

Corrélation et régression linéaire

Julie Milot

Professeure de mathématique
Département de mathématiques
Collège de Maisonneuve
jmilot@cmaisonneuve.qc.ca



Ressource développée dans le cadre du projet Mathéma-TIC
Financé par le ministère de l'Enseignement supérieur, de la Recherche et de la Science (MESRS)
du Québec dans le cadre du Programme d'arrimage universités-collèges

Mise en situation

Âge	Revenu annuel (\$)
15	2 500
16	2 600
17	3 200
18	3 750
19	7 750
20	11 250
21	13 750
22	15 250
23	17 500
24	20 000
25	23 230
26	24 000
27	26 630
28	27 560
29	29 000
30	31 500
31	31 400
32	33 780
33	34 650
34	35 780
35	37 000
36	35 600
37	38 540
38	45 200
39	43 000
40	44 600
41	48 550
42	48 000
43	49 700
44	50 340
45	52 310

Existe-t-il un lien statistique entre l'âge et le revenu annuel?

Si oui, quelle est la nature de ce lien? Peut-on l'exprimer à l'aide d'une fonction linéaire ?

Est-ce que la fonction illustrant le lien entre les variables permettrait de faire des prévisions de revenu pour un âge donné?

Identification des variables

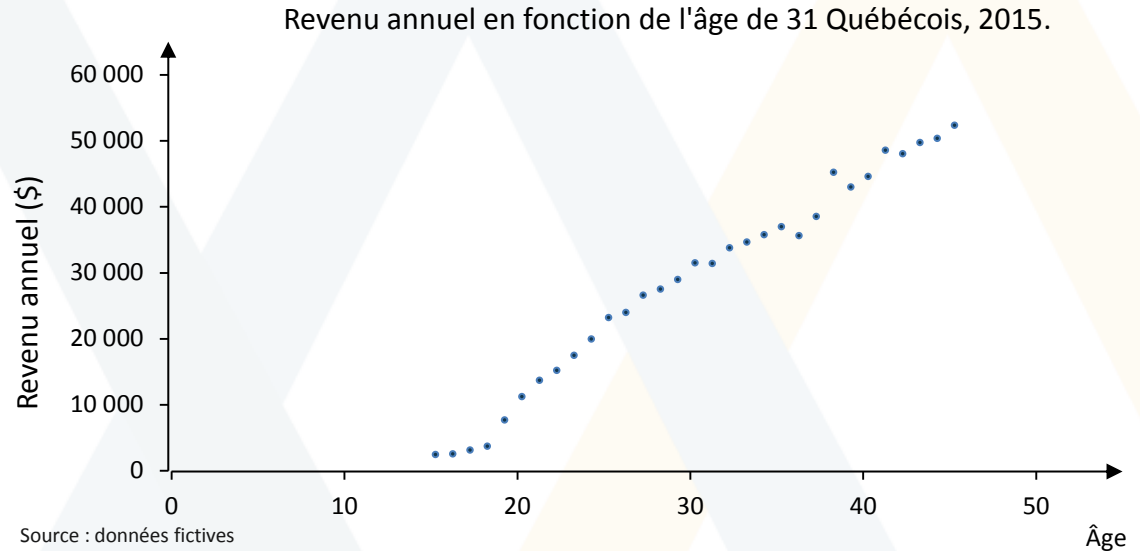
Âge	Revenu annuel (\$)
15	2 500
16	2 600
17	3 200
18	3 750
19	7 750
20	11 250
21	13 750
22	15 250
23	17 500
24	20 000
25	23 230
26	24 000
27	26 630
28	27 560
29	29 000
30	31 500
31	31 400
32	33 780
33	34 650
34	35 780
35	37 000
36	35 600
37	38 540
38	45 200
39	43 000
40	44 600
41	48 550
42	48 000
43	49 700
44	50 340
45	52 310

Quelle est la variable indépendante, la variable explicative?

Quelle est la variable dépendante, la variable expliquée?

