

DSFA

Spring 2019

Lecture 23

Correlation

Announcements

- HW6 due today
- Project 3 due Wednesday, May 8 (last day of classes)

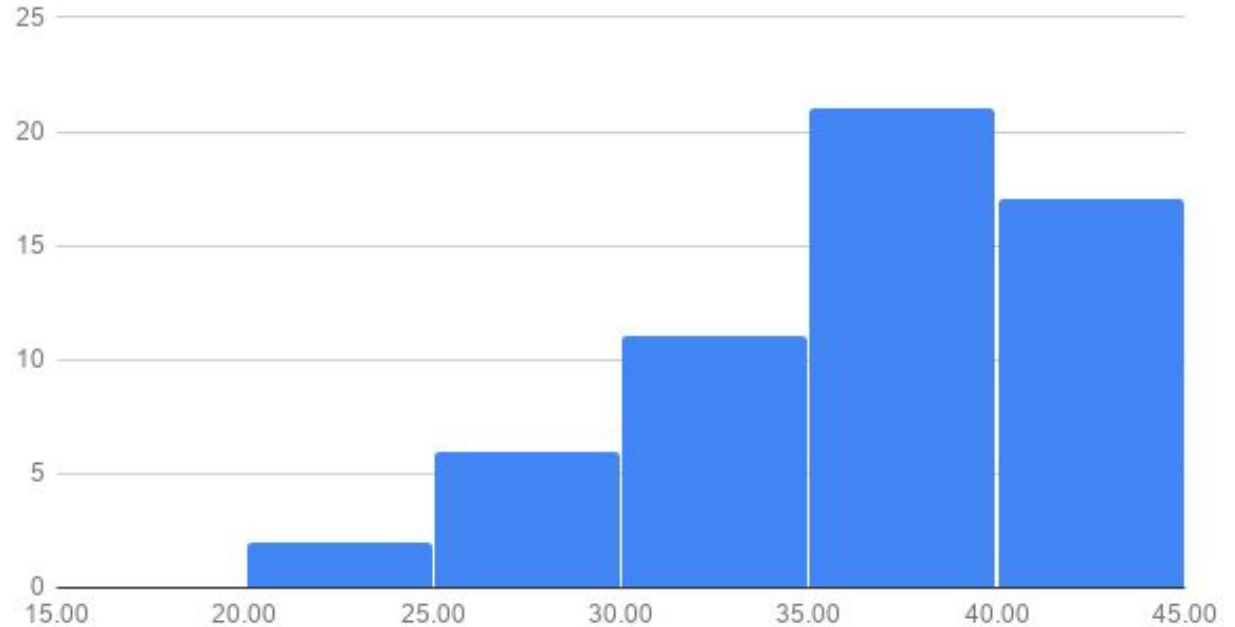
Prelim 2 Scores

N = 57

Mean 36.5 (81%)

