



Une filiale de DOCAPOSTE

EPICLIN 2025

du 14 au 16 mai 2025, Bordeaux, France

Prise en main et application de la méthode PLS-PM à l'aide de données provenant d'une étude observationnelle sur la drépanocytose : un tutoriel

Nada ASSI¹, Geoffrey CHEMINET², Jean-Benoit ARLET², Eléonore HERQUELOT¹

¹Heva, Lyon, France ; ²AP-HP, Paris, France



LA POSTE
SANTÉ &
AUTONOMIE

Qu'est-ce que la PLS-PM ?

Partial Least Squares - Path Modeling

Méthode multivariée conçue pour modéliser un **réseau de relations statistiques** entre des variables **latentes** indirectement mesurées au moyen d'un ensemble de variables observées et **mesurées** appelées variables **manifestes**.

Elle permet de **structurer des données hétérogènes** en explorant les relations pouvant exister entre différents blocs de variables.

L'approche **PLS-PM** n'impose **pas d'hypothèses distributionnelles** sur les données qui sont difficiles à satisfaire dans la réalité, et elle **ne s'appuie pas sur un processus de génération de données** ni sur des interprétations de modélisation causale.

L'objectif final de l'utilisation de PLS-PM est de **fournir un résumé pratique** de la façon dont l'ensemble des variables dépendantes est systématiquement expliqué par leurs ensembles de prédicteurs.

PLS path modeling by Michel Tenenhaus, Vincenzo Esposito Vinzi, Yves-Marie Chatelin, and Carlo Lauro (2005), Computational Statistics & Data Analysis (48: 159-205).



Définition des modèles

Chaque modèle PLS-PM est constitué de 2 sous-modèles

1. Modèle structurel (modèle interne)

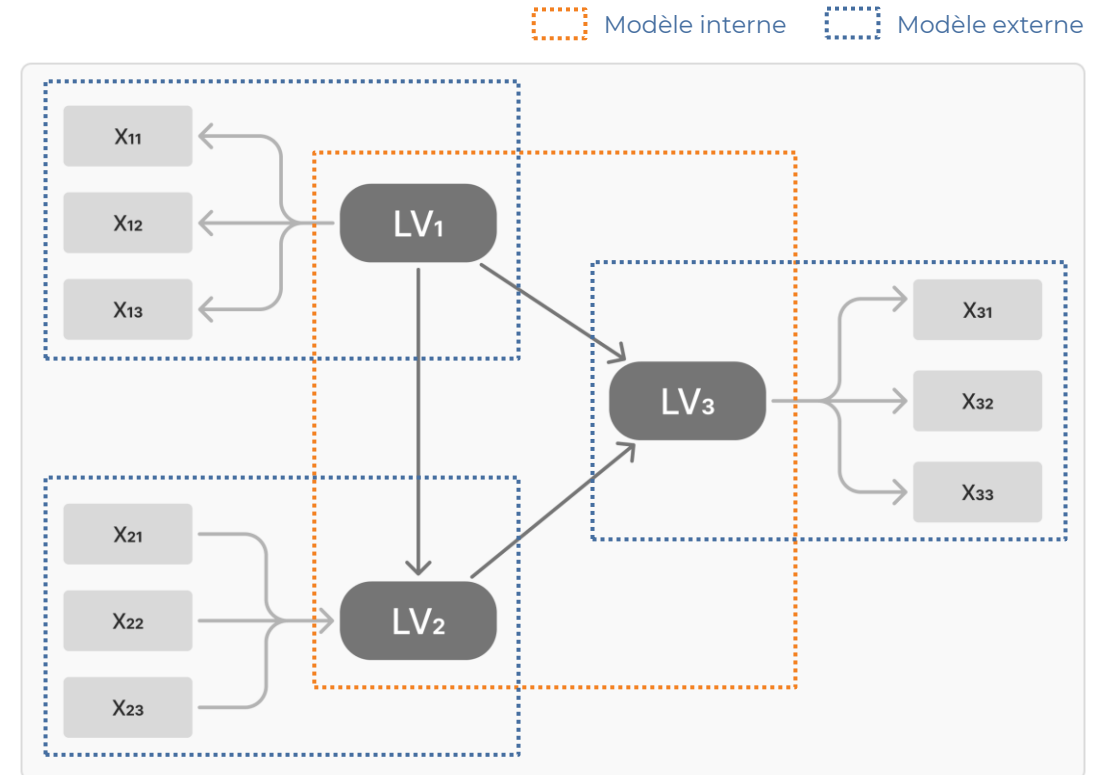
Décris les relations entre les variables latentes (LV)

Toutes les relations dans le modèle structurel sont considérées comme linéaires.

