 4 rows and 3 columns. The cell at row 1, column 2 contains the text 'single'. The 'Variables' panel on the left lists various variables, including 'married' and 'never_marri...'. The 'Modality' panel at the bottom left shows a list of modalities with '0' and '1' as values, and 'single' and 'married' as labels. The 'single' modality is highlighted in blue.

	1	2	3
1		single	
2			
3			
4			

Variables

Variable	Name	Label
	idcode	NLS id
	age	age in current year
	race	race
	married	married
	never_marri...	never married
	grade	current grade completed
	collgrad	college graduate
	south	lives in south
	smsa	lives in SMSA
	c_city	lives in central city
	industry	industry
	occupation	occupation
	union	union worker

Modality

Value	Label
0	single
1	married

Completing your column headers



The screenshot shows the Tabula software interface. The main window displays a data table with columns 1, 2, and 3, and rows 1, 2, 3, and 4. The cell at row 1, column 2 contains the text "single", and the cell at row 1, column 3 contains the text "married". A dashed blue arrow points from the "married" label in the Modality section of the Variables panel to the "married" cell in the data table.

Variables Panel:

Variable	Name	Label
	idcode	NLS id
	age	age in current year
	race	race
	married	married
	never_marri...	never married
	grade	current grade completed
	collgrad	college graduate
	south	lives in south
	smsa	lives in SMSA
	c_city	lives in central city
	industry	industry
	occupation	occupation
	union	union worker

Modality Panel:

Value	Label
0	single
1	married

Quickly dragging and dropping a set of modalities



Step 1: select target cells

The screenshot shows the Tabula software interface. The main window displays a data table with 4 rows and 3 columns. The first row contains the values 'single' and 'married' in the second and third columns, respectively. The second, third, and fourth rows have their first column cells selected with a light blue background. The interface includes a menu bar (File, Edit, Stata, View, Help), a toolbar with 'Result', 'Format', and 'Area' options, and a 'Variables' panel on the left. The 'Variables' panel lists various variables such as 'idcode', 'age', 'race', 'married', etc. The 'Modality' panel below it shows a list of values (1, 2, 3) and their corresponding labels (white, black, other).

	1	2	3
1		single	married
2			
3			
4			

Quickly dragging and dropping a set of modalities



Step 2: select all modalities

The screenshot shows the Tabula software interface. The 'Variables' panel on the left lists various variables, with 'race' selected. The 'Modality' panel below it shows the modalities for the selected variable: 1 (white), 2 (black), and 3 (other). A dashed blue arrow points to the 'Select all modalities' icon (a grid of dots) in the Modality panel.

Variable	Label
idcode	NLS id
age	age in current year
race	race
married	married
never_marri...	never married
grade	current grade completed
collgrad	college graduate
south	lives in south
smsa	lives in SMSA
c_city	lives in central city
industry	industry
occupation	occupation
union	union worker

Value	Label
1	white
2	black
3	other

Click here to select all modalities

Quickly dragging and dropping a set of modalities



Step 3: drag and drop the previously selected set of modalities on the target selection

The screenshot shows the Tabula software interface. On the left, the 'Variables' panel lists variables such as 'idcode', 'age', 'race', 'married', etc. Below this is the 'Modality' panel, which shows a list of modalities: 1 (white), 2 (black), and 3 (other). On the right, a data table is displayed with columns 1, 2, and 3. The table contains the following data:

	1	2	3
1		single	married
2	white		
3	black		
4	other		

A dashed blue arrow points from the 'black' modality in the 'Modality' list to the 'black' cell in the table, indicating a drag-and-drop action.

Creating your statistic



Click here to add a new statistic

Creating your statistic



Add statistic

Name:
mean

Command format:
summarize #var #if

Result
r(mean)

Result type

Scalar
 Macro
 Matrix

Result matrix coords

Row Col

OK Cancel

Tabula will replace “#var” with the variable name and “#if” with the condition

Dragging and dropping variable on the table

The screenshot shows the Tabula software interface. On the left, a 'Variables' panel lists various variables. The 'wage' variable is selected. On the right, a data table is displayed with columns 1, 2, and 3. The table contains the following data:

	1	2	3
1		single	married
2	white	mean(wage)	mean(wage)
3	black	mean(wage)	mean(wage)
4	other	mean(wage)	mean(wage)

A dashed blue arrow points from the 'wage' variable in the list to the 'mean(wage)' text in the table cells, indicating the process of dragging and dropping the variable onto the table.

Statistic “mean” must be selected before dragging and dropping

Let's have a look at the cells



summarize wage if race==1 & married==0

	1	2	3
1		single	married
2	white	mean(wage)	mean(wage)
3	black	mean(wage)	mean(wage)
4	other	mean(wage)	mean(wage)

summarize wage if race==1 & married==1

	1	2	3
1		single	married
2	white	mean(wage)	mean(wage)
3	black	mean(wage)	mean(wage)
4	other	mean(wage)	mean(wage)

summarize wage if race==2 & married==0

	1	2	3
1		single	married
2	white	mean(wage)	mean(wage)
3	black	mean(wage)	mean(wage)
4	other	mean(wage)	mean(wage)

summarize wage if race==2 & married==1

	1	2	3
1		single	married
2	white	mean(wage)	mean(wage)
3	black	mean(wage)	mean(wage)
4	other	mean(wage)	mean(wage)

summarize wage if race==3 & married==0

	1	2	3
1		single	married
2	white	mean(wage)	mean(wage)
3	black	mean(wage)	mean(wage)
4	other	mean(wage)	mean(wage)

summarize wage if race==3 & married==1

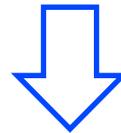
	1	2	3
1		single	married
2	white	mean(wage)	mean(wage)
3	black	mean(wage)	mean(wage)
4	other	mean(wage)	mean(wage)

How Tabula creates the Stata do file



What you see:

	1	2	3
1		single	married
2	white	mean(wage)	mean(wage)
3	black	mean(wage)	mean(wage)
4	other	mean(wage)	mean(wage)



What Tabula creates (extract):

```
Computations {
  [...]
  summarize wage if race==2 & married==0
  local t_2_3: display r(mean)
  [...]
Writing CSV file {
  forvalues row = 1/`nRow' {
    forvalues col = 1/`nCol' {
      file write outputFile "`t_`row' _ `col'" ;"
      [...]
    }
  }
  [...]
}
```

Let's tabulate!



