INSTRUCTOR'S SOLUTIONS MANUAL (ONLINE ONLY)

MARK SCHERVISH
Carnegie Mellon University

PROBABILITY AND STATISTICS
FOURTH EDITION

Morris DeGroot
Carnegie Mellon University

Mark Schervish
Carnegie Mellon University

Addison-Wesley
is an imprint of
PEARSON

Reproduced by Pearson Addison-Wesley from electronic files supplied by the author.

Publishing as Pearson Addison-Wesley, 75 Arlington Street, Boston, MA 02116.

ISBN-13: 978-0-321-71597-5
ISBN-10: 0-321-71597-7

Addison-Wesley
is an imprint of

PEARSON

www.pearsonhighered.com

Contents

Preface . . . vi

1 Introduction to Probability **1**
- 1.2 Interpretations of Probability . . . 1
- 1.4 Set Theory . . . 1
- 1.5 The Definition of Probability . . . 3
- 1.6 Finite Sample Spaces . . . 6
- 1.7 Counting Methods . . . 7
- 1.8 Combinatorial Methods . . . 8
- 1.9 Multinomial Coefficients . . . 13
- 1.10 The Probability of a Union of Events . . . 16
- 1.12 Supplementary Exercises . . . 20

2 Conditional Probability **25**
- 2.1 The Definition of Conditional Probability . . . 25
- 2.2 Independent Events . . . 28
- 2.3 Bayes' Theorem . . . 34
- 2.4 The Gambler's Ruin Problem . . . 40
- 2.5 Supplementary Exercises . . . 41

3 Random Variables and Distributions **49**
- 3.1 Random Variables and Discrete Distributions . . . 49
- 3.2 Continuous Distributions . . . 50
- 3.3 The Cumulative Distribution Function . . . 53
- 3.4 Bivariate Distributions . . . 58
- 3.5 Marginal Distributions . . . 64
- 3.6 Conditional Distributions . . . 70
- 3.7 Multivariate Distributions . . . 76
- 3.8 Functions of a Random Variable . . . 81
- 3.9 Functions of Two or More Random Variables . . . 85
- 3.10 Markov Chains . . . 93
- 3.11 Supplementary Exercises . . . 97

4 Expectation **107**
- 4.1 The Expectation of a Random Variable . . . 107
- 4.2 Properties of Expectations . . . 110
- 4.3 Variance . . . 113
- 4.4 Moments . . . 115
- 4.5 The Mean and the Median . . . 118

4.6 Covariance and Correlation . . . 121
4.7 Conditional Expectation . . . 124
4.8 Utility . . . 129
4.9 Supplementary Exercises . . . 134

5 Special Distributions **141**
5.2 The Bernoulli and Binomial Distributions . . . 141
5.3 The Hypergeometric Distributions . . . 145
5.4 The Poisson Distributions . . . 149
5.5 The Negative Binomial Distributions . . . 155
5.6 The Normal Distributions . . . 159
5.7 The Gamma Distributions . . . 165
5.8 The Beta Distributions . . . 171
5.9 The Multinomial Distributions . . . 174
5.10 The Bivariate Normal Distributions . . . 177
5.11 Supplementary Exercises . . . 182

6 Large Random Samples **187**
6.1 Introduction . . . 187
6.2 The Law of Large Numbers . . . 188
6.3 The Central Limit Theorem . . . 194
6.4 The Correction for Continuity . . . 198
6.5 Supplementary Exercises . . . 199

7 Estimation **203**
7.1 Statistical Inference . . . 203
7.2 Prior and Posterior Distributions . . . 204
7.3 Conjugate Prior Distributions . . . 207
7.4 Bayes Estimators . . . 214
7.5 Maximum Likelihood Estimators . . . 217
7.6 Properties of Maximum Likelihood Estimators . . . 220
7.7 Sufficient Statistics . . . 225
7.8 Jointly Sufficient Statistics . . . 228
7.9 Improving an Estimator . . . 230
7.10 Supplementary Exercises . . . 234

