An Overview of Distributions and How to Describe Them

Mike DeWitt
Institutional Research
Two Topics will be covered

• Overview of different types of distributions of data

• Some metrics to describe them
What’s in a name....
What’s in a name

• A **distribution** is the **form and frequency** that the data take
  • What is its **measure of central tendency**
    • Mean
    • Median
    • Mode
  • How “**spread out**” is it
    • Range
    • Standard Deviation
  • How “**peaky**” is the distribution
    • Kurtosis
Our friend the normal distribution

Normal Distribution
Our friend the normal distribution
Our friend the normal distribution

Normal Distribution

MEDIAN
Our friend the normal distribution

For symmetric, normal distributions the mean = median
Right Skew Distribution
Right Skew Distribution

MEAN
Right Skew Distribution

MEDIAN

MEAN
Right Skew Distribution

For Right Skew Distributions:
Mean > Median
Left Skew Distribution
Left Skew Distribution

MEAN
Left Skew Distribution

Mean

Median
For Right Skew Distributions
Median > Mean
But what about spread?
But what about spread?
But what about spread?

- **Standard Deviation**
  - Typically only appropriate for normal distribution which gives these nice guidelines
    - 68% of the data is within 1 standard deviation of the mean
    - 95% of the data is within 2 standard deviations of the mean
    - 99.97% of the data is within 3 standard deviation of the mean

- **Range**
  - Maximum Value – Minimum Value
  - Can be used to describe all kinds of distributions
Remind me of standard deviation....

Standard deviation = 20

Standard deviation = 5

Same mean, but standard deviation is 4x greater on right than left distribution
So why only normal distributions?
So why only normal distributions?

95% of the data is between -1.8 StdDev to +1.4 StdDev

95% of the data is between -1.4 StdDev to + 1.8 StdDev
So why only normal distributions?

95% of the data is between -1.8 StdDev to +1.4 StdDev

95% of the data is between -1.4 StdDev to +1.8 StdDev

Only use our 68-95-99.97 rule with normal distributions

This is why it is important to know the shape of the distribution
We’ve described central tendency, spread...

YOU KNOW WHAT A KURTOSIS IS?

TELL ME MORE
Kurtosis...how “peaky” is a distribution

- Kurtosis quantifies how much data exists in the tails of the distribution
Kurtosis...how “peaky” is a distribution

- Kurtosis quantifies how much data exists in the tails of the distribution.
Kurtosis...how “peaky” is a distribution

- Kurtosis quantifies how much data exists in the tails of the distribution
Kurtosis...how “peaky” is a distribution

- Kurtosis quantifies how much data exists in the tails of the distribution
- If there is more mass in the tails then more extreme results are likely
- Again...our standard deviation rules for a normal distribution fails with highly kurtotic data
Why do we need to know this...

- When we do hypothesis testing we need to ensure that the distribution of our data meets certain conditions
- It lets us use statistics to say “these are statistically different”