

*Responsable : Manasé Bezara*

*Personnes ressources :*

* *Taibi Salima (approche globale, données), UniLasalle*
* *Patrice Lepelletier (programmation, package plugin), UniLasalle*

M E T H O D E S E T O U T I L S A P P L I Q U E S.

( M O A - 2 )

Mars 2024

TRAITEMENT DE DONNEES ECONOMETRIQUES

***NB:*** *ce programme est recommandé aux etudiant.e.s qui avaient suivis le module optionnel MOA1- Data Mining* du mois de Fevrier 2024.

*m. bezara@istom.fr, 4 rue Joseph Lakanal ISTOM 49 000*

***Par MBZ***

******

**DEROULEMENT**

|  |  |  |
| --- | --- | --- |
| **Date** | **Matin** | **Après-midi** |
| Mardi  (05/03/23) | Méthode CART ( arbre de decision) | Analyse qualitative, chi-2 |
| Mercredi  (06/03/23) | Data temporelle | *Rcmdr* |
| Jeudi  (07/03/23) | Modelé logistique | Analyse sensorielle |
|  |  |  |

**Etude de cas 1 :**

Construction d’un arbre de régression par la méthode CART ( *Classification And Regression Tree).*

*NB :* On parle d’un arbre de décision dans le ca ou la variable à expliquer est qualitative.

La méthode CART permet d’expliquer et/ou prédire une variable réponse Y en fonction des variables explicatives X. Elle fonctionne avec deux types de variables Y. X peut être, quantitative pure, qualitative pure ou mixte.

Nous présentons ici une application de la méthode CART sur 680 observations de biomasse de maïs (g/m2) obtenues sur 40 sites différents en France, pour 17 années de 1995 à 2011.

L’objectif est de prédire cette biomasse en fonction des 6 variables suivantes :

* *T1 : température moyenne pendant la première partie de la saison de croissance (du jour 1 au jour 50) en Celsius.*
* *T2 : température moyenne pendant la seconde partie de la saison de croissance (du jour 51 au jour 100) en Celsius.*
* *T3 : température moyenne pendant la dernière partie de la saison de croissance ( du jour 101 au jour 150) en Celsius.*
* *RAD1, RAD2, RAD3 : rayonnements moyens en MJ/ m2 pendant les trois mêmes périodes que les températures moyennes.*

Les données sont dsponibles dans le package ZeBook : *maize.data\_MetaModelling*

**Etude de cas 2** (analyse qualitative)

1. Analyse de conformité :

Du point de vue théorique on savait que le population est composée de ¾ de phénotype A et ¼ de phénotype B.

Un échantillon est disponible: ( 32 A, 7 B)

Analyser sa conformité avec la théorie.

1. Etude de liens ( Salaire et Genre).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| SALAIRE | 1000-2000 | 2000-3000 | 3000-4000 | 4000-5000 |  |
| Hommes | **50** | **70** | **110** | **60** |  |
| Femmes | **80** | **75** | **100** | **30** |  |
|  |  |  |  |  |  |

**Etude de cas 3** (perfectionnement ACP)

* + - Package *factoextra, FactoMineR*
    - R-Jeu de données *decathlon2.*

*Sous-données*

- Individus actifs : lignes 1 à 23.

- Variables actives : colonnes 1 à 10.

* + - Validation de l’acp.( graphe éboulis des valeurs propres).
    - Plan factoriel des individus (classés les individus par ordre d’importance en code couleurs).
    - Plan factoriel des variables (définir les axes factoriels)
    - Plan Biplot.

*Remarque : normalement on devrait normaliser (centrer et réduire les données), scale.unit=TRUE*

