/*JOE BREW
PHC 6053
Exam 1*/

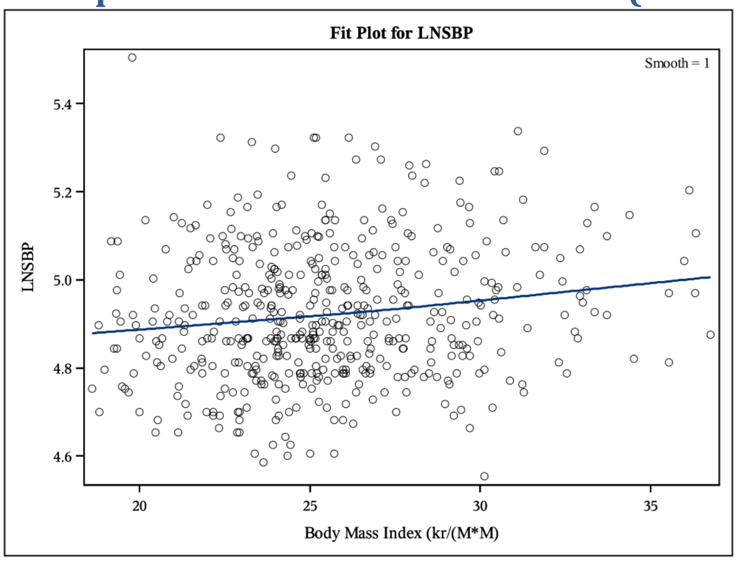
```
*READ IN THE DATA FROM ASSIGNMENT 1;
LIBNAME exam1 'E:\workingdirectory\phc6053\exam1';
DATA exam1.mydata;
SET exam1.fghm81;
RUN;

*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;
```

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

```
/* VARIABLE NAMES:
BMIGROUP (4 level categorical variable labeled as directed [1=under, 2=norm, 3=over, 4=obese)
BMIOVER (2 level categorical variable, labeled as 1=overweight, 2=not)
BMIOBESE (2 level categorical variable, labeled as 1=obese, 2=not)
BMIDUM (4 level categorical variable labeled in English as underweight (removed), normal, over,
obese)
*/

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)*/
```

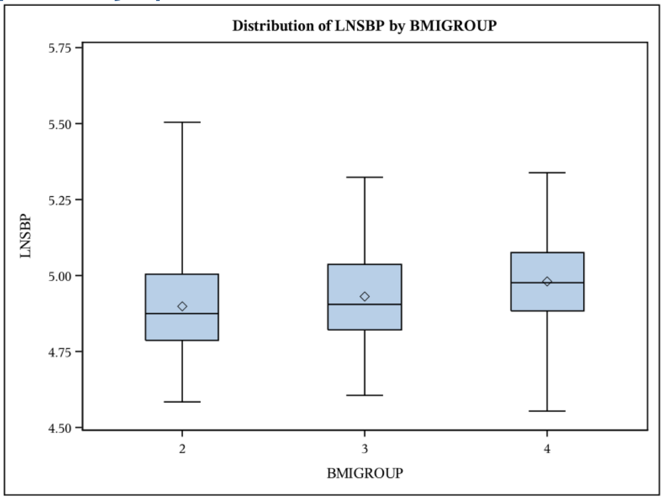
```
proc format;
   value bmigroupcode 2='normal'  3=' over'  4=' obese';
run;

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;
```

/* 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;
```

/* 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)*/
```

Parameter Estimates								
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**Variance Inflation**	**95% Confidence Limits**	
Intercept	1	4.89990	0.01006	487.18	<.0001	0	4.88014	4.91966
BMIOVER	1	0.03036	0.01451	2.09	0.0369	1.10006	0.00186	0.05887
BMIOBESE	1	0.08023	0.02312	3.47	0.0006	1.10006	0.03481	0.12565

```
proc reg data=dat2;
model LNSBP= BMIOVER BMIOBESE/clb vif ;;
run;
quit;
```

/*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 */
```

Parameter	Estimate		Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept	4.899897815	B	0.01005759	487.18	<.0001	4.880136189	4.919659442
BMIDUM obese	0.080226918	B	0.02311719	3.47	0.0006	0.034805178	0.125648658
BMIDUM over	0.030362247	B	0.01450858	2.09	0.0369	0.001855100	0.058869394
BMIDUM normal	0.000000000	B	.	.	.	.	.

