Statistiques et probabilités
Cours n°4

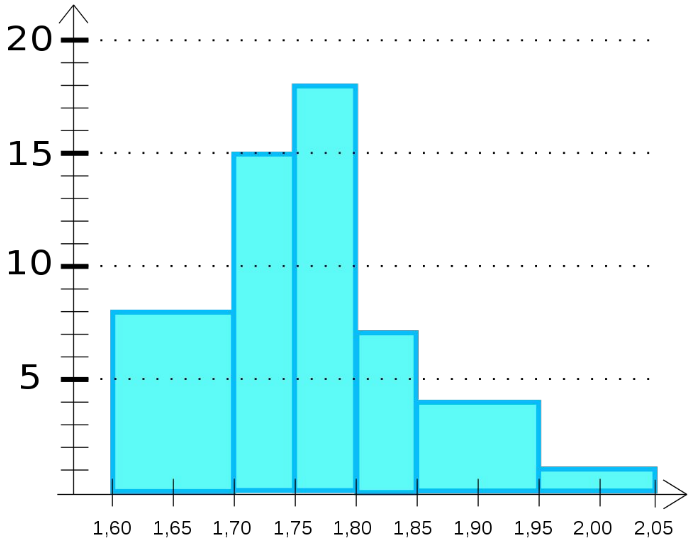
Guillaume Postic

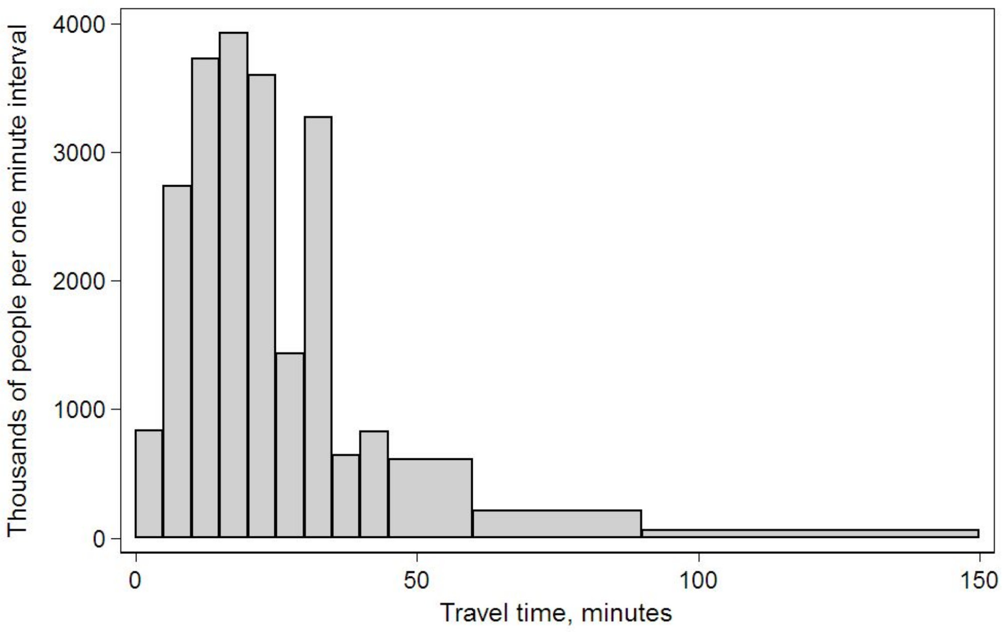
Université Paris-Saclay, Univ. Evry
Département informatique

