Statistiques en IUT

des œufs pour introduire les indicateurs de dispersion

samedi 2 juillet 2022, par Anne-Sophie Suchard

J'enseigne la bureautique et les mathématiques à l'IUT [1] de Cergy pour le B.U.T. [2] Management de la Logistique et des Transports. Au deuxième semestre de la première année, le cours **Statistiques et prévisions**, notamment, m'a été confié.

Le programme national [3] indique :

Développement de la capacité à utiliser les outils mathématiques et statistiques comme support d'argumentation :

- Mise en œuvre des outils répondant aux principaux problèmes quantitatifs liés à la gestion des entreprises
- Représentation graphiques de données
- Caractéristiques des distributions à un caractère (indicateurs de tendance centrale, de dispersion, etc.)
- Séries chronologiques
- Techniques de prévision (avec ou sans saisonnalité)

Mots clés : Tendance — dispersion — séries chronologiques — prévision — graphiques

Volume horaire défini nationalement : 20 heures dont 16 heures de TP.

Je n'ai aucune autre injonction : pas d'examen final, évaluation en contrôle continu pendant mes créneaux de cours avec des énoncés que je crée moi-même.

Je répète mon cours 5 fois à 5 groupes de TP [4] de 12 à 15 étudiants chacun.

Les étudiants n'ont pas de manuel. À chaque cours je leur distribue un « poly » à compléter qui comporte une activité, les formules à utiliser et des exercices d'application. Ils écrivent tout sur les « polys », ils n'ont ni classeurs ni cahiers.

Un diaporama du cours est à leur disposition sur l'ENT [5].

Les notions à présenter dans cette partie sont très utiles mais peu palpitantes. Elles sont, pour la plupart, présentes dans les programmes du collège et du lycée.

Je me suis demandée comment présenter ces notions et ces outils sans m'ennuyer, sans que les étudiants s'ennuient.

Un parcours sur la toile m'a amenée sur le site JavMath.ch où Jean-Philippe Javet met à disposition ses supports de cours, fruit de 30 ans d'enseignement au Gymnase de Morges dans le canton de Vaud en Suisse. Avec, notamment, les chapitres de statistiques.

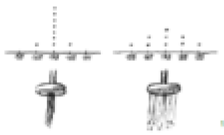
J’y ai trouvé ces schémas et illustrations :

§ 12.7 Les mesures de dispersion d'une variable statistique

Introduction Le paragraphe précédent a été consacré à l'étude de trois mesures de tendance centrale. Elles indiquent autour de quelle valeur se situent les données, mais ne donnent pas une description suffisante de la variable statistique. Par exemple, si on désire comparer les 2 groupes d'élèves proposés dans les diagrammes ci-dessous:

$\overline{x}$ = ... $\overline{x}$ = ...

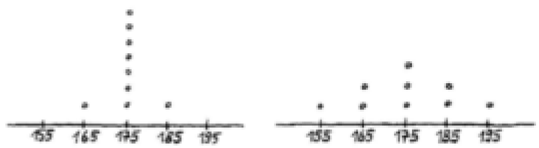
Mais pourtant, les 2 distributions ne sont pas identiques. Les distributions peuvent être comparées à une douche. Si elle est en position « jet étroit », presque toute l'eau est concentrée sur un seul point, c'est-à-dire le jet n'arrose pratiquement que la valeur moyenne. Si la douche est en position « pluie », l'eau est dispersée plus largement : il y a de grands écarts par rapport à la moyenne.



Pour mettre en évidence cette différence, il faut mesurer **la dispersion** des données autour de cette mesure de tendance centrale. Nous allons étudier deux mesures de dispersion.

et un peu plus loin :

En reprenant la situation d'introduction:



Calculer la variance et l'écart type

Cela m’a amenée à créer cette activité :

Je leur dis d’imaginer qu’ils ont décidé d’inviter cinq copains pour un repas et qu’ils vont dans un magasin acheter une boîte de six œufs. Ils doivent choisir une boîte parmi les sept boîtes.

On connaît le poids en grammes de chaque œuf.

Je leur demande d’abord de calculer le poids moyen d’un œuf pour chaque boîte.

