

# LSTAT2200

*Lucien Ledune*

*mai 26, 2018, 16:52*

## Contents

<b>Introduction</b>	<b>1</b>
Jeu de données . . . . .	1
Contraintes . . . . .	2
Questions de l'étude . . . . .	2
<b>Quel type de poivron est le plus consommé à Legoland ?</b>	<b>2</b>
Échantillonnage et calculs . . . . .	2
Discussion . . . . .	3
<b>Est-ce que la consommation de poivron est homogène au sein des trois provinces de Legoland ?</b>	<b>3</b>
Discussion . . . . .	5
<b>Le chiffre d'affaire diffère-t-il au sein des chaînes de supermarchés ?</b>	<b>5</b>

## Introduction

Dans le cadre du cours d'échantillonnage et sondage, nous allons réaliser un projet résumant la matière vue au cours. Ce travail sera divisé en plusieurs parties :

- Un descriptif de la problématique
- La description des méthodes de sondages utilisées
- Les statistiques descriptives de nos différents sondages
- Une discussion sur les résultats
- Une discussion sur les limites de notre approche et comment l'améliorer

## Jeu de données

Pour ce travail nous utiliserons le jeu de données **Legoland** qui représente la consommation de poivrons des habitants de Legoland. Legoland se divise en trois provinces, et les principaux supermarchés vendant les poivrons sont eux aussi au nombre de trois. Les habitants se nourrissent de poivrons de trois différentes sortes : Rouge, jaune et vert.

Une ligne représente une transaction, que l'on peut associer à la consommation de poivrons d'un ménage au cours d'une journée.

Voici les différentes variables :

Province : La province dans laquelle la transaction a été effectuée.

Enseigne : Le supermarché dans lequel elle a été faite.

Paielement : Le type de paiement

Vert, jaune, rouge : La quantité de poivrons de chaque type achetée.

N\_Personnes : Nombre de personnes au sein du ménage.

Il faut garder à l'esprit que même si nous avons la base de donnée pour ce projet, nous ne pouvons pas nous en servir directement puisque le but est d'échantillonner. Nous disposons cependant de certaines informations comme les effectifs des provinces et des enseignes. (Voir mail de Vincent Bremhorst).

## Contraintes

- Chaque province de Legoland doit être représentée dans l'échantillon, ainsi que chaque enseigne dans chaque province.
- L'obtention des informations à un coût : 1, 1.2 et 1.5\$ pour Grancub, P'ti rond et Toupla.
- Le budget pour chaque échantillon est de 25000\$.

## Questions de l'étude

Nous allons tenter de répondre à 4 questions principales au cours de ce travail :

- Quel type de poivron est le plus consommé à Legoland ?
- La consommation de poivrons est-elle homogène au sein des trois provinces ?
- Le chiffre d'affaire diffère-t-il au sein des chaînes de supermarchés ?
- Est-ce que l'installation de terminaux pour le paiement électronique serait rentable ?

## Quel type de poivron est le plus consommé à Legoland ?

### Échantillonnage et calculs

Puisque dans cette question nous ne nous intéressons pas aux différents groupes mais bien à la totalité de la population, un sondage aléatoire sans remise devrait être efficace.

Premièrement nous allons constituer notre échantillon. Celui-ci sera fait grâce à la fonction *srswor* afin de prendre 19250 échantillons.

Nous allons ensuite vérifier si celui-ci ne dépasse pas la limite de coûts, auquel cas nous devrions revoir ce chiffre à la baisse.

## [1] 24913.14

Nous sommes juste en dessous de la limite, nous pouvons donc conserver cet échantillon pour l'analyse.

Nous allons maintenant vérifier si cet échantillon représente bien les trois enseignes et les trois provinces. Pour ce faire nous allons afficher calculer les tables de proportions des provinces ainsi que des enseignes de la population (dont nous connaissons les totaux à priori) et les comparer avec les proportions de notre échantillon. Nous pourrons ensuite afficher la différence de ces proportions entre l'échantillon et l'information à priori pour voir si nous en sommes proche.

