STA 674

Regression Analysis And Design Of Experiments

Assessing Model Assumptions – Lecture 2

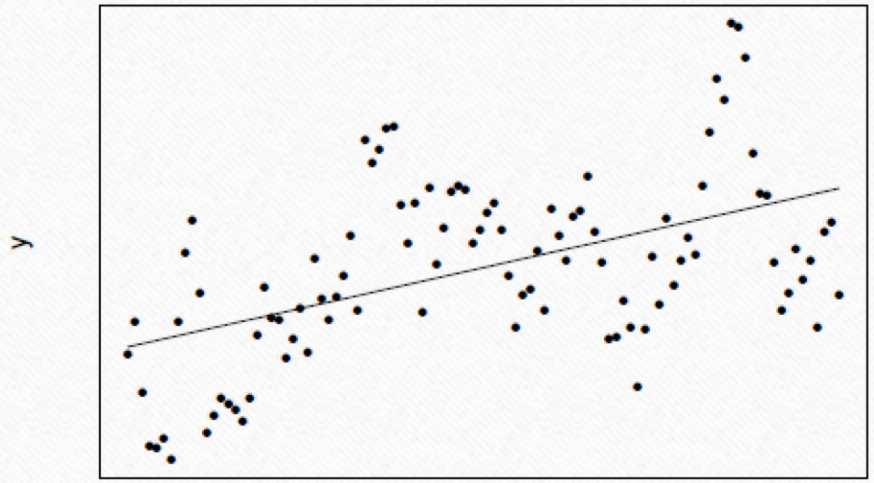
STA 674, RADOE:
Assessing Model Assumptions

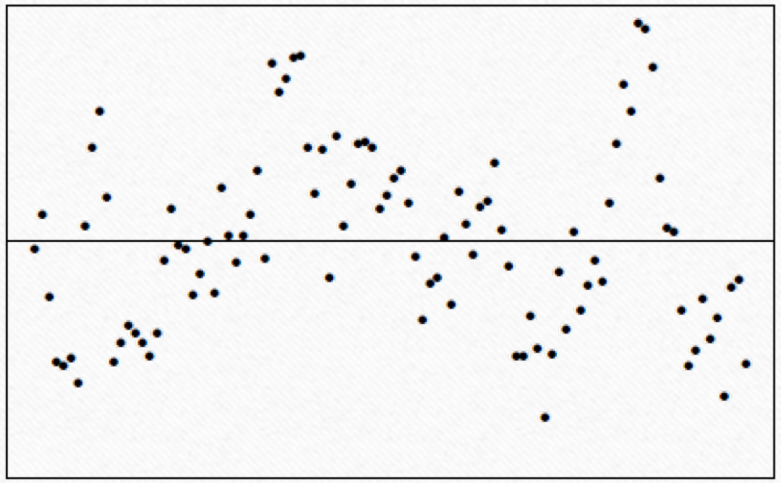
- Last time, we started our discussion of assessing model assumptions.
- This time, we cover the last assumption—normality of the errors.

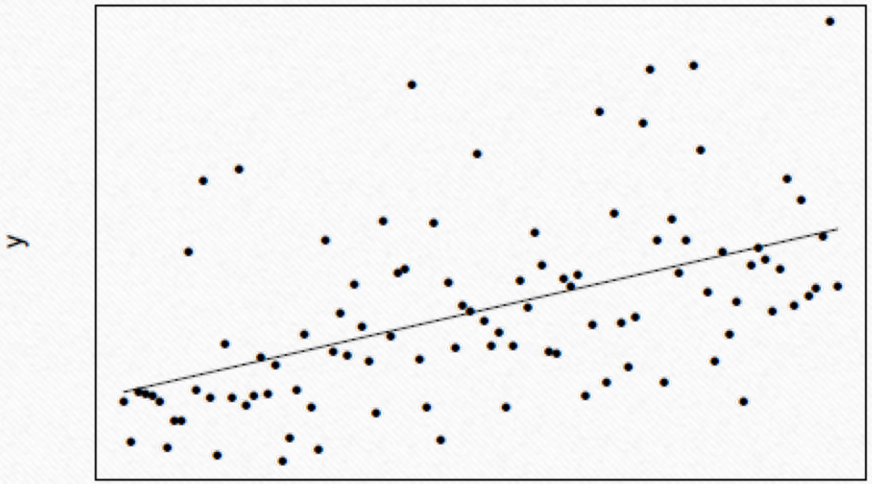
STA 674, RADOE:
Assessing Model Assumptions

