Creating customized tables in Stata

Gabriela Ortiz

Senior Applied Econometrician
StataCorp LLC

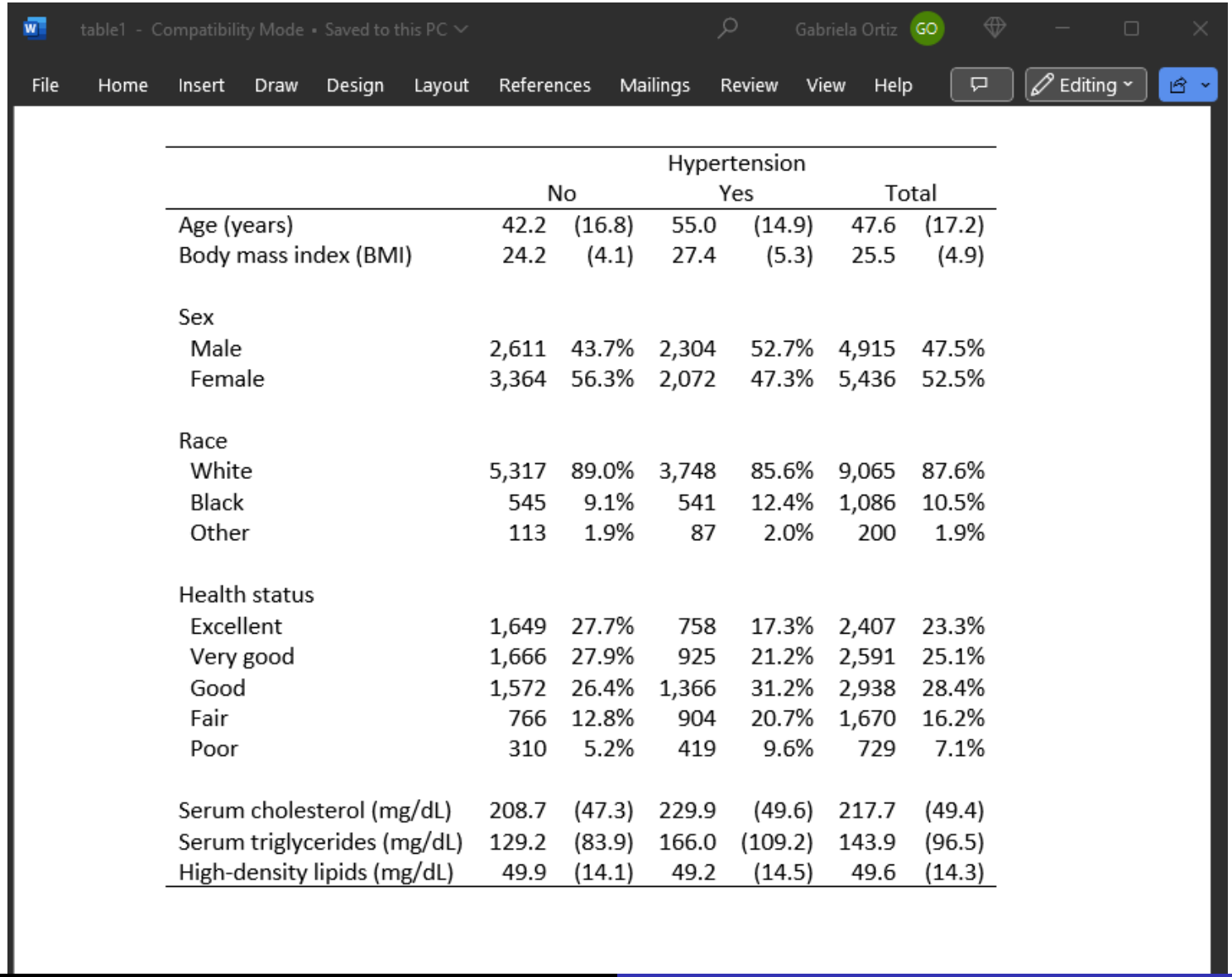
2023 Spanish Stata Conference

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
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

