DP-1127

AEC RESEARCH AND DEVELOPMENT REPORT

ASP - ANALYSIS OF SYNTHETICS PROGRAM FOR QUALITY CONTROL DATA

L. V. BAILEY
L. M. ARNETT

DUPONT
REG. U. S. PAT. OFF.

Savannah River Laboratory

Aiken, South Carolina

LEGAL NOTICE

Printed in the United States of America
Available from
Clearinghouse for Federal Scientific and Technical Information
National Bureau of Standards, U. S. Department of Commerce
Springfield, Virginia 22151
Price: Printed Copy $3.00; Microfiche $0.65

663564

DP-1127

Mathematics and Computers
(TID-4500)

ASP - Analysis of Synthetics Program for Quality Control Data

by

L. V. Bailey
L. M. Arnett

Approved by

J. E. Suich, Research Manager
Applied Mathematics Division

January 1968

E. I. DU PONT DE NEMOURS & COMPANY
SAVANNAH RIVER LABORATORY
AIKEN, S. C. 29801

CONTRACT AT(07-2)-1 WITH THE
UNITED STATES ATOMIC ENERGY COMMISSION

ABSTRACT

The computer program, ASP, which calculates bias, precision, and other statistics of analytical methods, was written in FORTRAN IV for use on the IBM System/360-65. The Savannah River Plant laboratories use ASP monthly and quarterly to evaluate and to report the bias and precision of analyses important to process control and accountability.

CONTENTS

ASP - Analysis of Synthetics Program for Quality Control Data

INTRODUCTION

Detailed knowledge of the reproducibility and accuracy of chemical and isotopic analyses is essential for process control and materials accountability. The analytical quality control program at the Savannah River Plant was designed to provide continual control over both the accuracy and the precision of analytical values reported by the various laboratories. Computer processing of the data greatly facilitates prompt and accurate statistical analysis and reports.

SUMMARY

The computer program, ASP, is used to calculate and to prepare reports of statistics covering quality control data consisting of analytical results of synthetic samples with values known to supervision, but not to the analyst. Calculated statistics include (1) the average known value, (2) the average determined value, (3) the average bias, (4) the average range in replicate determinations, (5) the precision, (6) the control limits (95% confidence interval), (7) the normality of data, (8) the statistical significance of the calculated bias, and (9) the statistical significance of changes in bias and precision from previous periods. Summary reports for specified combinations of analyses are presented, as well as detailed sorted listings for each analysis.

The program produces updated identification and statistics on cards for input to subsequent problems. The program processes current data in the following manner. Each determined value is the mean of duplicate analyses and thus each value has an associated range. All of the statistics are calculated first for the set of determined values that includes only the initially determined values, i.e., all repeated values are omitted. The maximum number of values that can be deleted from the statistical analyses is calculated as a function of the number of determinations. The individual determined values are compared with the value corresponding to a 5% confidence level in the t-distribution, and the value with the largest deviation is deleted if it exceeds the value in the t-distribution. If the analysis were repeated and if both the value and range of the repeated determination are less than the value in the t-distribution, the repeated value is substituted for the deleted value. All of the statistics are

then recalculated. This routine is repeated until the maximum number of values are deleted or until all values are within the 95% confidence limits. The retained values are then ordered by the magnitude of the bias, and the conformity to normality is determined by the chi-square test. Precision and control limits are calculated for 95% confidence limits in the t-distribution. The difference between the calculated bias and zero is tested for significance. Previous statistics are compared with current calculated statistics. The F-table (5% level) is used to determine significance of changes in variance, and the t-distribution (5% level) is used to determine significance of bias and significance of changes in bias.

