

Lecture 7: Tips & tricks, intro to epi concepts, & dynamic documents

BIOSTAT 212



Amanda Irish | amanda.irish@ucsf.edu
Department of Epidemiology and Biostatistics
University of California San Francisco

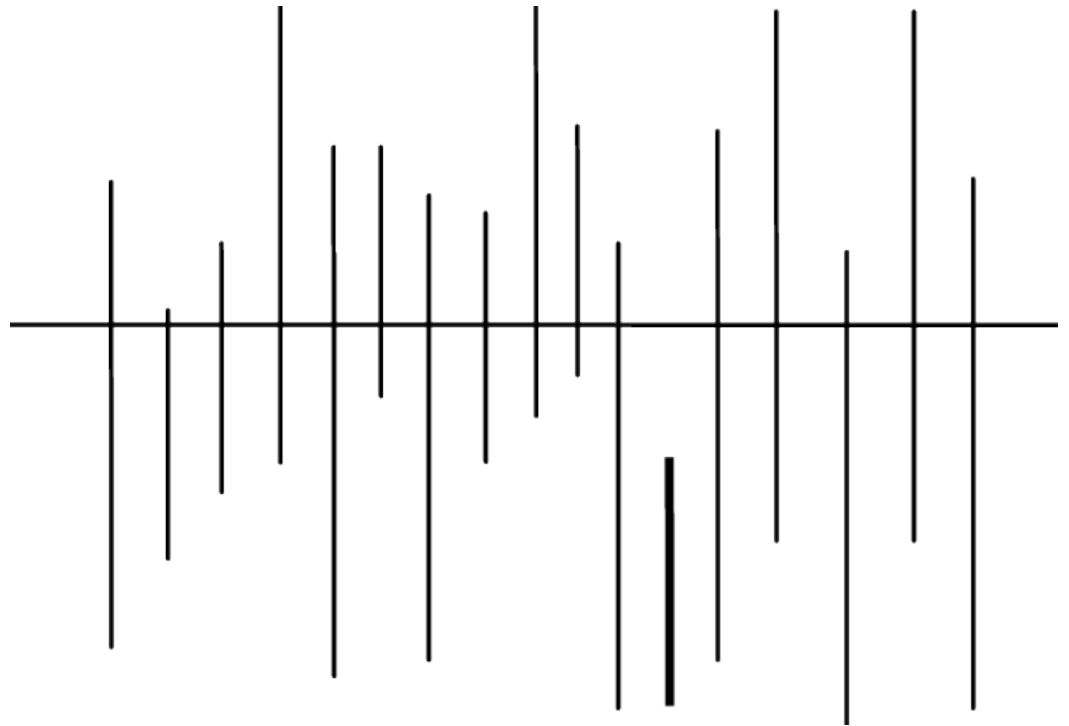
Agenda

- ◊ Tips and tricks
 - Confidence intervals
 - Macros
 - Loops
- ◊ Intro to basic epi concepts & epi immediate commands
- ◊ Dynamic documents

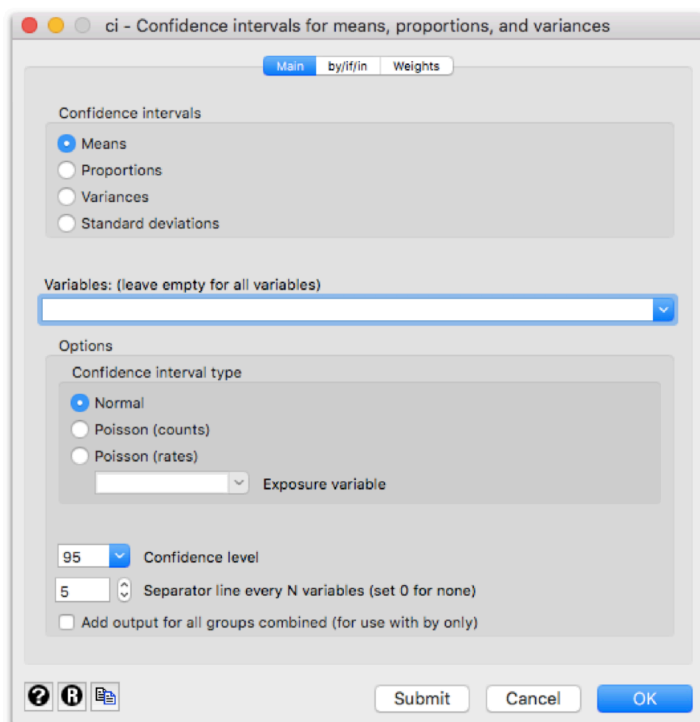
Confidence intervals

What does a 95% confidence interval mean?

