

```
-----
name: <unnamed>
log: D:\Jason\workshop\Key Stata commands\key stata command2.log
log type: text
opened on: 11 Jul 2022, 11:17:29
```

```
.
. *****
*****
* Example 1.
* Research question: how race and experience of violence predictd the number of depressive symptoms.
* Data: public data of Add Health at Wave III
* The number of respondents in the data: 6,504
* The numbers of respondents with valid values of racial backgrounds, experience of violence,
* or depressive symptoms change with the way these variables were constructed.
* The numbers of respondents in the analytic sample symptom also vary, depending on how variables were constructed
*****
*****
*Read in the data and check the number of respondents in the data and the format of data
*****
.
. use "D:\Jason\workshop\Key Stata commands\Add Health\depression.dta", clear
( )
.
. count
6,504
. duplicates report aid

Duplicates in terms of aid

-----
copies | observations      surplus
-----+-----
      1 |          6,504          0
-----
.
. *****
. * Examine the variables of racial backgrounds
. * Consider how the information of these variables should be aggregated
. *****
.
. des h3od2 h3od4a h3od4b h3od4c h3od4d h3od6

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

```

-----
h3od2          byte      %17.0g   h3od2          S1Q2 ARE YOU OF HISPANIC ORIGIN-W3
h3od4a         byte      %17.0g   race           S1Q4A RACE-WHITE-W3
h3od4b         byte      %17.0g   race           S1Q4B RACE-BLACK/AFRICAN AM-W3
h3od4c         byte      %17.0g   race           S1Q4C RACE-AMER INDIAN/NATIVE AM-W3
h3od4d         byte      %17.0g   race           S1Q4D RACE-ASIAN-W3
h3od6          byte      %37.0g   h3od6         S1Q6 RACE-SINGLE CATEGORY-W3

```

```
. sum h3od2 h3od4a h3od4b h3od4c h3od4d h3od6
```

Variable	Obs	Mean	Std. Dev.	Min	Max
h3od2	4,882	.1183941	.4297787	0	8
h3od4a	4,882	.7552233	.7993149	0	9
h3od4b	4,882	.2570668	.5049197	0	9
h3od4c	4,882	.056739	.3635974	0	9
h3od4d	4,882	.0553052	.3456249	0	9
h3od6	4,882	6.809709	1.004825	1	9

```
. tab1 h3od2 h3od4a h3od4b h3od4c h3od4d h3od6, mis
```

-> tabulation of h3od2

S1Q2 ARE YOU OF HISPANIC ORIGIN-W3	Freq.	Percent	Cum.
0 no	4,353	66.93	66.93
1 yes	522	8.03	74.95
8 don't know	7	0.11	75.06
.	1,622	24.94	100.00
Total	6,504	100.00	

-> tabulation of h3od4a

S1Q4A RACE-WHITE-W3	Freq.	Percent	Cum.
0 not marked	1,467	22.56	22.56
1 marked	3,376	51.91	74.46
6 refused	8	0.12	74.58
8 don't know	16	0.25	74.83
9 not applicable	15	0.23	75.06
.	1,622	24.94	100.00
Total	6,504	100.00	

-> tabulation of h3od4b

S1Q4B RACE-BLACK/AFRICA N AM-W3	Freq.	Percent	Cum.
0 not marked	3,664	56.33	56.33
1 marked	1,213	18.65	74.98
8 don't know	3	0.05	75.03
9 not applicable	2	0.03	75.06
.	1,622	24.94	100.00
Total	6,504	100.00	

-> tabulation of h3od4c

S1Q4C RACE-AMER INDIAN/NATIVE AM-W3	Freq.	Percent	Cum.
0 not marked	4,650	71.49	71.49
1 marked	226	3.47	74.97
8 don't know	3	0.05	75.02
9 not applicable	3	0.05	75.06
.	1,622	24.94	100.00
Total	6,504	100.00	

-> tabulation of h3od4d

S1Q4D RACE-ASIAN-W3	Freq.	Percent	Cum.
0 not marked	4,650	71.49	71.49
1 marked	227	3.49	74.98
8 don't know	2	0.03	75.02
9 not applicable	3	0.05	75.06
.	1,622	24.94	100.00
Total	6,504	100.00	

-> tabulation of h3od6

S1Q6 RACE-SINGLE CATEGORY-W3	Freq.	Percent	Cum.
1 white	93	1.43	1.43
2 black or African American	57	0.88	2.31
3 American Indian or N ative American	12	0.18	2.49
4 Asian or Pacific Islander	14	0.22	2.71
6 refused	3	0.05	2.75

7 legitimate skip	4,698	72.23	74.98
8 don't know	3	0.05	75.03
9 not applicable	2	0.03	75.06
.	1,622	24.94	100.00

Total	6,504	100.00	

```
. codebook h3od2 h3od4a h3od4b h3od4c h3od4d h3od6
```

```
-----
h3od2
S1Q2 ARE YOU OF HISPANIC ORIGIN-W3
-----
```

```

      type: numeric (byte)
      label: h3od2

      range: [0,8]                units: 1
unique values: 3                  missing .: 1,622/6,504
```

```

tabulation: Freq.  Numeric  Label
            4,353      0      0 no
            522       1      1 yes
            7         8      8 don't know
            1,622      .

```

```
-----
h3od4a
S1Q4A RACE-WHITE-W3
-----
```

```

      type: numeric (byte)
      label: race

      range: [0,9]                units: 1
unique values: 5                  missing .: 1,622/6,504
```

```

tabulation: Freq.  Numeric  Label
            1,467      0      0 not marked
            3,376      1      1 marked
            8         6      6 refused
            16        8      8 don't know
            15        9      9 not applicable
            1,622      .

```

```
-----
-----
h3od4b
S1Q4B RACE-BLACK/AFRICAN AM-W3
-----
-----
```

```

      type: numeric (byte)
      label: race

      range: [0,9]          units: 1
unique values: 4          missing .: 1,622/6,504

      tabulation: Freq.   Numeric   Label
                  3,664     0     0 not marked
                  1,213     1     1 marked
                   3         8     8 don't know
                   2         9     9 not applicable
                  1,622     .

```

```
-----
-----
h3od4c
S1Q4C RACE-AMER INDIAN/NATIVE AM-W3
-----
-----
```

```

      type: numeric (byte)
      label: race

      range: [0,9]          units: 1
unique values: 4          missing .: 1,622/6,504

      tabulation: Freq.   Numeric   Label
                  4,650     0     0 not marked
                   226     1     1 marked
                   3         8     8 don't know
                   3         9     9 not applicable
                  1,622     .

```

```
-----
-----
h3od4d
S1Q4D RACE-ASIAN-W3
-----
-----
```

```

      type: numeric (byte)
      label: race

```

```

range: [0,9] units: 1
unique values: 4 missing .: 1,622/6,504

```

tabulation:	Freq.	Numeric	Label
	4,650	0	0 not marked
	227	1	1 marked
	2	8	8 don't know
	3	9	9 not applicable
	1,622	.	.

```

-----
h3od6
S1Q6 RACE-SINGLE CATEGORY-W3
-----

```

```

type: numeric (byte)
label: h3od6

```

```

range: [1,9] units: 1
unique values: 8 missing .: 1,622/6,504

```

tabulation:	Freq.	Numeric	Label
	93	1	1 white
	57	2	2 black or African American
	12	3	3 American Indian or N ative American
	14	4	4 Asian or Pacific Islander
	3	6	6 refused
	4,698	7	7 legitim ate s kip
	3	8	8 don't know
	2	9	9 not applicable
	1,622	.	.

