TUT'RENTRÉE 2013-14

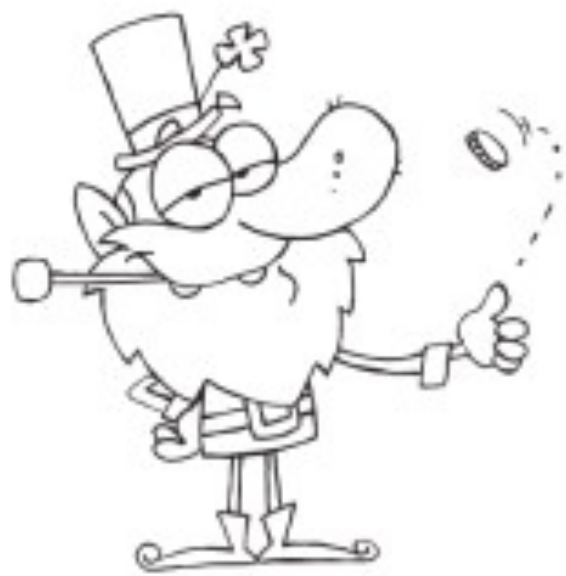
UE 4 : EVALUATION DES MÉTHODES D'ANALYSE APPLIQUÉES AUX SCIENCES DE LA VIE ET DE LA SANTÉ

PARTIE BIOSTASTIQUE (PR. BENOLIEL)

I-LA MÉTHODE STATISTIQUE EN MÉDECINE

II-STATISTIQUE DESCRIPTIVE

III-STATISTIQUE DÉDUCTIVE (PAS ABORDÉE)



"Maîtriser le hasard…"

I-La méthode statistique en médecine

Définitions générales

LA STATISTIQUE

- C'est l'art de collecter, d'analyser et d'interpréter des données.

- Elle dérive de l'étude des probabilités.

NE PAS CONFONDRE :

- LA statistique => une technique

- LES statistiques => une collection de valeurs

LA BIOSTATISTIQUE

- C’est l’application de la statistique au domaine biologique.

UNE DONNÉE

- C’est le résultat de l’observation d’un individu par :

  -un instrument (poids, taille…)

  -les sens de l’observateur (couleur de peau…)

UNE VARIABLE

• **Une donnée qui n'est pas strictement équivalente d'un individu à l'autre à un instant t**

=> variabilité INTER-INDIVIDUELLE

OU ENCORE

• **Une donnée qui n'est pas strictement équivalente d'un instant à l'autre pour un individu donnée**

=> variabilité INTRA-INDIVIDUELLE

Rappel : les différents types de variables !

Quantitatives		Qualitatives	
Discrète	Continue	• Nominale • Binaire	Ordinale
Nombre entier naturel	Nombre réel	Caractère non ordonnée	Caractère ordonné

3 QCM !

1) Indiquez la ou les réponse(s) juste(s) :

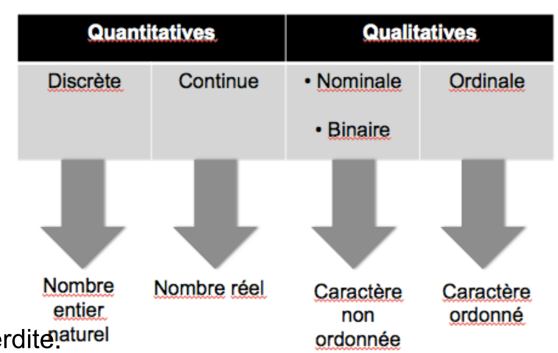
a) La masse d'une personne est une variable quantitative discrète.

b) Le nombre d'accouchements subits par une femme est une variable quantitative continue.

c) La couleur des yeux d'une personne est une variable qualitative ordinale.

d) La mention qu'un élève de PAES a reçu au bac est une variable qualitative nominale.

e) Aucune de ces affirmations n'est juste.



1) Indiquez la ou les réponse(s) juste(s) :

a) La masse d'une personne est une variable quantitative ~~discrète~~. CONTINUE.

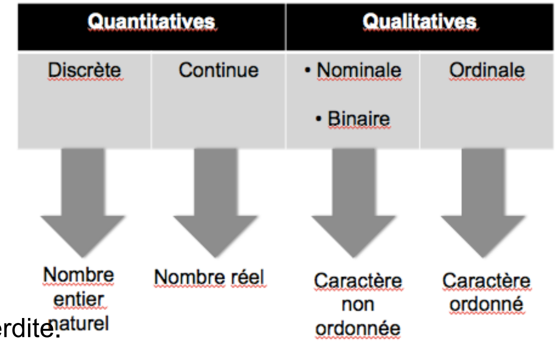
b) Le nombre d'accouchements subits par une femme est une variable quantitative ~~continue~~ DISCRETE.

c) La couleur des yeux d'une personne est une variable qualitative ~~ordinale~~ NOMINALE.

d) La mention qu'un élève de PAES a reçu au bac est une variable qualitative ~~nominale~~ ORDINALE.

e) Aucune de ces affirmations n'est juste.

Réponse : e !



2) Le sexe civil d'un individu est :

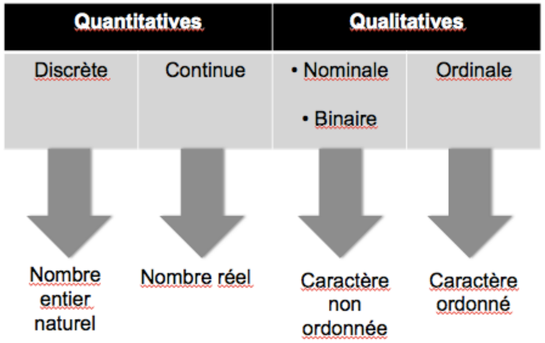
a) Une variable quantitative.

b) Une variable qualitative ordinale.

c) Une variable qualitative binaire.

d) Une variable qualitative nominale.

e) Aucune de ces affirmations n'est juste.



2) Le sexe civil d'un individu est :

a) Une variable ~~quantitative~~.

b) Une variable qualitative ~~ordinale~~.

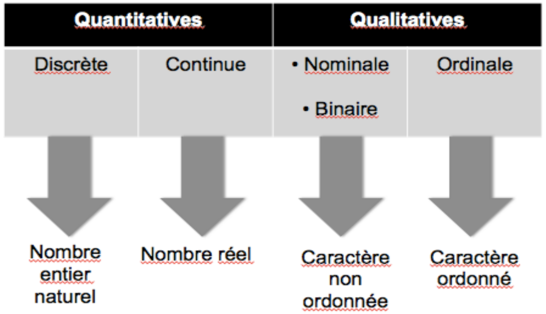
c) Une variable qualitative binaire.

d) Une variable qualitative nominale.

e) ~~Aucune de ces affirmations n'est juste~~.

Réponse : cd !

Ici la variable qualitative BINAIRE est un cas particulier de variable qualitative NOMINALE.



3) Un questionnaire de santé remis à un malade contient l’énoncé suivant : *“Evaluez votre douleur sur une échelle de 0 à 10 (0 : pas du tout douloureux et 10 : très douloureux)”*. La réponse attendue est :

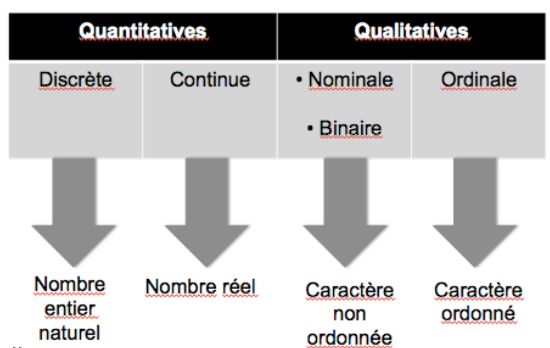
a) **Une variable qualitative ordinale.**

b) **Une variable qualitative nominale.**

c) **Une variable quantitative discrète.**

d) **Une variable quantitative continue.**

e) **Aucune de ces affirmations n’est juste.**



3) Un questionnaire de santé remis à un malade contient l'énoncé suivant : *"Evaluez votre douleur sur une échelle de 0 à 10 (0 : pas du tout douloureux et 10 : très douloureux)"*. La réponse attendue est :

a) **Une variable qualitative ordinale.**

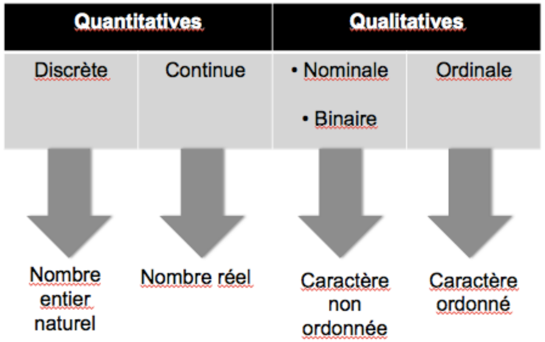
b) **Une variable qualitative ~~nominale~~.**

