Gertrude M. Cox

ANNUAL PROGRESS REPORT ON THE SOILS-WEATHER PROJECT

By

Jay T. Wakeley

Institute of Statistics
Mimeo. Series #19
For limited distribution

ANNUAL PROGRESS REPORT ON THE SOILS-WEATHER PROJECT

By

The Institute of Statistics

Work in the Soils-Weather project by the Institute of Statistics is of two sorts. The first is advisory with respect to procedural details in so far as statistical considerations form the proper basis for decisions about procedures. The second is the statistical analysis and summary of data collected in the project. This report is divided into sections dealing with these two aspects of the work.

Section I. Statistical Research.

Because of the scope of the Soils-Weather project and the many questions arising relative to procedures special studies of the statistical aspects of the project were initiated.

The Soils-Weather project is designed to provide data from which the effects of environmental variables on composition of the plant can be estimated in terms of regression coefficients. At numerous points in the conduct of the project alternative procedures could be followed. The actual choice of procedure must always stem in part from the expected consequence on precision of the regression estimates of soil and weather effects to be obtained from the data. Procedure for sampling the plant material serves to illustrate the point. The final choice was between using "just unfurled" leaves from as many plants (necessarily a fairly large number) as required to provide material for the various chemical analyses, or using all leaf material (excluding mid-ribs) from each of 5 to 10 plants. Three points were at issue: (a) For what portion of the plant is the fundamental information required? It was pointed out that the "just unfurled" leaf does not have the nutritional significance of the entire plant. (b) Which procedure lends itself to most uniform application from station to station? There is reason to doubt that the various workers would have the same concept of what a "just unfurled" leaf is. If they did not, this would be a source of composition differences between stations unrelated to differences in environment and hence a source of error in attempts to relate station differences in composition to differences in environment. (c) Which method leads to the most sampling error and what effect does the difference in sampling error have on the precision of estimation of regression coefficients.

Such questions as (c) above are involved in the statistical research. The first phase of this 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. (This is by no means a new field of statistical investigation, but the specific problems of the Soils-Weather project raises questions not answered by the literature in a form that is directly applicable.) The second phase is concerned with obtaining quantitative information on the factors which affect precision.

A. Factors affecting precision.

The important factors affecting precision of regression coefficients are

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

2. The magnitude of error in measurement of the independent variables (in this case the 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. Correlation among independent variables means that the variation in one independent of variation in the others is reduced and hence that precision of regression coefficients is reduced.

4. The number of indpendent observations in the data.

Knowledge of how these factors affect precision together with quantitative information about them in this projects forms the basis for many decisions of which some examples are given below.

1. How many samples should be collected from the field on a given day, and how many chemical determinations of a given sort should be made on each. As will be recorded below, rather good information has already been obtained relative to this question.

2. Is the agreement between workers in results of chemical analyses satisfactory?

3. In view of the relative magnitudes of variation in weather from station to station, year to year and season to season at the same station, and from week to week or day to day in a given season and station, what is likely to be the optimum distribution of effort in the project. To illustrate the point, if the variation in important weather variables within seasons at a station is small relative to that among stations, years, and seasons, effort directed at measuring plant composition frequently in each season will largely be wasted. As another example, if year to year variation in weather within stations is of a lower order than between stations in the same year, or if the correlations among weather variables are larger from year to year within stations than from station to station within years, it would be more effective, to collect data at more locations at the cost of reducing the number of years involved.

4. How much information (data) is likely to be required as a basis for concluding whether regression coefficients surpass or fall short of specified sizes and how will this vary depending on the manner in which the work is conducted.

A great deal of effort has been put into clarifying the basis for answering such questions as those listed above. A number of the basic equations relating to precision in this project have been formulated.

B. <u>Collection of quantitative information on factors affecting precision of regression estimates</u>.

Work to date has been primarily concerned with estimation of the components of variance in one group of the dependent variables, the vitamins. It is of prime importance to learn whether measurement errors are sufficiently low relative to the real variation in the nutritional components so that they will not cause serious loss of precision in estimating relationships with environmental variables. The vitamins were given first attention since the project workers have been more in doubt over ability to control errors in measuring the vitamins <u>than</u> in the cases of other nutrients.

