# 16

## STATISTICAL AND CORRELATIONAL TECHNIQUES

Stephen Gorard

## Introduction

This chapter presents a simple introduction to some of the uses of numbers in education research, illustrating a few of the many and varied research questions that can be addressed with numeric evidence. It is important to relaise that using numbers involves no kind of paradigmatic or epistemological assumptions – the supposed 'paradigms' of quantitative and qualitiative research are just red herrings. The chapter outlines some common sources of data and methods of analysis, before a relatively simple real-life example is presented. The bulk of the analysis was completed in less than two months and quickly led to several articles in high-prestige journals. Shorn of the schismic and other barriers that some commentators write about, apparently instead of doing research itself, doing research is really rather easy. And this is true even of work involving large-scale numeric datasets. The chapter ends by suggesting a few further examples of similar simple techniques.

## Statistical and correlational research

It is not possible to do justice to all of the approaches that might come under the heading of statistical and correlational research in a chapter of this brevity, especuially as the main focus is on a real-life case study of reserarch. There *is* a sub-set of statistical work that is based on random sampling theory and that is intended to help analysts estimate whether a result they have for a random sample is also true for the population from which that sample was drawn. It involves p-values, significance tests, confidence intervals and similar hard to comprehend ideas. That kind of work, while widespread, is not covered here for a number of reasons. The whole approach is unrealistic since true random samples are so rare and because it relies on a number of prior assumptions about measurement accuracy and a complete response rate that are even rarer in practice. The approach is also based on a fundamental error of confusing the probability of the data observed given a pre-specified hypothesis with the probability of that hypothesis being true given the data observed. The two values are very different, and one cannot be converted into the others without a third and unknown value (Gorard, 2010a). Finally, the approach is very limited in only being concerned with generalisability to a population. It does not help analysts decide the really important point, which is whether the result is substantively important (and this is what they have to do subsequently, whether they want to generalise that conclusion or not). In general, researchers working with numbers and reading the work of others can safely ignore p-values however portrayed - focusing instead on the number of cases, the problems caused by missing data, the quality of any measurements, and the fit between the research design and the research questions (Gorard 2013).

This chapter looks at the far more important second issue (about the substantive importance of results), an issue which is common to, and faced by, *all* analysts at some stage. The same kind of judgement about the importance and robustness of a numeric result is made when considering a non-numeric result (Gorard, 2010b). This logic of analysis is universal, and there are no paradigmatic differences in education research, any more than there are in real life. Take something everyday like the use-by date on food products in the UK. This is numeric information that requires no epistemological commitment or paradigmatic beliefs. Stores can use it to ensure produce is fresh and customers can use it to help decide whether to buy the product. Everyone will realise, if they think about it, that the date could be in error (mislabelled in printing for example), that nothing dramatic happens on that date (it is a limit stamped onto a continuous process of food becoming less fresh over time), that due to circumstances (like poor storage) the food may be beyond use before that date, and otherwise that much foodstuff is still safely edible the day after the use-by date. A customer using the date information to decide whether to buy or eat the product makes a subjective judgement. This, in summary, is what numeric analysis in social science also involves, and very similar steps can be used to describe textual and all other analyses. For example, a textual analyst knows that what they read could be a misprint, it could represent exactly what the writer intended to convey or it could be an attempt to mislead. In coding, they must make a subjective judgement about what the text portrays and hope that this does not mislead their own readers, and so on.

However, some different techniques of analysis *are* differentially suitable for certain kinds of research questions. Generally, researchers using numeric data want to know how strong their finding is, where that finding could be expressed as a difference, trend, or pattern. This estimate of the strength of a finding is usually computed as an ‘effect’ size. When considering a difference between two sets of measurements, a common approach would be to find the difference between the means of the two groups, and divide the result by the standard deviation of both groups combined. This ‘effect’ size is a standardized difference between means, and could be used to portray how much one group of learners was out-performing another group, or by how much one group of learners had improved over time, for example.

