

Exemplification méthodologique d'une analyse de classes latentes avec R

Agathe Bellemare-Lepage ^a , Marion Chatelois ^a & Pier-Olivier Caron ^b

^aUniversité du Québec à Montréal

^bUniversité TÉLUQ

Abstract ■ L'analyse de classes latentes (ACL) permet de partager et de distinguer des sous-groupes non observables (latents) d'individus sur la base de leurs réponses à un ensemble d'indicateurs observables (manifestes). Cette analyse permet de mieux comprendre la variabilité au sein d'une population. Or, il existe peu de documentation, surtout en français, sur la procédure à suivre pour réaliser une ACL sur la plateforme R. Ce logiciel statistique est accessible gratuitement et comporte de nombreux avantages en ce qui a trait à la programmation d'analyses, à la visualisation des données ainsi qu'à la gestion des variables et de l'environnement de travail. L'objectif du présent article est d'exemplifier la réalisation d'une ACL sur la plateforme R avec le package poLCA. Après une introduction sur l'origine et les principes de l'ACL, un tutoriel sur la réalisation d'une ACL avec R est présenté. Une situation hypothétique portant sur la perpétration de violence dans les relations amoureuses à l'adolescence est utilisée. La syntaxe R permettant de réaliser cette analyse est fournie et explicitée en détails. Dans une visée de partage des connaissances, similaire à la philosophie de R, cet article peut servir de guide pour tout étudiant ou chercheur voulant développer sa compréhension de l'ACL et ses compétences en statistiques sur cette plateforme.

Keywords ■ classes latentes, package poLCA, modélisation. **Tools** ■ R.

bellemare-lepage.agathe@courrier.uqam.ca

[10.20982/tqmp.19.2.p217](https://doi.org/10.20982/tqmp.19.2.p217)

Acting Editor ■ Denis Cousineau (Université d'Ottawa)

Ce texte a reçu le prix SQRP-TQMP remis lors de la conférence annuelle de la Société québécoise pour la recherche en psychologie en 2023.

Introduction

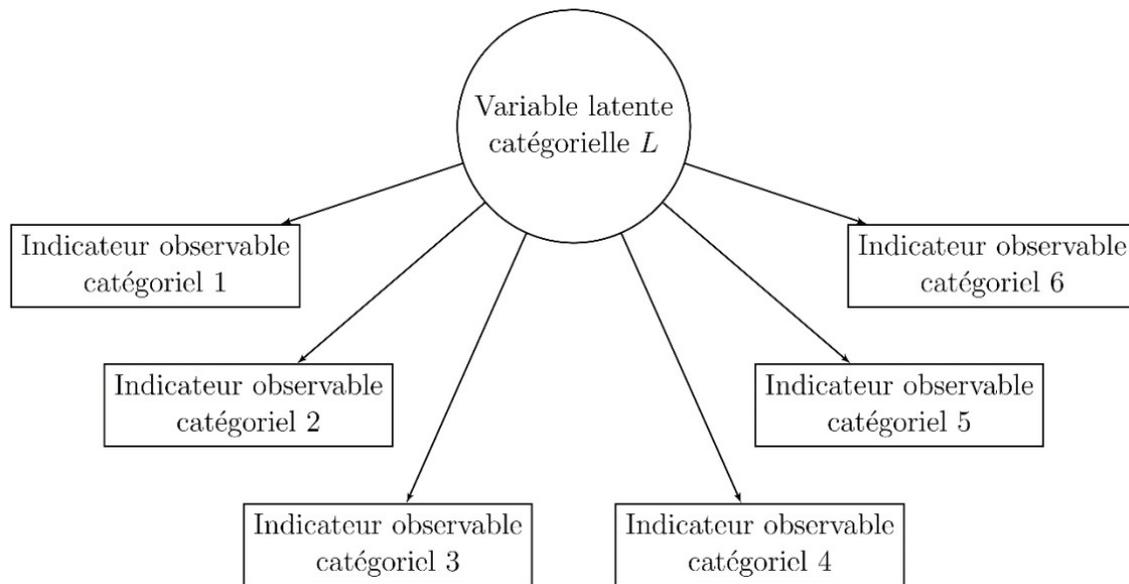
Dégager des variables latentes, c'est-à-dire des variables non observables, à partir du regroupement de plusieurs variables observables liées entre elles est pratique courante en psychologie (Laudy et collègues, 2005; Petersen et collègues, 2019). Ce type d'analyse permet d'étudier un phénomène complexe de même qu'identifier l'hétérogénéité au sein d'une population (Karnowski, 2017). À cette fin, l'analyse de classes latentes (ACL) permet d'inférer divers regroupements d'individus similaires sur certaines caractéristiques à partir de corrélations entre des variables observables et mesurables.

L'ACL est utilisée dans plusieurs domaines en sciences humaines. Par exemple, en sciences politiques, cette analyse permet d'identifier différents types de participants politiques (Johann et collègues, 2020), tandis qu'en sociologie, des chercheurs l'utilisent pour identifier des sous-groupes d'individus ayant différents vécus de discrimination (Karnowski, 2017). Dans le domaine de la recherche

en psychologie, l'ACL permet d'identifier des sous-groupes de personnes vivant une dépression selon leurs symptômes (Ulbricht et collègues, 2018), ou de perpétrateurs de certains types d'agression (p. ex., intimidation; Chung & Lee, 2020). Identifier l'hétérogénéité qui distingue des groupes a des retombées substantielles, notamment pour l'application de programmes de prévention (Nylund-Gibson & Hart, 2014). Certains programmes peuvent avoir des effets différentiels selon l'appartenance des individus à divers sous-groupes (P. O. Caron, 2019). Identifier ces sous-groupes permet alors la personnalisation des interventions selon les besoins et les enjeux des individus concernés, afin accroître les chances d'observer les effets escomptés (Butters et collègues, 2021).

Cet article a comme objectifs de faciliter la compréhension de l'ACL et d'illustrer son utilisation sur la plateforme R avec le package poLCA. Pour ce faire, l'article est divisé en deux sections. La première section est théorique et permet d'expliquer les fondements de l'ACL. La deuxième section vise à mieux illustrer comment uti-

Figure 1 ■ Variable latente catégorielle et indicateurs catégoriels



liser l'ACL sur la plateforme R à l'aide d'une situation hypothétique.

Fondements théoriques

Origine des structures latentes

Historiquement, l'origine de l'ACL est attribuée à Paul Lazarsfeld (1950). Ce sociologue américain cherche à classer les individus en plusieurs sous-groupes selon divers concepts abstraits, comme l'opinion publique ou encore le *Zeitgeist*, qui désigne les grandes lignes de la pensée d'une époque spécifique (Karnowski, 2017; Porcu & Giambona, 2017). Pour ce faire, il utilise leurs réponses à un ensemble de variables observables et dichotomiques, c'est-à-dire des variables qui ne peuvent prendre que deux valeurs. Cette idée est par la suite révisée et étouffée par de nombreux auteurs (p. ex., Goodman, 1974; Vermunt, 1997).

