

# Chapitre 9. Analyse de la variance

Dans ce chapitre nous étudions comment l'analyse de la variance de  $Y$  permet de tester l'égalité des moyennes conditionnelles de cette variable numérique dans les sous-populations induites par  $X$ . Dans cette problématique,  $X$  est appelée la *variable explicative*, ou le facteur explicatif, et  $Y$  la *variable expliquée*.

Dans la formule de la décomposition de la variance,

$$\text{“variance totale} = \text{variance intra} + \text{variance inter”},$$

la variance intra, moyenne des variances (conditionnelles), quantifie la part de la variabilité intrinsèque de  $Y$ , et la variance inter, variance des moyennes (conditionnelles), mesure l'hétérogénéité des sous-populations.

## 1) Variance expliquée

**Définition:** *La variance expliquée par la variable  $X$  est égale à la variance inter divisée par la variance globale de  $Y$ .*

C'est un nombre compris entre 0 et 1 puisque les variances sont des nombres positifs ou nuls, et que la variance inter est une part de la variance globale.

La variance expliquée est une mesure du lien entre le facteur  $X$  et la mesure numérique  $Y$ , pour apprécier comment  $Y$  dépend du fait d'appartenir à une sous-population ou à une autre.

*Exemple du nombre d'enfants de 0 à 6 ans selon la structure familiale (exemple 1 du chapitre 6) :*

Arrondi	4						
	Effectif	$\sum y_i$	$\sum y_i^2$	$\bar{y}_i$	$\sigma_i^2$	$n_i \cdot \bar{y}_i^2$	$n_i \cdot \sigma_i^2$
Mono	1035	1669	3199	1,6126	0,4903	2691,4955	507,4605
Couples	6536	11036	22802	1,6885	0,6376	18634,3468	4167,3536
Total	7571	12705	26001			21325,8423	4674,8141
Moyenne		1,6781	3,4343			2,8168	0,6175
	Variance	0,6183		Inter	0,0008	0,6175	Intra

Les moyennes dans les deux sous-populations diffèrent assez peu (de l'ordre de 5%), la variance inter est très faible ( $\approx 0,001$ ) de même que la variance expliquée par le facteur "structure familiale" ( $\approx 0.001$ ).

- Si la variance expliquée est égale à 1, la variance intra vaut 0, ce qui entraîne que toutes les variances conditionnelles sont nulles (la variance intra étant une somme de nombres positifs ou nuls, elle ne peut valoir 0 que si chaque terme est nul). Par conséquent, les individus de chaque sous-population ont tous la même mesure  $Y$ .

Dans ces conditions, le facteur  $X$  détermine entièrement la mesure  $Y$ : il suffit de connaître la sous-population dans laquelle l'individu se trouve pour connaître sa mesure  $Y$ .

Inversement il suffit de connaître la mesure  $Y$  d'un individu pour savoir dans quelle sous-population il se trouve, en supposant que les sous-populations ont toutes des valeurs de  $Y$  différentes.

- Si la variance expliquée est égale à 0, la variance inter est nulle, ce qui revient à dire que les moyennes conditionnelles de  $Y$  sont identiques :  $Y$  donne globalement les mêmes mesures sur toutes les sous-populations.

Cela ne signifie pas à proprement parler que  $X$  et  $Y$  sont indépendantes (il faudrait pour cela qu'en plus de l'égalité des moyennes conditionnelles on ait l'égalité des distributions conditionnelles), mais que le facteur  $X$  n'a pas d'effet global sur la mesure de  $Y$ .

Inversement, si  $X$  et  $Y$  sont indépendantes, les moyennes conditionnelles de  $Y$  sont identiques, la variance inter est donc nulle et la variance expliquée est égale à 0.

- Si la variance expliquée est proche de 0, la variance inter l'est également. Les moyennes de  $Y$  dans les sous-populations sont peu différentes les une des autres ; alors de deux choses l'une :
  - ou bien ces différences sont un effet des fluctuations d'échantillonnage pour des sous-populations de même moyenne, ce qui revient à considérer que le facteur  $X$  n'a pas d'effet sur la variable  $Y$ .
  - ou bien au contraire, elles sont dues au fait que  $X$  a un effet sur la  $Y$ , sans doute faible, mais suffisant pour rendre les mesures de  $Y$  différentes dans les sous-populations.

Ces observations ont donné l'idée d'un test pour apprécier si des petites différences entre les moyennes de  $Y$  dans les sous-populations peuvent être interprétées ou non comme l'effet du facteur  $X$ .

## 2) Test de l'égalité des moyennes par l'analyse de la variance

Soit  $T$  la valeur  $(n - k) * \frac{\text{var inter}}{\text{var intra}}$  calculée à partir de l'observation d'un échantillon de taille  $n$ ,  $k$  étant le nombre de sous-populations (le nombre de modalités de  $X$ ). Définissons l'hypothèse

*(H) : les moyennes de Y sont identiques dans les sous-populations.*

Si  $(H)$  est vraie alors les valeurs de  $T$  calculées sur différents échantillons vont varier au voisinage de 0, puisque du fait des fluctuations d'échantillonnage, les variances inter vont être proches de 0.

Plus précisément, on sait par des travaux mathématiques que sous cette hypothèse  $(H)$ , les valeurs de  $T$  varient approximativement comme la distribution du  $\chi^2$  à  $k - 1$  degrés de liberté.

Le principe du test est identique à celui du  $\chi^2$  pour l'indépendance statistique:

1. on détermine d'abord un seuil  $s$  au-delà duquel les valeurs de  $T$  sont considérées comme improbables si l'hypothèse ( $H$ ) est vraie, en utilisant une table des distributions du  $\chi^2$  qui est la distribution de  $T$  dans ce cas. Par exemple, si  $s$  est le 95ème centile, c'est-à-dire  $s = l_{k-1}(95\%)$ , alors 5% seulement des échantillons peuvent donner pour  $T$  une valeur supérieure à  $s$  (en supposant ( $H$ ) vraie).
2. on calcule la valeur  $t$  de  $T$  pour l'échantillon observé.
3. si  $t > s$  (i.e.,  $t$  fait partie des valeurs improbables si ( $H$ ) vraie), alors **on rejette l'hypothèse ( $H$ )**, en considérant que le facteur  $X$  a un effet global sur  $Y$  (les moyennes de  $Y$  sont différentes dans les sous-populations).
4. si  $t \leq s$  (i.e.,  $t$  ne fait pas partie des valeurs improbables si ( $H$ ) vraie), alors **on ne rejette pas l'hypothèse ( $H$ )**, en considérant qu'il est possible que le facteur  $X$  n'ait pas d'effet global sur  $Y$ .



