Chapitre 1 Notions de probabilités et de statistiques

Introduction

La statistique est la discipline qui étudie des phénomènes à travers la collecte de données, leur traitement, leur analyse, l'interprétation des résultats et leur présentation afin de rendre ces données compréhensibles par tous. C'est à la fois une science, une méthode et un ensemble de techniques.

Dans la pratique, les méthodes et outils statistiques sont utilisés dans des domaines tels que :

- Géophysique, pour les prévisions météorologiques, la climatologie, la pollution, les études des rivières et des océans ;
- Démographie : le recensement permet de faire une photographie à un instant donné d'une population et permettra par la suite des sondages dans des échantillons représentatifs.
- Sciences économiques et sociales, et en économétrie : l'étude du comportement d'un groupe de population ou d'un secteur économique s'appuie sur des statistiques. C'est dans cette direction que travaille l'Insee. Les questions environnementales s'appuient également sur des données statistiques ;
- Ecologie, pour l'étude des communautés végétales et des écosystèmes.

Les statistiques permettent d'exploiter les informations recueillies pour établir toute relation de causalité par l'interprétation et l'analyse.

Un phénomène aléatoire est un phénomène comporte des **variables aléatoire** c'est-à-dire des variables liées au hasard et dont les valeurs ne peuvent en conséquence, être connus en avance.

Les statistiques sont appliquées dans presque tous les domaines de l'activité scientifique. Lorsque l'analyse des données relatives à un groupe **d'individus** ou d'objectifs, par exemple les hauteurs d'eau, les débits dans un cours d'eau, intensités des pluies…, il est souvent impossible ou pas pratique d'examiner tous les éléments du groupe appelé **population** ; on examine alors une petite partie du groupe appelée **échantillon**.

Un échantillon peut être **représentatif** suivant une, deux, trois variables ou plus mais jamais totalement identique à la population totale. La représentativité d'un échantillon n'est toujours que partiellement vérifiable. C'est une notion relative.

1. Analyse statistique

Les caractéristiques d'une série statistique permettent de caractériser cette série en mettant en évidence des informations dont la valeur donne une indication importante sur cette série étudie.

Une série statistique est une suite de données individuelles par exemple, disposons de la série suivante de précipitations annuelles d'une station météorologiques, **P** en (mm) :

An	P(mm)	**An**	P(mm)	**An**	P(mm)	**An**	P(mm)	**An**	P(mm)
1996	794	**2001**	655	**2006**	804	**2011**	645	**2016**	549
1997	720	**2002**	720	**2007**	708	**2012**	645	**2017**	671
1998	652	**2003**	420	**2008**	755	**2013**	591	**2018**	800
1999	674	**2004**	645	**2009**	655	**2014**	698	**2019**	795
2000	804	**2005**	695	**2010**	549	**2015**	708	**2020**	704

Tableau 1 : Précipitation moyenne annuelle d'un bassin versant

En général ces données brutes ne sont pas organisées .pour pouvoir analyser une telle série et mettre en reliefs ses caractéristiques essentielles, l'on procède comme suit :

1) Ordre de série

On peut ranger les valeurs étudiées soit par l'ordre croissant soit dans l'ordre décroissant. La différence entre la plus grande valeur et la plus petite valeur appelé l'**amplitude** de la série.

Une valeur n'est inscrite qu'une seule fois et, en face, on indique le nombre de fois où l'on a observé cette valeur. Ce nombre est l'**effectif** de la valeur ou sa **fréquence absolue** (n_i) ; ainsi la fréquence absolue de précipitation 7**94** mm est **1**.

On peut également indiquer pour cette valeur la fréquence relative (fi) qui est le rapport entre la fréquence de la valeur et le total des fréquences absolues $N = \sum n_i = 25$ de la série, ainsi la fréquence relative de précipitation **794** est $\frac{1}{25} = 0.04$, le tableau N°2 récapitule tous les résultats de calcul.

Pi	Effectif ou fréquence absolue (ni)	Fréquence relative (fi)	Pi	Effectif ou fréquence absolue (ni)	Fréquence relative (fi)
420	1	0.04	698	1	0.04
547	2	0.08	704	1	0.04
591	1	0.04	708	2	0.08
645	3	0.12	720	2	0.08
652	1	0.04	755	1	0.04
655	2	0.08	794	1	0.04
671	1	0.04	795	1	0.04
674	1	0.04	800	1	0.04
695	1	0.04	804	2	0.08

Tableau 2 : Rangement des valeurs de précipitations

Pour mettre en reliefs les caractéristiques de la série étudiée on procède des groupements en classes de valeurs.

Lorsqu'on veut résumer une grande quantité de données brute il est commode de les distribuer en classes ou intervalles et de déterminer le nombre d'individus appartenant à chaque calsse, que l'on appelle fréquence ou effectif de la classe.

