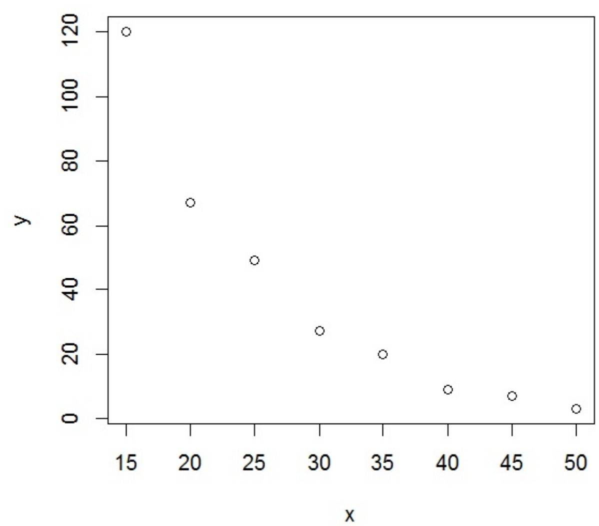
TD n° 2 : Régression linéaire simple 2

Exercice 1. On souhaite expliquer le nombre de survivants sur 10^6 bactéries (variable Y) à partir du temps d'exposition en minutes d'un agent microbien (variable X). Pour $n = 8$ expériences indépendantes, on observe les valeurs $(x_1, y_1), \ldots, (x_n, y_n)$ de (X, Y) suivantes :

x_i	15	20	25	30	35	40	45	50
y_i	120	67	49	27	20	9	7	3

1. Le nuage de points $\mathcal{N} = \{(x_1, y_1), \ldots, (x_n, y_n)\}$ est :



À partir de celui-ci, expliquer pourquoi une liaison linéaire entre Y et X n'est pas la plus judicieuse.

On considère la nouvelle variable $Z = \ln(Y)$ et, pour tout $i \in \{1, \ldots, n\}$, on pose $z_i = \ln(y_i)$.

2. Recopier et compléter le tableau ci-dessous :

x_i	15	20	25	30	35	40	45	50
z_i								

3. Représenter le nuage de points $\mathcal{N}_z = \{(x_1, z_1), \ldots, (x_n, z_n)\}$. À partir de celui-ci, expliquer pourquoi on peut envisager l'existence d'une liaison linéaire entre Z et X.

On adopte le modèle de *rls* : $Z = \beta_0 + \beta_1 X + \epsilon$, avec $\epsilon \sim \mathcal{N}(0, \sigma^2)$. Les paramètres β_0, β_1 et σ sont des réels inconnus.

4. Calculer $\overline{x}$, $\overline{z}$, s_x, s_z, sce_x et sce_z.

5. Calculer $\text{spe}_{x,z}$ et $r_{x,z}$.

6. Donner l'*emco* ponctuel b_1 de β_1 et l'*emco* ponctuel b_0 de β_0.

7. Tracer la droite de régression ajustant le nuage de point $\mathcal{N}_z$, puis la courbe corresponsante ajustant le nuage de point $\mathcal{N}$.

8. Donner une estimation ponctuelle de σ. Calculer ete_1 et ete_0.

9. Est-ce que la régression est significative ?

10. Donner un intervalle de confiance pour β_1 au niveau 95%, et un intervalle de confiance pour β_0 au niveau 95%.

Exercice 2. Un laboratoire veut expliquer le nombre de bactéries en 10^4 ufc (unités formant colonies) (variable Y) à partir de la température en degrés centigrades (variable X). Pour un échantillon de $n = 18$ cultures, des expérimentateurs observent les valeurs $(x_1, y_1), \ldots, (x_n, y_n)$ de (X, Y) suivantes :

x_i	10	10	10	12	12	12	14	14	14	16	16	16	18	18	18	20	20	20
y_i	16	12	11	16	17	13	18	14	13	21	12	17	25	18	21	20	26	23

1. Représenter le nuage de points $\{(x_1, y_1), \ldots, (x_n, y_n)\}$. À partir de celui-ci, expliquer pourquoi on peut envisager l'existence d'une liaison linéaire entre Y et X.

   On adopte le modèle de *rls* : $Y = \beta_0 + \beta_1 X + \epsilon$, avec $\epsilon \sim \mathcal{N}(0, \sigma^2)$. Les paramètres β_0, β_1 et σ sont des réels inconnus.

2. Calculer $\overline{x}$, $\overline{y}$, s_x, s_y, sce_x et sce_y.

3. On donne $\text{spe}_{x,y} = 209$. Calculer $r_{x,y}$.

4. Calculer le R^2 et le R^2 ajusté.

5. Donner l'*emco* ponctuel b_1 de β_1 et l'*emco* ponctuel b_0 de β_0.

6. Tracer la droite de régression.

7. Donner une estimation ponctuelle de σ. Calculer ete_1 et ete_0.

8. Est-ce que la regression est significative ?

9. Déterminer un intervalle de confiance pour β_1 au niveau 95%, et un intervalle de confiance pour β_0 au niveau 95%.

10. Un expérimentateur pense qu'un accroissement d'un degré centigrade provoque un accroissement moyen de 20000 bactéries. Peut-on affirmer, au risque 5%, qu'il se trompe ?

11. Ce même expérimentateur pense qu'à zéro degré, il n'y a, en moyenne, aucun développement de bactéries. Peut-on affirmer, au risque 5%, qu'il se trompe ?

12. Donner une estimation du nombre moyen de bactéries quand la température est à 19 degrés centigrades. Donner un intervalle de confiance pour celui-ci au niveau 95%.