Proceedings of the Iowa Academy of Science

Volume 76 | Annual Issue Article 38

1969

Statistical Methods for Integrating Populations of a Fauna Collected at Various Localities

D. E. Simon
Iowa State University

J. H. Elwell
Iowa State University

Statistical Methods for Integrating Populations of a Fauna Collected at Various Localities

D. E. Simon[1] and J. H. Elwell[2]

Abstract. Parameters characteristic of *Atrypa Histrix,* collected from various localities of the Cerro Gordo Member of the Upper Devonian Lime Creek Formation, were measured as:

1) Length, 2) Width, 3) Thickness, 4) Number of growth lines, and 5) Number of plications. The consistency of data obtained within a sample location was tested against similar data representing the total population by means of F-tests. Results suggested that the three sample locations represented at least two statistically random sample populations. These statistical differences in the single biologic population may be alternately interpreted as: 1) the population differed prior to preservaton, or 2) is attributed to post depositional changes at the sample locations. F-tests of the parameter ratios increased resolution sufficiently to suggest a single sample population statistically characteristic of the one biologic population.

Brachiopods identified by Vondra (1) as the species *Atrypa Histrix were* collected from three exposures of the Upper Devonian Lime Creek Formation. The sample locations are identified on the Index Map, Figure 1, as Rockford Brick and Tile Quarry, Bird

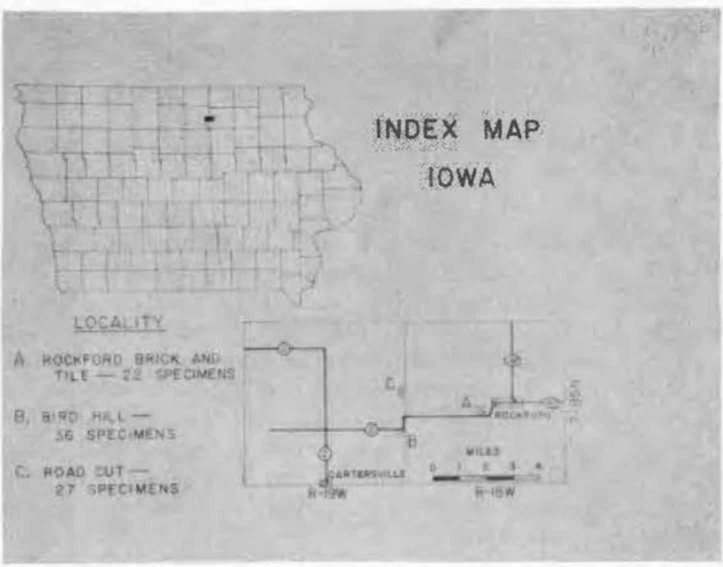

Fig. 1. Index may showing location of sampling sites.

[1] Graduate teaching Assistant, Geology, Iowa State University, Ames, Iowa.

[2] Instructor, Geology, Iowa State University, Ames, Iowa.

Proceedings of the Iowa Academy of Science, Vol. 76 [1969], No. 1, Art. 38

Hill and Roadcut. These locations are well known and it is generally agreed that these brachiopod samples form one biologic population. Several attributes of the brachiopod samples may be measured and recorded. The set of such attributes used in this study is shown in Figure 2. These data were compiled for 22