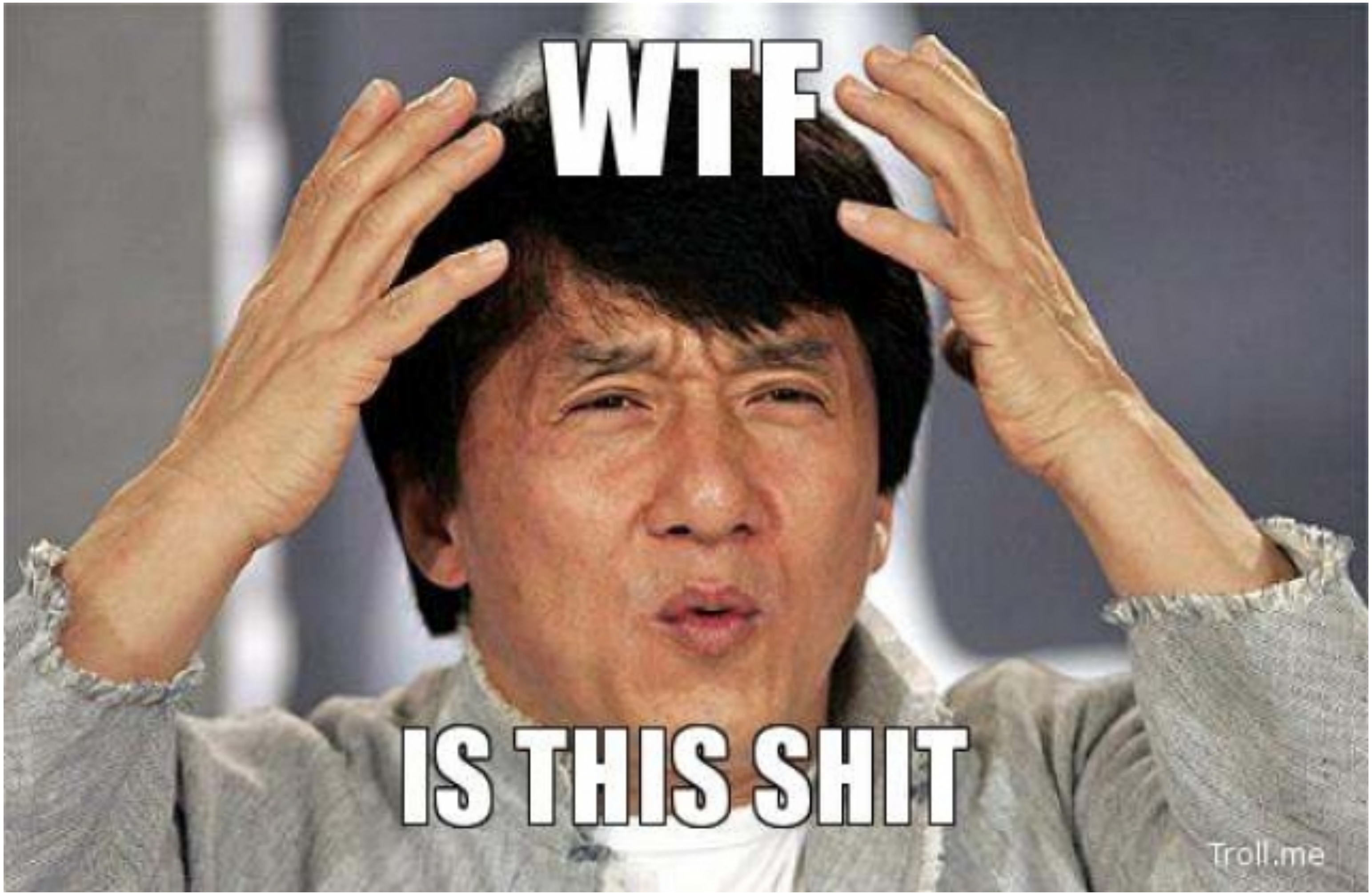
c) **Une variable ~~quantitative discrète~~.**

d) **Une variable ~~quantitative continue~~.**

e) **~~Aucune de ces affirmations n'est juste.~~**

Réponse : a !!!!!




WTF
IS THIS SHIT
Troll.me

A partir du moment où le degré de douleur ne se mesure pas à l'aide d'un instrument, il est difficile de le concevoir comme une variable quantitative. Pourtant son évaluation peut passer par une numérisation :

Evaluez votre douleur en plaçant un point sur l'échelle suivante :

Pas de douleur I-----------------------------------I Douleur intense

REVIENT au même que :

Evaluez votre douleur en entourant un chiffre :

0 1 2 3 4 5 6 7 8 9 10

Un patient souffrant d'une douleur "moyenne" répondra donc de la manière suivante :

Evaluez votre douleur en plaçant un point sur l'échelle suivante :

Pas de douleur I-----------------------------------I Douleur intense

REVIENT au même que :

Evaluez votre douleur en entourant un chiffre :

0 1 2 3 4 5 6 7 8 9 10

=> L'avantage de l'échelle numérique est de rendre les résultats plus facilement exploitables.

Ne JAMAIS oublier que :

Le fait que la variable soit numérique, n'implique pas nécessairement que ce soit une variable quantitative !

- **code postal…**
- **numéro étudiant…**

Une autre subtilité dans l'univers des variables

TENIR COMPTE DU CONTEXTE !!!!

Première situation

Question : Quelle est ta taille ?

Réponse : Je mesure 1m35 !

$\Rightarrow$ *Dans ce cas, quel type de variable représente la taille ?*

TENIR COMPTE DU CONTEXTE !!!!

Première situation

Question : Quelle est ta taille ?

Réponse : Je mesure 1m35 !

$\Rightarrow$ *Dans ce cas, quel type de variable représente la taille ?*

Variable quantitative !!!!!

TENIR COMPTE DU CONTEXTE !!!!

Seconde situation

Question :Quelle est ta taille ?

a) Moins de 1m60 (petit)

b) Entre 1m60 et 1m80 (moyen)

c) Plus de 1m80 (grand)

Réponse : a

⇒ *Dans ce cas, quel type de variable représente la taille ?*

TENIR COMPTE DU CONTEXTE !!!!

Seconde situation

Question :Quelle est ta taille ?

a) *Moins de 1m60 (petit)*
b) *Entre 1m60 et 1m80 (moyen)*
c) *Plus de 1m80 (grand)*

Réponse : a

⇒ *Dans ce cas, quel type de variable représente la taille ?*

Variable qualitative (ordinale) !!!!!

TENIR COMPTE DU CONTEXTE !!!!

Question : Quelle est ta taille ?

Réponse : Je mesure 1m50 !

Mesure d'une quantité (quantité de centimètres)

Variable QUANTITATIVE

Question :Quelle est ta taille ?

a) Moins de 1m60 (petit)

b) Entre 1m60 et 1m80 (moyen)

c) Plus de 1m80 (grand)

Réponse : a !

On catégorise (le fait d'être petit / moyen / grand)

Variable QUALITATIVE

Autres définitions

UN PARAMÈTRE

- C’est une grandeur apportant une information résumée sur la variablée étudiée.

Des exemples ?

UN PARAMÈTRE

- **C'est une grandeur apportant une information résumée sur la variablée étudiée.**

-moyenne

-variance

-médiane…

Surtout étudiés dans le contexte de la STATISTIQUE DESCRIPTIVE

HASARD OU VARIABILITÉ INTRINSÈQUE

• **Le hasard ou la variabilité intrinsèque est l'une des origines possibles à la variabilité.**

En effet, en statistique, l'idée centrale est que la variabilité a deux types d'origines possibles :

-le hasard = la variabilité intrinsèque

-une autre origine qui n'est pas forcément connue (en médecine, il peut s'agir d'une pathologie)

Surtout étudiés dans le contexte de la STATISTIQUE DÉDUCTIVE

Toute observation est soumise à une variabilité intrinsèque.

En effet, pour une observation répétée plusieurs de fois, le résultat est presque toujours variable.

L'observation d'une différence ne permet pas en soi d'en préciser la cause.

Constater une différence statistiquement significative ne donne pas la clé de son interprétation.

UNE SÉRIE STATISTIQUE

- **C'est une collection d'objets de même nature avec des caractèristiques différentes d'un objet à l'autre.**

-**population** : série de TOUS les individus étudiés, sur lesquels on veut inférer des décisions. La valeur de son effectif est souvent inconnue.

-**échantillon** : ensemble d'effectif limité et connu extrait de la population. L'objectif de la constitution d'un échantillon est qu'il soit représentatif de la population dont il est issu.

<u>**Un seul**</u> **moyen permet de sélectionner un échantillon représentatif au sein d'une population :**

LE TIRAGE AU SORT

= LA RANDOMISATION

Si le tirage au sort a été effectué correctement (<u>sans biais</u>), les observations menées sur l'échantillon sont généralisés à la population entière <u>à un risque d'erreur près</u>.

1 QCM !

1) L’ensemble des femmes qui habitent à Nice forment :

a) Une série statistique.
b) Une population.
c) Un échantillon représentatif des femmes qui habitent en France.
d) Un échantillon représentatif de la population niçoise.
e) Aucune de ces affirmations n’est juste.