```
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 */

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_i X_{pi} + \varepsilon_i$$

$$LNSBP_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_i X_{pi} + \varepsilon_i$$

$$\hat{Y} = \beta_0 + \beta_{obese} X_{obese} + \beta_{overweight} X_{overweight}$$

$$E(LNSBP|BMIGROUP) = \beta_0 + \beta_{obese}(obese) + \beta_{overweight}(overweight)$$

$$E(LNSBP|BMIGROUP) = 4.8999 + 0.08023(obese) + 0.03036(overweight)$$

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

The **intercept** (4.8999) is the mean log of systolic blood pressure (LNSBP) when both dummy variables (obese / over) are equal to 0. In other words, this is the mean LNSBP for a person of normal weight.

The **estimate for obese** (0.08023) is the mean increase in LNSBP when X_{obese} increases by one unit. Given that this is a binary dummy variable, this simply means that this is the mean difference between the LNSBP of obese relative to normal.

The **estimate for over** (0.0304) is the mean increase in LNSBP when X_{over} increases by one unit. Given that this is a binary dummy variable, this simply means that this is the mean difference between the LNSBP of overweight relative to normal.

/* 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))*/

One can estimate mean LNSBP by simply plugging a person's characteristics into the model. For example, a person of normal weight gets a 0 in the obese and a 0 in the over categories. Accordingly, their estimated mean LNSBP will be

$$E(LNSBP|BMIGROUP_{normal}) = 4.8999 + 0.08023(0) + 0.03036(0)$$
$$E(LNSBP|BMIGROUP_{normal}) = 4.8999$$

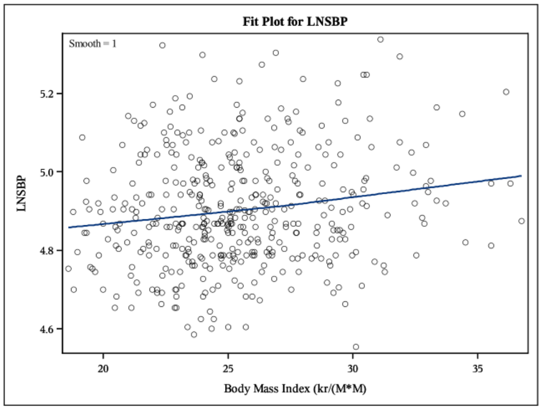
Likewise, one could estimate the mean LNSBP of an obese person by simply plugging in a 1 for the appropriate placeholder.

$$E(LNSBP|BMIGROUP_{obese}) = 4.8999 + 0.08023(1) + 0.03036(0)$$
$$E(LNSBP|BMIGROUP_{obese}) = 4.8999 + 0.08023$$
$$E(LNSBP|BMIGROUP_{obese}) = 4.98013$$

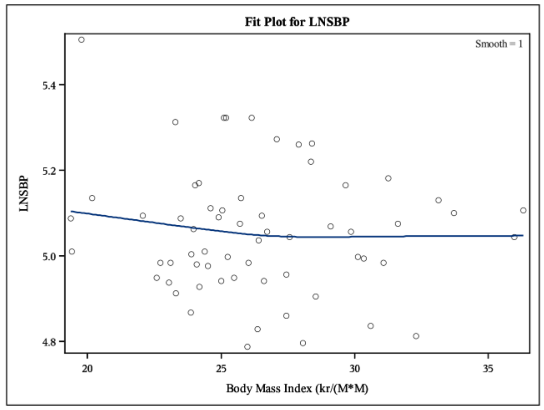
/* 7. Create a scatterplot of Y=LNSBP and X=BMI grouped by BPMEDS with LOWESS curves for each value of BPMEDS.

