

Unit 4 – Introduction to Stata *version 16*
Homework 1 of 2
SOLUTIONS

Log

```
. * From the top tool bar: FILE > OPEN. Following is what shows in command window
. use "http://people.umass.edu/biostat690c/data/wws1000.dta", clear
(Working Women Survey)

.
. * Preliminary ) Save to desktop
. cd/Users/cbigelow/Desktop
/Users/cbigelow/Desktop

. save wws1000.dta
file wws1000.dta saved

.
. * 1.) Review the data structure.

. describe

Contains data from wws1000.dta
  obs:      1,000                Working Women Survey
 vars:       30                8 Oct 2013 11:43
 size:     73,000                (_dta has notes)

-----
variable name  storage  display  value  variable label
              type    format   label
-----
idcode         int     %8.0g      Unique ID
age            byte     %8.0g      age in current year
race           byte     %8.0g      race
married        byte     %8.0g      married
collgrad       byte     %16.0g     college graduate
south          byte     %8.0g      lives in south
industry        byte     %23.0g     industry
occupation     byte     %22.0g     occupation
union          byte     %8.0g      union worker
wage           float    %9.0g      hourly wage
hours          byte     %8.0g      usual hours worked
nevermarried   byte     %8.0g      Woman never been married
yrschool       byte     %8.0g      Years of school completed
metro          byte     %9.0g      Does woman live in metro area?
ccity          byte     %8.0g      Does woman live a city center?
currexp        float    %9.0g      Years worked at current job
prevexp        float    %9.0g      Years worked at previous job
everworked     float    %9.0g      Has woman ever worked?
uniondues      float    %9.0g      Union Dues paid last week
marriedyrs     float    %9.0g      Years married (rounded to nearest year)
unempins       float    %9.0g      Under/Unemployment insur. received last week
numkids        float    %9.0g      Number of children
kidage1        float    %9.0g      Age of first child
```

kidage2	float	%9.0g				Age of second child
kidage3	float	%9.0g				Age of third child
grade	byte	%8.0g				current grade completed
grade4	byte	%9.0g				4 level Current Grade Completed
wage2	float	%9.0g				Wages, rounded to 2 digits
fwf	float	%9.0g				Frequency weight
networth	float	%9.0g				Net worth

Sorted by:						
.						
. codebook, compact						
Variable	Obs	Unique	Mean	Min	Max	Label

idcode	1000	1000	2590.679	1	5159	Unique ID
age	1000	28	36.276	21	83	age in current year
race	1000	3	1.275	1	3	race
married	1000	2	.64	0	1	married
collgrad	1000	2	.241	0	1	college graduate
south	1000	2	.422	0	1	lives in south
industry	991	12	8.088799	1	12	industry
occupation	994	13	4.592555	1	13	occupation
union	842	2	.236342	0	1	union worker
wage	1000	576	387.8163	0	380000	hourly wage
hours	998	59	37.40481	1	80	usual hours worked
nevermarried	1000	2	.104	0	1	Woman never been married
yrschool	998	11	13.15731	8	18	Years of school completed
metro	1000	2	.704	0	1	Does woman live in metro area?
ccity	1000	2	.297	0	1	Does woman live a city center?
currexp	994	26	5.114688	0	26	Years worked at current job
prevexp	994	23	6.031187	0	25	Years worked at previous job
everworked	1000	2	.972	0	1	Has woman ever worked?
uniondues	997	30	5.47342	0	29	Union Dues paid last week
marriedyrs	1000	12	3.558	0	11	Years married (rounded to nearest year)
unempins	1000	145	30.121	0	298	Under/Unemployment insur. received last week
numkids	1000	4	1.53	0	3	Number of children
kidage1	765	21	10.34771	0	21	Age of first child
kidage2	521	15	7.047985	0	14	Age of second child
kidage3	244	8	3.430328	0	7	Age of third child
grade	998	15	13.12425	4	18	current grade completed
grade4	998	4	2.533066	1	4	4 level Current Grade Completed
wage2	1000	513	7.82131	0	40.2	Wages, rounded to 2 digits
fwf	1000	10	4.356	0	9	Frequency weight
networth	1000	575	817.8472	-7000	33198.08	Net worth

```
.
. * 2b.) Print the first 5 observations of race, age, married and wage
```

```
. list race age married wage in 1/5
```

	race	age	married	wage
1.	1	35	1	10.53139
2.	1	37	1	13.55072
3.	1	41	1	10.46699
4.	1	41	1	7.004828
5.	1	35	1	5.297905

```
.
. ** Alternative Solution. Note the use of the letter "f"; "f" is for first.
```

```
. list race age married wage in f/5
```