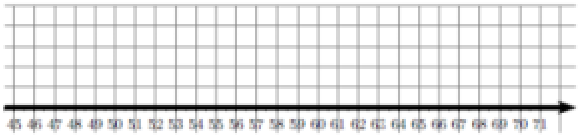
Je leur précise ensuite que les sept boîtes sont vendues au même prix.

Cela nous amène à dire « Il est préférable d’acheter une boîte parmi E, G, H, I, K plutôt que F ou J » ou encore « on élimine F et J ».

							Poids moyen	Variance	Écart type	Poids médian
Boîte E	57	64	66	51	70	64	62			
Boîte F	57	61	52	50	59	63	57			
Boîte G	60	64	64	64	60	60	62			
Boîte H	68	68	59	59	59	59	62			
Boîte I	60	63	63	60	63	63	62			
Boîte J	59	46	59	52	61	65	57			
Boîte K	67	67	57	57	67	57	62			

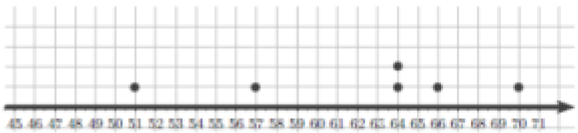
Nous remarquons aussi : « Si tous les œufs d'une boîte étaient identiques, chaque œuf aurait comme poids le poids moyen ».

Je leur demande ensuite de compléter un graphique de ce type pour chaque boite :

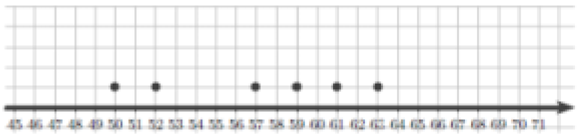


Ils obtiennent :

