INTERPRETATION AND PRESENTATION OF RESULTS

Thomas M. Little[1]
Biometrician Emeritus, University of California, Riverside, CA 92502

The purpose of an experiment is to answer questions. The truth of this seems so obvious, that it would not be worth emphasizing were it not for the fact that the results of many experiments are interpreted and presented with little or no reference to the questions that were asked in the first place. In other cases, it appears that the wrong questions were asked.

Interpretation of dose and response

Let us look at an example. An experiment was performed in which there were 7 treatments consisting of 7 levels of a material applied to the plants. The simple, straightforward, and useful question that might have been asked by the experimenter is: What is the relation between the amount of material applied and the plant response? Then there is a much more cumbersome and less useful question: Of the 21 possible pairs of treatments, which ones are significantly different from each other?

Can it be that this second question is really the one the experimenter had in mind when the experiment was planned? At any rate, that is the question that was answered when the results were reported like this:

Treatment Level	Plant Response
0	222 a
1	202 ab
2	205 ab
3	186 bc
4	164 cd
5	156 cd
6	147 d

Suppose the simpler question of the relation between dosage and response had been asked and answered. It would have been reported that there was a highly significant linear relation that accounted for over 96% of the variation in response. This could have been illustrated graphically, as in Fig. 1.

Partitioning of treatment sum of squares
in Example 1, Variable 1

Source of variation	df	SS	Percent of total
Treatments	6	23805	
Linear	1	22885	96%
Residual	5	920	4%

The question of which treatment responses are significantly different from each other is now irrelevant, and it need not, in fact, should not, have been asked. Once a significant linear trend is established, all treatment levels within the range of those used in the experiment are significantly different from one another in their effects. The best estimates of the treatment effects are the points on the regression line.

There are, of course, other kinds of relations between dosage and response besides simple linear ones, such as various curvilinear relations.

One of the variables measured in this same experiment showed a response to treatments that was obviously not linear. (Incidentally, variables should never be referred to as "parameters".) The data were presented as follows:

Treatment level	Response
0	9 d
1	12 cd
2	15 bc
3	22 a
4	19 ab
5	17 b
6	11 cd

Interpreting these results by examination of the letters is even more confusing than in the previous example. However, partitioning the treatment effects into individual degrees of freedom shows that over 85% of the variability among responses is accounted for by the linear and quadratic components. The data can therefore be simply summarized by a simple second-degree curve (Fig. 2).

Partitioning of treatment sum of squares
in Example 1, Variable 2

Source of variation	df	SS	Percent of total
Treatments	6	650.0	
Linear	1	71.5	11%
Quadratic	1	482.0	74%
Cubic	1	40.8	6%
Residual	3	55.7	9%

Whenever the treatments consist of a series of dosage levels, an effort should be made to find some meaningful relation between dosage and response, rather than resorting to a confusing and almost meaningless multiple comparison procedure.

Factorials

Let us look at another example. This is a 2 × 2 × 2 factorial. In other words, there are 3 factors, each applied in 2 different ways with all 8 possible combinations applied. This differs from the previous example in that no trends are involved. The logical questions to ask are: What is the effect of each factor? Are there interactions, or in other words, does the response to one factor depend on the level of another factor?

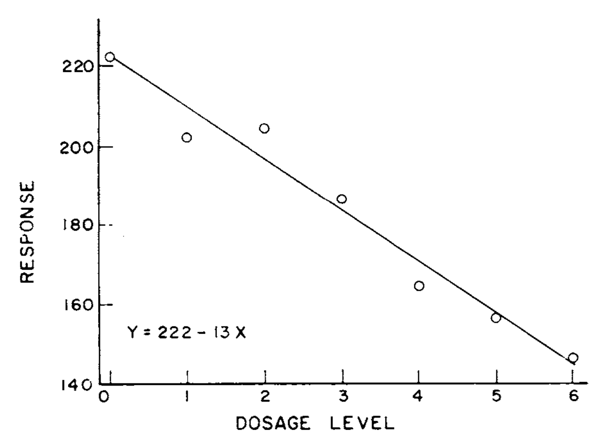

Fig. 1. Linear effect of 7 dosage rates on plant response.

[1]*Present address: 1488 Argyle Lane, Bishop, CA 93514.*

It wasn't these simple questions that were answered in the presentation of the data. Instead, the main question that was answered was: Which of the 28 possible pairs of treatment responses differed from each other significantly? Here is the way the data were presented:

Treatment combination	Response
ABC	10.4 b
ABc	8.5 b
AbC	15.2 ab
Abc	13.0 ab
aBC	21.4 a
aBc	18.6 ab
abC	22.6 a
abc	22.7 a

These letters don't tell us much, but if we partition the treatment sum of squares into main effects and interactions, we find that factor A had a highly significant effect, accounting for over 83% of the total treatment variation. Factor B accounted for another 12%, and there was no evidence of any effect of factor C or any of the interactions.

