

### Homework 5: Manipulating Datasets and Creating Variables

This homework assignment continues our replication of **Bansil et al., 2011** using the 20013-2014 data from NHANES. You may discuss the assignment with others but you **MUST** run your own code and write up the answers in your own words. Please **turn in a printed copy** of your write-up and SAS commands (attached as a printout to the back of your write-up) in class at **9:30AM on Friday, November 3<sup>rd</sup>** and **submit an electronic copy** of your write-up and code via the CANVAS Assignment tab. ONLY output that is **relevant to the questions** should be included.

- Be sure that:
- 1) Your SAS code runs from start to finish,
  - 2) Your results make sense (check your sample sizes and look for unreasonable, unlikely, or impossible answers),
  - 3) Your homework write-up has your name on EACH page. Submit your write-up on Canvas as an attachment to Assignment 5. Your file should be turned in to Canvas as a .doc, .docx, or .pdf in the following format:  
LASTNAME\_FIRSTNAME\_HW5.docx
  - 4) Your code is well commented. Commenting is done by `*type in any text here;` or `/* type in text here */`. Be sure to include the homework number and your name at the top of your homework code, identify each question in the code, and describe each new task. Also, format your code (indentation and carriage returns) to improve readability. Your code file should be turned in to Canvas as a .sas file in the following format: LASTNAME\_FIRSTNAME\_HW5.sas
  - 5) 5% will be deducted if either of tasks 3 or 4 above is not completed.
  - 6) All files needed for the homework can be found in the Files section of Canvas in the folder Homework 5 (including the homework questions, datasets, and formats).

**NOTE:** The Codebooks created in your past assignments may help you with this assignment.

---

Create a new folder on your computer or flash drive specifically for Homework 5. Make a permanent library called **nhanes** in SAS that points to the folder you just created. Download the included **sleepbp** dataset from Canvas and place them in the folder on your computer that corresponds to your nhanes library. Make sure that the nhanes library contains all of the downloaded **sleepbp** dataset. Include the code to create your library with the code submitted with the homework.

#### Table 1 Re-creation

We will continue working with the **sleepbp** dataset, which includes all merged data, recodes, and created variables from Homework 4. We will now finish creating all necessary variables to make a

descriptive Table 1. All new variables should be created in the same DATA step and saved to a dataset called **sleepbp\_final** in your **nhanes** library.

1. Create an average systolic (**sysbp\_ave**) and diastolic (**diabp\_ave**) blood pressure from the four NHANES measurements. Do not give respondents missing values unless all four of their measurements are missing.
2. Create an indicator variable (**bpmed**) for whether or not the respondent is currently taking a blood pressure medication. Be sure that those who have never been told they have hypertension or those that have never been told to take medication are coded as 0, not missing.
3. From **sysbp\_ave**, **diabp\_ave**, and **bpmed**, create a hypertension indicator (**htn**) as defined in the Bansil et al. paper.
4. Create the variables **smoker**, **diab**, **race**, **educ**, and **pir** corresponding to current smoking status, diabetes status, race/ethnicity, education, and poverty to income ratio, respectively, that match the categories in the table below. These should be coded as numeric variables. For **smoker**, be sure that never and former smokers are coded to 0. For **diab**, borderline/pre-diabetes should be coded to 0.
5. Create and assign formats for all the categorical variables in the table and save them to a permanent library. Indicator variables coded the same can all share a single yes/no format.
6. Check your recodes from problem 4 by producing cross-tabulations or PROC MEANS with CLASS statements.
7. Note that age and BMI are left as continuous for our table, but were categorized by Bansil et al. For what reason(s) might you want to categorize these variables? For what reason(s) might you want them to be continuous?

Categorizing them into distinct levels would allow us to perform calculations/analysis between the different groups (like linear regression) and allow for distinctions to be made between the levels (stratification), but having them be continuous is more specific and accurate than lumping people into groups, and you can use them for making histograms.

