

Comité d'éthique et de recherche de L'ATR

PROGRAMME DE FORMATION À LA RECHERCHE CLINIQUE DES
RÉSIDENTS DE RÉANIMATION MÉDICALE

« RÉGRESSION LINÉAIRE ET CORRÉLATION, ANALYSE
DE SURVIE »

Séance 5

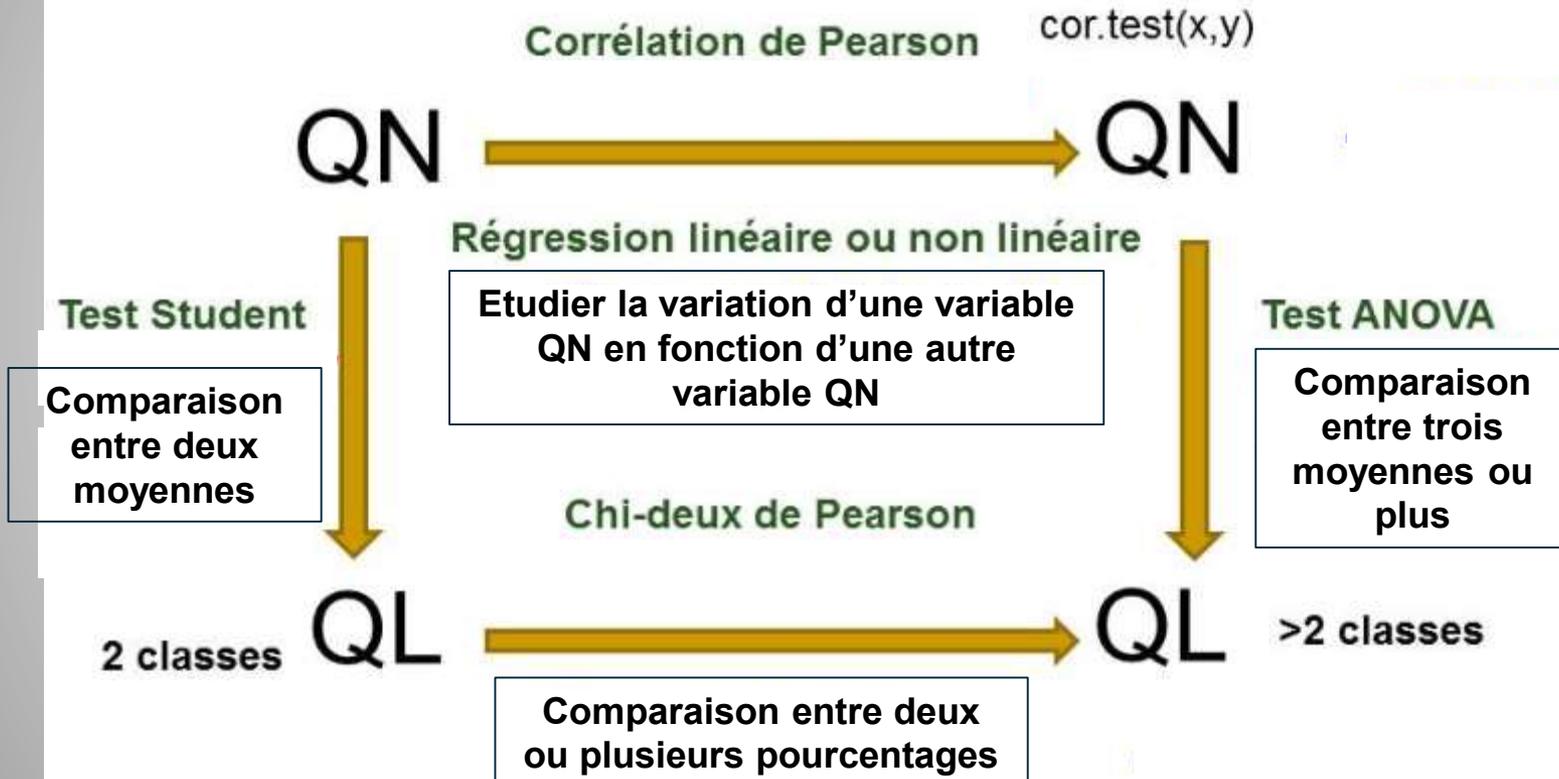
Jeudi 26 Mai 2022 à 14H

Salle des conférences. Service de Réa Med, La Rabta

Coordinateurs : Dr S Ayed, Dr A Mokline

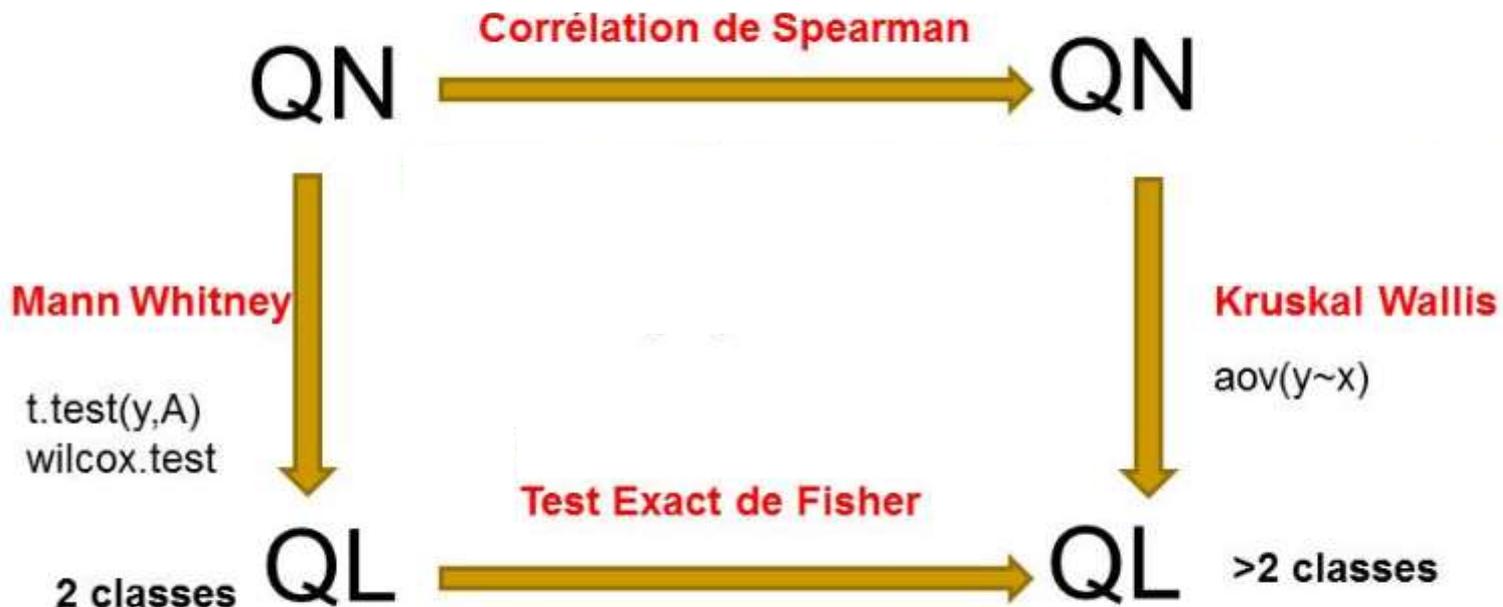
Orateur : Dr Rejaibi Salsabil

Analyses bivariées / Tests paramétriques



Analyses bivariées/ Tests non paramétriques

cor.test(x,y)



En rouge les méthodes à utiliser en cas de non normalité ou petits échantillons (n<30)

Corrélation

- Étudier la corrélation entre X et Y, revient à:
 1. Evaluer une relation statistique entre deux **variables quantitatives**;
 2. Calculer un paramètre qui mesure la force de la relation entre deux variables → le **coefficient de corrélation r**

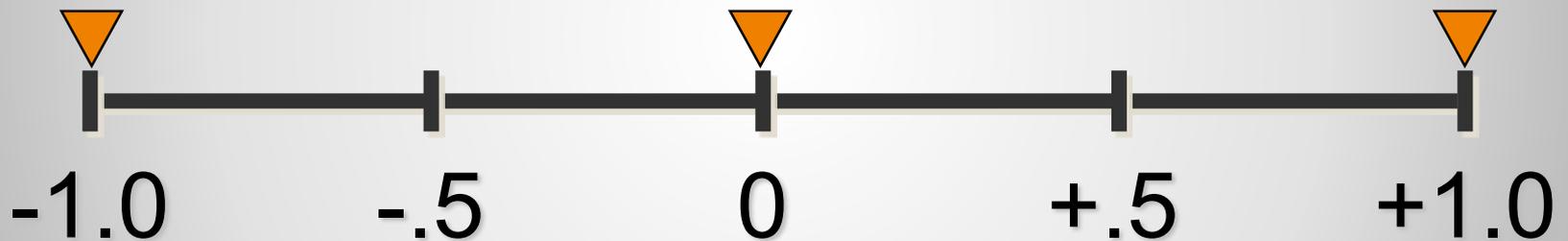
- **NB:** entre deux variables qualitatives (binaires ou à plusieurs modalités) → on ne parle pas de corrélation
 - on est en train de rechercher une association significative ou non entre deux variables QL

