



Modèle Linéaire

Analyse de
variance à un
facteur

Présentation
générale

Introduction

Régression et
analyse de variance

Modélisation

Estimation

Analyse de
variance

Equation d'analyse
de variance

Tests en analyse de
variance

Modèle Linéaire

Analyse de variance à un facteur

Thierry Dhorne

Institut Universitaire de Technologie de Vannes
Université de Bretagne Sud

Année Universitaire 2014-2015



Au programme !

Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

- Introduction
- Régression et analyse de variance
- Modélisation
- Estimation

Analyse de variance

- Equation d'analyse de variance
- Tests en analyse de variance

1 Présentation générale

Introduction

Régression et analyse de variance

Modélisation

Estimation

2 Analyse de variance

Equation d'analyse de variance

Tests en analyse de variance



Pourquoi cette méthode ?

Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

Introduction

Régression et analyse de variance

Modélisation

Estimation

Analyse de variance

Equation d'analyse de variance

Tests en analyse de variance

Objectif

l'analyse de variance est conçue pour étudier la structuration

- ▶ d'une variable quantitative
- ▶ selon une variable qualitative

- elle correspond donc à une « espèce » de régression
 - ▶ à « variable explicative qualitative »
- elle est utilisée **pour mettre en évidence** l'effet (éventuel ?) de la variable qualitative sur la variable quantitative



Par qui ?

Modèle Linéaire

Analyse de
variance à un
facteur

Présentation
générale

Introduction

Régression et
analyse de variance

Modélisation

Estimation

Analyse de
variance

Equation d'analyse
de variance

Tests en analyse de
variance



W.S. Gosset alias Student a été le précurseur de l'analyse de la variance



mais c'est Sir R.A. Fisher qui est le véritable père de cette méthode à laquelle il a consacré une partie de son génie. Il a en particulier développé

- ▶ le test qui porte son nom
- ▶ la planification expérimentale
- ▶ la théorie de la randomisation



Lien, causalité ou... différence ?

Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

Introduction

Régression et analyse de variance

Modélisation

Estimation

Analyse de variance

Equation d'analyse de variance

Tests en analyse de variance

- on utilise souvent le mot «effet»
- certains même «influence»
- ➔ qui traduisent une notion de causalité
- ce vocabulaire n'est pas parfaitement adapté
- ★ **car ce n'est pas toujours la causalité qui est étudiée**
- ➔ Y a-t-il un effet du résultat au Baccalauréat sur la moyenne de terminale ? serait une question stupide
- ➔ Y avait-il une différence de moyenne entre ceux qui ont eu le Baccalauréat et ceux qui ne l'ont pas eu ? est une question éventuellement pertinente (pour le formateur en particulier)
- ★ nous employons dans ce cours le mot «effet» issu du contexte expérimental mais le cadre général est plus large



Parlons correctement !

Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

Introduction

Régression et analyse de variance

Modélisation

Estimation

Analyse de variance

Equation d'analyse de variance

Tests en analyse de variance

- il existe un vocabulaire consacré d'origine anglo-saxonne
 - la variable qualitative est souvent appelée *facteur* (*factor*)
 - la variable quantitative est assez souvent appelée réponse (*response*)
 - les différentes valeurs prises par la variable qualitative sont appelées *modalités* (*categories*) ou niveaux (*levels*) lorsque la variable est ordonnée
- ★ on utilise parfois le terme *populations* dans le cas spécifique du modèle à un facteur



Les données

Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

Introduction

Régression et analyse de variance

Modélisation

Estimation

Analyse de variance

Equation d'analyse de variance

Tests en analyse de variance

unités expérimentales	facteur (X)	réponse (Y)
1	A	25
2	A	31
3	A	19
4	A	29
5	A	23
6	B	45
7	B	41
8	B	43
9	B	37
10	B	35
11	C	33
12	C	37
13	C	35
14	C	29
15	C	41
16	D	39
17	D	42
18	D	37
19	D	31
20	D	34



Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

Introduction

Régression et analyse de variance

Modélisation

Estimation

Analyse de variance

Equation d'analyse de variance

Tests en analyse de variance

	modalité A	modalité B	modalité C	modalité D
1	25	6 45	11 33	16 39
2	31	7 41	12 37	17 42
3	19	8 43	13 35	18 37
4	29	9 37	14 29	19 31
5	23	10 35	15 41	20 34



