4

The Mean and Variability as Affected by Continuous Selection for Composition in Corn

F. L. WINTER

INTRODUCTION

The effectiveness of selection in cross-fertilized crops such as maize may be explained upon the following principles:

1) There are heritable variations existing among the different individuals of the population at the beginning.

2) There is a gradual elimination of the undesirable type because the selected types consist of more desirable and fewer undesirable individuals in each generation.

3) Desirable mutations which may occur are retained and caused to combine with the desirable factors present.

4) Recombination of the desirable factors produces more desirable types.

Selection for a given type not only tends to bring the population to that type but is expected to decrease the variability. This decrease in variability of the population is brought about by a reduction in the percentage of heterozygous individuals. After the population becomes homozygous for the selected character, no further reduction in variability through selection can be expected. The variability that still remains is attributed to environment, upon which selection has no influence.

It is the purpose of this paper to present the effect of 28 years of continuous selection for composition in maize upon the mean and the variability of the selected character.

MATERIAL

In 1896 a series of breeding experiments was begun at the Illinois Agricultural Experiment Station to determine whether the chemical composition of corn could be influenced by selection (8). One hundred and sixty-three ears of a variety known as Burr's White were used as the foundation stock, from which selections were made in four different directions, namely, for high oil, low oil, high protein, and low protein.

These four strains were carried on in the same way. In the high protein, for example, the 24 ears highest in protein were selected for seed and planted in an isolated plot, each ear in a separate row. These rows were harvested separately and the seed for the next crop selected from the ears which were found to be highest in protein. Nine years later the system was modified somewhat in an attempt to prevent loss of vigor by inbreeding. Alternate rows were detasseled and seed was selected only from the highest yielding detasseled rows. In 1921 this system was again modified to reduce the amount of inbreeding. Two seed ears were taken from each of the detasseled rows regardless of yield. The high-oil, low-oil, and low-protein tests were similarly conducted, selection being made each year of ears highest in oil, lowest in oil, and lowest in protein, respectively.

The analytical methods employed have been described in detail in Illinois Agricultural Experiment Station Bulletins 43 and 53 (6, 7).

EFFECT OF SELECTION FOR PROTEIN CONTENT

The effect of selection for protein content in the corn grain is shown in Fig. 1. The average protein content which was 10.92 in 1896 had in 1924 been increased to 16.60% in the high-protein strain and had been decreased to 8.38% in low-protein strain. This is a difference of 8.22%. When measured by the best fitting straight lines, the difference is 9.49%. When compared with the original variety from which the strains were selected, the proportion increase is 50.01% and the proportional decrease is 23.26%.

EFFECT OF SELECTION FOR OIL CONTENT

From an average of 4.70% in 1896, the oil content had been increased to 9.86 in the high-oil strain and decreased to 1.51% in the low-oil strain in 1924. This is a difference of 8.35%. When measured by the best fitting straight lines, the difference is 8.85%. Relatively, selection has been much more effective in bringing about a change in oil content than in protein content. When compared with the original variety, the proportional increase in oil is 109.79% and the proportional decrease is 67.87%. (Fig. 2.)

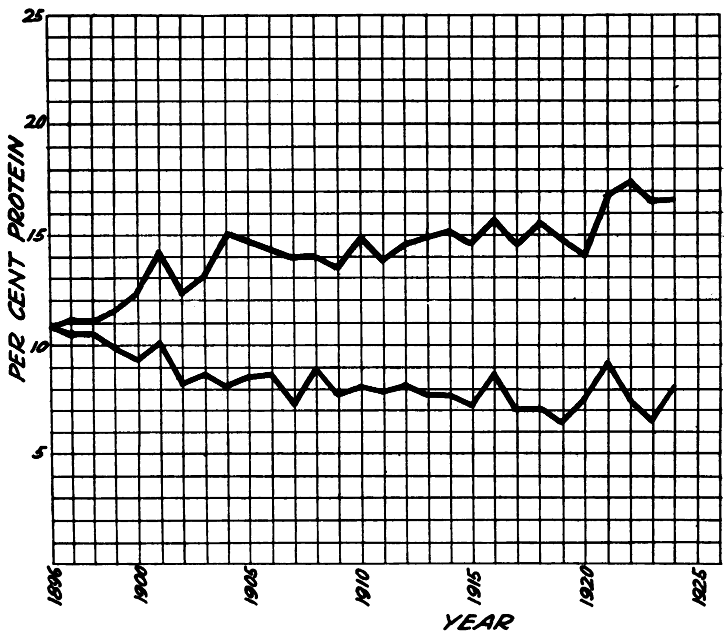

Figure 1. Progress of high-protein and low-protein corn breeding.

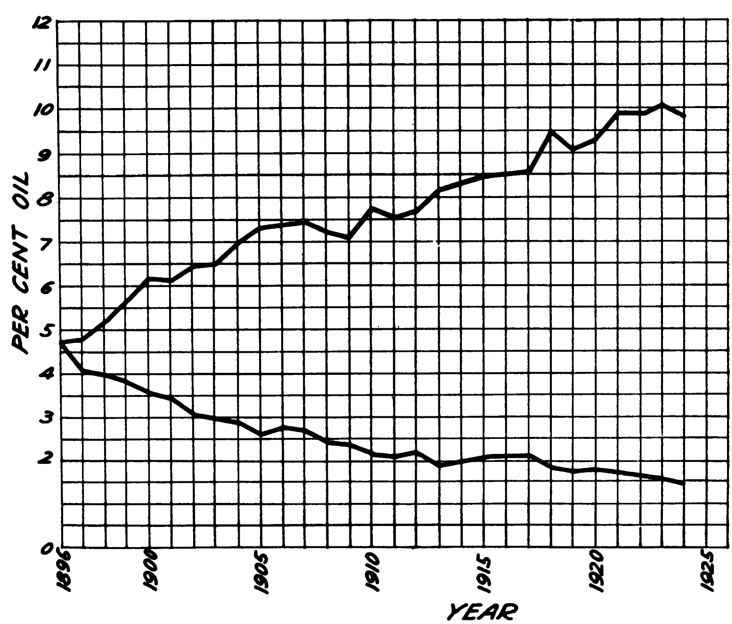

Figure 2. Progress of high-oil and low-oil corn breeding.

It is very probable that a more rapid shift in type with respect to both the oil and protein would have been obtained if only the ears that were highest and lowest in protein or in oil had been selected for seed, regardless of the yielding ability of the detasseled rows. On the other hand, if selection had been carried on without taking yield into account, inbreeding might have resulted in such a loss of vigor that it would have been impossible to continue the strains. When it is remembered that the original purpose of the experiment was to produce good-yielding strains having the desired composition, any great reduction in yield would have defeated the purpose.

PEDIGREES

Graphic presentations of the pedigrees of the strains of corn used in this work are on pages 99 and 100.

A study of the pedigrees of the four strains shows that there has been a rapid elimination of lines until the four strains are now represented by only a single ear each of the original seed stock. The high-protein strain traces back to ear 121 which on analysis showed 12.28% protein and 3.99% oil. This is slightly below the average composition of 12.54% for the stock seed ears but considerably above the average of the original 163 ears, 10.92%. The low-protein strain goes back to ear 106, which on analysis yielded 8.25% protein and 4.81% oil. The average for the stock seed ears of the low-protein strain was 8.96%. The high-oil strain goes back to ear 111, which on analysis showed 5.65% oil and 10.82% protein. The average for the stock seed ears for the high oil was 5.33% as compared to 4.70%, the average for the original 163 ears. The low-oil strain traces back to ear 110, which when analyzed was found to contain 4.10% oil and 11.13% protein. The average percentage of oil for the stock seed ears was 4.04. Thus, in the 28th year of selection all of the 96 ears of the 4 strains trace back to 4 ears of the original Burr's White.

Even though the ear has been used as the unit of selection and the history of the strains traced through the female side only, the reduction of the ancestry to a single ear for each strain indicates that the strains at the present date are likely to be more nearly homozygous than was the original material from which they came. Further evidence that the strains are more homozygous than open-pollinated varieties is furnished by selfing the strains. It has been shown at this station that upon selfing, a condition of uniformity is reached more quickly than with open-pollinated varieties. Although the individual inbred lines coming from any one of the strains differ among themselves in composition, they are always significantly different from any of the inbreds coming from the other strains.

That there is still some heterozygosity left in the individual within the strains is indicated by the fact that there is a reduction in vigor upon selfing.

