Institut de la Nutrition, de l'Alimentation et des Technologies Agro-Alimentaires
Département de Technologie Alimentaire
1e année Licence Sciences Alimentaires
Matière : **Statistiques Appliquées**

TD 03 : **Statistiques descriptives bivariées**

Exercice 1

Soit les variables X et Y donnes dans le tableau suivant :

X	1	3	4	6	8	9	11	14	16	19
Y	1	2	4	4	5	7	8	9	10	12

1. Calculer la moyenne et l'écart-type pour chaque variable ;
2. Calculer la covariance et le coefficient de corrélation de X et Y ;
3. Existe-il une relation entre X et Y ? quel est le type de cette relation ?
4. Déterminer l'équation de régression linéaire ;
5. Tracer le nuage de points représentant les données ;
6. Déterminer le point moyen et représenter-le dans le nuage de points ;
7. Tracer la droite de régression sur le nuage de points ;
8. Calculer les valeurs prédites ;
9. Calculer les variabilités entre :

   - les valeurs observées et leur moyenne.

   - les valeurs observées et les valeurs prédites.

   - les valeurs prédites et la moyenne des valeurs observées.

Exercice 2

Une expérience consiste à mesurer l'activité enzymatique après l'addition d'un additif X. Cette expérience est menée dans 10 tubes à essai et dans des conditions de laboratoire. Les résultats obtenus sont les suivants :

Tube	1	2	3	4	5	6	7	8	9	10
Activité enzymatique (UI)	15,9	14,9	15,1	14,9	14,9	14,7	15	14,8	14,8	14,8
Quantité du produit (mg)	8,2	11,3	11,9	11,1	11,5	11	10,9	10,9	10,2	10,9

Les mêmes questions que l'exercice précédent.

TD 03 / Corrigé-type / Exercice 1

1. Calculer les paramètres statistiques univariés pour chaque variable ;

$$\bar{x}\,;\bar{y}\,;\ x_i-\bar{x}\,;\ y_i-\bar{y}\,;\ (x_i-\bar{x})\ ;\ (y_i-\bar{y})^2\ ;\ (x_i-\bar{x})(y_i-\bar{y})$$

X	Y	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
1	1	-8,1	-5,2	65,61	27,04	42,12
3	2	-6,1	-4,2	37,21	17,64	25,62
4	4	-5,1	-2,2	26,01	4,84	11,22
6	4	-3,1	-2,2	9,61	4,84	6,82
8	5	-1,1	-1,2	1,21	1,44	1,32
9	7	-0,1	0,8	0,01	0,64	-0,08
11	8	1,9	1,8	3,61	3,24	3,42
14	9	4,9	2,8	24,01	7,84	13,72
16	10	6,9	3,8	47,61	14,44	26,22
19	12	9,9	5,8	98,01	33,64	57,42
$\bar{x} = 9,10$	$\bar{y} = 6,20$	////	///	**Σ = 312,9**	**Σ = 115,60**	**Σ = 187,8**

2. Calculer la covariance et le coefficient de corrélation de x et y ;

$$Cov\,(x,y) = \frac{\sum(x_i-\bar{x})\times(y_i-\bar{y})}{N} = \frac{187{,}8}{10} = \mathbf{18,78}\ ;$$

Coef. Corrélation : $r_{cal} = \frac{Coc\,(x,y)}{\sigma_x \times \sigma_y}$; Sachant que :

$$\sigma_x = \sqrt{\frac{\sum(x_i-\bar{x})^2}{N}} = \sqrt{\frac{312{,}9}{10}} = \mathbf{5,59}\ ; \qquad \sigma_y = \sqrt{\frac{\sum(y_i-\bar{y})^2}{N}} = \sqrt{\frac{115{,}60}{10}} = \mathbf{3,40}\ ;\ r_{cal} = \frac{18{,}78}{5{,}59\times 3{,}40} = \mathbf{0,9874}$$

3. Existe-il une relation entre X et Y ? que constatez-vous ?

   Comparer r_{cal} # r_{the}, (r_{the} Coef théorique)

   - Si **N≤100 :** le r_{tab} est à lire sur une table du coefficient de corrélation, avec ddl = N – 2
   - Si **N>100** : $r_{th} = Z_\alpha . S_m = Z_\alpha . \frac{1}{\sqrt{N-1}}$ avec $Z_\alpha = 1{,}96\ si\ \alpha = 5\ \%\ ou\ Z_\alpha = 2{,}58\ si\ \alpha = 1\%$
     - si $r_{cal} > r_{tab}$ ➔ il y'a une relation entre x et y
     - si $r_{cal} < r_{tab}$ ➔ il n'y a pas de relation entre x et y.

