Stata: An Introduction



BUILD Online Research Module

CSULB

Summer 2020



A Gentle Introduction to Stata

Sixth Edition

ALAN C. ACOCK



Acknowledgements

- Dr. Alan Acock, Author of "A Gentle Introduction to Stata"
- CSULB BUILD Program (NIH Award #RL5GM118978)



Agenda

- I. Background
- II. Getting Started
- III. Entering Data
- IV. Preparing Data for Analysis
- V. Working with commands, do-files and results
- VI. Descriptive Statistics and Graphs Single Variables
- VII. Statistics and Graphs Two Categorical Variables
- VIII. Tests for one or two means
- IX. Bivariate Correlation and Regression
- X. Multiple Regression
- XI. Logistic Regression
- XII. A Public Health Example
- XIII. What's Next

I. Background



- Public Health Background
 - Statistics training via coursework and real-world research projects
 - Quantitative Software experience:
 - Excel
 - MPlus
 - SPSS
 - <u>Stata</u>

I. Background





• Stata fan

- Why I like Stata:
 - Easy to use
 - Fast
 - Saves Time (Do-Files)
 - Helps you when you make a mistake
 - Efficient
 - Capable of complex analyses

I. Background



- Today's Workshop
 - Basic understanding of statistics
 - Created for the Stata novice
 - Covering the basics of Stata

- Before we begin
 - Short Term Trial for Students
 - <u>https://www.stata.com/customer-service/short-term-license/</u>
 - Student Plans
 - <u>https://www.stata.com/order/new/edu/gra</u> <u>dplans/student-pricing/</u>
 - Module material
 - Do File
 - Sample data set and codebook (upon formal request)
 - Personal Recommendation
 - "Pause" often to follow along as we go through different examples



• OBJECTIVES

- Understand the Stata interface
- Open an existing dataset
- Complete a short Stata session



Open Stata



• The Stata Interface



Open an Existing Dataset

tata/MP 13.1 - [Results]			File Edit History Help		
File Edit Data Graphics Statistics User Window Help		8			
☑ Open Ctrl+O □ ○ ○			🗢 🔿 🕃 🚍 🐕 help dta_examp	oles	
Save Ctrl+S	 Variables 	Ψ÷	help dta_examples ×		
Save As Ctrl+Shift+S	Variable Label				
View	There are no items to sh	ow.			
Do tatistics/Data Analysis StataCorp			Example datagets installs	d with Stata	
Filename 4905 Lakeway Drive				d with Stata	
Change Working Directory MP - Parallel Edition College Station, Texas 7/845 USA			The datasets listed h	ere are installed with S	tata Vo
Log 979-696-4600 stata8stata.com			Stata documentation t	hat are available via th	e Interne
Import 979-696-4601 (fax)			to the list of datase	ts for that manual.	
Export					
Print gle-user 2-core Stata perpetual license:					
Example Datasets		1	auto.dta	use describe	
Recent Datasets			auto2.dta	use describe	
			autornd.dta	use describe	
Exitestes:			bplong.dta	use describe	
 (/v# option or -set maxvar-) 5000 maximum variables 			bpwide.dta	use describe	
			cancer.dta	use describe	
			census.dta	use describe	
			citytemp.dta	use describe	
			citytemp4.dta	use describe	
			educ99gdp.dta	use describe	
	-		gnp96.dta	use describe	
			lifeexp.dta	use describe	
Command	*		network1.dta	use describe	
			networkia.dta	use describe	
			nisw88.dta		
			niswider.dta	use describe	
			sandstone dta	use describe	
			sp500 dta	use describe	
Title			surface dta	use describe	
[U] 1.2.2 Example datasets			tslinel.dta	use describe	
C:\Users\013733432\Documents			tsline2.dta	use describe	
Description			uslifeexp.dta	use describe	
			uslifeexp2.dta	use describe	
Example datasets installed with Stata			voter.dta	use describe	
This page contains links enabling you to describe or u	se the datasets that were installed	with Stata.	xtline1.dta	use describe	
		1			

h Stata are installed with Stata. You are available via the Internet that manual. se | describe e | describe se | describe

Stata 13 manual datasets

This page provides web access to all the datasets referred to in the Stata documentation.

