Reproducible reporting- Creating customized reports in Word, Excel, and PDF

Gabriela Ortiz

StataCorp LLC

July 12, 2022

1

Run this do-file

```
use low, clear
putdocx begin
putdocx textblock begin, style(Title)
Analysis of birthweights
putdocx textblock end
putdocx textblock begin
We have data on birthweights from Hosmer,
Lemeshow, and Sturdivant (2013, 24).
putdocx textblock end
graph hbar bwt, over(ht) over(smoke) asyvars ///
        title(Baby birthweights)
graph export bweight.png, replace
putdocx image bweight.png, width(4)
regress bwt i.smoke age, noheader cformat(%9.2f)
putdocx table bweight = etable
putdocx save bwreport, replace
```

Run this do-file

```stata
use low, clear
putdocx begin
putdocx textblock begin, style(Title)
Analysis of birthweights
putdocx textblock end
putdocx textblock begin
We have data on birthweights from Hosmer,
Lemeshow, and Sturdivant (2013, 24).
putdocx textblock end
graph hbar bwt, over(ht) over(smoke) asyvars ///
        title(Baby birthweights)
graph export bweight.png, replace
putdocx image bweight.png, width(4)
regress bwt i.smoke age, noheader cformat(%9.2f)
putdocx table bweight = etable
putdocx save bwreport, replace
```

Obtain this Word document
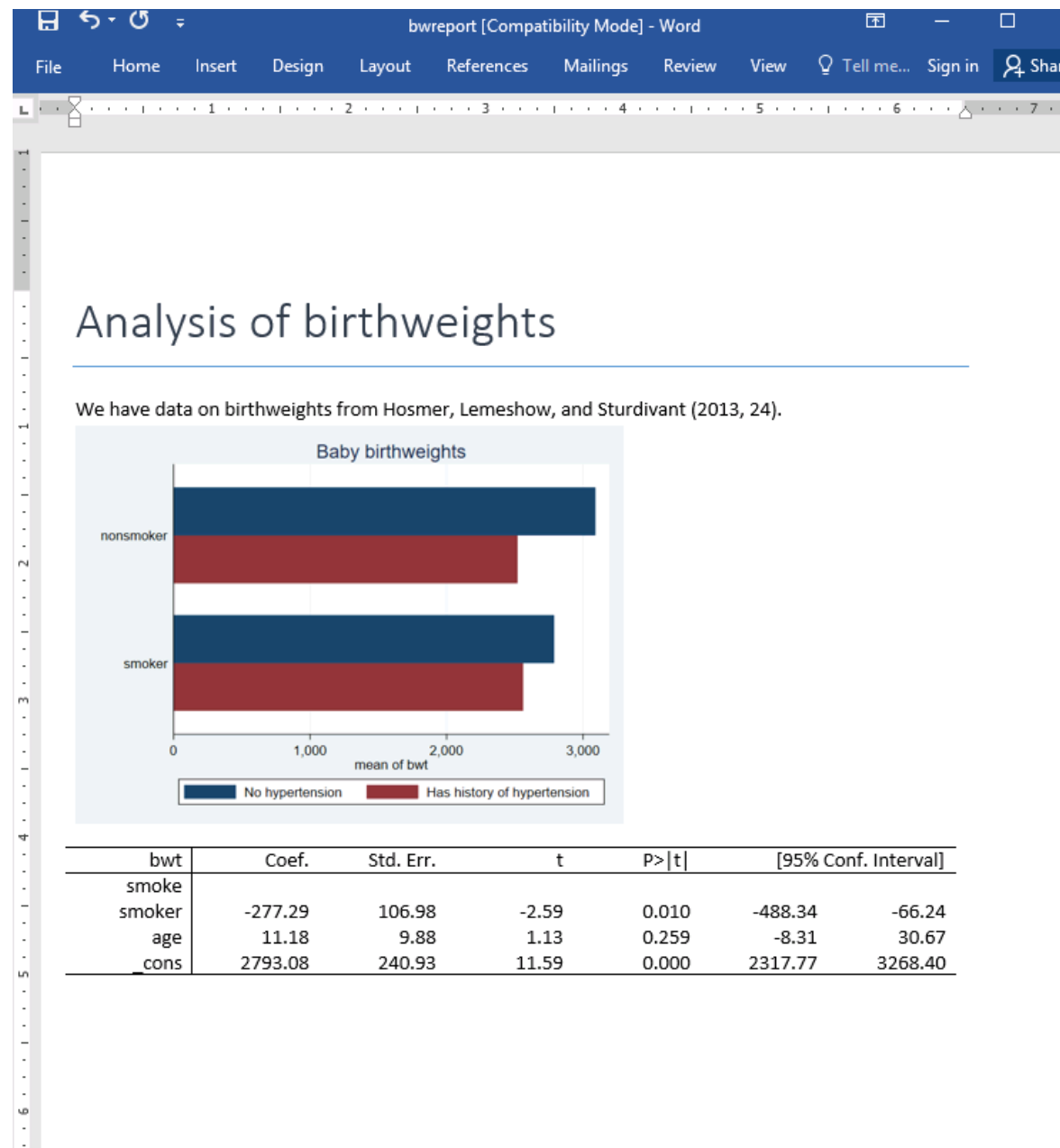
Run this do-file

```
use low, clear
putexcel set low.xlsx, replace
putexcel A1 = "Analysis of birthweights", bold
putexcel A3 = "We have data on birthweights from Hosmer, "
putexcel A4 = "Lemeshow, and Sturdivant(2013, 24). "
graph hbar bwt, over(ht) over(smoke) asyvars
graph export bweight.png, replace height(300)
putexcel A7 = image(bweight.png)
regress bwt i.smoke age, cformat(%9.2f)
putexcel A24 = etable
putexcel F24:G24, merge
putexcel save
```

Run this do-file

```
use low, clear
putexcel set low.xlsx, replace
putexcel A1 = "Analysis of birthweights", bold
putexcel A3 = "We have data on birthweights from Hosmer, "
putexcel A4 = "Lemeshow, and Sturdivant(2013, 24). "
graph hbar bwt, over(ht) over(smoke) asyvars
graph export bweight.png, replace height(300)
putexcel A7 = image(bweight.png)
regress bwt i.smoke age, cformat(%9.2f)
putexcel A24 = etable
putexcel F24:G24, merge
putexcel save
```