8 Sampling Distributions of Estimators **239**
8.1 The Sampling Distribution of a Statistic . . . 239
8.2 The Chi-Square Distributions . . . 241
8.3 Joint Distribution of the Sample Mean and Sample Variance . . . 245
8.4 The t Distributions . . . 247
8.5 Confidence Intervals . . . 250
8.6 Bayesian Analysis of Samples from a Normal Distribution . . . 254
8.7 Unbiased Estimators . . . 258
8.8 Fisher Information . . . 263
8.9 Supplementary Exercises . . . 267

9 Testing Hypotheses **273**
9.1 Problems of Testing Hypotheses . . . 273
9.2 Testing Simple Hypotheses . . . 278
9.3 Uniformly Most Powerful Tests . . . 284
9.4 Two-Sided Alternatives . . . 289
9.5 The t Test . . . 293
9.6 Comparing the Means of Two Normal Distributions . . . 296
9.7 The F Distributions . . . 299
9.8 Bayes Test Procedures . . . 303
9.9 Foundational Issues . . . 307
9.10 Supplementary Exercises . . . 309

10 Categorical Data and Nonparametric Methods **315**
10.1 Tests of Goodness-of-Fit . . . 315
10.2 Goodness-of-Fit for Composite Hypotheses . . . 317
10.3 Contingency Tables . . . 320
10.4 Tests of Homogeneity . . . 323
10.5 Simpson's Paradox . . . 325
10.6 Kolmogorov-Smirnov Tests . . . 327
10.7 Robust Estimation . . . 333
10.8 Sign and Rank Tests . . . 337
10.9 Supplementary Exercises . . . 342

11 Linear Statistical Models **349**
11.1 The Method of Least Squares . . . 349
11.2 Regression . . . 353
11.3 Statistical Inference in Simple Linear Regression . . . 356
11.4 Bayesian Inference in Simple Linear Regression . . . 364
11.5 The General Linear Model and Multiple Regression . . . 366
11.6 Analysis of Variance . . . 373
11.7 The Two-Way Layout . . . 378
11.8 The Two-Way Layout with Replications . . . 383
11.9 Supplementary Exercises . . . 389

12 Simulation **399**
12.1 What is Simulation? . . . 400
12.2 Why Is Simulation Useful? . . . 400
12.3 Simulating Specific Distributions . . . 404
12.4 Importance Sampling . . . 410
12.5 Markov Chain Monte Carlo . . . 414
12.6 The Bootstrap . . . 421
12.7 Supplementary Exercises . . . 425
R Code For Two Text Examples . . . 432

Preface

This manual contains solutions to all of the exercises in *Probability and Statistics*, 4th edition, by Morris DeGroot and myself. I have preserved most of the solutions to the exercises that existed in the 3rd edition. Certainly errors have been introduced, and I will post any errors brought to my attention on my web page `http://www.stat.cmu.edu/ mark/` along with errors in the text itself. Feel free to send me comments.

For instructors who are familiar with earlier editions, I hope that you will find the 4th edition at least as useful. Some new material has been added, and little has been removed. Assuming that you will be spending the same amount of time using the text as before, something will have to be skipped. I have tried to arrange the material so that instructors can choose what to cover and what not to cover based on the type of course they want. This manual contains commentary on specific sections right before the solutions for those sections. This commentary is intended to explain special features of those sections and help instructors decide which parts they want to require of their students. Special attention is given to more challenging material and how the remainder of the text does or does not depend upon it.

To teach a mathematical statistics course for students with a strong calculus background, one could safely cover all of the material for which one could find time. The Bayesian sections include 4.8, 7.2, 7.3, 7.4, 8.6, 9.8, and 11.4. One can choose to skip some or all of this material if one desires, but that would be ignoring one of the unique features of the text. The more challenging material in Sections 7.7–7.9, and 9.2–9.4 is really only suitable for a mathematical statistics course. One should try to make time for some of the material in Sections 12.1–12.3 even if it meant cutting back on some of the nonparametrics and two-way ANOVA. To teach a more modern statistics course, one could skip Sections 7.7–7.9, 9.2–9.4, 10.8, and 11.7–11.8. This would leave time to discuss robust estimation (Section 10.7) and simulation (Chapter 12). Section 3.10 on Markov chains is not actually necessary even if one wishes to introduce Markov chain Monte Carlo (Section 12.5), although it is helpful for understanding what this topic is about.

