**RUTGERS UNIVERSITY**

# Bloustein School of Planning and Public Policy

**Applied Multivariate Methods**

**Fall 2020**

**Dawne Mouzon, Ph.D.**

**Problem Set #2**

**NAME: HASSAN KHURSHID**

**SUBMISSION INSTRUCTIONS: Please upload to Canvas by 11:59 pm next Thursday night.**

**ADVANCED RECODING AND COMPUTING**

**(100 points total)**

**Please copy and paste your code and the output that we specify after each question. Minimize the font if necessary. Your output must be copied/pasted. Do not use screenshots or pictures. Copy and paste your code from your do-file only (not the log); copy and paste your output from the results window.**

**Remember that the output we specifically ask for does not include every consistency check you should be running to check your own work. For example, while we sometimes ask you to list only the first 40 observations (it fits nicely on one page!), in order to identify any problems in your data, you should be reviewing as many cases as needed to check all potential outcomes (e.g., all categories on the output variable). Generally, what we need to grade you is less than what you should do to check your code throughout the assignment.**

**Please use the Courier New font, size 10 to make your output line up. Reduce the font size if necessary to make the output line up. Do not bold the output.**

**Begin each question on a new page.**

**PART I (30 PTS)**

**\*Using 1972-2018 GENERAL SOCIAL SURVEY\***

1. **(15 pts) Please create a composite variable (race\_eth) based on the variables for race (race) and Hispanic/Latino ethnicity (hispanic). You will need to create an interim variable (hisp; a dummy variable for Hispanic/Latino ethnicity). Anyone who reports Hispanic/Latino ancestry should be coded as Hispanic/Latino on race\_eth, regardless of racial category reported. Any cases missing data on Hispanic/Latino ethnicity should be coded as “.u” for “unknown Hispanic/Latino ancestry) on the new race\_eth variable.**

**For this problem, submit:**

* **Your code for the whole problem**
* **Frequencies for new race\_eth variable and hisp**
* **Any basic consistency checks for the interim variable**
* **Complex consistency checks (list function) showing at least five cases for each potential outcome, including id, race, hisp, and race\_eth**
  + **Hint: use set more on temporarily, when gathering output for this step**

**tab1 race hispanic, miss**

**recode hispanic (1 =0) (2/50 = 1) (.d .i .n = .u), gen(hisp)**

**label variable hisp "dummy for hispanic/latino ethnicity"**

**label define yn 0 "No" 1 "Yes" .u "unknown Hispanic/Latino ancestry"**

**label values hisp yn**

**numlabel yn, add**

**tab hispanic hisp, miss**

**tab race hisp, miss**

**gen race\_eth =.**

**label variable race\_eth "recode of race variable"**

**label define race\_eth 1 "White" 2 "Black" 3 "Hispanic/Latino" 4 "Other" .u "unknown Hispanic/Latino ancestry"**

**replace race\_eth = 1 if race == 1 & hisp == 0**

**replace race\_eth = 2 if race ==2 & hisp ==0**

**replace race\_eth = 3 if hisp ==1**

**replace race\_eth = 4 if race == 3 & hisp == 0**

**replace race\_eth = .u if hisp == .u**

**label values race\_eth race\_eth**

**numlabel race\_eth, add**

**tab race\_eth, miss**

**list id race hisp race\_eth if hisp== .u in 52000/52800**

**list id race hisp race\_eth if race ==2 & hisp ==0 in 40000/40055**

**list id race hisp race\_eth if race ==3 & hisp ==0 in 41000/41100**

**list id race hisp race\_eth if hisp ==1 in 50000/50010**

**list id race hisp race\_eth if race ==1 & hisp ==0 in 40000/40005**

. tab race\_eth, miss

recode of race variable | Freq. Percent Cum.

