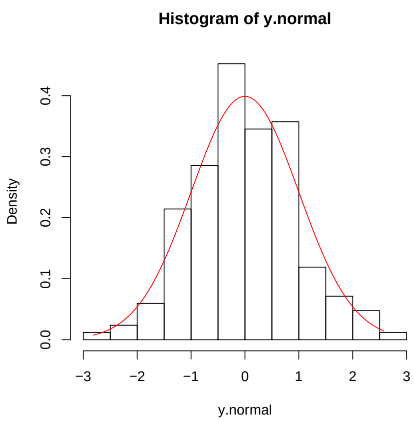
Cours 05

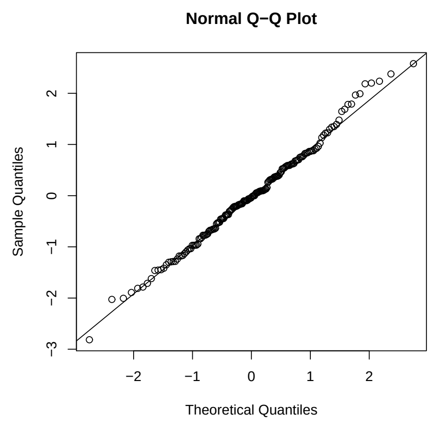
`qqnorm` et `qqplot`

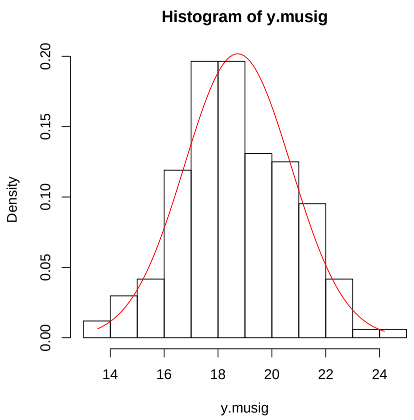
1 Cas d'échantillons normaux

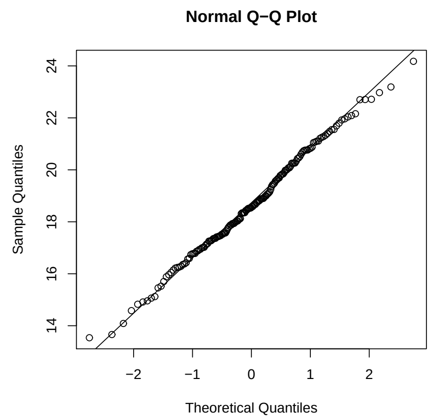
Si un échantillon `y` suit une loi normale de densité $f_{\mu,\sigma}$ la proportion p d'individus x_i tels que $y_i \leq q$ est approximativement $p = \int_{-\infty}^{q} f_{\mu,\sigma}(x)dx$, c'est-à-dire la probabilité pour cette loi d'être inférieur à q.

Supposons pour simplifier que les y_i =`y[i]` sont rangés de manière croissante [1]. Pour voir dans quelle mesure une loi normale est un bon modèle pour l'échantillon `y`, il convient de comparer les y_i au n-fractile d'indice i, c'est-à-dire à q_i =`qnorm(`p_i`,`μ`,`σ`)`, avec $p_i = \frac{i-0.5}{n}$. C'est ce que fait la commande `qqnorm(y)`, pour $\mu = 0$ et $\sigma = 1$ (loi normale centrée-réduite). Voici, à droite, ce que produit cette commande, d'abord pour un échantillon simulé normal centré réduit, puis pour une échantillon simulé avec $\mu = 18.73155$ et $\sigma = 1.919205$ [2]. On a matérialisé les deux échantillons simulés par leur histogramme, représenté à gauche à auquel on à superposé le graphe de la fonction $f_{\mu,\sigma}$.







