Critical evaluation of internet resources for teaching trend and variability in bivariate data

Pat Forster
Edith Cowan University
<p.forster@ecu.edu.au>

Introduction

A search on the Internet for resources for teaching statistics yields multiple sites with data sets, projects, worksheets, applets, and software. Often these are made available without information on how they might benefit learning. This paper addresses potential benefit from resources that target trend and variability relationships in bivariate data.

The paper is in five parts. The first is this introduction. In the second, trend and variability are defined. In the third, I quote research on the characteristics of data that influence students' discernment of trend and variability, and identify data sets that have particular characteristics and that can be freely downloaded from the web. The fourth part of the paper is a short review of free text-based resources. The final part is a review of spreadsheet programs and Java applets. The paper is relevant to teaching and learning in secondary schools.

Trend and variability

Trend encompasses the notions of covariation between a succession of data points, the possibility of unrealised data between the extremes of available data (interpolation), and the possibility of unrealised data beyond the extrema (extrapolation) (Ainley, Pratt & Nardi, 2001). Variability is the way data vary (Reading & Shaughnessy, 2004) and in this paper I use the term to mean variability which is manifested as deviation from a trend.

Issues for teaching are that students need to learn to notice and acknowledge trend and variability and, as well, represent them and describe them, measure and model them (for prediction and other purposes), explain them, and develop investigative strategies in relation to them (Reading & Shaughnessy, 2004; Wild & Pfannkuch, 1999). The focus in traditional curricula has been on trend but current recommendations are that variability should receive as much, if not more, attention (Cobb, McClain &

Gravemeijer, 2003; Reading & Shaughnessy, 2004). Analysis cannot be considered to be statistical unless variability is recognised.

Data sets

Contexts of data

It is widely accepted that working with real data benefits learning, however, students' recognition of trend relationships can hinge on the students being familiar with the contexts of the data (Cobb et al., 2003). Familiarity can be encouraged through class discussion, but certain situations are likely to be familiar to all students, most notably situations involving change over time: students generally possess intuitive ideas about change in phenomena (such as temperature) over time and these ideas support interpretation (Leinhardt, Zaslavsky & Stein, 1990).

In addition, interpretation is simplified if the dependent variable is measured vertically (Leinhardt et al., 1990). Examples are the height of students and plant height. In these cases, the heights of points above the horizontal axis on a scatter plot resemble the physical heights being represented. So, literal interpretation of the graph (e.g., that height increases as age increases) will be consistent with common-sense reasoning.

By way of contrast, literal interpretation of a graph is sometimes at odds with reality (Leinhardt et al., 1990). An example is a plot of swimming records over time. Swimming records are measured in seconds, minutes etcetera and present the contradiction that improvement, which is a positive outcome, corresponds with a negative trend. Analysis of such data might best be delayed until students gain confidence in discerning and describing simpler trend relationships.

Explaining variability (i.e., deviation of data from a trend) calls for knowledge of uncontrolled variables (Wild & Pfannkuch, 1999). These include environmental variables that are operational in the situation from which data are derived, and measurement and recording variables. For example, if corn is grown from seed and height data are generated each day of the growing cycle, then the heights recorded on any day could differ because of (a) differences in growing conditions (soil quality, and the amount of water, sunlight, fertiliser that individual plants receive), (b) greater or lesser accuracy in measurement, and (c) recording errors. Nicholson (1997) found that the provision of multivariate data from which students extract and analyse bivariate data can help them realise that uncontrolled variables are always present and act in combination so that data generally exhibit variability (i.e., are not perfectly correlated).

In summary, key considerations in selecting data sets from the internet for instruction are students' likely familiarity with the contexts of data or whether familiarisation can reasonably be brought about by discussion. Data with time as the independent variable and/or height as the dependent variable could be used initially, followed by other data for which interpretation is less intuitive.

Geometric configuration of graphed data

Trend is easily discerned if data are such that there is one value of the dependent variable for each value of the independent variable and covariance between adjacent points is always positive or always negative (Ainley et al., 2001). Outliers and local irregularity, where covariance between adjacent points changes direction (see Figure 1), can distract students from perceiving a trend — rather they focus on variability (Ben-Zvi & Arcavi, 2001). See Figure 1 in which the Olympic records over time exhibit positive covariation except between 1912 and 1920, 1968 and 1972, 1976 and 1980, and 1988 and 1992. As well, if data are horizontal or nearly horizontal then students may deny there is a trend (Noss, Pozzi & Hoyles, 1999): trend is more accessible if the rise (or fall) in data is steep.

