ON THE DISTRIBUTION OF STATISTICS SUITABLE FOR EVALUATING RAINFALL STIMULATION EXPERIMENTS

by

K. R. Gabriel, Hebrew University, Jerusalem
and University of North Carolina, Chapel Hill
and
Paul Feder, Stanford University

Institute of Statistics Mimeo Series No. 520

April 1967

This work was supported in part by the National Institutes of Health (Institute of General Medical Sciences) Grant No. GM-12868-03.

DEPARTMENT OF BIOSTATISTICS

UNIVERSITY OF NORTH CAROLINA

Chapel Hill, N. C.

ON THE DISTRIBUTION OF STATISTICS SUITABLE FOR EVALUATING RAINFALL STIMULATION EXPERIMENTS*

by

K. R. Gabriel, Hebrew University, Jerusalem
and University of North Carolina, Chapel Hill

and

Paul Feder, Stanford University

ABSTRACT

Two randomization tests of the null hypothesis in cloud seeding experiments are compared - the Wilcoxon-Mann-Whitney test and a test based on an average ratio of seeded to non-seeded amounts of precipitation. Data from the Israeli experiment suggest that the latter test is relatively more sensitive to apparent effects of seeding. The significance level of this test may be estimated by Monte Carlo methods or approximated by using the asymptotic normal distribution of the average ratio. Sampling trials show that this approximation is adequate only when the experiment is of several years' duration.

1. PROBLEMS OF STATISTICAL EVALUATION OF RAINFALL STIMULATION EXPERIMENTS.

The statistical evaluation of randomized rainfall stimulation experiments raises problems of choice of valid and suitable methods of analysis. Most standard statistical techniques are not applicable to such experiments since precipitation data do not generally fit the common textbook assumptions. Distributions of amounts of precipitation, especially for short periods such as 24 hours, are usually discontinuous at zero and have most pronounced tails. The discontinuity is due to a positive probability of no rainfall at all, and the tails show the presence of occasional very extreme amounts. Furthermore, measurements of precipitation are not independent either in space or in time, but there exist strong

* This work was supported in part by the National Institutes of Health (Institute of General Medical Sciences) Grant No. GM-12868-03.

correlations between stations which are scores of miles apart as well as appreciable dependence of precipitation on successive days. This lack of "good behaviour" of the data is especially troublesome when complex experimental designs are used, such as cross-over designs with random daily allocation of seeding to alternate areas and when detailed and sensitive statistical analyses are required, e.g., when concomitant variables are to be taken into account.

A further difficulty in applying standard statistical techniques is that the form of possible seeding effects is not known but may well be most irregular. Considerable differences in the outcomes of different rainfall stimulation experiments point to the existence of factors, not hitherto identified, which may sometimes further the effectiveness of cloud seeding but perhaps inhibit it at other times (Neyman and Scott [8]). A number of findings suggest that seeding may be highly effective on some occasions but have little or no effect on many other occasions (Siliceo et al. [10], Gabriel [4]). Since so little is known about the alternative one should be testing against, it is not only doubtful whether standard techniques are valid but it is difficult to decide what a good technique is. (For the derivation of optimal techniques under certain simple assumptions, see Neyman and Scott [7]).

Unless a satisfactory parametric model of precipitation becomes available, i.e., one which takes account of all the irregularities and dependences noted above, the safe course is to use randomization tests. (For an earlier discussion of the need for non-parametric tests, see Adderley [1]). Such tests compare a summary statistic based on the experimental results under the actual randomized allocation of treatments with all possible values this statistic might have assumed for the same experimental results had the allocation been different. To be specific, for each possible allocation under the randomization scheme, the observed experimental data are considered afresh and the resulting statistic compared with that found under the allocation used. This is a valid comparison under the null hypothesis that seeding has no effect, for in that case the same rainfalls would have occurred, no matter what the allocation. Clearly, randomization tests do not require any assumptions about the distribution and dependence of the precipitation data, and therefore provide valid analyses of rainfall experiments.

Different randomization tests are obtained by using this principle with different statistics. For example, one may take the difference between mean values on treated observations and on control observations, and compare the experimental difference with similar differences obtained by other allocations of the same data. Alternatively, one might take the difference between medians, or mean ranks, or proportions of observations above some constant, etc.. Each comparison statistic will yield a randomization test. For certain statistics the distribution over all allocations can be derived mathematically, and critical values for significance testing have been computed and tabulated. A well-known example is the WMW (Wilcoxon-Mann-Whitney) test whose statistic is the number of pairs consisting of a treated and a control observation, for which the treated observation has a larger variable value than the control observation.

