Fundamental Sampling Distributions

Fatih Cavdur

Population and Sample

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

A **sample** is a subset of a population.

Random Sample

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

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

Statistic

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

Sample Variance

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

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

Sampling Distribution

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

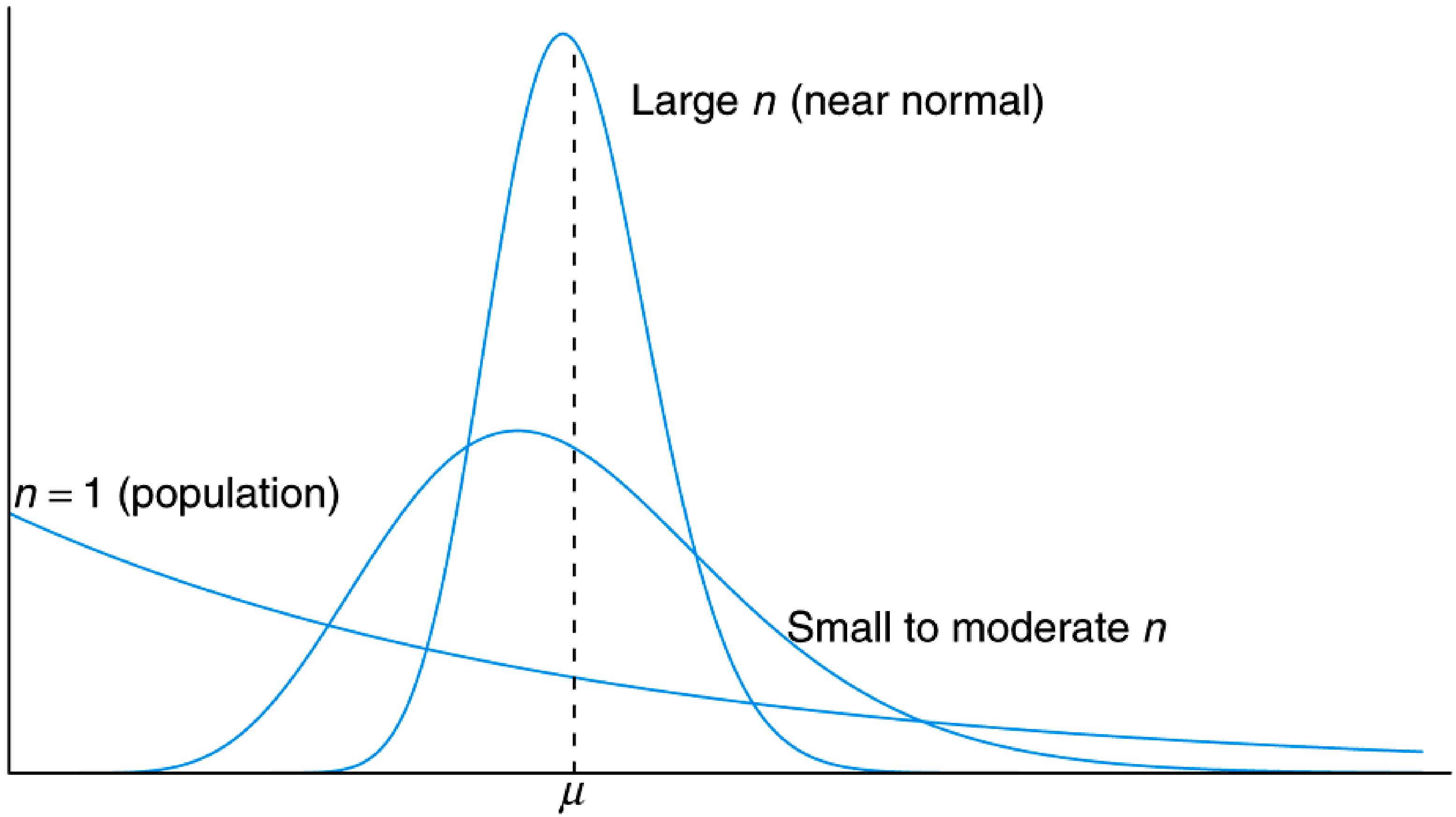
Central Limit Theorem

Central Limit Theorem: If $\bar{X}$ is the mean of a random sample of size n taken from a population with mean μ and finite variance σ^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 $n(z; 0, 1)$.

Central Limit Theorem



Example 8.4

- An electrical firm manufactures light bulbs that have a length of life that is approximately normal with mean 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.

Example 8.4

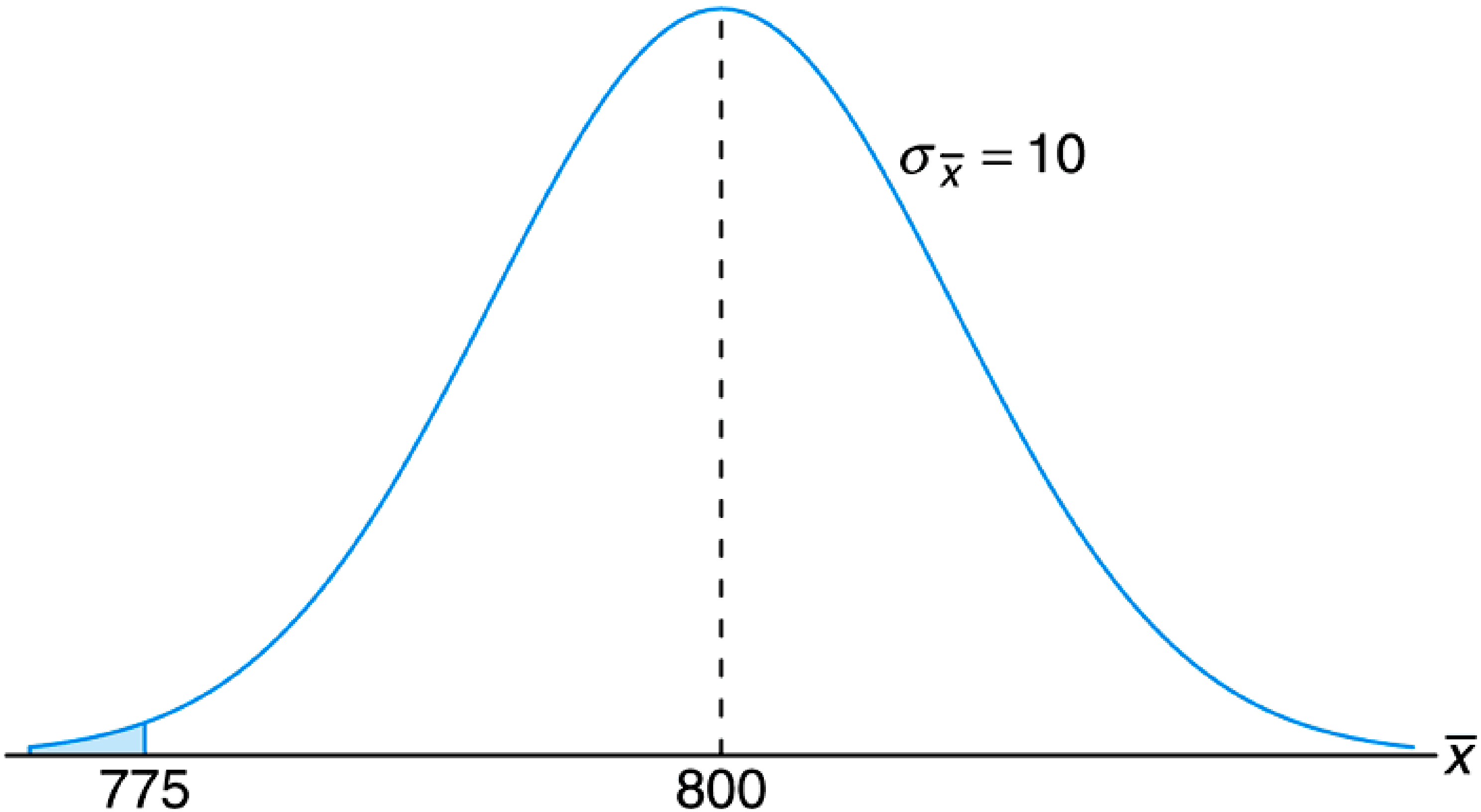
The sampling distribution of $\bar{X}$ is approximately normal with $\mu_{\bar{X}} = 800$ and $\sigma_{\bar{X}} = \frac{40}{\sqrt{16}} = 10$. We then have

