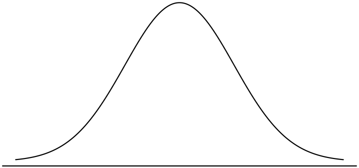
Lecture 5:
Normal Distribution
STAT 310, Spring 2021

- The normal distribution is one of the most common and important probability distributions.
- It is symmetric, unimodal, and bell-curve shaped.
- Many phenomena in nature approximately follow a normal distribution such as the height of individuals, the velocity in any direction of a molecule of gas, and measurement error.



- The normal distribution curve is a mathematical abstraction.
- Just as there is no such thing as a perfect circle, no real data set perfectly follows a normal distribution.
- However, many data sets *approximately* follow a normal distribution, and so the normal distribution provides a very useful approximation for a variety of problems.

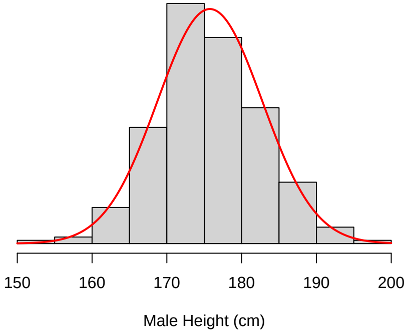
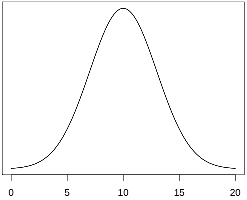


Figure: Histogram of male heights (cm) with normal distribution curve. We see that the distribution of height is approximately normal.

- The normal distribution is characterized by two parameters: the mean, μ, and standard deviation, σ.
- The mean specifies the center of the distribution. Changing the value of the mean shifts the bell-curve to the left or right.
- The standard deviation specifies the spread of the distribution. Changing the value of the standard deviation stretches or constricts the bell-curve.

- The notation $X \sim N(\mu, \sigma)$ means that the random variable X follows a normal distribution with mean μ and standard deviation σ.
- For example, the plot below shows the distribution of $N(\mu = 10, \sigma = 3)$



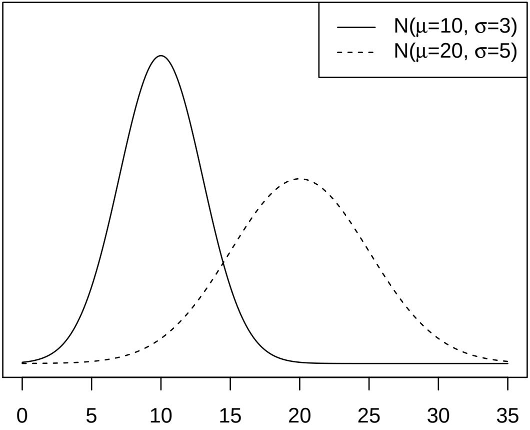
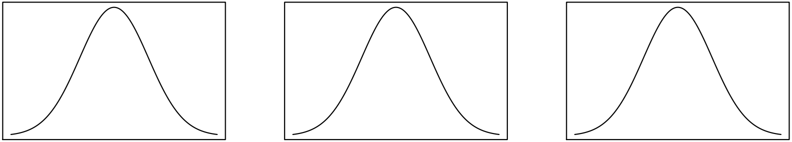
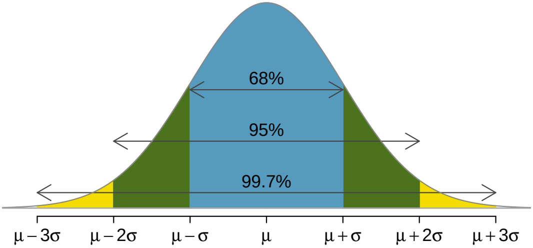


Figure: Plot of two normal distributions.

- Probabilities are computed as the area under the normal distribution curve.
- The total area under the normal distribution curve is always 1.



Empirical Rule



- About 68% of the distribution is contained within 1 standard deviation of the mean.
- About 95% of the distribution is contained within 2 standard deviations of the mean.
- About 99.7% of the distribution is contained within 3 standard deviations of the mean.

Standardizing with z-scores

- The normal distribution with mean $\mu = 0$ and standard deviation $\sigma = 1$ is called the **standard normal distribution** or **Z-distribution**.
- If x is an observation from $N(\mu, \sigma)$, we define the z-score as

$$z = \frac{x - \mu}{\sigma}$$

Standardizing with z-scores

- A z-score can be interpreted as the number of standard deviations an observation x lies away from the mean.
  - For instance, if a student has a z-score of 2 on an exam then that student is 2 standard deviations *above* the average score.
  - If a student has a z-score of -1.5 on an exam then that student is 1.5 standard deviations *below* the average score.

Example 1

The SAT score X of a students is normally distributed with mean $\mu = 1100$ and standard deviation $\sigma = 200$.

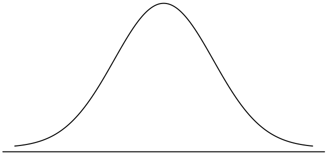
(a) Calculate and interpret the z-score for a student that scored a 1350 on the SAT.

(b) Calculate and interpret the z-score for a student that scored a 900 on the SAT.

Example 2

The amount X of the pollutant nitrogen oxide in automobiles is normally distributed with mean $\mu = 70$ ppb (parts per billion) and standard deviation $\sigma = 13$ ppb. That is, $X \sim N(\mu = 70, \sigma = 13)$.

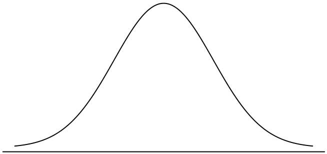
(a) What is the probability that a randomly selected vehicle will have an emission level less than 60 ppb?



Example 2

The amount X of the pollutant nitrogen oxide in automobiles is normally distributed with mean $\mu = 70$ ppb (parts per billion) and standard deviation $\sigma = 13$ ppb. That is, $X \sim N(\mu = 70, \sigma = 13)$.

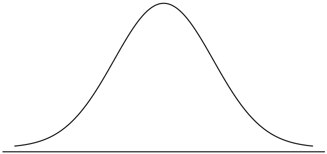
(b) What is the probability that a randomly selected vehicle will have an emission level greater than 90 ppb?



Example 2

The amount X of the pollutant nitrogen oxide in automobiles is normally distributed with mean $\mu = 70$ ppb (parts per billion) and standard deviation $\sigma = 13$ ppb. That is, $X \sim N(\mu = 70, \sigma = 13)$.

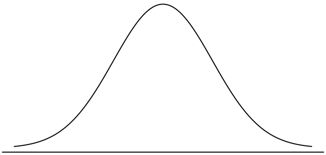
(c) What is the probability that a randomly selected vehicle will have an emission level between 60 and 90 ppb?



Example 3

Body temperatures are normally distributed with mean $\mu = 98.2$ and standard deviation $\sigma = 0.74$, in degrees Fahrenheit.
That is, $X \sim N(98.2, 0.74)$.

(a) Find the cutoff for the lowest 5% of body temperatures (the 5^{th} percentile)?



Example 3

Body temperatures are normally distributed with mean $\mu = 98.2$ and standard deviation $\sigma = 0.74$, in degrees Fahrenheit. That is, $X \sim N(98.2, 0.74)$.

(b) Find the cutoff for the highest 15% of body temperatures (the 85^{th} percentile)?