UE 4:

BIOSTATISTIQUES

VARIABLES ALÉATOIRES ET LOIS DE PROBABILITÉS

(PR STACCINI/SKIINI/P1 CONCENTRÉS)

AVANT TOUT: LES VARIABLES ALÉATOIRES

DEF: UNE VARIABLE ALÉATOIRE EST UNE ÉPREUVE MENANT À DES ÉVÉNEMENTS ÉLÉMENTAIRES QUI SONT DES NOMBRES

NOMBRE QUI VARIE..

- 2 TYPES :

- DISCRÈTES/À VALEURS (DÉNOMBRABLE)
- CONTINUES/À DENSITÉ (INDÉNOMBRABLE).

LES LOIS DE PROBABILITÉS:

1-LES LOIS DE PROBABILITÉS DISCRÈTES (VARIABLES ALÉATOIRES DISCRÈTES)

- LOI DE BERNOUILLI
- LOI BINOMIALE
- LOI DE POISSON
- LOI GEOMETRIQUE
- LOI HYPERGEOMETRIQUE

2-LES LOIS DE PROBABILITÉS CONTINUES (VARIABLES ALÉATOIRES CONTINUES)

- LOI EXPONENTIELLE
- LOI UNIFORME
- LOI NORMALE/LOI NORMALE CENTRÉE RÉDUITE

3-APPROXIMATIONS

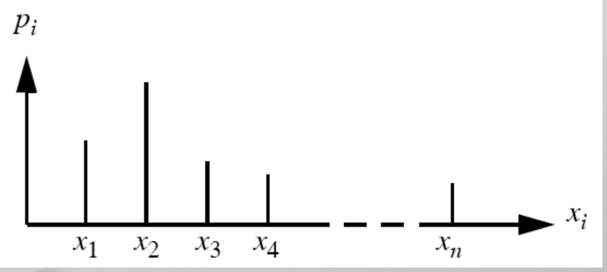
LOIS DE PROBABILITÉS DISCRÈTES

Une loi de probabilités d'une variable aléatoire X discrète est définie en donnant toutes les probabilités pi correspondante aux différentes éventualités xi.

$0 \leq pi \leq 1$ $\qquad$ $\sum_{0 \leq i \leq n} (pi) = 1$

<u>Représentation en tableau ou par diagramme en bâtons :</u>

x1	x2	x3	…	xn	…
p1	p2	p3	…	pn	…



DEFINITIONS

MOYENNE / ESPÉRANCE :

INDICATEUR DE **POSITION**, TRADUIT LA TENDANCE CENTRALE DE LA VARIABLE

$$E(X) = \mu = x1p1 + x2p2 + \ldots + xnpn = \Sigma(XI\ PI)$$

THEOREMES DE L'ESPÉRANCE : (k est un nombre réel, X est une variable aléatoire)

$$E(kX) = k\ E(X) \qquad E(X + k) = E(X) + k$$

$$E\left(\sum_{0\leq i\leq n} x_i\right) = \sum_{0\leq i\leq n}(E(x_i)) \qquad E(X+Y) = E(X) + E(Y)$$

VARIANCE ET ÉCART TYPE :

C'est un indicateur de dispersion.

On note la variance σ^2 et l'écart type σ est la racine carré de la variance.

σ^2 = Var (X) = $\sum_{0 \leq i \leq n} p_i$ (xi - μ)²

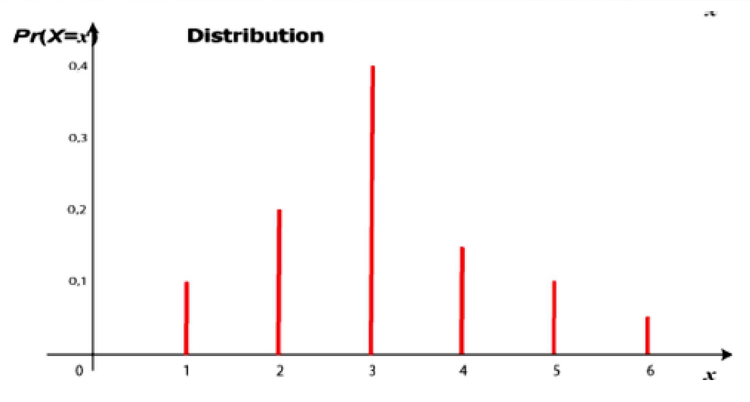
Pour « a » un réel quelconque, on aura :

