



MASTER 1 ECONOMÉTRIE-STATISTIQUES

Impact des préférences alimentaires sur le niveau de cholestérol

<i>Auteur</i>	<i>Id</i>
Nicolas TURPIN	12 011 092
Zhanike YEVNIYEVA	12 011 178

Présenté à :
Mme Hélène HUBER YAHY

15 janvier 2021

Table des matières

Introduction	1
1 Régression	2
1.1 Modèle de régression linéaire pour les lipoprotéines de haute densité et certaines covariables	2
1.2 Modèle de régression linéaire pour le cholestérol total et certaines covariables	3
2 Études sur la relation entre les habitudes alimentaires et la santé	4
2.1 Recherches existantes sur l'influence des régimes alimentaires	4
2.2 Origine du cholestérol et préférences alimentaires	5
3 Analyse coût efficacité aux Etats-Unis	7
3.1 Présentation des données et techniques utilisées	7
3.2 Résultats de l'analyse	8
Conclusion	9
Annexe	10
Références	13

Introduction

Ce rapport synthétise les hypothèses sur l'influence des préférences alimentaires sur le taux de cholestérol (lipoprotéine de haute densité (HDL), lipoprotéine de basse densité (LDL), cholestérol total) et son influence sur le risque de maladie cardiovasculaire. Nous avons également tenté de finaliser nos observations en discutant de la rentabilité des méthodes de traitement des maladies cardiovasculaires existantes. Une grande partie du rapport est concentrée sur les résultats des enquêtes menées par le Centre national des statistiques sanitaires des États-Unis. Dans un premier temps, nous allons fournir les résultats des modèles de régressions basés sur l'ensemble de données des enquêtes sur les dernières données disponibles. Ensuite, nous allons analyser les résultats des études existantes, y compris l'analyse des modèles alimentaires et son lien avec la maladie cardiovasculaire. Enfin, nous allons examiner l'analyse coût-efficacité du traitement des maladies cardiovasculaires aux États-Unis.

Il existe différentes opinions sur l'impact de la consommation de graisse en termes d'impact sur la santé (*Fats and Cholesterol* p. d.). Nous voulions examiner les recherches existantes décrivant les différents types de régimes alimentaires, les éléments nutritionnels et sa corrélation avec le volume de cholestérol dans l'organisme. Généralement, la plupart de nos hypothèses sont validées par des hypothèses énoncées et vérifiées dans les articles, cependant, il existe toujours des approches différentes qui peuvent conduire à des résultats différents.

1 Régression

Cette partie est consacrée à la description de certains des résultats de régression obtenus à l'aide du logiciel SAS. Afin de construire nos régressions, nous avons utilisé la base de données du Centre national des statistiques sanitaires des États-Unis, basée sur les résultats des enquêtes nationales sur la santé et la nutrition (NHANES) pour les années 2017-2018. Notre base de données comprenait 8704 observations qui ont été filtrées en fonction du sexe, de l'âge et de l'origine ethnique afin d'éliminer les entrées non éligibles. Nous allons procéder à l'analyse en deux étapes : d'abord, nous allons construire une régression linéaire pour les lipoprotéines de haute densité et certaines covariables, puis nous régresserons le taux total de cholestérol sur certains indicateurs alimentaires.

1.1 Modèle de régression linéaire pour les lipoprotéines de haute densité et certaines covariables

Nous étudions l'effet de nombreuses variables sur les lipoprotéines de haute densité (soi-disant «bon» cholestérol). Dans le modèle, nous prenons en compte le sexe, l'âge, l'appartenance ethnique, le niveau d'éducation, la catégorie d'IMC de la personne, des indicateurs nutritionnels spécifiques tels que les calories, les protéines, les glucides, les graisses saturées, et nous regardons également si la personne suit ou non un régime alimentaire spécifique.