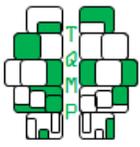
L'ACL est une analyse de structure latente. L'objectif d'une analyse de structure latente est d'interpréter les relations entre différentes variables observables, en considérant celles-ci comme divers indicateurs d'un même concept, soit d'une variable latente (Karnowski, 2017). Cette variable latente peut être identifiée à l'aide de différentes procédures, les plus communes étant l'analyse factorielle, l'analyse de profils latents, l'analyse de traits latents, ainsi que l'ACL. En comparaison à l'ACL, l'analyse factorielle permet d'identifier une variable latente continue à l'aide d'un ensemble de variables observables continues; l'analyse de profils latents utilise également des va-

riables observables continues, mais dans le but d'identifier une variable latente catégorielle; l'analyse de traits latents permet d'identifier une variable latente continue à partir de variables observables catégorielles. Enfin, l'analyse qui intéresse cet article, soit l'ACL, permet de faire ressortir une variable latente catégorielle à l'aide de variables observables également catégorielles (Karnowski, 2017). La Figure 1 illustre la relation entre la variable latente catégorielle étudiée et les indicateurs permettant de l'inférer.

Qu'est-ce que l'analyse de classes latentes

Une bonne façon de conceptualiser l'ACL est de la définir comme une approche « centrée sur la personne », dont l'intérêt est de regrouper des individus. Cette approche se distingue des analyses « centrées sur les variables » telles que l'analyse factorielle, qui concerne plutôt les relations entre des variables permettant de regrouper des items (Nylund-Gibson & Choi, 2018; Porcu & Giambona, 2017). L'ACL est utile pour étudier l'hétérogénéité au sein d'une population d'individus en créant des sous-groupes latents, c'est-à-dire des sous-groupes qui ne peuvent pas être directement observés ni mesurés (Porcu & Giambona, 2017). Préalablement inconnus, ces sous-groupes ou classes d'individus seront créés par l'entremise de corrélations entre des variables mesurables et observables, qualifiées d'indicateurs (Bartholomew et collègues, 2011).

Il est possible d'estimer la probabilité d'appartenance des individus aux classes identifiées (Collins & Lanza,



2010). Cela résulte en des classes mutuellement exclusives et exhaustives (Nylund-Gibson & Choi, 2018), au sein desquelles les individus sont homogènes entre eux sur le plan de leurs réponses à des questionnaires, et distincts des individus appartenant aux autres classes (Collins & Lanza, 2010). Cette analyse permet également d'estimer la probabilité qu'un individu endosse un certain item sur la base de la classe à laquelle il appartient (Porcu & Giambona, 2017).

Le modèle d'analyse de classes latentes

Postulat : l'indépendance locale. Le postulat théorique sous-jacent à l'ACL est l'indépendance locale. Suivant ce postulat, il est supposé que les indicateurs observables soient indépendants les uns des autres, conditionnellement à leur appartenance aux classes latentes (Porcu & Giambona, 2017). Ce postulat est aussi nommé « indépendance conditionnelle », la condition étant que l'appartenance à une classe explique toute la variance partagée entre les indicateurs (Nylund-Gibson & Choi, 2018). En d'autres termes, les interrelations entre les indicateurs observables ne pouvant pas être attribuées à la chance sont expliquées par la variable latente (Karnowski, 2017). Une fois la variable de classe latente modélisée, les indicateurs au sein de chaque classe latente sont indépendants les uns des autres (Karnowski, 2017; Nylund-Gibson & Choi, 2018). **Les paramètres du modèle.** Un modèle d'ACL sans covariables contient deux types de paramètres importants. Premièrement, les paramètres de proportion de classe désignent la proportion d'individus appartenant à une classe latente particulière, ou la taille relative de cette classe (Wurpts & Geiser, 2014). Représentons la variable latente catégorielle par L , et le nombre de classes latentes de L par C . La taille relative d'une classe c est notée $P(L = c)$. Tel qu'indiqué plus haut, le modèle d'ACL implique que les classes latentes soient mutuellement exclusives et exhaustives. Ainsi, chaque individu est membre d'une seule classe et la somme des tailles relatives des classes latentes équivaut à 1, soit $\sum P(L = c) = 1$ (Nylund-Gibson & Choi, 2018).

Deuxièmement, les paramètres d'items permettent de déterminer la probabilité qu'un individu endosse la catégorie spécifique d'un indicateur, considérant l'appartenance à une certaine classe (Wurpts & Geiser, 2014). Ces probabilités conditionnelles de réponse sont spécifiques à chaque item et à chaque classe, et sont utilisées pour interpréter les classes en termes du profil de réponse aux items. Elles décrivent les relations entre les indicateurs observables et la variable catégorielle latente (Nylund-Gibson & Choi, 2018). La probabilité conditionnelle de choisir la réponse y_j pour l'item j considérant l'appartenance à la classe c peut être représentée ainsi :

$$P(Y_j = y_j | L = c).$$

Afin d'illustrer l'ACL sous forme d'équation mathématique, le modèle d'ACL complet pour un schéma de réponses observées y parmi un ensemble de schémas de réponses Y est représenté par (Wurpts & Geiser, 2014) :

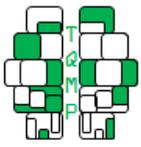
$$P(Y = y) = \sum_{c=1}^C P(L = c) \prod_{j=1}^J P(Y_j = y_j | L = c)$$

Considérations

Deux éléments importants à considérer avant de réaliser une ACL sont le choix des indicateurs mesurés et la taille de l'échantillon. Les indicateurs inclus dans une ACL sont choisis en fonction de la question de recherche et doivent être susceptibles de pouvoir distinguer les individus issus de la population hétérogène (Sinha et collègues, 2021). Il est recommandé d'utiliser au moins trois indicateurs (P. O. Caron, 2019). Utiliser un plus grand nombre d'indicateurs est généralement bénéfique, puisque cela augmente la précision de l'attribution des classes (Wurpts & Geiser, 2014). Or, un trop grand nombre d'indicateurs, surtout lorsque fortement corrélés entre eux, peut poser un problème en ce qui concerne le respect de l'indépendance locale, les indicateurs devant être indépendants entre eux une fois les classes identifiées. En ce qui concerne l'échantillon, une façon d'en estimer la taille nécessaire est de s'assurer d'avoir de 10 à 20 participants par indicateur (P. O. Caron, 2019). Il est généralement suggéré qu'une taille d'échantillon se situant entre 300 et 1000 participants est adéquate (Nylund-Gibson & Choi, 2018). Une taille d'échantillon trop petite peut mener à des enjeux de convergence, à des solutions instables ainsi qu'à une puissance insuffisante ne permettant pas de détecter des classes ou masquant les classes contenant moins de participants (Nylund-Gibson & Choi, 2018).

