STT 3410

Plans et analyses d'expériences

Christian Léger
Université de Montréal

Objectifs

Objectifs généraux:

- L'étudiant prendra plaisir à la statistique.
- L'étudiant maîtrisera les aspects essentiels, de même que les limites, de la planification d'expériences ainsi que de la théorie et la pratique de l'analyse statistique de telles expériences.

Objectifs particuliers:

À la fin du cours, l'étudiant devra être en mesure de:

- Comprendre et utiliser la méthode des moindres carrés dans les plans d'expérience;
- Comprendre les dangers associés à la méthode des moindres carrés;
- Juger du bien-fondé des hypothèses émises;
- Appliquer la méthodologie à des jeux de données en utilisant des logiciels appropriés, notamment SAS;

Objectifs

- Démontrer des résultats liés à la théorie;
- Appliquer la méthodologie à des plans autres que ceux vus en classe.

Contenu

- Introduction;
- Modèle à un facteur;
- Comparaisons multiples;
- Diagnostiques;
- Plans factoriels;
- Plans à mesures répétées;
- Plans incomplets;
- ...

Activités d'enseignement et d'apprentissage et évaluation

- Cours magistraux;
- Devoirs (4 à 6);
- La note sera déterminée à partir de la pondération suivante: 20% Devoirs + 35% Intra (mercredi le 20 octobre de 8h30 à 10h30) + 45% Final (mercredi le 15 décembre de 9h30 à 12h30 au 1207), **si** la moyenne pondérée de l'intra et du final est de 50% ou plus, **sinon** la note des devoirs est plafonnée au minimum de la moyenne des devoirs et de 50% **(menant à un échec)**.

Exemples d'introduction

- Expériences randomisées: Vaccin Salk contre la polio;
- Études empiriques: Logements à prix modique à Baltimore.

Vaccin Salk contre la polio

- Première épidémie aux États-Unis en 1916;
- Plusieurs centaines de milliers de victimes dans les 40 années suivantes;
- Plusieurs vaccins contre la polio ont été découverts dans les années 50;
- Celui de Salk semblait le plus prometteur; il semblait sans risque lors d'essais en laboratoire;
- Il était temps de passe à une expérience à grande échelle;
- Cette expérience a eu lieu en 1954 avec environ 2 000 000 d'enfants.

Comment planifier l'expérience

- Que veut-on savoir?
  - Si le vaccin est efficace;
  - S'il réduit la polio.
- Besoin d'un énoncé qu'on peut tester statistiquement;
- Proportion des vaccinés victimes de la polio est plus petite que ceux qui ne sont pas vaccinés.

Première approche

- Soit p_v, la proportion de vaccinés victimes de la polio;
- On donne le vaccin aux 2 000 000 d'enfants de l'expérience;
- On fait un test de $H_0 : \ p_v \geq p_{1953}$ versus $H_1 : \ p_v < p_{1953}$ où p_{1953} est la proportion d'enfants victimes de la polio en 1953.

Problème

Les épidémies de polio, comme plusieurs autres phénomènes, varient dans le temps. Il faudrait donc pouvoir comparer p_v à p_{1954}. En effet, si $p_{1954} < p_v < p_{1953}$, le vaccin semble réduire la proportion de victimes de la polio alors que c'est le contraire qui arrive. De la même façon, si $p_{1953} < p_v < p_{1954}$, le vaccin ne semble pas fonctionner alors qu'il est efficace.

Il faut donc utiliser deux groupes d'enfants en 1954: un qui reçoit le vaccin et un autre qui ne le reçoit pas.

Deuxième approche

Puisqu'il faut obtenir la permission des parents pour vacciner les enfants, on pourrait prendre comme groupe d'enfants vaccinés ceux dont les parents donnent leur accord et prendre comme groupe d'enfants non vaccinés ceux dont les parents refusent.

