

Problem Set 1 Solutions

Note: Text that is preceded by a “.” is the Stata code used in the analysis. Text enclosed in “*”s explains what each piece of code is doing. Where relevant, I have pasted the actual Stata output.

Part I

Using semicolon as delimiter

```
#delimit;
```

```
. log using ps1.log
```

```
. set more off
```

```
. cd "/Users/nlmiller/Desktop/Poli Sci Lab/PS1/"
```

Inputting Data

```
. insheet using "/Users/nlmiller/Desktop/Poli Sci Lab/PS1/scores.dat"
```

```
. list
```

```
+-----+-----+-----+-----+
|   name   |   age   |   test1  |   test2  |
+-----+-----+-----+-----+
1.   Bob    |    18   |    95    |    18    |
2.  Carol   |    21   |    43    |    27    |
3.   Ted    |    14   |    67    |    9     |
4.  Alice   |    12   |    23    |    31    |
+-----+-----+-----+-----+
```

```
. save "/Users/nlmiller/Desktop/Poli Sci Lab/PS1/scores.dta"
```

Part II

```
. clear
```

```
. use "/Users/nlmiller/Desktop/Poli Sci Lab/PS1/spaesubset2012.dta"
```

Describe variables to find which measures length of line

```
. describe
```

```
      obs:      10,200
     vars:         5
     size:    163,200 (99.9% of memory free)
                                8 Feb 2013 21:46
```

```
-----+-----+-----+-----+-----+
variable name   storage  display  value  variable label
                type    format    label
-----+-----+-----+-----+-----+
```

```

weight          double %10.0g          Case weights
regstate        byte   %8.0g          REGSTATE   State of Voter Registration
race            byte   %8.0g          RACE       Race
q4              byte   %8.0g          Q4         Mode of voting
q10             byte   %8.0g          Q10       Line length

```

Sorted by:

Tabulate line length variable to see its values/coding scheme

. tab q10, m

Line length	Freq.	Percent	Cum.
Not at all	3,037	29.77	29.77
Less than 10 minutes	2,394	23.47	53.25
10-30 minutes	1,375	13.48	66.73
31 minutes ñ 1 hour	593	5.81	72.54
More than 1 hour	211	2.07	74.61
I don't know	24	0.24	74.84
.	2,566	25.16	100.00
Total	10,200	100.00	

. tab q10, m nol

Line length	Freq.	Percent	Cum.
1	3,037	29.77	29.77
2	2,394	23.47	53.25
3	1,375	13.48	66.73
4	593	5.81	72.54
5	211	2.07	74.61
6	24	0.24	74.84
.	2,566	25.16	100.00
Total	10,200	100.00	

Generate a variable that equals 1 if the respondent reported waiting more than 30 minutes (when the variable= 4 or 5)

. gen thirtyplus_2012=1 if q10==4|q10==5.

Coding the variable as zero when the respondent waited 30 minutes or less

. replace thirtyplus_2012=0 if q10<=3

Collapse to create a dataset that records the mean of thirtyplus_2012 (equivalent to the percent who waited more than 30 minutes since it's a binary variable) by registration state, as well as how many observations were used in this calculation by counting the number of observations for thirtyplus_2012 (including analytical weights)

. collapse (mean) thirtyplus_2012 (count) n = thirtyplus_2012 [aweight=weight],by(regstate)

Listing the first 10 observations

```
. list in 1/10
```

```
+-----+
| regstate   thi~2012   n |
+-----+-----+-----+
| 1.   Alabama   .0760943  174 |
| 2.   Alaska    .0222348  165 |
| 3.   Arizona   .0761444   67 |
| 4.   Arkansas  .1117     170 |
| 5.   Californ .0210003   95 |
+-----+-----+-----+
| 6.   Colorado  .0249651   67 |
| 7.   Connecti .0667891  171 |
| 8.   Delaware  .0139119  190 |
| 9.   District  .3852806  178 |
| 10.  Florida   .3867461  131 |
+-----+-----+-----+
```

```
. save "/Users/nlmiller/Desktop/Poli Sci Lab/PS1/wait_by_state_2012.dta"
```

Part III

```
. clear
```

```
. use "/Users/nlmiller/Desktop/Poli Sci Lab/PS1/spaesubset2008.dta"
```

Describe variables to find which measures length of line

```
. describe
```

```
Contains data from /Users/nlmiller/Desktop/Poli Sci Lab/PS1/spaesubset2008.dta
  obs:      10,000
  vars:      5                               8 Feb 2013 22:03
  size:     120,000 (99.9% of memory free)
```

```
-----+-----+-----+-----+-----+
| variable name | storage | display | value | variable label |
|-----+-----+-----+-----+-----+
| weight        | float   | %9.0g   |       | case weight    |
| q5            | byte    | %8.0g   | q5    | mode of voting |
| q12           | byte    | %8.0g   | q12   | line length    |
| race          | byte    | %8.0g   | race  | race           |
| inputstate    | byte    | %8.0g   | inputstate | state of residence |
+-----+-----+-----+-----+-----+
```

