

# Thinking Statistically

# Ask Questions

-- What's the source of the data?

All of the possible units (the “universe”)

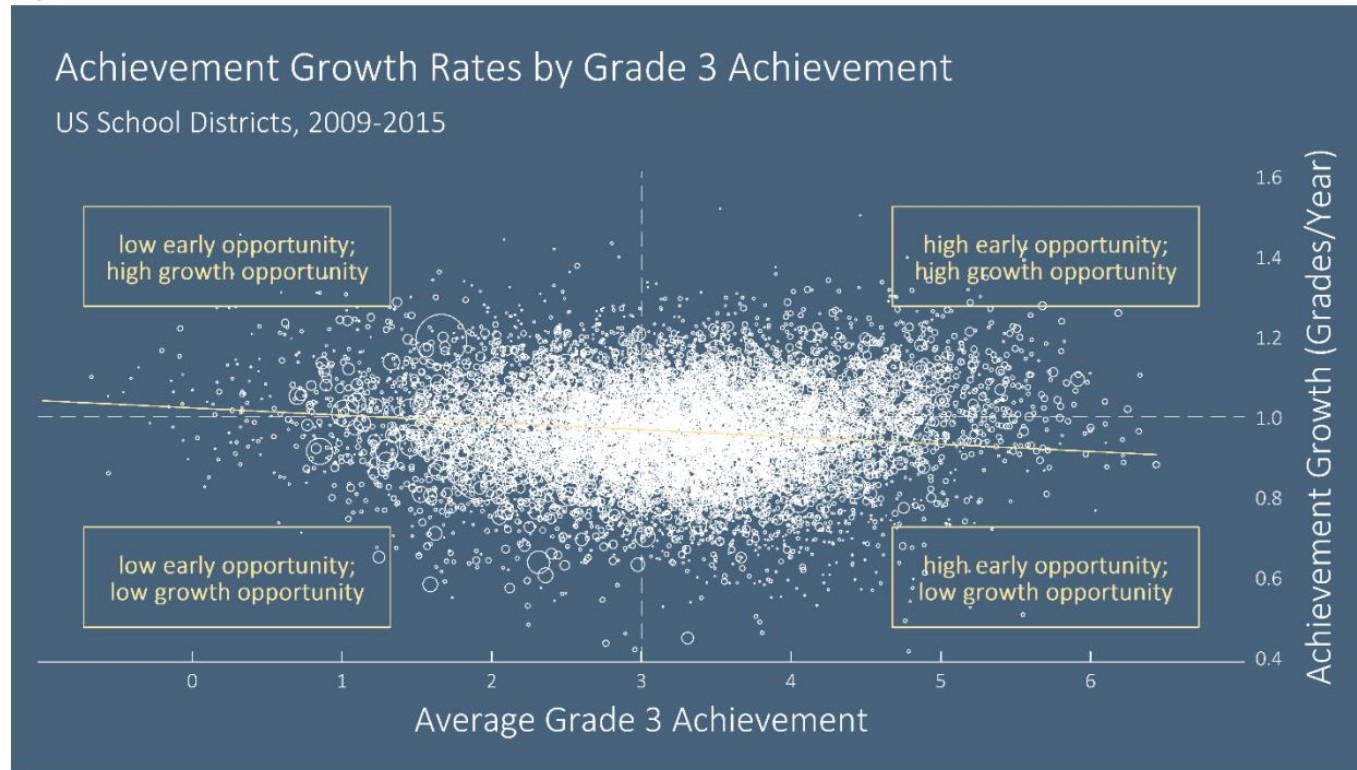
Some of the possible units, chosen for convenience

A random selection of the possible units

Some combination of the above

# The “Universe” of Data

Figure 4:



Source: Author's tabulations, Stanford Education Data Archive (Reardon et al. 2017a).

<https://>

# Ask Questions

What are we trying to do with the data?

Just describe what's in the data we have (pure descriptive)?