2. Modèle de mesure (modèle externe)

Il s'agit de la partie du modèle qui concerne les relations entre une variable latente (LV) et son bloc de variables manifestes (MVs).

 Des hypothèses sont faites sur la présence des flèches et leurs directions



Toutes les relations dans le modèle structurel sont linéaires

$$LV_j = \beta_0 + \sum_{i \rightarrow j} \beta_{ji} LV_i + erreur_j$$

Où :

- L'indice i de LV_i désigne toutes les variables latentes censées prédire LV_j ,
- β_{ji} sont les coefficients du chemin càd. la “force et la direction” des relations entre la réponse LV_j et les prédicteurs LV_i ,
- β_0 est simplement l'ordonnée à l'origine,
- $erreur_j$ représente le terme résiduel.

Le **système d'équations est récursif** càd. que les chemins formés entre les LV par les flèches du modèle structurel ne peuvent pas former de boucles.

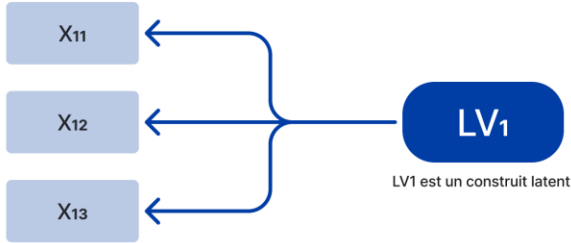
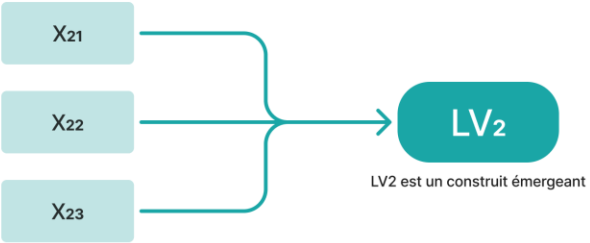
Spécifications de la régression/prédiction :

Nous voulons comprendre les valeurs conditionnelles attendues de la réponse LV_j déterminée par les prédicteurs LV_i avec le modèle $E(LV_j | LV_i) = \beta_{0i} + \Sigma \beta_{ji} LV_i$ la seule hypothèse supplémentaire étant la **décorrélacion entre une variable latente LV_j et l'erreur résiduelle $erreur_j$.**

$$cov(LV_j, erreur_j) = 0$$

Le Modèle de Mesure

Spécification des relations entre une LV et son bloc de MVs

	Mode réflexif	Mode formatif
	 LV1 est un construit latent	 LV2 est un construit émergent
Direction des relations entre LV et MVs	LV → MVs	MVs → LV
Nature des indicateurs	Les MVs « reflètent » la LV, ce sont des indicateurs " d'effets ".	Les MVs " forment " la LV, ce sont des indicateurs " causaux ".
Prend en compte la variance ou covariance observée	✅ Oui	✅ Oui
Les MVs covariant-elles ?	Généralement → Les changements dans une MV impliquent des changements dans les autres	Pas nécessairement → Les changements dans une MV n'impliquent pas de changements dans les autres
Le retrait d'un indicateur de la MV modifie-t-il le sens du construit ?	❌ Non, robuste au retrait d'indicateurs	✅ Oui, peut être sensible au changement d'indicateurs
La cohérence interne (CI) s'applique-t-elle ?	✅ Oui, les mesures de CI s'appliquent	❌ Non
Critères de validation	Vérifier l' unidimensionnalité : • Alpha de Cronbach • Rho de Dillon • Valeurs propres	Vérifier la redondance : • Variance moyenne extraite (AVE) • Communalité de bloc • Redondance moyenne

Spécification du modèle de mesure

Toutes les relations dans le modèle de mesure sont linéaires

$$X_{jk} = \lambda_{0jk} + \lambda_{jk}LV_j + \text{erreur}_{jk} \text{ (mode réflexif)}$$

$$LV_j = \lambda_{0j} + \lambda_{jk}X_{jk} + \text{erreur}_j \text{ (mode formatif)}$$

