DOCUMENT RESUME

ED 178 567 TM 009 711

AUTHOR McDermott, Paul A.; Watkins, Marley W.
TITLE Program STANDARD (Statistic of Conjoint Multiple Observer Agreement with a Standard).
PUB DATE 79
NOTE 19p.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Classification; *Computer Programs; *Goodness of Fit; Nonparametric Statistics; Observation; Rating Scales; Reliability; Statistical Analysis; *Tests of Significance
IDENTIFIERS Interrater Reliability; *STANDARD Computer Program

ABSTRACT

A computer program named Program STANDARD is presented and demonstrated. This program calculates the statistical significance of the overall agreement of the categorical assignments. The program is based on Light's statistic, G, for describing the conjoint agreement of many observers with correct or standard set of classifications on nominal scales. The present documentation includes a description of the program's purposes, input, output, language and processing, limitations and computational formulae. Also provided are instructions for preparing the job deck, a listing of the FORTRAN source deck, order of the card deck for computer processing, and a complete set of sample input and output data. (Author/CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED178567

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Program STANDARD

(Statistic of Conjoint Multiple Observer Agreement with a Standard)

by

Paul A. McDermott and Marley W. Watkins
University of Pennsylvania
University of Nebraska-Lincoln

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Paul A. McDermott
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

1979

TM009 711

SUMMARY

A computer program named Program STANDARD is presented and demonstrated. This program calculates the statistical significance of the overall agreement of the categorical assignments of two or more observers with a set of true or correct categorical assignments. The present documentation includes a description of the program's purposes, input, output, language and processing, limitations and computational formulae. Also provided are instructions for preparing the job deck, a listing of the FORTRAN source deck, order of the card deck for computer processing, and a complete set of sample input and output data.

PROGRAM DESCRIPTION, FEATURES,

AND

COMPUTATIONAL FORMULAE

ERIC

Description and Features

Introduction. Several measures of response agreement for observers' or raters' classifications on nominal scales appear to be finding increased popularity and a variety of applications among researchers and practitioners. Cohen's (1960) development of the statistic κ provided a means to determine the level and significance of the agreement between two observers in their assignments of objects or subjects to nominal scale categories. Later, Light (1971) refined and extended κ for use in the situation where multiple (i. e., more than two) observers would be making such assignments and further provided a method for testing the agreement among such multiple observers upon each specific nominal category considered. This multiple observer agreement statistic, known as κ m, was further extended by Fleiss (1971) to the case in which the sets of multiple observers would not remain identical from one case to another.

The several noted response agreement measures share two principal properties in common: first, they make no assumption that any true, correct or standard set of nominal scale classifications might exist against which the classifications offered by the various observers might be evaluated, and second, in cases where versions of κ m are applied, there being more than two observers involved, the classifications of all observers are weighted equally thus rendering a measure of the overall conjoint agreement of the observers with one another and not with any existing correct or standard set of classifications.

However, when considering response agreement among observers in applied and research settings, it is frequently desireable to test the conjoint agreement among observers relative to a standard set of classifications. This is true, for example, whenever it is necessary to assess the categorizing ability of trainee observers in the light of an expert's categorizations or when the classification accuracy of a categorical rating device must be sized up against the conjoint ratings of independent expert observers. For this purpose Light (1971) has defined the statistic G to test the significance of the conjoint agreement of many observers with a correct or standard set of classifications on nominal scales.

G is based upon a special variation of a multiple contingency table routine which first compares the obtained agreement of each of the observers' categorical choices with the standard categorical choice and thereafter compares this obtained level of agreement with what would be expected under the null hypothesis of random assignment of cases to categories. For purposes of testing statistical significance, G is distributed approximately according to the unit normal deviate.

The computer program presented and demonstrated in this paper, known as Program STANDARD, tests the statistical significance of the conjoint agreement of two or more observers with a standard set of classifications based upon Light's (1971) computational formulae for the statistic G.

Input. Each analysis requires four control cards and a data card deck as follows: (1) a title card; (2) a problem card to specify the number of cases being classified, number of categories and number of observers; (3) a pair of standard cards indicating the standard category choices for each case considered; and (4) a set of observer cards, one or two per each observer, specifying observers' category choices for each case.

Output. The information provided for each analysis includes: (1) an alphanumeric job title; (2) number of cases, categories and observers; (3) standard category choices for each case; (4) observers' category choices for each case; and (5) value of the G statistic (see Equation 1 ff., pp. 5-6) and level of statistical significance associated with the unit normal deviate.