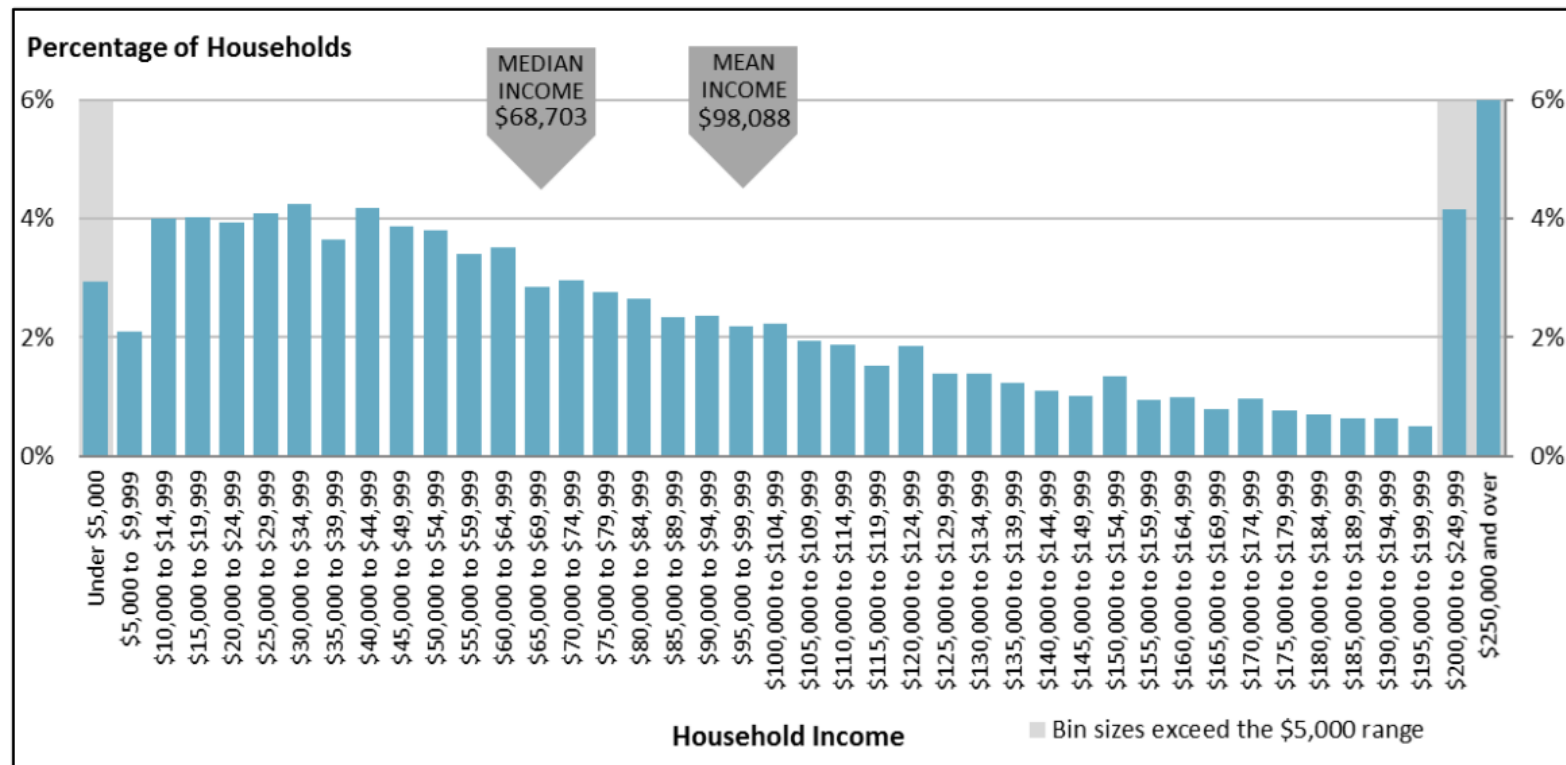
Use the data we have to describe a larger group (inferential)?

Use *relationships* in the data to make *predictions* about data we haven't seen yet (predictive)

# Variation in All Things

- There are two numbers we always *must* know
- What is the measure of central tendency-- what's the most likely number (mean, median mode)
- What is the measure of variation? (Standard deviation, range, interquartile range)
- Don't trust any analysis that doesn't tell or show you both

