

RAPPORT METHODOLOGIQUE

VERSION 3

ÉDITION 2024

PRESENTATION DES INTERVENANTS

La méthodologie mise en œuvre dans les études par catégorie alimentaire (dites études sectorielles), réalisées par l'Oqali, a été co-construite par les équipes Anses et INRAE de l'Oqali.

Rédaction du rapport

Blandine GIRAUDEAU – INRAE

Ghislaine NARAYANANE – INRAE

Marianne TRICHARD – INRAE

Révision du rapport

Olivier ALLAIS – INRAE

Shakila BOYER – INRAE

Johanna CALVARIN – INRAE

Aurore COUDRAY-Anses

Julie GAUVREAU-BEZIAT - Anses

Marine LANDREAU – INRAE

Hélène MATHIOT –Anses

Odeline MOLLE – INRAE

Anaëlle RABES - Anses

Karine VIN – Anses

Remerciements à Patrice Bertail (UMR Modal'X, CNRS - Université Paris-Nanterre) et Yannick Guyonvarch (UMR PSAE, INRAE - AgroParisTech) pour leurs expertises statistiques et leurs aides dans le développement des méthodologies mises en œuvre par l'Oqali.

VERSIONS :

Date	Numéro de version	Principaux changements
2025	3	Modifications partie 4.2 pour préciser le mode de calcul de l'effectif minimum, en distinguant le nombre de nutriments d'intérêt du secteur et celui de la famille, selon le type d'analyse
2025	2	Modifications partie 4.2 pour préciser le nombre d'échantillon minimum apparié nécessaire pour réaliser les analyses
2024	1	Création du document

SOMMAIRE

1	INTRODUCTION	1
1.1	Récapitulatif des paramètres statistiques de base.....	1
1.2	Tests et méthodes statistiques utilisés	3
2	TESTS DU CHI-2 D'INDEPENDANCE	4
2.1	Explication générale du test.....	4
2.2	Exemple appliqué à l'Oqali	4
3	TESTS DE PERMUTATION.....	7
3.1	Explication générale du test.....	7
3.2	Exemples appliqués à l'Oqali	7
3.2.1	Echantillons non appariés.....	7
3.2.2	Echantillons non appariés avec pondération	9
3.2.3	Echantillons appariés	11
4	METHODE DES DOUBLES DIFFERENCES	13
4.1	Explication générale de la méthode	13
4.2	Exemple appliqué à l'Oqali	14
5	CONTRIBUTION DE L'OFFRE ET DE LA DEMANDE A L'EVOLUTION DE LA QUALITE NUTRITIONNELLE	18
5.1	Explication générale de la méthode	18
5.2	Exemple appliqué à l'Oqali	21

Liste des figures

Figure 1 : Schéma d'une boîte à moustache ou boxplot.....	2
Figure 2 : Schéma expliquant la méthode des doubles différences.....	13
Figure 3 : Détail de calcul de la méthode des doubles différences	17
Figure 4 : Décomposition de l'évolution des teneurs moyennes pondérées en nutriments d'intérêt entre tn1 et tn	18
Figure 5 : Illustration de la méthode 1 de transformation des données.....	20
Figure 6 : Illustration de la méthode 2 de transformation des données.....	20
Figure 7 : Table de distribution du χ^2	24

Liste des tableaux

Tableau 1 : Distribution observée des produits étudiés par année de collecte et présence de portion indiquée.....	4
Tableau 2 : Distribution théorique des produits étudiés par année de collecte et présence de portion indiquée, sous l'hypothèse d'indépendance H0 (tableau dit de « contingence »)	5
Tableau 3 : Observations de la teneur en sucres des produits du groupe A (2010) et du groupe B (2016)	8
Tableau 4 : Teneurs en sucres et parts de marché des produits du groupe A (avec Nutri-Score) et du groupe B (sans Nutri-Score)	10
Tableau 5 : Teneurs en sucres, parts de marché et nombres de réplication associés, pour les produits du groupe A (avec Nutri-Score) et du groupe B (sans Nutri-Score).....	10
Tableau 6 : Teneurs en sucres des produits du groupe A (avec Nutri-Score) et du groupe B (sans Nutri-Score), après réplication.....	11
Tableau 7 : Teneurs en protéines des couples de produits appariés en 2016 (groupe A) et 2020 (groupe B).....	12
Tableau 8 : Exemple de permutation des teneurs en protéines des couples de produits appariés en 2016 (groupe A) et 2020 (groupe B).....	12
Tableau 9 : Teneurs moyennes en sucres des produits appariés avec et sans Nutri-Score en 2016 et 2020	16
Tableau 10 : Teneurs en matières grasses et parts de marché fictives des produits d'un secteur donné, pour 2010 (T0) et 2016 (T1)	21
Tableau 11 : Teneurs en matières grasses et parts de marché fictives des produits d'un secteur donné, pour 2010 (T0) et 2016 (T1), après inférence des valeurs nutritionnelles et transformation des parts de marché avec la méthode 1	22
Tableau 12 : Contributions des reformulations, renouvellement de l'offre et substitutions réalisées par les consommateurs (en pourcentage) à l'évolution de la teneur moyenne pondérée en matières grasses du secteur, entre 2016 et 2020 (méthode 1) – données fictives	23

Liste des Annexes

Annexe 1 : Table de distribution du χ^2	24
Annexe 2 : Détail de l'algorithme utilisé par Griffith et al. (2017) adapté aux données Oqali.....	25
Annexe 3 : Teneurs en matières grasses et parts de marché fictives des produits d'un secteur donné, pour 2010 (T0) et 2016 (T1), après inférence des valeurs nutritionnelles et transformation des parts de marché avec la méthode 2.....	26

1 INTRODUCTION

Les études par catégorie alimentaire (dites études sectorielles), réalisées par l'Oqali, ont pour objectif de suivre l'évolution de l'offre, des paramètres d'étiquetage et de la composition nutritionnelle, au cours du temps. Elles sont réalisées sur un ensemble de produits dont la représentativité du marché est estimée d'après les données Kantar Panel WorldPanel¹.

Les indicateurs s'appuient sur des paramètres statistiques de base (cf. partie 1.1.), des tests et méthodes statistiques sont mises en œuvre (cf. partie 1.2).²

1.1 Récapitulatif des paramètres statistiques de base

Les paramètres statistiques permettent de résumer, à l'aide de quelques valeurs clés, l'essentiel de l'information donnée par l'observation d'une variable quantitative (c'est-à-dire pouvant se traduire par une variable numérique). Par exemple, dans les études Oqali, ces paramètres sont calculés pour décrire au mieux les distributions des valeurs nutritionnelles des produits collectés.

- La **moyenne** se calcule en prenant la somme de toutes les données de la série et en la divisant par le nombre de données de la série.
- La **médiane** est le seuil pour lequel on retrouve un effectif égal de part et d'autre de celui-ci : autrement dit, 50% des valeurs d'une série sont inférieures ou égales à la médiane et 50% sont supérieures ou égales.
- Les **quartiles** sont les valeurs qui séparent un ensemble de données placées en ordre croissant :
 - Le 1^{er} quartile Q_1 est la plus petite valeur de la série telle qu'au moins 25% des valeurs sont inférieures ou égales à Q_1 ;
 - Le 2^e quartile Q_2 sépare la distribution en deux parties égales : il s'agit de la médiane.
 - Le 3^e quartile Q_3 est la plus petite valeur de la série telle qu'au moins 75% des valeurs sont inférieures ou égales à Q_3 .
- La **variance**, habituellement notée s^2 ou V , est une mesure de la dispersion des valeurs d'un échantillon autour de la moyenne. C'est la moyenne des carrés des écarts à la moyenne.
- L'**écart-type**, habituellement noté s , correspond à la racine carrée de la variance et mesure également la dispersion des données autour de la moyenne. Il est cependant plus aisé d'interpréter un écart-type que la variance car il s'exprime dans la même unité que la donnée étudiée : c'est donc ce paramètre qui est présenté dans les études de l'Oqali.

L'ensemble de ces paramètres peut être représenté graphiquement sur une boîte à moustache, ou boxplot, comme l'illustre la Figure 1 ci-dessous.

¹ Kantar – Panel Worldpanel : données d'achats des ménages représentatives de la population française.

² Ce rapport décrit les traitements statistiques mis en œuvre pour les études sectorielles d'évolution. Les traitements statistiques relatifs aux états des lieux des secteurs alimentaires sont décrits dans le *rapport méthodologique- Edition 2009*.

