A SEQUENTIAL ANALYSIS OF HAILFALL DATA FITTED TO A GAMMA DISTRIBUTION FUNCTION

By

William L. Cox
Participant
Research Participation Program
Sponsored by
National Science Foundation

Assigned to
Hailstorms in the High Plains
Project Sponsored by
National Science Foundation
Grant NSF G-23706

Dr. Richard A. Schleusener
Supervising Scientist
Civil Engineering Section
Colorado State University
Fort Collins, Colorado

August 1963

CER63WLC33

ABSTRACT

Three computer programs which were written to aid the statistical analysis of hailfall data are discussed. Gamma distribution parameters of raw data, scale changed, and non-scale changed data are determined by the first program. The goodness of fit is determined in the second program using a chi square test. The last program performs a sequential analysis test on data as it is collected. This is done by assuming certain Type I and Type II errors and testing various scale and non-scale changes of the data. Square root of the dents, cube root of the dents, and logarithm of the indicator energy numbers (hail fall data) were analyzed using these programs. The results of these tests and detailed notes on each program are presented.

TABLE OF CONTENTS

A SEQUENTIAL ANALYSIS OF HAILFALL DATA FITTED TO A GAMMA DISTRIBUTION FUNCTION

By
William L. Cox

INTRODUCTION

During the past four years (1960-1963) data on severe storms in the northeastern Colorado plains area have been collected from a hail network operated by Colorado State University. Basic information on the nature and characteristics of hailfalls has been obtained and published previously, (1) and (2).* This paper presents the results of work in the statistical design of an experiment leading to an evaluation of hail suppression by cloud seeding. The work was completed during the summer of 1963 while the author was on a ten-week Research Participation Program sponsored jointly by the National Science Foundation and Colorado State University.

OBJECTIVES

The principal objectives of this work were, utilization of Colorado State University's IBM 1620 digital computer to:

A. Develop a program to calculate gamma distribution parameters for raw hail data and data on which various scale and non-scale changes were performed.

---

*Numbers in parentheses refer to appended references.

B. Write a chi square "goodness of fit" program that would indicate agreement or lack of agreement between the observed hail data and a theoretical gamma distribution.

C. Write a sequential analysis program which would assume various Type I and Type II errors and test hail data as it is gathered.

GENERAL DISCUSSION

Three separate programs were written and debugged until all programs operated satisfactorily on the CSU computer. The source program printouts, notes on each program, and examples of the output are in Appendices I through III. Computer runs utilizing the square root of the dents and cube root of dents on hail indicators and the logarithm of energy numbers on hail indicators have been completed.

In program Number 1, the determination of gamma parameters, hail data was read in, 0.01 added to it, (to eliminate zero data) and the desired scale or non-scale change was performed on the data. Scale and non-scale changes were produced by coefficients punched on header cards. The two gamma parameters were calculated for each set of data with various scale and non-scale changes. Running time for this program with 720 data cards was approximately ten minutes.

Program Number 2 compares observed with expected values using a chi square goodness of fit test. Expected values were calculated using X values from 0.01 to 6.01, the gamma parameters computed from program Number 1, and the theoretical probability density function. The observed values were grouped into classes, and the chi square value computed. If the calculated chi square value exceeded the tabulated value, the hypothesis that the sample was drawn from the assumed distribution was rejected. Running time for program Number 2 with 720 cards was 14 minutes.

In program Number 3, the sequential analysis test, values for alpha and beta (Type I and Type II errors), the cumulative sum of the data, and the total number of observations were read into computer memory. Values for accepting or rejecting the hail suppression hypothesis were computed as was the cumulative sum of the data. The sum of the data was then compared with each of these values. A decision to accept or reject the hypothesis or to continue testing was made from this comparison. Running time on this program depended upon the number of observations required to make an accept or reject decision. For 30 observations running time was approximately three minutes.

RESULTS

Program No. 1.

Thirty computer runs were completed with this program. Sixteen of these were made using square root of dents data, ten using the sample rainfall data given in (3), one using the cube root of dents data, and three using logarithm of the indicator energy numbers. The mean values for the gamma distribution parameters are given in Table 1.

Program No. 2.

Four runs were made on the square root of dents data. These included two 0 and one 10 percent scale change and one with a level III non-scale change. One run using 0 scale changed cube root of dents data and two runs using logarithm of the indication energy numbers with 0 and 25 percent scale changes were completed. Results of these runs along with the tabulated values of chi square for four degrees of freedom and a 0.05 confidence level are presented in Table 2.

TABLE 1

GAMMA DISTRIBUTION PARAMETERS

Type of Data	No. of Runs	Type of Change	Q	QL
$\sqrt{\text{Dents}}$	4	0% Scale	4.2209	2.6830
$\sqrt{\text{Dents}}$	1	5% Scale	4.2210	2.8242
$\sqrt{\text{Dents}}$	3	10% Scale	4.2209	2.9811
$\sqrt{\text{Dents}}$	1	15% Scale	4.2210	3.1565
$\sqrt{\text{Dents}}$	2	25% Scale	4.2210	3.5774
$\sqrt{\text{Dents}}$	2	Level I Non-Scale[1]	3.3898	2.7660
$\sqrt{\text{Dents}}$	1	Level II Non-Scale	3.7553	2.8243
$\sqrt[3]{\text{Dents}}$	1	0% Scale	6.5017	3.2107
Ln.E.	1	0% Scale	0.9292	0.4817
Ln.E.	1	10% Scale	0.9292	0.5412
Ln.E.	1	25% Scale	0.9292	0.6494

