The Statistician (1997)
46, *No.* 1, *pp.* 105–110

Football standings and measurement levels

By DANIEL B. WRIGHT†

University of Bristol, UK

[Received December 1995. Revised November 1996]

SUMMARY
Levels of measurement are a topic that confuses many students and researchers. Much of the focus has been comparing interval and ordinal levels of measurement. Here I focus on the difference between interval and ratio levels and discuss why this often neglected distinction is important. Two examples are discussed. The first example concerns income differences for siblings. The second is the way in which football standings are calculated in the English Premier League. I go into detail with the second example and demonstrate the consequences of the choice of measurement level.

1. Introduction

The levels of measurement are a framework put forward by Stevens (see, for example, Stevens (1946, 1951)) and are discussed at length in most books on measurement theory. To a lesser extent these levels are described in *all* introductory statistics text-books for social scientists. Stevens's argument was that certain variables have particular qualities and that these qualities are important for deciding which statistical tests can be done on them. Undergraduates are taught these correspondences, often as if they are uncontroversial.

Beneath the surface there are complications. One problem is that the statistical tests themselves are blind to the type of data (i.e. where the numbers come from). Lord (1953) gave the example of players' numbers in American football. These have some meaning (those with the numbers 50–79 are ineligible to catch a pass), but in most circumstances they certainly do not constitute a scale in the sense that weight or height or even responses to a rating scale do. However, there are no statistical rules that prevent someone from calculating the mean of a set of football numbers. In Lord's fictitious example, first- and second-year students were assigned numbers. When they began quibbling over who had higher numbers, calculating the means for these groups was in some sense justified. The analyst must consider whether the statistic is 'meaningful' for the set of numbers in the particular situation.

The four most discussed levels of measurement are nominal, ordinal, interval and ratio. However, from the axiomatic approach, 'there is an infinity of scale types' (Coombs *et al.* (1970), p. 16) based on the various possible transformations mapping the theoretical construct onto the observed variable. This approach is now widely accepted by most statisticians and methodologists. I focus on the two most commonly assumed of the quantitative scale types: interval and ratio.

The purpose of this paper is to provide a concise description of these levels with accessible examples, addressing one specific methodological confusion. One of the most common confusions spread among students is that it makes no difference whether data are treated as interval or ratio. Booth (1995), for example, claims that 'there is no practical distinction

†*Address for correspondence*: Psychology Department, University of Bristol, 8 Woodland Road, Bristol, BS8 1TN, UK.
E-mail: D.B.Wright@Bristol.ac.uk

© 1997 Royal Statistical Society 0039–0526/97/46105

between interval and ratio scales in psychology or any other empirical discipline' (p. 196). To a large extent this folk belief has stymied discussion of the differences among all quantitative scale types outside measurement theory texts (like Coombs *et al.* (1970)). I shall concentrate on this confusion.

I shall describe two examples where the distinction between interval and ratio does affect conclusions. The first is a hypothetical example on children's spending money. Second, I show how the tie-breaking mechanism used in the Premier League of English football assumes a particular metric and I describe how this can affect team strategy. Both of these examples have 'paired' data, such that a paired *t*-test could conceivably be used if particular hypotheses were to be tested and some distributional assumptions were made. Paired data are common in the social sciences and therefore this seemed a reasonable choice for illustrative purposes.

2. Ratio and intervals levels of measurement

Before discussing these examples, it is worth noting what is meant by ratio and interval data. More technical definitions can be given, but here I shall keep with the simplest explanations. For ratio data, the ratio of two numbers is seen as the quantity of interest. Consider the example of money. Suppose that you were doing research on the amount of spending money which brothers and sisters have. You might choose families with brother–sister twin pairs and ask the parents how much each child receives. Suppose that in one pair the brother has £5 and the sister has £10. The sister has twice as much money as her brother. If you were going to treat money as ratio data, then these ratios would be the data that you would analyse and you should describe the data appropriately (e.g. using the geometric mean). Money, in some cases, can be treated as ratio data. I stress in 'some cases' because the level of measurement of the data is dependent on how the researcher wishes to treat the variable. The level is not an intrinsic characteristic of the data. The data are simply numbers and statistical tests are blind to their history (though the statistical analyst should not be).

For interval data, the difference between the two amounts becomes the critical value. Thus, the sister has £5 more than her brother. In some cases treating the data as interval data might be the appropriate level of measurement, in some cases ratio and in some cases neither. It could even be argued that if one child receives more than the other this ordering represents the critical theoretical construct and that the magnitude of the difference, however quantified, is unimportant to the construct under investigation (for example, perceived equity may simply be whether the sibling receives the same, more or less). The same variable could be best treated as ratio, interval, ordinal, nominal or none of these.

