

Projet de Statistique

Phase de statistiques descriptives

Au cours de ce projet nous allons étudier les éventuels problèmes de foie de différents patients en analysant leurs taux de différentes enzymes sanguines et leurs diagnostics de maladie. Notre étude porte sur un échantillon de 100 patients composé de 24 femmes et 76 hommes parmi lesquels 70 sont sains et 30 présentent des problèmes de foie. On peut donc se demander comment les patients ont été choisis pour l'étude et pourquoi une majorité d'hommes ont été sélectionnés.

Quelques données générales sur notre échantillon :

Ce tableau nous montre que chez les patients malades, tous les taux d'enzymes sont plus faibles que chez les patients sains. De plus, on remarque des valeurs extrêmes maximum pour les enzymes des patients sains. On peut alors se demander si seules les valeurs très basses impactent le statut du foie ?

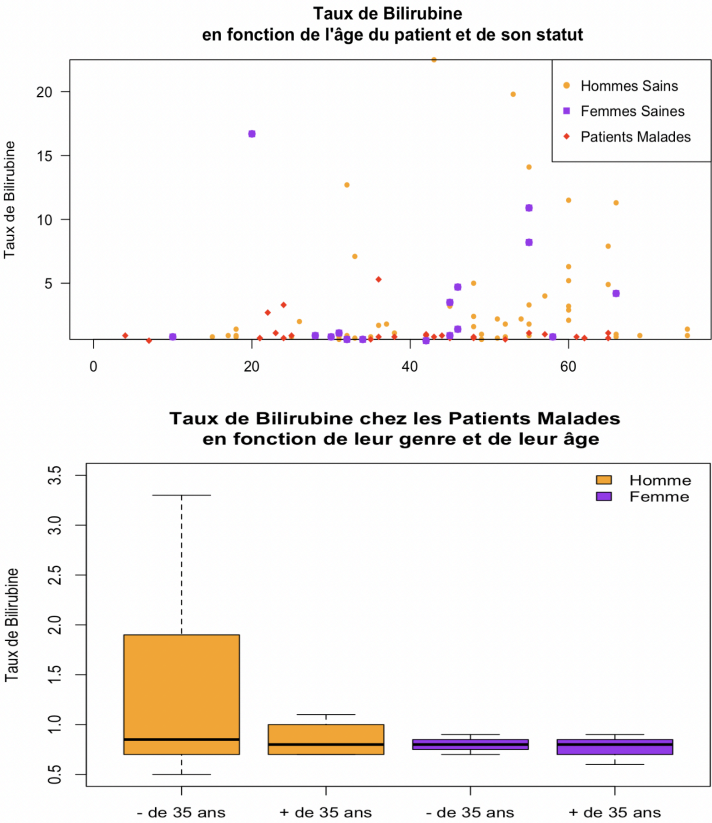
	Patients Sains Patients Malades					
	Âge	Bilirubine	Alkphos	Sgpt	Sgot	Protéines
Minimum	10 4	0.5 0.5	97 135	15 10	15 10	3.6 5.1
1 ^{er} Quart	32 24.25	0.8 0.7	190 162.8	26.5 22	33 22	5.925 6.05
Médiane	48 40	1.5 0.8	215 182	46.5 28	58 29.5	6.55 6.9
Moyenne	45.41 38.9	4.074 1.103	330.5 194.1	90.14 29	169.9 38.87	6.464 6.817
3 ^{ème} Quart	56.50 51	4.575 0.975	313 197.8	68.25 32	106.2 45.5	7.1 7.675
Maximum	75 65	27.7 5.3	1350 352	1250 82	4929 127	8.5 8.5

Notre moyenne d'âge chez les patients malades est de 38 ans, cependant après étude de notre échantillon et afin d'avoir des résultats plus visibles, nous séparerons la population entre les -35 ans et +35 ans.

Afin de déterminer l'origine des maladies du foie, nous allons étudier chaque caractéristique de notre échantillon, résumer les informations importantes et soulever les points notables.