The referee sample work was initiated by the project workers to furnish estimates of error variance inherent in the chemical determinations. Data collected in the regular course of the project and in certain special studies by the North Carolina workers have been used to obtain estimates of sampling variance and of the variance in true values of the vitamins. The important results are summarized in Tables 1, 2, and 3. With regard to error of chemical determinations three classes of variance were recognized and estimated. They were (a) variance among analyses made on the same day, (b) additional variance stemming from inability to get as comparable results on different days as on the same day, and (c) additional variance between analyses made by different workers at different stations

Table 1. Estimates of error variances of chemical determinations.

Vitamin	within day variance	day to day variance	station variance	Total
A(mg./100 gms)	.15	.31	6.70	7.16
B_1(mcg./100 gms)	682	542	4,900	6124
B_2(mcg./100 gms)	6230	16,308	23,787	46325

It should be pointed out that variances for vitamins B_1 and B_2 had turned out considerably higher in the results obtained with an earlier referee sample. However in the meantime the workers had discussed analytical problems and agreed on uniform procedures when they were at their annual meeting at Memphis. It should also be noted that the estimates in Table 1 are subject to large errors. However, they represent the best information available to date.

Table 2. Estimates of variance among field samples (within day variance of chemical determinations is also contained in these estimates).

Sampling method	Vitamin A	Vitamin B_1	Vitamin B_2	Vitamin C
"just unfurled" leaf	16.29(24)*		87,272(24)	4,068(22)
	3.99(25)	2,540(25)	30,248(25)	3,101(25)
	7.79(9)	17,597(9)	30,227(9)	2,710(9)
	8.21(27)	13,630(27)	73,274(27)	4,474(27)
Average	9.20(85)	9,670(61)	60,013(85)	3,761(83)
whole plants (5)	28.16(36)		120,631(36)	22,740(36)
	40.06(36)		135,500(36)	14,362(36)
Average	34.11(72)		128,065(72)	18,551(72)

*Figures in parentheses are degrees of freedom on which estimates are based.

In view of the larger variances for the whole plant samples (the method adopted) when they were based on 5 plants, it was decided that samples should contain a minimum of 10 plants. In view of the comparatatively low within day variances of chemical determination (see Table 1) this should mean a reduction of about 50% in variance among samples collected the same day.

Table 3. Estimates of variance in real vitamin contents from day to day within season and station.

	Vitamin			
	A	B_1	B_2	C
	21.74	43,112	229,685	8,551
	115.84		106,239	10,290
	39.08	85,939	862,635	8,579
Average	58.88	64,525	399,519	9,140
s	7.7	254	632	96

The average nutrient values of the crops for single days are based on duplicate samples. Consequently the error variance of these averages when used for within season, within station analyses is estimated as one half the within day sampling variance plus the day to day variance of chemical determinations (column 2, Table 1). Assuming 10 plant samples this becomes about

$$\frac{1}{2}\left(\frac{34}{2}\right) + .31 = 8.81$$

in the case of vitamin A. The ratio between this error variance of day means and the variance in true vitamin content, Table 3, is the key to whether measurement errors are sufficiently small. The approximate relative magnitudes of these two variances are listed below.

Vitamin	Estimated error variance of day means	Estimated variance of real vitmain content from day to day within season and station
A	1	8
B_1*	1	7
B_2	1	7
C	1**	2

*It was assumed that the ratio of variance for the two sampling methods would have been of about the same order as for the other vitamins.

**Day to day variance in chemical determination not available, hence error variance of day means may be underestimated.

If the magnitude of real variation is as large relative to error in measurement as indicated in the case of vitamins A, B_1, and B_2 the loss of precision due to measurement error will be of minor importance. The same can hardly be said with respect to ascorbic acid, however.

In the case of regression estimates based on differences from season to season, or from year to year the major source of error variance will be that among chemical determinations separated widely in time. (Sampling variance, along with day and day to day variance in chemical determinations will become unimportant because season means will be based on a number of samples collected and analyzed on different days.) If this can be kept as low as was estimated for station variance in the results of the last referee sample it will not be an important factor in precision of regression estimates. The workers seem to feel this should be, but as a precautionary measure have incorporated into the project some running checks on chemical analyses.

