DOCUMENT RESUME

ED 085 403 TM 003 344

AUTHOR	Dyer, Henry S.; Schrader, William B.
TITLE	Manual for Analyzing Results of an Educational Experiment (Analysis of Covariance).
INSTITUTION	Educational Testing Service, Princeton, N.J.
PUB DATE	60
NOTE	38p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Analysis of Covariance; Educational Experiments; *Guides; *Instructional Materials; *Manuals; Research Design; Research Methodology; *Statistical Analysis

ABSTRACT

The need for analysis of covariance is discussed in the context of present day education. . .new approaches to instruction including new methods, curricula, devices and patterns of classroom organization. A problem is then used as an example of a typical experiment and the reader is led step by step through the application, analysis and interpretation of analysis covariance. The explanation and example used attempt to reduce the complicated procedures to a process that can be handled by anyone with good command of ordinary arithmetic and some understanding of algebra. Worksheets for computing each step of the analysis are included. (RF)

Manual for

Analyzing Results of an Educational Experiment

(Analysis of Covariance)

HENRY S. DYER

WILLIAM B. SCHRADER

1960

EDUCATIONAL TESTING SERVICE, PRINCETON, NEW JERSEY

Additional copies of this manual may be ordered from Educational Testing Service, Box 592, Princeton, New Jersey. The cost is $1.00 per copy.

Published by Educational Testing Service,
Princeton, New Jersey, 1960

FOREWORD

This manual has been written with the hope and expectation that people who are not familiar with the complex statistics of educational experimentation will be able to use it in analyzing experimental results in a sound fashion.

The preparation of this manual was originally supported by the Fund for the Advancement of Education. The following persons helped the authors in one way or another in preparing this manual: Mrs. Anne H. Ferris, Miss Henrietta Gallagher, Dr. Martin Katz, and Dr. Marjorie Olsen. Acknowledgment is also due to Dr. Warren G. Findley, who made a number of useful suggestions on the basis of his use of the experimental edition of this manual.

Henry S. Dyer

William B. Schrader

Educational Testing Service

Princeton, New Jersey, 1960

ANALYZING THE RESULTS OF AN EDUCATIONAL EXPERIMENT

(Analysis of Covariance)

Introduction

The present ferment in American education is producing many new approaches to instruction -- new methods, new curricula, new devices, new patterns of classroom organization. As a result there is an increasing urge to find valid means for evaluating these new approaches by comparing them with one another and with the conventional ways of doing things. People wish to know whether there is any measurable difference between the old and the new in the amount of learning produced in pupils. They wish to know whether the gain in performance of a group of pupils treated one way differs significantly from the gain that would have occurred if the same group had been treated in another way.

Since it is impossible to treat one group in two ways simultaneously, it is necessary to deal with two or more groups each of which is treated differently from the others. A valid comparison of the gains made by the several groups requires that allowance must be made for initial differences between the groups. The statistical technique called analysis of covariance is generally regarded as the most rigorous means for making such adjustments and furnishing soundly interpretable results.

The method of analysis of covariance is one which makes the most of the available data and provides a valid interpretation of what the outcome of the experiment means. Hitherto this method has been accessible only to those with a sophisticated understanding of statistical formulas and procedures. The present explanation attempts to reduce all such complicated procedures to a step-by-step process that can be handled by anyone with good command of ordinary arithmetic and some understanding of algebra. The method to be described is based directly upon an original paper by Gulliksen and Wilks in the June 1950 issue of Psychometrika.[1]

The explanation that follows is built around a typical experiment. The nature of the experiment is described, the data obtained from it are given in full, and the analysis of the data is worked out in complete detail. The reader who wishes to use this approach on an experiment of his own is advised first to study the data carefully and then to work out each step of the analysis himself, checking his own results at every stage against those given. Once he is sure he has mastered the procedure, he may simply substitute the data from his own experiment for those given here, and then work through the same steps in analysis.

The Problem

Three classes of a course in chemistry were taught using special TV lectures and kinescopes. Three similar classes were taught by the conventional methods. The group taught by television made an average gain of 11 points on the final test as compared to their scores on a pretest of chemical knowledge. The group taught by the conventional method made an average gain of 4 points on the same test. The experimenter needs to know the answers to three questions before he can confidently evaluate the results of his experiment. First, he wants to

[1] Gulliksen, H. & Wilks, S. S. Regression Tests for Several Examples, Psychometrika, 1950, 15, 91-114.

