The Normal Distribution

INFO-1301, Quantitative Reasoning 1

University of Colorado Boulder

March 17, 2017

Prof. Michael Paul

Normal Distribution

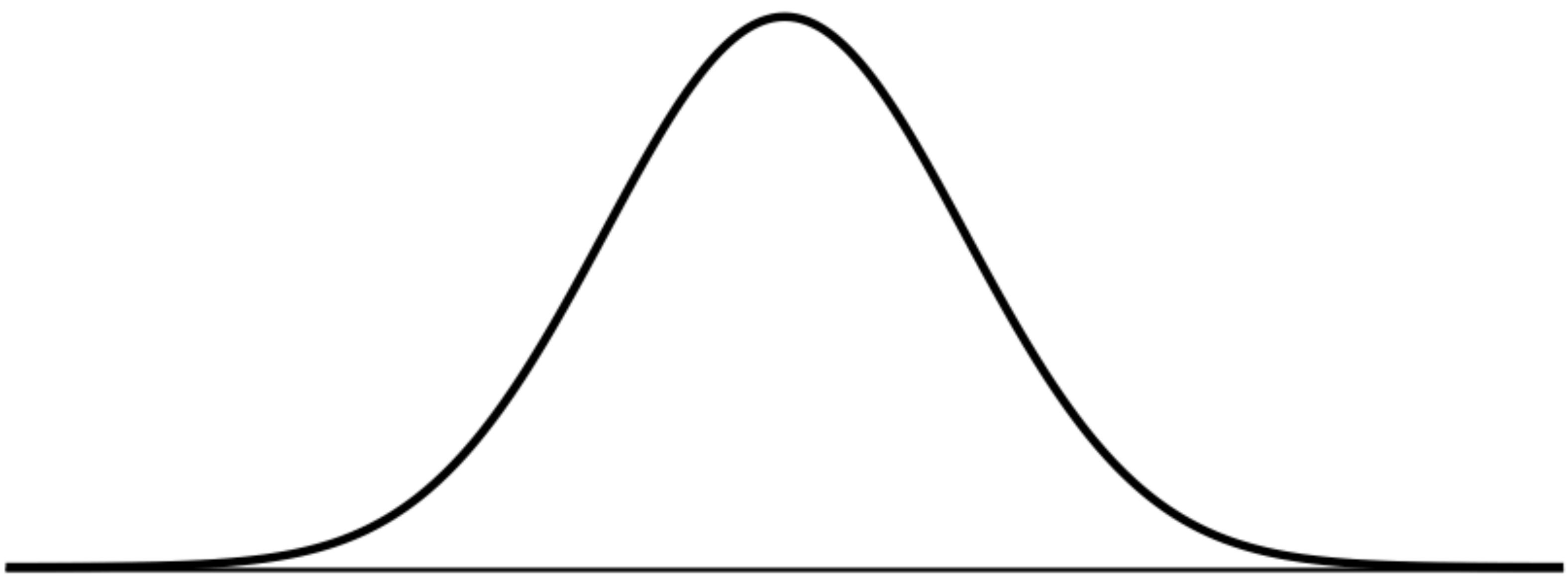
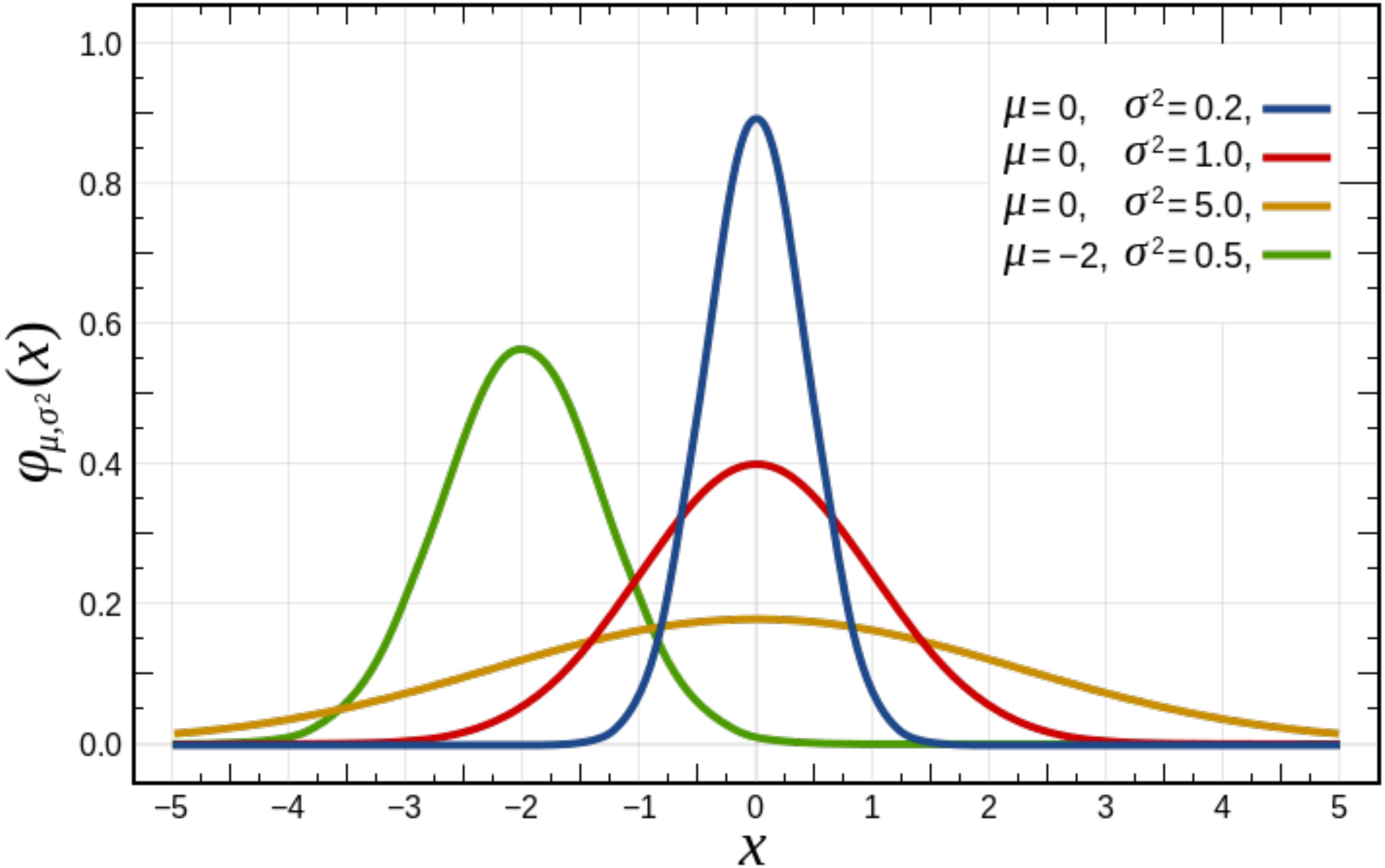