2. **Nombre d'intervalle**

Dans le cas d'un caractère quantitatif continu, l'établissement du tableau de fréquences implique d'effectuer au préalable une répartition en classes des données. Cela nécessite de définir le nombre de classes attendu et donc l'amplitude associée à chaque classe ou intervalle de classe. En règle générale, on choisit des classes de même **amplitude**.

Pour que la distribution en fréquence est un sens, il faut que chaque classe comprenne un nombre suffisant de valeurs **(n_i).**

Diverses formules empiriques permettent d'établir le nombre de classes pour un échantillon de taille ***N (notre cas N=25 valeurs)***

La règle de **STURGE** : Nombre de classes **K = 1+ (3,3 log *N*)**

La règle de **YULE :** Nombre de classes **K =** $2,5\sqrt[4]{N}$

3. **L'intervalle** entre chaque classe est obtenu ensuite de la manière suivante :

Intervalle de classe (IC)= (X max - X min) / Nombre de classes

avec X max et X min, respectivement la plus grande et la plus petite valeur de X dans la série statistique.

Pour notre cas :

Règle de Sturge : 1 + (3,3 log 25) = 5.61

Règle de Yule : $2,5\sqrt[4]{25}$ = 5.59

Les deux valeurs sont très peu différentes, on prend **6 intervalles**

4. Définition de l'intervalle de classe :

$$IC = \frac{804 - 420}{6} = 64 \text{ mm}$$

A partir de *X*min on obtient les limites de classes ou bornes de classes par addition successive de l'intervalle de classe

Pour la première classe de notre cas :

Limite inferieure = 420-0.5 = **419.5mm**

Limite supérieure = 419.5+64 = **483.5mm**

Les tableaux N°03 résume toutes ces opérations

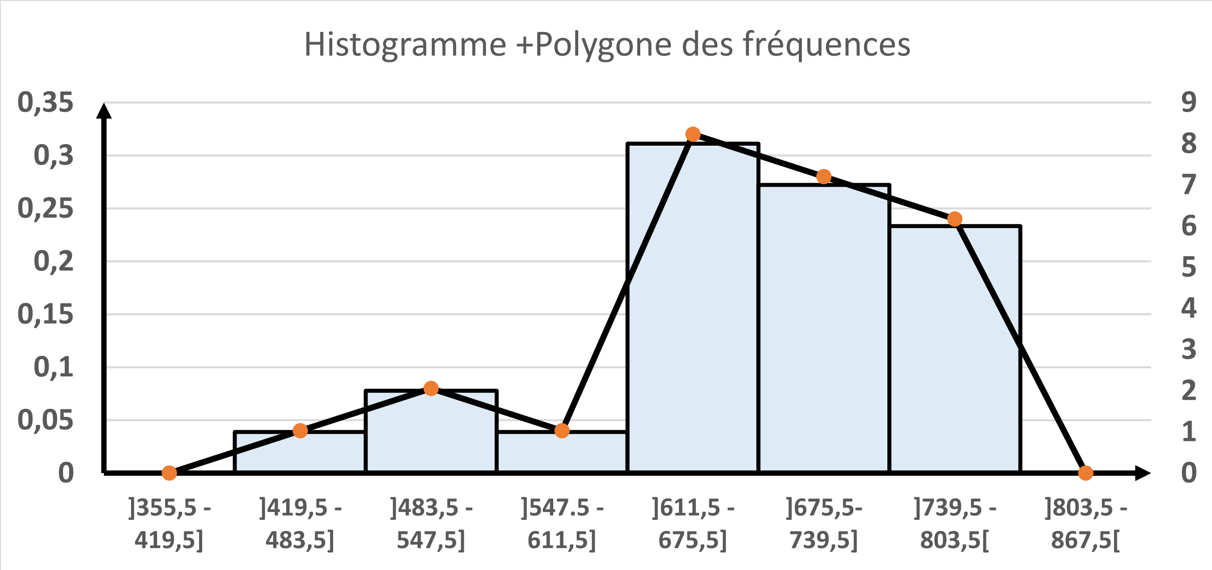
Numéros de classes	**Bornes des classes**	**Centre des classes**	**Effectif ou fréquence absolue (ni)**	**Fréquence relative (fi=ni/N)**
1	$419.5 \leq Pi < 483.5$	451.5	1	0.04
2	$483.5 \leq Pi < 547.5$	515.5	2	0.08
3	$547.5 \leq Pi < 611.5$	579.5	1	0.04
4	$611.5 \leq Pi < 675.5$	643.5	8	0.32
5	$675.5 \leq Pi < 739.5$	707.5	7	0.28
6	$739.5 \leq Pi < 803.5$	771.5	6	0.24

Tableau 3 : Groupement des valeurs des précipitations

5. Représentations graphiques

Les représentations graphiques ont l'avantage de renseigner immédiatement sur l'allure générale de la distribution. Elles facilitent l'interprétation des données recueillies.

