STA 674

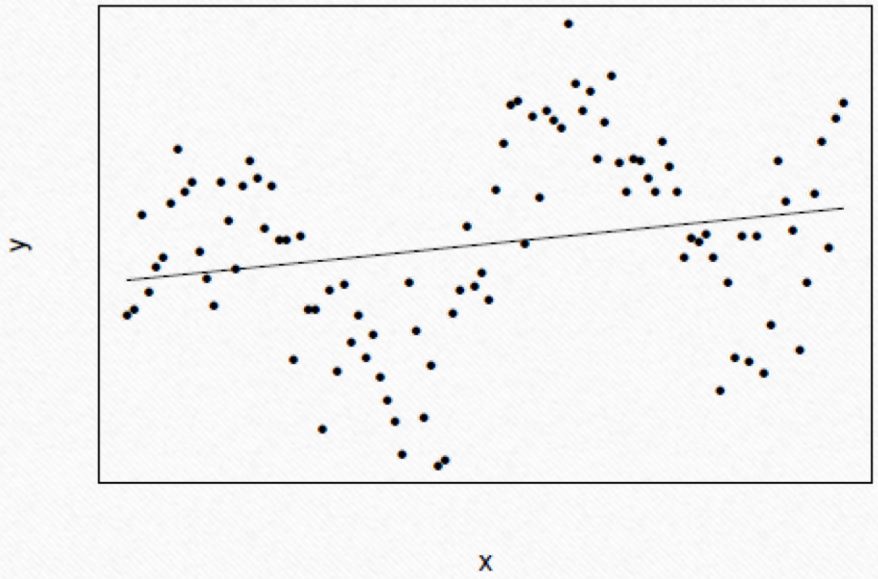
Regression Analysis And Design Of Experiments

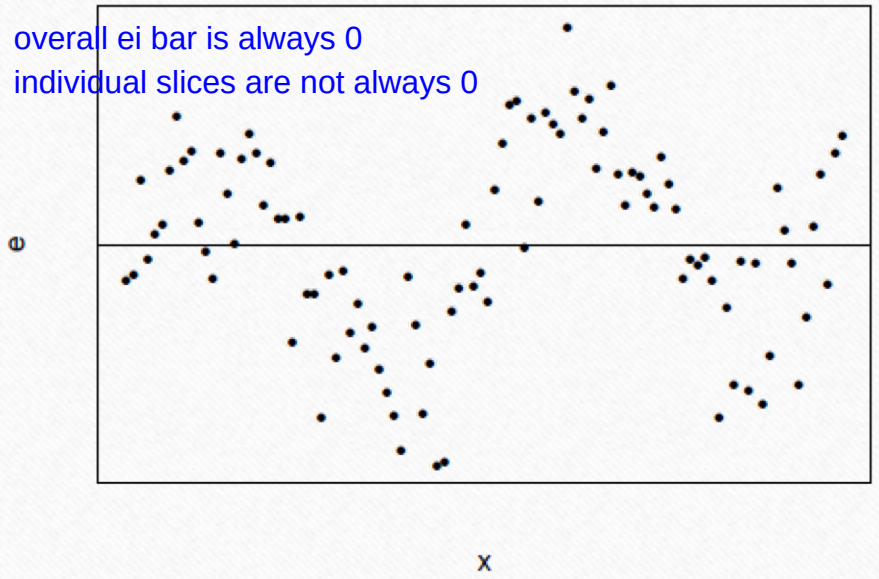
Assessing Model Assumptions – Lecture 1

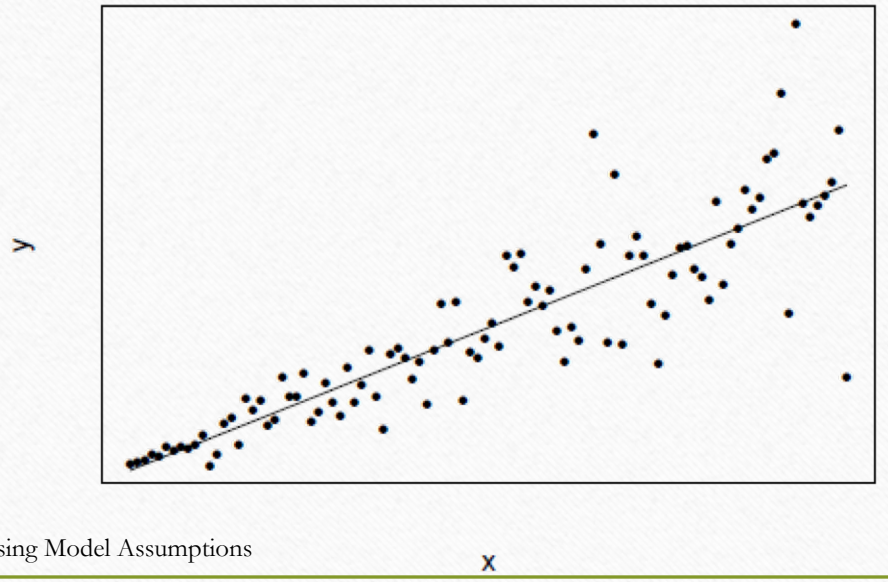
STA 674, RADOE:
Assessing Model Assumptions

