



INTITULÉ DU PROJET

Facteurs liés à la naissance d'un enfant.

MEMBRES

CISSE Lansana
KARAMOKO Awa
KPAMEGAN Falonne
LAOKPEZI Solim

— université
— LUMIÈRE
— LYON 2

Année-académique : 2022-2023

Sommaire

INTRODUCTION	2
I. CONTEXTE	3
II. DESCRIPTION	3
III. SOURCE DU DATASET.....	3
IV. PRESENTATION DES DONNNEES DU FICHIER D'ETUDE	3
V. Choix de la méthode	4
VI. Analyse des facteurs qui influence le poids d'un nouveau née	5
1. Analyse de la corrélation.....	5
2. L'utilisation de l'ACP et du cercle de corrélation	6
VII. Affichage du cercle de corrélation.....	7
Interprétation.....	8
VIII. Analyse des facteurs liée à la naissance prématuré d'un bébé	8
1. Représentation des 10 individus et les 10 catégories.....	8
2. Choix du nombre d'axes - Eboulis des valeurs propres.....	10
3. Interprétation de Chaque dimension de l'ACP.....	12
IX. Représentation graphique.....	13
CONCLUSION	15
BIBLIOGRAPHIE	16
TABLE DES MATIERE.....	17

INTRODUCTION

La naissance d'un nouveau-né est un événement passionnant et émouvant pour les parents, mais cela peut aussi être une source d'inquiétude. Le poids d'un nouveau-né et son état de santé à la naissance sont deux indicateurs importants pour évaluer la santé du bébé. Cependant, le poids de naissance d'un bébé et la probabilité de naissance prématurée sont influencés par de nombreux facteurs, tels que l'âge et l'état de santé de la mère, la génétique, le mode de vie, l'accès aux soins prénatals et les conditions environnementales. Dans ce travail, nous allons explorer les différents facteurs qui peuvent influencer le poids d'un nouveau-né et les facteurs liés à la naissance prématurée d'un bébé. Nous allons également discuter de l'importance de ces facteurs pour la santé du nouveau-né et de la mère, et des mesures qui peuvent être prises pour minimiser les risques liés à ces facteurs.

I. CONTEXTE

L'analyse de l'ensemble de données sur les naissances aux États-Unis en 2014 permettra de mieux comprendre les facteurs qui influencent le poids d'un nouveau-né et les facteurs liés à la naissance prématurée d'un bébé. Le poids de naissance est un indicateur important de la santé d'un nouveau-né et peut être influencé par plusieurs facteurs tels que l'âge de la mère, le nombre de visites à l'hôpital pendant la grossesse, le poids gagné par la mère pendant la grossesse, l'addiction de la mère au tabac, etc. De même, la naissance prématurée est associée à des risques de santé plus élevés pour le bébé et peut être liée à des facteurs tels que l'âge de la mère, le nombre de visites à l'hôpital pendant la grossesse, etc.

Dans ce rapport, nous analyserons les données sur les naissances de 2014 en utilisant des techniques d'analyse de données pour identifier les facteurs qui influencent le poids de naissance d'un nouveau-né et les facteurs liés à la naissance prématurée d'un bébé. Cette analyse aidera à mieux comprendre les tendances en matière de naissances aux États-Unis et à identifier les facteurs clés qui peuvent être utilisés pour informer les politiques et les interventions de santé publique pour améliorer la santé des nourrissons et des mères.

II. DESCRIPTION

Chaque année, les États-Unis publient à l'intention du public un grand ensemble de données contenant des informations sur les naissances enregistrées dans le pays. Cet ensemble de données a suscité l'intérêt des chercheurs médicaux qui étudient la relation entre les habitudes et les pratiques des futures mères et la naissance de leurs enfants. Il s'agit d'un échantillon aléatoire de 1 000 cas de l'ensemble de données publié en 2014.

III. SOURCE DU DATASET

United States Department of Health and Human Services. Centers for Disease Control and Prevention. National Center for Health Statistics. Natality Detail File, 2014 United States. Inter-university Consortium for Political and Social Research, 2016-10-07. <https://doi.org/10.3886/ICPSR36461.v1>.

