The Six Sigma Practitioner's Guide to Data Analysis

Second Edition

Donald J. Wheeler
Fellow of the American Statistical Association
Fellow of the American Society for Quality

SPC Press
Knoxville, Tennessee

Contents

	About the Author		*xi*
	Acknowledgments		*xii*
	Preface to the Second Edition		*xii*
	Introduction		*xiii*
PART ONE:	THE FOUNDATIONS OF DATA ANALYSIS		1
One	**Four Statistical Problems**		**3**
	1.1	Descriptive Statistics	3
	1.2	Probability Theory	5
	1.3	Statistical Inference	6
	1.4	The Homogeneity Question	8
	1.5	Two Perspectives on These Problems	9
	1.6	Axioms of Data Analysis	12
	1.7	Observational Studies and Experimental Studies	16
	1.8	Summary	17
Two	**Descriptive Statistics and Homogeneity**		**19**
	2.1	What Does the Standard Deviation Statistic Do?	19
	2.2	What Descriptive Statistics Do Not Do	29
	2.3	Local Measures of Dispersion	30
	2.4	Are the Data Homogeneous?	33
	2.5	The Difference between Global and Local	38
Three	**Process Behavior Charts**		**41**
	3.1	The Chart for Individual Values	41
	3.2	What Do We Gain from the *mR* Chart?	44
	3.3	What Makes the *XmR* Chart Work?	46
	3.4	How Many Values Do I Need for Limits?	50
	3.5	Rules for Detecting Nonhomogeneity	53
	3.6	Average and Range Charts	58
	3.7	Chunky Data	62
	3.8	Caution regarding Software	65
	3.9	Where Do We Go from Here?	66

Four **Statistics, Parameters, and Inference** **67**
4.1 The Concept of a Probability Model 67
4.2 Some Cautions regarding Probability Models 71
4.3 Elements of Statistical Inference 72
4.4 Interval Estimates of Location 76
4.5 Interval Estimates of Dispersion 84
4.6 Practical Statistical Inference 86
4.7 Interpreting "Degrees of Freedom" 88
4.8 Summary 91

PART TWO: THE TECHNIQUES OF DATA ANALYSIS 93

Five **Data Collected under One Condition** **95**
5.1 What Can We Say about Our Process? 95
5.2 But What about the Significance Levels? 99
5.3 Bead Board No. 3 100
5.4 The Data for NB10 105
5.5 Summary 107

Six **Data Collected under Two Conditions** **111**
6.1 Detecting a Difference with Histograms 111
6.2 Detecting a Difference with *XmR* Charts 113
6.3 Using an Average and Range Chart 116
6.4 The Analysis of Means 119
6.5 The Two-Sample Student's *t*-Test 122
6.6 The Paired *t*-Test 125
6.7 NB10 Revisited 128
6.8 Summary of Comparing Two Conditions 130

Seven **Data Collected under Three or More Conditions** **133**
7.1 *XmR* Charts for Each Treatment 133
7.2 Average and Range Charts 135
7.3 Analysis of Means 136
7.4 Analysis of Variance 137
7.5 The Tukey Post-Hoc Test 145
7.6 NB10 Again 146
7.7 Summary of Comparing Several Conditions 148

Eight	**Data Collected at Three or More Values for *X***		**151**
	8.1	The Universe Had a Definite Beginning	151
	8.2	Evaluating Terms in the Model	154
	8.3	Regression: One Line or Two?	156
	8.4	A Deterministic Relation with Controlled *X* Values	158
	8.5	A Deterministic Relation with Uncertain *X* Values	160
	8.6	*X* and *Y* Are Random Variables	162
	8.7	Data Snooping	166
	8.8	Trends	170
	8.9	The Role of the Scatterplot	170
	8.10	The Problem of Regression Models	171
Nine	**Count-Based Data**		**175**
	9.1	Two Types of Counts	175
	9.2	Interval Estimates for Universe Proportions	177
	9.3	Homogeneity for Counts	180
	9.4	Should I Compare Counts or Rates?	182
	9.5	A Caution regarding Counts on an *XmR* Chart	183
	9.6	Outliers or Signals?	186
	9.7	Comparing Two Proportions with Charts	187
	9.8	Comparing Two Single Proportions	189
	9.9	Comparing Proportions for Several Conditions	192
	9.10	Summary	196
Ten	**Counts of Events**		**197**
	10.1	Inference for Counts of Events: One Condition	197
	10.2	Using the Poisson Model	201
	10.3	Comparing Two Conditions	202
	10.4	Comparing Several Conditions	205
	10.5	Summary	206
Eleven	**Counts for Three or More Categories**		**207**
	11.1	Categorical Frequencies for One Condition	207
	11.2	Categorical Data for Two or More Conditions	215
	11.3	Summary	218

PART THREE: THE KEYS TO EFFECTIVE DATA ANALYSIS 219

Twelve The Dual Nature of Trouble 221
12.1 "We Are in Trouble" 221
12.2 A New Definition of Trouble 223
12.3 The Four Possibilities for Any Process 225
12.4 Research and Experimentation Are Not Enough 235
12.5 Operating a Process Predictably 241
12.6 Summary 242

