T1980AGRO03

CORRELATION AND PATH COEFFICIENT ANALYSIS OF THE COMPONENTS OF YIELD IN SUNFLOWER CULTIVARS, *HELIANTHUS ANNUUS* L.

V. E. GREEN*

ABSTRACT

Attempts were made to quantify the components of yield through the use of correlation coefficients, regression equations, and path coefficient analysis. Ten parameters of yield were subjected to correlation coefficient analysis for all possible combinations. All but two factors were highly significant and positive. One was highly significant and negative, and one was insignificant. "Dewey and Lu (Agronomy J. 51-515-518 of 1959) said that the path coefficient can be performed on the "r" values because it is simply a standardized partial regression coefficient and as such measures the direct influence of one variable upon another and permits the separation of the "r" into components of direct and indirect effects. The use of the method requires a cause and effect situation among the variables. As more variables are considered in the "r" table, these indirect associations become more complex, less obvious, and somewhat perplexing. At this point, the path-coefficient analysis provides an effective means of untangling direct and indirect causes of association and permits a critical examination of the specific forces acting to produce a give correlation and measures the relative importance of each causal factor". Since seed moisture, number of heads harvested per plot, and lodging did not seem to affect yields, they were omitted from the analysis. The sum of the direct and indirect effects equals

* Institute of Food and Agricultural Sciences, University of Florida, Gainesville, USA.

the "r" with yield and are shown in tables. The coefficients are either " + " or " − " and can be greater than 1, although there were none greater than unity in this test. Since the direct and indirect effects add to the "r" value, they were reduced to percentages and shown in another table. Each of the direct and indirect effects of the parameters is discussed in sufficient detail to show its importance toward final yield figures. The Matrix Procedure, SAS User's Guide, 1979 Edition, Part 3 pp. 271 ff. has made this method of analysis, which was once very laborious and often endless with calculating machines, a rapid, easy analysis that could prove very useful in the selection of parents and cultivars for use in areas not in the usual areas of sunflower production and selection.

A test with 20 cultivars of sunflower, *Helianthus annuus* L., from 15 proprietary sources in the USA was grown from 14 Aug. to 28 Nov. on a fine sand near Gainesville, FL, USA. The soil was treated with FMC Furadan 10% granules as insecticide/nematicide and Elanco Products Treflan as herbicide at 10 kg. and 2.5 liters per ha., respectively. Fertilizer was applied at a rate of 670 kg/ha 4-8-16 (N-P_2O_5-K_2O) at planting time. NH_4NO_3 was side-dressed at 250 kg/ha on 13 September and the stand was thinned to one plant every hill at 30 cm. in the row with 90 cm. between rows. Rows ran north and south. Prior to planting the rows had been subsoiled and chiselled with a Brown and Harden Superseeder subsoiler machine. The soil level remained flat following a shallow cultivation to cover the NH_4NO_3 and to control weeds. The sunflower head moth, *Homoeosoma electellum* Hulst, was controlled with applications of Ortho Orthene at 1.0 kg/ha in water on 27 Sept. and 5 Oct. Dupont Manzate 200 at 1 kg/ha controlled Alternaria leaf and stem spot enough to take the experiment to maturity successfully. Sprays were made on 5 and 12 Oct. Harvest of 4 cultivars each were made following physiological maturity on these dates: Nov. 15, 19, 26, 27, and 28. Plots consisted of 4 rows and there were 6 replications. The heads were harvested, dried to constant moisture 45° C, threshed in a Vogel Thresher, and the achenes aspirated in a Bates aspirator. The yields were weighed, moisture was determined with an electronic moisture meter, and weight per unit volume was determined with standard equipment. Yields were extrapolated to kg/ha. Weights per volume were calculated in kg/hl.

In the field data were collected for dates of emergence, dates of flowering, dates of physiological maturity, plant heights, head diameters, and numbers of active leaves before senescence. Meteorological data were used to calulate Growing Degree Days (GDD) from emergence to flowering, flowering to maturity, and emergence to maturity. The number of seed heads harvested per row was also recorded, as were the weights of 200 seed per plot.