**Figure I. Distribution of Household Income, 2019**



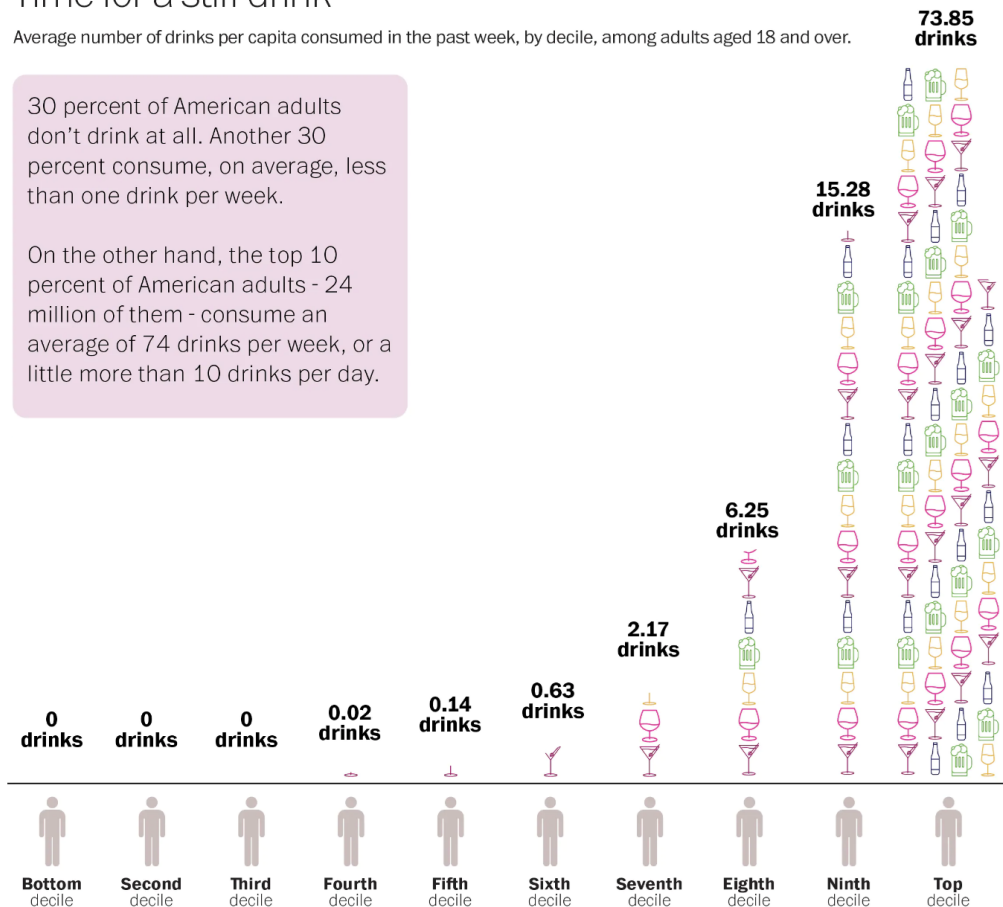
**Source:** U.S. Census Bureau, *Annual Social and Economic Supplement*, available at <https://www.census.gov/data/tables/time-series/demo/income-poverty/cps-hinc/hinc-06.html>.

# Time for a stiff drink

Average number of drinks per capita consumed in the past week, by decile, among adults aged 18 and over.

30 percent of American adults don't drink at all. Another 30 percent consume, on average, less than one drink per week.

On the other hand, the top 10 percent of American adults - 24 million of them - consume an average of 74 drinks per week, or a little more than 10 drinks per day.