Correlational research, on the other hand, addresses questions about the relationship between two or more variables, and the extent to which they co-vary. The most commonly used technique for correlation/regression is based on the Pearson’s R correlation coefficient. An R score of 1 means perfect correlation or even identity, an R of –1 means perfect inverse, and an R of 0 means no correlation at all between the two variables. Usually you will uncover R values between these extremes. Squaring the R value, to give R-squared, yields a different kind of ‘effect’ size to that above. Here, the R-squared represents how much of the variation in one score is common to the variation in the other score. (See Chapter 43 on

‘Multiple Linear Regression’ in this volume or Gorard (2001) for further explanation of correlation coefficients and examples of using Pearson’s R.)

Correlation is the basis of many more advanced techniques for analysis, such as factor analysis, regression and structural equation modelling. In the example below, it is used to help see the possible common patterns in 12 trends over time.

## An example: correlational research

Previous international work has shown that clustering pupils with similar characteristics in particular schools yields no clear academic benefit and can be disadvantageous to pupils both socially and personally (Gorard and Smith, 2010). It needlessly increases divisions between rich and poor in education and outcomes (Goldhaber et al., 2015; Yeung and Nguyen-Hoang, 2016). Understanding how and why this clustering happens, and how it may be reduced, is therefore important for policy. Yet previous work has tended to focus on only one kind of clustering at a time. In the USA, for example, black– white segregation of pupils has been the key issue. In the UK, and across Europe, the focus has been on social background, especially on the clustering or segregation in specific schools of pupils living in poverty. In the UK, segregation between schools by poverty has been considered an outcome of the regional stratification of economic activity, housing prices and social housing policies, increased diversity of schools and the process of school place allocation (Gorard et al., 2003). There is also evidence that changes in the overall number of schools, and changes in the prevalence of poverty, are related to the precise level of local between-school segregation. In the limited sense that segregation other than by poverty (such as by ethnicity) has been considered in the UK, it has been assumed that the same kinds of reasons apply for all measures. So the assumption has been that segregation by ethnicity and by poverty have the same determinants. But is this true? Is there one process, perhaps involving a number of indicators of disadvantage, that clusters similar pupils together in schools however their similarity is measured? Or do these factors operate differently, or perhaps not operate at all, in separate processes of segregation depending on which pupil characteristics are considered?

The analysis presented here is based on figures from the Annual Schools Census (ASC) for all state-funded secondary schools in England (Gorard and See 2013). It used official school-level figures for the number of full-time equivalent pupils on roll in each school for January of each year, the number eligible for and taking free school meals (FSM) which is a measure of family poverty, and those with a declared additional or special educational need with or without a statement (SEN), in each minority ethnic group, and those speaking a first language other than English. These are all indicators of possible educational disadvantage. For each indicator, there are two estimates of how clustered each pupil characteristic is. These estimates are the Gorard segregation index (GS) and the dissimilarity index (D). The two indices are very similar, with a higher value (nearer one) representing a very segregated system, while both would be zero if all schools have their proportionate share of potentially disadvantaged pupils. There is no space here to explain the calculation of these indices in more detail (but see Gorard et al., 2003; or Cheng and Gorard, 2010). The six indicators of disadvantage each summarised with these two indices yield 12 distinct measures of pupil segregation between schools which have been tracked for 14 years from 1996 to 2009.

The trends in between-school segregation, in terms of pupil backgrounds from 1996 to 2009, show several different characteristics (Table 16.1). Both indices (GS and D) tend to give very similar results for each indicator. However, the levels of clustering between schools in terms of different indicators are very different. Segregation by poverty is about 0.3, meaning that around a third of pupils with free school meals would have to exchange schools for poverty to be distributed between schools in proportion to their size. Segregation by pupil special need is a little less than this but is of the same order of magnitude (around 0.28). Segregation by minority ethnic group (non-white) and for those not speaking English as a first language is around twice these values, however (0.6 or more). Another difference is that segregation by FSM increased from 1996 to 2005/6 and subsequently dropped a little. All other indicators, on the other hand, have shown an annual decline in segregation.