Open an existing dataset National Longitudinal Study of Women, NLSW, 1988

ine contraction of optimizer of	austics user window help					•
) 🖬 🖷 🔲 🖻 • 📖 • 📝 • 🖻						
eview ▼ ₽ ×				 Variables 		ү म ×
Command _rc	$()^{(R)}$			Variable	Label	
use "C:\Program Files		1 Convright 1985-2	013 StataCorp LP	idcode	NLS id	
	Statistics/Data Analysis	StataCorp		age	age in current	:
	_	4905 Lakeway Dri	ve	race	race	
	MP - Parallel Edition	College Station,	Texas 77845 USA	married	married	l
		800-STATA-PC	http://www.stata.com	never_mar	never married	
		979-696-4600	stata@stata.com	grade	current grade	
		979-696-4601 (fa	x)	collgrad	college gradu	
				south	lives in south	
	Single-user 2-core Stata perpetua.	1 11cense:		smsa	lives in SMSA	
	Licensed to: Niloofar B	ov avarian		Proportion		
	CSULB	avarian				т
				Variables		
	Notes:			Name	idcode	
	 (/v# option or -set max 	var-) 5000 maximum va	riables	Label	NLS id	
				Type	int	
	. use "C:\Program Files (x86)\Sta	ta13\ado\base\n\nlsw8	8.dta", clear	Format	%8.0g	
	(NLSW, 1988 extract)			Value Labe		
				Notes		
				🗆 Data		
					nlsw88.dta	
				Label	NLSW, 1988 ex	xtract
				■ Notes		
				Variables	17	
				 Observatio 	ins 2,246	
				Size	59.22K	
	Command			a Memory	64M	
				Sorted by	idcode	

Open an existing dataset describe

dit Data Graphics Statistics User Win	dow Help							
T # x describe						Variables		
amond in the						Variable	Label	
nmand _rc Contains dat	a from c :\	Program Fil	es (x86)\Sta	ata13\ado\base\n\nlsw88.dta		Variable	Label	
"C:\Program Files obs:	2,246	2		NLSW, 1988 extract		idcode	NLS id	
cribe vars:	17			1 May 2011 22:52		age	age in current	
size:	60,642			(dta has notes)		race	race	
						married	married	
	storage	display	value			never_mar.	never married	
variable nam	e type	format	label	variable label		grade	current grade	
						collorad	college gradu	
idcode	int	%8.0g		NLS id		couth	lives in south	
age	byte	%8.0g		age in current year		south	lives in south	
race	byte	%8.0g	racelbl	race		smsa	lives in SMSA	
married	byte	%8.0g	marlbl	married		c_city	lives in central	
never_marrie	d byte	%8.0g		never married		industry	industry	
grade	byte	%8.0g		current grade completed		occupation	n occupation	
collgrad	byte	%16.0g	gradlbl	college graduate		union	union worker	
south	byte	%8.0g		lives in south		wage	hourly wage	
smsa	byte	%9.0g	smsalbl	lives in SMSA		bours	usual hours w	
c_city	byte	%8.0g		lives in central city	E	ttl. ovn	total work own	
industry	byte	%23.0g	indlbl	industry		tu_exp	total work exp	
occupation	byte	%22.0g	occlbl	occupation		tenure	Job tenure (ye	
union	byte	%8.0g	unionlbl	union worker				
wage	float	%9.0g		hourly wage				
hours	byte	%8.0g		usual hours worked				
ttl_exp	float	%9.0g		total work experience				
tenure	float	%9.0g		job tenure (years)				
Sorted by:	idcode				-			
Command								

