Maths 4 : Probabilités et Statistiques

Intitulé du domaine	Sciences et Technologie
Année	$2^{ème}$ année
Annuel ou semestriel	Semestriel
Unité d'enseignement	UEM3 Méthodologie
Chargé du module	Dr Djawad ZENDAGUI
Volume horaire global	45 heures
	1h30 /semaine Cours 1h30 /semaine TD
Nombre de crédits	4
Notation	(2*EMD+CC)/3 EMD= Epreuve de moyenne durée CC= Contrôle continu=N1+N2+N3+N4+N5 •N1, N2 et N3: Notes de 03 interrogations écrites de 10min max sur 04 points chacun •N4: Note de présence sur 04 points •N5: Note d'assiduité sur 04 points

Dr Djawad ZENDAGUI

Maître de conférences
Département de Génie Civil - Faculté des Sciences de l'Ingénieur-
Université Aboubekr Belkaid Tlemcen
BP 230 (13000)
Tel/Fax : 043 28 56 85
Email : d_zendagui@mail.univ-tlemcen.dz

Docteur d'Etat en Génie Civil	Ecole Nationale Polytechnique d'Alger (ENP)	Juillet 2006	Algérie
Certificate of Graduate Studies in Engineering and Construction Management	Université Missouri Rolla (UMR)	Juin 2003	USA
Magister en Génie Civil	Ecole Nationale Polytechnique d'Alger (ENP)	Juillet 1996	Algérie
Ingénieur d'Etat en Travaux Publics	Ecole Nationale des Travaux Publics d'Alger (ENTP)	Juin 1992	Algérie

Objectifs du cours

Le cours a pour but d'initier les étudiants aux principes de base de la probabilité et statistique.

Support pédagogique

Il est mis à la disposition des étudiants un support pédagogique sur papier du cours et des Travaux Dirigés (TD).

Plateforme Elearn de l'université (elearn.univ-tlemcen.dz).

Partie A Introduction générale et organisation du cours
Chapitre1

Partie B Traitement statistique de l'information
Chapitre2 à Chapitre 5

Partie C Traitement probabiliste de l'information
Chapitre6 à Chapitre 9

Chapitre 1

Introduction générale et organisation du cours

L'incertain est-il notre quotidien????

Ce qui est sûr, c'est que rien n'est sûr

Probabilité et Statistiques :

Outils au service de l'engineering

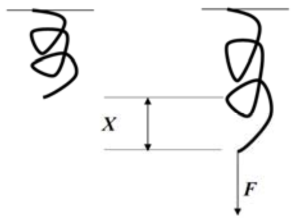
Place du cours dans votre futur métier

- Analyse des données
- Prédictions
- Simulations (processus stochastiques)
- Décisions (probabilités d'occurrence et risque)
- …

La conception d'un système donné nécessite trois étapes :

1. La compréhension du système : quelle est sa fonction
2. La modélisation du système : quel est le modèle à développer pour décrire ce système avec l'identification de l'input et l'output de ce modèle
3. Le recours aux données : l'utilisation de données nécessaires pour l'utilisation du modèle. Ces données sont l'input du modèle.

Considérons un exemple très simple……



Etape 1

Nul besoin de montrer le fonctionnement d'un ressort dans un système mécanique (suspension de voiture,…)

Etape 2

Essayons maintenant de voir comment modéliser ce ressort.

L'input du modèle est:

•La force agissante, dans notre cas *F*
•Le ou les caractéristiques du ressort, dans notre cas *K*
•L'output du modèle est par exemple la déformation du ressort, dans notre cas *X*.

Une relation toute simple a été mise en place **(modèle mathématique).**

$$F = K.X \qquad X = \frac{F}{K}$$

Etape 3

Donc en se basant sur ce modèle mathématique, il est possible de connaître la déformation du ressort connaissant *F* et *K*

En fait, il est possible de dire que la connaissance de l'output (Déformation du ressort) est acquise.

Est-ce vrai ?

NON

Pourquoi ?

1. Le premier problème qui se pose est tout d'abord: Est ce que le modèle mis en place est "exact" ?

2. Le deuxième problème auquel on est confronté est la validité de l'information de l'Input. En d'autres termes peut-on dire avec certitude dire que **les valeurs de *F* et *K* sont exactes et connues**

Un petit exemple

Un autre exemple mais celui là est dramatique

La statistique est un ensemble de méthodes permettant:

- ✓de recueillir des données “brutes”;
- ✓de présenter, résumer ces données;
- ✓de tirer des conclusions sur la population étudiée (sa structure, sa composition), d’aider à la prise de décision; en présence de données dépendant du temps, de faire de la prévision.

Les outils de la statistique et de la probabilité permettent de répondre à la question suivante:

Comment déterminer la valeur de l’Output si l’Input et/ou le modèle mathématique du phénomène étudié ne sont pas connus

Fin du Chapitre 1

Introduction générale et organisation du cours

Partie B
Traitement statistique de l’information

Chapitre 2

Collecte des données

•Population ou unité statistique et/ou échantillon

•Individu

•Caractère

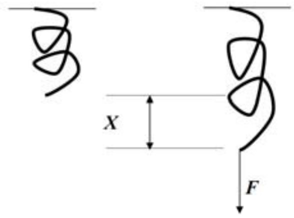
•Modalités

-Population: Employés d'une usine

-Individu: Un employé de cette usine

-Caractère: Salaire

-Modalités: 10000DA, 20000DA, 25000DA

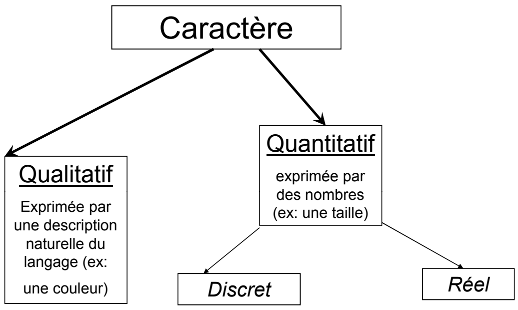


-Population: Ressorts

-Individu: Un ressort parmi ces ressort

-Caractère: Rigidité *K*

-Modalités: $K \in [10,20] \quad N/m$



Exemple:

On souhaite connaitre l'état des maisons

Choix entre les trois types de caractère

-Population: Maisons (100)

-Individu: Une maison parmi ces 100 maisons

-Caractère: L'état de la maison

-Modalités: Petite, moyenne, grande

Caractère qualitatif

-Population: Maisons (100)

-Individu: Une maison parmi ces 100 maisons

-Caractère: Nombre de pièces

-Modalités: 1, 2, 3, 4, 5

Caractère quantitatif discret

-Population: Maisons (100)

-Individu: Une maison parmi ces 100 maisons

-Caractère: Surface (notée S)

-Modalités: $S \in [60, 200]$ m²

Caractère quantitatif continu

Une compagnie achète 10 000 ampoules électriques d'un fabricant qui affirme que ses ampoules fonctionnent durant au moins 1 000 heures **(1 mois et 11 jours, sans arrêt).** Cette compagnie vérifie 15 ampoules et, suite à ces résultats doit décider si elle garde ou non les 10 000 ampoules.

Identifier la population, l'individu, le caractère et les modalités

Une compagnie achète 10 000 ampoules électriques d'un fabricant qui affirme que ses ampoules fonctionnent durant au moins 1 000 heures **(1 mois et 11 jours, sans arrêt).** Cette compagnie vérifie 15 ampoules et, suite à ces résultats doit décider si elle garde ou non les 10 000 ampoules.

Population	:	l'ensemble des *10 000 ampoules achetées*.
Échantillon	:	les *15 ampoules vérifiées*.
Individu	:	une *ampoule* parmi les 15
Caractère	:	*durée* de fonctionnement de l'ampoule
Modalités	:	*Durée* en heures

Variable statistique (VS) continue

Population	:	l'ensemble des *10 000 ampoules achetées*.
Échantillon	:	les *15 ampoules vérifiées*.
Individu	:	une ampoule parmi les 15
Caractère	:	l'*état* de l'ampoule
Modalités	:	Bon ou mauvais

Caractère qualitatif

Un même individu peut il avoir plusieurs caractères????

Oui

Population	:	l'ensemble des *10 000 ampoules achetées*.
Échantillon	:	les *15 ampoules vérifiées*.
Individu	:	une ampoule parmi les 15
Caractère	:	l'*état* de l'ampoule la *durée* de fonctionnement de l'ampoule
Modalités	:	Bon ou mauvais *Durée* en heures

Par rapport à un caractère un individu peut il avoir plusieurs modalités????

Non

- <u>Classification des données</u>
- Visualisation graphique
- Quantification à l'aide de paramètres statistiques

Les notes des étudiants

8,75	6,00	3,75	11,25	11,50	7,00	11,75	10,50	5,00	6,75	7,50
11,75	8,75	16,25	12,00	10,50	15,25	12,50	11,00	9,50	7,75	8,50
12,50	4,00	16,00	7,50	9,00	11,00	13,50	8,00	13,00	10,75	12,75
5,50	9,00	9,25	0,00	8,75	10,75	6,50	7,00	12,00	7,00	9,50
15,25	11,50	10,75	5,25	11,50	9,25	9,75	9,75	14,75	6,00	15,25
10,50	11,00	4,75	13,25	11,50	12,00	12,00	12,00	9,00	4,25	7,00
9,00	9,50	13,25	15,25	8,00	12,25	10,75	7,25	9,50	7,50	10,25
14,75	15,50	10,50	10,25	13,50	9,50	5,00	5,50	9,50	5,50	5,75
6,50	1,25	14,75	16,50	11,75	5,75	4,50	7,50	16,25	6,50	17,75
6,50	2,00	14,00	13,50	11,00	10,75	9,50	0,00	15,50	10,75	4,50
6,50	7,00	8,75	7,75	16,75	11,50	8,50	10,25	12,75	10,00	2,75

Les notes des étudiants

8,75	6,00	3,75	11,25	11,50	7,00	11,75	10,50	5,00	6,75	7,50
11,75	8,75	16,25	12,00	10,50	15,25	12,50	11,00	9,50	7,75	8,50
12,50	4,00	16,00	7,50	9,00	11,00	13,50	8,00	13,00	10,75	12,75
5,50	9,00	9,25	0,00	8,75	10,75	6,50	7,00	12,00	7,00	9,50
15,25	11,50	10,75	5,25	11,50	9,25	9,75	9,75	14,75	6,00	15,25
10,50	11,00	4,75	13,25	11,50	12,00	12,00	12,00	9,00	4,25	7,00
9,00	9,50	13,25	15,25	8,00	12,25	10,75	7,25	9,50	7,50	10,25
14,75	15,50	10,50	10,25	13,50	9,50	5,00	5,50	9,50	5,50	5,75
6,50	1,25	14,75	16,50	11,75	5,75	4,50	7,50	16,25	6,50	17,75
6,50	2,00	14,00	13,50	11,00	10,75	9,50	0,00	15,50	10,75	4,50
6,50	7,00	8,75	7,75	16,75	11,50	8,50	10,25	12,75	10,00	2,75

Population : l'ensemble des étudiants (121)
Individu : un étudiant parmi les 121.
Caractère : la note de l'examen
Modalités : les valeurs de la note [0,20]

Soit $\boldsymbol{P}$ une population formée de n individus (x_i, $i=1,..n$).

Soit $\boldsymbol{C}$ un caractère ayant k modalités c_1, $c_2,\ldots,c_k$. Ce caractère peut être qualitative, quantitative (discret ou continu)

Remarquez que $n \neq k$

Pourquoi????

La collecte de l'information relative au caractère $\boldsymbol{C}$ auprès de la population $\boldsymbol{P}$ consiste à observer pour chaque individu de $\boldsymbol{P}$ la modalité qui lui correspond.

Le résultat obtenu est la ***<u>série statistique</u>***

L'individu n°1 identifié par x_1 présente une modalité parmi les k modalités (avec par exemple k=10). Les autres individus vont avoir les modalités suivantes:

$x_1 = c_3$
$x_2 = c_{10}$
$x_3 = c_3$
.
.
.
$x_n = c_8$

Le traitement de cette information élémentaire consiste à dénombrer pour chaque modalité c_j, le nombre d'individus de la population P n_j qui présentent cette modalité.

La distribution statistique est donc formée par les couples

$$\left(c_j, n_j\right) \qquad j = 1,2,\ldots,k$$

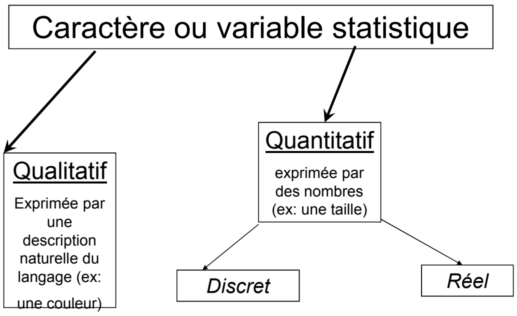
Notion de fréquence

Le nombre n_j est dit effectif ou <u>fréquence absolue</u> de la modalité c_i. Il est clair que :

$$n = \sum_{j=1}^{j=k} n_j$$

Le nombre f_j est dit <u>fréquence</u> de la modalité c_j.

$$f_j = \frac{n_j}{n} \qquad \sum_{j=1}^{j=k} f_j = 1$$



k modalités $c_1, c_2, \ldots, c_k$

Jours du mois de Janvier	**Population**
Climat	**Caractère**
Ensoleillé, pluvieux, orageux	**Modalité**

$n=31$.
$k=3$ modalités

c_1=*Ensoleillé*
c_2=*Pluvieux*
c_3=Orageux

Remarquez que $n \neq k$

Individu	Modalité
1	Ensoleillé
2	Ensoleillé
3	Ensoleillé
4	Pluvieux
5	Pluvieux
6	Pluvieux
7	Orageux
8	Orageux
9	Pluvieux
10	Pluvieux
11	Ensoleillé
12	Ensoleillé
13	Pluvieux
14	Pluvieux
15	Pluvieux
16	Pluvieux
17	Orageux
18	Orageux
19	Pluvieux
20	Pluvieux
21	Pluvieux
22	Ensoleillé
23	Ensoleillé
24	Ensoleillé
25	Pluvieux
26	Pluvieux
27	Pluvieux
28	Orageux
29	Orageux
30	Orageux
31	Ensoleillé

Série statistique

Individu	Modalité
1	Ensoleillé
2	Ensoleillé
3	Ensoleillé
4	Pluvieux
5	Pluvieux
6	Pluvieux
7	Orageux
8	Orageux
9	Pluvieux
10	Pluvieux
11	Ensoleillé
12	Ensoleillé
13	Pluvieux
14	Pluvieux
15	Pluvieux
16	Pluvieux
17	Orageux
18	Orageux
19	Pluvieux
20	Pluvieux
21	Pluvieux
22	Ensoleillé
23	Ensoleillé
24	Ensoleillé
25	Pluvieux
26	Pluvieux
27	Pluvieux
28	Orageux
29	Orageux
30	Orageux
31	Ensoleillé

Modalité	Total	Fréquence de la modalité
Total Ensoleillé	9	0.29
Total Orageux	7	0.23
Total Pluvieux	15	0.48

Tableau statistique

Les 10 séismes recensés durant une année au niveau d'une région **Population**

Intensité **Caractère**

Intensité I à XII **Modalité**

n=10.
k=12 modalités

$c_1=1$
$c_2=2$

$c_{12}=12$

Série statistique

Séisme n°	Intensité (Modalité)
1	2
2	4
3	6
4	8
5	2
6	1
7	6
8	7
9	5
10	5

Intensité (Modalité)	Effectif	Fréquence de la modalité
1	1	0.1
2	2	0.2
3	0	0
4	1	0.1
5	2	0.2
6	2	0.2
7	1	0.1
8	1	0.1
9	0	0
10	0	0
11	0	0
12	0	0

Tableau statistique

Les 10 séismes recensés durant une année au niveau d'une région **Population**

Magnitude de Richter **Caractère**

Valeur de la Magnitude **Modalité**

<u>Les valeurs de la magnitude sont réels</u>

n=10.
k=????

Comment construire le tableau statistique???

Série statistique

Séisme n°	Magnitude (Modalité)
1	2.2
2	3.1
3	5.5
4	7.5
5	2.2
6	1.5
7	4.3
8	6.8
9	4.9
10	4.7

Du moment que le caractère est quantitatif continu alors l'idée consiste à établir des classes et ce pour mettre en place le tableau statistique

Pas=1

Classe	Intervalle	Nombre d'individu	Fréquence
Classe 1	]1.5,2.5[	3	0.3
Classe 2	]2.5,3.5[	1	0.1
Classe 3	]3.5,4.5[	1	0.1
Classe 4	]4.5,5.5[	2	0.2
Classe 5	]5.5,6.5[	1	0.1
Classe 6	]6.5,7.5[	1	0.1
Classe 7	]7.5,8.5[	1	0.1

Séisme n°	Magnitude (Modalité)
1	2.2
2	3.1
3	5.5
4	7.5
5	2.2
6	1.5
7	4.3
8	6.8
9	4.9
10	4.7

Est-ce qu'il y a un seul tableau statistique ?