Le déroulement typique de l'analyse

Le déroulement d'une ACL consiste à tester séquentiellement des modèles contenant un nombre croissant de classes latentes afin de sélectionner la quantité permettant le meilleur ajustement du modèle. Pour commencer, il faut tester un modèle avec une seule classe, auquel est comparé un modèle à deux classes et ainsi de suite. Chaque nouveau modèle contenant une classe supplémentaire est comparé au précédent, soit au modèle contenant une classe de moins. Il est recommandé de cesser d'estimer des classes additionnelles lorsque les indices d'ajustement ne montrent plus d'amélioration du modèle, soit lorsque le dernier modèle testé n'est pas significativement préférable au modèle précédent. Il faut alors récolter les informations concernant l'ajustement des différents modèles afin de trouver celui qui décrit le mieux ce qui peut être observé dans les données empiriques (Nylund-



Gibson & Choi, 2018). Parmi les indices qui peuvent être utilisés pour guider ce choix, les critères d'information Akaike (AIC) et bayésien (BIC), le *maximum log-likelihood*, l'entropie relative et le test de Lo-Mendell-Rubin sont particulièrement utiles et fréquemment utilisés.

Pour l'AIC comme pour le BIC, une valeur plus petite indique un meilleur ajustement du modèle (Nylund-Gibson & Choi, 2018). Le BIC est souvent considéré comme l'indice d'ajustement le plus fiable (Nylund et collègues, 2007; Vermunt, 2002), permettant de cerner un modèle plus parcimonieux (Weller et collègues, 2020). Le package *poLCA* fournit également le *maximum log-likelihood*, qui désigne la vraisemblance du modèle. Une valeur absolue plus petite du *maximum log-likelihood* indique un modèle qui est mieux ajusté aux données observées (Linzer & Lewis, 2011).

L'entropie relative est une statistique diagnostique importante à considérer afin de vérifier à quel point les classes d'un modèle se distinguent l'une de l'autre (Weller et collègues, 2020). Une valeur d'entropie relative plus élevée, supérieure à .8, est souhaitée (Clark & Muthén, 2009, dans Nylund-Gibson & Choi, 2018). Un élément supplémentaire à considérer lors de l'analyse des résultats de différents modèles d'ACL est le test de ratio de vraisemblance Lo-Mendell-Rubin (LMR; Lo et collègues, 2001). Ce test permet de comparer des paires de modèles qui diffèrent d'une classe, afin de déterminer le modèle le plus vraisemblable.

La méthode présentée ici est exploratoire. Cette approche implique que le domaine de recherche étudié ne comporte pas de théories suffisamment développées pour permettre la formulation d'hypothèses précises au sujet des sous-groupes latents attendus (Hoiijtink, 2001; Laudy et collègues, 2005). Ainsi, le chercheur doit vérifier si les conclusions d'une analyse exploratoire suivent les théories existantes ou suggèrent l'émergence de nouvelles hypothèses (Laudy et collègues, 2005), en évitant le risque de surinterprétation et de suppositions (Hoiijtink, 2001). Bien que la plupart des ACL soient réalisées suivant cette approche (Finch & Bronk, 2011), réaliser une ACL confirmatoire est aussi possible. Le chercheur, à l'aide de différentes restrictions de paramètres, est en mesure de tester des hypothèses fondées sur une base théorique soutenue concernant le nombre et les caractéristiques des groupes latents (Finch & Bronk, 2011). Il incombe au chercheur de sélectionner l'approche la plus adéquate selon ses objectifs de recherche.

Exemplification

La prochaine section permet d'illustrer le contexte dans lequel les chercheurs peuvent être intéressés à recourir à l'ACL et la manière de procéder sur la plateforme R avec

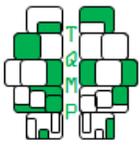
le package *poLCA* (Linzer & Lewis, 2011). Une situation hypothétique est présentée afin d'exemplifier cette procédure concrètement. Tout au long de cette démonstration, le script nécessaire pour reproduire les analyses exemplifiées est fourni à travers le texte. Le script complet se trouve en Listing 1 à la fin de l'article. Un jeu de données fictif adapté aux paramètres de l'étude hypothétique présentée a été créé aux fins de cette démonstration. Le script permettant de reproduire ce jeu de données est également fourni en Listing 2. Celui-ci peut être utilisé par le lecteur pour se pratiquer à effectuer des ACL sur R.

Situation hypothétique

Sophie, une étudiante au doctorat en psychologie, s'intéresse à la perpétration de violence dans les relations amoureuses chez les adolescents. L'hétérogénéité des perpétrateurs de ce type de violence est de plus en plus reconnue par les spécialistes du domaine (Adams et collègues, 2021; Reidy et collègues, 2016; Théorêt et collègues, 2021). La violence dans les relations amoureuses est définie comme l'utilisation de violence physique, psychologique et/ou sexuelle dans le contexte d'une relation intime (Anderson & Danis, 2007). La violence physique réfère à l'intention de faire mal à son partenaire, entre autres, en lui assénant des coups de pied ou en le frappant (Center for Disease Control and Prevention; CDC, 2017). La violence psychologique est définie comme l'intention de faire mal à son partenaire émotionnellement, soit avec des mots, des agissements ou en essayant de le ou la contrôler (CDC, 2017). Finalement, la violence sexuelle correspond à forcer un partenaire à poser des actes sexuels sans son consentement (CDC, 2017).

En outre, un partenaire peut perpétrer de la violence dans les relations amoureuses en réaction à une provocation perçue, ce qui se nomme de la violence réactive (Bushman & Anderson, 2001). À l'inverse, la violence proactive désigne des comportements violents perpétrés de façon plus stratégique, dans le but d'arriver à ses fins (Ramirez & Andreu, 2006). La violence dans les relations amoureuses (physique, psychologique et sexuelle) peut ainsi être perpétrée de façon réactive et/ou proactive. Plusieurs chercheurs identifient diverses classes de perpétrateurs qui se différencient sur la base des formes de violence utilisées (Adams et collègues, 2021; Reidy et collègues, 2016; Théorêt et collègues, 2021), sans nécessairement distinguer si cette violence est utilisée de façon proactive et/ou réactive.

À la suite de sa lecture de la littérature scientifique à ce sujet, Sophie se questionne à savoir s'il existe plusieurs classes de perpétrateurs de violence dans les relations amoureuses qui se différencient sur la base des formes de violence (physique, psychologique ou sexuelle)



utilisées de façon réactive et/ou proactive. L'étudiante détermine que l'ACL est l'analyse idéale pour réaliser son projet puisqu'elle s'intéresse à une population hétérogène. L'analyse lui permet d'identifier différentes classes de perpétrateurs en regroupant des individus qui se ressemblent sur la base de leurs réponses aux questionnaires administrés. L'étudiante souhaite étudier la répartition des perpétrateurs telle qu'elle se présente « naturellement », et ensuite proposer un modèle statistique afin de refléter cette répartition. Les classes de perpétrateurs sont des classes dites latentes puisqu'elles ne peuvent être qu'inférées par le biais d'indicateurs mesurables et observables.

