



# Statistical Methods in **Food and Consumer Research**

Second Edition

**Maximo Gacula, Jr.**  
**Jagbir Singh**  
**Jian Bi**  
**Stan Altan**

Food Science and Technology, International Series



# Statistical Methods in Food and Consumer Research

Second Edition

**Maximo C. Gacula, Jr.**

Gacula Associates Consulting, Scottsdale, Arizona

**Jagbir Singh**

Department of Statistics, Temple University, Philadelphia, Pennsylvania

**Jian Bi**

Sensometrics Research and Service, Richmond, Virginia

**Stan Altan**

J & J Pharmaceutical Research and Development, Raritan, New Jersey



AMSTERDAM • BOSTON • HEIDELBERG • LONDON  
NEW YORK • OXFORD • PARIS • SAN DIEGO  
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

Academic Press is an imprint of Elsevier



Academic Press is an imprint of Elsevier  
30 Corporate Drive, Suite 400, Burlington, MA 01803, USA  
525 B Street, Suite 1900, San Diego, California 92101-4495, USA  
84 Theobald's Road, London WC1X 8RR, UK

Copyright © 2009, Elsevier Inc. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from the publisher.

Permissions may be sought directly from Elsevier's Science & Technology Rights, Department in Oxford, UK: phone: (+44) 1865 843830, fax: (+44) 1865 853333, E-mail: [permissions@elsevier.com](mailto:permissions@elsevier.com). You may also complete your request online, via the Elsevier homepage (<http://elsevier.com>), by selecting "Support & Contact", then "Copyright and Permission" and then "Obtaining Permissions."

**Library of Congress Cataloging-in-Publication Data**  
APPLICATION SUBMITTED

**British Library Cataloguing-in-Publication Data**  
A catalogue record for this book is available from the British Library.

ISBN: 978-0-12-373716-8

For information on all Academic Press publications  
visit our Web site at [www.elsevierdirect.com](http://www.elsevierdirect.com)

Printed in the United States of America  
08 09 10 9 8 7 6 5 4 3 2 1

Working together to grow  
libraries in developing countries

[www.elsevier.com](http://www.elsevier.com) | [www.bookaid.org](http://www.bookaid.org) | [www.sabre.org](http://www.sabre.org)

ELSEVIER

BOOK AID  
International

Sabre Foundation

# **Statistical Methods in Food and Consumer Research**

Second Edition

# Food Science and Technology International Series

## **Series Editor**

**Steve L. Taylor**

*University of Nebraska – Lincoln, USA*

## **Advisory Board**

**Ken Buckle**

*The University of New South Wales, Australia*

**Mary Ellen Camire**

*University of Maine, USA*

**Roger Clemens**

*University of Southern California, USA*

**Hildegarde Heymann**

*University of California – Davis, USA*

**Robert Hutkins**

*University of Nebraska – Lincoln, USA*

**Ron S. Jackson**

*Quebec, Canada*

**Huub Lelieveld**

*Bilthoven, The Netherlands*

**Daryl B. Lund**

*University of Wisconsin, USA*

**Connie Weaver**

*Purdue University, USA*

**Ron Wrolstad**

*Oregon State University, USA*

*A complete list of books in this series appears at the end of this volume.*

*To my wife Mely and daughters Karen, Lisa, Elena.*

Maximo C. Gacula, Jr.

*To my wife, Veena, without whom my professional accomplishments would have been impossible  
and with whom each milestone is shared and enjoyed.*

Jagbir Singh

*To my wife Yulin and daughter Cindy.*

Jian Bi

*To Sodnam, Yelen and Tracy.*

Stan Altan

# Preface to the Second Edition

*Statistical Methods in Food and Consumer Research*, published in 1984, is the first book that dealt with the application of statistics to food science and sensory evaluation. Since then, statistical software has become indispensable in the statistical analysis of research data. The use of multivariate analysis in sensory evaluation has emerged in part due to the availability of statistical software that sensory and research analysts can effectively use. The use of nonparametric statistics in sensory evaluation resulted in several scientific research publications which are included in the second edition.