Boite E



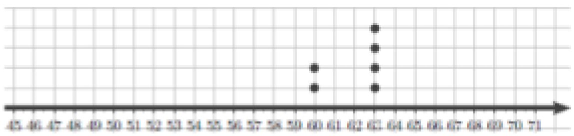
Boite F



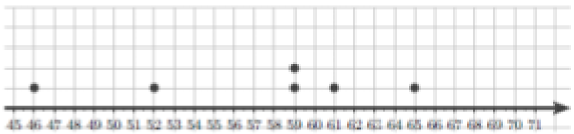
Boite G

Boite H

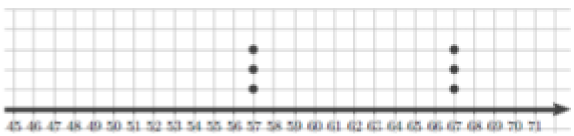
Boite I



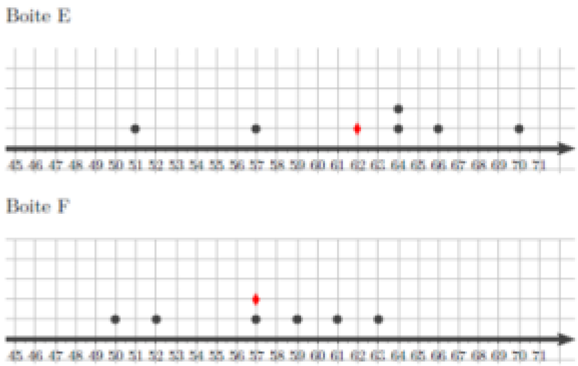
Boite J

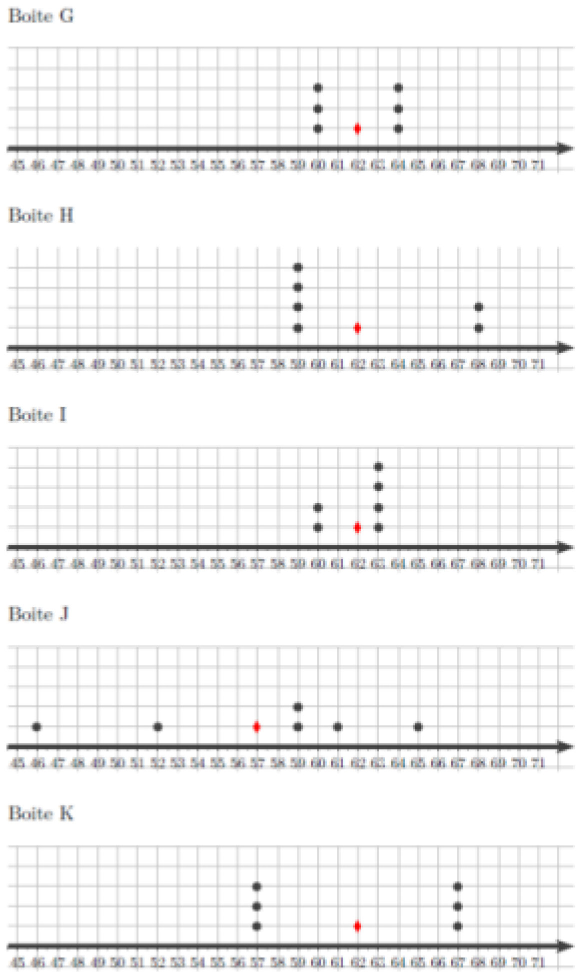


Boite K



Nous ajoutons les poids moyens (marqués en rouge) :





En utilisant ces graphiques je leur demande d'ordonner les sept boîtes d'œufs sachant qu'ils vont servir un œuf dans chaque assiette des cinq invités et qu'on souhaite ne pas créer d'injustice ; la taille des œufs doit être la plus identique possible.

Je note leur proposition au tableau sous forme d'une liste ordonnée avec des ex æquo.

$$\text{E et J} \qquad \text{K} \qquad \text{F} \qquad \text{H} \qquad \text{G} \qquad \text{I}$$

Je leur dis ensuite qu'on souhaite associer un nombre à chaque boîte, ce nombre étant un indicateur de dispersion des valeurs par rapport à la moyenne. Nous souhaitons attribuer le nombre 2 à la boite

G et le nombre 5 à la boite K.

Ils calculent alors variance et écart type en utilisant une calculatrice « collège ».

Les valeurs obtenues sont en général cohérentes avec le classement des boîtes qu'ils avaient effectué.

Lorsque leur classement comporte des erreurs, le calcul des écarts types sert à rectifier le classement.

$\textbf{Définition}$
La **variance** d'une série statistique est la moyenne des carrés des écarts à la moyenne.

$$\text{Variance} = \dfrac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dotsb + (x_n - \bar{x})^2}{n}$$

$\textbf{Propriété}$

$$\text{Variance} = \dfrac{(x_1)^2 + (x_2)^2 + \dotsb + (x_n)^2}{n} - (\bar{x})^2$$

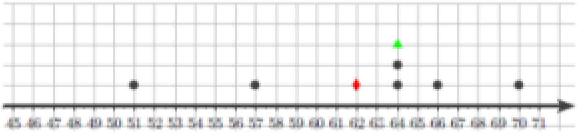
$\textbf{Définition}$
Plus l'**écart type** σ d'une série statistique est grand, plus les valeurs de la série sont éloignées de la moyenne de la série.

L'**écart type** σ est égal à la racine carrée de la variance.

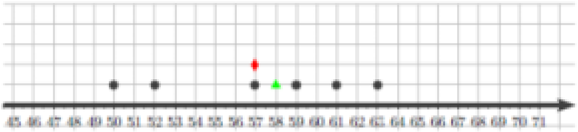
							Poids moyen	Variance	Ecart type
Boite E	57	64	66	51	70	64	62	39	6,24
Boite F	57	61	52	50	59	63	57	21,67	4,65
Boite G	60	64	64	64	60	60	62	4	2
Boite H	68	68	59	59	59	59	62	18	4,24
Boite I	60	63	63	60	63	63	62	2	1,41
Boite J	59	46	59	52	61	65	57	39	6,24
Boite K	67	67	57	57	67	57	62	25	5

Pour finir, nous déterminons le poids médian (marqué en vert) de chaque boite en utilisant les graphiques et la consigne donnée oralement : « 3 œufs doivent être plus légers que le poids médian et 3 œufs doivent être plus lourds que le poids médian ».

