Teaching methodology for smart students with less-developed abstract thinking skills

Vladimir Shelomovskii

vvsss@rambler.ru

Department of Math and Computer Science

University of Technology

Papua New Guinea, Lae

Abstract: *The author has been teaching statistics to students in Papua New Guinea for three semesters. These students are smart but their mathematical knowledge and abstract thinking skills require considerable development. The author is forced to build teaching on the basis of binding mathematical concepts with objects of the real world. The subject of the paper are examples of realizing this idea.*

1. Introduction

The author has taught mathematical disciplines in Russia for more than 25 years. The feature of training of most Russian higher school students is that they have a solid base of mathematical knowledge and well-developed abstract thinking. Usually the creativity of a lecturer lies in explaining links between abstract concepts. The author has used visualization and visual models in teaching as a subsidiary element of the learning process. In Papua New Guinea he has got a new situation. Students in Papua New Guinea University of Technology are still smart, but their mathematical knowledge and abstract thinking skills require considerable development. It might seem that the process of teaching here is easier than teaching in Russia, but it takes a little time to make sure that it employs the full knowledge, skills and experience of the lecturer.

At the very beginning of teaching, for example, in studying probabilities of events an attempt to present the multiplication law gives the result that surprises the author. We consider the following problem. We toss a coin; it can fall in two ways: heads or tails. In how many ways three coins may fall as a result of the toss? It seems evident for us that events (the coin dropped down by head or by tail) are independent, and then the number of events is $2^3 = 8$. However, Papua students claim in unison, the number of events is 6! The author sees that mathematical rule as a formula in a textbook or just uttered by a teacher is often an incomprehensible information for students, a set of numbers, letters, and mathematical symbols that require memorization. The teaching material which relies on abstract thinking, required in mathematics, is difficult to understanding for most of students. So the author is forced to construct visual models and to build teaching on the basis of binding abstract mathematical concepts with objects and phenomena of the real world. For example, in computer programming teaching, in particular the course of theory of algorithms while explaining concepts of binary trees, linked lists, recursion methods and sorting, the author widely uses game models with students taking part in it, computer models as well as models made from household goods, colour schemes and corresponding colour recording of computer running on the text of a program with a lot of recursion. Successful fragments of teaching, accumulated as a result of three semesters of work, are presented in the paper.

2. Statistics models

Model of data grouping. Mathematical statistics processes large data arrays, which are usually grouped according to a certain principle. A successful model of such grouping the author finds watching for ticket sales in PNG buses with unnumbered seats. A conductor collects money for the fare from each passenger and does not miss anyone.

Let's assign the numbers to passengers according to their place in the bus, starting from the row in the fore-part of the bus. The passenger number 1 sits at left in the first row in the fore-part of the bus. Further numbers are assigned to passengers to the right in the same row, and then in the same manner in the next row. Let's consider the variant when each passenger gives money to the conductor when he wants. Let's arrange the numbers of passengers according to the order in which they are delivering money to the conductor. In this case we get a disordered (random) sequence consisting of passenger numbers. If the conductor uses such method, it is easy to miss someone and it will be difficult to find it if the passenger does not want to pay.

Let's consider another variant. The conductor collects money according to rows, each containing 1-4 seats. At the first step, the conductor takes money from the passengers sitting at the first row, then he moves to the next group (second row), and so on. The psychological feature of people's thinking is used: we may easily remember the objects if their amount is no more than 7. If the number of objects is 8 or more, we usually lose something. It is clever solution of the problem, at the moment of payment service the conductor is controlling only a group of 4-8 people (1-2 rows). The mathematical aspect of this method is that the conductor sorts the passengers and makes data grouping using the rows. This illustrative example is used before the explanation of the model of data grouping. It helps students to carry out the grouping entirely consciously.

Model of mathematical point. Statistics deals with objects which have a lot of properties that can be measured quantitatively. Each measurement gives a certain number. Statistics typically combines a pair of numbers that characterize the object, in a form of a pair of coordinates of a point on the coordinate plane, and then considers the totality of these points, evaluate a mean, a standard deviation, quartiles and so on. The author is looking for visual models everywhere, in every moment of everyday life. He drinks a cup of tea with honey. Immersed in thought, he forgets to wash the cup. A small amount of honey remaining on the bottom of the cup attracts many ants. The author photographs them, uses the appropriated computer programs to process the picture and integrate it with the coordinate system. The computer illustration is used in the assignment for visual evaluation of the statistical distribution parameters.

