

Etude in vitro de la sensibilité bactérienne.

Présenté par: KHADIDJATOU THIAM

Master2 S.T.A.F.A.V UFR S.A.T

Université Gaston Berger

STAGE effectué au:

Laboratoire bactériologie-virologie fondamentale et appliquée
de l'U.C.A.D II

22 octobre 2008

Dédicaces

Je dédie ce travail à :

- mes parents pour l'éducation, l'amour et le soutien qu'ils m'ont donné durant tous ces années.

- a ma sœur Khayate et son mari Médoune Ba pour leur soutien et aide durant ces quatre mois de stage.

- tous mes frères et sœurs.

- tous mes camarades du DEA de Statistique Appliquée aux Vivants.

- tous mes amis de près ou de loin qui m'ont soutenu et assisté durant ces années d'étude.

Remerciements

Je souhaite adresser tous mes remerciements aux personnes qui ont apporté leur aide et qui ont ainsi contribué à l'élaboration de ce mémoire.

Tout d'abord le professeur Cheikh Saad-Bouh Boye doyen de la faculté de médecine, pharmacie et d'odonto-stomatologie, Monsieur Ngom professeur à la faculté de science et technologie de l'université Cheikh Anta Diop de Dakar, pour l'aide et le temps qu'ils ont bien voulu me consacrer et toutes les personnes du laboratoire de Bactériologie Virologie fondamentale et appliquée.

Enfin, j'adresse mes plus sincères remerciements à ma famille qui m'a toujours soutenu et encouragée au cours de la réalisation de ce mémoire.

Table des matières

Dédicaces	1
Remerciements	2
Résumé	5
Introduction	6
I Généralité sur les bactéries et le test de sensibilité	7
1 Les bactéries et les antibiotiques	8
1.1 Définition et exemple de bactéries	8
1.2 Définition et classification des antibiotiques	9
2 Méthodes d'étude de la sensibilité bactérienne	10
2.1 Introduction	10
2.2 Principe des méthodes de dilution	10
2.3 Principe des méthodes de diffusion	11
2.4 Principe de la méthode des E-test	11
II Modèle statistique d'étude de la sensibilité Bactérienne	13
3 Modèle statistique	14
3.1 Introduction	14
3.2 Objectif de l'étude et démarche de modélisation	14
3.3 Les données	15
3.4 Description mathématique des données	16
3.5 Ajustement d'un modèle d'étude	19
3.5.1 Introduction	19
3.5.2 La covariance	20
3.5.3 Le coefficient de corrélation	20
3.5.4 Le modèle de régression simple	21
3.5.5 Estimation des paramètres	22
3.5.6 Propriétés des estimateurs	23
3.5.7 Propriétés des résidus	24

3.5.8	Test de signification des paramètres estimés	24
3.5.9	La régression non paramétrique	26
3.5.10	Méthode des "courbes de concordance"	27
3.5.11	Validation du modèle	28
3.5.12	La prévision	28
III	Application aux données	29
4	Profil de sensibilité du Streptococcus Pyogenes	31
4.1	Etude de la liaison entre la CMI et le diamètre d'inhibition	31
4.2	Choix du modèle de régression linéaire simple	34
4.2.1	Application du modèle sur le test avec l'Erythromycine	34
4.2.2	Application du modèle sur le test avec le Teicoplanine	38
4.2.3	Application du modèle sur le test avec le Levofloxacin	40
4.3	Echec de l'application du modèle sur la Péniciline G et le Chloramphénicol	42
4.4	Choix du modèle des courbes de concordance	43
4.4.1	Application du modèle sur le test avec la Péniciline G	43
4.4.2	Application du modèle sur le test avec le Chloramphénicol	45
5	Profil de sensibilité du Streptococcus pneumoniae	47
5.1	Liaison entre la CMI et le diamètre d'inhibition	47
5.2	Modèle de régression simple pour le test avec la Péniciline G, l'Erythro- mycine et le Chloramphénicol	48
5.3	Modèle des courbes de concordance sur le test avec le Teicoplanine	49
6	Profil de sensibilité de l'Haemophilus Influenzae	51
6.1	Liaison entre la CMI et le diamètre d'inhibition	51
6.2	Modèle de régression simple pour le test avec l'Ampicilline et le Chloram- phénicol	51
6.3	Modèle des courbes de concordance pour le test avec le Lévofoxacin	53
	Conclusion	54
	Annexes	55
	Bibliographie	58

Résumé

Les méthodes les plus connues pour la détermination de la sensibilité bactérienne sont des méthodes expérimentales. Les tests permettent de récupérer des variables quantitatives, comme le diamètre d'inhibition et la concentration minimale inhibitrice (CMI). Une approche biométrique consiste donc à appliquer des modèles statistiques pour la détermination du profil d'une bactérie donnée. Mais il faut s'assurer de la validité de ces modèles choisis. Dans ce travail nous avons présenté les fondements mathématiques et statistiques de la modélisation, et enfin donné une manière d'estimer la CMI. La connaissance de cette valeur nous permettra par la suite de donner le profil de la bactérie.

Introduction

L'utilisation rationnelle des antibiotiques en clinique humaine devrait nécessairement passer par l'étude *in vitro* de la sensibilité des bactéries pathogènes aux différentes molécules antibiotiques. Le clinicien fait appel alors au bactériologiste pour l'aider à établir le diagnostic et le traitement d'une infection. La tâche du bactériologiste consiste alors à isoler, quantifier et identifier le ou les micro-organismes pathogènes responsables de l'infection. Pour orienter le choix de la thérapeutique, il est nécessaire de connaître la sensibilité du micro-organisme pathogène aux antibiotiques. C'est pourquoi ont été développées des méthodes de détermination de la sensibilité bactérienne aux antibiotiques connues sous le nom générique d'"antibiogramme". Le résultat pratique d'un antibiogramme est la classification du micro-organisme. Cette classification est établie à partir d'une grandeur de référence, la Concentration Minimale Inhibitrice (CMI), définie comme la plus faible concentration inhibant toute croissance significative de la population bactérienne.

Cette étude s'inscrit essentiellement dans une démarche méthodologique visant à recenser, proposer, et mettre en œuvre différents outils biométriques en particulier l'étude de modèle pour la détermination de la valeur de la CMI. Il existe plusieurs solutions au problème de détermination de cette valeur. La plus directe consiste à estimer cette valeur à partir de la connaissance d'une grandeur définie directement au niveau de l'antibiogramme. Dans le laboratoire de bactériologie virologie fondamentale et appliquée, pour le test de l'antibiogramme réalisé on récupère un diamètre d'inhibition, et la question que l'on se pose alors est, quelle relation exist-il entre le diamètre d'inhibition et la CMI ? En statistique les méthodes d'analyses comme la régression linéaire, l'analyse log-linéaire ou régression logistique, la comparaison de deux moyennes, l'analyse de variance à un facteur, et l'analyse de la covariance permettent de répondre à cette question. Mais le choix de la méthode dépend de plusieurs facteurs, dont le statut des variables (expliquée(s) et/ou explicative(s)), le type de variable (numériques vs. catégorielles) et la visée de l'analyse (caractérisation, explication, prédiction, confirmation). Pour des variables quantitatives, on applique la méthode de régression pour résoudre le problème d'estimation. L'objectif de ce mémoire est de montrer les méthodes statistiques d'étude de la sensibilité bactérienne et faire une application sur les données du laboratoire de bactériologie-virologie fondamentale et appliquée de la faculté de médecine, de pharmacie, et d'odontostomatologie de l'université Cheikh Anta Diop de Dakar. Nous organisons ce travail en trois parties.

Dans la première partie nous parlerons de la généralité sur les bactéries et le test de sensibilité, dans la deuxième partie on expose les modèles statistiques d'étude de la sensibilité bactérienne, et en troisième partie on fait une application sur les données.

Première partie

Généralité sur les bactéries et
le test de sensibilité

Chapitre 1

Les bactéries et les antibiotiques

1.1 Définition et exemple de bactéries

En microbiologie on distingue deux groupes de micro-organisme : les **procaryotes** et les **eucaryotes**. Les premiers sont des organismes unicellulaire dont le noyau est morphologiquement peu distinct du cytoplasme, quant aux eucaryotes ceux sont des organismes uni ou multi cellulaires dont le noyau est séparé du cytoplasme par une membrane nucléaire. Les bactéries font parti du groupe des procaryotes, on remarque rarement leur présence à cause de leur caractéristique microscopique. Quelques-unes d'entre elles causent des maladies chez les humains mais la plupart sont indispensables à toute vie sur terre. Les bactéries se reproduisent en se divisant en deux cellules identiques. Et cette croissance peut être modélisée par une suite géométrique de raison 2. Par exemple si on note par N_0 le nombre de bactérie au temps t_0 alors pour un temps t donné nous avons le nombte N_t de bactérie sera donné par l'équation suivante :

$$N_t = N_0 2^n$$

ou n est le nombre de génération au temps t .

Pour étudier une bactérie il est nécessaire de l'identifier. En premier lieu une observation microscopique de la forme nous permet de distinguer :

les coques, les bacilles, les vibrions, les fusiformes, et les bactéries spirilles.

Et enfin par des tests d'identification on pourra déceler celles qui sont à gram positif ou à gram négatif. Par exemple pour les coques à gram positif on peut citer :

les **staphylococcus** et les **streptococcus**.

Ainsi pour chaque type d'infection on pourra isoler l'agent pathogène et faire un test d'identification. Ce qui permettra par la suite de se prononcer sur le type de traitement à prendre. Cependant connaître seulement le type de bactérie n'est pas suffisant il faudrait en plus savoir le type d'antibiotique capable de l'inhiber.

1.2 Définition et classification des antibiotiques

Les antibiotiques sont des substances chimiques, produites par des micro-organismes et capables, à faible concentration, d'inhiber la croissance d'autres micro-organismes ou même de les détruire. En d'autre terme ceux sont les médicaments des maladies infectieuses bactériennes.

Chaque antibiotique est caractérisé par son spectre d'activité qui correspond aux différentes espèces bactériennes susceptibles d'être sensibles à son action. Selon les antibiotiques le spectre est limité ou large.

La classification est basée sur leur mode d'action. Ainsi on distingue sommairement :

1. les antibiotiques actifs sur la paroi bactérienne,
2. les antibiotiques actifs sur la membrane cytoplasmique,
3. les antibiotiques actifs sur des processus localisés dans le cytoplasme :
synthèse protéique, réplication de l'ADN ou les deux.

Dans chacun des groupes la classification par famille est fondée sur la structure chimique des différentes molécules. Par exemple on peut citer la famille des Bétalactamines dont la Pénicilline G, les polymyxines, les aminosides, les fluoroquinolones et enfin les inhibiteurs de Bétalactamases.

En connaissant leur mode d'action on peut pour une bactérie donnée se prononcer sur l'antibiotique susceptible de l'inhiber. Mais ceux ne sont pas tous les antibiotiques qui sont efficaces, car il y'a des phénomènes de résistance. Ainsi il serait plus judicieux de faire des tests pour déterminer lequel d'entre eux est le meilleur.

Chapitre 2

Méthodes d'étude de la sensibilité bactérienne

2.1 Introduction

La sensibilité bactérienne se définit comme étant la réaction d'une bactérie vis à vis d'un antibiotique donné. Pour l'étudier nous disposons de technique appelée antibiogramme. Celle-ci a pour but de prédire la sensibilité d'une bactérie à un ou plusieurs antibiotiques, grâce à la détermination de la concentration minimale inhibitrice (CMI). Par définition (O.M.S.), la CMI est la plus faible concentration d'antibiotique capable de provoquer une inhibition complète de la croissance d'une bactérie donnée, appréciable à l'œil nu, après une période d'incubation donnée. On distingue différentes méthodes d'antibiogramme dont on peut citer :

1. les méthodes de **dilution** en milieu liquide ou en milieu solide.
2. les méthodes de **diffusion** connu sous le nom d'antibiogramme standard
3. la méthode de **E-test** qui est une variante de l'antibiogramme standard.