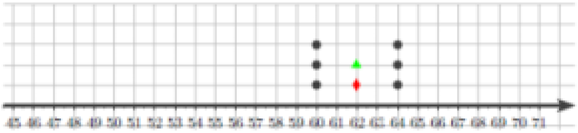
Boite E



Boite F

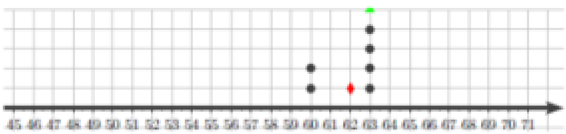


Boite G

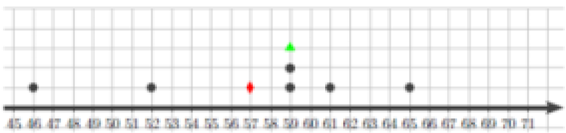


Boite H

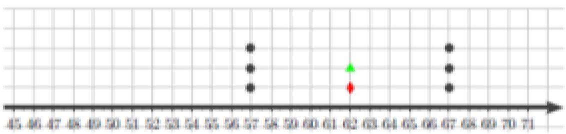
Boite I



Boite J



Boite K



							Poids moyen	Variance	Écart type	Poids médian
Boite E	57	64	66	51	70	64	62	39	6,24	64
Boite F	57	61	52	50	59	63	57	21,67	4,65	58
Boite G	60	64	64	64	60	60	62	4	2	62
Boite H	68	68	59	59	59	59	62	18	4,24	59
Boite I	60	63	63	60	63	63	62	2	1,41	63
Boite J	59	46	59	52	61	65	57	39	6,24	59
Boite K	67	67	57	57	67	57	62	25	5	62

Après cette activité, ils passent à des exercices d’application.

À une autre séance, je leur distribue un « poly » sur l’ajustement affine et la droite de régression par la méthode des moindres carrés.

$$\text{Covariance} = \dfrac{x_1y_1 + \dotsb + x_ny_n}{n} -\bar{x}\bar{y}$$

La droite d'ajustement de y en x obtenue par la méthode des moindres carrés a pour coefficient directeur :

$$m = \dfrac{\text{Covariance}}{\text{Variance des } x_i}$$

et elle passe par le point moyen.

Le coefficient de corrélation linéaire r :

$$r = \dfrac{\text{Covariance}}{\text{Écart type des } x_i \times \text{Écart type des } y_i}$$

$$r = \dfrac{\text{Covariance}}{\sqrt{\text{Variance des } x_i \times \text{Variance des } y_i}}$$

Ils font tous les calculs en utilisant une calculatrice « collège ».

Après l'évaluation (calculatrice « lycée » interdite, calculatrice « collège » prêtée), une séance de TP en salle informatique me permet de leur montrer comment obtenir, à l'aide d'un tableur (Excel en l'occurrence), la moyenne, la variance, l'écart type, la covariance, la droite de régression et le coefficient de corrélation.

Comme en BTS [6], nos étudiants proviennent de baccalauréats très variés : généraux avec ou sans maths, technologiques, professionnels. Avec l'activité des œufs, j'ai réussi à atteindre mes objectifs : parvenir à gérer au mieux cette hétérogénéité et répéter 5 fois le même cours sans m'ennuyer. J'espère que l'utilisation de ces graphiques et la situation concrète des œufs aident mes étudiants à mieux comprendre et mémoriser les différents indicateurs statistiques.

article suivant

retour au sommaire

Les chantiers de pédagogie mathématique n°193 juillet 2022
La Régionale Île-de-France APMEP, 26 rue Duméril, 75013 PARIS

Notes

[1] Institut Universitaire de Technologie

[2] Bachelor Universitaire de Technologie, c'est le DUT qui est devenu le B.U.T.

[3] Annexe 16 de l'Arrêté du 15-4-2022 — JO du 23-04-2022 paru au BO spécial n°4 du 26 mai 2022

[4] Travaux Pratiques

[5] Environnement Numérique de Travail

[6] Brevet de Technicien Supérieur