COMPARING GROUPS BY USING TINKERPLOTS AS PART OF A DATA ANALYSIS TASK – TERTIARY STUDENTS´ STRATEGIES AND DIFFICULTIES

Daniel Frischemeier
University of Paderborn, Germany
dafr@math.upb.de

The comparison of distributions of numerical variables is a fundamental idea in descriptive statistics. Preferably those processes are to be embedded in a data analysis-cycle. This emphasizes working with real and multivariate data and generating interesting statistical hypotheses. In the context of preservice teacher education, we designed an experimental course comprising 15 sessions on data analysis with TinkerPlots (Konold & Miller, 2011) in which Group comparisons played a fundamental role. After the course we have conducted a video study where we observed the participants while comparing groups with TinkerPlots. In the paper we will focus on several steps and the frequency of their occurrence, which can be identified when learners do group comparisons facilitated by software.

INTRODUCTION

The comparison of distributions of numerical variables is a fundamental idea in descriptive statistics (Konold & Higgins, 2003). Example questions motivating group comparisons are: “Do girls tend to spend more time on homework (per week in hours) than boys?” or “In which respect do men and women differ regarding their income?” Preferably those processes and questions are embedded in a data-analysis-cycle (like the PPDAC-Cycle, Wild & Pfannkuch, 1999) where learners have the opportunity to pose their own statistical questions, generate their own statistical hypotheses, construct an instrument for collecting data (mostly a questionnaire) and then analyze their collected data and make conclusions on it. When analyzing a huge amount of data, the use of adequate software becomes inevitable. These aspects motivated us to design a course for pre-service teachers to deepen their statistical knowledge combined with using adequate software. The course “Statistical reasoning with TinkerPlots” (for details see Frischemeier & Biehler, 2012) emphasizes working with real and multivariate data and generating interesting statistical hypotheses. Group comparisons played a fundamental role in this course, as well as the use of adequate software (TinkerPlots) for facilitating the data-exploring process. Consequently we want to get an insight in typical phases (“real problem”, “statistical problem/ statistical activity”, “use of software”) which are performed by learners when doing group comparisons facilitated by software.

GROUP COMPARISONS WITH SOFTWARE

Generally there is plenty of research relating to general conceptions and misconceptions of learners when comparing groups. Heaton and Mickelson (2002), for example, observed in their studies that their participants get lost in the process of data analysis and that they concentrate on the creation of plots but neglect the interpretation of them. Francis (2005) identified similar issues (description and interpretation of plots is neglected) when preservice teachers were doing a statistical project. This neglect of description and interpretation may be caused by difficulties learners have with describing and interpreting graphs as it can be also found in Biehler (1997), Biehler (2007) and Bruno and Espinel (2009).

In this article we want to focus on software use while comparing groups: Biehler (1997) describes a cycle of computer-supported statistical problem solving, which includes the following four components: *Statistical problem* -> *Problem for the software* -> *Results of software use* -> *interpretation of results in statistics*. He observed that “[…] we can often reconstruct in our students a direct jump from a real problem to a problem for the software without an awareness of possible changes” and further points out that “[…] students are satisfied with producing computer results that are neither interpreted in statistical nor subject matter terms”. Finally Biehler speaks from a “degenerate use of software for problem solving, where it only counts that the computer does it.”(Biehler, 1997, p. 175) There is also research in distinguishing between different types of using software while doing data analysis tasks. Makar and Confrey (2013) describe a possible approach when exploring the procedure of learners when doing a data analysis task with Fathom

In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, Arizona, USA.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org [© 2014 ISI/IASE]

and distinguished between three different types/approaches: “wondering”, “wandering” and “unwavering”. Wonderers for example have a view on the data with a theory in their mind, they seem to be goal-orientated as they try to go through the data and “seek evidence to support, refine and extend their theories”. Wanderers “have no particular evidence in mind when going into the data” (Makar & Confrey, 2013, p. 357), they look on the data, explore to see if anything “popped out” at them. The unwavering approach can be identified for example “by the decision pathway used: investigators looked for a particular piece of evidence to support or refute their original conjecture, and once they found it they were satisfied that they had answered the question put to them.” (Makar & Confrey, 2013, p. 357). The goal of this article is to refine the data analysis cycle of Biehler (1997) in the sense of adding potential new steps/phases and making a frequency analysis of the occurrence of the phases with the intention to identify patterns regarding learners´ procedure while conducting a group comparison with TinkerPlots.