Obtain this Excel file
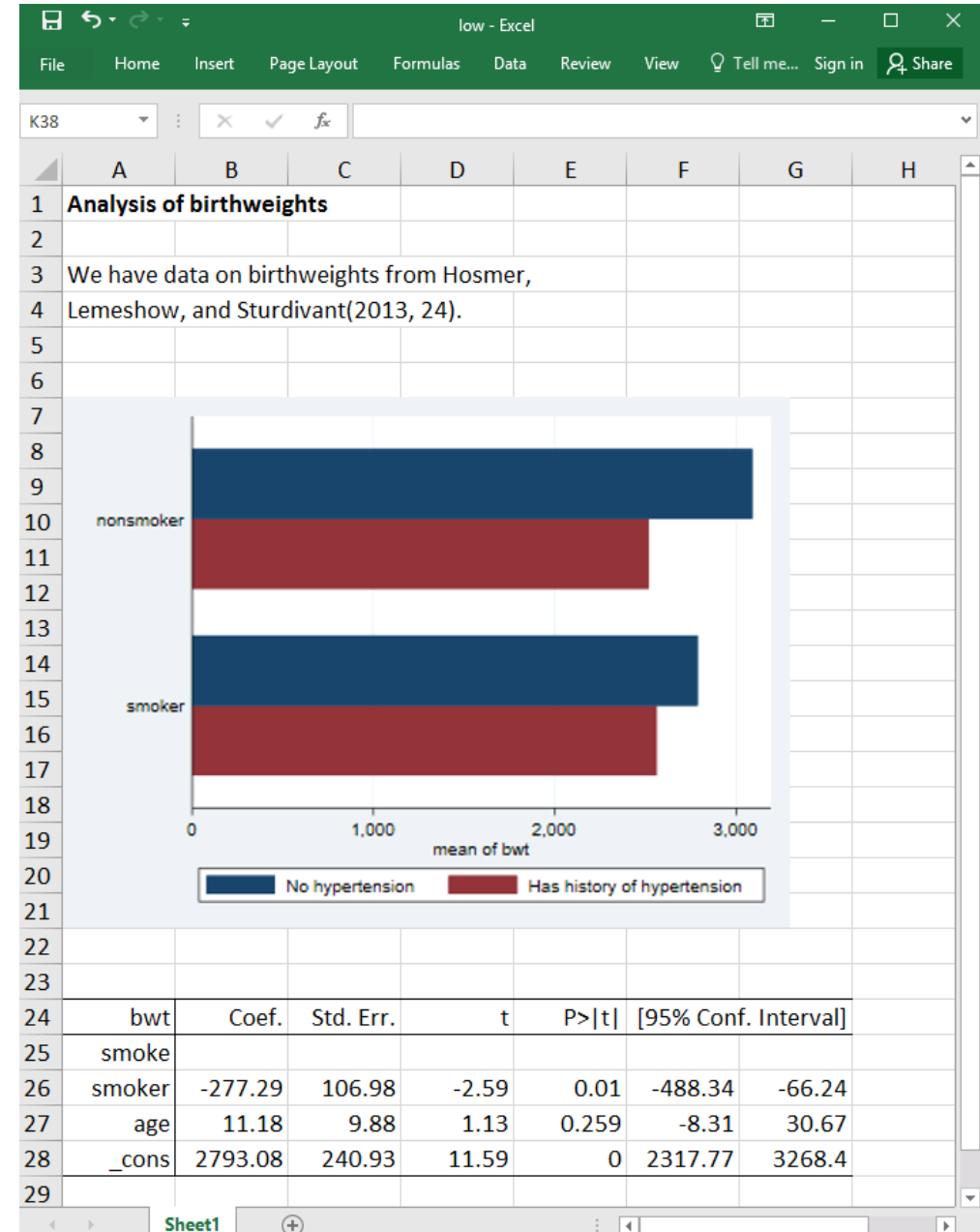
Overview

- Create reproducible reports with formatted results
  - Summary statistics, estimation results, and graphs
- Create reports in Word, Excel, and PDF format
  - **putdocx**, **putexcel,** and **docx2pdf**

Overview

- Create reproducible reports with formatted results
  - Summary statistics, estimation results, and graphs
- Create reports in Word, Excel, and PDF format
  - **`putdocx`**, **`putexcel,`** and **`docx2pdf`**
- Create a Word document with a header, footer, and page numbers
- Append multiple Word documents, and interact Stata's features with Word's features
- Create automated reports

Data on birthweights

```
Contains data from low.dta
  obs:           189                          Hosmer & Lemeshow data
  vars:           12                          15 Oct 2019 13:40
```

variable name	storage type	display format	value label	variable label
id	int	%8.0g		identification code
low	byte	%8.0g		birthweight<2500g
age	byte	%8.0g		age of mother
lwt	int	%8.0g		weight at last menstrual period
race	byte	%8.0g	race	race
smoke	byte	%9.0g	smoke	smoked during pregnancy
ptl	byte	%8.0g		premature labor history (count)
ht	byte	%27.0g	htlabel	has history of hypertension
ui	byte	%8.0g		presence, uterine irritability
ftv	byte	%8.0g		number of visits to physician during 1st trimester
bwt	int	%8.0g		birthweight (grams)
agegrp	float	%9.0g		

Analysis on birthweights

```
use low, clear

/* 1. How many observations in our dataset ? */
summarize bwt
```

Analysis on birthweights

```
use low, clear

/* 1. How many observations in our dataset ? */
summarize bwt
local total = r(N)
```

```
. summarize bwt
```

Variable	Obs	Mean	Std. Dev.	Min	Max
bwt	189	2944.286	729.016	709	4990

```
. return list

scalars:
            r(N) =  189
        r(sum_w) =  189
         r(mean) =  2944.285714285714
          r(Var) =  531464.3541033434
           r(sd) =  729.0160177275554
          r(min) =  709
          r(max) =  4990
          r(sum) =  556470
```

Analysis on birthweights

```
use low, clear

/* 1. How many observations in our dataset ? */
summarize bwt
local total = r(N)

/* 2. How many mothers in this dataset smoked during
pregnancy? */
count if smoke==1
```

```
. count if smoke==1
    74

. return list

scalars:
            r(N) =   74
```

Analysis on birthweights

```
use low, clear

/* 1. How many observations in our dataset ? */
summarize bwt
local total = r(N)

/* 2. How many mothers in this dataset smoked during
pregnancy? */
count if smoke==1

/* 3. How many didn't? */
display `total'-r(N)
```

```
. summarize bwt

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
         bwt |        189    2944.286     729.016        709       4990

. local total = r(N)

. count if smoke==1
   74

. display `total'-r(N)
115
```