- ◇ If we were to repeat a study an infinite amount of times and calculate a confidence interval for each repetition, 95% of those confidence intervals would contain the true parameter value
- ◇ There is a fixed true population parameter (e.g., a mean) and the confidence interval either contains it or doesn't



Confidence intervals in Stata: `ci`



```
. ci means mpg price, level(90)
```

Variable	Obs	Mean	Std. Err.	[90% Conf. Interval]	
mpg	74	21.2973	.6725511	20.17683	22.41776
price	74	6165.257	342.8719	5594.033	6736.48

```
sysuse auto  
ci means mpg price, level(90)
```

Statistics > Summaries, tables, and tests
> Summary and descriptive statistics > Confidence intervals

ci syntax

ci mean `varname`

- calculate confidence interval for a mean (continuous variable)

ci proportion `varname`

- calculate confidence interval for a proportion (dichotomous variable)

options

, level(`#`)

if

bysort `varname`:

-set confidence level (default is 95)

-subset data for CI using logic

-CI for each level of `varname`

Load
acupuncture
data set

use "acupuncture.dta", clear

Practice

- What is the 95% CI around the mean post-study headache score in the intervention group?
- What is the 95% CI around the mean post-study headache score in the control group?

Practice

- What is the 95% CI around the mean post-study headache score in the intervention group?

```
ci mean posths if group==1
```

- What is the 95% CI around the mean post-study headache score in the control group?

```
ci mean posths if group==0
```

Macros & Loops



USE MACROS TO
ABBREVIATE LISTS



USE LOOPS FOR REPETITIVE
TASKS

Local & global macros

- ◇ Abbreviations for a string of characters, number, or expression
- ◇ Local work within the given do-file
- ◇ Global work within the given Stata session
- ◇ Recommended to use local: if using multiple datasets and/or do-files within a single Stata session, global macros can cause errors & confusion

Macros: syntax & examples

```
local macroname "string1 string2 string3"  
`macroname'
```

```
webuse nhanes2
```

```
local continuous "age height weight bpsystol"  
sum `continuous'
```

Variable	Obs	Mean	Std. Dev.	Min	Max
age	10,351	47.57965	17.21483	20	74
height	10,351	167.6509	9.655916	135.5	200
weight	10,351	71.89752	15.35642	30.84	175.88
bpsystol	10,351	130.8817	23.33265	65	300

Practice

- ◇ Check how the categorical variables sex, migraine, group, and response are coded using a local macro.
- ◇ Say we want to compare 2 regression models for the continuous outcome “posths” on group and age in one model, and group, age, and sex in the second. Create local macros for the 2 sets of covariates and run linear regression models using the macros.

Practice

- Check how the categorical variables sex, migraine, group, and response are coded using a local macro.

```
local catvar "sex migraine group response"  
codebook `catvar'
```

- Compare 2 regression models for the continuous outcome “posths” on group and age in one model, and group, age, and sex in the second.

```
local model1var "group age"  
regress posths `model1var'
```

```
local model2var "group age sex"  
regress posths `model2var'
```

Loops

- ◇ Loops are commands that execute a group of commands multiple times
- ◇ Automate repetitive tasks
- ◇ Make your code shorter, clearer, and less prone to error
- ◇ Any time you find yourself copying and pasting code, ask yourself if you should be using a loop

Loops: foreach

- ◇ `foreach` is the most versatile way to create loops in Stata
- ◇ Allows for looping over a varlist, numlist, local or global macro
- ◇ Note placement of curly brackets in the syntax
- ◇ `set trace on` / `set trace off` before/after loop to locate errors

```
foreach var of varlist var1 var2 ... {  
    command `var'  
}
```

foreach examples

```
webuse nhanes2
```

```
foreach var of varlist age height weight bpsystol {  
    sum `var' if region==1  
    histogram `var' if region==1  
}
```



```
foreach var of varlist sex-hdresult {  
    replace `var'=. if `var'==999  
}
```

*** This is equivalent to:**

```
sum age if region==1  
histogram age if region==1  
  
sum height if region==1  
histogram height if region==1  
  
sum weight if region==1  
histogram weight if region==1  
  
sum bpsystol if region==1  
histogram bpsystol if  
region==1
```

Practice

- ◇ What percentage of study participants are female, report migraines, and showed >35% improvement in headache score by intervention group?
- ◇ What are the 95% CIs around the mean baseline headache score, post-study headache score, and one-year headache score in the intervention and control groups?