Objectif : identifier le groupe des meilleurs avec leur spécialité

*Corrélation aux axes Factoriels :*

*Habillage/coloration en fonction de Cos2*

*Constituer trois groupes de variables par K-Means*

*Res.km= kmeans(var$coord,centers=3,nstart =23)*

*+*

***Etude de cas 4***

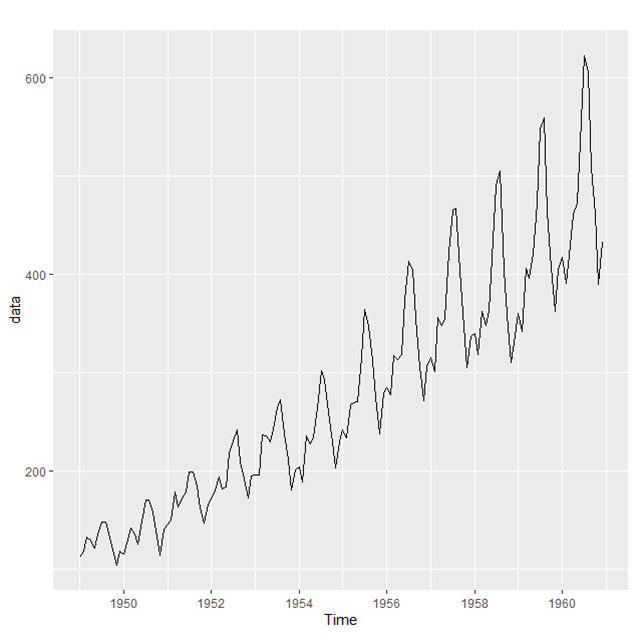
***A -Données temporelles***

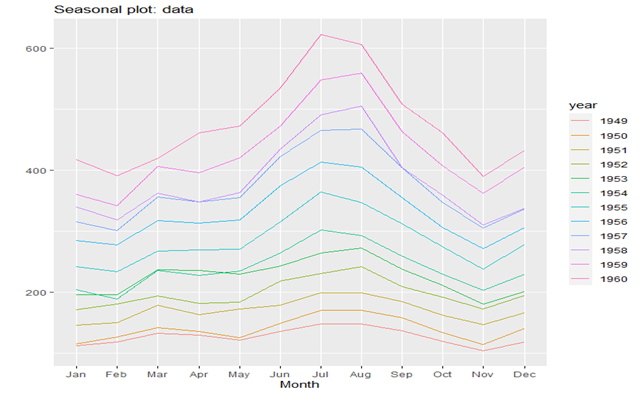
*Observation du nombre mensuel de passagers d’une compagnie aérienne Delta, de 1949-1960.*

