

# Contents

<b>1</b>	<b>Why Statistical Confidentiality?</b>	1
1.1	What Is Statistical Confidentiality?	2
1.2	Stakeholders in the Statistical Process	3
1.3	The Data Stewardship Organization’s Dilemma	3
1.4	The Value of Statistical Data	6
1.5	Why Are DSOs Concerned About Statistical Confidentiality?	8
1.5.1	A Difficult Context for a DSO	8
1.5.2	Providing Data and Protecting Confidentiality	11
1.5.3	Consequences of a Confidentiality Breach	12
1.5.4	What Motivates a DSO to Provide Confidentiality?	13
1.6	High-Quality Statistical Data Raise Confidentiality Concerns	18
1.6.1	Characteristics of High-Quality Statistical Data	18
1.6.2	Disclosure Risk Problems Stemming from Characteristics of High-Quality Statistical Data	21
1.7	Disclosure Risk and the Concept of the Data Snooper	22
1.8	Strategies of Statistical Disclosure Limitation	23
1.8.1	Restricted Access	23
1.8.2	Restricted Data	24
1.9	Summary	24
<b>2</b>	<b>Concepts of Statistical Disclosure Limitation</b>	27
2.1	Conceptual Models of Disclosure Risk	27
2.1.1	Elements of the Disclosure Risk Problem	29
2.1.2	Perceived and Actual Risk	35
2.1.3	Scenarios of Disclosure	36
2.1.4	Data Environment Analysis	42
2.2	Assessing the Risk	42
2.2.1	Uniqueness	42
2.2.2	Matching/Reidentification Experiments	43
2.2.3	Disclosure Risk Assessment for Aggregate Data	43

2.3	Controlling the Risk . . . . .	44
2.3.1	Metadata Level Controls . . . . .	44
2.3.2	Distorting the Data . . . . .	45
2.3.3	Controlling Access . . . . .	45
2.4	Data Utility Impact . . . . .	46
2.5	Summary . . . . .	47
<b>3</b>	<b>Assessment of Disclosure Risk . . . . .</b>	<b>49</b>
3.1	Thresholds and Other Proxies . . . . .	50
3.2	Risk Assessment for Microdata: Types of Matching . . . . .	51
3.2.1	File-Level Risk Metrics . . . . .	51
3.2.2	Record-Level Risk Metrics . . . . .	54
3.3	Record Linkage Studies . . . . .	56
3.3.1	Using an External Data Set . . . . .	57
3.3.2	Using the Pre-SDL Data Set . . . . .	58
3.4	Risk Assessment for Count Data . . . . .	60
3.5	What is at Risk?: Understanding Sensitivity . . . . .	62
3.6	Summary . . . . .	63
<b>4</b>	<b>Protecting Tabular Data . . . . .</b>	<b>65</b>
4.1	Basic Concepts . . . . .	67
4.1.1	Structure of a Tabular Array . . . . .	67
4.1.2	Risky Cells . . . . .	70
4.1.3	The Secondary Problem: The Data Snooper's Knowledge . . . . .	71
4.1.4	Disclosure Limitation . . . . .	75
4.1.5	Loss of Information . . . . .	76
4.1.6	The DSO's Problem . . . . .	76
4.1.7	Disclosure Auditing . . . . .	77
4.2	Four Methods to Protect Tables . . . . .	77
4.2.1	Cell Suppression . . . . .	78
4.2.2	Interval Publication . . . . .	81
4.2.3	Controlled Rounding . . . . .	82
4.2.4	Cell Perturbation . . . . .	85
4.2.5	All-in-One Method . . . . .	86
4.3	Other Methods . . . . .	86
4.3.1	Table Redesign . . . . .	87
4.3.2	Introducing Noise to Microdata . . . . .	87
4.3.3	Data Swapping . . . . .	88
4.3.4	Cyclic Perturbation . . . . .	88
4.3.5	Random Rounding . . . . .	89
4.3.6	Controlled Tabular Adjustment . . . . .	90
4.4	Summary . . . . .	92
<b>5</b>	<b>Providing and Protecting Microdata . . . . .</b>	<b>93</b>
5.1	Why Provide Access? . . . . .	95
5.2	Confidentiality Concerns . . . . .	99