Practice

- What percentage of study participants are female, report migraines, and showed >35% improvement in headache score by intervention group?

```
foreach var of varlist sex migraine response {  
    di "group x `var'"  
    tab group `var', row  
}
```

- What are the 95% CIs around the mean baseline headache score, post-study headache score, and one-year headache score in the intervention and control groups?

```
foreach var of varlist basehs posths oneyrhs {  
    ci mean `var' if group==1  
    ci mean `var' if group==0  
}
```


Additional resources on loops

- ◇ <https://stats.idre.ucla.edu/stata/modules/working-across-variables-using-foreach/>
- ◇ <https://data.library.virginia.edu/stata-basics-foreach-and-forvalues/>
- ◇ <https://jearl.faculty.arizona.edu/sites/jearl.faculty.arizona.edu/files/Introduction%20to%20Loops%20in%20Stata.pdf>
- ◇ <https://jearl.faculty.arizona.edu/sites/jearl.faculty.arizona.edu/files/Intro%20to%20loops,%20Year%202.pdf>

Intro to basic epi concepts

Motivating Research Question:

*Did Kevin Durant's presence on the Warriors
cause the SF Giants to be bad at baseball?*

The KD Curse?

- ◇ **Exposure:** Kevin Durant on the Warriors
- ◇ **Outcome:** SF Giants win percentage

Measures of Association

- ◇ An association between an exposure and outcome can be assessed by comparing the probability of outcome in the exposed group to the probability of outcome in unexposed group.
- ◇ **Ratio (division)**
$$\frac{[\text{Pr}(\text{outcome}) \text{ in exposed}]}{[\text{Pr}(\text{outcome}) \text{ in unexposed}]}$$
- ◇ **Difference (subtraction)**
$$[\text{Pr}(\text{outcome}) \text{ in exposed}] - [\text{Pr}(\text{outcome}) \text{ in unexposed}]$$

Ratio Measures of Association

Study Design	Ratio Measure	Stata command
Cross-sectional study	Prevalence Ratio (PR)	<code>cs O E, by(C)</code>
		<code>binreg O E C, rr</code> <code>glm O E C, family(bin) link(log) eform</code>
Cohort study	Risk Ratio (RR)	<code>cs O E, by(C)</code>
		<code>binreg O E C, rr</code> <code>glm O E C, family(bin) link(log) eform</code>
		<code>ir O E T</code> <code>poisson O E C, irr</code>
	Hazard Ratio (HR)	<code>stcox O E C</code>
Case-control study	Odds Ratio (OR)	<code>cc O E, by(C)</code>
		<code>mhodds O E, by(C)</code>
		<code>logistic O E C, or</code>
		<code>tabodds O E, adjust(C)</code>

O = outcomevar E = exposurevar C = covariates T = time

Difference Measures of Association

Study Design	Ratio Measure	Stata command
Cross-sectional study	Prevalence Difference	<code>cs O E</code> <code>binreg O E C, rd</code>
Cohort study	Risk Difference	<code>cs O E</code> <code>binreg O E C, rd</code>
	Rate Difference	<code>ir O E T, ird</code> <code>poisson O E C</code>
Case-control study	N/A	

O = outcomevar E = exposurevar C = covariates T = time

Risk Ratio and Risk Difference

		Disease	
		Yes	No
Exposure	Yes	a	b
	No	c	d

Risk Ratio =

$$\frac{a/(a+b)}{c/(c+d)}$$

Risk Difference =

$$a/(a+b) - c/(c+d)$$

Risk = Probability

cohort studies & RCTs

*similar concept for prevalence ratios, used in cross-sectional studies

Risk Ratios/Differences: `cs`

`cs outcomevar predictorvar`

– calculates risk ratios and risk differences

Options

`,by (varname)`

-stratify by a third variable

`[, missing]`

-include missing data as a category of the stratified variable

Limitations

❖ Cannot adjust for more than 1 covariate

Practice

- ◇ open “Kdcurse_data.dta”
- ◇ Create a table showing the Giants’ win probability by year
- ◇ Create a table showing the Giants’ win probability by era (pre KD, KD, post KD)
- ◇ Create a table showing the Giants’ win probability by home vs away (home)