Notons que chacune de ces méthodes sont expérimentales et applicables selon les laboratoires et les conditions de faisabilités. La fiabilité des résultats est influencée par de nombreux paramètres qui doivent être rigoureusement contrôlés. La standardisation est régie par des documents émanant de l'Organisation Mondiale de la Santé (O.M.S) et des divers Comités Nationaux.

2.2 Principe des méthodes de dilution

Les méthodes de dilution sont les méthodes de référence pour la détermination de la CMI. Elles consistent à mettre un inoculum bactérien standardisé au contact de concentrations croissantes d'antibiotiques selon une progression géométrique de raison 2.

En milieu liquide, l'inoculum bactérien est distribué dans une série de tubes (méthode de macrodilution) ou de cupules (méthode de microdilution) contenant l'antibiotique.

Après incubation, la CMI est indiquée par le tube ou la cupule qui contient la plus faible concentration d'antibiotique et où aucune croissance n'est visible.

En milieu solide, l'antibiotique est incorporé dans un milieu gélosé coulé en boîtes de pétri. La surface de la gélose est ensemencée avec un inoculum des souches à étudier (un inoculateur à têtes multiples, appareil de Steers, permet d'ensemencer de 20 à 30 souches différentes par boîte). Après incubation, la CMI de chaque souche est déterminée par l'inhibition de la croissance sur le milieu contenant la plus faible concentration d'antibiotique. La méthode de dilution en milieu gélosé, réalisée avec une gamme de concentrations en progression géométrique de raison 2 (ou de raison 1,25), est la méthode de référence.

2.3 Principe des méthodes de diffusion

Les méthodes de diffusion ou antibiogrammes standards sont les plus utilisées par les laboratoires de diagnostic. Des disques de papier buvard, imprégnés des antibiotiques à tester, sont déposés à la surface d'un milieu gélosé, préalablement ensemencé avec une culture pure de la souche à étudier. Dès l'application des disques, les antibiotiques diffusent de manière uniforme si bien que leurs concentrations sont inversement proportionnelles à la distance du disque. Après incubation, les disques s'entourent de zones d'inhibition circulaires correspondant à une absence de culture. Lorsque la technique est parfaitement standardisée, les diamètres des zones d'inhibition dépendent uniquement de la sensibilité du germe.

2.4 Principe de la méthode des E-test

Le E-test est une technique de détermination de la CMI fondée sur l'utilisation de bandelettes imprégnées d'un gradient exponentiel prédéfini de l'antibiotique à tester couvrant une zone continue de 0,016 à 256mg/l ou 0,002 à 32mg/l en fonction des molécules. Les bandelettes sont des supports inertes, hydrophobes de 5mm de large et 80mm de long. Elles sont appliquées sur la surface d'une gélose préalablement ensemencée avec la souche à étudier.

Après incubation, l'inhibition de la croissance bactérienne se traduit par la présence d'une ellipse dont le point d'intersection avec la bandelette définit la CMI. Une échelle de lecture imprimée sur la face supérieure de la bandelette permet une lecture rapide.

La plupart des méthodes de test de la sensibilité ne donnent que des résultats qualitatifs. C'est à dire nous renseigne directement sur le profil de la bactérie. Quelques unes d'entre elles mesurent soit la CMI (variable quantitative), soit un diamètre d'inhibition (mesure semi-quantitative).

L'interprétation du profil bactérienne se fait à partir de deux concentrations critiques : la **concentration critique supérieure** et la **concentration critique inférieure**. Ces concentrations sont fixées par les Comités Nationaux d'Antibiogramme. Par exemple CA-SFM (Comité de l'Antibiogramme de la Société Française de Microbiologie), EUCAST (European Committee for Antimicrobial Susceptibility Testing).

On distingue les trois catégories de profil suivants :

1. La croissance en présence de la plus forte concentration d'antibiotique est le fait de souches résistantes, en d'autre terme la probabilité d'échec thérapeutique est forte ;
2. La croissance en présence de la plus faible concentration seulement est caractéristique des souches intermédiaires, ce qui signifie que le succès thérapeutique est imprévisible ;
3. l'inhibition de la croissance aux deux concentrations d'antibiotique est spécifique des souches sensibles, qui signifie que la probabilité de succès thérapeutique est forte.

La méthode de référence pour la détermination de la CMI est les méthodes de dilution. Cependant celles-ci sont de mise en œuvre délicate et/ou onéreuses et elles sont réservées à des laboratoires spécialisés. De ce fait pour ce travail les tests expérimentaux seront faits à l'aide d'un antibiogramme standard (méthode de diffusion en gélose inoculum 10^6 bactéries/ml, gélose standardisée (Muller Hinton), disque de papier). Ce qui permettra de récupérer le diamètre d'inhibition comme donnée quantitative.

On remarquera que la variable CMI est très importante pour la détermination du profil de sensibilité d'une bactérie. Mais celle-ci n'est pas directement observable sur le test d'antibiogramme. Il est alors nécessaire de développer des méthodes de prédiction permettant à partir de la valeur du diamètre observée de donner rapidement et avec une moindre erreur une valeur de la CMI plus fiable que possible. Pour cela il faut disposer d'un modèle exprimant la liaison entre la CMI et diamètre.

Deuxième partie

Modèle statistique d'étude de
la sensibilité Bactérienne

Chapitre 3

Modèle statistique

3.1 Introduction

Selon Brown et Rothery, 1993 ; Witzjes, 1996 ; Pavé, 1994, un modèle est une représentation simplifiée de la réalité, facilitant la prédiction ou l'estimation et est exprimée en langage mathématique. Toute la réalité ne pouvant pas être modélisée, un choix doit être fait sur la partie de la réalité à modéliser. Et ainsi on définit l'espace du modèle, ce qui implique que les informations qui pourront être extraites de la modélisation ne seront valables que pour l'espace choisit. Il existe deux types de modèle selon l'objectif de la modélisation :

les modèle empiriques , les modèle mécanistes.

Si l'objectif est de faire de la prédiction alors le modèle empirique est largement suffisant. Par contre les modèles mécanistes sont construits à partir d'hypothèses expliquant les processus donnant lieu aux phénomènes observés.

3.2 Objectif de l'étude et démarche de modélisation

L'objectif de cette étude est de déterminer le profil d'une bactérie vis à vis de plusieurs antibiotiques. Cependant au lieu d'interpréter le profil à partir de données qualitatives, nous allons au contraire comparer la CMI avec les concentration critiques supérieures et inférieures.

Ainsi suivant que la valeur de la CMI est inférieure, supérieure ou se trouve entre ces deux valeurs critiques, on classera la bactérie selon les trois catégories définies précédemment. Dans cet objectif, un modèle empirique de description de la relation entre la CMI et le diamètre d'inhibition sera construit. Ceci permettrait d'estimer la valeur de la CMI, à partir de la valeur observée de celui-ci.

Dés lors la démarche de modélisation de l'étude de la sensibilité bactérienne suit les étapes suivantes :

1. Recueil des données,
2. Description mathématique des données par l'examen de la morphologie des nuages de points obtenus,
3. Ajustement d'un modèle d'étude,
4. Validation du modèle construit à l'étape précédente,
5. Prédiction

3.3 Les données

Les infections respiratoires aiguës (IRA) communautaires telles que l'otite moyenne aiguë (OMA), la rhino-sinusite, la bronchite et la pneumonie représentent la majorité des infections courantes chez les enfants. Ceux sont des maladies prévisibles et traitables, cependant elles peuvent causer des morts prématurés chez les enfants de moins de cinq ans si elles ne sont pas bien traitées.

Elles sont le premier motif de prescription d'antibiotique, et des études sur leur origine bactérienne ont montré qu'elles sont la cause de quatre bactéries :

Haemophilus influenzae, Moraxella catarrhalis, Streptococcus pneumoniae, Streptococcus pyogenes.

Durant ces dernières années, la résistance de ces bactéries aux principales classes d'antibiotiques a émergé et évolue de façon croissante dans le monde entier. L'émergence et l'expansion de cette résistance cause l'échec thérapeutique de certains antibiotiques, ce qui conduit à la recherche de nouvelles alternatives au traitement. Face à cette préoccupation, des études ont été menées ces dernières années pour déterminer la sensibilité aux antibiotiques de différentes souches de ces bactéries.

Dans le cadre de ce travail nous avons obtenu des données provenant d'une étude menée à Dakar entre janvier 2006 et Mai 2006. L'étude a été effectuée à l'unité de recherche et de biotechnologie microbienne du laboratoire de Bactériologie-Virologie du Centre Hospitalier Universitaire Aristide le Dantec. Des souches de Streptococcus pyogène, Streptococcus pneumoniae, et d'Haemophilus influenzae ont été isolé chez 133 patients atteints d'IRA dont 68 enfants et 65 adultes.

La détermination de la sensibilité aux antibiotiques est basée sur la recherche des diamètres d'inhibition et sur la détermination des concentrations minimales inhibitrices (CMI). Deux méthodes ont été utilisées :

- Diffusion en milieu gélosé (antibiogramme standard par la méthode des disques)
- E-test

Ainsi nous avons constitué une base de donnée contenant un échantillon de 71 souches dont 23 souches d'Haemophilus influenzae, 39 souches d'Streptococcus pyogène, 9 souches d'Streptococcus pneumoniae.

Des tests ont été faits avec les antibiotiques suivants :

Péniciline G, Chloramphénicole, Erythromycine, Teicoplanine, Ampicilline, Levofloxacin. Pour chaque test on a récupéré la CMI par le biais du E-test, et le diamètre d'inhibition grâce à l'antibiogramme standard par la méthode des disques. Les

données ont été saisies dans des tableaux Excel.

3.4 Description mathématique des données

On dispose de n souches sur lesquels on a mesuré deux variables quantitatives, la CMI et le diamètre. Par exemple le test du Chloramphénicol avec les souches des trois types de bactéries citées précédemment nous donne les tableaux suivants :

Code souche	Diamètre (en mm)	CMI (en mg/ml)
SN 001 091	30	0,5
SN 001 092	31	0,5
SN 001 093	32	0,75
SN 001 094	32	1,5
SN 001 095	32	1
SN 001 096	12	8
SN 001 097	33	0,5
SN 001 098	30	0,5
SN 001 099	32	1

Tableau : 3.1 - Test avec *Streptococcus pneumoniae*

Code souche	Diamètre (en mm)	CMI (en mg/ml)
SN 001 200	20	3
SN 001 201	20	3
SN 001 202	20	2
SN 001 203	21	3
SN 001 204	20	4
SN 001 205	20	4
SN 001 206	25	6
SN 001 207	21	3
SN 001 208	24	3
SN 001 209	25	2
SN 001 210	26	3
SN 001 211	26	2
SN 001 212	25	3
SN 001 213	28	3
SN 001 214	28	2
SN 001 215	24	3
SN 001 216	28	3
SN 001 217	24	3
SN 001 218	23	3
SN 001 219	24	4
SN 001 220	24	4
SN 001 221	26	3
SN 001 222	22	4
SN 001 223	28	2
SN 001 224	26	3
SN 001 225	29	3
SN 001 226	21	4
SN 001 227	20	3
SN 001 228	26	3
SN 001 229	22	3
SN 001 230	27	3
SN 001 231	26	4
SN 001 232	26	3
SN 001 233	20	2
SN 001 234	25	2
SN 001 235	20	3
SN 001 236	27	2
SN 001 237	21	3
SN 001 238	20	2

Tableau : 3.2 - Test avec Streptococcus pyogène

Code souche	Diamètre (en mm)	CMI (en mg/ml)
SN 001 121	18	8
SN 001 122	27	0,5
SN 001 123	20	16
SN 001 124	32	0,125
SN 001 125	31	0,064
SN 001 126	29	0,064
SN 001 127	20	8
SN 001 128	30	0,064
SN 001 129	30	0,5
SN 001 130	32	0,064
SN 001 131	29	0,125
SN 001 132	32	0,38
SN 001 133	19	4
SN 001 134	32	0,125
SN 001 135	30	0,125
SN 001 136	32	0,125
SN 001 137	32	0,064
SN 001 138	28	0,125
SN 001 139	15	8
SN 001 140	31	0,125
SN 001 141	32	0,19
SN 001 142	16	8
SN 001 143	34	0,064

Tableau : 3.3 - Test avec Haemophilus Influenzae

Dans les tableaux suivants le champ code souche nous permet d'identifier la souche. L'objectif d'un tel recueil de données est d'étudier la liaison entre la CMI et le diamètres. Pour cela une première idée consiste à réaliser un graphique cartésien, dont l'abscisse représente le diamètre et l'ordonnée la CMI. Dans ce repère chaque souche est représenté par un point. l'ensemble des souches constitue le nuage de point dont la forme révèle la liaison entre ces deux variables.