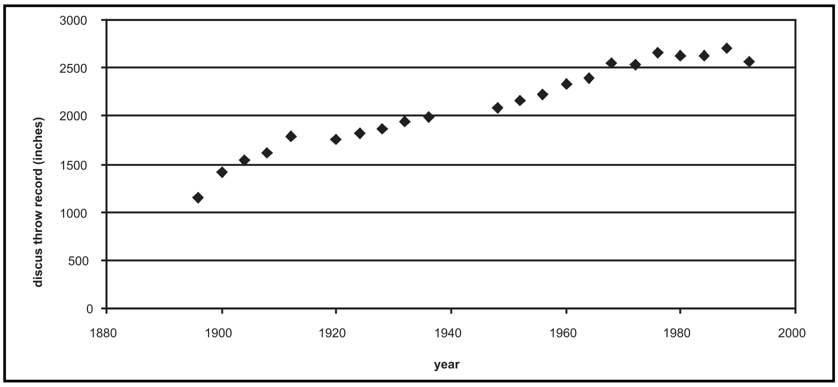

Figure 1. Olympic gold medal discus-throw 1896–1992 distances (Olympic Gold data, Exploring Data database, exploringdata.cqu.edu.au/datasets.htm).

Albeit, the appearance of data (rise and fall, distinctness of outliers, etc.) partly depends on the scales and dimensions of the graph, and manipulating these is feasible when technologies are used for the graphing. One option is to change the vertical scale. See Figure 2 with the scale chosen to emphasise trend, and Figure 3 with the same data and the scale chosen to emphasise variability. Otherwise, a graph can be reduced or enlarged by dragging its border, to respectively emphasise trend and variability. As well, when technologies are used for instructional purposes, outliers are easily removed, and consequently trend might be highlighted.

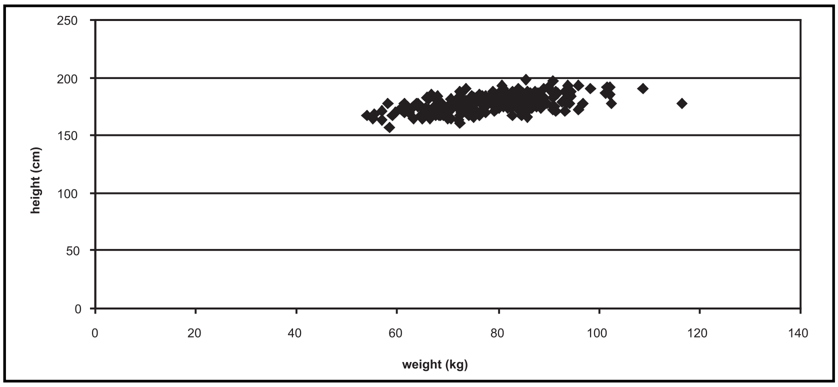

Figure 2. Height against weight for 260 women (body.dat, Journal of Statistics Education data archive, www.amstat.org/publications/jse/jse_data_archive.html).

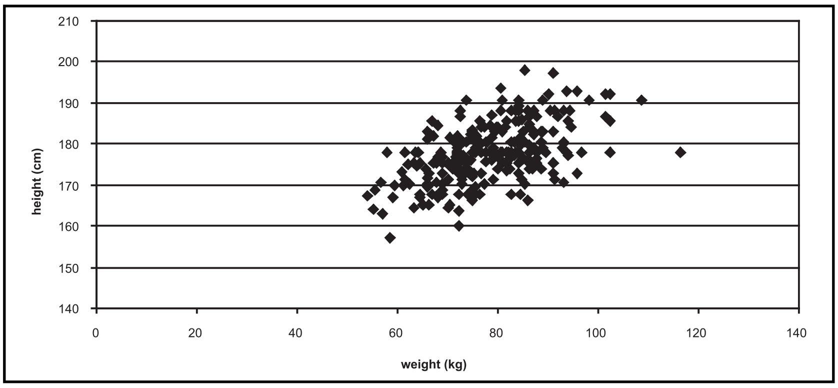

Figure 3. The graph in Figure 2 rescaled to emphasise variability.

