
Contents

Foreword	xix
Preface.....	xxi
Acknowledgments	xxiii
Author	xxv
1. Introduction	1
1.1 What Is Empirical Software Engineering?.....	1
1.2 Overview of Empirical Studies	2
1.3 Types of Empirical Studies	3
1.3.1 Experiment.....	4
1.3.2 Case Study	5
1.3.3 Survey Research.....	6
1.3.4 Systematic Reviews.....	7
1.3.5 Postmortem Analysis	8
1.4 Empirical Study Process	8
1.4.1 Study Definition.....	9
1.4.2 Experiment Design.....	10
1.4.3 Research Conduct and Analysis.....	11
1.4.4 Results Interpretation.....	12
1.4.5 Reporting	12
1.4.6 Characteristics of a Good Empirical Study	12
1.5 Ethics of Empirical Research.....	13
1.5.1 Informed Content	14
1.5.2 Scientific Value	15
1.5.3 Confidentiality	15
1.5.4 Beneficence.....	15
1.5.5 Ethics and Open Source Software.....	15
1.5.6 Concluding Remarks.....	15
1.6 Importance of Empirical Research	16
1.6.1 Software Industry.....	16
1.6.2 Academicians	16
1.6.3 Researchers.....	17
1.7 Basic Elements of Empirical Research	17
1.8 Some Terminologies	18
1.8.1 Software Quality and Software Evolution.....	18
1.8.2 Software Quality Attributes.....	20
1.8.3 Measures, Measurements, and Metrics	20
1.8.4 Descriptive, Correlational, and Cause-Effect Research.....	22
1.8.5 Classification and Prediction	22
1.8.6 Quantitative and Qualitative Data	22
1.8.7 Independent, Dependent, and Confounding Variables	23
1.8.8 Proprietary, Open Source, and University Software	24
1.8.9 Within-Company and Cross-Company Analysis.....	25

1.8.10	Parametric and Nonparametric Tests.....	26
1.8.11	Goal/Question/Metric Method.....	26
1.8.12	Software Archive or Repositories	28
1.9	Concluding Remarks.....	28
	Exercises	28
	Further Readings.....	29
2.	Systematic Literature Reviews.....	33
2.1	Basic Concepts.....	33
2.1.1	Survey versus SRs	33
2.1.2	Characteristics of SRs.....	34
2.1.3	Importance of SRs	35
2.1.4	Stages of SRs.....	35
2.2	Case Study.....	36
2.3	Planning the Review	37
2.3.1	Identify the Need for SR.....	37
2.3.2	Formation of Research Questions	38
2.3.3	Develop Review Protocol	39
2.3.4	Evaluate Review Protocol.....	45
2.4	Methods for Presenting Results.....	46
2.4.1	Tools and Techniques.....	46
2.4.2	Forest Plots	49
2.4.3	Publication Bias.....	51
2.5	Conducting the Review.....	51
2.5.1	Search Strategy Execution.....	51
2.5.2	Selection of Primary Studies.....	53
2.5.3	Study Quality Assessment.....	53
2.5.4	Data Extraction	53
2.5.5	Data Synthesis.....	53
2.6	Reporting the Review	56
2.7	SRs in Software Engineering.....	58
	Exercises	60
	Further Readings.....	61
3.	Software Metrics.....	65
3.1	Introduction.....	65
3.1.1	What Are Software Metrics?.....	66
3.1.2	Application Areas of Metrics.....	66
3.1.3	Characteristics of Software Metrics.....	67
3.2	Measurement Basics	67
3.2.1	Product and Process Metrics	68
3.2.2	Measurement Scale.....	69
3.3	Measuring Size.....	71
3.4	Measuring Software Quality.....	72
3.4.1	Software Quality Metrics Based on Defects.....	72
3.4.1.1	Defect Density	72
3.4.1.2	Phase-Based Defect Density.....	73
3.4.1.3	Defect Removal Effectiveness	73

