BLSS — THE BERKELEY INTERACTIVE STATISTICAL SYSTEM: AN OVERVIEW

D. Mark Abrahams
and
Fran Rizzardi

Technical Report No. 176
September 1988

Department of Statistics
University of California
Berkeley, California

BLSS — THE BERKELEY INTERACTIVE STATISTICAL SYSTEM: AN OVERVIEW

D. Mark Abrahams and Fran Rizzardi

Department of Statistics, University of California, Berkeley, CA 94720

Revised 27 September 1988

1. Introduction

BLSS (pronounced *bliss*) is a highly interactive statistics software system which runs on UNIX-based computers and workstations. It has grown out of the instructional needs of our department: We needed a single system which is both easy enough for students in freshman statistics courses (and first-year TA's) to learn and to use, but flexible and powerful enough for use in advanced undergraduate and beginning graduate classes. This paper is a brief overview of the software itself and its use in the Berkeley instructional program. For an extensive description of the software see Abrahams and Rizzardi [1988].

2. Instructional Philosophy

Why use statistical software in instruction? To some people the answer is obvious; but to some it may seem to be a passing fad. We see two main purposes, and they apply to all statistics courses at *all* levels — even our large introductory one-semester courses.

1). To help develop *statistical* intuition. Students should use real statistical techniques on real data — including large datasets (hundreds of observations). This develops their intuition about the art and science of analyzing data in a way which paper-and-pencil textbook problems cannot.

2). To help develop intuition for the *probability* and *mathematics* behind the statistics. Students should try out and experiment with statistical techniques, probability simulations, etc., both as a whole and — especially at the intermediate and advanced levels — in terms of their individual mathematical building blocks. This is the antidote to thinking of the techniques as 'black boxes'.

Developing intuition is crucial to understanding.

Of course, students should be able do all this without getting bogged down in the details of computer use!

3. User Interface and Datasets

BLSS provides a *simple, command-oriented* user interface which also recognizes *assignments* and algebraic *expressions*. For example, the command:

```
. regress x y
```

forms the regression of y on x, and the command:

```
. help regress
```

shows the on-line documentation for the **regress** command. The command:

```
. a = sqrt(b^2 + c^2)
```

assigns to a the value of the expression $\sqrt{b^2+c^2}$, and the command:

```
. s = t[4,5]
```

assigns to s the value of element [4,5] of the matrix t. (The '.' in these examples is the prompt character which BLSS types when it is ready to accept commands.)

Datasets (the objects a, b, c, s, t, x, and y in the preceding examples) are stored as familiar objects: they can be scalars, row vectors, column vectors, matrices, or three-way arrays. The work area can contain as many datasets as necessary, each with its own name, shape, and dimensions. The **list** command makes a list of datasets in the work area and describes their shape. For example, it might tell us:

```
. list

Contents of your work area:
a               dataset, dims=(1,7)      (row vector)
b               dataset, dims=(1,7)      (row vector)
c               dataset, dims=(1,7)      (row vector)
ozone           dataset, dims=(366,13)   (matrix)
plover          dataset, dims=(42,5)     (matrix)
s               dataset, dims=(1,1)      (scalar)
t               dataset, dims=(10,10)    (matrix)
x               dataset, dims=(24,1)     (column vector)
y               dataset, dims=(24,1)     (column vector)
```

In this example, and throughout this paper, we use *`slanted monospace`* font to show what the user types and `regular monospace` font to show what BLSS types.

4. Elementary Statistics

Elementary courses use BLSS for basic statistics commands such as descriptive statistics, stem-and-leaf diagrams, scatterplots, regression, etc. The following three commands, for example, fetch a dataset (of final examination scores) from the BLSS data library, show descriptive statistics for it, and make a stem-and-leaf diagram:

```
. load finalexam
Loaded "finalexam" from the BLSS system data area.

. stat finalexam

Statistics: finalexam
Col   N    Mean     SD      Min      25%      50%      75%      Max
 1   62   57.68   20.49   6.000   45.00    60.00    73.00    92.00

. stemleaf finalexam

N = 62,  min = 6,  25% = 45,  50% = 60,  75% = 73,  max = 92
Leaf digit unit (ldu) = 1  (1|2 represents 12.)

 0|69
 1|29
 2|99
 3|235699
 4|01455566
 5|02245678889
 6|11234556789
 7|0113356789
 8|00223346
 9|12
```

For these simple univariate statistics, the command forms are no more complex than:

command dataset

Subscripts can be used to specify columns, rows, and elements within a dataset. As a special case, single subscripts can be used to denote entire columns. For example, the command:

```
. scat plover[1] plover[2]
```

makes a scatterplot of the first column of the *plover* dataset (as the X variable) against the second (as the Y variable). Alternatively, *options* (which are enclosed in curly brackets {}) can be used to specify columns. Results of commands can be placed in *output datasets*. For example, the command:

```
. regress plover {x=1,2} {y=3} > y.hat resids coefs
```

forms the regression of column 3 of *plover* on columns 1 and 2. The command displays the results:

```
Dependent variable:      plover[3]
Independent variables:   plover[1 2]
Observations  42         Parameters  3

Parameter    Estimate      SE         t-Ratio     P-Value
intercept   -14.264      0.91567    -15.5779      0.0000
 coef 1       0.23824    0.017456    13.6480      0.0000
 coef 2       0.67424    0.043260    15.5859      0.0000

Residual SD  0.11289     Residual Variance   0.012744
Multiple R   0.97397     Multiple R-squared  0.94861
```

and it creates output datasets named *y.hat*, *resids*, and *coefs* which contain the fitted values, regression residuals, and estimated intercept and coefficients. (Of course, many other options and outputs are available for this command, but we do not discuss them here.)