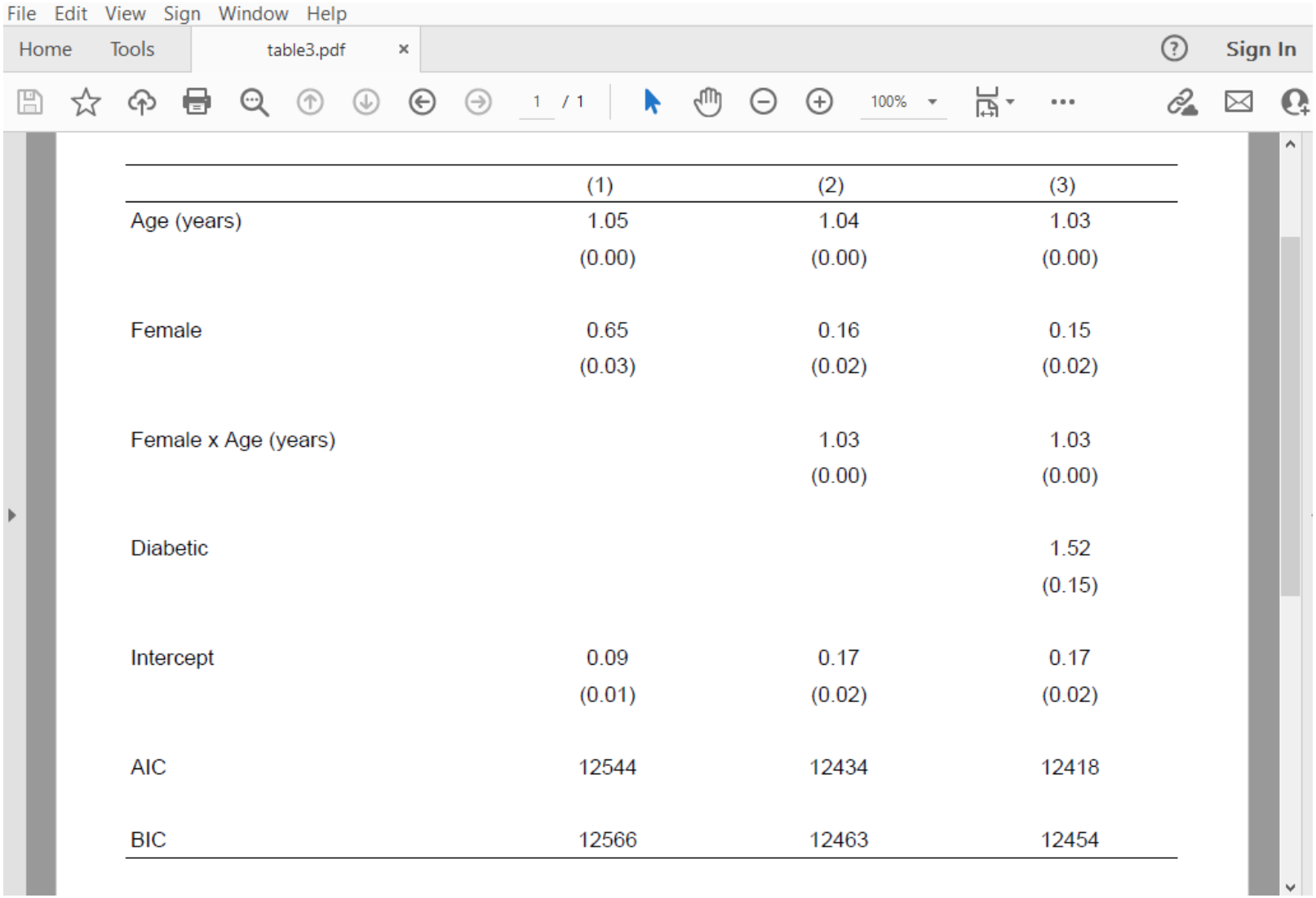
Then export your table to HTML.