Étude du taux de Bilirubine dans notre échantillon.

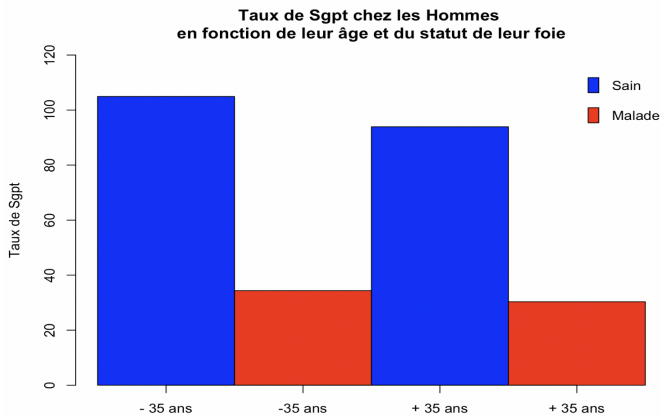
Nous remarquons que la moyenne de notre échantillon est environ 4 fois plus élevée chez les patients sains que chez les patients malades, et environ 2 fois plus pour la médiane. Nous allons donc étudier le taux de bilirubine en fonction de différents facteurs.



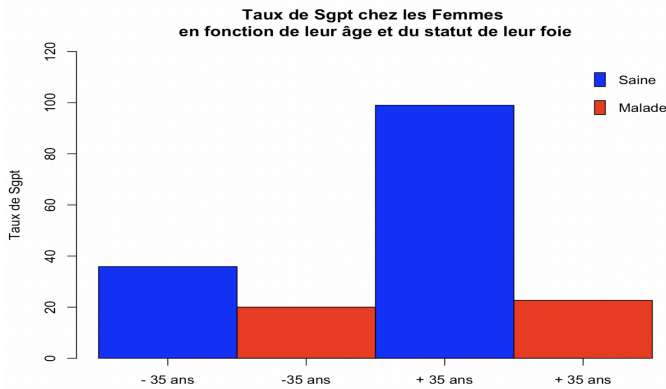
Lorsque l'on observe les taux de Bilirubine de nos patients, nous remarquons que ce taux augmente fortement avec l'âge chez les patients sains mais stagne chez les patients malades. On remarque chez nos patients sains que le genre a une influence négligeable sur les taux de bilirubine. Nous étudions donc les patients malades sous différentes conditions. Malgré quelques différences non significatives, nous voyons que la médiane ne varie pas ou presque. Ce qui signifie que la majorité des malades ont un taux de bilirubine aux alentours de 0.8. Le genre et l'âge ne semblent pas avoir d'influence sur le taux de bilirubine chez les malades. Nous observons sur notre nuage de points que plusieurs patients sains de tout âge ont des taux de bilirubine semblables à ceux des malades. De plus, des taux très élevés de bilirubine n'entraînent pas de maladie, on peut alors se demander si la baisse de la bilirubine seule peut entraîner la maladie ?

Étude du taux de Sgpt / Alanine aminotransférase dans notre échantillon

Comme on voit dans le tableau du début, la médiane des patients malades est inférieure de 40% à celle des patients sains, nous allons donc chercher quelles sont les causes de cette diminution.

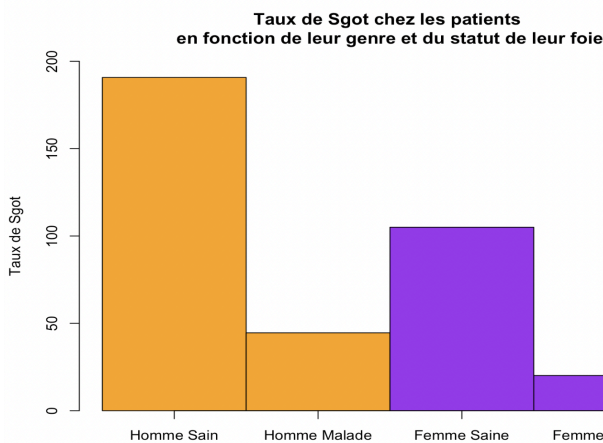


Pour la Sgpt nous avons ainsi analysé le taux chez les hommes sains et malades en fonction de leur âge. Nous pouvons voir que l'âge entraîne une légère diminution chez les hommes sains comme chez les malades. Ainsi nous pensons que chez les hommes, l'âge n'influe pas dans les variations de Sgpt.

