MATHÉMATIQUE

Sciences naturelles
FBD – 2e année du 2e cycle

Enseignement explicite

TRAJECTOIRE mathématique

Collecte de données en contexte fondamental
MAT-4172-2

Tiré à part

ÉDITIONS MARIE FRANCE

Claudie Lefebvre-Leblanc
Martin Francoeur

Enseignement explicite

Sciences naturelles
FBD – 2e année du 2e cycle

TRAJECTOIRE mathématique

Collecte de données
en contexte fondamental
MAT-4172-2

Tiré à part

ÉDITIONS MARIE FRANCE

Claudie Lefebvre-Leblanc
Martin Francoeur

Révision linguistique : Annie St-Germain
Révision scientifique : Alec Laporte
Correction d'épreuves : Doris Lizotte
Conception et réalisation : Interscript
Couverture : LaSo Design

Dépôt légal 2e trimestre 2014
Bibliothèque et Archives Canada
Bibliothèque et Archives nationales du Québec

Éditions Marie-France sont membres de ASSOCIATION NATIONALE DES ÉDITEURS DE LIVRES

ISBN : 978-2-89661-181-2

Imprimé au Canada

Nous reconnaissons l'aide financière du gouvernement du Canada par l'entremise du Fonds du livre du Canada pour nos activités d'édition.

Table des matières

SECTION 1

Distribution à deux caractères et corrélation

SECTION 2

Droites de régression

SECTION 3

Interpolation et extrapolation

SECTION 4

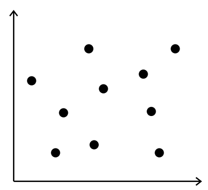
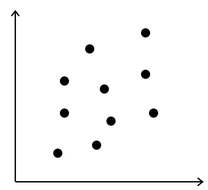
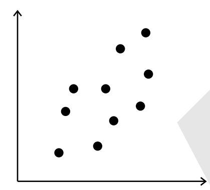
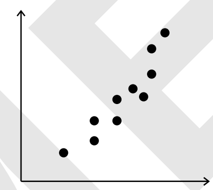
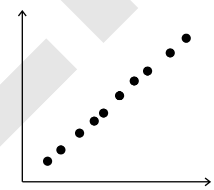
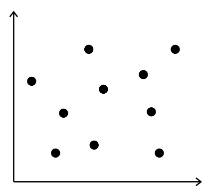
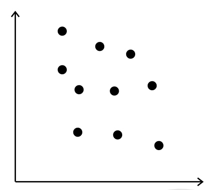
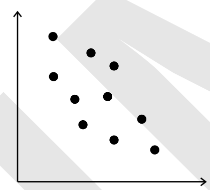
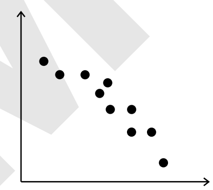
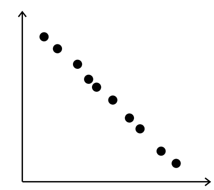
Situations-problèmes

 978-2-89661-181-2

Coefficient de corrélation linéaire

Il est possible de quantifier la force d'une corrélation, c'est-à-dire du lien entre deux variables, à l'aide du coefficient de corrélation, noté « *r* ». Ce coefficient doit se situer entre ⁻1 et 1, où un coefficient de 1 (ou ⁻1) indique une corrélation parfaite, et un coefficient de 0 indique une corrélation nulle.

Le tableau suivant présente les différents niveaux de corrélation, de même que leur sens.

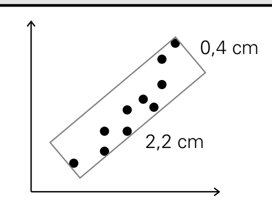
	$r = 0$	$0 < r < 1$			$r = 1$
	Nulle	**Faible**	**Moyenne**	**Forte**	**Parfaite 1**
POSITIVE					
NÉGATIVE					

Méthode du rectangle (ou de l'ellipse)

Il est possible d'estimer le coefficient de corrélation d'une distribution avec la méthode du rectangle (aussi appelée méthode de l'ellipse). Il faut représenter les données dans un nuage de points, encadrer tous les points de la distribution et employer la formule suivante :

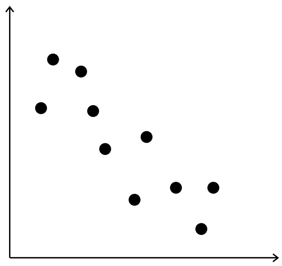
$$r = \pm\left(1 - \frac{\text{mesure de la largeur}}{\text{mesure de la longueur}}\right)$$

Ex. : Estimez le coefficient de corrélation de la distribution suivante.