NOTES

1. The non-scale changes were:

Data	Level I	Level II	Level III
0 to 1 incl.	30%	20%	15%
1 to 2	25	17	13
2 to 3	20	14	11
3 to 4	15	11	9
4 to 5	10	8	7
5 to 6	5	5	5

TABLE 2

COMPUTED VALUES FOR CHI SQUARE

Type of Data	Q	QL	Type change	Computed X^2	Tabulated X^2
$\sqrt{\text{Dents}}$	4.2209	2.6830	0% Scale	18.54	9.49
$\sqrt{\text{Dents}}$	4.2209	2.6830	0% Scale	18.54	9.49
$\sqrt{\text{Dents}}$	4.2209	2.9811	10% Scale	13.41	9.49
$\sqrt{\text{Dents}}$	3.9140	2.8268	Level III	13.90	9.49
$\sqrt[3]{\text{Dents}}$	6.5017	3.2107	0% Scale	275	9.49
Ln.E.	0.9292	0.4871	0% Scale	173	9.49
Ln.E.	0.9292	0.6494	25% Scale	237	9.49

Program No. 3

Five runs were made using square root of the dents data and various Type I and Type II errors. The data cards were either reshuffled or different cards used in each run. The results of these runs are presented in Table 3.

TABLE 3

NUMBER OF OBSERVATIONS REQUIRED FOR DECISION IN SEQUENTIAL ANALYSIS TEST

Alpha	Beta	Type of Data	Type of Change	No. of obs. Req'd for Decision
0.010	0.050	$\sqrt{\text{Dents}}$	25% Scale	20
0.010	0.050	$\sqrt{\text{Dents}}$	25% Scale	27
0.010	0.050	$\sqrt{\text{Dents}}$	25% Scale	22
0.010	0.010	$\sqrt{\text{Dents}}$	10% Scale	32
0.010	0.010	$\sqrt{\text{Dents}}$	10% Scale	64

DISCUSSION OF RESULTS

A. Program No. 1.

The consistent results of duplicate runs with this program indicate that the results were easily reproducible. Shuffling the input data deck does not effect the results, indicating that round-off errors were negligible. The parameters obtained with this program using the rainfall data suggested in (2), agree very closely with those presented in the paper. It is believed that the results of this program are quite reliable. The program is working satisfactorily.

B. Program No. 2.

The output from this program continues to indicate F3-type errors which are caused by "negative arguments in a logarithmic operation". However, there are no logarithms used in the program, and the source of this error indication is not known. It is felt that the results of the program are not effected by this error. The rainfall data mentioned above was also used in this program, producing the expected results. Runs with two different sets of class boundaries have produced essentially the same results for the square root of dents data. It appears that the square root of dents data do not fit the gamma distribution function. Results of the single run with the cube root of the dents should be considered only as a tenative indication, until further runs with this data can be completed. A slide rule approximation had to be used in determining QM1F, and a more accurate calculation of this value might change the computed chi square value. Both runs which utilized the logarithm of the energy number data produced very large computed chi square values. Thus, it is concluded that raw logarithm of the energy number data do not adequately fit a gamma distribution.

C. Program No. 3.

Although square root of the dents data do not fit the gamma distribution, they were used to check out this program. Duplicate results were obtainable with the program; however, shuffling the data deck or using a different series of data cards can significantly alter the results (see the 10 percent scale change in Table 3 for example). It appears that the results of this program are satisfactory and that the program is working as expected.

CONCLUSIONS

On the basis of this study it can be concluded that:

A. The gamma distribution parameters for the square root of dents, cube root of dents and logarithm of the energy numbers are as listed in Table 1.

B. Unmodified square root of dents data do not adequately fit a gamma distribution function.

C. Unmodified logarithm of the energy numbers data do not adequately fit a gamma distribution function.

D. The sequential analysis program cannot be used with square root of dents or logarithm of energy numbers data because of conclusion C.

E. To use the sequential analysis program, data must not only fit the gamma distribution, but also one of the distribution parameters, namely Q, must be constant for scale changes.

RECOMMENDATIONS

Future work in this area should include:

A. Consideration of other transformations for the dents data which might fit a gamma distribution.

B. Additional study of the cube root of dents data to accurately determine if it fits a gamma distribution.

C. Consideration of other types of hail data, or modification of data such as energy numbers and square root of dents, which might fit or be made to fit a gamma distribution.

D. Extension of the sequential analysis test to include non-scale changes (providing data fitting a gamma distribution can be found).

E. Consideration of other statistical tests, including those which do not involve any assumption of the distribution parameters, to evaluate the hail suppression effect.

LIST OF REFERENCES

1. Schleusener, Richard A. and Grant, Lewis O. Characteristics of hailstorms in the Colorado State University network, 1960-61. Proceedings of the Ninth Weather Radar Conference, 1961.

2. Schleusener, Richard A. and Marwitz, John D. Analysis of data on hailfalls as background for the design of an experiment in hail modification. Prepared for the Severe Local Storm Conference, November, 1963.

3. Hartley, H. O. and Lewish, W. T. Fitting of data to the two parameter gamma distribution with special reference to rainfall data. Iowa State University of Science and Technology, Ames, Iowa. June 29, 1959.