3.4.2	Usability Metrics	74
3.4.3	Testing Metrics	75
3.5	OO Metrics	76
3.5.1	Popular OO Metric Suites	77
3.5.2	Coupling Metrics	79
3.5.3	Cohesion Metrics.....	83
3.5.4	Inheritance Metrics.....	85
3.5.5	Reuse Metrics	86
3.5.6	Size Metrics.....	88
3.6	Dynamic Software Metrics.....	89
3.6.1	Dynamic Coupling Metrics.....	89
3.6.2	Dynamic Cohesion Metrics	89
3.6.3	Dynamic Complexity Metrics.....	90
3.7	System Evolution and Evolutionary Metrics.....	90
3.7.1	Revisions, Refactorings, and Bug-Fixes	91
3.7.2	LOC Based	91
3.7.3	Code Churn Based.....	91
3.7.4	Miscellaneous.....	92
3.8	Validation of Metrics	92
3.9	Practical Relevance	93
3.9.1	Designing a Good Quality System.....	93
3.9.2	Which Software Metrics to Select?	94
3.9.3	Computing Thresholds	95
3.9.3.1	Statistical Model to Compute Threshold	96
3.9.3.2	Use of ROC Curve to Calculate the Threshold Values	98
3.9.4	Practical Relevance and Use of Software Metrics in Research	99
3.9.5	Industrial Relevance of Software Metrics	100
	Exercises	100
	Further Readings.....	101
4.	Experimental Design	103
4.1	Overview of Experimental Design.....	103
4.2	Case Study: Fault Prediction Systems	103
4.2.1	Objective of the Study	104
4.2.2	Motivation.....	104
4.2.3	Study Context.....	105
4.2.4	Results	105
4.3	Research Questions.....	106
4.3.1	How to Form RQ?	106
4.3.2	Characteristics of an RQ	107
4.3.3	Example: RQs Related to FPS.....	108
4.4	Reviewing the Literature	109
4.4.1	What Is a Literature Review?	109
4.4.2	Steps in a Literature Review.....	110
4.4.3	Guidelines for Writing a Literature Review	111
4.4.4	Example: Literature Review in FPS.....	112
4.5	Research Variables	117
4.5.1	Independent and Dependent Variables.....	117

4.5.2	Selection of Variables.....	118
4.5.3	Variables Used in Software Engineering	118
4.5.4	Example: Variables Used in the FPS.....	118
4.6	Terminology Used in Study Types	118
4.7	Hypothesis Formulation	120
4.7.1	Experiment Design Types.....	120
4.7.2	What Is Hypothesis?.....	121
4.7.3	Purpose and Importance of Hypotheses in an Empirical Research	121
4.7.4	How to Form a Hypothesis?.....	122
4.7.5	Steps in Hypothesis Testing.....	124
4.7.5.1	Step 1: State the Null and Alternative Hypothesis.....	124
4.7.5.2	Step 2: Choose the Test of Significance	126
4.7.5.3	Step 3: Compute the Test Statistic and Associated p -Value	126
4.7.5.4	Step 4: Define Significance Level	127
4.7.5.5	Step 5: Derive Conclusions.....	127
4.7.6	Example: Hypothesis Formulation in FPS	129
4.7.6.1	Hypothesis Set A.....	130
4.7.6.2	Hypothesis Set B.....	130
4.8	Data Collection	131
4.8.1	Data-Collection Strategies	131
4.8.2	Data Collection from Repositories	132
4.8.3	Example: Data Collection in FPS	134
4.9	Selection of Data Analysis Methods	136
4.9.1	Type of Dependent Variable.....	137
4.9.2	Nature of the Data Set.....	137
4.9.3	Aspects of Data Analysis Methods	138
	Exercises	139
	Further Readings.....	140
5.	Mining Data from Software Repositories.....	143
5.1	Configuration Management Systems.....	143
5.1.1	Configuration Identification.....	144
5.1.2	Configuration Control.....	144
5.1.3	Configuration Accounting.....	146
5.2	Importance of Mining Software Repositories	147
5.3	Common Types of Software Repositories	147
5.3.1	Historical Repositories	148
5.3.2	Run-Time Repositories or Deployment Logs.....	149
5.3.3	Source Code Repositories	150
5.4	Understanding Systems	150
5.4.1	System Characteristics	150
5.4.2	System Evolution.....	151
5.5	Version Control Systems	151
5.5.1	Introduction.....	151
5.5.2	Classification of VCS.....	153
5.5.2.1	Local VCS	153
5.5.2.2	Centralized VCS	153
5.5.2.3	Distributed VCS.....	154
5.6	Bug Tracking Systems.....	155