1) L'ensemble des femmes qui habitent à Nice forment :

a) **Une série statistique.**
b) **Une population.**
c) **Un échantillon ~~représentatif~~ des femmes qui habitent en France.**
d) **Un échantillon ~~représentatif~~ de la population niçoise.**
e) ~~**Aucune de ces affirmations n'est juste.**~~

Réponse : ab

Fin de la première partie

II-Statistique Descriptive

1) Représentation des données

VARIABLES QUALITATIVES

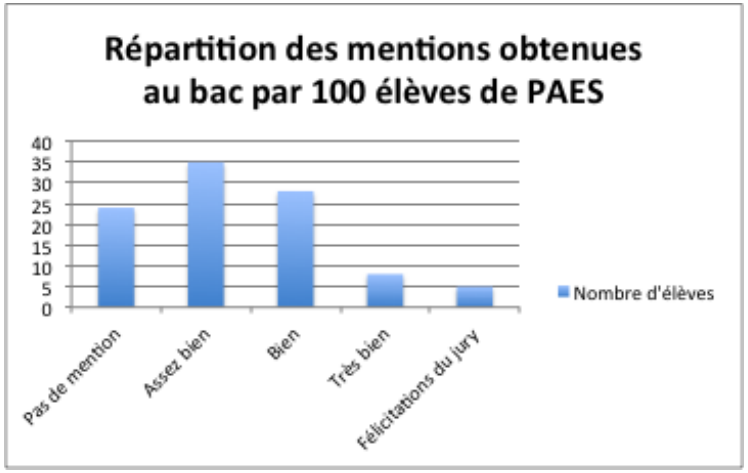
Deux manières de les représenter :

- **tableau**

Mentions	Nombre d'élèves
Pas de mention	24
Assez bien	35
Bien	28
Très bien	8
Félicitations du jury	5

- **histogramme**

(normalisé ou non)



VARIABLES QUANTITATIVES

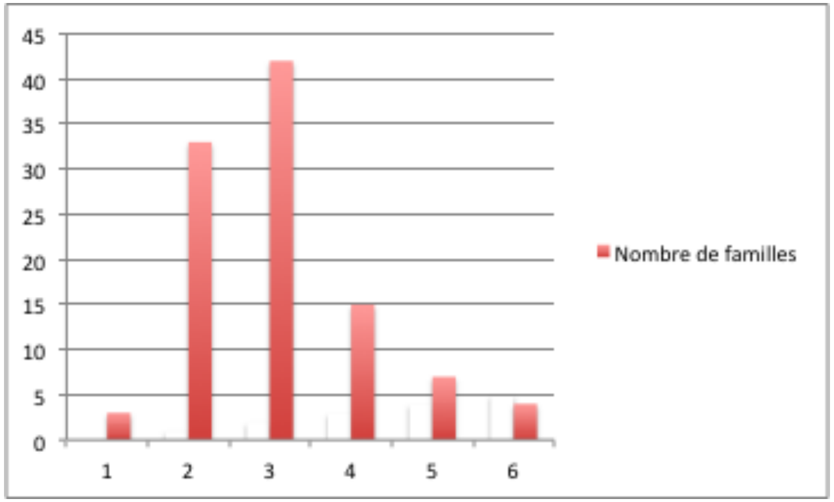
Deux manières de les représenter :

- **tableau**

Nombre d'enfants par famille	Nombre de familles
0	3
1	33
2	42
3	15
4	7
5	4

- **histogramme**

(normalisé ou non)



MAIS les variables QUANTITATIVES peuvent être synthétisées ou résumées par des paramètres :

-Indicateurs de position :

- **moyenne**
- **médiane**
- **quartiles** *(partagent une série ordonnée en 4 groupes de même effectif)*

-Indicateur de dispersion :

- **variance**
- **écart type**

variance = (écart type)^2

LA MOYENNE

- **Indicateur de <u>position</u>**

- **<u>Soit n données relevées par ordre croissant : x_1 ; x_i ; … x_n</u>**

$$m = \Sigma x_i / n$$

LA MEDIANE

- **Indicateur de <u>position</u>**

- **<u>Valeur centrale</u> d'une liste ordonnée par ordre croissant => la médiane sépare les premiers 50% de la série**

- **<u>Soit n données relevées par ordre croissant : x_1 ; x_i ; … x_n</u>**

-<u>Si n paire</u> : $M = (x_{n/2} + x_{n/2+1}) / 2$

-<u>Si n impaire</u> : $M = x_{(n+1) / 2}$

LES QUARTILES

- **Indicateurs de <u>position</u>**

- **<u>Valeurs qui partagent une série ordonnée en 4 groupes de même effectif</u>**

-Q1 (premier quartile) sépare les premiers 25% de la série

-Q2 (deuxième quartile) sépare les premiers 50% de la série

-Q3 (troisième quartile) sépare les premiers 75% de la série

Mais alors, quelle est la différence entre le Deuxième Quartile et la Médiane ?

Médiane

=

Deuxième Quartile !

- **<u>Soit n données relevées par ordre croissant : x_1 ; x_i ; … x_n</u>**

-<u>Si n est multiple de 4</u> :

- **$Q1 = x_{n/4}$**
- **$Q3 = x_{3n/4}$**

-<u>Si n n'est pas multiple de 4</u> :

- **$Q1 = (x_i + x_j) / 2$**

Avec i et j les deux valeurs les plus proches de n/4 tq : $i < n/4 < j$

- **$Q3 = (x_i + x_j) / 2$**

Avec i et j les deux valeurs les plus proches de 3n/4 tq : $i < 3n/4 < j$

<u>Exemple 1</u> : On s'intéresse à la médication d'un groupe de 12 patients âgés. Pour cela, on relève le nombre de types de médicaments pris par chacun des patients, et on range les valeurs de la façon suivante :

Numéro du patient	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Nombre de types de médicaments	2	3	3	3	4	5	5	7

- *Que vaut la médiane ?*

<u>Exemple 1</u> : On s'intéresse à la médication d'un groupe de 12 patients âgés. Pour cela, on relève le nombre de types de médicaments pris par chacun des patients, et on range les valeurs de la façon suivante :

Numéro du patient	1	2	3	4	5	6	7	8
Nombre de types de médicaments	2	3	3	3	4	5	5	7

- *Que vaut la médiane ?*

8 patients $\Rightarrow M = (x_{n/2} + x_{n/2+1}) / 2$

$\Rightarrow M = (x_{8/2} + x_{8/2+1}) / 2$

$\Rightarrow M = (x_4 + x_5) / 2$

$\Rightarrow M = (3+4)/2$

=> ***<u>M = 3,5 médicaments</u>***

<u>Exemple 1</u> : On s'intéresse à la médication d'un groupe de 12 patients âgés. Pour cela, on relève le nombre de types de médicaments pris par chacun des patients, et on range les valeurs de la façon suivante :

Numéro du patient	1	2	3	4	5	6	7	8
Nombre de types de médicaments	2	3	3	3	4	5	5	7

- *Que vaut Q1 ?*

<u>Exemple 1</u> : On s'intéresse à la médication d'un groupe de 12 patients âgés. Pour cela, on relève le nombre de types de médicaments pris par chacun des patients, et on range les valeurs de la façon suivante :

Numéro du patient	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Nombre de types de médicaments	2	3	3	3	4	5	5	7

- *Que vaut Q1 ?*

8 patients => Q1 = $x_{n/4}$
=> Q1 = $x_{8/4}$
=> Q1 = x_2
*=> **<u>Q1 = 3 médicaments</u>***

<u>Exemple 1</u> : On s'intéresse à la médication d'un groupe de 12 patients âgés. Pour cela, on relève le nombre de types de médicaments pris par chacun des patients, et on range les valeurs de la façon suivante :

Numéro du patient	1	2	3	4	5	6	7	8
Nombre de types de médicaments	2	3	3	3	4	5	5	7

- *Que vaut Q3 ?*

Exemple 1 : On s'intéresse à la médication d'un groupe de 12 patients âgés. Pour cela, on relève le nombre de types de médicaments pris par chacun des patients, et on range les valeurs de la façon suivante :

Numéro du patient	1	2	3	4	5	6	7	8
Nombre de types de médicaments	2	3	3	3	4	5	5	7

- *Que vaut Q3 ?*

8 patients => Q3 = $x_{3n/4}$
=> Q3 = $x_{3x8/4}$
=> *Q3* = $x_{24/4}$
=> *Q3* = x_6
=> ***Q3 = 5 médicaments***

Exemple 2 : on relève la masse de 5 enfants d'un service de pédiatrie

31 kg ; 30 kg ; 31 kg ; 33 kg ; 32 kg

=> On souhaite étudier ces données à l'aide de paramètres.

- **Quelle est la première étape ?**

Exemple : on relève la masse de 5 enfants d'un service de pédiatrie

~~*31 kg ; 30 kg ; 31 kg ; 33 kg ; 32 kg*~~

=> On souhaite étudier ces données à l'aide de paramètres.

- **Quelle est la première étape ? => ORDONNER DE FACON CROISSANTE !**

Numéro du patient	**1**	**2**	**3**	**4**	**5**
Masse (kg)	30	31	31	32	33

<u>Exemple</u> : on relève la masse de 5 enfants d'un service de pédiatrie

~~31 kg ; 30 kg ; 31 kg ; 33 kg ; 32 kg~~

=> On souhaite étudier ces données à l'aide de paramètres.

- **Quelle est la première étape ? => ORDONNER DE FACON CROISSANTE !**

Numéro du patient	1	2	3	4	5
Masse (kg)	30	31	31	32	33

- **Moyenne => ?**

Exemple : on relève la masse de 5 enfants d'un service de pédiatrie

~~*31 kg ; 30 kg ; 31 kg ; 33 kg ; 32 kg*~~

=> On souhaite étudier ces données à l'aide de paramètres.

• **Quelle est la première étape ? => ORDONNER DE FACON CROISSANTE !**

Numéro du patient	1	2	3	4	5
Masse (kg)	30	31	31	32	33

• **Moyenne => (30 + 31 + 31 + 32 + 33) / 5 = 31,4 kg**

<u>Exemple</u> : on relève la masse de 5 enfants d'un service de pédiatrie

~~31 kg ; 30 kg ; 31 kg ; 33 kg ; 32 kg~~

=> On souhaite étudier ces données à l'aide de paramètres.

- **Quelle est la première étape ? => ORDONNER DE FACON CROISSANTE !**

Numéro du patient	1	2	3	4	5
Masse (kg)	30	31	31	32	33

- **Moyenne => (30 + 31 + 31 + 32 + 33) / 5 = 31,4 kg**
- **Médiane => ?**

<u>Exemple</u> : on relève la masse de 5 enfants d'un service de pédiatrie

~~31 kg ; 30 kg ; 31 kg ; 33 kg ; 32 kg~~

=> On souhaite étudier ces données à l'aide de paramètres.