Median 38

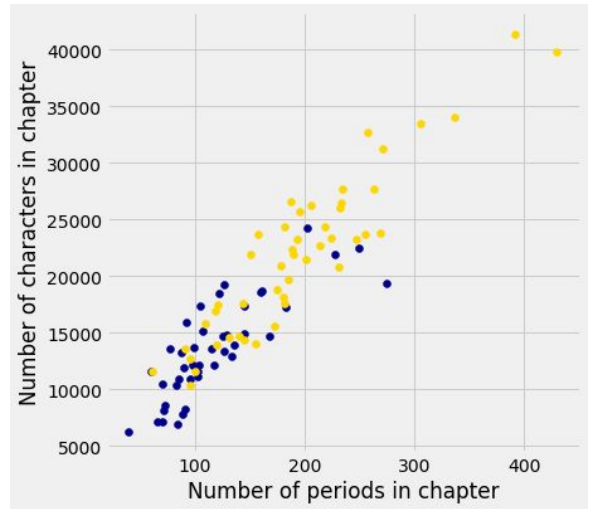
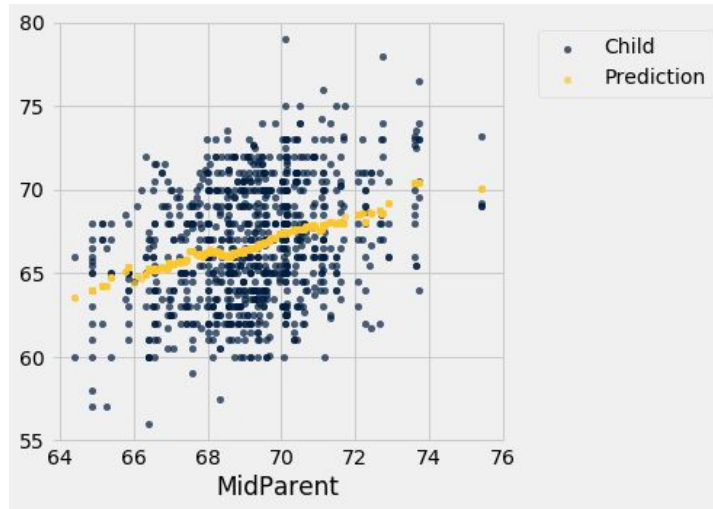
Std.Dev. 5.5



Prediction

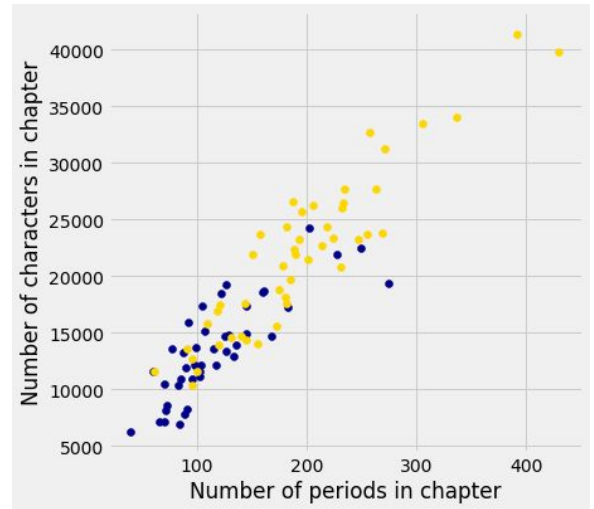
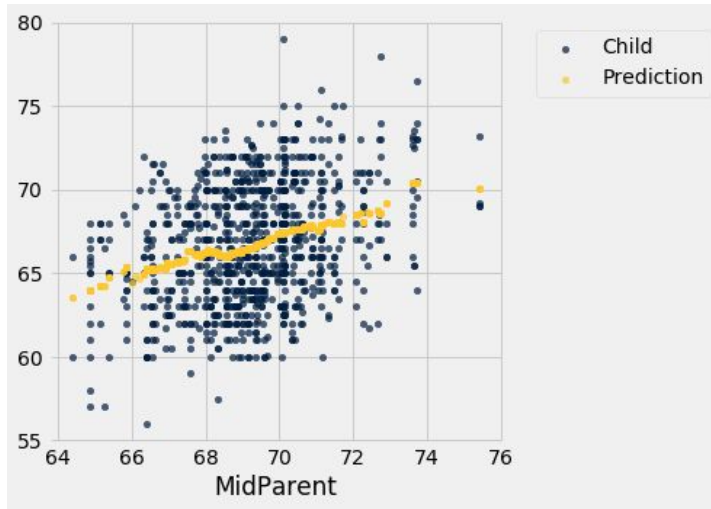
- Guess outcomes in the future, based on available data
- Our simple goal: predict value of one variable based on another

(Demo)



Prediction

If we have a line describing the relation between two variables, we can make predictions



Relation Between Two Variables

Visualize then quantify

- Any discernible pattern?
- Simplest kind of pattern: Linear? Non-linear?