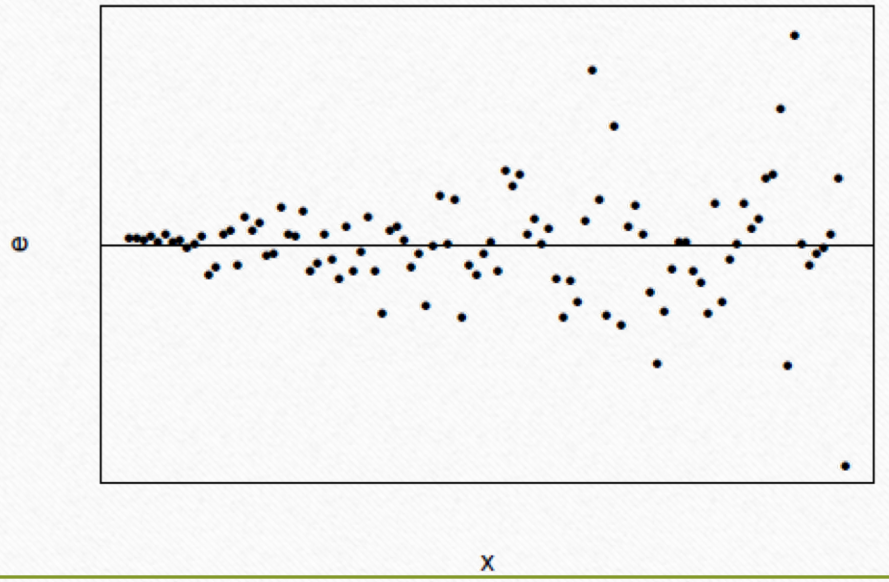
- Where does it fit in?
- What is it?
- Where next?

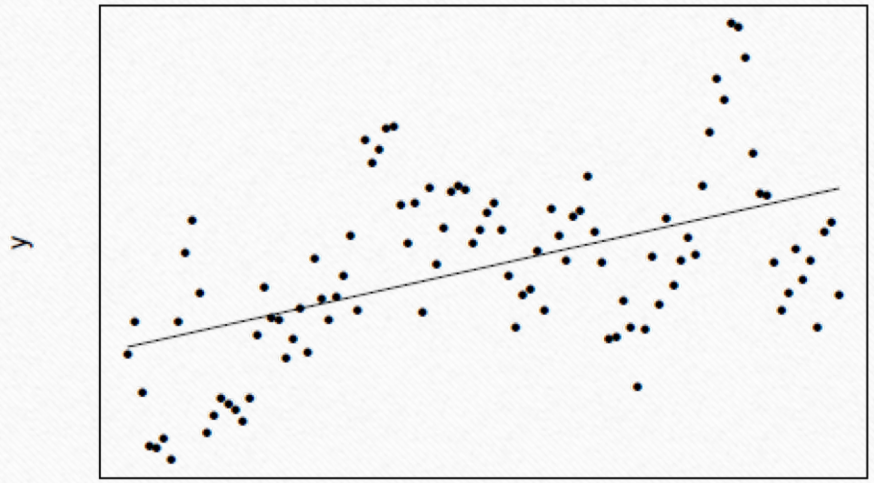
overall ei bar is always 0
individual slices are not always 0