$$z = \frac{775 - 800}{10} = -2.5$$

Hence,

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

Example 8.4



Sampling Distribution of Difference of Means

If independent samples 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

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

Hence,

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

is approximately a standard normal variable.

Case Study 8.1

2 independent experiments are run in which 2 different types of paint are compared. 18 units are painted using type A and B and drying times, in hours, are recorded. If the population standard deviation is known to be 1, what is the probability that $P\{\bar{X}_A - \bar{X}_B > 1\}$?

Case Study 8.1

We know that the distribution of the difference is approximately normal with mean and variance

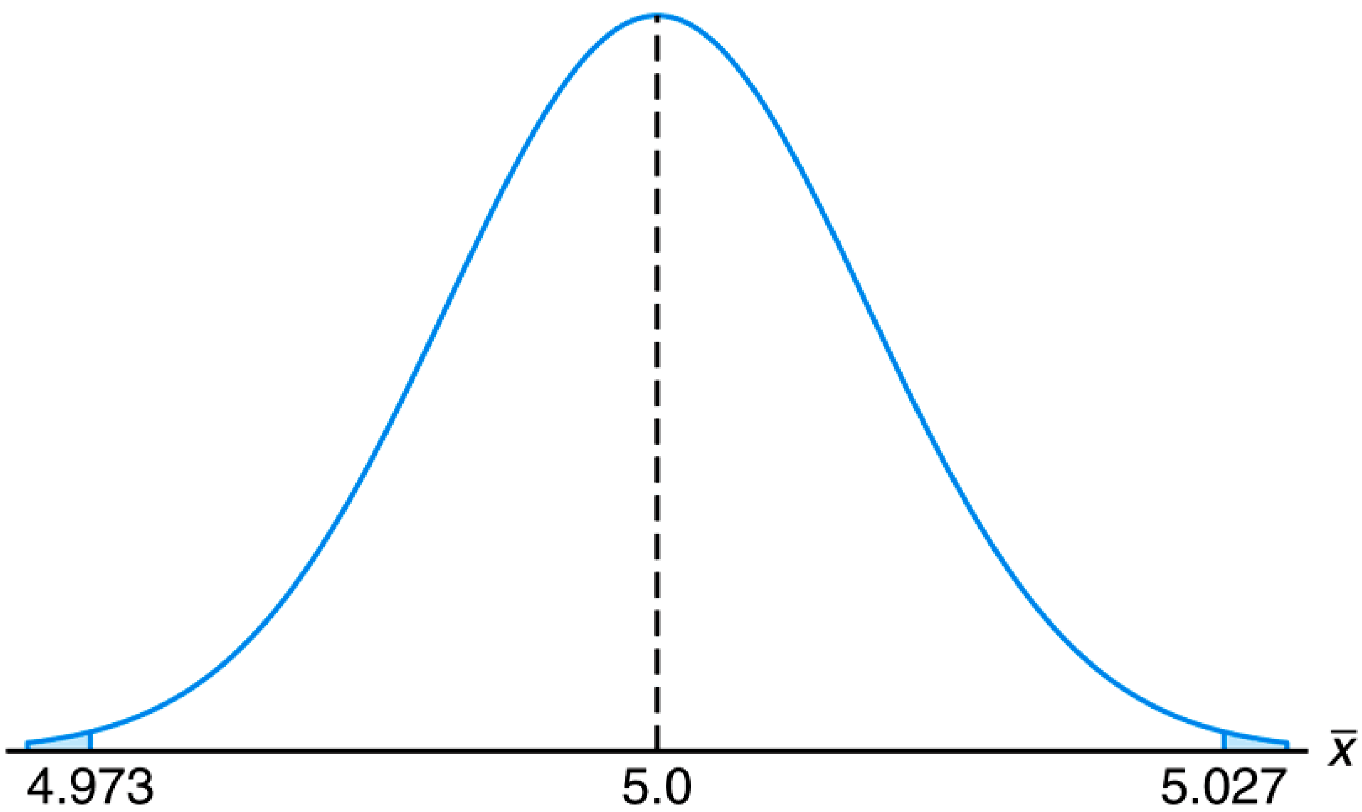
$$\mu_{\bar{X}_A - \bar{X}_B} = \mu_A - \mu_B = 0$$

$$\sigma^2_{\bar{X}_A - \bar{X}_B} = \frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B} = \frac{1}{18} + \frac{1}{18} = \frac{1}{9}$$

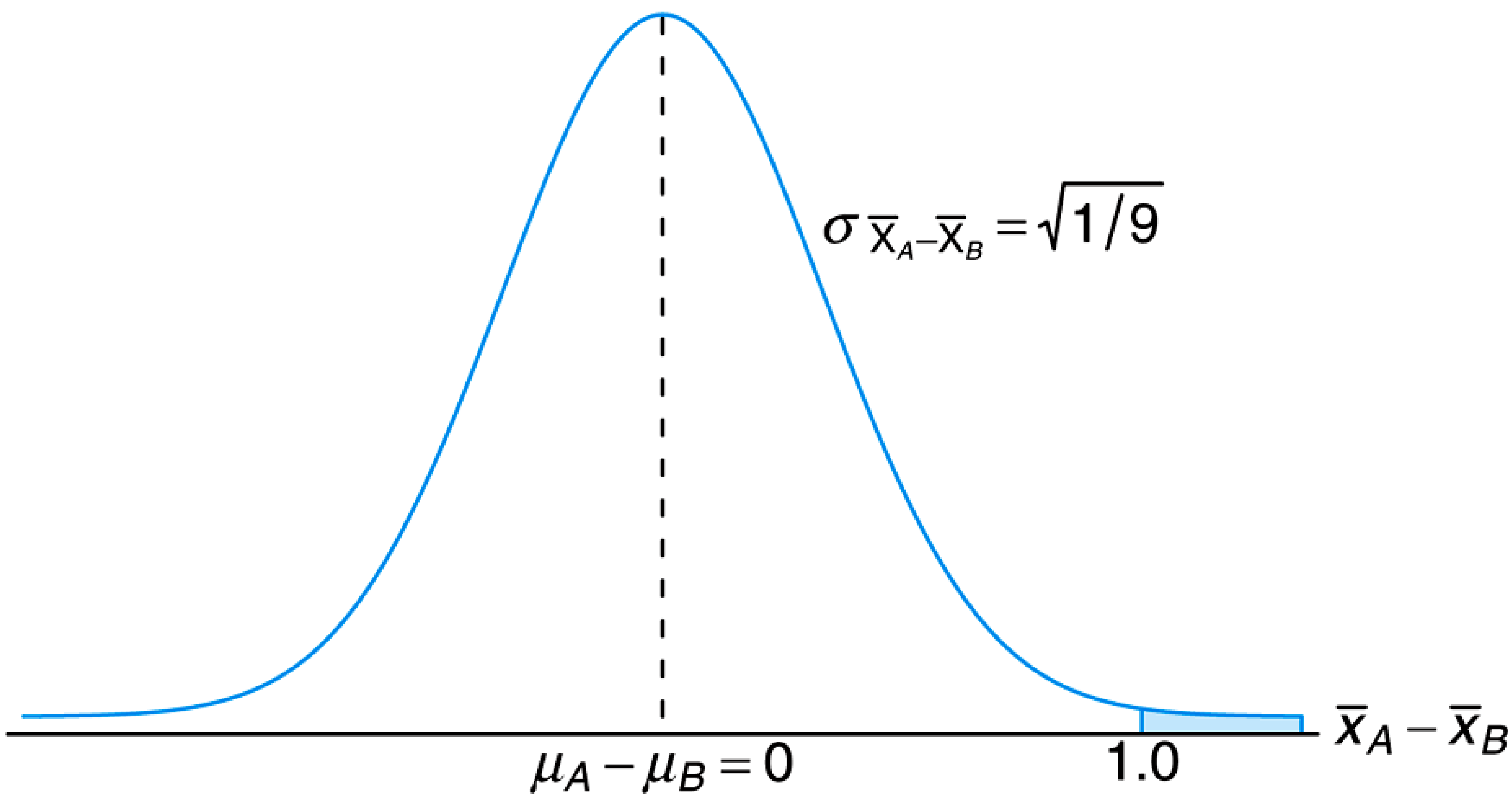
$$z = \frac{1 - (\mu_A - \mu_B)}{\sqrt{1/9}} = \frac{1 - 0}{\sqrt{1/9}} = 3.0$$

