Characterising the Belemnoid: A Graphical Device for Mapping Descriptive Statistics

by
James R Warren BSc MSc PhD PGCE

Imagine a sunny Summer's day. A flat and close-cropped meadow sits upon a valley flood-plain and is tenanted by sleeping sheep. The sheep lie at random but nevertheless show a clustering as is their gregarious habit. The disposition of the individuals is clear enough, but how are we to summarise their collective position?

We could mark each creature's location with a Cartesian co-ordinate in some mapping system, especially since our meadow conveniently approximates a Euclidean plane. We could then compute average co-ordinates in the ordinal and abscissal normals and pair these as a unique co-ordinate representative of the flock's mean position. Such a point is termed a CENTROID.

In most practical situations it is not, however, sufficient to consider the spatial deployments only. We also require to assign importance or "weight" to each individual in computing the mean position. Individuals with greater weight will tend to skew the mean position in their own directions. The mean position now embodies scalar as well as geometrical realities and is in some senses more representative of the substance of the group. The concept of the weighted spatial average is borrowed from inertial mechanics which set the convention of calling it the MASS CENTROID.

Associated with this MASS CENTROID is a weighted measure of dispersion, some adaptation of standard deviation which, following Stewart and Warntz[1] I shall call the DYNAMIC RADIUS.

The MASS CENTROID or WEIGHTED MEAN CENTER[2] has the co-ordinates (v_x,v_y) and can be computed using:-

$$v_x = \frac{\sum_{i=1}^{n} p_i x_i}{\sum_{i=1}^{n} p_i} \qquad \textbf{Eqn.1}$$

$$v_y = \frac{\sum_{i=1}^{n} p_i y_i}{\sum_{i=1}^{n} p_i} \qquad \textbf{Eqn.2}$$

Further, the DYNAMIC RADIUS is:-

$$r_{xy} = \sqrt{\frac{\sum_{i=1}^{n} p_i d_{ij}^2}{\sum_{i=1}^{n} p_i}} \qquad \textbf{Eqn.3}$$

It is possible to quote r_{xy} in terms involving (v_x,v_y) using:-

$$r_{xy} = \sqrt{\frac{\sum_{i=1}^{n} p_i \left[(x_i - v_x)^2 + (y_i - v_y)^2 \right]}{\sum_{i=1}^{n} p_i}} \qquad \textbf{Eqn.4}$$

Figure One shows the five sheep on our sunlit meadow in their co-ordinate positions annotated with their respective weights in kilograms. The mass centroid is marked as v_x,v_y and the dynamic radius plotted as r_{xy}.

Imagine now that the shepherdess suddenly enters the meadow from behind some trees and calls to her sheep. They are on the hoof. They run toward her, some faster some slower, some with rectilineal precision some wheeling in arcs. It is clear enough why the sheep are running because I have told you but if you were looking down upon the scene from a thousand meters the nature of the attractor and the character of her promise might be inapparent.

Clearly the picture has transformed from the static to the dynamic and the circular graphical device of Figure One now seems inadequate. Equiextensive in all directions it can tell us nothing about the collective direction of the flock.

We are also witnessing a weirder problem, one which would have bemused Virgil or Euclid, though not Dante and several of the Moderns who have followed him: And presumably not you or I. As they travel the sheep spontaneously famish or fatten, vanish permanently or temporarily, and are occasionally reinforced or abandoned by strangers. (After all, they are Space Sheep of the Imagination!)

Can we choose a graphical device simple enough for mapping which will yet statistically summarise the essence of these changes?

A circle is defined by one characteristic line but a quadrilateral by two, permitting the latter to represent an extra element of numerical information. Because a quadrilateral can be made to have only one (identity) axis of rotational symmetry it is also capable of indicating direction without ambiguity.

Let us chose a quadrilateral to plot instead of or as well as the dynamic radius circle. I propose a BELEMNOID (βελεμνον - dart, ειδος - form). This is a dart-like four-sided figure whose apex can be made to point to the successor mass centroid whilst its current mass centroid rests at the vertex of the figure's re-entrant. It is accordingly the case that a set of such belemnoids can point the course of a succession of distributional centroids and can give some impression, without animation and upon a single sheet of paper, of the flow of magnitudes across a surface.

