"USING PC AND TI-92 IN TEACHING STATISTICS IN AUSTRIAN SECONDARY SCHOOLS"

Otto WURNIG
Institute of Mathematics, University of Graz
A-8010 Graz, Heinrichstraße 36, Austria, Europe
otto.wurnig@uni-graz.at

ABSTRACT

After the recent reforms of the Austrian curricula in mathematics, **statistics** and **the use of the computer** were fixed in mathematical instruction **for ten to fourteen year-old students** (1993, 2000).

In grades 5 and 6 the concepts of absolute and relative frequency, mode, arithmetic mean, median and different possibilities to plot graphs (pictogram, pie graph, bar graph, line graph, polygon) were integrated. When they work on the computer the students are allowed to use hand calculators and they can use spreadsheets. As spreadsheet the teachers generally use **EXCEL** if computer science is a new subject in grade 5 or 6.

Since the Austrian CAS II project in 1997/98, the use of the **TI-92** has been tested in many classes. With the TI-92 it is possible to get a boxplot with the different quartiles of a set of data very quickly. But it also offers the teacher a good chance to acquaint the students of **grade 8** with such difficult concepts as **linear and geometrical regression** and **correlation**.

In my lecture I will show the way I have worked with students of grade 8 and with teacher students at university. It is very important not to take sets of data out of the school books only. I allow the students of grade 8 to work with their own data (length and mass) or I let them find real data with the help of **CBL (calculator-based laboratory)** in an experimental way. Thus they get a better understanding of the concepts of regression and correlation.

1. Introduction

The use of computers in mathematical education in schools depends on some very important conditions. The use of the computer has to be required in the **curriculum**, sufficient **hardware** and good **software** has to be bought and the computer is to be admitted in **oral and written exams**, including the final examinations. All these conditions have been fulfilled to a great extent for grammar schools, business academies, secondary technical and trade schools in Austria in the last ten years. Therefore it depends primarily on the mathematics teacher, in which way and how intensely the computer is used in mathematical education.

Based on the recent reform of the Austrian curricula in mathematics in 2000, **statistics** and **the use of the computer** have to be combined in the schoolbooks. This is only possible by putting parts of the book on the internet. In the schoolbook the students find a www-address and with it they can call for the additional parts called "schoolbook plus". If the twelve-year-old pupils want to call such parts of the widely used book by Reichel-Litschauer-Groß, they have to enter www.e-Lisa.at and can choose the hyperlink to the EXCEL-programs.

