# TP4 : Quantiles-quantiles plots et droite de Henry

**Exercice pour l'extraction de données :** Charger le jeu de données cardiaque.csv et l'affecter à `cardiaque`. En extraire l'échantillon des pressions systoliques chez les patients souffrant d'une pathologie cardiaque (modalité 1 de la variable `cardiaque`) et en donner les résumés numériques usuels : taille, moyenne, écart-type empirique corrigé (noté $s'$ dans le cours) et quartiles.

**Objectifs :** *Savoir apprécier graphiquement la normalité ou non d'une suite d'observations d'une variable $X$ à l'aide d'un histogramme ou d'un quantile-quantile plot accompagné de la droite de Henry.*

Dans cette partie $X$ est une variable aléatoire de loi $\mathcal{N}(2, 4)$ (soit $\mu$=2, $\sigma = 2$).

**Exercice 1 : quantiles empiriques ou théoriques**

Le quantile empirique d'ordre $\alpha$ d'un échantillon de $n$ tirages indépendants $x_1, x_2, ..., x_n$ de $X$ est la valeur pour laquelle, la proportion des $x_i$ qui se trouvent inférieurs ou égaux à cette valeur, vaut $\alpha$. Le quantile théorique d'ordre $\alpha$ appelé quantile (et calculé par la fonction de quantile `qnorm()`) est la valeur telle que la probabilité que $X$ soit inférieur ou égal à cette valeur, vaille $\alpha$.

1. Simuler un échantillon de taille $n = 50$ de $X$ et l'affecter à `x`.
2. Calculer les quantiles empiriques d'ordre $10\%, 20\%, ..., 90\%$ de `x` (appelés aussi les déciles empiriques car ils permettent de couper l'échantillon en dix) à l'aide de la fonction `quantile()` et en utilisant l'option `type=1`.
3. Calculer les quantiles théoriques d'ordre $10\%, 20\%, ..., 90\%$ de $X$ à l'aide de la fonction `qnorm()`.
4. Faire un graphe des 9 points d'abscisses les déciles et d'ordonnées les déciles empiriques. Ajouter la droite $y = x$. Commentaire.
5. Faire un histogramme des observations tirées et y superposer la densité de la loi $\mathcal{N}(2, 4)$. Commentaire.
6. Calculer les quantiles empiriques d'ordre $1/(n+1), 2/(n+1), ...., n/(n+1)$ de `x`. Les comparer aux valeurs ordonnées de l'échantillon que l'on obtiendra avec la commande `sort(x)`.
7. Calculer les quantiles théoriques d'ordre $1/(n+1), 2/(n+1), ...., n/(n+1)$ de $X$ et tracer les points d'abscisses quantiles théoriques et ordonnées quantiles empiriques. Y ajouter la droite $y = x$.
8. Calculer les quantiles théoriques d'ordre $1/(n+1), 2/(n+1), ...., n/(n+1)$ de la loi normale centrée réduite. Tracer les quantiles empiriques en fonction de ces nouveaux quantiles théoriques. Y ajouter la droite d'équation $y = \sigma x + \mu$ pour $\mu = 2$ et $\sigma = 2$. Cette droite s'appelle droite de Henry et permet d'apprécier la normalité d'un échantillon $x_1, ..., x_n$. Plus les points sont proches de la droite plus l'hypothèse de normalité faite sur la variable qui a généré ces observations est correcte.
9. Comparer le graphique précédemment obtenu avec le résultat de la commande `qqnorm(x)`. La commande `qqline(x)` permet-elle d'ajouter la droite de Henry ? Si non comment est construite cette droite ?

**Exercice 2 : diagnostiquer la normalité de données par un qqplot**

Le but ici est de comprendre comment interpréter des échantillons pour lesquels les points représentés ne sont pas alignés autour de la droite de Henry.

1. Choisir une valeur pour $n$ (supérieure à 100) et l'affecter à `n`.
2. Simuler un échantillon de taille $n$ de la loi uniforme sur $[-1, 1]$ avec la fonction `runif` et l'affecter à `x`. Rappeler son espérance que l'on notera $\mu$ et sa variance $\sigma^2$. Tracer le nuage des points croisant quantiles théoriques et quantiles empiriques avec `qqnorm(x)` et y superposer la droite de Henry obtenue avec `qqline(x)`. Interpréter.
3. Simuler un échantillon de taille $n$ de la loi de Student de paramètre 3 avec la fonction `rt` et refaire le même graphique que dans la question précédente. Pour aider l'interprétation, tracer sur un même graphe la densité de la loi $\mathcal{N}(0, 1)$ et celle de la Student à 3 degrés de liberté. Interpréter.
4. Refaire la même chose avec un échantillon de taille $n$ de la loi du Chi-deux de paramètre 2 avec la fonction `rchisq`. Commenter.