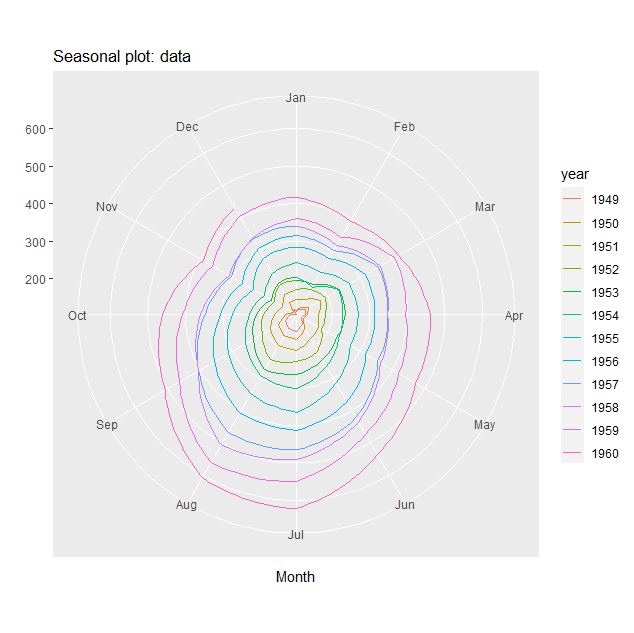
*Représentation et prédiction*

*library(tseries)*

*> data= AirPassengers*

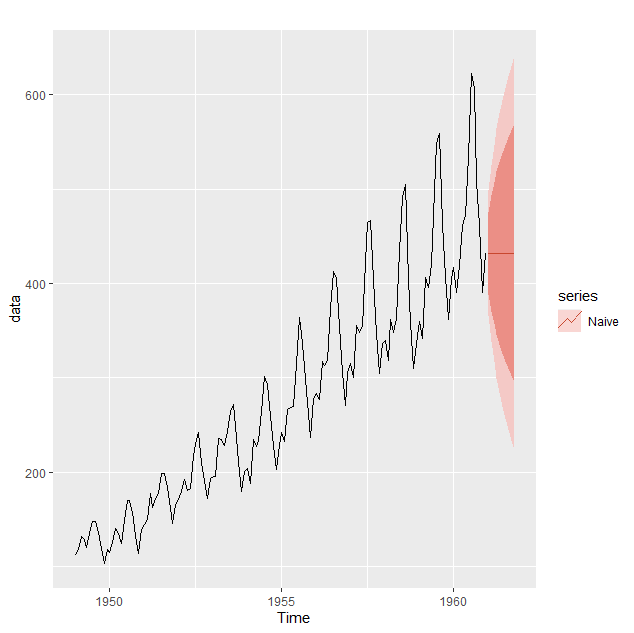
**

**

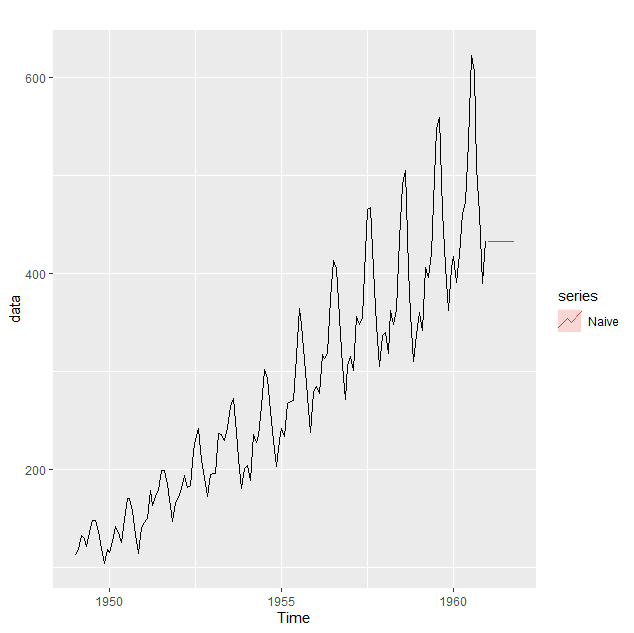


*Prévision naive*

*autoplot(data)+autolayer(naive(data,h=10),series="Naive")*



*autoplot(data)+autolayer(naive(data,h=10),series="Naive", PI = F)*

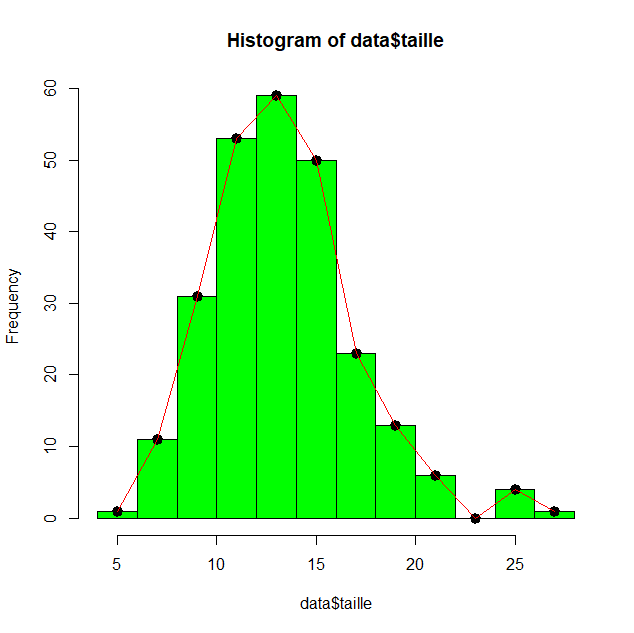


***Etude de cas 5***

*Package : BiostatR*

*Data = Mesures5*

1. *Représentation, histogramme*



***2-****Comparer les espèces selon la taille (Différence est-elle significative) ?*

**Etude de cas 6 (Classification, RCMDR)**

10 Juges expérimentés évaluent 4 types de Gâteau : CG (au Coco), VG ( à la vanille), HG ( au chocolat), FG ( à la Framboise), dans le but de savoir le gâteau à retenir.

Après une épreuve de dégustation, les résultats sont regroupés dans le tableau ci-dessous :

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Jury | CG | VG | HG | FG |
| Mr Henry  Mme Chantal  Mme Léa  Mr Théo  Mr Francis  Mme Isabelle  Mr Loic  Mme Maud  Mr René  Mme Ines | 6.4  6.8  7.2  8.3  8.4  9.1  9.4  7.5  8.5  7 | 8.1  8.2  2.5  3.7  5.4  5.9  4.9  6  7  6 | 8.2  1.3  4  4.9  5.2  5.5  6  7.5  7.5  5.6 | 8  8.5  7.9  8.5  6.3  8.4  7.5  8  6.8  5.9 |

***Mission :*** *établir une classification des dégustateurs en fonction de leur préférence ( notes)*

*1- Graphe (nuage des points (CG, VG), hist (VG), Graphe bandes, ….*

1. *ACP- CAH.*

**Etude de cas 7*: (Dispositif 2)*** Rcmdr et R basique

*Contexte*: un agriculteur se trouve face à un choix entre trois méthodes culturales et quatre type de terrain.

