

An ornate, golden frame with intricate carvings and scrollwork, surrounding the text. The frame is highly detailed, with floral and scroll motifs at the corners and midpoints of each side.

LE MONDE MERVEILLEUX
DES STATISTIQUES DESCRIPTIVES

by Pr. Maignant

PLAN :

♥ La Méthode Statistique en Médecine :

I) Définitions

II) Les types de Variables

III) Paramètres

♥ Les Statistiques Descriptives :

I) Notion de Variabilité

II) Estimation en Statistiques



LA MÉTHODE STATISTIQUE EN MÉDECINE

I) Définitions

🌀 Statistique : art de collecter, analyser et interpréter des données.

Lorsqu'elle est appliquée au domaine de la biologie, on parle de biostatistique

- Descriptive : description d'une situation à l'aide de paramètres

- Déductive : l'observation est-elle due au hasard ? Existe-t-il une autre explication ?

🌀 Donnée : résultat de l'observation d'un individu. *On observe une grande variabilité des données dans le domaine biologique qui peut être due au hasard ou qui peut être physiologique : inter sujet (comparaison de 2 sujets) ou intra sujet (comparaison du sujet à lui-même)*



☹ Variable quantitative : mesurable + + + , obtenue grâce à un appareil de mesure

☹ Variable qualitative : non mesurable + + +

☹ Population : série exhaustive de tous les individus étudiés + + +

☹ Echantillon : extrait de la population. Il doit être représentatif de la population d'où la nécessité du tirage au sort = randomisation + + +

II) LES TYPES DE VARIABLES

Variables qualitatives	Variables quantitatives
<p>Binaires : homme/femme Nominales : couleur des yeux Ordinales : douleur articulaire</p>	<p>Discrètes : âge Continues : poids, glycémie</p>

Une variable qualitative ordinale peut être approximée en une variable pseudo quantitative : la variable est QUALITATIVE mais ressemble à une quantitative.

III) PARAMÈTRES +++

☹ Moyenne :

- Variable quantitative discrète : $m = \Sigma xi / n$
- Variable quantitative continue : $m = \Sigma nixi / n$

☹ Variance : indique la dispersion des données autour de la moyenne

☹ Médiane : valeur de l'observation centrale qui sépare la série d'un effectif n en 2 sous séries de même effectif

Si n est pair : la médiane est donnée par la moyenne des deux valeurs correspondantes à $n/2$ et $(n/2)+1$

Si n est impair : la médiane est donnée par $(n+1)/2$

☹ Quartiles : valeurs de la variable qui partagent la série d'effectif n en 4 sous séries de même effectif

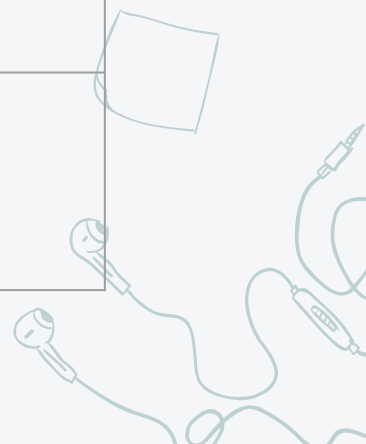
Ex : les notes de 7 PACES à l'épreuve d'UE4 : 15/12/13/9/18/15/10

1) moyenne : $m = (15*2 + 12 + 13 + 9 + 18 + 15 + 10) / 7 = 13,1$

2) médiane : on classe par ordre croissant : 9 ; 10 ; 12 ; 13 ; 15 ; 15 ; 18

7 notes → impair donc on prend la valeur $(n+1) / 2 = (7+1)/2 = 8/2 = 4$
C'est donc la 4ème valeur donc médiane = 13

3) 1er quartile : $0,25 * 7 = 1,75 \rightarrow Q1$ se trouve entre la 1ère et la 2ème valeur
Donc $Q1 = (9+10)/2 = 9,5$ 25% des PACES ont une note inférieure à 9,5.