Représentation des valeurs à l'aide d'un nuage de points

Âge	Revenu annuel (\$)
15	2 500
16	2 600
17	3 200
18	3 750
19	7 750
20	11 250
21	13 750
22	15 250
23	17 500
24	20 000
25	23 230
26	24 000
27	26 630
28	27 560
29	29 000
30	31 500
31	31 400
32	33 780
33	34 650
34	35 780
35	37 000
36	35 600
37	38 540
38	45 200
39	43 000
40	44 600
41	48 550
42	48 000
43	49 700
44	50 340
45	52 310



Coefficient de corrélation linéaire

Nombre permettant de déterminer l'intensité d'un lien linéaire entre deux variables quantitatives à échelle d'intervalle ou de rapport.

$$r = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{(n-1)s_x s_y}$$

Le coefficient de corrélation linéaire r est une valeur qui n'a pas d'unité et qui est toujours comprise entre -1 et 1.

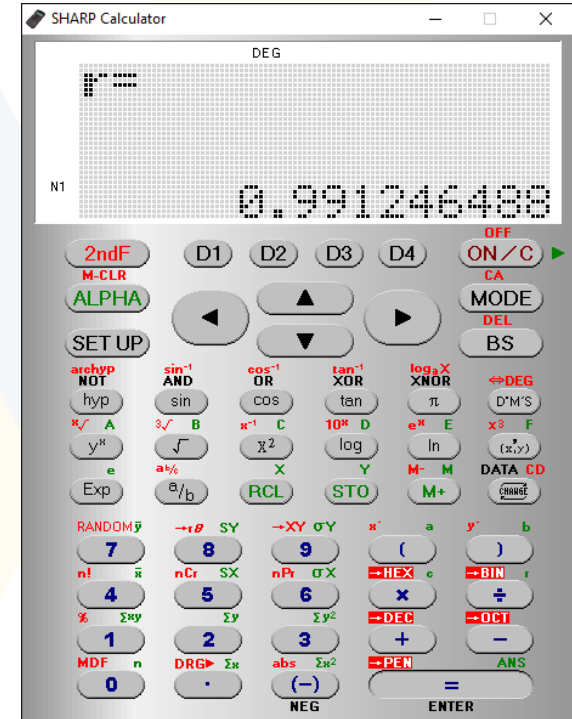
Un coefficient de corrélation linéaire positif indique un lien linéaire positif alors que si r est négatif, le lien linéaire entre les deux variables est négatif.

Plus la valeur de r est près de -1 ou 1, plus le lien linéaire entre les deux variables est fort.

Détermination du coefficient de corrélation linéaire r

Âge	Revenu annuel (\$)
15	2 500
16	2 600
17	3 200
18	3 750
19	7 750
20	11 250
21	13 750
22	15 250
23	17 500
24	20 000
25	23 230
26	24 000
27	26 630
28	27 560
29	29 000
30	31 500
31	31 400
32	33 780
33	34 650
34	35 780
35	37 000
36	35 600
37	38 540
38	45 200
39	43 000
40	44 600
41	48 550
42	48 000
43	49 700
44	50 340
45	52 310

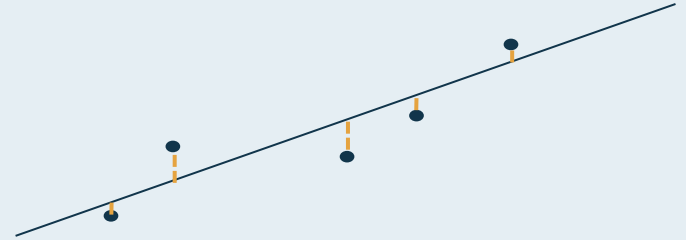
Interprétation :



Droite de régression

Droite minimisant la somme des carrés des distances verticales entre les points d'un nuage de points et elle-même.

$$y' = a + bx \quad \text{où} \quad b = r \frac{s_y}{s_x} \quad \text{et} \quad a = \bar{y} - b\bar{x}$$