	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
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 new **dtable** command
- Introduction to the **etable** command
- Introduction to the **table** command
- Introduction to the **collect** suite of commands

Introduction to the new **dtable** 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)
```

Table of descriptive statistics

```
. dtable age weight bpsystol i.sex i.race
```

	Summary
N	10,351
Age (years)	47.580 (17.215)
Weight (kg)	71.898 (15.356)
Systolic blood pressure	130.882 (23.333)
Sex	
Male	4,915 (47.5%)
Female	5,436 (52.5%)
Race	
White	9,065 (87.6%)
Black	1,086 (10.5%)
Other	200 (1.9%)

Statistics by groups

```
. dtable age weight bpsystol i.sex i.race,
>  by(diabetes)
```

	Diabetes status		
	Not diabetic	Diabetic	Total
N	9,850 (95.2%)	499 (4.8%)	10,349 (100.0%)
Age (years)	46.918 (17.193)	60.687 (11.475)	47.582 (17.216)
Weight (kg)	71.658 (15.220)	76.670 (17.175)	71.900 (15.357)
Systolic blood pressure	130.088 (22.759)	146.651 (28.387)	130.887 (23.332)
Sex			
Male	4,698 (47.7%)	217 (43.5%)	4,915 (47.5%)
Female	5,152 (52.3%)	282 (56.5%)	5,434 (52.5%)
Race			
White	8,659 (87.9%)	404 (81.0%)	9,063 (87.6%)
Black	1,000 (10.2%)	86 (17.2%)	1,086 (10.5%)
Other	191 (1.9%)	9 (1.8%)	200 (1.9%)

Tests of equality

```
. dtable age weight bpsystol i.sex i.race,
>  by(diabetes, nototals tests notestnotes)
>  continuous(age, test(none)) factor(race, test(none))
```

	Diabetes status		
	Not diabetic	Diabetic	Test
N	9,850 (95.2%)	499 (4.8%)	
Age (years)	46.918 (17.193)	60.687 (11.475)	
Weight (kg)	71.658 (15.220)	76.670 (17.175)	<0.001
Systolic blood pressure	130.088 (22.759)	146.651 (28.387)	<0.001
Sex			
Male	4,698 (47.7%)	217 (43.5%)	0.066
Female	5,152 (52.3%)	282 (56.5%)	
Race			
White	8,659 (87.9%)	404 (81.0%)	
Black	1,000 (10.2%)	86 (17.2%)	
Other	191 (1.9%)	9 (1.8%)	

Formatting and exporting the table

```
. dtable age weight bpsystol i.sex i.race,
>  by(diabetes, nototals tests notestnotes)
>  continuous(age, test(none)) factor(race, test(none))
>  sample(, statistics(freq) place(seplabels))
>  sformat("(N=%s)" frequency) note(Total sample: N = 10,349)
>  column(by(hide)) nformat(%7.2f mean sd)
>  title(Table 1. Demographics) export(table1.html, replace)
Table 1. Demographics
```

	Not diabetic (N=9,850)	Diabetic (N=499)	Test
Age (years)	46.92 (17.19)	60.69 (11.47)	
Weight (kg)	71.66 (15.22)	76.67 (17.18)	<0.001
Systolic blood pressure	130.09 (22.76)	146.65 (28.39)	<0.001
Sex			
Male	4,698 (47.7%)	217 (43.5%)	0.066
Female	5,152 (52.3%)	282 (56.5%)	
Race			
White	8,659 (87.9%)	404 (81.0%)	
Black	1,000 (10.2%)	86 (17.2%)	
Other	191 (1.9%)	9 (1.8%)	

```
Total sample: N = 10,349
(collection DTable exported to file table1.html)
```

The resulting HTML file:

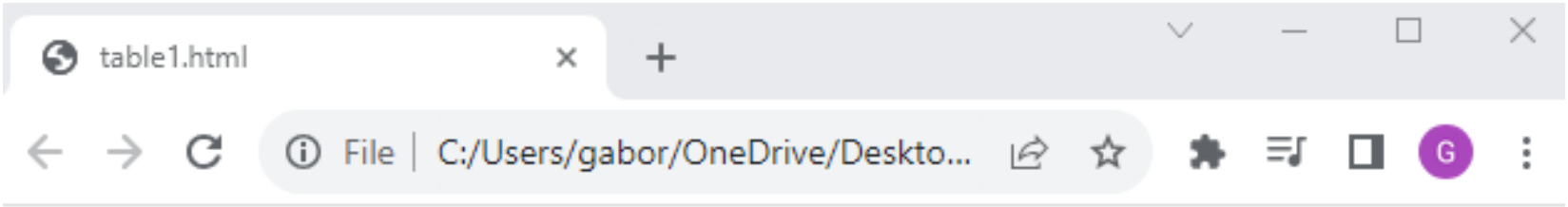


Table 1. Demographics

	Not diabetic (N=9,850)	Diabetic (N=499)	Test
Age (years)	46.92 (17.19)	60.69 (11.47)	
Weight (kg)	71.66 (15.22)	76.67 (17.18)	<0.001
Systolic blood pressure	130.09 (22.76)	146.65 (28.39)	<0.001
Sex			
Male	4,698 (47.7%)	217 (43.5%)	0.066
Female	5,152 (52.3%)	282 (56.5%)	
Race			
White	8,659 (87.9%)	404 (81.0%)	
Black	1,000 (10.2%)	86 (17.2%)	
Other	191 (1.9%)	9 (1.8%)	

Total sample: N = 10,349

Introduction to the **etable** command

Fitting a regression model