	Avantages	Inconvénients
Moyenne	<ul style="list-style-type: none">♥ Facile à calculer♥ Se manipule facilement dans les tests statistiques = adaptée aux calculs statistiques♥ Très significative si la répartition des données est assez symétrique et la dispersion faible	<ul style="list-style-type: none">♠ Sensible aux valeurs anormales (max et min)
Médiane	<ul style="list-style-type: none">♥ Calcul facile, peu sensible aux valeurs anormales♥ Utilisable pour les valeurs ordinales, les classes...	<ul style="list-style-type: none">♠ Se prête moins aux calculs statistiques



STATISTIQUES DESCRIPTIVES

I) Notion de Variabilité

Toutes les données biologiques possèdent une variabilité.

- ☹️ Une variabilité maîtrisée permet une estimation
- ☹️ Une variabilité non maîtrisée conduit à des biais

II) ESTIMATIONS EN STATISTIQUES

A) Définitions *(encore...)*

Les études en biostatistique sont réalisées sur un **échantillon représentatif** de la population après « échantillonnage »

On détermine des paramètres au niveau d'une population à partir d'observations réalisées sur un échantillon de cette population



Échantillon



ESTIMATION



Population cible



2 types d'estimations :

1. L'estimation ponctuelle : valeur unique jugée la meilleure à l'instant t (peu fiable+++).
2. L'estimation par intervalle : il y a un intervalle de valeurs comprenant la valeur recherchée, c'est l'**Intervalle de Confiance** ou IC (beaucoup plus fiable+++)
L'estimation par intervalle est plus juste.

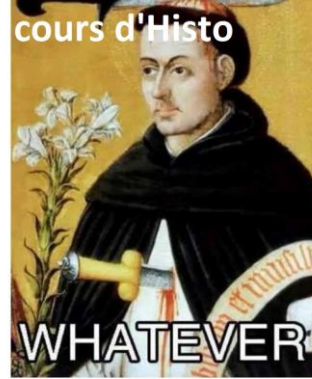
B) ESTIMATION DES DONNÉES QUANTITATIVES :

Méthodologie :

- 1) Définition précise de la population étudiée = Population cible
- 2) Tirage au sort+++ d'un échantillon représentatif
- 3) Calcul de l'intervalle de confiance

Pour les données quantitatives, on va estimer la moyenne !

Après le 15^{ème}
cours d'Histo



⊗ Écart-type : Il mesure la dispersion d'un ensemble de données autour de la moyenne

⊗ Intervalle de Confiance : L'IC c'est l'estimation de la moyenne vraie μ à partir de la moyenne calculée sur l'échantillon



$$\mu \in [m \mp \varepsilon s / \sqrt{n}] = \text{IC}$$

L'IC est aussi appelé intervalle au risque α . Le risque α c'est le risque d'erreur dans l'estimation de μ . On prend en général $\alpha = 5\%$

⊗ ε représente l'écart-réduit : C'est une valeur qui dépend du risque α

pour $\alpha = 5\%$; $\varepsilon = 1,96$

Les variations du risque α vont conditionner la précision de l'estimation et la largeur de l'intervalle de confiance

IC Large	IC Resserré
<p>Si $\alpha \searrow$ alors $\epsilon \nearrow$ donc l'IC \nearrow</p> <ul style="list-style-type: none">→ On a plus de chances que μ soit comprise dans l'IC→ Par contre on perd en précision	<p>Si $\alpha \nearrow$ alors $\epsilon \searrow$ donc l'IC \searrow</p> <ul style="list-style-type: none">→ On a moins de chance que μ soit dans l'IC→ Mais on diminue l'IC, on gagne en précision
 <p>La précision est mauvaise parce que les flèches ne sont pas au centre mais il n'y a pas de valeurs qui ne sont pas dans l'IC</p>	 <p>La précision est meilleure puisque les flèches sont + proches du centre mais les points verts ne sont pas dans l'IC</p>

Ici on visualise l'intervalle de confiance comme une cible

☉ Indice de précision « i » : Il permet de calculer la précision de l'estimation de μ . Cette valeur représente la largeur de l'IC.

- $i = \varepsilon s / \sqrt{n}$

- IC = [m \mp i]

Attention, quand l'indice de précision diminue la précision augmente !

