

# Illustration de la méthode des plans d'expériences sur la comparaison de boissons au cola

Jean-Marc Azaïs \*

14 mars 2005

## Résumé

Nous présentons une application de la méthode des plans d'expériences sur un sujet de tous les jours : la comparaison de certaines boissons au Cola. Notre méthodologie est basée sur des dégustations ternaires : le dégustateur doit détecter l'intrus parmi trois verres dont deux sont identiques. Nous étudions les différents facteurs influant sur cette comparaison pour construire une expérience randomisée. Enfin nous développons une méthodologie pour analyser conjointement la détection de l'intrus et les préférences du dégustateur.

## Abstract

This paper presents an application of the design of experiment to the comparison of some cola drinks. It is based on ternary tasting : the taster tries to identify the intruder among three glasses two of which contain the same product. We study the relevant factors and construct a randomized experiment. At the end we develop a new methodology for combining information on the intruder detection and on the preferences of the taster.

Mots clef : Données sensorielles, plans d'expériences factoriels, test du rapport de vraisemblance.

## 1 Introduction

Une expérience menée à l'Université Paul Sabatier par deux étudiants : Sylvan Rey et Julie Trappier [3] a pour but de comparer trois marques de boissons au Cola. Deux sont des marques très connues nous les noterons C et P pour respecter l'anonymat, la troisième est un produit dit "générique" : G.

La comparaison utilise la technique des plans d'expériences mais de manière atypique puisque la réponse est qualitative. En effet, pour maximiser la puissance de l'expérience, les deux étudiants ont utilisé la technique des "comparaisons ternaires" dite encore "solo-duo" : on présente à un dégustateur 3 verres identiques extérieurement dont deux sont remplis d'un même produit alors que le troisième contient un autre produit : "l'intrus".

Après avoir goûté les 3 verres, le dégustateur doit

- désigner l'intrus
- dire quel produit il préfère.

La méthodologie des plans d'expérience doit s'appliquer car la question n'est pas si simple qu'il paraît. En effet, il existe des produits

- sucrés et des produits dit "lights" aux édulcorants ;
- sous emballage plastique et sous emballage métallique ;
- parfumé (citron) ou non parfumé (nature) ;

---

\*Laboratoire de Statistique et Probabilités. UMR-CNRS C55830 Université Paul Sabatier. 118, route de Narbonne. 31062 Toulouse Cedex 4. France.azais@cict.fr. Expérience réalisée par Sylvan Rey et Julie Trappier

- enfin la dégustation peut se faire à température ambiante (ambient) ou a température du réfrigérateur (frais).

Pour des question de budget, puisqu’il s’agit d’une expérience amateur, nous n’avons pas pu aborder, en dehors de la comparaison plastique-métal, un point pourtant très important comme l’analyse de la variabilité inter-lots. Pour supprimer au contraire cette variabilité, tous les produits correspondant à une même dénomination ont été achetés le même jour dans un supermarché et ce à l’intérieur du même emballage plastique réunissant plusieurs bouteilles ou canettes métalliques.

Notre budget a permis de faire 48 comparaisons ternaires faites par 16 dégustateurs qui étaient des étudiants volontaires. La dégustation s’est faite à 11 Heures du matin.

En fait l’expérience se constitue de 5 sous-expériences ayant des buts différents.

1. La comparaison des deux marques C et P. En ce qui concerne les facteurs
  - sucre (light ou sucré)
  - température (ambient ou frais)
  - emballage (plastique ou métal)
  - “plateau” : on choisit le produit qui est présenté deux fois, et celui qui est présenté une fois,
 nous avons réalisé un plan factoriel complet<sup>1</sup>, ce qui nous amène à 16 unités.
2. La comparaison nature-parfumé. Elle comprend 8 unités avec structure de plan factoriel complet sur les facteurs : produit, température et plateau.
3. La comparaison marque - générique : 8 unités basées sur les facteurs produit, plateau et température.
4. La comparaison light-sucré, 8 unités. Il s’agit d’un plan fractionnaire <sup>2</sup> basé sur les 4 facteurs produit, plateau, température et emballage.
5. La comparaison plastique-métal. Il s’agit d’un plan fractionnaire  $2^{4-1}$  basé sur les 4 facteurs produit, plateau, température et sucre.

## 2 Construction de l’expérience

Chacun des 16 dégustateurs physiques tire au hasard un numéro de dégustateur pour l’expérience principale ainsi que deux numéros dans deux parmi les quatre expérience secondaires.