Bivariate data that form stacks (see Figure 4) have been found to invite discernment of variability and trend (Cobb et al., 2003; Forster, 2006). Stacks occur if data have a one-to-many correspondence, which arises if the independent variable is discrete (Figure 4) or can be brought about by rounding (see Figure 5 which shows the data from Figure 2 after weights were rounded to the nearest 5 kg). Successive rounding can reduce the number of stacks and result in slightly different regression models, and exploring this could highlight for students that rounding decisions affect analysis and the conclusions reached. Furthermore, stacks can be viewed as distributions of univariate data, and trend lines can be treated as lines through the central

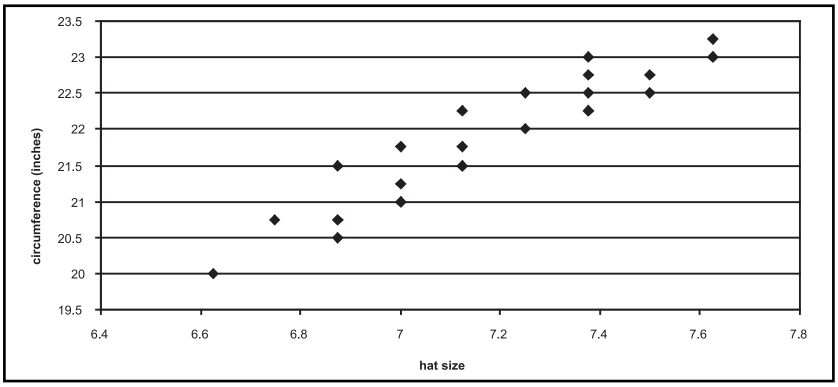

Figure 4. Hat circumference against hat size (a discrete variable) (hats.dat, Journal of Statistics Education, http://www.amstat.org/publications/jse/jse_data_archive.html).

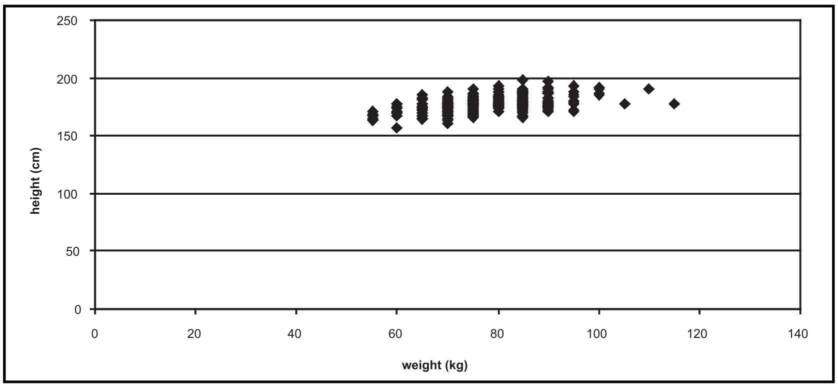

Figure 5. Height against weight (a continuous variable) with weights rounded to the nearest 5kg.

(mean or median) values. Regression analysis was founded on these ideas (Stanton, 2001) and they were used to guide instruction in the studies reported by Cobb et al. (2003) and Forster (2006). In both studies, students devised credible methods for locating trend lines when asked to analyse stacked data.

In summary, the geometric configuration of data warrants attention because it can favour discernment of trend or variability, or can invite discernment of both. Data with a one-to-one (x,y) correspondence and consistently positive or negative covariation would suit initial consideration of trend. Stacked data highlights vertical variability and trend relationships, and can support the definition of mathematical trend models.

Data sets with the various geometric configurations and a linear trend, or no trend, are listed in Table 1, together with their contexts and the internet databases from which they can be retrieved. The main criterion for selecting the sets was the contexts would probably be familiar to secondary school students. It is noted that most of the data files are in .xls format or can be imported into Excel by selecting the toolbar commands “Data” and “Import external data”. Data sets in contexts that would probably be familiar to secondary students and that exhibit non-linear trend patterns are listed in Table 2. Some sets lend themselves to testing multiple models (see column headed “trend”).

Anscombe’s (1973) four sets of decontextualised data also deserve mention as being useful for teaching. They each comprise eleven points and produce the same regression line and correlation coefficient but display different trend patterns (linear, quadratic, etc.). The sets can be used to highlight that trend models must be checked visually. They are available on the Exploring Data (ED) database <exploringdata.cqu.edu.au/datasets.htm>.

Checking the viability of the intercepts in relation to the contexts of data is another important aspect of choosing trend models. The diamond ring data set (diamond.dat, see Table 2) can be used to highlight the need for this. The intercepts for the linear model are incongruous in relation to the context, whereas the intercepts for quadratic and exponential models are reasonable. In fact, checking fit visually and checking the intercepts are two domains where students fall down: they tend to choose linear models over others and are satisfied that a line is appropriate if the correlation coefficient for data is high (Chu, 1996; Lingefjärd, 2002).

