Columbia University
in the City of New York

COLUMBIA UNIVERSITY STATISTICAL BUREAU
DOCUMENT No. 1

THE MENDENHALL-WARREN-HOLLERITH CORRELATION METHOD

By
Richard Warren
and
Robert M. Mendenhall

1929

THE MENDENHALL-WARREN-HOLLERITH CORRELATION METHOD

9

CONTENTS

1 9 2 9

Introduction.

This monograph has been prepared by Messrs. Warren and Mendenhall and is issued by the Columbia University Statistical Bureau, in answer to many requests for a detailed description of the Mendenhall-Warren-Hollerith correlation method, which was briefly described in a paper read before Section Q of the American Association for the Advancement of Science in December 1928.

The discovery of this economical method of calculating correlations was made by Messrs. Warren and Mendenhall while doing graduate work in Columbia University, and particularly in connection with their work as statistical consultants for the Carnegie Foundation for the Advancement of Teaching.

The M-W-H correlation method, using the standard Hollerith tabulating machine, produces Pearson product-moment correlation coefficients and related statistical constants at a much lower cost, and much more speedily, than any other methods or machines which have been described in statistical literature. The method is particularly economical in calculating inter-correlations based on large populations. Under certain conditions the machines will produce data for as many as twelve correlations at a single run of the cards, and normally for as many as five or six correlations.

It was partly in the interest of placing this notable discovery at the service of the research workers of the country that the Columbia University Statistical Bureau was established. The purposes of the Bureau are as follows:

1. To study and extend the adaptabilities of statistical machines to the special problems of educational and social science research.

2. To assist in training research directors and students of all branches of statistics in the use of such adaptabilities, and specifically to acquaint them immediately (by publications, demonstrations, and service) with the Mendenhall-Warren-Hollerith method of securing distributions and data in convenient form for calculating means, sigmas, Pearson coefficients of correlation, and also data facilitating the calculation of array-means, higher moments, and higher product-moments used in curve fitting.

3. To furnish consultation services to engineers and designers of special-purpose statistical machines, and of alterations of or attachments to existing machines to increase their flexibilith and adaptability to the more complex statistical procedures, including advanced business accounting procedures.

4. To furnish laboratory instruction on machine methods in statistical work to students registered in statistics courses in the University; and to provide demonstrations of the Mendenhall-Warren-Hollerith correlation method, and other recent adaptations of statistical machines, for visiting scientists, research directors, and machine operators.

The opportunity is taken here of making acknowledgment to the Carnegie Foundation for the Advancement of Teaching for its contribution to the development of the Mendenhall-Warren-Hollerith correlation method. The method was developed specifically as a means of bringing within feasible cost limits the statistical analysis of the 200,000 test returns and the enormous mass of other data collected by the Foundation in cooperation with the Joint Commission for the Study of the Relations of Secondary and Higher Education in Pennsylvania.

Ben D. Wood
Acting Director
COLUMBIA UNIVERSITY STATISTICAL BUREAU

THE COMPUTATION OF PRODUCT-MOMENT CORRELATIONS

ON THE

HOLLERITH TABULATING MACHINE

by

Richard Warren

When the heights of a group of people are measured we find that certain sizes tend to repeat themselves more frequently than others. If we write our measures, or scores, on index cards we can sort them to find how the size of the measure affects its frequency of occurrence. When we do this we find that very large measures and very small measures are comparatively infrequent and that the most frequent repetitions occur near the average of the group.

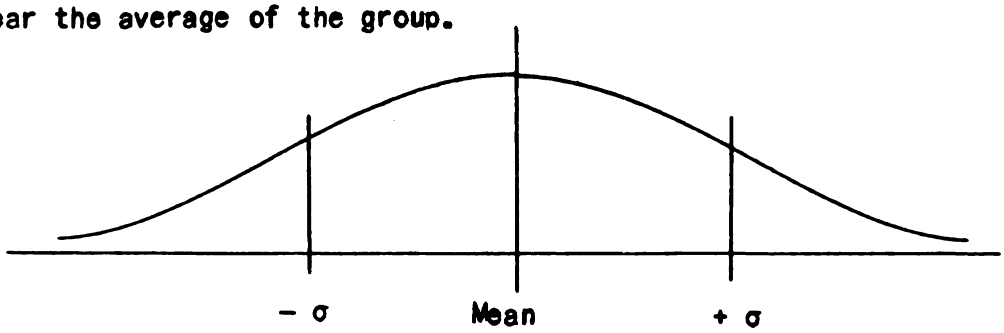


Figure 1.

The mathematics of frequency distributions has been studied by several investigators and it has been found that the formulas describing them are simplified if measures are expressed as deviations from the Average, or Arithmetic Mean of the group. Also it is found that one of the significant constants of the frequency curve is the distance from the Mean out to the point where the curve begins to flatten out, and that this distance is equal to the Root-Mean-Square of the deviations from the Mean, or the Square-root of the Average of the Squares of the Deviations. This number is called the Standard Deviation of the measures, and is an index of their variability. It is usually represented in formulas by the character σ, the small s of the Greek alphabet (Sigma). The Standard Deviation is also used to simplify formulas in the analysis of statistics since for many different traits, Height, Weight, Strength of Grip, Intelligence, etc., the different curves are all of the same size and shape when each person's measure is expressed as the number of Sigmas it is above or below the Mean. Measures expressed in these terms are called Standard Measures, or Standard Scores, and comparisons can be made between traits that were originally measured in different kinds of units, Heights compared with Weights, School Grades compared with Scores on an Intelligence Test, and so on.

If we measure two traits at the same time, Height and Weight for example, we find that certain *pairs* of Height-Weight measures repeat themselves more frequently than others, a high score in one occurring most frequently with a high score in the other. One number used to express the association between paired measures is the *Average Product* of the Standard Measures. It is called the Product-Moment Coefficient of Correlation, or Pearson r, referring to the English statistician who first used it extensively. It is written with subscripts indicating which pair it refers to, e.g. r_{12} is the correlation between traits 1 and 2, r_{35} the correlation between traits 3 and 5 etc., where traits 1 and 2 might be Height and Weight, and traits 3 and 5 Score on an Intelligence Test and Score on a Spelling Test, the r in each case being the average of the products of the Standard Scores on the two tests. r is always a fraction and it is negative when the two traits are so associated that a high score in one is usually found with a low score in the other.

Since we do not, in general, know the averages or the Standard Deviations of the measures we are studying until after we have gathered our material, we cannot express our measures, when originally made, as Standard Scores. The computation of r is made by averaging the products of the gross-scores in the ordinary units of measurement, pounds, points, or centimeters, and applying a correction formula to bring the result to the value it would have if we had used Standard Scores. The terms of the correction are the Means of the Scores and the Means of the Squares of the Scores.

The formula used is:

$$r_{12} = \frac{\frac{\Sigma X_1 X_2}{N} - \frac{\Sigma X_1}{N}\,\frac{\Sigma X_2}{N}}{\sqrt{\frac{\Sigma X_1^2}{N} - \left(\frac{\Sigma X_1}{N}\right)^2}\sqrt{\frac{\Sigma X_2^2}{N} - \left(\frac{\Sigma X_2}{N}\right)^2}}$$

in which the term $\frac{\Sigma X_1 X_2}{N}$ is the average product of the two scores, all the rest of the formula being the process of converting the result into terms of Standard Scores so that it may be compared with other measurements.

All of the terms in the formula are averages, either of scores or of score-products. The sums of the score-products used in computing the averages are called Product-Moments. The sum of the self-products of the scores of a variable is called the Second Moment of that variable. The labor of computing these sums on a Monroe or a Calculator is not prohibitive when the number of observations is less than a hundred, but when averages are to be computed for larger numbers of observations of several pairs of variables it is necessary to use greater economy of effort.

The index-card method is convenient for handling simultaneous measures of more than one trait. Cards are marked off in portions or *fields* similar to the columns of a book-keeper's ledger and the measures of the different traits are written in different fields of the card. Figure 2 shows one such card carrying the description of Person No. 125 in terms of his measurements in six different traits.

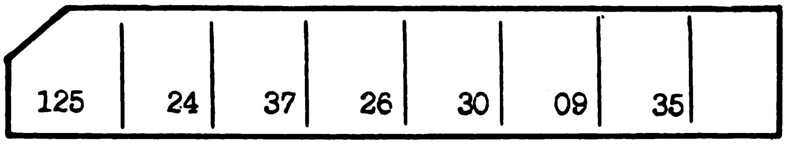


Figure 2.

This card is one of a set; all the cards of the set are alike in form. The cut corner prevents cards being inadvertently reversed when sorting. Note that there is but one horizontal line of data. This makes it possible to lay the cards in a vertical row and add the numbers in the column corresponding to a given trait just as columns are added on a data sheet. The sum of the numbers in the same field divided by the number of cards in the set is the average score in the corresponding trait.

In finding average products we would save time by bringing together cards that have common factors. For example, if we want the average product of scores on Traits 1 and 2, we can sort our cards into score-groups on variable 1 and multiply the sum of the variable 2 scores in each group by the common value of variable 1 in that group.

(X_1)	(X_2)	
24 ×	15 =	
24 ×	53 =	
24 ×	10 =	
24 ×	5 =	
24 ×	83 =	1992
23 ×	50 =	
23 ×	50 =	
23 ×	62 =	
23 ×	7 =	
23 ×	169 =	3887
. . .	. . .	
. . .	. . .	
	. . . =	. . . .
ΣX_1X_2 =		. . . .

We have only to multiply 24 × 83, 23 × 169, 22 × (the sum of score 2 for the 22 group), 21 × (the sum of score 2 for the 21 group), and so on, add the extensions and divide by the total number of cards to find the average product. The same kind of computation will give us the sum of the squares of the sorted variable; we have only to multiply the class-number of each group by the sum of the class-numbers in that group and add the extensions.

It will be noticed that sorting the cards on variable I prepares the set for computing any product-moment that involves variable I as one of the factors. If all the possible products involving variable I are to be computed, we must add and extend in each group all the fields of the card, including the field sorted on. The sum of each column of extensions will then be the sum of the products of variable I scores with the scores of the corresponding variable. In computing the set of products involving variable 2 we would sort the set of cards on variable 2 and omit the addition and extension of variable I products, since we have already found this sum on the first sort. Similarly, in finding the products with variable 3 we would omit the additions and extensions of both variables I and 2 since these have already been found.

Score 1 Classes		*Class Sub-totals of Score 2*		*Extensions*
5	×	81	=	405
4	×	32	=	128
3	×	47	=	141
2	×	19	=	38
1	×	85	=	85
				797

Score 1 Classes	*Cumulative Totals of Score 2*
5	81
4	113
3	160
2	179
1	264
	797

However, it is not necessary to make any extensions at all if we record, instead of group sub-totals in the other scores, *group cumulative-totals*. The sum of a complete series of cumulative totals is the same as the sum of the products of the sub-totals with their corresponding class-numbers.

In the numerical example at the left the sum of the five extensions is the same as the sum of the cumulative totals of the sub-totals.