For updating applications of statistics to advancement of sensory science we believe there is a need for a second edition of the book. We are pleased to have Jian Bi with years of experience in applications of statistics, in particular sensometrics, and Dr. Stan Altan with years of expertise in experimental designs as coauthors. Chapters 9–11 of the book were completely re-written and expanded from the basic principles of several statistical procedures to sensory applications. Two chapters were added—Chapter 12, Perceptual Mapping and Descriptive Analysis, and Chapter 13, Sensory Evaluation in Cosmetic Studies. For data analysis we have provided computer programs or codes in S-Plus® and SAS® (Statistical Analysis System) for Chapters 9–13. For Chapter 7 which deals with Response Surface Designs, Design-Expert® is highly recommended.

The revised book can be used as text for a first course in applied statistics for students majoring in Food Science and Technology, and for other students in agricultural and biological sciences. As a first course, Chapters 1–8 will be used. For graduate students majoring in sensory science, a second course would cover Chapters 9–13.

We thank the academic and industrial professionals who found the book useful and encouraged us to update it. We are grateful to Academic Press/Elsevier, Inc. for their interest in publishing the second edition, in particular to Nancy Maragioglio, Carrie Bolger, and Christie Jozwiak for their continuous and readily available support.

Maximo C. Gacula, Jr.  
Jagbir Singh  
Jian Bi  
Stan Altan

# Table of Contents

Preface . . . . . xiii

**1 Introduction . . . . . 1**

    1.1 A Brief Review of Tools for Statistical Inference. . . . . 1

        Notation and Symbolism. . . . . 2

        The Normal Distribution . . . . . 4

        Estimation. . . . . 8

        Testing of Hypotheses . . . . . 9

        Experimental Significance. . . . . 16

    1.2 Principles of Experimental Design . . . . . 16

        Randomization . . . . . 17

        Replication . . . . . 17

        Local Control . . . . . 17

        Blinding . . . . . 18

        Planning the Test Program. . . . . 19

    1.3 The Role of the Statistician in Research . . . . . 21

        Partnership . . . . . 21

        Client–Consultant Relationship . . . . . 21

    Exercises . . . . . 22

**2 Statistical Sensory Testing. . . . . 25**

    2.1 Psychophysical Aspects of Sensory Data . . . . . 25

        Weber’s Law . . . . . 27

        Fechner’s Law . . . . . 27

        Stevens’ Power Law . . . . . 28

        Thurstone’s Law of Comparative Judgment. . . . . 29

    2.2 Scales of Measurement. . . . . 30

        Hedonic Scale . . . . . 32

        Intensity Rating Scale. . . . . 34

        Rankings. . . . . 35

    2.3 Scale Structures . . . . . 35

    2.4 Construction of a Scale . . . . . 37

        Edwards’ Method of Successive Intervals. . . . . 38

        Scaling of Categories . . . . . 46

    2.5 Distribution of Sensory Data . . . . . 50

        Measures of Shape Distribution . . . . . 50

        Statistical Properties of Subjective Data. . . . . 52

        Transformation . . . . . 55

    2.6 Selection of Panel Members. . . . . 57

        Criteria Based on Binomial Response . . . . . 57

        Criteria Based on Rating Scale . . . . . 63



Sequential Analysis . . . . .	64
Quality Control Chart for Degree of Difference. . . . .	69
Exercises . . . . .	74
<b>3 The Analysis of Variance and Multiple Comparison Tests . . . . .</b>	<b>77</b>
3.1 Analysis of Variance. . . . .	77
Assumptions in the Analysis of Variance. . . . .	80
Fixed- and Random-Effects Models. . . . .	81
3.2 Multiple-Comparison Tests . . . . .	98
Least Significant Difference Test . . . . .	99
Dunnett's Test for Multiple Comparisons with a Control. . . . .	100
Tukey's Studentized Range Test . . . . .	101
Duncan's Multiple-Range Test . . . . .	103
Choice of Multiple-Comparison Tests. . . . .	105
3.3 Sample Size Estimation . . . . .	107
Sample Size for Confidence Interval Estimation . . . . .	109
Sample Size for Dichotomous Responses . . . . .	110
Exercises . . . . .	111
<b>4 Experimental Design . . . . .</b>	<b>113</b>
4.1 Simple Comparative Experiments. . . . .	113
Group-Comparison Designs. . . . .	113
Paired Comparison Designs. . . . .	115
4.2 Completely Randomized Designs. . . . .	121
4.3 Randomized Complete Block Designs . . . . .	124
Randomized Complete Block Designs with More Than One Observation per Experimental Unit. . . . .	128
4.4 Latin Square Designs. . . . .	131
Replicating LS Designs. . . . .	134
Partially Replicated Latin Square . . . . .	138
4.5 Cross-Over Designs . . . . .	142
4.6 Split Plot Designs . . . . .	153
Exercises . . . . .	166
<b>5 Incomplete Block Experimental Designs . . . . .</b>	<b>169</b>
5.1 Balanced Incomplete Block Designs . . . . .	169
Parameters of Balanced Incomplete Block Designs . . . . .	170
Intrablock Analysis . . . . .	171
Interblock Analysis . . . . .	177
Combining Intrablock and Interblock Estimates . . . . .	177
5.2 Balanced Incomplete Block Designs Augmented with a Control. . . . .	182
5.3 Doubly Balanced Incomplete Block Designs . . . . .	187
5.4 Composite Complete-Incomplete Block Designs . . . . .	194