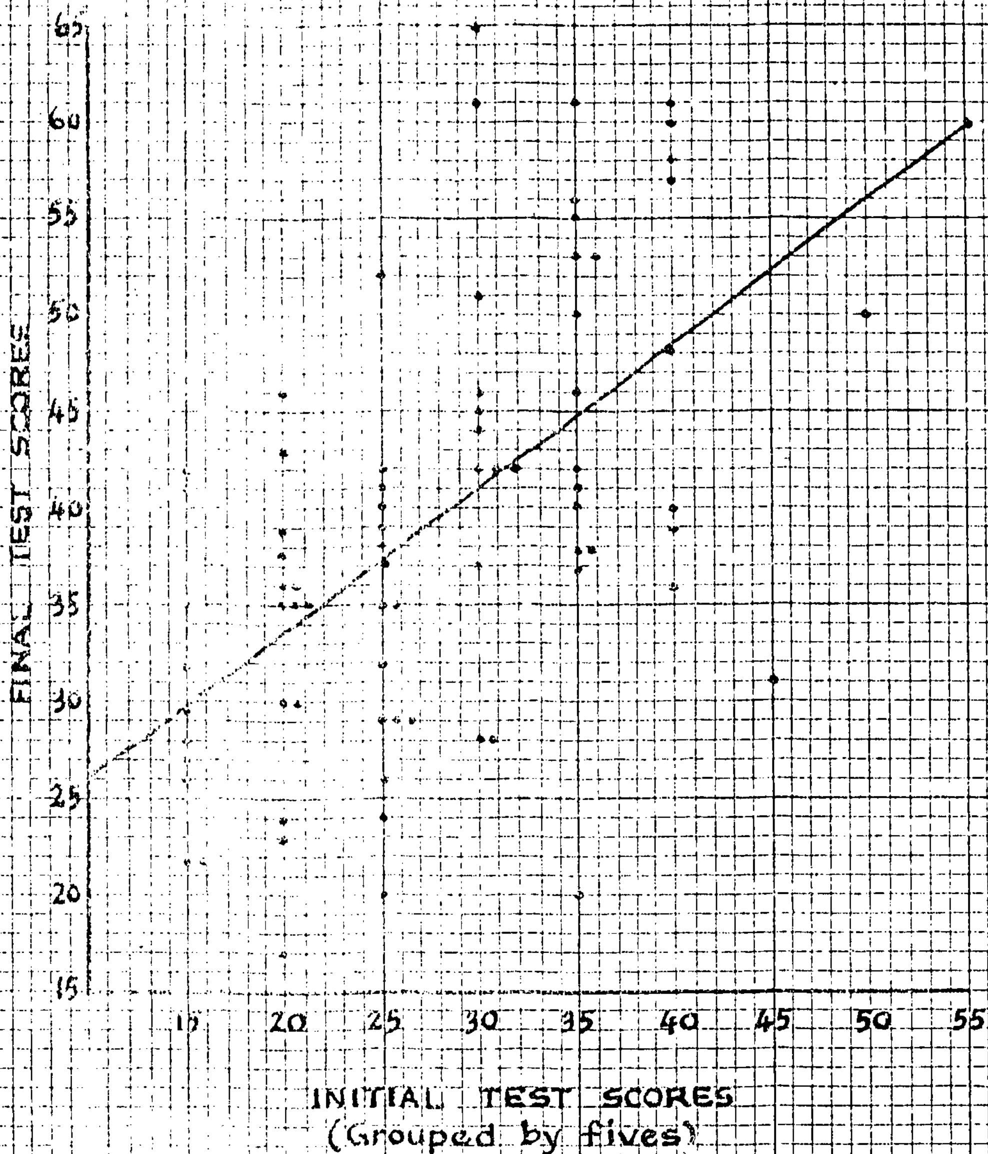
know whether differences in initial ability between the television and the conventional group may have accounted for all or part of the difference in results. Second, he wants to know whether the average difference is large enough to rule out the possibility that it arose merely by chance. Finally, and most important, he wants to know how big the average difference is after allowing for possible differences in ability of the two groups being compared. This manual provides a standard plan for getting the answers to all three questions.

Before going further, it may be well to explain what demands this plan places upon the person who uses it. First, all of the members in the two groups to be compared must have been given one or more tests prior to the training. Second, the experiment must have included at least fifty people in the experimental group and fifty people in the control group. Third, the person who has already invested a great deal of time, effort, and money in conducting an experiment must be willing and able to carry out some fairly tedious though not difficult arithmetic operations in order to be able to evaluate his results statistically. Fourth, the person who actually does the statistical analysis needs a reasonably good knowledge of certain topics in high school algebra including the use of logarithms for calculation. Fifth, the person who takes the primary responsibility for the task of statistical analysis must be willing to devote several hours of thoughtful study to the concepts involved in the method, unless he has had recent study in statistical methods as applied to educational data.*

The General Nature of the Analysis

What is the general nature of the analysis that follows? It works on the principle that there is a relationship between the score obtained by each student at the beginning of training and the score obtained at the

* One final comment may be made regarding the computing which is required. The work will be very much facilitated if it is done on a conventional desk calculating machine. A power-driven machine is very advantageous for this purpose. Many schools have such machines for the purpose of calculating grade averages and other numerical reports.



RESULTS FOR EXPERIMENTAL GROUP

Figure 1. How the line of relation summarizes the main trend of the relationship between initial scores and final test scores. (Each dot represents one student.)

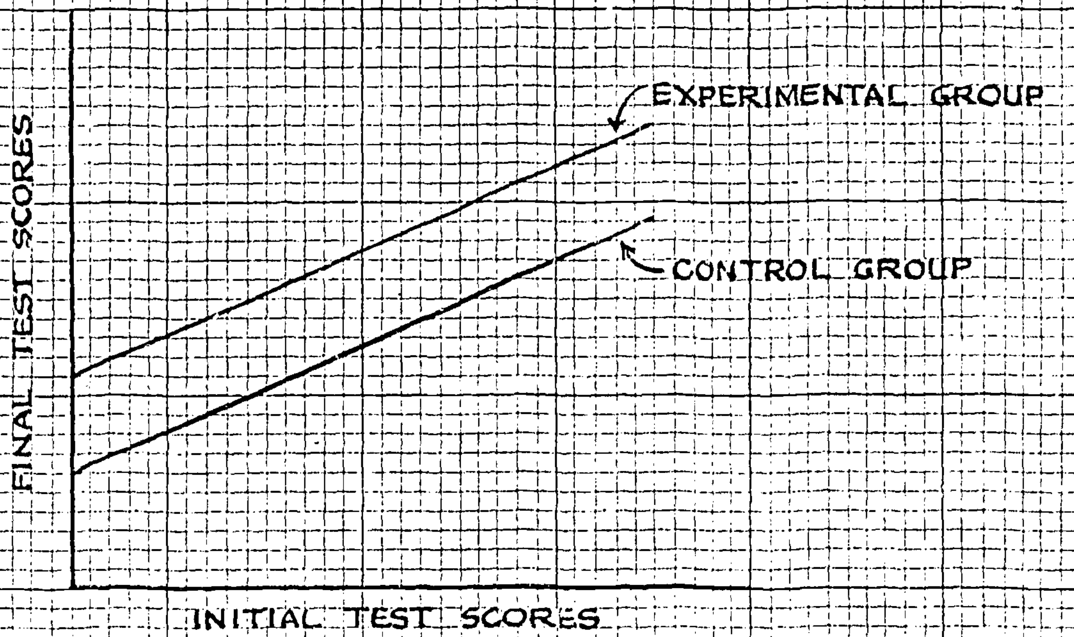


Figure 2. Lines of relation for both experimental and control groups when both have the same slope.