The reason for this is seen when we examine the formation of the sum of the cumulative totals, the 81 appearing in the total of the cumulative totals 5 times, the 32 4 times, and so on.

$$
\begin{array}{rl}
81 = & 81 \\
113 = & 81 + 32 \\
160 = & 81 + 32 + 47 \\
179 = & 81 + 32 + 47 + 19 \\
264 = & 81 + 32 + 47 + 19 + 85 \\
\hline
797 = & 5\times81 + 4\times32 + 3\times47 + 2\times19 + 1\times85
\end{array}
$$

This method of making the summations of the products may also be applied in series having gaps in the sorted field if we treat it as a complete series having some sub-totals equal to zero. We leave a space for each missing number and form cumulative totals for each step including the gaps. These will simply be repetitions of the last preceding total arrived at in the process of accumulation. For example:

5	81	81
4	32	113
		113
2	19	132
1	85	217

656 = 5×81 + 4×32 + 3×0 + 2×19 + 1×85

This method of correcting for gaps may also be applied to series terminating in some number other than 1. The correction consists in adding the last cumulative total as many times as there are lines in the series when it is extended down to 1; that is, the product of the last-line cumulative total times the next class-mark below the end of the table. It is convenient to indicate this correction by writing the *last* class-mark after the last cumulative total as a multiplier; in this way the addition of the column and the correction of its sum are accomplished in one operation.

In the example on the following page the sum of the cumulative totals is only 656. The correction, twenty lines of 217's, or 4340, added to this is 4996; the addition of the twenty 217's being performed immediately after adding the last number of the column, we are, in effect, adding twenty-one 217's as the last number of the sum.

25	81	81
24	32	113
		113
22	19	132
21	85	217 ×21
		4996 = 25×81 + 24×32 + 22×19 + 21×85

This method of computing product-moments involves only the operations of grouping and addition, and yields easily to the flexibility of the Hollerith Printing Tabulator.

The measures to be averaged are written on cards in holes, the distance of a hole from the top of the card indicating the value of the digit it represents, and the column of the card in which the hole is punched corresponding to the decimal place and field-position of the number. The card is divided into fields in exactly the same way as the index card of the preceding example, each field covering enough columns to provide for the number of digits in the highest score expected in that trait.

The tabulating card shown in Fig. 3 corresponds to the index card of Fig. 2.

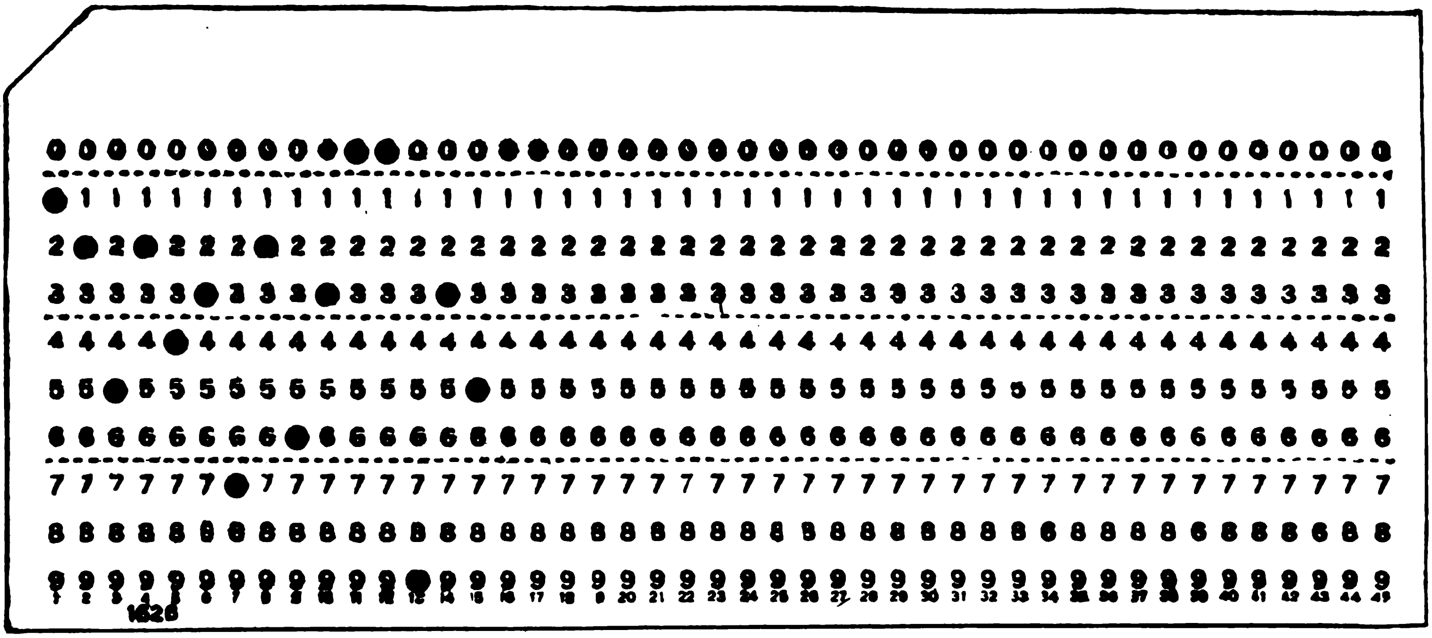


Figure 3.

The numbers are punched on the cards by a key-punch operator using a keyboard machine, and the set of punched cards is checked by repeating the motions of punching each card in a verifier, a machine which detects errors by failing to space the punched card when the key pressed does not agree with the number previously punched on the card. The cards are grouped by a sorting machine which deals cards into boxes corresponding to the number in any selected column, and they are added on an electric tabulator which is a set of five adding machines whose accumulating wheels are operated by electrical contact through the number-holes on the cards. The tabulator is equipped with a "control circuit" which prints the group designations, stops the process of feeding the cards when the class-number changes, prints the total of the selected fields and automatically starts feeding cards again. The tabulator may be set to clear its counters after every total or to print totals without clearing. The latter setting gives the necessary cumulative totals and is made by setting the counter re-set collars to the position for a grand total and turning the Progressive Total switches to the *On* position. The operation of the Hollerith Tabulator requires about the same amount of mechanical aptitude that is required to operate a typewriter; its principles and scope of application are best learned by direct observation.

Preparation of Data

Measures to be correlated should be recorded on data sheets large enough to permit writing all the measures for one person on one horizontal line. The form shown in Fig. 4 is satisfactory for general work.

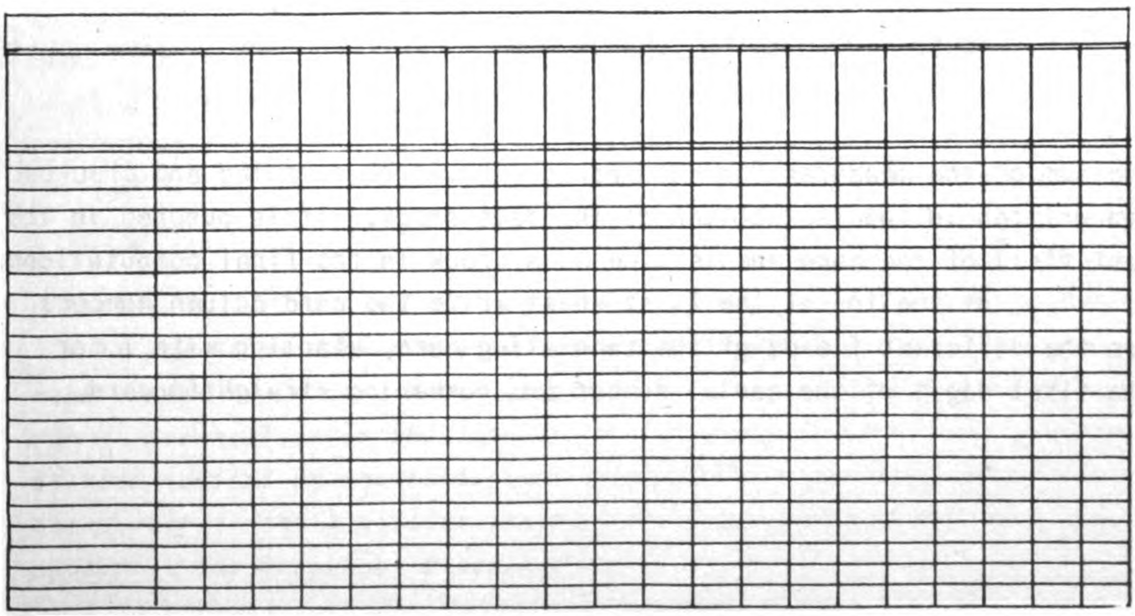


Fig. 4.

A convenient size is 8 1/2" by 11" with 40 lines parallel to the 11" edges and divided into 20 columns. Each line should be numbered in serial order for identification. If the original measures are not all positive numbers between 00 and 99 it is better to group them. Any measure can be expressed as a two figure positive number by the proper choice of scale. This allows a working range of 100 steps in the distribution; thirty is normally sufficient for all purposes.

If the original measures are not all positive two-digit numbers they may be transformed into all-positive numbers by adding to each score the estimated value of the lowest possible negative measure. This estimated value should be about ten points larger than actually expected in order to allow for errors in approximation. The measures should then be divided by about a fiftieth of their estimated range in order to make them all two digit numbers. The operations of adding the constant and dividing by the step-interval are made at the same time in preparing the data for the key-punch operator.

Key-punch operators are comparable to stenographers in intelligence and skill; the same care should be used in the preparation of data for the key-punch that is used in the preparation of manuscript for the typist. It will assist the operator if all numbers of the same field are written in the same number of digits; for example, if the serial number is expected to run as high as 1500 the lower serial numbers should be written as 0001, 0002, • • • 0035, • • • 0548, etc., and if the maximum score in a trait is a two digit number, all the scores in that trait should be written in two digits:

0001	03	15	22	47	69	00	12	01	10	179
0002	09	25	59	08	15	05	10	06	59	196
. .	.	.	.	.	.	.	.	.	.	.

The sum of the measures for each person should be computed and checked and written in the last column of the data sheet. It is punched in the last field of the card and is used as a check in the final computations.

At the top of the first sheet write the card column numbers for the different fields of the tabulating card, starting with I for the first digit of the serial number and numbering straightforward, one column per digit, to the last digit of the check sum[1]. Allow enough columns to a trait to take care of the maximum number of digits expected in that trait's score. It is not necessary to have lettered card forms printed for research work; they are of some assistance in accounting tabulating where the same set of forms and tables are used over a period of years. The Carnegie Foundation's Pennsylvania Study data were punched on the form shown in Fig. 5; the lines after every third column guide the eye when reading cards for identification.

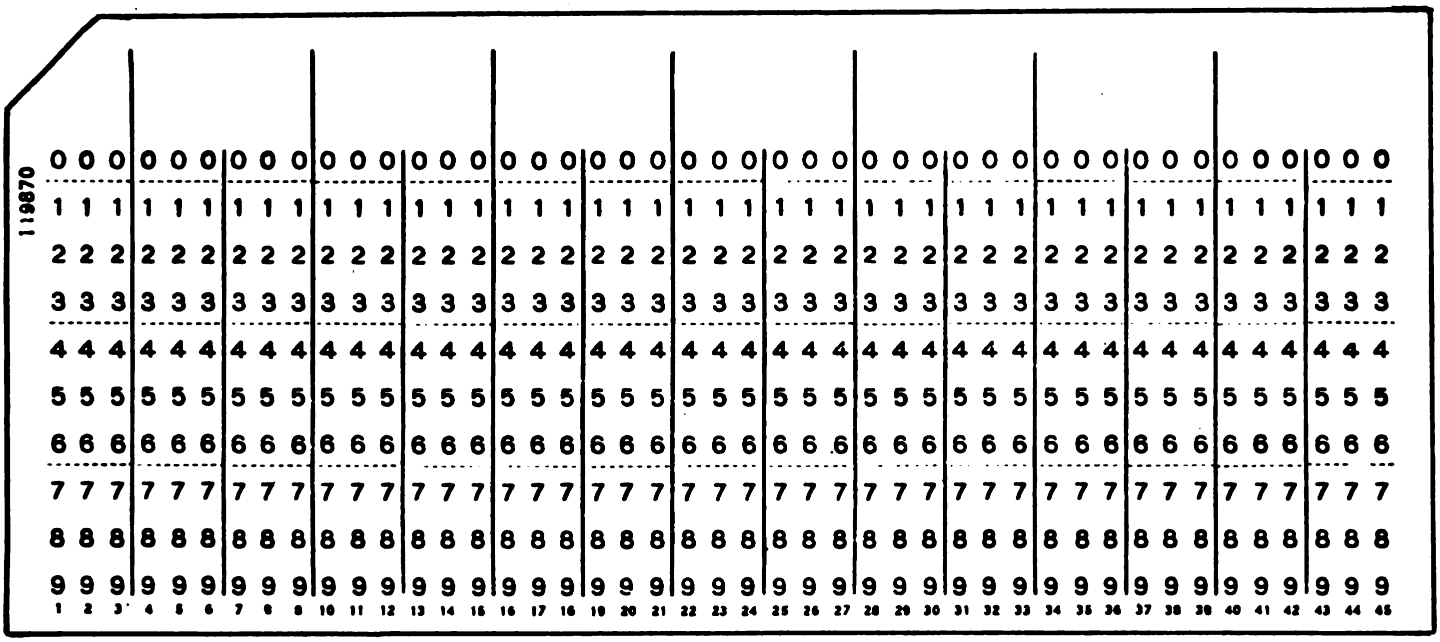


Fig. 5.

