L1 **Statistiques - cours 1 : définitions** H. Perrot

1. Définitions : - population
- échantillon
- unité statistique
- caractère : qualitatif ou quantitatif
- variable statistique : discrète ou continue
- série statistique

2. Séries statistiques d'un caractère quantitatif discret : - série statistique
- effectif total
- effectif partiel ou fréquence absolue
- fréquence relative
- étendue de la série

3. Séries statistiques d'un caractère quantitatif continu

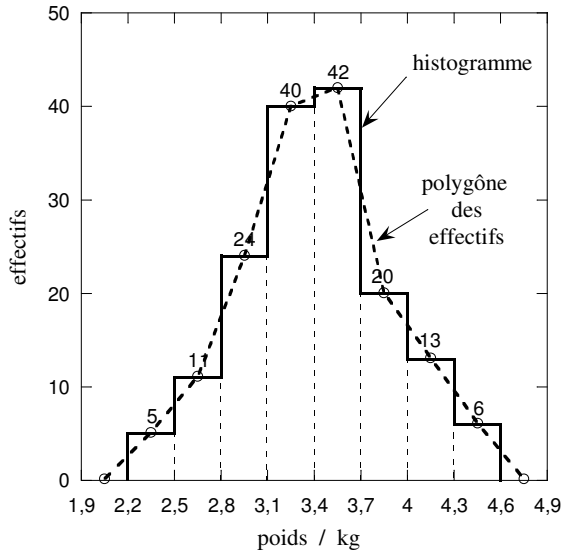
classe	1	2	3	4	5	6	7	8
limites	2,2-2,5	2,5-2,8	2,8-3,1	3,1-3,4	3,4-3,7	3,7-4,0	4,0-4,3	4,3-4,6
centre	2,35	2,65	2,95	3,25	3,55	3,85	4,15	4,45
effectif	5	11	24	40	42	20	13	6
fréq. relat. %	3,1	6,8	14,9	24,8	26,1	12,4	8,1	3,7

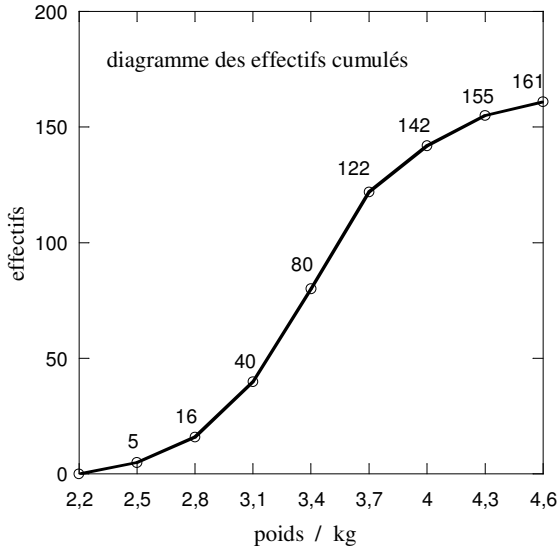
4. Séries statistiques d'un caractère qualitatif

5. Représentation graphique des séries statistiques

a) caractère discret
- diagramme en bâtons
- polygones des effectifs (fréquences)
- polygones des fréquences relatives
- diagramme cumulatif des effectifs
- diagramme cumulatif des fréquences relatives

b) caractère continu
- histogramme (classes égales, classes inégales)
- polygones des effectifs (fréquences)
- polygones des effectifs cumulés
- polygones des fréquences relatives cumulées





1. Moyenne arithmétique : $\bar{x} = \dfrac{x_1 + x_2 + \ldots + x_n}{n} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$ pour n valeurs x_i

$\bar{x} = \sum_{i=1}^{p} \dfrac{n_i}{n} x_i = \sum_{i=1}^{p} f_i x_i$ - caractère discret : p valeurs x_i, d'effectif n_i, de fréquence relative f_i

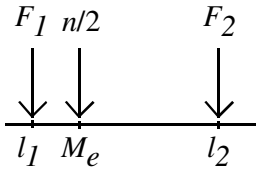
- caractère continu : p classes de centre x_i, d'effectif n_i, de fréquence relative f_i

calcul rapide : si $x_i = M + a\, z_i$ alors $\bar{x} = M + a\, \bar{z}$

2. Médiane

- on ordonne les valeurs de la série, M_e = valeur séparant la série en 2 groupes de même effectif
- série à caractère continu :

