

B) Stata interface (with data and commands, windows)

Command	Explanation
clear	cleans the previous data set <i>without</i> saving the data
input x	creates a new numeric variable called "x"
1	These are 3 values for x
2	
3	
end	end: the introduction of data has finished
<u>summarize</u> x	Descriptive statistics for variable x (the underlined text is enough to make the command work)

1. Description: summarize, detail

Command	Explanation
<u>set</u> more off, perm	Suppress the repeated presence of the word " -more- " (it will annoy you by tirelessly requesting that you may press the space bar to continue)
clear	Clears the previous data base
version 12	Do this only if you are using version 13 or higher, to keep the routines (specially the generation of random values) as they were in version 12.
<u>set</u> <u>obs</u> 100	Creates 100 empty rows
<u>set</u> <u>seed</u> 333	Set a random seed: new data will be randomly generated, but they will be the same when you run again this sequence of commands
<u>generate</u> speed=rnormal(80,20)	Generates a new random variable called "speed" with mean = 80 and standard deviation=20 (approximately)
replace speed=20 in 1 replace speed=33 in 2 replace speed=180 in 100	You have changed the values for the variable speed • in row 1, speed=20 • in row 2, speed=33 • in row 100, speed=180
<u>summarize</u> speed, <u>detail</u>	Descriptive statistics are requested

Please skip *always* this step if you are using versions 11 or older

4. Labels and codes of variables: label codebook

```
clear
input ///
smk  weight  sex
0    66      0
0    72      0
0    78      1
1    79      0
1    80      1
1    81      1
2    68      0
2    70      0
2    72      1
end
```

5. Categorical variables I: tabulate tables tabstat

```
clear
input ///
smoking weight
0    66
0    72
0    78
1    79
1    80
1    81
2    68
2    70
2    72
end
```

6. Categorical variables II: tab1 tab2

```
input ///
sex    smoke OH cancer    n
1      1      1      1    50
0      1      1      1    40
1      0      1      1    10
0      0      1      1     8
1      1      0      1    25
0      1      0      1    20
1      0      0      1     5
0      0      0      1     4

1      1      1      0    650
0      1      1      0    860
1      0      1      0    690
0      0      1      0    892
1      1      0      0    675
0      1      0      0    880
1      0      0      0    695
0      0      0      0    896
end
```

7. Categorization and quantiles: recode, xtile

```
clear
version 12
set obs 500
set seed 1234
gen BMI=rnormal(24,4)
replace BMI=20 if BMI>19.5 & BMI<20.5
```

8. Histograms: hist

```
clear
version 12
set obs 500
set seed 1234
gen x=rnormal(100,10)
```

9. Boxplots: graph box

```
clear
version 12
set obs 500
set seed 1234
gen x=rnormal(100,10)
```

10. Scatter plots: twoway

```
clear
input ///
hours      weight_chg
0          -1
0          0
1          0
2          .5
3          2.5
4          2
8          3
end
```

```
clear
input ///
str15 country Chocolate      Nobel
China      0.700      0.060
Brazil     2.900      0.050
Spain      3.600      1.701
Poland          3.600      3.124
Australia     4.500      5.451
Canada        3.900      6.122
Netherlands  4.500     11.356
USA           5.300     10.770
Ireland       8.800     12.706
Germany      11.600     12.668
UK           9.700     18.875
Switzerland 11.900     31.544
Sweden       6.480     31.855
end
```

11. More on twoway: bar line

```
clear
input ///
mark n
3      1
4      5
6      4
7      7
8      10
9      12
10     4
end
```

```
clear
input ///
lowfatdai averageSBP
1          140
2          132
3          128
4          122
5          110
end
```

12. Plots with confidence intervals and error bars: rcap

```
clear
version 12
set obs 500
set seed 999
g age=round(100*runiform())
replace age=age+20 if age<20
g SBP=95+(.4*age)+rnormal(0,15)
```

```
clear
input Q4 RR low upper
1 1 1 1
2 0.8 0.4 1.6
3 0.6 0.4 0.9
4 0.5 0.3 0.833
end
```

13. Text in graph, fine tuning of graphs

```
clear
input Q4 RR low upper
  1  1  1  1
  2 0.8  0.4  1.6
  3 0.6  0.4  0.9
  4 0.5  0.3  0.833
end
```

16. Dates and strings: mdy format substr display

```
clear
input ///
str10 name      day month year
      Mary_Ann      3      3      1981
      John           1      12     1979
      Peter          23      1      2009
      Grandpa        26      6      1927
end
```

17. Algebra: return list scalar + - * / ^

```
clear
version 12
set obs 500
set seed 999
g age=round(100*runiform())
```

20. Distributions: uniform binomial poisson

```
clear
version 12
set obs 1000
set seed 333
g x = uniform()
hist x, freq w(.2)
```

22. Graphics to assess

```
clear
version 12
se obs 999
se seed 1234
```

24. Hypothesis testing and p values

```
clear
input ///
married case      n
1         1       10
0         1       20
1         0       90
0         0       80
end
expand n
```

25. Testing normality: sktest, ladder, gladder

```
clear
se obs 150
se seed 1234
g z=rnormal(0,1)
g x=0.1+round(4*uniform()) in 1/50
replace x=abs(exp(rnormal(0.8,1.2))) if x==.
```

*in versions >13, you'll need to add version 12.0 at the beginning

```
clear
se obs 150
se seed 1234
g a=ceil(10*uniform())
g b=rnormal(20,4)
g x=a+b+abs(exp(rnormal(0.1,1)))
```

26. Comparing proportions: chi2, exact

```
clear
input ///
married case      n
1         1       10
0         1       20
1         0       90
0         0       80
end
expand n
```

```

clear
input ///
genotype case      n
1         1        5
1         0        1
0         0       11
end
expand n

```

27. Categorical variables, epidemiologic analyses: cs

```

clear
input ///
smoke  OH      cancer  n
0      0      0      1591
0      1      0      1582
0      0      1       9
0      1      1      18

1      0      0      1555
1      1      0      1510
1      0      1       45
1      1      1       90
end
expand n

```

28. Case-control studies, matched case-control: cc mcc

```

clear
input ///
case exposure      n
1         1       50
1         0       50
0         1       20
0         0       80
end
expand n
g sex=1 in 25/75
replace sex=1 in 110/175
replace sex=0 if sex==.

