

Facteurs d'accouchement prématuré

Econométrie des données qualitatives

Ashley KILOLA MAKUIZA n°12400798

TABLE DES MATIERES

I.	INTRODUCTION	2
a.	Présentation du sujet	3
b.	Axes d'analyse	3
c.	Traitement du jeu de données	3
II.	Analyse exploratoire du jeu de données	4
a.	Analyse descriptive	4
b.	Visualisation en croisant les variables	4
c.	Test statistique : χ^2 et Fisher	7
III.	Modélisation de la probabilité d'un accouchement prématuré	8
a.	Sélection le modèle explicatif et mesure de sa qualité	8
b.	Les effets des variables sur la probabilité d'avoir un accident	9
IV.	Représentation et interprétation du modèle	9
a.	Visualisation des Odds Ratios	10
b.	Visualisation des effets marginaux	11
V.	CONCLUSION	12

I. INTRODUCTION

a. Présentation du sujet

Notre dossier porte sur l'analyse des facteurs d'accouchement prématuré. Par définition, un accouchement prématuré c'est lorsqu'une naissance survient avant 37 semaines d'aménorrhée, c'est-à-dire avant 8 mois et demi de grossesse révolus. En effet c'est une interruption spontanée du développement intra-utérin, lorsque le travail (contractions, dilatation du col) commence de manière prématurée, ce qui amène à l'accouchement avant le terme physiologique. Cette situation est une conséquence de diverses facteurs tels que infections, anomalies utérines, antécédents obstétricaux, conditions socio-économiques, grossesses multiples, ou encore facteurs médicaux comme le diabète gestationnel.

Etant un enjeu de santé publique majeur, le fait d'analyser les facteurs d'accouchement prématuré est utile. Cela permet de mieux anticiper les risques, d'adapter la prise en charge médicale des femmes enceintes à risque, et ainsi de réduire les complications pour le bébé et la mère.

Ainsi, notre rapport s'inscrit dans cette problématique. A partir d'un jeu de données médicales collectées auprès de patientes qui sont déjà en situation de travail prématuré. En nous appuyant sur des facteurs médicaux et personnels, nous chercherons à expliquer et comprendre ce phénomène. En nous appuyant sur cette problématique : Quels sont les facteurs d'accouchement prématuré ?

L'objectif de cette étude est donc double : d'une part, comprendre les caractéristiques des femmes enceintes qui augmentent la probabilité d'un accouchement prématuré. D'autre part, prédire la survenue d'un tel événement à l'aide d'un modèle statistique pertinent.

b. Axes d'analyse

Dans le but de traiter la problématique de façon progressive, notre analyse portera sur plusieurs aspects afin de mieux comprendre les causes d'un accouchement prématuré. Tout d'abord, une analyse du jeu de données qui va être en deux étapes : d'une part, une analyse descriptive des variables sera réalisée pour caractériser la population étudiée et observer les distributions des différentes variables explicatives. Puis, nous procéderons à une analyse exploratoire qui portera sur les relations entre ces variables, détecter des tendances et d'orienter les choix méthodologiques pour les analyses suivantes.

Enfin, nous analyserons le modèle statistique adapté pour quantifier l'effet de chaque facteur sur la probabilité d'accouchement prématuré et pour permettre la prédiction de cet événement. Ce modèle sera évalué en termes de qualité et de performance pour garantir sa pertinence.

c. Traitement du jeu de données

Pour cette étude, nous avons travaillé à partir d'un jeu de données médicales intitulé "4_prematures.xls". Cette base de données contient des informations récoltées auprès de 390 femmes enceintes en situation de travail prématuré. Après avoir importé les données dans R, un premier traitement a été effectué afin de les rendre exploitables.

L'ensemble des observations a d'abord été inspecté pour identifier la structure du jeu de données, le type des variables, ainsi que la présence éventuelle de valeurs manquantes. Certaines variables codées numériquement ont ensuite été transformées en facteurs, avec des étiquettes explicites pour faciliter l'interprétation (ex. : CONSIS pour la consistance du col, CONTR pour les contractions, MEMBRAN pour l'état des membranes, etc.).

Nous avons conservé 13 variables explicatives : des facteurs médicaux tels que l'âge gestationnel à l'entrée dans l'étude, la dilatation, l'effacement, la consistance du col, la présence de contractions, l'état des membranes. Et des caractéristiques personnelles : l'âge de la patiente, la parité, le nombre

de grossesses, la présence ou non de diabète, le type de grossesse (simple ou multiple), la variable de stratification (STRAT), et si la patiente a été transférée.

