ANNUAL PROGRESS REPORT
ON THE
SOILS-WEATHER PROJECT, 1949

by

Jay T. Wakeley

Institute of Statistics
Mimeo. Series #24 (revised)
For Limited Distribution

REPORT OF THE INSTITUTE OF STATISTICS
on the
Soils-Weather Project, 1949

Southern Coop. Group

Work in the Soils-Weather project is of two sorts. The first is advisory with respect to procedural details in so far as statistical considerations form the logical basis for decisions. The second is the statistical analysis and summary of data collected in the project. This report is presented in two major sections dealing with these aspects of the work.

I. Statistical Research

The Soils-Weather project is designed to provide data from which the effects of soil and weather variables on composition of the plant can be estimated in terms of regression coefficients. The actual choice of procedure must depend in part on the expected consequence on precision of the regression estimates of soil and weather effects to be obtained from the data. As indicated in the report for 1948 many questions of procedure require statistical research as a basis for strengthening the decisions as to the use of one procedure or another, and to keep the cooperating workers informed as to what can be expected from the type of data being collected. This phase of the research is theoretical in nature and consists of identification of the factors which affect the precision of regression estimates and the quantitative effect of each on precision. Though this is not a new field of statistical investigation, specific problems of the Soils-Weather project raise questions not answered by the literature in directly applicable form.

A. Factors affecting precision.

The important factors affecting precision of regression coefficients are

1. The magnitude of error in measurement of the dependent variable (nutritional components of the plant) relative to the real variation present. This error consists of errors in chemical determination and error in sampling the plant population.

2. The magnitude of error in measurement of the independent variables (soil and weather variables).

3. The magnitude of real variation in the independent variable that is uncorrelated with the other independent variables to be involved in the regression equation. For greatest precision in the estimation of regression coefficients variation in one variable independent of variation in the others should be large; that is, correlation between them should be low.

4. The number of independent observations in the data.

Each of these factors was discussed in the 1948 report and is listed here merely to set the stage for presenting a brief resumé of past results and for presenting the results of some new research.

B. Results of research on the above factors.

Estimates of the magnitude of error in measurement of the dependent variable (nutritional components of the plant) relative to the real variation present were obtained for ascorbic acid, riboflavin and thiamine, since the project workers were more in doubt over ability to control errors in measuring vitamins than in the case of other nutrients. The results presented in the 1948 report of the Soils-Weather project can be summarized briefly as follows:

1. Loss of precision due to measurement error will be of minor importance if the magnitude of real variation is as large relative to error in measurement as indicated by the error variance of day means within season and station in the case of vitamins A, B_1 and B_2 in the referee sample work.

2. Regression estimates based on differences from season to season or from year to year derive the major portion of their error variance from among chemical determinations separated widely in time. If this can be kept as low as was estimated for station variance of the referee work referred to above, it will not be an important factor in precision of regression estimates.

3. Estimates of variance among field samples, including within day variance of chemical determinations, indicated that samples of five whole plants produced a much larger variance than did "just unfurled" leaves. Therefore, since other considerations specified that whole plants should be sampled, it was decided that samples should contain a minimum of ten plants. This should mean a reduction of about 50% in variance among samples collected the same day.

Equally important, and no so often overlooked as the relative error in measurement of the independent variable, is the magnitude of error in measurement of the independent variable (soil and weather variables). This has been greatly reduced by the use of precision instruments in the Soils-Weather project though it cannot be eliminated. It would be pertinent, then, to know the manner in which error in the independent variable affects precision of regression estimates. The two tables presented below should prove useful in better understanding the effect of errors of measurement in the independent variable. Suppose, on the one hand, that the independent variable (perhaps the amount of radiation received by the plants) could be measured entirely without error. Suppose, on the other hand, that even with the finest precision instruments available the independent variable can only be measured such that the ratio of the error variance in measurement of the variable to the real variance of the variable is 1 to 5. It is seen from Table 1 that the efficiency of the independent variable with error as compared with the independent variable perfectly measured varies from 98 percent to 40 percent depending upon how much of the variation in the dependent variable is "truly" associated with variation in the independent variable.

Table 1. Relative Efficiency of the Independent Variable <u>with</u> <u>Error</u> to the Independen
Variable <u>without</u> <u>Error</u> for Estimating the Dependent Variable.

Error Variance		"True" Reduction Due to Regression in Percent								
True Variance		10	20	30	40	50	60	70	80	90
2	(66)*	93	86	78	69	60	50	39	27	14
1	(50)	95	89	82	75	67	57	46	33	18
0.50	(33)	96	92	87	82	75	66	56	43	25
0.33	(25)	97	94	90	86	80	73	63	50	31
0.25	(20)	98	95	92	88	83	77	68	56	36
0.20	(17)	98	96	94	90	86	81	72	60	40
0.10	(9)	99	98	96	94	92	88	83	74	55
0.05	(5)	99	99	98	97	95	93	90	84	69

*Figures in parentheses are precent of possible reduction lost by presence of errors in the independent variable.

Table 2 presents the same information in terms of the affect on the error of estimate. It will be noticed that the error of estimate is always increased by the presence of errors of measurement of the independent variable. This has the ultimate affect of increasing the standard error of the regression estimate thereby lowering precision.