Var (X+a) = Var (X)

Var (aX) = a² Var (X)

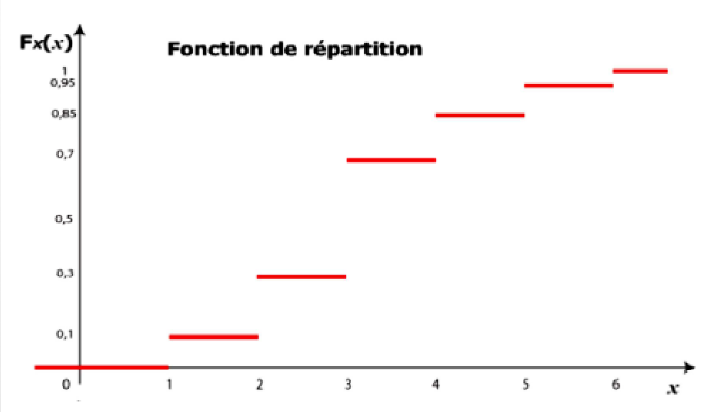
FONCTION DE RÉPARTITION/DISTRIBUTION

- FONCTION DE DISTRIBUTION :



- FONCTION DE RÉPARTITION:

Fonction **monotone croissante**



LOI DE BERNOULLI

<u>Paramètres:</u>

p : probabilité d'un succès

q : Echec : 1-p.

$\mu = p$

$\sigma^2 = pq$

<u>Formules :</u> $P(X=1) = p$ et $P(X=0) = q$

QCM BERNOULLI !

- **<u>QCM :</u> Je lance un dés a six faces, je cherche à avoir un chiffre impair, je suis donc une loi de Bernoulli de paramètres:**

A) $p = 7/14$

B) $q = 1/6$

C) $\mu = 0,5$

D) $\sigma^2 = 0,25$

E) Les propositions A, B, C et D sont fausses.

QCM BERNOULLI !

- **<u>QCM :</u> Je lance un dés a six faces, je cherche à avoir un chiffre impair, je suis donc une loi de Bernoulli de paramètres:**

A) $p = 7/14$

B) $q = 1/6$

C) $\mu = 0,5$

D) $\sigma^2 = 0,25$

E) Les propositions A, B, C et D sont fausses.

LOI BINOMIALE

C'est un enchainement de loi de Bernoulli :

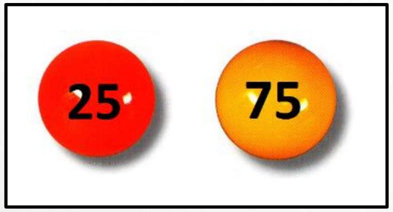
P (X = k) = $C_n^k \; p^k \; q^{n-k}$

0 $\leq$ k (= Nombre de succès) $\leq n$ (Nb d'épreuve répétée de Bernoulli)

- μ = np

- σ^2 = npq

LOI BINOMIALE



1[er] cas : Je tire 5 boules, j'en veux 4 rouges… Ici n=5
Donc n/N = 5/100 = 0,05
⇒ Je peux négliger la difference de proba entre chaque tirage donc loi binomiale **possible**

2[ème] cas : Je tire 20 boules, j'en veux 18 rouges. Au depart, on a une chance sur 4 d'avoir une rouge. Quand on en a tiré déjà 15, plus qu'une chance sur 8,5… on ne peut pas négliger la différence, donc loi hypergéométrique. n/N= 20/100= 0,2

Pour les qcms : Si le tirage est exhaustif (pas de remises), on définit le taux de sondage : n/N.
Si $n/N \geq 0{,}10$ *on utilisera la loi Hypergéometrique* N= population totale
Si n / N ≤ 0,10, on peut appliquer la loi Binomiale

QCM LOI BINOMIALE

<u>QCM:</u> La probabilité pour que Philippe, tuteur en histologie, fasse un qcm non ambiguë est de 0,15. Quelle est la probabilité pour que dans 10 qcms, il y ai 6 qcms ambiguës?

