

LSTAT2200 - Projet Legoland

Lucien Ledune

mai 26, 2018, 22:16

Contents

Introduction	1
Jeu de données	1
Contraintes	2
Questions de l'étude	2
Quel type de poivron est le plus consommé à Legoland ?	2
Échantillonnage et calculs	2
Discussion	3
Est-ce que la consommation de poivron est homogène au sein des trois provinces de Legoland ?	3
Discussion	5
Le chiffre d'affaire diffère-t-il au sein des chaînes de supermarchés ?	5
Discussion	7
Discussion : Conseils au gouvernement de Playmoland	7
Pays	7
Enseignes	7
Quelle est la proportion de transaction bancaires ?	8
Discussion	8
Conclusion	8
Annexes	8

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 pourrions 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
Nord	2150	2141	2176
Centre	2183	2065	2112
Sud	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 (Les statistiques descriptives se trouvent en annexe) :

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 pourrons 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.

Les statistiques descriptives se trouvent en annexe.

Table 9: Consommation de poivrons par région

	Vert	Jaune	Rouge	Somme
Nord	69468	49904	99341	218713
Centre	53413	54452	111716	219581
Sud	38644	59624	122364	220633
Somme	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.

Notre analyse ici va être très similaire à la précédente, à la différence près que ce n'est pas aux provinces que nous nous intéressons mais aux enseignes. Nous allons donc encore une fois utiliser les proportions des enseignes de la population (connues) pour répliquer celles-ci dans notre échantillon.

Pour ce faire nous allons encore utiliser un échantillon à deux degrés : Premièrement une sélection des groupes, ensuite un échantillonnage simple sans remise au sein de ces groupes en suivant les proportions de la population pour que les résultats collent un maximum à la réalité.

Dans cette analyse lorsque nous parlerons de chiffre d'affaire nous parlerons des ventes de l'échantillon uniquement, même s'il serait facile d'interpoller les valeurs pour coller au chiffre d'affaire "réel" de l'entreprise, ceci n'a pas vraiment de sens ici puisque l'objet de l'étude est la **différence** de chiffre d'affaire entre les enseignes.

[1] 24984.75

Encore une fois le coût de l'échantillon est de 24984\$, nous respectons donc bien le budget.

Regardons maintenant si les proportions sont bien respectées :

Table 10: Proportions de la population (données)

Grancub	P'ti rond	Toupla
0.3332	0.3348	0.332

Table 11: Proportions de l'échantillon

Grancub	P'ti rond	Toupla
0.3332	0.3347	0.332

Table 12: Compte des différentes enseignes dans l'échantillon

Grancub	P'ti rond	Toupla
6431	6460	6408

En effet les proportions sont parfaitement respectées, comme attendu.

Nous allons maintenant nous intéresser aux autres contraintes. Les provinces sont-elles bien représentées ?

Table 13: Compte des provinces dans l'échantillon

Nord	Centre	Sud
6518	6345	6436

Cela semble en effet être le cas. Il ne nous reste plus maintenant qu'à vérifier que chaque enseigne est bien représentée dans chaque province :

Table 14: Tableau croisé de l'échantillon

	Grancub	P'ti rond	Toupla	Total
Nord	2221	2144	2153	6518
Centre	2117	2137	2091	6345
Sud	2093	2179	2164	6436
Total	6431	6460	6408	19299

Il est clair que les enseignes sont bien représentées au vue des résultats de notre tableau croisé.

L'échantillon étant maintenant validé nous allons répondre à la question qui est de déterminer si le chiffre d'affaire est le même parmi les trois enseignes.

Les statistiques descriptives se trouvent en annexe.

Puisque le prix au kg est identique pour chaque poivron et ce dans chaque enseigne, nous pouvons considérer qu'1kg coûte 1\$, ce qui simplifiera grandement les calculs car nous ne devons pas nous soucier du prix dans ceux-ci. La surface vendant le plus de kg de poivrons sera donc celle qui réalisera le meilleur chiffre d'affaire.

