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

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*]

TOTAL SURVEY ERROR

- **SAMPLING ERROR**
- **NON SAMPLING ERROR**
  - **SPECIFICATION error**
  - **PROCESSING error**
  - **FRAME error**
  - **NONRESPONSE error**
  - **MEASURAMENT error**

(Biemer and Lyberg, 2003)

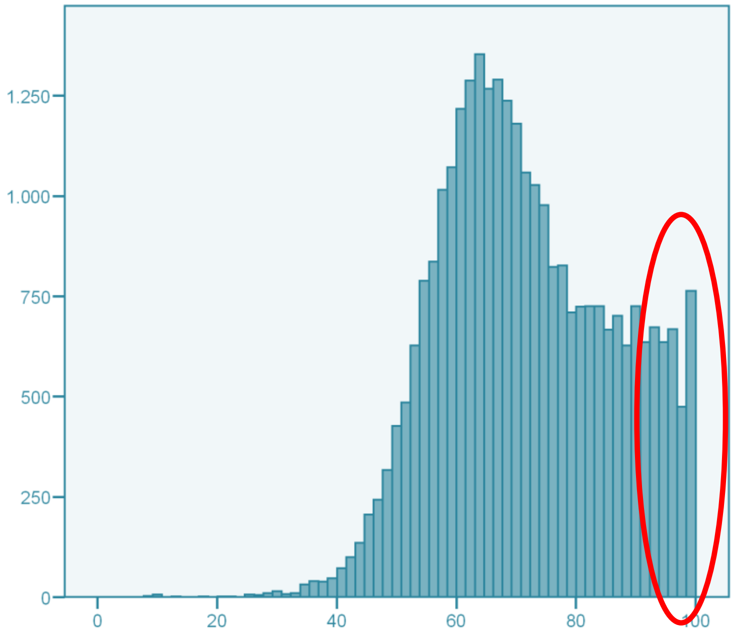
The excessive teacher support might be considered similar to an interviewer effect (Biemer, Groves, Lyberg, Mathiowetz and Sudman, 1991) and we might suppose that the student's score is affected by an error component which inflates the measurement errors (Braverman, 1996)

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

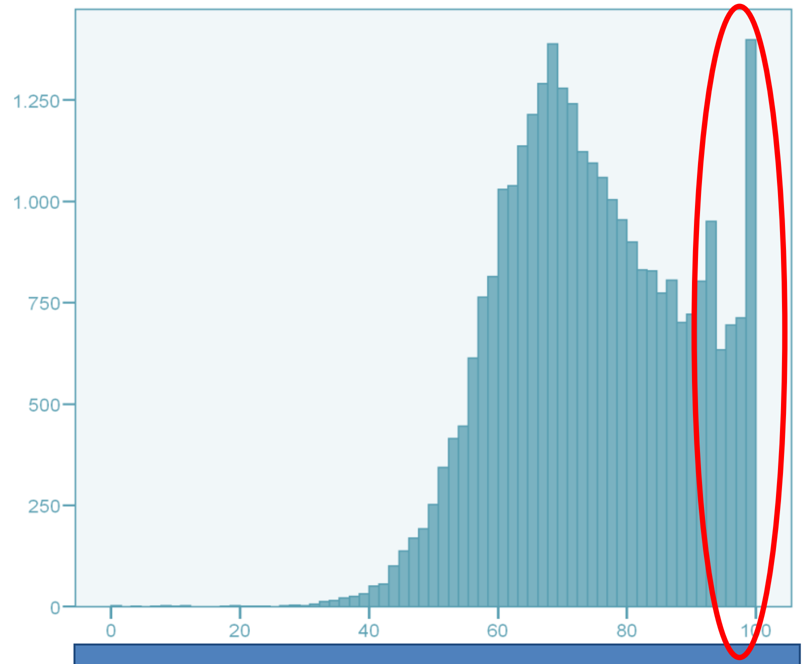
"Teacher cheating, especially in extreme cases, is likely to leave tell-tale signs" (Jacob B.A., Levitt S.D.; 2003)

MATHEMATICS CLASS MEAN SCORE - S.Y 2004/05



IV CLASS

PRIMARY SCHOOL

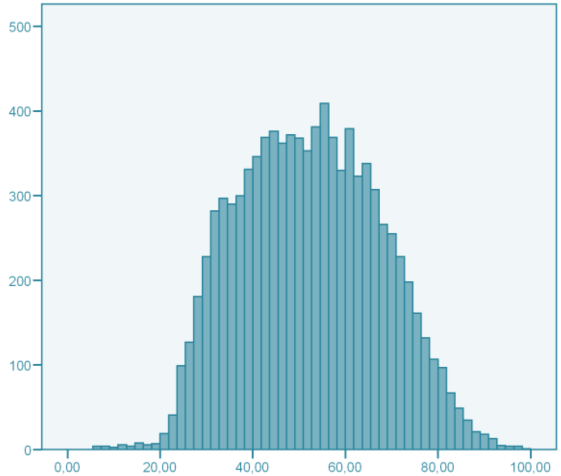


II CLASS

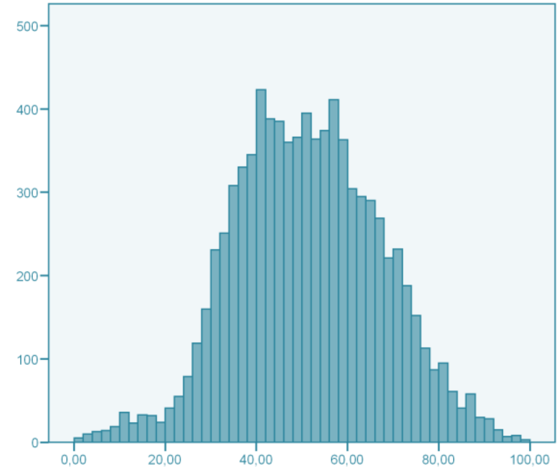
PRIMARY SCHOOL

"Teacher cheating, especially in extreme cases, is likely to leave tell-tale signs" (Jacob B.A., Levitt S.D.; 2003)

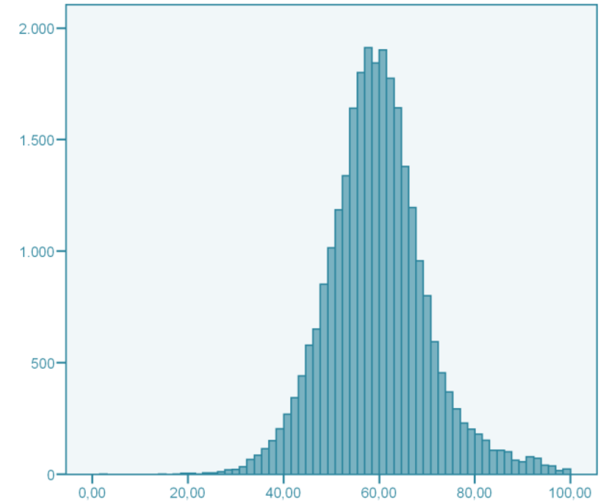
MATHEMATICS CLASS MEAN SCORE - S.Y 2004/05