Le préparateur verse les liquides dans les verres. Le serveur qui ne connaît pas le contenu du plateau l’amène et appelle le dégustateur concerné.

Le dégustateur teste les trois verres et coche sur sa fiche l’intrus puis le produit qu’il a préféré.

Chaque dégustateur a a remplir trois grilles du type suivant :

DEGUSTATEUR N°

Comparaison i :

	A	B	C
Intrus			
Préfééré			

Nous décrivons maintenant la construction par SAS<sup>®</sup> de l’expérience principale. Comme chaque dégustateur se voit attribuer un numéro au hasard, cela suffirait comme randomisation si l’on n’avait pas à choisir le numéro de l’intrus : facteur ”numéro” qui sera codé

<sup>1</sup>Le plan factoriel complet est le plan le plus simple qui consiste à réaliser une fois et une seule toutes les combinaisons entre les niveaux des différents facteurs. Dans notre cas tous les facteurs ont deux niveaux et donc s’il y a  $p$  facteurs ce plan demande  $2^p$  unités. Ce plan est appelé également plan  $2^p$ .

<sup>2</sup>Par plan fractionnaire  $2^{4-1}$  on entend un plan qui est la moitié du plan factoriel complet  $2^4$ . Il comprend donc 8 unités, il est construit à partir du plan factoriel complet pour 3 facteurs : les facteurs de base. On définit ensuite le 4ème facteur par une Clef. Voir Azaïs et Bardet [1] pour plus de détails.

A B C dans nos résultats (mais 1 2 3 par commodité au niveau SAS<sup>®</sup>).  
Voici donc le source correspondant à la création du plan. Il utilise la procédure Proc plan.

```
proc plan;
factors emballage =2 sucre=2 temperature=2 plateau =2 ;
/* definit 16 unites correspondant au plan factoriel complet: */
/* c'est a dire toutes les combinaisons dans un ordre aleatoire*/
treatment numero =3;
/* Sur les 16 unites precedemment definies on definit maintenant */
/*le facteur numero prenant 3 valeurs dans un ordre aleatoire */
output out=sasuser.colal1 emballage  sucre temperature plateau ;
/* Les donnees sont ecrites dans un table statistique standard */
/* dont le nom est sasuser.colal1. Elle comprend 16 lignes et 5
variables.*/
proc print;run;
/* On demande l'impression*/
```

Notons que l'on peut facilement obtenir le même résultat avec d'autres logiciels comme Splus<sup>®</sup> ou R<sup>®</sup>.

### 3 Les résultats

#### Expérience principale

unité	plateau	température	sucré	emballage	intrus	résultat	préfééré
1	2 C - 1 P	frais	sucré	métal	A	juste	C
2	1 C - 2 P	frais	sucré	métal	B	faux	
3	2 C - 1 P	ambient	sucré	métal	B	juste	C
4	1 C - 2 P	ambient	sucré	métal	C	faux	
5	2 C - 1 P	ambient	light	métal	A	faux	
6	1 C - 2 P	ambient	light	métal	C	juste	C
7	1 C - 2 P	frais	light	métal	B	juste	P
8	2 C - 1 P	frais	light	métal	A	faux	
9	1 C - 2 P	frais	sucré	plastique	C	faux	
10	2 C - 1 P	frais	sucré	plastique	C	faux	
11	2 C - 1 P	ambient	sucré	plastique	A	juste	C
12	1 C - 2 P	ambient	sucré	plastique	B	juste	C
13	2 C - 1 P	frais	light	plastique	B	faux	
14	1 C - 2 P	frais	light	plastique	A	faux	
15	2 C - 1 P	ambient	light	plastique	B	juste	C
16	1 C - 2 P	ambient	light	plastique	A	faux	

7 bonnes réponses sur 16.

### Comparaison nature-parfumé

unité	plateau	température	produit	intrus	résultat	préféré
17	2 citron - 1 nature	ambiant	C	A	faux	
18	1 citron - 2 nature	ambiant	C	C	juste	citron
19	1 citron - 2 nature	frais	C	B	juste	nature
20	2 citron - 1 nature	frais	C	A	juste	citron
21	1 citron - 2 nature	frais	P	B	juste	nature
22	2 citron - 1 nature	frais	P	B	juste	citron
23	2 citron - 1 nature	ambiant	P	A	faux	
24	1 citron - 2 nature	ambiant	P	A	juste	citron

6 bonnes réponses sur 8.