The problem is formulated as following. Let each ant corresponds to one observation, the point with coordinates (x,y). We have the set of pairs (x_i, y_i), $i = 1..n$. It represents the data in the set of observations. The axes of the coordinates are shown. Estimate: (a) The number $\boldsymbol{n}$ of the ants. (b) The mean $\boldsymbol{E(Y)}$. Show the line $\boldsymbol{y = E(Y)}$, (c) **2**-percentile for $\boldsymbol{x}$. Show the line $\boldsymbol{x}$ = 2-percentile. Etc.

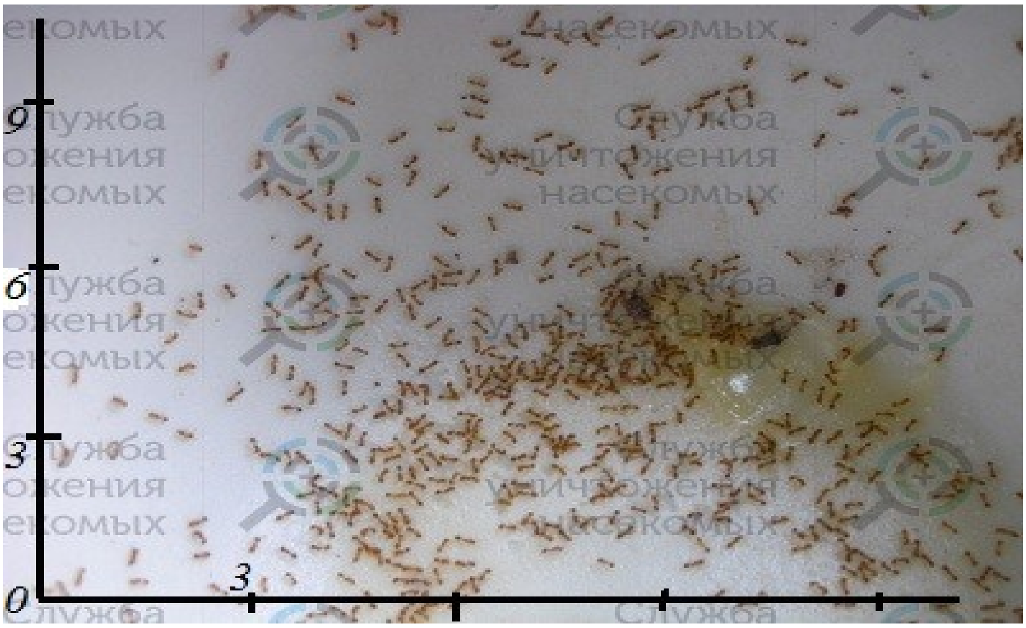


Figure 1 Ants used as mathematical points

Model of a random variable and a standard deviation. Internet resources devoted to the history and geography of the country give ideas to create exciting, emotive tasks of the mathematical course. The author tells the story about the geographical journey of the ship named Beagle with captain Robert FitzRoy, later the first governor of New Zealand. The model of a geographic point is used which we first need to draw onto the map. The point is drawn with a certain error, that is, in fact we get a region containing the valid point. Then following seafarers have to find that point in the ocean. A colorful description of the ship with a sailor in the cask at the top of the mast, its observation zone and the final cry "Land!" allows us to emotionally characterize the standard deviation, and its meaning. Accordingly, the concept of the necessary accuracy of a parameter estimation arises.

Model of the statistical distribution. The author takes advantage of the fact that large wall clocks of the same type are available in most university rooms: secretariats, library and many classrooms. These clocks allow to explain many aspects associated with a random variable. Accurate clocks must show the same actual time. However, the real clocks show what they show. The accuracy of the testimony of these clocks is significantly different. Typically, there is a large group of clocks, showing the time which differs from the actual time on 2-3 minutes. There are several precise clocks (within minutes, usually in the rooms of very neat secretaries) and a certain amount of clocks, which readings are different from the actual time on 10-15 minutes.

