Biostatistics 203A: Introduction to Data Management and Statistical Computing
Lab Assignment 3: Submission Template
Fall 2023
Name: Sebastian Martinez

Exercise 1

```
proc sort data=lb.hlth_2009;
by IND_ID HH_ID;
run;
proc transpose data= lb.hlth_2009 out=overall_set 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;
set overall_set (rename=(COL1=INDICATOR));
SYMPTOM = compress( WAVE, '_', 'd');
WAVE = compress(WAVE,'_','a');
run;
proc print data =hlth_very_long (obs=27) noobs;
var IND_ID HH_ID WAVE SYMPTOM INDICATOR;
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
4	INDICATOR	Num	8	
1	IND_ID	Num	8	INDIVIDUAL ID
5	SYMPTOM	Char	16	
3	WAVE	Char	16	NAME OF FORMER VARIABLE

Exercise 2

```
proc sort data=lb.hlth_2009;
by IND_ID HH_ID HH_Type;
run;
proc transpose data=lb.hlth_2009 out=hlth_long_headache (rename=(COL1=HEADACHE))
name=EXTRACT;
by IND_ID HH_ID HH_TYPE;
var HEADACHE_2004 HEADACHE_2006 HEADACHE_2009;
run;
data hlth_long_headache;
retain IND_ID HH_ID WAVE HEADACHE;
length WAVE $4;
```

```
set hlth_long_headache;
WAVE = compress(EXTRACT,'_','a');
drop EXTRACT;
run;
proc transpose data=lb.hlth_2009 out=hlth_long_sorethroat (rename=(COL1=SORETHROAT))
name=EXTRACT;
by IND_ID HH_ID HH_TYPE;
var SORETHROAT_2004 SORETHROAT_2006 SORETHROAT_2009;
run;
data hlth_long_sorethroat;
retain IND_ID HH_ID WAVE HEADACHE;
length WAVE $4;
set hlth_long_sorethroat;
WAVE = compress(EXTRACT,'_','a');
drop EXTRACT;
run;
proc transpose data=lb.hlth_2009 out=hlth_long_stomachache (rename=(COL1=STOMACHACHE))
name=EXTRACT;
by IND_ID HH_ID HH_TYPE;
var STOMACHACHE_2004 STOMACHACHE_2006 STOMACHACHE_2009;
run;
data hlth_long_stomachache;
set hlth_long_stomachache;
length WAVE $4;WAVE = compress(EXTRACT,'_','a');
drop EXTRACT;
run;
data hlth_long;
merge hlth_long_headache hlth_long_sorethroat hlth_long_stomachache;
by IND_ID HH_ID HH_Type;
run;
proc print data=hlth_long (obs=15) noobs;
var IND_ID HH_ID WAVE HEADACHE SORETHROAT STOMACHACHE HH_TYPE;
run;
proc contents data=hlth_long;
Run;
```

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Label
4	HEADACHE	Num	8	
2	HH_ID	Num	8	HOUSEHOLD ID
5	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	
3	WAVE	Char	4	

Exercise 3-

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

	Percentage of Individuals Endorsing the Symptom		
	2004	**2006**	**2009**
Symptom: Headache			
Urban-Dwelling	6.15	6.26	5.42
Rural-Dwelling	4.63	4.00	3.94
All	10.78	10.26	9.36
Symptom: Sore Throat			
Urban-Dwelling	11.73	9.70	8.83
Rural-Dwelling	7.91	6.75	8.05
All	19.64	16.45	16.88
Symptom: Stomachache			
Urban-Dwelling	4.67	3.80	2.04
Rural-Dwelling	2.88	2.39	1.50
All	7.55	6.19	3.54

Exercise 4

```
data array_head;
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;
HEADACHE = headache_array{i};
WAVE = wv{i};
output;
```

