



Que disent vos pieds? : diagrammes de dispersion et droites de meilleur ajustement

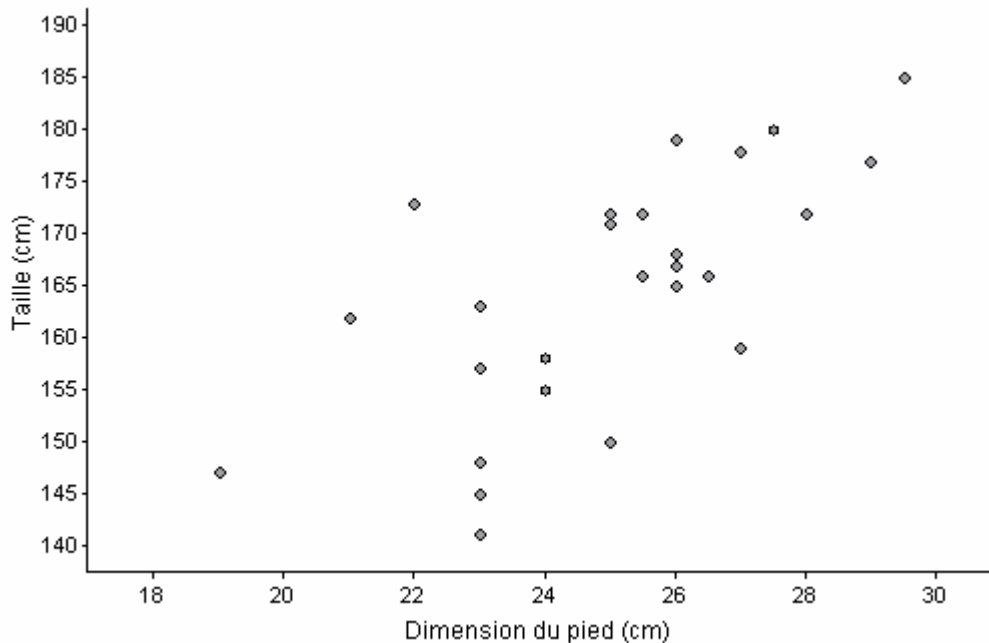
Feuille de travail de l'élève

Que disent vos pieds concernant votre taille? Pouvez-vous prédire la taille d'une personne par la dimension de ses pieds?

Si un élève de 10^e année a des pieds mesurant 27 cm de long, quelle pourrait bien être sa taille? Dans cette activité, vous utiliserez les données du Recensement à l'école pour déterminer s'il existe un rapport entre la dimension des pieds et la taille de l'élève. S'il y en a un, quel est-il?

Les élèves d'Afrique du Sud, du Royaume-Uni et de la Nouvelle-Zélande ont mesuré la dimension de leurs pieds ainsi que leur taille en centimètres, puis ils ont entré ces renseignements dans la base de données du Recensement à l'école. On a constitué un échantillon aléatoire à partir des données regroupées pour créer le diagramme de dispersion ci-dessous, où chaque point représente les paires ordonnées indiquant la taille en fonction de la dimension du pied.

Diagramme de dispersion de la taille d'élèves de 15 ans et de la dimension de leurs pieds



