TD Stats avec R

M2 – semestre automne 2021

Jeanne Charoy

Ces exercices sont a faire avec la base de donnees chronolex (a trouver sur le repository github du cours).

Probleme 1. On s’interesse aux performances on decision lexicales des participants en function de la frequence des mots. Les variables importantes sont donc:

* Frequence des mots: on va utiliser la variable freqbooks
* Les temps de reaction en decision lexicale: LDTrt

Question 1. Explore la variable freqbooks et la transformer en variable categorique (0.5 pts)

1. Quelle est la frequence moyenne pour les mots contenu dans la base de donnees? La mediane?
   1. HINT: si la valeur que vous obtenez est NA c’est parce qu’il y a des “missing values” qu’il faut donc exclure
2. Comment calcule-t-on la mediane d’un ensemble de valeur? Decris le procede.
3. Creer une nouvelle variable freqbooksCAT qui peut prendre les deux niveaus “high” et “low” a partir de la variable freqbooks.

Question 2. Calcule le temps de reaction moyen en decision lexicale en function de la frequence des mots (i.e., temps de reaction ~ frequence). Reporte le tableau obtenu ci-dessous (0.5pts)

|  |  |
| --- | --- |
| Frequence | Reaction times |
| High |  |
| low |  |

Question 3. Cree un barplot qui represente le temps de reaction en decision lexicale en function de la frequence des mots. Le barplot doit avoir des error bars. (bonus: change la couleur des barres) (0.5 pts).

Probleme 2. Maintenant on va s’interesser seulement aux mots de moins de 5 lettres.

1. Cree un subset de chronolex appele chronolex petit qui ne contient que les mots de moins de 5 lettres (la variable qui contient le nombre de lettre de chaque mot s’appelle Nlett). (0.25 points)
2. Cree une nouvelle colonne appele freqMoy qui donne la moyenne de freqfilms et freqbooks. (0.25 points)