The following text table shows the means, standard deviations, least significant differences, and the coefficients of varations for nine parameters that were measured in the fall test:

Parameter	Mean	SD	LSD	CV
Yield, kilos/hectare	1290	210	227	16.08
Seed moisture, %	6.79	0.55	0.58	8.13
Weight per volume, kg/hl	37.3	1.04	1.11	2.78
Head diameter, cm.	13.1	1.23	1.48	9.35
Heads per plot cut	30.0	3.27	3.73	10.92
Lodging, %	1.78	2.21	0.45	124.4
Plant height, cm.	155	11	12	7.06
200-seed weights, g.	8.10	0.97	1.07	11.96
Active leaves, number	9.74	1.64	1.74	16.8

It has been shown that in the 2 Feb. planting in 1979 that there is a correlation between the time required for flowering and yields. In that test, higher yields were obtained from later flowering entries. A correlation coefficient on the 14 Aug. data showed a similar trend.

Also, it was decided to ascertain whether the time required from flowering to maturity had any effect on the yields of the entries in this test. In addition, the other parameters measured in this experiment were tested for their effect on yield: stalk heights, weight of 200 seed, diameter of the head, weight per volume, and number of heads in the area chosen for harvest ina plot. All the paramenters affected yield with the exception of the last named factor. The correlation coefficients derived were:

Yield versus:	r:	Regression equation
Days from emergence to flowering	0.534**	$Y = 473.23 + 34.11X$
Days from flowering to maturity	-0.373**	$Y = 2439.80 + 38.98X$
Days from emergence to maturity	0.538**	$Y = -3733.05 + 60.51X$
Weight per unit volume	0.435**	$Y = -1404.84 + 88.07X$

Head diameter	0.466**	Y=	69.07 + 206.40X
Stalk height	0.593**	Y=	-321.34 + 23.66X
Weight in grams of 200 seed	0.424**	Y=	483.01 + 82.21X
Number heads harvested per plot	0.031n.s.	Y=	not calculated
Number of active leaves per plant	0.589**	Y=	593.37 + 57.01X

Weight per bushel, or the newer weight per volume measurement in the metric system, the kilograms per hectoliter, is a good indication of plumpness, and how well the seed is filled out within the pericarp. Lighter weights per volume indicate chaffiness or poor seed fill. The 200-seed weight is also a good indicator of seed quality. Heavier seeds seem to yield better oil quantities per unit weight of seed. But is there any connection between these two quality indicators? No attempts were made to correlate these two measurements until the fall test of 1979 oilseed cultivars. A correlation coefficient was computed between the 120 weight per volume figures and the 120 seed counts of 200 seed in grams. The resulting r value was 0.12347, insignificant up to the 17.9% level of significance.

Table 1 shows the results of a correlation coefficient analysis using the SAS procedures at the North East Florida Regional Data Center (NERDC), University of Florida. Each agronomic characteristic measured in a fall-planted sunflower test of 20 entries was subjected to correlation coefficient analysis with every other characteristic. The data consisted of 120 pairs of data, that is, data from 20 entries in six replications. Even though some meaningless correlations were tested with this method, many meaningful and informative "r" values were derived that assist in explaining relations among the data. Those coefficients that are highly significant for this test are marked with double asterisks (**) in the normal manner in the table.

In addition, a path-coefficient analysis was performed on the correlation coefficients, which, according to Dewey and Lu (Agronomy Journal 51:515-518. 1959) "is simply a standarlized partial regression coefficient and as such measures the *direct* influence of one variable upon an other and permits the separation of the correlation coefficient into components of direct and indirect effects. The use of the method requires a cause and effect situation among the variables, and the experimenter must assign direction in the causal system based upon *a priori* grounds or experimental evidence". They agree that it can be apparent that many characters such as those in Table 1 are correlated "because of a mutual association, positive or negative, with other characters. As more variables are considered in the correlation table, these indirect associations become more complex, less obvious, and somewhat perplexing. At this point, the path-coefficient analysis provides an effective means of untangling direct and indirect causes of association and permits a critical examination of the specific forces acting to produce a given correlation and measures the relative importance of each causal factor".

The Matrix Procedure, SAS User's Guide, 1979 Edition, Part 3, page 271 ff., was used to set up the matrix used to partition the direct and indirect effects of the parameters in the correlation coefficients of yield components. Since seed moisture, number of heads harvested per plot, and lodging did not seem to have any effect on yield in this test, see Table 1, these characteristics were omitted from the path-coefficient analysis.