Pas=1

Classe	Intervalle	Nombre d'individu	Fréquence
Classe 1	]1.5,2.5[	3	0.3
Classe 2	]2.5,3.5[	1	0.1
Classe 3	]3.5,4.5[	1	0.1
Classe 4	]4.5,5.5[	2	0.2
Classe 5	]5.5,6.5[	1	0.1
Classe 6	]6.5,7.5[	1	0.1
Classe 7	]7.5,8.5[	1	0.1

NON

Pas=1.5

Classe	Intervalle	Nombre d'individu	Fréquence
Classe 1	]1.5,3[	3	0.3
Classe 2	]3,4.5[	2	0.2
Classe 3	]4.5,6[	3	0.3
Classe 4	]6,7.5[	1	0.1
Classe 5	]7.5,9[	1	0.1

Pas=3.0

Classe	Intervalle	Nombre d'individu	Fréquence
Classe 1	]1.5,4.5[	5	0.5
Classe 2	]4.5,7.5[	4	0.4
Classe 3	]7.5,9.5[	1	0.1

Donc

- Pour un pas de 1 on a 7 classes
- Pour un pas de 1.5 on a 5 classes
- Pour un pas de 3 on a 3 classes

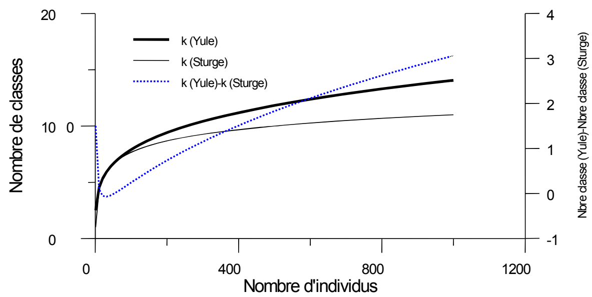
Questions :

- Est-ce que le choix du **pas** a une importance dans le traitement des données ?
- Quel **pas** choisir ? Cette question peut être posée autrement
- Quel est le nombre de classes que l'on doit choisir

Soit k le nombre de classes

STURGE $$k = 1 + 3.3\log_{10}(n)$$

YULE $$k = 2.5\sqrt[4]{n}$$



$$k = 1 + 3.3\log_{10}(10) = 4.3$$

$$k = 5$$

Fin du Chapitre 2

Collecte des données

Chapitre 3

Distribution statistique à un caractère

3.1 Description graphique
3.2 Description numérique
3.2.1 Caractéristique de tendance centrale
3.2.2 Caractéristique de dispersion
3.3 Caractéristiques de formes

•Classification des données

•Visualisation graphique

•Quantification à l'aide de paramètres statistiques

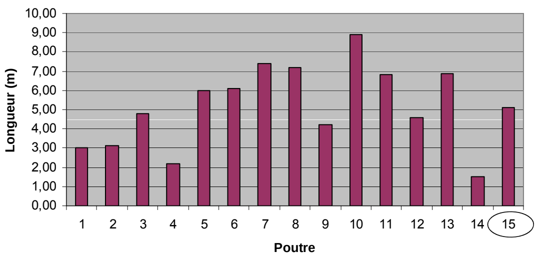
3.1 Description graphique

3.1.1 Variable quantitative continue

La série statistique

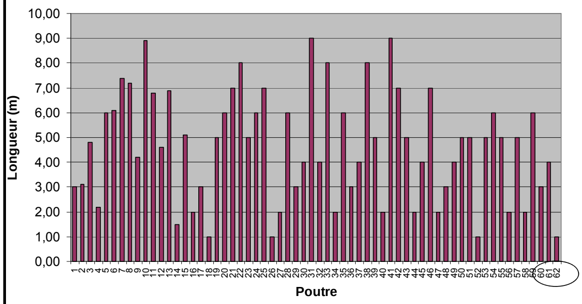
N° Individu	Longueur (m)
1	3.0
2	3.1
3	4.8
4	2.2
5	6.0
6	6.1
7	7.4
8	7.2
9	4.2
10	8.9
11	6.8
12	4.6
13	6.9
14	1.5
15	5.1

Histogramme



Exploitation difficile

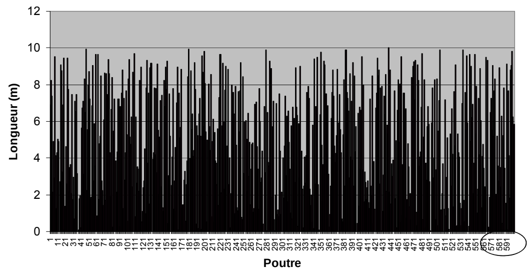
Histogramme



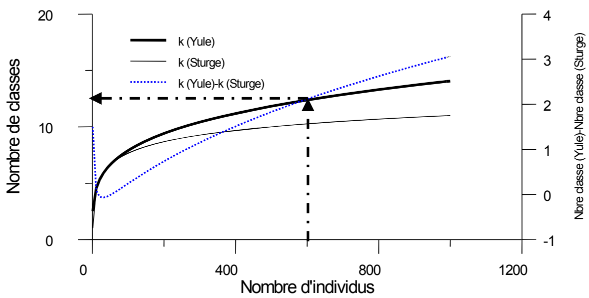
L'Exploitation devient plus difficile si le nombre d'individu augmente

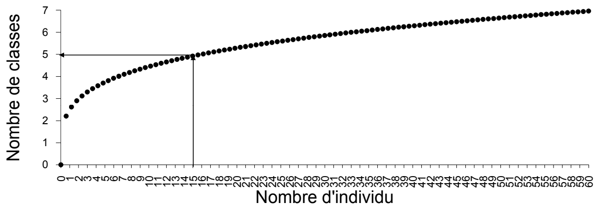
Dans le présent cas de 15 à 62

Histogramme



Avec un nombre d'individu qui avoisine les 600 individus, l'exploitation devient pratiquement **impossible**





Établissement du tableau (ou distribution) statistique

N° Individu	Longueur (m)
1	3.0
2	3.1
3	4.8
4	2.2
5	6.0
6	6.1
7	7.4
8	7.2
9	4.2
10	8.9
11	6.8
12	4.6
13	6.9
14	1.5
15	5.1

$$n = 15$$

$$k = 1 + 3.3\log_{10}(15)$$

$$= 4.88$$

$$k = 5$$

$$V_{\max} = 8.9\,\mathrm{m}$$

$$V_{\min} = 1.50\,\mathrm{m}$$

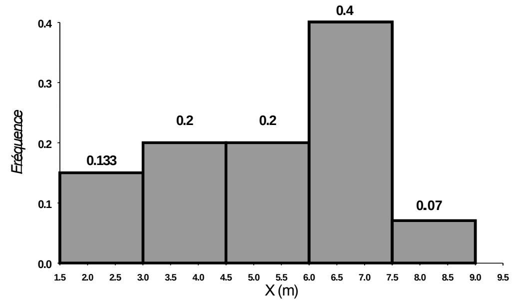
N° Individu	Longueur (m)
1	3.0
2	3.1
3	4.8
4	2.2
5	6.0
6	6.1
7	7.4
8	7.2
9	4.2
10	8.9
11	6.8
12	4.6
13	6.9
14	1.5
15	5.1

$$Pas = \frac{V_{\max} - V_{\min}}{k}$$

$$Pas = \frac{8.9 - 1.5}{5} = 1.48\,\mathrm{m}$$

$$Pas = 1.50\,\mathrm{m}$$

N° Classe	Intervalle	Nombre d'individu	Fréquence
Classe 1	[1.5, 3.0 [	2	0.13
Classe 2	[3.0, 4.5 [	3	0.2
Classe 3	[4.5, 6.0 [	3	0.2
Classe 4	[6.0, 7.5 [	6	0.4
Classe 5	[7.5, 9.0 [	1	0.07



Pas=1

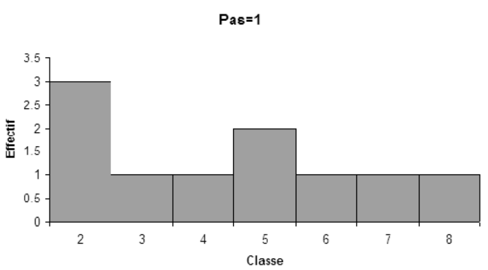
Classe	Intervalle	Nombre d'individu	Fréquence
Classe 1	[1.5,2.5[	3	0.3
Classe 2	[2.5,3.5[	1	0.1
Classe 3	[3.5,4.5[	1	0.1
Classe 4	[4.5,5.5[	2	0.2
Classe 5	[5.5,6.5[	1	0.1
Classe 6	[6.5,7.5[	1	0.1
Classe 7	[7.5,8.5[	1	0.1

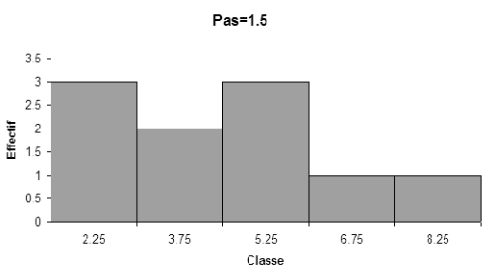
Pas=1.5

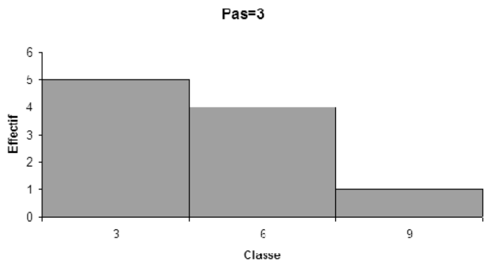
Classe	Intervalle	Nombre d'individu	Fréquence
Classe 1	[1.5,3[	3	0.3
Classe 2	[3,4.5[	2	0.2
Classe 3	[4.5,6[	3	0.3
Classe 4	[6,7.5[	1	0.1
Classe 5	[7.5,9[	1	0.1

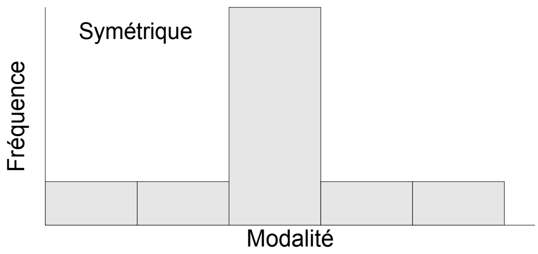
Pas=3.0

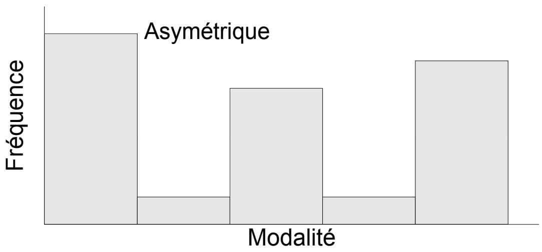
Classe	Intervalle	Nombre d'individu	Fréquence
Classe 1	[1.5,4.5[	5	0.5
Classe 2	[4.5,7.5[	4	0.4
Classe 3	[7.5,9.5[	1	0.1

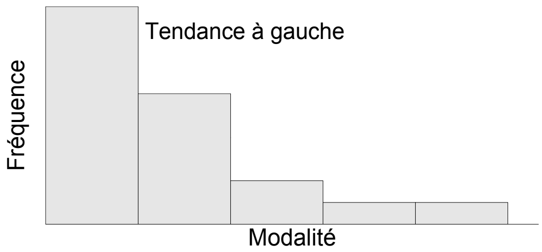


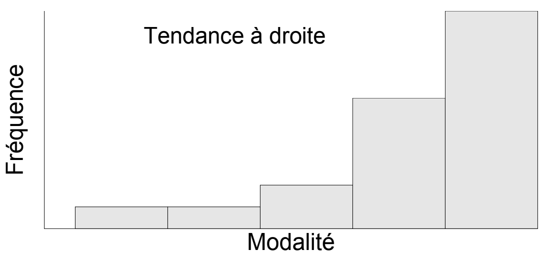


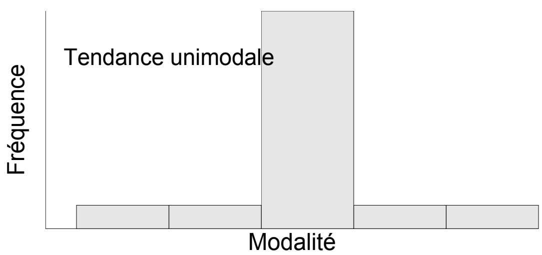

Pas=3
Effectif
0
1
2
3
4
5
6
3
6
9
Classe

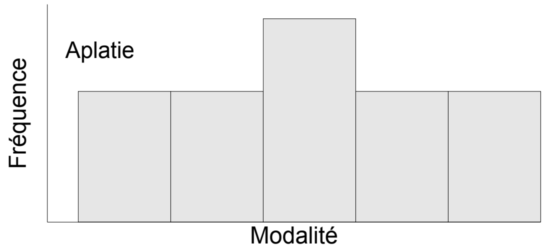

Symétrique
Fréquence
Modalité

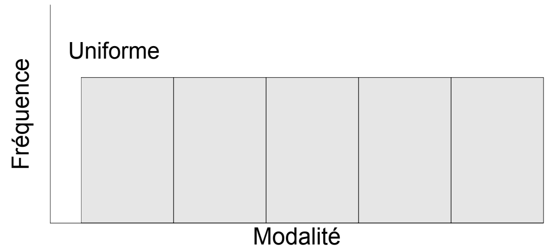

Asymétrique
Fréquence
Modalité


Tendance à gauche
Fréquence
Modalité


Tendance à droite
Fréquence
Modalité


Tendance unimodale
Fréquence
Modalité





Quel est le nombre d'individus ayant une valeur (modalité) inférieure à une certaine valeur (par exemple 6)

Fonction cumulative

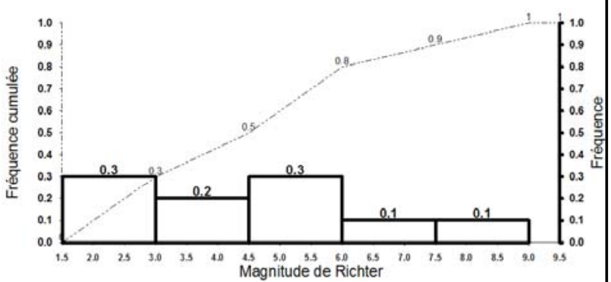
Il suffit pour cela de sommer le nombre de séismes dont la magnitude est inférieure à 6.

Ce nombre divisé par le nombre d'individus de la population s'appelle **Fréquence cumulée**.

Tableau 3.11 Pas=1.5

Classe	Intervalle	Nombre d'individu	Fréquence	Centre de classe	Fréquence cumulée
Classe 1	[1.5,3[	3	0.3	2.25	0.3
Classe 2	[3,4.5[	2	0.2	3.75	0.5
Classe 3	[4.5,6[	3	0.3	5.25	0.8
Classe 4	[6,7.5[	1	0.1	6.75	0.9
Classe 5	[7.5,9[	1	0.1	8.25	1.0

Pour chaque valeur, il existe une **fréquence cumulée**.



Le tracé des couples (valeur,**fréquence cumulée)** donne ce qu'on appelle la ***Fonction cumulative***.

Déterminer le nombre de séismes dont la magnitude est inférieure à 6.0.

Séisme n°	Magnitude (Modalité)
1	2.2
2	3.1
3	5.5
4	7.5
5	2.2
6	1.5
7	4.3
8	6.8
9	4.9
10	4.7

8 sur 10 soit 0.80 ou 80%

Question :
Déterminer le nombre de séismes dont la magnitude est inférieure à 4.5.

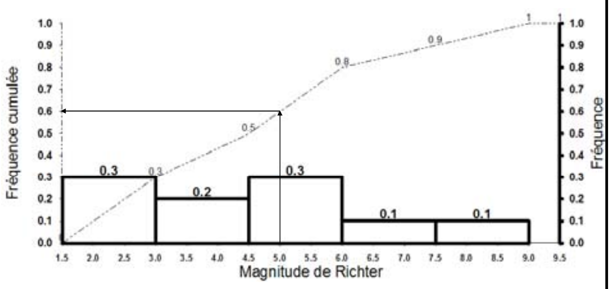
Séisme n°	Magnitude (Modalité)
1	2.2
2	3.1
3	5.5
4	7.5
5	2.2
6	1.5
7	4.3
8	6.8
9	4.9
10	4.7

5 sur 10 soit 0.50 ou 50%

Question :
Déterminer le nombre de séismes dont la magnitude est inférieure à 5.0.

Séisme n°	Magnitude (Modalité)
1	2.2
2	3.1
3	5.5
4	7.5
5	2.2
6	1.5
7	4.3
8	6.8
9	4.9
10	4.7

7 sur 10 soit 0.70 ou 70%



Soit 0.60 ou **60%**

7 sur 10 soit 0.70 ou **70%**

3.1 Description graphique

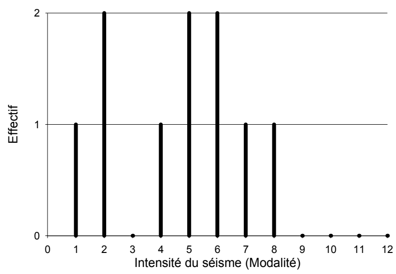
3.1.2 Variable quantitative discrète

La série statistique

Séisme n°	Intensité (Modalité)
1	2
2	4
3	6
4	8
5	2
6	1
7	6
8	7
9	5
10	5

La distribution statistique

Intensité (Modalité)	Effectif	Fréquence de la modalité
1	1	0.1
2	2	0.2
3	0	0
4	1	0.1
5	2	0.2
6	2	0.2
7	1	0.1
8	1	0.1
9	0	0
10	0	0
11	0	0
12	0	0

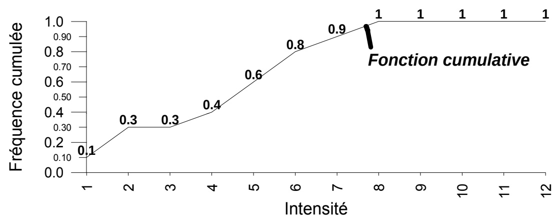


Représentation par bâtonnet

Intensité (Modalité)	Effectif	Fréquence de la modalité	Fréquence cumulée
1	1	0.1	0.1
2	2	0.2	0.3
3	0	0	0.3
4	1	0.1	0.4
5	2	0.2	0.6
6	2	0.2	0.8
7	1	0.1	0.9
8	1	0.1	1
9	0	0	1
10	0	0	1
11	0	0	1
12	0	0	1
	10	1	

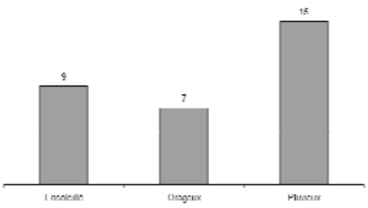
Le nombre de séismes dont l'intensité est inférieure à 5.

Ce nombre divisé par le nombre d'individus de la population s'appelle Fréquence cumulée.