$$P(Z > 3.0) = 0.0013$$

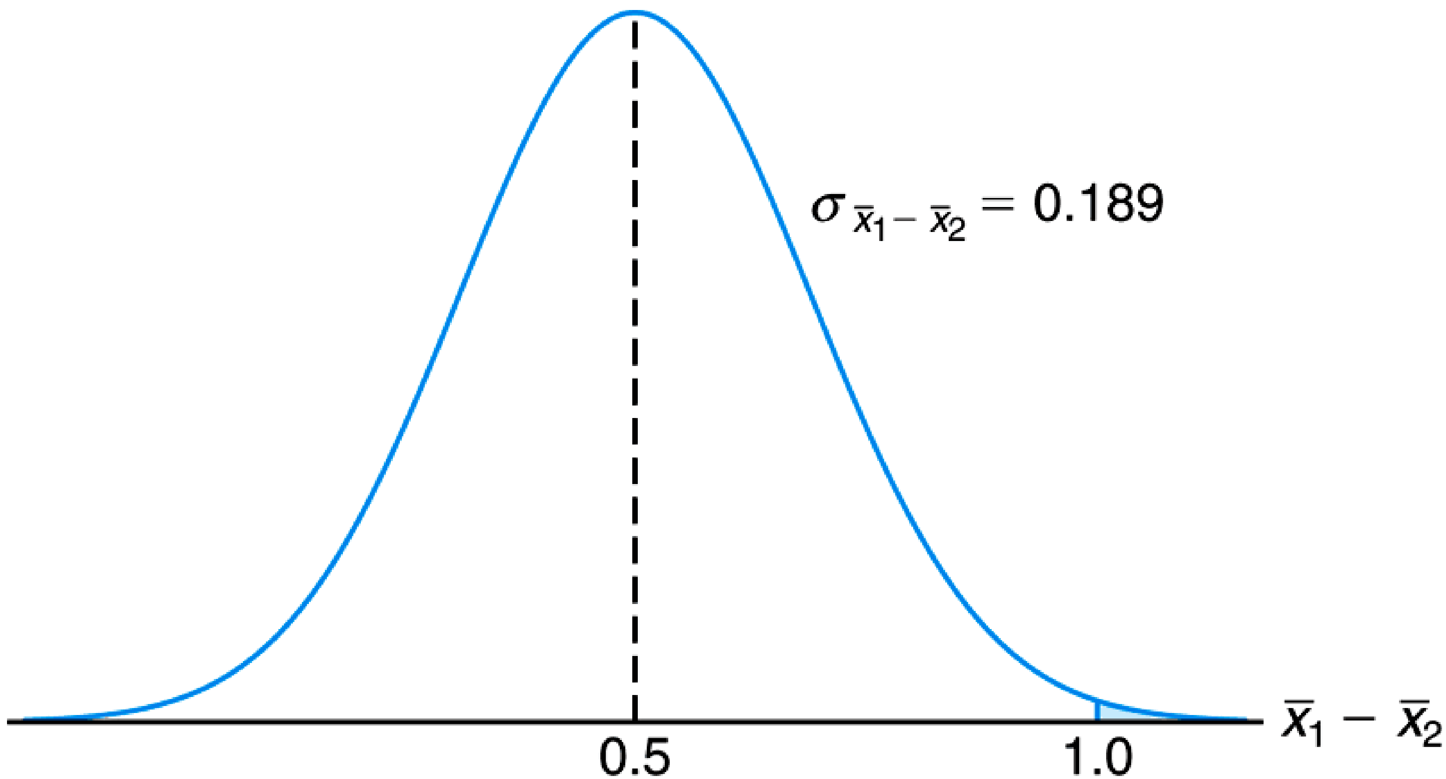
Case Study 8.1



Case Study 8.2



Example 8.6



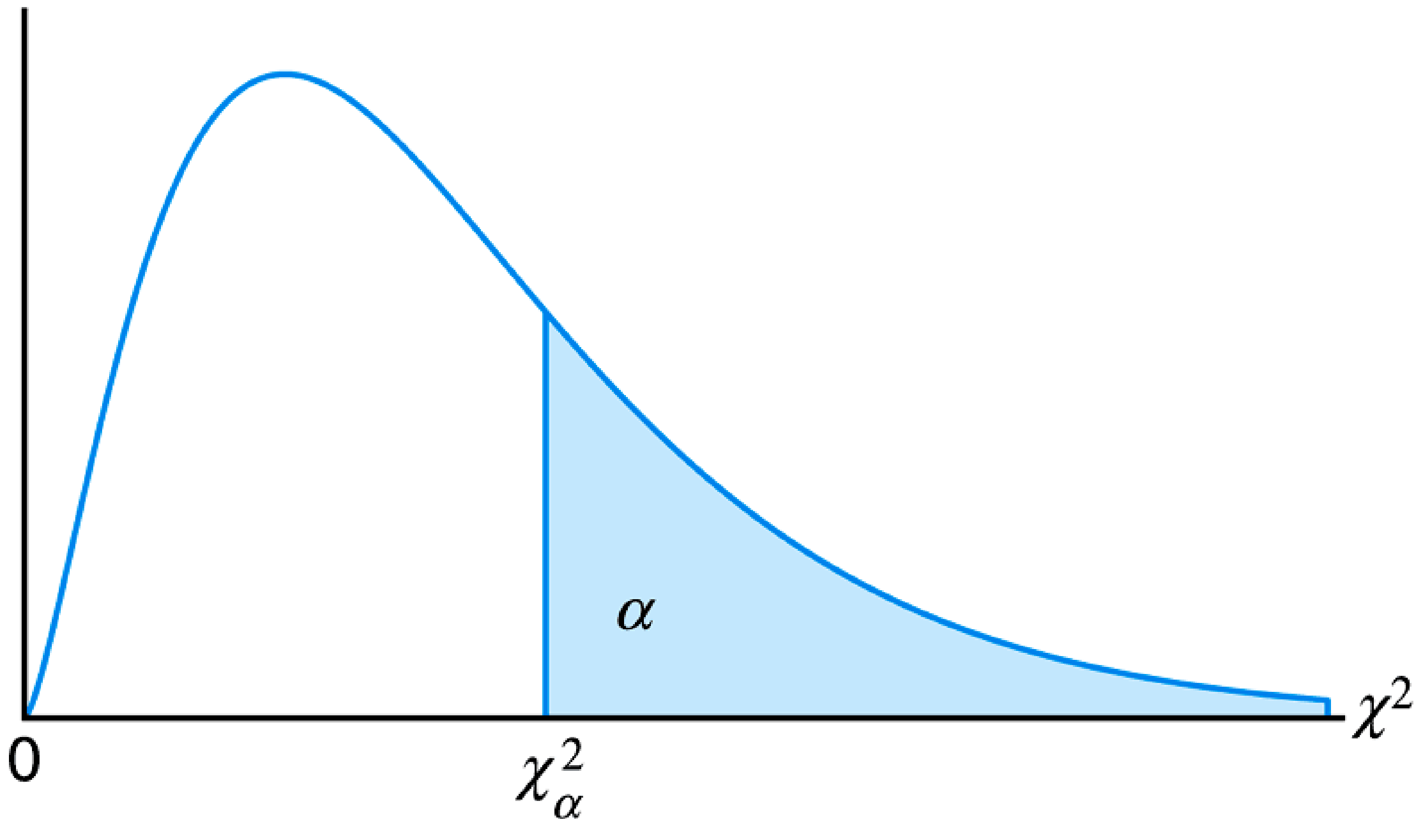
Sampling Distribution of Variance

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 $v = n - 1$ degrees of freedom.

Chi-Square Distribution



t-Distribution

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.