Pour les trois tableaux ci dessus on peut donner les graphes représentant le nuage de point.

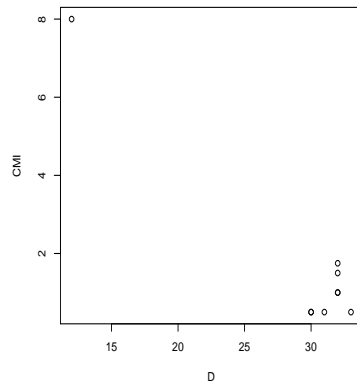


FIG. 3.1 – Nuage de point du Tab :3.1

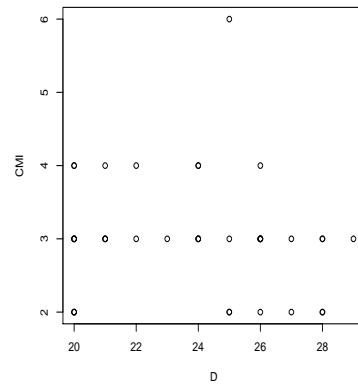


FIG. 3.2- Nuage de point du Tab :3.2

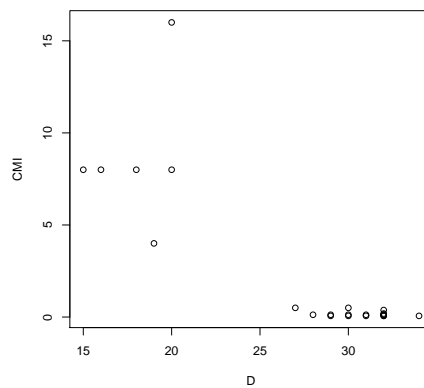


FIG. 3.3- Nuage de point du Tab : 3.3

Notons que le nuage de point nous donne seulement une idée sur la liaison, et que celle-ci doit pouvoir être vérifiée par des tests.

3.5 Ajustement d'un modèle d'étude

3.5.1 Introduction

Toute tentative de modélisation nécessite une étude descriptive préalable afin de s'assurer, au moins graphiquement, de la validité des hypothèses considérées. Ceci passe par une représentation du nuage de point, ce qui a été déjà fait dans l'étape précédente afin de déceler une éventuelle liaison linéaire ou non-linéaire entre les variables. Pour prédire la valeur de la CMI à partir de la connaissance de la valeur du diamètre d'inhibition, il faudrait alors disposer d'une équation exprimant la CMI à partir du diamètre soit :

$$CMI = f(D)$$

et tenter par la suite de spécifier la forme de cette fonction.

Le type de liaison le plus simple et le plus répandu entre deux variables quantitatives est la liaison linéaire ; celle-ci peut être positive (plus le diamètre est grand et plus la CMI est grande) ou négative (plus le diamètre est grand et plus la CMI est petite). Graphiquement, une telle liaison se traduit par un nuage de forme allongée. A la limite, lorsque la liaison linéaire est parfaite, les points du nuage sont alignés. Pour vérifier cette relation nous disposons d'indicateurs descriptifs de la liaison entre deux variables quantitatives : la **covariance** et le **coefficient de corrélation linéaire** de Bravais-Pearson.

3.5.2 La covariance

La covariance est une mesure de la force du lien entre deux variables aléatoires (numériques). Soit X, Y deux variables aléatoires on a :

$$\text{cov}(X, Y) = E[(X - \bar{X})(Y - \bar{Y})]$$

ou \bar{X} et \bar{Y} sont les moyennes arithmétiques respectifs de X et Y .

Si on prend un échantillon de taille n avec x, y les valeurs observées des variables X et Y la covariance s'exprime alors :

$$\text{cov}(x, y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

La covariance peut être faible parce que le lien entre les variables est effectivement faible, mais aussi parce que le lien est non linéaire. L'inconvénient de la covariance est que sa valeur dépend des unités dans lesquelles sont exprimées les variables X et Y , alors qu'une mesure de la force du lien entre deux variables ne devrait certainement pas en dépendre. C'est la raison pour laquelle, elle est souvent délaissée au profit de sa version standardisée, le coefficient de corrélation.

3.5.3 Le coefficient de corrélation

Le coefficient de corrélation linéaire a pour objet de quantifier l'allure plus ou moins linéaire d'un nuage de point. Il est défini simplement comme la moyenne des produits deux à deux des données centrées réduites.

$$r = \frac{1}{n} \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\left[\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2 \right]^{1/2}}$$

On rencontre aussi la définition suivante :

$$r = \frac{\text{cov}(x, y)}{\left[\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2 \right]^{1/2}}$$

Il est toujours compris entre -1 et 1 , valeurs atteintes lorsque la liaison linéaire est parfaite c'est à dire les points sont alignés.

Au coefficient de corrélation observé correspond la variable R^2 qui est le carré de celui-ci. Elle indique la part de variabilité de Y expliquée par la prise en compte de la liaison entre X et Y .

Une forte corrélation linéaire entre deux variables peut être décelée par un test appelé **test de signification du coefficient de corrélation**. En outre on dispose de tables qui fournissent les valeurs auxquelles comparer directement les valeurs absolues de r . Cette forte corrélation justifie le choix d'un modèle de régression linéaire simple.

3.5.4 Le modèle de régression simple

Deux variables quantitatives ont été mesurées sur n observation. Mais dans ce cas les deux variables ne jouent pas de rôle symétrique : on veut expliquer l'une par l'autre. Dans le cas de la régression linéaire simple la relation entre x et y est donnée par :

$$y = a + bx \tag{3.1}$$

Pour obtenir cette équation, on postule un modèle dont on estime les paramètres. Le de modèle de régression simple est :

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (\text{ML1})$$

Dans lequel β_0 est l'intercepte ou ordonnée à l'origine et β_1 est la pente en x_i . La variable y_i est dite variable à expliquer ou à prédire ou endogène ou encore (variable) réponse, x_i est dite variable explicative ou prédictrice ou exogène ou encore prédicteur. Quant à ε_i c'est une variable aléatoire représentant l'erreur commise en ajustant un modèle de régression simple sur les données, elle est encore appelée résidu.

L'écriture matricielle du modèle (ML1) est la suivante :

$$Y = \beta X + \varepsilon$$

avec

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{pmatrix} \quad \mathbf{X} = \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_i & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_0 \end{pmatrix} \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_i \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Le vecteur β est inconnu et doit être estimé. En outre il faut vérifier les hypothèses du modèle pour que celui-ci soit valable.

Soit $E(\varepsilon)$ le vecteur des espérances des n coordonnées de ε et $cov(\varepsilon)$ la matrice $n * n$ de terme (i,j) égal à $cov(\varepsilon_i, \varepsilon_j)$. Les résidus doivent vérifier les hypothèses suivantes :

1. $E(\varepsilon) = 0$
2. $cov(\varepsilon) = \sigma^2 I_n$
3. pour $i = 1 \dots n$ les variables ε_i sont indépendante et de loi gaussienne

Ce qui signifie que les erreurs doivent être centrée et non corrélées.

3.5.5 Estimation des paramètres

Le modèle étant postulé, il faut en estimer les paramètres, en régression simple on cherche le vecteur β qui ajuste aux mieux les données (i.e. les x_i et les y_i), autrement dit telles que les résidus soient petits.

L'estimateur de β par moindres carrés ordinaires (MCO) est une valeur $\hat{\beta}$ qui minimise la somme des carrés résiduels :

$$SCR(\beta) = \|\mathbf{Y} - \mathbf{X}\beta\|^2 = \sum_{i=1}^n (y_i - x_i\beta)^2$$

Notons qu'ici x_i représente un vecteur ligne de deux colonnes. Les dérivées partielles de $SCR(\beta)$ par rapport aux paramètres β_0 et β_1 donne :

$$\begin{cases} \frac{\partial SCR(\beta)}{\partial \beta_1} = -2 \sum_{i=1}^n x_i (y_i - (x_i \beta_1 + \beta_0)) \\ \frac{\partial SCR(\beta)}{\partial \beta_0} = 2 \sum_{i=1}^n (y_i - (x_i \beta_1 + \beta_0)) \end{cases}$$

Le point critique est obtenu en posant que les dérivées partielles sont nulles. Ce qui conduit au système suivant :

$$\begin{cases} \sum_{i=1}^n x_i y_i - \beta_1 \sum_{i=1}^n x_i^2 - \beta_0 \sum_{i=1}^n x_i = 0 \\ \sum_{i=1}^n y_i - \beta_1 \sum_{i=1}^n x_i + n\beta_0 = 0 \end{cases}$$

La solution de ce problème de minimisation est :

$$\begin{cases} \hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{cov(x, y)}{var(x)} \\ \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \end{cases}$$

avec $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$, $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$.

Remarquons que le coefficient de régression (β_1) et le coefficient de corrélation linéaire (r) sont étroitement liés :

$$\beta_1 = r \frac{s_y}{s_x}$$

avec s_x et s_y les écarts type respectifs de x et y .

3.5.6 Propriétés des estimateurs

On peut montrer que ces estimateurs sont sans biais soit :

$$E(\hat{\beta}_1) = \beta_1, \quad E(\hat{\beta}_0) = \beta_0.$$

En outre :

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sigma^2}{ns_x^2} = \sigma_1$$

avec σ^2 la variance résiduelle.

Cette variance exprime l'incertitude dans l'estimation de β_1 . On constate que celle-ci croît avec la variance résiduelle et décroît avec le nombre n de données.

La variance de l'estimateur $\hat{\beta}_0$ vaut :

$$Var(\hat{\beta}_0) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_1^n (x_i - \bar{x})^2} \right) = \frac{\sigma^2}{n} \left(1 + \frac{\bar{x}^2}{s_x^2} \right) = \sigma_0^2$$

Ici encore, la précision de l'estimateur croît avec le nombre de mesures n et décroît avec la variance résiduelle.

3.5.7 Propriétés des résidus

On note \hat{y}_i la valeur prédite par le modèle en x_i : $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$; le résidu observé s'écrit : $e_i = y_i - \hat{y}_i$.

Selon les hypothèses du modèle on a par construction que : $\sum_{i=1}^n e_i = 0$; intuitivement, ceci est dû à la présence d'une constante dans le modèle. En outre

$$\sum_{i=1}^n e_i x_i = ncov(e, x) = 0$$

Ceci exprime que les résidus sont non corrélés aux valeurs de x .

La variance résiduelle comme son nom l'indique est la variance des résidus. Elle est estimée en divisant la somme des carrés des résidus observés par son nombre de degrés de liberté (nombre de données diminué du nombre de paramètres estimés, soit ici $n - 2$)

$$\hat{\sigma}^2 = \frac{\sum_i (y_i - \hat{y}_i)^2}{n - 2} = \frac{\sum_i e_i^2}{n - 2}$$

On montre que cet estimateur est sans biais : $E(\hat{\sigma}^2) = \sigma^2$.