Text resources

Text resources for teaching and learning trend and variability include descriptions of the contexts and collection procedures for data. For example, detailed information is available for data sets from the *Journal of Statistics Education* (JSE) archive <www.amstat.org/publications/jse/jse_data_archive.html>. Briefer descriptions are provided on the ED database <exploringdata.cqu.edu.au/datasets.htm>. Such information can be important for valid interpretation of data.

Table 1. Data sets for exploring trend and variability in bivariate data: linear trend or no trend.

Data set	Source	Description	Variables			Structure		References
			time	height	other	distribution	trend	
Olympic Gold	[a]ED	Olympic high jump gold performance 1896-1992	✓	✓	-	[h]1-to-1	linear	
Olympic Gold	[a]ED	Olympic long jump and discus throw gold performance 1896-1992	✓	-	✓	1-to-1	linear	
Olympic Games 1984-2000	[b]ASD	Olympic track and field results 1984-2000	✓	✓	✓	1-to-1	linear	
Results by event	[c]BBC	Olympic track and field results 2004	✓	✓	✓	1-to-1	linear	
tdf.dat	[d]JSE	Tour de France data (multiple variables)						
		-year of race, age of winner	✓	-	✓	1-to-1	none	
		-year of race, average speed	✓	-	✓	1-to-1	linear	
Water resources	[e]QELP	DataSet#011, Columbia River velocity, depth	-	-	✓	1-to-1	linear	
hats.dat	[d]JSE	Hat size, circumference of hats	-	-	✓	stacked	linear	
Students' heights and weights	[f]PC	Data for 33 students (multiple variables)						
		-student's heights, mother's height	-	✓	-	stacked	linear	
Sweet pea data	[g]Stanton (2001)	Diameter of mother sweet-pea seed, diameter of daughter seed	-	-	✓	stacked	linear	Stanton (2001)
fruitfly.dat	[d]JSE	Data on fruitfly (multiple variables)						Hanley & Shapiro (1994)
		-number of companions, lifespan	-	-	✓	stacked	linear	Nicholson (1997)
cigarettes.dat	[d]JSE	Data for cigarettes by brand (multiple variables)						
		-tar content, nicotine content	-	-	✓	scattered	linear	

[a] ED: Exploring Data database <http://exploringdata.cqu.edu.au/datasets.htm>
[b] ASD: Athletics Statistics Database <http://www.sci.fi/~mapyy/tilastot.html>
[c] BBC: British Broadcasting Corporation sport website <http://news.bbc.co.uk/sport2/hi/olympics_2004/athletics/results/default.stm>
[d] JSE: Journal of Statistics Education data archive <http://www.amstat.org/publications/jse/jse_data_archive.html>
[e] QELP: Quantitative Environmental Learning Project database <http://www.seattlecentral.org/qelp/Data.html>
[f] PC: Pomona College database <http://www.economics.pomona.edu/StatSite/eclectic.html>
[g] Stanton (2001) <http://www.amstat.org/publications/jse/>
[h] 1-to-1 is to be taken to mean that for each x value there is a single value of y, as in Figure 1.

Dataset	Source	Description	Variables: time	Variables: height	Variables: other	Structure: distribution	Structure: trend	References
Forests	[a]QELP	DataSet#056, tree ring widths over time	✓	-	✓	[a] 1-to-1	exponential	
Water resources	[a]QELP	DataSet #071, discharge from flooded river over time	✓	-	✓	1-to-1	exponential	
Oil production.	[b]ED	World's oil production in barrels 1880-1988	✓	-	✓	1-to-1	exponential	
Used car prices	[c]Ound	Used car data (multiple variables)						
		-age of car (years), resale price	✓	-	✓	stacked	exponential	
diamond.dat	[d]JSE	Carat weight of diamonds, price of diamond rings		-	✓	stacked	multiple	Chu (1996, 2001)
fat.dat	[d]JSE	Body measurements						
		-age, body fat percentage	✓	-	✓	stacked	multiple	Johnson (1996)
		-weight, body fat percentage	-	-	✓	scattered	multiple	
body.dat	[d]JSE	Body measurements (multiple variables)						
		-age, height	✓	✓	-	stacked	multiple	Heinz, Johnson &
		-weight, height	-	✓	✓	scattered	multiple	Shoemaker (2003)
Gestation period	[e]CPA	Gestation period, average longevity for 43 animals	-	-	✓	scattered	multiple	
Global temp 1	[b]ED	Global mean temperature 1866-1996.	✓	-	✓	scattered	multiple	
normtemp.dat	[d]JSE	Temperature, gender and heart rate data (n=130)						
		-body temperature, heart rate	-	-	✓	stacked	multiple	
		-heart rate, body temperature	-	-	✓	scattered	multiple	Shoemaker (1996)
poverty.dat	[d]JSE	Population and economic indicators for 97 countries (multiple variables)						Rouncefield (1995)
		-gross national product, life expectancy for males	-	-	✓	scattered	polynomial	
Climate	[f]ABM	Daily, weekly, monthly temperature, rainfall etc						Weather
		-average monthly rainfall for Perth	✓	-	✓	1-to-1	sinusoidal	Predictions ([i]PC)
Astronomical information	[g]GAD	Sunrise/sunset, moonrise/moon set, planet rise and set times in each Australian State	✓	-	✓	1-to-1	sinusoidal	

