Scénario 6

AJUSTEMENT A UNE DROITE

1 Présentation générale

1.1 Objectifs :

- Tracer une droite d'ajustement
- Représenter graphiquement les écarts d'ajustement

1.2 Les étapes

1. Présentation d'un exemple
2. La droite : calcul et interprétation de a et b, tracé de la droite, calcul et représentation des points ajustés.
3. Le critère des moindres carrés , calcul des écarts (résidus) - représentation graphique, somme des moindres carrés
4. Les résidus
5. Domaine de validité de la régression (facultatif)
6. Evaluation
7. Biblio
8. Autres exemples

1.3 Le matériel

Pour chaque stagiaire :

- Régle graduée
- Calculatrice
- papier millimétré, feutres de couleur

Pour le formateur :

- tableau
- transparents et projecteur

Remarque : On a besoin d'avoir accès au tableau et à la projection de transparents en même temps.

1.4 La forme

Intéractive. On travaille à partir d'un exemple : influence du temps sur la longueur de cristaux de glace dans une chambre froide à 5° C, pression de gaz et humidité constantes.

1.5 Evaluation

1.6 Prérequis

- acquis dans d'autres scénarios

  1. notion de variance
  2. boîte à patte, stem-end-leaf, corrélation (scénario 4-5)
  3. variable contrôlée et variable aléatoire (scénario 1)

- autres :

  1. règle de 3
  2. positionner des points sur un graphe
  3. manipuler un repère de coordonnées
  4. utiliser une calculatrice
  5. nombres entiers relatifs : 5 - 8 = -3

1.7 Durée

Deux séances de 2 heures 1/2

2 Les étapes

2.1 Etape 1 : Présentation de l'exemple

- note pour la formateur : on a choisi délibéremment de ne pas utiliser le même jeu de données que pour le scénario : Liaison de 2 variables
- distribuer la fiche 1 (tableau de données)
- question "visualiser l'effet du temps sur l'accroissement de longueur des cristaux"
  distribuer 2 feuilles de papier millimétré à chacun
  chacun fait son graphique

1. corriger les problèmes d'échelle, si besoin (correction individuelle)
2. corriger le choix de la variable mis en abscisse, et de la variable mise en ordonnée (correction collective) Pour faire cette correction, ou pour bien faire comprendre le choix que l'on doit faire, aller au tableau, et demander quels sont les termes qu'on peut utiliser pour désigner les deux variables X et Y, les classer, et compléter.

X	Y
abscisse variable explicative variable contrôlé régresseur	ordonnée variable expliqué réponse variable dépendante (de X)

3. insister sur le rôle dissymétrique de X et Y
4. rappeler, par des exemples, la différence entre une variable contrôlée et une variable non contrôlée. Par exemple, dans notre cas, on contrôle la température. Si on avait fait des mesures de longueur de cristaux, dehors, en conditions naturelles, on n'aurait pas contrôlé la température ...
5. refaire le graphique définitif, si certains avaient inversé X et Y

2.2 Etape 2 : la droite

Calcul et interprétation de a et b
Tracé de la droite
Calcul et représentation des ajustés

- Question : On veut résumer, synthétiser, le nuage de points, par une courbe le plus simple, c'est une droite.
  $\Longrightarrow$ tracer une droite à la règle

  Si un stagiaire pose la question"comment choisir la droite", on répond : "au mieux, à l'oeil". Sur la feuille de papier millimétré chaque stagiaire trace sa droite.
- Etude d'une droite : faire au tableau, par le formateur : (ou avec les transparents 1 à 5, au choix).
- Chacun calcule les a et b de la droite qu'il a tracé. (estimation graphique).

  point stratégique : attention aux unités employées!! (ce doit être les unités réelles)

- Interprétation de a et b : faire interpréter par les stagiaires.

  a : longueur initiale des cristaux
  b : accroissement de longueur par unité de temps

- Demander aux stagiaires de positionner, sur leur droite, les Y ajustés correspondant aux X observés (notion d'ajustement).

  Chaque stagiaire complète la colonne "Y ajusté" (fiche stagiaire 2).

