

Méthodologie de construction de l'indicateur d'exposition au risque de précarité alimentaire

Note rédigée par :

- Alexandre Deprun, Ingénieur Data scientist, LARIIS-HETIS
- Claudia Ayika, Ingénieure Data scientist, LARIIS-HETIS
- Valeria Alfieri, Chargée de recherche, LARIIS-HETIS
- Julien Scheepers, responsable du LARIIS

En France, la précarité alimentaire est évaluée le plus souvent à travers les données remontées par les associations d'aide alimentaire, à savoir le nombre de bénéficiaires et les volumes distribués. Cette approche ne permet pas de rendre compte des déterminants de la précarité alimentaire, ni d'intégrer les situations de non-recours, raison pour laquelle des indicateurs ont commencé à voir le jour.

Un indicateur composite est un regroupement de variables pertinentes, triées à travers des méthodes statistiques et mathématiques, dont le but est de mesurer le risque et le niveau d'exposition au risque, d'un territoire donné à un phénomène donné. L'indicateur de précarité alimentaire mesure donc le risque d'exposition d'un territoire à l'insuffisance alimentaire. Ce qui signifie que les territoires présentant le risque le plus élevé n'ont pas forcément des hauts niveaux de précarité alimentaire, mais présentent plus de risques que d'autres territoires. Le risque est mesuré en fonction d'une moyenne de référence calculée à une échelle donnée (nationale, régionale, etc.).

Des indicateurs de précarité alimentaire existent et ont été expérimenté à différentes échelles en France, notamment en Ile-de-France et en Occitanie. Nous nous sommes appuyés sur trois séries de travaux de référence dans le champ de la précarité alimentaire :

- L'étude ABENA :
 - o Grange, D., Castetbon, K., Guibert, G., Vernay, M., Escalon, H., Delannoy, A., Féron, V., & Vincelet, C. (2013, mars). *Alimentation et état nutritionnel des bénéficiaires de l'aide alimentaire : Étude Abena 2011-2012 et évolutions depuis 2004-2005*. Observatoire régional de santé Île-de-France ; Institut de veille sanitaire ; Institut de prévention et d'éducation pour la santé.
- Les travaux conjoints de l'ANSA et du CREDOC :
 - o Agence nouvelle des solidarités actives (Ansa), & Centre de recherche pour l'étude et l'observation des conditions de vie (Crédoc). (2023, mars). *Diagnostic de la précarité alimentaire en Île-de-France : Document de synthèse*. Ansa ; Crédoc.

- L'Huillier, H. (2023, mars). *Diagnostic de la précarité alimentaire en Île-de-France : Revue de littérature*. Centre de recherche pour l'étude et l'observation des conditions de vie (Crédoc) ; Agence nouvelle des solidarités actives (Ansa).
- Bléhaut, M., Gressier, M., & Paquet, N. (2023, décembre). *La précarité alimentaire, en hausse, est liée à moins de diversité dans l'alimentation* (Cahier de recherche n° C357). Centre de recherche pour l'étude et l'observation des conditions de vie (CRÉDOC).
- Les travaux d'OBSOALIM (Observatoire des Solidarités Alimentaires de l'Hérault) :
 - Labarre, J., Néel, C., Perrin, C., & Bricas, N. (2022, septembre). *Une approche territoriale des facteurs de précarité alimentaire à l'aide de données en libre accès*. Communication présentée aux Rencontres « Pour des solidarités alimentaires » (22–23 septembre 2022).
 - Leur site internet apporte aussi des précisions depuis leur première version d'indicateur : <https://obso-alim.org/>

Ces études ne recouvrent pas toujours les mêmes thématiques ni les mêmes méthodologies. Nous avons fait le choix d'orienter notre travail vers la construction d'un indicateur de présomption de précarité alimentaire qui ne prendrait en compte que des variables socio-démographiques. D'autres indicateurs intègrent la question de l'offre de service dans leur calcul (ANSA-CREDOC). Nous avons préféré construire un indicateur d'indication du niveau de besoin potentiel avec la superposition d'une couche faisant apparaître l'offre (d'aide et généraliste). Ce choix s'inscrit dans la logique de la mission qui vise la coordination territoriale. Il s'agit de faciliter la mise à jour régulière du repérage des offres sans opérer de nouveaux calculs sur le niveau de risque d'exposition à la précarité alimentaire. Le niveau de de risque a été construit, comme dans la logique OBSOALIM, dans une logique d'accès open data aux variables mobilisées pour favoriser leur reproductibilité (notamment avec la mise à jour des données de recensement). Ces études nous ont servi essentiellement à mieux déterminer le choix des variables à retenir et à mieux comprendre les déterminants.

