

```
1
2 // Stata Workshop 1
3 // Department of Social Sciences, School of Arts and Social Sciences, The Open University of Hong
  Kong
4 // October, 2020
5
6 // Source of Data
7 // psychology.dta is used here. It is originally a SPSS data file (psychology.sav), which was
  downloaded from
8 // https://edge.sagepub.com/field5e/student-resources/datasets
9 // (website of Discovering Statistics Using IBM SPSS Statistics [by Andy Field]; access date: Oct
  14, 2020).
10
11 // Part 1 Descriptive Statistics
12
13 // 1.1
14 codebook group // check the coding of a categorical variable
15 tab1 group // show frequency distribution
16 tab1 group, nolabel // show the original code, not the label
17
18 // 1.2
19 codebook stats // "codebook" can also be applied to a continuous variable
20 sum stats // show descriptive statistics
21 h sum // seek help to learn more about a command
22 sum stats, detail // show more descriptive statistics
23 histogram stats // graphically illustrate the distribution
24
25 // 1.3
26 sort group // need this code before showing descriptive statistics by categories
27 by group: sum stats // show descriptive statistics by categories
28
29 sum stats if group==1 // show descriptive statistics of a certain category; need "==" for "if"
30
31 // Part 2 Data Management
32
33 // 2.1
34 label variable group "Group" // change variable label
35
36 // 2.2
37 rename group year // change variable name
38
39 label variable year "University Year"
40
41 // 2.3
42 codebook year
43
44 label define yearlabel 0 "year 1" 1 "year 2" 2 "year 3" // change labels of categories step 1:
  come up with a labeling scheme
45 label values year yearlabel // step 2: attach the labeling scheme
46
47 codebook year
48
49 // 2.4
50 gen freshman=. // recode: come up with a new categorical variable
51 replace freshman =1 if year==0
52 replace freshman =0 if year==1
53 replace freshman =0 if year==2
54
55 recode year (0=1) (1/2=0), gen (freshman2) // an alternative way
56
57 codebook year
58
59 recode year (0=1) (1=2) (2=3) // recode: same variable
60
61 codebook year
62
63 label define yearlabel2 1 "year 1" 2 "year 2" 3 "year 3"
64 label values year yearlabel2
```

```
65
66 codebook year
67
68 // 2.5
69 replace stats=stats/10 // do calculation on a variable: replace
70
71 replace stats=stats*10
72
73 gen newstats=stats/10 // do calculation on a variable: without replace
74
75 // 2.6
76 gen total = exper+stats+social+develop+person // total: listwise deletion
77
78 gen average = (exper+stats+social+develop+person)/5 // average: listwise deletion
79
80 egen total2 = rowtotal (exper stats social develop person) // total: ignore missing (make sense?)
81
82 egen average2 = rowmean (exper stats social develop person) // average: ignore missing (make good
sense)
```