### Comparaison marque-générique

unité	plateau	température	intrus	résultat	préféré
25	2 G - 1 P	frais	B	vrai	générique
26	1 G - 2 P	frais	C	vrai	P
27	2 G - 1 P	ambiant	A	vrai	P
28	1 G - 2 P	ambiant	A	faux	
29	1 G - 2 C	frais	C	vrai	C
30	2 G - 1 C	frais	C	faux	
31	2 G - 1 C	ambiant	B	vrai	générique
32	1 G - 2 C	ambiant	C	faux	

5 bonnes réponses sur 8.

### Comparaison light-sucré

exp	plateau	produit	température	emballage	intrus	résultat	préféré
33	2 sucre - 1 light	P	frais	plastique	B	faux	
34	1 sucre - 2 light	P	ambiant	plastique	C	vrai	light
35	2 sucre - 1 light	P	frais	métal	C	faux	
36	1 sucre - 2 light	P	ambiant	métal	B	faux	
37	1 sucre - 2 light	C	frais	plastique	C	vrai	sucre
38	2 sucre - 1 light	C	ambiant	plastique	C	vrai	sucre
39	2 sucre - 1 light	C	frais	métal	C	vrai	sucre
40	1 sucre - 2 light	C	ambiant	métal	C	vrai	sucre

5 bonnes réponses sur 8.

### Comparaison plastique-métal

Voici le tableau récapitulant les 8 tests de l'expérience concernant le conditionnement.

exp	plateaux	produit	température	sucre	intrus	résultat	préféré
41	2 plastique - 1 métal	P	ambiant	light	B	faux	
42	1 plastique - 2 métal	P	frais	light	C	faux	
43	2 plastique - 1 métal	P	ambiant	sucre	B	vrai	plastique
44	2 plastique - 1 métal	P	frais	sucre	C	faux	
45	2 plastique - 1 métal	C	frais	light	B	vrai	métal
46	2 plastique - 1 métal	C	ambiant	light	B	vrai	métal
47	2 plastique - 1 métal	C	frais	sucre	A	vrai	plastique
48	2 plastique - 1 métal	C	ambiant	sucre	A	vrai	métal

5 bonnes réponses sur 8.

## 4 Analyse des résultats

Dans un premier temps nous n'interprétons pas les préférences et considérons uniquement le nombre de bonnes réponses. Nous sommes amenés à réaliser un test unilatéral car, dans l'hypothèse d'absence d'effet, la probabilité de bonne réponse vaut  $1/3$  et il semble raisonnable de ne considérer que des alternatives où cette probabilité est supérieure. Le tableau ci-dessous donne le niveau exact du test en fonction du nombre d'essais et de la zone de rejet. Il a été calculé à l'aide d'un programme MATLAB <sup>®</sup>.

Nb d'essais	zone de rejet	niveau exact (avec 4 décimales)
8	$\geq 5$	0.0879
8	$\geq 6$	0.0197
16	$\geq 9$	0.05
24	$\geq 12$	0.0677
24	$\geq 13$	0.0284
48	$\geq 21$	0.086
48	$\geq 22$	0.0485
48	$\geq 23$	0.0254

Seulement deux expériences ou sous-expériences sont significatives :

- la comparaison citron-nature ( $P$ -Value =0.02),
- la comparaison sucré-light sur le produit C : 4 bonnes réponses ( $P$ -Value = $1/81 = 0.012$ ).

Par ailleurs les comparaisons : marque-générique, light-sucré, plastique-métal montrent une légère présomption de significativité ( $P$ -Value =0.088)

Par contre le nombre global de bonnes réponses : 28 est clairement significatif. Les 15 bonnes réponses à température ambiante comme les 13 à température fraîche le sont également.

Le score moyen, environ 58%, significativement différent de  $1/3$ , est à rapprocher des scores de comparaisons ternaires sur les vins cités dans la littérature (Morrot et Brochet [2]), avec un taux de découverte de l'intrus qui est de 62 %.

En conclusion, dans l'ensemble des produits comparés, il existe clairement des produits différents. Il est par contre très difficile de dire quels produits diffèrent puisque (si on met de coté light -sucré pour le produit C) seule la comparaison extrême nature-parfumé est significative.

Si on prend en compte maintenant les préférences, trois remarques s'imposent.