- 5.3 Why Protect Microdata? . . . . . 103
- 5.4 Restricted Data . . . . . 105
  - 5.4.1 In Order to Limit Disclosure, What Shall We Mask? . . . . . 108
- 5.5 Matrix Masking . . . . . 109
- 5.6 Masking Through Suppression . . . . . 110
- 5.7 Local Suppression . . . . . 112
- 5.8 Noise Addition . . . . . 112
- 5.9 Data Swapping . . . . . 114
  - 5.9.1 Implementations of Data Swapping . . . . . 115
  - 5.9.2 A Protocol for Data Swapping . . . . . 116
- 5.10 Masking Through Sampling . . . . . 118
- 5.11 Masking Through Aggregation . . . . . 119
  - 5.11.1 Global Recoding . . . . . 119
  - 5.11.2 Topcoding . . . . . 120
- 5.12 Microaggregation . . . . . 120
- 5.13 Synthetic Microdata . . . . . 120
- 5.14 Concluding Thoughts . . . . . 122
- 6 Disclosure Risk and Data Utility . . . . . 123**
  - 6.1 Basics of Disclosure Risk and Data Utility . . . . . 123
    - 6.1.1 Choosing the Parameter Values of an SDL Method . . . . . 124
  - 6.2 Data Utility Metrics . . . . . 125
  - 6.3 Direct Measurement of Utility . . . . . 126
  - 6.4 The R-U Confidentiality Map . . . . . 127
    - 6.4.1 Constructing an R-U Confidentiality Map: Multivariate Additive Noise . . . . . 129
    - 6.4.2 R-U Confidentiality Map for Topcoding . . . . . 131
  - 6.5 Discussion . . . . . 134
- 7 Restrictions on Data Access . . . . . 137**
  - 7.1 Who Can Have Access? . . . . . 138
  - 7.2 Where Can Access Be Obtained? . . . . . 139
  - 7.3 What Analysis Is Permitted? . . . . . 140
  - 7.4 Modes of Access . . . . . 141
    - 7.4.1 Free Access . . . . . 141
    - 7.4.2 Delivered Access . . . . . 141
    - 7.4.3 Safe Settings . . . . . 142
    - 7.4.4 Virtual Access . . . . . 142
    - 7.4.5 Licensing . . . . . 143
  - 7.5 Conclusion . . . . . 145
- 8 Thoughts on the Future . . . . . 147**
  - 8.1 New Meanings for Privacy and Statistical Confidentiality . . . . . 149
  - 8.2 Who Will Care About Statistical Data? . . . . . 151
  - 8.3 What New Forms of Data Stewardship Organizations Will Develop? . . . . . 152

- 8.4 Will Statistical Data Remain Valuable? . . . . . 154
- 8.5 New Data Types . . . . . 155
  - 8.5.1 Geospatial Data . . . . . 155
  - 8.5.2 Audio and Video Data . . . . . 156
  - 8.5.3 Biometric Recognition Data . . . . . 156
  - 8.5.4 Biological Material Data . . . . . 157
  - 8.5.5 Network Data . . . . . 158
- 8.6 Privacy Preserving Data Mining . . . . . 159
- 8.7 Other New Issues for Statistical Confidentiality . . . . . 160
  - 8.7.1 Technological Advances . . . . . 160
  - 8.7.2 Increased Expectations About Data Access . . . . . 161
  - 8.7.3 Sophisticated Privacy Advocates . . . . . 162
  - 8.7.4 New Confidentiality Legislation . . . . . 162
  - 8.7.5 Demand for Data from Researchers . . . . . 162
  - 8.7.6 Challenges in Communicating Confidentiality  
Protections . . . . . 163
- 8.8 Will There Be New Forms of Data Snooping? . . . . . 164
  - 8.8.1 The Data Snooper of the Future . . . . . 164
  - 8.8.2 New Attack Modalities . . . . . 165
- 8.9 What New Strategies of Disclosure Limitation Should  
Be Developed? . . . . . 167
- 8.10 Finally, an Exciting Vision for Statistical Confidentiality . . . . . 168
- Glossary** . . . . . 171
- References** . . . . . 181
- Index** . . . . . 195



<http://www.springer.com/978-1-4419-7801-1>

Statistical Confidentiality

Principles and Practice

Duncan, G.T.; Elliot, M.; Juan Jose Salazar, G.

2011, XII, 200 p., Hardcover

ISBN: 978-1-4419-7801-1