MONOTONE REGROUPING, REGRESSION
AND SIMPSON'S PARADOX

by

YOSEF RINOTT and MICHAEL TAM

Discussion Paper # 305 December 2002

CENTER FOR THE STUDY
OF RATIONALITY

Feldman Building, Givat-Ram, 91904 Jerusalem, Israel
PHONE: [972]-2-6584135 FAX: [972]-2-6513681
E-MAIL: ratio@math.huji.ac.il
URL: http://www.ratio.huji.ac.il/
Monotone Regrouping, Regression, and Simpson’s Paradox

Yosef Rinott and Michael Tam
Department of Statistics, Hebrew University, Jerusalem 91905, Israel.
University of Southern California, Los Angeles, CA 90089.

Abstract

We show in a general setup that if data $Y$ are grouped by a covariate $X$ in a certain way, then under a condition of monotone regression of $Y$ on $X$, a Simpson’s type paradox is natural rather than surprising. This model was motivated by an observation on recent SAT data which are presented.

Keywords and phrases: SAT data, monotone regression function, data grouping

Yosef Rinott is Professor, Department of Statistics, Hebrew University, Jerusalem 91905, Israel (email: rinott@mscc.huji.ac.il). This work was partially supported by a grant from the Israel Academy of Sciences. It was written during a Sabbatical leave at the University of Southern California.

Michael Tam is a student, University of Southern California, Los Angeles, CA, 90089 (email: mtam@usc.edu).
1 Introduction: data and motivation

The following data on SAT scores, which motivated this note, are taken from The College Board organization web-site collegeboard.com. US entering college students are divided into groups according to high school GPAs; Table 1 provides the percentage in each group and the groups’ average Verbal and Math SAT scores in 1992 and 2002. The last line combines all grade groups to provide the averages for the whole population.

<table>
<thead>
<tr>
<th>National High School Grade Averages</th>
<th>More students with top grade</th>
<th>Falling SAT scores</th>
<th>Verbal</th>
<th>Math</th>
</tr>
</thead>
<tbody>
<tr>
<td>A plus</td>
<td>5% 7%</td>
<td>619</td>
<td>607</td>
<td>629</td>
</tr>
<tr>
<td>A</td>
<td>12 17</td>
<td>575</td>
<td>565</td>
<td>583</td>
</tr>
<tr>
<td>A minus</td>
<td>14 17</td>
<td>546</td>
<td>538</td>
<td>554</td>
</tr>
<tr>
<td>B</td>
<td>52 47</td>
<td>486</td>
<td>479</td>
<td>486</td>
</tr>
<tr>
<td>C</td>
<td>17 11</td>
<td>434</td>
<td>424</td>
<td>428</td>
</tr>
<tr>
<td>All grades</td>
<td>100% 100%</td>
<td>500</td>
<td>504</td>
<td>501</td>
</tr>
</tbody>
</table>

Table 1: Falling SAT scores between 1992 and 2002 in each grade group

Note that in every grade group and in both Verbal and Math SATs, there is a decline in the average test scores between 1992 and 2002. Does this indicate a general decline in SAT scores? Should the averages of the combined groups (i.e., whole population averages) be consistent with the trend found in each group?

Often this apparent logical conclusion does hold and it is easy to find data sets in which group averages and total averages behave consistently. However, this is not always true. Indeed, the same web-site indicates that average SAT scores, when not divided into groups by high school grades, have been on a slow but consistent rise in the past decade or so, and there is definitely no general decline. In particular, as shown in the last line of Table 1, in 1992 the average SAT scores of entering college classes in the Verbal and Math tests were 500 and 501. In 2002 these numbers rose to 504 and 516 respectively. This rising trend also holds when the data is broken by gender or by states.

How is this contradiction explained, and how did it happen? It may be thought of as a special case or a subtle version of Simpson’s Paradox, which refers to the reversal of the direction of a comparison or an association when data from several groups are combined to form a single group. We will show that in some situations, this reversal of trend is not surprising; in fact, it is natural.

There are well-known simple cases of Simpson’s Paradox. For example, SAT scores of two years in a given area are divided by students’ ethnic groups and each group shows improvement in time while the total population SAT scores decline. Such a reversal is easily understood if between the two years there was a sufficiently large
increase in a relatively weak (in the sense of having low SAT averages) ethnic group, due to, say, immigration. However in the data and grouping above, this does not seem to be the explanation.