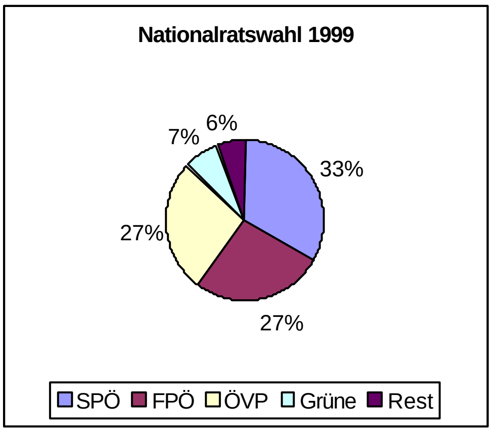
2. First use of a spreadsheet

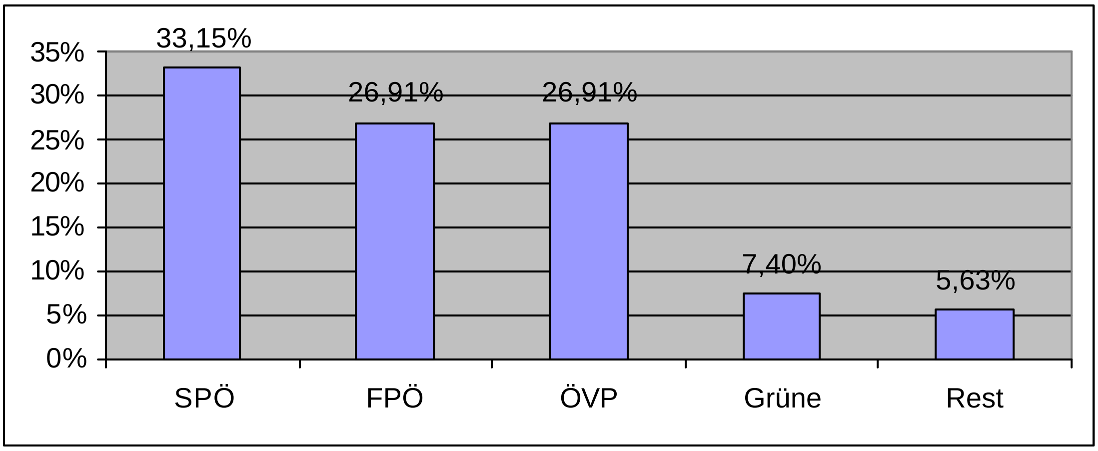
I would like to illustrate this with a problem taken from Reichel-Litschauer-Groß (2001) concerning the parliamentary election. Problem 213 asks the twelve-year-old pupils to calculate the relative frequency in %, the pie graph and the histogram. With a hyperlink they can study the solution given on the internet and try to get the same result.

Problem 213 - Nationalratswahl 1999

Right to vote:		5838373
valid:	79.17%	4622240

Partei	Prozente	Stimmen
SPÖ	33.15%	1532273
FPÖ	26.91%	1243845
ÖVP	26.91%	1243845
Grüne	7.40%	342046
Rest	5.63%	260232
Summe	**100.00%**	**4622240**





Histogram of the parliamentary election taken from the internet

The EXCEL-concept for ten to twelve-year-old pupils has been developed and tested by H. Groß and was presented at a meeting of the Austrian mathematics teachers in Vienna in 2001.

3. Linear regression line

In grade 8 the pupils make the first steps towards **Two-Variable-Statistics**. The first example in Reichel-Litschauer-Groß (1998) is the following:

The Millers want to buy a building plot and study the plots advertised in their newspaper. With the help of a map they have made the following table. In the first row they write the distance from the town center in km, in the second the size in m². ***x = distance, y = size***

Distance	2	3	5	10	20	25	30	41	49
Size	321	158	513	805	520	780	1800	1725	2540

Without a computer the pupils plot the data x and y as coordinate pairs by hand. By doing so they realize that y has the tendency to be directly proportional to x. This makes the pupils try to draw a straight line, which fits the points. It has proved useful to tell the pupils to draw this straight line through the point (mean of x, mean of y). Afterwards the pupils should try to rotate this line in such a way that it is close to the points.

If a computer is available in the classroom, you can demonstrate the solution of this problem easily with the help of a software program like MATHEASS. After the input of the coordinate pairs you get the function term of the approximation curve, the coefficient of determination, the correlation coefficient and the standard deviation together with the diagram.

