المدرسة العليا للتربية والتكوين
ECOLE SUPERIEURE D'EDUCATION ET DE FORMATION

جامعة شعيب الدكالي
Université Chouaïb Doukkali

Filière : Mathématiques Année Universitaire 2025/2026

Polycopié de Cours

Statistique descreptive et Initiaion à la probabilité

Pr. El Moutawakil Driss

0 Table des matières

Première partie

Statistique descriptive

1 Vocabulaire de la Statistique

1.1 Définitions et Notations

1.1.1 Statistique

Définition 1.1.1

La **statistique** désigne à la fois un ensemble de données et l'ensemble des activi tés consistant à collecter ces données, à les traiter et à les interpréter.

1.1.2 Statistiques

Définition 1.1.2

Les **statistiques**, l'ensemble des données numériques, interviennent pratiquement dans tous les domaines d'activité : gestion financière (états, banques, assurances, entreprises...), démographie, contrôles de qualité, études de marché, sciences expé rimentales (biologie, psychologie...)...

1.1.3 Statistique descriptive

Définition 1.1.3

Le traitement des données, pour en dégager un certain nombre de renseignements qualitatifs ou quantitatifs à des fins de comparaison, s'appelle la statistique des criptive.

1.1.4 Statistique inférentielle

Définition 1.1.4

Un autre but de la statistique consiste à extrapoler à partir d'un échantillon de la population à étudier, le comportement de la population dans son ensemble (son dages, contrôle de qualité comportant un test destructif...) C'est la statistique inductive ou encore appelée statistique inférentielle.

1.1.5 Population, individus et échantillon

Définition 1.1.5

La **population statistique** est l'ensemble des éléments sur lesquels porte l'étude.

Exemple 1.1.1 *En statistique, le terme de population est plus général et peut désigner des humains, mais aussi des objets, des villes, des pays, des entreprises, etc..*

Définition 1.1.6

On appelle **individu** statistique ou unité statistique tout élément de la population.

Exemple 1.1.2 *Si on effectue un contrôle sur les pièces défectueuses d'une production de 100 chemises d'une entreprise. La population est l'ensemble des 100 chemises et chacune des chemises constitue un individu.*

Remarque 1.1.1 *Les individus doivent appartenir à un ensemble bien déterminé et leur définition doit être parfaitement claire.*

Définition 1.1.7

Un **échantillon** de taille n est un sous-ensemble formé de n individus de la population.

Remarque 1.1.2 *La notion d'échantillon est fondamentale, car, en règle générale, la population entière n'est pas disponible ou observable. Dans ce cas, seul un échantillon est étudié et les résultats obtenus sont extrapolés à la population.*

Exemple 1.1.3 *Lorsqu'une entreprise souhaite faire une étude de marché pour un projet de lancement d'un nouveau produit, elle interroge seulement un échantillon de Marocains, et non tous les marocains, soit plus de* 35 *millions d'individus.*

1.1.6 Variable statistique et modalités

1.1.6.1 Notion de caractère ou variable statistique

Chaque individu d'une population peut être décrit relativement à un ou plusieurs caractères ou variables statistiques.

Définition 1.1.8

Une **variable statistique** est un aspect particulier des individus auxquels on s'intéresse, une caractéristique qui peut varier d'un individu à l'autre. Elle porte aussi le nom de caractère.

Exemple 1.1.4 *Les caractères étudiés peuvent être par exemple l'âge, le sexe, le poids, la masse, le nombre d'enfants d'une famille, le niveau d'étude, la marque ou le prix d'un véhicule, etc....*

1.1.6.2 Modalités

Définition 1.1.9

Les modalités d'une variable statistique sont les différentes valeurs que peut prendre celle-ci.

Exemple 1.1.5

1. *Si la variable est l'âge, les modalités sont l'ensemble des âges des individus observés : 19 ans, 20 ans, 21 ans, etc.*
2. *Si le caractère étudié est la couleur, les modalités du caractère seront des couleurs : bleu, vert, blanc, etc.*
3. *Si la variable est la situation familiale, les modalités du caractère seront : célibataire, marié, divorcé.*

Remarque 1.1.3 *À chaque individu doit être associée une modalité unique de la variable, c' est-à-dire "au plus une" et "au moins une".*

D'un point de vue mathématique, une variable statistique peut donc être définie comme suit

Définition 1.1.10

Une variable statistique, notée X, est une application définie sur une population statistique et à valeurs dans l'ensemble des modalités. Elle associe à chaque individu w_i de l'ensemble de départ (la population Ω) une unique modalité de l'ensemble d'arrivée notée $X(w_i)$ ou m_i.

1.1.7 Les variables qualitatives

Définition 1.1.11

Une variable statistique est dite de nature qualitative si ses modalités ne sont pas mesurables. Les modalités d'une variable qualitative sont les différentes catégories d'une nomenclature.

Remarque 1.1.4 *Une variable est qualitative lorsque l'ensemble des modalités n'est pas un ensemble de nombres.*

Exemple 1.1.6

1. *La profession (professeur, médecin, ...)*
2. *La situation matrimoniale (célibataire, veuf, ...)*
3. *Le sexe : (masculin, féminin)*
4. *La marque d'une voiture (Peugeot, Renault, Citroën, ...)*
5. *Les mois (Janvier, février, ...)*

Remarque 1.1.5 *Les modalités d'une variable qualitative peuvent être classées sur deux types d'échelle : nominale ou ordinale. À ces deux types d'échelle correspondent deux types de variables qualitatives.*

1.1.7.1 Les variables qualitatives nominales

Définition 1.1.12

Une variable statistique qualitative est dite définie sur une échelle nominale si ses modalités ne sont pas naturellement ordonnées.

Remarque 1.1.6 *L'ordre et l'origine des modalités, d'une variable qualitative nominale, sont arbitraires.*

Exemple 1.1.7 *Les différentes catégories de la variable nominale Professions et catégories socioprofessionnelles (CSP) :*

- *Agriculteurs exploitants*
- *Artisans, commerçants et chefs d'entreprise*
- *Professions intermédiaires*
- *Cadres et professions intellectuelles supérieures*
- *Employés*
- *Ouvriers*
- *Retraités*
- *Autres personnes sans activité professionnelle*

Dans cet exemple, il n'y a pas d'ordre naturel entre les huit catégories, ou modalités, qui sont de simples étiquettes, la variable qualitative (CSP) est définie sur une échelle nominale.

1.1.7.2 Les variables qualitatives ordinales

Définition 1.1.13

Une variable statistique qualitative est dite définie sur une échelle ordinale si l'ensemble de ses modalités peut être doté d'une relation d'ordre. C'est-à-dire que l'on peut opérer un classement de l'ensemble des catégories, de la plus petite à la plus grande (ou, inversement, de la plus grande à la plus petite).

Exemple 1.1.8

1. *Les échelles d'opinion : • Très précis • Précis • Assez précis • Peu précis • Imprécis.*
2. *Le rang : • un peu • moyen • beaucoup • énormément, ...*
3. *Modalités : • très souvent • souvent • parfois • rarement • jamais.*

1.1.8 Les variables quantitatives

Définition 1.1.14

Une variable statistique est dite de nature quantitative si ses modalités sont mesurables. Les modalités d' une variable quantitative sont des nombres liés à l'unité choisie, qui doit toujours être précisée.

Remarque 1.1.7 *Une variable est quantitative lorsque l'ensemble de ses modalités est un ensemble de nombres. Il existe deux types de variables quantitatives : les variables discrètes et les variables continues.*

1.1.8.1 Variables quantitatives discrètes

Définition 1.1.15

Une variable statistique quantitative est dite discrète si l'ensemble de ses modalités est un ensemble fini ou dénombrable. Ainsi, l'ensemble des modalités peut être donné sous la forme d'une liste de nombres, $M = \{x_1, x_2, \cdots, x_i\}$, finie ou infinie.

Exemple 1.1.9 *Le nombre d'enfants par ménage, âge, nombre d'habitants ...*

1.1.8.2 Variables quantitatives continues

Définition 1.1.16

Une variable statistique quantitative est dite continue si l'ensemble de ses modalités n'est pas dénombrable. Ainsi, une variable continue peut prendre toutes les valeurs d'un intervalle réel. Ces valeurs sont regroupées dans des intervalles de valeurs numériques appelés classes (elles constituent les modalités de la variable).

Remarque 1.1.8 *Dans un tableau, on reconnaît donc une variable quantitative continue au fait que les valeurs de la variable ont été regroupées en classes.*

Exemple 1.1.10

1. *Le tableau suivant indique la structure par âges d'une population :*

Âge (par ans)	$[4, 6[$	$[6, 8[$	$[8, 10[$	$[10, 12[$	$[12, 14[$	$[14, 14[$
Effectifs (n_i)	*10*	*18*	*5*	*7*	*23*	*2*

2. *Le revenu d'individus, nombre de salariés, chiffre d'affaires ...*

Remarque 1.1.9 *La distinction entre variable discrète et variable continue est parfois arbitraire.*

Les classes sont notées $[a_i, b_i[$. L'intervalle est fermé à gauche et ouvert à droite (à part pour la dernière classe) : il inclut toutes les valeurs de la variable supérieures ou égales à la borne inférieure et strictement inférieures à la borne supérieure b_i.

Définition 1.1.17

On appelle

1. Amplitude de la classe $[a_i, b_i[$ le réel noté A_i représentant la longueur de l'intervalle et défini par : $A_i = b_i - a_i$.
2. Le centre de classe de la classe $[a_i, b_i[$ est le réel noté x_i représentant le milieu de l'intervalle et donné par : $x_i = (a_i + b_i)/2$.

Définition 1.1.18

On appelle discrétisation le découpage en classes d'une série statistique quantitative.

Remarque 1.1.10 *Le nombre de classes ne doit pas être trop petit, perte d'informations, ni trop grand, le regroupement en classes est alors inutile et de plus, certaines classes pourraient avoir des effectifs trop faibles. En général, le nombre de classes est compris entre 5 et 20, il dépend du nombre n d'observations et de l'étalement des données. La formule de Sturges donne une valeur approximative du nombre k de classes :*

$$k \approx 1 + 3,222 \log_{10}(n).$$

2 Séries statistiques à une dimension

2.1 Les notions de base du calcul statistique

Les données statistiques sont issues de données brutes présentées sous forme de tableaux statistiques dans lesquels sont indiqués les effectifs et/ou les fréquences.

2.1.1 Effectifs et fréquences

Définition 2.1.1

Le nombre d'individus présentant la modalité m_i (variable qualitative) ou x_i (variable quantitative discrète) ou une modalité incluse dans $[e_i, e_{i+1}[$ (variable quantitative continue) s'appelle l'effectif. Il est noté n_i.
L'effectif total de la population, noté n, est la somme des effectifs

$$n = n_1 + n_2 + \cdots + n_k = \sum_{i=1}^{k} n_i,$$

avec k le nombre de modalités de la variable.

Définition 2.1.2

La fréquence associée à une modalité, ou à un ensemble de modalités regroupées en classes, indique la proportion d'individus présentant cette modalité, ou cet ensemble de modalités, par rapport à l'ensemble des individus. La fréquence associée à la i^{e} modalité, ou i^{e} classe, est notée f_i. Par définition :

$$f_i = \frac{n_i}{n}.$$

C'est un nombre compris entre 0 et 1, ou 0% et 100% et la somme des fréquences est égale à 1 ou 100%.

2.1.2 Distribution des effectifs et des fréquences cumulés

Cette section concerne les variables quantitatives pour lesquelles le tableau statistique est réalisé, les modalités étant ordonnées dans l'ordre croissant.