The data of the time shown by clocks of the same type in university rooms at different moments and the time shown by the checking watch at the same moments is transformed to the data set t_i measured by minutes (as the difference of the raw data). This data group forms bell-shaped distribution with a very long tail, which is typical for the actual data, but is absent in normal distributions, which are typically used in teaching.

The preparation requires an hour before the lecture: the author visits 10 easily accessible places (secretariats, rooms for academic conferences, the library, free accessible administrative offices). He records the time shown by the wall clocks and the time on the author's phone. Two rows of numbers are written on the whiteboard and the author begins data handling. The lecturer makes calculations using Scientific calculator in Scientific Mode. Students use the same calculators, each being able to make own calculations and check the results.

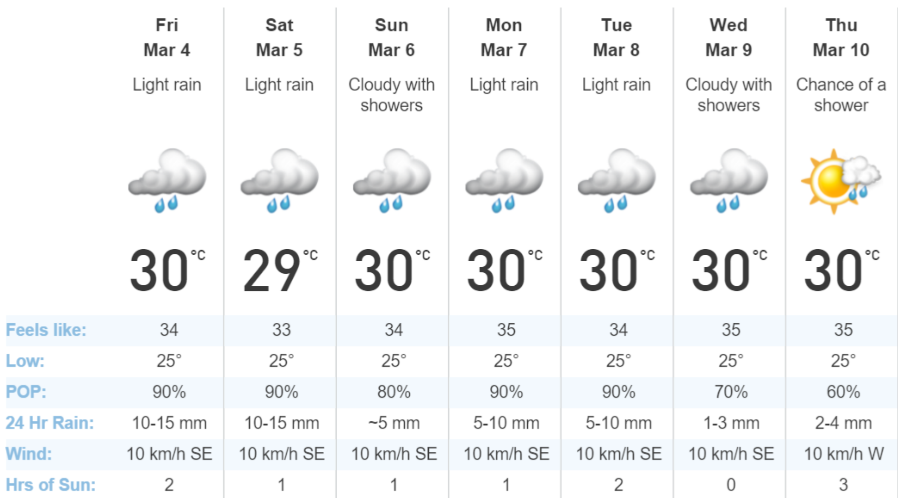
Location	MCS	ME	CE	MN	Rosa Kekedo	AS	Europa	Library	Sandover
Clock	10.07,5	10.18	10.09,5	10.03	10.14	10.15.5	10.17	10.20.5	10.25
Phone	10.06	10.07	10.09	10.12	10.14	10.16	10.19	10.20	10.23
t_i	1,5	11	0.5	-9	0	0	-2	0.5	2

In mathematical statistics it is important to understand the chain of inter-related activities, including data collection, selection of suitable data (Winsorisation or trimming the data set), data processing, evaluation of accuracy of the result, transition from a data set to a random variable. Students are learning to find the confidence interval for the mean value of the variable, to formulate conclusions, to use findings in practice. In the example with clocks the time difference is found as the deviation from the baseline time shown by the phone. We are sorting the data set. It turns out that clocks are working with very substantial errors in two secretariats, more than 10 minutes. That is outliers are found. Winsorisation is performed. As a result, the dispersion found for the original and the dispersion found for Winsorized series differ about five times. The example helps to explain the essence of the random variable, as the realization of the true time in the specific clock. A confidence interval is found for the true time specified with the clock. The author together with students form conclusions about how to evaluate the true time at the selected moment: by the use of one most accurate clock, or a group of several clocks. The recommendations on the level of trust to different clocks are made. Students may check the results on their watches. The explanation of the teaching material using this example is given on the general lecture and refined on tutorials in each group, that is, the total time is approximately two pairs (three hours) for each group.

Model of outliers. In mathematical statistics methods of data processing are significantly different in cases without outliers and with outliers presence. In the problem with clocks outliers are determined easily. The deviation from the true time is not more than 5 minutes for well tended clocks and more than 10 minutes for "orphan" clocks. The search of the average and the standard deviation for all indications and for data purified from outliers effectively and clearly shows the difference between robust methods, standard methods for purified data and standard methods, which can be formally applied for data with outliers.

Model of sources of information. The problem of information collecting also concerns the selection of data from the array. We must be able to discard redundant data and select those indications that meet certain conditions. One of the tasks in the statistics course assignment is associated with the selection of desired information from multiple datasets. We use publicly available Internet resources to obtain datasets. For example, sets of changing parameters that characterize weather in the area, in our case it is Lae city, PNG.