IV. PRESENTATION DES DONNÉES DU FICHIER D'ÉTUDE

Un cadre de données avec 1 000 observations sur les 13 variables suivantes :

[Tableau explicatif](#)

Variable	Description	Type de variable
----------	-------------	------------------

fage	Âge du père en années.	Quantitative
mage	Âge de la mère en années.	Quantitative
mature	Statut de maturité de la mère.	Qualitative
gained	Poids pris par la mère pendant la grossesse en livres.	Quantitative
weigth	Poids du bébé à la naissance en livres.	Quantitative
weeks	Durée de la grossesse en semaines.	Quantitative
visits	Nombre de visites à l'hôpital pendant la grossesse.	Quantitative
habit	Statut de la mère en tant que non-fumeuse ou fumeuse.	Qualitative
marital	Si la mère est mariée ou non mariée à la naissance.	Qualitative
whitemom	Si la maman est blanche ou non blanche.	Qualitative
premie	Si la naissance a été classée comme prématurée (premie) ou à terme.	Qualitative
sex	Sexe du bébé, féminin ou masculin.	Qualitative
lowbirthweight	Sexe du bébé, féminin ou masculin.	Qualitative

V. Choix de la méthode

Notre projet porte sur deux analyses :

- ✓ Analyse des facteurs qui influence le poids d'un nouveau née ;
- ✓ Analyse des facteurs liée à la naissance prématuré d'un bébé ;

Dans un premier temps, pour l'analyse des facteurs qui influence le poids d'un nouveau née, nous allons utiliser la méthode ACP et l'analyse de corrélation car les variables à étudier sont toute quantitatives.

Dans un second temps, pour l'analyse des facteurs liée à la naissance prématuré d'un bébé, nous allons utiliser la méthodes ACM car les variables à étudier sont quantitatives et qualitatives.

VI. Analyse des facteurs qui influence le poids d'un nouveau née

1. Analyse de la corrélation

La fonction `cor()` renvoie une matrice de corrélation indiquant la corrélation entre chaque paire de variables. Nous pouvons interpréter les valeurs de corrélation comme suit :

Si la valeur est proche de 1 ou -1, cela indique une forte corrélation positive ou négative entre les variables.

Si la valeur est proche de 0, cela indique une faible corrélation entre les variables

##	fage	mage	weeks	visits	gained
weight					
## fage	1.000000000	0.63743297	-0.01635314	-0.000302348	-0.03642077
0.02402910					
## mage	0.637432974	1.00000000	-0.01070783	0.058992905	-0.04179767
0.06277649					
## weeks	-0.016353142	-0.01070783	1.00000000	0.196235819	0.04073482
0.54804813					
## visits	-0.000302348	0.05899290	0.19623582	1.000000000	0.09353207
0.18078725					
## gained	-0.036420773	-0.04179767	0.04073482	0.093532066	1.00000000
0.14207451					
## weight	0.024029105	0.06277649	0.54804813	0.180787253	0.14207451
1.00000000					

Dans cette matrice de corrélation, on peut observer les points suivants :

- La corrélation entre l'âge du père et l'âge de la mère est de 0,6374, ce qui indique une corrélation positive forte. Cela signifie que plus l'âge du père est élevé, plus l'âge de la mère est susceptible d'être élevé également.
- La corrélation entre le poids de la mère pendant la grossesse et la durée de la grossesse est de 0,0407, ce qui indique une corrélation positive faible. Cela signifie que plus le poids de la mère pendant la grossesse est élevé, plus la durée de la grossesse est susceptible d'être élevée également, mais la corrélation est très faible.
- La corrélation entre le poids du bébé à la naissance et la durée de la grossesse est de 0,5480, ce qui indique une corrélation positive forte. Cela signifie que plus la durée de la grossesse est longue, plus le poids du bébé est susceptible d'être élevé.