8. Complete Table 1 below.

**PLEASE NOTE: Your numbers will NOT match those in the Table 1 of the paper because you are using different data. Please only keep two digits after the decimal point.**

<b>Table 1:</b> Distribution of Demographic and Health Characteristics Among US Adults by Hypertension Status, National Health and Nutrition Examination Survey, 2013-2014			
	Total N (%)	Normotensive N (%)	Hypertensive N (%)
Mean systolic BP (SD)	122.92 (17.92)	115.43 (10.87)	137.05 (19.96)
Mean diastolic BP (SD)	69.34 (12.47)	68.02 (10.14)	71.84 (15.66)
Female gender	2941 (51.52%)	1873 (32.81%)	1068 (18.71%)
Mean age (SD)	47.98 (18.43)	40.38 (16.23)	61.54 (13.74)
Race			
Non-Hispanic white	2417 (42.34%)	1515 (26.54%)	902 (15.80%)
Non-Hispanic black	1177 (20.62%)	625 (10.95%)	552 (9.67%)
Mexican American	787 (13.79%)	566 (9.92%)	221 (3.87%)
Other	1327 (23.25%)	953 (16.70%)	374 (6.55%)
Has health insurance	4515 (79.20%)	2717 (47.66%)	1798 (31.54%)
Education			
Less than high school	1147 (21.32%)	631 (11.73%)	516 (9.59%)
High school / GED	1215 (22.58%)	702 (13.05%)	513 (9.54%)
At least some college	3018 (56.10%)	2003 (37.23%)	1015 (18.87%)
Poverty level <sup>a</sup>			
<1.00	1221 (23.17%)	821 (15.58%)	400 (7.59%)
1.00 – 3.00	2085 (39.57%)	1260 (23.91%)	825 (15.66%)
>3.00	1963 (37.26%)	1298 (24.63%)	665 (12.62%)
Mean body mass index (SD)	28.89 (7.09)	27.79 (6.53)	30.91 (7.62)
Has diabetes	704 (12.34%)	199 (3.49%)	505 (8.85%)
Current smoker	1139 (47.26%)	767 (31.83%)	372 (15.44%)
Has sleep disorder	544 (9.55%)	256 (4.49%)	288 (5.06%)
Short sleep duration (<7 hours)	2148 (37.69%)	1361 (23.88%)	787 (13.81%)
Combination of sleep problems <sup>b</sup>			
Sleep disorder and short sleep	281 (4.94%)	136 (2.39%)	145 (2.55%)
Sleep disorder only	261 (4.59%)	120 (2.11%)	141 (2.48%)
Short sleep only	1862 (32.74%)	1222 (21.49%)	640 (11.25%)
None of the above	3283 (57.73%)	2173 (38.21%)	1110 (19.52%)

Abbreviations: BP, blood pressure; SD, standard deviation. <sup>a</sup>The ratio of a family's income to the federally defined poverty threshold for a family of the same size in the same calendar year. <sup>b</sup>Mutually exclusive sleep problem categories.

CODE:

\*Homework 5: Manipulating Datasets & Creating Variables | Stephanie Mecham | EPID 640  
Section 2;

\*Creating a permanent library;

```
libname nhanes "C:\Users\smecham\Desktop\ Nhanes";  
run;
```

