Mathematics and Biostatistics

Distribution of Values - 3

1st Semester 2022

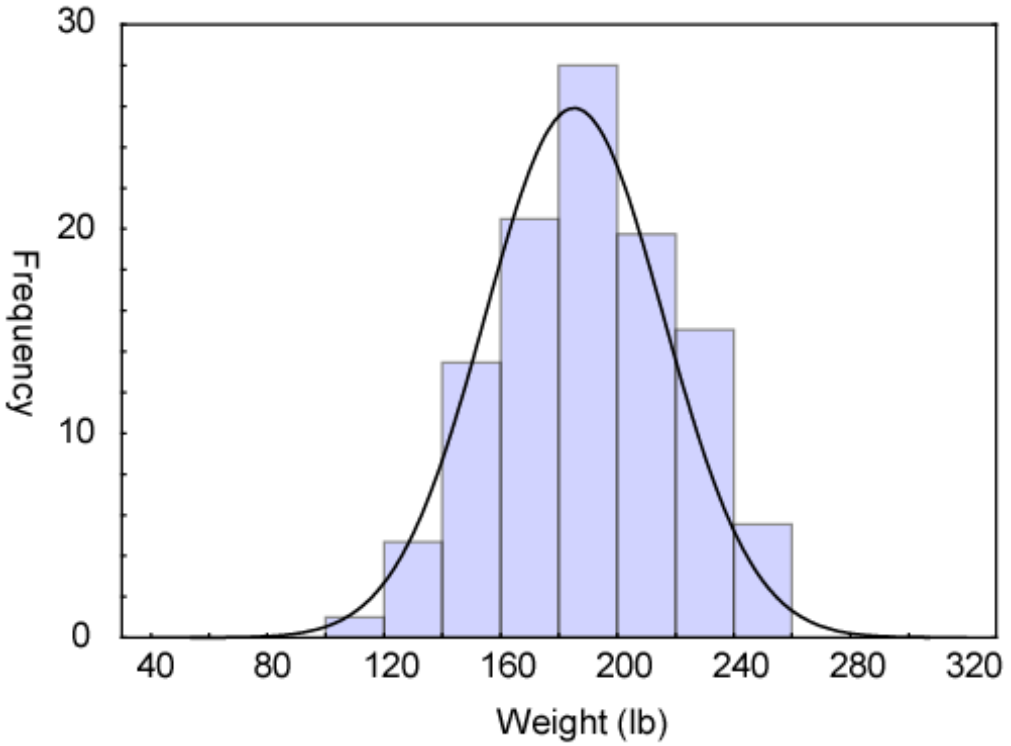
Lecture 4

Review

- Things we look for in a distribution
  1. Central Tendency
     a. MEAN -- average
     b. MEDIAN -- middle value
     c. MODE -- most frequently observed value.
  2. Variability of values and their spread
     a. Quartiles -- Median
     b. Standard deviation -- Mean
  3. Shape of the distribution.
  4. Gaps and clumping of values

