BIΩSTAT, A Computer Program Providing Simple Statistics For Biological Samples

J. M. Hillerud
Midland Lutheran College

Follow this and additional works at: https://digitalcommons.unl.edu/tnas


This Article is brought to you for free and open access by the Nebraska Academy of Sciences at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Transactions of the Nebraska Academy of Sciences and Affiliated Societies by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
BIostaT. A COMPUTER PROGRAM PROVIDING SIMPLE
STATISTICS FOR BIOLOGICAL SAMPLES

J. M. Hillerud
Department of Earth and Space Science,
Midland Lutheran College, Fremont, Nebraska 68025

ABSTRACT: A computer program developing simple statistics (N, MIN, MAX, RANGE, MEAN, MEDN, SDEV, CVAR, and standard errors of the latter four) is available at the University of Nebraska State Museum, Lincoln.

Data are read into an array of 40 measurements (six columns; maximum 9999.9mm) for a sample maximum of 50 specimens. The “shape” of the array may be adjusted within the maximum “area” of 2000 F-field nodes in each data-set. The data-card input field provides IDENTIFICATION (cc 1-8), CARD (an A, B, C or D in cc 9 indicates measurements 1-10, 11-20, 21-30 and 31-40 respectively within the array), SEX (M or F in cc 10), SITE (code in cc 11-12 indicates stratigraphic or geographic location), POSITION (L or R in cc 13 shows the skeletal geography), AGE-CODE (e.g., “JUV” in cc 15-17 indicates an immature specimen), BONE-CODE (consecutive numbers for all mammalian skeletal elements in cc 18-20), and ten MEASUREMENTS of cc 6 each. Verification subroutines with data-in-error messages are provided.

Output may be on cards, tape or printout. Preliminary statements (sample identification, measurement descriptions), data and statistical results are associated for each data-sequence. The program is open-ended.

In 1969-70, I took measurements on a large number of skull and postcranial elements of Holocene bison in the collections of the University of Nebraska State Museum, which had been obtained from several High Plains sites. These measurements, described and illustrated elsewhere (Hillerud, 1972b), were recorded directly on computer coding-forms and were punched as a series of card data-decks. See Hillerud (1972a, pp 39-40) for a discussion of the repeatability and reproducibility of these measurements. The computer program described below was used to generate simple univariate statistics on these data.

Although every computing center has an equivalent simple-statistics program in its files, most programs are not specifically designed to accommodate biological data. PROGRAM BIostaT is presented here because:

(a) its range of measurements, from 0.1 to 9999.9mm, covers a major part of the macrobiological size spectrum,

(b) measurement definitions, other pertinent descriptive information and a list of the raw data are printed in proximity to the statistical printout for each data-set, facilitating its interpretation,

(c) the raw measurements can be recorded directly upon computer coding-forms, and are printed in separated columns, providing easier checking for errors in data-cards,

(d) the verification subroutine provides machine identification of mixed
data-sets, helping to prevent the generation of improper statistics. If cards
with errors are encountered, the subroutine stores the contents of the cards
for later printout as DATA IN ERROR messages. But it allows the operator
to control grouping of selected data-sets, in order to obtain a larger sample,
by defining a few control-cards as “comment-cards.” The verification of raw
data is almost a necessary element in the generation of biostatistics, but it is
not commonly provided in standard programs.

(e) the program is economical of CPU time and is open-ended. It can
provide output on cards to be used as data-decks for more sophisticated
statistical programs.

Initially, the program was called “PROGRAM SIMPSTAT” (Hillerud,
1970b, pp. 50-56, Table 3, 1-12; Hillerud, 1970a, pp. 8-9) but when we
recognized that the data-array (the “z-array”) generated by the first three
subroutines could be used by computational subroutines other than simple
univariate statistics, the program was given its present title. A new subroutine,
INDEX, which divides each measurement from a single specimen by every
other measurement, is being developed.