**Table 16.1** Segregation 1996-2009, all indicators, secondary schools in England

| | *1996* | *1997* | *1998* | *1999* | *2000* | *2001* | *2002* | *2003* | *2004* | *2005* | *2006* | *2007* | *2008* | *2009* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| FSM takeup D | 0.35 | 0.34 | 0.35 | 0.36 | 0.37 | 0.37 | 0.39 | 0.38 | 0.38 | 0.38 | 0.39 | 0.39 | 0.38 | 0.38 |
| FSM takeup GS | 0.30 | 0.30 | 0.31 | 0.32 | 0.33 | 0.33 | 0.34 | 0.34 | 0.34 | 0.34 | 0.35 | 0.35 | 0.35 | 0.34 |
| FSM eligible D | 0.38 | 0.38 | 0.38 | 0.39 | 0.39 | 0.39 | 0.39 | 0.39 | 0.39 | 0.40 | 0.39 | 0.39 | 0.39 | 0.39 |
| FSM eligible GS | 0.31 | 0.31 | 0.32 | 0.32 | 0.33 | 0.33 | 0.33 | 0.34 | 0.34 | 0.34 | 0.34 | 0.34 | 0.34 | 0.33 |
| SEN statement D | 0.30 | – | 0.28 | 0.28 | 0.27 | 0.27 | 0.27 | 0.26 | 0.25 | 0.25 | 0.24 | 0.24 | 0.25 | 0.25 |
| SEN statement GS | 0.29 | – | 0.28 | 0.27 | 0.27 | 0.26 | 0.26 | 0.25 | 0.25 | 0.24 | 0.24 | 0.24 | 0.24 | 0.24 |
| SEN no statement D | 0.32 | – | 0.28 | 0.26 | 0.25 | 0.25 | 0.27 | 0.28 | 0.27 | 0.26 | 0.26 | 0.26 | 0.26 | 0.26 |
| SEN no statement GS | 0.27 | – | 0.24 | 0.22 | 0.21 | 0.21 | 0.22 | 0.24 | 0.24 | 0.23 | 0.22 | 0.22 | 0.22 | 0.21 |
| Non-white D | – | 0.68 | 0.67 | 0.65 | 0.65 | 0.64 | 0.65 | 0.59 | 0.57 | 0.55 | 0.54 | 0.54 | 0.54 | 0.55 |
| Non-white GS | – | 0.60 | 0.60 | 0.56 | 0.56 | 0.55 | 0.54 | 0.48 | 0.46 | 0.45 | 0.44 | 0.43 | 0.43 | 0.43 |
| ESL D | – | – | – | – | 0.70 | 0.70 | – | 0.66 | 0.66 | 0.66 | 0.64 | 0.63 | 0.63 | 0.63 |
| ESL GS | – | – | – | – | 0.65 | 0.64 | – | 0.61 | 0.60 | 0.60 | 0.59 | 0.56 | 0.56 | 0.55 |

*Notes*:Figures are presented to only two decimal places for ease of reading. The DfE figures for SEN in 1997 are only half those of 1996 and 1998, yielding much higher levels of segregation. They cannot be correct, and so we exclude them from our analysis. DfE can provide no figures for first language in 2002. Ethnicity was collected from 1997 onwards, and language from 2000 onwards.

So, perhaps there are three kinds of segregation going on here. The first is for FSM which shows a different level to ethnicity and language and a different trajectory of change over time to SEN. The second is SEN which shows a very different level of segregation to the third group of ethnicity and language but a similar trajectory over time. Why has segregation by poverty risen while segregation by ethnicity and language has fallen? Many of the kinds of factors that might affect segregation by poverty, including increased diversity in types of school or school closures, would surely also influence segregation in terms of other pupil characteristics. So are there genuinely different patterns of clustering in schools depending upon the kinds of pupil background measures used, with different determinants? One way of investigating this further is to calculate the *correlation* between the changes over time in each measure.