Grâce à cette simplification, nous pouvons donc suivre exactement la même méthodologie que dans la question précédente, en étudiant les enseignes au lieu des provinces.

Voici le tableau résultant de cette analyse :

Table 15: Consommation de poivrons par enseigne

	Vert	Jaune	Rouge	Somme
Grancub	44280	70875	99018	214173
P'ti rond	53185	54599	110936	218720
Toupla	62921	40623	123618	227161
Somme	160386	166096	333572	1320108

Discussion

La première chose que nous notons est la similarité entre cette analyse et la précédente ! Les conclusions semblent très similaires, voici ce que nous pouvons noter :

- Les trois enseignes semblent avoir un chiffre d'affaire relativement similaire, avec 214173, 218720 et 227161. Cela dit l'enseigne toupla semble tout de même en vendre légèrement plus que les autres, la différence étant d'environ 13000 kg avec Grancub.
- Le poivron rouge est celui qui rapporte le plus aux trois enseignes.
- Grancub vend plus de poivrons jaunes que de poivrons verts.
- P'ti rond vend plus ou moins autant de poivrons jaunes que de poivrons verts.
- Toupla vend plus de poivrons verts que de poivrons jaunes.

Encore une fois nous constatons une grande hétérogénéité dans les ventes lorsque l'on prend en compte les types de poivrons vendus par les enseignes.

Cependant en observant les totaux on voit que la différence de chiffre d'affaire même si elle est existante n'est pas si énorme. L'ordre décroissant est celui-ci : Toupla > P'ti rond > Grancub.

Discussion : Conseils au gouvernement de Playmoland

Cette discussion va être divisée en deux parties, les conseils concernant le pays et les conseils concernant les enseignes.

Pays

Notre première analyse nous apporte des renseignements importants quant à la demande des différents poivrons du pays. Il apparaît clairement que la demande de poivrons rouges est bien supérieure à la demande des autres types de poivrons, le conseil assez évident à donner ici est donc d'augmenter le prix du poivron rouge, car les habitants de Legoland seront prêts à payer plus pour continuer à profiter de leur aliment préféré.

Si il est possible de changer les prix d'exportation en fonction de la province, nous pouvons aussi en suivant la même logique augmenter les prix des poivrons verts dans le nord et le prix des poivrons jaunes dans le sud. Cependant cela reste hypothétique car il est peut probable qu'une telle discrimination soit applicable en réalité.

Enseignes

S'il nous est possible de différencier les prix des enseignes en leur vendant directement les poivrons, alors il serait utile de :

- Augmenter le prix des poivrons jaunes pour Grancub
- Augmenter le prix des poivrons verts pour Toupla
- Encore une fois, augmenter le prix des poivrons rouges de manière générale.

Quelle est la proportion de transaction bancaires ?

Pour répondre à cette question nous pouvons reprendre notre deuxième échantillon, qui était bien représentatif des provinces mais aussi des enseignes. De plus il est proportionné correctement par rapport aux proportions des provinces, ce qui nous permettra d'avoir une proportion **totale** proche de la réalité.

Nous allons regarder quelle est la proportion de transaction bancaires dans celui-ci.

Table 16: Tableau des types de paiement

	Espèce	Paiement Électronique	Somme	Proportion_province
Nord	1849	4609	6458	0.7137
Centre	2568	3818	6386	0.5979
Sud	3819	2586	6405	0.4037
Somme	8236	11013	19249	0.5721

Discussion

Nous voyons que les pourcentages théoriques donnés par Legoland sont confirmés (70%, 60% et 40%). Cependant la proportion quand au total des transactions n'atteint que 57%.

Le seuil de rentabilité n'étant pas atteint il ne faut pas rajouter les terminaux dans le contrat. Cependant ici aussi, si un contrat peu être négocié pour une seule province, alors celle du Nord pourrait être rentable car la proportion dépasse bel et bien le seuil de rentabilité fixé.