Partitioning of treatment sum of squares in Example 2

Source of variation	df	SS	Percent of total
Treatments	7	877	
Factor A	1	730	83%
Factor B	1	107	12%
Factor C	1	23	3%
A × B	1	8	
A × C	1	1	
B × C	1	3	2%
A × B × C	1	5	

The important facts to present are therefore the main effects of factor A and B:

Factor	Average effect
A	11.8
a	21.3
B	14.7
b	18.4

Actually, the means of the main effects were presented in a table, but 4 out of the 6 means were incorrect! Here are the means presented alongside the individual treatment means:

Individual treatment means		Main effects Calculated	Published
ABC 10.4	A	11.8	11.8
ABc 8.5	a	21.3	21.3
AbC 15.2			
Abc 13.0	B	14.7	17.8
aBC 21.4	b	18.4	15.3
aBc 18.6			
abC 22.6	C	17.4	20.5
abc 22.7	c	15.7	12.6

Unfortunately, mistakes of this kind are not uncommon in the pages of our journals. What is worse, they are seldom corrected in subsequent issues. Published papers are the permanent records of scientists' work, and every effort should be made to avoid presenting erroneous results. Tables and graphs should agree, and statements in the text should be borne out by the presented data.

It is all too easy to blame mistakes on a secretary, a statistical clerk, or the computer, but the ultimate responsibility for accuracy belongs to the authors. Some experimenters act as though operating a calculating machine is beneath them, and statistical analysis is a menial task that should be relegated entirely to secretaries or clerks. Their work often reflects this attitude.

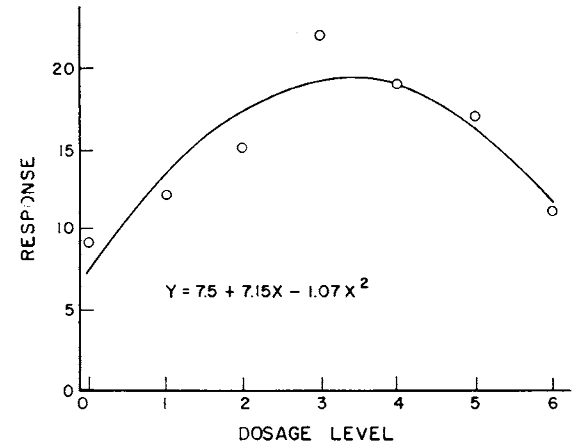

Fig. 2. Quadratic response of 1 variable to dosage level.

The next example combines both factorial and trend analysis, and illustrates how interactions can be interpreted and presented. It is a 2 × 2 × 4 factorial, with factor C consisting of 4 exposure times. Obvious questions would be: What are the effects of factors A and B? Is there a relation between response and exposure time? Are there any significant interactions among the 3 factors?

The question that was actually answered in the presentation of the data was: Within each level of factor C, which of the six pairs of treatments differed from each other significantly? Here are the results:

Level of A	Level of B	Level of C 0	1	2	3
1	1	74 c	92 c	108 b	134 b
1	2	48 b	108 d	156 c	292 d
2	1	12 a	18 a	76 a	92 a
2	2	18 a	58 b	90 ab	162 c

A partitioning of the 15 degrees of freedom for treatments tells us much more about the effects of the various factors and their interactions.

Partitioning of treatment sum of squares in Example 3

Source of variation	df	SS	Percent of total
Treatments	15	216,083	
Factor A	1	44,287	20.5%
Factor B	1	19,927	9.2%
A × B	1	817	0.4%
C linear	1	113,274	52.4%
C quadratic	1	2,977	1.4%
C cubic	1	163	0.1%
A × C linear	1	1,717	0.8%
B × C linear	1	21,094	9.8%
Residual	7	11,827	5.5%

There were only 4 comparisons that were significant, and these accounted for 92% of the variation among treatments. These were: the main effect of factor A, the main effect of factor B, the linear trend of factor C, and the interaction between factor B and the linear trend of factor C. These results can be summarized graphically (Fig. 3 and 4).

This experiment illustrates another important point. The overall F value for the treatment sum of squares based on 15 degrees of freedom is meaningless. This is because it is the average of 4 highly significant single degrees of freedom and 11 non-significant ones. The idea that one should proceed no further with an analysis, once a non-significant F-value for treatments is found, has led many experimenters to over-look important information in the interpretation of their data.

