\*READ IN THE DATA FROM ASSIGNMENT 1;

LIBNAME exam1 'E:\workingdirectory\phc6053\exam1';

**DATA** exam1.mydata;

SET exam1.fghm81;

RUN;

\*ASSIGN THE DATA TO &DAT;

%let dat=exam1.mydata;

RUN;

\*OPTIONAL: OUTPUT DIRECTLY AS A PDF FILE/;

ods pdf file = "E:\workingdirectory\phc6053\exam1\exam1.pdf" notoc;

/\* CREATE A BMIGROUP VARIABLE\*/

**data** &dat;

set &dat;

BMIGROUP=**.**;

if BMI<**18.5** then BMIGROUP=**1**;

if BMI >= **18.5** & BMI < **25** then BMIGROUP = **2**;

if BMI >= **25** & BMI < **30** then BMIGROUP = **3**;

if BMI >= **30** then BMIGROUP = **4**;

RUN;

/\* %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%\*/

/\* %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%\*/

/\* PART 1 \*/

/\* %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%\*/

/\* %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%\*/

/\* 1. Remove all the individuals in the underweight BMI group AND all the individuals with a BMI which is 40 or larger\*/

**data** dat2;

set &dat;

if BMIGROUP = **1** then delete;

if BMI > **40** then delete;

run;

/\* Create LN(SYSBP)\*/

**data** dat2;

set dat2;

LNSBP=**.**;

LNSBP=log(SYSBP);

run;

/\* Create DUMMY variables for each shift \*/

**data** dat2;

set dat2;

if BMIGROUP = **2** then BMIOVER = **0**;

if BMIGROUP = **2** then BMIOBESE = **0**;

if BMIGROUP = **3** then BMIOVER = **1**;

if BMIGROUP = **3** then BMIOBESE = **0**;

if BMIGROUP = **4** then BMIOVER = **0**;

if BMIGROUP = **4** then BMIOBESE = **1**;

run;

/\* Create BMIDUM (for use with GLM)\*/

**proc** **format**;

value bmigroupcode **2**='normal' **3**=' over' **4**=' obese';

**run**;

**data** dat2;

set dat2;

BMIDUM = BMIGROUP;

**data** dat2;

set dat2;

format BMIDUM bmigroupcode.;

run;

/\*TAKE A LOOK AT THE FIRST 7 OBSERVATIONS\*/

**proc** **print** data=dat2 (obs=**7**);

**run**;

/\* %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%\*/

/\* %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%\*/

/\* PART 2 \*/

/\* %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%\*/

/\* %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%\*/

/\* 2. Create a scatterplot with LOESS curve for Y=LN(SYSBP) by BMI \*/

/\* (Note: you found the simple linear regression equation for this relationship in an assignment) \*/

**proc** **loess** data=dat2;

model LNSBP=BMI;

**run**;

/\*3. Create side-by-side boxplots of Y= LNSBP by BMI groups (you should only have 3 groups now).\*/

/\* first sort the data\*/

**proc** **sort** data=dat2;

by BMIGROUP;

**run**;

**proc** **boxplot** data=dat2;

plot LNSBP\*BMIGROUP;

**run**;

/\* %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%\*/

/\* %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%\*/

/\* PART 3 Investigate the relationship between LNSBP and BMI using regression with the

categorical version of BMI (reference=normal)\*/

/\* %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%\*/

/\* %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%\*/

/\*ods pdf file = "E:\workingdirectory\phc6053\exam1\exam1question4.pdf" notoc;\*/

/\* 4. Conduct a linear regression analysis for Y = LNSBP using the indicator variables you created for BMI groups as predictors.

Provide only the table of parameter estimates

(note: if you want to convince yourself, run an ANOVA on this data and compare the results to those of the reg model)\*/

**proc** **reg** data=dat2;

model LNSBP= BMIOVER BMIOBESE/clb vif ;;

**run**;

**quit**;

/\*ods pdf close;\*/

/\*5. Conduct a linear regression analysis for Y=LNSBP using BMI groups.

In comparison to the previous model, you must use your software to create the indicator variables

for you while still using normal individuals as the reference group.

The goal is to obtain the same analysis as the previous question without coding

the indicator variables directly. Provide only the table of parameter estimates \*/

**proc** **glm** data=dat2;

class BMIDUM;

model LNSBP=BMIDUM / solution clparm; ;

**run**;

**quit**;

/\*6. Using the results from the model in question 4 (or 5, since they should be identical)......\*/

/\* a. Write the estimated regression model \*/

/\* b. Interpret all three parameter estimates in the model clearly in the words of the problem \*/

/\* c. Show EXPLICITLY how you can use the model to estimate the mean LNSBP for each of the three levels of BMI.

