

DSFA

Spring 2018

Lecture 29

Correlation

Announcements

- Prelim 2: Tue, 7:30 pm, Kimball B11

New course!

ORIE 2380, Urban Analytics, Fall 2018

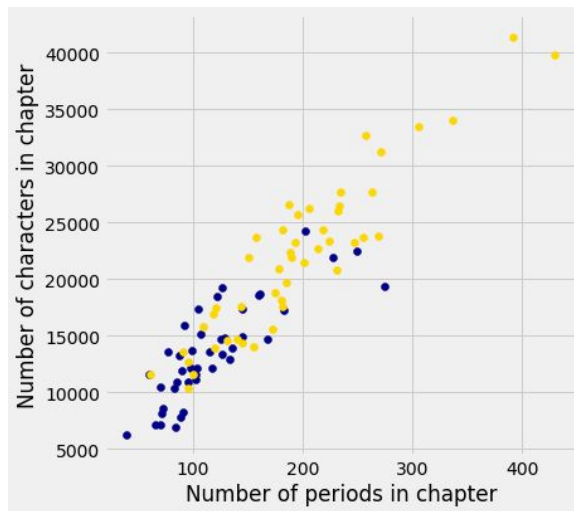
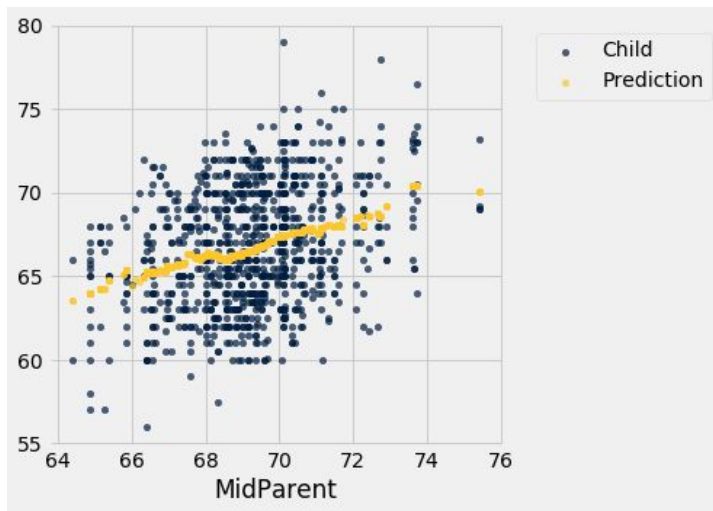
Data documenting urban life is being collected at a scale unimaginable just a few years ago. And yet, the use of urban data to improve our quality of life has its roots in stemming the 1854 cholera outbreak in London. Through a series of case studies, this course will explore a number of facets of the use of urban data in understanding, forecasting, and decision-making in our lives today, starting with this example in epidemiology, but continuing with modern transit (including ride- and bike-sharing), emergency services, criminology, and environmental monitoring/planning. The course will emphasize the role of mathematical, statistical and computational models in these settings.

Prereq: ORIE 1380 or equivalent

Prediction

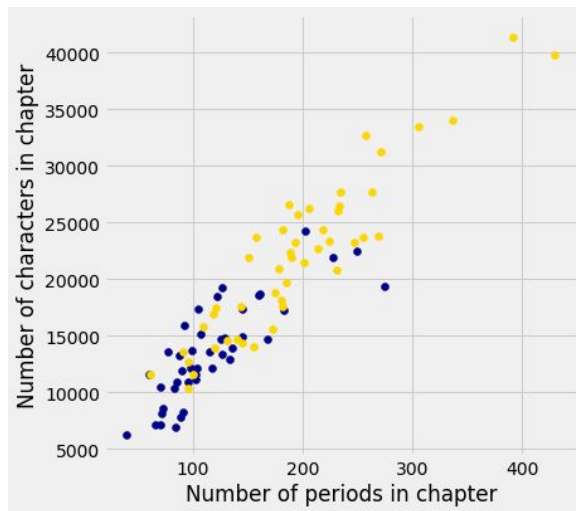
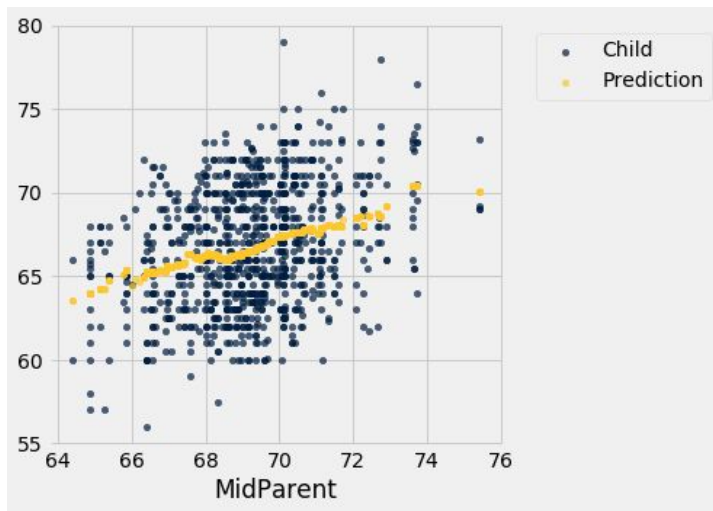
- Guess outcomes in the future, based on available data
- Our simple goal: predict value of one variable based on another

(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