Intrablock Analysis . . . . .	198
Exercises . . . . .	202
<b>6 Factorial Experiments . . . . .</b>	<b>205</b>
6.1 The $2^n$ Factorial Experiments . . . . .	206
6.2 The $3^n$ Factorial Experiments . . . . .	216
6.3 The $p \times q$ and $p \times q \times k$ Factorial Experiments . . . . .	225
6.4 Simple Confounding and Fractional Factorial Experiments . . . . .	233
Confounding in $2^n$ Factorial Experiments . . . . .	234
Fractional Factorial . . . . .	237
Exercises . . . . .	244
<b>7 Response Surface Designs and Analysis . . . . .</b>	<b>247</b>
7.1 General Concepts . . . . .	247
Construction of Response Surfaces . . . . .	251
7.2 Fitting of Response Surfaces and Some Design Considerations . . . . .	256
7.3 Illustrations of Fittings of First- and Second-Order Models . . . . .	263
7.4 Composite Designs . . . . .	275
Composite Designs from Fractional Factorials . . . . .	282
A $\frac{1}{2}$ Replicate of $2^4$ . . . . .	282
A $\frac{1}{4}$ Replicate of $2^6$ . . . . .	283
7.5 Rotatable Designs . . . . .	286
Composite Rotatable Designs . . . . .	292
Arrangements of Composite Designs in Blocks . . . . .	294
7.6 A Response Surface Analysis Approach for Sensory Data . . . . .	304
Exercises . . . . .	308
<b>8 Shelf Life Testing Experiments . . . . .</b>	<b>311</b>
8.1 Hazard Function . . . . .	312
8.2 Shelf Life Models . . . . .	316
Normal Distribution . . . . .	317
Lognormal Distribution . . . . .	318
Weibull Distribution . . . . .	320
Graphical Test for Failure Distributions . . . . .	328
8.3 Regression Analysis . . . . .	331
Confidence Intervals . . . . .	340
Nonlinear Relationships . . . . .	345
Exercises . . . . .	348
<b>9 Nonparametric Statistical Methods . . . . .</b>	<b>351</b>
9.1 Some Methods for Binary Data . . . . .	351
Tests for Independent Proportions . . . . .	351

Tests for Dependent Proportions . . . . .	359
Combining $2 \times 2$ Contingency Tables . . . . .	364
Measure of Association for $2 \times 2$ Tables . . . . .	369
<b>9.2 Some Methods Based on Ranks . . . . .</b>	<b>370</b>
Rank Tests for Single and Paired Samples . . . . .	371
Rank Tests for Two Independent Samples . . . . .	378
Rank Tests for Multiple Independent Samples . . . . .	389
Rank Tests for Multiple Related Samples . . . . .	395
Measures of Association for Ranking Data . . . . .	404
<b>9.3 Some Methods for Categorical Data . . . . .</b>	<b>414</b>
Contingency Tables and Chi-square Statistics . . . . .	414
Goodness of Fit Tests . . . . .	418
Analysis of $r \times c$ Contingency Tables . . . . .	422
Measures of Association for $r \times c$ Contingency Tables . . . . .	430
Analysis of Square ( $r \times r$ ) Contingency Tables . . . . .	433
<b>9.4 Multidimensional Contingency Tables . . . . .</b>	<b>438</b>
Introduction . . . . .	438
Simpson's Paradox . . . . .	439
Tests for Independence . . . . .	441
Numerical Example for Tests for Independence . . . . .	443
Exercises . . . . .	449
<b>10 Sensory Difference and Similarity Tests . . . . .</b>	<b>455</b>
<b>10.1 Sensory Difference Tests . . . . .</b>	<b>455</b>
Triangle Test . . . . .	455
Duo-Trio Test . . . . .	471
Pair Difference Test (2-AFC) . . . . .	473
Three-Alternative Forced-Choice Test (3-AFC) . . . . .	477
A-Not A Test . . . . .	478
Same-Different Test . . . . .	484
Comparisons of Difference Test Designs . . . . .	485
<b>10.2 Sensory Similarity (Equivalence) Tests . . . . .</b>	<b>496</b>
Introduction . . . . .	496
Similarity Tests Using Forced-Choice Methods . . . . .	497
Similarity Tests Using the Paired Comparison Method . . . . .	498
Similarity Tests Using A-Not A and Same-Different Methods . . . . .	501
Similarity Tests for Continuous Data . . . . .	503
Similarity Tests for Correlated Data . . . . .	506
Hypothesis Test and Confidence Interval for Similarity Evaluation . . . . .	510
<b>10.3 Replicated Difference and Similarity Tests . . . . .</b>	<b>514</b>
Introduction . . . . .	514
The Beta-Binomial (BB) Model . . . . .	515