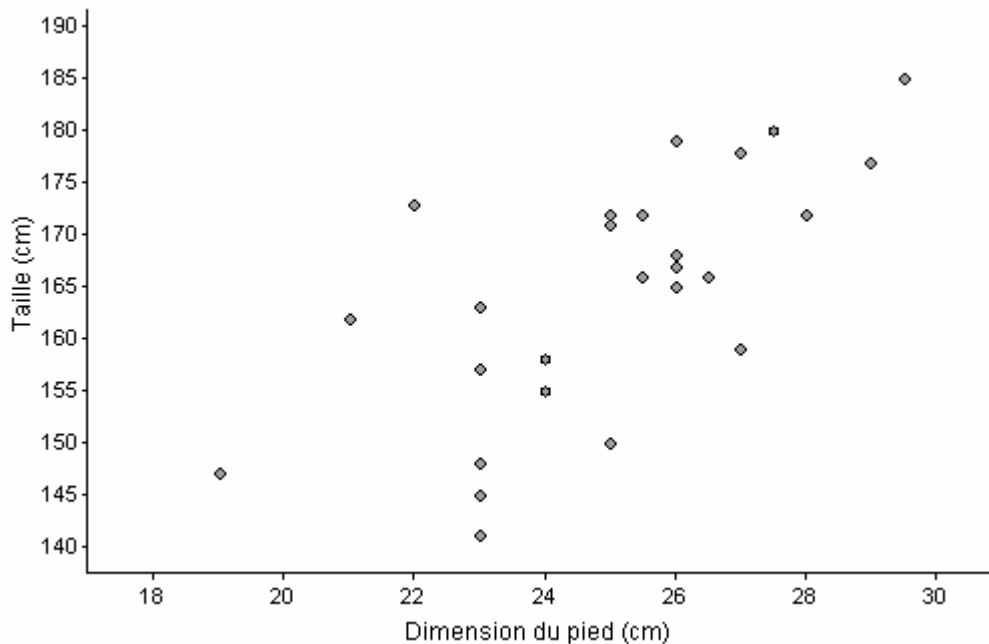
Examinez ce diagramme de dispersion

Existe-t-il une corrélation entre la dimension du pied et la taille?

Si oui, s'agit-il d'une corrélation forte ou faible, positive ou négative?

1^{re} partie : droite de meilleur ajustement

Figure 1. Taille d'élèves de 15 ans et dimension de leurs pieds



1. Manuellement, tracez une droite de meilleur ajustement à l'aide des points se trouvant à la figure 1.
2. Pour cette droite que vous venez de dessiner, trouvez :
 - a) sa pente;
 - b) son équation;
 - c) la hauteur du point d'intersection. (Expliquez ce que cela signifie et dites s'il s'agit ou non d'un scénario plausible.)
3. Comment avez-vous établi que cette droite offrait le « meilleur ajustement » ? (Note : Les échelles ne commencent pas à zéro dans la figure 1.)
4. Comparez votre droite à celles de plusieurs de vos camarades de classe. Est-ce que vous obtenez tous la même droite ? (Note : Comparez les droites visuellement, puis faites la comparaison des équations que vous et vos camarades de classe ont calculé pour obtenir ces droites.)
5. Lorsque des personnes différentes analysent les mêmes données, est-il important qu'elles obtiennent le même résultat ?

2^e partie : méthode de la droite médiane-médiane

La *droite médiane-médiane* est l'une des méthodes permettant de déterminer une droite de meilleur ajustement. Dans le tableau suivant, on donne la dimension des pieds ainsi que la taille de 30 élèves de 15 ans. (On a utilisé ces mêmes données dans la 1^{re} partie.)

Tableau 1. Données sur la taille d'élèves de 15 ans et sur la dimension de leurs pieds

Pays	Sexe	Date de naissance (aaaa-mm-jj)	Taille (cm)	Dimension du pied (cm)
Queensland	F	1986-02-25	180	27,5
Afrique du Sud	F	1986-09-07	163	23,0
Royaume-Uni	M	1985-01-25	185	29,5
Queensland	M	1986-03-26	166	26,5
Afrique du Sud	F	1986-04-17	172	25,0
Afrique du Sud	M	1986-05-10	178	27,0
Royaume-Uni	F	1985-04-30	162	21,0
Afrique du Sud	F	1986-06-20	148	23,0
Afrique du Sud	M	1986-06-15	167	26,0
Afrique du Sud	M	1986-08-08	179	26,0
Afrique du Sud	M	1985-11-10	147	19,0
Afrique du Sud	F	1986-07-15	158	24,0
Royaume-Uni	M	1985-09-15	155	24,0
Royaume-Uni	M	1985-12-29	172	28,0
Afrique du Sud	M	1986-01-29	141	23,0
Afrique du Sud	M	1986-07-23	150	25,0
Afrique du Sud	F	1986-04-06	145	23,0
Royaume-Uni	F	1985-08-10	171	25,0
Afrique du Sud	F	1985-11-27	157	23,0
Afrique du Sud	F	1985-11-26	157	23,0
Queensland	F	1986-04-03	166	25,5
Afrique du Sud	M	1986-01-24	159	27,0
Royaume-Uni	M	1985-02-09	172	25,5
Royaume-Uni	M	1985-07-09	168	26,0
Afrique du Sud	F	1986-09-06	163	23,0
Afrique du Sud	M	1986-07-22	150	25,0
Afrique du Sud	M	1985-12-19	177	29,0
Royaume-Uni	M	1985-02-08	172	25,5
Royaume-Uni	M	1985-02-22	172	22,0
Queensland	F	1985-11-09	165	26,0