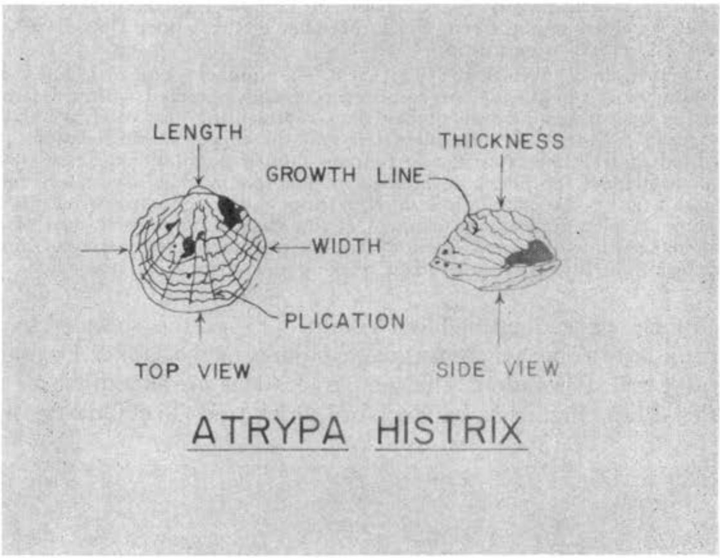

Fig. 2. Schematic line drawing of *Atrypa Histrix* showing parameters measured.

brachiopods from Rockford Brick and Tile Quarry, 36 from Bird Hill and 27 from Roadcut. Having generated these data we are concerned with what they represent. There are three possibilities: 1) they could represent only the specimens collected; 2) they could represent different ecological or preservation conditions characteristic of each sample location; or 3) they could be characteristic of the single biologic population. The purpose of this study is to illustrate how the self-consistency of the data may be used to determine a set of attributes that are characteristic of one statistical population. Such a set would be characteristic of the single biologic population.

F-Tests

Let us consider how to test for the self-consistency in data between sample locations. The criteria for acceptance as one useful statistical population are 1) samples may be considered as one random population; 2) characteristic relationships be linear or constant; and 3) characteristic relationships have a unique set of values for the biologic species. The first criterion may be evaluated by means of the variance ratio, F. Consider repeated measurements of a single specimen length. The sample standard deviation per degree of freedom or mean square (2) would be a measure of the uncertainty in length expected for a single observation. F-tests or ratios of mean squares calculated for different groupings estimate differences in the uncertainty of observations between groups of samples. Identical groups would have a ratio of one and the calculated value of F would be influenced by the number of observations and the sample uncertainty. Values of F have been tabulated for samples of different sizes drawn from a single random population. Critical tabulated values, Snedecor and Cochran (2) are used to test the self-consistency of these data. Calculated values that exceed the tabulated value for samples drawn from a single random population only 5% of the time are identified by*. Highly significant values which exceed the tabulated 1% value are identified by**. In our testing of self-consistency we will assume that more than one statistical population is present if either tabulated value is exceeded by the calculated F.

Table 1. Variance ratios of length, width, thickness, growth lines and plications between different groups.

	Total[a] / Rockford	Total[b] / Roadcut	Total[c] / Bird Hill	Rockford[d] / Roadcut & Bird Hill
Length	1.07	4.64**	2.03**	2.42**
Width	1.10	3.75**	1.86*	2.11*
Thickness	1.14	3.20**	1.88*	1.85*
Growth Lines	1.11	2.90**	2.61**	2.50**
Plications	1.35	5.61**	2.13**	2.03*

[a] *5% significance level(21,84 df) = 1.68
**1% significance level(21,84 df) = 2.09

[b] *5% significance level(84,27 df) = 1.75
**1% significance level(84,27 df) = 2.23

[c] *5% significance level(34,84 df) = 1.58
**1% significance level(34,84 df) = 1.90

[d] *5% significance level(21,62 df) = 1.73
**1% significance level(21,62 df) = 2.17

Measured-Populations

Calculated F values for the attributes measured are shown in Table 1. From the * symbols shown we conclude that no measured parameter is characteristic of one statistical population. Since they represent more than one population, none of these measurements may represent the single biologic population statistically. Because they lack self-consistency we reject these data as parameters characteristic of the biologic population.