Practice

- ◇ Create a table showing the Giants' win probability by year

```
tabstat win, by(year) stat(mean)
```

- ◇ Create a table showing the Giants' win probability by era (pre KD, KD, post KD)

```
tabstat win, by(era) stat(mean)
```

- ◇ Create a table showing the Giants' win probability by home vs away (home)

```
tabstat win, by(home) stat(mean)
```

Practice

- What is the risk ratio for the effect of KD on the Giants' win probability?
- What is the risk difference?

Practice

- What is the risk ratio for the effect of KD on the Giants' win probability?

`cs win KD`

- What is the risk difference?

`cs win KD`

Odds Ratio

		Outcome	
		Yes	No
Exposure	Yes	a	b
	No	c	d

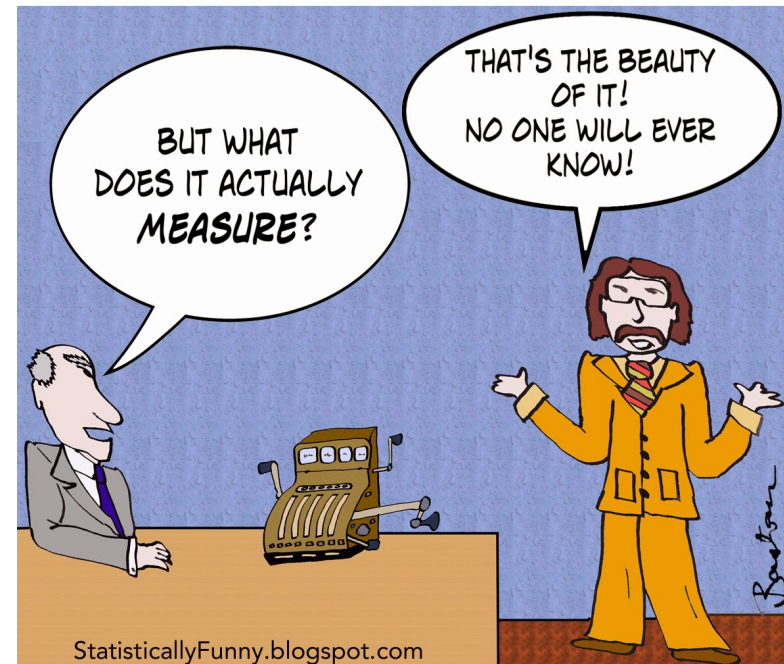
Odds Ratio =

$$\frac{a/b}{c/d}$$

For case-control studies

The odd thing about Odds Ratios....

- ❖ Odds are not intuitive, difficult to communicate and interpret
- ❖ They are farther from the null than the risk ratio
 - exaggerate the association if interpreted as a probability



Odds Ratios: `cc`

`cc outcomevar predictorvar`

– calculates odds ratios

Options

`,by (varname)`

-stratify by a third variable

`[, missing]`

-include missing data as a category of the stratified variable

Limitations

Cannot adjust for more than 1 covariate

Odds Ratios: mhodds

mhodds outcomevar predictorvar [covariates]

– calculates odds ratios and can adjust for covariates

Options

,by (varname)
[, missing]

-stratify by a third variable [to look for interaction]
-include missing data as a category of the stratified variable

Limitations

- ◇ Uses stratification to adjust for covariates
 - with several covariates, strata start to get “thin”

Practice

- What is the OR for the effect of KD on the Giants winning?
- What is the OR for the effect of KD on the Giants winning adjusting for home games?