- **Cas 1 : variables quantitatives discrètes :**

Définition 2.1.3

Soit X une variable quantitative discrète, on appelle effectif cumulé croissant à la modalité x_i, noté n_icc, le nombre d'individus statistiques pour lesquels X est inférieur ou égal à x_i. On a

$$n_1cc = n_1 \text{ et } n_icc = n_1 + n_2 + \cdots + n_i = \sum_{k=1}^{i} n_k.$$

Définition 2.1.4

Soit X une variable quantitative discrète, on définit la fréquences cumulées croissantes, notée f_icc, représentant la proportion d'individus statistiques pour lesquels X est inférieur ou égal à x_i. On a :

$$f_1cc = f_1 \text{ et } f_icc = f_1 + f_2 + \cdots + f_i = \sum_{k=1}^{i} f_k, \text{ ou encore } f_icc = \frac{n_icc}{n}.$$

Remarque 2.1.1 *Il est également possible de cumuler les effectifs et les fréquences dans le sens décroissant.*

- **Cas 2 : variables quantitatives continues :**

Dans le cas d'une variable quantitative continue, les données sont groupées en classes $[a_i, b_i[$, et on définit, de même que pour une variable discrète, n_icc le nombre d'individus statistiques pour lesquels X est inférieur ou égal à b_i, et f_icc la proportion d'individus statistiques pour lesquels X est inférieur ou égal à b_i.

⟨exp1.2.1⟩ **Exemple 2.1.1** *La répartition des salaries mensuels d'une entreprise est donnée par le tableau suivant :*

Salaire	n_i	n_icc	f_i	f_icc
$[4000, 5000[$	*100*	*100*	*0,2739*	*0,2739*
$[5000, 6000[$	*120*	*220*	*0,3287*	*0,6026*
$[6000, 7000[$	*95*	*315*	*0,2602*	*0,8628*
$[7000, 8000]$	*50*	*365*	*0,1369*	*1*
Total	*365*		*1*	

2.2 Les représentations graphiques

Les graphiques constituent une synthèse visuelle indispensable de l'information contenue dans le tableau statistique. Le choix du type du graphique utilisé dépendent de la nature des données et de la variable étudier.

2.2.1 Graphiques pour variables qualitatives

Les variables qualitatives -nominales ou ordinales- peuvent être représentées au choix à l'aide d'un diagramme à bandes ou à l'aide d'un diagramme circulaire.

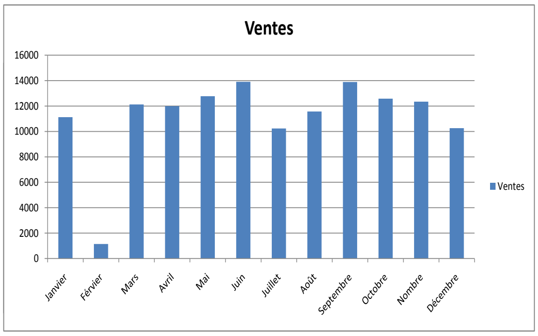
2.2.1.1 Diagramme à bandes

Définition 2.2.1

Dans un diagramme à bandes, une bande verticale (ou horizontale) est associée à chaque modalité. La distance entre chaque bande est constante. La largeur de chacune des bandes est la même, et sa hauteur (ou longueur) est égale à l'effectif ou à la fréquence de la modalité correspondante. Par conséquent, la surface d' une bande est proportionnelle à l' effectif (et donc à la fréquence) de la modalité associée.

Exemple 2.2.1 *On considère les ventes mensuelles réalisées par une entreprise au cours des douze mois de l'année 2021 :*

Mois	Ventes
Janvier	11120
Férvier	1140
Mars	12123
Avril	11993
Mai	12769
Juin	13904
Juillet	10234
Août	11567
Septembre	13893
Octobre	12576
Nombre	12345
Décembre	10254



2.2.1.2 Diagramme à secteur circulaire

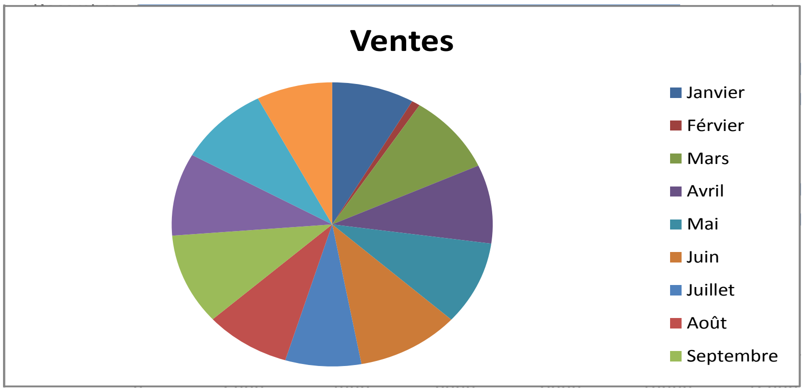
Définition 2.2.2

Un diagramme à secteurs circulaire est un graphique qui divise un disque en secteurs angulaires dont les angles aux centres sont proportionnels aux effectifs (ou aux fréquences) de chaque modalité. L'angle au centre α_i, en degrés, associé à la modalité m_i d' effectif n_i, est égal à :

$$\alpha_i = \frac{n_i}{n} \times 360 = f_i \times 360.$$

Dans ce graphique, l'aire du disque matérialise l' effectif total (ou la somme des fréquences).

Exemple 2.2.2 *On donne Dans le diagramme circulaire relatif à l'exemple précédent des ventes mensuelles réalisées par une entreprise au cours des douze mois de l'année 2021, l'angle au centre associé à chaque modalité est évalué ci-dessous :*



2.2.2 Graphiques pour variables quantitatives

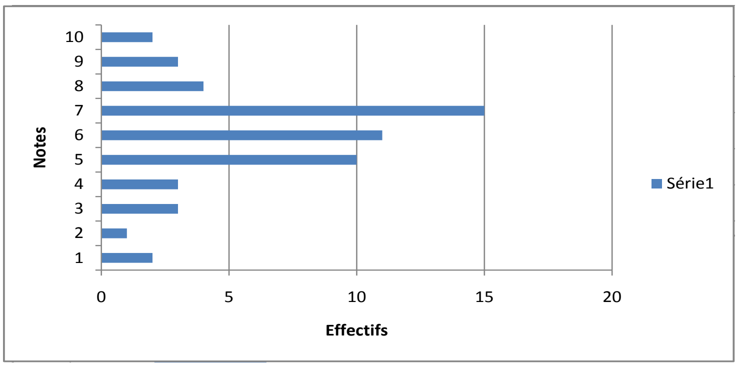
La représentation graphique d'une variable quantitative dépend de sa nature : discrète ou continue.

2.2.2.1 Cas discret : diagramme en bâtons

Définition 2.2.3

On appelle diagramme en bâtons un graphique qui à chaque modalité d' une variable quantitative discrète associe un segment (bâton) dont la hauteur est proportionnelle à l' effectif (ou à la fréquence).

Exemple 2.2.3 *On considère le diagramme en bâtons suivant qui représente la répartition des notes d'un contrôle des élèves d'une classe :*



2.2.2.2 Cas continue : Histogramme et polygone de fréquence

2.2.2.2.1 Histogramme et propriétés

Définition 2.2.4

Un histogramme est un diagramme composé de rectangles contigus dont les aires sont proportionnelles aux effectifs (ou aux fréquences) et dont les bases sont déterminées par les intervalles de classes.

Lors de la réalisation d'un histogramme, il est indispensable de distinguer deux cas.

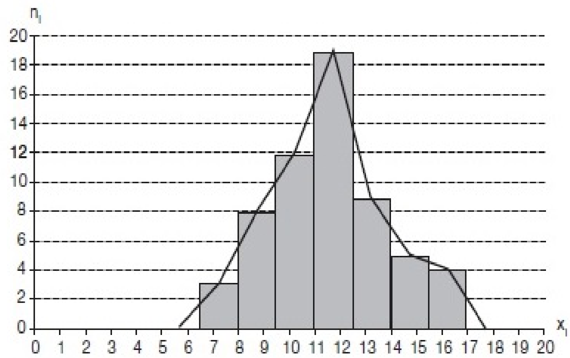
1. Si les amplitudes de classes sont égales, la hauteur des rectangles correspondra aux effectifs (ou aux fréquences) des classes.
2. Si les amplitudes sont différentes, afin de constituer l'histogramme, il est nécessaire de :
   — calculer, pour chaque classe, l' amplitude A_i.
   — calculer la densité $d_i = n_i/A_i$ pour un histogramme des effectifs, et $d_i = f_i/A_i$ pour un histogramme des fréquences.
   — affecter à chaque rectangle une hauteur proportionnelle à la densité d_i de la classe correspondante.
   — La hauteur des rectangles correspondra aux effectifs corrigés définies par $n_ic = d_i \times \min(A_i)$, avec $d_i = n_i/A_i$ ($\min(A_i)$ est l'amplitude minimal des classes). De même, il est possible de retenir comme hauteur la fréquence corrigée $f_ic = d_i \min(A_i)$, avec $d_i = f_i/A_i$ dans le cas d'un histogramme des fréquences.

2.2.2.2.2 Polygone des effectifs

Il permet de représenter la distribution sous la forme d'une courbe, quand les amplitudes de classes sont égales. Il est obtenu en joignant, par des segments de droite, les milieux des côtés supérieurs de chaque rectangle de l'histogramme. Pour fermer ce polygone, on ajoute à chaque extrémité une classe de fréquence nulle.

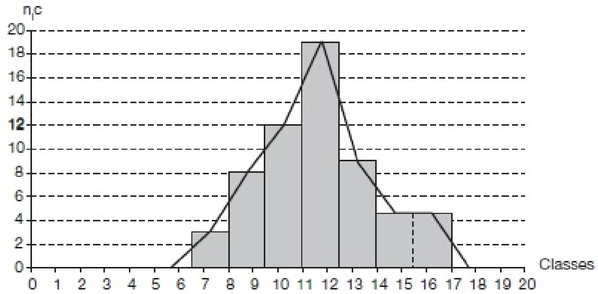
Exemple 2.2.4 *On considère la distribution statistique suivante correspondant à l'ancienneté du personnel cadre dans l'entreprise, exprimée en années :*

Classes	Effectifs
[6,5 ; 8[	3
[8 ; 9,5[	8
[9,5 ; 11[	12
[11 ; 12,5[	19
[12,5 ; 14[	9
[14 ; 15,5[	5
[15,5 ; 17[	4



Exemple 2.2.5 *Modifions légèrement l'exemple précédent en regroupant les deux dernières classes en une seule. Ce regroupement permet de traiter le cas de classes d'amplitudes différentes, puisque ainsi la dernière classe est d'amplitude 3 contre 1,5 pour toutes les autres classes.*

Classes	Effectifs	A_i	d_i	n_ic
[6,5 ; 8[	3	1,5	2	3
[8 ; 9,5[	8	1,5	5,33	8
[9,5 ; 11[	12	1,5	8	12
[11 ; 12,5[	19	1,5	12,67	19
[12,5 ; 14[	9	1,5	6	9
[14 ; 17[	9	3	3	4,5



2.3 Caractéristiques d'une distribution à un caractère

Une série de données peut être résumée par quelques valeurs numériques appelées caractéristiques des séries statistiques.

2.3.1 Les caractéristiques de position

Les caractéristiques de position ou de tendance centrale de la série statistique à une variable sont les principaux paramètres qui permettent de résumer une série statistique d'observations et d'éclairer sur la position du noyau (centre) de la série. Nous présenterons ici le mode, la moyenne, la médiane.

2.3.1.1 Mode et classe modale

2.3.1.1.1 Mode ou valeur dominante Mo

Définition 2.3.1

Le mode (ou valeur dominante) est la valeur de la variable qui a l'effectif (ou la fréquence) le plus grand. Graphiquement, c'est l'élément de la plus grande hauteur. On le note Mo.

Remarque 2.3.1

— *Le mode peut être calculé pour tous les types de variables (qualitative et quantitative).*

— *Le mode n'existe pas toujours et quand il existe, il n'est pas toujours unique.*

Exemple 2.3.1 (Cas discret) *On considère trois séries d'observations des notes de 6 étudiants pour trois matières différentes.*
Matière I : 6 observations classées par ordre croissant, 2, 3, 7, 10, 13, 14.
pas de mode.

Matière II : 6 observations classées par ordre croissant, 4, 4, 10, 18, 19, 19.
deux modes 4 et 19.
Matière III : 6 observations classées par ordre croissant, 4, 4, 10, 17, 18, 19.
Le mode est 4.