**Problématique**: *la méthode de culture et le type de champ impactent-ils le rendement de manière significative ?*

|  |  |  |  |
| --- | --- | --- | --- |
| Data :  ( à décrire)  Rendement | Méthode | Type de champ | |
| 20 | 1 | 1 |  |
| 7 | 1 | 1 |  |
| 39 | 1 | 2 |  |
| 17 | 1 | 2 |  |
| 34 | 1 | 3 |  |
| 13 | 1 | 3 |  |
| 13 | 1 | 4 |  |
| 5 | 1 | 4 |  |
| 35 | 2 | 1 |  |
| 52 | 2 | 1 |  |
| 30 | 2 | 2 |  |
| 28 | 2 | 2 |  |
| 58 | 2 | 3 |  |
| 73 | 2 | 3 |  |
| 64 | 2 | 4 |  |
| 62 | 3 | 1 |  |
| 44 | 3 | 1 |  |
| 82 | 3 | 2 |  |
| 81 | 3 | 2 |  |
| 69 | 3 | 3 |  |
| 84 | 3 | 3 |  |

Les données correspondent à une expérience où 3 méthodes de culture ont été testées sur quatre types de champs (*même sols, mais expositions différentes*). Les rendements sont mesurés après les moissons.

Note : méthode-3, n'a pu être testée ici sur le type de champ-4 pour insuffisance de graines, et qu'une expérience n'a pu être menée à bout (méthode 2, type de champ 4) à cause d'un orage de grêle.

**Mission :** « *significativité de l’effet de la méthode et du type de champ sur le rendement. »*

Travaux statistiques

1. Résolution R
2. Rcmdr

***Etude de cas 8***

*Traitement des données de comptage*

Position du problème

Le GLM est une approche à prioriser lorsque *la variable d’étude résulte d’un processus de comptage* (*nombre de cas positif, nombre d’œufs pondus, nombre de décès , nombre de vers vivants, nombre de visite d’un site, nombre d’arbres d’une forêt, etc…)*

Pourquoi? ( les hypothèses générales (normalité, sous- dispersion) sur les modèles classiques ne sont pas satisfaites).

Sur le plan pratique, étudions la version la plus populaire ( la regression de poisson) qui consiste à faire intervenir le lien logarithmique (fonction log) avec une structure d’erreur de type Poisson.

Nous profiterons pour revisiter *de manière détaillée* la méthode m-ANOVA

*data Ornstein du package (****car****)*

Interlocks : nombre de postes d’administrateurs partagés avec d’autres grandes entreprises.

Assets : actifs de la compagnie en millions de dollars

Sector : secteur industriel de la compagnie

Nation : nation contrôlant la compagnie. Il s’agit d’une variable catégorielle à 4 modalités.

**Mission :** *« étudier la liaison probable de la variable réponse « interlocks » avec la variable explicative « assets* ».

Régression de Poisson.

On dit qu’une variable aléatoire Y suit une distribution de Poisson de paramètre Lambda, si elle prend pour valeur y = 0 ,1,2,3,… avec une probabilité P dénie par :

La distribution de Poisson est ainsi définie par un seul paramètre : *Lambda*.



*Remarque :* plus le paramètre augmente, plus la loi de poisson se rapproche de la loi normale**. (**Mathématiquement à *30, elle est supposée normale).*

*Formulation du Glm Poisson.*

mod.pois1= glm (interlocks ~ log10(assets), family= « poisson », data =Ornstein)

Condition de validité de la régression de Poisson.

1. Indépendance des réponses : pas de relation entre une réponse et la suivante. *(Exemple : cas des données répétées).* ***OK***
2. Distribution des réponses selon Poisson. (Comparer la distribution des comptages réellement observés avec la distribution théorique de Poisson de paramètre lambda).

*Elle fait partie d’une hypothèse expérimentale. Si elle est rejetée suite à un décalage entre la distribution des comptages observés et théoriques. On devrait changer de structure d’erreur*.