The '>' symbol is called the *output dataset separator;* everything to its right is an output dataset. Think of it as an arrow. The general form is:

command input-datasets $\rightarrow$ *output-datasets*

In general, every number which is printed in an *output display* (such as in the **regress** display above) is also available as an *output dataset* should the user want it.

5. Elementary Probability and Sampling

BLSS provides several commands which illustrate concepts of probability and sampling in an elementary context. For example, students sometimes find the definition of a confidence interval confusing. The **confid** command illustrates the definition via simulation. Unless the information is provided (via input datasets and options on the command line), **confid** asks the student for the number and frequency of tickets to put in a box (using the vocabulary of Freedman, Pisani, and Purves [1978]), the number of draws per sample, and the number of samples:

```
. confid
Either:  enter the numbers on the tickets in the box;
    Or:  on separate lines, enter each ticket's number and frequency.
Finish with RETURN and CTRL-D.
5 2 4 8 6
(Control-D)
Box saved in:  box00
Number of draws per sample? 100
Number of samples? 10
```

```
10 samples of 100 draws each are made from box00.
Population mean is 5.  Population SD is 2.

 Sample Mean   Estimated SE               95% Confidence Interval
    4.8900        0.1974                  |-----------*----------|
    5.2000        0.2030                           |-----------*-----------|
    5.1000        0.1915                         |----------*----------|
    4.6200        0.1900            |-----------*----------|
    4.8000        0.1803                  |----------*----------|
    5.3500        0.2134                                |------------*------------|
    5.0000        0.2151                    |------------*------------|
    5.1300        0.1745                         |---------*---------|
    5.5300        0.2144                                     |------------*------->
    5.1000        0.1941                         |----------*----------|

 True mean + and - 4 SEs:    ........................x.........................

8 out of 10 confidence intervals (or 80.00%) covered the population mean.
```

BLSS provides several other such simulations. A coin-tossing simulation asks the student whether to use a fair coin or not, how many tosses to make, etc.; it keeps track of the cumulative number and proportion of heads, the difference from the expected value, etc. Another box-model simulation illustrates the sampling variation of the estimated average and standard deviation: it is similar to the **confid** simulation shown above, but instead displays summary statistics for each sample. (A separate **sample** command allows samples to be drawn — with or without replacement — from a real dataset.)

The **demoiv** command (named after De Moivre) provides an elementary illustration of the central limit theorem: it displays the scaled probability histogram for one observation from the box, and for the sum of 2, 4, 8, 16, . . . draws. Students find **demoiv** particularly entertaining — as they try (unsuccessfully) to design boxes which will defeat the central limit theorem.

6. Intermediate Use

BLSS provides many commands for specific statistical techniques, including: frequency counts, cross-tabulation, confidence intervals, hypothesis tests (t-tests, z-tests, χ^2-tests), probability functions, random number generation, matrix decompositions, time series analysis, etc.

However, its real power for instruction at intermediate and advanced levels comes from being able to assemble individual commands into larger sequences, combined with a number of useful matrix and array operators. We illustrate with a standard example from an intermediate level course — generating a random vector **x** from a multivariate Normal distribution with mean vector μ and covariance matrix Σ using the Cholesky decomposition: $\mathbf{x} = \mathbf{zT} + \mu$, where $\mathbf{T}'\,\mathbf{T} = \Sigma$ and **x**, **z**, and μ are row vectors, **z** of independent standard Normals. To create the vector *mu*, give the command:

```
. mu = 5, -2, 0
```

The comma operator '`,`' used here catenates the numbers together into a row vector (it can also catenate together column vectors and matrices which have the same number of rows); this is one way to create a new dataset. Another way to create a dataset is to read it in with the **read** command; we use this to enter the covariance matrix *sigma*:

```
. read > sigma
```

```
Type your data one row per line; finish with RETURN and CTRL-D.
 9  -6   3
-6  29  -2
 3  -2   5
(Control-D)
Read 9 values into "sigma"; dims=(3,3) (matrix).
```

(The **read** command automatically infers the dimensions of the new dataset from how it was typed in.) To compute the Cholesky square root of *sigma* and place it in *t* we use the **chol** command:

```
. chol sigma > t
```

To show *t*, simply type its name:

```
. t
    3.000    -2.000    1.000
    0.000     5.000    0.000
    0.000     0.000    2.000
```

To check that the Cholesky decomposition is correct we can look at $\Sigma - \mathbf{T}'\mathbf{T}$:

```
. sigma - t'#*t
    0.000     0.000    0.000
    0.000     0.000    0.000
    0.000     0.000    0.000
```

(The ' operator is matrix transposition, and the #* operator is matrix multiplication.)

Instead of generating a single random vector, we generate 100 simultaneously. We start by using the **rgau** command ('r' for random and 'gau' for Gaussian) to generate a matrix *z* of dimensions (100,3) containing random standard Normals:

```
. rgau {dims=100,3} > z
```

To obtain our 100 random 3-vectors as a matrix *x:*

```
. x = z#*t + mu
```

Note that the addition operation here added a 100-by-3 matrix to a length 3-row vector, and yielded a 100-by-3 matrix. The row vector *mu* was conceptually 'expanded' into a matrix of the appropriate dimension; that is, it set $x_{ij} = [\mathbf{zT}]_{ij} + \mu_j$ for all i and j. BLSS provides this 'dimension-expansion' feature for all elementwise binary operators. It is useful for many array operations, such as centering and scaling columns or rows of a matrix, etc.

