



User: Part1log

Largest Std. Dev. 973.9411
 name: <unnamed>
 log: F:\ShaziLapTop\Desktop\STATA 16\STATA 16 MP Full Portable Version\Stata16\part1.log
 log type: text
 opened on: 1 Jan 2000, 00:00:28

- 1 . *Importing .csv file
- 2 . import delimited "C:\Users\HP\Desktop\coding_challenge\Part 1\New Variables.csv"
(7 vars, 949 obs)
- 3 . describe

Contains data

obs: 949
 vars: 7

| variable name | storage type | display format | value label | variable label |
|-----------------|--------------|----------------|-------------|----------------|
| uniqueid | int | %8.0g | | |
| surveytime2 | float | %9.0g | | |
| thanavisitrea~2 | byte | %8.0g | | |
| exit | byte | %8.0g | | |
| base_age | int | %8.0g | | |
| base_income | int | %8.0g | | |
| base_educ | str27 | %27s | | |

Sorted by:

Note: Dataset has changed since last saved.

- 4 . rename base_educ educ
- 5 . merge 1:1 uniqueid using "C:\Users\HP\Desktop\coding_challenge\Part 1\Main Dataset.dta", force
(note: variable educ was byte in the using data, but will be str27 now)
(note: variable uniqueid was int, now float to accommodate using data's values)

| Result | # of obs. |
|-------------|-----------------|
| not matched | 60 |
| from master | 11 (_merge==1) |
| from using | 49 (_merge==2) |
| matched | 938 (_merge==3) |

- 6 . *Merge the new observations
- 7 . append using "C:\Users\HP\Desktop\coding_challenge\Part 1\New Observations.dta", force
(note: variable educ was byte in the using data, but will be str27 now)
(label treatment already defined)
(label gender already defined)
(label occup already defined)
(label castecode already defined)
(label standard already defined)
(label yesno already defined)
(label yes1no0 already defined)

8 . *New onservations are added and the two datasets are merged.

9 . *Average and median values of time spent surveying, considering only completed

10 .

11 . surveys.

surveys. is not a valid command name

r(199);

12 . sum surveytime if survey_complete == 1

| Variable | Obs | Mean | Std. Dev. | Min | Max |
|------------|-----|----------|-----------|------|-------|
| surveytime | 889 | 7206.845 | 973.9411 | 3757 | 10208 |

13 .

14 . sum surveytime if survey_complete == 1, d

| Length of Survey (in seconds) | | | | |
|-------------------------------|------|----------|-------------|----------|
| Percentiles | | Smallest | | |
| 1% | 4779 | 3757 | | |
| 5% | 5688 | 4337 | | |
| 10% | 6044 | 4433 | Obs | 889 |
| 25% | 6511 | 4517 | Sum of Wgt. | 889 |
| 50% | 7171 | | Mean | 7206.845 |
| | | Largest | Std. Dev. | 973.9411 |
| 75% | 7864 | 10042 | | |
| 90% | 8472 | 10078 | Variance | 948561.3 |
| 95% | 8799 | 10181 | Skewness | .0146197 |
| 99% | 9375 | 10208 | Kurtosis | 3.097474 |

15 . *Mean seconds is 7206.8 and Median is 7171 seconds

16 . *Average length of time spent surveying for this surveyor

17 . tab surveyor

| Surveyor | Freq. | Percent | Cum. |
|----------|-------|---------|--------|
| Anna | 125 | 12.64 | 12.64 |
| Benjamin | 141 | 14.26 | 26.90 |
| Caroline | 63 | 6.37 | 33.27 |
| David | 157 | 15.87 | 49.14 |
| Grace | 55 | 5.56 | 54.70 |
| Jane | 16 | 1.62 | 56.32 |
| John | 68 | 6.88 | 63.20 |
| Joseph | 68 | 6.88 | 70.07 |
| Mary | 153 | 15.47 | 85.54 |
| Peter | 44 | 4.45 | 89.99 |
| Sam | 99 | 10.01 | 100.00 |
| Total | 989 | 100.00 | |

```
18 . egen avgagemf=mean( surveytime ) if survey_complete == 1, by(surveyor)
    (111 missing values generated)
```

```
19 .
```

```
20 . tab avgagemf
```

| avgagemf | Freq. | Percent | Cum. |
|----------|-------|---------|--------|
| 6959.786 | 14 | 1.57 | 1.57 |
| 7076.86 | 57 | 6.41 | 7.99 |
| 7138.453 | 64 | 7.20 | 15.19 |
| 7155.367 | 90 | 10.12 | 25.31 |
| 7165.683 | 142 | 15.97 | 41.28 |
| 7194.862 | 58 | 6.52 | 47.81 |
| 7204.742 | 120 | 13.50 | 61.30 |
| 7237.973 | 37 | 4.16 | 65.47 |
| 7241.688 | 125 | 14.06 | 79.53 |
| 7292.577 | 137 | 15.41 | 94.94 |
| 7416.133 | 45 | 5.06 | 100.00 |
| Total | 889 | 100.00 | |

```
21 . *There are no signifincant observable differences between the surveyors.
```

```
22 . *Checking the uniqueness of our ID variables (hhid)
```

```
23 . describe hhid
```

| variable name | storage type | display format | value label | variable label |
|---------------|-----------------|-------------------|----------------|----------------|
| hhid | str7 | %9s | | Household ID |

```
24 . duplicates report hhid
```

Duplicates in terms of hhid

| copies | observations | surplus |
|--------|--------------|---------|
| 1 | 983 | 0 |
| 2 | 6 | 3 |
| 11 | 11 | 10 |

```
25 . duplicates list hhid
```