1. Absence de surdispersion

On dit qu’il a surdispersion lorsque la variance observée est supérieure à la moyenne. Sur le plan pratique on vérifie que **le ratio (variance observée/ moyenne) = (residual deviance/ddl** ) soit supérieur à 1.

**Visualisation de la liaison**.

*Package(ggplot2),*

*ggplot(Ornstein, aes(x=log10(assets),y=interlocks))+ geom\_point().*

Remarque : les variables assets étant étendue**, nous allons utiliser une transformation log10**.

ggplot(Ornstein,aes(x=log10(assets),y=interlocks))+ geom\_point().

**NB :** Cette transformation permet aussi de linéariser les relations.

Le modèle correspond à : log10(wy) = bo + b.log10(assets).

**Regression de Poisson**

Procédure (2):

Calculer la moyenne des comptages observés.

mean (Ornsteins$interlocks)

13.58

On simule des comptages théoriques selon une loi de Poisson de paramètre 13.58

*set.seed(1234) permet de simuler avec les mêmes comptages*

theoretic\_count = rpois (248, 13.58).

On incorpore ces comptages théoriques en une base de données.

tc\_df= data.frame(theoretic\_count).

On représente en simultané les comptages observés et les théoriques

ggplot (Ornstein, aes (interlocks))+

geom\_bar(fill= « #1E90FF »)+

geom\_bar(data=tc\_df,aes ( theoretic\_count, fill = « #1E90FF »,alpha=0.5))+

theme\_classic()+

theme (legend.position= « none »).

**Evaluation de la surdispersion**

summary(mod.pois1)

##

## Call:

## glm(formula = interlocks ~ log10(assets), family = "poisson",

## data = Ornstein)

##

## Deviance Residuals:

## Min 1Q Median 3Q Max

## -6.387 -2.468 -0.505 1.127 7.858

##

## Coefficients:

## Estimate Std. Error z value Pr(>|z|)

## (Intercept) -1.04156 0.09037 -11.53 <2e-16 \*\*\*

## log10(assets) 1.05477 0.02386 44.22 <2e-16 \*\*\*

## ---

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

##

## (Dispersion parameter for poisson family taken to be 1)

##

## Null deviance: 3737.0 on 247 degrees of freedom

## **Residual deviance: 1904.7 on 246 degrees** of freedom

## AIC: 2806.7

##

**Etude de cas 9 ( RLM cas matrice mal conditionnée)**

L’objectif principal est de déterminer l’influence de 7 composants sur l’indice d’octane pour un moteur de 12 mélanges.

Les données sont réelles et ont été récoltées dans de bonne condition

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| melange | x1 | x2 | x3 | x4 | x5 | x6 | x7 | Incoc |
| m1 | 0 | 0,23 | 0 | 0 | 0 | 0,74 | 0,03 | 98,7 |
| m2 | 0 | 0,1 | 0 | 0 | 0,12 | 0,74 | 0,04 | 97,8 |
| m3 | 0 | 0 | 0 | 0,1 | 0,12 | 0,74 | 0,04 | 96,6 |
| m4 | 0 | 0,49 | 0 | 0 | 0,12 | 0,37 | 0,02 | 95 |
| m5 | 0 | 0 | 0 | 0,62 | 0,12 | 0,18 | 0,08 | 96,6 |
| m6 | 0 | 0,62 | 0 | 0 | 0 | 0,37 | 0,01 | 91,2 |
| m7 | 0,17 | 0,27 | 0,1 | 0,38 | 0 | 0 | 0,08 | 81,9 |
| m8 | 0,17 | 0,19 | 0,1 | 0,38 | 0,02 | 0,06 | 0,08 | 83,1 |
| m9 | 0,17 | 0,21 | 0,1 | 0,38 | 0 | 0,06 | 0,08 | 82,4 |
| m10 | 0,17 | 0,15 | 0,1 | 0,38 | 0,02 | 0,1 | 0,08 | 83,2 |
| m11 | 0,21 | 0,36 | 0,12 | 0,25 | 0 | 0 | 0,06 | 81,4 |
| m12 | 0 | 0 | 0 | 0,55 | 0 | 0,37 | 0,08 | 88,1 |

Distillation directe

Reformat

Naphta de craquage thermique

Naphta de craquage catalytique

Polymere

Alkylat

Essence naturelle

> data=read.table("clipboard",dec=",",sep="\t",header=T)