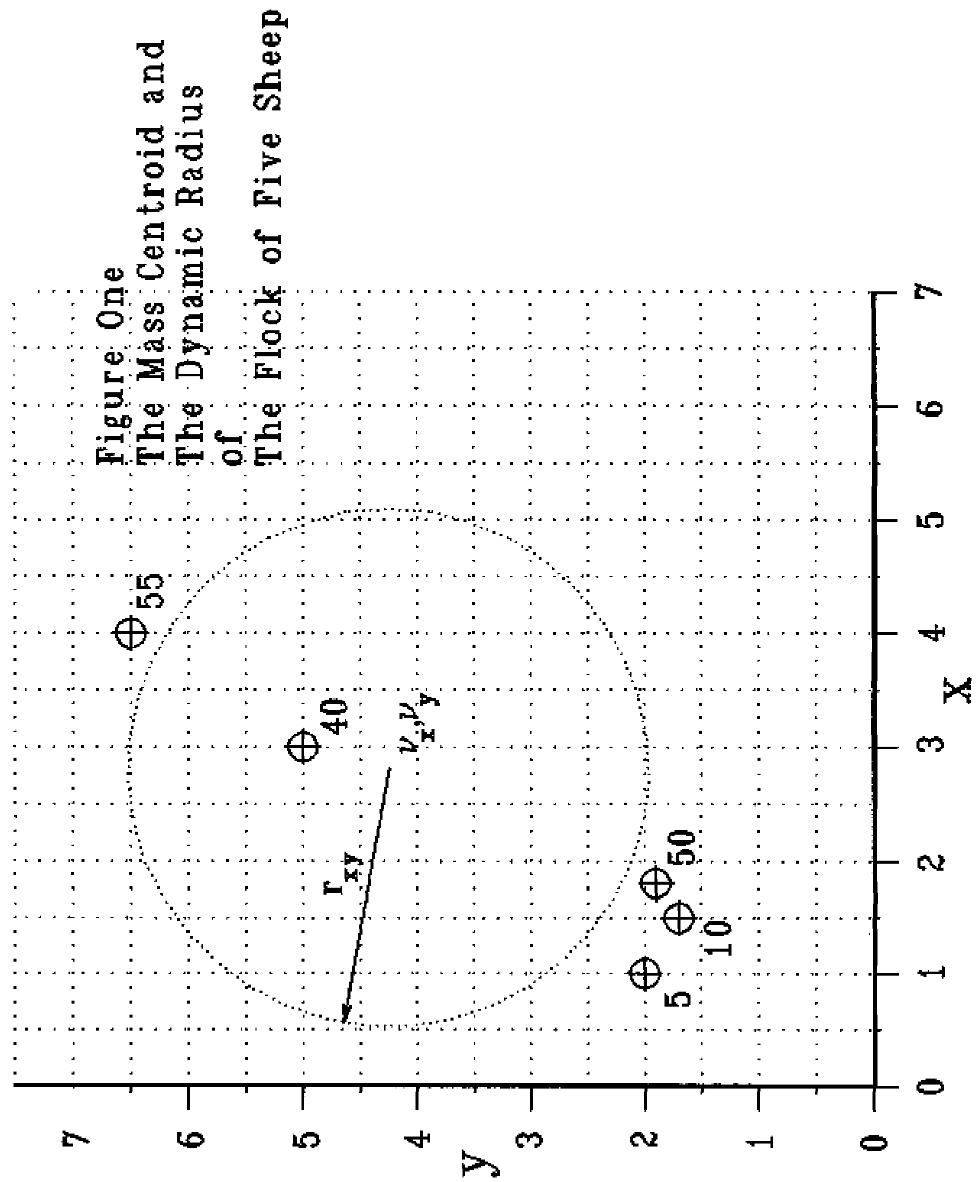


Figure One
The Mass Centroid and
The Dynamic Radius
of
The Flock of Five Sheep
55
40
50
10
5
r xy
X
y

This belemnoid shape is far from original as a graphical pointer. Indeed, cursory examination of the diagrams in this paper, prepared using the standard features of DraftChoice, will discover many examples employed as extension markers. What I believe to be new in this disquisition is the *way* in which the pointers are designed and used.

The belemnoid, being a quadrilateral of one axis can summarise two orthogonal values with its lineaments. I further propose that the axis of symmetry be employed to illustrate the Arithmetic Mean, μ, of the current set of populational magnitudes $p_1,...,p_n$ whilst its basal width represents their (Population) Standard Deviation, σ.

It should be noted that σ and r_{xy} are *not* equivalent, though the two statistics are mathematically related, in particular by summative magnitude. r_{xy} is essentially geometric whilst σ is scalar. It may make sense to compute r_{xy} for each successive data pattern and then if not to plot it, to use its average as a basis for scaling the belemnoids.

The Development of The Standard Belemnoid

Refer to Figure Two. The large stellate figure FHEAG is a pentagram or mullet whose five rays are equally spaced. Accordingly angle FOH is 2π/5 radians and angle FAH π/5 radians because angles at a center are double those at a circumference when on a common subtense.

φ is The Ratio of Phidias given by[3]:-

$$\phi = \frac{\sqrt{5} + 1}{2} \qquad \textbf{Eqn.5}$$

This ratio has a numeric value approximating 1.618033989.

Though of recondite scientific merit, φ has a variety of satisfying mathematical properties and is adopted here as a conventional and convenient basis for graphical design.

The angles of a triangle total π radians and triangle ABC is isosceles. Therefore angle ACB is 2π/5 radians.

By opposite angles ECD is also 2π/5 radians. Angles ACE and BCD are also mutually opposite whilst both pairs of angles; ACB with ECD and ACE with BCD must total 2π radians.

It therefore follows that angle ACE must be 3π/5 radians.

Now our Standard Belemnoid represented by the bold quadrilateral ACEF is of a particular aspect based upon The Ratio of Phidias. In particular, if line CD is unity then AC is φ.

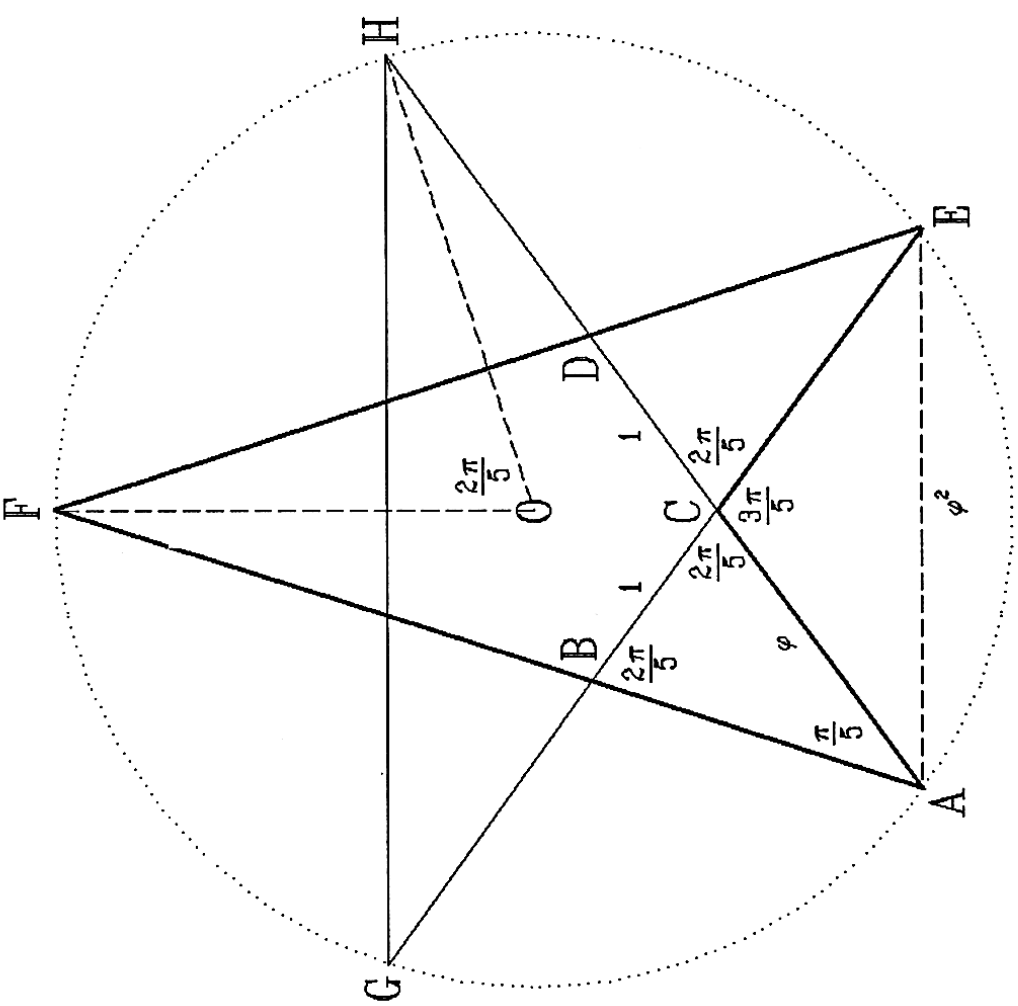
Appropriate applications of Carnot's Theorem confirm that:-

$$AE = \phi^2 \qquad \textbf{Eqn.6}$$

$$FC = \sqrt{5\phi^2 + 4\phi + 1 - (4\phi^2 + 2\phi)\mathrm{Cos}\,\pi/5} \qquad \textbf{Eqn.7}$$

Let us now take the step of identifying the (Population) Standard Deviation σ of a plottable data set with line AE and the data set's Arithmetic Mean μ with the distance FC. Clearly the above relationships will not be general to belemnoids adjusted in that way.

