# TD 03 : **Statistiques descriptives bivariées**

**Exercice 3**

Lors d'une enquête menée sur un échantillon de 15 familles, on a converti la consommation journalière en milliers de calories. Chaque adulte est compté pour une « unité de consommation » alors que les enfants sont comptés pour « une part d'unité », dépendant de son âge et de son sexe. Les résultats sont les suivants :

| **Familles** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Unités de consommation (Xi)** | 3,5 | 7,1 | 6,5 | 7,1 | 4,0 | 3,8 | 5,1 | 5,4 | 4,1 | 4,2 | 2,1 | 3,8 | 4,7 | 5,1 | 7,2 |
| **Calories par jour (Yi)** | 11 | 18 | 10 | 15 | 9 | 7 | 8 | 9 | 8 | 9 | 5 | 6 | 11 | 8 | 22 |

1. Calculer les paramètres statistiques univariés 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 ?
4. Quel est le type de relation qui existe entre les 2 variables ?
5. Calculer l'équation de la régression linéaire ?
6. Tracer le nuage de points représentant les données ;
7. Déterminer le point moyen et représenter-le dans le nuage de points ;
8. Tracer la droite de régression sur le nuage de points.
9. Calculer les valeurs prédites
10. 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 4**

Une expérience consiste à mesurer l'activité enzymatique avant et après l'addition d'un produit X. cette expérience est menée dans 23 tubes à essai et dans des conditions de laboratoire. Les résultats obtenus (en mg/l) sont les suivants :

| **Tube** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Avant** | 15,9 | 14,9 | 15,1 | 14,9 | 14,9 | 14,7 | 15 | 14,8 | 14,8 | 14,8 | 15,1 | 15,2 |
| **Après** | 8,2 | 11,3 | 11,9 | 11,1 | 11,5 | 11 | 10,9 | 10,9 | 10,2 | 10,9 | 10,2 | 11,5 |

| **Tube** | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Avant** | 14,8 | 14,9 | 15,2 | 15,4 | 14,9 | 14,5 | 14,9 | 15,1 | 15,5 | 15,4 | 15,4 |
| **Après** | 13,9 | 12,2 | 10,1 | 10,2 | 11,2 | 15,7 | 14,2 | 12,1 | 9,9 | 9,9 | 9,3 |

1. Calculer les paramètres statistiques univariés 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 ?
4. Quel est le type de relation qui existe entre les 2 variables ?
5. Calculer l'équation de la régression linéaire ?
6. Tracer le nuage de points représentant les données ;
7. Déterminer le point moyen et représenter-le dans le nuage de points ;
8. Tracer la droite de régression sur le nuage de points.
9. Calculer les valeurs prédites
10. 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

# TD 03 / Corrigé-type / Exercice 3

1. Calculer les paramètres statistiques univariés pour chaque variable ;

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

| Familles | UC (X) | Cal/J (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 | 3,5 | 11 | -1,41 | 0,60 | 2,00 | 0,36 | -0,85 |
| 2 | 7,1 | 18 | 2,19 | 7,60 | 4,78 | 57,76 | 16,62 |
| 3 | 6,5 | 10 | 1,59 | -0,40 | 2,52 | 0,16 | -0,63 |
| 4 | 7,1 | 15 | 2,19 | 4,60 | 4,78 | 21,16 | 10,06 |
| 5 | 4 | 9 | -0,91 | -1,40 | 0,83 | 1,96 | 1,28 |
| 6 | 3,8 | 7 | -1,11 | -3,40 | 1,24 | 11,56 | 3,79 |
| 7 | 5,1 | 8 | 0,19 | -2,40 | 0,03 | 5,76 | -0,45 |
| 8 | 5,4 | 9 | 0,49 | -1,40 | 0,24 | 1,96 | -0,68 |
| 9 | 4,1 | 8 | -0,81 | -2,40 | 0,66 | 5,76 | 1,95 |
| 10 | 4,2 | 9 | -0,71 | -1,40 | 0,51 | 1,96 | 1,00 |
| 11 | 2,1 | 5 | -2,81 | -5,40 | 7,91 | 29,16 | 15,19 |
| 12 | 3,8 | 6 | -1,11 | -4,40 | 1,24 | 19,36 | 4,90 |
| 13 | 4,7 | 11 | -0,21 | 0,60 | 0,05 | 0,36 | -0,13 |
| 14 | 5,1 | 8 | 0,19 | -2,40 | 0,03 | 5,76 | -0,45 |
| 15 | 7,2 | 22 | 2,29 | 11,60 | 5,23 | 134,56 | 26,53 |
| | $\bar{x}=4,91$ | $\bar{y}=10,40$ | //// | /// | **Σ = 32,06** | **Σ = 297,60** | **Σ = 78,12** |

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

$$Cov\,(x,y)=\frac{\sum(x_i-\bar{x})\times(y_i-\bar{y})}{N}=\frac{78{,}12}{15}=\mathbf{5,208}\;;$$