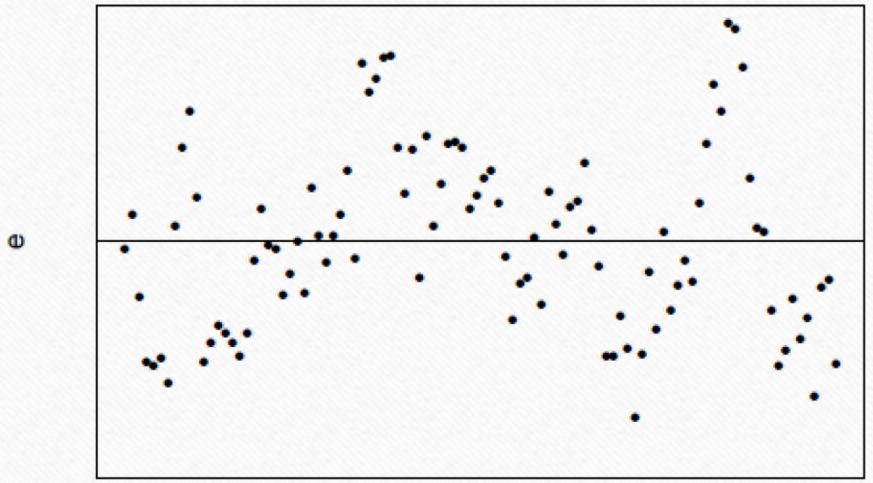


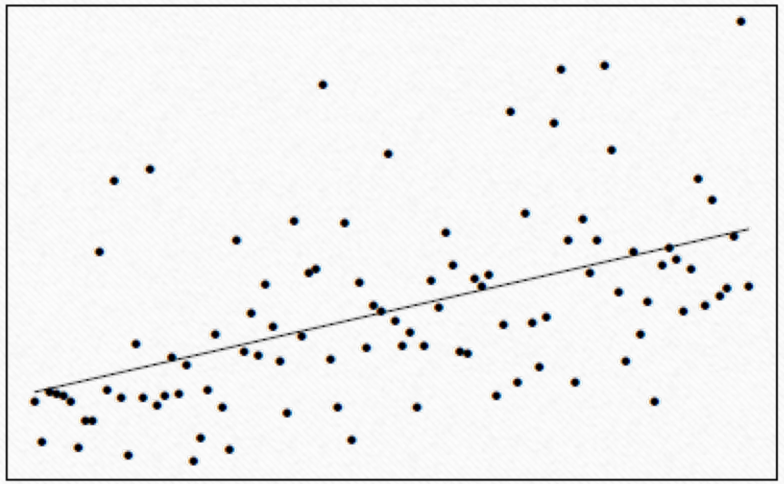


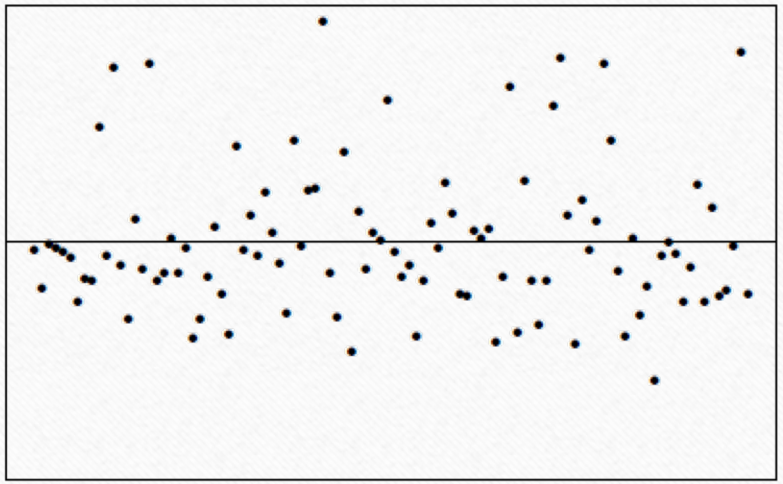





y
x


e
x


y
x


e
x

STA 674, RADOE:
Assessing Model Assumptions

Assumptions of the regression model:

- When fitting a linear regression model we assume that the errors satisfy the following assumptions:

1. The errors all have mean 0.
2. The errors all have variance σ_e^2.
3. The errors are normally distributed.
4. The errors are independent.

- If any of these assumptions are violated, then:
  - parameter estimates may be inaccurate,
  - standard errors may be too big or too small,
  - confidence intervals may not have the correct coverage probability, and
  - conclusions from tests may be mistaken.