- Complete a short Stata session
 - summarize

Stata/MP 13.1 - C:\Program File	es (x86)\Stata13\ado\bas	e\n\nlsw88.dta -	Results]		-	
File Edit Data Graphics St	tatistics User Window	Help				and in Stratistic
		noip				
	Sorted by: ide	ada				
eview f + ×	Sorted by. Ide	ode				
Command _rc	summarize					
use "C:\Program Files						
describe	Variable	Obs	Mean	Std. Dev.	Min	Max
summarize						
	idcode	2246	2612.654	1480.864	1	5159
	age	2246	39.15316	3.060002	34	46
	race	2246	1.282725	.4754413	1	3
	married	2246	. 6420303	.4795099	0	1
	never_marr~d	2246	.1041852	.3055687	0	1
	grade	2244	13.09893	2.521246	0	18
	collgrad	2246	.2368655	. 4252538	0	1
	south	2246	.4194123	. 4935728	0	1
	smsa	2246	.7039181	.4566292	0	1
	c_city	2246	.2916296	.4546139	0	1
	industry	2232	8.189516	3.010875	1	12
	occupation	2237	4.642825	3.408897	1	13
	union	1878	.2454739	.4304825	0	1
	wage	2246	7.766949	5.755523	1.004952	40.74659
	hours	2242	37.21811	10.50914	1	80
	ttl_exp	2246	12.53498	4.610208	.1153846	28.88461
	tenure	2231	5.97785	5.510331	0	25.91667

- Complete a short Stata session
 - summarize age
 - summarize age if
 south == 1

• . summarize age

•	Variable		Obs	Mean	Std. D	ev. Min	Max
•		-+					
•	age		2246 3	9.15316	3.0600	02 34	46

• . summarize age if south == 1

•	Variable		Obs	Mean	Std.	Dev.	Min	Max
•		+						
•	age	I	942	39.17834	3.11	8291	34	45

- Complete a short Stata session
 - histogram age, percent normal



- Objectives
 - Understand the importance of a codebook
 - Create variables in Stata
 - List resources for Data Management in Stata



• First start with a codebook

#	Variable Name	Question	Response Op	ptions/How	Scored		
	id	Student ID Number	Continuous				
	classid	Classroom ID Number	Continuous				
	deptid	Course Department	Descriptive	e			
			Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	_inatt1_t1	It is difficult for me to pay attention during classes	1	2	3	4	5
2	_hyper1_t1	I often feel restless	1	2	3	4	5
3	_inatt2_t1	It is difficult for me to concentrate on my academic work	1	2	3	4	5
4	_hyper2_t1	I am an impulsive person	1	2	3	4	5
5	_inatt3_t1	I have difficulty keeping track of my different school assignments	1	2	3	4	5
6	_hyper3_t1	I rarely plan ahead	1	2	3	4	5
7	_ss1_t1	I like "wild" parties	1	2	3	4	5
8	_ss2_t1	I enjoy getting into situations where I do not know how things will turn out	1	2	3	4	5
9	_ss3_t1	I prefer friends who are unpredictable	1	2	3	4	5

BEACH Q-V2 Codebook

- Use the Command Box to Create Variables in Stata
 - gen id=.
 - gen age = .
- Use the Command Box to Enter Variable Labels
 - . label variable id "Student ID"
 - . label variable age "Student Age"
- Go into the Data Editor (Edit) to enter data



- Use the Command Box to Create Variables in Stata (Categorical)
 - gen female=.
 - label variable female "Female"
- Create labels for response categories
 - label define sex 0 "male" 1 "female"
 - label values female sex
 - Try entering 0 and 1 for in your Data Editor



• Resources for Data Management



- Some of us may
 - 1) Prefer to enter data in Excel
 - 2) Work with secondary data that came in a different type of data file (e.g., SPSS)

