AMATYC National Conference

Long Beach, California

November 14, 1996

The Characterization of Data

Janet Heath, M.S.

William Pletsch, Ph.D.

Presentation: The Characterization of Data
Presenter: Dr. William Pletsch

Mathematics reform has come a long way over the last few years. Once the symbolical was the mainstay of traditional instructional methods. A few years ago, the Calculus Consortium based at Harvard, as well as other progressive groups, added the graphical and numerical as additional foci to allow a triangulation on the abstract concept. This was formulated into a major axiom of reform known as the Rule of Three. The Rule of Three was often augmented with a verbal approach and in time evolved to the Rule of Four which includes the verbal.

The Rule of Four: each mathematical concept should be developed descriptively, numerically, graphically, and symbolically.

Today, much progress has been made with the graphical, but less so with the descriptive and the numerical. Yet it is these very two approaches that in the end provide the most powerful tools for getting at the symbolic and simultaneously the underlying concept. The numerical points the way to the symbolical, which in turn points to the concept. The descriptive points directly to the concept itself. For instance the structure of a numerical difference quotient where x = 2 and the change in x is 0.1 for the function x^2, suggests very strongly the abstract difference quotient where the change in x is h for an arbitrary function $f(x)$. The description of what has happens between successive points in a linear data set, “The change is proportional”, points directly to the fundamental essence of the concept of linearity.

This workshop will show how the characterization of data sets can help to develop the numerical and the descriptive aspects of mathematics instruction as well as the symbolic and graphical. The workshop will do so using a group learning format, will be technology intensive, and will stress hands-on experimentation. The graphical and symbolical approaches will not be neglected as the Rule of Four will be fully integrated into the session.

The structure of a concrete manipulation frequently parallels the structure of the fully abstract symbolics of a concept. In this workshop, it will be demonstrated how the structure of a data set can enhance the understanding of a general concept and how an elementary manipulation can illuminate the abstract general symbolics. A simple example is that of an exponential data set. We begin even more simply, with a finite geometric sequence: 2, 6, 18, 54. A glimpse of the way the concrete numerical approach parallels the abstract symbolics is summarized in the table below.

	Concrete	Abstract
Geometric Sequence	2, 6, 18, 54	$a, ra, r^2a, \ldots, r^{n-1}a$
Iterative Step	2(3) = 6	$a_i r = a_{i+1}$
Exponential Function	$P(t) = 2(3)^t$	$P(t) = ar^t$

The geometric sequence can be converted to a data set of ordered pairs where the x-coordinate is the place in the sequence. Here the set is (1, 2), (2, 6), (3, 18), and (4, 54). Such a set can be plotted and otherwise manipulated on a graphing calculator or computer.

Quadratic, polynomial, exponential, Gaussians, exponential raised to polynomial powers, and sinusoidal data sets all point the way to the fundamental properties of, respectively, quadratic, polynomial, exponential, Gaussians, exponentials raised to polynomial powers, and sinusoidal functions. Quadratic, polynomial, exponential, Gaussians, exponentials raised to polynomial powers, and sinusoidal data sets can all be recognized and characterized through simple tests. A thorough understanding of one of these tests through verbal descriptions inevitably forces an increase in understanding of the class of functions indicated by the test. The third difference quotient of a cubic data set is a constant. Why? The third difference quotient of the cubic function that fits the cubic data set is a constant. Why? An analysis of these quotients indicates that each successive difference quotient knocks the power down one and so the cubic passes through the quadratic to the linear to, finally by the third iteration, the constant. Thus these tests can be examined to determine how their action sheds light on their respective class of functions.

Both the paralleling of concrete numerics and abstract symbolics demonstrated earlier and the simple tests discussed just above are manifested by the characterization of data. The paralleling of concrete numerics and abstract symbolics occurs extremely frequently in the course of mathematical investigations. Until recently this was not widely recognized, in part because there was no reason for doing so. The calculations were often too arduous to be meaningful. Similarly, the patterns indicated by the data tests were tedious to compute and of little practical value. Of course with the advent of the graphing calculator and computer all this has changed. Technology greatly facilitates the computations and allows application of the result. The second ratio of a Gaussian data set yields a constant but before technology this fact would be trying to discover and difficult to utilize. Today, with technology it is easy to discover (first graph the data which will suggest the Gaussian test) and easy to utilize (set up a matrix equation and solve to find the Gaussian that fits the data). Thus again we arrive at the conclusion so often arrived at before, that we are in an exciting age of mathematical exploration enabled by the new technologies.

