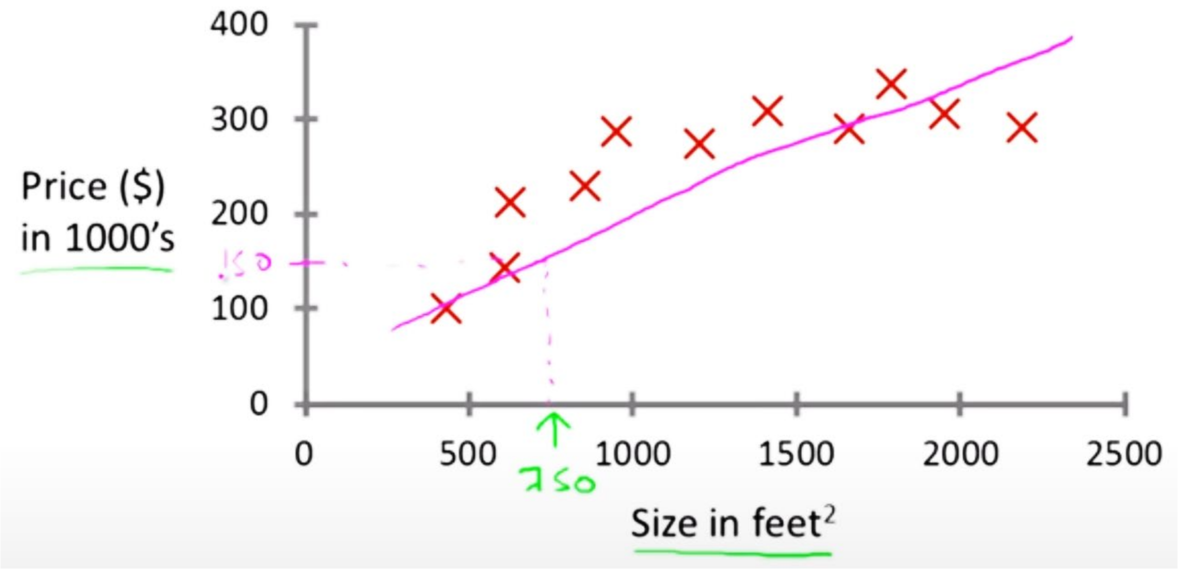
Simple Linear Regression

John Rios

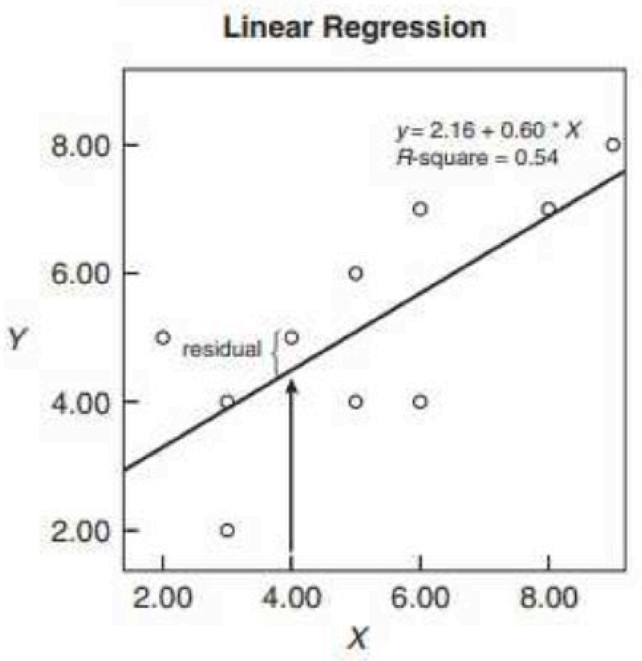
Business Intelligence

Terry College of Business
UNIVERSITY OF GEORGIA

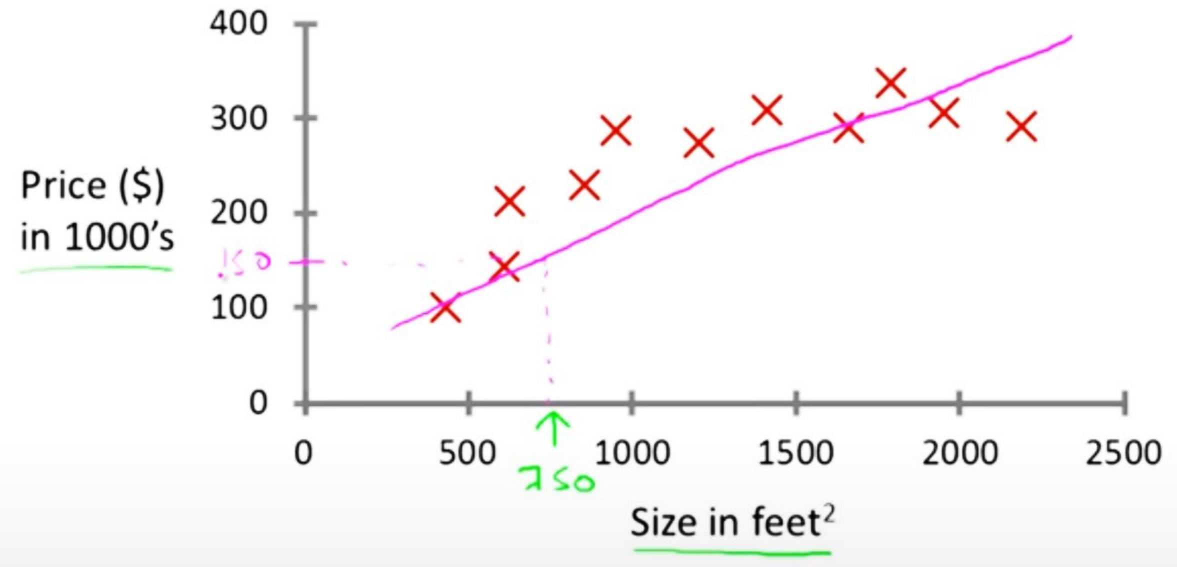


$$y = \alpha + \beta x$$

y = target (or dependent variable)
α = y-intercept
β = slope
x = predictor (or independent variable)



Linear Regression



The model function

$$y = \alpha + \beta x$$

y = target (or dependent variable)
α = y-intercept
β = slope
x = predictor (or independent variable)

Fitted line minimizes the sum or mean of the squares of the errors
Also known as ordinary least squares (OLS) regression – very popular!

Example

Find values of α and β that best fit y – try and get my predicted value of y as close as possible to the actual value of y

$$\boldsymbol{y = \alpha + \beta x}$$

y = wage
α = wage-intercept
β = slope
x = years of education

Years of Education	Wage
16	52,000
18	65,000
16	45,000
21	80,000
14	40,000
12	50,000
…	…

Interpretation

Find values of α and β that best fit y – try and get my predicted value of y as close as possible to the actual value of y

$$\boldsymbol{y = \alpha + \beta x}$$

y = wage

α = wage-intercept indicates the value of the target (wage) when all predictors are zero

β = slope indicates how much the target (wage) changes when the predictor (years of education) changes

x = years of education

R^2 = percentage of variance in the target explained by the predictors, ranges from 0 to 1

Disadvantages

High sensitivity to the data: **Erroneous** or otherwise **outlying data points** can severely skew the resultant linear function