Sorting

Sort the cards on the units column of variable I; pick up the cards with the highest number on the face of the stack. Sort on the tens column and again pick up with the highest number on top. Look through the first few groups and insert blank cards for each missing number in the sorted field. For example, if the highest measure is 69 and the next card is a 65 insert three blank cards for the 68, 67, and 66. The bottom of the stack may also have some missing measures and these should also be filled in with blank cards.

[1]*N.B. If an appreciable number of zero frequencies are to be expected in any score, the operator should be instructed to punch a 1 for a "checker" after the last column of every card. This column will be used instead of the card-count circuit and blank cards inserted by the sorter in the positions occupied by the missing scores will not appear in the frequency distribution.*

Tabulation

Wire the tabulator to control on, and indicate in the first counter, the field of variable 1. Wire the checker hole to Counter 2 (or use the card-count); in Counter 3 add variable 1, in Counter 4 add variable 2, in Counter 5 add variable 3. Turn on all the Progressive Total switches and set the re-set collar of Counter 2 to clear. All other counters are to be in the position for grand totals. Tabulate the cards and clear the machine. Re-wire for the same sort and control and to print cumulative totals of fields 4, 5, 6, and the Check Sum, in counters 2, 3, 4, and 5. It is not necessary to repeat the card-count or frequency distribution on this second run, and if there are more than 5 variables, all four counters (2, 3, 4, and 5) may be used to print cumulative totals for product-moments.

When all fields have been tabulated in cumulative totals with trait 1 as the control, sort the cards on the next field, field 2, inserting blank cards for the missing measures as before. Wire the tabulator to control on and indicate this field in the first counter, to count cards in the second, to print cumulative totals of field 2 in the third, and to print cumulative totals of fields 3 and 4 in counters 4 and 5.

Wiring directions for tabulator operators are written as follows:

	1	2	3	4	5	*Counters*
Sort 4–5	4–5	c.c.	4–5	6–7	8–9	*Card–Columns*
(Highest on Top)	C. & I.	Sub.	Cum.	Cum.	Cum.	*Re–set*
Table 1 Sheet 1	1	f	1	2	3	*Variables*
	1	2	3	4	5	*Counters*
Second run	4–5	10–11	12–13	14–15	16–17–18	*Card–Columns*
Same sort	C. & I.	Cum.	Cum.	Cum.	Cum.	*Re–set*
Table 1 Sheet 2	1	4	5	6	S	*Variables*

These directions show the counter numbers in the top row, and the card column numbers of the fields added in each counter, in the second row. The lowest row of figures are the column headings which are to be written on the tabulation sheets by the operator.

Since the tables for the computation of correlation coefficients are all of the same form, a simpler diagram may be drawn that will show all the settings for all the runs.

It is understood that (when a five-bank tabulator is used) all of the tables use Counter 1 to indicate the score-group of the variable sorted on and that on the first run of each sort Counter 2 is used for the card-count. It is not necessary to repeat this information. The diagram for all the settings for a 6 variable problem is shown in Fig. 6.

Job No.__________

	1	*2*	*3*	*4*	*5*	*6*	*S*	
(*Variables*)	*1*	*2*	*3*	*4*	*5*	*6*	*S*	
(*Card Columns*)	4–5	6–7	8–9	10–11	12–13	14–15	16–17–18	
(*Sorted on 1*) →	3	4	5	2	3	4	5	(*Counters*)
" 2 →		3	4	5	3	4	5	"
" 3 →			3	4	5	4	5	"
" 4 →				3	4	5	5	"
" 5 →					3	4	5	"
" 6 →						3	5	"

Figure 6.

The numbers at the top of the table are the numbers of the traits with the card-column numbers of their fields. The numbers in the body of the table give the number of the counter in which the trait is to be added. All the tables on the first line are made with the cards sorted on field 1, those on the second line are made with the cards sorted on field 2, and so on.

The tabulator operator will check the indicating column of the tables and write in pencil at the bottom of each column the correction factor to be applied to the last-line total to bring the sum of the cumulative totals to the value it would have if the table extended on down to control-group I. As was said before, this correction is made by writing the lowest class-mark opposite the last-line total of each column as a multiplier. If the series of measures for any of the sorted traits ends in zero, the 0-class totals are not included when the columns are added.

The tables are sent to the calculators for adding, and the computations of the coefficients are made as indicated in the directions on the form for computation of inter-correlations.

Variations of the procedure outlined here will at once suggest themselves. For example, if the grand totals of a pair of variables do not exceed four digits each, the two variables may be added in two halves of the same counter. If all the measures are of this size the capacity of the tabulator may be doubled by "splitting" counters in this way. There are nine wheels in each counter and any counter may be used to total two traits if one of the traits runs to a four digit total and the other to five. 4-4 splits are convenient to use but 4-5 splits should only be indicated when an appreciable number of runs will be saved by using them.

It will be noticed that all the tables are run to cumulative totals only. If array-means are wanted for the computation of η's or regression line ordinates, the cards should be tabulated to group sub-totals after every run for cumulative totals. The group sub-totals are the differences between the lines of the cumulative total columns. Sub-totals are obtained by setting the re-set collars to clear. Dividing the card-counts into the sub-totals of a column gives the array-means for that trait.

The computation of higher order product-moments may be made by punching total-cards from the cumulative total lines and running these to cumulative totals or by accumulating the second and higher summations by hand on the tabulation sheet.

Two-way frequency distributions are made in pairs by sorting on both traits, controlling on both fields simultaneously and running to card-counts. The Major-minor Control tabulator will print automatically in one run, the cell frequencies, the array frequencies, the array sums of the scores, and the total population.

ILLUSTRATION OF M-W-H CORRELATION METHOD

by

Robert M. Mendenhall

THE DATA SHEETS FOR JOB G-7

The data sent by Prof. G. are reproduced in full below in the same form as they appeared on his original copy, with the single exception that his figures were written with pen and ink. This form is ideal for data sheets, and clearly reflects the fact that this research scholar has had considerable experience in using Hollerith tabulating machines. The pages which follow are fac-similes of tables secured by running the 210 cards through a Hollerith printing tabulator, wired for listing. (Measures listed from the punched cards in this manner can be compared with the data sheets by unskilled clerks, and erroneously punched cards can be quickly located by means of the serial numbers. This method of verifying will be discussed later).

Serial numbers of students tested	Scores on each of six tests						Check-sum (Used only for checking purposes)
	Analogies	Sentences	Arith-metic	Number-series	Equations	Infor-mation	
Ser. No.	1	2	3	4	5	6	S
001	28	32	29	38	20	41	188
002	24	29	26	38	16	20	153
003	25	33	17	22	09	57	163
004	26	34	30	34	19	39	182
005	34	32	34	36	13	49	198
006	22	25	14	36	09	32	138
007	28	34	19	33	17	32	163
008	30	33	18	33	11	30	155
009	26	32	21	39	12	49	179
010	25	32	14	28	14	22	135
011	25	36	26	37	10	41	175
012	32	37	19	36	11	25	160
013	32	32	26	32	22	40	184
014	27	32	25	34	12	20	150
015	34	35	20	34	17	24	164
016	26	34	24	36	11	33	164
017	38	39	26	32	11	37	183
018	29	33	17	37	11	28	155
019	29	36	18	29	16	28	156
020	25	34	26	37	13	30	165
021	25	35	21	26	14	23	144
022	33	36	14	17	11	25	136
023	29	35	11	27	10	42	154
024	30	36	26	37	14	53	196
025	30	35	13	35	11	30	154
026	28	33	21	25	14	38	159
027	25	31	15	30	15	30	146
028	28	30	16	23	20	42	159
029	29	30	21	32	18	45	175
030	22	32	21	31	14	46	166
031	25	34	13	18	13	55	158
032	25	34	10	36	17	53	175
033	25	32	30	38	13	31	169
034	31	24	29	31	11	37	163
035	39	38	39	40	17	39	212
036	24	35	20	35	11	53	178
037	30	35	30	37	10	57	199
038	28	33	11	26	11	31	140
039	23	34	26	37	19	43	182

DATA SHEETS FOR JOB G-7

Ser. No.	1	2	3	4	5	6	S
040	20	33	18	29	17	23	140
041	25	24	20	38	16	28	151
042	26	26	18	39	10	57	176
043	26	30	16	33	11	25	141
044	24	33	20	25	11	50	163
045	34	33	16	28	10	20	141
046	29	34	17	36	14	34	164
047	35	36	10	16	10	32	139
048	31	31	27	39	20	57	205
049	20	29	15	36	10	46	156
050	20	30	17	27	20	46	160
051	33	28	19	22	09	45	156
052	28	34	14	31	10	27	144
053	26	38	31	36	16	57	204
054	30	33	15	40	14	31	163
055	26	29	28	39	20	50	192
056	23	33	21	34	18	37	166
057	33	35	31	36	16	55	206
058	22	38	28	38	18	47	191
059	25	32	34	38	21	25	175
060	28	33	21	23	22	39	166
061	25	30	17	32	11	32	147
062	29	37	23	33	11	38	171
063	26	37	21	27	11	33	155
064	33	31	18	32	10	30	154
065	27	36	33	40	23	48	207
066	27	37	21	34	10	40	169
067	36	33	31	37	17	48	202
068	34	34	24	38	11	36	177
069	23	31	19	39	08	44	164
070	33	35	20	32	20	56	196
071	26	34	18	30	13	58	179
072	24	31	17	38	13	32	155
073	27	33	11	39	13	43	166
074	21	32	26	15	10	45	149
075	22	32	13	30	09	34	140
076	25	30	27	37	12	36	167
077	24	37	34	34	18	51	198
078	27	33	28	39	15	45	187
079	29	36	36	29	22	57	209