East (2), in fitting curves to the data for the first 10 generations of selection, found that at first the curve was concave, showing great progress,

PEDIGREE OF ILLINOIS LOW-PROTEIN CORN

GENERATION NUMBER

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
																											01
																									02	02	02
																											03
																									04	04	04
																											05
																									06	06	06
																											07
																								02	08	08	08
																						04	06				09
																					08			08	10	10	10
													06	06	14	06	04	06	04	02		10	12				11
												10									18			12	12	12	12
*06	10	14	03	06	02	15	09	20	10	04	08		18	10	18	02	08	08	12	08		14	14				13
												22									20			14	14	14	14
													24	14	22	14	22	12	14	24		18	16				15
																					22			16	16	16	16
																						22	20				17
																								18	18	18	18
																											19
																									20	20	20
																											21
																									22	22	22
																											23
																									24	24	24

PEDIGREE OF ILLINOIS HIGH-PROTEIN CORN

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
																											01
																									02	02	02
																											03
																									04	04	04
																											05
																									06	06	06
																											07
																								06	08	08	08
																											09
																							04	08	10	10	10
																	08	16	02	12	04	08					11
												06	10	10	14	06							06	10	12	12	12
*21	07	23	13	07	07	17	18	22	22	12	24						20	20	06	18	06	14					13
												18	14	14	24	12							10	12	14	14	14
																	22	23	16	20	10	22					15
																							22	20	16	16	16
																											17
																								24	18	18	18
																											19
																									20	20	20
																											21
																									22	22	22
																											23
																									24	24	24

*These numbers, preceded by generation numbers, denote the seed ear numbers.

PEDIGREE OF ILLINOIS HIGH-OIL CORN

GENERATION NUMBER

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
																											01
																									02	02	02
																											03
																									04	04	04
																											05
																									06	06	06
																											07
																								06	08	08	08
																											09
																					06	12	02	08	10	10	10
																				04							11
														04	10	04	20	04	20		12	16	04	10	12	12	12
*11	10	03	07	08	09	09	14	16	08	14	02	08	10							06							13
														16	16	10	24	06	22		18	18	06	12	14	14	14
																				16							15
																					22	22	24	20	16	16	16
																											17
																								24	18	18	18
																											19
																									20	20	20
																											21
																									22	22	22
																											23
																									24	24	24

PEDIGREE OF ILLINOIS LOW-OIL CORN

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
																											01
																									02	02	02
																											03
																									04	04	04
																											05
																									06	06	06
																											07
																								06	08	08	08
																											09
																							10	08	10	10	10
																06	08	02	02	02	02	04					11
									06	04	08	04	10	06	04								12	10	12	12	12
*10	06	10	07	04	05	13	07	10								14	12	10	08	06	04	08					13
									18	22	18	10	20	12	20								14	12	14	14	14
																18	16	12	12	20	20	10					15
																							24	18	16	16	16
																											17
																								22	18	18	18
																											19
																									20	20	20
																											21
																									22	22	22
																											23
																									24	24	24

*These numbers, preceded by generation numbers, denote the seed ear numbers.

later convex, showing that progress became slower, and at last horizontal, indicating that no more progress would result from selection. An inspection of Fig. 1 and 2 presented here will show that, contrary to East's results, there has been considerable progress since the 10th generation for all of the strains with the possible exception of the Illinois low protein. The Illinois high-oil strain has shown greater progress since the 10th generation than it did before.

It does not seem possible to predict a limit to the progress that selection will make in the Illinois high-oil and the Illinois high-protein strains by the application of curves to the data at hand. Apparently the low-protein strain has made but little change in the last 20 years. If the average protein content for the years beginning with 1902 be compared with the protein content of ear 106, to which the strain now traces, no significant difference appears.

The low-oil strain is approaching a physiological limit. The greater percentage of the oil in a grain of corn is contained in the germ (9). Hence, in selection for low-oil content the size of the germ has been decreased both absolutely and relatively in comparison with the size of the endosperm. In the ear containing the lowest percentage of oil on record, namely, 0.69%, 80% of the grains were germless. The necessity of using ears having grains that will germinate naturally tends, therefore, to check the progress of selection, and eventually may stop it altogether.

EFFECT OF SELECTION UPON VARIABILITY

It seems to be accepted by most biologists that selection for a given character in a cross-fertilized crop like corn leads to a lower variability. That this reduction in variability may have its limits is brought out by Davenport (1) when he states that selection in a cross-fertilized crop such as corn simply shifts the type but does not appreciably change the variability unless a physiological limit is reached. As evidence for this belief, he cites Karl Pearson as stating that 10 to 13% reduction in variability obtained by the selection of two parents is almost the limit that can be reached, even if the complete ancestry had been selected. Davenport also uses the first 8 years' data on the Illinois "chemical" strains of corn, employing the coefficient of variation as a measure of variability, and states that there is no significant change in variability. It is of interest to see how the matter stands after the results of 20 more years of continued selection have been secured.

Variability Measured by the Coefficient of Variation

Protein

The effect of selection upon the variability in protein content as measured by the coefficient of variation from the beginning of the experiment through the 28 years of selection can be seen in Fig. 3. Although the yearly variations as measured by this coefficient are rather large, there are decided trends that run in directions opposite to the line of selection. As selection

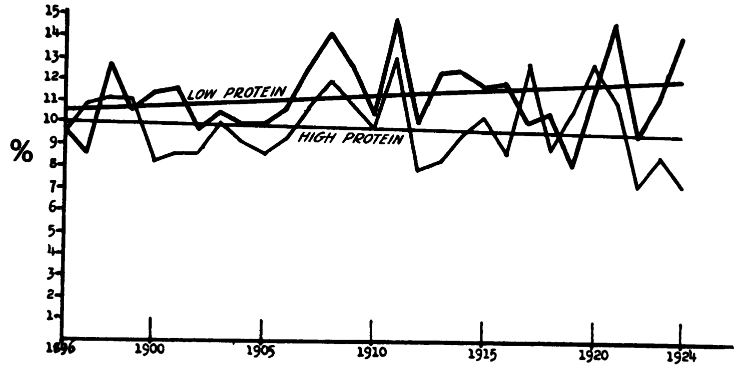

Figure 3. Yearly variations in protein content of corn and their straight-line trends as measured by the coefficient of variation.

leads to a low mean, the variability increases, and vice versa. As measured by the best fitting lines, the variability has increased 14.05% in the low-protein strain over that at the beginning and decreased 4.83% in the high-protein strain.

Oil

Similar results have been obtained for the oil content; only here the divergence is more marked. The variability, according to the coefficient of variability, has increased 69.84% for the low-oil strain and decreased 23.21% for the high-oil strain. (Fig. 4.)

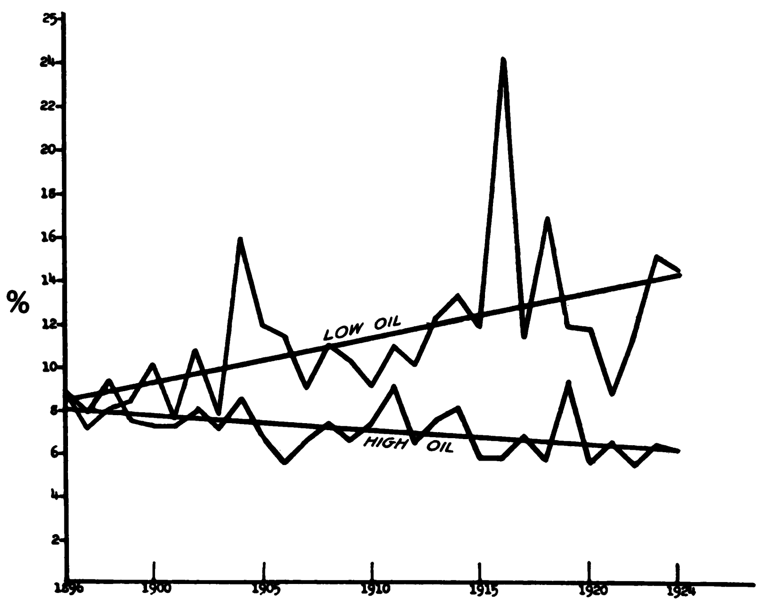

Figure 4. Yearly variations in the oil content of corn and their straight-line trends as measured by the coefficient of variation.

It appears from these results that selection can change variability, as measured by the coefficient of variation, more than 13% and that the variability may either increase or decrease, depending upon the magnitude of the mean. Although the coefficient of variation has been the most commonly used means of comparing variability when different types of variation are involved in the comparison, it is recognized to have decided limitations. Pearl (12) states that, ". . . the coefficient of variation has never been an entirely satisfactory constant to biologists, at least," and also that (11, p. 275) "one should always remember that this constant simply measures the degree of scatter of the distribution in relation to the mean value of the thing varying." Such a relation may have a real and significant meaning but sometimes it does not have, for reasons inherent in the nature of the facts themselves.

