

A proposal for score development from a questionnaire with a ranked one dimensional structure in MCA

Maidier Mateo-Abad, Arantza Urkaregi, Jone Lázaro, Inmaculada Arostegui

Universidad del País vasco/ Euskal Herriko Unibertsitatea

mateomaider@gmail.com

07/09/2016



Index

- Motivation
- Aim of the study
- Methodology
- Application to real data
- Conclusion
- Future work

Motivation

- **Surveys** are one of the **most frequently used assessment tools**.
- They are useful for **gathering information** directly from a **large number of people** in a systematic and standardized way.
- The questionnaires are often developed to **collect and summarize** demographic **data**, opinions, knowledge, behaviors, experiences, needs, etc.
- Surveys are very **useful to evaluate** the process and impacts of programs, interventions or policies.

Motivation

- The most widely used rating scales are the **ordinal scales** such as *Likert type* scales.
- A **total score** of this type of scales is commonly calculated by **just summing up** the numerical value assigned to each response of the multiple items.
 - Limitations:
 - It assumes that all items and the responses of each are **uniformly distributed**, with the **same weights**
 - It assumes an **underlying linear structure**.

Aim of the study

- **Propose a method** to create an **overall score based on the MCA coordinates** of the first factor for each item response.
- **Apply** this methodology to a **real data and compare** this overall score with
 - An existing general assessment
 - A classic score
- **Develop a R package** for the easy implementation of the method.

Methodology

- **Multiple correspondence analysis (MCA)** is mainly used for the statistical analysis of **qualitative variables**.
- MCA is an **explanatory dimension-reduction technique** designed to analyze underlying structures in a data set.
- MCA produces **results in terms of maps**: the interpretation of the results is done based on the **relative position of the individuals or variable categories**.
- By applying MCA to a questionnaire, the responses usually have an **arch shape** structure, usually named *Guttman effect*.
- A good feature is that the **missing values** can be retained as **additional categories** for each active variable.

Methodology

$Q \equiv$ Number of active items $\rightarrow q=1, \dots, Q$

$J_q \equiv$ Number of categories for the q variables $\rightarrow j_q=1, \dots, J_q$

- **Steps:**

1. **Apply the MCA** to the items.
2. **Prove** that the categories of each item are **ordered along the first factor**.
3. **Assign the coordinate** of the first factor to each item response option (C_{q,j_q})
4. **Calculate the weights** for each item response, in order to define, by summing them, an overall score from 0 to 100. [1]
5. **Calculate the Overall score** [2]

Methodology

Overall Score

$$Weight_{q,j_q} = \frac{(C_{q,j_q} - min_q) \times 100}{(range)}, q = 1, \dots, Q; j_q = 1, \dots, J_q \quad [1]$$

where

$$range = \sum_{q=1}^Q range_q = \sum_{q=1}^Q (max_q - min_q)$$

Then,

$$S_{overall} = \sum_{q=1}^Q Weight_{q,k_q} \quad [2]$$

Methodology

Classic Score

$$S_{classic} = \frac{(\sum_{q=1}^Q j_q) * 100}{maxSc}, \quad q = 1, \dots, Q; j_q = 0, \dots, J_q - 1 \quad [3]$$

where

$$maxSc = \sum_{q=1}^Q (J_q - 1)$$

Application to real data

MIR survey was created in order to monitor and evaluate the postgraduate medical education.

The aim: built an overall score that summarize the evaluation of the MIR program.

