University of Ferrara

DIPARTIMENTO DI ECONOMIA E MANAGEMENT

Stefano Bonnini

Composite Indicators

Summary

- Procedure for the computation of a composite indicator (CI)
- Variable transformations
- Aggregation
- Final remarks

Procedure for the computation of a composite indicator (CI)

In the presence of a multivariate dataset with $k>1$ informative variables $X_1, X_2, \ldots, X_k$, we could be interested in summarize the information provided by these variables

Goal: quantify a complex phenomenon which cannot be directly measured (latent variable) by assuming that it can be broken down into k measurable components or dimensions or items

Example 1: composite index to quantify the student performance according to the examination marks, about 5 courses (Mechanics, Vectors, Algebra, Analysis, Statistics), for each of 88 students

Example 2: composite index to quantify the satisfaction of 386 wine drinkers toward Passito wine, according to how much they like Passito itself, aroma and smell, sweetness, alcohol content, intensity of taste and according to how much they can pay for one bottle of Passito

Procedure for the computation of a composite indicator (CI)

The procedure for computing a CI consists in two steps:

a) Transformation of the k informative variables to allow comparability: $T_j(X_j)$, $j=1,2,\ldots,k$

b) Aggregation (synthesis) with a suitable function

$$\Psi[T_1(X_1),\ldots,\ T_k(X_k);w_1,\ldots,w_k\ ]$$

with $w_1,\ldots,w_k$ suitable weights which represent the degrees of importance of the k informative variables

Procedure for the computation of a composite indicator (CI)

Typical methods for detecting latent variables: Principal Component Analysis (PCA) and Factor Analysis (FA)

In the case of PCA e FA the number *q* of latent variables (components or factors) can be greater than one, instead in the case of CI we are dealing with only one latent variable

With PCA and FA we use:

a) Variable transformation: standardization

b) Aggregation: linear combination of the *k* informative variables (additive method) according to the correlations or maximum likelihood methods

Step (b) assumes that the dependence between the informative variables can be represented only by the correlations or assumes that the informative variables follow a specific multivariate distribution

Procedure for the computation of a composite indicator (CI)

The starting point of a procedure for computing a CI is the dataset, that is the $n \times k$ matrix of raw data, with *n*=number of units and *k*=number of variables

Units	Variables				
	X_1	…	X_v	…	X_k
1	x_{11}	…	x_{1v}	…	x_{1k}
…	…	…	…	…	…
u	x_{u1}	…	x_{uv}	…	x_{uk}
…	…	…	…	…	…
n	x_{n1}	…	x_{nv}	…	x_{nk}

After variable transformations we have new data: $z_{uv} = T_v(x_{uv})$

Summary

- Procedure for the computation of a composite indicator (CI)
- Variable transformations
- Aggregation
- Final remarks

Variable transformations

<u>Classic standardization</u>:

$$z_{uv} = \frac{x_{uv} - M(X_v)}{\sqrt{Var(X_v)}} = \frac{x_{uv} - \bar{x}_v}{s_v}$$

$$\bar{x}_v = \frac{\sum_{u=1}^{n} x_{uv}}{n}$$

$$s_v = \sqrt{\frac{\sum_{u=1}^{n}(x_{uv} - \bar{x}_v)^2}{n-1}}$$

- $x_{uv} \in (-\infty, +\infty) \Rightarrow z_{uv} \in (-\infty, +\infty)$
- $M(Z_v) = \bar{z}_v = 0$ and $Var(Z_v) = \sqrt{Var(Z_v)} = 1$

Variable transformations

<u>Comparison with the maximum (for nonnegative variables)</u>:

$$z_{uv} = \frac{x_{uv}}{\max(x_{1v}, x_{2v}, \cdots, x_{nv})}$$

$$0 \leq \frac{min(x_{1v}, \cdots, x_{nv})}{max(x_{1v}, \cdots, x_{nv})} \leq z_{uv} \leq 1$$

Variable transformations

Comparison with the minimum (for nonnegative variables):

$$z_{uv} = 1 - \frac{\min(x_{1v}, x_{2v}, \cdots, x_{nv})}{x_{uv}}$$

$$0 \leq z_{uv} \leq \frac{max(x_{1v}, \cdots, x_{nv}) - min(x_{1v}, \cdots, x_{nv})}{max(x_{1v}, \cdots, x_{nv})} \leq 1$$

This is a nonlinear transformation

Variable transformations

<u>Re-scaling</u>:

$$z_{uv} = \frac{x_{uv} - \min(x_{1v}, x_{2v}, \cdots, x_{nv})}{\max(x_{1v}, x_{2v}, \cdots, x_{nv}) - \min(x_{1v}, x_{2v}, \cdots, x_{nv})}$$

$$0 \leq z_{uv} \leq 1$$

Variable transformations

Rank transformation:

$$z_{uv} = \sum_{\substack{i=1 \\ i \neq u}}^{n} I_{(-\infty, x_{uv}]}(x_{iv})$$

$$1 \leq z_{uv} \leq n$$

- With this method the effect of possible outliers or asymmetry is eliminated
- This method is suitable for categorical variables
- Normalized ranks can be obtained by combining rank and re-scaling transformations: $\tilde{z}_{uv} = (z_{uv} - 1)/(n - 1)$, $0 \leq \tilde{z}_{uv} \leq 1$

