

STATA 17

How to make publication ready tables
using the “table” and “collect”
commands

WHAT'S NEW?

- A system that allows you to:
 - **Collect** results from any Stata command
 - Create custom table **layouts** and **styles**
 - **Save** and **use** those layouts and styles
 - **Export** your tables to the most popular document formats (e.g. Word, Excel, HTML, LaTeX, PDF)

THE "TABLE" COMMAND

- Table types
 - One-way, two-way, multiway, summary statistics, regression comparisons, estimation and postestimation results

- syntax: table (RowVars)(ColVars)
- **statistic()** : Frequencies are displayed by default but you can use the statistic() option to specify other statistics
- **nformat()** specifies the numerical display format for the statistic
- **sformat()** adds strings to the statistics

```
. table (var) (highbp), ///
>     statistic(fvfrequency sex)   ///
>     statistic(fvpercent  sex)   ///
>     statistic(mean age)   ///
>     statistic(sd age)   ///
>     nototals ///
>         nformat(%9.0fc fvfrequency) ///
>         nformat(%6.2f mean sd) ///
>         sformat("%s%%" fvpercent) ///
>         sformat("(%s)" sd)
```

	High blood pressure	
	0	1
Sex=Male		
Factor variable frequency	2,611	2,304
Factor variable percent	43.70%	52.65%
Sex=Female		
Factor variable frequency	3,364	2,072
Factor variable percent	56.30%	47.35%
Age (years)		
Mean	42.17	54.97
Standard deviation	(16.77)	(14.91)

TABLE VS. COLLECT COMMANDS

- We could simply change the variable labels and value labels of our variables to change the labels in our tables
- Collections allow us to specify custom labels for a particular table without altering our original data
- We might want to change the labels for some tables and not others, or, we may wish to use different labels for our graphs
- Collections allow us to create and label dimensions that are not based on variables in our dataset (i.e. result)

THE “COLLECT” COMMAND

```
. collect dims
```

```
Collection dimensions
```

```
Collection: Table
```

```
-----  
                Dimension    No. levels  
-----  
Layout, style, header, label  
                across        2  
                cmdset        1  
                colname       1  
                command       1  
                highbp        2  
                result        4  
                sex           2  
                statcmd       4  
                var           2
```

```
Style only
```

```
                border_block  4  
                cell_type     4  
-----
```

- The **table** command automatically creates a **collection**
- You can view the dimensions of the collection using [collect dims](#)
 - Tells us number of levels of each dimension, etc.
- You can view the levels of a variable using [collect levelsof](#)
- You can view the level labels using [collect label list](#)
- You can add labels to the levels using [collect label levels](#)

```
. collect label list highbp, all
```

```
Collection: Table
```

```
Dimension: highbp
```

```
Label: High blood pressure
```

```
Level labels:
```

```
.m Total  
0  
1
```

```
. collect label levels highbp 0 "No" 1 "Yes"
```

```
. collect label list highbp, all
```

```
Collection: Table
```

```
Dimension: highbp
```

```
Label: High blood pressure
```

```
Level labels:
```

```
.m Total  
0 No  
1 Yes
```

RESULT

- **Result** is created by the **statistic()** option in our table command

```
. collect label list result, all
```

```
Collection: Table
```

```
Dimension: result
```

```
Label: Result
```

```
Level labels:
```

```
fvfrequency Factor variable frequency
```

```
fvpercent Factor variable percent
```

```
mean Mean
```

```
sd Standard deviation
```

- Use [collect recode](#) to create 2 new levels named column 1 and column 2
- Place them side by side using [collect layout](#) (next slide)

```
. collect recode result fvfrequency = column1
```

```
> fvpercent = column2
```

```
> mean = column1
```

```
> sd = column2
```

```
(18 items recoded in collection Table)
```

- These 4 dimensions are stacked on top of each other, what if we want them side by side?