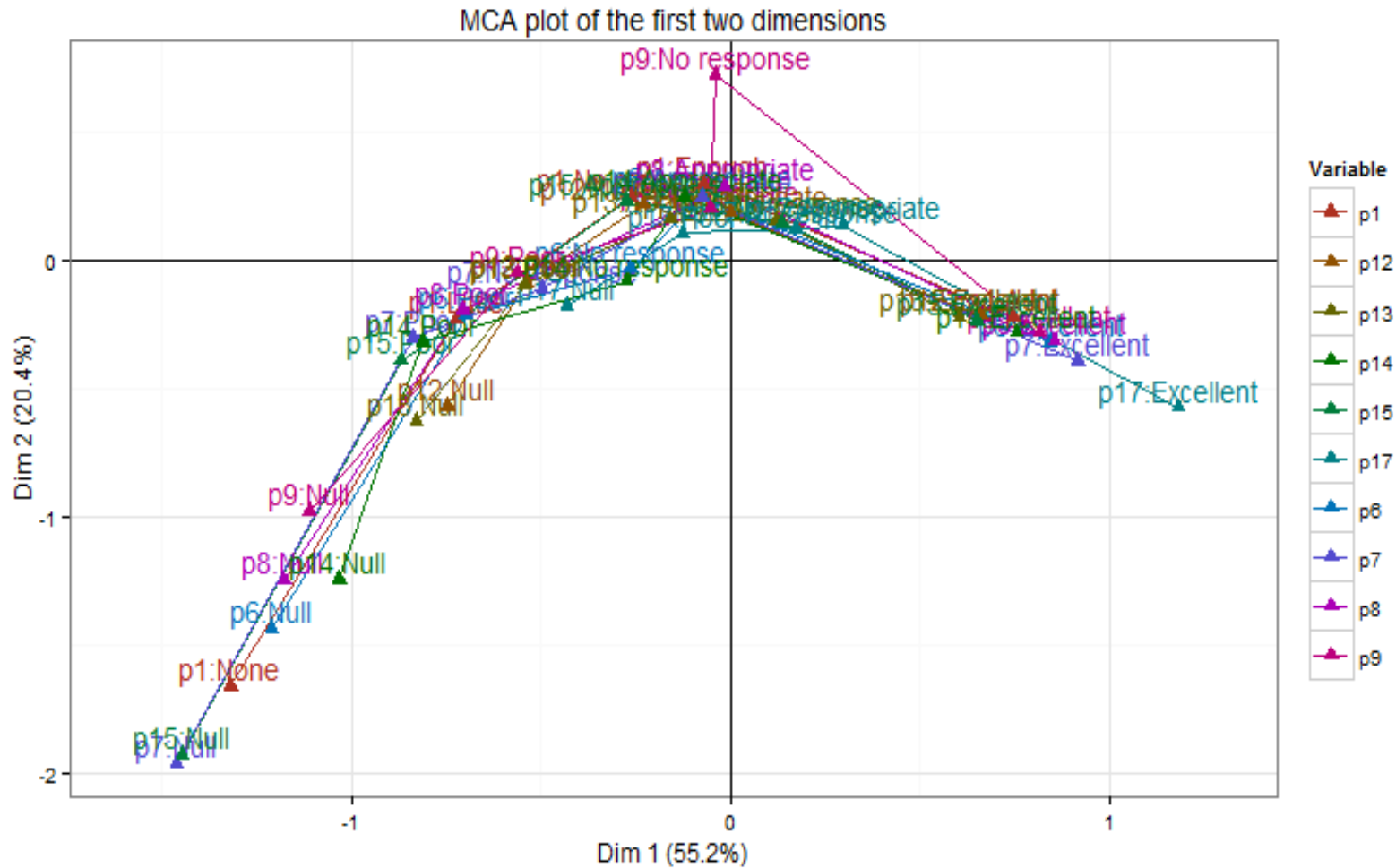
Data base:

- 849 MIR surveys
- 10 active variables in ordinal scale:
Issues related to service concern and supervision, tutor, research, formative program and sessions.
- 1 global assessment in ordinal scale

Application to real data

- $Q = 10$
- $J = \sum_{q=1}^Q J_q = 50$ where $J_q = 4 + 1$ (“no response”)
- Application of the methodology by the zscore function of the developed R package:
 - > `zscore(MIR, c(1:10), nr.consider=TRUE, nr.name="No response")`
 - `data` ≡ data base
 - `columns` ≡ where the active variables are
 - `nr.consider` ≡ if the MCA have to consider the no responses
 - `nr.name` ≡ the name in the data base for the no responses

Results



Results

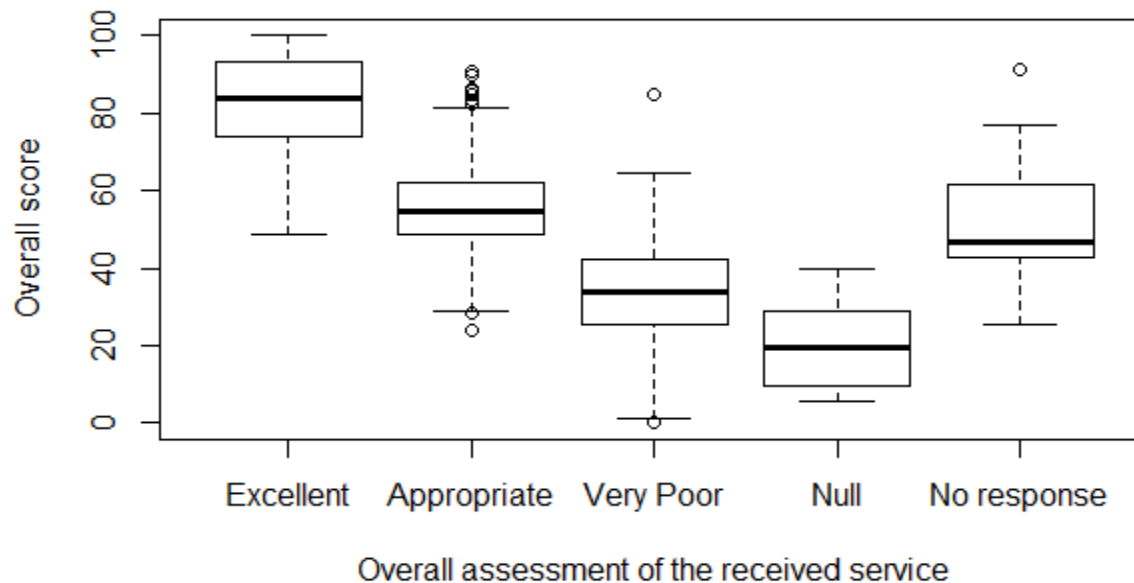
Items with their modalities and the weights associated.