2. THE TWO TESTS IN THE ISRAELI EXPERIMENT.

The Israeli rainfall stimulation experiment uses a cross-over design with daily random allocation of seeding to either the North or the Centre of Israel. (For detailed descriptions of the experiment, see Gabriel [3], [4], [5]).

Most analyses of this experiment are confined to rainy days - defined as having some precipitation in the always unseeded buffer area between the North and the Centre. In effect some 98% of all precipitation occurs on such days so that the omission of the other, dry, days is unlikely to hide any seeding effects.

For a day indexed by subscript i, out of a total of N days, x_i and y_i denote the mean amounts of precipitation per station in the North and the Centre respectively. The random allocation variable Θ_i is defined as 1 if seeding is allocated to the North and as 0, i.e., $1 - \Theta_i = 1$, if seeding is allocated to the Centre. With this notation, the WMW test uses the count statistic

$$U = \sum_{i=1}^{N} \sum_{j=1}^{N} \Theta_i(1-\Theta_j)\phi_{ij} \tag{1}$$

where ϕ_{ij} indicates whether the ith day's variable exceeds the jth day's variable. In the Israeli experiment $x_i - y_i$ was chosen as experimental variable so that

$$\phi_{ij} = \begin{cases} 1 & \text{if } x_i - y_i > x_j - y_j \\ \frac{1}{2} & \text{if } x_i - y_i = x_j - y_j \\ 0 & \text{if } x_i - y_i < x_j - y_j . \end{cases} \tag{2}$$

Note that even though the $x_i - y_i$ are not independent and identically distributed under the hypothesis of no seeding effects, the WMW test is still valid since the seeding allocation was determined by randomization.

This statistic is easily computed and its distribution has been tabulated in detail*. The WMW test can, therefore, be used in over-all analyses of entire

* See, for example, the tables by Owen [9]. For large samples an asymptotic approximation is available.

experiments as well as for detailed investigations which require testing in each of many categories of days. Such detailed analyses are of considerable importance in increasing sensitivity by using categories defined by well correlated concomitant variables (Gabriel [4]) and in providing breakdowns from which one may learn under what conditions seeding may be more or less effective (Neyman and Scott [8]).

Alternatively, a randomization test may be based on a quantitative measure of apparent seeding effect. The choice of a suitable measure will be guided by simplicity and by what one considers relevant as a useful or economically valuable effect. A ratio of amounts of precipitation under seeding to amounts in the absence of seeding may serve this purpose. In cross-over designs such a ratio can be obtained as the geometric mean of the seeded to unseeded ratios of both experimental areas. In the above notation, one may write the total amounts as

$$\text{in North:} \quad \sum_{i=1}^{N} \theta_i x_i \quad \text{seeded;} \quad \sum_{i=1}^{N} (1-\theta_i) x_i \quad \text{unseeded,}$$

$$\text{in Centre:} \quad \sum_{i=1}^{N} (1-\theta_i) y_i \quad \text{seeded;} \quad \sum_{i=1}^{N} \theta_i y_i \quad \text{unseeded.}$$

This average ratio is defined as R, where*

$$R^2 = \left[\sum_{i=1}^{N} \theta_i x_i \Big/ \sum_{i=1}^{N} (1-\theta_i) x_i \right] \left[\sum_{i=1}^{N} (1-\theta_i) y_i \Big/ \sum_{i=1}^{N} \theta_i y_i \right] \tag{3}$$

(R has been referred to as the Root-Double-Ratio by the Australian cloud seedint team, who seem to have been the first to use it - Adderley and Twomey [2]). One notices that each day's $\theta_i = 1$ or $1 - \theta_i = 1$ appears twice in the expression

---

* If all θ's are equal, R may be arbitrarily defined as 1.

for R, once in the numerator with one of the variables x, y and once in the denominator with the other variable. And since precipitation in the North and the Centre, i.e., x and y, is highly correlated ($r=.8$), this ensures that variations due to the random allocation of seeding will not greatly affect the statistic R. (An approximation to R is discussed in the Appendix).

The statistic R is an average ratio, but its use must not be understood to imply that the actual effect of cloud seeding on precipitation is multiplicative. R is an average, and a particular value of R might be the result of a variety of effects on different days of the experiment.

In comparing the two statistics it is clear that R has the advantage of depending directly on amounts of seeded versus unseeded rainfall, whereas U depends on these amounts only indirectly through the ranking of the differences $(x_i-y_i) - (x_j-y_j)$. If none of these differences are particularly large, either positively or negatively, as compared with the rest, it will not matter much that ranks are used instead of actual values, and the WMW statistic may be as appropriate as R but simpler to use. Indeed, the WMW test is known to be powerful in many standard situations. For these reasons the WMW test was originally chosen for the analysis of the Israeli experiment.

