

Société de Calcul Mathématique SA

Outils d'aide à la décision

depuis 1995



Méthodes mathématiques pour le traitement de données médicales imprécises

par Rima Abdenbi et Bernard Beauzamy
Société de Calcul Mathématique SA

octobre 2023

Nous avons eu récemment un contrat avec une entreprise qui met au point un vaccin destiné à lutter contre une maladie très commune. L'activité n'en est qu'à ses débuts : ce qu'on appelle "phase 1" dans les essais cliniques. Un petit nombre de patients (24 au total) ont fait l'objet d'examens et de traitements.

Le point intéressant, qui se rencontre dans la totalité des études épidémiologiques à ce stade, est que les données sont très pauvres et très imprécises. Il est évident que le panel n'est pas complet : il n'y a pas, en proportion suffisante, des jeunes et des vieux, des gros et des maigres, des malades et des bien portants, etc.

Le panel a été divisé en trois groupes : à l'un on donne une forte dose, au second une dose plus faible, et au troisième on ne donne rien (placebo) ; on observe ensuite l'évolution de certains marqueurs : dans un cas, tel pourcentage est à 10%, dans l'autre à 60%. Mais lequel est le mieux ? a priori, on n'en sait rien. Des doses trop fortes peuvent se révéler dangereuses, pour certaines personnes et dans certaines conditions : comment être sûr que le médicament n'aura pas d'effet secondaire ?

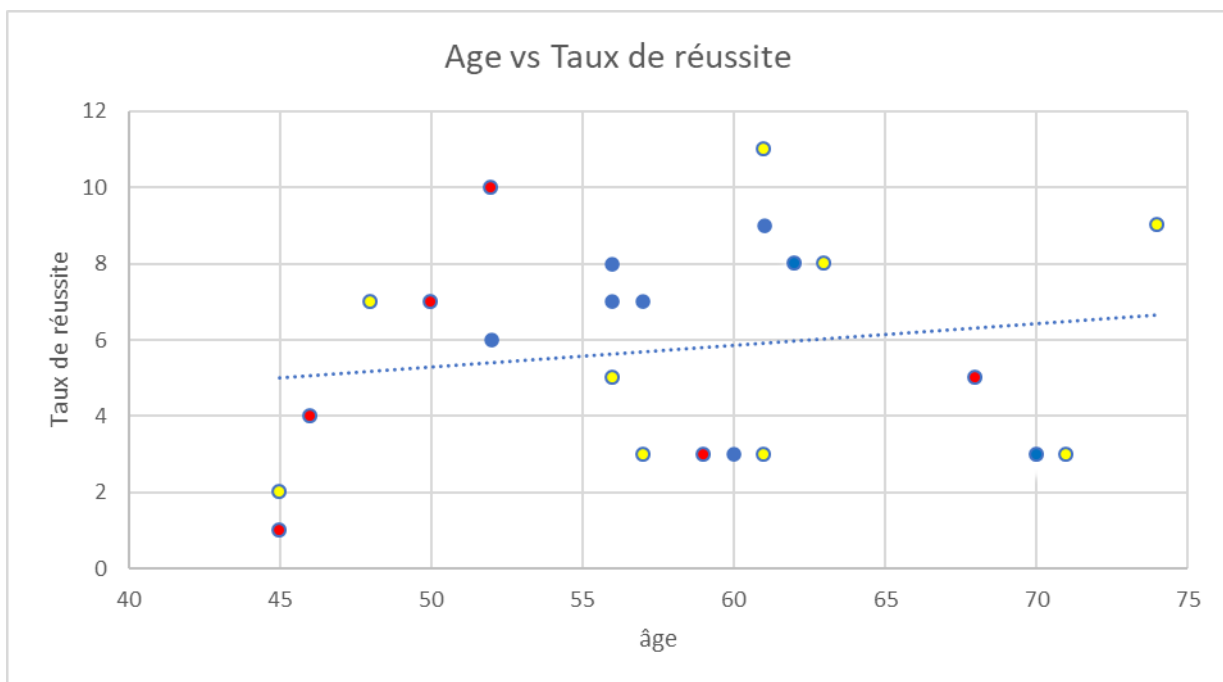
Dans ces conditions, la plupart des équipes médicales sont tentées de s'en remettre à des logiciels spécialisés : on entre les données, et le logiciel fait les calculs nécessaires ; il procède par ajustements linéaires (droite de régression) ; cela permet d'avoir une conclusion à partir de peu de données : deux données suffisent pour générer une droite.

Cette façon de traiter les données, au moyen d'une boîte noire, nous paraît structurellement dangereuse, parce qu'on a le sentiment de disposer, en sortie, de connaissances que l'on n'avait pas en entrée : le logiciel ajoute une connaissance factice, que l'on a tôt fait d'accepter pour vraie. La Nature, surtout dans le domaine médical, s'accommode mal des prolongements linéaires.

