

A complex network graph visualization with numerous nodes and edges, rendered in shades of blue and yellow, forming a large, abstract shape that resembles a map of Europe. The nodes are represented by small circles, and the edges are thin lines connecting them. The overall structure is dense and interconnected, with some clusters of nodes being more prominent than others.

# Introduction aux Statistiques

*MEDAS - 2017*

Thibault ALLART



# TOUR D'HORIZON

---

- Rappels (ou pas) sur les probabilités
- Statistiques descriptives
- Inférence statistique : intervalles de confiance, TCL, ...
- Tests statistiques
- Régression et classification
- Programmation R

*Probabilités*

*Statistiques*

*A/B-Testing*

*Tracking*

*Part #1*

# PROBABILITES



## QUESTION

---

- Soit un générateur de nombres aléatoires à valeur  $\{1, 2, 3, 4, 5, 6\}$   
Quelles sont les chances que le résultat soit 6 ?

Aléatoire  $\neq$  Équiprobable

*Probabilités*

*Statistiques*

*A/B-Testing*

*Tracking*

*Part #1.1*

# PROBABILITES : DEFINITIONS



# PROBABILITÉ

---

- Soit  $A$  un évènement, alors

$$0 \leq P(A) \leq 1$$

Rappel :  $60\% = \frac{60}{100} = 0.60$

- On note souvent  $\omega$  l'ensemble de tous les évènements et l'on a

$$P(\omega) = 1$$

Sur un dèss à 6 faces

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

$$P(\omega) = P(\text{dèss} = 1) +$$

$$\dots P(\text{dèss} = 6) = 1$$





# COMPLÉMENTAIRE

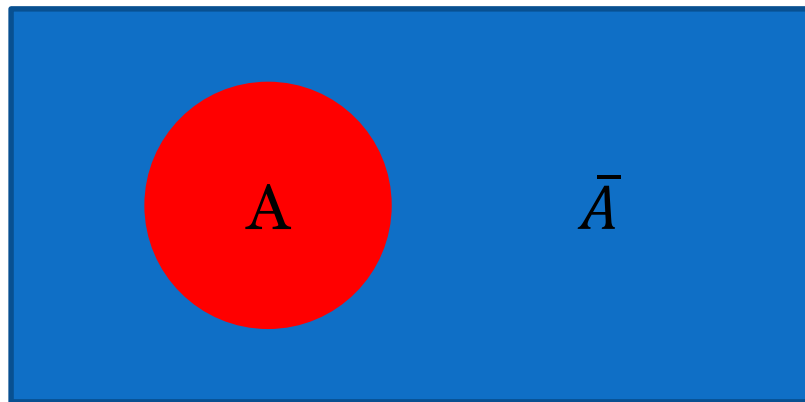
- Soit  $A$  un évènement, on note  $\bar{A}$  son complémentaire

$$P(\bar{A}) = 1 - P(A)$$

Exemple :

$A$  : le résultat du lancer est pair.

$\bar{A}$  : le résultat du lancer est impair.