Practice

- What is the OR for the effect of KD on the Giants winning?

```
cc win KD
```

```
mhodds win KD
```

```
cs win KD, or
```

- What is the OR for the effect of KD on the Giants winning adjusting for home games?

```
cc win K, by(home)
```

```
mhodds win KD, by(home)
```

Logistic regression

```
logistic outcomevar predictorvar covariates
```

```
logit outcomevar predictorvar covariates, or
```

```
glm outcomevar predictorvar covariates, family(binomial)  
    link(logit) eform
```

Options

-Specify an interaction between 2 variables

If using any categorical predictors with >2 categories, use `i.catvar`

Practice

- What is the OR for the effect of KD on the Giants winning?
- What is the odds ratio (OR) adjusting for whether it was a day or night game?

Practice

- What is the OR for the effect of KD on the Giants winning?

```
logistic win KD
```

```
logit win KD, or
```

```
glm win KD, family(binomial) link(logit) eform
```

- What is the odds ratio (OR) adjusting for whether it was a day or night game?

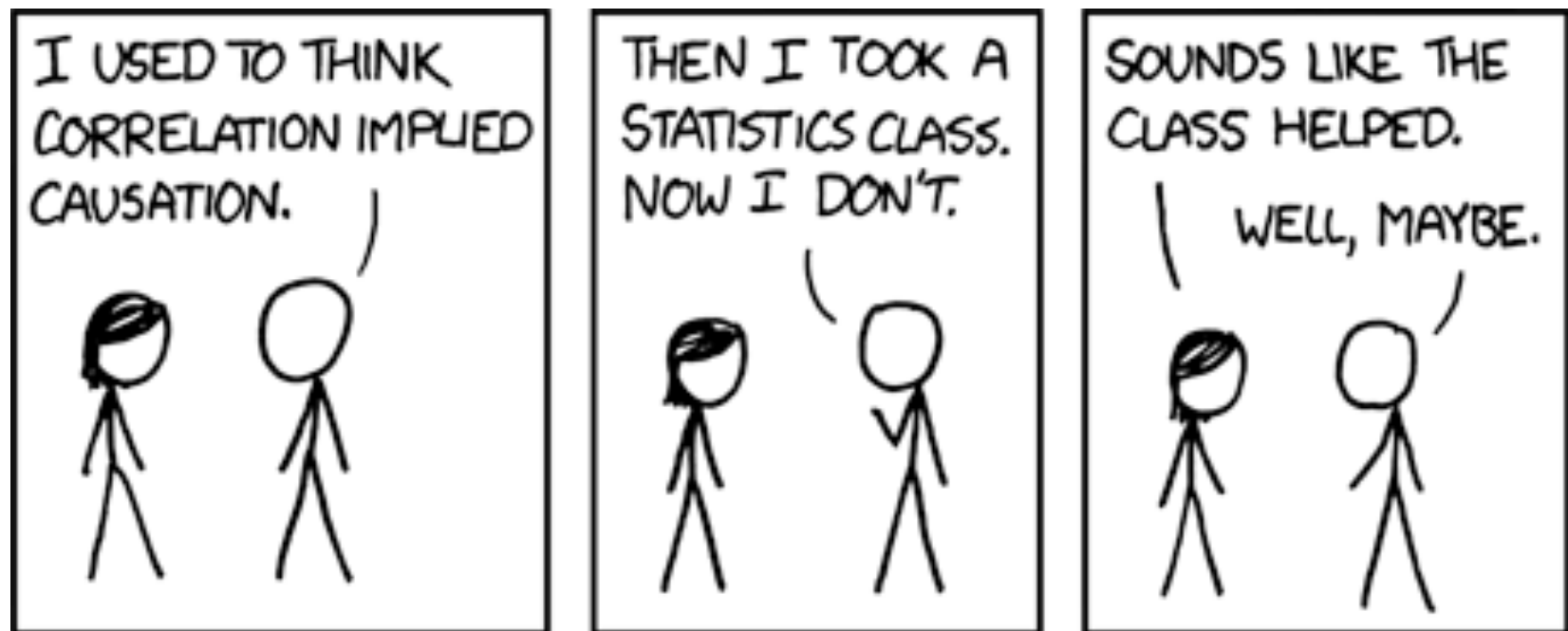
```
logistic win KD daygame
```

```
logit win KD daygame, or
```

```
glm win KD daygame, family(binomial) link(logit) eform
```

```
mhodds win KD daygame
```

Why do we care about associations?



