# /\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# THIS PROGRAM RUNS CODE TO COMPLETE

# SESSION 14, PH490KR, SPRING 2018

# 

# KATHERINE REEVES

# ORIGINAL: MARCH 22, 2016

# MODIFIED: MARCH 20, 2018

# \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

\*3.use a libname statement to tell SAS where your dataset is stored;

libname ph490kr "Z:\Katherine's Docs\Courses\Pubhlth 490kr\Datasets";

\*4.basic data step to read in the dataset;

data sess14; /\*this will put the dataset into "work", if I want to save it I can write data ph490kr.sess16\*/

set ph490kr.VitDSess14;

run;

\*5. use proc contents to view info about the dataset;

proc contents data=sess14;

run;

\*6. use proc print;

\*a. print ev\_smk and cur\_smk;

proc print data=sess14;

var ev\_smk cur\_smk;

run;

\*b. as above, but only first 100 observations;

proc print data=sess14 (obs=100);

var ev\_smk cur\_smk;

run;

\*c. brand for current smokers;

proc print data=sess14;

where cur\_smk=1;

var brand;

run;

\*d. cig\_day for those starting smoking after age 15;

proc print data=sess14;

where age\_cig>15 & age\_cig^=99;

var cig\_day;

run;

\*7. sort dataset by cur\_smk;

proc sort data=sess14;

by cur\_smk;

run;

\*8. print cig\_day by current smoking status;

\*note: you need to have the data sorted before you can use "by";

proc print data=sess14;

by cur\_smk;

var cig\_day;

run;

\*9. save a copy of the dataset including only current smokers;

\*code below will create a "temporary" dataset stored in the library "work";

data cursmk;

set sess14;

where cur\_smk=1;

run;

\*to make a permament dataset, we just specify the name of a permanent SAS library;

data ph490kr.cursmk;

set sess14;

where cur\_smk=1;

Run;

# /\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# THIS PROGRAM RUNS CODE TO COMPLETE

# SESSION 15, PH490KR, SPRING 2018

# 

# KATHERINE REEVES

# ORIGINAL: MARCH 22, 2016

# MODIFIED: MARCH 20, 2018

# \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

\*1.use a libname statement to access the dataset;

libname ph490kr "Z:\Katherine's Docs\Courses\Pubhlth 490kr\Datasets";

/\*3. assign labels to each variable\*/

\*first, I will read the data into a temporary dataset, then I'll start adding labels within the same data step;

data sess15;

set ph490kr.VitDsess14;

label ev\_smk="ev\_smk:ever smoked 20+ packs cigarettes"

cur\_smk="cur\_smk:currently smoke"

brand="brand:cigarette brand"

cig\_day="cig\_day:cigarettes per day smoked"

age\_cig="age\_cig:age started smoking"

age\_qt="age\_qt:age quit smoking";

run;

/\*4. write formats for the categorical variables and assign them. This takes two separate steps:

1--proc format to define the labels

2--assign the format within a data step

\*/

proc format library=ph490kr;

value ynf 0="0\_No" 1="1\_Yes";

value cigdayf 1="1\_1-10" 2="2\_11-20" 3="3\_21-30" 4="4\_31-40" 5="5\_41+" 99="99\_missing";

run;

options fmtsearch=(ph490kr);

data sess15;

set sess15; /\*note that I am overwriting my existing temporary dataset\*/

format ev\_smk ynf. cur\_smk ynf. cig\_day cigdayf.;

run;

/\*5. save as a permanent dataset and describe contents\*/

data ph490kr.VitDSess14\_kwr;

set sess15;

run;

proc contents data=ph490kr.VitDSess14\_kwr;

run;

/\*\*\*HERE IS HOW WE COULD INCORPORATE EVERYTHING WE'VE DONE INTO A SINGLE PROC FORMAT AND DATA STEP\*\*\*/

proc format library=ph490kr;

value ynf= 0 "0\_No" 1="1\_Yes";

value cigdayf 1="1\_1-10" 2="2\_11-20" 3="3\_21-30" 4="4\_31-40" 5="5\_41+" 99="99\_missing";

value smokef 0="0\_never smoker" 1="1\_past smoker" 2="2\_current smoker";

run;

options fmtsearch=(ph490kr);

data ph490kr.VitDSess14\_kwr;/\*note that this will save a permanent SAS dataset in my folder\*/

set PH490KR.VitDSess14;

label ev\_smk="ev\_smk:ever smoked 20+ packs cigarettes"

cur\_smk="cur\_smk:currently smoke"

brand="brand:cigarette brand"

cig\_day="cig\_day:cigarettes per day smoked"

age\_cig="age\_cig:age started smoking"

age\_qt="age\_qt:age quit smoking";

format ev\_smk ynf. cur\_smk ynf. cig\_day cigdayf.;

run;

proc contents data=ph490kr.VitDSess14\_kwr;

run;