Nous avons retenu comme échelle de comparaison/référence, celle départementale car, s'agissant de mieux orienter les actions d'aide alimentaire dans les Bouches-du-Rhône, cette option nous a semblé la plus pertinente. Cela signifie qu'on ne peut pas comparer un territoire des Bouches-du-Rhône et un territoire d'un département différent dans la version actuelle. En effet, la situation d'une commune classée en « risque très élevé » dans un département cumulant plus de facteurs de risque (comme cela est le cas des Bouches-du-Rhône) ne sera pas la même que celle d'une commune classée en « risque très élevé » dans un département cumulant moins de facteurs de risque. Pour ce type de comparaison, il faudrait utiliser une échelle de référence nationale.

Nous avons également fait le choix de mesurer le risque de précarité alimentaire au niveau IRIS (infracommunal). En effet, plus l'échelle territoriale observée est petite, plus l'information donnée sera précise et détaillée, ce qui aide davantage la prise de décision et l'orientation des actions. Cependant, la fiabilité d'une information est fonction de l'accès aux données, certaines variables pouvant ne pas être accessibles à cause du secret statistique. Ce qui signifie que l'IRIS peut être considéré comme le

niveau d'analyse le plus pertinent pour répondre à des enjeux opérationnels à condition que le biais pouvant être engendré par le manque de données soit moindre (disponibilité des données au niveau le plus fin).

1. Méthodologie du scoring : comment avons-nous procédé ?

La construction d'un indicateur passe par plusieurs étapes qui peuvent se résumer comme suit :

- Choix et sélection des variables
- Finalisation du scoring

Choix des variables

Un indicateur fiable demande avant tout de déterminer les variables les plus pertinentes. Ce processus de tri s'opère en plusieurs phases :

- Pré-sélection des variables à partir d'études et enquêtes existantes
- Recensement des sources de données disponibles.
- Calcul des parts pour homogénéiser les données.
- Analyse des corrélations et de la redondance des variables.
- Application d'une analyse en composantes principales (ACP) pour optimiser la représentation des dimensions.
- Sélection finale des variables en fonction des besoins de l'étude.

Nous allons détailler ces étapes de manière simplifiée ci-dessous. Une note méthodologique technique est disponible sur demande.

- Sélection initiale des variables

La première étape consiste à repérer toutes les variables pertinentes à travers une analyse des études et enquêtes existantes en lien avec la précarité alimentaire (Obsalim, ANSA, ABENA). Ce faisant, nous avons identifié 35 variables qui ont toutes été prises en compte dans l'élaboration du scoring. Cette liste (ci-dessous) a constitué notre base de départ :

- **P21_POPH** : Population masculine
- **P21_POPF** : Population féminine
- **P21_POP1529** : Population âgée de 15 à 29 ans
- **P21_POP3044** : Population âgée de 30 à 44 ans
- **P21_POP4559** : Population âgée de 45 à 59 ans
- **P21_POP6074** : Population âgée de 60 à 74 ans
- **P21_POP75P** : Population âgée de 75 ans et plus
- **P21_ACT1564** : Population active de 15 à 64 ans

- **P21_CHOM1564** : Population au chômage de 15 à 64 ans
- **P21_ETUD1564** : Population en études de 15 à 64 ans
- **P21_AINACT1564** : Population inactive de 15 à 64 ans
- **P21_SAL15P_TP** : Salariés à temps partiel
- **P21_SAL15P_INTERIM** : Salariés en intérim
- **P21_SAL15P_CDD** : Salariés en CDD
- **P21_CHOM_DIPLMIN** : Chômeurs ayant un diplôme minimum
- **P21_NSCOL15P** : Population non scolarisée de 15 ans et plus
- **P21_NSCOL15P_DIPLMIN** : Population non scolarisée avec un diplôme minimum
- **P21_NSCOL15P_BEPC** : Population non scolarisée ayant au moins le BEPC
- **P21_NSCOL15P_CAPBEP** : Population non scolarisée ayant au moins un CAP/BEP
- **DISP_TP6021** : Taux de pauvreté au seuil de 60% du revenu disponible
- **DISP_PPSOC21** : Part de l'ensemble des prestations sociales
- **DISP_PPMINI21** : Part des minimas sociaux
- **DISP_PPFAM21** : Parts des prestations familiales
- **DISP_PPLOGT21** : Part des prestations logement
- **P21_RP_PROP** : Résidences principales occupées par des propriétaires
- **P21_RP_LOC** : Résidences principales occupées par des locataires
- **P21_NPER_RP_LOCHLMV** : Nombre de personnes des résidences principales HLM loués vides
- **HSTU1P** : résidences principales (hors studio de 1 personne)
- **P21_RP_VOIT1P** : Nombre de ménages avec au moins une voiture
- **RP_NPER** : Nombre de personnes en résidence principale
- **taille_moy_menage** : Taille moyenne des ménages
- **A** : Nombre d'allocataires d'aides sociales
- **PERCOU** : Nombre de personnes couvertes
- **AI** : Nombre d'allocataires isolés sans enfants
- **AM** : Nombre d'allocataires mono-parents

Le découpage territorial utilisé est le découpage par IRIS qui nous permet d'identifier les disparités locales à une échelle plus fine que la commune et de cibler ainsi avec plus de précision les secteurs où la précarité est la plus marquée. Ce choix a été possible car nous avons pu avoir accès aux sources des données.

