Customizable tables with Stata

Kristin MacDonald

Executive Director of Statistical Services
StataCorp LLC

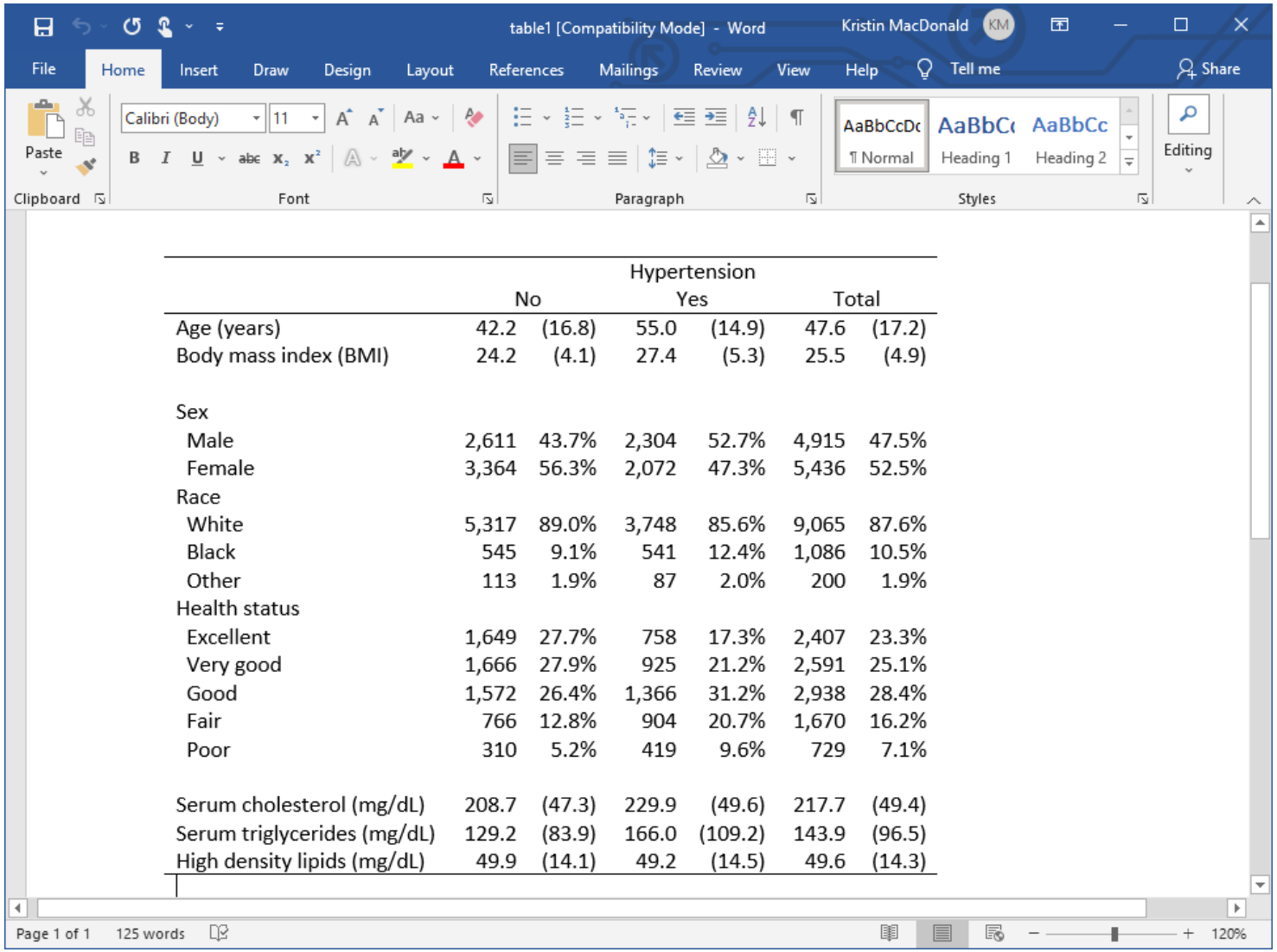
2021 Stata Conference

With Stata's features for customizable tables, you can . . .

Create a table of summary statistics.

	Hypertension					
	No		Yes		Total	
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)

Then export your table to a Word document.