Item	Levels	Coordinate	Weights	Item	Levels	Coordinate	Weights
1. Service concern for the formation of the MIR	A lot	0.765	10.99	12. Tutor responsibility about teaching	Excellent	0.667	7.48
	Enough	-0.069	6.64		Appropriate	0.001	3.97
	Little	-0.748	3.17		Poor	-0.543	1.1
	None	-1.385	0		Null	-0.738	0
	No response*	-0.208	5.63		No response*	-0.244	2.74
6. Supervision of the welfare work	Excellent	0.864	10.93	13. Satisfaction with the tutor	Excellent	0.609	7.63
	Appropriate	-0.074	6.04		Appropriate	-0.157	3.57
	Poor	-0.725	2.72		Poor	-0.536	1.52
	Null	-1.253	0		Null	-0.822	0
	No response*	-0.181	5.04		No response*	0.113	5.05
7. Evaluation of the welfare training	Excellent	0.948	12.65	14. Welfare supervision of the medical staff	Excellent	0.768	9.51
	Appropriate	-0.076	7.36		Appropriate	-0.123	4.85
	Poor	-0.876	3.32		Poor	-0.827	1.18
	Null	-1.511	0		Null	-1.024	0
	No response*	-0.507	5.13		No response*	-0.306	4.03
8. Evaluation of the teacher training	Excellent	0.879	10.82	15. Satisfaction with the medical staff	Excellent	0.664	11.16
	Appropriate	-0.015	6.19		Appropriate	-0.27	6.24
	Poor	-0.723	2.52		Poor	-0.884	3.07
	Null	-1.228	0		Null	-1.451	0
9. Evaluation of the ethical training	Excellent	0.838	10.25	17. Evaluation of the training received in research	Excellent	1.208	8.58
	Appropriate	-0.053	5.64		Appropriate	0.299	3.86
	Poor	-0.573	2.92		Poor	-0.124	1.62
	Null	-1.146	0		Null	-0.437	0
	No response*	-0.051	5.71		No response*	0.16	3.21

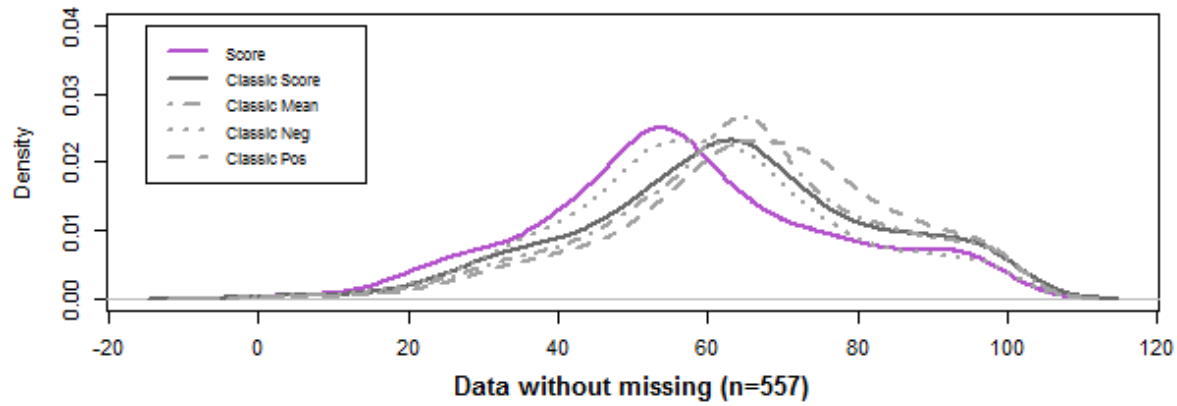
*No response categories are not considered to proof the order condition, but are for the score development.

