

Statistiques à deux variables

I. Série statistique à deux variables

La statistique descriptive à deux variables a pour but de mettre en évidence une relation éventuelle qui peut exister entre **deux variables** d'une population, considérées **simultanément**.

1) Définition

Définition

On appelle série statistique à deux variables X et Y, le relevé simultané de la valeur de deux caractères statistiques X et Y. Elle est donc constituée d'une liste de couples de nombres $(x_i ; y_i)$.

2) Nuage de points, point moyen

Définition

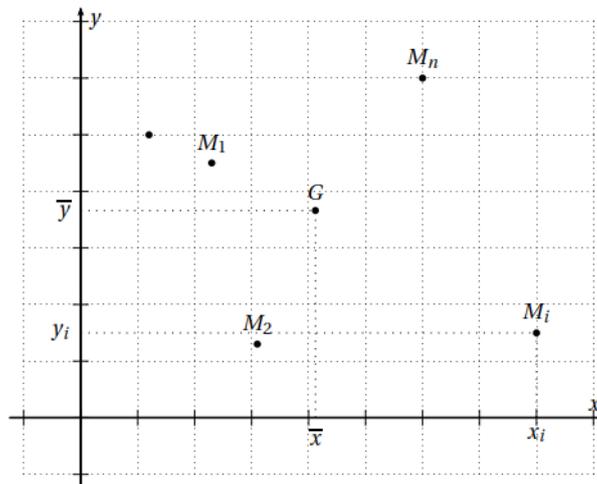
Le plan étant muni d'un repère orthogonal, on peut associer à chaque couple $(x_i ; y_i)$ de la série statistique le point M de coordonnées $(x_i ; y_i)$.

Le graphique ainsi obtenu constitue un **nuage de points**.

Définition

Le point moyen G du nuage de points est le point de coordonnées $(\bar{x} ; \bar{y})$ c'est-à-dire le point dont :

- L'abscisse est la moyenne de la série (x_i) : $\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$
- L'ordonnée est la moyenne de la série (y_i) : $\bar{y} = \frac{y_1 + y_2 + \dots + y_n}{n}$



II. Ajustement

1) Corrélation

Définition

Il y a corrélation entre deux variables X et Y observées sur les individus d'une même population lorsqu'il y a une relation entre X et Y.

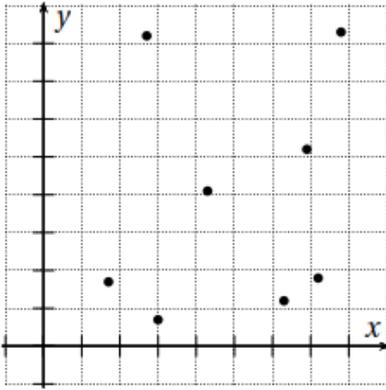
REMARQUE

L'existence d'une corrélation entre deux variables peut être décelée à l'aide d'un nuage de points.

Exemple

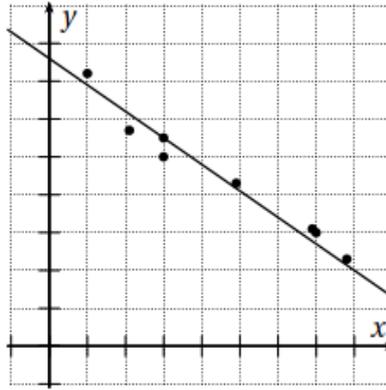
Considérons les diagrammes suivants :

Diagramme A



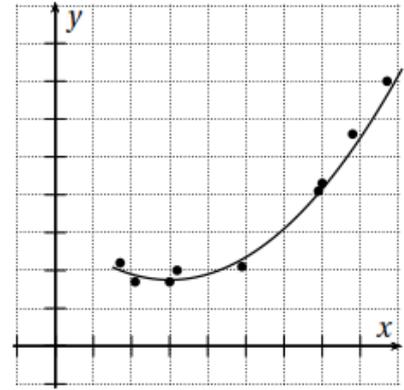
Absence de corrélation entre X et Y

Diagramme B



Corrélation entre X et Y

Diagramme C



Corrélation entre X et Y

2) Ajustement de y en x

Définition

Lorsque les valeurs de x sont connues, effectuer un ajustement de y en x d'un nuage de points consiste à trouver une fonction dont la courbe représentative d'équation $y = f(x)$ est la plus « proche » du nuage.

REMARQUE

L'existence d'une corrélation entre deux variables peut être décelée à l'aide d'un nuage de points.

