Fiche pratique

Régression linéaire

I Position du problème

A Énoncé

La tension et l'intensité sont mesurées au travers d'une résistance de $R = 1$ kΩ d'après le constructeur.

I (en A)	0,01	0,02	0,03	0,04	0,05	0,06
U (en V)	10,1	20,0	29,8	40,2	50,0	60,1

1. Montrer que la loi d'Ohm $U = R \times I$ est vérifiée.

2. Déterminer expérimentalement la valeur de la résistance R. La comparer à la valeur fournie par le constructeur.

B Approche naïve

C Principe de la régression linéaire

Dès lors que l'expérience fait apparaître une série de mesures bidimensionnelle (une mesure en fonction d'une autre, comme ici U en fonction de I), la manière la plus courante d'analyser les données est d'utiliser une régression linéaire.

En effet, si on place dans un plan les points correspondants aux couples formés par deux séries de données, on obtient ce qu'on appelle un nuage de points. Effectuer une régression linéaire entre les deux séries consiste à déterminer la droite qui passe au plus près de l'ensemble de ces points. Passer au plus près signifie minimiser la somme des distances entre chacun des points et la droite.

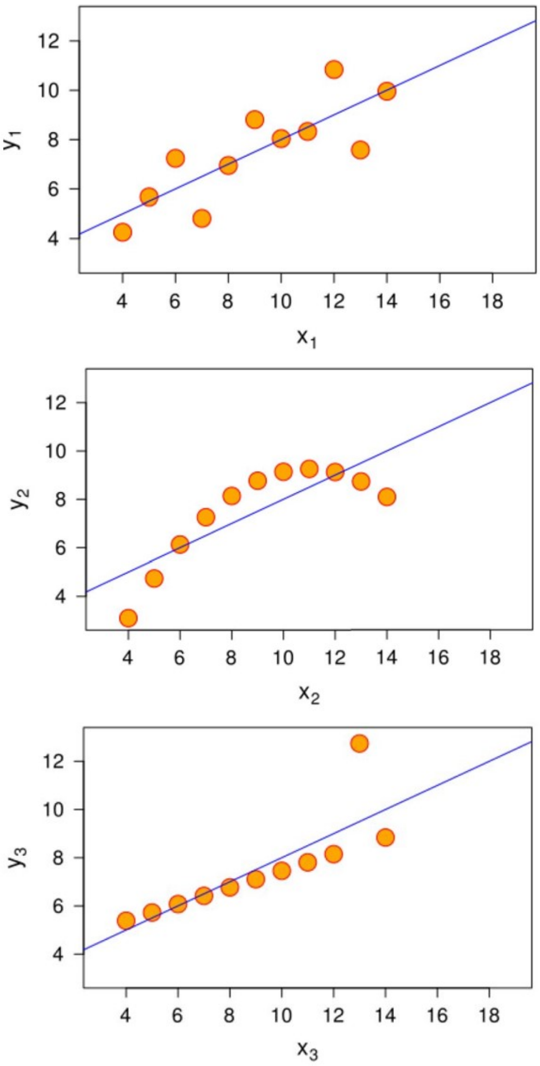
Cette droite a pour équation $y = ax + b$. On peut alors déterminer trois paramètres :

1. Le coefficient directeur de la droite a (la pente) ;
2. L'ordonnée à l'origine b ;
3. Les coefficients de corrélation (ou de régression), r ou r^2 (compris tous les deux entre 0 et 1).

C'est la calculatrice (ou l'ordinateur) qui effectue cette opération. Les coefficients de corrélation permettent d'évaluer la proximité de la droite par rapport au nuage de points, et donc **dans quel mesure un modèle linéaire est pertinent pour décrire le comportement expérimental du système étudié**.

D Critère d'évaluation de la pertinence du modèle

La calculatrice est une machine ! Elle n'émet pas de jugement sur les opérations qu'on lui fait faire. Si une régression linéaire est demandée, la calculatrice proposera l'équation d'une droite, même si les points ne sont pas du tout alignés. Il faut donc évaluer la pertinence du modèle linéaire. Pour ce faire, on utilise deux critères :