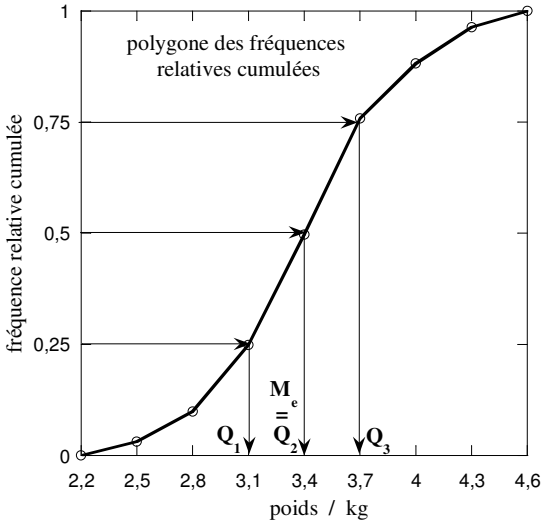
M_e dans la classe (l_1, l_2)
n effectif de la série
F_1 effectif cumulé en l_1
F_2 effectif cumulé en l_2



$$M_e = l_1 + (\frac{n}{2} - F_1)\frac{l_2 - l_1}{F_2 - F_1}$$

3. Quartiles, déciles, centiles

- on ordonne les valeurs de la série :
  - les quartiles Q_i séparent la série en 4 groupes de même effectif ($Q_2 = M_e$)
  - les déciles D_i séparent la série en 10 groupes de même effectif
  - les centiles C_i séparent la série en 100 groupes de même effectif
- M_e, Q_1, Q_2 et Q_3 sur l'exemple des nourrissons



4. Mode ou valeur dominante

- mode = valeur la plus fréquente

- série à caractère continu : mode dans la classe modale (l_1, l_2) Δ_1 (Δ_2) excédent d'effectif de la classe modale par rapport à la classe inférieure (supérieure)	$\text{mode} = l_1 + \dfrac{\Delta_1}{\Delta_1 + \Delta_2}(l_2 - l_1)$

Statistiques - cours 3 : paramètres de dispersion

1. Etendue : $x_{max} - x_{min}$ pour une série de valeurs x_i

2. Ecart moyen

- écart de x_i : $e_i = |x_i - \bar{x}|$
- écart moyen = moyenne arithmétique des écarts : $\bar{e} = \frac{1}{n}\sum_{i=1}^{n} e_i = \frac{1}{n}\sum_{i=1}^{p} n_i\, e_i = \sum_{i=1}^{p} f_i\, e_i$
  - pour une série à caractère discret : p valeurs x_i, d'effectif n_i, de fréquence relative f_i
  - pour une série à caractère continu : p classes de centre x_i, d'effectif n_i, de fréquence relative f_i

3. Variance σ_x^2 et écart-type σ_x

- variance σ_x^2 = moyenne des carrés des écarts :

$$\sigma_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{p} n_i\,(x_i - \bar{x})^2 = \sum_{i=1}^{p} f_i\,(x_i - \bar{x})^2$$

- formule de calcul plus rapide : $\sigma_x^2 = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right) - \bar{x}^2 = \left(\sum_{i=1}^{p} f_i\, x_i^2\right) - \bar{x}^2$

- propriété : si $x_i = M + a\, z_i$ alors $\sigma_x = |a|\, \sigma_z$

4. Ecart interquartile

- écart interquartile : $Q_3 - Q_1$ (contient 50 % des valeurs de la série)
- écart semi-interquartile : $\frac{Q_3 - Q_1}{2}$

5. Coefficient de variation

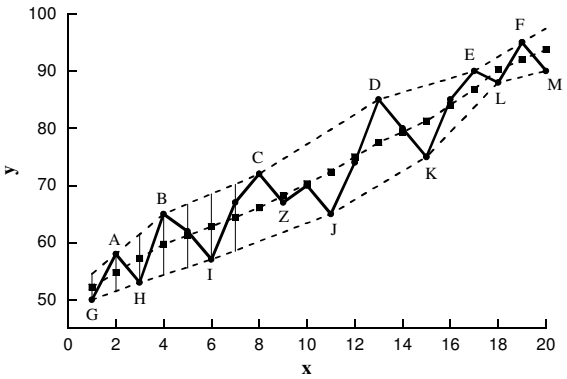
$CdV = \frac{\sigma_x}{\bar{x}}$ - sans unité, permet de comparer les dispersions de différentes séries

- problème si $\bar{x}$ proche de 0

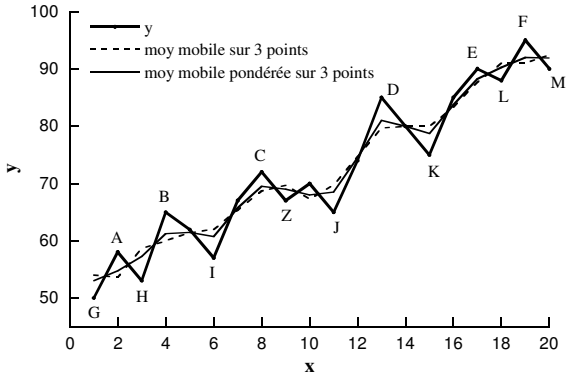
série de valeurs x_i (i = 1 à p), d'effectifs notés y_i (si série continue : x_i = centre de classes)

