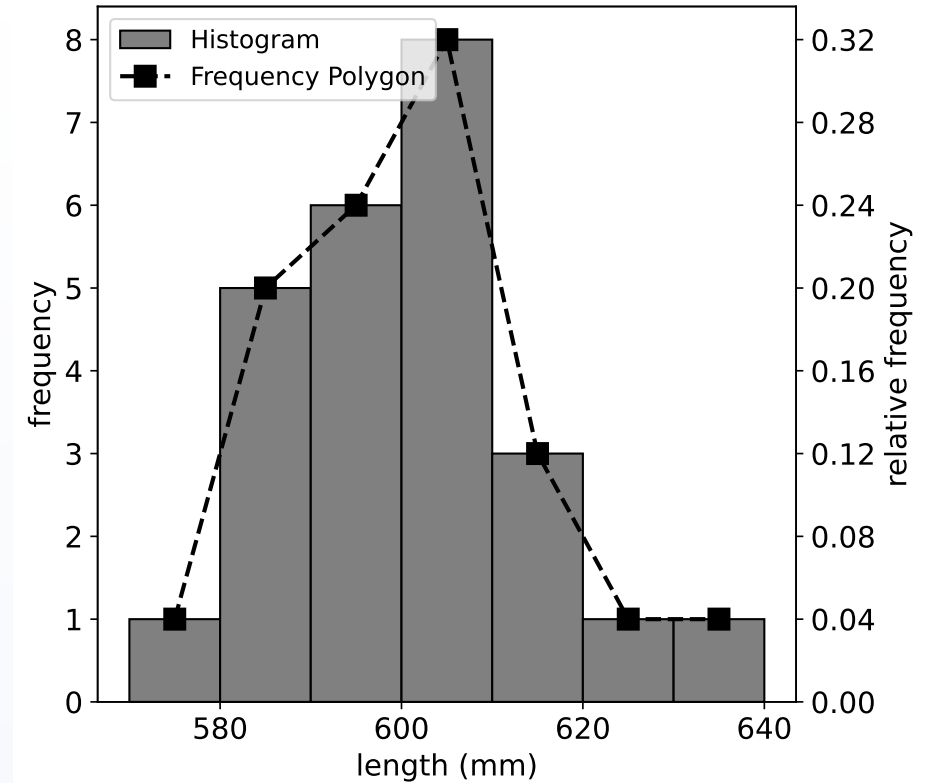


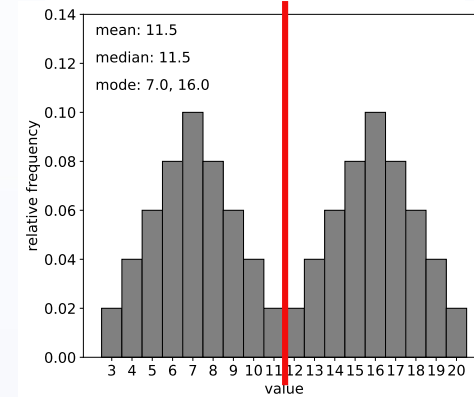
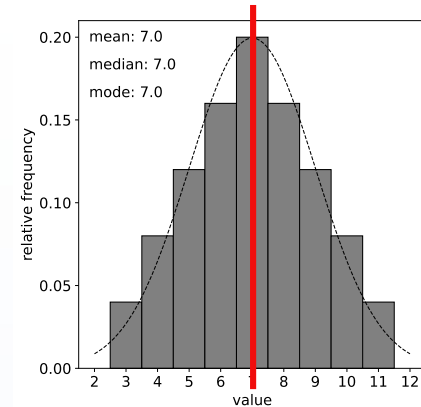
EGM101 – Skills Toolbox

Week 5, Part 6: Data Distributions

- “Shape” of data tells us something about the data
- Helps us compare different datasets
- Can help draw conclusions:
 - Effects of medical treatments
 - Outcomes of lab experiments
 - Presence of fraud/tampering

