

*Advisors:*

Stephen Fienberg   Ingram Olkin

## Springer Texts in Statistics

---

<i>Alfred</i>	Elements of Statistics for the Life and Social Sciences
<i>Chow and Teicher</i>	Probability Theory: Independence, Interchangeability, Martingales Second Edition
<i>Christensen</i>	Plane Answers to Complex Questions: The Theory of Linear Models
<i>du Toit, Steyn, and Strumpf</i>	Graphical Exploratory Data Analysis
<i>Kalbfleisch</i>	Probability and Statistical Inference: Volume I: Probability. Second Edition
<i>Kalbfleisch</i>	Probability and Statistical Inference: Volume 2: Statistical Inference. Second Edition
<i>Keyfitz</i>	Applied Mathematical Demography. Second Edition
<i>Kiefer</i>	Introduction to Statistical Inference
<i>Madansky</i>	Prescriptions for Working Statisticians
<i>Peters</i>	Counting for Something: Statistical Principles and Personalities

Albert Madansky

# Prescriptions for Working Statisticians

With 25 Illustrations



Springer-Verlag  
New York Berlin Heidelberg  
London Paris Tokyo

Albert Madansky  
Graduate School of Business  
University of Chicago  
Chicago, IL 60637  
U.S.A.

*Editorial Board*

Stephen Fienberg  
Department of Statistics  
Carnegie Mellon University  
Pittsburgh, PA 15213  
U.S.A.

Ingram Olkin  
Department of Statistics  
Stanford University  
Stanford, CA 94305  
U.S.A.

---

Mathematics Subject Classification (1980): 62-01

---

Library of Congress Cataloging-in-Publication Data

Madansky, Albert

Prescriptions for working statisticians.

(Springer texts in statistics)

Includes index.

1. Statistics. I. Title. II. Series.

QA276.M27 1988 519.5 87-28346

ISBN-13: 978-1-4612-8354-6

e-ISBN-13: 978-1-4612-3794-5

DOI: 10.1007/978-1-4612-3794-5

© 1988 by Springer-Verlag New York Inc.

Softcover reprint of the hardcover 1st edition 1988

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer-Verlag, 175 Fifth Avenue, New York, NY 10010, U.S.A.), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden. The use of general descriptive names, trade names, trademarks, etc. in this publication, even if the former are not especially identified, is not to be taken as a sign that such names, as understood by the Trade Marks and Merchandise Marks Act, may accordingly be used freely by anyone.

Typeset by Asco Trade Typesetting Ltd., Hong Kong.

9 8 7 6 5 4 3 2 1

To my daughters—Susan, Cynthia, Noreen, and Michele

רבות בנוחי עשו חיל

adapted from Proverbs 31:29

# Contents

Preface	xi
Acknowledgments	xix
CHAPTER 0	
A Thoughtful Student's Retrospective on Statistics 101	1
0. Introduction	1
1. The Introductory Model	1
2. The Regression Model	4
References	12
CHAPTER 1	
Testing for Normality	14
0. Introduction	14
1. Normal Plots	16
2. Regression Procedures	20
(a) Shapiro–Wilk Statistic	20
(b) Filliben Statistic	30
(c) D'Agostino Statistic	31
3. Studentized Range	34
4. Moment Checking	36
5. Standard Tests of Goodness-of-Fit	40
(a) Chi-Square Test	41
(b) Kolmogorov Test	44
(c) Durbin Test	50
6. Evaluation	50
Appendix I	53
References	53

## CHAPTER 2

Testing for Homoscedasticity	56
0. Introduction	56
1. Comparing Variances of Two Normal Distributions	59
2. Testing Homoscedasticity of Many Populations	61
(a) Bartlett Test	61
(b) Normal Score Procedure (Fligner–Killeen)	65
(c) Use of Analysis of Variance (Levene)	67
(d) Evaluation	68
3. Regression Residuals	68
(a) Standard Estimates	68
(b) Homoscedastic Estimated Residuals	70
(i) BLUS	70
(ii) Recursive Residuals	73
4. Testing Homoscedasticity of Regression Residuals	75
(a) Goldfeld–Quandt Procedure	78
(b) Lagrange Multiplier Test	81
(c) White Procedure	82
(d) Robust Procedures	83
(e) Nonparametric Tests	85
(f) Adapting Tests of Homogeneity of Variances	88
(g) Evaluation	88
Appendix I	89
References	89

