

machine at a rate exceeding one per second; the total of the squares of the numbers appearing on consecutive cards may be obtained automatically; if the variates x and y be punched in respective columns on each card, the total of the xy products for all the cards is likewise made available; the cards may be arranged in order of magnitude according to card number, date, or variates at a rate exceeding six per second, and finally the data on the cards may be printed on a scroll—the cards passing through the listing or printing machine at about 80 per minute.

In order to provide an actual problem to serve as an illustration, I secured the anthropometric records of one thousand first year male students who entered the University of Michigan in the fall of 1928. The 1,000 cards, of which Figure 1 is a sample, were punched by an operator of average ability in slightly less than three hours. The data for the students were selected at random and the cards, as punched, were numbered consecutively from 1 to 1,000,—the card selected for Figure 1 being the 275th. The weight to the nearest pound of this individual was 125 pounds, and the height, width of shoulders, and the circumferences of chest, waist, hips and right thigh were, respectively, 64.0, 16.8, 34.2, 26.0, 34.3 and 20.2,—linear measurements being made to the nearest one-tenth inch.

These 1,000 cards may now be placed in a *tabulating and listing* machine which can total all seven of the data fields simultaneously. If desired, the machine will also print all of the information of each card on a scroll together with the totals. The first and last parts of this scroll are reproduced below photographically,—the names of the individuals being omitted purposely. Because of both the number of columns involved and the magnitudes of the totals, the listed totals unfortunately run together. The vertical lines, inserted with a pen, facilitate the reading of the totals. The cards are totaled and listed simultaneously at the rate of 80 per minute.

1	184	681	165	338	285	368	198
2	181	674	164	348	265	387	190
3	188	677	177	368	306	388	188
4	134	698	166	380	265	348	196
5	138	698	168	358	287	346	195
6	143	718	168	356	278	360	190
7	166	683	171	397	385	383	228
8	138	688	167	368	280	344	194
9	148	691	176	378	300	361	208
10	118	648	187	388	240	330	180
991	128	688	158	360	253	342	192
992	149	684	160	340	284	364	211
993	138	664	168	388	281	354	190
994	149	698	168	355	285	370	212
995	128	630	176	365	280	345	205
996	133	695	167	387	267	353	191
997	130	667	168	330	278	345	200
998	148	718	166	343	275	370	201
999	141	680	174	370	300	360	218
1000	138	686	174	345	280	365	206

139200|678096|165337|385241|281510|355165|201096

Fig. 2

An investigation of the correlation that may exist between height and weight will involve the numerical value of

$$\sum_{i=1}^{1000} x_i y_i$$

where x_i and y_i designate the height and weight, respectively, of the i^{th} individual. The plugboard of an *Automatic Multiplying Punch* may be wired in a few seconds so that

- (a) the data of columns 34, 35 and 36 of Figure 1 will feed into the multiplier of the punch,
- (b) the data of columns 38, 39 and 40 will feed into the multiplicand, and then
- (c) the product, $x y$, for any card run through the machine will appear on the *product summary counter*. As the cards pass through the machine, the total of the products is accumulated on this counter.

If desired, each product may be punched automatically in the card, provided of course the card contains a sufficiently large number of otherwise vacant columns. The maximum number of digits in current models that may occur in either multiplier or multiplicand is eight. The number of digits in the multiplicand does not affect the speed of the multiplication; for three or less digits in the multiplier the cards feed through the machine at the rate of three seconds per card,—for eight digits in the mul-

tiplier the speed is twelve cards per minute. One may therefore place our cards in the machine, press a button, resume other duties, and some fifty minutes later the 1,000 cards will have yielded the total

$$\sum xy = 9\ 477\ 433.6$$

To obtain the sum of the squares of the variates in question it is necessary only to double-wire the machine,—one wire going to the multiplier and the other to the multiplicand. We obtain then

$$\sum x^2 = 4\ 615\ 312.12 \quad \sum y^2 = 19\ 692\ 452$$

By permitting the machine to punch each value of x in the card, we may treat x^2 as the multiplicand and x as the multiplier and then obtain the sum of the cubes of the variates; or by double-wiring x^2 obtain the sum of the fourth powers of the variates. If, while accumulating the cubes of the variates, we let the machine also punch each cube in the card, we may then obtain the sum of the powers of the variates up to and including that of the sixth order, etc. We are limited, of course, by the fact that the card contains eighty columns.

