STUDENTS - STATISTICAL SOFTWARE - STATISTICAL TASKS: A STUDY OF PROBLEMS AT THE INTERFACES

Rolf Biehler, Institut für Didaktik der Mathematik (IDM),
Universität Bielefeld, Germany

When students work on a data analysis problem using software, we can briefly speak of two interface problems: the interface „statistical task - task for the software” and the interface „subject matter task - statistical task”. We studied these problems on the basis of videotaped and transcribed sessions with 4 pairs of students who co-operatively worked on statistical problems using data analysis software and a complex data set. We were able to reconstruct different patterns of software use: computer-centred patterns and statistically more reflective patterns. The statistics-software interface caused problems because students were not aware of the new meanings of variable types, data tables and other concepts in a computational environment. With regard to the subject matter - statistics interface, some students were reluctant to use the distribution concept when further explaining variables for variability were at hand.

INTRODUCTION

The empirical data on which this exploratory study is based are videos and transcripts of 130 minutes sessions with pairs of student teachers who had participated in an introductory statistics course finished five weeks earlier. We presented some statistical tasks to them and gave them time to explore the data by themselves with the software tool (BMDP New System for Windows 1.0, *Bmdp_ns*) they were acquainted with from the course. Secondly, the students were asked to present results to the interviewers. Thirdly, we interviewed them about the data and their results. The students were permitted to resume their data analysis work and use the software during the interview. Four pairs of students participated in the study.

We were interested in students' statistical thinking, in the patterns of software use they had developed and whether the software use supported (or impeded) their thinking about a statistical problem. Using pairs of students allowed us to analyze how they negotiated work on the task and software use. The study is exploratory in the sense that our intent was to explore the "transcript data" to develop hypotheses and conceptualizations, remaining open to unexpected results.

When students work on a statistical problem using analysis software, we can perceive two interface problems. A given task has to be transformed into a task for the software, and output from the software has to be interpreted in terms of the original task. A second aspect consists of transforming a subject matter task into a statistical task and transforming back a statistical result into subject matter terms. Both processes are intertwined and partly circular. We know that the statistical knowledge available to the student will influence the way a subject matter problem is transformed, and the same is true with regard to software whose options will constrain statistical thinking (Biehler, 1997b).

TASKS AND DATA TO BE ANALYZED BY THE STUDENTS

The data set contained data on traffic accidents in Germany. It consisted of 365 cases representing the 365 days of the year 1987. For each day, 29 variables were reported: 4 variables representing "time": *day number* (1 to 365), *month*, *date* (1 to 30, or. 31, or 28 in February) and *weekday* (Monday to Sunday). The location of an accident was classified as occurring inside communities (in), outside communities without highways (out), and highways (hi). Two derived variables were added, namely the total number *tot=in+out+hi* , and *hi+out*. So, location was available in 5 types. For every day and location type, the number of accidents with injured people and those with only heavy material damage, and the number of lightly injured, heavily injured, and deaths were reported. In sum, for each of the 5 location types 5 variables were collected. These are 25 variables plus the 4 time variables. The students knew the similar data set for 1986 from the course and had explored these data with regard to differences between location types, seasonal trends, and differences between weekdays. I have developed an analysis of these data elsewhere to provide a prototypical example of data analysis in the sense of Tukey's detective work. From this analysis, we were able to develop expectations concerning the data-analytical and didactical potential of this data set (Biehler, 1992).

We gave the students four tasks related to the data. The first task was the following:

> John Datalover asks: "How many accidents happen every day on average?" Perform a statistical investigation for him and present him your results. Hints: The question concerning the average is ambiguous and insufficient. Answer the question concerning the average by using several adequate "averages" and explain them to the inquirer. Moreover, show him that averages provide only limited information by presenting to him further informative numerical summaries and graphs.

We wanted to explore how an everyday question was transformed into a reasonable statistical one by the students. The software was a means to enable students to make their method choices for the open problem. On the other hand, difficulties using the software were also observable. We will briefly sketch some of our results.