Figure 3.1: A normal curve.

Normal Distribution



Normal Distribution

- The most common curve in all of statistics and in all of the applications of statistics to science
- Unimodal, symmetric, bell curve
- Few data sets are perfectly normal in real life, but many are almost normal and many applications benefit from treating the distribution as normal
- First mathematical analysis of the normal distribution by Carl Frederic Gauss (1809)

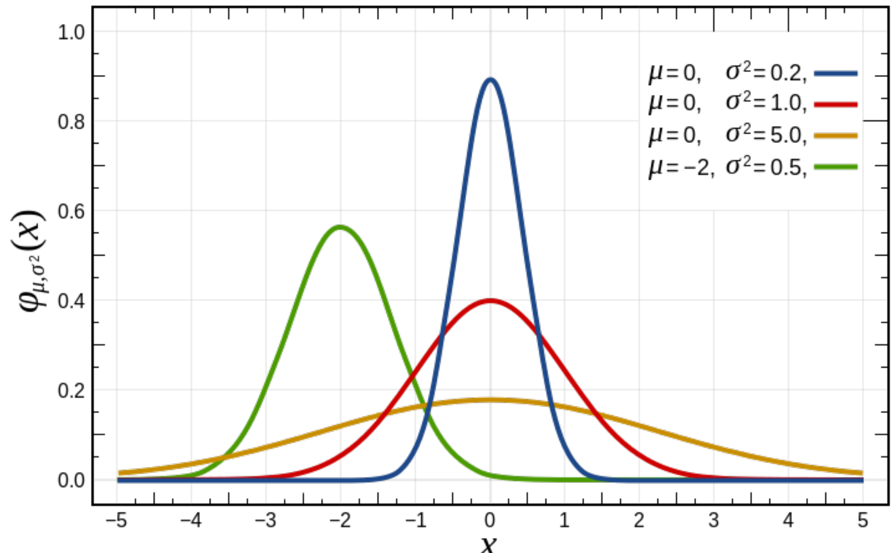
Also called the
Gaussian distribution

Normal Distribution

- The normal distribution is defined by the mean (mu, written as **μ**) and the standard deviation (sigma, written as **σ**)
- Written as N(**μ**, **σ**)
  - Or sometimes N(**μ**, **σ^2**) to show variance instead of standard deviation
- **μ** and **σ** are called **parameters**.
- http://students.brown.edu/seeing-theory/distributions/index.html#second
- N(0, 1) is called the **standard** normal distribution

Probability Density

- What does the normal distribution tell us?

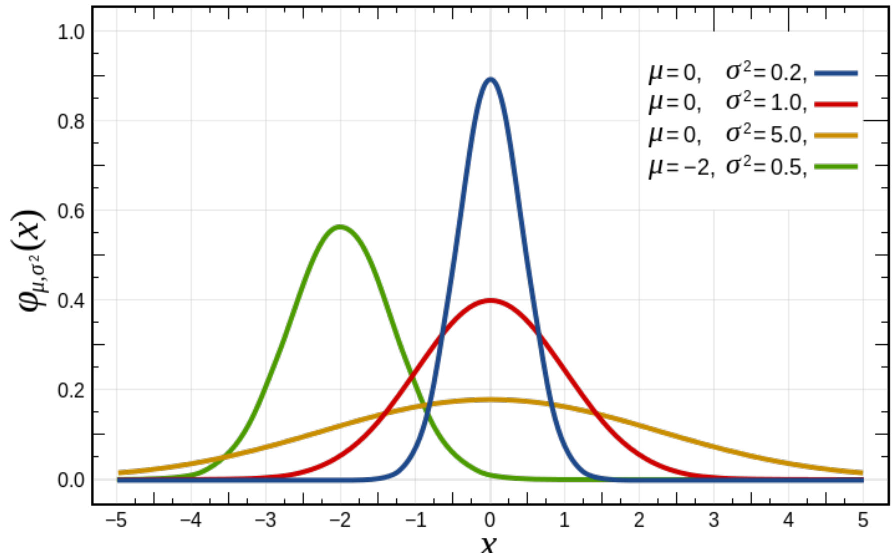


- The sample space is continuous: our concept of probability doesn't quite apply
- Instead: probability **density**

$$\frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Probability Density

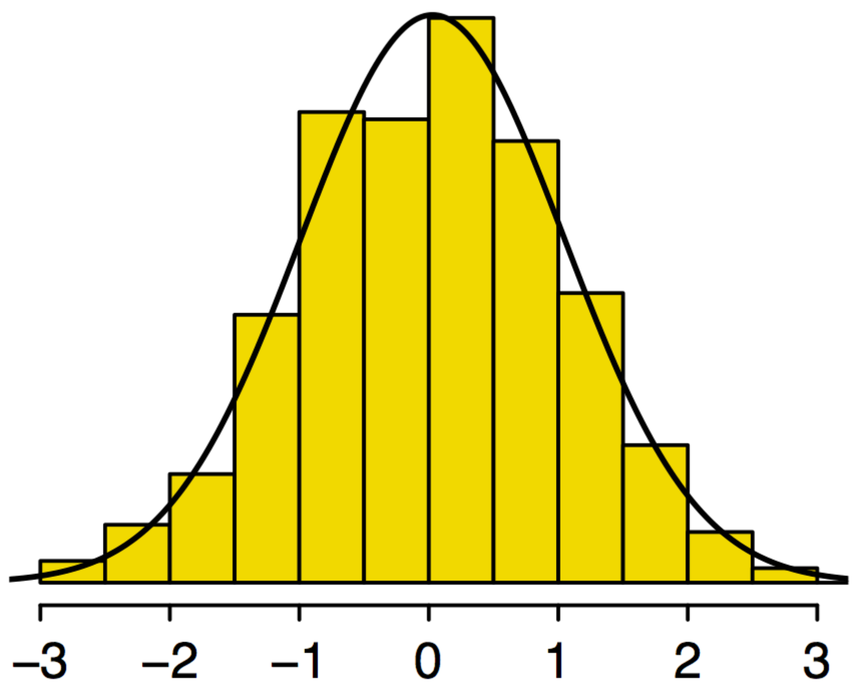
- What does probability density tell us?