> lm1=lm(data$Incoc~data$x1+data$x2+data$x3+data$x4+data$x5+data$x6+data$x7,data=data)

> summary(lm1)

Comment vérifier que la matrice Xt.X est mal conditionnée. Il faut construire la matrice de corrélation avec Incoc

> attach(data)

> data2=cbind(x1,x2,x3,x4,x5,x6,x7,Incoc)

> cor(data2)

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

*Théorie* : Rappel

Partant d’un modèle de régression linéaire classique, Y = aX + b, on prédit l’espérance E(Y) (linéaire)

Du fait de la binarité de Y, la linéarité de E(Y) ne peut pas s’appliquer.

On devrait passer à une forme généralisée g( E(Y)) , ou g est une **fonction de lien** d’où le

*Logit (p) = log( p / (1 – p)).*

Noter que *la fonction de lien* Logit transforme une valeur p comprise entre [ 0, 1 ] à R.

Construction d’une Matrice de confusion ( d’un modèle logistique)

Problème : « lesquelles des variables « âge, comorbidité, détresse respiratoire expliquent le plus une hospitalisation » (petit échantillon de 20 patient.e.s).

Data : Fichier data3covid (Moodle).

**Def** : Le modèle Logit a pour but d’expliquer une variable binaire Y à partir de plusieurs variables explicatives xi (quantitatives).

y = (infecté ou non), logit permet d’Identifier les facteurs liés à l’infection.

*Théoriq*uement, c’est un cas particulier du modèle linéaire généralisé (glm). Mais au lieu de travailler Y, *on va chercher E(Y) = p*, avec p une probabilité de réalisation de Y.

On construit ensuite une fonction lien *logit (p) = log(p / (1 – p)), q*ui transforme p comprise entre [0, 1] en un intervalle de R.

Autrement dit : p = e(logit(p))/ [ 1+ e(logit(p)) ]

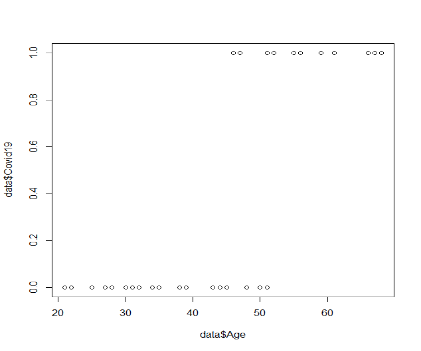
On peut visualiser cette construction de la manière suivante :

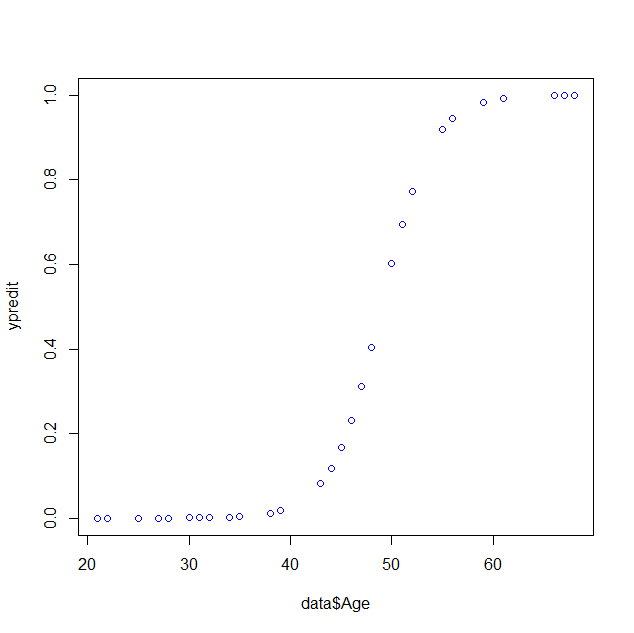
|  |  |
| --- | --- |
| Age | PCR |
| 47 | 1 |
| 35 | 0 |
| 22 | 0 |
| 39 | 0 |
| 30 | 0 |
| 46 | 1 |
| 45 | 0 |
| 45 | 0 |
| 34 | 0 |
| 52 | 1 |
| 66 | 1 |
| 35 | 0 |
| 52 | 1 |
| 61 | 1 |
| 28 | 0 |
| 67 | 1 |
| 25 | 0 |
| 31 | 0 |
| 68 | 1 |
| 47 | 1 |
| 67 | 1 |
| 32 | 0 |
| 38 | 0 |
| 51 | 1 |
| 51 | 0 |
| 27 | 0 |
| 38 | 0 |
| 48 | 0 |
| 52 | 1 |
| 50 | 0 |
| 55 | 1 |
| 56 | 1 |
| 31 | 0 |
| 48 | 0 |
| 50 | 0 |
| 66 | 1 |
| 21 | 0 |
| 44 | 0 |
| 43 | 0 |
| 59 | 1 |