-------------------------------------+-----------------------------------

1. White | 18,613 28.72 28.72

2. Black | 3,877 5.98 34.70

3. Hispanic/Latino | 3,095 4.78 39.47

4. Other | 1,065 1.64 41.12

.u. unknown Hispanic/Latino ancestry | 38,164 58.88 100.00

-------------------------------------+-----------------------------------

Total | 64,814 100.00

. list id race hisp race\_eth if hisp== .u in 52000/52800

+--------------------------------------------------------+

52014. | id | race | hisp |

| 994 | 1. white | .u. unknown Hispanic/Latino ancestry |

|--------------------------------------------------------|

| race\_eth |

| .u. unknown Hispanic/Latino ancestry |

+--------------------------------------------------------+

+--------------------------------------------------------+

52429. | id | race | hisp |

| 1409 | 1. white | .u. unknown Hispanic/Latino ancestry |

|--------------------------------------------------------|

| race\_eth |

| .u. unknown Hispanic/Latino ancestry |

+--------------------------------------------------------+

+--------------------------------------------------------+

52566. | id | race | hisp |

| 1546 | 3. other | .u. unknown Hispanic/Latino ancestry |

|--------------------------------------------------------|

| race\_eth |

| .u. unknown Hispanic/Latino ancestry |

+--------------------------------------------------------+

. list id race hisp race\_eth if race ==2 & hisp ==0 in 40000/40055

+------------------------------------+

| id race hisp race\_eth |

|------------------------------------|

40018. | 1902 2. black 0. No 2. Black |

40045. | 1929 2. black 0. No 2. Black |

40047. | 1931 2. black 0. No 2. Black |

40048. | 1932 2. black 0. No 2. Black |

40049. | 1933 2. black 0. No 2. Black |

|------------------------------------|

40050. | 1934 2. black 0. No 2. Black |

40051. | 1935 2. black 0. No 2. Black |

40053. | 1937 2. black 0. No 2. Black |

40054. | 1938 2. black 0. No 2. Black |

40055. | 1939 2. black 0. No 2. Black |

+-----------------------------------+

. list id race hisp race\_eth if race ==3 & hisp ==0 in 41000/41100

+-----------------------------------+

| id race hisp race\_eth |

|-----------------------------------|

41034. | 101 3. other 0. No 4. Other |

+-----------------------------------+

. list id race hisp race\_eth if hisp ==1 in 50000/50010

+-----------------------------------------------+

| id race hisp race\_eth |

|-----------------------------------------------|

50008. | 3498 1. white 1. Yes 3. Hispanic/Latino |

50009. | 3499 1. white 1. Yes 3. Hispanic/Latino |

+-----------------------------------------------+

. list id race hisp race\_eth if race ==1 & hisp ==0 in 40000/40005

+------------------------------------+

| id race hisp race\_eth |

|------------------------------------|

40000. | 1884 1. white 0. No 1. White |

40001. | 1885 1. white 0. No 1. White |

40002. | 1886 1. white 0. No 1. White |

40003. | 1887 1. white 0. No 1. White |

40004. | 1888 1. white 0. No 1. White |

|------------------------------------|

40005. | 1889 1. white 0. No 1. White |

+------------------------------------+

1. **(15 pts) Please create a new variable that computes the number of weekday mins spent on the internet (wwwminwkdays) based on the number of weekday hours spent on the internet (intwkdyh) and the number of weekday minutes on the internet (intwkdym). Carry over extended missing values to the new variable.**

**For this problem, submit:**

* **Your code for the whole problem**
* **Frequencies of the new variable (wwwminwkdays)**
* **Extended descriptives (sum, detail) for the new variable (wwwminwkdays)**
* **Complex consistency checks of the first 40 non-missing observations**

**tab1 intwkdyh intwkdym, miss**

**tab intwkdyh intwkdym , miss**

**gen wwwminwkdays = (intwkdyh\*60) + intwkdym**

**label variable wwwminwkdays "weekday minutes on the internet"**

**replace wwwminwkdays = .i if intwkdyh == .i & intwkdym == .i /\*don't place 'or' sign \*/**

**replace wwwminwkdays = .n if intwkdyh == .n & intwkdym == .n**

**replace wwwminwkdays = .d if intwkdyh == .d & intwkdym == .d**

**tab wwwminwkdays, miss**

**sum wwwminwkdays, detail**

**list id intwkdyh intwkdym wwwminwkdays if !missing(wwwminwkdays), separator(40)**

tab wwwminwkdays, miss

weekday |

minutes on |

the |

internet | Freq. Percent Cum.

------------+-----------------------------------

0 | 6 0.01 0.01

1 | 1 0.00 0.01

2 | 1 0.00 0.01

3 | 1 0.00 0.01

5 | 2 0.00 0.02

10 | 12 0.02 0.04

15 | 14 0.02 0.06

20 | 15 0.02 0.08

25 | 1 0.00 0.08

30 | 81 0.12 0.21

40 | 3 0.00 0.21

45 | 11 0.02 0.23

55 | 1 0.00 0.23

60 | 301 0.46 0.69

90 | 39 0.06 0.75

119 | 1 0.00 0.76

120 | 245 0.38 1.13

150 | 5 0.01 1.14

180 | 127 0.20 1.34

210 | 3 0.00 1.34

240 | 81 0.12 1.47

270 | 2 0.00 1.47

300 | 61 0.09 1.56

330 | 3 0.00 1.57

360 | 51 0.08 1.65

390 | 1 0.00 1.65

420 | 14 0.02 1.67

480 | 40 0.06 1.73

510 | 1 0.00 1.73

540 | 11 0.02 1.75

570 | 1 0.00 1.75

600 | 30 0.05 1.80

660 | 2 0.00 1.80

720 | 22 0.03 1.84

780 | 2 0.00 1.84

810 | 1 0.00 1.84

840 | 4 0.01 1.85

900 | 9 0.01 1.86

960 | 1 0.00 1.86

1200 | 4 0.01 1.87

1440 | 8 0.01 1.88

. | 6 0.01 1.89

.d | 4 0.01 1.90

.i | 63,577 98.09 99.99

.n | 8 0.01 100.00

------------+-----------------------------------

Total | 64,814 100.00

. sum wwwminwkdays, detail

weekday minutes on the internet

-------------------------------------------------------------

Percentiles Smallest

1% 10 0

5% 30 0

10% 30 0 Obs 1,219

25% 60 0 Sum of Wgt. 1,219

50% 120 Mean 193.1829

Largest Std. Dev. 212.7356

75% 240 1440

90% 480 1440 Variance 45256.45

95% 600 1440 Skewness 2.627613

99% 960 1440 Kurtosis 12.23784

-------------------------------------------------------------

list id intwkdyh intwkdym wwwminwkdays if !missing(wwwminwkdays), separator(40)