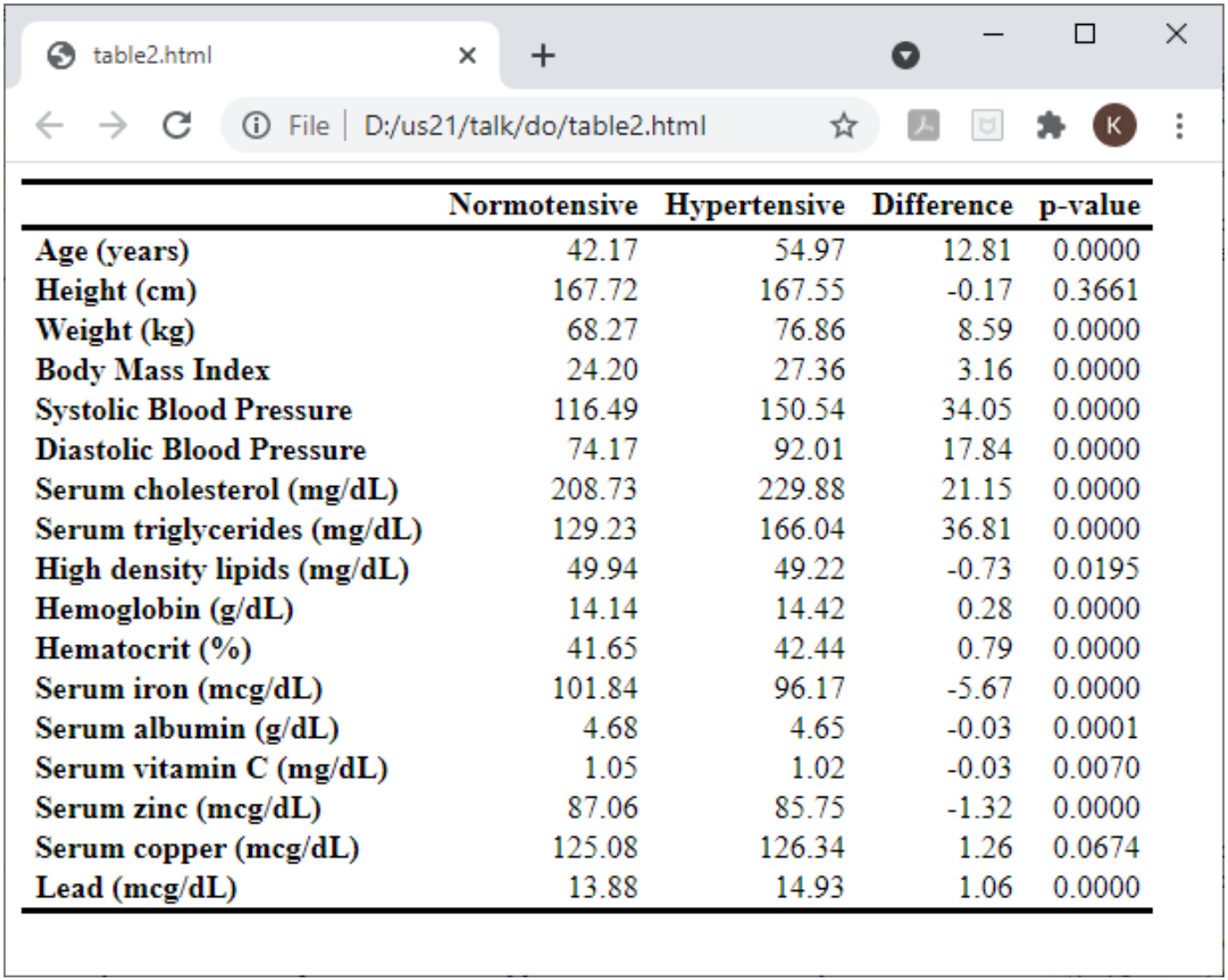


	Hypertension					
	No		Yes		Total	
Age (years)	42.2	(16.8)	55.0	(14.9)	47.6	(17.2)
Body mass index (BMI)	24.2	(4.1)	27.4	(5.3)	25.5	(4.9)
Sex						
Male	2,611	43.7%	2,304	52.7%	4,915	47.5%
Female	3,364	56.3%	2,072	47.3%	5,436	52.5%
Race						
White	5,317	89.0%	3,748	85.6%	9,065	87.6%
Black	545	9.1%	541	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)

Create a table of means and *t* tests of differences.

	Normotensive	Hypertensive	Difference	p-value
Age (years)	42.17	54.97	12.81	0.0000
Height (cm)	167.72	167.55	-0.17	0.3661
Weight (kg)	68.27	76.86	8.59	0.0000
Body Mass Index	24.20	27.36	3.16	0.0000
Systolic Blood Pressure	116.49	150.54	34.05	0.0000
Diastolic Blood Pressure	74.17	92.01	17.84	0.0000
Serum cholesterol (mg/dL)	208.73	229.88	21.15	0.0000
Serum triglycerides (mg/dL)	129.23	166.04	36.81	0.0000
High density lipids (mg/dL)	49.94	49.22	-0.73	0.0195
Hemoglobin (g/dL)	14.14	14.42	0.28	0.0000
Hematocrit (%)	41.65	42.44	0.79	0.0000
Serum iron (mcg/dL)	101.84	96.17	-5.67	0.0000
Serum albumin (g/dL)	4.68	4.65	-0.03	0.0001
Serum vitamin C (mg/dL)	1.05	1.02	-0.03	0.0070
Serum zinc (mcg/dL)	87.06	85.75	-1.32	0.0000
Serum copper (mcg/dL)	125.08	126.34	1.26	0.0674
Lead (mcg/dL)	13.88	14.93	1.06	0.0000

Then export your table to HTML.

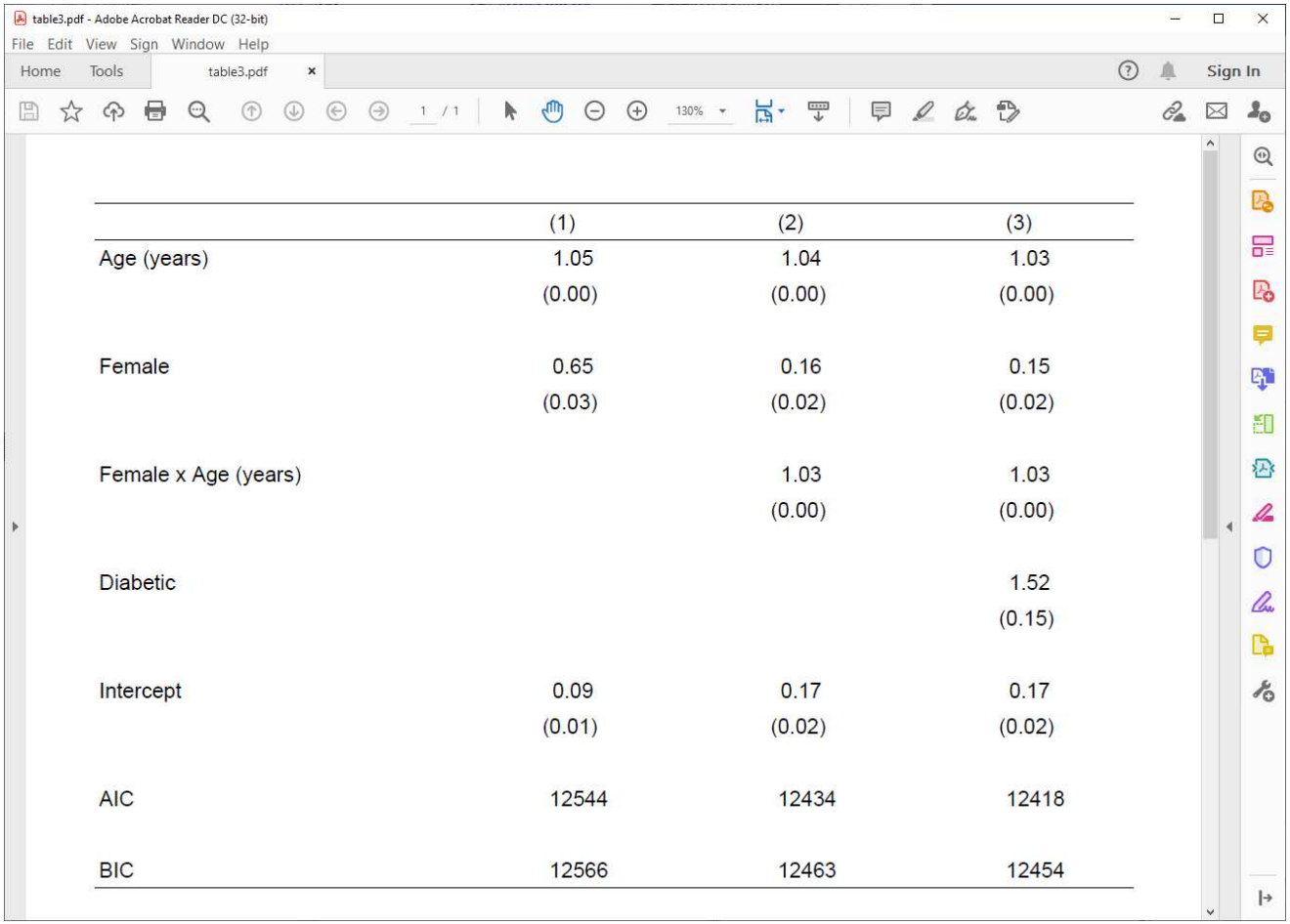


	Normotensive	Hypertensive	Difference	p-value
Age (years)	42.17	54.97	12.81	0.0000
Height (cm)	167.72	167.55	-0.17	0.3661
Weight (kg)	68.27	76.86	8.59	0.0000
Body Mass Index	24.20	27.36	3.16	0.0000
Systolic Blood Pressure	116.49	150.54	34.05	0.0000
Diastolic Blood Pressure	74.17	92.01	17.84	0.0000
Serum cholesterol (mg/dL)	208.73	229.88	21.15	0.0000
Serum triglycerides (mg/dL)	129.23	166.04	36.81	0.0000
High density lipids (mg/dL)	49.94	49.22	-0.73	0.0195
Hemoglobin (g/dL)	14.14	14.42	0.28	0.0000
Hematocrit (%)	41.65	42.44	0.79	0.0000
Serum iron (mcg/dL)	101.84	96.17	-5.67	0.0000
Serum albumin (g/dL)	4.68	4.65	-0.03	0.0001
Serum vitamin C (mg/dL)	1.05	1.02	-0.03	0.0070
Serum zinc (mcg/dL)	87.06	85.75	-1.32	0.0000
Serum copper (mcg/dL)	125.08	126.34	1.26	0.0674
Lead (mcg/dL)	13.88	14.93	1.06	0.0000

Create a table of regression results.

