Phase 2 Expectations:

Statistics

// FLATIRON SCHOOL

Agenda

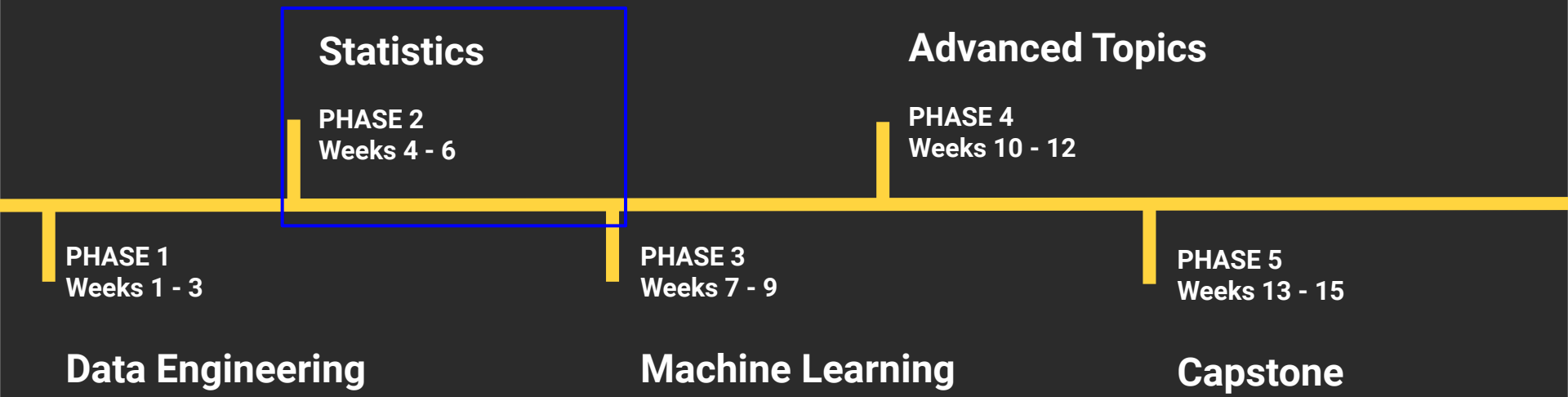
1. Phase 2 Overview
2. Week 1: SQL & Probability
3. Week 2: Hypothesis Testing & Simple Linear Regression
4. Gating & Assessments
5. Questions/Feedback

Date or other relevant info

Phase 2: Overview

// FLATIRON SCHOOL

Overall Phase Timeline



Josh Wills
@josh_wills

Follow

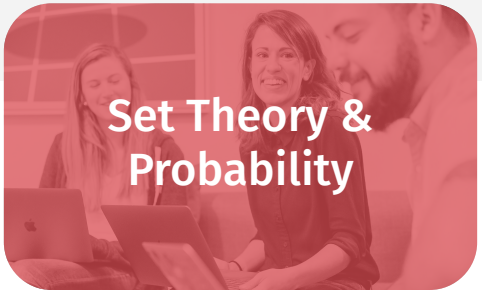
Data Scientist (n.): Person who is better at statistics than any software engineer and better at software engineering than any statistician.

Reply Retweet Favorite More

9:55 AM - 3 May 12

Statistics for Data Science

- Common data storage method
- Relational databases
- New way to access information
- Critical skill
- Integrate with python - pandas



- Unions, intersections, joins
- Uncertainty, confidence
- Predictive probability
- Conditional probability
- Distributions

- Hypothesis testings: Z-test, t-test, ANOVA, Chi-Squared
- Using distributions
- Beginning of data modeling
- Simple Linear regressions
- Quantifying relationships

Probability

Probability of Simple Events

$$P(A) = \frac{n}{N} = \frac{\#\,outcomes\ in\ A}{\#\,outcomes\ in\ Sample\ Space}$$