Thirteen Capability and Performance Indexes 243
13.1 The Voice(s) of the Process 243
13.2 The Voice of the Customer 245
13.3 The Capability Ratio, C_p 248
13.4 The Centered Capability Ratio, C_{pk} 250
13.5 The Performance Ratio, P_p 253
13.6 The Centered Performance Ratio, P_{pk} 256
13.7 Using the Four Ratios Together 257
13.8 The Ratios and the Four Possibilities 259
13.9 Operational Improvement 260
13.10 Short-Term and Long-Term Capability 262
13.11 Interval Estimates for Capability Indexes 262
13.12 Summary 267

Fourteen Using the Effective Cost of Production and Use 269
14.1 The Effective Cost of Production and Use 269
14.2 The Baseline Cost of Production and Use 271
14.3 The Centered Cost of Production and Use 272
14.4 The Predictable Cost of Production and Use 273
14.5 The Minimum Cost of Production and Use 275
14.6 Using the Effective Cost of Production and Use 276
14.7 Summary 280

Fifteen The Basis for the Effective Cost of Production and Use 281
15.1 The Structure of the *Effective Cost of Production and Use* 281
15.2 The Average Excess Costs 282
15.3 Finding the *Effective Cost of Production and Use* 288
15.4 Summary of *ECP&U* for Measurements 289
15.5 The Effective Cost of Production for Counts of Items 290
15.6 The Effective Cost of Production for Counts of Events 291

Sixteen	**The Six Sigma Zone**		**293**
	16.1	The *Effective Cost of Production and Use* Curves	293
	16.2	An Illustration	299
	16.3	The Six Sigma Zone	301
	16.4	Summary	302
Seventeen	**Some Problems**		**305**
	17.1	Problems with Defects per Million	305
	17.2	Problems with Defects per Million Opportunities	310
	17.3	Problems with FMEA Risk Priority Numbers	311
	17.4	Problems with Special Causes	316
	17.5	Problems with DMAIC Models	318
	17.6	Do We Need a Gauge R&R Study?	321
	17.7	Problems with Narrowly Defined Projects	323
	17.8	Summary	331
Eighteen	**Two Models for Process Improvement**		**333**
	18.1	Getting Started	333
	18.2	Characterize the Status of the Process Outcomes	334
	18.3	Outcomes in the State or Brink of Chaos	335
	18.4	Outcomes in the Threshold State	337
	18.5	Outcomes in the Ideal State	338
	18.6	Focusing Improvement on Strategic Objectives	338
	18.7	Summary	341
Nineteen	**An Honest Gauge R&R Study**		**343**
	19.1	The AIAG Gauge R&R Study	343
	19.2	The “Percentages” of the Total Variation	346
	19.3	The “Percentages” of the Specified Tolerance	350
	19.4	The “Number of Distinct Categories”	353
	19.5	The AIAG Guidelines	353
	19.6	What Can You Learn from the AIAG Study?	357
	19.7	An Honest Gauge R&R Study	358
	19.8	Summary	363

Appendix **367**

Table A.1 Bias Correction Factors for Measures of Dispersion 368-369
Table A.2 *XmR* Charts: Charts for Individual Values 370-371
Table A.3 Average and Range Charts Based on the Average Range 372-373
Table A.4 Average and Range Charts Based on the Median Range 374-375
Table A.5 Average and Standard Deviation Charts Based on the Average Standard Deviation Statistic 376-377
Table A.6 Average and Standard Deviation Charts Based on the Median Standard Deviation Statistic 378-379
Table A.7 Critical Values for Student's *t* Distributions 380-381
Table A.8 Percentiles of Chi-Square Distributions 382-383
Table A.9 Overall Alpha-Levels for Average Charts 384
Table A.10 Overall Alpha-Levels for Range Charts 385
Table A.11 Quick Scale Factors for Range-Based ANOM 386-388
Table A.12 Quick Scale Factors for Pooled Variance ANOM 389-391
Table A.13 Percentiles of the F-Distribution 392-397
Table A.14 Percentiles of the Studentized Range Distribution, *q* 398-403
Table A.15 Factors for 90% Interval Estimates of C_p and P_p 404
Table A.16 Factors for 90% Interval Estimates of C_{pk} and P_{pk} 405
Table A.17 The *Effective Cost of Production and Use* When *All* Nonconforming Units Are Scrapped 406-408
Table A.18 The *Effective Cost of Production and Use* When *All* Nonconforming Units Are Reworked 409-411
Table A.19 The *Effective Cost of Production and Use* When the Cost of Rework Is Half of the Cost of Scrap and the Average Is on the SCRAP Side of the Target 412-414
Table A.20 The *Effective Cost of Production and Use* When the Cost of Rework Is Half of the Cost of Scrap and the Average Is on the REWORK Side of the Target 415-417

Index **419**