The problem is formulated as following. Use Lae Short Term Forecast (Figure 2). Let the start of observations be at 2 March and the finish at 9 March. 4 March is considered as ”Today“. Make the bar chart for sun days, days with rain and showers days. Explain, how do you get quantities of the days.



	Fri Mar 4	Sat Mar 5	Sun Mar 6	Mon Mar 7	Tue Mar 8	Wed Mar 9	Thu Mar 10
	Light rain	Light rain	Cloudy with showers	Light rain	Light rain	Cloudy with showers	Chance of a shower
	30°C	29°C	30°C	30°C	30°C	30°C	30°C
Feels like:	34	33	34	35	34	35	35
Low:	25°	25°	25°	25°	25°	25°	25°
POP:	90%	90%	80%	90%	90%	70%	60%
24 Hr Rain:	10-15 mm	10-15 mm	~5 mm	5-10 mm	5-10 mm	1-3 mm	2-4 mm
Wind:	10 km/h SE	10 km/h SE	10 km/h SE	10 km/h SE	10 km/h SE	10 km/h SE	10 km/h W
Hrs of Sun:	2	1	1	1	2	0	3

Figure 2 Lae Short Term Forecast

The solution may be formulated as following. I remember, that sunny days were at 2nd March, 3th March, and today (4th March). I see in the table, that 5 – 9 March will be rainy. 10 March is out of the time of the problem.

3. Conscious data research

Evaluations for unobserved values. Mathematical statistics is widely used for the evaluations for unobserved variables. Models of such evaluations are very useful. At the Papua University fluorescent lamps are used in the classrooms. It is known that the typical service life of fluorescent lamps reaches 2000 - 3000 hours in good supply voltage. In Papua it is typically that the voltage is quite unstable, voltage spikes often occur at the emergency shut down of buildings. Thus, the typical lifetime of fluorescent lamps is close to a half of year. A typical situation is when about a half of lamps work in the classrooms. One of the tasks in the assignment was associated with the evaluation of the following event "A half of lamps work in the room today and passport lifetime of a fluorescent lamp is 2000 hours. Evaluate when maintenance services was carried out in the room."

The author expected difficulties with the transformation of hours in days and months, but suddenly faced with the students answers "18 hours" (ie, the time after the lectures) or "last year, because this year the lamps have been not changed". Throughout the course, students learn to use statistics in order to find answers to questions for which the ordinary worldly knowledge is no longer enough. But perhaps it is even more difficult to understand that the serrated mathematical rule, a computer and the Scientific calculator with the corresponding functions do not magically give the only possible answer to any question.

Estimation of lie by using correlation In magazines and newspapers we frequently come across data representations in a variety of forms. Data representations need to be looked at critically. Reading and interpreting graphical representations of data is not a trivial task. Many diagrams, charts or graphs in newspapers and magazines have been designed to magnify differences or to emphasize minor points. We should ask ourself questions such as: How is the data collected? Does the representation give a fair picture of the data? Is the data reliable? What purpose do the presenters of the data have? The author creates tasks, encouraging students to ask similar questions. Internet offers us an infinite number of resources on various topics, which would seem to have nothing to do with mathematics. But statistics is ubiquitous, it can deal even with poetry. The estimation for correlation can be applied to various objects and phenomena, and even to a poem.
There are two first verses of the poem of William Blake.

Tyger! Tyger! burning bright
In the forests of the night,
What immortal hand or eye
Could frame thy fearful symmetry?

In what distant deeps or skies
Burnt the fire of thine eyes?
On what wings dare he aspire?
What the hand dare sieze the fire?

The problem is formulated as following. In the verses of the poem the total number of words in each verse, ***y***, and the number of words with three of fewer letters, ***x***, are given in the table below. Let ***m*** be the number formed by two latest digits of your ID (if ID = 12345678, then ***m*** = 78).

Verse	***1***	***2***	***3***	***4***	***5***	***6***	***7***
x	7	7	9	7	11	7	10
y	20	25	27	23	28	20	***m***

1. Calculate the value of the product moment correlation coefficients in two cases: for first 6 verses, and for 7 verses.

2. How many verses are there in the poem? Explain how do you make the estimation.

In solving the problem students notice that statistics for first 6 verses and for 7 verses are very different. This leads to wonder, why? And how can the number of words in the old poem depend on ID of a given student?