Tests Based on the BB Model . . . . .	522
The Corrected Beta-Binomial (CBB) Model . . . . .	532
Tests Based on the CBB Model . . . . .	536
The Dirichlet-Multinomial (DM) Model . . . . .	540
Tests Based on the DM Model . . . . .	545
Exercises . . . . .	552
<b>11 The Method of Paired Comparisons in Sensory Tests and Thurstonian Scaling . .</b>	<b>559</b>
11.1 Paired Comparison Designs . . . . .	559
Completely Balanced Paired Comparison Designs . . . . .	559
Incomplete Balanced Paired Comparison Designs . . . . .	564
11.2 Paired Comparison Models . . . . .	571
The Scheffé Model . . . . .	571
The Bradley—Terry Model . . . . .	582
The Thurstone—Mosteller Model . . . . .	596
Choice of Models . . . . .	605
11.3 Thurstonian Discriminal Distance $d'$ . . . . .	607
Introduction . . . . .	607
Estimation of $d'$ . . . . .	609
Variance of $d'$ . . . . .	614
Tables and S-PLUS Codes for $d'$ and Variance of $d'$ . . . . .	620
Confidence Intervals and Tests for $d'$ . . . . .	621
11.4 Area Under ROC Curve and $R$ -Index . . . . .	624
Introduction . . . . .	624
Estimating $R$ -Index and Its Variance . . . . .	625
Difference Testing Using $R$ -Index . . . . .	627
Similarity Testing Using $R$ -Index . . . . .	629
Linking $R$ -Index with $d'$ . . . . .	630
Same-Different Area Theorem . . . . .	632
Exercises . . . . .	635
<b>12 Descriptive Analysis and Perceptual Mapping . . . . .</b>	<b>641</b>
12.1 Descriptive Analysis . . . . .	641
12.2 Consumer Tests . . . . .	642
In-House Consumer Test . . . . .	642
Home-Use Consumer Test . . . . .	643
Central Location Consumer Test . . . . .	644
12.3 Hedonic and Intensity Rating Scales . . . . .	644
Just-About-Right Rating Scale . . . . .	645
Signal-to-Noise Ratio . . . . .	646
Questionnaire Design . . . . .	653
12.4 Perceptual Mapping . . . . .	659
Factor Analysis . . . . .	660

Principal Component Analysis . . . . .	674
12.5 Preference Mapping . . . . .	679
Exercises . . . . .	685
<b>13 Sensory Evaluation in Cosmetic Studies . . . . .</b>	<b>687</b>
13.1 Experimental Designs . . . . .	687
Clinical Study Data . . . . .	688
Self-Controlled Design . . . . .	691
Parallel Design . . . . .	698
Cross-Over Design . . . . .	701
13.2 Regression Analysis . . . . .	706
Polynomial Regression . . . . .	706
Nonlinear Regression . . . . .	710
Fit of Nonlinear Regression Model . . . . .	718
Exercises . . . . .	719
Appendix . . . . .	721
References . . . . .	829
Index . . . . .	845
Food Science and Technology: International Series . . . . .	851

This introductory chapter is written to familiarize the reader with common symbols, notation, concepts, and some probability tools that are used in the subsequent chapters. This chapter, therefore, is a minimal prerequisite for the remainder of this book. The details of the material presented here can be found in many beginning statistical methods textbooks, such as those by Dixon and Massey (1969) and Huntsberger and Billingsley (1981).