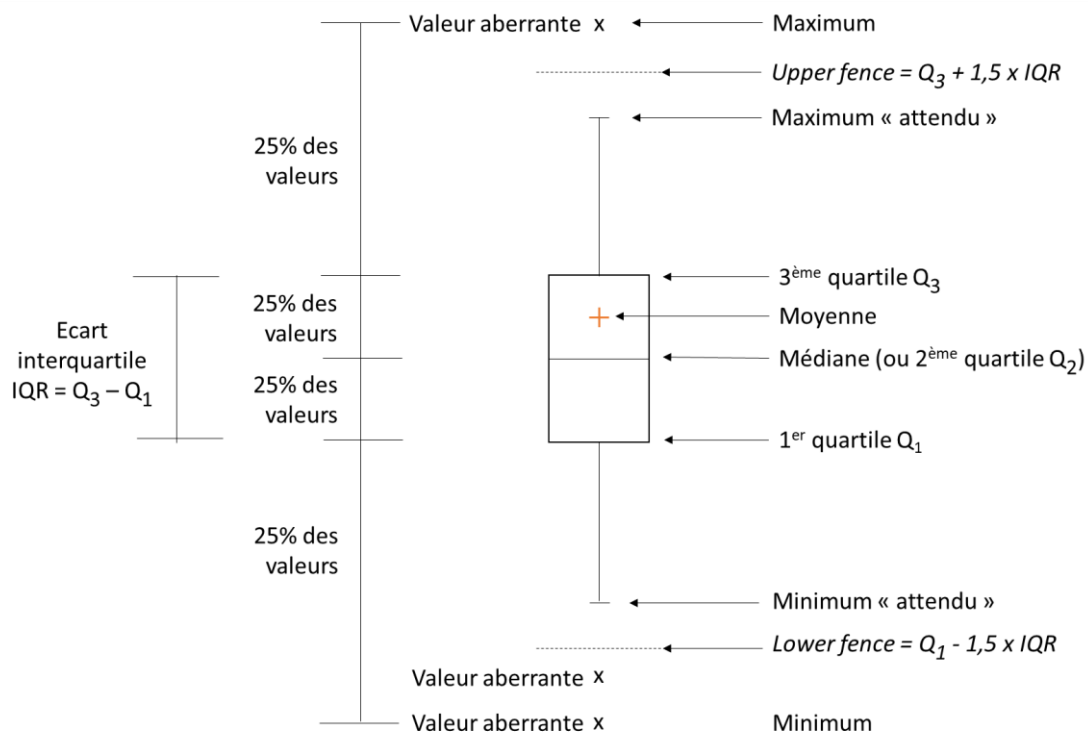


Figure 1 : Schéma d'une boîte à moustache ou boxplot³

A noter que, sur la Figure 1 :

- Les lignes qui s'étendent de part et d'autre de la boîte sont appelées « moustaches » et représentent la variation attendue des données. Elles sont calculées en utilisant 1.5 fois l'écart interquartile IQR (soit $Q_3 - Q_1$) :
 - Le maximum « attendu » correspond à la 1^{ère} observation de l'échantillon située juste en-dessous de l'*upper fence* (valeur théorique correspondant à : $Q_3 + 1,5 \times IQR$) ;
 - Le minimum « attendu » correspond à la 1^{ère} observation de l'échantillon située juste au-dessus du *lower fence* (valeur théorique correspondant à : $Q_1 - 1,5 \times IQR$).
- Les valeurs inférieures au minimum « attendu » ou supérieures au maximum « attendu », symbolisées par des croix noires, correspondent à des valeurs aberrantes.

³ Source : https://www.researchgate.net/figure/The-main-components-of-a-boxplot-median-quartiles-whiskers-fences-and-outliers_fig6_303779929

1.2 Tests et méthodes statistiques utilisés

Les tests statistiques mis en œuvre dans les études d'évolution de l'Oqali sont :

- Les **tests du chi-2**, utilisés lors de l'étude des paramètres d'étiquetage : ils permettent de montrer si la fréquence de présence d'un paramètre d'étiquetage évolue de manière significative entre deux années ;
- Les **tests de permutation**, utilisés pour le suivi de l'évolution de la composition nutritionnelle depuis 2018, en remplacement des tests de Mann-Whitney, de Wilcoxon et de Student⁴, mis en œuvre dans les précédentes études par catégorie alimentaire. Ces tests statistiques permettent, notamment, de montrer si la composition nutritionnelle, des familles de produits du secteur étudié, évolue de manière significative entre deux années. Ils sont également utilisés, depuis 2020, pour montrer si la composition nutritionnelle des produits porteurs du Nutri-Score est significativement différente des produits non porteurs du logo ;
- La **méthode des doubles différences**, utilisée, depuis 2020, pour évaluer l'effet de l'apposition du Nutri-Score sur l'évolution des parts de marché et des valeurs nutritionnelles pour les produits appariés. Elle permet, notamment, de montrer si l'apposition du Nutri-Score a un effet significatif sur l'évolution des parts de marché et des valeurs nutritionnelles.

Quel que soit le test statistique considéré, il est défini :

- Un **risque d'erreur α** qui représente la probabilité de conclure à une différence significative, à tort. Il est défini à 5% dans les études Oqali, soit $\alpha = 0,05$;
- Une **p-value** afin de déterminer le degré de significativité de la différence observée : plus la p-value est inférieure au risque d'erreur α , plus le degré de significativité est grand. Dans le cadre des études d'évolution de l'Oqali, il sera défini 3 seuils de significativité :
 - Lorsque $0,01 \leq pvalue < 0,05$, la différence est significative (symbolisée sur les graphiques de l'Oqali par une étoile *) ;
 - Lorsque $0,001 \leq pvalue < 0,01$, la différence est très significative (symbolisée sur les graphiques de l'Oqali par 2 étoiles **) ;
 - Lorsque $pvalue < 0,001$, la différence est extrêmement significative (symbolisée sur les graphiques de l'Oqali par 3 étoiles ***).

En plus de ces tests statistiques, un algorithme, développé et éprouvé par Griffith et al. (2017)⁵, peut être appliqué depuis 2018, afin d'évaluer la contribution de l'offre et de la demande à l'évolution de la qualité nutritionnelle. Cette méthode permet de mesurer la contribution relative des changements de l'offre (via des reformulations de produits ou un renouvellement de l'offre) et de la demande (changements des comportements de consommation) à l'évolution de la qualité nutritionnelle moyenne d'une catégorie alimentaire.

⁴ Suite à une expertise statistique, ces tests ont été remplacés par des tests de permutation afin d'augmenter la robustesse des résultats.

⁵ Griffith, R., O'Connell, M., & Smith, K. (2017). The Importance of Product Reformulation Versus Consumer Choice in Improving Diet Quality. *Economica*, 84(333), 34-53. doi:<https://doi.org/10.1111/ecca.12192>

2 TESTS DU CHI-2 D'INDEPENDANCE

2.1 Explication générale du test

Le test du chi-2 ou χ^2 d'indépendance est une méthode de test, utilisée pour vérifier s'il existe une relation entre deux variables qualitatives⁶. Le principe est de calculer l'écart entre la distribution observée et une distribution théorique que l'on obtiendrait si les deux variables étaient totalement indépendantes. Plus l'écart entre ces deux distributions est grand, plus la probabilité de rejeter l'hypothèse d'indépendance des deux variables (l'hypothèse nulle H_0) est forte.

2.2 Exemple appliqué à l'Oqali

Nous souhaitons comparer les résultats de 1263 produits classés d'après leur année de collecte et la présence ou non de portion indiquée, autrement dit vérifier le lien entre 2 variables qualitatives : l'année de collecte et la présence (ou non) de portion indiquée.

Etape 1 : Définition de l'hypothèse nulle H_0 et l'hypothèse alternative H_1 :

H_0 : l'année et la présence de portion indiquée sont indépendantes

H_1 : l'année et la présence de portion indiquée ne sont pas indépendantes

La mise en œuvre du test du chi-2 revient à vérifier que la fréquence de produits possédant une portion indiquée par année ne diffère pas statistiquement selon l'année (**H_0**) :

- si la statistique du test calculée ($T\chi^2$) est inférieure à la valeur χ^2 correspondante au risque d'erreur α choisi (χ^2 théorique) alors la fréquence de produits possédant une portion indiquée ne diffère pas statistiquement selon l'année ;
- si la statistique de test ($T\chi^2$) est supérieure à celle obtenue pour le risque d'erreur α choisi (χ^2 théorique) alors la fréquence de produits possédant une portion indiquée diffère statistiquement selon l'année.

Etape 2 : Calcul des valeurs observées et théoriques

Le Tableau 1 présente la **distribution observée** des produits étudiés, selon leur année de collecte et la présence d'une portion indiquée sur leur emballage.

Tableau 1 : Distribution observée des produits étudiés par année de collecte et présence de portion indiquée

	Année 2016	Année 2020
Présence de portion indiquée	191	230
Absence de portion indiquée	371	565

⁶ Une variable qualitative est une variable qui prend des valeurs exprimées sous forme de catégories distinctes (ex : homme, femme pour la variable genre)

Le Tableau 2 présente la **distribution théorique** des produits, selon leur année de collecte et la présence d'une portion indiquée sur leur emballage, dans le cas où ces deux variables ne dépendent pas l'une de l'autre.

Tableau 2 : Distribution théorique des produits étudiés par année de collecte et présence de portion indiquée, sous l'hypothèse d'indépendance H0 (tableau dit de « contingence »⁷)

	Année 2016	Année 2020
Présence de portion indiquée	174	247
Absence de portion indiquée	388	548

A partir des distributions observée et théorique, on calcule la statistique du test, notée $T\chi^2$. Pour déterminer cette statistique, on calcule :

1. Les différences entre les effectifs observés et théoriques pour chaque combinaison année et présence (ou non) de portion indiquée ;
2. On élève au carré chaque différence et la divise par l'effectif théorique pour la combinaison associée ;
3. La statistique est la somme de ces différences au carré et divisée par les effectifs théoriques.