Lies, big lies, statistics and regressions. American writer Mark Twain said: there are lies, big lies and statistics. Statistics is included in this list, as people often do not understand the basics of statistics and use its methods formally in situations where it is not permissible. To demonstrate this phenomenon the author examines the same data group (a total of five pairs of numbers) which is processed by a dozen different ways, getting very different results. The lecturer makes calculations using the Scientific calculator in Scientific Mode. Students use calculators to check the results.

Linear Regressions

Let the pairs (x_i, y_i), $i = 1 .. n$ be given. The pair (x_i, y_i) represents information about the object *i*. We will use two examples: ***Case a***: (0,2),(1,1),(2,2),(3,4),(4,13), $n = 5$. ***Case b***: We suppose that the point (4,13) is an outlier and use only the points (0,2),(1,1),(2,2),(3,4), $n = 4$.

We want to find the best fit straight line (regression line) $y = y_0 + b\,(x - x_0)$ such that the sum of the squared distances d_i^2 from points to the line be minimum. The point (x_0, y_0) is the midpoint of the set. We use the points M and N as follows: N = (med x_i, med y_i), $M = (\bar{x}, \bar{y})$.

Case a:* *N = (2, 2), ***M*** = (2, 4.4). ***Case b:* *M*** = (1.5, 2.25).

The author explains the Ordinary Least Square Regression (OLS) estimate of Y on X and X on Y. We suppose that coordinates $\boldsymbol{x}$ and $\boldsymbol{y}$ have the same errors and find the Major Axis Regression (MAR). We suppose that the errors of the coordinates x (y) are proportional to σ_x (σ_y) and find the Geometric Mean Regression (*GMR*). In Figures 3 and 4.a we see two lines. The first (blue) straight line has the bigger slope and is connected with all points. The second (red) line is the best fit for 4 points.

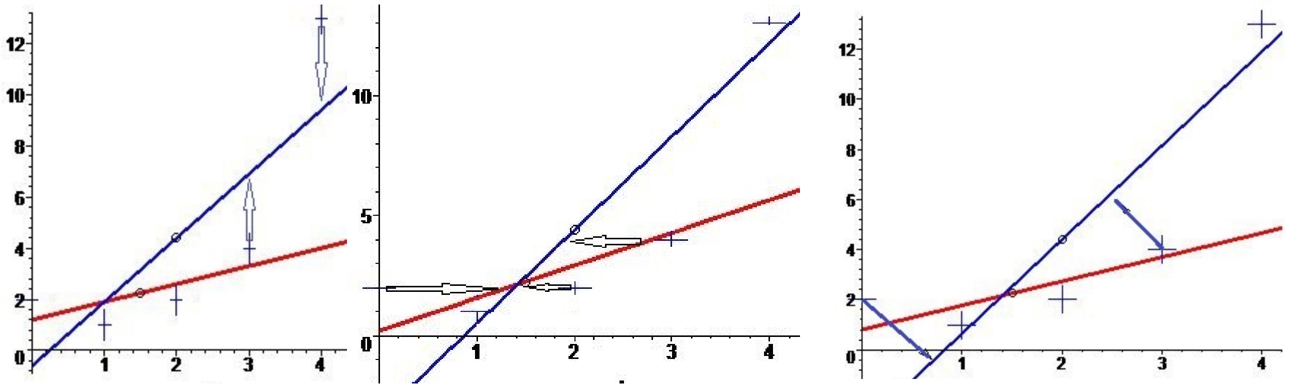


Figure 3 a) OSL estimation $y(x)$, b) OSL estimation $x(y)$, c) Major Axis Regression $y(x)$