-	🔏 Cut 🛛 🔭 🛙	Layout * Calibri (Body) * 28 * A A A
		Parat
SI	ata/MP 13.1 - [Results]	alian Milandaria Unite
File	Edit Data Graphics Statis	sucs oser window Help
	Open Ctri+O	abel wariable age "Student Age"
	Save Ctrl+S	aber variabre age stadent nge
	Save As Ctrl+Shift+S	lo "C:\Users\013733~1\AppData\Local\Te
1	View	
-	Do	en female=.
1		
	Filename	missing values generated)
	Filename Change Working Directory	missing values generated)
	Filename Change Working Directory Log	<pre>missing values generated) , abel variable female "Female" ,</pre>
	Filename Change Working Directory Log Import	<pre>missing values generated) abel variable female "Female"</pre>
	Filename Change Working Directory Log Import Export	<pre>missing values generated) .abel variable female "Female" . Excel spreadsheet (*xls;*xlsx) . Text data (delimited, *.csv,)</pre>
	Filename Change Working Directory Log Import Export Print	<pre>missing values generated) abel variable female "Female" Excel spreadsheet (*.xls*.xlsx) Text data (delimited, *.csv,) Vext data in fixed format </pre>
	Filename Change Working Directory Log Import Export Print Example Datasets	<pre>missing values generated) abel variable female "Female" Excel spreadsheet (*xls;*xlsx) Text data (nfixed format Text data in fixed format Text data in fixed format with a dictionary</pre>
	Filename Change Working Directory Log Import Export Print Example Datasets Recent Datasets	<pre>missing values generated) abel variable female "Female" Excel spreadsheet (*xls*xlsx) Text data (delimited, *csr,) Text data in fixed format Text data in fixed format with a dictionary Unformatted text data</pre>
8	Filename Change Working Directory Log Import Export Print Example Datasets Recent Datasets Exit	<pre>missing values generated) abel variable female "Female" Excel spreadsheet (*xls;*xlsx) Text data (delimited, *csv,) Text data in fixed format Text data in fixed format with a dictionary Unformatted text data SAS XPORT</pre>
	Filename Change Working Directory Log Import Export Print Example Datasets Recent Datasets Exit Exit	<pre>missing values generated) abel variable female "Female" Excel spreadsheet (*xls*xlsx) Text data (delimited, *.csv,) Text data in fixed format Text data in fixed format Unformatted text data SAS XPORT Haver Analytics database</pre>
18 c	Filename Change Working Directory Log Import Export Print Example Datasets Recent Datasets Exit terror = Irop var1	<pre>missing values generated) abel variable female "Female" Excel spreadsheet (*xls*xlsx) Text data (delimited, *.csv,) Text data in fixed format with a dictionary Unformatted text data SAS XPORT Haver Analytics database ODBC data source</pre>

- If this is the case
- 1)Stata has an IMPORT function
- 2) Stat Transfer is another option

Transfer Variables Obser	rvations Options Run Program Log PSS Data File	About	
Input File Type:	PSS Data File		
Input File Type:	PSS Data File		
Input File Type: S	PSS Data File		
_			• _7
File Specification:			Browse
			View
	() () () () () () () () () () () () () (
Output File Type: 15	tata (Standard)		Version 12
File Specification:			✓ Browse
			Save Program
Transf	er Reset	Help	Exit

Objectives

- Understand tools to prepare for data analysis such as
 - Checking for entry error
 - Recoding missing values
 - Reverse coding variables
 - Recoding variables
 - Checking alpha
 - Creating composite variables



Checking for entry Error

Practice with opening NLSW data set

- Use "fre" command to scan all variables for entry error
 - fre age-tenure

Recoding Missing Values

- Pretend the grade variable had a few responses with a -9 response
- Use mvdecode
- mvdecode grade, mv(-9=.)

Reverse Coding Variables

- Imagine you have a variable called "safety" with response options
 - 1=Very safe 5 = Not safe at all
- You may choose to reverse code so that a higher score on safety reflects more feelings of safety
 - revrs safety
 - Will create a new variable called revsafety
 - The revsafety variable will now be reverse coded

Recoding variables

- Imagine you have a variable "packs" – number of packs smoked a day
 - 1 pack 50
 - 2 packs 20
 - 3 packs 10
 - 4 packs 2
- You may want to combine category 3 and 4; to do so
 - recode packs (1=1 "1 pack")(2=2 "2 packs") (3/4=3 "3+ packs"), generate (packs_re)

Checking alpha

- Imagine you have five items you think measure an underlying concept of happiness
 - happy1 happy2 happy3 happy4 happy5
- You want to check their internal consistency before you make a composite variable
 - alpha happy1-happy5, item asis

Creating composite variables

- Imagine those five items have a strong alpha, to create on composite variable
 - egen happiness =
 rowmean (happy1-happy5)



- Objectives
 - Understand why the Do-File is your new best friend



- What is a Do-File?
 - A text file in Stata where you can save (and run!) all your commands and notes
 - As opposed to typing everything in the command box
 - Allows you to go back to a dataset and quickly repeat your analysis
 - Why is this important?