Variability Measured by the Standard Deviation

The standard deviation has usually been accepted as the standard method of measuring in absolute terms the degree of variability. It has the advantage that it is a constant of the mathematical formula for the curve of variation representing the distribution of a population.

Protein

The variability for protein content in the Illinois high-protein and Illinois low-protein strains, as measured by the standard deviation (Fig. 5), shows trends opposite to those obtained when the coefficient of variation is used to measure the variation. Although showing considerable yearly fluctu-

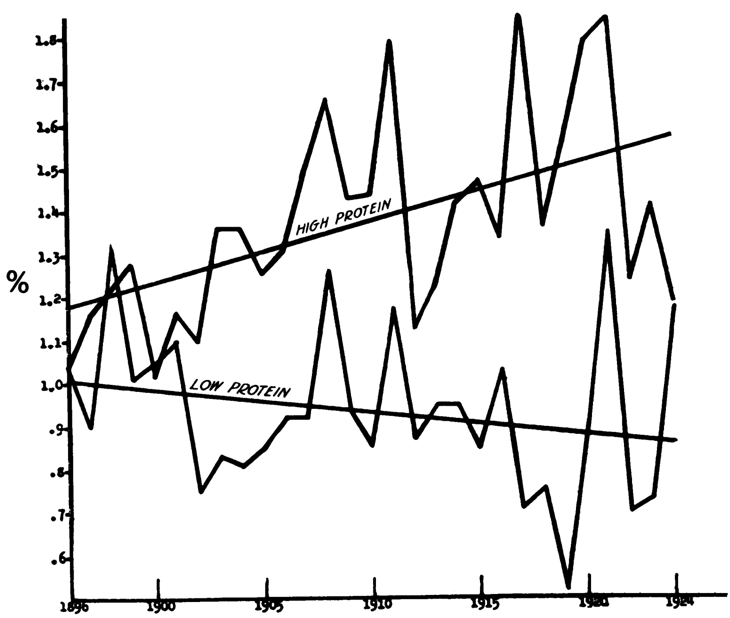

Figure 5. Yearly variations in protein content of corn and their straight-line trends as measured by the standard deviation, 1896-1924.

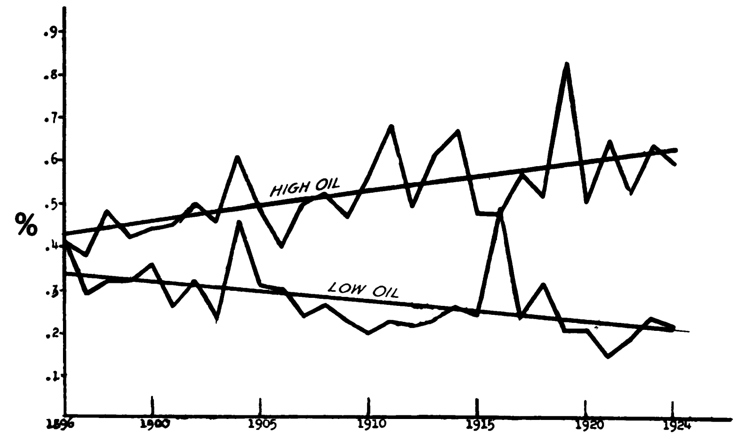

Figure 6. Yearly variations in oil content of corn and their straight-line trends as measured by the standard deviation, 1896-1924.

ation, the variability tends to move in the same direction as the mean; e. g., the high-protein strain is more variable than the low-protein strain for protein content. During the course of the experiments an increase in variability, according to this index, of 32.97% for the high-protein strain and a decrease of 14.71% for the low-protein strain have been obtained.

Oil

The variability measured by the standard deviation for the oil content shows similar results (Fig. 6). Here also the variability is greater in the strain having the higher content. An increase in variability of 50.95% in the high-oil strain and a decrease of 36.44% in the low-oil have been obtained. Again, the greater divergence in variability occurs between the oil strains.

When the standard deviation is used as a measure of variability, greater changes are exhibited than when the coefficient of variation is employed except for the low-oil strain. However, the standard deviation has its limitations also. It can not be used to compare variability except where like things are measured in like units.

Variability Measured by the Weinberg Formula

Weinberg (14) has recently proposed a method for measuring variation which is free from most of the limitations of the coefficient of variation and standard deviation. The coefficient W =

$$\frac{\sigma \ (M_n - M_o)^{1/2}}{[(M_a - M_o)\ (M_n - M_a)]^{1/2}}$$

when M_n is the highest value in the distribution, M_o is the lowest value in the distribution, M_a is the mean value of all variants.

Table 1. Comparison of coefficient of variation with Weinberg's formula for measuring variation.

Example 1		Example 2	
Class	f	Class	f
1	1	11	1
2	1	12	1
3	1	13	1
4	1	14	1
5	1	15	1
6	1	16	1
7	1	17	1
8	1	18	1
9	1	19	1
Mean = 5 S. D. = 2.582 C. V. = 51.64% W. = 1.83		Mean = 15 S. D. = 2.582 C. V. = 17.21% W. = 1.83	

The denominator of the formula measures roughly the skewness of the variability curve, while the second term in the numerator measures the range. The greater the skewness or the wider the range, or both, the greater the variability. The coefficient thus obtained is not affected by the magnitude of the mean as is the coefficient of variation, nor is it limited by unlike material or unlike units of measure as is the standard deviation. Like the coefficient of variation, however, it expresses variation as an abstract figure.

In Table 1 Weinberg's method is compared with the coefficient of variation for measuring variation by means of two examples. In example 1 the mean is 5, the standard deviation, 2.582, and the coefficient of variation is 51.64%. By the Weinberg method the variation is 1.83. In example 2 each class is 10 units higher. The mean is 15, the standard deviation is the same as in example 1, but the coefficient of variation is 17.21%. Variation, as expressed by Weinberg's formula, is the same as in example 1. It is believed that for comparative purposes the Weinberg formula gives a better conception of variability than does the coefficient of variation.

Protein

The effect of selection upon variability for protein content, as measured by the Weinberg formula (Fig. 7), is similar to the results obtained when the standard deviation is used in that variability increases as selection leads to a high mean and decreases as selection leads to a low mean. The increase in variability for the high-protein strain was 25.49%. However, the decrease in variability for the low-protein strain was only 3.94%.

Oil

Similar results were obtained for variability in oil content of the oil strains. (Fig. 8.) The increase in the high-oil strain was 23.97% and the decrease in the low-oil strain was 30.18%. Again, the divergence in variability is greater between the oil strains than between the protein strains.

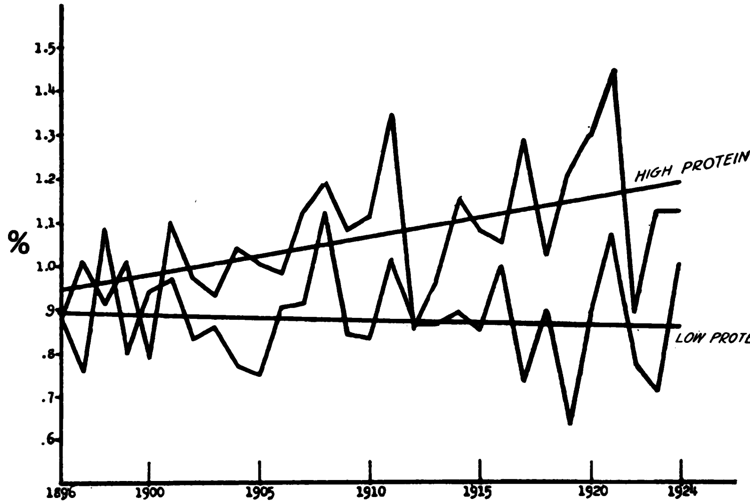

Figure 7. Yearly variations in protein content of corn and their straight-line trends as measured by Weinberg's formula, 1896-1924.

