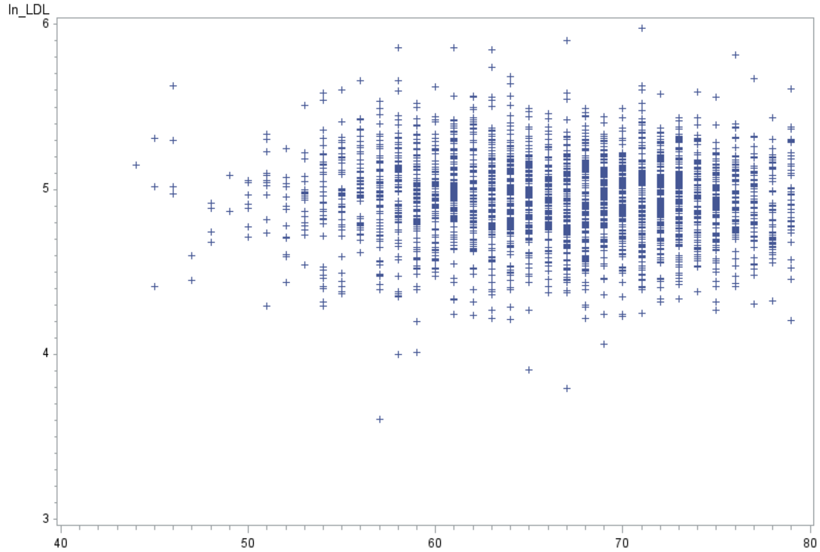
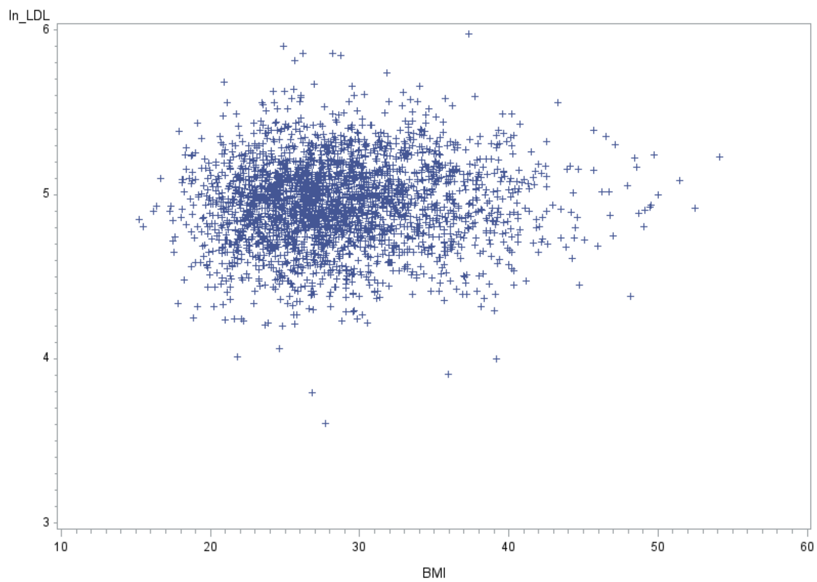
1.

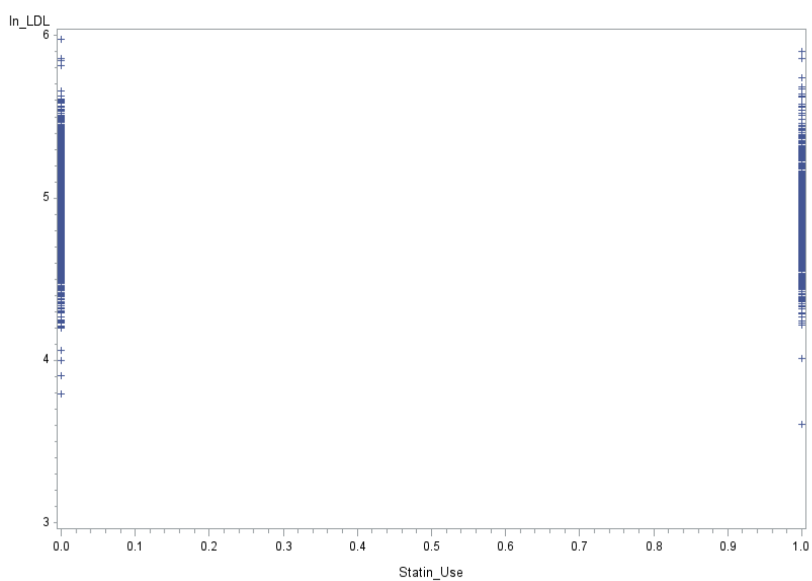
ln(LDL) versus Age



ln(LDL) versus BMI



ln(LDL) versus Statin use



From the scatter plots, ln(LDL) seems doesn't have significant association with age and BMI, respectively. If examine the correlation between each pair of them, I would expect the correlations all be around zero. ln(LDL) also seems doesn't have association with statin use status, since at both level of statin use (1=yes, o=no), the distributions of ln(LDL) don't differentiate, at least can't be detected by eyes.