[e] QELP: Quantitative Environmental Learning Project database <http://www.seattlecentral.org/qelp/Data.html>
[b] ED: Exploring Data database <http://exploringdata.cqu.edu.au/datasets.htm>
[c] Ound: Oundle School website, Probability/Statistics/Datasets <http://www.tsm-resources.com/mlink.html#stats>
[d] JSE: Journal of Statistics Education data archive <http://www.amstat.org/publications/jse/jse_data_archive.html>
[e] CPA: Contemporary Precalculus through Applications data base <http://courses.ncssm.edu/math/CPTA/data/ >
[f] ABM: Australian Bureau of Meteorology <http://www.bom.gov.au/climate/how/ >
[g] GAD: Geoscience Australia Database <http://www.ga.gov.au/geodesy/astro/index.jsp>
[h] 1-to-1 is to be taken to mean that for each x value there is a single value of y, as in Figure 1.
[i] PC: Pomona College database <http://www.economics.pomona.edu/StatSite/eclectic.html>

Table 2. Data sets for exploring trend and variability in bivariate data: non-linear and linear trend.

Other text resources include articles in online journals such as:

- Teaching Statistics <www.rsscse.org.uk/ts> and the
- Journal of Statistics Education <www.amstat.org/publications/jse>.

They provide a diverse range of ideas for teaching, together with reasons why particular approaches might benefit learning, and critique on the implementation of the approaches in actual classes. For example, Hunt (1994) presents a project where students collected their own data. The project called for sampling, measurement, formulation of a trend model, and explanation of variability in order to explain intercepts of the model. Sowey (2001) lists applications of statistics and statistical propositions that, in his experience, surprise students and motivate learning. Weldon (2005) discusses a variety of strategies for the graphical analysis of bivariate data. Rouncefield (1995), Hanley and Shapiro (1994), and Chu (1996, 2001) report on classroom use of data sets from the JSE data archive. Other articles are listed in Tables 1 and 2 next to the data sets to which they relate.

Other types of text resource include:

- suggestions for collaborative projects on the Maths Forum website <mathforum.org/workshops/sum96/data.collections/datalibrary>
- case studies (e.g., "Physical Strength" and "Job Performance"), Rice Virtual Lab in Statistics <www.ruf.rice.edu/~lane/rvls.html>
- free-response examination questions with solutions, Advanced Placement Statistics, College Board <www.collegeboard.com/testing>.

The case studies and examination solutions demonstrate that multiple approaches can be used for analysing any data set.

Spreadsheets

Spreadsheets have been found to benefit the teaching and learning of trend and variability in several ways. Firstly, scatter plots can be produced quickly on spreadsheets and the plots can be projected onto a whiteboard for the purpose of class discussion. Secondly, if students are using a spreadsheet for analysis, they can graph their data and focus on interpreting the graph, without the distraction of having to plot individual points.

Thirdly, interactive capabilities of spreadsheets can be used to highlight trend and/or variability relationships (Ainley et al., 2001; Ben-Zvi & Arcavi, 2001; Forster, 1997). Options include adding or deleting data from a table and observing changes in the graph; dragging the graph space in order to enlarge it, and changing the vertical scale on the graph so that data appear more or less dispersed (see Figures 3 and 4); calculating summary statistics and observing how they change when data change; superimposing estimated trend lines on the scatter plot using a line-drawing tool; and superimposing least squares regression models on the scatter plot. Commercial spreadsheets such as Excel offer all these capabilities. Free spreadsheet programs on the web usually offer some of them. Two examples are:

- the Winstats Peanut Software program, Parris (2005) <math.exeter.edu/rparris>

- the Calico Statistical Spreadsheet program from the Mathematics Department, California State University <www.math.csusb.edu/faculty/stanton/calico>

Customised spreadsheets and Java applets

Customised spreadsheets and Java applets that target particular aspects of trend and variability are also available for free downloading. Topics include the least squares regression principle, correlation, and random variation in samples. A search on each topic produced the following.