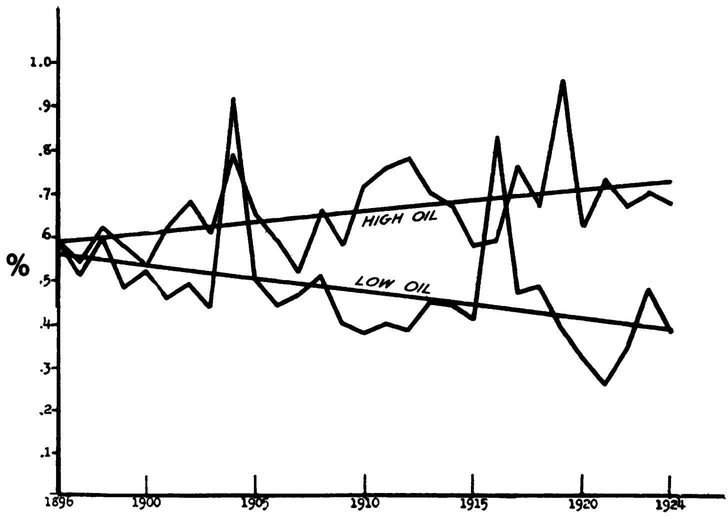

Figure 8. Yearly variations in oil content of corn and their straight-line trends as measured by Weinberg's formula, 1896-1924.

Variability Measured by the Modal Class

Since the type of the population is represented by the modal class, it is of interest to know what proportion of the population resides within the modal class as selection continues.

A graphical representation of the modal classes in respect to composition for the four different strains taken at periodic intervals is seen in Fig. 9.

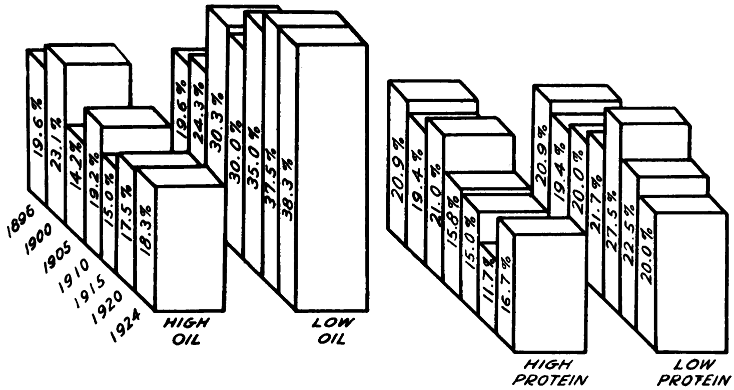

Figure 9. Graphic presentation of the modal classes with respect to the percentage composition of Illinois high and low oil, and high and low protein corn, at periodic intervals between 1896 and 1924.

The blocks represent the relative proportion of the population residing in the modal class for the respective years. Although there is considerable fluctuation, it can be seen that the trend is downward for the high-oil and the high-protein strains, i. e., as selection leads toward a higher mean fewer and fewer individuals of the population reside within the modal class. In the low-oil strain there is a decided trend upward. The trend in the low-protein strain is slightly upward. Hence, selection for a low mean increases the percentage of individuals lying in the modal class.

The percentage of the population in the modal class may be taken as a rough measure of the uniformity, and, correspondingly, the percentage outside of the modal class as an expression of variability. In general, the greater the percentage of the population lying in the modal class the less the variability, and vice versa. This variability may be expressed in the form of a ratio $\frac{100\text{-}Y}{Y}$ where Y is the percentage of the population lying in the modal class. The smaller Y is the greater the ratio; hence, the greater the variability.

Variability may also be expressed by the inverted straight-line trend for the modal classes. For the purposes of this work such trends were determined by the use of the formulas $\frac{100\text{-}Y_o}{Y_o} = X_o$ and $\frac{100\text{-}Y_d}{Y_d} = X_d$ when Y_o is the origin and Y_d the destination of the best fitting straight line for the modal classes, and X_o and X_d are the origin and destination, respectively, for the lines to be determined. The quotient thus obtained may be called the extramodal coefficient because it takes into account the population outside the modal class. The trends thus obtained (Figs. 10, 11, 12, 13) indicate that variability becomes less as a low mean is approached and greater as a high mean is approached. The method is empirical and therefore has the limitations that most empirical formulas have.

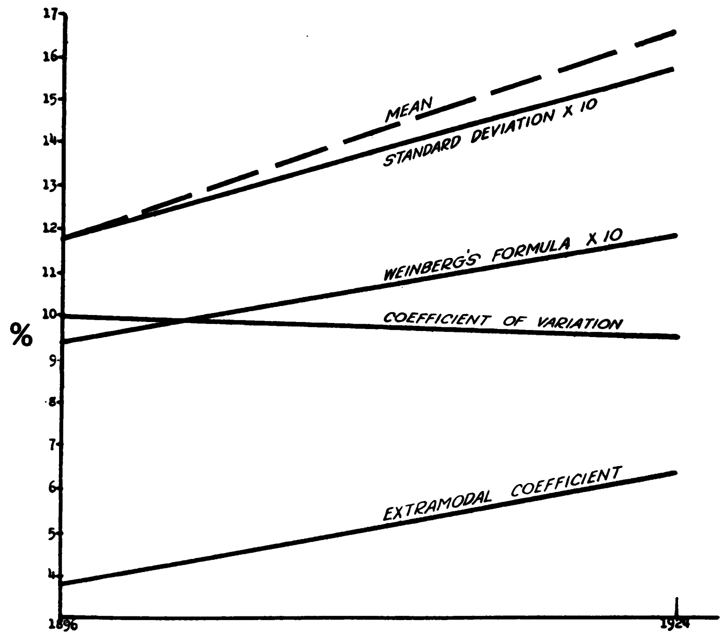

Figure 10. Comparison of four different measures of variation for protein content of Illinois high-protein corn as expressed by the best fitting straight lines; the broken line shows the trend of the mean protein content.

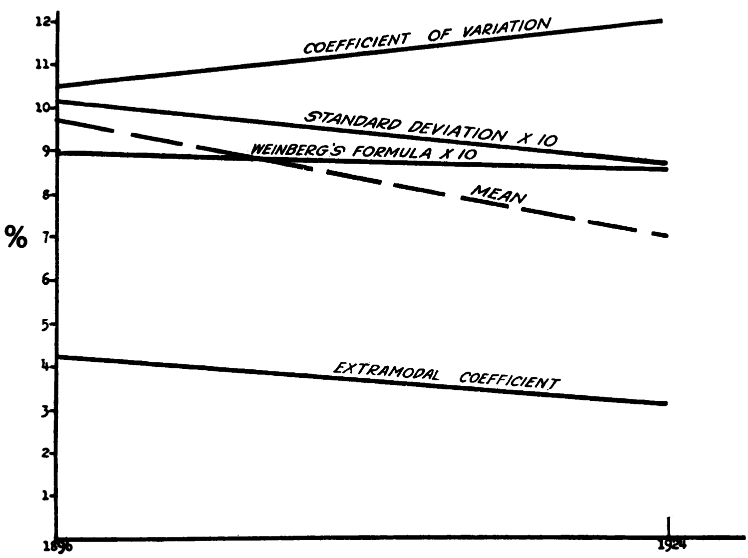

Figure 11. Comparison of four different measures of variation for protein content of Illinois low protein corn as expressed by the best fitting straight lines; the broken line shows the trend of the mean protein content.

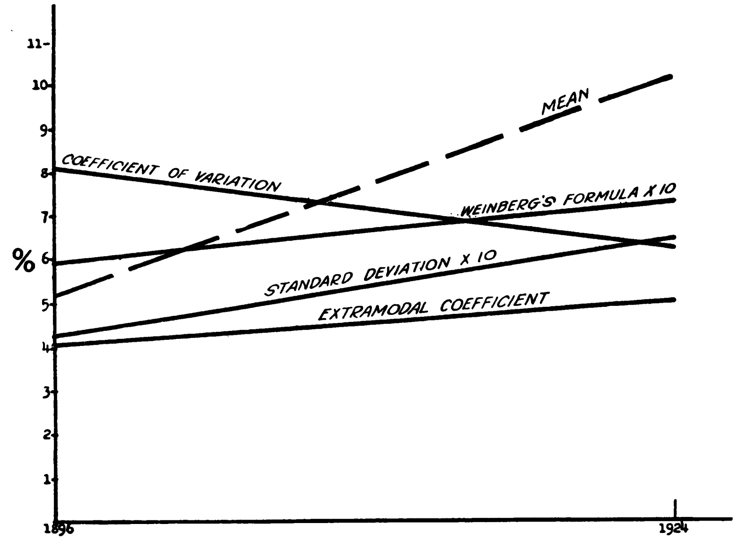

Figure 12. Comparison of four different measures of variation for oil content of Illinois high-oil corn as expressed by the best fitting lines; the broken line shows the trend of the mean oil content.