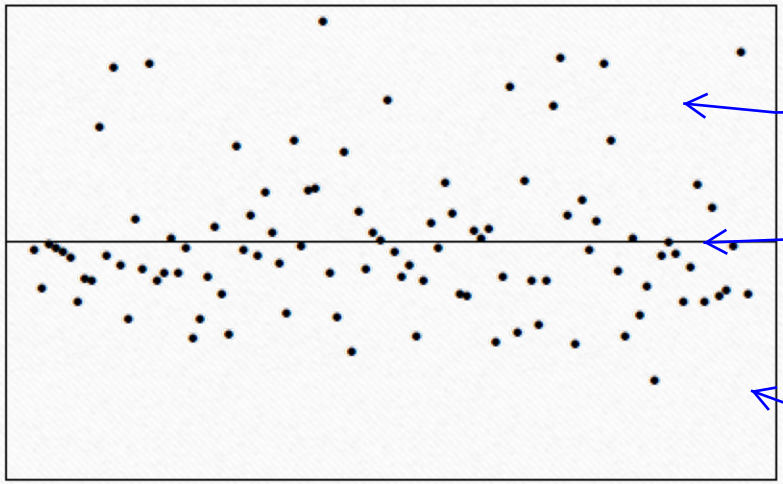
Assumptions

1. Mean Zero (Linearity): The average value of the errors is 0 regardless of the values of any of the predictors or of the response. data points should have a mean of 0...evenly distributed around residual plot of regression
2. Homoscedasticity (Equal variance): The errors all have variance σ_e^2. variance should be even across the residual plot
3. Independence: The errors are uncorrelated.
4. Normality: The errors are normally distributed.