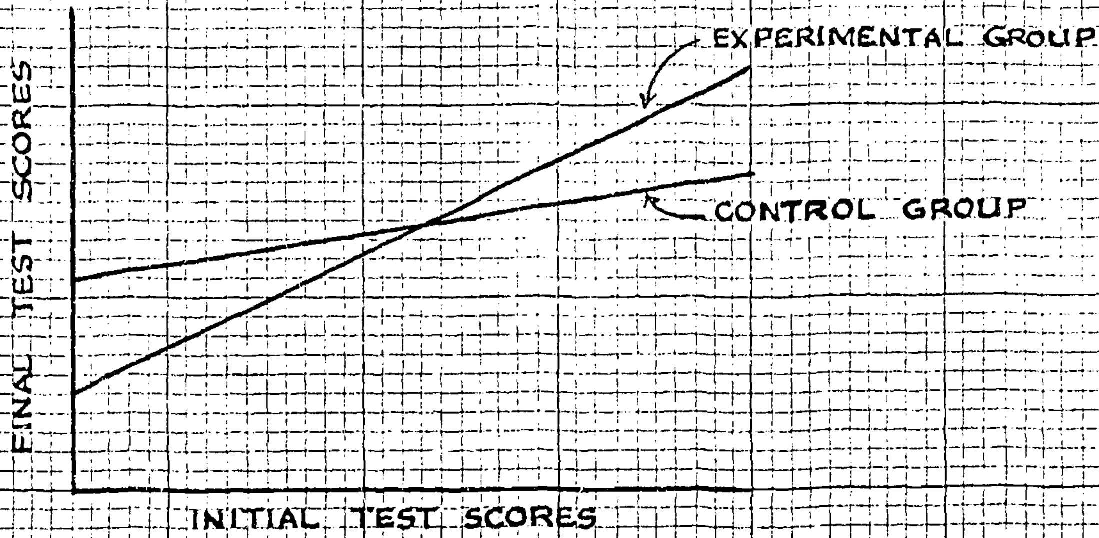


Figure 3. Lines of relation for both groups when the slopes are different.

end of training. In general, students with the higher scores at the beginning will tend to have the higher scores at the end, regardless of whether such students are in the experimental or the control group. This relationship may probably best be visualized as a "line of relation" between initial and final scores. Figure 1 shows what such a line of relation would look like for the group that studied chemistry by TV, that is, the experimental group. Notice that the relationship is far from perfect. Some students with low scores on the initial test got relatively high scores (above the line of relation) on the final test, while others with high scores on the initial test got relatively low scores (below the line of relation) on the final test. The vertical distance between any dot and the line of relation is an "error of prediction." That is, it shows how much the initial score is in error in predicting the final score.

A similar line of relation might be drawn between the initial and final scores of the control group. Figure 2 shows how the two lines of relations might look when drawn on the same chart.

The first question the experimenter asks is this: Are the errors of prediction greater as a whole in one group than in the other to an extent which cannot be attributable to chance? It is possible in some experiments that the results of the training would greatly lower or greatly increase the errors of prediction from the pretest. Usually, this is not the case, so we shall suppose that the experimenter finds that the errors of estimate in predicting final scores from pretest scores are no greater for one group than for the other.

He then proceeds to the second question. Is the line of relation <u>steeper</u> for one group than for the other? For example, it might happen that students of high initial ability in the experimental group would gain more than comparable students in the control group as the result of the kind of training given to the experimental group while students of low ability would do worse than comparable students in the control group. If this happened, one of the lines of relation would be noticeably steeper than the other. (See Figure 3) If the difference in slope proved to be statistically significant, the experimenter would have to conclude that the relative effects of the training were different for different levels of ability as measured by the initial test. Again, however, it is more likely that it would turn out that the steepness of the two lines would not be significantly different. Let us assume that it is not different. Now, it is clear that the two lines for the groups being compared may be regarded as parallel, as in Figure 2. At this point, the experimenter can ask the third question. Typically, this is the heart of the results. On the average, has the television group done better relative to its ability as measured by the initial test than has the conventionally-trained group? By applying a third test, let us suppose the experimenter finds that the difference is indeed statistically significant. He can then determine the typical amount of difference between the two groups. This difference is the vertical distance between the two lines of relation and will be uniform for all levels of initial test performance. These three questions are essentially hypotheses that will be tested in the course of the analysis.

If the procedure outlined below seems rather elaborate, it is well to keep in mind the particular pitfalls which this method is designed to avoid. If the scores used in adjusting for differences between the two groups do not show the same errors of estimate or the same slopes in relation to the measure of success in the experiment, the plan outlined here is especially valuable. Thus, if it should turn out that a particular method of training was superior for high scoring students but inferior for low scoring students, the results obtained by simpler methods would obscure these two results by merging them into a net result which would depend only on the proportion of able and inferior students in the particular groups studied. Further advantages of this plan are (1) that the statistical significance tests are made step by step as a direct part of the procedure and (2) that allowance is made for differences in ability at each step of the way.

MANAGING THE STATISTICAL ANALYSIS

Selecting the Best Predictor *

If several initial test scores are available as predictors of final scores (e.g., pretest, intelligence test, reading comprehension test), it is necessary either to decide upon which one to use or to find a suitable way to combine information from each. The following procedure offers a relatively quick and easy way to develop data to aid in making a choice, before beginning the analysis proper.

The first step in this work is to copy data for all the available initial tests on the answer sheets of the final test being careful to label each score with the designation of the test. Make up a combined set of final test answer sheets including the same number of papers from the experimental as from the control group. (If the number of papers differs for the two groups, eliminate papers at random from the larger group.) From this set of papers, choose the half which shows the highest score on the final test. (It may happen that the lowest score in the top half and the highest score in the bottom half are the same. Simply assign papers showing this score to the two groups at random.) Place a distinctive mark (a red check mark) on each paper in the top half. Now, using all the papers, sort them into an upper and a lower half on Initial Test A. Count the number of students in the top half on the final test. A similar routine should be followed for each of the other tests. The initial test which has the most candidates who score in the top half in it who are also in the top half in final score is the best predictor of the final score.

* This section may be skipped if only one predictor is to be used in the experiment.

The extension of the foregoing idea to the combination of predictors is direct. For example, the students may be divided into an upper and a lower half on the basis of the sum of scores on two predictors. If the number of students in both the top half on the sum of the two predictor scores and the top half on the final test scores is larger than when the predictors are used separately, then the combination scores should be used as the predictor in the main study. A more elaborate approach would give more weight to one predictor than the other in arriving at a combined total. Here again, however, the evaluation of the effectiveness of the combination would follow the procedure described above.