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

Then export your table to a PDF file.



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

You can create tables of

- Summary statistics, including a classic Table 1
- Results of classical hypothesis tests
- Regression results
- Postestimation tests
- Combinations of the above
- Results returned by any Stata commands

You can customize your table

- Table layout
- Numeric formats
- Labels appearing on rows and columns
- Stars and other added text
- Font type, size, and color
- Shading, borders, margins, alignment, and more

You can export your customized table to

- Word
- Excel
- LaTeX
- PDF
- Markdown
- HTML
- SMCL
- Plain text

You can also include the table in a report created by **putdocx**, **putexcel**, or **putpdf**.

Overview

- Introduction to the reimagined **table** command
- Introduction to the new **collect** suite of commands
- Examples
  - Table of summary statistics—Classic Table 1
  - Table of regression results

Introduction to the **table** command

To demonstrate, we will use NHANES II data.

```
. webuse nhanes2l, clear
(Second National Health and Nutrition Examination Survey)
. describe age sex race height weight bmi highbp
>          bpsystol bpdiast tcresult hdresult
Variable      Storage   Display    Value
    name         type    format    label      Variable label
------------------------------------------------------------------------------
age                 byte    %9.0g                 Age (years)
sex                 byte    %9.0g      sex        Sex
race                byte    %9.0g      race       Race
height              float   %9.0g                 Height (cm)
weight              float   %9.0g                 Weight (kg)
bmi                 float   %9.0g                 Body mass index (BMI)
highbp              byte    %8.0g                * High blood pressure
bpsystol            int     %9.0g                 Systolic blood pressure
bpdiast             int     %9.0g                 Diastolic blood pressure
tcresult            int     %9.0g                 Serum cholesterol (mg/dL)
hdresult            int     %9.0g                 High density lipids (mg/dL)
```

The **table** dialog box:

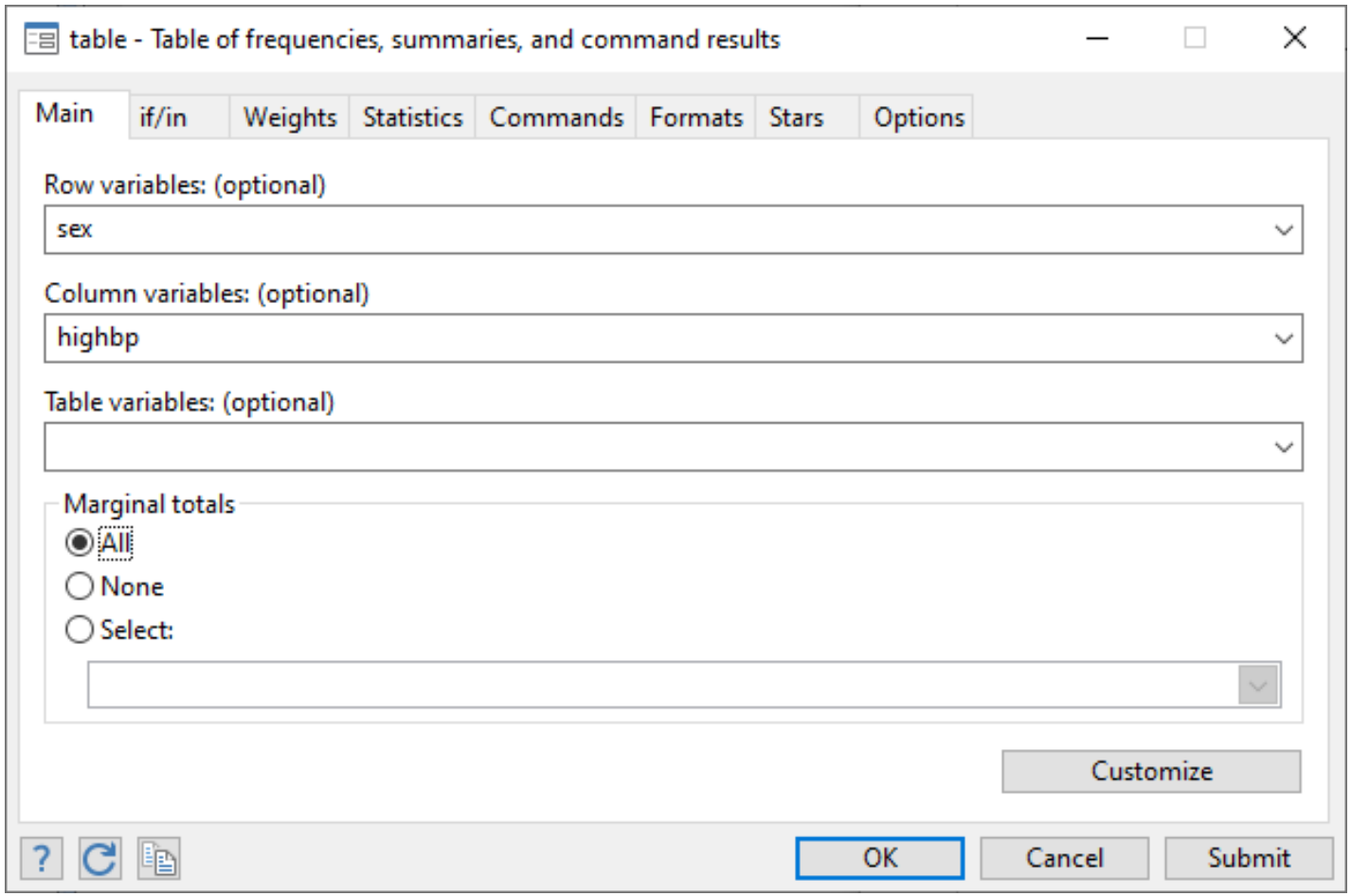


table command basics

Simplified **table** syntax:

. `table` (*row variables*) (*column variables*)

One-way tabulation

```
. table (highbp) ()
```

	Frequency
High blood pressure	
0	5,975
1	4,376
Total	10,351

One-way tabulation

```
. table () (highbp)
```

	High blood pressure		
	0	1	Total
Frequency	5,975	4,376	10,351

Two-way tabulation

```
. table (sex) (highbp)
```

	High blood pressure		
	0	1	Total
Sex			
Male	2,611	2,304	4,915
Female	3,364	2,072	5,436
Total	5,975	4,376	10,351

Two-way tabulation

```
. table (sex) (highbp), nototals
```

	High blood pressure	
	0	1
Sex		
Male	2,611	2,304
Female	3,364	2,072

Two-way tabulation

```
. table (sex) (highbp), totals(highbp)
```

	High blood pressure	
	0	1
Sex		
Male	2,611	2,304
Female	3,364	2,072
Total	5,975	4,376

```
. table (sex) (highbp), totals(sex)
```

	High blood pressure		
	0	1	Total
Sex			
Male	2,611	2,304	4,915
Female	3,364	2,072	5,436

Summary statistics

. table (*row variables*) (*column variables*),
statistic(*statspec*)

Summary statistics

statistic(*freqstat***)** requests frequency statistics.

