0.1 Sampling Distribution of Means (Continued)

- **Example 8.6**: An electric firm manufactures light bulbs that have a length of life that is approximately normally distributed, with mean equal to 800 hours and a standard deviation of 40 hours.
- Find the probability that a random sample of 16 bulbs will have an average life of less than 775 hours.
- Solution:
  - Even though $16 < 30$, the central limit theorem can be used because it is stated that the population distribution is approximately normal.
  - The sampling distribution of $\bar{X}$ will be approximately normal, with

$$\mu_{\bar{X}} = 800, \ \sigma_{\bar{X}} = 40/\sqrt{16} = 10$$

$$\bar{x} = 75 \Rightarrow z = \frac{775 - 800}{10} = -2.5$$

$$P(\bar{X} < 775) = P(Z < -2.5) = 0.0062$$

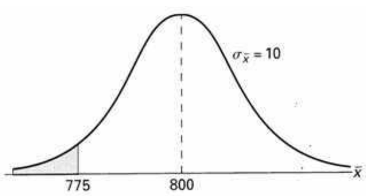

Figure 1: Area for Example 8.6.

- **Example 8.7**: A engineer conjectures that the population mean of a certain component parts is 5.0 millimeters. An experiment is conducted in which 100 parts produced by the process are selected randomly and the diameter measured on each.
- It is known that the population standard deviation $\sigma = 0.1$. The experiment indicates a sample average diameter $\bar{X} = 5.027$ millimeters.
- Does this sample information appear to support or refute the engineer's conjecture?

- Solution:

$$\begin{aligned} & P\left[|(\bar{X}-5)| \geq 0.027\right] \\ = & P[(\bar{X}-5) \geq 0.027] + P[(\bar{X}-5) \leq -0.027] \\ = & 2P(Z \geq 2.7) = 2 * 0.0035 = 0.007 \end{aligned}$$

Strongly refutes the conjecture!

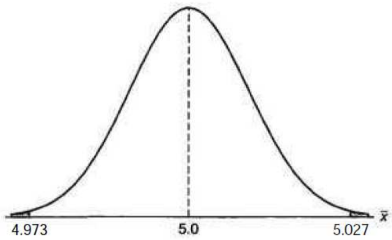

Figure 2: Area for Example 8.7.

- Sometimes we are interested in comparing two populations (i.e., one manufacturing process better than the other).
- Suppose we have two populations, the first with μ_1 and σ_1 and the second with μ_2 and σ_2.
- Let the statistic $\bar{X}_1$ represent the sample mean selected from the first population and the statistic $\bar{X}_2$ represent the sample mean selected from the second population.
- How about the sampling distribution
- Solution: Using Theorem 7.11, $\bar{X}_1 - \bar{X}_2$ is approximately normally distributed with mean

$$\mu_{\bar{X}_1-\bar{X}_2} = \mu_{\bar{X}_1} - \mu_{\bar{X}_2} = \mu_1 - \mu_2$$

and variance

$$\sigma^2_{\bar{X}_1-\bar{X}_2} = \sigma^2_{\bar{X}_1} + \sigma^2_{\bar{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

- **Theorem 8.3:**

If independent sample of size n_1 and n_2 are drawn at random from two populations, discrete or continuous, with means μ_1 and μ_2 and variances σ_1^2 and σ_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

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

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.

- **Example 8.8**: Two independent experiments are being run in which two different types of paints are compared.

- Eighteen specimens are painted using type A and the drying time in hours is recorded on each. The same is done with type B.

- The population standard deviations are both known to be 1.0. Assuming that the mean drying time is equal for the two types of paint,

- find $P(\bar{X}_A - \bar{X}_B > 1.0)$ where $\bar{X}_A$ and $\bar{X}_B$ are average drying times for samples of size $n_A = n_B = 18$.

- Solution:

$$\mu_{\bar{X}_A - \bar{X}_B} = \mu_A - \mu_B = 0$$

$$\sigma^2_{\bar{X}_A - \bar{X}_B} = \frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}$$

$$z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{1-0}{\sqrt{1/9}} = 3.0$$

$$P(Z > 3.0) = 1 - P(Z < 3.0)$$

$$= 1 - 0.9987 = 0.0013$$

Low probability. Assumption?

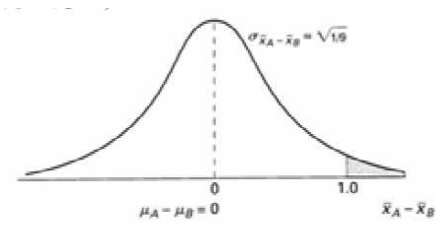

Figure 3: Area for Example 8.8.

