

Experimental Design and Analysis

Howard J. Seltman

November 1, 2009

Preface

This book is intended as required reading material for my course, Experimental Design for the Behavioral and Social Sciences, a second level statistics course for undergraduate students in the College of Humanities and Social Sciences at Carnegie Mellon University. This course is also cross-listed as a graduate level course for Masters and PhD students (in fields other than Statistics), and supplementary material is included for this level of study.

Over the years the course has grown to include students from dozens of majors beyond Psychology and the Social Sciences and from all of the Colleges of the University. This is appropriate because Experimental Design is fundamentally the same for all fields. This book tends towards examples from behavioral and social sciences, but includes a full range of examples.

In truth, a better title for the course is Experimental Design and Analysis, and that is the title of this book. Experimental Design and Statistical Analysis go hand in hand, and neither can be understood without the other. Only a small fraction of the myriad statistical analytic methods are covered in this book, but my rough guess is that these methods cover 60%-80% of what you will read in the literature and what is needed for analysis of your own experiments. In other words, I am guessing that the first 10% of all methods available are applicable to about 80% of analyses. Of course, it is well known that 87% of statisticians make up probabilities on the spot when they don't know the true values. :)

Real examples are usually better than contrived ones, but real experimental data is of limited availability. Therefore, in addition to some contrived examples and some real examples, the majority of the examples in this book are based on simulation of data designed to match real experiments.

I need to say a few things about the **difficulties of learning** about experimental design and analysis. A practical working knowledge requires understanding many concepts and their relationships. Luckily much of what you need to learn agrees with common sense, once you sort out the terminology. On the other hand, there is no ideal logical order for learning what you need to know, because everything relates to, and in some ways depends on, everything else. So be aware: many concepts are only loosely defined when first mentioned, then further clarified later when you have been introduced to other related material. Please try not to get frustrated with some incomplete knowledge as the course progresses. If you work hard, everything should tie together by the end of the course.

In that light, I recommend that you create your own “concept maps” as the course progresses. A concept map is usually drawn as a set of ovals with the names of various concepts written inside and with arrows showing relationships among the concepts. Often it helps to label the arrows. Concept maps are a great learning tool that help almost every student who tries them. They are particularly useful for a course like this for which the main goal is to learn the relationships among many concepts so that you can learn to carry out specific tasks (design and analysis in this case). A second best alternative to making your own concept maps is to further annotate the ones that I include in this text.

This book is on the world wide web at <http://www.stat.cmu.edu/~hseltman/309/Book/Book.pdf> and any associated data files are at <http://www.stat.cmu.edu/~hseltman/309/Book/data/>.

One key idea in this course is that you cannot really learn statistics without doing statistics. Even if you will never analyze data again, the hands-on experience you will gain from analyzing data in labs, homework and exams will take your understanding of and ability to read about other peoples experiments and data analyses to a whole new level. I don’t think it makes much difference which statistical package you use for your analyses, but for practical reasons we must standardize on a particular package in this course, and that is SPSS, mostly because it is one of the packages most likely to be available to you in your future schooling and work. You will find a chapter on learning to use SPSS in this book. In addition, many of the other chapters end with “How to do it in SPSS” sections.

There are some typographical conventions you should know about. First, in a non-standard way, I use capitalized versions of Normal and Normality because I don’t want you to think that the Normal distribution has anything to do with the ordinary conversational meaning of “normal”.

Another convention is that optional material has a gray background:

I have tried to use only the minimally required theory and mathematics for a reasonable understanding of the material, but many students want a deeper understanding of what they are doing statistically. Therefore material in a gray box like this one should be considered optional extra theory and/or math.

Periodically I will summarize key points (i.e., that which is roughly sufficient to achieve a B in the course) in a box:

Key points are in boxes. They may be useful at review time to help you decide which parts of the material you know well and which you should re-read.

Less often I will sum up a larger topic to make sure you haven't "lost the forest for the trees". These are double boxed and start with "In a nutshell":

In a nutshell: You can make better use of the text by paying attention to the typographical conventions.

Chapter 1 is an overview of what you should expect to learn in this course. Chapters 2 through 4 are a review of what you should have learned in a previous course. Depending on how much you remember, you should skim it or read through it carefully. Chapter 5 is a quick start to SPSS. Chapter 6 presents the statistical foundations of experimental design and analysis in the case of a very simple experiment, with emphasis on the theory that needs to be understood to use statistics appropriately in practice. Chapter 7 covers experimental design principles in terms of preventable threats to the acceptability of your experimental conclusions. Most of the remainder of the book discusses specific experimental designs and corresponding analyses, with continued emphasis on appropriate design, analysis and interpretation. Special emphasis chapters include those on power, multiple comparisons, and model selection.