DISCUSSION

DESCRIPTION OF THE QUALITY CONTROL PROGRAM USED BY THE SAVANNAH RIVER PLANT LABORATORIES

Synthetic samples, similar to production samples, are analyzed as part of the quality control program. Standard values are established for these synthetic samples, either by preparing them from weighed amounts of material or by analyzing selected production samples under carefully controlled conditions. These synthetic samples are submitted along with routine production samples to personnel analyzing production samples. The supervisor records the result of the analysis and plots the value on a control chart. If the value exceeds control limits, the sample is again submitted for analysis. The repeated value is recorded and plotted; if it also exceeds control limits, the supervisor takes action intended to correct the condition responsible for out-of-limit results. All types of analyses at the several concentrations required for process control or accountability are included in this program. Bias corrections, submitted each month to the Production Department, are used for materials accountability. Computer processing greatly facilitates the preparation of prompt reports that condense the large volume of data and meet the requirement for accurate statistical analysis of each unique type of analysis. The computer program, ASP, provides accurate mathematical analysis and prompt summary and detailed reports for management and laboratory personnel.

EQUATIONS USED IN COMPUTER PROGRAM

Standard formulas are used to calculate arithmetic means and unbiased estimates of variances and standard deviations.

Given a set of data,

$$x_i; \quad i = 1, 2, 3, \ldots N$$

the arithmetic mean is

$$\bar{X} = \frac{\sum_{1}^{N} x_i}{N}$$

the estimate of variance is

$$S^2 = \frac{\sum_{1}^{N} (x_i - \bar{X})^2}{N-1}$$

and the estimate of standard deviation is $\sqrt{S^2}$.

The maximum number of deletions is $D = .05N + \sqrt{(.05)(.95)N}$, truncated to an integer. This formula for D is the expected number of erroneous analyses plus one standard deviation given that the population contains 5% erroneous analyses. This formula is chosen arbitrarily and has the merit of easy application. Given that the sample contains 5% erroneous analyses, the probability that there are D erroneous analyses in a sample of size N is always equal to or greater than 0.16. This lower limit on the probability is approached as N approaches infinity. In the region of usual application, the probabilities are higher. For example, in the region D = 1 for $21 > N > 7$, the probability that the sample contains at least one erroneous analysis starts at 0.34 and rises to 0.64. At the point D = 2 for N = 21, the probability that the sample contains at least two erroneous analyses is 0.28.

The mean and estimated standard deviation are calculated for the initial set of determinations. The value, x_i, deviating most from the arithmetic mean is compared with the t-distribution.

If

$$t_i = \frac{x_i - \bar{X}}{S} > t_{.05}$$

for the corresponding degrees of freedom, and the number of values previously deleted is less than the maximum permissible number of deletions, then x_i is deleted. If the value were repeated and the t-value of the repeated value is less than $t_{.05}$, then the repeated value is included and the mean and standard deviation are recalculated. This routine is continued until the maximum number of deletions is reached or until all values have $t_i \leqslant t_{.05}$.

The average bias is the average determined value minus the average known value.

$$\overline{BIAS} = \overline{RV} - \overline{SV}$$

$$\% \text{ Bias} = \frac{(100)(\overline{BIAS})}{\overline{RV}}$$

The precision (95% confidence interval) equals $(t_{.05})(S)$, where $t_{.05}$ is the value from the t-distribution for P = .05, and m (degrees of freedom) = N-1. If m > 30, then the precision is 2.042 S.

The control limits are $\overline{X}$ ± precision.

The significance of the $\overline{BIAS}$ is determined by the comparison of the observed

$$t(\overline{BIAS}) = \frac{(\overline{BIAS})\sqrt{N}}{S}$$

with $t_{.05}$. If $t(\overline{BIAS}) < t_{.05}$ for m degrees of freedom, then $\overline{BIAS}$ is not significant; otherwise, the $\overline{BIAS}$ is significant.

The significance of a change in the $\overline{BIAS}$ is calculated in a similar manner from the following formula:

$$t(\overline{BIAS}_1 - \overline{BIAS}_2) = \frac{\overline{BIAS}_1 - \overline{BIAS}_2}{\sqrt{\left(\frac{m_1 S_1^2 + m_2 S_2^2}{m_1 + m_2}\right)\left(\frac{1}{N_1} + \frac{1}{N_2}\right)}}$$

where the subscripts designate the two biases being compared. The degrees of freedom are

$$m_1 = N_1 - 1$$

and

$$m_2 = N_2 - 1$$

The difference is considered significant if for m degrees of freedom $(m_1 + m_2)$,

$$t(\overline{BIAS}_1 - \overline{BIAS}_2) > t_{.05}$$

Significant changes in the precision of the analyses are detected by comparing the ratios of variances to values in the F-distribution.

