BRITISH OCCUPATIONAL HYGIENE SOCIETY

STATISTICAL ANALYSIS OF MONITORING DATA BY MICROCOMPUTER

Technical Guide Series, No. 1

August 1983

TECHNICAL GUIDE SERIES

No. 1 Statistical Analysis of Monitoring Data by Microcomputer. August 1983

Copies of the Technical Guide series are obtainable price £2.00 post free (cash with order) from the British Occupational Hygiene Society, Honorary Business Secretary, Mr P.J. Hewitt, University of Bradford. Bradford BD7 1DP

STATISTICAL ANALYSIS OF MONITORING DATA

BY MICROCOMPUTER

by D.T. Coker

Esso Europe Inc,
Industrial Hygiene Unit, Esso Research Centre,
Abingdon, Oxfordshire OX13 6AE.

A Basic programme for calculating means, standard deviations, Filliben fit factors, and probability plots for normal and log-normal distributions of monitoring data.

British Occupational Hygiene Society Technical Guide No. 1.

The BOHS Technical Guide series is intended for reports of practical use in occupational hygiene, mainly produced by the BOHS Technology Committee. The responsibility for accuracy and for opinions stated remains with the author, and not with the Society.

British Occupational Hygiene Society,
London School of Hygiene and Tropical Medicine,
Keppel Street, London WC1E 7HT

1983

STATISTICAL ANALYSIS OF MONITORING DATA BY MICROCOMPUTER

by D.T. Coker

The cost of microcomputers has now become so low that relatively powerful machines are within the budget means of the average occupational hygienist.

The applications of these computers in the hygiene area may not be obvious and programmes of use to the hygienist are not abundant at present.

In order to present applications and programmes for the hygienist the BOHS Technology Committee decided to start collecting and publishing these.

This is the first such paper and gives a programme which enables hygiene data to be analysed and presented in an understandable manner.

The author would welcome any programmes from hygienists to include in this series, covering any aspect of hygiene.

The programme given here is for processing any set of data for graphical presentation as a probability plot. This is a very convenient way of presenting data as it enables observations and predictions on the relationship between the data and exposure limits to be made.

Typical information which is required from hygiene data may include:-

- what is the typical exposure level?
- what is the average exposure?
- does the workplace comply with the TLV?
- what is the probability of exceeding the TLV?
- if the TLV is lowered, will the workplace then be out of compliance?

The probability plot is a simple understandable way of presenting data so that such information can be easily extracted from what may otherwise be a confusing mass of numbers.

Most statistical procedures are for normally (gaussian) distributed data but in general hygiene measurements tend to be lognormally distributed. However, this is not a hard and fast rule and so it is advisable to first test the data distribution using fit factors. The statistical parameters of the mean and standard deviation can then be calculated for the appropriate type of distribution, and the datum points can then be plotted graphically using probability paper with the appropriate concentration axis.

There are many ways of testing for distribution and this paper does not intend to review or compare these. The particular procedure - the Filliben probability plot correlation coefficient - is used here as it is applicable to this type of data, convenient for incorporating into computer programmes and the significance level of the coefficient, which is maximum at unity, can be read off one table.

Ref- Filliben, J.J. (1975) Technometrics 17, No.1, 111-117 Feb. 1975.

The programme is written in basic, originally for a Commodore PET 2001, but this should be easily translatable to other variants of basic. Line nos. 460 and 470 have been left vacant to allow commands to be inserted directing output to a line printer; line 1300 is also left vacant for printer termination commands.

The Programme Procedure is as follows:-

Line nos.

80 - 130	inputs name of data set and number of values
160 - 180	inputs each value
200 - 270	rearranges the values into ranked order, lowest to highest
280 - 290	converts each value to a logarithm
300 - 340	calculates means
350 - 430	calculates geometric standard deviation
480 - 580	prints table heading, means and GSD
700 - 740	calculates the plotting position for each value on the probability axis
750 - 1020	calculates and prints the fit coefficient for lognormal distribution

Line nos.

