CANADA

Department of Forestry

FOREST RESEARCH BRANCH

Simplified form of computation in a two-variable linear regression

(using the least squares method)

by

Luc Valiquette

Quebec District
December 1962

Que. 62-15

INTRODUCTION

To establish relationships between two variables is a problem which is often encountered in forest research and practice. The most scientific method of establishing it is by doing a regression analysis.

Despite the large amount of work that has been done to solve a two-variable linear regression problem, it is believed that there is a need for a simplified form of computation which combines and standardizes the statistical computations relating to such a problem.

The "Simplified form of computation in a two-variable linear regression" has been conceived not only to combine and standardize the computation procedures as well as the presentation of the inferences, but mainly to simplify the computation work (even a person without any statistical background is able to compute such a problem) in order to render it accessible to almost anybody. The suggested form is here presented and briefly explained, and a practical example is worked out.

METHODS

The least squares method which requires that the sum of the squares of the deviations of the observed points from the straight - line moving average be a minimum has been used; and, the conventional methods of computation of the statistics, which are well explained in most textbooks dealing with regression analysis, have been utilized.

THE SIMPLIFIED FORM

The "Simplified form of computation in a two-variable linear regression" is presented in Appendix 2. It is composed of two parts. Part I is a prerequisite dealing with the three characteristics of the populations sampled. These are (after Snedecor, 1957) :

a) for any selected X there is a normal distribution of Y from which the sample Y has been taken at random;

b) the true means of all the sampled populations lie on a straight regression line;

c) all sampled populations are normally distributed and have a common standard deviation (independent of X).

As it is suggested by most authors on this subject, a scatter diagram relating the two variables should be made before starting with computation. It will allow to check ocularly the characteristics of the populations sampled. This does not mean that one should always be content with just an ocular check. It is suggested that this is sufficient in most instances. There are particular cases in which it would be in order to carry out some important statistical tests such as uniformity of variances.

Part II of the computation form deals with the computation itself. The main statistics computed are the sample means, the sample regression coefficient, the sample regression constant, the sample standard deviations from regression and sample standard errors of the estimated Y at any selected value within the possible range of the independent variable. A table of "t" values for different probability levels is also given (Goulden, 1956).

AN EXAMPLE

To show how the computation is done, an example has been worked out. It is planned to find out the relationship between diameter (outside bark) at breast height and stump diameter (inside bark) at one foot. For this purpose, 44 balsam fir trees of various stump diameters have been sampled and their corresponding d.b.h. has been measured (Appendix 1).

As stipulated on part I of the form, and illustrated in Fig. 1, a scatter diagram has been made relating the data at hand. The three characteristics of the populations sampled have been checked ocularly and the computation has been done.

All the important inferences about the linear regression of diameter at breast height outside bark on stump diameter at one foot inside bark, are now available. The main inferences are graphically presented on Fig. 2 and, within the range of the experiment, the estimates of the most important parameters of the populations sampled are possible.

LITERATURE CITED

1. FISHER, R.A. 1950. Statistical Methods for Research Workers. 11th ed. Oliver and Boyd, Edinburgh.

2. GOULDEN, C.H. 1956. Methods of Statistical Analysis. 2nd ed. John Wiley & Sons, Inc., New York.

3. SNEDECOR, G.W. 1957. Statistical Methods Applied to Experiments in Agriculture and Biology. 5th ed. The Iowa State College Press, Ames, Iowa.

The individual observations on 44 balsam fir trees

and

the preliminary computations

Tree No.	Stump diameter (i.b.) -X-	Diameter at breast height (o.b.) -Y-
1	9.3	9.0
2	7.2	7.0
3	5.3	4.8
4	9.3	8.2
5	6.1	6.1
6	9.2	8.2
7	7.8	7.3
8	5.2	5.2
9	10.1	10.3
10	6.5	6.4
11	8.7	8.4
12	7.5	7.4
13	13.3	12.1
14	7.6	7.6
15	7.9	6.6
16	6.3	6.4
17	8.2	8.0
18	10.3	9.1
19	5.2	5.0
20	10.7	9.6
21	7.4	6.9
22	6.5	6.6

Tree No.	Stump diameter (i.b.) - X -	Diameter at breast height (o.b.) - Y -
23	10.1	10.0
24	8.5	8.0
25	7.8	7.4
26	9.5	9.0
27	6.9	6.8
28	8.6	7.7
29	4.9	5.7
30	9.8	9.0
31	8.4	8.0
32	6.8	6.7
33	10.3	9.7
34	6.7	6.6
35	6.7	6.1
36	12.6	11.6
37	7.1	6.8
38	11.3	10.4
39	8.6	8.5
40	4.6	4.5
41	5.1	5.2
42	7.9	7.7
43	8.7	8.7
44	6.0	5.9
	$\Sigma X = 352.5$	$\Sigma Y = 336.2$
	$\Sigma X^2 = 2999.19$	$\Sigma Y^2 = 2702.08$

$\Sigma XY = 2843.25$

$n = 44$