III CLASS UPPER SECONDARY SCHOOL

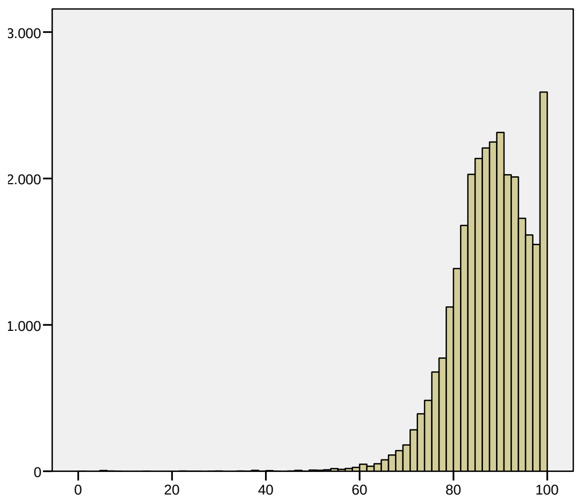


I CLASS UPPER SECONDARY SCHOOL

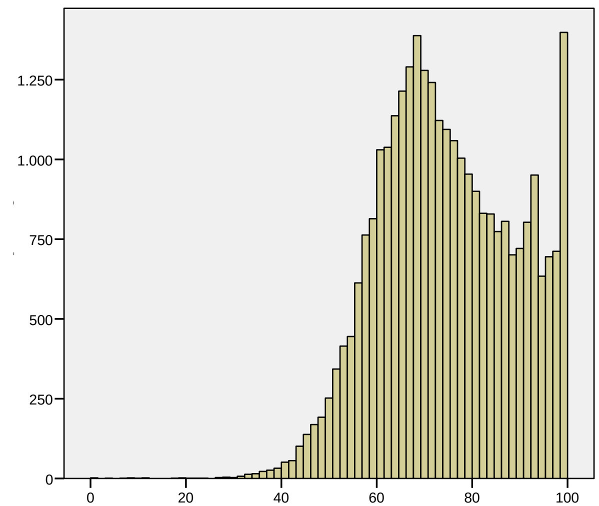


I CLASS LOWER SECONDARY SCHOOL

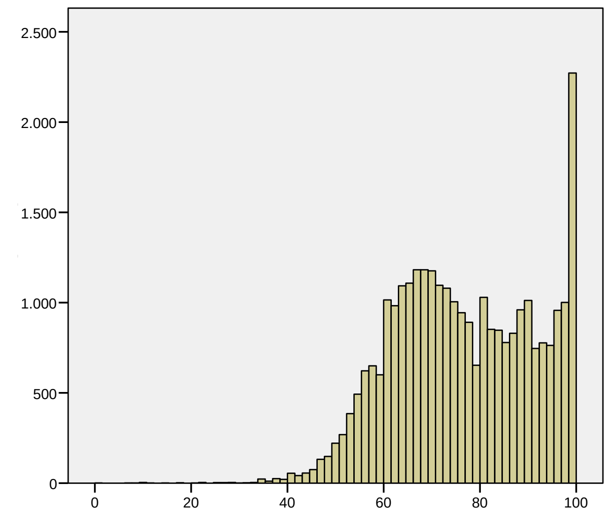
II CLASS - PRIMARY SCHOOL



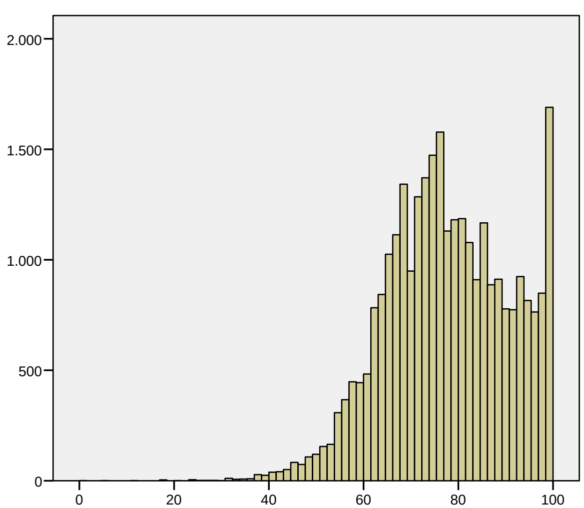
Reading s.y. 2004/05



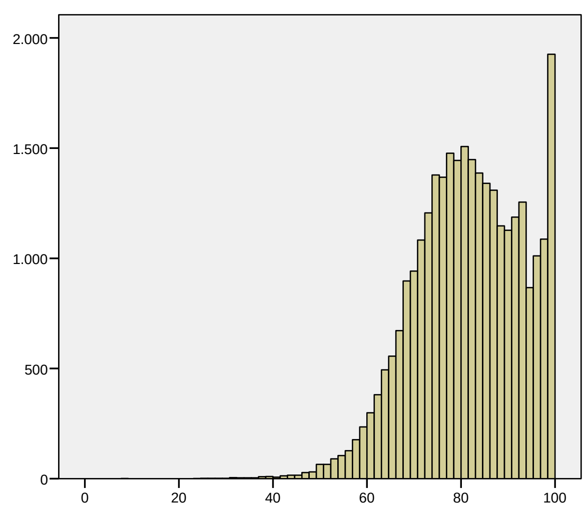
Mathematics s.y. 2004/05



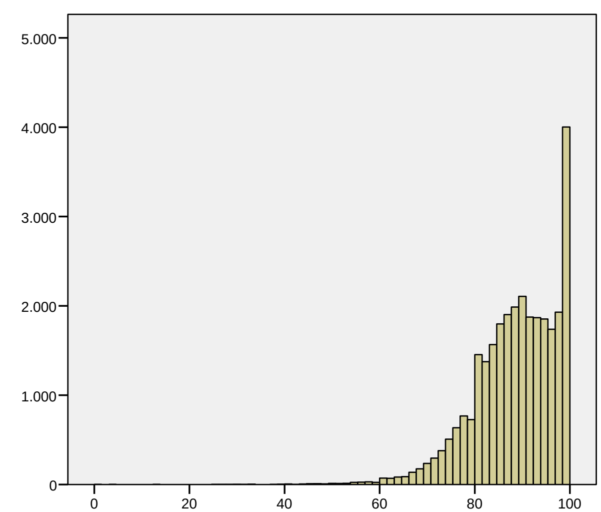
Science s.y. 2004/05



Reading s.y. 2005/06



Mathematics s.y. 2005/06



Science s.y. 2005/06

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: $\overline{p}_j$
- Class standard deviation score: σ_j
- Class index of answer homogeneity: $\overline{E}_j$
- 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 kth student of jth class

N_j denotes number of respondent students of jth class

$$\overline{E}_j = \frac{\sum_{s=1}^{Q} E_{sj}}{Q}$$

Index of answer homogeneity

Mean of h Gini indexes →

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

Gini's index of heterogeneity

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

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)