• Opening your do-file

	2 - Stata/MP 13.1 - [Results]	C. C. & Connex (States (A. S.	Re Dane	 	
F	ile Edit Data Graphics Statistics	User Window Help				
lew C	; d a i 🛛 • L • 🗹 • 🖬 :	8 0 8				
Re	view New Do-	file Editor				~
ori #	Command _rc	· ·				
^{ox} 1	doedit "C:\Users\01373343					
read 2	use "C:\Program Files (x86)					
t Ite 3	fre age- tenure	•				
eter 4	do "C:\Users\013733~1\Ap					
5	mvdecode grade, mv (0=.)					
6	clear					
7	use "C:\Program Files (x86)					
ox 8	fre collgrad					
fts 9	fre c_city					
t Ite 10	fre wage					
eter 11	fre tenure					
k 12	sysuse auto.dta					
k E 13	sysuse census.dta					
:bo: 14	fre region					
Fe 15	findit rvrscode					
war 16	revrs region					
rch 17	revregion 199	•				
18	fre revregion					
19	use "C:\Program Files (x8b)					
20	fre race					
21	recode race (1=1) (2/3 = 2)					
22	fre face					
25	use "Cul Brogram Files (v96)	•				
24	recode race (1-1) (2/3-2)					
26	fre race re					
27	clear					
28	use "C:\Program Files (x86)					
29	recode race (1=1 "White") (
30	fre race_re					
31	alpha collgrad occupation	. ciear				_
32	clear					1
						-
		Command				ф.
		1				

_ 00-	The cartor - do_me_spring 2010	
File	Edit View Project Tools	
	🖥 🖳 🛤 🐰 🖻 🕲 🕐 🗠 । 🕂 🐨 🕹 🖗 📮	
do	_file_spring 2016 Untitled.do	
1	//Stata Workshop	
2	//A Gentle Introduction to Stata	
3		
4	//I. Background	
5		
6	//II. Getting Started	
7	//First, open your dataset	
8	use "C:\Program Files (x86)\Stata13\ado\base\n\nlsw88.dta", clear	
9		
10	//Use describe command	
11	describe	
12		
13	//A short Stata session	
14	summarize	
16		
17	// III Entering Data in Stata	
1.8	clear	
19	//Creating Variables (Continuous)	
20	gen id=.	
21	gen age = .	
22	//Creating Variables Lables	
23	label variable id "Student ID"	
24	label variable age "Student Age"	
25		
26	//Creating variables (Categorical)	
27	gen female=.	
28	label variable female "Female"	
29	//Creating labels for response options	
30	label define sex 0 "male" 1 "female"	
31	label values female sex	
32	clear	
33		
34	//IV. Preparing Data for Analysis	
35	use "u:\Frogram Files (X86)\Stata13\ado\base\n\n1sw88.dta", clear	
30	//Charling for antry arran	
30	fre age-tenure	
30	//Decoding missing variables	
40	//Recouring missing variable had a couple regnondents with a -0 regnone	-
10	mudecode grade mu (_0=)	~

- Running the Do-File
 - Highlight and Execute

File	Edit View Project Loois		ii}
	🗿 🕼 🕼 🕍 🔊 🗠 🖅 🐨 🕭 🖓 😱		
do	p file spring 2016 Untitled.do	- ×	
1	//Stata Workshop		· ·
2	//A Gentle Introduction to Stata		
3			
4	//I. Background		
5			
6	//II. Getting Started		
7	//First, open your dataset		
8	use "C:\Program Files (x86)\Stata13\ado\base\n\nlsw88.dta", clear		
9			
10	//Use describe command		· ·
11	describe	=	· ·
12			· ·
13	//A short Stata session		· ·
14	summarize		· ·
15			· ·
17	// III Ferencies Deep is Serve		
10	alogn		
10	//Creating Variables (Continuous)		
20	gen id=		
21	den ade =		
22	//Creating Variables Lables		
23	label variable id "Student ID"		
24	label variable age "Student Age"		
25			. clear
26	//Creating variables (Categorical)		
27	gen female=.		. do "C
28	label variable female "Female"		
29	//Creating labels for response options		. use "
30	label define sex 0 "male" 1 "female"		(NLSW,
31	label values female sex		
32	clear		
33			lena or
34	//IV. Preparing Data for Analysis		
35	use "C:\Program Files (x86)\Stata13\ado\base\n\nlsw88.dta", clear		<u> </u>
36			
37	//Checking for entry error		Command
38	fre age-tenure		Commente
39	//Recoding missing variables		
40	//Fretend the grade variable had a couple respondents with a -9 response	*	
1	mudacoda drada mu (-ve)	•	
_			