*Commandes : plot( x, y), myreg=glm(y~x, family= binomial ( link=logit)), transcrire logit\_ypredit = -012\*x+5.95*

*ypredit=exp((logit\_ypredit)/(1+exp(logit\_ypredit))*

*points(x,ypredit,col= »red »)*





Courbe de tendance

Avec *data3covid*, construire le modèle explicatif cité plus haut. Identifier les variables qui expliquent le plus l’hospitalisation et annoncer les prédictions avec matrice de confusion**.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Patient** | AGE | Deg Inf | Comorb | RPCR | Hospi |
| Pas1 | 61 | 4 | 3 | Pos | 1 |
| Pas2 | 49 | 1 | 3 | Pos | 0 |
| Pas3 | 75 | 2 | 1 | Pos | 1 |
| Pas5 | 29 | 3 | 1 | Pos | 0 |
| Pas6 | 62 | 1 | 1 | Pos | 0 |
| Pas7 | 35 | 2 | 1 | Pos | 0 |
| Pas8 | 67 | 2 | 2 | Pos | 1 |
| Pas9 | 65 | 3 | 3 | Pos | 1 |
| Pas10 | 47 | 2 | 1 | Pos | 0 |
| Pas11 | 58 | 4 | 4 | Pos | 1 |
| Pas12 | 57 | 3 | 2 | Pos | 0 |
| Pas13 | 59 | 3 | 4 | Pos | 1 |
| Pas14 | 23 | 3 | 2 | Pos | 0 |
| Pas15 | 41 | 3 | 3 | Pos | 0 |
| Pas16 | 74 | 3 | 3 | Pos | 1 |
| Pas17 | 62 | 2 | 3 | Pos | 0 |
| Pas18 | 26 | 2 | 2 | Pos | 1 |
| Pas19 | 55 | 2 | 3 | Pos | 1 |
| Pas20 | 31 | 3 | 4 | Pos | 1 |
| Pas21 | 49 | 2 | 1 | Pos | 0 |

*Remarque :* ***matrice de confusion et significativité.***

*Pour évaluer la qualité du modèle on confronte les valeurs prédites avec les vraies valeurs observées par Y par la matrice de confusion. Pour la significativité des coefficients nous utilisons un test analogue à l’évaluation de la RLM. H_0 :  b_1 = b_2 = \dots = b_J = 0,*

***Etapes essentielles.***

* str(data), summary(data), attach(data)
* modele= glm( Hospi~ AGE+ Deg.Inf+ Comorb, family= binomial)
* summary(modele) ( par les p-values vous pouvez classer les variables par performance)
* Pred.proba= predict (modele, newdata=data, type= « response »)
* Pred.mod= factor (ifelse(Pred.proba<0.5, « Non Hospital », « Hospitalisation »)
* Matconf= table(Hospi,Pred.mod)

Etudes de cas 11 Analyse sensorielle

La perception sensorielle résulte de l’intégration des informations provenant des organes des sens. Pour simplifier, il existe 6 types de sensations.

* **La vision** : forme, couleur, apparence.
* **L’olfaction** : odeurs (sucres, acres, florales, herbacees).
* **La gustation** : sucre/sale, acide/amer, flaveurs (saveurs, parfum en bouche).
* **L’audition** : croustillant, craquant.
* **Le toucher** : (en bouche) texture, température, fondant, moelleux, ferme.
* **La perception trigéminale** : sensations piquantes, rafraichissantes, astringentes.

La notion de **qualité** peut se décomposer en qualité agronomique, commerciale, nutritionnelle et

gustative. Les 4/5 des mesures relatives à la qualité peuvent être effectuées en un temps record grâce à un panel d’instruments, témoins de l’avancement des techniques de mesure.

En revanche, certains critères ne peuvent être mesures que par l’homme et nécessitent un recours à l’analyse sensorielle.