- **Sources des données**

Les données utilisées proviennent principalement des sources suivantes :

- **INSEE** : Recensement de la population, bases locales de revenus déclarés
- **CAF** : Données sur les prestations sociales et les bénéficiaires.

Nous avons ensuite procédé au calcul des parts, c'est-à-dire à l'estimation du pourcentage de la population concernée par chaque variable afin de garantir leur comparabilité.

Prenons l'exemple des chômeurs âgés de 15 à 64 ans. Plutôt que de considérer uniquement le nombre absolu de chômeurs dans une zone donnée, nous avons calculé leur proportion par rapport à l'ensemble de la population en âge de travailler (15-64 ans). Cette approche est essentielle pour comparer différentes zones de manière pertinente.

En effet, une commune peut avoir un grand nombre de chômeurs en chiffre brut, mais si elle a aussi une population très nombreuse, la part des chômeurs par rapport à la population active peut rester faible. À l'inverse, une petite commune avec moins de chômeurs en valeur absolue peut avoir un taux de chômage plus élevé si la proportion de chômeurs par rapport à la population en âge de travailler est plus importante.

En utilisant des pourcentages plutôt que des chiffres bruts, on obtient donc une comparaison plus fiable entre les territoires, quelles que soient leurs tailles.

- **Analyse des corrélations et réduction des variables**

L'analyse des corrélations consiste à identifier les variables les plus pertinentes en opérant un tri parmi toutes les variables recensées à travers des calculs mathématiques. Cela permet d'éviter les biais engendrés par une trop grande proximité de certaines variables, ce qui produirait des redondances dans les valeurs.

Gestion des variables manquantes

Afin d'assurer la cohérence et la comparabilité des données et d'avoir des résultats le plus précis possible, nous avons avant toute chose procédé au traitement des valeurs manquantes. En effet, certaines variables présentent un taux de valeurs manquantes de 11,4% en raison du secret statistique. Autrement dit, certaines données ne sont pas accessibles dans un certain nombre de IRIS étant protégées par le secret statistique qui assure l'anonymat des personnes concernées.

Les variables en question sont les suivantes :

- **DISP_TP6021** : Taux de pauvreté au seuil de 60% du revenu disponible
- **DISP_PPSOC21** : Part de l'ensemble des prestations sociales
- **DISP_PPMINI21** : Part des minimas sociaux
- **DISP_PPFAM21** : Parts des prestations familiales

Étant donné que le taux de valeurs manquantes reste relativement faible (très peu d'IRIS concernés), nous avons choisi une imputation par la moyenne. Cette approche mathématique consiste à remplacer chaque valeur manquante d'une variable par la moyenne des valeurs existantes de cette même variable. Simple et efficace, cette méthode permet de limiter la perte d'information tout en

maintenant la structure statistique des données, garantissant ainsi une cohérence optimale de l'échantillon pour les analyses à venir.