A) $0{,}15^4$

B) $C_{10}^{4}\ 0{,}15^4\ 0{,}85^6$

C) $C_{10}^{6}\ 0{,}85^6\ 0{,}15^4$

D) $0{,}15^4 \times 0{,}85^6$

E) Les propositions A, B, C et D sont fausses.

QCM LOI BINOMIALE

<u>QCM:</u> La probabilité pour que Philippe, tuteur en histologie, fasse un qcm non ambiguë est de 0,15. Quelle est la probabilité pour que dans 10 qcms, il y ai 6 qcms ambiguës?

A) $0{,}15^4$

B) $C_{10}^{4}\ 0{,}15^4\ 0{,}85^6$

C) $C_{10}^{6}\ 0{,}85^6\ 0{,}15^4$

D) $0{,}15^4 \times 0{,}85^6$

E) Les propositions A, B, C et D sont fausses.

LOI DE POISSON

λ : Nombre moyen d'événements survenant dans une période

k : nombre de succès.

$$\mathbf{P(X = k) = \frac{e^{-\lambda}\lambda^{k}}{k!}}$$

$\mu = \lambda$

$\sigma^2 = \lambda$

QCM POISSON

<u>QCM:</u> Jean est a la gare. Il y a en moyenne 3 trains qui passent par heure. Quelle est la probabilité pour que Jean ne voit qu'un seul train passer en 2 heures ?

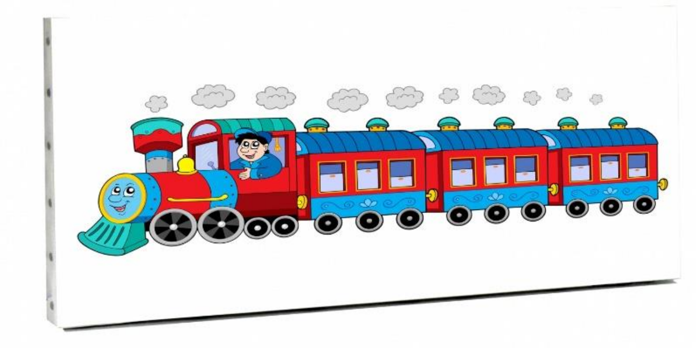
A) $3e^{-3}$

B) $\frac{e^{-3}\,3^1}{1!}$

C) $\frac{e^{-6}\,6^1}{1!}$

D) $\frac{e^{-1}\,1^6}{6!}$

E) $6e^{-6}$



QCM POISSON

<u>QCM:</u> Jean est a la gare. Il y a en moyenne 3 trains qui passent par heure. Quelle est la probabilité pour que Jean ne voit qu'un seul train passer en 2 heures ?

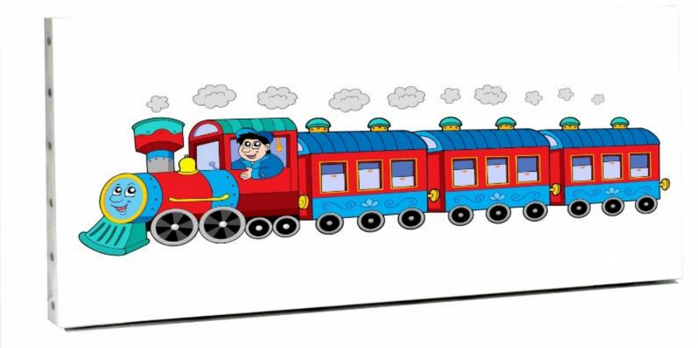
A) $3e^{-3}$

B) $\frac{e^{-3}\,3^{1}}{1!}$

C) $\frac{e^{-6}\,6^{1}}{1!}$

D) $\frac{e^{-1}\,1^{6}}{6!}$

E) $6e^{-6}$



LOI GÉOMÉTRIQUE

C'est une succession d'épreuves de Bernoulli.

k = Nombre d'essais nécessaire pour arriver a un succès. (donc >0)

$\mu = \frac{1}{p}$

$\sigma^2 = \frac{1-p}{p^2}$

$$\mathbf{P(X = k) = p}\boldsymbol{q^{k-1}}$$

QCM LOI GÉOMÉTRIQUE

<u>QCM:</u> Jean est toujours a la gare. La probabilité qu'il y reste encore est de 1/3 par heure. Quelle est la probabilité pour que Jean parte lors de la troisième heure?