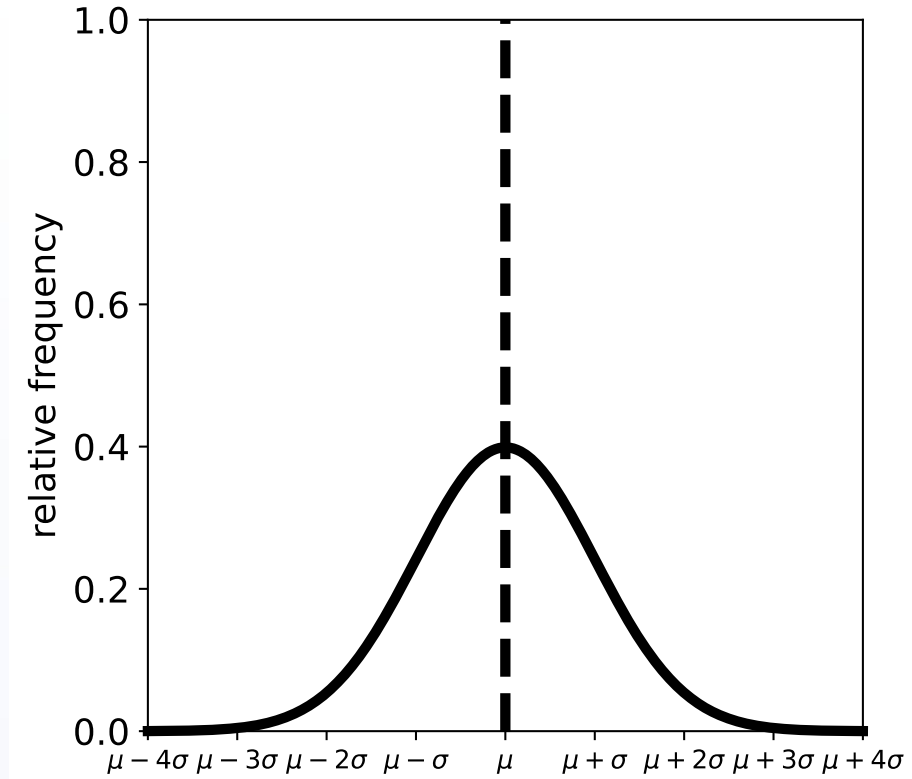


- **Symmetrical** distribution: left side mirrors right side
- Perfectly symmetrical: mean, median are same
 - Unimodal: mode equals median, mean
 - Multimodal: mode(s) not equal to median, mean
- **Skewed**: distribution is not symmetrical



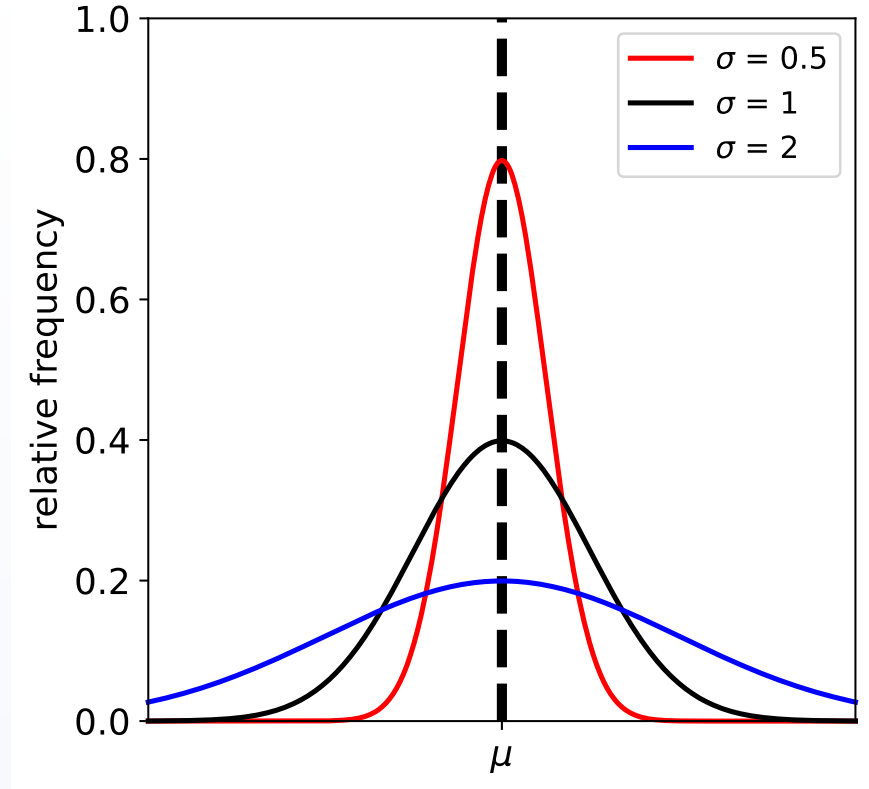
The Normal Distribution

- The most important distribution for statistics/probability*
 - Often used to approximate unknown distributions
- “Bell(-shaped) curve”
 - Not the only one!
- Symmetric: mean, median, mode coincide
- More on this later...