L'objectif de l'étudiante est de déterminer le nombre et la structure des différentes classes latentes de perpétration de violence dans les relations amoureuses, chez les adolescents âgés de 14 à 17 ans, selon les formes de violence utilisées et la façon dont elles sont émises (réactive et/ou proactive). Elle a recours à un échantillon mixte de 800 adolescents québécois. Cette taille d'échantillon est adéquate, puisqu'elle est supérieure à la taille de 300 participants recommandée par Nylund-Gibson et Choi (2018). Sophie utilise six indicateurs observables, soit la violence psychologique réactive, la violence psychologique proactive, la violence sexuelle réactive, la violence sexuelle proactive, la violence physique réactive et la violence physique proactive. Selon les écrits dans son domaine, ces indicateurs peuvent distinguer les diverses classes. Dans une population non clinique, la violence dans les relations amoureuses se présente souvent sous forme d'une distribution de Poisson, soit une majorité d'individus n'en perpétrant peu ou pas et une petite proportion en perpétrant beaucoup. Suivant la procédure adoptée par la majorité des chercheurs qui s'intéressent à la violence dans les relations amoureuses (p. ex., Diaz-Aguado & Martinez, 2015; Reyes et collègues, 2017), Sophie décide de dichotomiser les indicateurs. Cela implique que sur une échelle de violence allant de 1 à 5, où 1 désigne l'absence de perpétration, tous les participants ayant répondu 2 ou plus seront identifiés comme perpétrateurs.

Démonstration de l'analyse de classes latentes avec R à l'aide du package `poLCA`

Sophie veut recourir à un logiciel puissant et gratuit pour réaliser ces analyses. Elle choisit le logiciel R. Elle adopte également Rstudio, une interface graphique à utiliser conjointement avec R, qui lui permettra de plus facilement faire ces analyses. Tant R que Rstudio comportent de nombreux avantages en ce qui a trait à la programmation d'analyses, à la visualisation des données ainsi qu'à la gestion des variables et de l'environnement de travail.

Sophie commence par importer son jeu de données nommé `jd` (voir Caron, en écriture, pour plus de rensei-

gnements sur comment importer un jeu de données) :

```
jd <- read.csv2("violencerelamour.csv")
```

Ce jeu de données contient 800 participants étudiés à l'aide de six indicateurs dichotomiques. Le premier indicateur se nomme `pro_sex` et désigne la perpétration de violence sexuelle proactive. Le second indicateur, `rea_sex`, désigne la perpétration violence sexuelle réactive. `rea_phys` représente la perpétration de violence physique réactive, tandis que `pro_phys` désigne la perpétration de violence physique proactive. Enfin, les deux derniers indicateurs se nomment `rea_psy` et `pro_psy` et représentent respectivement la perpétration de violence psychologique réactive et de violence psychologique proactive.

Ensuite, elle installe et importe le package `poLCA` à l'aide des commandes suivantes :

```
install.packages("poLCA")
library(poLCA)
```

Sophie doit avant tout créer la formule qui est utilisée au sein de l'analyse et y inclure les indicateurs observables nécessaires (Linzer & Lewis, 2011). Puisque l'ACL implique plusieurs indicateurs observables, ceux-ci doivent être liés à l'aide de la commande `cbind(Y1, Y2, Y3, ...)` au sein de la formule. L'étudiante inclue les six indicateurs de violence dans les relations amoureuses, tels que nommés dans son jeu de données `jd`. Elle peut spécifier son modèle de la façon suivante.

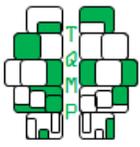
```
f1 <- as.formula(cbind(
  rea_sex, rea_psy, rea_phys,
  pro_phys, pro_sex, pro_psy) ~ 1)
```

Dans cette syntaxe, `cbind()` joint les indicateurs et `as.formula()` donne la propriété de formule (équation) à l'ensemble d'indicateurs. Le symbole `~1` indique à R qu'il s'agit d'un modèle de classes latentes de base. Le tout est assigné à la variable `f1` pour être ensuite utilisé avec la fonction `poLCA`.

Son modèle étant décrit, Sophie peut maintenant procéder à l'ACL. La commande permettant d'estimer un modèle de classes latentes dans `poLCA` est :

```
poLCA(formula, data, nclass)
```

Pour que l'analyse fonctionne, Sophie doit minimalement fournir les informations pour l'argument `formula` et l'argument `data`. L'argument `formula` correspond à la variable créée préalablement (ici appelée `f1`) et l'argument `data` correspond à son jeu de données (ici appelé `jd`). L'autre information importante à inclure est l'argument `nclass`, qui permet de déterminer le nombre de classes latentes que le modèle testé devrait supposer. Si `nclass`



n'est pas spécifié, le nombre de classes supposé par défaut est 2. La fonction `poLCA` contient aussi d'autres options qui peuvent être utilisées pour préciser le modèle testé au besoin.

Tel que mentionné précédemment, il est nécessaire de tester plusieurs modèles avec un nombre de classes différent et de comparer les résultats de chaque modèle pour déterminer celui qui présente le meilleur ajustement. Sophie utilise les commandes suivantes pour estimer des modèles contenant une seule classe, puis 2 et 3 classes :

```
LCA1 <- poLCA(f1, data = jd, nclass =1)
LCA2 <- poLCA(f1, data = jd, nclass =2)
LCA3 <- poLCA(f1, data = jd, nclass =3)
```

Toutefois, lorsqu'elle tente de réaliser l'analyse d'un modèle à 4 classes, le message d'erreur suivant s'affiche :

```
ALERT : iterations finished,
        MAXIMUM LIKELIHOOD NOT FOUND
```

Lorsque l'ACL est réalisée par `poLCA`, l'algorithme d'estimation est répété plusieurs fois (plusieurs itérations), jusqu'à ce que l'analyse converge sur le modèle ayant le meilleur ajustement possible. Si la convergence n'a pas été atteinte avant la fin de ces itérations, ce message d'erreur apparaît (Linzer & Lewis, 2011). Lorsqu'un modèle devient plus complexe, le nombre d'itérations fixé par défaut peut être insuffisant pour que l'analyse converge. Ce nombre d'itérations peut être augmenté pour améliorer la convergence. Cependant, si le problème persiste, d'autres explications doivent être investiguées (p. ex., taille de l'échantillon; Nylund-Gibson & Choi, 2018).

Le nombre d'itérations est déterminé par l'argument `maxiter`, qui est fixé à 1000 itérations par défaut. De façon à remédier à son problème, Sophie augmente le nombre d'itérations maximal, et le message d'erreur cesse d'apparaître :

```
LCA4 <- poLCA(f1, data = jd,
              nclass = 4, maxiter = 10000)
```

Dans la sortie de `poLCA` lors de la réalisation d'une ACL, l'étudiante retrouve premièrement les probabilités conditionnelles de réponse aux items pour chaque indicateur, au sein de chaque classe. La somme des probabilités conditionnelles de réponse dans chaque classe est égale à un (Linzer & Lewis, 2011). Deuxièmement, les résultats montrent la taille relative de chaque classe latente. Troisièmement, des informations concernant l'ajustement du modèle sont fournies. Le package `poLCA` calcule entre autres divers indices d'ajustement (p. ex., AIC, BIC), le nombre de paramètres estimés, les degrés de liberté résiduels et le *maximum log-likelihood*. Voici la sortie obtenue par Sophie pour le modèle à 3 classes latentes, LCA3 :