THE SUBJECT MATTER - STATISTICS TRANSPOSITION

Average and distribution

The task contained the question concerning the inadequacy of the overall average. We had expected that students would tend to use the mean and mention the median only secondarily. We did not expect that some would interpret the “ambiguity of the average” as related to the means of the different variables concerning accidents that were in the data table. Most students grouped the number of accidents by weekday and found different averages for each weekday, which were partly explainable by traffic behavior (Friday above average, Sunday inside communities below average). Some recognized a nonlinear trend when plotting the number of accidents against day number (higher in the period May to October, partly explainable by higher amounts of private traffic and less cautious traffic behavior than in winter time). Students used both phenomena to criticize the value of an overall average. Students seem to hold the intuitive notion that there should be no systematic deviation from the overall average (with regard to time or another further variable) for it to be considered a legitimate summary. This should be explored in further studies. Other student arguments referred to the common sense opinion that the average does not inform us about single exceptional cases or circumstances. However, none of the students gave arguments that located the average within a system of location and spread parameters that would give a more adequate summary of the distribution than only one number. No student offered a measure of spread except one group that pointed to the spreading out between minimum and maximum. After a further request from the interviewer, only one student explicitly used a distribution graph (box plot or histogram) to show the limited distribution information provided by an average, despite the fact that all of them had produced these graphs. In another study, we had observed high school students who probably had not constructed the notion of a statistical group with certain group properties as legitimate objects (see Biehler, 1997b; Konold et al., 1997). We could see the findings in this light. However, we tend to regard our new findings as indicating a more mature judgment in the spirit of EDA, namely to deliberately decide which variation will be described as distribution and which variation will be “explained” by further variables accepting residual variation in a second step (Biehler, 1995).

What is a daily average?

The statistical transposition is concerned with the interface between subject matter knowledge on the one hand and concepts and statistical concepts on the other. We expected that translating the question as one requiring the use of the program to determine the mean or median of the respective data column would be straight forward. That was not the case. The understanding of time as continuous interfered with the discrete structure of days, weeks, months. We can have weekly averages, daily averages for each weekday, daily averages for each week of the year. One group superimposed a linear regression line on a scatter plot of accident number against day number and interpreted it as showing how the “daily average” changed over time during the year. This can be regarded as an indication of understanding the regression line as providing a model for describing how the dependent variable changes with the independent variable “on average”.

Another pair of students wanted to calculate the mean but did not remember where to find the menu option for the mean. They reinvented other “middle values” by using the software feature to group continuous variables into groups of equal frequency or into intervals of equal length. They selected grouping into two groups and came up with a number equaling the median in the first choice and equaling the *(min+max)/2* in the second. Although the students were not able to theoretically relate the various averages, the episode shows how even a limited expressive power of software can be used for defining new avarages. It also illustrates how solutions are contrained by available software options.

THE COMPUTATIONAL TRANSPOSITION: THE MEANING OF VARIABLE TYPES

The implementation of knowledge in software is a transformation that usually includes a change of meaning. Balacheff (1993, pp. 147) has called this general problem the “computational transposition”. This change can be source of problems for the software user and at the same time the basis for encouraging new types of thinking if the software offers new objects, operations and representations “to think with” (Kaput, 1992). A typical problem is the concept of variable types. We know that several classifications are in use in statistics, and every software offers a slightly transposed representation of such a classification. *Bmdp_ns* distinguishes two major types with subtypes: character (pure, nominal, ordinal) and numerical (continuous, nominal, ordinal). The meaning of different variable types is often seen with regard to the measurement process, but the software implementation adds a new dimension: the “meaning” is also dependent on the kind of operations in the software that can be done with the different variable types. Statistical transformations between variable types and how available statistical methods require certain variable types are relevant here. For instance, the menu command “bar chart” is only executable for nominal and ordinal variables. For continuous variables, the command “histogram” has to be used. Nominal variables are formally represented by a list of “values” with an “attached” list of corresponding “group names”. The user can assign the same group name to different values, for instance, to collapse categories. Some procedures use the values, some the group names. The user can permanently “attach” a categorization of a continuous variable to such a variable. This discretization is used by the program for default histograms. Another use is analysis by group. Nominal and ordinal variables can be used as grouping variables. A continuous variable can only be used as a grouping variable if an attachment exists, but this does not mean that all procedures for nominal variables would accept the nominal attachment as input. Ordinal-numerical variables can be used as one of the variables in scatter plots, nominal-numerical and ordinal-character variables cannot. For producing a side by side dot plot of a variable grouped by a nominal-numerical variable, it helps to change the type to ordinal-numerical first. The “nature” of the variable has to be changed to make a certain operation possible. It is clear that a user needs a very differentiated mental model for operating successfully. Quite a few difficulties students had were related to this aspect of the computational transposition. The problems were exacerbated by further sources of difficulties: an unclear conceptual distinction between bar charts for nominal variables and histograms for continuous and difficulties with interpreting (density) histograms with unequal classes, which appeared unexpectedly because students had chosen an attached categorization based on intervals with equal frequency. The latter problem is also related to the computational transposition, namely that the software agent “histogram” behaves differently than expected because of some previously changed options whose “software meaning” was not completely clear.