The Data from the TV Experiment in Chemistry Instruction

On the next two pages (10 and 11) are the data from the TV experiment that we are using as a basis for demonstrating the method of analysis. In this case the "initial score," which will be used for predicting the "final score," happens to be the score on Form A of an achievement test in chemistry. The "final score" is the score on Form B of the same achievement test.

DATA USED IN WORKED EXAMPLE

Experimental Group

Student	Initial Score	Final Score	Student	Initial Score	Final Score
1	13	27	41	18	30
2	33	38	42	30	51
3	51	50	43	21	30
4	40	39	44	19	43
5	20	46	45	34	53
6	16	41	46	17	32
7	23	39	47	25	40
8	18	35	48	28	28
9	19	17	49	36	61
10	36	20	50	43	31
11	35	50	51	17	42
12	33	38	52	41	57
13	30	46	53	27	37
14	42	36	54	31	42
15	21	39	55	22	35
16	31	61	56	28	28
17	38	40	57	42	48
18	18	38	58	31	37
19	26	52	59	27	35
20	25	29	60	23	26
21	25	35	61	17	35
22	14	26	62	21	35
23	33	46	63	33	41
24	32	42	64	30	44
25	27	42	65	18	36
26	35	40	66	28	45
27	14	28	67	18	24
28	25	38	68	20	36
29	27	29	69	23	24
30	13	22	70	20	23
31	27	29	71	34	42
32	37	56	72	39	58
33	40	61	73	25	20
34	37	37	74	39	60
35	36	55	75	37	53
36	14	22			
37	23	32			
38	31	42			
39	26	41			
40	32	65			

DATA USED IN WORKED EXAMPLE

Control Group

Student	Initial Score	Final Score	Student	Initial Score	Final Score
1	26	34	41	31	42
2	12	16	42	44	40
3	27	25	43	17	27
4	28	34	44	18	23
5	30	41	45	31	33
6	14	14	46	28	30
7	17	27	47	46	43
8	34	41	48	20	35
9	27	25	49	30	16
10	55	62	50	32	35
11	23	35	51	42	40
12	26	29	52	47	40
13	26	41	53	47	31
14	18	38	54	42	31
15	43	47	55	41	47
16	31	30	56	21	24
17	33	35	57	23	33
18	18	24	58	38	48
19	37	48	59	50	55
20	16	20	60	32	40
21	40	44	61	26	38
22	21	20	62	21	29
23	39	52	63	25	19
24	29	40	64	32	34
25	30	35	65	40	51
26	13	11	66	42	56
27	23	33	67	40	33
28	28	28	68	36	36
29	43	35	69	28	38
30	25	37	70	32	53
31	37	42	71	32	29
32	26	37	72	18	30
33	28	30	73	50	47
34	47	40	74	51	57
35	52	57			
36	26	15			
37	32	37			
38	16	09			
39	31	37			
40	36	46			

The Analysis

The worksheets provided with this manual specify, step by step, all the calculations needed to evaluate the three hypotheses mentioned on pages 2 and 3 and to make a simple graphic presentation of the results. Each worksheet is organized to accomplish the computation of a particular set of basic figures. Thus, Worksheet A is for computing the necessary variances. In these worksheets, figures which are to be copied later are marked by an asterisk. The space into which they are to be copied gives the source from which the number is to be copied [worksheet (by letter) and the line (by number), e.g.,"A-L3" means line 3 on worksheet A]. It is a matter of the greatest importance that the work be carefully checked. The best method would require re-doing the entire analysis and comparing the results. In any case, however, all copying should be carefully checked because it happens all too often, even with skilled computers, that errors in copying figures occur.

One question in computing arises in deciding how many decimal places to carry. Four decimal places should insure adequate precision in the final results and it is recommended that the work be carried to that number of places. The slightly added work will be more than repaid in the added confidence you will have in the final results.

WORKSHEET A

COMPUTING VARIANCES OF MEASURES

LINE	EXPERIMENTAL GROUP[a]		CONTROL GROUP[a]		HOW OBTAINED
	FINAL	INITIAL	FINAL	INITIAL	
1	2931 *	2078 *	2614 *	2316 *	Add all scores in each column.
2	123,761 *	63,038 *	101,716 *	80,438 *	Square each score. Add all squares of scores.
3	75 *	75 *	74 *	74 *	Count number of students.
4	5625 *	5625 *	5476 *	5476 *	Square each number in line 3.
5	9,282,075	4,727,850	7,526,984	5,952,412	Multiply line 2 by line 3.
6	8,590,761	4,318,084	6,832,996	5,363,856	Square line 1.
7	691,314	409,766	693,988	588,556	Subtract line 6 from line 5.
8	122.9003 *	72.8473 *	126.7327 *	107.4792 *	Divide line 7 by line 4.
9	9217.5200 *	5463.5467 *	9378.2162 *	7953.4595 *	Divide line 7 by line 3.
10	39.0800 *	27.7067 *	35.3243 *	31.2973 *	Divide line 1 by line 3.

[a] Numbers followed by an asterisk are those numbers which will be copied on subsequent worksheets.

WORKSHEET B

COMPUTING COVARIANCES OF MEASURES

LINE	EXPERIMENTAL GROUP FINAL AND INITIAL	CONTROL GROUP FINAL AND INITIAL	HOW OBTAINED
11	* 85284	* 88357	Multiply each student's final score by his initial score; add these products for all students.
12	A-L1 2931	A-L1 2614	Copy the sum of the final scores from Worksheet A, line 1. ⓑ
13	A-L1 2078	A-L1 2316	Copy the sum of the initial scores from Worksheet A, line 1.
14	A-L3 75	A-L3 74	Copy the number of students from Worksheet A, line 3.
15	A-L4 5625	A-L4 5476	Copy the square of the number of students from Worksheet A, line 4.
16	6,396,300	6,538,418	Multiply line 11 by line 14.
17	6,090,618	6,054,024	Multiply line 12 by line 13.
18	305,682	484,394	Subtract line 17 from line 16.
19	* 54.3435	* 88.4576	Divide line 18 by line 15.
20	4075.7600	6545.8649	Divide line 18 by line 14.