590 - 660	calculates and prints normal standard deviation
1090	calculates and prints the fit coefficient for normal distribution
1120 - 1200	calculates and prints the coordinates for drawing the theoretical fit lines for both distributions
1220 - 1280	prints table of data values and their probability plotting position so that the graph can be plotted

The fit coefficient closest to unity indicates which distribution the data fits best. This programme calculates normal and lognormal fits but it may be possible to incorporate other distributions if required, but the fit factor may not be applicable to all distributions.

The Filliben paper referenced above gives signficance levels for the coefficient.

An example of 20 points is given below showing the computer printout and the probability plot.

'Test Set of 20 Data Points'

3.0, 33, 1.7, 4.4, 1.3, 4.1, 12, 17, 6.2, 2.5, 4.2, 0.47, 4.2, 7.9,

1.4, 0.70, 10, 19, 2.2, 2.3

Program listing

```
10 PRINT"DATACRUNCH"
20 PRINT"THIS PROGRAMME CALCULATES MEANS,STANDARDDEVIATIONS AND DISTRIBUTION"
30 PRINT"FIT FACTORS FOR LOGNORMAL AND NORMAL     DISTRIBUTIONS"
40 PRINT"THEN TABULATES DATA IN RANKED ORDER WITHPROBABILITY PLOTTING POSITION
50 PRINT"FOR EACH POINT"
60 PRINT
70 PRINT
80 PRINT"ENTER NAME OF DATA SET"
90 PRINT
100 INPUTA$
110 PRINT
120 PRINT"ENTER NO OF DATA POINTS"
130 INPUTN
140 PRINT"ENTER EACH DATA POINT FOLLOWED BY RETURN"
150 DIMA(N),B(N),T(N),X(N),K(N),H(N),L(N),Y(N),Z(N)
160 FORI=1TON
170 INPUTH(I)
180 NEXTI
190 F=0:S=0:T=0:G=0:W=0:Z1=0:Z2=0:Z3=0
200 FORI=1TO(N-1)
210 IFH(I)<=H(I+1)THEN260
220 Y=H(I)
230 H(I)=H(I+1)
240 H(I+1)=Y
250 F=1
260 NEXTI
270 IFF=1 THEN190
280 FORI=1TON
290 A(I)=LOG(H(I))
300 S=S+A(I)
310 T=T+H(I)
320 NEXTI
330 M=S/N
340 R=T/N
350 FORI=1TON
360 B(I)=A(I)-M
370 D=ABS(B(I))
380 G=G+D*D
390 NEXTI
400 V=G/(N-1)
410 C=SQR(V)
420 P=EXP(C)
430 Q=EXP(M)
440 PRINT
450 PRINT
480 PRINTA$
490 PRINTN"VALUES,RANGE"H(1)"TO"H(N)
500 PRINT"------------------------------------------------------------------------"
510 PRINT"FOR LOGNORMAL DISTRIBUTION-"
520 PRINT
530 PRINT
540 PRINT"GEOMETRIC MEAN=";INT(Q*1000+.5)/1000
550 PRINT"GEOMETRIC SD=";INT(P*100+.5)/100
560 GOTO690
570 J=0:Z1=0:Z2=0:Z3=0
580 PRINT"ARITHMETIC MEAN=";INT(R*100+.5)/100
590 FORI=1TON
600 B(I)=H(I)-R
610 D=ABS(B(I))
620 J=J+D*D
630 NEXT I
640 V=J/(N-1)
650 C=SQR(V)
660 PRINT"STANDARD DEVIATION=";INT(C*1000+.5)/1000
670 W=W+1
680 GOTO960
690 W=W+1
```

```
700 K(N)=.5↑(1/N)
710 K(1)=1-K(N)
720 FORI=2TO(N-1)
730 K(I)=(I-.3175)/(N+.365)
740 NEXT I
750 L1=2.515517
760 L2=.802853
770 L3=.010328
780 D1=1.432788
790 D2=.189269
800 D3=.001308
810 FORI=1TON
820 O=O+K(I)
830 IFK(I)>.5THEN860
840 E=K(I)
850 GOTO870
860 E=1-K(I)
870 T(I)=SQR(LOG(1/(E*E)))
880 NEXTI
890 H=O/N
900 FORI=1TON
910 X(I)=T(I)-(L1+L2*T(I)+L3*T(I)*T(I))/(1+D1*T(I)+D2*T(I)*T(I)+D3*T(I)↑3)
920 NEXTI
930 FORI=1TON
940 IFK(I)<.5THENX(I)=-X(I)
950 NEXTI
960 FORI=1TON
970 Z1=Z1+B(I)*X(I)
980 Z2=Z2+B(I)*B(I)
990 Z3=Z3+X(I)*X(I)
1000 NEXTI
1010 IFW=2THEN1090
1020 PRINT"FILLIBENS FIT FACTOR,R=";INT((Z1/SQR(Z2*Z3))*10000+.5)/10000
1030 PRINT
1040 PRINT
1050 PRINT"FOR NORMAL DISTRIBUTION-"
1060 PRINT
1070 PRINT
1080 GOTO570
1090 PRINT"FILLIBENS FIT FACTOR,R=";INT((Z1/SQR(Z2*Z3))*10000+.5)/10000
1100 PRINT
1110 PRINT
1120 PRINT"THEORETICAL FIT LINE PLOTTING POINTS"
1130 PRINT"VALUES","PLOTTING POSITIONS"
1140 PRINT"FOR LOGNORMAL DISTRIBUTION"
1150 PRINTINT((Q/(P↑1.64513))*10000+.5)/10000,"5.0% PROBABILITY"
1160 PRINTINT((Q*(P↑1.64513))*100+.5)/100,"95.0% PROBABILITY"
1170 PRINT
1180 PRINT"FOR NORMAL DISTRIBUTION"
1190 PRINTINT((R-C*1.64513)*10000+.5)/10000,"5.0% PROBABILITY"
1200 PRINTINT((R+C*1.64513)*100+.5)/100,"95.0% PROBABILITY"
1210 PRINT
1220 PRINT"DATA PLOTTING POSITIONS"
1230 PRINT
1240 PRINT"VALUES","PLOTTING POSITIONS"
1250 PRINT
1260 FORI=1TON
1270 L(I)=INT(K(I)*1000+.5)/10
1280 PRINTH(I),L(I)
1290 NEXTI
1310 END
READY.
```