Tests d'ajustement	
R carré	0.2367
Racine MSE	13.0948
DDL du dénominateur	15

Tableau 1: Ajustement du modèle

Tests des effets du modèle			
Effet	DDL num.	Valeur F	Pr >F
Modèle	15	290.19	<.0001
Intercept	1	511.12	<.0001
Sexe	1	73.37	<.0001
Ethnicité	4	12.96	<.0001
Âge au dépistage	1	35.48	<.0001
Niveau d'éducation	2	5.11	0.0203
Catégorie IMC	3	75.11	<.0001
Régime spécial	2	5.30	0.0182
Énergie (kcal)	1	30.28	<.0001
Protéine (gm)	1	15.45	0.0013
Glucides (g)	1	38.90	<.0001
Acides gras saturés totaux (g)	1	11.75	0.0037

Tableau 2: La significativité des variables

Alors que les variables semblent statistiquement significatives et fournissent des résultats attendus, on observe que le fait de ne pas suivre un régime spécifique a un effet positif plus important que le fait de le suivre (les coefficients estimés sont consécutivement égaux à 6,4782 et 4,3330). Modèle proche a été discuté dans l'article sur l'exploration des facteurs qui affectent le cholestérol sanguin et le risque de maladie cardiaque (KANTER et al. 2012). L'article suppose que de nombreux pays occidentaux n'ont pas de recommandations diététiques pour l'apport en cholestérol. Cependant, leurs directives diététiques se concentrent sur la réduction des graisses saturées, car elles pourraient avoir un effet puissant sur les facteurs de risque de maladies cardiovasculaires. Nous devons également tenir compte du fait que la question de savoir si la personne suit un régime correctif spécifique peut être liée au fait que la personne a déjà connu des problèmes de santé.

1.2 Modèle de régression linéaire pour le cholestérol total et certaines covariables

La deuxième partie de notre analyse de régression consiste à régresser le taux de cholestérol total sur certaines variables nutritionnelles. Dans le modèle, nous avons considéré les calories, les protéines, les glucides, les acides gras saturés, le lycopène, le magnésium et la niacine.

Tests d'ajustement	
R carré	0.5644
Racine MSE	171.77
DDL du dénominateur	15

Tableau 3: Ajustement du modèle

Tests des effets du modèle			
Effet	Estimation	Valeur F	Pr >F
Modèle		217.15	<.0001
Intercept	32.496	2.97	0.1054
Énergie (kcal)	0.087	17.15	0.0009
Protéine (gm)	4.260	202.83	<.0001
Glucides (g)	-0.580	22.61	0.0003
Acides gras saturés (g)	1.157	6.04	0.0266
Lycopène (mcg)	-0.001	10.51	0.0055
Magnésium (mg)	-0.313	13.55	0.0022
Niacine (mg)	-1.737	28.10	<.0001

Tableau 4: Ajustement du modèle

Toutes les variables de notre modèle semblent statistiquement significatives. Nous pouvons observer que la valeur du test t pour l'estimation de la variable «Protéines» est significativement plus élevée que pour les autres variables. La corrélation entre la consommation de protéines et le taux de cholestérol est positive, ce qui est attendu. L'article sur l'impact des protéines sur le taux de cholestérol note que les protéines, notamment d'origine animale, augmentent les concentrations de cholestérol total (FORSYTHE, GREEN et ANDERSON 1986). Sur

la base d'une étude dans le Journal du Collège américain de nutrition, un apport total plus élevé en glucides est lié au cholestérol total et aux lipoprotéines de basse densité plus faibles, ce qui est également le cas dans notre modèle (MA et al. 2006). On peut également observer une corrélation positive directe entre l'apport en graisses saturées et le cholestérol total. Cette observation est également étayée par les résultats de la recherche sur les acides gras saturés et le risque de maladie coronarienne : la consommation de graisses saturées augmente à la fois les taux de cholestérol à basse et à haute densité (SIRI-TARINO et al. 2010). Nous souhaitons maintenant nous concentrer sur l'interprétation des signes négatifs des coefficients des variables dans notre modèle. Cependant, l'impact de la valeur de l'ingestion de lycopène n'est pas important, la variable est toujours statistiquement significative. L'effet favorable du lycopène sur le cholestérol et la pression artérielle a été étudié dans le journal "Maturitas" (RIED et FAKLER 2011). Il a été suggéré dans l'article que l'apport en lycopène pourrait être efficace pour réduire le cholestérol de basse densité. Un autre modèle de corrélation négative est associé à l'apport en magnésium. Selon les recherches le magnésium peut être nécessaire pour convertir une lipoprotéine de basse densité en lipoprotéine de haute densité (SEELIG et ROSANOFF 2003). Une autre observation importante de notre modèle est liée à l'importance élevée de la variable niacine sur le cholestérol total. La niacine a été largement utilisée pour réguler les non-normalités du métabolisme des lipides plasmatiques et des lipoprotéines et dans le traitement des maladies cardiovasculaires athéroscléreuses, comme indiqué par une revue dans le journal de biochimie nutritionnelle ce que pourrait être observé dans les résultats de notre modèle de régression (GANJI, KAMANNA et KASHYAP 2003).

