# Why Normal is Better than Average

Get more comfortable using probability distributions to describe your data.

Presented 2/10/2016 to ASQ0511 by
Kristine Hejna

# About this talk

- Benefits of probability distributions  instead of single values.
  - Better representations of reality for forecasting
  - Provide more information for decision making
  - Individual acceptance
- Examples of probability distributions instead of single values
  - Normal examples
  - Examples with other distributions

# Probability Distributions Are Useful Models

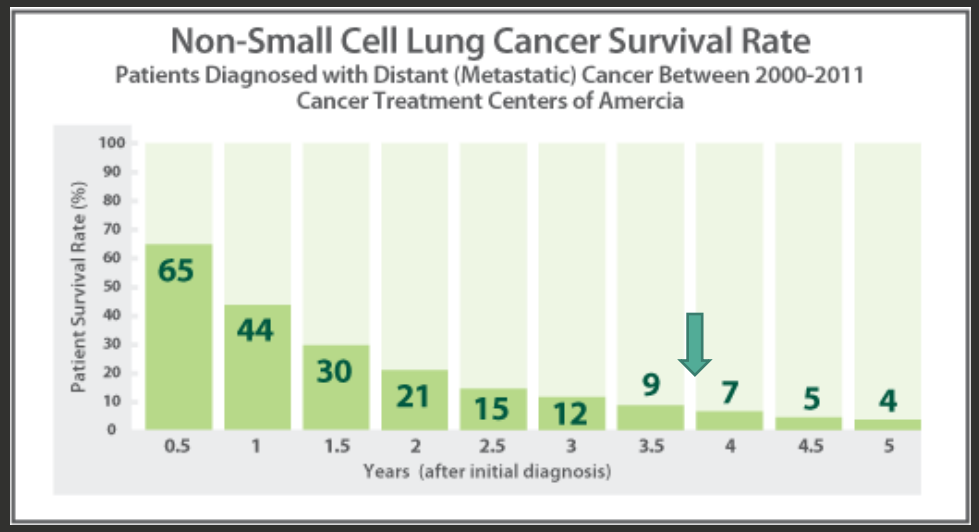
- Normal distributions
- Almost
  - Poisson
  - Weibull, Rayleigh
- Others
  - Bathtub
  - Multimodal

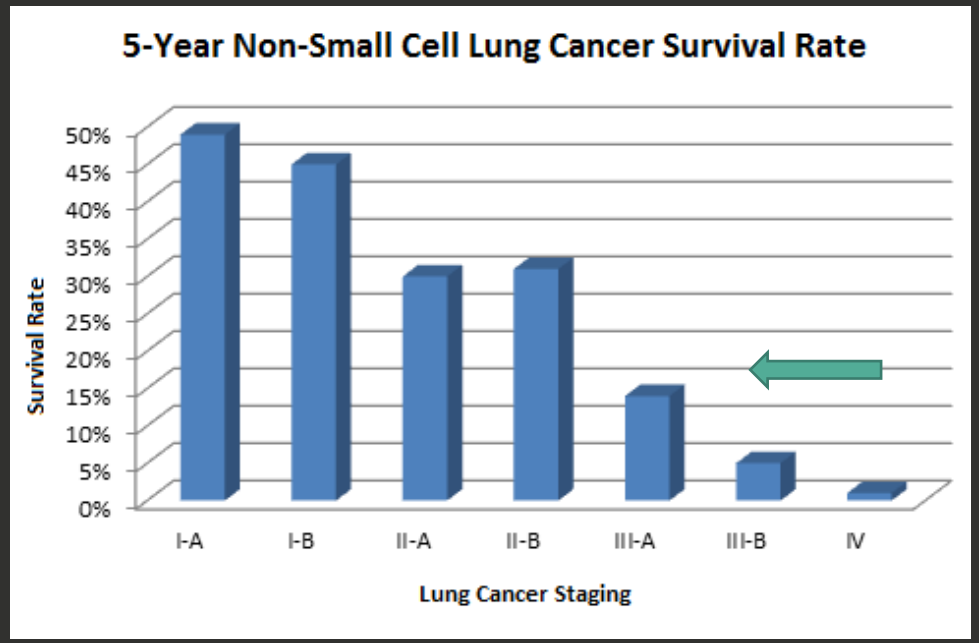
When you can find and describe a function that fits your data observations you can use it to predict future results.