`frequency`	frequency
`sumw`	sum of weights

Summary statistics

statistic(*sumstat varlist***)** requests summary statistics for the variables in *varlist*.

`mean`	mean
`semean`	standard error of the mean
`sebinomial`	standard error of the mean, binomial
`sepoisson`	standard error of the mean, Poisson
`variance`	variance
`sd`	standard deviation
`skewness`	skewness
`kurtosis`	kurtosis
`cv`	coefficient of variation

Summary statistics

`count`	number of nonmissing values
`median`	median
`p#`	#th percentile
`q1`	first quartile
`q2`	second quartile
`q3`	third quartile
`iqr`	interquartile range
`min`	minimum value
`max`	maximum value
`range`	range
...	first, last, total, factor-variable proportion, more

Summary statistics

statistic(*ratiostat* **[***varlist***])** requests ratio statistics.

`proportion`	proportion
`percent`	percentage
`rawproportion`	proportion ignoring weights
`rawpercent`	percentage ignoring weights

Summary statistics

```
. table () (highbp),
>       statistic(frequency)
>       statistic(percent)
```

	High blood pressure		
	0	1	Total
Frequency	5,975	4,376	10,351
Percent	57.72	42.28	100.00

Summary statistics

```
. table (sex) (highbp),
>       statistic(frequency)
>       statistic(percent)
```

	High blood pressure		
	0	1	Total
Sex			
Male			
Frequency	2,611	2,304	4,915
Percent	25.22	22.26	47.48
Female			
Frequency	3,364	2,072	5,436
Percent	32.50	20.02	52.52
Total			
Frequency	5,975	4,376	10,351
Percent	57.72	42.28	100.00

Summary statistics

```
. table (sex) (highbp),
>       statistic(frequency)
>       statistic(percent)
>       nototals
```

	High blood pressure	
	0	1
Sex		
Male		
Frequency	2,611	2,304
Percent	25.22	22.26
Female		
Frequency	3,364	2,072
Percent	32.50	20.02

Summary statistics

```
. table (sex) (highbp),
>       statistic(frequency)
>       statistic(percent)
>       statistic(mean age)
>       statistic(sd age)
>       nototals
```

	High blood pressure	
	0	1
Sex		
Male		
Frequency	2,611	2,304
Percent	25.22	22.26
Mean	42.8625	52.59288
Standard deviation	16.9688	15.88326
Female		
Frequency	3,364	2,072
Percent	32.50	20.02
Mean	41.62366	57.61921
Standard deviation	16.59921	13.25577

Formatting

. table (*row variables*) (*column variables*),
nformat() sformat()

```
. table (sex) (highbp),
>       statistic(frequency)
>       statistic(percent)
>       statistic(mean age)
>       statistic(sd age)
>       nototals
>       nformat(%9.0fc frequency)
>       sformat("%s%%" percent)
>       nformat(%6.2f mean sd)
>       sformat("(%s)" sd)
```

	High blood pressure	
	0	1
Sex		
Male		
Frequency	2,611	2,304
Percent	25.22%	22.26%
Mean	42.86	52.59
Standard deviation	(16.97)	(15.88)
Female		
Frequency	3,364	2,072
Percent	32.50%	20.02%
Mean	41.62	57.62
Standard deviation	(16.60)	(13.26)

Results from other commands

```
. table ( rowspec ) ( colspec ),
        command( cmdspec )
```

Results from other commands

```
. table (...) (...),
        command(correlate weight height age)
```

correlate stores the correlation matrix in **r(C)**.

```
. correlate weight height age
(obs=10,351)
```

	weight	height	age
weight	1.0000		
height	0.4775	1.0000	
age	0.0388	-0.2062	1.0000

```
. return list
scalars:
                  r(N) =  10351
                r(rho) =  .4774663496739867
matrices:
                  r(C) :  3 x 3
. matrix list r(C)
symmetric r(C)[3,3]
            weight      height         age
weight           1
height    .47746635           1
   age    .03881324  -.20616954           1
```

Results from other commands

```
. table (...) (...),
        command(r(C): correlate weight height age)
```

Row and column keywords

```
. table ( rowspec ) ( colspec ),
        command( cmdspec )
```

`result`	requested statistics
`var`	variables from statistic() option
`across`	index across() specifications
`colname`	column names for matrix statistics
`rowname`	row names for matrix statistics
`coleq`	column equation names for matrix statistics
`roweq`	row equation names for matrix statistics
`command`	index option command()
`statcmd`	index options statistic() and command()

```
. table (rowname) (colname),
>       command(r(C): correlate weight height age)
```

	Weight (kg)	Height (cm)	Age (years)
Weight (kg)	1	.4774663	.0388132
Height (cm)	.4774663	1	-.2061695
Age (years)	.0388132	-.2061695	1

Results from other commands

Specify row or column variables along with the **command()** option to compare results across groups.

```
. table () (sex), nformat(%6.2f)
>       command(regress bpsystol age weight)
```

	Sex		
	Male	Female	Total
Age (years)	0.48	0.77	0.64
Weight (kg)	0.33	0.46	0.41
Intercept	84.08	61.70	71.27

