Examples: To begin with, let us turn to a few examples and their solutions.

(a) A random sample of size 25 is taken from a normal population having a mean of 80 and a standard deviation of 5. A second random sample of size 36 is taken from a different normal population having a mean of 75 and a standard deviation of 3. Find the probability that the sample mean computed from the 25 measurements will exceed the sample mean computed from the 36 measurements by at least 3.4 but less than 5.9. Assume the means to be measured to the nearest tenth.

\textbf{Sol.} Let } X_1 \text{ denote the sample mean of the samples taken from the first population with mean } \mu_1 = 80 \text{ and standard deviation } \sigma_1 = 5. \text{ Let } X_2 \text{ denote the sample mean of the samples taken from the second population with mean } \mu_2 = 75 \text{ and standard deviation } \sigma_2 = 3. \text{ Here } n_1 = 25 \text{ and } n_2 = 36. \text{ Note also that the population distributions are normal. Let}

\[ Z = \frac{(X_1 - X_2) - (\mu_1 - \mu_2)}{\sqrt{(\sigma_1^2/n_1) + (\sigma_2^2/n_2)}}. \]

It is known that the sampling distribution of } Z \text{ is standard nor-
mal. Thus, the desired probability is given by
\[
\Pr\{3.4 \leq \bar{X}_1 - \bar{X}_2 < 5.9\} = \Pr\{3.35 < \bar{X}_1 - \bar{X}_2 < 5.85\} \\
\text{(continuity correction)} \\
= \Pr\left\{\frac{3.35 - 5}{\sqrt{25/25 + 9/36}} < Z < \frac{5.85 - 5}{\sqrt{25/25 + 9/36}}\right\} \\
\approx \Pr\{-1.48 < Z < 0.76\} \\
= 1 - \Pr\{Z > 1.48\} - \Pr\{Z > 0.76\} \\
= 1 - 0.0823 - 0.2236 = 0.7070.
\]

(b) A soft drink machine is being regulated so that the amount of drink dispensed averages 240 milliliters with a standard deviation of 15 milliliters. Periodically, the machine is checked by taking a sample of 40 drinks and computing the average content. If the mean of 40 drinks is a value within the interval \(\mu_{\bar{X}} \pm 2\sigma_{\bar{X}}\), the machine is thought to be operating satisfactorily; otherwise, adjustments are made. If the company official found that the mean of 40 drinks to be \(\bar{x} = 236\) milliliters and concluded that the machine needed no adjustment. Was this a reasonable decision?

**Sol.** Recall that \(\mu_{\bar{X}} = \mathbb{E}(\bar{X}) = \mu\), which is 240 (milliliters); and \(\sigma_{\bar{X}} = \text{Var}(\bar{X}) = \sigma/\sqrt{n}\), which is \(15/\sqrt{40} \approx 2.3717\). According to the rule adopted by the company, no adjustment is needed if the mean (i.e., the average) of 40 drinks is a value within the interval \(\mu_{\bar{X}} \pm 2\sigma_{\bar{X}}\). Here, \(\mu_{\bar{X}} \pm 2\sigma_{\bar{X}} = 240 \pm 2 \cdot \frac{15}{\sqrt{40}} = 240 \pm 4.7434\). As \(\bar{x} = 236\) which is within this range \((235.2566, 244.7434)\), it was a reasonable decision.
(c) The television picture tubes of manufacturer A have a mean lifetime of 6.5 years and a standard deviation of 0.9 year, while those of manufacturer B have a mean lifetime of 6.0 years and a standard deviation of 0.8 year. What is the probability that a random sample of 36 tubes from manufacturer A will have a mean lifetime that is at least 1 year more than the mean lifetime of a sample of 49 tubes from manufacturer B?