By running the punched cards through a sorting machine, we may obtain very readily the frequency distribution of the weights, and also the corresponding median, quartiles, etc. To accomplish this the cards must be run through a *sorting machine* three times, first sorting to column 35 of Figure 1, then to column 34 and finally to column 33. The cards pass through the sorting machine at the rate of 400 per minute, so that in approximately eight minutes—including time spent in handling the cards between sorts—these 1,000 cards will be perfectly arranged according to magnitude in weight. If the numbers with respect to which the sort is to be made contain n digits, the cards must be run through the sorter n times. We reproduce on the following page a photograph of the first part of a printed scroll obtained by running the cards through the listing machine after they had been sorted according to weight.

Figure 3.

Card #	Weight	Height	Shoulder	Chest	Waist	Hip	R. Thigh
232	89	597	140	295	253	305	171
146	100	607	157	300	242	312	170
691	101	690	150	301	245	308	165
358	102	676	154	327	262	322	174
555	102	662	153	303	242	307	165
941	102	640	154	313	262	313	168
209	103	663	154	310	285	313	172
801	103	649	150	302	247	302	168
14	104	638	147	313	253	300	178
513	105	635	148	310	250	308	187
720	105	637	155	330	258	319	178
563	106	648	152	302	245	320	169
672	106	638	162	323	254	320	181
153	107	623	143	317	250	322	172
75	108	630	160	335	245	332	183
235	108	669	157	320	247	324	172
322	108	624	165	305	260	325	188
505	108	637	152	330	246	330	183
31	109	620	161	334	265	340	192
393	109	625	152	314	233	336	187
30	110	631	162	345	265	327	180
160	110	667	153	320	247	322	175
185	110	627	152	323	274	320	172
631	110	630	153	335	255	330	180
802	110	623	149	317	257	328	181
15	111	650	158	312	265	344	194
151	111	637	157	322	258	325	172
273	111	651	156	332	270	326	180
447	111	655	160	314	252	324	176
20	112	626	155	300	230	335	174
426	112	647	154	353	262	328	170
507	112	691	161	323	262	322	180
716	112	667	148	330	254	322	185
831	112	651	147	373	262	333	182
308	113	661	155	318	250	330	180
383	113	665	154	338	240	333	178
449	113	651	151	328	258	311	163
541	113	675	152	316	257	340	186
591	113	650	166	341	265	333	180
826	113	654	147	331	260	333	183
898	113	677	156	320	253	335	187
933	113	656	153	315	240	326	179
947	113	637	146	315	262	337	188
967	113	696	151	320	258	330	169
148	114	654	153	320	247	322	183
177	114	633	160	342	255	321	172
257	114	653	162	318	270	320	185
368	114	679	145	318	243	322	173
462	114	653	153	322	261	321	171
464	114	627	150	310	262	332	180
545	114	612	157	334	260	318	181
741	114	662	161	338	262	332	186
951	114	681	153	311	242	333	174
987	114	667	152	342	250	330	188
10	115	645	157	328	248	330	180
139	115	665	146	332	265	339	188
159	115	636	160	328	297	318	174
336	115	663	153	321	252	323	167
743	115	639	158	322	278	338	182
949	115	686	154	330	242	330	182
970	115	653	152	313	250	337	172
988	115	692	155	330	252	352	172
52	116	640	157	316	260	337	193
186	116	606	159	310	258	337	186
226	116	665	166	341	262	335	185
268	116	636	162	320	265	350	192
347	116	630	163	321	265	325	195
510	116	662	162	337	275	330	190
523	116	695	158	325	243	340	182
926	116	641	154	338	257	340	182

A rough notion of the functional dependence that exists between weight and the other six variables recorded on the cards may be obtained by permitting the machine to total these ordered-with-respect-to-weight cards in consecutive groups of 100. That is, we obtain the averages for numerically equal groups selected according to the weight-deciles. The six regression lines may therefore be plotted, approximately, from the following results:

TABLE 1.
ANTHROPOMETRIC AVERAGES BASED ON WEIGHT DECILES.

<i>Inter-decile Range</i>	<i>Weight</i>	<i>Height</i>	<i>Sh'der</i>	<i>Chest</i>	<i>Waist</i>	<i>Hips</i>	<i>Rt.Th.</i>
First	112.98	65.133	15.576	32.607	25.748	32.968	18.133
Second	122.41	66.659	15.980	33.663	26.777	33.927	18.926
Third	127.85	67.087	16.161	34.421	26.978	34.387	19.206
Fourth	131.98	67.381	16.334	34.816	27.622	34.858	19.554
Fifth	135.62	67.937	16.406	34.860	27.954	35.081	19.893
Sixth	139.54	68.189	16.651	35.608	28.065	35.511	20.112
Seventh	143.87	68.576	16.789	35.766	28.513	36.006	20.438
Eighth	149.43	68.895	16.807	36.116	28.780	36.420	20.712
Ninth	156.01	69.185	17.022	36.788	29.537	37.181	21.444
Tenth	173.19	69.854	17.611	38.596	31.536	38.826	22.678