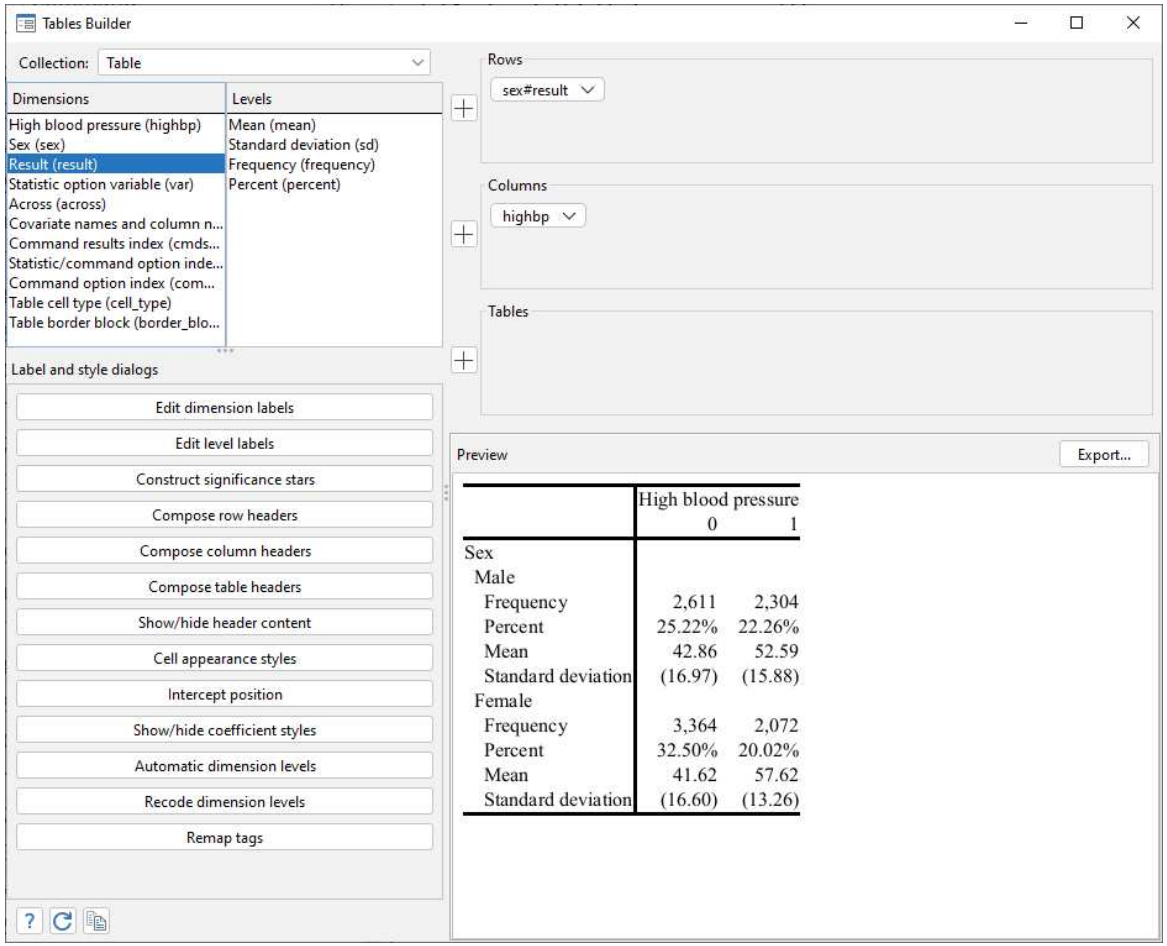
Introduction to the **collect** suite

Collection basics: Workflow

Basic workflow for collecting results and building tables:

- Collect results from Stata commands—**collect** command, **collect:** prefix, or **table**
- Explore the collection—**collect dims**, **collect levelsof**, and **collect label list**
- Define the rows and columns of the table—**collect layout** or **table**
- Customize your table, specifying formats, labels, font, shading, and more—**collect label**, **collect style**, **collect stars**, . . .
- Export your table to Word, Excel, LaTeX, PDF, Markdown, HTML, SMCL, or plain text—**collect export**
- Save your style, labels, and collection to use and modify later—**collect label save**, **collect style save**, **collect save**

The **Tables Builder**



Collection basics, step 1: Collect results

table automatically puts results into a collection.

```
. table (sex) (highbp),
>       statistic(frequency)
>       statistic(percent)
>       statistic(mean age)
>       statistic(sd age)
>       nototals
>       nformat(%9.0fc frequency)
>       sformat("%s%%" percent)
>       nformat(%6.2f mean sd)
>       sformat("(%s)" sd)
```

	High blood pressure	
	0	1
Sex		
Male		
Frequency	2,611	2,304
Percent	25.22%	22.26%
Mean	42.86	52.59
Standard deviation	(16.97)	(15.88)
Female		
Frequency	3,364	2,072
Percent	32.50%	20.02%
Mean	41.62	57.62
Standard deviation	(16.60)	(13.26)

Collection basics, step 2: Explore the collection

Values in a collection are organized according to their *tags*. A tag consists of a *dimension* and a *level* within the dimension. A tag is written as **dimension[level]**.

Value	Tag 1	Tag 2	Tag 3
2,611	`sex[1]`	`highbp[0]`	`result[frequency]`
2,304	`sex[1]`	`highbp[1]`	`result[frequency]`
25.22	`sex[1]`	`highbp[0]`	`result[percent]`

Our collection includes dimensions **sex**, **highbp**, and **result**, among others.

collect dims lists all dimensions in the collection.

```
. 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

Collection basics, step 2: Explore the collection

collect levelsof lists the levels of the specified dimension.
collect label list lists the levels and their associated labels.

```
. collect levelsof highbp
Collection: Table
 Dimension: highbp
    Levels: 0 1
. collect label list highbp, all
  Collection: Table
   Dimension: highbp
       Label: High blood pressure
Level labels:
          .m  Total
           0
           1
```

Collection basics, step 3: Specify the table layout

We specified the layout with **table**. Alternatively, we use **collect layout** to define the rows and columns.

```
. collect layout
Collection: Table
      Rows: sex#result
   Columns: highbp
   Table 1: 11 x 2
```

	High blood pressure	
	0	1
Sex		
Male		
Frequency	2,611	2,304
Percent	25.22%	22.26%
Mean	42.86	52.59
Standard deviation	(16.97)	(15.88)
Female		
Frequency	3,364	2,072
Percent	32.50%	20.02%
Mean	41.62	57.62
Standard deviation	(16.60)	(13.26)

Collection basics, step 4: Customize the table

collect label dim specifies the label for a dimension.
collect label levels specifies labels for the levels of a dimension.

```
. collect label dim highbp "Hypertension", modify
. collect label levels highbp 0 "No" 1 "Yes"
. collect label list highbp, all
  Collection: Table
   Dimension: highbp
       Label: Hypertension
Level labels:
          .m  Total
           0  No
           1  Yes
```

Collection basics, step 4: Customize the table

collect preview shows us our table with the new labels.

```
. collect preview
```

	Hypertension	
	No	Yes
Sex		
Male		
Frequency	2,611	2,304
Percent	25.22%	22.26%
Mean	42.86	52.59
Standard deviation	(16.97)	(15.88)
Female		
Frequency	3,364	2,072
Percent	32.50%	20.02%
Mean	41.62	57.62
Standard deviation	(16.60)	(13.26)

Collection basics, step 4: Customize the table

We can also modify the labels for our results.