**Sol.** We are given the following information:

<table>
<thead>
<tr>
<th>Population 1</th>
<th>Population 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_1 = 6.5$</td>
<td>$\mu_2 = 6.0$</td>
</tr>
<tr>
<td>$\sigma_1 = 0.9$</td>
<td>$\sigma_2 = 0.8$</td>
</tr>
<tr>
<td>$n_1 = 36$</td>
<td>$n_2 = 49$</td>
</tr>
</tbody>
</table>

Note that the sample sizes of both random samples are $\geq 30$. Hence the CLT can be applied and the sampling distribution of $\overline{X}_1 - \overline{X}_2$ will be approximately normal and will have a mean and standard deviation given by

$$\mu_1 - \mu_2 = 6.5 - 6.0 = 0.5,$$

and

$$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{0.81}{36} + \frac{0.64}{49}} = 0.189.$$
That is, 

\[
Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \approx N(0, 1).
\]

The probability that the mean of 36 tubes from manufacturer A will be at least 1 year longer than the mean of 49 tubes from manufacturer B is given by

\[
\text{IP}\{\bar{X}_1 - \bar{X}_2 \geq 1.0\} = \text{IP}\{Z > \frac{1.0 - 0.5}{0.189}\} \\
= \text{IP}\{Z > 2.65\} \\
\approx 0.0040,
\]

according to the normal table.

(d) A certain type of thread is manufactured with a mean of tensile strength of 78.3 kilograms and a standard deviation of 5.6 kilograms. Assume that a random sample of size \(n\) is taken.

(i) How is the standard deviation of the sample mean changed when the sample size \(n\) is increased from 64 to 196?

**Sol.** Let \(\bar{X}_1\) denote the sample mean of the samples of size \(n_1 = 64\) taken from the population of this certain thread (with mean \(\mu = 78.3\) and standard deviation \(\sigma = 5.6\)). Let \(\bar{X}_2\) denote the sample mean of the samples of size \(n_2 = 196\) taken from the same population.

Now the standard deviation of \(\bar{X}_1\) is given by

\[
\sigma_{\bar{X}_1} = \text{Var}(\bar{X}_1) = \sigma/\sqrt{n_1} = \frac{5.6}{8}.
\]

Similarly, the standard deviation of \(\bar{X}_2\) is given by

\[
\sigma_{\bar{X}_2} = \text{Var}(\bar{X}_2) = \sigma/\sqrt{n_2} = \frac{5.6}{16}.
\]
Observe that \[
\frac{\sigma_{\bar{X}_2}}{\sigma_{\bar{X}_1}} = \frac{8}{16} = \sqrt{\frac{n_1}{n_2}} = \frac{1}{2}.
\]
Thus, the standard deviation of the sample mean is reduced by half when the sample size is increased from 64 to 196.

(ii) When the sample size \(n\) is decreased from 784 to 49, how will the standard deviation of the sample mean be changed?

\textbf{Sol.} As above. Now \(n_1 = 784\) and \(n_2 = 49\). Note that \(\sigma_{\bar{X}_1} = \text{Var}(\bar{X}_1) = \sigma/\sqrt{n_1} = \frac{5.6}{28}\). Similarly,

\[
\sigma_{\bar{X}_2} = \text{Var}(\bar{X}_2) = \sigma/\sqrt{n_2} = \frac{5.6}{7}.
\]

Observe that \[
\frac{\sigma_{\bar{X}_2}}{\sigma_{\bar{X}_1}} = \frac{28}{7} = \sqrt{\frac{n_1}{n_2}} = 4.
\]
Thus, the standard deviation of the sample mean is increased by 4 times when the sample size is decreased from 784 to 49.

(e) If the standard deviation of the sample mean for the sampling distribution of random samples of size 36 is 2, how large must the size of the sample become if the standard deviation is to be reduced to 1.2?

\textbf{Sol.} Refer to the above example. It has been shown that there that the sample size must be increased in order to get the standard deviation reduced. Say, the sample size become \(n\). Then, \(\frac{1.2}{2} = \sqrt{\frac{36}{n}}\). Solving for \(n\), we find that \(n = 100\). That is, the size of the sample must become 100 if the standard deviation is to be reduced to 1.2.
(f) The amount of time that a bank teller spends on a customer is a r.v. with a mean \( \mu = 3.2 \) minutes and a standard deviation of \( \sigma = 1.6 \) minutes. If a random sample of 64 customers is observed, find the probability that their mean time at the teller’s counter is

(i) at most 2.7 minutes;

(ii) at least 3.2 minutes but less than 3.4 minutes.