- Probability of **ranges**:
  - "P(-1 ≤ X ≤ 1) = 0.68"
- **Relative** probability:
  - "It is twice as likely that X will be 0 than X will be 1.5"

Normal Approximation

Real data often naturally forms a normal curve



- Not an exact match, but a good approximation

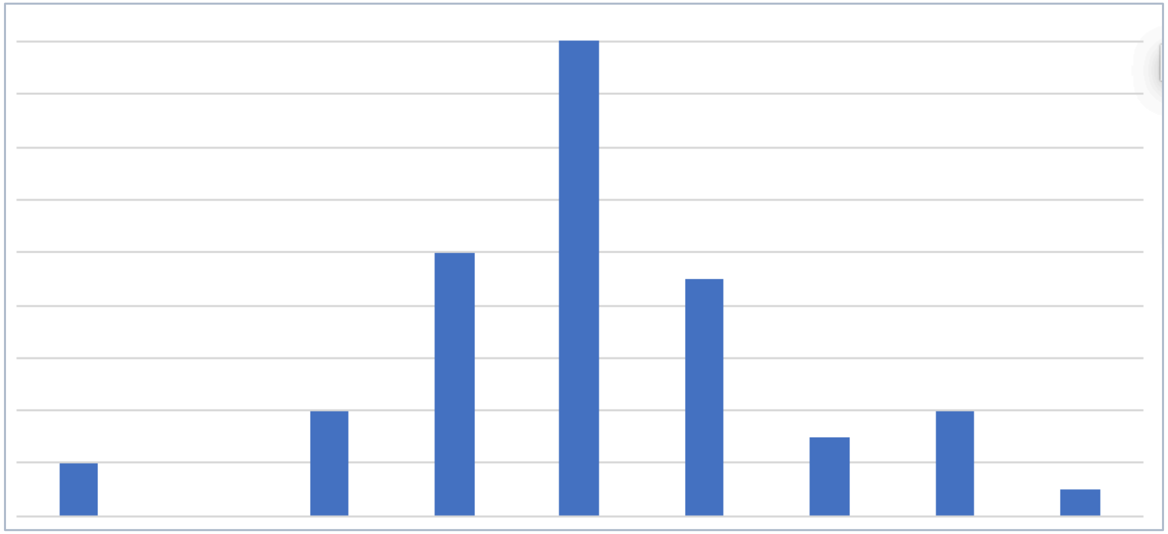
Normal Approximation

Using the normal distribution as an approximation to your data can help answer questions like:

- Probability of **ranges**:
  - “$P(-1 \leq X \leq 1) = 0.68$”
- **Relative** probability:
  - “It is twice as likely that X will be 0 than X will be 1.5”