Problème

Ceci pourrait créer un biais. Les gens plus riches, plus éduqués, pourraient être plus favorables à la vaccination. Ceci pourrait jouer contre le vaccin. La polio est en partie un problème d'hygiène et les plus pauvres ont pu être en contact avec une version plus faible de la polio lorsqu'ils étaient plus jeunes permettant ainsi de développer des anticorps. S'il y a une différence à la fin, peut-on l'attribuer à la vaccination?

Conclusion: On ne peut pas laisser les gens décider quel traitement ils recevront car on doit s'arranger pour que les deux groupes soient le plus identiques possible sauf pour ce qui est du traitement reçu. Sinon, l'effet du traitement pourrait être confondu avec d'autres effets.

Troisième approche

Plan du *National Foundation for Infantile Paralysis* (NFIP)

- Groupe traitement: Tous les enfants de 2e année dont les parents sont consentants;
- Groupe contrôle: Tous les enfants de 1ère et 3e années.

Problèmes

- La polio est une maladie contagieuse. Si la maladie a une incidence plus élevée dans une classe de 1ère ou 2e année, ceci pourrait créer un biais;
- Le groupe vacciné requiert le consentement des parents, ce qui n'est pas le cas pour le groupe contrôle. Les groupes ne sont peut-être pas comparables quant au revenu, race, etc. À nouveau, l'effet du vaccin pourrait être confondu avec d'autres variables.

Quatrième approche

Utiliser le hasard pour déterminer le groupe auquel chaque enfant est associé.

- On demande le consentement des parents pour faire partie de l'expérience;
- Pour chaque enfant qui participe à l'expérience, on tire à pile ou face (ou quelque chose d'équivalent) pour déterminer le groupe auquel l'enfant fera partie.

Ceci est une expérience contrôlée aléatoirement.

Autres précautions

- Utilisation d'un placebo. Dans ce cas-ci, il faudrait donner une injection de solution saline afin que le patient ne sache pas s'il a reçu le traitement ou non. Suite à une chirurgie grave, un tiers des patients qui reçoivent une pilule de sucre se sentent mieux!!!
- Le docteur doit éventuellement déterminer si le patient a la polio ou non. Certains cas ne sont pas toujours évidents à diagnostiquer. Dans les cas limites, le docteur pourrait se laisser influencer par son biais, pour ou contre le vaccin, s'il est au courant si le patient a reçu le traitement ou le placebo.

Lorsque ni le patient, ni le médecin ne savent si le patient a reçu le traitement ou le placebo, on parle alors d'une expérience à double insu.

Résultats

Expérience contrôlée aléatoirement à double insu

	# patients	taux/100 000
Vaccin	200 000	28
Contrôle	200 000	71
Sans consentement	350 000	46

Plan NFIP

	# patients	taux/100 000
2e année (vaccin)	225 000	25
1ère ou 3e (contrôle)	725 000	54
2e année (sans consentement)	125 000	44

Études empiriques

En anglais, on dit *observational studies*. On parle aussi d'études événementielles ou d'études d'observations. Les études empiriques sont en opposition avec les expériences randomisées (contrôlées). Dans ce type d'expériences (ou d'études), l'expérimentateur ne contrôle pas le traitement qui sera donné à chaque sujet et le hasard n'est pas utilisé pour assigner un sujet à un traitement.

Par exemple, une étude sur l'usage du tabac sur des humains ne peut pas être randomisée.

Dans de telles études, les conclusions des analyses statistiques subséquentes (par exemple, qu'il y a une différence statistiquement significative entre les différents traitements) doivent être interprétées avec prudence étant donné la possibilité de présence de d'autres facteurs confondus. Il est à noter que l'analyse statistique sera la même!

Les études empiriques ne permettent pas de conclure à des relations de

Études empiriques

cause-à-effet.

Étude sur les logements à prix modique à Baltimore

Questions d'intérêt pour les chercheurs: Est-ce que les taudis devraient être détruits ou rénovés? Est-ce que l'état devrait fournir des logements publics à prix modique ou des subventions pour la location de logements privés?