+---------------------------------------+

| id intwkdyh intwkdym wwwmin~s |

|---------------------------------------|

59604. | 5 1 0 60 |

59605. | 6 1 0 60 |

59607. | 8 2 0 120 |

59609. | 10 6 0 360 |

59610. | 11 7 0 420 |

59612. | 13 12 0 720 |

59613. | 14 2 0 120 |

59618. | 19 3 0 180 |

59621. | 22 1 0 60 |

59622. | 23 3 0 180 |

59624. | 25 10 0 600 |

59627. | 28 3 0 180 |

59628. | 29 1 0 60 |

59629. | 30 2 0 120 |

59630. | 31 1 0 60 |

59631. | 32 0 30 30 |

59633. | 34 3 0 180 |

59635. | 36 2 0 120 |

59636. | 37 2 0 120 |

59638. | 39 10 0 600 |

59642. | 43 1 0 60 |

59643. | 44 6 0 360 |

59645. | 46 10 0 600 |

59646. | 47 5 0 300 |

59648. | 49 4 30 270 |

59649. | 50 7 0 420 |

59650. | 51 3 0 180 |

59652. | 53 4 0 240 |

59653. | 54 2 0 120 |

59655. | 56 2 0 120 |

59656. | 57 1 0 60 |

59657. | 58 2 0 120 |

59658. | 59 7 0 420 |

59660. | 61 2 0 120 |

59661. | 62 6 0 360 |

59663. | 64 2 0 120 |

59665. | 66 3 0 180 |

59666. | 67 24 0 1440 |

59678. | 79 4 0 240 |

59681. | 82 5 0 300 |

|---------------------------------------|

**PART II (70 PTS)**

**\*Using MIDUS\***

1. **(10 pts) Based on the rounded BMI variable (w2bmi\_rd), please create a new variable (w2bmi\_cat) for which 1=underweight (below 18.5); 2=normal weight (18.5-24.9); 3=overweight (25.0-29.9); 4=obese (30.0 and above). Be sure to run the appropriate consistency check to check your own work.**

**For this problem, submit:**

* **Your code for the whole problem**
* **Frequencies for w2bmi\_cat**
* **Complex consistency checks (list function) showing at least five cases for each potential outcome, including M2ID, w2bmi\_rd, and w2bmi\_cat**

**gen w2bmi\_cat =.**

**label define w2bmi\_cat 1 "underweight (below 18.5)" 2 "normal weight (18.5-24.9)" 3 "overweight (25.0-29.9)" 4 "obese (30.0 and above)"**

**label values w2bmi\_cat w2bmi\_cat**

**numlabel w2bmi\_cat, add**

**replace w2bmi\_cat =1 if w2bmi\_rd < 18.5**

**replace w2bmi\_cat =2 if w2bmi\_rd >= 18.5 & w2bmi\_rd <=24.9**

**replace w2bmi\_cat =3 if w2bmi\_rd >=25 & w2bmi\_rd <= 29.9**

**replace w2bmi\_cat =4 if w2bmi\_rd >= 30 & !missing(w2bmi\_rd)**

**tab w2bmi\_cat, miss**

**list M2ID w2bmi\_rd w2bmi\_cat if w2bmi\_rd == 41**

**list M2ID w2bmi\_rd w2bmi\_cat if w2bmi\_rd > 35 & !missing(w2bmi\_rd)**

**list M2ID w2bmi\_rd w2bmi\_cat if w2bmi\_rd > 15 & w2bmi\_rd <20**

**list M2ID w2bmi\_rd w2bmi\_cat if w2bmi\_rd < 18**

**list M2ID w2bmi\_rd w2bmi\_cat if w2bmi\_rd > 20 & !missing(w2bmi\_rd) in 4450/4470**

tab w2bmi\_cat, miss

w2bmi\_cat | Freq. Percent Cum.