```

.
. *****
. * generate new race variables and code the missing values of these variables
. *****
.
. clonevar hispanic = h3od2
(1,622 missing values generated)
.
. clonevar white = h3od4a
(1,622 missing values generated)
.
. clonevar black = h3od4b
(1,622 missing values generated)
.
. clonevar native = h3od4c

```

(1,622 missing values generated)

```
. clonevar asian = h3od4d
(1,622 missing values generated)
```

```
. clonevar additional = h3od6
(1,622 missing values generated)
```

```
.
.
. mvdecode hispanic white black native asian additional, mv(6=.\7=.\8=.\9=.)
hispanic: 7 missing values generated
white: 39 missing values generated
black: 5 missing values generated
native: 6 missing values generated
asian: 5 missing values generated
additional: 4706 missing values generated
```

```
. tab1 hispanic white black native asian additional, mis
```

-> tabulation of hispanic

S1Q2 ARE YOU OF HISPANIC ORIGIN-W3	Freq.	Percent	Cum.
0 no	4,353	66.93	66.93
1 yes	522	8.03	74.95
.	1,629	25.05	100.00
Total	6,504	100.00	

-> tabulation of white

S1Q4A RACE-WHITE-W3	Freq.	Percent	Cum.
0 not marked	1,467	22.56	22.56
1 marked	3,376	51.91	74.46
.	1,661	25.54	100.00
Total	6,504	100.00	

-> tabulation of black

S1Q4B RACE-BLACK/AFRICA N AM-W3	Freq.	Percent	Cum.
0 not marked	3,664	56.33	56.33

1 marked		1,213	18.65	74.98
.		1,627	25.02	100.00

Total		6,504	100.00	

-> tabulation of native

S1Q4C RACE-AMER		Freq.	Percent	Cum.
INDIAN/NATIVE				
AM-W3				

0 not marked		4,650	71.49	71.49
1 marked		226	3.47	74.97
.		1,628	25.03	100.00

Total		6,504	100.00	

-> tabulation of asian

S1Q4D		Freq.	Percent	Cum.
RACE-ASIAN-W3				

0 not marked		4,650	71.49	71.49
1 marked		227	3.49	74.98
.		1,627	25.02	100.00

Total		6,504	100.00	

-> tabulation of additional

S1Q6 RACE-SINGLE CATEGORY-W3		Freq.	Percent	Cum.

1 white		93	1.43	1.43
2 black or African American		57	0.88	2.31
3 American Indian or Native American		12	0.18	2.49
4 Asian or Pacific Islander		14	0.22	2.71
.		6,328	97.29	100.00

Total		6,504	100.00	

```
.
.
. *****
. * count the number of respondents with valid information on their racial backgrounds
. *****
.
. egen valid_race = rownonmiss(hispanic white black native asian additional)
.
. tab1 valid_race, mis
```

-> tabulation of valid_race

valid_race	Freq.	Percent	Cum.
0	1,622	24.94	24.94
1	4	0.06	25.00
4	46	0.71	25.71
5	4,656	71.59	97.29
6	176	2.71	100.00
Total	6,504	100.00	

. list aid hispanic white black native asian additional valid_race in 1/20, nol

	aid	hispanic	white	black	native	asian	additi~l	valid_~e
1.	57100270	0	0	1	0	0	.	5
2.	57101310	0	0	1	0	0	.	5
3.	57103171	0
4.	57103869	0	0	1	0	0	.	5
5.	57104553	0
6.	57104649	0
7.	57104676	0	0	1	0	0	.	5
8.	57109625	0	1	0	0	0	.	5
9.	57110897	0
10.	57111071	0	1	0	1	0	1	6
11.	57111786	0	0	1	0	0	.	5
12.	57113943	0	0	1	0	0	.	5
13.	57116359	0
14.	57117542	0	1	0	0	0	.	5
15.	57117997	0	0	1	0	0	.	5
16.	57118381	0	1	0	0	0	.	5
17.	57118943	0	1	0	0	0	.	5
18.	57120005	0	1	0	0	0	.	5
19.	57120046	1	1	0	0	0	.	5
20.	57120371	0	1	0	0	0	.	5

.
 . *****
 . * check the combination of four racial backgrounds, including being white, black, native, and asian
 . *****
 . egen race_temp1 = concat(white black native asian), punct(_)

```
. label variable race_temp1 "the combination of racial backgrounds"
```

```
. tab1 race_temp1, mis
```

```
-> tabulation of race_temp1
```

the combination of racial backgrounds	Freq.	Percent	Cum.
._._._.	1,626	25.00	25.00
.0_0_0_0	35	0.54	25.54
0_.0_0_0	1	0.02	25.55
0_0_.0_0	2	0.03	25.58
0_0_0_.	1	0.02	25.60
0_0_0_1	185	2.84	28.44
0_0_1_0	114	1.75	30.20
0_1_0_0	1,138	17.50	47.69
0_1_0_1	5	0.08	47.77
0_1_1_0	21	0.32	48.09
1_0_0_0	3,218	49.48	97.57
1_0_0_1	30	0.46	98.03
1_0_1_0	76	1.17	99.20
1_0_1_1	3	0.05	99.25
1_1_0_0	35	0.54	99.78
1_1_0_1	2	0.03	99.82
1_1_1_0	10	0.15	99.97
1_1_1_1	2	0.03	100.00
Total	6,504	100.00	

```
.
.
. gen race_temp2 =.
(6,504 missing values generated)
```

```
. label variable race_temp2 "the racial backgrounds, using four racial backgropund variables"
```

```
. label define race_temp2 1 "1 white" ///
> 2 "2 black or African American" ///
> 3 "3 American Indian or Native American" ///
> 4 "4 Asian or Pacific Islander" ///
> 5 "5 hispanic" ///
> 6 "6 mixed racial backgrounds"
```

```
. label value race_temp2 race_temp2
```

```
. replace race_temp2 = 1 if race_temp1 == "1_0_0_0"
(3,218 real changes made)

. replace race_temp2 = 2 if race_temp1 == "0_1_0_0"
(1,138 real changes made)

. replace race_temp2 = 3 if race_temp1 == "0_0_1_0"
(114 real changes made)

. replace race_temp2 = 4 if race_temp1 == "0_0_0_1"
(185 real changes made)

. replace race_temp2 = 6 if inlist(race_temp1, "0_1_1_0",
"1_0_0_1","1_0_1_0","1_0_1_1" ,"1_0_1_0","1_1_0_0","1_1_0_1","1_1_1_0","1_1_1_1")
(179 real changes made)
```

```
. tab1 race_temp2, mis
```

```
-> tabulation of race_temp2
```

```
the racial backgrounds, using four |
racial backgropund variables |      Freq.      Percent      Cum.
-----+-----+-----+-----+-----+
                1 white |      3,218      49.48      49.48
                2 black or African American |      1,138      17.50      66.97
3 American Indian or Native American |           114      1.75      68.73
                4 Asian or Pacific Islander |           185      2.84      71.57
                6 mixed racial backgrounds |           179      2.75      74.32
                . |           1,670      25.68      100.00
-----+-----+-----+-----+
                        Total |           6,504      100.00
```

```
.
.
. *****
. * Incorporating the information of being hispanic
. *****
```

```
. tab2 race_temp2 hispanic, mis
```

```
-> tabulation of race_temp2 by hispanic
```

```
the racial |
backgrounds, using |      S1Q2 ARE YOU OF HISPANIC
four racial |      ORIGIN=W3
backgropund variables |      0 no      1 yes      . |      Total
-----+-----+-----+-----+-----+
                1 white |      2,870      344      4 |      3,218
```

2 black or African Am	1,104	33	1	1,138
3 American Indian or	40	74	0	114
4 Asian or Pacific Is	167	16	2	185
6 mixed racial backgo	161	18	0	179
.	11	37	1,622	1,670

Total	4,353	522	1,629	6,504

```
. clonevar race_temp3 = race_temp2
```

```
(1,670 missing values generated)
```

```
. replace race_temp3 = 5 if hispanic ==1
```

```
(522 real changes made)
```

```
. tab2 race_temp2 race_temp3 if hispanic ==1, mis
```

```
-> tabulation of race_temp2 by race_temp3 if hispanic ==1
```

	the racial backgrounds, using four racial background variables	5 hispani	Total
1 white	344		344
2 black or African Am	33		33
3 American Indian or	74		74
4 Asian or Pacific Is	16		16
6 mixed racial backgo	18		18
.	37		37

Total	522		522

```
. tab2 race_temp2 race_temp3 if hispanic ~=1, mis
```

```
-> tabulation of race_temp2 by race_temp3 if hispanic ~=1
```

	the racial backgrounds, using four racial background variables	1 white	2 black o	3 America	4 Asian o	6 mixed r	.	Total
1 white	2,874	0	0	0	0	0	0	2,874
2 black or African Am	0	1,105	0	0	0	0	0	1,105
3 American Indian or	0	0	40	0	0	0	0	40
4 Asian or Pacific Is	0	0	0	169	0	0	0	169
6 mixed racial backgo	0	0	0	0	161	0	0	161

```

-----
. |          0          0          0          0          0          1,633 | 1,633
Total | 2,874    1,105        40        169        161    1,633 | 5,982
-----

```

```

.
. *****
. * Incorporating the information of additional racial information
. *****
.
. tab2 race_temp3 additional, mis

```

-> tabulation of race_temp3 by additional