2.3.1.1.2 Classe modale

Une variable statistique continue est définit par un intervalle, ce qui signifie qu'il n y a pas une valeur précise!!, on parle ici d'une classe modale.

Définition 2.3.2

La classe modale est la classe d'effectif le plus élevé ou de fréquence la plus élevée

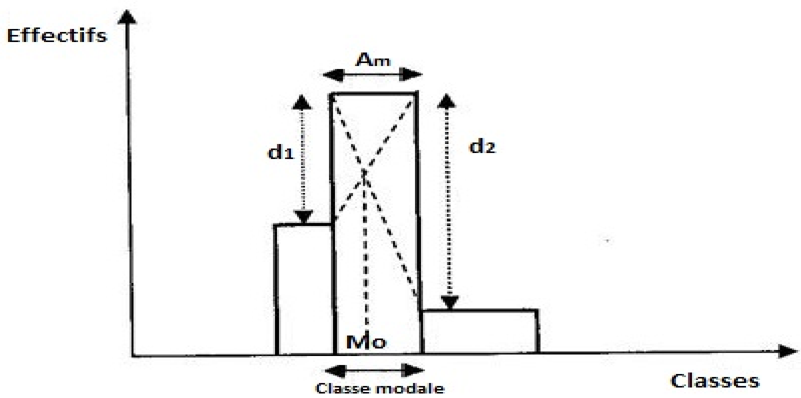
Exemple 2.3.2 *Si en reprend le tableau de l'exemple 2.1.1 on voit que la classe modale est la classe* $[5000, 6000[$.

Remarque 2.3.2 *Il est possible de déterminer la valeur précise du mode, via la relation suivante :*

$$Mo = x_{p-1} + A_m \times \frac{d_1}{d_1 + d_2} \tag{2.1}$$ 1.14

où x_{p-1} désigne la valeur de l'extrémité inférieure de la classe modale, A_m l'amplitude de cette même classe, d_1 la différence entre l'effectif de la classe modale et l'effectif de la classe précédente et d_2 la différence entre l'effectif de la classe modale et l'effectif de la classe suivante. Si en considère le tableau de l'exemple 2.1.1, la valeur du mode est donnée par :

$$Mo = 5000 + 1000 \times \frac{20}{20 + 25} = 5444,4444.$$



Remarque 2.3.3 *Dans le cas où les classes sont d'amplitudes différentes, il convient de corriger les effectifs ou les fréquences préalablement à la détermination du mode en utilisant la formule (2.1), avec d_1 et d_2 désignent alors des effectifs corrigés.*

2.3.1.2 Moyenne

Définition 2.3.3

La moyenne arithmétique est la somme des valeurs observées rapportée au nombre d'observations. Elle se note $\overline{x}$.
• La moyenne arithmétique simple (données en tableau brut). Soit $x_1, x_2, \cdots, x_n$ les n observations de la variable X : la moyenne arithmétique se note

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

• La moyenne arithmétique pondérée de r réels (distincts) $x_1, x_2, \cdots, x_r$ (données en tableau statistique), affectés respectivement des coefficients n_i, tels que $\sum_{i=1}^{r} n_i = n$, se note

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{r} n_i x_i = \sum_{i=1}^{r} f_i x_i.$$

Exemple 2.3.3

1. *Les notes d'un examen de Statistiques sont réparties comme suit :*

$$10 \;\; 10 \;\; 10 \;\; 11 \;\; 12 \;\; 12 \;\; 15 \;\; 15 \;\; 15 \;\; 15 \;\; 17 \;\; 17 \;\; 17 \;\; 17 \;\; 17.$$

la moyenne arithmétique (simple) dans ce cas (variable discrète) est

$$\overline{x} = \frac{1}{15}\left(10 + 10 + 10 + 11 + 12 + 12 + 15 + 15 + 15 + 15 + 17 + 17 + 17 + 17 + 17\right).$$

2. *Le tableau suivant donne la répartition de notes d'examen de Mathématiques.*

Note	*11*	*15*	*16*	*17*
Effectif	*5*	*4*	*6*	*2*

La moyenne arithmétique (pondérée) dans ce cas (variable discrète) est

$$\overline{x} = \frac{1}{17}\left(11 \times 5 + 15 \times 4 + 16 \times 6 + 17 \times 2\right).$$

3. *Le tableau suivant indique la structure par âges d'une population :*

Âge (par ans)	$[4,6[$	$[6,8[$	$[8,10[$	$[10,12[$	$[12,14[$	$[14,16]$
Effectifs (n_i)	*10*	*18*	*5*	*7*	*23*	*2*

Pour calculer la moyenne arithmétique (pondérée) dans ce cas (variable continue), il faut préalablement calculer les centres de classes x_i. On obtient ainsi est

$$\overline{x} = \frac{1}{65}\left(10 \times 5 + 18 \times 7 + 5 \times 9 + 7 \times 11 + 23 \times 13 + 2 \times 15\right).$$

2.3.1.3 Médiane

Définition 2.3.4

La médiane, notée Me, est la plus petite valeur de la série pour laquelle le nombre d'observations inférieures ou égales à cette valeur représente au moins 50% de l'effectif total de la série.
Convention : dans le cas d'une série discrète comportant un nombre pair d'observations, la médiane n'est pas nécessairement une valeur observée.

Pour calculer la médiane, deux cas sont à distinguer selon la façon dont les données ont été recueillies.
Cas 1 : La médiane d'une série de données brutes :

— Tout d'abord la série doit être classée dans l'ordre croissant des valeurs.

— Si la série brute comporte un nombre impair d'observations, noté $n = 2p+1$, la médiane est la valeur centrale de la série (ordonnée en sens croissant), donc la $(p+1)^{ime}$ observation ($Me = x_{p+1}$).

— Pour une série ayant un nombre pair $n = 2p$ de données, on peut prendre pour valeur médiane, indifféremment l'une ou l'autre des valeurs centrales ou n'importe quelle valeur intermédiaire entre ces deux valeurs, par exemple, la moyenne arithmétique de ces deux valeurs $\left(Me = \dfrac{x_p + x_{p+1}}{2}\right)$, mais, dans ces conditions, ce n'est pas une valeur observée.

Exemple 2.3.4

— *On considère la série d'observation des notes de 5 étudiants pour le module de Probabilités et Statistique : 11, 9, 18, 10, 13.*
Classons tout d'abord les modalités par ordre croissant. Dans notre exemple, ces modalités sont au nombre de $n = 5$*, c'est-à-dire un nombre impair, et* $p = 2$*, donc la médiane est la valeur centrale de la série ordonnée, c'est-à-dire la* 3^{e} *observation : 9, 10, 11, 13, 18. La médiane est* $Me = 11$.

— *Modifions légèrement l'exemple précédent en rajoutant un étudiant dont la note s'élève à 14. La série ordonnée s'écrit donc, 9, 10, 11, 13, 14, 18. Dans ce cas, on définit un intervalle médian, donné par [11 , 13], la médiane étant quant à elle égale à* $Me = \dfrac{11+13}{2} = 12$.

Cas 2 : La médiane dans un tableau statistique.
Pour calculer la médiane à partir d'un tableau statistique, il convient de distinguer deux cas :

1. Soit la variable est présentée comme un caractère discret.
2. Soit la variable est présentée comme un caractère continu.

Calcul 1 : Les modalités de la variable sont des valeurs isolées. La médiane est la valeur de la variable à laquelle est associé un effectif cumulé égal à $n/2$, ou une fréquence cumulée égale à $1/2$, n étant effectif total de la population.
Si ces valeurs d'un effectif cumulé égal à $n/2$ ou d'une fréquence cumulée égale à $1/2$ ne

correspondent pas exactement à une valeur du tableau et "tombent entre deux lignes" de la distribution, aucune valeur possible de la variable ne partage alors exactement la population en deux sous-ensembles égaux. Par convention, on retient dans ce cas comme médiane la valeur de la variable immédiatement supérieure.

Exemple 2.3.5 *On considère le tableau suivant*

Note	*11*	*15*	*16*	*17*
Effectif	*5*	*4*	*6*	*2*
Effectif cumulé	*5*	*9*	*15*	*17*

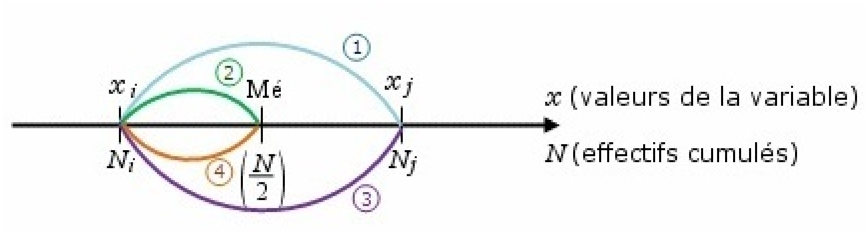
On a : $n = 17$ *; donc :* $n/2 = 8,5$*. Le calcul des effectifs cumulés montre que la moitié de l'effectif (*$n/2 = 8,5$*) se situe entre deux valeurs de la variable : 5 et 9 ans. Par convention, on retiendra comme médiane la valeur de la variable immédiatement supérieure, soit la note 15.*

Calcul 2 : Le calcul de la médiane se fait en deux temps :

— La détermination de la classe médiane (la classe de valeurs de la variable contenant la médiane). Elle est déterminée de la même manière que la médiane dans le cas d'une variable discrète, à partir des effectifs et des fréquences cumulés.

— Calcul de la valeur précise de la médiane à l'intérieur de la classe médiane. La méthode généralement utilisée pour ce faire est celle de l'interpolation linéaire, c'est mathématiquement une application simple du théorème de Thalès.

En effet soit $[x_i, x_j[$ la classe médiane déterminée à l'étape précédente. Appelons respectivement N_i et N_j les effectifs cumulés associés aux deux bornes de cette médiane : x_i et x_j. Le théorème de Thalès permet d'écrire, à partir de :



$$\frac{\text{Segment } 2}{\text{Segment } 1} = \frac{\text{Segment } 4}{\text{Segment } 3}$$

Soit

$$\frac{(Me - x_i)}{(x_j - x_i)} = \frac{\left(\frac{N}{2} - N_i\right)}{(N_j - N_i)}$$

D'où

$$Me = x_i + \frac{(\frac{N}{2} - N_i)(x_j - x_i)}{(N_j - N_i)}$$

Exemple 2.3.6 *Le tableau suivant indique la structure par âges d'une population :*

Âge (par ans)	$[4,6[$	$[6,8[$	$[8,10[$	$[10,12[$	$[12,14[$	$[14,14[$
Effectifs (n_i*)*	*10*	*18*	*5*	*7*	*23*	*2*
Effectif cumulé	*10*	*28*	*33*	*40*	*63*	*65*

Ici $N = 65$ *et* $\frac{N}{2} = 32,5$*, se situe entre 28 et 33. Ainsi la classe médiane est la classe* $[8,10[$*. Détermination de la valeur de la médiane : On* $x_i = 8$*,* $x_j = 10$ *(bornes de la classe médiane), on a aussi :* $N_i = 28$ *et* $N_j = 33$ *(effectifs cumulés associés). Ainsi on obtient*

$$Me = 8 + \frac{(32,5-28)(10-8)}{33-28} = 9,8$$

Remarque 2.3.4 *Le calcul de la valeur de la médiane dans la classe médiane* $[x_i, x_j[$ *peut aussi être calculer en utilisant les fréquences cumulées, par interpolation linéaire, via la relation suivante :*

$$Me = x_i + \frac{(x_j - x_i)\left(\frac{1}{2} - f_i\right)}{f_j - f_i}.$$

avec f_i *et* f_j *sont respectivement les fréquences cumulées associés aux deux bornes de cette médiane :* x_i *et* x_j

Remarque 2.3.5 *La présence de classes d'amplitudes inégales n'affecte pas le calcul de la médiane. Il n'est donc pas nécessaire de corriger les effectifs ou les fréquences pour déterminer la médiane.*

2.3.2 Les caractéristiques de dispersion

Ces caractéristiques quantifient les fluctuations des valeurs observées autour de la valeur centrale et permettent d'apprécier l'étalement de la série. Les principales sont : l'étendue, l'écart-type ou son carré appelé variance et le coefficient de variation.

