TEACHING STATISTICAL THINKING USING SPREADSHEETS

Alexander Vaninsky,
Hostos Community College, CUNY
500 Grand Concourse, Room B 409, Bronx, NY 10451, USA.
email: avaninsky@hostos.cuny.edu

Introduction. A famous French mathematician Augustin-Louis Cauchy proved in 1821 that the only continuous function that satisfies condition $f(x+y) = f(x) + f(y)$ is a direct variation function $f(x) = Cx$. This problem attracted attention of mathematical community for more than 150 years. At the Second International Congress in Paris in 1900, David Hilbert included generalization of the Cauchy functional equation in the list of principal problems that would lead to the furthering of mathematics in 20^{th} century. Among mathematicians involved in the research of the Cauchy functional equation are Banach, Darboux, John von Neumann, Sierpinski and many others. The last result known to the author is Shapiro (1973).

In this paper we focus on two more functional equations that, among several others, follow from the original one. The equations are as follows. First: if for any positive values of x and y condition $f(xy) = f(x)\ f(y)$ is held, then $f(x)=x^C$. Second: if for any values of x and y we have $f(x+y) = f(x)\ f(y)$, then $f(x)=C^x$. Our interest for these equations stems from the fact that they represent essentially different rates of expansion of an observed phenomenon: linear, polynomial, or exponential, respectively. Roughly speaking, each equation corresponds to three qualitatively different scenarios of development of a situation: slow, moderate, or avalanche-like. Importance of timely determination of the rate of expansion is self-evident in view of examples of unknown heavy diseases, global warming, poverty expansion, etc. The earlier the rate of growth is recognized, the more efficient may be the reaction of a society.

In this paper below we present statistical procedure based on the Cauchy functional equation and its consequences that allows testing series of observations for the rate of growth. We demonstrate how the procedure can be implemented using spreadsheet statistical functions. Finally, we apply suggested approach to testing the Malthusian theory regarding exponential growth of the world population, and discuss obtained results. The main goal of this paper is to provide a teaching tool for extracurricular activity that demonstrates how abstract mathematics paves the way for development of a statistical procedure leading to practically important conclusions.

Statistical background. To explain our approach, start with an example. Suppose that observations are made in days 2, 3, 5, and 6 and obtained data are as shown in table 1. What might be the estimation of the rate of growth? Calculating the values of $f(x) + f(y) = 7 + 10 = 17$ and $f(x) \cdot f(y) = 7 \cdot 10 = 70$, we can easily see that the only correspondence that is likely held is $f(x \cdot y) = 71 \approx f(x) \cdot f(y) = 70$. Based on this observation, we conclude that the rate of growth is probably polynomial.

Table 1. Example of statistical observations

Day (*x or y*)	$x = 2$	$y = 3$	$x+y = 5$	$x \cdot y = 6$
# Obs $f(x)$, $f(y)$	$f(x) = 7$	$f(y) = 10$	$f(x+y) = 48$	$f(x \cdot y) = 71$

In general case, situation is not so simple. We need to compare many observations and choose the most probable relationship. In this case we suggest applying statistical technique known as paired *t*-test. To apply this test we first pair observations as appropriate, and then apply *t*-test to the series of differences calculated by pairs. Finally, we test a hypothesis that average value of the series of the differences equals to 0. If the hypothesis is rejected, then the pairs do not satisfy the suggested criterion. In case that the hypothesis cannot be rejected at a given confidence level, we state that the rate of growth is likely found based on our general understanding of the problem.

Thus, the first problem that we face is forming convenient pairs of observations suitable for comparison of observations corresponding to x, y, $x+y$, and $x \cdot y$. This may be done in a way shown in table 2. The table is for a case of ten observations available, but it be continued in a similar was.

Table 2. Forming pairs of observations.

Ordinal number of observation		Observations combined for paired *t*-test	
x	y	$x+y$	$x \cdot y$
2	3	2+3 = 5	2*3=6
2	4	2+4=6	2*4=8
2	5	2+5 =7	2*5= 10

Recall that paired *t*-test statistic is

$$t = \frac{\bar{d}}{\left(\frac{s_d}{\sqrt{n}}\right)}, \qquad s_d = \sqrt{\frac{\sum_{i=1}^{n} d_i^2 - n\bar{d}}{n-1}},$$

where $\bar{d}$ is an average value of the difference calculated by pairs, and n is a number of observations. Statistic t has t-distribution with $(n\text{-}1)$ degrees of freedom and standard deviation s_d.

It may be noted that suggested approach is preferable over regression approach for samples of small or medium sizes. It analyzes internal dynamics of data change, while exponential regression might fail because for any finite interval $e^x \approx 1 + x + x^2/2! + \ldots + x^n/n!$ for some n, that is indistinguishable from a polynomial.