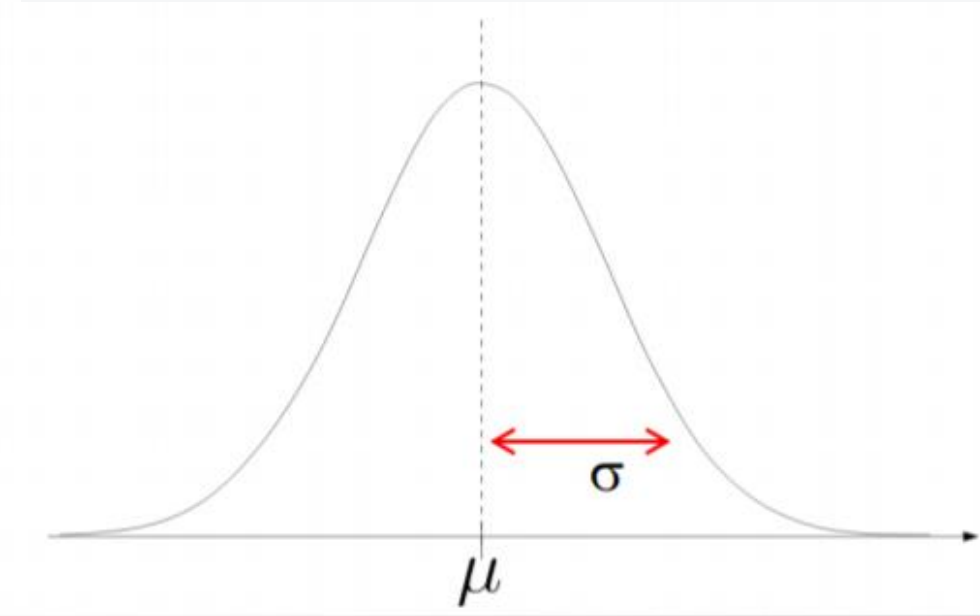
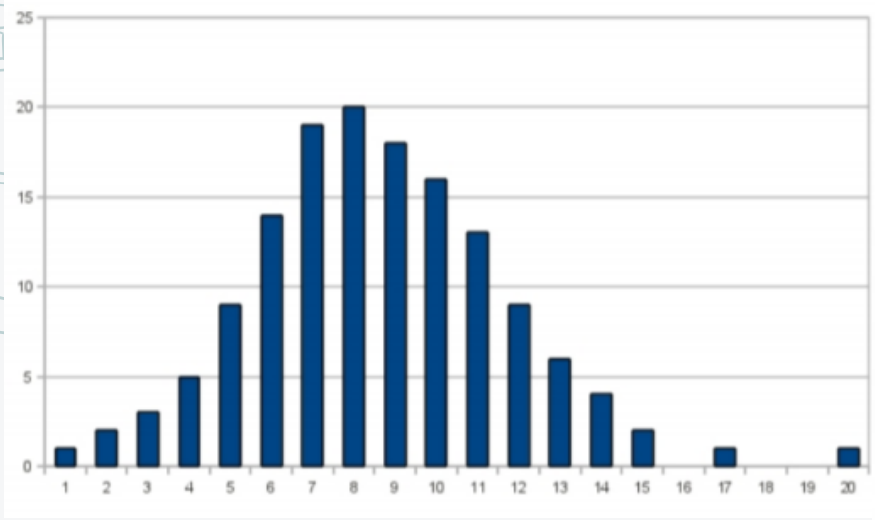
☉ Donc +++ :

Si $n \nearrow$ alors $i \searrow$ donc l'IC \searrow donc la précision \nearrow

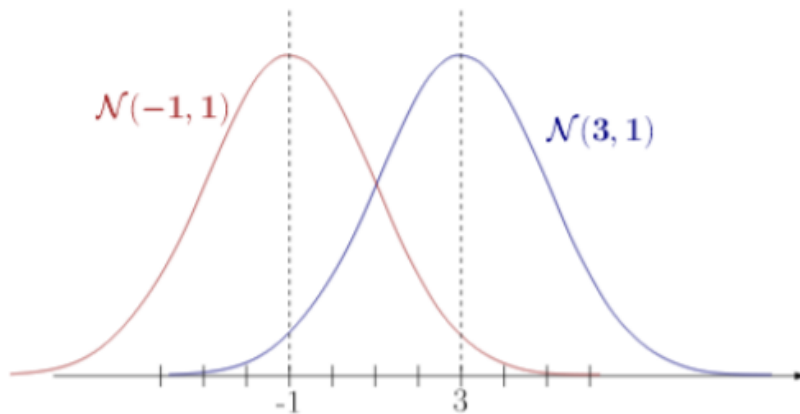
Si $\alpha \nearrow$ alors $\varepsilon \searrow$ donc $i \searrow$ donc l'IC \searrow donc la précision \nearrow