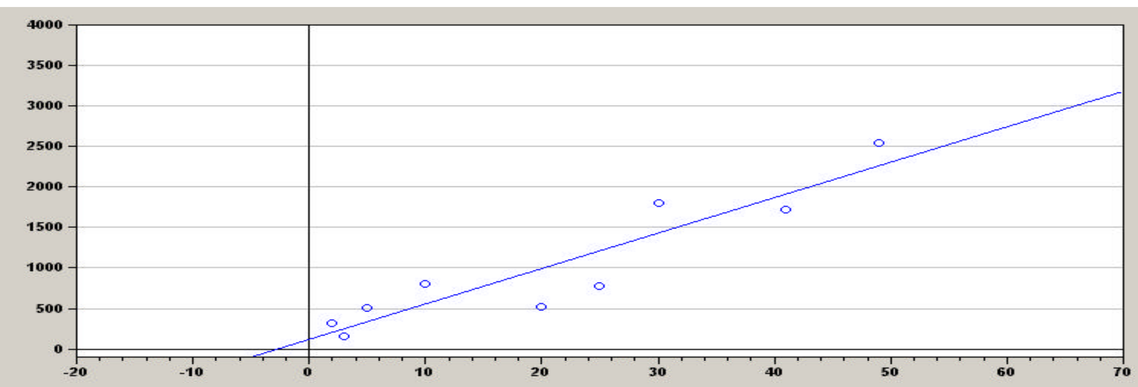
Linear Regression: y = a . x + b

y = 43.755882 x + 118.57353

9 Values
Coeff. of Determination = 0.85281527
Coeff. of Correlation = 0.92347998
Standard Deviation = 332.5105

Inverse Function : y = 0.022854x - 2.709888

x:	f(x):
2	206.08529
3	249.84118
5	337.35294
10	556.13235
20	993.69118
25	1212.4706
30	1431.2500
41	1912.5647
49	2262.6118



Scatter and regression line found with the software MATHEASS

The regression line thus found fits the data points. If all points were on the regression line, the coefficient of regression r and the coefficient of determination r^2 would be 1.

In this situation the pupils normally ask two questions:

How do you calculate: 1) the slope a and the y-intercept b, 2) the regression coefficient r?

The derivation of the formula for a and b is too difficult for 14-year-old-pupils. Therefore only the formula for the calculation of a and b is given in the schoolbooks. The formula for the calculation of r can be explained with the **concept of the covariance** without a and b. R. Diepgen has made such a suggestion in the journal "*Stochastik in der Schule*" (Heft 3, 2000).

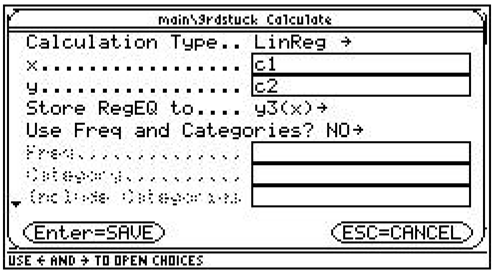
In their schoolbook "Angewandte Mathematik 4" for 18-year-old-students Kronfellner/Peschek show a concept with two regression lines. With the slopes of these two lines, r^2 can easily be calculated. With a computer algebra system (CAS) like DERIVE (TI-92) you can realize this concept even with 14-year-old pupils, if they have experiences with the handling of the data-matrix-editor and the graphic window of the TI-92.

4. Two linear regression lines – Referendum Temelin

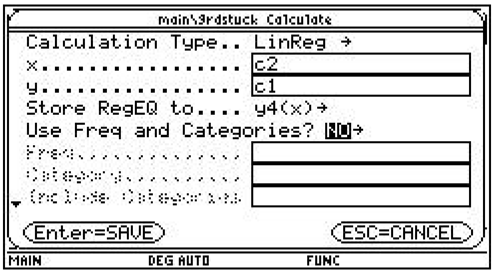
The citizens of Austria and Germany are very afraid of radioactive radiation. In Temelin (CR) an atomic reactor has been built near the borders in the last years and the test phase began. In January 2002 an Austrian party started the referendum "Veto against Temelin" to stop the operation. In the results published in the newspapers on January 22, 2002, the tendency the nearer to Temelin, the higher the percentage of people who went to sign the referendum, was remarked by the students of form 11. They suggested to check this. They took the data published in the newspaper and found the distance to Temelin with the help of the software program geothek which they often used in their geography lessons. The distance they put into the table was always the distance from Temelin to the county capital town.

Counties	B-land	Vienna	U. Austria	L. Austria	Salzburg	Tirol	Carinthia	V-berg	Styria
main towns	E-stadt	Vienna	Linz	St.Pölten	Salzburg	I-bruck	Kla-furt	Bregenz	Graz
percentage	14.8	15.4	23.5	16.9	13.4	8.7	15.5	6.7	12.0
distance	203	170	74	116	169	295	254	381	223

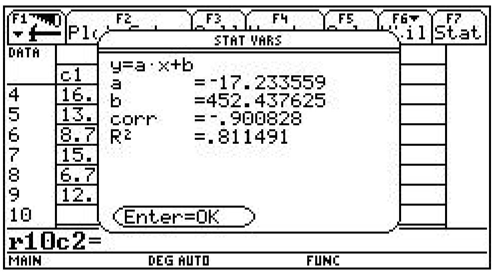
The data were entered as columns c1 (percentage), c2 (distance in km) in the data-matrix editor.



