

ACP: Centre de gravité-Inertie- Indicateurs de qualité

Prof. Mohamed El Merouani

1

Point moyen ou centre de gravité:

- On appelle le point moyen ou centre de gravité le vecteur G des moyennes arithmétiques de chaque variable:

$$G = (\bar{x}_1, \bar{x}_2, \bar{x}_3, \dots, \bar{x}_q)$$

- Lorsqu'on analyse des variables centrées, ce point moyen G sera le centre du repère considéré:

$$G \equiv O$$

Prof. Mohamed El Merouani

2

Inertie totale du nuage de points:

- On appelle inertie totale du nuage de points, et on la note Ig , la moyenne pondérée des carrés des distances des points au centre de gravité.
- Quand $G \equiv O$ et tous les points ont même poids, Ig est la somme des carrés des distances des points au centre de repère.

$$Ig = tr(Y'Y) = tr(X'X) = tr(\Lambda)$$

$$Ig = \sum_{i=1}^q \lambda_i = \sum_{i=1}^q Var(x_i)$$

3

Qualité des représentations sur les axes ou plans principaux:

- Le but de l'ACP étant d'obtenir une représentation des individus dans un espace de dimension plus faible que p , la question se pose d'apprécier la perte d'information subie et de savoir combien de facteurs retenir:
- Indicateurs de la qualité globale.
- Indicateurs de la qualité locale.

Indicateurs de la qualité globale:

- Inertie expliquée:

$$Ie = \lambda_1 + \lambda_2 + \dots + \lambda_s = \sum_{i=1}^s \lambda_i$$

- Indicateur de qualité globale:

$$Iq = \frac{\sum_{i=1}^s \lambda_i}{\sum_{i=1}^q \lambda_i} = \frac{Ie}{Ig}$$

5

Indicateurs de la qualité globale:

- On a: $0 \leq Iq \leq 1$

- Si, par exemple, $\frac{\lambda_1 + \lambda_2}{Ig} = 0,9$ on conçoit

clairement que le nuage de points est presque aplati sur un sous-espace à deux dimensions et qu'une représentation du nuage dans le plan des deux premiers axes principaux sera très satisfaisante.

Prof. Mohamed El Merouani

6

Indicateurs de la qualité locale:

- Contributions à l'inertie totale de l'individu i:

$$C_{ITi} = \frac{\lambda_i}{I_g}$$

– Les individus qui ont un CIT plus grand sont ceux qui sont les plus éloignés du G.

- Contributions à l'inertie expliquée de l'individu i:

$$C_{IEi} = \frac{\lambda_i}{I_e}$$

- En général, elles sont présentées sous la forme d'un pourcentage.

7

Point moyen ou centre de gravité:

- On appelle le point moyen ou centre de gravité le vecteur G des moyennes arithmétiques de chaque variable:

$$G = (\bar{x}_1, \bar{x}_2, \bar{x}_3, \dots, \bar{x}_q)$$

- Lorsqu'on analyse des variables centrées, ce point moyen G sera le centre du repère considéré:

$$G \equiv O$$

Prof. Mohamed El Merouani

8

Inertie totale du nuage de points:

- On appelle inertie totale du nuage de points, et on la note Ig , la moyenne pondérée des carrés des distances des points au centre de gravité.
- Quand $G \equiv O$ et tous les points ont même poids, Ig est la somme des carrés des distances des points au centre de repère.

$$Ig = tr(Y'Y) = tr(X'X) = tr(\Lambda)$$

$$Ig = \sum_{i=1}^q \lambda_i = \sum_{i=1}^q Var(x_i)$$

9

Qualité des représentations sur les axes ou plans principaux:

- Le but de l'ACP étant d'obtenir une représentation des individus dans un espace de dimension plus faible que p , la question se pose d'apprécier la perte d'information subie et de savoir combien de facteurs retenir:
- Indicateurs de la qualité globale.
- Indicateurs de la qualité locale.

Indicateurs de la qualité globale:

- Inertie expliquée:

$$Ie = \lambda_1 + \lambda_2 + \dots + \lambda_s = \sum_{i=1}^s \lambda_i$$

- Indicateur de qualité globale:

$$Iq = \frac{\sum_{i=1}^s \lambda_i}{\sum_{i=1}^q \lambda_i} = \frac{Ie}{Ig}$$

11

Indicateurs de la qualité globale:

- On a: $0 \leq Iq \leq 1$

- Si, par exemple, $\frac{\lambda_1 + \lambda_2}{Ig} = 0,9$ on conçoit

clairement que le nuage de points est presque aplati sur un sous-espace à deux dimensions et qu'une représentation du nuage dans le plan des deux premiers axes principaux sera très satisfaisante.

Prof. Mohamed El Merouani

12

Indicateurs de la qualité locale:

- Contributions à l'inertie totale de l'individu i:

$$C_{ITi} = \frac{\lambda_i}{I_g}$$

– Les individus qui ont un CIT plus grand sont ceux qui sont les plus éloignés du G.

- Contributions à l'inertie expliquée de l'individu i:

$$C_{IEi} = \frac{\lambda_i}{I_e}$$

- En général, elles sont présentées sous la forme d'un pourcentage.