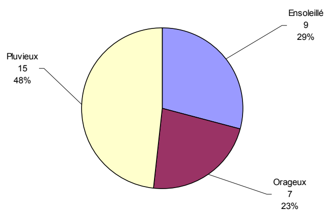


3.1 Description graphique

3.1.3 Variable qualitative



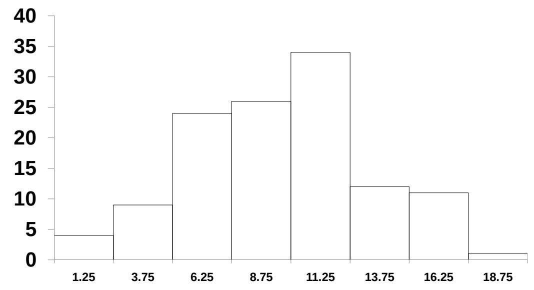
Représentation par diagramme



Représentation Diagrammes circulaires

Chapitre 2: Collecte des données

Les notes des étudiants



97

Chapitre 3: Distribution statistique à un caractère

Est-ce la description graphique est suffisante ?

98

Chapitre 3: Distribution statistique à un caractère

Les modèles mathématiques nécessitent l'introduction d'une valeur et non l'appréciation sur la variation des modalités

99

Chapitre 3: Distribution statistique à un caractère

3.2 Description numérique

3.2.1 Caractéristique de tendance centrale

Exemple: X=F/K

F: Force (mesurée en Newton) et K: Raideur: Connue=10 N/cm

F: Force (mesurée en Newton)

10	10.3	22.1	23	26	15.2	12.3	18	18.3	14.5

100

Chapitre 3: Distribution statistique à un caractère

$$n = 10 \qquad k = 1 + 3.3\log_{10}(10) = 4.33$$

Individus	Force (Modalité) N
1	10
2	10.3
3	22.1
4	23
5	26
6	15.2
7	12.3
8	18
9	18.3
10	14.5

$$k = 5$$

$$V_{\min} = 10\ \mathrm{N}$$

$$V_{\max} = 26\ \mathrm{N}$$

$$Pas = \frac{26-10}{5} = 3.2\ \ \mathrm{N}$$

$$Pas = 4\ \ \mathrm{N}$$

101

Chapitre 3: Distribution statistique à un caractère

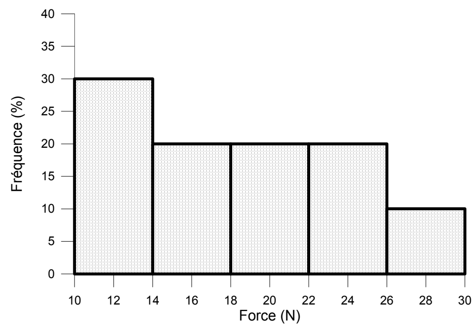
Individus	Force (Modalité) en Newton (N)
1	10
2	10.3
3	22.1
4	23
5	26
6	15.2
7	12.3
8	18
9	18.3
10	14.5

Limites des classes (N)		Centre de classe (N)	Effectif	Fréquence (%)
Limite inf	Limite sup			
10	14	12	3	30
14	18	16	2	20
18	22	20	2	20
22	26	24	2	20
26	30	28	1	10

102

Individus	Force (Modalité) en Newton (N)
1	10
2	10.3
3	22.1
4	23
5	26
6	15.2
7	12.3
8	18
9	18.3
10	14.5

Individus	Force (Modalité) en Newton (N)
1	12
2	12
3	24
4	24
5	28
6	16
7	12
8	20
9	20
10	16



10 valeurs de F sont observées.

Question:

Quelle valeur de F doit-on utiliser dans l'équation X=F/K

La plus petite ?
La plus grande?

La plus significative

Que recherche le concepteur

•Réduire un ensemble de données
•Conserver une partie de l'information

Résumer l'ensemble des valeurs par une seule valeur

Parfois c'est difficile à accepter

Caractéristique de tendance centrale

Elle décrit l'ordre de grandeur des valeurs et aussi la valeur centrale autour de laquelle se regroupent les observations.

Caractéristique de tendance centrale
(paramètres de position)

Choisir

•Série statistique

•Tableau statistique

1. Mode
2. Médiane
3. Moyenne (s)
4. Quantile

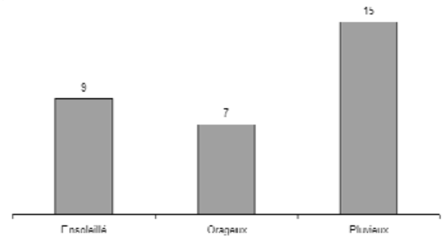
Mode

Le mode, noté *Mo*, est la modalité qui admet **la plus grande fréquence** :

Il est parfaitement défini pour une variable qualitative ou une variable quantitative discrète.

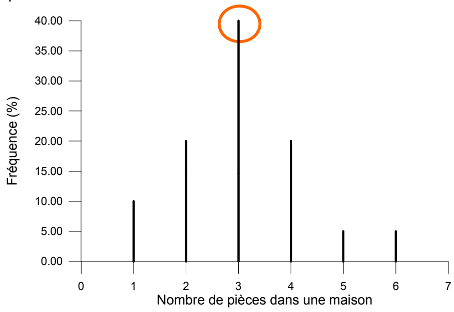
Pour une variable quantitative continue nous parlons de **classe modale** : c'est la classe dont la densité de fréquence est maximum.

Variable qualitative



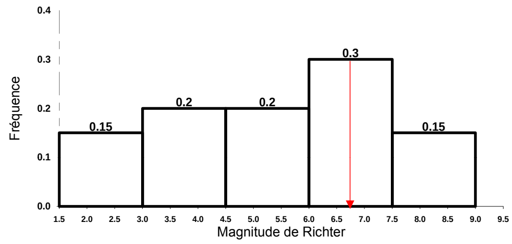
Mo= Pluvieux

Variable quantitative discrète



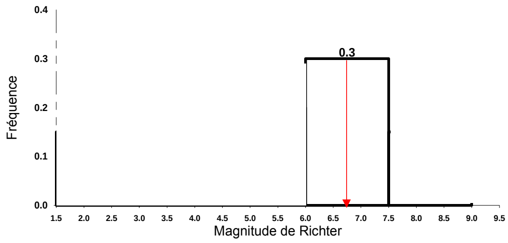
Mo=3 pièces

Variable quantitative continue



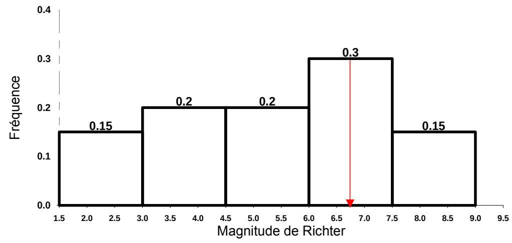
Milieu de la classe dont la fréquence est la plus grande
Mo=6.75

Variable quantitative continue



Milieu de la classe dont la fréquence est la plus grande
Mo=6.75

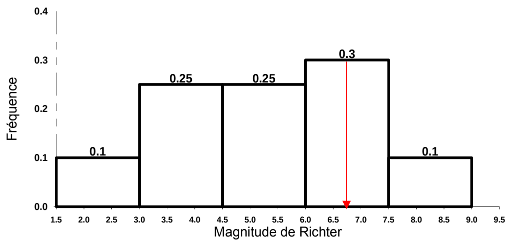
Variable quantitative continue



Milieu de la classe dont la fréquence est la plus grande
Mo=6.75

Chapitre 3: Distribution statistique à un caractère

Variable quantitative continue

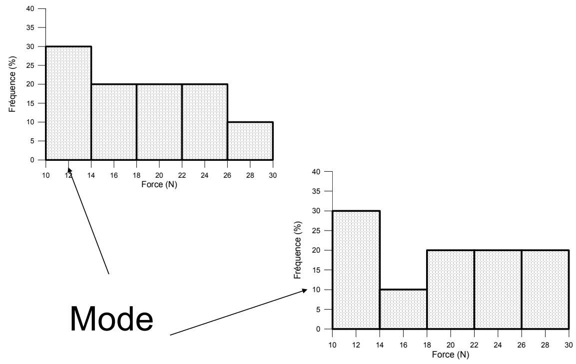


Mo=6.75

Le même résultat a été obtenu
pourtant les deux distributions ne sont pas identiques

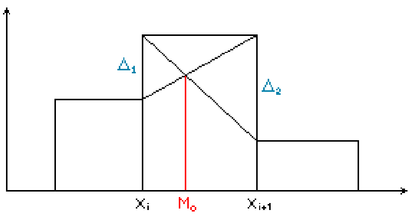
115

Chapitre 3: Distribution statistique à un caractère



116

Chapitre 3: Distribution statistique à un caractère

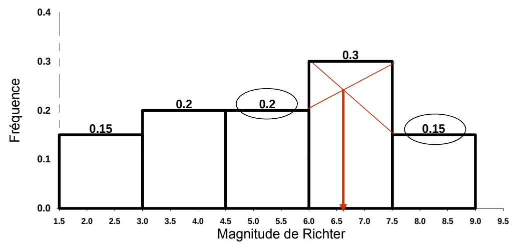


$$\frac{M_o - x_i}{\Delta_1} = \frac{x_{i+1} - M_o}{\Delta_2}$$

$$\boxed{M_o = x_i + \frac{\Delta_1}{\Delta_1 + \Delta_2}(x_{i+1} - x_i)}$$

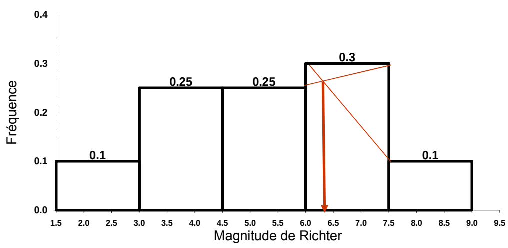
117

Chapitre 3: Distribution statistique à un caractère



118

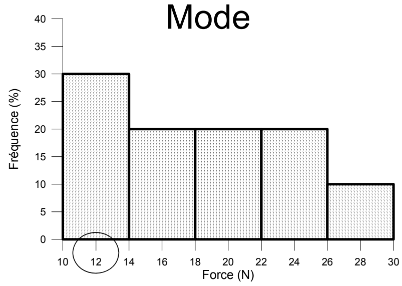
Chapitre 3: Distribution statistique à un caractère



119

Chapitre 3: Distribution statistique à un caractère

Mode



F=12 N

X=12/10=1.2cm

120

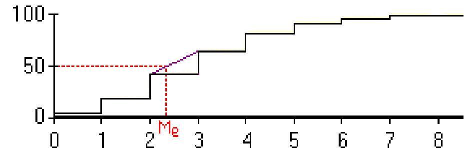
Médiane

La médiane *Me* est telle que l'effectif des observations dont les modalités sont inférieures à *Me* est égal à l'effectif des observations dont les modalités sont supérieures à *Me*.

Utilise la notion de fonction cumulative

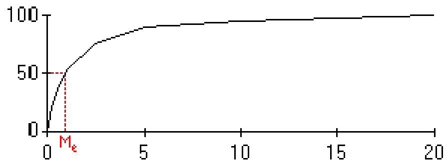
Quantitative discret

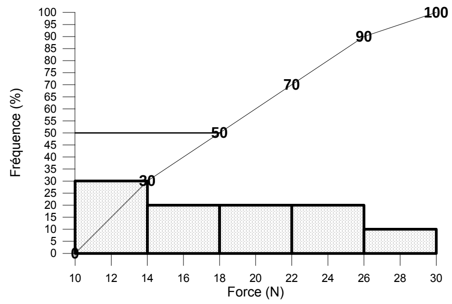
Courbe en escalier



Quantitative continue

Courbe cumulative





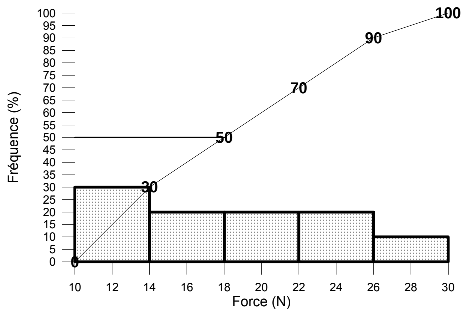
F=18 N

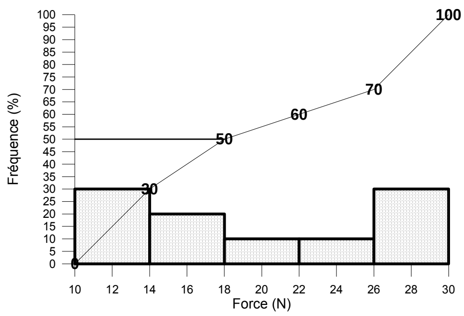
X=18/10=1.8cm

Médiane

Avantage:
- simple à calculer
- Prend en compte une partie des valeurs

Inconvénient:
ne tient pas compte de la distribution des modalités supérieures à la médiane

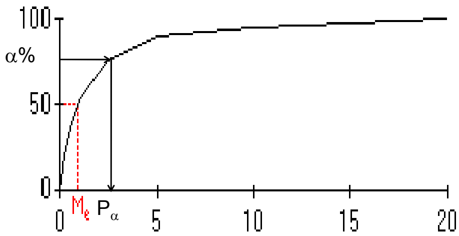




Valeur de tendance centrale

Médiane=18 N
Mode=12 Nm

Quantile



Quantile

Le quantile α est la valeur P_α qui laisse α% des observations en-dessous et $(1-\alpha)$% des observations au-dessus d'elle.

Les deux "quartiles" les plus importants sont P_{25} (qui laisse 25 % des observations en-dessous) et P_{75}.

Moyenne <u>(s)</u>

La moyenne ne se définit que pour une variable statistique quantitative.

Moyenne arithmétique dite par abréviation MOYENNE

Individus	Force (Modalité) N
1	10
2	10.3
3	22.1
4	23
5	26
6	15.2
7	12.3
8	18
9	18.3
10	14.5

$$\overline{F} = \frac{\sum_{i=1}^{10} F_i}{10} = \sum_{i=1}^{10}\left(\frac{1}{10}F_i\right)$$

$$= 16.97\,N$$

	Limites des classes (N)		Centre de classe (N) F_j	Effectif	f_j Fréquence (%)	$f_j F_j$
	Limite inf	Limite sup				
$j=1$	10	14	12	3	30	3.6
$j=2$	14	18	16	2	20	3.2
$j=3$	18	22	20	2	20	4
$j=4$	22	26	24	2	20	4.8
$j=5$	26	30	28	1	10	2.8

$$\overline{F} = \sum_{j=1}^{5} f_j F_j$$
$$= (3.6 + 3.2 + 4 + 4.8 + 2.8)$$
$$= 18.4N$$

Série statistique: Moyenne arithmétique=16.97 N

Tableau statistique: Moyenne arithmétique=18.40 N

DIFFERENCE ?????????????

Individus	Force (Modalité) en Newton (N)
1	10
2	10.3
3	22.1
4	23
5	26
6	15.2
7	12.3
8	18
9	18.3
10	14.5

$$\overline{F} = 16.97N$$

Individus	Force (Modalité) en Newton (N)
1	12
2	12
3	24
4	24
5	28
6	16
7	12
8	20
9	20
10	16

$$\overline{F} = 18.4N$$

Valeur de tendance centrale

⇩

Existe-t-il une seule moyenne

$$\overline{F}_R = \left(\sum_{j=1}^{j=k} \left[f_j \left(F_j \right)^R \right] \right)^{\frac{1}{R}}$$

R=1----> Moyenne arithmétique

$$\overline{F}_1 = \left(\sum_{j=1}^{j=k} \left[f_j \left(F_j \right)^1 \right] \right)^{\frac{1}{1}}$$

$$\overline{F}_1 = \left(\sum_{j=1}^{j=k} \left[f_j F_j \right] \right)$$

R=0----> Moyenne géométrique

$$\overline{F}_0 = \left(\sum_{j=1}^{j=k} \left[f_j \left(F_j \right)^0 \right] \right)^{\frac{1}{0}}$$

$$\log\left(\overline{F}_0\right) = \sum_{j=1}^{j=k} f_j \log\left(F_j\right)$$

R=-1----→ Moyenne harmonique

$$\overline{F}_{-1} = \left(\sum_{j=1}^{j=k}\left[f_j\left(F_j\right)^{-1}\right]\right)^{\frac{1}{-1}}$$

$$\overline{F}_{-1} = \frac{1}{\sum_{j=1}^{j=k}\left[f_j\,\frac{1}{F_j}\right]}$$

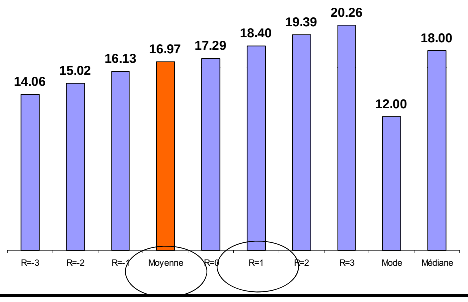
R=2----→ Moyenne quadratique

$$\overline{F}_{2} = \left(\sum_{j=1}^{j=k}\left[f_j\left(F_j\right)^{2}\right]\right)^{\frac{1}{2}}$$

$$\overline{F}_{2} = \sqrt{\sum_{j=1}^{j=k}\left[f_j\left(F_j\right)^{2}\right]}$$

Quelle moyenne choisir????

Etude comparative entres les moyennes et la moyenne réelle obtenue directement du tableau brut



Démontrer que

$$\sum_{j=1}^{j=k}\left[f_j\left(F_j-\overline{F}\right)\right]=0$$

$$\phi(a)=\sum_{j=1}^{j=k}\left[f_j\left(F_j-\overline{F}\right)^2\right]+\left(\overline{F}-a\right)^2$$

$$=\sum_{j=1}^{j=k}\left[f_j\left(F_j-a\right)^2\right]$$

Est-ce la réduction d'un ensemble de valeurs à un seule valeur est une étape suffisante

•Réduire un ensemble de données

•Conserver une partie de l'information

On souhaite donner un prix:
Meilleur classe

Critère de classement: **Moyenne**

Pour faire simple, on suppose que deux classes sont candidates pour ce prix et que chacune des deux classes est composée de 10 élèves

12	13	08	07	11	09	10,5	9,5	10	10

CLASSE –A-

17	13	03	07	19	18	01	02	16	04

CLASSE –B-

MOYENNE=10

Qui mérite de recevoir le prix? ???