2.3 Etape 3 : Critère des moindres carrés.

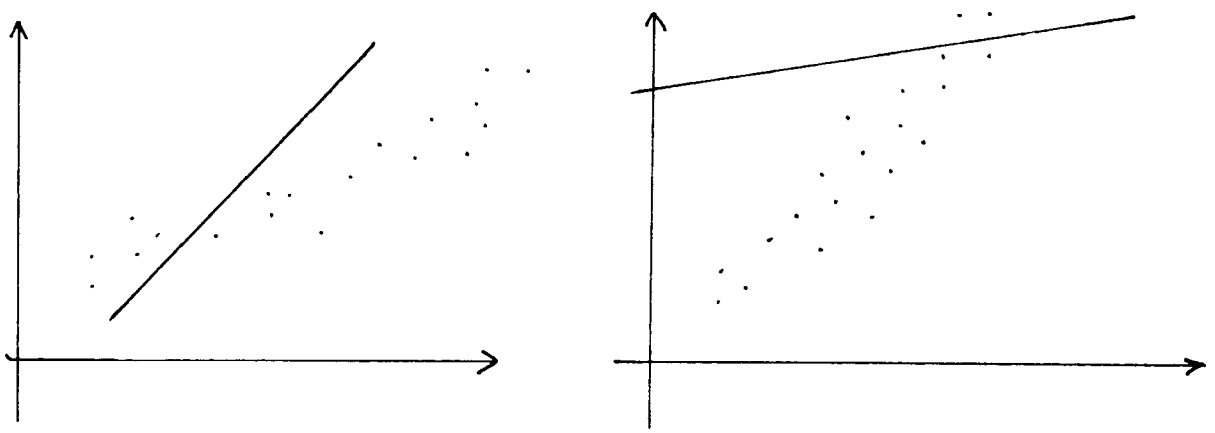
Calcul des écarts (résidus). Représentation graphique.

Somme des carrés des écarts

- On écrit au tableau les différentes valeurs de b trouvées par les stagiaires.
- **Question :** "Quelle est la meilleure valeur?"

  "sur quel critère peut-on se baser?"

  Si aucune réponse n'est donnée, tracer au tableau, ou sur transparent, une droite qui ne convient visiblement pas. Par exemple



Les réponses possibles :

- droite passant par les points extrêmes
- droit ayant le plus de points sur la droite
- droite ayant autant de points au dessus qu'en dessous
- droite passant par l'origine
- ...

La réponse du formateur :

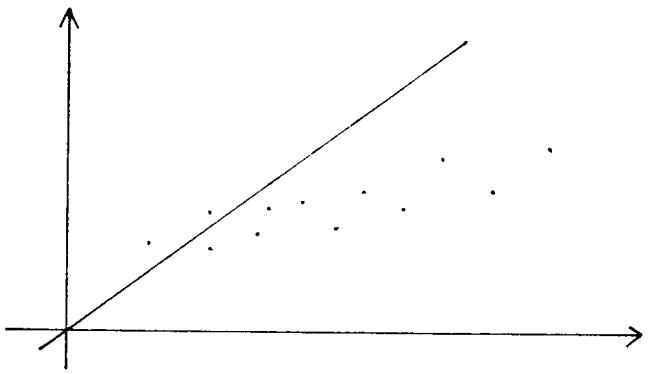
- supprimer les réponses qui conduisent à plusieurs droites possibles
- supprimer les réponses qui ne tiennent pas compte de tous les points

  (par exemple : avec les 2 points extrêmes on n'utilise que 2 points, alors qu'on dispose de plus d'informations).

Exemple :

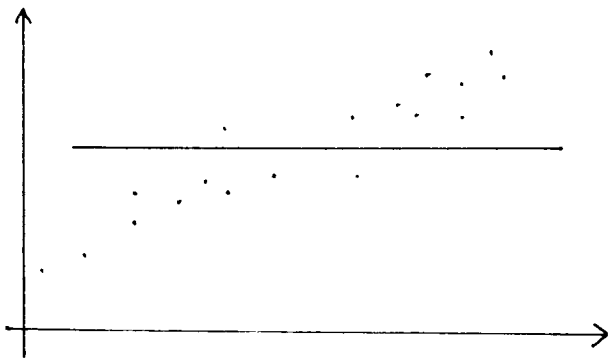
Si réponse : "droite passant par l'origine"

donner un contrexemple d'une droite qui passe par l'origine et ne convient pas.



