The Data Says What?!

Applying Data Analysis to a Complicated World

I. Olympic Medal Counts: Who "WON" the Olympics?

The Olympic Games are a source of national pride and, of course, bragging rights. The medal counts are regularly worked into such conversations:

Country	Gold	Silver	Bronze	Total Metals
United States	9	15	13	37
Germany	10	13	7	30
Canada	14	7	5	26
Norway	9	8	6	23

<u>Country Data:</u>

United States		
Number of Athletes Participating	214	
2006 Olympic Medal Count	Gold	9
	Silver	9
	Bronze	7
	Total	25
Population	296,000,000	
Land Area (square miles)	3,718,000	

Germany		
Number of Athletes Participating	152	
2006 Olympic Medal Count	Gold	11
	Silver	12
	Bronze	6
	Total	29
Population	82,000,000	
Land Area (square miles)	138,000	

Canada		
Number of Athletes Participating	205	
2006 Olympic Medal Count	Gold	7
	Silver	10
	Bronze	7
	Total	24
Population	33,000,000	
Land Area (square miles)	3,852,000	

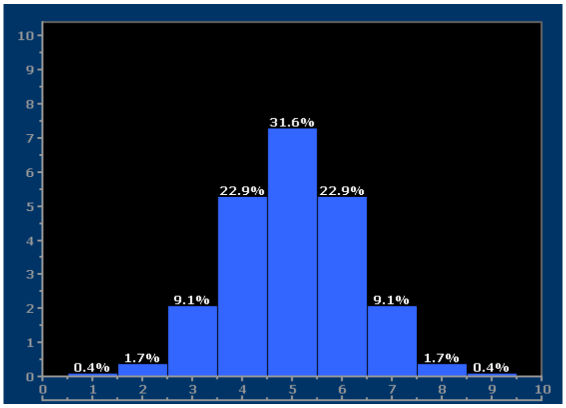
Norway		
Number of Athletes Participating	100	
2006 Olympic Medal Count	Gold	2
	Silver	8
	Bronze	9
	Total	19
Population	4,593,000	
Land Area (square miles)	125,000	

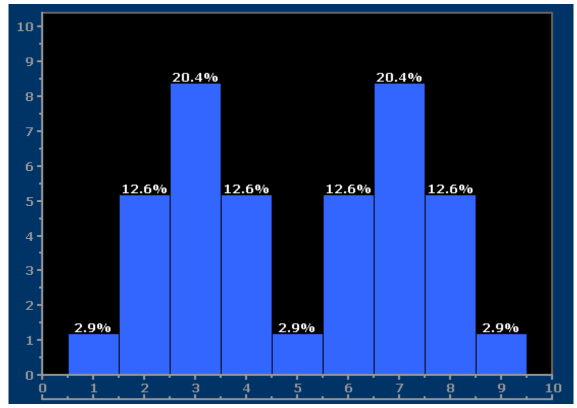
The Data Says What?!

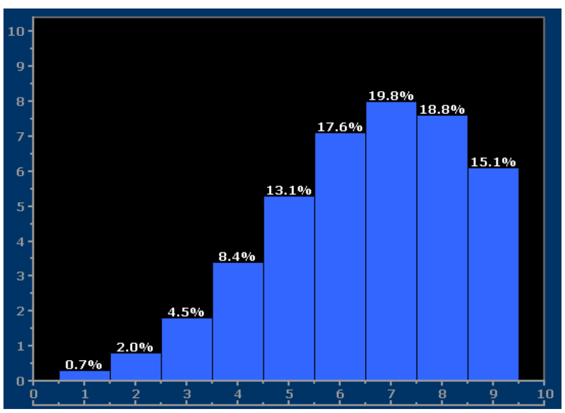
Applying Data Analysis to a Complicated World

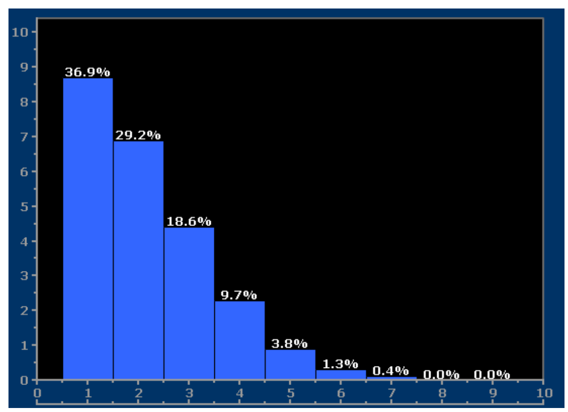
II. What is average anyways?

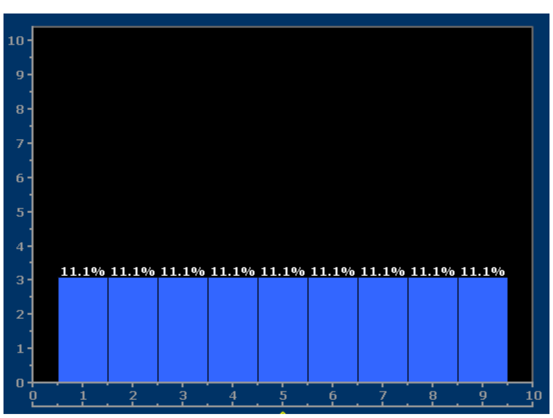
Consider some distributions from a 10 point quiz. Where is the "average" student?