Table 2. Percentage Increase in the Error of Estimate Due to Errors of Measurement of the Independent Variable.

Error Variance / True Variance	True Reduction Due to Regression in Percent								
	10	20	30	40	50	60	70	80	90
2	7	17	29	45	67	100	156	268	600
1	6	12	21	33	50	75	117	200	450
0.5	4	8	14	22	34	50	78	134	300
0.33	3	6	11	17	25	38	58	100	225
0.25	2	5	9	13	20	30	47	80	180
0.20	2	4	7	11	17	23	39	66	149
0.10	1	2	4	6	9	14	21	36	81
0.05	.06	1	2	3	5	7	13	19	44

Since there are no estimates at this time of the relative error in any of the independent variables being studied in the Soils-Weather project, it is difficult to draw any specific conclusions on the manner in which the variables are now being measured. It is clear, however, that the final success of the project depends to a large extent on the accuracy with which observations are taken. Greater care in the measurement of certain variables such as soil moisture and soil temperature, observation of which is known to be more inaccurate than most others, will certainly yield benefits to the project.

The last factor affecting precision of the estimate of a regression coefficient is the number of independent observations upon which the estimate is based. Quite commonly the researcher interested in estimating a regression coefficient is instructed merely to take as many observations as possible, the assumption being that the greater the number of observations the better the estimate. If the desired precision can be defined in advance, much energy and expense in the collection of the observations can be saved. There is nothing to be gained by continually estimating coefficients with greater accuracy than is necessary for the purpose of the experiment. The purpose of the Soils-Weather project is to formulate fundamental relationships between plant response (as measured by nutritive value) and factors of the environment. For this purpose, the experiment should be of such size (number of independent observations) as to yield the best possible estimates of the regression coefficients. That is, the desired regression estimates are those with the smallest possible confidence interval. Entries in Table 3 are the number of independent observations which will yield a confidence interval of size a or less (where a is expressed as a fraction of β) on the estimate of a true regression coefficient, β , ninety-five percent of the time for selected values of ϕ = (error of estimate/variance of indep. variable) written as a fraction of β^2.

Table 3. Sample Size, n, Which Will Result (P = 0.95) In a Confidence Interval of Size a(expressed as a fraction of β) for values of ϕ/β^2.

	ϕ/β^2						
$(a/\beta)^2$	10	5	2	1	.50	.25	.10
1	64	38	20	13	9	7	5
0.25	200	113	53	32	20	13	9
0.06	>1000	400	175	96	55	33	18
0.02	>1000	>1000	468	250	124	76	37

For example, assume for the moment that the total variation in ascorbic acid content of the plants is $\sigma_y^2 = 32$ and that it would be logical to expect 50 percent of this variation to be explained by regression of ascorbic acid content on daily radiation received by the plants. It is known also that the variation in the ascorbic acid content is sixteen times as great as the variation in the independent variable, radiation. From this information it follows that the true value of the regression coefficient, β , must be equal to four, though from the experimenters point of view it may or may not be equal to four depending upon the validity of his expecting 50 per cent of the variation in ascorbic acid to be associated with variation in radiation. The purpose of the experiment is to estimate the regression coefficient sufficiently well to be reasonably certain that the confidence interval will be one (i.e. $\beta = 4 \pm .5$). How many independent observations should be collected? Obtaining the values necessary to enter Table 3, we have

$$c^2 = \left(\frac{a}{\beta}\right)^2 = \left(\tfrac{1}{4}\right)^2 = 0.06$$

$$\Phi / \beta^2 = \left(\frac{16/1}{16}\right) = 1 \quad .$$

From the table it is seen that for $c^2 = 0.06$ and $\Phi / \beta^2 = 1$, 96 independent observations must be collected if a confidence interval of ± 0.5 about the estimate of the regression coefficient is to result 95 percent of the time.

For the Soils-Weather project, Table 3 emphasizes the fact that rather large numbers of independent observations must be accumulated before any very precise estimates of the regression coefficients can be obtained. Up to the present time about 70 observations from all sources are available for estimating regression coefficients on any given plant factor; an additional 70-80 will result from data now being processed, though they could not be included in this report. From the range of sample sizes shown in Table 3, it is apparent that with the number of observations available at this time only a small number of possibilities can be investigated with relatively low precision.

To summarize, information obtained from the data to date and from theoretical considerations indicates that, in general, accuracy of measuring vitamins is adequate when considered in terms of effect on precision of the regression estimates for which the data are being collected. It should be emphasized that extreme care must be exercised in measuring weather factors with high accuracy. Errors in measurement of these factors can reasonably be expected to reduce efficiency in estimation of regression coefficients 10-30 percent if care is not taken. Finally, the number of independent observations must be very large before any very precise estimates of the regression coefficients can be obtained, particularly since it is the purpose of the Soils-Weather project to formulate fundamental relationships between plant response and factors of the environment.

II. Analysis of the Data.

Analyses and results of the data collected and edited to date are given below.

A. Variations in nutritive value.

Since there is no assurance that the regression of any dependent variable on a given independent variable will be the same from station to station, and since the experimental errors must be reasonably homogeneous among stations before data from different stations can be combined, analyses of variation in nutritive value is presented here on a within station basis. Results from the nutrient data are presented by stations taken in alphebetical order, and include only data collected in the spring and fall seasons, 1948.