RESEARCH QUESTIONS

In general this research question emerges: *Which typical phases can be identified when doing a group comparison task with software?* From this question the following sub-questions can be derived: *What is the proportion of the phases in their data analysis process? How far do learners make conclusions or interpretations of their findings?*

METHOD

We conducted an exploratory video study. The participants of this study were the participants of the course for pre-service teachers, described in the introduction. The intention was to observe the solving and cognitive processes when doing a group comparison task with TinkerPlots with special focus on the devolution of the steps and the use of the software in their exploration process. For the interview study task, we used the dataset VSE_2006 taken from the German Bureau of Statistics which contains 861 cases sampled at random from German employees from all levels including variables such as gender, wage per month, kind of employment agreement, etc. We gave a task (see Figure 1) related to this data to the participants.

Current stats: Women are way behind men with salary

The salary gap between women and men remains high: about 23 percent is the difference between them. The reasons for the imbalance are varied; the lack of opportunities for leadership positions is not always the cause of the poor earnings of women.

YOUR TASK

In the TinkerPlots datafile you find the monthly salaries of 861 women and men of the year 2006. In which aspects do the distributions of the variable “salary” of the men and women differ? Carve out differences in both distributions!

Figure 1: VSE-task

This task was handed out to the participants alongside with a TinkerPlots-file including the dataset. The participants had to do the task in pairs where there was no intervention of us.

DATA COLLECTION

Fourteen participants (7 pairs) took part in the study. All of them had attended the course “Statistical reasoning with TinkerPlots” (see introduction) and an elementary course about statistics and probability. For the data analysis we collected several data: we recorded the screen activities via Camtasia-Studio (TechSmith Corporation, 1999-2009), recorded the communication of the participants while doing the task and collected the written forms and the TinkerPlots-files. The activities and communication were completely transcribed. In a subsequent interview-phase (stimulated recall) the participants were interviewed to elicit their thoughts and strategies while working on the task. We do not want to relate to the stimulated recall phase in this paper because for the use of software while comparing groups this does not seem to be crucial.

DATA ANALYSIS

We analyzed the data with qualitative methods. To answer our research questions we have had a look on the Camtasia-recordings and on the video- and audio-recordings and on the transcripts of the Camtasia-recordings (communication and the interaction with software were transcribed). In the following, we will present a method called "qualitative content analysis" by Mayring (2010) which has become more and more famous in empirical social research in Germany and can be applied in the case of having the intention to analyze a huge amount of transcribed data (in our case we have about 200 pages of transcribed data). A main goal of "qualitative content analysis" is the reduction of the huge amount of data in form of category systems. Kohlbacher (2006) presents an overview of the method "Qualitative content analysis": "Mayring´s qualitative content analysis tries to overcome these shortcomings of classical quantitative content analysis by applying a systematic, theory-guided approach to text analysis using categories" (Kohlbacher, 2006, p. 12). This analysis method has the aim "to filter out a particular structure from the material. Here the text [transcript] can be structured according to content, form and scaling". The basis of this analysis method is a category system that consists of variables and an exact definition of the variables. Furthermore coding rules and key examples are given for an exact assignment between coding and data material. In the following, we want to point out our procedure when constructing our categories (which we call "phases" in the sense of the phases of a group comparison process) in more detail. We constructed our categories in a mixed approach "deductively" and "inductively" (Kuckartz, 2012, p. 69). As a kind of a deductive approach - we took into account the theoretical aspects we knew about the existing research and generated categories which arose from the theory – in this case we took the model of Biehler (1997) as seen in Figure 1 as the basis of our theoretical, deductive approach and got the phases (some of them were renamed) "statistical activity", "software use", "results of software" and "conclusions". In a second step -with an inductive approach- we went through our data to refine our deductively developed categories and looked for further categories emerging from the data (Kuckartz, 2012). We found the phases "statistical activity", "software use", "results of software" and "conclusions" and also two more phases which occurred when learners compare groups – one on the beginning of the exploration process where a "real problem" is expressed and one on the end of it, when looking for "reasons" which might explain the findings in the data. In addition it seemed appropriate to rename some of the prior phases, so that we finally work with these categories, when facing a group comparison task with software use: *real problem, statistical activity, software use, results of software use (reading off / documentation), conclusions, reasons*. These phases are not meant to be in a chronological order. So it would also be possible to "overleap" some phases or to alter the order of phases. All in all, the cycle visualized by the scheme in Figure 2 is assumed to be a kind of norm we would expect from our participants when doing a data-analysis task and it also displays a potential approach an expert may follow when facing a group comparison task with software.