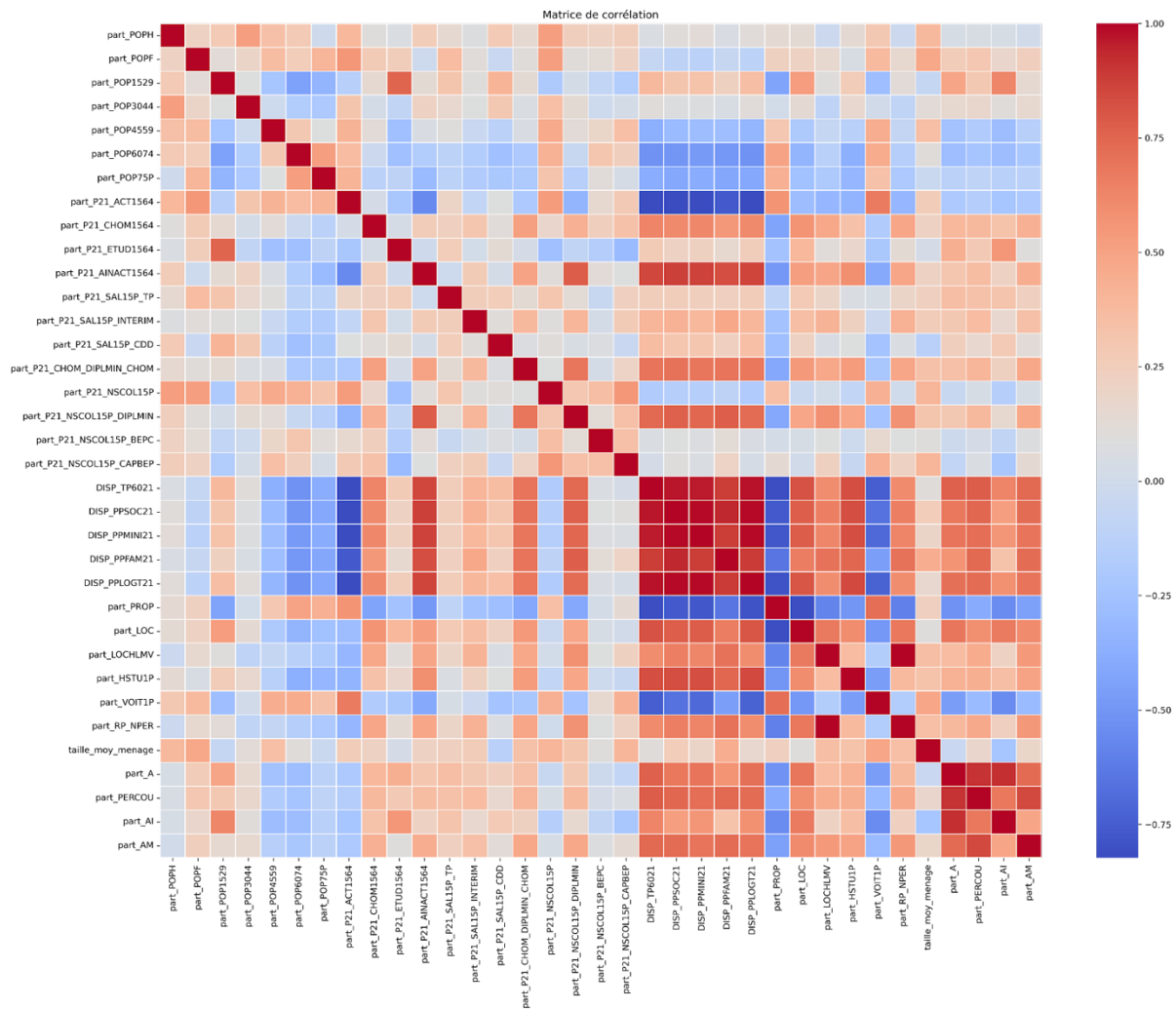
Toutefois, après la sélection des variables, seul le taux de pauvreté au seuil de 60% du revenu disponible (**DISP_TP6021**) a été retenu et intégré dans le calcul du scoring. Ainsi, bien que plusieurs variables aient fait l'objet d'une imputation, l'impact potentiel du biais introduit par l'imputation reste minimisé dans l'analyse finale.

Les autres variables de l'analyse étaient entièrement renseignées et n'ont pas nécessité de traitements particuliers.

A la suite de cela, une analyse des corrélations a été réalisée pour identifier les relations entre variables et éviter les redondances.

Corrélation et réduction des variables

L'analyse de la corrélation des variables permet d'identifier, parmi un certain nombre de variables qui se ressemblent, la variable la plus pertinente pour participer au calcul du scoring. Pour commencer, nous avons défini un seuil de corrélation (80%) au-delà duquel deux variables sont considérées comme trop similaires, car elles varient de la même manière. Nous avons ensuite réalisé une matrice des corrélations.



En parcourant chaque colonne de la matrice, nous récupérons toutes les variables ayant une corrélation supérieure au seuil défini. Les variables repérées sont ensuite réparties en groupes corrélés, ce qui nous permet d'avoir un traitement plus affiné.

Un groupe corrélé est un ensemble de variables qui varient de manière similaire, c'est-à-dire que lorsqu'une variable du groupe augmente ou diminue, les autres ont tendance à évoluer dans le même sens. Ce groupe est défini de manière à ce que les variables à l'intérieur soient plus liées entre elles qu'avec les variables d'un autre groupe. Les variables d'un même groupe partagent une information similaire et sont donc redondantes.

En guise d'illustration, les variables **DISP_TP6021**, **DISP_PPSOC21** et **DISP_PPMINI21** sont fortement corrélées entre elles. Elles appartiendront donc au même groupe corrélé.