1. Afin de simplifier le travail, nous examinerons seulement les élèves de sexe masculin du tableau 1.

Garçons		
Dimension du pied (cm)	Taille (cm)	Étapes
19,0	147	<ol style="list-style-type: none"> 1. Organisez les données de façon à ce que les valeurs x (dimension du pied) soient triées par ordre croissant. 2. Divisez les données en trois groupes. Si le nombre de données ne peut être divisé également, assurez-vous que les premier et troisième groupes contiennent exactement le même nombre de données et rajuster le groupe du milieu de façon à ce qu'il contienne seulement une donnée de plus ou de moins. Comme il y a 17 données, faites des groupes de 6, 5 et 6 éléments. 3. Identifiez les données en utilisant une couleur différente pour chaque groupe. 4. Examinez le premier groupe. <ul style="list-style-type: none"> • Quelle est la valeur médiane x? • Quelle est la valeur médiane y? Inscrivez ces valeurs sous forme de paire ordonnée (x,y). Il s'agit du point médian pour le premier groupe. Appelons-le S_1. Identifiez la paire ordonnée (utilisez le signe $(+)$ ou un carré plutôt qu'un point, mais servez-vous de la même couleur que celle du reste du groupe). 5. Reprenez l'étape n° 4 pour les groupes de données du milieu et le dernier groupe de données pour obtenir les points S_2 et S_3, respectivement. 6. Tracez (légèrement) la droite passant par S_1 et S_3. Trouvez la pente de la droite S_1S_3. 7. Calculez l'équation d'une deuxième droite passant par le point S_2 et qui est parallèle à la droite S_1S_3. 8. Déplacez la droite S_1S_3 d'un tiers vers S_2. Vous obtenez la droite médiane-médiane. Votre enseignant fera les étapes avec vous.
22,0	173	
23,0	141	
24,0	155	
25,0	150	
25,0	150	
25,5	172	
25,5	172	
26,0	167	
26,0	179	
26,0	168	
26,5	166	
27,0	159	
27,0	178	
28,0	172	
29,0	177	
29,5	185	

Tracez la droite médiane-médiane en suivant les mêmes étapes, mais en utilisant les données se rapportant aux élèves de sexe féminin du tableau 1.

Filles		
Dimension du pied (cm)	Taille (cm)	
		Le 1 ^{er} groupe contient _____ paires de données.
		Le 2 ^e groupe contient _____ paires de données.
		Le 3 ^e groupe contient _____ paires de données.
21,0	162	
23,0	163	$S_1 = (\quad , \quad)$ $S_2 = (\quad , \quad)$ $S_3 = (\quad , \quad)$
23,0	148	
23,0	145	La pente de la droite passant par S_1 et S_2 est _____.
23,0	157	
23,0	157	L'équation de la droite passant par S_1 et S_2 est
23,0	163	_____.
24,0	158	L'équation de la droite passant par S_3 parallèle à la
25,0	171	droite reliant S_1 et S_2 est
25,0	172	_____.
25,5	166	L'équation de la droite médiane-médiane est
26,0	165	_____.
27,5	180	

3. Pour déterminer la droite médiane-médiane pour les élèves de sexe masculin et ceux de sexe féminin regroupés ensemble, reportez-vous au tableau 1 et entrez les données sur la dimension du pied et la taille dans L_1 et L_2 de votre calculatrice. Ensuite, utilisez l'option médiane-médiane du menu STAT de votre calculatrice. (Utilisez une calculatrice graphique comme la calculatrice T1-83 ou T1-84 pour obtenir des résultats plus rapidement, surtout s'il y a une foule de données.)
4. Tracez la droite médiane-médiane dans votre diagramme de dispersion et identifiez-la clairement.
5. Comment pouvez-vous la comparer à la première droite que vous avez tracée manuellement sur les mêmes données de la figure 1? Permet-elle un meilleur ajustement?
6. Selon l'équation, si les pieds de l'élève mesurent 27 cm de long, quelle sera la taille probable de cet élève?