Because intermediate results can be examined and variations can be experimented with, assembling sequences of commands to perform more elaborate operations is an excellent way for students to build intuition.

7. Convenience Features

BLSS provides a number of convenience features which make life easier.

Aliases allow you to create abbreviations for commands and command-and-option combinations. As a trivial example, if the word 'regress' is too long to type, the command:

```
. alias reg regress
```

creates an alias name 'reg' for it. Thereafter, whenever you type the command **reg**, it is just as if you had typed the command **regress**; the new command **reg** can accept any of the inputs, options, or outputs of **regress**. More usefully, aliases can include inputs and options. For example, the **stat** command (which shows descriptive statistics) normally computes variances and standard deviations using the divide-by-(N–1) formula. The option **{dn=0}** makes it use the divide-by-N formula instead. To always use this option, give the command:

```
. alias stat stat {dn=0}
```

(This alias value contains two words, 'stat' and '{dn=0}'. In general, alias values can contain any number of words.) Most BLSS commands provide options which change their default behavior or output displays. Thus, aliases can be used to customize BLSS commands and displays to one's own taste.

Strings are similar to aliases, except that they allow for abbreviations to be placed in the middle of the command line as well as the beginning of it. Thus, they can provide abbreviations for long sets of options, long dataset names, etc.

BLSS keeps a record of commands typed in the current session; this is known as the *command history*. Commands from the history can be displayed, repeated, and edited (for example, to change an option or parameter value without re-typing the entire command).

Command files (or *macros*) can be written which combine existing commands into new commands; the new commands are invoked in the same way (using the same syntax, etc.) as existing commands. A convenient way to construct command files is to edit the record of commands in your history after you have given some set of commands that you want to repeat.

Aliases, strings, and command files — as well as datasets, on-line help files, and a startup message — can be installed on a class-wide basis. Thus, the instructor can provide a customized environment for all students in the class. For example, if the instructor is using a textbook which defines variances and standard deviations using the divide-by-N formula, he can install the alias shown above, so that the students automatically get the same results as in the textbook.

8. Software Interfacing

BLSS provides software interfaces to the operating system (UNIX) at several levels.

For common tasks — such as re-routing text output displays from the terminal screen to a disk file, sending files to a lineprinter (or laserprinter), and editing files (using the editor of one's choice) — BLSS has builtin interfaces to the appropriate UNIX software; to perform these tasks it is unnecessary to leave BLSS and enter UNIX. (This also means that students need not learn any UNIX commands; they can 'live' within BLSS.)

In addition, BLSS and UNIX can be interfaced at the command level: UNIX commands can be called directly from within BLSS, and vice versa. This allows for rapid integration of BLSS with other application software.

Finally, BLSS can be interfaced at the subroutine level — with UNIX routines, or with the user's own subroutines. In fact, BLSS commands are nothing other than UNIX commands (written in either C or Fortran) which call the BLSS subroutine library. The result is that — although we have been emphasizing its instructional uses — BLSS provides a ready-made frontend for running one's own Fortran (or C) subroutines.

Here is a simple example of a complete Fortran program which does so. It takes one input dataset and creates one output dataset.

```
c     Very simple Fortran calling program
c         to a user-supplied Fortran subroutine.
c     One input dataset (x) and one output dataset (y).
c     Minimal BLSS library support:
c         No option handling.  No dynamic memory allocation.
      real x(10000), y(10000)
c     ... Open the input dataset ...
      ix = iopeni("x", "-ERR-")
c     ... Find out its number of rows and columns ...
      nr = inr(ix)
      nc = inc(ix)
c     ... Check dataset size against allocated memory ...
      if (nr*nc .gt. 10000)
     +   call error("Dimensions cannot exceed nr*nc > 10000.")
c     ... Open the output dataset ...
      iy = iopeno("y", "-ERR-", 1, nr, nc)
c     ... Read the input data matrix ...
      call mread(ix, x, nr, nc, "", 0)
c     ... Check for no NA's (missing values) ...
      call mchk (ix, x, nr, nc, "-ERR-")
c     ... User-supplied subroutine to obtain y from x ...
      call mysubr(x, y, nr, nc)
c     ... Write the output data matrix ...
      call mwrite(iy, y, nr, nc)
c     ... Clean up and exit ...
      call iexit(0)
      end

c     ... This could be any user-supplied subroutine ...
c     ... For the purpose of illustration, we supply a trivial one ...
      subroutine mysubr(x, y, nr, nc)
      real    x(nr,nc), y(nr,nc)
      do 10 j = 1, nc
              do 10 i = 1, nr
                      y(i,j) = -x(i,j)
10    continue
      return
      end
```

The input and output datasets are called *x* and *y* inside the program, but can have any name as far as the user is concerned. The functions *iopeni* and *iopeno* take care of opening the input and output datasets and discovering their names from the command line; the subroutines *mread* and *mwrite* take care of reading and writing data one matrix-worth at a time. Other subroutines and functions (*inr*, *inc*, *mchk*) perform utility tasks such as returning the number of rows and columns in the input dataset, checking to see whether it contains NA's (missing values), etc. The BLSS library provides many more capabilities such as: checking for options, performing dynamic memory allocation for Fortran, a variety of error-condition handling, etc. Thus the calling program can be as simple or as complex as desired, in terms of options, defaults, and so forth.

The calling program can be invoked directly from UNIX as well as from within BLSS — this is helpful if you are in the process of writing your subroutine and want to make a lot of use of the editor, the compiler, and the debugger.