I am obliged to C. R. Montgomerie, Northern Alberta Institute of
Technology, Edmonton, Canada, for initial discussions in 1968 which led to
the first draft of “Program Simpstat.” J. D. Smith, University of Nebraska
Computing Center, debugged and amended the first draft, and wrote the
verification subroutine to my specifications and complete satisfaction
in 1970. J. M. Inguanzo, of the same department, aided in editing the program
for publication in 1972.

I am grateful to the University of Nebraska State Museum, which
provided funds for computer consultation and CPU time to develop the
program in 1970 as a part of my graduate research, and to The Society of the
Sigma Xi, which awarded me a Grant-in-Aid of research of $200.00 to
partially defray publishing costs.

Drs. C. B. Schultz and H. L. Gunderson, University of Nebraska State
Museum, and D. F. Costello, University of Nebraska Computing Center,
critically read this report. I am indebted to these gentlemen for their interest
and editorial criticism, but I accept responsibility for all errors remaining.

Characteristics of the Program:

The computer PROGRAM BIOSTAT utilizes raw data coded in the
format described below (see also Figure la), placing the numerical data in an
array of 2000 F-field nodes, initially 40 measurements by 50 specimens per
sample, accepting the data as follows:

cc 1-8 : specimen identification number,
c 9 : card-identification code: the letters refer to datum-positions; “A”
MATHEMATICS AND COMPUTER SCIENCE

to measurements 1-10, “B” to 11-20, “C” to 21-30 and “D” to 31-40,
c 10 : sex: “M” refers to male, “F” to female, “Blank” is not determined,
cc 11-12: site: a letter-code for geographic or stratigraphic location,
c 13 : position: skeletal geography; “L” refers to left, “R” to right, “Blank” is not applicable,
cc 15-17: age: a code for tooth-replacement or tooth-wear; for postcranial elements, “JUV” indicates juvenile specimens, “Blank” is mature,
cc 18-20: bone-code: refers the numerical data to a specific element, as listed in Table 1. Verification of this column reduces the possibility of mixed-element statistics being generated,
cc 21-80: (ten) measurements (per card), written in an F6.2 format (xxxx.x).

The program recognizes data-sets by the initial code-word “DATA” in cc 1-4. The end of an individual data-deck is identified by four “Y’s” (YYYY) in cc 1-4. The conclusion of the complete data-deck is identified by eight “Y’s” (YYYYYYYY) in cc 1-8.

The main program “calls” five subroutines sequentially:

(1) GETPREL accepts and prints explanatory data of any length in an alphanumeric field. These data are immediately available as printout to the researcher. The bone-code (Table 1) should be included in this section.

(2) GETDAT accepts and prints numerical data-sets presented for computation, and assigns the individual measurements to the computation (Z-) array. If an improper datum is encountered, this subroutine stores the contents of the card for later printout as a DATA IN ERROR message at the end of the run.

(3) VERIFY checks card order and homogeneity of the data-set. For example, if a B-card (“B” in c 9) with data for array-spaces 11-20 is encountered without a correct preceding A-card (with data for spaces 1-10), all information on the B-card is stored and later printed as a DATA IN ERROR message. The subroutine also verifies the data-sets by sex, site, position, age and bone-code. Sex, site, position and age-categories are removed from the program as it is reproduced here, allowing mixed data-sets of sufficient size for valid statistical analysis to be run. These control-cards remain in the subroutine deck as comment-cards, and can be re-activated by duplicating them without the “C***” in cc 1-4.

(4) PROCES generates statistics upon the z-array data-sets in several sequences. The formulae used in this subroutine are those presented in

207
Simpson, et. al. (1960, chapters 5-6 and pp. 166-167). The statistical parameters N, RMIN, RMAX and RMEDN are developed by computer counting techniques. All formulae are translated into FORTRAN-IV computer language.

(5) PUTDAT prints the statistical computations on the line-printer, and as a punched deck if desired. The printed array includes space for 40 lines of information, of which only the number of measurement parameters are used, and the rest show as "zero". Line 41 records the total number of specimens in the sample. Figure 1(b,c,d) illustrates the printout of an abbreviated data-set.