Où :

- X_{jk} sont les variables manifestes de la j-ème LV
- λ_{jk} sont les “loadings” de chacune des MV de la j-ème LV
- λ_{0jk} and λ_{0j} sont des termes d’intercept

Spécifications de la régression/prédiction :

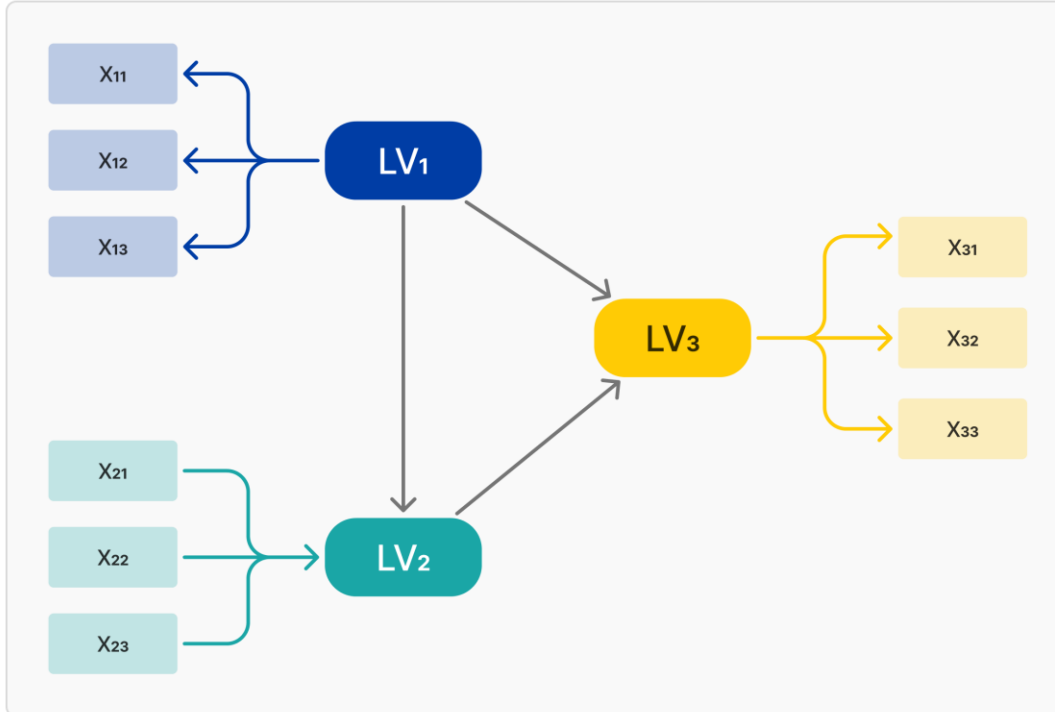
Notre objectif est de comprendre autant que possible les valeurs conditionnelles attendues des variables de réponse (manifestes ou latentes) en fonction des variables prédictives avec les modèles suivants :

$$E(X_{jk}|LV_j) = \lambda_{0jk} + \lambda_{jk}LV_j \text{ (mode réflexif)}$$

$$E(LV_j|X_{jk}) = \lambda_{0j} + \lambda_{jk}X_{jk} \text{ (mode formatif)}$$

L’approche PLS-PM est basée sur les composantes, où chaque LV est estimée comme une combinaison linéaire des MVs de son bloc (qui le compose/réfléchit). Une estimation $\widehat{LV}_j$ est appelée score et se note Y_j :

$$\widehat{LV}_j = Y_j = \sum_k \omega_{jk}X_{jk}$$



Modèle de mesure

LV_1 et LV_3 sont des blocs réflexifs

- $X_{1k} = \lambda_{1k} LV_1 + \text{erreur}$ avec $k = 1, 2, 3$
- $X_{3k} = \lambda_{3k} LV_3 + \text{erreur}$ avec $k = 1, 2, 3$