1. On effectue une vérification visuelle : on trace le nuage de points à la calculatrice et on vérifie qu'ils semblent alignés. Les trois nuages de points à droite présentent un r^2 identiques. Pourtant, ils ont des comportements très différents. Le premier a une dispersion élevée mais suit une tendance linéaire. Il est donc pertinent d'appliquer une régression linéaire. Dans le second cas, une tendance quadratique est visible. Tenter de réaliser une régression linéaire est alors absurde (même si le r^2 reste proche de l'unité). Dans ce cas, on peut supposer que le modèle que l'on a choisi ne rend pas bien compte de la réalité physique. Enfin, dans le troisième cas, l'avant dernier point est de toute évidence aberrant (erreur de relevé de l'expérimentateur par exemple). Il génère une dispersion artificielle et modifie le coefficient directeur de la droite de régression. Il convient de le supprimer (ou de reprendre la mesure) avant de réaliser la régression.
2. On utilise le coefficient de corrélation r : plus la valeur de r est proche de 1, meilleure est la qualité de la régression linéaire ($r = 1$ signifierait que tous les points sont parfaitement alignés). On considérera qu'une régression linéaire est satisfaisante (et que par voie de conséquence le modèle est pertinent) si

$$\boxed{|r| > 0,999 \qquad \text{ou} \qquad r^2 > 0,99}$$

Remarque importante : La valeur de r s'écrit avec tous les chiffres « 9 » après la virgule ainsi que le premier chiffre différent de « 9 » sans arrondi, même s'il faut mettre un grand nombre de chiffres significatifs. Lorsque vous donnez le résultat d'une régression, il convient de toujours indiquer r ou r^2 pour vérifier que le modèle est pertinent pour rendre compte des données expérimentales.

II Vérifier une loi linéaire

Il faut maintenant être en mesure d'utiliser sa calculatrice pour réaliser la régression linéaire. Voici une aide à l'utilisation sur Casio ou sur Texas Instrument.

Remarque : Une aide en ligne pour les calculatrices Casio Graph 35 et TI 83 : `http://www.sesabac.net/theories/CMP_Droite_Regression.swf`

TI 83+ :
Appuyer sur `stats`.
Dans le menu `EDIT` choisir `1 : Editer`.
Entrer les deux séries de données ; l'abscisse dans la première colonne `L1`, et l'ordonnée dans la 2ème colonne `L2`.
Dans le menu `graph stats` choisir la première ligne.
Choisir `Aff` pour afficher, puis le 1er type, la liste `L1` pour `ListeX` et la liste `L2` pour `ListeY`.
Appuyer sur `graphe` pour afficher le nuage de points.
Pour déterminer les coefficients a, b et r ou r^2 il faut retourner dans `stats` puis aller dans le menu `CALC` et choisir `4 : RegLin(ax+b)`.
En appuyant sur `ENTER` on obtient les coefficients.

Si le coefficient r ou r^2 n'est pas donné :

- Appuyer sur `2nde` puis `catalog`.
- Choisir `Diagnostic On` (en anglais) ou `correlAff` (en français) et appuyer sur `ENTER` 2 fois (fait).

Pour tracer la courbe d'ajustement :

- Aller dans `f(x)`.
- Appuyer sur `var`.
- Choisir `5 : Statistiques`.
- Dans le menu `EQ` choisir 1.
- Puis `ENTER`.
- Pour voir la droite aller dans `graphe`.

Casio Graph 35+ :
Aller dans le mode `STAT`
Entrer les deux séries de données d'origine, chacune dans une liste différente `List1` et `List2`.
Appuyer sur `GRAPH` puis `SET`.
Choisir le type de graphique `GraphType : Scatter`. Mettre `List1` dans `XList` en abscisse et `List2` dans `YList` en ordonnée.
Appuyer sur `ENTER`.
Appuyer sur `GRAPH 1` pour visualiser le nuage de points.
Appuyer sur `CALC` puis `X` puis sur `ax+b` pour obtenir une régression linéaire $y = ax + b$.
Une fenêtre `LinearReg` affiche la pente a, l'ordonnée à l'origine b et le coefficient de corrélation r^2 de la droite de régression.
Pour visualiser la droite de régression linéaire, choisir `DRAW` en bas à droite de la fenêtre `LinearReg`.

ATTENTION : Une erreur très courante consiste à inverser les deux listes (la régression ne s'effectuant alors pas dans le bon sens...).

III Vérifier une loi non linéaire

Dans le cas présenté en début de cette fiche, la loi que l'on se proposait de vérifier était linéaire : $U = R \times I$. Mais dans de nombreux cas, la loi à vérifier n'est pas linéaire. Supposons par exemple que l'on cherche à vérifier la loi de Snell Descartes sur la réfraction entre deux milieux d'indice optique n_1 et n_2 :

$$n_1 \sin(i_1) = n_2 \sin(i_2)$$

La relation entre i_1 et i_2 n'est pas linéaire (la fonction sinus n'étant pas linéaire). Aussi, il est impossible de tenter directement une régression linéaire sur $i_1 = f(i_2)$. En revanche, si vous tracez

$$\sin(i_1) = f(\sin(i_2))$$

vous devriez obtenir un nuage de points qui s'organise selon une droite. Vous allez alors pouvoir réaliser une régression linéaire pour vous assurer que la loi de Snell-Descartes est validée (en montrant que $r^2 > 0,99$). Il est possible, grâce aux calculatrices de réaliser directement des opérations sur les listes.