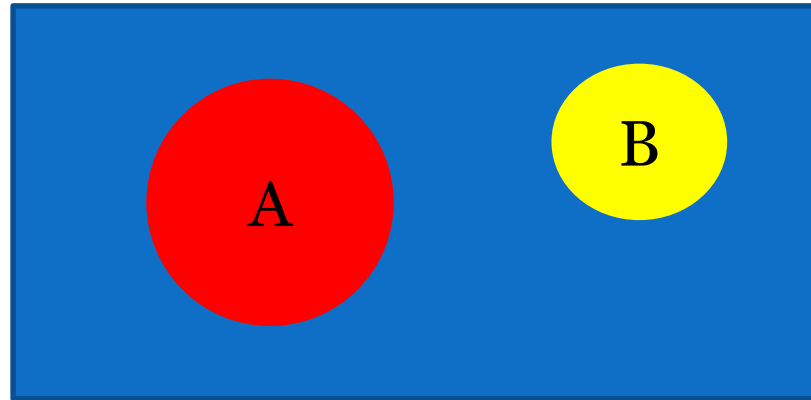


$I_2$

# SOMME D'ÉVÈNEMENTS INDÉPENDANTS

- Soit  $A$  et  $B$  deux évènements indépendants, alors

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



# UNION ET INTERSECTION

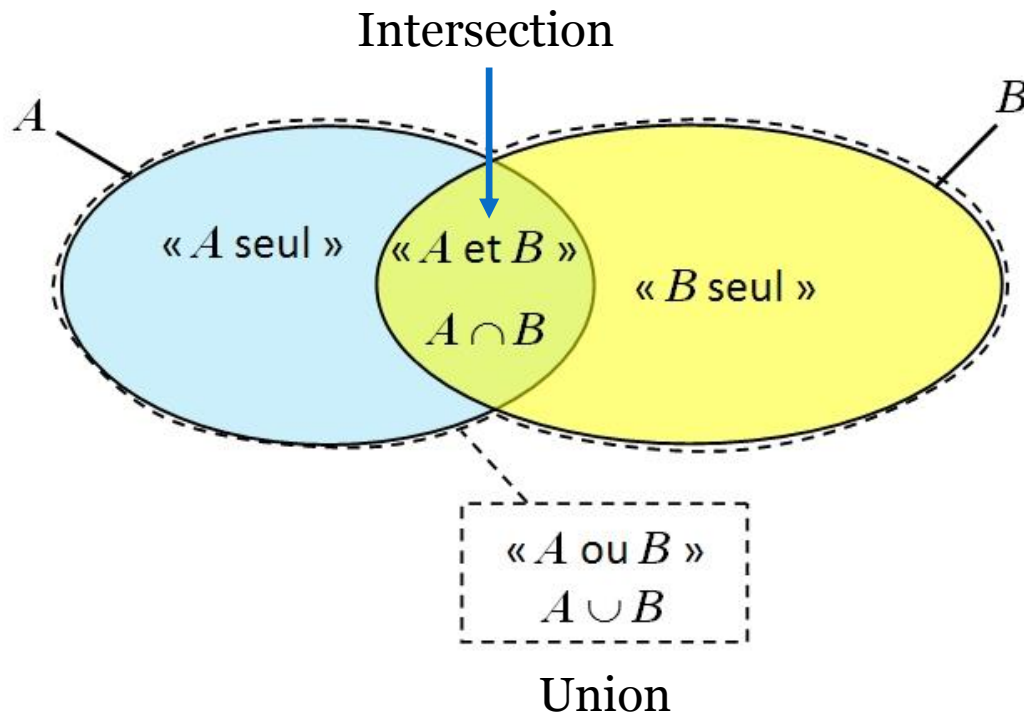
Exemple :

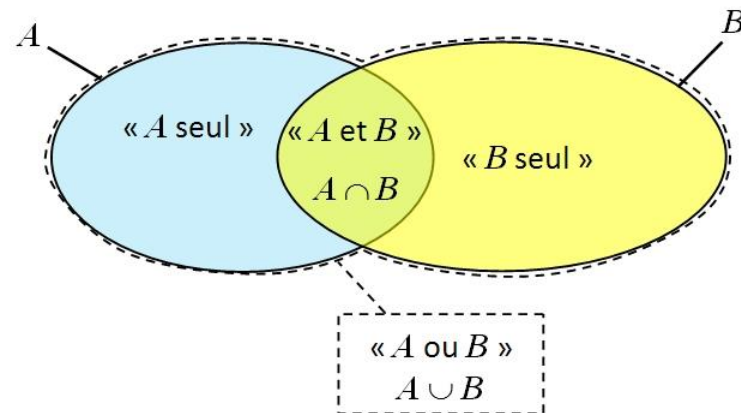
$$A = \{1, 2, 3\}$$

$$B = \{2, 4, 6\}$$

$$A \cup B = \{1, 2, 3, 4, 6\}$$

$$A \cap B = \{2\}$$





Formule des probabilités totales :

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$



# PROBABILITÉS TOTALES

---

Dans un jeu de 32 carte, quelle est la probabilité de tirer un roi ou un cœur ?

- + La probabilité de tirer un roi :  $4/32$
- + La probabilité de tirer un cœur :  $1/4=8/32$
- La probabilité de tirer le roi de cœur :  $1/32$

Résultat :  $11/32$

1) Calculez  $P(\overline{A \cup B})$

2) Calculez  $P(\overline{A \cap B})$

Faire un dessin

Réponses :

1)  $\bar{A} \cap \bar{B}$

2)  $\bar{A} \cup \bar{B}$

To do : Exercice 1

# PROBABILITÉS CONDITIONNELLES

---

La probabilité conditionnelle de A sachant B est notée :

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

Formule de Bayes :