2.

Model: $E(\ln(LDL)\ |Age) = \beta_1 + \beta_2 * Age$

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	**1**	5.04919	0.04958	101.84	<.0001
age	**1**	-0.00158	0.00074026	-2.13	0.0333

Ln(LDL) and Age are significantly negatively associated (p=0.0333), that is, disregard all other variables that could potentially influence their relationship, every 1-unit increase in Age is associated with a 0.00158-unit decrease in the mean ln(LDL).

Model: $E(\ln(LDL)\ |Age) = \beta_1 + \beta_2 * BMI$

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	**1**	4.86546	0.02589	187.93	<.0001
BMI	**1**	0.00275	0.00088959	3.10	0.0020

Ln(LDL) and BMI are significantly positively associated (p=0.0020), that is, disregard all other variables that could potentially influence their relationship, every 1-unit increase in BMI is associated with a 0.00275-unit increase in the mean ln(LDL).

Model: $E(\ln(LDL)\ |Age) = \beta_1 + \beta_2 * Statin\ Use$

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	**1**	4.98771	0.00601	829.99	<.0001
Statin_Use	**1**	-0.12021	0.00998	-12.05	<.0001

Ln(LDL) and Statin Use have significantly association (p<0.0001), that is, the mean ln(LDL) in having statin use group is 0.12021 lower than the mean ln(LDL) in the group without statin use, disregard all other variables that could potentially influence their relationship.

3.

a.

Assume that observations of ln(LDL) from different individuals are independent, and the conditional distribution of ln(LDL) given covariates (Age, BMI, Statin Use) has a univariate normal distribution, with mean

$$E(\ln(LDL)\,|Age, BMI, Statin\ Use) = \beta_1 + \beta_2 * Age + \beta_3 * BMI + \beta_4 * Statin\ Use$$

And also assume homogeneity of variance.

b.

Based on the definition given above, the mean ln(LDL) is a linear combination of the three covariates Age, BMI, and Satin use. The association between ln(LDL) and Age is conditional, that is, given BMI and statin use status, or in other words, holding BMI and statin use constant. Same for other covariates: the association between ln(LDL) and BMI is conditional, i.e., holding Age and statin use constant.; the association between ln(LDL) and statin use constant is conditional, i.e., holding Age and BMI constant.

c.

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept	1	4.98512	0.05846	85.28	<.0001	4.87049	5.09975
age	1	-0.00101	0.00073048	-1.38	0.1686	-0.00244	0.00042648
BMI	1	0.00243	0.00087837	2.76	0.0058	0.00070498	0.00415
Statin_Use	1	-0.11946	0.00997	-11.98	<.0001	-0.13900	-0.09991

Interpretation of Association:

After adjusting for BMI and Statin use status (holding BMI and Statin use status constant), age is unrelated to ln(LDL) (p=0.1686). After accounting for potential confounders, the relationship has been modified.

After adjusting for Age and Statin use status (holding Age and Statin use status constant), BMI is still positively correlated to ln(LDL) (p=0.0058). After accounting for potential confounders, the relationship doesn't change much.

After adjusting for Age and BMI (holding Age and BMI constant), statin use status is still significantly associated with ln(LDL) (p<0.0001). After accounting for potential confounders, the relationship doesn't change much.

Interpretation of estimates & confidence intervals:

β_2:Holding BMI and Statin use status constant, every 1-unit increase in age is associated with 0.00101 decrease in ln(LDL), which is not significant (p=0.1686); and we are 95% confident that based on our sample data, the "true" constant rate of change or slope is between -0.00244 and 0.00042648 (not sure positive or negative).

β_3:Holding Age and Statin use status constant, every 1-unit increase in age is associated with 0.00243 increase in ln(LDL), which is significant (p=0.0058); and we are 95% confident that based on our sample data, the "true" constant rate of change or slope is between 0.00070498 and 0.00415 (all positive).

β_4:Holding Age and BMI constant, the mean ln(LDL) in having statin use group is 0.11946 lower than this in without statin use group, which is significant (p<.0001); and we are 95% confident that based on our sample data, the "true" difference is between -0.139 and -0.09991 (all negative).

4.

The objective here is to examine the association between a set of CHD risk factors, in particular, the association between LDL with Age, BMI, and Statin use, respectively, using data from the Heart and Estrogen/Progestin Replacement study(HERS). From preliminary examination of the data by simple pairwise plots, we found no association between any of them. However, when using ordinary linear regression models to estimate pairwise associations, we found that, disregard all other variables that could potentially influence their relationship, LDL and Age are significantly negatively associated (p=0.0333); LDL and BMI are significantly positively associated (p=0.0020); LDL and Statin Use have significantly association (p<0.0001). Furthermore, we examined multivariate association by a single ordinary linear regression model, which presents some modification about their relationship. After adjusting for BMI and Statin use status, age is unrelated to LDL (p=0.1686); after adjusting for Age and Statin use status, BMI is still positively correlated to LDL (p=0.0058), every 1-unit increase in age is associated with 1.002433(95% CLs:1.000705, 1.004159) increase in LDL;and after adjusting for Age and BMI, statin use status is still significantly negatively associated with ln(LDL) (p<0.0001), the mean LDL in having statin use group is 1.126888 (95% CLs: 1.105071, 1.149124) lower than this in without statin use group. In conclusion, we found that, LDL is positively associated with Age and negatively associated with Statin use, while LDL has no association with Age. The effect in the univariate model (LDL vs Age) is misleading because of not accounting for confounder BMI.