Table 16.2 shows the correlation coefficients for all 12 national measures of segregation over time (as in Table 16.1). A correlation of 1 means that the two variables are, in effect, measuring the same thing (like Centigrade and Fahrenheit for temperature). Of course, each variable has a correlation of 1 with itself (the diagonal). We also learn from Table 16.2 that, for most practical purposes, the two indices of D and GS serve the same purpose. Whatever their theoretical differences, their values for each of the six indicators correlate very highly. Indeed, the correlation between D for segregation by free school meal take-up and GS for the same indicator is 1 (top left of table). The correlation between D for segregation by English as a second language and GS for the same indicator is + 0.99 (bottom right of table).

**Table 16.2** Correlations between trends in all 12 measures of segregation, secondary schools in England

| | **FSM takeup D** | **FSM takeup GS** | **FSM eligible D** | **FSM eligible GS** | **SEN statement D** | **SEN statement GS** | **SEN no statement D** | **SEN no statement GS** | **Non-white D** | **Non-white GS** | **ESL D** | **ESL GS** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| FSM takeup D | 1 | 1 | 0.79 | 0.96 | –0.93 | –0.93 | –0.54 | –0.47 | –0.82 | –0.87 | –0.83 | –0.79 |
| FSM takeup GS | 1 | 1 | 0.79 | 0.97 | –0.95 | –0.95 | –0.58 | –0.51 | –0.86 | –0.89 | –0.86 | –0.82 |
| FSM eligible D | 0.79 | 0.79 | 1 | 0.89 | –0.7 | –0.72 | –0.62 | –0.47 | –0.54 | –0.54 | 0.29 | 0.36 |
| FSM eligible GS | 0.96 | 0.97 | 0.89 | 1 | –0.96 | –0.97 | –0.65 | –0.55 | –0.87 | –0.89 | –0.66 | –0.61 |
| SEN statement D | –0.93 | -0.95 | –0.7 | –0.96 | 1 | 1 | 0.59 | 0.53 | 0.97 | 0.98 | 0.88 | 0.84 |
| SEN statement GS | –0.93 | –0.95 | –0.72 | –0.97 | 1 | 1 | 0.6 | 0.54 | 0.96 | 0.97 | 0.86 | 0.82 |
| SEN no statement D | –0.54 | –0.58 | –0.62 | –0.65 | 0.59 | 0.6 | 1 | 0.98 | 0.07 | 0.05 | –0.2 | –0.16 |
| SEN no statement GS | –0.47 | –0.51 | –0.47 | –0.55 | 0.53 | 0.54 | 0.98 | 1 | 0.03 | 0.01 | –0.02 | 0.03 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Non-white D | –0.82 | –0.86 | –0.54 | –0.87 | 0.97 | 0.96 | 0.07 | 0.03 | 1 | 0.99 | 0.92 | 0.89 |
| Non-white GS | –0.87 | –0.89 | –0.54 | –0.89 | 0.98 | 0.97 | 0.05 | 0.01 | 0.99 | 1 | 0.94 | 0.92 |
| ESL D | –0.83 | –0.86 | 0.29 | –0.66 | 0.88 | 0.86 | –0.2 | –0.02 | 0.92 | 0.94 | 1 | 0.99 |
| ESL GS | –0.79 | –0.82 | 0.36 | –0.61 | 0.84 | 0.82 | –0.16 | 0.03 | 0.89 | 0.92 | 0.99 | 1 |

For ease of analysis, therefore, Table 16.3 shows the same values as Table 16.2 but with the duplication of indices eliminated (only GS is retained). For a fuller analysis, see Gorard and Cheng (2011). What becomes clearer in this simplified table is that the values of GS for free school meal take-up and for eligibility are very strongly related (top left of table). Whichever way we measure free school meals, the results and their correlations with the other four indicators are similar. So using correlation, making some justifiable assumptions about correlations near 1 and ignoring FSM eligibility, we have 'reduced' 12 measures to five only. This kind of data reduction can make seeing the patterns in the data much easier.