Analysis on birthweights

```
use low, clear

/* 1. How many observations in our dataset ? */
summarize bwt
local total = r(N)

/* 2. How many mothers in this dataset smoked during
pregnancy? */
count if smoke==1

/* 3. How many didn't? */
display `total'-r(N)

/* 4. How does birthweight vary across mothers' smoking
habits while pregnant, and their history of hypertension
? */
graph hbar bwt, over(ht) over(smoke) asyvars
```
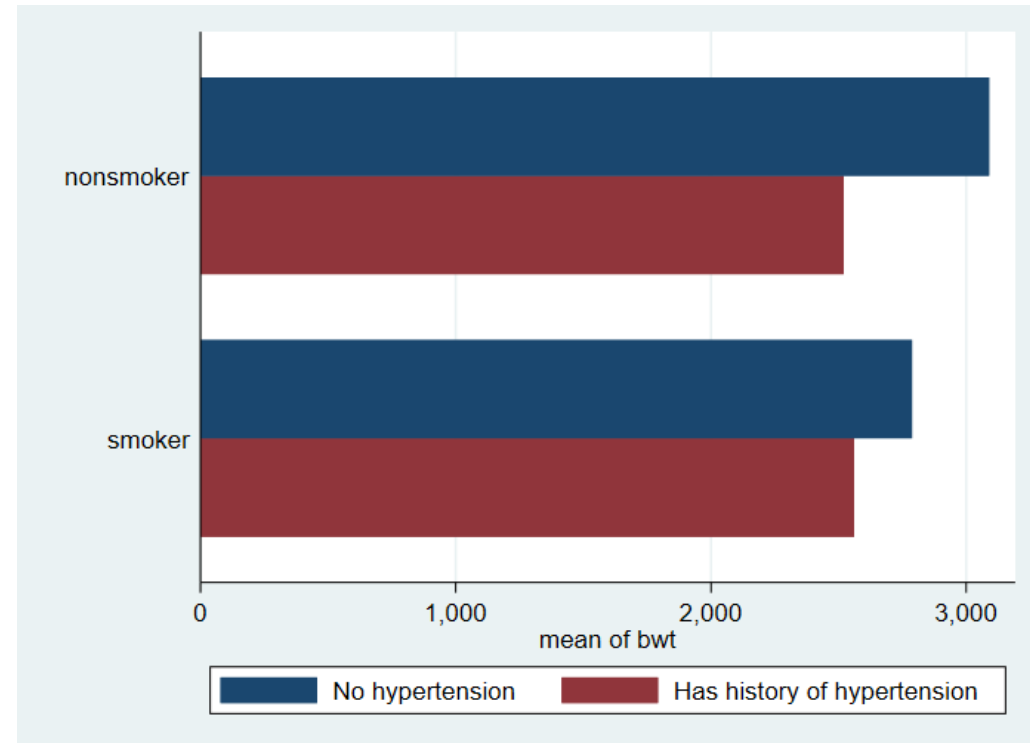
Analysis on birthweights

```
use low, clear

/* 1. How many observations in our dataset ? */
summarize bwt
local total = r(N)

/* 2. How many mothers in this dataset smoked during
pregnancy? */
count if smoke==1

/* 3. How many didn't? */
display `total'-r(N)

/* 4. How does birthweight vary across mothers' smoking
habits while pregnant, and their history of hypertension
? */
graph hbar bwt, over(ht) over(smoke) asyvars

/* 5. Model birthweight as a function of smoking and
maternal age */
regress bwt i.smoke age, cformat(%9.2f)
```

```
. use low, clear
(Hosmer & Lemeshow data)

. summarize bwt

    Variable |       Obs      Mean    Std. Dev.     Min      Max
-------------+---------------------------------------------------
         bwt |       189   2944.286    729.016      709     4990

. local total = r(N)

. count if smoke==1
    74

. display `total'-r(N)
115

. graph hbar bwt, over(ht) over(smoke) asyvars

. regress bwt i.smoke age, cformat(%9.2f)
```

Source	SS	df	MS
Model	4255739.01	2	2127869.51
Residual	95659559.6	186	514298.707
Total	99915298.6	188	531464.354

Number of obs	= 189
F(2, 186)	= 4.14
Prob > F	= 0.0175
R-squared	= 0.0426
Adj R-squared	= 0.0323
Root MSE	= 717.15

bwt	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
smoke						
smoker	-277.29	106.98	-2.59	0.010	-488.34	-66.24
age	11.18	9.88	1.13	0.259	-8.31	30.67
_cons	2793.08	240.93	11.59	0.000	2317.77	3268.40

Creating a Word document in Stata

```
putdocx begin
```

Creating a Word document in Stata

```
putdocx begin

putdocx paragraph
putdocx text (" Hello my webinar friends.")
putdocx text (" How are you doing today? ")
```

Creating a Word document in Stata

putdocx begin

putdocx textblock begin
Hello my webinar friends. How are you
doing today?
putdocx textblock end

Creating a Word document in Stata

putdocx begin

putdocx textblock begin
Hello my webinar friends. How are you doing today?
putdocx textblock end

putdocx image *emoji.png*

Creating a Word document in Stata

putdocx begin

putdocx textblock begin
Hello my webinar friends. How are you doing today?
putdocx textblock end

putdocx image *emoji.png*

putdocx collect

your_estimation_command
putdocx table *mytable***= etable**

19

Creating a Word document in Stata

putdocx begin

putdocx textblock begin
Hello my webinar friends. How are you doing today?
putdocx textblock end

putdocx image *emoji.png*

putdocx collect

your_estimation_command
putdocx table *mytable***= etable**

putdocx save *myfile*

Analysis on birthweights

```
use low, clear

/* 1. How many observations in our dataset ? */
summarize bwt
local total = r(N)

/* 2. How many mothers in this dataset smoked during
pregnancy? */
count if smoke==1

/* 3. How many didn't? */
display `total'-r(N)

/* 4. How does birthweight vary across mothers' smoking
habits while pregnant, and their history of hypertension
? */
graph hbar bwt, over(ht) over(smoke) asyvars

/* 5. Model birthweight as a function of smoking and
maternal age */
regress bwt i.smoke age, cformat(%9.2f)
```

```
. use low, clear
(Hosmer & Lemeshow data)

. summarize bwt
```

Variable	Obs	Mean	Std. Dev.	Min	Max
bwt	189	2944.286	729.016	709	4990

```
. local total = r(N)

. count if smoke==1
    74

. display `total'-r(N)
115

. graph hbar bwt, over(ht) over(smoke) asyvars

. regress bwt i.smoke age, cformat(%9.2f)
```

Source	SS	df	MS		
Model	4255739.01	2	2127869.51	Number of obs = 189	
Residual	95659559.6	186	514298.707	F(2, 186) = 4.14	
				Prob > F = 0.0175	
				R-squared = 0.0426	
				Adj R-squared = 0.0323	
Total	99915298.6	188	531464.354	Root MSE = 717.15	

bwt	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
smoke						
smoker	-277.29	106.98	-2.59	0.010	-488.34	-66.24
age	11.18	9.88	1.13	0.259	-8.31	30.67
_cons	2793.08	240.93	11.59	0.000	2317.77	3268.40