Analyse graphique

Données de base

Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

Introduction

Régression et analyse de variance

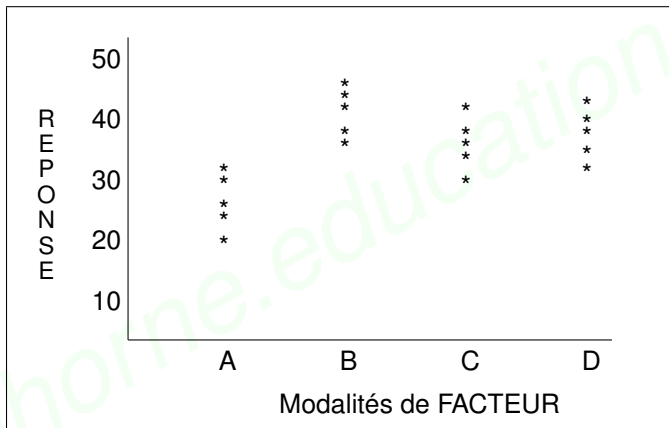
Modélisation

Estimation

Analyse de variance

Equation d'analyse de variance

Tests en analyse de variance





Analyse graphique

Avec deux données manquantes

Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

Introduction

Régression et analyse de variance

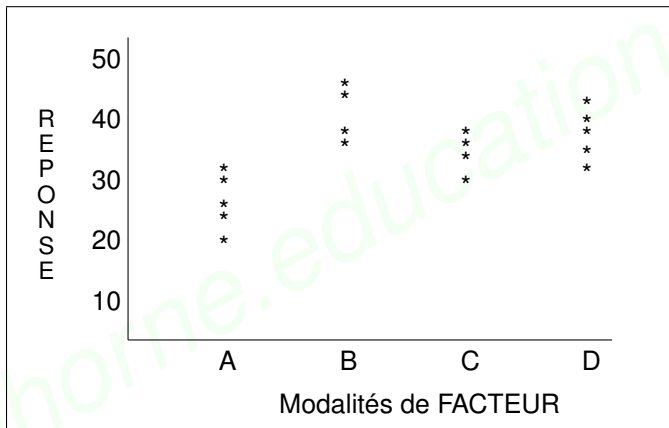
Modélisation

Estimation

Analyse de variance

Equation d'analyse de variance

Tests en analyse de variance





Analyse graphique

Un autre jeu de données

Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

Introduction

Régression et analyse de variance

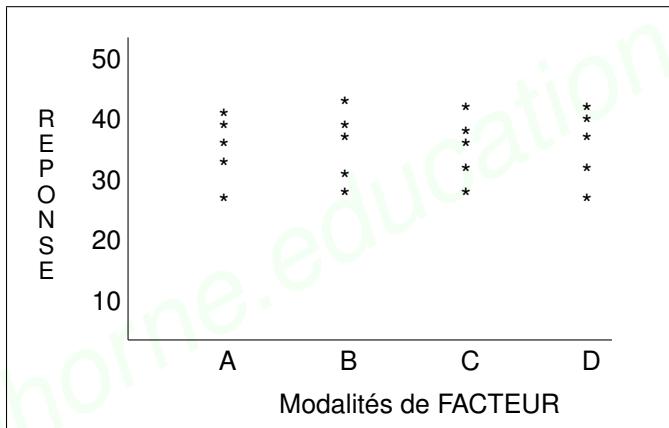
Modélisation

Estimation

Analyse de variance

Equation d'analyse de variance

Tests en analyse de variance





Modèle Linéaire

Analyse de
variance à un
facteur

Présentation
générale

Introduction

Régression et
analyse de variance

Modélisation

Estimation

Analyse de
variance

Equation d'analyse
de variance

Tests en analyse de
variance

Différence ?

Identité ?

Opposition ?

Question posée



Régression et analyse de variance

Le point de vue du technicien

Modèle Linéaire

Analyse de
variance à un
facteur

Présentation
générale

Introduction

Régression et
analyse de variance

Modélisation

Estimation

Analyse de
variance

Equation d'analyse
de variance

Tests en analyse de
variance

- une première différence a été notée
- ➔ les variables prédictrices sont
 - ▶ quantitatives pour les modèles de régression
 - ▶ qualitatives pour les modèles d'analyse de variance



Régression et analyse de variance

Une différence formelle ?

Modèle Linéaire

Analyse de
variance à un
facteur

Présentation
générale

Introduction

Régression et
analyse de variance

Modélisation

Estimation

Analyse de
variance

Equation d'analyse
de variance

Tests en analyse de
variance

- cette distinction entre variables qualitatives et quantitatives est plutôt une question de forme
- on peut faire intervenir des variables qualitatives dans les modèles de régression
 - ▶ le prix d'une propriété dépend de la surface construite et de la surface de terrain (variables quantitatives)
 - ▶ mais aussi du quartier (variable qualitative)
- on peut aussi remplacer une variable quantitative dans le contexte de l'analyse de variance
 - ▶ en la transformant en classes (par exemple)
 - ▶ on étudie l'effet de la température sur la qualité de l'emmental en la fixant successivement à 40°, 50°, 60°



Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

Introduction

Régression et analyse de variance

Modélisation

Estimation

Analyse de variance

Equation d'analyse de variance

Tests en analyse de variance

- d'un point de vue théorique les deux modèles sont deux variantes du même modèle linéaire général

→ ceci fera l'objet d'une leçon ultérieure

Vers le modèle linéaire général

- ce point de vue formel ne correspond pas à toutes les réalités



Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

Introduction

Régression et analyse de variance

Modélisation

Estimation

Analyse de variance

Equation d'analyse de variance

Tests en analyse de variance

- les praticiens font beaucoup plus la distinction
- car les domaines d'utilisation de la régression et de l'analyse de variance sont assez distincts
- les objectifs sont aussi assez différents
- ★ la distinction repose en partie sur l'opposition expériences planifiées ↔ expériences non planifiées



Expériences planifiées et expériences non planifiées

Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

Introduction

Régression et analyse de variance

Modélisation

Estimation

Analyse de variance

Equation d'analyse de variance

Tests en analyse de variance

- on appelle expérience planifiée une expérience pour laquelle on a construit un plan expérimental en choisissant les modalités du facteur et les répétitions

→ par exemple nous avons choisi

- ▶ d'étudier toutes les modalités du facteur «type de baguette»
- ▶ de réaliser 8 répétitions par modalité

- on appelle expérience non planifiée une expérience pour laquelle les modalités sont obtenues «au hasard» de l'échantillonnage

→ par exemple pour étudier

- ▶ l'effet du niveau d'étude
- ▶ sur le temps passé sur un site web



Régression et analyse de variance

Les objectifs

Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

Introduction

Régression et analyse de variance

Modélisation

Estimation

Analyse de variance

Equation d'analyse de variance

Tests en analyse de variance

- en général quand on fait des expériences planifiées
- c'est plutôt pour étudier l'effet de facteurs sur des variables
- l'analyse de variance s'attache donc à mettre en évidence cet effet **quand il existe !**
- souvent cet effet est suspecté
 - ▶ les filles ont-elles des résultats scolaires meilleurs que les garçons ?
- parfois il est espéré
 - ▶ notre nouveau médicament est-il plus efficace que l'ancien ?
- la forme de cet effet est moins importante, au moins dans un premier temps