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

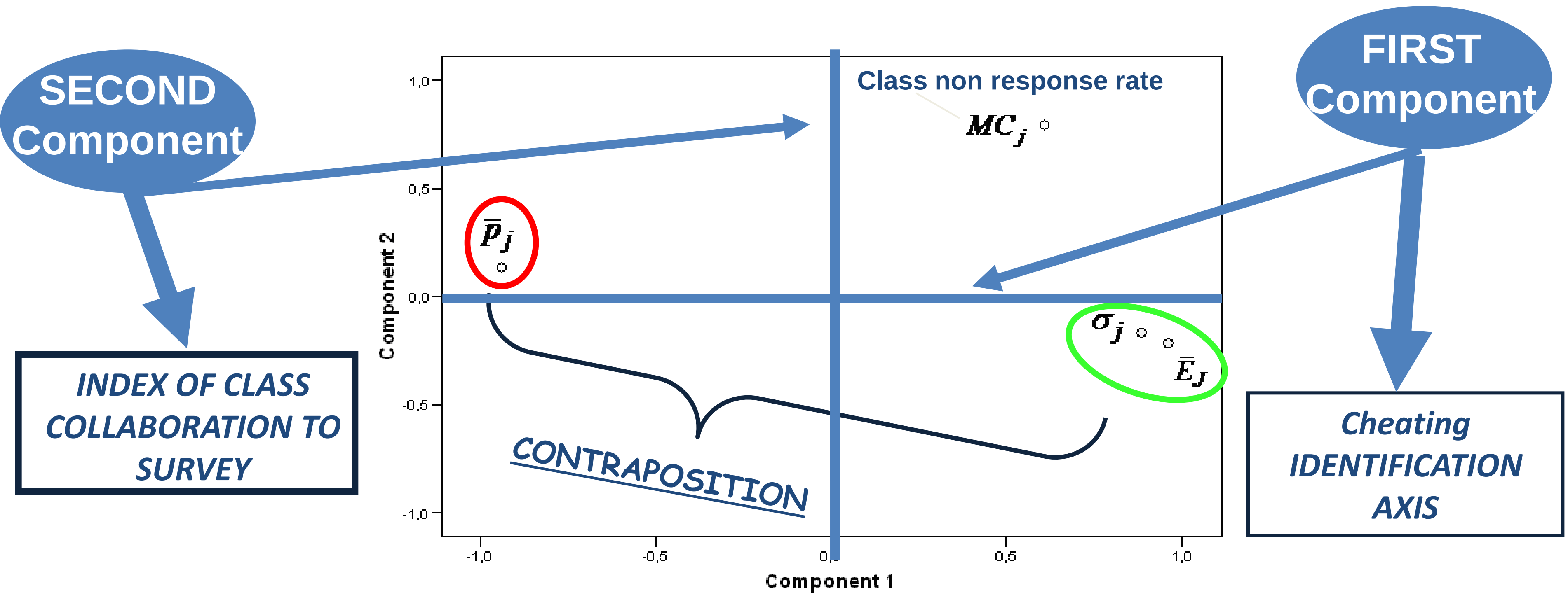
Q denotes the number of administered item to j^{th} class. It is a constant for each assessment area (reading, mathematics and science) and for each school level.

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).

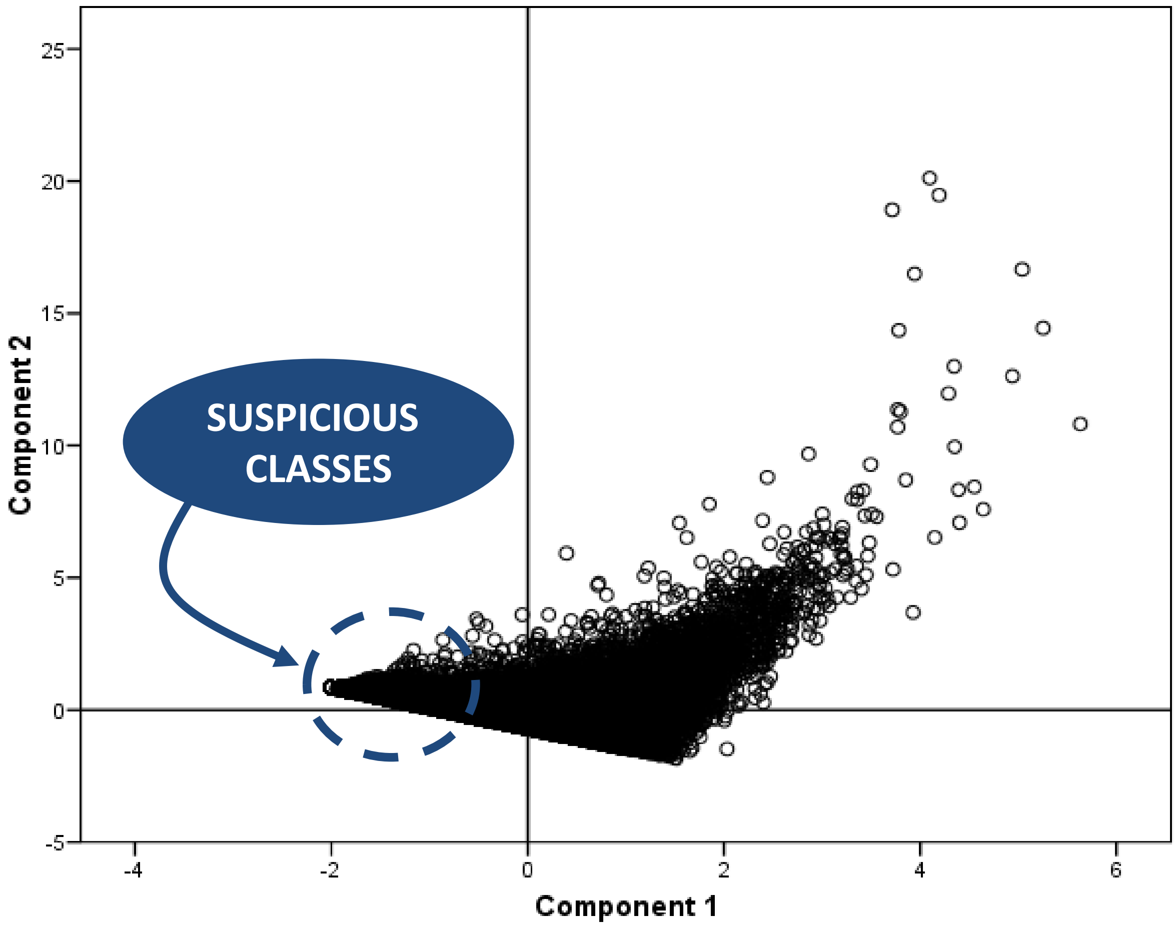
Eigenvalues of correlation matrix R, simple and cumulative percentage of explained variability by the principal component analysis

COMPONENT	Initial Eigenvalues		
	TOTAL	% of Variance	Cumulative %
1	2.956	73.911	73.911
2	0.723	18.086	91.997
3	0.288	7.211	99.208
4	0.032	0.792	100.000

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)

Overcoming the hard clustering → ***The fuzzy k-means***

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)

$$J_{FKM} = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^{m} \left\| x_j - v_i \right\|^2$$

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)

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

$$NCE = \frac{PE}{\log c}$$

Normalized Classification Entropy (Roubens, 1982)

→

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

Partition entropy (Bedzek, 1981)