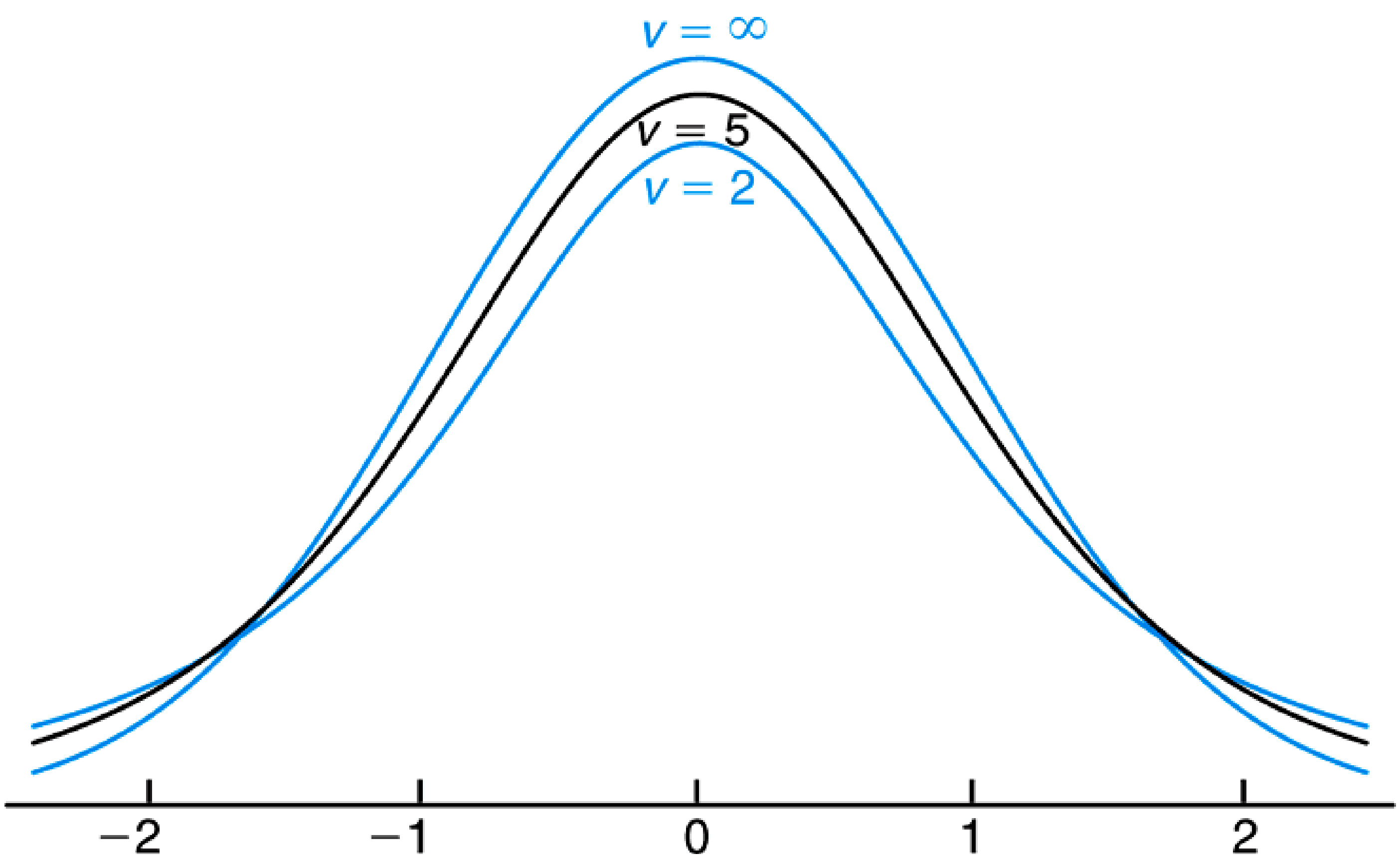
t-Distribution

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

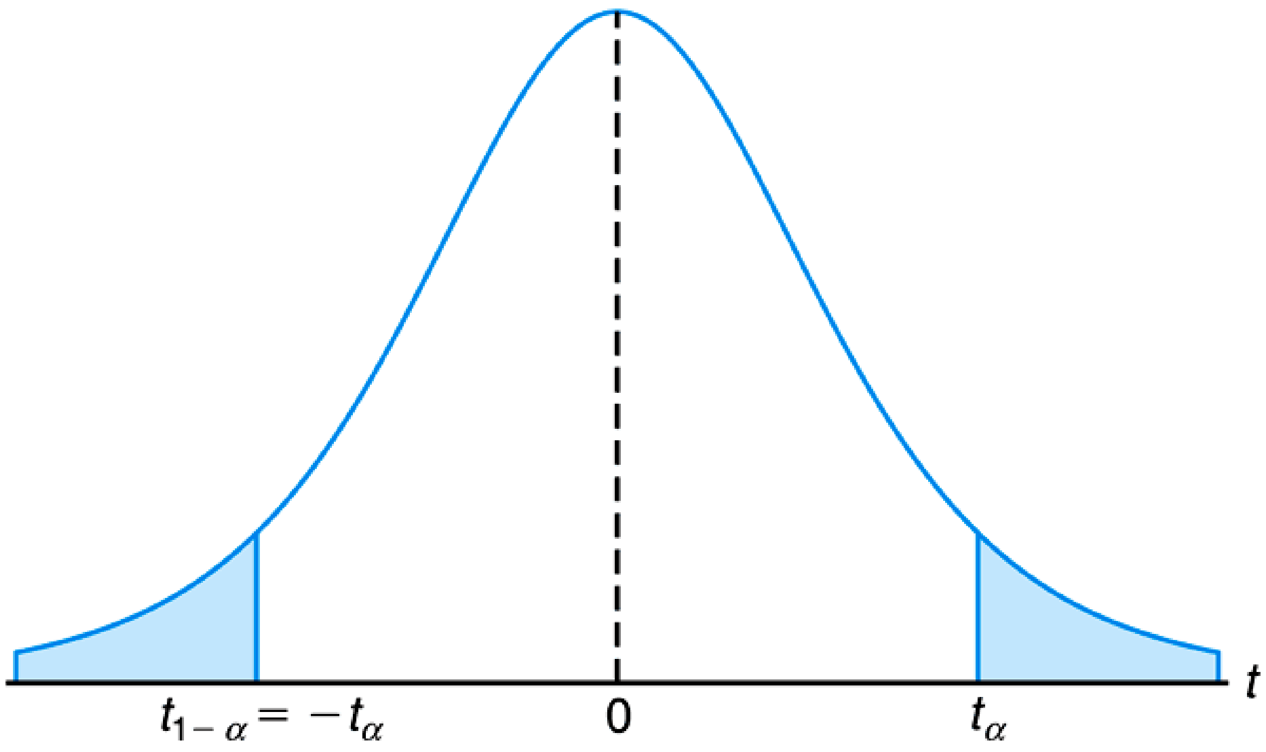
$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \text{and} \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