```

the racial |
backgrounds, using |
four racial |
backgropund variables | S1Q6 RACE-SINGLE CATEGORY-W3
-----
1 white | 0 | 0 | 0 | 0 | 2,874 | 2,874
2 black or African Am | 0 | 0 | 0 | 0 | 1,105 | 1,105
3 American Indian or | 0 | 0 | 0 | 0 | 40 | 40
4 Asian or Pacific Is | 0 | 0 | 0 | 0 | 169 | 169
5 hispanic | 8 | 3 | 3 | 2 | 506 | 522
6 mixed racial backgo | 85 | 49 | 9 | 12 | 6 | 161
. | 0 | 5 | 0 | 0 | 1,628 | 1,633
-----
Total | 93 | 57 | 12 | 14 | 6,328 | 6,504
-----

```

```

.
. clonevar race_final = race_temp3
(1,633 missing values generated)

. replace race_final = 1 if race_temp3 ==6 & additional ==1
(85 real changes made)

. replace race_final = 2 if race_temp3 ==6 & additional ==2
(49 real changes made)

. replace race_final = 2 if race_temp3 ==. & additional ==2
(5 real changes made)

. replace race_final = 3 if race_temp3 ==6 & additional ==3
(9 real changes made)

. replace race_final = 4 if race_temp3 ==6 & additional ==4
(12 real changes made)

.
. tab1 race_final, mis

-> tabulation of race_final

```

```

the racial backgrounds, using four |
racial backgropund variables |
-----+-----+-----+-----

```

	Freq.	Percent	Cum.
1 white	2,959	45.50	45.50
2 black or African American	1,159	17.82	63.31
3 American Indian or Native American	49	0.75	64.07
4 Asian or Pacific Islander	181	2.78	66.85
5 hispanic	522	8.03	74.88
6 mixed racial backgrounds	6	0.09	74.97
.	1,628	25.03	100.00
Total	6,504	100.00	

```

.
.
. *****
. * Violence variables
. *****
.
.
. des h3ds18b h3ds18c h3ds18d h3ds18e h3ds18f h3ds18g

```

variable name	storage type	display format	value label	variable label
h3ds18b	byte	%17.0g	violence	S26Q18B 12 MO,ONE PULLED GUN ON YOU-W3
h3ds18c	byte	%17.0g	violence	S26Q18C 12 MO,ONE PULLED KNIFE ON YOU-W3
h3ds18d	byte	%17.0g	violence	S26Q18D 12 MO,SOMEONE SHOT YOU-W3
h3ds18e	byte	%17.0g	violence	S26Q18E 12 MO,SOMEONE STABBED YOU-W3
h3ds18f	byte	%17.0g	violence	S26Q18F 12 MO,YOU WERE BEATEN UP-W3
h3ds18g	byte	%17.0g	violence	S26Q18G 12 MO,WERE BEATEN UP/THG STOL-W3

```

.
. tab1 h3ds18b h3ds18c h3ds18d h3ds18e h3ds18f h3ds18g, mis

```

```

-> tabulation of h3ds18b

```

	Freq.	Percent	Cum.
S26Q18B 12 MO,ONE PULLED GUN ON YOU-W3			
0 not marked	4,633	71.23	71.23
1 marked	204	3.14	74.37
6 refused	16	0.25	74.62
8 don't know	9	0.14	74.75
9 not applicable	19	0.29	75.05
.	1,623	24.95	100.00

Total | 6,504 100.00

-> tabulation of h3ds18c

S26Q18C 12 MO,ONE PULLED KNIFE ON YOU-W3	Freq.	Percent	Cum.
0 not marked	4,668	71.77	71.77
1 marked	177	2.72	74.49
6 refused	14	0.22	74.71
8 don't know	8	0.12	74.83
9 not applicable	14	0.22	75.05
.	1,623	24.95	100.00
Total	6,504	100.00	

-> tabulation of h3ds18d

S26Q18D 12 MO,SOMEONE SHOT YOU-W3	Freq.	Percent	Cum.
0 not marked	4,819	74.09	74.09
1 marked	25	0.38	74.48
6 refused	14	0.22	74.69
8 don't know	7	0.11	74.80
9 not applicable	16	0.25	75.05
.	1,623	24.95	100.00
Total	6,504	100.00	

-> tabulation of h3ds18e

S26Q18E 12 MO,SOMEONE STABBED YOU-W3	Freq.	Percent	Cum.
0 not marked	4,804	73.86	73.86
1 marked	40	0.62	74.48
6 refused	14	0.22	74.69
8 don't know	7	0.11	74.80
9 not applicable	16	0.25	75.05
.	1,623	24.95	100.00
Total	6,504	100.00	

-> tabulation of h3ds18f

S26Q18F 12 MO,YOU |

WERE BEATEN UP-W3	Freq.	Percent	Cum.
0 not marked	4,730	72.72	72.72
1 marked	114	1.75	74.48
6 refused	14	0.22	74.69
8 don't know	7	0.11	74.80
9 not applicable	16	0.25	75.05
.	1,623	24.95	100.00
Total	6,504	100.00	

-> tabulation of h3ds18g

S26Q18G 12 MO, WERE BEATEN UP/THG STOL-W3	Freq.	Percent	Cum.
0 not marked	4,812	73.99	73.99
1 marked	33	0.51	74.49
6 refused	14	0.22	74.71
8 don't know	7	0.11	74.82
9 not applicable	15	0.23	75.05
.	1,623	24.95	100.00
Total	6,504	100.00	

```
. mvdecode h3ds18b h3ds18c h3ds18d h3ds18e h3ds18f h3ds18g, mv(6=. \8=. \9=.)
h3ds18b: 44 missing values generated
h3ds18c: 36 missing values generated
h3ds18d: 37 missing values generated
h3ds18e: 37 missing values generated
h3ds18f: 37 missing values generated
h3ds18g: 36 missing values generated
```

```
.
.
. *****
. * two ways of aggregating the items
. *****
.
. gen sum_violence = h3ds18b + h3ds18c + h3ds18d + h3ds18e + h3ds18f + h3ds18g
(1,669 missing values generated)
. egen rowtotal_violence = rowtotal (h3ds18b h3ds18c h3ds18d h3ds18e h3ds18f h3ds18g)
. egen miss_violence = rowmiss (h3ds18b h3ds18c h3ds18d h3ds18e h3ds18f h3ds18g)
.
.
```



```

.
. sort miss_violence
. list aid h3ds18b h3ds18c h3ds18d h3ds18e h3ds18f h3ds18g sum_violence rowtotal_violence miss_violence if inlist(miss_violence,
1,3,5), nol sepby(miss_violence)

```

	aid	h3ds18b	h3ds18c	h3ds18d	h3ds18e	h3ds18f	h3ds18g	sum_violence	rowtotal_violence	miss_violence
4836.	95719977	.	0	0	0	0	0	.	0	1
4837.	57217613	.	0	0	0	0	0	.	0	1
4838.	90576336	0	0	0	0	.	0	.	0	1
4839.	90719939	.	0	0	0	0	0	.	0	1
4840.	93715937	.	0	0	0	0	0	.	0	1
4841.	90570992	.	0	0	0	0	0	.	0	1
4842.	94575026	.	0	0	0	0	0	.	0	1
4843.	98572373	.	.	.	0	0	0	.	0	3
4844.	93508623	0	0	0	0	3
4845.	92572924	.	.	.	0	0	0	.	0	3
4846.	97574031	.	0	.	.	0	0	.	0	3
4847.	92573816	.	0	0	5
4848.	93718100	.	.	0	0	5