	race	age	married	wage
1.	1	35	1	10.53139
2.	1	37	1	13.55072
3.	1	41	1	10.46699
4.	1	41	1	7.004828
5.	1	35	1	5.297905

```
.
. * 3.) Print the last 5 observations of south, industry, and occupation
```

```
. list south industry occupation in 996/1000
```

	south	industry	occupa~n
996.	1	11	13
997.	1	.	10
998.	0	6	4
999.	0	11	1
1000.	0	11	2

```
. ** Alternative Solution. Note the use of the letter "el"; "el" is for last.
```

```
. list south industry occupation in -5/1
```

	south	industry	occupa~n
996.	1	11	13
997.	1	.	10
998.	0	6	4
999.	0	11	1
1000.	0	11	2

```
.
. * 4b, c, d.) Produce a 2-way classification of race and married
. ** With row percentages
. tab2 race married, row
```

-> tabulation of race by married

Key
frequency
row percentage

race	married		Total
	0	1	
1	229 31.07	508 68.93	737 100.00
2	129 51.39	122 48.61	251 100.00
3	2 16.67	10 83.33	12 100.00
Total	360 36.00	640 64.00	1,000 100.00

```
. ** With column percentages
```

```
. tab2 race married, column
```

```
-> tabulation of race by married
```

Key
frequency
column percentage

race	married		Total
	0	1	
1	229 63.61	508 79.38	737 73.70
2	129 35.83	122 19.06	251 25.10
3	2 0.56	10 1.56	12 1.20
Total	360 100.00	640 100.00	1,000 100.00

```
. ** With cell percentages
```

```
. tab2 race married, cell
```

```
-> tabulation of race by married
```

Key
frequency
cell percentage

race	married		Total
	0	1	
1	229 22.90	508 50.80	737 73.70
2	129 12.90	122 12.20	251 25.10
3	2 0.20	10 1.00	12 1.20
Total	360 36.00	640 64.00	1,000 100.00

```
. * 5c.) Produce summary statistics for wage, separately for strata defined by race
. ** Sort on race
```

```
. sort race
```

```
. ** Solution using command table
```

```
. table race, contents(mean wage sd wage min wage max wage)
```

```
-----+-----
      race | mean(wage)   sd(wage)   min(wage)   max(wage)
-----+-----
      1 |    8.183961   6.100691   1.032247   40.19808
      2 |   1520.682   23984.96         0   380000
      3 |    7.79104    4.413307   1.80602   17.52817
-----+-----
```

```
. ** Alternative solution using command tabstat
```

```
. tabstat wage, by(race) stat(mean sd min max)
```

```
Summary for variables: wage
by categories of: race (race)
```

```
-----+-----
      race |      mean      sd      min      max
-----+-----
      1 |    8.183961  6.100691  1.032247  40.19808
      2 |   1520.682  23984.96         0   380000
      3 |    7.79104   4.413307   1.80602   17.52817
-----+-----
    Total |   387.8163  12016.41         0   380000
-----+-----
```

```
* 6c.) Obtain a simple random sample of size 10. Save.
```

```
** Good practice preliminary: Be sure to have saved source data first
```

```
. save wws1000.dta
```

```
file wws1000.dta already exists
r(602);
```

```
. ** Obtain sample of size 10
```

```
. sample 10, count
```

```
(990 observations deleted)
```

```
. ** Save size 10 data set to wws_sample10.dta
```

```
. save wws_sample10.dta
```

```
file wws_sample10.dta saved
```

* 7.) Export STATA data set to EXCEL using approach on p 34 of unit 5 notes

```
. export excel using "wvs_sample10", firstrow(variables)
file wvs_sample10.xls saved

. log close
```

Saving data to Excel – solution

The screenshot shows the STATA software interface. On the left, the 'File' menu is open, and 'Export' is selected. The 'Export' submenu is visible, showing options like 'Excel spreadsheet (*.xls;*.xlsx)', 'Comma- or tab-separated data', 'Text data (fixed- or free-format)', 'SAS XPORT', 'ODBC data source', and 'XML data'. The 'Excel spreadsheet (*.xls;*.xlsx)' option is highlighted. On the right, the 'export excel - Export to Excel file' dialog box is open. The 'Main' tab is selected. The 'Variables' field is empty. The 'Excel filename' field contains 'wvs_sample10'. The 'Worksheet' section has several options: 'Worksheet name' (unchecked), 'Starting cell' (unchecked), 'Allow worksheet to be modified if it already exists' (unchecked), and 'Replace worksheet if it already exists' (unchecked). The 'Save variable names to first row in Excel file' option is checked. Other options like 'Save variable labels to first row in Excel file', 'Output numeric values (not labels) of labeled variables', and 'Overwrite Excel file' are unchecked. The 'Submit', 'Cancel', and 'OK' buttons are at the bottom.