3.5.8 Test de signification des paramètres estimés

Ce test permet de se prononcer sur l'importance d'un paramètre dans le modèle. Ce qui revient à tester la nullité du paramètre.

Test de signification du coefficient de régression

Sachant que l'on observe jamais exactement $\hat{\beta}_1 = 0$, on peut se demander si cet écart est le fruit du hasard ou d'une réelle influence de x sur y . Autrement dit, l'estimateur de β_1 obtenue est-elle significative ? Ce test nous permet de répondre à cette question. Le principe du test consiste à choisir entre les deux hypothèses suivantes :

$$\begin{cases} H_0 : \beta_1 = 0 \\ H_1 : \beta_1 \neq 0 \end{cases}$$

Le problème de choix sera plutôt abordé sous l'angle : faut-il accepter ou non H_0 . La statistique de test du coefficient est construite en le rapportant à son écart type soit :

$$\hat{T} = \frac{\hat{\beta}_1}{\hat{\sigma}_1}$$

ou $\hat{\sigma}_1$ est estimateur de σ_1 qui est obtenu en remplaçant la variance résiduelle par son estimateur.

$$\hat{\sigma}_1 = \sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n (x_i - \bar{x})}}$$

Sous H_0 la statistique de test suit une loi de Student à $n - 2$ degrés de liberté qui est celui de la variance résiduelle.

Règle de décision

Pour un risque de première espèce α (généralement 5%) fixé on définit à partir de la loi sous H_0 , une zone d'acceptation de H_0 . Les bornes de celle-ci sont les quantiles correspondant aux probabilités $\alpha/2$ et $1 - \alpha/2$. Ainsi si on a :

$$|\hat{\beta}| \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})}{\hat{\sigma}^2}} > T_{n-2, 1-\alpha/2}$$

l'hypothèse H_0 est rejeté, on dit alors le test est significatif, sinon on accepte H_0 et on dira que le test n'est pas significatif. l'autre façon de procéder consiste à calculer la probabilité critique associée à la statistique de test et à la comparer à α .

On peut construire un test analogue à propos du paramètre β_0 . La statistique de test associée dans ce cas sera :

$$\hat{T} = \frac{\hat{\beta}_0}{\hat{\sigma}_0}$$

$$\text{avec } \hat{\sigma}_0 = \sqrt{\frac{\hat{\sigma}}{n} \left(1 + \frac{\bar{x}^2}{s_x^2}\right)}.$$

Sous l'hypothèse nulle ($\beta_0 = 0$) cette statistique suit une loi de Student à $n-2$ degrés de liberté. La règle de décision reste la même que pour le test sur le coefficient de régression.

Dans le cadre de notre étude, si une forte corrélation linéaire est décelée entre la CMI et le diamètre d'inhibition, ce qui se traduit graphiquement par un nuage de forme allongée, alors le modèle de régression simple sera approprié pour monter la relation qui existe entre ces deux variables. Cependant si telle n'est pas le cas on essaiera de trouver une courbe qui représente le mieux possible le nuage de point. En présence d'une liaison non-linéaire qui peut conduire néanmoins à un coefficient de corrélation linéaire très faible, le remède consiste à rechercher des transformations des variables permettant de rendre linéaire la relation.

Ces vingt dernières années, à la suite du développement des moyens de calcul, sont apparues des méthodes d'estimation dites fonctionnelles ou non-paramétriques qui proposent d'estimer la relation entre deux variables par une fonction construite point par point (noyau) ou dans une base de fonction splines.

3.5.9 La régression non paramétrique

On considère un modèle de régression de la forme

$$y_i = f(x_i) + \varepsilon_i \quad (3.2)$$

où les erreurs sont centrées et la fonction f est supposée régulière ce qui veut dire les dérivées existent jusqu'à un certain ordre. Dans ce contexte de nombreux estimateurs de celle-ci ont été proposés dont on peut citer :

Le lissage spline élémentaire qui consiste à chercher dans l'espace des fonctions continûment différentiables et avec une dérivée seconde de carré intégrable, le minimum d'un critère combinant ajustement des observations et régularité de la solution.

$$\hat{f}_\lambda = \underset{f}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^n (y_i - f(x_i))^2 + \lambda \int_{-\infty}^{+\infty} (f''(x))^2 dx$$

On montre que l'on obtient une fonction polynômiale (de degré 3) par morceau. La valeur du paramètre de lissage est fixée par validation croisée généralisée (GCV).

La régression non-paramétrique par la méthode du noyau qui consiste à calculer une moyenne pondérée autour de chaque observation. La pondération est fixée par une fonction K du même type que celle pour l'estimation par la méthode du noyau de la densité.

$$\hat{f}_\lambda = \sum_{i=1}^n \frac{K\left(\frac{x-x_i}{\lambda}\right) y_i}{\sum_{j=1}^n K\left(\frac{x-x_j}{\lambda}\right)}$$

La transformation qui consiste à rechercher des transformations élémentaires (logarithme, puissance) des variables susceptibles de rendre linéaire la relation entre les deux variables. Ceci amène à étudier les modèles des exemples suivants :

$$\begin{aligned}
 Y &= \beta_0 + \beta_1 \ln X \\
 \ln Y &= \beta_0 + \beta_1 X \quad \text{ou} \quad Y = ab^X \quad \text{avec} \quad \beta_0 = \ln a \quad \text{et} \quad \beta_1 = \ln b \\
 \ln Y &= \beta_0 + \beta_1 \ln X \quad \text{ou} \quad Y = aX^{\beta_1} \quad \text{avec} \quad \beta_0 = \ln a \\
 Y &= \beta_0 + \beta_1 X^\alpha
 \end{aligned}$$

Dans ce travail nous avons choisi la méthode de transformation pour résoudre le cas non-linéaire décelé par un coefficient de corrélation linéaire très faible. Notons que le caractère faible de celui-ci est donné en le comparant avec une valeur fixe. Un tel choix a été motivé par des recherches menées dans le cadre de l'antibiogramme. Celle-ci a débuté par les travaux de Cooper et Woodman dès les années 1946 et une autre voie sera donnée dès les années 50 par Tunneval et Ericson ; Bauer et al ; Chabbert.

3.5.10 Méthode des "courbes de concordance"

Ces auteurs ont établi une relation linéaire reliant pour une charge d'antibiotique donnée, le diamètre de la zone d'inhibition au logarithme de la CMI déterminée par la méthode de référence. Ainsi des **courbes de concordance** ont été établies par régression linéaire pour un antibiotique donné, sur un grand nombre de souches bactériennes. L'établissement de ces courbes et leur emploi ont ensuite été normalisés par une étude collective internationale patronnée par l'O.M.S. Dans ce contexte nous allons considérer les équations suivantes :

liaison entre le logarithme de la CMI et le rayon r

$$\ln CMI = \beta_0 + \beta_1 r \quad (3.3)$$

ou une liaison avec le rayon au carré

$$\ln CMI = \beta_0 + \beta_1 r^2 \quad (3.4)$$

Et ainsi on postule un modèle de régression linéaire simple pour obtenir ces types d'équation. Mais contrairement au premier cas celui-ci se fait sur des transformations du diamètre et de la CMI. Le modèle étant bien spécifié, les hypothèses bien décrites, il est alors important que celui-ci soit valide.

3.5.11 Validation du modèle

La validation consiste à vérifier les hypothèses. Pour le modèle de régression linéaire vu précédemment nous allons vérifier les points suivants :

1- linéarité

On peut vérifier visuellement par affichage en nuage de point.
Les signes des coefficients.

2- normalité des ε

Affichage de l'histogramme des résidus.

Hétéroscédasticité

Est-ce que les termes d'erreur ont une variance constante ?
Vérifier l'affichage des résidus versus l'affichage de Y avec X .

Auto-corrélation

Est-ce que les termes d'erreur sont indépendants ?
On affiche les résidus de façon à vérifier les régularités.

Le modèle étant valide, on peut alors passer à l'objectif qui est de pouvoir prédire la CMI à partir de la valeur du diamètre connue.

3.5.12 La prévision

L'un des intérêts pratiques des modèles est la prévision. Ce terme de prévision se réfère au futur, mais en statistique, il est plus général on se demande à quelle valeur de la variable à expliquer peut-on s'attendre pour une valeur de la variable explicative donnée. Une fois les paramètres estimés, on peut ensuite prévoir les valeurs y_i pour x_i donné. La prévision est dite ex post s'il s'agit de prévoir des valeurs déjà connues ou ex ante sinon.

Ces deux dernières sections sont plutôt pratiques, raison pour laquelle nous allons plus les développer dans la partie pratique.

Troisième partie

Application aux données

Introduction

Le logiciel R est un environnement de programmation statistique interfaçable avec C et Fortan. Il a été initialement créé par Robert Gentleman et Ross Ihaka du département de statistique de l'Université d'Auckland en Nouvelle Zélande. Depuis 1997, il s'est formé une équipe (la **R Core Team**) qui développe R.

R permet de réaliser des analyses statistiques et possède :

1. un système efficace de manipulation et de stockage de données
2. différents opérateurs pour le calcul sur tableaux, en particulier les matrices
3. des moyens graphiques pour visualiser les analyses
4. un langage de programmation simple et performant comportant conditions, boucles, moyens d'entrées sorties, possibilité de définir des fonctions récursives.

La gestion des fonctions se fait à l'aide de la notion de modules (packages).

R est un logiciel libre disponible sous Windows, Unix, Linux et Macintosh. Il est diffusé sous licence GNU. Et le logiciel est disponible depuis le site CRAN (Comprehensive R Archive Network).

Le CRAN est un ensemble de site qui fournit ce qui est nécessaire à la distribution de R, ses extensions, sa documentation, ses fichiers sources et ses fichiers binaires. Le site maitre est situé en Autriche à Vienne. On peut y accéder grâce à l'URL suivant :

`http://cran.r-project.org/`

Il existe des sites miroirs, et pour télécharger les applications du projet R, il est conseillé d'accéder au site miroir le plus proche géographiquement de l'endroit de votre connexion. C'est avec ce logiciel que nous allons faire notre application qui consiste à stocker les données, tracer les graphes, appliquer le modèle et afin faire des analyses.

Chapitre 4

Profil de sensibilité du Streptococcus Pyogenes

Nous disposons d'un échantillon de 39 souches de cette bactéries testée avec les antibiotiques suivants :

Erythromycine, Péniciline G, Chloramphénicol, Teicoplanine, Lévofloxacine

4.1 Etude de la liaison entre la CMI et le diamètre d'inhibition

Les données recueillies ont été enregistrées dans un fichier Excel, cependant pour pouvoir l'utiliser sous **R** on l'a transformé en fichier "**nom.csv**".

Caractéristique des données

La caractéristique des données nous renseigne sur le type et le champ de variation des données que nous manipulons.

Pour la plupart des antibiotiques testés, la CMI est de type numérique et le diamètre est de type entier, sauf pour le test avec le **Chloramphénicol** qui donne des entiers pour les deux variables. Nous récupérons sous forme de tableau le minimum et le maximum pour la CMI et le diamètre.