```
end;
keep HEADACHE IND_ID HH_ID WAVE HH_TYPE WAVE;
run;
data array_sore;
set lb.hlth_2009;
array sore_array{3}
SORETHROAT_2004
SORETHROAT_2006
SORETHROAT_2009;
array wv{3}$ wv1-wv3 ('2004','2006','2009');
do i = 1 to 3;
SORETHROAT = sore_array{i};
WAVE = wv{i};
output;
Symptom: Sore Throat
Urban-Dwelling 11.73 9.70 8.83
Rural-Dwelling 7.91 6.75 8.05
All 19.64 16.45 16.88
Symptom: Stomachache
Urban-Dwelling 4.67 3.80 2.04
Rural-Dwelling 2.88 2.39 1.50
All 7.55 6.19 3.54end;
keep SORETHROAT IND_ID HH_ID WAVE HH_TYPE WAVE;
run;
data array_sto;
set lb.hlth_2009;
array sto_array{3}
STOMACHACHE_2004
STOMACHACHE_2006
STOMACHACHE_2009;
array wv{3}$ wv1-wv3 ('2004','2006','2009');
do i = 1 to 3;
STOMACHACHE = sto_array{i};
WAVE = wv{i};
output;
end;
keep STOMACHACHE IND_ID HH_ID WAVE HH_TYPE WAVE;
run;
data hlth_array_long;
merge array_head array_sore array_sto;
by IND_ID HH_ID HH_Type;
run;
proc print data=hlth_array_long (obs=15) noobs;
var IND_ID HH_ID WAVE HEADACHE SORETHROAT STOMACHACHE HH_TYPE;
run;
proc contents data=hlth_array_long;
Run;
```

Alphabetic List of Variables and Attributes				
#	Variable	Type	Len	Label
4	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	
5	WAVE	Char	8	

Exercise 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 print data=hlth_wide_headache (obs = 8) noobs;
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 print data=hlth_wide_sorethroat (obs = 8) noobs;
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;
proc print data=hlth_wide_stomachache (obs = 8) noobs;
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) noobs;
var IND_ID HH_ID HEADACHE_2004 HEADACHE_2006 HEADACHE_2009 SORETHROAT_2004
SORETHROAT_2006 SORETHROAT_2009 STOMACHACHE_2004 STOMACHACHE_2006
STOMACHACHE_2009 HH_TYPE;
run;
proc contents data=hlth_wide;
Run;
```

IND_ID	HH_ID	HEADACHE_2004	HEADACHE_2006	HEADACHE_2009	SORETHROAT_2004	SORETHROAT_2006	SORETHROAT_2009	STOMACHACHE_2004	STOMACHACHE_2006	STOMACHACHE_2009	HH_TYPE
211101003002	211101003	0	.	1	0	.	1	0	.	1	1
211101003101	211101003	.	0	.	.	0	.	.	0	.	1
211101003102	211101003	.	0	.	.	0	.	.	0	.	1
211101008001	211101008	0	0	.	0	0	.	0	0	.	1
211101008002	211101008	0	0	0	0	0	0	0	0	0	1
211101008003	211101008	0	0	.	0	0	.	0	0	.	1
211101008005	211101008	.	0	0	.	0	0	.	0	0	1
211101008021	211101008	0	0	0	0	0	0	0	0	0	1

Exercise 6
data question;

	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	51.5
2004, 2006, and 2009	6733	38.03
2006 and 2009 (but not 2004)	1532	8.65

```
set lb.hlth_2009;
MISS_COUNT_0406 = CMISS(HEADACHE_2004, HEADACHE_2006);
MISS_COUNT_040609 = CMISS(HEADACHE_2004, HEADACHE_2006,HEADACHE_2009);
MISS_COUNT_0609no04 = CMISS(HEADACHE_2006, HEADACHE_2009);
if MISS_COUNT_0609no04 = 0 & (HEADACHE_2004 = .) then WV06_AND_09NOT04 = 1;
else WV06_AND_09NOT04 = 0;
run;
proc freq data=question;
tables MISS_COUNT_0406 MISS_COUNT_040609 WV06_AND_09NOT04;
run;
```

Exercise 7
Proc statement where headache = 0 or missing
Freq wave
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 51.5
2004, 2006, and 2009 6733 38.03
2006 and 2009 (but not 2004) 1532 8.65
N %
Individuals with surveys completed at the
following time points:
2004 12147 33.95
2006 11714 32.74data hlth_long_filtered;
set hlth_long;
where not missing(HEADACHE);
run;
proc freq data=hlth_long_filtered;
tables WAVE / nocum;
run;

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