Question
Est-ce que les caractéristiques de tendance centrale sont suffisantes pour identifier une valeur de F à utiliser

Tableau n°1

15.25	16.25	18.36	19.26	19.23	15.2	11.2	20.35	18.3	16.3

Tableau n°2

10.00	10.30	22.10	23.00	26.00	15.20	12.30	18.00	18.30	14.50

Les deux tableaux présentent la même moyenne arithmétique=16.97 N

Question: Est-ce que les deux tableaux sont identiques

Trouver une valeur qui reflète

au mieux

La dispersion des valeurs de notre échantillon

Valeur de dispersion

3.2 Description numérique

3.2.2 Caractéristique de dispersion

Caractéristique de dispersion

Choisir

•Série statistique

•Tableau statistique

1. Etendue
2. L'intervalle inter-quartiles
3. Ecart moyen
4. Ecart type

Tableau n°1

15.25	16.25	18.36	19.26	19.23	15.2	11.2	20.35	18.3	16.3

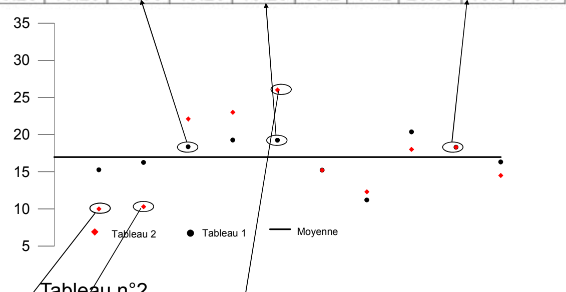


Tableau n°2

10.00	10.30	22.10	23.00	26.00	15.20	12.30	18.00	18.30	14.50

Caractéristiques de dispersion

Mesurer la différence qui existe entre la valeur max et min
ETENDUE

Tableau n°1

15.25	16.25	18.36	19.26	19.23	15.2	11.2	20.35	18.3	16.3

ETENDUE=20.35-11.2=**9.15 N**

Tableau n°2

10.00	10.30	22.10	23.00	26.00	15.20	12.30	18.00	18.30	14.50

ETENDUE=26-10=**16.0 N**

Caractéristiques de dispersion

ETENDUE (Tableau 1)= **9.15 N**

ETENDUE (Tableau 2)= **16.0 N**

ETENDUE (Tableau 2)>ETENDUE (Tableau 1)

C'est logique mais attention cette valeur peut vous fausser l'interprétation car elle se base sur les valeurs extrêmes

Tableau n°1

15.25	16.25	18.36	19.26	19.23	15.2	11.2	20.35	18.3	16.3

Moyenne= 16.97 N et ETENDUE=20.35-11.2=**9.15 N**

Tableau n°2

10.00	10.30	22.10	23.00	26.00	15.20	12.30	18.00	18.30	14.50

Moyenne= 16.97 N et ETENDUE=26-10=**16.0 N**

Tableau n°3

10.00	17.00	17.40	18.00	26.00	16.00	14.30	17.20	17.30	16.50

Moyenne= 16.97 N et ETENDUE=26-10=**16.0 N**

Tableau n°1

Moyenne= 16.97 N et ETENDUE=20.35-11.2=**9.15 N**

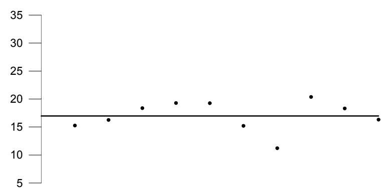


Tableau n°2

Moyenne= 16.97 N et ETENDUE=26-10=**16.0 N**

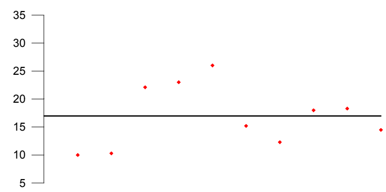
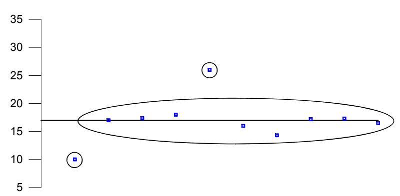
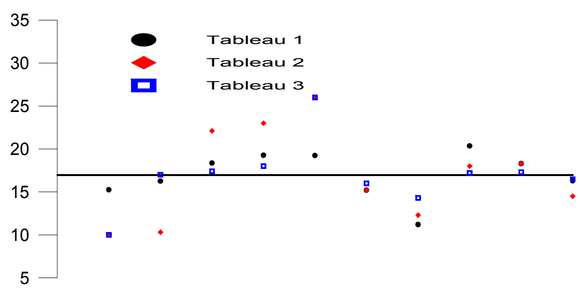


Tableau n°3

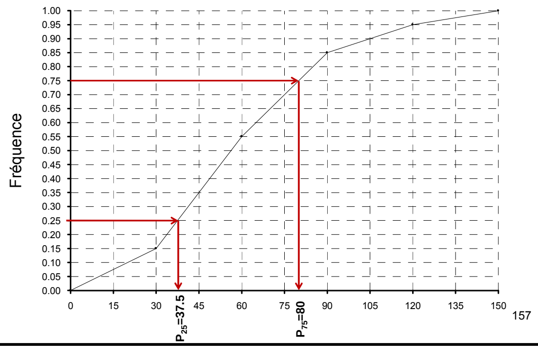
Moyenne= 16.97 N et ETENDUE=26-10=**16.0 N**





Caractéristiques de dispersion

L'intervalle inter-quartiles $H = P_{75} - P_{25}$



Caractéristiques de dispersion

Mesurer la dispersion des valeurs par rapport à la moyenne

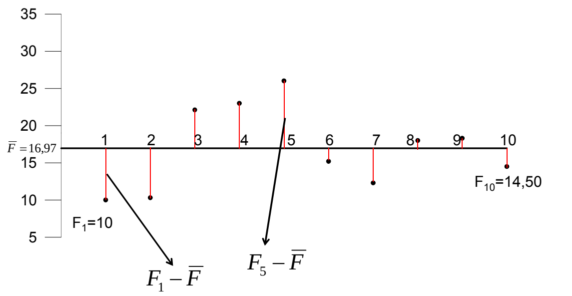
Estimer la différence entre

les valeurs observées

Et

la moyenne

10.00	10.30	22.10	23.00	26.00	15.20	12.30	18.00	18.30	14.50



Valeur de dispersion

Cette valeur DOIT être très petite

$$V = (F_1 - \bar{F}) + (F_2 - \bar{F}) + \ldots + (F_{10} - \bar{F})$$

$$V = (F_1 + F_2 + \ldots + F_{10} - 10\bar{F})$$

$$V = (10\bar{F} - 10\bar{F}) = 0$$

Valeur de dispersion

Cette valeur DOIT être très petite

$$V = (F_1 - \bar{F}) + (F_2 - \bar{F}) + \ldots + (F_{10} - \bar{F}) = 0$$

Or cette valeur est en fait nulle et ne peut donc mesurer la ***dispersion***

Autre méthode

Valeur de dispersion

$$V = |F_1 - \bar{F}| + |F_2 - \bar{F}| + \ldots + |F_{10} - \bar{F}|$$

Cette valeur permet d'estimer la différence

La valeur absolue est une fonction mathématique difficilement dérivable

Valeur de dispersion

$$V = (F_1 - \overline{F})^2 + (F_2 - \overline{F})^2 + \ldots + (F_{10} - \overline{F})^2$$

Cette valeur est toujours positif et permet d'estimer la différence entre les valeurs observées et leur moyenne

Valeur de dispersion

Mesurer la dispersion des valeurs par rapport à la moyenne

$$V = \sum_{i=1}^{10} (F_i - \overline{F})^R \text{ et } R \text{ pair}$$

Série statistique — ***Tableau statistique***

Moyenne

$$\overline{F} = \sum_{i=1} \frac{F_i}{n} \qquad \overline{F} = \sum_{j=1}^{j=k} f_j F_j$$

Écart moyen

$$EM_F = \sum_{i=1}^{10} \left[\frac{1}{n} \left| F_i - \overline{F} \right| \right] \qquad EM_F = \sum_{j=1}^{j=k} \left[f_j \left| F_j - \overline{F} \right| \right]$$

Série statistique — ***Tableau statistique***

Variance

$$(\sigma_F)^2 = \sum_{i=1} \left[\frac{(F_i - \overline{F})^2}{n} \right] \qquad (\sigma_F)^2 = \sum_{j=1}^{j=k} \left[f_j (F_j - \overline{F})^2 \right]$$

Ecart type

$$\sigma_F$$

Série statistique — ***Tableau statistique***

Différence d'ordre R

$$V_R = \left(\sum_{i=1}^{i=n} \left[\frac{1}{n} (F_i - \overline{F})^R \right] \right)^{\frac{1}{R}} \qquad V_R = \left(\sum_{j=1}^{j=k} \left[f_j (F_j - \overline{F})^R \right] \right)^{\frac{1}{R}}$$

Exercice: Déterminer l'Ecart Moyen et l'Ecart type

10.00	10.30	22.10	23.00	26.00	15.20	12.30	18.00	18.30	14.50

Limites des classes (N)		**Centre de classe (N)**	**Effectif**	**Fréquence (%)**
Limite inf	Limite sup			
10	14	12	3	30
14	18	16	2	20
18	22	20	2	20
22	26	24	2	20
26	30	28	1	10

Lim inf	Lim sup	Centre de classe F_k	Effectif	Fréquence f_k (%)	$f_k F_k$	$F_k - \overline{F}$	$f_k(F_k - \overline{F})$
10	14	12	3				
14	18	16	2				
18	22	20	2				
22	26	24	2				
26	30	28	1				

Lim inf	Lim sup	Centre de classe F_k	Effectif	Fréquence f_k (%)	$f_k F_k$	$F_k - \overline{F}$	$f_k(F_k - \overline{F})$
10	14	12	3	30			
14	18	16	2	20			
18	22	20	2	20			
22	26	24	2	20			
26	30	28	1	10			

Lim inf	Lim sup	Centre de classe F_k	Effectif	Fréquence f_k (%)	$f_k F_k$	$F_k - \overline{F}$	$f_k(F_k - \overline{F})$
10	14	12	3	30	3.6		
14	18	16	2	20	3.2		
18	22	20	2	20	4		
22	26	24	2	20	4.8		
26	30	28	1	10	2.8		
					$\overline{F}$ = 18.4 N		

Lim inf	Lim sup	Centre de classe F_k	Effectif	Fréquence f_k (%)	$f_k F_k$	$F_k - \overline{F}$	$f_k(F_k - \overline{F})$
10	14	12	3	30	3.6	-6.4	
14	18	16	2	20	3.2	-2.4	
18	22	20	2	20	4	1.6	
22	26	24	2	20	4.8	5.6	
26	30	28	1	10	2.8	9.6	
					$\overline{F}$ = 18.4 N		

Lim inf	Lim sup	Centre de classe F_k	Effectif	Fréquence f_k (%)	$f_k F_k$	$F_k - \overline{F}$	$f_k(F_k - \overline{F})$
10	14	12	3	30	3.6	-6.4	-1.92
14	18	16	2	20	3.2	-2.4	-0.48
18	22	20	2	20	4	1.6	0.32
22	26	24	2	20	4.8	5.6	1.12
26	30	28	1	10	2.8	9.6	0.96
					$\overline{F}$ = 18.4 N		0.0

$$\sum_{j=1}^{j=k}\left[f_j\left(F_j-\overline{F}\right)\right]=0$$

Lim inf	Lim sup	Centre de classe F_k	Effectif	Fréquence f_k (%)	$f_k F_k$	$F_k - \overline{F}$	$\lvert F_k - \overline{F}\rvert$	$f_k\lvert F_k - \overline{F}\rvert$
10	14	12	3	30	3.6	-6.4	6.4	1.92
14	18	16	2	20	3.2	-2.4	2.4	0.48
18	22	20	2	20	4	1.6	1.6	0.32
22	26	24	2	20	4.8	5.6	5.6	1.12
26	30	28	1	10	2.8	9.6	9.6	0.96
					$\overline{F}$ = 18.4 N			4.8 N

$$V=\sum_{j=1}^{j=k} f_j\left|F_j-\overline{F}\right|=4.8N$$

Ecart Moyen= 4.8 N

Lim inf	Lim sup	Centre de classe	Effectif	Fréquence f_k (%)	$f_k F_k$	$F_k - \bar{F}$	$(F_k - \bar{F})^2$	$f_k(F_k - \bar{F})^2$
10	14	12	3	30	3.6	-6.4	40.96	12.288
14	18	16	2	20	3.2	-2.4	5.76	1.152
18	22	20	2	20	4	1.6	2.56	0.512
22	26	24	2	20	4.8	5.6	31.36	6.272
26	30	28	1	10	2.8	9.6	92.16	9.216
					$\bar{F}$ =18.4 N	8		29.44

Ecart type= $\sqrt{29.44} = 5.42N$

Lim inf	Lim sup	Centre de classe F_k	Effectif	Fréquence f_k (%)	$f_k F_k$	$F_k - \bar{F}$	$(F_k - \bar{F})^2$	$f_k(F_k - \bar{F})^2$
11	13.4	12.2	1	10	1.22	-5.04	25.4016	2.54016
13.4	15.8	14.6	2	20	2.92	-2.64	6.9696	1.39392
15.8	18.2	17	2	20	3.4	-0.24	0.0576	0.01152
18.2	20.6	19.4	5	50	9.7	2.16	4.6656	2.3328
					17.24 N			6.2784

Ecart Type= $\sqrt{6.2784} = 2.50$ N

Tableau n°1

15.25	16.25	18.36	19.26	19.23	15.20	11.20	20.35	18.30	16.25

$$(\sigma)^2 = \sum_{j=1}^{j=k} f_j \left(F_j - \bar{F}\right)^2 = 6.28\, N^2$$

Tableau n°2

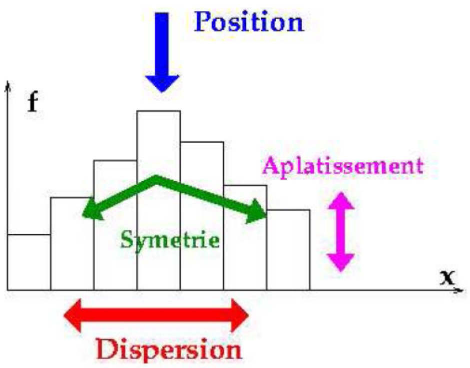
10.00	10.30	22.10	23.00	26.00	15.20	12.30	18.00	18.30	14.50

$$(\sigma)^2 = \sum_{j=1}^{j=k} f_j \left(F_j - \bar{F}\right)^2 = 29.44\, N^2$$

3.3 Caractéristiques de formes

1. Coefficient de variation
2. Coefficient de symétrie
3. Coefficient d'aplatissement

Coefficient de variation (noté **CV**)=Ecart type / Moyenne

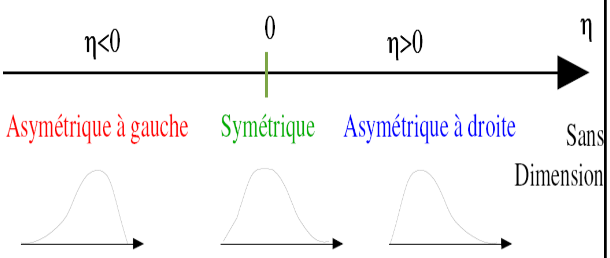


Coefficient de symétrie **(Cœfficient of Skewness)**

$$\eta = \sum_{i=1}^{i=n} \frac{1}{n} \frac{\left(F_i - \bar{F}\right)^3}{\left(\sigma_F\right)^3}$$

$$\eta = \sum_{j=1}^{j=k} f_j \frac{\left(F_j - \bar{F}\right)^3}{\left(\sigma_F\right)^3}$$

Coefficient de symétrie **(Cœfficient of Skewness)**

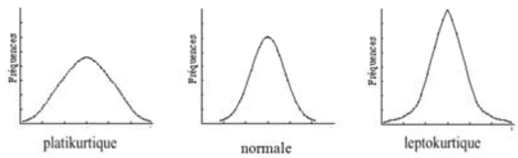


Coefficient d'aplatissement **(Cœfficient of Kurtosis)**
généralement comparé à la valeur de 3 qui est celui de la distribution

$$\kappa = \sum_{i=1}^{i=n} \frac{1}{n} \frac{\left(F_i - \overline{F}\right)^4}{\left(\sigma_F\right)^4}$$

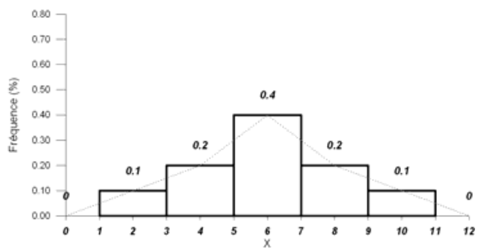
$$\kappa = \sum_{j=1}^{j=k} f_j \frac{\left(F_j - \overline{F}\right)^4}{\left(\sigma_F\right)^4}$$

Coefficient d'aplatissement **(Cœfficient of Kurtosis)**
généralement comparé à la valeur de 3 qui celle d'une distribution normale



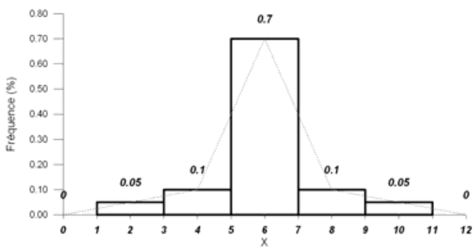
$\kappa - 3 < 0$ $\kappa - 3 = 0$ $\kappa - 3 > 0$

Exemple:



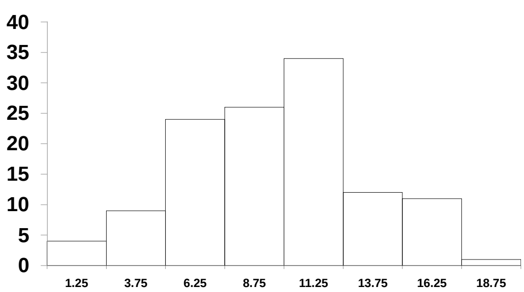
$$\kappa - 3 = -0{,}5 < 0$$

Exemple: La distribution normale est symétrique



$$\kappa - 3 = 2 > 0$$

Les notes des étudiants



Fin du Chapitre 3

Distribution statistique à un caractère

3.1 Description graphique
3.2 Description numérique
3.2.1 Caractéristique de tendance centrale
3.2.2 Caractéristique de dispersion
3.3 Caractéristiques de formes

Chapitre 4

Distribution statistique à deux caractères

4.1 Introduction
4.2 Description numérique
4.3 Principe de la méthode des moindres carrées "Least Square method »
4.4 Covariance

Caractère ou variable statistique

Qualitatif

Exprimée par une description naturelle du langage (ex: une couleur)

Quantitatif

exprimée par des nombres (ex: une taille)

Discret

Réel

k modalités $c_1, c_2, \ldots, c_k$

4.1 Introduction

Un même individu peut il avoir plusieurs caractères????