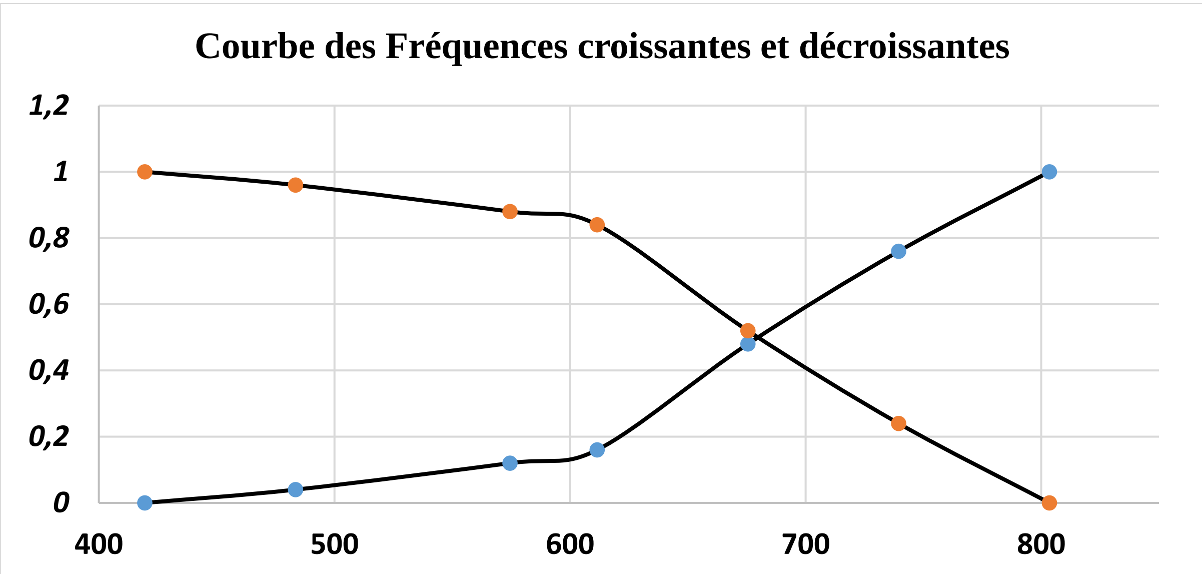
Pour les caractères quantitatifs continus, la représentation graphique est ***l'histogramme*** où la hauteur du rectangle est proportionnelle à l'effectif n_i. Ceci n'est vrai que si l'intervalle de classe est constant. Dans ce cas l'aire comprise sous l'histogramme s'avère proportionnelle à l'effectif total.



Courbe des fréquences cumulées croissantes et décroissantes

Précipitations	**Effectif cumulé**	**fréq cumulée croissant**	**Précipitations**	**Effectif cumulé**	**fréq cumulée décroissant**
<419.5	0	0	>419.5	25	1
<483.5	1	0.04	>483.5	24	0.96
<547.5	3	0.12	>547.5	22	0.88
<611.5	4	0.16	>611.5	21	0.84
<675.5	12	0.48	>675.5	13	0.52
<739.5	19	0.76	>739.5	6	0.24
<803.5	25	1	>803.5	0	0

Tableau 4 : Calcul des fréquences cumulées



Dans le tableau N°4 on a calculé les fréquences cumulées jusqu'aux bornes des intervalles

La somme des fréquences de toutes les valeurs plus petites que la limite supérieurs d'un intervalle est appelée fréquence cumulée au non dépassement **(FND)** : ainsi **48 %** des précipitations annuelles considérées sont inférieures à **675.5 mm**

En outre, la somme des fréquences de toutes les valeurs plus grandes que la limite inférieure d'un intervalle est appelée fréquences cumulées au dépassement **(FD)** : ainsi **52 %** des précipitations annuelles de notre échantillon sont supérieure à **675.5 mm**

On constate que : **FND+FD = 48% + 52% = 100**

A) Les paramètres centrales

Il existe trois valeurs centrales :

- le mode
- la médiane
- la moyenne

1. Le mode ou **valeur dominante** : est la valeur la plus fréquente d'une distribution, dans notre cas, cette valeur est 645 mm.

