CS 147: Computer Systems Performance Analysis

Summarizing Data
Overview

“Standard” Indices of Central Tendency
  Definitions
  Characteristics
  Selecting an Index

Other Indices
  Geometric Mean
  Harmonic Mean

Dealing with Ratios
  Case 1: Two Physical Meanings
  Case 1a: Constant Denominator
  Case 1b: Constant Numerator
  Case 2: Multiplicative Relationship
Summarizing Data With a Single Number

- Most condensed form of presentation of set of data
- Usually called the **average**
  - Average isn’t necessarily the mean
- Must be representative of a major part of the data set
Indices of Central Tendency

- Mean
- Median
- Mode
- All specify center of location of distribution of observations in sample
Sample Mean

- Take sum of all observations
- Divide by number of observations
- More affected by outliers than median or mode
- Mean is a linear property
  - Mean of sum is sum of means
  - Not true for median and mode
Sample Median

- Sort observations
- Take observation in middle of series
  - If even number, split the difference
- More resistant to outliers
  - But not all points given “equal weight”
Sample Mode

- Plot histogram of observations
  - Using existing categories
  - Or dividing ranges into buckets
  - Or using kernel density estimation
- Choose midpoint of bucket where histogram peaks
  - For categorical variables, the most frequently occurring
- Effectively ignores much of the sample
Characteristics of Mean, Median, and Mode

- Mean and median always exist and are unique
- Mode may or may not exist
  - If there is a mode, may be more than one
- Mean, median and mode may be identical
  - Or may all be different
  - Or some may be the same
Mean, Median, and Mode Identical
Median, Mean, and Mode All Different

- Mean
- Median
- Mode
So, Which Should I Use?

- Depends on characteristics of the metric
- If data is categorical, use mode
- If a total of all observations makes sense, use mean
- If not (e.g., ratios), and distribution is skewed, use median
- Otherwise, use mean

... but think about what you’re choosing
Some Examples

- Most-used resource in system
Some Examples

- Most-used resource in system
  - Mode
- Interarrival times
Some Examples

- Most-used resource in system
  - Mode
- Interarrival times
  - Mean
- Load
Some Examples

- Most-used resource in system
  - Mode

- Interarrival times
  - Mean

- Load
  - Median
Don’t Always Use the Mean

- Means are often overused and misused
  - Means of significantly different values
  - Means of highly skewed distributions
  - Multiplying means to get mean of a product
    - Only works for independent variables
  - Errors in taking ratios of means
  - Means of categorical variables
Geometric Means

- An alternative to the arithmetic mean

\[ \bar{x} = \left( \prod_{i=1}^{n} x_i \right)^{1/n} \]

- Use geometric mean if product of observations makes sense
Other Indices

Geometric Mean

Good Places To Use Geometric Mean

- Layered architectures
- Performance improvements over successive versions
- Average error rate on multihop network path
- Year-to-year interest rates
Harmonic Mean

- Harmonic mean of sample \( \{x_1, x_2, \ldots, x_n\} \) is

\[
\bar{x} = \frac{n}{1/x_1 + 1/x_2 + \cdots + 1/x_n}
\]

- Use when arithmetic mean of \( 1/x_i \) is sensible
Example of Using Harmonic Mean

- When working with MIPS numbers from a single benchmark
  - Since MIPS calculated by dividing constant number of instructions by elapsed time
    \[ x_i = \frac{m}{t_i} \]
  - Not valid if different \( m \)'s (e.g., different benchmarks for each observation)
Means of Ratios

- Given $n$ ratios, how do you summarize them?
- Can’t always just use harmonic mean
  - Or similar simple method
- Consider numerators and denominators
Both numerator and denominator have physical meaning

Then the average of the ratios is the ratio of the averages
Dealing with Ratios

Case 1: Two Physical Meanings

Example: CPU Utilizations

<table>
<thead>
<tr>
<th>Measurement Duration</th>
<th>CPU Busy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>40</td>
</tr>
<tr>
<td>1</td>
<td>50</td>
</tr>
<tr>
<td>1</td>
<td>40</td>
</tr>
<tr>
<td>1</td>
<td>50</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
</tr>
</tbody>
</table>