COLLECT LAYOUT

- [collect layout](#) changes the layout
- Row dimension is still (var), but we nest our new column 1 and 2 under highbp using “ # ”

```
-----+-----
                |      High blood pressure
                |      0          1
-----+-----
Sex=Male
  Factor variable frequency |      2,611      2,304
  Factor variable percent  |      43.70      52.65
Sex=Female
  Factor variable frequency |      3,364      2,072
  Factor variable percent  |      56.30      47.35
Age (years)
  Mean                    |      42.16502   54.97281
  Standard deviation      |      16.77157   14.90897
-----+-----
```

```
. collect layout (var) (highbp#result[column1 column2])
```

```
Collection: Table
```

```
Rows: var
```

```
Columns: highbp#result[column1 column2]
```

```
Table 1: 3 x 4
```

```
-----+-----
                |      High blood pressure
                |      0          1
                |      column1   column2   column1   column2
-----+-----
Sex=Male      |      2611     43.69874     2304     52.65082
Sex=Female    |      3364     56.30126     2072     47.34918
Age (years)   |      42.16502  16.77157     54.97281  14.90897
-----+-----
```

CUSTOMIZE THE DISPLAY OF NUMBERS

- Use [collect style cell](#) to apply different formats to different cells in the table
 - use `nformat()` and `sformat()` as we did in the table command

```
. collect style cell var[sex race hlthstat]#result[column2],
>         nformat(%6.1f) sformat("%s%%")

. collect style cell
>         var[age bmi tcresult tgresult hdresult]#result[column1 column2],
>         nformat(%6.1f)

. collect style cell
>         var[age bmi tcresult tgresult hdresult]#result[column2],
>         sformat("(%s)")

. collect preview
```

	High blood pressure			
	0		1	
	column1	column2	column1	column2
Age (years)	42.2	(16.8)	55.0	(14.9)
Body mass index (BMI)	24.2	(4.1)	27.4	(5.3)
Sex=Male	2,611	43.7%	2,304	52.7%
Sex=Female	3,364	56.3%	2,072	47.3%
Race=White	5,317	89.0%	3,748	85.6%
Race=Black	545	9.1%	541	12.4%
Race=Other	113	1.9%	87	2.0%
Health status=Excellent	1,649	27.7%	758	17.3%
Health status=Very good	1,666	27.9%	925	21.2%
Health status=Good	1,572	26.4%	1,366	31.2%
Health status=Fair	766	12.8%	904	20.7%
Health status=Poor	310	5.2%	419	9.6%
Serum cholesterol (mg/dL)	208.7	(47.3)	229.9	(49.6)
Serum triglycerides (mg/dL)	129.2	(83.9)	166.0	(109.2)
High density lipids (mg/dL)	49.9	(14.1)	49.2	(14.5)

CUSTOMIZE COLUMN LABELS

	High blood pressure			
	0		1	
	column1	column2	column1	column2
Age (years)	42.2	(16.8)	55.0	(14.9)
Body mass index (BMI)	24.2	(4.1)	27.4	(5.3)
Sex=Male	2,611	43.7%	2,304	52.7%
Sex=Female	3,364	56.3%	2,072	47.3%
Race=White	5,317	89.0%	3,748	85.6%
Race=Black	545	9.1%	541	12.4%
Race=Other	113	1.9%	87	2.0%
Health status=Excellent	1,649	27.7%	758	17.3%
Health status=Very good	1,666	27.9%	925	21.2%
Health status=Good	1,572	26.4%	1,366	31.2%
Health status=Fair	766	12.8%	904	20.7%
Health status=Poor	310	5.2%	419	9.6%
Serum cholesterol (mg/dL)	208.7	(47.3)	229.9	(49.6)
Serum triglycerides (mg/dL)	129.2	(83.9)	166.0	(109.2)
High density lipids (mg/dL)	49.9	(14.1)	49.2	(14.5)



```
. collect label dim highbp "Hypertension", modify
```