**Table 16.3** Correlations between trends in all six indicators, using GS index of segregation, secondary schools in England

| | **FSM takeup** | **FSM eligible** | **SEN statement** | **SEN no statement** | **Non-white** | **ESL** |
|---|---|---|---|---|---|---|
| FSM takeup | 1 | 0.97 | –0.95 | –0.51 | –0.89 | –0.82 |
| FSM eligible | 0.97 | 1 | –0.97 | –0.55 | –0.89 | –0.61 |
| SEN statement | –0.95 | –0.97 | 1 | 0.54 | 0.97 | 0.82 |
| SEN no statement | –0.51 | –0.55 | 0.54 | 1 | 0.01 | 0.03 |
| Non-white | –0.89 | –0.89 | 0.97 | 0.01 | 1 | 0.92 |
| ESL | –0.82 | –0.61 | 0.82 | 0.03 | 0.92 | 1 |

A fairer estimate of the strength of the relationship between any two variables is the effect size found by squaring the R correlation coefficient. R-squared shows how much of the variance in one variable is common to the other. Table 16.4 shows the R-squared values from Table 16.3, but ignoring the second measure of free school meals. Many of these values are very small. For example, the R-squared between special needs with no statement and non-white ethnic origin is 0.0001. All such values less than 0.5 have been ignored for ease of analysis. What Table 16.4 now makes clear is that measures of segregation by special needs with no statement are unrelated to any other measure used here. This is a somewhat different and more sophisticated analytical conclusion than that suggested by Table 16.1 which initially led to both indicators of SEN being treated together (see above). This finding could be important, because it suggests that whatever causes changes in segregation by SEN without statements is not the same thing that causes segregation by poverty, language and ethnicity, or by SEN with statements.

**Table 16.4** R-squared between trends in five indicators, using GS index of segregation, secondary schools in England

| | **FSM takeup** | **SEN statement** | **SEN no statement** | **Non-white** | **ESL** |
|---|---|---|---|---|---|
| FSM takeup | 1 | 0.90 | – | 0.79 | 0.67 |
| SEN statement | 0.90 | 1 | – | 0.94 | 0.67 |
| SEN no statement | – | – | 1 | – | – |

| | | | | | |
|---|---|---|---|---|---|
| Non-white | 0.79 | 0.94 | – | 1 | 0.84 |
| ESL | 0.67 | 0.67 | – | 0.84 | 1 |

*Note:* Values less than 0.5 have been suppressed.

The other four indicators have substantial variation in common over time (reasonably high values of R), and so it may be that whatever causes change in these values has some similarity for all of them. The values for free school meal segregation and the other three measures are negatively related (see Table 16.3), which means that whatever drives changes does so in opposite directions for these two groups of indicators. It is reasonable to assume, for simplicity at present, that whatever causes segregation by ethnic origin is also related to what causes segregation by language in England. Segregation of pupils from families living in poverty (FSM) is to some extent a separate process from segregation by language/ethnicity, having a very different scale and a near opposite trend over time, and to some extent it is an inverse. So, their determinants might be related but in an opposite direction, although this seems an unlikely situation. For example, if selection by aptitude (known as ‘tracking’ in some countries) is a process likely to segregate pupils by poverty, it seems unlikely that it would also *desegregate* them by language/ethnicity (since origin and socio-economic status (SES) are often strongly related).

Although simple, this is a valuable analysis which will assist in the search for the causes of, and so the solutions to, segregation by disadvantage in schools. This is because it shows that the clustering of pupils with similar characteristics in schools is not just one process, but at least two. It is important because segregation is important, and because understanding how it occurs is a key part of overcoming its dangers. What these processes are and how they differ cannot be estimated using these same data, so as usual numeric analysis is not the end of an investigation but merely the start of a more detailed study.

