Linear Regression (cont.)

—

Multiple Linear Regression, ANOVA, Multicollinearity

COGNIBOT

AI meets Industry

Multiple Linear Regression

- Multivariate data: (xi,1,xi,2, ...,xi,m,yi) (ndata points:i= 1,...,n)
- Model

$$\hat{y}_i = a_1 x_{i,1} + a_2 x_{i,2} + \ldots + a_m x_{i,m}$$

- xi,j are the explanatory (or predictor) variables.
- yi is the response variable.
- The total squared error is

$$\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - a_1 x_{i,1} - a_2 x_{i,2} - \ldots - a_m x_{i,m})^2$$

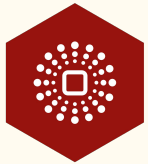
COGNIBOT
AI meets Industry

ANOVA

COGNIBOT

AI meets Industry

- Are the differences between the conditions significant?
- ANOVA tests the following hypotheses:
  - *H*o (null hypothesis): The means of all the groups are equal.
  - *H*a: Not all the means are equal

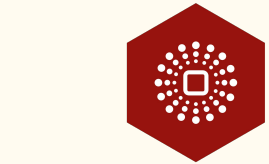


COGNIBOT

AI meets Industry

How ANOVA works

- ANOVA measures two sources of variation in the data and compares their relative sizes
  - variation BETWEEN groups
  - variation WITHIN groups

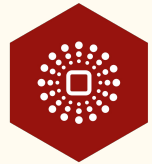


COGNIBOT

AI meets Industry

F-Score

- The ANOVA F-statistic is a ratio of the Between Group Variation divided by the Within Group Variation:
- A large F is evidence against Ho, since it indicates that there is more difference between groups than within groups.



COGNIBOT

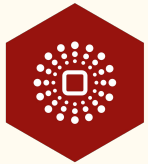
AI meets Industry

Multicollinearity

COGNIBOT

AI meets Industry

- *What does it mean?* A high degree of correlation amongst the explanatory variables
- *What are its consequences?* It may be difficult to separate out the effects of the individual regressors. Standard errors may be overestimated and t-values depressed.
- *How can you detect it?* Look at the Variance-inflation factor



COGNIBOT

AI meets Industry