*Sol.* Note that the population distribution is not specified, however, one can make use of the CLT to the sample mean \( \overline{X} \) as the sample size is \( n = 64 \geq 30 \). For part (i), the desired probability is

\[
\Pr \left\{ \overline{X} < 2.7 \right\} = \Pr \left\{ Z < \frac{2.7 - 3.2}{1.6/\sqrt{64}} \right\} = \Pr \left\{ Z > 2.5 \right\} = 0.0062.
\]

For part (ii),

\[
\Pr \left\{ 3.2 \leq \overline{X} < 3.4 \right\} = \Pr \left\{ 0 \leq Z < \frac{3.4 - 3.2}{1.6/\sqrt{64}} \right\}
\]
\[
= \Pr \left\{ Z > 0 \right\} - \Pr \left\{ Z > 1 \right\}
\]
\[
= \frac{1}{2} - 0.1587 = 0.3413.
\]

*End of Examples.*
Statistics is a theory of deducing estimates and properties of quantities which can not be observed directly. For example, consider the population of heights of students in NUS. We may want to know the average height ($\mu$ - a parameter). To estimate $\mu$, a practical way is selecting randomly, say, 100 students and then computing their average height. Also, we can set up a hypothesis

$$H_0 : \mu > 165 \text{ cm}$$

and then “test” it in terms of the data we have. In this simple example, we have encountered some basic aspects of statistics:

(1) *Sampling Design*: How to make a survey or collect data?

(2) Estimate the parameters or test hypotheses about the parameter.

(3) Make statistical conclusions or inference.

In the following, we shall deal with the first problem. Since the conclusion is based on the sample (data), we need to know the probabilistic behavior of a sample before we can make any reliable conclusion.

§1. **Basic Concepts.**

A **population** consists of the totality of the observations with which we are concerned. The number of observations in the population is defined to be the **size** of the population. Any numerical value describing a characteristic of the population is called a **parameter**.

**Examples:**
(a) If we are interested in studying the heights of students in NUS, then the population is the set of heights of all students, say,

\[\{160, 156, 170, 172, 163, \ldots\}\].

The average height \( \mu \) is a parameter.

(b) If we flip a coin indefinitely and record the heads that occur, the set of values obtained is the population.

Each observation in a population is a value of a random variable \( X \) having some probability distribution \( f(x) \). In Example (b), each observation in the population might be a value 0 or 1 of the binomial r.v. \( X \) with probability density function

\[ b(x; 1, p) = p^x q^{1-x}, \quad x = 0, 1. \]

A sample is a subset of a population.

In Example (a), a set of 100 students’ heights is a sample of the total students’ heights. Also, a set of 100 male students’ heights is a sample. Certainly, this sample does not reflect the characteristic of the original population. Any sampling procedure that produces inferences that consistently overestimate or consistently underestimate some characteristic of the population is said to be biased. To eliminate any possibility of bias in sampling procedure, it is desirable to choose a random sample in the sense that the observations are made independently and at random.

In selecting a random sample of size \( n \) from a population that has probability distribution (or probability density function) \( f(x) \), let us define the r.v.’s \( X_i, i = 1, 2, \ldots, n \), to represent the \( i^{th} \) sample value that we observe. The random variables \( X_1, X_2, \ldots, X_n \)
will then constitute a random sample from the population $f(x)$ with numerical values $x_1, x_2, \ldots, x_n$ if the values are obtained by repeating the experiment $n$ independent times under essentially the same conditions. Therefore, it is reasonable to assume that the $n$ random variables $X_1, X_2, \ldots, X_n$ are independent and that each has the same probability distribution $f(x)$.

$X_1, X_2, \ldots, X_n$ are said to be a random sample of size $n$ from the population $f(x)$.

A random sample is a sample selected randomly from a population.

The most important random sample is a sample obtained with replacement sampling. In this case, if the population has probability distribution (or probability density function) $f(x)$, then a random sample of size $n$ is given by $n$ independent r.v.’s $X_1, X_2, \ldots, X_n$, and each $X_i$ has the same probability distribution $f(x)$.