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}.$$

Formule des probabilités totales :

$$\mathbb{P}(A) = \sum_{i \in I} \mathbb{P}(A|B_i)\mathbb{P}(B_i).$$

To do : exercice 2



# INDÉPENDANCE

---

Deux évènements A et B sont indépendants si et seulement si :

$$P(A|B) = P(A)$$

Ce qui est équivalent à

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



*Probabilités*

*Statistiques*

*A/B-Testing*

*Tracking*

*Part #1.2*

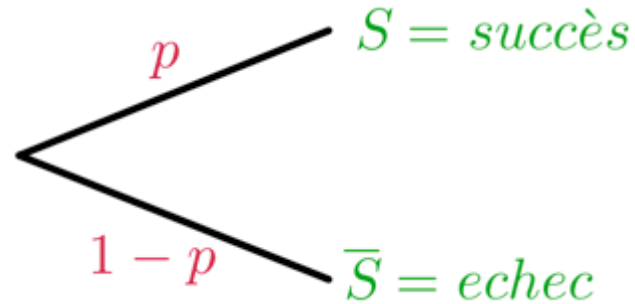
# PROBABILITES : LOIS ET CALCUL

# $I_6$

## BERNOULLI

Si je lance une pièce, quelles sont mes chances qu'elle tombe sur pile?

$$\mathbb{P}(X = 1) = 1 - \mathbb{P}(X = 0) = p.$$

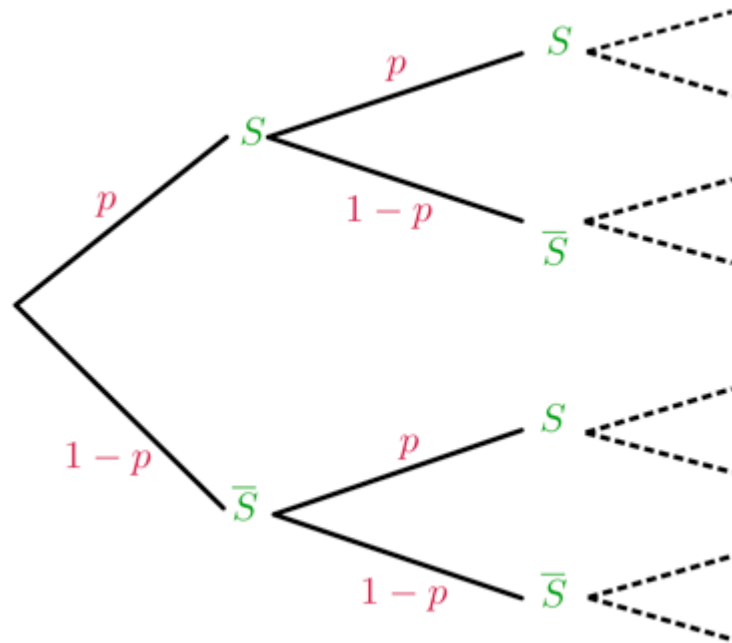


$I_7$

# ARBRE DE BERNOULLI

Quelles sont les chances qu'elle tombe 2 fois sur pile?

D'avoir une fois pile et une fois face?