Figure Two
The Development of the Standard Belemnoid from the Pentagram



But let us retain the basal re-entrant angle ACE = 3π/5 = 1.884955592 for *all* the devices we plot. This will prevent high σ - low μ belemnoids collapsing into flat figures of uncertain tendency and retain space at the rear end of our graphical pointer for titling information.

Pivotally, let us identify point C in the apex of the re-entrant with the distributional mass centroid.

We now have to hand a plottable pointing device with both scalar and vector attributes, and one susceptible of intuitive interpretation:-

(1) Two Statistical Moments (or their surrogates) μ and σ are explicit as readily-measurable orthogonal lineaments of the figure

(2) Geometrical Skewness is implied by the position of the mass centroid within the point field of populational entities

(3) The figure is strongly asymmetric normal to its axis of symmetry. It can therefore point to a successor, the next populational mass centroid.

Sizing the Belemnoid

The area of the graphical belemnoid does not intuitively portray any simple statistic. Nevertheless, we may quantify the area for the general case because it may assist in the semi-automation of computer plotting so as to avoid on the one hand absurdly little pointers difficult to find or follow and on the other enormous arrows dominating maps or even transgressing their borders.

We could, of course, fix a linear scale for μ and σ without recourse to area. Such an approach may commend itself in many applications where we could simply say "one map centimeter equals ten thousand units of production" or something like that.

Such an approach, however, requires cartographic insight of which machines have little.

In a practical situation where there are more than two mapped data points the Mean will almost always equal or exceed the Standard Deviation.

Figure Three illustrates this extreme case.

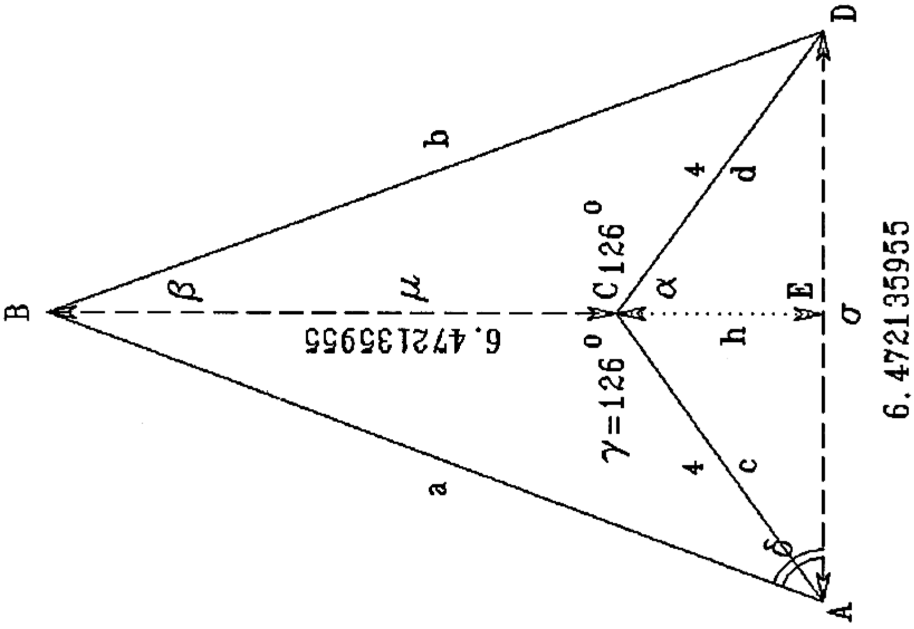
Allow that A is The Area of the Belemnoid, whilst A_{XYZ} is The Area of the General Triangle XYZ. Then:-

$$A = A_{ABD} - A_{ACD} \qquad \textbf{Eqn.8}$$

so:-

$$A = \frac{1}{2} ab \,\mathrm{Sin}\, 2\beta - \frac{1}{2} \mathrm{cdSin}\, 2\alpha \qquad \textbf{Eqn.9}$$

Figure Three
The Belemnoid with Mean equal to Standard Deviation



Therefore:-

$$A = \frac{1}{2}\left[a^2 \operatorname{Sin} 2\beta - c^2 \operatorname{Sin} 2\alpha\right] \qquad \textbf{Eqn.10}$$

Now:-

$$a^2 = \mu^2 + c^2 - 2\mu c \operatorname{Cos}\gamma \qquad \textbf{Eqn.11}$$

whilst:-

$$c^2 = \frac{\sigma^2}{4 \operatorname{Sin}^2 \alpha} \qquad \textbf{Eqn.12}$$

Substitution of Equations Eleven and Twelve into Equation Ten yields:-

$$A = \frac{1}{2}\left[a^2 \operatorname{Sin} 2\beta - \frac{\sigma^2}{4 \operatorname{Sin}^2 \alpha}.\operatorname{Sin} 2\alpha\right] \qquad \textbf{Eqn.13}$$

To express Sin 2β in terms of μ and σ we may first define the auxiliary variable h as The Height of Triangle ACD given by:-

$$h = \sqrt{c^2 - \frac{\sigma^2}{4}} = \sqrt{\frac{\sigma^2}{4}\left(\frac{1}{\operatorname{Sin}^2 \alpha} - 1\right)} \qquad \textbf{Eqn.14}$$

This allows us to quantify Belemnoid Height H=BE as:-

$$H = \mu + \sqrt{\frac{\sigma^2}{4}\left(\frac{1}{\operatorname{Sin}^2 \alpha} - 1\right)} \qquad \textbf{Eqn.15}$$

whereupon:-

$$\operatorname{Sin}\delta = \frac{1}{a}\left[\mu + \sqrt{\frac{\sigma^2}{4}\left(\frac{1}{\operatorname{Sin}^2 \alpha} - 1\right)}\right] \qquad \textbf{Eqn.16}$$

By the Sine Rule this implies that:-

$$\frac{a}{\operatorname{Sin}\delta} = \frac{\sigma}{\operatorname{Sin} 2\beta} \qquad \textbf{Eqn.17}$$

or:-

$$\operatorname{Sin} 2\beta = \frac{\sigma}{a}\operatorname{Sin}\delta \qquad \textbf{Eqn.18}$$

Accordingly:-

$$\text{Sin } 2\beta = \frac{\sigma}{a^2}\left[\mu + \sqrt{\frac{\sigma^2}{4}\left(\frac{1}{\text{Sin}^2\alpha} - 1\right)}\right] \qquad \textbf{Eqn.19}$$

Substitution of Equation Nineteen into Equation Thirteen now allows us to express Belemnoid Area in terms of μ,σ and α only:-

$$A = \frac{1}{2}\left\{\sigma\left[\mu + \sqrt{\frac{\sigma^2}{4}\left(\frac{1}{\text{Sin}^2\alpha} - 1\right)}\right] - \frac{\sigma^2 \text{Sin } 2\alpha}{4\text{Sin}^2\alpha}\right\} \qquad \textbf{Eqn.20}$$

Equation Twenty is the general equation for The Area of the Belemnoid.

