USDA

U. S. DEPARTMENT OF AGRICULTURE

AGRICULTURAL MARKETING SERVICE
Technical Services Division
WASHINGTON, D.C. 20250

Memorandum

TO : Karl H. Norris, Agricultural Engineer
Agricultural Research Service

DATE : November 13, 1974

FROM : Ron Moen, Mathematical Statistician
Statistical Services Group, TSD

SUBJECT: Recommended Statistical Techniques for Evaluation of the Infrared grain analyzers

Attached is our recommended statistical techniques to use in a lab's evaluation of their testing of the infrared grain analyzers (or more generally any new test equipment that provides a quantitative measurement). Since some of the readers may have limited statistical training, I have attempted to keep it simple. The calculations can be done by hand. To recommend more sophisticated techniques would require knowledge of the reader's background and computing facilities that are available to him.

The appendix contains a discussion of some of the techniques discussed at the Chicago meeting. Many of these are quite acceptable, but, taken by themselves, may give misleading information. We think our recommended procedures minimizes this risk.

I hope this is the type of information requested at the meeting. Please give me a call after you've had a chance to read the paper.

Attachments

FORM AMS-400 (4-3-72)

STATISTICAL TECHNIQUES FOR DETERMINING THE ACCURACY AND PRECISION OF THE INFRARED GRAIN ANALYZERS

The purpose of this paper is to recommend statistical techniques to provide a standard procedure for reporting your results of experimentation with the IR grain analyzers. Other techniques may be used but this will give us some common basis for comparing studies.

The experimental design to be conducted is to make up n pairs of homogeneous samples, and then randomly assign each half of the pair to the infrared (IR) and official (OF) methods. Measurements are then taken on each of two portions from the samples. Further, we will assume that the IR method has been calibrated previously with different samples. The samples for the study should cover the full range of measurements that would normally be encountered. Classes and/or varieties of grain chosen should be identified. A layout for the data is given below.

Sample No.

		1	2		n
OF	1st portion				
	2nd portion				
IR	1st portion				
	2nd portion				

A test method is characterized by its precision and accuracy (precision must be studied both within and between labs). We will address each of these concepts separately, giving our recommended procedure. An example is given to illustrate the procedures. Finally, our comments regarding the procedures discussed at the Chicago meeting are given in the appendix.

Recommended Procedure for Precision. Precision is agreement among repeated observations made under the same conditions. Let X denote a measurement from the official method and Y a measurement from the IR instrument. The data layout is given below.

Sample No.

	1	2		j		n
OF - 1st portion	X_{11}	X_{12}		X_{1j}		
OF - 2nd portion	X_{21}	X_{22}		X_{2j}		X_{2n}
IR - 1st portion	Y_{11}	Y_{12}		Y_{13}		Y_{1n}
IR - 2nd portion	Y_{21}	Y_{22}		Y_{2j}		Y_{2n}

The first subscript denotes the first or second measurement and the second subscript denotes the sample number. Precision is studied by comparing the variation of each pair of measurements on each sample. The best measurement of this "within sample" variability for the two methods is a standard deviation. The formulae are:

$$S_{OF} = \sqrt{\frac{1}{2n}\sum_{1}^{2n} d_j^2}$$ where $d_j = X_{1j} - X_{2j}$ (standard deviation

$$S_{IR} = \sqrt{\frac{1}{2n}\sum_{1}^{2n} d'^2_j}$$ $d'_j = Y_{1j} - Y_{2j}$ for official method)

These are simplified versions of the usual standard deviation formula when there are only two observations per sample.

A statistical test for differences in the standard deviations can be performed. The test statistic $F = S^2_{IR}/S^2_{OF}$ has a F distribution with n-1 degrees of freedom in the numerator and denominator. A significant F value would imply the OF method is more precise than the IR method.