A) 4/27

B) 2/27

C) 1/27

D) 1/9

E) 1/3

QCM LOI GÉOMÉTRIQUE

<u>QCM:</u> Jean est toujours a la gare. La probabilité qu'il y reste encore est de 1/3 par heure. Quelle est la probabilité pour que Jean parte lors de la troisième heure?

A) 4/27

B) 2/27

C) 1/27

D) 1/9

E) 1/3

$1/3 \times 1/3 \times 2/3 = 2/27$

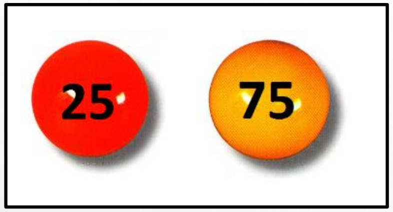
LOI HYPERGEOMETRIQUE

$\mu = \frac{nD}{N}$ $\sigma^2 = \frac{nD}{N} \times \frac{N-D}{N} \times \frac{N-n}{N-1}$

Sans remise.

On utilise la loi hypergeometrique lorsqu'on veut connaitre la probabilité d'obtenir « k » individus présentant ce caractère dans un échantillon de « n » individus, issu de la population « N ». **« D » étant le nombre d'individus présentant ce caractère dans « N »**

$$\mathbf{P(X = k) = \frac{C_D^k \times C_{N-D}^{n-k}}{C_N^n}}$$



Je tire 20 boules, j'en veux 18 rouges. Au depart, on a une chance sur 4 d'avoir une rouge.

<u>Ici :</u>

k= 18

D=25

n=20

N=100

$$\mathbf{P(X = k) = \frac{C_D^k \times C_{N-D}^{n-k}}{C_N^n}}$$

QCM LOI HYPERGEOMETRIQUE

<u>QCM:</u> Dans un zoo, il y a 15 animaux. 6 d'entre eux sont des mammifères, 4 sont des reptiles et 5 des poissons. Quelles est la probabilité en prenant 4 animaux au hasard, que 2 d'entre eux soient des poissons? (Rappel EXCEPTIONNEL: $C_n^k = \frac{n!}{k!(n-k)!}$)

A) 8/27

B) $\frac{C_5^2 \times C_{10}^2}{C_{15}^4}$

C)30/273

D) $\frac{C_4^2 \times C_{11}^2}{C_{15}^4}$

E) 30/91

QCM LOI HYPERGEOMETRIQUE

<u>QCM:</u> Dans un zoo, il y a 15 animaux. 6 d'entre eux sont des mammifères, 4 sont des reptiles et 5 des poissons. Quelles est la probabilité en prenant 4 animaux au hasard, que 2 d'entre eux soient des poissons? (Rappel EXCEPTIONNEL: $C_n^k = \frac{n!}{k!(n-k)!}$)

$$P(X = k) = \frac{C_D^k \times C_{N-D}^{n-k}}{C_N^n}$$

A) $8/27$

B) $\frac{C_5^2 \times C_{10}^2}{C_{15}^4}$

$$\frac{C_5^2 \times C_{10}^2}{C_{15}^4} = \frac{\frac{5*4}{2*1} \times \frac{10*9}{2*1}}{\frac{15*14*13*12}{4*3*2*1}}$$