ⓑ Note that source is referred to as A-L1 within the space wherein the number is to be copied.

WORKSHEET C

COMBINING RESULTS FOR EXPERIMENTAL AND CONTROL GROUPS

LINE	EXPERIMENTAL	CONTROL	SUM OF EXPERIMENTAL AND CONTROL	HOW OBTAINED
21	Final A-L1 2931	Final A-L1 2614	* 5545	Copy from Worksheet A, line 1. Add.
22	Initial A-L1 2078	Initial A-L1 2316	* 4394	Copy from Worksheet A, line 1. Add.
23	Final A-L2 123761	Final A-L2 101716	* 225477	Copy from Worksheet A, line 2. Add.
24	Initial A-L2 63038	Initial A-L2 80438	* 143476	Copy from Worksheet A, line 2. Add.
25	A-L3 75	A-L3 74	* 149	Copy from Worksheet A, line 3. Add.
26	Final A-L9 9217.5200	Final A-L9 9378.2162	18595.7362	Copy from Worksheet A, line 9. Add.
27	Initial A-L9 5463.5467	Initial A-L9 7953.4595	13417.0062	Copy from Worksheet A, line 9. Add.
28	B-L11 85284	B-L11 88357	* 173641	Copy from Worksheet B, line 11. Add.
29	B-L20 4075.7600	B-L20 6545.8649	10621.6249	Copy from Worksheet B, line 20. Add.
30	X	X	Final * 124.8036	Divide line 26 by line 25.
31	X	X	Initial * 90.0470	Divide line 27 by line 25.
32	X	X	* 71.2861	Divide line 29 by line 25.

WORKSHEET D

COMPUTING TOTAL GROUP VARIANCES AND COVARIANCES

DATA FOR TOTAL GROUP. COPY FROM WORKSHEET C	
5545	L21
4394	L22
225,477	L23
143,476	L24
149	L25
173,641	L28

LINE	TOTAL GROUP	HOW OBTAINED
33	25,872,509	Multiply line 28 b,: line 25.
34	24,364,730	Multiply line 21 by line 22.
35	1,507,779	Subtract line 34 from line 33.
36	22,201	Square line 25.
37	67.9149 *	Divide line 35 by line 36.
38	33,596,073	Multiply line 23 by line 25.
39	30,747,025	Square line 21.
40	2,849,048	Subtract line 39 from line 38.
41	Final 128.3297 *	Divide line 40 by line 36.
42	21,377,924	Multiply line 24 by line 25.
43	19,307,236	Square line 22.
44	2,070,688	Subtract line 43 from line 42.
45	Initial 93.2700 *	Divide line 44 by line 36.

WORKSHEET E

COMPUTING ERRORS OF ESTIMATE

LINE	EXPERIMENTAL	CONTROL	WEIGHTED	TOTAL	HOW OBTAINED
46	Worksheet A-L3* 75	Worksheet A-L3* 74	* 149	* 149	For "weighted" and "total" columns, use the sum of the first two columns.
47	B-L19 54.3435	B-L19 88.4576	C-L32 71.2861	D-L37 67.9149	Copy.
48	Initial A-I8 72.8473	Initial A-I8 107.4792	C-L31 90.0470	D-L45 93.2700	Copy.
49	* .7460	* .8230	* .7917	.7282	Divide line 47 by line 48.
50	2953.2160	7824.7470	5081.7081	4612.4336	Square line 47.
51	40.5398	72.8024	56.4340	49.4525	Divide line 50 by line 48.
52	Final A-I8 122.9003	Final A-I8 126.7327	C-L30 124.8036	D-L41 128.3297	Copy.
53	* 82.3605	* 53.9303	* 68.3696	* 78.8772	Subtract line 51 from line 52.

WORKSHEET F

Having reached this point, we are now ready to test the three hypotheses which were stated on page 2 in the form of questions. The following computations provide the necessary information for making the three significance tests.

TESTING HYPOTHESIS A

Do the errors of prediction in the experimental group differ significantly from those in the control group?

LINE	RESULT	HOW OBTAINED
54	Experimental E-L46 75	Copy results for experimental group from Worksheet E, line 46.
55	Experimental E-L53 82.3605	Copy results for experimental group from Worksheet E, line 53.
56	6177.0375	Multiply line 54 by line 55.
57	Control E-L46 74	Copy results for control group from Worksheet E, line 46.
58	Control E-L53 53.9303	Copy results for control group from Worksheet E, line 53.
59	3990.8422	Multiply line 58 by line 57.
60	10,167.8797	Add line 56 and line 59.
61	149.	Add line 54 and line 57.
62	68.2408 *	Divide line 60 by line 61.
63	1.2069	Divide line 55 by line 62.
64	.08167	Obtain logarithm of line 63.
65	6.1252	Multiply line 64 by line 54.
66	.7903	Divide line 58 by line 62.
67	9.89779-10	Obtain logarithm of line 66.
68	-7.5635	Multiply line 67 by line 57.
69	-1.4383	Add line 65 and line 68. (The result will be a minus quantity.)

WORKSHEET G

TESTING HYPOTHESIS B - Do the slopes of the two lines of relation differ significantly?

LINE	RESULT	HOW OBTAINED
70	Weighted E-L46 149	Copy result for the weighted group from Worksheet E, line 46.
71	Weighted E-L53 68.3696	Copy result for the weighted group from Worksheet E, line 53.
72	F-L62 68.2408	Copy result from Worksheet F, line 62.
73	.9981	Divide line 72 by line 71. (The result will be less than 1.0000.)
74	9.99917-10	Obtain logarithm of line 73.
75	- .1237	Multiply line 74 by line 70.

TESTING HYPOTHESIS C - Is the distance between the two lines of relation significantly greater than zero?

LINE	RESULT	HOW OBTAINED
76	Total E-L46 149	Copy result for the total group from Worksheet E, line 46.
77	Total E-L53 78.8772	Copy result for the total group from Worksheet E, line 53.
78	.8668	Divide line 71 by line 77. (The result will be less than 1.0000.)
79	9.93792-10	Obtain logarithm of line 78.
80	-9.2499	Multiply line 79 by line 76.