   Puisque N = 10 < 100 → ddl = N-2 = 8 ; α = 5% => **r_{the} = 0,6318 ;** α = 1% => **r_{the} = 0,7645.**

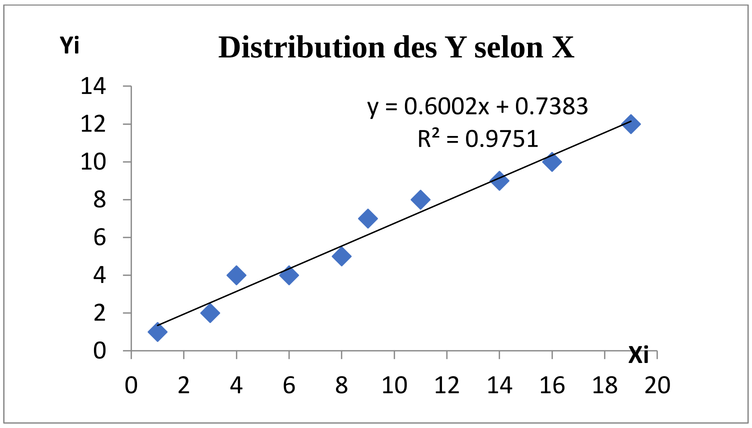
Nous constatons que $r_{cal} > r_{the}$, cela montre qu'il existe une relation entre X et Y.

4. Quel est le type de relation qui existe entre les 2 variables ? **Corrélation positive** <u>forte</u>
5. Déterminer l'équation de régression de type Y = a X + b   ($y_i = a.x_i + b + e_i$) :
   - « a » représente la pente de la droite de régression.

     $$a = \frac{cov(x,y)}{var(X)} = \frac{18{,}78}{5{,}59} = \mathbf{0,60} \quad \text{ou}\ a = \frac{\sum(x_i-\bar{x}).(y_i-\bar{y})}{\sum(x_i-\bar{x})^2} = \frac{187{,}8}{312{,}9} = \mathbf{0,60}$$

   - « b » représente l'ordonnée à l'origine (valeur de Y quand X = 0). On la détermine à partir du point central du nuage de point qui correspond aux moyennes des deux paramètres, ce qui coïncide avec : $b = \bar{y} - a\,\bar{x} = 6{,}20 - 0{,}60 \times 9{,}1 = 0{,}74$
   - L'équation de la droite de régression est donc : **Y = 0,60 X + 0,74.**

6. Nuage de points / Point moyen **m (9,1 - 6,20)** / Droite de régression.



7. Valeurs de prédiction ou valeurs prédites $(\hat{Y})$ et calcul des variabilités.

X	**Y**	$\hat{Y}$	$(Y-\bar{Y})^2$	$(Y-\hat{Y})^2$	$(\hat{Y}-\bar{Y})^2$
1	1	1,34	27,04	0,1156	23,6196
3	2	2,54	17,64	0,2916	13,3956
4	4	3,14	4,84	0,7396	9,3636
6	4	4,34	4,84	0,1156	3,4596
8	5	5,54	1,44	0,2916	0,4356
9	7	6,14	0,64	0,7396	0,0036
11	8	7,34	3,24	0,4356	1,2996
14	9	9,14	7,84	0,0196	8,6436
16	10	10,34	14,44	0,1156	17,1396
19	12	12,14	33,64	0,0196	35,2836
$\bar{x} = 9,1$	$\bar{y} = 6,2$	$\hat{y} = 6,2$	**Σ = 115,60**	**Σ = 2,884**	**Σ = 112,664**