The screenshot shows the Stata software interface with the 'Tabula' menu open. The menu options are:

- Tabula!
- Tabula...
- View output do file
- View last log file
- Stata executable...
- Stata dataset for computation...
- Output csv file name...
- Start Stata

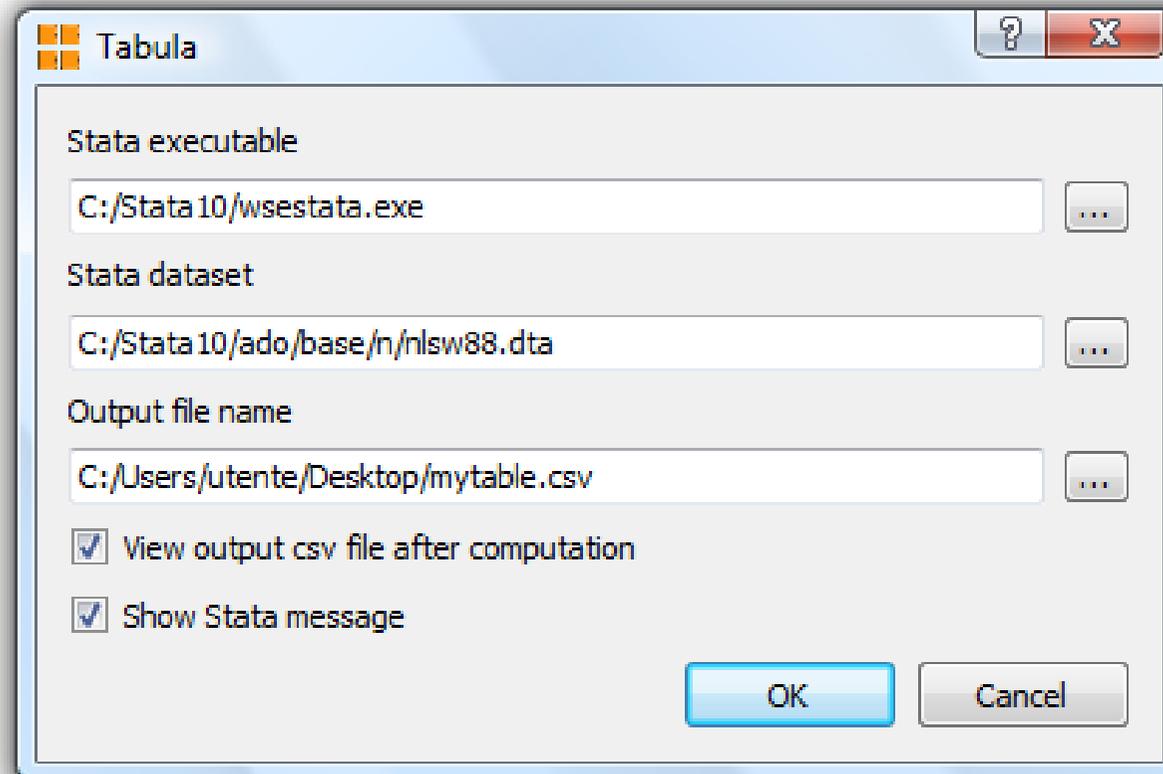
The background data table is as follows:

	1	2
		single
white		mean(wage)
black		mean(wage)
other		mean(wage)

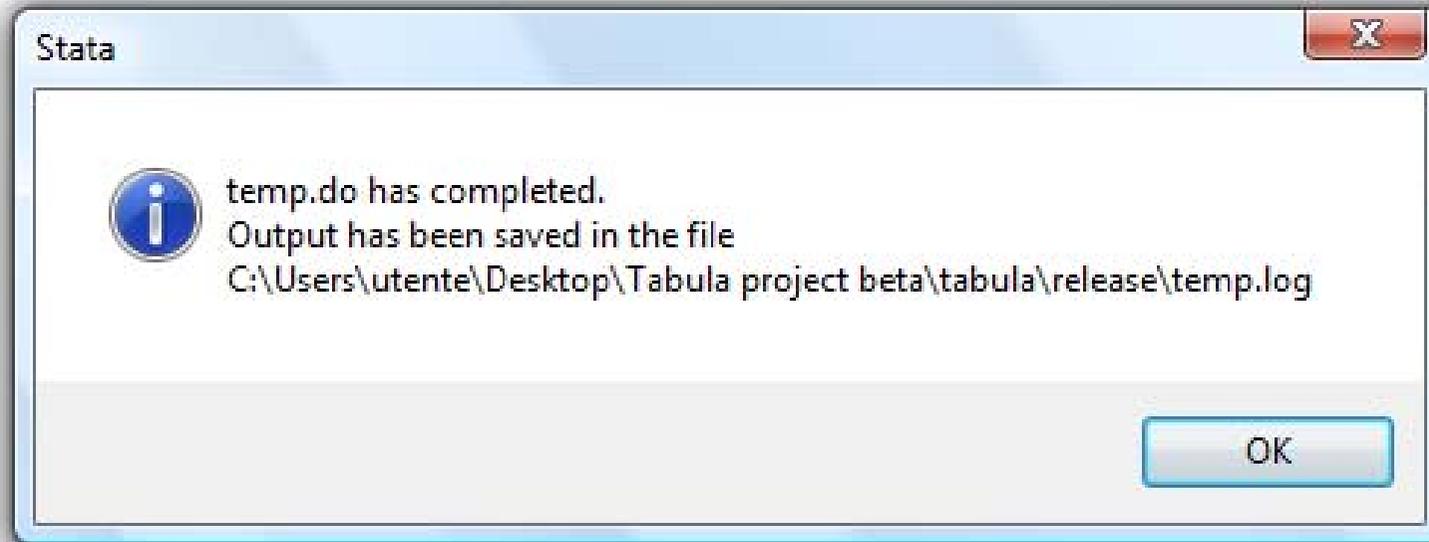
Below the table, the following variables are listed:

- race
- married
- never_m
- grade
- collgrad
- south
- current grade completed
- college graduate
- lives in south

The “Tabula” dialog window



The Stata message after computation



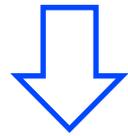
- Stata is executed in batch mode (Stata's GUI is not showed).
- You can suppress this message by unchecking the "Show Stata message" option in the "Tabula" dialog window.
- Stata saves a temp.log file containing the commands created by Tabula and the Stata's output.

The CSV table



The output CSV file

```
;single;married  
white;8.9292885;7.7246144  
black;6.7343413;6.968853  
other;8.4325389;8.6033334
```



	1	2	3
1		single	married
2	white	8.9292885	7.7246144
3	black	6.7343413	6.968853
4	other	8.4325389	8.6033334

- Tabula shows the output CSV file after Stata's computations.
- You can edit the text or copy cells to another software (for example Microsoft Excel or Microsoft Word).