La présentation que nous faisons ici prend en compte cette difficulté : elle est beaucoup plus modeste, étant purement qualitative, et non quantitative. Pour chaque paramètre et chaque patient, nous notons 1 si l'on constate que le paramètre a évolué favorablement chez ce patient et 0 dans le cas contraire (y compris si l'on n'en sait rien). Le tableau des résultats est donc complètement binaire : 0 ou 1. Il y a deux colonnes qui restent quantitatives : l'âge et l'indice de masse corporelle (BMI) dont on peut supposer qu'il est correctement mesuré.

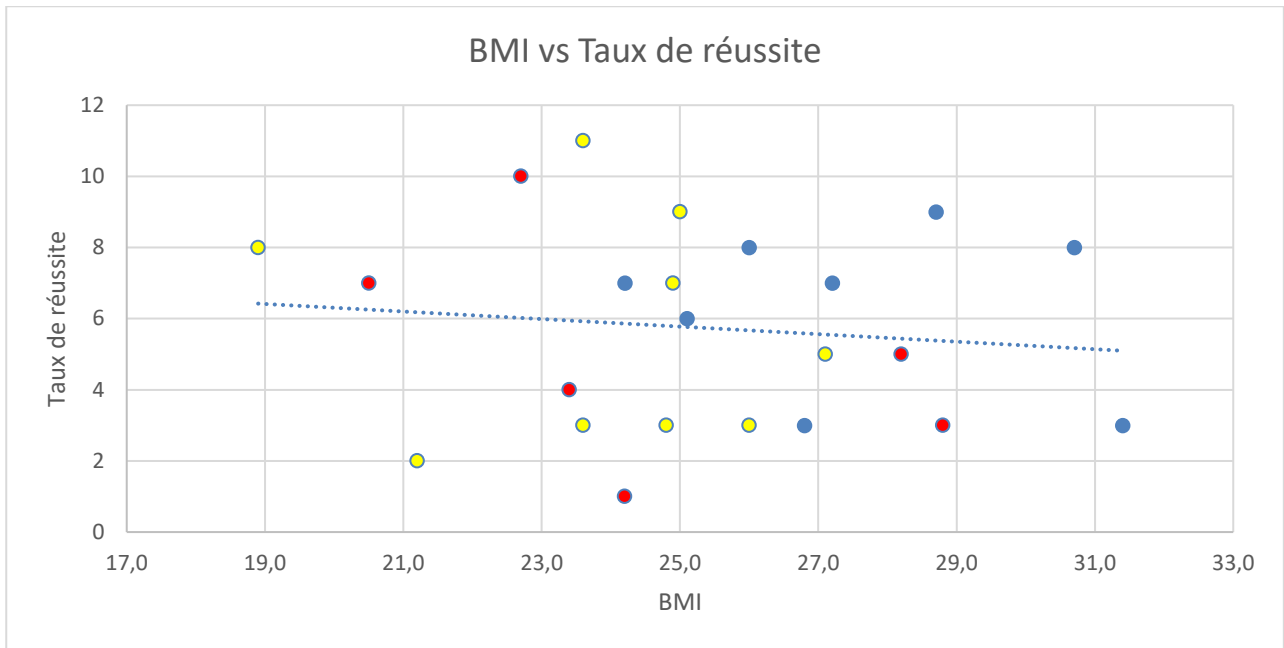
Le nombre de "1" dans la ligne de chaque patient peut s'appeler "taux de réussite du traitement" : chaque "1" indique qu'un test a été passé avec succès. Pour nous, à ce stade, les tests sont indifférenciés : il se peut que certains soient plus importants que d'autres, mais nous ne le savons pas.

Nous réalisons deux nuages de points : le premier est âge, en abscisse, versus taux de réussite, en ordonnée. Les points rouges sont ceux qui ont reçu le placebo ; les fortes doses sont en jaune et les faibles doses en bleu.



Quel que soit le groupe, il n'y a pas de relation claire entre l'âge et le taux de réussite : si on fixe un taux de réussite à 6, on constate que ce taux se retrouve indépendamment de l'âge. La courbe de régression (en pointillés) ne donne aucune indication significative.

Le second nuage de points représente le BMI (en abscisse) versus le taux de réussite (en ordonnée).



Là encore, il n'y a pas de dépendance claire et la droite de régression n'est pas significative.

Cette approche par nuage de points est très utile : elle montre bien le danger des approches par "régression", où l'on essaie à toute force de montrer un lien entre deux quantités.

Ensuite, notre traitement statistique est simple et purement factuel : pour chacun des trois groupes, combien de 0 et combien de 1 ?