Summary

- Procedure for the computation of a composite indicator (CI)
- Variable transformations
- Aggregation
- Final remarks

Aggregation

- Choice of a suitable aggregation function $\Psi: \Re^{2k} \rightarrow \Re$ to reduce the dimension of variables from k to 1:

$$y_u = \Psi[z_{u1}, z_{u2}, \ldots, z_{uk}; w_1, w_2, \ldots, w_k]$$

- Usually normalized weights are considered: $0 \leq w_v \leq 1$, $w_1 + \ldots + w_k = 1$

- Additive method: $y_u = w_1 z_{u1} + w_2 z_{u2} + \ldots + w_k z_{uk} = \sum_v w_v z_{uv}$

- Multiplicative method: $y_u = (z_{u1})^{w_1} \times (z_{u2})^{w_2} \times \cdots \times (z_{uk})^{w_k} = \prod_v (z_{uv})^{w_v}$

Aggregation

<u>NonParametric Combination (NPC) methodology</u>

- It is not based on correlations or distributional assumptions between variables
- Variable transformations into the open interval *(0,1). E.g.* rescaling-type transformation:

$$z_{uv} = \frac{x_{uv} - \min(x_{uv}) + c_1}{\max(x_{uv}) - \min(x_{uv}) + c_2}$$

with $0 < c_1 < c_2$

- Assumptions:

1. $X_1, \ldots, X_k$ are nondegenerate variables which can be transformed into the interval *(0,1)*
2. Transformed variables $Z_1, \ldots, Z_k$ have nondecreasing monotonic relationship
3. The relationship between Z_v and Y must be known and coherent for each *v*

Aggregation

<u>NonParametric Combination (NPC) methodology</u>

Function $\Psi(\cdot)$ must satisfy the following properties:

1. $\Psi(\cdot)$ is a continuous function respect to all its arguments (original informative variables and weights)

2. $\Psi(\cdot)$ is a nondecreasing function of the transformed variables $Z_1,\ldots,Z_k$:

   if $0 < z_{uv} < z'_{uv} < 1$ then $y_u = \Psi(\cdots, z_{uv}, \cdots) \leq y'_u = \Psi(\cdots, z'_{uv}, \cdots)$

3. $\Psi(\cdot)$ is invariant respect to permutations of the couples $(z_{u1}, w_1), \cdots, (z_{uk}, w_k)$

Aggregation

NonParametric Combination (NPC) methodology

Some suitable combining functions:

1. FISHER: $y_u = -\sum_{v=1}^{k} w_v \cdot \log(1 - z_{uv})$
2. *LIPTAK:* $y_u = \sum_{v=1}^{k} w_v \cdot \varphi^{-1}(z_{uv})$ with $\varphi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt$
3. LOGISTIC: $y_u = \sum_{v=1}^{k} w_v \cdot \log\left(\frac{z_{uv}}{1-z_{uv}}\right)$
4. TIPPETT: $y_u = \max\ (w_1 z_{u1}, w_2 z_{u2}, \cdots, w_k z_{uk})$
5. ADDITIVE: $y_u = \sum_{v=1}^{k} w_v \cdot z_{uv}$

Summary

- Procedure for the computation of a composite indicator (CI)
- Variable transformations
- Aggregation
- Final remarks

Final remarks

…on the combining functions:

1. TIPPETT: it rewards the units which excell in one aspect (respect to one variable) even if very poor in all the others
2. *LIPTAK:* it rewards the units with excellent results respect to several aspects but it penalizes the units with poor results respect to some aspects
3. FISHER: it rewards the units with excellent results respect to one or few aspects and does not eccessively penalize the units with poor results respect to some aspects

4. Combining functions can be used to combine significance levels and p-values in the case of multiple or multivariate tests

Final remarks

- A transformation is linear when it can be expressed by the equation $Y = a + bX$. Hence the following transformations are linear:
  - ☐ Standardization: $a = -\bar{x}_v / s_v \qquad b = 1/s_v$
  - ☐ Comparison with max: $a = 0 \qquad b = 1/\max_u(x_{uv})$
  - ☐ Rescaling: $a = -\min_u(x_{uv}) \,/\, \left[\max_u(x_{uv}) - \min_u(x_{uv})\right]$
    $b = 1/\left[\max_u(x_{uv}) - \min_u(x_{uv})\right]$

Final remarks

- A final normalization of the values of the composite indicator is useful for the interpretation of the results. E.g.

$$\tilde{y}_u = T(y_u) = \frac{y_u - \min_u(y_{uv})}{\max_u(y_{uv}) - \min_u(y_{uv})}$$

R exercises

Problem 1 - Passito

- Compute a composite indicator to quantify the satisfaction of 386 wine drinkers toward Passito wine, according to how much they like Passito itself (LIKE_PAS), aroma and smell (LIKE_AROMA), sweetness (LIKE_SWEET), alcohol content (LIKE_ALCOHOL), intensity of taste (LIKE_TASTE) and according to how much they can pay for one bottle of Passito (PRICE)

R exercises

Problem 2 - Mall

- Compute a composite index of satisfaction of 29 customers according to their satisfaction for the 5 evaluated partial aspects of the shopping center in the customer satisfaction survey

R exercises

Problem 3 - Students

- Compute a composite indicator to quantify the global evaluation of the three university programs by the 20 interviewed students