## CHAPTER 3

Testing for Independence of Observations	92
0. Introduction	92
1. Parametric Procedures	93
(a) Independence of Observations	93
(b) Autocorrelation	101
(c) Independence of Regression Residuals	102
(d) Box–Pierce Statistic	104
2. Nonparametric Procedures	106
(a) Runs Above and Below the Median	106
(b) Runs Up-and-Down	113
(c) Rank von Neumann Ratio	116
References	119

## CHAPTER 4

Identification of Outliers	120
0. Introduction	120
1. Normal Distribution	122
(a) One Outlier	122
(b) Multiple Outlier Indication	126
(c) Multiple Outlier Detection—Known Number	127
(d) Multiple Outlier Detection—Unknown Number	134

2. Nonparametric Procedures	137
3. Outliers in Regression	140
References	147
CHAPTER 5	
Transformations	148
0. Introduction	148
1. Deflating Heteroscedastic Regressions	148
2. Variance Stabilizing Transformations	156
3. Power Transformations (Box–Cox)	158
(a) Maximum Likelihood	160
(b) Hinkley Estimation Procedure	168
(c) Graphic Procedure	170
4. Letter-Values and Boxplots	173
5. Power Transformations of Regression Independent Variables	175
Appendix I	178
References	179
CHAPTER 6	
Independent Variable Selection in Multiple Regression	181
0. Introduction	181
1. Criteria for Goodness of Regression Model	181
2. Stepwise Procedures	191
(a) Order of Selection of Variables	192
(b) Theoretical Description	193
(c) Cutoff Rules	196
(d) Example	197
3. Multicollinearity	199
(a) Description of the Phenomenon	199
(b) Diagnostics	205
(c) Ridge Regression	208
References	211
CHAPTER 7	
Categorical Variables in Regression	214
0. Introduction	214
1. Two Sample Tests	214
(a) Parametric Procedures	215
(b) Nonparametric Procedures	219
2. Analysis of Variance via Regression (Model I)	228
(a) One-Way Anova	228
(b) Unbalanced Two-Way Anova with Only Main Effects	232
(c) Unbalanced Two-Way Anova with Interaction	235
(d) Median Polish	241
3. Components of Variance (Model II)	244
4. Dichotomous Dependent Variables	246
References	249



CHAPTER 8	
Analysis of Cross-Classified Data	252
0. Introduction	252
1. Independence in the $r_1 \times r_2$ Table	254
(a) General	254
(b) $2 \times 2$ Tables	256
2. Log-Linear Models in the $r_1 \times r_2$ Table	259
3. The Three-Dimensional Table	261
4. Analysis of Cross-Classifications with Ordered Categories	266
5. Latent Class Model	270
Appendix I	274
Appendix II	282
References	285
Index of Reference Tables	289
Index	291

# Preface

The first course in statistics, no matter how “good” or “long” it is, typically covers inferential procedures which are valid only if a number of preconditions are satisfied by the data. For example, students are taught about regression procedures valid only if the true residuals are independent, homoscedastic, and normally distributed. But they do not learn how to check for independence, homoscedasticity, or normality, and certainly do not learn how to adjust their data and/or model so that these assumptions are met.

To help this student out<sup>1</sup> I designed a second course, containing a collection of statistical diagnostics and prescriptions necessary for the applied statistician so that he can deal with the realities of inference from data, and not merely with the kind of classroom problems where all the data satisfy the assumptions associated with the technique to be taught. At the same time I realized that I was writing a book for a wider audience, namely all those away from the classroom whose formal statistics education ended with such a course and who apply statistical techniques to data. Thus the book begins with four chapters of *diagnostics*, answering in turn the following questions:

- How do I find out whether my data came from normal distributions?
- How do I find out whether my data are homoscedastic?
- How do I find out whether my data are independent?
- How do I find out whether there are outliers in my data?

I know of no book that treats the material in Chapters 1, 2, and 3, let alone combines those topics under one cover. Two books are devoted to the topic covered in Chapter 4; we include the most relevant material between these covers as well.