- Un ajustement permet de faire des estimations : *interpolation* (dans l'intervalle d'étude) et *extrapolation* (en dehors).
- Lorsque les points du nuage sont *presque alignés*, comme pour le diagramme B, on recherche une droite qui passe le plus près possible des points. On effectue alors un *ajustement affine*.
- On verra qu'il existe des ajustements qui ne sont pas affines, comme sur le diagramme C.
- Lorsque les valeurs de y sont connues, on effectue un ajustement de x en y du nuage (on cherche une fonction g dont la courbe d'équation $x = g(y)$ est la plus proche du nuage). C'est le contexte qui donne le type d'ajustement.

3) Ajustement affine

Une droite d'ajustement affine est une droite qui passe au plus près du nuage de points.

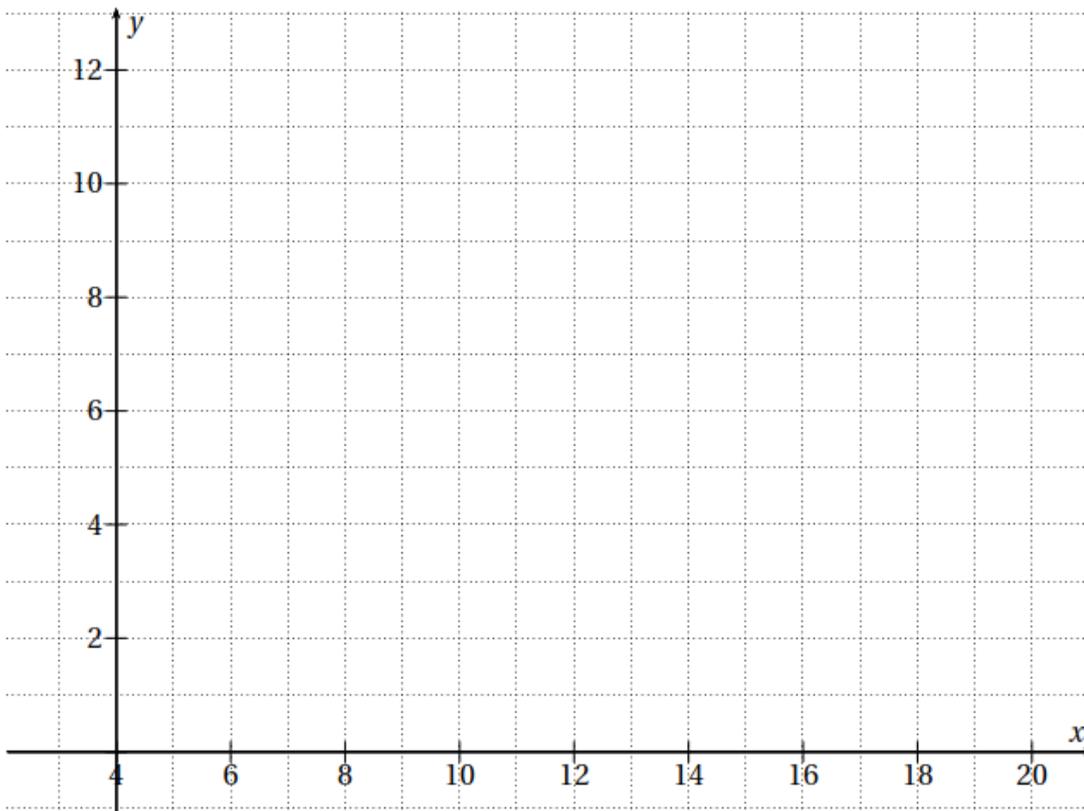
Ajustement au jugé

Le gérant d'un hypermarché, disposant d'un potentiel de vingt-huit caisses enregistreuses, a fait réaliser une étude statistique sur le temps moyen (en minutes) d'attente d'un client à la caisse.

On note x_i le nombre de caisses ouvertes, y_i le temps moyen correspondant en minutes.

x_i	4	5	6	7	8	9	10	11	12	13
y_i	12,25	12	11,5	11,75	10	10	9,75	9	8,25	8

- 1) (a) Construire dans le repère de la figure ci-dessous le nuage de points associé à la série.



- (b) Calculer les coordonnées du point moyen G du nuage. Placer G dans le repère.
- 2) Peut-on penser à un ajustement affine ? Justifier.
- 3) On propose comme droite d'ajustement affine la droite D qui a pour équation $y = -0,5x + 14,5$.
- (a) Tracer la droite D sur le graphique.
- (b) Le point G appartient-il à la droite D ? Justifier par un calcul.
- (c) En utilisant cette droite :
- déterminer le temps moyen d'attente d'un client à la caisse lorsque 20 caisses sont ouvertes
 - déterminer le nombre de caisses à ouvrir pour que le temps moyen d'attente d'un client à la caisse soit de trois minutes.