1. Méthodes de lissage

Méthode des points médians



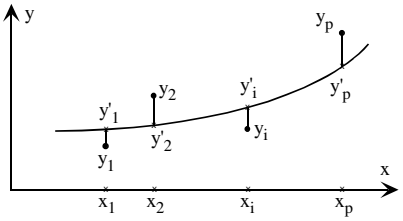
Méthode de la moyenne mobile



2. Méthodes d'ajustement

a) *méthode graphique* (pour une droite)

b) *méthode des moindres carrés :*



- on veut déterminer l'équation de la courbe ajustée $y'_i = f(x_i)$ pour minimiser $\sum_{i=1}^{p} (y_i - y'_i)^2$
- qualité de l'ajustement donnée par le coefficient de corrélation : $-1 \le R \le 1$

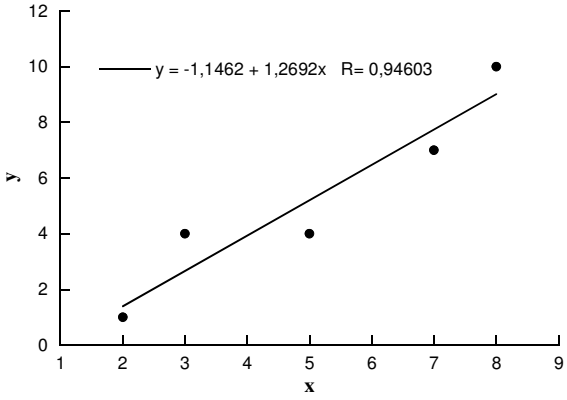
$$R = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}} = \frac{\sum_i x_i y_i - \frac{1}{p}\sum_i x_i \sum_i y_i}{\sqrt{\left(\sum_i {x_i}^2 - \frac{1}{p}(\sum_i x_i)^2\right)\left(\sum_i {y_i}^2 - \frac{1}{p}(\sum_i y_i)^2\right)}}$$

α) ajustement à l'aide d'une droite : $y = a\,x + b$ où a et b sont donnés par :

$$a = \frac{p\sum_i x_i\, y_i - \sum_i x_i \sum_i y_i}{p\sum_i {x_i}^2 - (\sum_i x_i)^2} = \frac{\overline{xy} - \bar{x}\,\bar{y}}{\overline{x^2} - \bar{x}^2} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2} \quad \text{et} \quad b = \bar{y} - a\bar{x}$$

exemple :

	x_i	y_i	${x_i}^2$	$x_i\, y_i$	${y_i}^2$
	2	1	4	2	1
	3	4	9	12	16
	5	4	25	20	16
	7	7	49	49	49
	8	10	64	80	100
somme	25	26	151	163	182

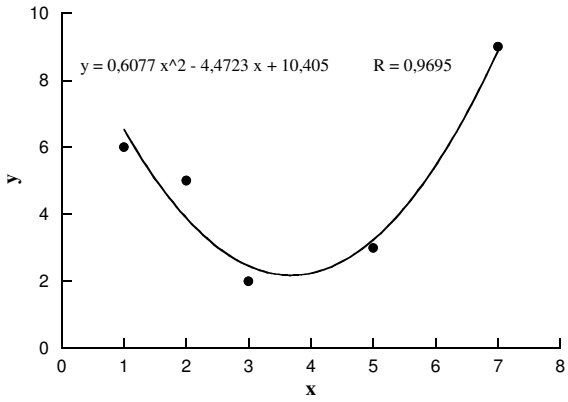


β) <u>ajustement à l'aide d'une parabole</u> : $y = a\,x^2 + b\,x + c$ où a, b et c sont donnés par :

$$a\sum_i x_i^2 + b\sum_i x_i + cp = \sum_i y_i$$
$$a\sum_i x_i^3 + b\sum_i x_i^2 + c\sum_i x_i = \sum_i x_i\,y_i$$
$$a\sum_i x_i^4 + b\sum_i x_i^3 + c\sum_i x_i^2 = \sum_i x_i^2\,y_i$$

exemple :

	x_i	y_i	x_i^2	x_i^3	x_i^4	$x_i\,y_i$	$x_i^2\,y_i$	y_i^2
	1	6	1	1	1	6	6	36
	2	5	4	8	16	10	20	25
	3	2	9	27	81	6	18	4
	5	3	25	125	625	15	75	9
	7	9	49	343	2401	63	441	81
Σ	18	25	88	504	3124	100	560	155

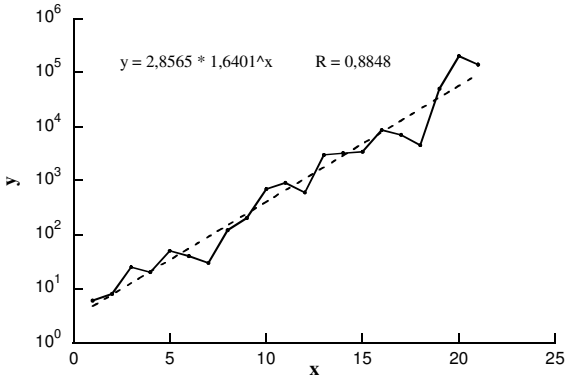


γ) <u>ajustement à l'aide d'une exponentielle</u> : $y = b\,a^x$

donc $Y = A\,x + B$
avec $Y = \log y$, $B = \log b$, $A = \log a$.