To summarize, information obtained from the data collected to date indicates that, in general, accuracy of measuring the vitamins is adequate when considered in terms of effect on precision of the regression estimates for which the data are being collected. The one possible exception is in the case of ascorbic acid. Further, information on this point will be accumulated as rapidly as data becomes available.

Section II. Analysis of Data

The Institute of Statistics is responsible for statistical analysis and summary of all data collected in the Soils-Weather project. This includes data on numerous supplementary studies such as reported in North Carolina Nutrition Series, Progress Reports 9-22. Analyses and results completed recently, and hence not available for inclusion in reports of cooperating workers, are given below.

A. Effect of pre-sampling.

Considering the limited numbers of plants to be grown in standard-soil pots and local-soil pots, it was necessary under the original sampling method to take leaf samples from the same plants on successive weeks during the growing season. It was believed by several cooperators that this procedure might alter the physiological balance of the plants so as to give spurious results on succeeding samples. A replicated experiment involving four treatments to test the validity of the sampling procedure was conducted at the Oklahoma station. Treatments consisted of "pulling" and discarding zero, one, three and five leaves from a number of sets of plants one week prior to sampling. Samples were taken from each set and analyzed for ascorbic acid and dry matter content. The results were extremely uniform for both constituents, there being no evidence that pre-sampling affected either ascorbic acid or dry matter content of the plants. The mean values are shown below (dry basis).

Table 4. Treatment means for vitamin C and dry matter.

Treatment	Vit. C (mg./100 gms)	Dry matter (%)
no leaves	957.9	16.1
1 leaf	916.2	15.5
3 leaves	931.9	16.4
5 leaves	983.3	15.3

B. Effect of soil and weather.

The experimental plan of the Soils-Weather project permits a within station comparison of plant response on two different soils under the same weather conditions, as well as a comparison between response of plants grown in the field and plants grown in pots containing the field soil under the same weather conditions. Notable differences in vitamin content have been observed between plants grown in local soil in pots and standard soil in pots (Table 5). In general, soil difference had less effect on vitamin content than the weather. Vitamins C, B_2, and A exhibited particular indifference to the two soil types at all stations when grown in pots.

Table 5. Mean vitamin content of turnip greens grown in local-soil in pots, standard-soil in pots and in the field. (Dry basis)

Constituent	Station	Local pots	Standard pots	Field
Vitamin A (mg./100 gms)	Georgia	40.06	39.84	64.33
	North Carolina	32.58	34.81	51.82
	Oklahoma	24.33	28.67	42.14
	Texas	35.36	40.28	37.21
Vitamin B_2 (mcg./100 gms)	North Carolina	2165	2292	3340
	Oklahoma	2251	2514	2345
	Texas	4010	4068	3232
Vitamin B_1 (mcg./100 gms.	Oklahoma	1770	2061	1699
	Texas	1481	1605	1924
Vitamin C (mg./100 gms)	Georgia	1192	1168	1369
	North Carolina	1214	1159	1307
	Oklahoma	1220	1252	1239
	Texas	1159	1184	1075

There were indication, however, that certain soil characteristics such as type of sub-soil, drainage and aeration must affect vitamin content since nutritive value of plants grown in the field differed significantly from those grown in the same soil in pots. Plants grown in the field at Georgia, North Carolina and Oklahoma showed significantly greater amounts of vitamin A. Vitamin B_2 content was significantly higher in field-grown plants at both North Carolina and Texas, while Texas and Georgia showed significantly higher vitamin B_1 and C respectively for plants grown under field conditions. With regard to mineral content of turnip greens grown in pots and in the field, it is interesting to note that where significant differences existed they were between the field and the pots (Ca, Mg and Fe); not between the two soil types confined to pots. A table of means for the Texas station are given below.

Table 6. Means

Constituent	Local pots	Standard pots	Field
Calcium	5.59	5.72	2.99
Magnesium	0.55	0.54	0.43
Phosphorus	1.12	1.18	1.15
Iron	56.68	50.61	44.74

The wisdom of planting in local-soil pots as well as standard-soil pots and the field is apparent from the fact that results from the local-soil pots are not consistently above or below results in the field (Table 5). The availability of data from local-soil pots will permit better definition of variation in response due to soil differences.