Dans le but de tenter de répondre à ces questions, des chercheurs de l'Université Johns Hopkins ont planifié une expérience ou étude.

Plan de l'étude

- Groupe "contrôle": 300 familles qui vivaient dans des taudis;
- Groupe "traitement": 300 familles qui vivaient dans *Lafayette Courts*, un projet de logements publics à prix modique;
- On a suivi les 600 familles pendant 3 ans;
- On a mesuré l'état de santé et les attitudes sociales des membres des familles à plusieurs reprises à l'aide de questionnaires et d'entrevues;
- Les résultats scolaires des enfants ont également été compilés.

Conclusion statistique: Les deux groupes étaient très similaires sauf sur deux points: le groupe traitement était plus satisfait de son environnement physique et il avait un plus faible taux de mortalité avec seulement deux morts contre 10 dans le groupe contrôle.

Problèmes

Est-ce que les différences entre les deux groupes (ou le fait qu'il n'y ait pas de différences!) sont dues à l'effet de traitement? Puisqu'il ne s'agit pas d'une expérience randomisée, il faut mieux comprendre comment les groupes ont été formés.

Plus de 1000 familles ont soumis leur candidature pour habiter dans *Lafayette Courts* lorsque ceux-ci ont ouvert. Le *Housing Authority* a choisi 800 familles parmi les 1000 pour devenir locataires. Tous vivaient dans des taudis avant.

Les chercheurs ont choisi les sujets du groupe traitement parmi ces 800 et les sujets du groupe contrôle parmi ceux qui n'ont pas été choisis comme locataires.

Autre problème

Les critères de sélection pour résider dans *Lafayette Courts* devaient être reliés à ceux faisant l'objet de cette étude. Il y a de bonnes chances que ceux qui n'ont pas été choisis étaient jugés moins désirables. Donc, s'il y a une différence entre les deux groupes, ça pourrait tout autant avoir été causé par le processus de sélection que par le traitement. De la même façon, une différence réelle entre les deux groupes a pu disparaître suite au processus de sélection. Il n'y aura jamais moyen de distinguer entre le processus de sélection et le traitement.

Comme dans plusieurs études ou expériences, plusieurs sujets ont quitté l'étude: certains se sont tannés, d'autres ont déménagé de la région de Baltimore. Ces deux causes peuvent affecter également les deux groupes.

Par contre, on a éliminé une centaine de familles du groupe contrôle parce qu'elles avaient réussi à sortir de leur taudis! Ceci ne peut que créer un biais dans l'étude.

Qu'est-ce qu'on aurait pu faire?

Il aurait fallu que les chercheurs et le *Housing Authority* se mettent d'accord et adoptent une procédure semblable à la suivante. Trouver 1000 locataires potentiels acceptables, les séparer aléatoirement entre les deux groupes: un groupe contrôle qui va demeurer dans son taudis et un groupe traitement qui va déménager dans *Lafayette Courts*. (Il n'y a pas de problème éthique puisque de toute façon, il y avait plus de locataires acceptables que de places).

Il faut aussi tenter de suivre tout le monde, même ceux qui déménagent.

Dans une étude empirique, il faut toujours se demander si le "traitement" est tout ce qui distingue les groupes, ou si d'autres facteurs d'importance pourraient être confondus.

Rôle de la planification d'expériences

Les chercheurs, ingénieurs, médecins, etc. planifient des expériences afin de mieux comprendre des phénomènes, processus, traitements, etc. Ils poursuivent en général l'un des deux buts suivants:

- Confirmer, c'est-à-dire vérifier des connaissances par rapport à un système, ou;
- Explorer, c'est-à-dire étudier l'effet de nouvelles conditions sur le système.

En général, lorsque c'est le second but qui est poursuivi, les expériences sont faites de façon itérative: on explore, puis selon ce qu'on découvre, on explore dans de nouvelles directions.