Ser. No.	1	2	3	4	5	6	S
080	32	31	09	18	14	25	129
081	20	34	21	35	16	33	159
082	22	36	25	37	12	41	173
083	27	36	19	21	11	51	165
084	24	30	11	29	05	25	124
085	24	14	22	28	10	36	134
086	28	33	24	36	18	52	191
087	30	35	19	27	08	30	149
088	18	29	15	32	10	48	152
089	28	34	23	37	17	54	193
090	27	33	20	32	16	48	176
091	27	27	19	35	10	51	169
092	25	31	29	40	07	36	168
093	33	35	23	39	11	40	181
094	23	28	21	33	20	28	153
095	24	37	14	29	17	33	154
096	28	36	18	37	12	57	188
097	26	33	27	40	11	56	193
098	27	22	17	17	14	34	131
099	22	30	16	32	18	36	154
100	31	38	29	39	18	40	195
101	25	31	16	26	11	23	132
102	34	34	24	31	17	34	174
103	29	33	22	33	10	33	160
104	28	39	17	35	13	20	152
105	21	33	14	31	11	40	150
106	34	37	28	34	10	48	191
107	24	33	11	29	11	25	133
108	29	33	30	40	17	49	198
109	21	37	14	22	09	32	135
110	28	35	25	37	16	52	193
111	23	32	18	29	17	42	161
112	27	39	22	33	13	38	172
113	25	34	30	09	12	38	148
114	28	36	21	38	10	34	167
115	31	34	22	37	19	57	200
116	27	35	14	24	15	29	144
117	30	26	25	28	16	50	175
118	21	33	15	23	12	53	157
119	27	37	32	32	11	40	179

DATA SHEETS FOR JOB G-7

Ser. No.	1	2	3	4	5	6	S
120	33	39	27	30	16	44	189
121	23	34	25	35	15	57	189
√122	27	36	10	16	12	30	131
123	23	36	19	16	09	36	139
124	29	32	28	29	16	49	183
125	35	31	25	40	11	42	184
126	28	32	15	28	11	20	134
127	29	38	17	30	10	33	157
128	32	34	15	22	15	21	139
129	33	34	20	34	11	25	157
130	33	33	22	32	12	28	160
131	22	17	19	23	11	33	125
132	26	34	32	34	12	53	191
133	29	31	28	38	17	56	199
134	36	34	20	34	11	48	183
135	26	33	19	25	10	23	136
136	35	38	25	31	12	49	190
137	30	28	17	38	10	39	162
138	29	37	31	32	11	30	170
139	28	31	12	18	09	30	128
140	26	30	37	38	19	56	206
141	23	29	18	18	11	32	131
142	29	34	34	38	21	43	199
143	24	38	15	27	17	01	122
144	27	36	19	34	20	44	180
145	24	36	22	19	10	39	150
146	29	35	33	36	19	43	195
147	28	34	12	29	17	27	147
148	26	37	19	30	11	25	148
149	22	31	23	31	11	43	161
150	32	32	19	36	15	34	168
151	25	31	33	30	11	47	177
152	24	32	19	33	13	34	155
153	30	23	23	36	16	57	185
154	28	35	19	38	11	32	163
155	32	34	20	24	21	49	180
156	21	35	25	22	12	30	145
157	33	31	13	27	09	15	128
158	30	31	18	32	11	52	174
159	23	22	24	37	11	52	169

Ser. No.	1	2	3	4	5	6	S
160	18	19	15	36	20	39	147
161	34	38	30	37	19	35	193
162	26	35	14	32	18	38	163
163	24	32	16	21	09	27	129
164	31	37	25	34	14	29	170
165	27	32	17	33	16	50	175
166	27	34	18	37	18	58	192
167	32	34	29	37	09	36	177
168	26	32	21	27	15	28	149
169	25	33	11	25	07	27	128
170	33	36	24	37	12	31	173
171	30	33	23	36	10	32	164
172	23	22	25	39	10	49	168
173	34	36	28	40	18	51	207
174	24	23	28	38	19	45	177
175	38	39	31	38	20	42	208
176	29	37	15	24	11	35	151
177	34	37	29	35	13	22	170
178	28	33	15	16	08	23	123
179	39	36	26	34	11	47	193
180	32	34	23	32	14	55	190
181	19	33	17	33	15	48	165
182	21	35	19	19	16	25	135
183	30	34	23	28	14	51	180
184	22	37	13	39	09	29	149
185	30	35	22	37	15	54	193
186	25	31	20	37	10	32	155
187	24	29	14	16	07	33	123
188	23	34	13	26	10	49	155
189	21	33	22	31	16	45	168
190	28	30	25	36	11	40	170
191	24	36	31	39	19	34	183
192	26	36	18	24	13	35	152
193	33	35	26	36	15	56	201
194	30	35	22	38	15	57	197
195	31	38	18	27	15	52	181
196	23	35	21	22	11	32	144
197	23	31	17	13	14	33	131
198	24	28	21	32	17	46	168
199	26	28	22	40	14	45	175

DATA SHEETS FOR JOB G-7

Ser. No.	1	2	3	4	5	6	S
200	29	26	21	19	14	17	126
201	24	31	13	27	11	41	147
202	24	37	18	29	08	25	141
203	25	37	10	22	16	24	134
204	29	36	24	33	13	28	163
205	24	35	19	27	11	31	147
206	32	31	17	31	09	22	142
207	26	32	33	35	19	57	202
208	31	24	19	19	10	29	132
209	25	31	21	18	15	29	139
210	26	29	12	25	10	34	136

The important features which make the form of the data sheets reproduced here ideal are as follows:

1. ADAPTATION FOR CARD PUNCHING - All the measures for a given individual are placed in a single horizontal line exactly in the order in which they are to be punched on the card. This is the most convenient form for the punch operator. If the measures are not in the order in which they are to be punched on the card or if they are presented in any other form than the horizontal line the speed of the punching is reduced.

2. SPACING - Figures are liberally spaced both laterally and vertically, which makes reading easier and correspondingly reduces the chance of error from this reading. It is, of course, desirable that data sheets be typewritten, but it is not essential. Data sheets filled out by pen and ink in average or better handwriting are adequate; but the use of pencil on data sheets is to be avoided.

3. SERIAL NUMBER - Every line of data should begin with the serial number of the individual on whom the measurements are taken. The serial number is always necessary. If the client omits the serial numbers it is necessary to supply these before giving the data-sheets to the operators. The serial number is the only convenient link between the punched card and the original data sheet. If a card is lost it can easily be identified by the aid of the serial number.

4. CHECK SUMS OF SCORES - In the data sheets reproduced above, the last column shows the sums of all the measures on each line. This check sum is useful in verifying the punching and in verifying the calculations; and if the check sums are not supplied they must be entered by the Bureau before the data sheets are handed to the punch operator.

5. MEASURES STATED IN TWO-PLACE NUMBERS - All the measures on the data sheets reproduced above are two-place numbers. In this particular job these measures are actual test scores so that no coding was necessary in order to stay within the limits of two-place numbers. Three and four-place numbers can be handled on the Hollerith machines, but they add at least 35% to the cost of the average job. Investigators using measures in three-place numbers will save time and money by coding these measures in two-place numbers in such a way

as to maintain an adequate number of significant steps for correlations and other statistical purposes. The only exception to the "two-place number rule" on the data sheets reproduced above appears in the last column. The check sums of the scores must always be the exact sums of the scores in each line.

6. ACCURACY OF DATA SHEETS - The data sheets reproduced above bore evidence that they had been checked twice independently. The Statistical Bureau cannot, of course, go back of the data sheets received and since very small errors may sometimes make it necessary to do most or all of a job over again the only protection of the owner against loss is to make sure that the data sheets are absolutely accurate and legible.

PUNCHING

The first step after receiving the data sheets and making sure that they were in proper form was to have the serial number, the various measures and the check sum for each individual punched on cards, using a separate card for each individual. In the job under consideration this task was performed with the aid of an electric duplicating key-punch, illustrated in Plate 2. The punching of the 210 cards involved in this job was done in less than ninety minutes by an operator who was receiving sixty-five cents an hour. The average operator can punch between 150 and 175 forty-five column cards per hour, or about 1000 per day. The best operators, when the data sheets are in good form can average two hundred cards per hour. The card reproduced below is punched for the first line of the data sheet above, that is, for the individual whose serial number is 001.

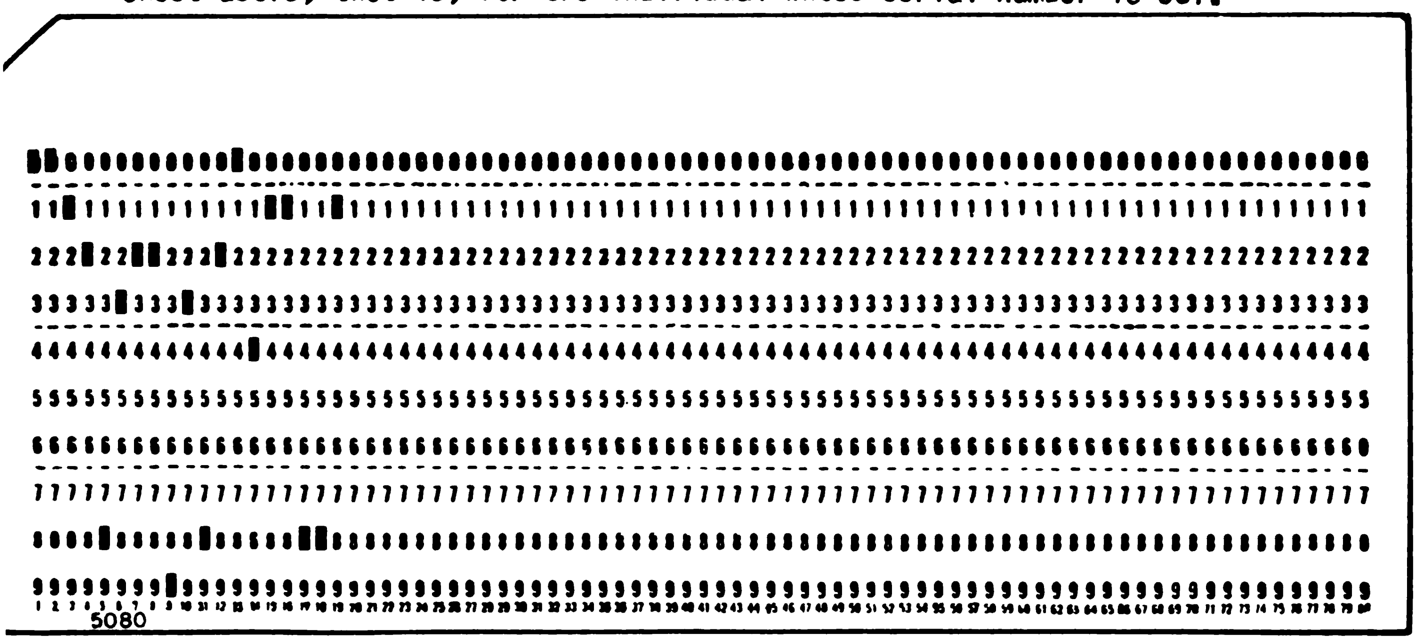


Figure 8.

VERIFYING

An essential step in the procedure is verifying the punching of the cards. The importance of this step cannot be overestimated. The process of verifying cards takes as long as the original punching or longer, depending on the method used. The following five methods are commonly used:

1. Two sets of cards, one white and one red, are punched by two different operators working independently. These cards are then superimposed, one pair at a time over a strong diffused light. A disagreement between the first and second punching shows through the covered spot.