Source: xkcd

We often* care about *causal*
questions

* I'd argue almost always

Causal Inference

6 ways a statistical association can occur:

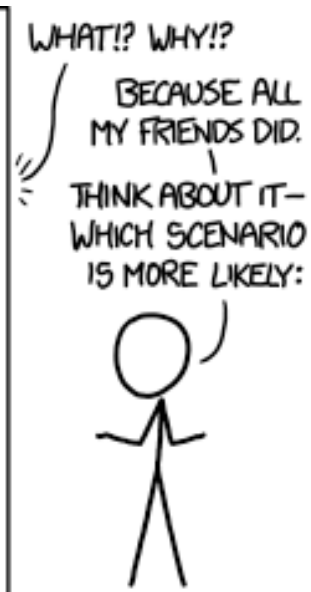
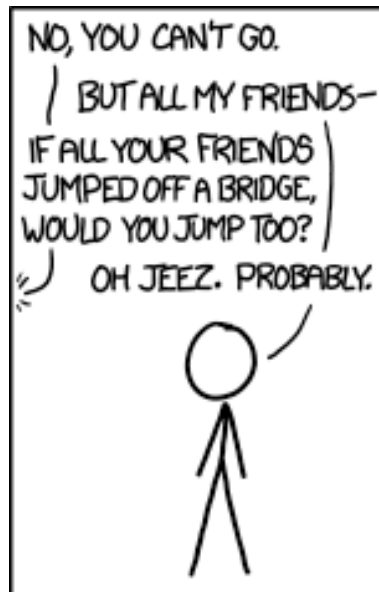
1. Chance
2. Selection bias
3. Measurement bias
4. Confounding
5. Reverse causation
6. **True causal relationship**

Causal Inference

6 ways a statistical association can occur:

1. Chance
2. Selection bias
3. Measurement bias
4. Confounding
5. Reverse causation
6. **True causal relationship**

If we can eliminate all other possible explanations for an association, then association is likely to be causal



dyndoc and producing
living reports in STATA

What is a dynamic document?

**BEWARE
LIVING
DOCUMENT!**



- ◇ Mix of text and code that automatically embeds results into text documents
- ◇ Markdown is a commonly used language for writing dynamic documents, and STATA incorporates the syntax into dyndoc
- ◇ **Allows you to have automatically produced reports that update to match latest data and changes to analyses**

Getting Started

```
<<dd_version: 2>>

BIOSTAT 212: INTRODUCTION TO DYNDOC
=====
This is an example of a dyndoc document. It's written in Markdown syntax, and
can be converted into a living MS Word or HTML document that will update the
results if the data change.

To turn this file into a .docx file, save this as dyndoc_example_biostat212.txt
. Then use the command dyndoc "dyndoc_example_biostat212.txt", saving("Dyndoc
Example Output.docx") docx

##Setting up
At the top of your .txt file, you should include the
<<dd_ignore>>
<<dd_version:2>> command to tell STATA the version of dyndoc that you're using.
Use version 2 if you are using STATA 16, and use version 1 if you are using
STATA 15.

To embed STATA commands and output, use:
~~~~
<<dd_do>>
commands go here
output will appear here (except graphs. more later)
<</dd_do>>
~~~~
<</dd_ignore>>

### Set working directory and open new log
~~~~
<<dd_do:quietly>>
cd "/Users/mmehrotra/Documents/biostat212/2019/week5"
clear
capture log close
log using "output/week5_lecture.smcl", replace
<</dd_do>>
~~~~
```

BIOSTAT 212: INTRODUCTION TO DYNDOC

This is an example of a dyndoc document. It's written in Markdown syntax, and can be converted into a living MS Word or HTML document that will update the results if the data change.

To turn this file into a .docx file, save this as dyndoc_example_biostat212.txt. Then use the command `dyndoc "dyndoc_example_biostat212.txt", saving("Dyndoc Example Output.docx") docx`

Setting up

At the top of your .txt file, you should include the `<<dd_version:2>>` command to tell STATA the version of dyndoc that you're using. Use version 2 if you are using STATA 16, and use version 1 if you are using STATA 15.

