CHAPITRE 8 :STATISTIQUES

Table des matières

1 Quelques rappels sur les statistiques à une variable 2
1.1 Moyenne ,Médiane et écart type 2
1.1.1 Moyenne 2
1.1.2 Médiane 2
1.1.3 Variance et écart type 2
1.2 les Quartiles 2
1.3 Déciles 2
1.4 Diagramme en boîte 3
1.5 Exercice 3

2 Adéquation à une loi équirépartie 5
2.1 Problématique 5
2.2 Répartition des données 5
2.3 Mesure de l'écart entre les 2 distributions 5
2.3.1 L'indicateur d^2_{obs} 5
2.3.2 L'exemple 5
2.4 Critères de décision 6
2.4.1 Fixation d'un seuil de décision 6
2.4.2 Prise de décision 6
2.4.3 L'exemple 6
2.5 Exercices 6
2.5.1 Exercice 1 :Extrait de l'épreuve de juin 2003 6
2.5.2 Exercice 2 :Extrait de l'épreuve de mars 2003 à Pondichéry 7

3 Statistiques à 2 variables. 8
3.1 Point moyen - Covariance 8
3.1.1 Nuage de points - Point moyen 8
3.1.2 variance - covariance 9
3.2 Ajustement affine par la méthode des moindres carrés 9
3.3 Exemples d'ajustements non affines 11
3.3.1 Ajustement logarithmique 11
3.3.2 Ajustement exponnetielle 12

1 Quelques rappels sur les statistiques à une variable

1.1 Moyenne ,Médiane et écart type

1.1.1 Moyenne

Définition

On la note $\overline{x}$ et elle est définie par : $\overline{x} = \dfrac{\sum_{i=1}^{k} n_i x_i}{\sum_{i=1}^{k} n_i} = \dfrac{n_1 x_1 + n_2 x_2 + \ldots + n_k x_k}{n_1 + n_2 + \ldots + n_k}$

$n_1, n_2, n_3, \ldots, n_k$ sont les effectifs correspondants aux valeurs $x_1, x_2, x_3, \ldots, x_k$, si la série est discrète , ou les centres de chaque classe, si la série est continue.

1.1.2 Médiane

Définition :
La médiane M_e est un paramètre de position
Il permet de couper la population étudiée en deux groupes contenant le même nombre d'individus.
Par définition 50 % de la population étudiée a des valeurs inférieure à M_e et 50 % une valeur supérieure à la M_e.

Ce paramètre est utile pour donner la répartition du caractère étudié.

1.1.3 Variance et écart type

Variance :
Pour calculer la variance V d'une série statistique on utilise la formule :
$V = \dfrac{\sum_{i=1}^{k} n_i (x_i - \overline{x})^2}{\sum_{i=1}^{k} n_i} = \dfrac{n_1 (x_1 - \overline{x})^2 + n_2 (x_2 - \overline{x})^2 + \ldots + n_k (x_k - \overline{x})^2}{n_1 + n_2 + \ldots + n_k}$

Pour calculer la variance , il faut donc calculer d'abord la moyenne.

Ecart-type
L'écart-type est le nombre noté σ tel que :$\sigma = \sqrt{V}$

1.2 les Quartiles

Définitions :
On appelle **premier quartile** d'une série la plus petite valeur Q_1 de la série
pour laquelle au moins un quart (25%) des données sont inférieures ou égales à Q_1.

On appelle **troisième quartile** d'une série la plus petite valeur Q_3 de la série
pour laquelle au moins trois quarts (75%) des données sont inférieures ou égales à Q_3.

On appelle **intervalle interquartile** l'intervalle $[Q_1; Q_3]$.

On appelle **écart interquartile** l'amplitude de l'intervalle $[Q_1; Q_3]$ c'est-à-dire le nombre $Q_3 - Q_1$.

1.3 Déciles

Définitions :
On appelle **premier décile** d'une série la plus petite valeur D_1 des termes de la série
pour laquelle au moins un dixième (10%) des données sont inférieures ou égales à D_1.

On appelle **neuvième décile** d'une série la plus petite valeur D_9 des termes de la série
pour laquelle au moins neuf dixièmes (90%) des données sont inférieures ou égales à D_9.

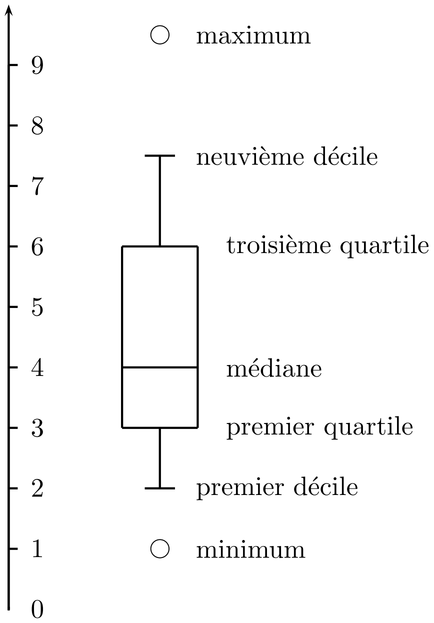
On appelle **intervalle interdécile** l'intervalle $[D_1; D_9]$.

On appelle **écart interdécile** l'amplitude de l'intervalle $[D_1; D_9]$, c'est-à-dire le nombre $D_9 - D_1$.

1.4 Diagramme en boîte

Ce type de diagramme est aussi appelé diagramme de Tuckey, boîte à moustaches ou boîte à pattes.

Il utilise le 1er et le 3ème quartile, les valeurs extrêmes, le 1er et le 9ème décile et éventuellement la médiane d'une série.



Le corps du diagramme, c'est-à-dire la boîte est formée d'un rectangle ayant pour extrémité inférieure le 1er quartile et pour extrémité supérieure le 3ème quartile.
A l'intérieur de ce rectangle on pourra tracer un segment représentant la médiane.