Conclusion

Dans ce travail, nous avons formé différents échantillons dans le but de répondre à différentes questions posées par le gouvernement de Playmoland.

Premièrement, nous avons montré que les poivrons rouges étaient les plus consommés, ce qui nous a permis de fournir des recommandations quant aux nouveaux prix à fixer.

Pour aller plus loin, nous avons examiné ce que devenait la consommation de ces poivrons au sein des différentes provinces et avons montré que celles-ci différaient assez bien. Dans la même optique nous avons analysé les ventes des différentes enseignes et avons trouvé des conclusions similaires.

Enfin, nous avons montré que la négociation des terminaux dans le contrat n'était pas rentable car elle n'atteignait pas le seuil nécessaire à sa rentabilité.

Annexes

Table 17: Statistiques descriptives du premier échantillon (continued below)

Province	Enseigne	Paie ment	Vert
Nord :6467	Grancub :6502	EspÃ¨ce : 8415	Min. : 0.000
Centre:6360	P'ti rond:6324	Paie ment Ã©lectronique:10835	1st Qu.: 0.720
Sud :6423	Toupla :6424	NA	Median : 7.090
NA	NA	NA	Mean : 8.371
NA	NA	NA	3rd Qu.:13.557
NA	NA	NA	Max. :44.650

Jaune	Rouge	N_personnes
Min. : 0.000	Min. : 0.00	Min. : 1.000
1st Qu.: 0.050	1st Qu.: 9.52	1st Qu.: 3.000
Median : 6.960	Median :16.97	Median : 4.000
Mean : 8.501	Mean :17.32	Mean : 4.524
3rd Qu.:13.877	3rd Qu.:24.51	3rd Qu.: 6.000
Max. :54.600	Max. :63.20	Max. :12.000

Table 19: Statistiques descriptives du deuxi me  chantillon (continued below)

Province	Enseigne	Paie ment	Vert
Nord :6458	Grancub :6398	EspÃ¨ce : 8236	Min. : 0.000
Centre:6386	P'ti rond:6372	Paie ment Ã©lectronique:11013	1st Qu.: 0.800
Sud :6405	Toupla :6479	NA	Median : 7.120
NA	NA	NA	Mean : 8.391
NA	NA	NA	3rd Qu.:13.560
NA	NA	NA	Max. :44.650

Jaune	Rouge	N_personnes
Min. : 0.000	Min. : 0.00	Min. : 1.000
1st Qu.: 0.000	1st Qu.: 9.59	1st Qu.: 3.000
Median : 7.000	Median :16.99	Median : 4.000
Mean : 8.519	Mean :17.32	Mean : 4.527
3rd Qu.:13.940	3rd Qu.:24.49	3rd Qu.: 6.000
Max. :51.270	Max. :64.15	Max. :12.000

Table 21: Statistiques descriptives du troisi me  chantillon (continued below)

Province	Enseigne	Paie ment	Vert
Nord :6518	Grancub :6431	EspÃ¨ce : 8335	Min. : 0.000
Centre:6345	P'ti rond:6460	Paie ment Ã©lectronique:10964	1st Qu.: 0.630
Sud :6436	Toupla :6408	NA	Median : 7.020
NA	NA	NA	Mean : 8.311

Province	Enseigne	Paielement	Vert
NA	NA	NA	3rd Qu.:13.500
NA	NA	NA	Max. :44.650

Jaune	Rouge	N_personnes
Min. : 0.000	Min. : 0.00	Min. : 1.000
1st Qu.: 0.000	1st Qu.: 9.53	1st Qu.: 3.000
Median : 7.130	Median :17.04	Median : 4.000
Mean : 8.606	Mean :17.28	Mean : 4.523
3rd Qu.:14.030	3rd Qu.:24.42	3rd Qu.: 6.000
Max. :54.600	Max. :59.91	Max. :12.000