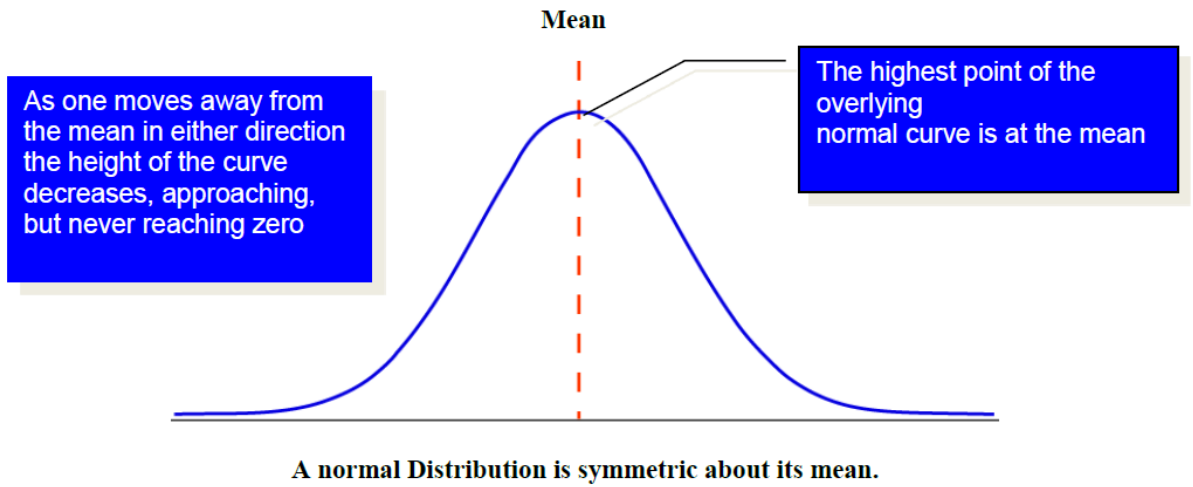
Shape of the distribution

- The plot of frequency distributions have shape.
- Distributions are often symmetrical with most scores falling in the middle and fewer toward the extremes.
- Most biological data are symmetrically distributed and form a normal curve, (bell-shaped curve).

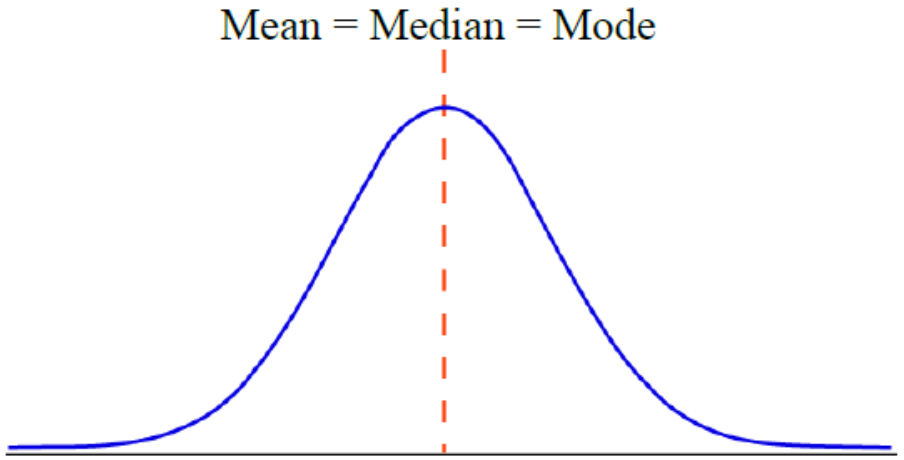


The normal distribution

- The area under a normal curve has a normal distribution (Gaussian distribution).
- Properties of a normal distribution:
  1. It is symmetric about its mean.
  2. The highest point is at its mean.
- The height of the curve decreases as one moves away from the mean in direction,
- approaching, but never reaching zero.