```
. quietly: logistic heartatk age bmi
. etable
```

	heartatk
Age (years)	1.088
	(0.005)
Body mass index (BMI)	1.013
	(0.010)
Intercept	0.000
	(0.000)
Number of observations	10349

Fitting a second model

```
. quietly: logistic heartatk age bmi bpsystol
. etable, append
```

	heartatk	heartatk
Age (years)	1.088	1.090
	(0.005)	(0.006)
Body mass index (BMI)	1.013	1.017
	(0.010)	(0.010)
Systolic blood pressure		0.997
		(0.002)
Intercept	0.000	0.000
	(0.000)	(0.000)
Number of observations	10349	10349

Customizing our estimation table

```
. etable, replay column(index) mstat(aic) mstat(bic)
>  showstars showstarsnote
```

	1		2	
Age (years)	1.088	**	1.090	**
	(0.005)		(0.006)	
Body mass index (BMI)	1.013		1.017	
	(0.010)		(0.010)	
Systolic blood pressure			0.997	
			(0.002)	
Intercept	0.000	**	0.000	**
	(0.000)		(0.000)	
AIC	3342.84		3342.77	
BIC	3364.57		3371.75	

** p<.01, * p<.05

Customizing our estimation table

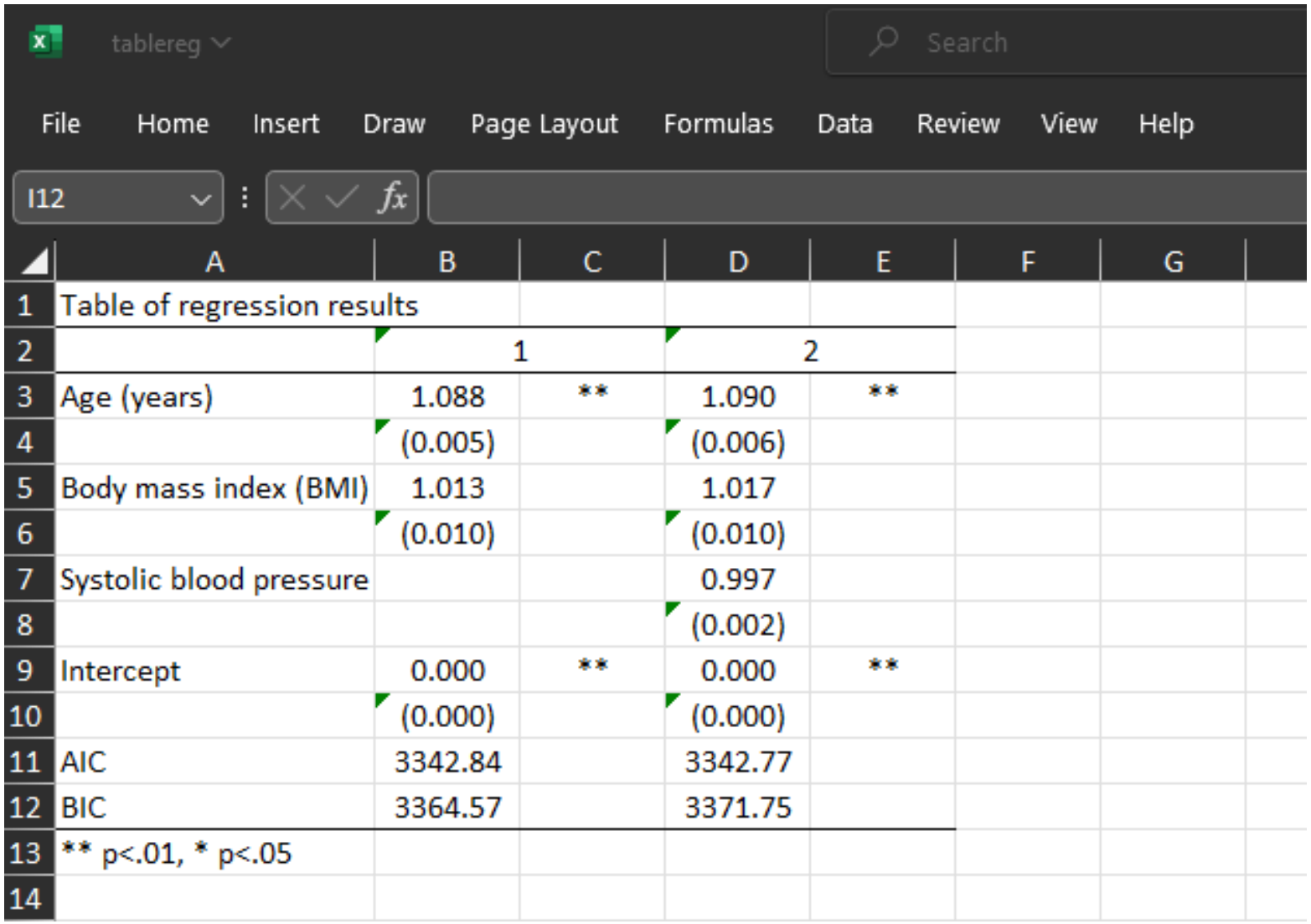
```
. etable, replay column(index) mstat(aic) mstat(bic)
>  showstars showstarsnote center
>  title(Table of regression results)
>  export(tablereg.xlsx)
```

Table of regression results

	1		2	
Age (years)	1.088	**	1.090	**
	(0.005)		(0.006)	
Body mass index (BMI)	1.013		1.017	
	(0.010)		(0.010)	
Systolic blood pressure			0.997	
			(0.002)	
Intercept	0.000	**	0.000	**
	(0.000)		(0.000)	
AIC	3342.84		3342.77	
BIC	3364.57		3371.75	

** p<.01, * p<.05

The resulting Excel file:



	A	B	C	D	E	F	G
1	Table of regression results						
2		1		2			
3	Age (years)	1.088	**	1.090	**		
4		(0.005)		(0.006)			
5	Body mass index (BMI)	1.013		1.017			
6		(0.010)		(0.010)			
7	Systolic blood pressure			0.997			
8				(0.002)			
9	Intercept	0.000	**	0.000	**		
10		(0.000)		(0.000)			
11	AIC	3342.84		3342.77			
12	BIC	3364.57		3371.75			
13	** p<.01, * p<.05						
14							

Introduction to the **table** command

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

Summary statistics

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

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)
```