Least squares regression principle

Items on the least squares regression principle include the:

- Matti Associates' (2003) Scatterplot applet <matti.usu.edu/nlvm/nav/category_g_4_t_5.html>
- National Council of Teachers' (NCTM, 2000) Least squares regression applet <standards.nctm.org/document/eexamples/chap7/7.4>
- West's (n.d.) Regression applet <www.stat.sc.edu/~west/javahtml/Regression.html>
- Finzer's (1995) Least squares applet <www.keypress.com/sketchpad/javasketchpad/gallery/index.php>
- Proquest Information and Learning Company's (2005) Least squares activities A and B <www.explorelearning.com/index.cfm?method=cResource.dspResourcesForCourse&CourseID=266>
- Hunt, Tyrrell, and Nicholson's (2001) spreadsheet, What is least squares? <www.mis.coventry.ac.uk/~nhunt/regress/listof.htm>

The properties of these applications are summarised in Table 3. In particular, the Matti and NCTM applets have been found valuable for teaching (Forster, 2006): use of the Matti applet can stimulate students in thinking about how to calculate a regression line and the NCTM applet illustrates the least squares relationship numerically and graphically. The applet by West could also support conjecture on calculating regression lines and Finzer's applet illustrates the least squares relationship, but both have fewer interactive capabilities and lack detail that is available on the Matti and NCTM applets (see Table 3). The Proquest applets offer a range of options for manipulating data and accessing summary statistics (see Table 3), but free use is limited to 5 minutes. The spreadsheet by Hunt et al. lacks the dynamic affordances of Java applets (e.g., ability to drag points on the graph) but includes an interactive worksheet that guides the user in locating regression lines through data that the user enters.

Table 4 summarises other applets and customised spreadsheets that address the development of skills to do with fitting trend models and predicting from them, and still others that target understanding to do with trend models. Potentially, the applications could benefit teaching and learning, but reports on realised benefit from them do not seem to be available.

Table 3. Java applets and customised spreadsheets on the least squares regression principle.

	Author or website					
Properties and interactive capabilities	Matti	NCTM	West	Finzer	Proqu.	Hunt
Graph	-	✓	✓	✓	✓	✓
Graph and table	✓	-	-	-	-	✓
Co-ordinate system						
-change scales	✓	-	-	-	-	-
-drag graph space	-	-	-	-	-	-
Data						
-add points to a graph by clicking	✓	-	✓	-	-	-
-drag points on a graph	-	✓	-	✓	✓	-
-enter data in a table	✓	-	-	-	-	✓
Estimate trend-line on scatter plot						
-drag existing line	-	✓	-	✓	✓	-
-enter *m* and *c* values, line appears	-	-	-	-	-	✓
-select *m* and *c* with sliders, line appears	-	-	-	-	✓	-
Graphics on scatter plot						
-regression line	✓	-	✓	✓	✓	
-mean point on regression line	✓	-	-	-	-	-
-residuals to estimated trend line	-	✓	-	-	-	-
-squares on the residuals to estimated trend line	-	✓	-	✓	✓	-
-shortest distances to estimated trend line	-	✓	-	-	-	-
Summary values near plot						
-equation of estimated trend line	-	✓	-	-	✓	-
-equation of regression line	✓	-	✓	-	✓	-
-mean point	✓	-	-	-	-	-
-correlation coefficient	✓	-	✓	-	-	-
-predictions from trend model	-	-	-	-	-	✓
-residuals	-	-	-	-	-	✓
-squares on individual residuals	-	✓	-	✓	✓	✓
-sum of the squares-on-the-residuals	-	✓	-	-	-	✓

Table 4. Applets and spreadsheets that target skills and understanding to do with trend models

Applet/spreadsheet title	Author / website	Topic
Skill development		
Regression by eye	[a]Rice	Drawing trend lines by eye
Trendlines in Excel	[b]Hunt	Producing trend models and r values in Excel
Solving using trend lines	[c]Proquest	Predicting from a trend model.
Understanding		
Polynomial regression	[d]Vestac	Best-fit polynomial models
Investigate transformations	[b]Hunt	Effects of transforming data on the scatter plot, regression line and correlation coefficient
Experiment with other criteria	[b]Hunt	Trend lines that fit various criteria.
Linear regression	[e]Stanton	Effects of additional data on residuals and summary statistics.
Residual plots	[b]Hunt	Residuals for data with seasonal and other patterns.

