# - TD n° 3 - Statistiques -

**Exercice 1** Pour étudier le poids moyen des enfants de 10 ans, et la variabilité de ce poids, on pèse $n = 130$ enfants. On obtient les résultats suivants :

| Poids | 22 | 23 | 25 | 26 | 27.5 | 28 | 28.5 | 29 | 29.5 | 30 | 32 | 33 | 34 | 37.5 | 38 | 40 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Effectifs | 1 | 1 | 2 | 3 | 8 | 15 | 20 | 21 | 23 | 22 | 5 | 2 | 1 | 4 | 1 | 1 |

1. Quel est le poids moyen d'un enfant de 10 ans ?
2. Ce poids est-il très variable ?
3. Plutôt que de donner une estimation "pontuelle" (question 1) du poids moyen d'un enfant de 10 ans, on voudrait prendre en compte la variabilité du poids et donner une "fourchette d'estimation".
   (a) Donner un intervalle de confiance contenant le poids moyen d'un enfant de 10 ans avec probabilité 95%.
   (b) Donner un intervalle de confiance contenant le poids moyen d'un enfant de 10 ans avec probabilité 99%.
   (c) Si on augmente la taille $n$ de l'échantillon, l'intervalle de confiance de niveau 99% sera t'il plus large ou plus resserré ?
   (d) On cherche le nombre $n$ de mesures nécessaires pour estimer avec une probabilité de 99% le poids moyen $\mu$ d'un enfant de 10 ans à + ou − 500 grammes, i.e avec une marge d'erreur de 500 grammes et un niveau de confiance de 99%. Déterminier ce nombre $n$.

**Exercice 2** On souhaite estimer la proportion de graines défectueuses dans une production de céréales. Pour cela on prélève un échantillon de $n = 40$ graines, et on trouve que 6 graines sont défectueuses.

1. Donner une estimation ponctuelle de la proportion de graines défectueuses dans la production.
2. Donner un intervalle de confiance contenant la proportion de graines défectueuses de la production avec probabilité 95%.
3. Quel est le nombre $n$ de graines à prélever si on veut estimer la proportion de graines défectueuses de la production avec une marge d'erreur de 10% et un niveau de confiance de 95% ? Et avec une marge d'erreur de 3% et toujours un niveau de confiance de 95% ?

**Exercice 3**
Le taux de rechute d'une certaine maladie est de 25%. Un laboratoire pharmaceutique annonce qu'un médicament permet de diminuer ce taux de rechute.

1. Des médecins testent ce médicament pour savoir si réellement il permet de diminuer ce taux de rechute, et peut être commercialisé.

   Ces médecins ne commercialiseront ce médicament que si son efficacité est réellement prouvée. Quel test statistique vont-ils faire ? Donner $H_0$, $H_1$ et justifier en interprétant les erreurs de première et deuxième espèce.
2. Ces médecins proposent alors ce médicament à 55 personnes et constatent que 13 d'entre elles rechutent.

   Quelle reproche pouvez-vous faire sur cette expérience ?

   Peuvent-ils affirmer, au risque de 5% de se tromper, que le médicament permet de diminuer ce taux de rechute ?
3. Rappeler la définition de la puissance d'un test. Comment s'interprète-t'elle ici ?
4. Supposons qu'effectivement ce médicament permet de faire baisser le taux de rechute et que le taux de rechute avec ce médicament vaut par exemple 23%. Avec quelle probabilité le test effectué par les médecins permet-il de détecter que le médiament est efficace ? Commenter.
5. Les médecins effectuent maintenant cette étude à une plus grande grande échelle : ils proposent le médicament à $n = 3000$ personnes et constatent que 691 d'entre elles rechutent (soit environ 23% ce qui est sensiblement le même pourcentage que dans la question 2).
   (a) A partir de cette étude, les médecins peuvent-ils affirmer, au risque de 5% de se tromper, que le médicament permet de diminuer le taux de rechute ?
   (b) Calculer la puissance du test, pour $n = 3000$ et lorsque le taux de rechute avec ce médicament vaut 23%.
   (c) Si le taux de rechute avec ce médicament ne vaut pas 23%, mais 22%, la puissance va être plus grande ou plus petite que pour $\theta = 0.23$ ? Donner l'allure de la courbe de la puissance en fonction de $\theta$.