Now we are ready to see what the outcome of the experiment is. Here is the way we go about it.

Hypothesis A: Do the errors of prediction in the experimental group differ significantly from those in the control group?

(1) If the value on Worksheet F, line 69, equals or exceeds 2.882, then the errors of prediction for the experimental group differ from the errors of prediction for the control group at the 1 per cent level of confidence. This means that there is less than one chance in 100 that such a difference would arise by chance. Such a difference is regarded as "very significant."

(2) If the value on Worksheet F, line 69, equals or exceeds 1.668, but is less than 2.882, then the errors of prediction for the experimental group differ from the errors of prediction for the control group at the 5 per cent level of confidence. This means that there is less than five chances in 100, but more than one chance in 100, that such a difference would arise by chance. Such a difference is regarded as "significant."

If the difference found in testing Hypothesis A is either significant or very significant, one should conclude that the results of the experiment are indeterminate. Nothing more can be said.

If, however, the difference found in testing Hypothesis A is not significant, one should then look at the outcome for Hypothesis B.

In the case of the present experiment, the difference found in testing Hypothesis A is 1.4383 (Worksheet F, line 69). (The minus sign should be disregarded.) This means that the difference found in testing Hypothesis A is not significant. Therefore we move on to test Hypothesis B.

Hypothesis B: Do the slopes of the two lines of relation differ significantly?

(1) If the value on Worksheet G, line 75, equals or exceeds 2.882, then the difference between the slopes of the lines of relation is very significant as defined above.

(2) If the value on Worksheet G, line 75, equals or exceeds 1.668, but is less than 2.882, then the difference between the slopes in the lines of relation is significant.

If the difference found in testing Hypothesis B turns out to be significant or very significant, one concludes that the effects of instruction differ from students of different ability. One cannot make any general statement about any general difference between the experimental and control groups.

If, however, the difference found in testing Hypothesis B is not significant, then one should look at the outcome for Hypothesis C.

In the case of the present experiment the difference found in testing Hypothesis B is .1237 (Worksheet G, line 75, disregarding the negative sign). This means that the difference between the slopes of the lines of relation are not significant. Therefore, we move on to test Hypothesis C -- the pay-off hypothesis.

Hypothesis C: Is the distance between the two lines of relation significantly greater than zero?

(1) If the value on Worksheet G, line 80, equals or exceeds 2.882, then the overall difference between the experimental and control groups is very significant.

(2) If the value on Worksheet G, line 80 equals or exceeds 1.668, but is less than 2.882, then the overall difference between the experimental and control group is significant.

If the difference is either significant or very significant one can conclude that the effect of instruction on one group is in all probability really different from the effect of instruction on the other group.

In the present experiment the difference found in testing Hypothesis C is 9.2499 (Worksheet G, line 80, disregarding the minus sign). This is a very significant difference. This means that in all probability the different kinds of instruction have had genuinely different effects.

SUMMARY STATEMENT

For each of the three hypotheses, the following applies:

LEVEL OF SIGNIFICANCE	MINIMUM CALCULATED VALUE
1 per cent	2.882
5 per cent	1.668

WORKSHEET H

In order to get a picture of the difference between the two groups one may set up a chart showing the lines of relation for each. It is done as follows:

CALCULATIONS FOR LINES OF RELATION

LINE	EXPERIMENTAL	CONTROL	HOW OBTAINED
81	Final A-L10 39.0800	Final A-L10 35.3243	Copy from appropriate column of Worksheet A, line 10.
82	Initial A-L10 27.7067	Initial A-L10 31.2973	Copy from appropriate column of Worksheet A, line 10.
83	E-L49 .7917	E-L49 .7917	Copy from Worksheet E, line 49. If Hypotheses A and B are <u>not</u> rejected, copy from "weighted" column for both groups. If either hypothesis is rejected, use appropriate value for each group.
84	21.9354	24.7781	Multiply line 82 by line 83.
85	17.1446	10.5462	Subtract line 84 from line 81.
86	60	60	Choose a convenient, high score on the predictor. Use the same score for both groups.
87	47.5020	47.5020	Multiply line 86 by line 83.
88	64.6466	58.0482	Add line 87 to line 85.
89	10	10	Choose a convenient, low score on the predictor. Use the same score for both groups.
90	7.9170	7.9170	Multiply line 89 by line 83.
91	25.0616	18.4632	Add line 90 to line 85.

SUMMARY OF POINTS FOR LINES OF RELATION

EXPERIMENTAL GROUP		CONTROL GROUP	
Initial	Final	Initial	Final
line 86: 60	line 88: 64.6	line 86: 60	line 88: 58.0
line 89: 10	line 91: 25.1	line 89: 10	line 91: 18.5

Graphing the Final Results

If the first two hypotheses do not yield "significant" differences, but the third hypothesis does, a highly effective graphical presentation of the results is possible. In this case, lines of relation between the measure of achievement and the predictive measure can be regarded as parallel. The vertical distance between the two lines can be taken as an indication of the extent to which one group excels the other. If either Hypothesis A or Hypothesis B is rejected, the two lines of relation will not be parallel. It may be useful, nevertheless, to draw the lines for the two groups.

All the basic calculations needed for drawing the two lines are included on Worksheet H. In making the graph, it is necessary to lay out a vertical and a horizontal scale. The vertical scale should begin at a value somewhat smaller than the lowest of the four final values shown in the summary table at the foot of Worksheet H and extend a bit higher than the highest value. The vertical scale should have low scores at the bottom end and high scores at the top end. The horizontal scale should begin with a score somewhat lower than the lower of the two selected values of the initial measure and extend to a score somewhat higher than the higher of the two. On the horizontal scale, low scores should be placed at the left and high scores at the right end of the scale.

Once the scales have been laid out, it is necessary to plot the four points given in the summary table on Worksheet H. To locate the first point, proceed along the horizontal scale until you reach the value of the initial measure, and then proceed upward until you reach a point at the same vertical level as the computed value of the final measure for that point. For example, the first point in Worksheet H would be plotted by going to the right until 60 is reached, and then going upward until 64.6 is reached. When the point is located, it may be marked by a heavy dot. This point will be near the upper right-hand corner of the graph. Then, the other point for the experimental group should be located, and marked by a heavy dot. A straight line drawn through these points is the line of relation for the experimental group. Exactly the same procedure may be followed for the other line. Figure 4 shows the results for the worked example.