Changes the label of the dimension highbp to "Hypertension"

```
. collect label levels highbp 0 "No" 1 "Yes"
```

Labels the level 0 "No" and level 1 "Yes"

```
. collect style header result, level(hide)
```

```
. collect preview
```

Hides the levels of dimension result in the column header (i.e., removes column 1 and column 2)

	Hypertension			
	No		Yes	
Age (years)	42.2	(16.8)	55.0	(14.9)
Body mass index (BMI)	24.2	(4.1)	27.4	(5.3)
Sex=Male	2,611	43.7%	2,304	52.7%
Sex=Female	3,364	56.3%	2,072	47.3%
Race=White	5,317	89.0%	3,748	85.6%
Race=Black	545	9.1%	541	12.4%
Race=Other	113	1.9%	87	2.0%
Health status=Excellent	1,649	27.7%	758	17.3%
Health status=Very good	1,666	27.9%	925	21.2%
Health status=Good	1,572	26.4%	1,366	31.2%
Health status=Fair	766	12.8%	904	20.7%
Health status=Poor	310	5.2%	419	9.6%
Serum cholesterol (mg/dL)	208.7	(47.3)	229.9	(49.6)
Serum triglycerides (mg/dL)	129.2	(83.9)	166.0	(109.2)
High density lipids (mg/dL)	49.9	(14.1)	49.2	(14.5)

CUSTOMIZE ROW LABELS

stack stacks the categories of the levels on top of each other instead of side by side

```
. collect style row stack, nobinder spacer
```

nobinder removes the "=" that appeared between each level and its categories

```
. collect style cell border_block, border(right, pattern(nil))
```

border(right, pattern(nil)) removes vertical line; **border_block** applies changes to all right borders

```
. collect preview
```

	Hypertension			
	No		Yes	
Age (years)	42.2	(16.8)	55.0	(14.9)
Body mass index (BMI)	24.2	(4.1)	27.4	(5.3)
Sex=Male	2,611	43.7%	2,304	52.7%
Sex=Female	3,364	56.3%	2,072	47.3%
Race=White	5,317	89.0%	3,748	85.6%
Race=Black	545	9.1%	541	12.4%
Race=Other	113	1.9%	87	2.0%
Health status=Excellent	1,649	27.7%	758	17.3%
Health status=Very good	1,666	27.9%	925	21.2%
Health status=Good	1,572	26.4%	1,366	31.2%
Health status=Fair	766	12.8%	904	20.7%
Health status=Poor	310	5.2%	419	9.6%
Serum cholesterol (mg/dL)	208.7	(47.3)	229.9	(49.6)
Serum triglycerides (mg/dL)	129.2	(83.9)	166.0	(109.2)
High density lipids (mg/dL)	49.9	(14.1)	49.2	(14.5)



	Hypertension			
	No		Yes	
Age (years)	42.2	(16.8)	55.0	(14.9)
Body mass index (BMI)	24.2	(4.1)	27.4	(5.3)
Sex				
Male	2,611	43.7%	2,304	52.7%
Female	3,364	56.3%	2,072	47.3%
Race				
White	5,317	89.0%	3,748	85.6%
Black	545	9.1%	541	12.4%
Other	113	1.9%	87	2.0%
Health status				
Excellent	1,649	27.7%	758	17.3%
Very good	1,666	27.9%	925	21.2%
Good	1,572	26.4%	1,366	31.2%
Fair	766	12.8%	904	20.7%
Poor	310	5.2%	419	9.6%
Serum cholesterol (mg/dL)	208.7	(47.3)	229.9	(49.6)
Serum triglycerides (mg/dL)	129.2	(83.9)	166.0	(109.2)
High density lipids (mg/dL)	49.9	(14.1)	49.2	(14.5)