<sup>1</sup> This student’s thoughts as he takes the first course are recorded in Chapter 0, entitled “A Thoughtful Student’s Retrospective on Statistics 101”.

Next the book presents two chapters of *prescriptions*. Chapter 5 answers the questions:

- How do I fix up my data so that we can treat them as if they were normally distributed?
- How do I fix up my regression variables so that the residuals are homoscedastic?
- How do I concoct an appropriate set of independent variables for a regression?

Chapter 6 answers the most frequently asked questions by the student in that “first course”:

- How do I know how many independent variables to put into a regression model?
- What do I do about multicollinearity?

The material in Chapter 5 is, to my knowledge, not to be found in any other book. Answers to the questions covered in my Chapter 6 are well presented in Chapter 6 of Draper and Smith [1966] or Chatterjee and Price [1977]; again my justification for writing this chapter is to include it between the covers of a single book of the timbre of this one.

Finally, the book contains two chapters on *techniques* for making inferences from specialized data, mixed categorical and measured data (Chapter 7), and cross-classified data (Chapter 8). Though the material in these chapters can be found in a number of other books, our approach to this material may be a little different from the usual fare, and this slant may give the reader added insight into the techniques covered in these chapters—in the main, into analysis of variance and into  $\log r$  linear modeling of contingency tables. Chapter 7 was motivated by the observation that a first course teaches the student both the two-sample  $t$  test and multiple regression. What would then be more natural in a second course than to use the multiple regression model with dummy independent variables as the vehicle for teaching students the generalization of the  $t$  test, the analysis of variance? Once dummy variables are introduced in regression, a natural further topic is the use of a “dummy” variable as a dependent variable. This, too, is included in Chapter 7.

The last topic, along with the technology of the analysis of variance, leads naturally to a topic not treated in a first course but which is probably as important as  $t$  tests and regression, namely the analysis of counted data. Chapter 8 gives a brief summary of  $\log r$  linear model analysis for counted data, more to give the reader an acquaintance with these methods, and whet his appetite for more, than replace the many excellent books devoted to this subject.

In summary, I have written a book that is ostensibly for a “second course” in applied statistics but actually may have merit as a ready handbook for the practitioner. To this end, it is self-contained in that all necessary tables are included in this book. When there was a number of competitive procedures for the same problem, I was faced with an editorial choice: Should I only

include the best procedure and disregard the rest, or include them all? My answer to that question was the latter, for a few reasons. First, in some cases we don't yet know which procedure is best. (Of course, for situations in which something is known about the properties of the procedure, I include that information.) Second, sometimes a "better" procedure may be more complex to implement than a "worse" procedure; by describing both the reader can select the one he chooses to use based not merely on "goodness" but on practicality. Finally, some of the "not-so-good" procedures are so well known and well ingrained in statistical practice, or are intellectual forerunners of the "better" procedures, that their absence would be noticed (and have been, in earlier drafts).

One institutional idiosyncrasy which creeps into this book, but in a minor way and can be brushed aside by the disinterested reader, is my occasional use of IDA, an interactive data analysis system available to students at the University of Chicago Graduate School of Business (see Roberts and Ling [1982]). These uses have purposely not been edited out, partly to enhance this book's direct value to my students, but also partly to stimulate the reader's interest in interactive data analysis systems as the "way of life" in current statistical practice. Moreover, as MINITAB (see Ryan, Joiner, and Ryan [1985]) embodies commands which implement some of the procedures described herein, I include some examples on the use of that data analysis system as well.

As I was writing these chapters, I had to grapple with the problem as to what level of mathematical and statistical background to presuppose of the reader. Some of the material requires no more mathematics than high school algebra; others require a good knowledge of matrix algebra. To cope with this range of prerequisites, I have adopted a number of expository ground rules:

- (1) describe the procedures in as clear a manner as I can consonant with the maturity level necessary for the user of the procedure;
- (2) provide motivation for the procedure, but stop short of a mathematical derivation;
- (3) provide an example of the use and computation of each procedure;
- (4) provide a FORTRAN computer program for implementing the more mathematically complex procedures.