Ajustement par la méthode des moindres carrés

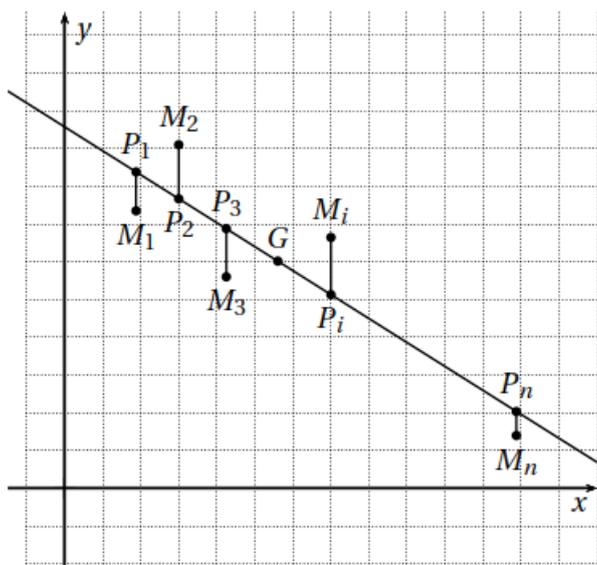
On connaît les valeurs x_i , on cherche à obtenir une droite d'ajustement dont les valeurs y sont les plus proches possibles des y_i « verticalement ».

Les points M_1, M_2, \dots, M_n sont les points du nuage.

Les points P_1, P_2, \dots, P_n sont les points d'une droite D de mêmes abscisses que, respectivement,

M_1, M_2, \dots, M_n , d'équation $y = ax + b$ qui est telle que la somme $S = M_1P_1^2 + M_2P_2^2 + \dots + M_nP_n^2$ soit minimale.

On admet qu'une telle droite existe toujours et on dit que cette droite réalise un ajustement affine du nuage de y en x par la méthode dite des moindres carrés. Elle passe par le point moyen G du nuage.



Définition

La droite D d'équation $y = ax + b$ obtenue par la méthode des moindres carrés est appelée droite de régression de y en x .

REMARQUE

- La droite de régression de y en x minimise la somme des carrés des distances en ordonnée.
- La droite de régression de x en y minimise la somme des carrés des distances en abscisse. Elle a pour équation $x = a'y + b'$. Elle permet d'estimer les valeurs de X lorsqu'on connaît celles de Y .

Définition

Pour apprécier la qualité d'un ajustement affine, on peut déterminer le coefficient de corrélation linéaire, noté r .

On admet que :

- r est toujours compris entre -1 et 1
- r est positif lorsque la droite d'ajustement monte et négatif lorsqu'elle descend
- plus r est proche de -1 ou 1 et meilleure est la qualité de l'ajustement.

Remarque

On détermine a , b et r avec la calculatrice ou le tableur.

Exemple

On reprend l'exemple du gérant d'hypermarché.

- 1) Avec la calculatrice, déterminer l'équation de la droite d'ajustement Δ obtenue par la méthode des moindres carrés. On arrondira a et b à 10^{-2} .
- 2) Déterminer le coefficient de corrélation r , arrondi à 10^{-2} . Que peut-on penser de la qualité de cet ajustement ?

III. Utilisation de la calculatrice

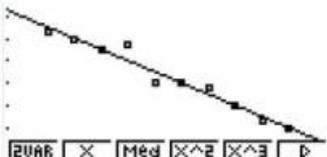
On peut retrouver tous ces paramètres statistiques en utilisant les listes d'une calculatrice.

Calculatrice Casio

- Effacer les données précédentes : Dans le menu **STAT**, DELA (F6), YES et EXE.
- Saisir les valeurs de X dans la 1^{ère} colonne (LIST1) et celles de Y dans la 2^{ème} colonne (LIST2).