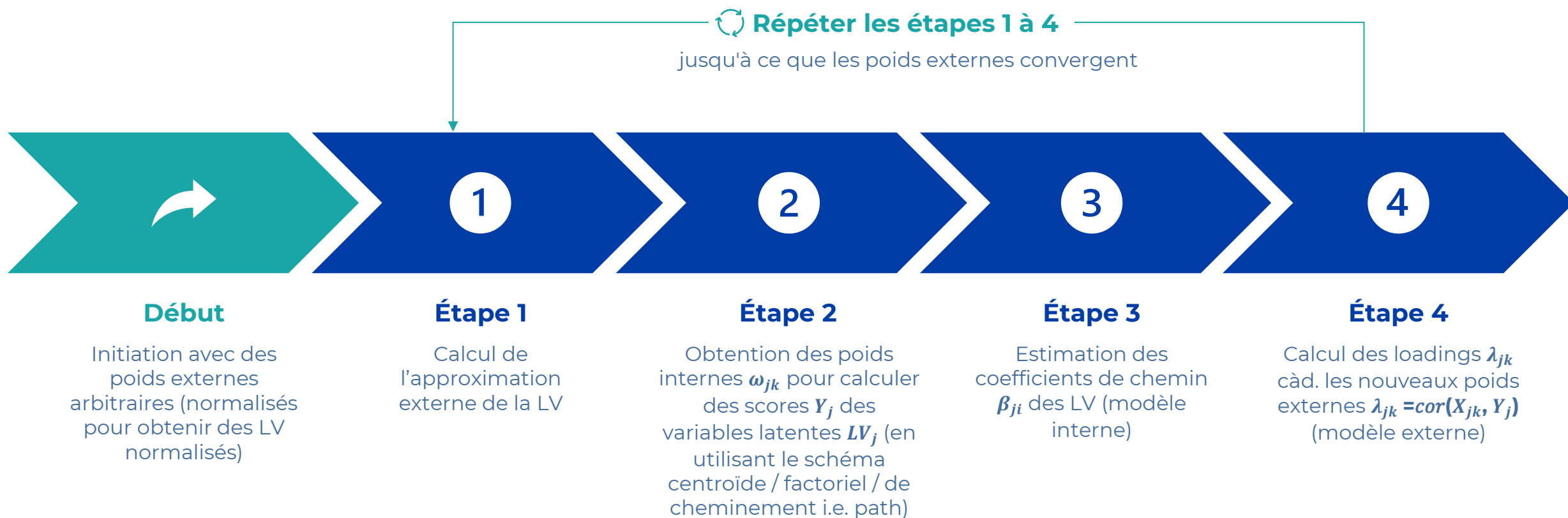
LV_2 est un bloc formatif

- $LV_2 = \sum_k \lambda_{jk} X_{2k} + \text{erreur}$ avec $k = 1, 2, 3$

Modèle Structurel

- $LV_2 = \beta_{21} LV_1 + \text{erreur}$
- $LV_3 = \beta_{31} LV_1 + \beta_{32} LV_2 + \text{erreur}$

 Toutes les variables sont standardisées pour éliminer les termes d'intercept.