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

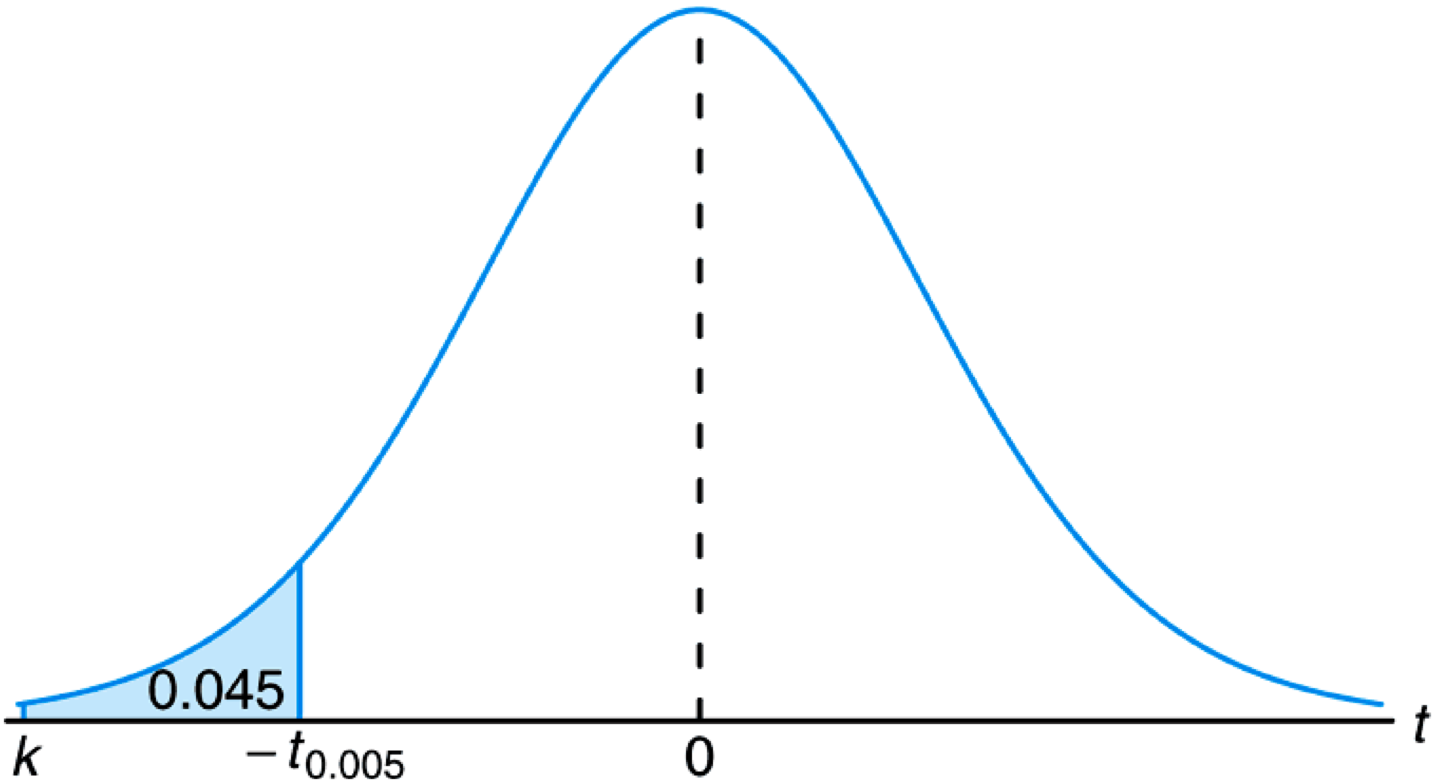
t-Distribution



t-Distribution



Example 8.10



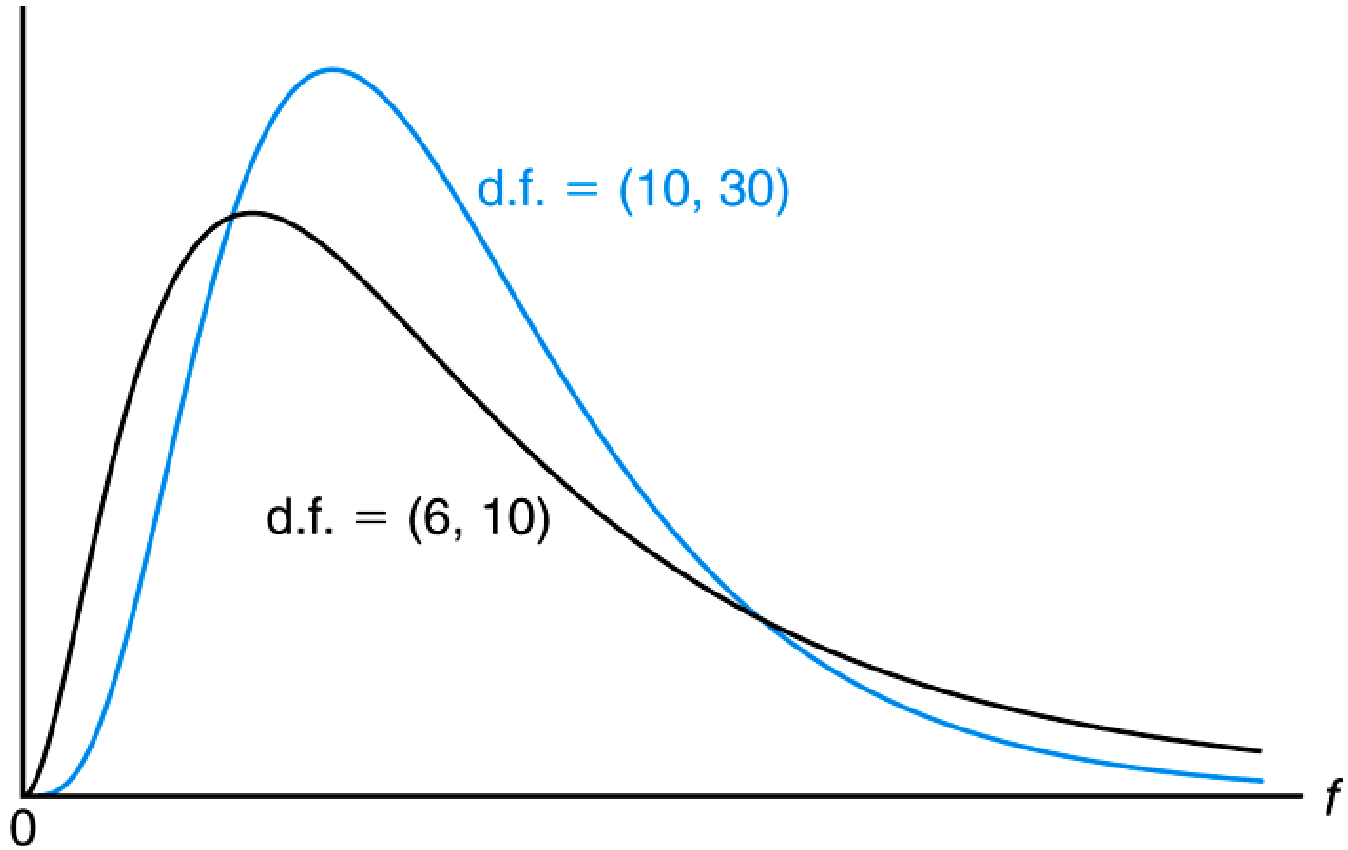
F-Distribution

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

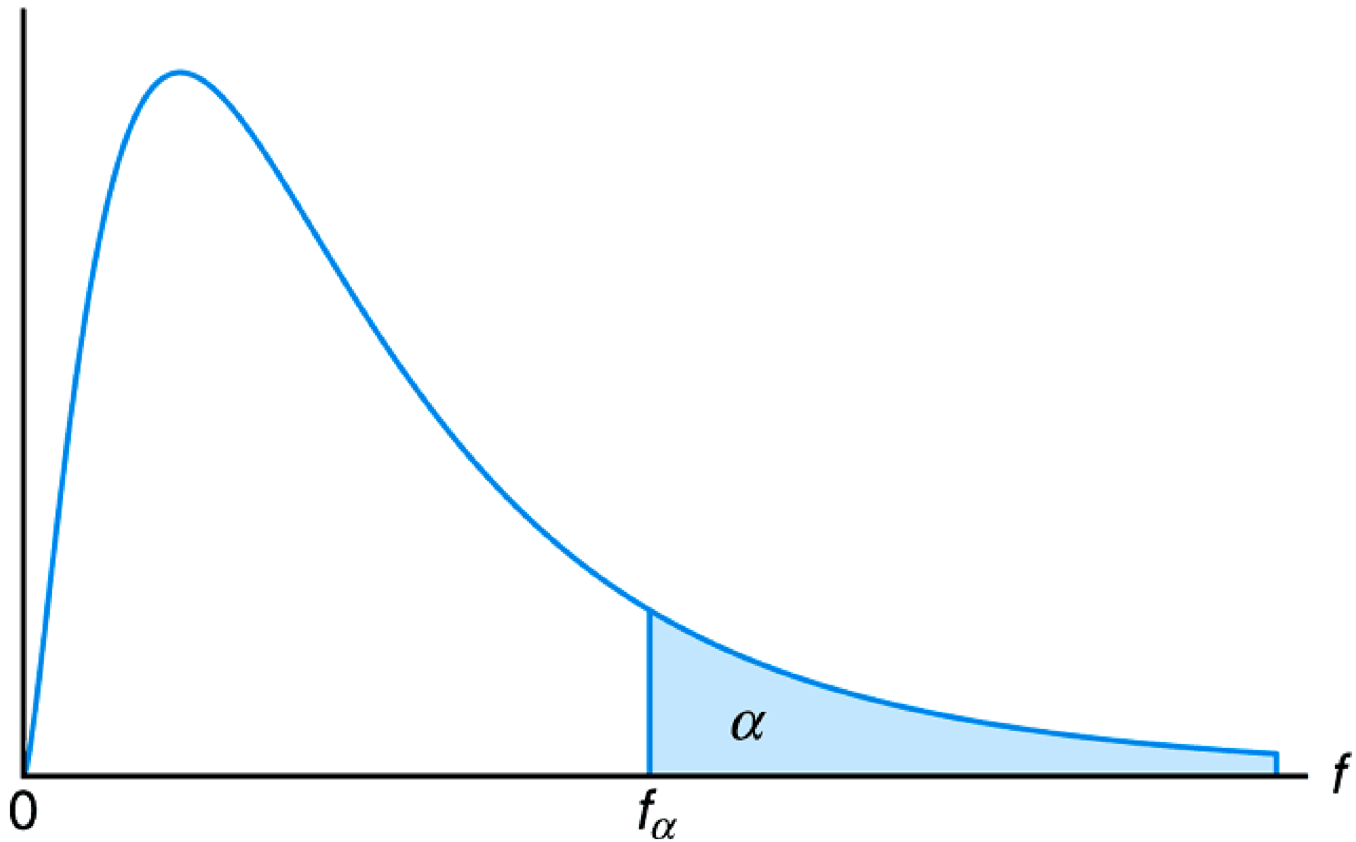
$$h(f) = \begin{cases} \frac{\Gamma[(v_1+v_2)/2](v_1/v_2)^{v_1/2}}{\Gamma(v_1/2)\Gamma(v_2/2)} \frac{f^{(v_1/2)-1}}{(1+v_1 f/v_2)^{(v_1+v_2)/2}}, & f > 0, \\ 0, & f \leq 0. \end{cases}$$

