This lab we discussed estimation and hypothesis testing in multiple linear regression. Estimation is relatively straightforward using SAS. We will go over the details of doing hypothesis

tests using SAS.

* Overall test
* Test for addition of a single variable
* Test for addition of a group of variables

1. Example

We will use the HERS data set which was used in class.

* 1. Overall Test

We will begin by asking the question whether BMI or Age is useful in explaining the variability of LDL.

LDLi = ß0 + ß1BMIi + ß2Agei + εi

H0: ß1 = ß2 = 0

* + 1. What is the alternative hypothesis H1?

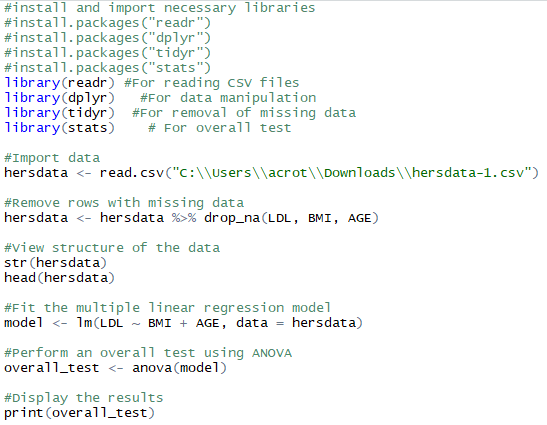
If the null hypothesis is that ß1 is equal to ß2 is equal to 0, then it is saying that there is no impact of those coefficients on LDL cholesterol. Therefore, the alternative hypothesis is that at least one, if not both, of ß1 or ß2 is not equal to zero, and has an impact on LDL cholesterol.

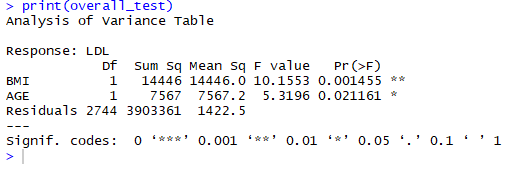
H1 : ß1 ≠ 0

H2 : ß2 ≠ 0

H3 : ß1 AND ß2 ≠ 0

* + 1. Perform the test in SAS and interpret the results.





An analysis of variance (ANOVA) was utilized to perform the overall test. The ANOVA results show that BMI has an F-value of 10.1553 with a p-value of 0.001455 and 1 Degree of Freedom. The results also show that Age has an F-value of 5.3196 with a p-value of 0.021161 and 1 Degree of Freedom. Both BMI and Age are indicated to have a statistically significant relationship with LDL based on their p-values being below the typical threshold of 0.05. The F-values indicate the strength of the relationship, which implies that the relationship between BMI and LDL is stronger than that of Age and LDL due to its higher F-value and lower p-value.

* 1. Test of Individual Regression Coefficients

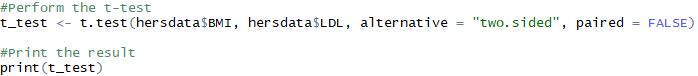
Once we have determined that at least one of the regressors is important, we can then test individual coefficients.

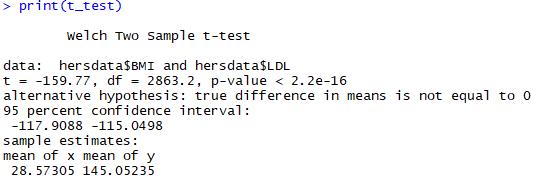
Given the model:

E(LDLi) = ß0 + ß1BMIi + ß2Agei

We want to test H0 : ß1 = 0 vs H1 : ß1 ≠ 0, given that Agei is in the model. We will do it in different ways.

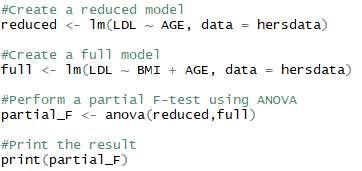
* + 1. T-test

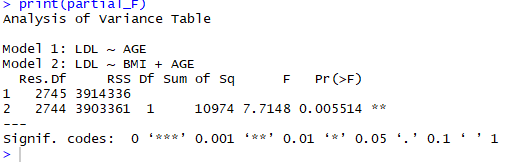




The t-test resulted in a t-statistic of -159.77 with a p-value of 2.2e-16, which is extremely small. This falls below the threshold of 0.05, providing evidence to reject the null hypothesis and accept the alternative hypothesis that there is a statistically significant relationship between LDL and BMI, given that Age is in the model.

