

# Naissances états-uniennes

## Contexte des données

Chaque année, les États-Unis donne accès à un large jeu de données avec des information sur les naissances au niveau fédéral. Ces données permettent de relier les habitudes et pratiques de personnes enceintes et la naissance de leur(s) enfant(s). L'enquête est très complète et comprend de nombreuses données associées aux accouchements. Elle permet une meilleure prise en compte des risques et une amélioration des pratiques médicales autour de l'accouchement. Ce jeu de données est un tirage aléatoire de 5000 naissances avec une extraction de 25 variables importantes sur les 126 mesurées pour les naissances ayant eu lieu en 2014.

## Descriptif des données

Ce jeu de données contient **5,000 lignes** et **25 colonnes** :

Nom de la colonne	Type de variable	Description
<code>birth_month</code>	Chaîne de caractères	Mois de naissance
<code>birth_day</code>	Chaîne de caractères	Jour de naissance
<code>birth_time</code>	Nombre entier	Heure de naissance (selon le format <code>XXYY</code> avec <code>XX</code> indiquant les heures après minuit et <code>YY</code> les minutes)
<code>birth_facility</code>	Chaîne de caractères	Lieu de l'accouchement : '(1) <code>Hospital</code> ' = hôpital, '(2) <code>Freestanding Birth Center</code> ' = maison de naissance, '(3) <code>Home (intended)</code> ' = à domicile (intentionnel), '(4) <code>Home (not intended)</code> ' = à domicile (non intentionnel), '(5) <code>Home (unknown if intended)</code> ' = à domicile (statut intentionnel inconnu) (), '(6) <code>Clinic / Doctor's Office</code> ' = clinique ou cabinet de médecin, '(7) <code>Other</code> ' = autre , <code>NA</code> = donnée manquante
<code>mother_age</code>	Chaîne de caractères	Âge de la mère, donné dans un intervalle en années
<code>mother_place_birth</code>	Chaîne de caractères	Lieu de naissance de la mère : "(1) <code>Born in the U.S. (50 US States)</code> " = aux États-Unis, "(2) <code>Born outside the U.S. (includes possessions)</code> " = en dehors des États-Unis, <code>NA</code> = donnée manquante
<code>mother_race</code>	Chaîne de caractères	Race sociale de la mère : '(1) <code>White</code> ' = blanche, '(2) <code>Black</code> ' = noire, '(3) <code>American Indian or Alaskan Native</code> ' = indienne américaine ou native d'Alaska, '(4) <code>Asian or Pacific Islander</code> ' = asiatique ou des îles du Pacifique
<code>marital_status</code>	Chaîne de caractères	Statut marital de la mère : "(1) <code>Married</code> " = mariée, "(2) <code>Unmarried</code> " = non mariée