- Variabilité entre les valeurs observées et leur moyenne

$$\text{Var (obs-moy)} = \sum(y_i - \bar{y})^2/(N) = 11,56$$

- Variabilité entre les valeurs observées et les valeurs prédites

$$\text{Var (obs-pred)} = \frac{\sum(y_i - \hat{y})}{N} = 0,2884$$

- Variabilité entre les valeurs prédites et la moyenne des valeurs observées

$$\text{Var (pred-moy)} = \sum(\hat{y} - \bar{y})^2/(N) = 11,2664$$

8. Que constatez-vous ?

Var (obs-moy) = Var (obs-pred) + Var (pred-moy)

TD 03 / Corrigé-type / Exercice 2

2. Calculer les paramètres statistiques univariés pour chaque variable ;

$$\bar{x}\,;\bar{y}\,;\ x_i-\bar{x}\,;\ y_i-\bar{y}\,;\ (x_i-\bar{x})\ ;\ (y_i-\bar{y})^2\ ;\ (x_i-\bar{x})(y_i-\bar{y})$$

X	Y	$x_i-\bar{x}$	$y_i-\bar{y}$	$(x_i-\bar{x})^2$	$(y_i-\bar{y})^2$	$(x_i-\bar{x})(y_i-\bar{y})$
15,9	8,2	0,92	-2,59	0,8464	6,7081	-2,3828
14,9	11,3	-0,08	0,51	0,0064	0,2601	-0,0408
15,1	11,9	0,12	1,11	0,0144	1,2321	0,1332
14,9	11,1	-0,08	0,31	0,0064	0,0961	-0,0248
14,9	11,5	-0,08	0,71	0,0064	0,5041	-0,0568
14,7	11	-0,28	0,21	0,0784	0,0441	-0,0588
15	10,9	0,02	0,11	0,0004	0,0121	0,0022
14,8	10,9	-0,18	0,11	0,0324	0,0121	-0,0198
14,8	10,2	-0,18	-0,59	0,0324	0,3481	0,1062
14,8	10,9	-0,18	0,11	0,0324	0,0121	-0,0198
$\bar{x}=14,98$	$\bar{y}=10,79$	////	///	**Σ = 1,06**	**Σ = 9,23**	**Σ = -2,36**

9. Calculer la covariance et le coefficient de corrélation de x et y ;

$$Cov\,(x,y)=\frac{\sum(x_i-\bar{x})\times(y_i-\bar{y})}{N}=\frac{-2{,}362}{10}=\mathbf{-0,24}\ ;$$

Coef. Corrélation : $r_{cal}=\frac{Cov\,(x,y)}{\sigma_x\times\sigma_y}$; Sachant que :

$$\boldsymbol{\sigma_x}=\sqrt{\frac{\sum(x_i-\bar{x})^2}{N}}=\sqrt{\frac{1{,}056}{10}}=\mathbf{0,32}\ ;\qquad \boldsymbol{\sigma_y}=\sqrt{\frac{\sum(y_i-\bar{y})^2}{N}}=\sqrt{\frac{9{,}229}{10}}=\mathbf{0,96}\ ;\quad r_{cal}=\frac{-0{,}24}{0{,}32\times0{,}96}=-\mathbf{0,76}$$

10. Existe-il une relation entre X et Y ? que constatez-vous ?

Comparer r_{cal} # r_{the}, (r_{the} Coef théorique)

- à partir de la table du coefficient de corrélation (ddl = N – 2 = 8 et α = 0,05), on lit que r_{tab}=
  - si $r_{cal} > r_{tab}$ ➔ il y'a une relation entre x et y
  - si $r_{cal} < r_{tab}$ ➔ il n'y a pas de relation entre x et y.

Nous constatons que $r_{cal} > r_{the}$, cela montre qu'il existe une relation entre X et Y.