A et B sont donnés par :

$$A = \frac{p\sum_i x_i Y_i - \sum_i x_i \sum_i Y_i}{p\sum_i x_i^2 - (\sum_i x_i)^2} \quad \text{et} \quad B = \overline{Y} - A\,\overline{x}$$

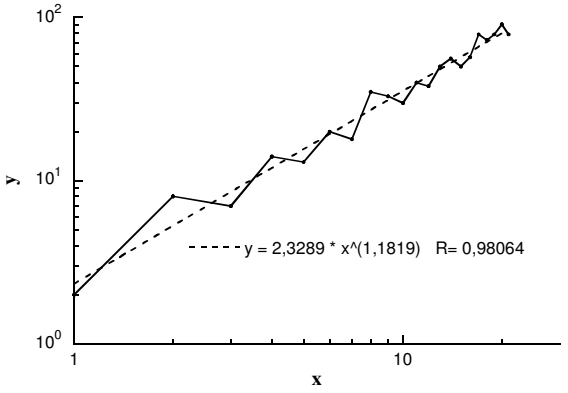


δ) <u>ajustement à l'aide d'une fonction puissance</u> : $y = b\,x^a$

donc $Y = a\,X + B$
avec $Y = \log y$, $X = \log x$, $B = \log b$.

a et B sont donnés par :

$$a = \frac{p\sum_i X_i Y_i - \sum_i X_i \sum_i Y_i}{p\sum_i X_i^2 - (\sum_i X_i)^2} \quad \text{et} \quad B = \overline{Y} - a\,\overline{X}$$



Statistiques - cours 5 : indices statistiques

1. Indice simple (ou élémentaire)

si : P_0 est la valeur d'une grandeur à la date t_0

P_1 est la valeur de la grandeur à la date t_1 : indice simple : $i_{1/0} = \frac{P_1}{P_0}$

indice simple de la grandeur à la date t_1, calculé sur la base 100 à la date t_0 : $I_{1/0} = \frac{P_1}{P_0} \times 100$

Propriétés :

- un indice simple est réversible : $\forall t : i_{0/t} = \frac{1}{i_{t/0}}$ ou : $\forall t : \frac{I_{0/t}}{100} = \frac{100}{I_{t/0}}$
- un indice simple est transférable (ou transitif, ou enchaînable, ou circulaire) :

$$\forall\, t_0, t_1, t_2 : \quad i_{t_2/t_1} \times i_{t_1/t_0} = i_{t_2/t_0} \text{ ou: } \frac{I_{t_2/t_1}}{100} \times \frac{I_{t_1/t_0}}{100} = \frac{I_{t_2/t_0}}{100}$$

2. Indice synthétique

soient plusieurs produits de prix P_i et quantités Q_i : on définit - indice des prix (P_i)
- indice des volumes (Q_i)
- indice des valeurs ($P_i\, Q_i$)

- indice des prix : n produits P_{i0} prix du produit i à la date de référence t_0
  P_{i1} prix du produit i à la date t_1

  pondérations par les quantités : Q_{i0} quantité consommée du produit i à la date t_0
  Q_{i1} quantité consommée du produit i à la date t_1

	Pondération Laspeyres (Q_i à la date t_0)	*Pondération Paasche* (Q_i à la date t_1)
• *indice des moyennes* (arithmétiques pondérées)	$L_{1/0} = \frac{\sum_i P_{i1} Q_{i0}}{\sum_i P_{i0} Q_{i0}} \times 100$	$P_{1/0} = \frac{\sum_i P_{i1} Q_{i1}}{\sum_i P_{i0} Q_{i1}} \times 100$
• *moyenne* (arithmétique pondérée) *des indices*	$L'_{1/0} = \frac{\sum_i \frac{P_{i1}}{P_{i0}} Q_{i0}}{\sum_i Q_{i0}} \times 100$	$P'_{1/0} = \frac{\sum_i \frac{P_{i1}}{P_{i0}} Q_{i1}}{\sum_i Q_{i1}} \times 100$

- indice de Fischer : $F_{1/0} = \sqrt{L_{1/0} \times P_{1/0}}$ (moyenne géométrique de $L_{1/0}$ et $P_{1/0}$)

série de n couples (x_i, y_j) avec $i = 1$ à p et $j = 1$ à q : - d'effectifs partiels notés n_{ij}

(si caractères continus : x_i et y_j = centres de classes) - de fréquences partielles : $f_{ij} = \frac{n_{ij}}{n}$