ADVANTAGES	DISADVANTAGES
Only method of checking errors of double-punching, i.e., two holes punched in a single column of a card.	*Hard on operator's eyes. Extra time for matching cards. Doubles number of cards used (not a disadvantage when two sets of cards are desired for other reasons). Errors may be overlooked by careless or tired operator.*

2. Reading cards directly against the data sheets.

May be done by unskilled clerks. Requires no machine.	*Requires two people, one reading and one checking. Reading cards takes longer than punching them.*

3. Listing the cards on the printing tabulator and comparing the list with the data sheets.

May be done by unskilled clerks List may be useful for other purposes.	*Requires two people, one reading and one checking. Requires tabulating machine time for listing.*

4. Running cards to be verified through the tabulator to a control balance.

Rapid method of proving accuracy where no error exists.	*May have compensating errors in set. Difficult to locate error card after fact of error is established.*

5. Verifying with Mechanical Verifier.

Rapid and accurate. Locates errors automatically.	*Will not locate errors in a double punched code.*

Verifying with the mechanical verifier is normally the best method of checking punching. This method has no rival when one set of cards is desired and when there is no double punching. The cost of the mechanical verifier is only $5.00 per month and it may be rented on a monthly basis. The verifying of the cards for this job was done by means of the mechanical verifier.

SORTING

As soon as the cards were verified, they were prepared for the first run through the tabulator by being sorted on variable 1, the highest scores being on top. The capacity of the sorter is 350 cards per minute, and of the Hollerith printing tabulator about 150 per minute. As soon as Table 1 (see below) was secured (from the first run of the cards through the tabulator), they were prepared for the second run (Table 2) by being sorted on variable 2; and so on until all six tables were done. Since in this job array-means were desired, Table 1A (see below) was run immediately after Table 1, and before sorting on Variable 2, in order to avoid re-sorting on Variable 1.

TABULATING

The tables produced by means of the sorting machine and the Hollerith Printing Tabulator are reproduced in full below, with explanatory notes. In a job of this sort the table number is always the same as, or is derived from, the number of the variable on which the cards have been sorted, i.e., the number which the tabulator operator puts at the top of the extreme left hand column of each table. (See below, Table 1 and Table 1A). If the number of variables is large, or if the measures are expressed in three or four-place numbers, there may be two or more sheets for each table. In such cases we have Table 1 sheet 1, Table 1 sheet 2, etc. Since this job was done on a seven-bank Hollerith Printing Tabulator, having two indicating banks and five adding banks, and since all measures were expressed in two-place numbers, and only six variables were involved, the whole of Table 1 was secured at one run and appears on one sheet, as shown in the slightly reduced facsimile of Table 1 below. This result was obtained by "splitting" the second, third, and fourth counters, making each counter carry the cumulative sums for two variables. If the measures are expressed in one-place numbers (as in correlating High School or College letter-grades, A, B, C, etc.), each counter may carry three variables, the total capacity of the 7-bank machine then being fifteen cross-products, or fifteen cumulative sums. Under certain limited conditions each counter may carry three variables when the measures are two-place numbers.

In running Table 1, the second indicating bank was used for indicating the control groups on Variable 1, the control variable, the first indicator not being used at all. The first counting bank was used for the card-count (CC), that is, to show the number of cases in each control group on Variable 1 (frequency distribution). The second counting bank was "split", the first half carrying the cumulative totals of Variable 1, and the second the cumulative totals of Variable 2. Counters 3 and 4 were split to carry the cumulative totals of Variables 3, 4, 5, and 6; and counter 5 carried the cumulative totals of the Check Sum, S.

Table 2, in the same form as the preceding table, is sorted, controlled, and indicated on Variable 2, and omitting all figures referring to Variable 1, since all of the summations involving Variable 1 have been obtained in Table 1.

ADDING OPERATION

When Tables 1-6 had been run they were attached to the "Operating Directions" given to the tabulator operator with the verified cards[1], and were sent to the adding machine operator, who entered in pencil the sums of the various columns, taking account of the corrections for assumed origin at zero. In Table 1 the last figure in each column was added in the equivalent of 18 times; in Table 2 the last figure in each column was added in 14 times. In order to avoid possible error or confusion on the part of the adding-machine operator, it is desirable to write the multiplier near the last figure in each and every column of a given table before the operator starts adding any of the columns. In Table 2, "14×" is written at the left of the last figure in every column to be added, except the frequency column.

Verifying the sums of the columns is essential, and must be done with the greatest care, preferably by a second operator. This

[1]*The cards were filed and will be kept for a year, or longer by arrangement with the investigator.*

is best done by doing the additions over again, *de novo,* and writing the sums below the first operator's sum-entries. In case of disagreement, the column should be added anew until three consecutive additions agree. The wrong answer may then be scratched out, not erased.

CALCULATOR COMPUTATIONS

When the additions had been checked, Tables 1-6 were passed on to the Monroe Calculator operator, who calculated all means, sigmas, and intercorrelations in about one hour, using the Mendenhall-Warren-Hollerith Form for Computing Intercorrelations. This form includes detailed directions for a convenient method of computing and checking. The reproduction below is a facsimile of an exact copy of the form which was used by our operator in the course of making the calculations for the job under consideration. (See Fig. 9).

TABLES FOR COMPUTING ARRAY-MEANS

As indicated above, immediately after running Table I, and before the cards were sorted on Variable 2, Table IA was run. The cards being already sorted on Variable I, the tabulator was wired to print non-cumulative sub-totals of all the variables, thus showing for each class-interval of Variable I the sums of the corresponding scores on Variables 2, 3, 4, 5, and 6, respectively, with the card-count or frequency distribution carried by the right-hand half of the fifth counter.

Tables 2A to 6A, inclusive, were run in a similar way. Only Table IA is reproduced below.

REPORT TO INVESTIGATOR

Carbon copies of all tables were made for this job, as in the case of most jobs. The report to Prof. G. included the originals of Tables IA to 6A, inclusive, with all array-means calculated, and the following summary taken from the M-W-H Form reproduced above, and from Tables 1-6:

COLUMBIA UNIVERSITY STATISTICAL BUREAU

Report to Prof. J. G. G.

Variables	(1)	(2)	(3)	(4)	(5)	(6)
Sigmas	4.1824	3.9742	6.2358	6.7348	3.7023	11.1606
Means	27.3714	32.8667	21.2143	31.2857	13.5333	38.6238
r_{1n}		0.2857	0.2478	0.1660	0.0540	0.0356
r_{2n}			0.1161	0.0403	0.0701	-0.0255
r_{3n}				0.4517	0.3754	0.3728
r_{4n}					0.2512	0.3083
r_{56}						0.2702

TABLE I

Facsimile of table resulting from the first run of the cards through the Hollerith 7-bank printing tabulator. This table includes one distribution (fX_1) data for one sigma (ΣX_1^2), and for seven means and for six cross-products. The first two rows of notations at top of the table are entered only for purposes of this exposition; the tabulating operator enters only the notations in the third line, these being sufficient to identify the columns. The comptometer operator enters only the sums at the bottom of the columns, all other notations appearing below having been entered only for the convenience of the reader.

Indicator#2	Counter#1	Counter #2		Counter #3		Counter #4		Counter #5
Analogies Test Scores	Card-Count f₁	ΣX_1^2	Products of variable 1 scores with scores of the remaining five variables and with the sums of all six variables.					
1	CC	1	2	3	4	5	6	S
39	2	78	0074	65	0074	28	0086	405
38	2	154	0152	122	0144	59	0165	796
		154	0152	122	0144	59	0165	796
36	2	226	0219	173	0215	87	0261	1181
35	3	331	0324	233	0302	120	0384	1694
34	9	637	0640	466	0615	248	0703	3309
33	12	1033	1048	723	0989	400	1153	5346
32	9	1321	1347	900	1257	530	1460	6815
31	7	1538	1573	1069	1483	637	1761	8061
30	14	1958	2025	1363	1965	812	2384	10507
29	18	2480	2638	1789	2540	1073	3072	13592
28	19	3012	3275	2146	3124	1340	3763	16660
27	16	3444	3813	2471	3624	1569	4432	19353
26	20	3964	4456	2932	4287	1844	5283	22776
25	21	4489	5142	3372	4911	2111	6002	26027
24	20	4969	5778	3763	5504	2364	6683	29061
23	13	5268	6179	4030	5882	2537	7217	31113
22	9	5466	6457	4202	6179	2648	7558	32510
21	7	5613	6695	4337	6342	2734	7828	33549
20	4	5693	6821	4408	6469	2797	7976	34164
19	1	ΣX_1 → 5712	6854	ΣX_3 → 4425	6502	ΣX_5 → 2812	8024	34329
18	2	18X → 5748	6903 ← ΣX_2	4455	6570 ← ΣX_4	2842	8111 ← ΣX_6	34628 ← ΣS
	210	161,004	189,908	123,301	180,812	77,965	222,358	955,348
	ΣfX_1	ΣX_1^2	ΣX_1X_2	ΣX_1X_3	ΣX_1X_4	ΣX_1X_5	ΣX_1X_6	ΣX_1S

Corrected sums of product-moments, origin arbitrarily fixed at zero. The columns are added and the corrections are made by the comptometer operator.

The correction for assumed origin at zero is made by adding the last figure in each column as many times as indicated by the magnitude of the smallest score in variable 1. The last figure in each column is the sum of all the scores of the variable indicated at the top of the column.

TABLE 2

Facsimile of table resulting from second run of cards. For this run the cards were sorted on variable 2, and the tabulator was wired to control on variable 2, thus securing the distribution of variable 2 scores (fX_2) data for sigma of variable 2 (ΣX_2^2), and the cross-products of variable 2 with all other variables except 1 (ΣX_1X_2 already being given in Table I)

2	f_2	2	3	4	5	6	S
39	5	195	123	0168	73	0181	904
38	9	537	355	0473	215	0534	2549
37	17	1166	721	0999	416	1093	5294
36	21	1922	1188	1629	704	1935	8876
35	22	2692	1655	2325	1006	2824	12621
34	29	3678	2280	3250	1432	4052	17627
33	29	4635	2837	4149	1809	5154	22295
32	20	5275	3296	4779	2100	5897	25584
31	20	5895	3686	5383	2336	6616	28702
30	10	6195	3889	5702	2481	6999	30305
29	7	6398	4017	5906	2565	7262	31348
28	5	6538	4117	6071	2635	7465	32162
27	1	6555	4136	6106	2645	7516	32331
26	3	6643	4200	6192	2685	7640	32808
25	1	6658	4214	6228	2694	7672	32946
24	3	6740	4282	6316	2731	7766	33392
23	2	6786	4333	6390	2766	7868	33754
22	3	6852	4399	6483	2801	8003	34222
		6852	4399	6483	2801	8003	34222
		6852	4399	6483	2801	8003	34222
19	1	6871	4414	6519	2821	8042	34369
		6871	4414	6519	2821	8042	34369
17	1	6888	4433	6542	2832	8075	34494
		6888	4433	6542	2832	8075	34494
		6888	4433	6542	2832	8075	34494
14	1	14x6902	14x4455	14x6570	14x2842	14x8111	14x34628
	210	230,118	147,023	216,159	93,622	266,346	1,143,176
	Σf_2	ΣX_2^2	ΣX_2X_3	ΣX_2X_4	ΣX_2X_5	ΣX_2X_6	ΣX_2S

