

Learning descriptive statistics with the use of dynamic content generated by CindyJS

Masataka Kaneko
 Faculty of Pharmaceutical Sciences
 Toho University

Takeo Noda
 Faculty of Science
 Toho University

1 Introduction

The importance of learners' manipulations on physical models has been repeatedly emphasized in mathematics education. For instance, J. Piaget [1] pointed out that

The figurative aspects of the cognitive functions, though obviously useful and necessary to the knowledge of states, are incapable of accounting for thought, because by themselves they cannot succeed in assimilating the transformations of reality. The reason is that knowledge is not a static copy of reality. To know an object is not to furnish a simple copy of it: It is to act on it so as to transform it and grasp within these transformations the mechanism by which they are produced. To know, therefore, is to produce or reproduce the object dynamically; but to reproduce it is necessary to know how to produce, and this is why knowledge derives from the entire action, not merely from its figurative aspects.

In line with this perspective, S. Papert proposed many “models” and “ingredients” which learners can use to experience mathematizing for themselves. Figure 1 shows his proposal for a physical model with which children write the program (as seen in the right panel) to balance the inverted pendulum [2].

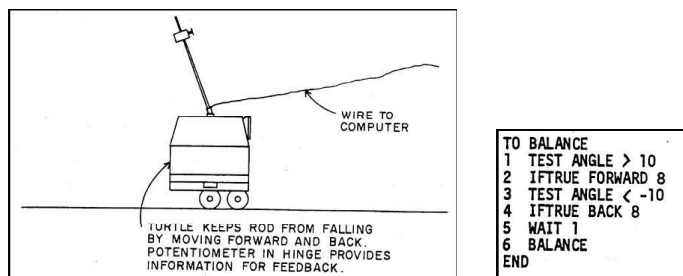


Figure 1. A “formal physical” model of the stick balancing situation

Some recently developed mathematical software systems have enabled learners to manipulate a large variety of mathematical models. Among them, dynamic geometry systems like

Cinderella (<https://www.cinderella.de>) have created an environment where a wide range of mathematical objects including geometric shapes and function graphs can be handled by dragging a mouse. However, in order for learners to handle mathematical models via these technologies, they must install those systems into their PCs and master their usage. Fortunately, Cinderella has been equipped with a function which converts graphical outputs into HTML format at the push of a button. The generated materials can be moved via the CindyJS system (<http://cindyjs.org>) on ordinary web browsers [3][4][5]. Some case studies using CindyJS system have been made in the authors' previous works [6][7].

In this paper, the methods and results of the other case study are described. The theme of this study is the fundamental concepts in descriptive statistics including correlation coefficients and line regressions. The learners' worksheets are analyzed in order to clarify what the barriers are to their reasoning with the aid of physical operations on models.

2 Theoretical Background

Besides learning materials related to elementary topics like correlation coefficients or line regression, the authors prepared a content related to a more advanced topic, orthogonal line regression. Firstly, we will examine its theoretical background.

While y - x regression or x - y regression is usually used in descriptive statistics, orthogonal regression is also important in some fields of statistics including principal component analysis. Regarding the meaning of "regression" as the best approximation of the data set, it is quite natural that the least squares method is applied to not only the errors in the values of x or y but also to the distance between each data and the resulting line. Thus, in the case of an orthogonal regression line, the targeting function is the square sum of the distance between each data (x_i, y_i) ($i = 1, \dots, m$) and the line $y = ax + b$ given as follows.

$$f(a, b) = \frac{1}{1 + a^2} \sum_{i=1}^m (ax_i - y_i + b)^2$$

Hence, at the extremal point of the function $f(a, b)$, the following conditions should be satisfied.

$$\begin{aligned} \frac{\partial f}{\partial b} &= \frac{2}{a^2 + 1} \sum_{i=1}^m (ax_i - y_i + b) = 0 \\ \frac{\partial f}{\partial a} &= -\frac{2a}{(a^2 + 1)^2} \sum_{i=1}^m (ax_i - y_i + b)^2 + \frac{2}{a^2 + 1} \sum_{i=1}^m x_i(ax_i - y_i + b) = 0 \end{aligned}$$

The abbreviated form of the first condition

$$\sum_{i=1}^m (ax_i - y_i + b) = 0$$

is the same as that of the usual line regression which implies that the resulting line should pass through the point $(\bar{x}, \bar{y}) = \left(\frac{1}{m} \sum_{i=1}^m x_i, \frac{1}{m} \sum_{i=1}^m y_i \right)$ and that the following equation should hold.