Creating a reproducible Word document

```
version 17
putdocx clear
```

Create a document in memory

```
use low, clear

putdocx begin
```
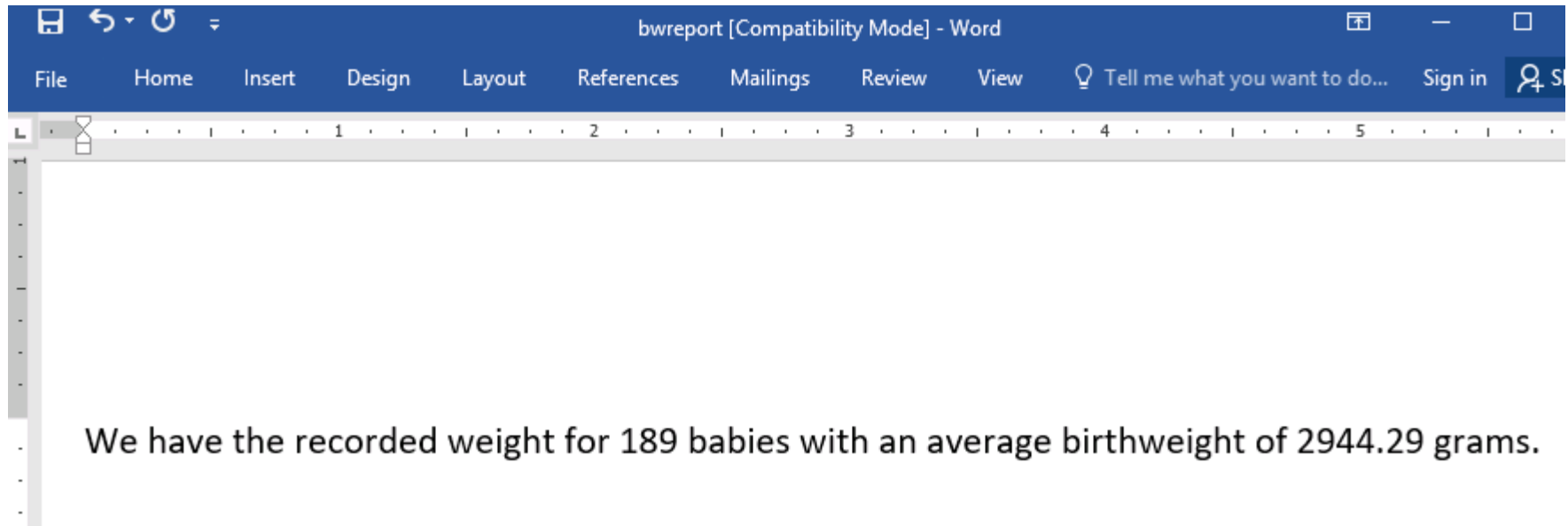
23

Embed text and results in a Word document

```
summarize bwt

local total = r(N)

putdocx textblock begin

We have the recorded weight for <<dd_docx_display: `total'>> babies with an

average birthweight of <<dd_docx_display: %5.2f r(mean)>> grams.
putdocx textblock end
```

Embed text and results in a Word document

```
summarize bwt

local total = r(N)

putdocx textblock begin

We have the recorded weight for <<dd_docx_display: `total'>> babies with an

average birthweight of <<dd_docx_display: %5.2f r(mean)>> grams.

putdocx textblock end
```


We have the recorded weight for 189 babies with an average birthweight of 2944.29 grams.

Embed text and results in a Word document
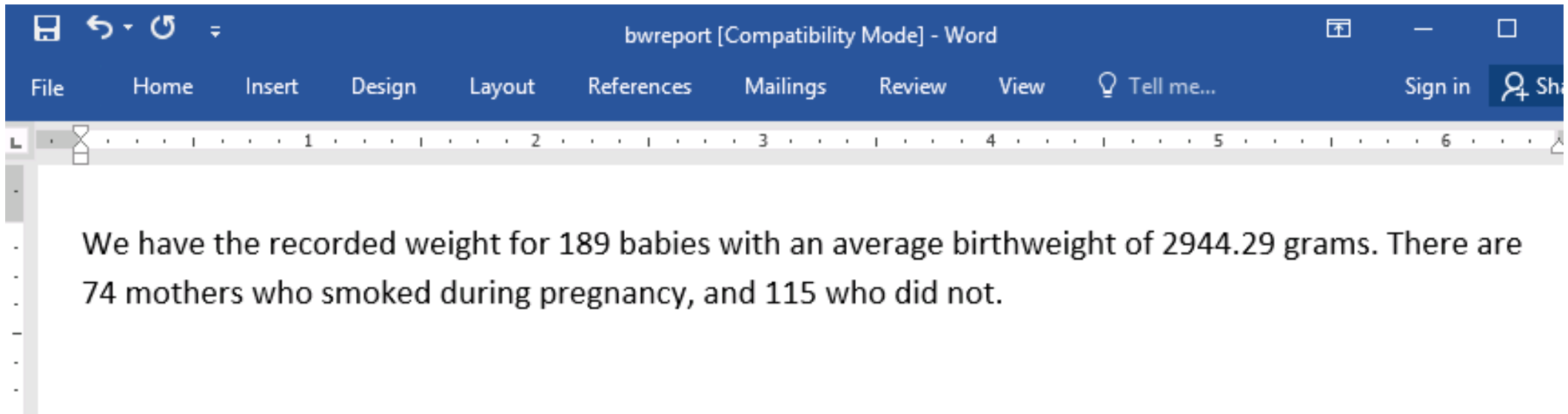
```
version 17
putdocx clear
use low, clear
putdocx begin
summarize bwt
local total = r(N)
putdocx textblock begin
We have the recorded weight for <<dd_docx_display: `total'>> babies with an
average birthweight of <<dd_docx_display: %5.2f r(mean)>> grams.
putdocx textblock end
```

count if smoke==1

putdocx textblock append

There are <<dd_docx_display: r(N)>> mothers who smoked during pregnancy,

and <<dd_docx_display: `total'-r(N)>> who did not.