Régression et analyse de variance

Les objectifs

Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

Introduction

Régression et analyse de variance

Modélisation

Estimation

Analyse de variance

Equation d'analyse de variance

Tests en analyse de variance

- la régression s'intéresse à la forme du lien
 - ▶ comment expliquer au mieux le lien entre le prix d'une voiture et ses caractéristiques techniques ?
- ce qui suppose que l'existence d'un lien est connu
 - ▶ les voitures grosses et puissantes sont plus chères que les petites voitures moins puissantes
- ★ l'analyse de variance peut en réalité mettre en évidence des effets peu décelables
- elle n'a donc pas en général d'objectif de prédiction



Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

Introduction

Régression et analyse de variance

Modélisation

Estimation

Analyse de variance

Equation d'analyse de variance

Tests en analyse de variance

- l'analyse de variance est utilisée avant la régression dans la construction cognitive

1. phase de découverte (\rightarrow analyse exploratoire)
2. phase de sélection-validation (\rightarrow analyse de variance)
3. phase de modélisation fine (\rightarrow régression)

\rightarrow l'analyse de variance trie, hiérarchise des sources de variabilité

★ le test de l'intérêt d'une variable dans un modèle de régression est un test d'analyse de variance

★ en analyse de variance on est souvent «obnubilé» par le test F pour mettre en évidence un effet du facteur

- en régression on s'intéresse à la qualité du modèle, ce qui dépasse largement le simple test F



Régression et analyse de variance

Le point de vue pédagogique

Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

Introduction

Régression et analyse de variance

Modélisation

Estimation

Analyse de variance

Equation d'analyse de variance

Tests en analyse de variance

- l'analyse de variance est présentée dans ce cours (comme dans beaucoup d'autres) après la régression
 - pour faciliter la progression dans la compréhension de l'étudiant
 - car il est reconnu que l'analyse de variance est plus abstraite et donc moins médiatique que la régression
 - elle pose aussi quelques problèmes de technique mathématique plus aigus
 - ▶ elle est naturellement multivariée (**d'où la forte dimensionnalité des problèmes**)
 - ▶ elle est travaillée sur des variables qualitatives (**la culture mathématique des étudiants est très quantitative**)
 - ▶ elle induit des difficultés algébriques (**voir estimabilité**)



Modèle et interprétation

Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

Introduction

Régression et analyse de variance

Modélisation

Estimation

Analyse de variance

Equation d'analyse de variance

Tests en analyse de variance

le modèle d'analyse de variance à un facteur s'écrit

Modèle

$$Y_{ir} = \mu_i + E_{ir}$$

- ▶ Y est la variable à étudier (réponse)
- ▶ les μ_i représentent les espérances de chaque modalité
- ▶ les E_{ir} représentent les écarts au modèle
- ▶ i est l'indice de la modalité du facteur
 $i = 1, \dots, I$ et I est donc le nombre de modalités
- ▶ r est l'indice de répétition à l'intérieur de la modalité
 $r = 1, \dots, n_i$, n_i est donc le nombre de répétitions de la modalité i ,
on a $\sum_{i=1}^I n_i = n$: taille de l'échantillon ou nombre d'observations



- le modèle

$$Y_{ir} = \mu_i + E_{ir}$$

s'accompagne de trois postulats

Postulats du modèle à un facteur

- ▶ $E(E_{ir}) = 0 \Rightarrow E(Y_{ir}) = \mu_i$,
- ▶ $V(E_{ir}) = V(Y_{ir}) = \sigma_E^2$,
- ▶ $C(E_{ir}, E_{i'r' \neq ir}) = 0$,



Postulat d'espérance

Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

Introduction

Régression et analyse de variance

Modélisation

Estimation

Analyse de variance

Equation d'analyse de variance

Tests en analyse de variance

- le postulat $E(Y_{ir}) = \mu_i$,
- indique qu'il n'y a pas de biais dans le modèle
- et ne paraît donc pas très restrictif en soi
- ★ mais attention car on peut avoir oublié
 - ▶ un facteur
 - ▶ ou une variable



Postulat de variance commune

Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

Introduction

Régression et analyse de variance

Modélisation

Estimation

Analyse de variance

Equation d'analyse de variance

Tests en analyse de variance

- le postulat $V(Y_{ir}) = \sigma_E^2$
- implique que la variance résiduelle est constante
- et donc commune à toutes les observations
- ★ ceci est une contrainte assez forte
- qu'il conviendrait de vérifier
- ★ nous indiquerons plus loin comment tester l'hypothèse d'égalité des variances résiduelles
- ce postulat s'appelle souvent postulat d'homoscédasticité
- terme d'origine grecque qui signifie isovariance ou equivariance



Postulat de non covariance

Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

Introduction

Régression et analyse de variance

Modélisation

Estimation

Analyse de variance

Equation d'analyse de variance

Tests en analyse de variance

- le postulat $C(E_{ir}, E_{i'r' \neq ir}) = 0$,
 - implique qu'il n'y a pas de lien de covariance entre les observations
 - donc pas de structure particulière sur les données
 - en particulier pas de structure
 - ▶ temporelle (voir séries chronologiques)
 - ▶ ou spatiale (voir statistique spatiale)
- on peut obtenir la non covariance par un échantillonnage adapté



Estimation des paramètres de l'espérance

Méthode

Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

Introduction

Régression et analyse de variance

Modélisation

Estimation

Analyse de variance

Equation d'analyse de variance

Tests en analyse de variance

- Identité entre l'estimation par la méthode du maximum de vraisemblance gaussien et par celle des moindres carrés
- ➔ il faut donc minimiser la quantité

$$\sum_{i=1}^I \sum_{r=1}^{n_i} (Y_{ir} - \mu_i)^2$$

- ➔ ce minimum est atteint quand chacun des minimums $\sum_{r=1}^{n_i} (Y_{ir} - \mu_i)^2$ est atteint
- ➔ on retrouve le problème classique de l'estimation de l'espérance pour chacune des modalités