Voici une représentation binaire des résultats des tests : 1 si l'on estime que le test a été passé avec succès (en vert), 0 sinon.

DOSES	Participants	caractéristiques			Conditions												Taux de réussite
		SEXE	AGE	BMI	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	
Faible dose	P1	M	60-69	25-30	1	1	1	1	1	0	1	1	0	1	1	0	9
	P2	M	70-79	25-30	1	0	1	1	0	0	0	0	0	0	0	0	3
	P4	M	60-69	30-35	1	1	1	0	0	0	1	1	1	1	0	1	8
	P5	F	50-59	25-30	0	0	1	0	1	1	1	1	1	1	0	0	7
	P7	M	60-69	30-35	0	0	1	0	0	0	1	0	0	0	0	1	3
	P8	F	50-59	18,5-25	0	0	1	0	1	1	0	1	1	1	1	1	7
	P10	M	50-59	25-30	0	0	1	0	1	1	1	1	1	1	0	1	8
	P11	M	50-59	25-30	0	1	1	1	0	0	1	0	0	0	1	1	6
Forte dose	P13	M	50-59	25-30	0	0	1	0	1	0	0	1	0	0	0	0	3
	P14	F	70-79	18,5-25	0	0	1	0	0	0	1	0	0	0	1	0	3
	P15	F	60-69	18,5-25	1	0	1	0	1	1	1	1	1	1	0	0	8
	P18	M	60-69	18,5-25	1	1	1	1	1	1	1	1	1	1	1	0	11
	P19	F	50-59	25-30	0	0	1	1	1	0	0	1	0	0	1	0	5
	P20	M	40-49	18,5-25	0	1	1	0	1	1	0	1	1	1	1	0	7
	P21	M	40-49	18,5-25	0	0	1	1	0	0	0	0	0	0	0	0	2
	P23	M	70-79	25-30	0	0	1	1	1	1	1	1	1	1	1	0	9
Placebo	P3	M	60-69	25-30	0	0	0	0	1	1	0	1	1	1	0	0	5
	P6	F	40-49	18,5-25	0	0	1	1	1	0	0	0	0	0	0	1	4
	P9	F	50-59	25-30	0	1	1	1	0	0	0	0	0	0	0	0	3
	P16	M	40-49	18,5-25	0	0	1	0	0	0	0	0	0	0	0	0	1
	P17	M	50-59	18,5-25	0	0	1	1	0	1	1	1	0	0	1	1	7
	P22	F	50-59	18,5-25	0	1	1	1	1	1	1	1	1	1	1	0	10

On peut faire la moyenne des taux de réussite selon les groupes :

	faible dose	forte dose	placebo
	9	3	5
	3	3	4
	8	8	3
	7	11	1
	3	5	7
	7	7	10
	8	2	
	6	9	
		3	
somme	51	51	30
moyenne	6,38	5,67	5,00

On constate que c'est le groupe "faible dose" qui obtient les meilleurs résultats en moyenne.

On peut s'intéresser aux situations "taux de réussite ≥ 6 " :

DOSES	Participants	caractéristiques			Conditions												Taux de réussite
		SEXE	AGE	BMI	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	
Faible dose	P1	M	60-69	25-30	1	1	1	1	1	0	1	1	0	1	1	0	9
	P4	M	60-69	30-35	1	1	1	0	0	0	1	1	1	0	1	1	8
	P5	F	50-59	25-30	0	0	1	0	1	1	1	1	1	1	0	0	7
	P8	F	50-59	18,5-25	0	0	1	0	1	1	0	1	1	1	1	0	7
	P10	M	50-59	25-30	0	0	1	0	1	1	1	1	1	1	0	1	8
Forte dose	P11	M	50-59	25-30	0	1	1	1	0	0	1	0	0	0	1	1	6
	P15	F	60-69	18,5-25	1	0	1	0	1	1	1	1	1	1	0	0	8
	P18	M	60-69	18,5-25	1	1	1	1	1	1	1	1	1	1	1	0	11
	P20	M	40-49	18,5-25	0	1	1	0	1	1	0	1	1	1	1	0	7
Placebo	P23	M	70-79	25-30	0	0	1	1	1	1	1	1	1	1	1	0	9
	P17	M	50-59	18,5-25	0	0	1	1	0	1	1	1	0	0	1	1	7
	P22	F	50-59	18,5-25	0	1	1	1	1	1	1	1	1	1	1	0	10

Là encore, le groupe "faible dose" l'emporte, puisque 6 individus sur 8 ont un score d'au moins 6, tandis que c'est seulement 4 sur 9 dans le groupe "forte dose" et 2 sur 6 dans le groupe placebo.

On constate que, à partir de données très incomplètes, on peut obtenir des conclusions assez précises, et cela sans faire d'hypothèse de modèle.