2.3.2.1 L'étendue

Définition 2.3.5

L'étendue d'une série est la différence entre la plus grande et la plus petite valeur observée. Elle est notée :

$$E = \max(x_i) - \min(x_i).$$

L'étendue permet une approche aisée de la dispersion d'une variable, mais sa signification reste très limitée. Car elle ne tient compte que des valeurs extrêmes de la série, elle ne dépend ni du nombre, ni des valeurs intermédiaires, elle est très peu utilisée dès que le nombre de données dépasse 10.

Exemple 2.3.7 *Si on considère la répartition, des notes d'un examen de Statistiques, suivante :*

$$10 \;\; 11 \;\; 12 \;\; 12 \;\; 15 \;\; 15 \;\; 15 \;\; 17 \;\; 17 \;\; 18 \;\; 18 \;\; 19.$$

Alors

$$E = \max(x_i) - \min(x_i) = 19 - 10 = 9.$$

2.3.2.2 Variance et écart-type

Définition 2.3.6

La variance, notée $V(x)$, d' une variable statistique X est donnée par la moyenne arithmétique des carrés des écarts des observations.

— Dans le cas de n observations, la variance est donnée par

$$V(x) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2.$$

— Dans le cas de n observations, ordonnées dans un tableau statistique (x_i, n_i), présentant k modalités :

$$V(x) = \frac{1}{n}\sum_{i=1}^{k} n_i(x_i - \overline{x})^2.$$

L'écart-type, noté σ_x, est la racine carrée de la variance et s'exprime ainsi dans la même unité que la variable étudiée :

$$\sigma_x = \sqrt{V(x)}.$$

Remarque 2.3.6 *Plus l'écart-type est faible, plus les valeurs sont regroupées autour de la moyenne. Ainsi, si l'écart-type d'une série de notes des étudiants est faible, cela signifie que la promotion est relativement homogène. Inversement, un écart-type élevé témoigne d'une forte dispersion au sein de la classe.*

Afin de faciliter les différentes étapes de calcul de la variance, il est possible d'utiliser la formule développée de la variance.

Définition 2.3.7

Formules développées de la variance :

— Cas de n observations :

$$V(x) = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right) - \overline{x}^2$$

— Cas d'un tableau statistique avec k modalités :

$$V(x) = \left(\frac{1}{n}\sum_{i=1}^{k} n_i x_i^2\right) - \overline{x}^2$$

Exemple 2.3.8 *Le tableau suivant indique la structure par âges d'une population :*

Âge (par ans)	$[4, 6[$	$[6, 8[$	$[8, 10[$	$[10, 12[$
Effectifs (n_i)	*10*	*18*	*5*	*7*

Ainsi

$$V(x) = \frac{1}{40}\left(10\times 5^2 + 18\times 7^2 + 5\times 9^2 + 7\times 11^2\right) - \left(\frac{1}{40}\left(10\times 5 + 18\times 7 + 5\times 9 + 7\times 11\right)\right)^2$$

2.3.2.3 Le coefficient de variation

Définition 2.3.8

Le coefficient de variation est le rapport noté $CV(x)$ et défini par : $CV(x) = \frac{\sigma_x}{\overline{x}}$, ce coefficient s'exprime en pourcentage de la moyenne.

Remarque 2.3.7 *le coefficient de variation, est utilisé pour comparer la dispersion de séries dont les unités sont différentes.*

Exemple 2.3.9 *Considérons qu'à la suite d'une étude statistique portant sur le poids x des voyageurs et sur celui y des bagages, une compagnie aérienne ait obtenu les résultats suivants :*

Paramètres	x	y
Moyenne	*70 kg*	*15 kg*
Écart-type	*8 kg*	*6 kg*

Pour la série des voyageurs $CV(x) = \frac{8}{70} = 0,1143$*, soit 11,43%, et pour la série des bagages :* $CV(x) = \frac{6}{15} = 0,40$*, soit 40%.*
Alors que l'écart-type de la série des voyageurs est plus grand que celui des bagages ($\sigma_x > \sigma_y$), la série des poids des bagages est plus dispersée que celle des poids des voyageurs, car $CV(y) > CV(x)$.

2.3.3 Les caractéristiques de forme

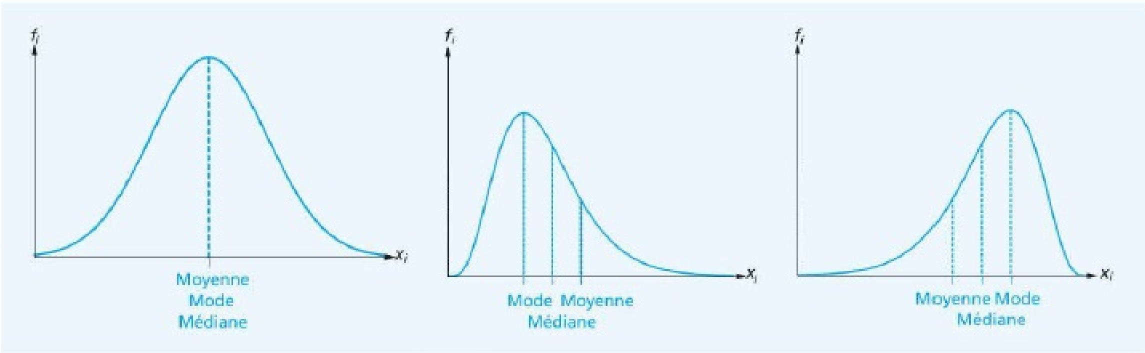
L'asymétrie est un paramètre de forme qui permet d'apprécier la répartition, régulière ou non, des observations autour d'une caractéristique de tendance centrale. Ainsi

— Une distribution est dite symétrique si les trois caractéristiques de tendance centrale sont égales c-à-d :$M_o = M_e = \bar{x}$

— Si $M_o < M_e < \bar{x}$, la distribution est dite asymétrique et étalée vers la droite

— Si $\overline{x} < M_e < M_o$ la distribution est dite asymétrique étalée vers la gauche.

Remarque 2.3.8 *Souvent, l'analyse du diagramme en bâtons ou de l'histogramme permet de se rendre compte du caractère symétrique ou non d'une distribution.*



Distribution
symétrique

Distribution
étalée vers la droite

Distribution
étalée vers la gauche

Il est possible de quantifier l'asymétrie d'une distribution en calculant des coefficients. On distingue trois principaux coefficients d'asymétrie (ou coefficients de skewness (en anglais)) : le coefficient de Yule, le coefficient de Pearson et le coefficient de Fisher.

Par exemple le coefficient de Fisher est donné par :

$$\gamma_1 = \frac{\mu_3}{\sigma_x^3} \text{ où } \mu_3 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^3.$$

La distribution est symétrique si $\gamma_1 = 0$, étalée vers la droite si $\gamma_1 > 0$ et étalée vers la gauche si $\gamma_1 < 0$.

3 Séries statistiques à deux dimensions

Dans la mesure où deux variables seront considérées simultanément, par exemple le salaire et l'âge, il sera possible de s'interroger sur l'existence de liens ou de relations entre les données. Cela nous conduira à l'analyse de régression et l'étude de la corrélation.

3.1 Liens entre deux variables : régression et corrélation

3.1.1 L'analyse de régression et le principe des moindres carrés ordinaires

3.1.1.1 Analyse de régression

Définition 3.1.1

L'analyse de régression renvoie à un ensemble de méthodes visant à analyser la relation existant entre deux ou plusieurs variables. Plus spécifiquement, si l'on considère le cas de deux variables x et y, l'objectif est d'expliquer la variable y (appelée variable expliquée) par la variable x (appelée variable explicative).

Remarque 3.1.1 *Il existe plusieurs types de modèles de régression, le plus connu étant le modèle de régression linéaire :* $y = ax + b$ *où* a *et* b *sont des constantes désignant respectivement le coefficient de pente de la droite de régression de* y *sur* x *et l'ordonnée à l'origine.*

3.1.1.2 Principe général

Exemple 3.1.1 *Considérons, un échantillon de 26 pays* $i(i = 1, \ldots, N$ *avec* $N = 26)$ *pour lesquels on observe les deux variables suivantes pour l'année 2010 : le taux d'investissement en pourcentage du PIB, noté* x_i*, et le taux de croissance économique (exprimé en pourcentage), noté* y_i*, avec* $i = 1, \ldots, 26$*. Les données sont reportées dans le tableau 3.1.*

x_i	y_i	x_i	y_i
15,39	2,45	23,18	5,15
17,91	2,72	23,73	5,32
18,07	2,87	24,07	4,68
18,15	2,95	24,37	5,95
18,58	2,11	25,34	4,64
18,97	2,50	28,08	6,91
19,20	2,03	29,20	5,22
20,92	5,04	30,24	7,82
21,22	4,96	31,57	7,70
21,64	5,84	31,76	7,74
21,72	5,44	32,22	7,72
21,93	5,62	33,57	9,53
22,11	6,03	43,15	9,67

TABLE 3.1 – Taux d'investissement en pourcentage du PIB et taux de croissance économique de 26 régions d'un pays

Afin de de quantifier la relation entre les deux variables, le taux d'investissement et le taux de croissance économique, il convient de représenter dans le plan (x, y) l'allure générale de la distribution à deux caractères.

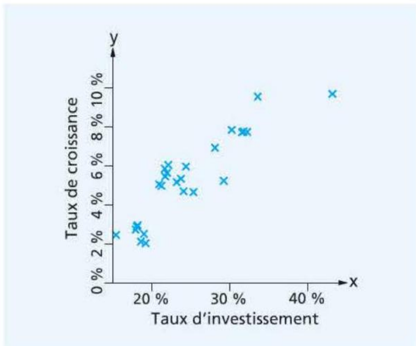


FIGURE 3.1 – Nuage de points, taux d'investissement et taux de croissance économique
?⟨nuage1⟩?

À cette fin, la méthode la plus utilisée consiste à ajuster le nuage de points par une droite (D), comme cela est reproduit sur la figure 3.2. On parle de droite de régression ou de droite d'ajustement ou encore de droite des moindres carrés. La droite (D) ne passant pas par tous les points du nuage, il existe naturellement des écarts entre les points du nuage et les points situés sur cette droite.

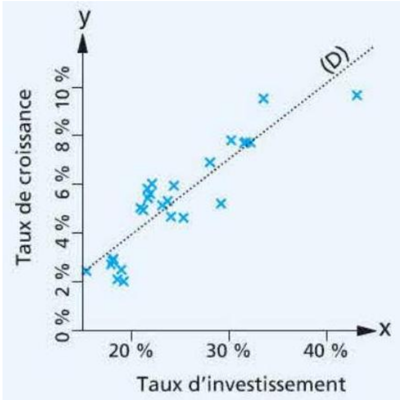


FIGURE 3.2 – Taux d'investissement et taux de croissance économique, ajustement du nuage de points par une droite

⟨nuage2⟩

Plus généralement, comme reproduit sur la figure 3.3

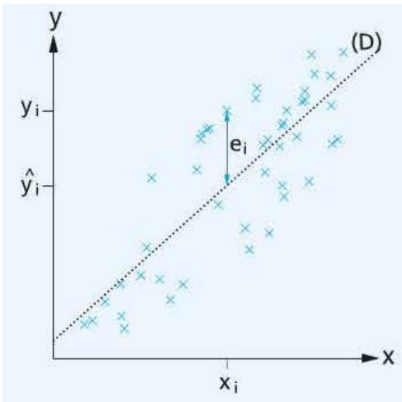


FIGURE 3.3 – Ajustement du nuage de points par une droite de régression

⟨nuage3⟩

notons $\hat{y}_i$ l'ordonnée d'un point de la droite (D) dont l'abscisse est x_i et désignons par e_i les écarts entre la valeur observée y_i de y et la valeur $\hat{y}_i$ située sur la droite :

$$e_i = y_i - \hat{y}_i$$

L'expression de la droite de régression est alors donnée par :

$$\hat{y}_i = ax_i + b$$

où a et b sont des constantes, et le problème consiste à identifier la droite (D) qui ajuste au mieux le nuage de points considéré. En d'autres termes, il s'agit de trouver la droite (D) telle que les écarts e_i soient les plus faibles possibles, c'est-à-dire telle que les valeurs $\hat{y}_i$ situées sur la droite soient les plus proches possibles des valeurs observées y_i.

