I found that when we add a page and record in this page, when we export the ppt into video, there is some bug in this page, so you can export your ppt after change to find is there any similar bug.

1.We add backward and both direction search here, so the title is changed

Hello everyone, I am HanGyu KANG, now let me continue our presentation, in three directions search method, here we only give the results of forward direction with BIC of all four full models, compared with that from AIC, results from BIC will give simpler models, just like that from subsets method, abdomen is also the most important variables.

2. I add the picture and first 6 important variables here, you can design how to interpret this page, meanwhile you can open our code and add more results about tree regression if you want, but attention that too small pictures may not clear in low quality mp4.

Add your interpretation.

3. 19 alternative models instead of 12.

Therefore, in variables selection, we find 19 alternative model, we only choose models less than 6 variables for simplicity. And you will find that there is no obvious improvement to add too many variables.

4. Here I put part of 19 results not all of them, it is impossible to put all results here

It is time to select models, our method is 30 repeated with 10-fold cross validation. Cross validation method divide dataset into training part for train and test part, which test model's ability to predict new data that was not used in train. We found that there is great improvement from model with 1 variable to 2 or more variables one. But there is almost no difference between model with

and without transformation. There is also no obvious improvement from model with 2 or 3 variables by adding more variables. Therefore we focus on model 3 and 4, who has no transformation and have 2 or 3 variables. Note that there is 0.003 Rsquared difference between them, so we decide to choose weight and abdomen as our final predictors.

5. The final model found ID39 is influential observation in Cooks' distance, so the real final model remove ID39

After we fit the model, the ID 39 is obvious influential observation in cook's distance, therefore we decide to remove it and fit again, here is the results of our final model, our models can explain about 72% variability of bodyfat, no serious multicollineraity and large residual error, there also exists clear linear relationship between predictors and bodyfat. We also gives an example how to use our model.

6.No change instead of picture

Now we conduct model diagnosis. There is no special pattern and obvious outliers in standardized residual plots. qq plots also perform well except some outliers, therefore our model meets normality, linearity and constant variance assumptions in linear regression well.

7.I remove the refit model without outliers in Leverage. It seems that this step is not nececssary. In fact I still do not know how to interpret Leverage, because when I fit the model without outliers in Leverage and plot Leverage again, there still some new outliers.

Leverage can show the outliers when we take abdomen and weight as predictors. as you can see, there is no very outstanding outliers in Leverage,