# /\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# THIS PROGRAM COMPLETES ACTIVITY FOR

# SESSION 16, PH490KR SPRING 2018

# 

# KATHERINE REEVES

# ORIGINAL: MARCH 20, 2018

# MODIFIED: MARCH 23, 2018

# \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

libname ph490kr "Z:\Katherine's Docs\Courses\Pubhlth 490kr\Datasets";

data sess16; /\*read in data into a temporary dataset that is stored in work and called sess16\*/

set ph490kr.vitdsess16; /\*this is the dataset being read in\*/

\*1. create a new variable coding as never, past, or current smokers;

\*first, we will look at ev\_smk & cur\_smk in a table, see \*A\*. below;

if cur\_smk=1 then smoke=2; /\*current smokers\*/

else if ev\_smk=1 & cur\_smk=0 then smoke=1; /\*past smokers\*/

else if ev\_smk=0 then smoke=0;

\*now check our work, see \*B\* below;

\*2. dichtomize weight at the median, first need to find the median, see \*C\* below;

if wt\_lbs<=136 then wt\_med=1;

else if wt\_lbs>136 then wt\_med=2;

if wt\_lbs=. then wt\_med=.;

\*check our work, see \*D\* below;

\*3. create a new variable calculating BMI;

bmi=wt\_lbs/(ht\_in\*ht\_in)\*703;

\*check our work, see \*E\* below;

\*4. group bmi into who categories;

if (bmi<18.5) then bmicat=1;

else if (bmi>=18.5) & (bmi<25) then bmicat=2;

else if (bmi>=25) & (bmi<30) then bmicat=3;

else if (bmi>=30) then bmicat=4;

else if bmi=. then bmicat=.;

\*check our work, see \*F\* below;

\*5. convert height to meters;

ht\_m=ht\_in\*.0254;

\*check our work, see \*G\* below;

run;

\*6. save as a permanent SAS dataset;

data ph490kr.vitdsess16\_kr;

set sess16;

run;

\*A\*. look at cross-tabulation of ev\_smk and cur\_smk;

proc freq data=sess16;

table ev\_smk\*cur\_smk /missing;

run;

\*B\* check our work for creating smoke;

proc freq data=sess16;

table smoke;

run;

proc print data=sess16 (obs=100);

var ev\_smk cur\_smk smoke;

run;

\*C\* find the median value of weight;

proc univariate data=sess16;

var wt\_lbs;

run;

/\*median wt\_lbs=136.00\*/

\*D\* check our work in dichotomizing weight;

proc print data=sess16 (obs=50);

var wt\_lbs wt\_med;

run;

proc sort data=sess16;

by wt\_med;

run;

proc univariate data=sess16;

by wt\_med;

var wt\_lbs;

run;

\*E\* check work in creating bmi;

proc print data=sess16 (obs=50);

var wt\_lbs ht\_in bmi;

run;

proc univariate data=sess16;

var bmi;

run;

\*F\* check work in categorizing BMI;

proc print data=sess16 (obs=75);

var bmi bmicat;

run;

\*G\* check work in converting height to meters;

proc print data=sess16 (obs=25);

var ht\_m ht\_in;

run;

\*\*\*EXTRA: LABEL VARIABLES AND THEIR VALUES\*\*\*;

proc format library=ph490kr;

value smokef 0="0\_never" 1="1\_past" 2="2\_current";

value wt\_medf 1="1:weight <=136 lbs" 2="2:weight >136 lbs";

value bmicatf 1="1\_underweight" 2="2\_normal weight" 3="3\_overweight" 4="4\_obese";

run;

options fmtsearch=(ph490kr);

data ph490kr.vitdsess16\_kr; /\*note, this code will overwrite my permanent SAS dataset\*/

set ph490kr.vitdsess16\_kr;

format smoke smokef. wt\_med wt\_medf. bmicat bmicatf.;

label smoke="smoke: smoking status";

label wt\_med="wt\_med: weight relative to median";

label bmi="bmi: BMI, kg/m2";

label bmicat="bmicat: BMI category";

label ht\_m="ht\_m: height in meters";

run;

proc contents data=ph490kr.vitdsess16\_kr;

run;