2. La médiane

On ne peut la calculer que pour les caractères quantitatifs. Les valeurs étant classées par ordre croissant, la médiane est la valeur du caractère qui partage celui-ci en deux ensembles d'effectifs égaux : 50 % des valeurs lui sont supérieures et 50 % lui sont inférieures.

$$Médiane = L_1 + \left(\frac{\frac{N}{2} - (\sum f_i)}{f_{médiane}} \right) * C$$

Où :

Classe médiane est celle qui correspond à N/2 =12.5

L_1 : limite inférieure de la classe médiane = 675.5 mm

$\sum f_i$: Somme des fréquences absolues de toutes les classes inférieure à la classe médiane : 12

$f_{médiane}$: Fréquence de la classe médiane = 7

C : grandeur de la classe médiane = 64

Dans notre cas la **médiane égale à 680.07 mm**

6. La moyenne arithmétique

$$\overline{X} = \frac{\sum P_i}{N}$$

Dans notre cas la moyenne arithmétique égale à $\overline{X}$ =682.08 mm

B) Les paramètres de dispersion

Le résumé d'une distribution que donne une valeur centrale ne nous renseigne pas sur la dispersion des valeurs autour de cette valeur centrale, c'est-à-dire sur la tendance qu'elles-ont à se concentrer ou se disperser autour de celle-ci.

1. Ecart-type

$$\sigma_{Pi} = \sqrt{Variance} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(P_i - \bar{P}\right)^2} \text{ = 90.92 mm}$$

2. Variance

$$Variance = \frac{1}{N}\sum_{i=1}^{N}\left(P_i - \bar{P}\right)^2 \text{ = 8312.82 mm}$$

3. Le Coefficient de variation

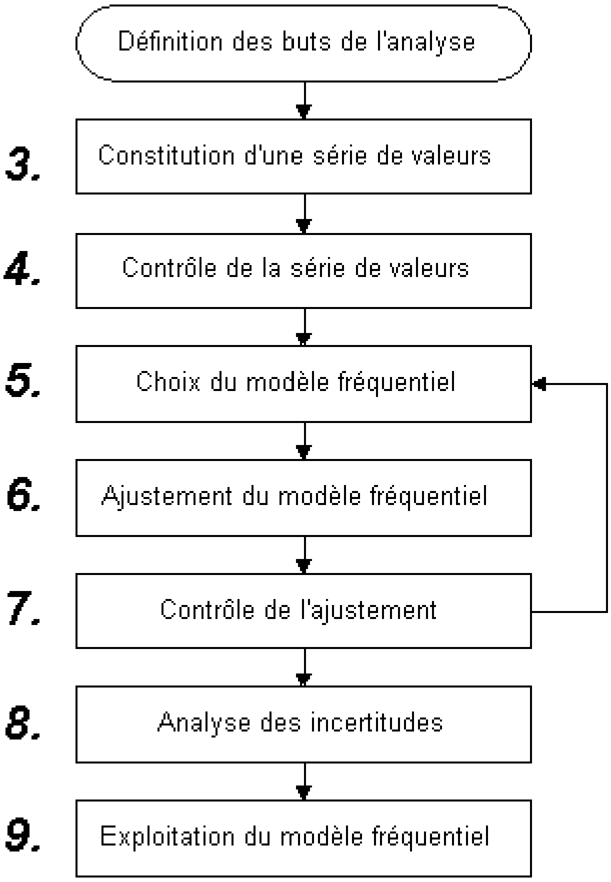
$$C_v = \frac{\sigma_{Pi}}{\bar{P}} \text{ = 0.133 mm}$$

2. **Analyse fréquentielle**

L'analyse fréquentielle est une méthode statistique de prédiction consistant à étudier les événements passés, caractéristiques d'un processus donné (hydrologique ou autre), afin d'en définir les probabilités d'apparition future.

Cette prédiction repose sur la définition et la mise en œuvre d'un modèle fréquentiel, qui est une équation décrivant le comportement statistique d'un processus. Ces modèles décrivent la probabilité d'apparition d'un événement de valeur donnée.

L'analyse fréquentielle fait appel à diverses techniques statistiques et constitue une filière complexe qu'il convient de traiter avec beaucoup de rigueur. Ses diverses étapes peuvent être schématisées très simplement selon le diagramme suivant :



Principales étapes de l'analyse fréquentielle.

2 Choix du modèle fréquentiel

La validité des résultats d'une analyse fréquentielle dépend du choix du modèle fréquentiel et plus particulièrement de son type. Diverses pistes peuvent contribuer à faciliter ce choix, mais il n'existe malheureusement pas de méthode universelle et infaillible

1 Loi normale

La loi normale se justifie, théoriquement par le théorème central-limite, comme la loi d'une variable aléatoire formée de la somme d'un grand nombre de variables aléatoires. En hydrologie fréquentielle des valeurs extrêmes, les distributions ne sont cependant pas symétriques, ce qui constitue un obstacle à son utilisation. Cette loi s'applique toutefois généralement bien à l'étude des modules annuels des variables hydrométéorologiques en climat tempéré

2 Loi log-normale

La loi log-normale est préconisée par certains hydrologues dont V.-T. Chow qui la justifie en argumentant que l'apparition d'un événement hydrologique résulte de l'action combinée d'un grand nombre de facteurs qui se multiplient. Dès lors la variable aléatoire $X = X_1 \cdot X_2 \cdot \cdots \cdot X_r$ suit une loi log-normale. En effet le produit de r variables se ramène à la somme de r

logarithmes de celles-ci et le théorème central-limite permet d'affirmer la log-normalité de la variable aléatoire.