Let's look at one more example. This example consisted of 7 treatments. There was a "control" and 3 levels of a material in the ratio of 2:5:10 applied with and without an additive. It is not my assignment to criticize design, but good interpretation starts with good design. This was not a very good design. It was not a complete factorial. A treatment consisting of the additive alone would have made it so. Furthermore, there seems to be no logical justification for the particular series of treatment levels chosen. Generally, when studying the relation between treatment level and response, a series of rates in arithmetic progression is the most efficient.

Here is the way the results were presented:

Treatment	Response
0 level	81 a
2 level	77 ab
2 level + additive	74 ab
5 level	67 bc
5 level + additive	66 bc
10 level	50 d
10 level + additive	56 cd

Actually, in spite of the poor design, the experimenter was lucky and didn't know it. Partitioning the treatment effects shows that there was no evidence whatsoever of any effect of the additive, or any interaction between additive and level of material.

Partitioning of treatment sum of squares
Example 4, Initial Partitioning

Source of variation	df	SS	Percent of total
Treatments	6	3,216	
Control vs. others	1	915	28%
Linear among others	1	2,249	70%
Additives	1	0	0%
Remainder	3	52	2%

This being the case, we can disregard additives and consider the 7 treatments as consisting of 1 treatment at the zero level and 2 treatments each of 3 other levels. We can then carry out a regression analysis, and we find that linear regression accounted for over 98% of the variability among treatments. This result can be neatly summarized on a graph (Fig. 5).

Partitioning of treatment sum of squares
Example 4, Final Partitioning

Source of variation	df	SS	Percent of total
Treatments	6	3,216	
Linear	1	3,164	98%
Deviation From linear	5	52	2%

Partitioning of treatments

In each example I have given, I have mentioned the partitioning of treatment effects as though the technique for doing this were common knowledge. Unfortunately, this may not be the case. Over many years of participation in short statistical refresher courses and seminars for agricultural research workers, I have made an alarming observation. Nearly every participant knew how to calculate LSD, and in recent years, Duncan's multiple range test. Still, less than 10% knew how to partition a treatment sum of squares into meaningful comparisons. This is too bad, for the technique is so powerful and yet so simple.

I am glad to note that in the last couple of years there has been an in-

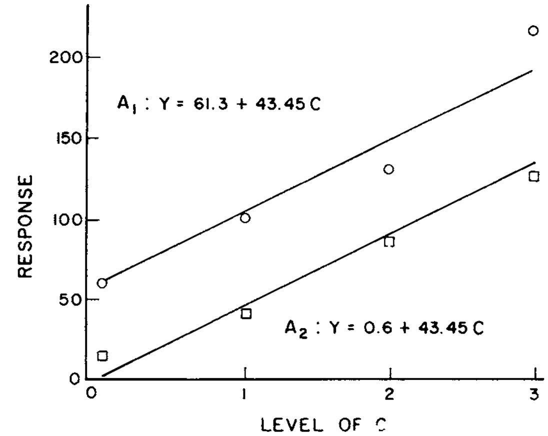

Fig. 3. Linear response to factor C at 2 levels of factor A, showing lack of interaction.

crease in the use of this technique reflected in our journals. There may be several reasons why it is not more widely used. Although it is described in nearly all statistics texts, it is usually under some formidable name such as "orthogonal comparisons", "orthogonal linear forms", or "single degrees of freedom." Furthermore, the discussions often tend to be clothed in unnecessarily complex mathematical jargon and symbolism.

Let's face it, there is probably another reason why the technique is not more widely used. People in any profession tend to copy each other, and horticulturists are no exception. "Dr. John Doe used Duncan's multiple range test in presenting his results, so that's good enough for me."

I should hasten to add that there are situations where multiple comparison procedures, such as Duncan's multiple range test, are appropriate. Such would be the case when testing a random assortment of cultivars or chemicals. Even in these cases, the investigator should ask whether the treatments fall into groups, the comparison of which would provide important information. Cultivars, for example, might be classified into those which are resistant and those which are susceptible to a certain disease, and a comparison made between the 2 groups.