$X_1, X_2, \ldots, X_n$ are said to be a random sample of size $n$ from a population with the probability density function $f(x)$ if $X_1, X_2, \ldots, X_n$ are $n$ independent random variables each having the same probability density $f(x)$. The joint probability density function of $X_1, X_2, \ldots, X_n$ is given by

$$f(x_1, x_2, \ldots, x_n) = f(x_1) f(x_2) \cdots f(x_n)$$

**Remarks:**

(i) A typical way of getting a random sample is as follows: Let $f(x)$ be the probability distribution of the population of the students' heights (say, a normal distribution with mean $\mu = 165$ cm and standard deviation $\sigma = 10$ cm). Choose $n = 100$ students with
replacement and let $X_1, X_2, \ldots, X_{100}$ be the possible heights of the 100 students, then each $X_i \sim N(\mu, \sigma^2)$, $X_1, X_2, \ldots, X_{100}$ are independent. We say that $X_1, X_2, \ldots, X_{100}$ form a random sample of size 100. Now, if we actually go and record the heights of 100 randomly chosen students, then we obtain $x_1, x_2, \ldots, x_{100}$, a realization of the random sample of $X_1, X_2, \ldots, X_{100}$. In practice, we also call $x_1, x_2, \ldots, x_{100}$ as a random sample.

(ii) If the population size is very large, we treat the sampling without replacement the same as sampling with replacement.

Our main purpose in selecting random samples is to elicit information about the unknown population parameters. Suppose, for example, we wish to arrive at a conclusion concerning the proportion of coffee-drinking people in Singapore who prefer a certain brand of coffee. It would be impossible to question every coffee-drinking Singaporean in order to compute the value of the parameter $p$ representing the population proportion. Instead, a large random sample is selected and the proportion $\hat{p}$ of this sample favoring the brand of coffee is calculated. Now, $\hat{p}$ is a function of the observed values in the random sample. Since many random samples are possible from the same population, $\hat{p}$ would vary from sample to sample. That is, $\hat{p}$ is a value of a random variable. Such a random variable is called a statistic.

Any function of a random sample is called a statistic.
§2. **Sampling Distribution of the Mean \( \bar{X} \).**

1. The probability distribution of a statistic is called a **sampling distribution**.

2. **Remarks:** Refer to the Central Limit Theorem (CLT).

   (a) If the population distribution is non-normal, yet the sample size \( n \geq 30 \), then we can use the CLT and in such a case, \( \bar{X} \approx N(\mu, \sigma^2/n) \).

   (b) If the population has a normal distribution, then
   \[
   \bar{X} \sim N(\mu, \sigma^2/n).
   \]

§3. **Sampling Distribution of \( \bar{X}_1 - \bar{X}_2 \).**

1. Suppose that we now have two populations, the first with mean \( \mu_1 \) and variance \( \sigma_1^2 \), and the second with mean \( \mu_2 \) and variance \( \sigma_2^2 \). Let the statistic \( \bar{X}_1 \) represent the mean of a random sample of size \( n_1 \) selected from the first population, and the \( \bar{X}_2 \) represent the mean of a random sample of size \( n_2 \) selected from the second population, independent of the sample from the first population. What can we say about the sampling distribution of \( \bar{X}_1 - \bar{X}_2 \)?

2. The reason that we want to know the sampling distribution of \( \bar{X}_1 - \bar{X}_2 \) is that we often need to compare two means. For example, we may want to know if the average score of two classes, \( \mu_1 \) and \( \mu_2 \) are the same, i.e., \( \mu_1 = \mu_2 \); or if the average score of the first class is higher than that of the second one, i.e., \( \mu_1 > \mu_2 \).

3. If independent samples of size \( n_1 \) and \( n_2 \) are drawn at random from two populations, with means \( \mu_1 \) and \( \mu_2 \) and variances \( \sigma_1^2 \) and \( \sigma_2^2 \), respectively, then
(i) \( \mu_{\bar{X}_1 - \bar{X}_2} = E (\bar{X}_1 - \bar{X}_2) = \mu_1 - \mu_2. \)

(ii) \( \sigma^2_{\bar{X}_1 - \bar{X}_2} = \sigma^2_{\bar{X}_1} + \sigma^2_{\bar{X}_2} = \frac{\sigma^2_1}{n_1} + \frac{\sigma^2_2}{n_2}. \)

(iii) If \( n_1 \) and \( n_2 \) are both large, then

\[
\bar{X}_1 - \bar{X}_2 \approx N \left( \mu_1 - \mu_2, \frac{\sigma^2_1}{n_1} + \frac{\sigma^2_2}{n_2} \right). 
\]

\[
\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{(\sigma^2_1/n_1) + (\sigma^2_2/n_2)}} \approx N(0, 1)
\]
A Brief Summary – Some Useful Facts

(a) If $X_1, X_2, \ldots, X_n$ are a random sample of size $n$ from a population with mean $\mu$, and variance $\sigma^2$, then

(i) $\mu_{\bar{X}} = E(\bar{X}) = \mu$;

(ii) $\sigma^2_{\bar{X}} = \text{Var}(\bar{X}) = \frac{\sigma^2}{n}$.

