1. Dans cet exercice, $x_1, \ldots, x_n$ est un échantillon numérique, et $y_1, \ldots, y_n$ est obtenu par transformation affine de $x_1, \ldots, x_n$:
$$y_i = \sigma x_i + \mu$$
avec $\sigma > 0$ et $\mu \in \mathbb{R}$. F_n désigne la fonction de répartition associée à $x_1, \ldots, x_n$, G_n la fonction de répartition associée à $y_1, \ldots, y_n$. Calculer la fonction $G_n^{\leftarrow} \circ F_n$ (la fonction qui définit le `qqplot`). L'échantillon est muni d'une pondération $p_1, \ldots, p_n$ ($\sum_{i=1}^n p_i = 1; \quad \forall i \leq n \colon p_i \geq 0$).
2. Montrer que si $x_1, \ldots, x_n$ est un échantillon numérique de variance σ^2 et d'écart inter-quartile IQR alors
$$\sigma^2 \geq \frac{\text{IQR}^2}{8} .$$
Ce résultat reste-t-il valable si l'échantillon est pondéré ?
3. En exhibant un exemple, montrer que le résultat précédent ne peut pas être amélioré.
4. Sur la planète Aldorande vivent deux populations les Nabous et les Hots. Les individus sont soit pauvres soit riches. On sait que 90% des Nabous sont pauvres et que 90% des pauvres sont des Nabous. Peut-on en conclure que la richesse est inégalement répartie entre les deux peuples ?
5. Si $x_1, \ldots, x_n$ est un échantillon numérique de moyenne $\bar{X}_n$ de variance V, est-il vrai que $V = \left((1/n) \sum_{i=1}^n x_i^2 - (\bar{X}_n)^2\right)$? Savez-vous le démontrer ?
6. Définir le coefficients de corrélation linéaire entre deux variables sur un même échantillon pondéré. Montrer que sa valeur absolue est comprise entre 0 et 1.
7. Dans un pays imaginaire, les conditions de mortalité sont décrites par une suite de quotients de mortalité $q_0, q_1, \ldots, q_N$ ($N = 150$). Le quotient q_i représente la proportion d'individus qui décèdent entre l'âge i et l'âge $i+1$ pendant une année (parmi les individus d'âge compris entre i et $i+1$ au premier janvier de l'année). Expliquer comment calculer à partir de la fonction de survie $\bar{F}(i) = \prod_{j=0}^{i-1} q_j$ ($F = 1 - \bar{F}$ est la fonction de répartition de la loi de survie),
   (a) la durée de vie médiane associée ;
   (b) les quartiles de la durée de vie ;
   (c) la variance de la durée de vie associée.

Corrélations

On veut étudier l'association entre l'espérance de vie à la naissance et les différents quotients de mortalité dans les départements français au XIX^{eme}

On utilise toujours la table de l'INED : (`tmort <- read.csv2("tables-mortalite-france.csv")`). On reprend les calculs d'espérance de vie à la naissance effectués durant les séances précédentes.

1. Calculer le coefficient de corrélation linéaire entre le quotient de mortalité juvénile et l'espérance de vie à la naissance.
2. Calculer le coefficient de corrélation linéaire entre les quotients de mortalité et l'espérance de vie à la naissance.
3. Calculer année par année les coefficients précédents (pensez à la fonction **aggregate**).
4. A l'aide de la fonction **pairs** (et en créant éventuellement un **data.frame** approprié), visualiser les nuages de points dont les coordonnées sont les espérances de vie à la naissance et les différents coefficients de mortalité (pour chaque âge, visualisez le nuage des points dont l' abscisse est le quotient de mortalité observé à cet âge dans le département et l'année alors que l'ordonnée est l'espérance de vie à la naissance dans ce même département et cette même année).