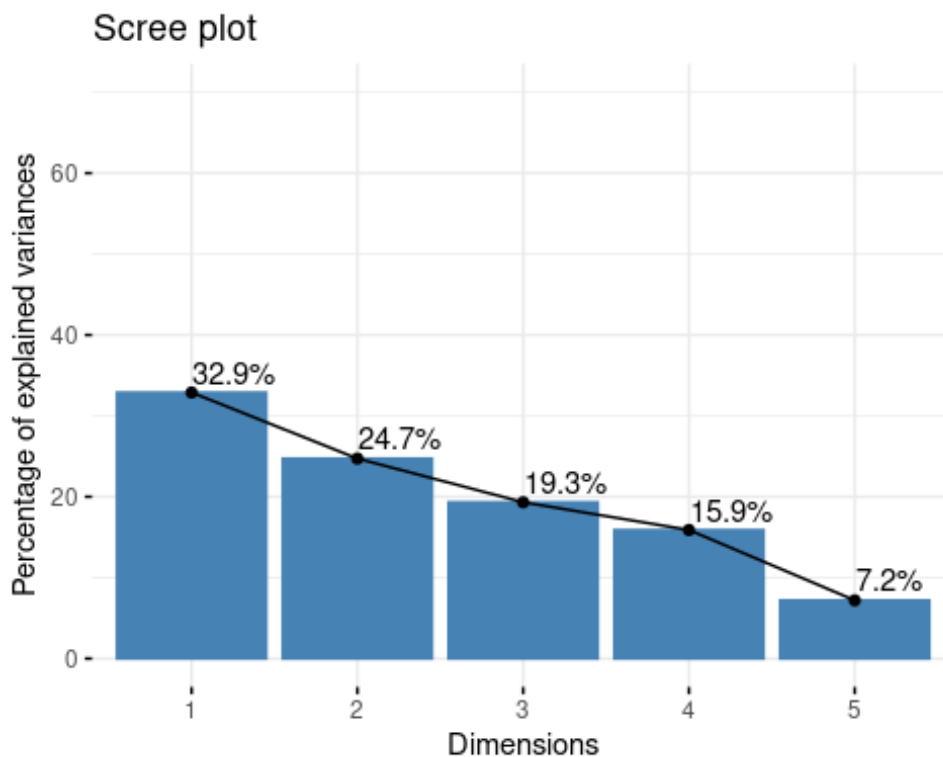
En conclusion, cette matrice de corrélation indique qu'il y a une corrélation forte et positive entre l'âge du père et l'âge de la mère, une corrélation faible entre le poids de la mère pendant la

grossesse et la durée de la grossesse, et une corrélation forte et positive entre le poids du bébé à la naissance et la durée de la grossesse.

2. L'utilisation de l'ACP et du cercle de corrélation

Affichage de la contribution

##	eigenvalue	variance.percent	cumulative.variance.percent
## Dim.1	1.6442073	32.884146	32.88415
## Dim.2	1.2360521	24.721041	57.60519
## Dim.3	0.9657317	19.314634	76.91982
## Dim.4	0.7944868	15.889736	92.80956
## Dim.5	0.3595222	7.190443	100.00000



Il y a 5 valeurs propres, autant que de variable actives. Nous retenons les 2 premiers axes auxquels nous allons donner une sémantique.

Analyse des résultats :

##	coord	contrib	cos2
## fage	0.90108582	49.38280430	0.8119556631
## mage	0.90415197	49.71944854	0.8174907921
## weeks	-0.02627738	0.04199595	0.0006905005
## visits	0.05946833	0.21508737	0.0035364821
## gained	-0.10263451	0.64066385	0.0105338416

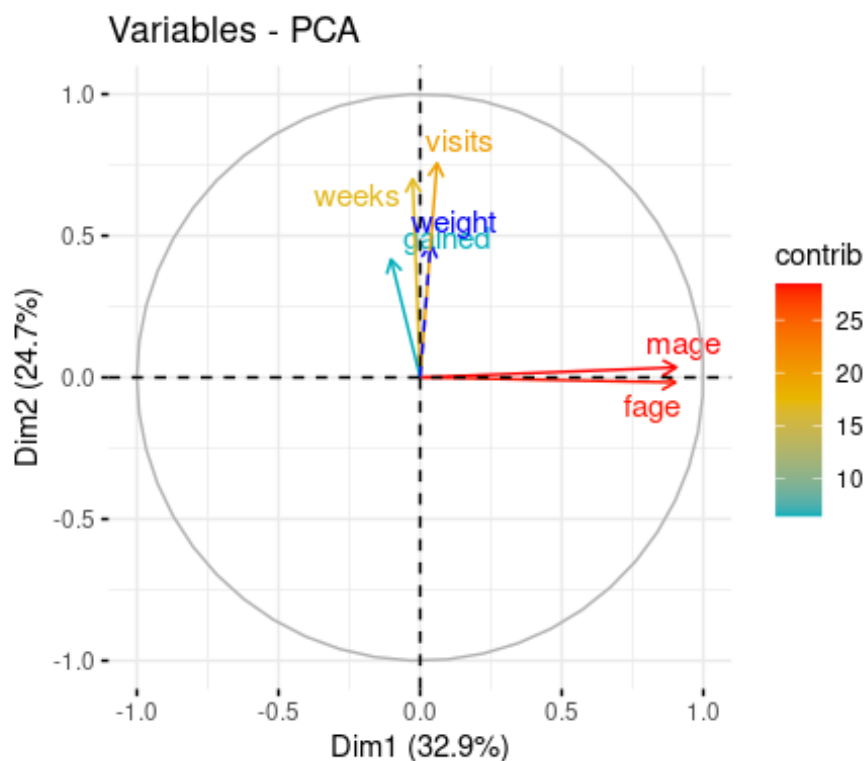
##	coord	contrib	cos2
## fage	-0.01769763	0.02533924	0.0003132061
## mage	0.03556373	0.10232409	0.0012647790
## weeks	0.69979163	39.61874595	0.4897083239
## visits	0.75575619	46.20900977	0.5711674153
## gained	0.41665133	14.04458096	0.1735983317