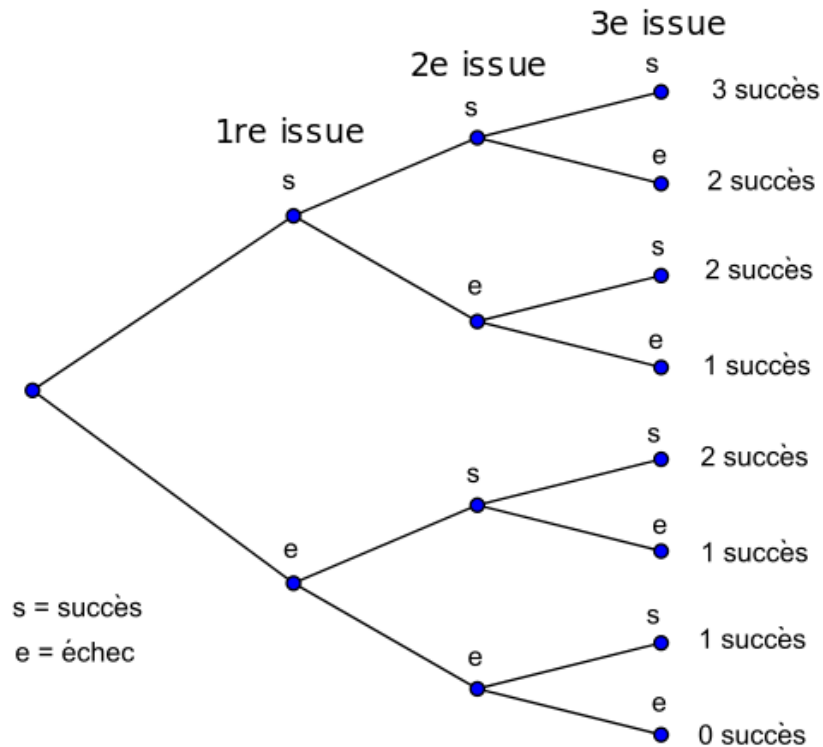




# LOI BINOMIALE

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

To do : exercice 3



# $I_9$ LOTO

Quelle est la probabilité de gagner à l'Euro Millions?

- 5 numéros entre [1,50]
- 2 numéros entre [1,11]

To do : exercice 2

$$\binom{50}{5} * \binom{11}{2} = \frac{50 * 49 * 48 * 47 * 46}{5 * 4 * 3 * 2 * 1} * \frac{11 * 10}{2 * 1} \\ = 116\,531\,800$$

Sachant que j'ai joué 100 fois au loto par le passé, quelles sont mes chances aujourd'hui?

To do : Ex 4





# EXERCICE TYPE BAC

Un joueur débute un jeu vidéo et effectue plusieurs parties successives.

On admet que :

- la probabilité qu'il gagne la première partie est de 0,1
- s'il gagne une partie, la probabilité de gagner la suivante est égale à 0,8
- s'il perd une partie, la probabilité de gagner la suivante est égale à 0,6

On note, pour tout entier naturel  $n$  non nul :

- $G_n$  l'évènement « le joueur gagne la  $n$ -ième partie »
- $p_n$  la probabilité de l'évènement  $G_n$

On a donc  $p_1 = 0, 1$ .

1. Montrer que  $p_2 = 0, 62$ . On pourra s'aider d'un arbre pondéré.
2. Le joueur a gagné la deuxième partie. Calculer la probabilité qu'il ait perdu la première.
3. Calculer la probabilité que le joueur gagne au moins une partie sur les trois premières parties.

Correction : [Exercices\Correction proba conditionnelles.docx](#)

To do : Ex 5

# $I_{17}$

## ESPÉRANCE D'UNE LOI DISCRÈTE

Soit une variable aléatoire  $X$  prenant pour valeur  $x_1, \dots, x_n$  avec probabilité  $p_1, \dots, p_n$ , alors l'espérance de  $X$  est définie comme

$$\mathbb{E}[x] = x_1p_1 + x_2p_2 + \dots + x_np_n$$

$$\mathbb{E}[x] = \sum_i x_i p_i$$

Calculer l'espérance d'un lancer de dé.

Calculer l'espérance de la somme de 2 lancers de dés.



To do ex 6

$$\mathbb{E}[x] = x_1p_1 + x_2p_2 + \cdots + x_np_n$$

Calculer l'espérance d'un lancer de dé.