To avoid possible typographical errors, the working PROGRAM BI⁸STAT is presented in facsimile as Table 2.

As here presented, PROGRAM BI⁸STAT accepts data in an array of 40 measurements by 50 specimens per data-set sample. A worker wishing to process fewer measurements on a larger sample can change the "shape" of the array by rewriting only a few statements in the program. For example, the following changes will produce a revised program which will accept 20 measurements on 100 specimens. The lines to be changed are marked with marginal dots in the program listing, Table 2.

In JCL machine control, rewrite the job name "BI⁸STAT40" as "BI⁸STAT20", to define the change.

In the main program, line 4, COMM8N line 3, change "Z(S1,40)" to "Z(IO1,20)". Duplicate six cards, and insert one for the main program COMM8N statement, and one for the COMM8N statements in each of the five subroutines.

In the main program, line 9, change "N40=40" to "N40=20", and line 10, change "NSO=SO" to "NSO=100".

No further changes are necessary. Other adjustments to the "shape" of the array can be easily developed, by changing the nine cards listed above.
TRANSACTIONS OF THE NEBRASKA ACADEMY OF SCIENCES

Coding form with data prepared for PROGRAM BIOSTAT (a). The printout shows measurement definitions (b), listing of measurement data (c) and statistics printout (d) on a data-set of Bison radius-ulnae processed by BIOSTAT.

---

(a) Table:

<table>
<thead>
<tr>
<th>AUDIT</th>
<th>CARD</th>
<th>SEX</th>
<th>SITE</th>
<th>DISED</th>
<th>AGE</th>
<th>ZONE</th>
<th>MEAS (10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3052-65A</td>
<td>L</td>
<td>.6</td>
<td>4.59</td>
<td>0.1</td>
<td>219</td>
<td>0</td>
<td>344</td>
</tr>
<tr>
<td>3053-65A</td>
<td>L</td>
<td>.6</td>
<td>4.11</td>
<td>0.1</td>
<td>393</td>
<td>0</td>
<td>316</td>
</tr>
<tr>
<td>3064-65A</td>
<td>L</td>
<td>.6</td>
<td>4.16</td>
<td>0.1</td>
<td>393</td>
<td>0</td>
<td>321</td>
</tr>
<tr>
<td>3065-65A</td>
<td>L</td>
<td>.6</td>
<td>4.16</td>
<td>0.1</td>
<td>393</td>
<td>0</td>
<td>315</td>
</tr>
<tr>
<td>3068-65A</td>
<td>L</td>
<td>.6</td>
<td>4.10</td>
<td>0.1</td>
<td>393</td>
<td>0</td>
<td>320</td>
</tr>
<tr>
<td>3060-65A</td>
<td>L</td>
<td>.6</td>
<td>4.30</td>
<td>0.1</td>
<td>319</td>
<td>0</td>
<td>319</td>
</tr>
</tbody>
</table>

---

(b) Material preparation:

- **Preliminary Statement**: RADIUS-ULNA - SPECIMEN CODE NO. 66
  - TOTAL LENGTH OF ELEMENT
  - LENGTH OF ULNA
  - LENGTH OF RADIUS
  - ANTERIOR-POSTERIOR WIDTH OF ELBOW JOINT
  - LENGTH OF ELECRANON TO SEMILUNAR NOTCH
  - LATERAL WIDTH OF PROXIMAL EXTREITY
  - WIDTH OF PROXIMAL ARTICULATING SURFACE, RADIUS
  - ANTERIOR-POSTERIOR WIDTH OF RADIUS, PROXIMAL EXTREITY
  - LATERAL WIDTH OF RADIUS SHFT AT SAME POINT
  - ANTERIOR-POSTERIOR WIDTH OF RADIUS SHFT AT SAME POINT
  - WIDTH OF RADIUS-ULNA, DISTAL EXTREITY
  - ANTERIOR-POSTERIOR WIDTH, DISTAL EXTREITY OF RADIUS-ULNA

(c) Data:

- **Mixed Deck**: ADULT AND JUVENILE RADIUS-ULNAE, MILBURN

(d) Statistics:

- **SECVAR**
Table 1. The mammalian bone-code, listing the elements of the mammalian skeleton with identifying code-numbers assigned to each element. The number is placed in cc 18-20 on each data card. Note the aberrant 900-numbers for unidentified teeth; these illustrate a technique for revision of the code, presented here as a “third approximation.”