$$b = \bar{y} - a\bar{x}$$

By substituting $\bar{y} - a\bar{x}$ for b in the abbreviated form of the second condition

$$\sum_{i=1}^m (ax_i - y_i + b)(ay_i - ab + x_i) = 0$$

we obtain

$$(\overline{xy} - \bar{x} \cdot \bar{y})a^2 + (\overline{x^2} - \bar{x}^2 - \overline{y^2} + \bar{y}^2)a - (\overline{xy} - \bar{x} \cdot \bar{y}) = 0$$

It is a quadratic equation of the form

$$Ka^2 + La - K = 0$$

in which K is the covariance s_{xy} and L is the difference $v_x - v_y$ in the variances of x and y . Therefore there are two real solutions to this equation which yield candidates for the extremum positions of $f(a, b)$. Further computations lead to the conclusion that only one of these two candidates which has the same signature as s_{xy} is minimal and the other is not any extremum. As the equation degenerates when $K = s_{xy} = 0$, a discontinuous phenomenon may occur if we vary the data so that the value of s_{xy} passes through 0 and its sign changes.

In order to simplify the reasoning process, the authors chose the condition that the data points are situated symmetrically regarding line $y = x$. In this case, \bar{x} is the same as \bar{y} and L should be 0 because of its definition. Thus, in this case, the solutions are $a = \pm 1$. Moreover, if $a = 1$, b should be 0. These results imply that there can be some discontinuous phenomenon of the extremal position of $f(a, b)$ when we move the data points while keeping the above mentioned symmetry. Though direct computations would show which of ± 1 gives the minimal value of $f(a, b)$, the authors made a lesson in which students tried to find the correct choice and the conditions controlling the choice while moving dynamic content generated by CindyJS.

3 Methods

To enable students to find the discontinuous phenomena and the conditions associated with it, the authors prepared the dynamic content shown below by using CindyJS. The screen for physical operation is shown in Figure 2.

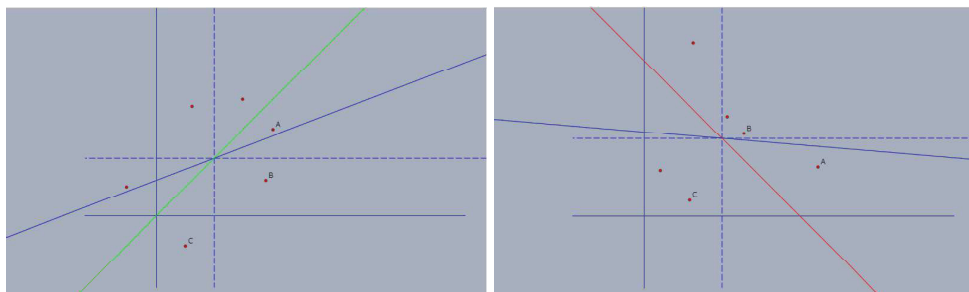


Figure 2. Screenshot of the main content

In this content, students can move three data points A, B, and C freely. The remaining three data points are moved automatically following the symmetry rule. These movements are controlled via the quite simple scripting below.

```

D.xy=[A.y,A.x]; E.xy=[B.y,B.x]; F.xy=[C.y,C.x];
x1=A.x; x2=B.x; x3=C.x; y1=A.y; y2=B.y; y3=C.y;
m=(x1+x2+x3+y1+y2+y3)/6;
draw([m,-13],[m,13],dashtype->2); draw([-13,m],[20,m],dashtype->2);
sx2=x1^2+x2^2+x3^2+y1^2+y2^2+y3^2; vx=sx2/6-m^2;
cv=((x1-m)*(y1-m)+(x2-m)*(y2-m)+(x3-m)*(y3-m))/3;
plot([m+vx*t,m+cv*t],start->-50,stop->50);
Length1=(x1-y1)^2+(x2-y2)^2+(x3-y3)^2;
Length2=(x1+y1-2*m)^2+(x2+y2-2*m)^2+(x3+y3-2*m)^2;
if(Length1<Length2,
  plot([m+t,m+t],start->-50,stop->50,color->[0,1,0]);
  ,
  plot([m+t,m-t],start->-50,stop->50,color->[1,0,0]);
);

```

As a result, the usual regression line is drawn in blue. To specify the orthogonal regression line, green or red is used in accordance with the choice of ± 1 as a . Together with some other contents, this content was implemented on iPads.