Normal Approximation

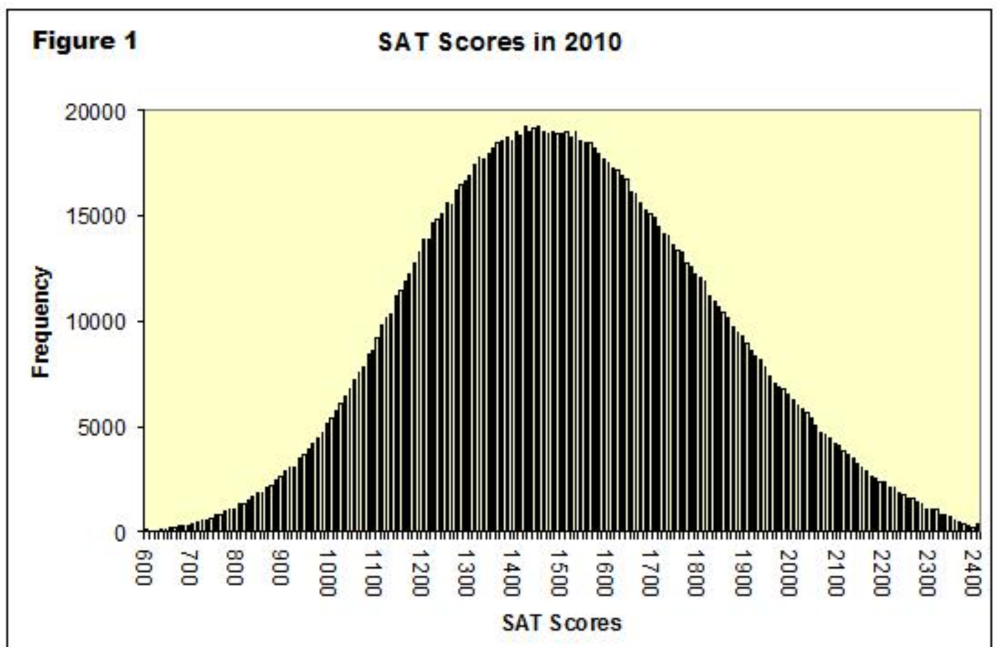
If your data is not reliable, the normal distribution will be a “smoother” curve, potentially more accurate than your actual data.