Estimation des paramètres de l'espérance

Solution

Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

Introduction

Régression et analyse de variance

Modélisation

Estimation

Analyse de variance

Equation d'analyse de variance

Tests en analyse de variance

- La solution générale est donc

Estimateurs des paramètres μ_i

$$M_i = \frac{1}{n_i} \sum_{r=1}^{n_i} Y_{ir}$$

→ notés

$$\begin{aligned} M_i &= \frac{1}{n_i} Y_{i+}, \\ &= Y_{i.}. \end{aligned}$$



Estimation de σ^2

Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

Introduction

Régression et analyse de variance

Modélisation

Estimation

Analyse de variance

Equation d'analyse de variance

Tests en analyse de variance

- L'estimateur intuitif (et du maximum de vraisemblance gaussien) de la variance résiduelle est

$$S_{MV}^2 = \frac{1}{n} \sum_{i=1}^I \sum_{r=1}^{n_i} (Y_{ir} - Y_i.)^2$$

- Chaque quantité $\sum_{r=1}^{n_i} (Y_{ir} - Y_i.)^2$ a pour espérance $(n_i - 1)\sigma^2$
- l'estimateur du corrigé pour le non biais de σ^2 est donc

Estimateur de la variance résiduelle

$$S^2 = \frac{1}{n - I} \sum_{i=1}^I \sum_{r=1}^{n_i} (Y_{ir} - Y_i.)^2$$

car $\sum_{i=1}^I (n_i - 1) = n - I$



Somme des Carrés des Écarts Totale

Forme

Modèle Linéaire

Analyse de
variance à un
facteur

Présentation
générale

Introduction

Régression et
analyse de variance

Modélisation

Estimation

Analyse de
variance

Equation d'analyse
de variance

Tests en analyse de
variance

- La «variabilité» totale des variables aléatoires observables peut être mesurée par la quantité

$$\sum_{i=1}^I \sum_{r=1}^{n_i} (Y_{ir} - Y_{..})^2$$

→ la notation $Y_{..}$ est utilisée pour être cohérent avec les notations précédentes Y_i .

★ mais, bien sûr, $Y_{..}$ est l'habituel \bar{Y} .

- Cette quantité est appelée Somme des Carrés des Écarts Totale : *SCET*.



Somme des Carrés des Écarts Totale

Décomposition

Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

Introduction

Régression et analyse de variance

Modélisation

Estimation

Analyse de variance

Equation d'analyse de variance

Tests en analyse de variance

On peut décomposer cette *SCET*

$$\begin{aligned}\sum_{i=1}^I \sum_{r=1}^{n_i} (Y_{ir} - Y_{..})^2 &= \sum_{i=1}^I \sum_{r=1}^{n_i} (Y_{ir} - Y_i + Y_i - Y_{..})^2 \\ &= \sum_{i=1}^I \sum_{r=1}^{n_i} (Y_{ir} - Y_i)^2 + \sum_{i=1}^I \sum_{r=1}^{n_i} (Y_i - Y_{..})^2 \\ &\quad + 2 \sum_{i=1}^I \sum_{r=1}^{n_i} (Y_{ir} - Y_i)(Y_i - Y_{..})\end{aligned}$$



Somme des Carrés des Écarts Totale

Annulation du double produit

Modèle Linéaire

Analyse de
variance à un
facteur

Présentation
générale

Introduction

Régression et
analyse de variance

Modélisation

Estimation

Analyse de
variance

Equation d'analyse
de variance

Tests en analyse de
variance

$$\begin{aligned} \sum_{i=1}^I \sum_{r=1}^{n_i} (Y_{ir} - Y_{..})^2 &= \sum_{i=1}^I \sum_{r=1}^{n_i} (Y_{ir} - Y_{i.})^2 + \sum_{i=1}^I (Y_{i.} - Y_{..})^2 \sum_{r=1}^{n_i} 1 \\ &\quad + 2 \sum_{i=1}^I (Y_{i.} - Y_{..}) \sum_{r=1}^{n_i} (Y_{ir} - Y_{i.}) \\ &= \sum_{i=1}^I \sum_{r=1}^{n_i} (Y_{ir} - Y_{i.})^2 + \sum_{i=1}^I n_i (Y_{i.} - Y_{..})^2 + 0. \\ &= \sum_{i=1}^I \sum_{r=1}^{n_i} (Y_{ir} - Y_{i.})^2 + \sum_{i=1}^I n_i (Y_{i.} - Y_{..})^2 \end{aligned}$$



Équation d'analyse de variance

Forme

Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

Introduction

Régression et analyse de variance

Modélisation

Estimation

Analyse de variance

Equation d'analyse de variance

Tests en analyse de variance

- Cette équation

$$\begin{aligned} \sum_{i=1}^I \sum_{r=1}^{n_i} (Y_{ir} - Y_{..})^2 &= \sum_{i=1}^I n_i (Y_{i.} - Y_{..})^2 + \sum_{i=1}^I \sum_{r=1}^{n_i} (Y_{ir} - Y_{i.})^2 \\ \text{SCET} &= \text{SCE Inter} + \text{SCE intra} \end{aligned}$$

- correspond au théorème de Pythagore
- est un cas particulier de l'équation d'analyse de variance caractéristique du modèle linéaire

$$\text{SCET} = \text{SCEF} + \text{SCER}$$



Équation d'analyse de variance

Interprétation

Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

Introduction

Régression et analyse de variance

Modélisation

Estimation

Analyse de variance

Equation d'analyse de variance

Tests en analyse de variance

- La somme des carrés des écarts intra-population mesure la dispersion des variables aléatoires observables à l'intérieur des populations c'est-à-dire des modalités du facteur
 - la somme des carrés des écarts inter populations mesure la dispersion des variables aléatoires entre les populations (en plus de la dispersion à l'intérieur)
 - on peut donc utiliser les abréviations $SCEI$ et $SCEi$, à l'instar de ce qui est fait en analyse discriminante
- ★ on utilise parfois y compris dans les ouvrages de langue française, les termes between (abrégé en B) et within (abrégé en W)