Georgia:

Table 4: Table of Means (dry basis).

Nutrient[1]	Season	Planting 1	Planting 2	Planting 3	Season Mean
Dry Matter	S	13.48	11.64	13.04	12.84
	F	12.16	15.08	-	13.62
Ascorbic Acid	S	1360.41	1378.07	1483.57	1399.84
	F	1436.81	1507.80	-	1472.30
Thiamine	S	1467.28	1692.00	1736.36	1617.57
	F	1956.96	1436.25	-	1699.60
Riboflavin	S	3428.45	3315.71	3515.21	3420.88
	F	3249.21	2601.30	-	2925.26
Carotene	S	60.85	67.21	63.94	63.50
	F	64.93	58.18	-	61.56

[1] Those nutrients for which two season's data are available.

Table 5. Analyses of Variance for Nutrients (Dry Basis) Grown in Spring and Fall, 1948.

Source of Variation	d.f.	Mean Squares D.M.	Mean Squares Vit. C.	Mean Squares Vit. A.	d.f.	Mean Squares Vit. B_1	Mean Squares Vit, B_2
Seasons	1	9.84	28,600	42.88	1	2,004,959	586,530
Plantings in Season	2	14.83*	69,496	174.79	2	340,841	140,284
Days in Planting and Season	39	4.58	16,569	74.73	37[1]	169,127	355,110
Field Duplicates	43	0.14	2,817	7.52	39[2]	4,383	22,654

[1] There are 33 d.f. from this source for Vit. B_1 .

[2] There are 37 d.f. from this source for Vit. B_1 .

Dry Matter: No seasonal differences were observed that were statistically significant. The second planting in the spring was significantly lower (14%) than the average of the first and third plantings (Table 1). Day to day variation in dry matter was high in all plantings (highly significant) and c.v. was 2-4 percent based on field duplicates.

Ascorbic Acid: Planting three in the spring was significantly higher than plantings one and two (.05) though the mean differences appear to be small for practical purposes. Day to day variation is of the same magnitude in spring and fall being much larger than the duplicate error. There is some indication of a positive trend in ascorbic acid content with later planting.

Thiamine: A positive trend is also evident with later planting date for this constituent in the spring data, and the latest planting was significantly (.05) higher than the first.

Riboflavin: No significant differences between seasons or plantings within seasons were observed. c.v. = 3-5%.

Carotene: Planting one of the spring season yielded a significantly lower carotene content than the second and third plantings. No trend was observed.

North Carolina:

Table 6. Table of Means (dry basis)

Nutrient[1]	Planting 1	2	3	Mean
Dry Matter	12.42	11.94	9.35	11.71
Ascorbic Acid	1,289.20	1,298.50	1,342.50	1,298.27
Riboflavin	3,233.00	3,484.00	3,850.00	3,440.42
Carotene	54.01	57.52	71.25	58.35

[1] Nutrients for the fall season only.

Table 7. Analyses of Variance for Nutrients in the Fall, 1948.

Source of Variation	d.f.	Mean Squares D.M.	Vit. B_2	Vit. A.	d.f.	Mean Squares Vit. C.
Plantings	2	13.92	560,123	41,284**	2	2,368
Days in Planting	9	5.87**	224,633*	149**	8	28,616**
Field Duplicates	12	0.57	69,646	14	11	3,777

Dry Matter: Though no significant differences were observed between plantings, it is interesting to note that there is some indication of a negative regression between dry matter content and planting date. Day to day variation is significantly large for all constituents listed.

Ascorbic Acid: The same remarks can be made for ascorbic acid except that the relationship between nutrient content and planting date is positive. That is, the later dates showed greater amounts of ascorbic acid. c.v. was 5 percent.

Riboflavin: No significant differences between planting dates occurred though the late planting yielded 19 percent more riboflavin than did the first planting. A positive trend was evident. c.v. was 8 percent.

Carotene: A highly significant difference was observed between the third and two earlier plantings; the late planting was 28 percent higher in carotene than the averag of the first and second plantings. Day to day variation was also high. c.v. was 6 percent.

Oklahoma:

Table 8. Table of Means (dry basis)

Nutrient[1]	Season Spring	Fall
Dry Matter	15.37	16.83
Ascorbic Acid	1,238.93	1,311.67
Thiamine	1,527.71	1,440.67
Riboflavin	2,344.67	2,519.17
Carotene	42.14	47.32
Fat	2.09	2.75
Ash	11.46	13.33
Fiber	7.23	5.99

[1] Those nutrients for which two season's data are available. One planting in each season.

Table 9. Analyses of Variance for Selected Nutrients in Spring and Fall, 1948.

Source of Variation	d.f.	Mean Squares Vit. C.	Fat	Ash	Fiber
Seasons	1	34,186**	2.784*	22.624*	9.989**
Days in Seasons	11	2,319	0.910**	4.652**	0.095
Field Duplicates	13	5,067	0.055	0.149	0.045

General: All nutrients listed in Table 1, except those discussed below showed highly significant (.01) day to day variation over and above the combined error of chemical determinations and field sample variation. Coefficients of variation ranged from 3-10%. No seasonal differences of notable magnitude were observed.