Exemple

Un manufacturier de tissus veut mieux comprendre son processus de fabrication et voir jusqu'à quel point il pourrait être avantageux de le modifier afin d'optimiser ses profits tout en tenant compte de diverses restrictions, notamment du point de vue de la qualité.

La mesure d'intérêt peut en être une qu'on veut maximiser (ou minimiser) par exemple la quantité de tissu produit, ou encore être contrôlée (tout en minimisant les coûts pour y arriver) comme une caractéristique demandée par le client, par exemple l'élasticité du tissu.

Ainsi l'élasticité du tissu peut dépendre de la tension sur le fil exercée par la tricoteuse. Si le fait de baisser la tension sur le fil ne change pas vraiment l'élasticité, cette information peut être très importante parce qu'une augmentation de la tension est associée à une augmentation des bris de fils et ainsi des coûts.

On pourrait donc vouloir faire une expérience avec deux niveaux de tension différents.

Questions

1. Est-ce que ces deux niveaux de tension sont les seuls d'intérêt?
2. Y a-t-il d'autres facteurs qui pourraient affecter l'élasticité et qui devraient être étudiés ou contrôlés dans l'expérience?
3. Combien de rouleaux de fils devrions-nous tester?
4. Dans quel ordre devrions-nous prendre les mesures?
5. Quelle méthode d'analyse des données devrions-nous utiliser?
6. Quelle différence dans la moyenne des résultats entre les deux méthodes devrait être considérée comme étant importante?

Importance de la planification statistique de l'expérience

Dans toute expérience, les résultats et conclusions dépendent énormément de la manière dont les données ont été recueillies.

Supposons que nous commencions l'expérience avec le bas niveau de tension. Nous obtenons le nombre de données désiré, puis alors que nous allons passer au haut niveau de tension, il faut changer le rouleau de fil, le nouveau provenant d'un nouveau fournisseur. Le niveau de tension sera alors confondu avec le type de fil utilisé.

Principes de base

Les trois principes de base de la planification d'expériences sont:

- les répliques;
- la randomisation;
- la formation de blocs.

Répliques

Par répliques, nous voulons dire la répétition d'un élément de base de l'expérience, par exemple, la répétition de mesures d'élasticité au niveau bas de la tension (par opposition à faire l'expérience une seule fois à ce niveau).

Les répliques permettent d'estimer l'erreur expérimentale (la variance des erreurs).

Elles mènent également à des estimations plus précises. Ainsi la variance de la moyenne de n observations est la variance d'une observation divisée par n.

Randomisation

La randomisation est la pierre d'assise des méthodes statistiques pour l'analyse des expériences. On parle ici tant de l'utilisation du hasard pour l'assignation des sujets aux différentes combinaisons de traitements qu'à l'ordre dans lequel les observations seront prises.

Les modèles statistiques nécessitent (en général) que les erreurs soient indépendantes et identiquement distribuées (i.i.d.) avec une moyenne nulle. Ceci pourrait ne pas être le cas si les observations d'un premier groupe ont été prises lors d'une journée chaude alors que celles du second groupe l'ont été lorsque la température était plus basse dans une expérience où la température pourrait avoir un effet sur la variable étudiée. En randomisation l'ordre des observations entre les deux groupes, il y aura alors des journées plus chaudes et moins chaudes dans les deux groupes, selon le hasard.

Blocs

La variance d'une moyenne dépend de la variance de la distribution et du nombre d'observations. On peut la diminuer soit en augmentant le nombre d'observations, soit en "diminuant" la variance de la distribution.

Comment peut-on diminuer la variance de la distribution? En utilisant des sujets plus homogènes.

Exemple de plusieurs types de peinture pour les lignes sur les autoroutes. Quatre types de peinture et 20 segments peinturés au total, versus quatre types de peinture regroupés dans cinq endroits où chaque type de peinture est appliqué une fois. Les cinq endroits deviennent des "blocs". Comme la durée de vie de la peinture dépend sans doute de la circulation et des conditions météo, les quatre bouts d'asphalte à un endroit précis d'une autoroute sont sans doute beaucoup plus homogènes que 20 segments d'autoroute pris partout au Québec.