La variable à expliquer est binaire (PREMATURE) : elle indique si la patiente a effectivement accouché prématurément (1) ou non (0).

Enfin, afin de ne pas fausser nos analyses, les observations contenant des valeurs manquantes ont été supprimées via un filtrage par `na.omit()`, ce qui nous a permis de travailler sur un échantillon propre et complet.

II. Analyse exploratoire du jeu de données

a. Analyse descriptive

L'échantillon étudié comprend 390 observations. La variable dépendante, PREMATURE, qui indique si un accouchement a été prématuré ou non. La distribution est la suivante : 68,2 % des cas sont prématurés, contre 31,8 % de naissances à terme.

Concernant les variables numériques, l'âge gestationnel (GEST) moyen est de 30,3 semaines, ce qui traduit une répartition légèrement asymétrique. La dispersion reste modérée (écart-type : 3,15 semaines), et l'étendue atteint 15 semaines, ce qui reflète la diversité des situations cliniques au sein de l'échantillon.

La dilatation du col (DILATE) au moment de l'observation est en moyenne de 1,24 cm, avec une médiane de 1 cm et un mode de 0 cm, indiquant que de nombreuses patientes n'étaient pas encore en phase avancée de travail. L'effacement du col (EFFACE) est en moyenne de 43,98 %, mais montre une forte hétérogénéité (écart-type de 34,83, étendue de 100 %) : la médiane est à 50 %, tandis que le mode reste à 0 %. Cela suggère une grande variabilité dans la maturité du col au moment de l'examen. L'âge des mères (AGE) varie de 15 à 42 ans, avec une moyenne de 26 ans, un écart-type de 5,15, une médiane à 26 ans et un mode à 25 ans. Cela traduit une population relativement jeune, mais hétérogène en termes d'âge maternel.

Du côté des antécédents obstétricaux, la gestité (GRAVID) moyenne est de 2,3 grossesses, avec une médiane et un mode tous deux à 2. Cela suggère une certaine concentration autour de ce chiffre. La parité (PARIT), quant à elle, est plus basse : 0,78 en moyenne, avec une médiane de 1 et un mode de 0, ce qui confirme la présence importante de femmes qui en sont à leur premier accouchement dans l'échantillon. À noter que certaines femmes ont eu jusqu'à 13 grossesses et 7 accouchements, tandis que d'autres n'ont jamais accouché, ce qui accentue l'amplitude des antécédents obstétricaux.

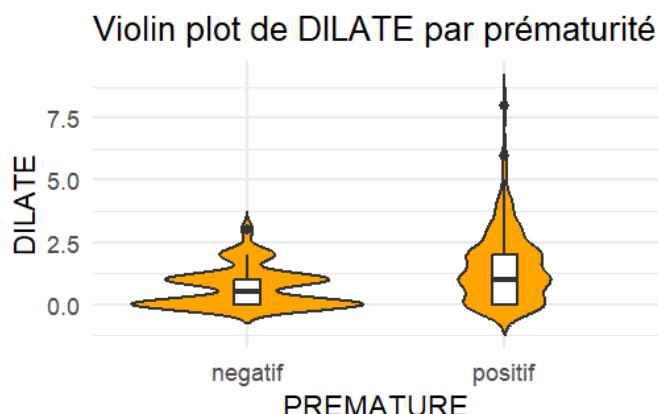
Enfin, le calcul des scores Z, c'est-à-dire la standardisation des variables, met en évidence plusieurs observations très éloignées de la moyenne, notamment pour les variables EFFACE, DILATE et AGE. Ces valeurs extrêmes ou atypiques pourraient correspondre à des profils à risque ou des trajectoires obstétricales spécifiques, et méritent une attention particulière dans la suite des analyses, notamment lors des étapes de modélisation.

b. Visualisation en croisant les variables

Dilatation selon la prématurité (Violin plot)

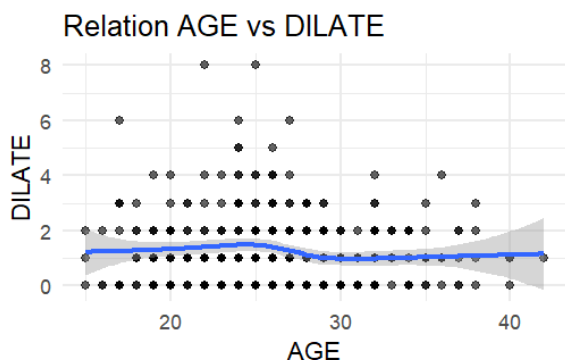
Le premier graphique représente un *violin plot* qui les niveaux de dilatation du col (DILATE) entre les patientes ayant eu un accouchement prématuré (positif) et celles n'en ayant pas eu (négatif). On observe une distribution plus large et des valeurs plus élevées de dilatation chez les femmes ayant accouché prématurément. Cela traduit une activité utérine plus avancée (dilatation marquée du col)