We require three items of statistics background:

- (1) familiarity with multiple regression modeling and at least a nodding acquaintance with the matricial representation of the regression parameter estimates;
- (2) exposure to the  $t$ ,  $F$ , chi-square, beta, gamma, binomial, Poisson, and uniform densities, at least to the point of knowing their mean and variance and how to read tables of their distributions; and
- (3) understanding of the workings of the expectation ( $\mathcal{E}$ ), variance ( $\mathcal{V}$ ), and covariance ( $\mathcal{C}$ ) operators.

These are for the most part covered in a good first course in statistics. Any

gap between what the reader has learned in such a course and a particular fact needed to read this book can be bridged by the reader with a minimum amount of individual effort.

Thus, the book is not written toward a single prototypical reader with a specific background. The applied statistician with a background in matrix algebra and a prior background in regression will find this book, at even its most difficult points, right at his level and in its easiest parts quite tedious. The reader whose background is merely high school algebra, and a first course in statistics which treated only the normal distribution, will find some of the material opaque. To him, I can only recommend that he read the opaque material, at least to ascertain the general drift of the described material, in order that, knowing of its existence, he can invoke it at appropriate occasions with the help of a more mathematically trained statistician. I have helped the reader along by segregating the material in each section of the book into two categories, labeled “Prescription” and “Theoretical Background”, respectively. I have tried to make the “Prescription” material more intuitive, readable to the practitioner, and “cook-bookish”; and even in the “Theoretical Background” most theoretical results are stated without proof. Of course, one can pursue any topic in the book in greater depth by going back to the cited references.

The reader should be forewarned that I am a bit idiosyncratic in my notation. I use lowercase letters to denote scalar variables and uppercase letters to denote matrix (and vector) variables. I use regular type to denote nonrandom variables and boldface type to denote random variables. Thus we can look at a symbol and know whether or not it is a matrix and whether or not it is random. (No longer need we write  $X$  for a random variable and  $x$  for its value, or equations like  $X = x$ , and no longer will we be confused about what symbols to use for a random matrix and its value.) Schematically, the notation is the following:

	Nonrandom	Random
Scalar	$x$	<b><math>x</math></b>
Matrix	$X$	<b><math>X</math></b>

At the blackboard one can easily denote randomness by using the printer’s symbol  $\sim$  under the letter to indicate randomness, e.g., what appears in this book as  $\mathbf{u}$  would be written on the blackboard as  $\underset{\sim}{u}$ . (This is by contrast to those who are already denoting randomness by using the  $\sim$  above the letter to indicate randomness.) Finally, I have attempted to name all parameters with Greek letters, and here again the case of the letter indicates whether it is a scalar or a matrix.

To make this book as usable as possible, I have appended to each chapter all the specialized tables necessary to perform the statistical procedures described therein, though in most cases only for the 5% significance level. Since

tables of the  $\chi^2$  and  $F$  distribution are necessary for almost all the chapters of this book, I append them for ready reference in this Preface.

## References

- Chatterjee, S. and Price, B. 1977. *Regression Analysis by Example*. New York: Wiley.  
 Draper, N. R. and Smith, H. 1966. *Applied Regression Analysis*. New York: Wiley.  
 Pearson, E. S. and Hartley, H. O., 1966 *Biometrika Tables for Statisticians*, Vol. I, 3rd ed. Cambridge: Cambridge University Press.  
 Ryan, B. F., Joiner, B. I., and Ryan, T. R., Jr. 1985. *Minitab Handbook*. Florence, KY: Wadsworth.  
 Roberts, H. V. and Ling, R. F. 1982. *Conversational Statistics with IDA*. New York: Scientific Press and McGraw-Hill.

