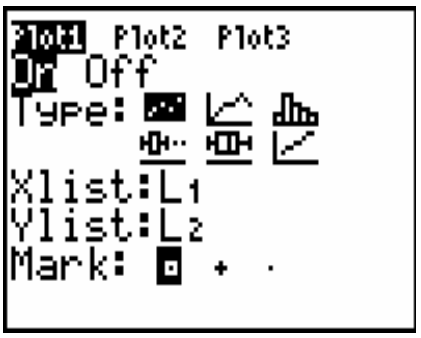
T.P. 7 - Exercice supplémentaire 1
Régression linéaire simple (Corrigé)

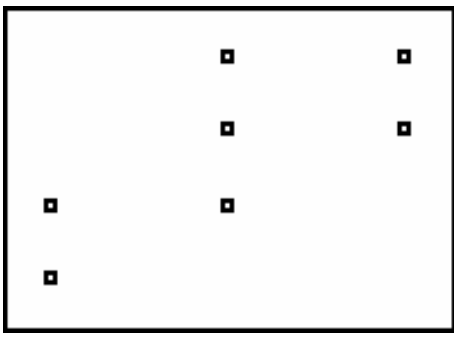
On a administré un test de lecture à 12 enfants âgés de 7, 8 et 9 ans. Voici les résultats obtenus par ces sujets :

i	Variable X Âge	Variable Y Note au test
1	7	6
2	8	8
3	9	8
4	7	7
5	9	9
6	8	8
7	7	6
8	9	9
9	8	7
10	9	8
11	8	9
12	7	7

1. Représentez cette série statistique par un nuage de points.

Réponse :



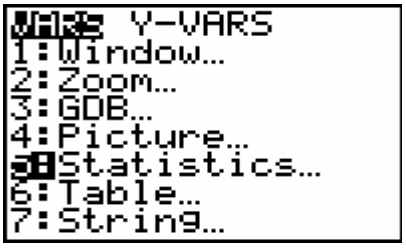


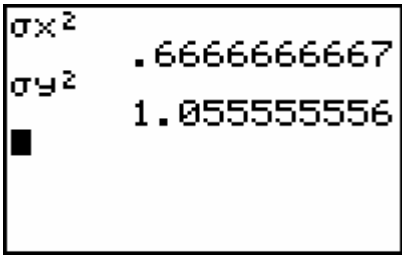
2. Calculez à l'aide de votre machine la moyenne, l'écart –type et la variance de la variable X et de la variable Y:

Réponse :

Moyenne de X = 8
Écart type de X =0,82
Moyenne de Y=7,67
Écart type de Y =1,07

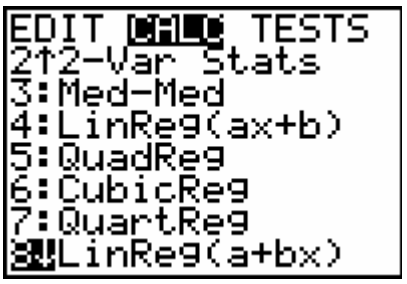
Pour la variance on ouvre Statistics dans le menu Vars :

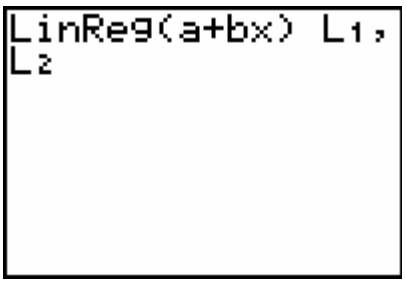


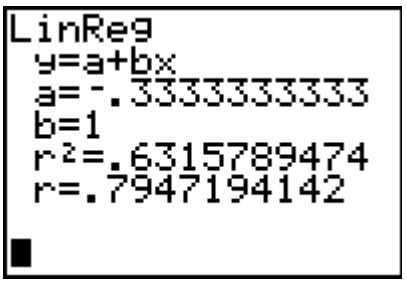


3. Déterminez l'équation de la droite de régression de Y en X.

Réponse :