Conditional item response (column) probabilities, by outcome variable, for each class (row)

```
$rea_sex
      Pr(1) Pr(2)
class 1: 0.1650 0.8350
class 2: 0.9591 0.0409
class 3: 0.2173 0.7827
```

```
$rea_psy
      Pr(1) Pr(2)
class 1: 0.1069 0.8931
class 2: 0.9070 0.0930
class 3: 0.1833 0.8167
```

```
$rea_phys
      Pr(1) Pr(2)
class 1: 0.1943 0.8057
class 2: 0.9166 0.0834
class 3: 0.2151 0.7849
```

```
$pro_phys
      Pr(1) Pr(2)
class 1: 0.2390 0.7610
class 2: 0.8663 0.1337
class 3: 0.9848 0.0152
```

```
$pro_sex
      Pr(1) Pr(2)
class 1: 0.3089 0.6911
class 2: 0.9616 0.0384
class 3: 0.9876 0.0124
```

```
$pro_psy
      Pr(1) Pr(2)
class 1: 0.1861 0.8139
class 2: 0.8800 0.1200
class 3: 0.9595 0.0405
```

```
Estimated class population shares
0.1367 0.4245 0.4388
```

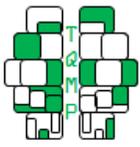
```
Predicted class memberships (by modal posterior
prob.)
0.1262 0.4425 0.4312
```

```
=====
Fit for 3 latent classes:
=====
number of observations: 800
number of estimated parameters: 20
residual degrees of freedom: 43
maximum log-likelihood: -2195.821
```

```
AIC(3): 4431.642
BIC(3): 4525.334
G^2(3): 51.86892 (Lik. ratio/deviance statistic)
X^2(3): 43.65539 (Chi-square goodness of fit)
```

Sophie aimerait aussi utiliser l'entropie relative pour guider ses décisions, mais cette statistique n'est pas fournie directement par `poLCA`. Elle doit créer une fonction supplémentaire :

```
poLCA.relentropy <- function(model){
  pik <- as.data.frame(model$posterior)
```



```
E <- 1 - (sum(-1 * pik * log(pik), na.rm = TRUE
) / (nrow(pik) * log(ncol(pik))))
return(E)
}
```

Il lui est ensuite possible de calculer la valeur de l'entropie relative pour chaque modèle testé, ici appelés LCA1, LCA2, LCA3 et LCA4 :

```
poLCA.relentropy(LCA1)
poLCA.relentropy(LCA2)
poLCA.relentropy(LCA3)
poLCA.relentropy(LCA4)
```

Sophie veut maintenant effectuer le test de ratio de vraisemblance Lo-Mendell-Rubin (LMR; Lo et collègues, 2001). Il est nécessaire d'installer et d'importer le package tidyLPA :

```
install.packages("tidyLPA")
library(tidyLPA)
```

Par la suite, il faut comparer chaque paire de modèles entre eux. Sophie doit comparer le modèle LCA1 avec LCA2, LCA2 avec LCA3, puis LCA3 avec LCA4. Voici la

Lo-Mendell-Rubin ad-hoc adjusted likelihood ratio test:
 LR = 338.928, LMR LR (df = 7) = 322.830, $p < 0.001$

Si les résultats sont significatifs (si la valeur p est inférieure à seuil prédéterminé, par exemple, 0,05), le modèle ayant une classe de plus, soit le LCA3, est significativement meilleur que le LCA2 (Nylund-Gibson & Choi, 2018).

Enfin, Sophie veut produire un graphique des schémas de réponses pour chaque classe latente d'un modèle. Elle utilise la commande `plot` qui est une fonction générique pour produire des graphiques et pour laquelle une version spécifique permet de représenter des sorties `poLCA`. Voici comment l'étudiante génère le graphique des réponses pour les classes latentes du modèle à 3 classes. La Figure 2 montre le graphique réalisé à l'aide de cette commande :

```
plot(LCA3)
```

Interprétation fictive

Sophie rassemble les indices d'ajustement choisis et les statistiques diagnostiques de chaque modèle dans un seul tableau pour faciliter leur comparaison. Pour ce faire, elle rapporte les indices d'ajustement fournis dans les résultats de chaque modèle (p. ex., BIC, AIC, LMR) et les statistiques

fonction utilisée pour comparer LCA2 avec LCA3 :

```
calc_lrt(n = LCA2$N,
  null_ll = LCA2$llik,
  null_param = LCA2$npar,
  null_classes = length(LCA2$P),
  alt_ll = LCA3$llik,
  alt_param = LCA3$npar,
  alt_classes = length(LCA3$P))
```

L'argument `n` désigne le nombre de participants de l'échantillon. L'argument `null_ll` requiert le *maximum log-likelihood* du modèle contenant le plus petit nombre de classes, soit LCA2. L'argument `null_param` désigne le nombre de paramètres estimés pour LCA2, et le `null_classes` désigne le nombre de classes de ce modèle. En comparaison, l'argument `alt_ll` requiert le *maximum log-likelihood* du modèle LCA3, le `alt_param`, le nombre de paramètres estimés pour ce modèle, et le `alt_classes`, le nombre de classes de ce modèle. Voici les résultats obtenus lorsque LCA2 est comparé avec LCA3 :

diagnostiques, dans ce cas-ci, l'entropie relative. Elle inclut le tableau 1 dans la section « Résultats » de son article. Le dernier modèle qu'elle teste est celui à 4 classes, puisqu'elle remarque que la plupart des indices d'ajustement indiquent que le modèle ne s'améliore plus.

L'étudiante remarque que le *maximum log-likelihood* n'est pas le meilleur indice d'ajustement pour guider son choix, puisqu'il ne cesse de suggérer que la vraisemblance augmente à mesure qu'il y a ajout d'une classe. Le BIC et l'AIC du modèle à trois classes sont ceux avec les valeurs les plus petites, ce qui suggère que ce modèle présente le meilleur ajustement. Ensuite, le LMR indique que le modèle à deux classes est significativement meilleur que celui à une classe ($p < .001$) et que celui à trois classes est significativement meilleur que celui à deux classes ($p < .001$). Le modèle à quatre classes n'est cependant pas significativement meilleur que le modèle à trois classes ($p = .248$), ce qui suggère que le modèle à trois classes est le plus approprié. L'entropie relative du modèle à trois classes est au-dessus de .80 et représente la valeur la plus élevée, ce qui indique que ce modèle distingue bien les classes entre elles. Ainsi, les indices d'ajustement et l'entropie convergent tous

