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*

```stata
#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
```

<table>
<thead>
<tr>
<th>name</th>
<th>age</th>
<th>test1</th>
<th>test2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bob</td>
<td>18</td>
<td>95</td>
<td>18</td>
</tr>
<tr>
<td>Carol</td>
<td>21</td>
<td>43</td>
<td>27</td>
</tr>
<tr>
<td>Ted</td>
<td>14</td>
<td>67</td>
<td>9</td>
</tr>
<tr>
<td>Alice</td>
<td>12</td>
<td>23</td>
<td>31</td>
</tr>
</tbody>
</table>

.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)

```

<table>
<thead>
<tr>
<th>variable name</th>
<th>type</th>
<th>format</th>
<th>label</th>
<th>variable label</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
*Tabulate line length variable to see its values/coding scheme*

```
. tab q10, m

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

Total | 10,200 | 100.00 |
```

```
. tab q10, m nol

<table>
<thead>
<tr>
<th>Line length</th>
<th>Freq.</th>
<th>Percent</th>
<th>Cum.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3,037</td>
<td>29.77</td>
<td>29.77</td>
</tr>
<tr>
<td>2</td>
<td>2,394</td>
<td>23.47</td>
<td>53.25</td>
</tr>
<tr>
<td>3</td>
<td>1,375</td>
<td>13.48</td>
<td>66.73</td>
</tr>
<tr>
<td>4</td>
<td>593</td>
<td>5.81</td>
<td>72.54</td>
</tr>
<tr>
<td>5</td>
<td>211</td>
<td>2.07</td>
<td>74.61</td>
</tr>
<tr>
<td>6</td>
<td>24</td>
<td>0.24</td>
<td>74.84</td>
</tr>
<tr>
<td></td>
<td>2,566</td>
<td>25.16</td>
<td>100.00</td>
</tr>
</tbody>
</table>

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

+---------------------------+
<table>
<thead>
<tr>
<th>regstate   thi-2012     n</th>
</tr>
</thead>
</table>
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)
----------------------------------------------------------------
storage  display     value
variable name   type   format      label      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
. tab q12, m nol

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

*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

<table>
<thead>
<tr>
<th>inputs-e</th>
<th>thi-2008</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>alabama</td>
<td>0.135972</td>
<td>175</td>
</tr>
<tr>
<td>alaska</td>
<td>0.019277</td>
<td>169</td>
</tr>
<tr>
<td>arizona</td>
<td>0.239623</td>
<td>92</td>
</tr>
<tr>
<td>arkansas</td>
<td>0.207640</td>
<td>178</td>
</tr>
<tr>
<td>californ</td>
<td>0.081499</td>
<td>97</td>
</tr>
<tr>
<td>colorado</td>
<td>0.123161</td>
<td>73</td>
</tr>
<tr>
<td>connecti</td>
<td>0.096325</td>
<td>181</td>
</tr>
</tbody>
</table>
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
size: 163,200 (99.9% of memory free)

<table>
<thead>
<tr>
<th>variable name</th>
<th>type</th>
<th>format</th>
<th>label</th>
</tr>
</thead>
<tbody>
<tr>
<td>weight</td>
<td>double</td>
<td>%10.0g</td>
<td>Case weights</td>
</tr>
<tr>
<td>regstate</td>
<td>byte</td>
<td>%8.0g</td>
<td>REGSTATE</td>
</tr>
<tr>
<td>race</td>
<td>byte</td>
<td>%8.0g</td>
<td>RACE</td>
</tr>
<tr>
<td>q4</td>
<td>byte</td>
<td>%8.0g</td>
<td>Q4</td>
</tr>
<tr>
<td>q10</td>
<td>byte</td>
<td>%8.0g</td>
<td>Q10</td>
</tr>
</tbody>
</table>

Sorted by:

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

.tab q4, m

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

.tab q4, m nol

<table>
<thead>
<tr>
<th>Mode of voting</th>
<th>Freq.</th>
<th>Percent</th>
<th>Cum.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5,979</td>
<td>58.62</td>
<td>58.62</td>
</tr>
<tr>
<td>2</td>
<td>1,654</td>
<td>16.22</td>
<td>74.83</td>
</tr>
<tr>
<td>3</td>
<td>1,770</td>
<td>17.35</td>
<td>92.19</td>
</tr>
<tr>
<td>4</td>
<td>27</td>
<td>0.26</td>
<td>92.45</td>
</tr>
<tr>
<td>.</td>
<td>770</td>
<td>7.55</td>
<td>100.00</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>10,200</td>
<td>100.00</td>
<td></td>
</tr>
</tbody>
</table>

.tab q10, m

<table>
<thead>
<tr>
<th>Line length</th>
<th>Freq.</th>
<th>Percent</th>
<th>Cum.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not at all</td>
<td>3,037</td>
<td>29.77</td>
<td>29.77</td>
</tr>
<tr>
<td>Less than 10 minutes</td>
<td>2,394</td>
<td>23.47</td>
<td>53.25</td>
</tr>
<tr>
<td>10-30 minutes</td>
<td>1,375</td>
<td>13.48</td>
<td>66.73</td>
</tr>
<tr>
<td>31 minutes ñ 1 hour</td>
<td>593</td>
<td>5.81</td>
<td>72.54</td>
</tr>
<tr>
<td>More than 1 hour</td>
<td>211</td>
<td>2.07</td>
<td>74.61</td>
</tr>
<tr>
<td>I don't know</td>
<td>24</td>
<td>0.24</td>
<td>74.84</td>
</tr>
<tr>
<td></td>
<td>2,566</td>
<td>25.16</td>
<td>100.00</td>
</tr>
</tbody>
</table>
. tab q10, m nol

<table>
<thead>
<tr>
<th>Line length</th>
<th>Freq.</th>
<th>Percent</th>
<th>Cum.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3,037</td>
<td>29.77</td>
<td>29.77</td>
</tr>
<tr>
<td>2</td>
<td>2,394</td>
<td>23.47</td>
<td>53.25</td>
</tr>
<tr>
<td>3</td>
<td>1,375</td>
<td>13.48</td>
<td>66.73</td>
</tr>
<tr>
<td>4</td>
<td>593</td>
<td>5.81</td>
<td>72.54</td>
</tr>
<tr>
<td>5</td>
<td>211</td>
<td>2.07</td>
<td>74.61</td>
</tr>
<tr>
<td>6</td>
<td>24</td>
<td>0.24</td>
<td>74.84</td>
</tr>
<tr>
<td></td>
<td>2,566</td>
<td>25.16</td>
<td>100.00</td>
</tr>
</tbody>
</table>

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)
. list
+--------------------------------------------+
<table>
<thead>
<tr>
<th>regstate   th-20121    n1   th-20122    n2</th>
</tr>
</thead>
</table>
2. | Alaska    0     136    .1432126    29 |
3. | Arizona   .0881236   61           0    6 |
4. | Arkansas  .0971476    87   .1297159    83 |
5. | Califor n .0213493    94          0    1 |
6. | Colorado  .0436208    37          0    30 |
7. | Connecticut .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 |
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. | Massach u .0571125   178          0    4 |
23. | Michigan  .2343819   127    .5325436   4 |
24. | Minnesot  .0398434   169          0    3 |
25. | Mississippi .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 |
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. | Tennessee .170697    70    .1314724  101 |
44. | Texas    .0625682    54    .1217225  120 |
<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>45.</td>
<td>Utah</td>
<td>.0820119</td>
<td>91</td>
<td>.0392503</td>
</tr>
<tr>
<td>46.</td>
<td>Vermont</td>
<td>.00388</td>
<td>126</td>
<td>0</td>
</tr>
<tr>
<td>47.</td>
<td>Virginia</td>
<td>.3114677</td>
<td>149</td>
<td>.0625045</td>
</tr>
<tr>
<td>48.</td>
<td>Washington</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>49.</td>
<td>West Virginia</td>
<td>.1088866</td>
<td>113</td>
<td>.0157209</td>
</tr>
<tr>
<td>50.</td>
<td>Wisconsin</td>
<td>.0522848</td>
<td>123</td>
<td>.0111778</td>
</tr>
<tr>
<td>51.</td>
<td>Wyoming</td>
<td>.0187056</td>
<td>149</td>
<td>0</td>
</tr>
</tbody>
</table>

```
.log close
```