3 Loi de Gumbel

E.-J. Gumbel postule que la loi double exponentielle, ou loi de Gumbel, est la forme limite de la distribution de la valeur maximale d'un échantillon de n valeurs. Le maximum annuel d'une variable étant considéré comme le maximum de 365 valeurs journalières, cette loi doit ainsi être capable de décrire les séries de maxima annuels.

Il est à remarquer que plus le nombre de paramètres d'une loi est grand, plus l'incertitude dans l'estimation est importante. Pratiquement il est par conséquent préférable d'éviter l'utilisation de lois à trois paramètres ou plus.

Ajustement du modèle fréquentiel

Dans notre partie ce chapitre nous étudierons les techniques de l'ajustement ou du calage d'un modèle fréquentiel à une série de données : il s'agit de définir les paramètres de la loi retenue. Nous utiliserons comme support pédagogique la loi de **Gumbel,** fréquemment utilisée en hydrologie, pour modéliser les événements extrêmes, les pluies notamment.

1 Présentation de la loi de Gumbel

La fonction de répartition de la loi de Gumbel s'exprime de la manière suivante :

$$F(x) = \exp\left(-\exp\left(-\frac{x-a}{b}\right)\right) \quad . \qquad (1)$$

Où a est le paramètre de position, b le paramètre d'échelle

Posons la variable réduite suivante $u = \dfrac{x-a}{b}$ (2)

. La distribution s'écrit alors comme suit : $F(x) = \exp\left(-\exp(-u)\right)$ (3)

et $u = -\ln\left(-\ln\left(F(x)\right)\right)$ (4)

L'avantage d'utiliser la variable réduite est que l'expression d'un quantile est alors linéaire. En effet pour trouver la valeur x_q d'un quantile, correspondant à la distribution $F(x_q) = q$, en fonction des deux paramètres a et b, il suffit d'utiliser la relation suivante : $x_q = a + bu_q$

En conséquence, dès lors que les points de la série à ajuster peuvent être reportés dans un système d'axes x - u il est possible d'ajuster une droite qui passe le mieux par ces points et d'en déduire les deux paramètres a et b de la loi. Il existe différentes méthodes d'ajustement :

méthode graphique (ajustement à l'œil ou à l'aide d'une régression statistique), méthode des moments ect.

En pratique, il s'agit essentiellement d'estimer la probabilité de non dépassement F(xi) qu'il convient d'attribuer à chaque valeur xi. Il existe de nombreuses formules d'estimation de la fonction de répartition à l'aide de la fréquence empirique. Elles reposent toutes sur un tri de la série par valeurs croissantes permettant d'associer à chaque valeur son rang r. Des simulations ont montré que pour la loi de Gumbel, il faut utiliser la fréquence empirique de Hazen :

$$\frac{r-0.5}{n} \quad (5)$$

où r est le rang dans la série de données classée par valeurs croissantes, n est la taille de l'échantillon, x[r] la valeur de rang r.

Rappelons encore que le temps de retour T d'un événement est défini comme étant l'inverse de la fréquence d'apparition de l'événement. Soit :

$$T = \frac{1}{1-F(x_i)} \quad (6)$$

2/ Démarche et résultats :

Etape 1 : Préparation de la série de données.

* Trier les valeurs dans l'ordre croissant.
* Attribuer un rang à chaque valeur.

Etape 2 : Calcul de la fréquence empirique pour chaque rang (Hazen, équation (5)).

Etape 3 : Calcul de la variable réduite « u » du Gumbel (équation (4)).

Etape 4 : Représentation graphique des couples (ui, xi) de la série à ajuster

Etape 5 : Ajustement d'une relation linaire de type aux couples (ui, xi) (figure 1) et en déduire les deux paramètres a et b). Avec un **ajustement de type graphique** (à l'œil), on a alors une estimation des paramètres a et b

<u>Exercice</u>

Pour le bassin versant X on vous demande :

1) Ajuster la série de précipitations maximums annuelles selon une distribution de Gumbel.

Ajuster les données graphiquement.

