The influence of teacher support on national standardized student assessment.

A fuzzy clustering approach to improve the accuracy of Italian students data

Claudio Quintano  Rosalia Castellano  Sergio Longobardi

sergio.longobardi@uniparthenope.it

Department of Statistics and Mathematics for Economic Research
UNIVERSITY OF NAPLES “PARTHENOPE”
“To err is human, to forgive divine ....... but to include errors in your design is statistical”

[Leslie Kish, 1978]
The excessive teacher support might be considered similar to an interviewer effect (Biemer, Groves, Lyberg, Mathiowetz and Sudman, 1991) and we might suppose that the student’s score is affected by an error component which inflates the measurement errors (Braverman, 1996)
This work considers data on students’ performance assessments collected by the Italian National Evaluation Institute of the Ministry of Education (INVALSI) in the s.y. 2004/05.

A multistage method, combining the factorial analysis with a fuzzy clustering approach, is implemented for evaluating and correcting the cheating in primary classes.
“Teacher cheating, especially in extreme cases, is likely to leave tell-tale signs” (Jacob B.A., Levitt S.D.; 2003)
“Teacher cheating, especially in extreme cases, is likely to leave tell-tale signs” (Jacob B.A., Levitt S.D.; 2003)
II CLASS - PRIMARY SCHOOL

Reading s.y. 2004/05
Mean = 87.81
Std. Dev. = 8.14
N = 30,031

Mathematics s.y. 2004/05
Mean = 74.71
Std. Dev. = 14.133
N = 30,097

Science s.y. 2004/05
Mean = 76.66
Std. Dev. = 14.615
N = 30,094

Reading s.y. 2005/06
Mean = 77.72
Std. Dev. = 13.029
N = 29,802

Mathematics s.y. 2005/06
Mean = 81.5
Std. Dev. = 11.439
N = 29,816

Science s.y. 2005/06
Mean = 88.54
Std. Dev. = 8.832
N = 29,831
The procedure is aimed at managing the presence of cheating at **class level**.

A class is considered “**suspicious**” if:

- *the within (class) variability of the final score is close to zero*
- *the answer behaviour is “homogeneous”*
- *the percentage of missing data is very low*
For each class, four indexes are computed:

- Class mean score: $\bar{p}_j$
- Class standard deviation score: $\sigma_j$
- Class index of answer homogeneity: $\bar{E}_j$
- Class non-response rate: $MC_j$


\[ \bar{p}_j = \frac{\sum_{k=1}^{N_j} p_{kj}}{N_j} \]

**Class mean score**

\[ \sigma_j = \sqrt{\frac{\sum_{k=1}^{N_j} (p_{kj} - \bar{p}_j)^2}{N_j}} \]

**Class standard deviation score**

\( p_{kj} \) denotes the score of \( k^{th} \) student of \( j^{th} \) class

\( N_j \) denotes number of respondent students of \( j^{th} \) class
The influence of teacher support on national standardized student assessment

Workshop INVALSI
“Metodi di identificazione, analisi e trattamento del cheating”
8 febbraio 2013, ROMA

Index of answer homogeneity

\[ E_j = \frac{\sum_{s=1}^{Q} E_{sj}}{Q} \]

Gini’s index of heterogeneity

\[ E_{sj} = 1 - \sum_{t=1}^{h} \left( \frac{n_t}{N_j} \right)^2 \]

\( h \) \( n_t/N_j \) denotes the relative frequency of students of \( j^{th} \) class that has given the \( t^{th} \) answer to \( s^{th} \) question.

The index of answer homogeneity is equal to zero when all students of \( j^{th} \) class have given the same answers to all test questions, while it reaches the value \((h-1)/h\) when in the \( j^{th} \) class the answer heterogeneity is maximum
$$MC_j = \frac{\sum_{k=1}^{N_j} M_{kj}}{N_j Q}$$

Class non-response rate

$M_{kj}$ denotes the number both of item non-responses and of invalid responses for the $k^{th}$ student of the $j^{th}$ class.

$Q$ denotes the number of administered item to $j^{th}$ class. It is a constant for each assessment area (reading, mathematics and science) and for each school level.

0 (there are no missing or invalid responses for the $j^{th}$ class)

1 (all students of $j^{th}$ class have given only missing or invalid answers)
At the second step, the size of the data matrix, composed of the four indexes at class level, is reduced to two components by using Factor Analysis with a principal component extraction (Jolliffe, 2002).

Eigenvalues of correlation matrix $R$, simple and cumulative percentage of explained variability by the principal component analysis

<table>
<thead>
<tr>
<th>COMPONENT</th>
<th>Initial Eigenvalues</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOTAL</td>
<td>% of Variance</td>
</tr>
<tr>
<td>1</td>
<td>2.956</td>
</tr>
<tr>
<td>2</td>
<td>0.723</td>
</tr>
<tr>
<td>3</td>
<td>0.288</td>
</tr>
<tr>
<td>4</td>
<td>0.032</td>
</tr>
</tbody>
</table>
The PCA allows to describe the answer behaviour of each student class through two variables.
Projection on the first two factorial axes plane of second class primary students

- High score
- High homogeneity
- Low variability
- Missing responses close to zero
Fuzzy k-Means (FKM)

Developed by Bezdek (1981) and Dunn (1974), it is a fuzzy version of the non-overlapping partition model - hard k-means.
Steps of the correction procedure:

- clustering the classes by **fuzzy k-means** algorithm
- *projection* on the factorial axes of the cluster centroids
- detection of “suspicious” cluster centroid
- computation of a **cheating index**
  (on the basis of degree of belonging to suspicious cluster)
The results of FKM are affected by two parameters:

- **number of the clusters (c)**
- **fuzziness index (m)**
Calibration strategy

A two-step approach:

Computation of some **validity measures** (number of cluster $c$)

**Sensitivity of FKM results** to the level of the fuzziness (fuzziness parameter $m$)
The validity measures to assess the goodness of the fuzzy partitions and to obtain the optimal number of $c$ are:

- **Fuzziness Performance Index (FPI)**
- **Normalized Classification Entropy (NCE)**
- **Separation index ($S$)**

These indices are calculated for several thresholds of the fuzziness parameters $m$ (1.5, 2.0, 2.5).
Fuzzy Performance Index (Roubens, 1982)

\[ FPI = 1 - \frac{cPC}{c - 1} \]

Partition coefficient (Bedzek, 1974)

\[ PC = \frac{1}{n} \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^2 \]

Normalized Classification Entropy (Roubens, 1982)

\[ NCE = \frac{PE}{\log c} \]

Partition entropy (Bedzek, 1981)

\[ PE = -\frac{1}{n} \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij} \log u_{ij} \]

The FPI and NCE are used for evaluating the fuzziness of the solutions. The lower the FPI and MPE values are, the more suitable is the corresponding solution (McBratney & Moore, 1985).
Separation index $S$ (Xie and Beni, 1991)

$$S = 1 - \frac{\sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^2 \| x_j - v_i \|^2}{n \min_{i,j} \| v_i - v_j \|^2}$$

Where the numerator denotes the compactness by the sum of square distances within clusters, while the denominator denotes separation by the minimal distance between clusters. The smaller the value of $S$, the better the compactness and separation between the clusters $v$.
Fuzzy performance Index (FPI), Normal Partition Entropy (NPE) and Separation index (S)
in correspondence of $m=1.5$ and for $c \in [2,20]$
Fuzzy performance Index (FPI), Normal Partition Entropy (NPE) and Separation index (S)
in correspondence of $m=2$ and for $c \in [2,20]$
Fuzzy performance Index (FPI), Normal Partition Entropy (NPE) and Separation index (S) 
in correspondence of $m=2.5$ and for $c \in [2,20]$
Number of classes with a weight close to zero (lower than 0.001) in correspondence of increasing values of fuzziness level: $1.1 \leq m \leq 2.5$ and for $c=3$
Some weighted central tendency measures of class mean score in correspondence of increasing values of fuzziness level: $1.1 \leq m \leq 2.5$ and for $c=3$

The correction procedure tends to equalize the central tendency measures for $m$ in the range $1.7-1.9$
Cluster 2 gathers the classes that present a cheating profile:

- high class average scores
- minimum within variability
- low presence of missing items
Indicating with “a” the cheating cluster, the degree of belonging to this cluster is:

\[ \mu_{ja} \]

This measure is considered as the “cheating probability” of \( j^{\text{th}} \) class

Otherwise it can be interpreted as the “cheating level” of each class
On the basis of $\mu_{ja}$, a weighting factor is developed:

$$W_j = 1 - \mu_{ja}$$

The students’ class with high probability to belong to outlier cluster will have a low weight while the class very far from this cluster will have a weight close to 1.
Some Results

Graphics by box plot of the index $u_{aj}$ distributions

(j-th unit degree of belonging to cheating cluster)
Some results

ORIGINAL DISTRIBUTION

ADJUSTED DISTRIBUTION
Some results

North-West

North-East

Centre

South

South and Islands

ADJUSTED DISTRIBUTION

ADJUSTED DISTRIBUTION

ADJUSTED DISTRIBUTION

ADJUSTED DISTRIBUTION

ADJUSTED DISTRIBUTION

ADJUSTED DISTRIBUTION

ADJUSTED DISTRIBUTION

ADJUSTED DISTRIBUTION

The influence of teacher support on national standardized student assessment

C. Quintano, R. Castellano, S. Longobardi
References

## Supplementary Slides

*Comparison between unweighted average score per class and the weighted score per class according to the factor $w_j=1- u_{aj}$*

<table>
<thead>
<tr>
<th></th>
<th>Original distribution</th>
<th>Adjusted distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MEAN</strong></td>
<td>74.71</td>
<td>67.39</td>
</tr>
<tr>
<td><strong>MODE</strong></td>
<td>100.00</td>
<td>68.75</td>
</tr>
<tr>
<td><strong>I QUARTILE</strong></td>
<td>64.42</td>
<td>60.93</td>
</tr>
<tr>
<td><strong>MEDIAN</strong></td>
<td>73.61</td>
<td>67.78</td>
</tr>
<tr>
<td><strong>III QUARTILE</strong></td>
<td>85.94</td>
<td>74.25</td>
</tr>
</tbody>
</table>
Supplementary Slides

The cluster centroids and the respective membership functions that solve the minimization problem of the $J_{FKM}$ function are:

$$v_i = \frac{\sum_{j=1}^{n} u_{ij}^m x_j}{\sum_{j=1}^{n} u_{ij}^m}, \quad 1 \leq i \leq c$$

$$u_{ij} = \left[ \sum_{k=1}^{c} \left( \frac{\|x_j - v_i\|^2}{\|x_j - v_k\|^2} \right)^{1/(m-1)} \right]^{-1}, \quad 1 \leq i \leq c, \quad 1 \leq j \leq n$$
Supplementary Slides

Invalsi macro areas (SNV 2004/05-2005/06)

North West=Valle d’Aosta, Piemonte, Liguria, Lombardia

North East=Trentino Alto Adige, Veneto, Friuli V.G., Emilia R.

Centre=Toscana, Umbria, Marche, Lazio

South=Abruzzo, Molise, Campania, Puglia

South and Islands= Basilicata, Calabria, Sicilia, Sardegna