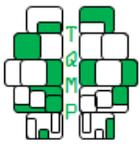


Figure 2 ■ Schémas de réponses pour le modèle à trois classes latentes

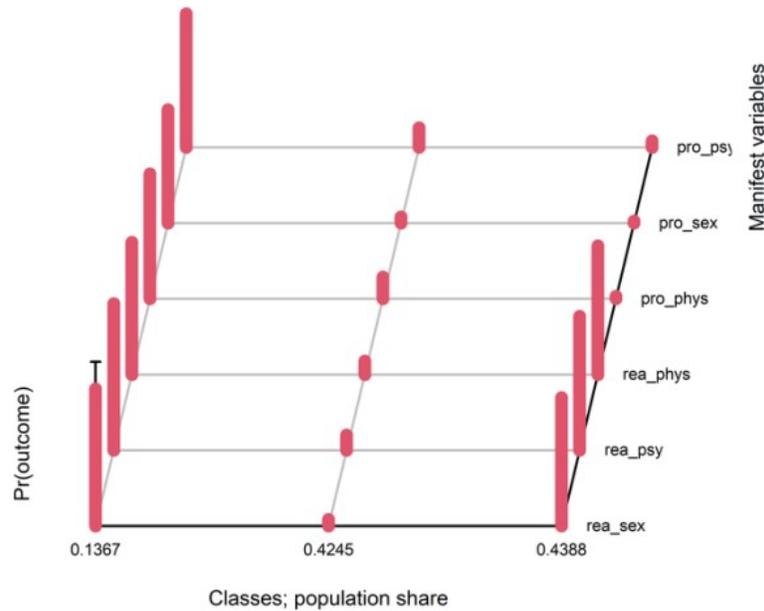


Tableau 1 ■ Indices d’ajustement du modèle et statistique diagnostique

| Modèles testés | Maximum log-likelihood | AIC | BIC | LMRa | Entropie relative | Degrés de liberté |
|----------------|------------------------|-----------------|-----------------|-----------------|-------------------|-------------------|
| 1 classe | -2687.896 | 5387.793 | 5415.9 | N/A | N/A | 57 |
| 2 classes | -2365.285 | 4756.57 | 4817.47 | 614.576* | 0.802721 | 50 |
| 3 classes | -2195.821 | 4431.642 | 4525.334 | 322.830* | 0.8396914 | 43 |
| 4 classes | -2191.063 | 4436.127 | 4562.611 | 9.064 | 0.7452709 | 36 |

Note. ^a = Lo-Mendell-Rubin ad-hoc adjusted likelihood ratio test. * = $p < .001$. Le gras indique la meilleure valeur pour l’indicateur concerné.

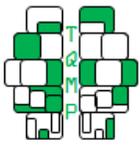
vers le modèle à 3 classes.

Il peut arriver que certains indices suggèrent différentes solutions, auquel cas Sophie devra user de son jugement critique et de ses connaissances dans son domaine. En effet, bien que les indices précédemment relevés guident la décision de Sophie, il est aussi important qu’elle base son interprétation sur le sens théorique des modèles trouvés. Dans cet exemple, les trois classes identifiées dans le modèle vers lequel les indices convergent sont jugées théoriquement acceptables par l’étudiante et sont semblables à celles trouvées dans des articles antérieurs (ce qu’elle soulignera dans la discussion). En s’appuyant sur la Figure 2, Sophie peut aussi analyser la distribution des participants dans chaque classe pour guider sa décision d’inclure une classe ou non. Lorsqu’une classe est composée d’une petite proportion de l’échantillon, il faut considérer qu’elle pourrait être difficile à retrouver dans un plus petit échantillon (Weller et collègues, 2020). Ainsi, son impor-

tance théorique et son soutien statistique sont cruciaux pour justifier son inclusion.

Comment rapporter les résultats

Selon les résultats, Sophie détermine qu’il y a trois classes, car ce modèle fournit les meilleurs indices d’ajustement (voir tableau 1). Les trois classes identifiées sont les suivantes (voir Figure 2) : (1) violence multiforme ($n = 109$; 13,67%); (2) peu ou pas de violence ($n = 340$; 42,45%); (3) violence réactive ($n = 351$; 43,88%). Les jeunes appartenant à la première classe ont des probabilités élevées d’endosser tous les items de perpétration de violence, soit les items de violence sexuelle réactive (.84) et proactive (.69), de violence psychologique réactive (.89) et proactive (.81), et de violence physique réactive (.81) et proactive (.76). Les jeunes appartenant à la seconde classe ont des probabilités moindres d’endosser les items de violence sexuelle proactive (.04) et réactive (.04), de violence psychologique



proactive (.12) et réactive (.01) et de violence physique proactive (.13) et réactive (.08). Enfin, les jeunes appartenant à la troisième classe ont des probabilités élevées de répondre aux items de violence sexuelle réactive (.78), de violence psychologique réactive (.82) et de violence physique réactive (.78) et des probabilités faibles d'endosser les items de violence sexuelle proactive (.01), de violence psychologique proactive (.04) et de violence physique proactive (.02).

Dans la section « Discussion » de son article, Sophie rappelle d'abord l'importance de son objectif de recherche. Ensuite, elle rapporte les classes trouvées, accompagnées de la proportion d'adolescents appartenant à ces classes entre parenthèses. Finalement, elle fait des liens avec des classes identifiées dans les résultats d'autres articles. Par exemple, la première classe est composée d'adolescents qui perpétuent de la violence sexuelle, psychologique et physique de façon proactive et réactive, ce qui est semblable à la classe de violence multiforme identifiée dans des études antérieures (Diaz-Aguado & Martinez, 2015; Reidy et collègues, 2016; Reyes et collègues, 2017).

Conclusion

L'ACL est une analyse fréquemment utilisée en raison des avancées qu'elle permet de réaliser pour mieux comprendre la complexité d'un phénomène. Les classes identifiées peuvent ensuite être utilisées pour créer un système de classification ou pour guider la création de programmes d'intervention variés en fonction des caractéristiques propres à chaque sous-groupe d'individus (Nylund-Gibson & Hart, 2014). Il est aussi intéressant de mentionner que le modèle d'ACL de base présenté dans cet article peut être bonifié pour inclure des covariables, qui peuvent prédire la variable latente étudiée (Porcu & Giambona, 2017). Le modèle s'appelle alors une régression de classes latentes, et considère l'effet des covariables sur la probabilité que des individus appartiennent à une certaine classe latente. Ces analyses supplémentaires permettent de mieux cerner le profil des individus se retrouvant dans certains sous-groupes. En outre, il est possible d'examiner les conséquences associées à l'appartenance à une classe, notamment en effectuant une régression des variables d'intérêt sur les variables de classification dans les différentes classes. Ces extensions de l'ACL n'ont pas été illustrées dans le présent article, mais gagneraient à être exemplifiées avec R dans le futur. Par ailleurs, un nouveau package prometteur, StepMixR, vient tout juste de paraître (Morin et collègues, 2023).

Pour conclure, l'illustration d'une ACL avec R contribue à diffuser les possibilités statistiques de R et à les rendre accessibles à la communauté derrière cette plateforme libre accès, qui ne cesse de gagner en popularité. Par la commo-

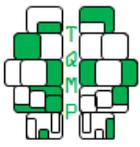
dité de son utilisation et l'ampleur de ses capacités, R permet aux utilisateurs de programmer leurs propres analyses et de les rendre aisément transmissibles et reproductibles, permettant le partage des connaissances entre chercheurs (Caron, en écriture). Il est avantageux d'accroître ses habiletés de programmation avec R afin de réaliser des analyses statistiques de qualité.

Note des auteurs

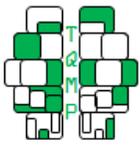
Les deux premières autrices ont contribué également au présent article.