Oui

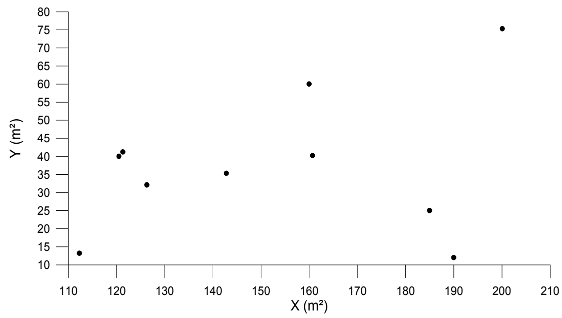
Individu	Maison
Caractère 1	Surface habitable (notée X)
Caractère 2	Surface non habitable (notée Y)

Surface non habitable: Le jardin et autres…

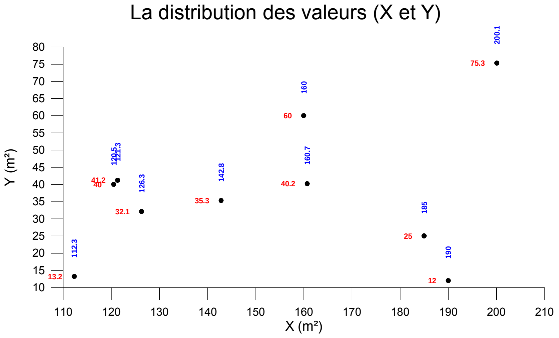
Ce qui est recherché c'est

De possibles relations entre le caractère 1 et le caractère 2.

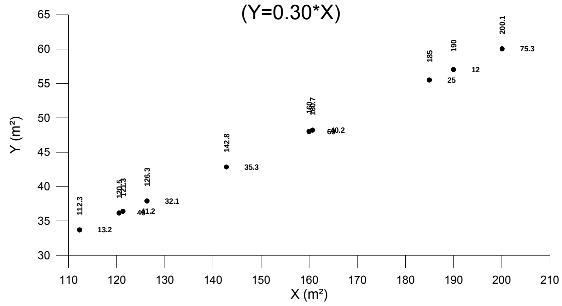
Maison	X (m²)	Y (m²)
1	200,1	75,3
2	121,3	41,2
3	160,7	40,2
4	112,3	13,2
5	190	12
6	120,5	40
7	160	60
8	185	25
9	126,3	32,1
10	142,8	35,3



La distribution des valeurs (X et Y)



Voilà ce que prévoit les codes de l'urbanisme (Y=0.30*X)



4.2 Description numérique

Individu	X	Y
1	x_1	y_1
2	x_2	y_2
⋮	⋮	⋮
e	x_e	y_e
⋮	⋮	⋮
n	x_n	y_n

Il est possible de passer de la série statistique vers le tableau statistique

Comment?

Tableau statistique

Classe Y / Classe X		*(1)* [CY_1, CY_2[	…	*(j)* [CY_j, CY_{j+1}[	…	*(k)* [CY_k, CY_{k+1}[
		Y_1		Y_j		Y_K
(1) [CX_1, CX_2[	X_1	n_{11}		n_{1j}		n_{1k}
⋮						
(i) [CX_i, CX_{i+1}[	X_i	n_{i1}		n_{ij}		n_{ik}
⋮						
(m) [CX_m, CX_{m+1}[	X_m	n_{m1}		n_{mj}		n_{mk}

n_{11}: l'ensemble des individus ayant la modalité X_1 de X et Y_1 de Y

$$n = \sum_{i=1}^{i=m} \sum_{j=1}^{j=k} n_{ij}$$

Tableau statistique

Classe Y / Classe X		*(1)* [CY_1, CY_2[	…	*(j)* [CY_j, CY_{j+1}[	…	*(k)* [CY_k, CY_{k+1}[
		Y_1		Y_j		Y_K
(1) [CX_1, CX_2[	X_1	f_{11}		f_{1j}		f_{1k}
⋮						
(i) [CX_i, CX_{i+1}[	X_i	f_{i1}		f_{ij}		f_{ik}
⋮						
(m) [CX_m, CX_{m+1}[	X_m	f_{m1}		f_{mj}		f_{mk}

f_{11}: Fréquence des individus ayant la modalité x_1 de X et y_1 de Y

$$f_{ij} = \frac{n_{ij}}{n}$$

Individu	X	Y
1	x_1	y_1
2	x_2	y_2
⋮	⋮	⋮
e	x_e	y_e
⋮	⋮	⋮
n	x_n	y_n

$$\overline{\overline{X}} = \sum_{e=1}^{e=n} \frac{x_n}{n} \qquad \overline{\overline{Y}} = \sum_{e=1}^{e=n} \frac{y_n}{n}$$

$$\sigma_X = \sqrt{\sum_{e=1}^{e=n} \left[\frac{1}{n} \left(x_i - \overline{\overline{X}} \right)^2 \right]} \qquad \sigma_Y = \sqrt{\sum_{e=1}^{e=n} \left[\frac{1}{n} \left(y_i - \overline{\overline{Y}} \right)^2 \right]}$$

Fréquence marginale

	Y_1	...	Y_j		Y_K	
X_1	f_{11}		f_{1j}		f_{1k}	$f_{1.}$
⋮						
X_i	f_{i1}		f_{ij}		f_{ik}	$f_{j.}$
⋮						
X_m	f_{m1}		f_{mj}		f_{mk}	$f_{m.}$
	$f_{.1}$		$f_{.j}$		$f_{.k}$	

$f_{.1}$: Fréquence des individus ayant la modalité y1 de Y

$$f_{.1} = \sum_{i=1}^{i=m} f_{i1}$$

Fréquence conditionnelle

Répond à la question suivante:

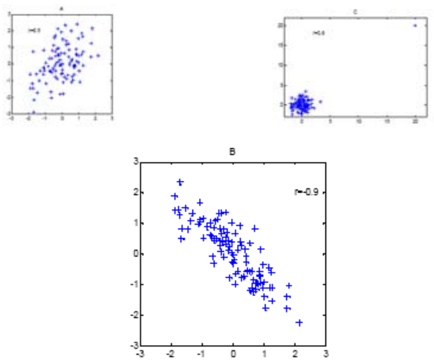
Quelle la fréquence des individus ayant la modalité x_1 de X sachant que $Y=y_1$

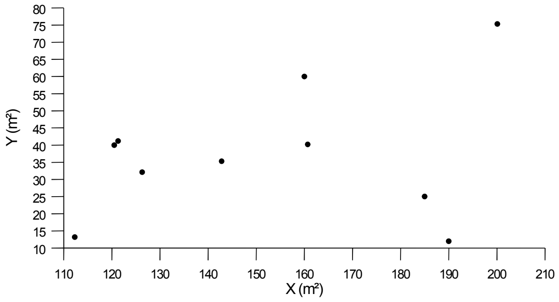
	Y_1	...	Y_j		Y_K	
X_1	f_{11}		f_{1j}		f_{1k}	$f_{1.}$
⋮						
X_i	f_{i1}		f_{ij}		f_{ik}	$f_{j.}$
⋮						
X_m	f_{m1}		f_{mj}		f_{mk}	$f_{m.}$
	$f_{.1}$		$f_{.j}$		$f_{.k}$	

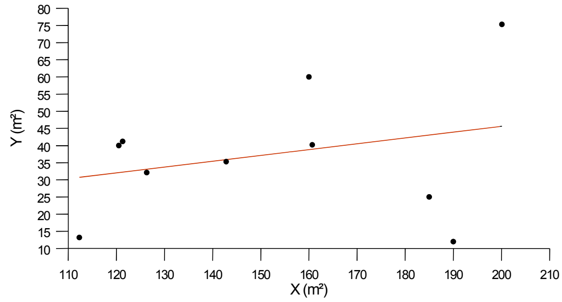
	Y_1	...				
X_1	f_{11}					
⋮						
X_i	f_{i1}					
⋮						
X_m	f_{m1}					
	$f_{.1}$					

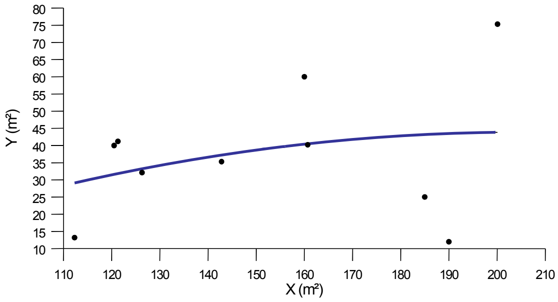
$$\frac{f_{11}}{f_{.1}} = \frac{\dfrac{n_{11}}{n}}{\dfrac{\sum_{i=1}^{i=m} n_{i1}}{n}} = \frac{n_{11}}{\sum_{i=1}^{i=m} n_{i1}}$$

•4.3 Principe de la méthode des moindres carrées "Least Square method »





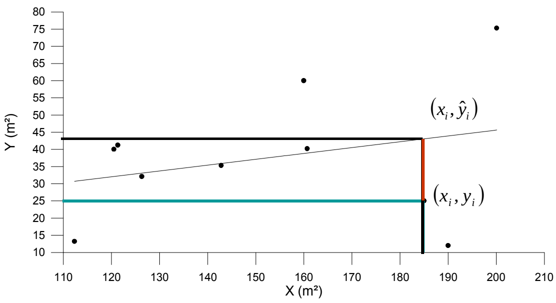




Problème étudié:
Chercher une relation de corrélation entre deux variables statistiques X et Y.

Quelles sont nos données:
On connaît l'ensemble des couples (x_i,y_i) $i=1,..,n$

Comment faire:
Utiliser la méthode des moindres carrés



$er_i = y_i - \hat{y}_i$ Soit très petite

En considérant une régression linéaire: La méthode des moindre carrés vous permet d'écrire la relation:

y=ax+b

Et de ce fait vous permet de déterminer les coefficients a et b (la démonstration sera donnée au niveau du cours)

Une corrélation n'a de valeurs que s'il y'a détermination du coefficient de corrélation

•4.4 Covariance

$$Cov(X,Y)=\sum_{i=1}^{i=m}\sum_{j=1}^{j=k} f_{ij}\left(x_i-\overline{\overline{X}}\right)\left(y_j-\overline{\overline{Y}}\right)$$

$$Cov(X,Y)=\sum_{i=1}^{i=n}\frac{1}{n}\left(x_i-\overline{\overline{X}}\right)\left(y_i-\overline{\overline{Y}}\right)$$

Fin du Chapitre 4

Distribution statistique à deux caractères

4.1 Introduction
4.2 Description numérique
4.3 Principe de la méthode des moindres carrées "Least Square method »
4.4 Covariance

Chapitre 5

Distribution statistique à plusieurs caractères

Partie C
Traitement probabiliste de l'information

Chapitre 6 Théorie de la probabilité

6.1 Espace de probabilité
6.2 Axiome de probabilité
6.3 Propriétés des évènements
6.4 Quelques notions supplémentaires

6.1 Espace de probabilité

On lance un dé , l'ensemble de tous les résultats possibles est identifié par

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

Un résultat possible de l'expérience est noté ω.
Ainsi, $\omega \in \Omega$

D'autres exemples:

On lance une pièce: $\Omega = \{P,F\}$

Durée de vie d'une pièce $\Omega = [0,+\infty[$

Une expérience aléatoire se décrit mathématiquement par la donnée d'un espace Ω dont les points notés ω sont les résultats possibles de l'expérience, ainsi que d'une probabilité P sur Ω.

Un événement A lié à l'expérience Ω est représenté par une partie de Ω noté A.

Chaque événement possède une probabilité P[A] qui est un nombre compris entre 0 et 1

Par exemple, considérons le jeu de dé:

Un <u>événement</u> A peut décrit comme suit:

- Obtenir un nombre pair
- Obtenir 4

A chaque évènement correspond une probabilité notée P[A]

P[A]: Probabilité d'avoir l'évènement A

(Ω, A, P) espace de probabilité

6.2 Axiome de probabilité

1. Soit un évènement A alors $0 \le P[A] \le 1$

2. $P[\Omega]=1=$ *événement certain*

3. Si A et B sont deux évènements mutuellement exclusive alors

P[AUB]=P[A]+P[B]

6.3 Propriétés des évènements

Soit A et B deux évènements qui ne sont nécessairement mutuellement exclusive alors

$$P[A \cup B]=P[A]+P[B]-P[A \cap B]$$

6.4 Quelques notions supplémentaires

-1- <u>Probabilité conditionnelle</u> :

Probabilité d'avoir l'évènement A sachant que l'évènement B s'est produit
Ne pas confondre entre évènements mutuellement exclusive et indépendants
Voilà un petit exemple en statistique

Population: Maisons (20)
Individu : Une maison parmi les 20
Caractères: X (Surface habitable), Y (Surface non habitable)
Modalités : Valeurs de X et Y

$$P[A/B]=\frac{P[A \cap B]}{P[B]}$$

	Y_1	Y_2	Y_3	
X_1	$f_{11}=0.05$	$f_{12}=0.05$	$f_{13}=0.3$	$\mathbf{f_{1.}=0.4}$
X_2	$f_{21}=0.3$	$f_{22}=0.1$	$f_{23}=0.2$	$\mathbf{f_{2.}=0.6}$
	$\mathbf{f_{.1}=0.35}$	$\mathbf{f_{.2}=0.15}$	$\mathbf{f_{.3}=0.5}$	

$$\frac{f_{11}}{f_{.1}}=\frac{0.05}{0.35}=0.143$$

-2- <u>Théorème de la probabilité Totale</u>

<u>Je l'appelle le théorème de la Vie</u>

Considérons B_i (i=1,…,n) évènements ME et CE alors:

$$P[A] = \sum_{i=1}^{n} P[A \cap B_i] = \sum_{i=1}^{n} P[A|B_i]\, P[B_i]$$

-3- Théorème de Bayes

$$P\left[A_i / B\right] = \frac{P\left[B / A_i\right]P\left[A_i\right]}{\sum_k P\left[B / A_k\right]P\left[A_k\right]}$$

Fin du Chapitre 6 Théorie de la probabilité

6.1 Espace de probabilité
6.2 Axiome de probabilité
6.3 Propriétés des évènements
6.4 Quelques notions supplémentaires

Chapitre 7 Variable aléatoire

7.1 Définitions
7.2 Variable aléatoire discrète
7.3 Variable aléatoire continue
7.4 Caractéristiques de tendance centrale
7.5 Caractéristiques de dispersion

7.1 Définitions

Une variable aléatoire est une fonction qui à chaque évènement élémentaire d'une expérience aléatoire associe un nombre réel ou entier.

Soit X une variable aléatoire et *k* un réel ou entier, l'évènement noté (X = *k*) est l'antécédent de *k* par la fonction

X : c'est l'ensemble de tous les événements élémentaires dont l'image par X est égale à *k*.

7.2 Variable aléatoire discrète

k est un nombre entier

Soit une épreuve qui ne comporte que deux résultats possibles, incompatibles.

Exemples:
•Arrivée d'un avion: A l'heure ou non
•Pièce de monnaie: Pile ou face

A évènement décrit par:
L'avion arrive à l'heure
B évènement décrit par:
L'avion n'arrive pas à l'heure prévue

Les évènement A et B sont

•Mutuellement exclusive(ME)
•Collectivement exhaustive (CE)

Pourquoi???

Car:
Si l'évènement A se produit alors l'évènement B ne peut pas se produire et inversement →ME

Car on ne peut avoir que les évènements A ou B. On ne peut avoir un autre évènement →CE

La probabilité de réalisation de l'événement A est

p

La probabilité de réalisation de l'événement B est

q

Alors $p+q=1$

Dans le cas d'une pièce de monnaie

p=0.5 ***q=0.5***

On note $P_X(x)$ la loi de probabilité de la variable aléatoire X.
Cette dernière peut prendre deux valeurs:

$x=1$ ou $x=0$

La fonction $P_X(x)$, dite **<u>Fonction de Probabilité de masse (FPM)</u>** dépend de x:

$P_X(x=1)$ écrite aussi $P_X(x)$ est la probabilité pour que l'évènement $x=1$ se réalise

A évènement décrit par $x=1$
B évènement décrit par $x=0$

Du moment que les deux évènements A et B sont ME et CE alors

P[A]+P[B]=1

$P_X(1)+P_X(0)=1$

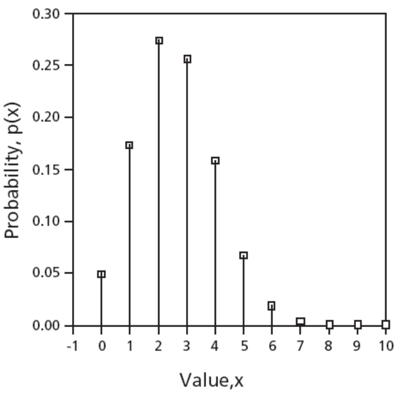
Considérons une VA X qui peut prendre plusieurs valeurs x_i, $(i=1,\dots,n)$

$P_X(x_1)+P_X(x_2)+\dots+P_X(x_n)=1$

$P_X(x_1)+P_X(x_2)+\dots+P_X(x_n)=1$

$$\sum_{i=1}^{n} P_X(x_i)=1$$

$$0 \le P_X(x_i) \le 1 \qquad \forall x_i$$



Fonction de répartition ou Cumulative Distribution Functions

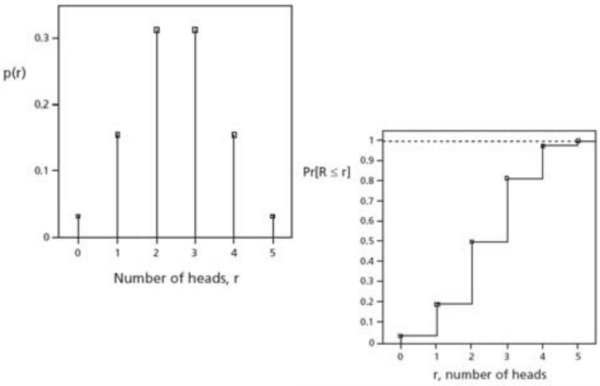
$$\Pr[X \le x] = \sum_{x_i \le x} p(x_i)$$

$$\Pr[X \le 3] = p(0) + p(1) + p(2) + p(3)$$

$$\Pr[X \le 2] = p(0) + p(1) + p(2).$$

$$p(3) = \Pr[X \le 3] - \Pr[X \le 2]$$

$$p(x_i) = \Pr[X \le x_i] - \Pr[X \le x_{i-1}]$$



Quelques exemples de loi de probabilité de VA discrète

VA de BERNOULLI

Une variable aléatoire qui prend comme valeur seulement 0 et 1 est appelé une variable indicatrice, ou une variable aléatoire de Bernoulli, ou un essai de Bernoulli

Soit X une VA discrète de type Bernoulli

$$P_X(x) = \begin{cases} p & x = 1 \\ q = 1 - p & x = 0 \end{cases}$$

Loi binomiale

Soit X la VAD qui décrit l'arrivée ou non d'un avion à l'heure

X suit une loi de type Bernoulli car on a soit succès ou échec.