This is known as the ***F*-distribution** with v_1 and v_2 degrees of freedom (d.f.).

F-Distribution



F-Distribution



F-Distribution

Writing $f_\alpha(v_1, v_2)$ for f_α with v_1 and v_2 degrees of freedom, we obtain

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

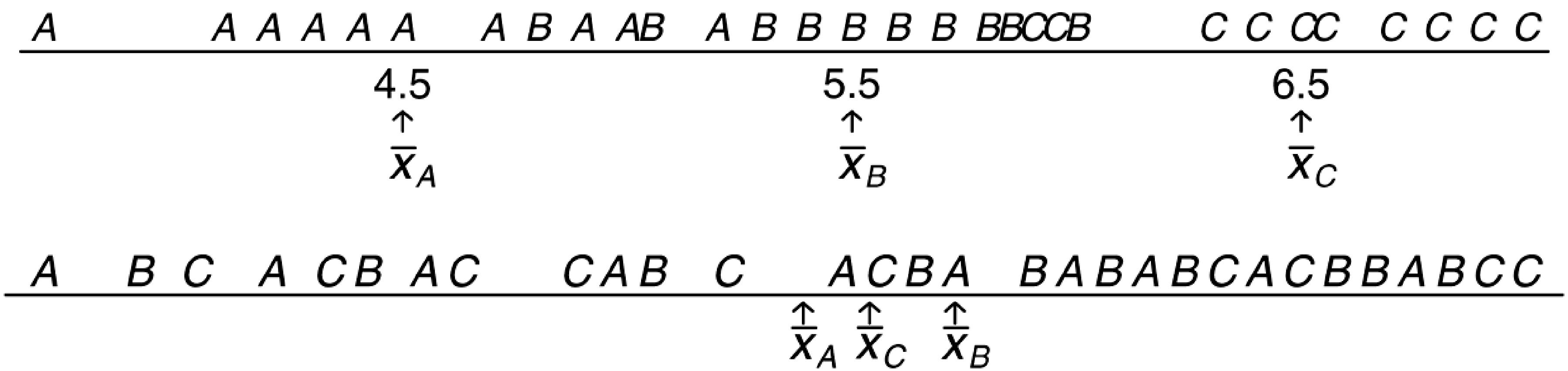
F-Distribution

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

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

has an F-distribution with $v_1 = n_1 - 1$ and $v_2 = n_2 - 1$ degrees of freedom.