- Pour la comparaison light-sucré sur le produit C, la préférence va toujours au produit sucré. Sous l'hypothèse nulle, la probabilité de trouver 4 fois l'intrus et de toujours préférer le même produit est de

$$(1/3)^4(1/2)^3 = 1.54 \cdot 10^{-3}.$$

Cela renforce donc considérablement la significativité de l'effet. Ce phénomène peut être dû également à une variabilité inter-lot, variabilité que nous n'avons pu quantifier faute de moyens.

- Sur la plupart des expériences, les préférences sont équilibrées et ne peuvent être utilisées pour renforcer la significativité. La comparaison nature-parfumé indique que les consommateurs font une différence entre les produits mais que leur préférence semble aller de manière équilibrée à un produit comme à l'autre.

- Il reste le cas de l'expérience principale où, d'une part, les 7 réponses justes et, d'autre part, la répartition des préférences 6 à 1, donnent séparément de légères présomptions sans être significatives. Comment combiner les deux informations ? C'est ce que nous allons aborder au paragraphe suivant.

### Combinaison des informations

Une première idée serait de réaliser un test du  $\chi^2$ . En effet le résultat de l'expérience peut être :

- l'intrus n'est pas trouvé avec probabilité  $p_1$ ,
- l'intrus est trouvé et on préfère le produit C avec probabilité  $p_2$ ,
- l'intrus est trouvé et on préfère le produit P avec probabilité  $p_3$ ,

et on désire tester l'hypothèse nulle

$$p_1 = 2/3, p_2 = 1/6, p_3 = 1/6. \quad (1)$$

Dans notre exemple, on dispose de  $n = 16$  observations et des effectifs  $Y_1 = 9$ ,  $Y_2 = 6$ ,  $Y_3 = 1$ . La statistique du test du  $\chi^2$  vaut 5.4. La valeur critique asymptotique à 5% vaut 5.99 mais elle est très approximative compte tenu du faible effectif. Il faudrait calculer une valeur critique non asymptotique, mais nous ne l'avons pas fait car le test du  $\chi^2$  n'exploite pas l'information que l'alternative peut seulement prendre la forme  $p_1 < 2/3$ . Comme le test du  $\chi^2$  est proche du test du rapport de vraisemblance nous avons donc préféré effectuer un test du rapport de vraisemblance entre l'hypothèse nulle  $H_0$  donnée par (1) et l'hypothèse générale  $H_1$  donnée par  $p_1 < 2/3$ .

Il est facile de vérifier que la log-vraisemblance sous  $H_0$  vaut

$$Lv_0 = Y_1 \log(2/3) + Y_2 \log(1/6) + Y_3 \log(1/6).$$

On a négligé les coefficients multinomiaux qui sont constants. Dans le cas où  $n = 16$ , il est facile de voir que les estimateurs du maximum de vraisemblance sous  $H_1$  sont définis par

- si  $Y_1 \leq 10$  :  $\hat{p}_i = n_i/n$ ,  $i=1,2,3$ .
- si  $Y_1 \geq 11$  :  $\hat{p}_1 = 2/3$ ;  $\hat{p}_2 = \frac{Y_2}{3(Y_2 + Y_3)}$ ;  $\hat{p}_3 = \frac{Y_3}{3(Y_2 + Y_3)}$

et la log-vraisemblance sous  $H_1$  vaut

$$Lv_1 = Y_1 \log(\hat{p}_1) + Y_2 \log(\hat{p}_2) + Y_3 \log(\hat{p}_3).$$

Dans le cas de nos observations, le calcul montre que la statistique du rapport de vraisemblance vaut  $d = -13.8358 - (-16.1915) = 2.3557$ .

Le programme MATLAB<sup>®</sup> donné en annexe permet de simuler un échantillon de taille  $n$  de la distribution d'échantillonnage sous l'hypothèse nulle, ce dans le cas de 16 observations. Il est donc facile d'en déduire la "P-value" associée à la valeur 2.3557. Là encore d'autres logiciels permettraient aisément d'obtenir le même résultat.

Sur 4000 simulations, on trouve que la probabilité pour la statistique de test de dépasser 2.3557 est 0.072. Cela donne une légère présomption mais ce n'est pas significatif. Le mystère reste donc entier...

