FINDING A FOOTPATH IN A FOREST OF FIGURES

A Beginner's Guide to Interpreting Numerical Data

By Emil J. Berger, Jr.

When confronted with a mass of data from a published article, or their own records, many banders utter sighs of desperation and wonder: What do all these numbers tell me?

The object of this discussion is to shed some light on the basics of data interpretation.

Let's assume that over a period of time you have banded 64 Carolina chickadees, and have carefully measured and recorded the length of the tail on each bird. You have also banded and measured the tails of 39 black-capped chickadees. (The data I shall be using is taken from EBBA Workshop Manual, Vol. II.)

About now you begin to suspect that the tail length MIGHT be a means of differentiating between these species when the decision is difficult, so interpretation is called for.

The first step is organizing the data for examination. A tally sheet is convenient for this. Arrange the tail lengths in a vertical column and tally the number of times each particular length was recorded to the right of the length value. Round off the lengths to the nearest whole millimeter.

Now, it appears that you MAY be on the track of something significant. The clusters of measurements for the two species are in different locations on the length scale.

The next thing you want to know is the average value for each group of measurements. Average is synonymous with mean in this case. Let's do the black-caps first.

Refer to the computation sheet.

Column 1 repeats a tabulation of the rounded-off tail length values.

Column 2 is labeled "N". This is an abbreviation for the number of individuals with each particular rounded-off tail length, obtained from the tally sheet.

Column 3 is headed "Coded Length L". It is possible to make these calculations using the true length (for example:-59 mm.), but far easier to convert the true length to a coded length. All this requires is calling the lowest measurement 1, the next higher 2, the next higher 3, and so forth, so long as the intervals are equal.

BLACK-CAPPED CHICKADEE COMPUTATIONS

1	2	3	4	5	6	7
MM	N	L	NxL	D	D^2	NxD^2
57	2	1	2	4	16	32
58	0	2	0	3	9	0
59	7	3	21	2	4	28
60	10	4	40	1	1	10
61	5	5	25	0	0	0
62	2	6	12	1	1	2
63	8	7	56	2	4	32
64	3	8	24	3	9	27
65	2	9	18	4	16	32
	TOTAL 39		TOTAL 198			TOTAL 163

$$L \text{ average} = \frac{\text{Total}(NxL)}{\text{Total}(N)} = \frac{198}{39} = 5.1$$

Round off to 5 so actual average length is 61 mm.

$$\text{Std. Deviation} = \sqrt{\frac{\text{Total}(NxD^2)}{\text{Total}(N)}}$$

$$S.\ D. = \sqrt{\frac{163}{39}} = \sqrt{4.2} = 2.1 \text{ mm.}$$

CAROLINA CHICKADEE COMPUTATIONS

48	1	1	1	5	25	25
49	1	2	2	4	16	16
50	5	3	15	3	9	45
51	7	4	28	2	4	28
52	8	5	40	1	1	8
53	9	6	54	0	0	0
54	15	7	105	1	1	15
55	9	8	72	2	2	18
56	7	9	63	3	3	21
57	1	10	10	4	4	4
58	1	11	11	5	5	5
	TOTAL 64		TOTAL 401			TOTAL 185

$$L \text{ average} = \frac{401}{64} = 6.3$$

Round off to 6 so actual average length is 53 mm.

$$S.\ D. = \sqrt{\frac{185}{64}} \quad \sqrt{2.9} = 1.7 \text{ mm}$$

After we have the lengths coded, we move to Column 4, (N x L). This again is an abbreviation for "Number" (Column 2) multiplied by "Coded Length" (Column 3).

In order to find the average tail length, we will divide the total (N x L), that is, the sum of all the (N x L) values in Column 4, by the total (N), that is, the sum of all the (N) values in Column 2. Or written in arithmetical form:-

$$L \text{ average} = \frac{\text{Total } (N \times L)}{\text{Total } (N)} = \frac{198}{39} = 5.1.$$

Rounding this off gives the average coded length of 5.

This corresponds to an average actual length of 61 mm.

Exactly the same procedure is followed for the Carolina chickadee data, yielding the average actual length of 53 mm.

A graph is a convenient way of showing the relationship between measurable phenomena. For instance, a car going 50 miles per hour goes 50 miles in one hour, 100 miles in two hours, and 150 miles in three hours. See Figure 1.