Pour chaque groupe corrélé, une méthode de calcul mathématique, à savoir l'**ACP** (Analyse en Composantes Principales) est réalisée. Ce qui nous permet d'identifier la variable avec le poids le plus élevé et qui par conséquent représente le mieux l'information contenue dans l'ensemble du groupe corrélé. C'est avec cette variable, considérée comme la plus pertinente du groupe que nous poursuivrons la construction de l'indicateur. Les autres variables du groupe sont donc abandonnées

car nous n'en avons désormais plus besoin. Cette procédure a donc été réalisée pour l'ensemble des groupes corrélés.

Groupe corrélé	Variable retenue
DISP_TP6021, DISP_PPSOC21	DISP_TP6021
DISP_TP6021, DISP_PPSOC21, DISP_PPMINI21	DISP_TP6021
DISP_TP6021, DISP_PPSOC21, DISP_PPMINI21, DISP_PPFAM21	DISP_TP6021
DISP_TP6021, DISP_PPSOC21, DISP_PPMINI21, DISP_PPFAM21, DISP_PPLOGT21	DISP_TP6021
part_PROP, part_LOC	part_LOC
part_LOCHLMV, part_RP_NPER	part_RP_NPER
part_A, part_PERCOU	part_PERCOU
part_A, part_AI	part_A
part_PERCOU, part_AM	part_PERCOU

Les variables retenues au sein des groupes corrélés sont donc : **DISP_TP6021, part_LOC, part_RP_NPER, part_PERCOU, part_A.**

A l'issue de toute cette étape d'analyse de corrélations et de réduction des variables, la liste initiale de **35 variables** a été réduite à **26 variables** qui sont :

- **part_POPH** : part d'hommes
- **part_POPF** : part de femmes
- **part_POP6074** : part des 60 à 74 ans
- **part_POP3044** : part des 30 à 44 ans
- **part_POP75P** : part des 75 ans et plus
- **part_POP1529** : part des 15 à 29 ans
- **part_POP4559** : part des 45 à 59 ans
- **part_P21_SAL15P_TP** : Part de personnes salariées de 15 ans ou plus à temps partiel
- **part_P21_SAL15P_INTERIM** : Part de personnes salariées de 15 ans ou plus intérimaires
- **part_PERCOU** : Part de personnes couvertes

- **part_P21_ACT1564** : parts d'actifs de 15 à 64 ans
- **part_P21_CHOM_DIPLMIN_CHOM** : part de chômeurs ayant un diplôme minimum
- **DISP_TP6021** : Taux de pauvreté au seuil de 60 % du revenu disponible
- **part_VOIT1P** : part des ménages avec au moins une voiture
- **taille_moy_menage** : taille moyenne des ménages
- **part_P21_ETUD1564** : Part d'élèves, étudiants et stagiaires non rémunérés de 15 à 64 ans
- **part_RP_NPER** : part de personnes en résidence principale
- **part_P21_NSCOL15P_DIPLMIN** : Part de personnes non scolarisées de 15 ans ou plus titulaires d'aucun diplôme ou au plus un CEP
- **part_P21_SAL15P_CDD** : Part de personnes salariées de 15 ans ou plus ayant un contrat à durée déterminée
- **part_P21_NSCOL15P** : part de personnes de 15 ans ou plus non scolarisées
- **part_P21_NSCOL15P_CAPBEP** : non scolarisée ayant au moins le CAP/BEPC
- **part_LOC** : part de locataires
- **part_P21_AINACT1564** : Part d'autres inactifs de 15 à 64 ans
- **part_P21_NSCOL15P_BEPC** : part des personnes non scolarisée ayant au moins le BEPC
- **part_P21_CHOM1564** : Part de chômeurs de 15 ans à 64 ans titulaires d'aucun diplôme ou au plus un CEP
- **part_HSTU1P** : part des résidences principales (hors studio de 1 personne)

Sélection finale des variables

A partir de la liste de ces 26 variables qui décrivent divers aspects socio-économiques et démographiques de la population, nous avons effectué un dernier tri manuel afin d'éliminer celles qui nous semblaient moins directement liées à notre problématique.

Nous avons ainsi écarté certaines variables qui, bien qu'informatives, n'apportent pas de valeur ajoutée déterminante pour le scoring. Par exemple, les variables décrivant la répartition des tranches d'âge (part des 15-29 ans, 30-44 ans, 45-59 ans, etc.) ont été supprimées car elles ne sont pas directement représentatives d'une situation de vulnérabilité alimentaire.

Certaines tranches d'âge peuvent être plus touchées par la vulnérabilité alimentaire (par exemple, les jeunes adultes ou les personnes âgées), mais cela dépend souvent d'autres facteurs, comme le niveau de revenus, la situation professionnelle ou l'accès aux aides sociales. C'est pourquoi nous avons préféré sélectionner des variables qui décrivent ces facteurs de manière plus explicite.