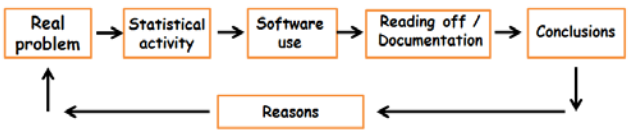

Figure 2: Group comparison cycle

PHASES WHICH OCCUR WHEN COMPARING GROUPS

In the following we describe the phases of our group comparison cycle (Fig. 2) and will give key examples for each phase. Note that we tried to construct our phases and their definition in a way that they are more or less disjoint and do not overlap:

- *Phase "Real problem"*
  At the "real problem" stage learners formulate, in their own words, what they want to explore without using statistical terms. An example for a typical quote, which was coded as "real problem" is: "Let us investigate whether men earn more than women."
- *Phase "Statistical activity"*
  One step further is when the learners enter into the "statistical world". At this stage -we call it "statistical problem/statistical activity" - they try expressing their plan and proceeding on a statistical level, for example: "We could compare the means of the two distributions". Another example is "What is the difference between the mean of the distribution of the salaries of men

and women?” Consequently a section is coded “statistical problem/statistical activity” if the students express their plan or proceeding in connection with statistical terms (like mean, median, spread, skew, distribution, etc…). This phase also includes that learners talk about their upcoming activity (“Let´s use dividers”) in the software but there is no use of the software at all at this stage.

- *Phase “Software use”*
  In this step the software is used actively in the way that the leaners use it to display or order the data, calculate values (like means or medians), etc. Example: A boxplot is displayed in TinkerPlots.
- *Phase “Results of software use (“Reading off and documentation of findings”)”*
  This stage of the exploration process covers the interaction and reaction of learners when the software provides results (such as displays or numbers). So here the results shown in the software are read off (“this is about 70.5%”, “the difference of means is 832.8€”) and documented on a sheet of paper. There are no interpretational elements at this stage; it is the documentation of observations and statistical summaries only.
- *Phase “Conclusions”*
  Following up the “reading off”-phase interpretations, comparisons and conclusions can be made by the learners. A segment coded “conclusions” was for example “The mean of the variable salary of men are higher so we can say that the men earn more than women”. Another key example that is also coded as “conclusions” is “the distributions of the variable salary of men and women are shifted, so we can conclude that the men earn more than the women”.
- *Phase “Reasons”*
  In this phase learners validate conclusions and interpretations and try to find reasons/arguments for their findings with their daily-life-knowledge. A key example for this step is “the salary of women is less because women take care of their children and therefore they can only work part-time”.
- *Phase “Other”*
  In this phase we coded elements which could not be related to the group comparison process or to the statistical process at all.

With the above categories (phases) taken as our basis we structured the data in the sense of a structural qualitative content analysis (Mayring, 2010): For our data analysis we took the whole amount of transcripts of the Camtasia-recordings and the video- and audio-recordings. We used the software MaxQDA for a computer-supported qualitative content analysis (Kuckartz, 2012) to filter out the structure of our transcripts and coded them. Mayring (2010) postulates to define coding units, in this case we defined our minimal coding unit as a word and a “unit of meaning” as our maximal coding unit. The coding units were assigned to the codes disjointly; we did not do multiple coding. After coding we made a frequency analysis (Mayring, 2010).

RESULTS

The frequencies of the occurrence of the codings are displayed in Table 1 (overview of frequency distribution of all codings) and Table 2 (frequency of codings separated by pair) below.

Table 1: Absolute and relative frequencies of phases

Pair	Real Problem	Statistical activity	Software use	Reading off / Documentation	Conclusions	Reasons	Other	Overall
Frequency	7	75	139	103	35	11	38	408
Relative Frequency	0.0172	0.1838	0.3407	0.2525	0.0858	0.0269	0.0931	1.0000

Table 1 displays the number and the relative frequencies of all codings distinguished by the phases. As we see there are huge proportions of codings belonging to “software use” (0.3407) and “Results of software (reading off/documentation)” (0.2525) but just small numbers of codings relating to “interpretational” or “reasons” level (0.0858 and 0.0269). So this table gives us the impression that “interpretation”, “conclusions” and “reason” - elements are rare when doing data analysis. This is also displayed in the proportions of codings separated by all pairs.