If, however, effects of seeding vary a great deal from day to day, i.e., if seeding has little or no effect on most days but on a few days it has very large effects, this will hardly affect the ranking of the $(x_i-y_i) - (x_j-y_j)$ differences even though some of them become extremely large. The WMW test will be quite insensitive to effects of this kind, but the R statistic will show them clearly. A hypothetical example will illustrate this:

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
θ_i	1	0	0	1	0	0	1	1	0	1	0	0	1	1	0	0
x_i	5	7	3	4	6	9	25	7	5	6	3	5	17	8	7	7
y_i	6	6	12	4	5	7	8	8	4	8	10	4	5	8	5	7

$$\sum_{i=1}^{16} \theta_i = 7, \quad \sum_{i=1}^{16} \theta_i x_i = 72, \quad \sum_{i=1}^{16} (1-\theta_i)x_i = 52, \quad \sum_{i=1}^{16} \theta_i y_i = 47, \quad \sum_{i=1}^{16} (1-\theta_i)y_i = 60$$

$$R = \sqrt{(72/52)(60/47)} = 1.33 \qquad U = 29.$$

(Approximation - see Appendix - $1 + 2(S-T) = 1 + 2(\frac{72}{72+52} - \frac{47}{47+60}) = 1.28$).
The median values under the null seeding effect hypothesis are Med $R = 1$ and Med $U = 7 \times 9/2 = 31.5$, so that the two statistics deviate from their medians in different directions. R indicates positive seeding effects because of the few apparently large effects (on days 3, 7, 11 and 13) whereas U indicates the contrary because all twelve other days had apparent small negative effects or no effects at all.

Another drawback of the WMW technique is that it is merely a test of significance and does not provide a quantitative estimate of the size of seeding effects. To obtain estimates and confidence bounds one must have recourse to rather cumbersome iterative techniques whose results lean heavily on one's assumptions regarding the form (e.g., additive, multiplicative, etc.) of the effects - assumptions which unfortunately have little to be based upon (Gabriel [4], section 5.1).

Variation of the statistic R from allocation to allocation depends on the actual amounts of precipitation observed during the experiment. It can therefore not be studied generally as was the U statistic of the WMW test.

A complete enumeration of values of R obtained under all possible allocations is not practical either, even with an electronic computer (with 300 rainy days 2^{300} values would have had to be computed). However, one may use the Monte Carlo technique to sample from all these allocations and thereby obtain an estimate of the probability of exceeding the actually observed R, R_0 say, by chance. Such sampling is readily performed on an electronic computer. For each sample a (pseudo) random allocation is generated and the experimental data used to compute the sample value of R. The proportion of sample R values which exceed the experimental value R_0 is an estimate of $\alpha = P(R > R_0)$. Simple random sampling theory applies, so that the estimate is unbiased and has variance $(1-\alpha)\alpha/n$ where n is the number of samples generated. A few hundred samples usually suffice to give a fairly good idea of the level of significance and the cost of generating them on a modern computer is negligible as compared with the expenses involved in a cloud seeding experiment.

For the data of the entire Israeli experiment, as well as for a number of categories of days separately, Monte Carlo trials of several hundred samples each were run on the Hebrew University's I.B.M. 7040 computer. The resulting estimates of significance levels by the randomization test are compared in Table 1 with those obtained by the WMW test. (The "R-asymptotic" levels of significance of Table 1 are discussed in section 3, below. The 1 + 2(S-T) approximation to R is discussed in the Appendix).

Each row of Tables 1 and 2 contains the results of different and independent Monte Carlo randomization samples, but the same daily rainfall data was used repeatedly. Thus, there are 400 and then another 200 independent randomizations on the same 1961-5 data, 500 more randomizations run after the 1965/6 season's results were added to the data, and a further 400 randomizations for the same