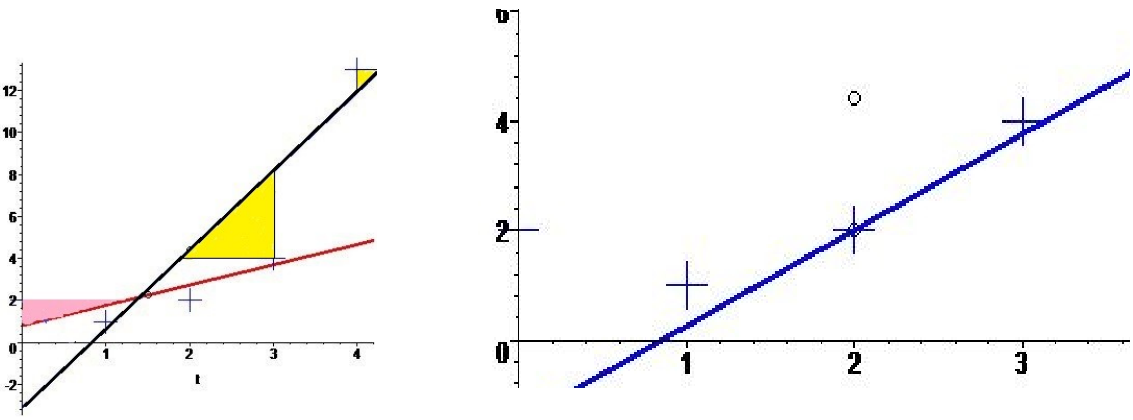


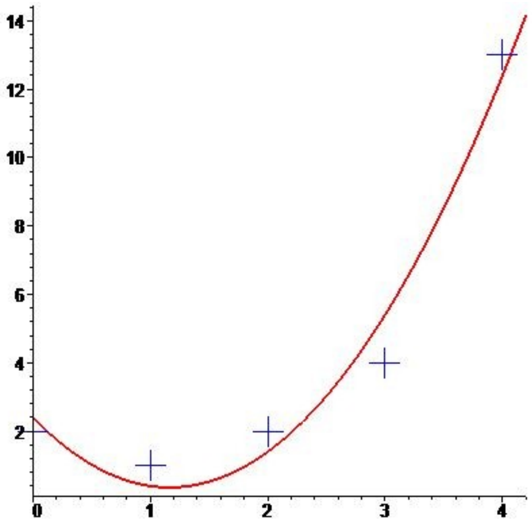
Figure 4 a) Geometric Mean Regression, b) Linear nonparametric regression

We suppose that there is outlier and find the nonparametric estimator of the slope of the regression line (Figure 4.b) as a median of coefficients of slopes of the lines passing through the pairs of points (x_i, y_i) and (x_j, y_j), $1 \le i < j \le n,\ \ x_i < x_j$. $b_{Sp} = med\,\dfrac{y_j - y_i}{x_j - x_i}$.

We suppose that the fitted curve has the second order, we have the equation $y = b_0 + b_1x + b_2x^2$. We use a Scientific calculator and get $y = 2.4 - 3.5\,x + 1.5\,x^2$. Students understand the crib:

(SC: ***Mode Mode*** (SD REG BASE) 2 ►(Lin Log Exp) ►(Pwr Inv Quad) **3** A B C)

We suppose that the fitted curve may be $y = a \times b^x$. We take natural logarithms on both sides, get a linear model: $\ln y = \ln a + x \ln b$, solve the linear model equation and get: $y = 0.23 \times 2.73^x$.



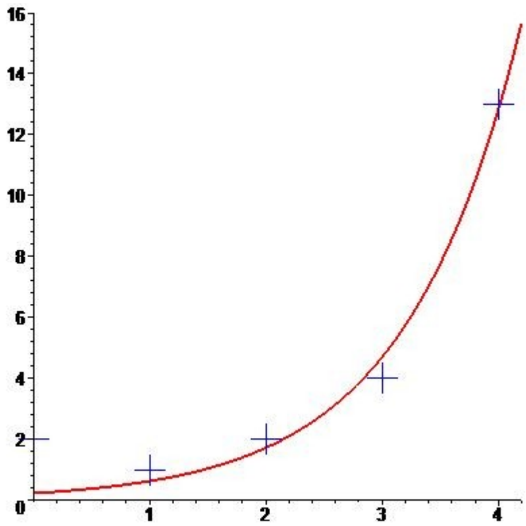


Figure 5 a) Quadratic Model, b) Exponential model

What is the true model? Students study these examples and understand that the same data can produce different results.

4. Teaching and Practice. Mining industry is developed in Papua New Guinea, so the problem of land rehabilitation is acute. For this aim acacia trees planting is applied. Acacia trees are planted in neat rows with a fixed distance between plants in the row and the same distance between rows. If the distance is too small, plants suppress each other. If the distance is large, the soil is not restored effectively; plants do not form a single whole, and damage to plants frequently occurs. There is a question about the optimal distance between plants. Students of the Agriculture Faculty are looking for the solution to this problem.