## 5 Conclusion

Notre étude reste, et c'est peut-être mieux ainsi, "a toy example". Il conviendrait, si l'on disposait d'un budget plus conséquent :

- d'étudier la variabilité inter-lot qui n'est approchée ici que par la comparaison métal-plastique,

– de faire des comparaisons avec un effectif plus important.

Faisons un calcul d'effectif pour obtenir un niveau de 5% et une puissance de 90% sous l'alternative où la probabilité de bonne réponse est de 60% (ce qui correspond grosso-modo au score moyen). En utilisant l'approximation gaussienne de la binomiale, le nombre de succès  $S$  sous l'hypothèse nulle  $p = 1/3$  peut être représenté approximativement par

$$S = n/3 + Z\sqrt{\frac{2n}{9}},$$

où  $Z$  est une variable de loi normale centrée réduite. Comme la valeur critique à 5% pour une telle loi vaut 1.65, on rejettera l'hypothèse nulle par un test unilatéral à 5% si

$$S > n/3 + 1.65\sqrt{\frac{2n}{9}}. \quad (2)$$

Sous l'alternative  $p = 0.6$  le nombre de succès a maintenant la représentation approchée :

$$S = 0.6n + Z\sqrt{n \times 0.4 \times 0.6}.$$

Cette loi dépasse avec probabilité 90 % la valeur  $0.6n - 1.28\sqrt{n \times 0.4 \times 0.6}$  car  $-1.28$  est le fractile associé à la probabilité 0.10 de la loi normale centrée réduite. Pour obtenir la puissance recherchée il fut donc que cette dernière quantité coïncide avec la valeur donnée par (2). En résolvant on trouve  $n=36$ .

Pour effectuer les 5 comparaisons de l'expérience ci-dessus ainsi que la comparaison inter-lot il faudrait donc environ  $6 \times 36 = 216$  unités. Cela demande un vrai budget et une vraie organisation.

Notons que le paragraphe 4 propose une méthodologie pour l'analyse de comparaisons ternaires.

En conclusion, nous avons voulu considérer une problématique très facile à comprendre et sur laquelle on trouve dans la conversation courante des opinions très diverses. Nous avons montré comment la méthodologie des plans d'expériences s'appliquait soulignant ainsi la nécessité permanente d'une réflexion statistique au moment de la construction de l'expérience.

## 6 Programme

On commence par prolonger la fonction  $x \log(x)$  en zéro par

```
function t= xlog(x)
    if x== 0, t=0; ,
        else, t =x*log(x); , end
```

puis la simulation se fait par :

```
function t= cola(n)

z= [0.5 4.5 5.5 7];
m =[ 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6];
t = zeros(1,n);
    for i=1:n,
        x = unidrnd(m);
        Y= histc(x,z);
% La solution est un peu acrobatique mais, au final, Y est un vecteur
% de dimension 3 qui contient une realisation de la multinomiale
```

```

% M(16,(2/3,1/6,1/6))
  if Y(1) >= 11, p1= 2/3;
    if ( Y(3) +Y(2))==0, p2= 1/6; p3=1/6;,
    else ,
    p2 = Y(2)/ (3*(Y(3) +Y(2)));
    p3= Y(3)/ (3*(Y(3) +Y(2)));
    ,end
  lv1 = Y(1) *log( 2/3)  +  3*(Y(3) +Y(2))*( xlog( p2) + xlog( p3));,
  else,
%estimateurs et vraisemblance si  Y(1) >=11

  p1 = Y(1)/16;
  p2 = Y(2)/ 16;
  p3= Y(3)/ 16;
  lv1 = 16* ( xlog(p1)+xlog(p2)+xlog(p3) );
%estimateurs et vraisemblance si  Y(1) <11
  ,end
  lv0 = Y(1)* log( 2/3) + Y(2)* log( 1/6) + Y(3)* log( 1/6);
%vraisemblance sous l'hypoth\`ese nulle.
  t(i) = lv1-lv0;  ,end

```

## Références

- [1] J.-M. Azaïs et J.-M. Bardet, (2005). *Le modèle linéaire par l'exemple : Régression, Analyse de la Variance et Plans d'Expérience*. A paraître.
- [2] G. Morrot et F. Brochet, (2004). Cognition et Vin. *Première journée scientifique Vigne-Vin*, p57-66, Agro-Montpellier, France.
- [3] S. Rey et J. Trappier, (2004). *Étude du goût*. Mémoire de DESS, Université Paul Sabatier, Toulouse, France.