EXPERIMENT

A ball is dropped from a tower 256 feet tall.

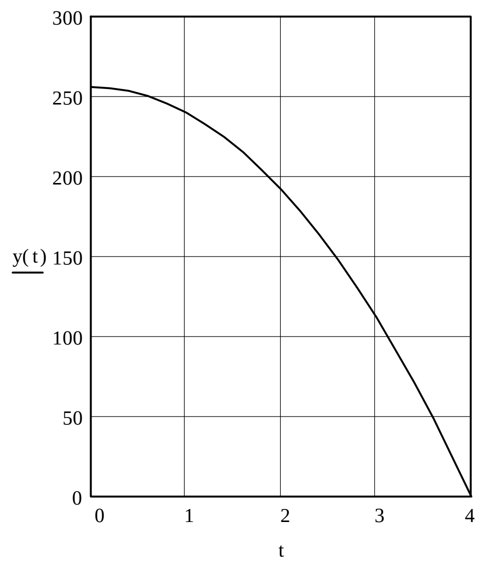
NUMERICAL

Data set

t	y(t)
0	256
0.2	255.36
0.4	253.44
0.6	250.24
0.8	245.76
1	240
1.2	232.96
1.4	224.64
1.6	215.04
1.8	204.16
2	192
2.2	178.56
2.4	163.84
2.6	147.84
2.8	130.56
3	112
3.2	92.16
3.4	71.04
3.6	48.64
3.8	24.96
4	0

GRAPHICAL

A Mathematical Picture



ALGEBRAIC

$y(t) = -16t^2 + 256$

DESCRIPTIVE

The ball drops 32 feet faster every second from 256 feet.

LAB: Power Function Difference Quotients

I) Compute the indicated difference quotients (**DQ**) for the given function by completing the tables.

A) f(x) = x

x	**f(x)**	**DQ1**	**DQ2**	**DQ3**
1	1	1	0	0
2	2	1	0	0
3	3	1	0	0
4	4	1	0	0
5	5	1	0	-
6	6	1	-	-
7	7	-	-	-

B) g(x) = x^2

x	**g(x)**	**DQ1**	**DQ2**	**DQ3**
1	1	3	2	0
2	4	5	2	0
3	9	7	2	0
4	16	9	2	0
5	25	11	2	-
6	36	13	-	-
7	49	-	-	-

C) $h(x) = x^3$

x	h(x)	DQ1	DQ2	DQ3
1	1	7	12	6
2	8	19	18	6
3	27	37	24	6
4	64	61	30	6
5	125	91	36	-
6	216	127	-	-
7	343	-	-	-

D) $i(x) = x^4$

x	i(x)	DQ1	DQ2	DQ3	DQ4
1	1	15	50	60	24
2	16	65	110	84	24
3	81	175	194	108	24
4	256	369	302	132	-
5	625	671	434	-	-
6	1296	1105	-	-	-
7	2401	-	-	-	-

II) What is the fourth difference quotient (**DQ5**) of $\mathbf{x^5}$? **Hint**: Use the above tables to spot a pattern.

III) What is the n^{th} difference (**DQn**) of $\mathbf{x^n}$?

LAB: Ratios

I) Compute the indicated ratios for the given function by completing the tables.

A) f(x) = exp(x) = e^x

x	**f(x)**	**Ratio1**	**Ratio2**	**Ratio3**
1	e^1	e^1	1	1
2	e^2	e^1	1	1
3	e^3	e^1	1	1
4	e^4	e^1	1	1
5	e^5	e^1	1	-
6	e^6	e^1	-	-
7	e^7	-	-	-

B) g(x) = exp(x^2)

x	**g(x)**	**Ratio1**	**Ratio2**	**Ratio3**
1	e^1	e^3	e^2	1
2	e^4	e^5	e^2	1
3	e^9	e^7	e^2	1
4	e^{16}	e^9	e^2	1
5	e^{25}	e^{11}	e^2	-
6	e^{36}	e^{13}	-	-
7	e^{47}	-	-	-