The author has created the three-dimensional model using visualizing program GinMA [1]. The land rehabilitation work is carried out widely enough, so it was not difficult to get the real data for the model. The computer model shows the real separate section of the landings, a field with a group of plants, shown in Figures 6,7 below. The model is interactive in the sense that different steps of the model show the state of plants at different moments of time. Plants are modelled in the form of a cone with the height and the diameter which correspond to the measured in field conditions parameters. The green color is used for the alive plant, yellow for the morbid plant, purple for the dead plant. The pair of control points ***A*** and ***B*** allows us to change proportionally the distance between plants and their sizes.

Students enjoy entering data and watching an interactive diagram of a conditional image of the garden, which they observe in carrying out practical work on the ground. This model establishes close visual link between real living plants and their computer models, helping to see that mathematical laws which govern the growth of model landings act not only in the computer but operate everywhere in the real world.

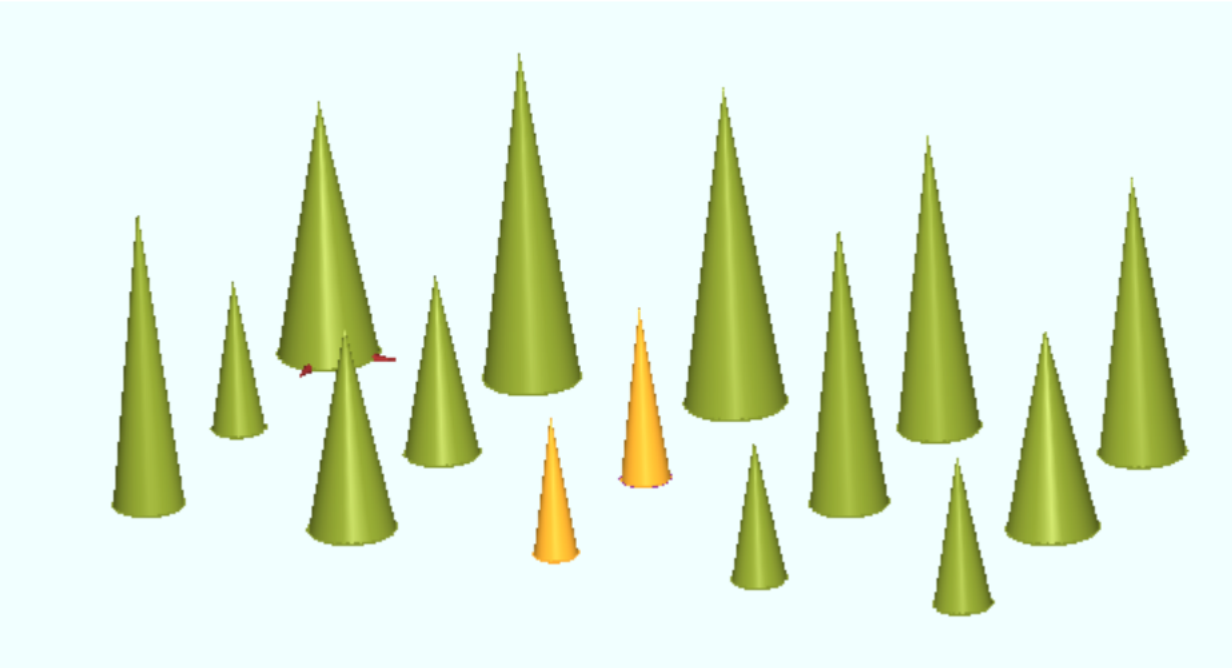


Figure 6 a) Plants at the age of 2 month

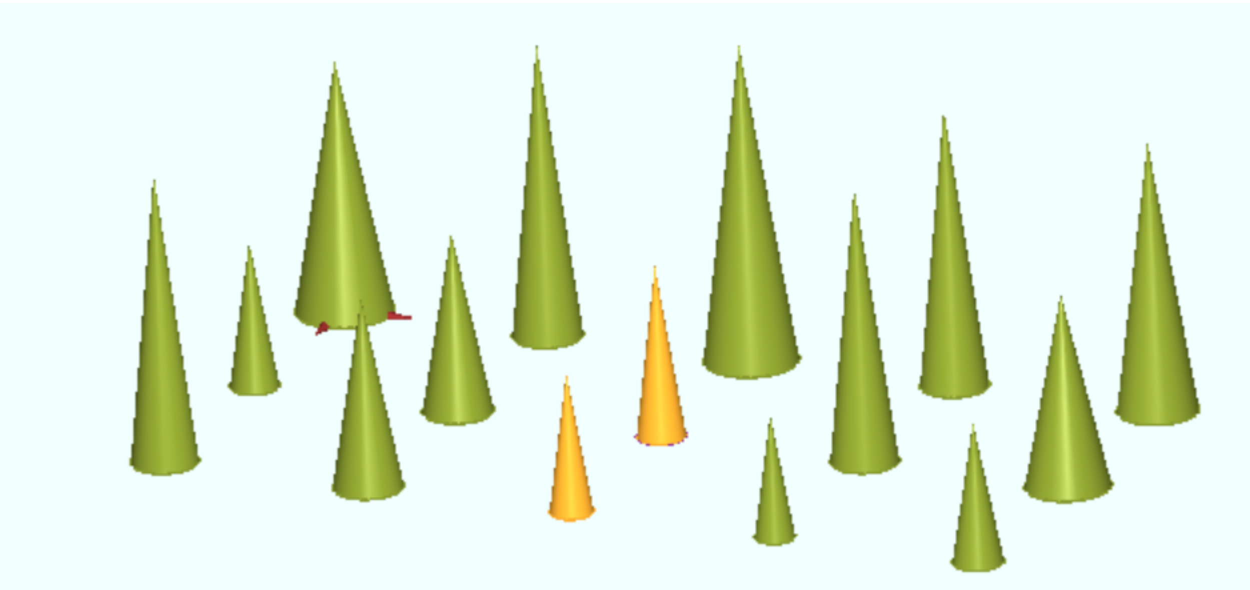


Figure 6 b) Plants at the age of 4 month

5. Conclusions

The result of such work is usually given in the form of certain numerical values that characterize the level of knowledge. Control measurements were carried in the form of two assignments. The model of the wind which is dangerous for bridges was used, the photographs of the wind destruction were included in the text of the assignment. The data individual for each student was used, each data set depending the student's ID number (to prevent mass cribbing). The reference group comprised 140 students. There were over 100 students on the lecture. Approximately 75% of students correctly fulfilled transformations of the data set and found the confidence interval on the proposed data.