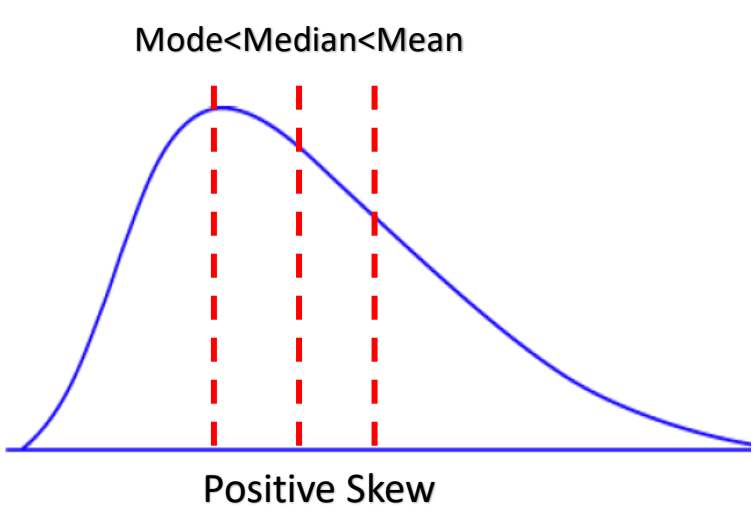


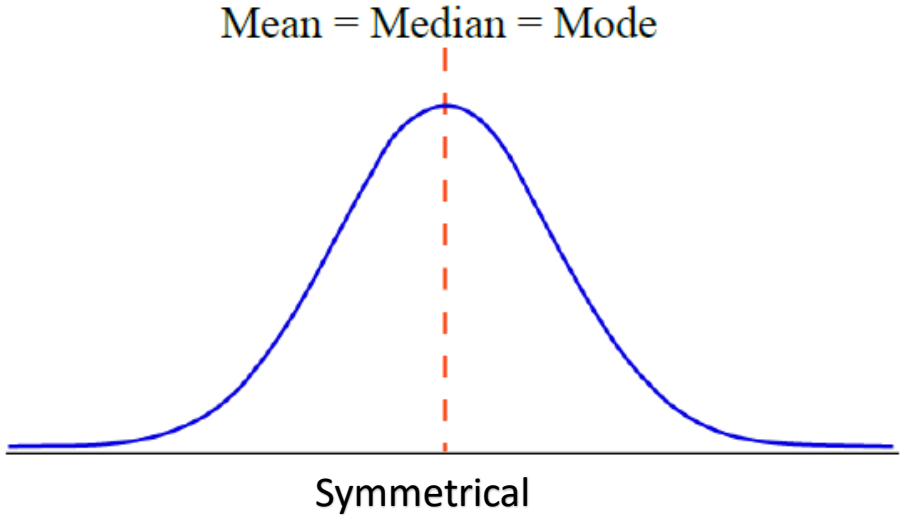
A normal Distribution is symmetric about its mean.

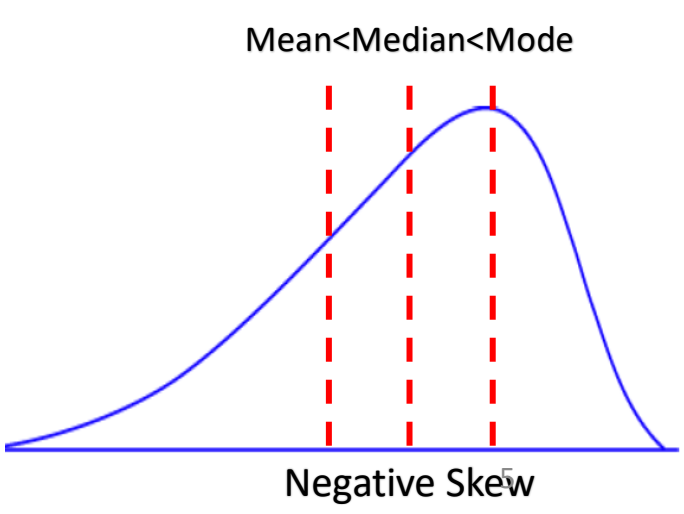


Skewed distributions

- The data are not distributed symmetrically in skewed distributions
- The mean, median, and mode are not equal and are in different positions.
- Scores are clustered at one end of the distribution.
- A small number of extreme values are located in the limits of the opposite end.