Équation d'analyse de variance

Forme générale

Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

Introduction

Régression et analyse de variance

Modélisation

Estimation

Analyse de variance

Equation d'analyse de variance

Tests en analyse de variance

Equation d'analyse de variance - Forme analytique

$$\sum_{i=1}^I \sum_{r=1}^{n_i} (Y_{ir} - Y_{..})^2 = \sum_{i=1}^I n_i (Y_{i.} - Y_{..})^2 + \sum_{i=1}^I \sum_{r=1}^{n_i} (Y_{ir} - Y_{i.})^2$$

- on peut facilement calculer les degrés de liberté associés à chacun des termes
 $n - 1$, $I - 1$ et $\sum_{i=1}^I (n_i - 1) = n - I$
- obtient donc la

Forme synoptique et degrés de liberté associés

$$\begin{aligned} SCET &= SCEF + SCER \\ n - 1 &= I - 1 + n - I \end{aligned}$$



Lien avec l'estimation

Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

Introduction

Régression et analyse de variance

Modélisation

Estimation

Analyse de variance

Equation d'analyse de variance

Tests en analyse de variance

- la somme de carrés des écarts résiduelle est la somme de carrés utilisée pour estimer la variance résiduelle

Estimateur de la variance résiduelle et SCER

$$S^2 = \frac{SCER}{n - I}$$

- cette quantité est appelée carré moyen résiduel : *CMR*
- ★ par analogie on appelle carré moyen factoriel le rapport d'une somme de carrés factorielle au nombre de degrés de liberté associé
- dans le modèle à un facteur

$$CMF = \frac{SCEF}{I - 1}$$



Test des différences entre espérances

Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

Introduction

Régression et analyse de variance

Modélisation

Estimation

Analyse de variance

Equation d'analyse de variance

Tests en analyse de variance

- on l'appelle souvent test de l'effet du facteur
- car les modèles d'analyse de variance sont utilisés avant tout pour mettre en évidence l'effet de facteurs
- ★ la prédiction est assez rare
- le test de la différence des μ_i est donc toujours le premier point de l'analyse et souvent le seul
- tous les logiciels d'analyse de variance mettent en avant le test d'analyse de variance



- le test correspond à l'alternative suivante

- ▶ $H_0 : \forall i, i' \in \{1, \dots, I\}, \mu_i = \mu_{i'}$,

- ▶ $H_1 : \exists i, i' \in \{1, \dots, I\}, \mu_i \neq \mu_{i'}$,

- ➔ souvent exprimée de la manière suivante

- ▶ H_0 : il n'y a pas d'effet du facteur

- ▶ H_1 : il y a effet du facteur



Statistique de test - I

Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

- Introduction
- Régression et analyse de variance
- Modélisation
- Estimation

Analyse de variance

- Equation d'analyse de variance
- Tests en analyse de variance

- il n'existe pas de statistique optimale pour ce test d'hypothèses quand il y a plus de 2 modalités du facteur

- ★ rappelons qu'il en existe un pour 2 modalités dans le cas gaussien

- il s'agit du test de Student étudié précédemment

- ce test est uniformément plus puissant



Statistique de test - II

Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

Introduction

Régression et analyse de variance

Modélisation

Estimation

Analyse de variance

Equation d'analyse de variance

Tests en analyse de variance

- l'étude de la puissance des tests est assez complexe
 - ▶ car l'hypothèse critique est multiforme
 - ▶ donc la manière de la prospecter ne peut pas être définie de manière unique
- ➔ il a donc fleuri au cours du siècle dernier un grand nombre de tests
 - on se limite dans cette leçon au test de Fisher
 - ➔ qui est facilement généralisable au cas de plusieurs facteurs
 - ★ dans la leçon suivante on étudiera d'autres solutions
 - ➔ comparaison multiple de moyennes



Statistique de test - III

Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

Introduction
Régression et analyse de variance
Modélisation
Estimation

Analyse de variance

Equation d'analyse de variance
Tests en analyse de variance

- la statistique de test est construite sur le rapport
 - ▶ d'une quantité mesurant l'effet éventuel des facteurs : le carré moyen factoriel
 - ▶ à une quantité mesurant la variance résiduelle : le carré moyen résiduel

Statistique de test

$$d = \frac{CMF}{CMR} = \frac{SCEF/(I - 1)}{SCER/(n - I)}$$



Postulat distributionnel

Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

Introduction

Régression et analyse de variance

Modélisation

Estimation

Analyse de variance

Equation d'analyse de variance

Tests en analyse de variance

- un postulat supplémentaire est nécessaire pour obtenir la loi de la statistique de test

Postulat gaussien

les Y_{ir} sont des variables aléatoires gaussiennes

$$Y_{ir} \rightsquigarrow \mathcal{G}(\mu_i, \sigma^2)$$

- ➔ ce postulat s'ajoute aux trois précédents mentionnés plus haut
- ★ il peut en réalité être légèrement assoupli
- ➔ loi à symétrie radiale



Loi de la *SCER*

Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

Introduction

Régression et analyse de variance

Modélisation

Estimation

Analyse de variance

Equation d'analyse de variance

Tests en analyse de variance

→ la *SCER* est une somme de carrés de n gaussiennes centrées de variance σ^2

★ équivalentes à $n - I$ gaussiennes centrées indépendantes de variance σ^2

→ si on divise chaque variables par σ (normalisation), on obtient donc

$$SCER/\sigma^2 \rightsquigarrow \chi_{n-I}^2$$

⇓

Loi de la somme de carrés résiduelle

$$SCER \rightsquigarrow \sigma^2 \chi_{n-I}^2$$



Décomposition théorique de la *SCEF* - I

Modèle Linéaire

Analyse de
variance à un
facteur

Présentation
générale

Introduction

Régression et
analyse de variance

Modélisation

Estimation

Analyse de
variance

Equation d'analyse
de variance

Tests en analyse de
variance

- la somme de carrés des écarts factoriels est égale à

$$\begin{aligned} SCEF &= \sum_{i=1}^I \sum_{r=1}^{n_i} (Y_{i.} - Y_{..})^2, \\ &= \sum_{i=1}^I \sum_{r=1}^{n_i} (\mu_i + E_{i.} - \bar{\mu} - E_{..})^2 \\ &= \sum_{i=1}^I \sum_{r=1}^{n_i} (\mu_i - \bar{\mu} + E_{i.} - E_{..})^2 \end{aligned}$$