$$m = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3,5.$$





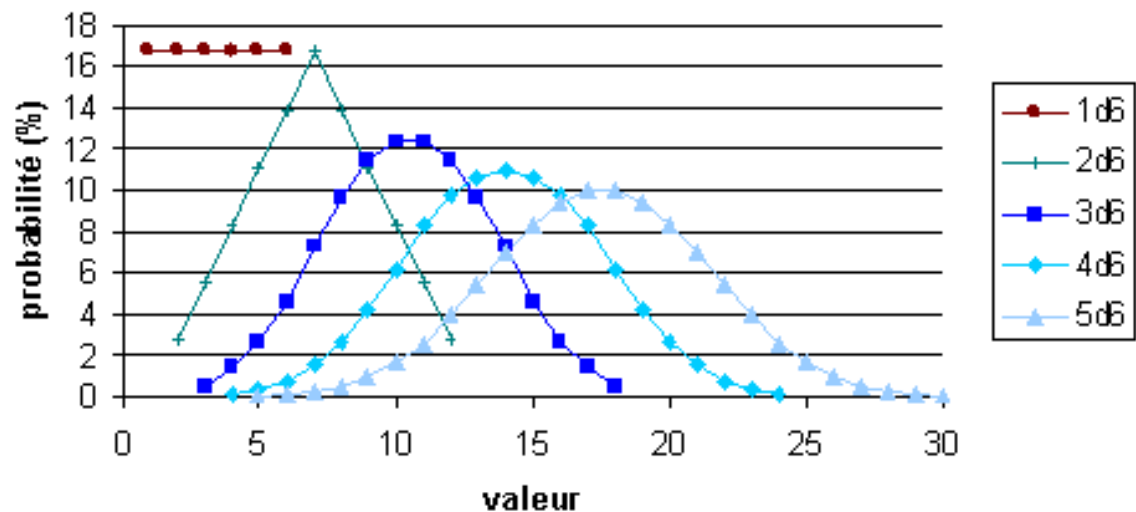
# ESPERANCE D'UNE LOI DISCRETE

Calculer l'espérance de la somme de 2 lancers de dés.

Somme des dés	2	3	4	5	6	7	8	9	10	11	12
Probabilité	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

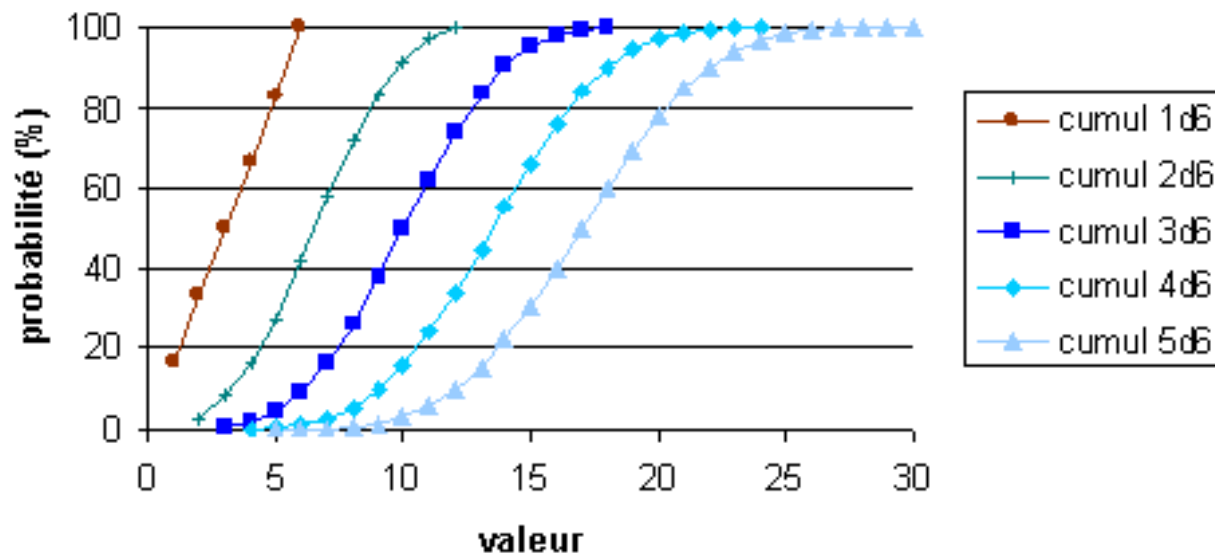
# ESPERANCE D'UNE LOI DISCRETE

Probabilité d'avoir une valeur



# ESPERANCE D'UNE LOI DISCRETE

Probabilité de faire moins qu'une valeur



# ESPERANCE D'UNE LOI DISCRETE

---

Distribution de 7 dés de jeux de rôle

- Calculez l'espérance d'un lancer de dés
- Calculez les probabilités associées au produit de 2 lancers de dés.

To do ex 7

# VARIANCE D'UNE LOI DISCRETE

---

Soit une variable aléatoire  $X$  prenant pour valeur  $x_1, \dots, x_n$  avec probabilité  $p_1, \dots, p_n$ , alors la variance de  $X$  est définie comme

$$Var[x] = \frac{1}{n} \sum_{i=1}^n (x_i - \mathbb{E}[x])^2$$



# EXPECTED VALUE

---

- Quelle est l'attaque offrant les meilleurs dégâts au long terme?
- Si l'ennemi a 15HP, cette attaque est elle toujours la meilleure?

Attack Name	Chance of hitting	Damage
Wind	100%	4
Fireball	80%	5
Lighting bolt	20%	40

Code : [Exercices\Expected value.xlsx](#)

To do : Ex 8



# LEARNING R

---

Pour les lois plus complexes, nous allons procéder par simulations.

R est un langage largement utilisé par la communauté statistique, voici un tutoriel pour en apprendre les bases.