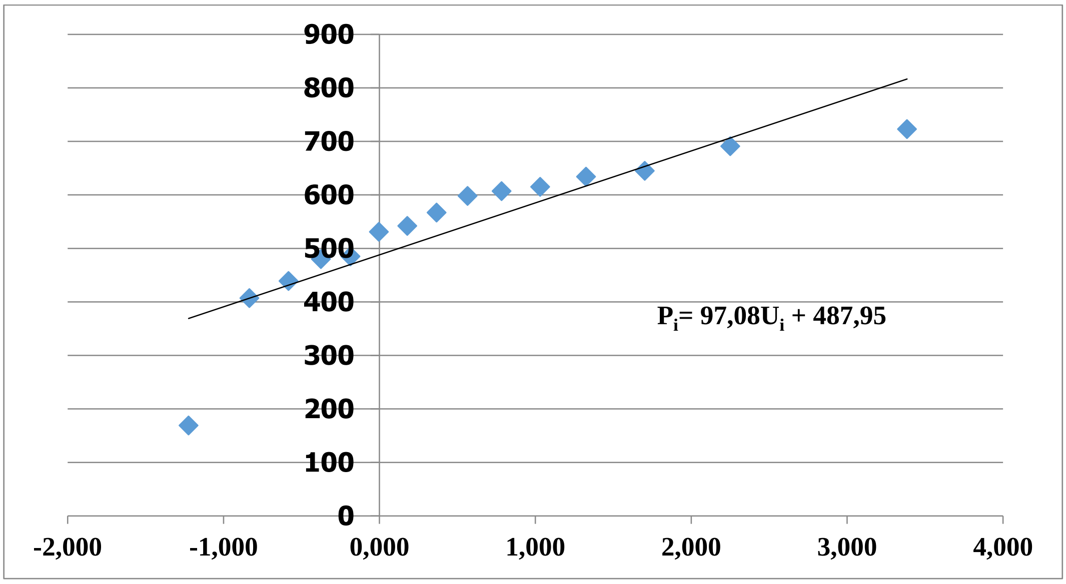
2) Estimer les précipitations annuelles de temps de retour, 5, 20, 50, 100 ans.

Année	**Précipitation (mm)**
2005	607
2006	169
2007	407
2008	439
2009	480
2010	485
2011	531
2012	542
2013	567
2014	598
2015	615
2016	634
2017	645
2018	691
2019	723

Tableau 5 : Valeurs de précipitations moyennes annuelles d'un bassin versant X

<u>Réponse</u>

Précipitation	**P par ordre croissant**	**rang**	**Fréquence (fxi)**	**variable réduite Ui**
607	169	1	0,0333	-1,224
169	407	2	0,1000	-0,834
407	439	3	0,1667	-0,583
439	480	4	0,2333	-0,375
480	485	5	0,3000	-0,186
485	531	6	0,3667	-0,003
531	542	7	0,4333	0,179
542	567	8	0,5000	0,367
567	598	9	0,5667	0,566
598	607	10	0,6333	0,784
615	615	11	0,7000	1,031
634	634	12	0,7667	1,325
645	645	13	0,8333	1,702
691	691	14	0,9000	2,250
723	723	15	0,9667	3,384



Période de retour T	100	50	20	5	2
F(xi)	0,99	0,98	0,95	0,8	0,5
variable Ui	4,6001	3,9019	2,9702	1,4999	0,3665
P (mm) pour T	934,48	866,70	776,25	633,51	523,48

2 Présentation de la loi de Normale (loi de Laplace Gauss)

Les lois normales ont une grande importance en statistiques. La courbe représentative de leur fonction de densité est appelée courbe de Gauss ou courbe en cloche du fait de sa forme. Elle possède un axe de symétrie en la moyenne ou la médiane

La variable utilisée est continue, C'est-à-dire qu'elle peut prendre un nombre indéfini de valeurs.

Cette courbe a deux paramètres : μ et σ.

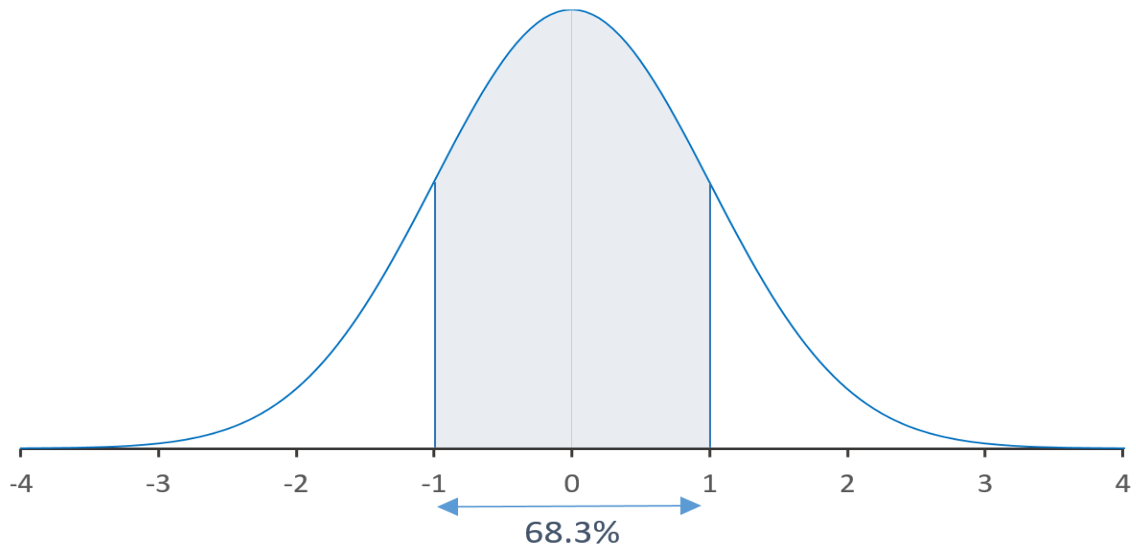
$$p(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}$$