Ascorbic Acid: At Oklahoma the day to day variability of ascorbic acid content was noticeably less than at Texas, Georgia, or North Carolina which makes the error appear large. The fall planting yielded significantly higher (.01) ascorbic acid content than did the spring planting, though it appears of little practical importance. This again is due in part to the unusually low day to day variability. c.v. = 6%.

Fat: This is the only station that has data on seasonal differences on fat, ash and fiber at this time. The turnip greens grown in the fall were observed to have a significantly higher fat content than the spring crop; the increase was 32 percent of the spring mean value. Day to day variability was high.

Ash: Results were the same as for fat, the fall crop yielding 16 percent more ash than the spring crop.

Fiber: The spring crop, in opposition to results for fat and ash, was observed to yield a significantly (.01) higher amount of fiber on the average amounting to 21 percent of the fall mean. Day to day variability was low.

Puerto Rico:

Table 10. Table of Means (dry basis).

Nutrient[1]	Planting 1	2	3	Mean
Dry Matter	16.65	14.41	11.81	14.29
Ascorbic Acid	1149.62	1107.46	1098.22	1118.43
Riboflavin	1869.27	2088.20	2789.94	2249.14
Thiamine	1398.67	1672.91	1474.02	1515.20
Carotene	25.90	27.97	36.21	30.03

[1] Nutrients for the summer season only, 1949.

Table 11. Analyses of Variance for Nutrients in Summer, 1949.

Source of Variation	d.f.	Mean Squares D.M.	Vit. C.	Vit. A.	d.f.	Mean Squares Vit. B_2	Vit. B_1
Plantings	2	75.91*	12,509	321	2	2,490,133**	324,230
Days in Plantings	21	15.47**	120,267**	146**	20	416,355**	222,713**[2]
Field Duplicates	24	0.83	38,527	37[1]	23	33,056	34,924[3]

[1] Based on 22 d.f. [2] Based on 18 d.f. [3] Based on 21 d.f.

<u>Dry Matter</u>: As for Texas and North Carolina, there is indication that late planting result in low dry matter content; the third planting for Puerto Rico is significantl lower (.05) than planting one or planting two. Day to day variation in nutritive value of sampled plants is significantly large for all nutrients. c.v. is 6 percent

<u>Ascorbic Acid</u>: The relationship between ascorbic acid content and planting date is positive, though differences are very small. c.v. is 17 percent which is high.

<u>Carotene</u>: Even though the mean for the third planting is considerably large than for plantings one and two, no inference should be drawn because of the greater variation between field duplicates for the third planting. The c.v. of 22. percent is high.

<u>Riboflavin</u>: Differences between planting dates are highly significant (.01) amounti to 49 percent more riboflavin in plants grown late in the summer than those grown earlier. c.v. is 9 percent. Day to day variation is also large.

<u>Thiamine</u>: No trend is even suggested and mean differences between planting are not significant. c.v. is 12 percent.

Texas:

Table 12. Table of Means (dry basis)

Nutrient[1]	Season	Planting 1	Planting 2	Season Mean
Dry Matter	S	18.86	14.83	16.85
	F	12.02	12.80	12.19
Ascorbic Acid	S	1092.27	1058.38	1087.43
	F	1389.86	1415.08	1395.27
Thiamine	S	2062.57	1806.37	2023.15
	F	1690.70	1582.75	1654.72
Riboflavin	S	3909.23	4747.87	4038.25
	F	3038.62	2661.92	2913.06
Carotene	S	42.62	48.80	43.57
	F	55.90	52.96	54.92

[1] Those nutrients for which two season's data are available.

Table 13. Analyses of Variance for Nutrients (dry basis) Grown in Spring and Fall, 1948.

Source of Variation	d.f.	Mean Squares D.M.	Mean Squares Vit. C.	d.f.	Mean Squares Vit. B_1	Mean Squares Vit. B_2	Vit.
Seasons	1	1039.56	2,653,421**	1	2,887,600	26,932,603	2,738.
Plantings in Season	2	58.32*	6,938	2	268,769	2,948,138	163.
Days in Planting and Season	52	12.13	86,421	40	238,562	1,192,453	96.
Field Duplicates	55	0.34	4,296	43	18,850	165,628	5.

<u>Dry Matter</u>: There were no significant differences between spring and fall seasons, though the means for spring and fall were 16.85 and 12.19 respectively. The significance of plantings in seasons (Table 2) was due entirely to the difference observed between the two spring plantings, the first planting being 27 percent higher in dry matter than the second planting. coefficient of variation was 3-4 percent.

Ascorbic Acid: Though no significant difference was observed between plantings in season, a highly significant difference between seasons was observed which was consistent in direction and magnitude; the fall season averaged 13 percent higher in ascorbic acid content, and plantings within seasons were very nearly the same. As for dry matter, day to day variation within seasons was significantly (.01) large. c.v. of 5-6 percent.

Thiamine: Differences between days in planting were highly significant in spring and fall. No planting or seasonal differences were observed which were statistically significant, though spring plantings had 22 percent higher thiamine content on the average (Table 1).

Riboflavin: There was quite a marked difference between the riboflavin content of turnip greens grown in the spring and in the fall, the spring crop including both plantings being 38 percent higher on the average; this was not statistically significant though of practical interest. The data were somewhat more variable with c.v. of 10-11 percent.