De même, certaines données relatives aux caractéristiques du logement, comme la part des ménages possédant au moins une voiture ou la taille moyenne des ménages, ont été jugées moins pertinentes que des indicateurs plus directement liés à la précarité, comme la part des ménages monoparentaux ou la proportion de locataires en HLM.

De plus, bien que la part des personnes couvertes (part_PERCOU) ait été mathématiquement retenue, la part des allocataires (part_A) est un meilleur choix conceptuel et méthodologique pour identifier les zones où la précarité économique (et donc la vulnérabilité alimentaire) est la plus marquée.

En effet, la part des personnes couvertes regroupe l'ensemble des personnes vivant dans un foyer où au moins une personne est allocataire de la CAF, incluant l'allocataire principal, son conjoint et ses enfants. Par exemple, dans une famille avec deux parents et deux enfants où un seul parent perçoit des allocations, les quatre membres du foyer sont considérés comme "couverts". Cela donne une vision large des bénéficiaires indirects d'aides sociales, sans pour autant distinguer ceux réellement en situation de précarité. À l'inverse, la part des allocataires ne prend en compte que les personnes recevant directement des prestations (RSA, APL, AAH, etc.), ce qui en fait un indicateur plus précis pour la précarité économique. Par exemple, une femme seule avec deux enfants qui perçoit le RSA et l'APL est une allocataire, mais ses enfants ne le sont pas. Ainsi, en remplaçant la part des personnes couvertes par la part d'allocataires, on affine l'analyse en ciblant uniquement les personnes directement bénéficiaires d'aides sociales, qui sont plus susceptibles d'être en situation de vulnérabilité alimentaire.

À l'issue de cette dernière sélection, **13 variables finales ont été retenues** pour le calcul du scoring :

- **part_MENPSEUL** : Part de ménages d'une personne
- **part_MENFAMMONO** : Part de ménages dont la famille principale est une famille monoparentale
- **part_P21_CHOM1564** : Part de chômeurs de 15 à 64 ans
- **part_P21_CHOM_DIPLMIN_CHOM** : Part de chômeurs de 15 ans à 64 ans titulaires d'aucun diplôme ou au plus un CEP
- **part_P21_AINACT1564** : Part d'autres inactifs de 15 à 64 ans
- **part_P21_SAL15P_TP** : Part de personnes salariées de 15 ans ou plus à temps partiel
- **part_P21_SAL15P_INTERIM** : Part de personnes salariées de 15 ans ou plus intérimaires
- **part_P21_SAL15P_CDD** : Part de personnes salariées de 15 ans ou plus ayant un contrat à durée déterminée
- **part_P21_NSCOL15P_DIPLMIN** : Part de personnes non scolarisées de 15 ans ou plus titulaires d'aucun diplôme ou au plus un CEP
- **part_P21_ETUD1564** : Part d'élèves, étudiants et stagiaires non rémunérés de 15 à 64 ans
- **part_A** : Part d'allocataires
- **DISP_TP6021** : Taux de pauvreté au seuil de 60 % du revenu disponible
- **part_LOCHLMV** : Part de résidences principales HLM loué vide

Finalisation du scoring

L'objectif de ce scoring est d'évaluer le niveau de vulnérabilité alimentaire des IRIS en combinant les 13 variables sélectionnées. Pour cela, plusieurs étapes sont suivies :

- Choix de la méthode de calcul du score : Agrégation des variables par une moyenne.
- Standardisation des variables pour assurer une contribution équitable de chacune.
- Calcul du score final par moyenne des valeurs standardisées.
- Regroupement des IRIS en classes homogènes grâce à la méthode de Jenks et choix du nombre optimal de groupes.

- **Méthode de calcul du scoring : la moyenne**

Pour obtenir un score global de précarité alimentaire par IRIS, il est nécessaire d'agréger les informations fournies par les 13 variables sélectionnées. La méthode choisie est la moyenne des valeurs de ces variables.

Cette méthode est simple et intuitive, offrant un indicateur global qui reflète la situation de chaque IRIS. Elle permet de considérer toutes les variables de manière équivalente sans donner plus de poids à l'une ou l'autre, tout en facilitant l'interprétation et la comparaison entre les territoires.