Since the acceptance of multiple populations conflicts with the

Table 2

Key to coding of comparisons

Symbol	Test Comparisons
L/W	Length to Width
L/T	Length to Thickness
W/T	Width to Thickness
G/P	Growth Lines to Plications
G/L	Growth Lines to Length
G/W	Growth Lines to Width
G/T	Growth Lines to Thickness
P/L	Plications to Length
P/W	Plications to Width
P/T	Plications to Thickness

Table 3. Variance ratios for the parameter ratios compared between groups.

	Total[a] / Rockford	Total[b] / Roadcut	Total[c] / Bird Hill	Rockford[d] / Roadcut & Bird Hill
L/W	1.02	1.24	1.10	1.14
L/T	1.06	1.45	1.07	1.30
T/W	1.05	1.17	1.02	1.17
G/P	1.22	1.18	1.23	1.24
G/L	1.04	1.59	1.29	1.03
G/W	1.00	1.29	1.41	1.19
G/T	1.03	1.11	1.38	1.35
P/L	1.26	1.19	1.05	1.19
P/W	1.22	1.48	1.14	1.05
P/T	1.13	1.74	1.13	1.09

[a] * 5% significance level(21,84 df) = 1.68
** 1% significance level(21,84 df) = 2.09

[b] * 5% significance level(84,27 df) = 1.75
** 1% significance level(84-27 df) = 2.23

[c] * 5% significance level(34,84 df) = 1.58
** 1% significance level(34,84 df) = 1.90

[d] * 5% significance level(21,62 df) = 1.73
** 1% significance level(21,62 df) = 2.17

original assumption of a single biologic population, grouping of the measurements was tried in an attempt to find parameters characteristic of the single biologic population. The new ratio parameters are shown in Table 2, which also serves as a key to the coding of parameters. Calculated F values for the ratio parameters are shown in Table 3. It should be noted that the same sample groupings that were significantly different in Table 1 as indicated by the * are now consistent with the assumption of a single statistical population. The absence of * symbols indicates that any of the ratios will satisfy the single statistical population criterion required of the useful statistical parameters sought to describe the single biologic population. By changing the form of the parameter, and testing the self-consistency of the sample data, we have found candidates for useful parameter description.

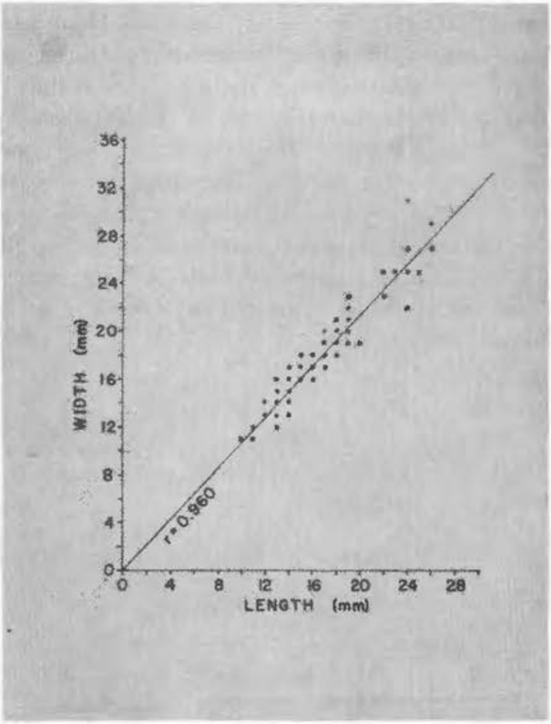

Fig. 3. Plot of length versus width.

Correlation Coefficient Tests

For the statistical population to satisfy the second criterion, a linear relationship characteristic of the population is required. A plot of the ratio numerator vs the denominator should be a straight line if this requirement is to be met by the ratio. The correlation coefficient (r) is a numerical measurement of the ability of these data to fit a straight line. The absolute value of (r) will lie between zero, a perfect circle, and 1, a straight line. The larger the absolute value of (r) the better the data fit a straight line. (2) For the purposes of indicating a linear relationship in the population sampled, an arbitrary limit of 0.8 was selected. Values of (r) greater than 0.8 will satisfy the linear relationship required of useful parameter ratios.