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

BPMEDSREC=no



BPMEDSREC=yes




```
/* 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.*/

Parameter Estimates						
Variable	**Label**	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	Intercept	1	4.72349	0.05231	90.29	<.0001
BMI	Body Mass Index (kr/(M*M)	1	0.00703	0.00204	3.45	0.0006
BPMEDS	Anti-hypertensive meds Y/N	1	0.43304	0.14213	3.05	0.0024
intBMI_BPMEDS		1	-0.01088	0.00536	-2.03	0.0427

```
/* 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 */

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_i X_{pi} + \varepsilon_i$$

$$LNSBP_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_i X_{pi} + \varepsilon_i$$

$$E(LNSBP|BMI, BPMEDS) = \beta_0 + \beta_{BMI} X_{BMI} + \beta_{BPMEDS} X_{BPMEDS} + \beta_{intBMIBPMEDS} X_{intBMIBPMEDS}$$

$$\hat{Y} = \beta_0 + \beta_{BMI}(BMI) + \beta_{BPMEDS}(BPMEDS) + \beta_{intBMIBPMEDS}(intBMIBPMEDS)$$

$$E(LNSBP|BMI, BPMEDS) = 4.72349 + 0.00703(BMI) + 0.43304(BPMEDS) - 0.01088(intBMIBPMEDS)$$

/* 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 */

BPMEDS = 0

$$E(LNSBP|BMI, BPMEDS_0) = 4.72349 + 0.00703(BMI) + 0.43304(0) - 0.01088(0)$$

$$E(LNSBP|BMI, BPMEDS_0) = 4.72349 + 0.00703(BMI)$$

BPMEDS = 1

$$E(LNSBP|BMI, BPMEDS_1) = 4.72349 + 0.00703(BMI) + 0.43304(1) - 0.01088(BMI)$$

$$E(LNSBP|BMI, BPMEDS_1) = 5.15653 - 0.00385(BMI)$$

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

Among those **not** on BPMEDS, for each 1 unit increase in BMI, the population mean LNSBP is estimated to **increase** by 0.00703 units.

Among those on BPMEDS, for each 1 unit increase in BMI, the population mean LNSBP is estimated to **decrease** by 0.00385 units.

/* 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. */

BPMEDS = 0

BMI=20
- $E(LNSBP|BMI_{20}, BPMEDS_0) = 4.72349 + 0.00703(20) + 0.43304(0) - 0.01088(0)$
- $E(LNSBP|BMI_{20}, BPMEDS_0) = 4.86409$

BMI=30
- $E(LNSBP|BMI_{30}, BPMEDS_0) = 4.72349 + 0.00703(30) + 0.43304(0) - 0.01088(0)$
- $E(LNSBP|BMI_{30}, BPMEDS_0) = 4.93439$

BMI=40
- $E(LNSBP|BMI_{40}, BPMEDS_0) = 4.72349 + 0.00703(30) + 0.43304(0) - 0.01088(0)$
- $E(LNSBP|BMI_{40}, BPMEDS_0) = 5.00469$

BPMEDS = 1

BMI=20
- $E(LNSBP|BMI_{20}, BPMEDS_1) = 4.72349 + 0.00703(20) + 0.43304(1) - 0.01088(20)$
- $E(LNSBP|BMI_{20}, BPMEDS_1) = 5.07953$

BMI=30
- $E(LNSBP|BMI_{30}, BPMEDS_1) = 4.72349 + 0.00703(30) + 0.43304(1) - 0.01088(30)$
- $E(LNSBP|BMI_{30}, BPMEDS_1) = 5.04103$