Valeurs du coefficient de corrélation

Correlation
négative
parfaite

Pas de Correlation

Correlation
positive
parfaite

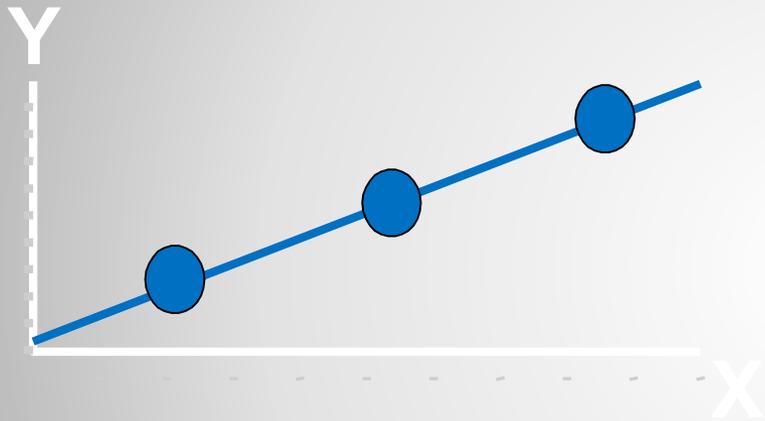


correlation négative croissante

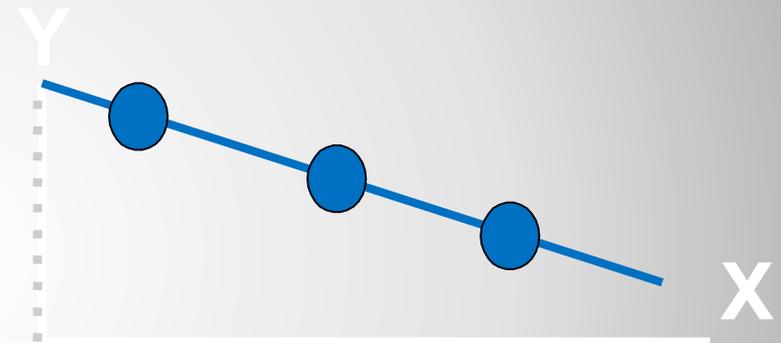
corrélation positive croissante

Coefficient de Correlation de Pearson

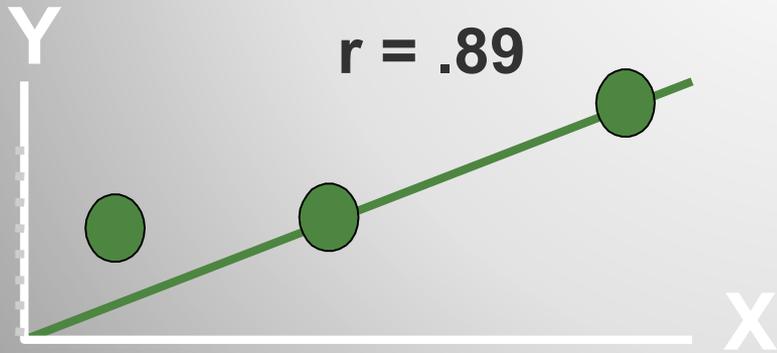
$r = 1$



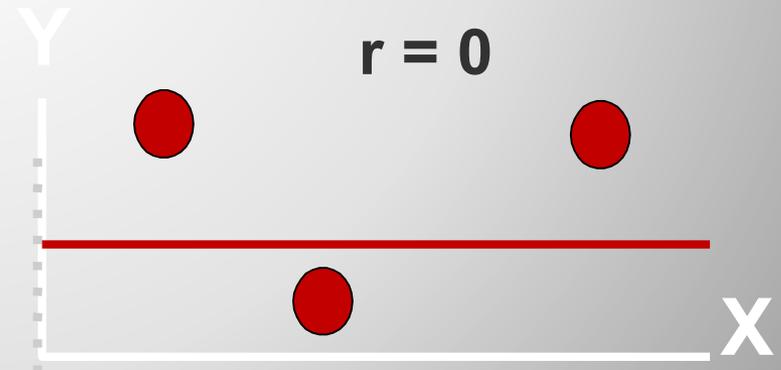
$r = -1$



$r = .89$



$r = 0$



Estimation et interprétation de ρ

- L'estimateur est noté r :

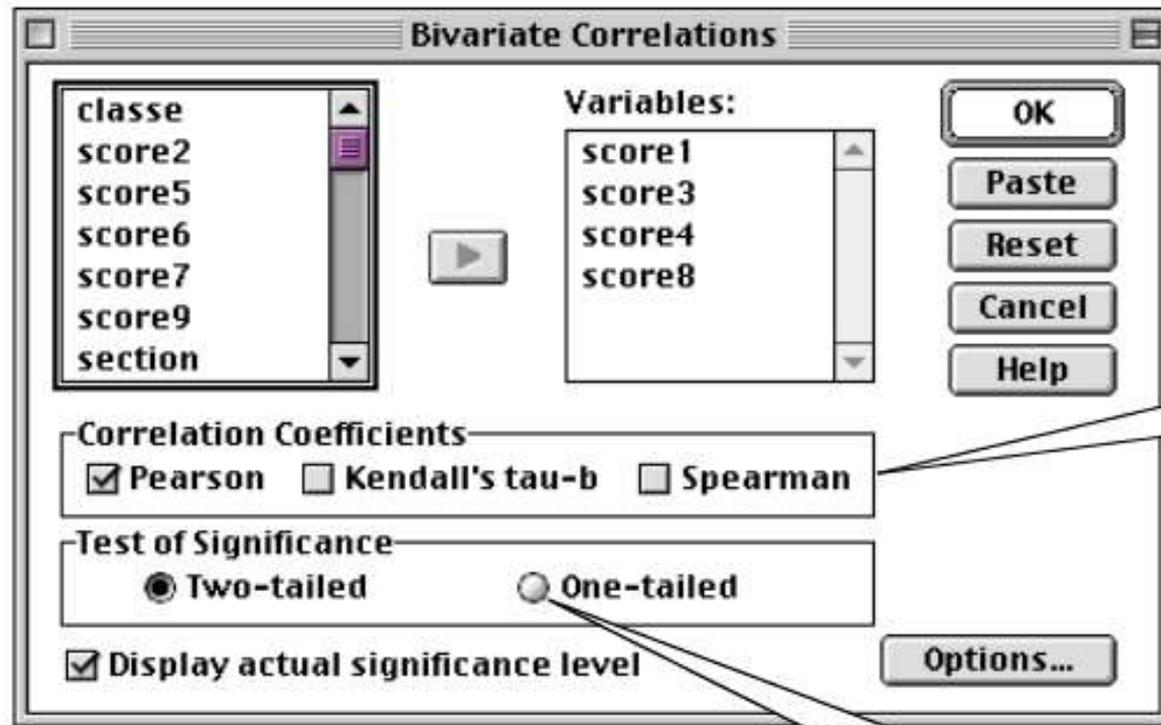
$$r = \frac{\text{cov}(x, y)}{s_x s_y} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

- C'est un nombre sans unité, qui s'interprète comme suit lorsqu'il est positif:

r	Interprétation de r
$r < 0.2$	faible
$0.2 < r < 0.5$	moyenne
$0.5 < r < 0.7$	forte
$r > 0.7$	très forte

- Les coefficients de corrélation mesurent la « force » de la relation entre les variables.
- N'ont de valeurs que s'ils sont **significatifs** donc toujours faire le test ($r \neq 0$).
- Corrélation n'est pas une preuve de relation cause-effet mais un appui.

Output SPSS(corrélation-1)



Permet de choisir le test adéquat.

A cocher si l'hypothèse est orientée.

Output SPSS(corrélation-2)

Résultats

```
-- Correlation Coefficients --
      SCORE1      SCORE3      SCORE4      SCORE8
SCORE1      1.0000      .0894      -.1288      .0841
      ( 250)      ( 250)      ( 250)      ( 250)
      P= .          P= .159      P= .042      P= .185
SCORE3      .0894      1.0000      -.2354      .1689
      ( 250)      ( 250)      ( 250)      ( 250)
      P= .159      P= .          P= .000      P= .007
SCORE4      -.1288      -.2354      1.0000      -.0797
      ( 250)      ( 250)      ( 250)      ( 250)
      P= .042      P= .000      P= .          P= .209
SCORE8      .0841      .1689      -.0797      1.0000
      ( 250)      ( 250)      ( 250)      ( 250)
      P= .185      P= .007      P= .209      P= .
```

(Coefficient / (Cases) / 2-tailed Significance)

" . " is printed if a coefficient cannot be computed

Coefficient de
corrélation r .

Nombre de sujets.

Probabilité
associée au r .