Language and Processing. Written in FORTRAN IV, Program STANDARD is compatible with machines in the IBM 360 series. Variables are in mnemonic form according to Light's (1971) computational formulae. Input editing and output specifications are provided for user's syntactical errors.

Limitations. Currently Program STANDARD will permit up to 160 cases to be assigned by 100 or fewer observers to a maximum of 10 nominal categories.

ERIC

Computational Formulae

The Program STANDARD algorithms are as follows:

$$\underline{G} = \frac{\underline{tm} - \underline{E}(\underline{tm})}{(\xi\ \underline{tm})^{\frac{1}{2}}} \sim \underline{N}\ (0,1); \qquad (1)$$

$$\underline{tm} = \sum_{\underline{p}=1}^{\underline{m}} \sum_{\underline{i}=\underline{j}}^{\underline{C}} \underline{n}_{\underline{ij}}^{(\underline{p})}; \qquad (2)$$

$$\underline{E}(\underline{tm}) = \frac{1}{\underline{n}} \sum_{\underline{i}=\underline{j}}^{\underline{C}} [\underline{n}_{\underline{i}+} (\sum_{\underline{p}=1}^{\underline{m}} \underline{n}_{+\underline{j}}^{(\underline{p})})]; \qquad (3)$$

and

$$\xi\ \underline{tm} = \frac{1}{\underline{n} - 1} \sum_{\underline{i}=\underline{j}}^{\underline{C}} [\underline{n}_{\underline{i}+} (\sum_{\underline{p}=1}^{\underline{m}} \underline{n}_{+\underline{j}}^{(\underline{p})})] + \frac{1}{\underline{n}^2 (\underline{n} - 1)} \underline{Z}$$

$$- \frac{1}{\underline{n} (\underline{n} - 1)} \sum_{\underline{i}=\underline{j}}^{\underline{C}} [\underline{n}_{\underline{i}+}^{2} (\sum_{\underline{p}=1}^{\underline{m}} \underline{n}_{+\underline{j}}^{(\underline{p})})]$$

$$- \frac{1}{\underline{n} (\underline{n} - 1)} \sum_{\underline{i}=\underline{j}}^{\underline{C}} \underline{n}_{\underline{i}+} (\sum_{\underline{p}=1}^{\underline{m}} (\underline{n}_{+\underline{j}}^{(\underline{p})})^2) \qquad (4)$$

and where $\underline{Z}$ in the former equation equals

$$\underline{n}_{1+} [\underline{n}_{1+} \sum_{\underline{p}} (\underline{n}_{+1}^{(\underline{p})}) (\underline{n}_{+1}^{(\underline{p})}) + \underline{n}_{2+} \sum_{\underline{p}} (\underline{n}_{+1}^{(\underline{p})}) (\underline{n}_{+2}^{(\underline{p})}) + \ldots$$

$$+ \underline{n}_{C+} \sum_{\underline{p}} (\underline{n}_{+1}^{(\underline{p})}) (\underline{n}_{+C}^{(\underline{p})})]$$

$$+ \underline{n}_{2+} [\underline{n}_{1+} \sum_{\underline{p}} (\underline{n}_{+2}^{(\underline{p})}) (\underline{n}_{+1}^{(\underline{p})}) + \underline{n}_{2+} \sum_{\underline{p}} (\underline{n}_{+2}^{(\underline{p})}) (\underline{n}_{+2}^{(\underline{p})}) + \ldots$$

$$+ \underline{n}_{\underline{C}+} \sum_{\underline{p}} (\underline{n}_{+2}^{(\underline{p})}) (\underline{n}_{+\underline{C}}^{(\underline{p})})]$$

$$\vdots \qquad \qquad \vdots$$

$$+ \underline{n}_{\underline{C}+} [\underline{n}_{1+} \sum_{\underline{p}} (\underline{n}_{+\underline{C}}^{(\underline{p})}) (\underline{n}_{+1}^{(\underline{p})}) + \underline{n}_{2+} \sum_{\underline{p}} (\underline{n}_{+\underline{C}}^{(\underline{p})}) (\underline{n}_{+2}^{(\underline{p})}) + \cdots$$

$$+ \underline{n}_{\underline{C}+} \sum_{\underline{p}} (\underline{n}_{+\underline{C}}^{(\underline{p})}) (\underline{n}_{+\underline{C}}^{(\underline{p})})]. \qquad (5)$$

ERIC

PROGRAM INSTRUCTIONS

AND SOURCE LISTING

Instructions for Preparing Job Deck

The following cards must be prepared:

Card 1:

Title Card--A job title may be punched in columns 1-80

Card 2:

Problem Card--Specifications defined below. All numbers must be right-adjusted.

