TIPE
séance méthode 2

Expérimentation et statistiques

Fabien DACHY
Jean-Charles DUPRIEZ
Delphine HADOUX
Joseph NICOLAS
François PICAVEZ

BCPST1 Châtelet – Douai – 2019-2020

Pourquoi cette séance ?

- Comparer des données
- Mettre en évidence des corrélations
- Ne pas faire dire n'importe quoi à des nombres
- Répondre à un problème scientifique de façon quantitative

Situation 1 : comparer des moyennes

- Problème scientifique : je cherche à montrer que le blé pousse mieux avec engrais que sans engrais

- Expérience validée, biais limités

- Lot 1 : 1 jardinière, arrosé 2 fois par semaine (eau du robinet), 50 grains de blé

- Lot 2 : 1 jardinière, arrosé 2 fois par semaine (eau du robinet + 0,1 g/L de nitrate de potassium), 50 grains de blé

Situation 1 : comparer des moyennes

- Je mesure chaque plant, et ensuite ?
- Calcul des moyennes à 14 jours et à 21 jours :

- Lot 1 (sans engrais) :
- 14 jours : $m_1 = 12,1$ cm
- 21 jours : $m_1' = 22,2$ cm

- Lot 2 (avec engrais) :
- 14 jours : $m_2 = 13,0$ cm
- 21 jours : $m_2' = 25,2$ cm

Situation 1 : comparer des moyennes

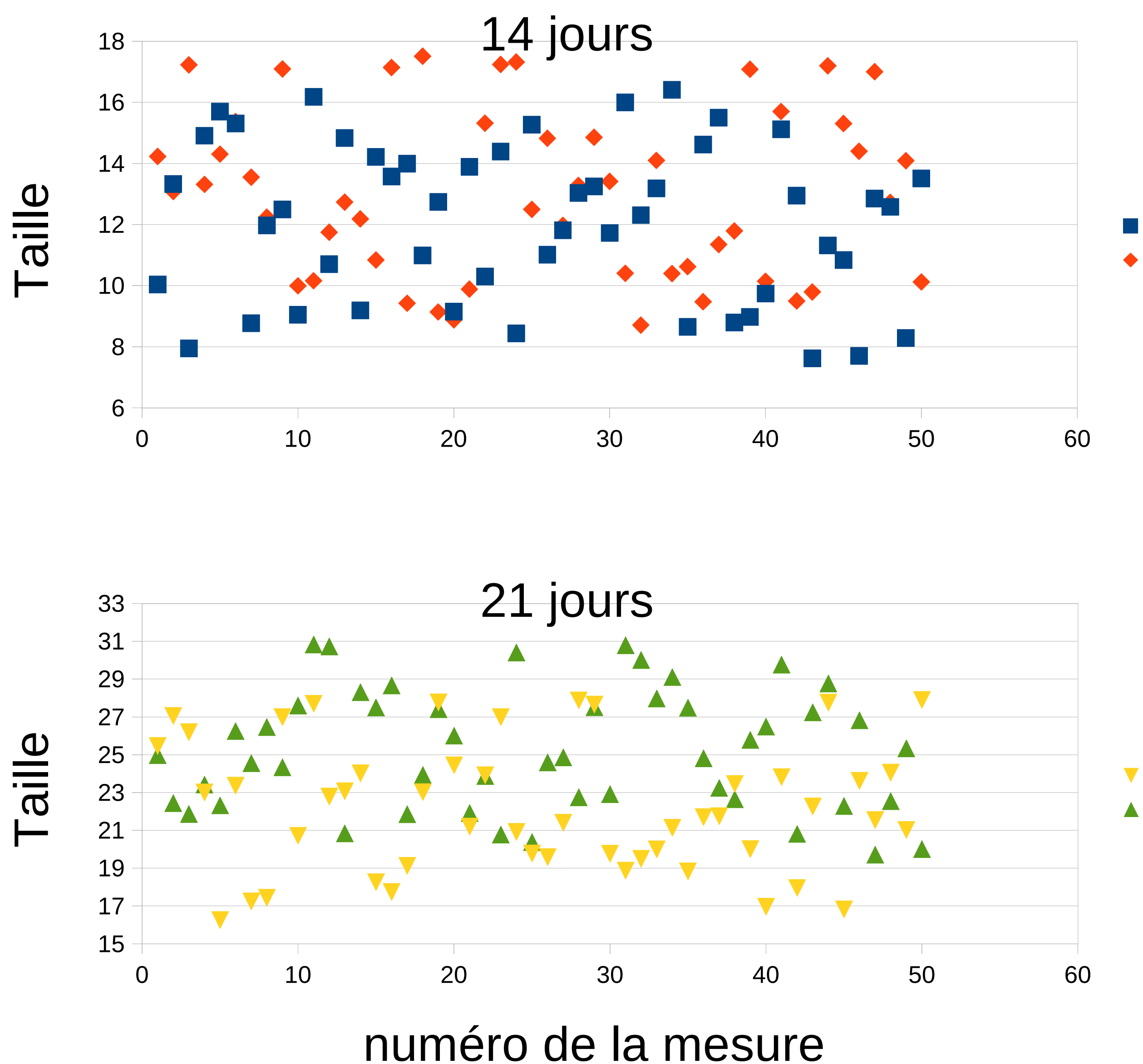
- Conclusion : le blé pousse mieux avec engrais que sans…
- … ou pas