Note that if neither Hypothesis A (Question 1) nor Hypothesis B (Question 2) shows a significant difference, the two lines of relation will be parallel. If, however, there is a significant difference for either hypothesis, the two lines of relation will not be parallel.

Presenting Final Results in a Table

At times, it may be convenient to summarize the main statistical results of an experiment in a table. Table 1 has been prepared to serve as an illustration of a table giving a fairly full account of the design and results. The mean scores reported in Table 1 can be obtained directly from Worksheet H, lines 81 and 82.

It must be noted that the summary statement shown in Table 1 should be given only if the tests for both Hypothesis A and Hypothesis B are not significant and if the test for Hypothesis C is significant. In that case, the numerical value of the difference reported can readily be obtained by subtracting the control group value in line 88 from the experimental group value in that line. As a check, the difference may also be determined by subtracting the control group value in line 91 from the experimental group value.

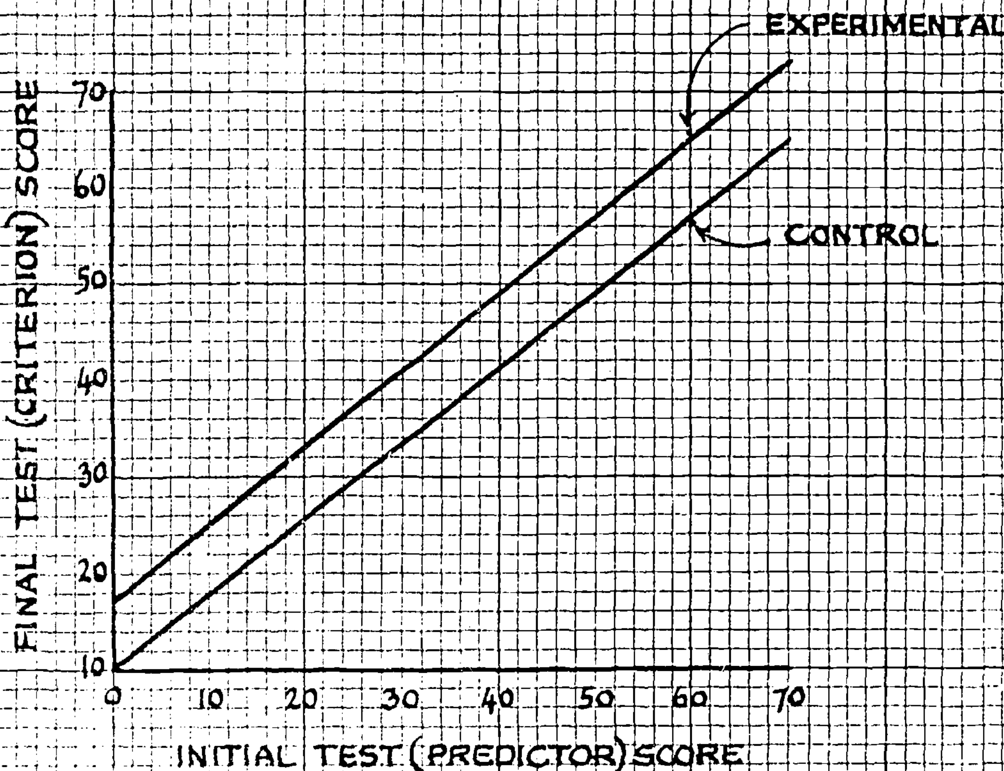


Figure 4. Results of experiment analyzed in worked example. The experimental group is 6.6 score points higher, when ability is held constant, than the control group.

TABLE I

RESULTS OF CHEMISTRY TRAINING EXPERIMENT

Groups Studied:

Experimental: 75 students who were taught chemistry using special TV lectures and kinescopes

Control: 74 students who were taught chemistry by conventional methods

Measures Used:

Initial: Form A of an achievement test in chemistry

Final: Form B of an achievement test in chemistry

Mean Scores:

	Experimental Group	Control Group
Initial Test	27.71	31.30
Final Test	39.08	35.32

Analysis of Covariance

A. Equality of errors of estimate: Not significant

B. Equality of slopes: Not significant

C. Equality of intercepts: Significant at one per cent level

Summary: The advantage of the experimental group on final scores, after allowing for differences between the groups on initial score, was 6.6 points.

WORKSHEET A

COMPUTING VARIANCES OF MEASURES

LINE	EXPERIMENTAL GROUP [a]		CONTROL GROUP [a]		HOW OBTAINED
	FINAL	INITIAL	FINAL	INITIAL	
1	*	*	*	*	Add all scores in each column.
2	*	*	*	*	Square each score. Add all squares of scores.
3	*	*	*	*	Count number of students.
4					Square each number in line 3.
5					Multiply line 2 by line 3.
6					Square line 1.
7					Subtract line 6 from line 5.
8	*	*	*	*	Divide line 7 by line 4.
9	*	*	*	*	Divide line 7 by line 3.
10	*	*	*	*	Divide line 1 by line 3.

[a] Numbers followed by an asterisk are those numbers which will be copied on subsequent worksheets.

WORKSHEET B

COMPUTING COVARIANCES OF MEASURES

LINE	EXPERIMENTAL GROUP FINAL AND INITIAL	CONTROL GROUP FINAL AND INITIAL	HOW OBTAINED
11	*	*	Multiply each student's final score by his initial score; add these products for all students.
12	A-L1	A-L1	Copy the sum of the final scores from Worksheet A, line 1. ⓑ
13	A-L1	A-L1	Copy the sum of the initial scores from Worksheet A, line 1.
14	A-L3	A-L3	Copy the number of students from Worksheet A, line 3.
15	A-L4	A-L4	Copy the square of the number of students from Worksheet A, line 4.
16			Multiply line 11 by line 14.
17			Multiply line 12 by line 13.
18			Subtract line 17 from line 16.
19	*	*	Divide line 18 by line 15.
20			Divide line 18 by line 14.