5.7	Extracting Data from Software Repositories	156
5.7.1	CVS	157
5.7.2	SVN.....	159
5.7.3	Git.....	162
5.7.4	Bugzilla	166
5.7.5	Integrating Bugzilla with Other VCS	169
5.8	Static Source Code Analysis	170
5.8.1	Level of Detail	171
5.8.1.1	Method Level.....	171
5.8.1.2	Class Level	171
5.8.1.3	File Level	171
5.8.1.4	System Level	172
5.8.2	Metrics.....	172
5.8.3	Software Metrics Calculation Tools.....	173
5.9	Software Historical Analysis.....	175
5.9.1	Understanding Dependencies in a System.....	176
5.9.2	Change Impact Analysis	177
5.9.3	Change Propagation.....	177
5.9.4	Defect Proneness	178
5.9.5	User and Team Dynamics Understanding	178
5.9.6	Change Prediction.....	178
5.9.7	Mining Textual Descriptions.....	179
5.9.8	Social Network Analysis	179
5.9.9	Change Smells and Refactoring	180
5.9.10	Effort Estimation	180
5.10	Software Engineering Repositories and Open Research Data Sets.....	180
5.10.1	FLOSSmole	180
5.10.2	FLOSSMetrics.....	181
5.10.3	PRedictOr Models In Software Engineering.....	182
5.10.4	Qualitas Corpus.....	182
5.10.5	Sourcerer Project.....	182
5.10.6	Ultimate Debian Database	183
5.10.7	Bug Prediction Data Set.....	183
5.10.8	International Software Benchmarking Standards Group	184
5.10.9	Eclipse Bug Data	184
5.10.10	Software-Artifact Infrastructure Repository.....	184
5.10.11	Ohloh.....	185
5.10.12	SourceForge Research Data Archive	185
5.10.13	Helix Data Set	186
5.10.14	Tukutuku	186
5.10.15	Source Code ECO System Linked Data.....	186
5.11	Case Study: Defect Collection and Reporting System for Git Repository	187
5.11.1	Introduction.....	187
5.11.2	Motivation	188
5.11.3	Working Mechanism.....	188
5.11.4	Data Source and Dependencies.....	190
5.11.5	Defect Reports.....	191
5.11.5.1	Defect Details Report	191
5.11.5.2	Defect Count and Metrics Report.....	193

	5.11.5.3	LOC Changes Report	194
	5.11.5.4	Newly Added Source Files	195
	5.11.5.5	Deleted Source Files	196
	5.11.5.6	Consolidated Defect and Change Report.....	197
	5.11.5.7	Descriptive Statistics Report for the Incorporated Metrics	198
5.11.6		Additional Features.....	199
	5.11.6.1	Cloning of Git-Based Software Repositories.....	199
	5.11.6.2	Self-Logging.....	201
5.11.7		Potential Applications of DCRS	202
	5.11.7.1	Defect Prediction Studies	202
	5.11.7.2	Change-Proneness Studies	203
	5.11.7.3	Statistical Comparison	203
5.11.8		Concluding Remarks	203
		Exercises	203
		Further Readings.....	204
6.		Data Analysis and Statistical Testing	207
6.1		Analyzing the Metric Data	207
	6.1.1	Measures of Central Tendency	207
		6.1.1.1 Mean	207
		6.1.1.2 Median.....	208
		6.1.1.3 Mode	209
		6.1.1.4 Choice of Measures of Central Tendency.....	209
	6.1.2	Measures of Dispersion.....	211
	6.1.3	Data Distributions	212
	6.1.4	Histogram Analysis	213
	6.1.5	Outlier Analysis.....	213
		6.1.5.1 Box Plots	214
		6.1.5.2 Z-Score.....	216
	6.1.6	Correlation Analysis	218
	6.1.7	Example—Descriptive Statistics of Fault Prediction System	218
6.2		Attribute Reduction Methods	219
	6.2.1	Attribute Selection.....	221
		6.2.1.1 Univariate Analysis	222
		6.2.1.2 Correlation-Based Feature Selection.....	222
	6.2.2	Attribute Extraction	222
		6.2.2.1 Principal Component Method	222
	6.2.3	Discussion.....	223
6.3		Hypothesis Testing	223
	6.3.1	Introduction.....	224
	6.3.2	Steps in Hypothesis Testing.....	224
6.4		Statistical Testing	225
	6.4.1	Overview of Statistical Tests.....	225
	6.4.2	Categories of Statistical Tests.....	225
	6.4.3	One-Tailed and Two-Tailed Tests	226
	6.4.4	Type I and Type II Errors	228
	6.4.5	Interpreting Significance Results	229

