PHC 6053: Regression Methods for the Health and Life Sciences
Dataset Information for Assignments

Dataset Information:

You will be provided a random sample from a larger dataset from the Framingham Heart Study. The dataset has the following variables.

#	Variable	Label
1	SEX	SEX
2	RANDID	Random ID
3	TOTCHOL	Serum Cholesterol mg/dL
4	AGE	Age (years) at examination
5	SYSBP	Systolic BP mmHg
6	DIABP	Diastolic BP mmHg
7	**CURSMOKE**	Current Cig Smoker Y/N
8	CIGPDAY	Cigarettes per day
9	BMI	Body Mass Index (kr/(M*M)
10	**DIABETES**	Diabetic Y/N
11	**BPMEDS**	Anti-hypertensive meds Y/N
12	HEARTRTE	Ventricular Rate (beats/min)
13	GLUCOSE	Casual Glucose mg/dL
14	**PREVCHD**	Prevalent CHD Y/N
15	**PREVAP**	Prevalent Angina Y/N
16	**PREVMI**	Prevalent MI Y/N
17	**PREVSTRK**	Prevalent Stroke Y/N
18	**PREVHYP**	Prevalent Hypertension Y/N
19	PERIOD	Examination cycle
20	HDLC	HDL Cholesterol mg/dL
21	LDLC	LDL Cholesterol mg/dL

Codes for Categorical Variables:

Yes/No Variables: (In Bold)	0 = No	1 = Yes

Sex/Gender	1 = Male	2 = Female

BMI Groups: Given in data or created by student depending on the assignment.

1 = Underweight	$(0 \leq BMI < 18.5)$
2 = Normal	$(18.5 \leq BMI < 25)$
3 = Overweight	$(25 \leq BMI < 30)$
4 = Obese	$(BMI \geq 30)$

You are highly encouraged to set up formats and meaningful labels for all variables but it is not required.