Table 2: Relative frequencies (absolute frequencies in brackets) of phase separated by pairs

Pair	Real Problem	Statistical activity	Software use	Reading off / Documentation	Conclusions	Reasons	Other	Overall
Hilde & Iris	0.0260 (02)	0.1818 (14)	0.4286 (33)	0.2468 (19)	0.0359 (03)	0.0260 (02)	0.0519 (04)	1.0000 (77)
Conrad & Maria	0.0143 (01)	0.2143 (15)	0.4286 (30)	0.3000 (21)	0.0285 (02)	0.0000 (00)	0.0143 (01)	1.0000 (70)
Erik & Simon	0.0000 (00)	0.1833 (11)	0.3333 (20)	0.1167 (07)	0.2333 (14)	0.0167 (01)	0.1167 (07)	1.0000 (60)
Laura & Ricarda	0.0111 (01)	0.1667 (15)	0.3445 (31)	0.3111 (28)	0.0444 (04)	0.0444 (04)	0.0778 (07)	1.0000 (90)
Martin & Wilma	0.0238 (01)	0.0952 (04)	0.1667 (07)	0.3333 (14)	0.0714 (03)	0.0714 (03)	0.2381 (10)	1.0000 (42)
Sandra & Luzie	0.0286 (01)	0.2000 (07)	0.3143 (11)	0.1143 (04)	0.2000 (07)	0.0000 (00)	0.1428 (05)	1.0000 (35)
Rico & Trudi	0.0294 (01)	0.2647 (09)	0.2059 (07)	0.2941 (10)	0.0588 (02)	0.0294 (01)	0.1176 (04)	1.0000 (34)

As we mentioned above, Table 2 reveals more precisely and separated by pairs that "conclusion"- (apart from Sandra & Luzie and Erik & Simon) and "reason" - elements (apart from Martin & Wilma and Laura & Ricarda) are rare in the processes. We also see slight differences between the pairs for example concerning their intensity of software use. Looking at Table 2, the transcripts and the Camtasia-recordings we want to present short summaries of four selected pairs of the process to the reader. We choose these four pairs to contrast two different approaches of learners when comparing groups with software use.

Hilde and Iris have in mind several summary statistics (mean, median, 1st quartile and 3rd quartile) they want to have. This is what they articulate on their "statistical activity" phase. They use TinkerPlots primarily to collect all statistical summaries and write them down ("Reading off and documentation"). Finally the results (summary statistics) produced by TinkerPlots seem to satisfy them; they do not make many connections or comparisons between the distributions nor make any interpretations. Their extensive software use (0.4286 of their codings are related to "software use") can be explained by the aspect that Hilde and Iris make further investigations in form of exploring relationships to other variables (as employment-agreement).

Similar to Hilde & Iris, Conrad & Maria use TinkerPlots primarily as a collector for summary statistics as well. They also have in mind what they want to work out, use the software to get their desired results but do not tend to make conclusions or looking for reasons to explain their findings. It is very notable that they collect summary statistics (as mean, median, 1st quartile, 3rd quartile) of both distributions but do neither connect nor compare them (just 0.0285 of their codings are related to "Conclusions" and 0.0000 of their codings are related to "Reasons").

Sandra & Luzie is a pair with a "medium" proportion on "software use" (0.3143 of their codings are related to "software use"). They use TinkerPlots in a different way as Hilde & Iris and Conrad & Maria did. Sandra & Luzie make a standard display ("histogram"). After constructing this graph with bin width of 1000€, they use it and during the further procedure they use dividers on a fully-separated plot in TinkerPlots to make extensive explorations and afterwards conclusions (0.2000 of their codings are related to "conclusions"). Sandra and Luzie have the "histogram" as a kind of "standard"-display for group comparisons in mind and try to work out differences of the distributions from this display.

Laura & Ricarda do it in a similar way as Sandra and Luzie did but used stacked dotplots as their "standard"-display in TinkerPlots. With the display itself and an extensive use of dividers they read off results (0.3111 of their codings are related to "results of software use") and make conclusions out of their findings. In contrast to Sandra & Luzie, they spend more time in their observations and also make further investigations in form of exploring relationships to other variables.