In many, but not all, cases, measurements of similar items, if plotted so as to show the number of times a particular value occurs in the group of items measured, will lie along a curve having the general shape of a bell. This curve is known as the normal, or Gaussian curve. See Figure 2. The peak of the curve is at the average measurement. The larger the number of individual measurements (sample size), the more accurately the average of the sample represents the average value for ALL measurements (population), and the more closely the curve represents the distribution of all measurements of this item.

Our normal curve can now supply us with additional information. The most useful tidbit is called the standard deviation.

68.3% of the tails measured will be found between one standard deviation greater than the average, and one standard deviation less than the average. The interval of average plus or minus 2 standard deviations will contain 95.5% of the measurements. Between plus or minus 3 standard deviations about the average, 99.7% of the measurements will lie. Thus the standard deviation gives us a number which immediately tells us the limits between which we may expect to find MOST of the measurements, while still allowing for the possibility of a genuine, correct, but quite unusual measurement.

Returning to our computation sheets, we refer again to our black-capped chickadee data and Column 5. We calculate "D". This is shorthand

mm	CAROLINA	BLACK-CAPS
48	|	
49	|	
50	卌	
51	卌 ||	
52	卌 |||	
53	卌 ||||	
54	卌 卌 卌	
55	卌 ||||	
56	卌 ||	
57	|	||
58	|	
59		卌 ||
60		卌 卌
61		卌
62		||
63		卌 |||
64		|||
65		||

TALLY

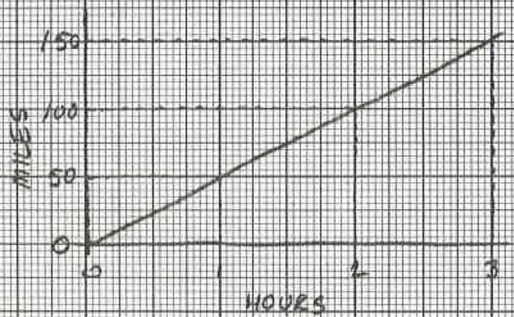

FIG. 1. GRAPH SHOWING DISTANCE TRAVELED AT 50 MPH.