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

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, **table**, **dtable**, or **etable**
- 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**

Collection basics, step 1: Collect results

```
. collect create reg
(current collection is reg)

. collect: quietly: logistic highbp c.age##sex
```

Collection basics, step 2: Specify the table layout

We use **collect layout** to define the rows and columns.

```
. collect layout (colname) (result[_r_b _r_ci])
Collection: reg
      Rows: colname
   Columns: result[_r_b _r_ci]
   Table 1: 6 x 2
```

	Coefficient	95% CI	
Age (years)	1.035184	1.031572	1.038808
Male	1		
Female	.1556985	.1173677	.2065477
Male # Age (years)	1		
Female # Age (years)	1.028811	1.02335	1.034302
Intercept	.1690035	.1413957	.2020018

Collection basics, step 3: Customize the table

Omit the base levels from the table.

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

	Coefficient	95% CI	
Age (years)	1.035184	1.031572	1.038808
Female	.1556985	.1173677	.2065477
Female # Age (years)	1.028811	1.02335	1.034302
Intercept	.1690035	.1413957	.2020018

Collection basics, step 3: Customize the table

Use **x** as the delimiter for interactions.

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

	Coefficient	95% CI	
Age (years)	1.035184	1.031572	1.038808
Female	.1556985	.1173677	.2065477
Female x Age (years)	1.028811	1.02335	1.034302
Intercept	.1690035	.1413957	.2020018

Collection basics, step 3: Customize the table

Remove the border from the right of the row labels.

