

# Contents

<i>Preface</i>	v
<i>About the Author</i>	vii
<b>1. Input Manipulation</b>	<b>1</b>
1.1 Manipulating Input Data . . . . .	1
1.1.1 Data Snooping . . . . .	1
1.1.2 Biased Sampling . . . . .	3
1.1.3 Eliminating Outliers . . . . .	5
1.1.4 Just Use False Inputs: The Hwang Woo-suk Cloning Scandal . . . . .	8
1.1.5 Data Transformation: Is Standardization or Transformation to Percentages Statistical Cheating? . . . . .	9
1.2 Manipulating the Research Timeframe . . . . .	16
1.2.1 Timeframe Snooping . . . . .	16
1.2.2 Lengthening and Shortening the Time Frame	17
Conclusion: How to Catch Input Manipulators — Piece of Cake . . . . .	20
Questions and Problems . . . . .	21
Appendix A.1: The Complete List Why Mathematicians Look Down on Statisticians . . . . .	22
Appendix A.2: Data Transformation in the Copula Model	23

Appendix A.3: The Nitty-Gritty When Calculating Standard Deviation and Volatility . . . . .	26
References . . . . .	30
<b>2. Manipulating Statistical Calculations</b>	<b>31</b>
2.1 Regression Analysis . . . . .	31
2.1.1 Applying the Linear Pearson Model to Nonlinear Data . . . . .	39
2.1.2 Polynomial Regression: Under- and Overfitting . . . . .	44
2.1.3 Correlation Does Not Imply Causation! . . . .	48
2.1.4 Kitchen Sink Regression . . . . .	54
2.1.5 Multicollinearity Can Distort Individual Regressors . . . . .	60
2.1.6 Heteroskedasticity to Manipulate Standard Errors . . . . .	66
Summary and Conclusion: How to Catch Regression Cheaters — Pretty Easy . . . . .	70
2.2 Inferential Statistics . . . . .	72
2.2.1 Increasing the Sample Size $n$ to Increase Significance . . . . .	77
2.2.2 Manipulating the Population Mean $\mu$ . . . . .	78
2.2.3 Manipulating the Population Standard Deviation Parameter, the “Nuisance Parameter” $\sigma$ . . . . .	80
2.2.4 Changing a Two-tailed Test into a One-tailed Test . . . . .	80
Summary and Conclusion: How to Catch Inference Cheaters — Piece of Cake? . . . . .	83
Questions and Problems . . . . .	85
Appendix A.1: Nasty Math Calculations for the Standard Deviation, Covariance, and the Correlation Coefficient . . . . .	86
Appendix A.2: On the Slope of the Regression function	87

Appendix A.3: Derivation of $SST = SSE + SSR$ . . . . .	89
Appendix A.4: Derivation of the $z$ -Score . . . . .	91
References . . . . .	92
<b>3. “Creative Reporting”, i.e., Distorting Outputs</b>	<b>93</b>
3.1 Numerical Bias . . . . .	93
3.1.1 Profits from Nothing: Adding Percentage Changes . . . . .	93
3.1.2 Reporting Bias in Academia . . . . .	96
3.1.3 Reporting Bias in Practice . . . . .	102
3.2 Visual Bias . . . . .	108
3.2.1 Image Distortion . . . . .	108
3.2.2 Truncating the $Y$ -Axis . . . . .	110
3.2.3 Different Axes Units . . . . .	112
3.2.4 Logarithmic Display . . . . .	114
3.2.5 Omitting Data . . . . .	116
3.2.6 2D Versus 3D Data Representation . . . . .	118
Summary: How to Identify Creative Reporters — No Problem . . . . .	120
Questions and Problems . . . . .	122

### 1.1.1. Data Snooping

There are researchers with a Machiavellian attitude. The objective of publication justifies the means, statistical cheating. One form of statistical cheating is data snooping.

What is Data Snooping, also called Data Fishing, Data Dredging or Data Enticery?

Data Snooping is selecting Data for which the Research Hypothesis works.

In clinical statistical research, first a research hypothesis is formulated. Then the hypothesis is tested with real-world data. Performing a data analysis first and selecting specific data for which the hypothesis works is Data Snooping and constitutes statistical cheating. Here are some examples of Data Snooping.