Etape 3 : Détermination du rejet ou de l'acceptation de l'hypothèse d'indépendance H_0 :

Dans notre exemple, nous avons :

$$T\chi^2 \approx 4,1$$

Afin de conclure sur le test, cette statistique doit être comparée à une valeur du χ^2 théorique, qui est déterminée (à partir de la table du χ^2 en Annexe 1) en fonction de 2 paramètres :

- Le risque d'erreur α (5%) qui représente la probabilité de rejeter l'hypothèse H_0 à tort, c'est-à-dire de conclure que les deux variables sont indépendantes alors qu'en réalité elles ne le sont pas.
- Le degré de liberté (d.d.l) qui est égal au produit du nombre de catégories de chaque variable moins 1. Pour ce test, il est de $1=(2-1)*(2-1)$.

Ainsi, la table du χ^2 nous indique que, pour un risque d'erreur α à 5% et 1 d.d.l, le χ^2 théorique est de 3,84.

On constate que :

$$T\chi^2 > \chi^2 \text{ théorique}$$

⁷ Méthode de représentation de données issues d'un comptage permettant d'estimer la dépendance entre deux caractères. Ces effectifs théoriques sont calculés de manière à avoir une proportion de produits avec une portion indiquée similaire quelle que soit l'année de collecte.

L'hypothèse H_0 est donc rejetée au seuil de 5%. Le test suggère que l'année et la présence de portion indiquée ne sont pas indépendantes, autrement dit que la fréquence de présence des portions indiquées diffère statistiquement selon l'année de collecte.

Etape 4 : Détermination de la p-value

La p-value correspond à la probabilité d'obtenir, sous l'hypothèse H_0 , des différences, entre les échantillons, aussi importantes que celles observées.

Elle permet de mesurer le degré de significativité du test, tel que défini par l'Oqali (cf. partie 1.2) : il n'est donc utile de la calculer que dans le cas où $T\chi^2 > \chi^2$ théorique (rejet de l'hypothèse H_0).

Pour obtenir la p-value, on regarde où se situe $T\chi^2$ dans la table du χ^2 (en Annexe 1) pour un d.d.l égal à 1 :

$$3,84 < T\chi^2 = 4,1 < 5,02$$

$$\text{Soit } 0,05 < \text{p-value} < 0,025$$

La fréquence de présence des portions indiquées diffère donc statistiquement selon l'année de collecte (avec un degré de significativité correspondant à 1 étoile *).

3 TESTS DE PERMUTATION

3.1 Explication générale du test

Les tests de permutation sont des tests non paramétriques⁸ qui permettent de mesurer si la distribution d'un échantillon est significativement différente d'une autre. Ces tests se basent sur le principe suivant : sous l'hypothèse nulle H_0 , les observations des deux échantillons sont échangeables. C'est pourquoi ce test consiste à une comparaison entre la statistique de test **observée** et les statistiques de test générées en **permutant les données** des deux échantillons.

L'Oqali utilise les tests de permutation pour différents types d'échantillons :

- 1) Un échantillon de produits **non appariés** (tous les produits sans prendre en compte les liens d'appariement⁹).
- 2) Un échantillon de produits **non appariés pondérés par les parts de marché** (tous les produits, sans prendre en compte les liens d'appariement⁹ mais en prenant en compte les parts de marché).
- 3) Un échantillon de produits **appariés** (seulement les produits retrouvés les deux années d'étude, sous une forme identique¹⁰ ou modifiée¹¹, en prenant en compte ce lien d'appariement⁹).

A l'Oqali, les tests de permutation ne sont réalisés que si la taille de l'échantillon total est supérieure ou égale à 6 (en-dessous, la robustesse des résultats est considérée comme trop faible) **et le risque d'erreur α est fixé à 5%.**

3.2 Exemples appliqués à l'Oqali

3.2.1 Echantillons non appariés

Le test de permutation des échantillons **non appariés** correspond à une comparaison de deux échantillons indépendants.

Dans notre exemple, nous allons comparer les teneurs moyennes en sucres des produits de l'année 2010 et des produits de l'année 2016 pour une famille d'un secteur donné. Nous avons un groupe A composé de 3 produits de l'année 2010 et un groupe B composé de 4 produits de l'année 2016. Les produits des deux groupes sont considérés comme indépendants.

⁸ Les tests non paramétriques ne nécessitent aucune hypothèse sur la distribution des données, contrairement aux tests paramétriques qui supposent que les données suivent un type de loi de distribution connu (ex : loi normale)

⁹ Un lien d'appariement correspond au lien entre un produit vu l'année A et sa version, identique ou modifiée, retrouvée l'année B.

¹⁰ Les produits identiques sont présents sur le marché à la fois l'année A et l'année B, et sont semblables en tous points.

¹¹ Les produits modifiés sont présents sur le marché l'année A et également l'année B mais dans une version modifiée, impliquant au moins un changement de l'emballage (portions, repères nutritionnels, allégations, autres informations) et/ou de la composition (valeurs nutritionnelles ou listes d'ingrédients).

Tableau 3 : Observations de la teneur en sucres des produits du groupe A (2010) et du groupe B (2016)

Groupe A Année 2010		Groupe B Année 2016	
Produit	Teneur en sucres (g/100g)	Produit	Teneur en sucres (g/100g)
1	2,6	4	2,9
2	1,1	5	1,8
3	2,6	6	2,9
		7	2,4

La somme des effectifs des 2 groupes A (n=3) et B (n=4) est bien supérieure ou égale à 6 donc le test de permutation peut être appliqué à cet échantillon.

Etape 1 : Formulation des hypothèses

Les deux hypothèses émises pour ce test sont les suivantes :

H₀ : les teneurs moyennes en sucres des deux groupes sont les mêmes. Si une différence est observée, celle-ci est due aux fluctuations d'échantillonnage.

H₁ : Il existe une différence entre les teneurs moyennes en sucres des deux groupes.

Ici, le rejet de l'hypothèse H₀ impliquera que les teneurs moyennes en sucres des deux groupes sont significativement différentes.

Etape 2 : Statistique de test observée

Pour les échantillons non appariés, la statistique de test est la valeur absolue de la différence des deux moyennes :

$$T_0 = |m_A - m_B|$$

$$T_0 = |2,1 - 2,5|$$

$$T_0 = 0,4$$

avec m_A et m_B, respectivement la moyenne du groupe A et la moyenne du groupe B.

Etape 3 : Permutations

Supposons que l'hypothèse H₀ est vraie : les teneurs moyennes en sucres des deux groupes sont donc les mêmes ; autrement dit, la teneur moyenne en sucres est indépendante du groupe étudié. Dans ce cas, il est tout aussi probable qu'un produit se trouve dans le groupe A ou dans le groupe B.

Nous allons considérer tous les agencements possibles, en permutant aléatoirement les observations entre les groupes A et B. Le nombre total de permutations possibles correspond à la factorielle du nombre d'observations, soit 5 040 dans notre exemple de 7 observations.

Pour chacune des permutations, la statistique de test T_n est calculée, en suivant la formule utilisée pour la statistique de test observée T_0 ¹².

Etape 4 : Calcul de la p-value et rejet de H_0

La p-value est estimée en calculant la proportion des statistiques de test supérieures ou égales à la statistique de test T_0 observée :

$$p\text{-value} = \frac{\text{nombre de fois où } T_n \text{ est supérieure ou égale à } T_0}{\text{nombre de permutations}}$$

Dans cet exemple, nous avons 2520 statistiques de test supérieures à T_0 donc la p-value vaut 0,5.

$$p\text{-value} = \frac{2520}{5040} = 0,5$$

Cette p-value de 0,5 étant supérieure au risque α de 0,05, l'hypothèse H_0 est acceptée : il n'y a donc pas de différence significative entre les moyennes des deux groupes.

Remarque : dans le cas de comparaisons multiples (plusieurs tests appliqués à un même échantillon de produits par exemple), une étape supplémentaire d'ajustement de la p-value est nécessaire. En effet, lorsque des comparaisons multiples sont réalisées, le risque de se tromper (en déclarant que deux moyennes sont différentes alors qu'en réalité elles ne le sont pas) augmente avec le nombre de tests réalisés.

Pour l'Oqali, cet ajustement se fait selon la méthode de Benjamini and Hochberg :

$$p\text{-value}_{\text{ajustée}} = \frac{p\text{-value} * \text{nombre de tests}}{\text{rang de la p-value}}$$

3.2.2 Echantillons non appariés avec pondération

De même que dans la partie 3.2.1, le test de permutation des échantillons **non appariés avec pondération** correspond à une comparaison de deux échantillons indépendants. **La seule différence entre le test de permutation présenté ci-dessous et celui de la partie 3.2.1 réside dans la prise en compte de pondération ou poids associés à chaque observation.**

Dans notre exemple, nous allons comparer des teneurs moyennes en sucres de produits avec Nutri-Score et de produits sans Nutri-Score pour une famille donnée, en tenant compte des parts de marché des produits (en volumes de ventes)¹³. Nous avons un groupe A composé de 3 produits avec Nutri-Score et un groupe B composé de 3 produits sans Nutri-Score : les produits des deux groupes sont considérés comme indépendants.

