

Classification Supervisée: Régression Logistique

Sandro Gazzo & Théo Dellouve

23/10/2020

Etude sur le diabète: Identification des risques associés

Ici, nous allons tenter d'identifier quels sont les facteurs de risques associés à la présence du diabète. Pour cela, on va utiliser une base de données comportant certaines caractéristiques cliniques d'un certain nombre de patients. On va, dans un premier temps, décrire la base de données.

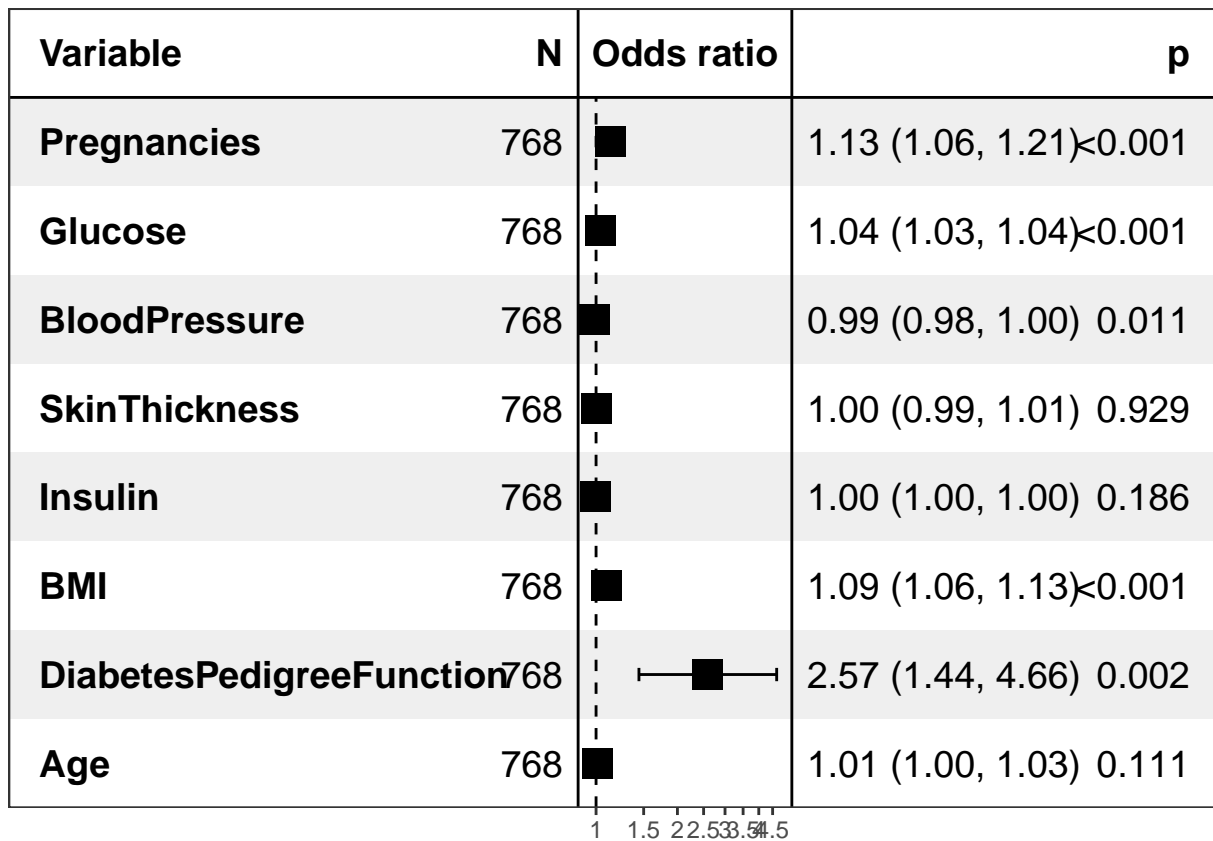
On a 768 patients pour lesquels on a relevé 9 caractéristiques:

- **Pregnancies:** nombre de grossesses;
- **Glucose:** taux de glucose;
- **BloodPressure:** pression sanguine;
- **SkinThickness:** épaisseur de la peau;
- **Insulin:** taux d'insuline;
- **BMI:** Indice de Masse Corporelle (IMC);
- **DiabetesPedigreeFunction:** taux d'hérédité du diabète;
- **Age:** âge;
- **Outcome:** si le patient est atteint du diabète ou non (1: a le diabète, 0: n'a pas le diabète).

On procède à une régression logistique en fonction de **Outcome**. La régression logistique a pour objectif de déterminer si un futur patient est atteint du diabète en fonction des caractéristiques cliniques.

Après avoir créé le modèle, on va pouvoir déterminer les facteurs de risques à l'aide des odds-ratios. Les odds-ratios permettent de mesurer l'effet d'un facteur sur le fait d'avoir le diabète.

```
## Warning in recalculate_width_panels(panel_positions, mapped_text =  
## mapped_text, : Unable to resize forest panel to be smaller than its heading;  
## consider a smaller text size
```



On remarque à la suite des odds-ratios