By definition

$$F = \frac{S_1^2}{S_2^2}$$

where $S_1^2 > S_2^2$.

If the observed ratio of variances, F, exceeds F for 5% probability corresponding to m_1 and m_2 degrees of freedom, the two variances are considered to be significantly different and thus the precision has changed significantly.

Each set of N retained determined values for which a mean, $\overline{X}$, and an estimated standard deviation, S, are calculated is tested for normality. The chi-square test is applied in the following manner. By definition

$$\chi^2 = \sum_{i=1}^{K} \left[\frac{(OF_i - EF_i)^2}{EF_i} \right]$$

where

OF_i = observed number of values in the i^{th} group

and

EF_i = expected number of values in the i^{th} group

The observed values retained in the statistical analysis are ordered from the smallest x_1 to the largest x_N and divided into K groups equal to N/S truncated to the nearest integer. However, K is never allowed to exceed 30 even though N exceeds 150. This set of observed values is compared with a normal distribution that is arbitrarily augmented in the tails to partially compensate for the values that may have been deleted from the original set of values. The total number of values in the augmented distribution is

$$N_T = N + \frac{N}{\sqrt{2\pi}\,\sigma} \left[\int_{-\infty}^{-U_1\sigma+\overline{X}} \exp\left(-\frac{u^2}{2}\right) du + \int_{U_N\sigma+\overline{X}}^{\infty} \exp\left(-\frac{u^2}{2}\right) du \right]$$

where

$$U_i = \frac{x_i - \overline{X}}{\sigma}$$

The boundaries of the ranges that include expected numbers of values equal to N_T/K are obtained by integrating the normal distribution with mean, $\overline{X}$, and standard deviation, σ. The last group is combined with the previous group so as to avoid having any group with less than five expected values. The chi-square is then calculated and compared with tabulated values. If the chi-square exceeds the value corresponding to a probability of 0.05, the hypothesis of normality is rejected. The degrees of freedom are equal to the number of groups minus two.

DESCRIPTION OF OUTPUT AND INPUT FOR COMPUTER PROGRAM

OUTPUT

Printed Reports (See Appendix B for examples of reports)

Each analysis and specified groups of analyses are identified alphabetically. For ease in interpreting results, statistics are reported to six significant digits or less with three decimal places multiplied by the power of 10 exponent requested for the analysis. The following reports are issued:

- A statistical summary of specified combinations of analyses including alphabetic identification of each analysis, average standard value, average bias (arithmetic amount), current and previous bias as percent of the average reported value, current and previous precision (95% confidence interval) as percent of the average reported value, statistical significance of bias, and statistical significance of changes in bias and precision.
- Control limits for bias and for range in replicate determinations for the specified combinations of analyses.
- For each analysis the calculated statistic, a listing of data in bias sequence with indication as to whether the data were included or deleted from calculations, and results of the chi-square analysis.

Punched Cards

Updated alphabetic identification cards and calculated statistics are punched for use as input data for future analyses. After the initial run of the program, control alphabetic and comparison cards are punched manually only if new analyses are added.

INPUT

Input consists of the following types of cards:

- Alphabetic identification of type of analysis and sample identification
- Previous quarterly statistics
- Previous monthly statistics
- Alphabetic identification of area or report groupings
- Current analytical results submitted for statistical analysis

Alphabetic Identification of Analysis

There should be a card for each unique type of analysis, since the alphabetic identification is the only printed identification for the analysis. These cards are updated and punched by the program. After initial preparation, it is unnecessary to keypunch cards as input for the next processing unless changes are desired or unless new analyses are added.