All models are wrong. Some are useful.
*(George E.P. Box)*

# More Numbers, more information

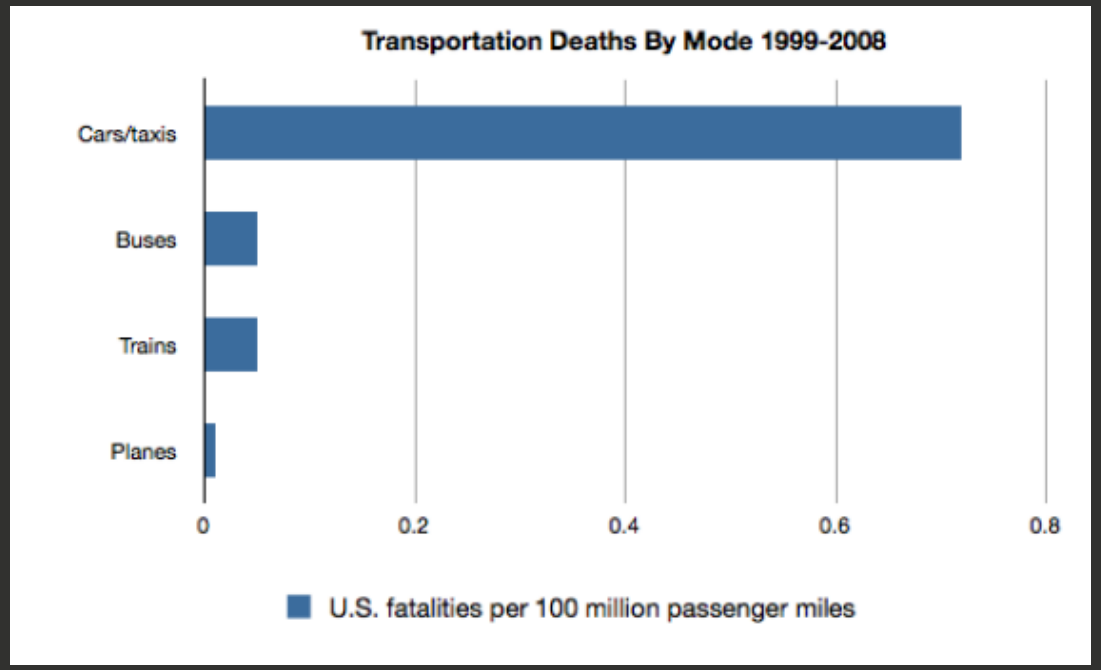
- The most recent lung cancer statistics show the overall five-year survival for lung cancer is **16 percent** — that is, 16 of every 100 people diagnosed with the disease will still be alive 5 years after diagnosis. (cancer.gov)
- These other stats say that if a person is diagnosed at Stage IV,  his/her chances of being alive after 5 years is less than 5%, and if he/she was diagnosed at Stage 1A it is more than 45%
- You would want to know.



# Don’t sweat the small stuff: Is this the small stuff?

- People often worry about unlikely events.
- People are often surprised by undesirable outcomes when they have tacitly accepted a risk.
- Practice developing a sense of probability and probability distributions of various events and conditions you will encounter.
  - Accept the things you cannot change, and the things that aren’t worth dealing with
  - Make the changes that will have an impact.