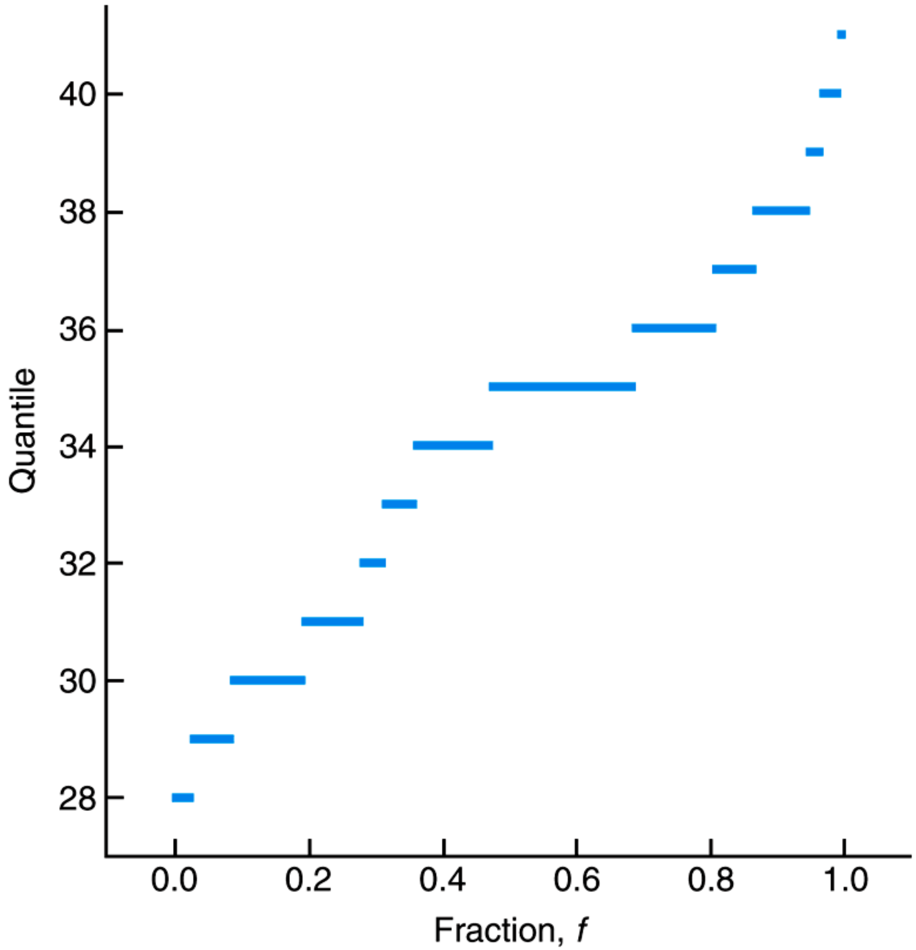
Example Datasets



Quantile

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 $q(f)$.