Carotene: Days in planting were highly significant for both the spring and fall seasons indicating the more marked influence of short-time environmental changes. Though non-significant statistically for few degrees of freedom, the fall season averaged 26 percent higher in carotene content, due primarily to the lower observed content of carotene in the first spring planting (Table 1).

Table 14. Table of Means.

Nutrient[1]	Planting 1	Planting 2
Nitrogen	4.94	6.00
Calcium	1.605	2.611
Phosphorus	0.612	0.631
Magnesium	0.224	0.328
Iron	24.30	24.74
Fat	3.46	4.36
Ash	16.71	20.41
Fiber	7.91	7.65

[1] Nutrients for the spring season only.

Table 15. Analyses of Variance[1] for Nutrients in the Spring, 1948

Source of Variation	d.f.	Mean Squares Ca.	Mean Squares Mg.	d.f.	Mean Squares N
Plantings	1	6.0798**	0.06521	1	7.7532**
Days in Planting	14	0.0511**	0.00088	26	1.0411**
Field Duplicates	16	0.0065	0.00007	28	0.0204

[1] Analyses of variance for P, Fe, Fat, Ash, and Fiber have been omitted because the results were not of particular note. Day to day variation was high in all of these constituents. The c.v.'s ranged from 4 to 11 percent.

Calcium: Calcium content of turnip greens showed a highly significant increase from the first to second planting amounting to 63 percent. Day to day variation was also highly significant. The c.v. was 4 percent.

Nitrogen: Same as for calcium with the increase in the second planting amounting to 21 percent. c.v. was 3 percent.

Magnesium: Same as for calcium with the increase in the second planting amounting to 46 percent. c.v. was 3 percent.

Ash: This constituent also showed a fairly marked increase from the first to second planting; it was not statistically significant due, in part, to the unusually high day to day variation. c.v. was 7 percent.

In summary for all stations, in general, real variation from sampling day to sampling day was considerably above the combined errors of chemical determination and field sampling. This combined error was unusually high, however, for the Texas station in their riboflavin determinations and for the Puerto Rico station in thiamine, carotene, and ascorbic acid. Significant differences between plantings were much less in prominence, the exceptions being for dry matter, carotene and riboflavin contents. Seasonal differences were generally non-significant with notable exception to this rule being observed for ascorbic acid at the Texas and Oklahoma stations. Oklahoma also showed large seasonal differences in the fat, ash and fiber contents of the plants grown under field conditions.

B. Weather effects.

In the first steps taken to date in the estimation of regression coefficients from data collected in the Soils-Weather project, multiple linear regression models have been used. This must appear rather naive to the plant physiologist, so one should hasten to add that the step is a preliminary one dictated by the amount of data compiled and by the belief that much can be learned of the general nature of the real realtionships. Assisted by this general knowledge, more realistic models descriptive of the relationship between response and environment may be formulated and tested by the experimental data as it accumulates. At this time regression studies involve only two constituents, ascorbic acid and riboflavin, since it is generally held that these may be more sensitive to changes in weather.

Tables 16 and 17 summarize regression results for ascorbic acid and riboflavin respectively for 15 different linear regression equations involving both soil and weather factors. Regression estimates for Puerto Rico have been omitted because their first season was in the summer of 1949 and the data have not been completely processed. The values within the tables are partial regression coefficients which may be used for comparative purposes. Each line in the table represents a regression equation. Column 2 designates the season for which data was used (spring 1; fall 2). Column 3 designates the period of time immediately preceding sampling for which the environmental factor has been summarized. The remaining columns designate the factors themselves. If a cell in the table is blank, it means that the factor heading the column in which that cell is found is not included as one of the independent variables. The last three columns of both tables give the percent reduction due to regression, a measure of the power of the experiment for estimating the regression relationship, and the same size.

Table 16. Regression Results for Ascorbic Acid, Dry Basis; 1948.

Station	Season (1)	Weather Summary Period	Days from Planting	Mean Air Temp.			Precip.	Evap.	Wind Flow	Rel. Humidity		Radiation		Soil: 2" depth		Soil Moist. 8"	Soil Temp. 8"	% Reduction Due Regression	% Reduction Attainable (3) by Chance (.05)	Sample Size
				Range	14-20	20-08				08-14	14-20	Before 0800	0800-0800	Max. Temp.	Range					
N.C.	2	2 day	N.S. -3.19	N.S. -4.62						N.S. -0.86		N.S. -14.22	N.S. 0.42					N.S. 68	33	11
	2	5 day	N.S. -1.95	N.S. 22.68						N.S. 2.01		N.S. 9.32	N.S. -0.51					N.S. 71	75	11
Okla.	1,2	2 day			N.S. 2.83	* -2.83			N.S. -0.11		* -0.87							** 73	41	26
	1,2	2 day			* 3.15	* -4.36			N.S. -0.16									** 60	41	26
Texas	2	2 day	N.S. 7.37		N.S. 4.37						N.S. 1.17		N.S. 0.52	N.S. -3.23	* 22.89			** 58	42	28
	2	2 day	N.S. 6.34		N.S. 2.63				* 1.12		N.S. -4.08		N.S. 0.12					** 55	38	28
	2	5 day	N.S. 7.32		N.S. -1.25				N.S. 0.16		N.S. 0.80		N.S. 0.42					** 57	39	28
Ga.	1	2 day	N.S. -0.06		N.S. -3.96			* 824.04	N.S. -1.56				N.S. 0.02					N.S. 40	43	24
	1	5 day	N.S. -6.32		* -5.23			* 725.26	* -1.49				N.S. -0.17					N.S. 37	43	24
	2	2 day	N.S. 1.51		** -4.37			N.S. 218.68	N.S. 0.02				N.S. 0.02					* 43	38	28
	2	5 day	N.S. 5.51		N.S. -1.02			N.S. -26.31	N.S. 0.03				N.S. 0.19					* 38	38	28
	1,2	2 day	N.S. -0.55					N.S. 362.24					N.S. 0.04	N.S. 5.48	*** -7.80			** 52	33	32
St'd Pots(2)	1,2	2 day	N.S. 1.11		N.S. -5.13	N.S. 2.59					N.S. 3.72		N.S. 0.50					N.S. 29	59	15