```

.
.
. *****
. * Depression variables
. *****
.
. des h3sp5 h3sp6 h3sp7 h3sp8 h3sp9 h3sp10 h3sp11 h3sp12

```

variable name	storage type	display format	value label	variable label
h3sp5	byte	%34.0g	depressed	S12Q5 PAST 7 DAYS BOTHERED BY THINGS-W3
h3sp6	byte	%34.0g	depressed	S12Q1 PAST 7 DAYS SHAKE OFF BLUES-W3
h3sp7	byte	%34.0g	depressed	S12Q7 PAST 7 DAYS FELT AS GOOD AS OTH-W3
h3sp8	byte	%34.0g	depressed	S12Q8 PAST 7 DAYS TROUBLE CONCENTRTNG-W3
h3sp9	byte	%34.0g	depressed	S12Q9 PAST 7 DAYS WERE DEPRESSED-W3
h3sp10	byte	%34.0g	depressed	S12Q10 PAST 7 DAYS TOO TIRED DO THNGS-W3
h3sp11	byte	%34.0g	depressed	S12Q11 PAST 7 DAYS ENJOYED LIFE-W3
h3sp12	byte	%34.0g	depressed	S12Q12 PAST 7 DAYS WERE SAD-W3

. tab1 h3sp5 h3sp6 h3sp7 h3sp8 h3sp9 h3sp10 h3sp11 h3sp12, mis

-> tabulation of h3sp5

S12Q5 PAST 7 DAYS BOTHERED BY THINGS-W3	Freq.	Percent	Cum.
0 Never/rarely	2,760	42.44	42.44
1 Sometimes	1,726	26.54	68.97
2 A lot of the time	315	4.84	73.82
3 Most of the time/all of the time	77	1.18	75.00
6 Refused	1	0.02	75.02
8 Don't know	2	0.03	75.05
9 Not applicable	1	0.02	75.06
.	1,622	24.94	100.00
Total	6,504	100.00	

-> tabulation of h3sp6

S12Q1 PAST 7 DAYS SHAKE OFF BLUES-W3	Freq.	Percent	Cum.
0 Never/rarely	3,695	56.81	56.81
1 Sometimes	894	13.75	70.56
2 A lot of the time	185	2.84	73.40
3 Most of the time/all of the time	99	1.52	74.92
6 Refused	1	0.02	74.94
8 Don't know	7	0.11	75.05
9 Not applicable	1	0.02	75.06
.	1,622	24.94	100.00
Total	6,504	100.00	

-> tabulation of h3sp7

S12Q7 PAST 7 DAYS FELT AS GOOD AS OTH-W3	Freq.	Percent	Cum.
0 Never/rarely	285	4.38	4.38
1 Sometimes	682	10.49	14.87
2 A lot of the time	1,066	16.39	31.26
3 Most of the time/all of the time	2,832	43.54	74.80
6 Refused	2	0.03	74.83
8 Don't know	12	0.18	75.02
9 Not applicable	3	0.05	75.06
.	1,622	24.94	100.00
Total	6,504	100.00	

-> tabulation of h3sp8

S12Q8 PAST 7 DAYS TROUBLE CONCENTRTNG-W3	Freq.	Percent	Cum.
0 Never/rarely	2,498	38.41	38.41
1 Sometimes	1,872	28.78	67.19
2 A lot of the time	364	5.60	72.79
3 Most of the time/all of the time	142	2.18	74.97
6 Refused	2	0.03	75.00
8 Don't know	4	0.06	75.06
.	1,622	24.94	100.00
Total	6,504	100.00	

-> tabulation of h3sp9

S12Q9 PAST 7 DAYS WERE DEPRESSED-W3	Freq.	Percent	Cum.
0 Never/rarely	3,623	55.70	55.70
1 Sometimes	960	14.76	70.46
2 A lot of the time	202	3.11	73.57
3 Most of the time/all of the time	89	1.37	74.94
6 Refused	1	0.02	74.95
8 Don't know	6	0.09	75.05
9 Not applicable	1	0.02	75.06
.	1,622	24.94	100.00
Total	6,504	100.00	

-> tabulation of h3sp10

S12Q10 PAST 7 DAYS TOO TIRED DO THNGS-W3	Freq.	Percent	Cum.
0 Never/rarely	2,402	36.93	36.93
1 Sometimes	1,983	30.49	67.42
2 A lot of the time	369	5.67	73.09
3 Most of the time/all of the time	122	1.88	74.97
8 Don't know	5	0.08	75.05
9 Not applicable	1	0.02	75.06
.	1,622	24.94	100.00
Total	6,504	100.00	

-> tabulation of h3sp11

S12Q11 PAST 7 DAYS ENJOYED LIFE-W3	Freq.	Percent	Cum.
------------------------------------	-------	---------	------

	0 Never/rarely	106	1.63	1.63
	1 Sometimes	706	10.85	12.48
	2 A lot of the time	1,267	19.48	31.96
3	Most of the time/all of the time	2,800	43.05	75.02
	6 Refused	1	0.02	75.03
	8 Don't know	2	0.03	75.06
	.	1,622	24.94	100.00
Total		6,504	100.00	

-> tabulation of h3sp12

	S12Q12 PAST 7 DAYS WERE SAD-W3	Freq.	Percent	Cum.
	0 Never/rarely	2,885	44.36	44.36
	1 Sometimes	1,671	25.69	70.05
	2 A lot of the time	243	3.74	73.79
3	Most of the time/all of the time	77	1.18	74.97
	6 Refused	1	0.02	74.98
	8 Don't know	4	0.06	75.05
	9 Not applicable	1	0.02	75.06
	.	1,622	24.94	100.00
Total		6,504	100.00	

```
. mvdecode h3sp5 h3sp6 h3sp7 h3sp8 h3sp9 h3sp10 h3sp11 h3sp12, mv(6=.\8=.\9=.)
  h3sp5: 4 missing values generated
  h3sp6: 9 missing values generated
  h3sp7: 17 missing values generated
  h3sp8: 6 missing values generated
  h3sp9: 8 missing values generated
  h3sp10: 6 missing values generated
  h3sp11: 3 missing values generated
  h3sp12: 6 missing values generated

. gen sum_depression = h3sp5 + h3sp6 + h3sp7 + h3sp8 + h3sp9 + h3sp10 + h3sp11 + h3sp12
(1,653 missing values generated)

. egen rowtotal_depression = rowtotal (h3sp5 h3sp6 h3sp7 h3sp8 h3sp9 h3sp10 h3sp11 h3sp12)

. egen miss_depression = rowmiss (h3sp5 h3sp6 h3sp7 h3sp8 h3sp9 h3sp10 h3sp11 h3sp12)

. *****
. * check the data
. *****
. sum sum_depression rowtotal_depression
```

Variable	Obs	Mean	Std. Dev.	Min	Max
sum_depres~n	4,851	7.636776	2.597382	0	24
rowtotal_d~n	6,504	5.72417	3.999446	0	24