The effect of differences in weather can be seen in the results on standard-soil pots at all stations. Large differences were indicated for vitamins A, B_1, and B_2. The available data are insufficient to permit regression analysis of the effect of weather on vitamin content from station to station, though the results of several within station regressions of this form are presented in a later section of this report.

C. Diurnal Effects.

Within station data yield information on the effect of diurnal variation in weather on nutritive content. This was accomplished by taking field samples at 8 a.m., 12 noon and 4 p.m. on a number of days throughout the growing season, and analyzing them for vitamin content.

At the Texas station vitamin B_2 content was significantly lower for samples taken in the late afternoon. The same difference was shown by the Oklahoma station though it was not statistically significant.

Table 7. Means for diurnal samplings of vitamin B_2 (dry basis) and dry matter.

Constituent	Station	8 a.m.	12 noon	4 p.m.
Vit. B_2 (mcg./100 gm)	Texas	4465	4182	3982
	Oklahoma	2542	2546	2122
Dry matter (%)	Texas	16.64	19.34	19.80
	Oklahoma	14.25	15.46	16.58

Dry matter, on the other hand, showed significantly lower values at 8 a.m. at both stations. Vitamins A, B_1 and C also showed marked differences depending on the time of sampling though they were not statistically significant due to the highly significant interactions of "Time of Sampling" with "Date of Sampling" which were evident in every case.

Table 8. Means for diurnal samplings (dry basis).

Constituent	Station	8 a.m.	12 noon	4 p.m.
Vit. A (mg./100 gm)	Texas	47.16	44.56	42.86
	Oklahoma	49.48	42.62	38.45
Vit. B (mcg/100 gm)	Texas	2126	1990	1843
	Oklahoma	1442	1436	1378
Vit. C (mg./100 gm)	Texas	1106	1056	1084
	Oklahoma	1190	1243	1120

D. Regression Studies

A number of multiple regression analyses have been attempted using the data collected at Texas for study of diurnal variation. These analyses, though meager in comparison to the possibilities when more data are available, are encouraging. A multiple regression analysis of vitamin B_1 content of turnip greens on total radiation during the half day of sunlight preceding sampling and air temperature at the time of sampling indicated a significant increase in vitamin B_1 content as radiation decreased, and a significant increase in vitamin B_1 content as air temperature increased. Thus, there was a negative regression between response and radiation, and a positive regression between response and air temperature. In this case, 35 percent of the diurnal variation in vitamin B_1 content was associated with radiation and air temperature. This is a very sizeable reduction in variation considering the rather low magnitude of the total variation in the data. There was a significant increase in vitamin B_2 content as radiation decreased. In contrast to the effect on vitamin B_1, the regression on air temperature at the time of sampling was significantly negative. The regression of vitamin B_2 on these two weather factors alone was responsible for a 20 percent reduction in total variation. The regression of vitamin C content on these variables showed no significant relationships.

A number of multiple regression studies also were made on daily data from each of two cooperating stations using the total radiation recieved the day preceding sampling and the maximum temperature on the sampling day as independent variables. Undoubtedly the results obtained on vitamin B_1 were the most outstanding. There was a significant negative regression of vitamin B_1 content on radiation, and a significant positive regression on maximum temperature at the Georgia station. These two weather factors accounted for a 41 percent reduction in variation. These results are in close agreement with those obtained for vitamin B_1 at the Texas station in the multiple regression study of the data on diurnal effects. Also pertinent is the fact that the partial regression coefficients were in good agreement in the two cases. The data from Georgia showed a significant positive regression (daily data) between vitamin C content of turnip greens and total radiation recieved during the day preceding sampling, though no well defined relationship was shown between vitamin C content and the maximum temperature. This was substantiated by the results for Texas, though statistical significance was not indicated.

These results are based on a very small amount of data relative to that which will become available from the project in even one more year. They should therefore be considered only as suggestive of findings possible in the project. However, these results (particularly the consistent findings for vitamin B_1) are definitely encouraging.

From now on data adapted to regression studies will accumulate rapidly, and the major portion of the statistical work will be concerned with such studies. The more extensive data to become available will make possible more elaborate and detailed analyses than those reported above from which much better information should be obtained than has been possible to date.