• **Quelle est la première étape ? => ORDONNER DE FACON CROISSANTE !**

Numéro du patient	1	2	3	4	5
Masse (kg)	30	31	31	32	33

• **Moyenne => (30 + 31 + 31 + 32 + 33) / 5 = 31,4 kg**

• **Médiane => 31 kg**

<u>Exemple</u> : on relève la masse de 5 enfants d'un service de pédiatrie

31 kg ; 30 kg ; 31 kg ; 33 kg ; 32 kg

=> On souhaite étudier ces données à l'aide de paramètres.

- **Quelle est la première étape ? => ORDONNER DE FACON CROISSANTE !**

Numéro du patient	1	2	3	4	5
Masse (kg)	30	31	31	32	33

- **Moyenne => (30 + 31 + 31 + 32 + 33) / 5 = 31,4 kg**
- **Médiane => 31 kg**
- **Q1 et Q3=> ?**

Numéro du patient	1	2	3	4	5
Masse (kg)	30	31	31	32	33

- **Calcul de Q1 pour un effectif non-multiple de 4 :**

=> $i < n/4 < j$

=> $i < 5/4 < j$

=> $i < 1,25 < j$

=> $i < 1,25 < j$

=> $1 < 1,25 < 2$

D'où : $Q1 = (x_i + x_j) / 2 = (x_1 + x_2) / 2 = (30+31)/2 =$ **30,5 kg**

1 QCM PORTANT ENCORE SUR CET EXEMPLE

1) **<u>Le tableau suivant est dressé dans le cadre d'une étude pédiatrique :</u>**

Numéro du patient	1	2	3	4	5
Masse (kg)	30	31	31	32	33

<u>La valeur du Troisième Quantile (Q3) est de :</u>

a) 31,5

b) 3,5

c) 32,5

d) 4,5

e) 30,5

1) **Le tableau suivant est dressé dans le cadre d'une étude pédiatrique :**

Numéro du patient	1	2	3	4	5
Masse (kg)	30	31	31	32	33

La valeur du Troisième Quantile (Q3) est de :

a) 31,5 => 3x5/4 = 15 / 4 = 3,75

D'où Q3 = (31+32) / 2 = 31, 5

~~b) 3,5~~

~~c) 32,5~~

~~d) 4,5~~

~~e) 30,5~~

Réponse : a

LA VARIANCE (ÉCART TYPE)^2

- Indicateurs de dispersion

- Paramètre important pour la Loi de Gauss que l'on va bientôt aborder

SURTOUT NE JAMAIS CONFONDRE :

	Avantages	Inconvénients
Moyenne	• Facile à calculer • Facile à manipuler => **adaptée aux calculs statistiques** • Significative si la répartition des données est symétrique et la dispersion faible	• Sensible aux valeurs anormales (mini ou maxi)
Médiane	• Facile à calculer • Peu sensible aux valeurs anormales • Utilisable pour les valeurs ordinales	• **Peu adaptée aux calculs statistiques**

Et ne pas oublier que…

La synthèse par des paramètres n'est bien évidemment PAS APPLICABLE pour l'étude de données qualitatives !

2) L'estimation statistique

OBJECTIF

• **Déterminer une <u>grandeur définie sur une population</u> à partir d'observations réalisées sur un <u>échantillon représentatif</u> de cette population.**

DEUX TYPES D'ESTIMATION

- **Estimation ponctuelle : valeur jugée la meilleure à un instant t. Très peu fiable.**

- **Estimation par intervalle : un intervalle de valeurs contenant la valeur recherchée (à un risque d'erreur près) ; c'est ce qu'on appelle l'Intervalle de Confiance (=IC) ou encore Intervalle au risque α** (avec le plus souvent α=5%). **Beaucoup plus fiable.**

METHODOLOGIE

1) Détermination précise de la population à étudier

2) Tirage au sort d'un échantillon représentatif

3) Etude de l'échantillon

4) Extrapolation des résultats à l'ensemble de la population

Estimation
(le plus souvent par intervalle)

<u>Exemple</u> : Une étude vise à déterminer la valeur moyenne de la glycémie dans la population française.

1) Population à étudier : la population française

2) Tirage au sort d'un échantillon A d'effectif connu (50 membres par ex)

3) On effectue les prises de sang et le relevé des valeurs

4) L'extrapolation donne :

- *estimation ponctuelle => 0,95 g/L*
- *estimation par intervalle à 95% => [0,90 – 1,04] g/L*

La valeur moyenne VRAIE de la glycémie d'un membre de la population française a donc de fortes chances d'appartenir à cet IC.

Pour plus de sûreté, on décide de réitérer l'opération sur un nouvel échantillon B tiré au sort, et de même effectif que l'échantillon A. L'extrapolation donne cette fois :

- *estimation ponctuelle => 1,03 g/L*
- *estimation par intervalle à 95% => [0,95 – 1,10] g/L*

On a donc :

	Echantillon A	**Echantillon B**
Estimation ponctuelle	0,95 g/L	1,03 g/L
Estimation par intervalle (95%)	[0,90 g/L ; 1,04 g/L]	[0,95 g/L ; 1,10 g/L]

Deux remarques :

-les estimations ponctuelles sont voisines

-les estimations par intervalle se recouvrent

=> Les résultats semblent se confirmer.

Deux règles primordiales :

-Soient A et B deux échantillons **représentatifs d'une même population**, alors :

- Deux estimations ponctuelles d'une même variable réalisée sur les échantillons A et B donneront des valeurs ponctuelles voisines, mais pas nécessairement la même valeur.

- Deux estimations par intervalle d'une même variable réalisée sur les échantillons A et B donneront des IC se recouvrant, mais pas nécessairement le même IC.

Attention ! Les notions qui suivent sont *SUPER IMPORTANTES !*

ESTIMATION DE DONNÉES QUANTITATIVES

ECHANTILLON

n = effectif

m = moyenne

s = écart type

$\xrightarrow{\text{Estimation}}$

POPULATION TOTALE

N = effectif

μ = moyenne VRAIE

σ = écart type VRAI

$$\mu \in IC_{(1-\alpha)}$$

$$IC_{(1-\alpha)} = [\ m - (\varepsilon.s) / \sqrt{n}\ \ ;\ m + (\varepsilon.s) / \sqrt{n}\ ]$$

α : risque d'erreur

ε : écart réduit => *facteur dépendant du risque α*

s : écart type

Deux valeurs au moins sont à connaître :

-Pour <u>α=5%</u> => <u>ε=1,96</u>
-Pour <u>α=1%</u> => <u>ε=2,6</u>

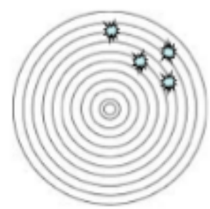
<u>Attention</u> : α et ε varient en SENS INVERSE…

Ce facteur ε va modéliser la LARGEUR ou encore la PRECISION de l'IC :

Rappel de la formule : **$IC_{(1-\alpha)} = [\ m - (\varepsilon.s) / \sqrt{n}\ ;\ m + (\varepsilon.s) / \sqrt{n}\ ]$**

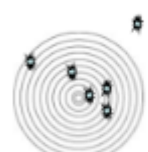
• si le risque α d'erreur diminue => ε augmente => l'IC s'élargit => l'IC contient davantage de valeurs, mais il est moins précis

Large = plus de chances de l'atteindre, mauvaise précision de l'estimation



• si le risque α d'erreur augmente => ε diminue=> l'IC s'affine => l'IC contient moins de valeurs, mais il est plus précis

Resserré = meilleure précision de l'estimation



<u>Exemple :</u> Une étude vise à déterminer la valeur moyenne de la glycémie dans la population française. Pour cela, on dispose d'un échantillon représentatif de 100 personnes tirées au sort dans la population. On choisit le risque tel que α=5%. Les résultats sont consignés :

α	**5,00%**	**1,00%**
m	*m= 0,96 g/L*	*m=0,96 g/L*
s	*s=0,5*	*s=0,5*
$IC_{(1-\alpha)}$	$IC_{95\%}$ *= [0,862-1,058 g/L]*	$IC_{99\%}$ *= [0,830-1,090 g/L]*

- *Il y a 95% de chance que μ appartienne à l'intervalle [0,862-1,058 g/L]*
- *Il y a 99% de chance que μ appartienne à l'intervalle [0,830-1,090 g/L]*
  <u>=> moins précis !</u>

INDICE DE PRÉCISION DE L'ESTIMATION DE DONNÉES QUANTITATIVES

$$i = (\varepsilon . s) / \sqrt{n}$$

La précision est d'autant plus grande que i est faible.

On peut en effet dire que :

$$IC_{(1-\alpha)} = [\ m - (\varepsilon.s) / \sqrt{n}\ ;\ m + (\varepsilon.s) / \sqrt{n}\ ] = [\ m - i\ ;\ m + i\ ]$$

$\Rightarrow$ **La PRECISION et la TAILLE de l'IC varient bien en SENS INVERSE !**

$\Rightarrow$ **La PRECISION augmente dans le MEME SENS que l'effectif de l'échantillon !**

NB : i = indice de précision de l'estimation

Calcul du nombre de sujets nécessaires pour une précision donnée :

$$n = \varepsilon^2 s^2 / i^2$$

2 QCM !

1) <u>La précision d'un IC augmente lorsque</u> :

a) L'indice de précision i diminue.

b) L'effectif n augmente.

c) L'écart réduit diminue.

d) Le risque α augmente.

e) Aucune de ces réponses n'est juste.