The resulting path-coefficients are shown in Table 2. The direct effects of each factor on yield is underlined in the table. The sum of the direct and indirect factors for a given parameter equals the correlation coefficient with yield. The coefficients are either plus or minus and can be greater than unity, although there were none greater than unity in this test.

As an example, the path-cofficient of Weight per volume vs. Yield (r= 0.435**) is derived by summing the direct and indirect effects, as follows:

$$\mathit{0.1646} + 0.0520 + 0.0155 + 0.0297 + 0.0455 + 0.0185 + 0.0094 + 0.0998 = 0.435$$

1. The direct effect, Weight per volume, on yield was the greatest of any of the yield components. This shows that well filled, heavy, plump achenes have a positive effect on the yield of sunflower in this test per unit area. Since the indirect effects of the other factors were all positive and smaller than the direct effect, we can assume that the other factors contributed less to the correltion coefficient and that none of them counterbalanced the direct effect of weight per volume on yield.

2. The direct effect of head diameter on yield and the indirect effects of the other contributing factors follows the same pattern as described above for weight per unit volume.

3. The direct effect of height of plant on yield presents a different picture. Although the highest correlation coefficient was obtained in this test was for height of plant, 0.593**, plant height was

TABLE 1

Correlation coefficients of all possible combinations of the parameters measured as agronomic characteristics of the entries in an August planted ilsced sunflower test at the Green Acres Farm, Gainesville, Florida, 1979

	WT/VOL	HEAD DIAM	HDS PLOT	LOD GING	HEIGHT	SEED WT	ACTV LVS	DAYS EF	DAYS FM	DAYS EM	YIELD
MOISTURE	0.223*	-0.052	-0.011	-0.113	0.093	-0.117	0.399**	0.099	-0.086	0.081	0.111
WT/VOL		0.255**	-0.109	-0.058	0.533**	0.123	0.150	5.69**	-0.473**	0.492**	0.435**
HD DIAM			-0.305*	0.088	0.320**	0.298**	0.314**	0.241**	-0.253**	0.151	0.466**
HDS PLOT				0.018	-0.096	-0.217*	-0.136	-0.037	0.221*	0.172	0.031
LODGING					-0.023	0.143	-0.005	-0.085	0.101	0.041	0.106
HEIGHT						0.233*	0.586**	0.731**	-0.544**	0.702**	0.593**
SEED WT							0.236**	0.081	-0.044	0.096	0.424**
ACTV LVS								0.546**	-0.451**	0.475**	0.589**
DAYS EF									-0.860**	0.835**	0.534**
DAYS FM										-0.436**	-0.373**
DAYS EM											0.538**

* and ** denotes statistical significance at the 0.01 and the 0.001 levels, respectively.

MOISTURE = Seed moisture; WT/VOL = Weight per volume; HD DIAM = Seedhead diameter; HDS PLOT = Number of harvested seedheads per plot; LODGING = Lodging, %; HEIGHT = Plant height; SEED WT = Weight of 200 seed; ACTV LVS = Number of active leaves per plant on 30 October; DAYS EF = Days from emergence to flowe.ing; DAYS FM = Days from flowering to maturity; DAYS EM = Days from emergence to maturity; and YIELD = Yield of achenes.

also highly correlated with all the other factors. The path coefficient analysis revealed that plant height had a direct effect only to the extent of a coefficient of *0.291*. Every other indirect factor contributed more to the "r" value of 0.593 except the number of days to flowering and to maturity. The partitioning of the "r" value into its components yielded the following figures, as seen in Table 2 for height:

0.0877 + 0.0653 + *0.291* + 0.0563 + 0.1776 + 0.0237 + 0.0108 + 0.1424 = 0.593

The greatest indirect effect was for Active Leaves (0.1776), followed by Days from Emergence to Maturity (0.1424), then Weight per Volume (0.0877), then Head Diameter (0.0653), then Weight of 200 seed (0.0563). These indirect effects might be explained thusly: plant height is not a yield component, but it has a strong effect on total possible leaf number that the plants tend to form in putting on a large number over a long period from emergence to maturity on the taller hybrids. These taller entries would contain more leaves that would be photosynthetically active even with the progress of a leaf disease and would allow the plants to mature good yields of plump, heavy seed even though the leaf disease was advancing up the plant from the soil level toward the flower.