putdocx textblock end

Embed text and results in a Word document

```
version 17
putdocx clear
use low, clear
putdocx begin
summarize bwt
local total = r(N)
putdocx textblock begin
We have the recorded weight for <<dd_docx_display: `total'>> babies with an
average birthweight of <<dd_docx_display: %5.2f r(mean)>> grams.
putdocx textblock end
```

count if smoke==1

putdocx textblock append

There are <<dd_docx_display: r(N)>> mothers who smoked during pregnancy,

and <<dd_docx_display: `total'-r(N)>> who did not.
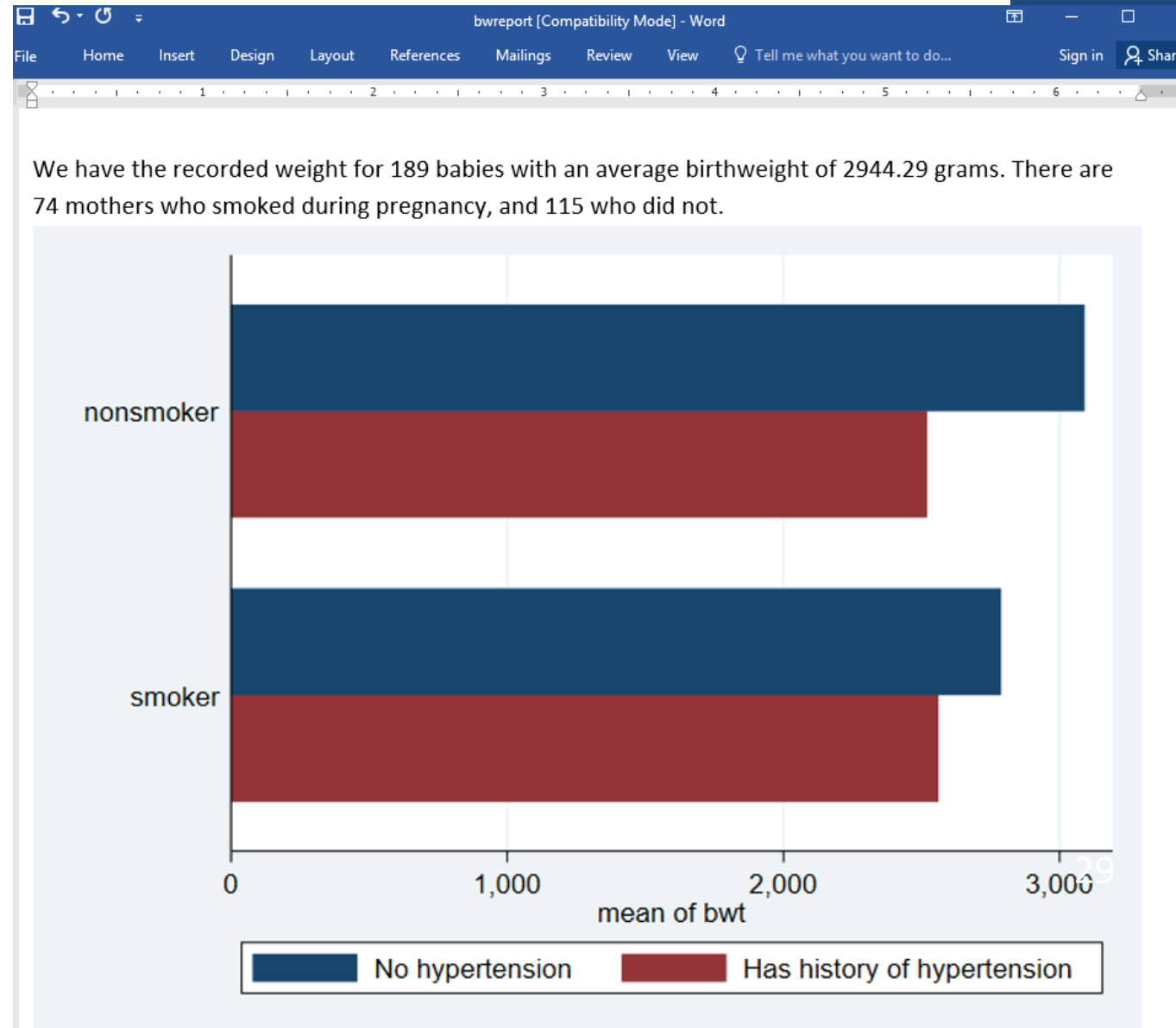
putdocx textblock end

Embed a Stata graph in a Word document

```
version 17
putdocx clear
use low, clear
putdocx begin
summarize bwt
local total = r(N)
putdocx textblock begin
We have the recorded weight for <<dd_docx_display: `total'>> babies with an
average birthweight of <<dd_docx_display: %5.2f r(mean)>> grams.
putdocx textblock end
count if smoke==1
putdocx textblock append
 There are <<dd_docx_display: r(N)>> mothers who smoked during pregnancy,
 and <<dd_docx_display: `total'-r(N)>> who did not.
putdocx textblock end
```

```
graph hbar bwt, over(ht) over(smoke) asyvars

graph export bweight.png

putdocx image bweight.png
```