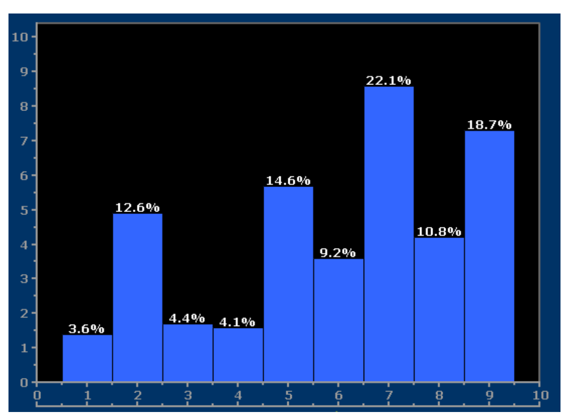












The Data Says What?!

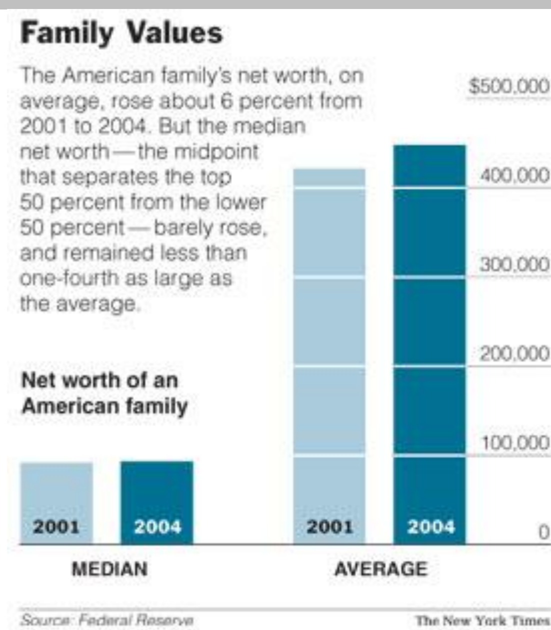
Applying Data Analysis to a Complicated World

"If you put one foot in a tub of hot water and the other in a tub of cold water and take the average, everything is fine."

Mark Zandi
Chief Economist, economy.com

How does the American populous react to such averages?

Economic Agragates:		
Inflation Rate for 2009 (Consumer Price Index)	-0.4%	Nine months of 2009 had deflation: July 2009 had 2.1% deflation
National Unemployment	9.7%	Down from 10% during Oct. – Dec.
Access to Employer Assistance Child Care	16%	Up from 4% in 2000
"Average" sleep per evening	8.6 hrs	Up from 8.55 hrs
"Average" time spent caring children	1.86 hrs	Down from 1.94 hrs in 2004
"Average" family net worth between 2001 and 2004	Rose 6%	



Sources:
Bureau of Labor Statistics
Federal Reserve

The Data Says What?!

Applying Data Analysis to a Complicated World

An excerpt from a June 2006 New York Times Article examines this phenomenon further:

> "In the first quarter of 2006, the U.S. economy grew at an annual rate of 5.3 percent, the fastest growth in two and a half years," George W. Bush said, as Mr. Paulson, the chief executive of Goldman Sachs, looked on. "We added 5.2 million new jobs since August of 2003. The national unemployment rate is down to 4.7 percent. Productivity is high, and that's leading to higher wages and a higher standard of living for the American people."
>
> Yet in the latest New York Times/CBS News poll, only 28 percent of the respondents said they approved of President Bush's handling of the economy, while 66 percent disapproved — the worst such ratings of his presidency. Only 6 percent rated the economy as very good, while 46 percent said it was fairly bad or very bad. And consumer confidence plummeted last month, according to the Conference Board.
>
> This strange and unlikely combination — strong and healthy aggregate macroeconomic indicators and a grumpy populace — has been a source of befuddlement to the administration and its allies. Gas prices and the Iraq war have surely contributed to this disconnect. But a lesser-known factor is also at work: the misleading aggregates.
>
> Aggregates — big-picture figures like the unemployment rate, productivity and growth in the gross domestic product — are highly useful to economists. But to most people, they're abstractions. You can't use a low unemployment rate to pay a mortgage.
>
> As a result, large aggregates "are something that people may hear about in the news, but don't have a direct impact on how people feel," said Lynn Franco, director of the Consumer Research Survey at the Conference Board.
>
> Aggregates generally are averages, which are of declining utility in an economy characterized by greater inequality of income and assets. In an interview with The Wall Street Journal in March, Mr. Snow took pains to point out that there had been substantial gains in per-capita income (8.2 percent, after inflation) and net worth (24 percent, before inflation) from the beginning of 2001 to the end of 2005.
>
> The data he cited were averages, or means, and that can be misleading. "The average wage is a useful indicator if you want to know what's happening to the tax base, but it might not tell you what's going on for the individual worker," said Alan B. Krueger, an economics professor at Princeton and a former chief economist at the Labor Department.
>
> To see how typical workers are doing, it's better to look at median wages and incomes — the midpoint that separates the top 50 percent from the lower 50 percent. And median income, which was stagnant during President Bush's first term, is struggling to keep pace with inflation. "Median household income has gone nowhere since the turn of the decade," said Mark Zandi.
>
> From 2001 to 2004, the average net worth of an American family rose 6.3 percent, according to the Federal Reserve's Survey of Consumer Finances. But not everybody grew richer. For the bottom 40 percent of families by income, the median net worth fell. "It just doesn't resonate with people when the Treasury secretary says everything is fine," Mr. Zandi said. "It's fine for half the population, and it's clearly not for the other half."
>
> There's a final reason that the aggregates may not accurately capture the public mood. Aggregates shed light on the performance of the economy in the last month, or in the last quarter. By contrast, measures of sentiments and polls gauge feelings about the present and expectations for the future.
>
> "Consumers tend to view current conditions as quite favorable, but their expectations for six months down the road are rather pessimistic," said Ms. Franco of the Conference Board. "It could be that we're nearing a peak, and this is as good as it gets."