To embed STATA commands and output, use:

```
<<dd_do>>
commands go here
output will appear here (except graphs. more later)
<</dd_do>>
```

Set working directory and open new log

Load Acupuncture Dataset

```
. use "data/acupuncture.dta"
. describe, short
```

```
Contains data from data/acupuncture.dta
obs:      401
vars:      83                      26 Aug 2016 16:46
Sorted by:
```



Tags

Dynamic tag	Description
<code><<dd_version>></code>	specify the minimum version required to convert the dynamic document
<code><<dd_do>></code>	execute a block of Stata code and optionally include its output
<code><</dd_do>></code>	end <code><<dd_do>></code>
<code><<dd_display>></code>	include output of Stata expression as shown by Stata's <code>display</code> command
<code><<dd_docx_display>></code>	include output of Stata expression in a <code>.docx</code> file as shown by Stata's <code>display</code> command and format text within a block
<code><<dd_graph>></code>	export a Stata graph and include a link to the file
<code><<dd_ignore>></code>	disable processing of dynamic tags except <code><<dd_remove>></code>
<code><</dd_ignore>></code>	end <code><<dd_ignore>></code>
<code><<dd_include>></code>	include the contents of a text file
<code><<dd_remove>></code>	remove the following text until <code><</dd_remove>></code> is specified
<code><</dd_remove>></code>	end <code><<dd_remove>></code>
<code><<dd_skip_if>></code>	skip text based on condition
<code><<dd_skip_else>></code>	skip text based on condition
<code><<dd_skip_end>></code>	end <code><<dd_skip_if>></code> block

`<<dd_docx_display>>` is only for use with `putdocx` `textblock` commands in a `do`-file.

Creating your document

After you've created your .txt file including commands and comments, run (in the command line)

Word Document (STATA 16):

```
dyndoc "filename.txt", saving(filename.docx) docx  
replace
```

HTML (STATA 15):

```
dyndoc "filename.txt", saving(filename.html) replace
```


Embedding results in text

Including output in lines of text

Let's embed some statistics about our data in our text. Instead of copying and pasting the results from `summarize`, we can use the `dd_display` dynamic tag with `r(min)` and `r(max)` stored results:

```
. summarize basehs
```

Variable	Obs	Mean	Std. Dev.	Min	Max
basehs	401	26.50998	16.16322	6.75	94.75

The variable `basehs` has a minimum value 6.75 and has a maximum value 94.75

```
##Including output in lines of text
```

Let's embed some statistics `about` our data in our text. Instead of copying and pasting the results from `summarize`, we can `use` the `**dd_display**` dynamic tag with `**r(min)**` and `**r(max)**` stored results:

```
~~~~
```

```
<<dd_do>>
```

```
summarize basehs
```

```
<</dd_do>>
```

The variable `basehs` has a minimum value `<<dd_display: %4.2f `r(min)'\>>` and has a maximum value `<<dd_display: %4.2f `r(max)'\>>`

```
~~~~
```

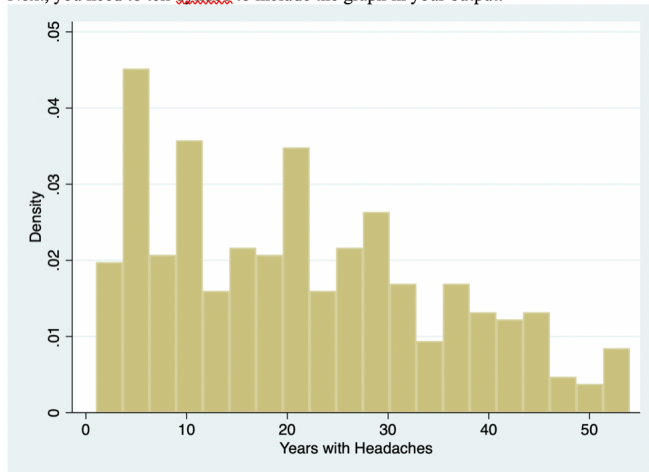
Figures

Including Graphs

Including a graph requires 2 steps. First, you need to create the graph in the code:

```
. histogram chronicity  
(bin=20, start=1, width=2.65)
```

Next, you need to tell `dyndoc` to include the graph in your output:



```
##Including Graphs
```

Including a `graph` requires 2 steps. First, you need to create the `graph` in the code:

```
~~~~~  
<<dd_do>>  
histogram chronicity  
<</dd_do>>  
~~~~~
```

Next, you need to tell `dyndoc` to `include` the `graph` in your output:

```
<<dd_graph: saving(histogram.png) replace height(400)>>
```

Demo

Thank you!