```
. collect label list result
  Collection: Table
   Dimension: result
       Label: Result
Level labels:
   frequency  Frequency
        mean  Mean
     percent  Percent
          sd  Standard deviation
```

Collection basics, step 4: Customize the table

```
. collect label levels result frequency "Freq."
>                              mean      "Mean (age)"
>                              percent   "Percent"
>                              sd        "SD (age)",
>                              replace
. collect preview
```

	Hypertension	
	No	Yes
Sex		
Male		
Freq.	2,611	2,304
Percent	25.22%	22.26%
Mean (age)	42.86	52.59
SD (age)	(16.97)	(15.88)
Female		
Freq.	3,364	2,072
Percent	32.50%	20.02%
Mean (age)	41.62	57.62
SD (age)	(16.60)	(13.26)

Collection basics, step 4: Customize the table

Many customizations are performed via **collect style cell**. Here we remove the vertical border to the right of the row headers.

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

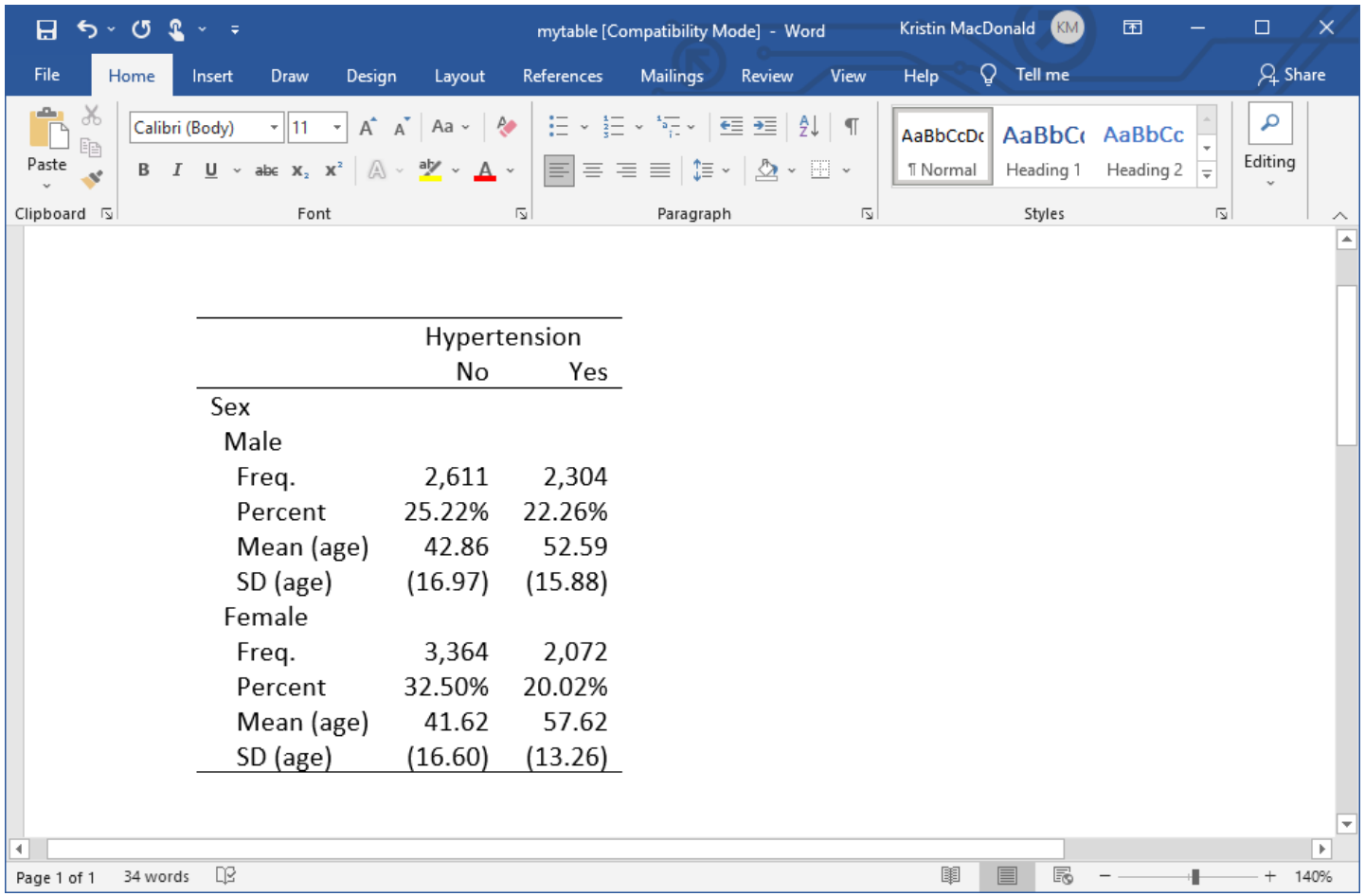
	Hypertension	
	No	Yes
Sex		
Male		
Freq.	2,611	2,304
Percent	25.22%	22.26%
Mean (age)	42.86	52.59
SD (age)	(16.97)	(15.88)
Female		
Freq.	3,364	2,072
Percent	32.50%	20.02%
Mean (age)	41.62	57.62
SD (age)	(16.60)	(13.26)

Collection basics, step 5: Export the table

collect export exports the customized table to our chosen format. When exporting to Word, **collect style putdocx** allows additional customizations specifically for this format.

```
. collect style putdocx, layout(autofitcontents)
. collect export mytable.docx, replace
(collection Table exported to file mytable.docx)
```

Collection basics, step 5: Export the table



	Hypertension	
	No	Yes
Sex		
Male		
Freq.	2,611	2,304
Percent	25.22%	22.26%
Mean (age)	42.86	52.59
SD (age)	(16.97)	(15.88)
Female		
Freq.	3,364	2,072
Percent	32.50%	20.02%
Mean (age)	41.62	57.62
SD (age)	(16.60)	(13.26)

Examples

- Table of summary statistics—Classic Table 1
- Table of regression results

Example 1: Classic Table 1

Table 1: Descriptive statistics by hypertensive status

	Hypertension					
	No		Yes		Total	
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)

Create a simple table using **table** with one categorical variable and one continuous variable.

```
. table (var) (highbp),
>       statistic(fvfrequency sex)
>       statistic(fvpercent sex)
>       statistic(mean age)
>       statistic(sd age) nototals
```

	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

Begin our customizations by changing labels.

```
. collect label dim highbp "Hypertension", modify
. collect label levels highbp 0 "No" 1 "Yes"
. collect preview
```

	Hypertension	
	No	Yes
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

Modify the levels of the **result** dimension by using **collect recode**.

```
. collect recode result fvfrequency = column1
>                      fvpercent   = column2
>                      mean        = column1
>                      sd          = column2
(12 items recoded in collection Table)
```

Request that the new levels appear as columns in the table.

```
. collect layout (var) (highbp#result[column1 column2])
Collection: Table
      Rows: var
   Columns: highbp#result[column1 column2]
   Table 1: 3 x 4
```

	Hypertension			
	No		Yes	
	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

Hide the labels of the **result** dimension so that **column1** and **column2** are not displayed.

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

	Hypertension			
	No		Yes	
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

Modify the row labels so that the levels of **sex** appear below the dimension label and we add extra vertical space.

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

	Hypertension			
	No		Yes	
Sex				
Male	2611	43.69874	2304	52.65082
Female	3364	56.30126	2072	47.34918
Age (years)	42.16502	16.77157	54.97281	14.90897

Remove the border that appeared to the right of the row labels.

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

	Hypertension			
	No		Yes	
Sex				
Male	2611	43.69874	2304	52.65082
Female	3364	56.30126	2072	47.34918
Age (years)	42.16502	16.77157	54.97281	14.90897

Specify a numeric format for frequencies of **sex**.