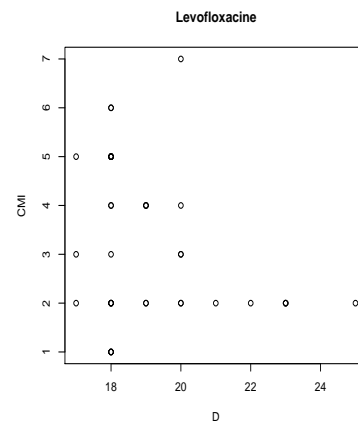
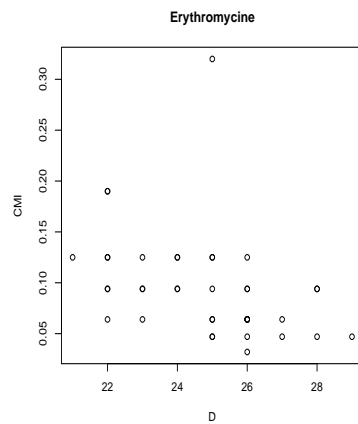
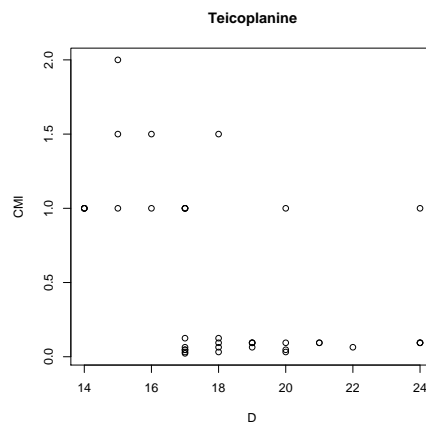
CMI		
Antibiotiques	Min	Max
Péniciline G	0,004	0,032
Erythromycine	0,032	0,32
Teicoplanine	0,023	2
Chloramphénicol	2	6
Levofloxacine	0,38	2

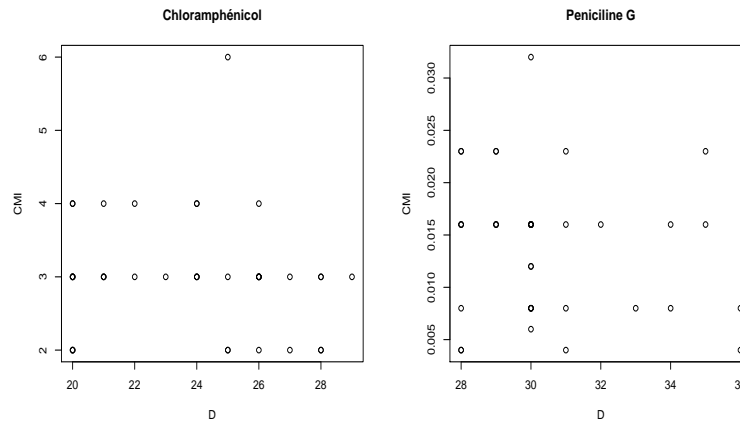
D		
Antibiotiques	Min	Max
Péniciline G	28	36
Erythromycine	21	29
Teicoplanine	14	24
Chloramphénicol	20	29
Levofloxacin	17	25

Une analyse de ces deux tableaux montre que l'on a de grandes valeurs de diamètre face à de très petites valeurs de la CMI. Ce qui signifie que si la liaison entre ces deux variables est linéaire, elle sera alors négative.

Pour déterminer le type de liaison qui existe entre ces deux variables nous allons en premier lieu tracer les graphes représentant les nuages de points :

Nuage de point





Un premier regard de ces graphes ne nous renseigne pas vraiment sur le type de relation car on a pas des formes allongées. Cependant nous disposons d'indicateurs pour décèler une relation linéaire. Nous avons utiliser un test pour déterminer si le coefficient de corrélation est significative ou pas. Nous obtenons les résultats suivants.

Pénicilline G

$cor(D, CMI) = -0,1646$
 $p\text{-value} = 0,3165$

Erythromycine

$cor(D, CMI) = -0,3783$
 $p\text{-value} = 0,01754$

Teicoplanine

$cor(D, CMI) = -0,5071$
 $p\text{-value} = 0,00098$

Chloramphénicole

$cor(D, CMI) = -0,1195$
 $p\text{-value} = 0,4684$

Lévofloxacine

$cor(D, CMI) = -0,3688$
 $p\text{-value} = 0,02088$

Au vu de ces tableaux on peut utiliser le critère psychométrique pour donner des conclusions. Ainsi on va comparer le coefficient de corrélation avec les valeurs 0,20 et 0,40 qui va s'interpréter comme suit

- $|cor| < 0,20$, corrélation faible
- $|cor| > 0,40$, corrélation importante
- $0,20 < |cor| < 0,40$, corrélation intermédiaire

Mais aussi pour un seuil α de 0,05 fixé on va comparer cette valeur avec la p-value obtenue à l'issu du test. En combinant ces deux méthodes on dira que l'on a une relation linéaire simple entre le diamètre et la CMI pour les antibiotiques suivants :

Teicoplanine, Erythromycin et Levofloxacine.

4.2 Choix du modèle de régression linéaire simple

Dans cette section nous allons par le biais des commandes de **R** calculer les estimateurs des coefficients vus précédemment et vérifier les hypothèses du modèle.

4.2.1 Application du modèle sur le test avec l'Erythromycine

On considère l'équation suivante

$$CMI = \beta_0 + \beta_1 D \quad (4.1)$$

En ajustant le modèle sur les données nous obtenons les résultats suivants :

Call:

```
lm(formula = CMI ~ D)
```

Residuals:

```
      Min       1Q   Median
-0.058586 -0.027750 -0.009117
      3Q       Max
 0.017016  0.226617
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.336746	0.097084	3.469	0.00134 **
D	-0.009735	0.003915	-2.487	0.01754 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.04888 on 37 degrees of freedom

Multiple R-squared: 0.1432, Adjusted R-squared: 0.12

F-statistic: 6.183 on 1 and 37 DF, p-value: 0.01754

D'après le listage, les estimateurs des moindres carrés des paramètres sont :

$\hat{\beta}_0 = 0.336746$, $\hat{\beta}_1 = -0.009735$. On a aussi les estimateurs des variances et la p-value du test de signification des paramètres. Et ainsi on obtient l'équation suivante qui pour une valeur du diamètre nous donne la valeur de la CMI.

$$CMI = 0,336746 - 0,009735D \quad (4.2)$$

La probabilité critique associée au test de signification du paramètre β_1 est de 0.01754, valeur inférieure à 0.05. On considère donc que ce paramètre est significativement différent de 0, et par suite il existe une liaison linéaire entre la CMI et le diamètre d'inhibition.

Le test sur β_0 permet de choisir entre un modèle de proportionnalité ($\beta_0 = 0$) et un modèle de régression ($\beta_0 \neq 0$). Dans cet exemple il semble qu'un modèle de régression est plus adapté car on a la probabilité critique associée au test de β_0 est de 0.00134 qui est inférieure à la valeur du seuil fixé par défaut (0.05).

La part de variabilité de la CMI expliquée par la prise en compte de la liaison linéaire entre le diamètre et la CMI est donnée par le "Multiple R-squared". Cette valeur vaut 0.1432 ce qui signifie que 14.32% de la variabilité de la CMI est expliquée par le diamètre. Ce pourcentage n'est pas très élevé, mais il est significatif (i.e trop important pour être raisonnablement attribué au hasard). On complète l'étude en construisant la table suivante, encore appelée table de la variance.

Anova Table (Type II tests)

Response: CMI

	Sum Sq	Df	F value	Pr(>F)
D	0.014773	1	6.1829	0.01754 *
Residuals	0.088405	37		

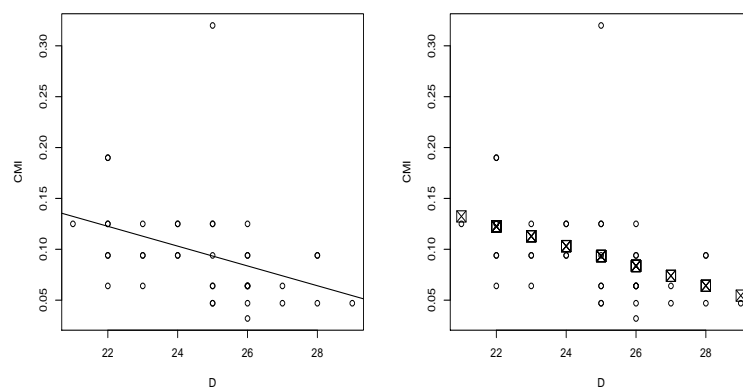
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

La valeur 0.014773 représente la part expliquée par le modèle ou encore la part due au modèle, et 0.088405 la somme des erreurs résiduelles. Et on retrouve le test de nullité du coefficient de régression.

Nous allons passer à l'étape de validations en vérifiant les hypothèses.

Test des hypothèses du modèle

Linéarité

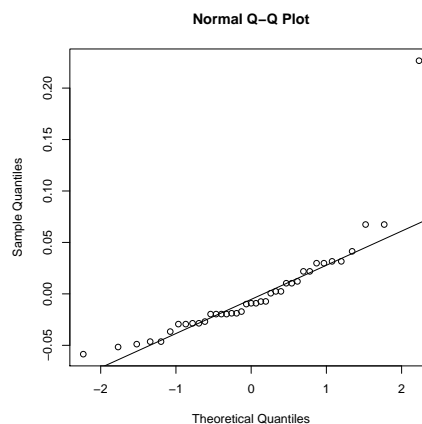


Visuellement nous remarquons que la droite est proche du maximum de point, et que

la distance entre les points prédites et les points du nuage sont petits. Ce qui signifie que la droite ajuste au mieux le nuage, d'où la relation est linéaire.

Normalité des résidus

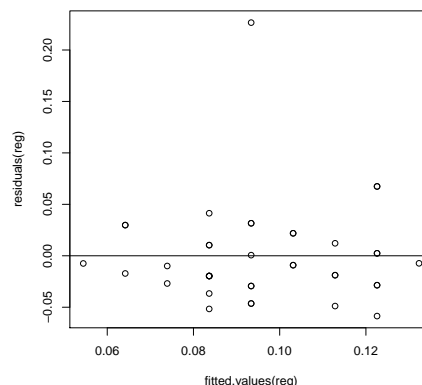
On peut vérifier cette hypothèse en analysant le résultat d'un test de normalité ou par le moyen d'un graphe en regardant comment les données sont distribuées sur la droite de Henri. Le test de Shapiro - Wilk est l'un des tests permettant de vérifier la normalité d'un échantillon. Cependant il est plus efficace d'utiliser les graphes. Ainsi on choisit de vérifier l'alignement des données sur la droite de Henri ou q-qplot.



Ce graphe confirme bien l'hypothèse de normalité des résidus, avec les données qui sont bien rangées suivant la droite.

Hétéroscédasticité

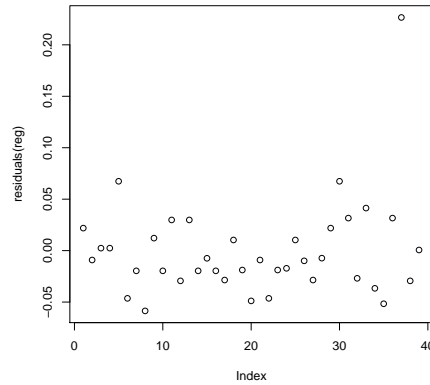
Ici aussi nous allons conclure à partir du graphe des résidus avec les valeurs ajustées. Les valeurs ajustées sont obtenues en calculant la CMI à partir de l'équation avec les paramètres estimés.



On remarque qu'il n'y a pas d'évidence d'hétéroscédasticité. Ce qui signifie en d'autre terme que la variance des résidus est constante.

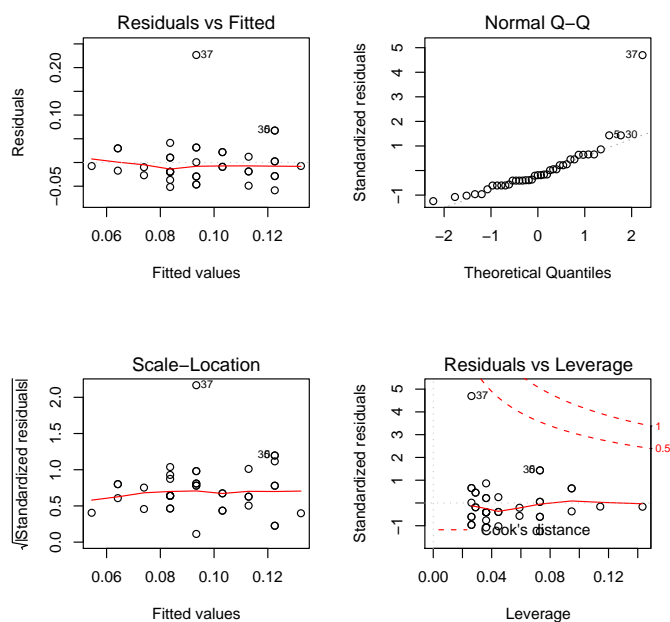
Auto-corrélation

Notons que l'autocorrélation peut être présente si les observations ont un ordre séquentiel naturel (par exemple, le temps).



Pas d'évidence d'auto-corrélation ce qui signifie que les termes d'erreur sont indépendants.

Nous résumons les tests de validités du modèle par le graphe suivant :



Par la suite nous utiliserons ce type de graphe pour la vérification des hypothèses du modèle.

Prévision de la CMI

Pour une valeur du diamètre donnée on obtient la valeur prédite de la CMI en remplaçant celle-ci dans l'équation qui contient les estimateurs. Dans le logiciel **R** nous

disposons de la fonction "**predict()**" qui nous permet d'avoir la valeur prédite et l'intervalle de prévision. Voici un exemple d'appel et on obtient le résultat suivant.