(iii) **CLT:**

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{\bar{X} - E\bar{X}}{\sqrt{\text{Var}(\bar{X})}} \approx N(0, 1)$$

(b) If the population has a normal distribution, then

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{\bar{X} - E\bar{X}}{\sqrt{\text{Var}(\bar{X})}} \sim N(0, 1).$$

Equivalently, $\bar{X} \sim N(\mu, \sigma^2/n)$.

(c) Sampling distribution of the difference between two means is as follows:

(i) $\mu_{\bar{X}_1 - \bar{X}_2} = E(\bar{X}_1 - \bar{X}_2) = \mu_1 - \mu_2$.

(ii) $\sigma^2_{\bar{X}_1 - \bar{X}_2} = \sigma^2_{\bar{X}_1} + \sigma^2_{\bar{X}_2} = \frac{\sigma^2_1}{n_1} + \frac{\sigma^2_2}{n_2}$.

(iii) $\bar{X}_1 - \bar{X}_2 \approx N \left( \mu_1 - \mu_2, \frac{\sigma^2_1}{n_1} + \frac{\sigma^2_2}{n_2} \right)$, and

$$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma^2_1/n_1 + \sigma^2_2/n_2}} \approx N(0, 1).$$
§4. **Sampling Distribution of $S^2$.**

1. First of all, recall that if a random sample of size $n$ is drawn from a population with mean $\mu$ and variance $\sigma^2$, then the sample variance is given by

\[
S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2,
\]

which gives a measurement of spreading of the sample $X_1, X_2, \ldots, X_n$. And we have

\[
E[S^2] = \sigma^2.
\]

2. In order to find the sampling distribution of $S^2$, we introduce a $\chi^2$-distribution.

A continuous r.v. $X$ has a chi-square ($\chi^2$) distribution, with $\nu$ degrees of freedom, if its density function is given by

\[
f_X(x) = \begin{cases} 
\frac{1}{2^{\nu/2} \Gamma(\nu/2)} x^{\nu/2-1} e^{-x/2}, & x > 0, \\
0, & x \leq 0.
\end{cases}
\]

For a chi-square r.v. $X$, we can use the table for chi-square distribution to find the value of $c$ such that

\[
\Pr\{X \geq c\} = \alpha
\]

for certain given $\alpha$ and $\nu$. The value of $c$ is usually denoted as $\chi^2_{\alpha,\nu}$, i.e.,

\[
\Pr\{X \geq \chi^2_{\alpha,\nu}\} = \alpha.
\]
3. **Examples:** Suppose that $X$ has a chi-square distribution with 7 degrees of freedom.

(i) Find $a$ such that $\mathbb{P}\{X \geq a\} = 0.05$.

(ii) Find $b$ such that $\mathbb{P}\{X \leq b\} = 0.75$.

(iii) Find $c$ such that $\mathbb{P}\{c > X > 4.671\} = 0.4$.

4. **Theorem:**

If $X_1, X_2, \ldots, X_n$ are independent random variables having the same normal distribution with mean $\mu$ and variance $\sigma^2$, then the r.v.

$$Y = \sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma^2} = \sum_{i=1}^{n} Z_i^2$$

has a chi-square distribution with $\nu = n$ degrees of freedom.

5. **Theorem:**

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

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

has a chi-square distribution with $\nu = n - 1$ degrees of freedom.
6. **Remarks:**

(a) The "degrees of freedom" of a chi-square r.v. indicates the number of independent normal random variables. For \( S^2 \), we have \( X_1 - \overline{X}, X_2 - \overline{X}, \ldots, X_n - \overline{X} \), but

\[
X_1 - \overline{X} + X_2 - \overline{X} + \cdots + X_n - \overline{X} = 0.
\]

So, there are actually \( n - 1 \) pieces of independent information instead of \( n \). Therefore, the degrees of freedom for \( S^2 \) is \( n - 1 \).

(b) The result in the second theorem above is a base for statistical inference concerning \( \sigma^2 \).