Sept étapes d'une expérience complète

En général, il est souhaitable que la statisticienne et les chercheurs du domaine d'application travaillent ensemble pour une résolution efficace du problème

1. Identification et énoncé du problème;
2. Choix des facteurs et des niveaux;
   - Facteurs quantitatifs (comment les contrôler) versus qualitatifs;
   - Nombre de niveaux différents;
   - Niveaux pré-sélectionnés versus choisis aléatoirement;
3. Sélection de la variable réponse ou dépendante (comment la mesurer);
4. Choix du plan d'expérience;
   - Quelle différence entre les moyennes des différents groupes veut-on détecter;
   - Quel risque sommes-nous prêt à prendre pour détecter une telle différence (puissance);

Sept étapes d'une expérience complète

- Nombre de répliques;
- Ordre des observations et méthode de randomisation (p.e., rats);
- Précision statistique versus coûts;

5. Réalisation de l'expérience (maintenir l'uniformité de l'environnement expérimental);
6. Analyse des données (l'item auquel nous consacrerons la plus grande partie de notre temps, avec l'item 4);
7. Conclusions et recommandations (incluant la suggestion de nouvelles expériences).

Distributions (révision)

Soit $Z_1, \ldots, Z_n$, des variables aléatoires (v.a.) indépendantes de loi normale standard, dénotée par $N(0,1)$, alors $U = \sum_{i=1}^{n} Z_i^2$ est une v.a. de loi khi-deux à n degrés de liberté, dénotée par χ^2_n.

Soit Z une v.a. $N(0,1)$ et U une v.a. indépendante de Z de loi χ^2_n, alors $Z/\sqrt{U/n}$ est une v.a. de loi t à n degrés de liberté, dénotée par t_n.

Soit U et V, deux v.a. indépendantes de loi khi-deux à m et n degrés de liberté, respectivement, alors

$$W = \frac{U/m}{V/n}$$

est une v.a. de loi F à m et n degrés de liberté, dénotée par $F_{m,n}$.

Note: Soit T une v.a. de loi t_n, alors T^2 est une v.a. de loi $F_{1,n}$.

Distribution normale multivariée

Soit un vecteur aléatoire X de longueur n distribué selon une loi normale multivariée de moyenne μ et de matrice de variance-covariance V, dénotée par $N(\mu, V)$. Ainsi, $E(X_i) = \mu_i$ et $Cov(X_i, X_j) = V_{ij}$.

Alors, la densité du vecteur aléatoire X est

$$f(x_1, x_2, \ldots, x_n) = \frac{\exp\{-(1/2)(x-\mu)'V^{-1}(x-\mu)\}}{(2\pi)^{n/2}|V|^{1/2}}.$$

Notez que les contours de la densité, c'est-à-dire les valeurs de x telles que la densité prend la même valeur, sont les ellipses

$$(x-\mu)'V^{-1}(x-\mu).$$

On peut démontrer que les distributions marginales et conditionnelles de composantes de X sont également normales.

Khi-deux décentrée

Soit $X \sim N(\mu, I)$ où I est la matrice identité de dimension $n \times n$, alors $X'X$ est distribué selon une loi khi-deux décentrée à n degrés de liberté et paramètre de décentralité $\lambda = \mu'\mu$, dénotée par $\chi^2(n, \lambda)$.

Attention: Il y a des bouquins et/ou des logiciels statistiques dont la définition de λ est $(1/2)\mu'\mu$. C'est le cas de Searle (1971). Ceci a des répercussions partout.