```
. tab2 sum_depression rowtotal_depression, mis
```

-> tabulation of sum_depression by rowtotal_depression

sum_depres sion	rowtotal_depression												Total	
	0	1	2	3	4	5	6	7	8	9	10	11	12	
0	10	0	0	0	0	0	0	0	0	0	0	0	0	10
1	0	8	0	0	0	0	0	0	0	0	0	0	0	8
2	0	0	21	0	0	0	0	0	0	0	0	0	0	21
3	0	0	0	81	0	0	0	0	0	0	0	0	0	81
4	0	0	0	0	184	0	0	0	0	0	0	0	0	184
5	0	0	0	0	0	381	0	0	0	0	0	0	0	381
6	0	0	0	0	0	0	1,084	0	0	0	0	0	0	1,084
7	0	0	0	0	0	0	0	945	0	0	0	0	0	945
8	0	0	0	0	0	0	0	0	734	0	0	0	0	734
9	0	0	0	0	0	0	0	0	0	525	0	0	0	525
10	0	0	0	0	0	0	0	0	0	0	311	0	0	311
11	0	0	0	0	0	0	0	0	0	0	0	208	0	208
12	0	0	0	0	0	0	0	0	0	0	0	0	118	118
13	0	0	0	0	0	0	0	0	0	0	0	0	0	75
14	0	0	0	0	0	0	0	0	0	0	0	0	0	63
15	0	0	0	0	0	0	0	0	0	0	0	0	0	31
16	0	0	0	0	0	0	0	0	0	0	0	0	0	27
17	0	0	0	0	0	0	0	0	0	0	0	0	0	15
18	0	0	0	0	0	0	0	0	0	0	0	0	0	14
19	0	0	0	0	0	0	0	0	0	0	0	0	0	7
20	0	0	0	0	0	0	0	0	0	0	0	0	0	6
22	0	0	0	0	0	0	0	0	0	0	0	0	0	2
24	0	0	0	0	0	0	0	0	0	0	0	0	0	1
.	1,625	0	2	4	3	3	1	3	6	3	1	0	0	1,653
Total	1,635	8	23	85	187	384	1,085	948	740	528	312	208	118	6,504

sum_depres sion	rowtotal_depression											Total
	13	14	15	16	17	18	19	20	22	24		
0	0	0	0	0	0	0	0	0	0	0	0	10
1	0	0	0	0	0	0	0	0	0	0	0	8
2	0	0	0	0	0	0	0	0	0	0	0	21
3	0	0	0	0	0	0	0	0	0	0	0	81
4	0	0	0	0	0	0	0	0	0	0	0	184
5	0	0	0	0	0	0	0	0	0	0	0	381
6	0	0	0	0	0	0	0	0	0	0	0	1,084
7	0	0	0	0	0	0	0	0	0	0	0	945
8	0	0	0	0	0	0	0	0	0	0	0	734
9	0	0	0	0	0	0	0	0	0	0	0	525
10	0	0	0	0	0	0	0	0	0	0	0	311
11	0	0	0	0	0	0	0	0	0	0	0	208
12	0	0	0	0	0	0	0	0	0	0	0	118
13	75	0	0	0	0	0	0	0	0	0	0	75
14	0	63	0	0	0	0	0	0	0	0	0	63
15	0	0	31	0	0	0	0	0	0	0	0	31
16	0	0	0	27	0	0	0	0	0	0	0	27
17	0	0	0	0	15	0	0	0	0	0	0	15
18	0	0	0	0	0	14	0	0	0	0	0	14
19	0	0	0	0	0	0	7	0	0	0	0	7
20	0	0	0	0	0	0	0	6	0	0	0	6

22	0	0	0	0	0	0	0	0	0	2	0	2
24	0	0	0	0	0	0	0	0	0	0	1	1
.	0	1	1	0	0	0	0	0	0	0	0	1,653

Total	75	64	32	27	15	14	7	6	2	1	1	6,504

. sort miss_depression

. list aid h3sp5 h3sp6 h3sp7 h3sp8 h3sp9 h3sp10 h3sp11 h3sp12 sum_depression rowtot_depression miss_depression if inlist(miss_depression, 1,3,5), nol sepby(miss_depression)

	aid	h3sp5	h3sp6	h3sp7	h3sp8	h3sp9	h3sp10	h3sp11	h3sp12	sum_depression	rowtot_depression	miss_depression
4852.	95719936	1	0	.	1	0	0	3	0	.	5	1
4853.	57138056	0	0	3	0	0	.	3	1	.	7	1
4854.	97572102	1	0	.	1	0	2	3	1	.	8	1
4855.	90579959	0	0	.	0	0	0	3	0	.	3	1
4856.	93572118	0	0	.	0	0	0	2	0	.	2	1
4857.	92716333	0	3	2	0	0	1	3	.	.	9	1
4858.	95716862	0	0	.	2	0	2	2	1	.	7	1
4859.	99713501	1	.	1	0	0	1	1	0	.	4	1
4860.	91579964	1	.	3	1	0	1	2	0	.	8	1
4861.	97716162	2	0	1	.	2	2	1	0	.	8	1
4862.	97577053	0	0	3	1	0	1	1	.	.	6	1
4863.	90716290	0	0	.	0	0	1	2	0	.	3	1
4864.	96575249	0	1	.	2	2	1	1	1	.	8	1
4865.	90716390	.	1	3	0	2	3	3	2	.	14	1
4866.	91716096	0	.	3	0	0	2	3	0	.	8	1
4867.	91573320	0	0	.	0	0	1	2	0	.	3	1
4868.	57249576	3	1	3	3	1	.	1	3	.	15	1
4869.	99571273	1	.	2	2	0	1	3	0	.	9	1
4870.	99500276	0	0	3	1	0	.	3	0	.	7	1
4871.	99579956	1	0	.	3	0	3	3	0	.	10	1
4872.	96574438	1	0	.	0	0	0	3	0	.	4	1
4873.	98606929	0	.	2	0	0	1	1	0	.	4	1
4877.	96676757	1	1	.	.	.	1	1	1	.	5	3

. *****
 . * Analysis
 . *****

. reg sum_depression i.race_final sum_violence

Source	SS	df	MS	Number of obs	=	4,800
Model	136.912769	6	22.8187949	F(6, 4794)	=	3.43
Residual	31932.112	4,794	6.6608494	Prob > F	=	0.0022
				R-squared	=	0.0043
				Adj R-squared	=	0.0030
Total	32069.0248	4,800	6.68104683	Root MSE	=	2.5809

	sum_depression	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]

race_final						
2	black or African American	.0278164	.0904767	0.31	0.759	-.1495594 .2051923
3	American Indian or N ative American	.0281369	.375876	0.07	0.940	-.7087525 .7650263
4	Asian or Pacific Islander	-.2573486	.1987463	-1.29	0.195	-.6469826 .1322854
5	hispanic	-.0177193	.1237177	-0.14	0.886	-.2602628 .2248243
6	mixed racial backgrounds	-.3742357	1.054849	-0.35	0.723	-2.442224 1.693753

```

sum_violence | .3138815 .0738075 4.25 0.000 .1691849 .4585782
_cons | 7.602942 .0483163 157.36 0.000 7.50822 7.697664

```

```
. sum sum_depression race_final sum_violence
```

```

Variable | Obs Mean Std. Dev. Min Max
-----|-----
sum_depres~n | 4,851 7.636776 2.597382 0 24
race_final | 4,876 1.803527 1.30779 1 6
sum_violence | 4,835 .1226474 .5048933 0 6

```

```
. count
6,504
```

```
. count if sum_depression ~=.
4,851
```

```
. count if sum_depression ~=. & race_final ~=.
4,845
```

```
. count if sum_depression ~=. & race_final ~=. & sum_violence ~=.
4,801
```

```
. reg rowtotal_depression i.race_final rowtotal_violence
```

```

Source | SS df MS Number of obs = 4,876
-----|-----
Model | 144.878505 6 24.1464176 F(6, 4869) = 3.56
Residual | 33019.8089 4,869 6.78164078 Prob > F = 0.0016
Total | 33164.6874 4,875 6.80301281 R-squared = 0.0044
Adj R-squared = 0.0031
Root MSE = 2.6042

```

```

rowtotal_depression | Coef. Std. Err. t P>|t| [95% Conf. Interval]
-----|-----
race_final |
2 black or African American | .011609 .090317 0.13 0.898 -1.1654531 .1886711
3 American Indian or N ative American | .1037915 .3753963 0.28 0.782 -.6321546 .8397376
4 Asian or Pacific Islander | -.2973038 .1994287 -1.49 0.136 -.688274 .0936665
5 hispanic | -.0041158 .1237031 -0.03 0.973 -.2466297 .2383982
6 mixed racial backgrounds | -.3674605 1.064359 -0.35 0.730 -2.454085 1.719164
rowtotal_violence | .3185107 .0743859 4.28 0.000 .1726808 .4643406
_cons | 7.594624 .0484757 156.67 0.000 7.499589 7.689658

```

```
. sum rowtotal_depression race_final rowtotal_violence
```

```

Variable | Obs Mean Std. Dev. Min Max
-----|-----
rowtotal_d~n | 6,504 5.72417 3.999446 0 24
race_final | 4,876 1.803527 1.30779 1 6
rowtotal_v~e | 6,504 .0911747 .4385915 0 6

```

```
. count
6,504
```

```

. count if rowtotal_depression ~=.
6,504

. count if rowtotal_depression ~=. & race_final ~=.
4,876

. count if rowtotal_depression ~=. & race_final ~=. & rowtotal_violence ~=.
4,876

.
.
.
.
. *****
. * Example 2: NSFG
. * Research question: how age and smoking is related to the probability of live birth.
. * Data: public data of NSFG 2017-2019
. * The number of respondents in the data: 3,709
. * The number of pregnancy records: 10, 215.
. * 3,662 respondents reported a total of 10,007 pregnancy with iive births
. * 3,709 respondents have valid information on age, which have impact on 20,215 pregnancy records.
. * 1,719 respondents have valid information on soming during pregnancy, which have impact on 2,293 pregnancy records
. * The final analytic sample has 2,572 pregnancy records from 1,719 respondents.
. *****
.
.
.
. use "D:\jason\workshop\Key Stata commands\nsfg\data\preg\example2.dta", clear
(National Survey of Family Growth (NSFG), 2017-2019)

.
. count
10,215

. duplicates report caseid

Duplicates in terms of caseid

-----
copies | observations      surplus
-----+-----
      1 |           903         0
      2 |          2060        1030
      3 |          2472        1648
      4 |          1884        1413
      5 |          1270        1016
      6 |           672         560
      7 |           329         282
      8 |           264         231
      9 |           117         104
     10 |           120         108

```

```

11 |          33          30
12 |          24          22
13 |          39          36
14 |          28          26

```