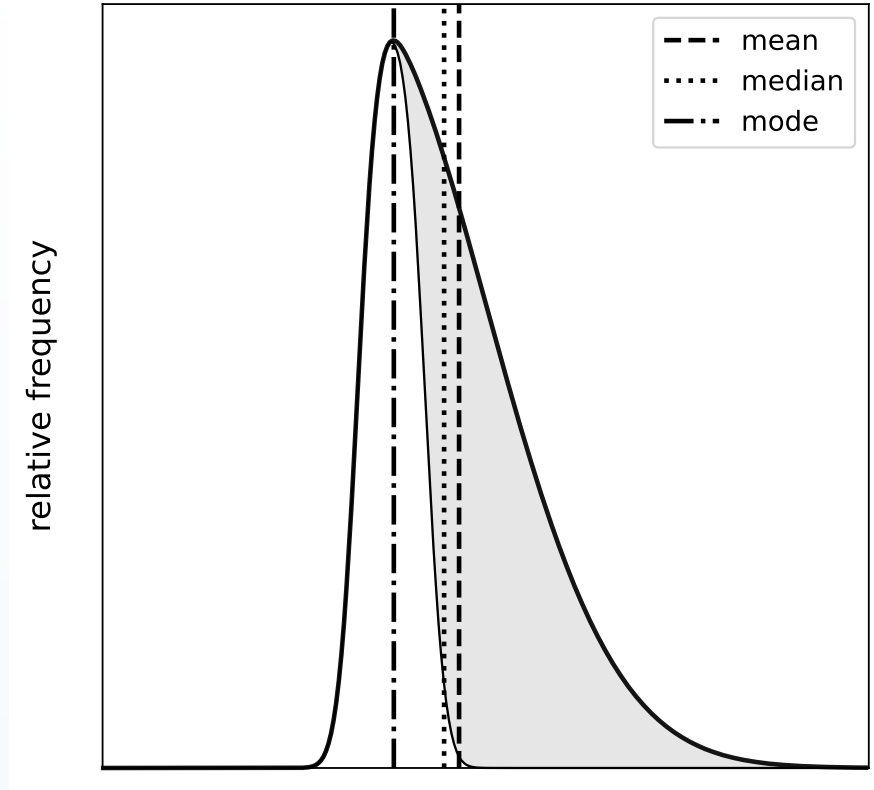


*probably.

- Can infer dispersion from shape of distribution
- Example: increasing standard deviation (σ):
 - Lower peak
 - Wider curve
- In general:
 - Low dispersion: high, narrow peak
 - High dispersion: short, wide peak