Les résultats de l'ACP indiquent que les 6 variables étudiées (l'âge du père, l'âge de la mère, le gain de poids pendant la grossesse, le nombre de semaines de grossesse, le nombre de visites et le poids du bébé) sont étroitement liées et peuvent être résumées en deux composantes principales. La première composante principale (Dim.1) est principalement liée au gain de poids pendant la grossesse, au nombre de semaines de grossesse, le poids du bébé et au nombre de visites, tandis que la deuxième composante principale (Dim.2) est principalement liée à l'âge du père et l'âge de la mère.

La première composante principale explique 32,9% de la variance totale, alors que la deuxième explique 24,7%, pour un total de 57,7%. Cela signifie que ces deux composantes principales représentent la majeure partie de la variance totale des données.

L'interprétation des axes de l'ACP est la suivante : la première composante principale (Dim.1) représente la taille du bébé, avec des valeurs élevées indiquant un bébé plus gros, alors que la deuxième composante principale (Dim.2) représente la santé de la mère pendant la grossesse, avec des valeurs élevées indiquant une mère plus en santé.

VII. Affichage du cercle de corrélation



Interprétation

L'analyse en composantes principales (ACP) et l'analyse de corrélation suggèrent plusieurs conclusions :

Les variables visits (nombre de visites prénatales), gained (le poids de la mère pendant la grossesse) et weeks (nombre de semaines de la grossesse) sont modérément corrélées avec le poids du bébé (weight) à la naissance sur l'axe 1. Par contre, les variables mage (age de la mère) et fage (age du père) sont positivement corrélés entre elles sur l'axe 2.

Le nombre de visites pendant la grossesse (visits) est la variable la plus importante pour la première dimension de l'ACP (Dim.1) avec une contribution positive élevée.

Les résultats de l'ACP et de l'analyse de corrélation suggèrent que le poids du bébé à la naissance est influencé à la fois par les facteurs maternels (le poids de la mère) et ceux de la grossesse (le nombre de visites et le nombre de semaine de la grossesse).

En conclusion

Pour améliorer le poids du bébé à la naissance, il peut être utile de se concentrer sur la gestion du poids pendant la grossesse et de surveiller attentivement les facteurs maternels tels que le poids de la mère, les visites prénatales ainsi que le nombre de semaines de la grossesse.

VIII. Analyse des facteurs liée à la naissance prématuré d'un bébé

1. Représentation des 10 individus et les 10 catégories.

```
##                               Dim.204 Dim.205
## Variance                      0.000    0.000
## % of var.                     0.000    0.000
## Cumulative % of var. 100.000 100.000
##
## Individuals (the 10 first)
##      Dim.1   ctr   cos2   Dim.2   ctr   cos2   Dim.3   ctr   cos2
## 1 |  0.077  0.002  0.001 | -0.264  0.030  0.009 | -0.138  0.009  0.002
## 2 |  0.252  0.025  0.006 | -0.236  0.024  0.005 | -0.196  0.018  0.003
## 3 |  0.991  0.383  0.101 |  0.194  0.016  0.004 | -0.066  0.002  0.000
## 4 | -0.340  0.045  0.005 |  0.181  0.014  0.001 |  0.345  0.055  0.005
## 5 |  0.209  0.017  0.002 |  0.495  0.105  0.012 |  0.351  0.057  0.006
## 6 |  0.075  0.002  0.001 | -0.300  0.039  0.010 | -0.163  0.012  0.003
## 7 |  1.183  0.546  0.127 |  0.945  0.384  0.081 |  0.308  0.044  0.009
```