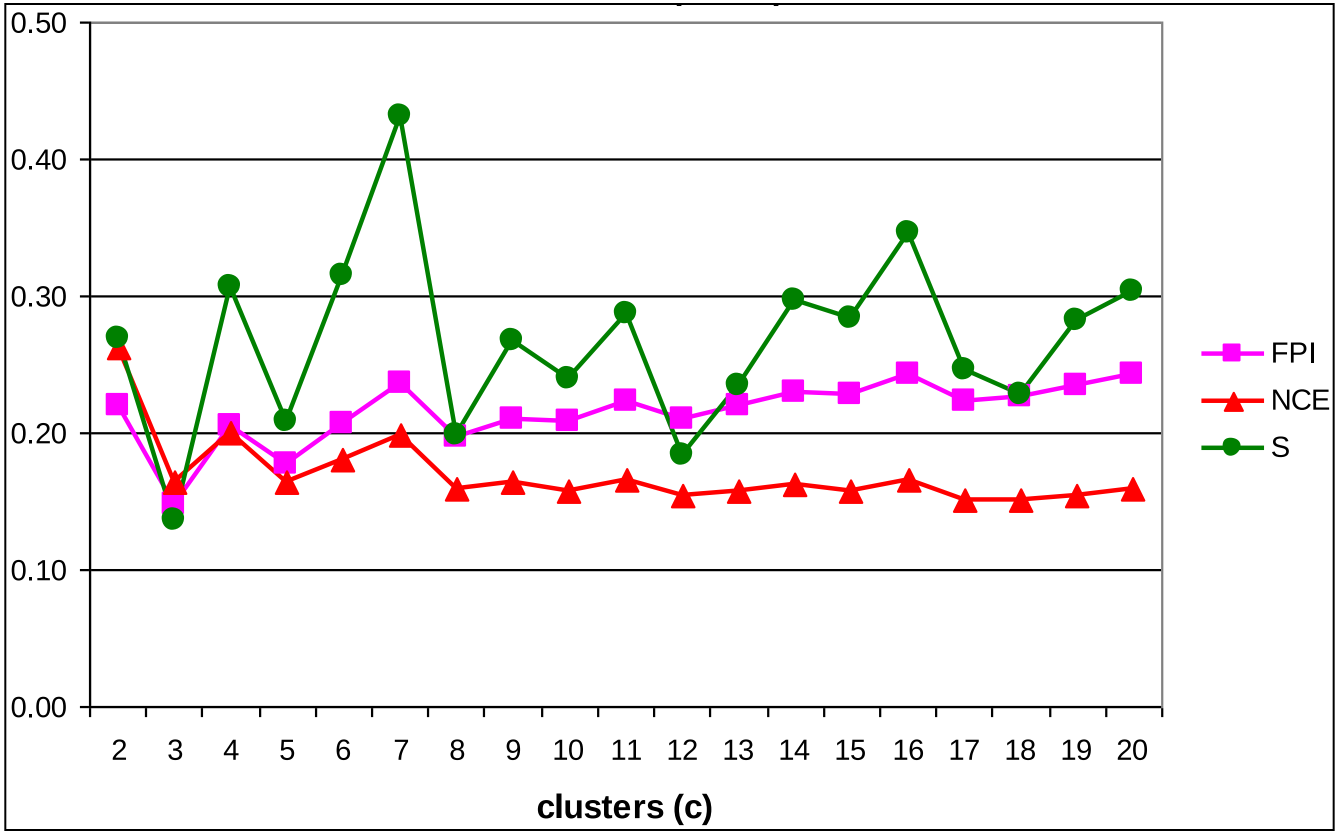
The FPI and NCE are used for evaluating the fuzziness of the solutions. The lower the FPI and MPE values are, the more suitable is the corresponding solution (McBratney & Moore, 1985).

Separation index S (Xie and Beni, 1991)

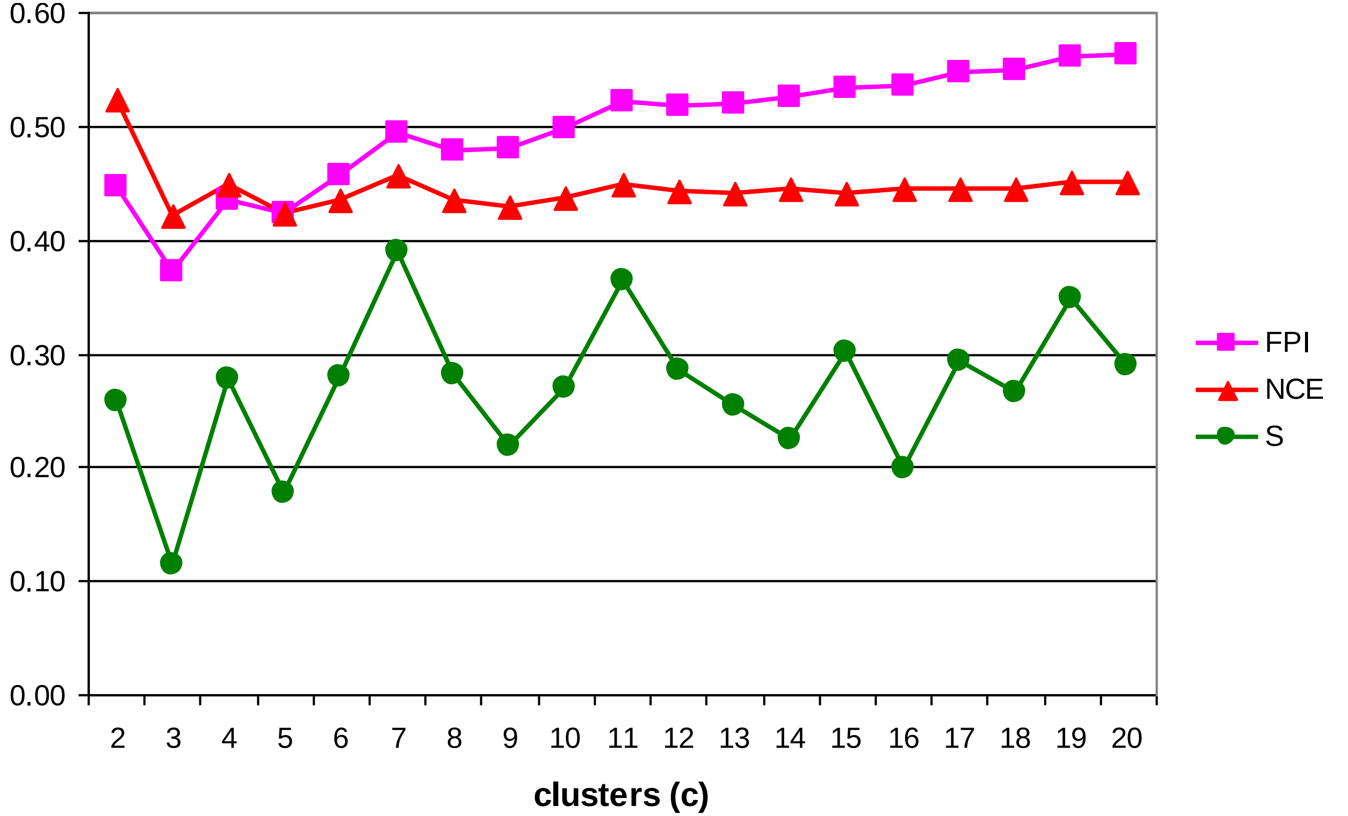
$$S = 1 - \frac{\sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^{2}\left\|x_j - v_i\right\|^2}{n \min_{i,j}\left\|v_i - v_j\right\|^2}$$