Probability

Multiplication Rule

Independent Events

$$P(X \cap Y) = P(X) \cdot P(Y)$$

Dependent Events

$$P(X \cap Y) = P(Y) \cdot P(X \mid Y)$$

Bayes' Theorem

$$P(X \mid Y) = \frac{P(X \cap Y)}{P(Y)}$$

Probability of Compound Events

Independent Events

$$P(A \text{ and } B) = P(A) \times P(B)$$

Dependent Events

$$P(A \text{ and } B) = P(A) \times P(B \mid A)$$

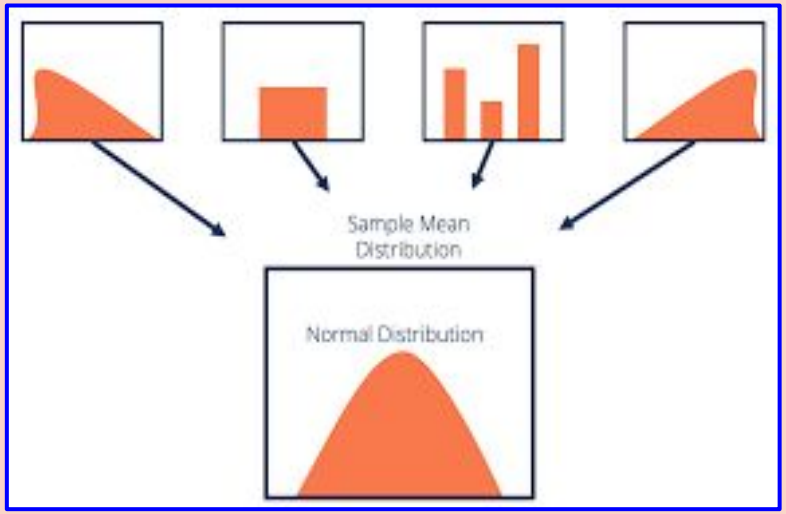
Mutually Exclusive

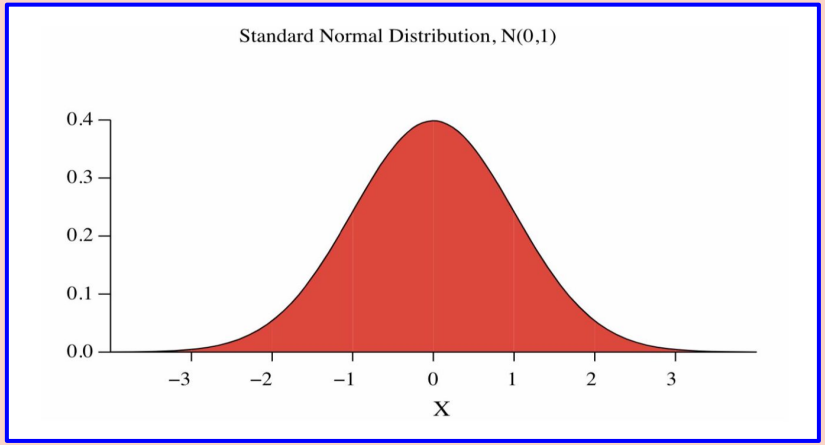
$$P(A \text{ or } B) = P(A) + P(B)$$

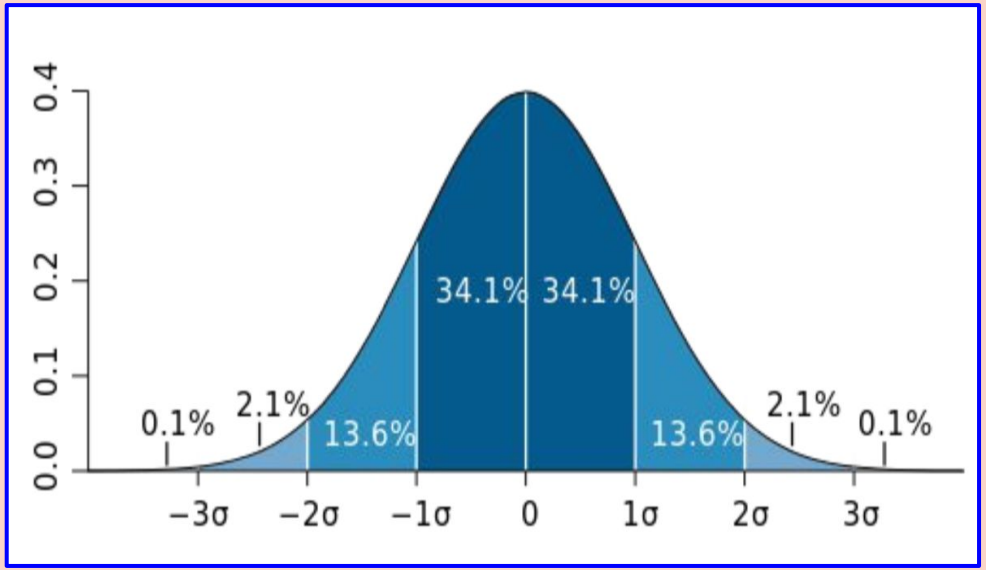
Mutually Inclusive

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