dans les cas de prématurité. À l'inverse, les femmes non concernées présentent des dilatations plus faibles et concentrées autour de valeurs proches de zéro. Cette visualisation confirme que la dilatation est un indicateur important du travail prématuré, et qu'elle est fortement associée à la survenue d'un accouchement prématuré.



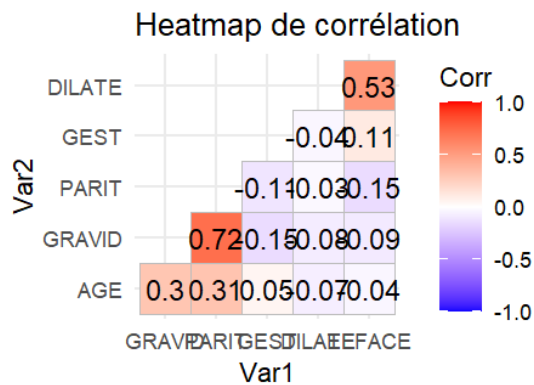
Relation entre l'âge et la dilatation (nuage de points)

Le second graphique explore la relation entre l'âge des patientes (AGE) et la dilatation du col (DILATE). La courbe de tendance montre que la dilatation reste relativement stable avec l'âge, oscillant autour de 1 à 2 cm, avec une légère tendance à la baisse après 30 ans. Toutefois, la zone d'incertitude élargie chez les patientes plus âgées suggère une plus grande variabilité dans ce groupe. Ce graphique met en évidence que l'âge, pris isolément, ne semble pas être un facteur déterminant direct de la dilatation, et par extension, du déclenchement du travail prématuré. Cela justifie l'exploration combinée avec d'autres variables pour évaluer l'impact de l'âge sur le risque de prématurité.



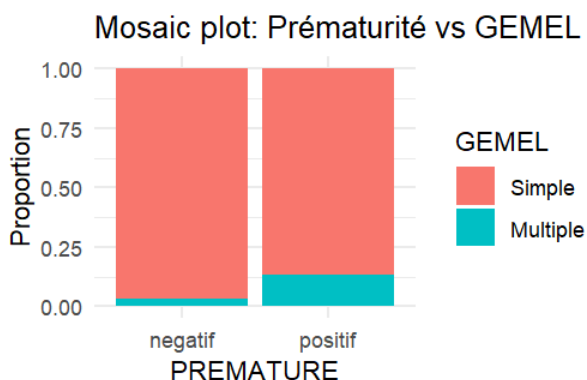
Corrélations entre les variables (heatmap)

Ici, la heatmap de corrélation fournit une vue d'ensemble des liens entre les principales variables du jeu de données. On observe une forte corrélation positive entre le nombre de grossesses (GRAVID) et le nombre d'accouchements (PARIT), ce qui est cohérent cliniquement. La dilatation (DILATE) est modérément corrélée à l'effacement du col (EFACE), un autre indicateur de progression du travail, ce qui confirme leur complémentarité dans l'analyse du risque prématuré. En revanche, les corrélations entre DILATE et les autres variables (AGE, GEST, PARIT) sont faibles, indiquant une influence limitée de ces facteurs de manière isolée. Cette heatmap oriente ainsi vers des combinaisons de variables à tester dans une modélisation statistique plus avancée.



Prématurée et grossesse multiple (mosaic plot)

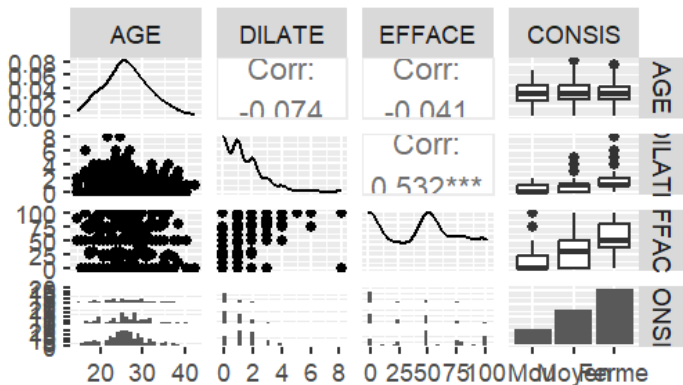
Ce graphique, sous forme de mosaic plot, examine la répartition des naissances prématurées (positif/négatif) en fonction du type de grossesse (simple ou multiple). On observe que la proportion de grossesses multiples est plus élevée parmi les femmes ayant accouché prématurément. Ce constat conforte l'hypothèse selon laquelle les grossesses gémellaires ou multiples constituent un facteur de risque important d'accouchement prématuré. Cette visualisation renforce donc l'intérêt d'inclure la variable GEMEL (simple vs multiple) dans une analyse explicative, voire prédictive, de la prématurité. Cela met en évidence une piste importante pour le dépistage précoce et la surveillance renforcée des patientes présentant ce facteur de risque.