1. Distributions marginales - distributions conditionnelles

a) *tableau de contingence*

$$n_{i\bullet} = \sum_{j=1}^{q} n_{ij}$$

$$n_{\bullet j} = \sum_{i=1}^{p} n_{ij}$$

x \ y	y_1	y_j	y_q	somme
x_1	n_{11}	n_{1j}	n_{1q}	$n_{1\bullet}$
x_i	n_{i1}	n_{ij}	n_{iq}	$n_{i\bullet}$
x_p	n_{p1}	n_{pj}	n_{pq}	$n_{p\bullet}$
somme	$n_{\bullet 1}$	$n_{\bullet j}$	$n_{\bullet q}$	n

b) *distributions marginales*

- distribution marginale de x : valeurs x_i, effectif marginal de $x_i = n_{i\bullet}$
- distribution marginale de y : valeurs y_j, effectif marginal de $y_j = n_{\bullet j}$
- fréquence marginale de x_i : $f_{i\bullet} = \frac{n_{i\bullet}}{n}$ et fréquence marginale de y_j : $f_{\bullet j} = \frac{n_{\bullet j}}{n}$

$$n = \sum_{i=1}^{p} n_{i\bullet} = \sum_{j=1}^{q} n_{\bullet j} = \sum_{i=1}^{p}\sum_{j=1}^{q} n_{ij} \qquad \text{et} \qquad \sum_{i=1}^{p} f_{i\bullet} = \sum_{j=1}^{q} f_{\bullet j} = \sum_{i=1}^{p}\sum_{j=1}^{q} f_{ij} = 1$$

c) *distributions conditionnelles*

- distribution conditionnelle de x pour $y = y_j$: valeurs x_i, effectif = n_{ij} (j$^{\text{ème}}$ colonne du tableau)
- distribution conditionnelle de y pour $x = x_i$: valeurs y_j, effectif = n_{ij} (i$^{\text{ème}}$ ligne du tableau)
- fréquence conditionnelle de x_i si $y = y_j$: $f_{i|j} = \frac{n_{ij}}{n_{\bullet j}}$ (% d'unités ayant x_i parmi ceux ayant y_j)
- fréquence conditionnelle de y_j si $x = x_i$: $f_{j|i} = \frac{n_{ij}}{n_{i\bullet}}$ (% d'unités ayant y_j parmi ceux ayant x_i)
- *propriétés* : $f_{ij} = f_{i|j} \times f_{\bullet j}$, $f_{ij} = f_{j|i} \times f_{i\bullet}$, et $\sum_{i=1}^{p} f_{i|j} = \sum_{j=1}^{q} f_{j|i} = 1$

2. Caractéristiques des séries à 2 caractères

a) *caractéristiques des distributions marginales*

- moyenne de x : $\bar{x} = \sum_{i=1}^{p} x_i \, f_{i\bullet} = \sum_{i=1}^{p}\sum_{j=1}^{q} x_i \, f_{ij}$ moyenne de y : $\bar{y} = \sum_{j=1}^{q} y_j \, f_{\bullet j} = \sum_{i=1}^{p}\sum_{j=1}^{q} y_j \, f_{ij}$
- variance de x : $\sigma_x^2 = \sum_{i=1}^{p} (x_i - \bar{x})^2 \, f_{i\bullet} = \overline{x^2} - \bar{x}^2$ et de y : $\sigma_y^2 = \sum_{j=1}^{q} (y_j - \bar{y})^2 \, f_{\bullet j} = \overline{y^2} - \bar{y}^2$

b) *caractéristiques des distributions conditionnelles*

- moyenne de x si y_j : $\bar{x}_j = \sum_{i=1}^{p} x_i \; f_{i|j}$ moyenne de y si x_i : $\bar{y}_i = \sum_{i=1}^{q} y_j \; f_{j|i}$

variance de x si y_j: $\sigma_j^2(x) = \sum_{i=1}^{p} (x_i - \bar{x}_j)^2 \; f_{i|j} = \overline{x_j^2} - \bar{x}_j{}^2$ et de y si x_i : $\sigma_i^2(y) = \sum_{i=1}^{p} (y_j - \bar{y}_i)^2 \; f_{j|i} = \overline{y_i^2} - \bar{y}_i{}^2$

c) *relations entre les caractéristiques marginales et conditionnelles*

- sur les moyennes : $\bar{x} = \sum_{j=1}^{q} \bar{x}_j \; f_{\bullet j}$ $\bar{y} = \sum_{i=1}^{p} \bar{y}_i \; f_{i\bullet}$

- sur les variances : $\sigma_x^2 = \sum_{j=1}^{q} \sigma_j^2(x) \; f_{\bullet j} + \sum_{j=1}^{q} (\bar{x}_j - \bar{x})^2 \; f_{\bullet j}$ et $\sigma_y^2 = \sum_{i=1}^{p} \sigma_i^2(y) \; f_{i\bullet} + \sum_{i=1}^{p} (\bar{y}_i - \bar{y})^2 \; f_{i\bullet}$

