

TICE Probabilités et statistiques 1 : Tableur et statistiques avec Geogebra

Prenez une feuille pour écrire vos commentaires.

1 Stabilisation des fréquences : la loi des grands nombres

Ouvrir geogebra, afficher les fenêtres Tableur et Graphique (menu Affichage en haut de la fenêtre). Enregistrer le fichier sous LGN.ggb.

On veut expliquer la notion de probabilité à l'aide du lancer d'un dé.

1. Simulation des lancers, calcul des fréquences
 - (a) Dans la colonne A du tableur, entrer les numéros des lancers jusqu'à 400.
 - (b) Dans la colonne B, simuler 400 lancers d'un dé à 6 faces. *Utiliser les commandes Alea.*
 - (c) Dans la colonne C, pour tout $1 \leq N \leq 400$, compter le nombre des 6 obtenus lors des N premiers lancers. *On peut utiliser les commandes Si() ou NbSi().*
 - (d) Indiquer dans la colonne D les fréquences associées.
2. Observation de la stabilisation des fréquences quand le nombre de lancers augmente
 - (a) Supprimer l'étiquetage des nouveaux points (*accessible dans Options/Etiquetage*). Afficher dans la fenêtre graphique les points indiquant la fréquence en fonction du nombre de lancers. *Créer les points dans la colonne E du tableur.*
 - (b) Complétez le graphe et commenter.
3. Observation des fluctuations des fréquences en fonction du nombre de lancers
 - (a) Afficher la grille de la fenêtre graphique et la trace des points d'abscisses 25, 100 et 400 qu'on peut mettre en couleur, puis recalculer plusieurs fois la série de lancers (utiliser le raccourci Ctrl + R). Que peut-on remarquer ?
 - (b) Pourquoi ces choix d'indices sont-ils particulièrement intéressants ?

2 Principaux éléments de statistiques descriptives

Ouvrir un nouveau fichier, Stat1Var.ggb, dans lequel on affiche le tableur et une fenêtre graphique. On s'intéresse à la répartition des salaires dans une entreprise A.

1. Collecte des données : la série brute

On recueille les salaires annuels en k€ par ordre alphabétique :

12, 17, 84, 25, 28, 16, 16, 70, 21, 23, 16, 32, 56, 24, 23, 21, 17, 22, 18, 25, 33, 16 .

Entrer la série ci-dessus dans la colonne A du tableur. Indiquer la légende dans A1. Exporter la liste sous le nom Salaires (sélectionner les valeurs et faire un clic droit).

2. Organisation des données : la série échantillonnée

Dans la colonne B, afficher la liste triée précédée de sa légende.

La commande Trier permet de trier une liste numérique par ordre croissant. On peut remplir les cellules du tableur avec la liste triée en utilisant la commande RemplirColonne dans la ligne de saisie générale. Attention, les colonnes sont numérotées.

3. Représentation des données : tableaux, courbes et graphiques

(a) Diagramme en bâtons

Afficher un diagramme en barres associée à la série brute, à l'aide de la commande Barres, avec des barres de largeur 0.1, à entrer dans la ligne de commande globale¹.

Quels sont les avantages et les inconvénients de cette représentation ? Donner des exemples de données pour lesquelles cette représentation serait adaptée.

Un diagramme en barres peut être utilisé pour représenter des données qualitatives, ou des données quantitatives ayant peu de valeurs distinctes.

(b) Histogrammes

Sélectionner les données de la colonne A, cliquer sur l'icône , choisir Statistiques à une variable puis Histogramme dans le menu déroulant.

1. Choisir une petite largeur permet d'émuler un diagramme en bâtons, non implémenté dans certaines versions de Geogebra.

i. Les classes

Faire varier le curseur. Préciser la définition d'une classe, ce qu'indique le curseur et la construction de l'histogramme (en particulier le rôle des abscisses et des ordonnées).

L'histogramme est une représentation privilégiée pour des données quantitatives continues ou discrètes lorsque le nombre de valeurs distinctes est grand.

- ii. Faire varier le nombre de classes. Quels problèmes rencontre-t-on si le nombre de classes est trop petit ou trop grand ? Pour cette série, quel nombre de classes est, à votre avis, idéal ?
- iii. En faisant un clic droit sur l'histogramme, on peut l'envoyer vers la fenêtre graphique. Envoyer les histogrammes obtenus avec des classes de pas 5, 10 et 20 vers la fenêtre graphique. Qu'observe-t-on ? Expliquer.
- iv. Histogrammes par effectifs ou fréquences relatives
Ouvrir les options avec le bouton . Dans les options, 3 types d'histogrammes sont proposés : par effectifs, relatives ou normalisées. Indiquer dans chacun des cas la quantité représentée en ordonnée, ainsi que son unité.
- v. De la même façon que pour les histogrammes par effectifs, envoyer vers la fenêtre graphique 3 histogrammes normalisés avec 5, 10 et 20 classes. En déduire un intérêt de l'histogramme des fréquences normalisées.
L'histogramme des fréquences normalisées est enseigné en fin de cycle 4.
- vi. *Facultatif* : Il est possible, et parfois pertinent, de faire des histogrammes avec des classes de tailles différentes. Dans Geogebra, cela se fait avec la commande `Histogramme`, en mettant en paramètres la liste des limites entre les classes et la liste des données (effectifs, effectifs relatifs ou densités). Expliquer dans ce cadre un deuxième intérêt des histogrammes relatifs ou normalisés.
- vii. Décrire la loi dont un histogramme normalisé est la densité.