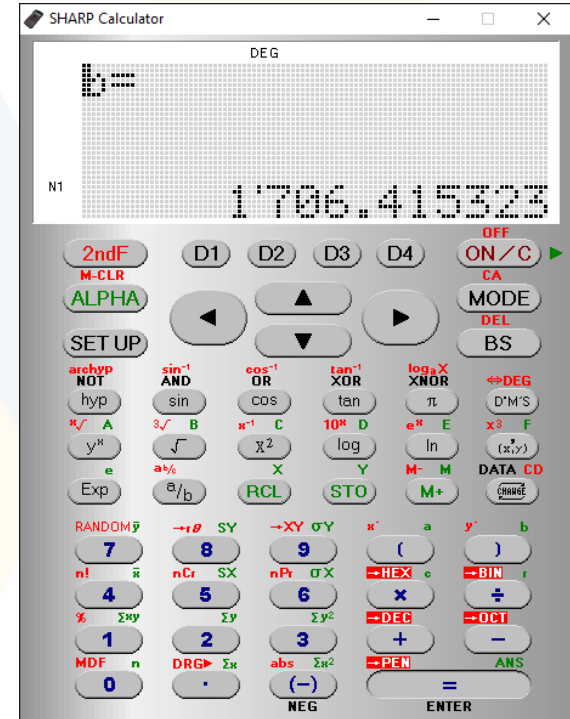


Si le lien linéaire entre les deux variables est suffisamment fort, on peut se servir de l'équation de la droite de régression comme modèle mathématique pour prévoir la valeur de la variable dépendante qui serait associée à une certaine valeur de la variable indépendante.

Cependant, il faut limiter nos prévisions à des valeurs proches de celles qui ont été obtenues lors de l'échantillonnage.

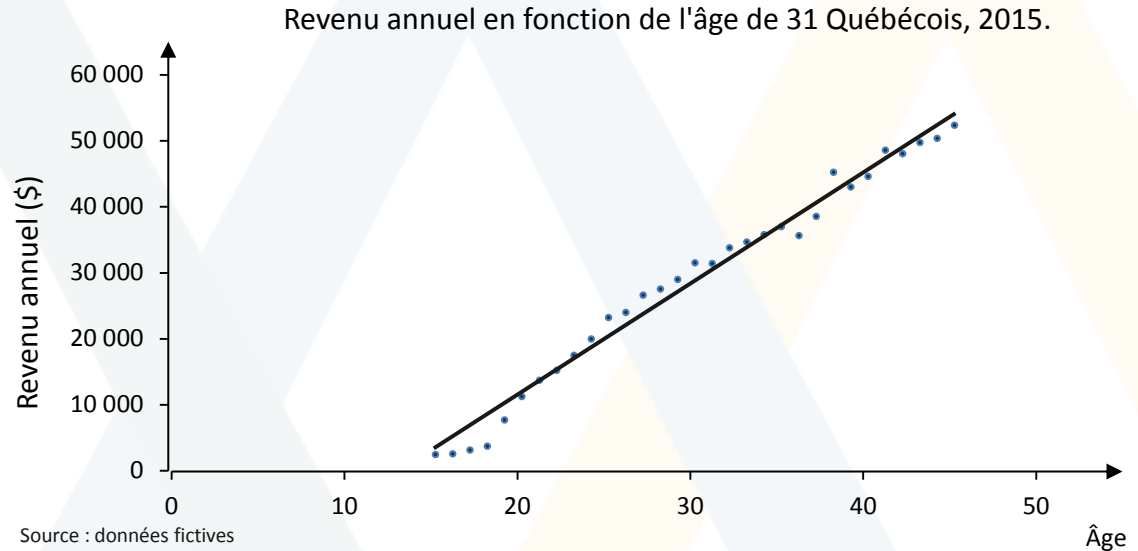
Détermination de l'équation de la droite de régression

Âge	Revenu annuel (\$)
15	2 500
16	2 600
17	3 200
18	3 750
19	7 750
20	11 250
21	13 750
22	15 250
23	17 500
24	20 000
25	23 230
26	24 000
27	26 630
28	27 560
29	29 000
30	31 500
31	31 400
32	33 780
33	34 650
34	35 780
35	37 000
36	35 600
37	38 540
38	45 200
39	43 000
40	44 600
41	48 550
42	48 000
43	49 700
44	50 340
45	52 310