This subjectivity makes some people think that the choice of level is unimportant. However, it does make a difference. If you treat the above data as ratio data, the difference between 5p and 10p is equivalent to the difference between £5 and £10. Because the ratios will probably tend to be more variable for low overall amounts (sister + brother totals), these will be weighted highly in the subsequent analysis. If the data are treated as interval data, then the difference between £5 and £10 will be 100 times larger than the difference between 5p and 10p. This will tend to lead to pairs with more total money being weighted more highly. There is no single correct answer for how to treat these data. Probably, here, neither the ratio nor the interval level would be appropriate because of these problems with differential weightings. More complex transformations would probably be appropriate and these transformations would imply that the data are neither ratio nor interval. However, the desired transformation will depend on the theorized mapping between the variable and the construct, not just on the observed data.

The appropriateness of measurement level can be questioned for almost all variables. The intelligence quotient IQ is an example given in some introductory text-books as an interval variable, but it is difficult to argue that it is *exactly* interval for any psychological theory. It is unlikely that, for whatever notion of intelligence that IQ is being used to measure, the difference between 60 and 70 will be the same as between 90 and 100. This is a common text-book example because of its manufactured normal distribution (an assumption for many statistical tests), not because of its level of measurement.

Cliff (1993a) argued that there are few if any interval level variables in the social sciences. This is because it is unlikely that the scale has the characteristic that equal distances at any point on the scale correspond to exactly equal distances for the theoretical construct under investigation. The same argument can be applied to ratio scales which also specify a precise correspondence between the variable and the construct.

3. Football standings

There are numerous examples where issues of the level of measurement are used outside individual academic disciplines. Using these examples is advantageous for teaching because specialist knowledge about the construct is not required.

The final standings for the 1995–96 English Premier League football season are listed in Table 1. In the English Premier League, positions are firstly determined by the number of wins and draws. Teams receive three points for a win and one point for a draw (none for losing). This algorithm also involves measurement issues; a win has been determined to be three times more valuable than a draw. It is designed to encourage teams not to 'settle for a draw'.

If two teams have the same number of points at the end of the season, league positions are then determined by the number of goals scored and goals conceded. The football authority has decreed that the team with the highest goal difference (goals scored minus goals conceded) is placed in the higher position. After the 1995–96 season, three teams were relegated from the Premier League. Although Manchester City had the same number of points as Coventry and Southampton, because they had scored 25 fewer goals than their opponents, and that this deficit is larger than for Coventry and Southampton, Manchester City were relegated. Other

TABLE 1
Premier League standings for the 1995–96 season

Team	*Wins*	*Losses*	*Draws*	*Points*	*Goals*			
					For	*Against*	*Difference*	*Ratio*
Manchester United	25	6	7	82	73	35	38	2.086
Newcastle	24	8	6	78	66	37	29	1.784
Liverpool	20	7	11	71	70	34	36	2.059
Aston Villa	18	11	9	63	52	35	17	1.486
Arsenal	17	9	12	63	49	32	17	1.531
Everton	17	11	10	61	64	44	20	1.455
Blackburn	18	13	7	61	61	47	14	1.298
Tottenham	16	9	13	61	50	38	12	1.316
Nottingham Forest	15	10	13	58	50	54	−4	0.926
West Ham	14	15	9	51	43	52	−9	0.827
Chelsea	12	12	14	50	46	44	2	1.045
Middlesbrough	11	17	10	43	35	50	−15	0.700
Leeds United	12	19	7	43	40	57	−17	0.702
Wimbledon	10	17	11	41	55	70	−15	0.786
Sheffield United	10	18	10	40	48	61	−13	0.787
Coventry	8	16	14	38	42	60	−18	0.700
Southampton	9	18	11	38	34	52	−18	0.654
Manchester City	9	18	11	38	33	58	−25	0.569
Queen's Park Rangers	9	23	6	33	38	57	−19	0.667
Bolton	8	25	5	29	39	71	−32	0.549

league positions, including the championship and lucrative spots in European competitions, can also be decided by this mechanism. Using goal difference assumes that a 3–1 victory is the same as a 6–4 victory. This assumes an interval scale.

