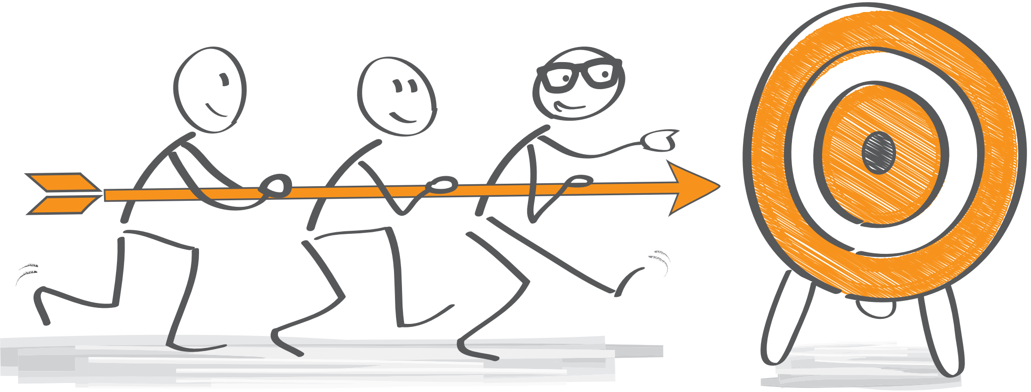
Regression Analysis
Other Regression Methods

Nicoleta Serban, Ph.D.
Associate Professor
Stewart School of Industrial and Systems Engineering

Robust Regression

About this lesson



Multiple Linear Regression

What if there are outliers?

- **If one or two, remove the outliers and fit again ⇒ Compare models with and without outliers**
- **If many, it is an indication that the normality assumption does not hold ⇒ Use an approach that provider robust estimates to outliers**

- *Linearity/Mean Zero Assumption*: $E(\varepsilon_i) = 0$
- *Constant Variance Assumption*: $\mathrm{Var}(\varepsilon_i) = \sigma^2$
- *Independence Assumption*: $\{\varepsilon_1, \ldots, \varepsilon_n\}$ are independent random variables
- *Normality Assumption*: $\varepsilon_i \sim$ Normal

Example: Departure from Normality

- Assume Y_i has a pdf given as

$$f(y|\mu,\sigma) = \frac{1}{2\sigma} e^{-|y-\mu|/\sigma}$$

  This has heavier tails than the normal distribution.

- MLE for μ: $\hat{\mu}$ to minimize $\min \sum_{i=1}^{n} |Y_i - \mu|$
  - The estimate of μ is the sample median
- Assuming $Y_i \sim f(y|\mu_i,\sigma)$ in regression analysis:
  - Estimate $(\beta_0, \beta_1, \beta_2, \ldots, \beta_p)$ by minimizing

$$\sum_{i=1}^{n} |Y_i - \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip}|$$

OLS vs Robust Regression

- **Ordinary Least Squares (OLS):** Estimate by minimizing

$$\sum_{i=1}^{n}(Y_i - \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip})^2$$

Estimate expectation: $E(Y_i | x_{i1}, x_{i2}, \ldots, x_{ip})$

- **Robust Regression**: Estimate by minimizing:

$$\sum_{i=1}^{n}|Y_i - \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip}|$$

Estimate median: $median(Y_i | x_{i1}, x_{i2}, \ldots, x_{ip})$

Why not always use Robust Regression?

Estimation Algorithm:

- Not close form expression $\Rightarrow$ Use numeric algorithm to estimate the regression parameters: Iteratively re-weighted least squares

Statistical Inference:

- The estimated variance is $\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} \widehat{w}_i r_i}{n-p-1}$
- Efficiency comes with a cost: Confidence intervals for Robust Regression are wider than for OLS

Summary