-----------------------------+-----------------------------------

1. underweight (below 18.5) | 42 0.57 0.57

2. normal weight (18.5-24.9) | 1,177 15.99 16.56

3. overweight (25.0-29.9) | 1,498 20.35 36.91

4. obese (30.0 and above) | 1,097 14.90 51.81

. | 3,548 48.19 100.00

-----------------------------+-----------------------------------

Total | 7,362 100.00

list M2ID w2bmi\_rd w2bmi\_cat if w2bmi\_rd > 35 & !missing(w2bmi\_rd)

+----------------------------------------------+

| M2ID w2bmi\_rd w2bmi\_cat |

|----------------------------------------------|

30. | 10038 39.6 4. obese (30.0 and above) |

51. | 10066 42.9 4. obese (30.0 and above) |

70. | 10094 36 4. obese (30.0 and above) |

76. | 10100 36.9 4. obese (30.0 and above) |

115. | 10145 38.1 4. obese (30.0 and above) |

. list M2ID w2bmi\_rd w2bmi\_cat if w2bmi\_rd > 15 & w2bmi\_rd <20

+-------------------------------------------------+

| M2ID w2bmi\_rd w2bmi\_cat |

|-------------------------------------------------|

224. | 10287 19.1 2. normal weight (18.5-24.9) |

258. | 10330 19.2 2. normal weight (18.5-24.9) |

334. | 10419 19.4 2. normal weight (18.5-24.9) |

344. | 10431 18.9 2. normal weight (18.5-24.9) |

368. | 10461 19.8 2. normal weight (18.5-24.9) |

. list M2ID w2bmi\_rd w2bmi\_cat if w2bmi\_rd == 41

+----------------------------------------------+

| M2ID w2bmi\_rd w2bmi\_cat |

|----------------------------------------------|

2652. | 13319 41 4. obese (30.0 and above) |

3246. | 14064 41 4. obese (30.0 and above) |

6701. | 18441 41 4. obese (30.0 and above) |

+----------------------------------------------+

list M2ID w2bmi\_rd w2bmi\_cat if w2bmi\_rd < 18

+------------------------------------------------+

| M2ID w2bmi\_rd w2bmi\_cat |

|------------------------------------------------|

573. | 10707 16.5 1. underweight (below 18.5) |

716. | 10887 15.1 1. underweight (below 18.5) |

814. | 11021 17.8 1. underweight (below 18.5) |

945. | 11179 16.5 1. underweight (below 18.5) |

1242. | 11568 17 1. underweight (below 18.5) |

list M2ID w2bmi\_rd w2bmi\_cat if w2bmi\_rd > 20 & !missing(w2bmi\_rd) in 4450/4470

+-------------------------------------------------+

| M2ID w2bmi\_rd w2bmi\_cat |

|-------------------------------------------------|

4451. | 15606 22.6 2. normal weight (18.5-24.9) |

4452. | 15607 30 4. obese (30.0 and above) |

4453. | 15608 21 2. normal weight (18.5-24.9) |

4454. | 15609 23.6 2. normal weight (18.5-24.9) |

4458. | 15616 30.2 4. obese (30.0 and above) |

|-------------------------------------------------|

4465. | 15625 35.1 4. obese (30.0 and above) |

4466. | 15626 25.1 3. overweight (25.0-29.9) |

4467. | 15627 26.1 3. overweight (25.0-29.9) |

4469. | 15629 26.7 3. overweight (25.0-29.9) |

4470. | 15630 41.6 4. obese (30.0 and above) |

+-------------------------------------------------+

1. **(30 pts) In Wave 3 of MIDUS, respondents were asked to rate their overall life satisfaction now (C1SQ1) and their life satisfaction 10 years ago (C1SQ2). Create a new variable (retrolifesat) comparing their life satisfaction 10 years ago to their life satisfaction currently. The resulting variable should be interval-ratio, with extended missing values for .i=”R does not have SAQ data” and .r=”refused.” Negative values should reflect worse life satisfaction now than 10 years ago. Carry over any extended missing values. *Remember: garbage in, garbage out. You will have to construct interim variables.***

**For this problem, submit:**

* **A detailed summary of your approach to this problem, including variable names**
* **Your code for the whole problem**
* **Frequencies and extended descriptives for the new variable retrolifesat (there should be value labels on the smallest value, largest value, and extended missing values on this variable)**
* **Frequencies and consistency checks for any new interim variables created**
* **Consistency check (list function) for the first 50 observations, including M2ID, life satisfaction now, life satisfaction 10 years ago and life satisfaction compared**

**Approach:**

First I made the recodes of variables C1SQ1 and C1SQ2 by cleaning them up and assigning them labels as per the instructions. After running crosstabs on input variables, I generated a new variable by subtracting life satisfaction currently from life satisfaction 10 years back to give difference. To incorporate the extended missing data, I used replace function and did that in two such lines.

Lastly, I listed the new variable and summarized it in detail.