READY TO EXPORT TO MS WORD

putdocx begin → use putdocx to add title, section header, text

putdocx paragraph, style(Title)

putdocx text ("Hypertension in the United States")

putdocx paragraph, style(Heading1)

putdocx text ("The National Health and Nutrition Examination Survey (NHANES)")

putdocx paragraph

putdocx text ("Hypertension is a major cause of morbidity and mortality in ")

putdocx text ("the United States. This report will explore the predictors ")

putdocx text ("of hypertension using the NHANES dataset.")

collect style putdocx, layout(autofitcontents) /// → use layout(autofitcontents) to retain original width of table

title("Table 1: Descriptive Statistics by Hypertension Status")

putdocx collect → exports our table to the document

putdocx save MyTable1.docx, replace

TABLE I

	High blood pressure		
	0	1	Total
Sex=Male			
Factor variable frequency	2,611	2,304	4,915
Factor variable percent	43.70	52.65	47.48
Sex=Female			
Factor variable frequency	3,364	2,072	5,436
Factor variable percent	56.30	47.35	52.52
Age (years)			
Mean	42.16502	54.97281	47.57965
Standard deviation	16.77157	14.90897	17.21483

AutoSave Off MyTable1.docx - Com... Sign in

File Home Insert Draw Design Layout References Mailings Review View Help ACROBAT

Hypertension in the United States

The National Health and Nutrition Examination Survey (NHANES)

Hypertension is a major cause of morbidity and mortality in the United States. This report will explore the predictors of hypertension using the NHANES dataset.

Table 1: Descriptive Statistics by Hypertension Status

	No		Hypertension		Total	
			Yes			
Age (years)	42.2	(16.8)	55.0	(14.9)	47.6	(17.2)
Body mass index (BMI)	24.2	(4.1)	27.4	(5.3)	25.5	(4.9)
Sex						
Male	2,611	43.7%	2,304	52.7%	4,915	47.5%
Female	3,364	56.3%	2,072	47.3%	5,436	52.5%
Race						
White	5,317	89.0%	3,748	85.6%	9,065	87.6%
Black	545	9.1%	541	12.4%	1,086	10.5%
Other	113	1.9%	87	2.0%	200	1.9%
Health status						
Excellent	1,649	27.7%	758	17.3%	2,407	23.3%
Very good	1,666	27.9%	925	21.2%	2,591	25.1%
Good	1,572	26.4%	1,366	31.2%	2,938	28.4%
Fair	766	12.8%	904	20.7%	1,670	16.2%
Poor	310	5.2%	419	9.6%	729	7.1%
Serum cholesterol (mg/dL)	208.7	(47.3)	229.9	(49.6)	217.7	(49.4)
Serum triglycerides (mg/dL)	129.2	(83.9)	166.0	(109.2)	143.9	(96.5)
High density lipids (mg/dL)	49.9	(14.1)	49.2	(14.5)	49.6	(14.3)

Page 1 of 1 170 words Focus 100%

Table of one regression model

```
. collect label list result, all
```

```
Collection: Table  
Dimension: result  
Label: Result
```

```
Level labels:
```

```
  N Number of observations  
N_cdf Number of completely determined failures  
N_cds Number of completely determined successes  
_r_b Coefficient  
_r_ci __LEVEL__ % CI  
_r_df df  
_r_lb __LEVEL__ % lower bound  
_r_p p-value  
_r_se Std. error  
_r_ub __LEVEL__ % upper bound  
_r_z z
```

```
collect clear  
collect _r_b _r_ci _r_p : logistic highbp c.age##i.sex i.diabetes  
collect layout (colname)(result[_r_b _r_ci _r_p])  
collect style cell result[_r_b _r_ci], nformat(%5.2f)  
collect style cell result[_r_p], nformat(%5.4f)  
collect style cell result[_r_ci], sformat("[%s]") cidelimiter(,)
```

```
table () (command result), /// no row dimensions, column dimensions are command and result  
command(_r_b _r_ci _r_p: logistic highbp c.age##i.sex i.diabetes) /// pull the OR, CI and p-value  
nformat(%5.2f _r_b _r_ci ) /// formatting 2 digits to the right  
nformat(%5.4f _r_p) /// formatting 4 digits to the right  
sformat("[%s]" _r_ci ) cidelimiter(, ) // ", " between LCI and UCI
```

```
collect label levels result _r_b "Odds Ratio", modify // changes label Coefficient to Odds Ratio  
collect label levels command 1 "Logistic Regression Model for Hypertension", modify  
collect style showbase off // hides base levels/reference groups of all factors  
collect style row stack, delimiter(" x ") nobinder // change interaction delimiter to x  
collect style cell border_block, border(right, pattern(nil))  
collect preview
```