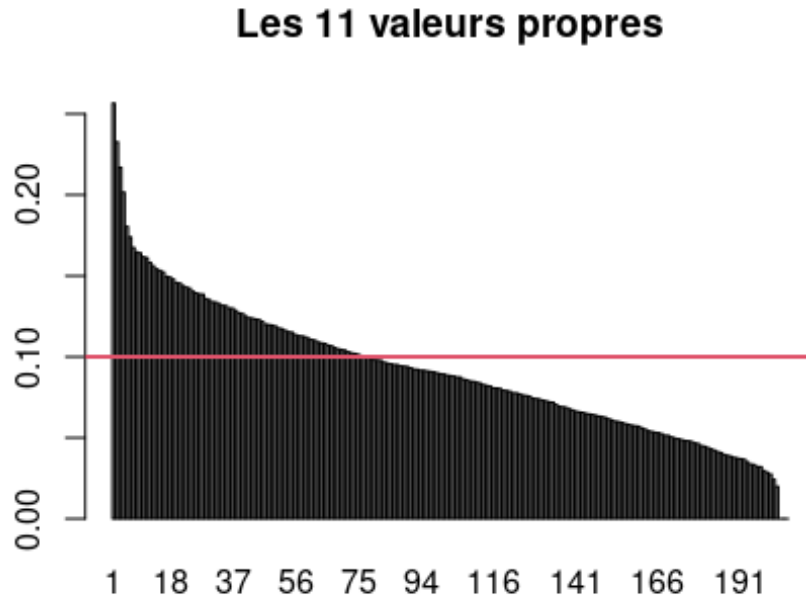
```

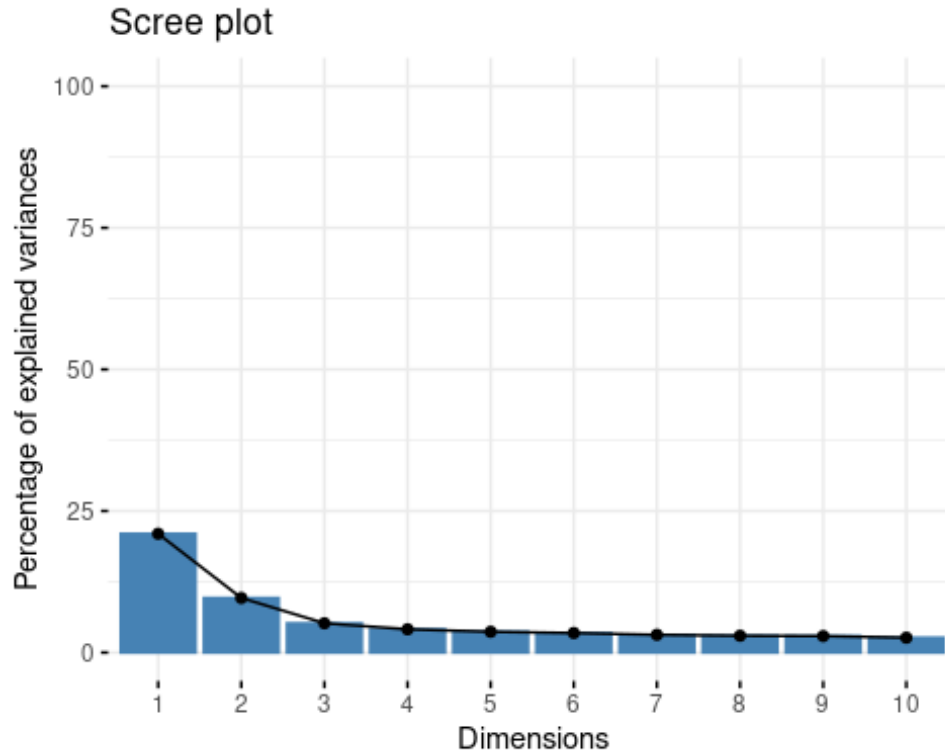
## 8      | -0.613  0.146  0.004 | -0.385  0.064  0.002 | -0.213  0.021  0.000
|
## 9      |  0.117  0.005  0.002 | -0.305  0.040  0.013 | -0.172  0.014  0.004
|
## 10     | -0.344  0.046  0.015 | -0.211  0.019  0.006 | -0.008  0.000  0.000
|
##
## Categories (the 10 first)
##          Dim.1    ctr    cos2 v.test    Dim.2    ctr    cos2 v.test    Dim.3
## 15      | -1.141  0.185  0.005 -2.286 |  0.719  0.081  0.002  1.441 |  1.463
## 16      | -0.369  0.005  0.000 -0.369 | -0.626  0.015  0.000 -0.626 | -0.292
## 17      | -1.066  0.121  0.003 -1.849 |  1.155  0.156  0.004  2.002 |  1.180
## 18      | -1.098  0.256  0.007 -2.696 |  0.649  0.099  0.003  1.594 |  1.521
## 19      | -0.625  0.042  0.001 -1.083 |  1.275  0.190  0.005  2.210 |  1.452
## 20      | -0.900  0.775  0.022 -4.738 | -0.183  0.035  0.001 -0.961 |  0.066
## 21      | -1.055  1.340  0.039 -6.255 | -0.129  0.022  0.001 -0.767 |  0.327
## 22      | -0.929  0.765  0.022 -4.704 |  0.177  0.031  0.001  0.896 |  0.351
## 23      | -0.814  0.352  0.010 -3.173 | -0.255  0.038  0.001 -0.993 |  0.190
## 24      | -0.738  0.502  0.015 -3.811 |  0.218  0.048  0.001  1.126 | -0.097
##          ctr    cos2 v.test
## 15      | 0.359  0.009  2.931 |
## 16      | 0.004  0.000 -0.292 |
## 17      | 0.175  0.004  2.046 |
## 18      | 0.582  0.014  3.736 |
## 19      | 0.265  0.006  2.517 |
## 20      | 0.005  0.000  0.349 |
## 21      | 0.153  0.004  1.941 |
## 22      | 0.129  0.003  1.777 |
## 23      | 0.023  0.001  0.740 |
## 24      | 0.010  0.000 -0.499 |
##
## Categorical variables (eta2)
##          Dim.1 Dim.2 Dim.3
## fage      | 0.644 0.124 0.171 |
## mage      | 0.812 0.153 0.107 |
## gained    | 0.214 0.446 0.657 |
## weeks     | 0.097 0.813 0.790 |
## mature    | 0.662 0.044 0.010 |
## habit     | 0.026 0.059 0.002 |
## marital   | 0.267 0.019 0.034 |
## whitemom  | 0.000 0.012 0.029 |
## premie    | 0.004 0.571 0.171 |
## sex       | 0.003 0.003 0.000 |