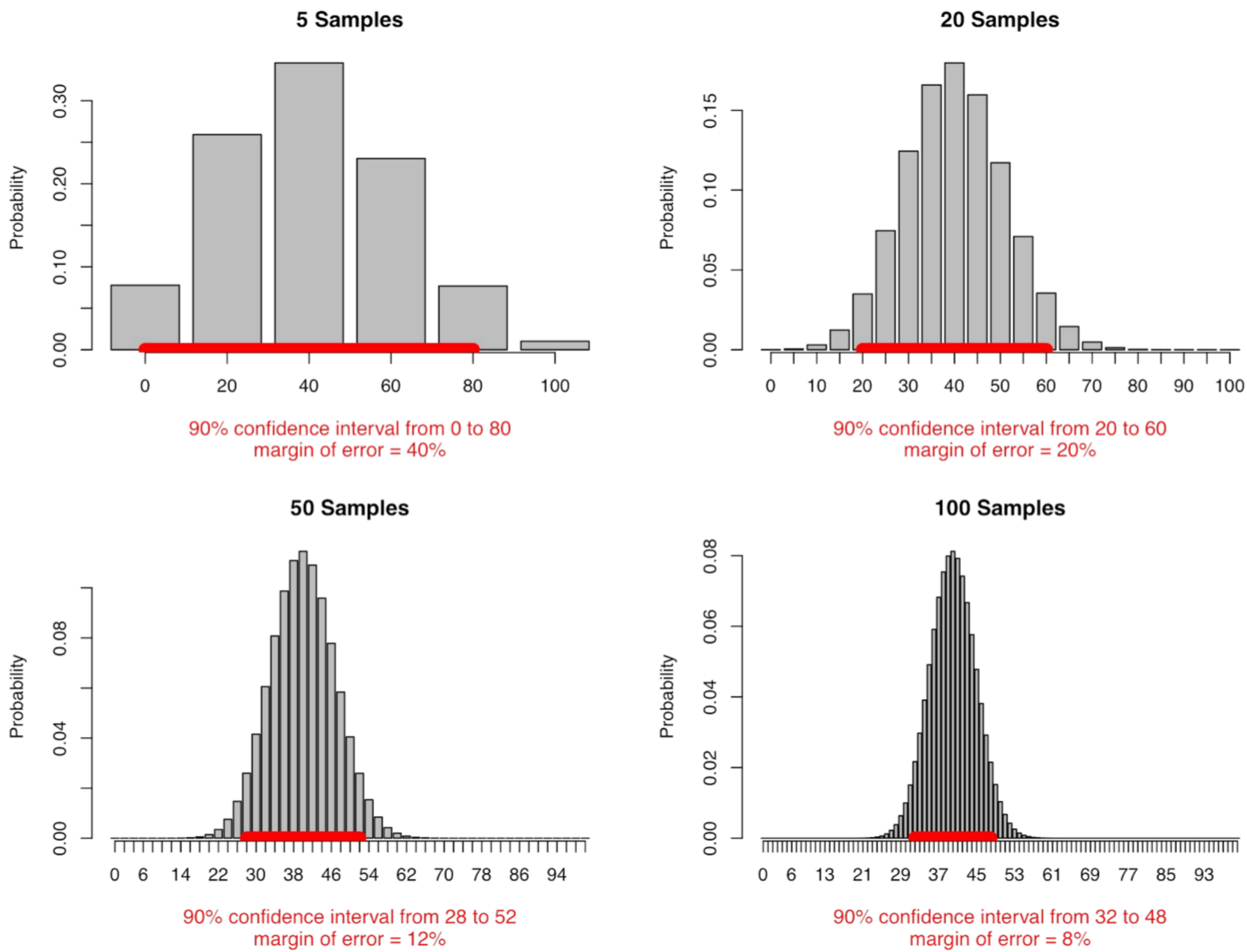


Normal Approximation

Examples of normally distributed data:

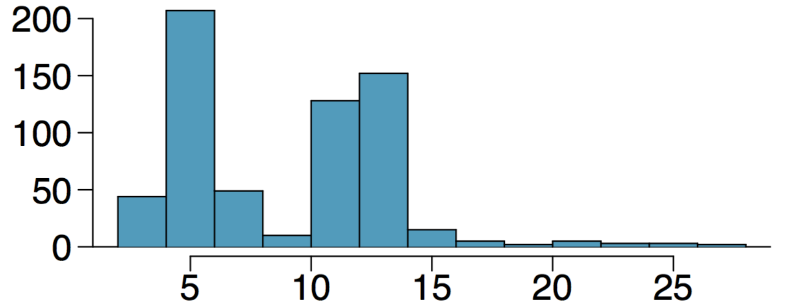
- Speeds of different cars at a spot on a highway
- Physical attributes (e.g., height of people)
- Measurement error (e.g., radar speed guns)
- Test scores



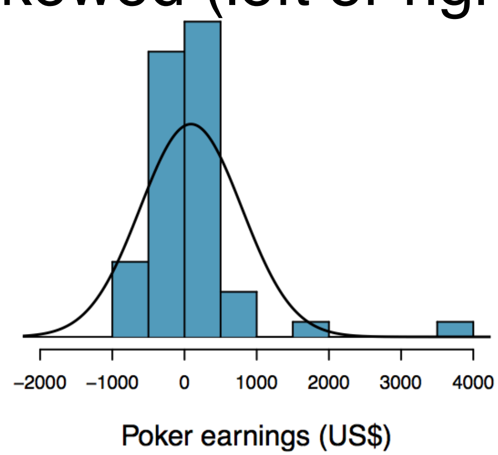

5 Samples
Probability
0.00 0.10 0.20 0.30
0 20 40 60 80 100
90% confidence interval from 0 to 80
margin of error = 40%
20 Samples
Probability
0.00 0.05 0.10 0.15