*Exemple du nombre d'enfants de 0 à 6 ans selon la structure familiale :*

- si la moyenne du nombre d'enfants est identique dans les deux structures familiales (Hypothèse ( $H$ )), la variable  $T$  suit une loi du  $\chi^2$  à  $2 - 1 = 1$  degré de liberté ; le 95ème centile vaut 3,84, ce qui signifie que dans cette hypothèse ( $H$ ), 5% seulement des échantillons donnent une valeur de  $T$  supérieure à  $s = 3,84$ .
- pour l'échantillon observé, la valeur  $t$  calculée pour  $T$  est égale à  $(7571 - 2) * \frac{0,0008}{0,6175} \approx 9,8$  ; si ( $H$ ) était vraie, l'échantillon serait un de ces 5% d'échantillons atypiques, ce que nous n'avons pas de raison de penser : nous rejetons donc cette hypothèse en admettant que le facteur "structure familiale" a un (faible) effet global sur le nombre d'enfants, puisque les moyennes sont considérées comme différentes dans les deux sous-populations.

On remarquera en passant que la faible valeur de la variance inter conduit quand même à considérer comme significative la différence entre les moyennes des deux sous-populations (elle n'est pas l'effet des fluctuations d'échantillonnage) ; cela est dû à la grande taille de l'échantillon.

En supposant celui-ci de taille 1000,  $t$  vaudrait

$(1000 - 2) * \frac{0,0008}{0,6175} \approx 1,3$ , ce qui nous aurait amené à ne pas rejeter l'hypothèse d'égalité des moyennes.

# Chapitre 10. Deux variables numériques

Dans ce chapitre nous étudions la liaison entre deux variables numériques. On peut bien entendu appliquer aux deux variables les procédures étudiées dans le cas où une seule des deux variables est numérique. La nouveauté est que chaque observation étant un couple de nombres  $(x, y)$ , elle peut être représentée graphiquement par un point d'un plan ; on peut alors faire appel à des procédures géométriques pour visualiser l'ensemble des observations, et analyser la forme du **nuage** ainsi formé. Nous présenterons ensuite la **covariance**, un indice de covariation linéaire des deux variables puis le **coefficient de corrélation linéaire**.

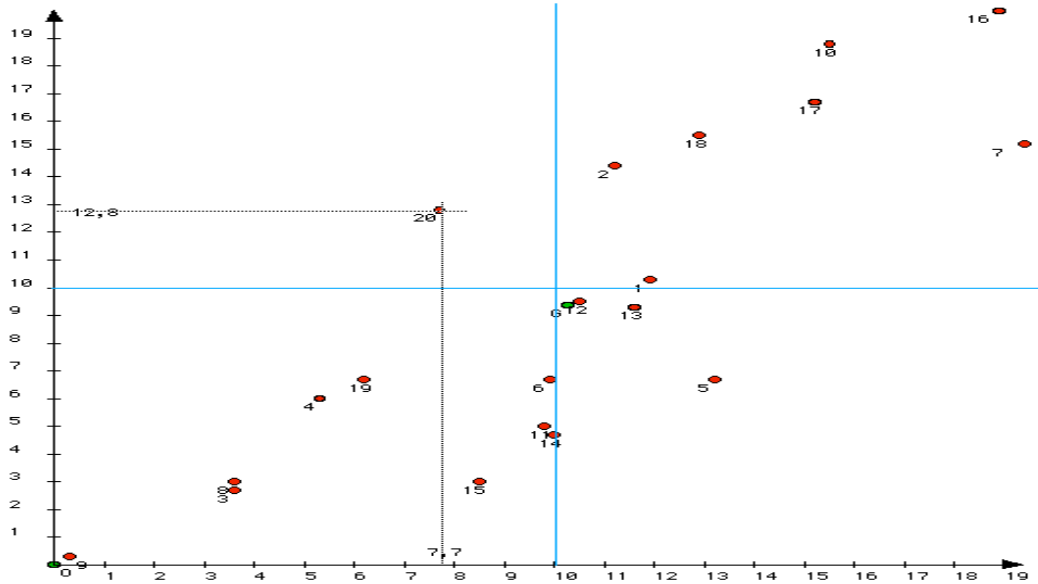
## 1) Nuage des observations.

Sur un plan, on place un repère orthonormé : l'axe horizontal gradué des abscisses qui identifie les valeurs de la variable  $X$ , et l'axe vertical gradué des ordonnées qui identifie celles de la variable  $Y$ . L'observation d'un individu  $e$  est représentée par le point  $e$  d'abscisse  $x_e$  et d'ordonnée  $y_e$ . L'ensemble des points est le **nuage des observations**. La “moyenne conjointe”  $M$  (point de coordonnées  $\bar{x}$  et  $\bar{y}$ ) est appelée le **centre de gravité du nuage**.