Application aux données observationnelles

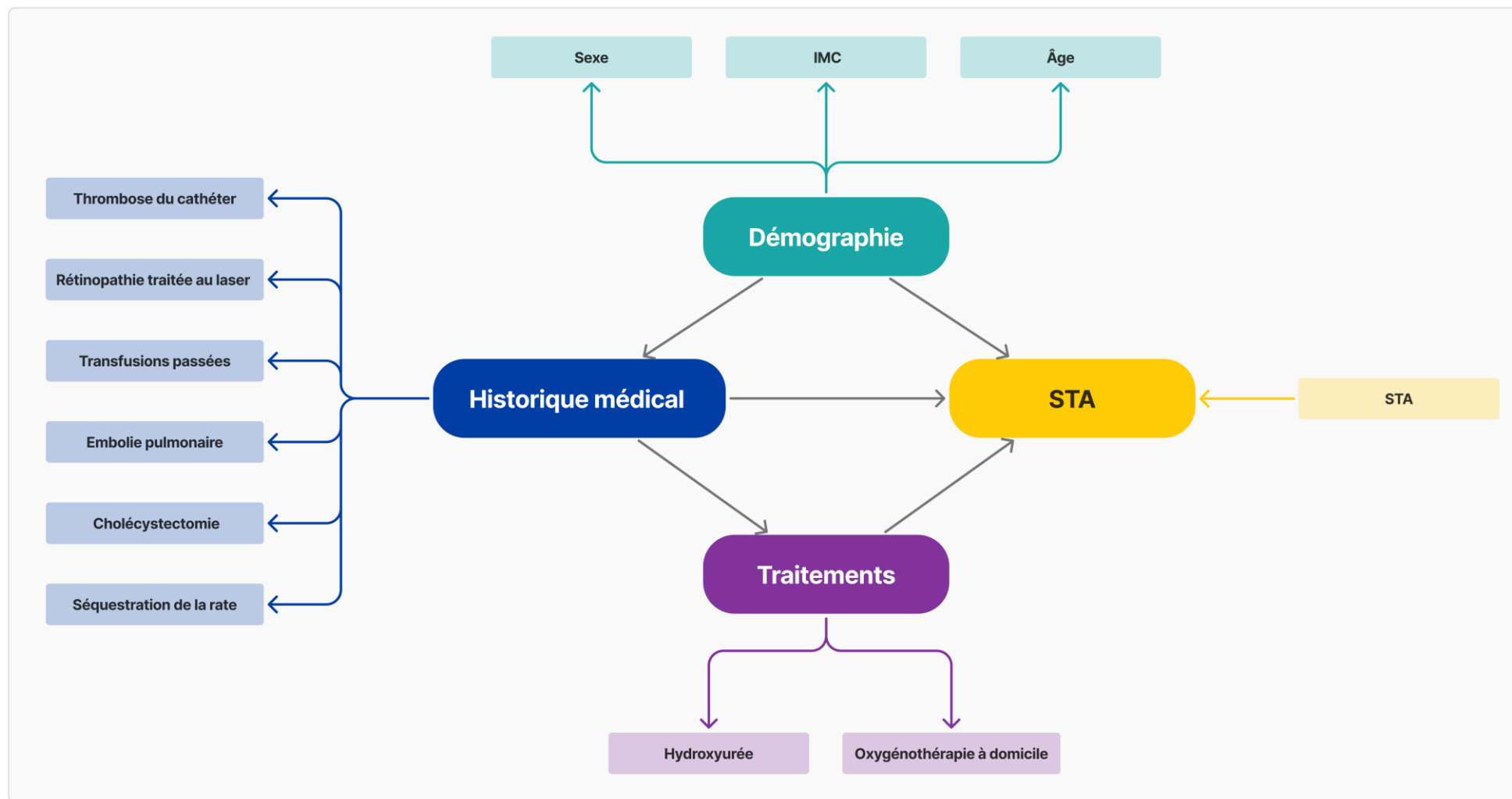
L'étude DREPANO

Une étude observationnelle menée sur un registre de **drépanocytose** de l'**AP-HP**.

Objectif : **structurer** un groupe de variables significativement associés au **Syndrome Thoracique Aigu (STA)** dans un modèle de régression multinomiale et **caractériser les interrelations existantes** entre elles.

- Nous définissons **4 variables latentes (LV)**, les liens entre elles et la direction de l'association supposée.
- Chaque LV est composée d'un bloc de **variables manifestes (MV)** ou mesurées.
- Le jeu de données utilisé dans cette application comprend 263 patients drépanocytaires âgés de 18 ans et plus, inclus de 2015 à 2024 dans le registre de l'AP-HP, avec les génotypes Hemoglobine SS (HbSS) et HbS beta thalassemia 'zero' (HbS beta⁰).





Comment implémente-t-on cela avec R ?

Le package `plspm` (Gaston Sanchez, Laura Trinchera, Giorgio Russolillo)

Matrice du modèle interne

```
demo <- c(0, 0, 0, 0)
hist <- c(1, 0, 0, 0)
trts <- c(0, 1, 0, 0)
acs <- c(1, 1, 1, 0)
# Path matrix created by row binding
inner_path_matrix <- rbind(demo, hist, trts,
acs)
colnames(inner_path_matrix) <-
rownames(inner_path_matrix)

innerplot(inner_path_matrix)
```

Matrice du modèle externe

```
demo_block <- c("age", "gender", "bmi")
hist_block <- c("past_transfusion",
"retina_laser", "spleen_sequestration",
"cholecystectomy",
"pulm_embolism", "cath_thrombosis")
trts_block <- c("hydroxyurea",
"home_oxygenotherapy")
acs_block <- c("acs_nb_life_cl")

outer_path_blocks <- list(demo_block,
hist_block, trts_block, acs_block)

outer_modes= c("A", "A", "A", "B")
demo_scale <-c("num","nom","num")
hist_scale <-c(rep("nom",
length(hist_block)))
trts_scale <- c(rep("ord",2))
acs_scale <- c("ord")
outer_scale=list(demo_scale, hist_scale,
trts_scale, acs_scale)
```

