TD4

Test d’hypotheses en R (correlation, t-test)

M2 – semestre automne 2021

3pts total

Reponds aux questions suivantes et envois moi une copie electronique de tes reponses par email avant le 10 Janvier 2022. Les reponses peuvent etre courtes et peuvent etre sous la forme de bullet points sauf si indication contraire. Je vous encourage a faire le travail en groupe et a discuter – cela dit assures toi que tu comprends les reponses que tu ecris. Tu peux inclure des screenshots des outputs R ou bien tu peux aussi m’envoyer ton script R avec ce word doc (ou juste le script R si tu as repondu aux questions ecrites dessus en commentaire).

Questions ecrites sur le cours (pas de R). (1 point)

Un groupe de chercheur.euses conduit une experience sur la comprehension de la parole accentee. En particulier iels s’interessent au lien entre familiarite avec l’accent et comprehension. Leur experience se deroule de la sorte:

* Tous les participants ecoutent un texte en francais lu par une personne dont l’espagnol est la premiere langue. Leur comprehension du texte est evalue avec un score do 0 a 20 (score 1).
* Un groupe de participants (le groupe experimentale) est ensuite expose a la meme voix de manière repete pendant que l’autre groupe (le groupe control) ecoute de la musique.
* Tous les participants sont a nouveau evalue sur la comprehension d’un texte lu par la meme voix avec un score de 0 a 20 (score 2).

1. Les chercheur.euses font l’hypothese que le groupe qui a ete plus expose a la voix aura une plus grande amelioration de la comprehension. Quelle est l’hypothese nulle? (0.10)
2. Quelle est la variable independante? De quel type est elle (categorique ou continue)? (0.10)
3. Quelle est la variable dependante? De quell type est elle? (0.10).
4. Les chercheur.euses decident d’utiliser un t-test pour analyser la difference des scores de comprehension moyen du groupe experimentale (score 1 vs score 2). Doivent iels utiliser un “independent t-test” ou un “paired t-test”? (0.20)

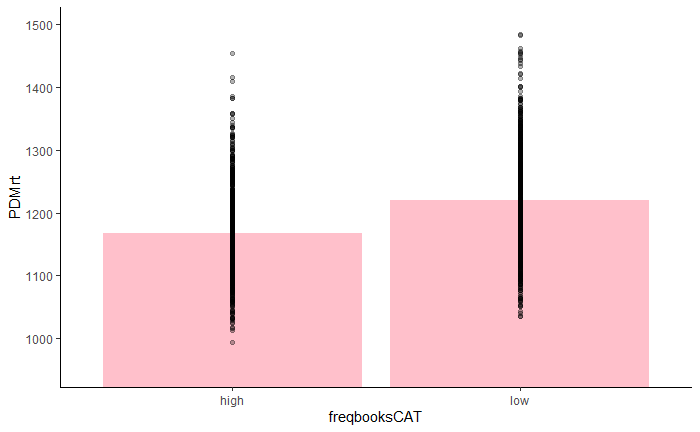
Trouver les erreurs dans les lignes de codes suivantes. Je vous encourage a ecrire les lignes dans R pour voir quell message d’erreur il vous donne si vous ne trouvez pas en regardant. (0.5)

1. Je veux creer un scatterplot. J’ai ecris (0.1): A picture containing text

   Description automatically generated
2. Je veux creer une nouvelle colonne qui contient les temps de reaction en LD multiplie par 10. J’ai ecris (0.1): 
3. Je veux connaitre le nombre de voisin de l’item “air”. J’ai ecris (0.2 – deux erreurs): 
4. Quel est le probleme ici? (0.10)

Probleme 2. Le dataset Ponari: effet des troubles du developpement du langage sur la perception des mots abstraits et concrets. (1pt)

Commencez par importer le dataset Ponari (sur Discord).

1. On s’interesse aux differences de traitement des mots concrets par le groupe DLD (trouble du langage) et le groupe TD (typical development). C’est a dire qu’on veut voir si les scores sur la variable Concrete different en function du groupe des participants.
   1. Quelle est la variable independante? De quel type est elle? (0.10)
   2. Quelle est la variable dependante? De quell type est elle? (0.10)
   3. Creez un objet meansConcrete qui contient les moyennes de reconnaissance des mots concrets (i.e., concrete) par groupe (i.e., group). (0.15)
   4. Creez un bar plot qui represente les moyennes par groupe que vous venez de calculer ainsi que les points de chaque observations (i.e. quelque chose comme ca: ) (0.25)
   5. Observe-t-on des outliers simplement avec un examen visual? (0.05)
   6. On veut s’avoir si il y a une difference significative entre les deux groupes pour les mots concrets. Quel test est approprie etant donne les variables auquelles on s’interesse ici (i.e., group et concrete)? (0.20)
   7. Applique le test que tu as choisis et reporte l’output de R ici - Est ce qu’il y a une difference significative entre les deux groupes? (0.15).

Probleme 3. Toujours avec le dataset Ponari (1pt)

1. On s’interesse maintenant aux liens entre reasoningSkills et la reconnaissance des mots abstraits (abstract) pour nos deux groupes de participants.
   1. De quel type sont les variables reasoningSkills et Abstract? (0.1)
   2. Quel test statistique est approprie pour analyser la relation entre deux variables de ce type? (0.1)
   3. Cree un subset de chronolex qui ne contient que les participants du groupe TD. (0.10)
   4. Cree un subset de chronolex qui ne contient que les participants du groupe DLD (0.10)
   5. Cree un scatterplot de reasoningSkills par abstract pour chaque groupe, avec une droite de regression (0.20)
   6. Observe tu des points suspects (des outliers potentiels)? (0.1)
   7. Quel est le coefficient de correlation entre reasoningSkills et abstract pour le groupe DLD? Pour le groupe TD? (0.1)
   8. Ces coefficients sont ils significatifs, oui ou non? Qu’est ce que cela nous indique sur la relation entre les deux variables pour chaque groupe? (0.20)