```
. collect style cell var[sex]#result[column1], nformat(%6.0fc)
. collect preview
```

	Hypertension			
	No		Yes	
Sex				
Male	2,611	43.69874	2,304	52.65082
Female	3,364	56.30126	2,072	47.34918
Age (years)	42.16502	16.77157	54.97281	14.90897

Specify a numeric format and add % to the percentages.

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

	Hypertension			
	No		Yes	
Sex				
Male	2,611	43.7%	2,304	52.7%
Female	3,364	56.3%	2,072	47.3%
Age (years)	42.16502	16.77157	54.97281	14.90897

Specify a numeric format for the means and standard deviations of **age**.

```
. collect style cell var[age]#result[column1 column2],
>         nformat(%6.1f)
. collect preview
```

	Hypertension			
	No		Yes	
Sex				
Male	2,611	43.7%	2,304	52.7%
Female	3,364	56.3%	2,072	47.3%
Age (years)	42.2	16.8	55.0	14.9

Add parentheses around the standard deviations of **age**.

```
. collect style cell var[age]#result[column2],
>         sformat("(%s)")
. collect preview
```

	Hypertension			
	No		Yes	
Sex				
Male	2,611	43.7%	2,304	52.7%
Female	3,364	56.3%	2,072	47.3%
Age (years)	42.2	(16.8)	55.0	(14.9)

Add more variables to our table.

```
. table (var) (highbp),
>       statistic(mean age bmi)
>       statistic(sd   age bmi)
>       statistic(fvfrequency sex race hlthstat)
>       statistic(fvpercent   sex race hlthstat)
>       statistic(mean tcresult tgresult hdresult)
>       statistic(sd   tcresult tgresult hdresult)
```

	High blood pressure		
	0	1	Total
Age (years)			
Mean	42.16502	54.97281	47.57965
Standard deviation	16.77157	14.90897	17.21483
Body mass index (BMI)			
Mean	24.20231	27.36081	25.5376
Standard deviation	4.100279	5.332119	4.914969
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
(output omitted)			

```
. collect label dim highbp "Hypertension", modify
. collect label levels highbp 0 "No" 1 "Yes"
. collect recode result fvfrequency = column1
                        fvpercent   = column2
                        mean        = column1
                        sd          = column2
. collect layout (var) (highbp#result[column1 column2])
. collect style header result, level(hide)
. collect style row stack, nobinder spacer
. collect style cell border_block, border(right, pattern(nil))
. collect style cell var[sex race hlthstat]#result[column1],
     nformat(%6.0fc)
. 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)")
```

	Hypertension					
	No		Yes		Total	
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)

Example 2: Table of regression results

We want to create a table reporting odds ratios and standard errors from logistic regression models.

```
. logistic highbp c.age i.sex
Logistic regression                                     Number of obs =  10,351
                                                        LR chi2(2)    = 1563.54
                                                        Prob > chi2   =  0.0000
Log likelihood = -6268.9975                             Pseudo R2     =  0.1109
```

highbp	Odds ratio	Std. err.	z	P>|z|	[95% conf.	interval]
age	1.049042	.0013945	36.02	0.000	1.046313	1.051779
sex						
Female	.648767	.0280172	-10.02	0.000	.5961141	.7060706
_cons	.0887874	.0063561	-33.83	0.000	.0771641	.1021615

Note: _cons estimates baseline odds.

We also want to include the AIC and BIC for each model.

```
. estat ic
Akaike´s information criterion and Bayesian information criterion
```

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	10,351	-7050.765	-6268.998	3	12544	12565.73

```
Note: BIC uses N = number of observations. See [R] BIC note.
. return list
matrices:
                  r(S) :  1 x 6
. mat list r(S)
r(S)[1,6]
             N         ll0          ll          df         AIC         BIC
.        10351  -7050.7655  -6268.9975           3   12543.995    12565.73
. display r(S)[1,"AIC"]
12543.995
```

We create a new collection **MyModels** and store results from the our logistic regression, tagging the results with dimension **model** and level **(1)**.

```
. collect clear
. collect create MyModels
(current collection is MyModels)
. collect _r_b _r_se, tag(model[(1)]) : logistic highbp c.age i.sex
Logistic regression                                     Number of obs =  10,351
                                                        LR chi2(2)    = 1563.54
                                                        Prob > chi2   =  0.0000
Log likelihood = -6268.9975                             Pseudo R2     =  0.1109
```

highbp	Odds ratio	Std. err.	z	P>|z|	[95% conf.	interval]
age	1.049042	.0013945	36.02	0.000	1.046313	1.051779
sex						
Female	.648767	.0280172	-10.02	0.000	.5961141	.7060706
_cons	.0887874	.0063561	-33.83	0.000	.0771641	.1021615

Note: _cons estimates baseline odds.

We add the results from **estat ic** to the collection, also tagging them results dimension **model** and level **(1)**.

```
. collect AIC=r(S)[1,"AIC"] BIC=r(S)[1,"BIC"], tag(model[(1)]) : estat ic
Akaike´s information criterion and Bayesian information criterion
```

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	10,351	-7050.765	-6268.998	3	12544	12565.73

Note: BIC uses N = number of observations. See [R] BIC note.

In total, we collect results from three models.

```
. collect create MyModels

. collect _r_b _r_se, tag(model[(1)]):
                    logistic highbp c.age i.sex
. collect AIC=r(S)[1,5] BIC=r(S)[1,6], tag(model[(1)]): estat ic

. collect _r_b _r_se, tag(model[(2)]):
                    logistic highbp c.age##i.sex
. collect AIC=r(S)[1,5] BIC=r(S)[1,6], tag(model[(2)]): estat ic

. collect _r_b _r_se, tag(model[(3)]):
                    logistic highbp c.age##i.sex i.diabetes
. collect AIC=r(S)[1,5] BIC=r(S)[1,6], tag(model[(3)]): estat ic
```

Specify the layout with covariates (**colname**) and statistics (**result**) on the rows and our **model** dimension on the columns.

```
. collect layout (colname#result) (model)
Collection: MyModels
      Rows: colname#result
   Columns: model
   Table 1: 24 x 3
```

	(1)	(2)	(3)
Age (years)			
Coefficient	1.049042	1.035184	1.034281
Std. error	.0013945	.0018459	.0018566
Male			
Coefficient	1	1	1
Std. error	0	0	0
Female			
Coefficient	.648767	.1556985	.1549363
Std. error	.0280172	.0224504	.0223461
Male # Age (years)			
Coefficient		1	1
Std. error		0	0
(output omitted)			

Omit the base levels from the table.

```
. collect style showbase off
. collect preview
```

	(1)	(2)	(3)
Age (years)			
Coefficient	1.049042	1.035184	1.034281
Std. error	.0013945	.0018459	.0018566
Female			
Coefficient	.648767	.1556985	.1549363
Std. error	.0280172	.0224504	.0223461
Female # Age (years)			
Coefficient		1.028811	1.028856
Std. error		.002794	.0027958
Diabetic			
Coefficient			1.521011
Std. error			.154103
Intercept			
Coefficient	.0887874	.1690035	.1730928
Std. error	.0063561	.0153794	.0157789

Remove the border from the right of the row labels.

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

	(1)	(2)	(3)
Age (years)			
Coefficient	1.049042	1.035184	1.034281
Std. error	.0013945	.0018459	.0018566
Female			
Coefficient	.648767	.1556985	.1549363
Std. error	.0280172	.0224504	.0223461
Female # Age (years)			
Coefficient		1.028811	1.028856
Std. error		.002794	.0027958
Diabetic			
Coefficient			1.521011
Std. error			.154103
Intercept			
Coefficient	.0887874	.1690035	.1730928
Std. error	.0063561	.0153794	.0157789

Specify a numeric format for all of our results.