C) $30/273$

D) $\frac{C_4^2 \times C_{11}^2}{C_{15}^4}$

$$= \frac{5*2*5*3*3}{15*7*13}$$

E) 30/91

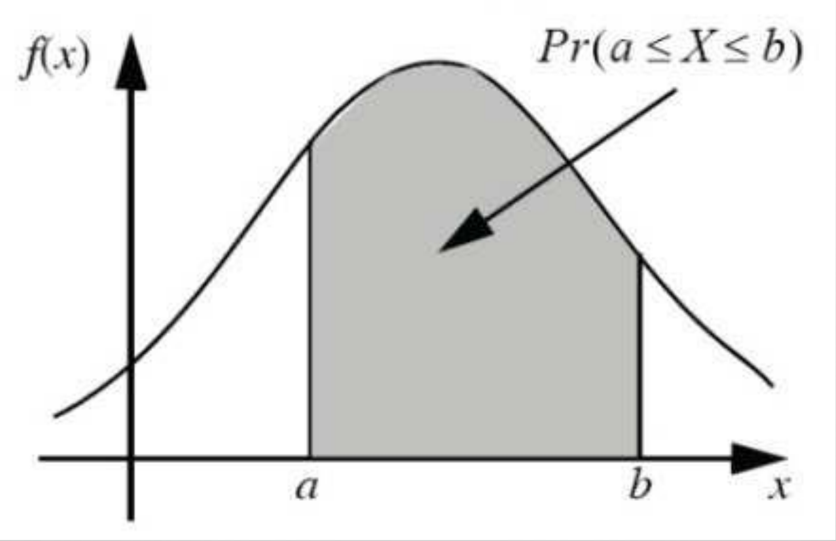
$$= \frac{5*2*3}{7*13}$$

<u>=30/91</u>

LES LOIS DE PROBABILITÉS CONTINUES

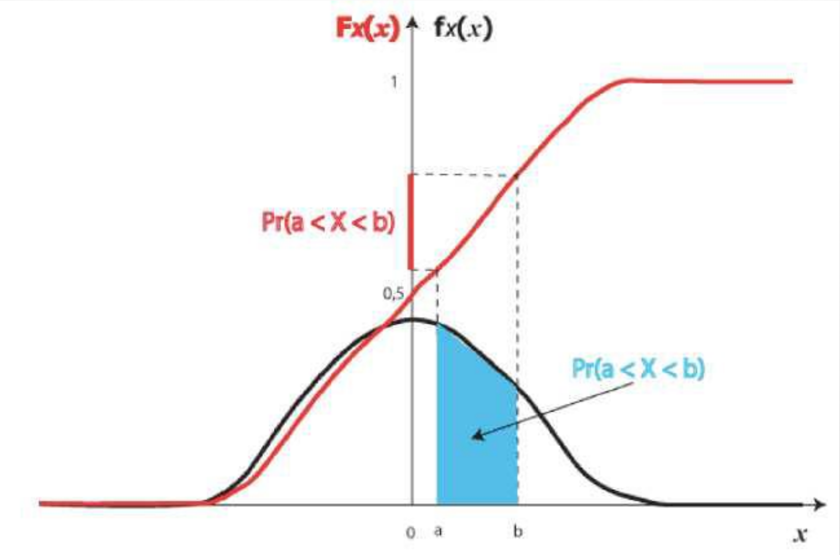
FONCTION DE DENSITÉ

De la même manière que la fonction de distribution pour les variables discretes, la fonction de densité **permet de distribuer** les probabilités pour les variables continues. On parle de densité, car on prend des intervalles.



FONCTION DE RÉPARTITION:

LA FONCTION DE RÉPARTITION F(X) « EST L'INTÉGRALE » DE LA FONCTION DE DENSITÉ. MONOTONE CROISSANTE.



LOI EXPONENTIELLE

λ : Taux de défaillance instantané (par unité de temps)

La loi exponentielle est utilisée pour décrire un phénomène de « mortalité » (ou survenue d'un événement).