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

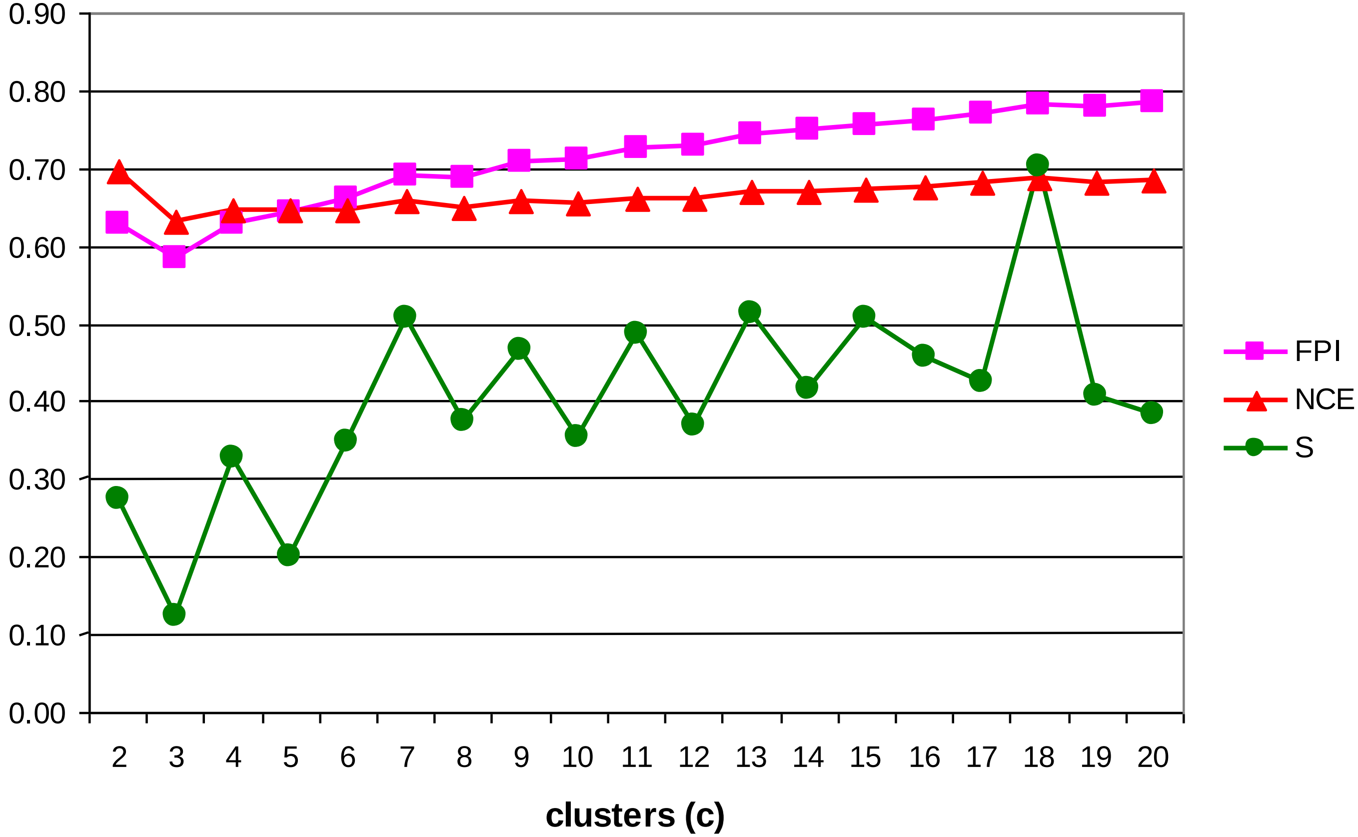
Fuzzy performance Index (FPI), Normal Partition Entropy (NPE) and Separation index (S) in correspondence of ***m=1.5 and for c ∈[2,20]***



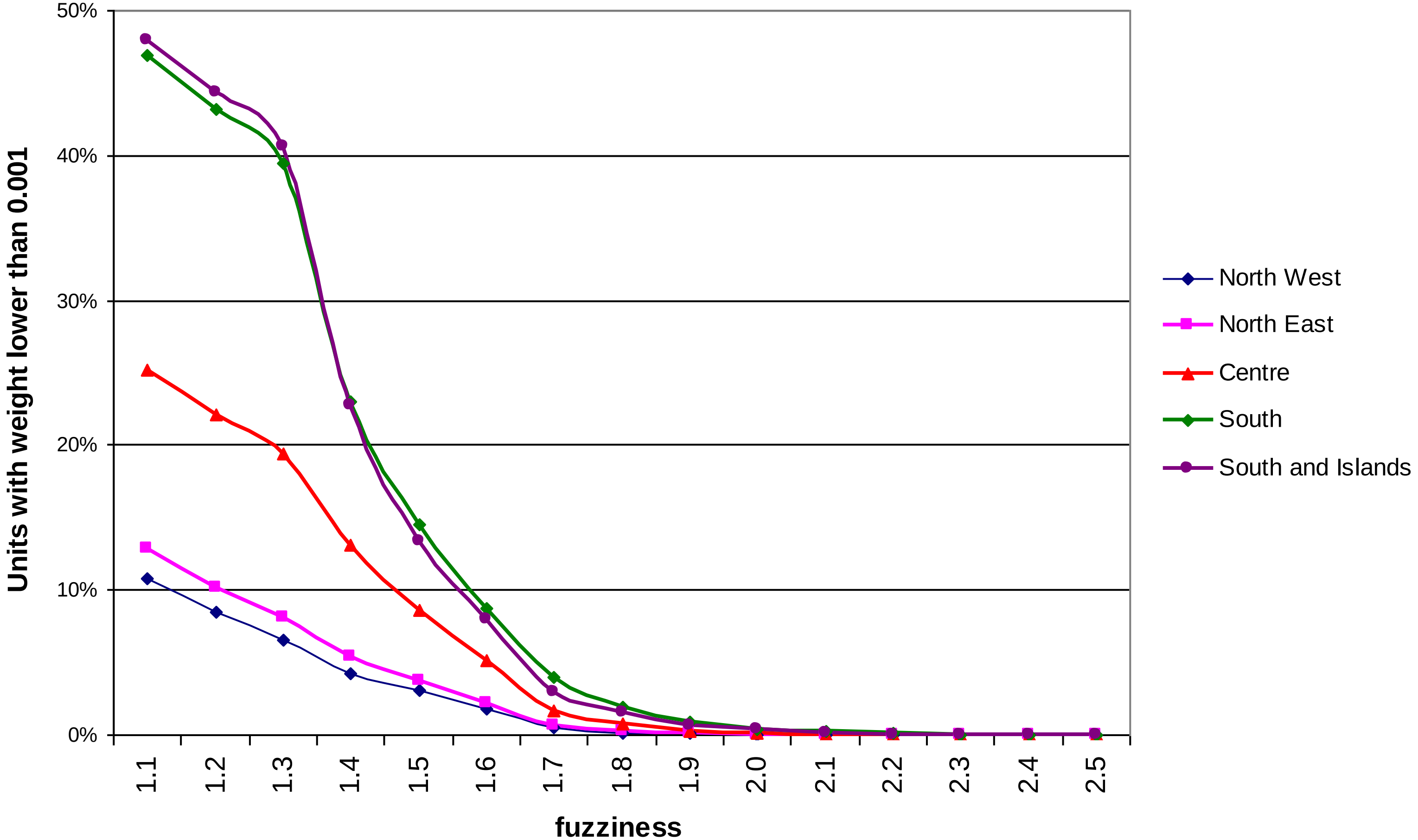
Fuzzy performance Index (FPI), Normal Partition Entropy (NPE) and Separation index (S) in correspondence of m=2 and for c ∈[2,20]