(Demo)

The Correlation Coefficient r

- Developed by Karl Pearson (1857-1936) based on work of Francis Galton (1822-1911)
 - Measures linear association
 - $-1 \leq r \leq 1$
 - $r = 1$: scatter is perfect straight line sloping up
 - $r = -1$: scatter is perfect straight line sloping down
 - $r = 0$: No linear association; *uncorrelated*
(Demo)
-

Definition of r

Correlation Coefficient (r) =

average of	(array) product of	x in standard units	and	y in standard units
---------------	-----------------------	---------------------------	-----	---------------------------

Measures how clustered the scatter is around a straight line

Classification

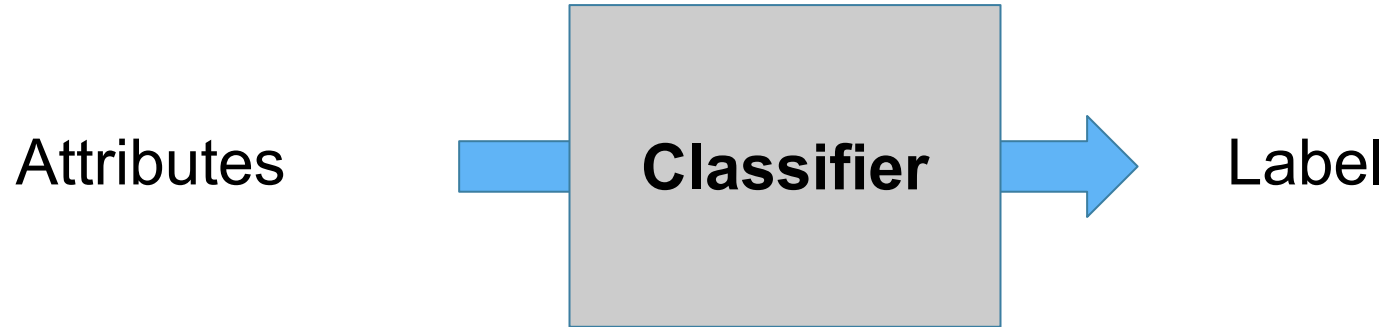
- Our study of **regression/correlation**:
 - One quantitative variable (x)
 - Predicts another quantitative variable (y)

 - Now, **classification**:
 - Many quantitative variables
 - Predict a **categorical** variable
-

Classification Terminology

- **Response variable:** the categorical variable we try to classify
 - **Classes or labels:** possible values of response variable
 - **Binary response:** 0 or 1
 - **Attributes:** variables used to make classification
-

Classifier



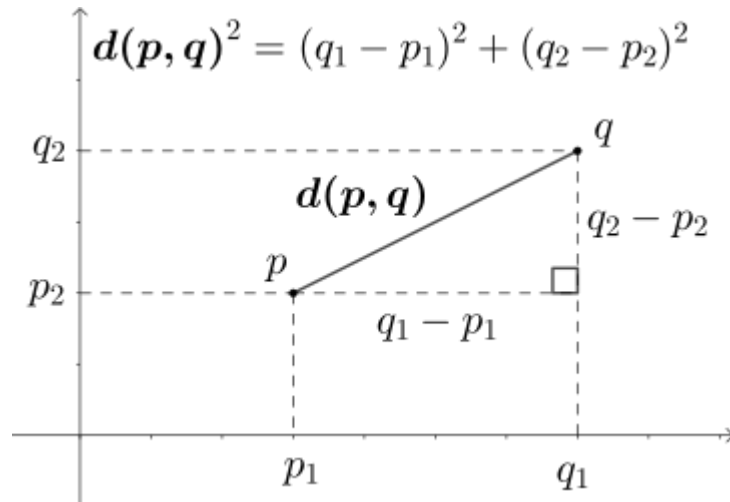
(Demo)

Nearest Neighbor

How to classify a new individual:

- Find their **nearest neighbor**: the individual closest to them in the data set
 - Assign the new individual the **same** label as that nearest neighbor
-

Distance



(Demo)