Références

- Adams, T. R., Handley, E. D., Warmingham, J. M., Manly, J. T., Cicchetti, D., & Toth, S. L. (2021). Patterns of dating violence moderate the effect of child maltreatment on suicide risk among disadvantaged minority female adolescents with depressive symptoms. *Journal of Family Violence*, 36(1), 5-16. doi : [10.1007/s10896-020-00153-5](https://doi.org/10.1007/s10896-020-00153-5).
- Anderson, K. M., & Danis, F. S. (2007). Collegiate sororities and dating violence : An exploratory study of informal and formal helping strategies. *Violence Against Women*, 13(1), 87-100. doi : [10.1177/1077801206294808](https://doi.org/10.1177/1077801206294808).
- Bartholomew, D. J., Knott, M., & Moustaki, I. (2011). *Latent variable models and factor analysis : A unified approach (3rd ed.)* Wiley.
- Bushman, B. J., & Anderson, C. A. (2001). Media violence and the American public : Scientific facts versus media misinformation. *American Psychologist*, 56, 477. doi : [10.1037/0003-066X.56.6-7.477](https://doi.org/10.1037/0003-066X.56.6-7.477).
- Butters, R. P., Droubay, B. A., Seawright, J. L., Tollefson, D. R., Lundahl, B., & Whitaker, L. (2021). Intimate partner violence perpetrator treatment : Tailoring interventions to individual needs. *Clinical Social Work Journal*, 49(3), 391-404. doi : [10.1007/s10615-020-00763-y](https://doi.org/10.1007/s10615-020-00763-y).
- Caron, P. O. (2019). *La modélisation par équations structurelles avec Mplus*. PUQ.
- Caron, P.-O. (en écriture). *Méthodes quantitatives avec R*. <https://mqr.teluq.ca>
- Centers for Disease Control and Prevention. (2017). *Preventing Teen Dating Violence*. <https://www.cdc.gov/>
- Chung, J. Y., & Lee, S. (2020). Are bully-victims homogeneous? Latent class analysis on school bullying. *Children and Youth Services Review*, 112(10492-1), 2. doi : [10.1016/j.childyouth.2020.104922](https://doi.org/10.1016/j.childyouth.2020.104922).
- Collins, L. M., & Lanza, S. T. (2010). *Latent Class and Latent Transition Analysis : With Applications in the Social, Behavioural, and Health Sciences*. John Wiley.
- Diaz-Aguado, M. J., & Martinez, R. (2015). Types of adolescent male dating violence against women, self-esteem, and justification of dominance and aggression. *Jour-*



- nal of Interpersonal Violence*, 30(15), 2636-2658. doi : [10.1177/0886260514553631](https://doi.org/10.1177/0886260514553631).
- Finch, W. H., & Bronk, K. C. (2011). Conducting Confirmatory Latent Class Analysis Using Mplus. *Structural Equation Modeling*, 18(1), 132-151. doi : [10.1080/10705511.2011.532732](https://doi.org/10.1080/10705511.2011.532732).
- Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, 61(2), 215-231. doi : [10.1093/biomet/61.2.215](https://doi.org/10.1093/biomet/61.2.215).
- Hojtink, H. (2001). Confirmatory latent class analysis : Model selection using Bayes factors and (pseudo) likelihood ratio statistics. *Multivariate Behavioral Statistics*, 36(4), 563-588. doi : [10.1207/S15327906MBR3604_04](https://doi.org/10.1207/S15327906MBR3604_04).
- Johann, D., Steinbrecher, M., & Thomas, K. (2020). Channels of participation : Political participant types and personality. *Plos one*, 15(10), e0240671-1. doi : [10.1371/journal.pone.0240671](https://doi.org/10.1371/journal.pone.0240671).
- Karnowski, V. (2017). Latent class analysis. In J. Matthes (Éd.). Wiley. doi : [10.1002/9781118901731.iecrm0130](https://doi.org/10.1002/9781118901731.iecrm0130).
- Laudy, O., Zoccolillo, M., Baillargeon, R. H., Boom, J., Tremblay, R. E., & Hoijtink, H. (2005). Applications of confirmatory latent class analysis in developmental psychology. *European Journal of Developmental Psychology*, 2(1), 1-15. doi : [10.1080/17405620444000193](https://doi.org/10.1080/17405620444000193).
- Lazarsfeld, P. F. (1950). *The Logical and Mathematical Foundation of Latent Structure Analysis*. Dans S. Princeton University Press.
- Linzer, D. A., & Lewis, J. B. (2011). polCA : An R package for polytomous variable latent class analysis. *Journal of Statistical Software*, 42(10), 1-29. doi : [10.18637/jss.v042.i10](https://doi.org/10.18637/jss.v042.i10).
- Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, 88(3), 767-778. doi : [10.1093/biomet/88.3.767](https://doi.org/10.1093/biomet/88.3.767).
- Morin, S., Legault, R., Bakk, Z., Giguère, C. É., de la Sablonnière, R., & Lacourse, É. (2023). A Python Package for Pseudo-Likelihood Estimation of Generalized Mixture Models with External Variables. arXiv. doi : [10.48550/arXiv.2304.03853](https://doi.org/10.48550/arXiv.2304.03853).
- Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling : A Monte Carlo simulation study. *Structural Equation Modeling*, 14(4), 535-569. doi : [10.1080/10705510701575396](https://doi.org/10.1080/10705510701575396).
- Nylund-Gibson, K., & Choi, A. Y. (2018). Ten frequently asked questions about latent class analysis. *Translational Issues in Psychological Science*, 4(4), 440-449. doi : [10.1037/tps0000176](https://doi.org/10.1037/tps0000176).
- Nylund-Gibson, K., & Hart, S. R. (2014). *Latent class analysis in prevention science*. Dans Sloboda.
- Petersen, K. J., Qualter, P., & Humphrey, N. (2019). The application of latent class analysis for investigating population child mental health : A systematic review. *Frontiers in psychology*, 10, 1214-1. doi : [10.3389/fpsyg.2019.01214](https://doi.org/10.3389/fpsyg.2019.01214).
- Porcu, M., & Giambona, F. (2017). Introduction to latent class analysis with applications. *The Journal of Early Adolescence*, 37(1), 129-158. doi : [10.1177/0272431616664845](https://doi.org/10.1177/0272431616664845).
- Ramirez, J. M., & Andreu, J. M. (2006). Aggression, and some related psychological constructs (anger, hostility, and impulsivity) Some comments from a research project. *Neuroscience & Biobehavioral Reviews*, 30(3), 276-291. doi : [10.1016/j.neubiorev.2005.04.015](https://doi.org/10.1016/j.neubiorev.2005.04.015).
- Reidy, D. E., Ball, B., Houry, D., Holland, K. M., Valle, L. A., Kearns, M. C., Marshall, K. J., & Rosenbluth, B. (2016). In search of teen dating violence typologies. *Journal of Adolescent Health*, 58(2), 202-207. doi : [10.1016/j.jadohealth.2015.09.028](https://doi.org/10.1016/j.jadohealth.2015.09.028).
- Reyes, H., Foshee, V. A., Chen, M. S., & Ennett, S. T. (2017). Patterns of dating violence victimization and perpetration among Latino youth. *Journal of Youth and Adolescence*, 46(8), 1727-1742. doi : [10.1007/s10964-016-0621-0](https://doi.org/10.1007/s10964-016-0621-0).
- Sinha, P., Calfee, C. S., & Delucchi, K. L. (2021). Practitioner's Guide to Latent Class Analysis : Methodological Considerations and Common Pitfalls. *Critical Care Medicine*, 49(1), e63-e79. doi : [10.1097/CCM.0000000000004710](https://doi.org/10.1097/CCM.0000000000004710).
- Théorêt, V., Hébert, M., Fernet, M., & Blais, M. (2021). Gender-specific patterns of teen dating violence in heterosexual relationships and their associations with attachment insecurities and emotion dysregulation. *Journal of Youth and Adolescence*, 50(2), 246-259. doi : [10.1007/s10964-020-01328-5](https://doi.org/10.1007/s10964-020-01328-5).
- Ulbricht, C. M., Chrysanthopoulou, S. A., Levin, L., & Lapane, K. L. (2018). The use of latent class analysis for identifying subtypes of depression : A systematic review. *Psychiatry Research*, 266(1), 228-246. doi : [10.1016/j.psychres.2018.03.003](https://doi.org/10.1016/j.psychres.2018.03.003).
- Vermunt, J. K. (1997). *LEM : A general program for the analysis of categorical data*. Department of Methodology ; Statistics.
- Vermunt, J. K. (2002). Latent class analysis of complex sample survey data : Application to dietary data : Comment. *Journal of the American Statistical Association*, 97(459), 736-737. <https://pure.uvt.nl/ws/portalfiles/portal/487923/jasa.pdf>
- Weller, B. E., Bowen, N. K., & Faubert, S. J. (2020). Latent class analysis : a guide to best practice. *Journal of Black Psychology*, 46(4), 287-311. doi : [10.1177/0095798420930932](https://doi.org/10.1177/0095798420930932).
- Wurpts, I. C., & Geiser, C. (2014). Is adding more indicators to a latent class analysis beneficial or detrimental? Results of a Monte-Carlo study. *Frontiers in Psychology*, 5, 920-1. doi : [10.3389/fpsyg.2014.00920](https://doi.org/10.3389/fpsyg.2014.00920).