	- Vi	ariables		
		Variable	Label	
		idcode	NLS id	
		age	age in current year	
		race	race	
		married	married	
		never_married	d never married	
		grade	current grade completed	
		collgrad	college graduate	
		south	lives in south	
		smsa	lives in SMSA	
		c_city	lives in central city	
		industry	industry	
		union	union worker	
		ware	bourly ware	
		hours	usual hours worked	
		ttl exp	total work experience	
	P	roperties		
		Variables		
		Name		idcode
		Label		NLS id
clear		Туре		int
		Format		%8.0g
do "C:\Users\013733~1\AppData\Local\Temp\STD01000000.tmp"		Value Label		
		Notes		
use "C:\Program Files (x86)\Statal3\ado\base\n\nlsw88.dta", clear	Ξ	Data		
ALDW, 1900 EXCLACC)	E	Filename		nlsw88.dta
		Label		NLSW, 1988 extract
nd of do-file		Variables		17
	=	Observations		2.246
	*	Size		59.22K
		Memory		64M
		Sorted by		idcode

- Objectives
 - Calculate descriptive statistics
 - Construct graphical depictions of data



Descriptive Statistics

• What are examples of descriptive statistics?



Using Stata

- use "C:\Program Files (x86)\Stata13\ado\base \n\nlsw88.dta", clear
- summarize wage, detail

Descriptive Statistics - Output

• . summarize wage, detail

•			hourly was	ge		
•						
•		Percentiles	Smallest			
•	1%	1.930993	1.004952			
•	5%	2.801002	1.032247			
•	10%	3.220612	1.151368	Obs	2246	
•	25%	4.259257	1.344605	Sum of Wgt.	2246	
•	50%	6.27227		Mean	7.766949	
•			Largest	Std. Dev.	5.755523	
•	75%	9.597424	40.19808			
•	90%	12.77777	40.19808	Variance	33.12604	
•	95%	16.52979	40.19808	Skewness	3.096199	
•	99%	38.70926	40.74659	Kurtosis	15.85446	

Descriptive Statistics - Output

. sktest wage

Skewness/Kurtosis tests for Normality ------ joint -----Variable | Obs Pr(Skewness) Pr(Kurtosis) adj chi2(2) Prob>chi2 wage | 2.2e+03 0.0000 0.0000 . .

Descriptive Statistics - Graphs

- What kind of variable is "wage"?
- What graphic illustration is most appropriate

histogram wage, percent normal



VII. Statistics and Graphs – Two Categorical Variables

- Objectives
 - Conduct a cross-tabulation
 - Conduct a two-variable Chi-Square test



VII. Statistics and Graphs – Two Categorical Variables

Categorical Variables

 What are examples of categorical variables/response options?



NLSW Dataset

- Two examples of categorical variables
 - Married (Yes/No)
 - Race (White/Black/Other)
- Research Question
 - Is there an association between race and marriage?

VII. Statistics and Graphs – Two Categorical Variables

Stata Command

• tabulate race married, chi2 expected row

Stata Output

tabulate race married, chi2 expected row

Key
frequency
expected frequency
row percentage

	mar		
race	single	married	Total
white	487	1,150	1,637
	586.0	1,051.0	1,637.0
	29.75	70.25	100.00
black	309	274	583
	208.7	374.3	583.0
	53.00	47.00	100.00
other	8	18	26
	9.3	16.7	26.0
	30.77	69.23	100.00
Total	804	1,442	2,246
	804.0	1,442.0	2,246.0
	35.80	64.20	100.00

Pearson chi2(2) = 101.4215 Pr = 0.000

VIII. Tests for one or two means

- Objectives
 - Conduct a one-sample test of means
 - Conduct a two-sample test of group means



VIII. Tests for one or two means

One-sample test of means

- Let's pretend the average hourly compensation rate in the U.S. in 1988 was \$10
 - You want to know if the average hourly rate wage for your sample is equal to national average

