On rappelle les définitions et résultats suivants :

- Si $(v_i)_{1\leqslant i\leqslant n}$ est une série statistique, la moyenne et la variance empiriques, notées respectivement $\overline{v}$ et s_v^2, sont données par : $\overline{v} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} v_i$ et $s_v^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(v_i-\overline{v})^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} v_i^2 - \overline{v}^2$.
- Si $(v_i)_{1\leqslant i\leqslant n}$ et $(w_i)_{1\leqslant i\leqslant n}$ sont deux séries statistiques, la covariance empirique de la série double $(v_i, w_i)_{1\leqslant i\leqslant n}$, notée $\mathrm{Cov}(v,w)$, est donnée par : $\mathrm{Cov}(v,w) = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(v_i-\overline{v})(w_i-\overline{w}) = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} v_i w_i - \overline{v}\,\overline{w} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(v_i-\overline{v})w_i$.

9.a) Montrer que pour tout $i \in [\![1,n]\!]$, la variable aléatoire T_i suit la loi normale $\mathcal{N}(au_i+b,\, \sigma^2)$.

b) Les variables aléatoires $T_1, T_2, \ldots, T_n$ sont-elles indépendantes ?

Pour tout $i \in [\![1,n]\!]$, soit φ_i la densité continue sur $\mathbf{R}$ de T_i : $\forall\, d \in \mathbf{R}$, $\varphi_i(d) = \dfrac{1}{\sigma\sqrt{2\pi}}\exp\left(-\dfrac{1}{2\sigma^2}\left(d-(au_i+b)\right)^2\right)$.

Soit $\mathcal{F}$ l'ouvert défini par $\mathcal{F} =]\,0,1\,[\times\mathbf{R}$ et M la fonction de $\mathcal{F}$ dans $\mathbf{R}$ définie par : $M(a,b) = \ln\left(\displaystyle\prod_{i=1}^{n}\varphi_i(t_i)\right)$.

On suppose que : $0 < \mathrm{Cov}(u,t) < s_u^2$.

10.a) Calculer le gradient $\nabla(M)(a,b)$ de M en tout point $(a,b)\in\mathcal{F}$.

b) En déduire que M admet sur $\mathcal{F}$ un unique point critique, noté $(\hat{a},\hat{b})$.

c) Exprimer $\hat{a}$ et $\hat{b}$ en fonction de $\mathrm{Cov}(u,t)$, s_u^2, $\overline{t}$ et $\overline{u}$.

($\hat{a}$ et $\hat{b}$ sont les estimations de a et b par la méthode dite du maximum de vraisemblance)

11.a) Soit $\nabla^2(M)(a,b)$ la matrice hessienne de M en $(a,b)\in\mathcal{F}$. Montrer que : $\nabla^2(M)(a,b) = -\dfrac{n}{\sigma^2}\begin{pmatrix} s_u^2+\overline{u}^2 & \overline{u} \\ \overline{u} & 1\end{pmatrix}$.

b) En déduire que M admet au point $(\hat{a},\hat{b})$ un maximum local.

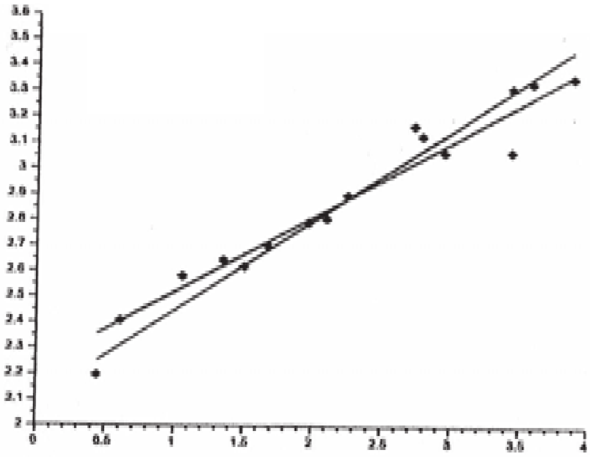
12. Soit (h,k) un couple de réels non nuls. Calculer $M(\hat{a}+h, \hat{b}+k) - M(\hat{a},\hat{b})$.
En déduire que M admet en $(\hat{a},\hat{b})$ un maximum global.