La densité de la khi-deux décentrée est

$$f(u) = \exp\{-\lambda/2\} \sum_{k=0}^{\infty} \frac{(\lambda/2)^k}{k!} \frac{u^{n/2+k-1} \exp\{-u/2\}}{2^{n/2+k}\Gamma(n/2+k)},$$

où Γ est la fonction gamma.

En utilisant les propriétés de la normale univariée et l'indépendance des composantes, il est facile de démontrer que la moyenne de $X'X$ est $n + \lambda$ alors que sa variance est de $2n + 4\lambda$.

Distribution F décentrée

Soit $U \sim \chi^2(m, \lambda)$ et $V \sim \chi^2_n$, deux v.a. indépendantes, alors

$$W = \frac{U/m}{V/n}$$

est une v.a. de loi F décentrée de paramètres m, n et λ, dénotée par $F(m, n, \lambda)$. Sa densité est

$$\begin{aligned} f(w) &= \exp\{-\lambda/2\} \sum_{k=0}^{\infty} \frac{(\lambda/2)^k}{k!} \frac{m^{m/2+k} n^{n/2+k} \Gamma(m/2+n/2+k)}{\Gamma(m/2+k)\Gamma(n/2)} \\ &\quad \times \frac{w^{m/2+k-1}}{(m+nw)^{m/2+n/2+k}}. \end{aligned}$$

La moyenne de W est $\frac{n}{n-2}(1+\lambda/m)$ alors que sa variance est

$$\frac{2n^2}{m^2(n-2)} \left(\frac{(m+\lambda)^2}{(n-2)(n-4)} + \frac{m+2\lambda}{n-4} \right).$$

Interprétation géométrique dans le modèle linéaire général

Considérons le modèle linéaire général

$$y = X\beta + \epsilon$$

où X est une matrice de dimension $n \times p$ et $\epsilon \sim N(0, \sigma^2 I)$.

Soit $r = \text{rang}(X)$, L_M, l'espace vectoriel engendré par les colonnes de X.

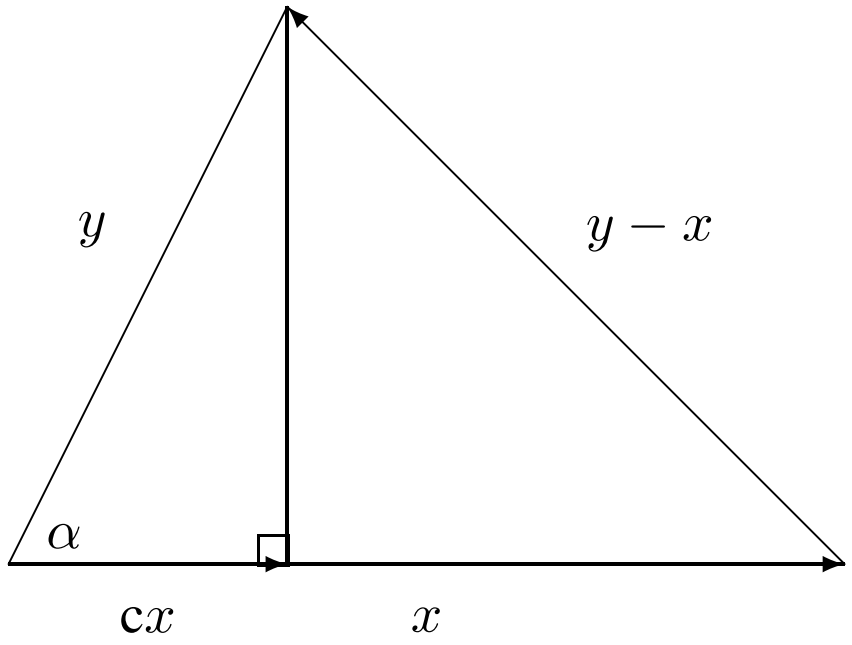
Notons par $d_2(x, y) = \left(\sum_{i=1}^{n}(x_i = y_i)^2\right)^{1/2} = \|x - y\|$, la norme L_2 entre les vecteurs x et y.