¹² Pour l'Oqali, lorsqu'un des deux groupes présentent plus de 10 observations, un test approximatif est réalisé avec un nombre de permutations fixé à 1 000 000. Sinon, un test exact est réalisé (ensemble des permutations possibles réalisées).

¹³ Les parts de marché, en volumes de ventes, sont fictives pour cet exemple.

Tableau 4 : Teneurs en sucres et parts de marché des produits du groupe A (avec Nutri-Score) et du groupe B (sans Nutri-Score)

Groupe A Avec Nutri-Score			Groupe B Sans Nutri-Score		
Produit	Teneur en sucres (g/100g)	Parts de marché ¹	Produit	Teneur en sucres (g/100g)	Parts de marché ¹
1	2,6	0,003	4	2,9	0,005
2	1,1	0,004	5	1,8	0,002
3	1,7	0,001	6	2,9	0,003

¹ ratio des volumes des produits identifiés par l'Oqali versus le volume total des produits identifiés par l'Oqali selon Kantar-Panel Worldpanel (données d'achats de ménages représentatives de la population française)

La somme des effectifs des 2 groupes A (n=3) et B (n=3) est bien supérieure ou égale à 6 donc le test de permutation peut être appliqué à cet échantillon.

Afin de tenir compte de la part de marché des produits considérés, les échantillons sont formés en répliquant chaque teneur en sucres selon un nombre de réplication. Pour chaque teneur en sucres, le nombre de réplication correspond à la part de marché dudit produit, multipliée par un facteur de réplication B. Pour l'Oqali, le facteur de réplication B est fixé à 1000 et on obtient ainsi :

$$\text{Nombre de réplication} = (B * \text{parts de marché})$$

Ce calcul est appliqué à chaque produit afin d'obtenir les résultats suivants dans notre exemple.

Tableau 5 : Teneurs en sucres, parts de marché et nombres de réplication associés, pour les produits du groupe A (avec Nutri-Score) et du groupe B (sans Nutri-Score)

Groupe A Avec Nutri-Score				Groupe B Sans Nutri-Score			
Produit	Teneur en sucres (g/100g)	Parts de marché ¹	Nombre de réplications	Produit	Teneur en sucres (g/100g)	Parts de marché ¹	Nombre de réplications
1	2,6	0,003	3	4	2,9	0,005	5
2	1,1	0,004	4	5	1,8	0,002	2
3	1,7	0,001	1	6	2,9	0,003	3

¹ ratio des volumes des produits identifiés par l'Oqali versus le volume total des produits identifiés par l'Oqali selon Kantar-Panel Worldpanel (données d'achats de ménages représentatives de la population française)

Pour chaque produit, les teneurs en sucres sont répliquées en fonction du nombre de réplication pour former l'échantillon final. Par exemple pour le groupe A, la teneur en sucres du produit 1 sera répliquée 3 fois. Cette méthode de réplication est appliquée à chaque produit des deux groupes étudiés pour obtenir les échantillons finaux du Tableau 6.

Tableau 6 : Teneurs en sucres des produits du groupe A (avec Nutri-Score) et du groupe B (sans Nutri-Score), après répllication

Groupe A après répllication Avec Nutri-Score		Groupe B après répllication Sans Nutri-Score	
Produit	Teneur en sucres (g/100g)	Produit	Teneur en sucres (g/100g)
1	2,6	4	2,9
1	2,6	4	2,9
1	2,6	4	2,9
2	1,1	4	2,9
2	1,1	4	2,9
2	1,1	5	1,8
2	1,1	5	1,8
3	1,7	6	2,9
		6	2,9
		6	2,9

A partir de ces échantillons, les différentes étapes de formulation des hypothèses, statistique de test observée, permutations et calcul de la p-value sont similaires à celles de la partie 3.2.1.

3.2.3 Echantillons appariés

Le test de permutation des échantillons **appariés** correspond à une comparaison de deux échantillons dépendants l'un de l'autre.

Dans le cas de l'Oqali, les échantillons appariés correspondent aux produits collectés par l'Oqali l'année A et l'année B, sous une forme identique (produits identiques) ou modifiée (produits modifiés). Pour les tests de permutation, l'un des échantillons étudiés comporte les produits appariés retrouvés l'année A et l'autre rassemble les produits correspondants, retrouvés sous une forme identique ou modifiée, l'année B (groupe B). Contrairement aux tests de permutation présentés dans les parties 3.2.1 et 3.2.2, les deux échantillons sont donc dépendants.

Ci-dessous, un exemple d'échantillons appariés, correspondant à la teneur moyenne en protéines des produits de l'année 2016 et leurs versions, modifiées ou identiques, de l'année 2020 pour une famille d'un secteur donné. Nous avons 4 couples de produits appariés, avec un groupe A composé de leurs versions de l'année 2016 et un groupe B composé de leurs versions de l'année 2020 : les deux groupes sont donc bien considérés comme dépendants.

Tableau 7 : Teneurs en protéines des couples de produits appariés en 2016 (groupe A) et 2020 (groupe B)

	Groupe A Année 2016	Groupe B Année 2020
Couple de produits appariés	Teneur en protéines (g/100g)	Teneur en protéines (g/100g)
1	22,9	24,0
2	28,8	24,0
3	28,0	28,0
4	25,3	26,5

La somme des effectifs des 2 groupes A (n=4) et B (n=4) est bien supérieure à 6 donc le test de permutation peut être appliqué à cet échantillon.

Le principe et les étapes sont les mêmes que pour la partie 3.2.1, à l'exception de la mise en œuvre des permutations. Contrairement aux échantillons non appariés pour lesquels toutes les observations sont permutées dans un ordre aléatoire, pour les échantillons appariés les permutations se réalisent **au sein d'un couple de produits**. Cela revient à réarranger les observations de l'année A et de l'année B, **en gardant les couples de produits ensemble**.¹⁴

Par exemple, une première permutation pourrait être une inversion des teneurs au sein du couple 1.

Tableau 8 : Exemple de permutation des teneurs en protéines des couples de produits appariés en 2016 (groupe A) et 2020 (groupe B)

	Groupe A Année 2016	Groupe B Année 2020	
Couple de produits appariés	Teneur en protéines (g/100g)	Teneur en protéines (g/100g)	Delta (Teneur B – Teneur A)
1	24,0	22,9	-1,1
2	28,8	24,0	-4,8
3	28,0	28,0	0
4	25,3	26,5	+1,2

Les autres étapes de formulation des hypothèses, statistique de test observée et calcul de la p-value sont similaires à celles de la partie 3.2.1.

¹⁴ Pour l'Oqali, le nombre de permutations pour les tests des produits appariés est fixé à 1 000 000.

4 METHODE DES DOUBLES DIFFERENCES

4.1 Explication générale de la méthode

La méthode des doubles différences est une méthode statistique utilisée pour estimer l'effet d'un traitement sur une variable donnée (variable de résultat) : elle consiste à comparer la différence entre un groupe de contrôle (sans traitement) et un groupe traité avant et après l'introduction du traitement.

Cette méthode permet d'obtenir un estimateur de l'effet du traitement, appelé **estimateur DID** (estimateur des doubles différences), ainsi qu'une **p-value** indiquant la significativité de cet effet, au sens statistique.

A noter que cette méthode n'est valable que sous les 4 hypothèses suivantes :

- **L'hypothèse des tendances parallèles communes** doit être respectée, autrement dit si le traitement n'avait pas été appliqué au groupe exposé au traitement, la variable de résultat aurait eu, en moyenne, la même évolution que celle du groupe témoin. Des covariables¹⁵ pourront néanmoins être ajoutées au modèle afin de tenir compte de paramètres identifiables, pouvant influencer la variable de résultat ;
- **Sur la période étudiée, le traitement correspond à la seule modification observée** pour le groupe traité.
- Les individus du groupe traité ont tous reçu le **traitement au même moment** ;
- Il n'y a **pas d'effet d'anticipation du traitement** par le groupe traité.

La Figure 2 ci-dessous représente graphiquement la méthode des doubles différences.