4. Seed Weight per 200 achenes was highly correlated with yield (r= 0.424**). The path-coefficient analysis revealed that a value of *0.2416* was the direct effect. The highest coefficient contributing as an indirect effect was Active Leaves with a value of 0.0715. This indicates that a higher number of green leaves near maturity contributed toward producing a high yield of heavy seed.

5. The direct effect of Active Leaves contributing to a highly significant correlation coefficent with yield (r= 0.589**) was *0.3031*. The greatest indirect effect contributing was Days from Emergence to Maturity with a value of 0.0963. Other indirect effects were practically negligible. This can be atributable to the fact that there were a higher number of active leaves on long season hybrids which resisted disease longer as the disease worked its way up the plant as time progressed.

6. Days from Emergence to Flower had little effect in itself on the very high correlation between that character and yield (r= 0.534**) with a value of *0.325*. Contributing well as an indirect effect was Days from Emergence to Maturity (O.1693) and the Test Weight per Bushel (0.0937) and the number of Active Leaves (0.1655). The

TABLE 2

Path coefficient analysis of the direct and indirect effects of the agronomic characteristics affecting yield in a fall planted oilseed sunflower test at the Green Acres Farm, Gainesville, Florida, 1979

		WT/VOL (1)	HEAD DIAM (2)	HEIGHT (3)	SEED WT (4)	ACTV LVS (5)	DAYS EF (6)	DAYS FM (7)	DAYS EM (8)
WT/VOL	(1)	*0.1646*	0.0240	0.0877	0.0202	0.0247	0.0937	−0.0779	0.0810
HD DIAM	(2)	0.0520	*0.2040*	0.0653	0.0608	0.0641	0.0492	−0.0516	0.0308
HEIGHT	(3)	0.0155	0.0093	*0.0291*	0.0068	0.0171	0.0213	−0.0159	0.0205
SEED WT	(4)	0.0297	0.0720	0.0563	*0.2416*	0.0570	0.0196	−0.0106	0.0232
ACT LVS	(5)	0.0455	0.0952	0.1776	0.0715	*0.3031*	0.1655	−0.1367	0.1440
DAYS EF	(6)	0.0185	0.0078	0.0237	0.0026	0.0177	*0.325*	0.0279	0.0271
DAYS FM	(7)	0.0094	0.0050	0.0108	0.0009	0.0089	−0.0170	*−0.0198*	0.0086
DAYS EM	(8)	0.998	0.0306	0.1424	0.0195	0.0963	0.1693	−0.0884	*0.2028*
Correlation with yield		0.435	0.466	0.593	0.424	0.589	0.534	-0.373	0.538

Note: Values underlined denote direct effects. Residual factor= 0.6420. The sum of the direct an indirect factors for a given parameter equals the correlation coefficient with yield.
Example: Column 1 adds to 0.435, the "r" value for WT/BUSH vs. Yield.
Please also refer to the footnotes to Table 20, this report.

days from emergence to flower are most of the days from emergence to maturity, the difference being only about 30 days. The longer the time from emergence to flower the longer the time the plants have to add active leaves, which in turn, gives more photosynthetically active leaf surface, usually higher on the plant above the diseased portion of the plant, thereby giving a higher yield of plump, heavy achenes.

7. Yield was highly correlated with the number of days from Flowering to Maturity, and was negative: r= -0.373**. The direct effect of the number of days from flowering to maturity was very low: *-0.0198*. The most important indirect effects contributing to the correlation were the Number of Active Leaves (-0.1367) and the number of Days From Emergence to Maturity (0-0884). The only positive indirect effect was that from Days from Emergence to Flower with a value of only an insignificant 0.0279. These data would suggest that the shorter time from flowering to maturity permitted the plants toquickly mature seed before the leaf and stem disease could do major damage to the plants. This would also hold true that the quicker the plants completed their life cycles from emergence to maturity the less disease they might encounter and the less leaves they would possibly add to their canopy.

8. The number of Days from Emergence to Maturity was highly correlated with yield (r= 0.538**). The direct effect of this charac-

ter was *0.2028*. Important indirect effects were Number of Active Leaves (0.1440) and test Weight per volume (0.0810). These data are very similar to those found in (6) above in the correlation for Days from Emergence to Flower. This should not be surprising as the two are complementary in that the number of days to maturity consists of all the days to flower plus the days from flower to physiologic maturity. More active leaves are added on long season plants which give more photosynthetic surface to produce heavy, plump seed yields of high quality. The more the number of leaves, especially those high on the plant, the more surface to photosynthesize even with disease incidence which worhs its way up the plant from those lowest leaves after flowering.