(moyenne des variances conditionnelles + variance des moyennes conditionnelles)

d) *covariance de x et y*

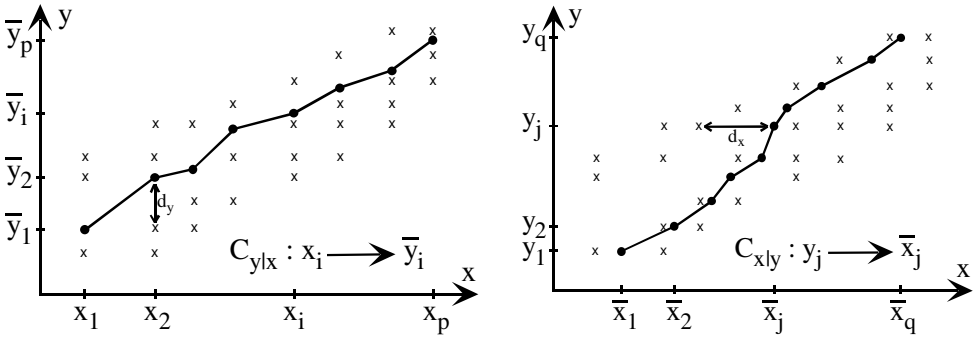
$$\mathrm{Cov}(x,y) = \sum_{i=1}^{p}\sum_{j=1}^{q} (x_i - \bar{x})(y_j - \bar{y}) \; f_{ij} = \sum_{i=1}^{p}\sum_{j=1}^{q} x_i \; y_j \; f_{ij} - \bar{x} \; \bar{y} = \overline{xy} - \bar{x} \; \bar{y}$$

$$\mathrm{Cov}(x,x) = \overline{x^2} - \bar{x}^2 = \sigma_x^2$$

3. Représentation graphique des séries à 2 caractères

courbes de régression $C_{y|x}$ et $C_{x|y}$:

"meilleures" courbes (polygonales) passant au milieu des points



somme des d_y^2 minimale somme des d_x^2 minimale

4. Séries à caractères x et y indépendants

$$f_{i|j} = f_{i\bullet} \text{ et } f_{j|i} = f_{\bullet j}$$

d'où les conséquences :

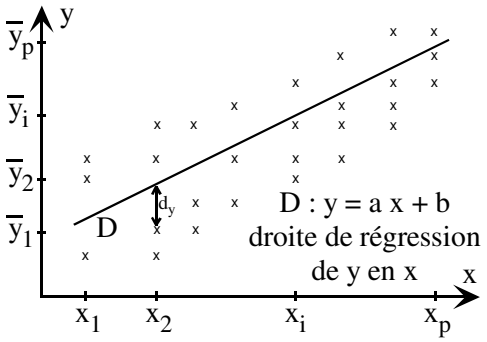
- $f_{ij} = f_{i\bullet} \times f_{\bullet j}$ les fréquences marginales de x et y donnent la fréquence partielle de (x_i, y_j)
- $\bar{x} = \bar{x}_j$ et $\bar{y} = \bar{y}_i$ la courbe de régression $C_{x|y}$ est verticale, $C_{y|x}$ est horizontale
- $\overline{xy} = \bar{x} \; \bar{y}$ d'où Cov(x,y) = 0 (attention : si Cov(x,y) = 0, x et y peuvent être dépendants)

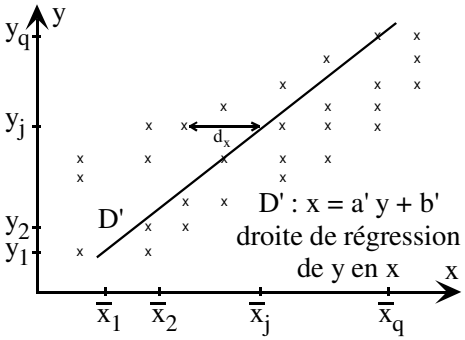
5. Ajustement, corrélation, régression

a) *ajustement linéaire - droites de régression*

- *droites de régression* $C_{y|x}$ et $C_{x|y}$:

"meilleures" droites passant au milieu des points





$$a = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\overline{x^2} - \overline{x}^2} = \frac{\mathrm{Cov}(x,y)}{\sigma_x^2},\ b = \overline{y} - a\overline{x} \qquad a' = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\overline{y^2} - \overline{y}^2} = \frac{\mathrm{Cov}(x,y)}{\sigma_y^2},\ b' = \overline{x} - a'\overline{y}$$

a = coefficient de régression de y en x $\qquad$ a' = coefficient de régression de x en y