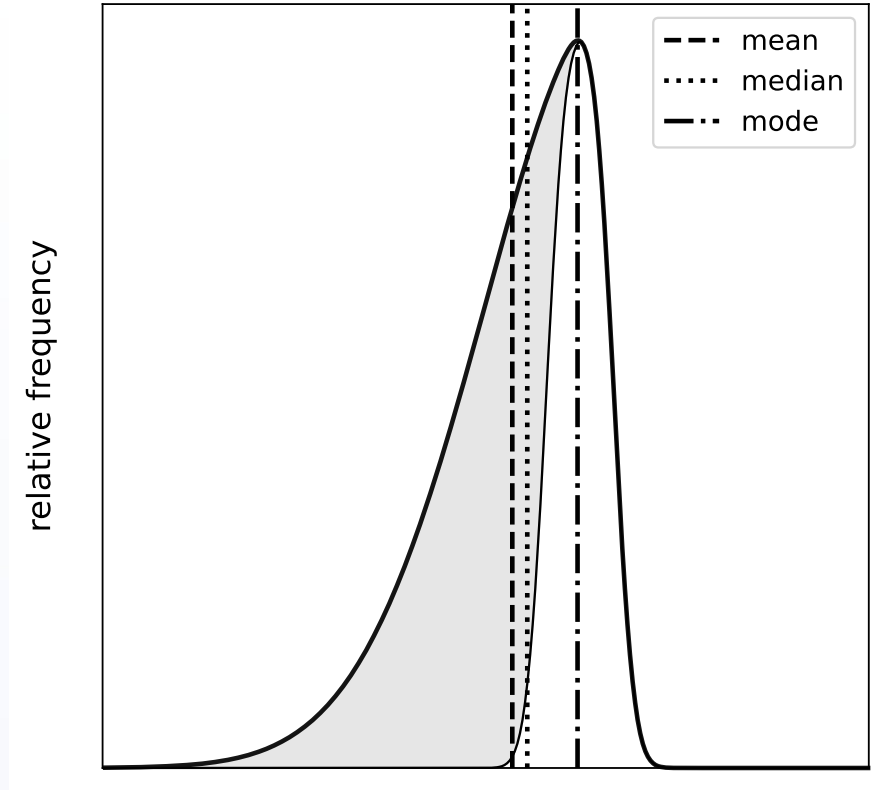


- “Tail” of excess values to the right:
 - “right-skew”
 - Left-leaning
- Typically (but not always!):
 - $\text{mean} > \text{median} > \text{mode}$



Negative Skew

- “Tail” of excess values to the left
 - “left-skew”
 - Right-leaning
- Typically (but not always!):
 - $\text{mean} < \text{median} < \text{mode}$



Measures of Skewness

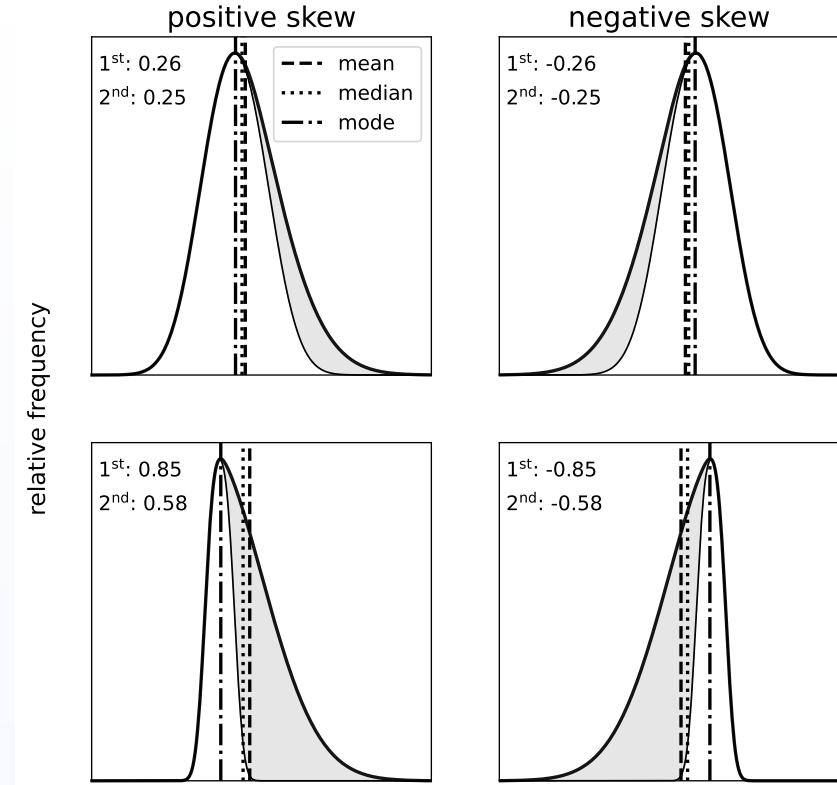
- Goal: compare skewness of different distributions
- Pearson's (1st) coefficient (mode skewness):

$$K_1 = \frac{\mu - \text{mode}}{\sigma}$$

- Pearson's (2nd) coefficient (median skewness):

$$K_2 = \frac{3(\mu - \text{median})}{\sigma}$$

- Values range between -1, 1 (-3, 3)
- NB: not completely interchangeable!



- Shape of data/frequency distribution is important
- Symmetrical:
 - Left side mirrors right side
 - Mean/median are same (perfectly symmetrical)
- If data are not symmetrical: skewed
 - Use relationship between mean, median, mode

- Illowsky and Dean, Chapters 2.2, 2.6
- Caswell, Chapters 7, 8
- Weiss, Chapter 2.4