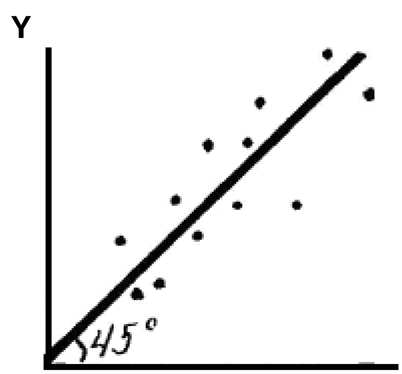
A useful measure of repeatability is "critical difference". Critical difference is the amount that will be exceeded about 5 percent of the time if measurements are made on each of two portions of a sample. Assuming normality the formula is:

Critical difference= $\pm 1.96\sqrt{2}S = \pm 2.77S$

The advantage of expressing precision this way is that the meaning can be directly related to the purpose for which the information is needed. With this measure the operator can check his technique.

Recommended Procedure for Accuracy. Accuracy is defined as closeness of an observed result from the actual or true value. Since this true value is impossible to ascertain, we can only compare the IR results to the OF results.

We recommend a simple plot of the within sample averages ($X_{.j}$ $Y_{.j}$) as a means of looking at accuracy. If the points fall on either side of a 45° line in a somewhat random manner, the IR method is accurate with respect to the official method. Such a graph would look like this:

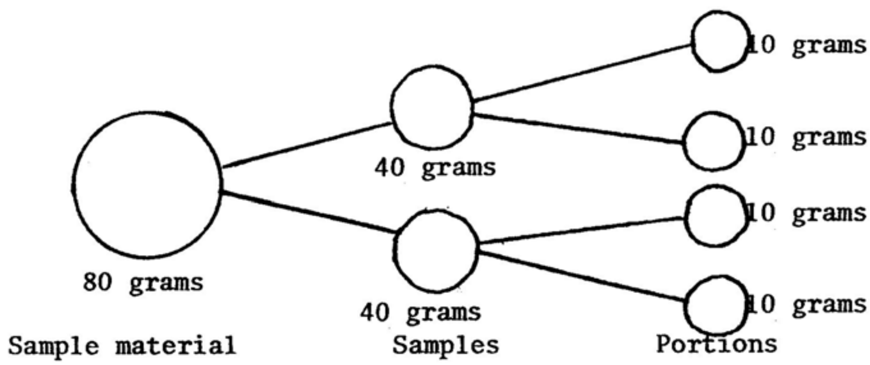


If nearly all points are above or below the line a bias exists and a calibration of the IR instrument is probably needed. How many will it take? This can be tested by counting the points above or below the line (below if there are fewer points below the line) and determining the probability of this number or less out n from binomial tables with P=q=l/2. If this probability is less than .025, a calibration is probably needed.

Example for Protein in Wheat

Material-10 D.N.S. wheat samples
11.0-16.0 protein range
10.1-10.5 moisture range

Sample Preparation



Design—Randomly assign each of the two 40 gram samples to the two methods. Run analysis on each 10 gram portion.

Sample No.

		1	2	3	4	5	6	7	8	9	10
OF	1st portion	9.9	10.0	11.3	ll.0	12.0	12.0	13.1	13.9	14.9	16.1
	2nd portion	9.8	10.2	11.0	11.2	12.2	12.3	12.9	14.0	14.9	16.2
IR	1st portion	10.2	10.5	11.7	10.7	11.6	11.9	12.8	13.9	14.7	15.8
	2nd portion	9.6	9.3	11.4	10.7	11.6	11.7	12.9	14.1	15.2	15.8

Calculations—Precision

OF d_i	.1	-.2	.3	-.2	-.2	-.3	.2	-.1	0	-.1	
D_i^2	.01	.04	.09	.04	.04	.09	.04	.01	0	$\Sigma d_i^2.$ = .37	
IR d_i'	.6	1.2	.3	0	0	.2	-.1	-.2	-.5	0	
$D_i'^2$	.36	1.44	.09	0	0	.04	.01	.04	.25	0	$\Sigma d_i'^2$ = 2.23

S_{OF}= 0.14 (standard deviation for OF method)

