SAE2 .04 ANALYSE DE DONNÉES

Site Vente aux Enchères

RAVIART ALEXANDRE

2022-2023

7 Analyse statistique

1) Le nombre de ventes effectuées en 2022

SELECT MONTH(finve) mois, YEAR(finve) annee, count(distinct idve) from VENTE natural join ENCHERIR where idSt = 4 and YEAR(finve) = 2022 and YEAR(debutve) = 2022 group by mois, annee;

Mois	Janvie	Février	Mars	Avril	Mai	Juin	Juillet	Aout	Septembre	Octobre	Novemebre	Décembre
V	14	25	22	40	31	31	35	46	40	38	32	26

<u>Moyenne de V</u> = (14+25+22+40+31+31+35+46+40+38+22+26)* 1/12 = 31.666

<u>Mediane de V</u> = (31+32)/2 = 31.5

Cela signifie que la moitié des valeurs sont plus grandes que 31.5.

rang	1	2	3	4	5	6	7	8	9	10	11	12
V	14	22	25	26	31	31	32	35	38	40	40	46

<u>Le mode de V</u> : 31

La valeur 31 apparaît deux fois ce qui est plus que tout autre valeur dans cette série.

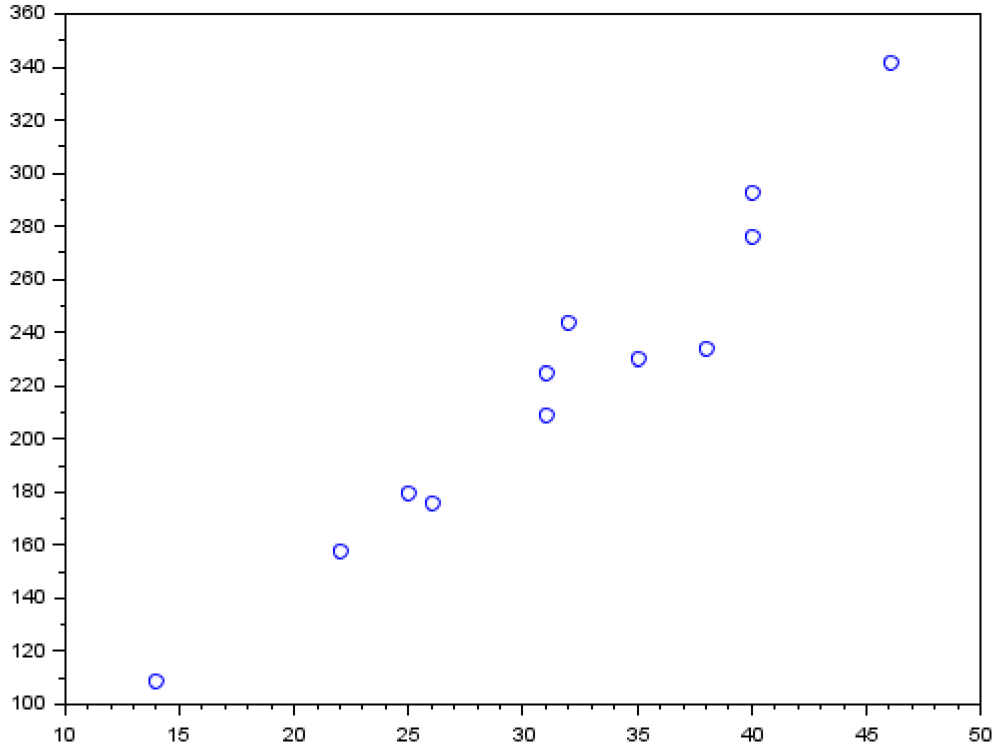
2) Le nombre d'enchères de 2022 par mois

La même requête que précedemment mais sans le distinct dans la close du count().

Mois	Janvie	Février	Mars	Avril	Mai	Juin	Juillet	Aout	Septembre	Octobre	Novemebre	Décembre
Ɛ	109	180	158	276	225	209	230	342	293	234	244	176

Tracer un nuage de points et faire apparaitre une corrélation linéaire pourrait indiquer une relation entre le nombre de ventes et le nombre d'enchères.

Nuage de point : scatter(V, Ɛ) =



A première vue, les points adopte des similarités.

Pour vérifier cela, il faut calculer le coefficient de corrélation, et pour cela, nous avons besoin de calculer la variance des statistiques puis la covariance. Qui signifie calculer la différence de chaque donnée par rapport à la moyenne, élevée au carré :

<u>Calcul des variances :</u>

varV=mean((V-moyenneV).*(V-moyenneV))

Var(X) = $(14 - 31.66)^2 + (25 - 31.66)^2 + (22 - 31.66)^2 + (40 - 31.66)^2 + (31 - 31.66)^2 + (31 - 31.66)^2 + (35 - 31.66)^2 + (46 - 31.66)^2 + (40 - 31.66)^2 + (38 - 31.66)^2 + (32 - 31.66)^2 + (26 - 31.66)^2/12 = 73.222$

Mean(Ɛ) = 223

Var(Ɛ) =   var Ɛ =mean((Ɛ -moyenne Ɛ).*(Ɛ -moyenne Ɛ)) = 3663

La variance de la deuxième statistique indique des valeurs très différentes les une des autres.