STA 674, RADOE:
Assessing Model Assumptions

Residuals

- The fit of a model to the data is assessed via the residual values:

$$e_i = y_i - \hat{y}_i$$
$$= \text{Observed} - \text{Expected}$$

- Residual Plots: Assumptions of the model should always be assessed by plotting the residuals against:

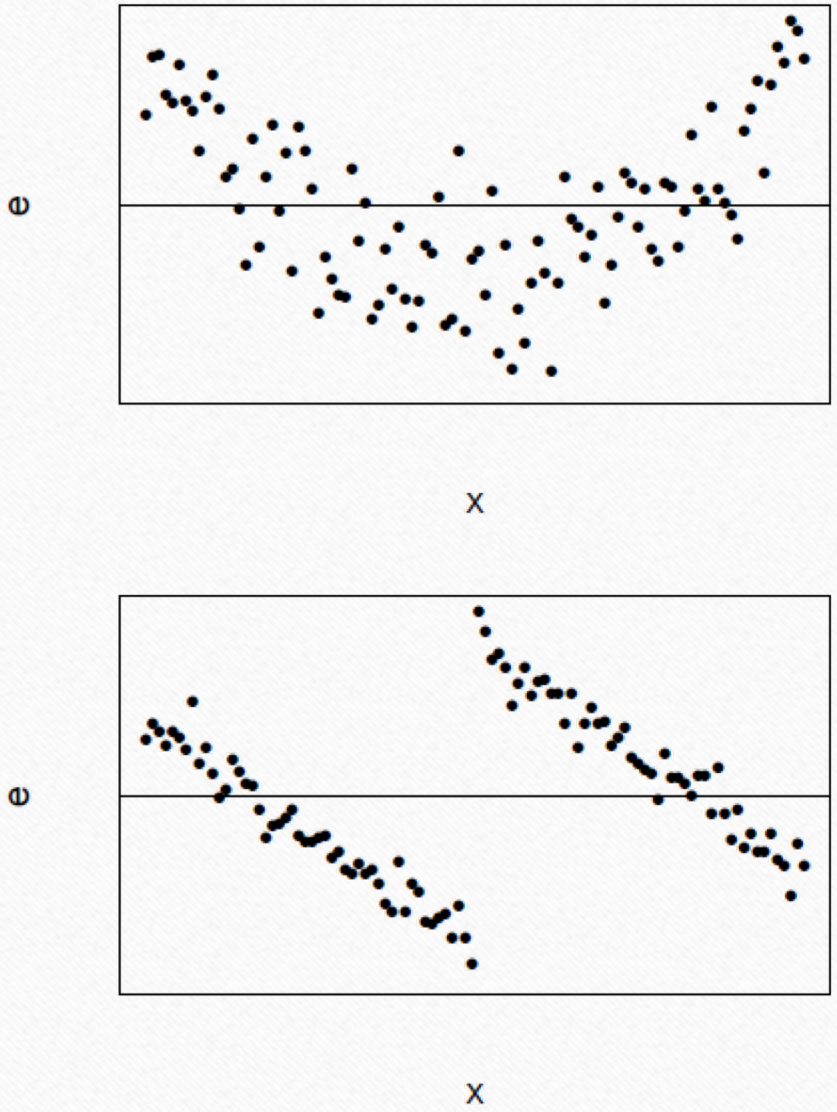
ei vs y, x1, x2, etc

  - the response,
  - each predictor variable, and
  - any other factors that might affect the relationship between the response and predictor.

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.


e
X
e
X
e
X

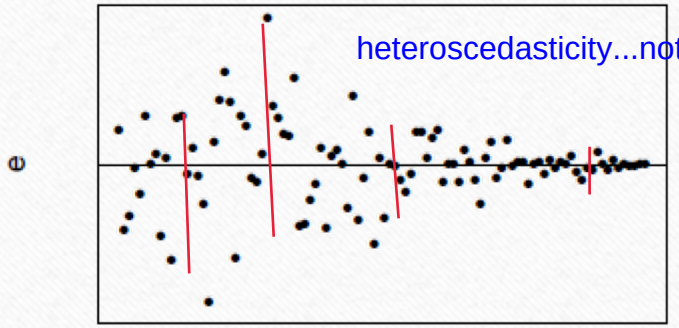
STA 674, RADOE:
Assessing Model Assumptions

Assumptions

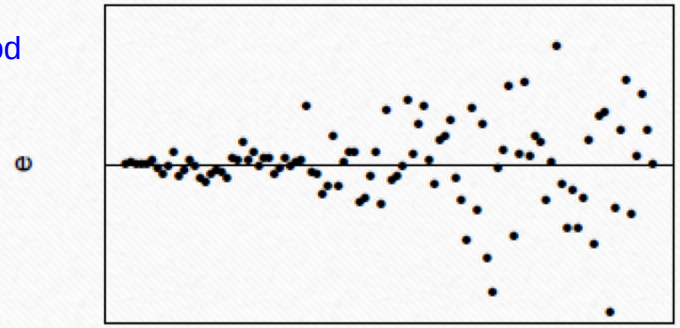
1. Mean Zero (Linearity): The average value of the errors is 0 regardless of the values of any of the predictors or of the response.
2. Homoscedasticity (Equal variance): The errors all have variance σ_e^2.