Using Statistical Software

The text was written without reference to any particular statistical or mathematical software. However, there are several places throughout the text where references are made to what general statistical software might be able to do. This is done for at least two reasons. One is that different instructors who wish to use statistical software while teaching will generally choose different programs. I didn't want the text to be tied to a particular program to the exclusion of others. A second reason is that there are still many instructors of mathematical probability and statistics courses who prefer not to use any software at all.

Given how pervasive computing is becoming in the use of statistics, the second reason above is becoming less compelling. Given the free and multiplatform availability and the versatility of the environment R, even the first reason is becoming less compelling. Throughout this manual, I have inserted pointers to which R functions will perform many of the calculations that would formerly have been done by hand when using this text. The software can be downloaded for Unix, Windows, or Mac OS from
`http://www.r-project.org/`
That site also has manuals for installation and use. Help is also available directly from within the R environment.

Many tutorials for getting started with R are available online. At the official R site there is the detailed manual: `http://cran.r-project.org/doc/manuals/R-intro.html`
that starts simple and has a good table of contents and lots of examples. However, reading it from start to finish is *not* an efficient way to get started. The sample sessions should be most helpful.

One major issue with using an environment like R is that it is essentially programming. That is, students who have never programmed seriously before are going to have a steep learning curve. Without going into the philosophy of whether students should learn statistics without programming, the field is moving in the direction of requiring programming skills. People who want only to understand what a statistical analysis is about can still learn that without being able to program. But anyone who actually wants to do statistics as part of their job will be seriously handicapped without programming ability. At the end of this manual is a series of heavily commented R programms that illustrate many of the features of R in the context of a specific example from the text.

Mark J. Schervish

Chapter 1

Introduction to Probability

1.2 Interpretations of Probability

Commentary

It is interesting to have the students determine some of their own subjective probabilities. For example, let X denote the temperature at noon tomorrow outside the building in which the class is being held. Have each student determine a number x_1 such that the student considers the following two possible outcomes to be equally likely: $X \le x_1$ and $X > x_1$. Also, have each student determine numbers x_2 and x_3 (with $x_2 < x_3$) such that the student considers the following three possible outcomes to be equally likely: $X \le x_2$, $x_2 < X < x_3$, and $X \ge x_3$. Determinations of more than three outcomes that are considered to be equally likely can also be made. The different values of x_1 determined by different members of the class should be discussed, and also the possibility of getting the class to agree on a common value of x_1.

Similar determinations of equally likely outcomes can be made by the students in the class for quantities such as the following ones which were found in the 1973 World Almanac and Book of Facts: the number of freight cars that were in use by American railways in 1960 (1,690,396), the number of banks in the United States which closed temporarily or permanently in 1931 on account of financial difficulties (2,294), and the total number of telephones which were in service in South America in 1971 (6,137,000).

1.4 Set Theory

Solutions to Exercises

1. Assume that $x \in B^c$. We need to show that $x \in A^c$. We shall show this indirectly. Assume, to the contrary, that $x \in A$. Then $x \in B$ because $A \subset B$. This contradicts $x \in B^c$. Hence $x \in A$ is false and $x \in A^c$.

2. First, show that $A \cap (B \cup C) \subset (A \cap B) \cup (A \cap C)$. Let $x \in A \cap (B \cup C)$. Then $x \in A$ and $x \in B \cup C$. That is, $x \in A$ and either $x \in B$ or $x \in C$ (or both). So either ($x \in A$ and $x \in B$) or ($x \in A$ and $x \in C$) or both. That is, either $x \in A \cap B$ or $x \in A \cap C$. This is what it means to say that $x \in (A \cap B) \cup (A \cap C)$. Thus $A \cap (B \cup C) \subset (A \cap B) \cup (A \cap C)$. Basically, running these steps backwards shows that $(A \cap B) \cup (A \cap C) \subset A \cap (B \cup C)$.