Si réponse : "droite ayant autant de points au dessus et en dessous"

contrexemple.



Mettre les stagiaires sur la piste, en parlant d'écart entre les ajustés et les observés. On arrive à la conclusion : le critère doit tenir compte des écarts entre ajustés et observés.

- les stagiaires calculent les écarts $Y_{\text{observés}} - Y_{\text{ajustés}}$ (résidus)

  ils complètent la colonne écart , fiche 2.

  point stratégique : attention au signe des écarts

ex : 5-8 qui donne 3!

Les stagiaires représentent les écarts sur leur graphique (cf. transparent 7) ces écarts reflètent la qualité de l'ajustement.

Facultatif :

Remarquer qu'il existe d'autres types d'écarts, et qu'on ne s'intéresse ici qu'à un type d'écart bien particulier (cf. transparent 7).

Montrer (sur transparent 7, ou au tableau, ou par le calcul) que les écarts (positifs et négatifs) se compensent si on les somme.
Donc, on les élève au carré, puis on en fait la somme.

$\Longrightarrow$ somme des carrés des écarts = mesure globale de la qualité de l'ajustement.

Chaque stagiaire termine le calcul de la somme des carrés des écarts (fiche stagiaire 2).

Le formateur écrit au tableau, en face des valeurs de b, les différentes sommes de carrés d'écarts trouvés par les stagiaires. Elles sont toutes différentes, il en existe une plus petite, qui correspond à la meilleure droite, parmi celles qui ont été tracées, (au sens des moindres carrés).

Question : "Peut-on trouver mieux?"

il faut une méthode qui permette de trouver a et b donnant la plus petite somme des carrés d'écarts.

On distribue le formulaire régression linéaire (cf. feuille stagiaire 3 et 4).
Expliquer le formulaire.

Chaque stagiaire complète la fiche 4, et calcule ainsi les valeurs estimées de a et b qui minimisent le critère des moindres carrés.

Chacun termine ainsi le calcul de la somme des carrés des écarts.

On compare aux sommes des carrés d'écarts calculés à partir des droites tracées à la main. C'est plus petit.
Chacun trace sur son graphique, la droite des moindres carrés (transparent 8).

2.4 Etape 4 : Les résidus.

Le formateur présente le transparent 9 : représentations graphiques des résidus. On remarque avec les stagiaires qu'il n'y a pas de structure particulière des résidus.

Le formateur présente le transparent 10 :

représentations graphiques des observés, et de la droite, ainsi que des résidus en fonction de x dans différents cas :

1. pas de structure des résidus
2. un point extrême
3. variance hétérogène
4. relation non linéaire

On remarque, avec les stagiaires, la structure des résidus, on la décrit. Insister sur l'importance de la visualisation des résidus.

2.5 Etape 5 : Domaine de validité de la régression.

Faire des remarques de bon sens, sur le domaine de validité de la droite, sur l'extrapolation.

Le formateur explique qu'il ne faut pas extrapoler, tirer de conclusion en dehors de la zone d'étude.

Le formateur présente le transparent 11 : (données tirées du polycopié de R. Tomassone, p. 262-263. épiceas)

Volume du tronc des épiceas en fonction de la surface du tronc à 1m 30. L'ordonnée à l'origine est non nulle.
Ça n'a pas de sens, de signification.
En fait, la relation entre volume et surface du tronc à 1m 30 est forcément non linéaire, lorsqu'on extrapole à gauche de la zone d'observation : la courbe doit passer par l'origine.

Autre exemple p. 18 (ibid.,)

X = nombre de composants
Y = temps de réparation

lorsque le nombre de composants est nul, le temps de réparation est nul.

2.6 Etape 6 : Evaluation.

- tracé d'une droite à partir de a et b
- distribuer la fiche 5 et une feuille de papier millimétré. Solution sur transparent 12.
- représentation des résidus. Distribuer la feuille stagiaire 6. Solution sur transparent 13.