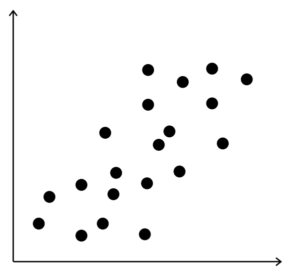
1	Si ce n'est pas fait, tracer le nuage de points.	
2	Construire le plus petit rectangle possible autour du nuage de points et en mesurer la longueur et la largeur.	
3	Déterminer le sens de la corrélation : • Positive si croissante ; • Négative si décroissante.	La corrélation est positive.
4	Calculer le coefficient à l'aide de la formule $r = \pm\left(1 - \frac{\text{mesure de la largeur}}{\text{mesure de la longueur}}\right)$	$r = +\left(1 - \frac{0,4}{2,2}\right) \approx 0,82$

8. Estimez le coefficient de corrélation de ces distributions à l'aide de la méthode du rectangle et qualifiez la corrélation.

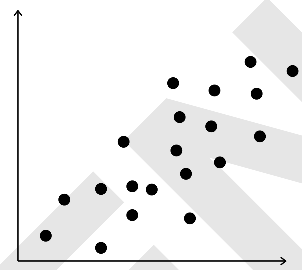
a)



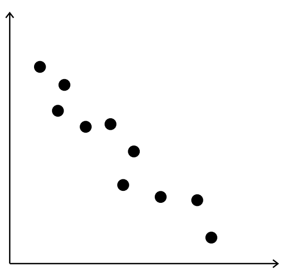
b)



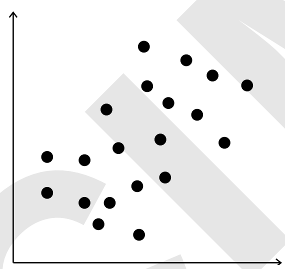
c)



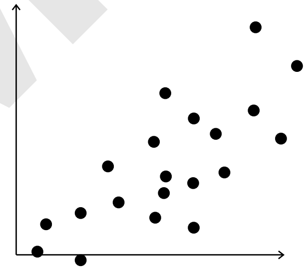
d)



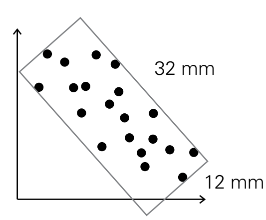
e)



f)

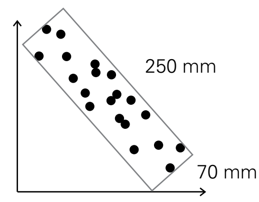


9. Quelle est la valeur approximative du coefficient de corrélation estimé à l'aide de la méthode du rectangle?



a) 0,625 **b)** 0,375 **c)** $^{-}0,625$ **d)** $^{-}0,375$

10. Quelle est la valeur approximative du coefficient de corrélation estimé à l'aide de la méthode du rectangle?



a) 0,72 **b)** 0,28 **c)** $^{-}0,72$ **d)** $^{-}0,28$

978-2-89661-181-2

Méthode de la droite médiane-médiane

La méthode de la droite médiane-médiane est moins sensible aux données aberrantes et est plus efficace que la méthode de la droite de Mayer lorsque le nombre de données est élevé. Voici comment procéder.

Ex. : Déterminez l'équation de la droite de régression de la distribution suivante à l'aide de la méthode de la droite médiane-médiane.