Sum 200%

Mean? 20/30
### Example: CPU Utilizations

<table>
<thead>
<tr>
<th>Measurement</th>
<th>CPU Busy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duration</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>40</td>
</tr>
<tr>
<td>1</td>
<td>50</td>
</tr>
<tr>
<td>1</td>
<td>40</td>
</tr>
<tr>
<td>1</td>
<td>50</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
</tr>
<tr>
<td><strong>Sum</strong></td>
<td><strong>200%</strong></td>
</tr>
</tbody>
</table>

**Mean?** Not 40%
### Case 1: Two Physical Meanings

#### Example: CPU Utilizations

<table>
<thead>
<tr>
<th>Measurement</th>
<th>CPU Busy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duration</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>40</td>
</tr>
<tr>
<td>1</td>
<td>50</td>
</tr>
<tr>
<td>1</td>
<td>40</td>
</tr>
<tr>
<td>1</td>
<td>50</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
</tr>
</tbody>
</table>

**Sum:** 200%

**Mean?** Nor 1.92%! 

---

2015-06-15
Why not 40%?
Because CPU-busy percentages are ratios
  - So their denominators aren’t comparable
The duration-100 observation must be weighted more heavily than the duration-1 ones
Go back to the original ratios:

$\text{Mean CPU Utilization} = \frac{0.40 + 0.50 + 0.40 + 0.50 + 20}{1 + 1 + 1 + 1 + 100} = 21\%$
Consider the Mean of Ratios: Case 1a

- Sum of numerators has physical meaning
- Denominator is a constant
- Take arithmetic mean of the ratios to get overall mean
For Example,

- What if we calculated CPU utilization from last example using only the four duration-1 measurements?
- Then the average is

\[
\frac{1}{4} \left( \frac{.40}{1} + \frac{.50}{1} + \frac{.40}{1} + \frac{.50}{1} \right) = 0.45
\]
Sum of denominators has a physical meaning
Numerator is a constant
Take harmonic mean of the ratios
Numerator and denominator are expected to have a multiplicative, near-constant property

\[ a_i = cb_i \]

Estimate \( c \) with geometric mean of \( a_i/b_i \)
An optimizer reduces the size of code

What is the average reduction in size, based on its observed performance on several different programs?

Proper metric is percent reduction in size

And we’re looking for a constant $c$ as the average reduction
Dealing with Ratios  
Case 2: Multiplicative Relationship  

Program Optimizer Example, Continued

<table>
<thead>
<tr>
<th>Program</th>
<th>Code Size Before</th>
<th>Code Size After</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>BubbleP</td>
<td>119</td>
<td>89</td>
<td>.75</td>
</tr>
<tr>
<td>IntmmP</td>
<td>158</td>
<td>134</td>
<td>.85</td>
</tr>
<tr>
<td>PermP</td>
<td>142</td>
<td>121</td>
<td>.85</td>
</tr>
<tr>
<td>PuzzleP</td>
<td>8612</td>
<td>7579</td>
<td>.88</td>
</tr>
<tr>
<td>QueenP</td>
<td>7133</td>
<td>7062</td>
<td>.99</td>
</tr>
<tr>
<td>QuickP</td>
<td>184</td>
<td>112</td>
<td>.61</td>
</tr>
<tr>
<td>SieveP</td>
<td>2908</td>
<td>2879</td>
<td>.99</td>
</tr>
<tr>
<td>TowersP</td>
<td>433</td>
<td>307</td>
<td>.71</td>
</tr>
</tbody>
</table>
Why Not Use Ratio of Sums?

- Why not add up pre-sizes and post-optimized sizes and take the ratio?
  - Benchmarks of non-comparable size
  - No indication of importance of each benchmark in overall code mix
  - When looking for constant factor, not the best method
So Use the Geometric Mean

- Multiply the ratios from the 8 benchmarks
- Then take the $1/8$ power of the result

\[
\bar{x} = (0.75 \times 0.85 \times 0.85 \times 0.88 \times 0.99 \times 0.61 \times 0.99 \times 0.71)^{1/8} \\
= 0.82
\]