# /\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# THIS PROGRAM PERFORMS ACTIVITIES FOR

# SESSION 17, PH490KR SPRING 2018

# 

# KATHERINE REEVES

# ORIGINAL: MARCH 23, 2018

# MODIFIED: MARCH 26, 2018

# \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

\*START BY ASSIGNING LIBNAME TO FOLDER WHERE DATASET IS STORED;

libname ph490kr "Z:\Katherine's Docs\Courses\Pubhlth 490kr\Datasets";

\*read in dataset to a temporary dataset;

data sess17;

set ph490kr.VitDsess17;

run;

\*1a. check age\_cig;

proc means data=sess17;

var age\_cig;

run;

proc freq data=sess17;

table age\_cig\*ev\_smk;

run;

\*fix by setting 99 to missing;

data sess17;

set sess17;/\*note: this code will overwrite our temporary dataset\*/

age\_cigkr=age\_cig;

if age\_cig=99 then age\_cigkr=.;

if age\_cig=0 then age\_cigkr=.;

run;

\*check our work;

proc means data=sess17;

var age\_cigkr;

run;

proc print data=sess17 (obs=50);

var age\_cig age\_cigkr;

run;

\*1.b. check age\_qt;

proc means data=sess17;

var age\_qt;

run;

proc freq data=sess17;

table age\_qt\*cur\_smk;

run;

\*set values of 99 to missing;

data sess17;

set sess17;

age\_qtkr=age\_qt;

if age\_qt=99 then age\_qtkr=.;

run;

\*check our work;

proc means data=sess17;

var age\_qtkr;

run;

proc print data=sess17 (obs=50);

var age\_qt age\_qtkr;

run;

\*1c.check cig\_day;

proc freq data=sess17;

table cig\_day\*cur\_smk/missing;

run;

\*set 99 to "0" for those who answer "no" to current smoking;

data sess17;

set sess17;

cig\_daykr=cig\_day;

if cig\_day = 99 and (cur\_smk eq 0) then cig\_daykr=0;

if cig\_day = 99 and ((cur\_smk eq .) or (cur\_smk eq 1)) then cig\_daykr=.;

run;

\*check our work;

proc freq data=sess17;

table cig\_day\*cig\_daykr cig\_daykr\*cur\_smk cig\_daykr/missing;

run;

proc freq data=ph490kr.vitdsmk;

table cig\_day;

run;

\*1d. check yrs\_oc;

proc means data=sess17;

var yrs\_oc;

run;

proc freq data=sess17;

table yrs\_oc\*ever\_oc/missing;

run;

\*need to set 99 to "0" for those who answer "no" to ever OC use;

/\*also, many said no to ever\_oc but then provided a number of years of OC use--how to handle?

1. could set to "0," essentially applying skip pattern (if they had skipped the question as

they were supposed to, they would have been set to 0)

2. set to missing, since answers are inconsistent

3. keep answers and set ever\_oc to yes since they are indicating having used OCs

Any of these options are ok, as long as you are consistent, document, and have a good rationale.

Let's go with option #1\*/

data sess17;

set sess17;

yrs\_ockr=yrs\_oc;

if yrs\_oc=99 & ever\_oc=0 then yrs\_ockr=0;

if yrs\_oc=99 & ever\_oc ^=0 then yrs\_ockr=.;

run;

\*check our work;

proc print data=sess17;

var yrs\_ockr;

where ever\_oc=0;

run;

proc print data=sess17;

var yrs\_ockr ever\_oc;

where ever\_oc ^=0;

run;

\*1.e. check curr\_oc;

proc freq data=sess17;

table curr\_oc\*ever\_oc/missing;

run;

/\*have 14 who answered yes to current but no to ever;

we will set to no to be consistent with our decision of how to handle yrs\_oc above

also have values of 99 that need to be set to no

and some values of 99 when ever\_oc=1 or missing that need to be set to missing\*/

data sess17;

set sess17;

curr\_ockr=curr\_oc;

if curr\_oc=1 & ever\_oc=0 then curr\_ockr=0;

if curr\_oc=99 & ever\_oc=0 then curr\_ockr=0;

if curr\_oc=99 & ever\_oc^=0 then curr\_ockr=.;

run;