(1) Spring, 1; Fall, 2.

(2) Three stations: N. C., Texas, Ga.

(3) Measure of the power of the experiment; a low value means greater power.

Table 17. Regression Results for Riboflavin, Dry Basis, 1948.

Station	Season (1)	Weather Summary Period	Days from Planting	Mean Air Temp.			Precip.	Evap.	Wind Flow	Rel. Humidity		Radiation		Soil: 2" depth		Soil Moist. 8"	Soil Temp. 8"	% Reduction Due Regression	% Reduction Attainable (3) by Chance (.05)	Sample Size
				Range	14-20	20-08				08-14	14-20	Before 0800	0800-0800	Max. Temp.	Range					
N.C.	2	2 day	* -23.36	N.S. -39.20						N.S. 1.44		* -63.28	N.S. 1.90					** 86	78	12
	2	5 day	* -21.65	** 89.91						N.S. 0.52		* 47.93	** -3.99					** 92	78	12
Okla.	1,2	2 day	* -32.49		N.S. -25.12	N.S. 17.41			N.S. -1.03									N.S. 36	42	25
	1,2	2 day			N.S. 21.66	N.S. -25.69			** -0.69									N.S. 11	42	25
Texas	2	2 day	N.S. 2.46		N.S. 25.46						N.S. 21.35		N.S. 0.20	* -46.92	* 57.77			* 63	63	18
	2	2 day	N.S. 0.46		N.S. 7.95				* 3.20		N.S. 3.30		N.S. -1.36					* 63	56	18
	2	5 day	N.S. -6.69		N.S. -0.74				N.S. -0.76		N.S. -6.34		N.S. -0.55					N.S. 21	56	18
Georgia	1	2 day	N.S. 50.23		N.S. -0.31			N.S. -0.84	N.S. 9.94				N.S. 1.12					N.S. 23	45	23
	1	5 day	N.S. -21.16		N.S. -3.01			N.S. 1.14	N.S. 1.03				N.S. -0.78					N.S. 13	45	23
	2	2 day	** 19.23		* 10.48			N.S. -171.18	N.S. 0.18				N.S. 0.29					** 72	39	27
	2	5 day	* -21.13		N.S. 2.78			N.S. 652.87	N.S. -0.46				N.S. -0.51					** 70	39	27
	1,2	2 day	N.S. -0.18					N.S. 707.33					N.S. -0.53			*** -142.24	*** 25.21	** 74	33	32
St'd Pots(2)	1,2	2 day	N.S. 6.00		N.S. 0.40						N.S. -1.99		N.S. -0.75					N.S. 38	80	10

(1) Spring, 1; Fall 2.
(2) Three stations: N.C., Texas, Ga.
(3) Measure of the power of the experiment; a low value means greater power.

There are several specific results and general consistencies evident from these summary tables which merit emphasis. Consider the summary for ascorbic acid.

1. In eight instances the variables used in the regression equation explained a greater portion of the variation in ascorbic acid content than could be expected by chance.

2. Most notable are the results for the Texas station in which the reductions, as a percent of the variation in ascorbic acid content observed from sampling to sampling, are 55, 57 and 58 percent, the result for Georgia where 52 percent is explained by measured environmental factors, and Oklahoma where 60 and 73 percent is explained.

3. In the order as given above, the independent variables most responsible for the reductions (as tested by Student's t) are range of soil temperature at the 2 inch depth, daily radiation and number of days from planting which adjusts for age effect, mean air temperature and relative humidity. The other variables contribute little or nothing. For the 55 percent reduction observed at Texas, relative humidity, wind flow and days from planting contribute most. In the third regression for Texas, while none of the independent variables were significant when tested individually, the highly significant reduction of 57 percent is due primarily to daily radiation and days from planting. The 52 percent reduction for Georgia is due largely to range of soil temperature at the two inch depth and evaporation rate. The large reductions for Oklahoma were due primarily to mean air temperature and relative humidity, 14-20.

4. To obtain proper perspective for judging these results, it must be noticed that the sample sizes are quite small. For example, in the first regression for Texas a 42 percent reduction due to regression would be obtained one time out of twenty by chance alone (next to last column) with a sample of 28 observations (last column). All other percentage reductions given in the table should be judged in relation to the percent attainable by chance for the sample size used.