9. Current Status

At U.C. Berkeley, BLSS is used in statistics courses at all levels, from freshman through graduate, as well as in quantitative courses taught by other departments. The following table summarizes its use in courses at Berkeley in the 1987-88 academic year.

	Number of One-Semester Courses	Total Enrollment
Statistics Department		
Freshman/Sophomore level	4	912
Junior/Senior level	13	342
Graduate level	3	134
Other Departments		
Anthropology	2	44
Chemistry	3	164
Electrical Engineering	1	61
Forestry	1	5
TOTAL	27	1662

10. Future Directions

BLSS is under continuing development by the Department; the design and programming effort involves several people. For the immediate future, we are concentrating on:

- Improved graphics. Our plan is to add high-quality, flexible, command-driven graphics capabilities which work with a variety of output devices including Tektronix-compatible graphics terminals, X-windows, and laserprinters, as well as character devices such as lineprinters and regular terminals.

- A full-screen, menu-driven user interface is being written: Some of our instructors believe that menus are preferable to commands for use in elementary level classes.

- More statistics capabilities. Obviously we cannot add everything; so, in order to help set priorities, we made a widespread survey of universities earlier this year. Three new capabilities currently being worked on are: elementary (rank-based) nonparametric methods, analysis of variance for factorial designs, and ARMA and ARIMA routines.

11. Why UNIX?

Why did we choose UNIX as the operating system on which to run BLSS? By limiting ourselves to a single operating system, we can concentrate on developing better software rather than on satisfying the idiosyncrasies of many different operating systems; and from our standpoint, UNIX is the most important operating system for the foreseeable future. It is the only vendor-independent operating system which runs on a wide variety of hardware (from desktop machines to mainframes and even supercomputers), and its networking and resource-sharing capabilities far exceed those of any other we know of. Our department has found that UNIX provides an excellent environment for both research and teaching. In particular, running classes is much easier on UNIX than on PC's, because UNIX eliminates the need for multiple copies of software, datasets, help files, and other resources on each PC, and because it provides easy communication between instructor, teaching assistants, and students.

Three years ago, our department considered switching over from UNIX to a network of PC's or Macintoshes for instruction — but we decided against it based on the strengths of UNIX.

12. Summary

BLSS meets a wide range of instructional needs, from elementary courses for freshman nonmajors through beginning courses for graduate students.

In our view, the purpose of BLSS (or the computer, in general) in statistical instruction is to help students develop *intuition:* for both statistics itself and for the probability and mathematics behind the statistics.

BLSS provides a *command-oriented* user interface which also recognizes *assignments* and *algebraic expressions*.

BLSS stores datasets intuitively (as scalars, row vectors, column vectors, matrices, etc.). The work area can contain as many datasets as needed, each with its own name and shape.

For elementary courses, BLSS provides standard statistics commands and also elementary probability demonstrations and simulations.

For intermediate and advanced courses, BLSS provides many specific statistical techniques — but its real power comes from being able to assemble sequences of commands and low-level matrix and array operations to accomplish more elaborate tasks.

Convenience features allow the user to customize the BLSS environment to his own liking. Course instructors can apply such customizations to an entire class.

The lineprinter, the editor, and UNIX command-level capabilities can be used directly from within BLSS. Commands can be written which interface BLSS and UNIX at the subroutine level.

BLSS enjoys widespread use at U.C. Berkeley. Over 1600 students used it here last year in courses at all levels, both within the Statistics Department and outside it.

References

Abrahams, D. Mark, and Rizzardi, Fran (1988). *BLSS: The Berkeley Interactive Statistical System.* Norton, New York.

Freedman, David, Pisani, Robert, and Purves, Roger (1978). *Statistics.* Norton, New York.

TECHNICAL REPORTS

Statistics Department

University of California, Berkeley

1. BREIMAN, L. and FREEDMAN, D. (Nov. 1981, revised Feb. 1982). How many variables should be entered in a regression equation? Jour. Amer. Statist. Assoc., March 1983, 78, No. 381, 131-136.

2. BRILLINGER, D. R. (Jan. 1982). Some contrasting examples of the time and frequency domain approaches to time series analysis. Time Series Methods in Hydrosciences, (A. H. El-Shaarawi and S. R. Esterby, eds.) Elsevier Scientific Publishing Co., Amsterdam, 1982, pp. 1-15.

3. DOKSUM, K. A. (Jan. 1982). On the performance of estimates in proportional hazard and log-linear models. Survival Analysis, (John Crowley and Richard A. Johnson, eds.) IMS Lecture Notes - Monograph Series, (Shanti S. Gupta, series ed.) 1982, 74-84.

4. BICKEL, P. J. and BREIMAN, L. (Feb. 1982). Sums of functions of nearest neighbor distances, moment bounds, limit theorems and a goodness of fit test. Ann. Prob., Feb. 1982, 11. No. 1, 185-214.

5. BRILLINGER, D. R. and TUKEY, J. W. (March 1982). Spectrum estimation and system identification relying on a Fourier transform. The Collected Works of J. W. Tukey, vol. 2, Wadsworth, 1985, 1001-1141.

6. BERAN, R. (May 1982). Jackknife approximation to bootstrap estimates. Ann. Statist., March 1984, 12 No. 1, 101-118.

7. BICKEL, P. J. and FREEDMAN, D. A. (June 1982). Bootstrapping regression models with many parameters. Lehmann Festschrift, (P. J. Bickel, K. Doksum and J. L. Hodges, Jr., eds.) Wadsworth Press, Belmont, 1983, 28-48.

