Biostatistics 203A: Introduction to Data Management and Statistical Computing

Lab Assignment 3: Submission Template

Fall 2023

Name: Yuhui Wang
Student ID Number: 606332401

Exercise 1

[Insert code used to accomplish this exercise. Additionally, insert Proc Contents output here by copying and pasting the table titled 'Alphabetic List of Variables and Attributes'.]

```
*1;
proc transpose data=lb.hlth_2009 out=hlth_very_long (rename=(COL1=INDICATOR)) name=WAVE;
by IND_ID HH_ID;
var HEADACHE_2004
   HEADACHE_2006
   HEADACHE_2009
   SORETHROAT_2004
   SORETHROAT_2006
   SORETHROAT_2009
   STOMACHACHE_2004
   STOMACHACHE_2006
   STOMACHACHE_2009;
run;

data hlth_very_long;
retain IND_ID HH_ID WAVE SYMPTOM INDICATOR;
set hlth_very_long;
SYMPTOM=compress(WAVE, '_', 'd');
WAVE = compress(WAVE, '_', 'a');
run;

proc contents data=hlth_very_long;
run;
```

Alphabetic List of Variables and Attributes				
#	Variable	Type	Len	Label
2	HH_ID	Num	8	HOUSEHOLD ID
5	INDICATOR	Num	8	
1	IND_ID	Num	8	INDIVIDUAL ID
4	SYMPTOM	Char	16	
3	WAVE	Char	16	NAME OF FORMER VARIABLE

Exercise 2

[Insert code used to accomplish this exercise. Additionally, insert Proc Contents output here by copying and pasting the table titled 'Alphabetic List of Variables and Attributes'.]

```
*2;
proc transpose data=lb.hlth_2009 out=hlth_head (rename=(col1=HEADACHE)) name=WAVE;
by IND_ID HH_ID HH_TYPE;
var HEADACHE_2004
   HEADACHE_2006
   HEADACHE_2009;
run;

data hlth_head;
set hlth_head;
WAVE = compress(WAVE, '_', 'a');

proc transpose data=lb.hlth_2009 out=hlth_sore (rename=(col1=SORETHROAT)) name=WAVE;
by IND_ID HH_ID HH_TYPE;
var SORETHROAT_2004
   SORETHROAT_2006
   SORETHROAT_2009;
run;

data hlth_sore;
set hlth_sore;
WAVE = compress(WAVE, '_', 'a');

proc transpose data=lb.hlth_2009 out=hlth_stomach (rename=(col1=STOMACHACHE)) name=WAVE;
by IND_ID HH_ID HH_TYPE;
var STOMACHACHE_2004
   STOMACHACHE_2006
   STOMACHACHE_2009;
```

```
run;

data hlth_stomach;
set hlth_stomach;
WAVE = compress(WAVE, '_', 'a');

data hlth_long;
merge hlth_head hlth_sore hlth_stomach;
by IND_ID HH_ID WAVE HH_TYPE;
run;

proc contents data=hlth_long;
run;
```

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Label
5	HEADACHE	Num	8	
2	HH_ID	Num	8	HOUSEHOLD ID
3	HH_TYPE	Num	8	1=URBAN SITE(U) 2=RURAL SITE(R)
1	IND_ID	Num	8	INDIVIDUAL ID
6	SORETHROAT	Num	8	
7	STOMACHACHE	Num	8	
4	WAVE	Char	13	NAME OF FORMER VARIABLE

Exercise 3

[This is just one example of a table format that would suffice. Other formats are acceptable.]

		NAME OF FORMER VARIABLE		
		2004	2006	2009
		%	%	%
1=URBAN SITE(U) 2=RURAL SITE(R)				
Urban	HEADACHE	6.15%	6.26%	5.42%
	SORETHROAT	11.73%	9.70%	8.83%
	STOMACHACHE	4.67%	3.80%	2.04%
Rural	HEADACHE	4.63%	4.00%	3.94%
	SORETHROAT	7.91%	6.75%	8.05%
	STOMACHACHE	2.88%	2.39%	1.50%
All	HEADACHE	5.14%	4.75%	4.43%
	SORETHROAT	9.18%	7.73%	8.31%
	STOMACHACHE	3.47%	2.86%	1.68%

```
*3;
proc format;
value dwelling 1="Urban"
                              2="Rural";
run;

proc tabulate data=hlth_long;
format hh_type dwelling.;
class wave hh_type;
var headache sorethroat stomachache;
table (hh_type All)*(headache sorethroat stomachache),
        wave*(mean="%"*F=percent8.2);
run;
```

Exercise 4

[Insert code used to accomplish this exercise. Additionally, insert Proc Contents output here by copying and pasting the table titled 'Alphabetic List of Variables and Attributes'.]

```
*4;
data hlth_array_long_headache;
 set lb.hlth_2009;
 array headache_array{3}
      HEADACHE_2004
```

```
      HEADACHE_2006
      HEADACHE_2009;
 array wv{3}$ wv1-wv3 ('2004','2006','2009');
 do i = 1 to 3;
  WAVE = wv{i};
  HEADACHE = headache_array{i};
  output;
 end;
 keep IND_ID HH_ID WAVE HEADACHE HH_TYPE;
run;

data hlth_array_long_sorethroat;
 set lb.hlth_2009;
 array sorethroat_array{3}
      SORETHROAT_2004
      SORETHROAT_2006
      SORETHROAT_2009;
 array wv{3}$ wv1-wv3 ('2004','2006','2009');
 do i = 1 to 3;
  WAVE = wv{i};
  SORETHROAT = sorethroat_array{i};
  output;
 end;
 keep IND_ID HH_ID WAVE SORETHROAT HH_TYPE;
run;

data hlth_array_long_stomachache;
 set lb.hlth_2009;
 array stomachache_array{3}
      STOMACHACHE_2004
      STOMACHACHE_2006
      STOMACHACHE_2009;
 array wv{3}$ wv1-wv3 ('2004','2006','2009');
 do i = 1 to 3;
  WAVE = wv{i};
  STOMACHACHE = stomachache_array{i};
  output;
 end;
 keep IND_ID HH_ID WAVE STOMACHACHE HH_TYPE;
run;

data hlth_array_long;
merge hlth_array_long_headache hlth_array_long_sorethroat hlth_array_long_stomachache;
by IND_ID HH_ID WAVE HH_TYPE;
```

```
run;

proc contents data=hlth_array_long;
run;
```

