

Contents

<i>List of Figures</i>	xiii
<i>List of Tables</i>	xv

Part I. The Limitations of Descriptive Methodology

1. Why Social Sciences Are Not Scientific Enough	3
2. Can Social Science Approaches Find the Law of Gravitation?	14
3. How to Construct Predictive Models: Simplicity and Nonabsurdity	23
4. Example of Model Building: Electoral Volatility	34
5. Physicists Multiply, Social Scientists Add—Even When It Does Not Add Up	52
6. All Hypotheses Are Not Created Equal	71
7. Why Most Numbers Published in Social Sciences Are Dead on Arrival	82

Part II. Quantitatively Predictive Logical Models

8. Forbidden Areas and Anchor Points	95
9. Geometric Means and Lognormal Distributions	120
10. Example of Interlocking Models: Party Sizes and Cabinet Duration	130
11. Beyond Constraint-Based Models: Communication Channels and Growth Rates	139
12. Why We Should Shift to Symmetric Regression	154
13. All Indices Are Not Created Equal	176

Part III. Synthesis of Predictive and Descriptive Approaches

14. From Descriptive to Predictive Approaches	187
15. Recommendations for Better Regression	199
16. Converting from Descriptive Analysis to Predictive Models	215
17. Are Electoral Studies a Rosetta Stone for Parts of Social Sciences?	225
18. Beyond Regression: The Need for Predictive Models	236
<i>References</i>	241
<i>Index</i>	249

List of Figures

3.1. Best linear fits to different patterns that all satisfy the directional prediction $dy/dx < 0$	26
3.2. Curves for quantitative model $y = k/x$, and an unsatisfactory data set	27
4.1. Individual-level volatility of votes vs. effective number of electoral parties—conceptually forbidden areas, anchor point, and expected zone	37
4.2. Individual-level volatility of votes vs. effective number of electoral parties—data and best linear fit from Heath (2005), plus coarse and refined predictive models	42
4.3. Individual-level volatility of votes vs. effective number of electoral parties—truncated data (from Heath 2005) and two models that fit the data and the anchor point	46
5.1. Typical ways variables interact in physics and in today's social science	57
5.2. A sequence for introducing new variables in electricity	67
6.1. The hierarchy of hypotheses	73
8.1. Fixed exponent functions—the simplest full family of curves allowed when x and y are conceptually restricted to positive values	98
8.2. Exponential functions—the simplest full family of curves allowed when y is conceptually restricted to positive values while x is not	102
8.3. The simplest full family of curves allowed when x and y are conceptually restricted to the range from 0 to 1, with three anchor points	108
10.1. The main causal sequence leading from population, assembly size, and district magnitude to mean duration of cabinets	131
12.1. The OLS regression line underreports the expected slope, whichever way you graph the two variables	156
12.2. The same data-set yields two distinct regression lines— y vs. x and x vs. y	170

List of Figures

12.3. The two OLS regression lines under- and overreport, respectively, the expected slope	172
14.1. Logical sequence and “gas station approach” for mean duration of cabinets	191
14.2. The opposite meanings of “theory”	193
14.3. Cycles and sub-cycles in scientific procedure	196
15.1. Graphing the data from Table 15.1 checks on whether linear regression makes sense	201
15.2. Proportionality profiles for elections in New Zealand and the United States: data and conceptual range	203
15.3. Will outliers follow the regression slope?	212
17.1. Direct interconnections of scientific disciplines	227

List of Tables

1.1. Predictive vs. descriptive models	7
2.1. Synthetic data where $y = 980x_1x_3/x_2^2$	18
4.1. How does the number of parties (N) affect volatility (V)?—predictive and descriptive approaches	44
5.1. The 20 equations voted the most important for physics (Crease 2004), by rank	53
5.2. Typical mathematical formats in physics and in today's social sciences	68
6.1. Predictive vs. descriptive models	80
7.1. Thinking patterns during the course of solving an intellectual problem	83
7.2. Total government expenditure in percent of GDP: How can it be predicted?	85
7.3. The number of constants/coefficients vs. their impact	88
8.1. Simplest formats resulting from conceptual constraints on ranges of occurrence of input and output variables	110
9.1. The relationships of arithmetic mean (x_A), median (x_M), and geometric mean (x_G) as the ratio of largest to smallest entry widens	123
10.1. Logical connections (and R^2 of logarithms) between characteristics of party systems	133
13.1. Degree of agreement with predictive models of mean cabinet duration for standard OLS and symmetric regressions of logarithms	179
15.1. Four data-sets that lead to the same linear fit and R^2 when linear regression is (mis)applied	200
15.2. A typical table of regression results	207
15.3. A typical table of regression results, with suggested complements	209
16.1. Approximate values of constants in predictive model for vote loss by incumbent's party, calculated from regression coefficients in Samuels (2004)	220