Appel principal

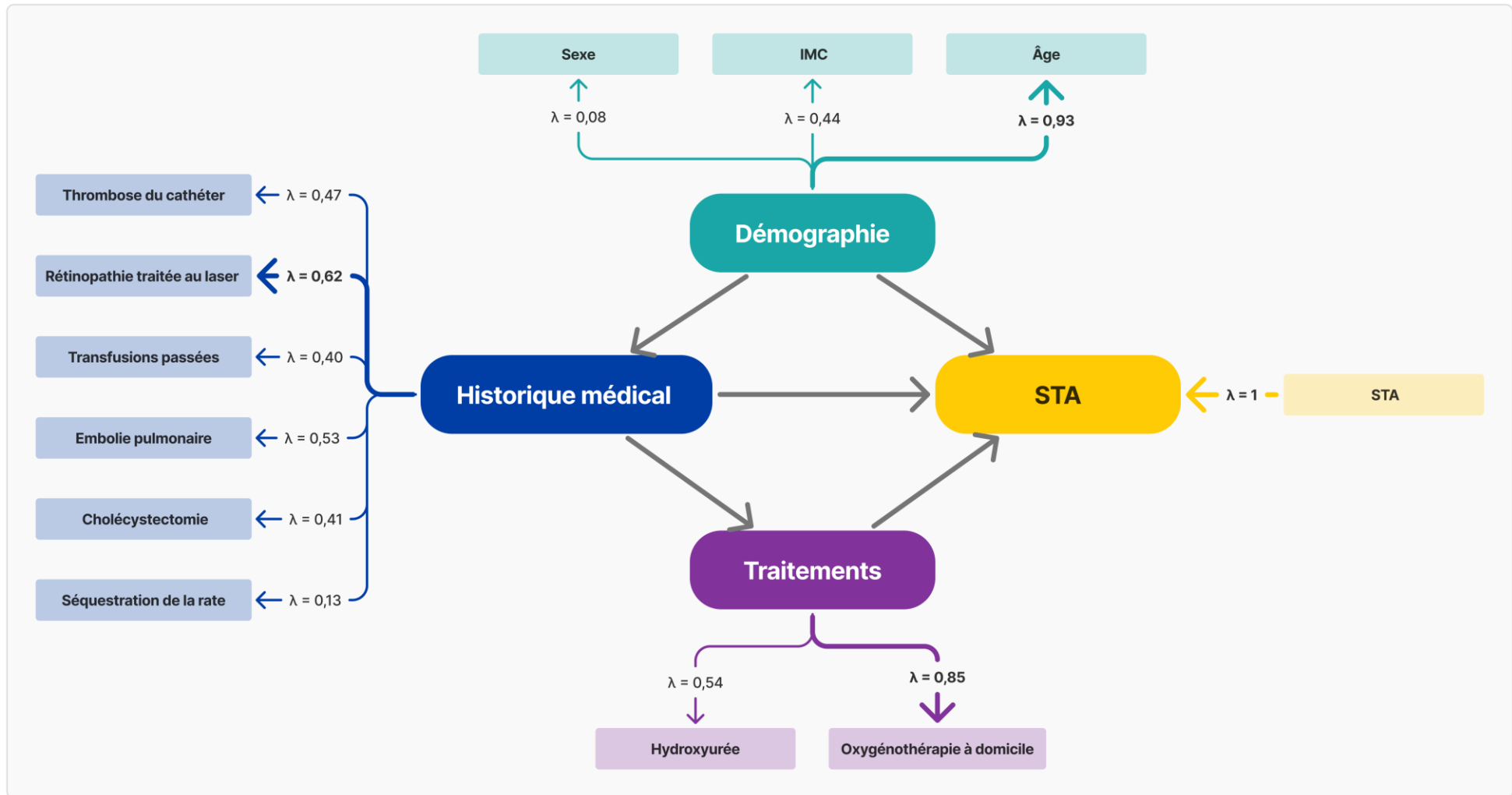
```
t1 <- proc.time()

mod_plspm_cc <- plspm(
  data_ssb_cc, inner_path_matrix,
  outer_path_blocks,
  modes = outer_modes, scaling =
  outer_scale,
  boot.val = TRUE, br = 100, scheme = "path"