- angle entre D et D' : $\theta = \mathrm{Arc\,tan}\frac{1}{a'} - \mathrm{Arc\,tan}\,a$
- θ mesure le degré de corrélation linéaire entre x et y :
  - $\theta = 0$: D et D' confondues : x et y liés par une fonction linéaire
  - $\theta = \pi/2$: D et D' perpendiculaires : x et y indépendants
- règle : $0 \le a\,a' \le 1$ a pour conséquence que la droite D' est toujours plus pentue que D

b) *coefficient de corrélation linéaire*

- $R = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\sigma_x\,\sigma_y} = \frac{\mathrm{Cov}(x,y)}{\sigma_x\,\sigma_y}$ avec $-1 \le R \le 1$, R mesure le degré de corrélation linéaire entre x et y
- $R = a\,a'$: donc l'angle entre D et D' mesure aussi ce degré de corrélation linéaire

c) *corrélation non linéaire*

- rapport de corrélation de y en x :

$$\eta^2_{y|x} = \frac{\sum_{i=1}^{p}\sum_{j=1}^{q} n_{ij}(\overline{y}_i - \overline{y})^2}{\sum_{i=1}^{p}\sum_{j=1}^{q} n_{ij}(y_j - \overline{y})^2}$$

avec : $0 \le \eta^2_{y|x} \le 1$

- rapport de corrélation de x en y :

$$\eta^2_{x|y} = \frac{\sum_{i=1}^{p}\sum_{j=1}^{q} n_{ij}(\overline{x}_i - \overline{x})^2}{\sum_{i=1}^{p}\sum_{j=1}^{q} n_{ij}(x_i - \overline{x})^2}$$

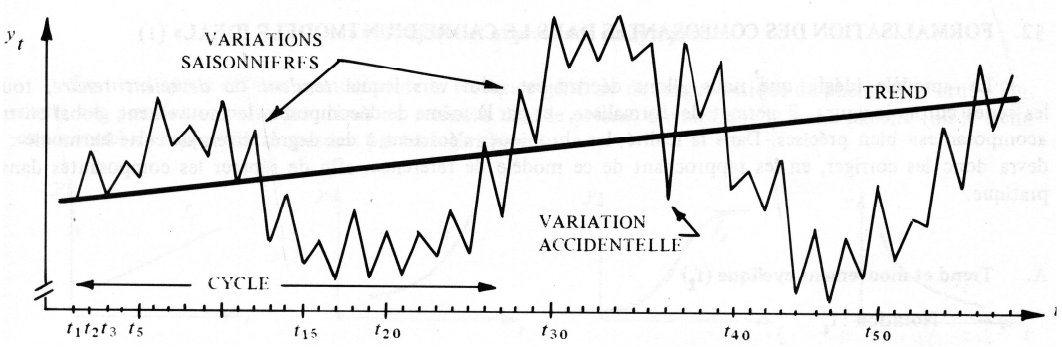
avec : $0 \le \eta^2_{x|y} \le 1$

- ces rapports mesurent le degré de corrélation directement sur les courbes de régression $C_{y|x}$ et $C_{x|y}$
- plus ces rapports sont proches de 1, plus la liaison est forte entre x et y
- si $\eta^2_{y|x} = 0$: alors $\overline{y} = \overline{y}_i$, $C_{y|x}$ est horizontale, y ne dépend pas de x (en général)
- si $\eta^2_{x|y} = 0$: alors $\overline{x} = \overline{x}_j$, $C_{x|y}$ est verticale, x ne dépend pas de y (en général).

1. Définition - composantes des séries chronologiques

a) *définition* : suite d'observations (notées y_t ou y_i) d'une variable quantitative ordonnées dans le temps

b) *composantes d'une série chronologique*



Bernard Py,

Statistique descriptive

(Economica)

3 composantes : - composante conjoncturelle (trend + cycle) : c_t
- composante saisonnière : s_t
- variations résiduelles : ε_t

2. Les modèles de composition

- *hypothèse* : composante saisonnière périodique $s_{t+p} = s_t$

schéma additif :

$y_t = c_t + s_t + \varepsilon_t$

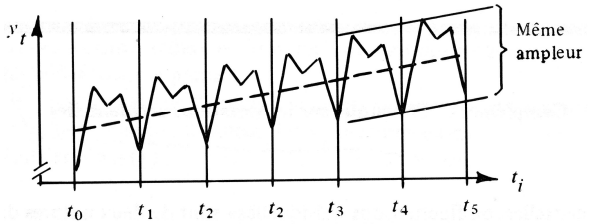


schéma multiplicatif :