\*check our work;

proc freq data=sess17;

table curr\_ockr\*ever\_oc/missing;

run;

\*2. label each variable and its values;

proc format library=ph490kr;

value ynf 0 ="0\_no" 1= "1\_yes";

value cig\_dayf 1= "1\_1-10" 2= "2\_11-20" 3= "3\_21-30" 4= "4\_31-40" 5= "5\_41+";

run;

options fmtsearch=(ph490kr);

data sess17;

set sess17;

format ev\_smk cur\_smk ever\_oc curr\_ockr ynf.;

label ev\_smk="ev\_smk: ever smoked 20+ packs cig";

label cur\_smk="cur\_smk:currently smoke cig";

label cig\_daykr="cig\_daykr:cigarettes per day";

label age\_cigkr="age\_cigkr:age started smoking";

label age\_qtkr="age\_qtkr:age quit smoking";

label ever\_oc="ever\_oc:ever used OCs";

label yrs\_ockr="yrs\_ockr:total years of OCuse";

label curr\_ockr="curr\_ockr:current OC use";

run;

proc contents data=sess17;

run;

\*let's clean up our dataset by dropping the variables we recoded;

drop cig\_day age\_cig age\_qt yrs\_oc curr\_oc;

run;

\*3. save a permanent copyl;

data ph490kr.VitDsess17\_working;

set sess17;

run;

proc contents data=ph490kr.VitDsess17\_working;

run;

# /\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# THIS PROGRAM RUNS ACTIVITIES FOR

# SESSION 18, PH490KR SPRING 2018

# 

# KATHERINE REEVES

# ORIGINAL: MARCH 29, 2018

# MODIFIED: MARCH 29, 2018

# \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

\*tell SAS where to look for the data by defining a library;

libname ph490kr "Z:\Katherine's Docs\Courses\Pubhlth 490kr\Datasets";

\*1. label all variables and their values;

\*first, write the formats;

proc format library=ph490kr;

value ynf 0="0:no" 1="1:yes";

value tanf 1="1:>1/week" 2="2:weekly" 3="3:bi-weekly" 4="4:monthly" 5="5:bi-monthly" 6="6:only once or twice";

\*additional formats for new variables;

value tan2f 5="5:bi-monthly or more" 6="6:only once or twice" 7="7:never tans";

value wstqf 1="1:Q1 23-<28in" 2="2:Q2 28-<30in" 3="3:Q3 30-<33in" 4="4:Q4 33-45 in";

value vitdf 1="1:deficient, <30" 2="2:insuffficient, 30-<50" 3="3:sufficient, 50-125" 4="4:toxic, >125";

run;

\*tell SAS where to look for the formats;

options fmtsearch=(ph490kr);

\*start a datastep;

data sess18;

set ph490kr.VitDsess18;

\*label each variable;

label id="id:participant id";

label alc\_ev="alc\_ev:ever had 12+ drinks alcohol";

label alc\_num="alc\_num: num alcohol drinks per day";

label tan="tan: used tanning bed/booth in past 6 months";

label tan\_freq="tan\_freq:frequency of tanning use";

label wst\_in="wst\_in: waist circumference, inches";

label OHD="OHD: 25OH vitamin d level";

\*assign formats;

format alc\_ev tan ynf.;

format tan\_freq tanf.;

\*2. fix alc\_num, first need to run some procs, see \*A\* below;

\*there are 19 who answers no to alc\_ev and were appropriately skipped from alc\_num--> can infer that they had 0 drinks

\*there are 4 who answered no to alc\_ev and were missing on alc\_num-->can infer that they had 0 drinks

\*there are 62 who were missing on both questions, and should be missing on alc\_num-->recode to .

\*there is 1 who answered yes on alc\_ev and missing on alc\_num, should be missing on alc\_num-->recode to .

\*1 said no to alc\_ev and then 1 to alc\_num--> set to 0 to be consistent with applying skip pattern

\*there is 1 who was missing on alc\_ev and then said 8 for alc\_num-->set to missing to be consistent

\*also, one person said "61 drinks," which is probably inaccurate (she would be dead!)-->recode to .;

alc\_numkr=alc\_num;

if alc\_num=99 & alc\_ev=0 then alc\_numkr=0 ;

if alc\_num=. & alc\_ev=0 then alc\_numkr=0 ;

if alc\_ev=99 then alc\_numkr=. ;

if alc\_num=99 & alc\_ev=99 then alc\_numkr=. ;

if alc\_num=99 & alc\_ev=1 then alc\_numkr=. ;

if alc\_num=1 & alc\_ev=0 then alc\_numkr=0 ;

if alc\_numkr=61 then alc\_numkr=. ;