)

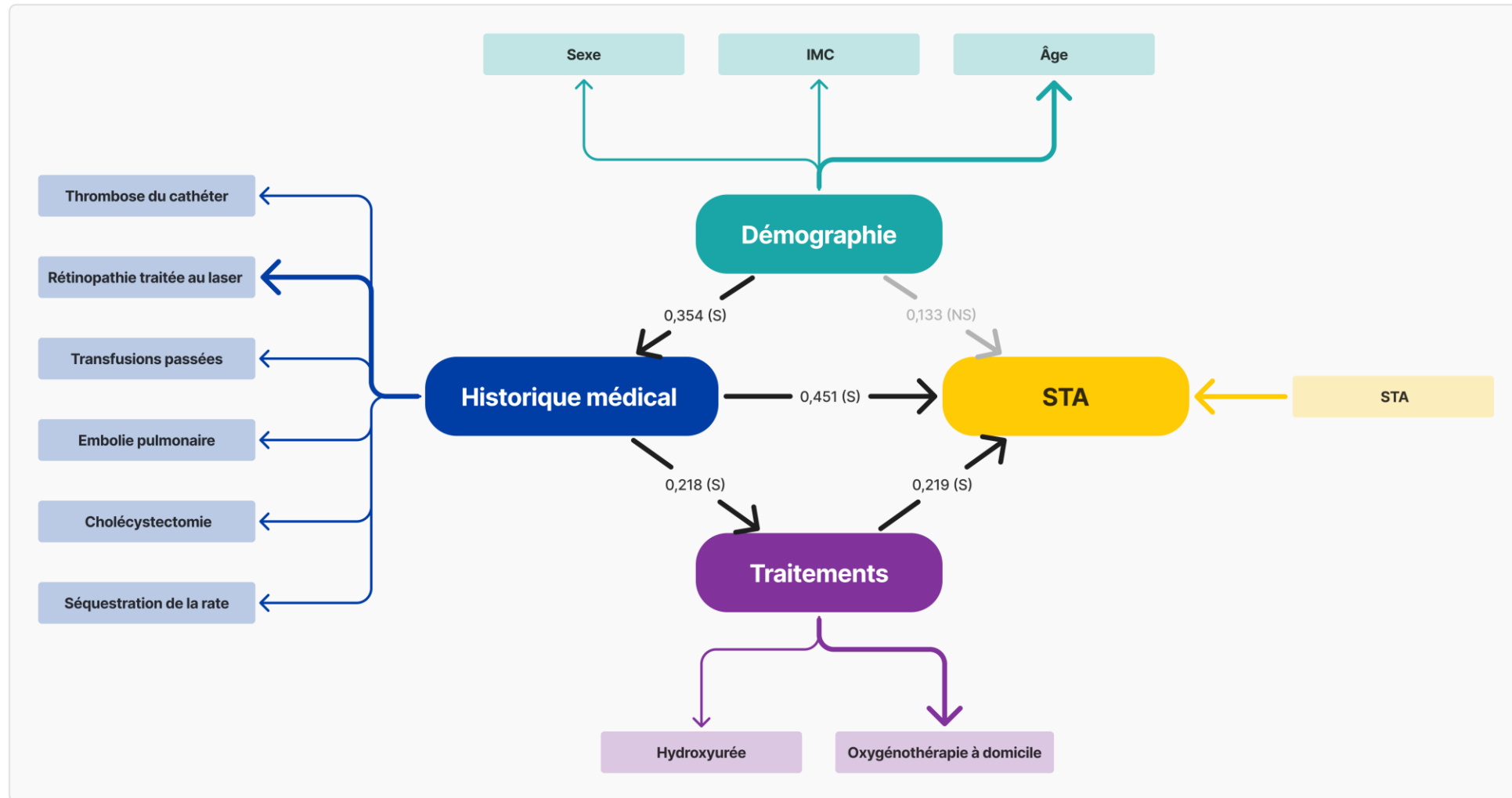
t2 <- proc.time()
t2 - t1

summary(mod_plspm_cc)
```

```
outerplot(mod_plspm_cc, what = "loadings")
```



```
innerplot(mod_plspm_cc, show.values = TRUE, box.size = 0.1)
```