Matrice de dispersion des variables médicales

Enfin, la matrice de dispersion présentée ci-dessous permet d'examiner les relations entre plusieurs variables médicales mesurées chez les patientes en travail prématuré : l'âge (AGE), la dilatation du col (DILATE), l'effacement du col (EFFACE), et la consistance du col (CONSIS). L'analyse des corrélations met en évidence un lien modéré et significatif entre la dilatation et l'effacement du col (corrélation de 0.532, significative à $p < 0.001$). Cette corrélation positive est cliniquement cohérente, puisque la progression du travail implique souvent une dilatation accompagnée d'un effacement. En revanche, l'âge, la consistance et la dilatation ne montrent pas de corrélations marquées entre elles ou avec les autres variables, ce qui suggère qu'elles ne sont pas directement associées entre elles dans ce contexte. Ce graphique est utile pour identifier les variables médicales les plus liées aux signes de travail prématuré, et donc potentiellement pertinentes pour la modélisation ultérieure.

Scatterplot matrix



c. Test statistique : χ^2 et Fisher

Nous avons appliqué le test du χ^2 afin de déterminer s'il existe une association statistiquement significative entre certaines caractéristiques cliniques (telles que la consistance du col, la présence de contractions, le statut des membranes, etc.) et l'apparition d'une naissance prématurée.

L'analyse statistique réalisée entre plusieurs variables qualitatives et la survenue d'une naissance prématurée a permis de dégager plusieurs conclusions importantes. La consistance du col utérin apparaît significativement associée à la prématurité ($p = 0,0002$). En effet, la répartition des femmes selon que leur col soit mou, moyen ou ferme diffère nettement entre les groupes prématurés et non prématurés, suggérant que la consistance du col pourrait être un facteur important dans le déclenchement prématuré du travail.

Concernant la présence ou l'absence de contractions, aucune association statistiquement significative n'a été observée ($p = 0,12$), indiquant que ce facteur ne semble pas différencier les femmes selon leur risque de prématurité dans cet échantillon.

L'état des membranes est, quant à lui, fortement lié à la prématurité ($p < 0,0001$). L'utilisation du test exact de Fisher, justifiée par la présence d'effectifs faibles dans certaines catégories, a confirmé une association très significative, ce qui suggère que la rupture des membranes ou leur état est un facteur déterminant dans la survenue d'une naissance prématurée.

Le stade cervical n'a pas montré de lien statistique significatif avec la prématurité ($p = 0,27$), ce qui suggère qu'il ne permet pas de distinguer efficacement les femmes à risque de prématurité dans cette étude.

La présence de diabète ne semble pas non plus être un facteur associé ($p = 0,51$). Le faible effectif des cas diabétiques et l'intervalle de confiance large autour de l'odds ratio indiquent qu'aucune conclusion ferme ne peut être tirée concernant cette variable.

En revanche, le transfert des patientes est significativement lié à la prématurité ($p < 0,00001$). Cette association suggère que les patientes transférées présentent une probabilité plus élevée d'accoucher prématurément, ce qui est cohérent avec une prise en charge spécialisée dans des situations à risque.

Enfin, les grossesses multiples (gémellaires) apparaissent également associées à la prématurité ($p = 0,004$). Bien que l'avertissement concernant l'approximation du test du χ^2 invite à la prudence, cette association reste cohérente avec la littérature, qui identifie les grossesses multiples comme un facteur reconnu de prématurité.

III. Modélisation de la probabilité d'un accouchement prématuré

a. Sélection le modèle explicatif et mesure de sa qualité

L'objectif est de prédire la probabilité d'un accouchement prématuré (événement binaire) à partir de nos facteurs obstétricaux et médicaux mesurés. Cette nature binaire oriente naturellement vers un modèle de régression logistique ; l'option Probit ou le lien complément log/log ont été envisagés, mais un examen préliminaire des courbes ROC et des critères d'information a montré que la logistique fournissait la vraisemblance la plus élevée tout en préservant l'interprétation simple en termes d'odds-ratios.