Setting variables' output format



Tabula

File Edit Stata View Help

summarize wage if race==1 & married==0

Result r(mean) Format  Area

Variables  

Variable

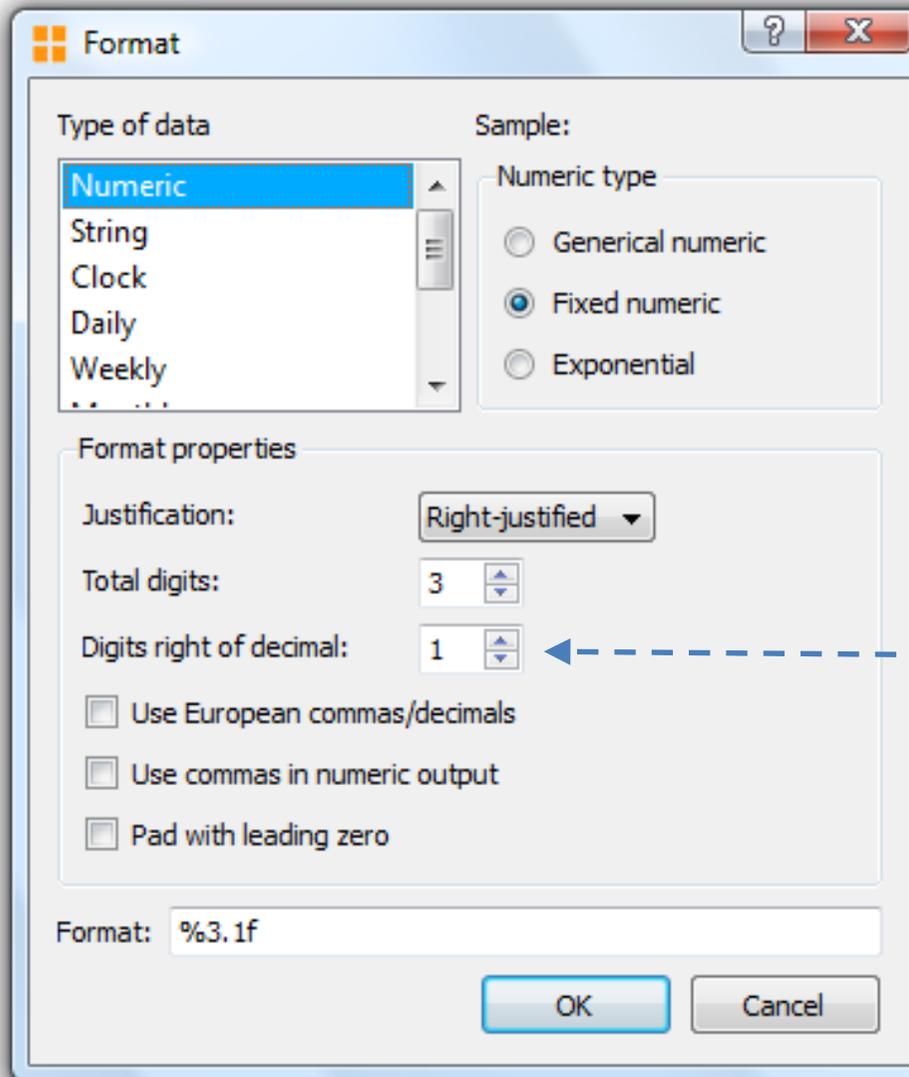
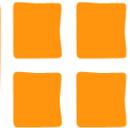
Name	Label
race	race
married	married
never_marri...	never married
grade	current grade completed

	1	2	3
1		single	married
2	white	mean(wage)	mean(wage)
3	black	mean(wage)	mean(wage)
4	other	mean(wage)	mean(wage)