5. Though the consistency or inconsistency of regression estimates from station to station cannot be emphasized at this stage, it is interesting to note that some consistency exists. The estimated regression coefficients for daily radiation (08-08) are very nearly the same magnitude for all stations and the majority of them indicate the relationship to be positive in slope.

Now consider the summary for riboflavin given in Table 17.

1. In seven instances the variables used in the regression equation accounted for a greater portion of the day to day variation in riboflavin content than could be expected by chance.

2. The most notable results are given by the regression involving a 5 day weather summary for North Carolina, and the last three regression equations for the Georgia station. The 92 percent reduction for North Carolina was due to days from planting, range of mean air temperature, radiation for early morning hours just prior to sampling, and daily radiation. This result was unique in that each of these variables tested significant considerably above the one percent level of significance.

3. Taking the last three regressions for Georgia (72, 70 and 74 percent reductions respectively), days from planting, mean air temperature (first regression) and days from planting, mean air temperature, daily radiation (second regression) and soil moisture at the 8 inch depth, mean temperature at the 8 inch depth, and daily radiation (third regression) were primarily responsible for the rather large reductions due to regression. Here again the reduction must be considered in light of the size reduction attainable by chance one time out of twenty as given in the next to last column.

4. No consistency in the magnitude or direction (+ or -) of the regression slopes is observed for any of the independent variables with regard to riboflavin content.

Each of the regression problems given in Tables 16 and 17 can be presented also in analysis of variance form. By so doing, a station to station comparison can be made of error, and some light can be shed on the adequacy or inadequacy of the linear models used. The analyses (not actually shown here) are of the form

Source of Variation	d.f.
Total	nk - 1
Samples	n - 1
Regression	p
Deviations from Regression	n - p - 1
Within Samples	n

where

n = number of sampling days
k = number of observations per sampling day (k = 2 for field duplicates)
p = number of parameters fitted.

Consider the last two sources of variation, keeping in mind that we are interested in how well the model has been chosen and whether the duplicate observations on the dependent variables are independent. Dependence may occur if the result on the first field duplicate biases the observation on the second field duplicate.

1. If the model is correct and duplicates on the dependent variable are independent, the last two sources of variation are estimates of the same thing, random error.

2. If the model is incorrect but duplicates are independent, "deviations from regression" will be greater than "within".

3. If the model is correct and duplicates are not independent, "within" will be underestimated and "deviations from regression" will appear too large.

4. If the model is incorrect and duplicates are not independent the whole analysis breaks down. "Deviations from regression" will be too large and "within" will be too small (underestimated).

Thus, it is apparent that some basis for judging whether the "within" source (error of chemical determination plus error of field duplicates) is large or small must be reached. This can be done by comparing the "within" error on the data used in regression estimates against the duplicate error estimated from all of the data accumulated for each station. Table 18 gives the best estimates available of the "within" error by stations for riboflavin and ascorbic acid. Table 19 gives the mean squares for both "within" and "deviations from regression" for the two constituents for each regression equation listed in the same order as Tables 16 and 17.

Table 18. Estimates of "within" Error (σ_e^2) Based on all Available Data by Stations.

	d.f.	B_2	d.f.	C
North Carolina	12	69,646	11	3,777
Texas	43	165,628	55	4,296
Georgia	39	22,654	43	2,817

Table 19. Mean Squares for "Within" and "Deviations from Regression" for Riboflavin and Ascorbic Acid Obtained from the Regression Analysis.

			B_2		C	
Identification			Within	Deviations	Within	Deiviations
N.C.	(2)	2 day	69,646	70,970	3,700	14,773
	(2)	3 day	69,646	265,980	5,887	14,902
	(2)	4 day	62,210	226,174	4,473	10,112
	(2)	5 day	69,646	43,936	3,777	13,519
Texas	(2)	2 day	56,890	236,311	4,579	73,679
	(2)	2 day	56,890	221,088	4,578	74,591
	(2)	5 day	56,890	470,154	4,578	71,224
Ga.	(1)	2 day	26,995	509,701	3,000	16,146
	(1)	5 day	26,995	576,888	3,000	16,860
	(2)	2 day	11,280	119,437	2,094	13,880
	(2)	5 day	11,280	130,094	2,094	15,493
	(1,2)	2 day	23,147	144,069	3,663	13,832

Comparing the magnitude of estimates of "within" variation in Table 19 with the better estimates given in Table 18, there is some indication that

1. Texas underestimated "within" error on riboflavin though the estimates for ascorbic acid appear to be good.

2. Due to the few degrees of freedom little can be inferred from the North Carolina results particularly on riboflavin. Estimates for ascorbic acid appear to be good.

3. Georgia underestimated "within" error on data collected in the fall season. Estimates for both riboflavin and ascorbic acid appear to be good on data collected in the spring.

Wherever combined estimates of chemical determination and field duplicate error (within) appear to be reasonably good, they can be compared with "deviations" in Table 19. If the two are about the same it suggests that the linear model is not bad. If "deviations" is much larger, the inference is that the model is poor. It is indicated, as would be expected from the simplicity of the models, that

1. In general, the linear model, while sometimes accounting for large and statistically significant portions of the variation in the dependent variable, does not adequately define the relationship for the purpose of the Soils-Weather project.