# The Law of Large Numbers

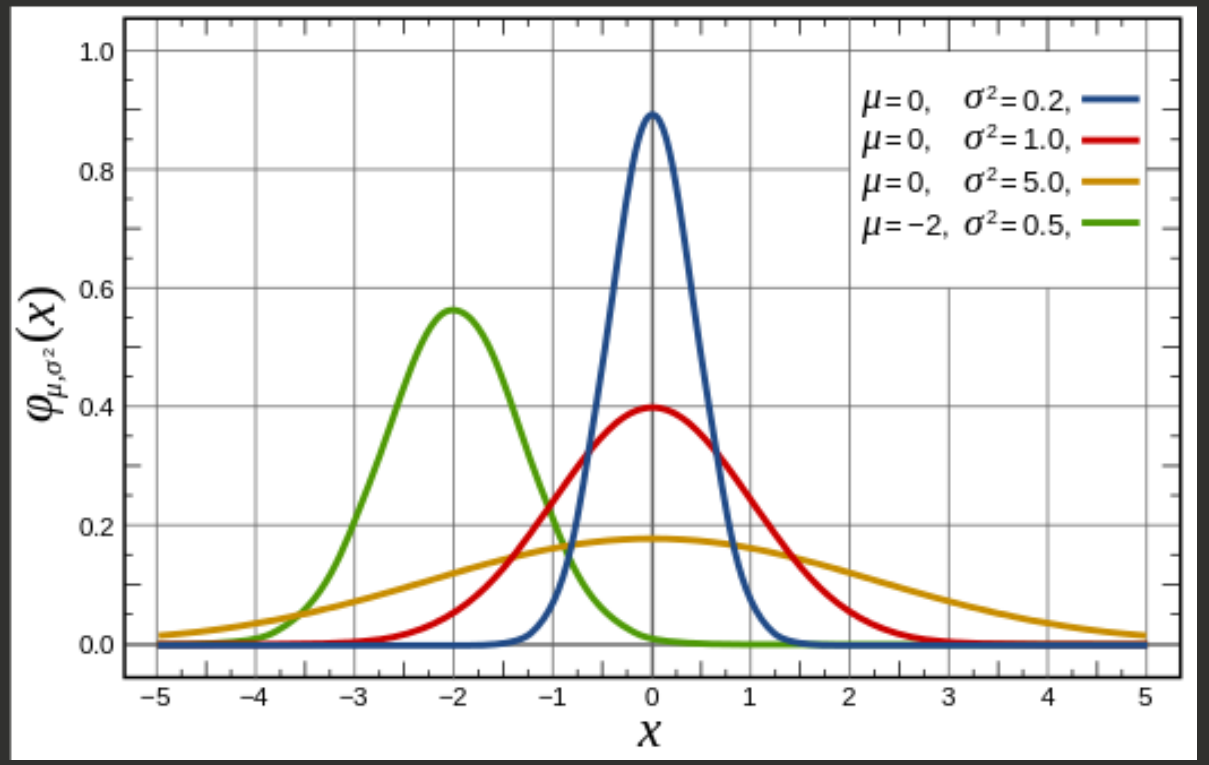
- The **strong law of large numbers** states that the sample average converges almost surely to the expected value.
- The **weak law of large numbers** states that the sample average converges in probability towards the expected value.
- Not much help when the numbers aren't large!