• ttest wage == 10

Output

. ttest wage == 10

• One-sample t test

•							
•	Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
•	+- wage	2246	7.766949	.1214451	5.755523	7.528793	8.005105
•	mean = :	mean(wage)			 t :	
•	Ho: mean =	10			degrees	of freedom :	= 2245
•	Ha: mea	n < 10		Ha: mean != 3	10	Ha: m	ean > 10

 $Pr(T < t) = 0.0000 \qquad Pr(|T| > |t|) = 0.0000 \qquad Pr(T > t) = 1.0000$



VIII. Tests for one or two means

Two-sample test of group means

• Let's say we want to know if the mean wage is the same for single versus married participants

ttest wage, by (married)

Output

. ttest wage, by (married)

Pr(T < t) = 0.9732

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
single married	804 1442	8.080765 7.591978	.223456 .1421835	6.336071 5.399229	7.642138 7.313069	8.519392 7.870887
combined	2246	7.766949	.1214451	5.755523	7.528793	8.005105
diff		.4887873	.2531718		0076882	.9852627
diff = Ho: diff =	= mean(sing: = 0	le) – mean(m	arried)	degrees	t = of freedom =	= 1.9307 = 2244
Ha: di	iff < 0		Ha: diff !=	0	Ha: di	iff > 0

Pr(|T| > |t|) = 0.0537



Pr(T > t) = 0.0268

Objectives

- Construct a scattergram
- Calculate correlation
- Estimate a regression model



Example

- You are interested in examining the relationship between total work experience (ttl_exp) and hourly wage (wage)
- You hypothesize that wage is dependent on work experience



Visual Learners

- May want to see a scattergram
- scatter wage ttl_exp



Visual Learners

- Some may want to see the fitted line through the scattergram
- twoway (lfit wage ttl_exp) (scatter wage ttl_exp)



٠

By the Numbers

- May prefer the correlation matrix
- pwcorr wage ttl_exp, obs sig star(0.05)

By The Numbers



. regress wage ttl exp

Is the IV associated with the DV?

- Stata can run various regressions **Output** based on the distribution
- If we assume the DV has a normal distribution, simply use the following format:

regress dv iv

regress wage ttl exp

Source | SS df MS Number of obs = 2246 F(1, 2244) = 170.14Prob > F Model | 5241.29609 1 5241.29609 = 0.0000 Residual | 69126.6713 2244 30.805112 R-squared = 0.0705 Adj R-squared = 0.0701 = 5.5502Total | 74367.9674 2245 33.1260434 Root MSE wage | Coef. Std. Err. t P>|t| [95% Conf. Interval] .3314291 .0254087 13.04 .2816021 ttl exp | 0.000 .3812562 cons | 3.612492 .3393469 10.65 0.000 2.947026 4.277959

X. Multiple Regression

- Objectives
 - Estimate a multiple regression model



IX. Multiple Regression

Example

- You are interested in examining correlates of wage
- In addition to work experience, you think job tenure (in years), and whether someone is a college grad may also be associated with wage







X. Multiple Regression

Are the IVs associated with the DV?

Output

- If we assume the DV has a normal distribution, simply use the following format:
- regress dv iv1 iv2 ...ivn

regress wage ttl_exp
tenure collgrad

regress wage ttl_exp tenure collgrad

Source	SS	df	MS		Number of obs	= 2231
	+				F(3, 2227)	= 108.04
Model	9414.34005	3 3138	3.11335		Prob > F	= 0.0000
Residual	64687.4876	2227 29.0	469185		R-squared	= 0.1270
	+				Adj R-squared	= 0.1259
Total	74101.8276	2230 33.2	295191		Root MSE	= 5.3895
wage	Coef.	Std. Err.	t	 P> t	[95% Conf.	Interval]
wage	Coef.	Std. Err.	t	P> t	[95% Conf.	Interval]
wage ttl_exp	Coef. +	Std. Err.	t 9.01	P> t 0.000	[95% Conf. 2142298	Interval] .3334983
wage ttl_exp tenure	Coef. + .273864 .0321168	Std. Err. .0304097 .0253609	t 9.01 1.27	P> t 0.000 0.206	[95% Conf. .2142298 0176167	Interval] .3334983 .0818502
wage ttl_exp tenure collgrad	Coef. + .273864 .0321168 3.251781	Std. Err. .0304097 .0253609 .2698484	t 9.01 1.27 12.05	<pre>P> t 0.000 0.206 0.000</pre>	[95% Conf. .2142298 0176167 2.7226	Interval] .3334983 .0818502 3.780962
wage ttl_exp tenure collgrad _cons	Coef. +	Std. Err. .0304097 .0253609 .2698484 .3397438	t 9.01 1.27 12.05 9.98	<pre>P> t 0.000 0.206 0.000 0.000 0.000</pre>	[95% Conf. .2142298 0176167 2.7226 2.72339	Interval] .3334983 .0818502 3.780962 4.055886

