TD 2 – Covariance & Correlation

M2 – semestre automne 2021

2pts total

Reponds aux questions suivantes et envois moi une copie electronique de tes reponses par email avant Vendredi 26/11 a 8:30 AM. 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.

**Questions ecrites (1pts)**

1. est l’equation de quelle mesure de la dispersion? Explique brievement les termes de cette equation (0.5)
2. Tu transformes les valeurs d’une variable quantitative en z-score. La distribution des valeurs de la variable *apres* la transformation (compare a la distribution avant) est elle (0.25):
   1. Transformee en distribution normale
   2. Restee la meme
   3. Decalee vers la gauche (i.e., plus negative)
   4. Decalee vers la droite (i.e., plus positive)
   5. Aucune de ces reponses
3. Quelle est la relation entre variance et covariance? (0.25)

**Exercices avec R (1 pts)**

Comme la derniere fois on travaille avec le dataset chronolex.

1. Importe le dataset chronolex dans une dataframe appelee chrono sur R. On va s’interesser en particulier au temps de reaction sur une tache de decision lexical LDTrt et au nombre de voisins des mots (OLD20 – il y a plusieurs mesure de voisinage, s’en est une assez connue). (total of 0.45 points for question 1)
   1. Quelle sont la moyenne et l’ecart type de LDTrt? (0.10)
      1. Moyenne:
      2. Ecart Type:
   2. Quelles sont la moyenne et l’ecart type de OLD20? (0.10)
      1. Moyenne =
      2. Ecart Type =
   3. Cree un dataframe qui contient la moyenne et l’ecart type de LDTrt et OLD20 (0.25). Ca doit ressembler a:

|  |  |
| --- | --- |
| Mean | Sd |
| 665.48… (c’est pas grave si il y a plus/moins de decimale) | 64.028… |
| 1.501… | 0.317…. |

1. Dessine un scatterplot des temps de reaction (LDTrt) en function du nombre de voisins (OLD20). (0.15)
   1. Bonus: change les labels des axes pour que l’abscisse soit “nombre de voisins” et l’ordonnee “temps de reaction en LD”
2. Cree une matrice de variance-covariance pour LDTrt, OLD20 et freqbooks (hint: pense a creer un subset de chrono qui ne contient que les variables dont tu as besoin). (0.15)
   * 1. Quelle est la variance de
        1. LDTrt:
        2. OLD20
        3. Freqbooks
     2. Quelle est la covariance de:
        1. LDTrt et OLD20
        2. LDTrt et freqbooks
        3. OLD20 et freqbooks
3. Quel est le coefficient de correlation entre LDTrt et OLD20? Qu’indique-t-il sur la relation entre LDTrt et OLD20, c’est a dire entre les temps de reaction en decision lexicale et le nombre de voisins des mots. (0.25)