因为lasso我们去掉了, 所以你这边只有五页了, 黑色标题是我的note, 尽量简单一点, 因为韩国同学的应该变长了, 讲的时候尽量避免关键数字, 因为后面不太可能有修改了

1. 我们有两种方法把outlier揪出来, 一个是regression, 一个是重新用bmi公式计算然后比较两个值, 这里放的图是regression的, 但是我们code是两个都做了

Hello everyone, I am CHENYANG JIANG, now let me continue our presentation, just like Siri equation, there is strong connection among adiposity, weight and height, which called BMI equation, here is the results of linear regression with adiposity and the new variable with weight and height, and find 2 outliers, both adiposity can be updated by new one from BMI equation.

2.之前我们通过boxplot删点了四个点, 后来觉得不妥, 所以我们只删除了182.

Here we give a table on how we deal with our outliers in our dataset, because we only have about 250 observation, we want to fix as many points as we can. From heat plot of correlation, you can see that most variables are highly correlated with each other. Remind us to select essential variables in the following steps.

3.PCA, Factor analysis删除, 增加了一点为什么我们要用linear regression的理由.

Our group decide to use linear regression to our model, linear regression can give us simple and clear model which can be explained and rebuild for the unprofessional easily. From boxcox we can see that it is unnecessary to transformation. And there is serious multicollinearity in simple linear

regression of all variables, remind us to select variables.

4. Lasso , Ridge删除, 换成tree regression, 这边method你不用解释太多,

Our groups use these three methods to select variables, with tree regression, including bagging and boosting one, subsets method and direction search including forward, backward and both directions. We have four full models for method 2 and 3. The model with all raw variables, variables after logarithmic transformation, square transformation and with all variables from above three model. Since all variables is positive, we want to figure out any improvement after transformation on x. Note that we prefer simple models.

5. 没有明显更改

This is the results of subsets method for the fourth full model, subsets method will select model with good BIC, Cp and R square. From the whole results, we found that abdomen and its transformation is the most important variables. Followed by weight and wrist.