6.4.6	<i>t</i> -Test.....	229
	6.4.6.1 One Sample <i>t</i> -Test.....	229
	6.4.6.2 Two Sample <i>t</i> -Test.....	232
	6.4.6.3 Paired <i>t</i> -Test.....	233
6.4.7	Chi-Squared Test.....	235
6.4.8	<i>F</i> -Test.....	242
6.4.9	Analysis of Variance Test.....	244
	6.4.9.1 One-Way ANOVA.....	244
6.4.10	Wilcoxon Signed Test.....	247
6.4.11	Wilcoxon–Mann–Whitney Test (<i>U</i> -Test).....	250
6.4.12	Kruskal–Wallis Test.....	254
6.4.13	Friedman Test.....	257
6.4.14	Nemenyi Test.....	259
6.4.15	Bonferroni–Dunn Test.....	261
6.4.16	Univariate Analysis.....	263
6.5	Example—Univariate Analysis Results for Fault Prediction System.....	263
	Exercises.....	265
	Further Readings.....	271
7.	Model Development and Interpretation.....	275
7.1	Model Development.....	275
	7.1.1 Data Partition.....	276
	7.1.2 Attribute Reduction.....	277
	7.1.3 Model Construction using Learning Algorithms/Techniques.....	277
	7.1.4 Validating the Model Predicted.....	277
	7.1.5 Hypothesis Testing.....	278
	7.1.6 Interpretation of Results.....	278
	7.1.7 Example—Software Quality Prediction System.....	278
7.2	Statistical Multiple Regression Techniques.....	280
	7.2.1 Multivariate Analysis.....	280
	7.2.2 Coefficients and Selection of Variables.....	280
7.3	Machine Learning Techniques.....	281
	7.3.1 Categories of ML Techniques.....	281
	7.3.2 Decision Trees.....	282
	7.3.3 Bayesian Learners.....	282
	7.3.4 Ensemble Learners.....	282
	7.3.5 Neural Networks.....	283
	7.3.6 Support Vector Machines.....	285
	7.3.7 Rule-Based Learning.....	286
	7.3.8 Search-Based Techniques.....	287
7.4	Concerns in Model Prediction.....	290
	7.4.1 Problems with Model Prediction.....	290
	7.4.2 Multicollinearity Analysis.....	290
	7.4.3 Guidelines for Selecting Learning Techniques.....	291
	7.4.4 Dealing with Imbalanced Data.....	291
	7.4.5 Parameter Tuning.....	292
7.5	Performance Measures for Categorical Dependent Variable.....	292
	7.5.1 Confusion Matrix.....	292
	7.5.2 Sensitivity and Specificity.....	294

7.5.3	Accuracy and Precision.....	295
7.5.4	Kappa Coefficient.....	295
7.5.5	F-measure, G-measure, and G-mean.....	295
7.5.6	Receiver Operating Characteristics Analysis	298
	7.5.6.1 ROC Curve	298
	7.5.6.2 Area Under the ROC Curve.....	300
	7.5.6.3 Cutoff Point and Co-Ordinates of the ROC Curve.....	300
7.5.7	Guidelines for Using Performance Measures.....	302
7.6	Performance Measures for Continuous Dependent Variable.....	304
7.6.1	Mean Relative Error.....	304
7.6.2	Mean Absolute Relative Error	304
7.6.3	PRED (A)	305
7.7	Cross-Validation	306
7.7.1	Hold-Out Validation.....	307
7.7.2	K-Fold Cross-Validation	307
7.7.3	Leave-One-Out Validation	307
7.8	Model Comparison Tests	309
7.9	Interpreting the Results	310
7.9.1	Analyzing Performance Measures.....	310
7.9.2	Presenting Qualitative and Quantitative Results	312
7.9.3	Drawing Conclusions from Hypothesis Testing.....	312
7.9.4	Example—Discussion of Results in Hypothesis Testing Using Univariate Analysis for Fault Prediction System	312
7.10	Example—Comparing ML Techniques for Fault Prediction.....	315
	Exercises	323
	Further Readings.....	324
8.	Validity Threats	331
8.1	Categories of Threats to Validity	331
8.1.1	Conclusion Validity	331
8.1.2	Internal Validity	333
8.1.3	Construct Validity.....	334
8.1.4	External Validity	335
8.1.5	Essential Validity Threats.....	337
8.2	Example—Threats to Validity in Fault Prediction System.....	337
8.2.1	Conclusion Validity	337
8.2.2	Internal Validity	339
8.2.3	Construct Validity.....	339
8.2.4	External Validity	340
8.3	Threats and Their Countermeasures	341
	Exercises	350
	Further Readings.....	350
9.	Reporting Results.....	353
9.1	Reporting and Presenting Results.....	353
9.1.1	When to Disseminate or Report Results?.....	354
9.1.2	Where to Disseminate or Report Results?.....	354