## Questions for further investigation

1. Find a dataset in your own area of interest that contains a large number of cases and at least two real-number variables. Select two variables, and draw a cross-plot graph of their relationship. Is the relationship anything like a straight line? Calculate the correlation coefficient.
2. Find a dataset in your own area of interest that contains a large number of cases and at least onereal-number variable. Create a second variable ‘Group’, giving half of the cases the value 1 and half the value 0. Now find the mean score of the first variable for all cases labeled 1 in the second variable. And find the mean score for all cases labeled 0. Note the standard deviation for each mean, and find the average of these two standard deviations. Find the difference between the two mean scores, and divide by their average standard deviation. This is a standardised ‘effect’ size. Why is it not necessarily evidence of a cause:effect relationship here?
3. Find an article in your area of interest that uses correlation or factor analysis. Prepare a critique, noting how well and fully the paper presents the methods, whether the paper includes undigested computer output or whether the tables are made easy to read, whether the paper uses

significance incorrectly (with population data or a convenience sample) and whether the paper uses causal words like 'influence' or 'impact' without justification.

## Suggested further reading

Department for Education, England, School Performance Tables (2015) – online at: http://www.education.gov.uk/schools/performance/index.html /. This fantastic UK website provides data relevant to the performance of every school and college in England for as many years as these are available. Try some of the ideas in this chapter, by correlating scores for schools over time, for progress from one formal assessment to another, or examine the results in terms of other useful data provided, such as the level of student absence. Many other countries will have their own versions of this dataset.

Gigerenzer, G. (2002) *Reckoning with Risk.* London: Penguin. This is a brilliant book for anyone who wants to think more clearly about numbers and the use of numeric evidence in social science. It shows how experts and advisers frequently present real evidence in ways that are deeply misleading. And it does so in a way that is easy for any reader. Primary school arithmetic only required.

Gorard, S. (2001) *Quantitative Methods in Educational Research: The Role of Numbers Made Easy,* London: Continuum. This is a popular introduction to reasoning with statistics, including how to calculate and use correlations. The book has become a standard for many courses because it presents everything from the outset so simply and without the clutter of technical language.

## References


Cheng, S. C. and Gorard, S. (2010) 'Segregation by poverty in secondary schools in England 2006–2009: a research note', *Journal of Education Policy,* 25(3): 415–18.

Goldhaber, D., Lavery, L. and Theobald, R. (2015) Uneven playing field? Assessing the teacher quality gap between advantaged and disadvantaged students, *Educational Researcher*, 44, 5, 293-307.

Gorard, S. (2001) *Quantitative Methods in Educational Research: The Role of Numbers Made Easy.* London: Continuum.

Gorard, S. (2010a) 'All evidence is equal: the flaw in statistical reasoning', *Oxford Review of Education,* 36(1): 63–77.

Gorard, S. (2010b) 'Research design, as independent of methods', in Teddlie, C. and Tashakkori, A. (eds) *Handbook of Mixed Methods.* Los Angeles: Sage.

Gorard, S. and Cheng, S. C. (2011) 'Pupil clustering in English secondary schools: one pattern or several?', *International Journal of Research and Method in Education* (in press).

Gorard, S. (2013) *Research Design*, London: SAGE.

Gorard, S. and See, BH. (2013) *Overcoming disadvantage in education*, London: Routledge.

Gorard, S. and Smith, E. (2010) *Equity in Education: An International Comparison of Pupil Perspectives.* London: Palgrave.

Gorard, S., Taylor, C. and Fitz, J. (2003) *Schools, Markets and Choice Policies.* London: RoutledgeFalmer.

Yeung**,** R. and Phuong Nguyen-Hoang, P. (2016) Endogenous peer effects: Fact or fiction?, The Journal of *Educational Research*, 109, 1, 37-49.