# About the Central Limit Theorem

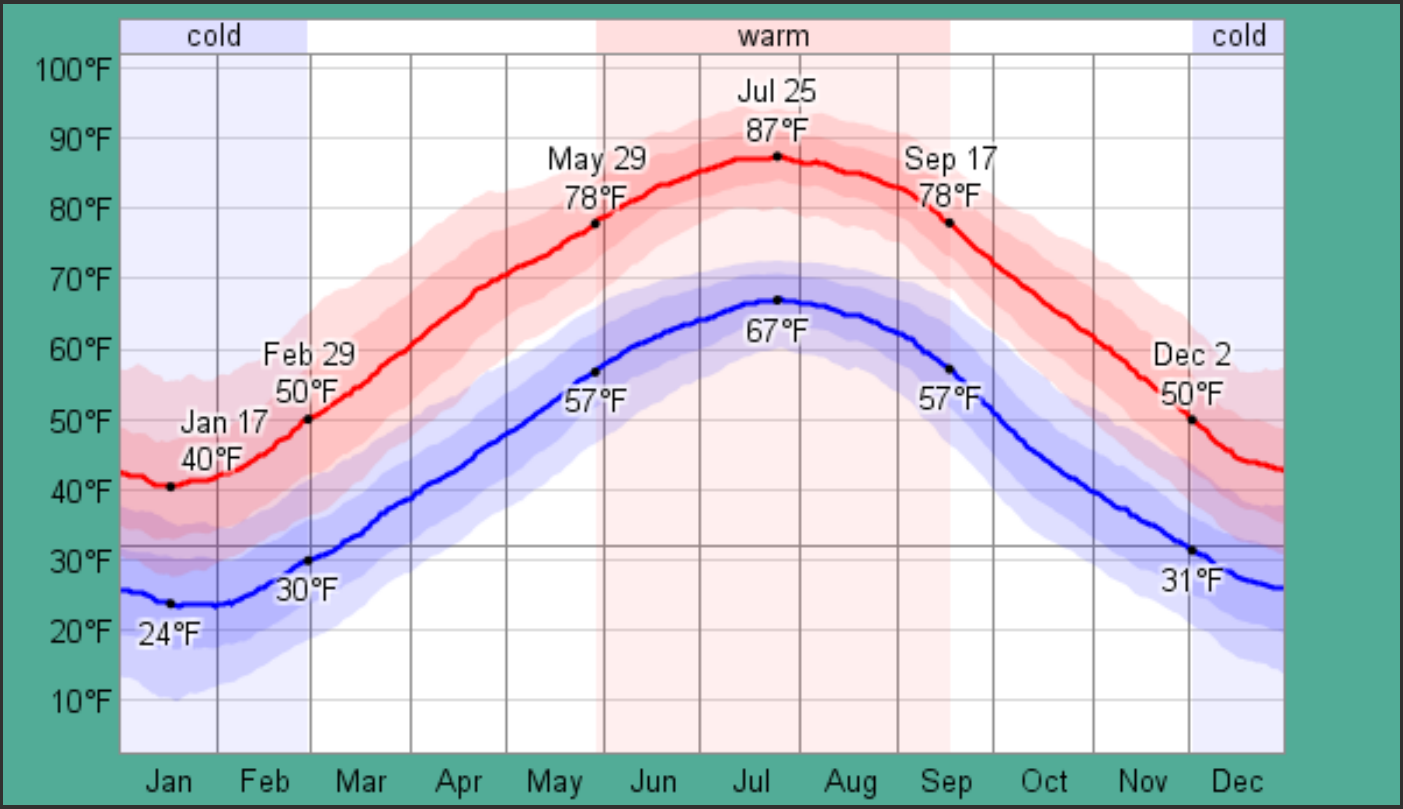
- Take any population, whether it's normally distributed or not. Randomly select at least 30 members from that population, measure them for some characteristic, and then find the average of those measures. That average is one data point. Return the samples, select another random sample of the same number, and find the average of *their* measures. Do the same again and again. The Central Limit Theorem says that those averages tend to have a normal distribution.

# Bell Curve, Gaussian Distribution, Normal

- Normal distributions are symmetric, and the mean, median, mode are all the same point.
- Normal is usual. For a single population we get a graph like these.
- Many animals (including people) have a strong sense of  normal and will reject individuals with observable characteristics that differ too much from the central tendency.  People can recognize a non-normal result from familiar processes.