Pour les blocs réflexifs (A),
→ vérifier l'unidimensionalité

	Mode	MVs	C.alpha	DG.rho	eig.1	eig.2
Démographie	A	3	0,3383	0,691	1,30	0,949
Historique médical	A	6	0,2353	0,554	1,30	1,152
Traitements	A	2	0,0413	0,676	1,02	0,979
STA	B	1	1,0000	1,000	1,00	0,000

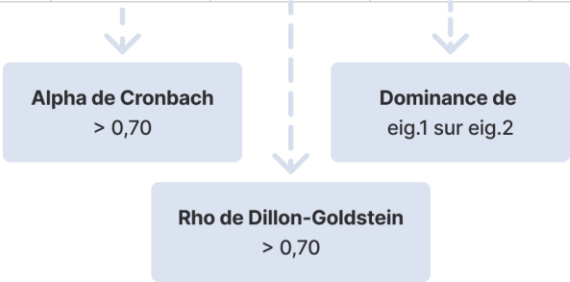
Pour les blocs formatifs (B),
→ vérifier la redondance

	Type	R2	Communalité du bloc	Redondance moyenne	AVE
Démographie	Exogène	0,0000	0,355	0,0000	0,355
Historique médical	Endogène	0,1253	0,206	0,0259	0,206
Traitements	Endogène	0,0475	0,509	0,0242	0,509
STA	Endogène	0,2208	1,000	0,2208	0,000

Goodness of fit
0.1991

Pour les blocs réflexifs (A),
→ vérifier l'unidimensionalité

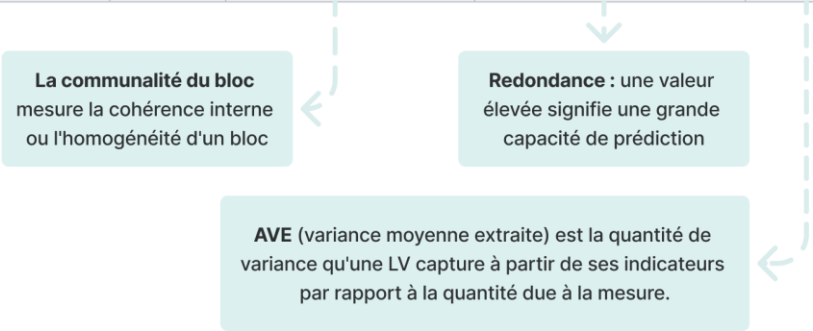
	Mode	MVs	C.alpha	DG.rho	eig.1	eig.2
Démographie	A	3	0,3383	0,691	1,30	0,949
Historique médical	A	6	0,2353	0,554	1,30	1,152
Traitements	A	2	0,0413	0,676	1,02	0,979
STA	B	1	1,0000	1,000	1,00	0,000



Goodness of fit
0.1991

Pour les blocs formatifs (B),
→ vérifier la redondance

	Type	R2	Communalité du bloc	Redondance moyenne	AVE
Démographie	Exogène	0,0000	0,355	0,0000	0,355
Historique médical	Endogène	0,1253	0,206	0,0259	0,206
Traitements	Endogène	0,0475	0,509	0,0242	0,509
STA	Endogène	0,2208	1,000	0,2208	0,000



Concernant l'application

Points forts

- Ce modèle permet de **quantifier** la contribution de chaque variable latente à la survenue du STA.

Limites

- Les variables latentes créées ont montré une **cohérence limitée** (particulièrement pour les antécédents médicaux), ce qui restreint l'interprétation possible de la contribution de ces variables (cela peut être lié à la petite taille de l'échantillon/aux données manquantes puisque l'analyse a été effectuée sur l'ensemble des cas complets, et au nombre et types limités de variables incluses par LV).

Concernant la méthode

Points forts

- La **PLS-PM** est une méthode utile qui réduit la dimensionnalité tout en structurant les relations entre variables, permettant ainsi de répondre au mieux aux questions scientifiques.
- Semblable aux SEM mais avec **moins d'hypothèses fortes** sur les données (structure de covariance).

Limites

- Suppose que toutes les relations entre les variables sont **linéaires**.



Une filiale de DOCAPOSTE

**Merci pour votre
attention**

 bd@hevaweb.com

 [@hevaweb.com](https://www.instagram.com/hevaweb.com)

 [linked.com/company/heva](https://www.linkedin.com/company/heva)