Embed a Stata graph in a Word document

```
version 17
putdocx clear
use low, clear
putdocx begin
summarize bwt
local total = r(N)
putdocx textblock begin
We have the recorded weight for <<dd_docx_display: `total'>> babies with an
average birthweight of <<dd_docx_display: %5.2f r(mean)>> grams.
putdocx textblock end
count if smoke==1
putdocx textblock append
There are <<dd_docx_display: r(N)>> mothers who smoked during pregnancy,
and <<dd_docx_display: `total'-r(N)>> who did not.
putdocx textblock end
```

```
graph hbar bwt, over(ht) over(smoke) asyvars
graph export bweight.png
putdocx image bweight.png
```
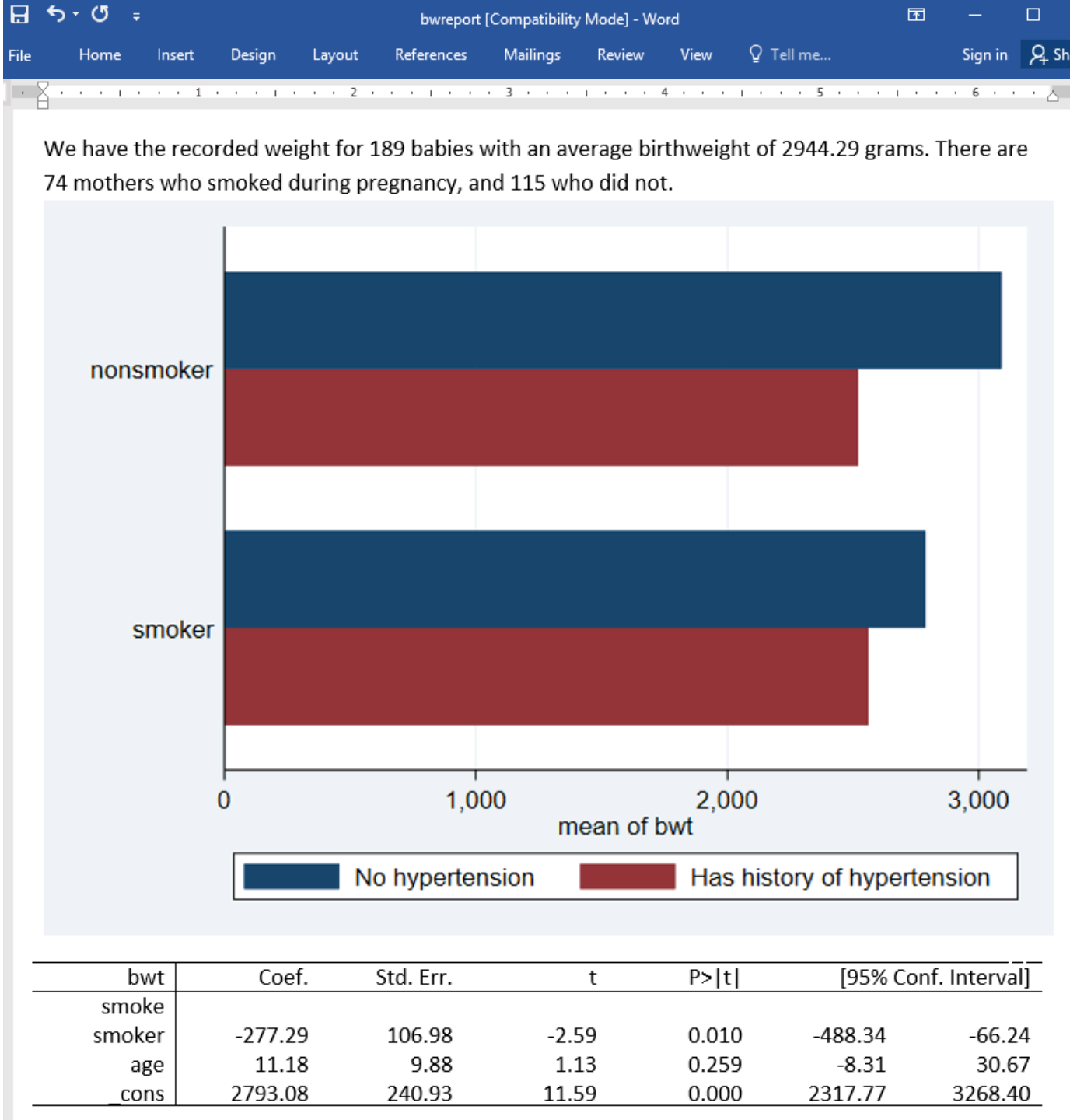
Embed a table of estimation results

```
version 17
putdocx clear
use low, clear
putdocx begin
summarize bwt
local total = r(N)
putdocx textblock begin
We have the recorded weight for <<dd_docx_display: `total'>> babies with an
average birthweight of <<dd_docx_display: %5.2f r(mean)>> grams.
putdocx textblock end
count if smoke==1
putdocx textblock append
 There are <<dd_docx_display: r(N)>> mothers who smoked during pregnancy,
 and <<dd_docx_display: `total'-r(N)>> who did not.
putdocx textblock end
graph hbar bwt, over(ht) over(smoke) asyvars
graph export bweight.png
putdocx image bweight.png
```

```
regress bwt i.smoke age, cformat(%9.2f)

putdocx table bweight = etable
```

Embed a table of estimation results

```
version 17
putdocx clear
use low, clear
putdocx begin
summarize bwt
local total = r(N)
putdocx textblock begin
We have the recorded weight for <<dd_docx_display: `total'>> babies with an
average birthweight of <<dd_docx_display: %5.2f r(mean)>> grams.
putdocx textblock end
count if smoke==1
putdocx textblock append
 There are <<dd_docx_display: r(N)>> mothers who smoked during pregnancy,
 and <<dd_docx_display: `total'-r(N)>> who did not.
putdocx textblock end
graph hbar bwt, over(ht) over(smoke) asyvars
graph export bweight.png
putdocx image bweight.png
```

```
regress bwt i.smoke age, cformat(%9.2f)
putdocx table bweight = etable
```


We have the recorded weight for 189 babies with an average birthweight of 2944.29 grams. There are 74 mothers who smoked during pregnancy, and 115 who did not.

bwt	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
smoke						
smoker	-277.29	106.98	-2.59	0.010	-488.34	-66.24
age	11.18	9.88	1.13	0.259	-8.31	30.67
_cons	2793.08	240.93	11.59	0.000	2317.77	3268.40

Add comments on the output

```
version 17
putdocx clear
use low, clear
putdocx begin
summarize bwt
local total = r(N)
putdocx textblock begin
We have the recorded weight for <<dd_docx_display: `total'>> babies with an
average birthweight of <<dd_docx_display: %5.2f r(mean)>> grams.
putdocx textblock end
count if smoke==1
putdocx textblock append
 There are <<dd_docx_display: r(N)>> mothers who smoked during pregnancy,
 and <<dd_docx_display: `total'-r(N)>> who did not.
putdocx textblock end
graph hbar bwt, over(ht) over(smoke) asyvars
graph export bweight.png
putdocx image bweight.png
regress bwt i.smoke age, cformat(%9.2f)
putdocx table bweight = etable
```

```
putdocx textblock begin
We find that on average, infants whose mothers smoked tend to weigh less.
putdocx textblock end
putdocx save bwreport, replace
```

Create a report in Word

```
version 17
putdocx clear
use low, clear
putdocx begin
summarize bwt
local total = r(N)
putdocx textblock begin
We have the recorded weight for <<dd_docx_display: `total'>> babies with an
average birthweight of <<dd_docx_display: %5.2f r(mean)>> grams.
putdocx textblock end
count if smoke==1
putdocx textblock append
There are <<dd_docx_display: r(N)>> mothers who smoked during pregnancy,
and <<dd_docx_display: `total'-r(N)>> who did not.
putdocx textblock end
graph hbar bwt, over(ht) over(smoke) asyvars
graph export bweight.png
putdocx image bweight.png
regress bwt i.smoke age, cformat(%9.2f)
putdocx table bweight = etable
putdocx textblock begin
We find that on average, infants whose mothers smoked tend to weigh less.
putdocx textblock end
putdocx save bwreport, replace
```

Create a report in Excel

Create a report in Excel

putexcel set *myfile*

Create a report in Excel

```
putexcel set myfile

putexcel A2 = "Hello webinar friends."
putexcel A3 = `r(mean)'
```

Create a report in Excel

```
putexcel set myfile

putexcel A2 = "Hello webinar friends."
putexcel A3 = `r(mean)'

putexcel A5 = image(emoji.png)

putexcel A7 = formula(Excel_formula)
```

Create a report in Excel

```
putexcel set myfile

putexcel A2 = "Hello webinar friends."
putexcel A3 = `r(mean)'

putexcel A5 = image(emoji.png)

putexcel A7 = formula(Excel_formula)

putexcel A8 = collect

your_estimation_command
putexcel A12 = etable
```

Create a report in Excel

```
putexcel set myfile

putexcel A2 = "Hello webinar friends."
putexcel A3 = `r(mean)'

putexcel A5 = image(emoji.png)

putexcel A7 = formula(Excel_formula)

putexcel A8 = collect

your_estimation_command
putexcel A12 = etable

putexcel save
```