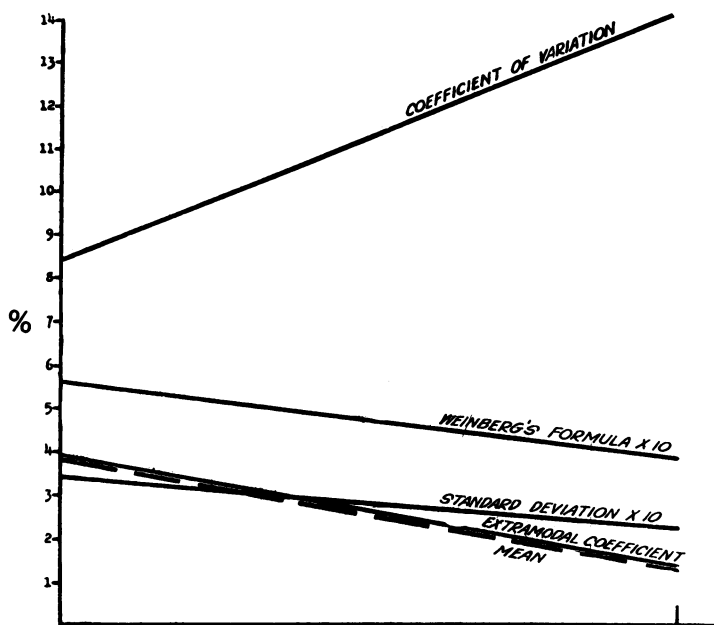

Figure 13. Comparison of four different measures of variation for oil content of Illinois low-oil corn as expressed by the best fitting straight lines; the broken line shows the trend of the mean oil content.

Table 2. Comparison of change in variability for chemical composition in four Illinois strains of corn as determined by the best fitting lines for four different measures of variability; 1896 and 1924.

Strain	Method used to measure variability	Year		Percentage of increase (+) or decrease (−)
		1896	1924	
High protein	Coefficient of variation	10.03	9.54	[a] −4.86
	Standard deviation	1.18	1.57	+33.05
	Weinberg's formula	.95	1.19	+25.26
	Extramodal coefficient	3.79	6.37	+68.07
Low protein	Coefficient of variation	10.53	12.02	+14.15
	Standard deviation	1.01	.86	−14.85
	Weinberg's formula	.89	.86	−3.37
	Extramodal coefficient	4.19	3.14	−25.06
High oil	Coefficient of variation	8.11	6.23	−23.18
	Standard deviation	.43	.64	+48.84
	Weinberg's formula	.59	.73	+23.73
	Extramodal coefficient	4.09	5.02	+22.74
Low oil	Coefficient of variation	8.50	14.44	+69.88
	Standard deviation	.34	.21	−38.24
	Weinberg's formula	.56	.39	−30.36
	Extramodal coefficient	3.68	1.50	−59.24

[a] These figures differ slightly from those in the text. The figures in the text were based on four figures after the decimal.

COMPARISON OF THE DIFFERENT MEASURES OF VARIABILITY

A comparison of the four different measures of variability discussed above for each of the four strains is shown in Figs. 10, 11, 12, and 13. The standard deviation and the coefficient obtained by the Weinberg method are multiplied by 10 so that they may be plotted on a common scale with the other measures. It is to be noted that in all cases variability progresses in the same direction as the mean when measured by the standard deviation, Weinberg formula, and the extramodal coefficient. If variability is expressed as a percentage of the mean it decreases as the mean increases, and vice versa. Of the four strains, the low protein shows on an average the least change in variability (Table 2).

RELATIVE VARIABILITY IN PROTEIN AND OIL

The protein strains show greater variability than do the oil strains (Tables 3 to 6), thus suggesting that the former may be affected to a greater extent by the environment, Hopkins states (8, p. 239) "the fat content of corn is even more susceptible to the influence of seed selection than is the protein content, doubtless due to the fact that the primary materials from which fat is manufactured, namely, carbon dioxide and water, are usually furnished to the plant in unlimited supply, while the formation of protein is essentially dependent upon the supply of available nitrogen in the soil." Variability is usually augmented when the living material under study is grown at a minimum or maximum rather than at the optimum condition.

Table 3. Average protein content and its percentage variability in Illinois high-protein strain corn, 1896-1924, as measured by different methods, together with extreme variates.

Year	Ears analyzed	Average protein content	Standard deviation	Coefficient of variation	Variability by Weinberg's formula	Lowest variate	Highest variate	Population in modal class
	Number	*Per cent*			*Per cent*			*Per cent*
1896	163	[a] 10.93	[a] 1.04	[a] 9.50	0.881	8.3	13.9	20.9
1897	112	10.99	1.16	10.90	1.008	8.3	13.6	---------
1898	252	10.98	1.22	11.15	.912	7.7	14.9	---------
1899	216	11.62	1.28	11.00	1.012	8.4	14.8	---------
1900	216	12.62	1.02	8.09	.795	9.3	15.7	19.4
1901	114	13.78	1.17	8.48	1.103	11.5	16.0	---------
1902	90	12.90	1.10	8.50	.965	9.5	15.0	---------
1903	100	13.51	1.36	10.04	.925	8.5	17.3	---------
1904	100	15.03	1.36	9.05	1.040	10.6	17.8	---------
1905	119	14.73	1.26	8.55	.998	10.8	17.4	21.0
1906	120	14.26	1.31	9.19	.978	10.5	17.7	---------
1907	120	13.90	1.49	10.72	1.119	10.3	17.4	---------
1908	119	13.94	1.66	11.91	1.193	9.4	17.3	---------
1909	120	13.29	1.43	10.76	1.076	9.2	16.4	---------
1910	120	14.87	1.44	9.68	1.109	11.2	18.0	15.8
1911	120	13.79	1.79	12.98	1.344	10.3	17.4	---------
1912	120	14.49	1.13	7.80	.854	10.3	17.5	---------
1913	120	14.83	1.22	8.23	.965	11.6	18.0	---------
1914	120	15.04	1.42	9.44	1.148	11.5	17.8	---------
1915	120	14.54	1.47	10.19	1.076	10.7	18.2	15.0
1916	120	15.66	1.34	8.56	1.047	12.7	19.3	---------
1917	120	14.45	1.85	12.80	1.277	10.2	18.6	---------
1918	120	15.49	1.36	8.78	1.022	11.8	18.9	---------
1919	120	14.70	1.55	10.54	1.201	11.1	17.8	---------
1920	120	14.01	1.79	12.78	1.287	9.5	17.4	11.7
1921	120	16.66	1.84	11.04	1.441	9.4	18.8	---------
1922	120	17.34	1.24	7.15	.890	12.6	20.6	---------
1923	120	16.53	1.41	8.50	1.116	13.1	19.7	---------
1924	120	16.60	1.19	7.17	1.120	14.6	19.2	16.7

[a] Data for the years 1896-1903 are taken from Davenport's Principles of Breeding (*1*, *p. 446*) means and standard deviations for the remaining years are calculated by the nongrouping method.

Table 4. Average protein content and its percentage variability in Illinois low-protein strain corn, 1896-1924, as measured by different methods, together with extreme variates.

Year	Ears analyzed	Average protein content	Standard deviation	Coefficient of variation	Variability by Weinberg's formula	Lowest variate	Highest variate	Population in modal class
	Number	*Per cent*			*Per cent*			*Per cent*
1896	163	[a] 10.93	[a] 1.04	[a] 9.50	0.881	8.3	13.9	20.9
1897	60	10.63	.90	8.47	.759	8.2	14.0	---------
1898	126	10.49	1.32	12.61	1.089	7.5	13.4	---------
1899	144	9.59	1.01	10.50	.802	6.7	13.1	---------
1900	144	9.13	1.04	11.34	.937	7.1	12.3	19.4
1901	126	9.63	1.10	11.47	.975	7.6	13.1	---------
1902	90	7.86	.75	9.60	.829	6.4	9.7	---------
1903	100	8.00	.83	10.41	.862	6.4	10.2	---------
1904	100	8.17	.81	9.91	.773	6.1	10.5	---------
1905	120	8.58	.85	9.91	.753	6.6	12.1	20.0
1906	120	8.65	.92	10.64	.899	6.7	10.9	---------
1907	120	7.32	.92	12.57	.910	5.8	10.5	---------
1908	120	8.96	1.26	14.06	1.117	6.3	11.4	---------
1909	120	7.48	.94	12.57	.842	5.5	10.8	---------
1910	120	8.26	.86	10.41	.827	6.5	11.0	21.7
1911	120	7.90	1.17	14.81	1.005	5.9	12.1	---------
1912	120	8.23	.82	9.96	8.860	6.8	10.8	---------
1913	120	7.71	.95	12.32	.861	5.7	10.8	---------
1914	120	7.67	.95	12.39	.890	5.9	10.8	---------
1915	120	7.27	.85	11.69	.854	5.7	9.9	27.5
1916	120	8.68	1.03	11.86	.994	6.6	10.9	---------
1917	120	7.09	.71	10.01	.233	5.6	9.6	---------
1918	120	7.13	.75	10.52	.894	5.9	8.8	---------
1919	120	6.46	.52	8.05	.629	5.4	8.3	---------
1920	120	7.54	.89	11.80	.890	6.0	10.5	22.5
1921	120	9.14	1.35	14.77	1.074	6.6	13.4	---------
1922	120	7.42	.70	9.43	.774	6.1	9.6	---------
1923	120	6.48	.73	11.27	.708	5.0	9.4	---------
1924	120	8.38	1.17	13.96	.998	6.1	11.8	20.0