This equation yields, for ACDB of Figure Three an Area of 20.94427191 square units in exact agreement with the 20.94427191 found be simple areal subtraction.

By separating roots Equation Twenty may be re-arranged as:-

$$A = \frac{1}{2}\left\{\sigma\left[\mu + \frac{\sigma}{2}\sqrt{\frac{1}{\text{Sin}^2\alpha} - 1}\right] - \frac{\sigma^2}{2} \text{ x } \frac{\text{Sin } 2\alpha}{2\text{Sin}^2\alpha}\right\} \qquad \textbf{Eqn.21}$$

Noting that:-

$$\frac{\text{Sin } 2\alpha}{2\text{Sin}^2\alpha} = \sqrt{\frac{1}{\text{Sin}^2\alpha} - 1} \qquad \textbf{Eqn.22}$$

Equation Twenty-One reduces to:-

$$A = \frac{1}{2}\left\{\sigma\left[\mu + \frac{f\sigma}{2}\right] - \frac{f\sigma^2}{2}\right\} \qquad \textbf{Eqn.23}$$

where f represents either interchangeable side of Equation Twenty-Two.

Accordingly:-

$$A = \frac{1}{2}\left\{\sigma\mu + \frac{f\sigma^2}{2} - \frac{f\sigma^2}{2}\right\} \qquad \textbf{Eqn.24}$$

or:-

$$A = \frac{\mu\sigma}{2} \qquad \textbf{Eqn.25}$$

which is the simplified general equation for The Area of a Belemnoid.

Allow that the radius of the circumscribed circle in Figure Three is the mean of the distributional dynamic radii, r_m.

Clearly, r_m=OA. For scaling purposes a Standard Belemnoid can be computed and plotted relative to this characteristic size. (We are assuming here that computer-optimised rather than manual scaling is being employed).

From Figure Two angle HAE must be π/5 radians since triangle ACE is isosceles and angle ACE is 3π/5 where the angles of a triangle total π radians. Since OA bisects angle BAC, angle OAC is π/10 radians and therefore angle OAE = angle OEA = 3π/10 radians.

Since angle AOE quintisects the circle it is clearly 2π/5 radians and application of Carnot's Theorem yields (in units of CD):-

$$(\theta^2)^2 = r_m^2 + r_m^2 - 2\, r_m \cdot r_m \,\mathrm{Cos}\frac{2\pi}{5} \qquad \textbf{Eqn.26}$$

Setting $\varphi^2 \equiv \sigma$ it is clear that:-

$$\sigma_s = \sqrt{2\, r_m^2\left(1 - \mathrm{Cos}\frac{2\pi}{5}\right)} \qquad \textbf{Eqn.27}$$

Equation Seven defines FC≡μ in terms of φ as:-

$$\mu_s = \sqrt{5\phi^2 + 4\phi + 1 - (4\phi^2 + 2\phi)\,\mathrm{Cos}\frac{\pi}{5}} \qquad \textbf{Eqn.28}$$

From which we may educe that:-

$$\mu_s = \sqrt{5\sigma + 4\sqrt{\sigma} + 1 - (4\sigma + 2\sqrt{\sigma})\,\mathrm{Cos}\frac{\pi}{5}} \qquad \textbf{Eqn.29}$$

But since[3]:-

$$\frac{FO}{OC} = \frac{r_m}{r} = \phi^2 \qquad \textbf{Eqn.30}$$

allowing r to be the radius of the circle intersecting points BCD then:-

$$\mu_s = FO + OC = r_m + r = r_m + \frac{r_m}{\phi^2} \qquad \textbf{Eqn.31}$$

therefore:-

$$\mu_s = r_m \left(1 + \frac{1}{\phi^2}\right) \qquad \textbf{Eqn.32}$$

We may compute the Ratio of the Area of the Standard Belemnoid, A_s, to that of its Circumscribed Circle, A_c, in the following terms:-

$$\frac{A_s}{A_c} = \frac{\mu_s \sigma_s}{2\pi\, r_m^2}$$

$$= \frac{r_m}{2\pi\, r_m^2} . \left(1 + \frac{1}{\phi^2}\right) . \sqrt{2\, r_m^2 \left(1 - \mathrm{Cos}\frac{2\pi}{5}\right)} \qquad \textbf{Eqn.33}$$

$$\therefore \frac{A_s}{A_c} = \left(1 + \frac{1}{\phi^2}\right) . \frac{\sqrt{2\, r_m^2 \left(1 - \mathrm{Cos}\frac{2\pi}{5}\right)}}{2\pi\, \mathrm{r_m}} \qquad \textbf{Eqn.34}$$

Setting r_m to unity allows this ratio to reduce to:-

$$\frac{A_s}{A_c} = \left(1 + \frac{1}{\phi^2}\right) . \frac{\sqrt{2\left(1 - \mathrm{Cos}\frac{2\pi}{5}\right)}}{2\pi} \qquad \textbf{Eqn.35}$$

$$= 0.258562878$$

This ratio will *not* of course apply to plotted belemnoids representing actual statistics.

Once the plottable length of the Standard Belemnoid medial μ_s line has been computed according to Equation Thirty-Two it will probably be desired to calibrate it against a magnitude. Unless a round-number calibratory value is to be assigned on a manual or automatic basis it may prove convenient to have a machine assign it the mean of the distributional μ values and scale the distributional belemnoids accordingly.

I recommend that the Standard Belemnoid is plotted at the lower right-hand corner of the map for reference purposes with a caption indicative of scale.

Programmatically, this may be accomplished by setting Vertex E of Figure Two to the lowest Northing and highest Easting and back-plotting the Standard Belemnoid appropriately.