C) $h(x) = \exp(x^3)$

x	h(x)	Ratio1	Ratio2	Ratio3	Ratio4
1	e^{1}	e^{7}	e^{12}	e^{6}	1
2	e^{8}	e^{19}	e^{18}	e^{6}	1
3	e^{27}	e^{37}	e^{24}	e^{6}	1
4	e^{81}	e^{61}	e^{30}	e^{6}	-
5	e^{125}	e^{91}	e^{36}	-	-
6	e^{216}	e^{127}	-	-	-
7	e^{343}	-	-	-	-

D) $i(x) = \exp(x^4)$

x	i(x)	Ratio1	Ratio2	Ratio3	Ratio4
1	e^{1}	e^{15}	e^{50}	e^{60}	e^{24}
2	e^{16}	e^{65}	e^{110}	e^{84}	e^{24}
3	e^{81}	e^{175}	e^{194}	e^{108}	e^{24}
4	e^{256}	e^{369}	e^{302}	e^{132}	-
5	e^{625}	e^{671}	e^{434}	-	-
6	e^{1296}	e^{1105}	-	-	-
7	e^{2401}	-	-	-	-

II) What can you say about the 5^{th} ratio of $\mathbf{\exp(x^5)}$? **Hint**: Use the above tables to spot a pattern.

III) What can you say about the n^{th} ratio of $\mathbf{\exp(x^n)}$?

We use the way of Archimedes from the previous labs and generalize to the following theorem.

Theorem

Let A = $\{a_1, a_2, a_3, \ldots , a_n\}$ be a sequence of equally spaced numbers.
Let D = $\{(a_1,b_1), (a_2,b_2),(a_3,b_3), \ldots , (a_n,b_n)\}$ be a set of points.

1) If the n^{th} difference sequence is equal to a constant then there exists a polynomial of degree *n*, $p(x) = c_n x^n + c_{n-1} x^{n-1} + \ldots + c_n$ such that

$$p(a_k) = b_k.$$

2) If the n^{th} ratio sequence is equal to a constant then there exists a polynomial of degree *n*, $p(x) = c_n x^n + c_{n-1} x^{n-1} + \ldots + c_n$ such that

$$\exp(p(a_k)) = b_k.$$

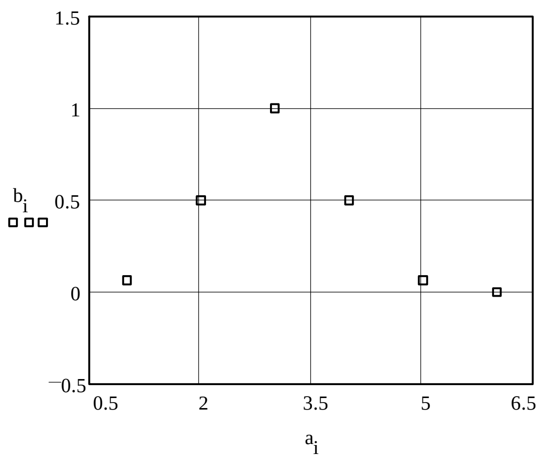
Gaussian Data Lab (Mathcad)

The data set (1, 0.0625), (2, 0,5), (3, 1.0), (4, 0.5), (5, 0.0625), (6, 0.00195) is received.

The student vectorizes it...

$$a := \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{bmatrix} \qquad b := \begin{bmatrix} 0.063 \\ 0.5 \\ 1 \\ 0.5 \\ 0.063 \\ 0.002 \end{bmatrix}$$

then plots it.



The student begins testing the data. From the graph it is suspected that the function that fits the data is a Gaussian. The student runs a series of ratios. The second ratio is a near constant. The third ratio is almost 1.
Conclusion: By a previous theorem, a Gaussian or bell curve, will fit the data closely.

$$c_j := \frac{b_j}{b_{j-1}} \qquad d_k := \frac{c_k}{c_{k-1}} \qquad e_m := \frac{d_m}{d_{m-1}}$$

b_i	c_j	d_k	e_m
0.063	7.937	0.252	0.992
0.5	2	0.25	1.008
1	0.5	0.252	1
0.5	0.126	0.252	
0.063	0.032		
0.002			

Gaussian Data Lab (continued)

Problem: Find the function $\mathbf{e^{P(x)}}$, $\mathbf{P(x) = ax^2 + bx + c}$, whose graph contains the points of the data set. Since there are only three unknowns, only three points are needed. The student uses (2 , 0.5) , (3 , 1.0) , (4 , 0.5).