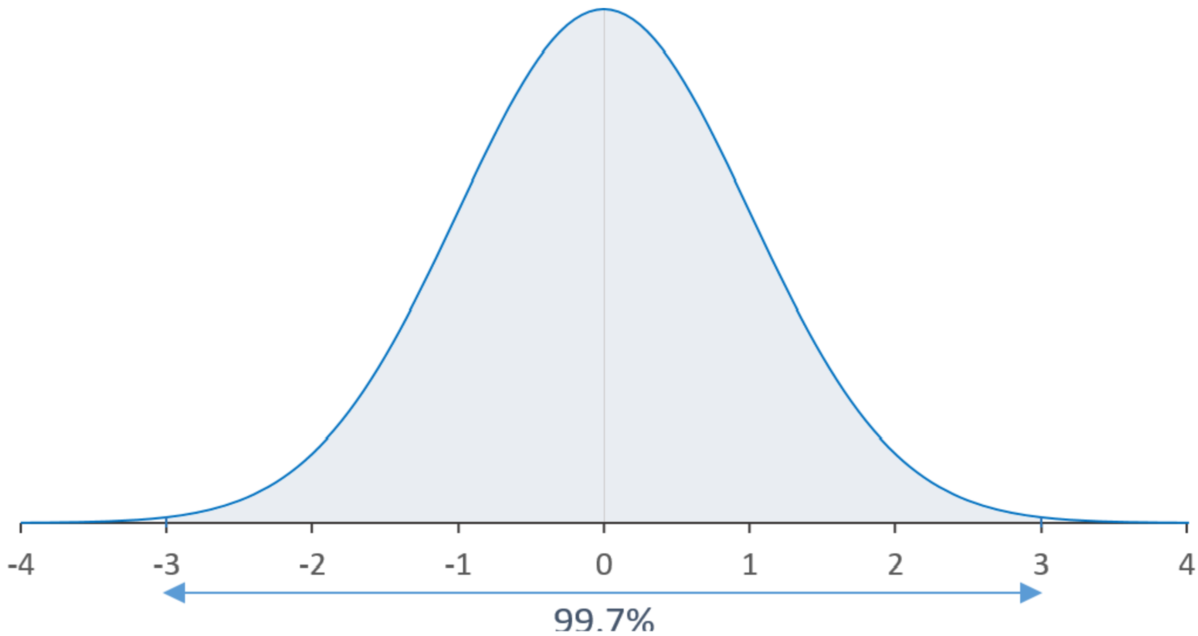
Son équation finalement très simple puisque ces 2 seuls paramètres suffisent : μ et σ. Les autres éléments de l'équation sont des constantes : Le **nombre d'Euler *e*** (2,71828) et **Pi *π*** (3.14159)

Les mathématiciens ont simplifié les choses en calculant les aires sous une loi normale spéciale de paramètres : μ=0 et de σ=1. Cette distribution est connue sous le nom de **loi normale centrée réduite**.

$$P(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}U^2}$$

On dit alors que U est distribué suivant une loi normale centrée réduite de moyenne nulle et de variance 1



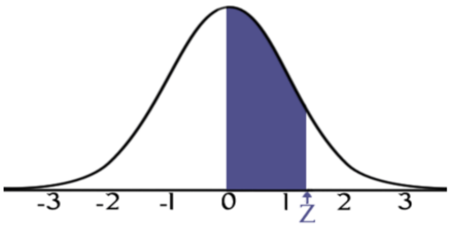


Les deux figures représente les courbes normales centrées réduites, on a reporté sur une graphique les aires comprises entre U = -1 et U = +1 et U = -2 et U = +2 et U = -3 et U = +3 qui sont respectivement «égale à 68.27%, 95.45%, 99.73% de l'aire totale

La table suivante indique les aires comprise ou la courbe normale entre u = 0, et toutes valeurs positives de U. A partir de cette table on peut obtenir l'aire comprise entre deux valeurs quelconques en utilisant la symétrie de la courbe par rapport à U =0.

Utilisation de la table de la loi normale centre réduite

Les tables sont constituées généralement d'un schéma et d'un tableau.