# Average temperature, normal for here.

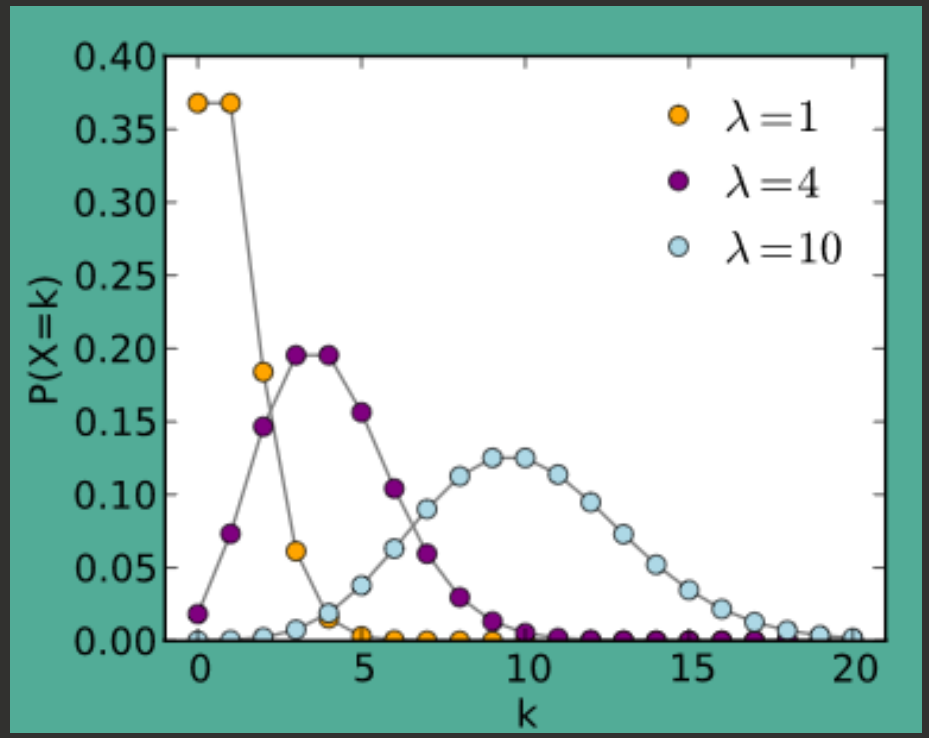



- Fairfax County
- These are normal temperatures across the shaded widths.
- Radio stations and newspapers use the averages and call them *normal.*
- We don't actually get "average" temperatures very often.
- We get "normal" temperatures all the time.

# Not Normal may be Expected

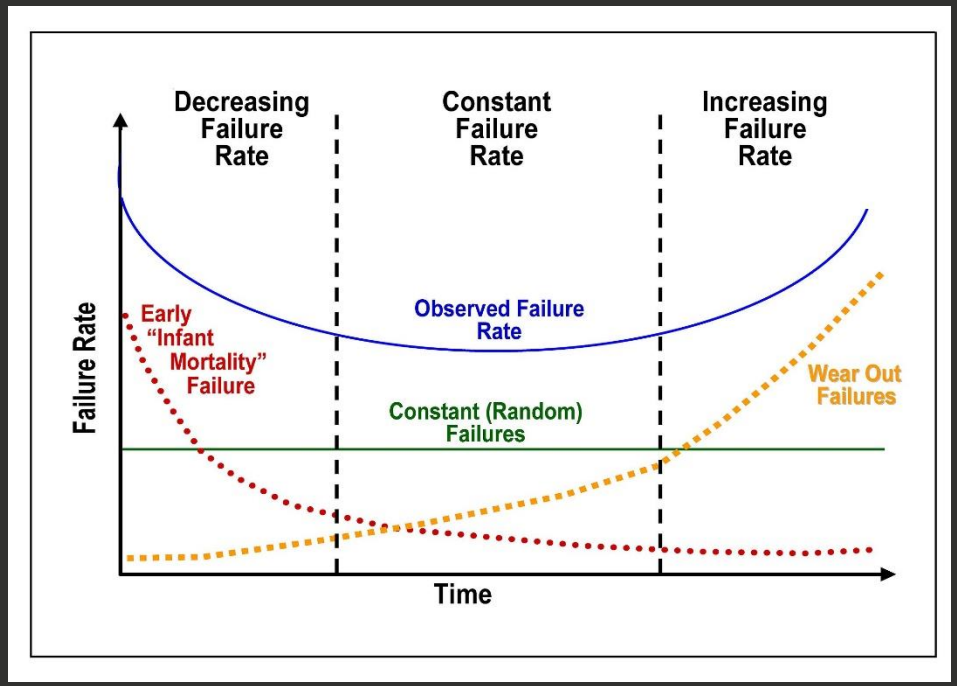
- Poisson Distributions describe occasional, independent, events.
  - Soccer goals
  - Calls to a call center
  - Radioactive decay
  - Overflow floods