LOI DE GAUSS OU LOI NORMALE :

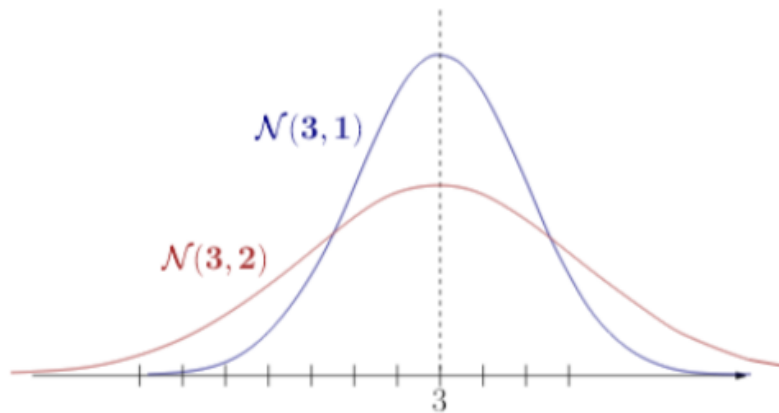
- ☹ En sciences humaines, on observe souvent des distributions des variables plutôt symétriques autour de la moyenne avec une forme de cloche : c'est la **courbe de Gauss**
- ☹ La loi de Gauss permet de visualiser l'IC autour de la moyenne, l'écart type, la dispersion autour de cette valeur moyenne et la moyenne
- ☹ Pour pouvoir faire des calculs on va supposer que notre variable X (quantitative continue) suit une distribution « modèle » : la **loi Normale**. Ainsi, Pour chaque (μ, σ) il existe une loi normale de moyenne μ et d'écart type σ : on la note **$N(\mu, \sigma)$**

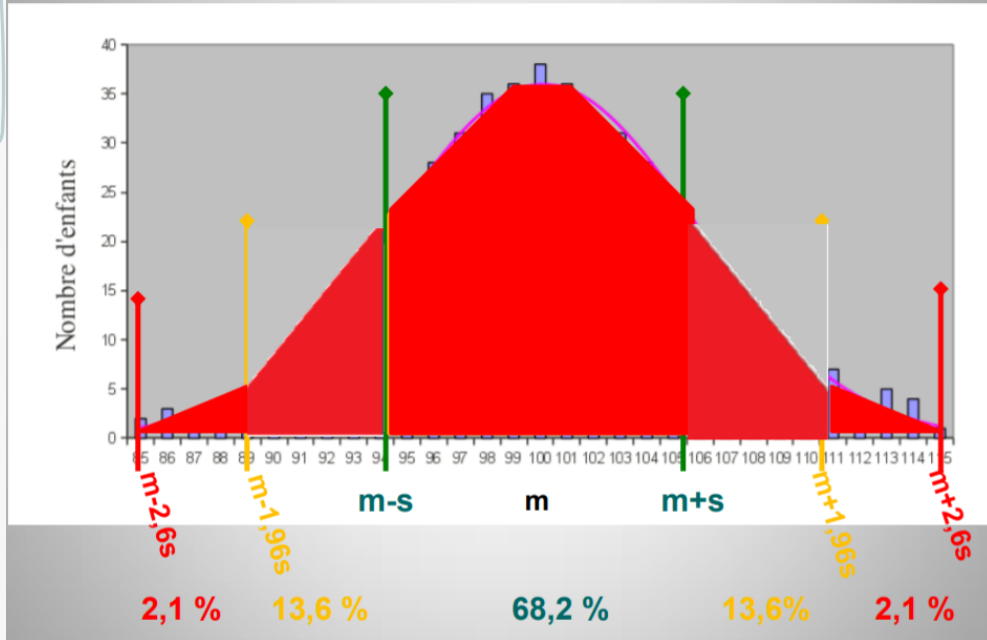


Exemples de lois normales avec **moyennes différentes**, même écart-type :