```

```

clear
input ///
exp_case exp_contr  n
1         1       10
1         0       75
0         1       25
0         0      110
end
expand n

```

29. Comparison between two means: ttest sdtest

```
clear
version 12.0
set obs 300
g group=_n>120
set seed 1234
g bmi=rnormal(25,4) in 1/120
replace bmi=rnormal(24,4) in 121/300
```

```
version 12
set seed 1234
g e=rnormal(0,2)
g bmi_2yr=bmi+0.25+e
```

30. Non-parametric tests: ranksum signrank

```
clear
version 12
set obs 300
g group=_n>120
set seed 1234
g bmi=rnormal(25,4) in 1/120
replace bmi=rnormal(24,4) in 121/300
```

```
version 12
set seed 1234
g e=rnormal(0,2)
g bmi_2yr=bmi+0.25+e
```

31. Comparison between >2 means: oneway kwallis

```
clear
input ///
x      group
4      1
4      1
6      1
8      1
8      1

4      2
8      2
8      2
8      2
8      2

8      3
8      3
10     3
12     3
12     3
end
```

32. Simple linear regression: regress

```
clear
input ///
hours      weight_chg
0          -1
0          0
1          0
2          .5
3          2.5
4          2
8          3
end
```

35. Multiple regression, dummy variables

```
clear
input ///
w_chg      ses    smk    age
9          1      1      68
8          1      1      65
7          1      1      65
6          1      1      63
3          2      0      71
3          2      1      71
0          1      1      58
-2         4      1      53
-3         2      0      49
-5         2      2      52
-5         3      0      48
-5         4      2      55
-6         4      0      52
-8         3      2      45
end
```

36. Factorial anova, repeated measures anova

```
clear
input ///
w_chg      ses    smk    age
9          1      1      68
8          1      1      65
7          1      1      65
6          1      1      63
3          2      0      71
3          2      1      71
0          1      1      58
-2         4      1      53
-3         2      0      49
-5         2      2      52
-5         3      0      48
-5         4      2      55
-6         4      0      52
-8         3      2      45
end
```

clear				
input ///				
id	time	weight	diet	
1	1	76	1	4
1	2	65	1	4
1	3	63	1	4
				5
2	1	82	1	5
2	2	70	1	5
2	3	68	1	6
				6
3	1	80	1	6
3	2	78	1	6
3	3	70	1	6
				end

37. ANCOVA: margins marginsplot

clear			
input ///			
SBP	MedDiet	age	
145	12	68	
165	9	65	
153	10	66	
105	14	63	
115	5	45	
123	6	38	
98	13	43	
133	3	35	
160	5	69	
133	7	52	
140	6	48	
102	12	35	
166	4	72	
120	7	55	
end			

38. Transforming columns/rows: reshape

clear				
input ///				
id	weight1	weight2	weight3	diet
1	76	65	63	1
2	82	70	68	1
3	80	78	70	1
4	83	83	84	2
5	79	80	79	2

```
6      84      84      84      2
end
```

```
clear
input ///
subject      SBP      hour
1           120        9
1           125       12
1           121       18
1           116       24
2           118        9
2           122       12
2           122       18
2           124       24
end
```

39. Protecting your data: preserve restore

```
clear
sysuse citytemp4.dta
```

```
clear
input ///
id weight1 weight2 weight3 diet
1    76      65      63      1
2    82      70      68      1
3    80      78      70      1
4    83      83      84      2
5    79      80      79      2
6    84      84      84      2
end
```

41. Logistic regression: logistic logit

```
clear
input ///
OH car_crash      n
0   0           50
0   1           10
1   0           40
1   1           20
end
expand n
```

```
clear
sysuse cancer
```

42. ROC curves: roctab roccomp

```
clear
sysuse cancer
quietly logistic died i.drug age
predict p_death
```

```
clear
sysuse cancer
quietly logistic died age
predict p_death1
quietly logistic died i.drug age
predict p_death2
```

44. Kaplan-Meier curves: stset sts graph

```
clear
sysuse cancer
```

44. Confounding and interaction: est store lrtest

```
clear
input ///
smoke alcohol      CHD      n
1          1          1      400
1          0          1      120
1          1          0     7600
1          0          0     1880
0          1          1       30
0          0          1       84
0          1          0     2970
0          0          0     6916
end
expand n
```

```
clear
input ///
FVL      OC      DVT      n
1          1          1       61
1          0          1        6
1          1          0      710
1          0          0     940

0          1          1       30
0          0          1        8
0          1          0     9700
0          0          0     9920
end
expand n
```

45. Loops: foreach

```
clear
```

```

input ///
date2 date4 date6 date8 date10 date12 date14
15500 16265 17008 17770 18518 19288 20032
14850 15590 16330 17070 17810 18550 19290
14950 15601 16430 17179 17925 18550 .
16100 16800 17500 18000 18500 . .
17290 17903 20431 20655 . . .
14911 15601 . 18001 . . .
15120 16043 . . . . .
19999 . . . . .
end
format date* %td

```