\*Question 1-4: Creating and coding new variables;

```
data nhanes.sleepbp_final;  
set nhanes.sleepbp;
```

```
*Creating sysbp_ave;  
sysbp_ave= mean(BPXSY1, BPXSY2, BPXSY3, BPXSY4);
```

```
*Creating diabp_ave;  
diabp_ave= mean(BPXDI1, BPXDI2, BPXDI3, BPXDI4);
```

```
*Creating bpmed;  
if BPQ020=2 or BPQ040A=2 or BPQ050A=2 then bpmed= 0;  
else if BPQ050A=1 then bpmed=1;  
else if BPQ050A=. or BPQ050A=7 or BPQ050A=9 then bpmed=.;
```

```
*Creating htn;  
if sysbp_ave=. and diabp_ave=. and bpmed=. then htn=.;  
else if sysbp_ave >= 140 or diabp_ave>= 90 or bpmed=1 then htn=1;  
else htn=0;
```

```
*Creating smoker;  
if SMQ040=. or SMQ040=7 or SMQ040=9 then smoker=.;  
else if SMQ020=2 then smoker=0;  
else if SMQ020= 1 and SMQ040=3 then smoker= 0;  
else smoker=1;
```

```
*Creating diab;
if DIQ010=. or DIQ010=7 or DIQ010=9 then diab=.;
else if DIQ010= 1 then diab=1;
else if DIQ010=2 or DIQ010=3 then diab=0;

*Creating race;
if RIDRETH1=1 then race=3;
else if RIDRETH1=3 then race=1;
else if RIDRETH1=4 then race=2;
else race=4;

*Creating educ;
if DMDEDUC2=. or DMDEDUC2=7 or DMDEDUC2=9 then educ=.;
if DMDEDUC2=1 or DMDEDUC2=2 then educ=1;
else if DMDEDUC2=3 then educ=2;
else if DMDEDUC2=4 or DMDEDUC2=5 then educ=3;

*Creating PIR;
if INDFMPIR =. then pir=.;
else if INDFMPIR < 1 then pir=1;
else if INDFMPIR >=1 AND INDFMPIR <=3 then pir=2;
else if INDFMPIR >3 then pir=3;

*Re-coding health insurance (optional step);
if HIQ011 in (., 7, 9) then healthinsurance_new=.;
else if HIQ011 = 2 then healthinsurance_new=0;
else if HIQ011 = 1 then healthinsurance_new=1;

*Re-coding sleep disorder (optional step);
if SLQ060 in (., 7, 9) then sleepdisorder_new=.;
else if SLQ060 = 2 then sleepdisorder_new=0;
else if SLQ060= 1 then sleepdisorder_new=1;
run;

*Question 5: Assigning formats;

libname nhanes 'C:\Users\smecham\Desktop\nhanes';
options fmtsearch = (nhanes);

proc format library=nhanes;
value yn
```

```
. = 'Missing'  
0 = 'No'  
1 = 'Yes';
```

```
value sex  
1 = 'Male'  
2 = 'Female'  
. = 'Missing';
```

```
value race  
1 = 'Non-Hispanic White'  
2 = 'Non-Hispanic Black'  
3 = 'Mexican American'  
4 = 'Other'  
. = 'Missing';
```

```
value educ  
1 = 'Less than high school'  
2 = 'High school/GED'  
3 = 'At least some college'  
. = 'Missing';
```

```
value sleepcombo  
1 = Short sleep and sleep disorder  
2 = No short sleep and sleep disorder  
3 = Short sleep and no sleep disorder  
4 = No short sleep and no sleep disorder  
. = Missing;  
run;
```

```
data nhanes.Sleepbp_final;  
    set nhanes.sleepbp_final;  
    format bpped yn. htn yn. smoker yn. sleepdisorder_new yn. diab yn. healthinsurance_new  
    yn. sleepcombo sleepcombo. RIAGENDR sex. race race. educ educ.;  
run;
```

```
*Question 6: Checking re-codes;  
*Proc Freq step;  
proc freq data=nhanes.Sleepbp_final;  
table smoker*SMQ020 diab*DIQ010 race*RIDRETH1 educ*DMDEDUC2;  
run;
```

```
*Proc Means with CLASS statement step;  
proc means data=nhanes.Sleepbp_final N NMISS MIN MAX;  
var INDFMPIR;  
class PIR;  
run;
```

\*Question 8: Completing the table;

```
proc means data=nhanes.Sleepbp_final;  
var sysbp_ave diabp_ave RIDAGEYR BMXBMI;  
run;
```

```
proc means data=nhanes.Sleepbp_final;  
var sysbp_ave diabp_ave RIDAGEYR BMXBMI;  
class htn;  
run;
```

```
proc freq data=nhanes.Sleepbp_final;  
tables RIAGENDR*htn race*htn healthinsurance_new*htn educ*htn smoker*htn pir*htn diab*htn  
sleepdisorder_new*htn shortsleep*htn sleepcombo*htn;  
run;
```