Statistical Bulletin
Reliability & Variation Research

DETROIT RESEARCH INSTITUTE
P.O. BOX 36504 • GROSSE POINTE, MICHIGAN 48236 • (313) 886-7976

LEONARD G. JOHNSON
EDITOR

WANG H. YEE
DIRECTOR

Volume 17
Bulletin 3

July , 1987
Page 1

INVERSE REGRESSION ANALYSIS OF ORDERED MEASUREMENTS ON NORMAL PROBABILITY PAPER

INTRODUCTION

IN THE FIELD OF STATISTICAL PROCESS CONTROL THE SO-CALLED NORMAL (GAUSSIAN) DISTRIBUTION HAS PLAYED A DOMINANT ROLE OVER THE MANY DECADES SINCE THE PIONEERS OF QUALITY CONTROL INITIATED THE USE OF STATISTICS ON MANUFACTURING ASSEMBLY LINES. FURTHERMORE, IN THE INTEREST OF CLARIFYING THE BASIC IDEAS INVOLVED IN STATISTICAL QUALITY CONTROL WE HAVE FOUND THAT, THE GRAPHICAL APPROACH IS MOST EFFECTIVE IN ANY EDUCATIONAL EFFORT AT BUILDING UP AN UNDERSTANDING OF THIS MOST IMPORTANT DECISION TOOL IN MODERN INDUSTRY.

FOR THE ABOVE REASONS WE ARE PRESENTING THIS STATISTICAL BULLETIN TO CLARIFY THE CURVE FITTING PROCESS (LEAST SQUARES) WHICH CAN GENERATE ACCURATE NORMAL PARAMETERS FROM ACTUAL SAMPLES OF MEASUREMENTS WHICH ARE IN NEED OF CONTROL. ONCE THE MATHEMATICS OF REGRESSION ANALYSIS YIELDS THE BEST FITTING PARAMETERS FOR A NORMAL DISTRIBUTION WE PICTURE THE RESULTS ON NORMAL PROBABILITY PAPER FOR A CLEAR UNDERSTANDING OF THE PROCESS CONTROL PROBLEM.

INVERSE REGRESSION ANALYSIS ON NORMAL PAPER

THE PROPER PROCEDURE FOR FITTING A LINE ON NORMAL PROBABILITY PAPER TO A DATA SET CAN BE OUTLINED AS FOLLOWS:

EVERY LINEAR REGRESSION (IN THE INVERSE SENSE) IS EXPRESSED AS A LINEAR EQUATION OF THE TYPE

$$X = A + B*Y$$

WHERE X=SOME TRANSFORMED VALUE OF EACH MEASURED X (ABSCISSA)

AND Y=CORRESPONDING TRANSFORMED ORDINATE (PAIRED WITH X)

IN THE CASE OF A STATISTICAL PROCESS CONTROL PROBLEM WE TAKE

X=ACTUAL MEASUREMENT

AND WE TAKE Y=LN(F/(1-F)), WHERE

F=MEDIAN RANK OF THE ORDER STATISTIC NUMBER FOR MEASUREMENT X

THE BEST WAY TO DEMONSTRATE WHAT THIS INVERSE REGRESSION PROCESS IS ALL ABOUT IS TO TAKE AN ACTUAL EXAMPLE (SEE NEXT PAGE).

EXAMPLE

SUPPOSE WE MEASURE A SHAFT DIAMETER ON SEVERAL SPECIMENS, AS INDICATED BELOW:

8 SPECIMENS (SHAFTS) WERE MEASURED FOR DIAMETER (NOMINAL 2 IN.):

$$\left\{\begin{matrix} 2.011, & 2.005, & 1.997, & 1.998 \\ 1.995, & 2.002, & 2.003, & 1.991 \end{matrix}\right\} \quad (N=8)$$

THESE MEASUREMENTS ARE PUT INTO NUMERICAL ORDER, AS FOLLOWS:

ORDER NO.	X	F=MED.RANK*	Y=LN(F/(1-F))
1	1.991	.0833	-2.3983
2	1.995	.2024	-1.3714
3	1.997	.3214	-0.7473
4	1.998	.4405	-0.2391
5	2.002	.5595	+0.2391
6	2.003	.6786	+0.7473
7	2.005	.7976	+1.3714
8	2.011	.9167	+2.3983

*THE MEDIAN RANK OF ORDER STATISTIC NO. J IN A SAMPLE OF SIZE N IS GIVEN BY BENARD'S FORMULA F=(J-.3)/(N+.4).

IN THE LEAST SQUARES METHOD OF FITTING LINES TO DATA WE NEED THE FOLLOWING ITEMS:

SAMPLE SIZE = N

SUM OF Y'S = S1

SUM OF X'S = T1

SUM OF SQUARES OF Y'S = S2

SUM OF SQUARES OF X'S = T2

SUM OF XY PRODUCTS = P1

IN OUR COMPUTER PROGRAM NORPLOT WE USE THE SYMBOLS S1,T1,S2,T2, AND P1, AS INDICATED ABOVE.