2. The linear model involving certain environmental factors as independent variables does comparatively well in some cases. The model appears best for the first two regression equations for riboflavin at Texas (remembering that the error is underestimated), the first and last regression equations for riboflavin at North Carolina and the last three regression equations for ascorbic acid at Georgia (again error is underestimated).

Other comparisons might be mentioned though the relatively few degrees of freedom available for estimating both "within" and "deviations from regression" makes anything more than speculation rather hazardous even in the cases indicated for special attention.

C. Soil Effects.

The responses observed in field plantings are obviously produced by variations in soil and weather and to their interaction effects. The responses observed in pots of standard soil represent the effect of weather variation alone when comparisons are made among stations. Thus, the differences between results observed in standard soil and in the field are the reflection of soil effects and soil-weather interactions. These differences are free of any weather influence which could be considered a "main effect." It is the purpose of second field locations at each station, initiated in Spring, 1949, to obtain estimates of the effect of a greater number and types of soils. By utilizing data collected from a greater number of soils, the range of response due to soil variation can be increased thereby giving greater reliability and validity to inferences drawn. In fact, general inferences with a broad field of application cannot be made until information is obtained for a greater number of soils than is represented in the summary of data to follow.

Inspection of the results given in Table 20 under "standard soil" for spring and fall shows the variability in nutrient-content caused by variation in weather from station to station. Weather of the fall season generally produced a wider range of response for dry matter, ascorbic acid and thiamine than did the weather of the spring season (Table 21). Weather of the spring season, however, produced a wider range of results for riboflavin and carotene from station to station. The variation in nutritive value from station to station is generally as large or larger than within station variation. So that if the errors of chemical determination separated widely in time and undertaken by different workers can be kept as low as within station error of this kind, valuable additional information will be derived from comparisons among stations.

Table 20. Mean Vitamin Content of Turnip Green Grown in Standard Soil in Pots and in the Field; Spring and Fall, 1948 (dry basis).

Constituent	Station	Spring		Fall	
		Standard	Field	Standard	Field
Dry Matter (%)	Georgia	16.01	12.84	13.24	12.16
	North Carolina	-	-	34.81	11.71
	Oklahoma	16.08	15.37	17.10	16.83
	Texas	16.64	16.85	14.15	12.19
Ascorbic Acid (mg./100 gms.)	Georgia	1167.60	1399.84	1202.85	1436.81
	North Carolina	-	-	1159.00	1298.27
	Oklahoma	1252.07	1238.93	1260.17	1311.67
	Texas	1183.70	1087.43	1394.00	1395.27
Thiamine (mcg./100 gms.)	Georgia	-	1617.57	1137.80	1956.96
	North Carolina	-	-	-	-
	Oklahoma	1766.29	1527.71	1278.00	1440.67
	Texas	1605.10	2023.15	1512.33	1654.72
Riboflavin (mcg./100 gms.)	Georgia	-	3420.88	1551.65	3249.21
	North Carolina	-	-	2292.00	3440.42
	Oklahoma	2513.50	2344.67	2772.25	2519.17
	Texas	4067.80	4038.25	2630.67	2913.06
Carotene (µg./100 gms.)	Georgia	39.84	63.50	41.94	64.93
	North Carolina	-	-	34.81	58.35
	Oklahoma	28.67	42.14	41.03	47.32
	Texas	40.28	43.57	51.03	54.92

Table 21. Range of Station Means by Consitutents for Spring and Fall, 1948.

Constituent	Spring Standard	Spring Field	Fall Standard	Fall Field
Dry Matter	0.63	4.01	21.57	5.12
Ascorbic Acid	84.47	312.41	235.00	138.54
Thiamine	161.19	495.44	374.53	516.29
Riboflavin	1554.30	1693.58	1220.60	921.25
Carotene	11.61	21.36	10.00	17.61

The joint effect of soil and soil-weather interaction on nutririve value of the plants is reflected by the differences between observations on standard and in the field. Differences between mean responses in standard and field plantings are shown in Table 22.

Table 22. Differences[1] between Mean Response in Standard Soil and in the Field by Stations and Seasons, 1948.

	Spring				Fall			
Constituent	Ga.	N.C.	Okla.	Texas	Ga.	N.C.	Okla.	Texas
Dry Matter	-3.17	-	- 0.71	0.21	-1.08	-23.10	-0.27	-1.96
Ascorbic Acid	232.60	-	- 13.14	- 96.27	233.96	139.27	51.50	1.27
Thiamine	-	-	-238.58	418.05	819.16	-	162.67	142.39
Riboflavin	-	-	-168.83	- 29.55	1697.56	1148.42	-253.08	282.39
Carotene	23.66	-	13.47	3.29	22.99	23.54	6.29	3.89

[1] The negative sign indicates that the observed mean for standard soil was larger than for field planting.

In general, soil differences had less effect on vitamin content than variation in weather. The range of values for a given constituent shows that the field soils used at the various stations are sufficiently different in character to produce measureable effects in nutritive value above and beyond the direct effect of weather alone.

Regression techniques must be used in evaluation of the effect of soil variation on the nutritive value of the plants in much the same manner as has been done for weather variables. At this point in the project, however, there are an insufficient number of soils represented to make such study plausible.