

# SpringerBriefs in Statistics

For further volumes:  
<http://www.springer.com/series/8921>



**Photograph 1** Tiritiri Island, Auckland, New Zealand. (Photo: SP)

Simo Puntanen  
George P. H. Styan  
Jarkko Isotalo

# Formulas Useful for Linear Regression Analysis and Related Matrix Theory

It's Only Formulas But We Like Them

 Springer

Simo Puntanen  
School of Information Sciences  
University of Tampere  
Tampere, Finland

George P. H. Styan  
Department of Mathematics  
and Statistics  
McGill University  
Montréal, QC, Canada

Jarkko Isotalo  
Department of Forest Sciences  
University of Helsinki  
Helsinki, Finland

ISSN 2191-544X  
ISBN 978-3-642-32930-2  
DOI 10.1007/978-3-642-32931-9  
Springer Heidelberg New York Dordrecht London

ISSN 2191-5458 (electronic)  
ISBN 978-3-642-32931-9 (eBook)

Library of Congress Control Number: 2012948184

© The Author(s) 2013

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media ([www.springer.com](http://www.springer.com))

# Preface

*Lie la lie, lie la la-lie lie la-lie.*

*There must be fifty ways to leave your lover.*

*Oh, still crazy after all these years.*

PAUL SIMON<sup>1</sup>

Think about going to a lonely island for some substantial time and that you are supposed to decide what books to take with you. This book is then a serious alternative: it does not only guarantee a good night's sleep (reading in the late evening) but also offers you a survival kit in your urgent regression problems (definitely met at the day time on any lonely island, see for example Photograph 1, p. ii).

Our experience is that even though a huge amount of the formulas related to linear models is available in the statistical literature, it is not always so easy to catch them when needed. The purpose of this book is to collect together a good bunch of helpful rules—within a limited number of pages, however. They all exist in literature but are pretty much scattered. The first version (technical report) of the *Formulas* appeared in 1996 (54 pages) and the fourth one in 2008. Since those days, the authors have never left home without the *Formulas*.

This book is not a regular textbook—this is supporting material for courses given in linear regression (and also in multivariate statistical analysis); such courses are extremely common in universities providing teaching in quantitative statistical analysis. We assume that the reader is somewhat familiar with linear algebra, matrix calculus, linear statistical models, and multivariate statistical analysis, although a thorough knowledge is not needed, one year of undergraduate study of linear algebra and statistics is expected. A short course in regression would also be necessary before traveling with our book. Here are some examples of smooth introductions to regression: Chatterjee & Hadi (2012) (first ed. 1977), Draper & Smith (1998) (first ed. 1966), Seber & Lee (2003) (first ed. 1977), and Weisberg (2005) (first ed. 1980).

The term *regression* itself has an exceptionally interesting history: see the excellent chapter entitled *Regression towards Mean* in Stigler (1999), where (on p. 177) he says that the story of Francis Galton's (1822–1911) discovery of regression is “an exciting one, involving science, experiment, mathematics, simulation, and one of the great thought experiments of all time”.

---

<sup>1</sup> From (1) *The Boxer*, a folk rock ballad written by Paul Simon in 1968 and first recorded by Simon & Garfunkel, (2) *50 Ways to Leave Your Lover*, a 1975 song by Paul Simon, from his album “Still Crazy After All These Years”, (3) *Still Crazy After All These Years*, a 1975 song by Paul Simon and title track from his album “Still Crazy After All These Years”.

This book is neither a real handbook: by a handbook we understand a thorough representation of a particular area. There are some recent handbook-type books dealing with matrix algebra helpful for statistics. The book by Seber (2008) should be mentioned in particular. Some further books are, for example, by Abadir & Magnus (2005) and Bernstein (2009). Quick visits to matrices in linear models and multivariate analysis appear in Puntanen, Seber & Styan (2013) and in Puntanen & Styan (2013).

We do not provide any proofs nor references. The book by Puntanen, Styan & Isotalo (2011) offers many proofs for the formulas. The website <http://www.sis.uta.fi/tilasto/matrixtricks> supports both these books by additional material.

