Section 8.1
Random Sampling
Definition 8.1

A population consists of the totality of the observations with which we are concerned.
Definition 8.2

A sample is a subset of a population.
Definition 8.3

Let \( X_1, X_2, \ldots, X_n \) be \( n \) independent random variables, each having the same probability distribution \( f(x) \). Define \( X_1, X_2, \ldots, X_n \) to be a random sample of size \( n \) from the population \( f(x) \) and write its joint probability distribution as

\[
f(x_1, x_2, \ldots, x_n) = f(x_1)f(x_2) \cdots f(x_n).
\]
Section 8.2

Some Important Statistics
Definition 8.4

Any function of the random variables constituting a random sample is called a statistic.
Theorem 8.1

If \( S^2 \) is the variance of a random sample of size \( n \), we may write

\[
S^2 = \frac{1}{n(n-1)} \left[ n \sum_{i=1}^{n} X_i^2 - \left( \sum_{i=1}^{n} X_i \right)^2 \right].
\]
Section 8.3

Sampling Distributions
Definition 8.5

The probability distribution of a statistic is called a sampling distribution.
Section 8.4

Sampling Distribution of Means and the Central Limit Theorem
Theorem 8.2

Central Limit Theorem: If $\bar{X}$ is the mean of a random sample of size $n$ taken from a population with mean $\mu$ and finite variance $\sigma^2$, then the limiting form of the distribution of

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}},$$

as $n \to \infty$, is the standard normal distribution $n(z; 0, 1)$. 
Figure 8.1 Illustration of the Central Limit Theorem (distribution of $\bar{X}$ for $n = 1$, moderate $n$, and large $n$)
Figure 8.2 Area for Example 8.4

\[ \sigma_{\bar{x}} = 10 \]
Figure 8.3  Area for Case Study 8.1
Figure 8.4  Area for Example 8.5
Theorem 8.3

If independent samples of size $n_1$ and $n_2$ are drawn at random from two populations, discrete or continuous, with means $\mu_1$ and $\mu_2$ and variances $\sigma_1^2$ and $\sigma_2^2$, respectively, then the sampling distribution of the differences of means, $\bar{X}_1 - \bar{X}_2$, is approximately normally distributed with mean and variance given by

$$
\mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2 \quad \text{and} \quad \sigma^2_{\bar{X}_1 - \bar{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.
$$

Hence,

$$
Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}
$$

is approximately a standard normal variable.
Figure 8.5  Area for Case Study

\[ \sigma \bar{x}_A - \bar{x}_B = \sqrt{1/9} \]

[Diagram showing a normal distribution with the mean of the difference between \( \mu_A \) and \( \mu_B \) set to 0, and the standard deviation of the difference between \( \bar{x}_A \) and \( \bar{x}_B \) set to \( \sqrt{1/9} \).]
Figure 8.6 Area for Example 8.6

\[ \sigma_{\bar{x}_1 - \bar{x}_2} = 0.189 \]

\( \bar{x}_1 - \bar{x}_2 \)
Section 8.5

Sampling Distribution of $S^2$
Theorem 8.4

If $S^2$ is the variance of a random sample of size $n$ taken from a normal population having the variance $\sigma^2$, then the statistic

$$\chi^2 = \frac{(n - 1)S^2}{\sigma^2} = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{\sigma^2}$$

has a chi-squared distribution with $v = n - 1$ degrees of freedom.
Figure 8.7  The chi-squared distribution
Section 8.6

$t$-Distribution
Theorem 8.5

Let $Z$ be a standard normal random variable and $V$ a chi-squared random variable with $v$ degrees of freedom. If $Z$ and $V$ are independent, then the distribution of the random variable $T$, where

$$T = \frac{Z}{\sqrt{V/v}},$$

is given by the density function

$$h(t) = \frac{\Gamma[(v + 1)/2]}{\Gamma(v/2)\sqrt{\pi v}} \left(1 + \frac{t^2}{v}\right)^{-(v+1)/2}, \quad -\infty < t < \infty.$$  

This is known as the $t$-distribution with $v$ degrees of freedom.
Corollary 8.1

Let $X_1, X_2, \ldots, X_n$ be independent random variables that are all normal with mean $\mu$ and standard deviation $\sigma$. Let

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \quad \text{and} \quad S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2.$$ 