1) **La précision d'un IC augmente lorsque :**

a) L'indice de précision i diminue.

b) L'effectif de l'échantillon n augmente.

c) L'écart réduit ε diminue.

d) Le risque α augmente.

~~**e) Aucune de ces réponses n'est juste**~~**.**

- La précision est d'autant plus grande que i est faible (car lorsque i est faible, la largeur de l'IC diminue donc la précision augmente)
- $i = (\varepsilon . s) / \sqrt{n}$
- α et ε varient en sens inverses

Réponse : abcd

2) Un IC est d'autant plus large que :

a) La moyenne sur laquelle il est centré est faible.

b) La moyenne sur laquelle il est centre est élevée.

c) L'effectif de la population N est faible.

d) L'effectif de l'échantillon n est faible.

e) Aucune de ces réponses n'est juste.

2) Un IC est d'autant plus large que :

- ~~**a) La moyenne sur laquelle il est centré est faible.**~~
- ~~**b) La moyenne sur laquelle il est centre est élevée.**~~
- ~~**c) L'effectif de la population N est faible.**~~
- **d) L'effectif de l'échantillon n est faible.**
- ~~**e) Aucune de ces réponses n'est juste.**~~

$$IC_{(1-\alpha)} = [\ m - (\varepsilon.s) / \sqrt{n}\ ;\ m + (\varepsilon.s) / \sqrt{n}\ ]$$

Réponse : d

Illustration par la Loi Normale ou Loi de Gauss

PRINCIPE DE LA LOI

- **La <u>Loi Normale</u> ou <u>Loi de Gauss</u> permet de modéliser des phénomènes naturels issus de plusieurs événements aléatoires.**

- On retrouve ainsi sur une courbe en cloche :

  -la notion d'IC autour de la moyenne μ

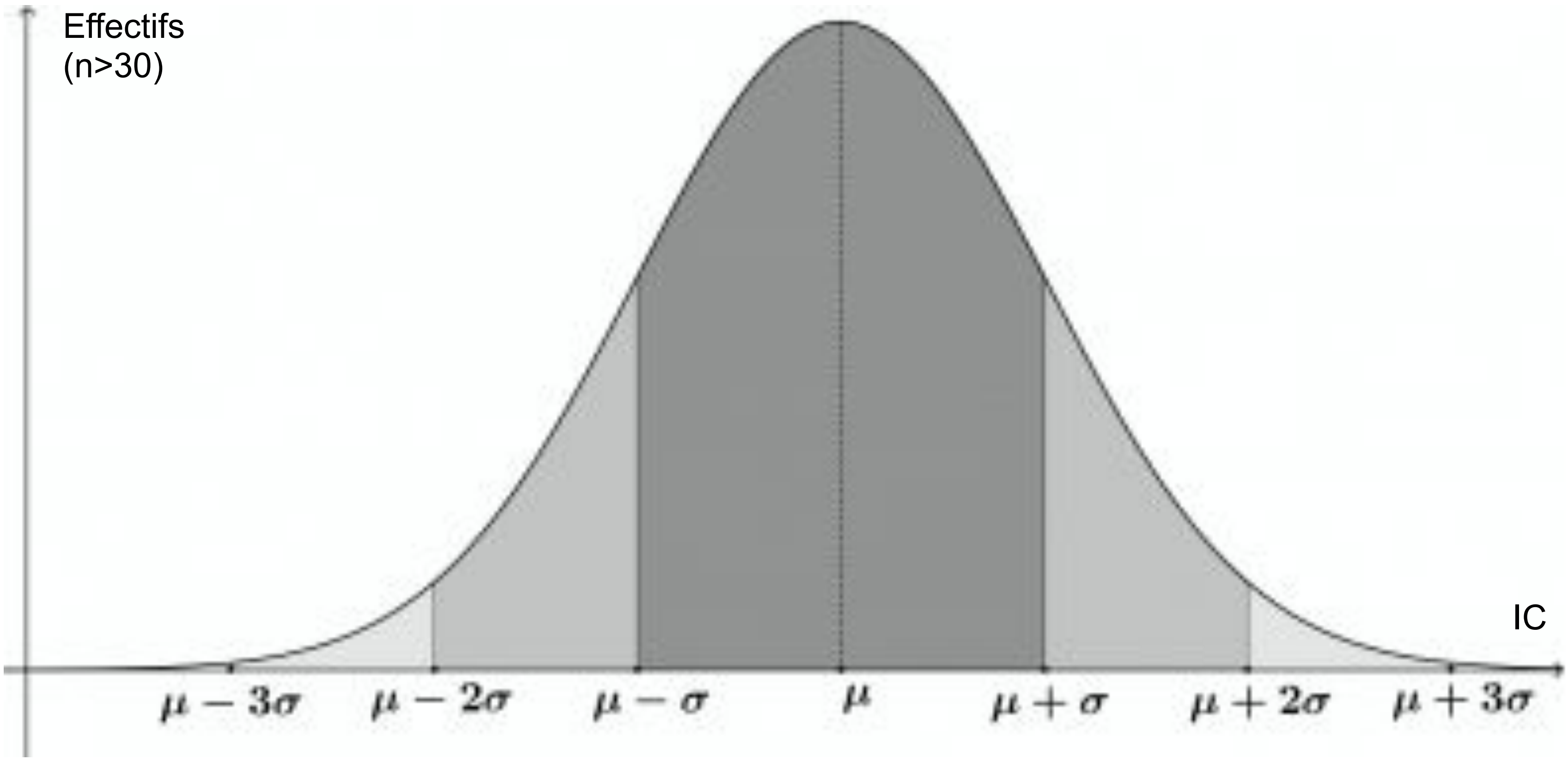
  -la notion d'écart type σ

  -la notion de dispersion autour de cette valeur moyenne

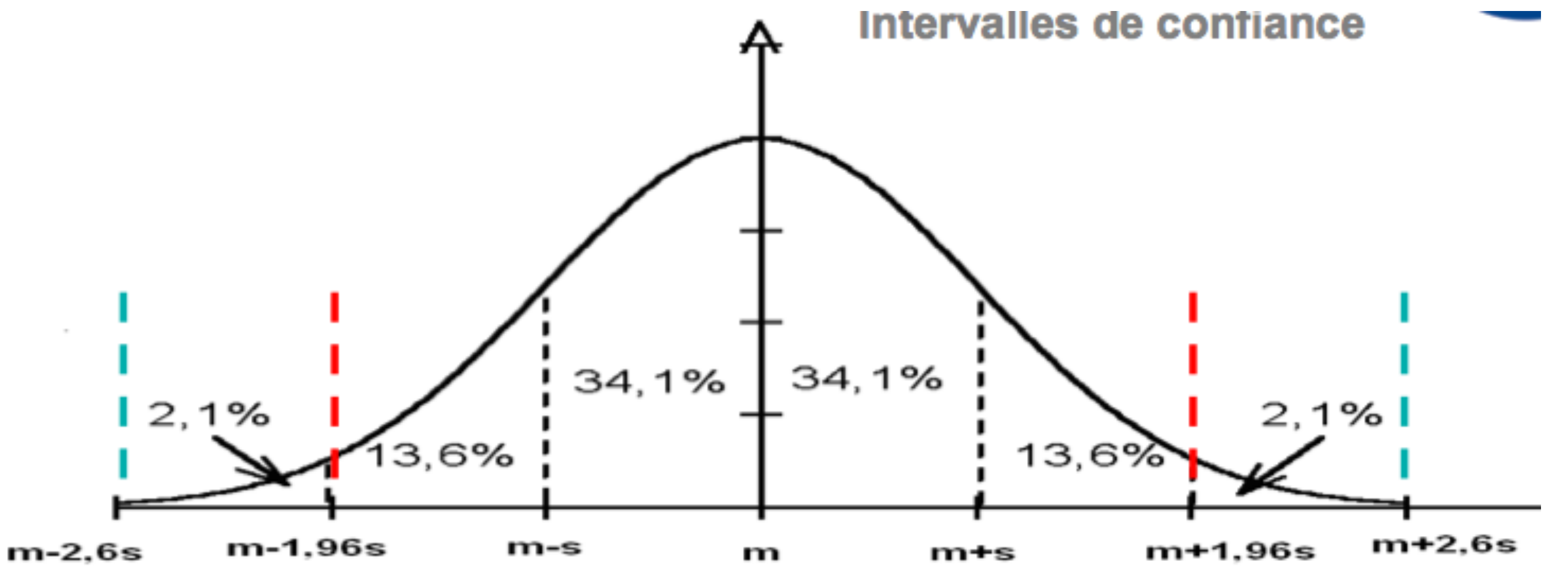
Pré-requis INDISPENSABLE :

La Loi Normale (Loi de Gauss) n'est applicable que sur des échantillons dont l'effectif est SUPERIEUR OU EGAL A 30 MEMBRES.