Cette technique impose d’accorder un soin particulier à l’échantillonnage pour garantir la fiabilité des

résultats.

L’objectif est d’obtenir un lot représentatif de ce qui est teste, homogène et dont les

caractéristiques suivantes sont connues : espèce, variété, essai, origine, itinéraire agronomique, date de récolte, itinéraire post récolte et données physico-chimique.

Les échantillons de différentes variétés testées devront être au même stade de maturité (préalablement défini).

Pour assurer l’homogéneité du lot, le tri peut être réalisé sur le stade de maturité (code couleur), le calibre (calibreuse, mesure du poids (melon), utilisation d’une calibrette (pêche), l’aspect, la couleur ou sur les caracteristiques intrinsèques (fermeté, acidité).

L**’analyse sensorielle** est l’examen des proprietés organoleptiques d’un produit par les organes des sens. Selon les objectifs, il est fondamental de bien distinguer les différentes démarches : **descriptive**, **discriminative**, **hédonique**.

On propose à 15 dégustateurs de tester l’intensité de la sensation « de moelleux », entre un biscuit à 5% et un biscuit à 9 %. Chaque consommateur est invité à évaluer le caractère gras en bouche par une note comprise entre 0 et 10.

Les résultats sont consignés dans le tableau.

|  |  |  |
| --- | --- | --- |
| Dégustateur | Biscuit 6% | Biscuit 9% |
| 1 | 4 | 6.5 |
| 2 | 4 | 7 |
| 3 | 7 | 3 |
| 4 | 5 | 5 |
| 5 | 3 | 5 |
| 6 | 2 | 2.5 |
| 7 | 4 | 5 |
| 8 | 5 | 4 |
| 9 | 5.5 | 5.5 |
| 10 | 4 | 4 |
| 11 | 5 | 5 |
| 12 | 5 | 5 |
| 13 | 3 | 6 |
| 14 | 5 | 4 |
| 15 | 4 | 2 |

L’objectif étant de voir s’il existe ou pas de différence significative par rapport au caractère susmentionné entre les deux produits.

Effectuer une évaluation sensorielle des résultats sachant que l’on accepte un risque de 5%.

A-Test Triangulaire (problématique : les produits peuvent-ils être considérés différents ou pas ?)

Objectif : chercher une différentielle sensorielle entre deux produits très proches.

Panel : 12 à 15 sujets (non entrainés).

Expérience : 3 échantillons dont 2 identiques.

Indiquer l’échantillon que vous percevez comme différent.

**Méthodologie**

1. Procédure expérimentale.

3 échantillons de produits provenant de 2 lots sont présentés dans 3 verres ( ou assiettes) marqués par de numéro..

Les ordres de présentation des produits aux dégustateurs sont aléatoires.

***Questionnaire :***

* Parmi les 3 échantillons, lequel percevez-vous différent des deux autres ?

Indiquer son code dans le cadre prévu à cet effet.

?

Code Ech2

Code Ech1

Code Ech3

……………

1. Tableau des réponses :

|  |  |  |
| --- | --- | --- |
| Jury | réponses C | réponses NC |
| Jury 1  Jury 2  Jury 3  Jury 4  Jury 5  Jury 6  Jury 7  Jury 8  Jury 9  Jury 10  Jury 11  Jury 12 | x  x  x  x  x  x  x  x | x  x  x  x |

1. Tableau de contingence des réponses.

|  |  |
| --- | --- |
| Nombre de réponses correctes | Nombre de réponses incorrectes |
| 8 | 4 |

1. Règle de décision du test (table du test triangulaire).

|  |  |
| --- | --- |
| Taille  du Jury | Seuil = 0.05 |
| Test triangulaire |
| 5  6  7  8  9  10  11  12  13  14  15  16  17  18  19  20  21  22  23  24  25  26  27  28 | 4  5  5  6  6  7  7  8  8  9  9  9  10  10  11  11  12  12  12  13  13  14  14  15 |

B- Epreuve d’appariement

1. boissons sont présentées comme témoins à un panel de dégustateurs- étudiants.

Les verres sont ensuite codés afin que les dégustateurs ne puissent pas reconnaître les boissons proposées (de nouveau) dans un ordre aléatoire.

Les 4 boissons sont alors représentées dont 2 parmi les 4 deux fois.

Les 28 dégustateurs doivent alors reconnaître la concordance entre « témoins » et « produits codés ».