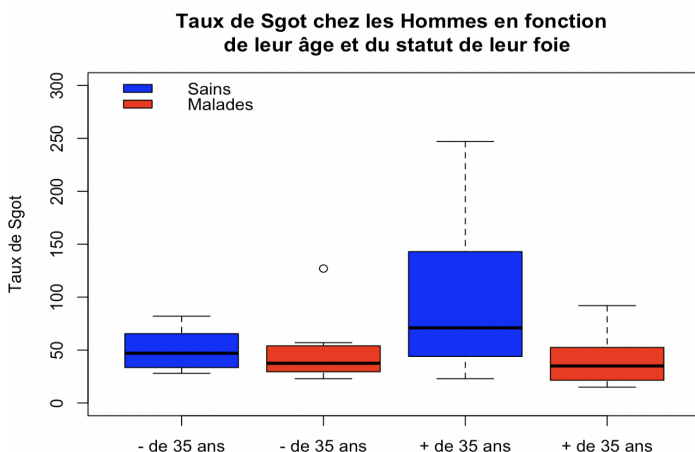


Lorsque nous faisons cette même analyse chez les femmes, nous obtenons tout l'inverse. Nous pouvons voir que chez les femmes saines le taux de Sgpt augmente fortement (60%), tandis que chez les femmes malades nous ne voyons pas de différence liée à l'âge. Ainsi comme pour la bilirubine, la maladie maintient le taux de Sgpt à des valeurs très basses. Nous pouvons alors nous demander si la baisse de Sgpt peut entraîner la maladie et s'il y a bien une différence liée au genre du patient ?

Étude du taux de Sgot / Aspartate aminotransférase dans notre échantillon



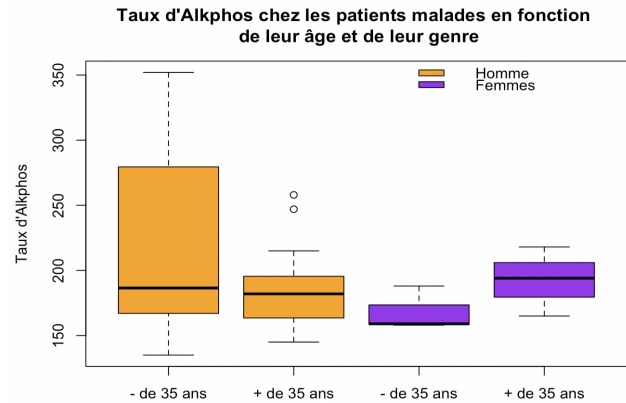
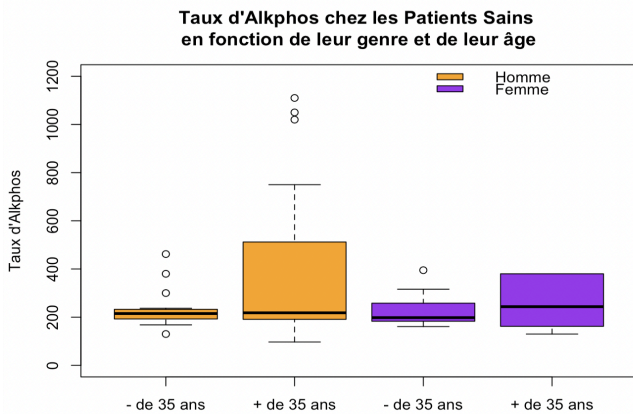
Pour la Sgot, nous pouvons voir que la maladie diminue fortement les taux enzymatiques, à peu près 70%. Ayant obtenu le même ratio pour les 2 genres entre sains et malades, nous représentons uniquement le taux de Sgot chez les hommes en fonction de leur âge et du statut de leur foie. Ce choix est appuyé par le fait qu'une majorité d'hommes a été choisie pour l'étude.