Nous pouvons conclure que, bien que dans le domaine scientifique, il existe de multiples arguments sur l'interprétation des variables influençant le cholestérol et les maladies connexes, notre modèle est toujours viable en ce qui concerne la signification statistique des variables utilisées. La prochaine partie de notre rapport se concentrera sur l'interprétation des résultats d'un autre modèle expliquant le modèle d'influence du régime alimentaire sur les maladies.

2 Études sur la relation entre les habitudes alimentaires et la santé

Dans cette section, nous allons discuter de quelques recherches intéressantes qui ont été menées afin d'étudier l'impact des habitudes alimentaires sur la santé.

2.1 Recherches existantes sur l'influence des régimes alimentaires

Des recherches récentes dans la revue BMC Medical Research Methodology se sont concentrées sur l'application d'une analyse de modèle telle que LASSO afin d'analyser les facteurs de risque liés à la maladie (ZHANG, TAPERA et GOU 2018). Afin d'obtenir de meilleurs résultats de prédiction, les auteurs de l'article ont utilisé la méthode LASSO (Least Absolute Shrinkage and Selection Operator) au lieu des techniques traditionnelles de clustering et de k-means. L'ensemble de données qui a été utilisé pour l'analyse statistique sont les résultats de NHANES de 2005 à 2006 de la même manière que les données que nous avons utilisées pour effectuer nos régressions (les résultats sont décrits dans la section précédente).

Dans le corps de l'article, les auteurs déclarent que, sur la base de l'analyse en composantes principales, il a été constaté que 10 composantes principales représentent ensemble 65% de la variation de l'ensemble de données. En outre, des composants principaux sélectionnés sont utilisés pour réaliser un modèle de régression linéaire multiple pour prédire le niveau de triglycérides, le cholestérol à lipoprotéines de basse densité, le cholestérol à lipoprotéines de haute densité et le cholestérol total. Ensuite, le modèle LASSO a été appliqué aux catégories d'aliments sélectionnées. Les résultats de l'analyse statistique ont montré que LASSO obtenait une précision de prédiction plus élevée et expliquait un plus grand pourcentage de variabilité des facteurs de risque qu'un modèle de régression multiple.

Les deux modèles ont prouvé leur viabilité et ont permis d'en tirer des informations : le composant principal caractérisé par une consommation élevée de produits laitiers riches en matières grasses, de viande rouge, de céréales raffinées et de sucreries était positivement associé au taux de triglycérides et au cholestérol total et négativement associé au cholestérol à lipoprotéines de haute densité, alors qu'un régime alimentaire sain représenté par une consommation élevée de fruits et légumes, corrélé négativement avec le taux de triglycérides et positivement corrélé avec le taux de cholestérol des lipoprotéines de haute densité.

Une autre recherche du Journal américain de la nutrition clinique a analysé les schémas alimentaires liés aux facteurs de risque de maladies cardiovasculaires (KERVER et al. 2003). L'étude visait à déterminer si les comportements alimentaires complexes sont regroupés selon des schémas liés aux risques de maladies cardiovasculaires. La recherche a révélé six modèles alimentaires dans lesquels il y avait 2 modèles prédominants caractérisant des apports élevés d'œufs, de viandes rouges et de produits laitiers riches en matières grasses et modèle sain caractérisé par des apports élevés de légumes et fruits.

2.2 Origine du cholestérol et préférences alimentaires

Les maladies cardiovasculaires (MCV) sont une des principales causes de décès aux Etats-Unis, nous pouvons en dénombrer environ un décès sur quatre (SOLIMAN 2018). Le point de repère de ces maladies est l'athérosclérose : il s'agit d'une affection inflammatoire chronique provoquée par le dépôt de cholestérol et de tissus fibreux dans les parois artérielles qui s'accumulent et conduisent finalement à un rétrécissement, un épaississement, ou un blocage de la lumière artérielle.