La méthode des moindres carrés pour estimer β consiste à trouver le vecteur $\hat{\beta}$ tel que

$$\hat{\beta} = \arg\min_{\beta} d_2(y, X\beta),$$

c'est-à-dire, c'est le vecteur qui minimise la distance euclidienne entre y et $X\beta$.

Projection d'un vecteur sur un autre



Loi des cosinus

$$\begin{aligned} \|y - x\|^2 &= \|y\|^2 + \|x\|^2 - 2\|y\|\|x\| \cos\alpha \qquad (1) \\ \|y - x\|^2 &= (y - x)'(y - x) = y'y - 2y'x + x'x \\ &= \|y\|^2 + \|x\|^2 - 2y'x. \qquad (2) \end{aligned}$$

Projection d'un vecteur sur un autre

De (1) et (2), on déduit que

$$\cos\alpha = \frac{y'x}{\|y\|\|x\|}.$$

La projection de y sur x est le vecteur dans l'espace engendré par x tel que la distance entre celui-ci et y est minimisée. De plus,

$$\begin{aligned}\cos\alpha &= \frac{\text{côté adjacent}}{\text{hypoténuse}} \\ &= c\frac{\|x\|}{\|y\|}.\end{aligned}$$

Ainsi,

$$c = \frac{\|y\|\cos\alpha}{\|x\|} = \frac{y'x\|y\|}{\|y\|\|x\|^2} = \frac{y'x}{\|x\|^2}.$$

Donc la projection de y sur x est

$$\left(\frac{y'x}{\|x\|^2}\right)x.$$

Projection d'un vecteur sur un espace

Soit $\{\alpha_1, \ldots, \alpha_r\}$ une base de L_M, alors la projection de y sur α_i est $(y'\alpha_i)\alpha_i = \alpha_i(\alpha_i' y)$. La projection de y sur L_M est la somme des projections de y sur chacun des éléments de la base, soit

$$\begin{aligned} \sum_{i=1}^{r} \alpha_i(\alpha_i' y) &= \left(\sum_{i=1}^{r} \alpha_i \alpha_i'\right) y \\ = (\alpha_1, \ldots, \alpha_r) \begin{pmatrix} \alpha_1' \\ \vdots \\ \alpha_r' \end{pmatrix} y & \\ &= TT'y, \end{aligned}$$

où $T = (\alpha_1, \ldots, \alpha_r)$ est de dimension $n \times r$.

Note: Une matrice de projection est unique et donc TT' est indépendant de la base choisie.