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

	Coefficient	95% CI	
Age (years)	1.035184	1.031572	1.038808
Female	.1556985	.1173677	.2065477
Female x Age (years)	1.028811	1.02335	1.034302
Intercept	.1690035	.1413957	.2020018

Collection basics, step 3: Customize the table

Specify a numeric format for our results.

```
. collect style cell result, halign(center)
. collect style cell result[_r_b], nformat(%6.3f)
. collect style cell result[_r_ci], nformat(%5.2f)
>  cidelimiter(,) sformat("[%s]")
. collect preview
```

	Coefficient	95% CI
Age (years)	1.035	[1.03, 1.04]
Female	0.156	[0.12, 0.21]
Female x Age (years)	1.029	[1.02, 1.03]
Intercept	0.169	[0.14, 0.20]

Collection basics, step 3: Customize the table

We can add stars for significance.

```
. collect stars _r_p 0.01 "***" 0.05 "** " 0.1 "* " 1 " ",
>  attach(_r_b) shownote
```

Collection basics, step 3: Customize the table

We can also modify the labels for our results.

```
. collect label levels result _r_b "OR", modify
. collect preview
```

	OR	95% CI
Age (years)	1.035***	[1.03, 1.04]
Female	0.156***	[0.12, 0.21]
Female x Age (years)	1.029***	[1.02, 1.03]
Intercept	0.169***	[0.14, 0.20]

*** p<.01, ** p<.05, * p<.1, p<1

Saving 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 ORstyle
. collect label save ORlbl
```

Applying style and label files

Now we can collect results from another logistic regression model. We can then use the saved styles and labels.

```
. collect create reg2
. collect: quietly: logistic highbp age i.sex
. collect style use ORstyle, override
. collect label use ORlbl, modify
```

```
. collect preview
```

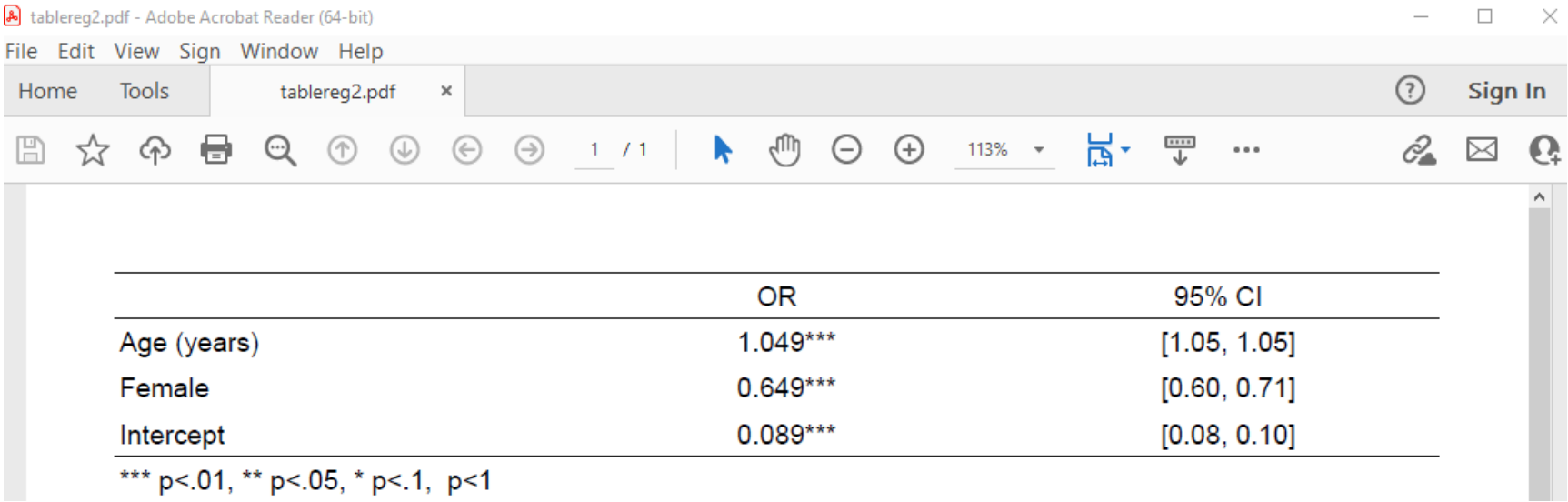
	OR	95% CI
Age (years)	1.049***	[1.05, 1.05]
Female	0.649***	[0.60, 0.71]
Intercept	0.089***	[0.08, 0.10]

*** p<.01, ** p<.05, * p<.1, p<1

Collection basics, step 4: Export the table