Table 1. Significance Levels Attained by Different Tests - Rainy Days Only*

	Number of Days	Level of Significance			Observed R	Observed 1+2(S-T)	Number of Permutations
		W-M-W	R-Randomization	R-Asymptotic			
1961-6	327	.057	.004	.004	1.190	1.174	500
1961-6 (Interior of Regions)	327	.009	.000	.0003	1.273	1.240	400
1961-5	281	.190	.03	.024	1.155	1.144	400
			.04				200
1961	26	.140	.0725	.011	1.737	1.492	400
1961/2	57	.387	.21	.173	1.185	1.164	400
1962/3	49	.532	.34	.242	1.102	1.095	400
1963/4A	20	.605	.47	.463	1.017	1.017	400
1963/4B	57	.032	.0175	.008	1.376	1.317	400
1964/5	72	.615	.175	.164	1.131	1.123	400
1965/6	46	.015	.0025	.001	1.544	1.391	400
1961-2	83	.263	.14	.082	1.244	1.215	200
1962-4	126	.165	.12	.074	1.154	1.143	200
1964-6	118	.133	.02	.014	1.244	1.213	200
1961-6 Buffer-South < -10	26	.123	.130	.085	1.138	1.126	300
" " -10 ≤ < - 5	22	.178	.087	.055	1.281	1.243	300
" " - 5 ≤ < - 1	27	.055	.027	.008	1.395	1.309	300
" " - 1 ≤ < 0	24	.074	.120	.029	1.415	1.209	300
" " 0 ≤ < 1	81	.855	.130	.058	1.433	1.355	300
" " 1 ≤ < 3	46	.843	.683	.685	0.918	0.936	300
" " 3 ≤ < 5	20	.741	.820	.813	0.709	0.700	300
" " 5 ≤ < 10	35	.008	.047	.008	1.394	1.293	300
" " 10 ≤ < 20	23	.019	.000	.000	1.792	1.507	300
" " 20 ≤	23	.189	.220	.163	1.177	1.159	300

* A detailed description of the dates and groupings in these tables is given elsewhere (Gabriel [4], [5]) and is not repeated here as it is not directly relevant to the question of choice of statistics. However, it may be mentioned that the annual periods are mid-October to mid-April (except 1961 which started in February). The split of the 1963/4 season was due to a change in the definition of the operational day from an 8 p.m. start and end to an 8 a.m. start and end of seeding.

days' rainfall in the interior areas which is highly correlated with that in the entire areas used before. Clearly, in as far as effects of cloud seeding are apparent, and one test is found more sensitive than another, these different calculations give dependent and very similar results. However, as regards the form of the distribution of the sample R values, the results are independent from one row to another, each being based on separate Monte Carlo samples. The same remarks apply also to the further repeated use of the same rainfall data in each of the next four parts of the Tables. As far as Table 1 is concerned these are repetitious uses of the same data and must not be mistaken as independent cumulative evidence. The rows and parts of Table 2, on the other hand, do provide independent information on the distribution of R under randomization.

For all seasons together as well as for each season by itself, the significance levels for the WMW test are seen to exceed the estimates of α for the R randomization test. In so far as the seeding effects observed in this and other experiments are real and not mere random fluctuations in rainfall, the difference in the behaviour of the two tests would indicate that the WMW test is indeed less sensitive to this type of effect than the test using the average ratio. One would conclude that the WMW test, though valid for rainfall stimulation experiments, is less powerful than the R-randomization test. Clearly, this conclusion is meaningful only if there really are seeding effects, whereas if seeding were completely ineffectual, no one test could be more powerful than another.

Table 1 also shows similar comparisons of levels within categories of days defined by means of a concomitant variable that is correlated with x-y but is unaffected by seeding. This variable is the difference in precipitation amounts between the unseeded buffer and South areas, which lie, respectively, between the North and Centre experimental areas and South of the Centre area.

No consistent pattern is found between the two levels of significance. For some categories the WMW level is higher, for other categories the R level is higher. In so far as real seeding effects exist and the above conclusions about the relative sensitivity of the two tests are meaningful, one would further conclude that the greater sensitivity of the R-randomization test holds only when all types of days are tested together. Within relatively homogeneous categories of days, neither test would appear consistently superior to the other.

3. THE ASYMPTOTIC DISTRIBUTION OF R AND ITS USE.

When the number N of experimental observations is large, the sampling distribution of R may be approximated by asymptotic theory. It is shown in the Appendix that for large N the distribution of R under random allocation of seeding tends to normality with expectation 1 and variance

$$\mathrm{Var}_0(R) = \sum_{i=1}^{N} (x_i/X - y_i/Y)^2 , \tag{4}$$

where

$$X = \sum_{i=1}^{N} x_i \quad \text{and} \quad Y = \sum_{i=1}^{N} y_i . \tag{5}$$

This asymptotic result holds provided no single value of either x_i^2/X^2 or y_i^2/Y^2 or $(x_i/X - y_i/Y)^2$ remains appreciable as N increases. In other words, it holds unless the rainfall amounts x, y and differences $\frac{x}{X} - \frac{y}{Y}$ on a very few days dominate the total amounts and differences over the entire experiment even when the experiment becomes increasingly long. (The exact conditions are given in the Appendix, equations (15) and (18)).

To find out whether this assumption is tenable for precipitation data or whether it is vitiated by the occasional occurrences of extreme amounts of rainfall, the distribution of sample R values was compared to the asymptotic normal distribution. Sample R values were obtained by the same Monte Carlo sampling described above in connection with Table 1. The results are presented in Table 2 for data of the entire Israeli experiment as well as for several separate categories of days.

