

Legends

- Week 2
- Week 3
- Week 4
- Week 5
- Week 6
- Week 7
- Week 8

Start with my full model for final project -- Predicting NBA player's salary based on player's own features

Divide the data into training and test data

Fitting the linear model and estimating the coefficients using test data

Check the two additional conditions

Check linearity by creating plots of residuals vs each predictor and fitted value and look for any systematic patterns

Check constant variance by plots of residuals vs each predictors and fitted values and look for any pattern, especially fanning pattern

Check the four assumptions of linear regression

Check normality by normal QQ plot and look for whether the points match the critical values of standard normal

Check uncorrelated error by plots of residuals vs predictor and fitted value and look for any patterns like large clusters

Plot the response against the fitted values and scatterplot of all the predictors in pairs

Do the plots show a severe deviation of the conditions?

The patterns we see from residual plots can not tell us what is wrong.

Fit the model and check assumptions again

Apply Box-Cox transformation on Y and X at a time or simultaneously. Find a close value of lamda that is easier for transformation

Depending on how variance is related to mean responses, we can use variance stabilizing transformation, such as ln, inverse

Are there any violations of assumptions?

Are constant variance the only violated assumption?

Do common transformations like square root or natural log on one or both the predictors and response work?

Perform hypothesis test (T test) on each coefficients of predictor with a null hypothesis coefficients equals to zero. Check the significance of each predictor.

Removing this predictor is reasonable and we can move on to look for other non-significantly predictors

A

B

C