- **Example 8.9**: 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?

Table 1: Data for Example 8.9.

Population 1	Population 2
$\mu_1 = 6.5$	$\mu_2 = 6.0$
$\sigma_1 = 0.9$	$\sigma_2 = 0.8$
$n_1 = 36$	$n_2 = 49$

- $P(\bar{X}_1 - \bar{X}_2 \geq 1.0) = ?$

Solution: Since both n_1 and n_2 is greater than 30, the sampling distribution of $\bar{X}_1 - \bar{X}_2$ will be approximately normal.

$$\mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2 = 6.5 - 6.0 = 0.5$$

$$\sigma^2_{\bar{X}_1 - \bar{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

$$= \frac{0.9^2}{36} + \frac{0.8^2}{49} = 0.0356$$

$$z = \frac{1 - 0.5}{\sqrt{0.0356}} = 2.65$$

$$P(\bar{X}_1 - \bar{X}_2 \geq 1.0) = P(Z > 2.65)$$

$$= 1 - P(Z < 2.65) = 1 - 0.9960$$
$$= 0.004$$

Low probability value.

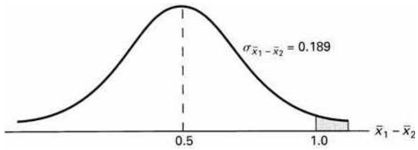

Figure 4: Area for Example 8.9.

0.2 Sampling Distribution of S^2

- If a random sample of size n is taken from a normal population with mean μ and variance σ^2, and the sample variance S^2 is computed.
- Consider the distribution of the statistics $\frac{(n-1)S^2}{\sigma^2}$

$$\sum_{i=1}^{n}(X_i - \mu)^2 = \sum_{i=1}^{n}\left[(X_i - \bar{X}) + (\bar{X} - \mu)\right]^2$$

$$= \sum_{i=1}^{n}(X_i - \bar{X})^2 + \sum_{i=1}^{n}(\bar{X} - \mu)^2 + 2(\bar{X} - \mu)\sum_{i=1}^{n}(X_i - \bar{X})$$

$$= \sum_{i=1}^{n}(X_i - \bar{X})^2 + n(\bar{X} - \mu)^2$$

- Dividing both sides by σ^2 and substituting $(n-1)S^2$ for $\sum_{i=1}^{n}(X_i - \bar{X})^2$, we obtain

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

chi-squared random variable with n degrees of freedom

$$\frac{(n-1)S^2}{\sigma^2} +$$

$$\frac{(\bar{X}-\mu)^2}{\frac{\sigma^2}{n}}$$

chi-squared random variable with 1 degrees of freedom

- **Theorem 8.4**:

  If S^2 is the variance of a random sample of size n taken from a normal population having the variance σ^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 $\nu = n - 1$ degrees of freedom.

- It is customary to let χ^2_α represent the χ^2-value above which we find an area of α.

- This is illustrated by the shaded region in Fig. 5.

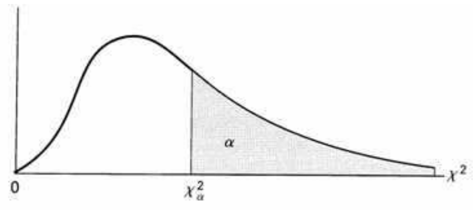

Figure 5: The chi-squared distribution.

For $\nu = 7, \chi^2_{0.05} = 14.067$, and $\chi^2_{0.95} = 2.1677$ (Table A.5)

TABLE A.5 Critical Values of the Chi-Squared Distribution

	α									
v	0.995	0.99	0.98	0.975	0.95	0.90	0.80	0.75	0.70	0.50
1	0.0^4393	0.0^3157	0.0^3628	0.0^3982	0.00393	0.0158	0.0642	0.102	0.148	0.455
2	0.0100	0.0201	0.0404	0.0506	0.103	0.211	0.446	0.575	0.713	1.386
3	0.0717	0.115	0.185	0.216	0.352	0.584	1.005	1.213	1.424	2.366
4	0.207	0.297	0.429	0.484	0.711	1.064	1.649	1.923	2.195	3.357
5	0.412	0.554	0.752	0.831	1.145	1.610	2.343	2.675	3.000	4.351
6	0.676	0.872	1.134	1.237	1.635	2.204	3.070	3.455	3.828	5.348
7	0.989	1.239	1.564	1.690	2.167	2.833	3.822	4.255	4.671	6.346
8	1.344	1.646	2.032	2.180	2.733	3.490	4.594	5.071	5.527	7.344
9	1.735	2.088	2.532	2.700	3.325	4.168	5.380	5.899	6.393	8.343
10	2.156	2.558	3.059	3.247	3.940	4.865	6.179	6.737	7.267	9.342