S_{IR}=0.33 (standard deviation for IR method)

Critical difference for two independent measurements (from 10 gram portions of a sample of 40 grams)

$\pm 1.96\ \sqrt{2}S_{OF} = \pm\ 0.39$ or OF method

$\pm 1.96\ \sqrt{2}S_{IR} = \pm\ 0.91$ for IR method

Calculations - Accuracy

Within Sample Averages

OF :9.85: 10.10: 11.15: 11.10:
12.10:12.15: 13.00: 13.95: 14.90:
16.15

IR :9.90: 9.90: 11.55: 10.70: 11.60: 11.80: 12.85: 14.00: 14.95:

15.80

Plot on attached graph.

Overall averages OF = 12.4
IR = 12.3

Conclusions--Based on this rather limited data the IR device appears unbiased and needs no calibration (probability of 4 or less out of 10 is .38) As for precision the OF method is more precise than the IR method (F = 5.56 is significant at α = .01).

APPENDIX

<u>Other Procedures Discussed at Chicago Meeting</u>

Let $d_{ij} = Y_{ij} - X_{.j}$ - Calculate

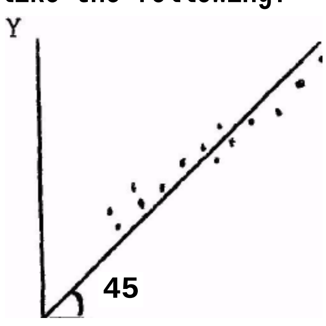
$$S_d = \sqrt{\frac{\sum_{j=1}^{n}\sum_{i=1}^{n}(d_{ij} - \bar{d})^2}{2n - 1}}$$

A large S_d would indicate one or more of the following

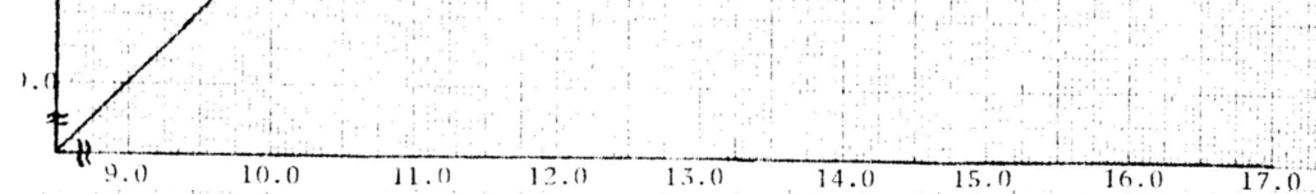
(1) the IR lacked precision
(2) the OF lacked precision
(3) lack of homogeneity in the sample material allocated to the two methods.