L'étudiant(e)

le prof

Situation 1 : comparer des moyennes

- Dispersion des données
- La différence paraît moins évidente !



Situation 1 : comparer des moyennes

- Comment décider si les moyennes calculées sont « différentes » ?
- Quantifier la dispersion : la variance et l'écart-type

$$V = \frac{1}{N} \sum_{i=1}^{N} \left(x_i - \bar{x} \right)^2 \quad \text{et} \quad \sigma = \sqrt{V}$$

Situation 1 : comparer des moyennes

- **Hypothèse nulle** : moyennes m_1 et m_2 **identiques**.

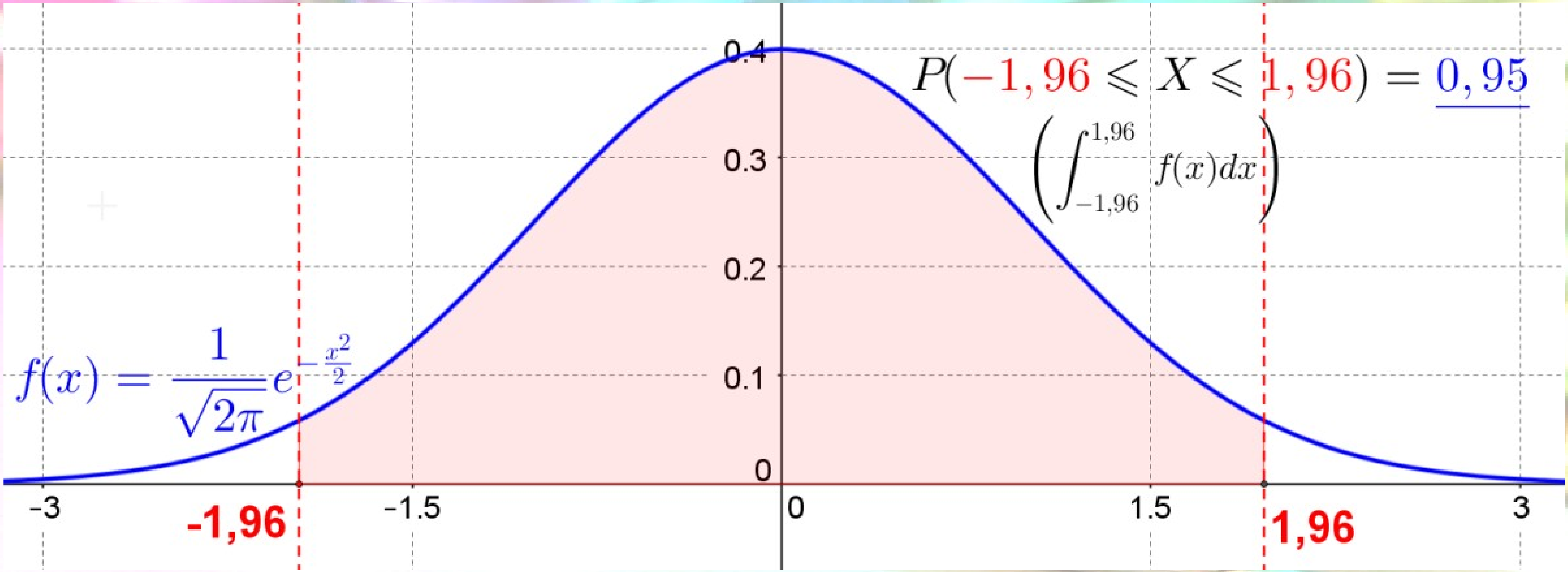
- Soit $$z = \frac{m_1 - m_2}{\sqrt{\frac{\sigma_1{}^2}{n_1} + \frac{\sigma_2{}^2}{n_2}}}$$

