

# DataBird: Statistiques

<https://infogram.com/fr/page/choisir-le-bon-graphique-visualisation-de-donnees>



Voici une série d'exercices sur les statistiques descriptives

## Exercice 1: Les statistiques descriptives

Supposons que tu aies recueilli des données sur le temps de sommeil quotidien de 20 personnes. Voici les données que tu as collectées (en heures) :

7.5, 8.2, 6.8, 7.1, 6.5, 7.9, 8.0, 6.3, 7.7, 7.0, 6.6, 8.5, 7.2, 6.9, 7.8, 8.1, 7.4, 7.6, 6.7, 7.3

Maintenant, effectue les étapes suivantes pour analyser ces données :

1. Calcule la moyenne du temps de sommeil.
2. Calcule la médiane du temps de sommeil.
3. Calcule l'écart type du temps de sommeil.
4. Calcule la variance du temps de sommeil.
5. Calcule l'étendue du temps de sommeil.
6. Calcule le premier quartile et le troisième quartile du temps de sommeil.
7. Trace un diagramme en boîte (box plot) pour représenter graphiquement les données.

Note : Tu peux effectuer les calculs manuellement ou utiliser un logiciel statistique de ton choix (Python)

## Exercice 2: Choisir le bon type de graphiques

Quel type de graphique devrais-tu utiliser pour représenter ces données

- Situation 1: Tu veux représenter l'évolution des températures quotidiennes moyennes d'une ville sur une période d'un an.
- Situation 2: Tu souhaites comparer les performances de vente de plusieurs produits dans différentes régions du pays.
- Situation 3: Tu analyses la répartition des âges des étudiants dans une université.
- Situation 4: Tu étudies la relation entre le niveau d'éducation et le salaire des employés.
- Situation 5: Tu veux montrer la répartition des membres d'un club sportif par catégorie d'âge.

Line chart

Bar Chart

Bar chart,  
pie chart

Scatter plot

Histogram

## Exercice 3: Le biais de centralité

Situation 1 : Tu collectes des données sur les revenus mensuels de 100 personnes dans une ville.

- Question : Tu souhaites calculer une mesure de tendance centrale pour représenter les revenus dans cette ville. Quelle mesure devrais-tu utiliser et pourquoi ? Mode

Situation 2 : Tu étudies les temps de réaction des athlètes dans une compétition sportive.

- Question 1 : Tu as recueilli les temps de réaction de 20 athlètes. Quelle mesure de tendance centrale devrais-tu utiliser pour représenter ces temps de réaction et pourquoi ?
- Question 2 : Tu as remarqué qu'un athlète a obtenu un temps de réaction extrêmement élevé par rapport aux autres. Comment cela pourrait-il affecter le choix de la mesure de tendance centrale et pourquoi ?
- Question 3 : Quels autres facteurs devrais-tu prendre en compte lors de l'analyse des temps de réaction des athlètes, en plus de la mesure de tendance centrale ?

#### Exercice 4: Les biais d'échelles

Situation : Tu compares les budgets de recherche de trois départements au sein d'une entreprise.

- Question 1 : Les budgets des trois départements sont respectivement de 500 000 €, 1 000 000 € et 2 000 000 €. Si tu présentes ces budgets sous forme de pourcentages, quel biais potentiel cela pourrait-il entraîner et pourquoi ?
- Question 2 : Si tu présentes les budgets des trois départements en termes absolus, comment cela pourrait-il affecter la perception des différences entre les budgets et pourquoi ?

Causation means that changes in one variable brings about changes in the other;  
there is a cause-and-effect relationship between variables. The two variables are correlated with each other and there is also a causal link between them.  
A correlation doesn't imply causation, but causation always implies correlation.

#### Exercice 5: Corrélation et causalité

Situation : Tu collectes des données sur le temps d'étude et les performances académiques des étudiants dans une classe.

- Question 1 : Tu observes une corrélation positive élevée entre le temps d'étude et les performances académiques des étudiants.

On peut émettre l'hypothèse que le temps passé à étudier influence les résultats scolaires. Mais il faudrait vérifier s'il n'existe pas des variables cachées pour pourr

Que peux-tu conclure de cette corrélation en termes de causalité ?

- Question 2 : Comment peux-tu différencier la corrélation de la causalité dans cette situation ? Quels autres facteurs pourraient influencer à la fois le temps d'étude et les performances académiques ?
- Question 3 : Quels types de preuves supplémentaires pourraient être nécessaires pour établir une relation de causalité entre le temps d'étude et les performances académiques des étudiants ?

### Exercice 6: Le paradoxe de Simpson

Situation : Tu analyses les taux de réussite aux examens dans deux groupes, A et B, composés d'étudiants de différents départements.

Données :

Groupe A :

- Département 1 : 90% de réussite (100 étudiants)
- Département 2 : 80% de réussite (200 étudiants)

Groupe B :

- Département 1 : 70% de réussite (300 étudiants)
- Département 2 : 60% de réussite (150 étudiants)

Question 1 : Si tu analyses les taux de réussite de chaque département séparément, lequel semble avoir de meilleurs résultats dans chaque groupe ?

Question 2 : Si tu combines les résultats des deux départements dans chaque groupe, qu'observe-tu concernant les taux de réussite globaux dans les groupes A et B ?

Question 3 : Comment le paradoxe de Simpson se manifeste-t-il dans cette situation ? Quelle conclusion importante peut-on en tirer ?

Une fois que tu as réalisé ces exercices tu peux venir télécharger le notebook d'exercices de la session

Saisissez du texte ici

Enoncé correct pour l'exercice 6 :

Il y a deux hôpitaux dans une ville, Hôpital A et Hôpital B.

Données :

Hôpital A :

Département enfant : taux de survivabilité : 30 % ( sur 100 patients)

Département adulte : taux de survivabilité : 96.66668 % ( sur 900 patients)

Hôpital B :

Département enfant : taux de survivabilité : 52.5% ( sur 400 patients)

Département adulte : taux de survivabilité : 98.3 % ( sur 600 patients)

Q1: Si tu analyses les taux de réussite de chaque département entre les 2 hôpitaux, quel hôpital semble être le meilleur?

Q2 : Si tu calcules le taux de survivabilité total de chaque hôpital, quel hôpital paraît être le meilleur ?

Q3: Comment le paradoxe de Simpson se manifeste-t-il dans cette situation ? Quelle conclusion importante peut-on en tirer ?