$y_t = c_t \times s_t + \varepsilon_t$

ou $y_t = c_t \times s_t \times \varepsilon_t$

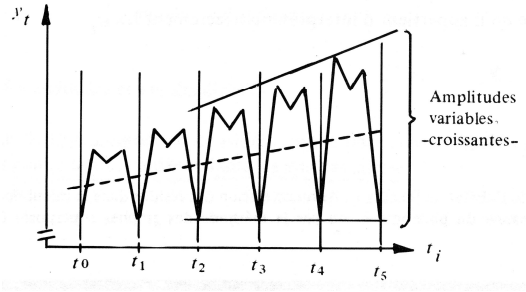
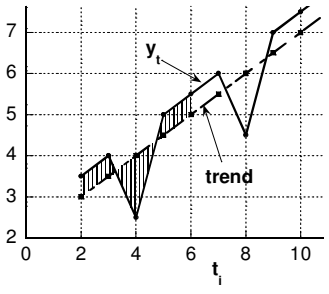


schéma additif :

$$\sum_{i=0}^{p-1} s_{t+i} = 0$$

soit : $\bar{s}_t = 0$



3. Analyse d'une série chronologique (limitée en cours au modèle additif)

a) *détermination du trend*

- méthode analytique : lorsqu'on devine l'expression analytique du trend (ex: $c_t = a\,t + b$ pour un trend linéaire) on détermine ses paramètres (ex : a et b) par la méthode des moindres carrés.

- <u>méthode empirique</u> : méthode de la moyenne mobile (MM)

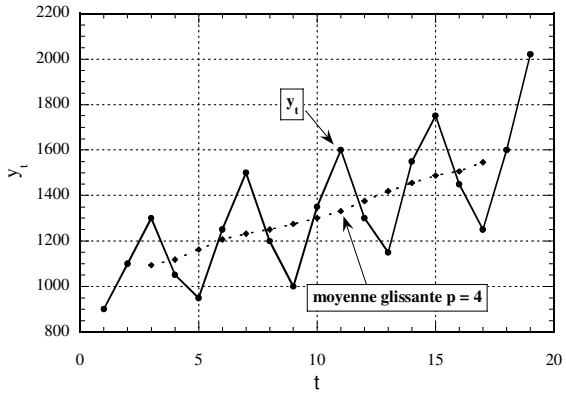
∗ *cas impair* : ex. sur 3 points

$$y'_t = \frac{1}{3}(y_{t-1} + y_t + y_{t+1})$$

∗ cas *pair* : sur 4 points (moyenne de $y_{t+0,5}$ et de $y_{t+1,5}$)

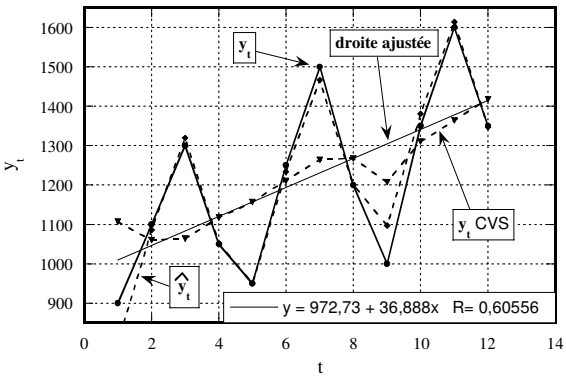
$$y'_{t+1} = \frac{1}{8}(y_{t-1} + 2y_t + 2y_{t+1} + 2y_{t+2} + y_{t+3})$$

* la MM d'ordre p élimine les variations saisonnières d'ordre p idéales (vérifiant : $\sum_{i=0}^{p-1} s_{t+i} = 0$)



b) *correction des variations saisonnières* (exemple pour un trend linéaire)

- détermination de l'équation de la droite ajustant le trend : $c_t = a\,t + b$ (méthode des moindres carrés)
- calcul de la composante saisonnière : $s_t = y_t - (a\,t + b)$
- calcul des coefficients saisonniers S_j (j = 1 à p) et des coefficients corrigés : $S'_j = S_j - \frac{1}{p}\sum_{i=1}^{p} S_i$
- calcul de la série chronologique ajustée : $\hat{y}_t = a\,t + b + S'_j$
- calcul de la série corrigée des variations saisonnières (CVS): $y^*_t = y_t - S'_j$
- d'où les variations résiduelles $\varepsilon_t = y^*_t - (at + b)$



t	y_t	c_t	$y_t - c_t$
1	900	1009,6	-109,62
2	1100	1046,5	53,494
3	1300	1083,4	216,61
4	1050	1120,3	-70,282
5	950	1157,2	-207,17
6	1250	1194,1	55,942
7	1500	1230,9	269,05
8	1200	1267,8	-67,834
9	1000	1304,7	-304,72
10	1350	1341,6	8,390
11	1600	1378,5	221,50
12	1350	1415,4	-65,386

S_1

coefficients saisonniers

S_1 = -207,170
S_2 = 39,2753
S_3 = 235,721
S_4 = -67,8340

total : -0,00804
d'où ρ = -0,0020

d'où les coefficients saisonniers corrigés :

S'_1 = -207,168
S'_2 = 39,2773
S'_3 = 235,723
S'_4 = -67,8320

4. <u>Comparaison de 2 séries</u> y_t et z_t

- on pose $Y_t = \frac{y_t - \bar{y}}{\sigma_y}$ et $Z_t = \frac{z_t - \bar{z}}{\sigma_z}$

- coefficient de covariation linéaire: $C = \frac{1}{n}\sum_{t=1}^{n} Y_t\, Z_t$