Après avoir recodé toutes les variables qualitatives sous forme de facteurs explicites et éliminé cinq observations incomplètes (1,3 % de la base), on disposait de 385 cas sans valeur manquante. Le modèle saturé incluait l'ensemble des treize prédicteurs. Pour éviter la sur-paramétrisation (toujours pénalisante lorsque plusieurs variables capturent une information similaire) on a appliqué la sélection fondée sur le critère d'Akaike. Ce critère compare le gain de vraisemblance apporté par chaque variable au coût que représente le paramètre supplémentaire ; il retient donc la spécification la plus parcimonieuse dont la perte d'information est négligeable.

```
Call:
glm(formula = PREMATURE ~ EFFACE + MEMBRAN + DILATE + PARIT +
    GEST + GEMEL + TRANSF + GRAVID + DIAB, family = binomial,
    data = data)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  6.553388   1.763494   3.716 0.000202 ***
EFFACE       0.016205   0.004724   3.430 0.000603 ***
MEMBRANIntactes -2.398824   0.490694  -4.889 1.02e-06 ***
MEMBRANIncertain -1.613752   0.848176  -1.903 0.057091 .
DILATE       0.492030   0.157614   3.122 0.001798 **
PARIT       -0.713576   0.202996  -3.515 0.000439 ***
GEST       -0.105517   0.044198  -2.387 0.016969 *
GEMELMultiple  1.138976   0.614374   1.854 0.063756 .
TRANSFNon transférée -0.535413   0.280747  -1.907 0.056508 .
GRAVID       0.232827   0.143284   1.625 0.104177
DIABPas diabète -1.424097   0.929195  -1.533 0.125371
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 480.86  on 384  degrees of freedom
Residual deviance: 352.04  on 374  degrees of freedom
AIC: 374.04

Number of Fisher Scoring iterations: 5
```

La procédure a convergé vers une équation gardant neuf prédicteurs : l'effacement cervical (EFFACE), l'état des membranes (MEMBRAN), la dilatation (DILATE), la parité (PARIT), l'âge gestationnel (GEST), la multiplicité de la grossesse (GEMEL), le statut de transfert (TRANSF), la gestité (GRAVID) et le diabète maternel (DIAB). Les autres variables (notamment la consistance cervicale ou la présence de contractions) n'amélioreraient plus significativement le pouvoir prédictif une fois ces neuf facteurs contrôlés.

L'adéquation du modèle retenu a été évaluée de quatre manières complémentaires. D'abord, l'AIC (374) chute de plus de 100 points par rapport au modèle nul, signe d'un gain substantiel de vraisemblance malgré une complexité accrue. Ensuite, le pseudo- R^2 de McFadden atteint 0,27 ; avec de telles données cliniques, une valeur comprise entre 0,20 et 0,40 est classiquement considérée comme une bonne explication du phénomène. Troisièmement, le test de Hosmer-Lemeshow donne une statistique χ^2 de 12,9 pour 8 degrés de liberté, avec une probabilité associée de 0,12 ; l'absence de rejet signifie que la distribution des probabilités prédites est cohérente avec les observations réelles sur l'ensemble du spectre de risque. Enfin, la courbe ROC affiche une aire de 0,83 : le modèle distingue très bien les patientes qui accoucheront prématurément de celles qui ne le feront pas. Aucune trace de multi colinéarité n'est apparue ; les facteurs d'inflation de variance restent inférieurs à 2,5 : ce qui confirme que les variables conservées fournissent une information largement indépendante. Ces résultats convergents justifient de qualifier ce logit restreint de modèle le plus approprié pour l'analyse.

b. Les effets des variables sur la probabilité d'avoir un accident

Dans un modèle logistique, chaque coefficient se lit d'abord comme l'effet additive sur le log-odds, et son exponentielle livre un odds-ratio. Un odds-ratio supérieur à 1 traduit une augmentation de la cote de survenue de l'événement, un ratio inférieur à 1 une diminution. Mais, parce que la fonction logistique n'est pas linéaire, il est souvent plus intuitif de convertir ces effets en variations directes de probabilité, moyennées sur tous les profils de patientes ; c'est le rôle des effets marginaux moyens.