opposite is heteroscedasticity

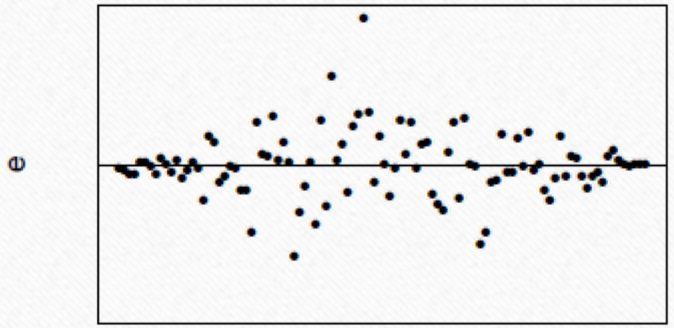
Left Funnel



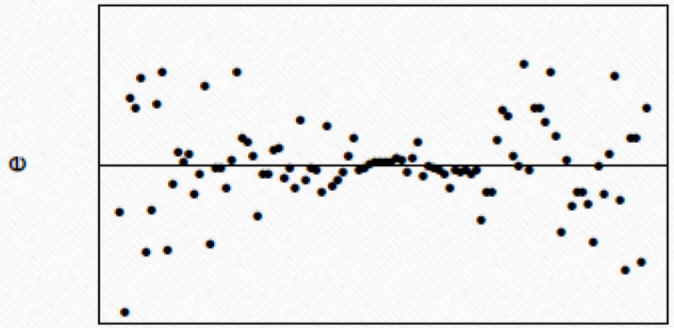
Right Funnel



Football



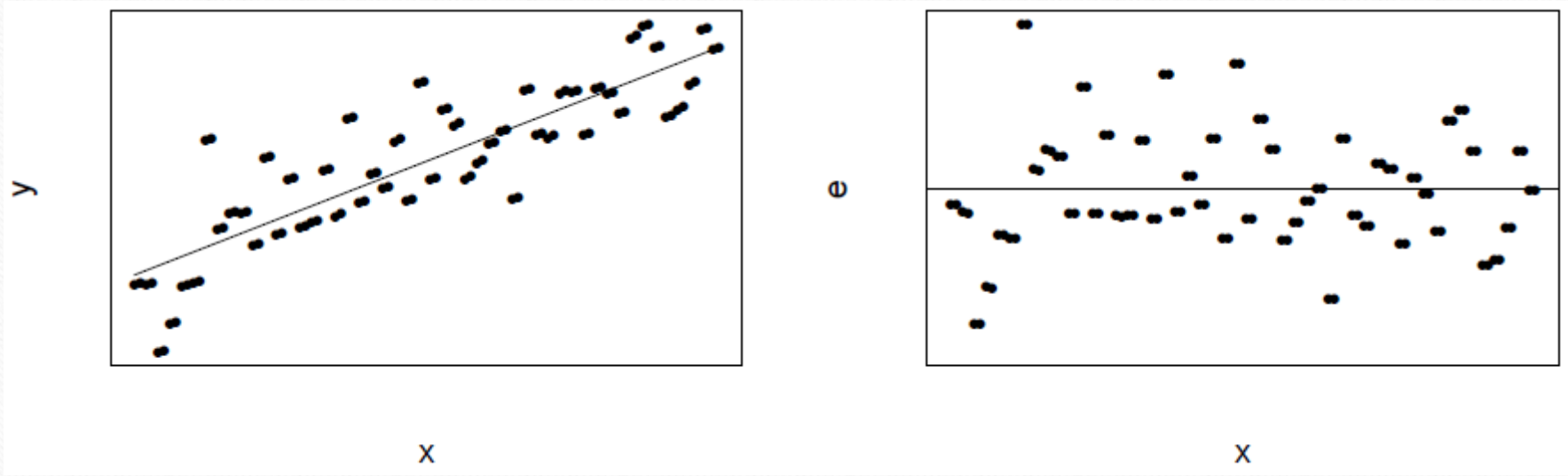
Bowtie



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.
2. Homoscedasticity (Equal variance): The errors all have variance σ_e^2.
3. Independence: The errors are uncorrelated.


y
x
e
x

STA 674, RADOE:
Assessing Model Assumptions

- Common Causes of Dependence (Correlation):
  - dependence over time
  - repeated observations on the same subject
  - replicates from batches of production
  - spatial dependence