

### Exercice n° 1. Régression linéaire (cf. exo de TD sur les eucalyptus)

Récupérer les fichiers de données (circonférences et hauteurs) aux adresses suivantes :

[https://www-fourier.ujf-grenoble.fr/~rossigno/Enseignement/ens\\_files/circ.txt](https://www-fourier.ujf-grenoble.fr/~rossigno/Enseignement/ens_files/circ.txt)

[https://www-fourier.ujf-grenoble.fr/~rossigno/Enseignement/ens\\_files/h.txt](https://www-fourier.ujf-grenoble.fr/~rossigno/Enseignement/ens_files/h.txt)

1. Représenter les données sous la forme d'un nuage de points (hauteur en fonction de la circonférence).
2. Effectuer une régression affine et représenter sur le graphique :
  - la droite de régression,
  - deux autres droites, parallèles à la droite de régression, et telles que 80% des points soient situés entre ces deux droites.
3. Calculer le coefficient de corrélation entre la hauteur et sa "prédiction affine".

### Exercice n° 2. Comparaison de moyennes pour des échantillons gaussiens de même variance

On traite 12 parcelles de terrain identiques avec un engrais A et 12 autres avec un nouvel engrais B. On obtient avec A une production moyenne de 4,8 quintaux avec un écart-type empirique observé égal à 0,36. Les quantités correspondantes pour B valent 5,7 et 0,4. On modélise cette expérience comme en TD.

1. Donner la valeur observée sur ces données pour l'intervalle de confiance construit en TD, au niveau 95%.
2. Pensez-vous que l'on puisse considérer les espérances de production comme différentes selon qu'on utilise l'engrais A ou l'engrais B?

### Exercice n° 3. Estimation

On peut construire plusieurs intervalles de confiance pour le paramètre  $p$  de la loi binomiale de paramètres  $n$  et  $p$ , sur la base de l'observation d'une variable aléatoire  $X$  de loi binomiale<sup>1</sup> de paramètres  $n$  et  $p$  :

1. en notant  $q_\alpha$  le quantile d'ordre  $\alpha$  de la loi  $\mathcal{N}(0, 1)$ , un intervalle de confiance  $[\frac{X}{n} \pm \frac{q_{1-\alpha/2}}{2\sqrt{n}}]$  dont on contrôle le niveau asymptotique (au moins  $1 - \alpha$ ) grâce au théorème central limite,
2. celui de niveau  $1 - \alpha$  qu'on peut obtenir grâce à l'inégalité de Tchebychev :  $[\frac{X}{n} \pm \frac{1}{2\sqrt{n\alpha}}]$ .

Pour différentes valeurs de  $n$  (par exemple,  $n = 30$ ,  $n = 100$  et  $n = 1000$  ou  $10000$ ) et pour  $p$  variant de  $\frac{5}{n}$  à  $\frac{1}{2}$ , déterminer une approximation numérique de la probabilité que  $p$  appartienne à chacun de ces intervalles de fluctuation, au niveau 95%. Pour chaque valeur de  $n$ , vous représenterez les résultats sur la forme d'un graphique, en traçant, en fonction de  $p$ , la probabilité évoquée.

---

1. ou encore pour le paramètre  $p$  de la loi de Bernoulli de paramètre  $p$  sur la base de l'observation d'un échantillon de taille  $(X_1, \dots, X_n)$ , et en posant  $X = \sum_{i=1}^n X_i$