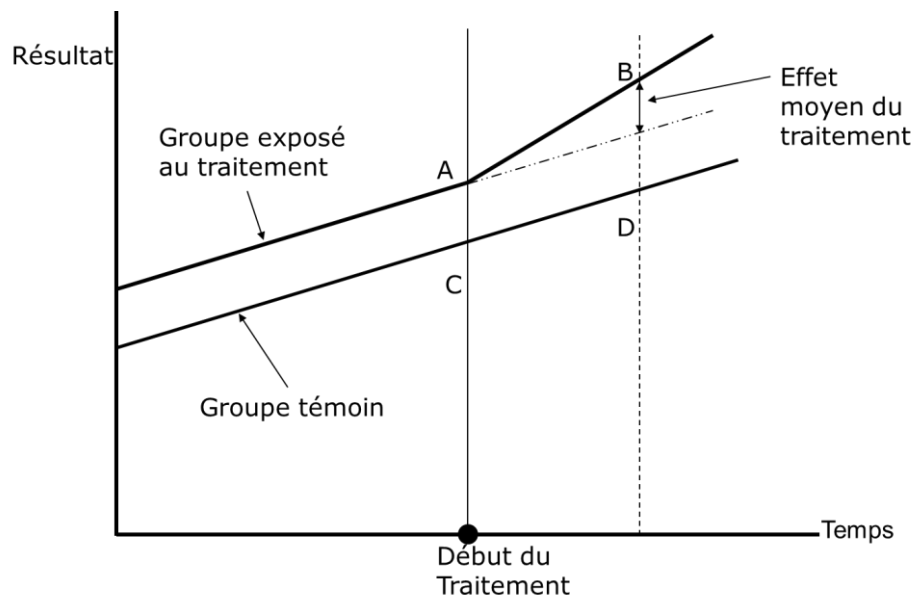


Figure 2 : Schéma expliquant la méthode des doubles différences¹⁶

¹⁵ Une covariable désigne une variable (continue ou discrète) qui est incluse dans une analyse statistique pour réduire l'effet de confusion ou pour tenir compte des variations dues à des facteurs externes. En d'autres termes, c'est une variable qui peut influencer à la fois le résultat (variable dépendante) et les autres variables présentes dans la régression statistique (variable indépendante). Elle est donc prise en compte pour s'assurer que l'effet observé est bien attribuable à la variable indépendante étudiée.

¹⁶ Source : http://cega.berkeley.edu/assets/miscellaneous_files/Difference_in_Differences_Rwanda.pdf

4.2 Exemple appliqué à l'Oqali

A partir de 2023, les études par catégorie alimentaire, réalisées par l'Oqali, intègrent, lorsque cela est pertinent, des indicateurs spécifiques au Nutri-Score, dont une partie vise à estimer l'effet de son apposition (sur l'emballage des produits) sur les parts de marché et les valeurs nutritionnelles, comparativement aux produits sans Nutri-Score.

La **méthode des doubles différences** est appliquée (uniquement aux produits présents les deux années d'étude¹⁷) et permet de comparer l'évolution des parts de marché ou des valeurs nutritionnelles, observée entre deux années d'étude, pour les produits appariés sans Nutri-Score aux deux temps étudiés (groupe témoin) et les produits appariés sans Nutri-Score la première année étudiée et avec le logo la seconde année (le groupe traité).

Comme énoncé dans la partie 4.1, cette méthode est mise en œuvre sous les quatre hypothèses suivantes :

- 1) L'hypothèse des **tendances parallèles communes**, autrement dit si les produits du groupe avec Nutri-Score, le groupe traité, n'avaient pas apposé le logo, l'évolution moyenne du paramètre d'intérêt (i.e. part de marché et valeurs nutritionnelles) des produits de ce groupe aurait suivi une tendance « parallèle » à celle du groupe témoin. Cette hypothèse a été partiellement atténuée en incluant des covariables dans les estimations (cf. ci-dessous). La considération de covariables permet de considérer que l'estimateur DID sera sans biais même si ces deux groupes peuvent avoir connu des évolutions différentes dans leurs parts de marché moyennes ou leurs valeurs nutritionnelles moyennes, à condition que ces évolutions différentielles puissent être expliquées par des changements dans ces covariables;
 - ➔ Pour l'étude de l'effet du Nutri-Score sur les parts de marché, les covariables retenues sont : le prix moyen, les teneurs des produits avant apposition du Nutri-Score pour les nutriments d'intérêt du secteur (parmi les nutriments suivants : acides gras saturés, protéines, sucres, matières grasses et sel), la famille de produits (lorsque la méthode est appliquée sur le secteur entier) et le segment de marché.
 - ➔ Pour l'étude de l'effet du Nutri-Score sur les valeurs nutritionnelles, les covariables retenues sont : le prix moyen avant apposition du Nutri-Score ainsi que la famille de produits (lorsque la méthode est appliquée sur le secteur entier) et le segment de marché. Les autres valeurs nutritionnelles avant apposition du Nutri-Score n'ont pas été intégrées car les modifications de recette sont complexes à appréhender (par exemple, une diminution de la teneur en sel peut être compensée par une hausse de la teneur en sucres, pour des raisons organoleptiques ou technologiques).
- 2) Pour le groupe traité, **la seule modification, observée sur la période étudiée, correspond à l'apposition du Nutri-Score, outre celles sur les covariables ;**
- 3) Les produits du groupe traité ont tous apposé le Nutri-Score au même moment : le potentiel décalage dans le temps des effets, suite à la réelle apposition du Nutri-Score sur les emballages, n'est pas pris en compte ;

¹⁷ Afin d'être pris en compte pour l'étude d'une variable (part de marché ou valeurs nutritionnelles), les couples de produits appariés doivent présenter une valeur pour toutes les variables et covariables considérées dans le modèle, pour les deux années d'étude.

- 4) Il est considéré qu'il n'y a pas d'effet d'anticipation de l'apposition du Nutri-Score : autrement dit si, en prévision d'une future apposition du Nutri-Score, certains produits du groupe sans Nutri-Score ont été reformulés, ces éventuelles modifications de positionnement (valeurs nutritionnelles par exemple) ne sont pas prises en compte.

Pour exécuter l'analyse, le nombre de couples de produits au sein de chacun des 2 groupes (avec et sans Nutri-Score) doit être strictement supérieur à la somme de toutes les covariables + 5 (qui correspond à une contrainte technique spécifique à la fonction utilisée dans R)¹⁸.

Ainsi, à l'échelle du secteur :

- pour les parts de marché : prix moyen + nombre de nutriments d'intérêt du secteur (parmi la liste suivante : acides gras saturés, protéines, sucres, matières grasses et sel) + 2 variables de segmentation¹⁹ + 5 (le nombre minimal de couples de produits appariés est donc compris entre 10 et 14 selon le nombre de nutriments d'intérêt) ;
- pour les valeurs nutritionnelles : prix moyen + 2 variables de segmentation¹⁹ + 5 (le nombre minimal de couples de produits appariés est donc de 9).

Ainsi, à l'échelle des familles :

- pour les parts de marché : prix moyen + nombre de nutriments d'intérêt de la famille²⁰ (parmi la liste suivante : acides gras saturés, protéines, sucres, matières grasses et sel) + 1 variable de segmentation²¹ + 5 (le nombre minimal de couples de produits appariés est donc compris entre 9 et 13 selon le nombre de nutriments d'intérêt) ;
- pour les valeurs nutritionnelles : prix moyen + 1 variable de segmentation²⁰ + 5 (le nombre minimal de couples de produits appariés est donc de 8).

La robustesse des résultats est considérée comme trop faible en-dessous de cet effectif.

Dans notre exemple, nous allons étudier l'effet de l'apposition du Nutri-Score sur les teneurs moyennes en sucres entre 2016 (T1) et 2020 (T2) pour une famille donnée. Les deux groupes de produits considérés seront donc le groupe sans Nutri-Score (absence du logo au T1 et T2) et le groupe avec Nutri-Score (absence du logo au T1 mais présence au T2).

¹⁸ Le nombre 5 correspond à une contrainte technique spécifique à la fonction utilisée dans R. Si le nombre de couples par groupe n'est pas strictement supérieur à la somme des covariables plus 5, la fonction ne peut pas s'exécuter.

¹⁹ Les variables de segmentation choisies sont : famille et segment de marché.

²⁰ Le nombre de nutriments d'intérêt de la famille peut différer de celui retenu pour l'ensemble du secteur. Par exemple, dans le secteur des Glaces et sorbets, les familles « Vrac sorbet » et « Sorbet cône » ne comptent qu'un seul nutriment d'intérêt (Sucres), tandis que le secteur dans son ensemble en comprend trois (sucres, AGS, MG). Certaines familles peuvent toutefois avoir les mêmes nutriments d'intérêt que ceux définis au niveau du secteur, comme c'est le cas pour les familles « Glace mini bâtonnet » et « Glace mini cône » (sucres, AGS, MG).

²¹ La variable de segmentation choisie est : segment de marché.

Tableau 9 : Teneurs moyennes en sucres des produits appariés avec et sans Nutri-Score en 2016 et 2020

Sucres	Moyenne des produits appariés pour le groupe sans Nutri-Score (g/100g)		Moyenne des produits appariés pour le groupe avec Nutri-Score (g/100g)	
	T1 (2016)	T2 (2020)	T1 (2016)	T2 (2020)
Famille 1	2,1	2,2	2,9	1,8

Etape 1 : Définition de l'hypothèse nulle H_0 et l'hypothèse alternative H_1 :

H_0 : L'apposition du Nutri-Score n'a pas eu d'effet sur les teneurs moyennes en sucres.

H_1 : L'apposition du Nutri-Score a impacté les teneurs moyennes en sucres.

