Université Iba Der THIAM de Thiès **Année Académique 2020-2021**
UFR SES/SET **Master 1 Sciences des données**

Statistique-Mathématique
Devoir maison numéro 1
Travail à faire en groupe de 3 personnes.
à rendre avant 15 août 2021 à $0h00$
(remise sur github)

Exercice 1 (Données covid-19 Sénégal)

1. Enregistrer les données dans un format adapté pour une lecture par la suite avec Python sachant que la première ligne du fichier correspond au noms des variables. Convertir la colonne date en type datetime et supprimer toutes les lignes ayant une valeur manquante.
2. Ajouter dans les données une nouvelle variable proportion qui représente le taux de positivité journalier.
3. Ajouter dans les données une nouvelle variable total qui représente le nombre cumulé de cas positifs du début de la pandémie jusqu'à une certaine date.
4. Représenter les variables tests et cas en fonction de la date dans le même graphique. Ajouter une légende dans le graphique.
5. Représenter les variables communautaire et contact en fonction de la date dans le même graphique. Ajouter une légende dans le graphique.
6. Tracer le boxplot et l'histogramme de la variable cas. Conclure sur la présence ou non d'éventuelles valeurs manquantes pour cette variable.
7. Discrétisez la variable cas. Pour ce faire on ajoutera une variable dans le DataFrame des données une nouvelle variable nommée `cas_classe`. Cette variable aura 4 classes :

$$[\min(cas), Q_1],\ \ ]Q_1, Q_2],\ \ ]Q_2, Q_3],\ \ ]Q_3, \max(cas)].$$

   ou Q_1, Q_2, Q_3 sont respectivement les 3 premiers quantiles de la variable cas, $\min(cas)$ et $\max(cas)$ respectivement la plus petite et la plus grande valeur de la variable cas.
8. Donner les fréquences des modalités de la nouvelle variable `cas_classe`.
9. On suppose que la variables cas suit une loi binomiale négative $\mathcal{NB}(\theta, p)$, $\theta > 0$ et $< 0 < p < 1$. Calculer les estimations de θ et p en utilisant la méthode des moments.
10. Même question avec la méthode du maximum de vraisemblance

Exercice 2 Soit (X, Y) un couple de variables aléatoires discrètes dont la loi de probabilité est indiquée dans le tableau ci-aprés :
a) Déterminer les lois marginales de X et Y et indiquer si ces variables aléatoires sont indépendantes.
b) Calculer Cov(X,Y).
c) Déterminer la loi du couple $[\inf(X, Y), \sup(X, Y)]$.

X \ Y	1	2	3	4
1	0,08	0,04	0,16	0,12
2	0,04	0,02	0,08	0,06
3	0,08	0,04	0,16	0,12

Exercice 3 La durée de vie d'un certain type de lampes de vidéoprojecteur suit une loi exponentielle de densité de probabilité

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0, \\ 0 \text{ sinon.} \end{cases}$$

1. Une variable aléatoire X est dite « sans mémoire », ou « sans vieillissement », si elle vérifie la propriété suivante :

   $$\mathbb{P}(X > x + x_0 | X > x_0) = \mathbb{P}(X > x),$$

   pour tous x, x_0 positifs.
   Montrer que la durée de vie de ces lampes de vidéoprojecteur est sans mémoire.
2. Déterminer la durée de vie moyenne d'une lampe en fonction de λ ?
3. Un fabriquant de lampes annonce à son client que la durée de vie moyenne des lampes qu'il produit est d'au moins 10000 heures alors que, dans les faits, une lampe sur deux a une durée de vie de moins de 7000 heures. Est-il honnête ? Justifier.