Prédiction et erreur

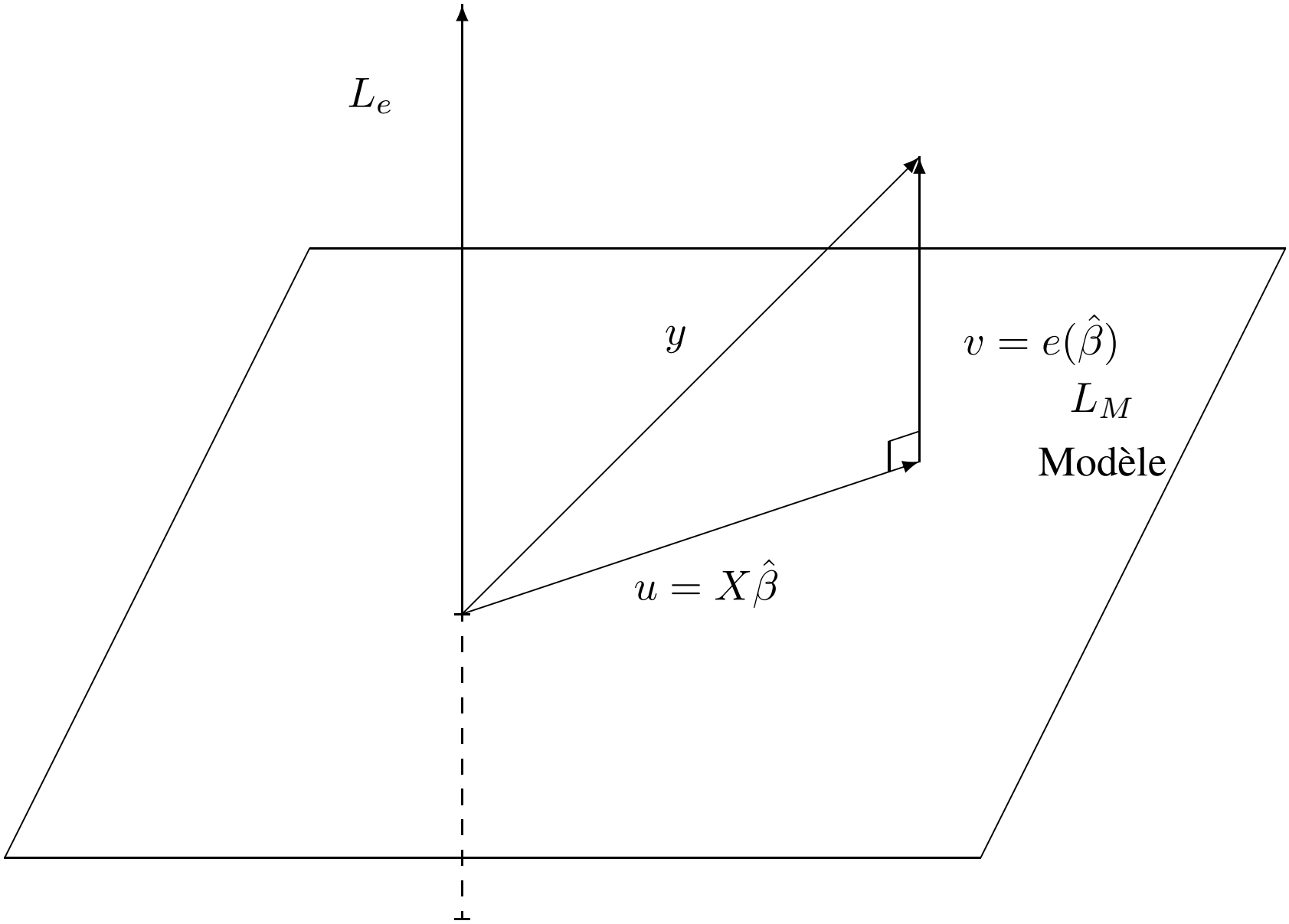
Soit $X\hat{\beta}$, le vecteur des prévisions et $e(\hat{\beta}) = y - X\hat{\beta}$, le vecteur des erreurs. Alors

$$\begin{aligned} y &= X\hat{\beta} + e(\hat{\beta}) \\ &= X(X'X)^{-}X'y + (I - X(X'X)^{-}X')y \\ &= u + v, \end{aligned}$$

où $u \in L_M$, $v \in L_e = L_M^{\perp}$, l'espace perpendiculaire à L_M et A^{-} est l'inverse généralisé de A, soit une matrice satisfaisant l'équation $AA^{-}A = A$.

Notez que $u'v = 0$.

Projection de y dans L_M



Moindres carrés et adéquation

L'estimation par la méthode des moindres carrés consiste donc en une décomposition orthogonale de y entre une composante modèle, $X\hat{\beta}$ et une composante erreur, $y - X\hat{\beta}$.

Comment juger de l'adéquation du modèle?

Si le modèle explique bien les observations y, alors $\hat{e} = e(\hat{\beta})$ devrait être court par rapport à $X\hat{\beta}$.

La statistique qui nous intéresse est donc

$$c\frac{\|X\hat{\beta}\|^2}{\|\hat{e}\|^2} = \frac{\|Py\|^2}{\|(I-P)y\|^2},$$

où c est une constante et P est la matrice de projection dans L_M.