A hint to an explanation of the above contradiction is given in the caption in The College Board web-site to the upper part of Table 1: “Rising Grades and Falling Test Scores May Signal Grade Inflation” which takes the SAT group averages decline as a signal of grade inflation. Reversing this logic we show that grade inflation between 1992 and 2002 can lead to this seeming paradox: decline of each grade group’s average SAT score and rise in the whole population. This trend reversal is natural under conditions which we analyze in Section 3 in a general theoretical setup. Section 2 provides a simulated example which sheds more light on the particular aspect of the paradox discussed here and its explanation.

2 A simulated example

In order to crystallize the issue, we next consider fictitious (simulated) data. The advantage of doing so is that we have a complete and simple data set with a known mechanism that generated it. Consider a simulated sample of 100 workers, where the X and Y-axes denote age (in years) and income (in thousands of $, say), in 1995.

Figure 1: Scatterplot of simulated values of age and income in 1995

Suppose that the same information was obtained again for the same 100 workers in 2000. Comparing the two columns of average income on the right of Table 2 below we see a decline between 1995 and 2000 in each age group. It would seem that the wages of these 100 workers decreased in the five year period. However,
in constructing these data the wages of all workers were kept constant, and hence the average (80.3) is unchanged. Moreover, the same group definitions were used in 1995 and 2000. The only difference between 1995 and 2000 is that each worker’s age increased by 5 years.

So how did this seemingly paradoxical decrease in group averages occur?

<table>
<thead>
<tr>
<th>Age Group</th>
<th>% in group</th>
<th>Average income</th>
</tr>
</thead>
<tbody>
<tr>
<td>50 &lt; age</td>
<td>9%</td>
<td>93.5</td>
</tr>
<tr>
<td>40 &lt; age ≤ 50</td>
<td>40</td>
<td>84.8</td>
</tr>
<tr>
<td>age ≤ 40</td>
<td>51</td>
<td>74.5</td>
</tr>
<tr>
<td>All groups</td>
<td>100%</td>
<td>80.3</td>
</tr>
</tbody>
</table>

Table 2: Average income of 100 workers in different age groups and the percentage in each group for two years. The average income is declining for each age group between 1995 and 2000, while the average over all groups is constant.

Before we discuss the paradox we make some remarks:

- We kept the \( Y \) values and their total average constant only for simplicity. It is easy to see that if salaries in 2000 would be about 5% higher than those of 1995, the group averages would still be lower than those of 1995, and with a different choice of parameters in the simulation, an increase higher than 5% can still yield the same result. Also, if in both years dollars are replaced by units that measure real purchasing power, we obtain the same result as above.

- The 1995 data of Figure 1, contains 100 i.i.d MATLAB generated observations from a Bivariate Normal(\( \mu, \Sigma \)) distribution, with mean vector \( \mu = (40, 80) \) and covariance matrix \( \Sigma = \begin{pmatrix} 49 & 49 \\ 49 & 100 \end{pmatrix} \), so the correlation = 0.7.

The 2000 data is exactly the same except that 5 years were added to each x-value. Table 2 is then obtained from these 1995 and 2000 simulated data.

The explanation to the seeming paradox is simple: workers who were between 35 and 40, or 45 and 50 in 1995, moved to a higher age group. If \( X \) denotes age in 1995, then the age groups in 2000 corresponds to grouping by \( X ≤ 35, \ 35 < X ≤ 45, \ 45 < X \). We constructed the data with positive correlation between age and income. The decrease in group averages is due to the fact that in each group, the older workers who tend to be its highest wage earners in 1995 moved to a higher age group in 2000, where they are the youngest and hence tend to be the lowest earners, thus lowering all averages. Likewise, in the opening example of SAT data, grade inflation moved the best B students to the A group. The B group lost its top students and thus its SAT average went down, while the A group gained students who are below its average.
and so its average also went down; hence both groups exhibit a decline of averages. In our two examples there was no overall decrease in the \( Y \) variable (SAT score, or income), in fact there was an increase in the first example, and it remained constant in the second. In both examples the decrease in the group averages of the variable \( Y \) was caused in the same way. Groups are defined by a variable \( X \) on which \( Y \) is positively dependent. Moving the top part of a group to the next higher group lowers the average in both groups. The lower group loses its best while the higher group gains those who did not qualify before, and are below its average.