La largeur du rectangle n'est pas fixée, elle sera choisie de façon à obtenir un graphique harmonieux.
Ce rectangle représente les données contenues dans l'intervalle interquartile.

On repère ensuite les hauteurs correspondant au 1er et au 9ème décile,
et on trace deux pattes représentant les données contenues dans l'intervalle interdécile.

(la largeur des pattes n'a pas d'importance).

On peut ensuite terminer le graphique, en faisant figurer par des points les données qui sont en dehors de l'intervalle interdécile.

Si certaines données, sont manifestement très éloignées, on ne les représentera pas, mais on écrira leurs valeurs au dessous du diagramme.

Remarques :
Une boîte avec des pattes courtes indique que la série est assez concentrée autour de sa médiane.
Au contraire des pattes longues indique que la série est assez dispersée.

Un des avantages de cette représentation, est qu'elle nécessite très peu de calculs.
La représentation peut aussi se faire horizontalement, la graduation se trouvant alors sur l'axe horizontal, d'où l'appellation de "boîte à moustaches".
Le graphique est parfois fait en dessinant des pattes correspondant au 1er et au 99ème centile, ou même aux valeurs extrêmes.

1.5 Exercice

On considère la série suivante :

$$3, 3, 3, 4, 5, 5, 5, 6, 6, 7, 8, 8, 9, 9, 9, 9, 9, 10, 10, 11, 11, 11, 12, 13, 13, 13, 14, 14, 15, 15, 17,$$

1. Déterminer la moyenne et l'écart type de cette série.
2. Déterminer la médiane et les quartiles de cette série.
3. Déterminer le premier et le neuvième décile de cette série.
4. construire la boîte à moustache de cette série.

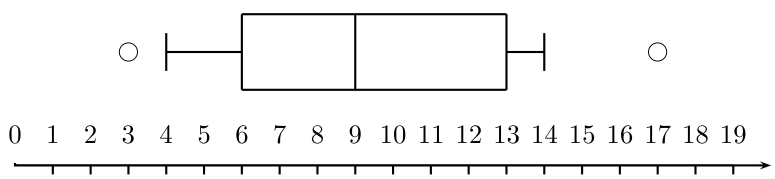
Solution :

1. $\overline{x} \approx 9,26$ et $\sigma \approx 3,85$.
2. L'effectif total est N=31 et 50% de 31 = 15,5 donc la médiane est la 16^{me} valeur de cette série ordonnée de manière croissante donc la médiane est $M_e = 9$.

   25% de 31 = 7,75 donc le premier quartile est la $8^{ième}$ valeur de cette série ordonnée de manière croissante donc le premier quartile est $Q_1 = 6$.

   75% de 31 = 23,25 donc le troisième quartile est la $24^{ième}$ valeur de cette série ordonnée de manière croissante donc le troisième quartile est $Q_3 = 13$.
3. 10% de 31 = 3,1 donc le premier décile est la $4^{ième}$ valeur de cette série ordonnée de manière croissante donc le premier décile est $D_1 = 4$.

   90% de 31 = 27,9 donc le neuvième décile est la $28^{ième}$ valeur de cette série ordonnée de manière croissante donc le neuvième décile est $D_9 = 14$.
4. On en déduit le diagramme suivant :



2 Adéquation à une loi équirépartie

2.1 Problématique

Exemple de problème :
Comment vérifier qu'un dé n'est pas truqué ?

Commentaires :

Lorsqu'on lance un grand nombre de fois un dé ,les nombres d'apparition des différentes faces ne sont pas en général égaux et **on peut donc se demander si la variation observée d'une face à l'autre est une variation « acceptable »(le dé sera considéré comme non truqué) ou « non acceptable »(le dé sera considéré comme truqué).**

2.2 Répartition des données

On lance donc un dé n fois où n est un grand nombre.On obtient alors la réparttion suivante :

Numéros de la face	1	2	3	4	5	6
Effectifs	n_1	n_2	n_3	n_4	n_5	n_6

Ici la répartition des fréquences expérimentales est donc :

Numéros de la face	1	2	3	4	5	6
Fréquences	f_1	f_2	f_3	f_4	f_5	f_6

où $f_i = \dfrac{n_i}{n}$.

La loi est équirépartie lorsque les fréquences théoriques[1] d'apparition de chacune des 6 faces valent toutes $\dfrac{1}{6}$:

Numéros de la face	1	2	3	4	5	6
Fréquences théoriques	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Pour savoir si la variation des fréquences observées d'une face à l'autre est une variation « *acceptable* » ,il s'agit de déterminer si les fréquences f_i sont suffisamment voisines de $\frac{1}{6}$.

Il faut donc pouvoir mesurer l'écart entre la distribution des fréquences observées et la distribution des fréquences théoriques.

2.3 Mesure de l'écart entre les 2 distributions

2.3.1 L'indicateur d^2_{obs}

On introduit l'indicateur suivant :

$$d^2_{obs} = \left(f_1 - \frac{1}{6}\right)^2 + \left(f_2 - \frac{1}{6}\right)^2 + \ldots + \left(f_6 - \frac{1}{6}\right)^2$$

Observons quelques propriétés de celui-ci :

1. Expliquez pourquoi les deux propriétés sont équivalentes :
   – Chaque f_i est voisin de $\frac{1}{6}$.
   – d^2_{obs} est voisin de 0.
2. Expliquez pourquoi lorsque n augmente, d^2_{obs} diminue.(On rappelle que n est la taille de l'échantillon[2].)

2.3.2 L'exemple

Après 10000 lancers on a obtenu la répartion suivante :

Numéro de la face	1	2	3	4	5	6
Effectifs observés	2066	1400	1650	1866	1534	1484
Fréquences observées	$0,2066$	$0,14$	$0,165$	$0,1866$	$0,1534$	$0,1484$

On peut alors vérifier que $d^2_{obs} \approx 0,0032$.