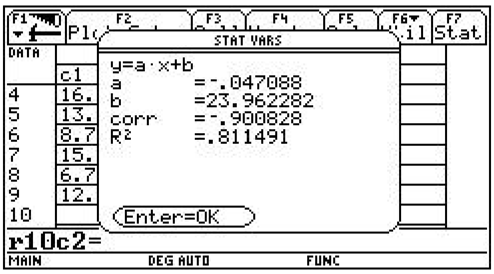
Linear regression with **x = c1, y = c2**



Linear regression with **x = c2, y = c1**



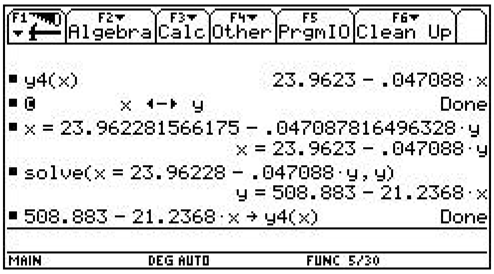
1. regression line has the slope **a = -17.233559**



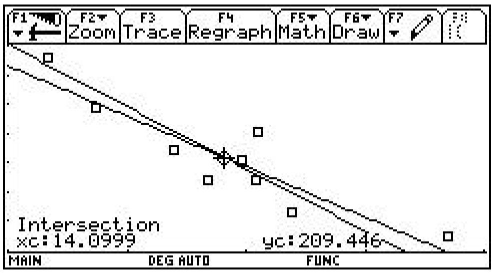
y4(x): slope **a** = **-0.047088**

Coefficient of determination **r^2 = (-17.233559) ⋅ (-0.047088) = 0.811491,** r = √(0.811491) = **0.90083**

To find the right **2. regression line** you now have to take the **inverse linear function of y4(x)**.



2. regression line: y = -21.2368 x + 508.883



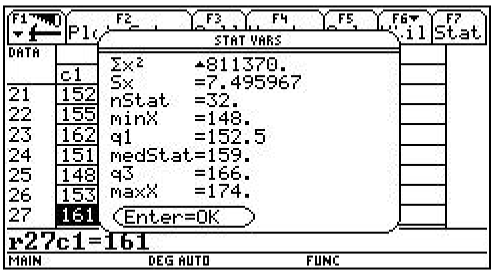
Intersection point: [mean of x, mean of y]

The coordinates of the intersection point of the two linear regression lines are the mean of x 14.1% and the mean of y 209.5 km. The angle between the two lines is small and the correlation coefficient r 0.90. From this follows that the inverse proportion is very strong.

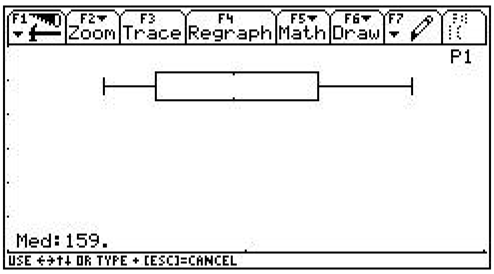
5. Relationship between length and mass of a person

It is very important not to take sets of data out of the school books only. I make my students of grade 8 work with their own data (length and mass). Looking at their own data allows the students to get a better understanding of the concepts of statistics, especially of linear regression and correlation.

The concepts mean, median and histogram have always been well known in Austria, but the concept of the quartiles has not. The quartiles q1, q2 and q3 divide the ordered set of data into four equal parts. The quartiles and the boxplot were contained in the new curriculum of form 7 in 1987 for the first time. In the commentary to the curriculum the chairman, H. Bürger, explained these concepts taking the data of population and area from the European countries 1988. In 1995/96 the TI-92 came on the market in Austria and for the first time a calculator could calculate q1, q2, q3 and could make a boxplot. In the boxplot you can see the value of minX, q1, medStat, q3 and maxX with the help of the cursor if you use the TRACE-Mode.