However, a complete study of the results of the method applying could not be carried for the following reason. Students became involved in the political process, they decided to replace Prime Minister. Students boycott began, which resulted in riots, buildings arsons and the murder of one of students. As a result, the academic year was interrupted for 4 months with the subsequent loss of purity of the experiment.

Nevertheless, the results of applying the described methods have already been observed in the classes. Students mutely looking at the teacher, gradually turn into students, occasionally laughing, coming after the lecture with questions that show that students understand the essence of the material and want to delve into the material further. These are expected positive changes, which could be explained by the fact that young people adapt to the lecturer, begin to better understand the language. But several cases are unexpected for the author and predispose to explain the reasons for positive changes in students behavior by his teaching methods. Students ask questions on topics not included in the courses taught by the author. The author notes, that these issues are outside the course, and it may be better to ask other teachers, more experienced in teaching in this country, better speaking language, but students answer: We have books and notes with all the formulas. Please, explain it to us in such a manner as you explain the material on your lectures, to become clear.

We may also give an example of another kind. The student brings the assignment on the statistics course. The author asks the question of how he calculates the correlation coefficient. The student takes scientific calculator and performs calculations quickly and beautifully, in scientific mode. Before studying this course, students in Papua used these calculators to perform only the four basic arithmetic operations. This course has caused a clear interest and many smart students easily mastered previously unknown to them methods of the calculator use in scientific mode. The technology skill appeared as the result of students' interest to learning.

Figure 7 Sketch of the experimental field

Acknowledgements The author would like to thank his wife Svetlana Nosulya for her help in selection of examples and translation to English.

References

[1] D. Shelomovskii. GInMA, http://deoma-cmd.ru/en/Products/Geometry/

[2] https://www.youtube.com/watch?v=ylh1d0o41Mk