

4) Converting continuous variables to categorical variables:

Example:

```
. use c:\sya6933\week7\age,clear
. count
  27
. sum age
```

Variable	Obs	Mean	Std. Dev.	Min	Max
age	27	40.92593	11.93226	21	59

```
. count if age>20 & age<=30
  6
. count if age>30 & age<=40
  6
. count if age>40 & age<=50
  8
. count if age>50 & age<=60
  7
```

/**normal way to do**/

```
. gen age_cat=0 if age>20 & age<=30
. replace age_cat=1 if age>30 & age<=40
. replace age_cat=2 if age>40 & age<=50
. replace age_cat=3 if age>50 & age<=60
. tab age_cat
```

age_cat	Freq.	Percent	Cum.
0	6	22.22	22.22
1	6	22.22	44.44
2	8	29.63	74.07
3	7	25.93	100.00
Total	27	100.00	

/**quick way 1 to do**/

```
. gen age_cat1=recode(age,20,30,40,50,60)
. tab age_cat1
```

age_cat1	Freq.	Percent	Cum.
30	6	22.22	22.22
40	6	22.22	44.44
50	8	29.63	74.07
60	7	25.93	100.00
Total	27	100.00	

/**quick way 2 to do**/

```
. gen age_cat2=autocode(age,4,21,59)
```

/**Note: autocode works like recode, except that all you tell the function is the range and the total number of cells that you want that range broken into. Stata broke age into 4 evenly spaced categories (30.5-21=40-30.5=49.5-40=59-49.5=9.5) from min 21 to max 59**/

```
. tab age_cat2
```

age_cat2	Freq.	Percent	Cum.
30.5	6	22.22	22.22
40	6	22.22	44.44
49.5	7	25.93	70.37
59	8	29.63	100.00
Total	27	100.00	

/of course you may use "sum age" to get the range of age, but you don't have to if you don't know the min and max of age**/**

```
. sort age          /**Note:you need to sort the age first**/
```

```
. gen age_cat3=autocode(age,4,age[1],age[_N])
```

```
. tab age_cat3
```

age_cat3	Freq.	Percent	Cum.
30.5	6	22.22	22.22
40	6	22.22	44.44
49.5	7	25.93	70.37
59	8	29.63	100.00
Total	27	100.00	

/quick way 3 to do**/**

```
. egen age_cat4=cut(age),at(20,30,40,50,60) label
```

```
. tab age_cat4
```

age_cat4	Freq.	Percent	Cum.
20-	6	22.22	22.22
30-	6	22.22	44.44
40-	7	25.93	70.37
50-	8	29.63	100.00
Total	27	100.00	

```
. /*****  
> 20-: 20 up to (but not including) 30  
> 30-: 30 up to (but not including) 40  
> 40-: 40 up to (but not including) 50  
> 50-: 50 up to (but not including) 60  
> *****/
```

/I don't like value:20-,30-,40-,50-**/**

```
. egen age_cat5=cut(age),at(20,30,40,50,60) icodes
```

```
. tab age_cat5
```

age_cat5	Freq.	Percent	Cum.
0	6	22.22	22.22
1	6	22.22	44.44

2	7	25.93	70.37
3	8	29.63	100.00

Total	27	100.00	

/**If you prefer, you can ask cut() to choose the cutoffs to form groups with approximately the same number per group. Below we request the creation of 4 (roughly) equally sized groups. ***/

```
. egen age_cat6=cut(age),group(4)
. table age_cat6
```

age_cat6	Freq.
0	6
1	7
2	7
3	7

2) Converting tables in Stata to Word:

Method 1:

Suppose that you want to save a table constructed using command “**tabulate**” in a form that makes it easy to convert into a table in MS Word. Stata has a **Copy Table** feature that you might find useful.

```
. sysuse auto,clear
(1978 Automobile Data)
```

```
. regress mpg foreign weight
```

Source	SS	df	MS	Number of obs = 74		
Model	1619.2877	2	809.643849	F(2, 71)	=	69.75
Residual	824.171761	71	11.608053	Prob > F	=	0.0000
-----				R-squared	=	0.6627
Total	2443.45946	73	33.4720474	Adj R-squared	=	0.6532
-----				Root MSE	=	3.4071

mpg	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
foreign	-1.650029	1.075994	-1.53	0.130	-3.7955	.4954422
weight	-.0065879	.0006371	-10.34	0.000	-.0078583	-.0053175
_cons	41.6797	2.165547	19.25	0.000	37.36172	45.99768

Make sure at the outset that you have set suitable options by clicking **Edit** in the menu bar and then **Table Copy Options**. Here removing all the vertical bars is advisable, so make sure **Remove all** is selected, and click **OK**. Now highlight the table in the Results window, and click **Edit** and then **Copy Table**.