x	14	6	27	14	3	19	22	9
y	10	5	17	6	2	14	21	7

<table>
<tr><td>1</td><td>Classer les données en ordre croissant selon la valeur de la première variable (x). Pour deux valeurs égales de la variable x, ordonner selon la variable y.</td><td><table><tr><td>x</td><td>3</td><td>6</td><td>9</td><td>14</td><td>14</td><td>19</td><td>22</td><td>27</td></tr><tr><td>y</td><td>2</td><td>5</td><td>7</td><td>6</td><td>10</td><td>14</td><td>21</td><td>17</td></tr></table></td></tr>
<tr><td>2</td><td>Diviser la distribution en trois groupes.
• Les 1^{er} et 3^e groupes doivent contenir le même nombre de données.
• Le nombre de données du 2^e groupe doit se rapprocher le plus possible de celui des 1^{er} et 3^e groupes.</td><td><table><tr><td>x</td><td>3</td><td>6</td><td>9</td><td>14</td><td>14</td><td>19</td><td>22</td><td>27</td></tr><tr><td>y</td><td>2</td><td>5</td><td>7</td><td>6</td><td>10</td><td>14</td><td>21</td><td>17</td></tr></table></td></tr>
<tr><td>3</td><td>Déterminer la médiane des abscisses et la médiane des ordonnées pour chacun des groupes, qui composeront les coordonnées des points M_1, M_2 et M_3.</td><td rowspan="2"><table><tr><td></td><td>X</td><td>Y</td></tr><tr><td>M_1</td><td>6</td><td>5</td></tr><tr><td>M_2</td><td>14</td><td>8</td></tr><tr><td>M_3</td><td>22</td><td>17</td></tr><tr><td>P</td><td>14</td><td>10</td></tr></table></td></tr>
<tr><td>4</td><td>Déterminer la moyenne des abscisses et la moyenne des ordonnées des points M_1, M_2 et M_3, qui formeront les coordonnées du point P.</td></tr>
<tr><td>5</td><td>Déterminer l'équation de la droite :
• Dont la pente est parallèle (identique) à celle de la droite passant par M_1 et M_3.
• Qui passe par le point P.</td><td>$m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{17 - 5}{22 - 6} = 0{,}75$
$y = 0{,}75x + b$
$10 = 0{,}75 \cdot 14 + b$
$^{-}0{,}5 = b$
$y = 0{,}75x - 0{,}5$</td></tr>
</table>

5. Déterminez l'équation de la droite de régression des distributions suivantes en utilisant la méthode de la droite médiane-médiane.

a)

x	27	15	19	13	25	17	16	14	11
y	9	8	6	1	3	6	6	3	4

4. Le tableau suivant présente le nombre total d'habitants dans les 10 villes québécoises les plus peuplées, de même que le nombre total d'habitants dans la province.

Sachant qu'en 2011 il y avait un total de 3 742 728 habitants dans les 10 plus grandes villes québécoises, estimez le nombre d'habitants total dans la province.

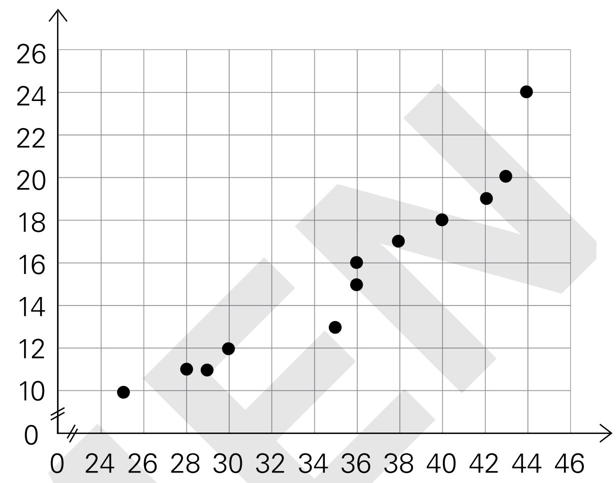
Année	N^bre total d'habitants des 10 villes les plus peuplées	N^bre total d'habitants dans la province
1951	1 945 121	4 055 679
1956	2 252 755	4 628 378
1961	2 643 279	5 259 211
1966	2 982 494	5 780 845
1971	3 134 754	6 027 764
1976	3 174 708	6 234 445
1981	3 145 295	6 438 400
1986	3 195 993	6 532 461
1991	3 338 571	6 895 963
1996	3 407 382	7 138 834
2001	3 467 745	7 238 345
2006	3 594 149	7 546 131

Source : http://www.stat.gouv.qc.ca/donstat/societe/demographie/dons_regnl/regional/Tableau_top_10.htm

978-2-89661-181-2

5. À l'aide de la droite de Mayer, complétez la table de valeurs suivante.

x	5		34	51	
y		3			42



 978-2-89661-181-2

Choisir les données à traiter

Dans les problèmes liés aux statistiques, on retrouve un nombre important de données, et il n'est souvent pas nécessaire de toutes les traiter. Dans le cas des modèles linéaires, il peut alors être utile de déterminer la variable ayant le plus d'influence sur la variable étudiée dans le problème en utilisant le coefficient de corrélation linéaire.