13. On rappelle qu'en *Scilab*, les commandes `variance` et `corr` permettent de calculer respectivement la variance d'une série statistique et la covariance d'une série statistique double.
Si $v = (v_i)_{1\leqslant i\leqslant n}$ et $w = (w_i)_{1\leqslant i\leqslant n}$ sont deux séries statistiques, alors la variance de $(v_i)_{1\leqslant i\leqslant n}$ est calculable par `variance(v)` et la covariance de $(v_i, w_i)_{1\leqslant i\leqslant n}$ est calculable par `corr(v,w,1)`.

On a relevé pour $n = 16$ entreprises qui produisent le bien considéré à l'époque donnée, les deux séries statistiques $(u_i)_{1\leqslant i\leqslant 16}$ et $(t_i)_{1\leqslant i\leqslant 16}$ reproduites dans les lignes (1) et (2) du code *Scilab* suivant dont la ligne (5) est incomplète :

```
(1)  u=[1.06,0.44,2.25,3.88,0.61,1.97,3.43,2.10,1.50,1.68,2.72,1.35,2.94,2.78,3.43,3.58];

(2)  t=[2.58,2.25,2.90,3.36,2.41,2.79,3.32,2.81,2.62,2.70,3.17,2.65,3.07,3.13,3.07,3.34]

(3)  plot2d(u,t,-4) // -4 signifie que les points sont représentés par des losanges.

(4)  plot2d(u,corr(u,t,1)/variance(u)*u+mean(t)-corr(u,t,1)/variance(u)*mean(u))// équation de la
     droite de régression de t en u.

(5)  plot2d(u,..............................................................)// équation de la
     droite de régression de u en t.
```

Le code précédent complété par la ligne (5) donne alors la figure suivante :



a) Compléter la ligne (5) du code permettant d'obtenir la figure précédente (*on reportera sur sa copie, uniquement la ligne (5) complétée*).

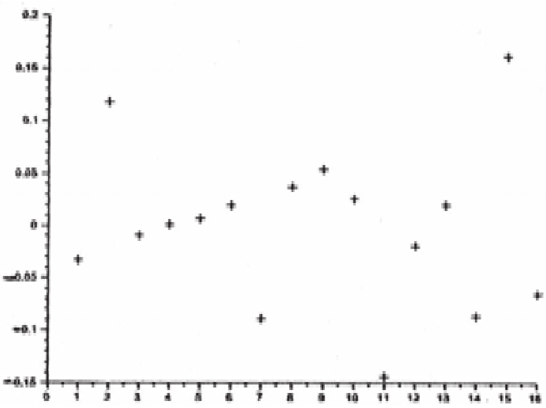
b) Interpréter le point d'intersection des deux droites de régression.

c) Estimer graphiquement les moyennes empiriques $\overline{u}$ et $\overline{t}$.

d) Le coefficient de corrélation empirique de la série statistique double $(u_i, t_i)_{1 \leqslant i \leqslant 16}$ est-il plus proche de -1, de 1 ou de 0 ?

e) On reprend les lignes (1) et (2) du code précédent que l'on complète par les instructions (6) à (11) qui suivent et on obtient le graphique ci-dessous :

```
(6)    a0=corr(u,t,1)/variance(u)
(7)    b0=mean(t)-corr(u,t,1)/variance(u)*mean(u)
(8)    t0=a0*u+b0
(9)    e=t0-t
(10)   p=1:16
```



```
(11)   plot2d(p,e,-1) // -1 signifie que les points sont représentés par des symboles d'addition.
```

Que représente ce graphique ? Quelle valeur peut-on conjecturer pour la moyenne des ordonnées des 16 points obtenus sur le graphique ? Déterminer mathématiquement la valeur de cette moyenne.

14. Pour tout entier $n \geqslant 1$, on pose : $A_n = \dfrac{1}{n s_u^2} \displaystyle\sum_{i=1}^{n} (u_i - \overline{u})\, T_i$. On suppose que le paramètre σ^2 est connu.

a) Calculer l'espérance $E(A_n)$ et la variance $V(A_n)$ de la variable aléatoire A_n. Préciser la loi de A_n.

b) On suppose que a est un paramètre inconnu. Soit α un réel donné vérifiant $0 < \alpha < 1$.
On note Φ la fonction de répartition de la loi normale centrée réduite et d_α le réel tel que $\Phi(d_\alpha) = 1 - \dfrac{\alpha}{2}$ ·
Déterminer un intervalle de confiance du paramètre a au niveau de confiance $1 - \alpha$.