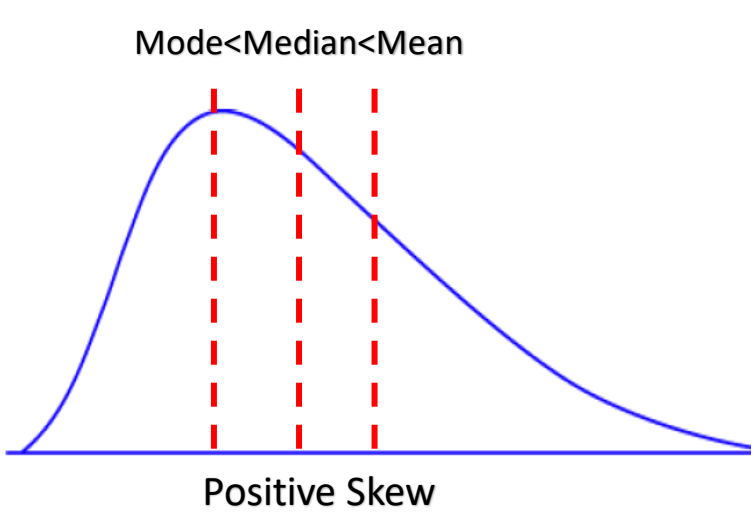


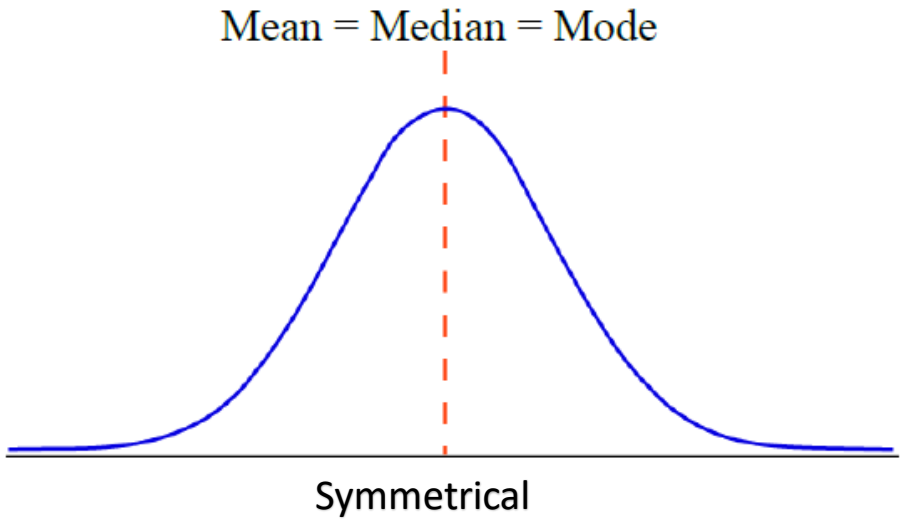


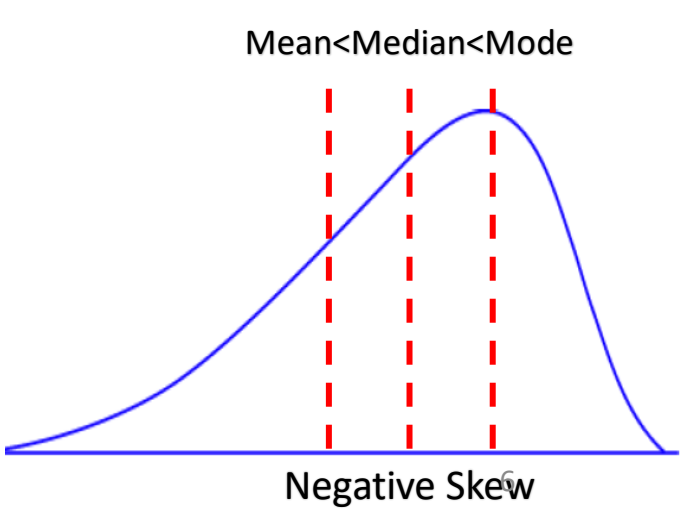


Skewed distributions

- Because the mean is shifted so much, it is not the best estimate of the average score for skewed distributions.
- The median is a better estimate of the center of skewed distributions
- It will be the central point of any distribution 50% of the values are above and 50% below the median.