[a] Data for the years 1896, 1898, 1899, 1900, 1901, 1902, and 1903 are taken from Davenport's Principles of Breeding (*1*, *p. 446*); means and standard deviations for the remaining years are calculated by the nongrouping method.

Table 5. Average oil content and its percentage variability in Illinois high-oil strain corn, 1896-1924, as measured by different methods, together with extreme variates.

Year	Ears analyzed	Average oil content	Standard deviation	Coefficient of variation	Variability by Weinberg's formula	Lowest variate	Highest variate	Population in modal class
	Number	*Per cent*			*Per cent*			*Per cent*
1896	163	[a] 4.68	[a] 0.41	[a] 8.83	0.585	3.9	6.0	19.6
1897	80	4.79	.38	7.87	.543	3.6	5.7	--------
1898	216	5.10	.48	9.33	.615	4.1	6.7	--------
1899	108	5.65	.42	7.47	.582	4.3	6.5	--------
1900	108	6.10	.44	7.26	.527	4.6	7.4	23.1
1901	126	6.24	.45	7.26	.621	4.9	7.1	--------
1902	90	6.25	.50	8.06	.680	5.0	7.2	--------
1903	100	6.51	.46	7.07	.608	5.5	7.6	--------
1904	101	7.11	.61	8.61	.792	6.0	8.4	--------
1905	120	7.30	.49	6.70	.651	6.3	8.6	14.2
1906	120	7.38	.40	5.46	.594	6.6	8.5	--------
1907	120	7.43	.51	6.82	.522	6.2	8.7	--------
1908	120	7.21	.53	7.36	.658	5.9	8.5	--------
1909	120	7.05	.47	6.62	.580	5.8	8.4	--------
1910	120	7.72	.57	7.36	.718	6.5	9.0	19.2
1911	120	7.52	.69	9.20	.763	5.9	9.2	--------
1912	120	7.71	.50	6.49	.778	6.5	8.7	--------
1913	120	8.16	.62	7.63	.700	6.4	9.6	--------
1914	120	8.30	.68	8.23	.668	5.8	10.1	--------
1915	120	8.47	.49	5.77	.576	6.9	9.8	15.0
1916	120	8.51	.49	5.80	.592	7.2	10.0	--------
1917	120	8.52	.58	6.78	.762	7.2	9.7	--------
1918	120	9.36	.53	5.67	.674	8.0	10.5	--------
1919	120	9.06	.84	9.25	.961	7.3	10.4	--------
1920	120	9.28	.52	5.57	.619	7.8	10.6	17.5
1921	120	9.94	.66	6.64	.729	8.4	11.7	--------
1922	120	9.86	.54	5.48	.674	8.7	11.3	--------
1923	120	10.08	.65	6.45	.695	8.3	11.8	--------
1924	120	9.86	.61	6.19	.676	8.4	11.7	18.3

[a] Data for the years 1896–1903 are taken from Davenport's principles of breeding (*1, p. 446*); means and standard deviations for the remaining years are calculated by the nongrouping method.

Table 6. Average oil content and its percentage variability in Illinois low-oil strain corn, 1896-1924, as measured by different methods, together with extreme variates.

Year	Ears analyzed	Average oil content	Standard deviation	Coefficient of variation	Variability by Weinberg's formula	Lowest variate	Highest variate	Population in modal class
	Number	*Per cent*			*Per cent*			*Per cent*
1896	163	[a] 4.68	[a] 0.41	[a] 8.83	0.585	3.9	6.0	19.6
1897	50	4.10	.29	7.10	.510	3.4	4.7	--------
1898	108	3.59	.32	8.13	.589	3.2	4.8	--------
1899	144	3.85	.32	8.42	.484	2.8	4.6	--------
1900	144	3.57	.36	10.13	.522	2.6	4.5	24.4
1901	126	3.45	.26	7.59	.456	2.8	4.1	--------
1902	90	3.00	.32	10.83	.492	2.1	3.8	--------
1903	90	2.99	.23	7.83	.441	2.5	3.6	--------
1904	100	2.89	.46	15.91	.920	2.4	3.4	--------
1905	119	2.58	.31	11.86	.502	1.8	3.1	30.3
1906	120	2.67	.30	11.35	.443	1.6	3.5	--------
1907	120	2.60	.24	9.04	.466	2.2	3.3	--------
1908	120	2.39	.27	11.09	.507	1.8	2.9	--------
1909	120	2.24	.23	10.40	.402	1.4	2.8	--------
1910	120	2.21	.20	9.05	.384	1.6	2.7	30.0
1911	120	2.06	.23	11.02	.398	1.4	2.7	--------
1912	120	2.19	.22	10.05	.386	1.3	2.7	--------
1913	120	1.91	.24	12.30	.451	1.3	2.4	--------
1914	120	1.98	.26	13.33	.446	1.3	2.7	--------
1915	120	2.07	.25	11.93	.410	1.4	3.1	35.0
1916	120	2.07	.50	24.30	.828	1.3	4.7	--------
1917	120	2.10	.24	11.38	.474	1.7	2.8	--------
1918	120	1.88	.32	16.86	.489	1.2	3.0	--------
1919	120	1.77	.21	11.92	.394	1.3	2.5	--------
1920	120	1.80	.21	11.83	.323	1.0	2.4	37.5
1921	120	1.71	.15	8.77	.264	1.0	2.3	--------
1922	120	1.68	.19	11.55	.347	.9	2.2	--------
1923	120	1.58	.24	15.19	.480	1.1	2.1	--------
1924	120	1.51	.22	14.57	.387	.9	2.2	38.3

[a] Data for the years 1896, 1898, 1899, 1900, 1901, 1902, and 1903 are taken from Davenport's Principles of Breeding (*1, p. 446*); means and standard deviations for the remaining years are calculated by the nongrouping method.

SYMMETRY OF DISTRIBUTION

A number of other statistical expressions have been proposed intended to describe the nature of a population with respect to its distribution. Certain of these expressions are of interest in connection with the present study. An important item of information in regard to a distribution is whether the variates are symmetrically distributed with reference to the mean or whether there is a bunching of variates on one side of the mean and a long tailing out of the variates on the other side; i. e., to know the amount of skewness.

Another item that should be known is whether the variates are densely grouped at the mean, giving a high peak to the frequency polygon, or whether the distribution is rather flat in the middle and contracted at the ends, or whether the distribution of the variates is intermediate between these two conditions; i. e., to know the amount of kurtosis. A normal distribution is said to be mesokurtic, a peaked curve leptokurtic, and a flat curve platykurtic.

The mean, median, mode, standard deviation, coefficient of variation, variation by Weinberg's formula, skewness, and kurtosis are given in Table 7 for each of the four strains taken at periodic intervals. The median, mode, skewness, and kurtosis were calculated on the basis of percentiles.[1] Kelley (10, p. 58–62, 75–77) states that "this method of determining curve types, although in general not as accurate as the longer method of Pearson, can be used where the populations are large and the standard errors are small." The variates of the high-oil and the low-oil strains appear to be fairly symmetrically distributed with reference to the mean. In no case is skewness significantly different from zero. In 5 of the 7 years studied the median is greater than the mean in the high-oil strain. In the low-oil strain the median is lower than the mean in 4 of the 7 years. The differences in all cases are very small.

The distributions for the low-protein strain are negatively skewed, i. e., tail out on the high side for each of the years studied except 1896. Although the skewness is much greater than in the case of the oil strains, in no instance is skewness significantly different from zero. Likewise, the distributions for the high-protein strain show a greater skewness than do the distributions for the oil strains, but in no case is the skewness significantly different from zero.