```
predict(reg,data.frame(D=45),interval="prediction")
      fit      lwr      upr
1 -0.1013082 -0.2908981 0.08828163
```

Avec "fit" qui nous donne la valeur prédite, "lwr" et "upr" donnent les deux bornes de l'intervalle de confiance.

En résumé nous dirons que nous avons obtenu une manière de déterminer la CMI en contournant la méthode expérimentale qui peut prendre du temps. Et ainsi pour se prononcer sur le profil de sensibilité du *Streptococcus pyogenes* vis à vis de l'Erythromycine, il suffira juste de comparer cette valeur avec celle de la concentration critique supérieure (CCS) et de la concentration critique inférieures (CCI). L'interprétation se fait comme suit :

1. Sensibles (S), les souches pour lesquelles la CMI est inférieure ou égale à la CCI ;
2. Résistantes (R), les souches vis-à-vis desquelles la CMI est supérieure à la CCS ;
3. De sensibilité intermédiaire (I), les souches vis-à-vis desquelles la CMI est comprise entre les deux concentrations critiques.

Le résultat que nous obtenons dans ce travail c'est la possibilité de donner le profil de sensibilité d'une souche d'*streptococcus* vis à vis de l'érythromycine. On obtient ce résultat par le biais de la prévision de la valeur de la CMI à partir de la valeur du diamètre d'inhibition obtenue par le test de l'antibiogramme.

4.2.2 Application du modèle sur le test avec le Teicoplanine

L'ajustement du modèle de régression donne les résultats suivants :

Call:

```
lm(formula = CMI ~ D)
```

Residuals:

```
      Min      1Q  Median      3Q      Max
-0.61570 -0.39083 -0.06347  0.31018
 1.15681
```

Coefficients:

```
              Estimate Std. Error  t value Pr(>|t|)
(Intercept)  2.37686    0.52544   4.524   6.08e-05 ***
D            -0.10224    0.02857  -3.579  0.000984 ***
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4995 on 37 degrees of freedom
Multiple R-squared: 0.2572, Adjusted R-squared: 0.2371
F-statistic: 12.81 on 1 and 37 DF, p-value: 0.0009843

Et ainsi on obtient l'équation suivante qui nous donne la valeur de la CMI pour le test avec le Teicoplanine.

$$CMI = 2.37686 - 0.10224D \tag{4.3}$$

Le test de signification des paramètres donne des probabilités critiques très inférieures à la valeur par défaut du seuil. Ce qui permet de garder tous les paramètres. En outre on a le diamètre qui explique 25.72% de la variabilité de la CMI, pourcentage significatif. La table d'analyse de la variance donne le résultat suivant :

Analysis of Variance Table

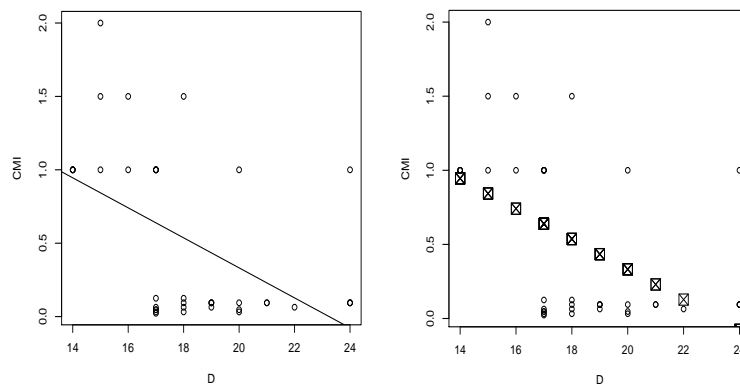
Response: CMI

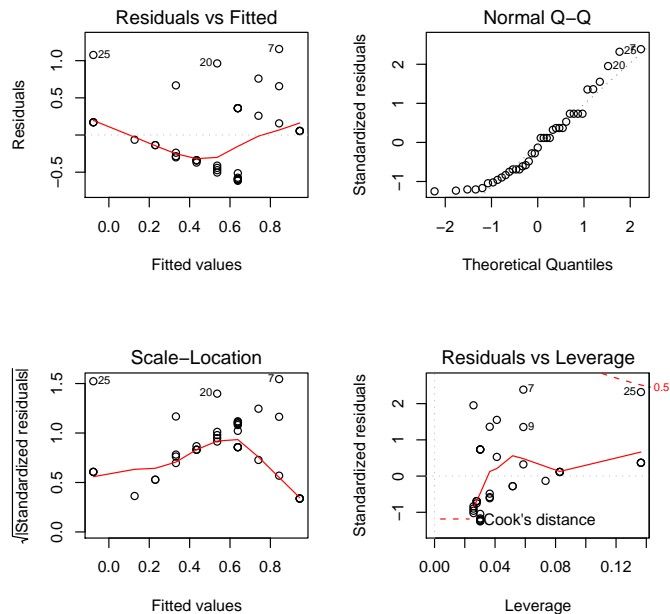
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
D	1	3.1962	3.1962	12.811	0.0009843 ***
Residuals	37	9.2312	0.2495		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Nous remarquons que l'on peut retrouver la valeur du "**Multiple R-squared**" à partir de la cette table. On l'obtient en prenant la somme des parts expliquées sur la somme total. Par exemple sur le tableau suivant on retrouve la valeur comme suit. $R^2 = 3.1962/(3.1962 + 9.2312)$.

Pour la validation du modèle nous allons analyser les graphes ci dessous.





Au vu de ces résultats on en conclut que les hypothèses sont vérifiées.

Pour la prévisions nous donnons l'exemple suivant. Toujours pour une valeur du diamètre récupérée par le biais de l'antibiogramme par diffusion avec le Teicoplanine, on donne la valeur de la CMI correspondante.

```
predict(reg,data.frame(D=26),interval="prediction")
      fit      lwr      upr
1 -0.2815049 -1.401967  0.8389573
```

Avec l'équation du Teicoplanine obtenue on pourra ainsi donner le profil de sensibilité d'une quelconque souche de streptococcus pyogène.

4.2.3 Application du modèle sur le test avec le Levofloxacin

L'ajustement du modèle sur les données donne les résultats suivants :

Call:

```
lm(formula = CMI ~ D)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.57569	-0.36262	-0.01956	0.21310	1.04431

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.63083	0.74014	3.555	0.00106 **
D	-0.09306	0.03856	-2.413	0.02088 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4457 on 37 degrees of freedom
Multiple R-squared: 0.136, Adjusted R-squared: 0.1126
F-statistic: 5.823 on 1 and 37 DF, p-value: 0.02088

Ce listage nous permet de donner l'équation permettant de trouver pour le test avec le levofloxacin la CMI en fonction du diamètre.

$$CMI = 2.63083 - 0.09306D \quad (4.4)$$

La table des variances donne

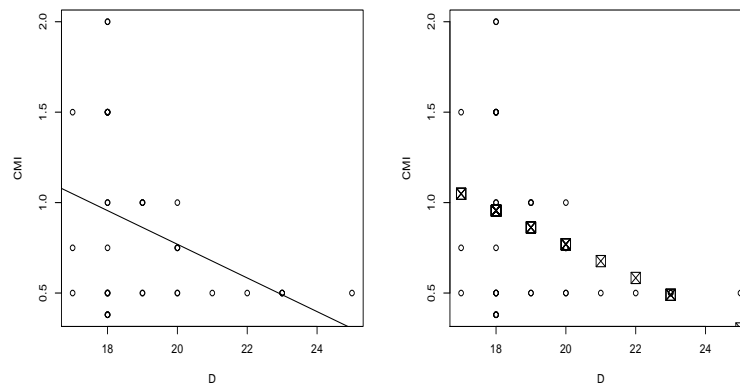
Analysis of Variance Table

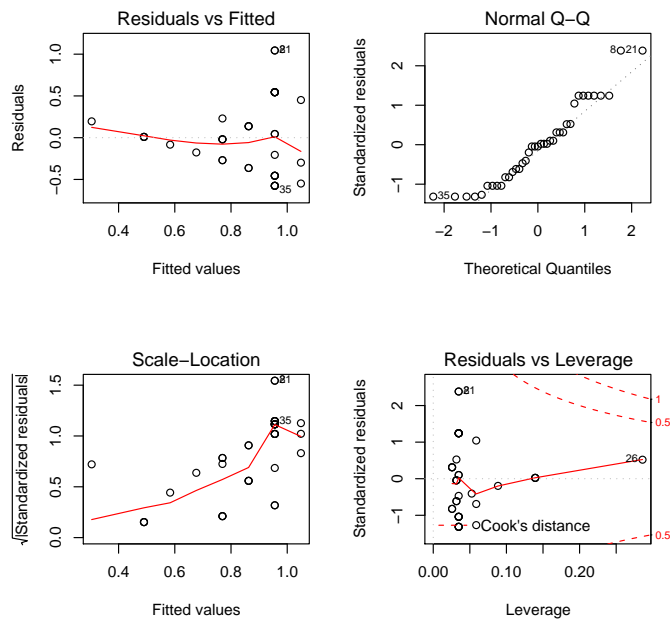
Response: CMI

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
D	1	1.1570	1.1570	5.8233	0.02088 *
Residuals	37	7.3512	0.1987		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' ,

Ici aussi on vérifie les hypothèses par les graphes.





4.3 Echec de l'application du modèle sur la Pénicilline G et le Chloramphénicol

Nous allons dans cette section montrer qu'un modèle de régression n'est pas adapté pour décrire la liaison entre la CMI et le diamètre. Étant donné que pour ces deux antibiotiques, les coefficients de corrélation calculés précédemment sont petits, on peut dire alors qu'il serait faux de considérer que la liaison entre la CMI et le diamètre est linéaire. Pour montrer cela tentons d'ajuster un modèle linéaire sur les données. On obtient pour les deux antibiotiques les résultats suivants :

Pénicilline G

```
Call:
lm(formula = CMI ~ D)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-0.011121 -0.005930  0.001355  0.002546  0.017832
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.0284520  0.0143233   1.986  0.0544 .
D            -0.0004761  0.0004689  -1.015  0.3165
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.006626 on 37 degrees of freedom
Multiple R-squared: 0.02711, Adjusted R-squared: 0.0008137
```

F-statistic: 1.031 on 1 and 37 DF, p-value: 0.3165

Chloramphénicol

Call:

```
lm(formula = CMI ~ D)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.15116	-0.15116	-0.01886	0.11345	3.01422

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.81270	1.08228	3.523	0.00115 **
D	-0.03308	0.04515	-0.733	0.46844

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8157 on 37 degrees of freedom

Multiple R-squared: 0.0143, Adjusted R-squared: -0.01234

F-statistic: 0.5367 on 1 and 37 DF, p-value: 0.4684

Le listage de ces tableaux donne des p-values supérieures à la valeur du seuil, ce qui signifie que les paramètres ne sont pas significatifs. D'où tenter de représenter les données par une droite est totalement faux. Cependant nous disposons des méthodes de transformation pour corriger le modèle.