Si le coefficient de corrélation linéaire entre deux variables est faible, cela peut indiquer que ces données ne sont pas pertinentes puisqu'elles ne sont que faiblement liées à la variable étudiée. Il est alors préférable de traiter la variable (ou les variables) étant le plus fortement liée à la variable étudiée, c'est-à-dire celle (ou celles) indiquant un coefficient de corrélation linéaire plus élevé.

Ex.: Voici des informations concernant 10 maisons situées dans le quartier de Sabin. Proposez à Sabin un moyen rapide et fondé sur une démarche statistique d'estimer la valeur de sa maison.

Valeur maison	Superficie habitable (m^2)	Superficie terrain (m^2)	N^bre^ chambres	N^bre^ salles de bain	Année construction
122 500	845	5 900	3	1	1994
128 000	950	6 150	2	1	1990
135 000	925	5 900	4	2	2000
142 500	1 000	6 275	2	2	1996
158 500	1 015	7 000	3	1	2005
165 000	1 085	6 305	2	1	2007
174 000	1 175	6 850	4	3	2009
178 500	1 150	6 615	2	2	2010
189 000	1 250	7 200	3	1	2006

Pour déterminer la variable pour laquelle le lien est le plus fort avec le prix de la maison, on calcule le coefficient de corrélation linéaire pour chacune d'elles à l'aide de la calculatrice.

Superficie habitable/Coût maison : 0,97
Superficie terrain/Coût maison : 0,85
N^bre^ de chambres/Coût maison : 0,03
N^bre^ de salles de bain/Coût maison : 0,22
Année construction/Coût maison : 0,89

Les coefficients de corrélation nous indiquent que c'est la superficie habitable qui est le plus fortement liée au prix de la maison. Toujours à l'aide de la calculatrice, on détermine l'équation de la droite de régression : $y = 172{,}9523x - 25\,553{,}1875$.

On pourrait donc conseiller à Sabin d'utiliser la règle suivante : $y = 172{,}9523x - 25\,553{,}1875$ où $\boldsymbol{x}$ est la superficie habitable de sa maison (m^2), et $\boldsymbol{y}$, la valeur de sa maison.

978-2-89661-181-2

1. Le tableau ci-dessous présente certaines mensurations de 10 hommes âgés de 27 ans. À partir de ces données, déterminez le moyen le plus efficace d'estimer la taille d'un homme de 27 ans.

Taille (m)	Longueur pied (cm)	Longueur pouce (cm)	Étendue bras (cm)	Largeur bouche (cm)	Tour de taille (cm)
1,54	25,8	6,1	150	5,2	84
1,65	24,7	5,8	162	6,1	90
1,72	28,7	6	170	4,9	81
1,73	26,4	5,9	175	5,8	87
1,75	28,2	5,5	178	6,4	78
1,77	29,1	5,7	175	5,7	85
1,79	27,4	6,1	181	6,1	95
1,81	25,6	5,8	184	5,9	94
1,85	28,5	6,4	183	5,8	92
1,88	26,9	5,1	189	5,7	98

a) Quels sont les coefficients de corrélation liant chaque variable à la taille des hommes ?

b) Quelle variable est le plus fortement liée à la taille des hommes ?

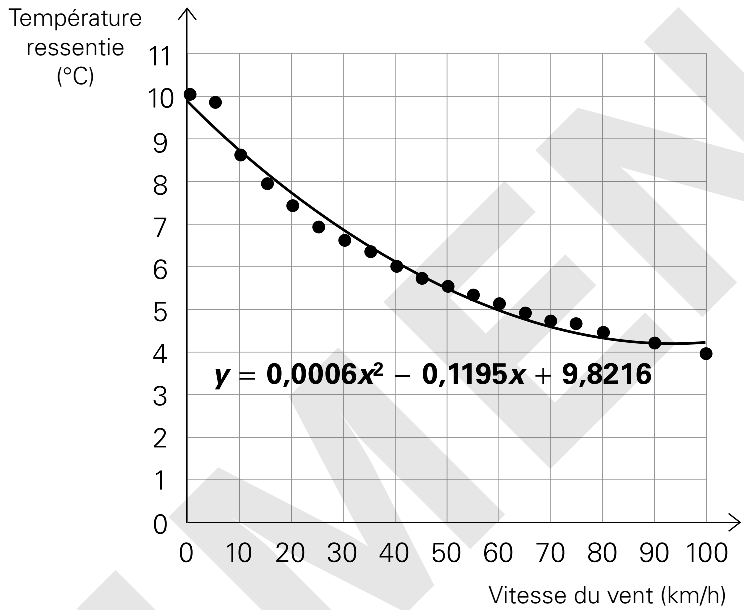
c) Quelle est l'équation de la droite de régression associant cette variable à la taille des hommes ?

