

## Colles série 4 : QQ plot d'adéquation à la loi normale et propriétés des quantiles d'une loi normale

---

### Sujet 1:

Soit  $X$  une variable de loi  $\mathcal{N}(\mu, \sigma^2)$  avec  $\mu = 1$  et  $\sigma^2 = 1$ .

1. Dans quel intervalle centré en  $\mu$  peut-on garantir que plus de 99 fois sur 100 on observera une réalisation de  $X$  ? Représenter sur un même graphique les fonctions de répartition et densité de  $X$  sur cet intervalle.
2. Ajouter les éléments permettant une lecture graphique des trois quartiles.

### Sujet 2:

Soit  $X$  une variable de loi  $\mathcal{N}(\mu, \sigma^2)$  avec  $\mu = 2$  et  $\sigma^2 = 3$  et  $U$  une variable de loi normale centrée et réduite.

1. Calculer les quantiles d'ordre  $(0.1, 0.2, \dots, 0.9)$  de  $U$  et les affecter à `qu` puis les quantiles d'ordre  $(0.1, 0.2, \dots, 0.9)$  de  $X$  qui seront affectés à `qx`. Représenter le nuage des points d'abscisses les quantiles de  $U$  et d'ordonnées ceux de  $X$ .
2. Quelle est l'équation de la droite obtenue (indication : on cherchera par une lecture graphique l'ordonnée à l'origine et la pente et on superposera la droite d'ordonnée à l'origine et de pente lues avec `abline`) ? Donner l'expression du quantile d'ordre  $\alpha$  de  $X$  en fonction du quantile d'ordre  $\alpha$  de  $U$ .

### Préambule sujets 3 à 6 :

Ces sujets portent sur les données du fichier `donneesSerie4.csv` qui sera téléchargé dans le répertoire qui sera ensuite choisi dans R studio comme répertoire de travail (via le menu `session/set working directory /...`). Ensuite pour charger les données sous R il faut exécuter :

```
read.table("donneesSerie4.csv", header=T, sep=" ", dec=".") -> dat
```

Le data.frame nommé `dat` ainsi créé se manipule comme `mtcars`. Il est constitué de six échantillons de taille  $n = 100$  tirés sous des lois différentes dont trois sont des lois normales, une est uniforme, une de Student et une du Chi-deux. Les paramètres de ces lois sont inconnus.

### Sujet 3:

Dans cet exercice nous n'utiliserons que les échantillons `e1` et `e2`.

1. Faire un résumé numérique de chacun des deux échantillons et les comparer. Calculer leurs écart-type empiriques corrigés ( $s'$ ) et les comparer. Ces comparaisons permettent-elles de dire que ces échantillons sont tirés sous la même loi ou le contraire ?
2. Faire le QQ-plot de l'échantillon `e1` en fonction des quantiles de la loi normale centrée réduite (avec `qqnorm()`). Y ajouter, en rouge, la droite de Henry dont l'ordonnée à l'origine et la pente sont respectivement la moyenne empirique et l'écart-type empirique corrigé  $s'$  (avec `abline()`). Faire de

même avec **e2** et afin de comparer les deux graphiques les représenter sur une même fenêtre coupée en deux parties (1 ligne x 2 colonnes) en imposant les mêmes limites en ordonnées pour chacun des deux `qqnorm()`. Que conclure (on pourra noter l'intervalle contenant tout l'échantillon **e2**) ? Sachant les lois utilisées, laquelle proposer pour chacun des deux échantillons ?

#### Sujet 4:

Dans cet exercice nous n'utiliserons que les échantillons **e3** et **e4**.

1. Faire un résumé numérique de chacun des deux échantillons et les comparer. Calculer leurs écart-type empiriques corrigés ( $s'$ ) et les comparer. Ces comparaisons permettent-elles de dire que ces échantillons sont tirés sous la même loi ou le contraire ?
2. Faire le QQ-plot de l'échantillon **e3** en fonction des quantiles de la loi normale centrée réduite (avec `qqnorm()`). Y ajouter, en rouge, la droite de Henry dont l'ordonnée à l'origine et la pente sont respectivement la moyenne empirique et l'écart-type empirique corrigé  $s'$  (avec `abline()`). Que conclure de ce nuage de points et en particulier du signe des quantiles empiriques ? Faire de même avec **e4** et afin de comparer les deux graphiques les représenter sur une même fenêtre coupée en deux parties (1 ligne x 2 colonnes) en imposant les mêmes limites en ordonnées pour chacun des deux `qqnorm()`. Sachant les lois utilisées, laquelle proposer pour chacun des deux échantillons ?

#### Sujet 5:

Dans cet exercice nous n'utiliserons que les échantillons **e5** et **e6**.

1. Faire un résumé numérique de chacun des deux échantillons et les comparer. Calculer leurs écart-type empiriques corrigés ( $s'$ ) et les comparer. Ces comparaisons permettent-elles de dire que ces échantillons sont tirés sous la même loi ou le contraire ?
2. Faire le QQ-plot de l'échantillon **e5** en fonction des quantiles de la loi normale centrée réduite (avec `qqnorm()`). Y ajouter, en rouge, la droite de Henry dont l'ordonnée à l'origine et la pente sont respectivement la moyenne empirique et l'écart-type empirique corrigé  $s'$  (avec `abline()`). Que conclure de ce nuage de points et en particulier de la symétrie des quantiles empiriques ? Faire de même avec **e6** et afin de comparer les deux graphiques les représenter sur une même fenêtre coupée en deux parties (1 ligne x 2 colonnes) en imposant les mêmes limites en ordonnées pour chacun des deux `qqnorm()`. Sachant les lois utilisées, laquelle proposer pour chacun des deux échantillons ?

#### Sujet 6:

Dans cet exercice nous n'utiliserons tous les échantillons du data.frame **dat**.

1. Partitionner la fenêtre graphique en deux lignes et six colonnes. Faire les histogrammes (en densité) de chacun des six échantillons sur la première ligne en ajoutant à chaque histogramme une verticale rouge passant par la valeur moyenne de l'échantillon. Sur la seconde ligne les six QQ-plot selon la loi normale (avec `qqnorm()`) auquel seront ajoutées les droites de Henry respectives (d'ordonnée à l'origine la moyenne de l'échantillon et pente l'écart-type corrigé  $s'$ ).
2. Trois échantillons sont obtenus avec des lois normales, lesquels ? Précisez-en approximativement les paramètres. Pour les trois autres échantillons, leur associer une des trois autres lois proposées (chacune n'étant utilisée que pour un échantillon).