THEN, WE CALCULATE COEFFICIENTS A AND B IN THE LINEAR RELATION

X = A + B*Y

THIS IS DONE BY USING THE FOLLOWING FORMULAS:

$$A = \frac{S2*T1 - S1*P1}{N*S2 - S1*S1}$$

$$B = \frac{N*P1 - S1*T1}{N*S2 - S1*S1}$$

IN THE CASE OF A NORMAL PROBABILITY PAPER PLOT WE CAN STATE THAT

MEAN = A

STANDARD DEVIATION (SIGMA) = .1296*B*(10+LN(N)) (N<50)

= 1.8*B (N≥50)

IN THE NUMERICAL EXAMPLE WE HAVE

N = 8

T1 = 16.002 (SUM OF X'S)

S1 = 0 (SUM OF Y'S)

T2 = 32.008278 (SUM OF SQUARES OF X'S)

S2 = 16.4964139 (SUM OF SQUARES OF Y'S)

P1 = .0671202 (SUM OF XY PRODUCTS)

FROM THE ABOVE SUMMATIONS WE CALCULATE

$$A = \frac{(16.4964139)(16.002) - 0(.0671202)}{8(16.4964169) - 0}$$

$$= \frac{263.9756152}{131.9713352}$$

$$= 2.00025$$

$$B = \frac{8(.0671202) - 0}{8(16.4964139)-0}$$

$$= .00406878$$

FROM THESE VALUES OF A AND B WE GET

MEAN = A = 2.00025

SIGMA = .1296(.00406878)(10+2.07944) = .006370

THE CORRELATION COEFFICIENT (GOODNESS OF FIT) IS GIVEN BY THE FORMULA

$$R = \frac{N*P1 - S1*T1}{\sqrt{(N*S2 - S1*S1)(N*T2 - T1*T1)}}$$

$$= \frac{8(.0671202) - 0}{\sqrt{8(16.4964139)\left[8(32.008278)-(16.002)^2\right]}}$$

$$= \frac{.0671202}{\sqrt{.0045778}}$$

$$= \frac{.0671202}{.067659} = ,992$$

FOR N POINTS WE REQUIRE A CORRELATION COEFFICIENT AT LEAST 1 - 1/2N.

FOR N = 8 THIS BECOMES 1 - 1/16 = 15/16 = ,9375.

SINCE .992 > .9375, WE CONCLUDE THAT THE NORMALTY ASSUMPTION IS JUSTIFIED.

CONCLUSION

BY PLOTTING A SET OF MEASURED ORDER STATISTICS VS LN $\left(\frac{\text{MEDIAN RANK}}{\text{1-MEDIAN RANK}}\right)$ ON NORMAL PROBABILITY PAPER WE GET GOOD ESTIMATES OF

(A) MEAN

(B) SIGMA

(C) GOODNESS OF FIT

THESE ESTIMATES ARE DETERMINED BY USING INVERSE REGRESSION ANALYSIS FOR A RELATION

X = A +BY, WHERE

X = MEASURED VARIABLE UNDER STUDY

AND $Y = \text{LN}\left(\frac{\text{MEDIAN RANK}}{1 - \text{MEDIAN RANK}}\right)$

THE NORMAL PROBABILITY PLOT FOR THE EXAMPLE APPEARS IN FIGURE 1. IT CAN BE SEEN THAT THE MEAN IS AT 2.00025 (CENTER), AND