Let us analyse the Belemnoid Area A of Equation Twenty-Five in statistical terms. By squaring all terms of Equation Twenty-Five we obtain:-

$$A^2 = \frac{\mu^2 \sigma^2}{4} \qquad \textbf{Eqn.36}$$

σ^2 is the (Population) Variance. Accordingly:-

$$A^2 = \frac{(\Sigma\, p\,)^2}{n^2} x \frac{1}{4} x \Sigma\left(p - \frac{\Sigma\, p}{n}\right)^2 x \frac{1}{n} \qquad \textbf{Eqn.37}$$

therefore:-

$$A^2 = \frac{(\Sigma\, p\,)^2}{4\, n^3} \Sigma\left(p - \frac{\Sigma\, p}{n}\right)^2 \qquad \textbf{Eqn.38}$$

or:-

$$A = \sqrt{\frac{(\Sigma\, p\,)^2}{4\,n^3}\,\Sigma\left(p - \frac{\Sigma\, p}{n}\right)^2} \qquad \textbf{Eqn.39}$$

Equation Thirty-Nine indicates that whilst Area is proportional to the Summated Magnitudes of the cluster, Σp, it is roughly proportional to the -1.5 power of the Number of Cluster Entities, n.

This will cause the belemnoid to show a strong tendency to shrink with increasing cluster population and researchers who wish to evade such effects should either certify that cluster populations are comparable or else install appropriate mathematical compensators.

Readers are further reminded that the descriptive parameters Arithmetic Mean and Standard Deviation may not be statistically significant for their own research purposes when the number of cluster objects is few and their dispersion relatively great. Further study of this problem requires tenable assumptions to be made about the nature of the distributions of object magnitudes and Appendix One elaborates some possible approaches to this problem.

The Average Belemnoid Direction

As aforementioned, our belemnoids are designed to illustrate the migration of the mass centroid by pointing along the cartographic azimuth to the successor centroid.

Such pointing is of course impossible at the last cluster whose successor is non-existent or unknown. At the final centroid we shall orient our belemnoid to point the average trajectory of the preceding k-1 migrations.

Determining the average azimuth is more problematic than it sounds. Clearly a simple mean of the angles will not do. In an absurd situation we would compute that the mean direction between 1° (almost due North) and 359° (almost due North) was 180° (exactly due South).

In fact none of the usual statistical averages will represent the collective tendency with any accuracy or reliability and we must revert to classical analytic geometry for a solution. That solution establishes the mean of a swarm of angles as the angle of their vectorial resultant where the length of each vector component is taken to be unity.

If this Resultant Angle, θ_m, averages the Angles θ_i

then:-

$$\theta_m = j\pi + \mathrm{ArcTan}\left(\frac{\sum_{i=1}^{k-1}\mathrm{Sin}\,\theta_i}{\sum_{i=1}^{k-1}\mathrm{Cos}\,\theta_i}\right) \qquad \textbf{Eqn.40}$$

where the Augmentation Indicator, j, may logically be determined from the signs of $\Sigma\mathrm{Sin}\theta_i$ and $\Sigma\mathrm{Cos}\theta_i$ using Table One:-

		ΣCos	
		+	-
ΣSin	+	0	1
	-	2	1

Table One
The Choice of Augmentation Indicator, j

Let us illustrate the computation of the Resultant Angle by recourse to the seven arbitrary angles listed in Table Two:-

Angle (Degrees)	Angle (Radians)	Cos	Sin
20	.3490659	.9396926	.3420201
25	.4363323	.9063078	.4226183
40	.6981317	.7660444	.6227876
45	.7853982	.7071068	.7071068
46	.8028515	.6946584	.7193398
215	3.7524579	-.8191520	-.5735764
217	3.7873645	-.7986355	-.6018150
Totals		+2.3960224	+1.6584811
Sign		1	1
j			0
ArcTan			.6054590
Mean (Radians)			.6054590
Mean (Degrees)			34.6902449

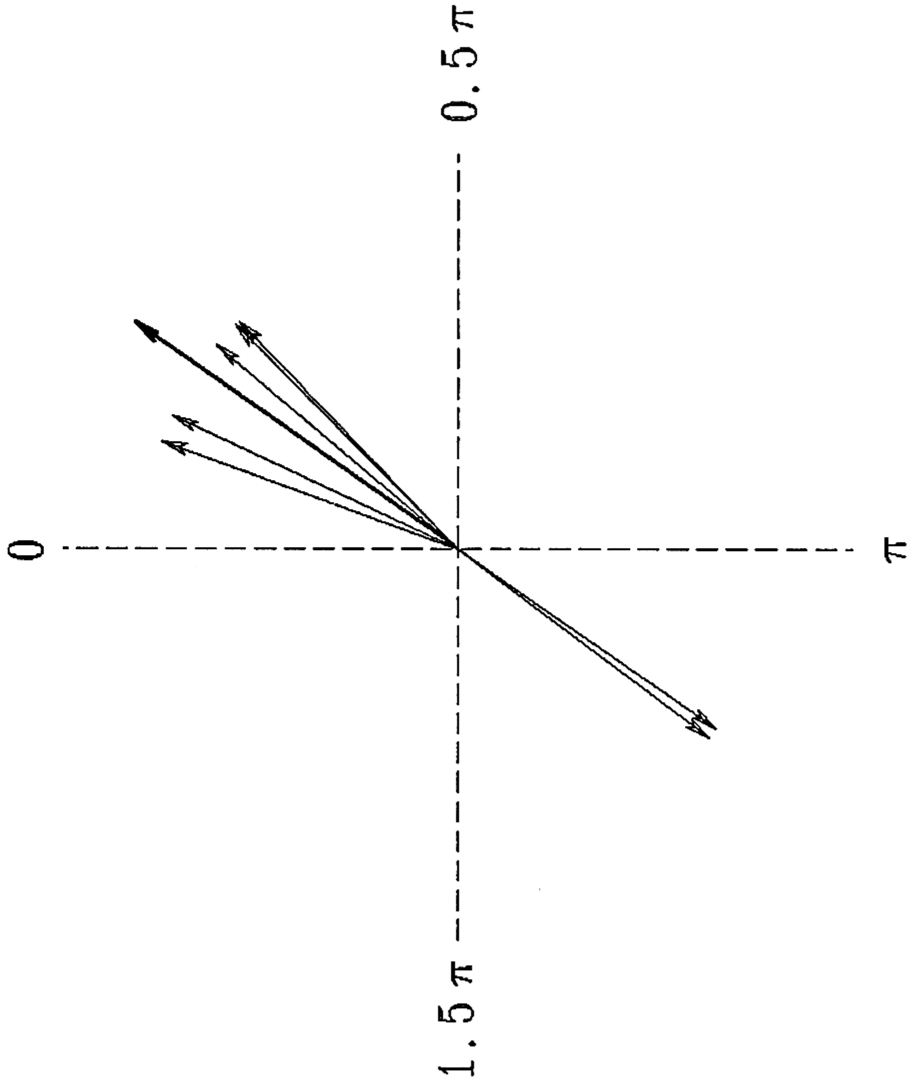
Table Two
The Resultant of a Set of Angles

These angles and their resultant mean are plotted as a rose in Figure Four. It is clear that this average is immune to the biasing effects that diametric components can exert upon some other means.