$\mu = \frac{1}{\lambda}$ $\sigma^2 = \frac{1}{\lambda^2}$

<u>nota :</u> *lien avec la loi de poisson : si un évènement se réalise selon une loi de poisson de paramètre* ***λ****, alors le temps entre 2 réalisations consécutives de l'événement est distribué selon une loi exponentielle de paramètre* ***1/λ.***

$$\boldsymbol{P(X \leq x) = \int_0^x \lambda e^{-\lambda t}\, dt}$$

$$\mathbf{= 1-} \, \boldsymbol{e^{-\lambda x}}$$

QCM LOI EXPONENTIELLE

<u>QCM:</u> La durée de vie d'un ordinateur suit une loi exponentielle de paramètre λ = 1 / 10 (l'unité de temps est l'année). Quelle est la probabilité qu'il ne fonctionne plus avant 5 ans d'utilisation ?

A) $\frac{e^{-5}\, 5^{0,1}}{(\frac{1}{10})!}$

B) e^{-5}

C) 1-e^{-5}

D) 1-$e^{-0,5}$

E) $e^{-0,5}$

QCM LOI EXPONENTIELLE

<u>QCM:</u> La durée de vie d'un ordinateur suit une loi exponentielle de paramètre λ = 1 / 10 (l'unité de temps est l'année). Quelle est la probabilité qu'il ne fonctionne plus avant 5 ans d'utilisation ?

A) $\frac{e^{-5}\,5^{0,1}}{(\frac{1}{10})!}$

B) e^{-5}

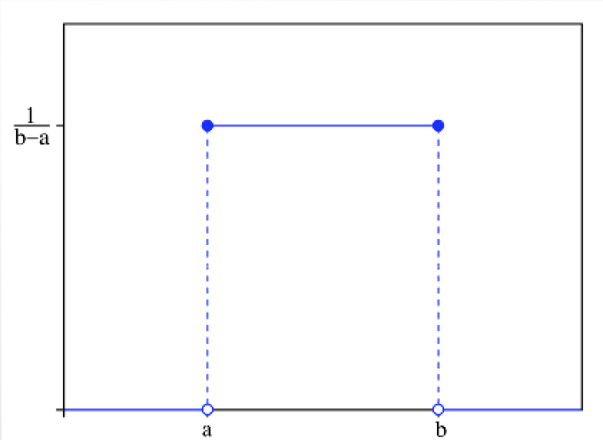
C) 1-e^{-5}

D) 1-$\boldsymbol{e^{-0,5}}$

E) $e^{-0,5}$

LOI UNIFORME

Définition : on utilise une loi uniforme lorsque entre un point *a* et un point *b*, la densité de probabilité est toujours égale entre ces deux points et nulle ailleurs.



$$\mathbf{P(X \leq x) = \int_a^x \frac{1}{b-a} dx}$$

$\mu = (a+b) / 2$

$$\sigma^2 = \frac{(b-a)^2}{2}$$

Une pizza doit arriver entre 12 et 13 heure. Si on suppose que la probabilité est uniforme, alors quelle est la probabilité que la pizza arrive entre 12h et 12h12?

1 chance sur 5 !

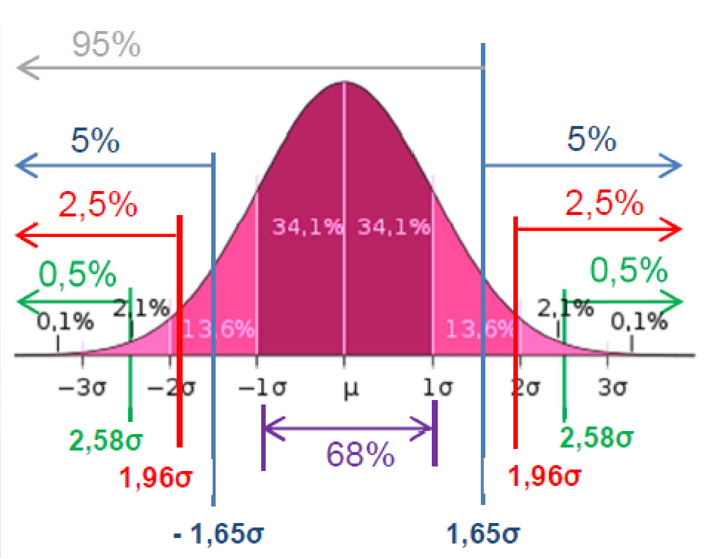
LOI NORMALE = N(μ,σ)

La courbe représentative de **la fonction de densité** est appelée courbe de gauss.