Exemples de lois normales avec même moyenne, **écart-types différents** :





- ★ $[m - 1 s ; m + 1s]$
- ★ $[m - 1,96 s ; m + 1,96s]$
- ★ $[m - 2,6 s ; m + 2,6s]$

contient 68,2% de la population
 contient 95,4% de la population
 contient 99,6% de la population

C) ESTIMATION DES DONNÉES QUALITATIVES :

Méthodologie :

- 1) Constitution d'un échantillon représentatif par TAS
- 2) Calcul du pourcentage p_{obs} de l'échantillon présentant un caractère A et de l'écart-type « s »
- 3) Estimation de la valeur vraie « p » du pourcentage de la population présentant A et de l'écart-type « σ »

Pour les données qualitatives, on va estimer un pourcentage !

SONDAGES

☪ Le sondage est une application directe de l'IC calculée sur des données qualitatives. Tout résultat de sondage doit être accompagné d'un IC.

☪ Pour une bonne estimation il nous faut +++ :

-Un échantillon représentatif constitué par TAS

-Pas de biais pendant la sélection

-Un IC qui accompagne toujours l'estimation (il montre la variabilité des données)



-Une taille importante de l'échantillon : Si $n \nearrow$ la précision \nearrow

QRU 1: A propos des définitions en biostatistiques, donnez la bonne réponse :

- A) Une variable quantitative est mesurable, est obtenue via un appareil de mesure et peut s'apparenter à une douleur articulaire.
- B) Une variable qualitative est non mesurable et peut être binaire, nominale ou ordinale.
- C) Une population représente la grande majorité des individus étudiés.
- D) Un échantillon fiable est un extrait de la population réalisé ou non avec l'aide d'un tirage au sort = randomisation.
- E) Les réponses A, B, C, D sont fausses.



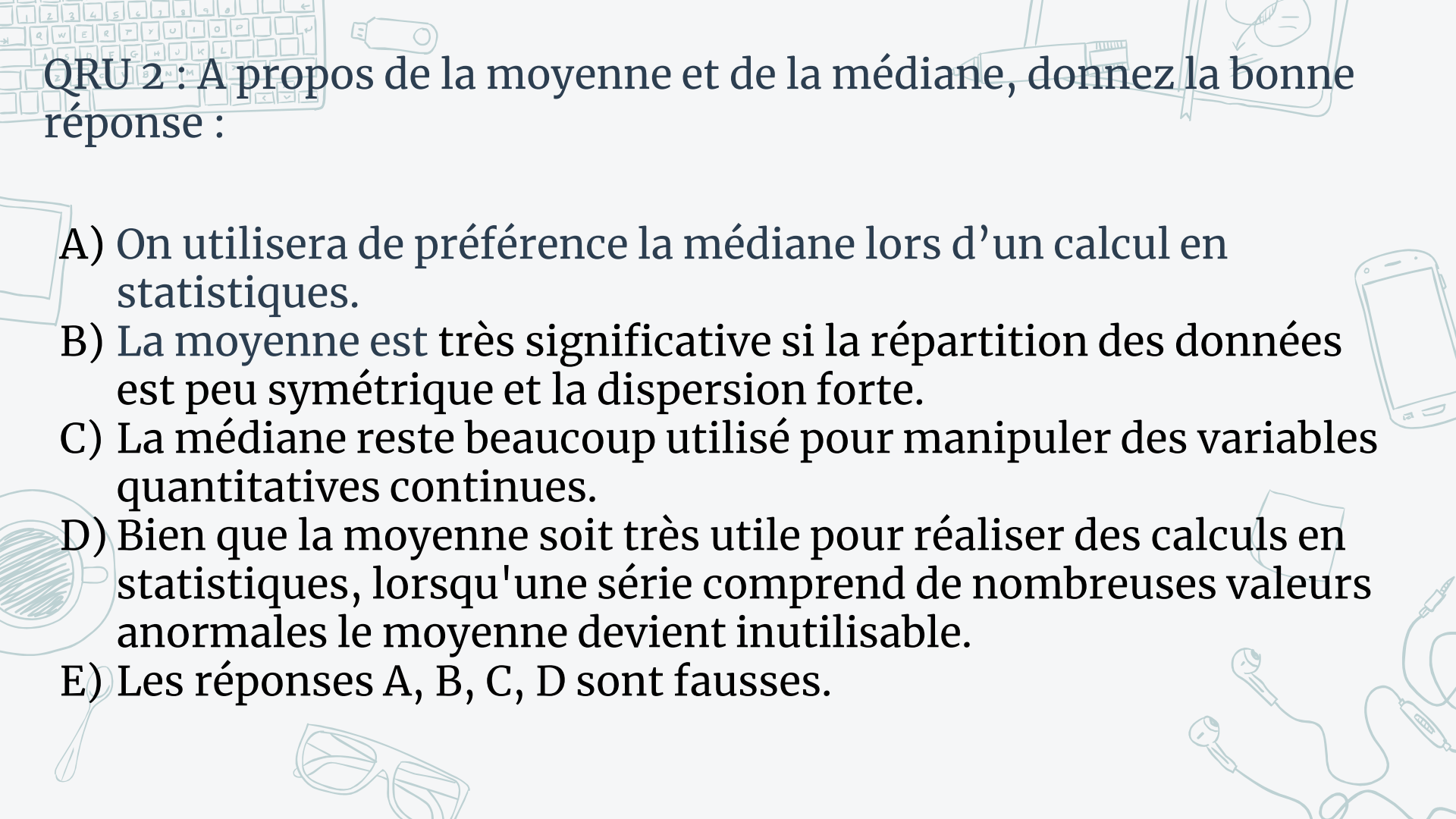
Correction QRU 1 :