0 10 20 30 40 50 60 70 80 90 100
90% confidence interval from 20 to 60
margin of error = 20%
50 Samples
Probability
0.00 0.04 0.08
0 6 14 22 30 38 46 54 62 70 78 86 94
90% confidence interval from 28 to 52
margin of error = 12%
100 Samples
Probability
0.00 0.02 0.04 0.06 0.08
0 6 13 21 29 37 45 53 61 69 77 85 93
90% confidence interval from 32 to 48
margin of error = 8%

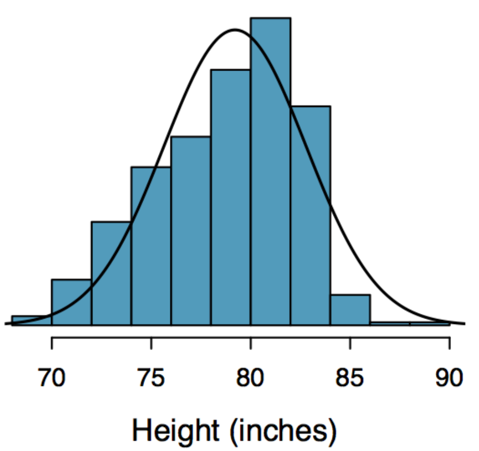
Normal Approximation

- When the normal distribution is a bad approximation:
  - Bimodal or multimodal distributions