This way of breaking ties is a measurement decision that is discussed by many football supporters. Other leagues or federations (and other sports) use different tie-breaking mechanisms. Many fans feel that a ratio scale might be better. Accordingly, a 3–1 score would be viewed as equivalent to a 6–2 score: the winner outscoring their opponent by a ratio of 3 to 1. These two mechanisms are clearly closely associated, but as illustrated below they can produce different league positions if some of the teams are involved in fewer goals than others.

There were four sets of point ties at the end of the 1995–96 season. Near the top of the table, Aston Villa and Arsenal were tied on points (63) and on goal difference (17). In these circumstances the higher position goes to the team that scored more goals. With teams scoring more than their opponents, this automatically means that the team with the lower goal ratio is placed on top. Aston Villa score about 1.49 goals for each of their opponents', whereas Arsenal scored 1.53 goals. Thus if a goal ratio had been used the order of these teams would be reversed and this would have affected seedings for European competitions (Arsenal had the lowest position of any of the English clubs to qualify for European competitions). Everton, Blackburn and Tottenham were also tied, and placed in this order based on goal difference. If goal ratio had been used, Tottenham would have finished above Blackburn. Similarly, the positions of Middlesbrough and Leeds would have been reversed. The final set of ties was the set which forced Manchester City to be relegated. Coventry and Southampton were tied on points and goal difference (−18). Coventry were placed above Southampton because of scoring more goals. Here, because their goal differences are negative, Coventry had the better goal ratio.

The question on the terraces and in the public houses around the grounds is whether the present mechanism is best, or whether something like a ratio measure would be better. The interval and ratio measures are two of the simplest tie-breaking mechanisms, and therefore the most realistic options.

Although the principal purpose of the tie-breaking mechanism is presumably to differentiate teams along some 'ability' dimension, it can serve other functions. For example, one function of these mechanisms could be to encourage teams to score more goals because the total number of goals in a game has been identified by research as the major predictor of enjoyment of televised football matches (Peter Ayton, personal communication).

Teams differ in quality and in strategies. Suppose that each team had an ability index and for argument that this index mapped onto a goal ratio. A good club may have a 2-to-1 ratio as their ability, a poor club a 1-to-2 ratio. Suppose also that teams could adjust their strategies to stress either scoring or defending, but that this ratio would remain constant. The effect of changing strategies would be to raise or to lower the total number of goals in the game. Fig. 1 shows in abstract the relationship between total goals, ability and the current tie-breaking mechanism. For clubs with ratios above 1, scoring more goals increases the goal difference. Clubs below this ratio would be encouraged to stress defence.

A club's ability will not be as simple as a ratio, and ability and strategy will not be independent. Still, the ratio ability index is a possibility that is worth exploring for illustrative purposes. Another possibility is that teams have the ability to keep the average goal difference constant, but to increase or decrease the total number of goals by moving attacking players to defence and vice versa. According to the present tie-breaking mechanism, if the goal difference per game was kept constant, teams would be neither encouraged nor discouraged from scoring more. The tie-breaking mechanism would be optimal for discriminating teams based on ability but would not influence strategy. The lines on Fig. 1 would be horizontal.

However, suppose that the tie-breaking mechanism was the ratio of goals scored to goals conceded. If the ability index was also the ratio of goals scored to goals conceded, as assumed for Fig. 1, then there would be no incentive for increasing the total number of goals (the lines

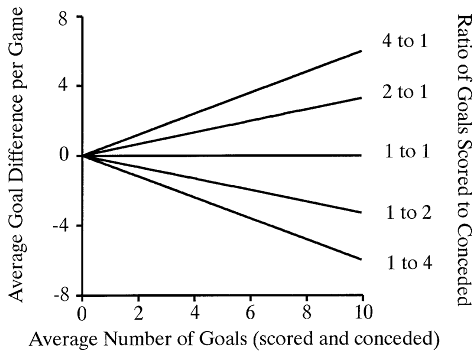

Fig. 1. Effect of total goals for teams with different goal ratios if the tie-breaking mechanism was the difference in goals scored and goals conceded

would be horizontal). The tie-breaking mechanism would be optimal for discriminating ability. However, if the ability index was a goal difference, then there would be a strategy incentive. As shown in Fig. 2, good clubs would try to minimize the total goals to keep their ratio high, whereas poor clubs would try to increase their total. As the total number of goals rises, assuming that the goal difference remains constant, the ratios for both good and bad clubs approach 1.

Before concluding, it is worth making one *caveat*. The importance of any tie-breaking mechanism on influencing strategy must be put in the context that the teams must be tied for it to have any effect. Thus, the three points for a win and one point for a draw system will have a larger effect than the tie-breaking mechanism. Arguments similar to those made here could be made about how the choice of this point system influences strategies.