<table>
<thead>
<tr>
<th>CRANIAL ELEMENTS</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1 SKULL FRAGMENT</td>
<td>8 UC-1</td>
<td>16 LI-1</td>
<td>24 LP-3</td>
</tr>
<tr>
<td>2 MANIDIAL HANUS</td>
<td>9 UP-1</td>
<td>17 LI-2</td>
<td>25 LP-4</td>
</tr>
<tr>
<td>TEETH</td>
<td>10 UP-2</td>
<td>18 LI-3</td>
<td>925 UNIDENT. LPM</td>
</tr>
<tr>
<td>3 UT-1</td>
<td>11 UP-3</td>
<td>19 LI-4</td>
<td>26 LM-1</td>
</tr>
<tr>
<td>4 UT-2</td>
<td>12 UP-4</td>
<td>20 LI-5</td>
<td>28 LM-2</td>
</tr>
<tr>
<td>5 UT-3</td>
<td>912 UNIDENT. UP-920 UNIDENT. LI-28 LM-3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 UT-4</td>
<td>13 UM-1</td>
<td>21 LC-1</td>
<td>928 UNIDENT. LM-29 IDENT. TOOTH</td>
</tr>
<tr>
<td>7 UT-5</td>
<td>14 UM-2</td>
<td>22 LP-1</td>
<td>29 IDENT. TOOTH</td>
</tr>
<tr>
<td>907 UNIDENT. UT-15 UM-3</td>
<td>23 LP-2</td>
<td>30 GREAT CORN. HYOID</td>
<td></td>
</tr>
<tr>
<td>915 UNIDENT. UM-31 MINOR CORN. HYOID</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AXIAL ELEMENTS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32 CV-1 ATLAS</td>
<td>40 TV-1</td>
<td>48 TV-9</td>
<td>56 LV-3</td>
</tr>
<tr>
<td>33 CV-2 AXIS</td>
<td>41 TV-2</td>
<td>49 TV-10</td>
<td>57 LV-4</td>
</tr>
<tr>
<td>34 CV-3</td>
<td>42 TV-3</td>
<td>50 TV-11</td>
<td>58 LV-5</td>
</tr>
<tr>
<td>35 CV-4</td>
<td>43 TV-4</td>
<td>51 TV-12</td>
<td>59 LV-6</td>
</tr>
<tr>
<td>36 CV-5</td>
<td>44 TV-5</td>
<td>52 TV-13</td>
<td>60 UNIDENT LV</td>
</tr>
<tr>
<td>37 CV-6</td>
<td>45 TV-6</td>
<td>53 UNIDENT TV</td>
<td>61 SACRUM</td>
</tr>
<tr>
<td>38 CV-7</td>
<td>46 TV-7</td>
<td>54 LV-1</td>
<td>62 CAUDAL V. UNDIFF.</td>
</tr>
<tr>
<td>39 UNIDENT CV</td>
<td>47 TV-8</td>
<td>55 LV-2</td>
<td>63 --------</td>
</tr>
<tr>
<td>FORE-LIMB ELEMENTS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>64 SCAPULA</td>
<td>70 ULNAR C.</td>
<td>76 MC-2</td>
<td></td>
</tr>
<tr>
<td>65 HUMERUS</td>
<td>71 ACCESSORY C.</td>
<td>77 MC-3 (CANNON)</td>
<td></td>
</tr>
<tr>
<td>66 RADIUS-ULNA</td>
<td>72 C12+3 C.</td>
<td>78 MC-4</td>
<td></td>
</tr>
<tr>
<td>67 ULNA</td>
<td>73 C-4 C.</td>
<td>79 MC-5</td>
<td></td>
</tr>
<tr>
<td>68 RADIAL C.</td>
<td>74 CARPAL UNDIFF.</td>
<td>80 UNIDENT MC (SPLINT)</td>