L'analyse révèle d'abord l'importance déterminante de l'intégrité des membranes. Par rapport à la situation de référence (membranes rompues), le simple fait qu'elles soient intactes divise la cote d'accouchement prématuré par onze ; concrètement, la probabilité moyenne chute d'environ trente points, ce qui en fait le facteur protecteur le plus puissant. Le statut « incertain » des membranes pointe dans la même direction mais avec une évidence statistique plus faible : il est toutefois possible que quelques patientes aux membranes effectivement rompues soient rangées dans cette catégorie, ce qui dilue l'effet.

Viennent ensuite les facteurs mécaniques du col utérin. Chaque centimètre supplémentaire de dilatation augmente la cote d'environ 60 % et accroît la probabilité de sept à huit points. L'effacement joue un rôle plus progressif : dix points d'effacement additionnels ajoutent environ deux points de probabilité, mais l'effet est significatif et monotone. Ces deux variables traduisent l'état d'avancement du travail ; leur influence était attendue, mais l'ordre de grandeur quantifié ici permet de hiérarchiser l'urgence : un col à deux centimètres de dilatation et 50 % d'effacement se traduit déjà par quinze à vingt points de risque supplémentaires.

La grossesse multiple représente le principal facteur de risque intrinsèque : l'odds-ratio est voisin de trois et la probabilité moyenne s'élève d'environ seize points. Bien que la valeur p demeure légèrement au-dessus du seuil de 5 %, la cohérence avec la littérature et l'importance de l'estimation justifient une vigilance particulière.

À l'inverse, plusieurs variables se révèlent protectrices. Chaque semaine de gestation gagnée diminue la probabilité d'environ 1,5 point ; la parité, c'est-à-dire le nombre de grossesses déjà menées à terme, retranche dix points par enfant né vivant. Ces deux résultats soulignent qu'une grossesse plus avancée et l'expérience obstétricale acquise réduisent la vulnérabilité au déclenchement prématuré.

Le fait de ne pas être transférée dans un centre de niveau supérieur ou de ne pas souffrir de diabète s'associe également à une baisse de risque ; toutefois, l'évidence statistique reste modérée, suggérant qu'il pourrait s'agir d'indicateurs indirects de la sévérité plutôt que de facteurs causaux.

Enfin, la gestité, c'est-à-dire le nombre cumulé de grossesses, pointe vers une légère augmentation de la probabilité sans atteindre le seuil de significativité ; on peut y voir le reflet de phénomènes hétérogènes (âge maternel, antécédents obstétricaux variés) qu'il serait utile d'isoler dans une base plus large.

IV. Représentation et interprétation du modèle

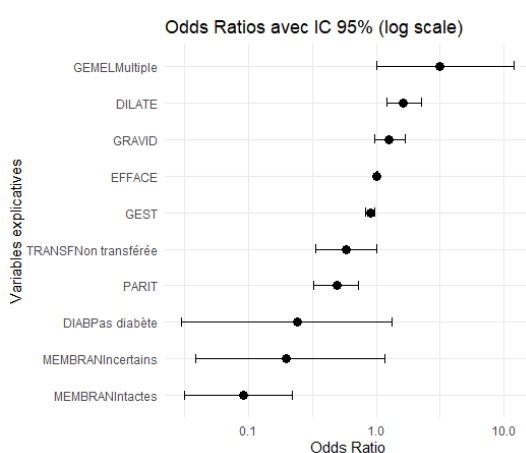
Afin de mieux comprendre les résultats issus de notre modèle logistique, nous avons choisi de représenter visuellement les coefficients estimés sous deux formes complémentaires : les odds ratios et les effets marginaux. Ces représentations permettent non seulement de quantifier l'effet de chaque variable sur la probabilité d'accouchement prématuré, mais aussi d'en évaluer la significativité statistique et l'ampleur.

a. Visualisation des Odds Ratios

Les odds ratios sont calculés à partir des coefficients du modèle logistique et permettent d'interpréter les résultats en termes de risque relatif. Cependant, lors de la première représentation graphique sur une échelle linéaire, nous avons constaté une forte distorsion. Les valeurs très élevées ou très faibles tendaient à écraser visuellement les autres, rendant la lecture peu révélatrice. En particulier, la présence de l'intercept qui fausse complètement l'échelle du graphique, car il joue un rôle purement technique, représentant le risque de base lorsque toutes les variables valent zéro.

