Introduction to Data Science and Analytics

Levels of Measurement

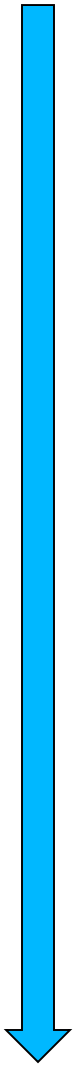
DATA SCIENCE
& ANALYTICS

Varieties of Data and Measurements

- Type of measurement determines the statistical techniques that are applicable
- For instance:
  - Which types of measurements are suitable in a regression analysis
    - for the independent / predictor variables?
    - for the dependent / predicted variables?
  - What about data used in classification?

Varieties of Data and Measurements

- Nominal measurements
  - Classes or factors of a data set
- Ordinal measurements
  - Ordered or ranked measures of a data set
- Interval measurements
  - Difference measured from end points
- Ratio measurements
  - Continuous magnitude compared to a unit magnitude



DATA SCIENCE
& ANALYTICS

Varieties of Data and Measurements

- Levels of measurements – lowest precision
- Example: Temperature
  - When you walked into work, was the temperature:
    - Cold ?
    - Cool ?
    - Warm ?
    - Hot ?

These are *Nominal* measures

DATA SCIENCE & ANALYTICS

Varieties of Data and Measurements

- Levels of measurements – lower but not lowest precision
- Example: Temperature
  - When you walked into work, was the temperature:
    - Coldest day of the year?
    - Hottest day of the year?
    - 5th coldest day of the year?

    These are *Ordinal* measures

DATA SCIENCE & ANALYTICS

Varieties of Data and Measurements

- Levels of measurements – better but not best precision
- Example: Temperature
  - When you walked into work, was the temperature:
    - 20° F ?
    - 35° F ?
    - 99° F ?

These are *Interval* measures

(e.g. relative to 0°F end point)

DATA SCIENCE & ANALYTICS

Varieties of Data and Measurements

- Levels of measurements – better but not best precision
- Example: Temperature
  - Is 98° F twice (2X) as hot as 49° F degrees?
  - If you say yes ….
    - Then what is twice cold as 32° F ?
    - Where does 0° F fit in ? -10° F & -20° F?

There is a lack of precision in the °F measurement is because the 0° reference point is arbitrary

DATA SCIENCE & ANALYTICS

Varieties of Data and Measurements

- Central tendency (e.g. mean) and dispersion (e.g. standard deviation) are descriptive statistics
- How central tendency is measured varies per the measurement level

Central Tendencies

- Nominal
  - The most common value, i.e., mode
- Ordinal
  - The middle rank, i.e., median
  - Other percentile/quantile
  - Also mode

DATA SCIENCE
& ANALYTICS

Central Tendencies

- Interval
  - Arithmetic mean, i.e., average
  - Also mode, median, and percentile/quantile

- Ratio
  - Mode, median, percentile/quantile, and arithmetic mean
  - Additional measures: geometric and harmonic means

Continuous vs. Discrete Variables

- Nominal and ordinal are discrete
  - Values come from a *limited set of values*
- Interval and ratio are continuous
  - Values have *infinite possible values*
- These variable types define suitable models for inferential statistics
  - Example : Regression vs. Classification