Previous Quarterly Statistics

These cards are punched by the program when results from more than 31 days are analyzed, and are retained as input for the next run of the program.

Previous Monthly Statistics

These cards are punched by the program when monthly statistics are calculated, and are used as input for the next month for comparison.

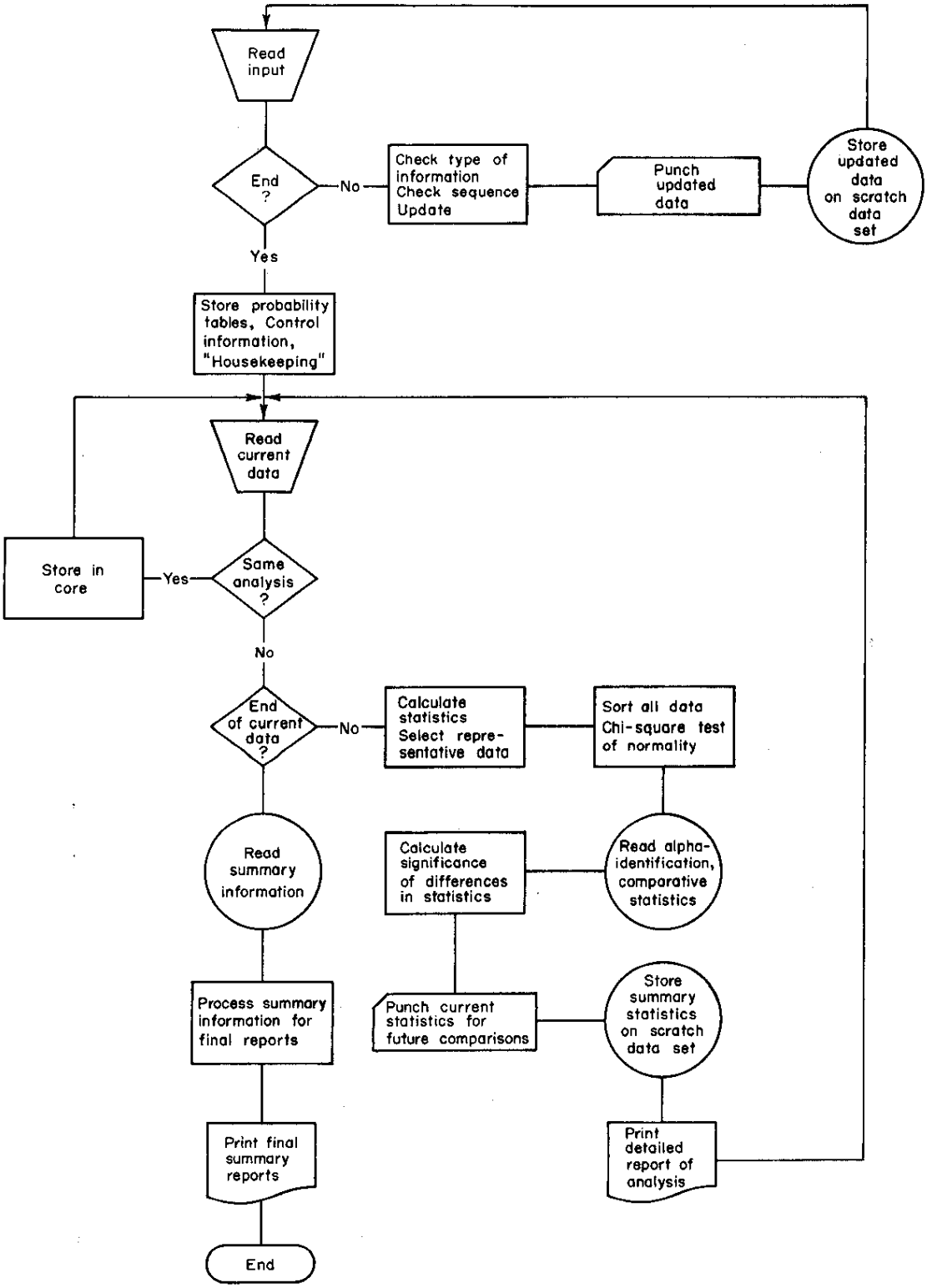
Report or Area Alphabetic Identification

There are exactly nine of these cards. The alphabetic identification shown on the card will be printed for each report grouping. These cards are punched by the program.