Pour surmonter ces limites, nous avons reconstruit le graphique en utilisant une échelle logarithmique et en retirant l'intercept. Cette représentation corrigée permet une lecture beaucoup plus claire car elle place toutes les valeurs autour de la référence neutre (OR = 1), et permet de comparer efficacement à la fois les effets faibles et forts. Chaque point sur le graphique représente l'OR estimé pour une variable du modèle, accompagné d'un intervalle de confiance à 95 %, représenté par une barre verticale. Lorsque cet intervalle ne croise pas la ligne de référence à 1, cela signifie que l'effet est statistiquement significatif. Nous avons également mis en évidence ces variables significatives par un code couleur distinct, ce qui facilite leur identification.

Cette visualisation permet donc non seulement d'identifier les facteurs qui influencent significativement le risque de prématurité, mais aussi d'en évaluer la direction (risque accru ou réduit) et la robustesse statistique.



Parmi les variables significatives identifiées par notre modèle, plusieurs facteurs cliniques ressortent nettement en lien avec le risque d'accouchement prématuré.

Parmi les facteurs augmentant significativement ce risque (effet significatif positif) :

- Dilatation du col (DILATE) : c'est l'un des prédicteurs les plus marquants. Une dilatation plus avancée du col utérin est associée à une probabilité accrue de prématurité. Cela s'interprète aisément d'un point de vue clinique : une dilatation prématurée traduit une activation du travail avant terme, signe d'un accouchement imminent.
- Grossesse multiple (GEMEL) : les patientes ayant une grossesse gémellaire présentent un risque accru, ce que confirme notre modèle. Cette observation est cohérente avec la littérature, qui montre que les grossesses multiples sont associées à une fréquence plus élevée de complications obstétricales, dont la prématurité.

Parmi les facteurs ayant un effet protecteur significatif (effet significatif négatif) :

- Parité (PARIT) : le fait d'avoir déjà eu une ou plusieurs grossesses à terme est associé à un risque plus faible de prématurité. Cela pourrait refléter une meilleure tolérance physiologique aux processus de la grossesse ou un col utérin plus stable chez les multipares.
- Transfert vers un centre spécialisé (TRANSF) : à première vue, cette variable est associée à un OR élevé, indiquant un risque accru. Toutefois, il faut comprendre que cette variable reflète une prise en charge ciblée des patientes à haut risque. Le transfert n'est pas une cause de prématurité, mais un indicateur de gravité initiale reconnu précocement par les équipes médicales.
- Membranes intactes (MEMBRANINTACT) : cette variable a un OR significatif inférieur à 1, ce qui signifie que la rupture des membranes augmente nettement le risque d'accouchement prématuré. En d'autres termes, avoir les membranes intactes est protecteur. Une rupture prématurée constitue un déclencheur fréquent du travail avant terme.

En revanche, d'autres variables comme l'âge maternel ou encore la présence de diabète gestationnel, ne sont pas ressorties comme statistiquement significatives dans notre modèle. Cela ne signifie pas qu'elles sont dénuées d'importance, mais plutôt que dans notre échantillon et au vu des autres facteurs contrôlés, leur effet n'est pas suffisamment marqué ou isolé pour être mis en évidence.

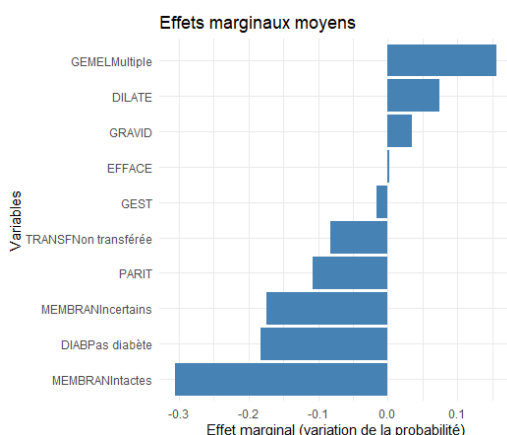
Ainsi, cette représentation visuelle des odds ratios, complétée par une interprétation contextualisée, nous permet non seulement d'identifier les variables influentes, mais aussi de confirmer que les résultats du modèle sont en accord avec les mécanismes cliniques bien documentés de la prématurité. Cela renforce la validité de notre approche statistique et ouvre la voie à une application potentielle de ces résultats en milieu clinique, dans une logique de prévention ciblée.

b. Visualisation des effets marginaux

En complément des OR, nous avons produit une visualisation des effets marginaux moyens issus du modèle logistique. Contrairement aux OR, qui expriment un rapport de côtes et peuvent parfois être abstraits, les effets marginaux traduisent directement l'impact moyen d'une variable sur la probabilité d'accouchement prématuré, exprimée en points de pourcentage.