  - Skewed (left or right) distributions

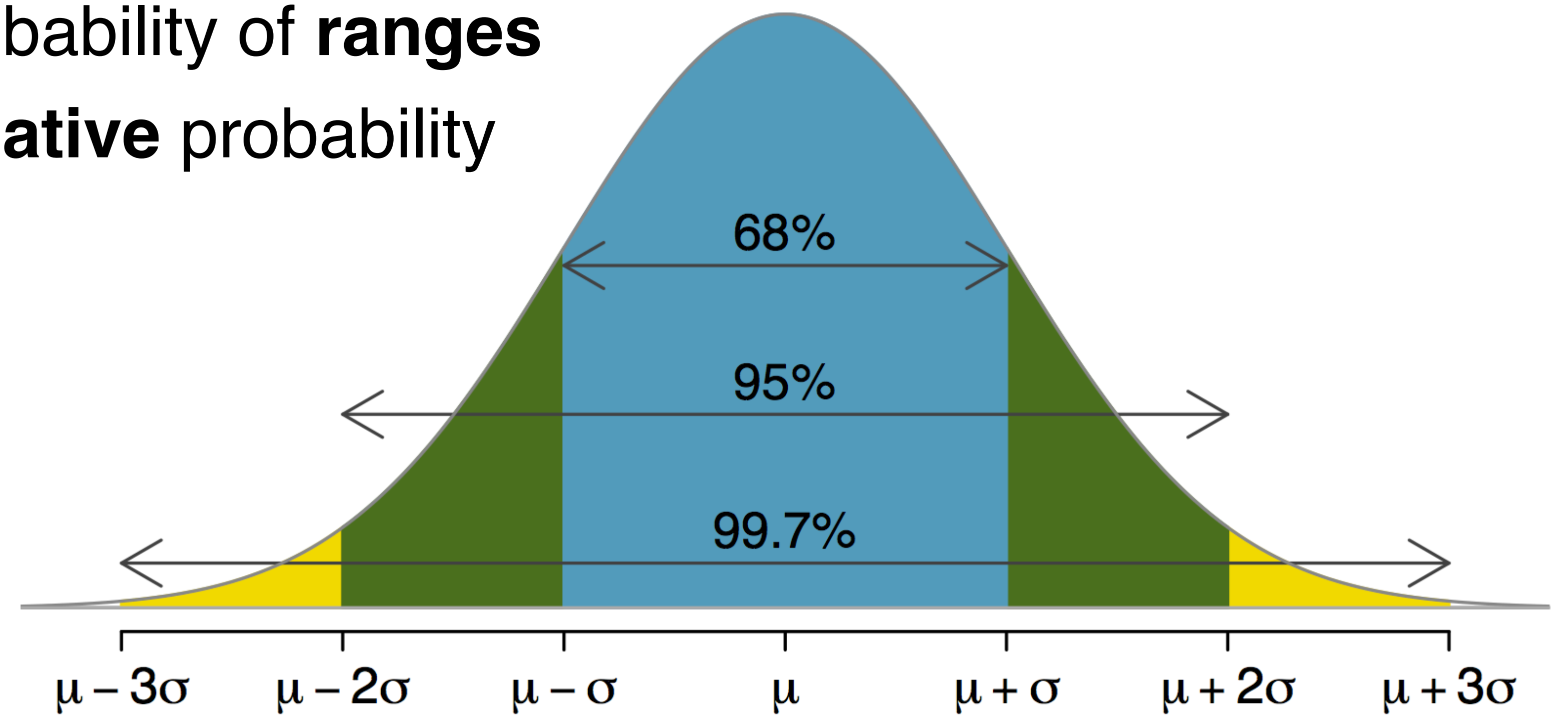




What can we do with this?

If the normal distribution is a good approximation, then we can use the math of the probability density to answer questions about the data:

- Probability of **ranges**
- **Relative** probability



What can we do with this?

Common use case: measurement error

- We can quantify the probability that our error is acceptably small