\*check our work, see \*B\* below;

label alc\_numkr="alc\_numkr:number of alcoholic drinks/day, recoded";

\*3. recode tan\_freq;

tan\_freqkr=tan\_freq;

if tan\_freq<=5 then tan\_freqkr=5;

if tan\_freq=9 & tan^=0 then tan\_freqkr=. ;

if tan\_freq=9 & tan=0 then tan\_freqkr=7 ;

\*go back and update formats in proc format above, check work in \*C\* below;

label tan\_freqkr="tanfreqkr:frequency of tanning, recoded";

format tan\_freqkr tan2f.;

\*4. group wst\_in into quartiles--first check it and find cutpoints, see \*D\* below;

\*cutpoints: 28,30,33;

if wst\_in<28 then wstquart=1;

else if wst\_in>=28 & wst\_in<30 then wstquart=2;

else if wst\_in>=30 & wst\_in<33 then wstquart=3;

else if wst\_in>=33 then wstquart=4;

if wst\_in=. then wstquart=.;

label wstquart="wstquart: waist circumference, quartiles";

format wstquart wstqf.;

\*check our work, see \*E\* below;

\*5. check OHD for errors, group into deficient/insufficient/sufficient/toxic, see \*F\* below;

OHDkr=OHD;

if OHD eq 999 then OHDkr=.;

if OHDkr <30 & OHDkr ne . then OHDcat=1;

else if OHDkr >=30 & OHDkr<50 then OHDcat=2;

else if OHDkr >=50 & OHDkr <=125 then OHDcat=3;

else if OHDkr >125 then OHDcat=4;

label OHDcat="OHDcat: vitamin D sufficiency status";

format OHDcat vitdf.;

\*check work, see \*G\* below;

run;

\*6. save a permanent version of modifed dataset;

data ph490kr.vitdsess18\_working;

set sess18;

run;

\*A\* check alc\_num for errors;

proc means data=sess18;

var alc\_num;

run;

proc freq data=sess18;

table alc\_num\*alc\_ev/missing;

run;

\*B\* check our work in recoding alc\_num;

proc freq data=sess18;

table alc\_num \*alc\_numkr/missing;

run;

proc print data=sess18 (obs=100);

var alc\_num alc\_ev alc\_numkr;

run;

\*C\* check our work for recoding tan\_freq;

proc freq data=sess18;

table tan\_freqkr\*tan\_freq/missing;

run;

proc print data=sess18;

var tan tan\_freq tan\_freqkr;

run;

\*D\* inspect waist circumference;

proc univariate data=sess18;

var wst\_in;

run;

\*E\* check work in grouping wst\_in into quartiles;

proc print data=sess18 (obs=50);

var wst\_in wstquart;

run;

proc sort data=sess18;

by wstquart:

run;

proc means data=sess18;

var wst\_in;

class wstquart;

run;

\*F\* look at OHD;

proc univariate data=sess18;

var OHD;

run;

\*G\* check work in categorizing vitamin D;

proc sort data=sess18;

by OHDcat;

run;

proc means data=sess18;

class OHDcat;

var OHDkr;

run;

proc freq data=sess18;

table OHDcat/missing;

run;

# /\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# THIS PROGRAM COMPLETES ACTIVITIES FOR

# SESSION 19, PH490KR, SPRING 2018

# 

# KATHERINE REEVES

# ORIGINAL: March 29, 2016

# MODIFIED: April 6, 2018

# \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

\*1. open dataset;

libname ph490kr "Z:\Katherine's Docs\Courses\Pubhlth 490kr\Datasets";

libname fmt "Z:\Katherine's Docs\Courses\Pubhlth 490kr\Datasets\sess19fmt";

options fmtsearch=(fmt);

data vitd;

set ph490kr.VitDsess19;

run;

proc contents data=vitd;

run;

\*2. calculate mean, etc for bmi;

proc univariate data=vitd;

var bmi;

run;

\*2.A.repeat by smoking status;

proc univariate data=vitd;

class smoke01;

var bmi;

run;

\*or can use proc means;

proc means data=vitd mean std p25 p50 p75;

var bmi;

run;

proc means data=vitd mean std p25 p50 p75;

class smoke01;

var bmi;

run;