3. To prove the first result, let $x \in (A \cup B)^c$. This means that x is not in $A \cup B$. In other words, x is neither in A nor in B. Hence $x \in A^c$ and $x \in B^c$. So $x \in A^c \cap B^c$. This proves that $(A \cup B)^c \subset A^c \cap B^c$. Next, suppose that $x \in A^c \cap B^c$. Then $x \in A^c$ and $x \in B^c$. So x is neither in A nor in B, so it can't be in $A \cup B$. Hence $x \in (A \cup B)^c$. This shows that $A^c \cap B^c \subset (A \cup B)^c$. The second result follows from the first by applying the first result to A^c and B^c and then taking complements of both sides.

3.10 Markov Chains

Commentary

Instructors can discuss this section at any time that they find convenient or they can omit it entirely. Instructors who wish to cover Sec. 12.5 (Markov chain Monte Carlo) and who wish to give some theoretical justification for the methodology will want to discuss some of this material before covering Sec. 12.5. On the other hand, one could cover Sec. 12.5 and skip the justification for the methodology without introducing Markov chains at all.

Students may notice the following property, which is exhibited in some of the exercises at the end of this section: Suppose that the Markov chain is in a given state s_i at time n. Then the probability of being in a particular state s_j a few periods later, say at time $n+3$ or $n+4$, is approximately the same for each possible given state s_i at time n. For example, in Exercise 2, the probability that it will be sunny on Saturday is approximately the same regardless of whether it is sunny or cloudy on the preceding Wednesday, three days earlier. In Exercise 5, for given probabilities on Wednesday, the probability that it will be cloudy on Friday is approximately the same as the probability that it will be cloudy on Saturday. In Exercise 7, the probability that the student will be on time on the fourth day of class is approximately the same regardless of whether he was late or on time on the first day of class. In Exercise 10, the probabilities for $n = 3$ and $n = 4$, are generally similar. In Exercise 11, the answers in part (a) and part (b) are almost identical.

This property is a reflection of the fact that for many Markov chains, the nth power of the transition matrix $\boldsymbol{P}^n$ will converge, as $n \to \infty$, to a matrix for which all the elements in any given column are equal. For example, in Exercise 2, the matrix $\boldsymbol{P}^n$ converges to the following matrix:

$$\begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{2}{3} & \frac{1}{3} \end{bmatrix}.$$

This type of convergence is an example of Theorem 3.10.4. This theorem, and analogs for more complicated Markov chains, provide the justification of the Markov chain Monte Carlo method introduced in Sec. 12.5.

Solutions to Exercises

1. The transition matrix for this Markov chain is

$$\boldsymbol{P} = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} \end{bmatrix}.$$

   (a) If we multiply the initial probability vector by this matrix we get

$$\boldsymbol{vP} = \left(\frac{1}{2}\frac{1}{3} + \frac{1}{2}\frac{2}{3}, \frac{1}{2}\frac{2}{3} + \frac{1}{2}\frac{1}{3}\right) = \left(\frac{1}{2}, \frac{1}{2}\right).$$

   (b) The two-step transition matrix is $\boldsymbol{P}^2$, namely

$$\begin{bmatrix} \frac{1}{3}\frac{1}{3} + \frac{2}{3}\frac{2}{3} & \frac{1}{3}\frac{2}{3} + \frac{2}{3}\frac{1}{3} \\ \frac{2}{3}\frac{1}{3} + \frac{1}{3}\frac{2}{3} & \frac{2}{3}\frac{2}{3} + \frac{1}{3}\frac{1}{3} \end{bmatrix} = \begin{bmatrix} \frac{5}{9} & \frac{4}{9} \\ \frac{4}{9} & \frac{5}{9} \end{bmatrix}.$$

2. (a) 0.4, the lower right corner of the matrix.

19. (a) First, insert s from (6.2.15) into the expression in (6.2.14). We get

$$n\left[\log(p)+\left(\frac{1-p}{p}+u\right)\log\left\{\frac{(1+u)p+1-p}{up+1-p}(1-p)\right\}-\log\left\{1-\frac{1-p}{\frac{(1+u)p+1-p}{up+1-p}(1-p)}\right\}\right].$$

The last term can be rewritten as

$$-\log\left\{1-\frac{up+1-p}{(1+u)p+1-p}\right\}=-\log(p)+\log\{(1+u)p+1-p\}.$$

The result is then

$$n\left[\left(\frac{1-p}{p}+u\right)\log\left\{\frac{(1+u)p+1-p}{up+1-p}(1-p)\right\}+\log\{(1+u)p+1-p\}\right].$$

This is easily recognized as n times the logarithm of (6.2.16).