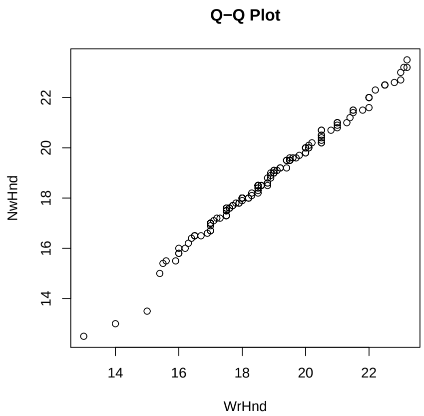


L'instruction `qqline(y)` permet d'ajouter une droite passant par le premier et le troisième quartile de la loi théorique (ici $f_{0,1}$) et de l'échantillon `y`. Voir `?qqnorm` pour savoir comment choisir d'autres valeurs des paramètre que celles par défaut pour cette commande. Si on pose `qqny=qqnorm(y)`, on retrouve les abscisse utilisées pour la figure dans `qqny$ x` (dans l'ordre des individus de `y`). On peut vérifier qu'elles correspondent bien aux fractiles des p_i indiqués en comparant `sort(qqny$x)` à `qnorms(probas)`, si on a posé `probas=qnorm((1:length(y)-0.5)/length(y))`.

1. sinon, il suffit de remplacer `y` par `sort(y)`
2. on reprend ici la moyenne et l'écart-type des l'échantillon mesuré de `NwHnd`

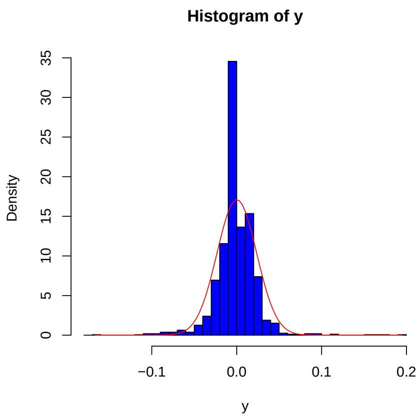
2 Quantile contre Quantile : `qqplot`

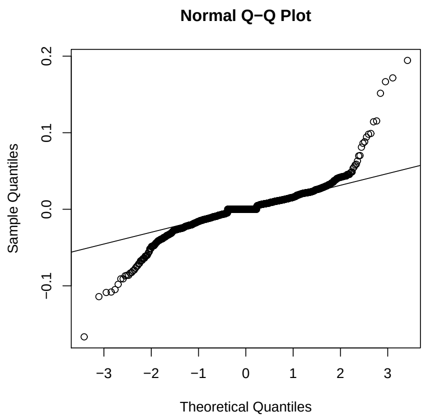
Si au lieu de comparer des quantiles empiriques de `y` à une loi théorique telle $f_{\mu,\sigma}$ on veut les comparer aux quantiles empiriques d'un autre échantillon `x`, il suffit d'utiliser la commande `qqplot(x,y)` qui compare les quantiles homologues des deux échantillons. Ici le résultat pour les échantillons `x=WrHnd` et `y=NwHnd` déjà considérés dans les cours antérieurs. Si, comme ici, les deux échantillons ont même taille n =`length(x)=length(y)`, cette commande revient simplement à `plot(sort(x),sort(y))`. Voyez-vous ce qu'il faudrait choisir comme argument à la fonction `plot` dans la cas où les échantillons ne sont pas de même taille, par exemple si `length(y)=2*length(x)` ?[3]



3 Rendements : un cas partiellement gaussien

Voici un exemple pratique. On a choisi pour `y` l'échantillon que nous avons déjà considérés de rendements sur les prix du caoutchouc sur le marché de Hatyai. On observe que le modèle gaussien est satisfaisant, sauf pour les rendements égaux à 0 et les rendement très éloignés de cette valeur.





3. Réponse : `plot(sort(x),sort(y)[seq(2,length(y),2)])`. Voir ce qui est dit de `type` dans `?quantile` pour découvrir les diverses solutions adoptées par la statistique pour traiter de la délicate question des quantiles lorsque le fractile considéré ne conduit pas à un indice entier.