Tests of goodness of fit of the asymptotic normal distribution with mean 1 and variance $\sum_{i=1}^{N} (x_i/X-y_i/Y)^2$ show no significant deviations when the data for the entire 1961-5 or 1961-6 period of the experiment are analyzed. However, when R values are obtained separately for each season, the fit is significantly poor - sum of chi-square statistics 389.00 with 133 d.f.. For pairs of seasons the fit is not good - sum of chi-square statistics 73.40 with 57 d.f.. One may conclude that asymptotic theory gives a good approximation when N is as large as 300 rainy days (5-6 seasons), that it is doubtful for 100-odd rainy days (2 seasons), and that it is clearly inadequate for as few as 50 or so days (single seasons).

Monte Carlo trials have also been carried out within ten categories of days defined by the buffer-South difference. The number of days per category was about N = 30, and the fit of the sample R distribution to the asymptotic normal was very poor - sum of chi-square statistics 427.4 with 190 d.f.. Clearly, for this size of N the asymptotic approximation is inadequate even within relatively homogeneous categories of days.

Further details of the sampled distributions of R are also presented in Table 2. These give some idea of how the sampled distributions of R deviate from the asymptotic normal. The distributions of R have slight positive skewness and a small positive bias in the expectation. The bias is of the order of 2% for single seasons and of less than 1% for the whole length of the experiment.

Table 2. Sampled Distributions of R, Goodness of Fit and Other Characteristics (Rainy Days Only)

	Number of Days	Number of Permutations	Chi-square for goodness of fit 19 d.f.†	Prop. $(R \geq 1)$	Mean	Variance	Asymptotic variance	α_3	α_4
1961-6	327	500	29.2	.52	1.0051	.0052	.0053	.006	2.81
1961-6*	327	400	14.1	.51	1.0034	.0066	.0063	.014	3.16
1961-5	281	400	21.1	.51	1.0075	.0061	.0061	.025	3.36
		200	23.4	.52	1.0042	.0064	.0061	.027	2.96
1961	26	400	153.7	.49	1.0739	.1596	.1018	.295	2.91
1961/2	57	400	50.3	.48	1.0221	.0519	.0387	.174	3.96
1962/3	49	400	62.9	.52	1.0214	.0262	.0212	.057	2.59
1963/4A	20	400	37.7	.51	1.0208	.0473	.0352	.092	3.34
1963/4B	57	400	35.1	.50	1.0117	.0285	.0243	.055	2.78
1964/5	72	400	18.1	.49	1.0025	.0200	.0180	.055	2.99
1965/6	46	400	31.2	.52	1.0131	.0316	.0298	.053	2.57
1961-2	83	200	29.8	.52	1.0312	.0367	.0307	.066	2.64
1962-4	126	200	21.2	.56	1.0163	.0130	.0113	.018	2.84
1964-6	118	200	22.4	.52	1.0152	.0129	.0122	.028	2.46
1961-6	26	200	23.0	.46	0.9985	.0114	.0101	.028	2.68
	22	200	32.4	.52	1.0226	.0377	.0310	.105	3.35
10 categories of rainy days according to buffer-south differences (See Table 1)	27	200	27.4	.50	1.0269	.0335	.0267	.108	3.29
	24	200	94.6	.46	1.0293	.0940	.0481	.192	2.52
	81	200	77.8	.50	1.0435	.1170	.0761	.202	2.72
	46	200	11.4	.52	1.0166	.0295	.0287	.016	2.65
	20	200	58.0	.44	1.0497	.1666	.1074	.429	3.95
	35	200	56.6	.56	1.0411	.0341	.0269	.122	3.09
	23	200	20.8	.46	1.0056	.0398	.0378	.103	3.02
	23	200	25.4	.54	1.0285	.0406	.0323	.064	2.94

* Amounts of precipitation in interior parts of areas only.

† Each distribution of Monte Carlo sample values of R was sorted into 20 classes for testing of goodness of fit.

On the other hand, the proportions of R values above 1 do not deviate systematically from 1/2. These characteristics are readily understood when one considers that R is a ratio which may assume any non-negative value and that the chance of any one value is equal to that of its reciprocal. Thus, 1 must be the median, but there is positive skewness and the expectation exceeds 1.

More crucial to the application of the asymptotic distribution is the difference between $\mathrm{Var}_0(R)$ and the true variance of R. Table 2 clearly shows that the sample estimates exceed the asymptotic expression for almost all the Monte Carlo trials that were run. For single seasons and for particular buffer-South cateogries of days the variance appears to be 20-40% larger than $\mathrm{Var}_0(R)$. For pairs of seasons it is 10-20% larger, but for the entire length of the experiment it is very close to the asymptotic value.

