Journal of Experimental Child Psychology 178 (2019) 30–40

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

On the preference for linear regression models in children: Results of a field study in elementary school children

Nadine Maussner [a], Martin J. Tomasik [a], Reinhard Schuster [b], Thomas Ostermann [a,*]

[a] Department of Psychology and Psychotherapy, Witten/Herdecke University, 58448 Witten, Germany
[b] Institute of Mathematics, University of Lübeck, 23562 Lübeck, Germany

ARTICLE INFO

Article history:
Received 15 November 2017
Revised 8 August 2018

Keywords:
Linear regression
Children
Performance
Preference
Horizontal–vertical minimization
Line of best fit

ABSTRACT

We aimed at showing that children are able to identify and draw a line of best fit through a cloud of points according to the concept of linear regression. A sample of 72 children in third and fourth primary school grades were given two different tasks. In a first task, 10 different sheets with three illustrations of the same scatterplot with different lines of regression to choose from were presented. In a second task, scatterplots were given and children were asked to draw a line of best fit with pencil and ruler. In the first task bisector regression was preferred (χ^2 = 15.21, *df* = 4, *p* = .004), whereas in the second conventional regression ordinary least squares ($X|Y$) was favored (χ^2 = 27.14, *df* = 4, *p* < .001). In summary, the preference of the traditional regression model using a minimization in the vertical dimension was only partly supported by our data.

Introduction

Covariation between variables is a core concept in the social and physical sciences and is fundamental for understanding more advanced concepts such as moderation and causal inference (see Pearl, 2009). Against the backdrop of Piaget's (1983) theory of cognitive development, understanding

* Corresponding author.
E-mail address: thomas.ostermann@uni-wh.de (T. Ostermann).

https://doi.org/10.1016/j.jecp.2018.09.009
0022-0965/

covariation between two variables requires an understanding of correspondence (i.e., to confirm identity), classification (i.e., to identify many items as one class), seriation (i.e., to order a series), and conservation (i.e., to be able to focus on different aspects or dimensions) (Moritz, 2004). Reasoning about covariation can thereby follow two approaches, namely (a) a pointwise approach in which bivariate pairs are compared in order to identify rules on how to translate one dimension to the other and (b) a variational approach in which change on one dimension is considered to infer change on the other (Nemirovsky, 1996). Moritz (2004) suggested that "these approaches are similar to two Piagetian schema which Wavering (1989) suggested to be developed in reasoning to create bivariate graphs: (a) one-to-one correspondence of bivariate data values and (b) seriation of values of a variable, necessary for scaling of graphs to produce a coordinate system" (p. 232). And indeed, most of the research into children's understanding of covariation has employed graphical representations (e.g., Coulombe & Berenson, 2001; Krabbendam, 1982; Leinhardt, Zaslavsky, & Stein, 1990; Moritz, 2000, 2004).

There are various approaches to depict the association between two variables such as contingency tables, scatterplots, and regression lines. Univariate linear regression is one of the oldest and most popular techniques to show the functional interdependency between two measurable variables, resulting in a fitting line (Sayago, Boccio, & Asuero, 2004). Although there seems to be evidence that this technique can be traced to the ancient Greek mathematicians, it appears that its popularity increased during the 18th century when it was applied to geographical and astronomical problems of distance measurements (Howarth, 2001). Legendre (1805) was the first to publish the principle of modern linear regression by introducing the concept of minimizing the sum of squares of the errors. In life science and later on in psychology, this technique of regression became popular by the work of Galton (1886) on the regression toward mediocrity in the hereditary stature in which this technique was applied. And even today this original method of minimizing the sum of squares of the vertical distances of points to a linear function $y = \alpha + \beta x$, also known as the method of ordinary least squares [OLS ($X|Y$)], still attracts researchers due to its mathematical ease and the interpretability of its results. In mathematical terms, this is denoted by

$$S(\alpha,\beta) = \sum_{i=1}^{n} u^2 = \sum_{i=1}^{n} [Y_i - (\alpha + \beta X_i)]^2 = \min.$$

However, it is hardly ever recognized that there is more than one possible metric to minimize the distance between the data and the linear function (Ludbrook, 2010). According to Isobe, Feigelson, Akritas, and Babu (1990), there are at least four other equally well justifiable metrics apart from (a) OLS($X|Y$):

(b) OLS($Y|X$) or ordinary least squares (with X as the dependent variable), where the regression line is defined by minimizing the square sum in the horizontal direction;
(c) OLS(bisector), which halves OLS($Y|X$) and the inverse of OLS($X|Y$);
(d) orthogonal regression (OR), which calculates the regression function using the perpendicular as a distance measure; and
(e) reduced major axis (RMA), which calculates the geometric mean of the slopes of OLS($X|Y$) and OLS($Y|X$).

Fig. 1 gives a geometric interpretation of four of these measures (method "e" does not have an intuitive geometric interpretation). Applying them leads to different lines with different slopes β_i and intercepts α_i. And although OLS($X|Y$) is the commonly used method, it is not known whether it is also the most intuitive one. Besides the different methods of regression, there might be some aspects that influence the intuitive decision to choose one of these models. A very famous effect in psychology of perception is the horizontal–vertical effect, which states that for humans the horizontal distance is less important than the vertical one. Tversky (1997) suggested that this might be explained by a "natural preference" of the vertical dimension caused by daily life experiences. Humans, animals, and plants grow vertically, and there is also a dominance of columns over rows. For instance, people tend to make a vertical shopping list rather than a horizontal one. This preference was also reflected in a