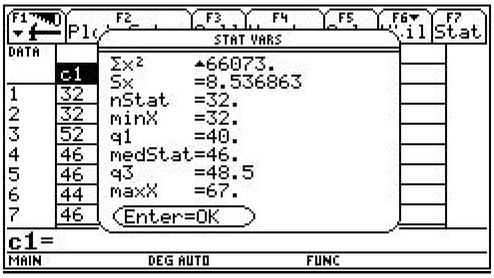


all data for the boxplot are calculated and shown

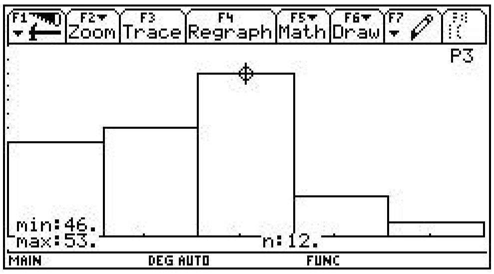


Boxplot of the length

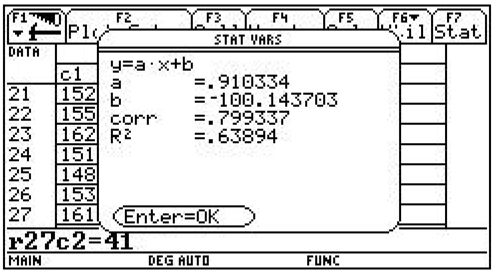
You can find all the important statistical data of the mass of the pupils with the command ShowStat on the screen of the TI-92. You can also very quickly make a histogram with the TI-92 and study it in the TRACE-Mode.



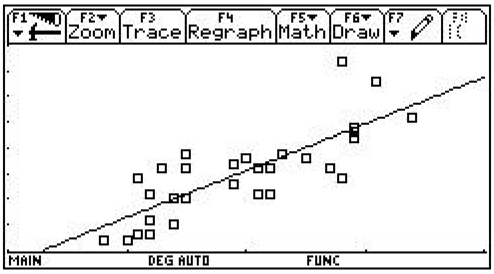
Calculation results of the mass



Histogram of the mass



correlation coefficient r = 0.799337

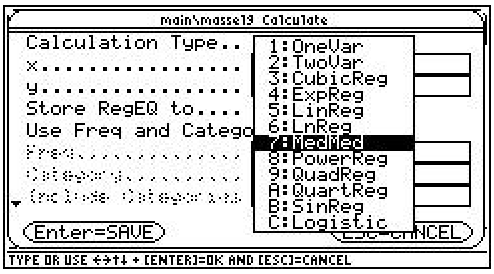


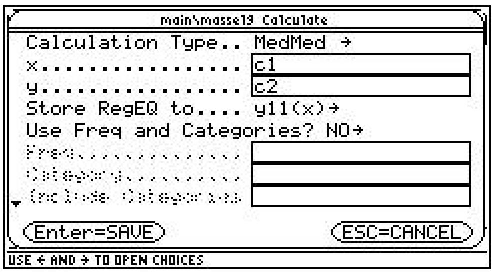
Points with the linear regression line

You can make a scatter plot with the TI-92 (x = length, y = mass) and find the equation of the linear regression line. The correlation coefficient is r   0.80. If r   0.8, you can rightly say that the mass becomes greater if the pupil is taller. But there are some exceptions, which have a great influence on the position of the regression line.