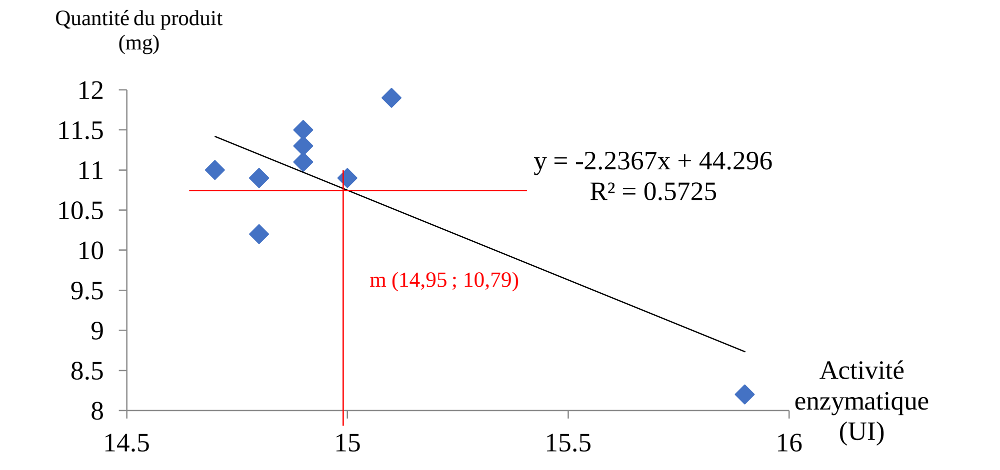
11. Quel est le type de relation qui existe entre les 2 variables ? **Corrélation négative**

12. Déterminer l'équation de régression de type Y = a X + b ($y_i = a.x_i + b + e_i$) :
    - « a » représente la pente de la droite de régression.

$$a=\frac{cov(x,y)}{var(X)}=\frac{-0{,}24}{0{,}11}=\mathbf{-2,18\ (ou\ -2,24)}\quad \text{ou}\ a=\frac{\sum(x_i-\bar{x}).(y_i-\bar{y})}{\sum(x_i-\bar{x})^2}=\frac{-2{,}362}{1{,}056}=\mathbf{-2,18}$$

    - « b » représente l'ordonnée à l'origine (valeur de Y quand X = 0). On la détermine à partir du point central du nuage de point qui correspond aux moyennes des deux paramètres, ce qui coïncide avec : $b=\bar{y}-a\,\bar{x}=10{,}79+2{,}18\times14{,}98=43{,}47$
    - L'équation de la droite de régression est donc : **Y = -2,18 X + 43,47.**

13. Nuage de points / Point moyen **m (14,98 - 10,79)** / Droite de régression.



Quantité du produit en fonction de l'activité enzymatique

14. Valeurs de prédiction ou valeurs prédites $(\hat{Y})$ et calcul des variabilités.

X	**Y**	$\hat{Y}$	$(Y-\bar{Y})^2$	$(Y-\hat{Y})^2$	$(\hat{Y}-\bar{Y})^2$
15,9	8,2	8,808	6,7081	0,37	3,93
14,9	11,3	10,988	0,2601	0,10	0,04
15,1	11,9	10,552	1,2321	1,82	0,06
14,9	11,1	10,988	0,0961	0,01	0,04
14,9	11,5	10,988	0,5041	0,26	0,04
14,7	11	11,424	0,0441	0,18	0,40
15	10,9	10,77	0,0121	0,02	0,00
14,8	10,9	11,206	0,0121	0,09	0,17
14,8	10,2	11,206	0,3481	1,01	0,17
14,8	10,9	11,206	0,0121	0,09	0,17
$\bar{x} = 14,98$	$\bar{y} = 10,79$	$\hat{y} = 10,81$	**Σ = 9,23**	**Σ = 3,95**	**Σ = 5,02**

- Variabilité entre les valeurs observées et leur moyenne

  Var (obs-moy) = $\sum(y_i-\bar{y})^2/(N) = 0,92$

- Variabilité entre les valeurs observées et les valeurs prédites

  Var (obs-pred) = $\frac{\sum(y_i-\hat{y})}{N} = 0,40$

- Variabilité entre les valeurs prédites et la moyenne des valeurs observées

  Var (pred-moy) = $\sum(\hat{y}-\bar{y})^2/(N) = 0,50$

15. Que constatez-vous ?

Var (obs-moy) = Var (obs-pred) + Var (pred-moy)