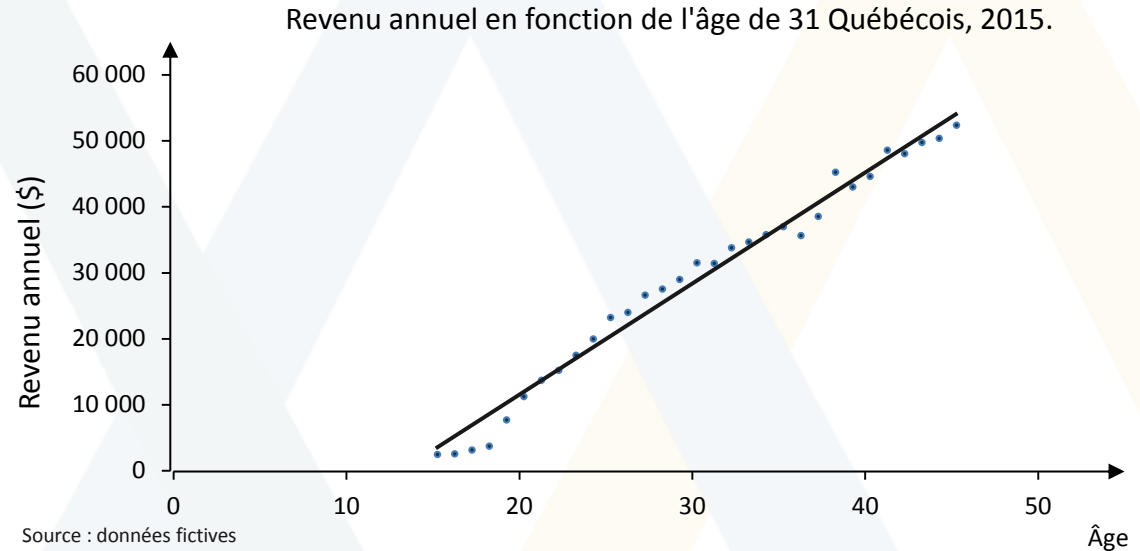
Représentation graphique de la droite de régression

Âge	Revenu annuel (\$)
15	2 500
16	2 600
17	3 200
18	3 750
19	7 750
20	11 250
21	13 750
22	15 250
23	17 500
24	20 000
25	23 230
26	24 000
27	26 630
28	27 560
29	29 000
30	31 500
31	31 400
32	33 780
33	34 650
34	35 780
35	37 000
36	35 600
37	38 540
38	45 200
39	43 000
40	44 600
41	48 550
42	48 000
43	49 700
44	50 340
45	52 310



Prévision à l'aide de l'équation de la droite de régression

Âge	Revenu annuel (\$)
15	2 500
16	2 600
17	3 200
18	3 750
19	7 750
20	11 250
21	13 750
22	15 250
23	17 500
24	20 000
25	23 230
26	24 000
27	26 630
28	27 560
29	29 000
30	31 500
31	31 400
32	33 780
33	34 650
34	35 780
35	37 000
36	35 600
37	38 540
38	45 200
39	43 000
40	44 600
41	48 550
42	48 000
43	49 700
44	50 340
45	52 310



Quel serait le revenu annuel d'un Québécois de 25 ans selon le modèle mathématique qui représente la situation ?

Résumé

Pour déterminer s'il existe un lien statistique linéaire entre deux variables quantitatives à échelle d'intervalle ou de rapport il faut :

1. déterminer les variables indépendante et dépendante;
2. représenter les valeurs à l'aide d'un nuage de points;
3. déterminer et interpréter le coefficient de corrélation linéaire r ;
4. si le lien linéaire est suffisamment fort, on peut déterminer l'équation de la droite de régression et s'en servir pour faire des prévisions.

Conception du contenu

Julie Milot

Collège de Maisonneuve
jmilot@cmaisonneuve.qc.ca

Révision du contenu

Samuel Bernard et Hélène Lambert

samuel.bernard@collanaud.qc.ca
hlambert@cmaisonneuve.qc.ca

Direction de projet

Samuel Bernard
Bruno Poellhuber

Postproduction

Marie-Ève Lanthier

Musique

Sébastien Belleudy

sebe.bandcamp.com

Calculatrice virtuelle

PC-Simulator

sharp-world.com

Conception graphique

Christine Blais

Production des modèles en LaTeX

Nicolas Beauchemin

nicolas.beauchemin@bdeb.qc.ca

Production

Samuel Bernard



Bruno Poellhuber



Vidéo mise à disposition selon les termes de la licence

Creative Commons internationale 4.0

Paternité / Pas d'utilisation commerciale / Partage dans les mêmes conditions

Les autorisations au-delà du champ de cette licence peuvent être obtenues à

Mathema-TIC.ca