2) Click here to set variables' output format

1) Select the cells containing statistics

The format dialog window



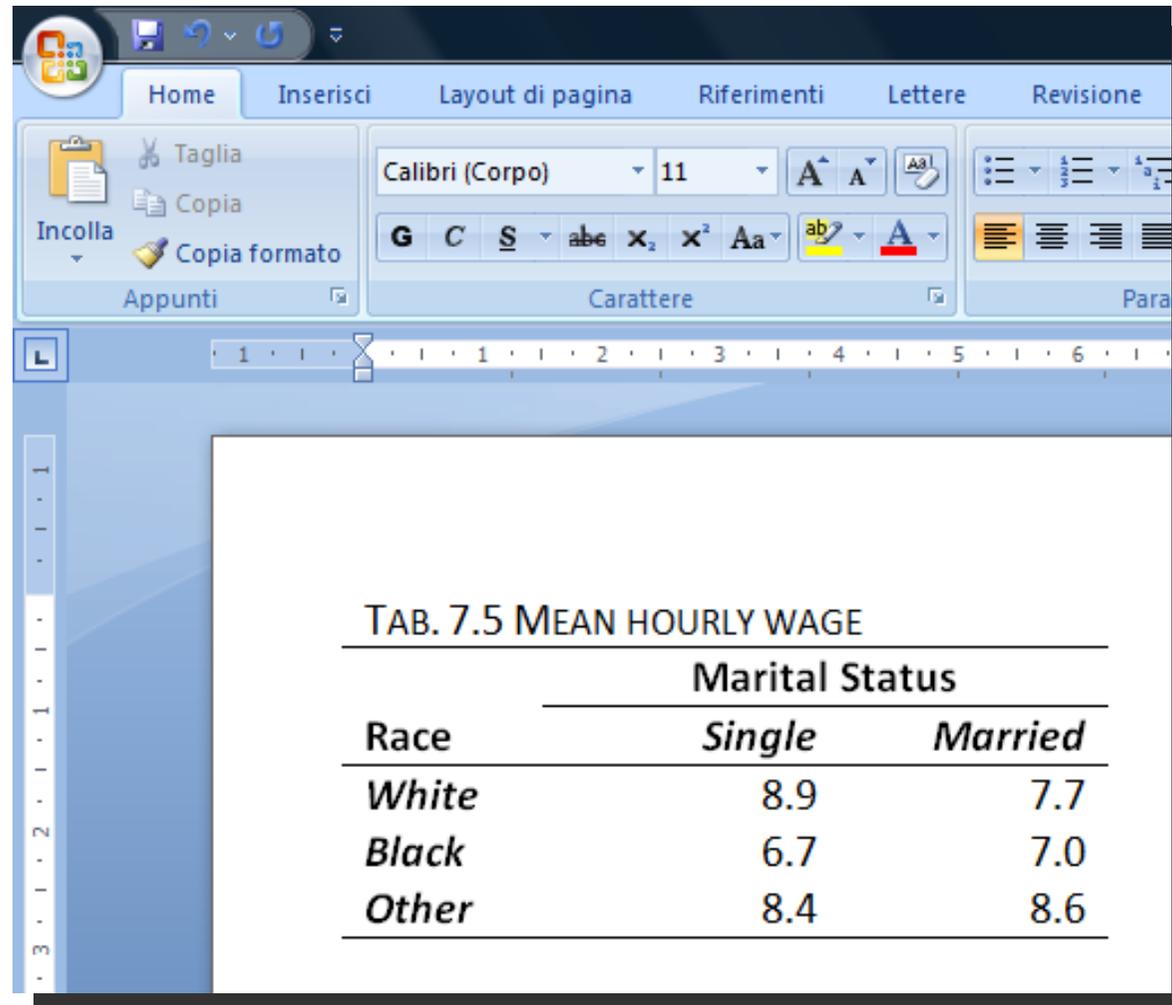
One digit after the decimal place

The table with the desired variable's output format



	1	2	3
1	single	married	
2	white	8.9	7.7
3	black	6.7	7.0
4	other	8.4	8.6

The final table edited in Microsoft Word



The screenshot shows the Microsoft Word interface with the 'Home' tab selected. The ribbon includes options for 'Inserisci', 'Layout di pagina', 'Riferimenti', 'Lettere', and 'Revisione'. The 'Carattere' group is visible, showing font size '11' and various formatting options. The table below is centered on the page and contains the following data:

TAB. 7.5 MEAN HOURLY WAGE		
	Marital Status	
Race	<i>Single</i>	<i>Married</i>
<i>White</i>	8.9	7.7
<i>Black</i>	6.7	7.0
<i>Other</i>	8.4	8.6



Ready for publication!

Adding marginal totals



Step 1: insert a new row

	1	2	3	
1		single	married	
2	white	mean(wage)	mean(wage)	
3	black	mean(wage)	mean(wage)	
4	other	mean(wage)	mean(wage)	

- Insert rows ▶
 - Up...
 - Down...
- Insert columns ▶
- Remove row(s)
- Remove column(s)
- Format statistic view ▶
- Format condition view ▶
- Set simple text
- Edit simple text

Adding marginal totals



Step 2: insert a new column

	1	2	3	
1		single	married	
2	white	mean(wage)	mean(wage)	
3	black	mean(wage)	mean(wage)	
4	other	mean(wage)	mean(wage)	
5				

- Insert rows
- Insert columns
 - Left...
 - Right...
- Remove row(s)
- Remove column(s)
- Format statistic view
- Format condition view
- Set simple text
- Edit simple text

Adding marginal totals



Step 3: set simple text for the last row header

	1	2	3	4
1		single	married	
2	white	mean(wage)	mean(wage)	
3	black	mean(wage)	mean(wage)	
4	other	mean(wage)	mean(wage)	
5				

Set simple text

Text:
Total

OK Cancel

Adding marginal totals



Step 4: set simple text for the last column header

	1	2	3	4
1		single	married	Total
2	white	mean(wage)	mean(wage)	
3	black	mean(wage)	mean(wage)	
4	other	mean(wage)	mean(wage)	
5	Total			

Adding marginal totals



Step 5: fill cells with statistics

	1	2	3	4
1		single	married	Total
2	white	mean(wage)	mean(wage)	mean(wage)
3	black	mean(wage)	mean(wage)	mean(wage)
4	other	mean(wage)	mean(wage)	mean(wage)
5	Total	mean(wage)	mean(wage)	mean(wage)

summarize wage if race==1

summarize wage if married==0

summarize wage

Creating a more complex table



The final table is wrong

	1	2	3	4
1		not college grad	not college grad	not college grad
2		single	married	Total
3	white	mean(wage)	mean(wage)	mean(wage)
4	black	mean(wage)	mean(wage)	mean(wage)
5	other	mean(wage)	mean(wage)	mean(wage)
6	Total	mean(wage)	mean(wage)	mean(wage)
7				
8		college grad	college grad	college grad
9		single	married	Total
10	white	mean(wage)	mean(wage)	mean(wage)
11	black	mean(wage)	mean(wage)	mean(wage)
12	other	mean(wage)	mean(wage)	mean(wage)
13	Total	mean(wage)	mean(wage)	mean(wage)

summarize wage if
 race==2 & col1grad==0
 & married==1



Right!

summarize wage if
 race==2 & col1grad==0
 & married==1 &
 col1grad==1 &
 married==1



Wrong!