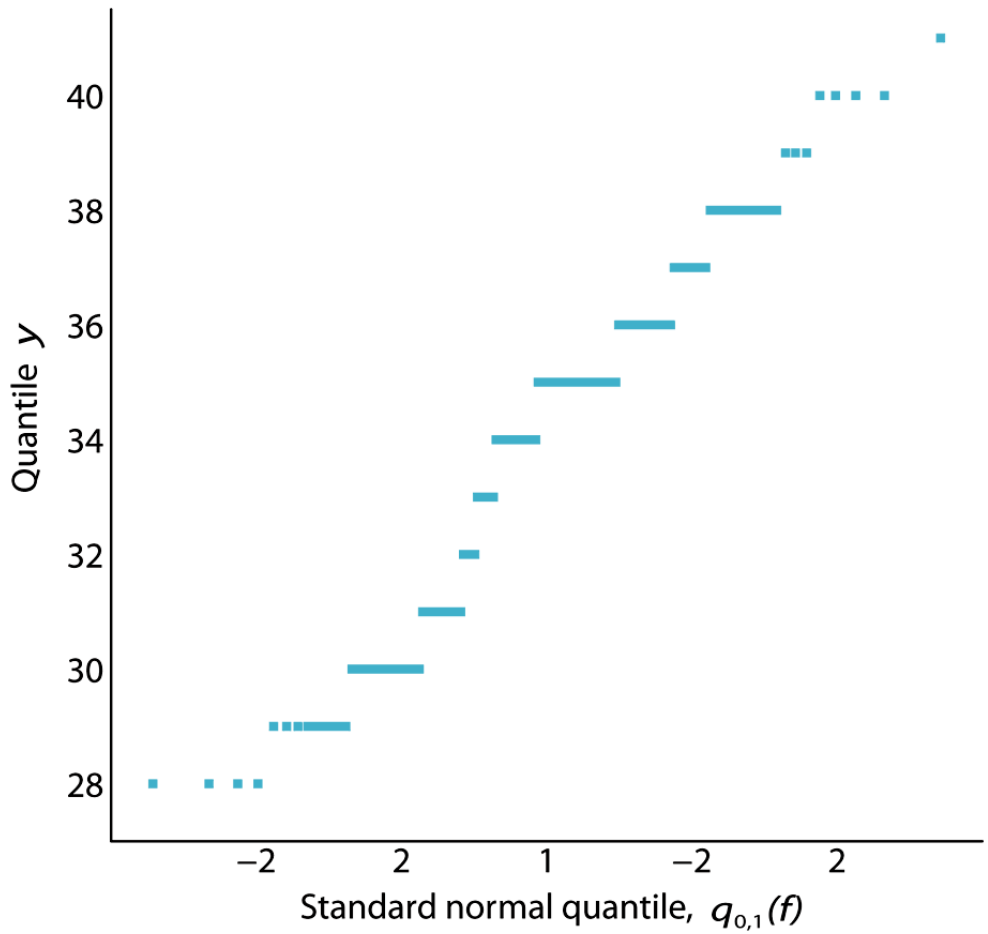
Quantile



Quantile

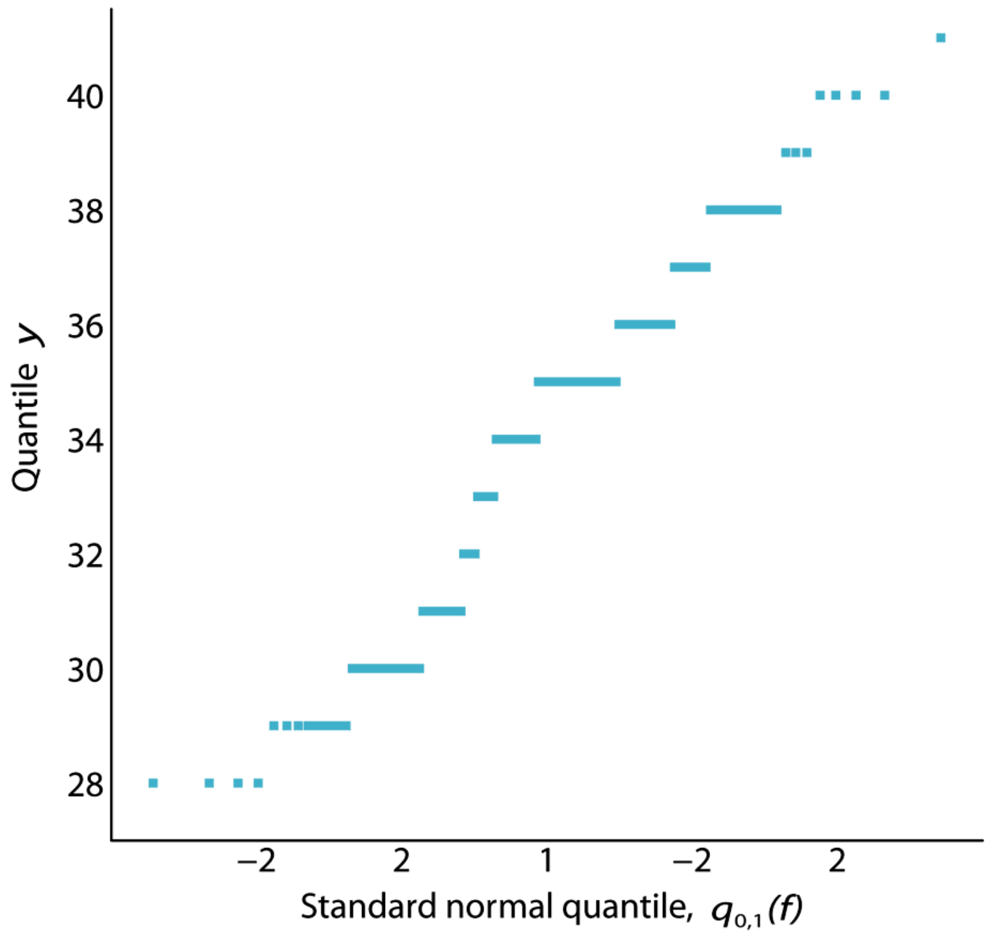
The **normal quantile-quantile plot** is a plot of $y_{(i)}$ (ordered observations) against $q_{0,1}(f_i)$, where $f_i = \frac{i-\frac{3}{8}}{n+\frac{1}{4}}$.

Quantile



40
38
36
34
32
30
28
Quantile y
−2
2
1
−2
2
Standard normal quantile, $q_{0,1}(f)$

Quantile

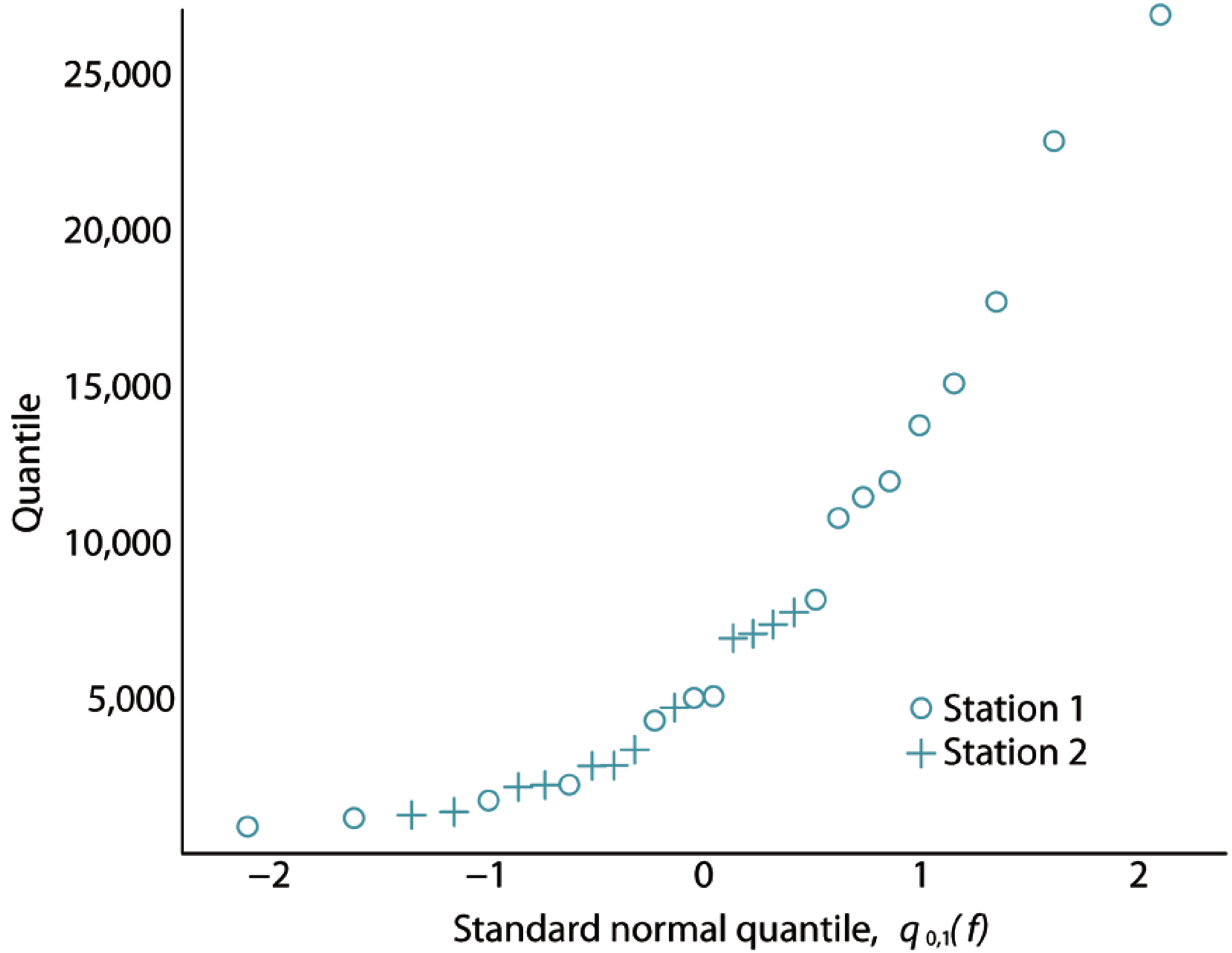


40
38
36
34
32
30
28
Quantile y
−2
2
1
−2
2
Standard normal quantile, $q_{0,1}(f)$

Example 8.12

Number of Organisms per Square Meter			
Station 1		Station 2	
5,030	4,980	2,800	2,810
13,700	11,910	4,670	1,330
10,730	8,130	6,890	3,320
11,400	26,850	7,720	1,230
860	17,660	7,030	2,130
2,200	22,800	7,330	2,190
4,250	1,130		
15,040	1,690		

Example 8.12



25,000
20,000
15,000
10,000
5,000
Quantile
−2
−1
0
1
2
Standard normal quantile, $q_{0,1}(f)$
Station 1
Station 2

End of Lecture ☺