The standardized fourth moment shows no further systematic deviation from normality.

In applying asymptotic theory to practical testing of significance, one would compute the normalized statistic

$$Z = (R-1)/\sqrt{\mathrm{Var}_0(R)} \qquad (6)$$

and enter it in a table of the normal probability integral. The resulting "R - asymptotic" levels of significance are compared in Table 1 with the unbiased estimates obtained by Monte Carlo sampling. For all comparisons except those of the entire length of the experiment, this asymptotic normal method is seen to underestimate α and makes the results appear more highly significant than they really are. This underestimate of α results from the underestimate of the variance of R in the denominator of the above expression for Z.

For finite N the moments of R might be approximated somewhat better by using the formulas

$$\mathcal{E}_1 R = 1 + \mathrm{Var}_0(R)/2 \qquad \mathrm{Var}_1(R) = \mathrm{Var}_0(R)\mathcal{E}_1 R. \tag{7}$$

Checking of these approximations against the sample moments in Table 2 shows that they deviate systematically somewhat less but in the same direction as the previous approximations, $\mathcal{E}_0 R = 1$ and $\mathrm{Var}_0(R)$ of (4). As far as the estimate of the variance is concerned, the new approximations do not reduce the bias much and are therefore little better than those used in (4).

We conclude that the asymptotic normal distribution may be applied safely only when data of some 5 seasons, i.e., 300-odd rainy days, are available for analysis. For shorter periods and for smaller categories of days, use of asymptotic formulas underestimates the variance and leads to spuriously significant results. This means that in effect the asymptotic distribution is useful only for overall evaluation of an entire experiment and cannot be used for detailed analyses within categories or shorter periods.

Our findings are based entirely on data of the Israeli experiment. One may ask how far the conclusions can be extended to other rainfall stimulation experiments. Clearly, one cannot infer from Israeli rainfall data to data in areas with very different rainfall regimes, but some inferences are possible to slightly different experimental designs. Though our evaluation of the Israeli experiment used only rainy days, we have made some additional checks using the data of all days of the rainy seasons. There were altogether 946 days, rainy and dry together. For these days we checked the applicability of the asymptotic normal distribution of R, and did the same for sets of 5, 10 and 20 successive days. The latter might correspond more closely to the analyses of some experiments which used units considerably longer than 24 hours.

The results of these additional Monte Carlo sampling trials are presented in Table 3. Tests of goodness of fit do not indicate deviation from the asymptotic distribution either for single days or for sets of days. Sample estimates of the variance of R do not differ much from asymptotic $Var_0(R)$ though the latter seems slightly too low for single days and possibly a little too large for sets of days. Our conclusion that the asymptotic distribution of R may be used safely when data of about 5-6 seasons are available evidently holds not only for single rainy days but also for all single days and for sets of days as well.

Table 3. Sampled Distributions of R, Goodness of Fit and Other Characteristics (All Single Days and Other Periods)

	Number of Periods	Number of Permutations	Chi-square for goodness of fit 19 d.f.†	Prop. (R > 1)	Mean	Variance	Asymptotic variance	α_3	α_4
1961-6 (Days)	946	200	26.40	.495	1.0065	.0048	.0051	.006	3.21
1961-6 (5-days)	189	200	24.00	.535	1.0054	.0050	.0044	.009	2.53
1961-6 (10-days)	94	200	25.20	.505	1.0056	.0049	.0047	.019	3.31
1961-6 (20-days)	47	200	12.40	.560	1.0085	.0051	.0047	.005	3.19

†See footnote to Table 2.

It is interesting to note that the variance of R based on rainy days is slightly larger than that based on all days. Since the value of R is practically the same in both cases, it appears that exclusion of "dry" days may slightly reduce the sensitivity of the R randomization test. Contrary results were found in some earlier calculations for the WMW test which appeared to gain in sensitivity when restricted to rainy days.

ACKNOWLEDGMENT:

The authors are happy to acknowledge Professor Lincoln E. Moses's original suggestion to study the distribution of the average ratio by both asymptotic and Monte Carlo methods and his very helpful comments.

REFERENCES

[1] Adderley, E. E., "Non-parametric methods of analysis applied to large-scale cloud seeding experiments," Journal of Meteorology 18: 692-694 (1961).

[2] Adderley, E. E. and Twomey, S., "An experiment on artificial stimulation of rainfall in the Snowy Mountains of Australia," Tellus 10: 275-280 (1958).

[3] Gabriel, K. R., "L'expérience de pluie provoquée en Israel. Quelques résultats partiels," Journal de Récherches Atmosphériques 2: 1-6 (1965).