higher positive variance

will result in positive skewed distribution

lower negative variance

more objective than a histogram

STA 674, RADOE:
Assessing Model Assumptions

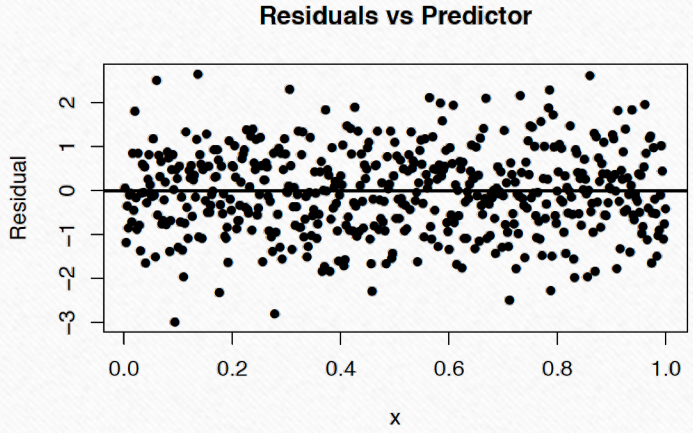
QQ-Plots

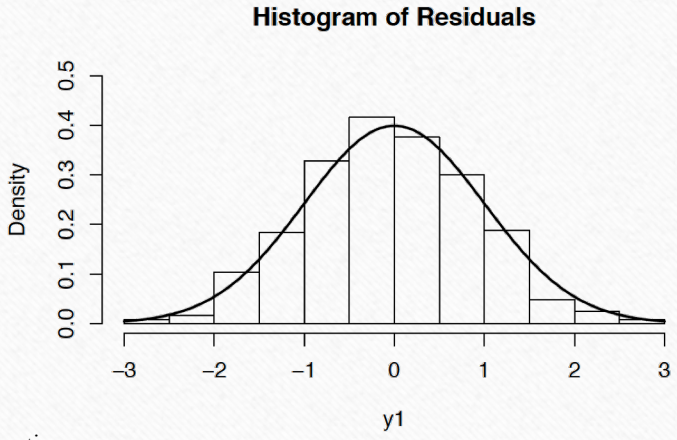
- A **normal QQ-plot** plots the quantiles of the residuals versus the theoretical quantiles of a normal distribution with the same mean and variance.
- If the residuals are normally distributed then the points should fall close to a straight line.

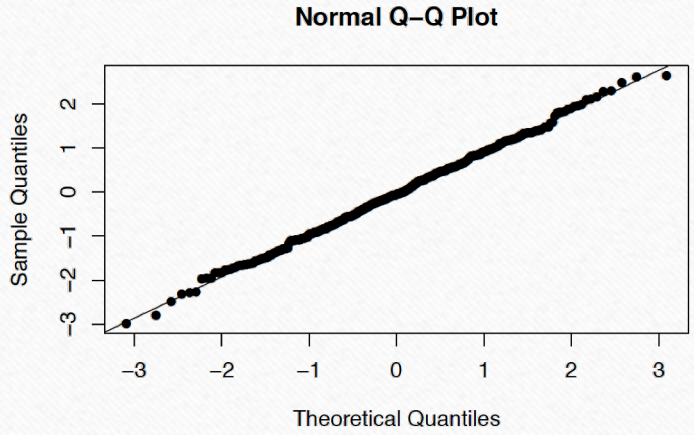
if there are a lot of points, a histogram can do a better job at assessing normality