Etape 2 : Calcul des différences et de l'estimateur DID

Les différences sont calculées selon les modalités de la Figure 3 :

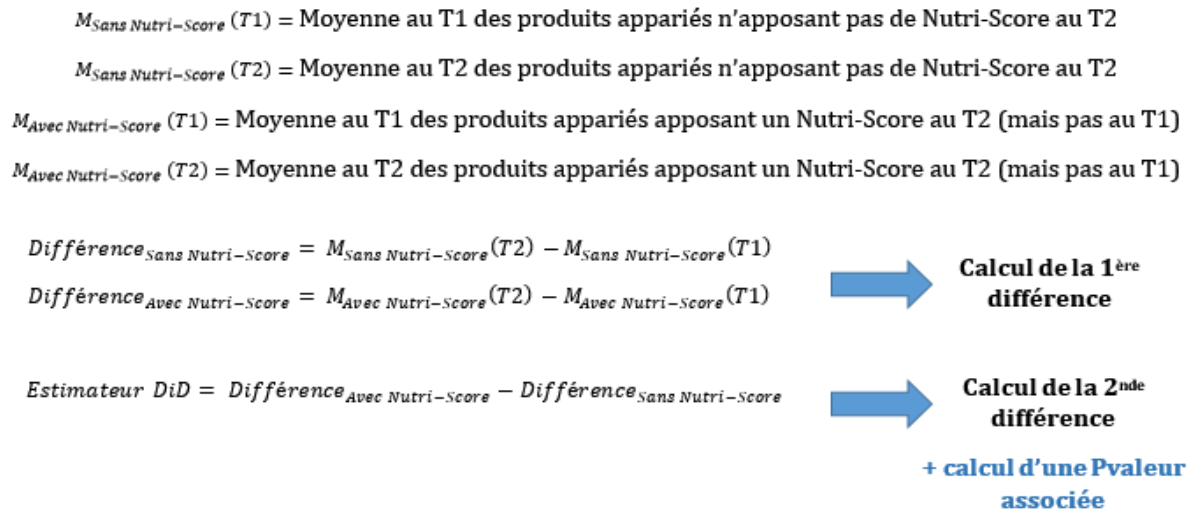


Figure 3 : Détail de calcul de la méthode des doubles différences

L'estimateur DID ainsi obtenu est de -1,2g/100g.

Etape 3 : Détermination de la p-value

A partir de cet estimateur DID, la p-value est calculée, selon la méthode du multiplier bootstrap ou ré-échantillonnage multiplicatif²².

Dans notre exemple, la p-value obtenue est environ égale à $4,34E^{-11}$. Celle-ci est inférieure au risque d'erreur α de 0,05 donc l'hypothèse H_0 est rejetée : l'apposition du Nutri-Score a significativement impacté les teneurs moyennes en sucres des produits.

Ici, la p-value est inférieure à 0,001 donc l'hypothèse H_0 est rejetée avec un degré de significativité très élevé (***) (cf. partie 1.2).

²² Le multiplier bootstrap est une méthode de calcul des erreurs standards des estimateurs par bootstrap qui prend en compte la non-linéarité de l'estimateur DID.

5 CONTRIBUTION DE L'OFFRE ET DE LA DEMANDE A L'EVOLUTION DE LA QUALITE NUTRITIONNELLE

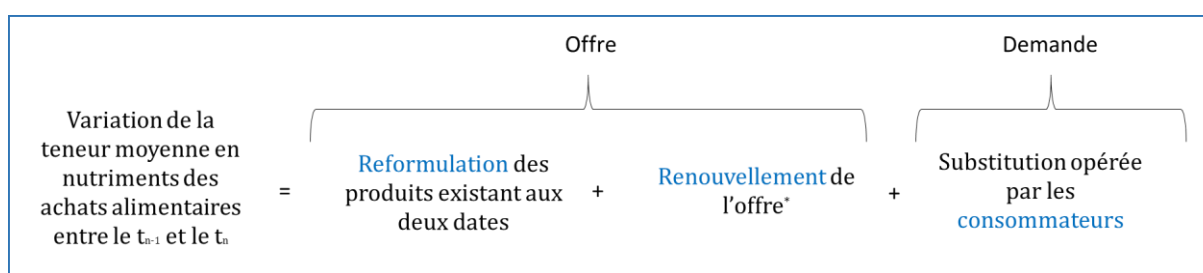
5.1 Explication générale de la méthode

Une méthode utilisée par Griffith et al. (2017)²³ a été adaptée pour pouvoir être appliquée aux données Oqali afin de mesurer la contribution relative des changements de l'offre (évolution de la qualité nutritionnelle des aliments via des reformulations ou un renouvellement de l'offre) et de la demande (changements des comportements de consommation) à l'évolution de la qualité nutritionnelle moyenne d'un secteur.

Cette méthode est appliquée à l'échelle d'un secteur (catégorie) et non d'une famille (sous-catégorie), contrairement aux autres analyses sur la composition nutritionnelle, réalisées par l'Oqali (cf. parties 3 et 4). De plus, elle nécessite de réaliser certaines transformations des données (cf. étapes 1 et 2 décrites ci-dessous), ce qui implique que ses résultats ne permettent pas de conclure sur l'évolution des teneurs moyennes en nutriments pondérées par les parts de marché à l'échelle du secteur. Ces traitements, complémentaires des autres indicateurs de suivi des valeurs nutritionnelles, permettent uniquement de décomposer les évolutions constatées afin de comprendre si elles viennent de reformulations, renouvellement de l'offre ou de substitutions opérées par les consommateurs.

Pour chaque nutriment d'intérêt d'un secteur, l'évolution de la teneur moyenne pondérée par les parts de marché est interprétée comme étant la résultante de trois effets (Figure 4) :

- la reformulation des produits par les industriels ;
- le renouvellement de l'offre, c'est-à-dire l'apparition/le retrait de références sur le marché ;
- les substitutions opérées par les consommateurs au sein des produits retrouvés les deux années étudiées.



* Le renouvellement de l'offre correspond au retrait/lancement des produits

Figure 4 : Décomposition de l'évolution des teneurs moyennes pondérées en nutriments d'intérêt entre t_{n-1} et t_n

La méthode mise en œuvre permet de quantifier ces trois effets et d'analyser leur contribution à l'évolution la qualité nutritionnelle moyenne des aliments consommés.

Pour quantifier chacun de ces effets, les produits Oqali sont affectés à 3 sous-groupes :

²³ Griffith, R., O'Connell, M., & Smith, K. (2017). The Importance of Product Reformulation Versus Consumer Choice in Improving Diet Quality. *Economica*, 84(333), 34-53. doi:<https://doi.org/10.1111/ecca.12192>

- **les produits retirés ou non captés au t_n (groupe X)** : ces produits ont été collectés au t_{n-1} mais n'ont pu être reliés à des produits collectés au t_n . Il s'agit de produits qui ont été retirés du marché, ou bien de produits toujours existants mais qui n'ont pas pu être captés par l'Oqali au t_n ;
- **les produits appariés (groupe C)** : ces produits sont présents dans la base de données de l'Oqali à la fois au t_{n-1} et au t_n . Ils peuvent présenter ou non des évolutions d'emballage ou de composition nutritionnelle entre le t_{n-1} et le t_n ;
- **les produits nouveaux ou non captés au t_{n-1} (groupe N)** : ces produits ont été collectés au t_1 mais n'ont pu être reliés à des produits collectés au t_{n-1} . Il s'agit d'innovations, d'extensions de gamme, ou bien de produits déjà existants mais qui n'avaient pas été captés par l'Oqali au t_{n-1} .

Les différentes étapes pour appliquer cette méthode sont décrites ci-dessous.

Etape 1 : Inférence des données de composition nutritionnelle manquantes

Afin d'appliquer cette méthode, une première condition doit être remplie : la composition nutritionnelle des produits pris en compte doit être connue pour tous les nutriments étudiés. Pour les produits Oqali, il arrive que certaines données nutritionnelles ne soient pas renseignées pour l'ensemble des nutriments d'intérêt²⁴. **Ces données manquantes sont alors inférées comme suit :**

- Pour les références appartenant au groupe C (références collectées au t_{n-1} et au t_n) :
 - si les informations disponibles pour son père ou son fils permettent d'inférer sa composition nutritionnelle, alors celle-ci lui est attribuée ;
 - si les informations disponibles pour son père ou son fils ne permettent pas d'inférer sa composition nutritionnelle, alors la composition nutritionnelle moyenne pondérée à t_{n-1} de sa famille d'appartenance lui est attribuée.

Dans les deux cas, les inférences de données sont réalisées sous l'hypothèse d'une non reformulation des références.

- Pour les références appartenant aux groupes N et X (produits nouveaux ou non captés au t_{n-1} et produits retirés ou non captés au t_n), la composition nutritionnelle moyenne pondérée à t_{n-1} de leur famille d'appartenance leur est attribuée.

Etape 2 : Transformation des données

Afin d'appliquer cette méthode, une seconde condition doit être remplie : chacune des sommes des parts de marché à t_{n-1} et à t_n doit être égale à 100%. A l'Oqali, les échantillons étudiés ne couvrent pas 100% des volumes de ventes des secteurs donc la somme des parts de marché est inférieure à 100%. Les données doivent être transformées pour satisfaire cette condition. Deux possibilités de modification des données ont été identifiées, issues de deux hypothèses différentes : elles constituent deux méthodes de traitement.