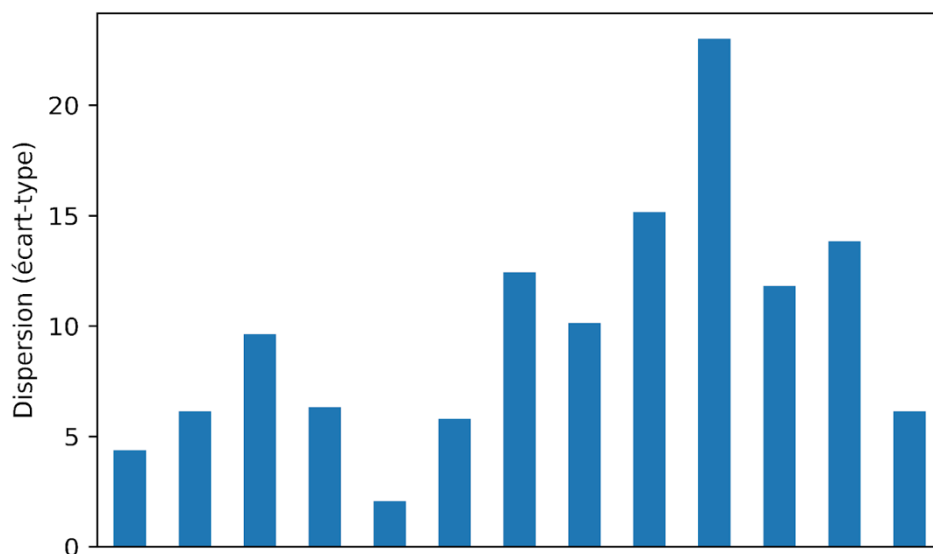
Cependant, avant d'effectuer ce calcul, il est nécessaire de s'assurer que toutes les variables sont sur une même échelle de mesure pour éviter qu'une variable plus dispersée n'ait un impact disproportionné sur le score final. C'est pourquoi une standardisation est appliquée.

- **Standardisation des variables**

Les variables sont exprimées en pourcentage, mais leur dispersion varie fortement d'un IRIS à l'autre. Certaines variables peuvent avoir une amplitude de variation importante, tandis que d'autres restent relativement stables.

Si nous calculons directement la moyenne des variables brutes, celles ayant une plus grande dispersion auront plus d'influence sur le score final car leurs valeurs varient plus que les autres.

Afin de corriger le biais introduit par les différences de dispersion entre les variables, et pour assurer que chaque variable contribue équitablement au score final, nous appliquons une standardisation. C'est une méthode qui consiste à transformer les variables pour qu'elles soient toutes comparables sur une échelle commune, indépendamment de leur plage de dispersion.



Dispersion des variables

Les variables sont exprimées en pourcentage, mais leur dispersion peut varier fortement d'un IRIS à l'autre. Certaines variables peuvent avoir une amplitude de variation importante, tandis que d'autres restent relativement stables. La figure suivante illustre ce phénomène (cf Figure ci-dessus)

Le graphique représente les écarts-types qui mesurent la dispersion autour de la moyenne pour chaque variable. En calculant simplement la moyenne des valeurs de ces variables, celles qui ont une plus grande dispersion, donc un écart-type élevé, auront une influence beaucoup plus forte sur le résultat final.

Pour éviter ce biais, une standardisation est appliquée. Cette transformation mathématique permet de rendre toutes les variables comparables, indépendamment de leur plage de valeurs. Elle repose sur le calcul d'une nouvelle valeur pour chaque variable, en fonction de sa moyenne et de son écart-type. Après standardisation, chaque variable a une moyenne de 0 et un écart-type de 1. Cela garantit que toutes contribuent de manière équilibrée au calcul final du score.

Calcul du scoring

Une fois les variables standardisées, la moyenne des valeurs standardisées est calculée pour un IRIS donné en sommant les valeurs standardisées de chaque variable, puis en divisant cette somme par 13 selon la formule suivante : $\frac{1}{13} \sum_{i=1}^{13} Z_i$ où Z_i représente la valeur standardisée de chaque variable pour un IRIS donné.

Ce score final permet d'évaluer la vulnérabilité alimentaire de chaque IRIS :

- Un score élevé signifie que l'IRIS est fortement exposé à la vulnérabilité alimentaire.
- Un score faible indique que l'IRIS est moins exposé à la vulnérabilité alimentaire

- **Regroupement des IRIS en classes homogènes**

Pour faciliter l'interprétation des résultats, les IRIS ont été regroupés en **6 classes** en fonction de leur score de précarité alimentaire.