2.7 Les points stratégiques

- écrire l'équation d'une droite, à partir de sa représentation graphique
- notion de critère d'ajustement
- notion de variable explicative, variable expliquée
- choix de l'échelle pour faire un graphique, à partir d'un tableau de données
- choisir les unités réelles, pour calculer la pente d'une droite, à partir de sa représentation graphique (et non pas de l'unité de représentation)
- le signe dans le calcul des écarts $Y_{observé} - Y_{ajusté}$
  exemple : 5-8 = -3 (et pas 3)
- notion de variable contrôlée : à comparer à la notion de facteur contrôlé du scénario 1.

  Exemple :

  avec/sans engrais $\Longrightarrow$ facteur contrôlé
  dose 10, dose 20, dose 30, dose 50 d'engrais $\Longrightarrow$ variable contrôlée.

2.8 Mots clés

Régression linéaire : ajustement à 1 droite.
Abscisse, ordonnée X, Y
Variable explicative, variables expliquée
Variable contrôlée, réponse
Variable dépendante (de X)
Régression
Liaison, relation
Relation linéaire
Proportionnalité
modèle, ajustement, régression
Paramètre
Pente, ordonnée à l'origine
Equation
Résidus
Minimiser un critère d'ajustement
Critère des moindres carrés
Domaine de validité (extrapolation)

2.9 Liste du vocabulaire

Ajustement ou calcul d'une valeur ajustée (fitting, smoothing)
Critère des moindres carrés (least squares method)
Ordonnée à l'origine (intercept)
Pente (slope)
Régression linéaire (linear regression)
Relation linéaire
Résidu (residual)
Variable expliquée et variable explicative

2.10 Références bibliographiques

ANDERSON J.B.A. *Interpreting Data. A first course in statistics* University of Aberdeen. Chapman and Hall, London, New-York.

VAN VYVE-GENETTE A., GOHY J.M., FEYTMANS E. *Statistiques élémentaires pour les sciences biomédicales, apprentissage par la micro informatique.* De Boeck University.

de TURCKHEIM-LESQUOY E. (1985-86) *Méthodes mathématiques en biologie.* Chapitre 3 : statistique de base.

NEMENYI P., DIXON K.S., Jr. N.B., HEDSTROM L.M.. *Statistics from scratch*

R. TOMASSONE. (1987). *Mathématique et informatique Biométrie.* INA-PG, Paris.

FREEDMAN D., PISANI R., PURVES R. *Statistics.* W.W. Norton & Company, New-York, London.

DRAPER N.R., SMITH H., *Applied Regression Analysis.* John Wiley & Sons, Inc.

TOMASSONE R..LESQUOY E., MILLIER C. *Actualités scientifiques et agronomiques 13. La régression. Nouveaux regards sur une ancienne méthode statistique.* Masson.

CHAMBERS J.M., CLEVELAND W.S., KLEINER B., P.A. *Graphical Methods for Data Analysis.* Wadsworth International Group. Belmont California, Duxbury Press, Boston.

WONNACOTT T.H., WONNACOTT R.T. *Economie, Gestion, Sciences, Médecine* (avec exercices d'application). ed. Economica, 49 rue Héricart, 75015 Paris.

3 Annexes

3.1 Documents stagiaires

* Présentation du scénario
* Fiche 1 : données
* Fiche 2 : formulaire à compléter, en représentant les ajustés, les écarts, les carrés des écarts, la somme des carrés des écarts (à partir de a et b estimés graphiquement).
* Fiche 3 : formulaire général pour le calcul de a et b minimisant le critère des moindres carrés
* Fiche 4 : même formulaire, partiellement rempli, sur le jeu de données de la fiche 1
* Fiche 5 et 6 : document d'évaluation
* Fiche 7 et 8 : corrigés :
* Vocabulaire

3.2 Documents formateurs

Solution de la feuille stagiaire 4
Transparents 1 à 5 avec commentaires
Transparents 6 à 11 étapes 3 à 5
Transparents 12 à 13, solution de l'évaluation

3.3 Transparents

13 transparents