US average is about 17

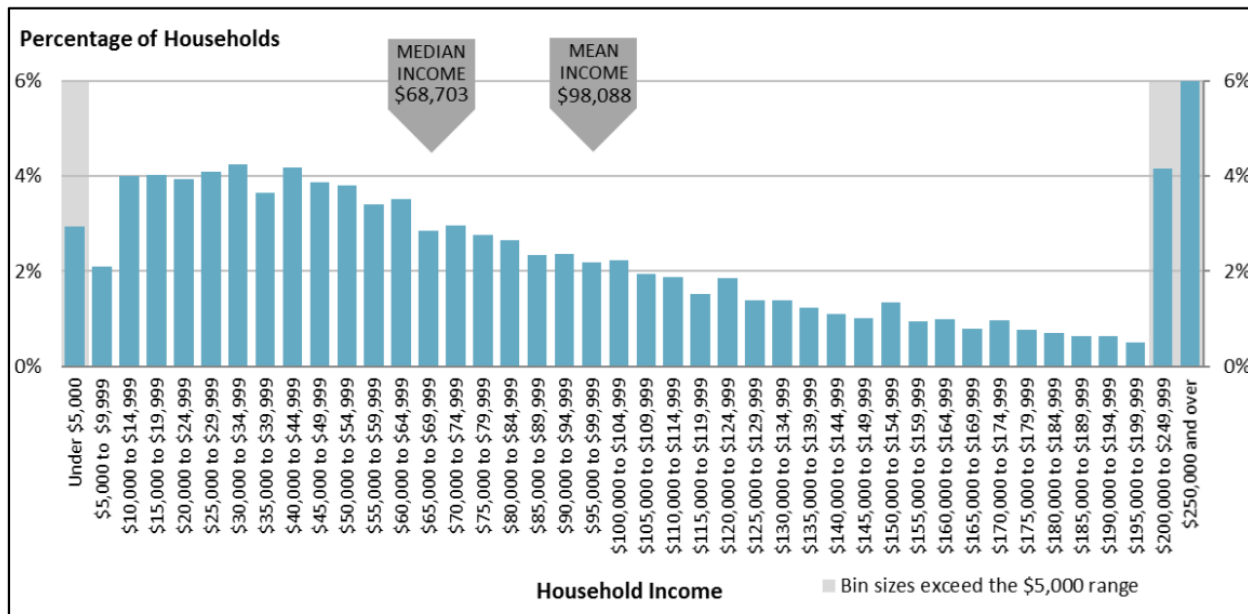
# Probability

- Probability is an abstract representation of the world that results in purely mathematical properties
- We can use probability to tell us what the world might look like *IF OUR ASSUMPTIONS ARE MET*
- Sometimes these assumptions are reasonable: coins really do tend to have a .5 probability of heads or tails
- Sometimes these assumptions are most definitely NOT reasonable



# Assumption vs. Reality

Figure 1. Distribution of Household Income, 2019



**Source:** U.S. Census Bureau, *Annual Social and Economic Supplement*, available at <https://www.census.gov/data/tables/time-series/demo/income-poverty/cps-hinc/hinc-06.html>.

# Descriptive Statistics

- In descriptive statistics we're simply summarizing the data we have on hand
- Many times this is enough, particularly if our data cover or mostly cover the universe of units
- It's crucially important that we be clear about what we're up to if we're doing descriptive stats.

# Inferential Statistics

- In inferential statistics we're using the data from our sample to describe the larger world
- The data in our sample are connected to the world through the probability of any given unit being sampled
- Once we know that, we can talk about the extent to which our sample represents the broader world

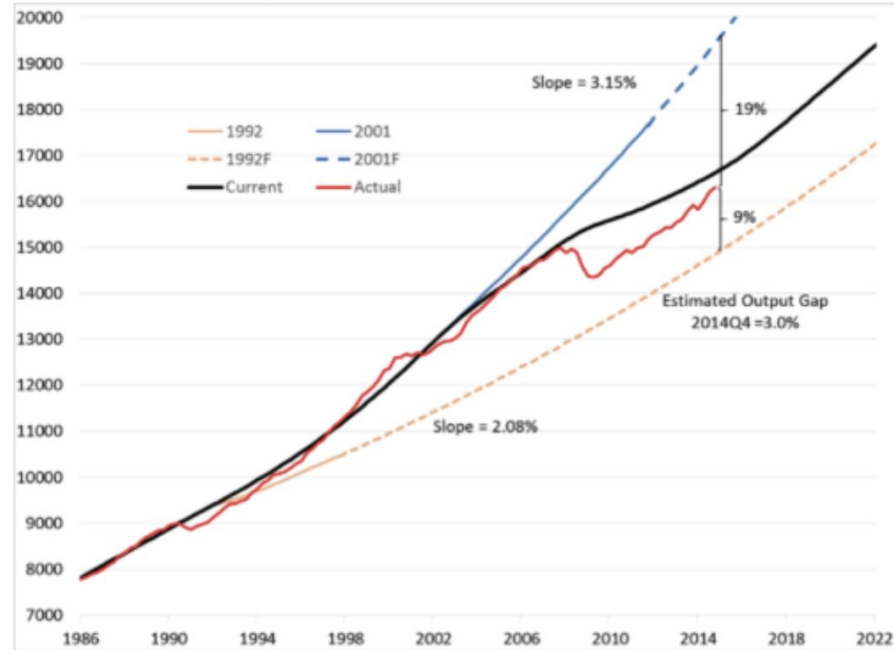
# Predictive Analysis

In predictive analysis we're trying to say what we think will happen in data we haven't seen yet based on relationships in data we have on hand

Data can be from the universe of data OR a sample

Many, many times (but not always) predictive analysis is about predicting what will happen in the future

Chart 2: Selected Vintages of U.S. Real Potential GDP with forecasts



Source: CBO, Federal Reserve Bank of St. Louis (ALFRED) and authors' calculations.

**Stephen G. Cecchetti and Kermit L. Schoenholtz:** [\*Forecasting Trend Growth: Living with Uncertainty\*](#):

“We should all be wary of anyone...