4.4 Choix du modèle des courbes de concordance

Cette méthode consiste à faire des transformations pour obtenir la relation linéaire. Nous allons rappeler les types de transformation trouvés dans la littérature, et qui ont été cités précédemment.

$$Y = \beta_0 + \beta_1 X \quad (4.5)$$

où Y est le logarithme de la CMI et X est le diamètre ou le carré du diamètre.

4.4.1 Application du modèle sur le test avec la Pénicilline G

En ajustant le modèle sur les données on se retrouve dans la même situation que la première. La valeur de la p-value du test de signification du paramètre est toujours supérieure à la valeur par défaut du seuil. Pour résoudre cela nous tentons de poser comme hypothèse la nullité de la constante de régression. Ce qui signifie que si la valeur du diamètre est nulle, on s'attend donc à une valeur de la CMI nulle. Un diamètre nul veut

dire qu'il n'y a pas d'inhibition observable en d'autre terme au contact de l'antibiotique la souche de streptococcus continu sa croissance. Donc pour un E-test on ne pourra pas récupérer une CMI. Et ainsi on considère le modèle sans constante. En appliquant le modèle sur les données on obtient le listage suivant :

Call:

```
lm(formula = y ~ x - 1)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.49003	-0.35157	0.04024	0.29368	1.26702

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
x	-0.14398	0.00313	-45.99	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5971 on 38 degrees of freedom

Multiple R-squared: 0.9824, Adjusted R-squared: 0.9819

F-statistic: 2115 on 1 and 38 DF, p-value: < 2.2e-16

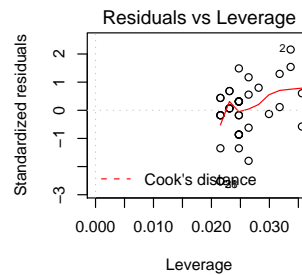
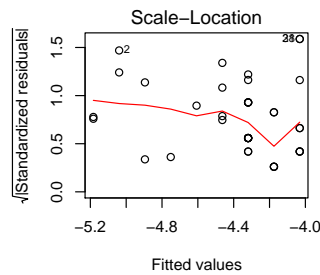
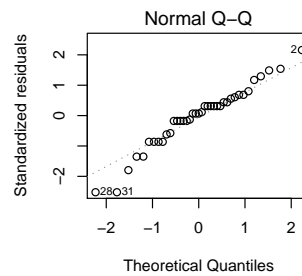
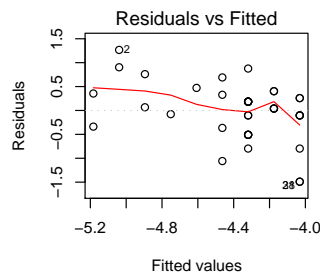
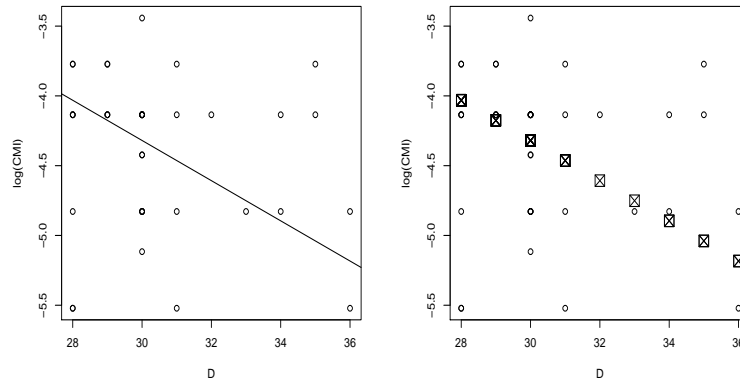
D'après le listage on peut donner l'équation suivante donnant la valeur de la CMI.

$$\ln(CMI) = -0.14398D \quad (4.6)$$

Et pour obtenir directement la valeur de la CMI on peut utiliser l'équation ci-après.

$$CMI = \exp(-0.14398D) \quad (4.7)$$

Nous remarquons que le diamètre explique 98,24% de la variabilité de la CMI. Ce qui est un pourcentage très significatif. Les graphes suivants vérifient les hypothèses du modèle.



4.4.2 Application du modèle sur le test avec le Chloramphénicol

Dans ce cas aussi nous nous sommes rendu compte le modèle sans constante était mieux adapter pour rendre la liaison linéaire.

En appliquant le modèle sur les données nous obtenons les résultats suivants :

Call:

```
lm(formula = y ~ x - 1)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.54792	-0.16462	0.03484	0.21213	0.68366

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
x	0.044324	0.002032	21.81	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3042 on 38 degrees of freedom

Multiple R-squared: 0.926, Adjusted R-squared: 0.9241

F-statistic: 475.6 on 1 and 38 DF, p-value: < 2.2e-16

Nous obtenons l'équation suivante pour le calcul de la valeur CMI.

$$\ln(CMI) = 0.044324D \quad (4.8)$$

On peut utiliser celle-ci aussi qui est une tranformation de la première.

$$CMI = \exp(0.044324D) \quad (4.9)$$

En résumé nous dirons que ce passage par le logarithme nous permet de retrouver le modèle linéaire et ainsi par une petite transformation à faire on retrouve la vraie valeur de la CMI. Et en connaissant cette valeur on pourra déterminer le profil de sensibilité de le streptococcus vis á vis de la Péniciline G et du Chloramphénicol.

Chapitre 5

Profil de sensibilité du *Streptococcus pneumoniae*

Nous disposons d'un échantillon de 9 souches de cette bactérie testée avec les antibiotiques suivants :

Chloramphénicol, Erythromycine, Péniciline G, Teicoplanine.

5.1 Liaison entre la CMI et le diamètre d'inhibition

Caractéristique des données

Nous avons les même caractéristiques pour les données recueillies avec le test sur le streptococcus pyogène. Appliquons un test sur le coefficient de corrélation pour quantifier la liaison linéaire entre les deux variables. Les résultats ci dessous sont obtenus.

Péniciline G

Erythromycine

$\text{cor}(D, \text{CMI}) = -0.7068932$

$\text{p-value} = 0.0332$

$\text{cor}(D, \text{CMI}) = -0.922256$

$\text{p-value} = 0.0003996$

Teicoplanine

Chloramphénicol

$\text{cor}(D, \text{CMI}) = -0.1135924$

$\text{p-value} = 0.771$

$\text{cor}(D, \text{CMI}) = -0.9568533$

$\text{p-value} = 5.266e-05$

Une analyse de ces tableaux nous permet de conclure que l'on peut ajuster un modèle de régression simple sur le test avec la péniciline g, l'erythromycine et le chloramphénicol.

5.2 Modèle de régression simple pour le test avec la Péniciline G, l'Erythromycine et le Chloramphénicol

Pour chacun de ces antibiotiques nous allons donner l'équation de régression qui donne la valeur de la CMI.

Pour la péniciline g nous avons les résultats suivants

Call:

```
lm(formula = CMI ~ D)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.8186	-0.3865	0.1299	0.2332	1.9532

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.38938	1.06582	3.180	0.0155 *
D	-0.10327	0.03906	-2.644	0.0332 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.049 on 7 degrees of freedom

Multiple R-squared: 0.4997, Adjusted R-squared: 0.4282

F-statistic: 6.992 on 1 and 7 DF, p-value: 0.03322

On donne l'équation pour le calcul des valeurs de la CMI

$$CMI = 3.38938 - 0.10327D \quad (5.1)$$

Pour l'erythromycine on donne les résultats ci dessous

Call:

```
lm(formula = CMI ~ D)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.5929	-0.6731	0.2481	0.2481	2.4071

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	9.57838	1.28701	7.442	0.000144 ***
D	-0.30658	0.04857	-6.312	0.000400 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.183 on 7 degrees of freedom
Multiple R-squared: 0.8506, Adjusted R-squared: 0.8292
F-statistic: 39.84 on 1 and 7 DF, p-value: 0.0003996

On donne l'équation

$$CMI = 9.57838 - 0.30658D \quad (5.2)$$

Pour le chloramphénicol on a

Call:

```
lm(formula = CMI ~ D)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.9607	-0.6100	0.2272	0.2407	0.9907

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	11.98089	1.20675	9.928	2.24e-05 ***
D	-0.35067	0.04025	-8.712	5.27e-05 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7487 on 7 degrees of freedom
Multiple R-squared: 0.9156, Adjusted R-squared: 0.9035
F-statistic: 75.91 on 1 and 7 DF, p-value: 5.266e-05

On obtiens la valeur de la CMI en remplaçant la valeur du diamètre dans l'équation suivante

$$CMI = 11.98089 - 0.35067D \quad (5.3)$$

Remarquons que pour le modèle ajusté la part de variabilité expliquée par le diamètre est très importante pour les trois antibiotiques. On 49.97% pour la péniciline g, 85.06% pour l'érythromycine et 91.56% pour le chloramphénicol. En outre on note que les tests sur les paramètres sont très significatifs.

Par les méthodes développés précédemment nous monter la validité du modèle.

5.3 Modèle des courbes de concordance sur le test avec le Teicoplanine

Dans cette section on utilise la transformation des variables pour rendre linéaire la relation. En appliquant la même démarche que sur le test avec le streptococcus pyogène

on obtient les résultats ci dessous.

Call:

```
lm(formula = y ~ x - 1)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.409261	-0.380873	-0.003796	0.283886	0.312274

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
x	-0.014194	0.005084	-2.792	0.0235 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.319 on 8 degrees of freedom

Multiple R-squared: 0.4935, Adjusted R-squared: 0.4302

F-statistic: 7.795 on 1 and 8 DF, p-value: 0.02349

Les équations suivantes nous donne la valeur de la CMI.

$$\ln(CMI) = -0.014194D \quad (5.4)$$

Et celle-ci nous permet d'obtenir la valeur directe de la CMI.

$$CMI = \exp(-0.014194) \quad (5.5)$$

Ces résultats nous permettent de donner le profil de sensibilité d'une souche de streptococcus pneumoniae vis à vis de ces antibiotiques. La détermination du profil se fait par interprétation des valeurs de la CMI estimée.

Chapitre 6

Profil de sensibilité de l'Haemophilus Influenzae

On a un échantillon de 23 souche de cette bactérie testée avec les antibiotiques suivants :

Ampicilline, Lévoﬂoxacine, Chloramphénicol

6.1 Liaison entre la CMI et le diamètre d'inhibition

Caractéristiques des données

L'application du test sur le coefficient de corrélation donne les résultats suivants :

Levoﬂoxacine

Chloramphénicol

$\text{cor}(D, \text{CMI}) = -0.4008414$

$\text{p-value} = 0.05802$

$\text{cor}(D, \text{CMI}) = -0.8363367$

$\text{p-value} = 6.66e-07$

Ampicilline

$\text{cor}(D, \text{CMI}) = -0.9272076$

$\text{p-value} = 2.038e-10$

Au vu de ces tableaux on peut appliquer un modèle de régression simple pour les tests avec l'ampicilline et le chloramphénicol et quant au test avec le lévoﬂoxacine on applique le modèle des courbes de concordance.

6.2 Modèle de régression simple pour le test avec l'Ampicilline et le Chloramphénicol

Dans cette section nous allons ajuster le modèle sur les données pour obtenir les paramètres estimés. Ce qui nous permettra de donner les équations.