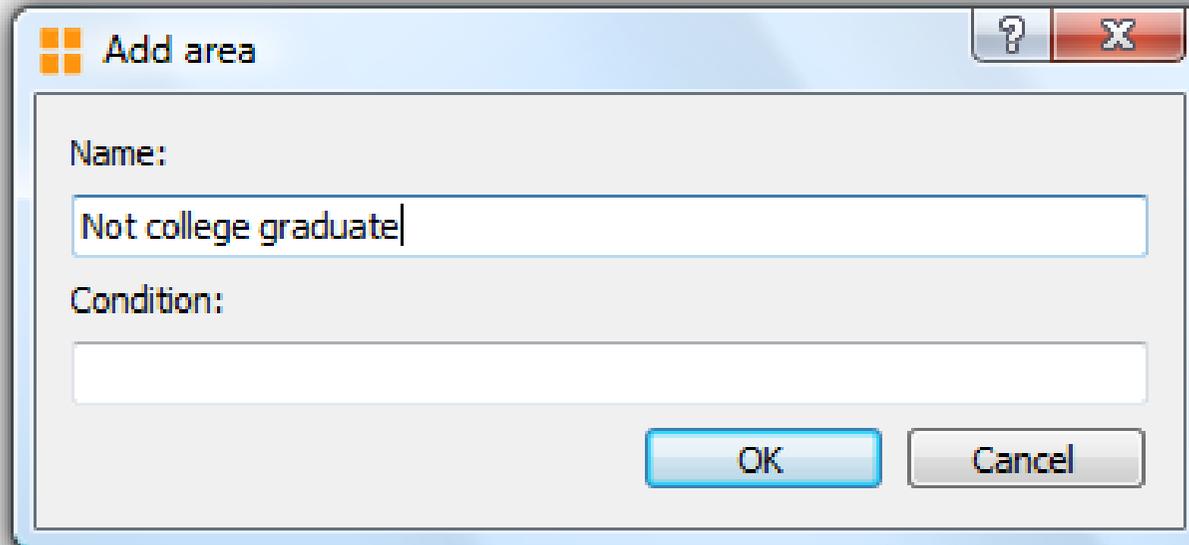
Assigning areas to cells



Step 1: add a new area



Click here



Assigning areas to cells



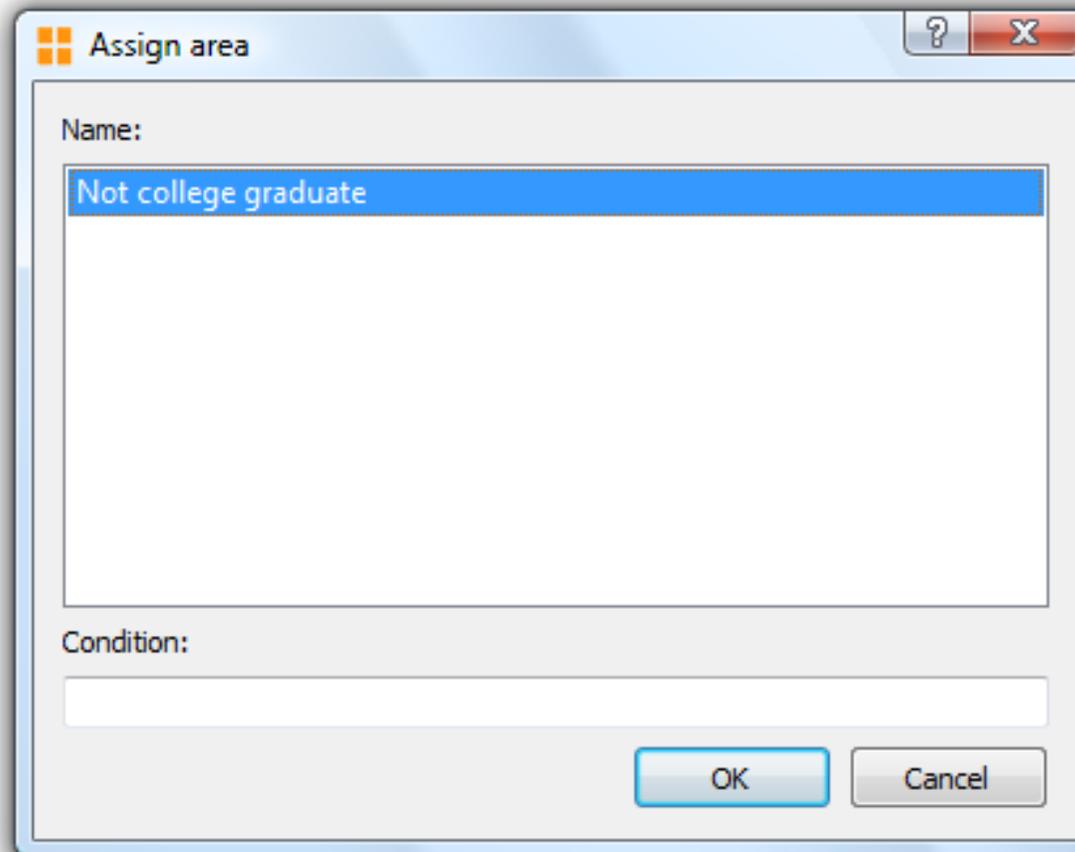
Step 2: select top part of the table

	1	2	3	4
1		not college grad	not college grad	not college grad
2		single	married	Total
3	white	mean(wage)	mean(wage)	mean(wage)
4	black	mean(wage)	mean(wage)	mean(wage)
5	other	mean(wage)	mean(wage)	mean(wage)
6	Total	mean(wage)	mean(wage)	mean(wage)
7				
8		college grad	college grad	college grad
9		single	married	Total
10	white	mean(wage)	mean(wage)	mean(wage)
11	black	mean(wage)	mean(wage)	mean(wage)
12	other	mean(wage)	mean(wage)	mean(wage)
13	Total	mean(wage)	mean(wage)	mean(wage)

Assigning areas to cells



Step 3: assign “Not college graduate” area to the top part of the table



Click here



Assigning areas to cells



Step 4: add a new area

The image shows a dialog box titled "Add area". It has a title bar with a question mark icon and a close button (X). The dialog contains two text input fields. The first field is labeled "Name:" and contains the text "College graduate". The second field is labeled "Condition:" and is currently empty. At the bottom of the dialog, there are two buttons: "OK" and "Cancel".

Add area

Name:
College graduate

Condition:

OK Cancel

Assigning areas to cells



Step 5: select bottom part of the table

	1	2	3	4
1		not college grad	not college grad	not college grad
2		single	married	Total
3	white	mean(wage)	mean(wage)	mean(wage)
4	black	mean(wage)	mean(wage)	mean(wage)
5	other	mean(wage)	mean(wage)	mean(wage)
6	Total	mean(wage)	mean(wage)	mean(wage)
7				
8		college grad	college grad	college grad
9		single	married	Total
10	white	mean(wage)	mean(wage)	mean(wage)
11	black	mean(wage)	mean(wage)	mean(wage)
12	other	mean(wage)	mean(wage)	mean(wage)
13	Total	mean(wage)	mean(wage)	mean(wage)

Assigning areas to cells



Step 6: assign “College graduate” area to the bottom part of the table

The image shows a dialog box titled "Assign area". It has a standard Windows-style title bar with a question mark icon and a close button (X). The dialog is divided into two main sections. The first section is labeled "Name:" and contains a list box with two items: "College graduate" and "Not college graduate". The "College graduate" item is currently selected and highlighted with a blue background. The second section is labeled "Condition:" and contains an empty text input field. At the bottom right of the dialog, there are two buttons: "OK" and "Cancel".

Assigning areas to cells



The final table is OK

	1	2	3	4
1		not college grad	not college grad	not college grad
2		single	married	Total
3	white	mean(wage)	mean(wage)	mean(wage)
4	black	mean(wage)	mean(wage)	mean(wage)
5	other	mean(wage)	mean(wage)	mean(wage)
6	Total	mean(wage)	mean(wage)	mean(wage)
7				
8		college grad	college grad	college grad
9		single	married	Total
10	white	mean(wage)	mean(wage)	mean(wage)
11	black	mean(wage)	mean(wage)	mean(wage)
12	other	mean(wage)	mean(wage)	mean(wage)
13	Total	mean(wage)	mean(wage)	mean(wage)

summarize wage if
race==2 & collgrad==0
& married==1



Right!

summarize wage if
race==2 & collgrad==1
& married==1



Right!