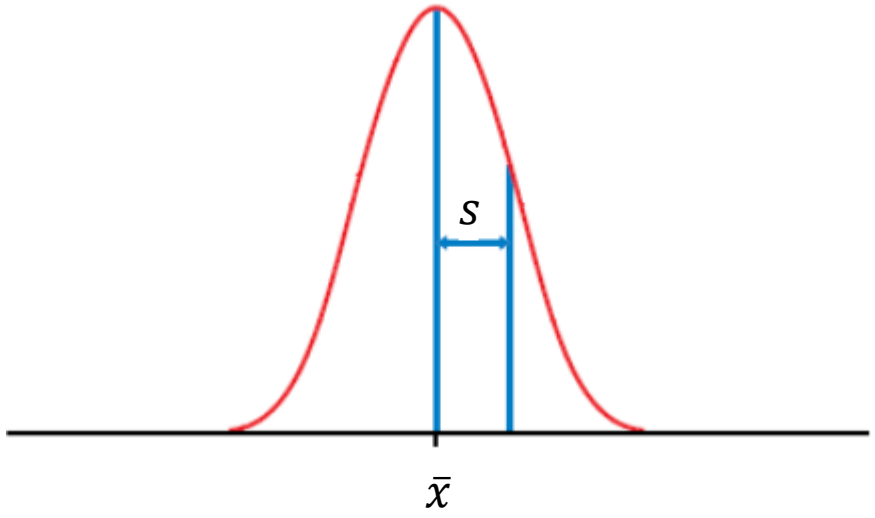






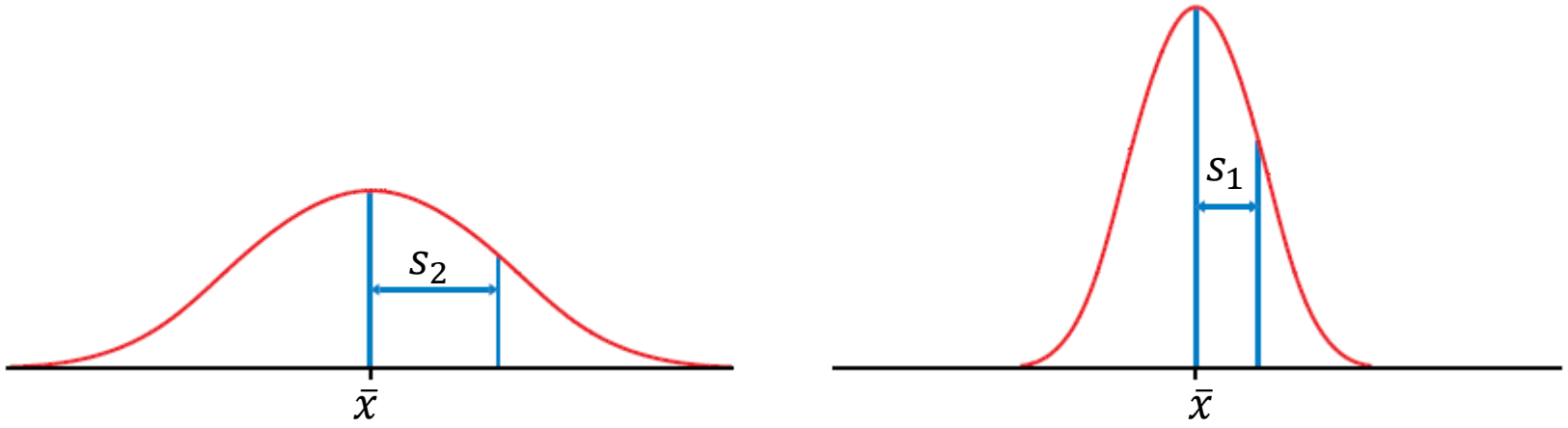
Normal distributions

- One particularly important class of density curves is the Normal (bell-shaped) curves.
- These curves are symmetric, unimodal.
- They are called Normal curves, and they describe Normal distributions.
- All Normal distributions have the same overall shape.
- The particular Normal distribution is specified by giving its mean $\bar{x}$ and its standard deviation s.
- The mean is located at the center of the symmetric curve and is the same as the median.