5.

a.

Model & Assumption:

Assume that observations of ln(LDL) from different individuals are independent, and the conditional distribution of ln(LDL) given covariates (BMI, Statin Use) has a univariate normal distribution, with mean

$$E(\ln(LDL)\,|BMI, Statin\ Use) = \beta_1 + \beta_2 * BMI + \beta_3 Statin\ Use$$

And also assume homogeneity of variance.

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept	**1**	4.91271	0.02554	192.34	<.0001	4.86263	4.96279
BMI	**1**	0.00262	0.00086728	3.02	0.0025	0.00091951	0.00432
Statin_Use	**1**	-0.11983	0.00997	-12.02	<.0001	-0.13937	-0.10029

From Q1 we see that, Ln(LDL) and Statin Use have significantly association (p<0.0001), that is, the mean ln(LDL) in having statin use group is 0.12021 lower than the mean ln(LDL) in the group without statin use, disregard all other variables that could potentially influence their relationship.

The scientific question of interest that, whether BMI modifies the association between ln(LDL) and statin use, can be assessed by looking at the parameter estimate and p-value of β_3. After adjusting for BMI, ln(LDL) and Statin Use still have significant association (p<0.0001), and the mean ln(LDL) in having statin use group is 0.119831 lower than the mean ln(LDL) in the group without statin use. Compare to unadjusted result, BMI doesn't modify the association between ln(LDL) and Statin Use Status.

b.

The main effect of statin use represent the difference of mean ln(LDL) between two group (Have Statin use v.s. No Statin use) with BMI=0, that is, the difference of average ln(LDL) values for these two subpopulation those BMI are 0. However, in reality, no such human has 0 BMI. We would want the main effect of statin use to represent the difference of mean ln(LDL) between two most reasonable or common groups. Thus, the main effect here doesn't have a meaningful effect.

c.

$$E(\ln(LDL)\,|BMI, Statin\ Use) = \beta_1 + \beta_2 * BMI\_Centered + \beta_3 Statin\ Use$$

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept	**1**	4.98758	0.00600	831.17	<.0001	4.97581	4.99934
bmi_centered	**1**	0.00262	0.00086728	3.02	0.0025	0.00091951	0.00432
Statin_Use	**1**	-0.11983	0.00997	-12.02	<.0001	-0.13937	-0.10029

Now the main effect of statin use represent the difference of mean ln(LDL) between two group (Have Statin use v.s. No Statin use) with average BMI values (about 28.573). So people with average BMI values and having statin use, on average, has 0.11983 lower ln(LDL) value than those with average BMI value but no statin use. In summary, the reason why the main effect of stain use represent a meaningful quantity is that, it has a meaningful comparison between two most common subgroups, while with un-centered BMI, it only represents difference between two "ghost" groups.

Appendix (SAS Program):

```
**********************************
Applied Longitudinal Analysis
Xiner Zhou
2/7/2015
********************************;
*import dataset;
data hers;
infile  'C:\data\Projects\APCD High Cost\Longitudinal\hers.txt';
input ln_LDL age BMI Statin_Use;
run;

*Q1:Provide a plot of ln(LDL) against each of age, BMI, and statin use ;
proc gplot data=hers;
title 'ln(LDL) versus Age';
plot ln_LDL*age /overlay;
run;
proc gplot data=hers;
title 'ln(LDL) versus BMI';
plot ln_LDL*BMI /overlay;
run;
proc gplot data=hers;
title 'ln(LDL) versus Statin use';
plot ln_LDL*statin_use /overlay;
run;

*Q2:Use PROC REG in SAS to fit a regression model to describe each of the three associations
plotted in response to question 1;
proc reg data=hers;
model ln_LDL=age;
run;
proc reg data=hers;
model ln_LDL=BMI;
run;
proc reg data=hers;
model ln_LDL=statin_use;
```

```
run;

*Q3:Use PROC REG is SAS to fit a single model describing the multivariable association between
ln(LDL) and age, BMI, and statin use;
proc reg data=hers;
model ln_LDL= age BMI statin_use/clb ;
run;

*Q5a: Assess whether the data provide strong evidence that BMI modifies the ln(LDL): statin use
association;
proc means data=hers;var bmi;run;
data hers;
set hers;
bmi_centered=bmi-28.5730506;
run;

*Q5c:Construct a new variable (centered bmi) defined as bmi – mean(bmi) Re-fit the model;
proc reg data=hers;
model ln_LDL=bmi_centered statin_use /clb ;
run;
```