Set an Excel workbook for export

```
use low2, clear
putexcel set low.xlsx, replace
```

Write expressions to an Excel workbook

```
use low, clear
putexcel set low.xlsx, replace
```

```
putexcel A1 = "Analysis of birthweights", bold

putexcel A1:C1, border(bottom, thick) merge hcenter

putexcel A3 = "We have data on birthweights from Hosmer, "

putexcel A4 = "Lemeshow, and Sturdivant(2013, 24). "
```

Write expressions to an Excel workbook

```
use low, clear
putexcel set low.xlsx, replace
```

```
putexcel A1 = "Analysis of birthweights", bold
putexcel A1:C1, border(bottom, thick) merge hcenter
putexcel A3 = "We have data on birthweights from Hosmer, "
putexcel A4 = "Lemeshow, and Sturdivant(2013, 24). "
```

Write expressions to an Excel workbook

```
use low, clear
putexcel set low.xlsx, replace
putexcel A1 = "Analysis of birthweights", bold
putexcel A1:E1, border(bottom, thick) merge hcenter
putexcel A3 = "We have data on birthweights from Hosmer, "
putexcel A4 = "Lemeshow, and Sturdivant(2013, 24). "
```

```
putexcel A6 = "Low birthweights by maternal age"
putexcel A6:C6, merge border(bottom,double)
```
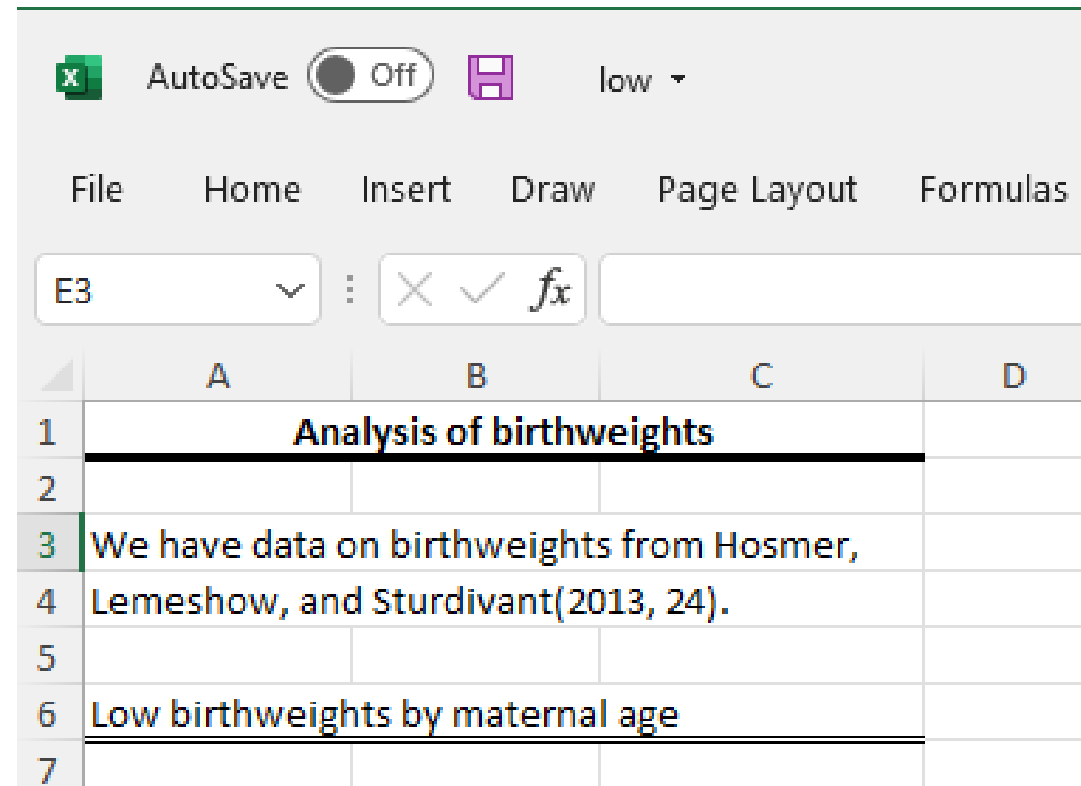
Write expressions to an Excel workbook

```
use low2, clear
putexcel set low.xlsx, replace
putexcel A1 = "Analysis of birthweights", bold
putexcel A1:C1, border(bottom, thick) merge hcenter
putexcel A3 = "We have data on birthweights from Hosmer, "
putexcel A4 = "Lemeshow, and Sturdivant(2013, 24). "
```

putexcel A6 = "Low birthweights by maternal age"

putexcel A6:C6, merge border(bottom,double)

Obtain percentages within each level of a factor variable

```
use low2, clear
putexcel set low.xlsx, replace
putexcel A1 = "Analysis of birthweights", bold
putexcel A1:C1, border(bottom, thick) merge hcenter
putexcel A3 = "We have data on birthweights from Hosmer, "
putexcel A4 = "Lemeshow, and Sturdivant(2013, 24). "
putexcel A6 = "Low birthweights by maternal age"
putexcel A6:C6, merge border(bottom,double)
```

```
. table (agegrp) (), stat(fvpercent low)
```

	Birthweight < 2500 g	
	No	Yes
Age group		
Teens	70.59	29.41
20s	63.96	36.04
30s	84.62	15.38
40s	100.00	0.00
Total	68.78	31.22

Customizing our table

```
use low2, clear
putexcel set low.xlsx, replace
putexcel A1 = "Analysis of birthweights", bold
putexcel A1:C1, border(bottom, thick) merge hcenter
putexcel A3 = "We have data on birthweights from Hosmer, "
putexcel A4 = "Lemeshow, and Sturdivant(2013, 24). "
putexcel A6 = "Low birthweights by maternal age"
putexcel A6:C6, merge border(bottom,double)
```

```
table (agegrp) (), stat(fvpercent low)
collect style cell border_block[corner row-header], border(right, pattern(nil))
collect style cell border_block[corner column-header], border(top, pattern(nil))
```

Customizing and exporting our table

```
table (agegrp) (), stat(fvpercent low)
collect style cell border_block[corner row-header], ///
  border(right, pattern(nil))
collect style cell border_block[corner column-header], ///
  border(top, pattern(nil))
putexcel A8 = collect
```

	File	Home	Insert	Draw	Page Layout	Formulas

A1		✕ ✓ *fx*	Analysis of birthweights

	A	B	C	D
1	Analysis of birthweights			
2				
3	We have data on birthweights from Hosmer,			
4	Lemeshow, and Sturdivant(2013, 24).			
5				
6	Low birthweights by maternal age			
7				
8		Birthweight < 2500 g		
9		No	Yes	
10	Age group			
11	Teens	70.59	29.41	
12	20s	63.96	36.04	
13	30s	84.62	15.38	
14	40s	100.00	0.00	
15	Total	68.78	31.22	