Using area conditions



Step 1: remove rows

	1	2	3	4
1		not college grad	not college grad	not college grad
2		single	married	Total
3	white	mean(wage)	mean(wage)	mean(wage)
4	black	mean(wage)	mean(wage)	mean(wage)
5	other	mean(wage)	mean(wage)	mean(wage)
6	Total	mean(wage)	mean(wage)	mean(wage)
7				
8		college grad	college grad	college grad
9		single	married	Total
10	white	mean(wage)	mean(wage)	mean(wage)
11	black	mean(wage)	mean(wage)	mean(wage)
12	other	mean(wage)	mean(wage)	mean(wage)
13	Total	mean(wage)	mean(wage)	mean(wage)

- Insert rows ▶
- Insert columns ▶
- Remove row(s)
- Remove column(s)
- Format statistic view ▶
- Format condition view ▶
- Set simple text
- Edit simple text

Using area conditions



Step 2: manage areas



Click here



Using area conditions



Step 3: set condition for “Not college graduate” area

The screenshot shows a dialog box titled "Area management" with a standard Windows-style title bar (minimize, maximize, close buttons). The dialog contains the following elements:

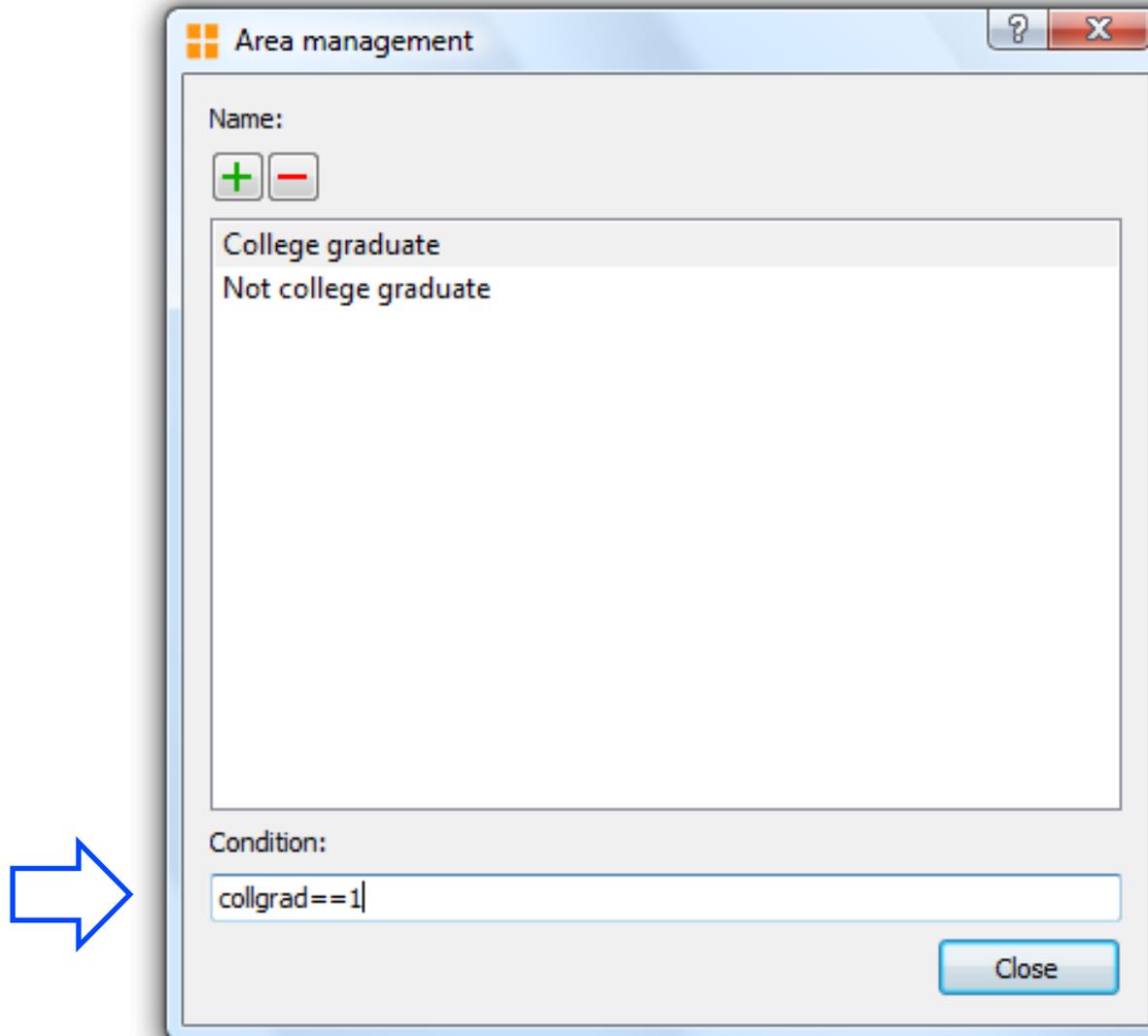
- Name:** A label above two small buttons: a green "+" button and a red "-" button.
- List:** A list box containing two items: "College graduate" and "Not college graduate". The "Not college graduate" item is currently selected and highlighted.
- Condition:** A label above a text input field containing the text "collgrad==0".
- Close:** A button labeled "Close" at the bottom right of the dialog.

A blue arrow points from the left side of the image towards the "Condition" field.

Using area conditions



Step 4: set area condition for “College graduate” area



Using area conditions



The final table is OK

	1	2	3	4
1		single	married	Total
2	white	mean(wage)	mean(wage)	mean(wage)
3	black	mean(wage)	mean(wage)	mean(wage)
4	other	mean(wage)	mean(wage)	mean(wage)
5	Total	mean(wage)	mean(wage)	mean(wage)
6				
7		single	married	Total
8	white	mean(wage)	mean(wage)	mean(wage)
9	black	mean(wage)	mean(wage)	mean(wage)
10	other	mean(wage)	mean(wage)	mean(wage)
11	Total	mean(wage)	mean(wage)	mean(wage)

summarize wage if
(col1grad==1) &
race==2 & married==1



Right!

summarize wage if
(col1grad==0) &
race==2 & married==1



Right!