*Exemple : les notes à deux partiels de statistique pour un échantillon de 20 étudiants en deuxième année de psychologie.*

	1	2	3	4	5	6	7	8	9	10
P1	11,9	11,2	3,6	5,3	13,2	9,9	19,4	3,6	0,3	15,5
P2	10,3	14,4	2,7	6	6,7	6,7	15,2	3	0,3	18,8
P1	11	12	13	14	15	16	17	18	19	20
P1	9,8	10,5	11,6	10	8,5	18,9	15,2	12,9	6,2	7,7
P2	5	9,5	9,3	4,7	3	20	16,7	15,5	6,7	12,8

Le nuage :



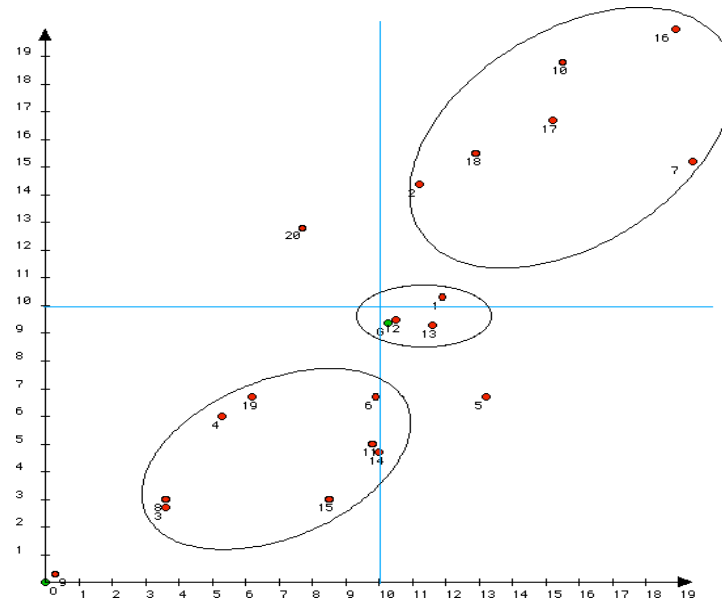
Pour représenter l'étudiant 20 par exemple, on trace une verticale d'abscisse  $x_{20} = 7,7$  et une horizontale d'ordonnée  $y_{20} = 12,8$  : le point est à l'intersection des deux droites.

Le centre de gravité  $M$  est le point de coordonnées  $x_M = 10,3$  et  $y_M = 9,4$ .

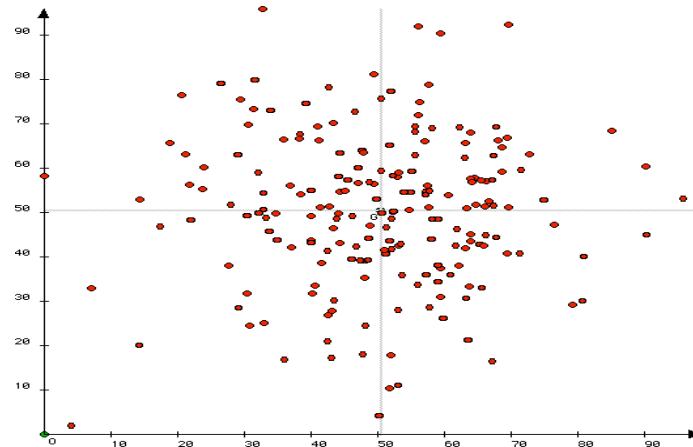
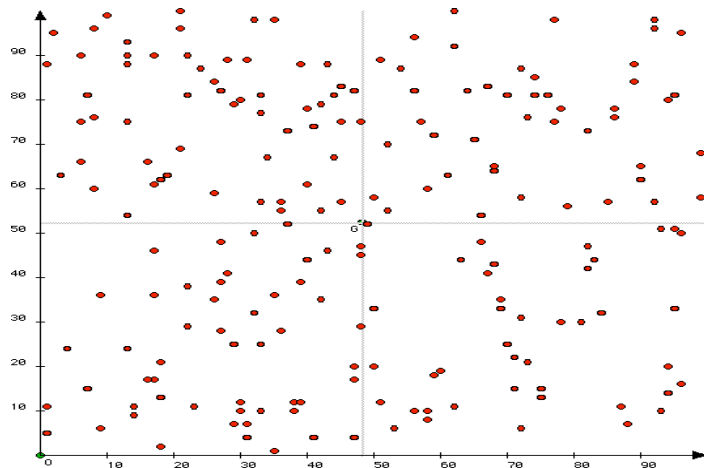
Si on trace une verticale d'abscisse 10 et une horizontale d'ordonnée 10, on découpe le plan en quartiers :

- le quartier nord-est contient les étudiants qui ont la moyenne aux deux partiels.
- le quartier sud-ouest contient les étudiants qui n'ont la moyenne à aucun des partiels.
- le quartier sud-est contient les étudiants qui ont la moyenne au premier partiel seulement.
- le quartier nord-ouest contient ceux qui ont la moyenne au second partiel seulement.

**Formes de nuage et liaison de variables.** Le nuage est l'image de l'échantillon vue à travers le prisme des variables  $X$  et  $Y$  ; l'analyse de sa forme permet d'apprécier la façon dont ces deux variables structurent l'échantillon, ou de manière équivalente la liaison de ces variables qui se manifeste dans l'échantillon. Dans l'exemple précédent on peut repérer trois sous-échantillons et 3 observations atypiques.

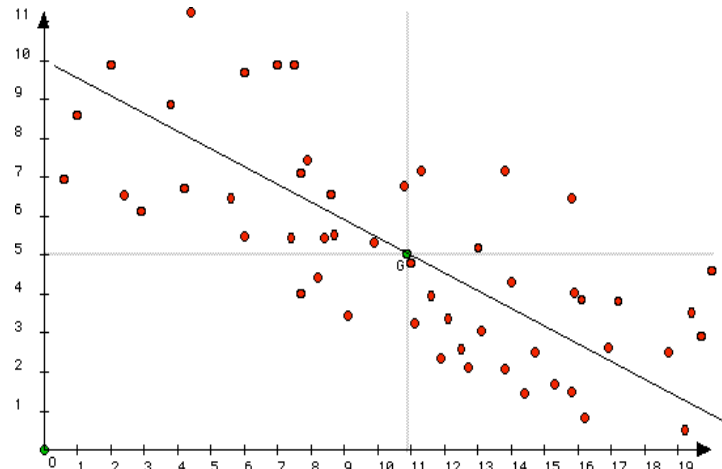
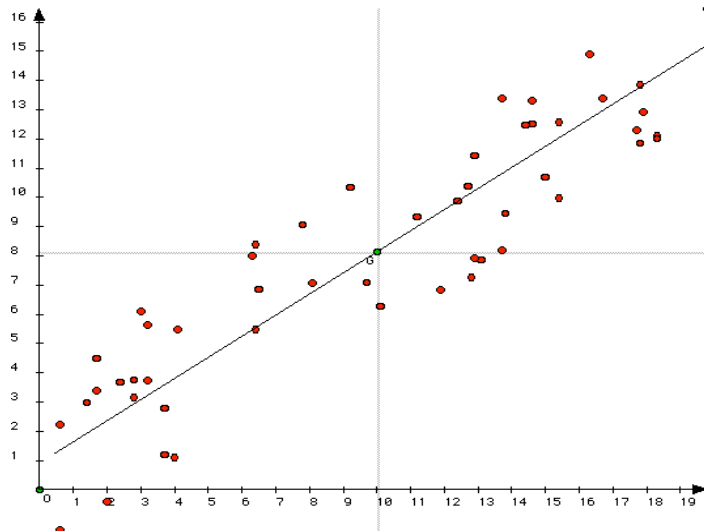