The Data Says What?!

Applying Data Analysis to a Complicated World

III. Starbucks: Walmart Ratio

The Starbucks: Walmart Ratio measures the number of Starbucks for every Walmart in a given state.

National Mean: 2.4764 Starbucks for every Walmart
National Median: 1.49 Starbucks for every Walmart

This statistic is used by some political analysts to predict how Democratic or Republican a state is. The higher the ratio (the more Starbucks for each Walmart), the more Democratic the state tends to be.

Focusing on the West Coast:

State	Starbucks : Walmart Ratio (Number of Starbucks per Walmart)	Percent of Voters Registered as Republican
Alaska	3.57	41
Arizona	0.69	44
California	12.48	34
Oregon	8.38	34
Washington	13.63	31

Plot the above data and look for a trend.

How could you use this trend to predict the percentage of voters registered as Republican in Idaho based off the knowledge that their Starbucks to Walmart Ratio is 2.29?

The Data Says What?!

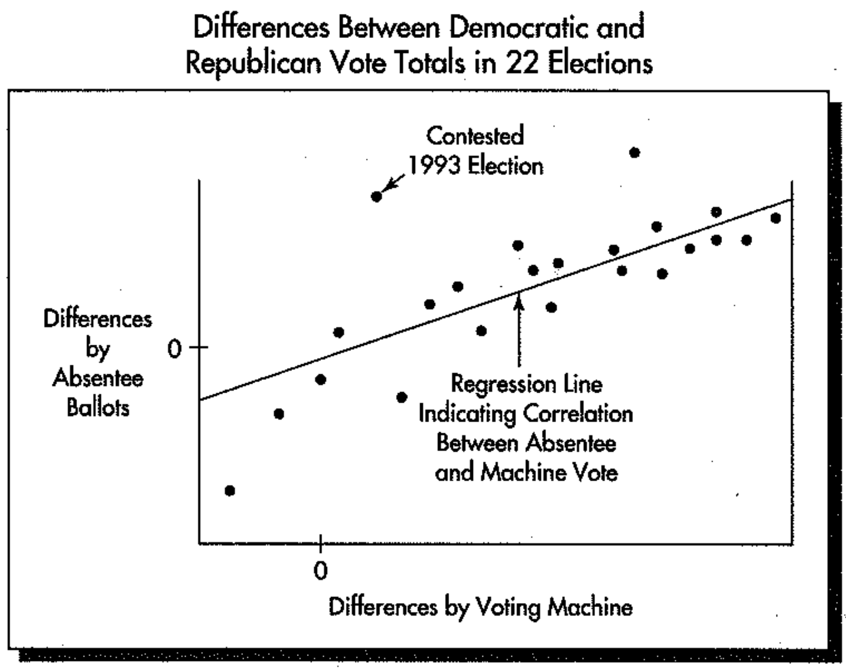
Applying Data Analysis to a Complicated World

IV. Lines of Best Fit (Trendlines): Oracles of Truth?

The media often tells us about "correlations" between various factors in our lives.

- Health: Coffee reduces my risk of developing diabetes.
- Economics: Furthering my education will increase my salary.
- Law: An apparent fraud involving absentee ballots in Pennsylvania.

  In a district crucial to the balance of power in the state legislature, the Republican candidate received more votes on the voting machines, but the Democratic candidate received an overwhelming number of absentee ballots, giving the Democrat the narrow victory. The case ended up in court with math experts for the judge, Democrats, and Republicans. The Republican math expert showed the following:



However, there are times when we must take regression with a grain of salt:

• Mental illness	Versus	Radios
• Risk of drowning in the Ocean	Versus	Ice cream consumption
• Proximity to high frequency power lines	Versus	Risk of developing cancer