[1]en fait il s'agit des probabilités
[2]c-à-d le nombre de lancers effectués

2.4 Critères de décision

La question qui se pose est :

L' indicateur d^2_{obs} est-il suffisamment voisin de 0 pour considérer le dé comme non truqué ou encore peut-on dire que sa loi est bien en adéquation avec la loi équirépartie ?

On voudrait savoir si ce dé se comporte comme la plupart des dés équilibrés.

2.4.1 Fixation d'un seuil de décision

En raison de la fluctuation d'échantillonnage [3] ,on va chercher à comparer le d^2_{obs} à d'autres valeurs obtenues sur des échantillons issus de dés équilibrés.

On va donc considérer une simulation d'un **grand nombre d'échantillons** de n lancers d'un dé équilibré et considérer la série des d^2_{obs} obtenus.On note alors D_9 le neuvième décile provenant de cette simulation,ainsi 90% des valeurs des d^2_{obs} sont inférieures ou égales à D_9.

On peut prendre le neuvième décile comme seuil de décision,cela signifiera que les 10% des valeurs de d^2_{obs} qui seront supérieures à D_9 seront considérés comme marginales.

2.4.2 Prise de décision

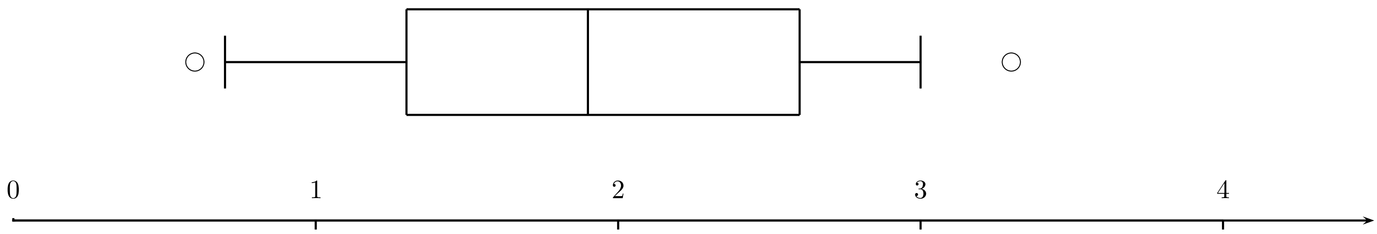
• Si $d^9_{obs} > D_9$ on déclare que le dé est truqué au risque de 10%.
Ainsi la série n'est pas déclarée équirépartie avec une marge d'erreur de 10 %.

• Si $d^9_{obs} \leqslant D_9$ on déclare que le dé est équilibré au risque de 10%.
Ainsi la série est déclarée équirépartie avec une marge d'erreur de 10 %.

Remarque : on pourrait fixer d'autres seuils de décisions.
- Pour un risque de 25% on utiliserait le 3^{ieme} quartile.
- Pour un risque de 1% on utiliserait le 99^{ieme} centile.
- ... etc.

2.4.3 L'exemple

Précédemment on a vu que $d^2_{obs} \approx 0,0032$.
Pour plus de lisibilité on multiplie toutes les valeurs des d^2_{obs} par 10 000.
Un étude statistique sur les valeurs des nombres « $10\,000 d^2_{obs}$ » a permis d'obtenir le diagramme de Tuckey suivant :



Donc le neuvième décile des « 10 $000d^2_{obs}$ » est 3 donc en divisant par 10 000,le neuvième décile des d^2_{obs} est $D_9 = 0,003$ c-à-d que 90% des d^2_{obs} pour un dé équilibré sont inférieurs ou égals à $0,003$.
Or ici $d^2_{obs} > D_9$,on peut donc déclarer le dé non équilibré au risque de 10 %.
Ainsi ce dé peut être considéré comme truqué avec une marge d'erreur de 10%.

2.5 Exercices

2.5.1 Exercice 1 :Extrait de l'épreuve de juin 2003

Les guichets d'une agence bancaire d'une petite ville sont ouverts au public cinq jours par semaine : les mardi, mercredi, jeudi, vendredi et samedi.

Le tableau ci-dessous donne la répartition journalière des 250 retraits d'argent liquide effectués aux guichets une certaine semaine.

[3] c-à-d de la variation des résulttas lorsque l'on recommence une série ,appelée souvent un échantillon de n lancers

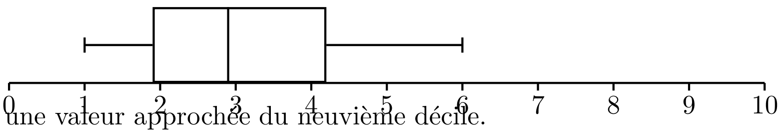
Jour de la semaine	mardi	mercredi	jeudi	vendredi	samedi
Rang i du jour	1	2	3	4	5
Nombre de retraits	37	55	45	53	60

On veut tester l'hypothèse :le nombre de retraits est indépendant du jour de la semaine. On suppose donc que le nombre des retraits journaliers est égal à $\frac{1}{5}$ du nombre des retraits de la semaine.

On pose $d^2_{obs} = \displaystyle\sum_{i=1}^{5}\left(f_i - \frac{1}{5}\right)^2$ où f_i est la fréquence des retraits du i-ème jour.

1. Calculer les fréquences des retraits pour chacun des cinq jours de la semaine.
2. Calculer alors la valeur de $1\,000 d^2_{obs}$ (la multiplication par 1 000 permet d'obtenir un résultat plus lisible).
3. En supposant qu'il y a équiprobabilité des retraits journaliers, on a simulé 2 000 séries de 250 retraits hebdomadaires.