<u>Calcul de la covariance :</u>

mean((V -moyenne V).*(Ɛ -moyenne Ɛ)) = 502.83

<u>Calcul du coefficient de corrélation :</u>

ρV, Ɛ = covV Ɛ /sqrt(V*var Ɛ) = 503.83/sqrt(73.222*3663.333) = 0.97

Cela signifie qu'il y a une tendance à une augmentation du nombre d'enchères en fonction de l'augmentation du nombre de ventes. 97 % de la variation observée dans le nombre d'enchères peut être expliquée par la variation dans le nombre de ventes.

Traçage de la droite ;

a = covV Ɛ /varV;

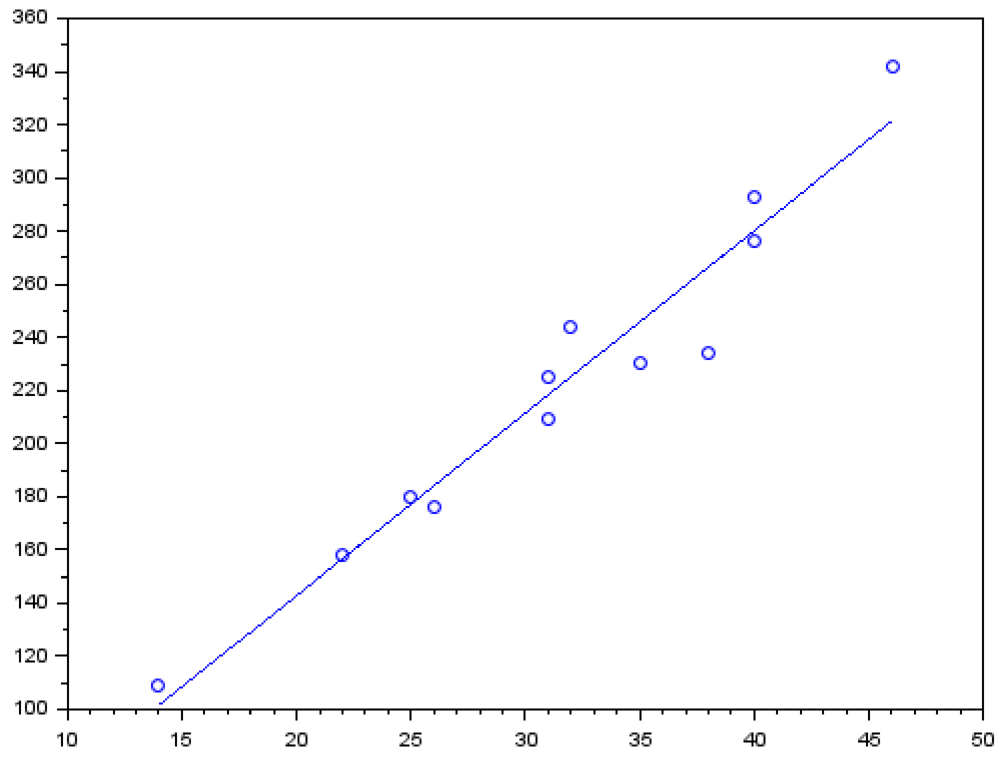
b = moyenne Ɛ - a*moyenne Ɛ;

y = a*V + b;

y-b/a

plot(V, Ɛ,"o")

plot(V,y)



Nous voyons que les points ont tous la distance la plus courte à la droite faible, comme en témoigne le coefficient de corrélation de 0.97.

Nous pouvons utiliser cette droite de régression pour prédire la valeur de V à partir de la valeur de Ɛ.

y = 285 ; a = 6.87 ; b = 5.54

V = (y - b) / a

V = (285 - 5.54) / 6.87 = 40.67

On anticipe et estime 40.67 vente pour 285 enchères.

Contenu du fichier Scilab :

```
// données

X = [14;25;22;40;31;31;35;46;40;38;32;26];
Y = [109;180;158;276;225;209;230;342;293;234;244;176];

// calcul des statistiques
moyenneX = mean(X);
varX = mean((X - moyenneX).^2);
moyenneY = mean(Y);
varY = mean((Y - moyenneY).^2);
covXY = mean((X - moyenneX).*(Y - moyenneY));
pente = covXY/varX;
ordonnee = moyenneY - pente*moyenneX;

// tracé de la droite de régression
plot(X,Y,"o")
plot(X, pente*X + ordonnee)
```