On s'intéresse maintenant non pas à un seul avion mais à plusieurs (*m* et *m*>1). Il est introduit une nouvelle VAD Y qui décrit le nombre d'avions qui arrivent à l'heure (succès).

En fait c'est comme si une pièce de monnaie est jeté *m* fois. Si le résultat de l'essai est Pile c'est un succès et si le résultat de l'essai est Face c'est un échec

Ce qui est recherché c'est la loi de probabilité de la VAD Y.

$$P_Y(y)$$

Considérons uniquement 3 avions (donc 3 essais).

Quelles sont les possibilités.
Y=0, 1, 2, 3.

0 succès donc tous les avions n'arrivent pas à l'heure
1 succès donc sur les trois avions un seul va arriver à l'heure
2 succès donc sur les trois avions deux vont arriver à l'heure
3 succès donc les trois avions vont arriver à l'heure

Les évènements Y=0, Y=1, Y=2, Y=3 sont des évènements ME et CE

$$0 \le P_Y(y_i) \le 1 \qquad \sum_{i=1}^{4} P_Y(y_i) = 1$$

Du moment qu'il y a indépendance entre les trois essais alors

$$P_Y(0) = q.q.q = q^3 = (1-p)^3$$

$$P_Y(3) = p.p.p = p^3$$

$$P_Y(1) = ??$$

$$P_Y(y) = C_n^y p^y (1-p)^{n-y}$$

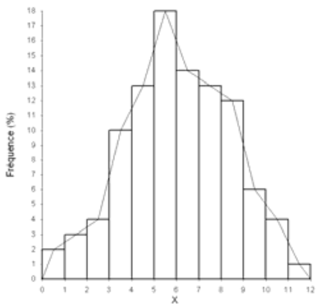
$$y = 0,1,2,\ldots,n$$

$$C_n^y = \frac{n!}{y!(n-y)!}$$

7.3 Variable aléatoire continue

k est un nombre réel

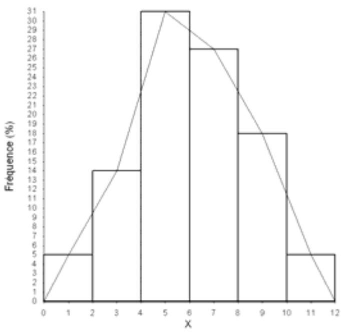
		Centre de classe	Fréquence (%)
0	1	0,5	2
1	2	1,5	3
2	3	2,5	4
3	4	3,5	10
4	5	4,5	13
5	6	5,5	18
6	7	6,5	14
7	8	7,5	13
8	9	8,5	12
9	10	9,5	6
10	11	10,5	4
11	12	11,5	1



		Centre de classe	Fréquence (%)
0	1	0,5	2
1	2	1,5	3
2	3	2,5	4
3	4	3,5	10
4	5	4,5	13
5	6	5,5	18
6	7	6,5	14
7	8	7,5	13
8	9	8,5	12
9	10	9,5	6
10	11	10,5	4
11	12	11,5	1

		Centre de classe	Fréquence (%)
0	2	1	5
2	4	3	14
4	6	5	31
6	8	7	27
8	10	9	18
10	12	11	5

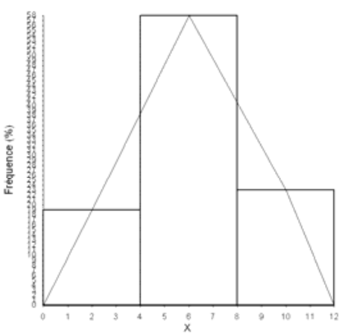
		Centre de classe	Fréquence (%)
0	2	1	5
2	4	3	14
4	6	5	31
6	8	7	27
8	10	9	18
10	12	11	5

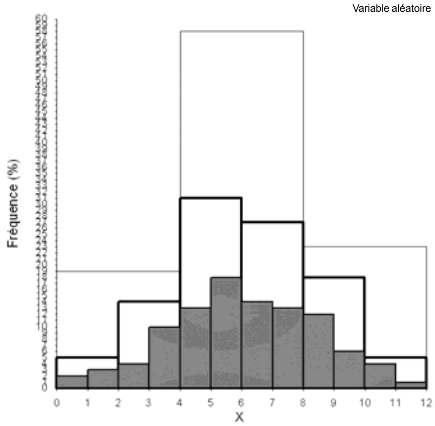


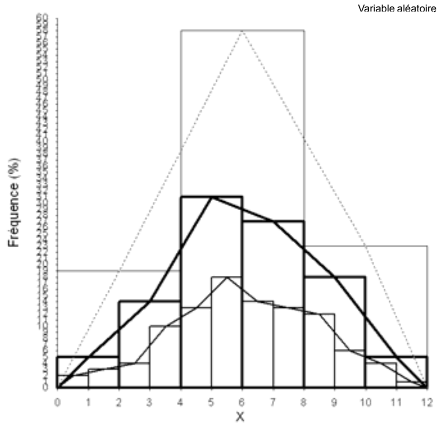
		Centre de classe	Fréquence (%)
0	2	1	5
2	4	3	14
4	6	5	31
6	8	7	27
8	10	9	18
10	12	11	5

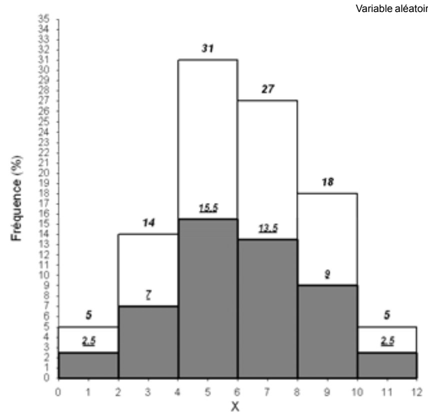
		Centre de classe	Fréquence (%)
0	4	2	19
4	8	6	58
8	12	10	23

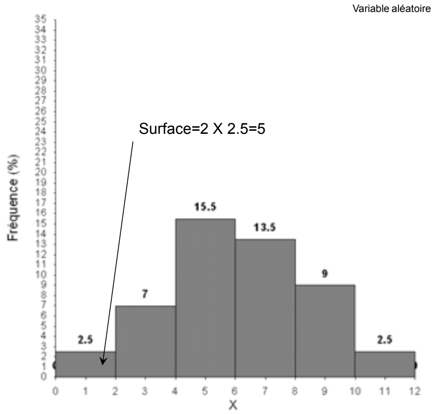
		Centre de classe	Fréquence (%)
0	4	2	19
4	8	6	58
8	12	10	23

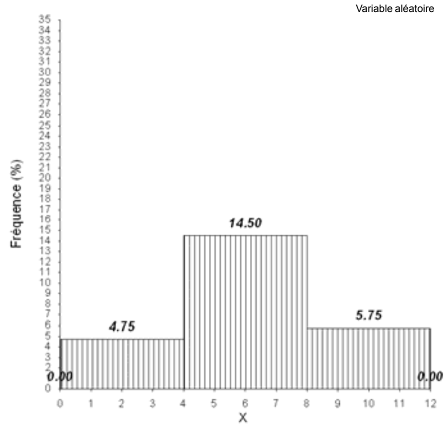


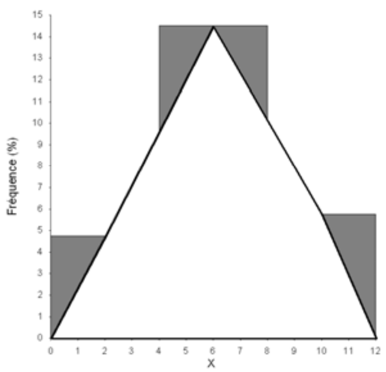






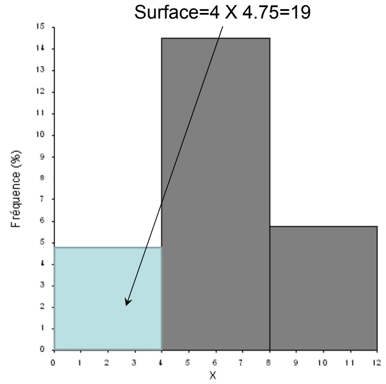






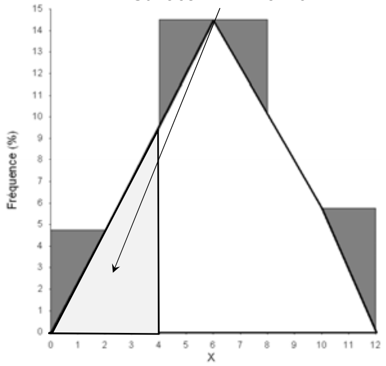
Fréquence des individus ayant des modalités comprises entre 0 et 4

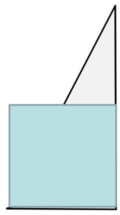
Surface=4 X 4.75=19



Fréquence des individus ayant des modalités comprises entre 0 et 4

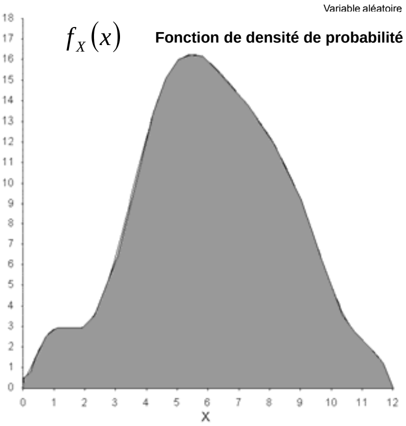
Surface=4 X 4.75=19





=

$f_X(x)$ **Fonction de densité de probabilité de la VA X**



Conditions

$$f_X(x)dx = P[x \le X \le x+dx]$$

Surface de la courbe $f_X(x)$ **entre** x **et** $x+dx$

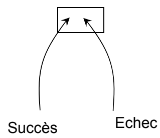
$$\int_{-\infty}^{\infty} f_X(x)\,\mathrm{d}x = 1$$

$F_X(x)$ **Fonction de distribution cumulative de la VA X**

$$\mathbf{F_X(x) = P[X \le x]} \qquad \boxed{F_X(x) = \int_{-\infty}^{x} f_X(u)\,\mathrm{d}u,}$$

$$F_X(-\infty) = 0 \qquad F_X(+\infty) = 1$$

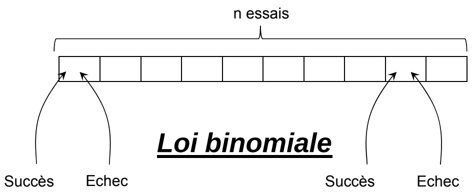
$$\mathbf{P[x_1 < X \le x_2] = F_X(x_2) - F_X(x_1)}$$



1 seul essai

<u>Loi de Bernoulli</u>

$$P_X(x) = \begin{cases} p & x = 1 \\ q = 1 - p & x = 0 \end{cases}$$



Loi binomiale

$$p_X(x) = C_n^x p^x (1-p)^{n-x} \quad x = 0,1,2,\ldots,n \quad \text{avec} \quad C_n^x = \frac{n!}{x!(n-x)!}$$

On s'intéresse à la probabilité d'avoir *X* séismes qui peuvent survenir dans un site bien précis avec des accélérations supérieures ou égale à a_{max} et ce pendant un temps *t*. Pour ce faire supposons que l'intervalle *t* est subdivisé en *n* intervalles (taille de l'intervalle Δt) de telle sorte que sur chaque intervalle $(t_1, t_1 + \Delta t)$ il ne peut y avoir qu'un séisme (succès) ou non (échec) (Sur l'intervalle $(t_1, t_1 + \Delta t)$ il ne peut y avoir plus d'un séisme). On suppose que la probabilité d'avoir un séisme est égal à *p* et que les séismes sont indépendants dans l'espace et dans le temps. Donc la probabilité d'avoir *x* séismes sur les *n* intervalles est égales à :

$$p_X(x) = C_n^x p^x (1-p)^{n-x} \quad x = 0,1,2,\ldots,n \quad \text{avec} \quad C_n^x = \frac{n!}{x!(n-x)!} \tag{3.16}$$

Pour le même cas étudié, si l'on augmente le nombre d'intervalles alors la probabilité de succès *p* au niveau de chaque intervalle diminue. En effet, si la probabilité de succès sur un intervalle $(t_1, t_1 + \Delta t)$ est de *p* alors la probabilité de succès sur l'intervalle $(t_1, t_1 + \Delta t/2)$ *p'* est forcément inférieure à *p*.

Ainsi si *n* tend vers l'infini alors la probabilité sur chaque intervalle (que l'on note *p*) tend vers 0. Toutefois le produit *np* reste constant. Ainsi

$$np = \nu \quad \text{ou} \quad p = \frac{\nu}{n} \tag{3.17}$$

$$p_X(x) = \frac{n!}{x!(n-x)!}\left(\frac{\nu}{n}\right)^x \left(1-\frac{\nu}{n}\right)^{n-x}$$

$$p_X(x) = \frac{\nu^x}{x!}\left(1-\frac{\nu}{n}\right)^n \frac{n!}{(n-x)!}\,\frac{1}{n^x(1-\nu/n)^x}$$

$$p_X(x) = \frac{\nu^x}{x!}\left(1-\frac{\nu}{n}\right)^n \left\{\frac{n(n-1)(n-2)\ldots(n-x+1)}{[n(1-\nu/n)]^x}\right\}$$

Lorsque *n* tend vers l'infini alors le terme entre accolade tend vers 1 alors que le terme $(1-\nu/n)^n$ tend vers $e^{-\nu}$. Par conséquent,

$$p_X(x) = \frac{\nu^x e^{-\nu}}{x!} \quad x = 0,1,2,\ldots,\infty$$

Cette distribution est connue sous le terme *Distribution de Poisson*,

$$p_X(x) = \frac{(\lambda t)^x e^{-\lambda t}}{x!} \quad x = 0,1,2,\ldots,\infty$$

Considérons la variable aléatoire *T* définie par

T: le temps pour lequel on a le 1^er séisme qui survient.

Donc le premier succès

La probabilité pour que $T > t$ est égale à la probabilité de ne pas avoir de séismes durant le laps de temps *t*.

De plus la probabilité pour que $T > t$ est égale à

$$1 - F_T(t)$$

La probabilité de ne pas avoir de séismes durant le laps de temps *t* est égale à

$$P_X(0)$$

$$p_X(x) = \frac{(\lambda t)^x e^{-\lambda t}}{x!} \quad x = 0,1,2,\ldots,\infty$$

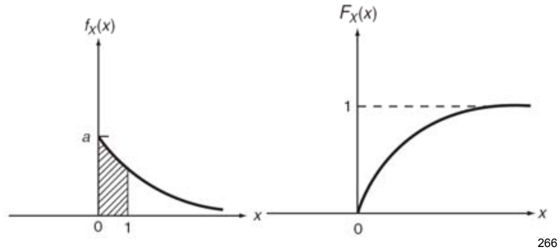
$$1 - F_T(t) = p_X(0) = \frac{(\lambda t)^0 e^{-\lambda t}}{0!}$$

$$e^{-\lambda t} = 1 - F_T(t)$$

$$F_T(t) = 1 - e^{-\lambda t}$$

$$f_T(t) = \lambda e^{-\lambda t}$$

$$f_X(x) = \begin{cases} a e^{-ax}, & \text{for } x \geq 0; \\ 0, & \text{elsewhere;} \end{cases}$$



Le nombre de météores dans un intervalle de 30 secondes est de 1.81. Supposez que l'apparition des météores est aléatoire et indépendante

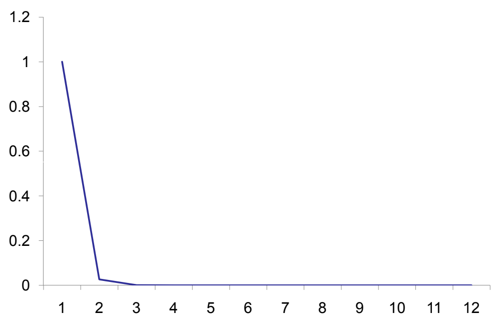
Quelle est la probabilité de ne pas avoir de météores dans un intervalle d'une minute?

$$P_X(x) = \frac{(\lambda t)^x e^{-\lambda t}}{x!} \quad x = 0,1,2,\ldots,\infty$$

$\lambda = (1.81) / (0.50 \text{ minute}) = 3.62 / \text{minute}.$

$$P_X(0) = \frac{(\lambda t)^0 e^{-\lambda t}}{0!} = e^{-\lambda t}$$

$$P_X(0) = e^{-3.62}$$



Le nombre moyen de collisions se produisant en semaine pendant les mois d'été à une intersection particulière est de 2.00. Supposez que les conditions de la distribution de Poisson sont satisfaites.

a) Quelle est la probabilité de ne pas avoir de collisions en une semaine particulière ?
b) Quelle est la probabilité qu'il y aura exactement une collision en semaine ?
c) Quelle est la probabilité d'avoir exactement deux collisions en semaine ?
d) Quelle est la probabilité de trouver pas plus de deux collisions en semaine ?

$$P_X(x) = \frac{(\lambda t)^x e^{-\lambda t}}{x!} \quad x = 0,1,2,\ldots,\infty$$

$$\lambda = 2 / semaine$$

Quelle est la probabilité de ne pas avoir de collision en une semaine particulière ?

$$P_X(0) = \frac{(\lambda t)^0 e^{-\lambda t}}{0!} \qquad P_X(0) = e^{-\lambda t}$$

$$P_X(0) = e^{-2*1}$$

$$P_X(0) = 0.135$$

Quelle est la probabilité qu'il y aura exactement une collision en semaine ?

$$P_X(1) = \frac{(\lambda t)^1 e^{-\lambda t}}{1!} \qquad P_X(1) = e^{-\lambda}\lambda$$

$$P_X(1) = e^{-2*1}2$$

$$P_X(1) = 0.271$$

Quelle est la probabilité qu'il y aura exactement deux collisions en semaine ?

$$P_X(2) = \frac{(\lambda t)^2 e^{-\lambda t}}{2!} \qquad P_X(2) = \frac{e^{-\lambda}\lambda^2}{2}$$

$$P_X(2) = 0.271$$

Quelle est la probabilité de trouver pas plus de deux collisions en semaine ?