```

2. Choix du nombre d'axes - Eboulis des valeurs propres.

##	eigenvalue	variance.percent	cumulative.variance.percent
## Dim.1	2.565827e-01	1.376785e+00	1.376785
## Dim.2	2.327018e-01	1.248644e+00	2.625429
## Dim.3	2.169600e-01	1.164175e+00	3.789604
## Dim.4	2.015837e-01	1.081669e+00	4.871273
## Dim.5	1.804431e-01	9.682315e-01	5.839504
## Dim.6	1.739649e-01	9.334702e-01	6.772974
## Dim.7	1.674402e-01	8.984597e-01	7.671434
## Dim.8	1.646701e-01	8.835957e-01	8.555030
## Dim.9	1.641101e-01	8.805905e-01	9.435620
## Dim.10	1.617206e-01	8.677690e-01	10.303389





L'ACP a été réalisée sur deux ensembles de données : les individus (10 premiers) et les catégories (10 premières). Les résultats de l'ACP sont présentés sous forme de tableau, avec les dimensions (Dim), les contributions (ctr) et les cosinus carrés (cos2) pour chaque variable dans chaque dimension.

Les individus sont représentés dans trois dimensions (Dim.1, Dim.2 et Dim.3) avec des contributions différentes pour chaque dimension. Les variables qui ont des contributions élevées pour chaque dimension sont les variables les plus importantes pour cette dimension. Les cosinus carrés indiquent la proportion de la variance totale qui est expliquée par chaque dimension pour chaque variable.

Les catégories sont également représentées dans trois dimensions (Dim.1, Dim.2 et Dim.3) avec des contributions différentes pour chaque dimension. Les résultats comprennent également des tests de signification pour chaque dimension (v.test), qui mesurent la significativité statistique de la dimension. Les variables qui ont des contributions élevées pour chaque dimension sont les variables les plus importantes pour cette dimension. Les cosinus carrés indiquent la proportion de la variance totale qui est expliquée par chaque dimension pour chaque variable.

