Fiche 6 - Statistiques

1 - Estimation et test de la moyenne

Exercice 1. On suppose que le poids d'un nouveau né est une variable normale d'écart-type égal à $0,5$kg. Le poids moyen des 49 enfants nés au mois de janvier 2018 dans la maternité de la Croix-Rousse a été de $3,3$kg.

1. Déterminer un intervalle de confiance à 95% pour le poids moyen d'un nouveau né dans cet hôpital.
2. Quel serait le niveau de confiance d'un intervalle de longueur $0,1$kg centré en $3,3$kg?

Exercice 2. On prélève indépendamment n individus d'une population divisée en deux sous-populations A et B de proportions respectives p et $1-p$ (ex: les électeurs de Jean Marie Le Pen lors du premier tour du 21 avril 2002, parmi l'ensemble des personnes ayant exprimées leur vote à cette élection).

1. Soit K le nombre d'individus de la sous-population A présents dans l'échantillon. Quelle est la loi de K?
2. Notons $F = \dfrac{K}{n}$ la fréquence empirique de la catégorie A. Donner l'espérance et la variance de cette fréquence empirique.
3. Quel est le comportement de F quand n devient grand?
4. Au premier tour de l'élection de 2002, Le Pen a receuilli 16,86% des votes exprimés. A Sainte-Consorce (Rhône), lors du premier tour, Le Pen a obtenu 199 suffrages des 927 exprimés. Le vote FN est-il significativement plus élevé dans ce village que dans le reste de la France.

Exercice 3. Je lance 100 fois une pièce de monnaie et je tombe 43 fois sur "Face". On cherche a tester l'hypothèse que cette pièce soit truquée. On suppose que la probabilité d'obtenir "Face" est de $0,6$ pour une pièce truquée. On souhaite donc trancher entre les deux hypothèses suivantes :

$$\begin{cases} H_0 & : \quad p = 0.5 \\ H_1 & : \quad p = 0.6 \end{cases}$$

1. Au risque d'erreur de première espèce $\alpha = 0.05$, doit-on rejeter l'hypothèse H_0?

2. Donner une valeur approchée du risque de seconde espèce β, c'est à dire la probabilité que l'on ne rejette pas H_0 alors que H_1 est vraie.
   Remarque : La probabilité de rejeter H_0 à raison est appelée "puissance du test". Il s'agit de $1 - \beta$.

Exercice 4. On veut comparer le salaire moyen des hommes avec le salaire moyen des femmes d'une grande multinationale qui se vante de sa politique d'égalité salariale entre hommes et femmes.

1. Parmi l'ensemble des salarié-e-s de cette entreprise, on tire 20 salaires au hasard parmi les hommes et autant parmi les femmes. La moyenne empirique du salaire des hommes est $\overline{y} = 2055$ et l'écart-type empirique $\sigma_y = 828$. La moyenne empirique du salaire des femmes est $\overline{x} = 2044$ et l'écart-type empirique $\sigma_x = 1811$. Effectuer un test statistique pour savoir si le salaire des hommes est significativement supérieur au salaire des femmes au risque d'erreur de 5%.

2. Deux des 40 salarié-e-s tiré-e-s au hasard gagnent en fait plus de 5000 euros : une femme qui gagne 9800 euros et un homme qui gagne 5269 euros. On décide de refaire les calculs sans tenir compte de ces deux personnes. Pour les 19 hommes restant, on obtient une moyenne empirique de $\overline{y} = 1886$ et un écart-type empirique $\sigma_y = 387$. Pour les femmes, la moyenne empirique devient $\overline{x} = 1636$ et l'écart-type empirique $\sigma_x = 348$. Vos conclusions sont-elles toujours les mêmes?

Exercice 5. On mesure sur 15 jeunes femmes la pression sanguine (en mmHg) avant qu'elles ne commencent à prendre la pilule puis on la remesure six mois après.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Avant	70	80	72	76	76	76	72	78	82	64	74	92	74	68	84
Après	68	72	62	70	58	66	68	52	64	72	74	60	74	72	74

Au vu de cet échantillon, peut-on conclure à un effet de la pilule sur la pression sanguine?