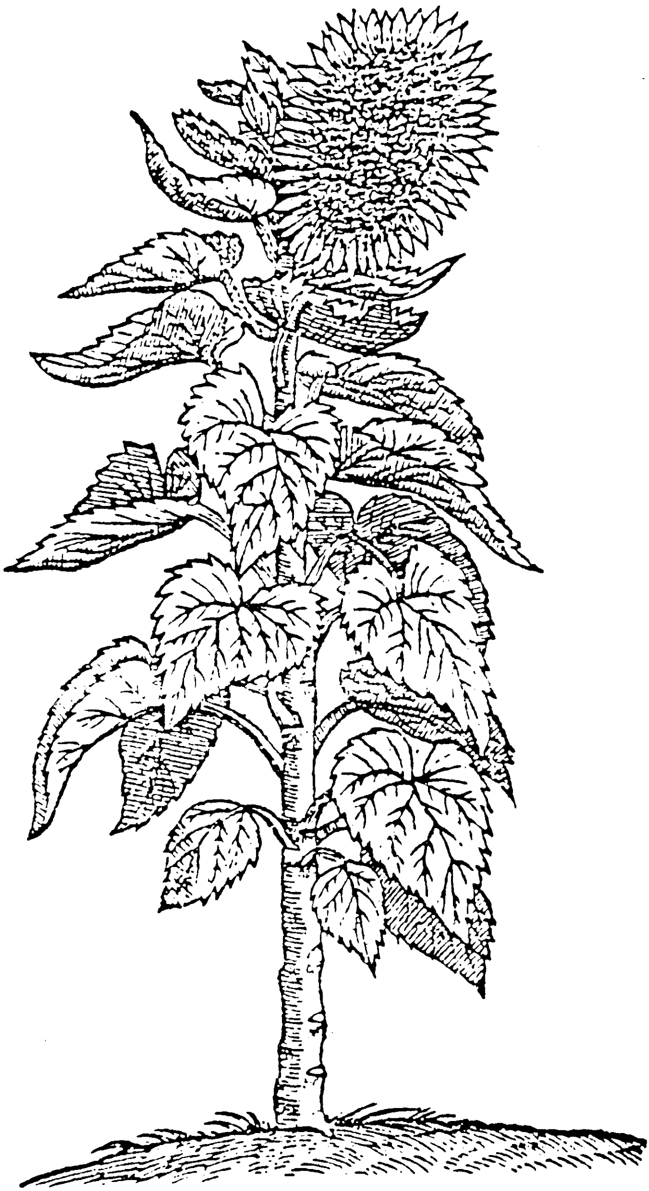
Another way of showing the contributions of the direct and indirect effects of the yield parameters, since they add arithmetically to the correlation coefficient, is to reduce them to percentages. Although I have never seen this done in the literature, I am presenting this method in Table 3, which please see. It is hoped that this table Will aid those persons who think better in percentages than in coefficients. I find myself in the former group.

TABLE 3

The percentage contribution of the direct and indirect effects of the agronomic characteristics effecting yield in a fall planted oilseed sunflower test at the Green Acres Farm, Gainesville, FL, 1979.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1)	37.8	9.0	14.8	4.8	4.2	17.5	20.9	15.1
(2)	12.0	43.8	11.0	14.3	10.9	9.2	13.8	5.7
(3)	3.6	2.0	4.9	1.6	2.9	4.0	4.3	3.8
(4)	6.8	15.5	9.5	57.0	9.7	3.7	2.8	4.3
(5)	10.5	20.4	29.9	16.9	51.5	31.0	36.6	26.8
(6)	4.3	1.7	4.0	0.6	3.0	6.1	-7.5	5.0
(7)	2.2	1.1	1.8	0.2	1.5	-3.2	5.3	1.6
(8)	22.9	6.6	24.0	4.6	16.3	31.7	23.7	37.7

Note: The figures in a column might not add to exactly 100.0%, due to rounding
See footnotes to Tables 20 and 21 for added information.



SUNFLOWER. JOHN GERARDE, LONDON, 1633