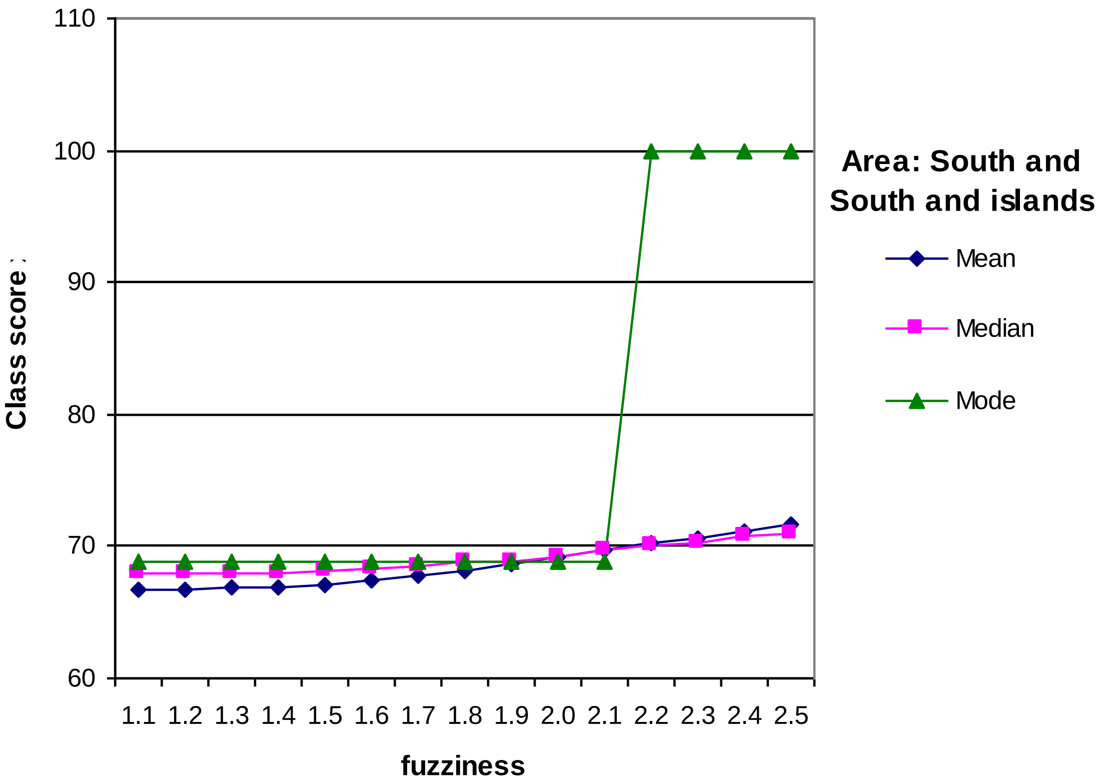
Fuzzy performance Index (FPI), Normal Partition Entropy (NPE) and Separation index (S) in correspondence of ***m=2.5 and for c ∈[2,20]***



*Number of classes with a weight close to zero (lower than 0.001) in correspondence of increasing values of fuzziness level: **1.1 ≤ m ≤ 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 ≤ m ≤ 2.5 and for c=3***

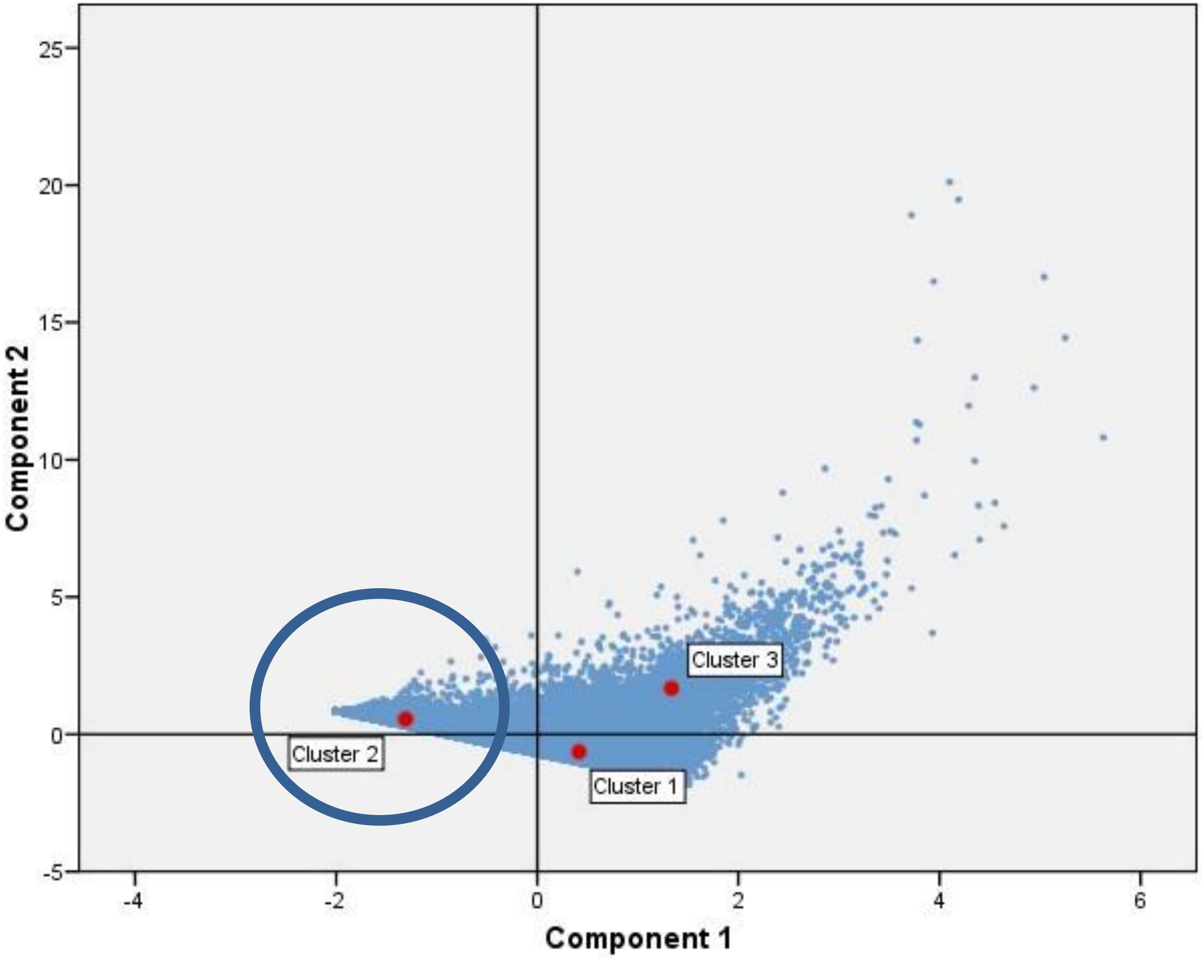


The correction procedure tends to equalize the central tendency measures for *m* in the range 1.7-1.9

Cluster 2 gathers the classes that present a cheating profile:

- high class average scores
- minimum within variability
- low presence of missing items

Projection on factorial plane of centroids computed by fuzzy k-means algorithm, c=3, m=1.7



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 jth class

Otherwise it can be interpreted as the "***cheating level***" of each class

On the basis of μ_{ja}, a weighting factor is developed:

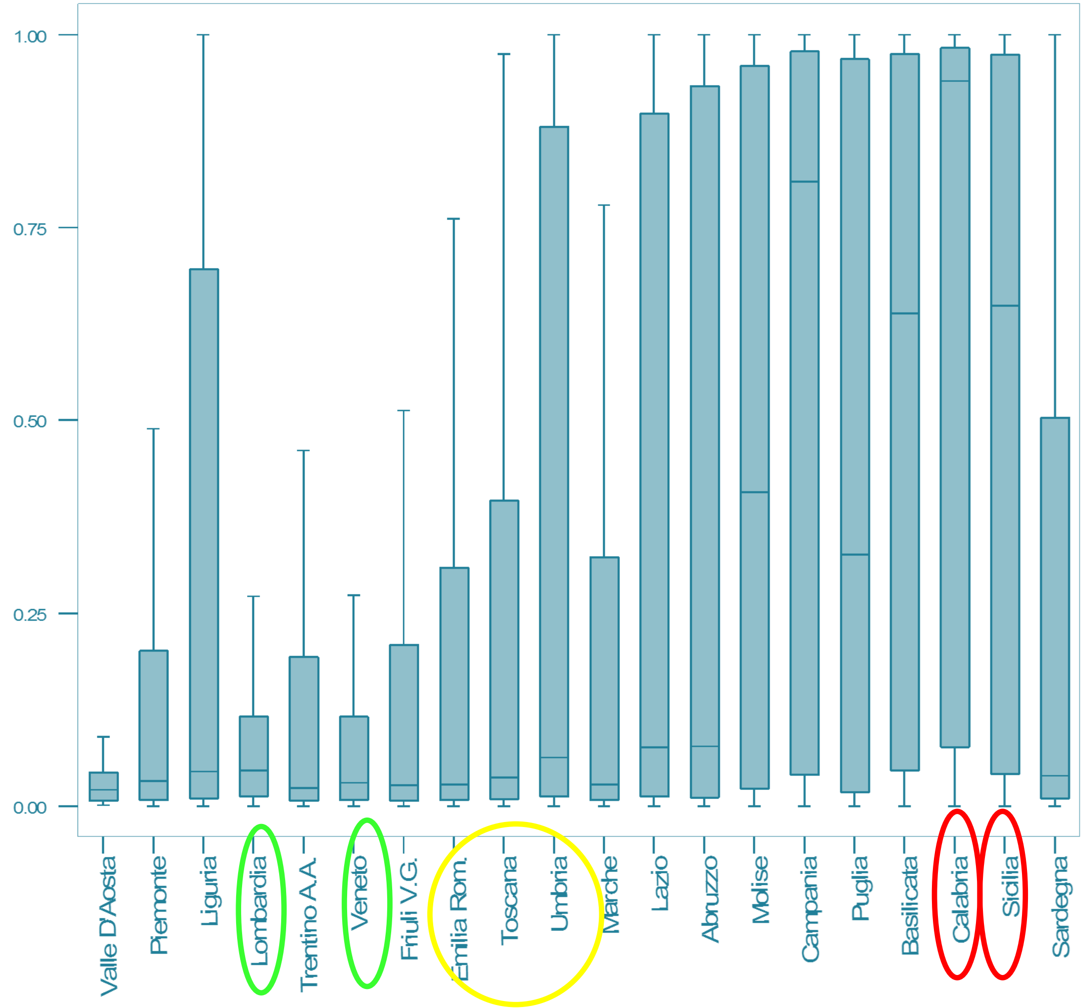
$$W_j = 1 - \mu_{ja}$$

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

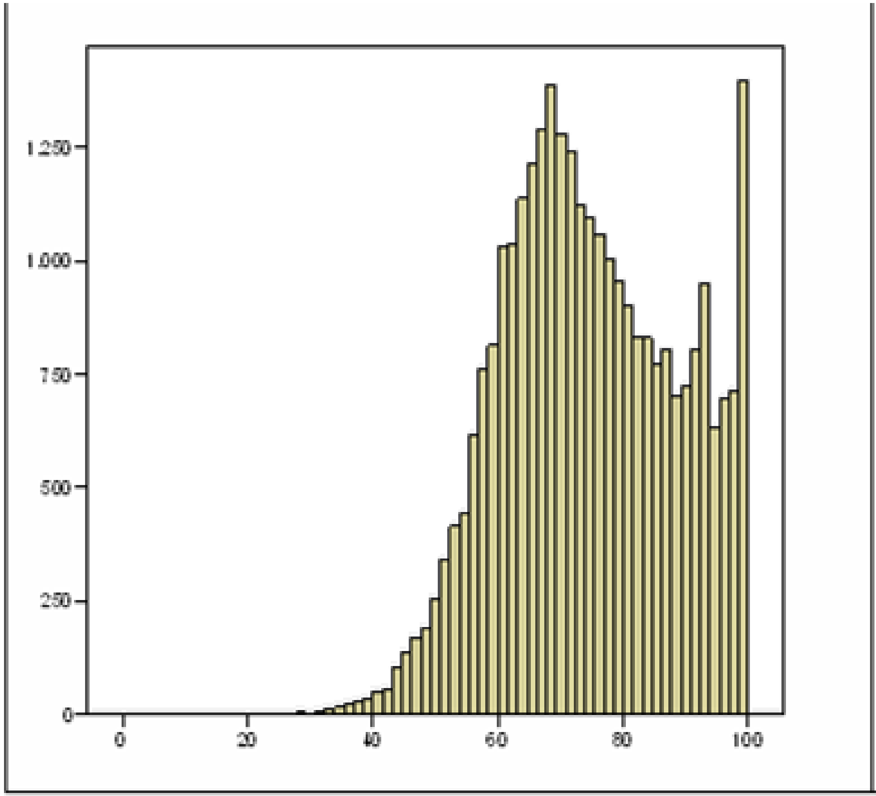
Graphics by box plot of the index u_{aj} distributions