[1]Mode = Mean − 3.03 (mean − median); Skewness = $P_{.50} - \frac{1}{2}D$ (S.D. of Sk = $0.55914\ D/N^{1/2}$, $D = P_{.50} - P_{.10}$); Kurtosis = Q/D (Q = quartile deviation, S.D. of Ku = $0.27779/N^{1/2}$). To determine the class in which the $pN + \frac{1}{2}$ measure lies, let f_p = the frequency in this class, i_p = the interval or range covered by this class, F_p = the sum of the frequencies in all classes below this class, v_p = the value of the lower boundary of this class, N = the total population, P_p = percentile (the value of which is to be calculated), p = proportion of classes having values smaller than P_p. Then:

$$P_p = v_p + [(pN - F_p)/f_p]\, i_p.$$

Table 7. Measurements of central or average tendencies, and of dispersions for distributions of the variates of four Illinois strains of corn taken at periodic intervals.

HIGH-PROTEIN STRAIN

Year	Number of ears analyzed	Mean	Median	Mode	Standard deviation [a]	Coefficient of variation [a]	Variability by Weinberg's formula [a]	Skewness [b]	Kurtosis [c]
1896	163	10.96	11.04	11.18	1.04	9.50	0.88	+0.1082±0.0874	−0.0008±0.0147
1900	216	12.69	12.64	12.55	1.02	8.09	.80	−.1312± .0704	+.0101± .0127
1905	119	14.77	14.94	15.29	1.26	8.55	1.00	+.3997± .2312	−.0320± .0172
1910	120	14.93	15.08	15.37	1.44	9.68	1.11	+.2159± .1313	+.0229± .0171
1915	120	14.55	14.67	14.92	1.47	10.19	1.08	+.2447± .1426	+.0106± .0171
1920	120	14.04	13.92	13.67	1.79	12.78	1.29	−.2885± .1775	+.0206± .0171
1924	120	16.68	16.65	16.58	1.19	7.17	1.12	−.1027± .1237	−.0065± .0171

LOW-PROTEIN STRAIN

Year	Number of ears analyzed	Mean	Median	Mode	Standard deviation [a]	Coefficient of variation [a]	Variability by Weinberg's formula [a]	Skewness [b]	Kurtosis [c]
1896	163	10.96	11.04	11.18	1.04	9.50	0.88	+0.1082±0.0874	+0.0008±0.0147
1900	144	9.19	9.07	8.83	1.04	11.34	.94	−.2044± .0922	+.0043± .0156
1905	120	8.64	8.48	8.15	.85	9.91	.75	−.1772± .0946	+.0398± .0171
1910	120	8.32	8.24	8.08	.86	10.41	.83	−.1443± .0870	−.0085± .0171
1915	120	7.30	7.16	6.85	.85	11.69	.85	−.2182± .0775	+.0018± .0171
1920	120	7.60	7.50	7.29	.89	11.80	.89	−.1543± .0835	+.0061± .0171
1924	120	8.44	8.34	8.14	1.17	13.96	1.00	−.1865± .1049	−.0050± .0171

HIGH-OIL STRAIN

Year	Number of ears analyzed	Mean	Median	Mode	Standard deviation [a]	Coefficient of variation [a]	Variability by Weinberg's formula [a]	Skewness [b]	Kurtosis [c]
1896	163	4.74	4.83	5.00	0.41	8.83	0.59	+0.0323±0.0391	−0.0368±0.0147
1900	108	6.15	6.20	6.31	.44	7.26	.53	+.0654± .0653	−.0051± .0180
1905	120	7.35	7.34	7.32	.49	6.70	.65	−.0053± .0519	+.0330± .0171
1910	120	7.78	7.72	7.60	.57	7.36	.72	−.1090± .0507	−.0274± .0171
1915	120	8.52	8.58	8.69	.49	5.77	.58	+.1098± .0574	+.0014+ .0171
1920	120	9.34	9.35	9.38	.52	5.57	.62	+.0850± .0491	−.0192± .0171
1924	120	9.92	10.00	10.17	.61	6.19	.68	+.0117± .0726	−.0301± .0171

LOW-OIL STRAIN

Year	Number of ears analyzed	Mean	Median	Mode	Standard deviation [a]	Coefficient of variation [a]	Variability by Weinberg's formula [a]	Skewness [b]	Kurtosis [c]
1896	163	4.74	4.83	5.00	0.41	8.83	0.59	+0.0323±0.0391	−0.0368±0.0147
1900	144	3.63	3.61	3.56	.36	10.13	.52	−.0565+ .0304	+.0007± .0153
1905	119	2.64	2.62	2.58	.31	11.86	.50	−.0197± .0262	+.0250± .0172
1910	120	2.26	2.27	2.29	.20	9.05	.38	+.0260+ .0241	+.0171± .0171
1915	120	2.13	2.12	2.12	.25	11.93	.41	−.0177± .0209	+.0155± .0171
1920	120	1.84	1.86	1.89	.21	11.83	.32	+.0257± .0227	−.0268± .0171
1924	120	1.56	1.54	1.52	.22	14.57	.39	−.0130± .0240	−.0322± .0171

[a] Taken from Tables 3, 4, 5, and 6.
[b] A positive sign (+) indicates that the distribution tails out on the low side.; a negative sign (−) indicates the reverse.
[c] Data reported equals Kü. less 0.26315. If sign is positive (+) curve is platykurtic; if sign is negative (−) curve is leptokurtic.

The four strains show but little deviation from a mesokurtic curve. The small deviations from the normal probability curve are as frequent in the direction of a leptokurtic curve as of a platykurtic curve.

A study of the extreme variates of the different strains for the period of selection (Fig. 14) shows that in the high-oil strain the variates deviate as far on the low side as they do on the high side. In the low-oil strain the deviation is slightly greater on the high side. However, the extreme variates in the high-oil strain are about twice as far from the mean as those in the low-oil strain. The range in the low-oil strain is becoming less while that in the high-oil strain is becoming greater.

The extreme variates in the low-protein strain are farther from the mean in a positive direction than in a negative direction, while in the high-

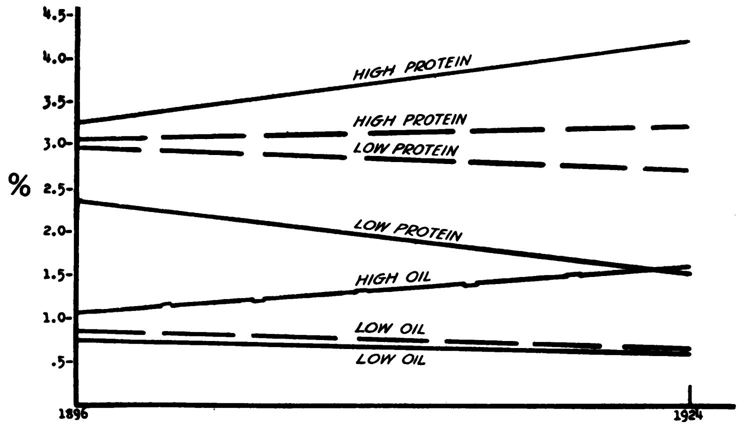

Figure 14. Comparison of the range between the lowest variate and the mean (solid line), and the highest variate and the mean (broken line), for high-protein, low-protein, high-oil, and low-oil Illinois corn as expressed by the best fitting straight lines.

protein strain the reverse is true. This means that the distribution tails out toward the mean of the original nonselected material in both cases.

GENERAL DISCUSSION

If, as has been stated by many, the population with the greatest variability offers the greatest chance for improvement by selection then it should be possible to make still greater progress in the Illinois high-protein and Illinois high-oil strains in future years than has been made in the past because the range of each was greater in 1924 than it was in 1896 (Fig. 15). Such a conclusion would be rather questionable. As has been stated, the strains now appear to be more nearly homozygous than was the original material, and our knowledge of genetics and the effects of selection would lead us to believe that they should be. Being more nearly homozygous, they should be less variable genetically. The apparent increase in variability must then be due either to the environment or to the methods of measuring variability.