If we had arranged the cards numerically with respect to height, instead of weight, we would have obtained the following results:

TABLE 2.
ANTHROPOMETRIC AVERAGES BASED ON HEIGHT DECILES

<i>Inter-decile Range</i>	<i>Weight</i>	<i>Height</i>	<i>Sh'der</i>	<i>Chest.</i>	<i>Waist</i>	<i>Hips</i>	<i>Rt.Th.</i>
First	123.95	63.339	16.036	34.201	27.371	34.217	19.592
Second	130.97	65.295	16.261	34.856	27.984	34.944	19.929
Third	133.75	66.367	16.282	34.787	27.731	35.152	19.926
Fourth	136.28	67.021	16.570	35.429	28.329	35.302	20.084
Fifth	139.81	67.623	16.587	35.379	28.206	35.499	20.223
Sixth	140.60	68.189	16.532	35.510	28.184	35.560	20.008
Seventh	142.65	68.806	16.659	35.550	28.380	35.821	20.315
Eighth	143.44	69.498	16.638	35.328	28.065	35.858	20.205
Ninth	145.71	70.412	16.776	35.778	28.236	35.960	20.068
Tenth	155.72	72.346	16.996	36.423	29.024	36.852	20.746

A comparison of tables 1 and 2 reveals clearly that the right thigh measurements are more highly correlated with weight than with height and this phenomena appears also to be true—though possibly less pronounced—for the shoulder, chest, waist and hip measurements.

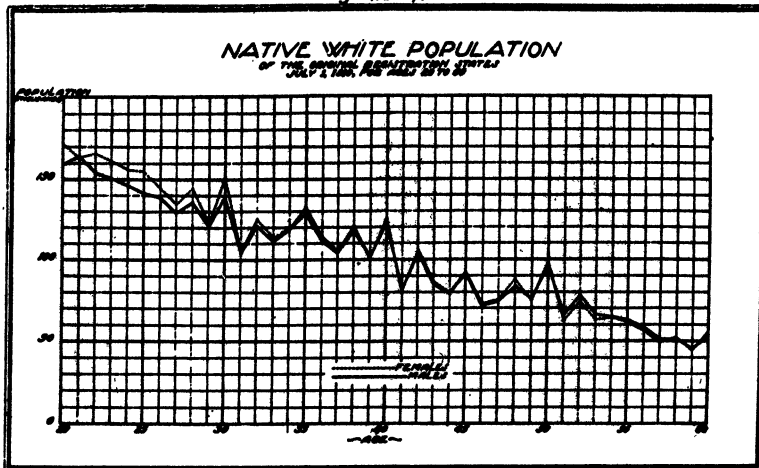
By sorting the cards according to the last recorded digit of each weight, the tendency of observers to state results as multiples of two and five is apparent. The results presented in table

TABLE 3.
RELATIVE FREQUENCY OF FINAL
DIGITS IN OBSERVED VARIATES.

<i>Final Digit</i>	<i>Frequency</i>
0	142
1	69
2	117
3	84
4	108
5	112
6	112
7	69
8	117
9	70
Total	1000

3 indicate that final digits of 1, 3, 7 and 9 are decidedly unpopular and this casts a reflection upon the accuracy of the recorded measurements. This type of bias is well known—in fact, census and mortality statistics usually present the same phenomena. Figure 4, which follows, clearly illustrates this fact.

Figure 4.



SUMMARY

The punching, sorting and listing machines provide a most economical and accurate method of recording and analyzing, non-mathematically, observational data. The punched-card system is especially effective in constructing frequency distributions and correlation tables when the data are very numerous.

The recent development of the Automatic Multiplying Punch is unquestionably the most important contribution to the mechanics of mathematical statistics since the invention of adding and multiplying machines. By enabling one to compute moments and product-moments exactly and without the grouping of variates about class-marks, corrections such as those due to Sheppard are unnecessary for practical computations. Indeed, it is even possible to evaluate linear functions of one or more variables with this machine, and subsequently print the graduated values on a scroll—these results being “rounded off” to any desired number of decimal places.

In this article I have described only a few of the various machines employed in statistical and accounting practice. Readers may secure additional information from the International Business Machines Corporation, Tabulating Machine Division, 270 Broadway, New York City.

B. A. C. 6-17-34.