<td></td>
</tr>
<tr>
<td>69 INTERMEDIATE C.75 METACARPAL-1</td>
<td>81 --------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHALANGES</td>
<td>( PHALANX ONE, DIGIT TWO)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHALANGES</td>
<td>( PHALANX ONE, DIGIT TWO)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>82 P-1-O-1</td>
<td>87 P-1-O-1</td>
<td>92 P-1-O-1</td>
<td>97 UNIDENT PHALANX</td>
</tr>
<tr>
<td>83 P-1-O-2</td>
<td>88 P-1-O-2</td>
<td>93 P-1-O-2</td>
<td></td>
</tr>
<tr>
<td>84 P-1-O-3</td>
<td>89 P-1-O-3</td>
<td>94 P-1-O-3</td>
<td></td>
</tr>
<tr>
<td>85 P-1-O-4</td>
<td>90 P-1-O-4</td>
<td>95 P-1-O-4</td>
<td></td>
</tr>
<tr>
<td>86 P-1-O-5</td>
<td>91 P-1-O-5</td>
<td>96 P-1-O-5</td>
<td></td>
</tr>
<tr>
<td>HIND-LIMB ELEMENTS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>97 PELVIS</td>
<td>103 LAT. MAL. FIN.</td>
<td>109 MT-2</td>
<td></td>
</tr>
<tr>
<td>98 FEMUR</td>
<td>104 T-(L)</td>
<td>110 MT-3 (CANNON)</td>
<td></td>
</tr>
<tr>
<td>99 TIBIA</td>
<td>105 T-(2+3)</td>
<td>111 MT-4</td>
<td></td>
</tr>
<tr>
<td>100 FIBULA</td>
<td>106 T-(C+4)</td>
<td>112 MT-5</td>
<td></td>
</tr>
<tr>
<td>101 CALCANEUM</td>
<td>107 UNIDENT TARS.113 UNIDENT MT (INCLUDES SPLINT).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>102 ASTRAGALUS</td>
<td>108 MT-1</td>
<td>114 --------</td>
<td></td>
</tr>
<tr>
<td>RIB-CAGE ELEMENTS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>115 R-1</td>
<td>119 R-5</td>
<td>123 R-9</td>
<td>127 R-13</td>
</tr>
<tr>
<td>116 R-2</td>
<td>120 R-6</td>
<td>124 R-10</td>
<td>128 UNIDENT. RIB</td>
</tr>
<tr>
<td>117 R-3</td>
<td>121 R-7</td>
<td>125 R-11</td>
<td>129 COSTAL CART.</td>
</tr>
<tr>
<td>118 R-4</td>
<td>122 R-8</td>
<td>126 R-12</td>
<td>130 --------</td>
</tr>
<tr>
<td>ACCESSORY ELEMENTS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>131 CLAVICLE</td>
<td>134 XIPHISSERTNUM</td>
<td>137 SESAMOID DIST.140 --------</td>
<td></td>
</tr>
<tr>
<td>132 STERNEBRA</td>
<td>135 PATELLA</td>
<td>138 BAGULUM</td>
<td>141 --------</td>
</tr>
<tr>
<td>133 MANUNHM</td>
<td>136 SESAMOID PROX.139 DS GLITORIS</td>
<td>142 --------</td>
<td></td>
</tr>
</tbody>
</table>
Table 2. PROGRAM BIOSTAT, in facsimile.