**Indépendance.** Lorsque  $X$  et  $Y$  sont indépendantes, le nuage se répartit régulièrement autour du centre de gravité, dans une forme qui dépend de la distribution des variables ; ci-dessous deux nuages de variables indépendantes, le premier quand elles sont uniformément distribuées, le second quand leur distribution est en en “cloche”.

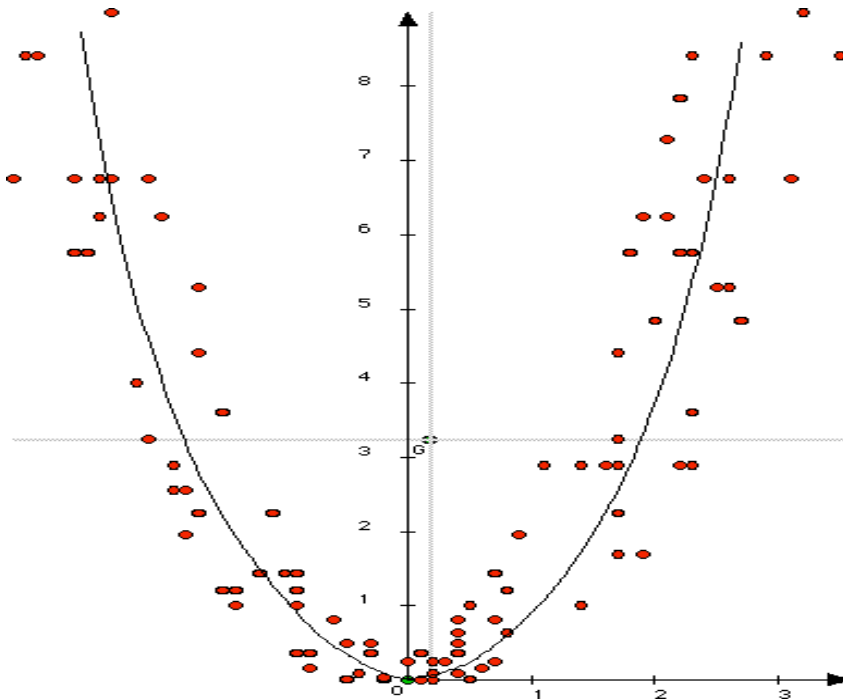




**Covariation linéaire.** Dans ce cas, le nuage s'étend le long d'une droite imaginaire qui formalise la liaison linéaire entre les deux variables. Le premier nuage allongé le long d'une droite ascendante rend compte d'une covariation positive ( $y$  croît globalement avec  $x$ ), le second d'une covariation négative ( $y$  décroît quand  $x$  croît).



**Liaison fonctionnelle.** Le nuage se disperse le long de la courbe d'une fonction qui exprime la liaison entre les deux variables.



## 2) Covariance

La **covariance** est une **mesure de la variation simultanée des deux variables**, de leur **covariation**.

**Définition** : *La covariance d'une série de  $n$  observations conjointes de deux variables numériques  $X$  et  $Y$  est la “moyenne des produits des écarts à la moyenne” :*

$$cov(x, y) = \frac{\sum_{i=1, \dots, n} (x_i - \bar{x}) * (y_i - \bar{y})}{n}.$$

On peut démontrer qu'elle est égale à “**la moyenne des produits moins le produit des moyennes**” :

$$cov(x, y) = \frac{\sum_{i=1, \dots, n} x_i * y_i}{n} - \bar{x} * \bar{y}.$$

*Exemple : Taille des pères et fils aînés dans 12 familles.*

	1	2	3	4	5	6	7	8	9	10	11	12	Moy.
x	165	160	170	163	173	157	178	168	173	170	175	180	169,33
y	173	168	180	165	175	168	173	165	175	170	173	178	171,92
x*y	28545	26880	30600	26895	30275	26376	30794	27720	30275	28900	30275	32040	29131,25

En utilisant la seconde formule, on obtient

$$cov(x, y) = 29131,25 - (169,33 * 171,92) = 20.036.$$

Lorsque les résultats sont fournis sous forme de tableau de contingence et non de données brutes et que les variables sont continues, on fait comme si les  $n_{ij}$  individus de la modalité conjointe  $(i, j)$  étaient tous groupés au point  $(c_i, c'_j)$  où  $c_i$  et  $c'_j$  sont les centres des modalités-intervalles, et on approxime la covariance par les deux formules :

$$\frac{\sum_{i=1, \dots, k, j=1, \dots, p} n_{ij} * (c_i - \bar{x}) * (c'_j - \bar{y})}{n}$$

et

$$\frac{\sum_{i=1, \dots, k, j=1, \dots, p} n_{ij} * c_i * c'_j}{n} - \bar{x} * \bar{y}$$

On remarquera que dans les deux premières formules on parcourt les  $n$  observations, alors que dans ces deux dernières on parcourt les  $k * p$  modalités conjointes.

Si  $X$  (respectivement  $Y$ ) est discrète, on remplace  $c_i$  (resp.  $c'_j$ ) par la modalité  $m_i$  (resp.  $m'_j$ ) dans les formules ci-dessus.