   Pour chaque série, on a calculé la valeur du $1\,000 d^2_{obs}$ correspondant. On a obtenu ainsi 2 000 valeurs de $1\,000 d^2_{obs}$. Ces valeurs ont permis de construire le diagramme en boîte ci-dessous où les extrémités des "pattes " correspondent respectivement au premier décile et au neuvième décile.

   Lire sur le diagramme une valeur approchée du neuvième décile.
4. En argumentant soigneusement la réponse, dire si pour la série observée au début, on peut affirmer, avec un risque d'erreur inférieur à 10 %, que le nombre de retraits est indépendant du jour de la semaine ?

2.5.2 Exercice 2 :Extrait de l'épreuve de mars 2003 à Pondichéry

Un pisciculteur possède un bassin qui contient trois variétés de truites : communes, saumonées et arc-en-ciel. Il voudrait savoir s'il peut considérer que son bassin contient autant de truites de chaque variété. Pour cela il effectue, au hasard, 400 prélèvements d'une truite avec remise et obtient les résultats suivants :

Variété	**Commune**	**Saumonée**	**Arc-en-ciel**
Effectifs	146	118	136

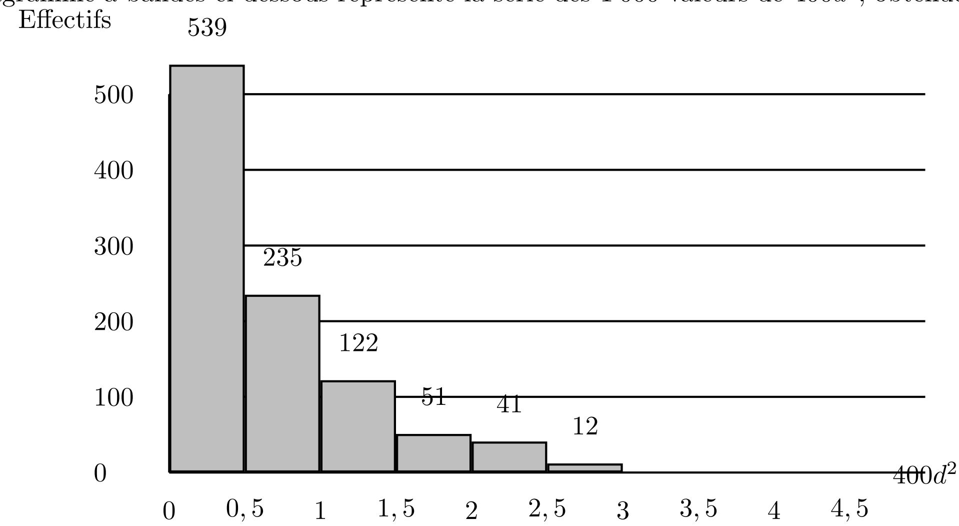
1. (a) Calculer les fréquences de prélèvement f_c d'une truite commune, f_s d'une truite saumonée et f_a d'une truite arc-en-ciel. On donnera les valeurs décimales exactes.

   (b) On pose $d^2 = \left(f_c - \frac{1}{3}\right)^2 + \left(f_s - \frac{1}{3}\right)^2 + \left(f_a - \frac{1}{3}\right)^2$.

   Calculer $400d^2$ arrondi à 10^{-2} ; on note $400d^2_{obs}$ cette valeur.

   À l'aide d'un ordinateur, le pisciculteur simule le prélèvement au hasard de 400 truites suivant la loi équirépartie. Il répète 1 000 fois cette opération et calcule à chaque fois la valeur de $400d^2$.

   Le diagramme à bandes ci-dessous représente la série des 1 000 valeurs de $400d^2$, obtenues par simulation.

2. Déterminer une valeur approchée à 0,5 près par défaut, du neuvième décile D9 de cette série.
3. En argumentant soigneusement la réponse dire si on peut affirmer avec un risque d'erreur inférieur à 10 % que « le bassin contient autant de truites de chaque variété ».
4. On considère désormais que le bassin contient autant de truites de chaque variété. Quand un client se présente, il prélève au hasard une truite du bassin.

   Trois clients prélèvent chacun une truite. Le grand nombre de truites du bassin permet d'assimiler ces prélèvements à des tirages successifs avec remise.

   Calculer la probabilité qu'un seul des trois clients prélève une truite commune.

3 Statistiques à 2 variables.

3.1 Point moyen - Covariance

Sur une population donnée, étudions deux caractères.

Pour chacun des n individus de cette population, notons x_i et y_i les valeurs prises par chacun de ces caractères, et présentons les données à l'aide de la série statistique à deux variable suivante :

Valeur x_i	x_1	x_2	...	x_n
Valeur y_i	y_1	y_2	...	y_n

3.1.1 Nuage de points - Point moyen

Définition :
Dans un repère orthogonal, l'ensemble des points M_i de coordonnées $(x_i; y_i)$ (avec $1 \leqslant i \leqslant n$) est appelé **le nuage de points** associé à cette série statistique à deux variables.

Notons $\overline{x} = \frac{1}{n}(x_1 + x_2 + \ldots + x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i$ et $\overline{y} = \frac{1}{n}(y_1 + y_2 + \ldots + y_n) = \frac{1}{n}\sum_{i=1}^{n} y_i$

Ainsi $\overline{x}$ et $\overline{y}$ représentent respectivement la moyenne des séries (x_i) et (y_i).

Définition :
Le point G de coordonnées $(\overline{x}; \overline{y})$ est appelé **le point moyen du nuage** de points associé à cette série statistique à deux variables.

Obtention des coordonnées du point moyen grâce à la calculatrice :

Texas Instrument (TI - 80) :

STAT **1 : Edit** ... *permet d'entrer les valeurs de x dans L1, puis celles de y dans L2*

STAT **CALC 2 : 2-VAR Stats** puis (2^{nd} L1, 2^{nd} L2)**ENTER***(Ceci nous donne $\overline{x}$ et $\overline{y}$).*

Casio Graph 25

Dans le menu **STAT**, entrer les valeurs de x dans **List 1**, puis celles de y dans **List 2**.