```
. duplicates report caseid pregodr
```

Duplicates in terms of caseid pregodr

```

-----
copies | observations      surplus
-----+-----
1 |             10215          0
-----

```

```
. list caseid pregodr in 1/50, sepby(caseid)
```

```

+-----+
| caseid      pregodr |
+-----+
1. | 88819  1ST PREGNANCY |
2. | 88819  2ND PREGNANCY |
+-----+
3. | 83055  1ST PREGNANCY |
4. | 83055  2ND PREGNANCY |
+-----+
5. | 92062  1ST PREGNANCY |
6. | 92062  2ND PREGNANCY |
7. | 92062  3RD PREGNANCY |
+-----+
8. | 84933  1ST PREGNANCY |
9. | 84933  2ND PREGNANCY |
+-----+
10. | 84368  1ST PREGNANCY |
11. | 84368  2ND PREGNANCY |
+-----+
12. | 88197  1ST PREGNANCY |
13. | 88197  2ND PREGNANCY |
14. | 88197  3RD PREGNANCY |
+-----+
15. | 81914  1ST PREGNANCY |
16. | 81914  2ND PREGNANCY |
17. | 81914  3RD PREGNANCY |
+-----+
18. | 81537  1ST PREGNANCY |
19. | 81537  2ND PREGNANCY |
+-----+
20. | 85655  1ST PREGNANCY |
21. | 85655  2ND PREGNANCY |

```

```

-----+-----
22. | 81415 1ST PREGNANCY |
23. | 81415 2ND PREGNANCY |
24. | 81415 3RD PREGNANCY |
25. | 81415 4TH PREGNANCY |
26. | 81415 5TH PREGNANCY |
27. | 81415 6TH PREGNANCY |
-----+-----
28. | 88505 1ST PREGNANCY |
-----+-----
29. | 88058 1ST PREGNANCY |
30. | 88058 2ND PREGNANCY |
31. | 88058 3RD PREGNANCY |
-----+-----
32. | 92032 1ST PREGNANCY |
33. | 92032 2ND PREGNANCY |
34. | 92032 3RD PREGNANCY |
-----+-----
35. | 90634 1ST PREGNANCY |
-----+-----
36. | 88067 1ST PREGNANCY |
-----+-----
37. | 82737 1ST PREGNANCY |
38. | 82737 2ND PREGNANCY |
39. | 82737 3RD PREGNANCY |
40. | 82737 4TH PREGNANCY |
41. | 82737 5TH PREGNANCY |
42. | 82737 6TH PREGNANCY |
-----+-----
43. | 91477 1ST PREGNANCY |
44. | 91477 2ND PREGNANCY |
-----+-----
45. | 82432 1ST PREGNANCY |
46. | 82432 2ND PREGNANCY |
-----+-----
47. | 89468 1ST PREGNANCY |
48. | 89468 2ND PREGNANCY |
49. | 89468 3RD PREGNANCY |
50. | 89468 4TH PREGNANCY |
-----+-----

```

```

.
.
. *****
. * Determine the number of respondents in the data
. * This file has 3,709 female respondents and 10,215 pregnancy records
. * On average, respondents have 2.75 pregnancies
. *****
.
. sort caseid pregordr

```

```

. by caseid: gen preg_n = _n
. by caseid: gen preg_N = _N

.
. label variable preg_n "indicator of pregnancy"
. label variable preg_N "total number of pregnancy of a respondent"

.
. list caseid pregordr preg_n preg_N in 1/50, sepby(caseid) nol

```

```

+-----+
| caseid  pregordr  preg_n  preg_N |
+-----+
1. | 80719          1         1         4 |
2. | 80719          2         2         4 |
3. | 80719          3         3         4 |
4. | 80719          4         4         4 |
+-----+
5. | 80720          1         1         4 |
6. | 80720          2         2         4 |
7. | 80720          3         3         4 |
8. | 80720          4         4         4 |
+-----+
9. | 80723          1         1         3 |
10. | 80723          2         2         3 |
11. | 80723          3         3         3 |
+-----+
12. | 80727          1         1         2 |
13. | 80727          2         2         2 |
+-----+
14. | 80728          1         1         3 |
15. | 80728          2         2         3 |
16. | 80728          3         3         3 |
+-----+
17. | 80729          1         1         1 |
+-----+
18. | 80731          1         1         3 |
19. | 80731          2         2         3 |
20. | 80731          3         3         3 |
+-----+
21. | 80733          1         1         1 |
+-----+
22. | 80742          1         1         3 |
23. | 80742          2         2         3 |
24. | 80742          3         3         3 |
+-----+
25. | 80743          1         1         2 |

```

26.	80743	2	2	2

27.	80747	1	1	3
28.	80747	2	2	3
29.	80747	3	3	3

30.	80756	1	1	4
31.	80756	2	2	4
32.	80756	3	3	4
33.	80756	4	4	4

34.	80760	1	1	2
35.	80760	2	2	2

36.	80764	1	1	3
37.	80764	2	2	3
38.	80764	3	3	3

39.	80766	1	1	1

40.	80768	1	1	1

41.	80772	1	1	2
42.	80772	2	2	2

43.	80773	1	1	2
44.	80773	2	2	2

45.	80776	1	1	3
46.	80776	2	2	3
47.	80776	3	3	3

48.	80785	1	1	2
49.	80785	2	2	2

50.	80787	1	1	2

```
. tab1 preg_n, mis
```

```
-> tabulation of preg_n
```

indicator of pregnancy	Freq.	Percent	Cum.
1	3,709	36.31	36.31
2	2,806	27.47	63.78
3	1,776	17.39	81.16

4	952	9.32	90.48
5	481	4.71	95.19
6	227	2.22	97.42
7	115	1.13	98.54
8	68	0.67	99.21
9	35	0.34	99.55
10	22	0.22	99.77
11	10	0.10	99.86
12	7	0.07	99.93
13	5	0.05	99.98
14	2	0.02	100.00

-----+-----
 Total | 10,215 100.00

. sum preg_N if preg_n ==1

Variable	Obs	Mean	Std. Dev.	Min	Max
preg_N	3,709	2.754112	1.678096	1	14

.
 .
 . *****
 . * The pregnancy outcomes
 . *****
 .
 .
 . tab1 pregend1 pregend2, mis

-> tabulation of pregend1

BC-1 How Pregnancy Ended - 1st	mention	Freq.	Percent	Cum.
Miscarriage		1,652	16.17	16.17
Stillbirth		107	1.05	17.22
Abortion		931	9.11	26.33
Ectopic or tubal pregnancy		124	1.21	27.55
Live birth by Cesarean section		2,098	20.54	48.09
Live birth by vaginal delivery		5,095	49.88	97.96
Refused		7	0.07	98.03
Don't know		7	0.07	98.10
.		194	1.90	100.00
-----+-----				
Total		10,215	100.00	

-> tabulation of pregend2

BC-1 How Pregnancy Ended - 2nd	mention	Freq.	Percent	Cum.
--------------------------------	---------	-------	---------	------

Miscarriage	1	0.01	0.01
Stillbirth	1	0.01	0.02
Abortion	2	0.02	0.04
Ectopic or tubal pregnancy	1	0.01	0.05
Live birth by Cesarean section	5	0.05	0.10
Live birth by vaginal delivery	3	0.03	0.13
.	10,202	99.87	100.00
Total	10,215	100.00	

```

.
. gen livebirth1 = .
(10,215 missing values generated)

. replace livebirth1 = 1 if inlist(pregend1,5,6)
(7,193 real changes made)

. replace livebirth1 = 0 if inlist(pregend1,1,2,3,4)
(2,814 real changes made)

. replace livebirth1 = . if inlist(pregend1,7,8,0)
(0 real changes made)

. label variable livebirth1 "outcome of a pregnancy"

. label define livebirth 0 "not a livebirth" 1 "livebirth"

. label value livebirth1 livebirth

.
.
. gen livebirth2 = .
(10,215 missing values generated)

. replace livebirth2 = 1 if inlist(pregend2,5,6)
(8 real changes made)

. replace livebirth2 = 0 if inlist(pregend2,1,2,3,4)
(5 real changes made)

. replace livebirth2 = . if inlist(pregend2,7,8,0)
(0 real changes made)

.
. label variable livebirth2 "additional outcome of a pregnancy"