Clearly, vector analytic techniques could be made to embody in this average any statistical values, whether quantitative by the Mean or qualitative by the Standard Deviation. I have not incorporated these because the belemnoid is of course representing them.

This final pointer consummates the tendency of the entire migratory history and is at the center of any forecasting methodology.

Figure Four
Seven Test Angles and their Vectorial Resultant as a Mean



The Cartographic Program

1 Record, or determine from records,
The Magnitude of Interest at precisely-located
Grid Points at two or more precisely-timed
Polling Events

2 For each Event compute the distributional
Mass Centroid and Dynamic Radius,
and for the collection of Event Entities
The Arithmetic Mean of Magnitudes and their
(Population) Standard Deviation

3 Compute the Arithmetic Mean of the Dynamic Radii, r_m.
Then compute the μ_s attaching to the Standard Belemnoid
using Equation Thirty-Two or its surrogate

4 Use a Search to locate the Lowest Northing, N_{min},
and Highest Easting, E_{max}, and express them in
compatible co-ordinates

5 Compute σ_s for the Standard Belemnoid using
Equation Twenty-Seven or its surrogate

6 Use:-

$$h_s = \frac{\sigma}{2} \mathrm{Tan} \frac{\pi}{5}$$

to define the Auxiliary Height, h_s

7 Erect the Standard Belemnoid Plot Vertices:-

(E_{max},N_{min}) $(E_{max}-\sigma_s/2,N_{min}+h_s)$
$(E_{max}-\sigma_s,N_{min})$ $(E_{max}-\sigma_s/2,N_{min}+h_s+\mu_s)$

8 μ_s and σ_s are expressed in Map Units:
Essentially whatever d_{ij} was measured in when
the dynamic radii were calculated.
It is required to determine a Scaling Constant K,
which will relate Cartographic distances to
Value Magnitudes. Accordingly compute the
Mean of Arithmetic Means, M_μ, and divide it by μ_s
to obtain K

9 Compose the plottable caption:
"Standard Belemnoid Mu = " M_μ "<UNITS NAME>"

10 For the First Event Distribution adjust the Distributional statistics using:-

$$\mu=\mu/K$$

$\sigma=\sigma/K$

to convert Magnitude Units to Map Units

11 Erect the Unoriented Belemnoid Vertices about the Distributional Mass Centroid as:-

(v_x,v_y)

$(v_x,v_y+\mu)$

$(v_x+\sigma/2,v_y-h)$

$(v_x-\sigma/2,v_y-h)$

12 Repeat Steps 10 and 11 for all k Event Distributions

13 Compute the Directional Angle, θ_1, of the Second Mass Centroid with respect to the First Mass Centroid using:-

$\theta_1 = \text{ArcTan}\,[(v_{x2}-v_{x1})/(v_{y2}-v_{y1})]$

If $v_{y2}-v_{y1}$ is negative increment θ_1 by π before application

14 Orient each of the Four Belemnoid Vertices using:-

$X = x\text{Cos}\theta_1 + y\text{Sin}\theta_1$

$Y = y\text{Cos}\theta_1 - x\text{Sin}\theta_1$

Where lower case co-ordinates are unrotated.
Because rotation is centered about v_x,v_y the centroid co-ordinates remain untransformed whilst the rotated apical co-ordinates are yielded by:-

$X = v_x + \mu\text{Sin}\theta_1$

$Y = v_y + \mu\text{Cos}\theta_1$

Meanwhile the rotated basal co-ordinates are given by:-

$X = v_x + (\sigma/2)\text{Cos}\theta_1 - h\text{Sin}\theta_1$

$Y = v_y - (\sigma/2)\text{Sin}\theta_1 - h\text{Cos}\theta_1$

and by:-

$X = v_x - (\sigma/2)\text{Cos}\theta_1 - h\text{Sin}\theta_1$

$Y = v_y + (\sigma/2)\text{Sin}\theta_1 - h\text{Cos}\theta_1$

15 Repeat Steps 13 and 14 for all except the last of the Event Distributions 1 through k-1

16 Find the Vectorial Resultant of the Available 1 to k-1 Directional Angles, θ. Call this average θ_m

17 Orient the last of the Distributional Belemnoids to the Average Direction of the others using:-

$X = x\text{Cos}\theta_m + y\text{Sin}\theta_m$

$Y = y\text{Cos}\theta_m - x\text{Sin}\theta_m$

on each of its four vertices

18 Manually plot, or employ a selected
Graphical Software Facility to graph
The k Swarms of Distributional Points,
The k Mass Centroids and their pivoting
Belemnoids.
Plot all required annotations.
Plot the Standard Belemnoid.

An Example

Let us examine how mining activity swarmed across a fictitious historical orefield.

Figure Five maps the sixteen metal mines involved with their productions of opprobrium (lets us say) in metric tonnes.

Production was sampled at ten-yearly intervals in 1960, 1970, 1980 and 1990. Production is shown against Year and Mine in Table Three.

MINE	GRID REFERENCE		YEAR			
	EASTING	NORTHING	1960	1970	1980	1990
ABLE	3	2	457			
BAKER	5	3	330	842		
CARTER	5.5	1.5	425			
DEACON	10.5	5		223		
EDWARDS	10.5	6.5		551	94	
FRANCIS	11	6		846	374	
GRANVILLE	11.5	5.5		967		
HARWELL	8	13.5			1	
ISMAY	7.5	14			4730	
JARROLD	7	13.5			323	
KELD	5	10				14
LATIMER	3	10				516
MORRIS	4	9.5				380
NUGENT	3	9				610
OSWALD	4	8.5				64
PYE	5	8				342