Master 1 MIAGE - 2022/2023

Histogramme

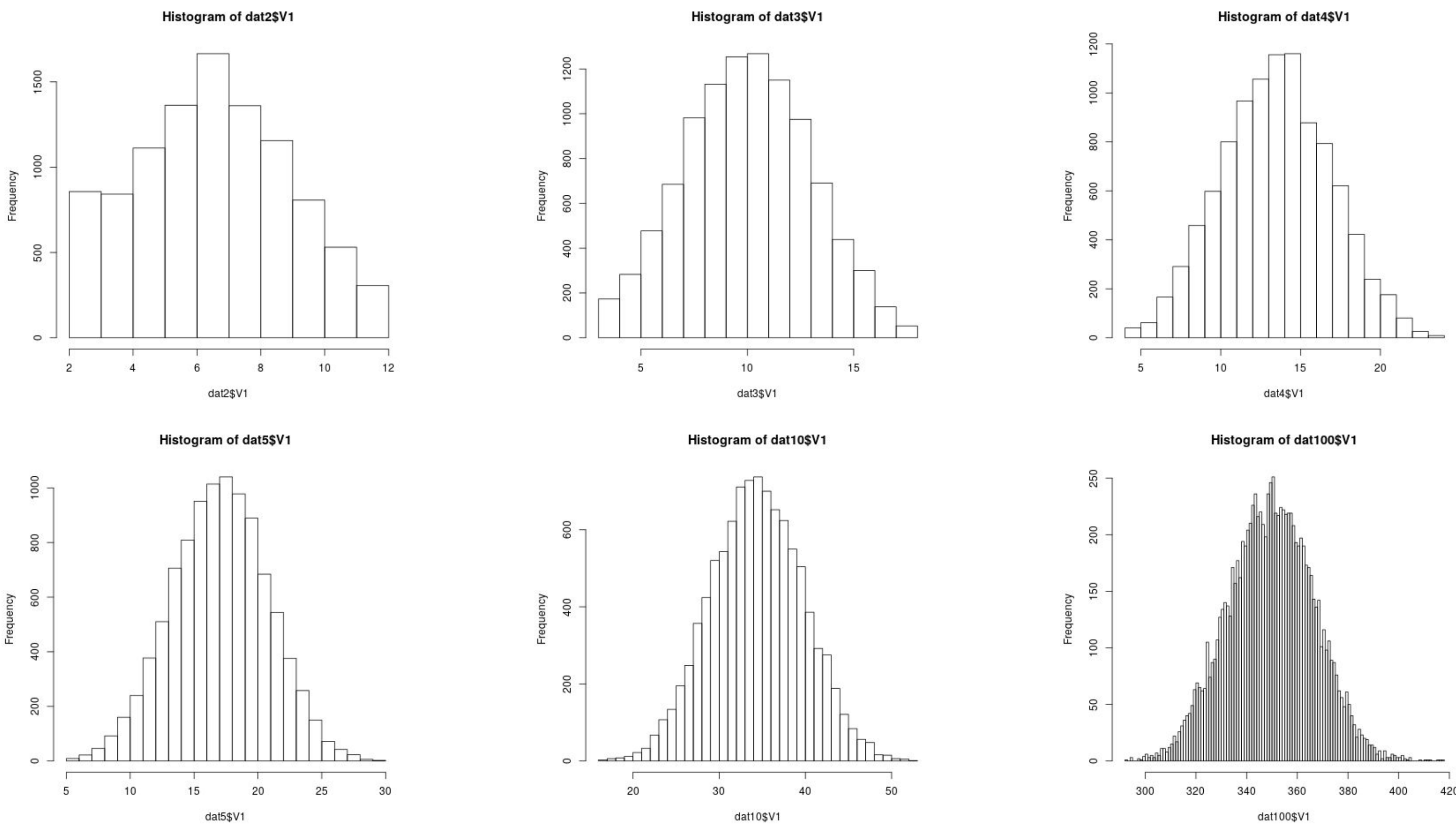
Représentation graphique approximant la distribution d'une variable aléatoire par groupement des données en classes représentées par des colonnes.





Théorème de la limite centrale (1)

On lance 10 000 fois *n* dés équilibrés à 6 faces (loi uniforme) ; à chaque lancer, on calcule **la somme *S* des *n* dés**. Ci-dessous les représentations par histogrammes des distributions de *S*, pour des valeurs de *n* de 2, 3, 4, 5, 10 ou 100 :



Théorème de la limite centrale (2)

Le **théorème de la limite centrale** établit la convergence vers la loi normale de la somme d'une suite de variables aléatoires.