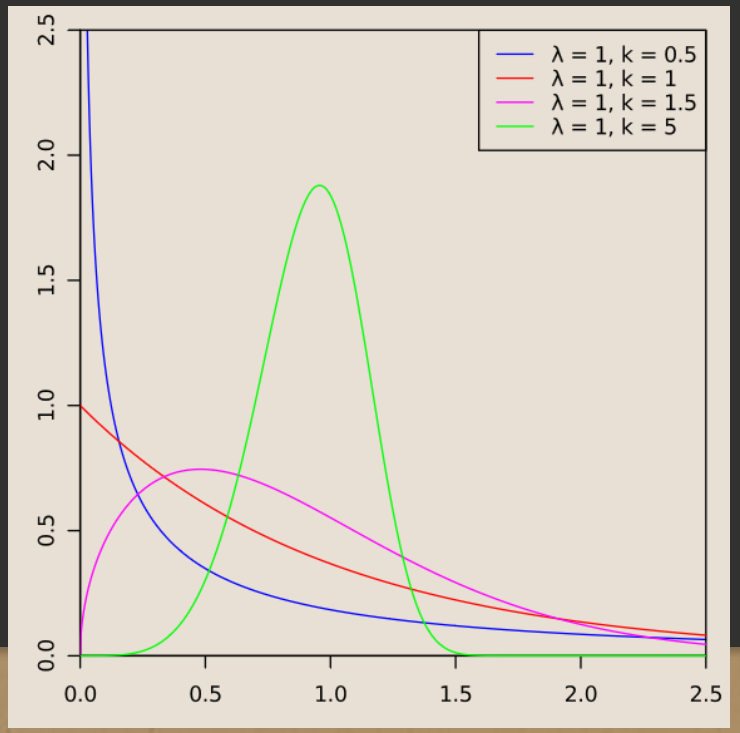
- The horizontal axis is the index *k*, the number of occurrences. λ is the expected value. The function is defined only at integer values of *k*. The connecting lines are only guides for the eye.



# Not Normal may be Expected

- Reliability engineering uses bathtub curves.
  - Failure rates over a product life
- Reliability engineering also uses Weibull Distributions

