

## Projet R et analyse de données

### 1 Présentation générale

La chaîne de magasins X possède deux supermarchés dans deux quartiers de la ville V. La direction de la chaîne voudrait s'assurer de l'opportunité de maintenir ces deux magasins dans la ville et pour cela a recruté un stagiaire en statistique pour mener une étude comparative des deux magasins. Le stagiaire a pour mission :

1. de réaliser une étude descriptive sur les ventes hebdomadaires de chacun des deux magasins (cf. section 2),
2. de calculer le risque pour chacun des deux magasins de ne pouvoir assurer les charges fixes qui se montent à 5 Meuro par an (cf. section 3),
3. de vérifier que l'objectif moyen qui est de 100 keuro peut être considéré comme atteint pour chacun des deux magasins (cf. section 3),
4. de déterminer si l'un des deux magasins est significativement plus performant que l'autre (cf. section 4),
5. d'étudier l'influence du type de temps (beau, variable, mauvais) sur le résultat des ventes (cf. section 5)
6. d'étudier le lien entre le type de produits vendus et la période de l'année (cf. section 6).

Pour cela le stagiaire dispose d'un fichier de données contenant le montant des ventes hebdomadaires en keuro de l'année précédente pour chaque magasin, le type de temps moyen pour chaque semaine et la répartition des ventes sur quelques produits type pour le premier magasin. La chaîne de magasins spécifie que tous les traitements doivent être effectués sous R.

Chaque étudiant doit télécharger le jeu de données qui lui a été affecté, le traiter sous R afin de réaliser un rapport suivant le plan défini (cf. section 7).

### 2 Étude descriptive

Faire une analyse descriptive des 3 premières variables.

Pour cela, produire quelques graphiques et calculs d'indices qui font sens par rapport à l'étude.

### 3 Risque sur les charges et objectif de vente

On s'intéresse au montant des ventes hebdomadaires.

- Pour chaque magasin, estimer la moyenne  $\mu$  et l'écart-type  $\sigma$  du montant des ventes hebdomadaires
- Avec ces valeurs de paramètres, calculer alors la probabilité pour que le montant annuel soit inférieur à 5 Meuro
- Construire l'intervalle de confiance du montant moyen des ventes hebdomadaires pour chacun des 2 magasins
- Réaliser le test pour savoir si l'objectif moyen de 100 keuro peut être considéré comme atteint pour chacun des deux magasins

### 4 Magasin le plus performant

Pour répondre à la question de savoir si un magasin est significativement plus performant que l'autre, on créera un vecteur qui indique, chaque semaine, si le magasin 1 a une vente plus élevée que le magasin 2.

À partir de ce vecteur, réaliser un test sur  $p$  qui permette de répondre à la question.

### 5 Influence du temps

Pour tester l'influence du temps sur le résultat des ventes :

- répartir le montant des ventes hebdomadaires en 3 classes : “< 90”, “90 – 110” et “> 100”,
- construire pour chaque magasin le tableau de contingence croisant ces 3 classes et le type de temps
- réaliser le test d'indépendance entre montant des ventes et type de temps

## 6 Lien entre le type de produits vendus et la période de l'année

Pour étudier le lien entre la vente de certains produits et le moment de l'année, on considère le tableau de taille  $52 \times 7$  formé des ventes hebdomadaires sur quelques produits.

- Réaliser une ACP centrée réduite
- Interpréter les composantes principales en argumentant
- Commenter les regroupements de semaines et de produits

## 7 Plan

Le rapport devra comporter les différentes parties :

- Introduction : Objet de l'étude et moyens mis en oeuvre
- Description des données : graphiques, indices, interprétation, commentaires
- Analyse : Risque sur les charges et objectif de vente, performance magasin, influence du temps
- Analyse des ventes selon la période
- Synthèse et conclusion de l'étude