LA COURBE DE GAUSS

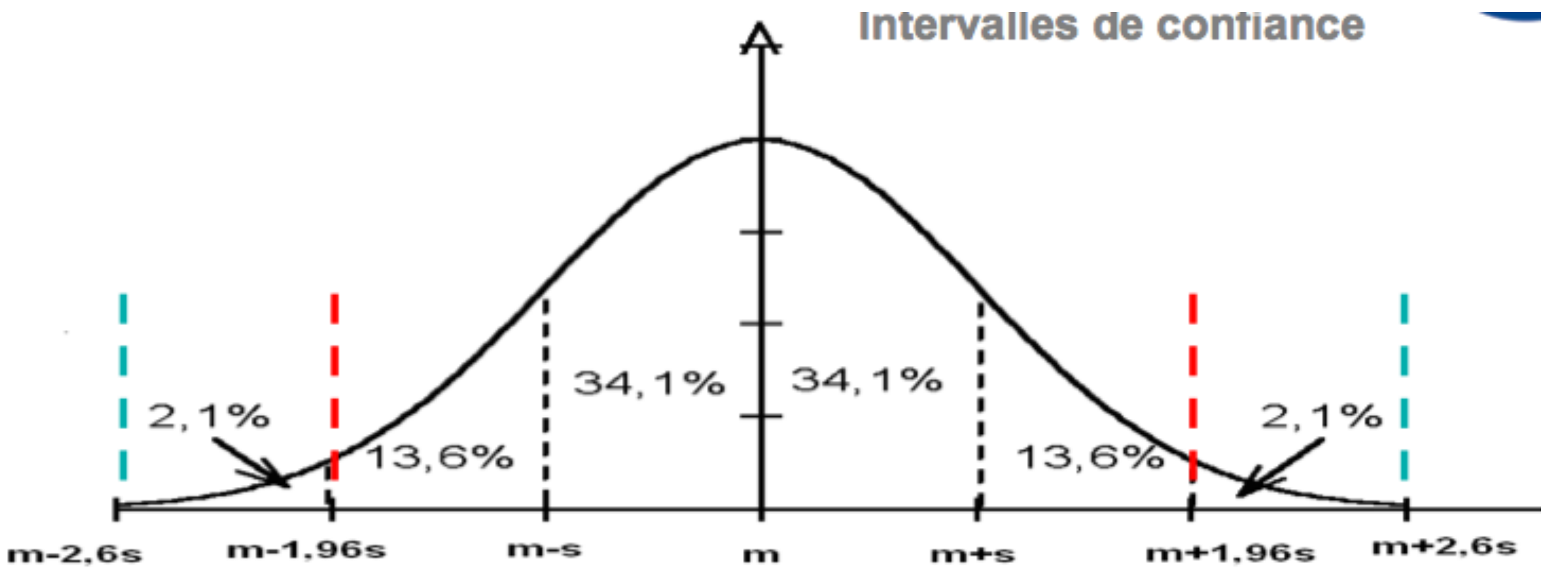


LA COURBE DE GAUSS



La répartition des effectifs est proportionnelle à la surface sous la courbe :

- **IC = [m-1s ; m+1s]** contient **68,2%** de la population
- **$IC_{95\%}$ = [m-1,96 s ; m+1,96s]** contient environ **95,4%** de la population
- **$IC_{99\%}$ = [m-2,6 s ; m+2,6s]** contient environ **99,6%** de la population

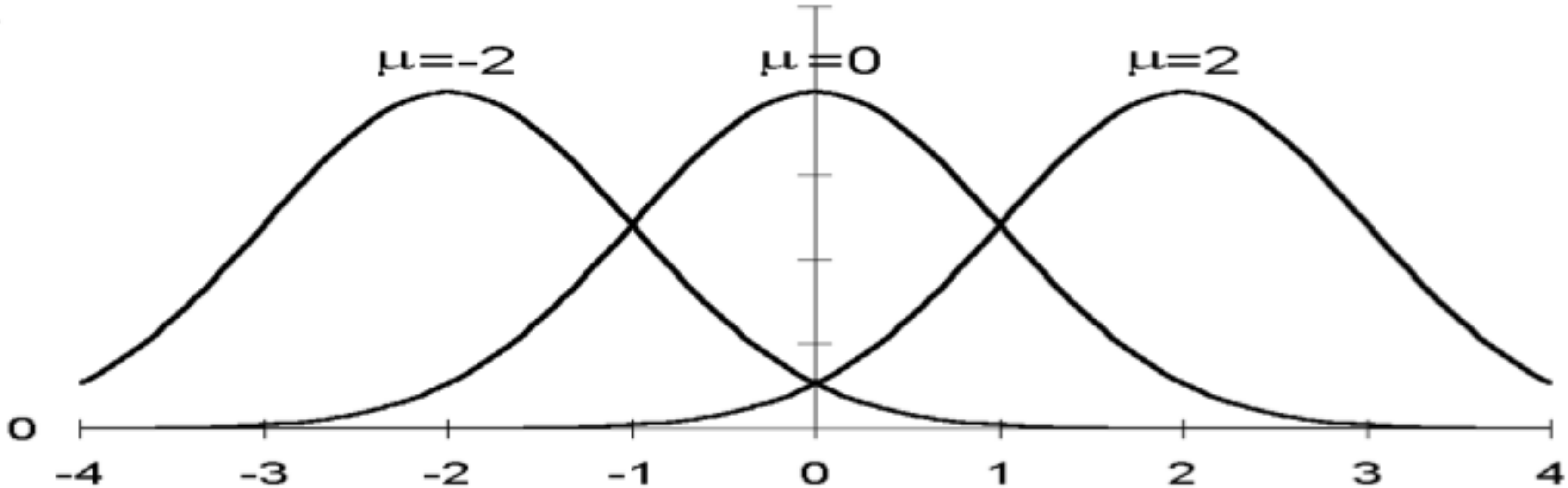


*<u>Exemple</u> : La taille « X » des hommes adultes suit une loi Normale de moyenne **μ = 180 cm** et d'écart type σ **= 6 cm**.*

- *La proportion d'homme dont la taille est **comprise entre** 174 cm (μ-σ) et 186 cm (μ+σ) est de 34,1% + 34,1% = 68,2%.*

- *La proportion d'homme dont la taille est **inférieure** à 168,2 cm (μ - 1,96σ) **ou supérieure** à 191,8 cm (μ + 1,96σ) est de 2,1% + 2,1% = 4,2%*

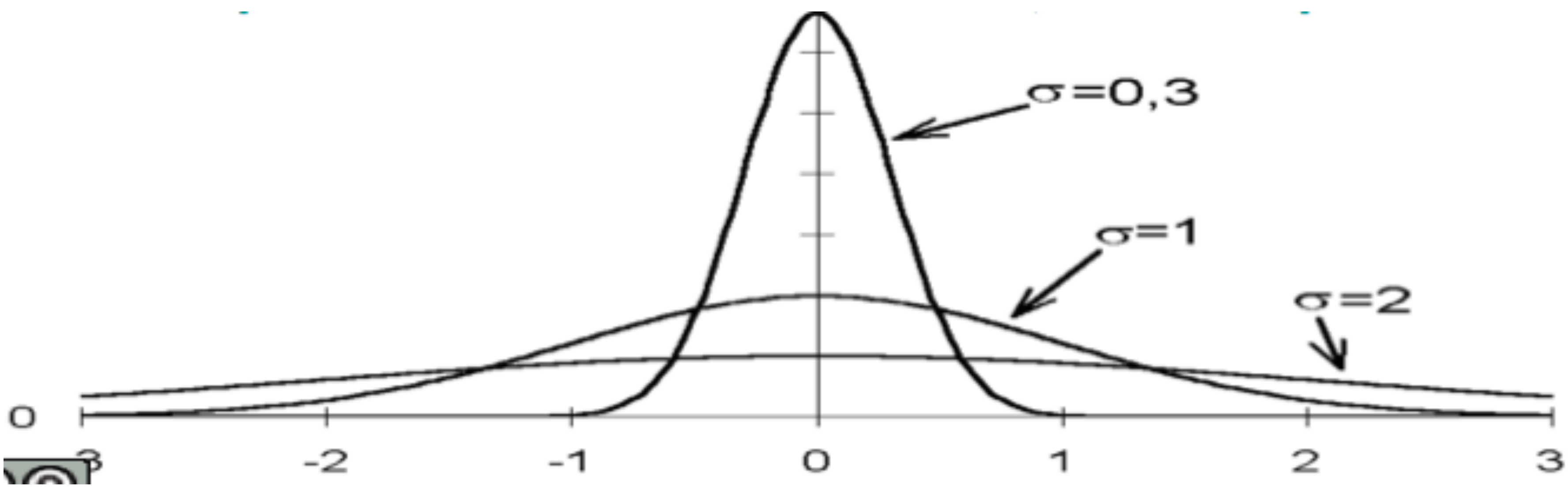
DIFFÉRENTS ASPECTS DE LA COURBE



La valeur de la MOYENNE modifie la POSITION de la courbe en cloche :

- La courbe se déplace vers la droite pour les moyennes élevées
- La courbe se déplace vers la gauche pour les moyennes faibles

DIFFÉRENTS ASPECTS DE LA COURBE



La valeur de l'ECART TYPE modifie la FORME de la courbe en cloche :

- La courbe s'aplatit pour des valeurs dispersées.
- La courbe se resserre pour des valeurs proches.

ESTIMATION DE DONNÉES QUALITATIVES

ECHANTILLON

- n = effectif
- po = pourcentage observé
- s = écart type

Estimation →

POPULATION TOTALE

- N = effectif
- p = pourcentage réel
- σ = écart type VRAI

$$\mu \in IC_{(1-\alpha)}$$

$$IC_{(1-\alpha)} = [\ po - (\varepsilon.s)\ ;\ po + (\varepsilon.s)\ ]$$

$$IC_{(1-\alpha)} = [\ po - \varepsilon.\sqrt{po.qo/n}\ ;\ po + \varepsilon.\sqrt{po.qo/n}\ ]$$

Avec s = $\sqrt{po.qo/n}$ *et* **qo = 1-po**

α : risque d'erreur

ε : écart réduit => <u>*facteur dépendant du risque α*</u>

Rappel :

-Pour $\underline{\alpha=5\%}$ => $\underline{\varepsilon = ?}$

-Pour $\underline{\alpha=1\%}$ => $\underline{\varepsilon = ?}$

Attention : α et ε varient.. comment ?

