

FD

**Techniques of Water-Resources Investigations of the United States Geological Survey**

**Book 4, Hydrologic Analysis and Interpretation**

**Chapter A3**

# **Statistical Methods in Water Resources**

**By D.R. Helsel and R.M. Hirsch**

U.S. DEPARTMENT OF THE INTERIOR  
GALE A. NORTON, Secretary

U.S. GEOLOGICAL SURVEY  
Charles G. Groat, Director

September 2002

The use of firm, trade, and brand names in this report is for identification purposes only and does not constitute endorsement by the U.S. Geological Survey.

---

Publication available at:  
<http://water.usgs.gov/pubs/twri/twri4a3/>

# ERRATA

Please attach to USGS TWRI Book 4 Chapter A3

**"Statistical Methods in Water Resources," by D.R. Helsel and R.M. Hirsch**

1. **Page 46, 51** — "Lees Ferry, Colorado" was changed to "Lees Ferry, Arizona."
2. **Page 279** — In figure 10.8, the slope of method LNS was changed so that the term  $(-A)$  was included in the numerator.
3. **Page 476** — In the third equation under the heading "Large-sample approximation,"  $s$  was changed to Greek lowercased sigma.
4. Various typographical changes were made that do not change the meaning or content of the report.



# Table of Contents

---

Preface	xi
<b>Chapter 1 Summarizing Data</b>	<b>1</b>
1.1 Characteristics of Water Resources Data	2
1.2 Measures of Location	3
1.2.1 Classical Measure -- the Mean	3
1.2.2 Resistant Measure -- the Median	5
1.2.3 Other Measures of Location	6
1.3 Measures of Spread	7
1.3.1 Classical Measures	7
1.3.2 Resistant Measures	8
1.4 Measures of Skewness	9
1.4.1 Classical Measure of Skewness	9
1.4.2 Resistant Measure of Skewness	10
1.5 Other Resistant Measures	10
1.6 Outliers	11
1.7 Transformations	12
1.7.1 The Ladder of Powers	12
<b>Chapter 2 Graphical Data Analysis</b>	<b>17</b>
2.1 Graphical Analysis of Single Data Sets	19
2.1.1 Histograms	19
2.1.2 Stem and Leaf Diagrams	20
2.1.3 Quantile Plots	22
2.1.4 Boxplots	24
2.1.5 Probability Plots	26
2.2 Graphical Comparisons of Two or More Data Sets	35
2.2.1 Histograms	35
2.2.2 Dot and Line Plots of Means, Standard Deviations	35
2.2.3 Boxplots	38
2.2.4 Probability Plots	40
2.2.5 Q-Q Plots	41
2.3 Scatterplots and Enhancements	45

2.3.1	Evaluating Linearity	45
2.3.2	Evaluating Differences in Location on a Scatterplot	47
2.3.3	Evaluating Differences in Spread	50
2.4	Graphs for Multivariate Data	51
2.4.1	Profile Plots	51
2.4.2	Star Plots	53
2.4.3	Trilinear Diagrams	56
2.4.4	Plots of Principal Components	58
2.4.5	Other Multivariate Plots	59
<b>Chapter 3</b>	<b>Describing Uncertainty</b>	<b>65</b>
3.1	Definition of Interval Estimates	66
3.2	Interpretation of Interval Estimates	67
3.3	Confidence Intervals for the Median	70
3.3.1	Nonparametric Interval Estimate for the Median	70
3.3.2	Parametric Interval Estimate for the Median	73
3.4	Confidence Intervals for the Mean	74
3.4.1	Symmetric Confidence Interval for the Mean	75
3.4.2	Asymmetric Confidence Interval for the Mean	76
3.5	Nonparametric Prediction Intervals	76
3.5.1	Two-Sided Nonparametric Prediction Interval	77
3.5.2	One-Sided Nonparametric Prediction Interval	78
3.6	Parametric Prediction Intervals	80
3.6.1	Symmetric Prediction Interval	80
3.6.2	Asymmetric Prediction Intervals	80
3.7	Confidence Intervals for Percentiles (Tolerance Intervals)	82
3.7.1	Nonparametric Confidence Intervals for Percentiles	83
3.7.2	Nonparametric Tests for Percentiles	84
3.7.3	Parametric Confidence Intervals for Percentiles	88
3.7.4	Parametric Tests for Percentiles	90
3.8	Other Uses for Confidence Intervals	90
3.8.1	Implications of Non-Normality for Detection of Outliers	90
3.8.2	Implications of Non-Normality for Quality Control	91
3.8.3	Implications of Non-Normality for Sampling Design	93
<b>Chapter 4</b>	<b>Hypothesis Tests</b>	<b>97</b>
4.1	Classification of Hypothesis Tests	99
4.1.1	Classification Based on Measurement Scales	99
4.1.2	Classification Based on the Data Distribution	100

Page on Disk

7362