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

$$\sigma_x=\sqrt{\frac{\sum(x_i-\bar{x})^2}{N}}=\sqrt{\frac{32{,}06}{15}}=\mathbf{1,46}\;;\qquad \sigma_y=\sqrt{\frac{\sum(y_i-\bar{y})^2}{N}}=\sqrt{\frac{297{,}60}{15}}=\mathbf{4,32}\;;\quad r_{cal}=\frac{5{,}208}{1{,}46\times4{,}32}=\mathbf{0,83}$$

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 = 15 < 100 → ddl = N-2 = 13 ; α = 5% => **$r_{the}$ = 0,5139 ;** α = 1% => **$r_{the}$ = 0,6411.**

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**

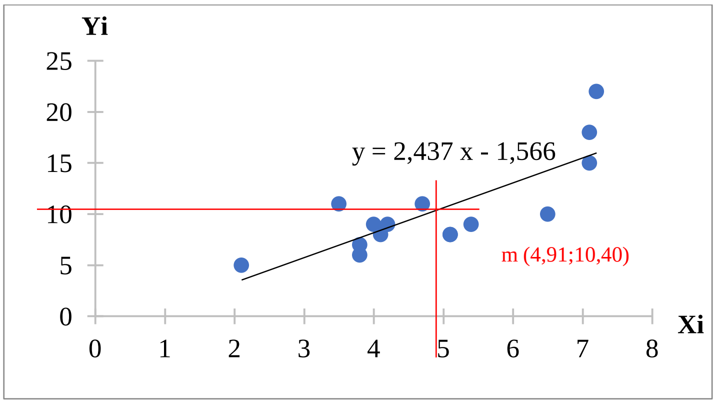
5. Déterminer l'équation de régression de type Y = a X + b ($y_i=a.x_i+b+e_i$) :

   - « a » représente la pente de la droite de régression.

$$a=\frac{cov(x,y)}{var(X)}=\frac{5{,}208}{2{,}137}=\mathbf{2,437}\quad ou\quad a=\frac{\sum(x_i-\bar{x}).(y_i-\bar{y})}{\sum(x_i-\bar{x})^2}=\frac{78{,}12}{32{,}06}=\mathbf{2,437}$$

- « 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 : $b = \bar{y} - a\,\bar{x} = 10{,}40 - 2{,}437 \times 4{,}91 = -1{,}566$
- L'équation de la droite de régression est donc : **Y = 2,437 X – 1,566.**

6. Nuage de points / Point moyen **m (4,91 - 10,40)** / Droite de régression.




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

| UC (X) | Cal/J (Y) | $\hat{Y}$ | $(Y - \bar{Y})^2$ | $(Y - \hat{Y})^2$ | $(\hat{Y} - \bar{Y})^2$ |
|---|---|---|---|---|---|
| 3,5 | 11 | 6,9635 | 0,36 | 16,29 | 11,81 |
| 7,1 | 18 | 15,7367 | 57,76 | 5,12 | 28,48 |
| 6,5 | 10 | 14,2745 | 0,16 | 18,27 | 15,01 |
| 7,1 | 15 | 15,7367 | 21,16 | 0,54 | 28,48 |
| 4 | 9 | 8,182 | 1,96 | 0,67 | 4,92 |
| 3,8 | 7 | 7,6946 | 11,56 | 0,48 | 7,32 |
| 5,1 | 8 | 10,8627 | 5,76 | 8,20 | 0,21 |
| 5,4 | 9 | 11,5938 | 1,96 | 6,73 | 1,43 |
| 4,1 | 8 | 8,4257 | 5,76 | 0,18 | 3,90 |
| 4,2 | 9 | 8,6694 | 1,96 | 0,11 | 2,99 |
| 2,1 | 5 | 3,5517 | 29,16 | 2,10 | 46,90 |
| 3,8 | 6 | 7,6946 | 19,36 | 2,87 | 7,32 |
| 4,7 | 11 | 9,8879 | 0,36 | 1,24 | 0,26 |
| 5,1 | 8 | 10,8627 | 5,76 | 8,20 | 0,21 |
| 7,2 | 22 | 15,9804 | 134,56 | 36,24 | 31,14 |
| $\bar{x} = 4{,}91$ | $\bar{y} = 10{,}4$ | **10,41** | **Σ = 297,60** | **Σ = 107,23** | **Σ = 190,39** |

- Variabilité entre les valeurs observées et leur moyenne

$$\text{Var (obs-moy)} = \sum(y_i - \bar{y})^2/(N) = \mathbf{19{,}84}$$

- Variabilité entre les valeurs observées et les valeurs prédites

$$\text{Var (obs-pred)} = \sum(y_i - \hat{y})^2/(N) = \mathbf{7{,}15}$$

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

$$\text{Var (pred-moy)} = \sum(\hat{y} - \bar{y})^2/(N) = \mathbf{12{,}69}$$

- Que constatez-vous ? **Var (obs-moy) = Var (obs-pred) + Var (pred-moy)**