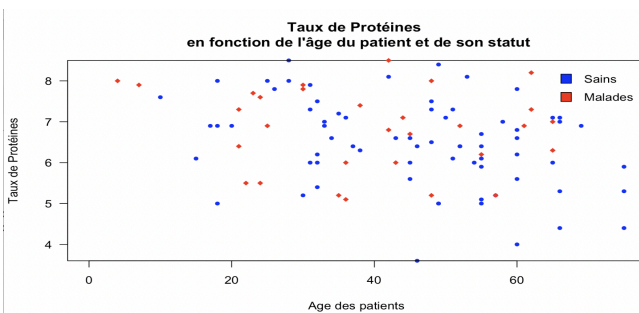
Chez les hommes sains, l'âge entraîne augmentation du taux de Sgot. Chez les hommes malades cependant, aucun changement n'est notable avec l'avancée de l'âge. Ainsi, la maladie maintient le taux de Sgot très bas. Comme pour les autres enzymes, la maladie semble maintenir le taux au plus bas quelque soit l'âge. La Sgpt joue-t-elle un rôle dans la maladie et si oui, dans quelles conditions ?

Étude de l'Alkphos / Phosphatase alcaline dans notre échantillon

On voit que nos médianes ne varient pas beaucoup (180~200). De plus, on observe chez les patients sains que certains hommes peuvent présenter des taux très forts, ce qui n'est jamais le cas chez les malades. Nous pouvons donc nous demander si le taux d'Alkphos est lié à la maladie ?



Étude du taux de protéines dans notre échantillon



Pour les protéines, comme le montre ce nuage de point et le tableau en 1ère page, nous ne voyons pas de grosses différences entre les patients sains et malades quelles que soient les conditions. Il s'agit peut-être d'un témoin.

En conclusion de cette partie d'analyse descriptive nous avons vu que le genre et l'âge n'ont pas l'air d'être impactants sur la contraction de la maladie dans la majorité des cas. De plus, nous remarquons qu'une valeur très basse dans une enzyme n'entraîne pas nécessairement la maladie, nous nous demanderons alors si une seule ou plusieurs enzymes sont responsables de cette maladie et si plusieurs, lesquelles et dans quelles conditions ?

Phase d'analyse d'une question biologique

Introduction

L'objectif de cette partie sera d'étudier si la diminution du taux d'une unique enzyme peut provoquer la maladie ou si la baisse de plusieurs d'entre elles simultanément en est la cause.

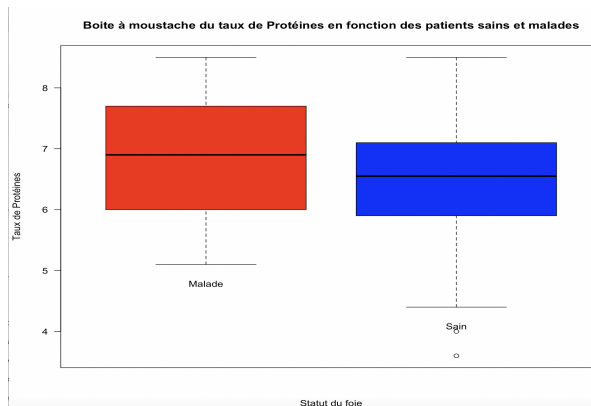
Nous déterminerons ensuite quels facteurs (âge, genre) favorisent la contraction de la maladie.

Nous avons observé qu'un patient présente des valeurs aberrantes (Sgot et Sgpt), les résultats sont donc faussés pour l'échantillon. Nous allons donc retirer cette personne de l'échantillon afin de garder le même effectif dans les différents tests.