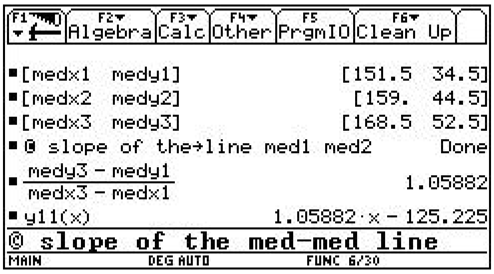
6. Med-Med Line – a new regression line

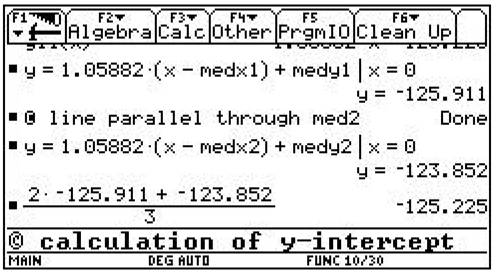
The TI-92 offers the possibility to choose another type to fit the points with the **med-med regression**. It is a linear regression, which is not so sensitive against run away data as the linear regression. The equation of the med-med line can easily be derived with the knowledge of the linear function only.



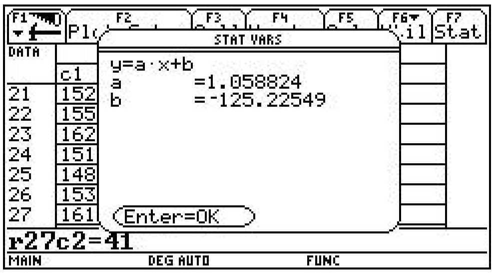


In the TI-92 guide book (1995) the calculation type MedMed is described in the following way:
"*Median-Median - Fits the data to the model* ***y = a . x + b*** *(where a is the slope, and b is the y-intercept) using the median-median line, which is part of the resistant line technique.*"
The ascended ordered set is divided into three subsets which have an equal number of elements (if possible). In each of the three subsets the median is taken. In any case it is possible to control them. The TI-92 calculates and stores them, but they are not displayed on the screen. You have to call them: [medx1, medy1] [medx2, medy2] [medx3, medy3]
Afterwards the slope of the med-med line can be calculated with the points med1 and med3, the y-intercept with the y-intercept of med1 med2 and the y-intercept of the parallel line through med2.

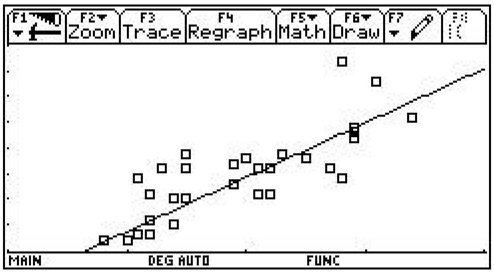




The med-med line is not only easier to calculate, it also fits the data better, which can be seen in the next figure.



slope a and y-intercept of the med-med line

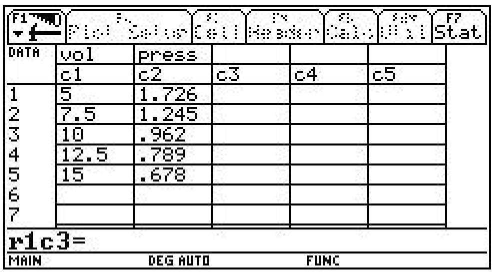


points with the med-med line

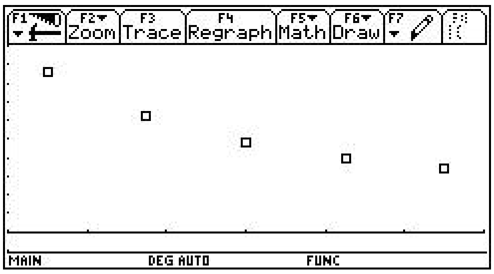
If you compare the med-med line with the linear regression line you see that the line fits the points better and is not so sensitive against the run away data.