QQ-Plots – Normally Distributed Residuals

ei normally distributed

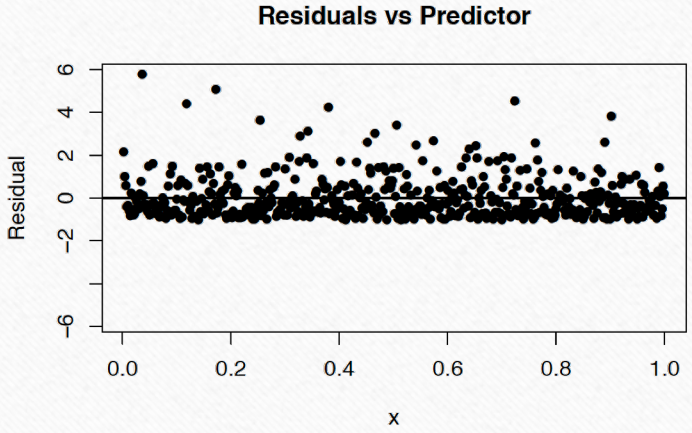


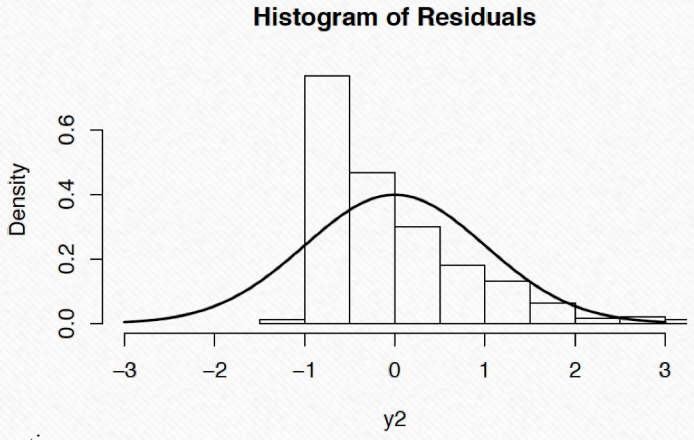


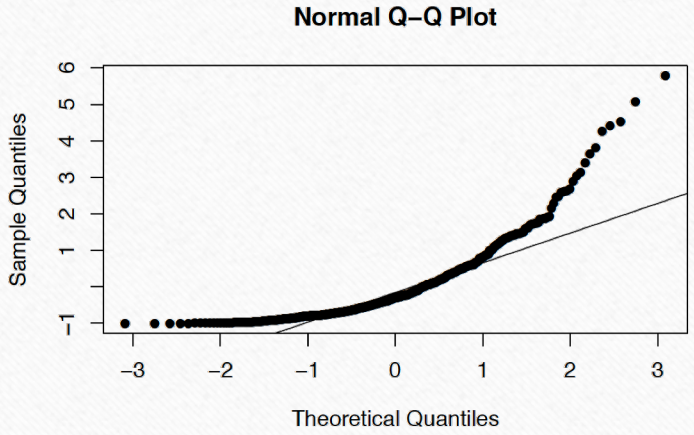


QQ-Plots – Right-Skewed Residual

ei not normally distributed

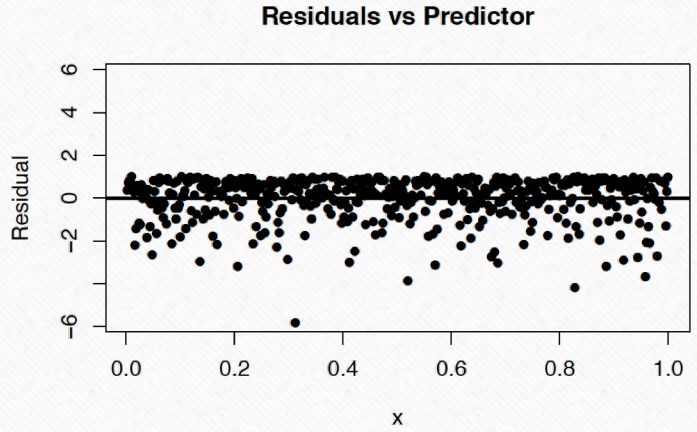


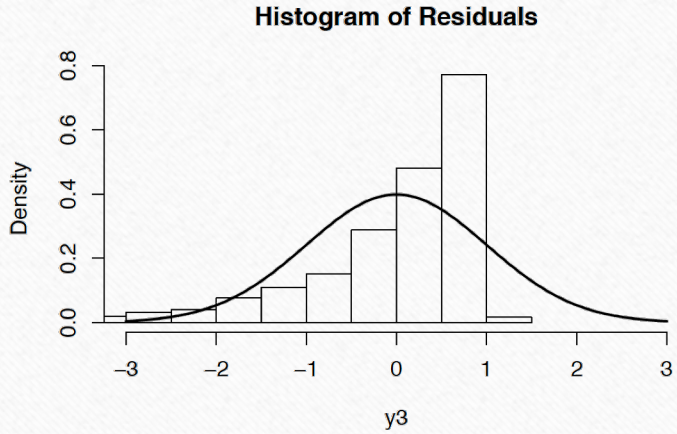


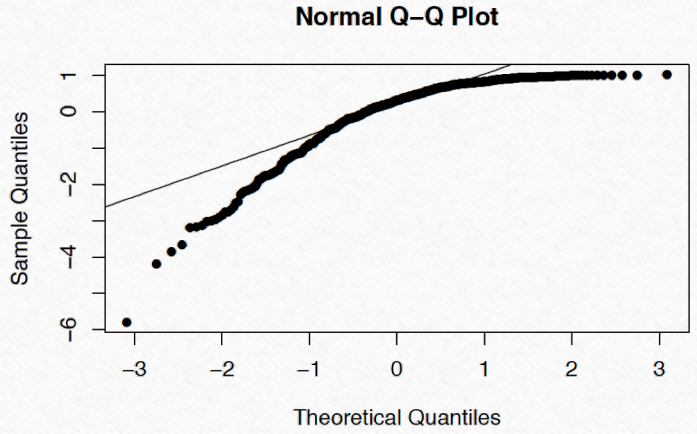


QQ-Plots – Left-Skewed Residuals

ei not normally distributed

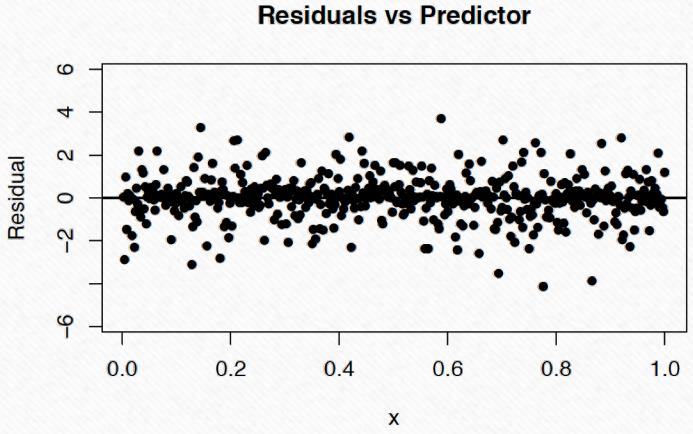