Columns	Punch
1-3	Number of cases (maximum = 160)
4-5	Number of categories (maximum = 10)
6-8	Number of observers (maximum = 100)

Cards 3 & 4:

Standard Cards--A standard or "correct" category choice must be indicated for each category specified in columns 4-5 of the Problem Card. If the standard category choice for the 1st case is category 1, then punch 1 in column 1. If the standard category choice for the 1st case is category 2, than punch 2 in column 1....If the standard category choice for the 1st case is category 9, then punch 9 in column 1. If the standard category choice for the 1st case is category 10, then punch 0 in column 1. Repeat standard category specifications with each column corresponding to each case specified in columns 1-3 of the Problem Card. Two standard cards must be provided. If the number of cases is less than 81, the second card must be a blank card.

Observer Cards--One or two observer cards must be punched for each case. If the problem calls for more than 80 cases, then two observer cards must be punched for each case. If the observer category choice for the first case is 1, then punch 1 in column 1. Continue as outlined in directions for Standard Cards.

For example: 5 observers classified 10 cases into 3 categories--

(Title)	EXAMPLE PROBLEM
(Problem)	01003005
(Standard 1)	1223332113
(Standard 2)	blank card
(Observer 1)	1233222123
(Observer 2)	3223311112
.	.
.	.
.	.
(Observer 5)	1233211133