8. BICKEL, P. J. and COLLINS, J. (March 1982). Minimizing Fisher information over mixtures of distributions. Sankhyā, 1983, 45, Series A, Pt. 1, 1-19.

9. BREIMAN, L. and FRIEDMAN, J. (July 1982). Estimating optimal transformations for multiple regression and correlation.

10. FREEDMAN, D. A. and PETERS, S. (July 1982, revised Aug. 1983). Bootstrapping a regression equation: some empirical results. JASA, 1984, 79, 97-106.

11. EATON, M. L. and FREEDMAN, D. A. (Sept. 1982). A remark on adjusting for covariates in multiple regression.

12. BICKEL, P. J. (April 1982). Minimax estimation of the mean of a mean of a normal distribution subject to doing well at a point. Recent Advances in Statistics, Academic Press, 1983.

14. FREEDMAN, D. A., ROTHENBERG, T. and SUTCH, R. (Oct. 1982). A review of a residential energy end use model.

15. BRILLINGER, D. and PREISLER, H. (Nov. 1982). Maximum likelihood estimation in a latent variable problem. Studies in Econometrics, Time Series, and Multivariate Statistics, (eds. S. Karlin, T. Amemiya, L. A. Goodman). Academic Press, New York, 1983, pp. 31-65.

16. BICKEL, P. J. (Nov. 1982). Robust regression based on infinitesimal neighborhoods. Ann. Statist., Dec. 1984, 12, 1349-1368.

17. DRAPER, D. C. (Feb. 1983). Rank-based robust analysis of linear models. I. Exposition and review.

18. DRAPER, D. C. (Feb 1983). Rank-based robust inference in regression models with several observations per cell.

19. FREEDMAN, D. A. and FIENBERG, S. (Feb. 1983, revised April 1983). Statistics and the scientific method, Comments on and reactions to Freedman, A rejoinder to Fienberg's comments. Springer New York 1985 Cohort Analysis in Social Research, (W. M. Mason and S. E. Fienberg, eds.).

20. FREEDMAN, D. A. and PETERS, S. C. (March 1983, revised Jan. 1984). Using the bootstrap to evaluate forecasting equations. J. of Forecasting. 1985, Vol. 4, 251-262.

21. FREEDMAN, D. A. and PETERS, S. C. (March 1983, revised Aug. 1983). Bootstrapping an econometric model: some empirical results. JBES, 1985, 2, 150-158.

22. FREEDMAN, D. A. (March 1983). Structural-equation models: a case study.

23. DAGGETT, R. S. and FREEDMAN, D. (April 1983, revised Sept. 1983). Econometrics and the law: a case study in the proof of antitrust damages. Proc. of the Berkeley Conference, in honor of Jerzy Neyman and Jack Kiefer. Vol I pp. 123-172. (L. Le Cam, R. Olshen eds.) Wadsworth, 1985.

24. DOKSUM, K. and YANDELL, B. (April 1983). Tests for exponentiality. Handbook of Statistics, (P. R. Krishnaiah and P. K. Sen, eds.) 4, 1984.

25. FREEDMAN, D. A. (May 1983). Comments on a paper by Markus.

26. FREEDMAN, D. (Oct. 1983, revised March 1984). On bootstrapping two-stage least-squares estimates in stationary linear models. Ann. Statist., 1984, 12, 827-842.

27. DOKSUM, K. A. (Dec. 1983). An extension of partial likelihood methods for proportional hazard models to general transformation models. Ann. Statist., 1987, 15, 325-345.

28. BICKEL, P. J., GOETZE, F. and VAN ZWET, W. R. (Jan. 1984). A simple analysis of third order efficiency of estimate Proc. of the Neyman-Kiefer Conference, (L. Le Cam, ed.) Wadsworth, 1985.

29. BICKEL, P. J. and FREEDMAN, D. A. Asymptotic normality and the bootstrap in stratified sampling. Ann. Statist. 12 470-482.

30. FREEDMAN, D. A. (Jan. 1984). The mean vs. the median: a case study in 4-R Act litigation. JBES. 1985 Vol 3 pp. 1-13.

31. STONE, C. J. (Feb. 1984). An asymptotically optimal window selection rule for kernel density estimates. Ann. Statist., Dec. 1984, 12, 1285-1297.

32. BREIMAN, L. (May 1984). Nail finders, edifices, and Oz.

33. STONE, C. J. (Oct. 1984). Additive regression and other nonparametric models. Ann. Statist., 1985, 13, 689-705.

34. STONE, C. J. (June 1984). An asymptotically optimal histogram selection rule. Proc. of the Berkeley Conf. in Honor of Jerzy Neyman and Jack Kiefer (L. Le Cam and R. A. Olshen, eds.), II, 513-520.

35. FREEDMAN, D. A. and NAVIDI, W. C. (Sept. 1984, revised Jan. 1985). Regression models for adjusting the 1980 Census. Statistical Science. Feb 1986, Vol. 1, No. 1, 3-39.

36. FREEDMAN, D. A. (Sept. 1984, revised Nov. 1984). De Finetti's theorem in continuous time.

37. DIACONIS, P. and FREEDMAN, D. (Oct. 1984). An elementary proof of Stirling's formula. Amer. Math Monthly. Feb 1986, Vol. 93, No. 2, 123-125.

38. LE CAM, L. (Nov. 1984). Sur l'approximation de familles de mesures par des familles Gaussiennes. Ann. Inst. Henri Poincaré, 1985, 21, 225-287.