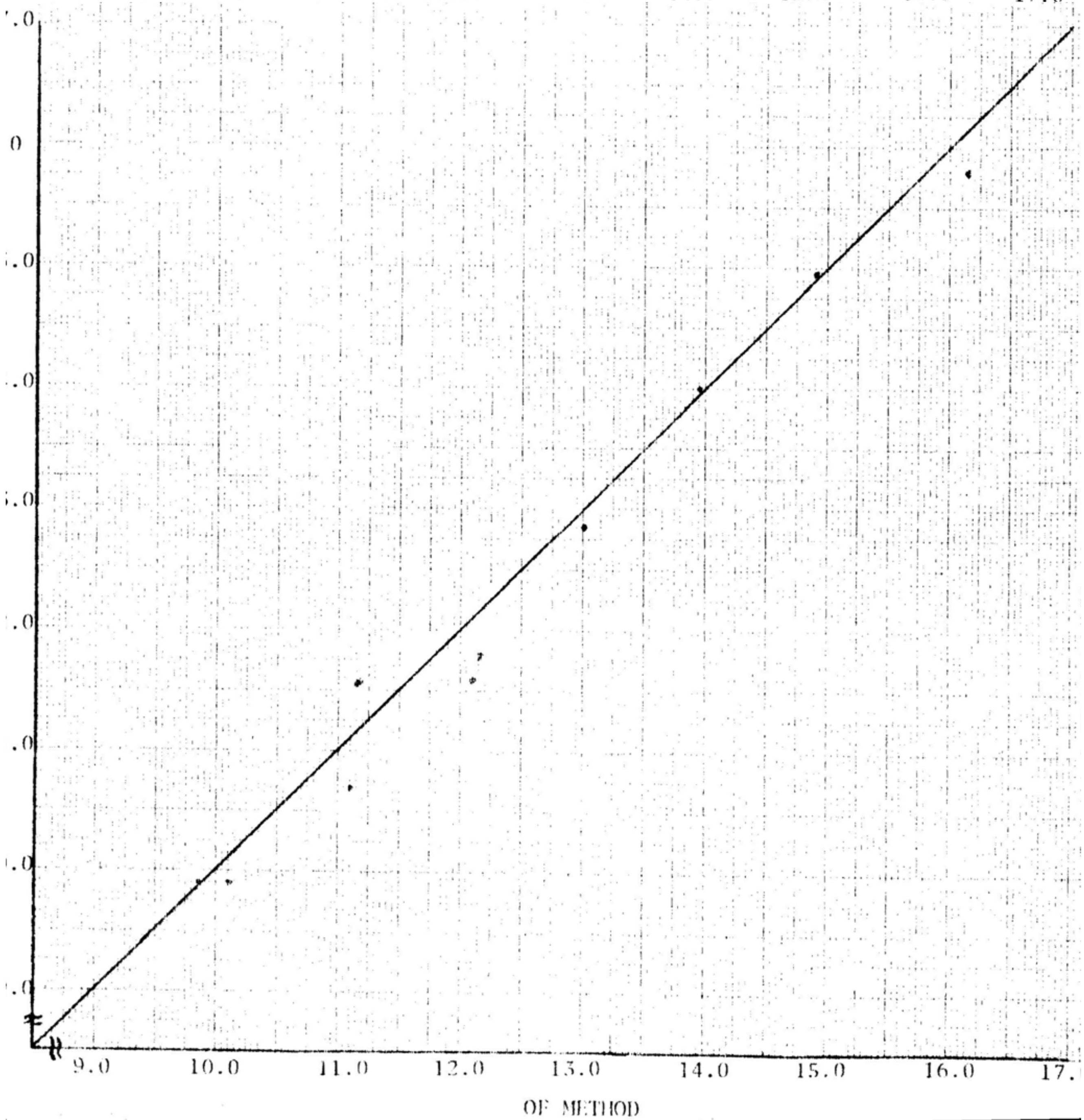
In our recommended procedure we are looking at the precision of each method separately. Lack of homogeneity in the sample material will inflate each of the standard deviations. However, we still will be able to determine *if* the IR precision is as good as the OF's by a direct comparison of the standard deviations. We do not recommend calculating S_d.

The average $\bar{d} = \sum^{n} d_j$ where $d_j = \bar{Y}_{.j} - \bar{X}_{.j}$ - can be calculated. This average difference should be close to zero. A statistical hypothesis that the true difference is zero can be tested with a paired t-test. Failure to reject this hypothesis may not imply accuracy if the plot with the 45° line looks like the following:



A simple linear regression analysis can be run on the pairs $(X_{.j}, Y_{.j})$. The resulting equation $Y = a + bX$ and data should be plotted. Ideally, the fitted line will coincide with the 45° line; that is, the slope b will equal 1 and the Y intercept a will equal 0. However, regression analysis assumes no measurement error of the independent variable. In this case that means the OF method is assumed precise. A good fit of the regression line (as measured by the standard error of estimate $S_{y.x}$) doesn't imply accuracy of the IR method if this regression line is not close to the 45° line. A high positive coefficient of correlation (r) may be misleading for the same reason.


9.0
10.0
11.0
12.0
13.0
14.0
15.0
16.0
17.0


9.0
10.0
11.0
12.0
13.0
14.0
15.0
16.0
OF METHOD