Program Source Listing

```
      DIMENSION MATRIX(1000,10),SUMROW(10),SUMCLM(10),RATCLM(100,10),
     XSUBSIX(100,10),SEVEN1(10),A(10),SEVEN2(10),B(10),NRAW(100,160),
     XKEY(160),TITLE(20)
      INTEGER TM,RATCLM,NRAW,MATRIX,TITLE
      DOUBLE PRECISION PART1,ETM,ONE,TWO,THREE,FOUR,FIVE,SIX,
     XSEVEN1,SEVEN2,ZETA,SEVEN3,SEVEN,ETATM,A
      DATA TM/0/,ROW/0.00/,COLUMN/0.00/,PART1/0.00/,ETM/0.00/,ONE/0.00/,
     XTWO/0.00/,THREE/0.00/,ZETA/0.00/,FOUR/0.00/,FIVE/0.00/,SIX/0.00/,
     XSEVEN3/0.00/,SEVEN/0.00/,ETATM/0.00/,G/0.00/,SEVEN1/10*0.00/,
     XSUMROW/10*0.00/,SUMCLM/10*0.00/,A/10*0.00/,B/10*0.00/,
     XRATCLM/1000*0/,SUBSIX/1000*0.00/,MATRIX/10000*0/
      READ (5,1)TITLE
    1 FORMAT(20A4)
      READ (5,4)NCASE,NCAT,NRATE
    4 FORMAT(I3,I2,I3)
      READ (5,6)KEY
    6 FORMAT(80I1,/,80I1)
C THIS READS THE DATA
      DO 9 I=1,NRATE
      READ (5,8) (NRAW(I,J),J=1,NCASE)
    8 FORMAT(80I1)
    9 CONTINUE
      WRITE(6,800)
  800 FORMAT('1',55X,'PROGRAM STANDARD',//,32X,
     X'STATISTIC OF CONJOINT MULTIPLE OBSERVER AGREEMENT WITH A STANDARD
     X',/////,40X,'MARLEY W. WATKINS',9X,'AND',9X,'PAUL A. MCDERMOTT',
     X//,33X,
     X'UNIVERSITY OF NEBRASKA-LINCOLN',10X,'UNIVERSITY OF PENNSYLVANIA',
     X//////////)
      WRITE (6,900)TITLE
  900 FORMAT('0',20A4)
      WRITE(6,901)NCASE,NCAT,NRATE
  901 FORMAT('0',5X,'WITH',4X,I4,2X,'CASES',4X,I3,2X,
     X'CATEGORIES',4X,I4,2X,'OBSERVERS')
      IF (NCASE.GT.160)GO TO 500
      IF (NCAT.GT.10)GO TO 503
      IF (NRATE.GT.100)GO TO 505
      WRITE(6,902)
  902 FORMAT('0',/////,9X,'THE STANDARD CARD WAS READ AS')
      WRITE(6,903) (KEY(J),J=1,NCASE)
  903 FORMAT(' ',80I1)
      WRITE(6,904)
  904 FORMAT(' ',/////,9X,'THE FIRST DATA CARD WAS READ AS')
      I=1
      WRITE(6,905) (NRAW(I,J),J=1,NCASE)
  905 FORMAT(' ',80I1)
   11 CONTINUE
      DO 20 I=1,NRATE
      DO 20 J=1,NCASE
```

```
      IF(NRAW(I,J).EQ.0)NRAW(I,J)=10
      IF(KEY(J).EQ.0)KEY(J)=10
 20   CONTINUE
      NO=NCAT * NRATE
C THIS SERIES OF CARDS(N=1 THRU 60) CALCULATES THE CELLS OF
C AGREEMENT AND DISAGREEMENT AND PLACES THEM IN MATRIX
      N=1
      I=1
      M=1
 30   IF(I.GT.NCASE) GO TO 40
      J=KEY(I)
      K=NRAW(N,I)
      IF(N.GT.1) J=NCAT * (N-1) + J
      MATRIX(J,K)=MATRIX(J,K) + 1
      I=I + 1
      GO TO 30
 40   N=N + 1
      M=M + 2
      I=1
      IF (N.GT.NRATE)GO TO 60
      GO TO 30
 60   CONTINUE
      J=1
      I=1
 70   IF(I.GT.NCAT) I=1
      TM=TM + MATRIX(J,I)
C THIS CALCULATES THE SUM OF DIAGONALS(AGREEMENT CELLS)
      I=I + 1
      J=J + 1
      IF (J.GT.NO)GO TO 73
      GO TO 70
 73   CONTINUE
C     CALCULATES THE SUM OF ROW AND TOTAL ROW VALUE
      DO 100 I=1,NCAT
      DO 100 J=1,NCAT
      SUMROW(I)= SUMROW(I) + MATRIX(I,J)
      ROW= ROW + MATRIX(I,J)
100   CONTINUE
      DO 120 J=1,NCAT
      DO.120 I=1,NO
      SUMCLM(J)= SUMCLM(J) + MATRIX(I,J)
      COLUMN= COLUMN + MATRIX(I,J)
120   CONTINUE
      DO 140 I=1,NCAT
      PART1 = PART1 + (SUMROW(I) * SUMCLM(I))
140   CONTINUE
      ETM= (1./100.) * PART1
      ONE=(1./(NCASE-1.)) * PART1
      TWO=1./(NCASE**2 * (NCASE-1.))
      FOUR=1./(NCASE * (NCASE - 1.))
      DO 200 I=1,NCAT
      FIVE= FIVE + SUMROW(I)**2 * SUMCLM(I)
C CALCULATES Z
200   CONTINUE
```

```
      SIX= FOUR * FIVE
      DO 240 J=1,NCAT
      N=1
      I=1
      M=1
210   IF (N .GT. NCAT) M=M + 1
      IF (N .GT. NCAT) N=1
      IF (I .GT. NO) GO TO 240
      RATCLM(M,J)= RATCLM(M,J) + MATRIX(I,J)
      SUBSIX(M,J)= RATCLM(M,J)**2
      I=I + 1
      N=N + 1
      GO TO 210
240   CONTINUE
      L=1
      I=1
      J=1
      K=2
160   IF(L .GT. NCAT) GO TO 180
      A(J)=  (RATCLM(I,J) * RATCLM(I,L)  +  RATCLM(K,J) *
     XRATCLM(K,L)) * SUMROW(L)
      B(J)= B(J) + A(J)
      L=L + 1
      GO TO 160
180   L=1
      J= J + 1
      IF (J .GT. NCAT) GO TO 190
      GO TO 160
190   CONTINUE
      DO 195 I=1,NCAT
      ZETA= ZETA + (SUMROW(I) * B(I))
195   CONTINUE
C FINAL CALULATIONS OF G
      THREE= TWO * ZETA
      DO 260 J=1,NCAT
      DO 260 I=1,NRATE
      SEVEN1(J)= SEVEN1(J) + SUBSIX(I,J)
260   CONTINUE
      DO 270 I=1,NCAT
      SEVEN2(I)= SEVEN1(I) * SUMROW(I)
270   CONTINUE
      DO 280 I=1,NCAT
      SEVEN3= SEVEN3 + SEVEN2(I)
280   CONTINUE
      SEVEN= FOUR * SEVEN3
      ETATM= ONE + THREE - (FOUR * FIVE) - SEVEN
      G= (TM - ETM)/ETATM**.5
      GO TO 599
500   WRITE(6,501)
501   FORMAT('0','NUMBER OF CASES SPECIFIED ON PROBLEM CARD
     XIS TOO LARGE')
      GO TO 99
503   WRITE(6,504)
504   FORMAT('0','NUMBER OF CATEGORIES SPECIFIED ON PROBLEM
```

```
     XCARD IS TOO LARGE')
      GO TO 99
505   WRITE(6,506)
506   FORMAT('0','NUMBER OF OBSERVERS SPECIFIED ON PROBLEM CARD
     XIS TOO LARGE')
      GO TO 99
599   CONTINUE
      IF (ABS(G).GE.3.21)GO TO 740
      IF (ABS(G).GE.2.81)GO TO 720
      IF (ABS(G).GE.2.58)GO TO 700
      IF (ABS(G).GE.1.96)GO TO 680
      IF (ABS(G).GE.1.76)GO TO 660
      IF (ABS(G).GE.1.65)GO TO 640
      IF (ABS(G).LT.1.65)GO TO 600
600   WRITE(6,80)G
 80   FORMAT('0', 9X,'G OF',F9.4,'------ NON-SIGNIFICANT')
      GO TO 99
640   WRITE (6,84)G
 84   FORMAT('0', 9X,'G OF',F9.4,'------ P LESS THAN .10')
      GO TO 99
660   WRITE (6,86)G
 86   FORMAT('0', 9X,'G OF',F9.4,'------ P LESS THAN .08')
      GO TO 99
680   WRITE (6,88)G
 88   FORMAT('0', 9X,'G OF',F9.4,'------ P LESS THAN .05')
      GO TO 99
700   WRITE(6,90)G
 90   FORMAT('0', 9X,'G OF',F9.4,'------ P LESS THAN .01')
      GO TO 99
720   WRITE (6,92)G
 92   FORMAT('0', 9X,'G OF',F9.4,'------ P LESS THAN .005')
      GO TO 99
740   WRITE (6,94)G
 94   FORMAT('0', 9X,'G OF',F9.4,'------ P LESS THAN .001')
 99   STOP
      END
```