STANDARD NORMAL TABLE (Z)

Entries in the table give the area under the curve between the mean and z standard deviations above the mean. For example, for $z = 1.25$ the area under the curve between the mean (0) and z is 0.3944.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0190	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2969	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3513	0.3554	0.3577	0.3529	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998

Exemple d'ajustement d'une loi normale à un échantillon

Pour le bassin versant on vous demande :

Ajuster les données graphiquement.

1) Estimer les précipitations annuelles de temps de retour, 5, 20, 50, 100 ans.

Année	**Précipitation (mm)**
2005	607
2006	169
2007	407
2008	439
2009	723
2010	531
2011	485
2012	634
2013	567
2014	615
2015	598
2016	542
2017	645
2018	691
2019	480

Réponse

On se propose d'ajuster une loi normale (loi de Gauss) à un échantillon donné de pluie annuelle. Les étapes à suivre sont les suivantes

1- Calcul des caractéristiques empiriques :

- Moyenne arithmétique $\bar{P}$
- Ecart-type σ

2- Classement des valeurs :

On classe les valeurs de l'échantillon par ordre croissant ou décroissant en attribuant à chacune des valeurs son numéro d'ordre $\mathbf{n_i}$ compté à partir de 1.

3- Calcul de la fréquence expérimentale

$$f(x) = \frac{n_i - 0.5}{N}$$

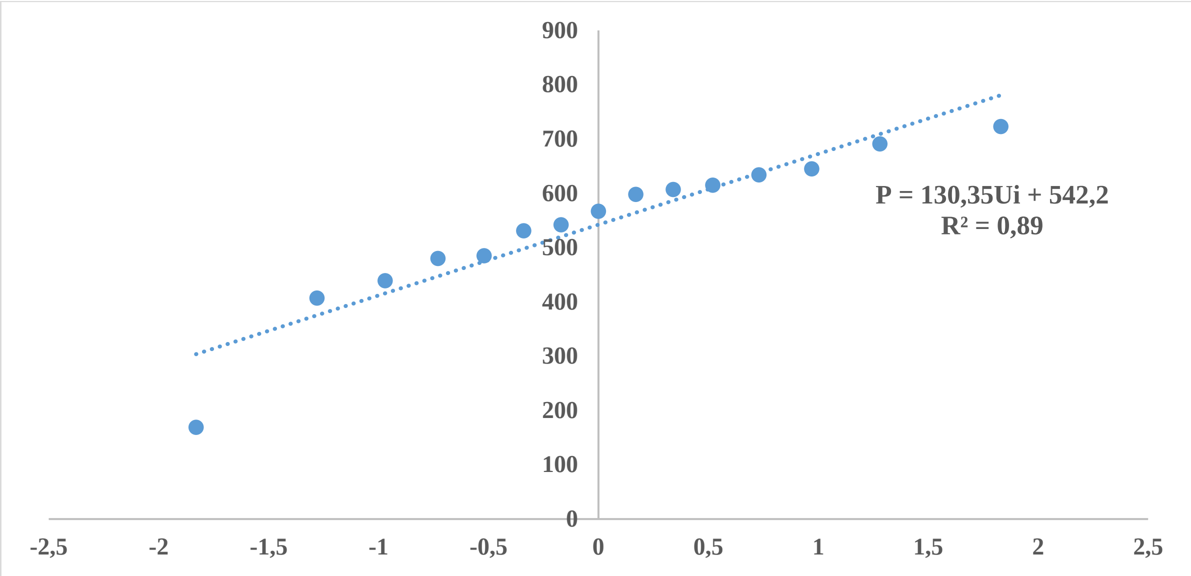
4- On déduit les valeurs de Ui à partir des tables à l'aide des fréquences empiriques F (x)

5- Tracer de la courbe de Henry : Pi = a Ui+b

6- A l'aide de ce modèle Pi = a Ui+b on déduit toues les précipitations probables pour différent temps de retour

Le tableau qui suit récapitule toutes les opérations de calcul

Précipitation (mm)	P par ordre croissant	rang	Fréquence f(x)	variable réduite U
607	169	1	0,0333	-1,83
169	407	2	0,1000	-1,28
407	439	3	0,1667	-0,97
439	480	4	0,2333	-0,73
480	485	5	0,3000	-0,52
485	531	6	0,3667	-0,34
531	542	7	0,4333	-0,17
542	567	8	0,5000	0
567	598	9	0,5667	0,17
598	607	10	0,6333	0,34
615	615	11	0,7000	0,52
634	634	12	0,7667	0,73
645	645	13	0,8333	0,97
691	691	14	0,9000	1,28
723	723	15	0,9667	1,83



Pmoy	542,2	mm
ecart type	136,93	mm

Période de retour T	100	50	20	5	2
F(xi)	0,99	0,98	0,95	0,8	0,5
Variable de Gauss Ui	2,33	2,05	1,65	0,84	0
Précipitation (mm) pour différent T	845,9155	809,4175	757,2775	651,694	542,2