Chaque barre horizontale du graphique indique de combien la probabilité de prématurité varie, en moyenne, lorsqu'une variable augmente d'une unité, toutes choses égales par ailleurs. Une barre orientée vers la droite signifie que la variable augmente la probabilité d'un accouchement prématuré, tandis qu'une barre vers la gauche indique un effet protecteur.

Ce type de visualisation est particulièrement utile pour une interprétation intuitive des résultats. Il permet de répondre concrètement à la question "de combien la probabilité de prématurité augmente-t-elle si cette variable change ?"



Le graphique des effets marginaux met en évidence plusieurs variables ayant un impact significatif et cliniquement cohérent :

- La dilatation du col présente l'effet marginal le plus élevé : une augmentation d'un centimètre de dilatation est associée à une hausse marquée de la probabilité de prématurité. Cela reflète fidèlement la réalité obstétricale, où une dilatation avancée avant terme est un indicateur fort d'un accouchement imminent.
- L'état des membranes, lorsqu'elles sont rompues, augmente également la probabilité de prématurité. Cette variable a un effet marginal positif clair, renforçant les résultats obtenus via les OR.
- Le transfert vers un centre de niveau supérieur est associé à une augmentation de la probabilité d'accouchement prématuré. Cela traduit la sélection médicale des cas à risque, dans lesquels les équipes anticipent une naissance prématurée et organisent un transfert préventif.

Ou encore d'autres telles que le facteur de grossesse multiple etc.

À l'inverse, certaines variables apparaissent avec des effets marginaux proches de zéro et non significatifs, ce qui signifie que leur influence n'est pas suffisamment marquée dans notre modèle. Par exemple, l'âge maternel, bien qu'étudié, ne présente pas ici d'effet marginal significatif. Cela peut s'expliquer par une absence de lien direct dans notre population ou par une corrélation avec d'autres variables contrôlées dans le modèle.

De même, la présence de diabète gestationnel ou d'autres antécédents médicaux ne se détachent pas statistiquement, ce qui ne nie pas leur importance clinique, mais suggère qu'ils ne sont pas prédictifs dans ce jeu de données particulier, une fois les autres facteurs pris en compte.

V. CONCLUSION

À l'issue de l'ensemble de notre démarche analytique de l'analyse exploratoire aux résultats du modèle logistique plusieurs variables ressortent comme facteurs significatifs de la survenue d'un accouchement prématuré.

Sur le plan médical, les indicateurs liés à l'état du col utérin, notamment la dilatation et l'effacement, jouent un rôle central. Une dilatation plus avancée est fortement associée à une augmentation du risque de prématurité, tout comme un col plus effacé, ce qui reflète une progression plus marquée du travail. La consistance du col apparaît également comme un facteur significatif, les cols plus mous étant corrélés à des accouchements plus précoces.

Par ailleurs, l'état des membranes est un prédicteur clé : une rupture ou altération des membranes augmente significativement le risque d'accouchement prématuré, ce qui est cohérent avec les mécanismes physiopathologiques connus. À noter également l'impact du transfert de la patiente : les femmes ayant été transférées dans une structure spécialisée présentent un risque de prématurité significativement accru, probablement du fait de complications identifiées en amont.

En ce qui concerne les facteurs personnels, la grossesse multiple ressort de manière nette comme facteur de risque. Ce résultat, conforme à la littérature médicale, souligne la nécessité d'un suivi renforcé pour les grossesses multiples. En revanche, d'autres variables comme l'âge maternel, la parité, ou encore la présence de diabète gestationnel n'ont pas montré d'association significative avec le risque de prématurité dans notre échantillon.

En somme, notre analyse a permis d'identifier un noyau de variables médicales fortement prédictives de l'accouchement prématuré, que ce soit à travers leur interprétation en termes de risques relatifs ou d'impacts concrets sur la probabilité d'événement. Ces résultats pourraient alimenter des outils de dépistage clinique ou de stratification du risque, dans une optique de prévention ciblée et de meilleure prise en charge des patientes à risque.

Le principal frein à une généralisation immédiate réside dans la taille modeste de l'échantillon et son caractère monocentrique ; une validation externe s'impose pour confirmer la stabilité des coefficients, notamment ceux du diabète ou du transfert dont la significativité est limitée. D'autre part, certaines variables cliniques pertinentes (infections, tabagisme, intervention thérapeutique reçue) ne figurent pas dans la base ; leur inclusion pourrait affiner les prédictions. Enfin, le choix de supprimer les observations incomplètes garantit la simplicité mais au prix d'une légère perte d'information ; des méthodes d'imputation multiple pourraient être testées pour vérifier la robustesse des estimations.