**tab1 C1SQ1 C1SQ2, miss**

**tab C1SQ1 C1SQ2, miss**

**recode C1SQ1 (-1 =.i) (98 = .r), gen(LifeNow)**

**recode C1SQ2 (-1 =.i) (98 = .r), gen(LifeThen)**

**label variable LifeNow "Life Satisfaction Currently"**

**label variable LifeThen "Life Satisfaction 10 years ago"**

**label define life .i "R does not have SAQ data" .r "refused"**

**label values LifeThen LifeNow life**

**numlabel life, add**

**tab1 C1SQ1 LifeNow, miss**

**tab1 C1SQ2 LifeThen, miss**

**tab LifeNow LifeThen, miss**

**gen retrolifesat = LifeNow - LifeThen**

**label variable retrolifesat "Comparison of current vs previous satisfaction"**

**replace retrolifesat = .i if LifeThen == .i & LifeNow == .i**

**replace retrolifesat = .r if LifeThen == .r & LifeNow == .r**

**label define retrolifesat -10 "Smallest Difference" 10 "Largest Difference" .i "R does not have SAQ data" .r "refused"**

**label values retrolifesat retrolifesat**

**numlabel retrolifesat, add**

**tab retrolifesat, miss**

**sum retrolifesat, detail**

**list M2ID LifeThen LifeNow retrolifesat in 1/50**

. tab1 C1SQ1 LifeNow, miss

-> tabulation of C1SQ1

Rate life overall currently | Freq. Percent Cum.

--------------------------------------+-----------------------------------

-1. RESPONDENT DOES NOT HAVE SAQ DATA | 576 7.82 7.82

0. WORST | 5 0.07 7.89

1 | 5 0.07 7.96

2 | 9 0.12 8.08

3 | 45 0.61 8.69

4 | 36 0.49 9.18

5 | 127 1.73 10.91

6 | 152 2.06 12.97

7 | 390 5.30 18.27

8 | 841 11.42 29.69

9 | 719 9.77 39.46

10. BEST | 312 4.24 43.70

98. REFUSED | 77 1.05 44.74

. | 4,068 55.26 100.00

--------------------------------------+-----------------------------------

Total | 7,362 100.00

-> tabulation of LifeNow

Life Satisfaction Currently | Freq. Percent Cum.

-----------------------------+-----------------------------------

0 | 5 0.07 0.07

1 | 5 0.07 0.14

2 | 9 0.12 0.26

3 | 45 0.61 0.87

4 | 36 0.49 1.36

5 | 127 1.73 3.08

6 | 152 2.06 5.15

7 | 390 5.30 10.45

8 | 841 11.42 21.87

9 | 719 9.77 31.64

10 | 312 4.24 35.87

. | 4,068 55.26 91.13

.i. R does not have SAQ data | 576 7.82 98.95

.r. refused | 77 1.05 100.00

-----------------------------+-----------------------------------

Total | 7,362 100.00

. tab1 C1SQ2 LifeThen, miss

-> tabulation of C1SQ2

Rate life overall 10 years ago | Freq. Percent Cum.

--------------------------------------+-----------------------------------

-1. RESPONDENT DOES NOT HAVE SAQ DATA | 576 7.82 7.82

0. WORST | 3 0.04 7.86

1 | 6 0.08 7.95

2 | 15 0.20 8.15

3 | 28 0.38 8.53

4 | 56 0.76 9.29

5 | 140 1.90 11.19

6 | 204 2.77 13.96

7 | 400 5.43 19.40

8 | 777 10.55 29.95

9 | 689 9.36 39.31

10. BEST | 343 4.66 43.97

98. REFUSED | 57 0.77 44.74

. | 4,068 55.26 100.00

--------------------------------------+-----------------------------------

Total | 7,362 100.00

-> tabulation of LifeThen

Life Satisfaction 10 years |

ago | Freq. Percent Cum.

-----------------------------+-----------------------------------

0 | 3 0.04 0.04

1 | 6 0.08 0.12

2 | 15 0.20 0.33

3 | 28 0.38 0.71

4 | 56 0.76 1.47

5 | 140 1.90 3.37

6 | 204 2.77 6.14

7 | 400 5.43 11.57

8 | 777 10.55 22.13

9 | 689 9.36 31.49

10 | 343 4.66 36.15

. | 4,068 55.26 91.40

.i. R does not have SAQ data | 576 7.82 99.23

.r. refused | 57 0.77 100.00

-----------------------------+-----------------------------------

Total | 7,362 100.00

tab LifeNow LifeThen, miss

Life Satisfaction | Life Satisfaction 10 years ago

Currently | 0 1 2 3 4 5 6 7 | Total

----------------------+----------------------------------------------------------------------------------------+----------

0 | 1 1 0 0 1 0 0 0 | 5

1 | 0 0 0 0 1 0 1 0 | 5

2 | 0 0 1 1 1 3 1 2 | 9

3 | 0 1 0 5 5 9 2 11 | 45

4 | 0 0 1 0 5 2 8 9 | 36

5 | 0 0 0 3 4 38 13 30 | 127

6 | 0 1 4 3 8 13 31 27 | 152

7 | 0 0 2 5 12 27 50 128 | 390

8 | 0 1 2 7 13 30 65 129 | 841

9 | 0 2 4 4 3 13 25 53 | 719

10 | 2 0 1 0 3 4 6 8 | 312

. | 0 0 0 0 0 0 0 0 | 4,068

.i. R does not have S | 0 0 0 0 0 0 0 0 | 576

.r. refused | 0 0 0 0 0 1 2 3 | 77

----------------------+----------------------------------------------------------------------------------------+----------