TABLE A.5 (continued) Critical Values of the Chi-Squared Distribution

	α									
v	0.30	0.25	0.20	0.10	0.05	0.025	0.02	0.01	0.005	0.001
1	1.074	1.323	1.642	2.706	3.841	5.024	5.412	6.635	7.879	10.827
2	2.408	2.773	3.219	4.605	5.991	7.378	7.824	9.210	10.597	13.815
3	3.665	4.108	4.642	6.251	7.815	9.348	9.837	11.345	12.838	16.268
4	4.878	5.385	5.989	7.779	9.488	11.143	11.668	13.277	14.860	18.465
5	6.064	6.626	7.289	9.236	11.070	12.832	13.388	15.086	16.750	20.517
6	7.231	7.841	8.558	10.645	12.592	14.449	15.033	16.812	18.548	22.457
7	8.383	9.037	9.803	12.017	14.067	16.013	16.622	18.475	20.278	24.322
8	9.524	10.219	11.030	13.362	15.507	17.535	18.168	20.090	21.955	26.125
9	10.656	11.389	12.242	14.684	16.919	19.023	19.679	21.666	23.589	27.877
10	11.781	12.549	13.442	15.987	18.307	20.483	21.161	23.209	25.188	29.588

- **Example 8.10**: 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 that the battery lifetime follows a normal distribution.
- Solution:

$$s^2 = \frac{n\sum_{i=1}^{n} X_i^2 - (\sum_{i=1}^{n} X_i)^2}{n(n-1)} = \frac{5 * 48.26 - 15^2}{5 * 4} = 0.815$$

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{4 * 0.815}{1} = 3.26$$

- Since $n = 5$, χ^2 has $\nu = n - 1 = 4$ degrees of freedom.
- From Table A.5 row $\nu = 4$, wee see that

$$\chi^2_{0.025} = 11.143 \; and \; \chi^2_{0.975} = 0.484$$

- Since 95% of the values with 4 degrees of freedom fall between 0.484 and 11.143, the computed value with $\sigma^2 = 1$ is reasonable (since our $\chi^2 = 3.26$ falls within this range).
- Therefore the manufacturer has no reason to suspect that the standard deviation is other than 1 year.

0.3 t-distribution

- Central Limit Theorem (Theorem 8.2) assumes σ is known in

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

- However, σ might not be known. Then, consider the random variable

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

- The value of sample variance S^2 fluctuate considerably from sample to sample, T does not follow the standard normal distribution but follows t-distribution with the degrees of freedom $n-1$.

- In developing the sampling distribution of T, we shall assume that our random sample was selected from a normal population.

$$T = \frac{(\bar{X} - \mu)/(\sigma/\sqrt{n})}{\sqrt{S^2/\sigma^2}} = \frac{Z}{\sqrt{V/(n-1)}}$$

- where $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ and $V = \frac{(n-1)S^2}{\sigma^2}$

- **Theorem 8.5**:

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

$$T = \frac{Z}{\sqrt{V/\nu}}$$

  is given by the density function

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

  This is known as the **t-distribution** with ν degrees of freedom.

- **Corollary 8.1**:

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

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

Then the random variable $T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$ has a t-distribution with $\nu = n-1$ degrees of freedom.

- Student t-distribution
  - The probability distribution of T was first published in 1908 in a paper by W. S. Gosset.
  - Employed by an Irish brewery, but disallowed publication.
  - Published his work secretly under the name "Student".
- The shape of T looks like the standard normal (Z) (depending on the degrees of freedom, $n-1$). Symmetric about $\mu = 0$, bell-shaped.
- Difference between T and Z: variance of $T \geq 1$ and depends on n
- T and Z are the same as $n \to \infty$
- t_α represents the t-value above which we find an area of α to the right.
- t-distribution is symmetric about 0: $t_{1-\alpha} = -t_\alpha$

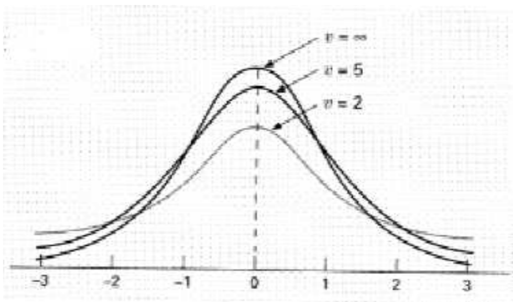

Figure 6: The t-distribution curves for $\nu = 2, 5$, and ∞

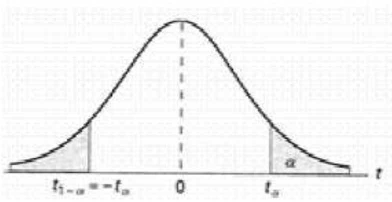

Figure 7: Symmetry property of the t-distribution.