```
. collect export tablereg2.pdf
(collection reg2 exported to file tablereg2.pdf)
```

Here is the PDF file:



	OR	95% CI
Age (years)	1.049***	[1.05, 1.05]
Female	0.649***	[0.60, 0.71]
Intercept	0.089***	[0.08, 0.10]

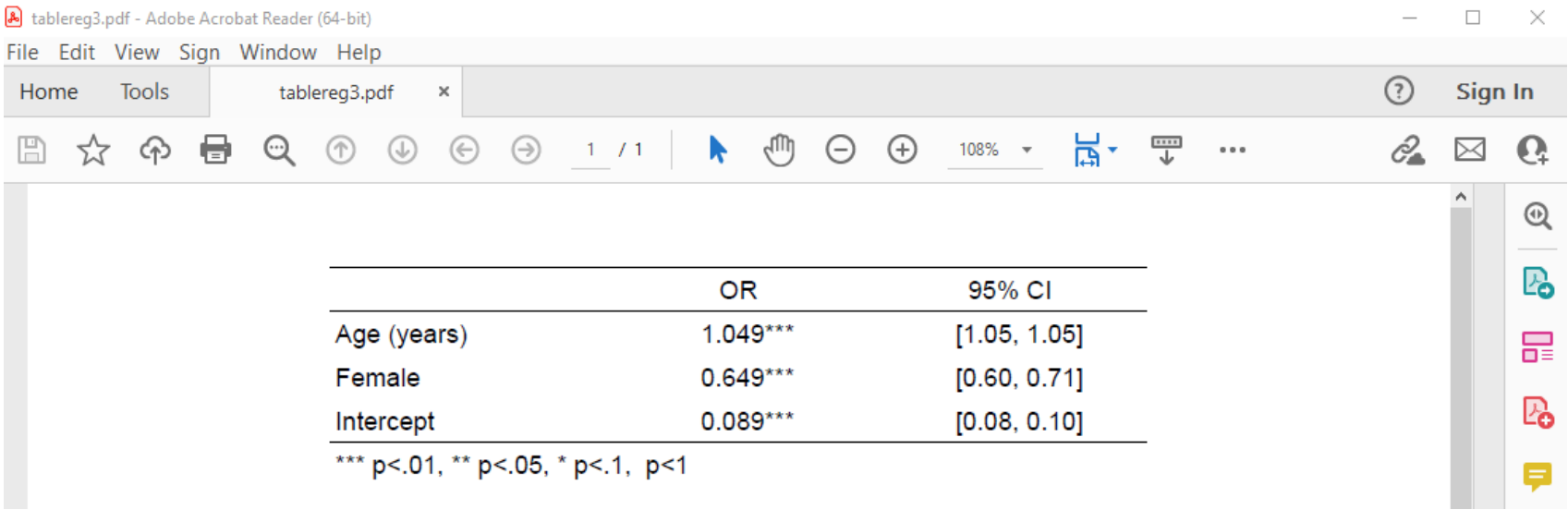
*** p<.01, ** p<.05, * p<.1, p<1

Collection basics, step 5: Specify the style for export

```
. collect style putpdf, halign(center) width(65%)

. collect export tablereg3.pdf, replace
(collection reg2 exported to file tablereg3.pdf)
```

Here is the updated PDF file:



	OR	95% CI
Age (years)	1.049***	[1.05, 1.05]
Female	0.649***	[0.60, 0.71]
Intercept	0.089***	[0.08, 0.10]

*** p<.01, ** p<.05, * p<.1, p<1

Summary

- The **dtable** command can easily create and export tables of descriptive statistics.
- The **etable** command can easily create and export tables of estimation results.
- The **table** command can easily create and format tabulations, tables of summary statistics, and tables of results from other Stata commands.
- The **collect** suite 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

- **dtable**

  https://www.stata.com/manuals/rdtable.pdf
- **etable**

  https://www.stata.com/manuals/retable.pdf
- **table**

  https://www.stata.com/manuals/rtableintro.pdf
- **collect**

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