- 
- A) Faux : Une douleur articulaire est une douleur QUALITATIVE, le fait qu'elle soit ordinale n'enlève en rien le fait qu'elle reste qualitative.
 - B) Vrai
 - C) Faux : La Population comprend la totalité des individus étudiés.
 - D) Faux : Un échantillon est TOUJOURS réalisé avec l'aide d'un tirage au sort. D'autant plus si il est fiable.
- 



Réponse QRU 1 : B





QRU 2 : A propos de la moyenne et de la médiane, donnez la bonne réponse :

- A) On utilisera de préférence la médiane lors d'un calcul en statistiques.
- B) La moyenne est très significative si la répartition des données est peu symétrique et la dispersion forte.
- C) La médiane reste beaucoup utilisé pour manipuler des variables quantitatives continues.
- D) Bien que la moyenne soit très utile pour réaliser des calculs en statistiques, lorsqu'une série comprend de nombreuses valeurs anormales le moyenne devient inutilisable.
- E) Les réponses A, B, C, D sont fausses.

Correction QRU 2 :

- A) Faux : On utilisera de préférence la moyenne lors d'un calcul en statistiques.
- B) Faux : La moyenne est très significative si la répartition des données est assez symétrique et la dispersion faible.
- C) Faux : Nimp. La médiane est surtout utilisée pour des variables quantitatives discrètes ou ordinales.
- D) Faux : Assez pute, même si la moyenne est sensible aux valeurs normales, elle reste utilisable lorsqu'il y en a, elle sera juste moins précise.

Aller c'est pour vous mettre dans le bain ;)

Réponse QRU 2 : E



QRU 3 : A propos de plein de trucs, donnez la bonne réponse :

- A) En sciences humaines, on observe souvent des distributions des variables plutôt asymétriques autour de la moyenne avec une forme de cloche : c'est la courbe de Gauss.
- B) $[m - 1,96s ; m + 1,96s]$ contient 90,8% de la population.
- C) Pour les données qualitatives, on va estimer un pourcentage.
- D) Un sondage est surtout réalisé à partir d'études portant sur des variables quantitatives.
- E) Les réponses A, B, C, D sont fausses.

Correction QRU 3 :

- A) Faux : La distribution des variables est souvent **symétrique** autour de la moyenne.
- B) Faux : $[m - 1,96s ; m + 1,96s]$ contient **95,4%** de la population. C'est une valeur à connaître !
- C) Vrai
- D) Faux : Un sondage est préférentiellement réalisé en utilisant des variables qualitatives. ex : sondage lors des élections.

Réponse QRU 3 : C