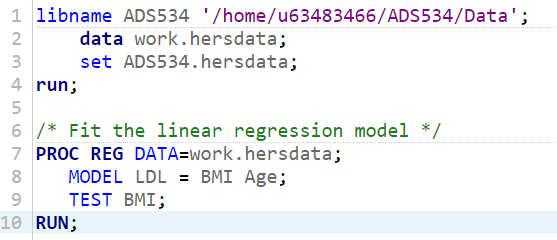
* + 1. Partial F-test by comparing SSR from two models: linear regression models with and without BMI.

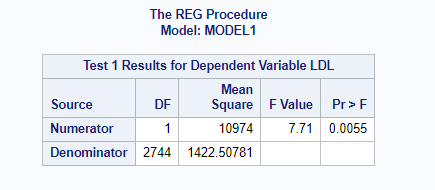




Since F = 7.7148 (p-value of 0.005514), we have evidence to reject the null hypothesis (ß1 = 0) at the 0.05 significance level. The variable BMI appears to have a statistically significant relationship with LDL, even with Age as a contributing variable.

* + 1. Partial F-test using the test statement in SAS





The output using the test statement in SAS produced an F-value of 7.71 (p-value of 0.0055). This gives us evidence to reject the null hypothesis and accept the alternative hypothesis that ß1 ≠ 0 given Age is included in the model. BMI has a statistically significant relationship with LDL even with the inclusion of Age as a variable.

* 1. Tests for Groups of Predictors

Often, it is of interest to determine if, collectively, a group of predictors significantly contribute to the variability in Y given another group of predictors are in the model.

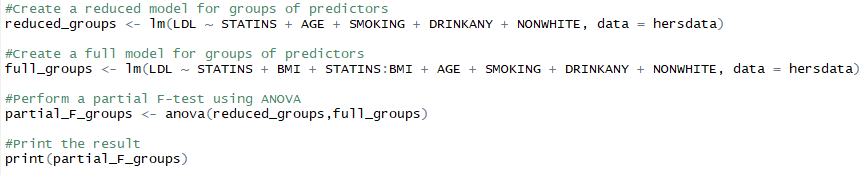
Given the model

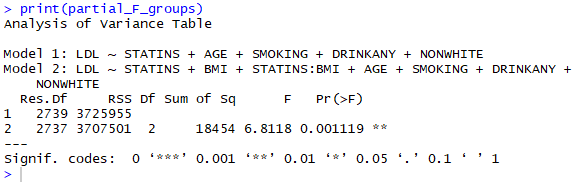
E(LDLi) = ß0 + ß1statinsi + ß2BMIi + ß3statinsiBMIi + ß4Agei + ß5Smokingi + ß6Drinkanyi + ß7Nonwhitei ,

where there are two terms associated with BMI. We would like to know if BMI is significantly associated with LDL levels, given the model that this association differs by statin use?

We want to test H0 : ß2 = ß3 = 0 vs H1: at least one of ß2, ß3 ≠ 0, given that other predictors are in the model. We will do it in different ways.

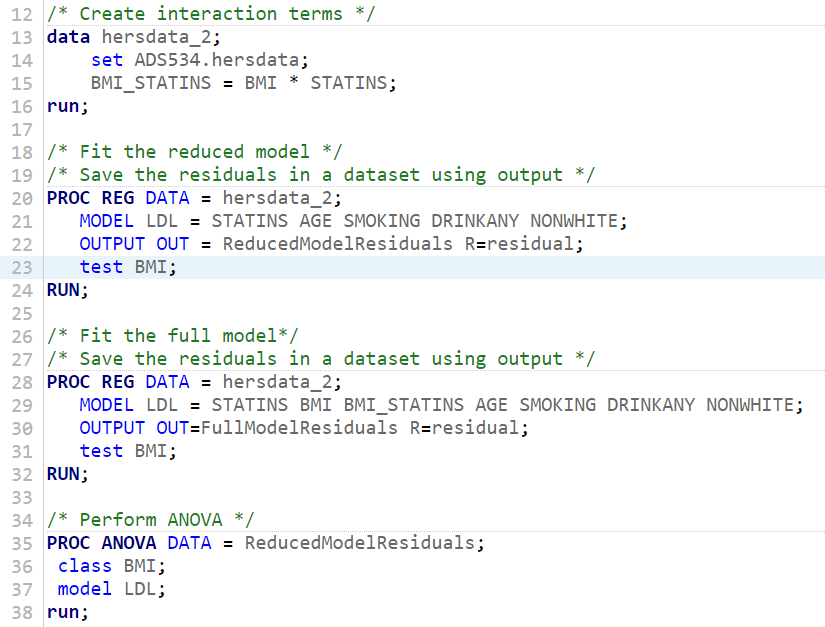
* + 1. Partial F-test by comparing SSR from two models: linear regression models with and without BMI.