Correction

TABLE 3

Facsimile of table resulting from the third run of the cards

3	f_3	3	4	5	6	S
39	1	39	0040	17	0039	212
		39	0040	17	0039	212
37	1	76	0078	36	0095	418
36	1	112	0107	58	0152	627
		112	0107	58	0152	627
34	4	248	0253	131	0320	1397
33	4	380	0394	203	0515	2178
32	2	444	0460	226	0608	2548
31	6	630	0678	325	0874	3721
30	6	810	0873	415	1123	4810
29	6	984	1093	493	1335	5871
28	8	1208	1388	626	1726	7398
27	4	1316	1534	685	1919	8152
26	10	1576	1869	826	2331	9933
25	11	1851	2242	967	2790	11840
24	7	2019	2490	1060	3056	13051
23	8	2203	2762	1164	3426	14476
22	10	2423	3090	1298	3858	16185
21	17	2780	3585	1549	4455	18869
20	10	2980	3910	1693	4868	20572
19	18	3322	4423	1904	5492	23326
18	14	3574	4848	2086	6071	25606
17	14	3812	5270	2265	6579	27765
16	6	3908	5433	2344	6752	28621
15	11	4073	5747	2487	7099	30211
14	10	4213	6013	2608	7410	31633
13	7	4304	6215	2680	7663	32664
12	3	4340	6287	2716	7754	33075
11	6	4406	6462	2773	7947	33920
10	4	4446	6552	2828	8086	34499
9	1	9X 4455	6570	2842	8111	34628
	210 Σf_3	102,723 ΣX_3^2	143,373 ΣX_3X_4	62,116 ΣX_3X_5	177,533 ΣX_3X_6	756,069 ΣX_3S

TABLE 4

Facsimile of table resulting from the fourth run of the cards

4	f_4	4	5	6	S
40	9	360	132	0397	1707
39	12	828	297	0934	3852
38	17	1474	567	1598	6889
37	20	2214	843	2493	10529
36	17	2826	1084	3230	13515
35	8	3106	1192	3553	14888
34	13	3548	1375	4058	17182
33	11	3911	1525	4442	18955
32	16	4423	1755	5103	21680
31	9	4702	1866	5446	23138
30	7	4912	1951	5717	24274
29	10	5202	2097	6051	25822
28	6	5370	2172	6250	26721
27	10	5640	2299	6569	28213
26	4	5744	2345	6695	28784
25	5	5869	2397	6867	29506
24	4	5965	2457	7015	30133
23	4	6057	2522	7182	30740
22	7	6211	2603	7423	31756
21	2	6253	2623	7501	32050
		6253	2623	7501	32050
19	4	6329	2673	7611	32593
18	5	6419	2735	7782	33278
17	2	6453	2760	7841	33545
16	5	6533	2806	7995	34200
15	1	6548	2816	8040	34349
		6548	2816	8040	34349
13	1	6561	2830	8073	34480
		6561	2830	8073	34480
		6561	2830	8073	34480
		6561	2830	8073	34480
9	1	9X 6570	2842	8111	34628
	210	215,072	90,229	258,625	1,104,270
	Σf_4	ΣX_4^2	$\Sigma X_4 X_5$	$\Sigma X_4 X_6$	$\Sigma X_4 S$

TABLE 5

Facsimile of table resulting from the fifth run of the cards

5	f_5	5	6	S
23	1	23	0048	207
22	3	89	0184	766
21	3	152	0301	1320
20	10	352	0746	3108
19	9	523	1155	4828
18	10	703	1610	6660
17	14	958	2171	9232
16	16	1214	2836	11974
15	13	1409	3376	14172
14	15	1619	3912	16531
13	13	1788	4387	18685
12	13	1944	4902	20837
11	42	2406	6427	27540
10	27	2676	7473	31784
9	12	2784	7868	33508
8	4	2816	7990	34085
7	3	2837	8086	34504
		2837	8086	34504
5	1	5x2842	8111	34628
	210	41,340	112,113	477,385
	Σf_5	ΣX_5^2	$\Sigma X_5 X_6$	$\Sigma X_5 S$

TABLE 6

Facsimile of table resulting from the sixth run of the cards

6	f_6
58	2
57	12
56	5
55	3
54	2
53	5
52	5
51	5
50	4
49	8
48	7
47	3
46	4
45	7
44	3
43	5
42	5
41	4
40	7
39	6
38	5
37	3
36	7
35	3
34	9
33	9
32	11
31	5
30	10
29	5
28	7
27	4
25	11
24	2
23	5
22	3
21	1
20	5
17	1
15	1
1	1
Σf_6	210

6	S
116	371
800	2688
1080	3683
1245	4237
1353	4623
1618	5520
1878	6428
2133	7347
2333	8052
2725	9503
3061	10779
3202	11340
3386	11990
3701	13177
3833	13710
4048	14613
4258	15479
4422	16162
4702	17390
4936	18409
5126	19222
5237	19734
5489	20850
5594	21346
5900	22764
6197	24092
6549	25726
6704	26518
7004	28015
7149	28749
7345	29836
7453	30384
7453	30384
7728	31963
7776	32261
7891	32936
7957	33383
7978	33522
8078	34252
8078	34252
8078	34252
8095	34378
8095	34378
8110	34506
8110	34506
8110	34506
8110	34506
8110	34506
8110	34506
8110	34506
8110	34506
8110	34506
8110	34506
8110	34506
8110	34506
8110	34506
8110	34506
8111	34628
8111	34628
339435	1376410
ΣX_6^2	$\Sigma X_6 S$

Mr. D. H. Leavens, of the Research Staff of the Graduate School of Business Administration at Harvard University, has used the cumulative total method in combination with the *"Digit System"* to sum extensions of four and five digit numbers.

The Digit System corresponds to the familiar partial-product arrangement of pencil and paper multiplication; the sums of the products with the units digits of the multipliers are found first, then the sums of the products with the tens digits are found, then the products with the hundreds digits, and so on, and these partial products are added in their respective decimal positions.

Instead of sorting completely through the given field the cards are sorted on the units column first and tabulated to cumulative totals, controlling on the units column only. They are then sorted on the tens column and tabulated to cumulative totals, controlling on the tens column only; then sorted and tabulated, controlling on the hundreds column, and so on for as many digits as there are in the largest multiplier. The sums of the columns for the successive sorts (omitting, in the addition, the numbers on the 0 lines as explained previously) are put in their proper decimal position by writing the correct number of zeros at their ends. Moving the connection to the indicating counter one hole to the left after each sort identifies the decimal position of each line of sums; the number of zeros to be written in at the end of each total is then the same as the number of zeros showing in the indicating column of the table. Adding these adjusted totals then gives the sums of the cross-products.

This device is particularly useful where there is a relatively large variability in any trait. If the range of any measure is over thirty class-marks a saving in time will be effected by computing the necessary moments for the sort by this method and tabulating the frequency distribution separately after the last run of cumulative totals, since the cards will then be in order on all digits of the sorted trait.

	Indicating Column Variable 6	Cumulative Totals of 6	Cumulative Totals of S
Sorted on Units	9	771	3204
	8	1609	6751
	7	2670	10815
	6	3386	13576
	5	4251	17520
	4	4855	20155
	3	5747	23958
	2	6635	27813
	1	7231	30468
		8111	34628
	Partial Products →	37165	154260
Sorted on Tens	50	2333	8052
	40	4702	17390
	30	7004	28015
	20	8078	34252
	10	8110	34506
		8111	34628
	Partial Products →	302270	1222150
		339435	1376410
		ΣX_6^2	$\Sigma X_6 S$

Table 6A

A short method for arriving at the results of Table 6

Table I-A

Facsimile of table printed by Hollerith 7-bank tabulator, providing data for computing array-means, $M_{2.1}$, $M_{3.1}$, $M_{4.1}$, etc. The figures in each line of columns 2 to 6, inclusive, are each divided by the corresponding figure in the column headed CC (card-count or frequency). Table 2-A is similar, giving $M_{1.2}$, $M_{3.2}$, $M_{4.2}$, etc., and need not be reproduced.

Indicator # 2	Counter # 1		Counter # 2		Counter # 3		Counter # 4		Counter # 5		
Variables: 1	2		3		4		5		6		CC
Class-groups	$\Sigma X_{2.1}$	$M_{2.1}$	$\Sigma X_{3.1}$	$M_{3.1}$	$\Sigma X_{4.1}$	$M_{4.1}$	$\Sigma X_{5.1}$	$M_{5.1}$	$\Sigma X_{6.1}$	$M_{6.1}$	f_1
39	74	37.00	65	32.50	74	37.00	28	14.00	86	43.00	02
38	78	39.00	57	28.50	70	35.00	31	15.50	79	39.50	02
36	67	33.50	51	25.50	71	35.50	28	14.00	96	48.00	02
35	105	35.00	60	20.00	87	29.00	33	11.00	123	41.00	03
34	316	35.11	233	25.93	313	34.78	128	14.33	319	35.44	09
33	408	34.00	257	21.42	374	31.17	152	12.67	450	37.50	12
32	299	33.22	177	19.67	268	29.78	130	14.44	307	34.11	09
31	226	32.29	169	24.14	226	32.29	107	15.29	301	43.00	07
30	452	32.29	294	21.00	482	34.43	175	12.50	623	44.50	14
29	613	34.06	426	23.67	575	31.94	261	14.50	688	38.22	18
28	637	33.53	357	18.79	584	30.74	267	14.05	691	36.37	19
27	538	33.62	325	20.31	500	31.25	229	14.31	669	41.81	16
26	653	32.65	461	23.05	663	33.15	275	13.75	851	42.55	20
25	676	32.19	440	20.95	624	29.71	267	12.71	719	34.24	21
24	636	31.80	391	19.55	593	29.65	253	12.65	681	34.50	20
23	401	30.85	267	20.54	378	29.08	173	13.31	534	41.08	13
22	278	30.89	172	19.11	297	33.00	111	12.33	341	37.89	09
21	238	34.00	135	19.29	163	23.29	86	12.29	270	38.57	07
20	126	31.50	71	17.75	127	31.75	63	15.75	148	37.00	04
19	33	33.00	17	17.00	33	33.00	15	15.00	48	48.00	01
18	48	24.00	30	15.00	68	34.00	30	15.00	87	43.50	02