[4] Gabriel, K. R., "The Israeli artificial rainfall stimulation experiment. Statistical evaluation for the period 1961-1965," Fifth Berkeley Symposium on Mathematical Statistics and Probability. Weather Modification Section, 1966.

[5] Gabriel, K. R., "Recent results of the Israeli artificial rainfall stimulation experiment," Journal of Applied Meteorology, 6, 437-438 (1967).

[6] Loève, Michel, Probability Theory. (3rd ed.) Princeton: Van Nostrand, 1963.

[7] Neyman, J. and Scott, E. L., "Asymptotically optimal tests of composite hypotheses for randomized experiments with non-controlled predictor variables," Journal of the American Statistical Association 60: 699-721 (1965).

[8] Neyman, J. and Scott, E. L., "Some outstanding problems relating to rain modification," Fifth Berkeley Symposium on Mathematical Statistics and Probability. Weather Modification Section, 1966.

[9] Owen, Donald B., Handbook of Statistical Tables. Reading, Mass.: Addison-Wesley, 1962.

[10] Siliceo, E. Perez, Ahumada, A. A. and Mosino, P. A., "Twelve years of cloud seeding in the Necaxa watershed, Mexico," Journal of Applied Meteorology 2: 311-323 (1963).

APPENDIX

THE ASYMPTOTIC DISTRIBUTION OF R

The statistic R is a function of weighted sums of independent Bernoulli variables $\theta_1, \theta_2, \ldots, \theta_N$, each being 1 or 0 with probability $\frac{1}{2}$. Using the notation of (5), and defining

$$\lambda_i = x_i/X \quad \text{and} \quad \mu_i = y_i/Y \ , \tag{8}$$

these sums may be written

$$S = \sum_{i=1}^{N} \lambda_i \theta_i \quad \text{and} \quad T = \sum_{i=1}^{N} \mu_i \theta_i \tag{9}$$

and the average ratio (3) becomes

$$R = [S/(1-S) \cdot (1-T)/T]^{\frac{1}{2}}. \tag{10}$$

One readily obtains

$$\begin{aligned} \mathcal{E}(S) &= \Sigma_i \lambda_i/2 = \tfrac{1}{2} \\ \mathcal{E}(T) &= \Sigma_i \mu_i/2 = \tfrac{1}{2} \\ \mathrm{Var}(S) &= \Sigma_i \lambda_i^2/4 \\ \mathrm{Var}(T) &= \Sigma_i \mu_i^2/4 \\ \mathrm{Cov}(S,T) &= \Sigma_i \lambda_i \mu_i/4. \end{aligned} \tag{11}$$

The difference $S-T = \Sigma_i(\lambda_i - \mu_i)\theta_i$ is a weighted sum of N independent Bernoulli variables. Its expectation is zero since $\Sigma_i(\lambda_i - \mu_i) = 1-1=0$ and its variance is $\Sigma_i(\lambda_i - \mu_i)^2/4$. To obtain its asymptotic distribution, define, for any N,

$$Z_{Ni} = \frac{(\lambda_i - \mu_i)(\theta_i - \frac{1}{2})}{\sqrt{\Sigma_i(\lambda_i - \mu_i)^2/4}} \qquad i=1,2,\ldots,N, \tag{12}$$

so that $\mathcal{E}Z_{Ni} = 0 \quad i=1,2,\ldots,N$ and $\Sigma_i \mathrm{Var}(Z_{Ni}) = 1$ for all N. Hence Loève's Normal Convergence Criterion ([6], p. 295) applies and

$$\Sigma_i Z_{Ni} = \frac{2(S-T)}{\sqrt{\Sigma_i(\lambda_i-\mu_i)^2}} \tag{13}$$

is asymptotically N(0,1) provided that for every $\epsilon > 0$

$$\Sigma_i \int_{|u|\geq\epsilon} u^2 dF_{Z_{Ni}}(u) \to 0 \text{ as } N\to\infty . \tag{14}$$

Condition (14) holds if, for every $\epsilon > 0$, there exists an N_o large enough so that for all $N > N_o$

$$\frac{|\lambda_i-\mu_i|}{\sqrt{\Sigma_i(\lambda_i-\mu_i)^2}} < \epsilon \qquad \text{for all } i=1,2,\ldots,N$$

(for in that case $P(|Z_{Ni}| \geq \epsilon) = 0 \quad i=1,2,\ldots,N$). Hence a sufficient condition for asymptotic normality in (13) is

$$\lim_{N\to\infty} \max_{1\leq i\leq N} \frac{|\lambda_i-\mu_i|}{\sqrt{\Sigma_i(\lambda_i-\mu_i)^2}} = 0 \quad . \tag{15}$$