3^e partie : méthode de régression des moindres carrés

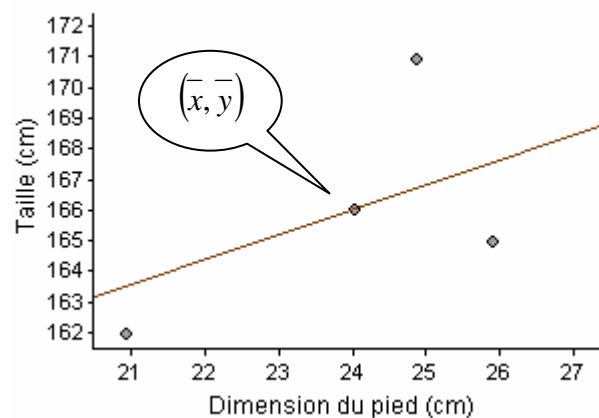
La *régression des moindres carrés* est une autre méthode permettant de déterminer une droite de meilleur ajustement. Cette méthode est fondée sur la moyenne. Elle est difficile à créer si l'on ne dispose pas d'un outil comme une calculatrice graphique ou un programme d'ordinateur. Toutefois, cette méthode est fort utile à connaître. Dans l'exemple ci-dessous, on montre en quoi consiste cette méthode.

Exemple de régression des moindres carrés

X Dimension du pied (cm)	Y Taille (cm)
21	162
25	171
26	165

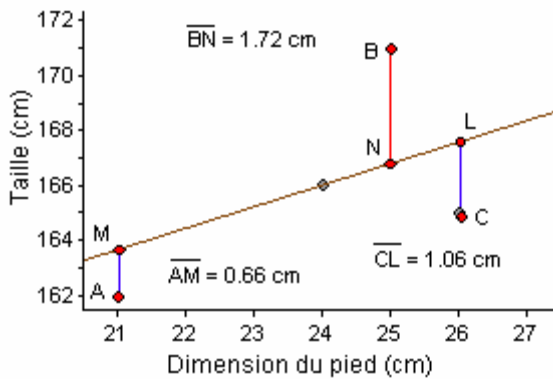
$$\bar{x} = \underline{\hspace{2cm}} \quad \bar{y} = \underline{\hspace{2cm}}$$

Il semble normal de faire passer la droite de meilleur ajustement par le point (\bar{x}, \bar{y}) . Il est possible de tracer plusieurs de ces lignes passant par le point (\bar{x}, \bar{y}) . Le diagramme montre l'une d'entre elles.



Comment détermine-t-on la droite de meilleur ajustement devant passer par le point (\bar{x}, \bar{y}) ?

Trouvez le déplacement vertical (distance) à partir de chacun des points de la droite. La droite dont la somme de ces déplacements verticaux est la plus petite possible représente la droite de meilleur ajustement.



Dans le graphique ci-contre, comme le point B est au-dessus de la ligne, son déplacement vertical à partir de la droite est positif. Les points A et C se trouvent sous la ligne, alors leurs déplacements verticaux sont donc négatifs. Quelle est la somme de ces déplacements verticaux?