```
. collect style cell, nformat(%5.2f)
. collect preview
```

	(1)	(2)	(3)
Age (years)			
Coefficient	1.05	1.04	1.03
Std. error	0.00	0.00	0.00
Female			
Coefficient	0.65	0.16	0.15
Std. error	0.03	0.02	0.02
Female # Age (years)			
Coefficient		1.03	1.03
Std. error		0.00	0.00
Diabetic			
Coefficient			1.52
Std. error			0.15
Intercept			
Coefficient	0.09	0.17	0.17
Std. error	0.01	0.02	0.02

We add parentheses around the standard errors.

```
. collect style cell result[_r_se], sformat("(%s)")
. collect preview
```

	(1)	(2)	(3)
Age (years)			
Coefficient	1.05	1.04	1.03
Std. error	(0.00)	(0.00)	(0.00)
Female			
Coefficient	0.65	0.16	0.15
Std. error	(0.03)	(0.02)	(0.02)
Female # Age (years)			
Coefficient		1.03	1.03
Std. error		(0.00)	(0.00)
Diabetic			
Coefficient			1.52
Std. error			(0.15)
Intercept			
Coefficient	0.09	0.17	0.17
Std. error	(0.01)	(0.02)	(0.02)

Request that the column headers and results be centered.

```
. collect style cell cell_type[item column-header], halign(center)
. collect preview
```

	(1)	(2)	(3)
Age (years)			
Coefficient	1.05	1.04	1.03
Std. error	(0.00)	(0.00)	(0.00)
Female			
Coefficient	0.65	0.16	0.15
Std. error	(0.03)	(0.02)	(0.02)
Female # Age (years)			
Coefficient		1.03	1.03
Std. error		(0.00)	(0.00)
Diabetic			
Coefficient			1.52
Std. error			(0.15)
Intercept			
Coefficient	0.09	0.17	0.17
Std. error	(0.01)	(0.02)	(0.02)

Hide the labels for the statistics.

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

	(1)	(2)	(3)
Age (years)	1.05	1.04	1.03
	(0.00)	(0.00)	(0.00)
Female	0.65	0.16	0.15
	(0.03)	(0.02)	(0.02)
Female # Age (years)		1.03	1.03
		(0.00)	(0.00)
Diabetic			1.52
			(0.15)
Intercept	0.09	0.17	0.17
	(0.01)	(0.02)	(0.02)

Add extra space between columns.

```
. collect style column, extraspace(1)
. collect preview
```

	(1)	(2)	(3)
Age (years)	1.05	1.04	1.03
	(0.00)	(0.00)	(0.00)
Female	0.65	0.16	0.15
	(0.03)	(0.02)	(0.02)
Female # Age (years)		1.03	1.03
		(0.00)	(0.00)
Diabetic			1.52
			(0.15)
Intercept	0.09	0.17	0.17
	(0.01)	(0.02)	(0.02)

Add space between rows and use **x** as the delimiter for interactions.

```
. collect style row stack, spacer delimiter(" x ")
. collect preview
```

	(1)	(2)	(3)
Age (years)	1.05	1.04	1.03
	(0.00)	(0.00)	(0.00)
Female	0.65	0.16	0.15
	(0.03)	(0.02)	(0.02)
Female x Age (years)		1.03	1.03
		(0.00)	(0.00)
Diabetic			1.52
			(0.15)
Intercept	0.09	0.17	0.17
	(0.01)	(0.02)	(0.02)

Append the AIC and BIC to the bottom of each column.

```
. collect layout (colname#result result[AIC BIC]) (model)
Collection: MyModels
      Rows: colname#result result[AIC BIC]
   Columns: model
   Table 1: 16 x 3
```

	(1)	(2)	(3)
Age (years)	1.05	1.04	1.03
	(0.00)	(0.00)	(0.00)
Female	0.65	0.16	0.15
	(0.03)	(0.02)	(0.02)
Female x Age (years)		1.03	1.03
		(0.00)	(0.00)
Diabetic			1.52
			(0.15)
Intercept	0.09	0.17	0.17
	(0.01)	(0.02)	(0.02)
	12544.00	12434.34	12417.74
	12565.73	12463.32	12453.97

Add the labels for AIC and BIC.

```
. collect style header result[AIC BIC], level(label)
. collect preview
```

	(1)	(2)	(3)
Age (years)	1.05	1.04	1.03
	(0.00)	(0.00)	(0.00)
Female	0.65	0.16	0.15
	(0.03)	(0.02)	(0.02)
Female x Age (years)		1.03	1.03
		(0.00)	(0.00)
Diabetic			1.52
			(0.15)
Intercept	0.09	0.17	0.17
	(0.01)	(0.02)	(0.02)
AIC	12544.00	12434.34	12417.74
BIC	12565.73	12463.32	12453.97

Specify a different format for AIC and BIC.

```
. collect style cell result[AIC BIC], nformat(%8.0f)
. collect preview
```

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

Saving and using styles and labels

After customizing one table, we can easily apply all of the same customizations to similar tables we create in the future.

```
. collect style save MyRegStyle
. collect label save MyRegLabel
```

Now we can collect results from logistic regression models fit to a different dataset. We only need to specify the table layout and use the saved styles and labels.

```
. collect clear
. webuse lbw, clear

. collect _r_b _r_se, tag(model[(1)]):
                    logistic low age i.smoke
. collect AIC=r(S)[1,5] BIC=r(S)[1,6], tag(model[(1)]): estat ic

. collect _r_b _r_se, tag(model[(2)]):
                    logistic low age i.smoke c.age#i.smoke
. collect AIC=r(S)[1,5] BIC=r(S)[1,6], tag(model[(2)]): estat ic

. collect layout (colname#result result[AIC BIC]) (model)

. collect style use MyRegStyle, override
. collect label use MyRegLabel
```

```
. collect preview
```

	(1)	(2)
Age of mother	0.95	0.92
	(0.03)	(0.04)
Smoker	2.00	0.38
	(0.64)	(0.58)
Smoker x Age of mother		1.08
		(0.07)
Intercept	1.06	2.24
	(0.80)	(2.29)
AIC	233	234
BIC	243	247

Summary

- The **table** command can now easily create and format tabulations, tables of summary statistics, and tables of results from other Stata commands.
- The **collect** suite is allows for building even more complex tables as well as customizing those tables and exporting them to many formats.
- Customized tables can also be included in complete reports.
- Saving styles and labels allows you to easily apply your desired customizations to tables you create in the future.

Learn more

- **table**

  https://www.stata.com/manuals/rtableintro.pdf

- **collect**

  https://www.stata.com/manuals/tables.pdf