. label value livebirth2 livebirth

.

```

```
. tab2 livebirth1 livebirth2, mis
-> tabulation of livebirth1 by livebirth2
```

outcome of a pregnancy	additional outcome of a pregnancy		.	Total
	not a liv	livebirth		
not a livebirth	3	4	2,807	2,814
livebirth	2	4	7,187	7,193
.	0	0	208	208
Total	5	8	10,202	10,215

```
.
. *****
. * Code the end outcome of pregnancy
. *****
.
. *****
. * Sum of the items
. *****
. gen sum_livebirth = livebirth1 + livebirth2
(10,202 missing values generated)
. recode sum_livebirth (2=1)
(sum_livebirth: 4 changes made)
.
. *****
. * the rowtotal function
. *****
. egen rowtotal_livebirth = rowtotal(livebirth1 livebirth2)
. recode rowtotal_livebirth (2=1)
(rowtotal_livebirth: 4 changes made)
.
. *****
. * Use the replace commands
. *****
. gen code_livebirth = livebirth1
(208 missing values generated)
```

```
. replace code_livebirth = 1 if livebirth1 ==0 & livebirth2 ==1
(4 real changes made)
```

```
. label value sum_livebirth livebirth
```

```
. label value rowtotal_livebirth livebirth
```

```
. label value code_livebirth livebirth
```

```
.
.
.
.
. *****
. * Compare differnt ways of constructin the LIVEBIRTH variable
. *****
```

```
. count
10,215
```

```
. sum sum_livebirth rowtotal_livebirth code_livebirth
```

Variable	Obs	Mean	Std. Dev.	Min	Max
sum_livebirth	13	.7692308	.438529	0	1
rowtotal_livebirth	10,215	.7045521	.4562662	0	1
code_livebirth	10,007	.7191966	.4494141	0	1

```
. tab2 livebirth1 livebirth2, mis
```

```
-> tabulation of livebirth1 by livebirth2
```

outcome of a pregnancy	additional outcome of a pregnancy		.	Total
	not a liv livebirth	livebirth		
not a livebirth	3	4	2,807	2,814
livebirth	2	4	7,187	7,193
.	0	0	208	208
Total	5	8	10,202	10,215

```
.
.
.
. *****
. * check the age variable
```

```
. *****
```

```
. tabl ager, mis
```

```
-> tabulation of ager
```

Age at interview	Freq.	Percent	Cum.
15 YEARS	1	0.01	0.01
16 YEARS	3	0.03	0.04
17 YEARS	9	0.09	0.13
18 YEARS	23	0.23	0.35
19 YEARS	36	0.35	0.70
20 YEARS	34	0.33	1.04
21 YEARS	67	0.66	1.69
22 YEARS	109	1.07	2.76
23 YEARS	119	1.16	3.93
24 YEARS	190	1.86	5.79
25 YEARS	220	2.15	7.94
26 YEARS	251	2.46	10.40
27 YEARS	366	3.58	13.98
28 YEARS	313	3.06	17.04
29 YEARS	364	3.56	20.61
30 YEARS	385	3.77	24.38
31 YEARS	388	3.80	28.17
32 YEARS	488	4.78	32.95
33 YEARS	540	5.29	38.24
34 YEARS	463	4.53	42.77
35 YEARS	486	4.76	47.53
36 YEARS	458	4.48	52.01
37 YEARS	399	3.91	55.92
38 YEARS	447	4.38	60.29
39 YEARS	405	3.96	64.26
40 YEARS	370	3.62	67.88
41 YEARS	356	3.49	71.37
42 YEARS	322	3.15	74.52
43 YEARS	355	3.48	77.99
44 YEARS	365	3.57	81.57
45 YEARS	337	3.30	84.87
46 YEARS	413	4.04	88.91
47 YEARS	407	3.98	92.89
48 YEARS	356	3.49	96.38
49	363	3.55	99.93
50	7	0.07	100.00
Total	10,215	100.00	

```
. tabl ager, mis nol
```

-> tabulation of ager

Age at interview	Freq.	Percent	Cum.
15	1	0.01	0.01
16	3	0.03	0.04
17	9	0.09	0.13
18	23	0.23	0.35
19	36	0.35	0.70
20	34	0.33	1.04
21	67	0.66	1.69
22	109	1.07	2.76
23	119	1.16	3.93
24	190	1.86	5.79
25	220	2.15	7.94
26	251	2.46	10.40
27	366	3.58	13.98
28	313	3.06	17.04
29	364	3.56	20.61
30	385	3.77	24.38
31	388	3.80	28.17
32	488	4.78	32.95
33	540	5.29	38.24
34	463	4.53	42.77
35	486	4.76	47.53
36	458	4.48	52.01
37	399	3.91	55.92
38	447	4.38	60.29
39	405	3.96	64.26
40	370	3.62	67.88
41	356	3.49	71.37
42	322	3.15	74.52
43	355	3.48	77.99
44	365	3.57	81.57
45	337	3.30	84.87
46	413	4.04	88.91
47	407	3.98	92.89
48	356	3.49	96.38
49	363	3.55	99.93
50	7	0.07	100.00
Total	10,215	100.00	

.
 .
 . *****
 . * check the smoke variable
 . *****
 .

```
. tab1 postsmks, mis
```

```
-> tabulation of postsmks
```

```

BE-4 R |
smoked at |
all after R |
knew she |
was |
pregnant |      Freq.      Percent      Cum.
-----+-----
    Yes |         279         2.73         2.73
    No  |        2,293        22.45        25.18
  Refused |          1         0.01        25.19
    .    |        7,642        74.81       100.00
-----+-----
  Total |       10,215       100.00

```

```
. tab1 postsmks, mis nol
```

```
-> tabulation of postsmks
```

```

BE-4 R |
smoked at |
all after R |
knew she |
was |
pregnant |      Freq.      Percent      Cum.
-----+-----
    1 |         279         2.73         2.73
    5 |        2,293        22.45        25.18
    8 |          1         0.01        25.19
    . |        7,642        74.81       100.00
-----+-----
  Total |       10,215       100.00

```

```
. clonevar smoke = postsmks
```

```
(7,642 missing values generated)
```

```
. recode smoke (5=0) (8=.)
```

```
(smoke: 2294 changes made)
```

```
. label define smoke 1 "was smoking" 0 "was not smoking"
```

```
. label value smoke smoke
```

```
.
```

```
. tab2 postsmks smoke, mis
```

```
-> tabulation of postsmks by smoke
```

		BE-4 R smoked at all after R knew she was pregnant			Total
		was not s	was smoki	.	
Yes		0	279	0	279
No		2,293	0	0	2,293
Refused		0	0	1	1
.		0	0	7,642	7,642

Total		2,293	279	7,643	10,215

```
.
.
. *****
. * Regression analysis
. *****
```

```
. logit code_livebirth smoke ager
```

```
Iteration 0: log likelihood = -1342.643
Iteration 1: log likelihood = -1336.4905
Iteration 2: log likelihood = -1336.4002
Iteration 3: log likelihood = -1336.4002
```

```
Logistic regression                Number of obs   =    2,572
                                   LR chi2(2)         =    12.49
                                   Prob > chi2        =    0.0019
                                   Pseudo R2          =    0.0046

Log likelihood = -1336.4002
```

code_livebirth	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
smoke	-.5089268	.1404032	-3.62	0.000	-.784112	-.2337416
ager	-.0021932	.0077043	-0.28	0.776	-.0172934	.012907
_cons	1.417825	.2434322	5.82	0.000	.9407071	1.894944

```
.
. count
10,215

. count if code_livebirth ~=.
10,007

. count if code_livebirth ~=. & smoke ~=.
2,572
```

```
. count if code_livebirth ~= . & smoke ~= . & ager ~= .
2,572
```

```
.
.
. *****
. * The number of respondents with valid information on livebirth
. * A total of 10,007 livebirths came from 3,662 female respondents
. *****
```

```
. tab1 code_livebirth
```

```
-> tabulation of code_livebirth
```

code_livebirth	Freq.	Percent	Cum.
not a livebirth	2,810	28.08	28.08
livebirth	7,197	71.92	100.00
Total	10,007	100.00	

```
.
.
. gen livebirth_temp = 1 if code_livebirth ~= .
(208 missing values generated)
```

```
. sort caseid livebirth_temp
```

```
. by caseid: gen valid_livebirth = _n
```

```
. tab1 valid_livebirth
```

```
-> tabulation of valid_livebirth
```

valid_liveb	Freq.	Percent	Cum.
1	3,709	36.31	36.31
2	2,806	27.47	63.78
3	1,776	17.39	81.16
4	952	9.32	90.48
5	481	4.71	95.19
6	227	2.22	97.42
7	115	1.13	98.54
8	68	0.67	99.21
9	35	0.34	99.55
10	22	0.22	99.77

11	10	0.10	99.86
12	7	0.07	99.93
13	5	0.05	99.98
14	2	0.02	100.00

Total	10,215	100.00	