One more notice should be made. To allow for a constant term in the raw data, we assign the first observation ordinal number 0 and use observations corresponding to x=0 or x=1 for normalization. Thus, if $f(x) = K + Cx$ rather than just $f(x) = Cx$, as required, the transformation $f(x) - f(0) = (Cx + K) - K = Cx$ leads to the required analytical form of the function. Similarly, $f(x) = Kx^C$ becomes $f(x) = x^C$ after division by $f(1) = K$, and $f(x) = Ke^{Cx}$ becomes $f(x) = e^{Cx}$ after division of all observations by $K = f(0)$. Note that spreadsheets provide tools for performing such transformations easily.

The final step is calculating the paired *t*-test statistic. Spreadsheets have a required statistical function *TTEST( )*. It should be applied with parameters 2, and 1, that mean *two-tailed distribution* and *paired t-test*, respectively. An example of Excel function for data located in ranges *P10:P60* and *Q10:Q60* is this "*=TTEST(P10:P60,Q10:Q60,2,1)*".

Example: testing the Malthusian theory of population growth. Thomas Robert Malthus, http://en.wikipedia.org/wiki/Thomas_Malthus, was born in 1766 in a prosperous family. His father was a friend of the philosophers Hume and Rousseau. Malthus started his education at home, then continued at Jesus College, Cambrige, UK. His principal subject was mathematics, in which he earned a master's degree. In 1805 he became Britain's first professor in political economy. His students affectionately referred to him as "Pop", or "Population" Malthus. One of his students wrote a responsive essay concerning population growth and criticizing many of his ideas.

Malthus's views were largely developed in reaction to the optimistic views of his father and his associates, notably Rousseau. He suggested the Principle of Population that was based on the idea that population if unchecked increases at a geometric rate (i.e. exponentially, as 2, 4, 8, 16, etc.) whereas the food supply grows at an arithmetic rate (i.e. linearly, as 1, 2, 3, 4, etc.). He wrote: "The power of population is so superior to the power of the earth to produce subsistence for man, that premature death must in some shape or other visit the human race." Malthus made a prediction that population would outrun food supply, leading to a decrease in food per person. He even predicted that this must occur by the middle of the 19th century. Fortunately, this prediction failed, in particular, due to his incorrect use of statistical analysis and ignoring development of industrial chemistry.

It is interesting to test Malthus's hypotheses using contemporary data. In this paper we focus on one of them, the hypothesis of exponential growth of population. We used data of the US Census Bureau for 1950 -2050 (actual data until 2000, interim projections for 2001-2002, and predictions starting with 2003), available on website http://www.census.gov/ipc/www/idb/worldpop.html. In calculations, data were separated into groups as follows: 1950 - 1975, 1975 - 2000, 2000 - 2025, 2025 - 2050, 1950 - 2000, 2000 - 2050, and 1950 - 2050. The first observation in each group was assigned an ordinal number 0 and used for normalization.

Obtained results are as follows. No one of the hypotheses turned out to be applicable for the long periods like 1950 - 2050, 1950 -2000, or 2000 - 2050. Results obtained for four shorter periods of 1950 -1975, 1975 - 2000, 2000 - 2025, and 2025 - 2050 lead to different conclusions as shown in table 3. Using data of the table and assuming 5%

Table 3. Probabilities of linear, polynomial, or exponential growth hypotheses, as measured by paired *t*-test, %

Period	Hypothesis of world population growth		
	Linear	Polynomial	Exponential
1950 - 1975	0.000	8.190	0.007
1975 - 2000	0.000	4.812	28.164
2000 - 2025	0.000	2.034	0.000
2025 - 2050	0.000	0.918	0.000

significance level, the following conclusions may be made. Polynomial rate of growth cannot be rejected for the period of 1950 -1975, and exponential rate, for 1975 - 2000. For periods of 2000 - 2025 and 2025 - 2050 results are inconclusive, though polynomial hypothesis seems more likely than the exponential one. It should be noted that data for 2002 and later years are just forecasts, and cannot be taken for granted. Thus, our estimations obtained for the periods of 2000 - 2025 and 2025 - 2050 are just expectations as well.

Summarizing, we can state that expected situation is not so dangerous as Malthus predicted. Periods of exponential growth of population alternate with the periods of polynomial growth. This alternation together with advances of industrial chemistry allows for the hope that there will be enough food in the world for everybody.

References.

Weiers, R. (2002). *Introduction to Business Statistics*. 4th Ed. CA: Wadsworth Group. Duxbury/Thompson Learning.

Cauchy, A. L. (1821). Cours d'Analyse de l'Ecole Royale Polytechnique. Chez Debure frères. Reproduced by Editrice Clueb Bologna, 1992. (In French.)

Santos, J. and E. Weisstein. (2005). Cauchy Functional Equation. *http://mathworld.wolfram.com/CauchyFunctionalEquation.html*.

Shapiro, H. (1973). A Micronote on a Functional Equation. *The American Mathematical Monthly*, 80(9), 1041.

Websites

http://en.wikipedia.org/wiki/Thomas_Malthus

http://www.census.gov/ipc/www/idb/worldpop.html