(b) For all u, q is given by (6.2.16). For $u=0$, $q=(1-p)^{(1-p)/p}$. Since $0<1-p<1$ and $(1-p)/p>0$, we have $0<q<1$ when $u=0$. For general u, let $x=p(1+u)+1-p$ and rewrite

$$\log(q)=\log(p+x)+\frac{p+x}{p}\log\frac{(1-p)(p+x)}{x}.$$

Since x is a linear increasing function of u, if we show that $\log(q)$ is decreasing in x, then q is decreasing in u. The derivative of $\log(q)$ with respect to x is

$$-\frac{p}{x(p+x)}+\frac{1}{p}\log\frac{(1-p)(p+x)}{x}.$$

The first term is negative, and the second term is negative at $u=0$ ($x=1$). To be sure that the sum is always negative, examine the second term more closely. The derivative of the second term is

$$\frac{1}{p}\left(\frac{1}{p+x}-\frac{1}{x}\right)=\frac{-1}{x(p+x)}<0.$$

Hence, the derivative is always negative, and q is less than 1 for all u.

20. We already have the m.g.f. of Y in (6.2.9). We can multiply it by $e^{-sn/10}$ and minimize over $s>0$. Before minimizing, take the logarithm:

$$\log[\psi(s)e^{-sn/10}]=n\left[\log(1/2)+\log[exp(s)+1]-\frac{3s}{5}\right]. \tag{S.6.7}$$

The derivative of this logarithm is

$$n\left[\frac{\exp(s)}{\exp(s)+1}-\frac{3}{5}\right].$$

The derivative is 0 at $s=\log(3/2)$, and the second derivative is positive there, so $s=\log(3/2)$ provides the minimum. The minimum value of (S.6.7) is -0.02014, and the Chernoff bound is $\exp(-0.02014n)=(0.98)^n$ for $\Pr(Y>n/10)$. Similarly, for $\Pr(-Y>n/10)$, we need to minimize

$$\log[\psi(-s)e^{-sn/10}]=n\left[\log(1/2)+\log[exp(-s)+1]+\frac{2s}{5}\right]. \tag{S.6.8}$$

The derivative is

$$n\left[\frac{-\exp(-s)}{\exp(-s)+1}+\frac{2}{5}\right],$$

9.5 The t Test

Commentary

This section provides a natural continuation to Sec. 9.1 in a modern statistics course. We introduce the t test and its power function, defined in terms of the noncentral t distribution. The theoretical derivation of the t test as a likelihood ratio test is isolated at the end of the section and could easily be skipped without interrupting the flow of material. Indeed, that derivation should only be of interest in a fairly mathematical statistics course.

As with confidence intervals, computer software can replace tables for obtaining quantiles of the t distributions that are used in tests. The R function `qt` can compute these. For computing p-values, one can use `pt`. The precise use of `pt` depends on whether the alternatoive hypothesis is one-sided or two-sided. For testing $H_0 : \mu \leq \mu_0$ versus $H_1 : \mu > \mu_0$ using the statistic U in Eq. (9.5.2), the p-value would be `1-pt(u,n-1)`, where `u` is the observed value of U. Fot the opposite one-sided hypotheses, the p-value would be `pt(u,n-1)`. For testing $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$, the p-value is `2*(1-pt(abs(u),n-1))`. The power function of a t test can be computed using the optional third parameter with `pt`, which is the noncentrality parameter (whose default value is 0). Similar considerations apply to the comparison of two means in Sec. 9.6.

Solutions to Exercises

1. We computed the summary statistics $\overline{x}_n = 1.379$ and $\sigma' = 0.3277$ in Example 8.5.4.

   (a) The test statistic is U from (9.5.2)
   $$U = 10^{1/2}\frac{1.379 - 1.2}{0.3277} = 1.727.$$
   We reject H_0 at level $\alpha_0 = 0.05$ if $U \geq 1.833$, the 0.95 quantile of the t distribution with 9 degrees of freedom. Since $1.727 \not\geq 1.833$, we do not reject H_0 at level 0.05.