- Sous l'**hypothèse nulle**, *z* suit la **loi normale centrée réduite**, donc :

$$P\left(\left|z\right| > 1{,}96\right) = 0{,}05$$

Situation 1 : comparer des moyennes

- Illustration par un graphe :



Situation 1 : comparer des moyennes

- Deux cas de figure :
  - Si $|z| \leq 1{,}96$ alors on accepte l'hypothèse nulle : **les moyennes ne sont pas significativement différentes**
  - Si $|z| > 1{,}96$ alors on rejette l'hypothèse nulle : **les moyennes sont significativement différentes**
- Cadre d'application : n > 30

Situation 1 : comparer des moyennes

- A vos calculatrices...

- Lot 1 (50 plants, sans engrais) :
- 14 jours : m_1 = 12,1 cm ; σ_1 = 2,6 cm
- 21 jours : m_1' = 22,2 cm ; σ_1' = 3,4 cm

- Lot 2 (50 plants, avec engrais) :
- 14 jours : m_2 = 13,0 cm ; σ_2 = 2,8 cm
- 21 jours : m_2' = 25,2 cm ; σ_2' = 3,2 cm

Situation 1 : comparer des moyennes

- Dans notre cas :

  - A 14 jours :
    - ***z = 1,67***
    - ***z < 1,96*** donc m_1 et m_2 **ne sont pas significativement différentes**
  - A 21 jours :
    - ***z = 3,54***
    - ***z > 1,96*** donc m_1' et m_2' **sont significativement différentes**
    - risque d'erreur : 5 %