1. Copy N (
the firs

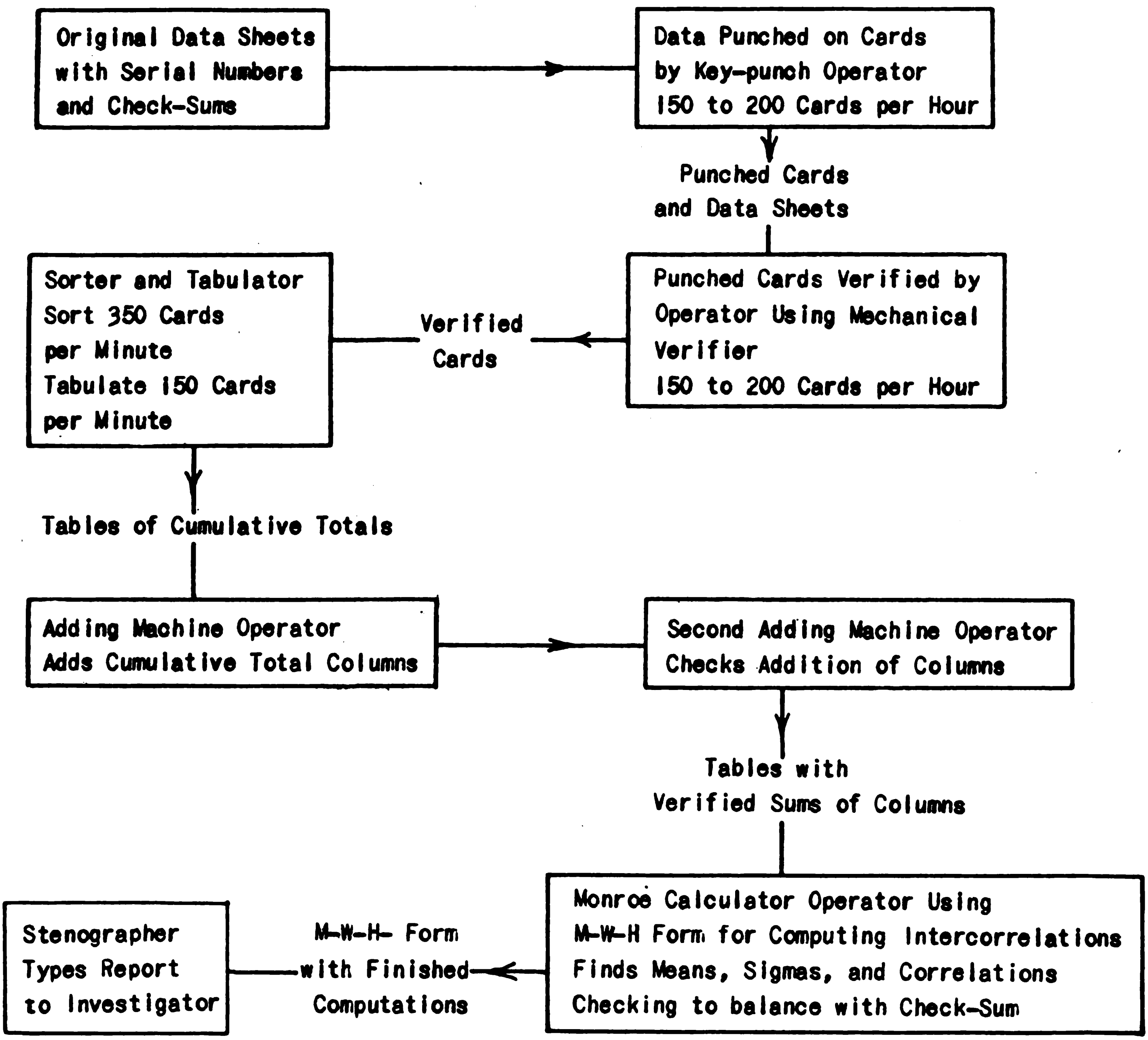
2. Copy in
the last

3.

9	10	S	Check		
		35.9811			
		164.8952	164.8952		
		34,628	34,628		
		955348	955.348	a	
		4549.2762	4549.2761	b	
				c	1
		35.8637	35.8636	d	
		150.4874	150.4874	e	
				f	
		1,143,176	1,143,176	a	
		5443.6952	5443.6952	b	
		5419.5611		c	2
		24.1341	24.1342	d	
		142.0246	142.0246	d	
				f	
		756,069	756,069	a	
		3600.3286	3600.3285	b	

DIAGRAM OF PROCEDURE - The following diagram summarizes graphically the steps in the six-variable job which we have used to illustrate the procedure of computing intercorrelations by the M-W-Hollerith method.

DIAGRAM OF M-W-H PROCEDURE
FOR COMPUTING INTERCORRELATIONS



PHOTOGRAPHS OF MACHINES

USED IN

MENDENHALL-WARREN-HOLLERITH CORRELATION METHOD

Plate I on the following page is a photograph of the first unit of machines installed for the Columbia University Statistical Bureau. It shows all of the machines mentioned in the preceding diagram of procedure.

While all these machines are a necessary convenience, it is obvious that the heart of the unit is the Hollerith Printing Tabulator, supported by the Sorter and Duplicating Key-Punch. The feature of this tabulating machine which is most essential for correlation purposes is its capacity for carrying progressive-totals, or cumulative totals, as well as group-totals or sub-totals. Other favorable features are: that it uses 80-column cards, thus permitting the computation of as many as 76 one-place variables, or 38 two-place variables, or 25 three-place variables with a single set of punched cards; that it is a 7-bank machine, having two indicating banks and five counting or adding banks, each of which may be "split" two or more ways, thus doubling or trebling the number of accumulations that may be carried at each run of the cards, according to whether the scores are two or one-place numbers; that it has a free switch-board, which can be quickly and easily wired for a great variety of operations, thus affording a degree of flexibility which will be especially valuable to those institutions that may find it convenient to use the same installation for both research and business accounting purposes; and finally, that it prints all results, thus affording a compact and convenient permanent record which greatly facilitates checking and follow-up work.

Plates 2, 3, 4, and 5 are enlarged photographs of the Key-Punch, the Verifier, the Sorter, and the Tabulator, pictured in Plate I.

Figure 10, following Plate 5, is a diagram of the switchboard on the tabulator pictured in Plate 5, wired for Table I, of which a slightly reduced facsimile is reproduced. Any good Hollerith machine operator can explain this diagram to a statistician who is not familiar with tabulating machines.

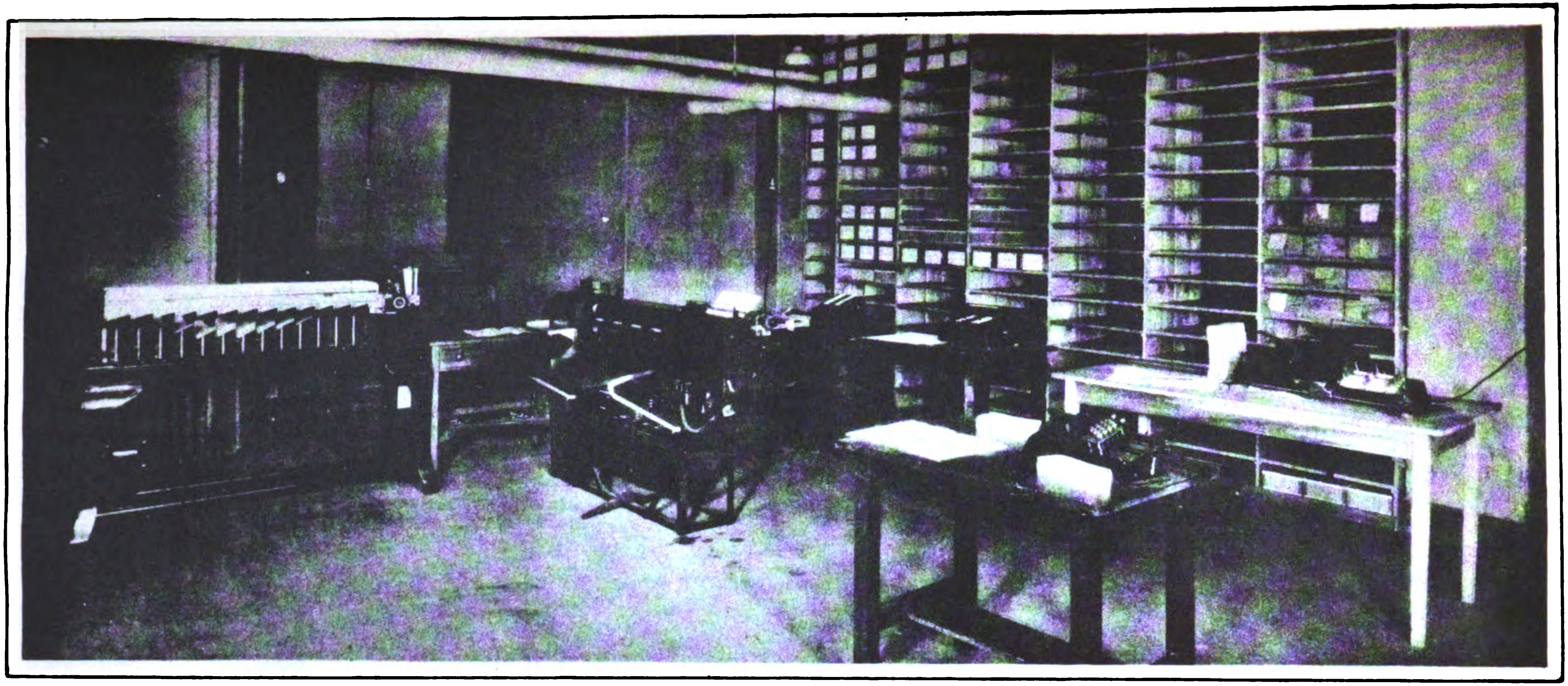


Plate I

THE FIRST UNIT OF MACHINES INSTALLED FOR THE COLUMBIA UNIVERSITY STATISTICAL BUREAU

Hollerith Sorter

Hollerith Printing Tabulator

Burroughs Calculator

Monroe Calculating Machine

Hollerith Electric Duplicating Key-punch

Hollerith Verifier

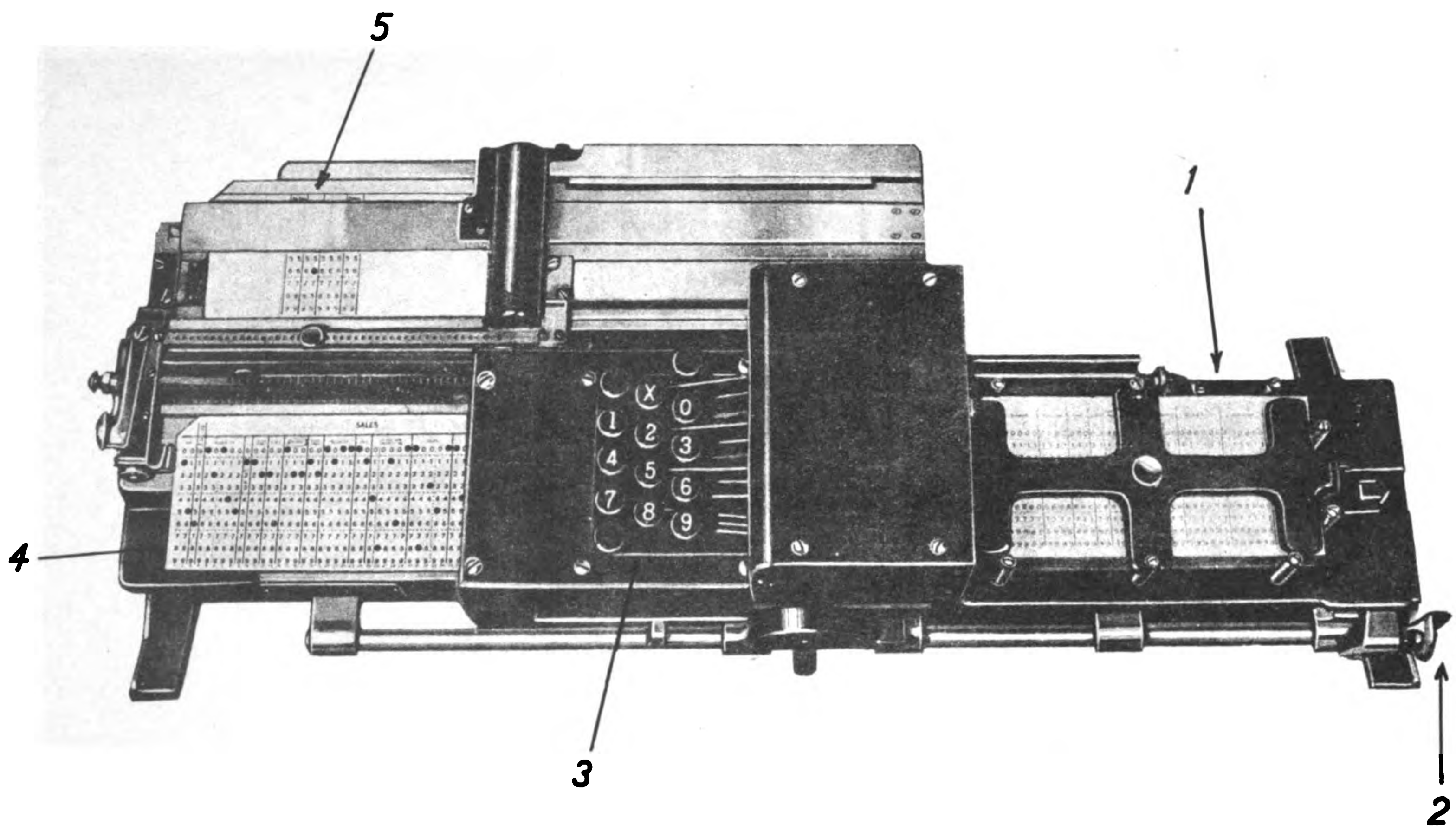


Plate 2

HOLLERITH ELECTRIC DUPLICATING KEY-PUNCH

1. Blank card magazine. 2. Card feed lever. 3. Standard Key-punch keyboard. 4. Punched card. 5. Punched card carrying items common to successive members of set being punched. This information is automatically duplicated on every card.