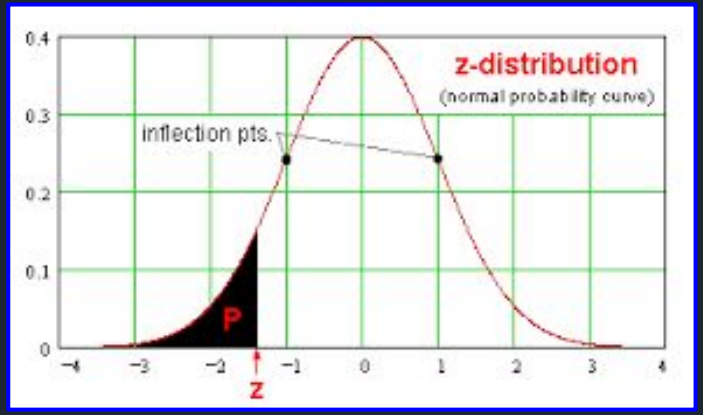
Distributions

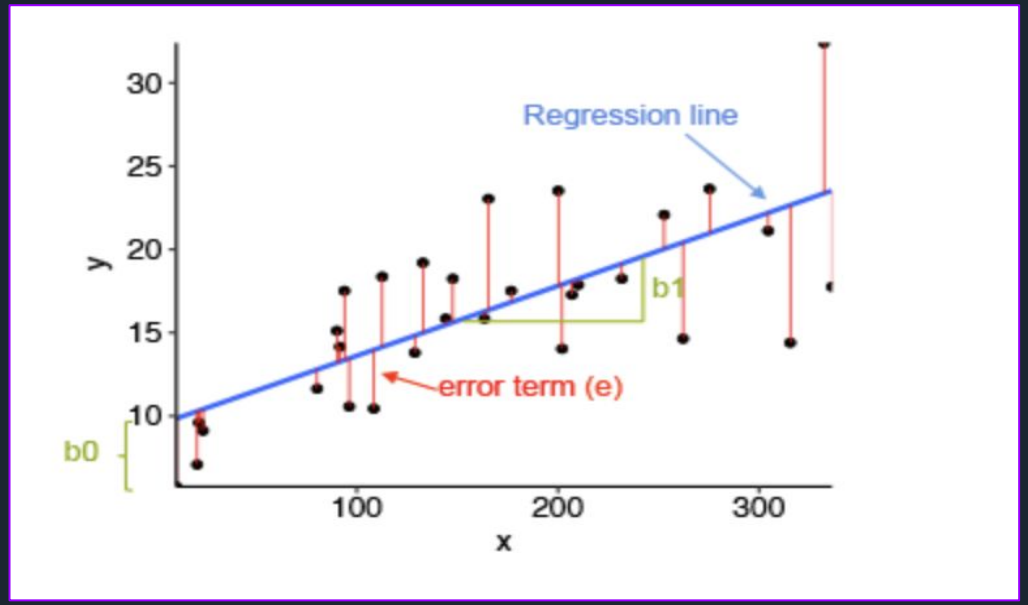






Hypothesis Testing





Measuring Student Progress

CP & CC

- 11/22 - Wed - SQL Checkpoint
- 11/27 - Thurs - Probability Checkpoint
- 11/29 - Tues - Hypothesis Testing Checkpoint
- 12/1 - Friday - Phase 2 Code Challenge

Blog Post

- Due Tuesday of 2nd Week
- 11/28
- Topic: Describe a DS tool or library with tutorial
- See canvas

Project

- Group Project
- Third week
- Movie data
- SQL database
- Tiny bit of Linear Regression

1:1

Link next week

- Phase 1 review
  - 1:1 Meet
- Project review
  - Group