Table Three
The Production of Opprobrium in Metric Tonnes
by Mine and Year

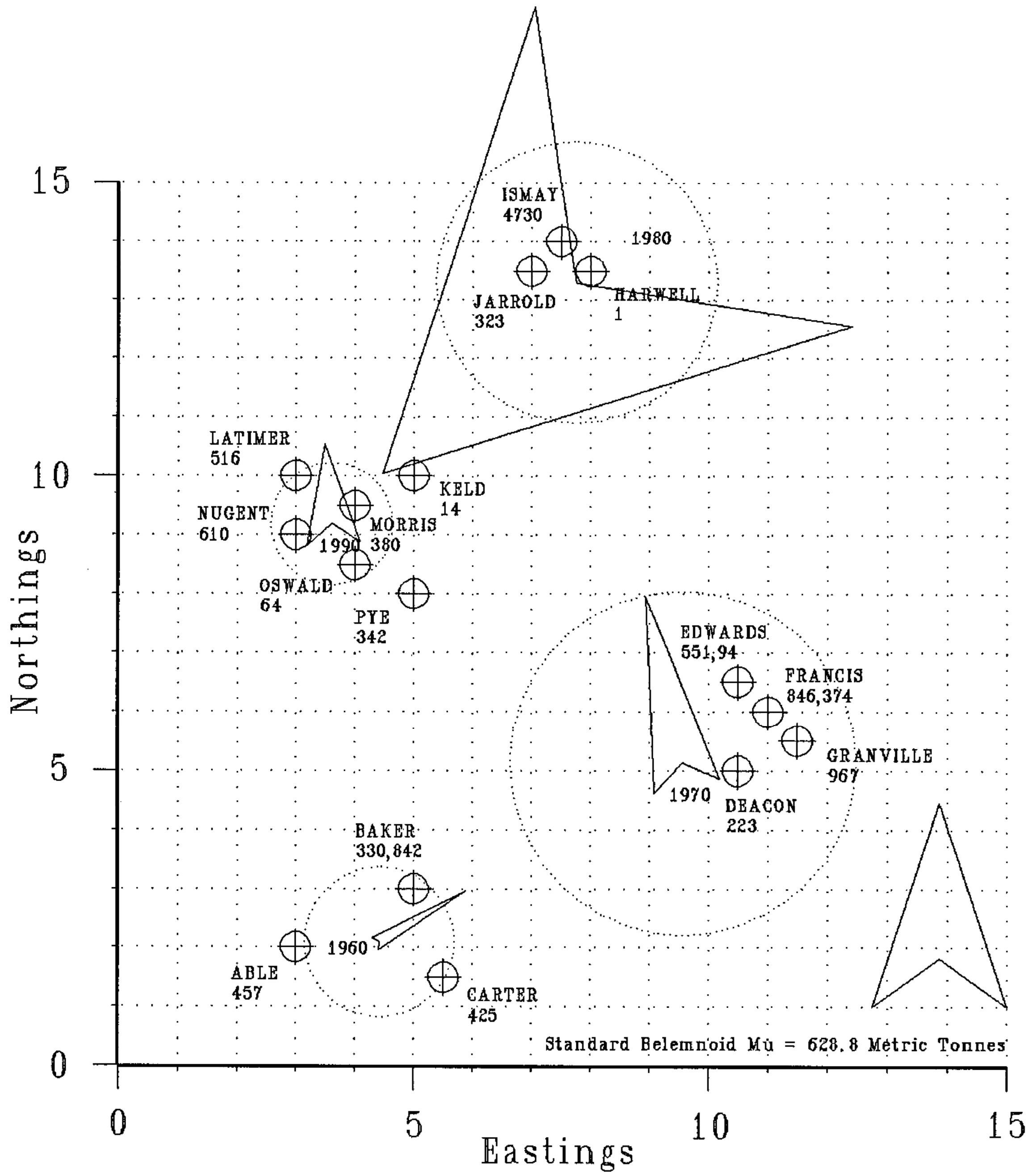


Figure Five
Migration of the Mass Centroid and
Changes in Production through Time
For the Mines of Opprobrium

The table shows the tendency of the mines conveniently to segregate into clusters separated in time as well as space. It can be seen, however, that Wheal Baker is still very much in production ten years after 1960, whilst Wheal Edwards and Wheal Francis are both producing in consecutive sampling Events 1970 and 1980.

The mass centroids for the four year-clusters are the centers of the circles whose diameters represent the respective dynamic radii.

The exaggerated area of the 1980 belemnoid illustrates the way in which combined high value mean and high value standard deviation can give a false impression of dominance. Nevertheless, the belemnoids give useful summaries with practiced interpretation and convey an excellent feeling for location and migration. Furthermore their aspect ratios aptly convey the relationships between respective means and standard deviations.

The final belemnoid pointing the resultant vector gives an excellent counter-intuitive report of the overall direction of values.

It can be appreciated that should the workings and their associated values be more promiscuously mixed in time and space then directional devices such as these belemnoids may constitute the only useful agents of developmental interpretation.

Notation

α	Half the Re-Entrant Angle of the Belemnoid
a	A Belemnoid Long Edge
A	The Area of the Belemnoid
A_c	The Area of the Circle Circumscribing the Standard Belemnoid
A_s	The Area of the Standard Belemnoid
A_{XYZ}	The Area of Triangle XYZ
β	Half the Apical Angle of the Belemnoid
b	A Belemnoid Long Edge
γ	The Angle Subtended by a or b
c	A Belemnoid Short Edge
δ	The Complete Basal Angle of a Belemnoid
d	A Belemnoid Short Edge
d_{ij}	The Displacement between a Location and its Mass Centroid
E_{max}	The Highest Easting
f	A Special Trigonometric Function of α
h	The Height of the Belemnoid Re-Entrant
h_s	The Re-Entrant Height for the Standard Belemnoid
H	The Total Belemnoid Height
θ_i	The Directional Angle between Entity Cluster i and Cluster i+1
θ_m	The Vectorial Resultant Mean Directional Angle
j	The Angular Augmentation Indicator
k	The Number of Entity Clusters
K	The Scaling Constant
μ	The Arithmetic Mean
μ_s	The Mean Characterising the Standard Belemnoid

M_μ	The Mean of Arithmetic Means
v_x	The Abscissal Co-ordinate of the Mass Centroid
v_y	The Ordinal Co-ordinate of the Mass Centroid
v_{x1}	The Centroid Abscissal Co-ordinate for the First Point
v_{x2}	The Centroid Abscissal Co-ordinate for the Second Point
v_{y1}	The Centroid Ordinal Co-ordinate for the First Point
v_{y2}	The Centroid Ordinal Co-ordinate for the Second Point
n	The Number of Values
N_{min}	The Lowest Northing
π	The Ludolphine Constant
p_i	The ith Locational Value ("Weight") (Populational Magnitudes)
r	The Radius of a Pentagram's Inner Circle
r_m	The Mean of the Distributional Dynamic Radii
r_{xy}	The Dynamic Radius
σ	The Population Standard Deviation
σ_s	The Standard Deviation Characterising the Standard Belemnoid
φ	The Ratio of Phidias
x	The Unrotated Abscissal Co-ordinate
x_i	The Abscissal Co-ordinate (Easting) of Value i
X	The Rotated Abscissal Co-ordinate
y	The Unrotated Ordinal Co-ordinate
y_i	The Ordinal Co-ordinate (Northing) of Value i
Y	The Rotated Ordinal Co-ordinate