Conditions d'applications

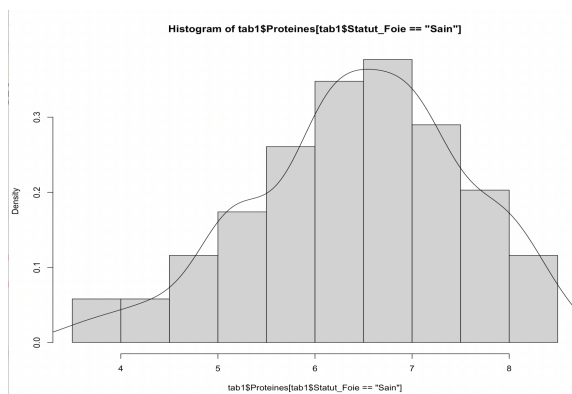
Afin de déterminer quels tests statistiques sont adaptés à notre échantillon, nous commençons par calculer les variances de chaque enzyme chez les patients sains et malades. Nous réalisons un test de Fisher dont les hypothèses sont :

- H_0 , les variances de nos enzymes chez les patients sains et malades sont égales
- H_1 , les variances de ces taux ne sont pas égales.



D'après les p-valeurs obtenues, toutes sont inférieures à α sauf pour les protéines, nous acceptons H_0 au risque β pour les protéines mais nous rejetons cette hypothèse au risque α pour les 4 enzymes étudiées, on peut dans ce cas parler d'hétéroscédasticité.

Nous étudions ensuite la normalité de nos échantillons enzymatiques sains et malades à l'aide d'histogrammes et de graphiques quantiles-quantiles et nous pouvons conclure que seules les protéines suivent une loi normale.



Au vu de ces résultats pour les variances et la distribution de nos données, il ne sera pas possible de faire des tests paramétriques sur l'intégralité de notre échantillon, pour ces cas là nous nous limiterons donc à des tests non paramétriques.

Comparaison de nos échantillons sains et malades

Nous allons tout d'abord vérifier les hypothèses faites durant la phase d'analyse. Le test de Wilcoxon Mann Whitney est un test non paramétrique qui permet de comparer deux moyennes dans nos échantillons contenant des données appariées.

Ainsi, nous posons nos 2 hypothèses :

- H_0 : le taux d'enzyme entre les sains et les malades est le même
- H_1 : le taux d'enzyme entre sains et malades à des différences significatives

Nous confirmons les observations faites en phase de statistiques descriptives, en effet pour la Bilirubine, l'Alkphos, la Sgpt et la Sgot nous obtenons une p-valeur $< \alpha$ ainsi nous acceptons H_1 au risque α .

Pour les protéines nous réalisons un test de Student car les conditions sont respectées. Nous gardons les mêmes hypothèses et nous observons que seuls les niveaux de protéines ne varient pas entre les patients sains et malades.

Comparaison de nos échantillons en fonction de l'âge

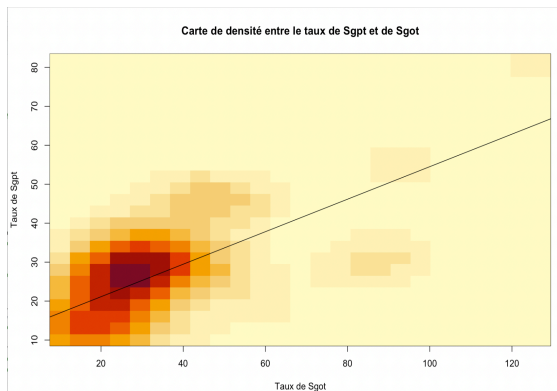
Cependant, dans notre échantillon nous remarquons que certains patients présentent un taux très faible dans une enzyme et ne sont pourtant pas malades, nous en déduisons qu'il existe des facteurs impactant ces observations. Nous allons donc séparer notre échantillon en fonction de l'âge des patients comme durant la phase d'analyses descriptives. Nous recommandons les mêmes tests (variance et normalité) que précédemment en vérifiant les conditions d'applications de nos tests pour chaque groupe.