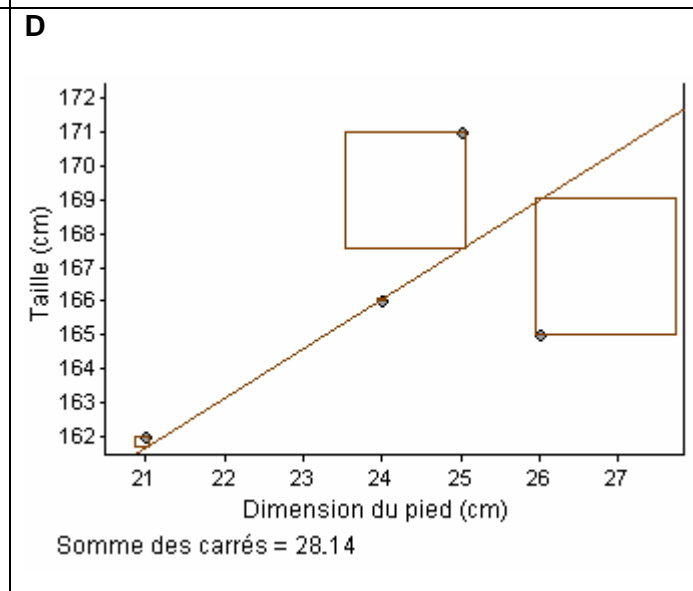
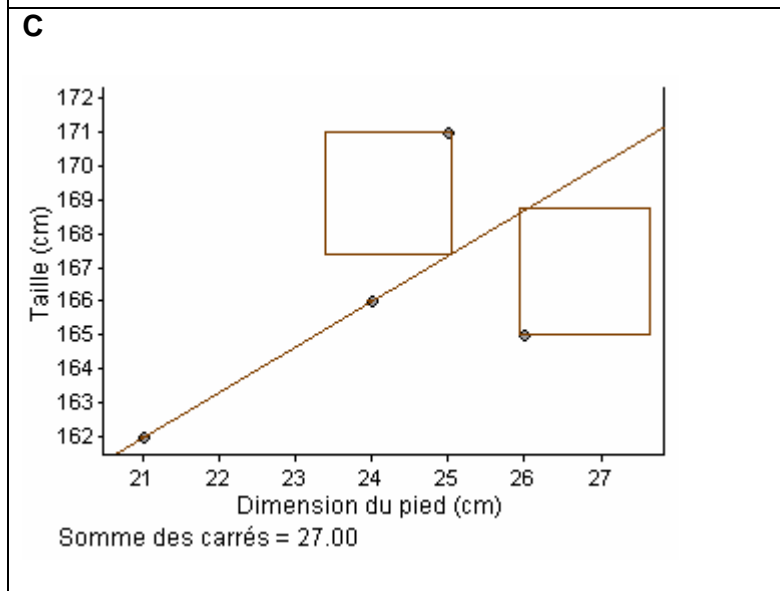
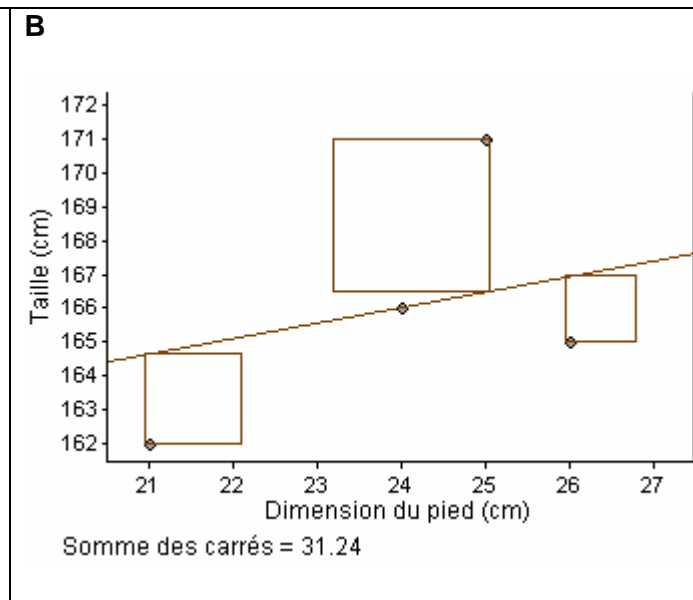
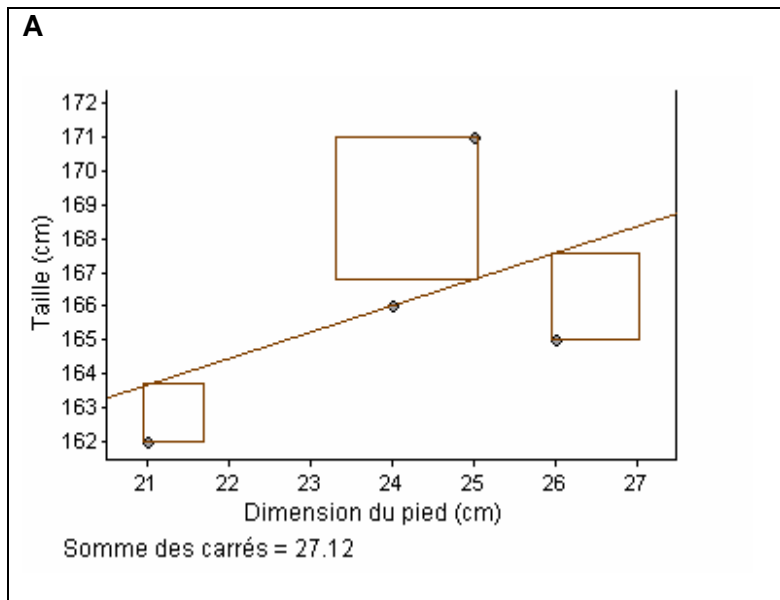
Toute ligne passant par le point (\bar{x}, \bar{y}) donnera le même résultat — zéro —, et ce, en raison de l'« annulation » qui se produit lorsque la somme des distances est positive et que les autres sont négatives. Par conséquent, une approche plus élaborée est nécessaire.