To do : <http://tryr.codeschool.com/>

or

<https://www.datacamp.com>

# Simulation Monte-Carlo



- Soit un jeu de carte dans lequel il est possible de collectionner N cartes différentes. Sachant que l'on possède déjà K cartes différentes, quelle est la probabilité d'en obtenir une nouvelle que l'on ne possède pas déjà?

Code R :

[Exercices\collectionneur.R](#)



To do : Ex 9



# FORGEMAGIE

---

- Quelle est la probabilité d'overmager un item avec une rune PA (proba=1/100) au bout de K tentatives?
  - $K=100$  ?
  - Trouver le K tel que la proba soit de  $\frac{1}{2}$
  - Simulation des revenus de l'agent en fonction de sa mise de départ et des prix proposés. Puis Monte-Carlo  $E[\text{gain}]$ ,  $P[\text{faillite}]$
- Idem pour la probabilité de Drop un item rare.
  - Quelle est la distribution des joueurs possédant l'item au bout de n jours, si ils ne peuvent attaquer le monstre qu'une fois par jour?

Code R : [Exercices\proba exo item.R](#)

To do : Ex 10 et 11



# LOIS USUELLES

---

- Slides [modeles\Slides - Lois usuelles.pdf](#)
- Trouver une application Design à chacune d'elle.



# KEY POINTS

- RANDOM  $\neq$  UNKNOWN
- LE DESIGNER MAITRISE L'ALEATOIRE !

*Probabilités*

*Statistiques*

*A/B-Testing*

*Tracking*

*Part #2*

# STATISTIQUES ELEMENTAIRES



Thibault ALLART



## QUELQUES DÉFINITIONS

---

- Population : ensemble d'éléments sur lequel portent les observations
- Individu : élément d'une population
- Variable : Propriété étudiée dans une étude statistique
  - Qualitatif. Ex: Classe du personnage, sexe
  - Quantitatif
    - Discret : Nombre d'items possédés
    - Continu : Position du joueur, Points de vie.



# STATISTIQUES ÉLÉMENTAIRES

---

x
10
8
13
9
11
14
6
4
12
7
5

Pour x calculez :

- La moyenne
- La médiane
- La Variance
- L'écart-type

[https://thibaultallart.github.io/teaching/statistics/correlation\\_Anscombe.txt](https://thibaultallart.github.io/teaching/statistics/correlation_Anscombe.txt)

# II<sub>3</sub>

## STATISTIQUES ÉLÉMENTAIRES

x	y
10	8.04
8	6.95
13	7.58
9	8.81
11	8.33
14	9.96
6	7.24
4	4.26
12	10.84
7	4.82
5	5.68

Pour y calculez :

- La moyenne
- La médiane
- La Variance
- L'écart-type

Calculez ensuite la corrélation entre x et y.

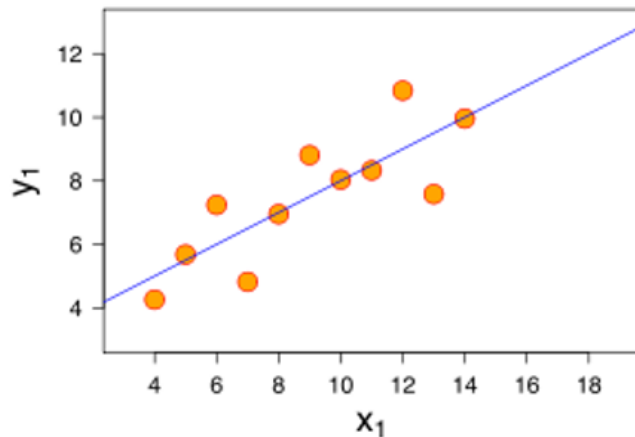
Code R : [Exercices\Corrélation et scatter plot.R](#)



# II<sub>4</sub>

## REGRESSION LINEAIRE

x	y
10	8.04
8	6.95
13	7.58
9	8.81
11	8.33
14	9.96
6	7.24
4	4.26
12	10.84
7	4.82
5	5.68



$$y_1 = ax + b$$

Que valent a et b?

Quelle prediction peut on donner pour  $x_1=12$  ?

Pour  $x_1=18$  ?



## CORRELATION

Refaire les calculs pour B, C et D.