Situation 1 : comparer des moyennes

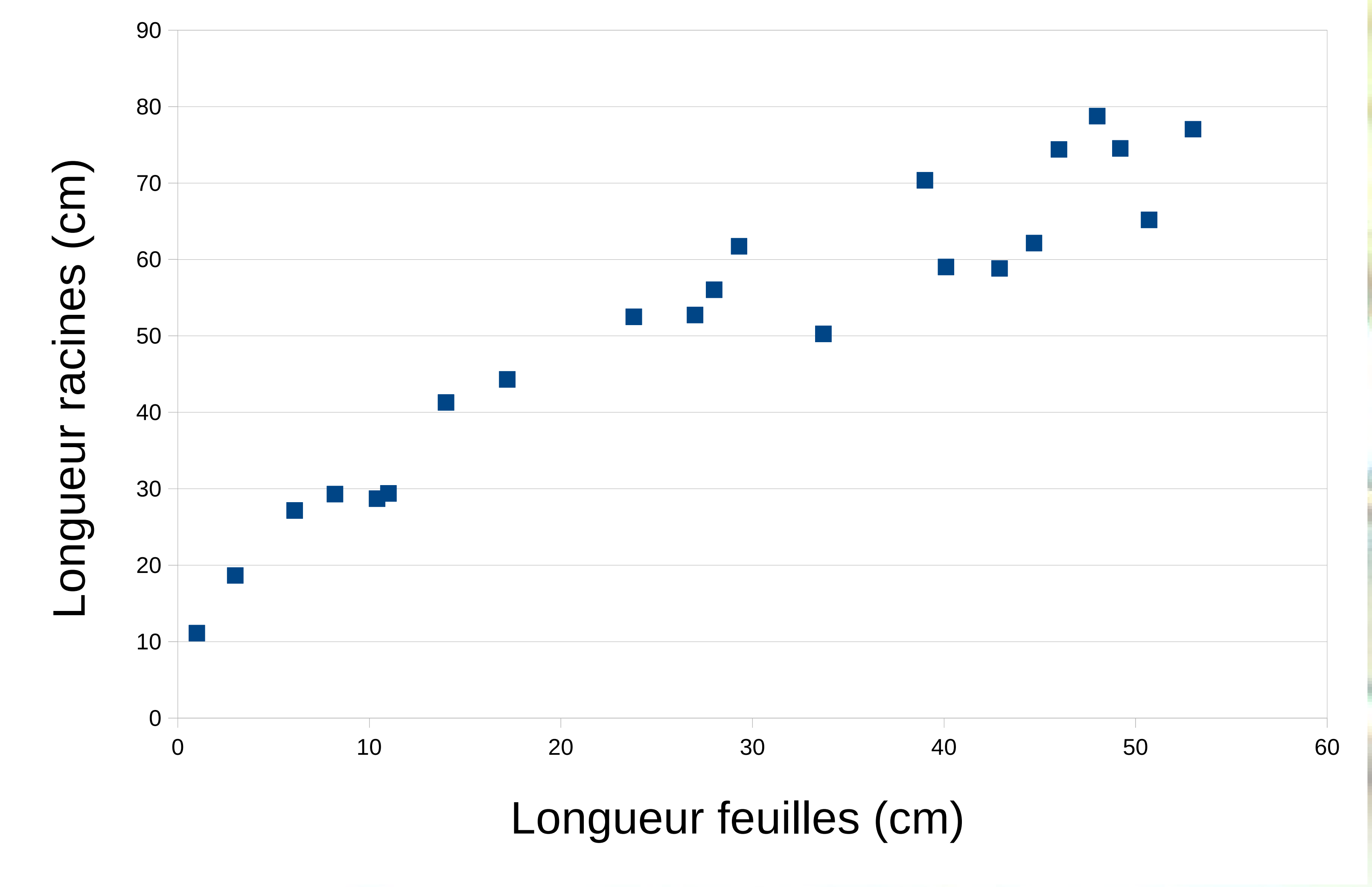
- **Interprétation biologique :**
  - Premiers jours : la plante puise dans ses réserves (graine) plutôt que dans le sol
  - Cet engrais n'agit pas sur les premières phases du développement

Situation 2 : établir une corrélation

- Question scientifique : je cherche à montrer qu'il y a proportionnalité entre la longueur totale des feuilles et la longueur totale des racines chez le blé

Situation 2 : établir une corrélation

- Je mesure la taille des feuilles et des racines sur une population de 22 plants de blé, et j'obtiens :



Situation 2 : établir une corrélation

- Conclusion : c’est bon, c’est proportionnel…
- …ou pas

L'étudiant(e)