The Linear Relationship

A plot of Width vs Length is shown to illustrate a parameter ratio relationship in Figure 3. The (r) value is accepted as evidence that the ratio satisfies the linear relationship required of a useful statistical population. In passing, it should be noted that if the line passes through the origin the ratio may be characterized by a single constant, the slope of the line. However, if the line does not pass through the origin both a slope and an intercept are required to characterize the linear relationship. Such constants serve as expected values of the parameters characteristic of the single biologic population. New sample locations which did not generate these expected constants would be suspected to contain a different biologic population.

Table 4

Correlation coefficients *r* as calculated for the various comparisons within different groups.

Comparison	Rockford	Roadcut	Bird Hill	Roadcut & Bird Hill	Total
L/W	0.914	0.897	0.953	0.938	0.960
L/T	0.778	0.791	0.657	0.710	0.854
W/T	0.612	0.716	0.704	0.718	0.836
G/P	0.529	0.534	0.698	0.602	0.791
G/L	0.806	0.494	0.703	0.618	0.851
G/W	0.708	0.347	0.638	0.535	0.797
G/T	0.861	0.476	0.566	0.516	0.823
P/L	0.655	0.442	0.887	0.670	0.834
P/W	0.559	0.319	0.776	0.576	0.779
P/T	0.640	0.576	0.469	0.370	0.726

A Useful Parameter Set

Correlation coefficients calculated for each parameter ratio at each sample location and for the total population are shown in Table 4. The individual sample locations are presented to show that the (r) calculated for the total population is larger than the individual (r's) for each parameter ratio. This qualitatively supports the concept of a single statistical population for ratio attribute in the biologic population sampled. To a first approximation several of the parameter ratios for the total population form a set that appear to satisfy the linear relationship required of a useful parameter. The set characteristic of the single biologic population includes ratios of L/W, G/L, W/T and L/T all which have a correlation coefficient of 0.8 or higher. The constants representing this set of parameters as slopes or intercepts are statistically characteristic of the biologic species.

Conclusions

1. The selected set of slopes and intercepts may be used as expected values for the biologic population.
2. Systematic deviation from the expected values may reveal anomalies in certain sample location related to environment or preservation.
3. Whatever the paleontologist's interest, these methods will enable parameters to be tested and grouped so that the parameter reported will indeed represent the statistcial population he desires to study.
4. The major contribution of the paper is the compilation of common statistical tests into a procedure for the statistical determination of what population each parameter estimates.

References Cited

Vondra, Carl F. Associate Professor of Geology, Iowa State University Personal communication.

Snedecor, George W. & William G. Cochran. *Statistical Methods.* Sixth Edition. The Iowa State University Press, Ames, Iowa, 1967.

Appendix I

Statistical formulas (Snedecor and Cochran, 1967).

1. $X =$ observation
2. $n =$ number of observations
3. $\overline{X} =$ mean
4. $SS_A = \Sigma x_A^2 = \Sigma(X_A - \overline{X})^2 =$ Sum of Squares
5. d.f. $=$ n-1 $=$ Degrees of Freedom
6. $MS_A = {}^{ss}{}_A/\text{d.f.} =$ Mean Square

7. $Fc,AB = \frac{\text{MS class A}}{\text{MS class B}} = \text{F-test}$

* $F_c > F_T$, exceeded by random population 5%

** $F_c > F_T$, exceeded by random population 1%

8. $SS_{AB} = \Sigma x_A x_B = \left[(\Sigma X_A X_B) - \frac{(\Sigma X_A)\ (\Sigma X_B)}{n} \right]$

9. $r = \Sigma x_A x_B / \Sigma x_A^2 \Sigma x_B^2$ = Correlation Coefficient

"A measure of the degree of linearity:

$\pm =$ perfect fit."