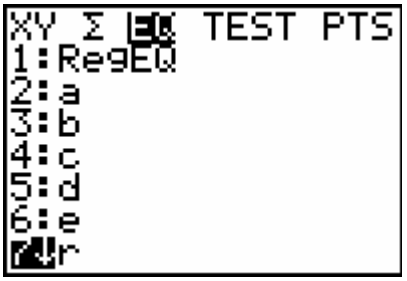


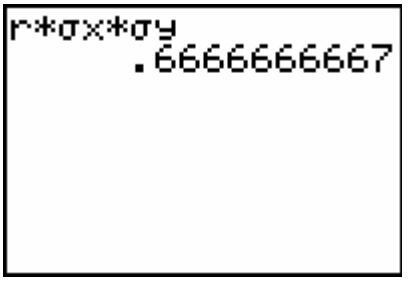
4. Déterminez la covariance entre X et Y.

Réponse :

$$\mathbf{cov}_{XY} = r_{XY} . S_X . S_Y$$

On ouvre Statistics dans le menu VARS pour chercher la valeur de r_{XY} , S_x et S_y :



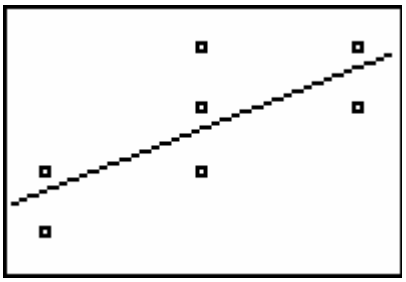


5. Déterminez le résultat prédit au test pour un enfant âgé de 10 ans. Représentez la droite de régression sur le nuage de points.

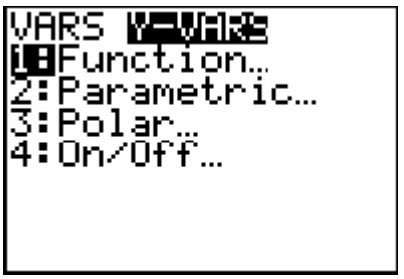
Réponse :

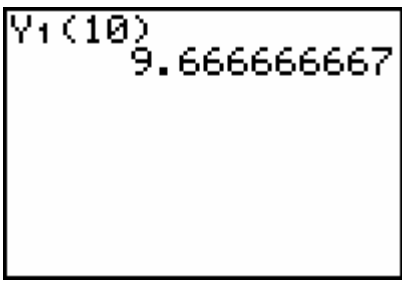
Nous allons introduire l'équation de la droite de régression dans grâce à la commande Y1 :

Le graphique :



Pour trouver Y1 on doit ouvrir VARS, placer le curseur sur Y-VARS et choisir function puis Y1 :





6. Interprétez vos résultats à partir des mesures calculées et à partir du graphique que vous avez tracé.

Réponse :

L'examen du graphique de dispersion nous indique qu'il y a une association positive entre la variable X, âge de l'enfant, et la variable Y, score obtenu au test. En effet, on constate que presque la totalité des points se trouvent soit dans le quadrant inférieur gauche soit dans le quadrant supérieur droit. Cette interprétation est renforcée par cov_{XY} qui est positive.

Nous observons aussi (et cette valeur est plus facile à interpréter que celle de cov_{XY}) que la valeur de r_{XY} est positive et que d'après le diagramme de dispersion, les points ont tendance à s'aligner selon une droite de pente positive.

Le coefficient de détermination est de 0,63. Ceci signifie que seulement 63% des variations de la variable Y (score au test) entre individus peuvent être expliquées par l'influence linéaire de X sur Y.