In MS Word, click **Edit** and then **Paste**. Highlight the pasted text and click **Table** and then **Convert and Text to Table**. Specify **Tabs** under the **Separate text at** if it is not already selected. Click **OK** to create your table.

Choose font type: Courier New, and size: 9

mpg	Coef.	Std. Err.	t	P>t	[95% Conf. Interval]

foreign	-1.654524	1.082448	-1.53	0.131	-3.8134	.5043508
weight	-.0064743	.0006999	-9.25	0.000	-.0078701	-.0050785
headroom	-.2188614	.5416975	-0.40	0.687	-1.299243	.8615202
_cons	41.99314	2.312433	18.16	0.000	37.38113	46.60514

Method 2: use outreg2.ado to do the job: PP5

```
cd c:\sya6933
sysuse auto,clear
regress mpg foreign weight
outreg2 using myfile, word bdec(2)
regress mpg foreign weight headroom trunk
outreg2 using myfile, word bdec(3)
regress mpg foreign weight headroom trunk length turn displacement
outreg2 using myfile, word bdec(4)
```

This will create a table including all three models' results in MS Word .doc document named myfile under current directory. If you change "word" to "excel", it will create the file in excel format instead.

"bdec(#)" means how many decimal digits you want to keep for reported coefficients.

COEFFICIENT	(1) mpg	(2) mpg	(3) mpg
foreign	-1.65 (1.08)	-1.607 (1.092)	-1.9666 (1.1814)
weight	-0.01*** (0.00)	-0.006*** (0.001)	-0.0042** (0.0020)
headroom		-0.044 (0.638)	-0.0592 (0.6449)
trunk		-0.079 (0.150)	-0.0122 (0.1590)
length			-0.0631 (0.0644)
turn			-0.1649 (0.1981)
displacement			0.0008 (0.0103)
Constant	41.68*** (2.17)	41.886*** (2.333)	53.1377*** (7.5836)
Observations	74	74	74
R-squared	0.66	0.66	0.68

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

3) Extracting first name and last name and switching their order

```
. cd c:\sya6933
. clear
. input str40 fullname

1. "John Adams"
2. "Adam Smiths"
3. "Mary Smiths"
4. "Charlie Wade"
5. end

. list

+-----+
|      fullname      |
+-----+
1. |      John Adams   |
2. |      Adam Smiths  |
3. |      Mary Smiths  |
```

```
4. | Charlie Wade |
+-----+
```

```
. /**some people use this way**/
```

```
. gen n2 = regexs(2) if regexm(fullname, "([a-zA-Z]+) [ ]*([a-zA-Z]+)")
. gen n3 = regexs(3) if regexm(fullname, "([a-zA-Z]+) [ ]*([a-zA-Z]+)")
. gen newname1=regexs(3)+", "+regexs(2) if regexm(fullname, "([a-zA-Z]+) [ ]*([a-zA-Z]+)")
. list
```

```
+-----+
|      fullname      n2      n3      newname1 |
+-----+
1. |   John Adams   John   Adams   Adams, John |
2. |  Adam Smiths  Adam   Smiths  Smiths, Adam |
3. |  Mary Smiths  Mary   Smiths  Smiths, Mary |
4. | Charlie Wade Charlie   Wade   Wade, Charlie |
+-----+
```

```
. /**I use this way**/
```

```
. gen first_n=word(fullname,1)
. gen last_n=word(fullname,2)
. gen newname2=word(fullname,2)+", "+word(fullname,1)
. list
```

```
+-----+
|      fullname      n2      n3      newname1  first_n  last_n  newname2 |
+-----+
1. |   John Adams   John   Adams   Adams, John   John   Adams   Adams, John |
2. |  Adam Smiths  Adam   Smiths  Smiths, Adam   Adam   Smiths  Smiths, Adam |
3. |  Mary Smiths  Mary   Smiths  Smiths, Mary   Mary   Smiths  Smiths, Mary |
4. | Charlie Wade Charlie   Wade   Wade, Charlie Charlie   Wade   Wade, Charlie |
+-----+
```