-SPSS calcule **toutes les corrélations possibles deux à deux** et les présente sous forme de tableau croisé.

-On ne s'intéresse donc qu'aux résultats au dessus (ou au dessous) de la diagonale puisque ceux-ci sont identiques.

Output SPSS (corrélation-3)

- On observe donc **trois corrélations significatives** (mises en gras):

- ❖ **Corrélation négative** entre Score1 et Score 4
- ❖ **Corrélation négative** entre Score 3 et Score 4.
- ❖ **Corrélation positive** entre Score 3 et Score 8

Résultats

--- Correlation Coefficients ---

	SCORE1	SCORE3	SCORE4	SCORE8
SCORE1	1.0000 (250) P= .	.0894 (250) P= .159	-.1288 (250) P= .042	.0841 (250) P= .185
SCORE3	.0894 (250) P= .159	1.0000 (250) P= .	-.2354 (250) P= .000	.1689 (250) P= .007
SCORE4	-.1288 (250) P= .042	-.2354 (250) P= .000	1.0000 (250) P= .	-.0797 (250) P= .209
SCORE8	.0841 (250) P= .185	.1689 (250) P= .007	-.0797 (250) P= .209	1.0000 (250) P= .

(Coefficient / (Cases) / 2-tailed Significance)

* . . is printed if a coefficient cannot be computed

Coefficient de corrélation r .

Nombre de sujets.

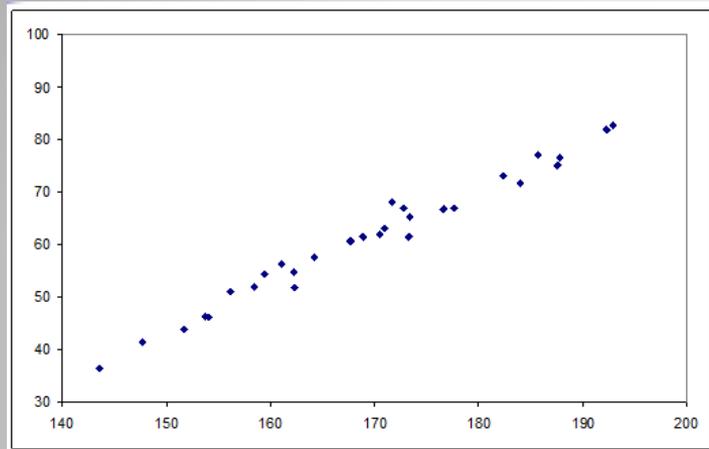
Probabilité associée au r .

Régression Linéaire / Notion de modèle en mathématiques

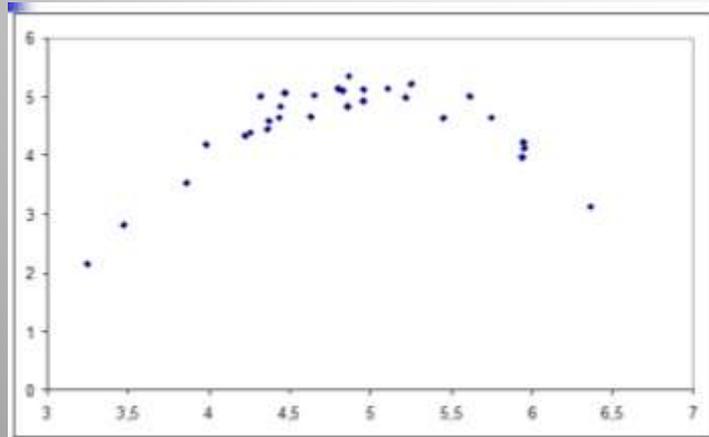
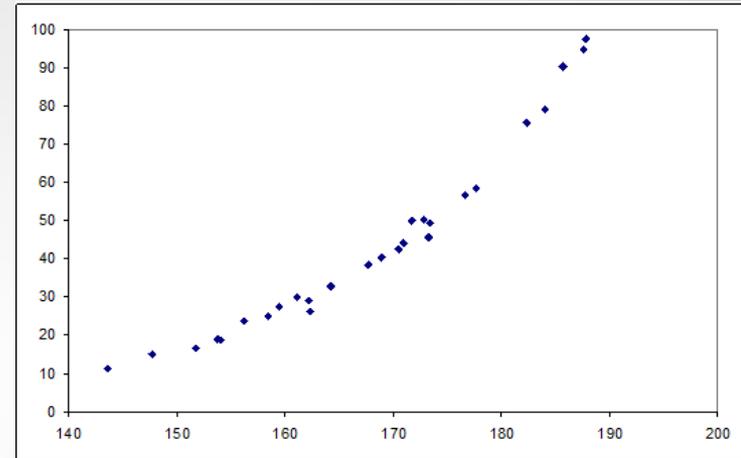
- ✓ « La régression » est un **modèle** permettant d'étudier la relation entre variables.
- ✓ Les mathématiques permettent de modéliser, c'est-à-dire de représenter, toutes sortes de situations, PUIS, l'étude mathématique ou les simulations informatiques de ces représentations nous informant – lorsque les représentations sont bonnes – sur le monde réel.

Régression Linéaire

« Linéaire »

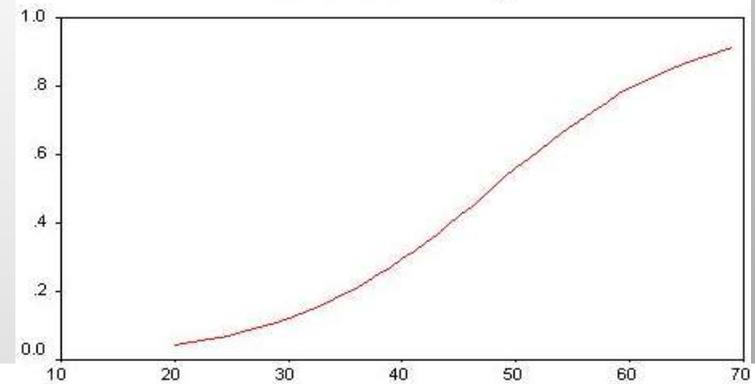


« Exponentielle »



« Parabolique »

Probabilité d'une maladie cardiaque
en fonction de l'âge



« Logistique »

Régression Linéaire

- **Le modèle** linéaire se traduit par une **droite de régression linéaire**

entre deux **variables quantitatives** : « $Y = \alpha + \beta X$ » OU:

- Y= variable dépendante à expliquer;
- X= variable indépendante explicative;
- β = pente de la droite de régression.

La variation de Y est expliquée par la variation de X;
il ne s'agit pas d'une simple corrélation

Exemple:

- Variation de la durée de survie pour une maladie donnée selon la note de gravité;
- Variation du poids de naissance du nouveau né selon le terme de la grossesse.

Application sur SPSS: corrélation

Analyse Marketing direct Graphiques Utilitaires Fenêtre Aide

Rapports
 Statistiques descriptives
 Tableaux personnalisés
 Comparer les moyennes
 Modèle linéaire général
 Modèles linéaires généralisés
 Modèles Mixtes
Corrélation
 Régression
 Log Linéaire
 Réseaux neuronaux
 Classifier
 Réduction des dimensions
 Echelle
 Tests non paramétriques
 Prévisions
 Survie
 Réponses multiples
 Analyse des valeurs manquantes
 Imputation multiple
 Echantillons complexes
 Simulation...
 Contrôle de qualité
 Courbe ROC...
 Modélisation spatio-temporelle...

Bivariée...
 Partielle...
 Distances...
 Corrélation canonique

ernorat	sexe	c
1		0
1		0
1		0
1		0
1		1
1		1
1		1
1		0
1		1
3		0
3		1
1		0
1		1
1		1
1		1
1		1
3		0
3		1
1		0

Corrélations bivariées

Variables :

- date_hospitalisat...
- fièvre_sup_39°
- céphalées
- Arthralgies
- Non_immunisé
- Statut_DCD
- Date_de_Point
- X_QN
- Missing_QNN

Taux_IgG1U/ml
Taux_IgG2U/ml

Options
Style...
Bootstrap...

Coefficients de corrélation

Pearson Tau-b de Kendall Spearman

Test de signification

Bilatéral Unilatéral

Repérer les corrélations significatives

OK Cojler Réinitialiser Annuler Aide

Corrélations

		Taux_IgG1U/ml	Taux_IgG2U/ml
Taux_IgG1U/ml	Corrélation de Pearson	1	,994**
	Sig. (bilatérale)		,000
	N	299	299
Taux_IgG2U/ml	Corrélation de Pearson	,994**	1
	Sig. (bilatérale)	,000	
	N	299	299

** La corrélation est significative au niveau 0,01 (bilatéral).

Régression linéaire : Remarques sur les données

Données : Les variables dépendantes et indépendantes doivent être quantitatives. Les variables catégorielles, comme la religion, la qualification, la zone de résidence, doivent être enregistrées sous forme de variables binaires (muettes) ou sous de tout autre type de variables de contraste.