Inferring Continuous Variables

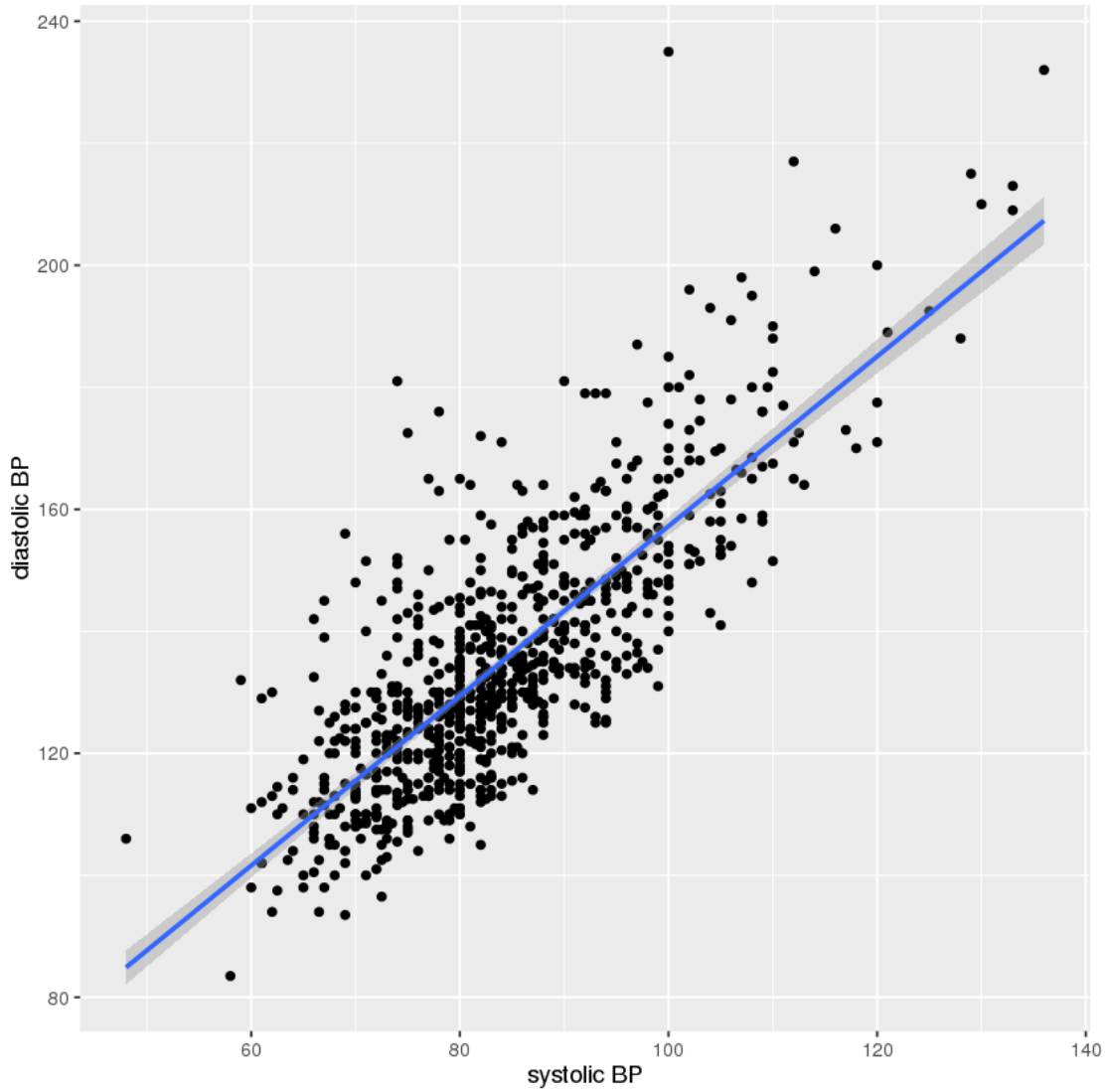
- Interval and ratio measurements are typically predicted through inferential statistics using a form of *regression analysis*
- A formula of independent (predictor) variables (x-axis) and model coefficients compute a dependent (predicted) value (y-axis)

