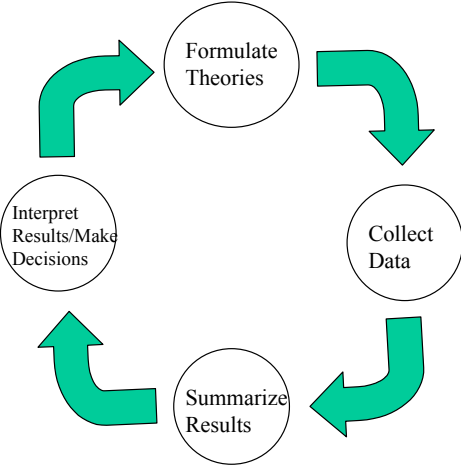
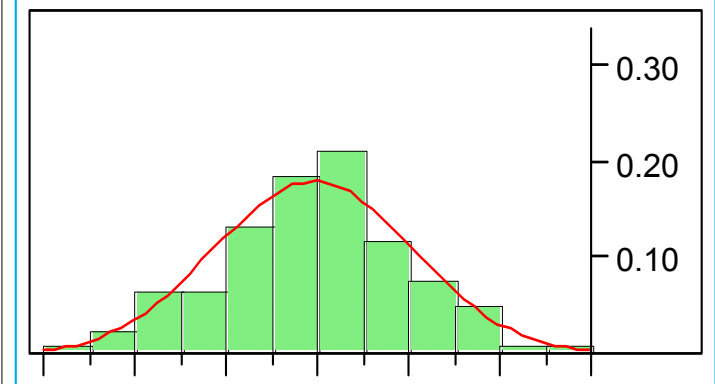


Chapter Goals

How to use models to make decisions.

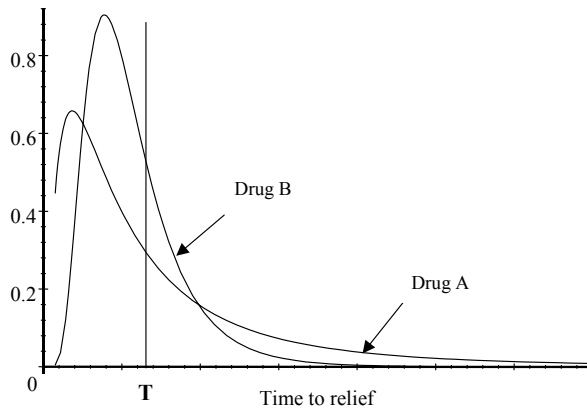


Why model?



Data: Sample
Model: Population

Example Suppose we wish to compare two drugs, Drug A and Drug B, for relieving arthritis pain. Subjects suitable for the study are randomly assigned one of the two drugs. Results of the study are summarized in the following model for the time to relief for the two drugs.



Random Variables

Definition A variable is *continuous* if it can take on any value on a continuum.

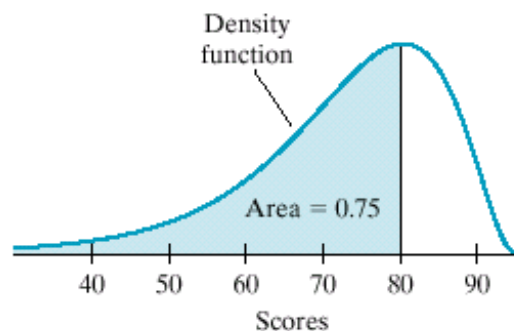
Definition A variable is *discrete* if it assumes only a countable number of values.

Example The number of customers served in a restaurant on any particular day.

Example The time it takes to come to school.

Modeling Continuous Variables

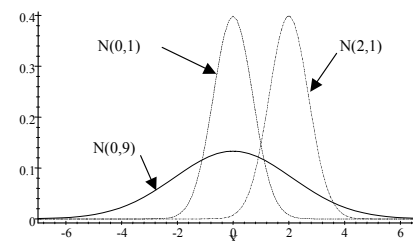
Definition A density function is a (nonnegative) function or curve that describes the overall shape of a distribution. The total area under the entire curve is equal to **one**, and proportions are measured as areas under the density function.



Normal Distribution

Definition $X \sim N(\mu, \sigma^2)$ means that the variable or characteristic X is normally distributed with mean μ and variance σ^2 (or standard deviation σ).

Definition $X \sim N(\mu, \sigma^2)$, then the standardized normal variable $Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$. Z is called the **standard normal**.