Les résultats présentent également des mesures d'association (eta2) entre des variables catégorielles. Chaque variable est représentée par trois dimensions (Dim.1, Dim.2, Dim.3), avec un coefficient eta2 pour chaque dimension. Le coefficient eta2 est une mesure de l'effet de chaque variable sur la variation de la réponse étudiée.

Les résultats indiquent que la variable "fage" a un effet important sur la réponse étudiée, avec un coefficient eta2 de 0,671 pour la première dimension. De même, les variables "mage", "mature" et "marital" ont également un effet important avec des coefficients eta2 respectifs de 0,854, 0,710 et 0,288.

En revanche, les variables “whitemom” et “sex” ont des coefficients η^2 très faibles, indiquant qu’elles ont un impact limité sur la variation de la réponse étudiée.

3. Interprétation de Chaque dimension de l’ACP

Pour la dimension 1, les variables les plus importantes sont l’âge de la mère (mage), le statut de maturité de la mère (mature) et l’âge du père (fage), avec des valeurs de R^2 très élevées (supérieures à 0,6) et des p-values très faibles (inférieures à 0,05). Cela indique que ces variables ont une forte influence sur la variation de la première dimension de l’ACP, qui semble être principalement liée à la maturité et l’âge des parents.

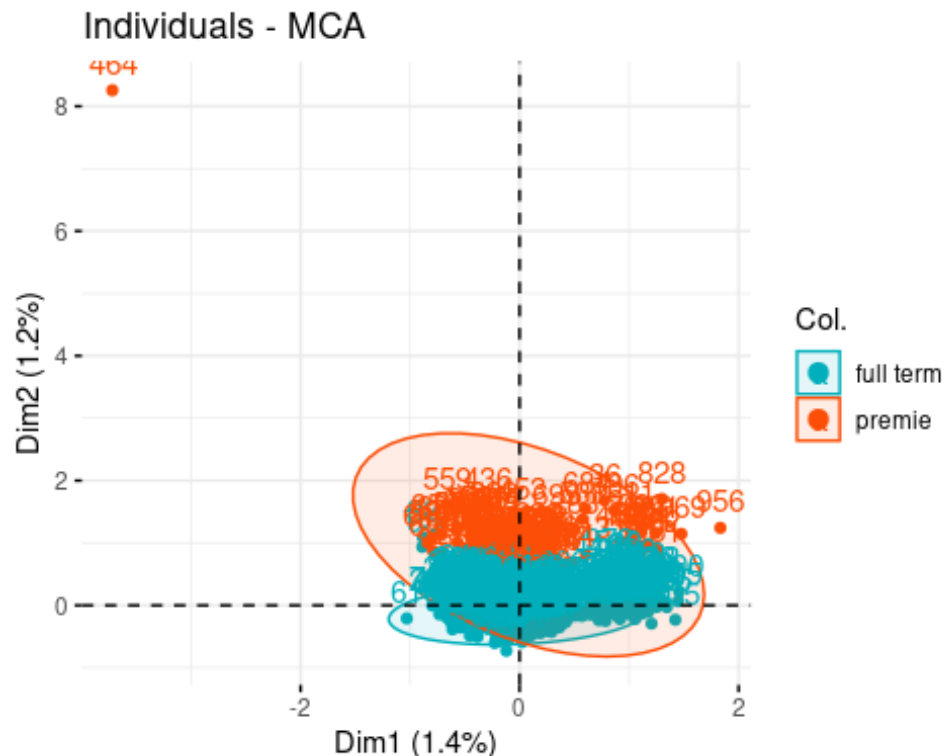
La dimension 2 est principalement influencée par la durée de la grossesse en semaines (weeks) et le fait que la naissance ait été classée comme prématurée (premie), avec des valeurs de R^2 très élevées et des p-values très faibles. Cela indique que la deuxième dimension de l’ACP est principalement liée au temps de gestation et au fait que la naissance soit prématurée ou non.

Les autres variables ont des contributions relativement faibles pour ces deux dimensions de l’ACP.

Tableaux de contingence des variables.

```
## , , sex = female
##
##          lowbirthweight
## premie      low not low
## full term   16      425
## premie      27       27
##
## , , sex = male
##
##          lowbirthweight
## premie      low not low
## full term   12      423
## premie      26       44
##
## , , sex = female
##
##          mature
## premie      mature mom younger mom
## full term           76           365
## premie              8           46
##
## , , sex = male
##
##          mature
## premie      mature mom younger mom
## full term           58           377
## premie              17           53
```

IX. Représentation graphique



L'ACM permet d'analyser les relations entre les variables catégorielles et de visualiser les relations entre les individus dans un espace de faible dimension. Dans notre cas, nous avons utilisé l'ACM pour évaluer l'influence de plusieurs variables sur la classification des naissances en terme ou prématuré.