Inferring Continuous Variables

Regression Plot Example

Understanding Variable Relationships

Example: Blood Pressure Measurements



DATA SCIENCE & ANALYTICS

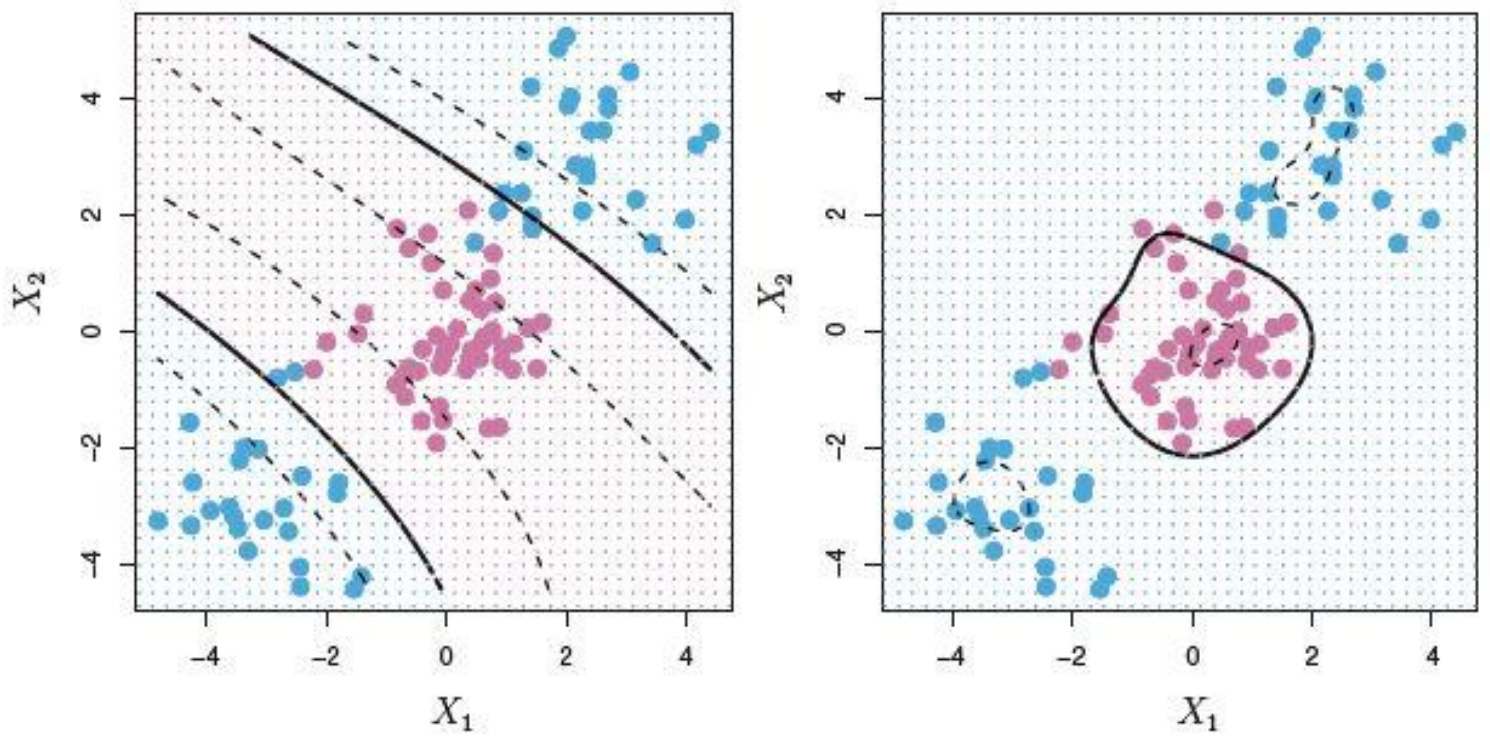
Inferring Discrete Variables

- Nominal measurements are typically predicted through inferential statistics using *classification techniques* (e.g. neural networks)
- A formula of independent (predictor) variables and model coefficients determine a dependent (predicted) value based on decision planes derived from statistical models

Inferring Nominal Variables

There are numerous ways to predict nominal values given input measurements

Classification examples from 2D scatter plot with decision boundaries



DATA SCIENCE & ANALYTICS

Introduction to Data Science and Analytics

Levels of Measurement

DATA SCIENCE & ANALYTICS