La méthode des moindres carrés ordinaires (MCO) nous permet précisément de répondre à cet objectif puisqu'elle consiste à trouver la droite (D), c'est-à-dire les valeurs des paramètres a et b, telles que la somme des carrés des écarts e_i soit minimale. Notons que l'on cherche à minimiser la somme des carrés des écarts et non pas directement les écarts puisque ceux-ci pouvant prendre des valeurs positives et négatives (et nulles), ils peuvent se compenser de sorte que leur somme - et donc leur moyenne - reste proche de zéro.

3.1.1.3 Détermination de l'équation de la droite de régression

Afin de déterminer la valeur des paramètres (ou coefficients) a et b de la droite de régression, on applique le principe des MCO consistant à minimiser la somme des carrés des écarts. On cherche ainsi à minimiser l'expression suivante par rapport aux paramètres a et b :

$$\sum_{i=1}^{N} e_i^2 = \sum_{i=1}^{N} (y_i - ax_i - b)^2$$

Propriété 3.1.1

Les paramètres a et b de la droite de régression sont donnés par :

$$a = \frac{\mathrm{Cov}(x,y)}{V(x)} \quad \text{et} \quad b = \bar{y} - a\bar{x}$$

$V(x)$ désignant la variance de x et $\mathrm{Cov}(x,y)$ la covariance entre x et y :

$$V(x) = \frac{1}{N}\sum_{i=1}^{N} (x_i - \bar{x})^2 = \frac{1}{N}\sum_{i=1}^{N} x_i^2 - \bar{x}^2$$

$$\mathrm{Cov}(x,y) = \frac{1}{N}\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{N}\sum_{i=1}^{N} x_i y_i - \bar{x}\bar{y}$$

Le paramètre a est la pente de la droite de régression de y sur x, b désignant l'ordonnée à l'origine.

Exemple 3.1.2 *Si l'on reprend l'exemple de la relation entre le taux d'investissement et le taux de croissance économique (tableau 3.1), on peut calculer les divers indicateurs nécessaires à la détermination de a et b :*

- *Moyenne du taux de croissance :* $\bar{y} = 5,33$
- *Moyenne du taux d'investissement :* $\bar{x} = 24,47$
- *Covariance entre le taux d'investissement et le taux de croissance :*

$$\mathrm{Cov}(x,y) = 12,14$$

- *Variance du taux d'investissement :* $V(x) = 39,11$

On obtient alors aisément les valeurs des paramètres de la droite de régression :

$$a = \frac{12,14}{39,11} = 0,31 \quad \textit{et} \quad b = 5,33 - 0,31 \times 24,47 = -2,26$$

D'où l'équation de la droite de régression reportée sur la figure 2.3 :

$$\hat{y} = 0,31x - 2,26$$

On constate ainsi que la valeur de a est positive, illustrant bien l'existence d'une relation positive entre les deux variables : lorsque le taux d'investissement augmente, la croissance augmente également toutes choses égales par ailleurs.

3.1.2 Mesure du degré de liaison entre deux variables et qualité d'une régression

3.1.2.1 Coefficient de corrélation linéaire

Le coefficient de corrélation linéaire est un nombre sans dimension permettant de quantifier le degré de liaison (linéaire) entre deux variables. Il s'agit ainsi d'un indicateur du degré de proximité entre les points du nuage et ceux figurant sur la droite de régression.

Définition 3.1.2

Le coefficient de corrélation linéaire $r(x, y)$ entre les variables x et y est donné par :

$$r(x, y) = \frac{\text{Cov}(x, y)}{\sigma_x \sigma_y}$$

où σ_x et σ_y désignent respectivement les écarts-types de x et y.

Le coefficient de corrélation linéaire est compris entre -1 et 1 :

$$-1 \leq r(x, y) \leq 1$$

Le coefficient de corrélation linéaire est :

• positif si les variables x et y évoluent dans le même sens ; la corrélation étant d'autant plus forte que $r(x, y)$ est proche de 1 . Les couples de points (x, y) sont alors concentrés autour d'une droite croissante, ainsi que cela est reproduit sur les figures 3.4 et 3.5.

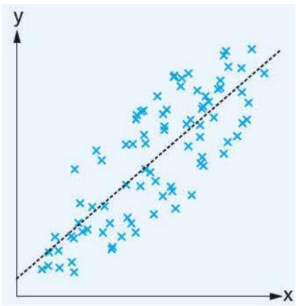


Figure 3.4 – Coefficient de corrélation linéaire positif

⟨nuage4⟩

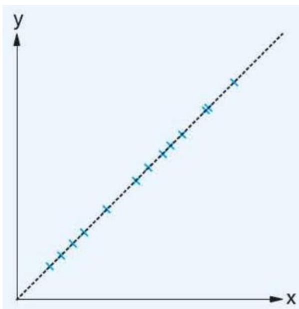


Figure 3.5 – Coefficient de corrélation linéaire égal à 1

⟨nuage5⟩

• négatif si les variables x et y évoluent en sens contraire ; la corrélation (négative) étant d'autant plus forte que $r(x, y)$ est proche de -1 . Les couples de points (x, y) sont alors concentrés autour d'une droite décroissante, comme représenté sur les figures 3.6 et 3.7.

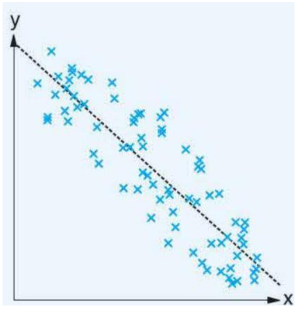


FIGURE 3.6 – Coefficient de corrélation linéaire négatif

⟨nuage6⟩

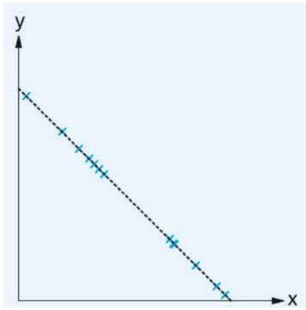


FIGURE 3.7 – Coefficient de corrélation linéaire égal à -1

⟨nuage7⟩

- nul si les variables x et y ne sont pas liées linéairement entre elles ($\vee$ figure 2.9). Le nuage de points est alors dispersé dans le plan (x, y) et il n'est pas possible de l'ajuster par une droite autre que la droite horizontale $\hat{y} = b$ (parallèle à l'axe des abscisses).

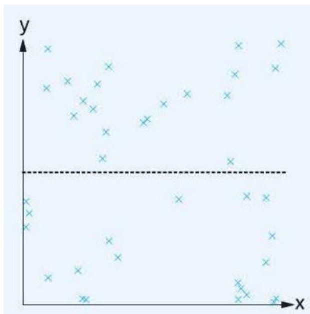


FIGURE 3.8 – Coefficient de corrélation linéaire nul

?⟨nuage8⟩?

Exemple 3.1.3 *Le calcul du coefficient de corrélation linéaire entre le taux d'investissement et le taux de croissance pour notre échantillon de 26 pays ($\triangleright$ tableau 2.5) donne :*

$$r(x, y) = \frac{12,14}{6,27 \times 2,16} = 0,90$$

$r(x, y)$ est ainsi relativement proche de 1 , témoignant d'une corrélation positive importante entre le taux d'investissement et le taux de croissance économique dans notre échantillon.

3.1.2.2 Qualité d'une régression

Même si $r(x, y)$ nous permet d'apprécier le degré de corrélation linéaire entre x et y, il est possible de recourir à un indicateur plus général permettant de mesurer l'intensité de la liaison entre deux variables. Cet indicateur, appelé coefficient de détermination, reste valable en cas de corrélation non linéaire et permet en outre d'apprécier la qualité d'une régression. Son expression est issue de l'équation d'analyse de la variance.

Équation d'analyse de la variance. Ainsi que nous l'avons vu, la variance marginale est égale à la somme de la variance des moyennes conditionnelles et de la moyenne des variances conditionnelles. Considérons nos deux variables x et y pour un échantillon de i individus, $i = 1, \ldots, N$. La variance marginale de y correspond à la **variance totale** (ou globale) de la série à une dimension, c'est-à-dire de la variable y. La variance des moyennes conditionnelles de y indique, par définition, la dispersion des moyennes conditionnelles entre elles. Il s'agit ainsi des moyennes des valeurs y_i pour chaque valeur de x. En d'autres termes, la variance des moyennes conditionnelles de y mesure la dispersion sur la droite de régression : c'est la **variance expliquée** par la régression. La moyenne des variances conditionnelles de y représente la dispersion moyenne des distributions conditionnelles de y : il s'agit de la dispersion moyenne des points autour de la droite de régression. En d'autres termes, c'est la variance qui n'est pas expliquée par la régression, on l'appelle **variance résiduelle**. On en déduit l'équation suivante :

$$\text{Variance totale} \;=\; \text{Variance expliquée} \;+\; \text{Variance résiduelle}$$

soit encore :

$$V(y) = V(\hat{y}) + V(e) \tag{3.1}$$ eq-variance

où $\hat{y}$ est la valeur située sur la droite de régression, $\hat{y} = ax + b$ (où a et b sont supposés estimés par la méthode des MCO), et e désigne l'écart entre la valeur observée de y et $\hat{y}, e = y - \hat{y}$. L'écart e porte également le nom de **résidu** et $\hat{y}$ de **valeur estimée** (ou ajustée) de y. L'équation (3.1) est connue sous le nom d'**équation d'analyse de la variance**.

La qualité d'une régression est ainsi d'autant meilleure que la variance expliquée par cette régression est élevée et, donc, que la variance résiduelle est faible. En d'autres termes, plus la variance expliquée est proche de la variance totale, meilleure est la régression. Il est possible de quantifier cela par le calcul d'un rapport, c'est-à-dire d'un nombre sans dimension, appelé coefficient de détermination.

Définition 3.1.3

Le **coefficient de détermination**, noté R^2, est défini comme le rapport entre la variance expliquée et la variance totale (ou marginale) : soit encore :

$$R^2 = \frac{\text{Variance expliquée}}{\text{Variance totale}} = 1 - \frac{\text{Variance résiduelle}}{\text{Variance totale}}$$

$$R^2 = \frac{V(\hat{y})}{V(y)} = 1 - \frac{V(e)}{V(y)}$$

Le coefficient de détermination mesure ainsi la part ou le pourcentage de variance expliquée par la régression : il fournit une mesure du pouvoir explicatif de la régression.

On a par construction :

$$0 \leq R^2 \leq 1$$

Trois principaux cas peuvent alors se présenter :

— Un coefficient de détermination égal à 0 signifie que la variance expliquée est nulle : la régression n'explique pas le nuage de points, les variables x et y ne sont pas liées entre elles.

— Un coefficient de détermination égal à 1 correspond au cas où la variance expliquée est égale à la variance totale : la régression explique parfaitement le lien entre x et y et la droite d'ajustement passe par tous les points du nuage.

— Dans le cas général, le coefficient de détermination prend une valeur située entre ces deux extrêmes. Plus la valeur de R^2 est proche de 1 , plus la variance expliquée est proche de la variance totale et meilleure est la qualité de la régression.

Remarque 3.1.2 *Dans le cas d'une régression linéaire entre deux variables x et y, il est également possible d'exprimer le coefficient de détermination par la relation suivante, aisément utilisable en pratique :*

$$R^2 = \frac{\mathrm{Cov}(x, y)^2}{V(x)V(y)}$$

et correspondant au carré du coefficient de corrélation linéaire entre x et y : $R^2 = r^2(x, y)$.