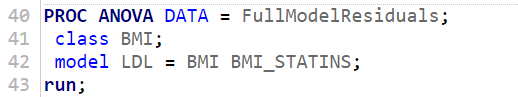


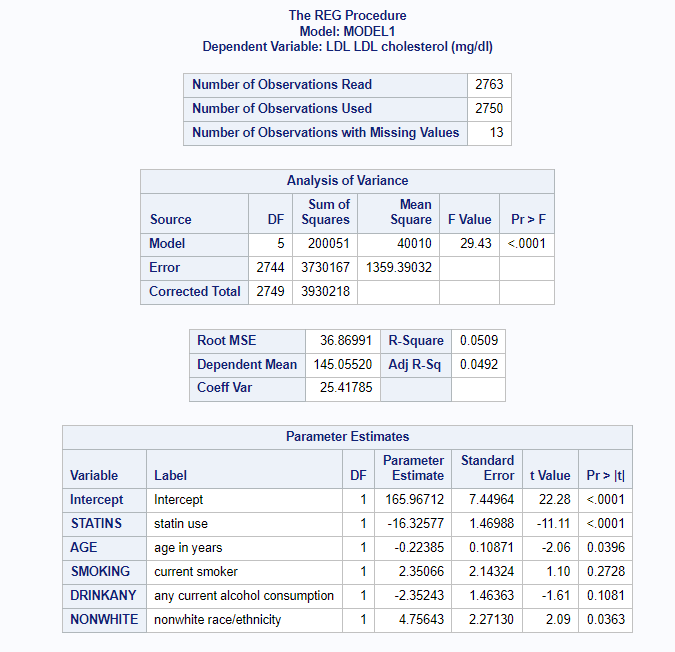


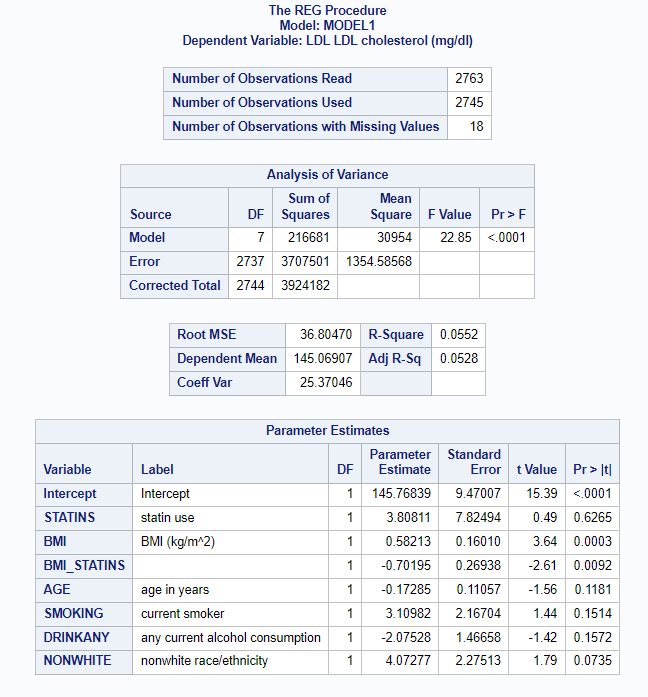
Since F = 6.8118 (p-value of 0.001119), we have evidence to reject the null hypothesis (ß2 = ß3 = 0) at the 0.05 significance level. The variable BMI and the relationship of statins with BMI appears to have a statistically significant relationship with LDL, even with a group of other predictor variables present within the model.

* + 1. Partial F-test using the test statement in SAS.









The output using the test statement in SAS produced an F-value of 29.43 (p-value of <0.0001) for the reduced model, and an F-value of 22.85 (p-value of <0.0001) for the full model. This gives us evidence to reject the null hypothesis and accept the alternative hypothesis that ß2 = ß3 = 0 given Age is included in the model. BMI has a statistically significant relationship with LDL even with the inclusion of Age as a variable.

What if we want to test whether BMI is significantly associated with LDL levels for those people receiving statin?

In other words, we want to test H0 : ß2 + ß3 = 0 vs H1: ß2 + ß3 ≠ 0, given that other predictors are in the model. Notice the differences in H0 from the above.

* + 1. Partial F-test using the test statement in SAS.