39. DIACONIS, P. and FREEDMAN, D. A. (Nov. 1984). A note on weak star uniformities.

40. BREIMAN, L. and IHAKA, R. (Dec. 1984). Nonlinear discriminant analysis via SCALING and ACE.

41. STONE, C. J. (Jan. 1985). The dimensionality reduction principle for generalized additive models.

42. LE CAM, L. (Jan. 1985). On the normal approximation for sums of independent variables.

43. BICKEL, P. J. and YAHAV, J. A. (1985). On estimating the number of unseen species: how many executions were there?

44. BRILLINGER, D. R. (1985). The natural variability of vital rates and associated statistics. Biometrics, to appear.

45. BRILLINGER, D. R. (1985). Fourier inference: some methods for the analysis of array and nonGaussian series data. Water Resources Bulletin, 1985, 21, 743-756.

46. BREIMAN, L. and STONE, C. J. (1985). Broad spectrum estimates and confidence intervals for tail quantiles.

47. DABROWSKA, D. M. and DOKSUM, K. A. (1985, revised March 1987). Partial likelihood in transformation models with censored data.

48. HAYCOCK, K. A. and BRILLINGER, D. R. (November 1985). LIBDRB: A subroutine library for elementary time series analysis.

49. BRILLINGER, D. R. (October 1985). Fitting cosines: some procedures and some physical examples. Joshi Festschrift, 1986. D. Reidel.

50. BRILLINGER, D. R. (November 1985). What do seismology and neurophysiology have in common? - Statistics! Comptes Rendus Math. Rep. Acad. Sci. Canada. January, 1986.

51. COX, D. D. and O'SULLIVAN, F. (October 1985). Analysis of penalized likelihood-type estimators with application to generalized smoothing in Sobolev Spaces.

52. O'SULLIVAN, F. (November 1985). A practical perspective on ill-posed inverse problems: A review with some new developments. To appear in Journal of Statistical Science.

53. LE CAM, L. and YANG, G. L. (November 1985, revised March 1987). On the preservation of local asymptotic normality under information loss.

54. BLACKWELL, D. (November 1985). Approximate normality of large products.

55. FREEDMAN, D. A. (June 1987). As others see us: A case study in path analysis. Journal of Educational Statistics. 12, 101-128.

56. LE CAM, L. and YANG, G. L. (January 1986). Replaced by No. 68.

57. LE CAM, L. (February 1986). On the Bernstein - von Mises theorem.

58. O'SULLIVAN, F. (January 1986). Estimation of Densities and Hazards by the Method of Penalized likelihood.

59. ALDOUS, D. and DIACONIS, P. (February 1986). Strong Uniform Times and Finite Random Walks.

60. ALDOUS, D. (March 1986). On the Markov Chain simulation Method for Uniform Combinatorial Distributions and Simulated Annealing.

61. CHENG, C-S. (April 1986). An Optimization Problem with Applications to Optimal Design Theory.

62. CHENG, C-S., MAJUMDAR, D., STUFKEN, J. & TURE, T. E. (May 1986, revised Jan 1987). Optimal step type design for comparing test treatments with a control.

63. CHENG, C-S. (May 1986, revised Jan. 1987). An Application of the Kiefer-Wolfowitz Equivalence Theorem.

64. O'SULLIVAN, F. (May 1986). Nonparametric Estimation in the Cox Proportional Hazards Model.

65. ALDOUS, D. (JUNE 1986). Finite-Time Implications of Relaxation Times for Stochastically Monotone Processes.

66. PITMAN, J. (JULY 1986, revised November 1986). Stationary Excursions.

67. DABROWSKA, D. and DOKSUM, K. (July 1986, revised November 1986). Estimates and confidence intervals for median and mean life in the proportional hazard model with censored data.

68. LE CAM, L. and YANG, G.L. (July 1986). Distinguished Statistics, Loss of information and a theorem of Robert B. Davies (Fourth edition).

69. STONE, C.J. (July 1986). Asymptotic properties of logspline density estimation.

71. BICKEL, P.J. and YAHAV, J.A. (July 1986). Richardson Extrapolation and the Bootstrap.

72. LEHMANN, E.L. (July 1986). Statistics - an overview.

73. STONE, C.J. (August 1986). A nonparametric framework for statistical modelling.

74. BIANE, PH. and YOR, M. (August 1986). A relation between Lévy's stochastic area formula, Legendre polynomial, and some continued fractions of Gauss.

75. LEHMANN, E.L. (August 1986, revised July 1987). Comparing Location Experiments.

76. O'SULLIVAN, F. (September 1986). Relative risk estimation.

77. O'SULLIVAN, F. (September 1986). Deconvolution of episodic hormone data.

78. PITMAN, J. & YOR, M. (September 1987). Further asymptotic laws of planar Brownian motion.

79. FREEDMAN, D.A. & ZEISEL, H. (November 1986). From mouse to man: The quantitative assessment of cancer risks. To appear in Statistical Science.

80. BRILLINGER, D.R. (October 1986). Maximum likelihood analysis of spike trains of interacting nerve cells.

81. DABROWSKA, D.M. (November 1986). Nonparametric regression with censored survival time data.

82. DOKSUM, K.J. and LO, A.Y. (November 1986). Consistent and robust Bayes Procedures for Location based on Partial Information.

83. DABROWSKA, D.M., DOKSUM, K.A. and MIURA, R. (November 1986). Rank estimates in a class of semiparametric two-sample models.

84. BRILLINGER, D. (December 1986). Some statistical methods for random process data from seismology and neurophysiology.

85. DIACONIS, P. and FREEDMAN, D. (December 1986). A dozen de Finetti-style results in search of a theory. Ann. Inst. Henri Poincaré, 1987, 23, 397-423.