*Exemple : Distribution en pourcentage des revenus (exprimés en centaines d'euros) de l'épouse (Y) et de l'époux (X) dans des ménages.*

X / Y	[5,15[	[15,25[	[25,35[	[35,45[	Total X
[5,15[	20	10	0	0	30
[15,25[	4	36	5	0	45
[25,35[	1	9	10	0	20
[35,45[	0	0	0	5	5
Total Y	25	55	15	5	100

On utilise la seconde formule : pour chaque modalité conjointe  $(i, j)$  on calcule le produit  $n_{ij} * c_i * c'_j$ , par exemple pour la modalité  $(2, 1)$  on aura le produit  $4*20*10=800$  :

Centres	10	20	30	40	10	20	30	40	
10	20	10	0	0	2000	2000	0	0	
20	4	36	5	0	800	14400	3000	0	
30	1	9	10	0	300	5400	9000	0	
40	0	0	0	5	0	0	0	8000	
									449

On fait la moyenne des  $k * p$  nombres calculés, ici 449, auquel on retranche le produit des moyennes  $20*20$  :  $cov(x, y) = 49$ .

## Propriétés de la covariance.

1.  $cov(x, x) = var(x)$ .

2.  $cov(x - a, y - b) = cov(x, y)$ .

En particulier  $cov(x - \bar{x}, y - \bar{y}) = cov(x, y)$  : on ne modifie pas la covariance en centrant les données.

3.  $cov(a * x, b * y) = a * b * cov(x, y)$ .

En particulier  $cov\left(\frac{x}{\sigma(x)}, \frac{y}{\sigma(y)}\right) = \frac{cov(x, y)}{\sigma(x) * \sigma(y)}$  : la covariance des données réduites est égale à la covariance des données divisée par le produit des écart-types.

4. La **covariance des données centrées et réduites** est égale à la **covariance des données divisée par le produit des écart-types** : c'est une conséquence des deux propriétés précédentes.

### 3) Corrélation linéaire.

Le coefficient de corrélation linéaire d'une série de  $n$  observations conjointes est la covariance des données centrées et réduites. On le note  $r$  et d'après la 4ème propriété on a :

$$r = \frac{\text{cov}(x, y)}{\sigma(x) * \sigma(y)}$$

$r$  est un **nombre sans dimension** (il ne dépend pas des unités de mesure) et on peut démontrer qu'il est **compris entre -1 et 1**.

Exemple : Taille des pères et fils.

$\text{cov}(x, y) = 20,036$ ,  $\sigma(x) = 6,76$  et  $\sigma(y) = 4,61$ , si bien que

$r = \frac{20,036}{6,76 * 4,61} = 0,643$  ; en utilisant le mètre plutôt que le centimètre on

aurait la même valeur aux arrondis près,  $\frac{0,002}{0,0676 * 0,0461} = 0,642$



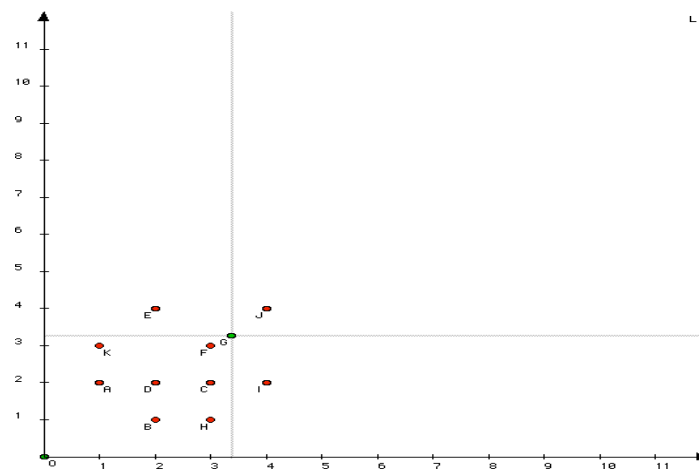
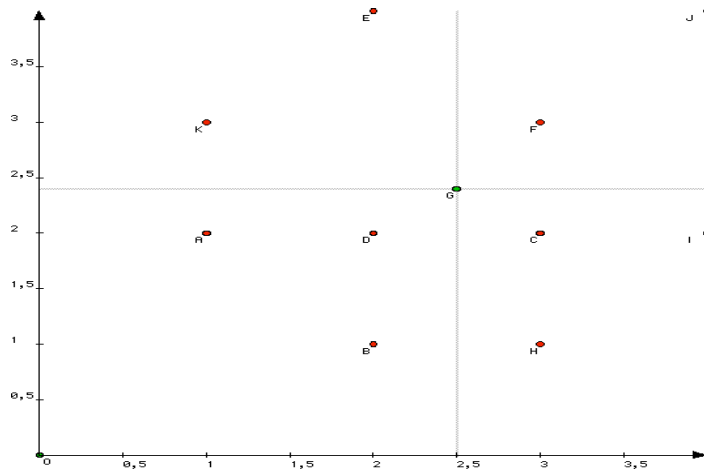
#### 4) **Corrélation et covariation linéaire.**

Si on trace des droites parallèles aux axes de coordonnées passant par le centre de gravité  $M$ , on partage le plan en quatre parties : NO (Nord-Ouest), NE (Nord-Est), SO (Sud-Ouest) et SE (Sud-Est).

Si les valeurs de  $X$  et  $Y$  sont liées par une relation linéaire et qu'elles covarient linéairement, le nuage s'étend le long d'une droite ascendante (dans les quadrants SO et NE) ou descendante (dans les quadrants NO et SE). Dans le premier cas le coefficient de corrélation linéaire  $r$  sera proche de 1 et dans le second proche de -1 ; à la limite il sera égal à 1 ou -1 si les points forment une droite.