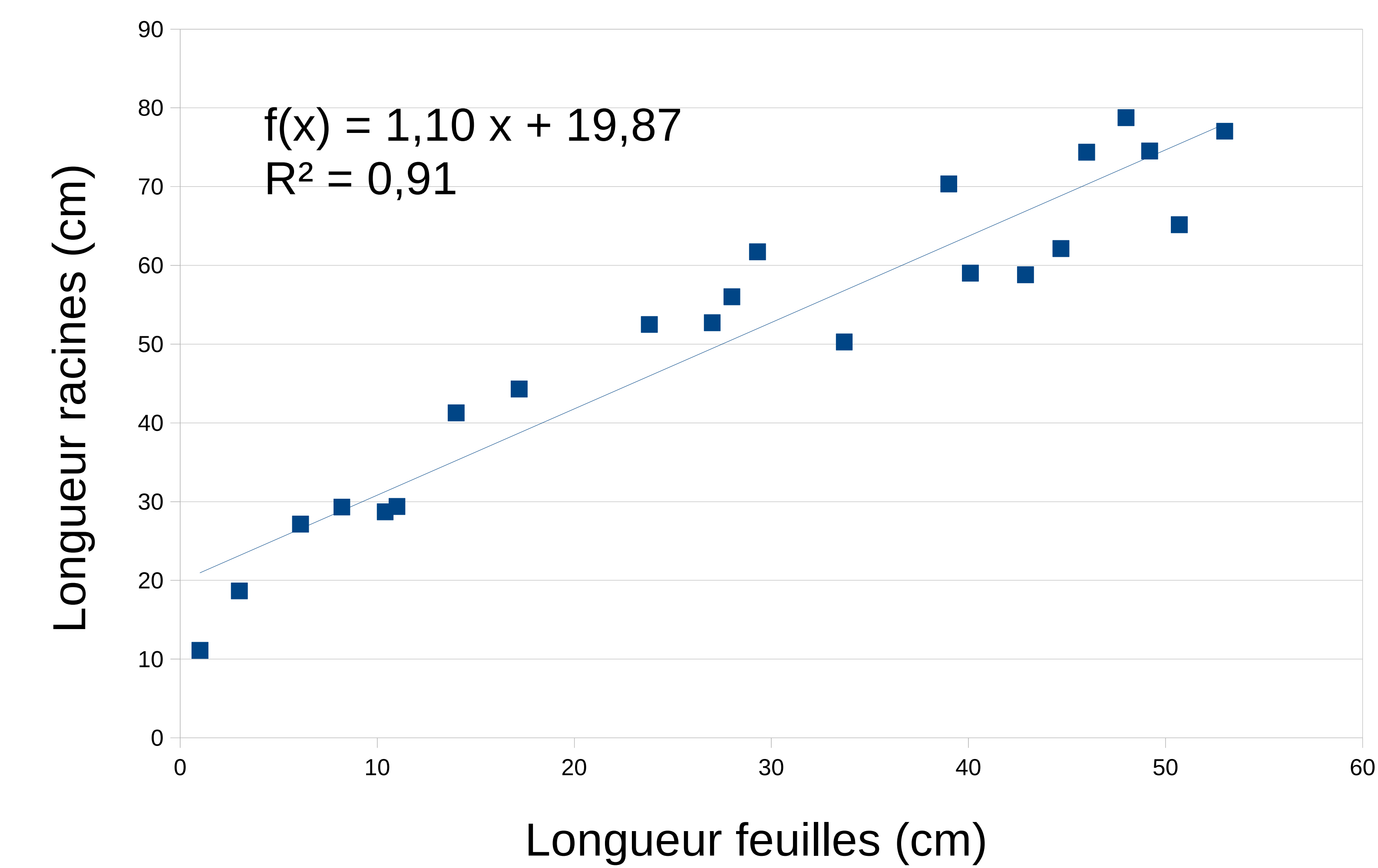
le prof

Situation 2 : établir une corrélation

- Objectif du scientifique :
  - Chercher à savoir s'il existe un lien entre deux variables
  - Préciser la nature mathématique de ce lien
  - Idée : tracer une droite de régression (LibreOffice, Regressi, Excel...)