Deuxième partie

Eléments de Probabilités

1 Eléments de Probabilités

1.1 Notions élémentaires de base

1.1.1 Expérience aléatoire et univers des possibles

Définition 1.1.1

Une expérience aléatoire est une expérience renouvelable, en théorie ou en pratique, et qui, renouvelée dans des conditions identiques ne donne pas forcément le même résultat à chaque renouvellement.

Exemple 1.1.1

1. *Lancer de pièce, d'un dé, d'une fléchette en direction d'une cible.*
2. *Choix d'une carte dans un jeu de cartes.*
3. *Prélever des boules dans une urne.*

Remarque 1.1.1 *D'autres expériences aléatoires ne peuvent pas être renouvelées en pratique dans les mêmes conditions. Par exemple si on considère la réussite d'un étudiant au baccalauréat. En effet c'est une expérience aléatoire qui peut aboutir à l'un des deux résultats "admis" ou "non admis". Bien évidemment, dans la pratique, cette expérience aléatoire ne peut pas être reproduite plusieurs fois dans les mêmes conditions, c'est-à- dire la même année, avec le même sujet et le même niveau de préparation et de maturité de l'étudiant, etc.*

Définition 1.1.2

L'univers des possibles (ou univers), noté Ω, est défini par l'ensemble de tous les résultats possibles qui peuvent être obtenus au cours d'une expérience aléatoire.

On distingue les univers comprenant un nombre fini de résultats de ceux comprenant un nombre infini de résultats. Parmi les univers infinis, on distingue les univers infinis non dénombrables des univers infinis dénombrables.

Exemple 1.1.2

1. *L'univers* $\Omega = \{w_1, \cdots, w_n, \cdots\} = \{w_i, i \in \mathbb{N}\}$ *est un univers infini dénombrable puisque l'on peut identifier chacun des éléments de* Ω*, même s' il en existe une infinité.*

2. $\Omega = \mathbb{R}$ *ou* $\Omega =]-\infty, a]$ *sont des exemples d' univers infinis non dénombrables.*

Définition 1.1.3

Dans le cas d'un univers fini ou infini dénombrable Ω, la taille de l'univers est appelée cardinal et est représentée par l'opérateur $\text{card}(\Omega)$ ou $|\Omega|$.

Exemple 1.1.3 *On considère une expérience aléatoire correspondant au lancer d'un dé à 6 faces. L'univers (fini) des possibles est alors défini par :*

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

Le cardinal de cet univers est égale à 6*, i.e.* $card(\Omega) = 6$.

Propriété 1.1.1

Soient E un ensemble fini et A et B deux sous-ensemble de Ω. Alors on a

— $\text{Card}(\overline{A}) = \text{Card}(\Omega) - \text{Card}(A) \quad (\text{où } \overline{A} = \{x \in \Omega : x \notin A\})$.

— $\text{Card}(A \cup B) = \text{Card}(A) + \text{Card}(B) - \text{Card}(A \cap B)$.

— $\text{Card}(A \setminus B) = \text{Card}(A) - \text{Card}(A \cap B)$.

1.1.2 événements

Définition 1.1.4

Un événement (ou une partie) A est un sous-ensemble de l'univers des possibles Ω, vérifiant1 $A \subset \Omega$. Un événement constitué d'un seul élément, i.e. pour lequel $\text{card}(A) = 1$, est un événement élémentaire (ou singleton).

Exemple 1.1.4 *On considère une expérience aléatoire correspondant au lancer d'un dé à 6 faces, telle que* $\Omega = \{1, 2, 3, 4, 5, 6\}$.

1. *L' événement "nombre pair", noté A, correspond au sous-ensemble de l' univers* Ω *défini par* $A = \{2, 4, 6\}$.
2. *L'événement "nombre impair", noté B, correspond au sous-ensemble* $B = \{1, 3, 5\}$.
3. *On peut en outre définir un autre événement* C*, sans lui attribuer un nom particulier, tel que* $C = \{1, 5\}$ *par exemple.*
4. *L'événement* $D = \{1\}$ *est un événement élémentaire ou singleton.*

Définition 1.1.5

• Un événement certain correspond à l'univers des possibles Ω.
• Un événement impossible est un événement qui ne se réalise jamais. Il correspond à l'ensemble vide, noté $\emptyset$.

Terminologie probabiliste	**Terminologie ensembliste**	**Notation**
événement certain	Espace entier	Ω
événement impossible	Partie vide	$\emptyset$
événement contraire	Complémentaire	$\overline{A}$ ou A^c
A et B	Intersection	$A \cap B$
A ou B (ou non exclusif)	Réunion	$A \cup B$
événements incompatibles	Parties disjointes	$A \cap B = \emptyset$
Système complet d'événements	Partition de Ω	$A_i \cap A_j = \emptyset_{\ (i \neq j)}$, $\cup A_i = \Omega$
Implication $A \Rightarrow B$	Inclusion	$A \subset B$

TABLE 1.1 – Terminologies probabiliste et ensembliste.

L'implication $A \subset B$ ou $A \Rightarrow B$: l'événement A ne peut être réalisé sans que B le soit.

Définition 1.1.6

L'ensemble des parties, noté $\mathcal{P}(\Omega)$, correspond à l'ensemble de tous les événements réalisables à partir des événements élémentaires de l'univers Ω. Par convention $\Omega \in \mathcal{P}(\Omega)$ et $\emptyset \in \mathcal{P}(\Omega)$.

Remarque 1.1.2 *Pour un univers des possibles Ω de dimension finie, de cardinal card$(\Omega) = n$, le cardinal de l'ensemble des parties $\mathcal{P}(\Omega)$ est égale à :*

$$card\ (\mathcal{P}(\Omega)) = 2^n.$$

Exemple 1.1.5 *Considérons l'exemple d' un lancer de dé à trois faces. L' univers des résultats possibles est $\Omega = \{1, 2, 3\}$. Ainsi, de cet univers on peut construire $2^3 = 8$ événements réalisables, on obtient donc*

$$\mathcal{P}(\Omega) = \{\{1\}, \{2\}, \{3\}, \{1 \cup 2\}, \{1 \cup 3\}, \{2 \cup 3\}, \{1 \cup 2 \cup 3\}, \emptyset\}$$

L'événement $\{1 \cup 2 \cup 3\}$, qui s' interprète comme le fait d'obtenir 1, 2 ou 3, peut être noté sous la forme $\{1, 2, 3\}$ et correspond à l'événement certain (univers) Ω.

1.2 Analyse combinatoire : Dénombrements

L'analyse combinatoire est l'étude mathématique de la manière de ranger des objets.

1.2.1 Principe de multiplication

Principe Soit ξ une expérience qui comporte 2 étapes : la $1^{\text{ère}}$ qui a p_1 résultats possibles et chacun de ces résultats donne lieu à p_2 résultats possibles lors de la $2^{\text{ème}}$ étape. Alors l'expérience ξ a $p_1 \times p_2$ résultats possibles.

Remarque 1.2.1 *Le principe multiplicatif peut s'énoncer ainsi : si un événement A peut se produire de p façons possibles et si un événement B peut se produire de q façons possibles, alors la réalisation de A* **suivie** *de B peut se produire de $p \times q$ façons possibles.*

Remarque 1.2.2 *Le principe de multiplication se généralise au cas où l'expérience ξ comporte n étapes. Le nombre de résultats possibles est alors :*

$$p_1 \times p_2 \times \cdots \times p_n.$$

Exemple 1.2.1

1. *On lance une pièce de monnaie deux fois successivement. Combien y-a-t-il de résultats possibles ? Réponse:* $2 \times 2 = 2^2$.
2. *On veut former des mots de 3 lettres distinctes (ayant un sens ou non !). Combien a-t-on de possibilités ? Réponse: On a*
   - *26 possibilités de choisir la* $1^{\text{ère}}$ *lettres.*
   - *25 possibilités de choisir la* $2^{\text{ème}}$ *lettres.*
   - *24 possibilités de choisir la* $3^{\text{ème}}$ *lettres*

   Ainsi, le nombre de mots possible est : $26 \times 25 \times 24$.
3. *Une urne contient quatre boules 1R, 1B, 1N, 1V. On effectue 2 tirages successifs avec remise. Combien y-a-t-il de résultats possibles ? Réponse:* $4 \times 4 = 16$.

Conséquence 1.2.1 *Si une expérience ξ consiste à répéter n fois de façon indépendante une même expérience qui a p résultats possibles, alors ξ a*

$$p^n = \underbrace{p \times p \times \cdots \times p}_{n \text{ fois}}$$

issues possibles.

1.2.2 Arrangements

1.2.2.1 Arrangements avec répétition

Définition 1.2.1

Soient E un ensemble non vide de cardinal fini n, $n \in \mathbb{N} \setminus \{0\}$, et $p \in \mathbb{N}$. On appelle arrangement avec répétition de p éléments parmi n tout p-uplet d'éléments de E c'est-à-dire tout élément du produit cartésien

$$E^p = \underbrace{E \times E \times \cdots \times E}_{p \text{ fois}}.$$

Exemple 1.2.2 *Soit* $E = \{0, 1, 2, \cdots, 9\}$ *alors*

- $(1, 2, 3)$ *est un arrangement de* 3 *éléments parmi* 10.
- $(0, 9, 8, 9)$ *est un arrangement de* 4 *éléments parmi* 10.
- $(1, 2, 1, 3, 4, 2)$ *est un arrangement de* 6 *éléments parmi* 10.

Remarque 1.2.3

- *Dans un arrangement avec répétition, les p objets de la liste ne sont pas nécessairement tous distincts.*

— *Cela correspond à un tirage avec remise et avec ordre.*

Nombre d'arrangements avec répétition :

Soit un arrangement avec répétition de p éléments parmi n.

— Pour choisir le $1^{\text{ère}}$ élément on a n possibilités.

— Pour choisir le $2^{\text{ème}}$ élément on a n possibilités.

$\vdots$

— Pour choisir le $p^{\text{ème}}$ élément on a n possibilités.

En conséquence du principe de multiplication, le nombre d'arrangement avec répétition de p éléments parmi n est n^p.

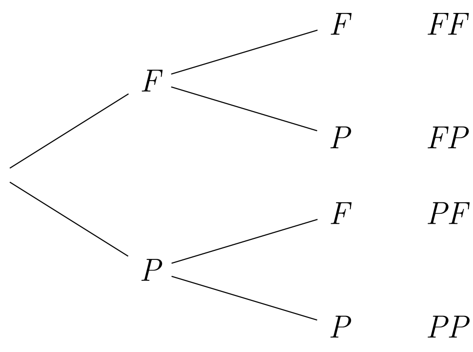
Propriété 1.2.1

Soit E un ensemble fini à n éléments (i.e $\text{Card}(E) = n$). Alors on a :

$$\text{Card}(E^p) = n^p.$$

Exemple 1.2.3

1. *On lance une pièce de monnaie* 2 *fois successives. Combien a-t-on de résultats possibles ? L'univers des éventualités est* $\Omega^2 = \{FF, FP, PF, PP\}$ *avec* $\Omega = \{F, P\}$.
   Réponse: $Card(\Omega^2) = 2^2 = 4$.
   On peut s'en apercevoir en utilisant l'arbre des éventualités :



2. *Un questionnaire à choix multiples, autorisant une seule réponse par question, comprend 15 questions. Pour chaque question, on propose 4 réponses possibles. De combien de façons peut-on répondre à ce questionnaire ?*
   Réponse: Chaque réponse est un arrangement avec répétition de 15 *éléments parmi* 4*. Ainsi le nombre de réponses possibles est donc :* 4^{15}.
3. *Tirage successif avec remise (échantillon non exhaustif)*
   Une urne contient 6 *boules. On effectue 2 tirages successifs avec remise. Combien y-a-t-il de résultats possibles ?*
   Réponse: 6^2.

1.2.2.2 Arrangements sans répétition

Définition 1.2.2

Soit E un ensemble de cardinal fini n et p un entier naturel tel que $0 \leq p \leq n$.
On appelle arrangement sans répétition de p éléments parmi n tout sous-ensemble **ordonné** de E ayant p éléments deux à deux **distincts**.
Autrement dit, tout p-uplet **ordonné** $(a_1, a_2, \cdots, a_p)$ d'éléments de E deux à deux **distincts**.