The subjects were the first grade university students (19 years old) who were non mathematics majors. After fundamental notions of orthogonal line regression and basic usage of the content were presented as shown in Figure 3, iPads were delivered to the subjects and they were asked to move the content with a partner.

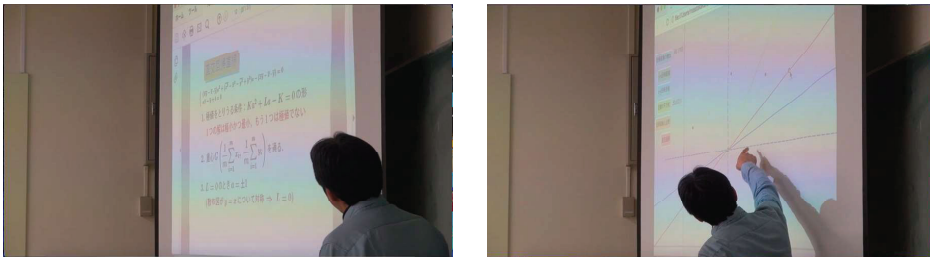


Figure 3. Scenes of experimental class

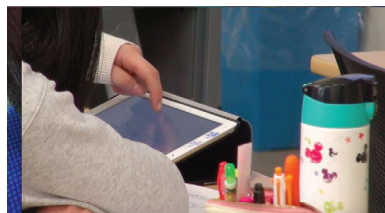


Figure 4. Students' physical operations

Students had a choice of lessons which were presented to them on iPads. Their tasks were the following.

1. Find some phenomena which they think are mathematically important and report them with the use of screenshots.
2. Explain the mathematical background of those phenomena.

Their worksheets were collected and the descriptions in them were analyzed qualitatively. The activities of some groups were recoded via VTR. Some hints were given to the groups whose operating process seemed to have gotten stuck.

4 Results

Among the 21 groups in the class, 6 groups used the content cited in the last section. As seen in Figure 5 which shows their worksheets for task 1, almost all groups found the discontinuous phenomenon together with its relation to the slope of the usual regression line.

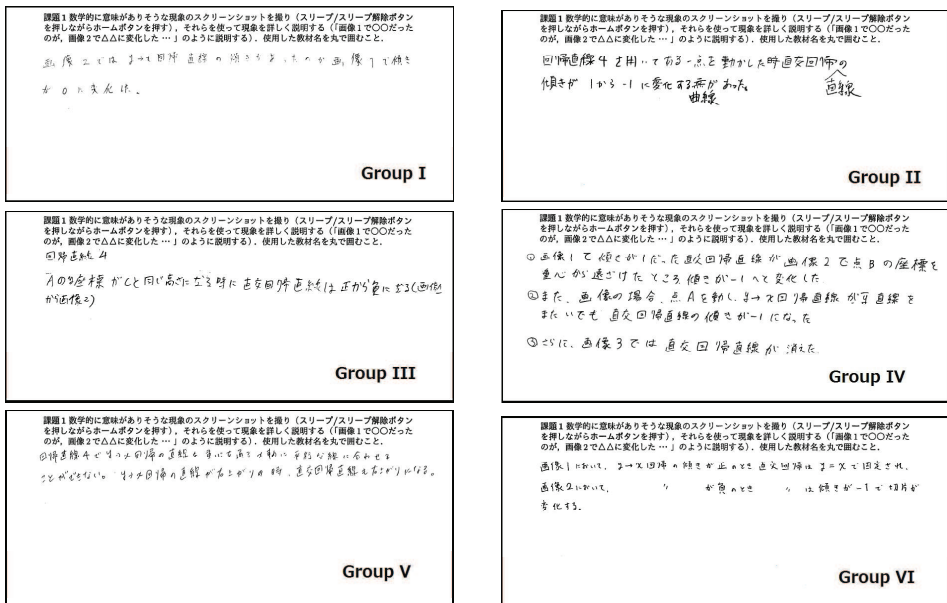


Figure 5. Worksheets for task 1

Contrarily, most groups seemed to have faced great difficulty in doing task 2. While groups III and V described their thinking as seen in Figure 6, groups I, II, and IV left almost no comments.