Descriptive statistics of the constructed scores

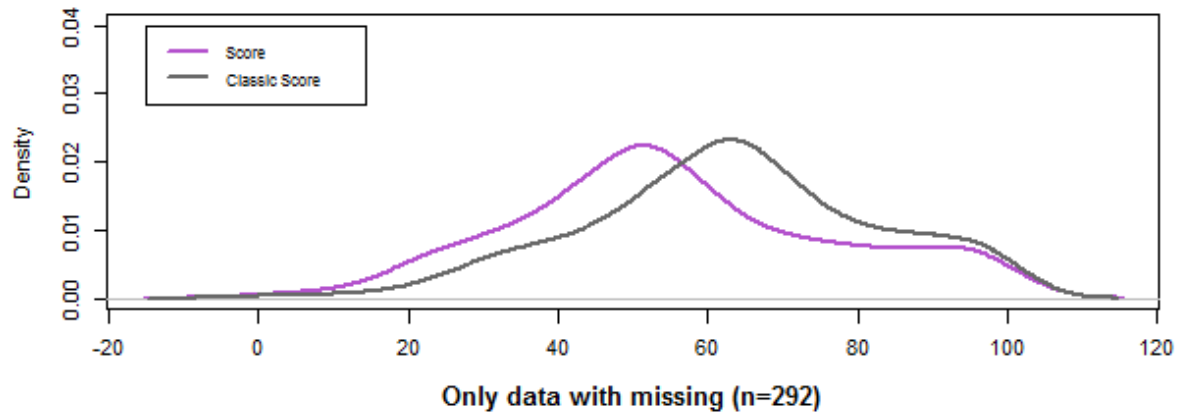
	N	Min	Max	Mean	SD	Q1	Median	Q3
Overall Score	849	0	100	57.2	19.6	45	55.6	69.3
POA: General assessment								
<i>Excellent</i>	184	48.8	100	82.2	12.6	74	84	93
<i>Appropriate</i>	493	23.8	90.7	55.5	11.2	48.6	54.7	61.9
<i>Very Poor</i>	139	0	85	34.7	12.7	25.7	33.7	42.2
<i>Null</i>	14	5.7	39.6	20	10.9	10.2	19.7	28.6
<i>No response</i>	19	25.4	91.4	52	17.3	42.6	46.8	61.6
Classic Score	557	0	100	63.1	19.4	50	63.3	76.7
Classic-mean	849	0	100	64.3	18.6	53.3	63.3	76.7
Classic-pos	849	0	100	66.7	18.2	56.7	66.7	80
Classic-neg	849	0	100	59.6	18.4	46.7	60	70



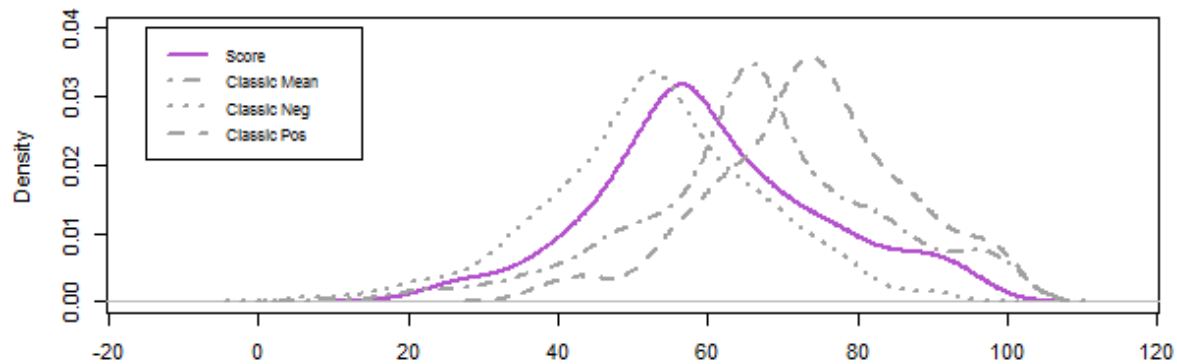
Complete Data (n=849)



Data without missing (n=557)



Only data with missing (n=292)



Conclusion

- ✓ The proposed methodology allows to construct a valid and effective overall score.
- ✓ It can be calculated for all the individuals regardless if it has any missing value.
- ✓ It has a good discriminatory ability, where the extreme values can be differentiated.
- ✓ This score could be used to determinate which factors are significant in the assessment of a questionnaire.
- ✓ The developed package in R gives the chance to implement this method by any person and in different fields.

Future work

- ❖ Generalize the methodology for questionnaires with two or more dimensions.
- ❖ Try to find a good approach when the survey has two types of variables: ordinals and continuous.
- ❖ Made more stable by applying bootstrap methodology.
- ❖ Implement all the advances in the R package.

Eskerrik asko!

Gràcies!

Gracias!