Solution: A point **[A,B]** on the graph of $\mathbf{e^{P(x)}}$ satisfies $\mathbf{e^{P(A)} = B}$. Taking the natural log of both sides results in **P(A) = ln(B)**. Thus

$$a\cdot(2)^2 + b\cdot(2) + c = \ln(0.5)$$
$$a\cdot(3)^2 + b\cdot(3) + c = \ln(1.0)$$
$$a\cdot(4)^2 + b\cdot(4) + c = \ln(0.5)$$

This results in the following matrix equation.

$$\begin{pmatrix} 4 & 2 & 1 \\ 9 & 3 & 1 \\ 16 & 4 & 1 \end{pmatrix} \cdot \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} \ln(0.5) \\ \ln(1.0) \\ \ln(0.5) \end{pmatrix}$$

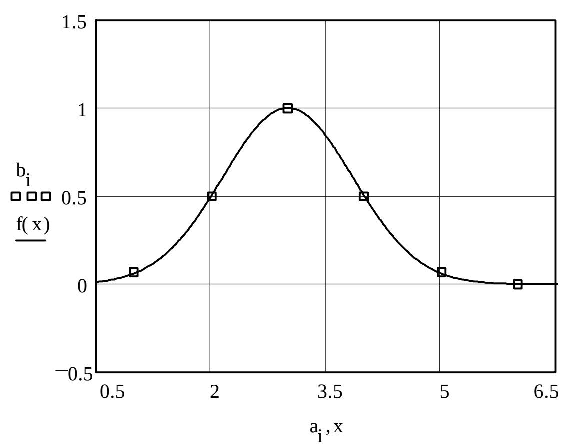
Solving the matrix equation yields

$$\begin{pmatrix} 4 & 2 & 1 \\ 9 & 3 & 1 \\ 16 & 4 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 0.5 & -1 & 0.5 \\ -3.5 & 6 & -2.5 \\ 6 & -8 & 3 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0.5 & -1 & 0.5 \\ -3.5 & 6 & -2.5 \\ 6 & -8 & 3 \end{pmatrix} \cdot \begin{pmatrix} \ln(0.5) \\ \ln(1.0) \\ \ln(0.5) \end{pmatrix} = \begin{pmatrix} -0.693 \\ 4.159 \\ -6.238 \end{pmatrix}$$

Thus $f(x) := e^{-0.693\cdot x^2 + 4.159\cdot x - 6.238}$

We note that if the original data set
(1, 0.0625), (2, 0,5), (3, 1.0), (4, 0.5), (5, 0.0625), (6, 0.00195) is indeed Gaussian, then the curve just found, by the theorm, must fit the entire set.

$$a = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{bmatrix} \qquad b = \begin{bmatrix} 0.063 \\ 0.5 \\ 1 \\ 0.5 \\ 0.063 \\ 0.002 \end{bmatrix}$$



Newton's Law of Cooling Lab (Mathcad)

An obvious lab involving the Law of Cooling is to have students collect the data themselves from something cooling to room temperature. This lab is a variant where the data has already been collected.

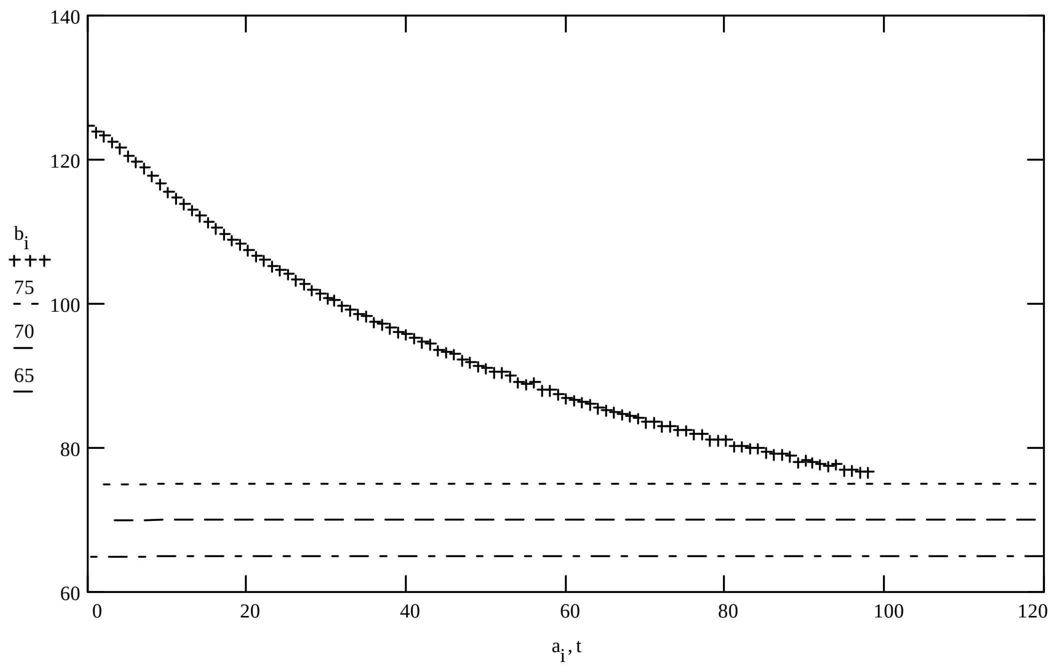
Mary Robinson, one of the co-investigators of this proposal, gave a workshop on a CBL or Calculator Based Laboratory, at a May 1996 conference at Ruidoso, New Mexico. Mary heated a quantity of water and let it cool. The CBL she used noted the temperature of the water every second for 99 seconds. The data Mary collected was read into the program Mathcad and is used below.