Hypothèses : Pour chaque valeur de la variable indépendante, la distribution de la variable dépendante doit être normale. La variance de la distribution de la variable dépendante doit être constante pour toutes les valeurs de la variable indépendante. La relation entre la variable dépendante et chaque variable indépendante doit être linéaire et toutes les observations doivent être indépendantes.

Obtenir une analyse de régression linéaire

Cette fonction nécessite l'option Statistiques de base.

1. A partir des menus, sélectionnez :

Analyse > Régression > Linéaire...

2. Dans la boîte de dialogue Régression linéaire, sélectionnez une variable numérique dépendante.

3. Sélectionnez une ou plusieurs variables indépendantes.

Application sur SPSS: Régression linéaire

Régression linéaire

Dépendant :
Taux_IgG1U/ml

Bloc 1 de 1

Précédent Suivant

Variables indépendantes :
Age_calculé

Méthode : Entrez

Variable de filtrage : Règle...

Libellés d'observation :

Pondération WLS :

OK Coïler Réinitialiser Annuler Aide

Statistiques...
Tracés...
Enregistrer...
Options
Style...
Bootstrap...

- date_naissance
- vaccination_inco...
- date_hospitalisat...
- fièvregt_sup_39°
- céphalées
- Arthralgies
- code_sérum_1
- code_sérum_2
- Taux_IgG2U/ml
- Non_immunisé
- Statut_DCD
- Date_de_Point
- X_QN
- Missing_QNN
- Y_QL
- Missing_QL
- SMEAN(Missing_...
- Age_calculé

1 35,0 6

0	0	64,0	12
1		45,0	11
1	1	35,0	6
1	1	50,0	6

Régression linéaire

Dépendant :

- date_naiss...
- vaccination...
- date_hospit...
- fièregt_sup...
- céphalées
- Arthralgies
- code_sérum...
- code_sérum...
- Taux_IgG2L...
- Non_immun...
- Statut_DCD
- Date_de_Pe...
- X_QN
- Missing_QN
- Y_QL
- Missing_QL
- SMEAN(Mis...
- Age_calculé

Régression linéaire : Statistiques

Coefficients de régressi...

- Estimations
- Intervalles de confiance
- Niveau (%) :
- Matrice de covariance
- Qualité de l'ajustement
- Variation de R-deux
- Descriptives
- Mesure et corrélations partielles
- Tests de colinéarité

Résidus

- Durbin-Watson
- Diagnostic des observations
- Points atypiques : écarts types
- Toutes les observations

Poursuivre Annuler Aide

Statistiques...

Tracés...

Enregistrer...

Options

Style...

Bootstrap...

OK Coller Réinitialiser Annuler Aide

Régression linéaire

Dépendant :
Taux_IgG1U/ml

Bloc 1 de 1

Précédent Suivant

Variables indépendantes :
Age_calculé

Méthode : Entrez

Variable de filtrage : Règle...

Libellés d'observation :

Pondération WLS :

OK Coïler Réinitialiser Annuler Aide

Statistiques...
Tracés...
Enregistrer...
Options
Style...
Bootstrap...

date_naissance
vaccination_inco...
date_hospitalisat...
fièvregt_sup_39°
céphalées
Arthralgies
code_sérum_1
code_sérum_2
Taux_IgG2U/ml
Non_immunisé
Statut_DCD
Date_de_Point
X_QN
Missing_QNN
Y_QL
Missing_QL
SMEAN(Missing_...
Age_calculé

1 35,0 6

→ Régression

[Jeu_de_données] C:\Users\user\Desktop\EPU ATRéanimation\Base_virose_2015_avec_duplicates.sav

Statistiques descriptives

	Moyenne	Ecart type	N
Taux_lgG1UIml	9,12	7,381	299
Age_calculé	21,47	9,268	299

Corrélations

		Taux_lgG1UIml	Age_calculé
Corrélation de Pearson	Taux_lgG1UIml	1,000	-,015
	Age_calculé	-,015	1,000
Sig. (unilatéral)	Taux_lgG1UIml	.	,397
	Age_calculé	,397	.
N	Taux_lgG1UIml	299	299
	Age_calculé	299	299

Variables introduites/éliminées^a

Modèle	Variables introduites	Variables éliminées	Méthode
1	Age_calculé ^b	.	Introduire

a. Variable dépendante : Taux_IgG1UImI

b. Toutes les variables demandées ont été introduites.

Récapitulatif des modèles

Modèle	R	R-deux	R-deux ajusté	Erreur standard de l'estimation
1	,015 ^a	,000	-,003	7,392

a. Prédicteurs : (Constante), Age_calculé

ANOVA^a

Modèle		Somme des carrés	ddl	Carré moyen	F	Sig.
1	Régression	3,762	1	3,762	,069	,793 ^b
	Résidu	16230,659	297	54,649		
	Total	16234,421	298			

a. Variable dépendante : Taux_IgG1UImI

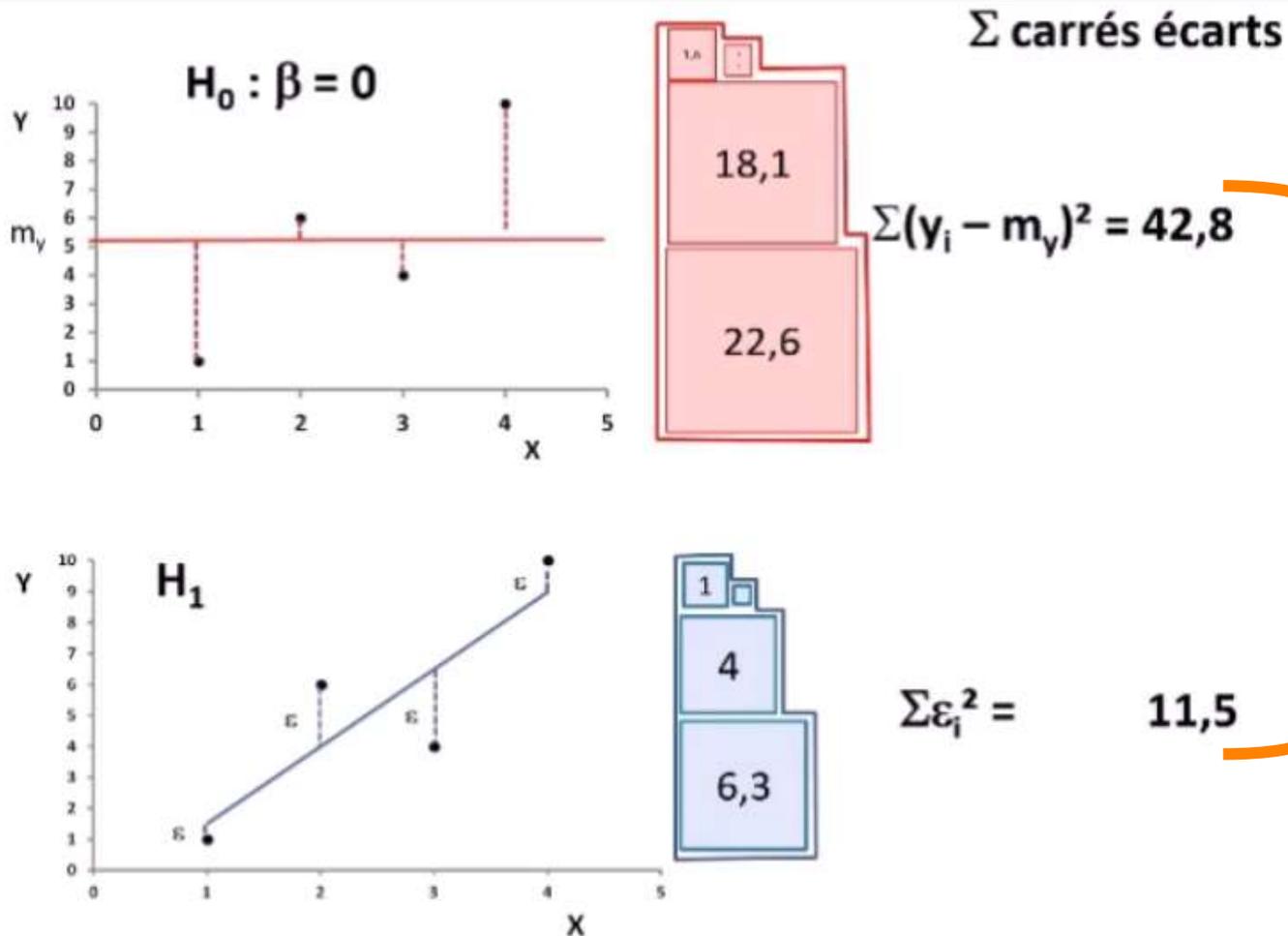
b. Prédicteurs : (Constante), Age_calculé