Exemple 1.2.4 *Si* $E = \{a, b, c, d\}$ *alors*

- abc, bac, cab, bcd, dca *sont des arrangements de* 3 *éléments parmi* 4.
- $abcd, abdc, bacd, dbca$ *sont des arrangements de* 4 *éléments parmi* 4.
- aab *n'est pas un arrangement (sans répétition) de* 3 *éléments parmi* 4.

Remarque 1.2.4

- *Dans un arrangement sans répétition, les* p *objets de la liste sont tous distincts.*
- *Cela correspond à un tirage sans remise et avec ordre.*

Le nombre d'arrangements sans répétition et le symbole A_n^p :

Soit un arrangement sans répétition de p éléments parmi n.

- Pour choisir le $1^{\text{ère}}$ élément on a n possibilités.
- Pour choisir le $2^{\text{ème}}$ élément on a $n-1$ possibilités.

  $\vdots$
- Pour choisir le $p^{\text{ème}}$ élément on a $n-p$ possibilités.

En conséquence du principe de multiplication, le nombre d'arrangement sans répétition de p éléments parmi n est

$$n \times (n-1) \times \ldots \times (n-p).$$

Théorème 1.2.1

Soient $n \in \mathbb{N} \setminus \{0\}$ et $p \in \mathbb{N}$. Alors le nombre d'arrangements sans répétition de p éléments d'un ensemble de cardinal n est

$$A_n^p = \begin{cases} 0 & \text{si } p > n \\ \dfrac{n!}{(n-p)!} = n\times(n-1)\times(n-2)\cdots\times(n-(p-1)) & \text{si } p \leq n. \end{cases}$$

Avec $n! = n\times(n-1)\times(n-2)\times\cdots\times2\times1$ et $0! = 1$.

Exemple 1.2.5 1. *Nous avons* 3 *enfants. De combien de façons différentes peuvent-ils s'asseoir sur un banc de deux places ?*
Réponse: Chaque façon de s'asseoir est un arrangement de 2 *éléments parmi* 3.
Le nombre de possibilités est donc $A_3^2 = 3 \times 2 = 6$.

2. *On veut former des nombres de trois chiffres différents parmi les chiffres :* $2, 3, 5, 6, 7$. *Combien a-t-on de possibilités ?*
   Réponse: Chaque nombre est un arrangement de 3 *éléments parmi* 5. *Le nombre de possibilités est donc* $A_5^3 = 5\times4\times3 = 60$.
3. *Une urne contient* 6 *boules. On effectue un tirage successif et sans remise de* 3 *boules. Combien a-t-on de possibilités ?*
   Réponse: $A_6^3 = 6\times5\times4 = 120$.

1.2.2.3 Permutations

Définition 1.2.3

Soit E un ensemble de cardinal n. On appelle permutation tout arrangement sans répétition des n éléments de E.

De l'étude des arrangements, on déduit immédiatement que le nombre de permutations d'un ensemble E de cardinal n est $A_n^n = n!$, et que toute permutation s'identifie à une bijection : $\{1, 2, \cdots, n\} \longrightarrow E$.

Exemple 1.2.6

1. *Combien y-a-t-il de nombres à trois chiffres choisi parmi les chiffres* $1, 2, 3$ *?*
   Réponse: Chaque nombre est une permutation de ces trois chiffres. Il y en a donc $3! = 3\times2\times1 = 6$.
2. *De combien de façons peut-on repartir* 7 *personnes sur* 7 *chaises ?*
   Réponse: Désignons par $p_1, p_2, \cdots, p_7$ *les* 7 *personnes et posons* $E = \{p_1, p_2, \cdots, p_7\}$. *Une répartition peut se voir comme un arrangement des* 7 *éléments de* E *c'est-à-dire une permutation de* E*, il y en a* $7! = 5040$.

1.2.3 Combinaisons

1.2.3.1 Définition et exemples

Définition 1.2.4

Soit E un ensemble de cardinal n et p un entier naturel tel que $0 \leq p \leq n$. On appelle combinaison de p éléments de E toute partie de E de cardinal p.

Exemple 1.2.7 *Si* $E = \{a, b, c, d\}$ *alors les parties :*

1. $\{a, b\}$, $\{a, c\}$, $\{d, c\}$ *et* $\{d, a\}$ *sont des combinaisons de deux éléments de* E.
2. $\{a, b, c\}$, $\{a, c, d\}$, $\{d, c, b\}$ *et* $\{b, d, a\}$ *sont des combinaisons de* 3 *éléments de* E.

Remarque 1.2.5

1. *Il est essentiel de noter que dans une combinaison :*
   - *On ne tient pas compte de l'ordre, les parties* $\{a, b, c\}$, $\{c, a, b\}$ *et* $\{b, c, a\}$ *sont les mêmes.*
   - *Les éléments sont deux à deux distincts.*
2. *Une combinaison correspond à un tirage sans remise et sans ordre.*

Notation On note $C_n^p := \binom{n}{p}$ le nombre de combinaisons de p éléments parmi n.

Théorème 1.2.2

Soit E un ensemble à n éléments. Le nombre de combinaisons de p éléments de E est égal à :

$$C_n^p = \begin{cases} 0 & \text{si } p > n \\ \dfrac{n!}{p!(n-p)!} = \dfrac{A_n^p}{p!} & \text{si } p \leq n. \end{cases}$$

Exemple 1.2.8

1. *Soit $E = \{a, b, c, d\}$. Le nombre de parties de E contenant 2 éléments est $C_4^2 = 6$. Le nombre de parties de E contenant 3 éléments est $C_4^3 = 4$.*
2. *De combien de façons différentes peut-on choisir 4 étudiants parmi 10 ? Réponse: C_{10}^4 façons différentes.*
3. *Combien de groupes de 4 élèves peut-on former d'une classe de 30 élèves ? Réponse: Chaque groupe possible est une combinaison de 4 éléments parmi 30. Le nombre de groupes possibles est : $C_{30}^4 = 27405$.*

Exemple 1.2.9 *(Modèle). Quand on effectue un tirage simultané (au même temps) de p boules d'une urne qui contient n boules ($p \leq n$) alors toute possibilité est une combinaison de p éléments parmi n.*

Remarque 1.2.6 *(Tirages avec remise et sans ordre) : Quand on effectue un tirage de p boules avec remise et sans ordre d'une urne qui contient n boules, alors le nombre de tirages différents est :*

$$C_{n+p-1}^p = \binom{n+p-1}{p}.$$

Conclusion : Le tableau suivant contient l'ensemble des résultats précédents

	avec ordre	sans ordre
avec répétitions	n^p	$C_{n+p-1}^p = \binom{n+p-1}{p}$
sans répétitions	A_n^p	$C_n^p = \binom{n}{p}$

1.3 Tribu d'événements et espace probabilisable

Définition 1.3.1

Une tribu ou σ-algèbre sur l'univers Ω est un sous-ensemble d'événements ou de parties, notée $\mathcal{F}$, vérifiant :

1. $\mathcal{F} \subset \mathcal{P}(\Omega)$, $\Omega \in \mathcal{F}$ et $\emptyset \in \mathcal{F}$.
2. L'ensemble $\mathcal{F}$ est stable par passage au complémentaire : pour tout événement A de $\mathcal{F}$, l'événement complémentaire $\overline{A}$ appartient à l'ensemble $\mathcal{F}$.
$$\forall A \in \mathcal{F} \text{ alors } \overline{A} \in \mathcal{F}.$$
3. L'ensemble $\mathcal{F}$ est stable par réunion dénombrable : pour toute suite d'événements $(A_n)_{n\in\mathbb{N}}$ appartenant à $\mathcal{F}$, l'union de ces événements appartient à l'ensemble $\mathcal{F}$.
$$(A_n)_{n\in\mathbb{N}} \in \mathcal{F} \text{ alors } \cup_{n\in\mathbb{N}} A_n \in \mathcal{F}.$$

Exemple 1.3.1 *Soit un univers $\Omega = \{1, 2, 3\}$, alors l'ensemble :$\mathcal{A} = \{\emptyset, \{1\}, \{2, 3\}, \Omega\}$ est une tribu sur Ω.*

Propriété 1.3.1: Ensemble des parties

Si l'univers Ω est fini ou infini dénombrable, l'ensemble des parties $\mathcal{P}(\Omega)$ est une tribu sur Ω.

Définition 1.3.2

Un univers probabilisable est un couple $(\Omega, \mathcal{F})$ où $\mathcal{F}$ est une tribu (ou σ-algèbre) sur l'univers Ω.

Exemple 1.3.2 *On considère l'expérience aléatoire qui consiste à lancer un dé à trois faces parfaitement équilibrées. L' univers des résultats possibles est $\Omega = \{1, 2, 3\}$. Ainsi l'ensemble des parties $\mathcal{P}(\Omega)$ est défini par :*

$$\mathcal{P}(\Omega) = \{\{1\}, \{2\}, \{3\}, \{1 \cup 2\}, \{1 \cup 3\}, \{2 \cup 3\}, \{1 \cup 2 \cup 3\}, \emptyset\}.$$

Puisque l'univers Ω est fini, l'ensemble des parties $\mathcal{P}(\Omega)$ est une tribu (ou σ-algèbre) sur Ω le couplet $(\Omega, \mathcal{P}(\Omega))$ est un univers probabilisable.

1.4 Probabilités

1.4.1 Définition générale d'une probabilité

Définition 1.4.1

Soit $(\Omega, \mathcal{F})$ un **univers probabilisable** fini ou infini dénombrable. Une probabilité (ou mesure de probabilité) est une application $Pr : \mathcal{F} \to [0, 1]$, telle que :

1. $P(\Omega) = 1$.
2. Pour toute suite d'événements disjoints $(A_n)_{n\in\mathbb{N}}$ de $\mathcal{F}$ on a (propriété de σ-additivité) :
$$P(\cup_{n\in\mathbb{N}} A_n) = \sum_{n\in\mathbb{N}} P(A_n).$$

Définition 1.4.2

Un **univers probabilisé** es un triplet $(\Omega, \mathcal{F}, Pr)$ où $\mathcal{F}$ est une σ-algèbre sur l' univers Ω et $P(.)$ une mesure de probabilité.

Propriété 1.4.1: Mesure de probabilité

Soit un univers probabilisé $(\Omega, \mathcal{F}, Pr)$, alors quels que soient les événements A et B appartenant $\mathcal{F} \subset \mathcal{P}(\Omega)$, la mesure de probabilité Pr vérifie :

1. $P(\emptyset) = 0$
2. $P(\Omega) = 1$
3. $P(\overline{A}) = 1 - P(A)$.
4. $A \subset B \Rightarrow P(B \backslash A) = P(B) - P(A)$.
5. $A \subset B \Rightarrow P(A) \leq P(B)$ (croissance de Pr).
6. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ (additivité forte).

Remarque 1.4.1 *Si deux événements A et B sont incompatibles (ou disjoints), c'est-à-dire si $A \cap B = 0$, alors :*

$$P(A \cup B) = P(A) + P(B).$$

1.4.2 L'hypothèse d'équiprobabilité, probabilité uniforme

Considérons l'épreuve du jet d'un dé non truqué (i.e toutes les faces ont la même chance de sortir). On définit sur l'univers $\Omega = \{1, 2, 3, 4, 5, 6\}$ la probabilité Pr satisfaisant l'hypothèse d'équiprobabilité :

$$P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\}) = P(\{5\}) = P(\{6\}).$$

Comme

$$1 = P(\Omega) = P(\{1\}) + P(\{2\}) + P(\{3\}) + P(\{4\}) + P(\{5\}) + P(\{6\}).$$

on en déduit que

$$P(\{i\}) = \frac{1}{6}, \quad \forall i \in \{1, 2, 3, 4, 5, 6\}.$$

Soit l'événement A : "avoir un nombre pair". On a $A = \{2, 4, 6\}$ et par suite

$$P(A) = P(\{2\}\cup\{4\}\cup\{6\}) = P(\{2\}) + P(\{4\}) + P(\{6\}) = \frac{3}{6} = \frac{\text{Card}(A)}{\text{Card}(\Omega)}.$$

D'où le théorème, suivant

Théorème 1.4.1

Soit $(\Omega, \mathcal{P}(\Omega))$ un espace probabilisable **fini**. L'hypothèse d'équiprobabilité (i.e. tous les événements élémentaires ont la même chance d'être réalisé) définit une probabilité Pr unique, donnée par

$$\forall A \subset \Omega, \quad P(A) = \frac{\text{Card}(A)}{\text{Card}(\Omega)}.$$

On dit alors que la probabilité Pr est uniforme et que tous les événements sont équiprobables. En fait, en pratique

$$P(A) = \frac{\text{nombre de cas favorables à la réalisation de } A}{\text{nombre total de cas possibles}}.$$

La relation dans le théorème précédent va nous permettre de calculer enfin des probabilités en se servant de toutes les techniques de dénombrement décrites précédemment.