CALC

SET , *entrer dans* **2VarXList : List 1** et **2VarYList : List 2**

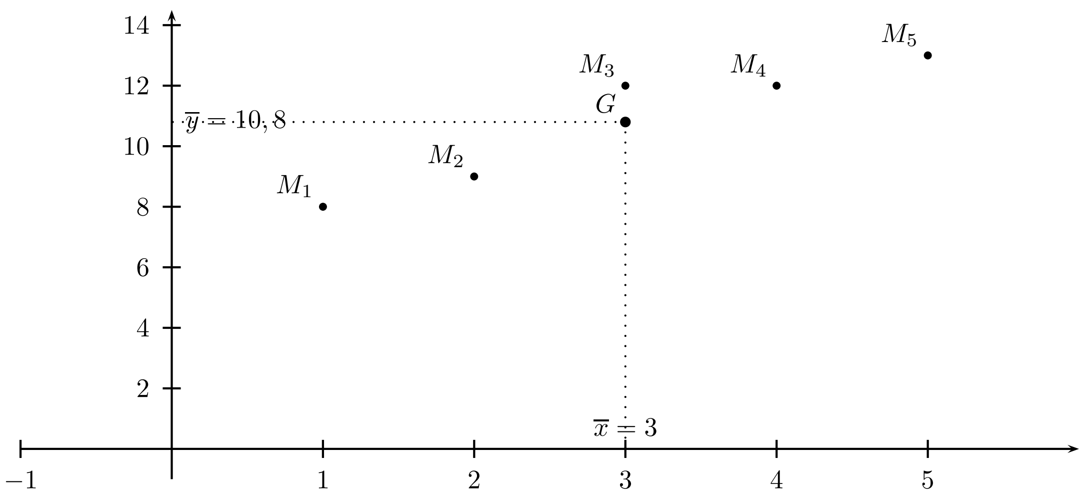
EXE puis **Calc 2-Var** *(On obtient alors $\overline{x}$ et $\overline{y}$.)*

Exemple :

La série statistique double suivante indique les notes mensuelles d'un élève au cours des cinq premiers mois de l'année scolaire numérotés de 1 à 5.

Valeur $Moisx_i$	1	2	3	4	5
Note y_i	8	9	12	12	13

$\overline{x} = 3$ et $\overline{y} = 10,8$ donc le point moyen G du nuage représenté ci-dessous a pour coordonnées $(3; 10,8)$.



3.1.2 variance - covariance

Pour étudier la dispersion de chaque variable x et y, on peut calculer leurs variances : $V_x = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i - \overline{x})^2$ et $V_y = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(y_i - \overline{y})^2$.

Mais il est utile d'introduire une quantité qui fasse intervenir à la fois les valeurs de x et de y.

Définition :

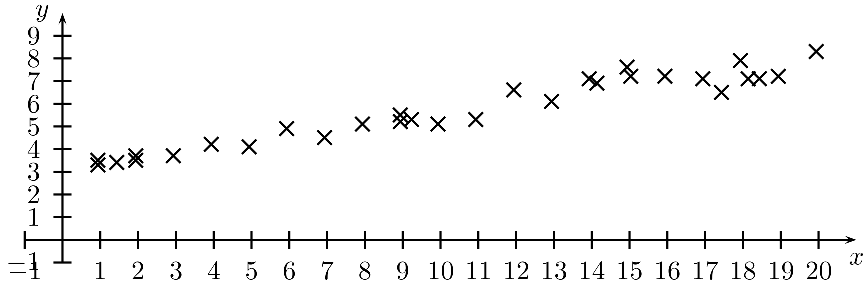
On appelle **covariance de x et y** le nombre : $C_{xy} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = \left(\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i y_i\right) - \overline{xy}.$

La seconde expression est plus commode pour les calculs à la main.

Dans l'exemple précédent, $C_{xy} = \frac{1}{5}(1 \times 8 + 2 \times 9 + 3 \times 12 + 4 \times 12 + 5 \times 13) - 3 \times 10,8 = 35 - 32,4 = 2,6$.