Table 1: Différence de proportion dans les provinces

Nord	Centre	Sud
0.0004559	-0.001353	0.0008971

Table 2: Différence de proportion dans les enseignes

Grancub	P'ti rond	Toupla
0.00454	-0.006231	0.001691

En effet il semble que les proportions soient bien respectées. Nous pouvons également afficher la table croisée de notre échantillon :

Table 3: Table croisée de l'échantillon

	Grancub	P'ti rond	Toupla
<b>Nord</b>	2150	2141	2176
<b>Centre</b>	2183	2065	2112
<b>Sud</b>	2169	2118	2136

L'échantillon étant maintenant constitué et validé nous pouvons maintenant répondre à la question en faisant la somme des achats des différents types de poivrons de l'échantillon :

Table 4: Consommation de poivron de l'échantillon

vert	jaune	rouge
161140	163636	333455

## Discussion

Comme nous le voyons très bien ici, les poivrons rouges sont largement majoritaires avec plus de 333.000 poivrons consommés. Les deux autres types de poivrons sont assez similaire dans leurs ventes avec 163636 et 161140 en faveur des poivrons jaunes. Cette information pourrait être intéressante pour la détermination des prix des poivrons, car elle nous informe sur la demande de ceux-ci. Cependant avant de prendre une conclusion trop hâtive nous allons aller plus loin en regardant si cette consommation est homogène entre les différentes provinces.

## Est-ce que la consommation de poivron est homogène au sein des trois provinces de Legoland ?

Nous allons maintenant nous intéresser à la deuxième question principale de ce projet, et tenter de déterminer si la consommation de poivrons est homogène au sein des trois provinces de Legoland.

Pour cette question nous pouvons changer notre plan de sondage.

Nous devons ici bien nous assurer que les trois provinces sont bien représentées car nous voulons avant tout analyser l'homogénéité dans la consommation de poivrons au sein de celles-ci.

Pour ce faire nous allons donc utiliser un échantillonnage aléatoire à deux degrés afin de bien cibler notre étude. Comme nous disposons des totaux des différentes provinces nous pourrions nous en servir afin de quantifier les échantillons de chaque province que nous allons récupérer. Cela va nous permettre de sélectionner toutes les provinces et par la suite de choisir le nombre d'échantillon à récupérer dans chaque province.

Nous allons pour le choix de ce nombre nous servir des proportions, que nous allons ensuite multiplier par

une constante  $k$  permettant de se rapprocher des 25000\$ de contrainte sans les dépasser afin de garantir la qualité de l'échantillon.

## [1] 24951.05

En effet nous restons bien dans les coûts avec 24951\$. Nous allons maintenant regarder les provinces et les enseignes représentées dans notre échantillon.

Table 5: Proportions de la population (données)

Nord	Centre	Sud
0.3355	0.3317	0.3328

Table 6: Proportions de l'échantillon

Nord	Centre	Sud
0.3355	0.3318	0.3327

Table 7: Compte des différentes provinces dans l'échantillon

Nord	Centre	Sud
6458	6386	6405

Nous voyons bien que les différentes proportions sont respectées. Regardons maintenant si les enseignes sont bien représentées.

Table 8: Compte des enseignes de l'échantillon

Grancub	P'ti rond	Toupla
6398	6372	6479

Encore une fois il est clair que les trois enseignes sont bien représentées dans l'échantillon, nous pouvons donc le valider et répondre à la question.

Pour ce faire nous allons regarder la consommation des différents poivrons grâce à une équation "logique" pour bien séparer les provinces.

Table 9: Consommation de poivrons par région

	Vert	Jaune	Rouge	Somme
<b>Nord</b>	69468	49904	99341	218713
<b>Centre</b>	53413	54452	111716	219581
<b>Sud</b>	38644	59624	122364	220633
<b>Somme</b>	161525	163980	333421	1317853

## Discussion

Le tableau ci-dessus mets en évidence plusieurs choses intéressantes. Premièrement, la consommation **totale** de poivrons au sein des différentes provinces semble en effet assez semblable avec 218713, 219581 et 220633. Cependant attention car si l'on observe la consommation par types de poivrons, celle-ci diffère énormément d'une province à l'autre. Ainsi :

- Les trois provinces consomment une majorité de poivrons rouges.
- La province du nord consomme plus de poivrons verts que les autres provinces mais moins de poivrons jaunes.
- La province de centre consomme des poivrons verts et jaunes dans une quantité assez similaire.
- La province du sud consomme plus de poivrons jaunes que les autres provinces mais moins de poivrons verts

Cette analyse nous permet donc très clairement de distinguer une hétérogénéité de la consommation des types de poivrons en fonction des provinces.

## Le chiffre d'affaire diffère-t-il au sein des chaînes de supermarchés ?

Pour répondre à cette question il faut rappeler que le coût au kg des différents types de poivrons sont actuellement identiques, nous pouvons donc nous concentrer sur la quantité de poivrons vendus au sein de chaque enseigne. Ici encore nous connaissons les totaux de la population, nous pourrions donc nous servir de ces totaux pour obtenir un échantillon de la même proportion.

Nous devons ici encore faire bien attention à prendre un échantillon aussi grand que possible tout en restant dans la limite, car comme les quantités achetées varient beaucoup d'une transaction à l'autre nous devons tenter de nous rapprocher un maximum de la réalité en prenant autant de lignes que possible.