Afin de permettre un traitement et convenir de sa nécessité pour une maladie, il faut pouvoir cibler sa source. Dans le cas du cholestérol, de nombreuses études et recherches ont été effectuées pour comprendre le lien existant entre les préférences alimentaires et son taux excessif dans l'organisme. Beaucoup de controverses existent à ce sujet, mais nous allons synthétiser cela.

Il est à noter que les aliments à teneur élevée en cholestérol, le sont aussi en acide gras saturés. Nous pouvons en déduire qu'une consommation excessive de ces graisses peut favoriser le risque de MCV. C'est de cette manière que l'American Heart Association avait recommandé une consommation maximale de 300mg de cholestérol par jour. Cependant, de nombreuses preuves scientifiques expérimentales ont invalidés l'hypothèse selon laquelle le cholestérol alimentaire augmentait le cholestérol LDL/sanguin (responsable des MCV). Aujourd'hui, il est encore préconisé par les directives de réduire au maximum sa consommation de cholestérol alimentaire, tout en maintenant un régime alimentaire sain.

COMPOSITION NUTRITIVE DES ALIMENTS CIBLES					
Aliments pour 100 g					
Oeufs	Boeuf	Frommage	Poulet	Beurre	Crevette
<u>Protéines(g)</u>					
12.56	8.21	25	18.6	0.9	24
<u>Total lipides(g)</u>					
9.51	15	32.14	15.06	78.57	13.27
<u>Cholestérol(mg)</u>					
372	99	107	75	214	124
<u>Total choline(mg)</u>					
293.8	NR	NR	59.7	NR	NR

Tableau 5: Aliments principaux contenant du Cholestérol.

Comme le montre le [Tableau 5] ci dessus, l'oeuf est un aliment très riche en cholestérol. De ce fait, il a été considéré comme un aliment à limiter dans sa consommation. Plusieurs revues indiquent de n'en consommer que trois, voir un seul maximum par semaine. Cependant, il est apparu que l'oeuf est la seule source alimentaire de cholestérol pauvre en acide gras saturés, et riche en nutriments. C'est pourquoi, d'autres scientifiques stipulent qu'il est possible d'en consommer plusieurs par jour sans que cela n'affecte négativement notre état de santé. De plus, nous pouvons voir que les oeufs contiennent de la choline. Elle joue un rôle très important : celui du développement du cerveau foetal et néonatal. Une choline inadéquate chez les femmes enceintes augmente le risque de malformations du tube neural chez la progéniture. Une étude portant sur plus de 130 400 sujets, âgés de 40 à 69 ans, à montré qu'une consommation de 7 oeufs par semaine réduisait le risque de syndrome métabolique par rapport à la consommation de 1 oeuf par semaine.

Aux Etats-Unis, la plupart des études de cohorte à ce sujet suggèrent soit une corrélation négative entre la consommation d'oeufs chez les patients diabétiques, soit aucun effet. On pourrait penser que la différence de ces résultats proviendrait de variables de confusions ou a des différences dans les habitudes alimentaires selon les régions du monde. Pour notre cas, ce qu'il faut retenir, c'est que la littérature actuelle n'approuve pas l'idée que le cholestérol alimentaire augmente le risque de maladie cardiaque chez une personne en bonne santé ; pourtant, beaucoup de preuves existent sur le fait que les acides gras saturés et les acides gras trans gonflent le risque de maladie cardiovasculaire. Comme nous pouvons le comprendre, il n'est pas aisé de se faire une idée précise sur la manière de se prévenir "alimentairement parlant", pour ce genre de maladie. Heureusement, plusieurs traitements sont apparus au fil des années pour venir en aide aux individus souffrant de MCV.

3 Analyse coût efficacité aux Etats-Unis

Nous nous intéressons ici à l'analyse de coût efficacité, c'est à dire à une analyse permettant de mieux comprendre les enjeux d'une décision gouvernementale quant à la prise en charge d'un nouveau traitement ou non par ce dernier, pour une maladie donnée. Dans le cadre de notre étude, un traitement face au cholestérol a suscité notre curiosité : le traitement par statines.