La **méthode 1** consiste à transformer la part de marché de chaque référence de manière à ce que la somme des parts de marché soit égale à 100% (Figure 5). Les produits non couverts par l'Oqali

²⁴ Cela est dû au fait que, en fonction des dates de collecte, la déclaration nutritionnelle affichée sur les emballages (principale source de données de l'Oqali) pouvait être absente ou présenter un niveau de détail variable.

sont supposés avoir la même évolution que l'échantillon étudié sur la période (cette méthode « distord » d'autant moins les données que les couvertures du marché des échantillons étudiés sont élevées et du même ordre de grandeur à chacun des temps de suivi et que les échantillons sont représentatifs du marché à t_{n-1} comme à t_n).

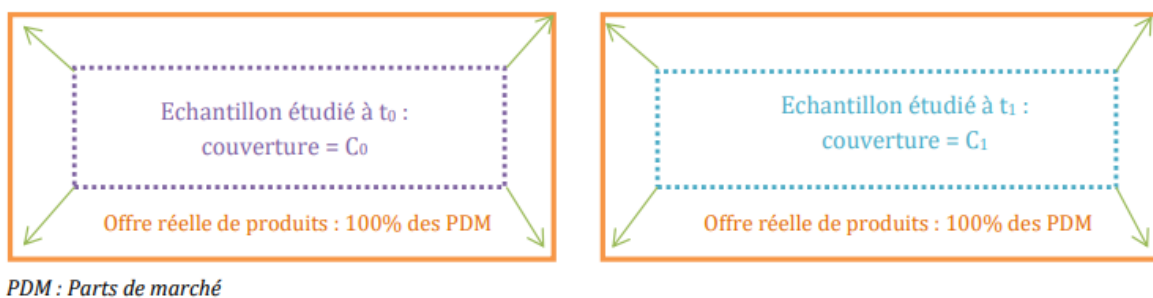


Figure 5 : Illustration de la méthode 1 de transformation des données

Les parts de marché de chaque référence sont transformées comme suit :

$$w_{it}^* = w_{it} / C_t$$

avec w_{it}^* : la part de marché corrigée de la référence i à l'instant t

w_{it} : la part de marché de la référence i à l'instant t

C_t : la couverture de l'échantillon Oqali à l'instant t

La **méthode 2** consiste à « ajouter », à l'échantillon étudié, une référence ayant comme part de marché celle du marché non couvert par l'Oqali. Cette référence est affectée au groupe C et sa composition nutritionnelle est supposée ne pas avoir évolué sur la période (Figure 6).

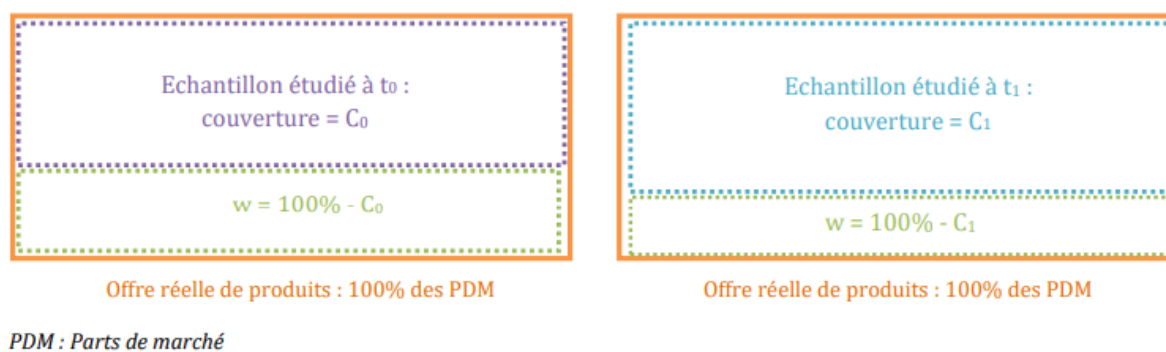


Figure 6 : Illustration de la méthode 2 de transformation des données

Les résultats de la méthode 2 étant globalement semblables à ceux obtenus via la méthode 1, ceux-ci ne sont la plupart du temps pas intégrés aux études par catégorie d'aliments.

Etape 3 : Application de l'algorithme

L'algorithme utilisé par Griffith et al. (2017)²⁵ a été adapté aux données Oqali. Le détail de cet algorithme et de son application se trouve en Annexe 2.

²⁵ Griffith, R., O'Connell, M., & Smith, K. (2017). The Importance of Product Reformulation Versus Consumer Choice in Improving Diet Quality. *Economica*, 84(333), 34-53. doi:<https://doi.org/10.1111/ecca.12192>

Attention, cette méthode ne permet pas d'obtenir l'évolution des teneurs moyennes en nutriments, pondérées par les parts de marché, à l'échelle du secteur. En effet, les parts de marché utilisées ont été remises à 100% et les valeurs nutritionnelles manquantes ont été inférées : elles ne reflètent donc pas exactement la réalité du marché. Comme cité ci-dessus, **cette méthode permet donc d'observer la contribution de l'offre et de la demande sur la qualité nutritionnelle globale d'un secteur mais ne donne pas d'élément sur des évolutions statistiquement significatives au cours du temps.**

5.2 Exemple appliqué à l'Oqali

Dans notre exemple, nous allons observer la contribution de l'offre et de la demande à l'évolution de la teneur moyenne en matières grasses, pondérée par les parts de marché, entre 2010 et 2016, pour un secteur donné. L'année 2010 correspond au t_0 (évoqué dans la partie 5.1 sous le nom t_{n-1}) et l'année 2016 correspond au t_1 (évoqué dans la partie 5.1 sous le nom t_n). L'échantillon étudié est présenté dans le Tableau 10 ; les produits qui le composent ont été classés selon les sous-groupes utilisés dans la méthode (cf. partie 5.1).

Tableau 10 : Teneurs en matières grasses et parts de marché fictives des produits d'un secteur donné, pour 2010 (T0) et 2016 (T1)

Produit	Année 2010 (T ₀)		Année 2016 (T ₁)		Sous-groupe
	Teneur en matières grasses (g/100g)	Parts de marché ¹	Teneur en matières grasses (g/100g)	Parts de marché ¹	
1	3,2	0,17	/	/	X
2	3,5	0,26	/	/	X
3	<i>Donnée manquante</i>	0,13	5,9	0,23	C
4	9,0	0,19	6,2	0,10	C
5	/	/	4,3	0,12	N
6	/	/	3,7	0,35	N

¹ ratio des volumes des produits identifiés par l'Oqali versus le volume total des produits identifiés par l'Oqali selon Kantar-Panel Worldpanel (données d'achats de ménages représentatives de la population française)

« / » : produit non existant ou non capté par l'Oqali

« Donnée manquante » : donnée non renseignée (la déclaration nutritionnelle affichée sur les emballages pouvait être absente ou présenter un niveau de détail variable)

Sous-groupes : « X » produits retirés ; « C » produits appariés ; « N » produits nouveaux

Etape 1 : Inférence des données de composition nutritionnelle manquantes

Dans notre exemple, le produit 3 appartient au groupe C mais il ne dispose pas d'une teneur en matières grasses renseignée pour l'année 2010. Comme expliqué dans la partie 5.1, la teneur en matières grasses de son produit « fils » de 2016 lui sera donc attribuée, soit 5,9 g/100g.

Etape 2 : Transformation des données

Afin d'appliquer cette méthode, les parts de marché des produits étudiés sont transformées de manière à ce que la somme des parts de marché soit égale à 100% pour chaque année.

Le Tableau 11 présente respectivement les données étudiées, après transformation des parts de marché selon la méthode 1 (extension proportionnelle des parts de marché). Les données étudiées, après transformation des parts de marché selon la méthode 2 sont présentées en Annexe 3.

Tableau 11 : Teneurs en matières grasses et parts de marché fictives des produits d'un secteur donné, pour 2010 (T0) et 2016 (T1), après inférence des valeurs nutritionnelles et transformation des parts de marché avec la méthode 1

Année 2010 (T0)				Année 2016 (T1)			Sous-groupe
Produit	Teneur en matières grasses (g/100g)	Parts de marché ¹	Parts de marché ¹ corrigées avec la méthode 1	Teneur en matières grasses (g/100g)	Parts de marché ¹	Parts de marché ¹ corrigées avec la méthode 1	
1	3,2	0,17	0,23	/	/	/	X
2	3,5	0,26	0,35	/	/	/	X
3	5,9	0,13	0,17	5,9	0,23	0,29	C
4	9,0	0,19	0,25	6,2	0,10	0,13	C
5	/	/	/	4,3	0,12	0,15	N
6	/	/	/	3,7	0,35	0,44	N

¹ ratio des volumes des produits identifiés par l'Oqali versus le volume total des produits identifiés par l'Oqali selon Kantar-Panel Worldpanel (données d'achats de ménages représentatives de la population française)

« / » : produit non existant ou non capté par l'Oqali

« Donnée manquante » : donnée non renseignée (la déclaration nutritionnelle affichée sur les emballages pouvait être absente ou présenter un niveau de détail variable)

Sous-groupes : « X » produits retirés ; « C » produits appariés ; « N » produits nouveaux