86. DABROWSKA, D.M. (January 1987). Uniform consistency of nearest neighbour and kernel conditional Kaplan - Meier estimates.

87. FREEDMAN, D.A., NAVIDI, W. and PETERS, S.C. (February 1987). On the impact of variable selection in fitting regression equations.

88. ALDOUS, D. (February 1987, revised April 1987). Hashing with linear probing, under non-uniform probabilities.

89. DABROWSKA, D.M. and DOKSUM, K.A. (March 1987, revised January 1988). Estimating and testing in a two sample generalized odds rate model.

90. DABROWSKA, D.M. (March 1987). Rank tests for matched pair experiments with censored data.

91. DIACONIS, P and FREEDMAN, D.A. (April 1988). Conditional limit theorems for exponential families and finite versions of de Finetti's theorem. To appear in the Journal of Applied Probability.

92. DABROWSKA, D.M. (April 1987, revised September 1987). Kaplan-Meier estimate on the plane.

92a. ALDOUS, D. (April 1987). The Harmonic mean formula for probabilities of Unions: Applications to sparse random graphs.

93. DABROWSKA, D.M. (June 1987, revised Feb 1988). Nonparametric quantile regression with censored data.

94. DONOHO, D.L. & STARK, P.B. (June 1987). Uncertainty principles and signal recovery.

95. CANCELLED

96. BRILLINGER, D.R. (June 1987). Some examples of the statistical analysis of seismological data. To appear in *Proceedings, Centennial Anniversary Symposium, Seismographic Stations, University of California, Berkeley.*

97. FREEDMAN, D.A. and NAVIDI, W. (June 1987). On the multi-stage model for carcinogenesis. To appear in Environmental Health Perspectives.

98. O'SULLIVAN, F. and WONG, T. (June 1987). Determining a function diffusion coefficient in the heat equation.

99. O'SULLIVAN, F. (June 1987). Constrained non-linear regularization with application to some system identification problems.

100. LE CAM, L. (July 1987, revised Nov 1987). On the standard asymptotic confidence ellipsoids of Wald.

101. DONOHO, D.L. and LIU, R.C. (July 1987). Pathologies of some minimum distance estimators. Annals of Statistics, June, 1988.

102. BRILLINGER, D.R., DOWNING, K.H. and GLAESER, R.M. (July 1987). Some statistical aspects of low-dose electron imaging of crystals.

103. LE CAM, L. (August 1987). Harald Cramér and sums of independent random variables.

104. DONOHO, A.W., DONOHO, D.L. and GASKO, M. (August 1987). Macspin: Dynamic graphics on a desktop computer. IEEE Computer Graphics and applications, June, 1988.

105. DONOHO, D.L. and LIU, R.C. (August 1987). On minimax estimation of linear functionals.

106. DABROWSKA, D.M. (August 1987). Kaplan-Meier estimate on the plane: weak convergence, LIL and the bootstrap.

107. CHENG, C-S. (August 1987). Some orthogonal main-effect plans for asymmetrical factorials.

108. CHENG, C-S. and JACROUX, M. (August 1987). On the construction of trend-free run orders of two-level factorial designs.

109. KLASS, M.J. (August 1987). Maximizing $E \max_{1\le k\le n} S_k^+/ES_n^+$: A prophet inequality for sums of I.I.D. mean zero variates.

110. DONOHO, D.L. and LIU, R.C. (August 1987). The "automatic" robustness of minimum distance functionals. Annals of Statistics, June, 1988.

111. BICKEL, P.J. and GHOSH, J.K. (August 1987, revised June 1988). A decomposition for the likelihood ratio statistic and the Bartlett correction — a Bayesian argument.

112. BURDZY, K., PITMAN, J.W. and YOR, M. (September 1987). Some asymptotic laws for crossings and excursions.

113. ADHIKARI, A. and PITMAN, J. (September 1987). The shortest planar arc of width 1.

114. RITOV, Y. (September 1987). Estimation in a linear regression model with censored data.

115. BICKEL, P.J. and RITOV, Y. (Sept. 1987, revised Aug 1988). Large sample theory of estimation in biased sampling regression models I.

116. RITOV, Y. and BICKEL, P.J. (Sept.1987, revised Aug. 1988). Achieving information bounds in non and semiparametric models.

117. RITOV, Y. (October 1987). On the convergence of a maximal correlation algorithm with alternating projections.

118. ALDOUS, D.J. (October 1987). Meeting times for independent Markov chains.

119. HESSE, C.H. (October 1987). An asymptotic expansion for the mean of the passage-time distribution of integrated Brownian Motion.

120. DONOHO, D. and LIU, R. (October 1987, revised March 1988). Geometrizing rates of convergence, II.

121. BRILLINGER, D.R. (October 1987). Estimating the chances of large earthquakes by radiocarbon dating and statistical modelling. To appear in *Statistics a Guide to the Unknown.*

122. ALDOUS, D., FLANNERY, B. and PALACIOS, J.L. (November 1987). Two applications of urn processes: The fringe analysis of search trees and the simulation of quasi-stationary distributions of Markov chains.

123. DONOHO, D.L., MACGIBBON, B. and LIU, R.C. (Nov.1987, revised July 1988). Minimax risk for hyperrectangles.

124. ALDOUS, D. (November 1987). Stopping times and tightness II.

125. HESSE, C.H. (November 1987). The present state of a stochastic model for sedimentation.

126. DALANG, R.C. (December 1987, revised June 1988). Optimal stopping of two-parameter processes on nonstandard probability spaces.