MEAN + 3 SIGMA IS AT 2.01936 (TOP),

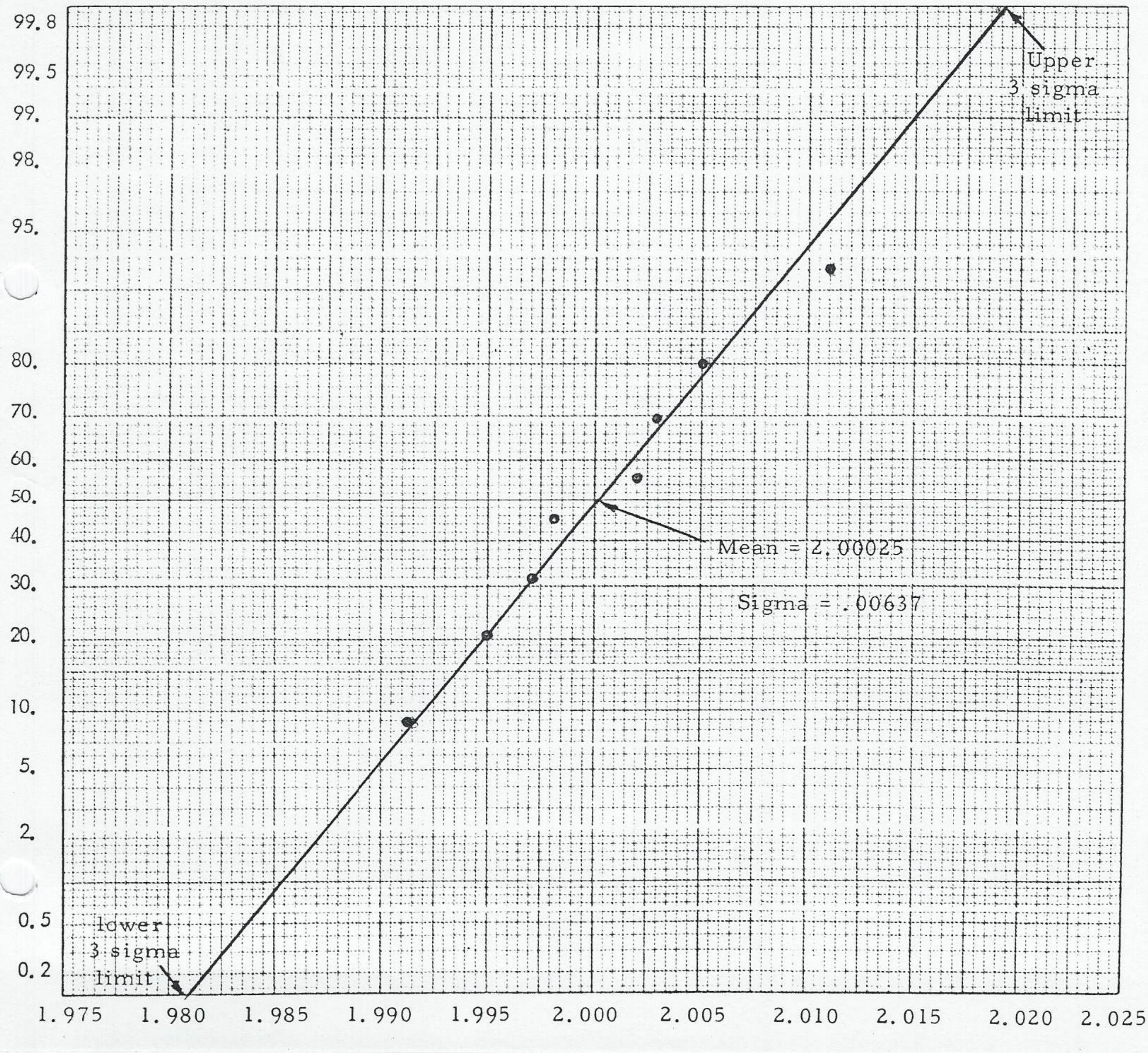
AND MEAN - 3 SIGMA IS AT 1.98114 (BOTTOM).

THE COMPUTER PROGRAM AND OUTPUT APPEAR IN THE APPENDIX.

NORMAL PROBABILITY PLOT FOR THE EXAMPLE

NOTE: Each data point is put at its Median Rank

FIGURE 1



APPENDIX (1)

PROGRAM LISTING

```
1 COLOR 14
10 DIM F(100),X(100),Y(100)
20 PRINT "NORMAL PLOT FROM ORDER STATISTICS"
30 LPRINT "NORMAL PLOT FROM ORDER STATISTICS"
40 PRINT
50 LPRINT
60 PRINT
70 LPRINT
90 LPRINT
100 DATA 8,1.991,1.995,1.997,1.998,2.002,2.003,2.005,2.011
110 READ N
120 FOR I=1 TO N
130 READ Y(I)
140 F(I)=(I-.3)/(N+.4)
150 X(I)=LOG((F(I))/(1-F(I)))
155 NEXT I
160 S1=0
170 S2=0
180 T1=0
190 T2=0
200 P1=0
210 FOR I = 1 TO N
220 U=X(I)
230 V=Y(I)
240 S1=S1+U
250 S2=S2+U*U
260 T1=T1+V
270 T2=T2+V*V
280 P1=P1+U*V
290 Q1=N*P1-S1*T1
300 Q2=N*S2-S1*S1
310 Q3=N*T2-T1*T1
320 Q4=S2*T1-S1*P1
325 NEXT I
330 A=Q4/Q2
340 B=Q1/Q2
350 S=.1296*B*(10+LOG(N))
355 REM FOR N>50 MAKE ABOVE S=1.8*B
360 R=Q1/(SQR(Q2*Q3))
370 PRINT "VALUE","MEDIAN RANK"
380 LPRINT "VALUE","MEDIAN RANK"
390 FOR I=1 TO N
400 PRINT Y(I),F(I)
410 LPRINT Y(I),F(I)
415 NEXT I
420 PRINT
430 LPRINT
450 LPRINT
460 PRINT "MEAN","SIGMA","GOODNESS OF FIT"
470 LPRINT "MEAN","SIGMA","GOODNESS OF FIT"
480 PRINT A,S,R
490 LPRINT A,S,R
500 END
```

APPENDIX (2)

COMPUTER SOLUTION TO THE EXAMPLE

NORMAL PLOT FROM ORDER STATISTICS

VALUE	MED.RANK
1.991	.0833333334
1.995	.202380952
1.997	.321428572
1.998	.440476191
2.002	.55952381
2.003	.678571429
2.005	.797619048
2.011	.916666667

MEAN	SIGMA	FIT
2.00025	6.37033904E-03	.992053571