3.2 Ajustement affine par la méthode des moindres carrés

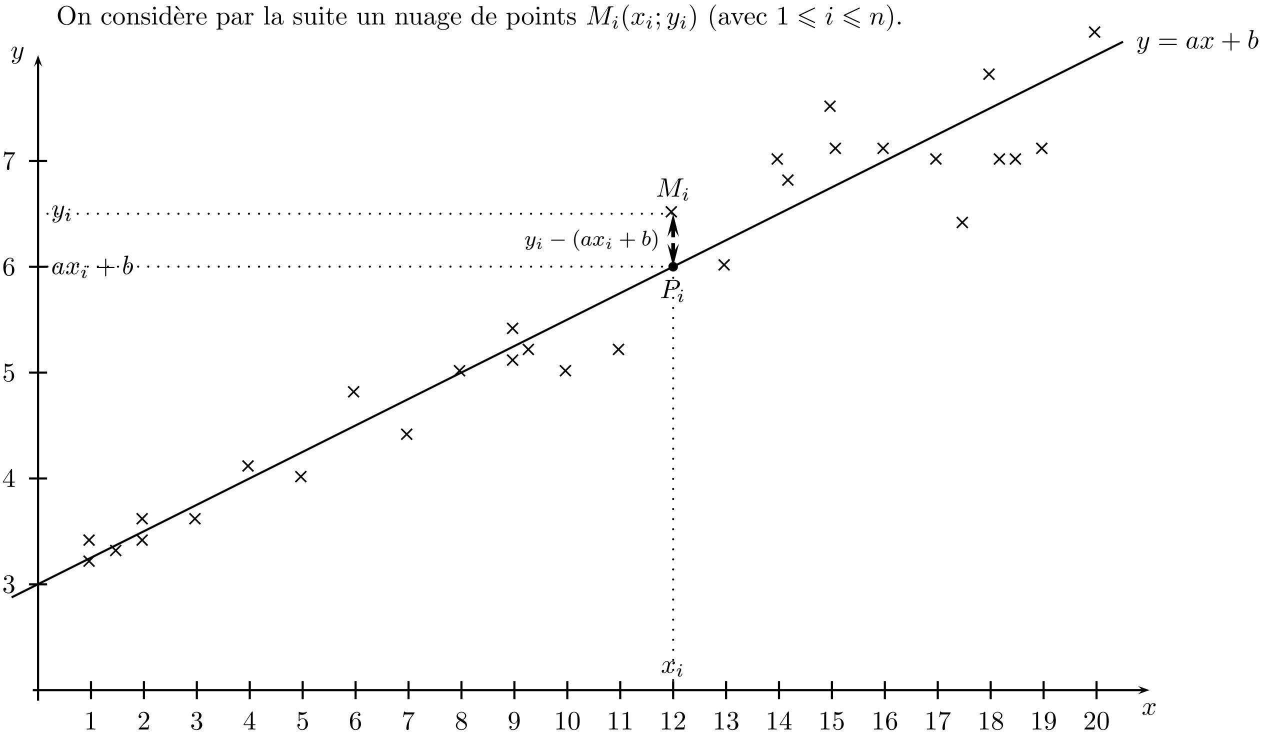
Lorsque **les points du nuage paraissent presque alignés,** on peut chercher une relation de la forme $y = ax + b$ qui exprime de façon approchée y en fonction de x, autrement dit, une fonction affine f telle que l'égalité $y = f(x)$ s'ajuste au mieux avec les données.



Graphiquement, cela signifie qu'on cherche une droite qui passe au plus près de tous les points du nuage.

Une telle relation permettrait notamment de faire des prévisions. Pour mesurer la qualité d'une telle formule, on considère, pour chaque valeur x_i, la différence entre la valeur observée, c'est à dire y_i, et la valeur calculée par la formule, c'est à dire $ax_i + b$. On souhaite que toutes les différences : $y_i - (ax_i + b)$ appelées **erreurs,** ou **résidus**, ou **perturbations**, soient **les plus petites possible**.La méthode la plus couramment employée, dite **méthode des moindres carrés**, consiste à choisir a et b de façon que la **somme des carrés des résidus soit la plus petite possible.**

On considère par la suite un nuage de points $M_i(x_i; y_i)$ (avec $1 \leqslant i \leqslant n$).



Définition :
Il existe une droite unique associée au nuage $M_i(x_i; y_i)$, avec $i = 1, \dots, n$ telle que la somme des ${M_iP_i}^2$ soit minimale.
- Cette droite passe par le point moyen $G(\overline{x}, \overline{y})$ du nuage.
- Elle a une équation du type $y = ax + b$ avec $a = \frac{C_{xy}}{V_x}$ et $b = \overline{y} - a\overline{x}$.
- Cette droite s'appelle **la droite de régression** de y en x.

Utilisation de la calculatrice pour la détermination de l'équation de la droite de régression
Texas Instrument (TI - 80) :
STAT **1 : Edit** … *permet d'entrer les valeurs de x dans L1, puis celles de y dans L2*
STAT **CALC 3 : LINREG(aX+b)** *puis* (2^{nd} L1, 2^{nd} L2)**ENTER***(Ceci nous donne a et b.*
Casio Graph 25
Dans le menu **STAT**, entrer les valeurs de x dans **List 1**, puis celles de y dans **List 2**.
CALC
SET , *entrer dans* **2VarXList : List 1** et **2VarYList : List 2**
EXE puis **Calc Reg 1-Linear** *(On obtient alors a et b.)*

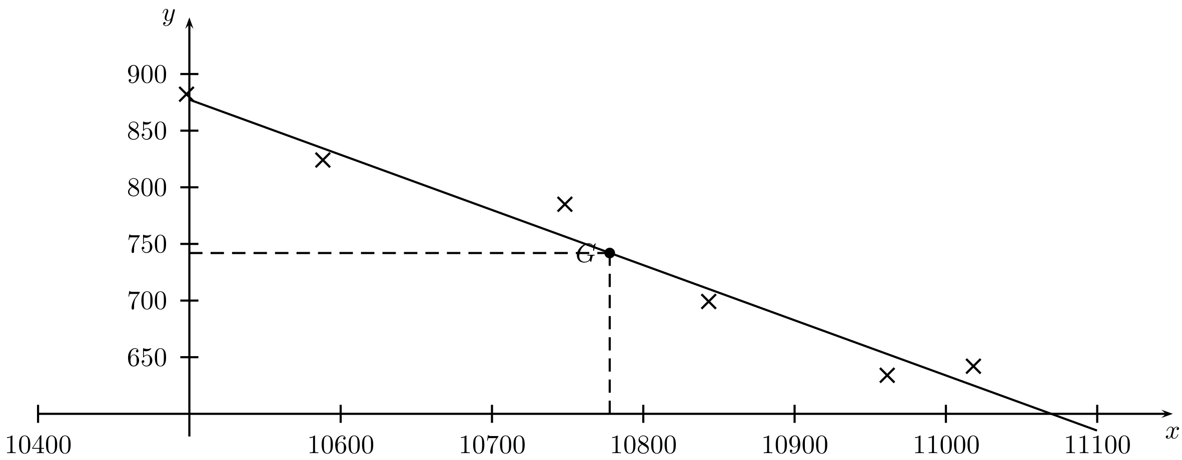
Exercice : Considérons la série statistique à deux variables $(x_i; y_i)$, pour $i = 1, 2, \dots, 6$:

x_i	10 500	10 590	10 750	10 845	10 963	11 020
y_i	880	822	783	697	632	640

1. Placer, dans un repère orthogonal, le nuage de points et constater que la forme "allongée" du nuage justifie un ajustement affine.
2. Placer le point moyen G de ce nuage.
3. Tracer la droite d de régression de y en x.