Tabulate line length variable to see its values/coding scheme

```
. tab q12, m
```

line length	Freq.	Percent	Cum.
not at all	3,264	32.64	32.64
less than 10 minutes	2,138	21.38	54.02
10-30 minutes	1,301	13.01	67.03
31minutes - 1 hour	710	7.10	74.13
more than 1 hour	365	3.65	77.78
don't know	2	0.02	77.80
.	2,220	22.20	100.00
Total	10,000	100.00	

. tab q12, m nol

line length	Freq.	Percent	Cum.
1	3,264	32.64	32.64
2	2,138	21.38	54.02
3	1,301	13.01	67.03
4	710	7.10	74.13
5	365	3.65	77.78
6	2	0.02	77.80
.	2,220	22.20	100.00
Total	10,000	100.00	

Generate a variable that equals 1 if the respondent reported waiting more than 30 minutes (when the variable= 4 or 5)

. gen thirtyplus_2008=1 if q12==4|q12==5

Coding the variable as zero when the respondent waited 30 minutes or less

. replace thirtyplus_2008=0 if q12<=3

Collapse to create a dataset that records the mean of thirtyplus_2008 by registration state as well as how many observations were used in this calculation by counting the number of observations of thirtyplus_2008. (including analytical weights)

. collapse (mean) thirtyplus_2008 (count) n= thirtyplus_2008 [aweight=weight], by(inputstate)

Listing the first 10 observations

. list in 1/10

	inputs~e	thi~2008	n
1.	alabama	.135972	175
2.	alaska	.0192774	169
3.	arizona	.239623	92
4.	arkansas	.2076403	178
5.	californ	.0814993	97
6.	colorado	.123161	73
7.	connecti	.0963251	181

```

8. | delaware      .102089   187 |
9. |  florida      .2867942   144 |
10. | georgia       .3538881   171 |
+-----+

```

Renaming input state variable to match name of variable from 2012 dataset (to merge, a common identifier is needed)

```
. rename inputstate regstate
```

```
. save "/Users/nlmiller/Desktop/Poli Sci Lab/PS1/wait_by_state_2008.dta"
```

Merging datasets using regstate identifier. Note below that 1 observation is not matched because there is no data for the District of Columbia in the 2008 dataset

```
. merge 1:1 regstate using "/Users/nlmiller/Desktop/Poli Sci Lab/PS1/wait_by_state_2012.dta"
```

Result	# of obs.
not matched	1
from master	0 (_merge==1)
from using	1 (_merge==2)
matched	50 (_merge==3)

List states where the proportion that waited 30+ minutes was greater in 2012 than in 2008

```
. list regstate if thirtyplus_2012>thirtyplus_2008
```

```

+-----+
| regstate |
+-----+
2.  alaska
9.  florida
11. hawaii
15. iowa
20. maryland
+-----+
21. massachu
22. michigan
26. montana
29. new hamp
32. new york
+-----+
34. north da
39. rhode is
41. south da
+-----+

```

```
. save "/Users/nlmiller/Desktop/Poli Sci Lab/PS1/wait_by_state_merge.dta"
```

Part IV

```
. clear
```

```
. use "/Users/nlmiller/Desktop/Poli Sci Lab/PS1/spaesubset2012.dta"
. describe
```

```
Contains data from /Users/nlmiller/Desktop/Poli Sci Lab/PS1/spaesubset2012.dta
  obs:      10,200
  vars:      5                               8 Feb 2013 21:46
  size:      163,200 (99.9% of memory free)
```

```
-----
```

variable name	storage type	display format	value label	variable label
weight	double	%10.0g		Case weights
regstate	byte	%8.0g	REGSTATE	State of Voter Registration
race	byte	%8.0g	RACE	Race
q4	byte	%8.0g	Q4	Mode of voting
q10	byte	%8.0g	Q10	Line length

```
-----
```

```
Sorted by:
```

```
*Tabulate line length and voting method variables to see their values/coding scheme*
```

```
. tab q4, m
```

Mode of voting	Freq.	Percent	Cum.
In person on Election Day (at polling p	5,979	58.62	58.62
In person before Election Day (early)	1,654	16.22	74.83
Voted by mail (or absentee)	1,770	17.35	92.19
I don't know	27	0.26	92.45
.	770	7.55	100.00
Total	10,200	100.00	

```
. tab q4, m nol
```

Mode of voting	Freq.	Percent	Cum.
1	5,979	58.62	58.62
2	1,654	16.22	74.83
3	1,770	17.35	92.19
4	27	0.26	92.45
.	770	7.55	100.00
Total	10,200	100.00	

```
. tab q10, m
```