(c) Courbe des fréquences cumulées

- i. Choisir une répartition en classes. Utiliser la colonne C pour afficher les limites des classes, la colonne D pour afficher les effectifs cumulés (utiliser la commande `NbSi`), la colonne E pour afficher les fréquences cumulées.
- ii. Afficher la courbe des fréquences cumulées dans la fenêtre graphique.
- iii. Comment trouver graphiquement la médiane et les quartiles à l'aide de cette courbe ?
- iv. Quel est le lien entre la courbe des fréquences cumulées et l'histogramme normalisé ? Interpréter la médiane à l'aide de l'histogramme normalisé.

(d) La commande Boite à Moustache : le diagramme en boîte

Exporter le dessin obtenu à l'aide de la commande dans la fenêtre graphique. Expliquer sa construction et indiquer les valeurs numériques nécessaires à la construction de ce graphique.

4. **Synthèse des données : les indicateurs statistiques.**

Sélectionner la série brute, cliquer sur  puis choisir `Statistiques` à une variable. Identifier les données numériques indiquées.

(a) Caractéristiques de position : mode, moyenne, médiane

Dans la colonne F, écrire sous forme texte les différents indicateurs de position suivis de leur valeur numérique, en utilisant les commandes pré-définies et la liste `Salaires`. Vérifier les valeurs des indicateurs statistiques à l'aide de la fenêtre.

(b) Sensibilité aux données des indicateurs de position

Modifier une ou plusieurs valeurs de la liste `Salaires`. Observer les modifications sur la médiane et la moyenne. Commenter.

(c) Caractéristiques de dispersion : étendue, écart interquartile, écart-type (variance)

Dans la colonne H, écrire sous forme texte les différents indicateurs de dispersion suivis de leur valeur numérique, en utilisant les commandes pré-définies.

Voir l'aide pour les formules. Il n'y a pas de commande "étendue" ni "interquartile". En revanche, on peut utiliser les fonctions `Min`, `Max`, `Q1`, `Q3`, `Centile`.

5. **Analyse des données : comparer plusieurs séries**

Dans une autre entreprise, comptant 19 salariés, la répartition des salaires est la suivante (salaires annuels en k€) : 11, 45, 22, 17, 25, 26, 23, 23, 25, 31, 18, 18, 15, 24, 27, 26, 42, 24, 27.

Entrer en colonne J ces dernières données précédées de leur légende. Utilisez les outils précédents pour comparer les répartitions des salaires dans les deux entreprises.

3 Outil statistiques à deux variables, ajustements

Créer maintenant un nouveau fichier appelé `Stat2Var.ggb`. Faire afficher dans la fenêtre principale le tableur et la zone graphique. On voudrait étudier la série des données couplées suivantes qui indiquent la taille du fils (ligne du bas) selon la taille du père. Cette série de données est issue des notes de Galton sur l'étude des tailles des enfants selon la taille des parents (1886). La taille est obtenue en pouces (2.54 cm) en ajoutant 60 aux valeurs indiquées dans le tableau.

Père	5	6.2	6.5	7	8	8	8.5	9	9	9	9.5	10	10	10	10.5	11	11	12	13	14.5
Fils	7.3	7.4	7.3	9.3	7.5	10	9.1	8.5	11.3	9.5	8.4	8.8	9	10.5	9.8	9.5	11	10.8	10	11.6

Entrer ces valeurs dans les colonnes A et B du tableur. On indiquera les légendes en ligne 1.

1. Indiquer les valeurs des données en mètres dans les colonnes C et D. On ne travaillera plus que sur ces données.
2. Sélectionner les deux séries de données, cliquer sur l'icône , puis choisir `Statistiques à deux variables` . Afficher les statistiques. Identifier les indicateurs de position et de dispersion de chacune des séries.
3. Retrouver les valeurs de la fenêtre statistique par les commandes directes, et les indiquer en colonne E.
4. Sélectionner l'ajustement affine dans le menu statistique. Calculer les valeurs des coefficients de la droite des moindres carrés des données D selon C dans la colonne F.

AIDE : Tableur et statistiques avec Geogebra

Avec le Tableur :

- Pour **étendre une formule** aux élément d'une ligne ou d'une colonne, on remplit la première cellule, puis on fait glisser la souris sur la zone où l'on souhaite propager la formule. On peut aussi compléter la colonne en double cliquant sur le coin de la cellule.
- Si on ne veut pas incrémenter l'indice d'un terme de la cellule lorsqu'on étend une formule, on doit mettre le symbole \$ devant l'indice qu'on veut garder fixe. Par exemple $B1=A1$ étendu à la colonne B donnera $B2=A2$, $B3=A3$.. alors que $B1=A$1$ donnera $B2=A1$, $B3=A1$,...
- **La liste des nombres** $A1, A2, \dots, A10$ du tableur s'écrit $A1:A10$. Pour effectuer une opération aux éléments d'une liste, on peut appliquer l'opération à la liste : par exemple " $A1:A10 +1$ " retourne la liste des valeurs $\{A1+1, \dots, A10+1\}$.

Avec Geogebra :

- Pour écrire les légendes sous forme texte, on peut mettre des guillemets. Pour inclure des formules, sélectionner `Formule` `Latex`. On met des \$ autour des formules.
- Pour afficher les coordonnées dynamiques d'un point dans l'étiquetage on utilise $\%x$, $\%y$.
- Pour un affichage dynamique dans un texte, cliquer sur la case `objet`.
- Pour appeler une cellule dans une colonne, on utilise la commande `Cellule`.
- Pour taper une commande Geogebra, il suffit en général de taper le début du nom de la commande, la complétion s'effectue automatiquement.
- On change l'échelle d'un seul axe avec l'outil  en cliquant sur l'axe puis en déplaçant la souris.
- Le clic droit dans la zone graphique permet d'accéder à la commande `recadrer`.
- Pour choisir **la zone d'affichage** aller dans le menu `Propriétés/avancé` après avoir sélectionné l'objet.