Décomposition théorique de la *SCEF* - II

Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

Introduction

Régression et analyse de variance

Modélisation

Estimation

Analyse de variance

Equation d'analyse de variance

Tests en analyse de variance

- la *SCEF* peut être conceptuellement décomposée en
 - ▶ une partie dépendant des effets des facteurs
 - ▶ une autre partie strictement résiduelle
- ★ la *SCEF* est donc une somme de carrés à la fois
 - ▶ structurelle (effet des espérances)
 - ▶ résiduelle (effet du bruit)
- il existe un cas limite pour lequel on peut facilement aller plus loin
- lorsque H_0 est vrai, c'est-à-dire lorsque les μ_i sont tous égaux



Loi de la *SCEF* sous H_0

Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

Introduction

Régression et analyse de variance

Modélisation

Estimation

Analyse de variance

Equation d'analyse de variance

Tests en analyse de variance

- lorsque les μ_i sont tous égaux, les premiers termes des carrés sont nuls

$$\begin{aligned} SCEF &= \sum_{i=1}^I \sum_{r=1}^{n_i} (\mu_i - \bar{\mu} + E_{i.} - E_{..})^2 \\ &= \sum_{i=1}^I \sum_{r=1}^{n_i} (E_{i.} - E_{..})^2 \end{aligned}$$

- ➔ la *SCEF* ne contient plus que du «bruit» résiduel
- ★ le carré moyen factoriel est un (nouvel) estimateur non biaisé de σ^2 noté S_F^2
- ➔ ceci est exploité par la statistique de test



Loi de la statistique de test sous H_0

Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

Introduction

Régression et analyse de variance

Modélisation

Estimation

Analyse de variance

Equation d'analyse de variance

Tests en analyse de variance

- sous l'hypothèse H_0 , la statistique vaut donc

$$\frac{S_F^2}{S^2} = \frac{S_F^2/\sigma^2}{S^2/\sigma^2}$$

- le numérateur et le dénominateur suivent des χ^2 respectivement à $I - 1$ et $n - I$ degrés de liberté

→ la statistique de test suit donc une loi de Fisher : $F(I - 1, n - I)$

Loi de la statistique de test sous H_0

$$\frac{CMF}{CMR} \rightsquigarrow \mathcal{F}_{I-1, n-I}$$



Loi de la statistique de test sous H_0

Etude graphique

Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

Introduction

Régression et analyse de variance

Modélisation

Estimation

Analyse de variance

Equation d'analyse de variance

Tests en analyse de variance

- bien que le problème des tests d'hypothèses soit supposé connu
 - il importe de disposer d'une représentation de la stratégie de test
- ceci permet de lever des difficultés classiques
 - ▶ méconnaissance de la loi de Fisher
 - ▶ en particulier de ses premiers moments
 - ▶ forme de la région de rejet
- nous utilisons dans la suite une loi de Fisher correspondant au test des différences de poids entre baguettes
 - c'est-à-dire un $\mathcal{F}_{5,7}$



Loi de la statistique de test sous H_0

Etude graphique

Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

Introduction

Régression et analyse de variance

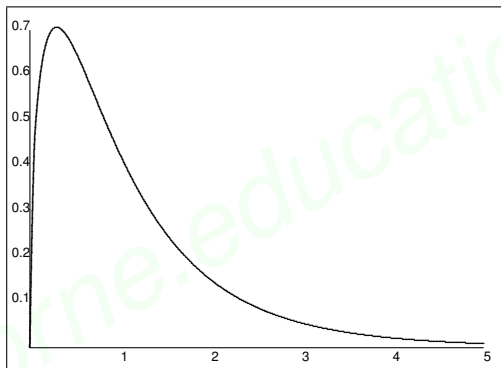
Modélisation

Estimation

Analyse de variance

Equation d'analyse de variance

Tests en analyse de variance





- La région critique est donc naturellement caractérisée par une inégalité de la forme :

$$\frac{CMF}{CMR} > f_{I-1, n-I, 1-\alpha}$$

- Cette région est définie strictement une fois le risque de première espèce α fixé.