[a] Rice Virtual Lab in Statistics <http://www.ruf.rice.edu/~lane/stat_sim/index.html>
[b] Hunt et al. (2001) <http://www.mis.coventry.ac.uk/~nhunt/regress/listof.htm>
[c] Proquest Information and Learning Company (2005) <http://www.explorelearning.com/index.cfm?method=cResource.dspResourcesForCourse&CourseID=266>
[d] Vestac Java applets <http://www.kuleuven.ac.be/ucs/java/index.htm>
[e] Stanton (n.d.) <http://www.math.csusb.edu/faculty/stanton/probstat/index.html>

Correlation

The applets/spreadsheets on correlation focus on the relationship between the distribution of data on the scatterplot and values for Pearson's correlation coefficient r. They include:

- Correlation revisited (Neuwirth, 1996) <sunsite.univie.ac.at/Spreadsite/statexamp/>
- Correlation (Proquest Information and Learning Company, 2005) <www.explorelearning.com/index.cfm?method=cResource.dspResourcesForCourse&CourseID=266>
- Components of rand regression by eye (Rice Virtual Lab in Statistics) <www.ruf.rice.edu/~lane/stat_sim/>
- Guessing correlations and Investigating correlation (Hunt et al., 2001) <www.mis.coventry.ac.uk/~nhunt/regress/listof.htm>

"Correlation revisited" and "Correlation" allow the user to change r with a slider and observe the effect on a scatter plot of given data. The "Correlation" applet graphs the regression line as well as the data, and the naïve user could be misled into thinking r depended solely on the slope of the regression line.

"Components of r" allows the user to explore relationships between r, the spread in X, spread in Y, slope of the regression line, and explained and unexplained variation. This applet supports identification of more properties of r than do "Correlation revisited" and "Correlation". "Regression by eye", "Guessing correlation", and "Investigating correlation" allow the user to guess r values for given data or data that they enter, then, check the guesses.

Random variation

Applets and spreadsheets on random variation in samples from the same population are listed in Table 5. They illustrate variation between scatterplots for samples, and variation between regression lines, parameters of regression lines, and residuals (see Table 5). Furthermore, while variation between samples may not be treated formally in school statistics courses, demonstration with the applets could highlight for students that deterministic conclusions based on single samples, especially small ones, are inappropriate: rather they might appreciate that trend for a sample is an estimate of underlying trend (Nicholson, 1996).

Concluding comments

In this paper I have suggested data sets for introducing students to trend and variability relationships. The criteria that I used in choosing them were (a) they were derived from contexts that might be familiar to secondary-school students, and (b) they exhibit different geometric structures when graphed. The Australian Bureau of Statistics CensusAtSchool database <www.abs.gov.au> is another source of data in contexts that are familiar to students. Schools can register to contribute and retrieve data.

I have also identified Java applets and spreadsheets for possible use in teaching. Other examples are available on the websites that were referenced

Table 5. Applets and spreadsheets on random variation in samples from the same population.

Applet/spreadsheet title	Author/ website	Operation
Regression plot	[a]Vestac	Select parameters for regression line for population, sample size. View scatterplots and regression lines for multiple samples.
Histograms of slope and intercept	[a]Vestac	Same as above but view histograms for slope and *y* intercepts for trend lines for samples.
Exploring variation in predictions	[b]Hunt	Choose sample size. View scatter plot and regression line for population (n=50), and data and regression lines for samples superimposed on the plot.
Sampling distribution of r	[b]Hunt	Choose population correlation coefficient _ and sample size. View scatter plots and histogram of r for samples.
Normal probability plots	[b]Hunt	Choose sample size and distribution of errors for population. View distribution of residuals for samples.

[a] Vestac Java applets <http://www.kuleuven.ac.be/ucs/java/index.htm>
[b] Hunt et al. (2001) <http://www.mis.coventry.ac.uk/~nhunt/regress/listof.htm>

in the paper, but the applications that I chose indicate the scope of what is available. It also needs to be acknowledged that the applets and spreadsheets target technical understanding of trend and variability (the least squares principle, etc.) and technical understanding is insufficient for meaningful analysis of data. Amongst other things, students need to learn strategies to reveal trend and variability in bivariate data and, as well, learn to interpret regression models and variability in relation to the contexts of data.

Acknowledgments

This paper is part of an Australian Research Council funded project (DP0345843). Some of the ideas expressed in it arose from studies in Year 12 classes of Craig Davis, St Hilda's Anglican School for Girls, Perth and Romaine Saunders, Presbyterian Ladies' College, Perth.