3.1 Présentation des données et techniques utilisées

Plusieurs traitements ont vu le jour pour répondre à la problématique des MCV. Un traitement s'est révélé assez bon : celui du traitement par statines (PROSSER et al. 2000, GARBER 2000). Nous avons voulu savoir si une extension de ce traitement pour des profils différents serait efficace, ou encore, coût efficace. L'American College of Cardiology et l'American Heart Association ont mené une étude de 100 cohortes pour répondre à cette question (KOHLI-LYNCH et al. 2019). L'étude de cohorte a été augmentée d'une procédure de micro-simulation. L'association de ces deux méthodes ressemble à ce que l'on fait par bootstraps : les résultats obtenues sont issus de créations par échantillonnages probabilistes à partir d'enquêtes nationales américaines sur la santé et la nutrition. Les enquêtes couvrent la période 1999-2014 sur la position du secteur américain par rapport aux soins de santé. Chaque cohorte comprend 1 000 000 d'individus répondants négatif à l'enquête ASCVD (premier événement cardiovasculaire athéroscléreux : il s'agit de l'apparition d'un premier symptôme ressenti et avéré de maladie cardiovasculaire athérosclérotique). Toutes les cohortes sont composées de 50% d'hommes et de 50% de femmes âgés de 40 ans au départ. Toutes les données analysées étaient désidentifiées et provenant d'études ayant achevé la collecte de données dans le passé. Cette analyse a été revue et approuvée par le comité d'examen institutionnel du Columbia University Medical Center. D'autres variables du modèle ont été dérivées d'enquêtes nationales, de méta-analyses et de publications ; les données ont été analysées pendant un peu plus d'un an, soit du 15 mai 2018 au 10 juin 2019.

On note "AR10" : le risque absolu en 10 ans ; et "LDL-C" : cholestérol à lipoprotéines de basse densité.

Les traitements par statines		
Numéro traitement	AR10	LDL-C(mg/dL)
0 (traitement en place)	$\geq 7.5\%$	≥ 190
1	[5%,7.4%]	[160,189]
2	[5%,7.4%]	[130,159]
3	$\geq 5\%$	\forall

Tableau 6: Les différentes stratégies de traitement par statines

Quatre stratégies de traitement par statines ont été comparées :

→ La première a été de traiter les patients avec un AR10 d'au moins 7.5% et un diabète ou un LDL-C d'au moins 190mg/dL (ceci correspond au traitement en vigueur).

→ La deuxième a été d'ajouter un traitement par statines d'intensité modérée pour un AR10 allant de 5 à 7.4% et un taux de LDL-C allant de 160 à 189 mg/dL.

→ La troisième correspond à l'ajout d'un traitement par statines d'intensité modérée avec le même AR10 que précédemment (de 5 à 7.4%) mais cette fois avec un taux de LDL-C allant de 130 à 159 mg/dL.

→ La quatrième correspond à l'ajout d'un traitement par statines d'intensité modérée avec un AR10 d'au moins 5%, quel que soit le niveau de LDL-C.

A noter que le traitement par statine a également été comparé à l'absence de traitement par statine dans les strates d'âge, de sexe, d'AR10 et de LDL-C. On s'attend également à ce que le coût par années de vie gagnées augmente au fur et à mesure qu'il touche plus de monde (nombre de cas plus important).

3.2 Résultats de l'analyse

Vont alors apparaître des résultats sur les années de vie ajustées sur la qualité de vie (QALY), et les coûts (dollars américain de 2019, projetés et actualisés de 3.0% par an) ; ainsi que le résultat principal attendu en terme de ratio coût-efficacité (ICER).

Nous allons présenter les résultats de ces 100 cohortes, relatifs aux stratégies du [Tableau 6] à l'aide du tableau suivant :

Tableau récapitulatif		
Numéro de Stratégie	Résultat	ICER
1	Efficace*	$\approx 22000\$/\text{QALY}$
2	Efficace**	$\approx 27000\$/\text{QALY}$
3	Efficace***	$\approx 33558\$/\text{QALY}$

Tableau 7: Résultat des stratégies de traitement par statines

Cette analyse a révélée que les femmes étaient moins touchées que les hommes, les ICER affichés ci-dessus correspondent à une moyenne "homme/femme" où le traitement des hommes tend à faire augmenter ce ratio, et inversement pour les femmes.