Rappel :

-Pour <u>α=5%</u> => <u>ε=1,96</u>
-Pour <u>α=1%</u> => <u>ε=2,6</u>

<u>Attention</u> : α et ε varient en SENS INVERSE…

INDICE DE PRÉCISION DE L'ESTIMATION DE DONNÉES QUALITATIVES

$$i = \varepsilon.s = \varepsilon.\ \sqrt{po.qo/n}$$

La précision est d'autant plus grande que i est faible.

On peut en effet dire que :

$IC_{(1-\alpha)} = [\ po - \varepsilon.\sqrt{po.qo/n}\ ;\ po + \varepsilon.\sqrt{po.qo/n}\ ] = [\ po - i\ ;\ po + i\ ]$

$\Rightarrow$ **La PRECISION et la TAILLE de l'IC varient bien en SENS INVERSE !**

$\Rightarrow$ **La PRECISION augmente dans le MEME SENS que l'effectif de l'échantillon !**

NB : i = indice de précision de l'estimation

Calcul du nombre de sujets nécessaires pour une précision donnée :

$$n = \varepsilon^2(po.qo) / i^2$$

<u>Exemple 1 :</u> On interroge un échantillon représentatif ***de 900 personnes*** *au sujet de leur intention de vote à une élection présidentielle opposant le candidat A au candidat B*. ***52% ont déclaré qu'ils voteraient pour A*** *et 48% se prononcent en faveur du candidat B.*

Les journaux ont-ils raison d'affirmer que le candidat A arrive en tête du sondage avec 52% des voix ?

Exemple 1 : *On interroge un échantillon représentatif* ***de 900 personnes*** *au sujet de leur intention de vote à une élection présidentielle opposant le candidat A au candidat B****. 52% ont déclaré qu'ils voteraient pour A*** *et 48% se prononcent en faveur du candidat B.*

Les journaux ont-ils raison d'affirmer que le candidat A arrive en tête du sondage avec 52% des voix ?

- ***po = 52%***
- ***qo = 1-po = 48%***
- ***α = 5%***

<u>Exemple 1 :</u> On interroge un échantillon représentatif ***de 900 personnes*** *au sujet de leur intention de vote à une élection présidentielle opposant le candidat A au candidat B****. 52% ont déclaré qu'ils voteraient pour A*** *et 48% se prononcent en faveur du candidat B.*

Les journaux ont-ils raison d'affirmer que le candidat A arrive en tête du sondage avec 52% des voix ?

- ***po = 52%***
- ***qo = 1-po = 48%***
- ***α = 5%*** *(donc ε = 1,96)*

$$IC_{(1-\alpha)} = [\ po - \varepsilon.\sqrt{po.qo/n}\ ;\ po + \varepsilon.\sqrt{po.qo/n}\ ]$$

<u>Exemple 1 :</u> On interroge un échantillon représentatif ***de 900 personnes*** *au sujet de leur intention de vote à une élection présidentielle opposant le candidat A au candidat B****. 52% ont déclaré qu'ils voteraient pour A*** *et 48% se prononcent en faveur du candidat B.*

Les journaux ont-ils raison d'affirmer que le candidat A arrive en tête du sondage avec 52% des voix ?

- ***po = 52%***
- ***qo = 1-po = 48%***
- ***α = 5%*** *(donc ε = 1,96)*

$$IC_{(1-\alpha)} = [\ po - \varepsilon.\sqrt{po.qo/n}\ ;\ po + \varepsilon.\sqrt{po.qo/n}\ ]$$

$$IC_{(95\%)} = [\ 0,52 - 1,96\sqrt{0,52x0,48/900}\ ;\ 0,52 + 1,96\sqrt{0,52x0,48/900}\ ]$$

*<u>Exemple 1 :</u> On interroge un échantillon représentatif **de 900 personnes** au sujet de leur intention de vote à une élection présidentielle opposant le candidat A au candidat B. **52% ont déclaré qu'ils voteraient pour A** et 48% se prononcent en faveur du candidat B.*

Les journaux ont-ils raison d'affirmer que le candidat A arrive en tête du sondage avec 52% des voix ?

- ***po = 52%***
- ***qo = 1-po = 48%***
- ***α = 5%** (donc ε = 1,96)*

$$IC_{(1-\alpha)} = [\ po - \varepsilon.\sqrt{po.qo/n}\ ;\ po + \varepsilon.\sqrt{po.qo/n}\ ]$$

$$IC_{(95\%)} = [\ 0,52 - 1,96\sqrt{0,52x0,48/900}\ ;\ 0,52 + 1,96\sqrt{0,52x0,48/900}\ ]$$

$$\boldsymbol{IC_{(1-\alpha)} = [\ 0,49\ ;\ 0,55\ ]}$$

ALOOOOOORS ???????? =D

<u>Exemple 1 :</u> On interroge un échantillon représentatif ***de 900 personnes*** *au sujet de leur intention de vote à une élection présidentielle opposant le candidat A au candidat B****. 52% ont déclaré qu'ils voteraient pour A*** *et 48% se prononcent en faveur du candidat B.*

Les journaux ont-ils raison d'affirmer que le candidat A arrive en tête du sondage avec 52% des voix ?

- ***po = 52%***
- ***qo = 1-po = 48%***
- ***α = 5%*** *(donc ε = 1,96)*

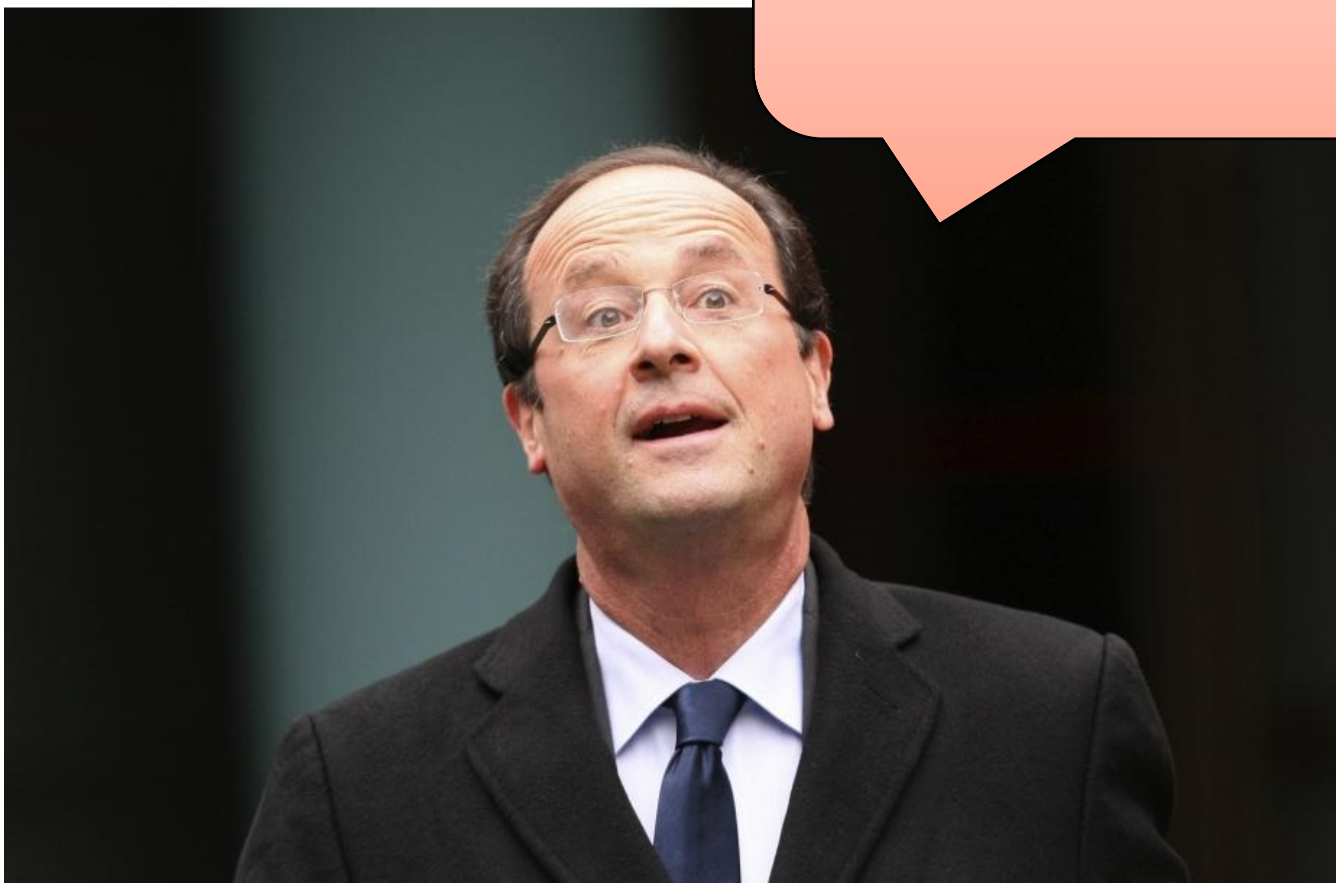
$$IC_{(1-\alpha)} = [\ po - \varepsilon.\sqrt{po.qo/n}\ ;\ po + \varepsilon.\sqrt{po.qo/n}\ ]$$

$$IC_{(95\%)} = [\ 0{,}52 - 1{,}96\sqrt{0{,}52x0{,}48/900}\ ;\ 0{,}52 + 1{,}96\sqrt{0{,}52x0{,}48/900}\ ]$$

$$\boldsymbol{IC_{(1-\alpha)} = [\ 0{,}49\ ;\ 0{,}55\ ]}$$

=> Il est FAUX d'affirmer que le candidat A arrive en tête des sondages !!!!!!

