

On note que les colonnes numériques sont: "age", "bmi", "children", "charges"

On peut dire qu'il y a 2 classes "male et "female" avec des proportions très proches.

Il y a largement plus de non fumeur que de fumeur concernant la variable "smoker".

On peut remarquer que les bénéficiaires sont issus de toute part des Etats-Unis mais ceux du Sud-Est et par la même occasion de l'Est, sont ceux qui profitent mieux de cette assurance.

On a donc tendance à accorder la prise en charge plus aux personnes dont les garants sont des hommes.

On peut donc retenir que les fumeurs reçoivent plus la prise en charges médicale que les non-fumeurs et pas seulement, lorsqu'ils en reçoivent, c'est vraiment plus élevé comparativement à ce qu'on pourrait donner à un non-fumeur. Généralement un fumeur a une prise en charge plus élevée qu'un non-fumeur. On pourrait déduire que l'état de santé de l'individu influence le montant de sa prise en charge. On peut constater que ceux résident dans l'Est américain reçoivent plus de prises en charges comparativement à ceux de l'Ouest. On peut constater que la prise en charge augmente avec l'âge de l'individu. On peut dire que généralement la prise en charge augmente avec la masse corporelle mais c'est beaucoup d'avoir moins de 30000 quand la masse corporelle varie entre 15 et 45. On constate que les familles avec 0 enfants sont celles qui reçoivent le plus la prise en charge et un peu moins en passant par les foyers à 1, 2 et 3 enfants. Par ailleurs, les foyers avec 4 ou 5 enfants sont rarement concernés par la prise en charge.

L'âge est directement corrélé à la prise en charge de l'assurance de même que le bmi alors que le nombre d'enfant est quand très faiblement corrélé à la prise en charge. On pourrait penser que le nombre d'enfant n'influence pas la prise.

## Interprétation

De l'ensemble des résultats de l'analyse descriptive on peut retenir que pour bénéficier d'une prise en charge médicale de façon générale il n'y a pas de distinction, mais plus vous êtes âgés, plus vous en recevrez cela multiplié par la même occasion les chances d'avoir une masse corporelle plus élevée également responsable de la hausse de la prise en charge. De façon parallèle votre état de santé, comme le fait d'être fumeur favorise l'obtention de cette prise en charge mais également le sexe du garant. Les personnes ayant des garants masculins ont plus de facilité à avoir cette prise en charge. Par ailleurs, les plus chanceux sont ceux qui vivent dans l'est des Etats-Unis. Cette préférence régionale est due à des facteurs externes comme nos recherches l'ont révélées avec ses sources entre autres:

- Kaiser Family Foundation : <https://www.kff.org/private-insurance/>
- Commonwealth Fund : <https://www.commonwealthfund.org/>
- Centre for Medicare and Medicaid Services : <https://www.cms.gov/> (<https://www.cms.gov/>) En effet, l'Est se distingue par une couverture d'assurance plus élevée, une réglementation plus stricte de l'assurance, des coûts de santé généralement plus bas et un accès plus facile aux soins, tandis que l'Ouest offre un marché de l'assurance privée plus vaste, une utilisation accrue des soins préventifs, mais avec des disparités plus importantes dans la couverture, les coûts et l'accès aux soins, en particulier dans les zones rurales.

C'est surtout à cause de ces raisons que l'Est est beaucoup plus touché par la prise en charge que l'Ouest comme on n'a pu le remarquer plutôt.

Le nuage de point des valeurs résiduelles ne montre aucune forme ou tendance, les points sont concentrés autour de 0, en montant on voit que les points se dispersent. On peut donc dire que le modèle est généralement bien ajusté mais qu'il y a des risques que l'erreur augmente lorsque le modèle a affaire à de grandes données.

### *Interprétation*

Grâce à ces nuages de point on peut dire que le fait qu'il y ait une corrélation positive entre les valeurs réelles et celles prédites, de même qu'une dispersion des points autour de 0 pour les résidus de modèle montre qu'il peut globalement généraliser.

### *Commentaires*

Au terme de cette étude nous pouvons retenir que l'obtention d'une prise en charge médicale par l'assurance dépend de nombreux paramètres à l'instar de la zone de provenance, l'état de santé de l'individu, le nombre d'enfant. Par ailleurs, le modèle obtenu est généralement performant mais comme le montre le nuage de point des résidus, on peut comprendre que l'erreur est plus grande avec des valeurs plus grandes, ce qui est provoqué par la rareté des grandes données dans le dataset de départ suggérant qu'il s'agissait de valeurs aberrantes que nous avons quand même gardé pour permettre au modèle de quand même mieux généraliser même si l'erreur est toujours présente. Aussi notons qu'en dehors des facteurs étudiés d'autres facteurs externes entre en jeu et pouvant permettre d'améliorer ce modèle, cela prouvé car la l'assurance est aussi conditionné par la zone de provenance des individus.

### *Conclusion*

Avoir une prise en charge médicale, c'est avoir de l'âge, un état de santé négatif, mais aussi le moins d'enfants possible à charge.

### *Solution*

Le plus gros problème surtout lié à la rareté des grandes sommes pour la prise en charge et la disparité entre les zones même s'il reflète la réalité des choix sont quand même des sources de biais qu'il faudra corriger avec plus de données et l'intégration de nouvelles variables explicatives en plus des anciennes afin d'avoir une meilleure description de la réalité.