Coefficients^a

Modèle		Coefficients non standardisés		Coefficients standardisés	t	Sig.	Intervalle de confiance à 95,0% pour B	
		B	Erreur standard	Bêta			Borne inférieure	Borne supérieure
1	(Constante)	9,384	1,080		8,688	,000	7,258	11,510
	Age_calculé	-,012	,046	-,015	-,262	,793	-,103	,079

a. Variable dépendante : Taux_IgG1UImI

Régression Linéaire: test de la pente par l'analyse de la variance

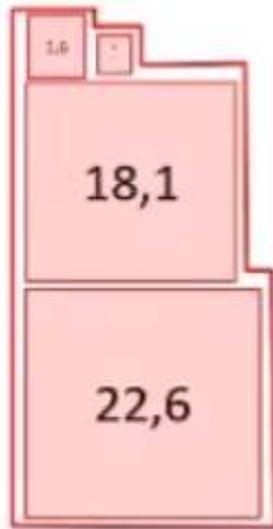


$$Y = \alpha + \beta x$$

ECART =
part de la
variabilité
due au
modèle de
régression

Régression Linéaire: test de la pente par l'analyse de la variance

Décomposition de la dispersion des Y (i)



Dispersion totale

=



Régression

+

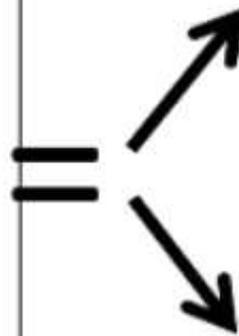
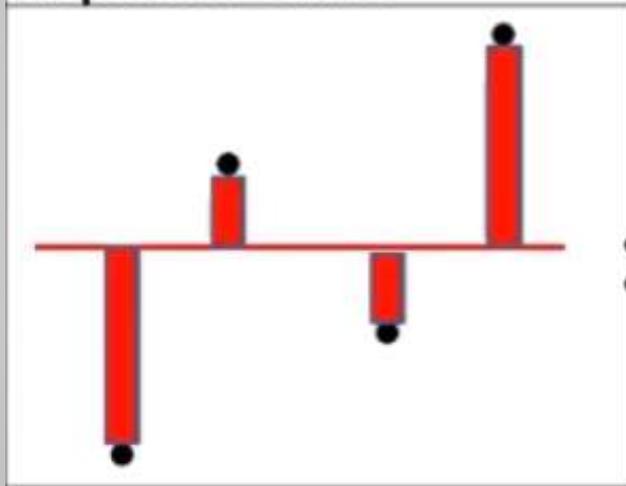


Résidus

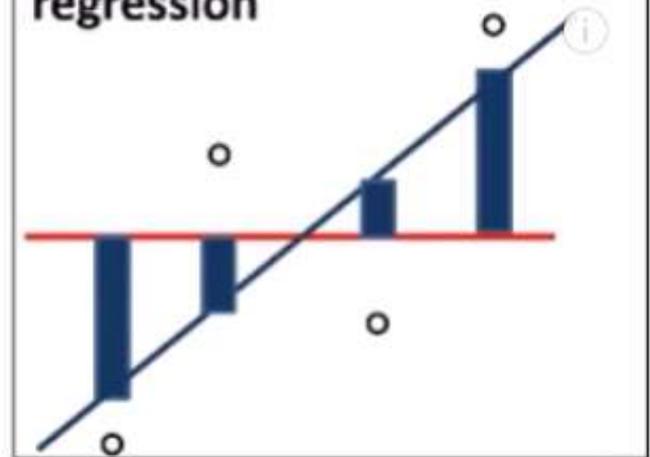
Régression Linéaire: test de la pente par l'analyse de la variance

Décomposition de la dispersion des Y

Dispersion totale

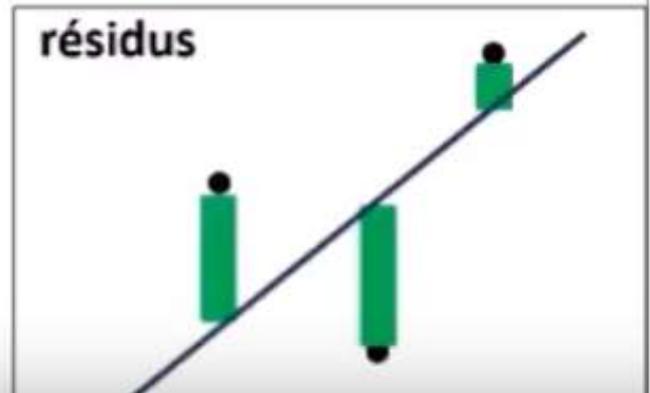


régression



+

résidus



Analyse des carrés des écarts

Dispersion	Somme des carrés des écarts	Part de la dispersion	R^2
Régression	31,3	73 %	0,73
Résiduelle	11,5	27 %	
Totale	42,8	100 %	



Part de la
variabilité
expliquée par
le modèle

Variables introduites/éliminées^a

Modèle	Variables introduites	Variables éliminées	Méthode
1	Age_calculé ^b	.	Introduire

a. Variable dépendante : Taux_IgG1UImI

b. Toutes les variables demandées ont été introduites.

Récapitulatif des modèles

Modèle	R	R-deux	R-deux ajusté	Erreur standard de l'estimation
1	,015 ^a	,000	-,003	7,392

a. Prédicteurs : (Constante), Age_calculé

ANOVA^a

Modèle		Somme des carrés	ddl	Carré moyen	F	Sig.
1	Régression	3,762	1	3,762	,069	,793 ^b
	Résidu	16230,659	297	54,649		
	Total	16234,421	298			

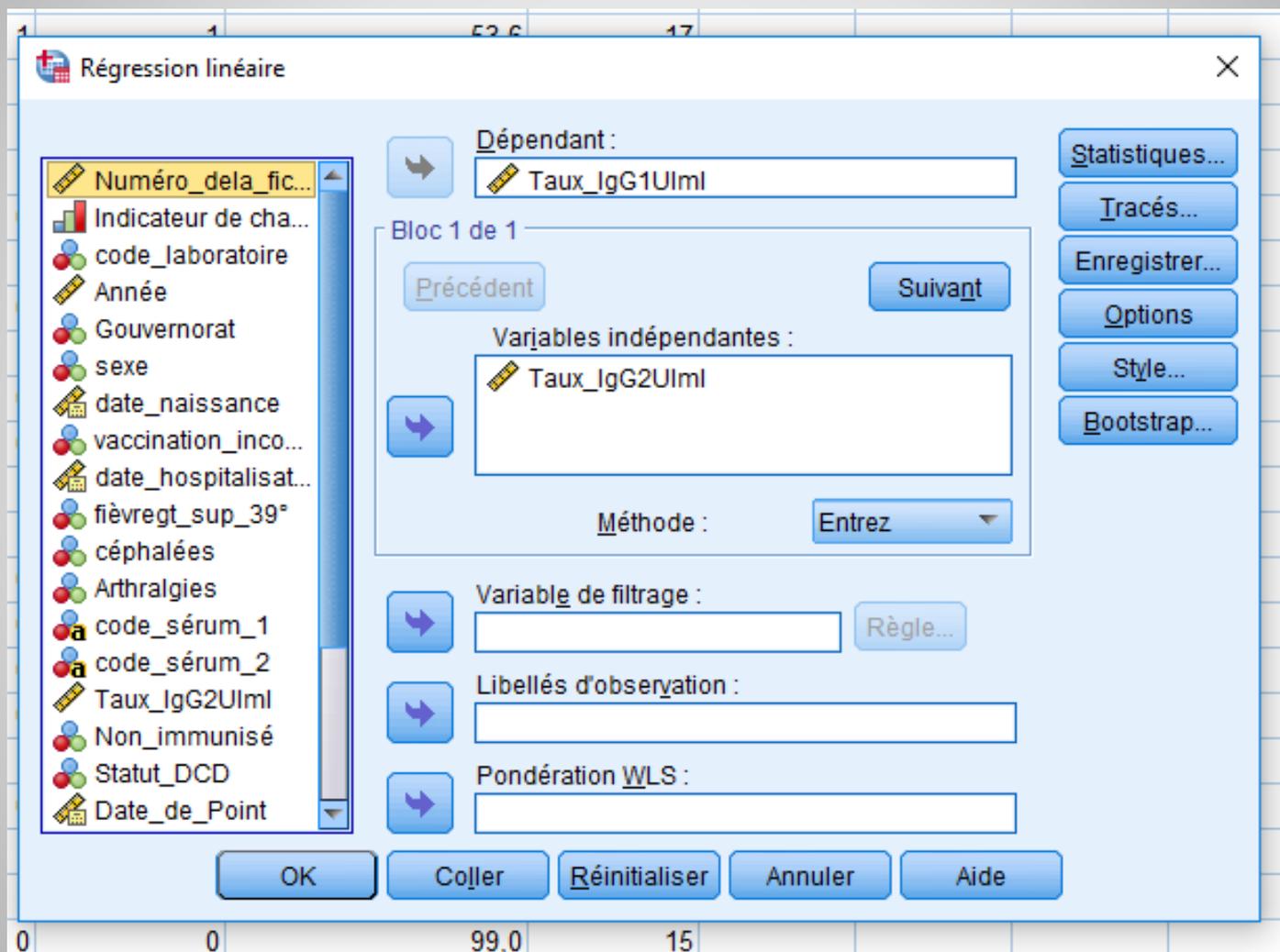
a. Variable dépendante : Taux_IgG1UImI

b. Prédicteurs : (Constante), Age_calculé

Coefficients^a

Modèle		Coefficients non standardisés		Coefficients standardisés	t	Sig.	Intervalle de confiance à 95,0% pour B	
		B	Erreur standard	Bêta			Borne inférieure	Borne supérieure
1	(Constante)	9,384	1,080		8,688	,000	7,258	11,510
	Age_calculé	-,012	,046	-,015	-,262	,793	-,103	,079

a. Variable dépendante : Taux_IgG1UImI



➔ Régression

Statistiques descriptives

	Moyenne	Ecart type	N
Taux_IgG1U/ml	9,12	7,381	299
Taux_IgG2U/ml	9,04	7,390	299

Corrélations

		Taux_IgG1U/ml	Taux_IgG2U/ml
Corrélation de Pearson	Taux_IgG1U/ml	1,000	,994
	Taux_IgG2U/ml	,994	1,000
Sig. (unilatéral)	Taux_IgG1U/ml	.	,000
	Taux_IgG2U/ml	,000	.
N	Taux_IgG1U/ml	299	299
	Taux_IgG2U/ml	299	299