Solution :

1. Représentation du nuage de points :



2. Le point moyen G a pour coordonnées $(\overline{x}; \overline{y})$.On obtient grâce à la calculatrice : $\overline{x} = 10778$ et $\underline{y} = 742,3$.
3. La droite d passe par G ;pour la tracer, il suffit de connaître son équation $y = ax + b$:d'après la calculatrice, $a \approx -0,487$ et $b \approx 5991$.

3.3 Exemples d'ajustements non affines

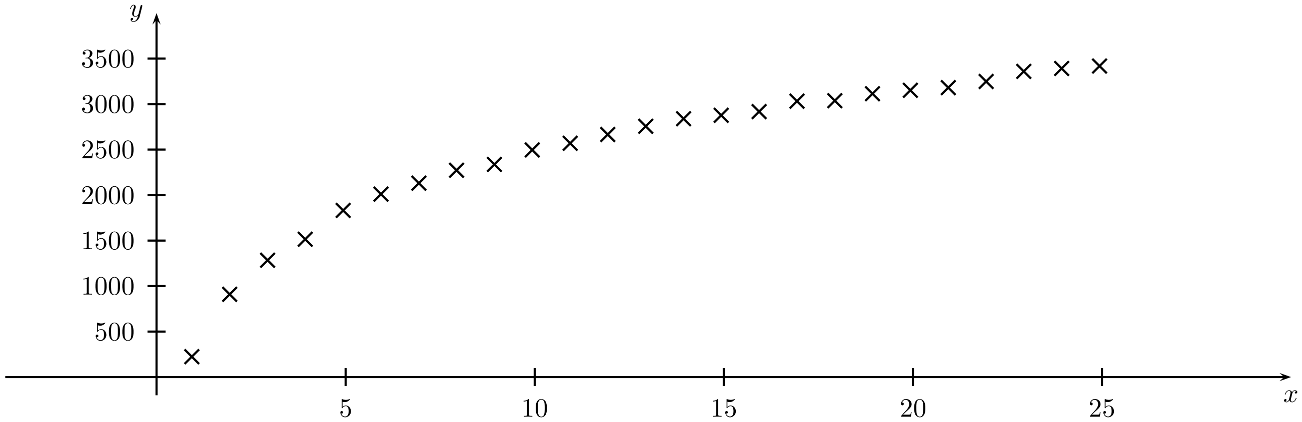
Il existe une multitude d'ajustements non affines,nous en présenterons deux des plus courants en Terminale ES.

3.3.1 Ajustement logarithmique

On considère les données suivantes :

x_i	1	2	3	4	5	6	7	8	9	10	11	12
y_i	198	881	1256	1489	1804	1983	2104	2247	2312	2468	2541	2639

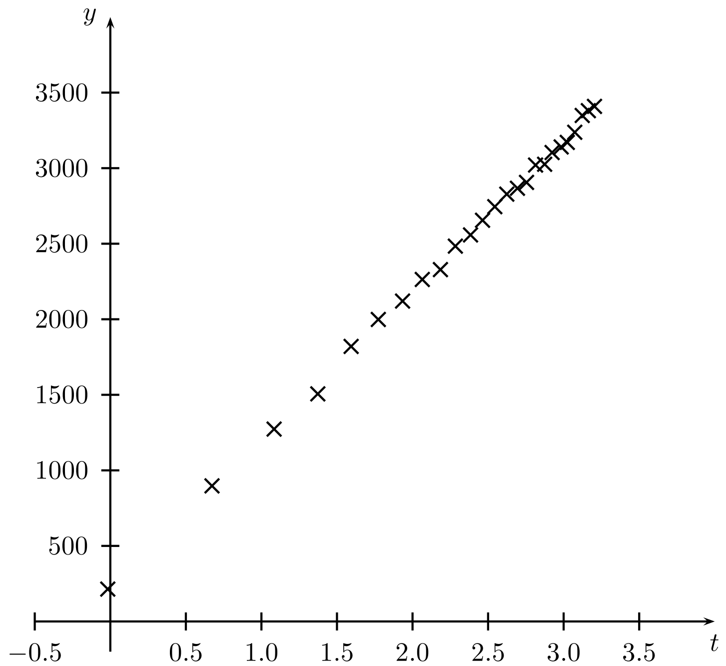
x_i	13	14	15	16	17	18	19	20	21	22	23	24	25
y_i	2728	2811	2850	2890	3005	3010	3087	3125	3155	3221	3333	3365	3392



La forme du nuage ne suggère pas un ajustement affine mais éventuellelemnt un ajustement de la forme $y = a\ln x + b$
On peut le vérifier en posant $t = \ln x$ et en plaçant les points de coordonnées $(t; y)$ dans un nouveau repère :

$t_i = \ln x_i$	0	0,69	1,1	1,39	1,61	1,79	1,95	2,08	2,20	2,30	2,40	2,48
y_i	198	881	1256	1489	1804	1983	2104	2247	2312	2468	2541	2639

$t_i = \ln x_i$	2,56	2,64	2,71	2,77	2,83	2,89	2,94	3	3,04	3,09	3,14	3,18	3,22
y_i	2728	2811	2850	2890	3005	3010	3087	3125	3155	3221	3333	3365	3392



Ces points sont presque alignés, ce qui permet d'envisager un ajustement affine du type $y = at + b$.