All in all, several phases (not necessarily in order), as "real problem", "statistical problem", etc. (see figure 2) can be identified when observing learners while doing group comparisons with software. Having a look at our data, we can say that we observed -as it also can be found in other studies (see literature review)- that interpretations or conclusions just cover a small proportion in the whole data analysis cycle (similar to observations made by Biehler (1997) as pointed out above). In addition -once again similar to the observation in Biehler (1997, p. 175)- it seems to be obvious that some learners (in our case especially Conrad & Maria) are satisfied with the results a software produced without having the need of interpreting or of finding reasons to explain them. Participants (like Martin & Wilma or Laura & Ricarda) having a small proportion of "statistical activity" codings tend to have the attitude to "jump directly into the software" as it was also mentioned by Biehler (1997). Looking on our pairs we can find two main attitudes while comparing groups in our data - Hilde & Iris and Conrad & Maria, for example, can be described as learners, who gather summary statistics but do not draw many conclusions from that. Sandra & Luzie and Laura & Ricarda, for example, can be described as learners who have a certain norm in their mind they want to follow when doing a group comparison.

DISCUSSION

The results show that learners may need support in a group comparison process with software. A possible support could address two aspects: on the one hand it could support to structure process of learners while comparing groups and to help them to document their findings in a kind of "data analysis scheme". On the other hand it also should enhance them to make conclusions and interpretations of their findings in the data. Further research in the development of such a material is needed.

REFERENCES

Biehler, R. (1997). Students' difficulties in practicing computer supported data analysis - Some hypothetical generalizations from results of two exploratory studies. In J. Garfield & G. Burrill (Eds.), *Research on the role of technology in teaching and learning statistics* (pp. 169-190). Voorburg, The Netherlands: International Statistical Institute.

Biehler, R. (2007). Students' strategies of comparing distributions in an exploratory data analysis context [Electronic Version]. CD-ROM. *Proceedings of 56th Session of the ISI* from http://www.stat.auckland.ac.nz/~iase/publications/isi56/IPM37_Biehler.pdf

Bruno, A., & Espinel, M. C. (2009). Construction and evaluation of histograms in teacher training. *International Journal of Mathematical Education in Science and Technology*, *40*(4), 473–493.

Camtasia Studio™. 6.0 TechSmith Corporation 1999-2009.

Francis, G. (2005). An approach to report writing in statistics courses. In B. Chance & B. Philipps (Eds*.), Proceedings of the IASE/ISI Satellite Conference on Statistics Education*, Sydney.

Frischemeier, D., Biehler, R. (2012). Statistisch denken und forschen lernen mit der Software TinkerPlots. In: *Beiträge zum Mathematikunterricht 2012* (pp. 257-260). WTM: Münster.

Heaton, R. M., & Mickelson, W. T. (2002). The learning and teaching of statistical investigation in teaching and teacher education. *Journal of Mathematics Teacher Education*. 5(1), 35–59.

Kohlbacher, F. (2006). The use of qualitative content analysis in case study research. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*. Volume 7, No. 1, Art. 21. (http://www.qualitative-research.net/index.php/fqs/article/view/75, retrieved on 10/07/2013)

Konold, C., & Higgins, T. (2003). Reasoning about data. In J. Kilpatrick, W.G. Martin & D. E. Schifter (Eds.), *A research companion to principles and standards for school mathematics* (pp. 193-215). Reston, VA: NCTM.

Konold, C., & Miller, C. D. (2011). *TinkerPlots: Dynamic data exploration* [Computer software, Version 2]. Emeryville, CA: Key Curriculum Press.

Kuckartz, U. (2012). *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung*.1. Auflage. Weinheim und Basel.

Makar, K., & Confrey, J. (2013). Wondering, wandering or unwavering? Learners' statistical investigations with Fathom. In T. Wassong, D. Frischemeier, P. R. Fischer, R. Hochmuth, & P. Bender (Eds.), *Mit Werkzeugen Mathematik und Stochastik lernen - Using Tools for Learning Mathematics and Statistics*. Wiesbaden: Springer Spektrum.

MaxQDA. Software for computer-supported qualitative content analysis. URL: http://www.maxqda.de (retrieved on 10/07/2013).

Mayring, P. (2010). *Qualitative Inhaltsanalyse*. Weinheim und Basel: Beltz. 11. Auflage.

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, *67*(3), 223-265.