Nom de la colonne	Type de variable	Description
mother_education	Chaîne de caractères	Plus haut niveau scolaire atteint par la mère : '(1) 8th grade or less' = collège ou moins, '(2) 9th through 12th grade with no diploma' = lycée mais sans baccalauréat, '(3) High school graduate or GED completed' = lycée avec baccalauréat, '(4) Some college credit, but not a degree.' = suivi de formation universitaire mais sans diplôme, '(5) Associate degree (AA,AS)' = équivalent BTS ou BUT, '(6) Bachelor's degree (BA, AB, BS)' = Licence universitaire, '(7) Master's degree (MA, MS, MEng, MEd, MSW, MBA)' = Master, '(8) Doctorate (PhD, EdD) or Professional Degree (MD, DDS, DVM, LLB, JD)' = Doctorat ou équivalent, NA = donnée manquante
father_age	Nombre entier	Âge du père en années (NA = donnée manquante)
father_race	Chaîne de caractères	Race sociale du père : '(1) White' = blanche, '(2) Black' = noire, '(3) American Indian or Alaskan Native' = indienne américaine ou native d'Alaska, '(4) Asian or Pacific Islander' = asiatique ou des îles du Pacifique, NA = donnée manquante
father_education	Chaîne de caractères	Plus haut niveau scolaire atteint par le père : '(1) 8th grade or less' = collège ou moins, '(2) 9th through 12th grade with no diploma' = lycée mais sans baccalauréat, '(3) High school graduate or GED completed' = lycée avec baccalauréat, '(4) Some college credit, but not a degree.' = suivi de formation universitaire mais sans diplôme, '(5) Associate degree (AA,AS)' = équivalent BTS ou BUT, '(6) Bachelor's degree (BA, AB, BS)' = Licence universitaire, '(7) Master's degree (MA, MS, MEng, MEd, MSW, MBA)' = Master, '(8) Doctorate (PhD, EdD) or Professional Degree (MD, DDS, DVM, LLB, JD)' = Doctorat ou équivalent, NA = donnée manquante
number_children	Nombre entier	Nombre d'enfants déjà présents (NA = donnée manquante)
time_last_pregnancy	Chaîne de caractères	Durée depuis la dernière grossesse : 'Zero to 3 months (plural delivery)' = 0 à 3 mois (grossesse multiple), autre niveau = intervalle en mois, NA = donnée manquante
mother_smoking	Chaîne de caractères	Mère ayant fumé pendant la grossesse : "N" = Non, "Y" = Oui, NA = donnée manquante
body_mass_index	Nombre réel	Indice de masse corporelle de la mère (NA = donnée manquante)
pregnancy_weight_gain	Nombre entier	Gain de poids pendant la grossesse en livres américaines (1lb = 0,45kg environ ; NA = donnée manquante)
no_infections	Chaîne de caractères	Pas d'infection durant la grossesse : '(0) False' = au moins une, '(1) True' = aucune infection, NA = donnée manquante
fetal_presentation	Nombre entier	Présentation du bébé : 1 = céphalique (= habituelle), 2 = en siège, 3 = autre, NA = donnée manquante
delivery_tool	Chaîne de caractères	Méthode d'accouchement : 1 = spontanée, 2 = forceps, 3 = ventouse, 4 = césarienne, NA = donnée manquante
delivery_method	Chaîne de caractères	Nature d'accouchement : "(1) Vaginal" = vaginal, "(2) C-Section" = césarienne, NA = donnée manquante

Nom de la colonne	Type de variable	Description
apgar5_score	Nombre entier	Score d'Apgar du bébé après 5 minutes (NA = donnée manquante)
plurality	Chaîne de caractères	gémellité : '(1) Single' = grossesse simple, '(2) Twin' = jumeaux, '(3) Triplet' = triplés
birth_weight	Chaîne de caractères	Intervalle de poids à la naissance (en grammes) (NA = donnée manquante)
no_abnormalities	Chaîne de caractères	Pas d'anormalité congénitale du bébé : "(0) False" = certaines anormalités constatées, "(1) True" = aucune anormalité observée, NA = donnée manquante

## Travail demandé

### Rapport à rendre

Vous réaliserez un court document (4 à 6 pages), introduisant le jeu de données et les problématiques que vous vous êtes posées. Vous ajouterez l'ensemble de votre notebook d'analyse en annexe de votre rapport.

Il n'est pas nécessaire d'analyser toutes les données présentes, vous pouvez vous concentrer sur un sous-ensemble de données.

Votre rapport devra être structuré de la manière suivante :

1. Introduction (présentant le contexte de votre étude et votre problématique)
2. Préparation des données (présentant des analyses descriptives des données et vos éventuels filtres sur les données)
3. Résultats (vous réalisez vos analyses et présentez vos résultats)
4. Discussion (vous recontextualisez vos résultats par rapport à votre problématique, vous en présentez les limites principales) avec une partie "Implications pour la pratique officinale" (que signifient vos résultats pour la Pharmacie et votre future pratique ?)

### Référence

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>.