   (b) We need to compute the probability that a t random variable with 9 degrees of freedom exceeds 1.727. This probability can be computed by most statistical software, and it equals 0.0591. Without a computer, one could interpolate in the table of the t distribution in the back of the book. That would yield 0.0618.

2. When $\mu_0 = 20$, the statistic U given by Eq. (9.5.2) has a t distribution with 8 degrees of freedom. The value of U in this exercise is 2.

   (a) We would reject H_0 if $U \geq 1.860$. Therefore, we reject H_0.

   (b) We would reject H_0 if $U \leq -2.306$ or $U \geq 2.306$. Therefore, we don't reject H_0.

   (c) We should include in the confidence interval, all values of μ_0 for which the value of U given by Eq. (9.5.2) will lie between -2.306 and 2.306. These values form the interval $19.694 < \mu_0 < 24.306$.

3. It must be assumed that the miles per gallon obtained from the different tankfuls are independent and identically distributed, and that each has a normal distribution. When $\mu_0 = 20$, the statistic U given by Eq. (9.5.2) has the t distribution with 8 degrees of freedom. Here, we are testing the following hypotheses:

   $$H_0 : \mu \geq 20,$$
   $$H_1 : \mu < 20.$$

   We would reject H_0 if $U \leq -1.860$. From the given value, it is found that $\overline{X}_n = 19$ and $S_n^2 = 22$. Hence, $U = -1.809$ and we do not reject H_0.

15. Let $Y_1 = W_1, Y_2 = W_2 - 5$, and $Y_3 = \frac{1}{2}W_3$. Then the random vector

$$\boldsymbol{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix}$$

satisfies the conditions of the general linear model as described in Sec. 11.5 with

$$\boldsymbol{Z} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & -1 \end{bmatrix}, \qquad \boldsymbol{\beta} = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}.$$

Thus,

$$\boldsymbol{Z}'\boldsymbol{Z} = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}, \qquad (\boldsymbol{Z}'\boldsymbol{Z})^{-1} = \begin{bmatrix} 3/8 & -1/8 \\ -1/8 & 3/8 \end{bmatrix},$$

and

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{Y} = \begin{bmatrix} \frac{1}{4}Y_1 + \frac{1}{4}Y_2 + \frac{1}{2}Y_3 \\ \frac{1}{4}Y_1 + \frac{1}{4}Y_2 - \frac{1}{2}Y_3 \end{bmatrix}.$$

Also,

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{3}(\boldsymbol{Y} - \boldsymbol{Z}\hat{\boldsymbol{\beta}})'(\boldsymbol{Y} - \boldsymbol{Z}\hat{\boldsymbol{\beta}}) \\ &= \frac{1}{3}\left[(Y_1 - \hat{\theta}_1 - \hat{\theta}_2)^2 + (Y_2 - \hat{\theta}_1 - \hat{\theta}_2)^2 + (Y_3 - \hat{\theta}_1 + \hat{\theta}_2)^2\right]. \end{aligned}$$

The following distributional properties of these M.L.E.'s are known from Sec. 11.5: $(\hat{\theta}_1, \hat{\theta}_2)$ and $\hat{\sigma}^2$ are independent; $(\hat{\theta}_1, \hat{\theta}_2)$ has a bivariate normal distribution with mean vector (θ_1, θ_2) and covariance matrix

$$\sigma^2(\boldsymbol{Z}'\boldsymbol{Z})^{-1} = \begin{bmatrix} 3/8 & -1/8 \\ -1/8 & 3/8 \end{bmatrix}\sigma^2;$$

$3\hat{\sigma}^2/\sigma^2$ has a χ^2 distribution with one degree of freedom.

16. Direct application of the theory of least squares would require choosing α and β to minimize

$$Q_1 = \sum_{i=1}^{n}(y_i - \alpha x_i^{\beta})^2.$$

This minimization must be carried out numerically since the solution cannot be found in closed form. However, if we express the required curve in the form $\log y = \log\alpha + \beta\log x$, and then apply the method of least squares, we must choose β_0 and β_1 to minimize

$$Q_2 = \sum_{i=1}^{n}(\log y_i - \beta_0 - \beta_1\log x_i)^2,$$