\*2.B. perform t test;

proc ttest data=vitd;

class smoke01;

var bmi;

run;

\*3. tabulate numbers of bmi categories;

proc freq data=vitd;

table bmi\_cat;

run;

\*3.A.B.distribution by smoking status with chi square test;

proc freq data=vitd;

table bmi\_cat\*smoke01/chisq;

run;

\*4. calculate summary statistics for weight;

proc means data=vitd mean std p25 p50 p75;

var wt\_lbs;

run;

\*4.A. stratification on nsaid use;

proc means data=vitd mean std p25 p50 p75;

class nsaid;

var wt\_lbs;

run;

\*4.B. perform t test;

proc ttest data=vitd;

class nsaid;

var wt\_lbs;

run;

\*5. tabulate nsaid status;

proc freq data=vitd;

tables nsaid;

run;

\*5.A.B. cross-tab with bmi category;

proc freq data=vitd;

tables bmi\_cat\*nsaid/chisq;

run;

\*6. calculate mean BMI among highest tertile of height;

proc means data=vitd;

where ht\_tert eq 3;

var bmi;

run;

\*8. among normal weight, t test sitting by smoking status;

proc ttest data=vitd;

where bmi\_cat eq 2;

class smoke01;

var tot\_sit;

run;

# /\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# Produces answer key for

# SESSION 20, PH490KR, SPRING 2018

# 

# katherine reeves

# original: april 11, 2016

# modified: april 10, 2018

# \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

libname ph490kr "Z:\Katherine's Docs\Courses\Pubhlth 490kr\Datasets";

libname fmt "Z:\Katherine's Docs\Courses\Pubhlth 490kr\Datasets\sess20fmt";

options fmtsearch=(fmt);

data temp;

set ph490kr.vitdsess20;

run;

proc contents;run;

\*1. RR for alcohol and PMS;

/\*note: first we need to sort the data so that we are calculating the appropriate relative risk\*/

proc sort data=temp;

by descending alc01 descending pmscc;

run;

proc freq data=temp order=data;

tables alc01\*pmscc/measures;

run;

/\*RR=3.8\*/

\*2. OR for smoking and PMS;

proc sort data=temp;

by descending smoke01 descending pmscc;

run;

proc freq data=temp order=data;

tables smoke01\*pmscc/ measures;

run;

/\*OR=5.35\*/

\*3. OR for smoke and PMS adjusted for sitting;

proc sort data=temp;

by descending smoke01 descending sitq descending pmscc;

run;

proc freq data=temp order=data;

\*syntax is adjustment\*exposure\*outcome, also use option "measures";

tables sitq\*smoke01\*pmscc/cmh measures;

run;

/\*stratum-specific ORs--T3: 3.88, T2: 12.67, T1: 4.67; adjusted 5.87

from above, unadjusted OR=5.35, so no evidence that sitting is a confounder\*/

/\*let's see how to do this in logistic regression, which allows us to easily control for

multiple variables and get adjusted odds ratios\*/

\*unadjusted;

proc logistic data=temp order=data;

model pmscc(event='1\_case') =smoke01;

run;

\*adjusted for for sitting;

proc logistic data=temp order=data;

model pmscc (event='1\_case')=smoke01 sitq;

run;

\*4. OR for weight and PMS;

proc sort data=temp;

by descending wt\_med descending pmscc;

run;

proc freq data=temp order=data;

tables wt\_med\*pmscc/measures;

run;

/\*OR=0.94\*/

\*5. OR for weight and PMS adjusted for smoking;

proc sort data=temp;

by descending wt\_med descending smoke01 descending pmscc;

run;

proc freq data=temp order=data;

tables smoke01\*wt\_med\*pmscc/cmh measures;

run;

/\*unadjusted OR=0.94, adjusted OR=1.07, CIs overlap--no evidence that smoking is a confounder\*/

\*6. OR for tertile of sitting and PMS;

/\*proc freq doesn't give ORs for exposure with >2 levels,

so we will subset the data to get the 2 separate ORs we want\*/

proc sort data=temp;

by descending sitq descending pmscc;

run;

proc freq data=temp order=data;

where sitq ne 3;/\*OR for T2 vs T1\*/

tables sitq\*pmscc/measures ;

run;

proc freq data=temp order=data;

where sitq ne 2;/\*OR for T3 vs T1\*/

tables sitq\*pmscc/measures ;

run;

/\*OR for T3 v T1 0.625, T2 v T1 0.81\*/