Another example of this phenomenon is the following famous (in some version) anecdote: a group of top students of university B transferred to university A. The average IQ in each university went down.

### 3 Brief theory

**Definition 1.** Let \((X, Y)\) be a random vector. We say that \( Y \) has a Monotone (increasing) Regression Function on \( X \) if \( E[Y|X = x] \) is increasing in \( x \).

Examples of monotone regression functions include the case that \((X, Y)\) are bivariate normal with a positive correlation, and many other bivariate positively dependent variables. We remark that Positive Regression Dependence of \( Y \) on \( X \) (see, e.g., Tong 1980 for definition and examples) implies \( E[Y|X = x] \) increasing in \( x \), which in turn implies \( \text{Cov}(X, Y) \geq 0 \).

The following Proposition explains the above examples formally. It follows readily from the relation

\[
E[Y|X \in (a, b)] = \int_a^b E[Y|X = x]dF_X(x)/P(X \in (a, b)).
\]

**Proposition 2.** The following conditions are equivalent: (a) \( Y \) has a Monotone Regression Function on \( X \). (b) \( E[Y|X \in (a, b)] \) is increasing in \( a \) and in \( b \) for all \( a, b \). (c) \( E[Y|X \in (a, b)] \geq E[Y|g(X) \in (a, b)] \) for any strictly increasing function \( g \) satisfying \( g(x) \geq x \).

Thus, if we think of \( X \) as age in 1995, and \( Y \) as income in both 1995 and 2000, then the middle age group, for example, changed from \((40, 50)\) to \((35, 45)\). This explains the observed decrease in the group (conditional) mean, when a Monotone Regression Function is assumed. The function \( g(x) \) of condition (c) above was \( x + 5 \) for our simulated example, and it is a function that reflects grade inflation in the SAT data. An inflation function \( g(x) = ax \) for some \( a > 1 \) and \( x > 0 \) clearly satisfies the condition. The function \( g \) is the regrouping function, and when the same data \( Y \) is averaged and compared in groups determined by \( X \) and \( g(X) \) being in the same family of intervals, we obtain the paradox in a natural way.

We summarize the above discussion of Proposition 2 and reformulate it in the following proposition which provides sufficient and in some sense necessary conditions
for a Simpson’s Paradox type reversal between the total expectation of $Y$ (which here is constant) and the conditional on group expectations, which are all decreasing.

**Proposition 3.** Let $Y$ have a Monotone Regression Function on $X$, and let $X' = g(X)$, where $g$ is a strictly increasing function satisfying $g(x) \geq x$. Let $a_0 < a_1 < \ldots < a_k$ and consider the intervals $A_i = (a_{i-1}, a_i)$, $i = 1, \ldots, k$. Then in Table 3 below the column of conditional expectations by $X$ is larger than the column of conditional expectations by $X'$. Conversely, if for each such $g$ and any set of intervals the first column is larger than the second column for some $k \geq 2$, then $Y$ has a Monotone Regression Function on $X$.

<table>
<thead>
<tr>
<th>Group</th>
<th>by $X$</th>
<th>by $X'$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1$</td>
<td>$E[Y</td>
<td>X \in A_1]$</td>
</tr>
<tr>
<td>$A_2$</td>
<td>$E[Y</td>
<td>X \in A_2]$</td>
</tr>
<tr>
<td>$\ldots$</td>
<td>$\ldots$</td>
<td>$\ldots$</td>
</tr>
<tr>
<td>$\ldots$</td>
<td>$\ldots$</td>
<td>$\ldots$</td>
</tr>
<tr>
<td>$A_k$</td>
<td>$E[Y</td>
<td>X \in A_k]$</td>
</tr>
<tr>
<td>All groups</td>
<td>$\overline{E}[Y]$</td>
<td>$\overline{E}[Y]$</td>
</tr>
</tbody>
</table>

Table 3:

Articles like Samuels (1993) and Scarsini and Spizzichino (1999), and references therein, provide related, but more elaborate and therefore complicated explanations to versions of Simpson’s Paradox.

**References**