- 9.1.3 Report Structure..... 355
 - 9.1.3.1 Abstract..... 356
 - 9.1.3.2 Introduction..... 357
 - 9.1.3.3 Related Work..... 357
 - 9.1.3.4 Experimental Design 357
 - 9.1.3.5 Research Methods 358
 - 9.1.3.6 Research Results..... 358
 - 9.1.3.7 Discussion and Interpretation..... 359
 - 9.1.3.8 Threats to Validity..... 359
 - 9.1.3.9 Conclusions 359
 - 9.1.3.10 Acknowledgment 359
 - 9.1.3.11 References..... 359
 - 9.1.3.12 Appendix..... 359
- 9.2 Guidelines for Masters and Doctoral Students 359
- 9.3 Research Ethics and Misconduct..... 361
 - 9.3.1 Plagiarism 362
- Exercises 362
- Further Readings..... 363
- 10. Mining Unstructured Data..... 365**
 - 10.1 Introduction..... 365
 - 10.1.1 What Is Unstructured Data?..... 366
 - 10.1.2 Multiple Classifications..... 366
 - 10.1.3 Importance of Text Mining..... 367
 - 10.1.4 Characteristics of Text Mining..... 367
 - 10.2 Steps in Text Mining 368
 - 10.2.1 Representation of Text Documents..... 368
 - 10.2.2 Preprocessing Techniques 368
 - 10.2.2.1 Tokenization..... 370
 - 10.2.2.2 Removal of Stop Words 370
 - 10.2.2.3 Stemming Algorithm..... 371
 - 10.2.3 Feature Selection 372
 - 10.2.4 Constructing a Vector Space Model 375
 - 10.2.5 Predicting and Validating the Text Classifier 377
 - 10.3 Applications of Text Mining in Software Engineering..... 378
 - 10.3.1 Mining the Fault Reports to Predict the Severity of the Faults..... 378
 - 10.3.2 Mining the Change Logs to Predict the Effort 378
 - 10.3.3 Analyzing the SRS Document to Classify Requirements into NFRs 378
 - 10.4 Example—Automated Severity Assessment of Software Defect Reports..... 379
 - 10.4.1 Introduction 379
 - 10.4.2 Data Source 380
 - 10.4.3 Experimental Design..... 380
 - 10.4.4 Result Analysis..... 382
 - 10.4.5 Discussion of Results..... 385

10.4.6	Threats to Validity	387
10.4.7	Conclusion.....	387
Exercises		387
Further Readings.....		388
11. Demonstrating Empirical Procedures		391
11.1	Abstract	391
11.1.1	Basic.....	391
11.1.2	Method.....	392
11.1.3	Results.....	392
11.2	Introduction.....	392
11.2.1	Motivation	392
11.2.2	Objectives	392
11.2.3	Method.....	393
11.2.4	Technique Selection	393
11.2.5	Subject Selection.....	394
11.3	Related Work	394
11.4	Experimental Design.....	395
11.4.1	Problem Definition.....	395
11.4.2	Research Questions.....	395
11.4.3	Variables Selection	396
11.4.4	Hypothesis Formulation	397
11.4.4.1	Hypothesis Set	397
11.4.5	Statistical Tests.....	398
11.4.6	Empirical Data Collection.....	399
11.4.7	Technique Selection	400
11.4.8	Analysis Process.....	401
11.5	Research Methodology	401
11.5.1	Description of Techniques	401
11.5.2	Performance Measures and Validation Method.....	404
11.6	Analysis Results.....	404
11.6.1	Descriptive Statistics.....	404
11.6.2	Outlier Analysis	408
11.6.3	CFS Results.....	408
11.6.4	Tenfold Cross-Validation Results.....	408
11.6.5	Hypothesis Testing and Evaluation	416
11.6.6	Nemenyi Results	419
11.7	Discussion and Interpretation of Results.....	420
11.8	Validity Evaluation	423
11.8.1	Conclusion Validity	423
11.8.2	Internal Validity	423
11.8.3	Construct Validity.....	423
11.8.4	External Validity	423
11.9	Conclusions and Future Work	424
Appendix.....		425

12. Tools for Analyzing Data	429
12.1 WEKA.....	429
12.2 KEEL.....	429
12.3 SPSS.....	430
12.4 MATLAB [®]	430
12.5 R.....	431
12.6 Comparison of Tools.....	431
Further Readings.....	434
 Appendix: Statistical Tables	 437
References	445
Index	459