Current Data Cards

Cards containing the current data are punched for each run of the program.

APPENDIX A - Flow Diagram of ASP



APPENDIX B - Output Examples

QUALITY CONTROL DATA - ANALYSES OF SYNTHETIC SAMPLES
TECHNICAL SERVICES SECTION - LABORATORIES DIVISION

ANALYSES FROM 6-29-66 THROUGH 9-29-66

SPECIAL PRODUCTS, PU-238

ANALYSIS AND METHOD	SAMPLE IDENTIF.	UNIT OF MEASURE	PERCENT CORR. AT SOURCE	AVERAGE STANDARD VALUE DIGITS	EXP	AVERAGE BIAS ARITHMETIC AMOUNT DIGITS	EXP	BIAS PERCENT OF R.V. PREV.	BIAS PERCENT OF R.V. CURR.	PRECISION (95 PERCENT CI) PREV.	PRECISION (95 PERCENT CI) CURR.	BIAS SIGNF. DIFF. FROM 0	SIGNF DIFF. IN BIAS	SIGNF DIFF. IN PRECSN
PLUTONIUM, TTA EXT.	10.1	D/M/ML		4.620	5	0.011	5	6.8	0.2	16.0	16.6	NO	YES	NO
PLUTONIUM, TTA EXT.	9.3	D/M/ML		1.740	6	-0.075	6	-3.4	-4.5	18.2	15.4	YES	NO	NO
PLUTONIUM, GROSS ALPHA	5.3-2	D/M/ML		1.680	9	-0.031	9	0.6	-1.9	6.4	4.9	YES	YES	BETTER
SPECIFIC GRAVITY, F.D.	5.3-2	SP. GR.		12.832	-1	-0.027	-1	-1.0	-0.9	2.1	3.1	YES	NO	WORSE
NEPTUNIUM, TIOA-TTA EXT.	10.1	D/M/ML	5.0	5.740	3	-0.185	3	-5.6	-3.3	12.4	18.3	NO	NO	WORSE
NEPTUNIUM, TIOA-TTA EXT.	5.3-4	D/M/ML		14.940	3	-0.568	3	-3.6	-4.0	4.3	8.3	YES	NO	WORSE
NEPTUNIUM, TIOA-TTA EXT.	5.1-8	D/M/ML		2.390	6	-0.051	6	-2.1	-2.2	5.6	5.2	YES	NO	NO
NEPTUNIUM, TIOA-TTA EXT.	5.3-2	D/M/ML	5.0	2.990	6	0.019	6	2.9	0.6	4.8	6.8	NO	YES	WORSE
ACID, PHENOL	5.3-2	NORMALITY		4.640		0.050		2.5	1.1	6.3	5.1	YES	NO	NO