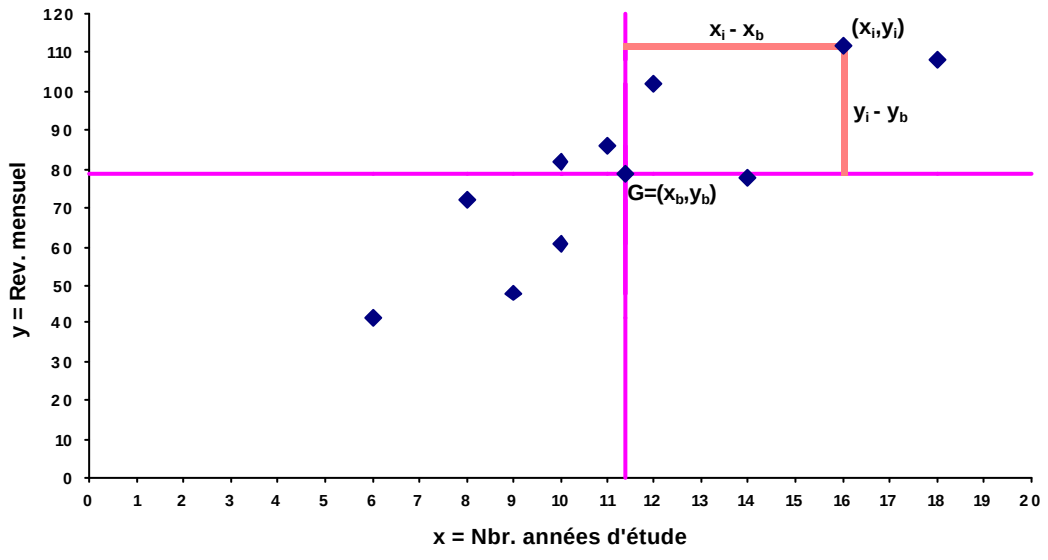
T.P. 7 – Exercice Supplémentaire 2
Série statistique bivariée – Nuages de points - Covariance
(Corrigé)

Considérons un échantillon de 10 employés du même age, d'une entreprise. Soit X le nombre d'années d'études effectuées et Y le revenu mensuel (en milliers de francs) touché par chacun d'entre eux. Les observations sont contenues dans le tableau suivant :

	X_i	Y_i	$X_i - \overline{X}$	$\left(X_i - \overline{X}\right)^2$	$Y_i - \overline{Y}$	$\left(Y_i - \overline{Y}\right)^2$	$\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)$
1	6	41	**-5.4**	**29.16**	**-38**	**1444**	**205.20**
2	8	72	**-3.4**	**11.56**	**-7**	**49**	**23.80**
3	9	48	**-2.4**	**5.76**	**-31**	**961**	**74.40**
4	10	82	**-1.4**	**1.96**	**3**	**9**	**-4.20**
5	10	61	**-1.4**	**1.96**	**-18**	**324**	**25.20**
6	11	86	**-0.4**	**0.16**	**7**	**49**	**-2.80**
7	12	102	**0.6**	**0.36**	**23**	**529**	**13.80**
8	14	78	**2.6**	**6.76**	**-1**	**1**	**-2.60**
9	16	112	**4.6**	**21.16**	**33**	**1089**	**151.80**
10	18	108	**6.6**	**43.56**	**29**	**841**	**191.40**
$\sum_{i=1}^{n}$	**114**	**790**	**0**	**122.40**	**0**	**5296**	**676**

1. Représentez cette série par un nuage de points.

Graphique :



2. Complétez les cellules vides du tableau des données.

3. Calculez manuellement les valeurs suivantes :

   NB : Vous pouvez vous baser sur les sommes calculées au point 2.

<table>
<tr>
<td>Réponse :

Moyenne de X : $\overline{X} = \mathbf{11,4}$

Variance de X :