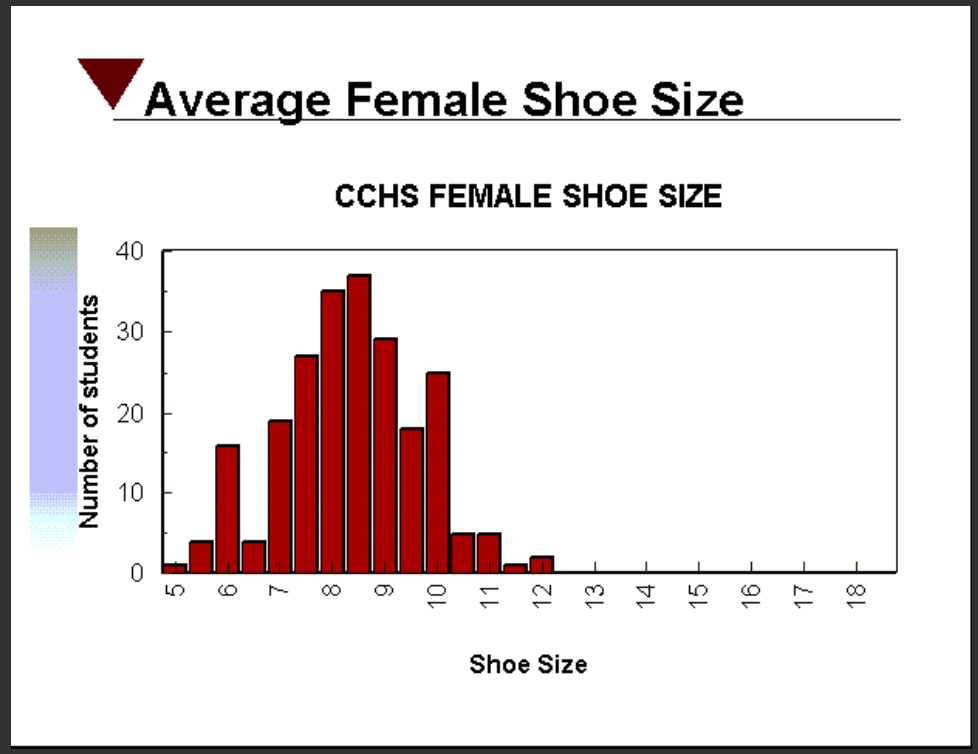



By Calimo - Own work, after Philip Leitch., CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=9671814

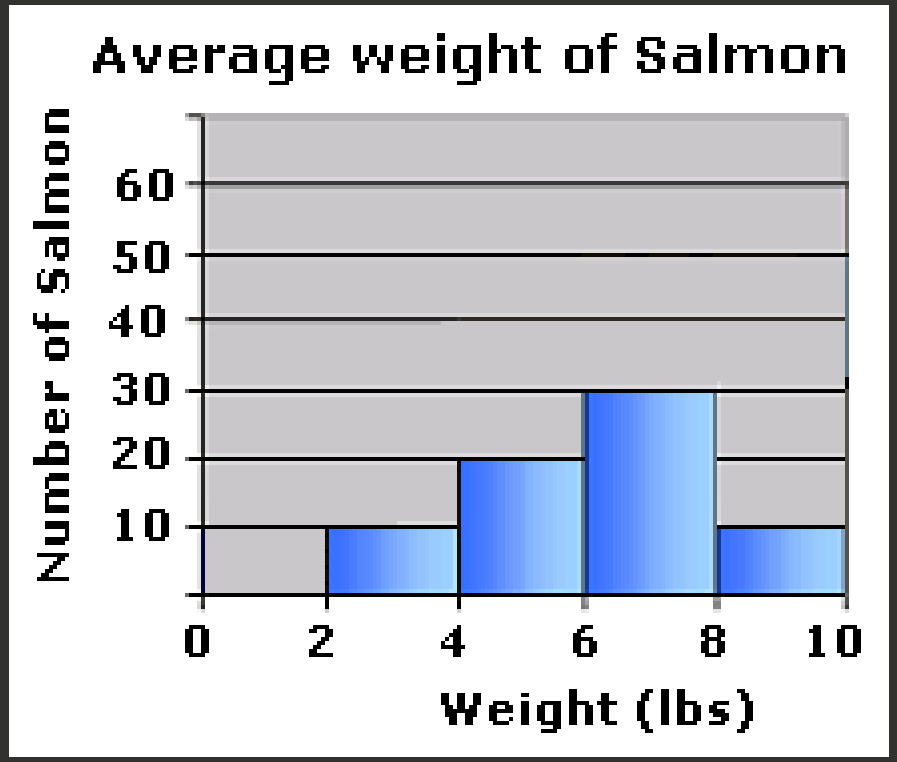
# Thinking in distributions

- Political predictions are all using distributions
- Candidate X is ahead of Candidate Y 28% to 23% with voters 18 to 24
- The average women's shoe size
  - American women average size 8
  - Companies don't make just the average size.



# Ask yourself – critical thinking

- Are you limiting your information by thinking in averages?
- Are you missing something important?
- Is someone limiting the information you are getting? Intentionally?
- Did they weigh salmon that weren't caught? Is this only adult salmon, spawning salmon, legally caught salmon (so no one reported the small ones?)



# Resources and recommended reading

- Probability Distributions
  - Wikipedia – not supposed to contain original content, but has some of the best write ups on various distributions and applications of them
- Critical Thinking
  - Paulos, John Allen, *A Mathematician Reads the Newspaper* , Basic Books, 1995

# Summary

- Benefits of probability distributions  instead of single values.
  - Better representations of reality for forecasting
  - Provide more information for decision making
  - Individual acceptance
- Examples of probability distributions instead of single values
  - Normal examples
  - Examples with other distributions
- Critical thinking refresher, too.