## Percentage Points of the $\chi^2$ Distribution.

$\chi^2$	0.995	0.990	0.975	0.950	0.900	0.750	0.500
1	392.704. 10 <sup>-10</sup>	157088. 10 <sup>-9</sup>	982069. 10 <sup>-8</sup>	393214. 10 <sup>-8</sup>	0.0157908	0.1015308	0.454937
2	0.0100251	0.0201007	0.0506356	0.102587	0.210720	0.575364	1.38629
3	0.0717212	0.114832	0.215795	0.351846	0.584375	1.212534	2.36597
4	0.206990	0.297110	0.484419	0.710721	1.063623	1.92255	3.35670
5	0.411740	0.554300	0.831211	1.145476	1.61031	2.67460	4.35146
6	0.675727	0.872085	1.237347	1.63539	2.20413	3.45460	5.34812
7	0.989265	1.239043	1.68987	2.16735	2.83311	4.25485	6.34581
8	1.344419	1.646482	2.17973	2.73264	3.48954	5.07064	7.34412
9	1.734926	2.087912	2.70039	3.32511	4.16816	5.89883	8.34283
10	2.15585	2.55821	3.24697	3.94030	4.86518	6.73720	9.34182
11	2.60321	3.05347	3.81575	4.57481	5.57779	7.58412	10.3410
12	3.07382	3.57056	4.40379	5.22603	6.30380	8.43842	11.3403
13	3.56503	4.10691	5.00874	5.89186	7.04150	9.29906	12.3398
14	4.07468	4.66043	5.62872	6.57063	7.78953	10.1653	13.3393
15	4.60094	5.22935	6.26214	7.26094	8.54675	11.0365	14.3389
16	5.14224	5.81221	6.90766	7.96164	9.31223	11.9122	15.3385
17	5.69724	6.40776	7.56418	8.67176	10.0852	12.7919	16.3381
18	6.26481	7.01491	8.23075	9.39046	10.8649	13.6753	17.3379
19	6.84398	7.63273	8.90655	10.1170	11.6509	14.5620	18.3376
20	7.43386	8.26040	9.59083	10.8508	12.4426	15.4518	19.3374
21	8.03366	8.89720	10.28293	11.5913	13.2396	16.3444	20.3372
22	8.64272	9.54249	10.9823	12.3380	14.0415	17.2396	21.3370
23	9.26042	10.19567	11.6885	13.0905	14.8479	18.1373	22.3369
24	9.88623	10.8564	12.4011	13.8484	15.6587	19.0372	23.3367
25	10.5197	11.5240	13.1197	14.6114	16.4734	19.9393	24.3366
26	11.1603	12.1981	13.8439	15.3791	17.2919	20.8434	25.3364
27	11.8076	12.8786	14.5733	16.1513	18.1138	21.7494	26.3363
28	12.4613	13.5648	15.3079	16.9279	18.9392	22.6572	27.3363
29	13.1211	14.2565	16.0471	17.7083	19.7677	23.5666	28.3362
30	13.7867	14.9535	16.7908	18.4926	20.5992	24.4776	29.3360
40	20.7065	22.1643	24.4331	26.5093	29.0505	33.6603	39.3354
50	27.9907	29.7067	32.3574	34.7642	37.6886	42.9421	49.3349
60	35.5346	37.4848	40.4817	43.1879	46.4589	52.2938	59.3347
70	43.2752	45.4418	48.7576	51.7393	55.3290	61.6983	69.3344
80	51.1720	53.5400	57.1532	60.3915	64.2778	71.1445	79.3343
90	59.1963	61.7541	65.6466	69.1260	73.2912	80.6247	89.3342
100	67.3276	70.0648	74.2219	77.9295	82.3581	90.1332	99.3341
X	-2.5758	-2.3263	-1.9600	-1.6449	-1.2816	-0.6745	0.0000

SOURCE: Pearson and Hartley [1958]. Reprinted from *Biometrika Tables for Statisticians*, Vol. I, with the permission of the Biometrika Trustees.

Percentage Points of the  $\chi^2$  Distribution (*continued*)