$S_X^2 = \frac{\mathbf{122,40}}{\mathbf{10}} = \mathbf{12,24}$

Écart type de X
$S_X = \sqrt{\mathbf{12,24}} = \mathbf{3,5}$</td>
<td>

Moyenne de Y : $\overline{Y} = \mathbf{79}$

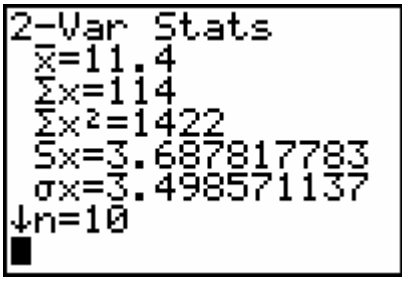
Variance de Y

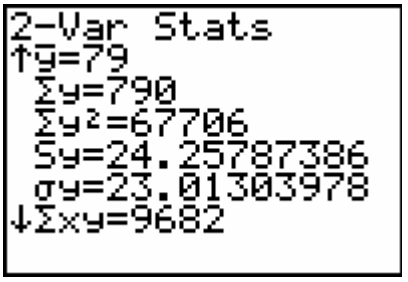
$S_Y^2 = \frac{\mathbf{5296}}{\mathbf{10}} = \mathbf{529,6}$

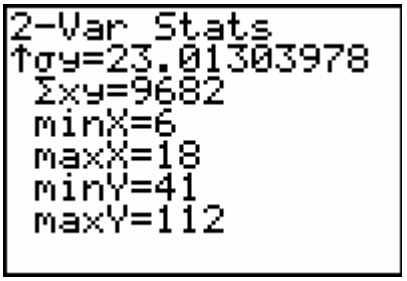
Écart type de Y
$S_Y = \sqrt{\mathbf{529,6}} = \mathbf{23,01}$</td>
</tr>
<tr>
<td colspan="2">Covariance de X et Y :
$\mathbf{cov}_{XY} = S_{XY} = \frac{\mathbf{676}}{\mathbf{10}} = \mathbf{67,6}$</td>
</tr>
</table>

4. Vérifiez vos résultats à l'aide de la TI 84 +

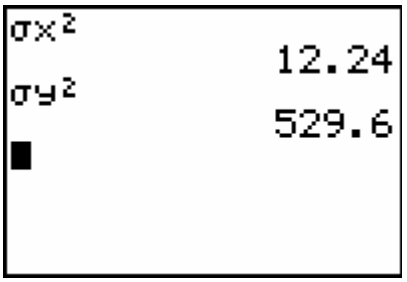
Réponse :







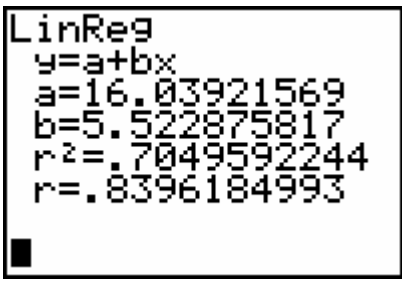
Pour la variance : VARS STATISTICS :



5. Sur le graphique du point 1, tracez les droites définissant les 4 quadrants dans le plan.

6. Déterminez l'équation de la droite de régression de Y en fonction de X.

Réponse :



$$\hat{Y} = \mathbf{16{,}04} + \mathbf{5{,}52}X$$

7. Interprétez vos résultats sur base du graphique que vous avez tracé et des valeurs que vous avez calculées.

Réponse :

Les valeurs de X et de Y sont très différentes entre elles : elles sont mesurées dans des unités différentes. Cependant nous pouvons évaluer la force du lien qui les unit. Nous voyons que la covariance entre les deux variables est positive : cela signifie que plus le nombre d'années d'études effectuées est élevé, plus le revenu mensuel des individus l'est.

En regardant le graphique, nous constatons en effet que de façon générale, les sujets qui se situent en dessous de la moyenne pour la variable X, se situent également en dessous de la moyenne pour la variable Y ; de même, si un sujet se trouve au-delà de la moyenne pour la variable X, il se situe également au-delà de la moyenne pour la variable Y. Ceci s'observe grâce au fait que sur le graphique de dispersion, les points se trouvent soit dans le quadrant inférieur gauche, soit dans le quadrant supérieur droit. Trois sujets sont dans les deux autres quadrants : ce sont des sujets qui sont proches de la moyenne pour les deux variables, elles sont proches du centre de gravité.

Le coefficient de détermination est de 0,70. Ceci signifie que seulement 70% des variations de la variable Y (le revenu) entre individus peuvent être expliquées par l'influence linéaire de X sur Y.