A		B		C		D	
x	y	x	y	x	y	x	y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.1	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.1	4	5.39	19	12.5
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89



# CORRELATION

---

Property	Value
<u>Mean</u> of $x$ in each case	9 (exact)
Sample <u>variance</u> of $x$ in each case	11 (exact)
Mean of $y$ in each case	7.50 (to 2 decimal places)
Sample variance of $y$ in each case	4.122 or 4.127 (to 3 decimal places)
<u>Correlation</u> between $x$ and $y$ in each case	0.816 (to 3 decimal places)
<u>Linear regression</u> line in each case	$y = 3.00 + 0.500x$ (to 2 and 3 decimal places, respectively)



## CORRELATION

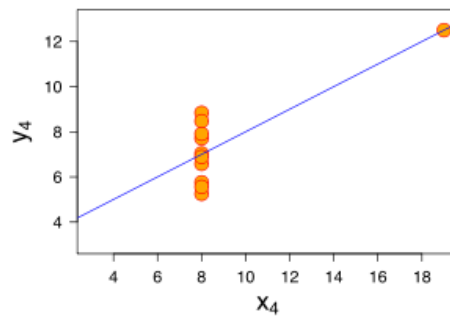
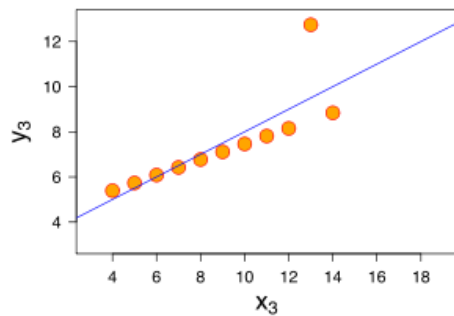
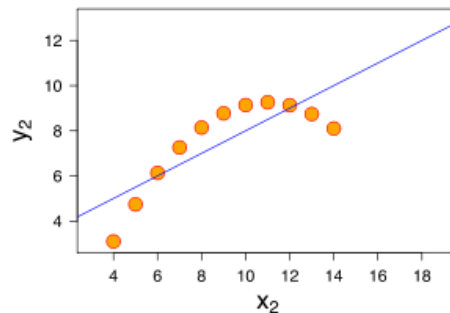
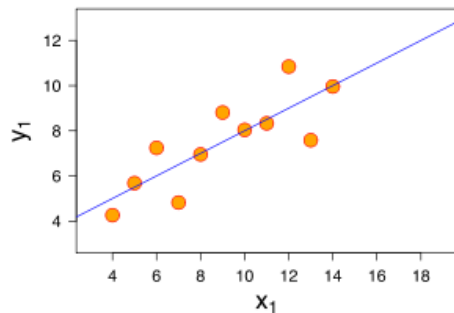
---

Affichez sous R les 4 graphiques de  $y \sim x$  et la droite de régression linéaire. Que constatez vous?