4. Summarizing remarks

This paper arose from what seems a paradox to many students when they learn statistics. Students are often taught that interval and ratio refer to different measurement levels, but then they come to believe that these types may be treated in the same way. This paradox creates confusion about many aspects of levels of measurement. By providing two examples I hope to have clarified some of the issues.

I used the sibling allowance and football examples to illustrate that deciding whether a

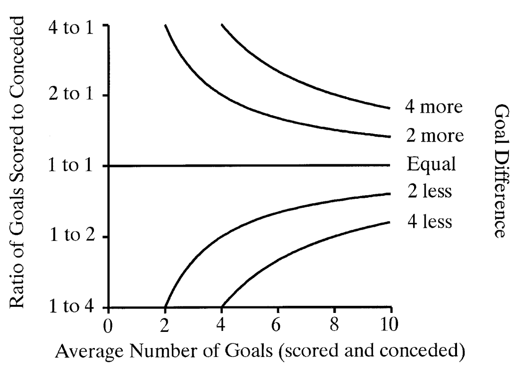

Fig. 2. Effect of total goals for teams with different goal differences if the tie-breaking mechanism was the ratio of goals scored

variable is ratio or interval does matter, that it can have real consequences and that although the issues are not statistically complex they do raise some difficult questions. Students seldom consider why a variable should be treated in a certain way and instead are simply taught how to plug numbers into equations. These examples I hope illustrate the types of questions that can be raised about any variable in empirical research.

Too often discussion about levels becomes combined with discussion of the distributional characteristics of a variable in relation to choosing a statistical test. These have little to do with each other. Cliff (1993b) described the level of a variable as a 'somewhat peripheral question' in deciding which test to use and that it 'is not a question which should be of central concern to the investigator' (p. 75). In the sibling money example, most researchers probably would be deciding between the paired *t*-test and the Wilcoxon matched pairs test. Both of these assume interval data—they both use the difference between the two siblings—but only the *t*-test makes distributional assumptions about this difference. If these are the tests that the researcher is considering, then the choice has nothing to do with the level of measurement.

Although I have focused on comparing interval and ratio scales, there are clearly similar arguments that can be made in differentiating most levels. I choose this comparison because it is often neglected. Most of the discussion in text-books concerns differentiating interval and ordinal data. Text-books tend either to advocate the use of ordinal statistics when there is *any* doubt about either the distribution or the level of the data, or they say that we do not need to worry too much about the assumptions. According to the first view, if there is any question about the level of measurement, then ordinal methods should be used. As noted, there are *always* questions about it; therefore, these books are in fact advocating that interval-based statistics should never be used.

With respect to the levels of measurement argument, I side towards the second view (see Wright (1997)). Since it is impossible in most cases to argue that the level of measurement of a variable is, for example, interval, students should be able to conceive of the variable at that level and to try other conceivable levels to make sure that their conclusions are robust. Students should be taught to construct statistical models that fit their theoretical hypotheses. In some cases the models will dictate ordinal level and the research hypotheses should reflect this (see Cliff (1993a)), but in other cases they will dictate some form of quantitative model. If students are planning to treat a variable as quantitative they should realize that there are many choices and that these can produce different results.

References

Booth, D. (1995) 'Tricky stats' no. 3. *Psychologist*, **8**, 196–197.

Cliff, N. (1993a) Dominance statistics: ordinal analyses to answer ordinal questions. *Psychol. Bull.*, **114**, 494–509.

———(1993b) What is and isn't measurement. In *A Handbook for Data Analysis in the Behavioral Sciences: Methodological Issues* (eds G. Keren and C. Lewis), pp. 59–93. Hove: Erlbaum.

Coombs, C. H., Dawes, R. M. and Tversky, A. (1970) *Mathematical Psychology: an Elementary Introduction*. Englewood Cliffs: Prentice Hall.

Lord, F. M. (1953) On the statistical treatment of football numbers. *Am. Psychol.*, **8**, 750–751.

Stevens, S. S. (1946) On the theory of scales of measurement. *Science*, **103**, 677–680.

———(1951) Mathematics, measurement and psychophysics. In *Handbook of Experimental Psychology* (ed. S. S. Stevens), pp. 1–49. New York: Wiley.

Wright, D. B. (1997) *Understanding Statistics: Introduction to Statistics for the Social Sciences*. London: Sage.