Etape 3 : Application de l'algorithme

Comme expliqué dans la partie 5.1, l'algorithme est appliqué aux données de l'échantillon inféré et transformé. Les résultats finaux sont exprimés en pourcentage (rapportés à la moyenne en matières grasses pondérée par les parts de marché à t_0) :

Tableau 12 : Contributions des reformulations, renouvellement de l'offre et substitutions réalisées par les consommateurs (en pourcentage) à l'évolution de la teneur moyenne pondérée en matières grasses du secteur, entre 2016 et 2020 (méthode 1) – données fictives

Nutriment	Moyenne pondérée 2010 (g/100g)	Moyenne pondérée 2016 (g/100g)	Evolution des moyennes pondérées	Reformulations	Renouvellement de l'offre	Substitutions
Matières grasses	5,2	4,8	-8%	-13%	+6%	-0,8%

Comme explicité précédemment (cf. partie 5.1), de par les diverses transformations des données réalisées pour l'application de cette méthode, les résultats obtenus ne permettent pas de conclure qu'à l'échelle du secteur étudié, la teneur moyenne en matières grasses pondérée par les parts de marché diminue entre 2010 et 2016 (-8%). Ils permettent de montrer que pour le secteur, la diminution estimée de la teneur moyenne pondérée en matières grasses est principalement expliquée par des reformulations (contribution de -13%) malgré un renouvellement de l'offre qui va dans le sens d'une augmentation des teneurs (+6%). Les substitutions réalisées par les consommateurs, au sein des produits retrouvés les deux années d'étude, participent très faiblement à cette évolution (-0,8%).

Annexe 1 : Table de distribution du χ^2

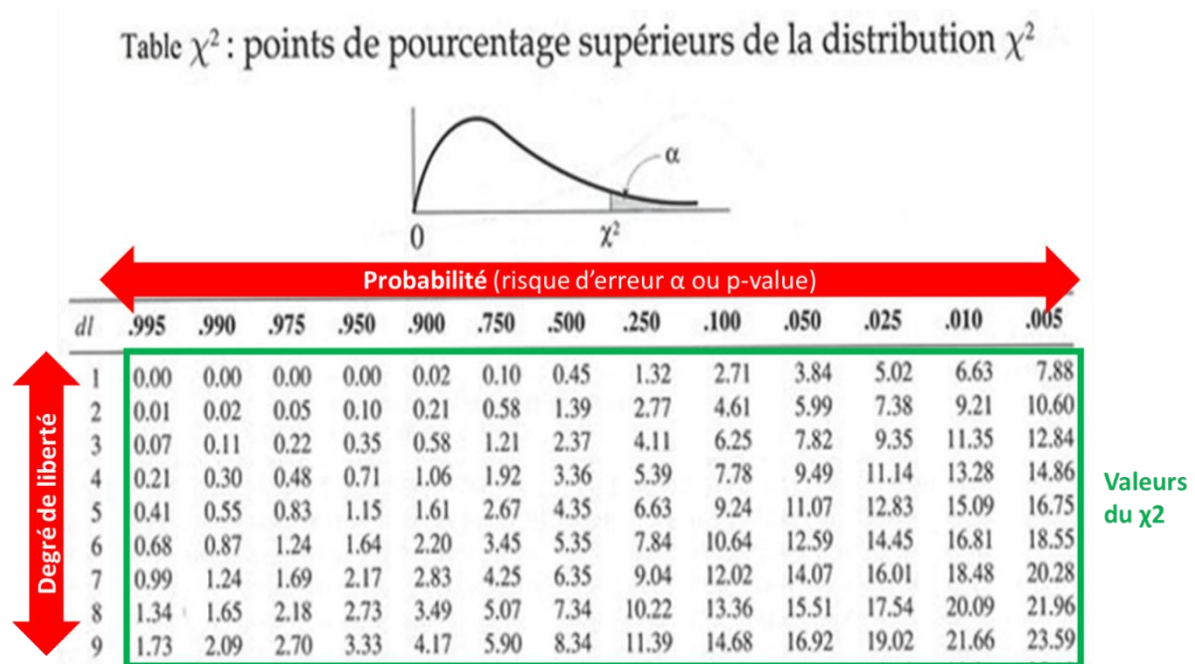


Figure 7 : Table de distribution du χ^2

Annexe 2 : Détail de l'algorithme utilisé par Griffith et al. (2017)²⁶ adapté aux données Oqali

Notons S_t la teneur moyenne en un nutriment pondérée par les parts de marché, sur un secteur, à l'instant t .

$$\text{On a } S_t = \sum_i w_{it} s_{it}$$

Avec w_{it} : la part de marché d'un produit i à l'instant t

s_{it} : la teneur en un nutriment donné d'un produit i à l'instant t

L'évolution de la moyenne pondérée en un nutriment est donnée par :

$$\Delta S_t = S_t - S_{t-1}$$

Celle-ci se traduit de la manière suivante lorsque l'on décompose cette évolution selon les différents sous-groupes N (produits nouveaux ou non captés au t_0), C (produits captés au t_0 et au t_1) et X (produits retirés ou non captés au t_1) :

$$\begin{aligned} \Delta S_t = & \sum_{i \in C} w_{it-1} \Delta s_{it} \\ & + \sum_{i \in N} w_{it} (s_{it} - S_{t-1}) - \sum_{i \in X} w_{it-1} (s_{it-1} - S_{t-1}) \\ & + \sum_{i \in C} (s_{it-1} - S_{t-1}) \Delta w_{it} + \sum_{i \in C} \Delta s_{it} \Delta w_{it} \end{aligned}$$

Le premier terme $\sum_{i \in C} w_{it-1} \Delta s_{it}$ rend compte de **la contribution de la reformulation des produits** captés au t_0 et au t_1 (groupe C), en considérant la part de marché du t_0 . Ainsi, la contribution à ce terme d'un produit reformulé sera d'autant plus importante que sa part de marché à t_0 était forte.

Le second terme $\sum_{i \in N} w_{it} (s_{it} - S_{t-1}) - \sum_{i \in X} w_{it-1} (s_{it-1} - S_{t-1})$ traduit **la contribution du renouvellement de l'offre de produits sur le marché**, c'est-à-dire l'apparition (groupe N) ou le retrait (groupe X) de références entre le t_0 et le t_1 . Il tient compte de la composition nutritionnelle relative de ces produits par rapport à la composition nutritionnelle moyenne pondérée du secteur à t_0 et de leur part de marché. Ainsi, par exemple, un produit issu du groupe N présentant une teneur en nutriment inférieure à la moyenne pondérée du secteur à t_0 contribuera à faire diminuer la moyenne pondérée du secteur, et ce d'autant plus que sa part de marché à t_1 est forte.

Le dernier terme $\sum_{i \in C} (s_{it-1} - S_{t-1}) \Delta w_{it} + \sum_{i \in C} \Delta s_{it} \Delta w_{it}$ reflète **l'évolution des choix des consommateurs**, à l'intérieur du groupe C (produits captés au t_0 et au t_1), plus précisément :

- le terme $\sum_{i \in C} (s_{it-1} - S_{t-1}) \Delta w_{it}$ rend compte de l'évolution de la part de marché d'un produit du groupe C ; il dépend également de l'écart de sa composition nutritionnelle à la composition moyenne pondérée du secteur à t_0 . Ainsi, si un produit présentant une teneur en nutriment supérieure à la moyenne pondérée du secteur à t_0 voit sa part de marché diminuer, cela contribuera à faire baisser la moyenne pondérée du secteur ;

le dernier terme $\sum_{i \in C} \Delta s_{it} \Delta w_{it}$ correspond à la covariance entre les teneurs en nutriments et les parts de marché des produits à l'intérieur du groupe C .

²⁶ Griffith, R., O'Connell, M., & Smith, K. (2017). The Importance of Product Reformulation Versus Consumer Choice in Improving Diet Quality. *Economica*, 84(333), 34-53. doi:<https://doi.org/10.1111/ecca.12192>

Annexe 3 : Teneurs en matières grasses et parts de marché fictives des produits d'un secteur donné, pour 2010 (T0) et 2016 (T1), après inférence des valeurs nutritionnelles et transformation des parts de marché avec la méthode 2

	Année 2010 (T ₀)		Année 2016 (T ₁)		
Produit	Teneur en matières grasses (g/100g)	Parts de marché ¹	Teneur en matières grasses (g/100g)	Parts de marché ¹	Sous-groupe
1	3,2	0,17	/	/	X
2	3,5	0,26	/	/	X
3	5,9	0,13	5,9	0,23	C
4	9,0	0,19	6,2	0,10	C
5	/	/	4,3	0,12	N
6	/	/	3,7	0,35	N
7	3,9	0,25	3,9	0,20	C

¹ ratio des volumes des produits identifiés par l'Oqali versus le volume total des produits identifiés par l'Oqali selon Kantar-Panel Worldpanel (données d'achats de ménages représentatives de la population française)

« / » : produit non existant ou non capté par l'Oqali

« Donnée manquante » : donnée non renseignée (la déclaration nutritionnelle affichée sur les emballages pouvait être absente ou présenter un niveau de détail variable)

Sous-groupes : « X » produits retirés ; « C » produits appariés ; « N » produits nouveaux