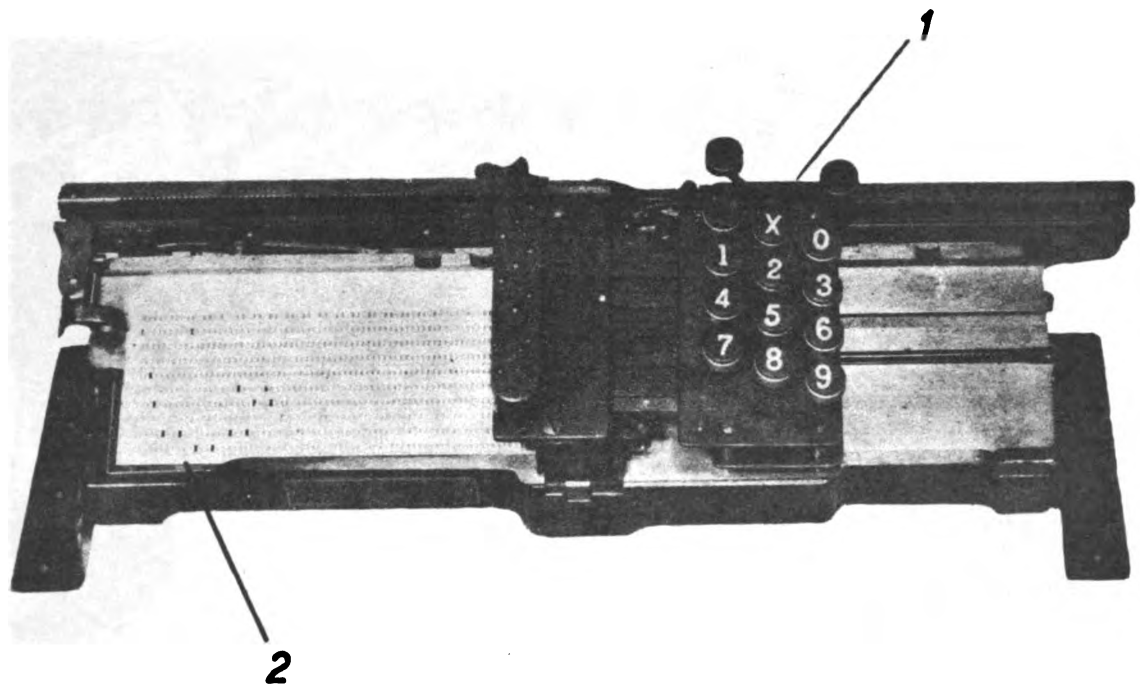


Plate 3

Hollerith Mechanical Verifier

1. Standard Key-punch Keyboard. 2. Punched Card whose Numbers are to be Compared with Data-sheet.

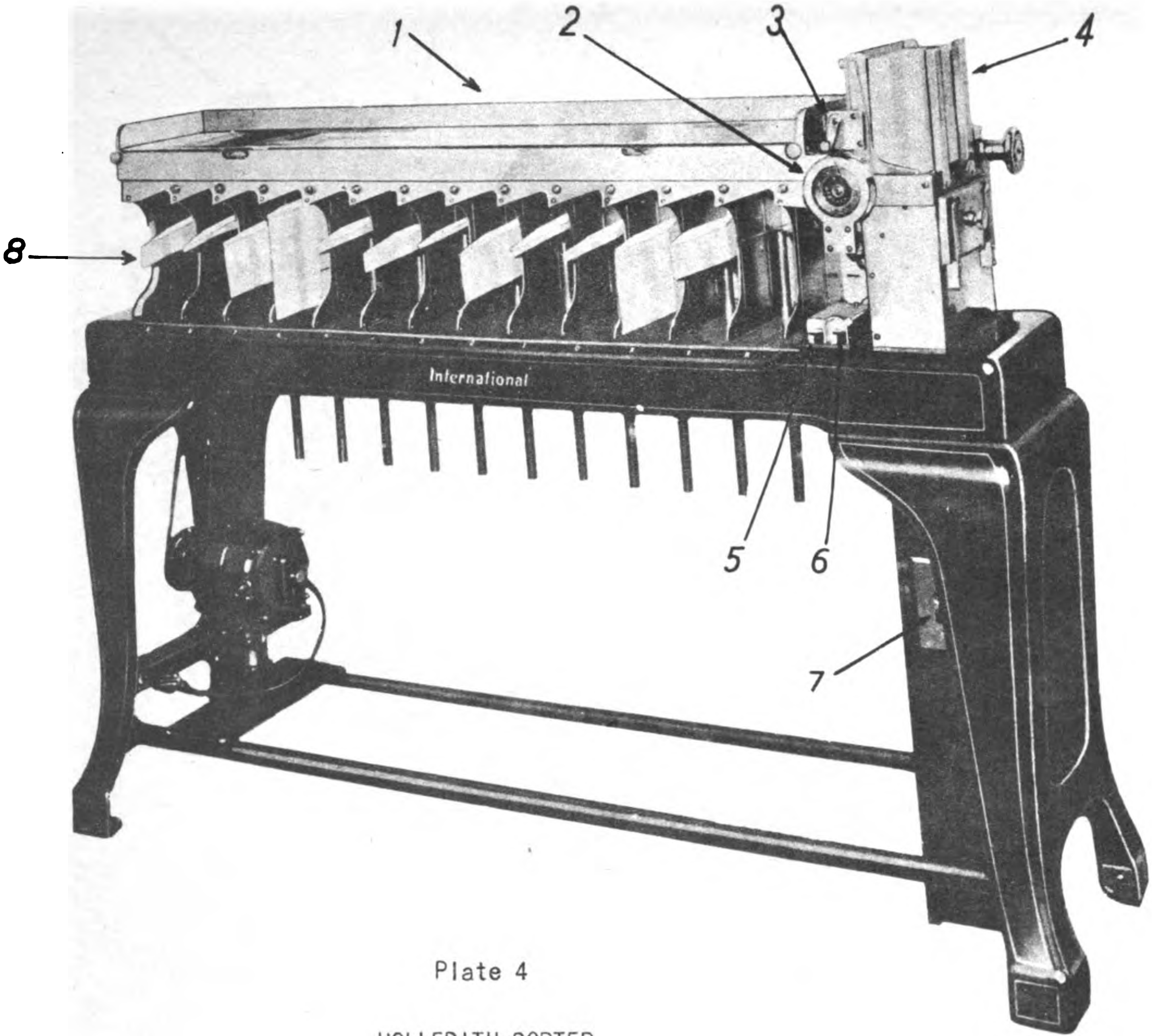


Plate 4

HOLLERITH SORTER

1. Card tray with rim for stacking. 2. Commutator cut-out for disconnecting selector mechanism of any or all boxes. 3. Brush shift-lever for selecting column to be sorted on. 4. Cards in feed-hopper. 5. Start key. 6. Stop-key. 7. Power switch. 8. Sorted cards in boxes.

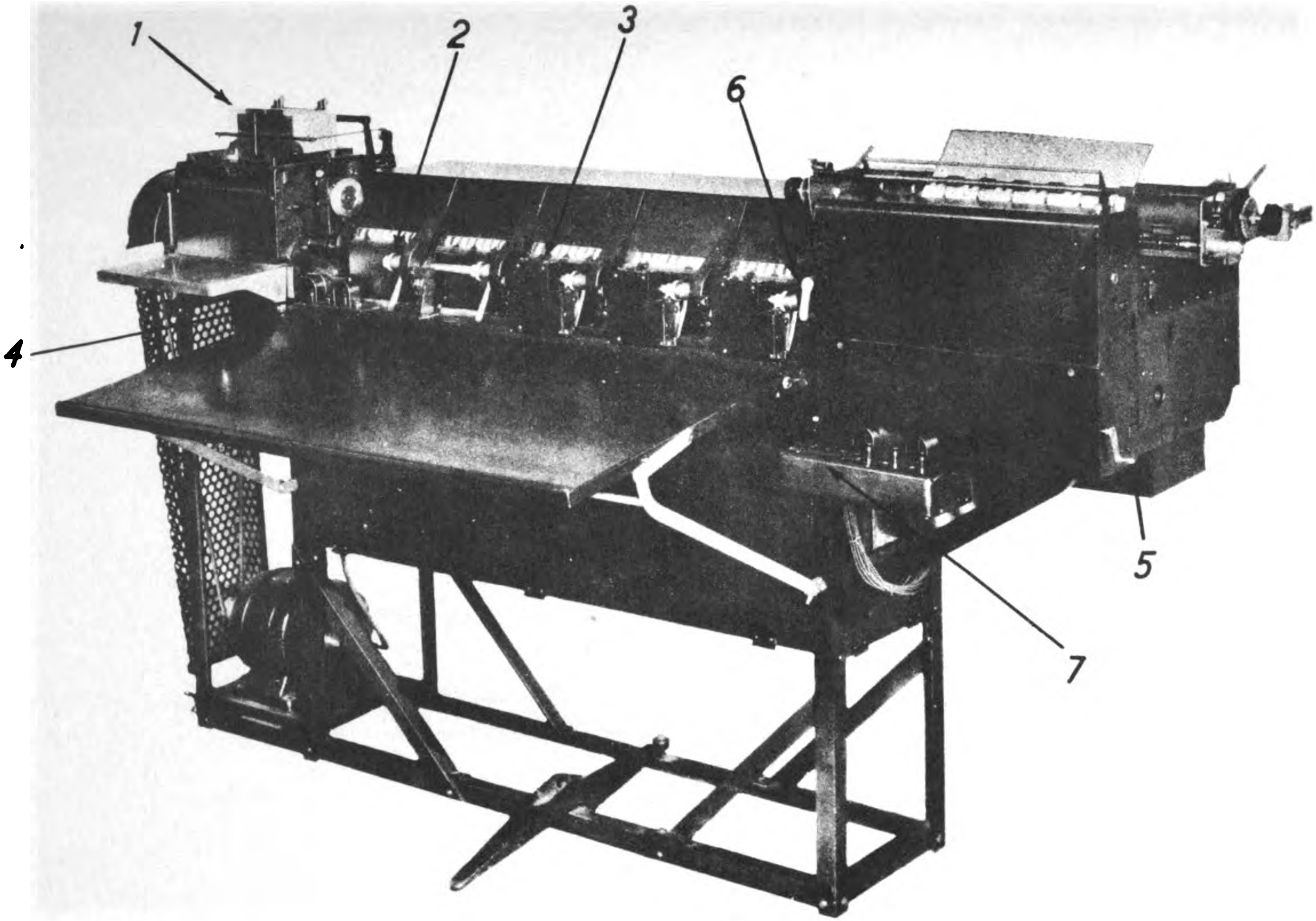


Plate 5

Hollerith Printing Tabulator

1. Card Hopper. 2. Reset Collar. 3. Magnetic Reset Clutch. 4. Start, Stop, and Reset Buttons. 5. Printing Hammer Tension Adjustment. 6. List-Tabulate Shift. 7. Progressive Total Switches. When these switches are in the forward position the Tabulator will print Cumulative Totals.