Alphabetic List of Variables and Attributes				
#	Variable	Type	Len	Label
5	HEADACHE	Num	8	
2	HH_ID	Num	8	HOUSEHOLD ID
3	HH_TYPE	Num	8	1=URBAN SITE(U) 2=RURAL SITE(R)
1	IND_ID	Num	8	INDIVIDUAL ID
6	SORETHROAT	Num	8	
7	STOMACHACHE	Num	8	
4	WAVE	Char	8	

Exercise 5

[Insert code used to accomplish this exercise. Additionally, insert Proc Print output displaying the first 8 observations of hlth_wide.]

```
*5;
proc transpose data=hlth_array_long out=hlth_wide_headache (drop=_NAME_) prefix= HEADACHE_;
by IND_ID HH_ID HH_TYPE;
id WAVE;
var HEADACHE;
run;

proc transpose data=hlth_array_long out=hlth_wide_sorethroat (drop=_NAME_) prefix= SORETHROAT_;
by IND_ID HH_ID HH_TYPE;
id WAVE;
var SORETHROAT;
run;

proc transpose data=hlth_array_long out=hlth_wide_stomachache (drop=_NAME_) prefix=
STOMACHACHE_;
by IND_ID HH_ID HH_TYPE;
id WAVE;
var STOMACHACHE;
run;
```

```
data hlth_wide;
merge hlth_wide_headache
        hlth_wide_sorethroat
        hlth_wide_stomachache;
by IND_ID HH_ID HH_TYPE;
run;

proc print data=hlth_wide (obs=8);
run;
```

Obs	IND_ID	HH_ID	HH_TYPE	HEADACHE_2004	HEADACHE_2006	HEADACHE_2009	SORETHROAT_2004	SORETHROAT_2006	SORETHROAT_2009	STOMACHACHE_2004	STOMACHACHE_2006	STOMACHACHE_2009
1	211101003002	211101003	1	0	.	1	0	.	1	0	.	1
2	211101003101	211101003	1	.	0	.	.	0	.	.	0	.
3	211101003102	211101003	1	.	0	.	.	0	.	.	0	.
4	211101008001	211101008	1	0	0	.	0	0	.	0	0	.
5	211101008002	211101008	1	0	0	0	0	0	0	0	0	0
6	211101008003	211101008	1	0	0	.	0	0	.	0	0	.
7	211101008005	211101008	1	.	0	0	.	0	0	.	0	0
8	211101008021	211101008	1	0	0	0	0	0	0	0	0	0

Exercise 6

	N	**%**
Individuals with surveys completed at each of the following time points:		
2004 and at least one subsequent time point	9826	55.5
2004 and 2006	9120	51.51
2004, 2006, and 2009	6733	38.03
2006 and 2009 (but not 2004)	1532	8.65

[Also insert code used to accomplish this task]

```
*6;
data question;
 set hlth_wide;
 MISS_COUNT_0609 = CMISS(HEADACHE_2006, HEADACHE_2009);
run;

data question;
 set question;
 if (HEADACHE_2004 ne .) and (MISS_COUNT_0609 ne 2)
        then WV04_AND_06OR09 = 1;
 else WV04_AND_06OR09 = 0;

 if (HEADACHE_2004 ne .) and (HEADACHE_2006 ne .)
        then WV0406 =1;
 else WV0406 =0;

 if (HEADACHE_2004 ne .) and (HEADACHE_2006 ne .) and (HEADACHE_2009 ne .)
```

```
        then WV040609=1;
 else WV040609=0;

 if (HEADACHE_2006 ne .) and (HEADACHE_2009 ne .) and (HEADACHE_2004 = .)
        then WV0609_NOT04=1;
 else WV0609_NOT04=0;
run;

proc freq data=question;
tables WV04_AND_06OR09 WV0406 WV040609 WV0609_NOT04/nocum list;
run;
```

Exercise 7

	N	%
Individuals with surveys completed at the following time points:		
2004	12147	33.95
2006	11714	32.74
2009	11922	33.32

[Also insert code used to accomplish this task]

```
*7;
proc freq data=hlth_long (where=(headache ne .));
tables wave/nocum;
run;
```

//////////////////////

If all surveys completed means all of these three surveys not equal to ., then the result should be below:

	N	%
Individuals with surveys completed at the following time points:		
2004	12146	33.94
2006	11712	32.73
2009	11912	33.29

Code:

```
data completed;
set hlth_long;
if HEADACHE=. then delete; miss=CMISS(SORETHROAT, STOMACHACHE); run;
```

```
data percentage;
set completed;
if (WAVE=2004) and (miss=0) then percent_2004=1; else percent_2004=0;

if (WAVE=2006) and (miss=0) then percent_2006=1; else percent_2006=0;

if (WAVE=2009) and (miss=0) then percent_2009=1; else percent_2009=0;
run;

proc freq data=percentage;
tables percent_2004 percent_2006 percent_2009/nocum list; run;
```