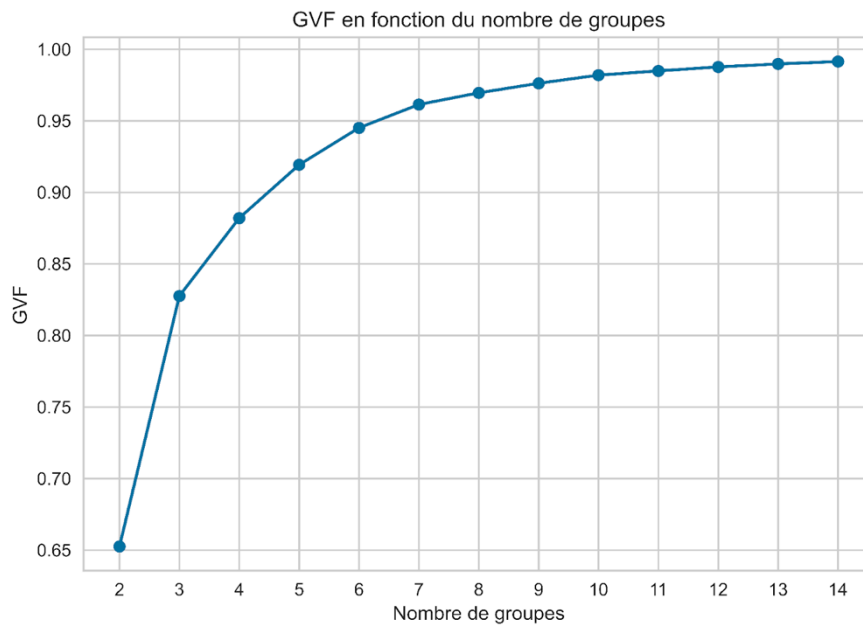
Pour effectuer cette classification, nous avons utilisé la **méthode des seuils naturels de Jenks**, qui est une méthode mathématique permettant d'optimiser la répartition des données en groupes homogènes.

Cette méthode fonctionne en recherchant les seuils les plus adaptés pour diviser les IRIS en groupes de manière à ce que :

- Les IRIS appartenant à un même groupe aient des scores aussi proches que possible les uns des autres (c'est-à-dire minimiser la variance à l'intérieur des groupes).
- Les différences de score entre les groupes soient maximisées, afin d'avoir une séparation nette entre les niveaux de précarité alimentaire.

Autrement dit, l'algorithme analyse les scores et identifie les coupures naturelles dans la distribution des données, permettant ainsi d'éviter des regroupements arbitraires ou déséquilibrés. Grâce à cette méthode, chaque groupe obtenu représente un niveau distinct de précarité alimentaire, allant des IRIS les moins exposés aux IRIS les plus vulnérables.

Pour évaluer la qualité de la classification, un indicateur statistique appelé **GVF** (Goodness of Variance Fit) a été utilisé. Cet indicateur mesure la capacité de la classification à réduire la variance au sein des groupes et à maximiser la différence entre eux.



En testant différents nombres de groupes (de 2 à 14), nous avons observé que le GVF augmente fortement jusqu'à un certain seuil, puis s'améliore très peu avec l'ajout de nouveaux groupes.

- Avec 2 à 5 groupes, la classification n'était pas assez fine. Ce qui ne permettait pas de bien différencier les niveaux de précarité alimentaire.
- Au-delà de 6 groupes, l'amélioration du GVF était minime, indiquant que l'ajout de nouvelles classes n'apportent pas d'information significative supplémentaire.

Ainsi, 6 classes ont été retenus, car ils permettent un bon compromis entre précision et lisibilité des résultats :

- 0 = Exposition non significative
- 1 = Très faible exposition
- 2 = Faible exposition
- 3 = Exposition modérée
- 4 = Forte exposition
- 5 = Très forte exposition

2. Précautions de lecture

Comme évoqué plus haut, les IRIS les plus exposés sont ceux plus à risque de précarité alimentaire compte tenu de la distribution des variables sélectionnées par rapport aux variables des autres IRIS du département. **Il n’y a pas d’adéquation parfaite entre les niveaux de risque et le nombre exact de personnes en situation réelle de précarité alimentaire** dans le sens où des populations concernées par la précarité alimentaire peuvent exister dans les territoires à faible risque ; et inversement tous les habitants d’un territoire fortement exposé ne sont pas en situation de précarité alimentaire. En effet, l’indicateur étant construit au niveau départemental, seules les comparaisons entre IRIS ou communes du département peuvent faire sens d’un point de vue analytique. Si nous changeons d’échelle d’analyse et nous nous situons au niveau régional ou national, il est fort probable qu’un territoire classé comme « très faiblement exposé » au risque de précarité alimentaire dans les Bouches-du-Rhône devienne plus exposé dans ce nouveau référentiel.

Cette précision est importante car nous supposons que si des actions d’aide alimentaire sont en cours sur un territoire apparaissant comme faiblement exposé, cela signifie potentiellement qu’elles répondent à des besoins.

Au sujet de l’analyse de la distribution de l’offre d’aide alimentaire sur le département par rapport aux résultats de l’indicateur, il est plus pertinent d’interroger l’absence d’actions sur des territoires très exposés que l’existence d’offres sur des territoires peu exposés.

Enfin, pour une lecture plus complète de l’estimation du niveau de besoin il sera nécessaire d’articuler le niveau de risque et le nombre d’habitants de l’IRIS (informations disponibles dans les tableaux de bord dynamiques).