The data may require **intensive manual manipulation** and **transformations**

Regression Analysis – Two Approaches

- Explanation (Inference) – Traditional Use

Determine which predictors are the most useful for estimating the outcome variable.

Determine the amount of variance in the outcome variable that is explained by the predictor variable(s).

- Prediction - Machine Learning

Predict values of the outcome variable from values of the predictor variable.

Traditional Use

Explanatory modeling

- The goal is to explain the relationship between predictors (independent variables) and target (dependent variable)

Model Evaluation

Fit the data well and understand the amount of variance explained as well as the statistical significance of each predictor

- Evaluation (of goodness of fit) involves the use of (Adjusted) R-squared and p-values

Significance

What does it mean for a predictor to be statistically significant?

- A measure of the probability that the observed effect was due to random chance.
- In practice, a p-value < 0.05.
- A $p < 0.05$ means we are 95% confident the result is not a mistake (i.e., not driven by randomness)

Significance

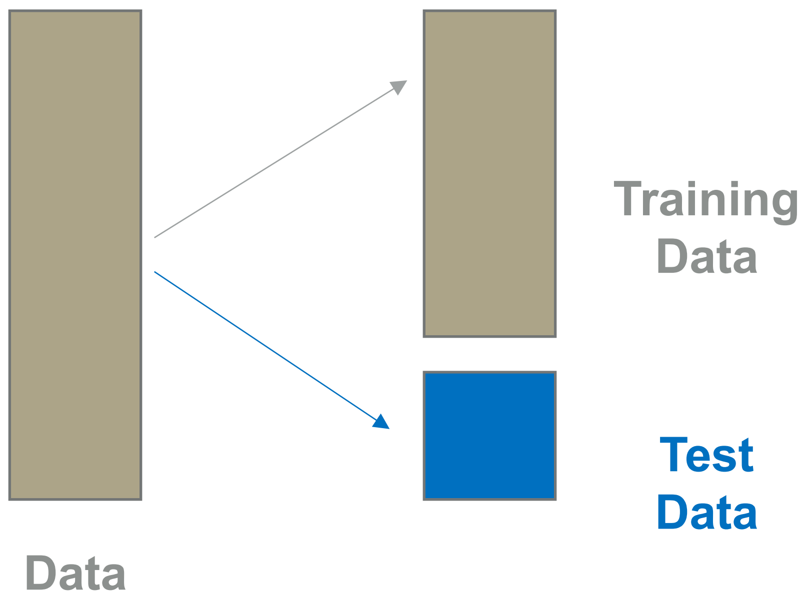
What does it mean for a predictor to be statistically significant?

- Can also look at confidence intervals

- You can claim statistical significance (i.e., reject the null hypothesis) when the CI does not include zero.

Machine Learning Use

- Predictive modeling
- Evaluate based on prediction error



Train and Test the Model