Exercice 6. Des relevés effectués pendant de nombreuses années ont permis d'établir que le niveau naturel des pluies dans la Beauce en millimètres par an suit une loi normale $\mathcal{N}(600, 100^2)$. Des entrepreneurs surnommés faiseurs de pluie, prétendent pouvoir augmenter de $50mm$ le niveau de pluie, ceci par insémination des nuages au moyen d'iodure d'argent. Leur procédé a été mis à l'essai durant 9 années et on a relevé les hauteurs de pluies suivantes

Année	1	2	3	4	5	6	7	8	9
mm	510	614	780	512	501	534	603	788	650

Deux hypothèses s'affrontent : ou bien l'insémination est sans effet, ou bien elle augmente réellement le niveau moyen de pluies de 50mm. Ces hypothèses peuvent se formaliser comme suit, si $m = E[X]$ où X est la variable aléatoire égale au niveau de pluie :

$$\begin{cases} H_0 & : \quad m = 600 \\ H_1 & : \quad m = 650 \end{cases}$$

Des agriculteurs hésitent à opter pour ce procédé en raison des possibles risques de pollutions dûs à l'argent et du coût onéreux du dispositif.

1. Au risque d'erreur de 5%, faut-il rejeter l'hypothèse nulle H_0 (c'est-à-dire adopter la technique d'insémination)?

2. Les agriculteurs sont même prêts à accepter H_1 si le résultat obtenu ces neufs dernières années font partie d'une éventualité improbable qui n'a que 5 chances sur 100 de se produire. Que doivent-ils faire?

2 - Test du χ^2

Exercice 7. On teste un médicament X destiné à soigner une maladie en phase terminale. On traite des patients avec ce médicament tandis que d'autres reçoivent un placebo. On compte le nombre de patients toujours vivants deux mois après le début du traitement. Voici le tableau obtenu :

	vivants	décédés
ayant reçu le médicament X	29	17
ayant reçu le placebo	38	7

Le médicament X est-il efficace?

Exercice 8. On lance un dé 600 fois de suite. On obtient les résultats suivants.

face	1	2	3	4	5	6
nombre d'occurences	88	109	107	94	105	97

Peut-on dire que ce dé est truqué?

Exercice 9. On mesure la taille du lobe frontal de 30 crabes *Leptograpsus variegatus*. Voici les 30 longueurs obtenues :

12,6	12,0	20,9	14,2	16,2	15,3	10,4	22,1	19,8	15
12,8	20	11,8	20,6	21,3	11,7	18	9,1	15	15,2
15,1	14,7	13,3	21,7	15,4	16,7	15,6	17,1	7,2	12,6

Est-ce que cette variable suit une loi normale?

3 - Modèle linéaire

Exercice 10. On considère les données suivantes qui représentent les altitudes (en m) et des températures moyennes mesurées (en $°C$) dans dix lieux différents :

X	1040	1230	1500	1600	1740	1950	2200	2530	2800	3100
Y	7,4	6	4,5	3,8	2,9	1,9	1	-1,2	-1,5	-4,5

1. Calculer $\sum x_i$, $\sum y_i$, $\sum x_i^2$, $\sum y_i^2$ et $\sum x_i y_i$.
2. Déterminer les valeurs de $\overline{x}$, $\overline{y}$, S_x^2, S_y^2, S_{xy}.
3. Déterminer les paramètres a et b de la régression linéaire et l'estimateur sans biais $\overline{\sigma^2}$.
4. Quelle température moyenne peut-on prévoir pour une altitude de 1100m en se basant sur un modèle linéaire?
5. Déterminer un intervalle de confiance pour la température correspondant à l'altitude de 1400m, puis un intervalle de prédiction pour la mesure de la moyenne de températures correspondant à l'altitude de 1000m à un niveau $0,95$.
6. Tester l'hypothèse que la pente de la droite est égale à 0,99, puis à 1, avec un risque d'erreur de $0,05$.