Sincere thanks go to Götz Trenkler, Oskar Maria Baksalary, Stephen J. Haslett, and Kimmo Vehkalahti for helpful comments. We give special thanks to Jarmo Niemelä for his outstanding L<sup>A</sup>T<sub>E</sub>X assistance. The Figure 1 (p. 126) was prepared using the Survo software, online at <http://www.survo.fi> (thanks go to Kimmo Vehkalahti) and the Figure 2 (p. 126) using PStricks (thanks again going to Jarmo Niemelä).

We are most grateful to Alice Blanck, Ulrike Stricker-Komba, and to Niels Peter Thomas of Springer for advice and encouragement.

This research has been supported in part by the Natural Sciences and Engineering Research Council of Canada.

SP, GPHS & JI  
June 7, 2012

*MSC 2000:* 15-01, 15-02, 15A09, 15A42, 15A99, 62H12, 62J05.

*Key words and phrases:* Best linear unbiased estimation, Cauchy–Schwarz inequality, column space, eigenvalue decomposition, estimability, Gauss–Markov model, generalized inverse, idempotent matrix, linear model, linear regression, Löwner ordering, matrix inequalities, oblique projector, ordinary least squares, orthogonal projector, partitioned linear model, partitioned matrix, rank cancellation rule, reduced linear model, Schur complement, singular value decomposition.

## References

- Abadir, K. M. & Magnus, J. R. (2005). *Matrix Algebra*. Cambridge University Press.
- Bernstein, D. S. (2009). *Matrix Mathematics: Theory, Facts, and Formulas*. Princeton University Press.
- Chatterjee, S. & Hadi, A. S. (2012). *Regression Analysis by Example*, 5th Edition. Wiley.
- Draper, N. R. & Smith, H. (1998). *Applied Regression Analysis*, 3rd Edition. Wiley.
- Puntanen, S., Styan, G. P. H. & Isotalo, J. (2011). *Matrix Tricks for Linear Statistical Models: Our Personal Top Twenty*. Springer.
- Puntanen, S. & Styan, G. P. H. (2013). Chapter 52: Random Vectors and Linear Statistical Models. *Handbook of Linear Algebra*, 2nd Edition (Leslie Hogben, ed.), Chapman & Hall, in press.
- Puntanen, S., Seber, G. A. F. & Styan, G. P. H. (2013). Chapter 53: Multivariate Statistical Analysis. *Handbook of Linear Algebra*, 2nd Edition (Leslie Hogben, ed.), Chapman & Hall, in press.
- Seber, G. A. F. (2008). *A Matrix Handbook for Statisticians*. Wiley.
- Seber, G. A. F. & Lee, A. J. (2006). *Linear Regression Analysis*, 2nd Edition. Wiley.
- Stigler, S. M. (1999). *Statistics on the Table: The History of Statistical Concepts and Methods*. Harvard University Press.
- Weisberg, S. (2005). *Applied Linear Regression*, 3rd Edition. Wiley.

# Contents

Formulas Useful for Linear Regression Analysis and Related Matrix Theory . .	1
1 The model matrix & other preliminaries . . . . .	1
2 Fitted values and residuals . . . . .	8
3 Regression coefficients . . . . .	10
4 Decompositions of sums of squares . . . . .	15
5 Distributions . . . . .	19
6 Best linear predictor . . . . .	27
7 Testing hypotheses . . . . .	33
8 Regression diagnostics . . . . .	39
9 BLUE: Some preliminaries . . . . .	44
10 Best linear unbiased estimator . . . . .	48
11 The relative efficiency of OLSE . . . . .	58
12 Linear sufficiency and admissibility . . . . .	63
13 Best linear unbiased predictor . . . . .	65
14 Mixed model . . . . .	68
15 Multivariate linear model . . . . .	72
16 Principal components, discriminant analysis, factor analysis . . . . .	74
17 Canonical correlations . . . . .	76
18 Column space properties and rank rules . . . . .	78
19 Inverse of a matrix . . . . .	82
20 Generalized inverses . . . . .	86
21 Projectors . . . . .	91
22 Eigenvalues . . . . .	96
23 Singular value decomposition & other matrix decompositions . . . . .	105
24 Löwner ordering . . . . .	109
25 Inequalities . . . . .	112
26 Kronecker product, some matrix derivatives . . . . .	115
Index . . . . .	117