Chapitre :
Statistiques à deux variables

⊛ **Activité :** Cp38 (utilisation des listes avec la calculatrice)

I. Nuage de points et point moyen

Plutôt que simplement étudier un seul caractère sur une population, on peut s'intéresser à l'étude de deux caractères en même temps. On obtient alors deux listes de valeurs : x_i et y_i.
L'ensemble des couples $(x_i; y_i)$ forme une **série statistique double**.

1. Nuage de points et point moyen

On peut représenter dans un repère orthogonal l'ensemble des points de coordonnées $(x_i; y_i)$. On obtient alors ce que l'on appelle un **nuage de points**.
Le **point moyen** de ce nuage est le point $G(\overline{x}; \overline{y})$ dont les coordonnées sont les moyennes des deux séries :

$$\overline{x} = \frac{x_1 + \cdots + x_n}{n} \qquad \overline{y} = \frac{y_1 + \cdots + y_n}{n}$$

Exemple voir livre page 40.

2. Transformation affine

On peut changer chacune des deux séries en leur appliquant une fonction affine.
Soit $f : x \mapsto ax + b$ et $g : y \mapsto cy + d$.
Soit alors $z_i = f(x_i)$ et $t_i = g(y_i)$.
La double série des $(z_i; t_i)$ a alors pour point moyen le point G' de coordonnées $(a\overline{x} + b; c\overline{y} + d)$.

Exemple voir livre page 40.

▶ **Exercices :** 1,3,4p47 (moyennes, utilisation de la calculatrice ou non)

▶ **Exercices :** 11,12,13p48, 9p48

▶ **Exercices :** 15,17p49

II. Ajustement affine

⊛ **Activité :** 1 et 2 p39 (choix entre deux droites)

Dans certains cas, le nuage de points semble être formé de points presque alignés. On cherche alors à exprimer cela en effectuant un **ajustement affine**, c'est à dire à déterminer une fonction affine f dont la droite représentative est « la plus proche possible » de l'ensemble des points.
L'idée consiste à trouver la droite qui minimise les carrés des distances (verticales) des points à cette droite.

Dessin avec les distances verticales

On admet qu'il existe une unique droite vérifiant cette propriété et qu'elle passe par le point moyen du nuage.
On appelle cette droite la **droite de régression de y en x**, et on dit qu'elle est obtenue par la méthode des **moindres carrés**.
Son équation est de la forme $y = ax + b$, où a et b sont obtenus grâce à la calculatrice.
On donne les formules suivantes :

$$a = \frac{cov(x,y)}{V(x)} = \frac{\frac{1}{n}\Sigma(x_i - \overline{x})(y_i - \overline{y})}{\frac{1}{n}\Sigma(x_i - \overline{x})^2} \qquad b = \overline{y} - a\overline{x}$$

L'intérêt d'une telle droite est par exemple de prévoir une valeur future (si les x_i sont des années).

▶ **Exercices :** 22,23,24p50

▶ **Exercices :** 26,27,28p51 (lire page 43 des méthodes)

III. Ajustement non affine

Dans les trois cas suivant, on remarque que l'on se ramène, par changement de variable (en appliquant des fonctions aux listes de données), à un ajustement affine.

1. Par une parabole

Si le nuage de points semble suivre une parabole, avec y dépendant de x de manière polynomiale de degré 2, on détermine d'abord l'abscisse du sommet x_S (c'est la valeur de x pour laquelle y est extremale).
On pose alors $t = (x - x_S)^2$, puis on fait un ajustement affine de y en t : $y = at + b$.
Par suite, on remplace alors t par sa définition, et on obtient

$$y = a(x - x_S)^2 + b$$

▶ **Exercice :** 1p44

2. Par une exponentielle

Si le nuage de points semble suivre une courbe exponentielle, avec y dépendant de x de manière exponentielle, donc en particulier forcément positif, on pose $z = \ln(y)$.
On fait ensuite un ajustement affine de z en x : $z = ax + b$.
Par suite, on remplace z par sa définition puis on applique l'exponentielle :

$$\ln(y) = ax + b \Leftrightarrow y = e^{ax+b} = e^b \times e^{ax}$$

Le plus souvent, on donne une valeur approchée de e^b. Parfois on peut utiliser le fait que $e^{ax} = (e^a)^x$, et donner une valeur approchée de e^a (voir dans le chapitre sur la fonction exponentielle la section sur les fonctions puissances).

⊛ **Activité :** lire l'exercicé corrigé H page 149

▶ **Exercice :** 81p160

3. Par un logarithme

Si le nuage de points semble suivre une courbe logarithmique, avec y dépendant de x de manière logarithmique, on pose $z = e^y$.
On fait ensuite un ajustement affine de z en x : $z = ax + b$.
Par suite, on remplace z par sa définition puis on applique l'exponentielle :

$$e^y = ax + b \Leftrightarrow y = \ln(ax + b)$$

▶ **Exercice :** 88p162

IV. Adéquation avec une loi équirépartie

Lire : livre page 112
On considère une expérience aléatoire ayant n issues. On fait cette expérience un nombre N de fois, ce qui fait que l'on obtient des fréquences f_i pour toutes les issues (i allant de 1 à n). La loi équirépartie est par définition telle que la probabilité de chaque issue est $\frac{1}{n}$. On cherche à savoir si l'expérience réalisé permet de confirmer que la loi qui concerne l'expérience aléatoire est bien équirépartie.
Pour cela, on calcule une « distance » entre les deux fréquences (obtenues et théoriques) :

$$d^2_{\text{obs}} = \left(f_1 - \frac{1}{n}\right)^n + \dots \left(f_n - \frac{1}{n}\right)^2$$

Ensuite, on simule un grand nombre de fois (disons 500) la simulation de N expériences où l'on calcule cette même distance. On obtient donc 500 distances, et l'on peut alors déterminer les valeurs statistiques de cette série, en particulier le 9ème décile (ou le 90ème centile).

Théorème 90% des valeurs de d^2 obtenues lors de la simulation de la loi équirépartie sont inférieures à D_9 (le neuvième décile) ; si la valeur observée d^2_{obs} trouvée lors de l'expérience sur la pièce testée est telle que $d^2_{\text{obs}} < D_9$, on conclut avec un risque d'erreur de 10% que la pièce est équilibrée.

Preuve : Admis □

▶ **Exercice :** 29,30p127