Normal distributions

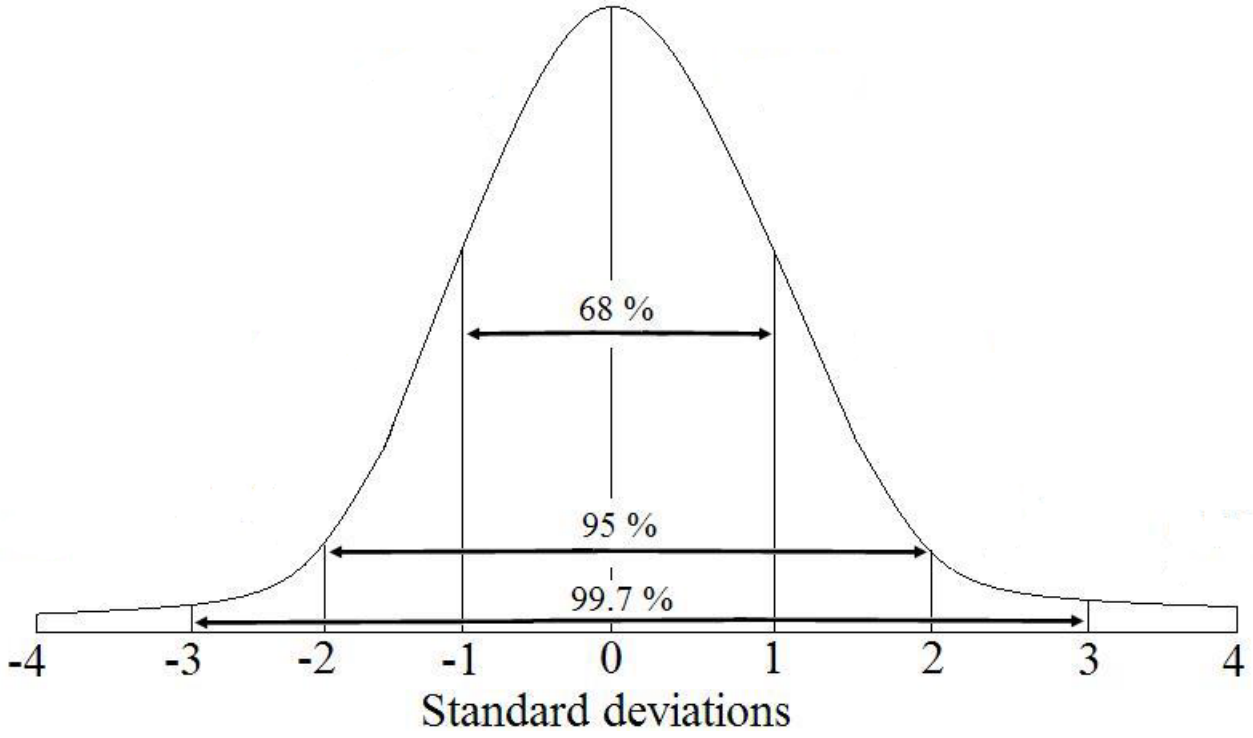
- Changing $\bar{x}$ without changing s moves the Normal curve along the horizontal axis without changing its spread.
- The standard deviation s controls the spread of a Normal curve.
- Two Normal curves with different values of s , $(s_1, and\ s_2)$. The curve with the larger standard deviation is more spread out.



The 68–95–99.7 rule

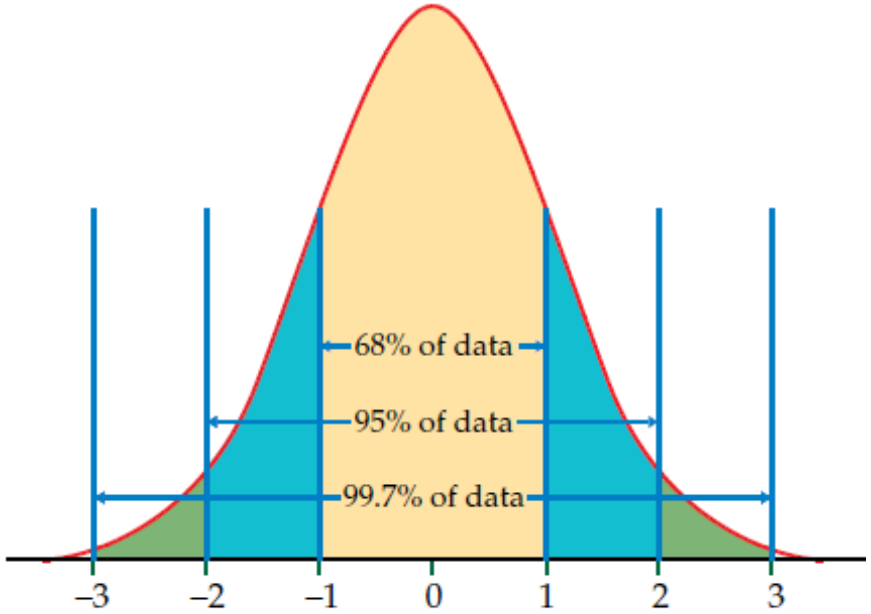
In the Normal distribution with mean $\bar{x}$ and standard deviation s:

- Approximately 68% of the observations fall within s of the mean $\bar{x}$.
- Approximately 95% of the observations fall within $2s$ of $\bar{x}$.
- Approximately 99.7% of the observations fall within $3s$ of $\bar{x}$.



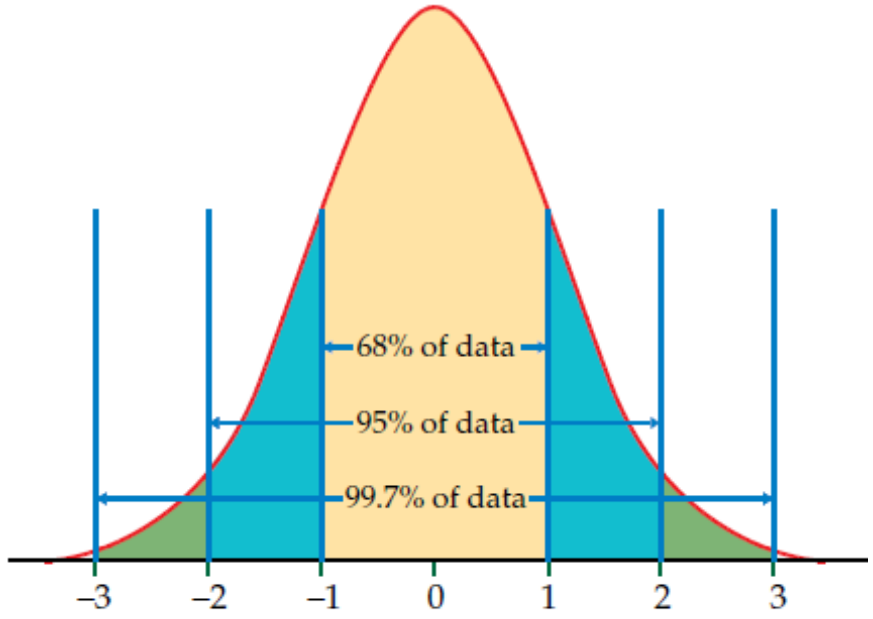
Weights of young women

- The distribution of weights of young women aged 18 to 24 is approximately Normal with mean $\bar{x} = 64.5\,Kg$ and standard deviation $s = 2.5\,Kg$. Use the $68-95-99.7$ rule to give a range of weights that includes $68\%, 95\%, 99.7\%$ of these women.