You may be interested in my background. I obtained my M.D. in 1979 and practiced clinical pathology for 15 years before returning to school to obtain my PhD in Statistics in 1999. As an undergraduate and as an academic pathologist, I carried

out my own experiments and analyzed the results of other people's experiments in a wide variety of settings. My hands on experience ranges from techniques such as cell culture, electron auto-radiography, gas chromatography-mass spectrometry, and determination of cellular enzyme levels to topics such as evaluating new radioimmunoassays, determining predictors of success in in-vitro fertilization and evaluating the quality of care in clinics vs. doctor's offices, to name a few. Many of my opinions and hints about the actual conduct of experiments come from these experiences.

As an Associate Research Professor in Statistics, I continue to analyze data for many different clients as well as trying to expand the frontiers of statistics. I have also tried hard to understand the spectrum of causes of confusion in students as I have taught this course repeatedly over the years. I hope that this experience will benefit you. I know that I continue to greatly enjoy teaching, and I am continuing to learn from my students.

Howard Seltman
August 2008

Contents

1	The Big Picture	1
1.1	The importance of careful experimental design	3
1.2	Overview of statistical analysis	3
1.3	What you should learn here	6
2	Variable Classification	9
2.1	What makes a “good” variable?	10
2.2	Classification by role	11
2.3	Classification by statistical type	12
2.4	Tricky cases	16
3	Review of Probability	19
3.1	Definition(s) of probability	19
3.2	Probability mass functions and density functions	24
3.2.1	Reading a pdf	27
3.3	Probability calculations	28
3.4	Populations and samples	34
3.5	Parameters describing distributions	35
3.5.1	Central tendency: mean and median	37
3.5.2	Spread: variance and standard deviation	38
3.5.3	Skewness and kurtosis	39

3.5.4	Miscellaneous comments on distribution parameters	39
3.5.5	Examples	40
3.6	Multivariate distributions: joint, conditional, and marginal	42
3.6.1	Covariance and Correlation	46
3.7	Key application: sampling distributions	50
3.8	Central limit theorem	52
3.9	Common distributions	54
3.9.1	Binomial distribution	54
3.9.2	Multinomial distribution	56
3.9.3	Poisson distribution	57
3.9.4	Gaussian distribution	57
3.9.5	t-distribution	59
3.9.6	Chi-square distribution	59
3.9.7	F-distribution	60
4	Exploratory Data Analysis	61
4.1	Typical data format and the types of EDA	61
4.2	Univariate non-graphical EDA	63
4.2.1	Categorical data	63
4.2.2	Characteristics of quantitative data	64
4.2.3	Central tendency	67
4.2.4	Spread	69
4.2.5	Skewness and kurtosis	71
4.3	Univariate graphical EDA	72
4.3.1	Histograms	72
4.3.2	Stem-and-leaf plots	78
4.3.3	Boxplots	79
4.3.4	Quantile-normal plots	83

4.4	Multivariate non-graphical EDA	88
4.4.1	Cross-tabulation	89
4.4.2	Correlation for categorical data	90
4.4.3	Univariate statistics by category	91
4.4.4	Correlation and covariance	91
4.4.5	Covariance and correlation matrices	93
4.5	Multivariate graphical EDA	94
4.5.1	Univariate graphs by category	95
4.5.2	Scatterplots	95
4.6	A note on degrees of freedom	98
5	Learning SPSS: Data and EDA	101
5.1	Overview of SPSS	102
5.2	Starting SPSS	104
5.3	Typing in data	104
5.4	Loading data	110
5.5	Creating new variables	116
5.5.1	Recoding	119
5.5.2	Automatic recoding	120
5.5.3	Visual binning	121
5.6	Non-graphical EDA	123
5.7	Graphical EDA	127
5.7.1	Overview of SPSS Graphs	127
5.7.2	Histogram	131
5.7.3	Boxplot	133
5.7.4	Scatterplot	134
5.8	SPSS convenience item: Explore	139
6	t-test	141