Normalisation de la région critique

Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

Introduction

Régression et analyse de variance

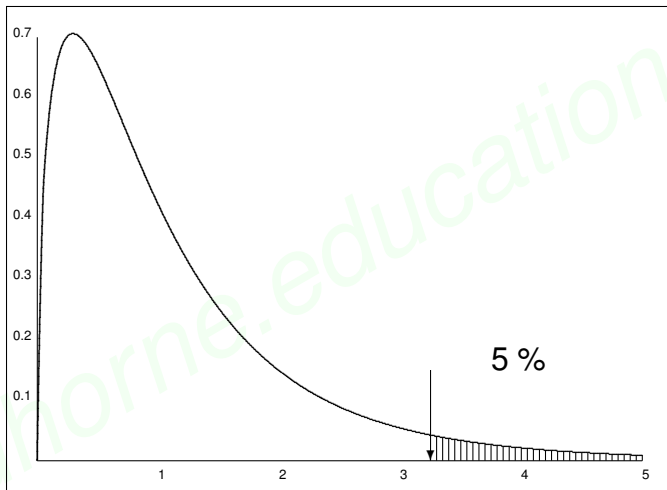
Modélisation

Estimation

Analyse de variance

Equation d'analyse de variance

Tests en analyse de variance





Construction des régions d'acceptation et de rejet

Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

Introduction

Régression et analyse de variance

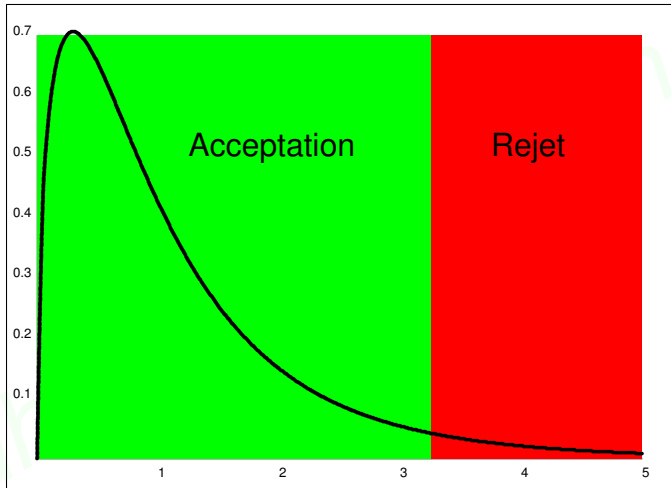
Modélisation

Estimation

Analyse de variance

Equation d'analyse de variance

Tests en analyse de variance





Etude de cas - I

Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

Introduction

Régression et analyse de variance

Modélisation

Estimation

Analyse de variance

Equation d'analyse de variance

Tests en analyse de variance

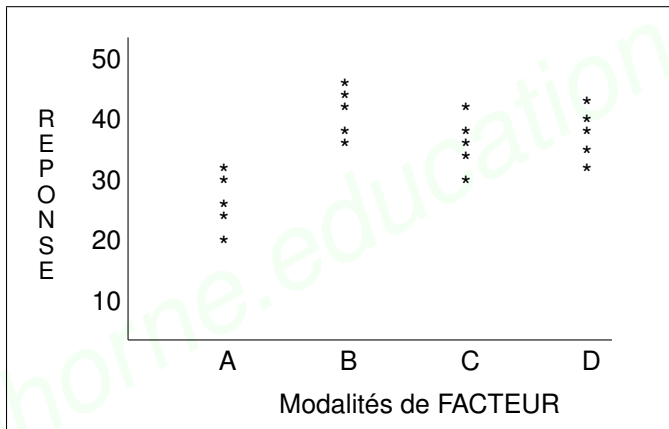




Table d'analyse de variance - I

Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

Introduction

Régression et analyse de variance

Modélisation

Estimation

Analyse de variance

Equation d'analyse de variance

Tests en analyse de variance

```
> summary(aov(Reponse~Facteur,aov.don))
```

```
MODEL REPONSES = FACTEUR ;
```

	Df	Sum of Sq	Mean Sq	F Value	Pr (F)
Facteur	3	599.0	199.6667	10.20009	0.0005372857
Residuals	16	313.2	19.5750		



Décision - I

Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

Introduction

Régression et analyse de variance

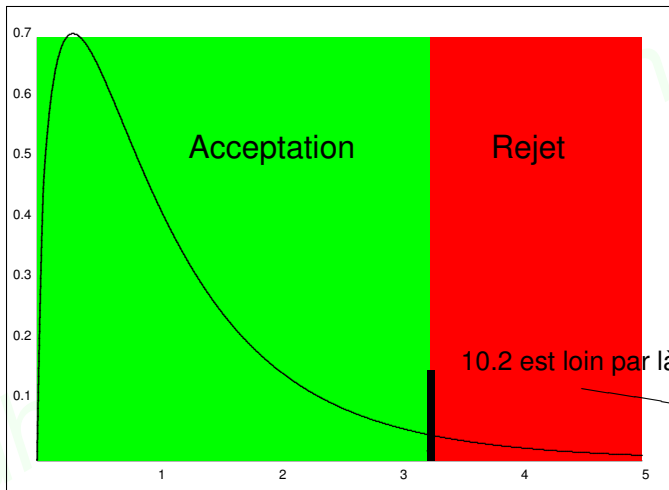
Modélisation

Estimation

Analyse de variance

Equation d'analyse de variance

Tests en analyse de variance





Etude de cas - II

Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

- Introduction
- Régression et analyse de variance
- Modélisation
- Estimation

Analyse de variance

- Equation d'analyse de variance
- Tests en analyse de variance

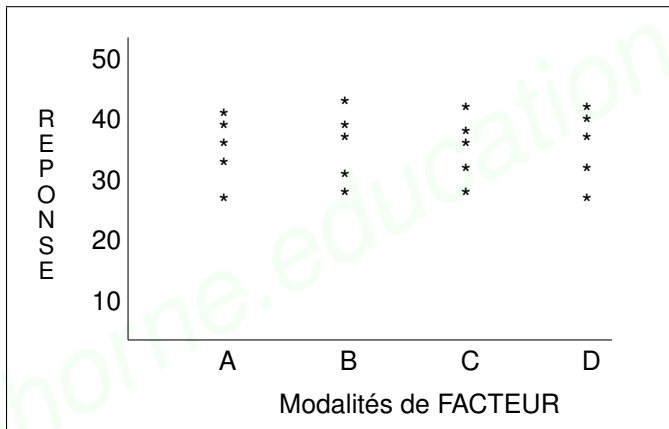




Table d'analyse de variance - II

Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

Introduction

Régression et analyse de variance

Modélisation

Estimation

Analyse de variance

Equation d'analyse de variance

Tests en analyse de variance

```
> summary(aov(Reponse~Facteur,aov1.don))
```

	Df	Sum of Sq	Mean Sq	F Value	Pr(F)
Facteur	3	0.8	0.26667	0.00799012	0.9989763
Residuals	16	534.0	33.37500		