Situation 2 : établir une corrélation

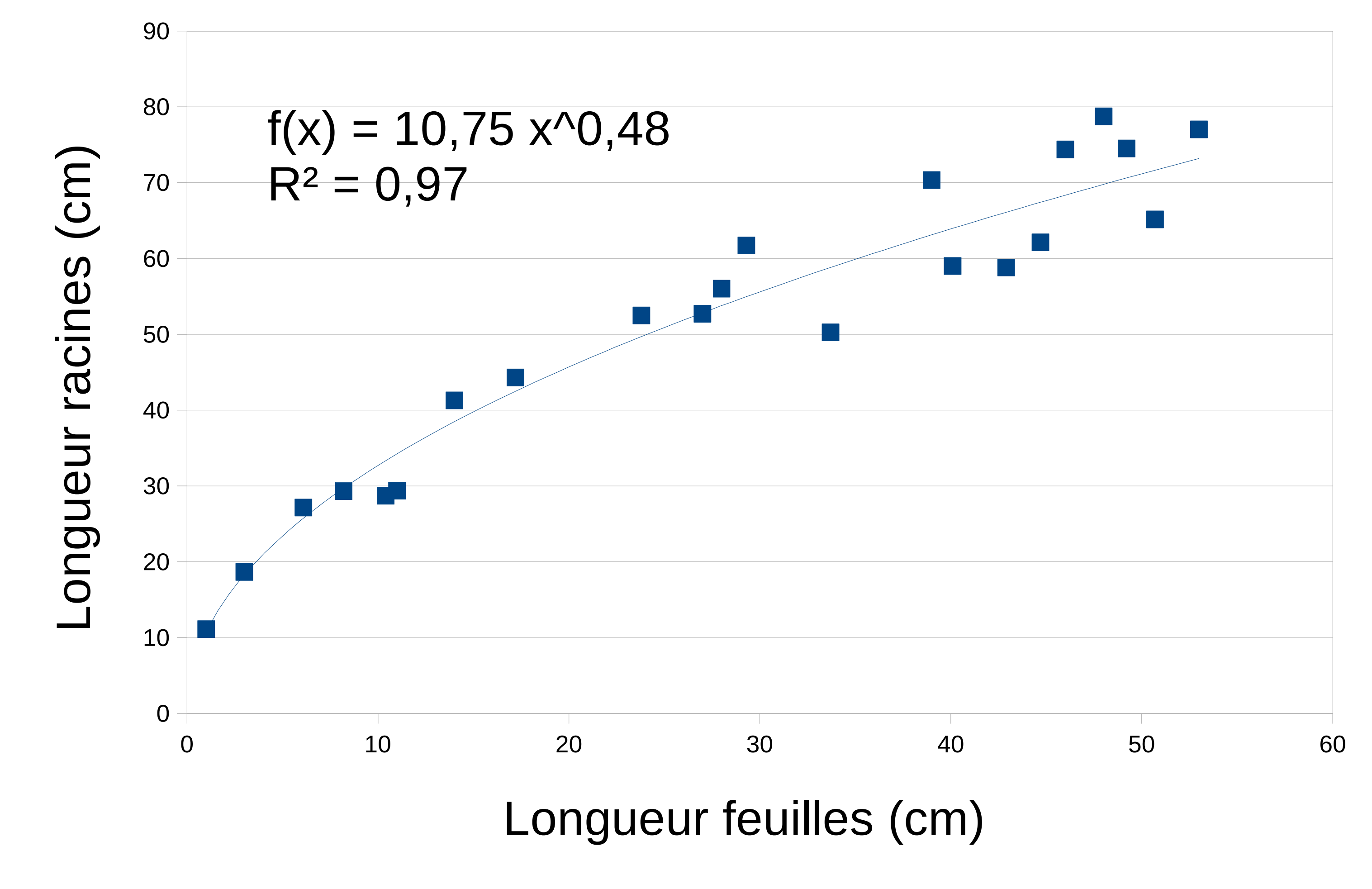
- Résultat :
- $R^2 = 0{,}91$...
- Points visiblement non répartis aléatoirement autour de la droite de tendance



- Peut-on faire mieux... ? la corrélation **linéaire** est-elle la seule solution ?

Situation 2 : établir une corrélation

- 2^e essai : courbe de tendance de type puissance
- $R^2 = 0,97$... c'est mieux !
- Meilleure modélisation des données



Situation 2 : établir une corrélation

- Bilan et éléments de réflexion :
  - Que conclure quant à ces régressions ? autres possibles ?
  - Conviennent-elles toutes les deux ?
  - Prouvent-elles une relation entre taille des racines et des tiges ?
  - Quand choisit-on que le R^2 est « bon » ?
  - Cherche-t-on une adéquation à une loi ou une corrélation inconnue *a priori* entre deux grandeurs ?

Situation 2 : établir une corrélation

- Interpréter une corrélation : ne pas confondre corrélation et causalité
  - Excellente corrélation entre niveau de maths d'un collégien et sa pointure de chaussure. Donc ?
  - Forte corrélation entre prévalence de l'obésité et du diabète de type II. Donc ?
  - Et pour rire : http://www.tylervigen.com/spurious-correlations

Au boulot !