7. Evaluation of Physical Experiments – The Law of Boyle

In many grammar schools a new subject, named **science-lab**, has been created in grade 8. In this subject the pupils make experiments in little groups. One new possibility is to work with the TI-92 and a **Calculator Based Laboratory (CBL)** which allows to collect data during physical and chemical experiments (B.&A. Aspetsberger, 2001). E.g. the students can discover Boyle's law, they can find the relationship between pressure and volume of a confirmed gas. In the subject science lab, the pupils start collecting data by varying the volume of the gas and simultaneously measuring the pressure. Both, volume and pressure, are stored to a data matrix of the TI-92: volume V in the column c1 and pressure p in c2. In mathematics the pupils define **a scatter plot** and visualize the data in a graphical window.

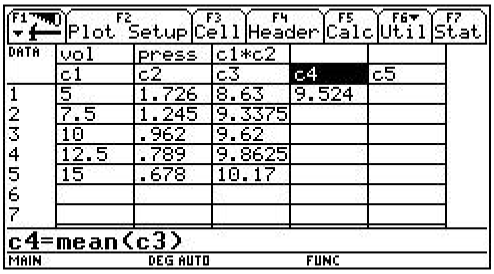


the data for volume and pressure

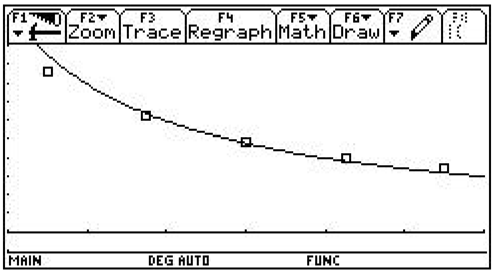


a scatter plot of the points [v, p]

Now they try to find a good regression type. The pupils can find out that the volume in c1 is increasing f the pressure in c2 is decreasing and they notice that the points are not lying on a straight line. The pupils of this age have the advantage that they know only one type which is not linear and in inverse proportion. It is the type **x . y = a (**=constant). They have sometimes solved such problems before, e.g. they had to find the time for different speeds if the distance was constant (e.g. 60 km). Therefore they try to multiply the volume with the pressure. The product in c3 is nearly constant. Which value is to be assumed for a? They try a plot with a = 9.524 (=mean(c3)).

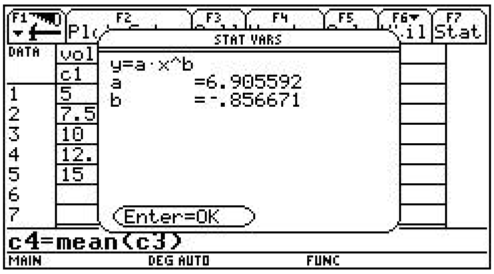


c3 = c1 ⌡c2 and mean(c3) = 9.524

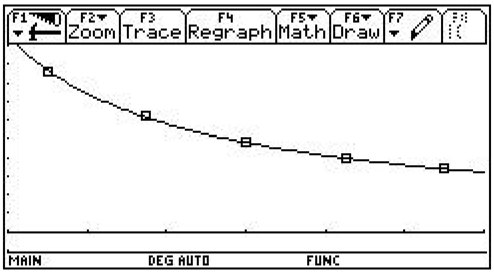


the graph of x ⌡y = 9.524 fits the points

B. & K. Aspetsberger (2001) have worked with students of 17 to 18. They made experiments in chemistry and physics in their science courses. The students were not content with the results because the data points for small volumes deviated from the graph. They tried next to calculate with the type **power regression**. The graph fitted a little better, but the exponent with -0,86 deviated too much from –1.

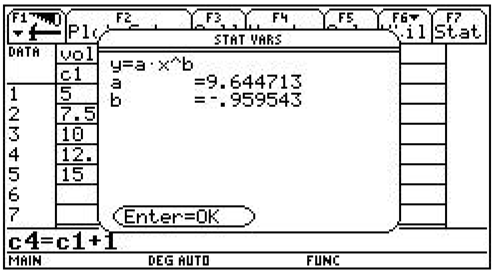


the exponent b 0.856671

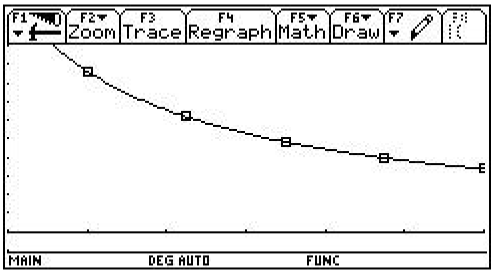


the graph fits better the points

The students controlled the experiment stepwise and found reasons to add 1 to c1. Now the calculation type **power regression** fitted the data points quite well.