Lorsque  $r$  est proche de 1 ou -1, on dit que les variables sont **linéairement corrélées** : la **covariation linéaire entraîne donc la corrélation linéaire.**

Réciproquement, si  $X$  et  $Y$  sont linéairement corrélées ( $r$  est proche de 1 ou -1) un nuage homogène s'étend généralement le long d'une droite, d'autant plus étroitement que  $r$  est proche de 1 ou -1, de sorte que  $X$  et  $Y$  covarient linéairement. Si le nuage n'est pas homogène, par exemple si il est composé de plusieurs sous-échantillons distincts, ou si il comporte un ou plusieurs points aberrants l'alignement selon une droite peut être fictif. Considérons par exemple les deux nuages fictifs suivants qui diffèrent seulement par la présence dans le second du point particulier  $L$  :



Pour le premier nuage on a  $\bar{x} = 2,50$  ;  $\bar{y} = 2,40$  ;  $\sigma(x) = 1,02$  ;  
 $\sigma(y) = 1,02$  ;  $cov(x, y) = 0,10$  et  $r = 0,10$  : cette faible valeur n'a rien  
d'étonnant comme on le verra dans le paragraphe suivant.

Pour le second nuage on a  $\bar{x} = 3,36$  ;  $\bar{y} = 3,27$  ;  $\sigma(x) = 2,90$  ;  
 $\sigma(y) = 2,93$  ;  $cov(x, y) = 7,63$  et  $r = 0,90$  : cette forte valeur qui  
suggère un alignement linéaire est seulement due à la présence éloignée  
du point  $L$ .

La conclusion est que si la **covariation linéaire (liaison par une relation linéaire) implique la corrélation linéaire** ( $r$  proche de 1 ou -1), la **corrélation linéaire n'implique pas toujours la covariation linéaire.**

## 5) Corrélacion et indépendance.

Quand les variables sont indépendantes, nous avons vu que la covariance est proche de 0, ce qui entraîne que  $r$  est lui aussi proche de 0. Dans ce cas, on dit que les variables sont **non-corrélées linéairement** : l'indépendance entraîne la non-corrélation linéaire.

La réciproque est fautive : une valeur de  $r$  proche de 0 n'entraîne pas toujours l'indépendance.

Par exemple, dans le cas du nuage du paragraphe "liaison fonctionnelle", la covariance est proche de 0, pareil pour  $r$ , alors que les variables  $X$  et  $Y$  sont fortement liées.

En conclusion, la **non corrélation linéaire** (un  $r$  proche de 0) équivaut à dire que  **$X$  et  $Y$  ne sont pas liées par une relation linéaire**, et non que  $X$  et  $Y$  ne sont pas liées.

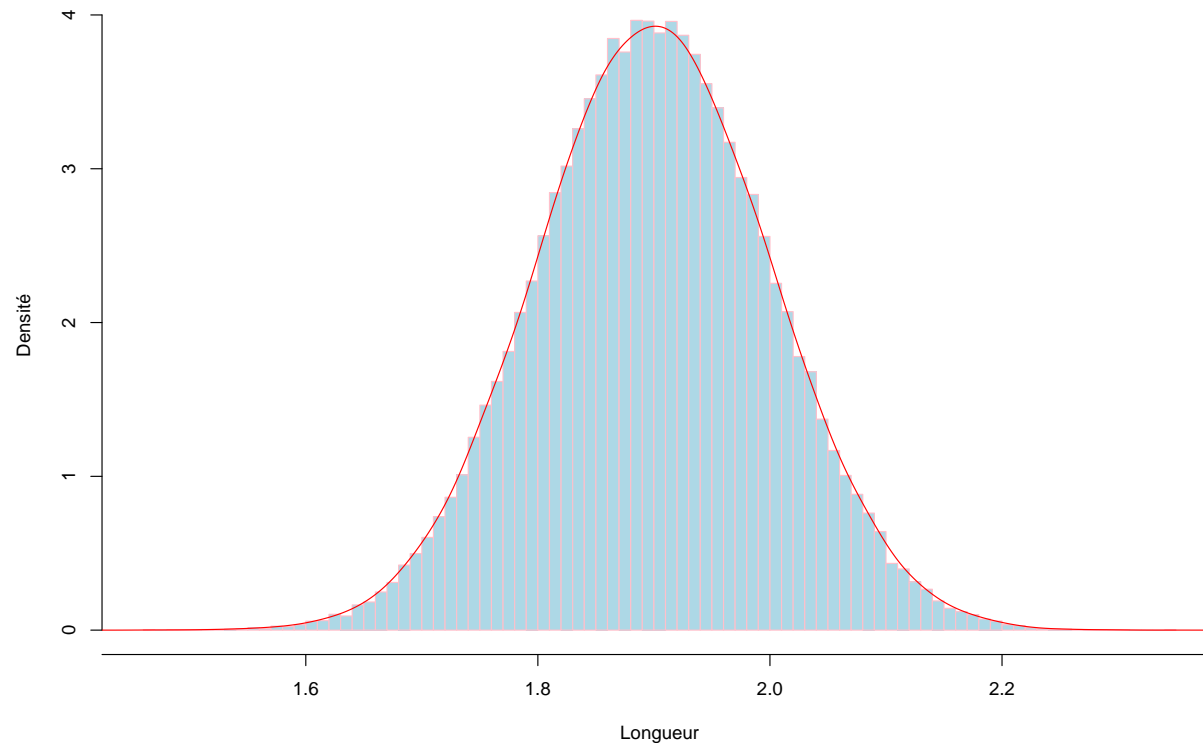
# Chapitre 11. La distribution gaussienne

## 1) Introduction

La **distribution gaussienne** ou **loi normale** permet de modéliser de nombreux phénomènes observés dans la pratique.