Emerson (4, p. 30–31) in speaking of the variability in number of rows on an ear of corn says: ". . . It is more reasonable to suppose that an ear which can vary in any one of eight spikes will show a greater degree of fluctuation than one which can vary only in any one of four spikes. For this reason it is likely that strains with a high number of rows will never show the low variability seen in strains with a low number of rows." One may reason in like manner about variability of the oil and of the protein content for the different strains. The high strains are more variable than the low strains because there is more material present for the environment to interact with. Supposing the germ of the high-oil strain to contain 600 cells and that of the low oil to contain 300 cells, we might expect twice the effect on the high-oil

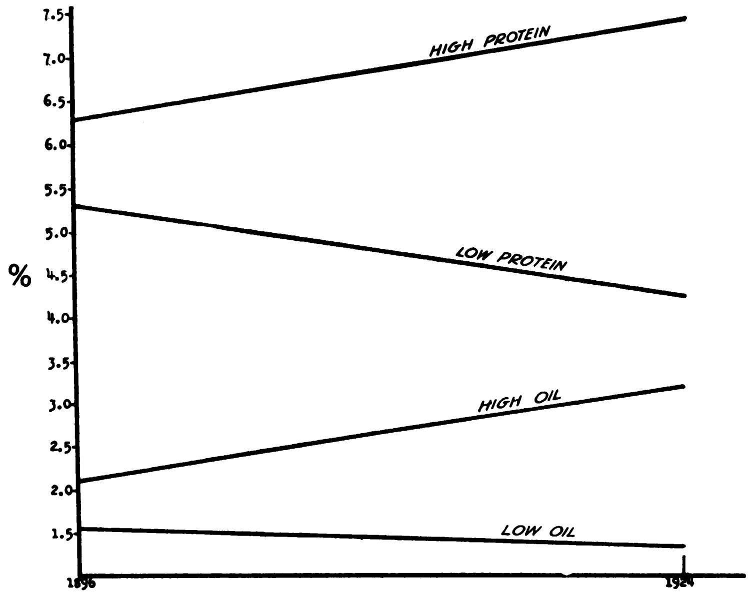

Figure 15. Comparison of the ranges between the highest variate and the lowest variate for high-protein, low-protein, high-oil, and low-oil Illinois corn as expressed by the best fitting straight lines.

strain from environmental action as on the low oil under identical conditions.

Zeleny (15, p. 15) states: "It is a common principle of embryology that a changed condition does not act by accretion, i. e., by addition or subtraction of individual parts without affecting the rest. On the contrary, the action is upon all of the preexisting parts of the organ." This he calls the theory of proportionate action. Zeleny, in the study of the effect of selection for eye facet number in the white bar-eye race of *Drosophila melanogaster,* found that a race with 300 facets was affected ten times as much as a race with only 30 facets by a 1C change in temperature. In order to measure the variability of races with different facet numbers he scaled his classes for grouping so that each class range was a fixed percentage of the mean of its class. He used as a mid or zero point the mean of the original population from which his selected material came. He (15, p. 15) goes on to explain that "the standard deviation, as determined by this method, is expressed in factorial units and serves directly as a coefficient of variation strictly comparable in all cases, regardless of the mean values of the different stocks that may be compared."

Should such a method of measuring variability be applied to the distributions of the oil and protein strains in the present study, it can be readily seen that the degree of deviation from the means would depend upon the value taken as a percentage of the means of the classes. The larger this figure the smaller the deviation of the high strain as compared to the low strain. Also, the farther the high strain is removed from the low strain the greater will be its comparative reduction in variability. The factorial method of measuring variability has not been used in an analysis of these data because the method affords no manner of determining the figure to be used in arriving at the class ranges with such data. Such a figure might be 10, 20, or 23%. Zeleny apparently uses such a figure as will give a normal distribution for the populations studied. As has been shown, the distributions for protein and for oil content thus far studied do not show a significant deviation from the normal curve.

Although it appears logical to attribute the increase in variation in the high strains to the fact that there is more material for the environment to interact with, it is impossible to prove it, because the variability due to the environment can not be separated from the variability due to the segregation and recombination of factors. The complexity of the inheritance of protein and oil content and the manner of conducting the selection work make it entirely impossible to analyze the four strains genetically. Others who have studied the protein content of corn have also found it impossible to determine the genetic factors involved. East (3) lists a large number of factors other than genetic that may and probably do affect protein content, e. g., number of seeds on the ear, size of pericarp, lack of phosphorus, and departure from optimum temperature and moisture at critical periods. In addition to these there may be considered the factors affecting size of germ, size of endosperm, absorption of different amounts of food elements through the roots and their translocation through the stalk and subsequent deposition within the pericarp. The total protein content is made up chemically of at least four protein groups (13) each of which may be represented by a single or several genes in the germ plasm. Some of these factors may be dominant and others recessive, or there may be any graduation of dominance. Hayes (5) states: "Protein content is therefore inherited in much the same way as other characters which are dependent for their full expression on many different inherited factors of the plant and likewise upon environmental conditions." What has been said of the protein content may likewise be said of the oil content.

SUMMARY

Twenty-eight years of continuous selection for protein and oil content in corn has produced four types which are distinctly different in their composition. When compared with the original nonselected material the high-protein and the high-oil strains show a proportional increase of 50.01 and 109.79%, respectively. The low-protein and low-oil strains show a proportional decrease of 23.26 and 67.87%, respectively.

The high-protein and high-oil strains show no indications of having reached a limit to further increases.

The low-protein strain has changed but little during the last 20 years.

The low-oil strain is approaching a physiological limit to further decreases. Ears with extremely low oil content have a high percentage of germless seeds.

The four different strains now trace back in their pedigrees to a single ear each.

Variability has been shown to change considerably following selection. The degree of change in variability depends somewhat upon the method used in measuring it.

Variability of oil or protein content appears to depend upon the magnitude of the mean of the selected character.

Variability as measured by the coefficient of variation increases when selection leads to a low mean, and vice versa.

Variability, as measured by the standard deviation, Weinberg's formula, and extramodal coefficient, increases when selection leads to a high mean, and vice versa.

The percentage of the population lying in the modal classes decreases when selection leads to a high mean, and vice versa.

The symmetry of the distribution curve as determined by the percentile method for the four strains taken at periodic intervals is not significantly different from that of the normal variability curve.

It is suggested that the apparent increase in variability of the high strains may be due to the fact that there is more material present for the environment to interact with.

ACKNOWLEDGMENT

The writer wishes to thank Dr. L. H. Smith for suggesting the problem. To Dr. C. M. Woodworth, under whose direction these studies were made, the writer expresses sincere thanks for suggestions for analyzing the data and for valuable assistance in the preparation of this paper.

LITERATURE CITED

1. Davenport, E. 1907. Principles of breeding. A treatise on thremmatology; or the principles and practices involved in the economic improvement of domesticated animals and plants. 727 p. Boston, New York.
2. East, E. M. 1910. The role of selection in plant breeding. Pop. Sci. Mo. 77:190-203.
3. - - - - - and Jones, D. F. 1920. Genetic studies on the protein content of maize. Genetics 5:543-610.
4. Emerson, R. A., and E. M. East. 1913. The inheritance of quantitative characters in maize. Nebr. Agr. Expt. Sta. Research Bul. 2, 120 p.
5. Hayes, H. K. 1922. Production of high-protein maize by Mendelian methods. Genetics 7:237-257.
6. Hopkins, C. G. 1896. Composition and digestibility of corn ensilage, cow pea ensilage, soja bean ensilage, and corn-fodder. Ill. Agr. Expt. Sta. Bul. 43, 181-201. (Appendix: Constituents and uses of food, p. 202-208.)

7. - - - - -. 1898. The chemistry of the corn kernel. Ill. Agr. Expt. Sta. Bul. 53, 127-180. (Abstract) p. 180a-180d.
8. Hopkins, C. G. 1899. Improvement in the chemical composition of the corn kernel. Ill. Agr. Expt. Sta. Bul. 55, 205-240.
9. - - - - -, Smith, L. H., and E. M. East. 1903. The structure of the corn kernel and the composition of its different parts. Ill. Agr. Expt. Sta. Bul. 87, 77-112.
10. Kelley, T. L. 1923. Statistical method. 390 p. New York.
11. Pearl, R. 1923. Introduction to medical biometry and statistics. 379 p. Philadelphia and London.
12. - - - - -. 1927. The graphic representation of relative variability. Science 65:237-241.
13. Showalter, M. F., and R. H. Carr. 1922. Characteristic proteins in high and low protein corn. (Abstract) Science 56:24-25.
14. Weinberg, W. 1926. Zur messung der streuung bei extensiven masszahlen. Arch. Rassen u. Gesell. Biol. 18:84.
15. Zeleny, C. 1922. The effect of selection of eye facet number in the white bar-eye race of Drosophila melanogaster. Genetics 7:1-115.