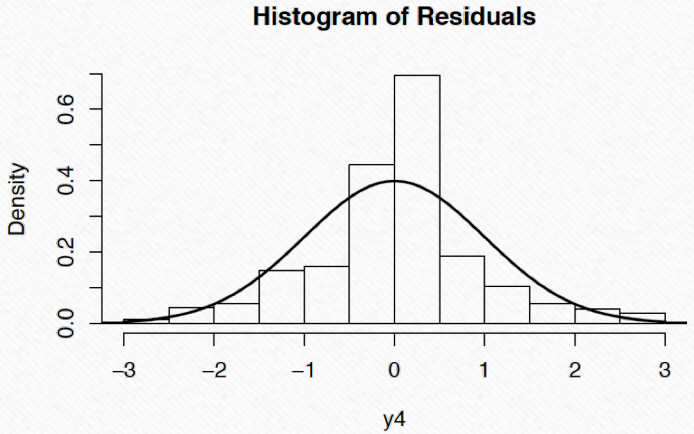


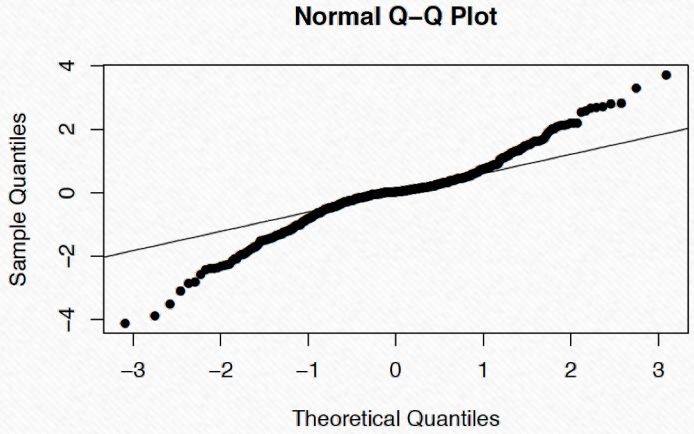


QQ-Plots –Heavy-Tailed Residuals

ei NOT normally distributed







STA 674, RADOE:
Assessing Model Assumptions

Examine assumptions, but if only a few points exclude the regression model, then it could be ok

Exercise: FEV Data

- The following plots display the residual diagnostics for the model of lung capacity (FEV) as a function of age, gender, and smoking status. Does the model appear to satisfy the regression assumptions?