Insignificant digits

For my final comments on the presentation of data, I am indebted to Dr. M. T. Vittum of Cornell for his suggestions. They deal with false accuracy in publishing results. The U.S. Coast and Geodetic Survey plaque at the summit of Mount Whitney is an extreme example. This shows the altitude to be 14,496.811 ft.! I was chided by an engineering dean for ridiculing this claim, because, he said, we biologists just didn't understand how precisely the engineers could measure things. I

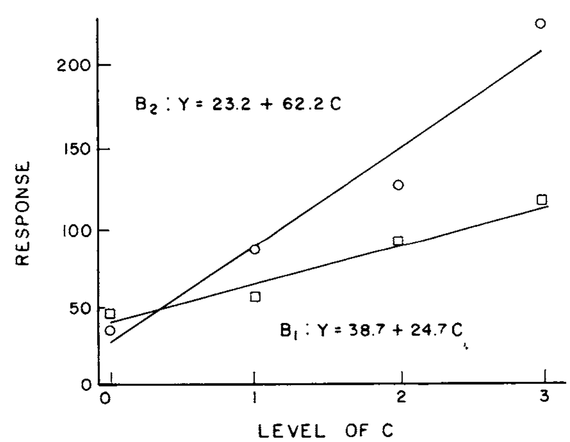

Fig. 4. Linear response to factor C at 2 levels of factor B, showing interaction.

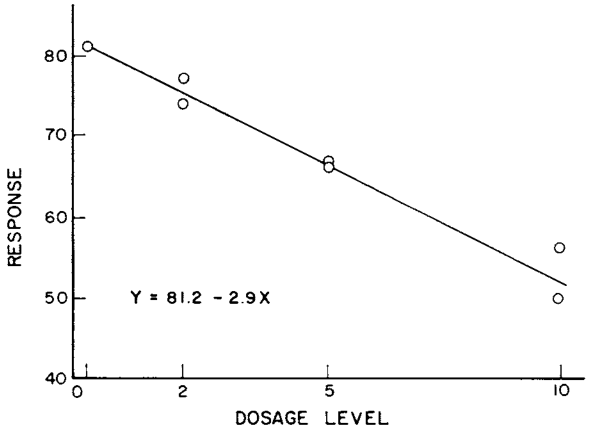

Fig. 5. Negative linear response to dosage level, disregarding non-significant treatment factor.

agreed. We biologists are not entirely blameless in this matter. While we do not generally go to the extremes in the above example, we do often pretend that our data are more accurate than they really are. In making measurements, more than 3 significant digits are almost never justified. Even in reporting means, only in cases where we are dealing with material with a very low coefficient of variability, or a very large number of replicates, are we justified in reporting 4 significant digits.

A common cause of reporting results with too many significant digits is the conversion of measurements from English to metric units. In questioning the results of one paper, I was told that part of the results had been given in pounds and part in kilograms, and if I would multiply the error mean square by $0.45359237^2 = 0.2057460381222169$, everything would come out all right! In one paper, we were told that the rate of application of spray was 1402 liters/hectare, but the author is at least to be commended for telling us that this was 150 gallons/acre. In another paper, it was stated that the turf grass was cut to a length of 0.635 cm. Evidently a mower was used which was calibrated in quarter-inch increments.

Even statisticians, who should know better, sometimes exaggerate the accuracy of their statements. For a statistician to tell an experimenter that the probability of obtaining an F value greater than that which was observed is 0.068, is inexcusable. Yet this statement was made in a recent paper.

Suggestions for improvement

In conclusion, I would make the following suggestions:

1) In planning an experiment, decide definitely what questions you want to answer, and design the experiment to answer these questions.

2) In presenting the results, tell the reader what questions the experiment was designed to answer.

3) Interpret the results as answers to the questions you asked in the beginning.

4) Don't deceive yourself or the reader with exaggerated claims of accuracy.

5) Strive to avoid mistakes and inconsistencies in the final presentation. If you take the credit for the paper, the mistakes are yours too!

Selected References

Chew, V. 1976. Uses and abuses of Duncan's multiple range test. Proc. Fla. State Hort. Soc. 89:251–253.

Chew, V. 1976. Comparing treatment means: a compendium. HortScience 11:348–357.

Little, T. M. 1978. If Galileo published in *HortScience*. HortScience 13:504–506.

Little, T. M. and F. J. Hills. 1978. Agricultural experimentation — design and analysis. Wiley, New York.

Petersen, R. G. 1977. Use and misuse of multiple comparison procedures. Agron. J. 69:205–208.