7. **Examples:**

(a) Find the probability that a random sample of 5 observations, from a normal population with variance \( \sigma^2 = 1 \), will have a sample variance \( S^2 \)

(i) greater than 2.78575;

(ii) less than 0.121.

(b) A manufacturer of car batteries guarantees that his batteries will last, on the average, 3 years with a standard deviation of 1 year. If five of these batteries have lifetimes of 1.9, 2.4, 3.0, 3.5, and 4.2 years, is the manufacturer still convinced that his batteries have a standard deviation of 1 year? (Assume the lifetime of a battery follows a normal distribution)
§5. *F*-distribution

One of the most important distributions in applied statistics is the *F*-distribution.

1. The statistic *F* is defined to be the ratio of two independent chi-square random variables, each divided by its number of degrees of freedom:

\[ F = \frac{U/\nu_1}{V/\nu_2}, \]

where \( \nu_1 \) and \( \nu_2 \) are the degrees of freedom of *U* and *V*, respectively. Such an *F* is said to have an *F*-distribution with degrees of freedom \( \nu_1 \) and \( \nu_2 \).

2. The density function of *F* defined above is given by

\[
f(x) = \begin{cases} 
\frac{\Gamma[(\nu_1 + \nu_2)/2](\nu_1/\nu_2)^{\nu_1/2} x^{\nu_1/2-1}}{\Gamma(\nu_1/2) \Gamma(\nu_2/2) (1 + \nu_1 x/\nu_2)^{(\nu_1+\nu_2)/2}}, & x > 0, \\
0, & x \leq 0.
\end{cases}
\]

From the table for an *F*-distribution, one can check out \( f_{\alpha;\nu_1,\nu_2} \), leaving an area of \( \alpha \) to the right, where \( \alpha \) takes on values 0.05, 0.025, 0.01 and 0.001, and \( \nu_1 \) is the degrees of freedom in the numerator and \( \nu_2 \) is the degrees of freedom in the denominator. That is, \( \Pr\{F > f_{\alpha;\nu_1,\nu_2}\} = \alpha \), for an *F* r.v. with degrees of freedom \( \nu_1 \) and \( \nu_2 \).
3. An important and useful property about $F$-distributions is that: If $W$ is an $F$ r.v. with degrees of freedom $\nu_1$ and $\nu_2$, then $\frac{1}{W}$ is also having an $F$ distribution with degrees of freedom $\nu_2$ and $\nu_1$.

Thus, we have the following relation:

$$f_{\alpha;\nu_1,\nu_2} = \frac{1}{f_{1-\alpha;\nu_2,\nu_1}}.$$

4. 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

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

has an $F$-distribution with $\nu_1 = n_1 - 1$ and $\nu_2 = n_2 - 1$ degrees of freedom.

5. **Examples:**

(a) If $S_1^2$ and $S_2^2$ represent the variances of independent random samples of $n_1 = 25$ and $n_2 = 31$, taken from normal populations with variances $\sigma_1^2 = 10$ and $\sigma_2^2 = 15$, respectively, find

(i) $\Pr(S_1^2/S_2^2 > 1.26)$;

(ii) Find $c$ such that $\Pr\{S_1^2/S_2^2 \leq c\} = 0.95$.

(b) Let $S_1^2$ and $S_2^2$ be the variances of independent random samples of $n_1 = 11$ and $n_2 = 8$, taken from normal populations with a common variance $\sigma^2$. Find $\Pr\{S_1^2/S_2^2 \leq 6.62\}$. 

18
§6. **Student t-distribution.**

1. Recall that the sample mean $\overline{X}$ of a random sample of size $n$ drawn from a normal population is normally distributed. Indeed, $Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \sim N(0, 1)$. When the value of $\sigma$ is known, one can of course make use of such a standardization. However, in practice, the population variance is typically unknown. For samples of size $n \geq 30$, a good estimate of $\sigma^2$ is provided by calculating the sample variance $S^2$. Thus, we introduce the following t-statistic

\[
T = \frac{\overline{X} - \mu}{S / \sqrt{n}}.
\]