TABLE OF A SINGLE REGRESSION

Hypertension in the United States

The National Health and Nutrition Examination Survey (NHANES)

Hypertension is a major cause of morbidity and mortality in the United States. This report will explore the predictors of hypertension using the NHANES dataset.

Table 4: Logistic Regression Model for Hypertension Status

	Logistic Regression Model for Hypertension		
	Odds Ratio	95% CI	p-value
Age (years)	1.03	[1.03, 1.04]	0.0000
Sex			
Female	0.15	[0.12, 0.21]	0.0000
Sex x Age (years)			
Female	1.03	[1.02, 1.03]	0.0000
Diabetes status			
Diabetic	1.52	[1.25, 1.86]	0.0000
Intercept	0.17	[0.14, 0.21]	0.0000

Table of multiple regression models

```
collect clear
collect create MyModels
collect _r_b _r_se, ///
        name(MyModels) ///
        tag(model[(1)]) ///
        : logistic highbp c.age i.sex
collect AIC=r(S)[1,"AIC"] ///
        BIC=r(S)[1,"BIC"], ///
        name(MyModels) ///
        tag(model[(1)]) ///
        : estat ic
collect _r_b _r_se, ///
        name(MyModels) ///
        tag(model[(2)]) ///
        : logistic highbp c.age##i.sex
...
collect layout (colname#result result[AIC BIC]) (model), name(MyModels)
collect style cell cell_type[item column-header], halign(center) // CENTER ITEMS AND COLUMN HEADERS
collect style header result, level(hide) // HIDE LABELS FOR ROW DIMENSION RESULT
// TURN OFF BASE LEVELS FOR FACTOR VARIABLES
collect style showbase off
// CHANGE THE INTERACTION DELIMITER
collect style row stack, spacer delimiter(" x ")
// REMOVE THE VERTICAL LINE
collect style cell border_block, border(right, pattern(nil))
// FORMAT THE NUMBERS
collect style cell, nformat(%5.2f)
collect style cell result[AIC BIC], nformat(%8.0f)
// PUT PARENTHESES AROUND THE STANDARD ERRORS
collect style cell result[_r_se], sformat("(%s)")
// LABEL AIC AND BIC
collect style header result[AIC BIC], level(label)
collect preview
```

* Use `tag(dimension(level))`

TABLE OF MULTIPLE REGRESSIONS

Table 3: Logistic Regression Models for Hypertension Status

	(1)	(2)	(3)
Age (years)	1.05 (0.00)	1.04 (0.00)	1.03 (0.00)
Female	0.65 (0.03)	0.16 (0.02)	0.15 (0.02)
Female x Age (years)		1.03 (0.00)	1.03 (0.00)
Diabetic			1.52 (0.15)
Intercept	0.09 (0.01)	0.17 (0.02)	0.17 (0.02)
AIC	12544	12434	12418
BIC	12566	12463	12454

Note: Odds ratio (standard error)

USEFUL FOR COMPARING MULTIPLE REGRESSION MODELS

Cox proportional hazards model for all-cause mortality Glucose, mg/dL

Incidence rate (per 1,000 person-years) of mortality by percentiles of glucose, mg/dL, 18-44 years old