Les stratégies ont été classées comme coûts-efficaces pour l'horizon temporel fixé (c'est à dire qu'elles sont sujettes à une hausse incrémentielle des QALY pour des coûts inférieurs à celui du seuil gouvernemental). Elles ont été classées comme ceci :

→ Très rentable ($\text{ICER} \geq 0$ \$ par QALY mais < 50000 \$ par QALY gagnée).

→ Moyennement rentable ($\text{ICER} \geq 50000$ \$ par QALY mais < 150000 \$ par QALY gagnée).

→ Pas rentable ($\text{ICER} \geq 150000$ \$ par QALY gagnée).

Dans les catégories d'âge, d'AR10 et de sexe, les personnes ayant des taux de LDL-C de base plus élevés, ont gagné plus de QALY grâce au traitement par statine. La stratégie numéro 3, avec la spécification d'un AR10 d'au moins 5%, semble être la plus efficace. En effet, cette stratégie présente un ratio cout-efficacité très important, (à hauteur de 33558\$/QALY), mais permettrait d'éviter la plupart des évènements ASCVD. Les résultats suggèrent donc que le traitement de tous les patients à bas risque (par extension du traitement en place), et ce, quel que soit le niveau de LDL-C, serait la stratégie adoptée si le gouvernement consacre un budget pour les dépenses de soins supérieur à ce ratio. Synthétiquement, si les décideurs étaient prêts à payer la somme de 50000\$ par QALY gagnés pour résoudre plus largement le problème des MCV, traiter les personnes avec un AR10 d'au moins 5% serait la stratégie

privilégiée, notamment parce-qu'elle touche un public plus large et qu'elle se révèle plus efficace que les stratégies 0,1,2 dans le cadre du traitement par statine.

Conclusion

En définitive, nous avons vu que le niveau de cholestérol chez un individu a tendance à augmenter en fonction du tabagisme, du sexe, de l'âge, de l'origine ethnique ou encore d'autres facteurs. De plus, bien que le taux de cholestérol semble également augmenter en fonction du poids corporel, nous avons constaté qu'il est difficile de faire un lien direct entre les préférences alimentaires et ce taux. Pour cette raison, les spécialistes de santé suggèrent encore de limiter notre consommation en aliments lipidiques : notamment pour l'influence qu'elle transmet sur le poids corporel ; et donc indirectement sur le taux de mauvais cholestérol. Enfin, le traitement de tous les patients à risque d'au moins 5% sur 10ans, serait très rentable en terme d'années de vies gagnées si la société était disposée à consacrer des ressources importantes aux soins de santé pour la prévention de maladie cardiovasculaire athérosclérotique.

Ces informations nous laissent à penser que nous sommes encore faiblement armés face aux maladies, mais également que le budget consacré aux besoins de soins est encore trop faible où non prioritaire. Il serait aisé de spéculer face aux décisions gouvernementales sur le budget accordé aux soins et un probable lien avec le surpeuplement actuel. En effet, il est à noter que plus nous sommes nombreux, plus nous sommes amenés à consommer les ressources de notre planète de manière inappropriée et massivement. Il en résulte alors des effets indésirables sur la santé de notre sphère bleue (réchauffement climatique). Comment pourrions-nous faire pour pallier à cette double problématique ? Peut-être que tout dépendrait, finalement, directement de notre propre consommation.