Export a table of estimation results

```
use low2, clear
putexcel set low.xlsx, replace
putexcel A1 = "Analysis of birthweights", bold
putexcel A1:C1, border(bottom, thick) merge hcenter
putexcel A3 = "We have data on birthweights from Hosmer, "
putexcel A4 = "Lemeshow, and Sturdivant(2013, 24). "
putexcel A6 = "Low birthweights by maternal age"
putexcel A6:C6, merge border(bottom,double)
table (agegrp) (), stat(fvpercent low)
collect style cell border_block[corner row-header], border(right,
collect style cell border_block[corner column-header], border(top,
putexcel A8 = collect
```

```
putexcel A18 = "Linear regression model of infant birthweights"
putexcel A18:C18, merge

collect create regression
collect: regress bwt i.smoke age
collect style use myreg, replace

putexcel A19 = collect
```

Export a Stata graph to Excel

```
use low2, clear
putexcel set low.xlsx, replace
putexcel A1 = "Analysis of birthweights", bold
putexcel A1:C1, border(bottom, thick) merge hcenter
putexcel A3 = "We have data on birthweights from Hosmer, "
putexcel A4 = "Lemeshow, and Sturdivant(2013, 24). "
putexcel A6 = "Low birthweights by maternal age"
putexcel A6:C6, merge border(bottom,double)
table (agegrp) (), stat(fvpercent low)
collect style cell border_block[corner row-header], border(right, pattern(nil))
collect style cell border_block[corner column-header], border(top, pattern(nil))
putexcel A8 = collect
putexcel A18 = "Linear regression model of infant birthweights"
putexcel A18:C18, merge
collect create regression
collect: regress bwt i.smoke age
collect style use myreg, replace
putexcel A19 = collect
```

```
graph hbar bwt, over(ht) over(smoke) asyvars ///
  title(Baby birthweights) ytitle(Average birthweight (grams)) ///
  subtitle(by mother's smoking status and history of hypertension)
graph export bweight.png, replace height(300)
putexcel E6 = image(bweight.png)
putexcel save
```

Create a report in Excel

50
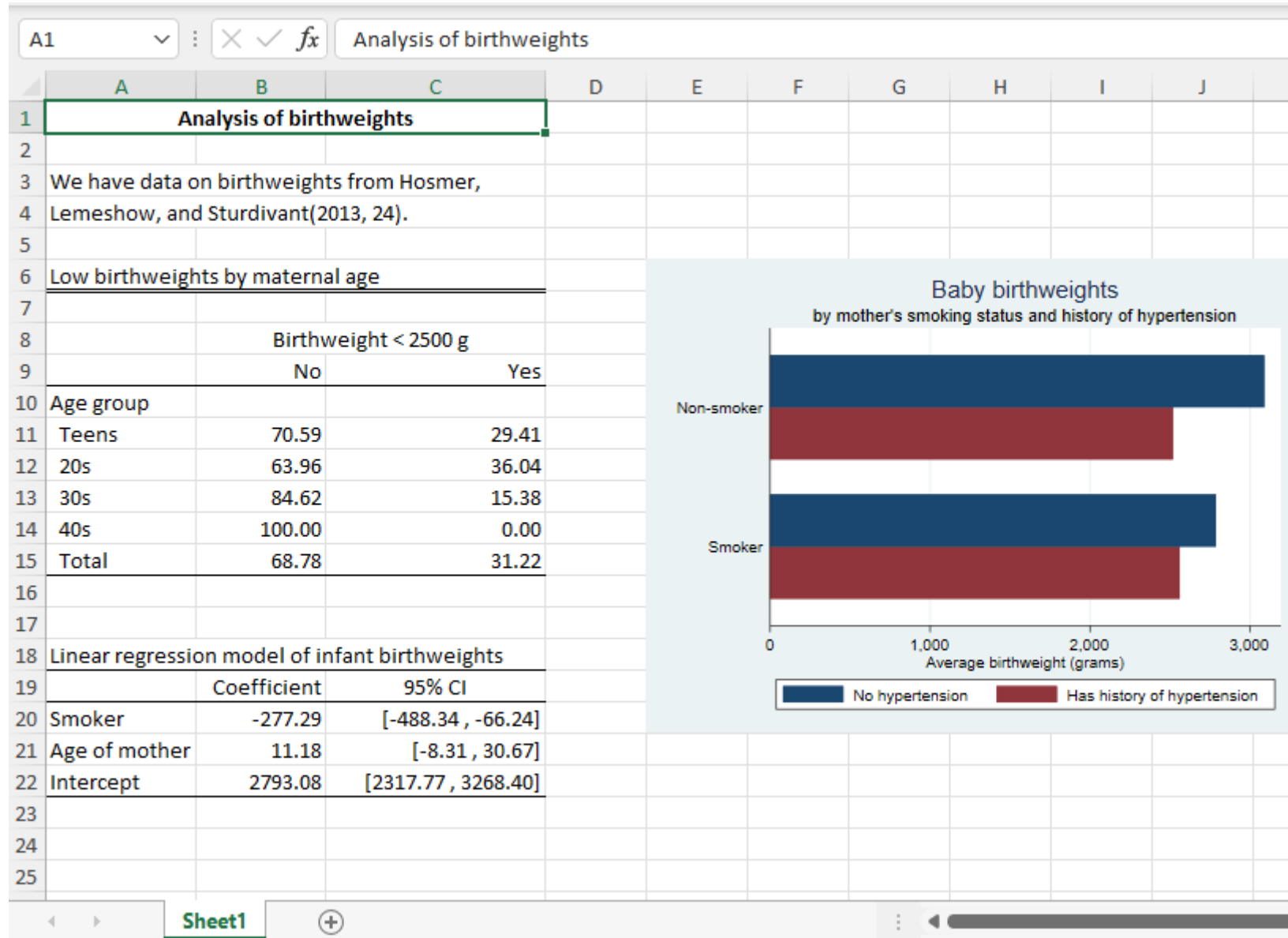
Today we learned how to

- Embed Stata results and graphs in Excel, Word, and PDF
- Customize Word documents with footers, page numbers, and headings
- Add hyperlinks in Word documents
- Append multiple files
- Generalize do-files for automated reporting

There is more you can do

- Set the page size and page margins

- Organize Word documents into sections, each with their own page layout

- Create tables in Word documents from matrices and the dataset in memory

- Nest tables and images within tables

- Learn more about reporting in the [Stata Reporting Reference Manual](#)

- [Create documents with Markdown-formatted text and Stata output](#)

- Get an overview of Stata's [reproducible and automated reporting](#) features

Thank you!