Vous avez dit, “élections présidentielles” ?



Exemple 2 : Un célèbre journal affirme que la côte de confiance du président François Hollande a reculé de 2 points à 27% entre juin et juillet. L'article précise que "le sondage a été réalisé du 27 juin au 1er juillet auprès d'un échantillon de 1.000 personnes représentatif de la population âgée de 18 ans et plus"

Exemple 2 : Un célèbre journal affirme que la côte de confiance du président François Hollande a reculé de 2 points à 27% entre juin et juillet. L’article précise que “le sondage a été réalisé du 27 juin au 1er juillet auprès d'un échantillon de 1.000 personnes représentatif de la population âgée de 18 ans et plus”

Décorticons ces données !

Exemple 2 : *Un célèbre journal affirme que la côte de confiance du président François Hollande a reculé de 2 points à 27% entre juin et juillet. L'article précise que "le sondage a été réalisé du 27 juin au 1er juillet auprès d'un échantillon de 1.000 personnes représentatif de la population âgée de 18 ans et plus"*

Décorticons ces données !

Juin 2013		***Juillet 2013***
• ***po = 0,29*** • ***qo = 1-po = 0,71*** • ***α = 5% (donc ε = 1,96)***	→	• ***po = 0,27*** • ***qo = 1-po = 0,73*** • ***α = 5% (donc ε = 1,96)***
$IC_{(95\%)} = [\ 0{,}26\ ;\ 0{,}\ 32\ ]$		$IC_{(95\%)} = [\ 0{,}24\ ;\ 0{,}\ 30\ ]$

Que peut-on en conclure ?

Exemple 2 : *Un célèbre journal affirme que la côte de confiance du président François Hollande a reculé de 2 points à 27% entre juin et juillet. L'article précise que "le sondage a été réalisé du 27 juin au 1er juillet auprès d'un échantillon de 1.000 personnes représentatif de la population âgée de 18 ans et plus".*

Décorticons ces données !

Juin 2013

- ***po = 0,29***
- ***qo = 1-po = 0,71***
- ***α = 5% (donc ε = 1,96)***

$IC_{(95\%)} = [\ 0,26\ ;\ 0,32\ ]$

→

Juillet 2013

- ***po = 0,27***
- ***qo = 1-po = 0,73***
- ***α = 5% (donc ε = 1,96)***

$IC_{(95\%)} = [\ 0,24\ ;\ 0,30\ ]$

Que peut-on en conclure ?

=> Les IC se recoupent ; la chute de la côte de confiance n'est donc, en réalité, pas du tout mise en évidence ! Le journal se contentera de dire que celle-ci a chuté de deux points (estimation ponctuelle)…

C’est quand même cool la biostat !

“le sondage a été réalisé du 27 juin au 1er juillet auprès d'un échantillon de 1.000 personnes représentatif de la population âgée de 18 ans et plus”

On peut émettre d’autres réserves au sujet de cet article :

-L’effectif de l’échantillon est-il suffisamment grand ?

-Par quel moyen les participants ont-ils été intérrogés (porte à porte, téléphone, internet…) ?

-La période durant laquelle s’est déroulée le sondage permet-elle une représentativité ?

-Et surtout, y’a-t-il eu des non-réponses ? => La plupart du temps, la somme de po et qo ne vaut pas 1 !!

A noter que, le plus souvent, les journalistes ne précisent même pas la nature de l’échantillon sur lequel s’appuie l’enquête…

Une <u>non-réponse</u> à un sondage constitue toujours un <u>biais</u> !

QCM FINAL !

<u>Indiquez lesquelles de ces affirmations sont justes</u> :

a) La formule suivante est adaptée à l'étude des données quantitatives : $IC_{(1-\alpha)} = [\ m - (\varepsilon.s) / \sqrt{n}\ ;\ m + (\varepsilon.s) / \sqrt{n}\ ]$

b) La formule suivante est adaptée à l'étude des données qualitatives : $IC_{(1-\alpha)} = [\ po - (\varepsilon.s)\ ;\ po + (\varepsilon.s)\ ]$

c) La formule suivante est adaptée à l'étude des données qualitatives : $IC_{(1-\alpha)} = [\ po - \varepsilon.\sqrt{po.qo/n}\ ;\ po + \varepsilon.\sqrt{po.qo/n}\ ]$

d) La formule suivante est adaptée à l'étude des données qualitatives : $IC_{(1-\alpha)} = [\ po - \varepsilon.\sqrt{po.qo}/n\ ;\ po + \varepsilon.\sqrt{po.qo}/n\ ]$

e) Aucune de ces affirmations n'est juste.

Indiquez lesquelles de ces affirmations sont justes :

a) La formule suivante est adaptée à l'étude des données quantitatives : $IC_{(1-\alpha)} = [\ m - (\varepsilon.s) / \sqrt{n}\ ;\ m + (\varepsilon.s) / \sqrt{n}\ ]$

b) La formule suivante est adaptée à l'étude des données qualitatives : $IC_{(1-\alpha)} = [\ po - (\varepsilon.s)\ ;\ po + (\varepsilon.s)\ ]$

c) La formule suivante est adaptée à l'étude des données qualitatives : $IC_{(1-\alpha)} = [\ po - \varepsilon.\sqrt{po.qo/n}\ ;\ po + \varepsilon.\sqrt{po.qo/n}\ ]$

~~**d) La formule suivante est adaptée à l'étude des données qualitatives :**~~ $\cancel{IC_{(1-\alpha)} = [\ po - \varepsilon.\sqrt{po.qo}/n\ ;\ po + \varepsilon.\sqrt{po.qo}/n\ ]}$

~~**e) Aucune de ces affirmations n'est juste.**~~

Réponse : abc

Fin de la deuxième partie

Une dernière chose avant un repos bien mérité !!

Il s'agit d'un petit moment de détente avec...

…les petits conseils pour la PACES !

Promis, ça dure moins d’une minute !

1 - Soyez ordonnés !

2 – Ne vous mettez pas en retard…

3 – Ayez l’esprit ouvert.

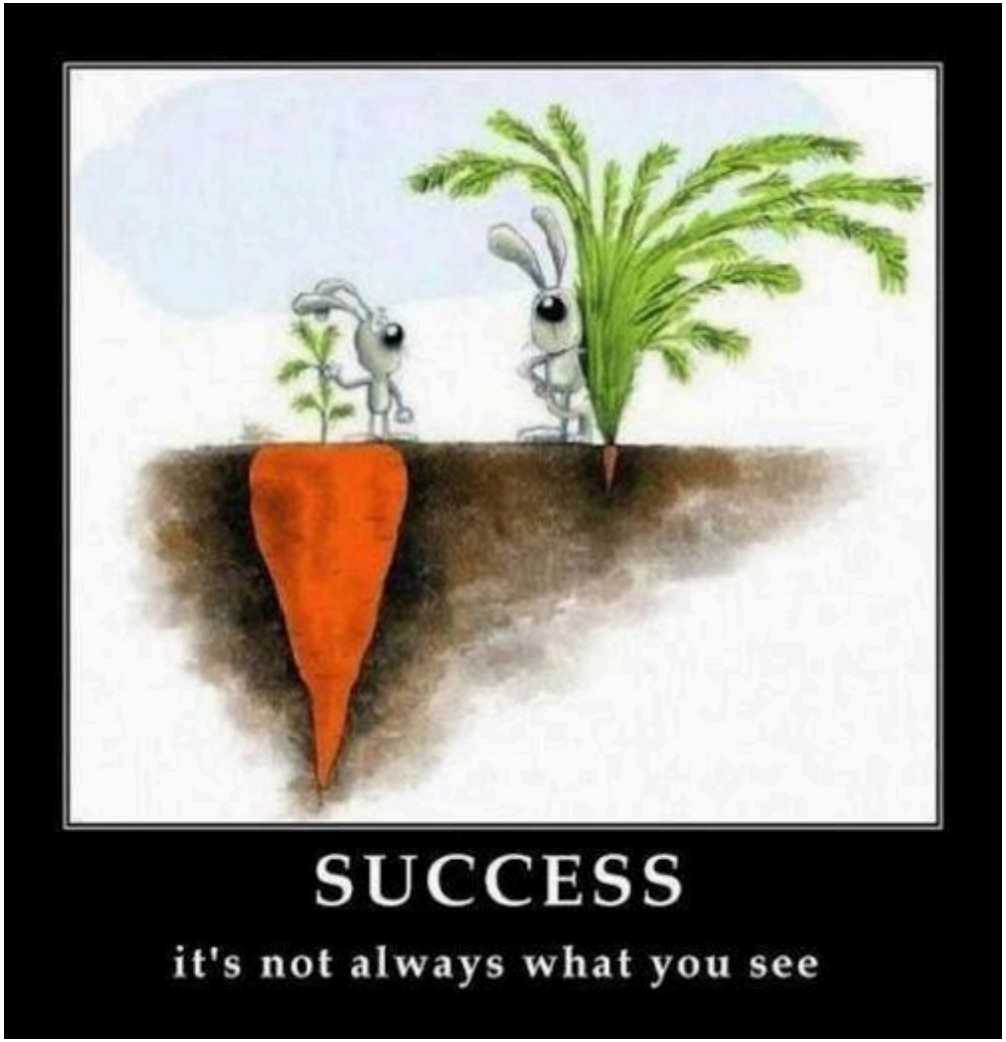
4 – Ne vous laissez pas distraire.

5 – Tenez vous à l’écart des trompeurs.

6 – Posez-vous des questions..

7 – Entraidez-vous !

8 – Gardez le sourire !



9 – Ne doutez pas trop…

10 – Et croyez en vous !


BON COURAGE A TOUS