QUALITY CONTROL DATA - ANALYSES OF SYNTHETIC SAMPLES
TECHNICAL SERVICES SECTION - LABORATORIES DIVISION

ANALYSES FROM 6-30-66 THROUGH 9-28-66

PUREX SYNTHETICS, PU-239

ANALYSIS AND METHOD	SAMPLE IDENTIF.	UNIT OF MEASURE	PERCENT CORR. AT SOURCE	CONTROL LIMITS FOR REPORTED VALUE ADD TO STANDARD VALUE AVERAGE DIGITS	EXP	MINIMUM DIGITS	EXP	MAXIMUM DIGITS	EXP	CONTROL LIMITS FOR RANGE AVERAGE DIGITS	EXP	MAXIMUM DIGITS	EXP
PLUTONIUM, GROSS ALPHA	8.1	D/M/ML	2.0	0.028	7	-0.264	7	0.320	7	0.093	7	0.238	7
PLUTONIUM, GROSS ALPHA	9.8	D/M/ML	0.0	0.029	8	-0.028	8	0.085	8	0.033	8	0.095	8
PLUTONIUM, GROSS ALPHA	C-5	D/M/ML	0.0	0.161	9	0.021	9	0.300	9	0.062	9	0.149	9
RATIO OF PU/URANIUM	8.1	D/M/ML / G/L	2.0	0.084	4	-0.627	4	0.795	4	0.477	4	1.215	4
SPECIFIC GRAVITY, F.D.	8.1	SP. GR.	0.0	-4.450	-6	-0.077	-1	0.077	-1	0.0	-1	0.0	-1
URANIUM, CALCULATED	8.1	G/L	0.0	-0.342	0	-6.266	0	5.582	0	0.0	0	0.0	0
URANIUM, FLUOROPHOTOM.	15.7	G/L	0.0	-0.121	-3	-0.950	-3	0.708	-3	0.101	-3	0.347	-3
ACID, CONDUCTOMETRI	8.1	NORMALITY	0.0	0.005	0	-0.042	0	0.051	0	0.022	0	0.057	0
ACID, OXALATE	C-5	NORMALITY	0.0	0.037	0	-0.126	0	0.201	0	0.050	0	0.142	0

QUALITY CONTROL DATA - ANALYSES OF SYNTHETIC SAMPLES
TECHNICAL SERVICES SECTION - LABORATORIES DIVISION

ANALYSES FROM 7- 5-66 THROUGH 9-19-66

PUREX SYNTHETICS, PU-239

PLUTONIUM, GROSS ALPHA C-5 D/M/ML

	DIGITS	EXP		
AVERAGE STANDARD VALUE =	5.780	9		
AVERAGE REPORTED VALUE =	5.941	9		
AVERAGE BIAS =	0.161	9	PERCENT OF AV. R.V. =	2.702
PRECISION (95 P.C. C.I.)=	0.139	9	PERCENT OF AV. R.V. =	2.341