ⓑ Note that source is referred to as A-L1 within the space wherein the number is to be copied.

WORKSHEET C

COMBINING RESULTS FOR EXPERIMENTAL AND CONTROL GROUPS

LINE	EXPERIMENTAL	CONTROL	SUM OF EXPERIMENTAL AND CONTROL	HOW OBTAINED
21	Final A-L1	Final A-L1	*	Copy from Worksheet A, line 1. Add.
22	Initial A-L1	Initial A-L1	*	Copy from Worksheet A, line 1. Add.
23	Final A-L2	Final A-L2	*	Copy from Worksheet A, line 2. Add.
24	Initial A-L2	Initial A-L2	*	Copy from Worksheet A, line 2. Add.
25	A-L3	A-L3	*	Copy from Worksheet A, line 3. Add.
26	Final A-L9	Final A-L9		Copy from Worksheet A, line 9. Add.
27	Initial A-L9	Initial A-L9		Copy from Worksheet A, line 9. Add.
28	B-L11	B-L11	*	Copy from Worksheet B, line 11. Add.
29	B-L20	B-L20		Copy from Worksheet B, line 20. Add.
30			Final *	Divide line 26 by line 25.
31			Initial *	Divide line 27 by line 25.
32			*	Divide line 29 by line 25.

WORKSHEET D

COMPUTING TOTAL GROUP VARIANCES AND COVARIANCES

DATA FOR TOTAL GROUP. COPY FROM WORKSHEET C
L21
L22
L23
L24
L25
L28

LINE	TOTAL GROUP	HOW OBTAINED
33		Multiply line 28 by line 25.
34		Multiply line 21 by line 22.
35		Subtract line 34 from line 33.
36		Square line 25.
37	*	Divide line 35 by line 36.
38		Multiply line 23 by line 25.
39		Square line 21.
40		Subtract line 39 from line 38.
41	Final *	Divide line 40 by line 36.
42		Multiply line 24 by line 25.
43		Square line 22.
44		Subtract line 43 from line 42
45	Initial *	Divide line 44 by line 36.

WORKSHEET E

COMPUTING ERRORS OF ESTIMATE

LINE	EXPERIMENTAL	CONTROL	WEIGHTED	TOTAL	HOW OBTAINED
46	Worksheet A-L3*	Worksheet A-L3*	*	*	For "weighted" and "total" columns, use the sum of the first two columns.
47	B-L19	B-L19	C-L32	D-L37	Copy.
48	Initial A-L8	Initial A-L8	C-L31	D-L45	Copy.
49	*	*	*	*	Divide line 47 by line 48.
50					Square line 47.
51					Divide line 50 by line 48.
52	Final A-L8	Final A-L8	C-L30	D-L41	Copy.
53	*	*	*	*	Subtract line 51 from line 52.

WORKSHEET F

Having reached this point, we are now ready to test the three hypotheses which were stated on page 2 in the form of questions. The following computations provide the necessary information for making the three significance tests.

TESTING HYPOTHESIS A

Do the errors of prediction in the experimental group differ significantly from those in the control group?

LINE	RESULT		HOW OBTAINED
54	Experimental	E-L46	Copy results for experimental group from Worksheet E, line 46.
55	Experimental	E-L53	Copy results for experimental group from Worksheet E, line 53.
56			Multiply line 54 by line 55.
57	Control	E-L46	Copy results for control group from Worksheet E, line 46.
58	Control	E-L53	Copy results for control group from Worksheet E, line 53.
59			Multiply line 57 by line 58.
60			Add line 56 and line 59.
61			Add line 54 and line 57.
62		*	Divide line 60 by line 61.
63			Divide line 55 by line 62.
64			Obtain logarithm of line 63.
65			Multiply line 64 by line 54.
66			Divide line 58 by line 62.
67			Obtain logarithm of line 66.
68			Multiply line 67 by line 57.
69			Add line 65 and line 68. (The result will be a minus quality.)

WORKSHEET G

TESTING HYPOTHESIS B - Do the slopes of the two lines of relation differ significantly?

LINE	RESULT	HOW OBTAINED
70	Weighted E-L46	Copy result for the weighted group from Worksheet E, line 46.
71	Weighted E-L53	Copy result for the weighted group from Worksheet E, line 53.
72	F-L62	Copy result from Worksheet F, line 62.
73		Divide line 72 by line 71. (The result will be less than 1.0000.)
74		Obtain logarithm of line 73.
75		Multiply line 74 by line 70.

TESTING HYPOTHESIS C - Is the distance between the two lines of relation significantly greater than zero?

76	Total E-L46	Copy result for the total group from Worksheet E, line 46.
77	Total E-L53	Copy result for the total group from Worksheet E, line 53.
78		Divide line 71 by line 77. (The result will be less than 1.0000.)
79		Obtain logarithm of line 78.
80		Multiply line 79 by line 76.

WORKSHEET H

In order to get a picture of the difference between the two groups one may set up a chart showing this lines of relation for each. It is done as follows:

CALCULATIONS FOR LINES OF RELATION

LINE	EXPERIMENTAL	CONTROL	HOW OBTAINED
81	Final A-L10	Final A-L10	Copy from appropriate column of Worksheet A, line 10.
82	Initial A-L10	Initial A-L10	Copy from appropriate column of Worksheet A, line 10.
83	E-L49	E-L49	Copy from Worksheet E, line 49. If Hypotheses A and B are <u>not</u> rejected, copy from "weighted" column for both groups. If either hypothesis is rejected, use appropriate value for each group.
84			Multiply line 82 by line 83.
85			Subtract line 84 from line 81.
86			Choose a convenient, high score on the predictor. Use the same score for both groups.
87			Multiply line 86 by line 83.
88			Add line 87 to line 85.
89			Choose a convenient, low score on the predictor. Use the same score for both groups.
90			Multiply line 89 by line 83.
91			Add line 90 to line 85.

SUMMARY OF POINTS FOR LINES OF RELATION

EXPERIMENTAL GROUP		CONTROL GROUP	
Initial	Final	Initial	Final
LINE 86:	LINE 88:	LINE 86:	LINE 88:
LINE 89:	LINE 91:	LINE 89:	LINE 91: