

Unit 5 – STATA for Data Description

Homework

SOLUTIONS

LOG – Code & Output

Note – In the solutions that follow, in some instances I have provided more than is actually needed. Thought you might enjoy ...

```
. *
. **** Before you begin: Download from the course website descriptive\_gss.dta
```

```
. **** 1a.) Review choices of options for command tabstat. Note – Below is a screen capture.
. help tabstat
```

statname	Definition
<u>mean</u>	mean
<u>count</u>	count of nonmissing observations
<u>n</u>	same as count
<u>sum</u>	sum
<u>max</u>	maximum
<u>min</u>	minimum
<u>range</u>	range = max - min
<u>sd</u>	standard deviation
<u>variance</u>	variance
<u>cv</u>	coefficient of variation (sd/mean)
<u>semean</u>	standard error of mean (sd/sqrt(n))
<u>skewness</u>	skewness
<u>kurtosis</u>	kurtosis
<u>p1</u>	1st percentile
<u>p5</u>	5th percentile
<u>p10</u>	10th percentile
<u>p25</u>	25th percentile
<u>median</u>	median (same as p50)
<u>p50</u>	50th percentile (same as median)
<u>p75</u>	75th percentile
<u>p90</u>	90th percentile
<u>p95</u>	95th percentile
<u>p99</u>	99th percentile
<u>iqr</u>	interquartile range = p75 - p25
<u>q</u>	equivalent to specifying p25 p50 p75

```
. **** 1b.) Command SUMMARIZE for continuous variables: detailed descriptives
. summarize hrs1, detail
```

```
. summarize hrs1, detail

              number of hours worked last week
-----
Percentiles      Smallest
1%                6          1
5%               16          2
10%              21          2      Obs          1729
25%              36          2      Sum of Wgt.    1729

50%              40
75%              50          89      Mean          41.77675
90%              60          89      Std. Dev.     14.62304
95%              68          89      Variance      213.8332
99%              88          89      Skewness      .2834814
                          Kurtosis      4.310339
```

```
. ***** 1c.) Command TABSTAT for continuous variables: detailed descriptives
. tabstat hrs1 age wwzhr, stat(n mean sd semean min q max skewness kurtosis) columns(variable)
```

```
. tabstat hrs1 age wwzhr, stat(n mean sd semean min q max skewness kurtosis) columns(variable)

+-----+-----+-----+
| stats | hrs1 | age | wwzhr |
+-----+-----+-----+
| N      | 1729 | 2751 | 1574 |
| mean   | 41.77675 | 46.28281 | 5.907878 |
| sd     | 14.62304 | 17.37049 | 8.866734 |
| se(mean) | .3516739 | .331182 | .2234917 |
| min    | 1     | 18   | 0     |
| p25    | 36    | 32   | 1     |
| p50    | 40    | 44   | 3     |
| p75    | 50    | 58   | 7     |
| max    | 89    | 89   | 112   |
| skewness | .2834814 | .4475504 | 3.997908 |
| kurtosis | 4.310339 | 2.304249 | 30.39248 |
+-----+-----+-----+
```

```
. **** 1e.) Command TABSTAT with option BY(sortvariable) for continuous variables
. sort polviews
. tabstat hrs1, by(polviews) stat(n mean sd semean min q max) columns(statistics) format(%8.2f)
```

```
. sort polviews
. tabstat hrs1, by(polviews) stat(n mean sd semean min q max) columns(statistics) format(%8.2f)
```

Summary for variables: hrs1
by categories of: polviews (think of self as liberal or conservative)

polviews	N	mean	sd	se(mean)	min	p25	p50	p75	max
extremely libera	27.00	37.89	16.83	3.24	5.00	20.00	40.00	50.00	72.00
liberal	93.00	43.84	16.46	1.71	12.00	35.00	40.00	55.00	89.00
slightly liberal	110.00	42.84	11.88	1.13	5.00	40.00	40.00	50.00	70.00
moderate	336.00	42.15	13.46	0.73	2.00	39.50	40.00	48.00	89.00
slightly conserva	129.00	42.36	16.23	1.43	9.00	36.00	40.00	50.00	89.00
conservative	125.00	43.90	15.49	1.39	2.00	40.00	40.00	50.00	89.00
extrmly conserva	25.00	37.00	10.26	2.05	10.00	35.00	40.00	42.00	52.00
Total	845.00	42.43	14.45	0.50	2.00	37.00	40.00	50.00	89.00

```
. *
. **** 2a.) Command NUMLABEL, ADD to display numerical codes together with labels
. ***** Command TAB1 for discrete variables: frequencies, relative frequencies, etc
. numlabel, add
. tab1 marital sex satjob7, sort missing
```

```
. numlabel, add
. tab1 marital sex satjob7, sort missing
```

-> tabulation of marital

marital status	Freq.	Percent	Cum.
1. married	1,269	45.90	45.90
5. never married	708	25.61	71.50
3. divorced	445	16.09	87.59
2. widowed	247	8.93	96.53
4. separated	96	3.47	100.00
Total	2,765	100.00	

-> tabulation of sex

respondents sex	Freq.	Percent	Cum.
2. female	1,537	55.59	55.59
1. male	1,228	44.41	100.00
Total	2,765	100.00	

```
-> tabulation of satjob7
```

job satisfaction in general	Freq.	Percent	Cum.
1. very satisfied	1,945	70.34	70.34
2. very satisfied	289	10.45	80.80
3. fairly satisfied	264	9.55	90.34
4. neither satisfied nor dissatisfied	127	4.59	94.94
5. fairly dissatisfied	53	1.92	96.85
6. very dissatisfied	47	1.70	98.55
7. completely dissatisfied	29	1.05	99.60
	11	0.40	100.00
Total	2,765	100.00	

```
. **** 2b.) Two way cross-tab with row percentages: sex x polviews
```

```
. tab2 sex polviews, row
```

```
. tab2 sex polviews, row
```

```
-> tabulation of sex by polviews
```

Key
frequency
row percentage

respondent s sex	think of self as liberal or conservative							Total
	1. extrem	2. libera	3. slight	4. modera	5. slghtl	6. conser	7. extrml	
1. male	24	71	71	244	100	107	12	629
	3.82	11.29	11.29	38.79	15.90	17.01	1.91	100.00
2. female	23	72	88	278	109	103	29	702
	3.28	10.26	12.54	39.60	15.53	14.67	4.13	100.00
Total	47	143	159	522	209	210	41	1,331
	3.53	10.74	11.95	39.22	15.70	15.78	3.08	100.00

```
. tab2 polviews sex, row
```

```
-> tabulation of polviews by sex
```

Key			
frequency			
row percentage			
think of self as liberal or conservative	respondents sex		Total
	1. male	2. female	
1. extremely liberal	24 51.06	23 48.94	47 100.00
2. liberal	71 49.65	72 50.35	143 100.00
3. slightly liberal	71 44.65	88 55.35	159 100.00
4. moderate	244 46.74	278 53.26	522 100.00
5. slghtly conservati	100 47.85	109 52.15	209 100.00
6. conservative	107 50.95	103 49.05	210 100.00
7. extrmly conservati	12 29.27	29 70.73	41 100.00
Total	629 47.26	702 52.74	1,331 100.00