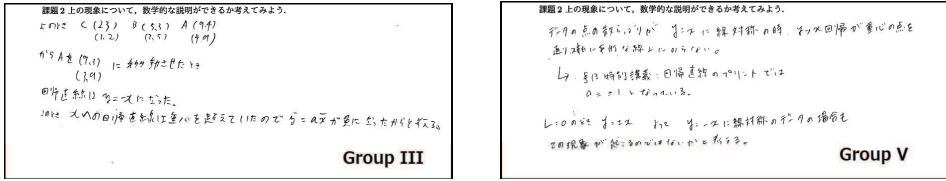


Figure 6. Worksheets for task 2

As is easily seen, these descriptions remain surface level “observations” which are not connected to the essential points given in section 2. As seen in Figure 7, only the discussion in group IV was linked to the mathematical condition

$$Ka^2 + La - K = 0$$

given in section 2 though they needed some advice from the teachers.

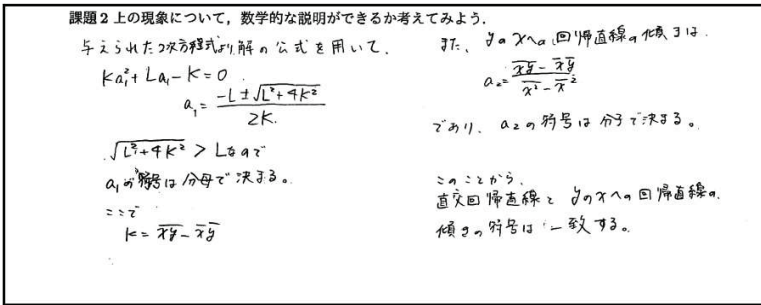


Figure 7. Worksheet of group VI

The two points where they needed advice are specified with red rectangles in Figure 8.

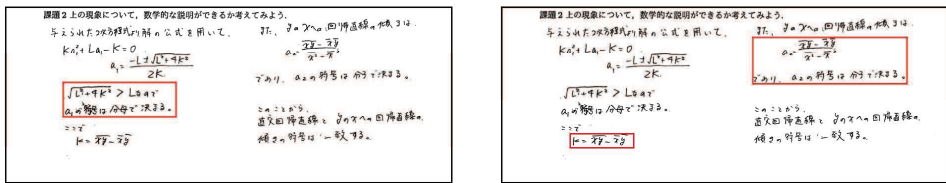


Figure 8. Points where advice was needed

The members could solve the quadratic equation and assume that the signature of the solution was related to the slope of the orthogonal regression line. However, they could not decide the sign based on the form of the solution in the first place. Similarly, in the second place, they could not infer that the slope of the usual regression line was decided by the sign of the covariance s_{xy} because the variance v_x is always positive.

5 Discussions and Conclusions

Whereas many groups found the discontinuous phenomenon, almost no group could infer its mathematical background. This result seems to give some insights for the effective use of dynamic content.

Together with the above six samples, the descriptions made by all groups did not include any inequalities though some groups treated the geometric situations which could be specified by some “range” of parameter. In the case of this task, many subjects were able to obtain the solution form $\frac{-L \pm \sqrt{L^2 + 4K^2}}{2K}$ and understood the necessity to consider its sign. However, their reasoning did not lead to the comparison between $\sqrt{L^2 + 4K^2}$ and L . This gap seems to be caused by the learners’ lack in experience of inferring with the use of inequalities.

Moreover, the comparison between two complicated quantities $a_1 = \frac{-L \pm \sqrt{L^2 + 4K^2}}{2K}$ and $a_2 = \frac{K}{v_x}$ which appear in Figure 7 and 8 is also problematic. In fact, in the curriculums of Japanese high school mathematics and university calculus courses, almost all of the descriptions related to some locus or region are given by inequalities with simple variables like x , y , or θ as seen in the samples of Figure 9.

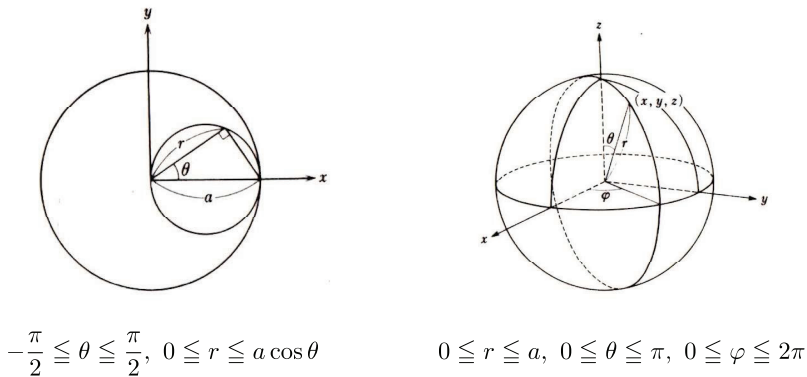


Figure 9. Inequalities appearing in a university text

(M. Sugiura: Introduction to Calculus I)

In the case of this task, both K and L are computed from the data set in a complicated manner and their relation to the position of data points is indirect. Thus it can be assumed that the link between mathematical expressions and their geometric images was lost along the way as the subjects worked out the solution.