the exponent b 0.96 is near -1



the graph fits the points quite well

8. Statistics and Computer – a Chance for Teacher Training

Probability & statistics are a compulsory part in the final grades of all schools in Austria which prepare students for university level. Up to 2001/02, **Computer science** has been a compulsory subject for all pupils in grade 9 only. The Ministry of Education now plans to make computer science a **compulsory subject for all pupils of grade 5 in autumn 2002**. In the other grades the pupils can choose computer science as a voluntary subject. The teacher students at university have a very different knowledge in working with a computer and they rarely have experience how to teach statistics with the help of a computer at school. Therefore many teacher students find it very difficult to plan such a statistics lesson.

The use of computer packages at school results in **more independent productive pupils activity**. Individual pupil activity has nearly erased pupils calculating on the blackboard. **Computer lessons imply less class teaching and more partner and individual work as well as less note taking and more production**. On the other hand a lot of time is devoted to the pupils to enable them to work with the computer and the program (Nocker, 1996). But the use of computer packages is a great temptation for the pupils, because such functions and programs can quickly be used as a blind tool. (Wurnig, 2001)

In one of my lectures at university, I work with the teacher students in a computer lab with statistics software and try to develop concepts of the school curriculum with them. In this lecture they do not only have to learn how to use the computer in the right way, but, in addition, they also have to learn that **a mathematical concept has different levels of precision** (R. Fischer, 1985). They have to experience personally how computer algebra systems change their learning. H. Heugl (1997) states **three stages in learning mathematics** if students use symbolic computation systems in the classroom: the heuristic stage, the exact stage, the application stage. He points out that the experimental or heuristic phase often does not exist in the traditional mathematics education.

REFERENCES

- Aspetsberger, B. & K., 2001, "*Experiences with CBL and the TI-92 in Austrian High School Classes integrating Math, Physics and Chemistry*", CDROM T³ Euope.
- Bürger, H., 1988, „*Statistik*", Mathematik AHS Kommentarheft 2, Wien, ÖBV, 84-101.
- Diepgen, R., 2000, „*Kovarianz oder Bestimmtheitsmaß*", Stochastik in der Schule, Heft 3, 29-32.
- Fischer, R., Malle, G., 1985, „*Mensch und Mathematik*", Mannheim/Wien/Zürich: BI, 174-177.
- Groß, H., 2001, „*Elektronische Hilfsmittel in der Unterstufe – insbesondere EXCEL*", ÖMG Didaktik-hefte, Heft 33, 58-70.
- Heugel, H., "Symbolic computation systems in the classroom", The International DERIVE Journal, Vol. 3, No. 1, 1996, 1-10.
- Kronfellner, M., Peschek, W., 2000, „*Angewandte Mathematik 4*", Wien, öbv & hpt, 67-84.
- R. J. Nocker, "*The impact of DERIVE on classroom methodology*", The International DERIVE Journal, Vol. 3, No. 1, 1996,. 73-89.
- Reichel, H-Ch., Litschauer, D., Groß, H.,2001, „*Lehrbuch der Mathematik 2*", Wien: hpt. problem 213.
- Reichel, H-Ch., Litschauer, D., Groß, H.,1998, „*Lehrbuch der Mathematik 4*", Wien: hpt, p. 158.
- TI-92 *Guidebook*, 1995, Texas Instruments, p. 196.
- Wurnig, O., 2001, "Vorteile und Gefahren des Einsatzes von CAS im Stochastikunterricht", in Beiträge zum Mathematikunterricht, Hildesheim/Berlin, 672-675.