Variables introduites/éliminées^a

Modèle	Variables introduites	Variables éliminées	Méthode
1	Taux_IgG2UIml ^b	.	Introduire

a. Variable dépendante : Taux_IgG1UIml

b. Toutes les variables demandées ont été introduites.

Récapitulatif des modèles

Modèle	R	R-deux	R-deux ajusté	Erreur standard de l'estimation
1	,994 ^a	,988	,988	,811

a. Prédicteurs : (Constante), Taux_IgG2UIml

ANOVA^a

Modèle		Somme des carrés	ddl	Carré moyen	F	Sig.
1	Régression	16039,190	1	16039,190	24399,926	,000 ^b
	Résidu	195,232	297	,657		
	Total	16234,421	298			

a. Variable dépendante : Taux_IgG1UIml

b. Prédicteurs : (Constante), Taux_IgG2UIml

Coefficients^a

Modèle		Coefficients non standardisés		Coefficients standardisés	t	Sig.	Intervalle de confiance à 95,0% pour B	
		B	Erreur standard	Bêta			Borne inférieure	Borne supérieure
1	(Constante)	,145	,074		1,959	,051	-,001	,291
	Taux_IgG2UIml	,993	,006	,994	156,205	,000	,980	1,005

a. Variable dépendante : Taux_IgG1UIml

Visualisation graphique des droites de régression sur SPSS

ketting direct **Graphiques** Utilitaires Fenêtre Aide

Générateur de graphiques...
 Sélecteur de modèles de représentations graphiques...
 + Tracé de Weibull...
 + Comparer les sous-groupes
 + Tracés de variables de régression

Boîtes de dialogue ancienne version

- Barres...
- Barres 3D...
- Courbes...
- Surfaces
- Circulaire...
- Plafond-plancher...
- Diagramme à surfaces
- Barre d'erreur
- Pyramide de population...
- Dispersion/Points...**
- Histogramme

Y_QL					L_1	UMCI_1
0					13,93345	14,15578
1						
1	1		53,6	17		
1	1		53,6	17		
1	1		53,6	17		
1	1		53,6	17		
0	.		77,0	17		
1	1		39,0	17		
0	.		74,0	17		
1	1		52,0	17		
0	0		62,0	17		
0	.		72,0	16		
1	1		49,0	16		
1	1		25,0	16		
0	0		85,0	16	2,00335	2,25851
0	0		69,0	16	1,00161	1,27464
0	0		68,0	16	6,99922	7,19072
0	.		75,0	16	7,99458	8,18097

.	1	1	53,6	17	5,00412	5,21458
.	1	1	53,6	17	14,91883	15,15602
77	0	.	77,0	17	6,99922	7,19072
39	1	1	39,0	17	18,85366	19,16365
74	0	.				4,22832
52	1	.				6,20198
62	0	.				24,18129
72	0	.				4,22832
49	1	.				5,21458
25	1	.				5,21458
85	0	.				2,25851
69	0	.				1,27464
68	0	.				7,19072
75	0	.	75,0	16	7,99458	8,18097
51	1	1	51,0	16	5,00412	5,21458
61	0	0	61,0	16	22,78242	23,17734

Dispersion/Points

Dispersion simple
 Dispersion de type Matrice
 Point simple

Dispersion de type Superpos
 Dispersion 3D

Nuage de points simple



- Numéro_dela_fic...
- Indicateur de cha...
- code_laboratoire
- Année
- Gouvernorat
- sexe
- date_naissance
- vaccination_inco...
- date_hospitalisat...
- fièvregt_sup_39°
- céphalées
- Arthralgies
- code_sérum_1
- code_sérum_2
- Taux_IgG2U/ml
- Non_immunisé
- Statut_DCD
- Date_de_Point
- X_QN
- Missing_QNN
- Y_QL
- Missing_QN

Axe des Y :
Taux_IgG1U/ml

Titres...

Axe des X :
Age_calculé

Options

Définir les marques par :
[Empty field]

Etiqueter les observations par :
[Empty field]

Panel par

Lignes :
[Empty field]

Variables imbriquées (pas de lignes vides)

Colonnes :
[Empty field]

Variables imbriquées (pas de colonnes vides)