**JCL Machine Control**

```plaintext
TRANSACTIONS OF THE NEBRASKA ACADEMY OF SCIENCES

//MILLERDU JOB MUSKRAT00200, IOBSTAT0000, MSGLEVEL=1, PRTY=7, CLASS=A
//MILLERDU JOB MUSKRAT00200
//MILLERDU JOB MUSKRAT00200
*** IF YOU ASW TO OMIT SPECIFY PLOT=VOL+PRIVATE ON THE EXEC CARD
*** AND PROGRAM=PRIVATE
***
//MILLERDU JOB MUSKRAT00200
//MILLERDU JOB MUSKRAT00200
//MILLERDU JOB MUSKRAT00200

**Table 2. PROGRAM BIOSTAT, in facsimile.**

**JCL Machine Control**

```
```

**Program**

```
```

---

211
CALL GTPREL
IF(0.O.EQ.2.0)GO TO 5
CALL PUTDAT
CALL PUTFPS
CALL PUTDAT
IF(O.O.EQ.2.0)GO TO 1
WRITE(IN,*)
4 FORMAT(**)
5 STOP
6 END

SUBROUTINE GTPREL
DIMENSION LINE(20)
INTEGER*4 DATA, WYEEL4*/DATA/*1,YYYY*/
INTEGER*4 CARD, SEX, SITE, Position, AGE, BONE
REAL*4 MEASURE
REAL*8 WYEEL4, WYEEL8
INTEGER ACRO
DATA YEE4, WYEEL8, ACRO/*YYYY*/
REAL*4 MEASURE
REAL*8 MU I O, HMUID
COMMON MUID(20U), MEASURE(20U, 10), RMIN(40), RMAX(40),
RANGE(40), RMEAN(40), SDNDEV(40), RCVAR(40), SEMEAN(40),
SEDEVI40, SECVAR(40), SEMEDN(40), CARD(20U), SEX(20U),
SITE(20U), POSID(20U), AGE(20U), BONE(20U), YEE20U,
CINTA, DONE, ILPP, IN, IE, SUM(40), N10, N40, N50,
N51, N55, N200, N201, I
1 I NECT=0
WRITE(10, Z)
ILPP=1
2 FORMAT(*'///',/)
3 READ(IN,100)NDATAM rubbish GT401GO TO 50
WRITE(10, Z)
3 FORMAT(*'///',/)
4 IF(I NECT.GT.40)GO TO 1
5 FORMAT(*'///',/)
6 ILPP=ILPP+1
7 IF(ILPP.LE.55)GO TO 5
8 WRITE(10, Z)
9 25 FORMAT(*')
10 ILPP=0
GO TO 5
11 DONE=2.0
12 WRITE(10, Z)
13 50 FORMAT(*'///',/)
14 60 FORMAT(*'///',/)
15 70 RETURN
16 END

SUBROUTINE GTPREL
REAL*8 WYEE4, WYEE8
INTEGER*4 ACRO
DATA WYEEL4, WYEEL8, ACRO/*YYYY*/
REAL*4 MEASURE
REAL*8 MU I O, HMUID
COMMON MUID(20U), MEASURE(20U, 10), RMIN(40), RMAX(40),
RANGE(40), RMEAN(40), SDNDEV(40), RCVAR(40), SEMEAN(40),
SEDEVI40, SECVAR(40), SEMEDN(40), CARD(20U), SEX(20U),
SITE(20U), POSID(20U), AGE(20U), BONE(20U), YEE20U,
CINTA, DONE, ILPP, IN, IE, SUM(40), N10, N40, N50,
N51, N55, N200, N201, I
110 FORMAT(*'///',/)
120 C
C ICNTA = THE NUMBER OF SPECIMENS...
212
TRANSACTIONS OF THE NEBRASKA ACADEMY OF SCIENCES

213
MATHEMATICS AND COMPUTER SCIENCE

THE VELOCITY SPECTRUM HAS ENCOUNTERED HOMOGENEOUS DATA.

THE ARRAY CONTINUE

THE ARRAY CONTINUE

THE ARRAY CONTINUE

THE ARRAY CONTINUE

THE ARRAY CONTINUE

THE ARRAY CONTINUE

THE ARRAY CONTINUE
TRANSACTIONS OF THE NEBRASKA ACADEMY OF SCIENCES

2-5
REFERENCES CITED