Using the OR logical operator



Step 1: remove rows 4 and 10

	1	2	3	4	
1		single	married	Total	
2	white	mean(wage)	mean(wage)	mean(wage)	
3	black	mean(wage)	mean(wage)	mean(wage)	
4	other	mean(wage)	mean(wage)	mean(wage)	<ul style="list-style-type: none">Insert rows ▶Insert columns ▶Remove row(s) ◀Remove column(s)Format statistic view ▶Format condition view ▶Set simple textEdit simple text
5	Total	mean(wage)	mean(wage)	mean(wage)	
6					
7		single	married	Total	
8	white	mean(wage)	mean(wage)	mean(wage)	
9	black	mean(wage)	mean(wage)	mean(wage)	
10	other	mean(wage)	mean(wage)	mean(wage)	
11	Total	mean(wage)	mean(wage)	mean(wage)	

Using the OR logic operator

Step 2: drag and drop “other” modality in labels named “black”

The screenshot shows a software interface with a 'Variables' panel on the left and a data table on the right. The 'Variables' panel has a 'Variable' section with a list of variables and a 'Modality' section with a list of modalities. A dashed blue arrow points from the 'other' modality in the 'Modality' section to the 'black' label in the data table.

	1	2	3	4
1		single	married	Total
2	white	mean(wage)	mean(wage)	mean(wage)
3	black	mean(wage)	mean(wage)	mean(wage)
4	Total	mean(wage)	mean(wage)	mean(wage)
5				
6		single	married	Total
7	white	mean(wage)	mean(wage)	mean(wage)
8	black	mean(wage)	mean(wage)	mean(wage)
9	Total	mean(wage)	mean(wage)	mean(wage)

Variables Panel:

Variable

Name	Label
idcode	NLS id
age	age in current year
race	race
married	married
never_marri...	never married
grade	current grade completed
collgrad	college graduate
south	lives in south
smsa	lives in SMSA
c_city	lives in central city
industry	industry
occupation	occupation
union	union worker

Modality

Value	Label
1	white
2	black
3	other

Using the OR logic operator



The final table

	1	2	3	4
1		single	married	Total
2	white	mean(wage)	mean(wage)	mean(wage)
3	black other	mean(wage)	mean(wage)	mean(wage)
4	Total	mean(wage)	mean(wage)	mean(wage)
5				
6		single	married	Total
7	white	mean(wage)	mean(wage)	mean(wage)
8	black other	mean(wage)	mean(wage)	mean(wage)
9	Total	mean(wage)	mean(wage)	mean(wage)

summarize wage if (col | grad==0) & (race==2 | race==3) & married==1

Combining text and numbers in a table



Step 1: create the “smallsize.ado” file and put it in an ADO path

smallsize.ado

```
program define smallsize, rclass
    syntax [if]
    quietly count `if'
    if (r(N) < 60) {
        return local star "(*)"
    }
    else {
        return local star " "
    }
end
```

This ado program returns “(*)” if the number of observations satisfying the “if” condition is less than 60

Combining text and numbers in a table



Step 2: create the “smallsize” statistic with “macro” result selected

Edit statistic

Name:
smallsize

Command format:
smallsize #if

Result:
r(star)

Result type

Scalar

Macro

Matrix

Result matrix coords

Row Col

OK Cancel

Combining text and numbers in a table



Step 3: create a table like this

	1	2	3	4	5
1		single	single	married	married
2	Ag/Forestry/Fisheries	mean(wage)	smallsize(wage)	mean(wage)	smallsize(wage)
3	Mining	mean(wage)	smallsize(wage)	mean(wage)	smallsize(wage)
4	Construction	mean(wage)	smallsize(wage)	mean(wage)	smallsize(wage)
5	Manufacturing	mean(wage)	smallsize(wage)	mean(wage)	smallsize(wage)
6	Transport/Comm/Utility	mean(wage)	smallsize(wage)	mean(wage)	smallsize(wage)
7	Wholesale/Retail Trade	mean(wage)	smallsize(wage)	mean(wage)	smallsize(wage)
8	Finance/Ins/Real Estate	mean(wage)	smallsize(wage)	mean(wage)	smallsize(wage)
9	Business/Repair Svc	mean(wage)	smallsize(wage)	mean(wage)	smallsize(wage)
10	Personal Services	mean(wage)	smallsize(wage)	mean(wage)	smallsize(wage)
11	Entertainment/Rec Svc	mean(wage)	smallsize(wage)	mean(wage)	smallsize(wage)
12	Professional Services	mean(wage)	smallsize(wage)	mean(wage)	smallsize(wage)
13	Public Administration	mean(wage)	smallsize(wage)	mean(wage)	smallsize(wage)

Combining text and numbers in a table



The final table

The screenshot shows a window titled 'Csv viewer' with a menu bar containing 'File' and 'Edit'. The window displays a table with 13 rows and 5 columns. The first row has a dotted border around the cell in column 1. The data in the table is as follows:

	1	2	3	4	5
1		single	single	married	married
2	Ag/Forestry/Fisheries	2.3	(*)	6.3	(*)
3	Mining	17.2	(*)	9.7	(*)
4	Construction	8.1	(*)	7.2	(*)
5	Manufacturing	7.7		7.3	
6	Transport/Comm/Utility	12.0	(*)	10.9	(*)
7	Wholesale/Retail Trade	6.5		5.9	
8	Finance/Ins/Real Estate	11.1	(*)	9.3	
9	Business/Repair Svc	5.6	(*)	8.5	(*)
10	Personal Services	4.4	(*)	4.4	(*)
11	Entertainment/Rec Svc	7.9	(*)	5.4	(*)
12	Professional Services	8.5		7.6	
13	Public Administration	9.5		8.9	

Further examples: a 5-way table



	1	2	3	4
1		single	married	Total
2	Race			
3	white	mean(wage)	mean(wage)	mean(wage)
4	black	mean(wage)	mean(wage)	mean(wage)
5	other	mean(wage)	mean(wage)	mean(wage)
6	Labor union			
7	nonunion	mean(wage)	mean(wage)	mean(wage)
8	union	mean(wage)	mean(wage)	mean(wage)
9	Degree			
10	not college grad	mean(wage)	mean(wage)	mean(wage)
11	college grad	mean(wage)	mean(wage)	mean(wage)
12	Population density			
13	nonSMSA	mean(wage)	mean(wage)	mean(wage)
14	SMSA	mean(wage)	mean(wage)	mean(wage)
15	Total	mean(wage)	mean(wage)	mean(wage)