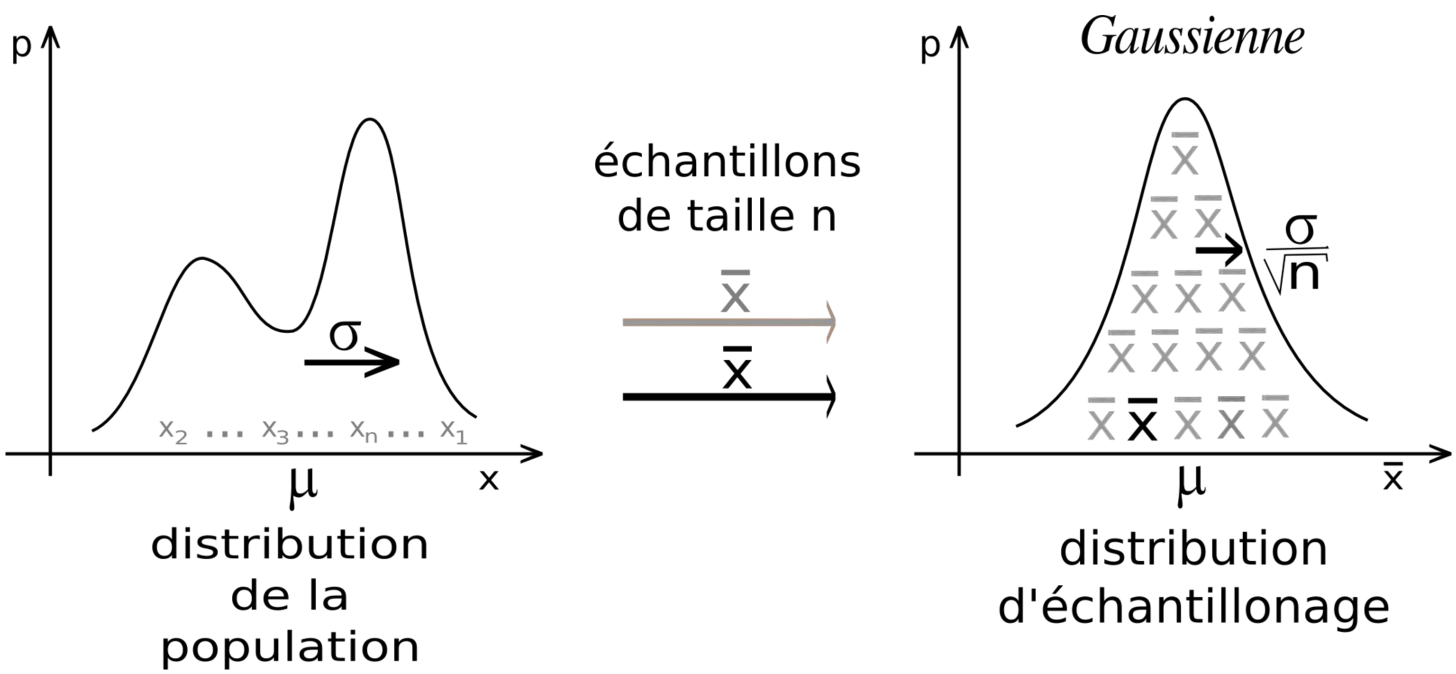
Conditions sur les variables aléatoires :

- Indépendantes
- Identiquement distribuées (c.-à-d. de même loi)
- Suite suffisamment longue (n > 20 ou 30, selon les auteurs)

<u>Note</u> : les conditions de Liapounov et Lindeberg permettent de supprimer l'hypothèse selon laquelle les variables aléatoires sont de même loi.

Théorème de la limite centrale (3)

Le théorème de la limite centrale s'applique donc à la distribution des moyennes d'échantillons de tailles égales (ces moyennes étant des sommes, toutes divisées par la même valeur).



Estimateur de l'écart-type : $S \sim N(n\mu, n\sigma^2)$

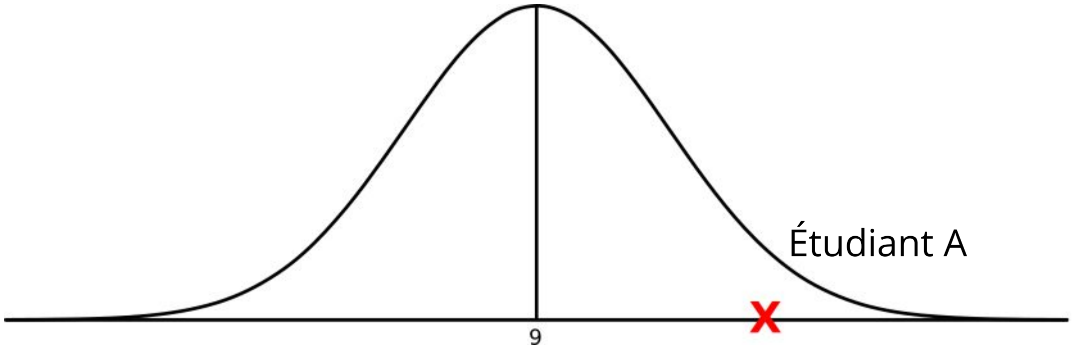
Variable centrée réduite (1)

Question) Deux étudiants veulent savoir qui est le meilleur, en comparant leurs notes obtenues à l’UE de statistiques. Ils appartiennent à deux groupes de TD différents, chacun noté par un enseignant différent.

