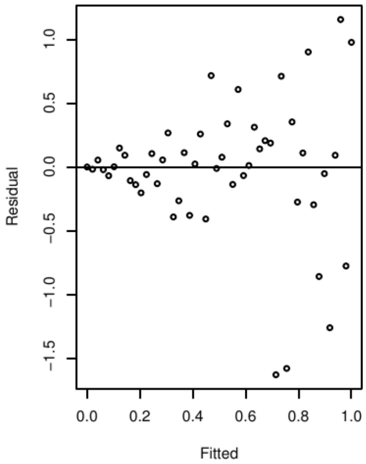
Quiz 6

04/15/2019

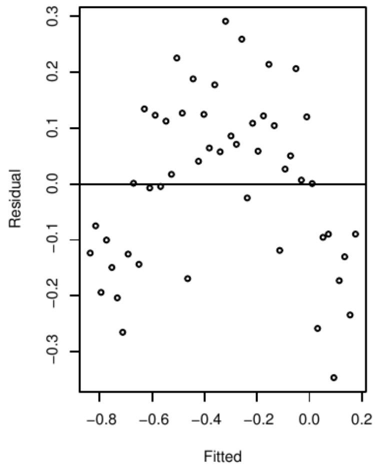
Instructions:
Write your name at the top right. You are to work on this quiz alone without any help from any other resource *except for a single* 8.5×11 *inch page of handwritten notes.*

Problem 1: Describe what regression assumption might fail, examining the plot below. What are the specific implications on our (ordinary) regression analysis of this assumption failing?



The plot shows non-constant variance which does not in itself bias the estimation of the model parameters, but can dilute our inference and explanatory power. Specifically, the hypothesis tests and confidence intervals will no longer be valid, such that we cannot tell "how good" our estimate is, even though it is unbiased.

Problem 2: Describe what regression assumption might fail, examining the plot below. What are the specific implications on our (ordinary) regression analysis of this assumption failing?

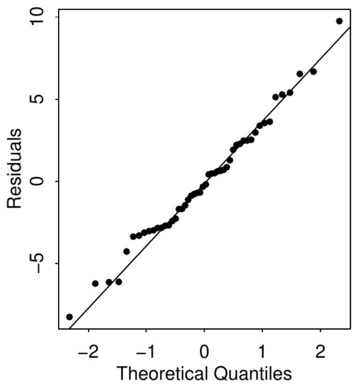


The hypothesis of the model structure itself should be in question. The residuals exhibit non-linear structure, so that our proposed form for the model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{1}$$

may not be valid. Particularly, this will mean that our inferences and explanations will be misleading and fundamentally biased, let alone the uncertainty of the parameters, etc...

Problem 3: Let the horizontal axis be defined as the standard normal quantiles. Name the type of plot below and describe what regression assumption is being evaluated. What are the specific implications on our (ordinary) regression analysis if this assumption fails? Describe two example situations when this assumption may not be strictly necessary.



This is a Q-Q plot, used to evaluate the assumption of the Gaussianity of the error $\boldsymbol{\epsilon}$. If Gaussianity fails, we may still perform the least-squares as a best, unbiased linear estimator, but this will no longer be a maximum-likelihood estimator. In addition, our confidence intervals and our hypothesis tests use this assumption, so it will dilute our uncertainty quantification. However, we can do "OK", when for instance, the error is short-tailed or when we have a large number of observations (using the Central Limit Theorem).

Problem 4: Explain what leverage is in terms of the values of $\mathbf{X}$. Explain what an outlier is and when: (i) an outlier **will** have a strong effect; and (ii) an outlier **will not** have a strong effect on our regression analysis.

A point will have high leverage if the associated explanatory variables are far away from the center of mass in the ($\mathbf{X}$) predictor space. A point is an outlier when it deviates significantly from the signal or the model fit. An outlier has a strong impact on the regression analysis when it has high leverage, and will not have a strong impact on the regression analysis when it has small leverage.