Several students confounded a classification for an analysis variable such as number of a*ccidents inside communities* and a discretization of the time axis. For instance, they wanted to determine the number of accidents for every month and to display this distribution as a bar chart suggesting to take 12 classes in the histogram assuming that this classification would provide months. The display they had in mind could have been produced by selecting a bar chart for the variable *month* and assigning the role “case frequency” to the variable *x* = *accidents inside communities.* This specification is obviously not very near to the semantics of the situation and another indication of the computational transposition. In the *Medass* prototype, we favoured specifying the same display by first calculating the sum of *x* grouped acccording to *month*, thereby producing a list of 12 numbers that can be plotted against month in a bar chart in a second step (see Biehler, 1997a). Although we can consider this as a better generalizable adaptation to such situations, the computational transposition shows up in a different way. The analysis of the transcripts also shows that students wanted to make different time aggregations to see structure. The representation of time by three variables was difficult to handle and time aggregating operations were not available in *Bmdp_ns*. We could ask for more adequate data structures that can better represent the semantics of time series. Such data structures could better support students’ conceptual thinking about time series, although it would make a software more complicated in the first place.

PATTERNS OF SOFTWARE USE AND STATISTICAL ACTIVITY

We intended to reconstruct different patterns of software use similar to Krummheuer’s (1993) approach. We were able to observe that software tools function as cognitive tools that empower and transform thinking in valuable ways, for instance by encouraging an exploratory approach to data analysis. We also observed, however, trial and error exploration driven by directly available software options. Most student groups were inclined to produce many windows or pages of output at the expense of constructing reasonable interpretations of a selection of graphs and of providing answers to the original subject matter problems. The planning and negotiation stage before using software commands was short, and much of the conversation was characterized by reference to software terminology and computer jargon. The temptation was great to change to another graph if some output needed more time for proper understanding. Nevertheless, we were also able to reconstruct different patterns and qualities of integrating software use. One group usually planned steps of analysis from a statistical point of view, negotiated and decided on how to use the software in a second step, formulated expectations before a new plot or summary was produced, and compared the result to their own expectations.

CONCLUSIONS AND PERSPECTIVES

Some results of students’ work on the first task were reported above. The other tasks were more concerned with using the data to answer subject matter questions.

Students were asked to find out different seasonal and weekly patterns of the accidents of different location types and develop explaining hypotheses. We found interesting student strategies for handling the complexity of the data and for interpreting multiple box plots. Difficulties in relating subject matter and statistical findings were revealed. The linking of windows was responsible for some unexpected student difficulties. The multiple window environment posed some unexpected obstacles for adequate comparisons of statistical graphs because of distorting icons and uncoordinated scales. Results will be reported elsewhere.

REFERENCES

Balacheff, N. (1993). Artificial intelligence and real teaching. In C. Keitel and K. Ruthven (Eds.), *Learning From Computers: Mathematics Education and Technology* (pp. 131-158). Berlin: Springer.

Biehler, R. (1992). *Intendierte Anwendungen und didaktische Begründungen zu einem Softwarewerkzeug zur Explorativen Datenanalyse und stochastischen Simulation für Schule und Ausbildung*. München: Inst. f. Film und Bild in Wissenschaft und Unterricht (FWU).

Biehler, R. (1995). Probabilistic thinking, statistical reasoning, and the search for causes - Do we need a probabilistic revolution after we have taught data analysis? In J. Garfield (Ed.) *Research Papers from ICOTS 4, Marrakech 1994* Minneapolis: University of Minnesota.

Biehler, R. (1997a). Software for learning and for doing statistics. *International Statistical Review, 65(2)*, 167-189.

Biehler, R. (1997b). Students' difficulties in practising computer supported data analysis - Some hypothetical generalizations from results of two exploratory studies. In J. Garfield and G. Burrill (Eds.), *Research on the Role of Technology in Teaching and Learning Statistics* (pp. 169-190). Voorburg: International Statistical Institute.

Kaput, J. J. (1992). Technology and mathematics education. In D. Grouws (Ed.) *Handbook of Research on Mathematics Teaching and Learning* (pp. 515-555). New York: Macmillan.

Konold, C., Pollatsek, A., Well, A., and Gagnon, A. (1997). Students analyzing data: Research of critical barriers. In J. Garfield and G. Burrill (Eds.), *Research on the Role of Technology in Teaching and Learning Statistics* Voorburg: International Statistical Institute.

Krummheuer, G. (1993). Orientierungen für eine mathematikdidaktische Forschung zum Computereinsatz im Unterricht. *Journal für Mathematikdidaktik, 14(1)*, 59-92.