A) QUESTION: What was the temperature of the room where Mary conducted the experiment?

B) PROBLEM: Find a function $f(t) = A + B\,C^t$ that fits the data .

A) Solution: The question is asking for the asymptote. Essentially, the student must graph the data, experiment and guess. A bad guess will lead to a poor fit in the problem of part B. Here a guess of 75 seems too high, whereas a guess of 65 seems too low. 70 is better and, as will be seen, leads to a good fit.

The data and these guesses are plotted below.



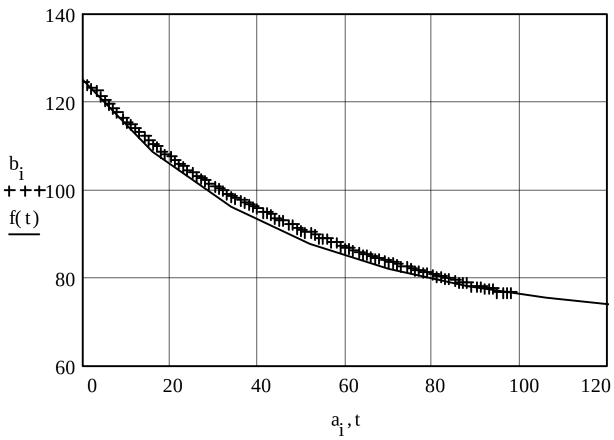
B) Solution: There are several ways to solve the problem. Two require knowing the asymptote. The first is algebraic. The first and last data points are, respectively (0, 124.63) and (98, 76.64). The asymptote is 70. The general solution is $y = A + B\,C^t$. Thus

$$76.64 = 70 + B\,C^{98}$$

$$124.63 = 70 + B\,C^{0}$$

Now both B and C can be computed exactly. B = 54.63, C = 0.9787247714

$$f(t) := 70 + 54.63 \cdot 0.9787247714^t$$

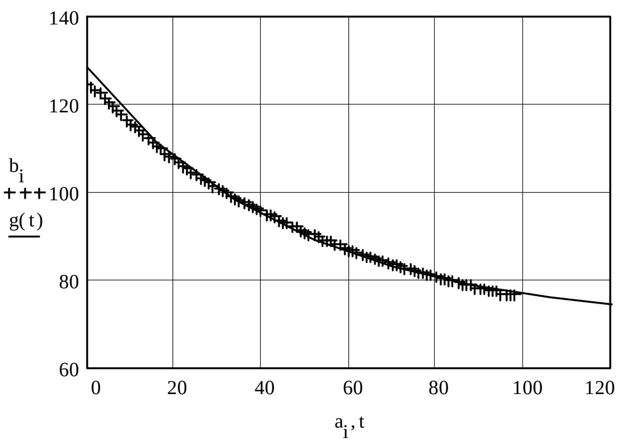


Squares Error

$$\sum_{i=0}^{98} \left(f(i) - b_i\right)^2 = 336.361$$

Another way is to move the y-data down 70, do an exponential regression to find B and C then add A = 70. The results are below. Predictably, the least squares error is lower.

$$g(t) := 70 + 58.085 \cdot 0.979^t$$



Least Squares Error

$$\sum_{i=0}^{98} \left(g(i) - b_i\right)^2 = 86.616$$