Duplicates in terms of hhid

| group: | obs: | hhid |
|--------|------|------|
| 1 | 150 | |
| 1 | 209 | |
| 1 | 609 | |
| 1 | 684 | |
| 1 | 729 | |
| 1 | 778 | |
| 1 | 791 | |
| 1 | 798 | |
| 1 | 806 | |
| 1 | 895 | |

| | | |
|-------|-----|---------|
| ----- | | |
| 1 | 902 | |
| 2 | 564 | 1802011 |
| 2 | 565 | 1802011 |
| 3 | 605 | 1807077 |
| 3 | 606 | 1807077 |
| ----- | | |
| 4 | 643 | 1813023 |
| 4 | 644 | 1813023 |
| ----- | | |

26 . *Removing duplicate values

27 . duplicates drop hhid, force

Duplicates in terms of hhid

(13 observations deleted)

28 . duplicates report hhid

Duplicates in terms of hhid

| | | |
|--------|--------------|---------|
| ----- | | |
| copies | observations | surplus |
| ----- | | |
| 1 | 987 | 0 |
| ----- | | |

29 . *Create histograms and overlay them with k-density plots

30 . histogram income, frequency
(bin=29, start=0, width=208.65517)

31 . graph save "Graph" "C:\Users\HP\Desktop\coding_challenge\Part 1\Hist1.gph"
(file C:\Users\HP\Desktop\coding_challenge\Part 1\Hist1.gph saved)
twoway (kdensity income) (histogram income), name(k1,replace)

32 . graph save "k1" "C:\Users\HP\Desktop\coding_challenge\Part 1\Kdensity1.gph"
(file C:\Users\HP\Desktop\coding_challenge\Part 1\Kdensity1.gph saved)
histogram surveytime, frequency
(bin=29, start=3757, width=222.44828)

33 . graph save "Graph" "C:\Users\HP\Desktop\coding_challenge\Part 1\Hist2.gph"
(file C:\Users\HP\Desktop\coding_challenge\Part 1\Hist2.gph saved)
twoway (kdensity surveytime) (histogram surveytime), name(k2,replace)

34 . graph save "k2" "C:\Users\HP\Desktop\coding_challenge\Part 1\Kdensity2.gph"
(file C:\Users\HP\Desktop\coding_challenge\Part 1\Kdensity2.gph saved)
*Removing personally identifiable information (PII)

```
35 . replace surveyor = "1" if surveyor == "Anna"
    (125 real changes made)
```

```
36 . tab surveyor
```

| Surveyor | Freq. | Percent | Cum. |
|----------|-------|---------|--------|
| 1 | 125 | 12.68 | 12.68 |
| Benjamin | 141 | 14.30 | 26.98 |
| Caroline | 63 | 6.39 | 33.37 |
| David | 157 | 15.92 | 49.29 |
| Grace | 55 | 5.58 | 54.87 |
| Jane | 16 | 1.62 | 56.49 |
| John | 68 | 6.90 | 63.39 |
| Joseph | 68 | 6.90 | 70.28 |
| Mary | 152 | 15.42 | 85.70 |
| Peter | 44 | 4.46 | 90.16 |
| Sam | 97 | 9.84 | 100.00 |
| Total | 986 | 100.00 | |

```
37 . replace surveyor = "2" if surveyor == "Benjamin"
    (141 real changes made)
```

```
38 . replace surveyor = "3" if surveyor == "Caroline"
    (63 real changes made)
```

```
39 . replace surveyor = "4" if surveyor == "David"
    (157 real changes made)
```

```
40 . replace surveyor = "5" if surveyor == "Grace"
    (55 real changes made)
```

```
41 . replace surveyor = "6" if surveyor == "Jane"
    (16 real changes made)
```

```
42 . replace surveyor = "7" if surveyor == "Joseph"
    too few quotes
    r(132);
```

```
43 . replace surveyor = "7" if surveyor == "Joseph"
    (68 real changes made)
```

```
44 . replace surveyor = "8" if surveyor == "Mary"
    (152 real changes made)
```

```
45 . replace surveyor = "9" if surveyor == "Peter"
    (44 real changes made)
```

```
46 . replace surveyor = "10" if surveyor == "John"
    (68 real changes made)
```

47 . replace surveyor = "11" if surveyor == "Sam"
(97 real changes made)

48 . tab surveyor

| Surveyor | Freq. | Percent | Cum. |
|----------|-------|---------|--------|
| 1 | 125 | 12.68 | 12.68 |
| 10 | 68 | 6.90 | 19.57 |
| 11 | 97 | 9.84 | 29.41 |
| 2 | 141 | 14.30 | 43.71 |
| 3 | 63 | 6.39 | 50.10 |
| 4 | 157 | 15.92 | 66.02 |
| 5 | 55 | 5.58 | 71.60 |
| 6 | 16 | 1.62 | 73.23 |
| 7 | 68 | 6.90 | 80.12 |
| 8 | 152 | 15.42 | 95.54 |
| 9 | 44 | 4.46 | 100.00 |
| Total | 986 | 100.00 | |

49 . *recode missing values.

50 . mvdecode burglaryyn vandalismyn trespassingyn, mv(-999)
burglaryyn: 4 missing values generated
vandalismyn: 5 missing values generated

51 . mvdecode burglaryyn vandalismyn trespassingyn, mv(-997)
burglaryyn: 7 missing values generated
trespassin~n: 8 missing values generated

52 . log close
name: <unnamed>
log: F:\ShaziLapTop\Desktop\STATA 16\STATA 16 MP Full Portable Version\Stata16\part1.log
log type: text
closed on: 1 Jan 2000, 01:06:28