Next, to obtain the asymptotic distribution of the average ratio R, one may expand it as a Taylor series about $(S,T) = (\frac{1}{2},\frac{1}{2})$. Noting that

$$\frac{\partial R}{\partial S} = \frac{R}{2S(1-S)} \quad \text{and} \quad \frac{\partial R}{\partial T} = -\frac{R}{2T(1-T)} ,$$

one obtains, at point $(\frac{1}{2},\frac{1}{2})$,

$$R=1 \quad \partial R/\partial S = 2 \quad \text{and} \quad \partial R/\partial T = -2.$$

The Taylor series therefore becomes

$$\begin{aligned} R &= 1 + 2(S-\tfrac{1}{2}) - 2(T-\tfrac{1}{2}) + W \\ &= 1 + 2(S-T) + W, \end{aligned} \tag{16}$$

where W is a sum of terms with two or more $(S-\frac{1}{2})$, $(T-\frac{1}{2})$ factors each.

By virtue of (11) $(S-\frac{1}{2})^2 = 0_p(\Sigma_i\lambda_i^2)$ and $(T-\frac{1}{2})^2 = 0_p(\Sigma_i\mu_i^2)$ so that $W = 0_p(\Sigma_i\mu_i^2 + \Sigma_i\lambda_i^2)$. Dividing both sides of (16) by $\sqrt{\Sigma_i(\lambda_i-\mu_i)^2}$ and rearranging, one obtains

$$\frac{R-1}{\sqrt{\Sigma_i(\lambda_i-\mu_i)^2}} = \frac{2(S-T)}{\sqrt{\Sigma_i(\lambda_i-\mu_i)^2}} + 0_p\left(\frac{\Sigma_i\lambda_i^2 + \Sigma_i\mu_i^2}{\sqrt{\Sigma_i(\lambda_i-\mu_i)^2}}\right) . \tag{17}$$

Provided, then, that

$$\lim_{N\to\infty} \frac{\Sigma_i\lambda_i^2 + \Sigma_i\mu_i^2}{\sqrt{\Sigma_i(\lambda_i-\mu_i)^2}} = 0 \tag{18}$$

it follows ([6], p. 174, item 16) that $(R-1)/\sqrt{\Sigma_i(\lambda_i-\mu_i)^2}$ and $2(S-T)/\sqrt{\Sigma_i(\lambda_i-\mu_i)^2}$ have asymptotically the same distribution, that is, if (15) also holds, a unit normal distribution. In particular, this establishes the asymptotic expectation $\mathcal{E}_0 R=1$ and variance in (4), above.

If, in (16), terms involving the second derivatives

$$\frac{\partial^2 R}{\partial S^2} = \frac{(4S-1)R}{4S^2(1-S)^2}, \quad \frac{\partial^2 R}{\partial S \partial T} = -\frac{R}{4ST(1-S)(1-T)}, \quad \frac{\partial^2 R}{\partial T^2} = \frac{(3-4T)R}{4T^2(1-T)^2}$$

were separated from W, one would obtain

$$R = 1+2(S-\tfrac{1}{2})-2(T-\tfrac{1}{2})+4(S-\tfrac{1}{2})^2/2-2\times 4(S-\tfrac{1}{2})(T-\tfrac{1}{2})/2+4(T-\tfrac{1}{2})^2/2 + W'$$
$$= 1+2(S-T)+2(S-T)^2+W' . \qquad (16a)$$

After some reduction, this will be seen to yield the approximations to the expectation and variance given in (7), above.

Finally, this development suggests that the statistic $1+2(S-T)$ might itself be used as a good practical approximation to R. Under the null hypothesis its expectation is one and its variance $\Sigma_i(\lambda_i-\mu_i)^2$ for any sample size. Condition (15) suffices for its asymptotic normality. It is also not difficult to argue the intuitive appeal of this statistic since S is the proportion of North area rainfall which falls on North seeded days, and T is the proportion of Centre area rainfall falling on the same days. Without seeding effects, the expected values of S and T are both $\frac{1}{2}$, and $\mathcal{E}\{1+2(S-T)\} = 1$, whereas with a proportionate increase of $(1+r)$ due to seeding, S would tend to be close to $\frac{1+r}{2+r}$ and T to $\frac{1}{2+r}$ so that $1+2(S-T)$ would be close to $1 + \frac{2r}{2+r} = 1+r - \frac{r^2}{2+r}$, i.e., quite near $1+r$ for small r. Table 1 shows that the observed values of this statistic indeed underestimate the average ratio R. The underestimate is very slight unless r is at least 0.25.