In conclusion, the results of this case study indicate the following should be taken into account when dynamic contents are used in the mathematics classroom.

1. Discontinuous phenomenon in this case study could not be found without learners’ interactive operations on mathematical models. The result that almost all groups could find it clearly illustrates the effectiveness of using a dynamic geometry system.

- Care should be taken that learners' interactive operations on the mathematical model can be seamlessly linked to the mathematical expressions of that model. When those expressions are complicated, some preliminary training is needed.

6 Future Work

With respect to the current state of education, it is not easy for average university students to construct these mathematical models using programming language by themselves. However, especially for talented students, it will be quite useful if they can input the geometric elements and the necessary “commands” through some script editor to handle those models. Now, a function called “Cindyeditor” which acts as a script editor is now being developed. Figure 10 shows a screenshot of the Cindyeditor associated with the content in this case study.

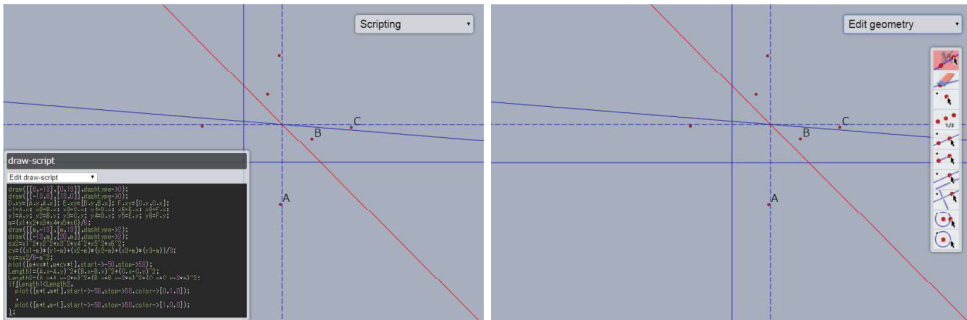


Figure 10. Screenshot of Cindyeditor

A web based editor for CindyJS widgets is currently being developed by Jürgen Richter-Gebert and Aaron Montag as part of the CindyJS-project. An online demo of the experimental state is available at <https://cindyjs.org/editor>. The users can observe how changing parts of their code would affect the widget in real-time. The authors think that the verification of the effect of using Cindyeditor for STEM education is an urgent need.

Moreover, investigating the real states in which dynamic geometry systems are used is needed. One of the most reliable ways is to track communications between learners as they move the content collaboratively. A serious problem is that it is not easy to grasp which member made which utterance. To overcome this difficulty, the authors are now planning to use a system which analyses communications with the aid of wearable recording devices as seen in Figure 11.



Figure 11. Wearable recording device

Correlations between learners' physical operations and communications should help us to clarify the above mentioned states.

References

- [1] J. Piaget, The Role of Action in the Development of Thinking, In: Overton W.F., Gallagher J.M. (eds), Knowledge and Development, pp.17–43, 1977
- [2] S. Papert, Teaching children to be mathematicians versus teaching about mathematics, International Journal of Mathematics Education in Science and Technology 3, pp.249–262, 1972
- [3] M. von Gagern, U. Kortenkamp, J. Richter-Gebert, M. Strobel: CindyJS – mathematical visualization on modern devices, Lecture Notes in Computer Science 9725, pp.319–326, 2016
- [4] M. von Gagern, J. Richter-Gebert: CindyJS plugins – extending the mathematical visualization framework, Lecture Notes in Computer Science 9725, pp.327–334, 2016
- [5] A. Montag, J. Richter-Gebert: CindyGL – authoring GPU-based interactive mathematical content, Lecture Notes in Computer Science 9725, pp.359–366, 2016
- [6] M. Kaneko: Using tangible contents generated by CindyJS and its influence on mathematical cognition, Lecture Notes in Computer Science 10407, pp.199–215, 2017
- [7] T. Noda, M. Kaneko: Collaborative use of mathematical content generated by CindyJS on tablets, Lecture Notes in Computer Science 10931, pp.379–388, 2018