Show all work. (Note: you should verify for yourself that these are simply the sample means for the three groups in analysis

(which does not adjust for any other covariates))\*/

/\* %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%\*/

/\* %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%\*/

/\* PART 4 The following analysis will investigate an interaction in the relationship between LNSBP

and two predictors: quantitative BMI and BPMEDS (reference = NO) \*/

/\* %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%\*/

/\* %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%\*/

/\* 7. Create a scatterplot of Y=LNSBP and X=BMI grouped by BPMEDS with LOWESS curves for each

value of BPMEDS. (Note: don't answer here but... what is "happening" behind this plot? Think about it!"\*/

/\* First, create a BPMEDSREC variable which uses English instead of numbers\*/

**proc** **format**;

value bpmedscode **0**='no' **1**=' yes' ;

**run**;

**data** dat2;

set dat2;

BPMEDSREC = BPMEDS;

**data** dat2;

set dat2;

format BPMEDSREC bpmedscode.;

run;

/\* Now sort the data\*/

**proc** **sort** data=dat2;

by BPMEDSREC;

**run**;

**proc** **loess** data=dat2;

by BPMEDSREC;

model LNSBP=BMI;

**run**;

/\* 8. Conduct a linear regression analysis for Y=LNSBP using BMI (quantitative), BPEMDS (reference=no),

and their interaction term. Allow software to handle reference category and interaction term. Provide only table of

parameter estimates.\*/

/\* First, create interaction term\*/

**data** dat2;

set dat2;

intBMI\_BPMEDS = BMI \* BPMEDS;

run;

**proc** **reg** data=dat2;

model LNSBP= BMI BPMEDS intBMI\_BPMEDS;

**run**;

**quit**;

/\* 9. Using the results from the model in question 8... \*/

/\* a. Write the complete estimated regression model \*/

/\* b. Show EXPLICITLY how to use the complete estimated model to find the estimated equations

for each BPMED group and provide the simplified equations relating LNSBP and BMI for each BPMED group \*/

/\* c. Provide an interpretation, in the words of the problem, of the effect of BMI on the mean LNSBP for

each BPMED group. \*/

/\* d. Use the current model to estimate the mean LNSBP within each BPMED group for BMI values of 20, 30 and 40.

Show your work and provide a summary table of your calculated estimates. \*/

/\* %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%\*/

/\* PART 5. The following analysis will investigate an interaction in the relationship between LNSBP

and two predictors: categorized BMI (reference = normal) and BPMDES (reference = no) \*/

/\* %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%\*/

/\*10. Conduct a linear regression analysis for Y = LNSBP using bmi groups (categorical, reference=normal),

BPMEDS (reference = no) and their interaction term. It is up to you to determine how to handle the reference

categories and interaction term, but I highly suggest allowing the software to handle these components instead of

creating variables yourself. Provide only the table of parameter estimates. \*/

/\* First, create multilevel interaction terms\*/

**data** dat2;

set dat2;

intBMIOVER\_BPMEDS = BMIOVER \* BPMEDS;

run;

**data** dat2;

set dat2;

intBMIOBESE\_BPMEDS = BMIOBESE \* BPMEDS;

run;

**proc** **reg** data=dat2;

model LNSBP= BMIOVER BMIOBESE BPMEDS intBMIOVER\_BPMEDS intBMIOVER\_BPMEDS;

**run**;

**quit**;

/\* 11. Using the results from the model in question 10... \*/

/\* a. Write the complete ESTIMATED regression model. \*/

/\* b. Use the current model to estiamte the mean LNSBP within each BPED group for each BMI category.

Show your work and provide a summary table of your calculated estimates. (Note: you can verify for yourself that these

are the sample means for the groups defined by each combination of BMI group and BPMEDS)

/\* %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%\*/

/\* PART 6. The following analysis adds the binary variable PREVSTRK (ref=no) to the previous investigation

using an interaction between the quantitative BMI and BPMEDS (ref=no) \*/

/\* %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%\*/

/\*12. Conduct a linear regression analysis for Y=LNSBP using PREVSTRK (ref=no) as only predictor \*/

**proc** **reg** data=dat2;

model LNSBP = PREVSTRK;

**run**;

**quit**;

/\* a. provide on the table of parameters from the output. \*/

/\* b. Interpret precisely, in the words of the problem, the parameter estimate for PREVSTRK and its confidence interval \*/

/\*13. Conduct a linear regression analysis for Y=LNSBP using PREVSTRK (ref=no), BMI (quantitative) and BPMEDS (ref=no)

and the itneraction term between BMI (quantitative) and BPMEDS (reference=no). Provide only the table of parameter estimates.\*/

**proc** **reg** data=dat2;

model LNSBP = PREVSTRK BMIOVER BMIOBESE BPMEDS intBMI\_BPMEDS;

**run**;

**quit**;

/\*14. Calculate the percent change in the parameter estimate for PREVSTRK between the two models in questions 12 and 13.

(note: You might also check the parameter estimates from the model in question 8 to see how "stable" the estiamtes are

for BMI, BPMEDS and their interaction;

/\* %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%\*/

/\* PART 7. The final analysis adds the variable AGE to our analysis from question 13 \*/

/\* %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%\*/

/\*15. Conduct a linear regression analysis for Y=LNSBP using AGE, PREVSTRK (ref=no), BMI (quant), BPMEDS (ref=no),

and the interaction term between BMI (quant) and BPMEDS (ref=no)\*/

**proc** **reg** data=dat2;

model LNSBP = PREVSTRK BMIOVER BMIOBESE BPMEDS AGE intBMI\_BPMEDS;

**run**;

**quit**;

/\* a. Provide only the table of parameter estimates. \*/

/\* b. Provide interpretations, in the words of the problem, of the parameter estimates for AGE and PREVSTRK \*/

/\* c. Write the full estimated regression model. \*/

/\* d. Explain precisely what the intercept represents in this analysis. Is this value meaningful in this situation? \*/

/\* e. Write the estimated regression model for the individuals with BPMEDS=no and interpret the partial slope of the BMI term,

in the words of the problem. \*/

/\* f. Write the estimated regression model for individuals with BPMEDS=Yes and interpret the partial slope of the BMI

term in the words of the problem. \*/

ods pdf close;