6.1	Case study from the field of Human-Computer Interaction (HCI) . . .	143
6.2	How classical statistical inference works	147
6.2.1	The steps of statistical analysis	148
6.2.2	Model and parameter definition	149
6.2.3	Null and alternative hypotheses	152
6.2.4	Choosing a statistic	153
6.2.5	Computing the null sampling distribution	154
6.2.6	Finding the p-value	155
6.2.7	Confidence intervals	159
6.2.8	Assumption checking	161
6.2.9	Subject matter conclusions	163
6.2.10	Power	163
6.3	Do it in SPSS	164
6.4	Return to the HCI example	165
7	One-way ANOVA	171
7.1	Moral Sentiment Example	172
7.2	How one-way ANOVA works	176
7.2.1	The model and statistical hypotheses	176
7.2.2	The F statistic (ratio)	178
7.2.3	Null sampling distribution of the F statistic	182
7.2.4	Inference: hypothesis testing	184
7.2.5	Inference: confidence intervals	186
7.3	Do it in SPSS	186
7.4	Reading the ANOVA table	187
7.5	Assumption checking	189
7.6	Conclusion about moral sentiments	189
8	Threats to Your Experiment	191

8.1	Internal validity	192
8.2	Construct validity	199
8.3	External validity	201
8.4	Maintaining Type 1 error	203
8.5	Power	205
8.6	Missing explanatory variables	209
8.7	Practicality and cost	210
8.8	Threat summary	210
9	Simple Linear Regression	213
9.1	The model behind linear regression	213
9.2	Statistical hypotheses	218
9.3	Simple linear regression example	218
9.4	Regression calculations	220
9.5	Interpreting regression coefficients	226
9.6	Residual checking	229
9.7	Robustness of simple linear regression	232
9.8	Additional interpretation of regression output	235
9.9	Using transformations	237
9.10	How to perform simple linear regression in SPSS	238
10	Analysis of Covariance	241
10.1	Multiple regression	241
10.2	Interaction	247
10.3	Categorical variables in multiple regression	254
10.4	ANCOVA	256
10.4.1	ANCOVA with no interaction	257
10.4.2	ANCOVA with interaction	260
10.5	Do it in SPSS	266

11 Two-Way ANOVA	267
11.1 Pollution Filter Example	271
11.2 Interpreting the two-way ANOVA results	274
11.3 Math and gender example	279
11.4 More on profile plots, main effects and interactions	284
11.5 Do it in SPSS	290
12 Statistical Power	293
12.1 The concept	293
12.2 Improving power	298
12.3 Specific researchers' lifetime experiences	302
12.4 Expected Mean Square	305
12.5 Power Calculations	306
12.6 Choosing effect sizes	308
12.7 Using n.c.p. to calculate power	309
12.8 A power applet	310
12.8.1 Overview	311
12.8.2 One-way ANOVA	311
12.8.3 Two-way ANOVA without interaction	312
12.8.4 Two-way ANOVA with interaction	314
12.8.5 Linear Regression	315
13 Contrasts and Custom Hypotheses	319
13.1 Contrasts, in general	320
13.2 Planned comparisons	324
13.3 Unplanned or post-hoc contrasts	326
13.4 Do it in SPSS	329
13.4.1 Contrasts in one-way ANOVA	329
13.4.2 Contrasts for Two-way ANOVA	335

14 Within-Subjects Designs	339
14.1 Overview of within-subjects designs	339
14.2 Multivariate distributions	341
14.3 Example and alternate approaches	344
14.4 Paired t-test	345
14.5 One-way Repeated Measures Analysis	349
14.6 Mixed between/within-subjects designs	353
14.6.1 Repeated Measures in SPSS	354
15 Mixed Models	357
15.1 Overview	357
15.2 A video game example	358
15.3 Mixed model approach	360
15.4 Analyzing the video game example	361
15.5 Setting up a model in SPSS	363
15.6 Interpreting the results for the video game example	368
15.7 Model selection for the video game example	372
15.7.1 Penalized likelihood methods for model selection	373
15.7.2 Comparing models with individual p-values	374
15.8 Classroom example	375
16 Categorical Outcomes	379
16.1 Contingency tables and chi-square analysis	379
16.1.1 Why ANOVA and regression don't work	380
16.2 Testing independence in contingency tables	381
16.2.1 Contingency and independence	381
16.2.2 Contingency tables	382
16.2.3 Chi-square test of Independence	385
16.3 Logistic regression	389

16.3.1	Introduction	389
16.3.2	Example and EDA for logistic regression	393
16.3.3	Fitting a logistic regression model	395
16.3.4	Tests in a logistic regression model	398
16.3.5	Predictions in a logistic regression model	402
16.3.6	Do it in SPSS	404
17	Going beyond this course	407