Soit F l'évènement décrit pas: Pas plus de deux collisions en semaine

Ce qui est recherché c'est la probabilité de F

Soit A l'évènement décrit pas:	0 collision en une semaine
Soit B l'évènement décrit pas :	1 collision en une semaine
Soit C l'évènement décrit pas:	2 collisions en une semaine

Pas plus de deux collisions en semaine cela veut dire 0 ou 1 ou 2 collisions par semaine

F = AUBUC **P[F] =P[AUBUC]**

Les évènements A, B et C sont ME donc:

P[F] =P[A]+P[B]+P[C]

$$P[F] = P_X(0) + P_X(1) + P_X(2)$$

$$P[F] = 0.135 + 0.271 + 0.271$$

$$P[F] = 0.677$$

Quelle est la probabilité de trouver plus de deux collisions en semaine ?

Soit F l'évènement décrit pas: Pas plus de deux collisions en semaine

Soit F^C l'évènement décrit pas: plus de deux collisions en semaine

Les évènement F et F^C sont complémentaires

P[F^C]=1-P[F]=1-0.677=0.323

Quelle est la probabilité qu'il y aura exactement deux collisions en deux semaine ?

$$P_X(2) = \frac{(\lambda t)^2 e^{-\lambda t}}{2!} \qquad t = 2$$

$$P_X(2) = 0.147$$

7.4 Caractéristiques de tendance centrale
7.5 Caractéristiques de dispersion

VA discrète

En statistique

X Variable statistique: Intensité d'un séisme

Séisme n°	Intensité (Modalité)
1	2
2	4
3	6
4	8
5	2
6	1
7	6
8	7
9	5
10	5

Intensité (Modalité)	Effectif	Fréquence de la modalité
1	1	0.1
2	2	0.2
3	0	0
4	1	0.1
5	2	0.2
6	2	0.2
7	1	0.1
8	1	0.1
9	0	0
10	0	0
11	0	0
12	0	0

$$\overline{X} = \sum_{i}^{12} f_i X_i \qquad \sigma_X^2 = \sum_{i}^{12} f_i \left(X_i - \overline{X}\right)^2$$

X VAD Intensité d'un séisme

$$P_X(x=1) \equiv f_1$$

La moyenne est notée en statistique $\overline{X}$ $\quad \overline{X} = \sum_{i}^{12} f_i X_i$

La moyenne est notée en probabilité $E[X]$

Espérance mathématique de la VA D X

$$E[X] = \sum_{i=-\infty}^{i=\infty} P_X(x_i) x_i$$

Loi de Bernoulli

$$P_X(x) = \begin{cases} p & x = 1 \\ q = 1 - p & x = 0 \end{cases}$$

$$E[X] = \sum P_X(x_i) x_i = p*1 + q*0$$

$$E[X] = p$$

$$E[X] = 0.5 \quad \text{si} \quad p = q = 0.5$$

La variance est notée en statistique $\sigma_X^2 = \sum_{i}^{12} f_i \left(X_i - \overline{X}\right)^2$

La variance est notée en probabilité σ_X^2

Variance de VA D X

$$\sigma_X^2 = \sum_{i=-\infty}^{i=\infty} P_X(x_i)(x_i - E[X])^2$$

Loi de Bernoulli

$$P_X(x) = \begin{cases} p & x = 1 \\ q = 1 - p & x = 0 \end{cases}$$

$$E[X] = p$$

$$\sigma_X^2 = \sum P_X(x_i)(x_i - p)^2 = p(1-p)^2 + q(0-p)^2$$

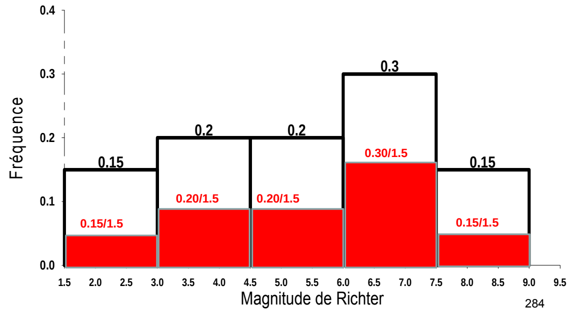
$$\sigma_X^2 = p(1-p)^2 + (1-p)p^2 = (1-p)p(1-p+p)$$

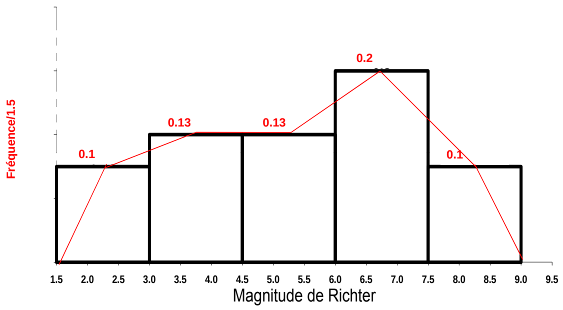
$$= (1-p)p = qp$$

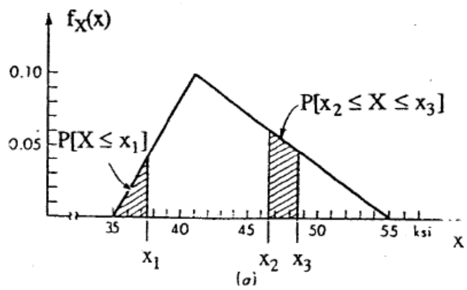
7.4 Caractéristiques de tendance centrale

7.5 Caractéristiques de dispersion

VA continue







Moyenne arithmétique dite par abréviation MOYENNE

Individus	Force (Modalité) N
1	10
2	10.3
3	22.1
4	23
5	26
6	15.2
7	12.3
8	18
9	18.3
10	14.5

$$\overline{F} = \frac{\sum_{i=1}^{10} F_i}{10} = \sum_{i=1}^{10}\left(\frac{1}{10} F_i\right)$$

$$= 16.97N$$

Limites des classes (N)		Centre de classe (N)	Effectif	Fréquence (%)	$f_k F_k$
Limite inf	Limite sup				
10	14	12	3	30	3.6
14	18	16	2	20	3.2
18	22	20	2	20	4
22	26	24	2	20	4.8
26	30	28	1	10	2.8

$$\overline{F} = \sum_{k=1}^{5} f_k F_k$$

$$= (3.6 + 3.2 + 4 + 4.8 + 2.8)$$

$$= 18.4N$$

$$\overline{F} = \sum_{j=1}^{k} f_j F_j$$

$$E[F] = \int_{-\infty}^{\infty} \left(f_F(f) df \right) f$$

$$E[F] = \int_{-\infty}^{\infty} f f_F(f) df$$

$$E[X] = \int_{-\infty}^{\infty} x f_X(x) dx$$

$$\sigma_X^2 = \int_{-\infty}^{\infty} (x - E[X])^2 f_X(x) dx$$

$$V = \sum \left[f_k \left| F_k - \overline{F} \right| \right]$$

$$EM_X = \int_{-\infty}^{\infty} |x - E[X]| f_X(x) dx$$

$$f_T(t) = \lambda e^{-\lambda t}$$

$$E[T] = \int_0^{+\infty} t f_T(t) dt = \frac{1}{\lambda}$$

$$\sigma_T^2 = \int_0^{+\infty} (t - E[T])^2 f_T(t) dt = \frac{1}{\lambda^2}$$

Example 4.2. Problem: the waiting time X (in minutes) of a customer waiting to be served at a ticket counter has the density function

$$f_X(x) = \begin{cases} 2e^{-2x}, & \text{for } x \geq 0; \\ 0, & \text{elsewhere.} \end{cases}$$

Calculer l'espérance mathématique de la VA X

$$E[X] = \int_{-\infty}^{\infty} x f_X(x) dx$$

$$E\{X\} = \int_0^{\infty} x(2e^{-2x}) dx = \frac{1}{2} \text{minute.}$$

$$E[a] = a$$

$$E[aX] = aE[X]$$

$$E[aX + b] = aE[X] + b$$

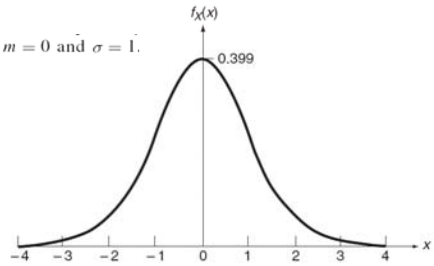
$$\sigma_X^2 = \int_{-\infty}^{\infty} (x - E[X])^2 f_X(x) dx = E\left[(X - E[X])^2 \right]$$
$$= E\left[X^2\right] - (E[X])^2$$

$$\begin{aligned} E\{g(X) + h(X)\} &= \int_{-\infty}^{\infty} [g(x) + h(x)] f_X(x) dx \\ &= \int_{-\infty}^{\infty} g(x) f_X(x) dx + \int_{-\infty}^{\infty} h(x) f_X(x) dx \\ &= E\{g(X)\} + E\{h(X)\}, \end{aligned}$$

Distribution normale ou gaussienne, $X \quad N(m,\sigma)$

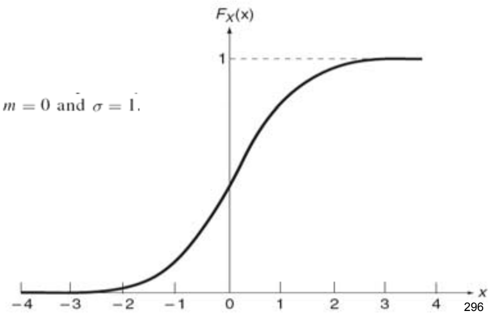
$$f_X(x) = \frac{1}{(2\pi)^{1/2}\sigma}\exp\left[-\frac{(x-m)^2}{2\sigma^2}\right], \quad -\infty < x < \infty$$

$m = 0$ and $\sigma = 1$.



$$F_X(x) = \frac{1}{(2\pi)^{1/2}\sigma}\int_{-\infty}^{x}\exp\left[-\frac{(u-m)^2}{2\sigma^2}\right]du, \quad -\infty < x < \infty,$$

$m = 0$ and $\sigma = 1$.



$$E[X] = m$$

$$E\left[(X - E[X])^2\right] = \sigma^2$$

$$U = \frac{X - m}{\sigma}$$

$$E[U] = 0$$

$$E\left[(U - E[U])^2\right] = 1$$

$X \quad N(m,\sigma) \qquad U = \dfrac{X - m}{\sigma}$

$U \quad N(0,1)$

$$f_X(x) = \frac{1}{(2\pi)^{1/2}\sigma}\exp\left[-\frac{(x-m)^2}{2\sigma^2}\right], \quad -\infty < x < \infty$$

$$f_U(u) = \frac{1}{(2\pi)^{1/2}}\exp\left[-\frac{u^2}{2\sigma^2}\right] \qquad -\infty < u < \infty$$

$$F_X(x) = \frac{1}{(2\pi)^{1/2}\sigma}\int_{-\infty}^{x}\exp\left[-\frac{(u-m)^2}{2\sigma^2}\right]du, \quad -\infty < x < \infty,$$

$$F_U(\alpha) = \frac{1}{(2\pi)^{1/2}}\int_{-\infty}^{\alpha}\exp\left[-\frac{u^2}{2\sigma^2}\right]du$$

$$F_U(-\alpha) = 1 - F_U(\alpha)$$

Fin du Chapitre 7 Variable aléatoire

7.1 Définitions
7.2 Variable aléatoire discrète
7.3 Variable aléatoire continue
7.4 Caractéristiques de tendance centrale
7.5 Caractéristiques de dispersion

Chapitre 8 Vecteurs aléatoires

8.1 Vecteurs aléatoires discrets
8.2 Vecteurs aléatoires continues

En engineering, il est souvent souhaitable de relier deux ou plusieurs ou variables aléatoires Pression, températures.

8.1 Vecteurs aléatoires discrets

JPMF: Joint Probability Mass Function
FDPC: Fonction de distribution de probabilité conjointe

$$\boxed{p_{XY}(x,y) = P(X = x \cap Y = y),}$$

$$0 < p_{XY}(x_i, y_j) \leq 1,$$

$$\sum_i \sum_j p_{XY}(x_i, y_j) = 1,$$

8.1 Vecteurs aléatoires discrets

$$\sum_i p_{XY}(x_i, y) = p_Y(y),$$

$$\sum_j p_{XY}(x, y_j) = p_X(x),$$

Fonction de distribution marginale

Fonction de distribution cumulative conjointe

$$\boxed{F_{XY}(x,y) = \sum_{i=1}^{i:x_i \leq x} \sum_{j=1}^{j:y_j \leq y} p_{XY}(x_i, y_j).}$$

8.2 Vecteurs aléatoires continues

$$f_X(x)dx = P[x \leq X \leq x + dx]$$

$$f_{XY}(x,y)dxdy = P[(x \leq X \leq x + dx) \cap (y \leq Y \leq y + dy)]$$

$$f_Y(y)dy = P[y \leq Y \leq y + dy]$$

Indépendance

$$f_X(x) f_Y(y) dxdy = P[(x \leq X \leq x + dx)] P[(y \leq Y \leq y + dy)]$$

$f_{XY}(x,y)$ (FDPC)

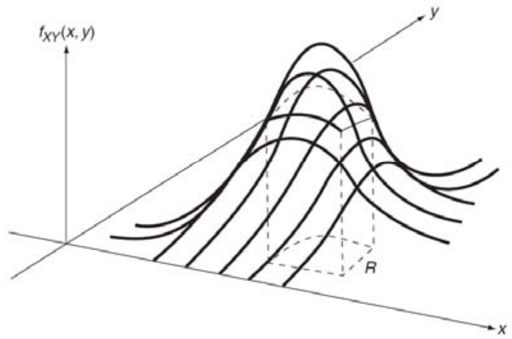
$$\boxed{F_{XY}(x,y) \equiv P(X \leq x \cap Y \leq y) = \int_{-\infty}^{y}\int_{-\infty}^{x} f_{XY}(u,v)\mathrm{d}u\mathrm{d}v.}$$

$$\boxed{f_{XY}(x,y) = \frac{\partial^2 F_{XY}(x,y)}{\partial x \partial y}.}$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\mathrm{d}x\mathrm{d}y = 1,$$

$$\int_{-\infty}^{\infty} f_{XY}(x,y)\mathrm{d}y = f_X(x),$$

$$\int_{-\infty}^{\infty} f_{XY}(x,y)\mathrm{d}x = f_Y(y).$$

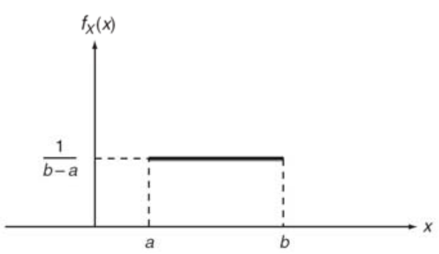


Deux personnes pensent se rencontrer entre 9h00 et 10h00. Chaque personne ne peut attendre plus de 10 minutes. C'est-à-dire qu'au-delà de 10 minutes elle partira si la deuxième personne ne vient pas. On suppose les deux personnes peuvent arriver à n'importe quel moment entre 9h00 et 10h00 et que leurs temps d'arrivée ne sont pas connus.

