Rapport SAE 2.04

Exploitation d'une base de données

VAE

AUTEUR
GNANESWARAN ROSHAN

PARTIE MATH
(SECTION 7)

Section 7 : Analyse statistique

1) Voici le tableau du nombre de ventes effectuées en 2022 par mois.

mois	ventes (V)
1	17
2	25
3	21
4	41
5	31
6	34
7	33
8	46
9	41
10	36
12	35
12	22

Nous noterons cette série statistique V.

Pour la moyenne nous avions à additionner toutes les valeurs et diviser par le total de valeur soit ici 12. Ce qui nous donne donc :

V barre = 31,83.

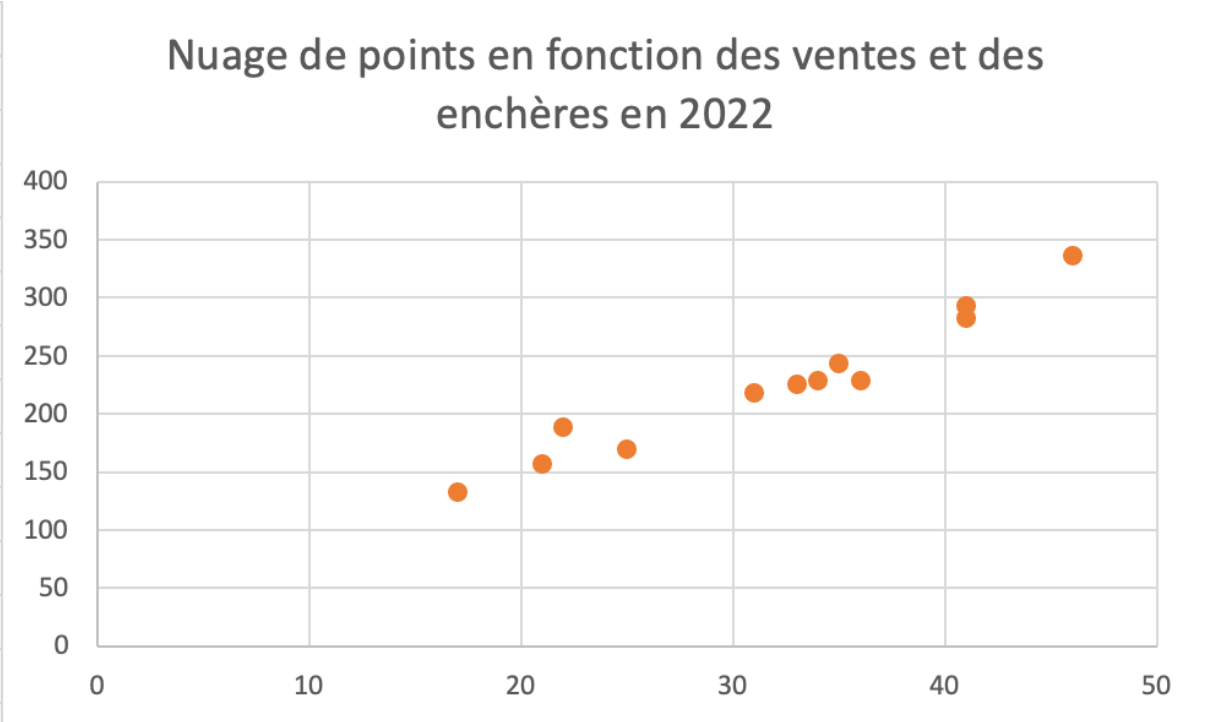
Pour avoir la médiane de V, il nous suffit de trier les valeurs par ordre croissant et de prendre la valeur à la position 6. Ce qui nous donne pour la médiane de V = 33 comme nous pouvions avec le tableau ci-dessous.

mois	nb
1	17
3	21
12	22
2	25
5	31
7	33
6	34
12	35
10	36
4	41
9	41
8	46

Nous allions maintenant extraire les données du nombre d’enchères par mois en 2022, ce qui va nous donner une nouvelle statistique, qu’on notera E.

mois	ventes (V)	encheres (E)
1	17	133
2	25	170
3	21	157
4	41	283
5	31	218
6	34	229
7	33	226
8	46	336
9	41	293
10	36	229
12	35	243
12	22	189

Ce qui nous donne la statistique double (V,E), avec ceci nous allons obtenir un nuage de points.



Nous pouvions constater visuellement une corrélation car les points sont très proches et forme comme une droite croissante.

Nous allons donc calculer le coefficient de corrélation ρV,E du couple (V,E) afin de confirmer notre théorie.

Nous aurions donc besoin des données comme la variance de V, la moyenne de E ainsi que la variance de E. De plus nous arions besoin de la covariance de VE.

La formule de la variance est V= (Σ (x-μ)²) / N.

Variance de V = 73.68
Variance de E = 3176,75

La formule de la covariance est :

$$: Co(X,Y) = \sum_{i=1}^{N} \frac{(X_i - \overline{X})(Y_i - \overline{Y})}{N}$$

Covariance de (V,E) = 470,66

Avec ces nouvelles données nous pouvions calculer le coefficient de corrélation avec la formule :

$$\rho(X;Y) = \frac{\Sigma(X-\bar{X}).(Y-\bar{Y})}{\sqrt{\Sigma(X-\bar{X})^2}.\sqrt{\Sigma(Y-\bar{Y})^2}} = \frac{Cov(X;Y)}{\sqrt{V(X)}.\sqrt{V(Y)}}$$

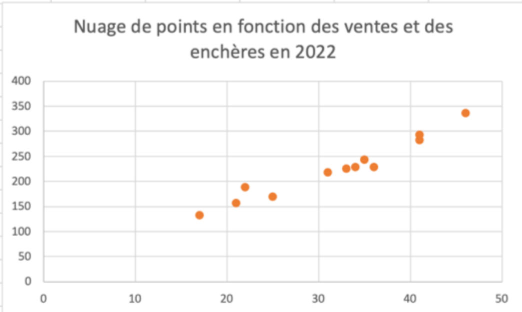
Soit le coefficient de corrélation pour la statique double est de 0,97. On peut donc en déduire que les données se corrèle très bien car le coefficient est très proche de 1. Ce qui semble normal car plus y’a des enchères plus y’aura des ventes.

Avec les données obtenues on peut anticiper que sur un mois dont le nombre d’enchères et de 285. On peut donc en déduire que le nombre de ventes sera de 37.

ANNEXE :

Tableau Excel pour les données :

	A	B	C
1	mois	ventes (V)	encheres (E)
2	1	17	133
3	2	25	170
4	3	21	157
5	4	41	283
6	5	31	218
7	6	34	229
8	7	33	226
9	8	46	336
10	9	41	293
11	10	36	229
12	12	35	243
13	12	22	189
14			
15	Moyenne	31,8333333	225,5
16	Variance	73,6388889	3176,75
17	Mode	41	
18	Mediane	33	
19			
20	Covariance(V,E)	470,666667	
21	Coeff de Corellation	0,97312428	
22			



Le fichier. Xlsx est aussi disponible dans le zip dans le dossier Math afin d'obtenir les formules. J'ai fait le choix d'utiliser Excel car je le trouve plus intuitif que Scilab et plus simple d'utilisation. Cependant, vous retrouveriez également le fichier. Sci avec les formules dans le même dossier

Version Scilab :

Le nuage de point :