*Exemple : Longueur (en cm) de  $n = 10000$  (petits) poissons.*

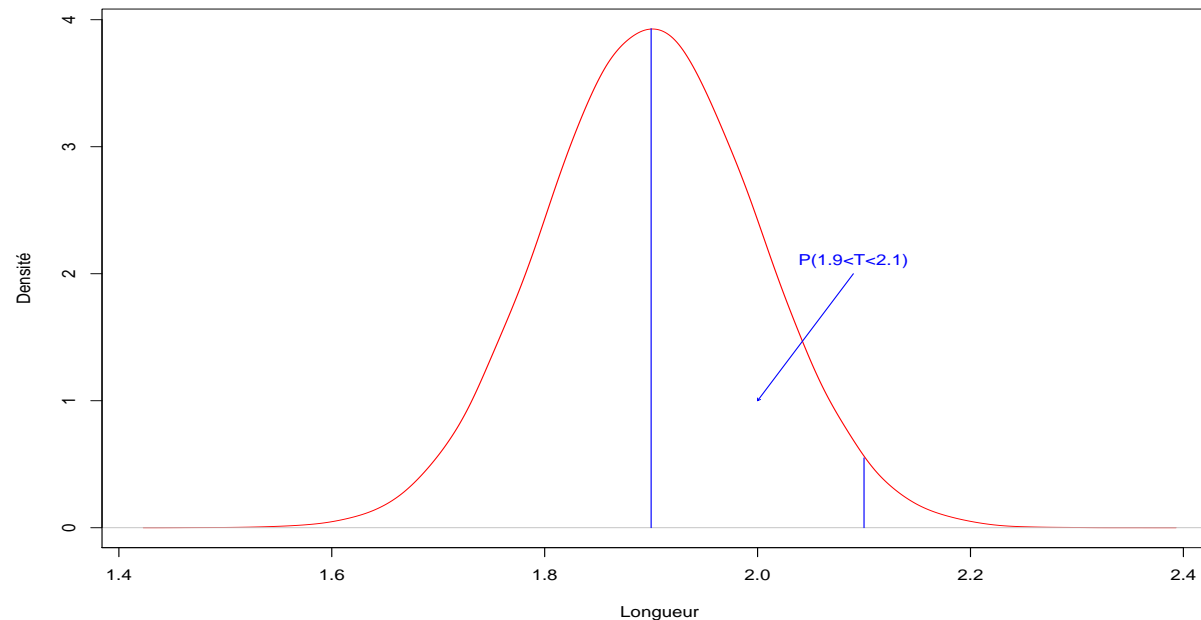
Après avoir relevé la longueur de chaque poisson, on trace l'histogramme des densités (de fréquence) avec  $K = 100$  modalités-intervalles.



On remarque que l’histogramme ci-dessus a la forme d’une courbe en “cloche” (courbe rouge). On dit que la variable  $X = \text{“longueur”}$  suit une loi normale. Il est alors possible de faire des calculs de probabilité (plus de détails sur ce thème en L3).

Par exemple, quelle est la probabilité que la longueur d'un poisson pris au hasard soit supérieure à 2 cm ? inférieure à 1.8 cm ? ou comprise entre 1.9 cm et 2.1 cm ? Ces probabilités sont notées, respectivement,  $P(X \geq 2)$ ,  $P(X \leq 1.8)$  et  $P(1.9 \leq X \leq 2.1)$ .

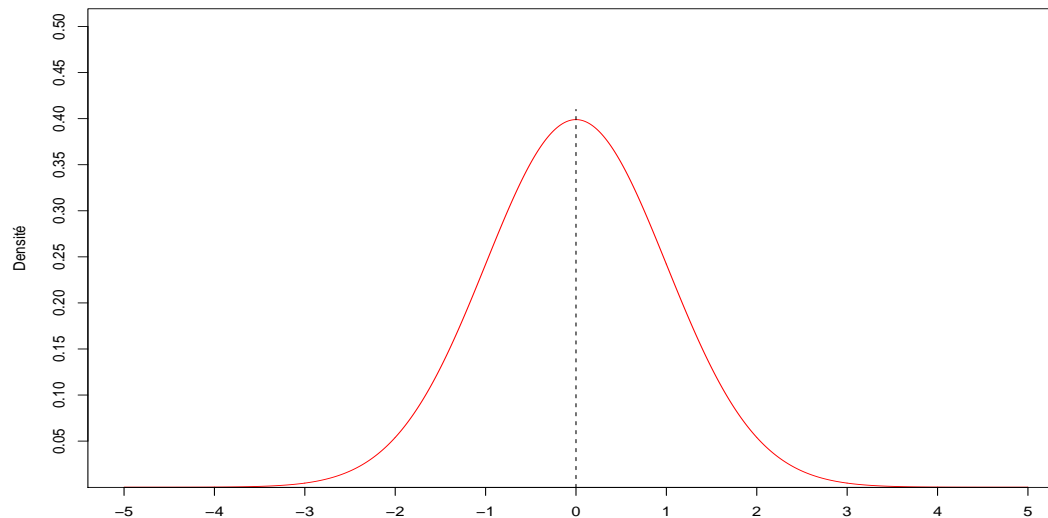
Pour cette dernière probabilité, elle est en fait égale à la **surface sous la courbe** entre 1.9 et 2.1



## 2) Loi normale centrée réduite : $Z \sim \mathcal{N}(0, 1)$

Pour calculer les probabilités d'intervalles on se ramène à une table déjà construite donnant les valeurs des “quantiles” d'une **loi normale centrée réduite**.

C'est la loi d'une variable de moyenne (théorique) égale à 0 et de variance (théorique) égale à 1. Par convention, on désignera cette variable par  $Z$ . Sa densité est représentée par :