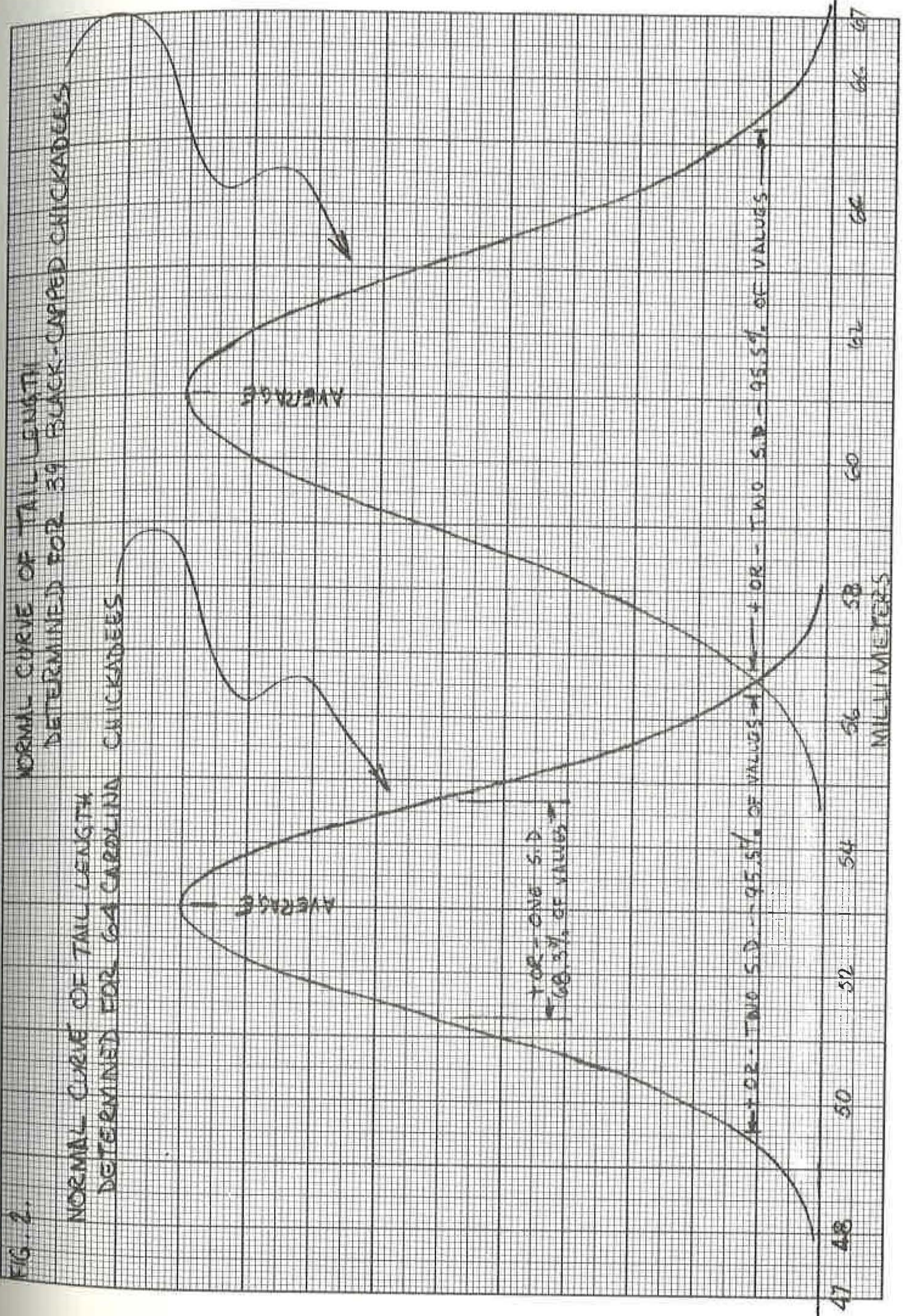

FIG. 2.

notation for "Deviation", and deviation is simply the difference between the particular length and the average length. Therefore, the deviation for the average length of 61 mm. is zero.

Column 6 says "D^2", which is algebraic notation for "D" multiplied by itself, or (D x D).

With Column 6 completed, we now multiply each number in it by the corresponding value for "N" in Column 2, giving us Column 7, labeled ($N \times D^2$).

Our next step is adding all of the values in Column 7 to obtain the value for total ($N \times D^2$) of 163.

The standard deviation may now be obtained by taking the square root of the number obtained when the total ($N \times D^2$) value from Column 7 is divided by the total (N) value from Column 2.

Undoubtedly, taking the square root is the most difficult arithmetic operation in this entire computation. Fortunately, it is also the last arithmetic operation. Square root is denoted by the algebraic symbol $\sqrt{\ }$.

The square root of a number is simply that number which, when multiplied by itself, gives the original number. For instance, 2 x 2 = 4, so 2 is the square root of 4. Again 3 x 3 = 9, so 3 is the square root of 9. The square root of 5 is about 2.24 (rounded off).

If you lack the necessary knowledge to calculate square roots, refer to bibliography reference A for a handbook containing very good tables of square roots.

Our set of data for black-caps gives a standard deviation for tail lengths of 2.1 mm. For Carolinas we get a standard deviation of 1.2 mm.

Going back to our definition for standard deviation, we can now say that 95.5% of black-caps tails will lie at 61 mm. plus or minus 2 x 2.1 mm. In other words, 95.5% will be between 56.8 and 65.2 mm. long. Similarly, we say that 95.5% of Carolina's tails will be at 53 mm. plus or minus 2 x 1.7 mm., that is, they will be between 49.6 and 56.4 mm. long. See Figure 3.

The temptation now is to say with authority that chickadees with tails longer than 56.6 mm. are black-caps, and chickadees with tails shorter than 56.6 mm. are Carolinas. However, the hardest part of the analysis has arrived. It is now mandatory to THINK. How good is your data?

Are the differences in length really due to species, or are they caused by other factors such as wear due to feather age, or perhaps age of the bird? Are they due to difference in banding locations? Are

combinations of these measurements with other measurements a better indicator? Are your samples large enough to be representative? And so forth.

When you have answered all of the questions brought to mind by your study, and still believe what these numbers say to you, THEN you have found your footpath through this forest of figures.

Author's Note:

This paper uses actual measurements, but is not to be construed as anything other than an example of data manipulation.

Bibliography:

A. "Handbook of Mathematical Tables and Formulas"
by R. S. Burington
Handbook Publishers, Inc.
Sandusky, Ohio

B. "Elementary Statistics"
by F. E. Croxton
Dover Publications, Inc.
New York 1959