	IR	95% CI
Glucose, percentile		
Glucose, mg/dL <50%	8.43	[3.95,17.98]
Glucose, mg/dL 50-<75%	31.91	[5.12,198.80]
Glucose, mg/dL 75-<90%	428.57	
Glucose, mg/dL >90%	107.82	[91.67,126.80]

Hazard Ratio (95%) of mortality by percentiles of Glucose, mg/dL, N=6757

	Model 1		Model 2	
	HR	95% CI	HR	95% CI
Glucose, percentile				
Glucose, mg/dL <50%	1.00		1.00	
Glucose, mg/dL 50-<75%	2.71	[0.58,12.59]	50.26	[0.09,26596.85]
Glucose, mg/dL 75-<90%	77.55	[6.24,964.06]	187.31	[0.09, 4.1e+05]
Glucose, mg/dL >90%	7.70	[1.55,38.28]	24.16	[0.02,23771.17]
Age (years)	1.07	[0.96, 1.20]	1.12	[0.94, 1.32]
Sex				
Male	1.00		1.00	
Female	1.48	[0.35, 6.27]	6.36	[0.11,364.74]
Race/Ethnicity				
NH White	1.00		1.00	
NH Black	2.32	[0.39,13.84]	2.59	[0.01,665.36]
Mex-Am	0.00	[0.00, 0.00]	0.00	[0.00, 0.00]
Other/Other Hisp	13.82	[2.20,86.82]	26.16	[0.04,15602.71]
Education - Recode (old version)				
Less than High School			1.00	
High School Diploma (including GED)			0.60	[0.05, 7.83]
More than High School			0.06	[0.00, 1.29]
Smoking status				
Current			1.00	
Former			17.29	[0.05,6314.71]
Never			2.60	[0.37,18.31]
Diagnosed diabetes				
No			1.00	
Yes			0.06	[0.00,64.10]
Hypertension				
No			1.00	
Yes			3.27	[0.68,15.80]

Model 1: Adjusted for age, sex, race

Model 2: Adjusted for Model 1 + education, smoking status, BMI categories, diabetes, and hypertension

SAVING AND USING CUSTOM STYLES

```
// TURN OFF BASE LEVELS FOR FACTOR VARIABLES
```

```
collect style showbase off
```

```
// STACK THE ROW LEVEL LABELS AND CHANGE THE INTERACTION DELIMITER
```

```
collect style row stack, delimiter(" x ") nobinder
```

```
// REMOVE THE VERTICAL LINE
```

```
collect style cell border_block, border(right, pattern(nil))
```

```
// FORMAT THE NUMBERS
```

```
collect style cell result[_r_b _r_ci], nformat(%8.2f)
```

```
collect style cell result[_r_p], nformat(%5.4f)
```

```
collect style cell result[_r_ci], sformat("[%s]") cidelimiter(,)
```

```
// HIDE THE LOGISTIC REGRESSION COMMAND
```

```
collect style header command, level(hide)
```

```
// SAVE THE STYLE
```

```
collect style save MyLogitStyle, replace
```

```
// USING THE STYLE
```

```
. table () (command result), command(_r_b _r_p _r_ci : logistic low c.age###i.smoke ht)
```

```
style(MyLogitStyle, override)
```

STATA has some predefined styles built in as well (eg. myreg.stjson, myregci .stjson)

USEFUL RESOURCES

STATA manual

- <https://www.stata.com/manuals/tables.pdf>
- <https://www.stata.com/bookstore/customizable-tables-and-collected-results-reference-manual/>

STATA blog (7 parts, very helpful)

- <https://blog.stata.com/category/reporting/>

YouTube videos

- <https://www.youtube.com/watch?v=OPyA8w26IGc>

QUESTIONS?

PLEASE FEEL FREE TO CONTACT ME:

NATALIE DAYA MALEK
NDAYAI@JH.EDU