$\nu \backslash Q$	0.250	0.100	0.050	0.025	0.010	0.005	0.001
1	1.32330	2.70554	3.84146	5.02389	6.63490	7.87944	10.828
2	2.77259	4.60517	5.99147	7.37776	9.21034	10.5966	13.816
3	4.10835	6.25139	7.81473	9.34840	11.3449	12.8381	16.266
4	5.38527	7.77944	9.48773	11.1433	13.2767	14.8602	18.467
5	6.62568	9.23635	11.0705	12.8325	15.0863	16.7496	20.515
6	7.84080	10.6446	12.5916	14.4494	16.8119	18.5476	22.458
7	9.03715	12.0170	14.0671	16.0128	18.4753	20.2777	24.322
8	10.2188	13.3616	15.5073	17.5346	20.0902	21.9550	26.125
9	11.3887	14.6837	16.9190	19.0228	21.6660	23.5893	27.877
10	12.5489	15.9871	18.3070	20.4831	23.2093	25.1882	29.588
11	13.7007	17.2750	19.6751	21.9200	24.7250	26.7569	31.264
12	14.8454	18.5494	21.0261	23.3367	26.2170	28.2995	32.909
13	15.9839	19.8119	22.3621	24.7356	27.6883	29.8194	34.528
14	17.1170	21.0642	23.6848	26.1190	29.1413	31.3193	36.123
15	18.2451	22.3072	24.9958	27.4884	30.5779	32.8013	37.697
16	19.3688	23.5418	26.2962	28.8454	31.9999	34.2672	39.252
17	20.4887	24.7690	27.5871	30.1910	33.4087	35.7185	40.790
18	21.6049	25.9894	28.8693	31.5264	34.8053	37.1564	42.312
19	22.7178	27.2036	30.1435	32.8523	36.1908	38.5822	43.820
20	23.8277	28.4120	31.4104	34.1696	37.5662	39.9968	45.315
21	24.9348	29.6151	32.6705	35.4789	38.9321	41.4010	46.797
22	26.0393	30.8133	33.9244	36.7807	40.2894	42.7956	48.268
23	27.1413	32.0069	35.1725	38.0757	41.6384	44.1813	49.728
24	28.2412	33.1963	36.4151	39.3641	42.9798	45.5585	51.179
25	29.3389	34.3816	37.6525	40.6465	44.3141	46.9278	52.620
26	30.4345	35.5631	38.8852	41.9232	45.6417	48.2899	54.052
27	31.5284	36.7412	40.1133	43.1944	46.9630	49.6449	55.476
28	32.6205	37.9159	41.3372	44.4607	48.2782	50.9933	56.892
29	33.7109	39.0875	42.5569	45.7222	49.5879	52.3356	58.302
30	34.7998	40.2560	43.7729	46.9792	50.8922	53.6720	59.703
40	45.6180	51.8050	55.7585	59.3417	63.6907	66.7659	73.402
50	56.3336	63.1671	67.5048	71.4202	76.1539	79.4900	86.661
60	66.9814	74.3970	79.0819	83.2976	88.3794	91.9517	99.607
70	77.5766	85.5271	90.5312	95.0231	100.425	104.215	112.317
80	88.1303	96.5782	101.879	106.629	112.329	116.321	124.839
90	98.6499	107.565	113.145	118.136	124.116	128.299	137.208
100	109.141	118.498	124.342	129.561	135.807	140.169	149.449
X	+ 0.6745	+ 1.2816	+ 1.6449	+ 1.9600	+ 2.3263	+ 2.5758	+ 3.0902

SOURCE: Pearson and Hartley [1958]. Reprinted from *Biometrika Tables for Statisticians*, Vol. I, with the permission of the Biometrika Trustees.

Percentage Points of the *F* Distribution.  
Upper 2.5% Points.