# CORRELATION

---



*Probabilités*

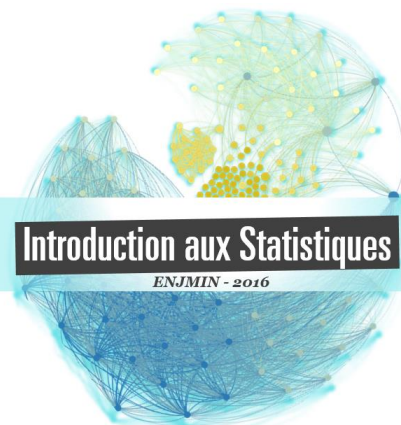
*Statistiques*

*A/B-Testing*

*Tracking*

*Part #3*

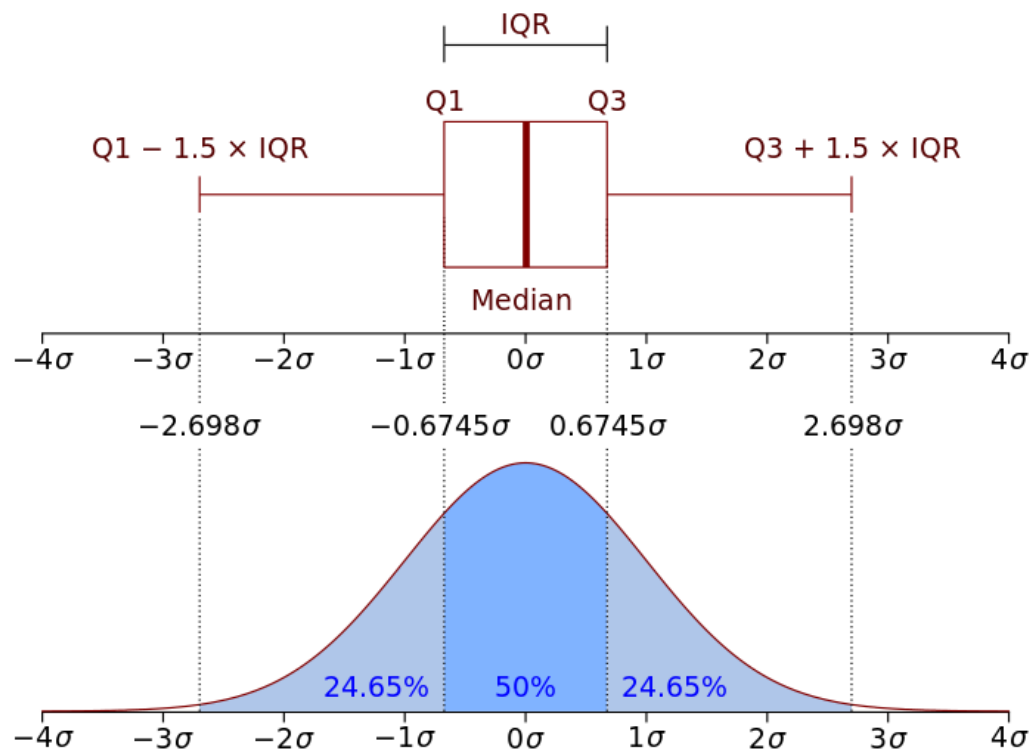
# INTERVALLES DE CONFIANCE



Thibault ALLART



# LOI NORMALE





# LOI NORMALE

---

- <http://sites.uclouvain.be/selt/shiny/xzp/>





## UTILITÉ DES STATISTIQUES

---

- J'ai lancé une pièce 100 fois et elle est tombée 59 fois sur pile. Peut on affirmer qu'elle est truquée?



## UTILITÉ DES STATISTIQUES

---

- J'ai lancé une pièce 100 fois et elle est tombée 59 fois sur pile. Peut on affirmer **avec un taux d'erreur de 5%** qu'elle est truquée?

## Théorème Central Limite

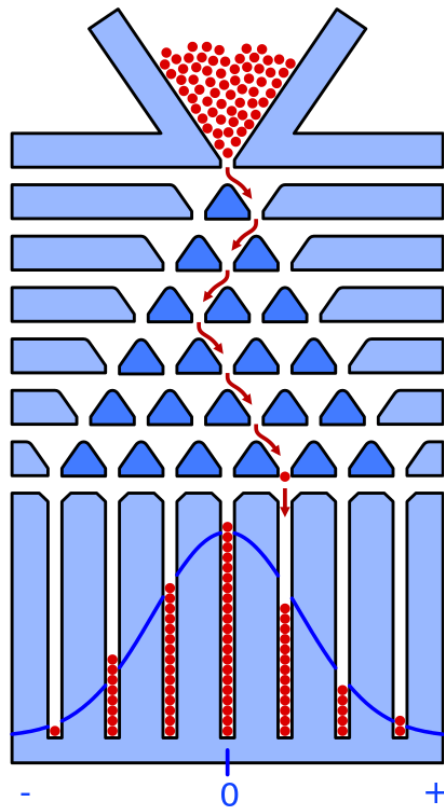
Soient  $X_1, \dots, X_n$  des variables aléatoires réelles indépendantes, de même loi

$$\frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \quad \text{lorsque } n \rightarrow +\infty.$$

En déduire une statistique de test d'une proportion

# III<sub>5</sub>

## CONVERGENCE DE LA LOI BINOMIALE



Code R :

[Exercices\TCL.R](#)

Voir somme et produit de lois.R



# INTERVALLES DE CONFIANCE

---

- $IC_{95\%} = \left[ \bar{x} - 1,96 \frac{\sigma_x}{\sqrt{n}}; \bar{x} + 1,96 \frac{\sigma_x}{\sqrt{n}} \right]$
- Shiny IC : <http://sites.uclouvain.be/selt/shiny/ic/>

*Probabilités*

*Statistiques*

*A/B-Testing*

*Tracking*

*Part #4*

# TESTS STATISTIQUES



Thibault ALLART



# TESTS STATISTIQUES

---

- Il existe une multitude de tests statistiques. Les plus utilisés sont sur :  
<http://marne.u707.jussieu.fr/biostatgv/?module=tests>

Test de Student sous R :

[Exercices\Test de Student.R](#)