127. Same as No. 133.

128. DONOHO, D. and GASKO, M. (December 1987). Multivariate generalizations of the median and trimmed mean II.

129. SMITH, D.L. (December 1987). Exponential bounds in Vapnik-Červonenkis classes of index 1.

130. STONE, C.J. (Nov.1987, revised Sept. 1988). Uniform error bounds involving logspline models.

131. Same as No. 140

132. HESSE, C.H. (December 1987). A Bahadur - Type representation for empirical quantiles of a large class of stationary, possibly infinite - variance, linear processes

133. DONOHO, D.L. and GASKO, M. (December 1987). Multivariate generalizations of the median and trimmed mean, I.

134. DUBINS, L.E. and SCHWARZ, G. (December 1987). A sharp inequality for martingales and stopping-times.

135. FREEDMAN, D.A. and NAVIDI, W. (December 1987). On the risk of lung cancer for ex-smokers.

136. LE CAM, L. (January 1988). On some stochastic models of the effects of radiation on cell survival.

137. DIACONIS, P. and FREEDMAN, D.A. (April 1988). On the uniform consistency of Bayes estimates for multinomial probabilities.

137a. DONOHO, D.L. and LIU, R.C. (1987). Geometrizing rates of convergence, I.

138. DONOHO, D.L. and LIU, R.C. (January 1988). Geometrizing rates of convergence, III.

139. BERAN, R. (January 1988). Refining simultaneous confidence sets.

140. HESSE, C.H. (December 1987). Numerical and statistical aspects of neural networks.

141. BRILLINGER, D.R. (January 1988). Two reports on trend analysis: a) An Elementary Trend Analysis of Rio Negro Levels at Manaus, 1903-1985 b) Consistent Detection of a Monotonic Trend Superposed on a Stationary Time Series

142. DONOHO, D.L. (Jan. 1985, revised Jan. 1988). One-sided inference about functionals of a density.

143. DALANG, R.C. (February 1988). Randomization in the two-armed bandit problem.

144. DABROWSKA, D.M., DOKSUM, K.A. and SONG, J.K. (February 1988). Graphical comparisons of cumulative hazards for two populations.

145. ALDOUS, D.J. (February 1988). Lower bounds for covering times for reversible Markov Chains and random walks on graphs.

146. BICKEL, P.J. and RITOV, Y. (Feb.1988, revised August 1988). Estimating integrated squared density derivatives.

147. STARK, P.B. (March 1988). Strict bounds and applications.

148. DONOHO, D.L. and STARK, P.B. (March 1988). Rearrangements and smoothing.

149. NOLAN, D. (March 1988). Asymptotics for a multivariate location estimator.

150. SEILLIER, F. (March 1988). Sequential probability forecasts and the probability integral transform.

151. NOLAN, D. (March 1988). Limit theorems for a random convex set.

152. DIACONIS, P. and FREEDMAN, D.A. (April 1988). On a theorem of Kuchler and Lauritzen.

153. DIACONIS, P. and FREEDMAN, D.A. (April 1988). On the problem of types.

154. DOKSUM, K.A. (May 1988). On the correspondence between models in binary regression analysis and survival analysis.

155. LEHMANN, E.L. (May 1988). Jerzy Neyman, 1894-1981.

156. ALDOUS, D.J. (May 1988). Stein's method in a two-dimensional coverage problem.

157. FAN, J. (June 1988). On the optimal rates of convergence for nonparametric deconvolution problem.

158. DABROWSKA, D. (June 1988). Signed-rank tests for censored matched pairs.

159. BERAN, R.J. and MILLAR, P.W. (June 1988). Multivariate symmetry models.

160. BERAN, R.J. and MILLAR, P.W. (June 1988). Tests of fit for logistic models.

161. BREIMAN, L. and PETERS, S. (June 1988). Comparing automatic bivariate smoothers (A public service enterprise).

162. FAN, J. (June 1988). Optimal global rates of convergence for nonparametric deconvolution problem.

163. DIACONIS, P. and FREEDMAN, D.A. (June 1988). A singular measure which is locally uniform.

164. BICKEL, P.J. and KRIEGER, A.M. (July 1988). Confidence bands for a distribution function using the bootstrap.

165. HESSE, C.H. (July 1988). New methods in the analysis of economic time series I.

166. FAN, JIANQING (July 1988). Nonparametric estimation of quadratic functionals in Gaussian white noise.

167. BREIMAN, L., STONE, C.J. and KOOPERBERG, C. (August 1988). Confidence bounds for extreme quantiles.

168. LE CAM, L. (August 1988). Maximum likelihood an introduction.

169. BREIMAN, L. (August 1988). Submodel selection and evaluation in regression-The conditional case and little bootstrap.

170. LE CAM, L. (September 1988). On the Prokhorov distance between the empirical process and the associated Gaussian bridge.

171. STONE, C.J. (September 1988). Large-sample inference for logspline models.

172. ADLER, R.J. and EPSTEIN, R. (September 1988). Intersection local times for infinite systems of planar brownian motions and for the brownian density process.

173. MILLAR, P.W. (October 1988). Optimal estimation in the non-parametric multiplicative intensity model.

174. YOR, M. (October 1988). Interwinings of Bessel processes.

175. ROJO, J. (October 1988). On the concept of tail-heaviness.

176. ABRAHAMS, D.M. and RIZZARDI, F. (October 1988). The Berkeley interactive statistical system: An overview.

Copies of these Reports plus the most recent additions to the Technical Report series are available from the Statistics Department technical typist in room 379 Evans Hall or may be requested by mail from:

Department of Statistics
University of California
Berkeley, California 94720

Cost: $1 per copy.