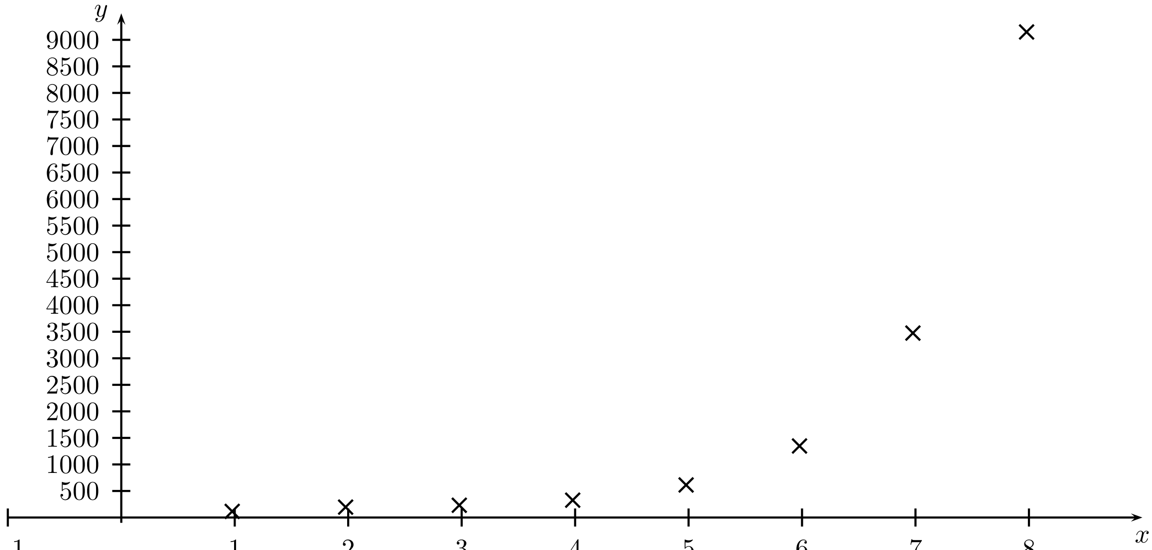
Par la méthode des moindres carrées on obtient à la calculatrice $a \approx 989$ et $b \approx 180$ doù $y = 989t + 180$ (en arrondissant les coefficients à l'unité)

Comme $t = lnx$, $\boxed{\text{on obtientl'ajustement logarithmique : } y = 989\ln x + 180 .}$

3.3.2 Ajustement exponnetielle

On considère les données suivantes :

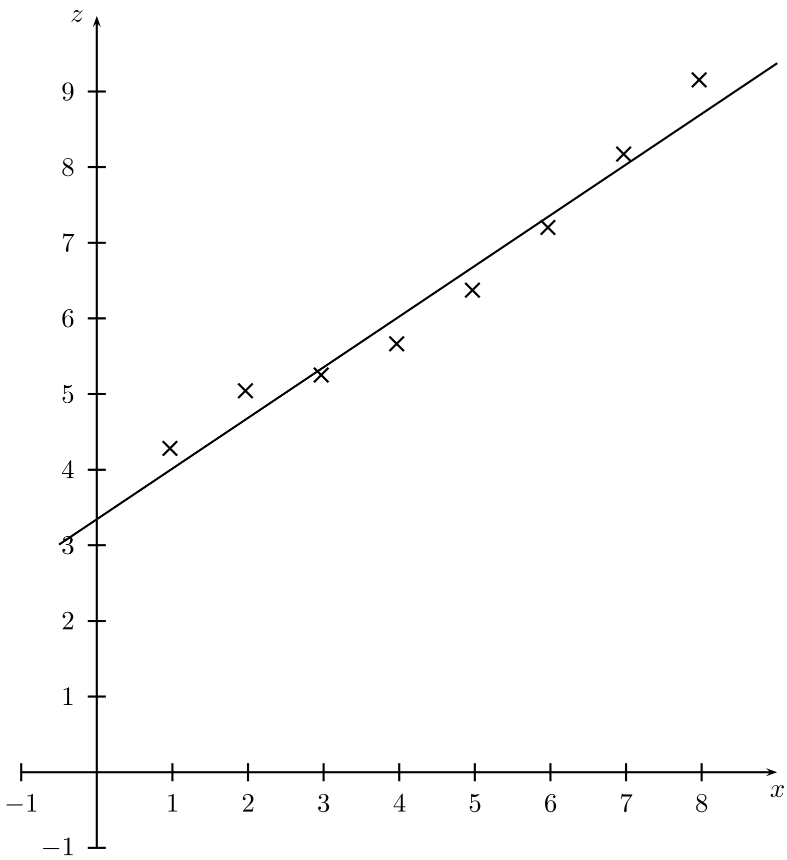
x_i	1	2	3	4	5	6	7	8
y_i	70	150	185	280	565	1300	3430	9100



La forme du nuage ne suggère pas un ajustement affine .On pose $z = \ln y$ pour obtenir des données en oronnées moins grandes.

x_i	1	2	3	4	5	6	7	8
$z_i = \ln y_i$	4,25	5,01	5,22	5,63	6,34	7,17	8,14	9,12

On place les points de coordonnées $(x; z)$ dans un nouveau repère :



Ces points sont presque alignés, ce qui permet d'envisager un ajustement affine du type $z = ax + b$.

Par la méthode des moindres carrées on obtient à la calculatrice $a \approx 0,670$ et $b \approx 3,344$ d'foù $z = 0,670x + 3,344$ (en arrondissant les coefficients au millième)

Comme $z = \ln y$ donc $e^z = e^{\ln y}$ donc $y = e^z$ d'où un ajustement exponnetielle du type : $y = e^{0,670x+3,344} = e^{0,670x}e^{3,344} = e^{3,344}e^{0,670x} = 28,332e^{0,670x}$.

Ainsi $\boxed{y = Ae^{Bx} \text{ avec } A \approx 28,332 \text{ et } B \approx 0,670.}$