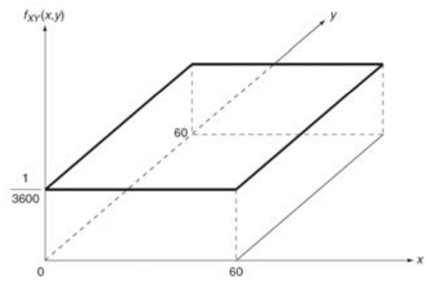
X décrite par : le nombre de minutes écoulé jusqu'à l'arrivée de la personne A

Y décrite par : le nombre de minutes écoulé jusqu'à l'arrivée de la personne B

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & a \leq x \leq b \\ 0 & elsewhere \end{cases}$$



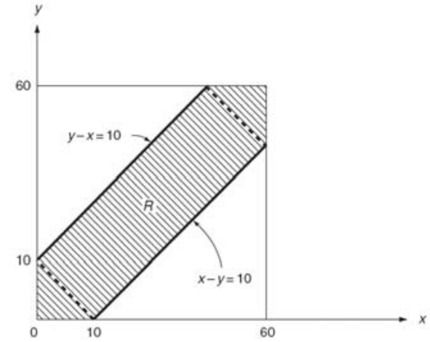
$$f_{XY}(x,y) = \begin{cases} \dfrac{1}{3600}, & \text{for } 0 \leq x \leq 60, \text{and } 0 \leq y \leq 60; \\ 0, & \text{elsewhere.} \end{cases}$$



Question :

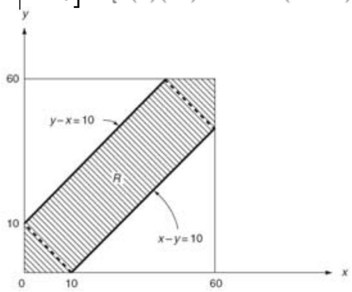
Il est recherché la probabilité pour que les deux personnes se rencontrent entre 9h00 et 10h00.

$$R : |X - Y| \leq 10$$



Probabilité que les deux personnes se rencontrent=

$$P[|X - Y| \leq 10] = [2(5)(10) + 10\sqrt{2}(50\sqrt{2})]/3600 = \frac{11}{36}$$



Exemple: Fiabilité d'un système

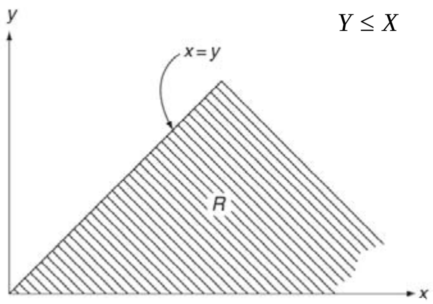
Y: Résistance VA

X: Action VA

$p_r = P[Y \leq X]$ Probabilité de rupture

$$f_{XY}(x, y) = \begin{cases} ab\mathrm{e}^{-(ax+by)}, & \text{for } (x, y) > 0 \\ 0, & \text{for } (x, y) \leq 0; \end{cases}$$

$Y \leq X$



$$p_f = \int_0^\infty \int_y^\infty ab\mathrm{e}^{-(ax+by)}\mathrm{d}x\mathrm{d}y = \frac{b}{a+b}$$

Distribution conditionnelle

$$\boxed{F_{XY}(x|y) = P(X \leq x | Y = y).}$$

$$\boxed{p_{XY}(x|y) = P(X = x | Y = y).}$$

$$p_{XY}(x|y) = P(X = x | Y = y) = \frac{P(X = x \cap Y = y)}{P(Y = y)},$$

$$p_{XY}(x|y) = \frac{p_{XY}(x, y)}{p_Y(y)}, \text{ if } p_Y(y) \neq 0,$$

$$\boxed{f_{XY}(x|y) = \frac{\mathrm{d}F_{XY}(x|y)}{\mathrm{d}x},}$$

$$P(x_1 < X \le x_2 | y_1 < Y \le y_2) = \frac{P(x_1 < X \le x_2 \cap y_1 < Y \le y_2)}{P(y_1 < Y \le y_2)}.$$

$$P(x_1 < X \le x_2 | y_1 < Y \le y_2) = \int_{y_1}^{y_2}\int_{x_1}^{x_2} f_{XY}(x,y)\,\mathrm{d}x\,\mathrm{d}y \Big/ \int_{y_1}^{y_2}\int_{-\infty}^{\infty} f_{XY}(x,y)\,\mathrm{d}x\,\mathrm{d}y$$
$$= \int_{y_1}^{y_2}\int_{x_1}^{x_2} f_{XY}(x,y)\,\mathrm{d}x\,\mathrm{d}y \Big/ \int_{y_1}^{y_2} f_Y(y)\,\mathrm{d}y. \quad (3.40)$$

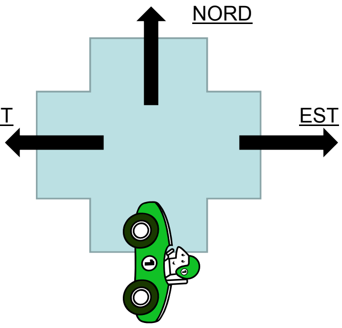
$x_1 = -\infty, x_2 = x, y_1 = y$, and $y_2 = y + \Delta y,\ \Delta y \to 0$,

$$F_{XY}(x|y) = \frac{\int_{-\infty}^{x} f_{XY}(u,y)\,\mathrm{d}u}{f_Y(y)},$$

$$f_Y(y) \neq 0.$$

$$f_{XY}(x|y) = \frac{\mathrm{d}F_{XY}(x|y)}{\mathrm{d}x} = \frac{f_{XY}(x,y)}{f_Y(y)}, \quad f_Y(y) \neq 0$$

n voitures qui arrivent au niveau d'une intersection.



X VA : Nombre de voitures qui tournent vers l'Est
Y VA : Nombre de voitures qui tournent vers l'Ouest

$$p_{XY}(x,y)$$

$$p_{XY}(x|y) = \frac{p_{XY}(x,y)}{p_Y(y)}, \text{ if } p_Y(y) \neq 0,$$

$$p_{XY}(x,y) = p_{XY}(x|y)p_Y(y)$$

Probabilité[Voiture tourne vers l'Est]=p X

Probabilité[Voiture tourne vers l'Ouest]=q Y

Probabilité[Voiture tourne vers le Nord]=r=1-p-q

Loi binomiale: Tournez vers l'Ouest ou non

$$p_Y(y) = \binom{n}{y} q^y (1-q)^{n-y}, \quad y = 1, 2, \ldots .$$

$$p_{XY}(x|y)$$

Sur les *n* voitures, on sait que déjà *y* voitures ont pris la direction de l'Ouest (à gauche). Donc il reste uniquement (n-y) qui doivent soit tourner à gauche (Est) soit se diriger vers le Nord.

$$p_{XY}(x|y) = \binom{n-y}{x}\left(\frac{p}{r+p}\right)^x\left(1 - \frac{p}{r+p}\right)^{n-y-x},$$

$$x = 0, 1, \ldots, n-y,\ y = 0, 1, \ldots, n.$$

Fin du Chapitre 8 Vecteurs aléatoires

8.1 Vecteurs aléatoires discrets
8.2 Vecteurs aléatoires continues

Chapitre 9 Transformation d'une Variable Aléatoire

9.1 Objectifs
9.2 Méthode générale
9.3 Application pour le cas d'une fonction linéaire
9.4 Fonctions à deux Variables aléatoires

9.1 Objectifs

Soit C le coût de construction d'un mur

C1: coût des matériaux (DA)

C2: coût de la Main d'œuvre (DA/Heure)

C=C1+ C2*H

H: Nombre d'heures de travail

C=10.000+100*100

C=20.000 DA

H VA notée X

C VA notée Y

$$Y = a + bX$$

$f_X(x)$ **Fonction de densité de probabilité de la VA X**

$$f_X(x)dx = P[x \le X \le x + dx]$$

$f_Y(y)$ **Fonction de densité de probabilité de la VA X**

$$f_Y(y)dy = P[y \le Y \le y + dy]$$

Si la loi de probabilité de la VA X est <u>connue</u> alors comment déterminer la loi de probabilité de Y

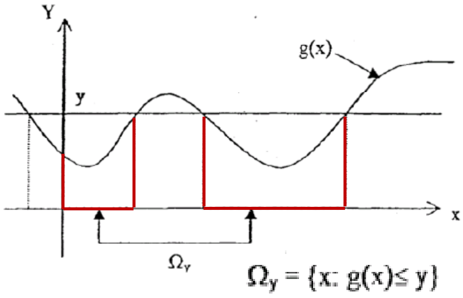
$$f_X(x) \qquad\qquad f_Y(y)$$

$$Y = a + bX$$

$$Y = g(X)$$

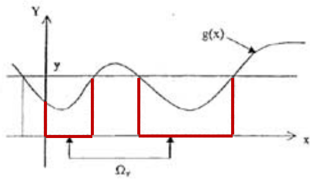
9.2 Méthode générale

$$F_Y(y) = P[Y \le y]$$



$$\Omega_y = \{x: g(x) \le y\}$$

$$F_Y(y) = P[Y \le y]$$



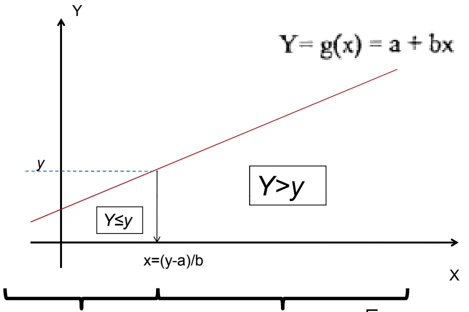
$$F_Y(y) = P[Y \le y] = P[x \in \Omega_y] = \int_{\Omega_y} f_X(x)dx$$

9.3 *Application pour le cas d'une fonction linéaire*

$$Y = g(x) = a + bx$$

If $b > 0$:

$$X(y) = \frac{y-a}{b};\ \Omega_y = \{x : a + bx \le y\} = \left(-\infty, \frac{y-a}{b}\right]$$



$$F_Y(y) = P[Y \le y] \Longrightarrow F_Y(y) = P\left[X \le \frac{y-a}{b}\right]$$

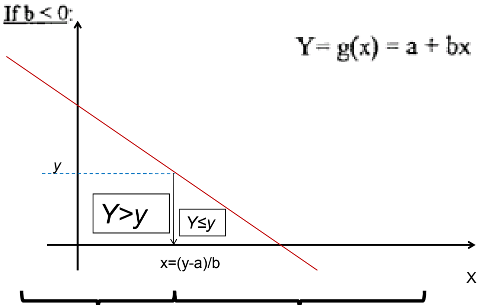
$$F_X(x) = P[X \le x]$$

$$F_Y(y) = P\left[X \le \frac{y-a}{b}\right] = F_X\left(\frac{y-a}{b}\right)$$

$$f_Y(y) = \frac{dF_Y(y)}{dy}$$

$$f_Y(y) = \frac{1}{b} f_X\left(\frac{y-a}{b}\right)$$

If $b < 0$:



$$F_Y(y) = P[Y \le y] \Longrightarrow F_Y(y) = P\left[X > \frac{y-a}{b}\right]$$

$$F_Y(y) = P\left[X > \frac{y-a}{b}\right] = 1 - P\left[X \le \frac{y-a}{b}\right]$$

$$F_Y(y) = 1 - P\left[X \le \frac{y-a}{b}\right] = 1 - F_X\left(\frac{y-a}{b}\right)$$

$$f_Y(y) = \frac{dF_Y(y)}{dy}$$

$$f_Y(y) = -\frac{1}{b} f_X\left(\frac{y-a}{b}\right)$$

$$Y = g(x) = a + bx$$

If b > 0:

$$f_Y(y) = \frac{1}{b}\, f_X\left(\frac{y-a}{b}\right)$$

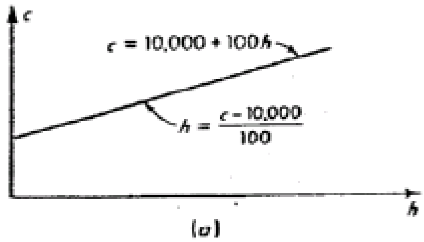
If b < 0:

$$f_Y(y) = -\frac{1}{b}\, f_X\left(\frac{y-a}{b}\right)$$

$$f_Y(y) = \frac{1}{|b|}\, f_X\left(\frac{y-a}{b}\right)$$

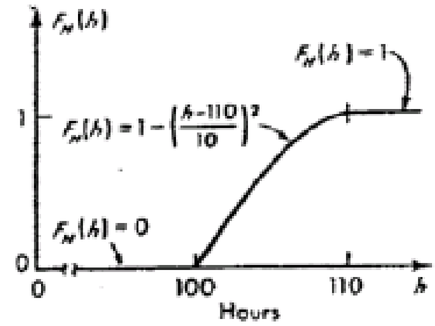
337

Chapitre 9: Transformation d'une Variable



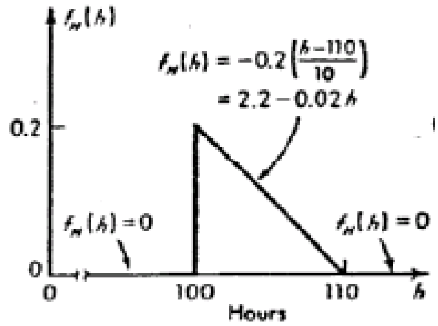
338

Chapitre 9: Transformation d'une Variable



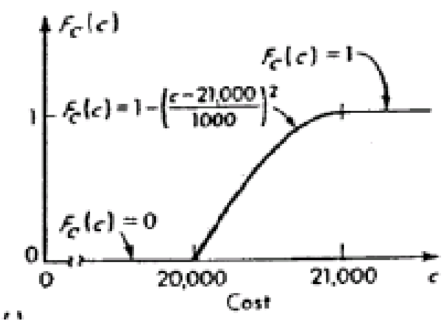
339

Chapitre 9: Transformation d'une Variable



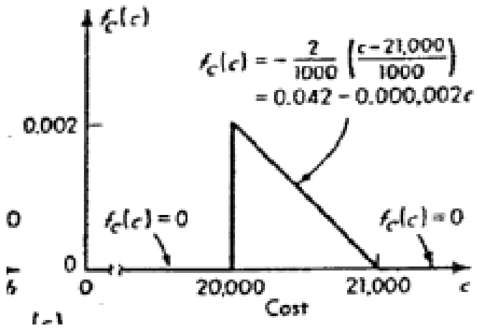
340

Chapitre 9: Transformation d'une Variable



341

Chapitre 9: Transformation d'une Variable



342

9.4 Fonctions à deux Variables aléatoires

Soit X_1 et X_2 deux variables aléatoires définies soit pat leurs PDF ou leur JPDF

$$f_{X_1}(x_1) \qquad f_{X_2}(x_2)$$

$$f_{X_1X_2}(x_1, x_2)$$

$$Y = X_1 + X_2$$

$$f_Y(y) \qquad\qquad F_Y(y)$$

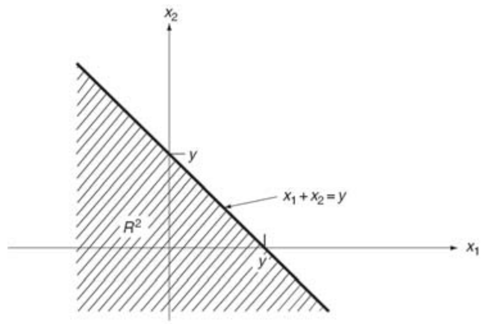


Figure 5.20 Region $R^2: x_1 + x_2 \le y$

$$F_Y(y) = \iint_{(R^2 : x_1 + x_2 \le y)} f_{X_1X_2}(x_1, x_2)\mathrm{d}x_1\mathrm{d}x_2$$

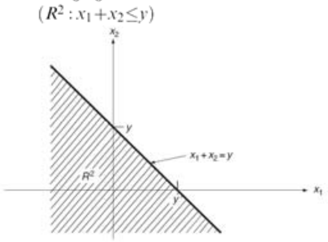


Figure 5.20 Region $R^2: x_1 + x_2 \le y$

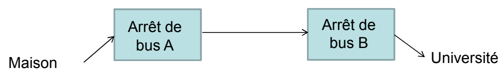
$$F_Y(y) = \int_{-\infty}^{\infty}\int_{-\infty}^{y-x_2} f_{X_1X_2}(x_1, x_2)\mathrm{d}x_1\mathrm{d}x_2$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X_1X_2}(y - x_2, x_2)\mathrm{d}x_2.$$

Si X_1 et X_2 sont indépendant

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X_1}(y - x_2) f_{X_2}(x_2)\mathrm{d}x_2.$$

$$\boxed{f_Y(y) = \int_{-\infty}^{\infty} f_{X_1}(y - x_2) f_{X_2}(x_2)\mathrm{d}x_2 = \int_{-\infty}^{\infty} f_{X_2}(y - x_1) f_{X_1}(x_1)\mathrm{d}x_1.}$$



X: VA temps d'attente au niveau de l'arrêt de bus A

$$f_X(x) = \lambda e^{-\lambda x}$$

Y: VA temps d'attente au niveau de l'arrêt de bus B

$$f_Y(y) = \beta e^{-\beta y}$$

$$T = X + Y$$

$$F_T(t) = ??????????$$

$$Y = X_1 + X_2$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X_1}(y - x_2) f_{X_2}(x_2)\mathrm{d}x_2.$$

$$T = X + Y$$

$$f_T(t) = \int_{-\infty}^{\infty} f_X(t-y) f_Y(y) dy$$

$$f_T(t) = \int_{-\infty}^{\infty} f_X(t-y) f_Y(y) dy$$

$$f_X(x) = \lambda e^{-\lambda x} \qquad f_Y(y) = \beta e^{-\beta y}$$

$$x \geq 0 \qquad y \geq 0$$

$$f_X(t-y) = \lambda e^{-\lambda(t-y)} \qquad t - y \geq 0$$

$$f_T(t) = \int_{-\infty}^{\infty} \lambda e^{-\lambda(t-y)} \beta e^{-\beta y} dy$$

$$y \geq 0 \qquad t - y \geq 0$$

$$0 \leq y \leq t$$

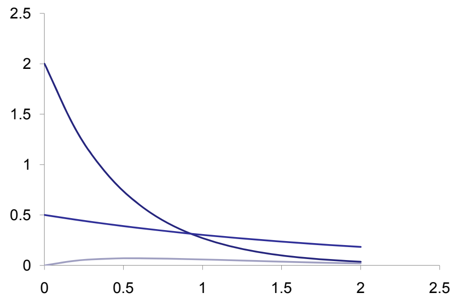
$$f_T(t) = \int_{0}^{t} \lambda e^{-\lambda(t-y)} \beta e^{-\beta y} dy$$

$$f_T(t) = \frac{\lambda\beta}{\lambda - \beta}\left(e^{-\beta t} - e^{-\lambda t}\right)$$

$$t \geq 0$$

$$f_X(x) = \lambda e^{-\lambda x} \qquad f_Y(y) = \beta e^{-\beta y}$$

$$f_T(t) = \frac{\lambda\beta}{\lambda - \beta}\left(e^{-\beta t} - e^{-\lambda t}\right)$$



Fin Chapitre 9 Transformation d'une Variable Aléatoire

9.1 Objectifs
9.2 Méthode générale
9.3 Application pour le cas d'une fonction linéaire
9.4 Fonctions à deux Variables aléatoires

Fin du cours