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

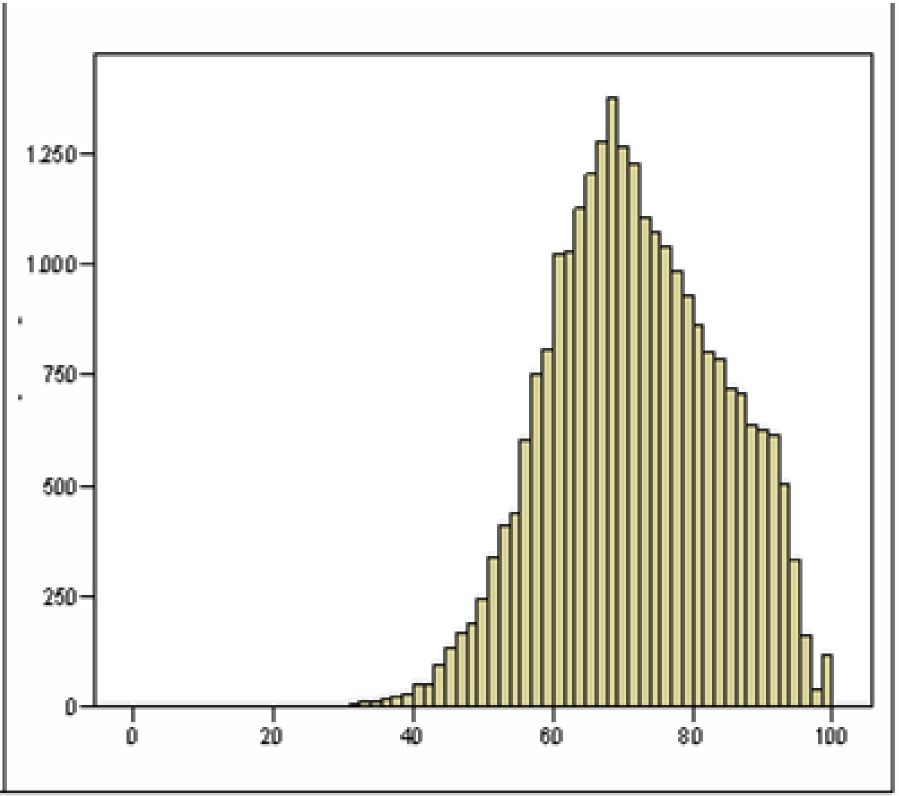


Some results

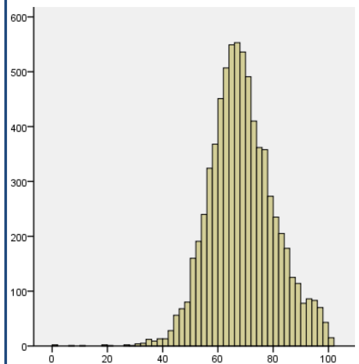
ORIGINAL DISTRIBUTION



ADJUSTED DISTRIBUTION

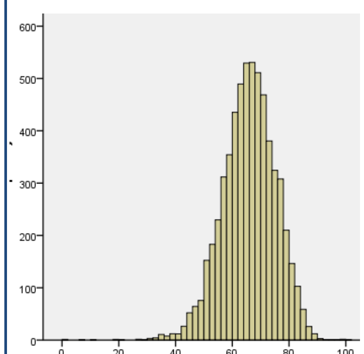


Some results

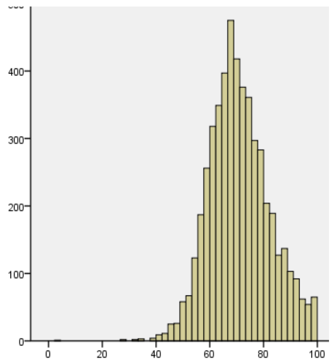


ORIGINAL DISTRIBUTION

North-West

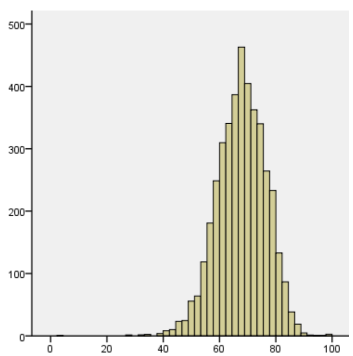


ADJUSTED DISTRIBUTION

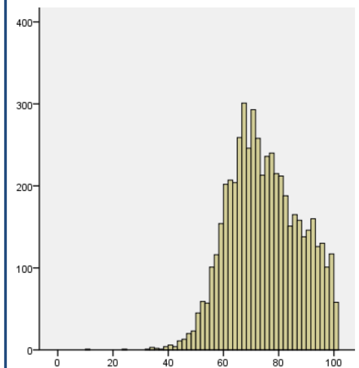


ORIGINAL DISTRIBUTION

North-East

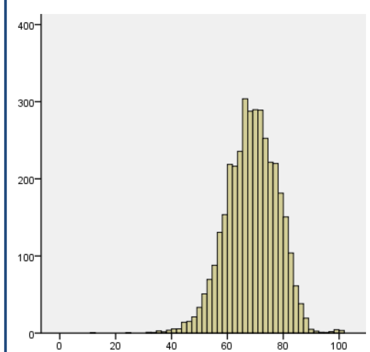


ADJUSTED DISTRIBUTION

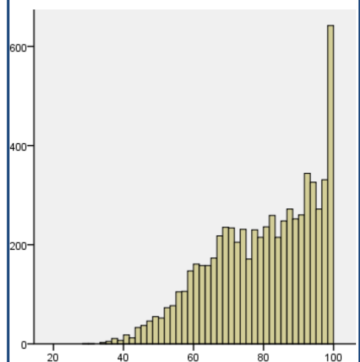


ORIGINAL DISTRIBUTION

Centre

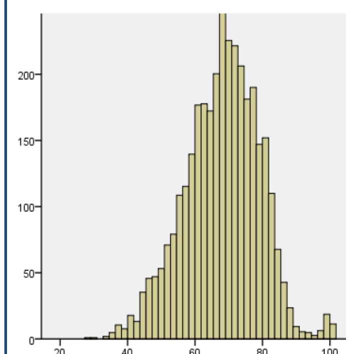


ADJUSTED DISTRIBUTION

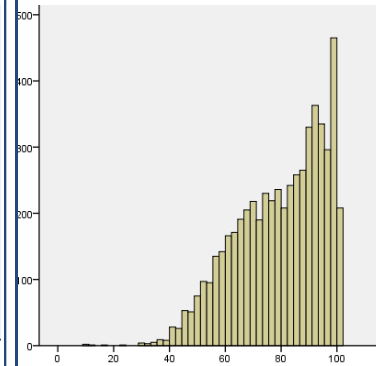


ORIGINAL DISTRIBUTION

South

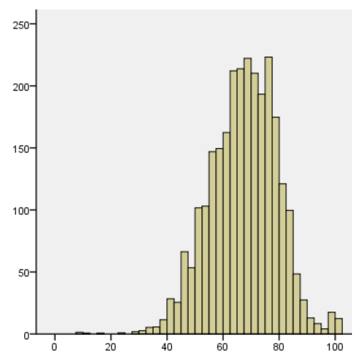


ADJUSTED DISTRIBUTION



ORIGINAL DISTRIBUTION

South and Islands



ADJUSTED DISTRIBUTION

References

Bezdek J.C., (1974). Cluster validity with fuzzy sets. *Journal of Cybernetics,* 3, 58-73.

Bezdek J.C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms.* Plenum Press

Biemer P.P., Groves R.M., Lyberg L.E., Mathiowetz N.A., Sudman S. (1991). *Measurement Errors in Surveys*. Wiley, New York.

Braverman M. (1996). Sources of Survey Error: Implications for Evaluation Studies. *New Directions for Evaluation: Advances in Survey Research*, 70, 17-28.

Dunn J.C. (1974). A Fuzzy Relative of the ISODATA Process and its Use in Detecting Compact Well Separated Clusters. *Journal of Cybernetics*, 3, 32-57.

Jacob B.A., Levitt S.D. (2003) Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating. The Quarterly Journal of Economics, 118(3), 843-877

Jolliffe I.T. (2002). *Principal Component Analysis*. Springer, New York.

McBratney A.B., Moore A.W. (1985). Application of fuzzy sets to climatic classification. *Agricultural and Forest Meteorology,* 35, 165-185.

Quintano, C., Castellano R., Longobardi S. (2009). A Fuzzy Clustering Approach to Improve the Accuracy of Italian Student Data. An Experimental Procedure to Correct the Impact of Outliers on Assessment Test Scores, Statistica & Applicazioni 7,149-171.

Roubens M. (1982). Fuzzy clustering algorithms and their cluster validity. *European Journal of Operational Research*, 10 (3), 294–301.

Supplementary Slides

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

	Original distribution	Adjusted distribution
MEAN	74.71	67.39
MODE	100.00	68.75
I QUARTILE	64.42	60.93
MEDIAN	73.61	67.78
III QUARTILE	85.94	74.25

Supplementary Slides

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

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

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

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