Order of Card Deck for Computer Processing

The job deck for each analysis is arranged sequentially as follows:

A. Computer systems job classification cards.

B. Program STANDARD source deck.

C. Computer systems end of source deck card.

D. Program STANDARD title card.

E. Program STANDARD problem card.

F. Program STANDARD standard card 1.

G. Program STANDARD standard card 2.

H. Program STANDARD observer card deck.

I. Computer systems end of job card.

SAMPLE INPUT AND OUTPUT DATA

Input for Example Problem

```
TEST RUN WITH LIGHT'S (1971) DATA
10003002
11111111111111111112222222222222222222222222222223333333333333333333333333333333
33333333333333333123
11111111111122222231111222222222223333333333333311111111111111111111111122223333
33333333333333333333
11111111113333333221111111111111112222222222222211111122222222222222222222223333
33333333333333333233
```

Output for Example Problem

PROGRAM STANDARD

(STATISTIC OF CONJOINT MULTIPLE OBSERVER AGREEMENT WITH A STANDARD)

TEST RUN WITH LIGHT'S (1971) DATA

WITH 100 CASES 3 CATEGORIES 2 OBSERVERS

THE STANDARD CARD WAS READ AS
11111111111111111111222222222222222222222222222233333333333333333333333333333333
33333333333333333123

THE FIRST DATA CARD WAS READ AS
11111111111112222231112222222222223333333333333311111111111111111111111122223333
33333333333333333333

G OF 3.7640------ P LESS THAN .001

ERIC

REFERENCES

Cohen, J.A. A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 1960, 20, 37-46.

Fleiss, J.L. Measuring nominal scale agreement among many raters. Psychological Bulletin, 1971, 76, 378-382.

Light, R.J. Measures of response agreement for qualitative data: Some generalizations and alternatives. Psychological Bulletin, 1971, 76, 365-377.