Total | 3 6 15 28 56 140 204 400 | 7,362

Life Satisfaction | Life Satisfaction 10 years ago

Currently | 8 9 10 . .i. R doe .r. refus | Total

----------------------+------------------------------------------------------------------+----------

0 | 0 0 2 0 0 0 | 5

1 | 1 1 1 0 0 0 | 5

2 | 0 0 0 0 0 0 | 9

3 | 9 3 0 0 0 0 | 45

4 | 6 5 0 0 0 0 | 36

5 | 23 10 6 0 0 0 | 127

6 | 40 18 7 0 0 0 | 152

7 | 104 52 10 0 0 0 | 390

8 | 412 147 35 0 0 0 | 841

9 | 141 424 50 0 0 0 | 719

10 | 36 24 228 0 0 0 | 312

. | 0 0 0 4,068 0 0 | 4,068

.i. R does not have S | 0 0 0 0 576 0 | 576

.r. refused | 5 5 4 0 0 57 | 77

----------------------+------------------------------------------------------------------+----------

Total | 777 689 343 4,068 576 57 | 7,362

. tab retrolifesat, miss

Comparison of current vs |

previous satisfaction | Freq. Percent Cum.

-----------------------------+-----------------------------------

-10. Smallest Difference | 2 0.03 0.03

-9 | 1 0.01 0.04

-8 | 1 0.01 0.05

-7 | 1 0.01 0.07

-6 | 3 0.04 0.11

-5 | 23 0.31 0.42

-4 | 36 0.49 0.91

-3 | 66 0.90 1.81

-2 | 175 2.38 4.18

-1 | 350 4.75 8.94

0 | 1,273 17.29 26.23

1 | 361 4.90 31.13

2 | 194 2.64 33.77

3 | 78 1.06 34.83

4 | 41 0.56 35.38

5 | 17 0.23 35.62

6 | 9 0.12 35.74

7 | 5 0.07 35.81

8 | 3 0.04 35.85

10. Largest Difference | 2 0.03 35.87

. | 4,088 55.53 91.40

.i. R does not have SAQ data | 576 7.82 99.23

.r. refused | 57 0.77 100.00

-----------------------------+-----------------------------------

Total | 7,362 100.00

. sum retrolifesat, detail

Comparison of current vs previous satisfaction

-------------------------------------------------------------

Percentiles Smallest

1% -5 -10

5% -3 -10

10% -2 -9 Obs 2,641

25% 0 -8 Sum of Wgt. 2,641

50% 0 Mean .0552821

Largest Std. Dev. 1.647993

75% 1 8

90% 2 8 Variance 2.715882

95% 3 10 Skewness .154942

99% 5 10 Kurtosis 7.981263

. list M2ID LifeThen LifeNow retrolifesat in 1/50

+----------------------------------------------------------------------------------------------------+

| M2ID LifeThen LifeNow retrolifesat |

|----------------------------------------------------------------------------------------------------|

1. | 10001 6 8 2 |

2. | 10002 .i. R does not have SAQ data .i. R does not have SAQ data .i. R does not have SAQ data |

3. | 10004 . . . |

4. | 10005 . . . |

5. | 10006 . . . |

|----------------------------------------------------------------------------------------------------|

6. | 10007 . . . |

7. | 10008 . . . |

8. | 10009 . . . |

9. | 10010 . . . |

10. | 10011 4 9 5 |

|----------------------------------------------------------------------------------------------------|

11. | 10013 . . . |

12. | 10014 . . . |

13. | 10015 5 7 2 |

14. | 10016 . . . |

15. | 10017 . . . |

|----------------------------------------------------------------------------------------------------|

16. | 10018 . . . |

17. | 10019 8 8 0 |

18. | 10020 .i. R does not have SAQ data .i. R does not have SAQ data .i. R does not have SAQ data |

19. | 10021 . . . |

20. | 10023 . . . |

|----------------------------------------------------------------------------------------------------|

21. | 10024 7 7 0 |

22. | 10025 . . . |

23. | 10027 . . . |

24. | 10028 . . . |

25. | 10030 10 10 0 |

|----------------------------------------------------------------------------------------------------|

26. | 10034 . . . |

27. | 10035 . . . |

28. | 10036 7 7 0 |

29. | 10037 6 5 -1 |

30. | 10038 4 7 3 |

|----------------------------------------------------------------------------------------------------|

31. | 10039 . . . |

32. | 10040 6 8 2 |

33. | 10041 . . . |

34. | 10042 . . . |

35. | 10044 . . . |

|----------------------------------------------------------------------------------------------------|

36. | 10046 8 9 1 |

37. | 10047 7 8 1 |

38. | 10049 . . . |

39. | 10050 . . . |

40. | 10051 . . . |

|----------------------------------------------------------------------------------------------------|

41. | 10052 . . . |

42. | 10053 . . . |

43. | 10055 . . . |

44. | 10056 . . . |

45. | 10059 . . . |

|----------------------------------------------------------------------------------------------------|