Nous commençons par des tests de comparaison de moyennes, avec le test de Wilcoxon Mann Whitney (car variances différentes et la normalité n'est pas respectée) pour la plupart des enzymes. Cependant pour

l'Alkphos chez -35 ans, nous obtenons des variances égales entre les patients sains et les patients malades, nous effectuons alors un test de Welch en précisant 'var=TRUE' pour ce cas là car la normalité n'est pas respectée. On pose nos hypothèses :

- H_0 : le taux d'enzyme entre les sains et les malades est le même
- H_1 : le taux d'enzyme entre sains et malades à des différences significatives

D'après nos résultats, chez les -35 ans, nous n'observons aucune différence significative dans les taux d'enzymes entre les patients sains et malades, contrairement aux +35 ans chez qui, chacune des 4 enzymes se retrouve en quantité très différente chez les sains et les malades. Nous cherchons maintenant à comprendre pourquoi certains patients -35 ans sont malades tout en présentant des taux qui n'engendrent pas forcément la maladie. Pour cela nous étudions les relations de corrélation entre les enzymes.



Nos échantillons ne suivent pas une loi normale et les variances ne sont pas égales, l'étude de la corrélation se fera donc grâce au test de Spearman, test non paramétrique. D'après nos coefficients de corrélation obtenus chez les patients malades en comparaison avec les sains, il y a de forts liens entre Sgot/Sgpt, Bilirubine/Alkphos et Bilirubine/Sgpt. Ces relations pourraient expliquer pourquoi certains patients sont malades même en présentant un taux faible dans une enzyme. En effet avec ces corrélations, si une enzyme est faible, il y a de fortes chances qu'une autre enzyme qui lui est corrélée soit faible également.

Afin de mettre ces enzymes en relation chez les patients de -35 ans nous réalisons un autre test non paramétrique, le test de Kruskal Wallis. D'après nos résultats : il y a de grosses différences entre les taux chez les malades et les sains pour la Sgpt, la Sgot et la Bilirubine. Nous pouvons donc supposer que des valeurs basses de ces 3 enzymes simultanément sont nécessaires pour qu'un patient de -35 ans soit malade.

Comparaison de nos échantillons en fonction du genre

Maintenant que nous savons que la maladie a plus de chances de se déclarer chez les patients plus âgés, nous allons chercher pourquoi une majorité d'hommes a été choisie pour cette étude. Les variances et la normalité de nos échantillons ne permettent, une nouvelle fois, de ne faire que des tests non paramétriques (sauf pour la Sgpt et l'Alkphos chez les femmes +35 ans). En effectuant les tests adaptés nous observons que, chez les +35 ans, seule la Sgot chez les femmes varie fortement entre les saines et malades, alors que chez les hommes, toutes les enzymes varient énormément. Nous pouvons en déduire que, si une majorité d'hommes a été choisie pour l'étude, c'est parce qu'ils sont bien plus touchés par la maladie que les femmes.

Conclusion

Nous avons étudié un échantillon des taux enzymatiques de 100 patients de tous âges. Cela nous a mené à nous poser la question 'Quelle(s) enzyme(s) pourrai(en)t causer la maladie et dans quelles circonstances ?' Nous avons alors réalisé différents tests afin de répondre à cette question.

Cette étude a démontré que la maladie apparaît sous des conditions différentes en fonction de l'âge, du genre et des taux d'enzymes. Elle a pour effet de bloquer tous les niveaux enzymatiques à leur plus bas, et la baisse de chacune des enzymes peut entraîner la maladie dans certaines conditions. Nous avons pu observer au cours de cette étude que les hommes âgés sont le plus susceptibles d'être touchés par des problèmes de foie. De plus, la Sgot, la Sgpt et la Bilirubine semblent liées entre elles, ce qui augmente donc la probabilité d'attraper la maladie. Certaines variations observées dans nos valeurs peuvent être expliquées par une carence en vitamines B6 (alcoolisme, médicaments...) ou une grossesse.