TABLE A.4 Critical Values of the t-Distribution

	α						
v	**0.40**	**0.30**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**
1	0.325	0.727	1.376	1.963	3.078	6.314	12.706
2	0.289	0.617	1.061	1.386	1.886	2.920	4.303
3	0.277	0.584	0.978	1.250	1.638	2.353	3.182
4	0.271	0.569	0.941	1.190	1.533	2.132	2.776
5	0.267	0.559	0.920	1.156	1.476	2.015	2.571
6	0.265	0.553	0.906	1.134	1.440	1.943	2.447
7	0.263	0.549	0.896	1.119	1.415	1.895	2.365
8	0.262	0.546	0.889	1.108	1.397	1.860	2.306
9	0.261	0.543	0.883	1.100	1.383	1.833	2.262
10	0.260	0.542	0.879	1.093	1.372	1.812	2.228
11	0.260	0.540	0.876	1.088	1.363	1.796	2.201
12	0.259	0.539	0.873	1.083	1.356	1.782	2.179
13	0.259	0.537	0.870	1.079	1.350	1.771	2.160
14	0.258	0.537	0.868	1.076	1.345	1.761	2.145
15	0.258	0.536	0.866	1.074	1.341	1.753	2.131

- **Example 8.11**: The t-value with $\nu = 14$ degrees of freedom that leaves an area of 0.025 to the left, and therefore an area of 0.975 to the right, is

  $$t_{0.975} = -t_{0.025} = -2.145$$

  Look up $t_{0.025}$, and then place a negative sign.

- **Example 8.12**: Find $P(-t_{0.025} < T < t_{0.05}) =?$

  $$P(-t_{0.025} < T < t_{0.05}) = 1 - 0.05 - 0.025 = 0.925$$

  Since $t_{0.05}$ leaves an area of 0.05 to the right, and $-t_{0.025}$ leaves an area of 0.025 to the left, we find a total area of 0.925.

- Find k such that $P(k < T < -1.761) = 0.045$, for a random sample of size 15 selected from a normal distribution and

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

- Solution:

$$\nu = 15 - 1 = 14$$

From Table A.4, $-t_{0.05} = -1.761$ Let $k = -t_\alpha$, $0.045 = 0.05 - \alpha \Rightarrow \alpha = 0.005$ see Fig. 8

$$k = -t_{0.05} = -2.977 \ (Table\ A.4)$$

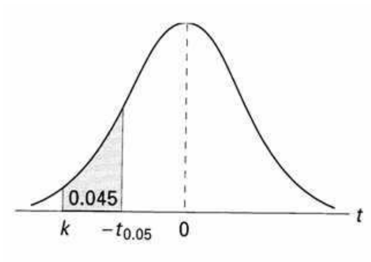

Figure 8: The t-values for Example 8.13.

- **Example 8.14**: A engineer claims that the population mean of a process is 500 grams. To check this claim he samples 25 batches each month.
- If the computed t-value falls between $-t_{0.05}$ and $t_{0.05}$, he is satisfied with his claim.
- What conclusion should he draw from a sample that has a mean $\bar{x} = 518$ grams and a sample standard deviation $s = 40$ grams? Assume the distribution of yields to be approximately normal.
- Solution: From Table A.4,

$$t_{0.05} = 1.711 \ (\nu = 24)$$

Assumption $\mu = 500 \Rightarrow$

$$t = \frac{518 - 500}{40/\sqrt{25}} = 2.25$$

$$2.25 > 1.711 \rightarrow error$$

if $\mu > 500$, t-value would be more reasonable. The process produces a better product than he thought.

- Exactly 95% of the values of a t-distribution with $\nu = n-1$ degrees of freedom lie between $-t_{0.025}$ and $t_{0.025}$.
- A t-value that falls **below** $-t_{0.025}$ **or above** $t_{0.025}$ would tend to make us believe that either a very rare event has taken place or perhaps our assumption about μ is in error.
- **What is the t-distribution used for?**. The t-distribution is used extensively in problems that deal with
  - Inference about the population mean.
  - Comparative samples (two sample means).
- Use of the t-distribution for the statistic

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

requires that $X_1, X_2, \ldots, X_n$ be normal.