BMI=40
- $E(LNSBP|BMI_{40}, BPMEDS_1) = 4.72349 + 0.00703(40) + 0.43304(1) - 0.01088(40)$
- $E(LNSBP|BMI_{40}, BPMEDS_1) = 5.00253$

	BMI		
BPMEDS	**20**	**30**	**40**
0	4.86409	4.93439	5.00469
1	5.07953	5.04103	5.00253

/*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. */
```

Parameter Estimates						
Variable	**Label**	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	Intercept	1	4.88191	0.01003	486.69	<.0001
BMIOVER		1	0.02791	0.01461	1.91	0.0567
BMIOBESE		1	0.08684	0.02428	3.58	0.0004
BPMEDS	Anti-hypertensive meds Y/N	1	0.17754	0.03151	5.63	<.0001
intBMIOVER_BPMEDS		1	-0.02423	0.04288	-0.56	0.5724
intBMIOBESE_BPMEDS		1	-0.12273	0.05786	-2.12	0.0344

```
/* 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 intBMIOBESE_BPMEDS;
run;
quit;
```

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

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

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_i X_{pi} + \varepsilon_i$$

$$LNSBP_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_i X_{pi} + \varepsilon_i$$

$$E(LNSBP|BMIOVER, BMIOBESE, BPMEDS) = \beta_0 + \beta_{BMIOVER} X_{BMIOVER} + \beta_{BMIOBESE} X_{BMIOBESE}\ \beta_{BPMEDS} X_{BPMEDS} + \beta_{intBMIOVERBPMEDS} X_{intBMIOVERBPMEDS} + \beta_{intBMIOBESEBPMEDS} X_{intBMIOVERBPMEDS}$$

$$\hat{\hat{Y}} = \beta_0 + \beta_{BMIOVER}(BMIOVER) + \beta_{BMIOBESE}(BMIOBESE) + \beta_{BPMEDS}(BPMEDS) + \beta_{intBMIOVERBPMEDS}(intBMIOVERBPMEDS) + \beta_{intBMIOBESEBPMEDS}(intBMIOBESEBPMEDS)$$

$$E(LNSBP|BMIOVER, BMIOBESE, BPMEDS) = 4.88191 + 0.02791(BMIOVER) + 0.08684(BMIOBESE) + 0.17754(BPMEDS) - 0.02423(intBMIOVERBPMEDS) - 0.12273(intBMIOBESEBPMEDS)$$

/ b. Use the current model to estimate the mean LNSBP within each BPMED group for each BMI category.*

I manually wrote this model into R to do the calculations:

```
model10 <- function(bmiover, bmiobese, bpmeds){
  4.88191 +
    (0.02791*bmiover) +
    (0.08684*bmiobese) +
    (0.17754*bpmeds) -
    (0.02423*(bmiover*bpmeds)) -
    (0.12273*(bmiobese*bpmeds))}
```

<u>BPMEDS = 1 BMI = NORMAL</u>

$$E(LNSBP|BMIOVER_0, BMIOBESE_0, BPMEDS_1)$$
$$= 4.88191 + 0.02791(BMIOVER) + 0.08684(BMIOBESE) + 0.17754(BPMEDS) - 0.02423(intBMIOVERBPMEDS) - 0.12273(intBMIOBESEBPMEDS)$$

$$E(LNSBP|BMIOVER_0, BMIOBESE_0, BPMEDS_1) = 4.8819 + 0.17754(1)$$
$$E(LNSBP|BMIOVER_0, BMIOBESE_0, BPMEDS_1) = 5.05945$$

<u>BPMEDS = 1 BMI = OVER</u>

$$E(LNSBP|BMIOVER_1, BMIOBESE_0, BPMEDS_1)$$
$$= 4.88191 + 0.02791(BMIOVER) + 0.08684(BMIOBESE) + 0.17754(BPMEDS) - 0.02423(intBMIOVERBPMEDS) - 0.12273(intBMIOBESEBPMEDS)$$
$$E(LNSBP|BMIOVER_1, BMIOBESE_0, BPMEDS_1) = 5.06313$$

<u>BPMEDS = 1 BMI = OBESE</u>

$$E(LNSBP|BMIOVER_0, BMIOBESE_1, BPMEDS_1)$$
$$= 4.88191 + 0.02791(BMIOVER) + 0.08684(BMIOBESE) + 0.17754(BPMEDS) - 0.02423(intBMIOVERBPMEDS) - 0.12273(intBMIOBESEBPMEDS)$$
$$E(LNSBP|BMIOVER_0, BMIOBESE_1, BPMEDS_1) = 5.02356$$

<u>BPMEDS = 0 BMI = NORMAL</u>

$$E(LNSBP|BMIOVER_0, BMIOBESE_0, BPMEDS_0)$$
$$= 4.88191 + 0.02791(BMIOVER) + 0.08684(BMIOBESE) + 0.17754(BPMEDS) - 0.02423(intBMIOVERBPMEDS) - 0.12273(intBMIOBESEBPMEDS)$$
$E(LNSBP|BMIOVER_0, BMIOBESE_0, BPMEDS_0) = 4.88191$

<u>BPMEDS = 0 BMI = OVER</u>

$$E(LNSBP|BMIOVER_1, BMIOBESE_0, BPMEDS_0)$$
$$= 4.88191 + 0.02791(BMIOVER) + 0.08684(BMIOBESE) + 0.17754(BPMEDS) - 0.02423(intBMIOVERBPMEDS) - 0.12273(intBMIOBESEBPMEDS)$$
$E(LNSBP|BMIOVER_1, BMIOBESE_0, BPMEDS_0) = 4.90982$

<u>BPMEDS = 0 BMI = OBESE</u>

$$E(LNSBP|BMIOVER_0, BMIOBESE_1, BPMEDS_0)$$
$$= 4.88191 + 0.02791(BMIOVER) + 0.08684(BMIOBESE) + 0.17754(BPMEDS) - 0.02423(intBMIOVERBPMEDS) - 0.12273(intBMIOBESEBPMEDS)$$
$E(LNSBP|BMIOVER_0, BMIOBESE_1, BPMEDS_0) = 4.96875$

	BMI		
BPMEDS	**NORMAL**	**OVERWEIGHT**	**OBESE**
0	4.88191	4.90982	4.96875
1	5.05945	5.06313	5.02356

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)

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

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	4.91733	0.00691	711.77	<.0001
PREVSTRK	Prevalent Stroke (Infarct,Hem)	1	0.17374	0.04415	3.94	<.0001

```
/* a. provide on the table of parameters from the output. */
proc reg data=dat2;
model LNSBP = PREVSTRK;
run;
quit;
```

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

The **parameter estimate for PREVSTRK** (0.17374) is the mean increase in LNSBP when $X_{PREVSTRK}$ increases by one unit. This simply means that this is the mean difference between the LNSBP of those with prevalent stroke relative to normal. Individuals for whom PREVSTRK=1 (ie, those with prevalent stroke) have a population mean LNSBP 0.17374 units greater than those for whom PREVSTRK=0 (ie, those without prevalent stroke).

(The **intercept** (4.91733) is the mean log of systolic blood pressure (LNSBP) when the PREVSTRK variable is 0. In other words, this is the mean LNSBP of those who do not have prevalent stroke.)

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

```
Provide only the table of parameter estimates.*/
```

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	4.72341	0.05202	90.80	<.0001
PREVSTRK	Prevalent Stroke (Infarct,Hem)	1	0.10941	0.04279	2.56	0.0109
BMI	Body Mass Index (kr/(M*M)	1	0.00699	0.00203	3.44	0.0006
BPMEDS	Anti-hypertensive meds Y/N	1	0.44702	0.14143	3.16	0.0017
intBMI_BPMEDS		1	-0.01183	0.00534	-2.22	0.0272