2. **Theorem.**

*Let Z be a standard normal r.v. and V a chi-square r.v. with $\nu$ degrees of freedom. If Z and V are independent, then the p.d.f. of the r.v. T defined by

\[
T = \frac{Z}{\sqrt{V/\nu}},
\]

is given by

\[
f_T(t) = \frac{\Gamma[(\nu + 1)/2]}{\Gamma(\nu/2)\sqrt{\pi\nu}} \left(1 + \frac{t^2}{\nu}\right)^{-(\nu+1)/2}.
\]

*This is known as the t-distribution with $\nu$ degrees of freedom.*

3. Now write

\[
T = \frac{\overline{X} - \mu}{S / \sqrt{n}} = \frac{Z}{\sqrt{V/(n-1)}},
\]

19
where \( Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \) and \( V = (n - 1)S^2 / \sigma^2 \). It can be shown that if the sample is drawn from a normal distribution, then \( Z \) and \( V \) are independent.

4. **Theorem.**

Let \( X_1, X_2, \ldots, X_n \) be a random sample of size \( n \) from a normal population with mean \( \mu \) and variance \( \sigma^2 \), then the random variable

\[
T = \frac{\bar{X} - \mu}{S / \sqrt{n}}
\]

has the \( t \)-distribution with \( \nu = n - 1 \) degrees of freedom. Its density function is given by

\[
f_T(t) = \frac{\Gamma(n/2)}{\Gamma((n - 1)/2)\sqrt{\pi(n - 1)}} \left(1 + \frac{t^2}{n - 1}\right)^{-n/2}.
\]

5. **Some Historical Notes.**

The probability distribution of \( T \) was first published in 1908 in a paper by W.S. Gosset. At the time, Gosset was employed by an Irish brewery that disallowed publication of research by members of its staff. To circumvent this restriction, he published his work secretly under the name “Student”. Consequently, the distribution of \( T \) is usually called the “Student \( t \)-distribution”, or simply the \( t \)-distribution.
6. **Remarks:**

(a) $T$ behaves more or less like a normal r.v. but heavily depends on $n$.

(b) The density function $f_T(t)$ is bell-shaped and symmetric about $x = 0$. When $n$ gets larger and larger, the t-distribution curve gets closer and closer to the normal curve which corresponds to the case of $n = \infty$.

(c) If $T$ has t-distribution with $\nu$ degrees of freedom, it is customary to let $t_{\alpha; \nu}$ represent the t-value such that $P\{T > t_{\alpha; \nu}\} = \alpha$. By symmetry, $\forall 0 < \alpha < 1$ and $\nu > 0$,

$$t_{1-\alpha; \nu} = -t_{\alpha; \nu}.$$  

7. **Examples:**

(a) If $T$ has t-distribution with $\nu$ degrees of freedom, find

$$\Pr\{-t_{0.05; \nu} < T < t_{0.1; \nu}\}.$$  

(b) Find $a$ such that $\Pr\{a < T < -1.761\} = 0.045$ for a random sample of size 15 selected from a normal distribution.

(c) A cigarette manufacturer claims that his cigarettes have an average nicotine content of 1.83 milligrams. If a random sample of 8 cigarettes of this type have nicotine contents of 2.0, 1.7, 2.1, 1.9, 2.2, 2.1, 2.0 and 1.6 milligrams, would you agree with the manufacturer’s claim?

21
8. **Theorem.**

If independent samples of size \( n_1 \) and \( n_2 \) are drawn at random from normal populations with means \( \mu_1 \) and \( \mu_2 \) and with unknown but equal variances, then

\[
T = \frac{(X_1 - X_2) - (\mu_1 - \mu_2)}{S_p \sqrt{1/n_1 + 1/n_2}}
\]

has the t-distribution with \( \nu = n_1 + n_2 - 2 \) degrees of freedom, where

\[
S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}.
\]

9. **Example:**

A taxi company is trying to decide whether to purchase brand A or brand B tires for its fleet of taxis. To estimate the difference in the two brands, an experiment is conducted using 12 of each brand. The tires are run until they wear out. The results are (in kilometers)

<table>
<thead>
<tr>
<th></th>
<th>Brand A</th>
<th>Brand B</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \overline{X}_1 )</td>
<td>38,100</td>
<td>36,300</td>
</tr>
<tr>
<td>( S_1 )</td>
<td>6100</td>
<td>5000</td>
</tr>
</tbody>
</table>

Would you agree with the claim that \( \mu_1 - \mu_2 = 0 \) kilometers, assuming the populations to be normally distributed?