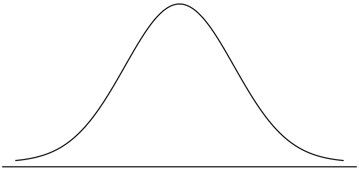
Lecture 5:
Normal Distribution
STAT 310, Spring 2023

- 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.

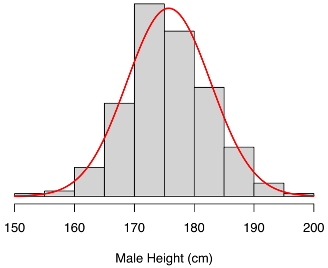
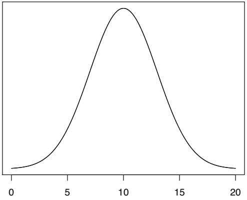


Figure: Histogram of male heights (cm) with normal distribution curve. We see that the distribution of height is approximately normal.

- 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.[1]

- For example, the plot below shows the distribution of $N(\mu = 10, \sigma = 3)$.



[1] Informally, a "random variable" is a variable that takes on numerical values that represent outcomes of a random process, such as the height or IQ of a randomly selected person.

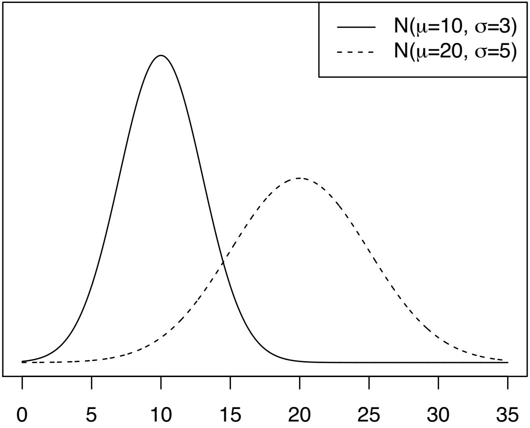
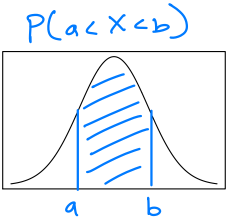
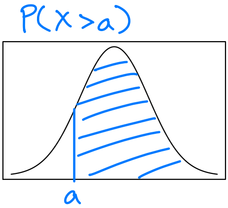
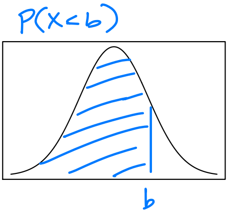


Figure: Plot of two normal distributions.

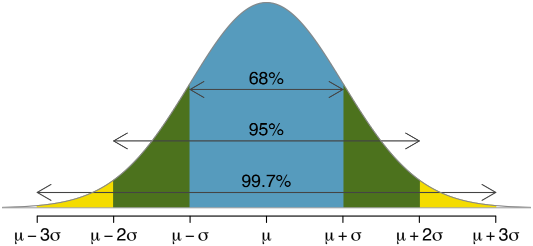
- Probabilities are computed as the area under the normal distribution curve.
- The total area under the normal distribution curve is always 1.







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

SAT scores are normally distributed with mean $\mu = 1100$ and standard deviation $\sigma = 200$.

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

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

Student scored 1.25 standard deviations **above** average SAT score.

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

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

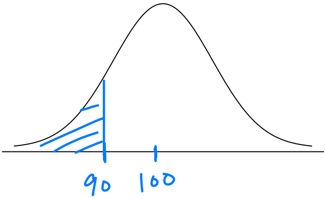
Student scored 1 standard deviation **below** average SAT score.

Example

pnorm() is an R function for computing probabilities from a normal distribution

Scores on an IQ test follow a normal distribution with mean $\mu = 100$ and standard deviation $\sigma = 15$. That is, $X \sim N(\mu = 100, \sigma = 15)$.

(a) What is the probability that a person has an IQ less than 90?



X = score on IQ test

$$z = \frac{90-100}{15} = -0.67$$

$$P(X<90)$$

$$= P(Z<-0.67)$$

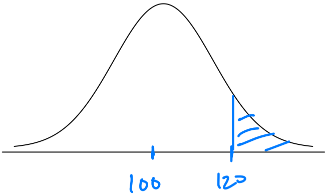
$$= \text{pnorm}(-0.67)$$

$$= \boxed{0.251}$$

Example

Scores on an IQ test follow a normal distribution with mean $\mu = 100$ and standard deviation $\sigma = 15$. That is, $X \sim N(\mu = 100, \sigma = 15)$.

(b) What is the probability that a person has an IQ greater than 120?



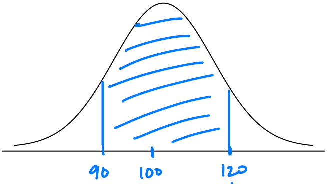
$$z = \frac{120 - 100}{15} = 1.33$$

$$\begin{aligned} P(X > 120) &= 1 - P(X < 120) \\ &= 1 - P(Z < 1.33) \\ &= 1 - \text{pnorm}(1.33) \\ &= \boxed{0.092} \end{aligned}$$

Example

Scores on an IQ test follow a normal distribution with mean $\mu = 100$ and standard deviation $\sigma = 15$. That is, $X \sim N(\mu = 100, \sigma = 15)$.

(c) What is the probability that a person has an IQ between 90 and 120?



$z = \dfrac{90 - 100}{15} = -0.67$

$z = \dfrac{120 - 100}{15} = 1.33$

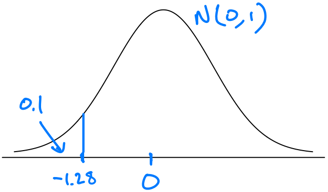
$$P(90 < X < 120)$$
$$= P(X < 120) - P(X < 90)$$
$$= P(z < 1.33) - P(z < -0.67)$$
$$= \text{pnorm}(1.33) - \text{pnorm}(-0.67)$$
$$= \boxed{0.657}$$

Example

qnorm() is an R function for computing percentiles (or quantiles) from a normal distribution

Scores on an IQ test follow a normal distribution with mean $\mu = 100$ and standard deviation $\sigma = 15$. That is, $X \sim N(\mu = 100, \sigma = 15)$.

(d) What is the cutoff for the lowest 10% of all IQ scores (i.e., find the 10th percentile)?



$z = \text{qnorm}(0.1) = -1.28$

Solve for x:

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

$$\Rightarrow -1.28 = \frac{x - 100}{15}$$

$$\Rightarrow (-1.28)(15) = x - 100$$

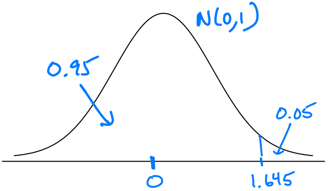
$$\Rightarrow x = 100 - 1.28(15)$$

$$= \boxed{80.8}$$

Example

Scores on an IQ test follow a normal distribution with mean $\mu = 100$ and standard deviation $\sigma = 15$. That is, $X \sim N(\mu = 100, \sigma = 15)$.

(e) What is the cutoff for the highest 5% of all IQ scores (i.e., find the 95th percentile)?



$z = \text{qnorm}(0.95) = 1.645$

Solve for x:

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

$$\Rightarrow 1.645 = \frac{x - 100}{15}$$

$$\Rightarrow 15(1.645) = x - 100$$

$$\Rightarrow x = 100 + 15(1.645)$$

$$= \boxed{124.7}$$