$\frac{v_1}{v_2}$	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	$\infty$
1	647.8	799.5	894.2	899.6	921.8	937.1	948.2	956.7	963.3	968.6	976.7	984.9	993.1	997.2	1001	1006	1010	1014	1018
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.41	39.43	39.45	39.46	39.46	39.47	39.48	39.49	39.50
3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42	14.34	14.25	14.17	14.12	14.08	14.04	13.99	13.95	13.90
4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.75	8.66	8.56	8.51	8.48	8.41	8.36	8.31	8.26
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.52	6.43	6.33	6.28	6.23	6.18	6.12	6.07	6.02
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.37	5.27	5.17	5.12	5.07	5.01	4.96	4.90	4.85
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.67	4.57	4.47	4.42	4.36	4.31	4.25	4.20	4.14
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.20	4.10	4.00	3.95	3.89	3.84	3.78	3.73	3.67
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.87	3.77	3.67	3.61	3.56	3.51	3.45	3.39	3.33
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.62	3.52	3.42	3.37	3.31	3.26	3.20	3.14	3.08
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53	3.43	3.33	3.23	3.17	3.12	3.06	3.00	2.94	2.88
12	6.56	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.28	3.18	3.07	3.02	2.96	2.91	2.85	2.79	2.72
13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31	3.25	3.15	3.05	2.95	2.89	2.84	2.78	2.72	2.66	2.60
14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21	3.15	3.05	2.95	2.84	2.79	2.73	2.67	2.61	2.55	2.49
15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.96	2.86	2.76	2.70	2.64	2.59	2.52	2.46	2.40
16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05	2.99	2.89	2.79	2.68	2.63	2.57	2.51	2.45	2.38	2.32
17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98	2.92	2.82	2.72	2.62	2.56	2.50	2.44	2.38	2.32	2.25
18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93	2.87	2.77	2.67	2.56	2.50	2.44	2.38	2.32	2.26	2.19
19	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88	2.82	2.72	2.62	2.51	2.45	2.39	2.33	2.27	2.20	2.13
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.68	2.57	2.46	2.41	2.35	2.29	2.22	2.16	2.09
21	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80	2.73	2.64	2.53	2.42	2.37	2.31	2.25	2.18	2.11	2.04
22	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76	2.70	2.60	2.50	2.39	2.33	2.27	2.21	2.14	2.08	2.00
23	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73	2.67	2.57	2.47	2.36	2.30	2.24	2.18	2.11	2.04	1.97
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70	2.64	2.54	2.44	2.33	2.27	2.21	2.15	2.08	2.01	1.94
25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68	2.61	2.51	2.41	2.30	2.24	2.18	2.12	2.05	1.98	1.91
26	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65	2.59	2.49	2.39	2.28	2.22	2.16	2.09	2.03	1.95	1.88
27	5.63	4.24	3.65	3.31	3.08	2.92	2.80	2.71	2.63	2.57	2.47	2.36	2.25	2.19	2.13	2.07	2.00	1.93	1.85
28	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61	2.55	2.45	2.34	2.23	2.17	2.11	2.05	1.98	1.91	1.83
29	5.59	4.20	3.61	3.27	3.04	2.88	2.76	2.67	2.59	2.53	2.43	2.32	2.21	2.15	2.09	2.03	1.96	1.89	1.81
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.41	2.31	2.20	2.14	2.07	2.01	1.94	1.87	1.79
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39	2.29	2.18	2.07	2.01	1.94	1.88	1.80	1.72	1.64
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.17	2.06	1.94	1.88	1.82	1.74	1.67	1.58	1.48
120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16	2.05	1.94	1.82	1.76	1.69	1.61	1.53	1.43	1.31
$\infty$	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05	1.94	1.83	1.71	1.64	1.57	1.48	1.39	1.27	1.00

SOURCE: Pearson and Hartley [1958]. Reprinted from *Biometrika Tables for Statisticians*, Vol. I, with the permission of the Biometrika Trustees.



Percentage Points of the  $F$  Distribution (*continued*)  
Upper 5% Points.

$\frac{v_1}{v_2}$	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	$\infty$
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.48	4.43	4.40	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.43	3.38	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	4.23	3.37	2.97	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.76	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.36	1.26
$\infty$	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

SOURCE: Pearson and Hartley [1958]. Reprinted from *Biometrika Tables for Statisticians*, Vol. I, with the permission of the Biometrika Trustees.

# Acknowledgments

My editors, Steve Fienberg and Ingram Olkin, are to be thanked for not only their useful comments on earlier drafts of this book, but also for their encouragement in seeing that this enterprise reached its conclusion. Michael Cohen and Diane Lambert contributed immeasurably with their detailed critical reading of an earlier draft of this book. The extent to which they do not recognize this version is almost wholly attributable to my response to their comments. Chapter 8 benefitted from the extensive comments of the “four most” experts in the field, Steve Fienberg, Zvi Gilula, Leo Goodman, and Shelby Haberman. Evelyn Shropshire typed each of the drafts of this book, always with a smile and a willingness to push on and get it out, and this made my job so much easier. Finally, I would like to express my appreciation to each of my students who suffered through the various versions of this book and my effort to “debug” it.

*Chicago, Illinois*  
*June 1987*

A. M.