Then the random variable $T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$ has a $t$-distribution with $v = n - 1$ degrees of freedom.
Figure 8.8 The $t$-distribution curves for $\nu = 2, 5, $ and $\infty$
Figure 8.9  Symmetry property (about 0) of the $t$-distribution

$t_{1-\alpha} = -t_\alpha$
Figure 8.10  The *t*-values for Example 8.10
Section 8.7

F-Distribution
Theorem 8.6

Let $U$ and $V$ be two independent random variables having chi-squared distributions with $v_1$ and $v_2$ degrees of freedom, respectively. Then the distribution of the random variable $F = \frac{U/v_1}{V/v_2}$ is given by the density function

$$h(f) = \begin{cases} \frac{\Gamma[(v_1+v_2)/2](v_1/v_2)^{v_1/2}}{\Gamma(v_1/2)\Gamma(v_2/2)} \frac{f^{(v_1/2)-1}}{(1+v_1 f/v_2)^{(v_1+v_2)/2}}, & f > 0, \\ 0, & f \leq 0. \end{cases}$$

This is known as the $F$-distribution with $v_1$ and $v_2$ degrees of freedom (d.f.).
Figure 8.11 Typical $F$-distributions

d.f. $= (10, 30)$

d.f. $= (6, 10)$
Figure 8.12  Illustration of the $f_\alpha$ for the $F$-distribution
Writing $f_\alpha(v_1, v_2)$ for $f_\alpha$ with $v_1$ and $v_2$ degrees of freedom, we obtain

$$f_{1-\alpha}(v_1, v_2) = \frac{1}{f_\alpha(v_2, v_1)}.$$
Theorem 8.8

If \( S_1^2 \) and \( S_2^2 \) are the variances of independent random samples of size \( n_1 \) and \( n_2 \) taken from normal populations with variances \( \sigma_1^2 \) and \( \sigma_2^2 \), respectively, then

\[
F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2}
\]

has an \( F \)-distribution with \( v_1 = n_1 - 1 \) and \( v_2 = n_2 - 1 \) degrees of freedom.
Figure 8.13  Data from three distinct samples

|   | A | A | A | A | A | A | B | A | AB | A | B | B | B | B | BBCBB |   | C | C | CC | C | C | C | C |
|   | 4.5 |   |   |   |   |   | 5.5 |   |   |   | 6.5 |   |   |   |   |   |   |   |
|   | \( \bar{x}_A \) |   |   |   |   |   | \( \bar{x}_B \) |   |   |   | \( \bar{x}_C \) |   |   |   |   |   |   |   |

Figure 8.14  Data that easily could have come from the same population

\[ \begin{array}{cccccccc}
A & B & C & A & CB & AC & CAB & C \\
\uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \uparrow \\
X_A & X_C & X_B & A & C & B & A & C
\end{array} \]
Definition 8.6

A quantile of a sample, \( q(f) \), is a value for which a specified fraction \( f \) of the data values is less than or equal to \( q(f) \).
Figure 8.15  Quantile plot for paint data
The normal quantile-quantile plot is a plot of $y_{(i)}$ (ordered observations) against $q_{0.1}(f_i)$, where $f_i = \frac{i - \frac{3}{8}}{n + \frac{1}{4}}$. 
Figure 8.16 Normal quantile-quantile plot for paint data
Table 8.1  Data for Example 8.12

<table>
<thead>
<tr>
<th>Number of Organisms per Square Meter</th>
<th>Station 1</th>
<th>Station 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5,030</td>
<td>2,800</td>
</tr>
<tr>
<td></td>
<td>13,700</td>
<td>4,670</td>
</tr>
<tr>
<td></td>
<td>10,730</td>
<td>6,890</td>
</tr>
<tr>
<td></td>
<td>11,400</td>
<td>7,720</td>
</tr>
<tr>
<td></td>
<td>860</td>
<td>7,030</td>
</tr>
<tr>
<td></td>
<td>2,200</td>
<td>7,330</td>
</tr>
<tr>
<td></td>
<td>4,250</td>
<td>2,190</td>
</tr>
<tr>
<td></td>
<td>15,040</td>
<td>1,690</td>
</tr>
</tbody>
</table>
Figure 8.17 Normal quantile-quantile plot for density data of Example 8.12