

### Exercice 1. -

Une entreprise fabrique un médicament sur deux chaînes de production. On s'intéresse aux variations de la quantité d'une certaine substance A contenue dans chaque médicament.

On a contrôlé le dosage de la substance A avec un échantillon de  $n = 100$  médicaments à la sortie de chacune des deux chaînes de fabrication.

On a trouvé un dosage moyen de  $10.75 \text{ mg}$  pour la première chaîne et  $10.70 \text{ mg}$  pour la deuxième. On sait par ailleurs que l'écart-type des chaînes de production est le même et est égal à  $0.2 \text{ mg}$ .

Construire un test à 1% permettant de savoir si la différence des moyennes observées est due à des fluctuations de l'échantillonnage ou bien si la chaîne de fabrication 1 produit des médicaments contenant davantage de substance A que la chaîne 2.

### Exercice 2. -

Un sondage effectué auprès de  $n = 2000$  personnes indique que 19% d'entre elles connaissent la marque de lessive Omopac.

Après une campagne publicitaire, un sondage analogue auprès de  $n = 1000$  personnes montre que 230 d'entre elles connaissent cette marque.

Peut-on considérer que la campagne a été efficace ?

### Exercice 3. -

Un archéologue utilise deux isotopes différents pour dater  $n = 130$  objets.

Pour chacun d'entre eux, il calcule la différence  $d_i$  des dates avec les deux isotopes. Ces 130 différences ont pour moyenne  $\bar{d} = 53$  ans et un écart-type  $s_d = 680$  ans.

1. Quel test doit-on effectuer pour comparer les deux méthodes de datation ?
2. Effectuer ce test et conclure.

### Exercice 4. -

On va utiliser R pour comparer les échantillons des fichiers Ech1, Ech2 et Ech3 deux à deux afin de savoir s'ils sont issus de la même population.

La commande principale est : **t.test(data1;data2;...)**

Il faut préciser si le test est unilatéral avec l'argument **alternative = "less"** ou **"greater"** ou bilatéral avec l'argument **alternative = "two.sided"**.

On rappelle que ce test n'est valable que si les échantillons sont gaussiens ou bien suffisamment grands pour appliquer le TCL, et si la moyenne est un indicateur pertinent sur les données traitées (par de valeurs extrêmes).

Faire un test bilatéral avec  $\alpha = 5\%$  ? Que remarque-t-on ?

### Exercice 5. -

1. Le fichier **Demographie.csv** contient l'espérance de vie de 50 pays classés par continent et choisis aléatoirement. L'objectif est d'utiliser R afin de déterminer si l'espérance de vie dépend du continent. On utilise pour cela l'instruction **anova(lm(EV~CONTINENT,data=Mydata))**
  - (a) Tracer les boxplot de l'espérance de vie suivant les continents.
  - (b) Justifier la possibilité d'effectuer un test de l'ANOVA.
  - (c) Conclure à partir des résultats de l'ANOVA.

2. Le fichier **Industries.csv** répertorie un certain nombre de caractéristiques (Capital, CA, Continent, ...) de 275 entreprises du numériques en 2005. L'objectif est de déterminer si l'investissement en R&D dépend du type d'industries.
  - (a) Pourquoi ne peut-on pas utiliser le test de l'ANOVA ? Quel test semble plus approprié ?
  - (b) Effectuez ce test avec R pour conclure. Quelle aurait été la conclusion si on avait effectué une ANOVA ?