- Objectives
 - Estimate a simple logistic regression model
 - Estimate a multiple logistic regression model



Simple Logistic Regression

- You are interested in examining whether being a union worker (union) is dependent on being a college graduate (collgrad)
- What is the dependent variable here?
- How is it measured?



Simple Logistic Regression

. logistic union collgrad

• logistic union collgrad

Logistic regre	Number	of obs	=	1878			
				LR chi	2(1)	=	17.30
				Prob >	chi2	=	0.0000
Log likelihood	= -1037.974	8		Pseudo	R2	=	0.0083
union	Odds Ratio	Std. Err.	Z	P> z	[95% Co	onf.	Interval]
+							
collgrad	1.64747	.1951016	4.22	0.000	1.30621	13	2.077883
_cons	.284287	.0182102	-19.64	0.000	.250745	52	.3223157

Multiple Logistic Regression

- You are interested in examining whether being a union worker (union) is dependent on being a college graduate (collgrad), age (age), and race/ethnicity (race)
- What is the dependent variable here?
- How is it measured?
- How are the IVs measured?



Multiple Logistic Regression . xi: logistic union collgrad age i.race

• xi: logistic union collgrad age i.race

i.race	(natura)	lly coded	; _Irace_	1 omi	tted)		
Logistic regr	ession			Numb	er of obs	=	1878
				LR c	hi2(4)	=	33.72
				Prob	> chi2	=	0.0000
Log likelihoo	d = -1029.	7628		Pseu	do R2	=	0.0161
union	Odds Rat	tio Std.Err	• Z	P> z	[95%	Conf.	Interval]
	+						
collgrad	1.7347	.2081249	4.59	0.000	1.371	228	2.194602
age	1.0137	.0181203	0.76	0.445	.9788	438	1.049889
_Irace_2	1.603	.191092	3.97	0.000	1.26	996	2.025852
_Irace_3	1.687	.7450893	1.18	0.236	.7100	424	4.009309
_cons	.14222	.1008741	-2.75	0.006	.0354	229	.5710734

XII. A Public Health Example

Prescription Stimulant Misuse



Background

- Prescription Stimulant Misuse: use of stimulants without a valid prescription, use in excess of a prescription, and/or use for purposes other than prescribed
- Health-related complications and Misuse
 - Increased blood pressure
 - Increased heart rate
 - Addiction
 - Psychosis
- Disproportionately affects college population
 - Research exploring prevalence and correlates important for prevention and intervention

XII. A Public Health Example

Misuse of Prescription Stimulants



Pilot Study

- Random Sampling
 - One Stage Cluster Sample
 - 19 classes
 - 94.71% response proportion
 - Analytic N = 499 students
 - Sample representative of undergraduate population
 - Participants completed 100-item BEACH-Q questionnaire

XII. A Public Health Example

Data Entry

- Survey
- Codebook
- Excel
- Import

Once in Stata

- Variable response options
- Variable labels
- Descriptive Data
- Inferential Data

XIII. What's Next

- Tobit regression
- Poisson regression
- Nested regression
- Factor analysis
- Longitudinal data analysis
- Hierarchical Models
- Structural Equation Modeling





XIII. What's Next

Resources

- Stata textbooks
- UCLA IDRE





Thank You!



Before & After Stata



A Gentle Introduction to Stata

Sixth Edition

ALAN C. ACOCK



Acknowledgements

- Dr. Alan Acock, Author of "A Gentle Introduction to Stata"
- CSULB BUILD Program (NIH Award #RL5GM118978)