46. | 10060 8 10 2 |

47. | 10061 9 9 0 |

48. | 10062 8 6 -2 |

49. | 10063 8 7 -1 |

50. | 10065 . . . |

+----------------------------------------------------------------------------------------------------+

1. **(30 pts) Create a new variable (prosplifesat for prospective life satisfaction) comparing respondents’ current life satisfaction to their anticipated life satisfaction 10 years from now (C1SQ3). The resulting variable should be interval-ratio, with extended missing values for .i=”R does not have SAQ data” and .r=”refused.” Negative values should reflect worse anticipated life satisfaction in 10 years than now. Carry over any extended missing values. *Remember: garbage in, garbage out. You will have to construct interim variables.***

**For this problem, submit:**

* **A detailed summary of your approach to this problem, including variable names**
* **Your code for the whole problem**
* **Frequencies and extended descriptives for the new variable prosplifesat (there should be value labels on the smallest value, largest value, and extended missing values on this variable)**
* **Frequencies and consistency checks for any new interim variables created**
* **Consistency check (list function) for the first 50 observations, including M2ID, life satisfaction now, life satisfaction 10 years in the future, and the new variable for prospective life satisfaction**

**Approach:**

Like the previous question, I recoded, this time, the prospective life satisfaction variable C1SQ3 by cleaning it up and introducing the proper labels. Then I generated a new label called prosplifesat which is the difference between prospective life satisfaction and current life satisfaction rating. To incorporate the extended missing data, I used replace function and did that in two such lines.

Lastly I listed the new variable

**tab1 C1SQ1 C1SQ3, miss**

**tab C1SQ1 C1SQ3, miss**

**recode C1SQ3 (-1 =.i) (98 = .r), gen(LifeAfterwards)**

**label variable LifeAfterwards "Anticipated Life Satisfaction"**

**label values LifeAfterwards life**

**numlabel life, add**

**tab1 C1SQ3 LifeAfterwards, miss**

**tab LifeAfterwards LifeNow, miss**

**gen prosplifesat = LifeAfterwards - LifeNow**

**label variable prosplifesat "Comparison of current vs prospective life satisfaction"**

**replace prosplifesat= .i if LifeAfterwards == .i & LifeNow == .i**

**replace prosplifesat = .r if LifeAfterwards == .r & LifeNow == .r**

**tab prosplifesat, miss**

**sum prosplifesat, detail**

**list M2ID LifeNow LifeAfterwards prosplifesat in 1/50**

tab1 C1SQ3 LifeAfterwards, miss

-> tabulation of C1SQ3

Rate life overall 10 years future | Freq. Percent Cum.

--------------------------------------+-----------------------------------

-1. RESPONDENT DOES NOT HAVE SAQ DATA | 576 7.82 7.82

0. WORST | 23 0.31 8.14

1 | 15 0.20 8.34

2 | 35 0.48 8.82

3 | 42 0.57 9.39

4 | 74 1.01 10.39

5 | 157 2.13 12.52

6 | 148 2.01 14.53

7 | 318 4.32 18.85

8 | 658 8.94 27.79

9 | 763 10.36 38.16

10. BEST | 402 5.46 43.62

98. REFUSED | 83 1.13 44.74

. | 4,068 55.26 100.00

--------------------------------------+-----------------------------------

Total | 7,362 100.00

-> tabulation of LifeAfterwards

Anticipated Life |

Satisfaction | Freq. Percent Cum.

-----------------------------+-----------------------------------

0 | 23 0.31 0.31

1 | 15 0.20 0.52

2 | 35 0.48 0.99

3 | 42 0.57 1.56

4 | 74 1.01 2.57

5 | 157 2.13 4.70

6 | 148 2.01 6.71

7 | 318 4.32 11.03

8 | 658 8.94 19.97

9 | 763 10.36 30.33

10 | 402 5.46 35.79

. | 4,068 55.26 91.05

.i. R does not have SAQ data | 576 7.82 98.87

.r. refused | 83 1.13 100.00

-----------------------------+-----------------------------------

Total | 7,362 100.00

. tab LifeAfterwards LifeNow, miss

Anticipated Life | Life Satisfaction Currently

Satisfaction | 0 1 2 3 4 5 6 7 | Total

----------------------+----------------------------------------------------------------------------------------+----------

0 | 3 1 2 3 2 1 1 6 | 23

1 | 0 2 1 2 1 4 0 1 | 15

2 | 0 1 1 6 8 7 1 3 | 35

3 | 0 0 0 5 5 12 3 10 | 42

4 | 0 0 1 2 4 16 16 12 | 74

5 | 1 1 0 8 4 46 21 26 | 157

6 | 0 0 2 5 1 9 33 40 | 148

7 | 0 0 1 7 4 16 30 123 | 318

8 | 0 0 1 4 3 11 30 108 | 658

9 | 0 0 0 2 1 1 11 51 | 763

10 | 1 0 0 0 1 1 3 6 | 402

. | 0 0 0 0 0 0 0 0 | 4,068

.i. R does not have S | 0 0 0 0 0 0 0 0 | 576

.r. refused | 0 0 0 1 2 3 3 4 | 83

----------------------+----------------------------------------------------------------------------------------+----------