### 1.1 A BRIEF REVIEW OF TOOLS FOR STATISTICAL INFERENCE

The purpose of this section is to review briefly the tools used in *statistical inference*. Statistical inference is the process of making inductive statements based on *random* samples from target populations. The inference process includes the estimation of population parameters and the formulation and testing of hypotheses about the parameters: a *parameter* is a numerical quantity that describes some characteristics of the population. For instance, the mean  $\mu$  and the variance  $\sigma^2$  are examples of population parameters. In the hypothesized simple linear regression model,  $Y = \alpha + \beta X + \varepsilon$ , the intercept  $\alpha$  and the slope  $\beta$  are the parameters of the model.

Because random sampling is a prerequisite for valid statistical inference, it is important that the samples be collected following an accepted random sampling procedure. If this is not done, the inferences about the population are not reliable. A precise and reliable inference can be reached only through the use of an appropriate *experimental design* for random sampling and data collection. The general principles of experimental designs are summarized in Section 1.2. The sample size plays a major role in ensuring that the selected random sample will lead to statistical inferences with a specified amount of confidence.

## Notation and Symbolism

The science of statistics is concerned with the collection, summarization, and interpretation of experimental data. The data may be summarized in the form of graphs, charts, or tables. Because of the many statistical operations to which the data are subjected, a general system of notation is used to facilitate the operations. A working knowledge of this system is essential in understanding the computational procedures involved in statistical analyses. A brief description of the notation that will be encountered in succeeding chapters is given here.

A commonly encountered symbol is sigma,  $\Sigma$ , used to denote a sum or total. For example, the sum of  $n$  numbers  $X_1, X_2, \dots, X_n$  is denoted by

$$\sum_{i=1}^n X_i = X_1 + X_2 + \dots + X_n. \quad (1.1-1)$$

The lower and upper limits on  $\Sigma$  indicate that the  $X_i$  is added from  $i = 1$  to  $i = n$ ; for simplicity, the limits on  $\Sigma$  may be omitted if no confusion is likely. Similarly, we have the following:

$$\begin{aligned} \sum X_i^2 &= X_1^2 + X_2^2 + \dots + X_n^2, \\ \sum X_i Y_i &= X_1 Y_1 + X_2 Y_2 + \dots + X_n Y_n, \\ \sum cX &= cX_1 + cX_2 + \dots + cX_n = c \sum_{i=1}^n X_i, \end{aligned} \quad (1.1-2)$$

where  $c$  is a constant. Note that  $(\Sigma X_i)^2$ , which denotes the square of the sum defined by (1.1-1), is not the same as that given by (1.1-2).

Whereas the population mean is symbolized by  $\mu$ , the mean of a sample  $X_1, X_2, \dots, X_n$  of  $n$  items is denoted by  $\bar{X}$ , where

$$\bar{X} = \sum X_i / n. \quad (1.1-3)$$

Similarly, the population variance is denoted by  $\sigma^2$ , where the sample variance is denoted by  $S^2$ , where

$$S^2 = \sum (X_i - \bar{X})^2 / n - 1. \quad (1.1-4)$$

Whenever the population parameters are unknown, as is usually the case, they are estimated by appropriate sample quantities called *statistics*. Thus, if population mean  $\mu$  and variance  $\sigma^2$  are unknown, they may be estimated by statistics  $\bar{X}$  and  $S^2$ , respectively. At times an estimate of a parameter, say  $\theta$ , is denoted by  $\hat{\theta}$ . Thus, an estimate of the population  $\mu$  can be denoted by  $\hat{\mu}$ , which may or may not be  $\bar{X}$ .

Depending on the populations under investigation, appropriate symbols can be made more explicit. For example, if more than one population is under investigation in a study, their means can be denoted by  $\mu_1$ ,  $\mu_2$ , and so on.