TI 83+ :
Aller sur la case `L3` puis `ENTER`.
Entrer, par exemple ici, la formule `L3=sin(L2)` dans la ligne de calcul (en bas de l'écran) puis `ENTER`.
Vous avez ainsi créé une série de données dans `L3` à partir du sinus des valeurs de la série `L2`.

Casio Graph 35+ :
Aller dans le menu `RUN`.
Appuyer sur `OPTION`.
Choisir `LIST`.
Entrer la formule `f(List2)` $\rightarrow$ `List3` par exemple `sin(List2)` $\rightarrow$ `List3`.
La calculatrice doit afficher DONE. Vous avez ainsi créé une série de données dans `List3` à partir des valeurs de la série `List2`.

IV Déterminer l'incertitude sur les coefficients de regression

A Position du problème

Les coefficients de la régression (ordonnée à l'origine et coefficient directeur) servent le plus souvent à déterminer une grandeur expérimentale. Dans le premier exemple de ce cours, le coefficient directeur permet par exemple de déterminer la valeur de la résistance R. Il est donc utile de déterminer l'incertitude-type sur ces coefficients de régression afin de déterminer l'incertitude-type sur la valeur de la grandeur expérimentale recherchée. Pour ce faire, une méthode Monte-Carlo est proposée.

B Principe général

L'idée est de générer un nombre N de sets de données (x, y) avec un tirage aléatoire (dans l'intervalle associé à l'incertitude-type sur chacune des grandeurs) pour chacune des valeurs de x et de y. Pour chacun de ces N sets, une régression linéaire est réalisée permettant d'obtenir une liste de N coefficients directeurs et de N ordonnées à l'origine. L'écart-type sur les N tirages est ensuite calculé afin d'obtenir l'incertitude-type sur les coefficients de régression. Le script suivant permet de comprendre les étapes en détail. Notez qu'il est nécessaire de déterminer les incertitudes-type *Delta_x* et *Delta_y* sur les variables x et y par une méthode autre que statistque (incertitudes de type B).

C Script modèle (à adapter)

```
import numpy as np
import matplotlib.pyplot as plt

############### Données expérimentales   ###############

```

```
# Données mesurées

x=np.array([0.01,2.5,5,7.5,10])
y=np.array([2.2,7.7,12.4,17.7,21.1])


# Précisions mesurées (à renseigner au cas par cas)

Delta_x = 0.05*x
Delta_y=1

# Possibilité de changer les labels des axes
plt.figure(figsize=(8,6))
plt.xlabel('x')
plt.ylabel('y')



############### Régression ##################


N=10000

 # regression linéaire sur la valeur mesurée
a,b = np.polyfit(x, y, 1)

 # Estimation de son incertitude-type par simulation Monte-Carlo
ta,tb=[],[] #liste des coef de régression Monte Carlo

# on réalise N tirages aléatoires des valeurs de x et de y dans
# leur intervalle (fixés par Delta_x et Delta_y)

for i in range(0, N):
    nbpts = len(x)
    mx = x + Delta_x*np.random.uniform(-1 , 1, nbpts)
    my = y + Delta_y*np.random.uniform(-1, 1, nbpts)

    p=np.polyfit(mx, my, 1) # on réalise une reg lin
    #sur chacun des tirages

    ta.append(p[0])
    tb.append(p[1])


# écart type (standard deviation en anglais) pente
# et ordonnée à l'origine
u_a = np.std(ta,ddof=1)
u_b = np.std(tb,ddof=1)

# Equation de la droite de régression (pour le tracé)
yfit = a*x + b
xfit = np.linspace(min(x), max(x), 2)
```

```python
# Tracé
plt.plot(xfit, a*xfit + b, 'r', label='Régression linéaire')
plt.errorbar(x, y, xerr=Delta_x, yerr=Delta_y, fmt='bo', label='Mesures')
plt.legend()
plt.show()

print ("Coef. directeur (et incertitude-type) :", a ,"+/-", u_a)
print ("Ordonnée à l'origine (et incertitude-type) :", b ,"+/-", u_b)
```