SUB	List 1	List 2	List 3	List 4
8	11	9		
9	12	8.25		
10	13	8		
11				

```
StatGraph1
Graph Type : Scatter
XList      : List1
YList      : List2
Frequency  : 1
Mark type  : *
```



Utilisation du graphique :

- Pour paramétrer le graphique : **GRPH** (F1), **SET** (F6). Choisir **Scatter** pour Graph Type, List1 pour XList et List2 pour YList et EXE.
- Pour afficher le nuage : **GPH1** (F1).
- Pour afficher les coordonnées du point G : **CALC** (F2), **2VAR** (F2). Faire défiler les valeurs.
- Pour afficher les résultats a, b et r, de la « régression linéaire » (ajustement affine) : **CALC** (F2), **REG** (F3) X (F1) et ax+b (F1).
- Pour dessiner la droite d'ajustement : **COPY** (F6), puis dans le menu **GRAPH**, la dessiner.

```
1Var XList : List1
1Var Freq  : 1
2Var XList : List1
2Var YList : List2
2Var Freq  : 1
```

```
2-Variable
x̄ = 8.5
Σx = 85
Σx² = 805
x̄σn = 2.87228132
x̄σn-1 = 3.02765035
n = 10
```

```
LinearReg
a = -0.5030303
b = 14.5257575
r = -0.9768932
r² = 0.95432034
MSe = 0.1249053
y = ax + b
```

Calcul des paramètres :

- Paramétrer les listes : **CALC** (F2) et **SET** (F6). Choisir List1 pour 2Var XList et List2 pour 2Var YList et List2 pour YList, puis EXE.
- Pour afficher les coordonnées du point G : **CALC** (F1) et **2VAR** (F1). Faire défiler avec la touche.
- Pour afficher les résultats a, b et r, de la « régression linéaire » (ajustement affine) : **REG** (F3) et X (F1).

Pour obtenir des estimations :

- Pour estimer la valeur de Y pour la valeur $X = x_0$: Dans le menu **RUN**, **OPTN** **STAT** (F5) \hat{y} (F2) et EXE.
- Pour estimer la valeur de X pour la valeur $Y = y_0$: Dans le menu **RUN**, **OPTN** **STAT** (F5) \hat{x} (F1) et EXE.

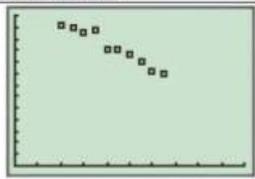
Calculatrice TI

- Effacer les données précédentes : **EDIT**, 4:EffacerList **entrer**, L1,L2 (2nde 1, 2nde 2) et **entrer**.
- Dans le menu **STATS** 1:EDIT **entrer**, saisir les valeurs de X dans L1, et celles de Y dans L2 et **quitter** (2nde Mode).

```
2nd F1 CALC TESTS
1:Edit...
2:SortA<
3:SortD<
4:ClrList
5:SetUPEditor
```

L1	L2	L3	2
8	10		
9	10		
10	9.75		
11	9		
12	8.25		
13	8		

```
2nd F1 P1ot2 P1ot3
Off
Type:
Xlist:L1
Ylist:L2
Mark: *
```



Pour utiliser le graphique :

- Pour paramétrer le graphique : **graph stats** (2nde f(x)), **Graph1** **entrer**.
- Mettre **On** en surbrillance, icône nuage en surbrillance, L1 pour Xlist et L2 pour Ylist, etc...
- Pour afficher le nuage : **Graphe**. Modifier la fenêtre graphique **fenetre** en prenant en compte les valeurs des x_i et y_i .

```
2-Var Stats
x̄=8.5
Σx=85
Σx²=805
Sx=3.027650354
σx=2.872281323
n=10
```

```
LinReg
y=ax+b
a=-.503030303
b=14.5257575
```

```
2nd F5 Y-VARS
1:window...
2:Zoom...
3:GDB...
4:Picture...
5:Statistics...
6:Table...
7:String...
```

```
XY Σ TEST PTS
1:RegEQ
2:a
3:b
4:c
5:d
6:e
7:r
```

Pour obtenir les paramètres :

- Paramétrer les listes :
- Pour afficher les coordonnées du point G : **CALC**, 2:STATS 2-Var **entrer** et L1,L2 (2nde 1, 2nde 2). Faire défiler les valeurs.
- Pour afficher les résultats a, b de la « régression linéaire » (ajustement affine) : **CALC**, 4:RéglLin(ax+b) **entrer** L1,L2 et **entrer**.
- Pour afficher le coefficient de corrélation r : **VAR**, 5:Statistiques **entrer**, mettre EQ en surbrillance 7:r **entrer** **entrer**.

Pour obtenir des estimations :

- Pour estimer la valeur de Y pour la valeur $X = x_0$: **def tbl** (2nde fenetre).
- Compléter avec la valeur de x_0 , **table** (2nde Graphe). Affichage sur la 1^{ère} ligne de la valeur de y_0 .