

Contents

1	Data Collection	1
1.1	Sample Surveys	2
1.2	Fitness Club: Example of a Sample Survey	4
1.3	Experiments	4
1.4	Experiments: An Example	5
1.5	Data Collection	6
1.6	Registers	6
1.7	Questionnaire Surveys	7
1.7.1	Background Questions	7
1.7.2	Study Questions	7
1.8	Sources of Errors in Surveys	10
1.9	Comparing Methods of Data Collection	11
1.10	Example Continued	13
2	Presentation of Data	15
2.1	Bar Charts	15
2.2	Histograms	17
2.3	Pie Charts	19
2.4	Scatter Plots	20
2.5	Line Charts	21
2.6	Bubble Plots	22
2.7	Tables	22
2.7.1	The Ingredients of a Table	23
2.7.2	Percentages	24
3	Description of Data	27
3.1	Systematic and Random Variation	27
3.2	Measures of Location	29
3.2.1	Average	29
3.2.2	Median	31

3.2.3	Mode	32
3.2.4	Choosing a Measure of Location	33
3.3	Measures of Dispersion	35
3.3.1	Range	35
3.3.2	Variance and Standard Deviation	36
3.3.3	Interquartile Range	38
3.3.4	Choosing a Measure of Dispersion	40
3.3.5	Relative Spread (Dispersion)	40
3.4	Example: Statistical Functions in Spreadsheets	41
3.5	Data Type and Descriptive Statistics	43
3.5.1	Data Types	44
3.5.2	Descriptive Statistics and Type of Data	44
4	The Normal Distribution	47
4.1	Characteristics of the Normal Distribution	47
4.2	Density Function and Distribution Function	49
4.3	Fractiles	50
4.4	Calculations in the Normal Distribution	51
4.5	The Normal Distribution and Spreadsheets	52
4.5.1	NORMDIST (X; Mean; Stdev; Cumulative)	53
4.5.2	NORMINV (Probability; Mean; Stdev)	53
4.5.3	Example	53
4.6	Testing for the Normal Distribution	54
4.6.1	Simple Methods	55
4.6.2	Skewness and Kurtosis	56
4.6.3	Normal Plot	59
4.7	Random Numbers	60
4.8	Confidence Intervals	62
4.8.1	Confidence Interval for the Mean	63
4.8.2	Confidence Interval for the Mean in Case of a Small Sample	66
4.8.3	Confidence Interval for the Standard Deviation	70
4.9	More About the Normal Distribution	72
4.10	Lognormal Distribution	74
4.10.1	Introduction	74
4.10.2	Example	75
4.11	Control Charts: Everything in Control?	76
4.11.1	Introduction	76
4.11.2	Statistical Control	77
4.11.3	Construction of a Simple Control Chart	78
4.11.4	Unusual Patterns	78
4.11.5	Practical Use of Control Charts	79
4.11.6	Example	79
4.12	Process Capability	80
4.12.1	Introduction	81
4.12.2	Example	81

4.12.3	Process Capability Indices	81
4.12.4	Process Capability Indices: Important!	83
4.12.5	Example, continued	84
5	Analysis of Qualitative Data	85
5.1	The Binomial Distribution	85
5.1.1	Example	86
5.2	The Binomial Distribution and the Normal Distribution	87
5.3	The Binomial Distribution in Spreadsheets	89
5.3.1	Example	89
5.4	Statistical Uncertainty in Sample Surveys	90
5.4.1	Example	91
5.5	Is the Sample Representative?	95
5.6	Statistical Tests	96
5.6.1	Example	97
5.6.2	Approximation with the Normal Distribution	98
5.6.3	Significance Level	99
5.6.4	Statistical Test or Confidence Interval	99
5.7	Frequency Tables	100
5.7.1	Introduction to Chi-Squared Test	100
5.7.2	Confidence Interval for Difference Between Two Proportions	103
5.7.3	Several Rows and/or Columns	104
5.7.4	Calculations in Spreadsheets	107
5.7.5	Calculations by Calculator	108
6	Error Sources and Planning	111
6.1	Two Kinds of Errors	111
6.2	Random Error and Sample Size	111
6.2.1	A Qualitative Variable	113
6.2.2	A Quantitative Variable	115
6.3	Bias (Systematic Errors)	116
6.3.1	Errors in the Sampling (Sample Selection)	117
6.3.2	Errors in the Definition of the Sample	117
6.3.3	What Is a Representative Sample?	118
6.4	Sampling (Sample Selection)	119
6.4.1	Simple Random Sampling	119
6.4.2	Stratified Sampling	120
6.4.3	Cluster Sampling	121
6.4.4	Systematic Sampling	123
6.4.5	Quota Sampling	123
6.4.6	Purposive Sampling	124
6.4.7	Convenience Sampling	125

7	Assessment of Relationship	127
7.1	Example	128
7.2	Linear Regression with Spreadsheets	131
7.3	Is There a Relationship?	134
7.3.1	Note	135
7.4	Multiple Linear Regression	136
7.5	Final Remarks	137
8	Comparing Two Groups	139
8.1	Matched Pairs: The Paired t-Test	139
8.1.1	Example	139
8.1.2	Description	140
8.1.3	Calculation	141
8.1.4	Spreadsheets	143
8.2	Comparing Two Groups Means	144
8.2.1	Example	144
8.2.2	Description	144
8.2.3	Calculation	145
8.2.4	Spreadsheets	147
8.2.5	Size of an Experiment	148
8.3	Other Statistical Tests for Two Groups	148
8.3.1	Test for the Same Variance in the Two Groups	148
8.3.2	Comparing Two Group Means: Two Samples with Equal Variances	149
8.4	Analysis of Variance, ANOVA	150
8.4.1	Introduction	150
8.4.2	Example	150
8.5	Final Remarks	151
9	Appendices	153
9.1	Probability Theory	153
9.1.1	Sample Space, Events, and Probability	154
9.1.2	Random Variables; the Binomial Distribution	159
9.1.3	Random Variables: Mean and Variance	161
9.1.4	Technical Note: The Binomial Coefficient	163
9.2	Summary of Statistical Methods	164
9.2.1	Quantitative Data	164
9.2.2	Qualitative Data	166
9.3	Statistical Functions in Spreadsheets	168
9.4	Statistical Tables	169
9.4.1	Fractiles in the Normal Distribution	169
9.4.2	Probabilities in the Normal Distribution	169
9.4.3	Table of the t-Distribution	169

9.4.4	Table of the Chi-Squared Distribution	170
9.4.5	Statistical Uncertainty in Sample Surveys	172
9.5	Fitness Club: Data from the Sample Survey	174
9.6	Where to Go from Here	175
9.6.1	Literature	175
9.6.2	Useful Links	176
9.6.3	Overview of Statistical Software	177
9.7	Glossary	178
Index		183