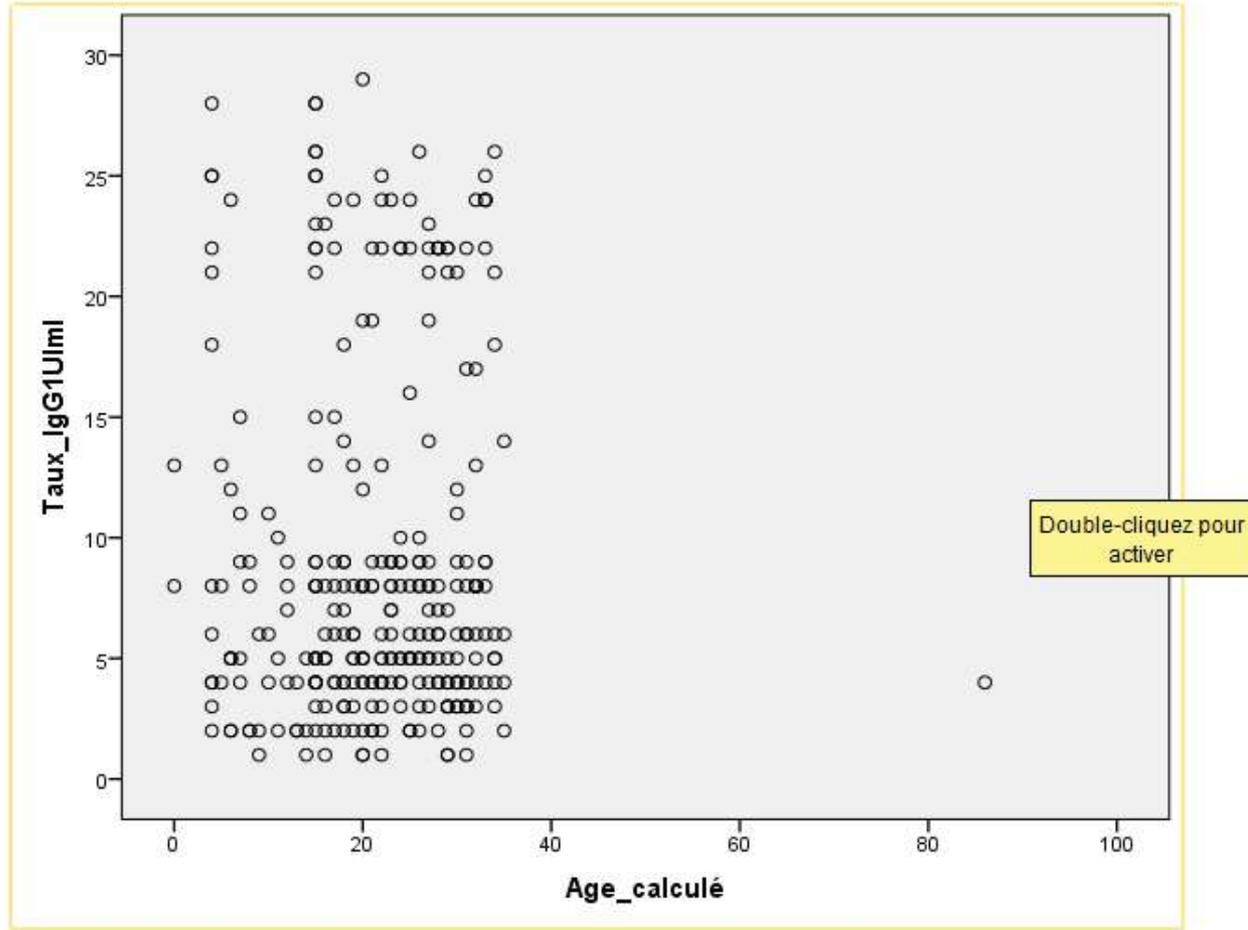
Modèle

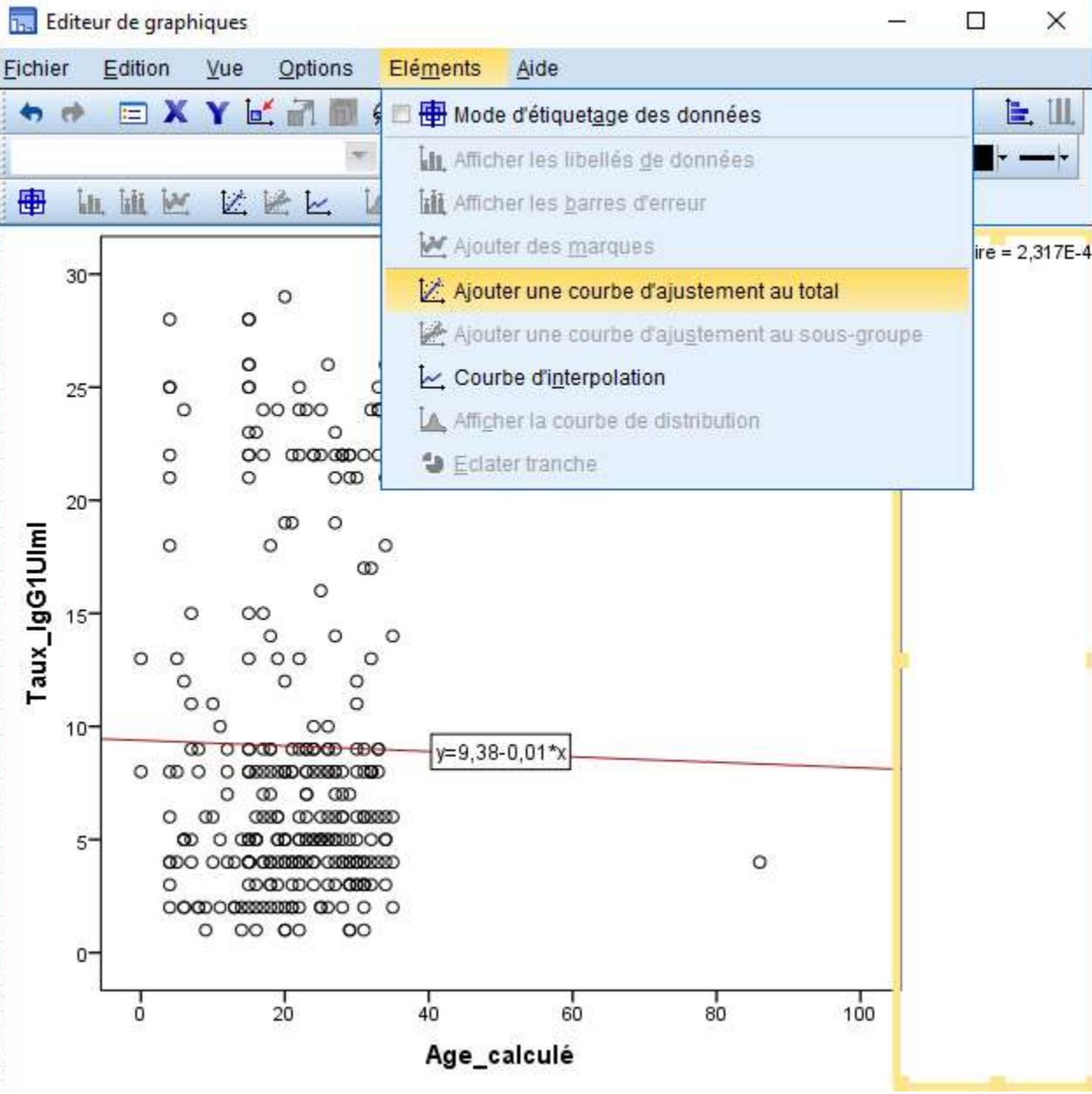
Utiliser les spécifications du graphique de :

Fichier...

- OK
- Coller
- Réinitialiser
- Annuler
- Aide

➔ **Graphique**





Régression linéaire

Dépendant :

Bloc 1 de 1

Variables indépendantes :

Méthode :

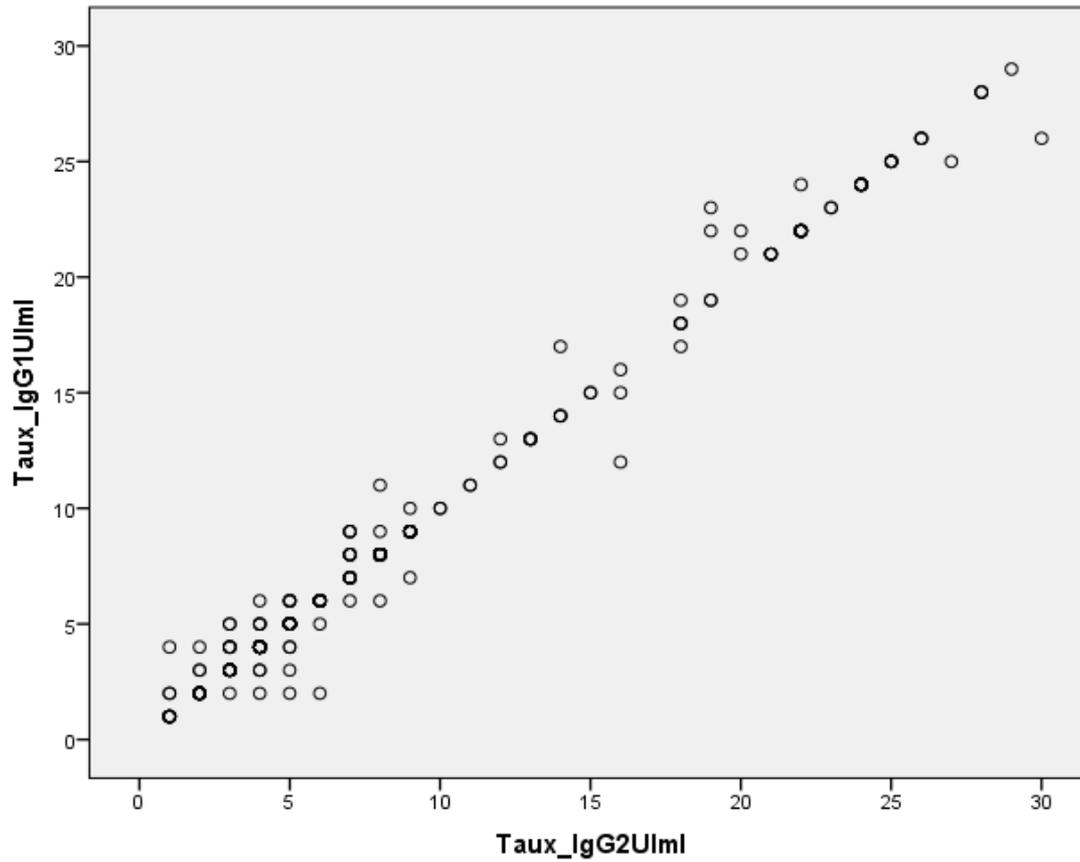
Variable de filtrage :

Libellés d'observation :