Pour éviter cet effet d'annulation, on doit élever les distances au carré. La droite de régression linéaire est également appelée *droite de régression des moindres carrés*, car celle-ci est créée lorsque la somme des carrés est à son minimum.

La quadrature de la ligne AM donne l'aire d'un carré dont chaque côté est égal à \overline{AM} .

Dans le diagramme A ci-dessous, on trace un carré dont chaque côté est égal à \overline{AM} . On dessine également un carré dont chaque côté est égal à \overline{BN} ainsi qu'un autre carré dont chaque côté est égal à \overline{CL} . Notre objectif est de déplacer la ligne passant par le point (\bar{x}, \bar{y}) jusqu'à ce qu'on obtienne la somme de ces carrés la plus petite.

Les diagrammes B, C et D montrent les carrés créés lorsqu'on déplace la ligne passant au point (\bar{x}, \bar{y}) . Examinez seulement les carrés. Pouvez-vous déterminer dans quel diagramme on trouve les carrés dont la somme des aires est la plus petite?



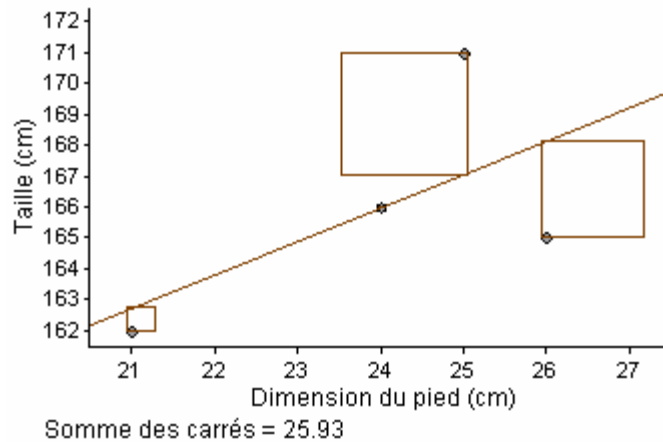
Comme les différences sont très petites, il peut être difficile de voir quel diagramme contient les carrés dont la somme des aires est minimale. Examinez-les de nouveau, mais en portant attention cette fois-ci aux valeurs de la somme des carrés de chaque diagramme.

Après les divers essais avec la droite, la somme minimale des carrés est de 25,93.

Lorsque la somme des carrés est de 25,93, l'équation de la droite est $y = 1,07x + 140,3$.

Celle-ci s'appelle *droite de régression des moindres carrés* ou *droite de régression linéaire*. On peut également utiliser une calculatrice TI-83 pour déterminer cette droite.

Si le coefficient de variation (r^2) est près de 1, la corrélation est forte. Si r^2 est près de zéro, il y a alors une faible corrélation entre les deux variables. Que pouvez-vous conclure concernant la corrélation entre les deux variables figurant dans le graphique ci-contre, où $r^2 = 0,38$?



1. Utilisez votre calculatrice ou un logiciel (p. ex. un tableur ou Fathom) pour déterminer l'équation de régression linéaire pour les données originales mettant en rapport la dimension du pied et la taille (tableau 1).
2. Tracez cette droite de régression des moindres carrés dans le diagramme de dispersion et identifiez-la clairement.
3. Utilisez-la pour prédire la taille d'une personne dont les pieds mesurent 27 cm de long. Cette réponse diffère-t-elle beaucoup de la réponse que vous avez trouvée à l'aide de la droite médiane-médiane? Selon vous, quelle méthode vous donne le meilleur ajustement? Pourquoi?

4^e partie : un projet à réaliser à l'aide des données de votre classe!

1. Utilisez les données du Recensement à l'école relatives à votre classe pour créer un diagramme de dispersion montrant la dimension des pieds (y) par rapport à l'empan de la main (x). La corrélation est-elle forte ou faible?
2. Déterminez l'équation de la droite médiane-médiane.
3. Tracez celle-ci dans votre diagramme de dispersion et identifiez-la clairement.
4. Déterminez l'équation de la droite de régression des moindres carrés.
5. Tracez celle-ci dans votre diagramme de dispersion et identifiez-la clairement.
6. Selon chaque modèle, quelle est la dimension du pied prédite pour une main dont l'empan est de 18 cm?

Collaboration : Anna Spanik, professeure de mathématiques, école secondaire Halifax West, Nouvelle-Écosse.