<u>References</u>

1 "Macrogeography and Social Science"
JQ Stewart and W Warntz 1958
The Geographical Review
V48 pp 167-184

2 "Statistics in Geography"
Second Edition 1985
David Ebdon
Blackwell of Oxford
ISBN 0-631-13688-6

3 "The Divine Proportion"
HE Huntly 1970
Dover Publications of New York
ISBN 0-486-22254-3

APPENDIX ONE

Sample Sizes and Confidence Limits for Small Samples

Throughout the text of the main paper I have assumed a requirement to compute a Population Standard Deviation σ according to:-

$$\sigma = \sqrt{\frac{\Sigma(x - \mu)^2}{n}} \qquad \textbf{Eqn.41}$$

In many research contexts investigators are likely to prefer the Sample Standard Deviation:-

$$s = \sqrt{\frac{\Sigma(x - \mu)^2}{n - 1}} \qquad \textbf{Eqn.42}$$

You can see that so long as the Number of Objects in the Data Cluster (or Sample Size), n, is large then the difference between σ and s will be negligible.

When we are working with rare objects there is however likely to be an appreciable difference between the two flavors of standard deviation. Experimental scientists and other students of mass action can create or at least gather a sufficient number of data to circumvent uncertainties due to small amounts. Historiographers must frequently be more careful: After all you cannot create anew extra Elizabethan sea battles or Victorian lead mines to suit your mathematical convenience.

One context in which the Sample Standard Deviation is convenient is the assessment of confidence limits. A confidence limit allows us to specify the degree of accuracy with which a Population Arithmetic Mean, μ, has been determined given a known Sample Standard Deviation, s, and whilst allowing for a limited supply of data.

In order to apply this treatment we must make some assumption about the character of the data distribution and that assumption may itself be substantiated or denied by other specialised tests which we shall not explore here.

We shall assume without qualification that our data magnitudes vary randomly about their population arithmetic mean. Accordingly, when there are more than about thirty data values we can apply The Theory of The Gaussian Distribution to estimating confidence limits. In our work however, and certainly for the Mines of Opprobrium example, our data are far fewer and we shall accordingly need to use The Theory of The Student's t-Distribution.

A confidence limit allows you to make statements like "I am 95% sure that the overall population mean would be in the range of the sample mean give or take d units if I had enough data to fix that population mean for certain (i.e. an infinite number of data)". It is therefore the case that the confidence limit is a measure of the *quality* of the mean, and by implication of the other descriptive statistics.

We can approach the confidence limit in two ways:-

(a) Compute a Normative Sample Size, n_r.
This is the number of Data magnitudes we

should have if we require a certain accuracy

(b) Compute the actual Confidence Limit, L. This specifies what accuracy we have *actually* achieved given the restrictions which apply

The Normative Sample Size, n_r, is given by:-

$$n_r = entier \left[\left(\frac{st_{\gamma,v}}{d}\right)^2\right] + 1 \qquad \textbf{Eqn.43}$$

and The Confidence Limit, L, by:-

$$L = m \pm t_{\gamma,v}\frac{s}{\sqrt{n}} \qquad \textbf{Eqn.44}$$

The parameter $t_{\gamma,v}$ is The Student's t-Value associated with a Confidence Interval of γ and v=n-1 Degrees of Freedom. $t_{\gamma,v}$ may be manually or automatically read from tables with or without interpolation or it may be computed algorithmically.

I shall example the use of Equations Forty-Three and Forty-Four by reference to the three 1960 Mines of Opprobrium, whose respective productions were 457, 330 and 425 tonnes.

Let us choose a γ=95% Confidence Interval which will give us a 1 in 20 chance of being wrong when we say that the true mean is in the range m±d. Because the Degrees of Freedom is equal to the number of values less one v must be two. The critical $t_{\gamma,v}$ for these parameters was read from a table as 4.303.

Meanwhile m=404 and s=66.05300902. (Note the large difference between this latter figure and σ=53.93205602).

Suppose that we wished to determine the mean within ±10 tonnes. Then:-

$$n_r = entier \left[\left(\frac{66.05300902x4.303}{10}\right)^2\right] + 1 = 808$$

If we must be content with what we have, but wish to know the range within which the true mean has a 95% chance of lying then:-

$$L = m \pm t_{\gamma,v}\frac{s}{\sqrt{n}}$$
$$= 404 \pm 164.0980141$$
$$= 239.9019859 \text{ or } 568.0980141$$

It can now be appreciated that there is some sense in which the mean of these three numbers is "not worth knowing".

In any event, if we need 808 items to compute an accurate mean we are well within the province of the normal distribution and can there apply The Theory of The Gaussian Distribution.

A full set of these statistics for all four clusters and for the nineteen yields forming the total output of the sixteen mines is reproduced in Table 1A:-

Year	1960	1970	1980	1990	19 Yields
Sample Mean	404	685.8	1104.4	321	636.2632
Sample Sd	66.0530	300.6205	2032.698	239.0674	1116.6222
t (95%)	4.303	2.776	2.776	2.571	2.101
d (tonnes)	50	50	50	50	50
Normative Sample Size	33	279	12737	152	2202
Upper Confidence Limit	568.0980	1059.0098	3627.9233	571.9266	1174.4778
Lower Confidence Limit	239.9020	312.5902	-1419.1233	70.0734	98.0485

Table 1A Confidence Statistics for Mine Clusters

Readers can readily apprehend the preposterous numbers of mines required in each polling year in order to estimate a mean within ±50 tonnes of the true value, as well as the uselessly wide bands of uncertainty associated with the estimate.

The major source of trouble is of course 1980 which includes the prolific Wheal Ismay and the failed trial hole Wheal Harwell. The uncertainties inherent in that episode propagate to the summative statistics for the nineteen yields invalidating any broad conclusions which might have been reached from those.

It is now even more firmly established that data magnitudes must be made concordant by some grouping criterion before summative statistics are computed.

Notation for Appendix One

γ	The Confidence Interval
d	The Required (±) Accuracy in Units
L	The (Upper or Lower) Confidence Limit
μ	The Population Mean
m	The Sample Mean
v	The Degrees of Freedom
n	The Number of Data
n_r	The Required Sample Size
σ	The Population Standard Deviation
s	The Sample Standard Deviation
$t_{\gamma,v}$	The Student's t-Value associated with γ and v
x	A Data Value