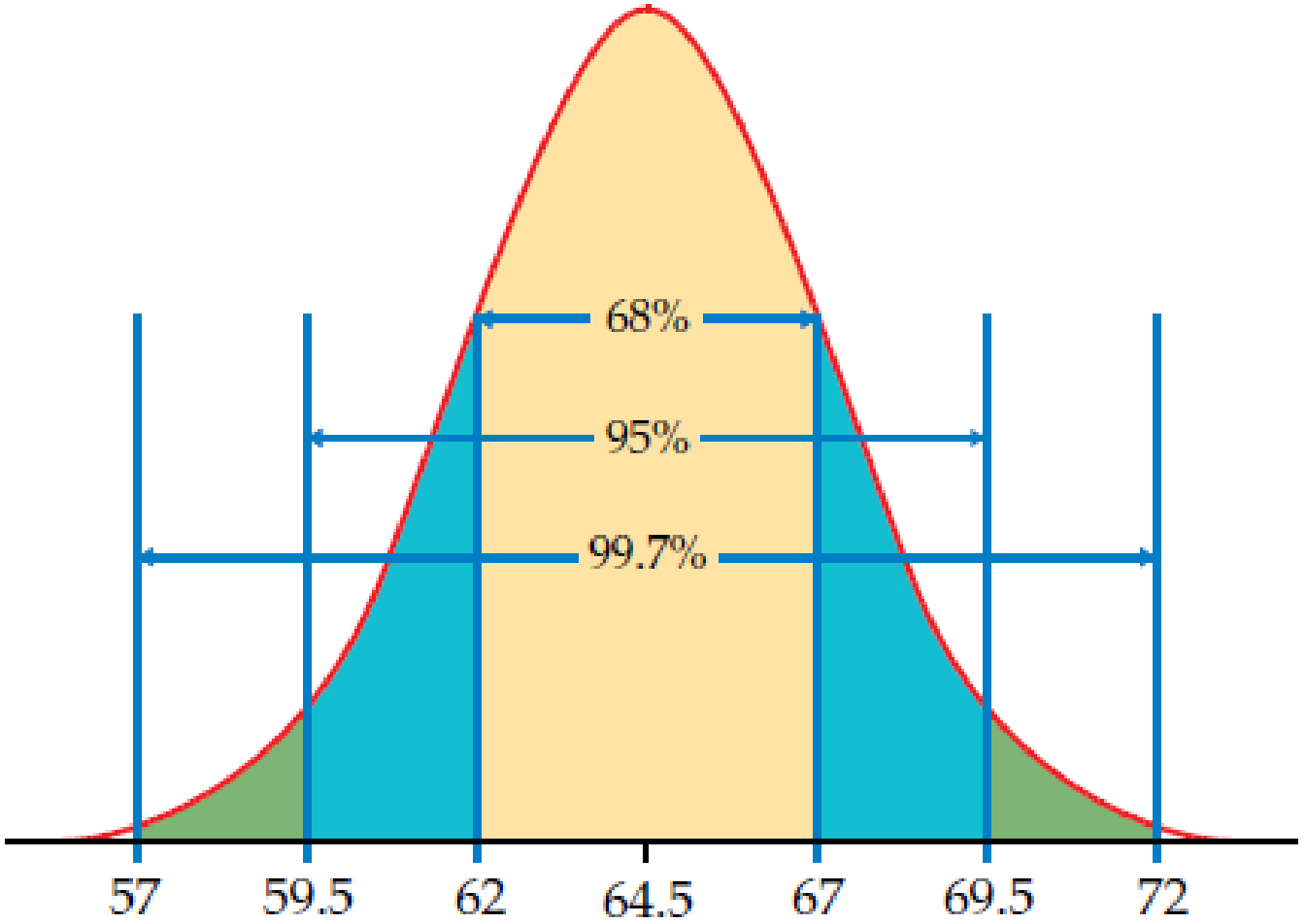


Weights of young women

- One standard deviation is $2.5\,Kg$ for this distribution. The 68 part of the $68–95–99.7$ rule says that the middle 68% of young women are between $64.5 - 2.5$ and $64.5 + 2.5\text{Kg}$, that is, between $62\,Kg$ and $67\,Kg$.
- Two standard deviations is $5\,Kg$ for this distribution. The 95 part of the 68–95–99.7 rule says that the middle 95% of young women are between 64.5–5 and 64.5+5Kg, that is, between $59.5\,Kg$ and $69.5\,Kg$.
- Three standard deviations is $7.5\,Kg$ for this distribution. The 99.7 part of the 68–95–99.7 rule says that the middle 99.7% of young women are between 64.5–7.5 and 64.5+7.5Kg, that is, between $57\,Kg$ and $72\,Kg$.



Weights of young women



Normal distributions importance in statistics

- First, Normal distributions are good descriptions for some distributions of real data. Distributions that are often close to Normal include scores on tests taken by many people ,repeated careful measurements of the same quantity, and characteristics of biological populations
- Second, Normal distributions are good approximations to the results of many kinds of chance outcomes, such as tossing a coin many times.
- Third, and most important, we will see that many statistical inference procedures based on Normal distributions work well for other roughly symmetric distributions.

The End of Lecture