Il faut une grande quantité de population pour obtenir cette courbe. (Beaucoup d'epreuves)
Elle permet de représenté les phénomenes naturels :

Si les étudiantes en médecine font en moyenne 55 kg et que l'écart type vaut 2,5 il y a alors 95% des étudiantes qui pèsent entre 50 et 60 kg

- $p(x < \mu - 1{,}65\ \sigma \text{ ou } x > \mu + 1{,}65\ \sigma) = 10\%$
- $\underline{p(x < \mu - 1{,}96\ \sigma \text{ ou } x > \mu + 1{,}96\ \sigma) = 5\%}$
- $\underline{p(x < \mu - 2{,}58\ \sigma \text{ ou } x > \mu + 2{,}58\ \sigma) = 1\%}$
- $p(x < \mu - 3{,}30\ \sigma \text{ ou } x > \mu + 3{,}30\ \sigma) = 0{,}1\%$
- $p(\mu - \sigma < x < \mu + \sigma) = 68\%$



VARIABLE CENTRÉE RÉDUITE

« Centrée »: consiste à soustraire la moyenne« µ »de la variable à chacune de ses valeurs initiales

« Réduite »: consiste à diviser toutes les valeurs que prend la variable par son écart-type « σ »

Soit X variable aléatoire de moyenne µ et d'écart type σ. On définit la variable centrée réduite Y :

$$\mathbf{Y} = \frac{\boldsymbol{X} - \mu}{\boldsymbol{\sigma}}$$

2 propriétés de la « Variable centrée réduite » : **E(Y) = 0 et Var (Y) = 1 → σ = 1**

Indépendante des unités

LOI NORMALE CENTRÉE RÉDUITE   R(0;1)

La loi Normale centrée réduite est un cas particulier de la loi Normale, où μ = 0 et σ = 1.

En effet, il existe une « table de la loi normale centrée réduite » qui donne la probabilité pour que la variable « Z » soit inférieure à z.

$$\mathbf{Z} = \frac{\mathbf{X} - \boldsymbol{\mu}}{\boldsymbol{\sigma}}$$

/!\ La table de la loi normale centrée réduite donne

les probabilités de Z pour $0 \leq z \leq 3,999$

La table de la loi normale centrée réduite

On Prend la variable poids des étudiantes en médecine. On cherche a savoir la proportion d'étudiante donc le poids est **supérieur** à 56,4kg. On modifie la variable en variable centrée réduite grâce a notre formule:

$$Z = \frac{56{,}4 - 55}{2{,}5} = 0{,}56$$

	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549

La table de la loi normale centrée réduite

$$Z = \frac{56{,}4-55}{2{,}5} = 0{,}56$$

	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549

Donne la probabilité **pour que la variable Z soit inférieure à z**

On a donc 71,23% des étudiantes qui pesent moins de 56,4 kilos, et 1-0,7123 soit 28,77% d'étudiantes en medecine ayant un poids **superieur** a 56,4 kilos.

APPROXIMATIONS !

loi binomiale → loi de poisson **B(n,p) → P(λ = np)**

<u>3 conditions :</u>

- n ≥ 50

- p ≤ 0,1

- np ≤ 5

Je décide de piocher puis de remettre dans le paquet une carte 100 fois. Si j'obtiens le deux de carreau, je considère que je gagne, sinon je considère que je perds.

Ici n = 100 donc >50 ; p<0,1 et np < 5 donc je peux appliquer la loi de Poisson telle que P(λ = $\frac{1}{52} \times 100$)

Elles doivent <u>toutes</u> être remplies !!

APPROXIMATIONS !

Loi Binomiale → Loi Normale **$B(n,p) \rightarrow N(np, \sqrt{npq})$**

2 conditions :

- $np \geq 5$

- $nq \geq 5$

APPROXIMATIONS !

Loi de poisson → Loi normale **$P(\lambda) \rightarrow N(\lambda, \sqrt{\lambda})$**

Une seule condition :

- **$\lambda > 25$**

FINI !!!!