```
proc reg data=dat2;
model LNSBP = PREVSTRK BMI 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;
```

Question 12 parameter estimate for PREVSTRK: 0.17374
Question 13 parameter estimate for PREVSTRK: 0.10941
Percent change = `(0.17374 - 0.10941) / 0.17374`
37% decrease (ie, when adjusted for other variables, the effect of prevstrk on LNSBP is reduced by 37%)

/*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)*/

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

Parameter Estimates						
Variable	**Label**	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	Intercept	1	4.34697	0.06759	64.31	<.0001
PREVSTRK	Prevalent Stroke (Infarct,Hem)	1	0.05274	0.04082	1.29	0.1970
BMI	Body Mass Index (kr/(M*M)	1	0.00771	0.00191	4.04	<.0001
BPMEDS	Anti-hypertensive meds Y/N	1	0.44404	0.13293	3.34	0.0009
AGE	Age (years) at examination	1	0.00603	0.00074757	8.07	<.0001
intBMI_BPMEDS		1	-0.01240	0.00502	-2.47	0.0139

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

AGE: Adjusting for other variables, for each one unit increase in age, the population mean for LNSBP is estimated to increase by 0.00603 units.

PREVSTRK: After adjustment for other variables, individuals for whom PREVSTRK=1 (ie, those with prevalent stroke) have a population mean LNSBP 0.05274 units greater than those for whom PREVSTRK=0 (ie, those without prevalent stroke)

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

$$E(LNSBP|PREVSTRK, BMIOVER\ \ , BPMEDS\ \ , AGE, intBMIBPMEDS\ ) = \\ 4.34697 + \ 0.05274(PREVSTRK) + 0.00771(BMI) + 0.44404(BPMEDS) + 0.00603(AGE) - 0.01240(intBMIBPMEDS)$$

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

The intercept is the hypothetical population mean of LNSBP for individuals for whom all of the parameter estimates are 0. In laymen's terms, this means the estimated population mean LNSBP for people without prevalent stroke, with a BMI of 0, who are not on BPMEDS, who are 0 years old.

This is **not** meaningful. The physical laws of the universe (as far as I know) do not allow for a BMI of 0.

/ 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. */*

$$E(LNSBP|PREVSTRK, BMIOVER\ \ , BPMEDS_0, AGE, intBMIBPMEDS\ ) = \\ 4.34697 + \ 0.05274(PREVSTRK) + 0.00771(BMI) + 0.44404(BPMEDS) + 0.00603(AGE) - 0.01240(0)$$

$$E(LNSBP|PREVSTRK, BMIOVER\ \ , BPMEDS_0, AGE, intBMIBPMEDS\ ) = \\ 4.34697 + \ 0.05274(PREVSTRK) + 0.00771(BMI) + 0.00603(AGE)$$

Among those not on BPMEDS, and holding constant prevalent stroke and age, for each 1 unit increase in BMI, the population mean LNSBP is estimated to increase by 0.00771 units.

/ 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. */*

$$E(LNSBP|PREVSTRK, BMIOVER\quad, BPMEDS_1, AGE, intBMIBPMEDS) =$$
$$4.34697 + \; 0.05274(PREVSTRK) + 0.00771(BMI) + 0.44404(BPMEDS) + 0.00603(AGE) - 0.01240(intBMIBPMEDS)$$
$$4.34697 + \; 0.05274(PREVSTRK) + 0.00771(BMI) + 0.44404(1) + 0.00603(AGE) - 0.01240(1 * BMI)$$

$$4.79101 + \; 0.05274(PREVSTRK) + 0.00603(AGE) - 0.00469(BMI)$$

Among those on BPMEDS, and holding constant prevalent stroke and age, for each 1 unit increase in BMI, the population mean LNSBP is estimated to decrease by 0.00469 units.

```
proc reg data=dat2;
model LNSBP = PREVSTRK BMI BPMEDS AGE intBMI_BPMEDS;
run;
quit;

ods pdf close;
```