SLIDE SETS AVAILABLE
American Society for Horticultural Science

Angiosperms of North America, No NA-2
83 slides - E. Memmler - - - - $54.78
Avocado Production & packing No. 78
65 slides - B. W. Lee - - - - - $42.90
Avocado Propagation, Nursery, No. 781
55 slides - B. W. Lee - - - - - $36.30
Bananas, No. ASHS 12 - 80 slides
L. G. Wilson & E. Memmler - - -.$52.80
Beans, Green Lima, No. 3
42 slides - R. Brendler - - - - $27.50
Bristle-Cone Pine Plant Community, Set
BC-1, 32 slides - - - - - - - - $26.40
Cabbage & Lettuce, No. ASHS 20
45 slides - R. Brendler - - - - $29.70
Cacti of the Southwest, No. D-1
30 slides - E. Memmler - - - - $26.40
Careers in Horticulture, No. ASHS 13
85 slides - G. M. Kessler & Com.$56.10
Celery Production, No. ASHS 15
57 slides - R. Brendler - - - - $37.62
Citrus Propagation, Nursery, No. 711
18 slides, B. Lee & R. Burns - $13.20
Citrus Identification, No. 712,
24 slides, - B. W. Lee - - - - $15.84
Dwarf Connifer Collection, Nat., Arb.
56 slides, - E. Memmler, ASHS 5 $36.96
Extension slide sets - Francis R. Gouin
Care & Culture of House plants
80 slides, ASHS 10 - - - - - $52.80
Care & culture of Garden Roses,
80 slides, ASHS 6 - - - - - - $52.80
Growing Annual Flowers,
72 slides, ASHS 4 - - - - - - $47.52
Pruning Oranmental Trees & shrubs
80 slides, ASHS 3 - - - - - - $52.80
Flower Set Illustrating Families,
206 slides, No. FF-2, Memmler $135.96
Forestry, E. Memmler - Set F-1
60 slides, - - - - - - - - - - $39.60
Gymnosperms of North America, No. NA-1
89 slides, - E. Memmler - - - - $58.30
Herbs, Spices & Fragrances, ASHS 7
139 slides, - Elmo W. Davis - - $91.74
Indoor landscaping, (House Plants),
92 slides, ASHS 1, F. Cochran - $60.72
Kiwifruit (Chinese Gooseberry), comm.
72 slides, ASHS 26, - - - - - $47.52
Land Preparation, Vegetable Crops
46 slides, ASHS 21, R. Brendler $30.36
Landscaping Illustrated, L-1
117 slides, E. Memmler - - - - $77.22
Leaf Shapes & Nomenclature, LS-1
55 slides, E. Memmler - - - - - $48.40
Lemon Production and Care, R. Burns
75 slides, ASHS 22, - - - - - - $49.50

Lily Hybridizing Techniques, No. LH-1
32 slides, E. Memmler - - - - - $26.40
Macadamia Nut Production, No. 65
36 slides, E. Memmler - - - - - $23.10
Major Types of Plant Communities,
32 slides, No. PC-1, E. Memmler $26.40
Mushroom Production, No. 46
32 slides, E. Memmler - - - - - $20.12
Orange Production & Packing, No. 74
82 slides, R. Burns - - - - - - $54.12
Orchid Propagation, No. 428
58 slides, E. Hetherington - - $38.28
Pineapple Production, No. 49
52 slides, B. Lee - - - - - - - $34.32
Plant Propagation, Wholesale Nurs.
60 slides, No. 152, E. Memmler $39.60
Protea Propagation, No. 160
43 slides, E. Memmler - - - - - $39.60
Rose Production, No. ASHS 8 - - - $90.42
137 slides, F. Gouin & J. Williams
Strawberry fruit Production, No. 96
35 slides, B. Lee, - - - - - - $23.10
Strawberry Nursery Plant Production
62 slides, No. 961, B. Lee - - $40.92
Sugar Cane Production, ASHS 158
20 slides, John E. Bowen - - - $13.20
Tissue Culture, Nurseries, ASHS 14
92 slides, Toshio Murishige - - $60.72
Tomato Production & Packing, ASHS 9
73 slides, R. Brendler, - - - - $48.18
Turf Grass Production, No. 157
40 slides, R. Brendler, - - - - $26.40

Horticulture Is, ASHS 11, 80 slides,
25 folders, Francis Gouin - - - $55.00

These slide sets are the result of cooperative efforts of members and friends of the ASHS. For more information and to learn how you too may join in on the ASHS Slide Production Program write to Eugene Memmler.

PRICES are subject to change without notice.

Orders outside the United States:- Payment MUST accompany the order + 20% for Air Postage. Orders originating and shipped to points in the U. S. purchase orders are accepted. Orders over $20.00 are sent Post Paid.

Send **ORDERS TO: EUGENE MEMMLER,**
Payments should be made out to:
EUGENE MEMMLER
Chairman ASHS Slide Collection,
3287 Dunsmere Rd. Glendale, CA. 91206 USA