Further examples: different statistics in the same table



	1	2	3	4	5	6	7
1		single	single	single	married	married	married
2		Mean	Median	Std. Dev.	Mean	Median	Std. Dev.
3	Professional/technical	mean(wage)	median(wage)	sd(wage)	mean(wage)	median(wage)	sd(wage)
4	Managers/admin	mean(wage)	median(wage)	sd(wage)	mean(wage)	median(wage)	sd(wage)
5	Sales	mean(wage)	median(wage)	sd(wage)	mean(wage)	median(wage)	sd(wage)
6	Clerical/unskilled	mean(wage)	median(wage)	sd(wage)	mean(wage)	median(wage)	sd(wage)
7	Craftsmen	mean(wage)	median(wage)	sd(wage)	mean(wage)	median(wage)	sd(wage)
8	Operatives	mean(wage)	median(wage)	sd(wage)	mean(wage)	median(wage)	sd(wage)
9	Transport	mean(wage)	median(wage)	sd(wage)	mean(wage)	median(wage)	sd(wage)
10	Laborers	mean(wage)	median(wage)	sd(wage)	mean(wage)	median(wage)	sd(wage)
11	Farmers	mean(wage)	median(wage)	sd(wage)	mean(wage)	median(wage)	sd(wage)
12	Farm laborers	mean(wage)	median(wage)	sd(wage)	mean(wage)	median(wage)	sd(wage)
13	Service	mean(wage)	median(wage)	sd(wage)	mean(wage)	median(wage)	sd(wage)
14	Household workers	mean(wage)	median(wage)	sd(wage)	mean(wage)	median(wage)	sd(wage)
15	Other	mean(wage)	median(wage)	sd(wage)	mean(wage)	median(wage)	sd(wage)
16	Total	mean(wage)	median(wage)	sd(wage)	mean(wage)	median(wage)	sd(wage)

Tip #1: How to call Tabula from Stata



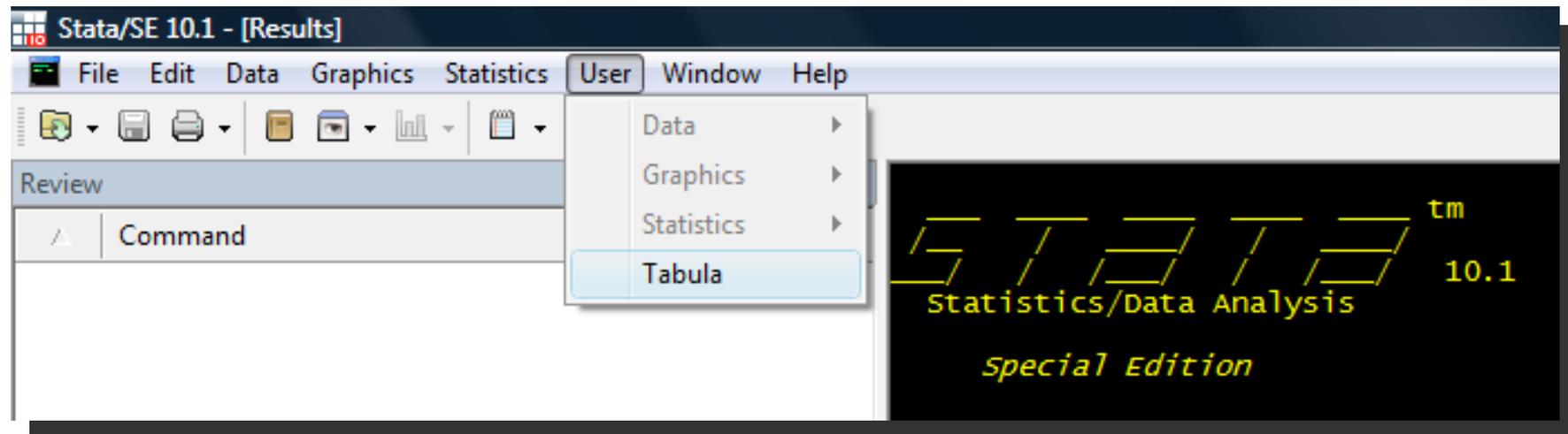
What you need is this useful ado file:

tabula_ado

```
program define tabula_  
  display  
  display as text "Note: Tabula will open the last saved dataset"  
  display as text "Check if your dataset was saved"  
  winexec "C:\Program Files\Tabula 0.3 beta\tabula 0.3 beta.exe" "$S_FN"  
end
```

- By typing “tabula_” from Stata you can call Tabula that open the last saved data set.
- Please modify the Tabula executable’s path in the ADO file according to your path.

Tip #2: How to call Tabula from the Stata's user menu



- 1) Create the `tabula_.ado` file (see tip #1).
- 2) Put the `tabula_.ado` file in a valid ado path (type `help adopath` for more informations).
- 3) Add this line to the `profile.do` file:
`window menu append item "stUser" "Tabula" `tabula_`"`

Tip #3: How to debug Tabula



Step 1: create a statistic with e(cmdline) macro result

Edit statistic

Name:
meancmd

Command format:
mean #var #if

Result:
e(cmdline)

Result type

Scalar
 Macro
 Matrix

Result matrix coords

Row Col

OK Cancel

Tip #3: How to debug Tabula



Step 2: drag and drop the statistic with “macro” result type

	1	2	3
1		single	married
2	white	meancmd(wage)	meancmd(wage)
3	black	meancmd(wage)	meancmd(wage)
4	other	meancmd(wage)	meancmd(wage)

Tip #3: How to debug Tabula: the self-explaining table



The final self-explaining table

The screenshot shows a window titled "Csv viewer" with a menu bar containing "File" and "Edit". The window displays a table with three columns labeled "1", "2", and "3". The table content is as follows:

	1	2	3
1		single	married
2	white	mean wage if race==1 & married==0	mean wage if race==1 & married==1
3	black	mean wage if race==2 & married==0	mean wage if race==2 & married==1
4	other	mean wage if race==3 & married==0	mean wage if race==3 & married==1

Advantages from using Tabula



Stop reinventing the wheel



Avoid error-prone complex do files



Self-explaining data analysis

Possible improvements to Tabula



- Allowing user to assign more than one area to each cell
- Adding “undo” and “redo” functions
- Adding copy, cut and paste functions for cells
- Allowing user to save and load “statistic” commands list
- Allowing user to choose delimiter for the CSV file
- Integrating Tabula with document markup languages like Latex or Html
- Creating the on line help
- Creating versions of Tabula for other operating systems
- Fixing many bugs

A possible future: Tabula with Numerics by Stata



Numerics by Stata is a technology that allows stand-alone software to use Stata as a statistical engine

<http://www.stata.com/products/numbystata.html>



ADePT, a software by World Bank, uses Numerics by Stata for poverty analysis

<http://www.stata.com/news/statanews.25.2.pdf>

TABULA

Tabula could be improved using Numerics by Stata allowing:

- Direct connection to Stata (no intermediate do file would be needed).
- Real-time computations.



What You See Is What You Tabulate