```
. tab1 valid_livebirth if livebirth_temp ==1
```

```
-> tabulation of valid_livebirth if livebirth_temp ==1
```

valid_liveb irth	Freq.	Percent	Cum.
1	3,662	36.59	36.59
2	2,756	27.54	64.14
3	1,727	17.26	81.39
4	927	9.26	90.66
5	468	4.68	95.33
6	217	2.17	97.50
7	109	1.09	98.59
8	63	0.63	99.22
9	35	0.35	99.57
10	20	0.20	99.77
11	10	0.10	99.87
12	6	0.06	99.93
13	5	0.05	99.98
14	2	0.02	100.00

Total	10,007	100.00	

```
.
.
.
. list caseid pregordr code_livebirth livebirth_temp valid_livebirth in 1/50, sepby(caseid) nol
```

	caseid	pregordr	code_l~h	livebi~p	valid_~h
1.	80719	1	1	1	1
2.	80719	2	1	1	2
3.	80719	3	1	1	3
4.	80719	4	1	1	4

5.	80720	1	0	1	1
6.	80720	2	0	1	2
7.	80720	3	1	1	3
8.	80720	4	1	1	4

9.	80723	1	1	1	1

10.	80723	2	1	1	2
11.	80723	3	1	1	3

12.	80727	1	0	1	1
13.	80727	2	1	1	2

14.	80728	1	0	1	1
15.	80728	2	0	1	2
16.	80728	3	1	1	3

17.	80729	1	0	1	1

18.	80731	1	1	1	1
19.	80731	2	0	1	2
20.	80731	3	1	1	3

21.	80733	1	0	1	1

22.	80742	1	1	1	1
23.	80742	2	1	1	2
24.	80742	3	1	1	3

25.	80743	1	1	1	1
26.	80743	2	1	1	2

27.	80747	1	1	1	1
28.	80747	2	1	1	2
29.	80747	3	1	1	3

30.	80756	1	0	1	1
31.	80756	2	1	1	2
32.	80756	3	1	1	3
33.	80756	4	0	1	4

34.	80760	1	1	1	1
35.	80760	2	1	1	2

36.	80764	1	0	1	1
37.	80764	2	1	1	2
38.	80764	3	1	1	3

39.	80766	1	1	1	1

40.	80768	1	1	1	1

41.	80772	1	1	1	1
42.	80772	2	1	1	2

43.	80773	1	0	1	1
44.	80773	2	1	1	2

```

-----+-----
45. | 80776      1      1      1      1 |
46. | 80776      2      0      1      2 |
47. | 80776      3      1      1      3 |
-----+-----
48. | 80785      1      1      1      1 |
49. | 80785      2      1      1      2 |
-----+-----
50. | 80787      1      1      1      1 |
-----+-----

```

```
. list caseid pregodr code_livebirth livebirth_temp valid_livebirth if inlist(caseid, 92033, 91773), sepby(caseid) nol
```

```

-----+-----
| caseid  pregodr  code_l~h  livebi~p  valid_~h |
-----+-----
9956. | 91773      1      .      .      1 |
-----+-----
10201. | 92033      1      1      1      1 |
10202. | 92033      2      1      1      2 |
10203. | 92033      3      .      .      3 |
-----+-----

```

```

.
.
.
. *****
. * The number of respondents with valid information on smoking
. * 2,293 pregnancy where mothers smoked during pregnancy.
. * These pregnancies were from 1,719 female respondents.
. *****
.
.

```

```
. tab1 smoke
```

```
-> tabulation of smoke
```

```

BE-4 R smoked |
at all after R |
knew she was |
pregnant |      Freq.      Percent      Cum.
-----+-----
was not smoking |      2,293      89.15      89.15
was smoking |      279      10.85      100.00
-----+-----
Total |      2,572      100.00

```

```
. gen smoke_temp = 1 if smoke ~=.
```

```
(7,643 missing values generated)
```

```
. sort caseid smoke_temp
```

```
. by caseid: gen valid_smoke = _n
```

```
. tab1 valid_smoke if smoke_temp ==1
```

```
-> tabulation of valid_smoke if smoke_temp ==1
```

valid_smoke	Freq.	Percent	Cum.
1	1,719	66.84	66.84
2	619	24.07	90.90
3	171	6.65	97.55
4	43	1.67	99.22
5	13	0.51	99.73
6	4	0.16	99.88
7	2	0.08	99.96
8	1	0.04	100.00
Total	2,572	100.00	

```
.
.
.
. *****
. * 3,709 female respondents reported age for 10,215 pregnancy records
. *****
```

```
. tab1 ager, mis
```

```
-> tabulation of ager
```

Age at interview	Freq.	Percent	Cum.
15 YEARS	1	0.01	0.01
16 YEARS	3	0.03	0.04
17 YEARS	9	0.09	0.13
18 YEARS	23	0.23	0.35
19 YEARS	36	0.35	0.70
20 YEARS	34	0.33	1.04
21 YEARS	67	0.66	1.69
22 YEARS	109	1.07	2.76
23 YEARS	119	1.16	3.93
24 YEARS	190	1.86	5.79
25 YEARS	220	2.15	7.94
26 YEARS	251	2.46	10.40
27 YEARS	366	3.58	13.98
28 YEARS	313	3.06	17.04
29 YEARS	364	3.56	20.61
30 YEARS	385	3.77	24.38

31 YEARS		388	3.80	28.17
32 YEARS		488	4.78	32.95
33 YEARS		540	5.29	38.24
34 YEARS		463	4.53	42.77
35 YEARS		486	4.76	47.53
36 YEARS		458	4.48	52.01
37 YEARS		399	3.91	55.92
38 YEARS		447	4.38	60.29
39 YEARS		405	3.96	64.26
40 YEARS		370	3.62	67.88
41 YEARS		356	3.49	71.37
42 YEARS		322	3.15	74.52
43 YEARS		355	3.48	77.99
44 YEARS		365	3.57	81.57
45 YEARS		337	3.30	84.87
46 YEARS		413	4.04	88.91
47 YEARS		407	3.98	92.89
48 YEARS		356	3.49	96.38
49		363	3.55	99.93
50		7	0.07	100.00

 Total | 10,215 100.00

. sort caseid ager

. by caseid: gen valid_ager = _n

. tab1 valid_ager if ager ~=. , mis

-> tabulation of valid_ager if ager ~=.

valid_ager		Freq.	Percent	Cum.
1		3,709	36.31	36.31
2		2,806	27.47	63.78
3		1,776	17.39	81.16
4		952	9.32	90.48
5		481	4.71	95.19
6		227	2.22	97.42
7		115	1.13	98.54
8		68	0.67	99.21
9		35	0.34	99.55
10		22	0.22	99.77
11		10	0.10	99.86
12		7	0.07	99.93
13		5	0.05	99.98
14		2	0.02	100.00

 Total | 10,215 100.00

```

.
.
.
. *****
. * valid cases of the analytic sample
. * 1,719 female respondents provide valid information on live births, smoking, and age for 2,572 pregnancies
. *****
.
. gen valid_temp = 1 if code_livebirth ~=. & smoke ~=. & age ~=.
(7,643 missing values generated)

. tab1 valid_temp, mis

-> tabulation of valid_temp

valid_temp |      Freq.   Percent   Cum.
-----+-----
      1 |      2,572    25.18    25.18
      . |      7,643    74.82   100.00
-----+-----
Total |     10,215   100.00

.
.
. sort caseid valid_temp

. by caseid: gen valid_all = _n

. tab1 valid_all if valid_temp ==1, mis

-> tabulation of valid_all if valid_temp ==1

valid_all |      Freq.   Percent   Cum.
-----+-----
      1 |      1,719    66.84    66.84
      2 |         619    24.07    90.90
      3 |         171     6.65    97.55
      4 |          43     1.67    99.22
      5 |          13     0.51    99.73
      6 |           4     0.16    99.88
      7 |           2     0.08    99.96
      8 |           1     0.04   100.00
-----+-----
Total |         2,572   100.00

.
.
. log close
name: <unnamed>
log: D:\Jason\workshop\Key Stata commands\key stata command2.log

```

log type: text
closed on: 11 Jul 2022, 11:17:29
