

Statistique Inférentielle - TD3

Test de normalité

La normalité d'un échantillon est d'une importance cruciale car elle permet d'utiliser certains tests paramétriques, comme par exemple le test de Student, lorsque les échantillons ne sont pas grands ($n < 30$). Il est indispensable d'utiliser ces tests paramétriques lorsque cela est possible, car ils sont généralement plus puissants que leur alternatives non paramétriques (test de Wilcoxon ou test du signe pour le test de Student). Nous allons voir dans cet exercice comment il est possible de juger graphiquement de la normalité des données puis de tester cela à l'aide du test de Shapiro-Wilk.

Simulation de données

- Simuler un échantillon de 100 données provenant de la loi normale d'espérance 20 et de variance 10, à l'aide de la commande `rnorm`.
- Représenter les données à l'aide d'un histogramme (`hist`) et d'une fonction de répartition empirique (`ecdf`).
- Donner une estimation de l'espérance et la variance des données.
- Superposer sur l'histogramme la densité de probabilité de la loi normale d'espérance et variance données par la question précédente. Faites de même avec la fonction de répartition sur le graphique de la fonction de répartition empirique.
- Tester la normalité des données à l'aide du test de Shapiro-Wilk (`shapiro.test`).
- Recommencer l'opération avec un échantillon de taille 10 puis de taille 1000.
- Recommencer cette fois en simulant les données à l'aide d'une loi exponentielle d'espérance 20.

Données réelles

Reprenez la deuxième question de l'exercice 1 du TD 2 en testant la normalité des données pour les hommes et pour les femmes.