Printout of results for test set (see text p.3)

TEST SET
20 VALUES, RANGE .47 TO 33

FOR LOGNORMAL DISTRIBUTION-

GEOMETRIC MEAN= 3.929
GEOMETRIC SD= 3.03
FILLIBENS FIT FACTOR, R= .9955

FOR NORMAL DISTRIBUTION-

ARITHMETIC MEAN= ~~[illegible]~~ 6.879
STANDARD DEVIATION= 8.084
FILLIBENS FIT FACTOR, R= .8505

THEORETICAL FIT LINE PLOTTING POINTS

VALUES	PLOTTING POSITIONS
FOR LOGNORMAL DISTRIBUTION	
.6337	5.0% PROBABILITY
24.36	95.0% PROBABILITY
FOR NORMAL DISTRIBUTION	
-6.4207	5.0% PROBABILITY
20.18	95.0% PROBABILITY

DATA PLOTTING POSITIONS

VALUES	PLOTTING POSITIONS
.47	3.4
.7	8.3
1.3	13.2
1.4	18.1
1.7	23
2.2	27.9
2.3	32.8
2.5	37.7
3	42.6
4.1	47.5
4.2	52.5
4.2	57.4
4.4	62.3
6.2	67.2
7.9	72.1
10	77
12	81.9
17	86.8
19	91.7
33	96.6

LOG PROBABILITY PLOT OF TEST SET