References

Ainley, J., Pratt, D. & Nardi, E. (2001). Normalising: Children's activity to construct meanings for trend. *Educational Studies in Mathematics, 45*, 131–146.

Anscombe, F. J. (1973), Graphs in statistical analysis. *American Statistician, 27*, 17–21

Ben-Zvi, D. & Arcavi, A. (2001). Junior high school students' construction of global views of data and data representations. *Educational Studies in Mathematics, 45*(1), 35–65.

Chu, S. (1996). Diamond ring pricing using linear regression. *Journal of Statistics Education, 4*(3). Accessed at http://www.amstat.org/publications/jse

Chu, S. (2001). Pricing the C's of diamond stones. *Journal of Statistics Education, 9*(2), Accessed at http://www.amstat.org/publications/jse

Cobb, P., McClain, K. & Gravemeijer, K. (2003). Learning about statistical variation. *Cognition and Instruction, 21*(1), 1–78.

Forster, P. (1997). Teaching through technology. *Australian Senior Mathematics Journal, 11*(1), 54–61.

Forster, P. A. (2006). Technologies for teaching and learning trend. *International Journal for Mathematical Education in Science and Technology.*

Forster, P. A. (in press). Technologies for teaching and learning trend. *International Journal for Mathematical Education in Science and Technology.*

Hanley, J. A. & Shapiro, S. H. (1994). Sexual activity and the lifespan of male fruitflies: A data set that gets attention. *Journal of Statistics Education, 2*(1). Accessed at http://www.amstat.org/publications/jse/

Heinz, G., Peterson, L. J., Johnson, R. W. & Kerk, C. J. (2003). Exploring relationships in body dimensions. *Journal of Statistics Education, 11*(2). Accessed at http://www.amstat.org/publications/jse/

Hunt, N. (1994). A tale of six cities. *Teaching Statistics, 16*(1), 5–8. Accessed at http://www.rsscse.org.uk/ts/

Johnson, R. W. (1996). Fitting percentage of body fat to simple body measurements. *Journal of Statistics Education, 4*(1). Accessed at http://www.amstat.org/publications/jse/

Leinhardt, G., Zaslavsky, O. & Stein, M. K. (1990). Function, graphs, and graphing: Tasks, learning, and teaching. *Review of Educational Research, 60*, 1–64.

Lingefjärd, T. (2002). Mathematical modelling for preservice teachers: A problem from anaesthesiology. *International Journal of Computers for Mathematical Learning, 7*(2), 117–143.

Nicholson, J. (1996). Exploring sampling. *Teaching Statistics, 18*, 49–51. Accessed at http://www.rsscse.org.uk/ts/

Nicholson, J. (1997). Developing probabilistic and statistical reasoning at the secondary level through the use of data and technology. In J. Garfield & G. Burrill (Eds), *Research on the Role of Technology in Teaching and Learning Statistics* (pp. 29–44). Voorburg: International Statistics Institute. Accessed at http://www.stat.auckland.ac.nz/~iase/publications.php?show=8

Noss, R., Pozzi, S. & Hoyles, C. (1999). Touching epistemologies: Meanings of average and variation in nursing practice. *Educational Studies in Mathematics, 40*(1), 25–51.

Reading, C. & Shaughnessy, J. M. (2004). Reasoning about variation. In D. Ben-Zvi & J. Garfield (Eds), *The Challenge of Developing Statistical Literacy, Reasoning and Thinking* (pp. 210–226). Dordrecht: Kluwer Academic Publishers.

Rouncefield, M. (1995). The statistics of poverty and inequality. *Journal of Statistics Education, 3*(2). Accessed at http://www.amstat.org/publications/jse/

Shoemaker, A. L. (1996). What's normal? — Temperature, gender, and heart rate. *Journal of Statistics Education, 4*(2). Accessed at http://www.amstat.org/publications/jse/

Sowey, E. R. (2001). Striking demonstrations in teaching statistics. *Journal of Statistics Education, 9*(1). Accessed at http://www.amstat.org/publications/jse/

Stanton, J. M. (2001). Galton, Pearson, and the peas: A brief history of linear regression for statistics instructors. *Journal of Statistics Education, 9*(3). Accessed at http://www.amstat.org/publications/jse/

Weldon, K. L. (2005). From data to graphs to words — but where are the models? *Proceedings of the International Association for Statistical Education/International Statistical Institute Satellite conference: Statistics Education and Communication of Statistics.* Accessed at http://www.stat.auckland.ac.nz/~iase/publications.php?show=14

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