Line length	Freq.	Percent	Cum.
Not at all	3,037	29.77	29.77
Less than 10 minutes	2,394	23.47	53.25
10-30 minutes	1,375	13.48	66.73
31 minutes ñ 1 hour	593	5.81	72.54
More than 1 hour	211	2.07	74.61
I don't know	24	0.24	74.84
.	2,566	25.16	100.00

Total | 10,200 100.00

```
. tab q10, m nol
```

Line length	Freq.	Percent	Cum.
1	3,037	29.77	29.77
2	2,394	23.47	53.25
3	1,375	13.48	66.73
4	593	5.81	72.54
5	211	2.07	74.61
6	24	0.24	74.84
.	2,566	25.16	100.00
Total	10,200	100.00	

Generate a variable that equals 1 if the respondent reported waiting more than 30 minutes (when the variable= 4 or 5)

```
. gen thirtyplus_2012=1 if q10==4|q10==5.
```

Coding the variable as zero when the respondent waited 30 minutes or less

```
. replace thirtyplus_2012=0 if q10<=3
```

Collapse to create a dataset with the mean of thirtyplus_2012 by different types of voting and state, as well as a count of the number of observations with data available

```
. collapse (mean) thirtyplus_2012 (count) n=thirtyplus_2012 [aweight=weight], by(regstate q4)
```

Drop unnecessary observations where mode of voting is missing and where mode of voting is not on election day or early voting

```
. drop if q4==.| q4>=3
```

Reshape to create a state-level dataset with separate variables for thirty_plus2012 by each mode of voting

```
. reshape wide thirtyplus_2012 n, i(regstate) j(q4)
```

(note: j = 1 2 3 4)

Data	long	->	wide
Number of obs.	169	->	51
Number of variables	4	->	9
j variable (4 values)	q4	->	(dropped)
xij variables:			

```

thirtyplus_20122 ...          thirtyplus_2012  ->  thirtyplus_20121
                                data          ->  data1 data2 ... data4
-----

```

. list

	regstate	th~20121	n1	th~20122	n2
1.	Alabama	.0760943	174	.	.
2.	Alaska	0	136	.1432126	29
3.	Arizona	.0881236	61	0	6
4.	Arkansas	.0971476	87	.1297159	83
5.	Californ	.0213493	94	0	1
6.	Colorado	.0436208	37	0	30
7.	Connecti	.0546858	168	.6057609	3
8.	Delaware	.0140393	188	0	2
9.	District	.3842785	129	.3883336	49
10.	Florida	.3249482	66	.4485258	65
11.	Georgia	.1564134	98	.202052	72
12.	Hawaii	.02441	72	.0976297	28
13.	Idaho	.0366039	124	.077841	23
14.	Illinois	.0470069	132	.2011549	49
15.	Indiana	.1627559	132	.0687497	39
16.	Iowa	.0658377	100	0	30
17.	Kansas	.0871042	121	.1569817	37
18.	Kentucky	.0645977	162	0	10
19.	Louisian	.1623183	131	.3101856	32
20.	Maine	.0115224	120	0	25
21.	Maryland	.2777531	134	.5463478	39
22.	Massachu	.0571125	178	0	4
23.	Michigan	.2343819	127	.5325436	4
24.	Minnesot	.0398434	169	0	3
25.	Mississi	.0601983	154	.1571319	13
26.	Missouri	.0864275	156	0	11
27.	Montana	.145807	82	.4420979	6
28.	Nebraska	.0410909	126	0	15
29.	Nevada	.0309614	51	.0767755	117
30.	New Hamp	.0858861	182	.	.
31.	New Jers	.0152495	159	0	5
32.	New Mexi	.0600595	61	.0154869	97
33.	New York	.0687448	161	0	5
34.	North Ca	.0493978	75	.2250464	96
35.	North Da	.0705447	66	0	35
36.	Ohio	.1329663	97	.0861069	41
37.	Oklahoma	.1642924	155	.159851	19
38.	Oregon	0	5	0	4
39.	Pennsylv	.0917703	175	.	.
40.	Rhode Is	.0928268	171	0	6
41.	South Ca	.2866086	140	.3107851	20
42.	South Da	.0285281	125	0	44
43.	Tennesse	.170697	70	.1314724	101
44.	Texas	.0625682	54	.1217225	120

45.	Utah	.0820119	91	.0392503	53

46.	Vermont	.00388	126	0	32
47.	Virginia	.3114677	149	.0625045	23
48.	Washingt	0	2	0	5
49.	West Vir	.1088866	113	.0157209	56
50.	Wisconsi	.0522848	123	.0111778	44

51.	Wyoming	.0187056	149	0	20

. log close