CONTROL LIMITS FOR BIAS

MINIMUM =	0.021	9
MAXIMUM =	0.300	9
AV. RANGE IN DUPLICATES =	0.062	9
MAX. CONTR. LIMIT,RANGE =	0.149	9

DATE MO DA YR	SHIFT	SAMPLE LOG NUMBER	REPEAT CODE	STANDARD VALUE DIGITS	EXP	REPORTED VALUE DIGITS	EXP	RANGE IN DUPLICATES DIGITS	EXP	BIAS (R.V.-STD.VAL.) ARITH. AMOUNT DIGITS	EXP	PERCENT OF R.V.	T-VALUE BIAS	T-VALUE RANGE	INCLUDED IN CALCULATIONS
7- 7-66	G	031412	2	5.780	9	5.600	9	0.270	9	-0.180	9	-3.2	-5.1	5.0	NO-REPEAT
8-18-66	G	031496	0	5.780	9	5.770	9	0.100	9	-0.010	9	-0.2	-2.6	0.9	NO-R.V.
7-25-66	E	031443	2	5.780	9	5.780	9	0.070	9	0.0	9	0.0	-2.4	0.2	NO-REPEAT
7-28-66	G	031454	0	5.780	9	5.830	9	0.110	9	0.050	9	0.9	-1.7	1.2	YES
9-12-66	F	034945	0	5.780	9	5.840	9	0.040	9	0.060	9	1.0	-1.5	-0.5	YES
7- 7-66	G	031412	1	5.780	9	5.850	9	0.300	9	0.070	9	1.2	-1.4	5.8	NO-RANGE
8-25-66	E	034911	0	5.780	9	5.850	9	0.090	9	0.070	9	1.2	-1.4	0.7	YES
8- 4-66	E	031467	2	5.780	9	5.850	9	0.030	9	0.070	9	1.2	-1.4	-0.8	NO-REPEAT
9- 1-66	F	034926	0	5.780	9	5.880	9	0.010	9	0.100	9	1.7	-0.9	-1.3	YES
7-18-66	G	031432	0	5.780	9	5.880	9	0.080	9	0.100	9	1.7	-0.9	0.4	YES
8-29-66	G	034918	0	5.780	9	5.900	9	0.050	9	0.120	9	2.0	-0.6	-0.3	YES
7-21-66	F	031440	2	5.780	9	5.900	9	0.060	9	0.120	9	2.0	-0.6	-0.1	NO-REPEAT
8- 8-66	G	031474	0	5.780	9	5.900	9	0.110	9	0.120	9	2.0	-0.6	1.2	YES
9- 8-66	G	034940	0	5.780	9	5.910	9	0.070	9	0.130	9	2.2	-0.5	0.2	YES
7-25-66	E	031443	1	5.780	9	5.940	9	0.060	9	0.160	9	2.7	-0.0	-0.1	YES
9-19-66	G	034960	2	5.780	9	5.940	9	0.120	9	0.160	9	2.7	-0.0	1.4	NO-REPEAT
8-11-66	F	031482	2	5.780	9	5.950	9	0.060	9	0.170	9	2.9	0.1	-0.1	NO-REPEAT
7-21-66	F	031440	1	5.780	9	5.960	9	0.080	9	0.180	9	3.0	0.3	0.4	YES
8- 1-66	F	031459	1	5.780	9	5.960	9	0.010	9	0.180	9	3.0	0.3	-1.3	YES
7-11-66	F	031417	2	5.780	9	5.960	9	0.120	9	0.180	9	3.0	0.3	1.4	NO-REPEAT
7-14-66	E	031425	1	5.780	9	5.980	9	0.010	9	0.200	9	3.3	0.6	-1.3	YES
7-11-66	F	031417	1	5.780	9	5.980	9	0.030	9	0.200	9	3.3	0.6	-0.8	YES
8-22-66	F	034903	1	5.780	9	5.980	9	0.070	9	0.200	9	3.3	0.6	0.2	YES
8- 4-66	E	031467	1	5.780	9	5.990	9	0.070	9	0.210	9	3.5	0.7	0.2	YES
8- 1-66	F	031459	2	5.780	9	5.990	9	0.110	9	0.210	9	3.5	0.7	1.2	NO-REPEAT
9- 5-66	E	034929	1	5.780	9	6.000	9	0.070	9	0.220	9	3.7	0.9	0.2	YES
8-11-66	F	031482	1	5.780	9	6.010	9	0.020	9	0.230	9	3.8	1.0	-1.0	YES
8-22-66	F	034903	2	5.780	9	6.020	9	0.010	9	0.240	9	4.0	1.2	-1.3	NO-REPEAT
7- 5-66	E	031401	1	5.780	9	6.020	9	0.170	9	0.240	9	4.0	1.2	2.6	YES
9- 5-66	E	034929	2	5.780	9	6.040	9	0.170	9	0.260	9	4.3	1.5	2.6	NO-REPEAT
7- 5-66	E	031401	2	5.780	9	6.040	9	0.050	9	0.260	9	4.3	1.5	-0.3	NO-REPEAT
9-19-66	G	034960	1	5.780	9	6.060	9	0.030	9	0.280	9	4.6	1.8	-0.8	YES
7-14-66	E	031425	2	5.780	9	6.200	9	0.110	9	0.420	9	6.8	3.9	1.2	NO-REPEAT

COMPARISON OF OBSERVED AND THEORETICAL FREQUENCIES

UPPER RANGE DIGITS	EXP	OBSERVED FREQUENCY	THEORETICAL FREQUENCY	PERCENT DIFFERENCE
0.0100	9	10.0	6.5	53.5
0.0130	9	4.0	5.5	-26.8
0.0160	9	2.0	5.2	-61.4
0.0200	9	7.0	7.3	-4.6
0.0420	9	10.0	8.5	17.6

NO SIGNIF. DIFF. FROM NORMAL DISTRIBUTION