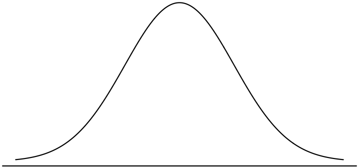
Lecture 5:
Normal Distribution
STAT 310, Spring 2021

- 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.

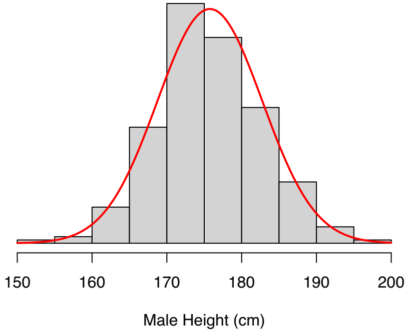
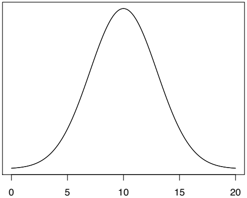


Figure: Histogram of male heights (cm) with normal distribution curve. We see that the distribution of height is approximately normal.

- 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 $N(\mu = 10, \sigma = 3)$



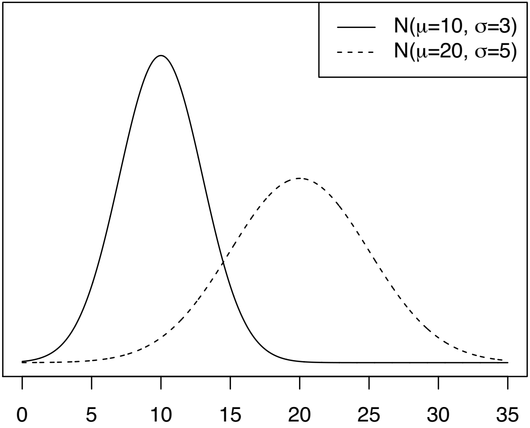
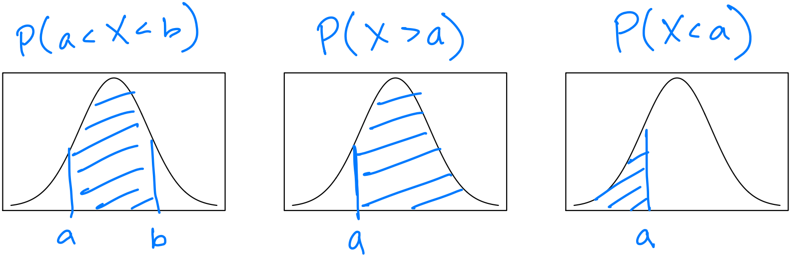
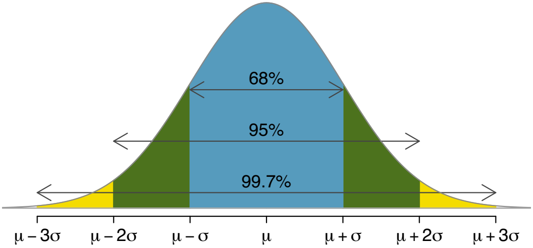


Figure: Plot of two normal distributions.

- 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

$N(\mu = 0, \sigma = 1)$

- 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.

$$z = \frac{X - \mu}{\sigma} = \frac{1350 - 1100}{200} = 1.25$$

student scored 1.25 st dev above average SAT score

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

$$z = \frac{X - \mu}{\sigma} = \frac{900 - 1100}{200} = -1$$

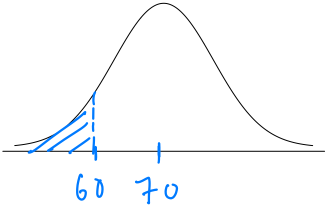
student scored 1 st dev below average SAT score

Example 2

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

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?



$$z = \frac{60-70}{13} = -0.77$$

$$P(X<60) = P(Z<-0.77)$$

$$= \text{pnorm}(-0.77)$$

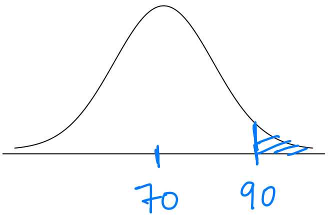
$$= \boxed{0.22}$$

pnorm( ) is R function for computing <u>probabilities</u> from a normal distribution

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?



$$z = \frac{90-70}{13} = 1.54$$

$$P(X > 90) = 1 - P(X < 90)$$

$$= 1 - P(Z < 1.54)$$

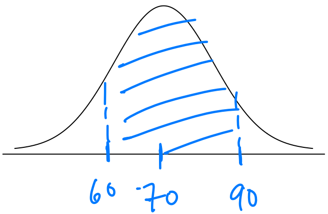
$$= 1 - \text{pnorm}(1.54)$$

$$= \boxed{0.062}$$

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?



$$z = \frac{60-70}{13} = -0.77 \qquad z = \frac{90-70}{13} = 1.54$$

$$P(60 < X < 90)$$
$$= P(X < 90) - P(X < 60)$$
$$= P(Z < 1.54) - P(Z < -0.77)$$
$$= \text{pnorm}(1.54) - \text{pnorm}(-0.77)$$
$$= \boxed{0.72}$$

Example 3

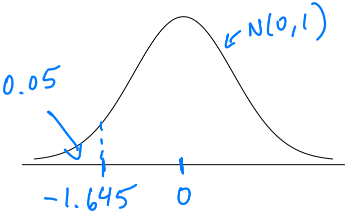
qnorm( ) ⇒ R function for computing percentiles from normal distribution

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)?

Solve for x:

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



$$\Rightarrow -1.645 = \frac{x - 98.2}{0.74}$$

$$\Rightarrow (-1.645)(0.74) = x - 98.2$$

$$\Rightarrow x = 98.2 - 1.645(0.74)$$

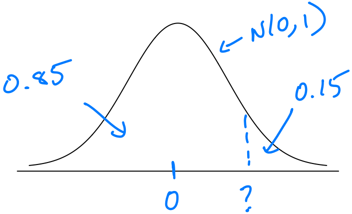
$$= \boxed{96.98}$$

$z = \text{qnorm}(0.05) = -1.645$

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)?



$z = \text{qnorm}(0.85)$
$= 1.036$

Solve for x:

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

$$1.036 = \frac{x - 98.2}{0.74}$$

$$\Rightarrow 1.036(0.74) = x - 98.2$$

$$\Rightarrow x = \underline{98.2 + 1.036(0.74)}$$

$$= \boxed{98.97}$$