0.4 F-distribution

- The F-distribution finds enormous application in comparing sample variances.
- **Theorem 8.6**:

Let U and V be two independent random variables having chi-squared distribution with ν_1 and ν_2 degrees of freedom, respectively.
Then the distribution of the random variable $F = \frac{U/\nu_1}{V/\nu_2}$ is given by the density

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

This is known as the F-**distribution** with ν_1 and ν_2 degrees of freedom (d.f.).

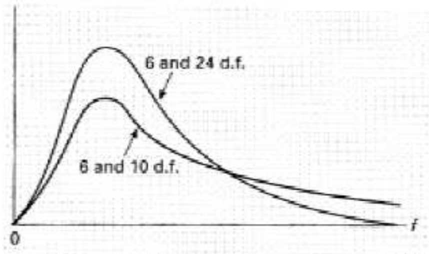

Figure 9: Typical F-distributions.

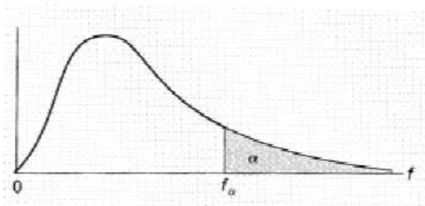

Figure 10: Illustration of the f_α for the F-distribution.

- **Theorem 8.7**:

Writing $f_\alpha(\nu_1, \nu_2)$ for f_α with ν_1 and ν_2 degrees of freedom, we obtain

$$f_{1-\alpha}(\nu_1, \nu_2) = \frac{1}{f_\alpha(\nu_2, \nu_1)}$$

- E.g., f-value with 6 and 10 degrees of freedom, leaving an area of 0.95 to the right,

$$f_{0.95}(6, 10) = \frac{1}{f_{0.05}(10, 6)} = \frac{1}{4.06} = 0.246$$

- **F-distribution with two sample variances**. Suppose that random samples of size n_1 and n_2 are selected from two normal populations with variances σ_1^2 and σ_2^2

$$X_1^2 = \frac{(n_1 - 1)S_1^2}{\sigma_1^2} \; and \; X_2^2 = \frac{(n_2 - 1)S_2^2}{\sigma_2^2} \; (from\ Theorem\ 8.4)$$

Let $U = X_1^2$ and $V = X_2^2$ having chi-squared distribution with $\nu_1 = n_1 - 1$ and $\nu_2 = n_2 - 1$ degrees of freedom.

- Using Theorem 8.6, we obtain the following result (theorem:)

- **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 σ_1^2 and σ_2^2, respectively, then

$$F = \frac{U/\nu_1}{V/\nu_2} = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2}$$

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

- **What is the F-distribution used for?**

- F-distribution is called the **variance ratio distribution**.
  - It is used two-sample situations to draw inferences about the population variances. (Theorem 8.8)
  - It is also applied to many other types of problems in which the sample variances are involved.

- Suppose there are three types of paints to compare. We wish to determine if the population means are equivalent.

Paint	Sample Mean	Sample Variance	Sample Size
A	4.5	0.20	10
B	5.5	0.14	10
C	6.5	0.11	10

The notion of the important components of variability is best seen through some simple graphics.

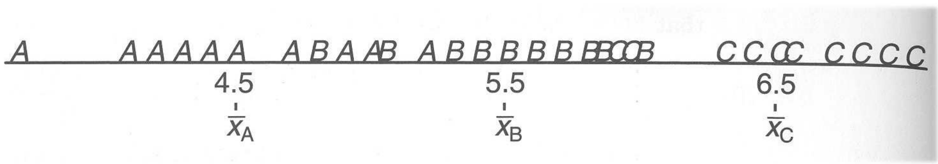

Figure 11: Data from three distinct samples.

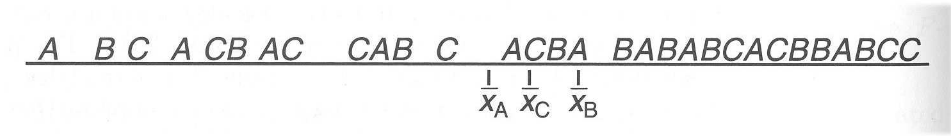

Figure 12: Data that easily could have come from the same population.

- Two key sources of variability:
  1. Variability within samples.
  2. Variability between samples.
- Clearly, if (1) >> (2), the data could all have come from a common distribution.
- The above two sources of variability generate important ratios of sample variances, which are used in conjunction with the F-distribution.
- The general procedure involved is called **analysis of variance.**