1. Refroidissement éolien

Au bulletin météo, on entend souvent parler du facteur éolien. Il s'agit en fait du vent qui abaisse la température ressentie à l'extérieur. Les météorologistes ont déterminé des équations leur permettant de calculer la température ressentie en fonction de la vitesse du vent et de la température initiale extérieure.

Le diagramme suivant présente la température ressentie (°C) selon la vitesse du vent (km/h) pour une température initiale de 10 °C. On peut modéliser cette situation à l'aide d'une parabole dont l'équation est présentée dans le diagramme.

Température ressentie (°C) selon la vitesse du vent (km/h) lorsqu'il fait 10 °C



Vitesse vent (km/h)	Temp. ressentie pour une temp. initiale de ...		
	0 °C	**−15 °C**	**−30 °C**
0	0,0	−15,0	−30,0
10	−3,3	−21,2	−39,2
20	−5,2	−24,2	−43,1
30	−6,5	−26,0	−45,6
40	−7,4	−27,4	−47,5
50	−8,1	−28,6	−49,0
60	−8,8	−29,5	−50,3
70	−9,3	−30,4	−51,4
80	−9,8	−31,1	−52,4
90	−10,2	−31,8	−53,3
100	−10,6	−32,4	−54,1

Le tableau ci-contre présente les températures ressenties selon la vitesse du vent pour des températures initiales de 0 °C, de −15 °C et de −30 °C.

On peut supposer que, pour une température initiale autre que 10 °C, il suffirait d'appliquer une translation verticale à cette parabole pour obtenir une nouvelle équation modélisant la température ressentie selon la vitesse du vent pour cette nouvelle température initiale.

À partir des données fournies, vérifiez cette supposition.

6. Mise en forme

L'IMC est l'indice de masse corporelle. Il permet d'évaluer les risques de maladies liées à un surplus ou à une insuffisance de poids. On le calcule en divisant le poids (en kg) par le carré de la taille (en m). Voici un tableau présentant les risques associés aux différentes valeurs de l'IMC.

IMC – Risques pour la santé	
< **20**	Peut être associé à des problèmes de santé ou de nutrition chez certaines personnes
[20, 25]	IMC sain pour la plupart des adultes
]25, 27]	Risques possibles de problèmes de santé et de maladies chez certaines personnes
> **27**	Risque accru de problèmes de santé et de maladies

Source : http://www1.pharmaprix.ca/fr/Health-And-Pharmacy/Health-Tools/BMICalculator.aspx

Le tableau ci-dessous présente différentes données concernant la taille, le poids et les habitudes de vie de 12 femmes de 30 ans.

Josiane mesure 1,60 m et pèse 66,6 kg. Elle mange en moyenne huit repas préparés et prend trois consommations d'alcool par semaine. Elle mange quotidiennement cinq portions de fruits et légumes. Elle souhaite modifier ses habitudes de vie pour obtenir un IMC sain. À l'aide d'une démarche statistique, conseillez Josiane sur les modifications à apporter à ses habitudes de vie.

	Taille (m)	Masse (kg)	N^{bre} mets préparés consommés hebdomadairement	N^{bre} consommations alcool hebdomadaires	N^{bre} portions fruits et légumes
1	1,54	45	2	1	7
2	1,72	68	3	8	5
3	1,65	75	4	9	3
4	1,8	90	6	4	2
5	1,6	86	10	8	1
6	1,7	95	9	5	0
7	1,55	65	5	3	4
8	1,85	71	2	0	6
9	1,75	82	7	1	5
10	1,68	76	4	7	7
11	1,9	88	5	2	5
12	1,78	78	4	3	6