La visualisation produite par le code R montre les individus dans le plan factoriel (composé des deux premières dimensions principales), colorés en fonction de leur classe (terme ou prématuré) et avec des ellipses représentant la dispersion des individus de chaque classe.

L'analyse des résultats de l'ACM peut être effectuée de plusieurs manières, par exemple :

L'observation des individus : Nous pouvons observer la répartition des individus dans le plan factoriel et identifier des groupes d'individus similaires. Nous pouvons remarquer que les individus prématurés (en rouge) semblent être plus dispersés que les individus à terme (en bleu) et qu'il n'y a pas de groupes clairement distingués. Cela pourrait indiquer que la classification à terme ou prématuré dépend de plusieurs facteurs et qu'il n'y a pas un seul groupe de facteurs qui explique la variation.

L'observation des variables : Nous pouvons observer la contribution des variables à la formation des axes principaux. Dans notre cas, nous pouvons remarquer que les variables "weeks" et "mature" sont les variables qui contribuent le plus à la formation du premier axe principal, tandis que les variables "habit", "marital" et "whitemom" contribuent davantage au deuxième axe principal. Cela pourrait indiquer que la durée de la grossesse et le statut de

maturité de la mère sont des facteurs importants dans la classification des naissances, tandis que le fait que la mère soit non-fumeuse, mariée ou blanche a une influence moindre.

L'observation des relations entre variables : Nous pouvons observer les relations entre les variables en regardant les corrélations entre les variables et leur position dans le plan factoriel. Par exemple, nous pouvons remarquer que la variable "gained" (poids pris par la mère pendant la grossesse) est positivement corrélée avec le premier axe principal, ce qui signifie qu'un gain de poids plus important pourrait être associé à une durée de grossesse plus longue et à un risque moindre de naissance prématurée.

En résumé, l'ACM nous permet de visualiser les relations complexes entre plusieurs variables catégorielles et d'identifier les facteurs qui influencent la classification des naissances à terme ou prématuré.

CONCLUSION

Au terme de notre analyse, il existe de nombreux facteurs qui influencent le poids d'un nouveau-né et les risques de naissance prématurée d'un bébé. Certains de ces facteurs sont liés à la santé de la mère, tels que son âge, son indice de masse corporelle et ses habitudes, tandis que d'autres facteurs sont liés à la grossesse elle-même, tels que les complications de la grossesse et la santé du fœtus. Il est important de souligner que, bien que certains facteurs puissent augmenter les risques de faible poids à la naissance ou de naissance prématurée, il est possible de prendre des mesures pour réduire ces risques. Cela peut inclure des soins prénatals réguliers, une alimentation saine et équilibrée, l'abstention de substances nocives telles que l'alcool et le tabac, et la gestion efficace des conditions médicales préexistantes. En fin de compte, il est essentiel que les femmes enceintes reçoivent les soins appropriés pour assurer la santé et le bien-être de leur bébé à naître. En travaillant ensemble avec les professionnels de la santé, les familles peuvent prendre les mesures nécessaires pour maximiser les chances d'une grossesse saine et d'une naissance réussie.

BIBLIOGRAPHIE

Cours L3 : ANALYSE EXPLORATOIRE DES DONNEES, Sabine LOUDCHER, 2023

Table des matières

SOMMAIRE.....	1
INTRODUCTION	2
I. CONTEXTE	3
II. DESCRIPTION	3
III. SOURCE DU DATASET.....	3
IV. PRESENTATION DES DONNNEES DU FICHIER D'ETUDE	3
V. Choix de la méthode	4
VI. Analyse des facteurs qui influence le poids d'un nouveau née	5
1. Analyse de la corrélation.....	5
2. L'utilisation de l'ACP et du cercle de corrélation	6
VII. Affichage du cercle de corrélation.....	7
Interprétation.....	8
VIII. Analyse des facteurs liée à la naissance prématuré d'un bébé	8
1. Représentation des 10 individus et les 10 catégories.....	8
2. Choix du nombre d'axes - Eboulis des valeurs propres.....	10
3. Interprétation de Chaque dimension de l'ACP.....	12
IX. Représentation graphique.....	13
CONCLUSION	15
BIBLIOGRAPHIE	16