Annexe

Extrait du code du logiciel SAS exécuté afin d'obtenir des résultats de régression (le code complet est joint en tant que fichier .sas séparé) :

```
/* LINEAR REGRESSION */
LIBNAME NH "C:\Users\zhani\Documents\NHANES_database";
OPTIONS NODATE NOCENTER;
options ls=72;
proc format;
  VALUE sexfmt      1 = 'Male'
                   2 = 'Female';
  VALUE sex2fmt     1 = 'Female'
                   2 = 'Male';
  VALUE race2fmt    1='Mexican Americans'
                   3='Non-Hispanic white'
                   4='Non-Hispanic black';
  VALUE race3fmt    1='Mexican American'
                   2='Non-Hispanic black'
                   5='Non-Hispanic white';
  VALUE smkfmt      1='Never smoker'
                   2='Past smoker'
                   3='Current smoker';
  VALUE educ        1="< HS"
                   2="HS/GED"
                   3="> HS";
  VALUE bmicat      1="underweight"
                   2="normal weight"
                   3="overweight"
                   4="obese";
  VALUE bmicatf     1="under weight"
                   2="overweight"
                   3="obese"
                   4="normal weight";
run;
data nh.analysis_data;
set nh.analysis_data;
if ridstatr=2; *all mec exam data;

/*set don't know and refused (7,9) to missing*/
if dmddeduc>3 then dmddeduc=.;

/*define smokers*/
if smq020 eq 2 then smoker=1;
else if smq020 eq 1 and smq040 eq 3 then smoker=2;
```

```
else if smq020 eq 1 and smq040 in(1,2) then smoker=3;

/*recode gender*/
if riagendr eq 1 then sex=2;
else if riagendr eq 2 then sex=1;

/*recode race/ethnicity*/
ethn=ridreth1;
if ridreth1 eq 3 then ethn=5;
else if ridreth1 eq 4 then ethn=2;
else if ridreth1 eq 2 then ethn=3;
else if ridreth1 eq 5 then ethn=4;

if 0 le bmxbmi lt 18.5 then bmicatf=1;
else if 18.5 le bmxbmi lt 25 then bmicatf=4;
else if 25 le bmxbmi lt 30 then bmicatf=2;
else if bmxbmi ge 30 then bmicatf=3;

if 0 le bmxbmi lt 18.5 then bmicat=1;
else if 18.5 le bmxbmi lt 25 then bmicat=2;
else if 25 le bmxbmi lt 30 then bmicat=3;
else if bmxbmi ge 30 then bmicat=4;

if (LBDHDD^=. and riagendr^=. and ridreth1^=. and smoker^=. and dmddeduc^=.
and bmxbmi^=.)and WTMEC2YR>0
    and (ridageyr>=20) then eligible=1;
    *else eligible=2;

label riagendr='Gender'
    sex = 'Gender - recode'
    ridreth1='Race/ethnicity'
    ridageyr='Age in years'
    ethn = 'Race/ethnicity - recode'
    dmddeduc='Education'
    bmicatf='BMI category';

run;

PROC SURVEYREG data=nh.analysis_data nomcar;
STRATA sdmvstra;
CLUSTER sdmvpsu;
WEIGHT WTMEC2YR;
CLASS sex ethn dmddeduc smoker bmicatf DRQSDIET;
DOMAIN eligible;
MODEL LBDHDD = sex ethn ridageyr dmddeduc bmicatf DRQSDIET DR1TKCAL DR1TPROT
DR1TCARB DR1TSFAT /CLPARM solution vadjust=none;
```

```
ESTIMATE 'Never vs past smoker' smoker 1 -1 0;
TITLE    'Linear regression model for high density lipoprotein and
selected covariates: NHANES 2017-2018';
run;

PROC SURVEYREG data=nh.analysis_data nomcar;
STRATA sdmvstra;
CLUSTER sdmvpsu;
WEIGHT WTMEC2YR;
CLASS sex ethn smoker dmddeduc bmicatf;
DOMAIN eligible;
MODEL DR1TCHOL = DR1TKCAL DR1TPROT DR1TCARB DR1TSFAT DR1TLYCO DR1TMAGN
DR1TNIAAC /CLPARM solution vadjust=none;;
TITLE    'Linear regression model for total cholesterol and selected
covariates: NHANES 2017-2018';
run;
```