Pour le test avec l'ampicilline on obtient les résultats suivants :

Call:

```
lm(formula = CMI ~ D)
```

Residuals:

Min	1Q	Median	3Q	Max
-2.4845	-1.1237	0.2470	0.8188	2.0141

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	13.96307	1.01187	13.80	5.32e-12 ***
D	-0.49857	0.04395	-11.34	2.04e-10 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.37 on 21 degrees of freedom

Multiple R-squared: 0.8597, Adjusted R-squared: 0.853

F-statistic: 128.7 on 1 and 21 DF, p-value: 2.038e-10

Ici aussi on a le test sur les paramètres est très significatif. En outre on donne l'équation ci dessous pour le calcul des valeurs estimées de la CMI.

$$CMI = 13.96307 - 0.49857D \quad (6.1)$$

Pour le test avec le chloramphénicol on a

Call:

```
lm(formula = CMI ~ D)
```

Residuals:

Min	1Q	Median	3Q	Max
-3.3738	-1.2407	-0.1497	0.4419	9.2178

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	18.61401	2.37433	7.840	1.14e-07 ***
D	-0.59159	0.08462	-6.991	6.66e-07 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.387 on 21 degrees of freedom

Multiple R-squared: 0.6995, Adjusted R-squared: 0.6851

F-statistic: 48.87 on 1 and 21 DF, p-value: 6.66e-07

L'équation suivante donne la valeur estimée de la CMI.

$$CMI = 18.61401 - 0.59159D \quad (6.2)$$

La valeur des probabilités critiques est très inférieure à la valeur par défaut du seuil. Ce qui veut dire que les paramètres sont significatifs.

6.3 Modèle des courbes de concordance pour le test avec le Lévofoxacine

Nous appliquons ici la première transformation citée précédemment avec l'hypothèse de nullité de la constante du modèle.

Call:

```
lm(formula = y ~ x - 1)
```

Residuals:

Min	1Q	Median	3Q	Max
-2.1758	-1.3058	0.9918	1.1290	1.8618

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
x	-0.088418	0.009446	-9.361	3.96e-09 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.41 on 22 degrees of freedom

Multiple R-squared: 0.7993, Adjusted R-squared: 0.7902

F-statistic: 87.62 on 1 and 22 DF, p-value: 3.961e-09

Les deux équations ci dessous donne la valeur de la CMI estimée

$$\ln(CMI) = -0.014194D \quad (6.3)$$

Et pour obtenir la valeur directe de la CMI on utilise cette équations qui est une transformation de la première.

$$CMI = \exp(-0.014194) \quad (6.4)$$

La détermination du profil de sensibilité d'une bactérie se fait par le biais de la connaissance de la valeur de la CMI. Cependant on se rencontre qu'il n'est pas toujours possible d'obtenir facilement cette valeur. Les modèles présentés précédemment nous donne la solution en permettant d'estimer cette valeur à partir de la valeur du diamètre d'inhibition, qui peut être obtenue grâce à un antibiogramme. Et ainsi on obtient le profil en comparant la valeur de la CMI avec la valeur des deux concentrations critiques.

Conclusion

Nous avons résumé dans ce mémoire les modèles statistiques d'étude de la sensibilité bactérienne. Il nous avait été demandé de prédire la valeur de la CMI à partir de la valeur du diamètre d'inhibition récupérée par le biais de l'antibiogramme. Il apparaît que la méthode la plus simple pour estimer cette valeur est le modèle de régression simple. Le coefficient de corrélation linéaire est un indicateur permettant de quantifier la liaison linéaire. Si elle est très importante (valeur > 0.4) on pourra alors ajuster un modèle de régression. Cependant ce n'est pas un indicateur fiable dans le cas où il est faible. D'où l'importance de déléguer le calcul de celui-ci au profit d'un test de signification du coefficient de corrélation. En comparant la valeur de la p-value avec la valeur par défaut du seuil on décidera du modèle le plus approprié pour ajuster les données.

En outre il se trouve que des erreurs peuvent être commises dans la manière d'obtention des données. Par exemple une différence entre le temps de récupération des deux variables peut entraîner une disconcordance entre les variables. Et dans ce cas la régression non-paramétrique permet de rendre linéaire la relation par le biais des transformations faites sur les variables.

Enfin il faudrait s'assurer de la validité des modèles ajustés, et pour cela on teste les hypothèses.

Annexes

Instruction et Codes sur R utilisés

Chargement des packages

La plupart des packages qui étaient nécessaires pour ce travail ont été chargés automatiquement dès l'ouverture du logiciel. Cependant pour obtenir les tables de variance il a fallu charger le package suivant :

```
library(car)
```

Importation des données

Nous avons rangé les données dans des répertoires différents. Et ainsi dans un premier temps on charge le répertoire de travail et après on passe à l'importation des données. Pour le travail sur le *Streptococcus pyogène* nous avons utilisé les commandes suivantes :

```
setwd("C:/Documents and Settings/khadidjatou thiam/Bureau/memoir \\  
de dea/donnees/streptococcus pyogène")  
data1=read.table("penicil.csv",header=T,sep=";",dec=",")  
attach(data1)  
data2=read.table("erythom.csv",header=T,sep=";",dec=",")  
attach(data2)  
data3=read.table("teicopla.csv",header=T,sep=";",dec=",")  
attach(data3)  
data4=read.table("chloramph.csv",header=T,sep=";",dec=",")  
attach(data4)  
data5=read.table("levofloxa.csv",header=T,sep=";",dec=",")  
attach(data5)
```

La première commande nous permet de charger le répertoire de travail. Pour le *Streptococcus pneumoniae* on a

```
setwd("C:/Documents and Settings/khadidjatou thiam/Bureau/memoir \\  
de dea/donnees/streptococcus pneumoniae")  
data1=read.table("penici-pneu.csv",header=T,sep=";",dec=",")  
attach(data1)  
data2=read.table("erythrom-pneu.csv",header=T,sep=";",dec=",")  
attach(data2)  
data3=read.table("teico-pneu.csv",header=T,sep=";",dec=",")  
attach(data3)
```



```
data4=read.table("chloram-pneu.csv",header=T,sep=";",dec=",")
attach(data4)
```

On change de répertoire pour pouvoir importer les données sur le test avec le *Streptococcus pneumoniae*. Et enfin pour le test avec l'*Haemophilus Influenzae*, on charge le répertoire suivant.

```
setwd("C:/Documents and Settings/khadidjatou thiam/Bureau/memoir de dea/donnees/h
data1=read.table("amp-hae.csv",header=T,sep=";",dec=",")
attach(data1)
data2=read.table("chloram-hae.csv",header=T,sep=";",dec=",")
attach(data2)
data3=read.table("levo-hae.csv",header=T,sep=";",dec=",")
attach(data3)
```

Notons que l'option `attach()` nous permet de pouvoir travailler avec le nom des colonnes pour appeler les variables. Cela donne comme résultat le nom de chaque colonne du tableau.

Instructions Pour le test sur le coefficient de corrélation nous avons utilisé la fonction,

```
cor.test(x, y, alternative = c("two.sided", "less", "greater"), method = c("pearson
```

par défaut on a la méthode de Pearson. Notons que les autres méthodes Kendall et Spearman sont utilisés dans le cas de test non paramétrique. On affiche les graphes des nuages de points par la fonction,

```
plot(x, y, ...)
```

Et on a les options suivantes pour le modèle de représentation du graphe,

```
"p" for points,
"l" for lines,
"b" for both,
"c" for the lines part alone of "b",
"o" for both 'overplotted',
"h" for 'histogram' like (or 'high-density') vertical lines,
"s" for stair steps,
"S" for other steps, see 'Details' below,
"n" for no plotting.
```

Pour la régression nous avons utilisé la fonction,

```
lm(formula, data, subset, weights, na.action,
    method = "qr", model = TRUE, x = FALSE, y = FALSE, qr = TRUE,
    singular.ok = TRUE, contrasts = NULL, offset, ...)
```

formula représente la formule, data est le data.frame contenant les données. Les autres arguments sont facultatifs. Dans la formule on peut utiliser l'option `-1`, pour indiquer que le modèle est sans constante. Ce qui signifie que c'est un modèle de proportionnalité. Pour avoir les détails sur le modèle on applique la fonction,

```
summary(object, ...)
```

Par exemple on a les commandes suivantes :

```
reg=lm(CMI~D)
summary(reg)
reg=lm(CMI~D-1)
```

Enfin pour les tables de variance on a la fonction,

```
anova(object, ...)
```

Objet représente ce qui a été obtenu en appliquant le modèle sur les données. Par exemple,

```
anova(reg)
```

Bibliographie

[1]Thèse Emma Diatta, Etude de l'effet inoculum et du temps d'incubation sur la sensibilité in vitro des Cocci á Gram positif aux antibiotiques, Juillet 2008.

[2] Thèse Kéne Bougoul Seck, Microméthode d'étude in vitro de la sensibilitéaux antibiotiques des mycoplasmes, des Staphylocoques, et des Entérobactéries, Juillet 2004.

[3] User's Guide, Neo-Sensitas, Susceptibility-Testing.

[4]Thèse Ibrahima Sagna, Macrolides apparentées, Etude prospective sur la sensibilité des bactéries á Gram positif, Juillet 2004 .

[5]Thèse Nouha Diabang, Souche de référence et contrôle in vitro de la sensibilité aux antibiotiques, juillet 2003 .

[6]Thèse Marie Laure Delignette, Méthode de prédiction des aptitudes de croissance des populations de micro-organisme, Mai 1995.

[7].Thèse Sophie Bréand, Etude biométrique de la réponse d'une population bactérienne á une variation défavorable de température ou de ph, Application en microbiologie prévisionnelle alimentaire, Juin 1998

[8]Jérôme Pages, Statistiques Générales pour utilisateurs, 1—Méthodologie.

[9]François Husson et Jérôme Pages, Statistique Générales pour utilisateurs, 2—Exercices et Corrigés .

[10]A. Philippon, Faculté de Médecine Cochin-Port-Royal Université Paris V, Cours de Bactériologie Général, Antibiotique : Etude in vitro, l'antibiogramme.

[11]Faculté de Médecine Lyon-Sud, laboratoire de Microbiologie, UMR CNRS 5558, Bases biologiques de l'antibiothérapie, Evaluation in vitro de la sensibilité des bactéries aux antibiotiques.

[12]Journal of Antimicrobial Chemotherapy : Supplement S1 to volume 48 july 2001. Antimicrobial Susceptibility Testing : a report of the working party on Susceptibility Testing of the British Society for Antimicrobial Chemotherapy.

http : //www.jac.oupjournals.org/content/vol48/Suppl1/

[13]B Berçot Faculté de Médecine U niversité Paris 7 - Denis Diderot, Cours Méthode d'étude de la résistance des bactéries au laboratoire.

[14]Université Paris 13, Laboratoire Analyse, Géométrie, et Application UMR 7539 Cours Analyse Statistique des données.

[15]Etude de la sensibilité aux antibiotiques -Détermination de la CMI.

[16]Vincent Jarlier DI4 Mal, Résistance aux antibiotiques, Rôle du laboratoire de bactériologie.

[17]Pierre Duchesne, Cours Méthode Prévision .
www.dms.umontreal.ca/ duchnesne

[18]Christophe Lalanne, Cours Biostatistique et Statistiques appliquées aux sciences expérimentales. Introduction, 2006 – 2007 .

[19]Jerémie Mattout et Pascal Bessonneau, Régression linéaire et corrélation simple,2006–2007.

[20]Philippe Bess, Université Paul Sabatier Toulouse II, Publication du laboratoire de Statistique et Probabilité, Pratique de la modélisation statistique .
www.lsp.ups – tlse.fr/Bess

[21]Nicolas Turenne INRA, Cours Régression linéaire, 2005.

[22]D.Chessel & A.B Dufour, Fiche Biostatistique : Principe des tests du modèle linéaire.

[23]François Jehl-Monique Chomarat-Michèle Weber-Alain Gérard, Edition Biomérieux, De l'antibiogramme à la prescription.