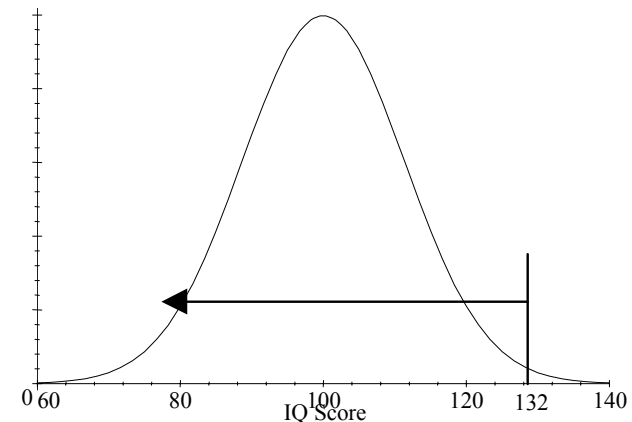


Properties of the Normal Distribution

- Symmetric about the mean μ .
- Bell-shaped
- Mean = Median = Mode
- Approximately 68% of the area under the curve is within ± 1 standard deviation of the mean.
- Approximately 95% of the area under the curve is within ± 2 standard deviation of the mean.
- Approximately 99.7% of the area under the curve is within ± 3 standard deviation of the mean.

Note: Any normal distribution $N(\mu, \sigma^2)$ can be transformed to a standard normal distribution $N(0, 1)$.

Example Let the variable X represent IQ scores of 12-year-olds. Suppose $X \sim N(100, 256)$. Jessica is a 12-year-old and has an IQ score of 132. What proportion of 12-year-olds have IQ scores less than Jessica's score of 132?



Let's do it! 6.1

- Find the area under a standard normal distribution between $z = 0$ and $z = 1.22$.

- Find the area under a standard normal distribution to the left of $z = -2.55$.

- Find the area under a standard normal distribution between $z = -1.22$ and $z = 1.22$.

Let's do it! 6.2

Consider the previous IQ Scores example, where $X \sim N(100, 256)$.

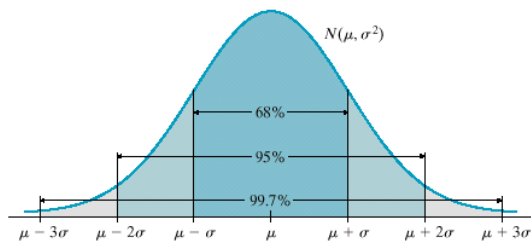
- What proportion of the 12-year-olds have IQ scores below 84?

- What proportion of the 12-year-olds have IQ scores 84 or more?

- What proportion of the 12-year-olds have IQ scores between 84 and 116?

Definition *The 68-95-99 rule for any normal distribution $N(\mu, \sigma^2)$:*

- Approximately 68% of the area under the curve is within ± 1 standard deviation of the mean $[\mu - \sigma, \mu + \sigma]$.
- Approximately 95% of the area under the curve is within ± 2 standard deviation of the mean $[\mu - 2\sigma, \mu + 2\sigma]$
- Approximately 99.7% of the area under the curve is within ± 3 standard deviation of the mean $[\mu - 3\sigma, \mu + 3\sigma]$



Exercise *Suppose cholesterol measures for healthy individuals have a normal distribution. Kyle's standardized cholesterol measure was $z = -2$. Using the 68-95-99 rule, what percentile does Kyle's measure represent?*

Exercise *Lee's standardized cholesterol measure was $z = 3.2$. Does Lee's cholesterol seem unusually high?*

Let's do it! 6.4

Different species of pine trees are grown at a Christmas-tree farm. It is known that the length of needles on a Species A pine tree follows a normal distribution. About 68% of such needles have lengths centered around the mean between 5.9 and 6.9 inches.

1. What are the mean and standard deviation of the model for Species A pine needle lengths?

2. A 5.2-inch pine needle is found that looks like a Species A needle but is somewhat shorter than expected. Is it likely that this needle is from a Species A pine tree?

Let's do it! 6.6

The finishing times for swimmers performing the 100-meter butterfly are normally distributed with a mean of 55 seconds and a standard deviation of 5 seconds.

1. The sponsors decide to give certificates to all those swimmers who finish in under 49 seconds. If there are 50 swimmers entered in the 100-meter butterfly, approximately how many certificates will be needed?

2. What time must a swimmer finish to be in the top fastest 2% of the distribution of finishing times?

Assessing Normality

- Look at a histogram or stem-and-leaf plot and check for non-normal features such as gaps, outliers, or pronounced skewness.
- Compare your observations to the 68-95-99 rule for normal distributions.
- Plot P-P and Q-Q plots.
- Calculate measures for Normality.

Modeling Discrete Variables

Definition *The probability mass function of a discrete variable is a graph, table, or formula that specifies the proportion associated with each possible value the variable can take. The mass function $p(X = x)$ (or just $p(x)$) has the following properties:*

- *All values of the discrete function $p(x)$ must be between 0 and 1, both inclusive, and*
- *if you add up all values, they should sum to 1.*

Example *Let X represent the number of books in a backpack for students enrolled at a particular college. The table below gives the discrete mass function:*

X	2	3	4
Proportion ($p(X = x)$)	0.5	0.3	0.2

Let's do it! 6.11

Consider the following discrete mass function:

X	-1	2	3	4
$p(X = x)$	0.1		0.6	0.2

1. Complete the above table.
2. $p(X < 3.1)$?

3. $p(X \geq -1.1)$?

4. $p(2 < X < 3)$?