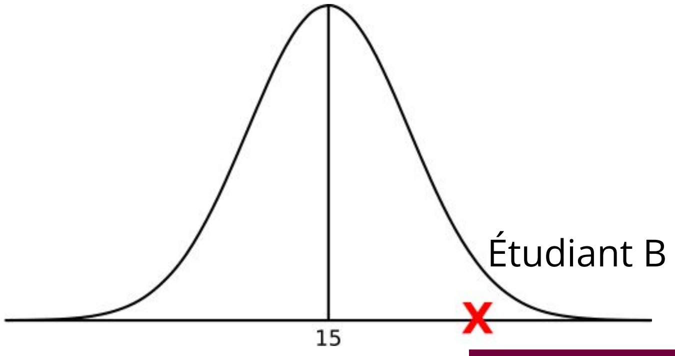
L’étudiant A a eu 15/20, dans le groupe 1, où la moyenne est de 9 et l’écart-type est de 5.

L’étudiant B a eu 19/20, dans le groupe 2, où la moyenne est de 15 et l’écart-type est de 3.

Illustration du problème avec des notes distribuées selon une loi normale et une représentation de la densité estimée :



Groupe 1



Groupe 2

Variable centrée réduite (2)

Pour comparer les notes des deux étudiants, il faut calculer leurs notes

- centrées, par soustraction avec la moyenne μ
- et réduites, en divisant par l'écart-type σ

Ainsi, pour toute variable centrée réduite : μ = 0 et σ = 1

La variable centrée réduite est également appelée variable standardisée, **Z-score**, valeur Z, ou (improprement) variable normalisée.

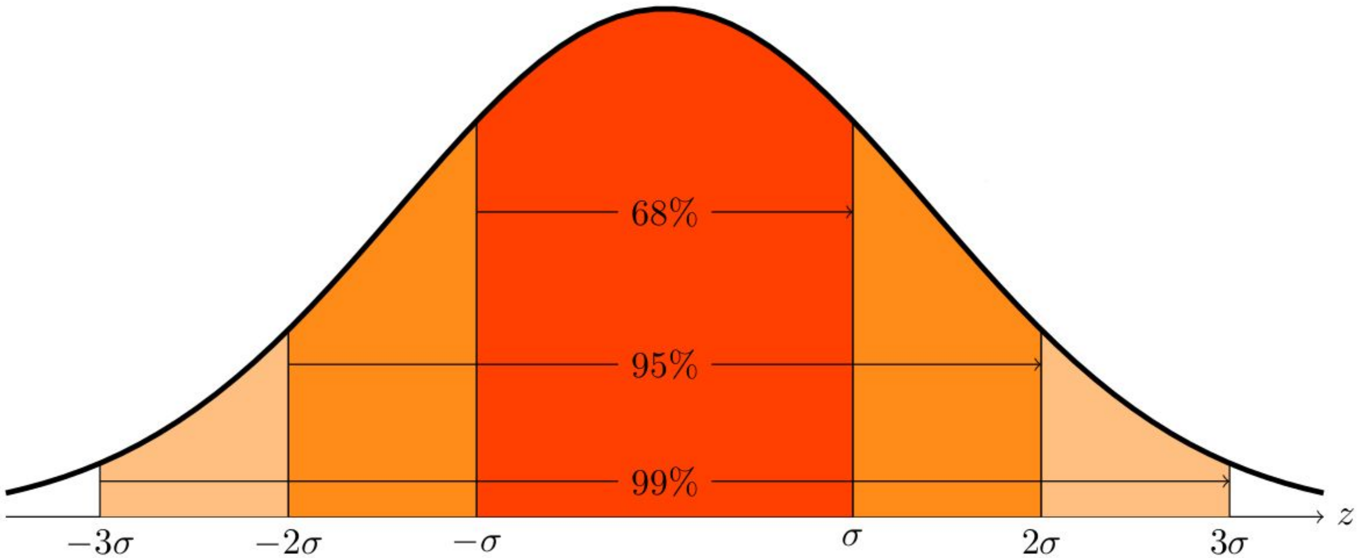
$$z = \frac{x - \mu}{\sigma}$$

Ainsi, si $X \sim N(\mu, \sigma)$, alors $Z \sim N(0, 1)$

Les notes centrées réduites suivent toutes la même distribution et sont directement comparables. L'étudiant A a une note standardisée de 6/5, inférieure à celle de l'étudiant B, 4/3.

Variable centrée réduite (3)

Lecture de la loi normale centrée réduite (ou standardisée) ou **fonction de distribution Z** :



1. $P(-1 < Z < 1)$ is
(a) 0.025 (b) 0.16 (c) 0.68 (d) 0.84 (e) 0.95

2. $P(Z > 2)$
(a) 0.025 (b) 0.16 (c) 0.68 (d) 0.84 (e) 0.95