Références

- [1] *Fats and Cholesterol*. URL : <https://www.hsph.harvard.edu/nutritionsource/what-should-you-eat/fats-and-cholesterol>. (accessed : 21/12/2020).
- [2] W. A. FORSYTHE, M. S. GREEN et J. J. ANDERSON. « Dietary protein effects on cholesterol and lipoprotein concentrations : a review. (Anglais) ». In : *Journal of the American College of Nutrition* 5.6 (1986), p. 533-549. DOI : <http://dx.doi.org/10.1080/07315724.1986.10720155>.
- [3] S. H. GANJI, V. S. KAMANNA et M. L. KASHYAP. « Niacin and cholesterol : role in cardiovascular disease (review). (Anglais) ». In : *The Journal of Nutritional Biochemistry* 14.6 (2003), p. 298-305. URL : <https://pubmed.ncbi.nlm.nih.gov/12873710/>.
- [4] A. M. GARBER. « Using Cost-Effectiveness Analysis To Target Cholesterol Reduction. (Anglais) ». In : *Annals of Internal Medicine* 132.10 (2000), p. 833.
- [5] M. M. KANTER, P. M. KRIS-ETHERTON, M. L. FERNANDEZ, K. C. VICKERS et D. L. KATZ. « Exploring the Factors That Affect Blood Cholesterol and Heart Disease Risk : Is Dietary Cholesterol as Bad for You as History Leads Us to Believe?. (Anglais) ». In : *Advances in Nutrition* 3.5 (2012), p. 711-717. DOI : <https://doi.org/10.3945/an.111.001321>.
- [6] J. M. KERVER, E. J. YANG, L. BIANCHI et W. O. SONG. « Dietary patterns associated with risk factors for cardiovascular disease in healthy US adults. (Anglais) ». In : *The American Journal of Clinical Nutrition* 78.6 (2003), p. 1103-1110. URL : <https://academic.oup.com/ajcn/article/78/6/1103/4677520>.
- [7] C. N. KOHLI-LYNCH, B. K. BELLOWES, G. THANASSOULIS, Y. ZHANG, M. J. PLETCHER, E. VITTINGHOFF, M. J. PENCINA, D. KAZI, A. D. SNIDERMAN et A. E. MORAN. « Cost-effectiveness of Low-density Lipoprotein Cholesterol Level-Guided Statin Treatment in Patients With Borderline Cardiovascular Risk. (Anglais) ». In : *JAMA Cardiology* 4.10 (2019), p. 969.
- [8] Y. MA, Y. LI, D. E. CHIRIBOGA, B. C. OLENDZKI, J. R. HEBERT, W. LI, K. LEUNG, A. R. HAFNER et I. S. OCKENE. « Association between Carbohydrate Intake and Serum Lipids. (Anglais) ». In : *Journal of the American College of Nutrition* 25.2 (2006), p. 155-163. URL : <https://pubmed.ncbi.nlm.nih.gov/16582033>.
- [9] L. A. PROSSER, A. A. STINNETT, P. A. GOLDMAN, L. W. WILLIAMS, M. G. HUNINK, L. GOLDMAN et M. C. WEINSTEIN. « Cost-Effectiveness of Cholesterol-Lowering Therapies according to Selected Patient Characteristics. (Anglais) ». In : *Annals of Internal Medicine* 132.10 (2000), p. 769.
- [10] K. RIED et P. FAKLER. « Protective effect of lycopene on serum cholesterol and blood pressure : Meta-analyses of intervention trials. (Anglais) ». In : *Maturitas* 68.4 (2011), p. 299-310. URL : [https://www.maturitas.org/article/S0378-5122\(10\)00446-9/fulltext](https://www.maturitas.org/article/S0378-5122(10)00446-9/fulltext).
- [11] M. S. SEELIG et A. ROSANOFF. *The magnesium factor*. Avery, 2003.

- [12] P. W. SIRI-TARINO, Q. SUN, F. B. HU et R. M. KRAUSS. « Saturated Fatty Acids and Risk of Coronary Heart Disease : Modulation by Replacement Nutrients. (Anglais) ». In : *Current Atherosclerosis Reports* 12.6 (2010), p. 384-390. URL : <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2943062>.
- [13] G. SOLIMAN. « Dietary Cholesterol and the Lack of Evidence in Cardiovascular Disease. (Anglais) ». In : *Nutrients* 10.6 (2018), p. 780. URL : <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6024687/>.
- [14] F. ZHANG, T. M. TAPERA et J. GOU. « Application of a new dietary pattern analysis method in nutritional epidemiology. (Anglais) ». In : *MC Medical Research Methodology* 18.1 (2018). URL : <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6206725/#CR6>.