Sujet d'examen 4

Exercice 1:
Soit X une variable aléatoire admettant la fonction f pour fonction de densité.
On souhaite obtenir des réalisations de X par la méthode du rejet, qui je vous le rappelle, est la suivante :

1. Soit $c > 0$ et g une fonction de densité pour laquelle on sait obtenir des réalisations et telle que $f \leq c.g$. On note Y la variable aléatoire de densité g.
2. Soit u une réalisation d'une loi uniforme sur $[0;1]$.
3. Soit y une réalisation de Y.
4. On pose $z = c.u.g(y)$
5. si $z \leq f(y)$ alors on pose $x = y$ sinon on recommence à l'étape 2.

La réalisation obtenue est alors une réalisation de X.

1. Application : Soit $X \sim \mathcal{N}(0;1)$
   On peut montrer que $|X|$ admet pour fonction de densité:
   $$f(x) = \begin{cases} \frac{2}{\sqrt{2.\pi}} \exp\left(-x^2/2\right) \text{ si x } >0 \\ 0 \text{ sinon} \end{cases}$$
   On peut montrer que
   $$\forall x \in \mathbb{R}^+,\ f(x) \leq \sqrt{\frac{2.e}{\pi}} \exp\left(-x\right)$$
   Simuler 10000 réalisations de $|X|$.
2. Soit T une variable aléatoire telle que $P(T=-1) = P(T=1) = 1/2$. On peut montrer que $T.|X|$ a même loi que X.
   Simuler 10000 réalisations de X.
3. Par un graphique sur lequel vous ferez apparaitre un histogramme de vos données et la fonction de densité de la loi normale centrée réduite, vérifiez que vos simulations sont correctes. Vous prendrez soin de rappeler comment on obtient un histogramme.

Exercice 2:
Soit le fichier temp.txt.
Réalisez une analyse en composantes principales de ce jeu de données.
Vous expliquerez le principe d'une telle analyse, le nombre d'axes conservés, et vous interpréterez les deux représentations graphiques usuelles.

Exercice 3:
Une entreprise se spécialise dans l'analyse de système et la programmation sur ordinateurs de problèmes techniques et de gestion. Elle veut utiliser al régression dans une étude sur le temps requis, par ses analystes, pour programmer des projets complexes.
Les données suivantes représentes le temps total en heure pour programmer différents projets (variable X) et le nombre d'instructions de chaque programme (varible Y).

x_i	40	55	62	58	62	94	120	134	128	140	152	174	167	218
y_i	60	82	100	142	190	220	285	354	400	425	440	500	530	640

1. Réalisez le nuage de points et expliquez pourquoi une modélisation par un modèle linéaire est envisageable.
2. On considère alors le modèle $Y = a_0 + a_1.X + \varepsilon$. Définir l'écriture matricielle associée à ce modèle et définir les objets.
3. Par le logiciel R, obtenez les estimations de a_0 et a_1 et tracez sur le nuage de point la droite de régression.
4. Evaluez la qualité du modèle.

Exercice 4: question de cours
Expliquer la méthode de k-means.