In a multiclassified data structure, the dot notation is used to indicate totals. For example, let  $X_{ijk}$  denote the  $k$ th observation in the  $(i, j)$ th cell. Then  $X_{ij\cdot}$  can be used to denote the sum of  $n_{ij}$  observations in the  $(i, j)$ th cell. That is  $X_{ij\cdot} = \sum_{k=1}^{n_{ij}} X_{ijk}$ . If there is only one observation per cell, that is,  $n_{ij} = 1$ , then there is no use in retaining the subscript  $k$ , and one may express the data in a row and column structure. For example, Table 1.1 consists of  $r$  rows and  $c$  columns; the observation in the  $i$ th row and the  $j$ th column is denoted by  $X_{ij}$ . As noted earlier, when  $i$  or  $j$  or both are replaced by a dot ( $\cdot$ ), the dot denotes the sum over the replaced subscript. Therefore,

$$X_{\cdot\cdot} = \sum_{j=1}^c \sum_{i=1}^r X_{ij} = G$$

is the grand total, and the grand mean is given by

$$\bar{X} = X_{\cdot\cdot}/rc = \hat{\mu}.$$

It follows that (see Table 1.1)  $X_{i\cdot}$  and  $X_{\cdot j}$  are the sums of the  $i$ th row and  $j$ th column, respectively. Thus,  $\bar{X}_{i\cdot} = X_{i\cdot}/c$ , and  $\bar{X}_{\cdot j} = X_{\cdot j}/r$ . This system is easily extended to a more complex data structure such as  $X_{ijk\cdots p}$ , where the number of subscripts corresponds to the number of classifications. If  $n_{ijk\cdots p}$  denotes the number of observations in the  $(i, j)$ th cell, the dot notation representation is similarly used. For

**Table 1.1** An  $r \times c$  Table with One Observation per Cell

Rows	Columns				Row sums	Row means
	1	2	...	c		
1	$X_{11}$	$X_{12}$	...	$X_{1j}$	$X_{1\cdot}$	$\bar{X}_{1\cdot}$
2	$X_{21}$	$X_{22}$	...	$X_{2j}$	$X_{2\cdot}$	$\bar{X}_{2\cdot}$
:	:	:	:	:	:	:
$r$	$X_{r1}$	$X_{r2}$	...	$X_{rj}$	$X_{r\cdot}$	$\bar{X}_{r\cdot}$
Column sums	$X_{\cdot 1}$	$X_{\cdot 2}$	...	$X_{\cdot c}$	$X_{\cdot\cdot}$ (grand total) = $G$	
Column means	$\bar{X}_{\cdot 1}$	$\bar{X}_{\cdot 2}$	...	$\bar{X}_{\cdot c}$	$\bar{X}$ (grand mean)	



example,  $n...$  denotes the total number of observations in the experiment;  $n_{i..}$  denotes the number of observations summed over the  $j$  and  $k$  classifications.

### The Normal Distribution

Researchers design and conduct experiments to observe one or more quantities whose numerical values cannot be predicted with certainty in advance. The quantities of interest may be the shelf life or the preference score of one item, or some sensory characteristics of, say, food products. The shelf life, the preference score, etc., of an item is not known in advance but could be any one of a number of possible values. A quantity that can assume any one numerical value from a range of possible values, depending on the experimental outcomes, is called a *random variable* (r.v.) in statistical terminology. Since the numerical values of an r.v. depend on the experimental outcomes, the values of an r.v. are governed by a chance model or a probability distribution, as it is called in statistical language. In short, associated with each r.v. is a probability distribution, and statisticians use this structure to study various aspects of the quantities (r.v.'s) of interest.

Some probability distributions can be described by close mathematical functions called the *probability density functions* (pdf's). Once the pdf of an r.v.  $X$  is given, it can be used to calculate probabilities of events of interest in decision making. In what follows now we introduce one important probability distribution through its pdf and a number of others without writing down their pdf's.

The most important probability distribution in applied statistics is the *normal* or *Gaussian probability distribution*. An r.v.  $X$  is said to have a normal probability distribution if its pdf is of the mathematical form

$$f(X) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[ -\frac{(X - \mu)^2}{2\sigma^2} \right], \quad (1.1-5)$$

where  $-\infty < X < \infty$ ,  $-\infty < \mu < \infty$ , and  $\sigma^2 > 0$ . The graph of  $f(X)$  is sketched in Figure 1.1. The graph is a bell-shaped curve that is symmetric around  $\mu$ . Hence, the parameter  $\mu$  is the center of the distribution of the values of  $X$ . The shape of the curve depends on  $\sigma^2$ . The curve tends to be flat in shape as the shape parameter  $\sigma^2$  increases. The normal probability distribution is characterized completely by