Total | 5 5 9 45 36 127 152 390 | 7,362

Anticipated Life | Life Satisfaction Currently

Satisfaction | 8 9 10 . .i. R doe .r. refus | Total

----------------------+------------------------------------------------------------------+----------

0 | 1 0 3 0 0 0 | 23

1 | 4 0 0 0 0 0 | 15

2 | 5 2 1 0 0 0 | 35

3 | 5 2 0 0 0 0 | 42

4 | 18 4 0 0 0 1 | 74

5 | 30 10 9 0 0 1 | 157

6 | 46 9 1 0 0 2 | 148

7 | 99 32 6 0 0 0 | 318

8 | 376 100 19 0 0 6 | 658

9 | 203 453 38 0 0 3 | 763

10 | 49 101 234 0 0 6 | 402

. | 0 0 0 4,068 0 0 | 4,068

.i. R does not have S | 0 0 0 0 576 0 | 576

.r. refused | 5 6 1 0 0 58 | 83

----------------------+------------------------------------------------------------------+----------

Total | 841 719 312 4,068 576 77 | 7,362

tab prosplifesat, miss

Comparison |

of current |

vs |

prospective |

life |

satisfactio |

n | Freq. Percent Cum.

------------+-----------------------------------

-10 | 3 0.04 0.04

-8 | 2 0.03 0.07

-7 | 12 0.16 0.23

-6 | 9 0.12 0.35

-5 | 22 0.30 0.65

-4 | 46 0.62 1.28

-3 | 71 0.96 2.24

-2 | 163 2.21 4.46

-1 | 327 4.44 8.90

0 | 1,280 17.39 26.28

1 | 458 6.22 32.50

2 | 156 2.12 34.62

3 | 37 0.50 35.13

4 | 17 0.23 35.36

5 | 8 0.11 35.47

6 | 4 0.05 35.52

10 | 1 0.01 35.53

. | 4,112 55.85 91.39

.i | 576 7.82 99.21

.r | 58 0.79 100.00

------------+-----------------------------------

Total | 7,362 100.00

. sum prosplifesat, detail

Comparison of current vs prospective life

satisfaction

-------------------------------------------------------------

Percentiles Smallest

1% -5 -10

5% -3 -10

10% -2 -10 Obs 2,616

25% -1 -8 Sum of Wgt. 2,616

50% 0 Mean -.1227064

Largest Std. Dev. 1.556533

75% 1 6

90% 1 6 Variance 2.422796

95% 2 6 Skewness -1.047892

99% 4 10 Kurtosis 8.975668

. list M2ID LifeNow LifeAfterwards prosplifesat in 1/50

+--------------------------------------------------------------------------------+

| M2ID LifeNow LifeAfterwards prospl~t |

|--------------------------------------------------------------------------------|

1. | 10001 8 8 0 |

2. | 10002 .i. R does not have SAQ data .i. R does not have SAQ data .i |

3. | 10004 . . . |

4. | 10005 . . . |

5. | 10006 . . . |

|--------------------------------------------------------------------------------|

6. | 10007 . . . |

7. | 10008 . . . |

8. | 10009 . . . |

9. | 10010 . . . |

10. | 10011 9 9 0 |

|--------------------------------------------------------------------------------|

11. | 10013 . . . |

12. | 10014 . . . |

13. | 10015 7 9 2 |

14. | 10016 . . . |

15. | 10017 . . . |

|--------------------------------------------------------------------------------|

16. | 10018 . . . |

17. | 10019 8 7 -1 |

18. | 10020 .i. R does not have SAQ data .i. R does not have SAQ data .i |

19. | 10021 . . . |

20. | 10023 . . . |

|--------------------------------------------------------------------------------|

21. | 10024 7 5 -2 |

22. | 10025 . . . |

23. | 10027 . . . |

24. | 10028 . . . |

25. | 10030 10 10 0 |

|--------------------------------------------------------------------------------|

26. | 10034 . . . |

27. | 10035 . . . |

28. | 10036 7 7 0 |

29. | 10037 5 8 3 |

30. | 10038 7 8 1 |

|--------------------------------------------------------------------------------|

31. | 10039 . . . |

32. | 10040 8 8 0 |

33. | 10041 . . . |

34. | 10042 . . . |

35. | 10044 . . . |

|--------------------------------------------------------------------------------|

36. | 10046 9 9 0 |

37. | 10047 8 8 0 |

38. | 10049 . . . |

39. | 10050 . . . |

40. | 10051 . . . |

|--------------------------------------------------------------------------------|

41. | 10052 . . . |

42. | 10053 . . . |

43. | 10055 . . . |

44. | 10056 . . . |

45. | 10059 . . . |

|--------------------------------------------------------------------------------|

46. | 10060 10 10 0 |

47. | 10061 9 9 0 |

48. | 10062 6 5 -1 |

49. | 10063 7 9 2 |

50. | 10065 . . . |

+--------------------------------------------------------------------------------+