Décision - II

Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

Introduction

Régression et analyse de variance

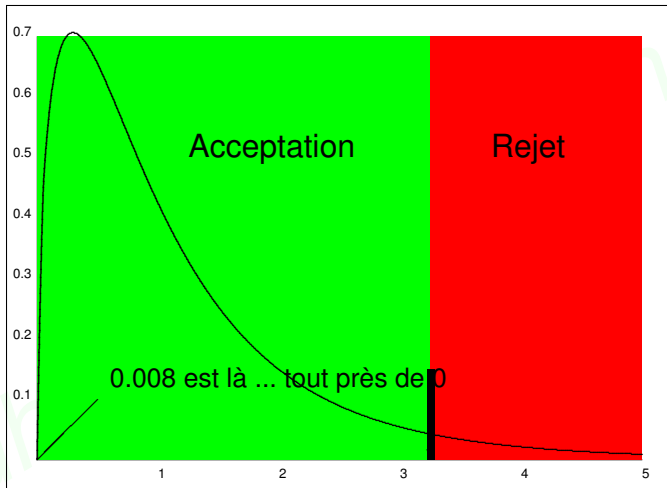
Modélisation

Estimation

Analyse de variance

Equation d'analyse de variance

Tests en analyse de variance





Puissance du test - I

Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

Introduction

Régression et analyse de variance

Modélisation

Estimation

Analyse de variance

Equation d'analyse de variance

Tests en analyse de variance

- La forme de la fonction de puissance du test F est donnée par les quantiles de la statistique de décision sous l'hypothèse critique.
- La $SCEF$ n'est plus une somme de carrés de gaussiennes centrées mais décentrées.
- La somme des carrés des espérances des gaussiennes vaut dans le cas présent :

$$\sum_{i=1}^I \sum_{r=1}^{n_i} (\mu_i - \bar{\mu})^2$$



Puissance du test - II

Modèle Linéaire

Analyse de
variance à un
facteur

Présentation
générale

Introduction

Régression et
analyse de variance

Modélisation

Estimation

Analyse de
variance

Equation d'analyse
de variance

Tests en analyse de
variance

→ La statistique de test est alors égale à :

$$\begin{aligned}d &= \frac{\sum_{i=1}^I \sum_{r=1}^{n_i} (\mu_i - \bar{\mu} + E_{i.} - E_{..})^2 / (I - 1)}{S^2}, \\ &= \frac{\sum_{i=1}^I \sum_{r=1}^{n_i} (\mu_i - \bar{\mu} + E_{i.} - E_{..})^2 / \sigma^2 (I - 1)}{S^2 / \sigma^2}.\end{aligned}$$



Puissance du test - III

Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

Introduction

Régression et analyse de variance

Modélisation

Estimation

Analyse de variance

Equation d'analyse de variance

Tests en analyse de variance

→ C'est la loi d'un χ^2 décentré :

$$\chi_{I-1, \lambda}^2,$$

où le paramètre de non centralité est

$$\lambda = \sum_{i=1}^I \sum_{r=1}^{n_i} (\mu_i - \bar{\mu})^2 / \sigma^2.$$

→ De la même façon, le rapport CMF/CMR est distribué selon un F décentré :

$$F_{I-1, n-I, \lambda},$$

de même paramètre de non centralité.



Loi de Fisher décentrée

Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

Introduction

Régression et analyse de variance

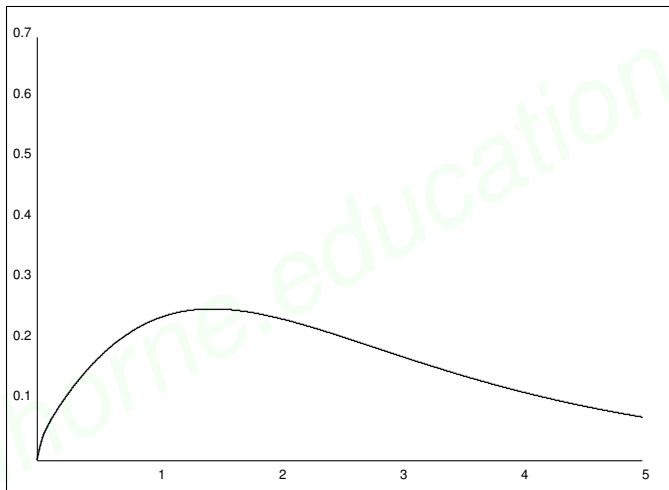
Modélisation

Estimation

Analyse de variance

Equation d'analyse de variance

Tests en analyse de variance





Calcul de la puissance - I

Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

Introduction

Régression et analyse de variance

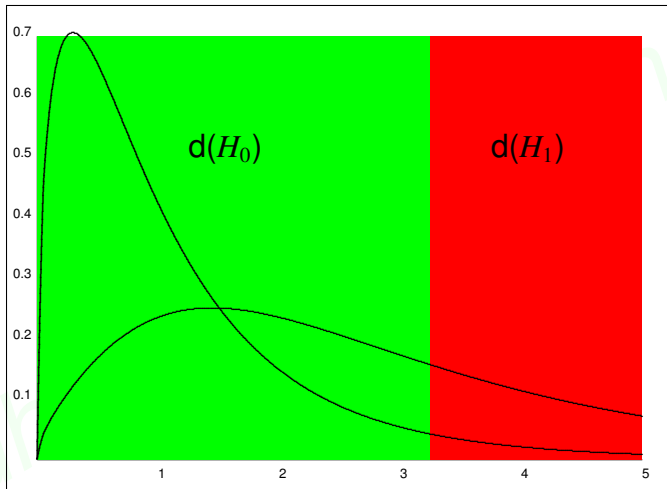
Modélisation

Estimation

Analyse de variance

Equation d'analyse de variance

Tests en analyse de variance





Calcul de la puissance - II

Modèle Linéaire

Analyse de variance à un facteur

Présentation générale

Introduction

Régression et analyse de variance

Modélisation

Estimation

Analyse de variance

Equation d'analyse de variance

Tests en analyse de variance