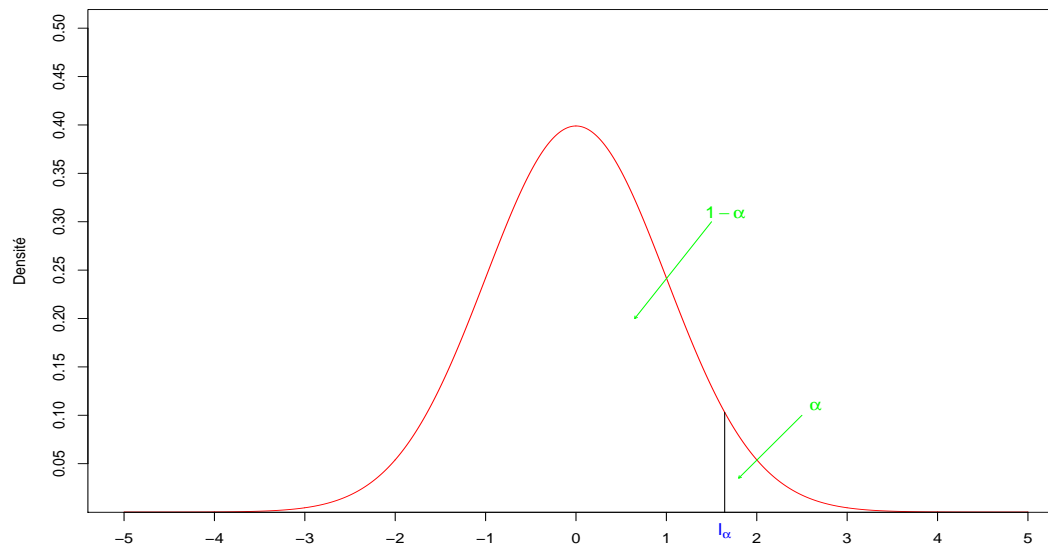


On notera  $l_\alpha$  le réel positif tel que  $P(Z \geq l_\alpha) = \alpha$  où  $\alpha$  est une probabilité comprise entre 0 et 0.5.

*Remarque :  $l_\alpha$  est le quantile d'ordre  $1 - \alpha$  :  $l_\alpha = q_{1-\alpha}$ . Ainsi,*

$$P(Z \leq l_\alpha) = 1 - \alpha$$

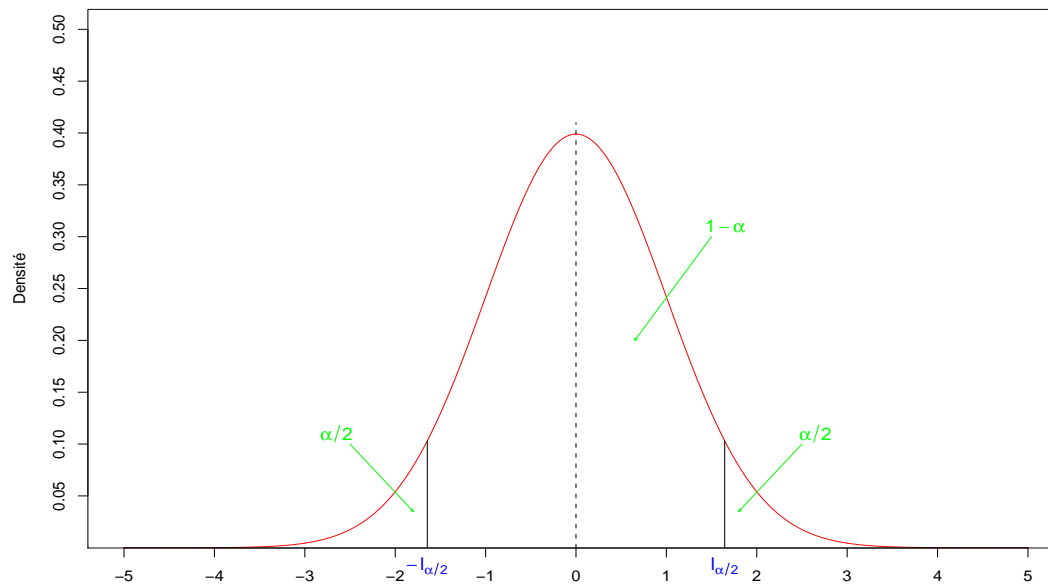
Exemple :  $\alpha = 0.05$  alors  $P(Z \leq l_{0.05}) = 0.95$



## Calcul d'une probabilité entre 2 valeurs symétriques

$$P(-l_{\alpha/2} \leq Z \leq l_{\alpha/2}) = 1 - \alpha \text{ et } P(Z \geq l_{\alpha/2}) = P(Z \leq -l_{\alpha/2}) = \alpha/2$$

Exemple :  $\alpha = 0.04$  alors  $P(-l_{0.02} \leq Z \leq l_{0.02}) = 0.96$



**Extrait de la table de la loi normale  $\mathcal{N}(0, 1)$  :**

P		+0.001	+0.002	+0.003	+0.004	+0.005	+0.006	+0.007	+0.008	+0.009
0	Inf	3.0902	2.8782	2.7478	2.6521	2.5758	2.5121	2.4573	2.4089	2.3656
0.0100	2.3263	2.2904	2.2571	2.2262	2.1973	2.1701	2.1444	2.1201	2.0969	2.0749
0.0200	2.0537	2.0335	2.0141	1.9954	1.9774	1.9600	1.9431	1.9268	1.9110	1.8957
0.0300	1.8808	1.8663	1.8522	1.8384	1.8250	1.8119	1.7991	1.7866	1.7744	1.7624
0.0400	1.7507	1.7392	1.7279	1.7169	1.7060	1.6954	1.6849	1.6747	1.6646	1.6546
0.0500	1.6449	1.6352	1.6258	1.6164	1.6072	1.5982	1.5893	1.5805	1.5718	1.5632
0.0600	1.5548	1.5464	1.5382	1.5301	1.5220	1.5141	1.5063	1.4985	1.4909	1.4833
0.0700	1.4758	1.4684	1.4611	1.4538	1.4466	1.4395	1.4325	1.4255	1.4187	1.4118
0.0800	1.4051	1.3984	1.3917	1.3852	1.3787	1.3722	1.3658	1.3595	1.3532	1.3469
0.0900	1.3408	1.3346	1.3285	1.3225	1.3165	1.3106	1.3047	1.2988	1.2930	1.2873
0.1000	1.2816	1.2759	1.2702	1.2646	1.2591	1.2536	1.2481	1.2426	1.2372	1.2319

## Exemples de calculs de probabilités et de quantiles :

$$P(-2.5758 \leq Z \leq 2.5758) = 0.99$$

$$P(-1.96 \leq Z \leq 1.96) = 0.95$$

$$P(-1.6449 \leq Z \leq 1.6449) = 0.90$$

$$P(Z \geq 2.5758) = 0.005$$

$$P(Z \geq 1.96) = 0.025$$

$$P(Z \geq 1.6449) = 0.05$$