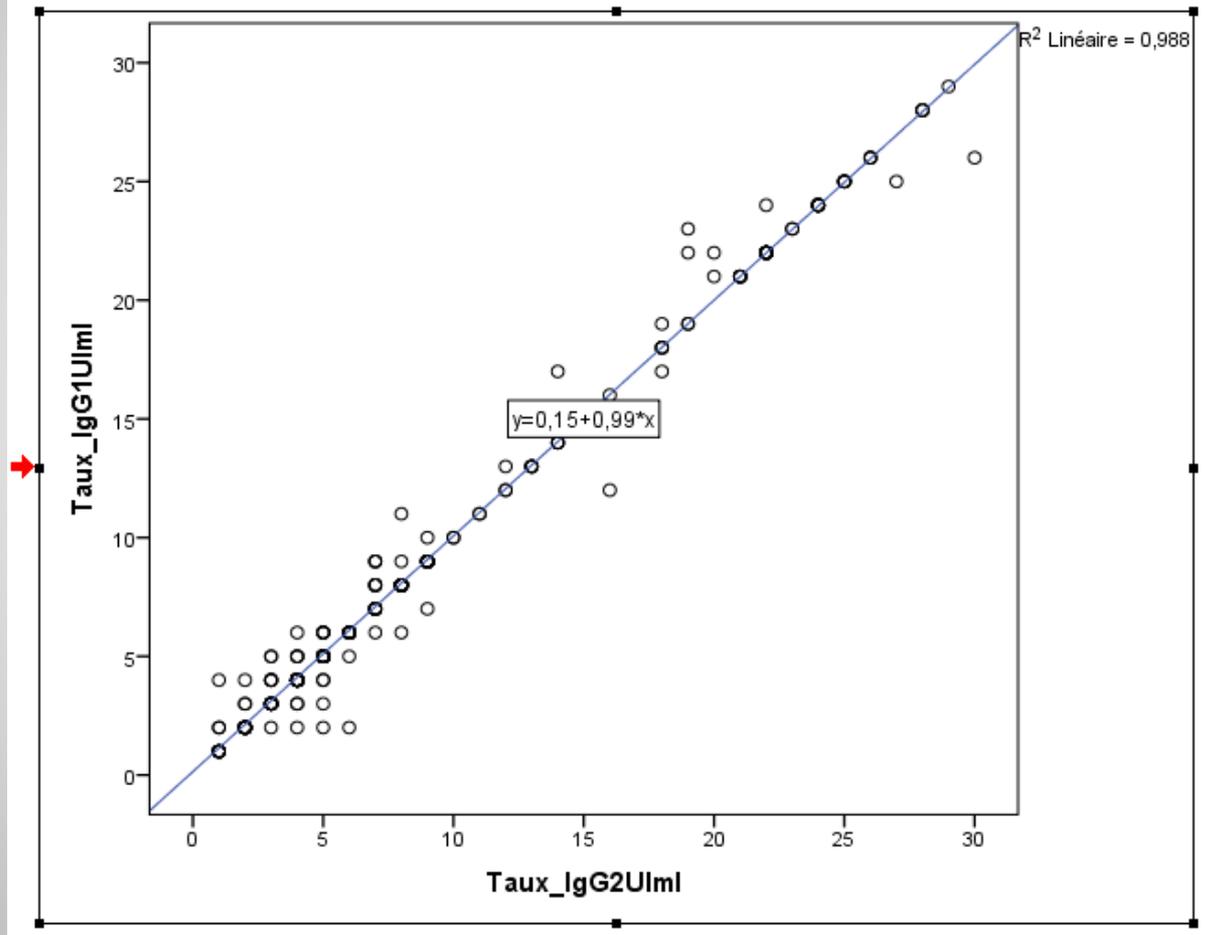
Pondération WLS :

)	.	73,0	15	7,99458	8,18097
---	---	------	----	---------	---------

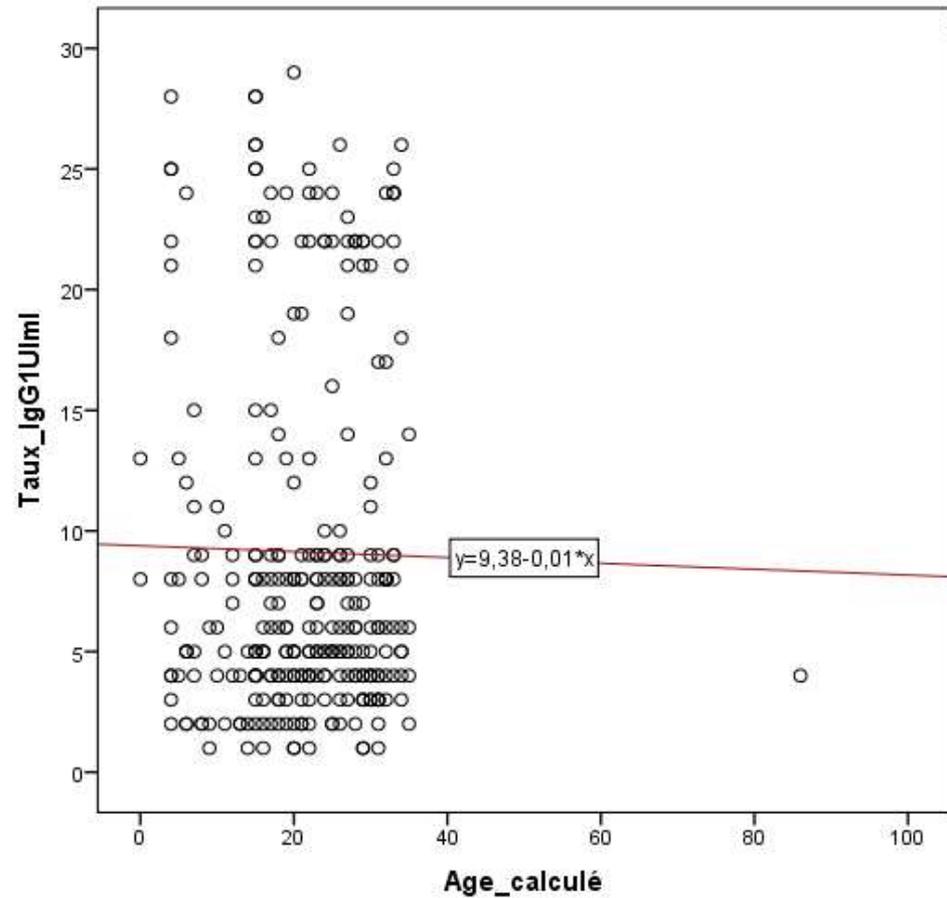
→ Graphique

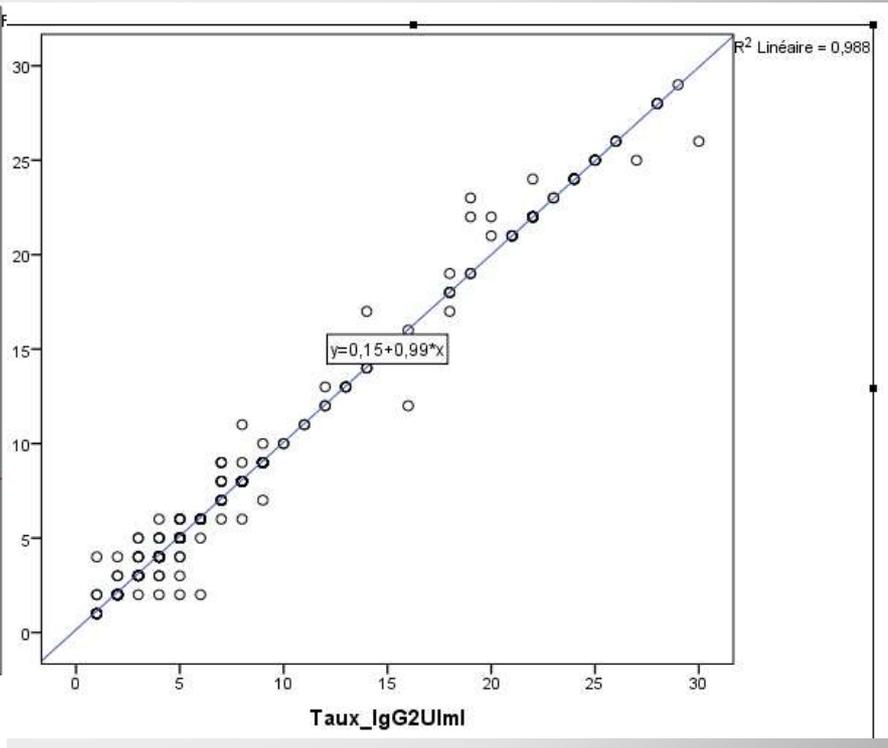
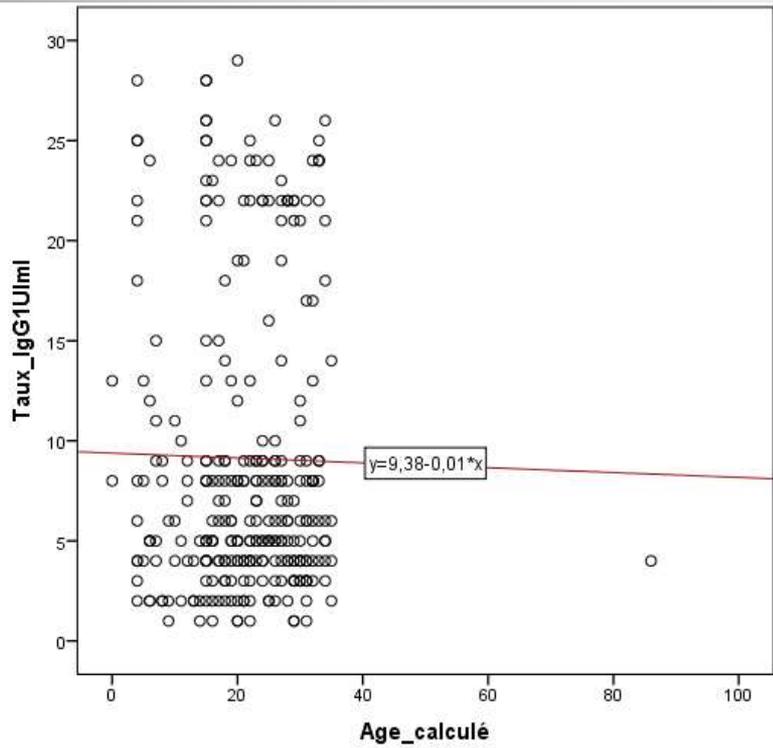


Graphique



Graphique





Merci de votre attention