**Annexe : Listings 1 et 2.****Listing 1 – Script R de l'ACL**

```
# Package
install.packages("poLCA")
library(poLCA)

# Analyse de classes latentes
f1 <- as.formula(cbind(rea_sex, rea_psy, rea_phys, pro_phys, pro_sex, pro_psy) ~ 1)

LCA1 <- poLCA(f1, data = jd, nclass = 1)
LCA2 <- poLCA(f1, data = jd, nclass = 2)
LCA3 <- poLCA(f1, data = jd, nclass = 3)
LCA4 <- poLCA(f1, data = jd, nclass = 4, maxiter = 10000)

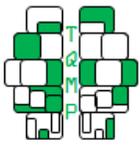
# Créer une fonction pour calculer l'entropie poLCA.relentropy <- function(model){
  pik <- as.data.frame(model$posterior)
  1 - (sum(-1 * pik * log(pik), na.rm = TRUE) / (nrow(pik)* log(ncol(pik)))) }

# Calculer les quatre entropies
poLCA.relentropy(LCA1)
poLCA.relentropy(LCA2)
poLCA.relentropy(LCA3)
poLCA.relentropy(LCA4)

# Test de ratio de vraisemblance Lo-Mendell-Rubin install.packages("tidyLPA")
library(tidyLPA)

# 1 classe vs 2 classes
calc_lrt(n = LCA1$N,
         null_ll = LCA1$llik,
         null_param = LCA1$npar,
         null_classes = length(LCA1$P),
         alt_ll = LCA2$llik,
         alt_param = LCA2$npar,
         alt_classes = length(LCA2$P))

# 2 classes vs 3 classes
calc_lrt(n = LCA2$N,
         null_ll = LCA2$llik,
         null_param = LCA2$npar,
         null_classes = length(LCA2$P),
         alt_ll = LCA3$llik,
         alt_param = LCA3$npar,
         alt_classes = length(LCA3$P))
```



```
# 3 classes vs 4 classes
calc_lrt(n = LCA3$N,
        null_ll = LCA3$llik,
        null_param = LCA3$npar,
        null_classes = length(LCA3$P),
        alt_ll = LCA4$llik,
        alt_param = LCA4$npar,
        alt_classes = length(LCA4$P))

# Créer le graphique
plot(LCA3)
```

Listing 2 – Script R pour la création d’un jeu de données fictif

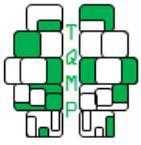
```
# Installer (au besoin) et appeler le package poLCA install.packages("poLCA")
library(poLCA)

# Déclarer une valeur de départ, nommée « graine » ou seed en anglais,
# pour s'assurer de toujours obtenir les mêmes valeurs
set.seed(42)

# Déterminer vers quelle solution l'ACL converge, ainsi que les probabilités
# de réponses à chaque item des six indicateurs, pour chacune des trois
# classes prévues, à l'aide de la fonction probs

# Créer un modèle à 3 classes composé de 6 indicateurs dichotomiques.
# Ainsi, il y a 6 matrices (indicateurs), de 3 lignes (classes) et 2
# colonnes (nombre d'options par indicateur)

probs <- list(matrix(c(.95, .05,
                      .99, .01,
                      .30, .70), ncol = 2, byrow = TRUE),
             matrix(c(.95, .05,
                      .20, .80,
                      .15, .85), ncol = 2, byrow = TRUE),
             matrix(c(.90, .10,
                      .20, .80,
                      .20, .80), ncol = 2, byrow = TRUE),
             matrix(c(.90, .10,
                      .98, .02,
                      .23, .77), ncol = 2, byrow = TRUE),
             matrix(c(.90, .10,
                      .20, .80,
                      .10, .90), ncol = 2, byrow = TRUE),
             matrix(c(.90, .10,
                      .97, .03,
                      .20, .80), ncol = 2, byrow = TRUE)
            )
```



```
# Utiliser la fonction simdata() pour créer un jeu de données fictif,  
# ici contenant 800 participants  
simdata <- poLCA.simdata(N = 800, probs)  
  
# Extraire le jeu de la liste de sortie de simdata()  
jd <- simdata$dat  
  
# Renommer les variables  
colnames(jd) <- c("pro_sex", "rea_sex", "rea_phys",  
                 "pro_phys", "rea_psy", "pro_psy")
```

Citation

Bellemare-Lepage, A., Chatelois, M., & Caron, P.-O. (2023). Exemplification méthodologique d'une analyse de classes latentes avec R. *The Quantitative Methods for Psychology*, 19(2), 217-229. doi : [10.20982/tqmp.19.2.p217](https://doi.org/10.20982/tqmp.19.2.p217).

Copyright © 2023, Bellemare-Lepage *et collègues*. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Received: 19/05/2023 ~ Accepted: 16/06/2023