Remarque 1.4.2 *Comme on l'a signalé avant, il ne faut pas croire que sur un espace probabilisable fini on ne définit que la probabilité uniforme. Ce pendant on peut définir des probabilités qui ne sont pas uniformes. Considérons le jet d'une pièce de monnaie truquée telle que*

$$P(F) = \frac{1}{3} \text{ et } P(P) = \frac{2}{3}.$$

Alors Pr est une probabilité non uniforme définit sur l'espace probabilisable fini $(\Omega, \mathcal{P}(\Omega))$ avec $\Omega = \{F, P\}$.

Exemple 1.4.1 *On effectue un tirage successif sans remise de 2 boules, indiscernables au toucher, d'une urne qui contient 3 boules rouges et 5 boules vertes. Calculer les probabilités d'avoir :*

$$\begin{array}{lcl} A & : & \text{"2 boules rouges"} \\ B & : & \text{"2 boules de même couleurs"} \end{array}$$

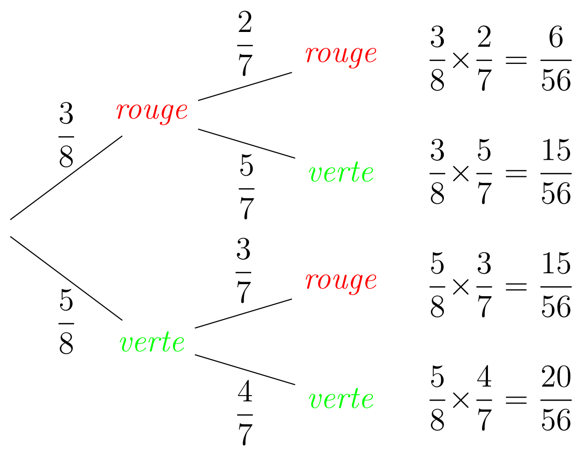
Solution 1.4.1 *Puisque qu'on a un tirage successif sans remise, toute tirage possible est un arrangement de 2 éléments parmi 8. L'univers des éventualités Ω compte $Card(\Omega) = A_8^2 = 8\times 7 = 56$ éléments. On a*

$$Card(A) = A_3^2 = 6 \text{ et } Card(B) = Card(F) + Card(G) = A_3^2 + A_5^2 = 26.$$

avec F : "l'événement d'avoir deux boules rouge" et G : "l'événement d'avoir deux boules vertes".
Comme l'hypothèse d'équiprobabilité est satisfaite, on a

$$P(A) = \frac{Card(A)}{Card(\Omega)} = \frac{6}{56} \text{ et } P(B) = \frac{Card(B)}{Card(\Omega)} = \frac{26}{56}.$$

On peut aussi calculer ces probabilités en utilisant l'arbre des éventualités suivant :



1.5 Indépendance d'événements

La notion de l'indépendance traduit le fait que la réalisation d'un événement n'a pas d'influence sur la réalisation d'un autre.

Définition 1.5.1: Indépendance de deux événements

Deux événements A et B d'un espace probabilisé $(\Omega, \mathcal{B}, P)$ sont dits (stochastiquement) indépendants si

$$P(A \cap B) = P(A) \times P(B).$$

Exemple 1.5.1 *On jette un dé équilibré. l'univers des éventualités est* $\Omega = \{1, 2, 3, 4, 5, 6\}$ *et* $(\Omega, \mathcal{P}(\Omega), P)$ *est l'espace probabilisé,* P *étant la probabilité uniforme. Soient les événements suivant :* $A = \{2, 4, 6\}$, $B = \{5, 6\}$ *et* $C = \{5\}$. *On a immédiatement :*

$$P(A) = \frac{1}{2}, \quad P(B) = \frac{1}{3}, \quad P(C) = \frac{1}{6}, \quad P(A \cap B) = \frac{1}{6}, \quad P(A \cap C) = 0 \quad \text{et} \quad P(B \cap C) = \frac{1}{6}.$$

On en déduit que A *et* B *sont indépendants, mais* B *et* C *ainsi que* A *et* C *ne le sont pas.*

Proposition 1.5.1

Soient A et B deux événements d'un espace probabilisé $(\Omega, \mathcal{B}, P)$. Alors les propositions suivantes sont équivalentes :

(i) A et B sont indépendants

(ii) A et $\overline{B}$ sont indépendants

(iii) $\overline{A}$ et $\overline{B}$ sont indépendants

(iv) $\overline{A}$ et B sont indépendants

Définition 1.5.2: Indépendance de n événements

Soient $A_1, A_2, \cdots, A_n$ n événements d'un espace probabilisé $(\Omega, \mathcal{B}, P)$.

(i) On dit que ces événements sont deux à deux indépendants si

$$\forall i, j \in \{1, 2, \cdots, n\},\ i < j,\ \ P(A_i \cap A_j) = P(A_i)P(A_j)$$

(ii) On dit que ces événements sont mutuellement indépendants si pour toute partie J de $\{1, 2, \cdots, n\}$, on a

$$P(\bigcap_{j \in J} A_j) = \prod_{j \in J} P(A_j)$$

1.6 Probabilité conditionnelle

Définition 1.6.1

Soient $(\Omega, \mathcal{F}, P)$ un espace probabilisé et B un événement de probabilité non nulle. Soit A un événement quelconque. On appelle probabilité de A sachant que B est réalisé, le nombre noté $P(A/B)$ ou encore $P_B(A)$ défini par :

$$P(A/B) = \frac{P(A \cap B)}{P(B)}.$$

Propriété 1.6.1

Soient $(\Omega, \mathcal{F}, P)$ un espace probabilisé. Alors

— $0 \leq P(A/B) \leq 1$.

— $P(A_1 \cup A_2/B) = P(A_1/B) + P(A_2/B)$ si $A_1 \cap A_2 = \varnothing$.

— $P(\overline{A}/B) = 1 - P(A/B)$.

Exemple 1.6.1 *On extrait sans remise 2 cartes d'un jeu de 32 cartes. La première carte tirée est un roi. Quelle est la probabilité que la $2^{\text{ème}}$ carte soit aussi un roi ?*

Solution 1.6.1 *On considère les deux événements*

$$R_1 : \text{" la } 1^{\text{ère}} \text{ carte tirée est un roi"}$$

R_2 : " *la* $2^{\text{ème}}$ *carte tirée est un roi*".

Il s'agit de calculer $P(R_2/R_1)$.

$$P(R_2/R_1) = \frac{P(R_1 \cap R_2)}{P(R_1)} = \frac{\frac{A_4^2}{A_{32}^2}}{\frac{4}{32}} = \frac{3}{31}.$$

Remarque 1.6.1 *On écrit souvent la probabilité conditionnelle de la manière suivante :*

$$P(A \cap B) = P(B) \times P(A/B).$$

Cette formule s'appelle **formule des probabilités composées** *et c'est sous cette forme que le conditionnement sera le plus souvent utilisé.*

Théorème 1.6.1

Soient $A_1, A_2, \cdots, A_n$ n événements d'un espace probabilisé tels que :

$$P(A_1 \cap A_2 \cap \cdots \cap A_{n-1}) > 0.$$

Alors on peut écrire :

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2/A_1)P(A_3/A_1 \cap A_2) \times \cdots \times P(A_n/A_1 \cap A_2 \cap \cdots \cap A_{n-1}).$$

Exemple 1.6.2 *Soit une urne contenant* 4 *boules blanches et* 3 *boules noires. On tire une à une et sans remise* 3 *boules de l'urne. Quelle est la probabilité que la* $1^{\text{ère}}$ *boule tirée soit blanche, la* $2^{\text{ème}}$ *blanche et la* $3^{\text{ème}}$ *noire ?*

Solution 1.6.2

— *Avec le conditionnement : Notons* B_i *l'événement "la* $i^{\text{ème}}$ *boule tirée est blanche" et* N_i *l'événement "la* $i^{\text{ème}}$ *boule tirée est noire". Il s'agit de calculer* $P(B_1 \cap B_2 \cap N_3)$.

$$P(B_1 \cap B_2 \cap N_3) = P(B_1)P(B_2/B_1)P(N_3/B_1 \cap B_2) = \frac{4}{7} \times \frac{3}{6} \times \frac{3}{5} = \frac{6}{35}.$$

— *Par calcul directe :*

$$P(B_1 \cap B_2 \cap N_3) = \frac{A_4^2 A_3^1}{A_7^3} = \frac{6}{35}.$$

1.7 La formule des probabilités totales

Théorème 1.7.1

Soient $(\Omega, \mathcal{B}, P)$ un espace probabilisé et $(A_i)_{i=1,\cdots,n}$ une famille d'événements telle que $\Omega = A_1 \cup A_2 \cdots \cup A_n$, $A_i \cap A_j = \varnothing$ pour tout $i \neq j$ et $P(A_i) > 0$. Si $B \in \mathcal{B}$ alors on a

$$P(B) = \sum_{i=1}^{n} P(A_i)P(B/A_i)$$

Remarque 1.7.1 *Dans le cas particulier* $\Omega = A \cup \overline{A}$, $A \in \mathcal{B}$ *avec* $P(A) > 0$, *la formule des probabilités totales s'écrit*

$$P(B) = P(A)P(B/A) + P(\overline{A})P(B/\overline{A}).$$

C'est sous cette forme qu'elle sera le plus souvent utilisée.

Exemple 1.7.1 *On considère une urne* $\mathcal{U}_1$ *contenant 2 boules blanches et une boule noire et une urne* $\mathcal{U}_2$ *contenant une boule blanche et une boule noire toutes indiscernables au toucher. On choisit une urne au hazard puis on prélève une boule dans cette urne. Quelle est la probabilité de tirer une boule blanche ?*

Solution 1.7.1 *On considère les événements suivants :*

$$\begin{array}{lcl} A_i & : & \text{"avoir l'urne } \mathcal{U}_i\text{"} \quad i \in \{1,2\}. \\ B & : & \text{"avoir une boule blanche".} \end{array}$$

Ainsi, l'univers des éventualités est $\Omega = A_1 \cup A_2$, *cette réunion est disjointe et on a* $B = (B \cap A_1) \cup (B \cap A_2)$. *La formule des probabilités totales nous permet d'écrire*

$$P(B) = P(A_1)P(B/A_1) + P(A_2)P(B/A_2) = \frac{1}{2}\times\frac{2}{3} + \frac{1}{2}\times\frac{1}{2} = \frac{7}{12}.$$

1.8 La formule de Bayes

Soient A et B deux événements de probabilité non nulle. En utilisant le conditionnement, nous pouvons écrire

$$P(A \cap B) = P(B)P(A/B) = P(A)P(B/A).$$

Nous en déduisons

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B/A)}{P(B)}.$$

Alors, on a la propriété suivante :

Proposition 1.8.1: Formule de Bayes

Soit $(A_i)_{i\in\{1,2,\cdots,n\}}$ une partition de Ω (i.e. $\Omega = \bigcup_{i=1}^{n} A_i$ et $\forall i \neq j \ A_i \cap A_j = \varnothing$).
Soit B un événement. Alors on a

$$\forall j \in \{1,2,\cdots,n\}, \quad P(A_j/B) = \frac{P(A_j)P(B/A_j)}{\sum_{i=1}^{n} P(A_i)P(B/A_i)}.$$