**Exercice 4**
Une entreprise fabrique, en grande quantité, des tiges métalliques. Soit X la variable aléatoire qui, à chaque tige de la production, associe sa longueur. On note $\mu$ la longueur moyenne des tiges de la production et $\sigma$ la variabilité de la longueur. Les paramètres $\mu$ et $\sigma$ sont inconnus. On prélève au hasard un échantillon de $n = 50$ tiges dans la production. On note $X_1, \ldots, X_n$ les variables aléatoires donnant les longueurs des 50 tiges.

1. Donner les estimateurs usuels de $\mu$ et de $\sigma^2$.
2. Pour l'échantillon prélevé, la longueur moyenne obtenue est de $\bar{x} = \frac{1}{50}\sum_{i=1}^{50} x_i = 69$ mm avec un écart-type estimé de $s = \sqrt{\frac{1}{49}\sum_{i=1}^{50}(x_i - \bar{x})^2} = 3$ mm. Donner un intervalle de confiance contenant la longueur moyenne $\mu$ d'une tige de la production avec probabilité 95%. Et au niveau 99.73% ?
3. Le directeur de l'usine voudrait s'assurer que la norme fixée par son client est respectée, c'est-à-dire que les tiges ont en moyenne une longueur de 70 mm. Tester, au niveau 5%, si la différence observée entre la longueur moyenne des tiges de l'échantillon (69 mm) et la norme (70 mm) peut être attribuée aux fluctuations d'échantillonnage ou doit être attribuee à un mauvais réglement de la machine.
4. Peut-on affirmer, ave un risque de 5% de se tromper, que la longueur moyenne des tiges de la production est inférieure à la norme ?

**Exercice 5**

On étudie la concentration de polluant contenu dans les déchets d'une usine. On note $X$ la variable aléatoire qui à chaque prélévement, associe sa concentration en polluant (en microgrammes par litre). $X$ est une variable aléatoire d'espérance inconnue $\mu$, et de variance inconnue $\sigma^2$. Pour savoir si le polluant contenu dans les déchets d'une usine n'est pas en quantité trop élevée, on effectue 40 prélèvements. On note $X_1, \ldots, X_n$ les variables aléatoires donnant les concentrations en polluant des $n = 40$ prélèvements. On a observé les résultats suivants :

$$\bar{x} = \frac{1}{40}\sum_{i=1}^{40} x_i = \ 6.3 \text{ mg/l}; \quad \sqrt{\frac{1}{39}\sum_{i=1}^{40}(x_i - \bar{x})^2} = 1.2 \text{ mg/l}$$

Si la concentration de polluant contenu dans les déchets dépasse 6 mg/l, l'usine est déclarée hors norme.

1. Donner le test que ferait le directeur de cette usine. Celui ci veut en priorité contrôler le risque de déclarer que son usine est hors norme, si ce n'est pas le cas. Donner $H_0$, $H_1$ et justifier en interprétant les erreurs de première et deuxième espèce.
2. Effectuer le test au niveau 5% et conclure : peut-on affirmer, au risque de 5% de se tromper que l'usine est hors norme ?
3. Rappeler la définition de la puissance d'un test. Comment s'interprète-t'elle ici ?
4. Supposons dans cette question, que l'usine n'est pas aux normes et que la concentration moyenne de polluant contenu dans ses déchets vaut $\mu = 6.4$ mg/l.
   (a) Quelle est la probabilité de le detecter avec les $n = 40$ prélévements ?
   (b) Que proposez vous pour détecter que l'usine n'est pas aux normes avec grande probabilité, tout en gardant un test de niveau 5% ?
   (c) Calculer le nombre de prélévements nécessaires pour que la puissance du test soit au moins égale à 90%. (On supposera que sur cet échantillon de $n$ prélévements, l'ecart-type estimé est encore égal à 1.2 mg/l).
5. Il y a un danger écologique majeur dès que la concentration en polluant est supérieure à 7 mg/kg.
   (a) Quel test effectuerait le député écologique de la région ? Celui ci, contrairement au directeur de l'usine, considère que l'erreur la plus grave serait de déclarer l'usine non dangereuse alors qu'il y a un danger écologique. Justifier en interprétant les erreurs de première et deuxième espèce.
   (b) Le député effectue 80 prélévements et obtient une concentration moyenne de 6.4 mg/l avec un écart-type estimé encore égal de 1.2 mg/l. Effectuer le test <u>au niveau 1%</u>. Quelle est la conclusion du député ?