

## DEVOIR MAISON

*À rendre pour le 8 juin 2018*

### 1 Enquête de fréquentation touristique dans l'hôtellerie de l'Insee (5 points)

À partir de la présentation de l'enquête de fréquentation touristique dans l'hôtellerie de l'Insee fournie en annexe (dernière page du sujet), répondez aux questions suivantes.

**Question 1.1** (1,5 points) Parmi ces objectifs, lesquels sont remplis par le dispositif décrit en Annexe ? Justifiez vos réponses.

- A. Savoir s'il y a plus ou moins de touristes que l'année dernière
- B. Savoir s'il y a plus ou moins de touristes français que l'année dernière
- C. Savoir si les touristes sont satisfaits de leur séjour en France

**Question 1.2** (0,5 point) Rappelez les caractéristiques d'une base de sondage parfaite.

**Question 1.3** (1 point) À votre avis, dispose t-on d'une telle base de sondage ici ? Justifiez votre réponse.

**Question 1.4** (0,5 point) Rappelez en quelques lignes ce qu'est un sondage stratifié.

**Question 1.5** (1,5 points) Proposez quelques variables de stratification pertinentes pour cette enquête en justifiant votre choix.

### 2 Enquête sur les clients d'un centre de vacances (12 points)

Les deux gérants d'un grand centre de vacances, Hadrien et Tahar, souhaitent mieux connaître le degré de satisfaction et le profil de leurs clients. Pour ce faire, ils organisent à partir de la saison 2016 une enquête auprès d'un échantillon de vacanciers le jour de leur départ.

Le premier objectif de l'enquête est de mesurer si les vacanciers sont dans l'ensemble satisfaits de leur séjour. Afin d'adapter leur offre de services dans le centre de loisirs, ils souhaitent par ailleurs connaître le montant moyen de dépenses par vacancier (y compris en dehors du centre de vacances).

#### Saison 2016 : Sondage aléatoire simple

**Question 2.1** (1 point) Sachant que Hadrien et Tahar ont l'impression qu'environ 1 vacancier sur 10 n'est pas totalement satisfait de son séjour, déterminez la taille d'échantillon nécessaire pour estimer la part de vacanciers insatisfaits à l'aide d'un sondage aléatoire simple et avec un coefficient de variation de 5 %.

L'enquête est finalement réalisée par sondage aléatoire simple sur 500 vacanciers sur les 10 000 qu'a accueilli le centre de vacances durant la saison 2016 :

- la part de vacanciers insatisfaits  $p$  dans l'échantillon est de 13 % ;
- le montant moyen de dépenses  $\bar{y}$  est de 264 euros par vacancier de l'échantillon et par semaine, avec une variance empirique  $s^2$  de 314 614.

**Question 2.2** (2 points) Calculez l'estimateur d'Horvitz-Thompson et son intervalle de confiance pour la proportion  $P$  de vacanciers insatisfaits et pour la moyenne  $\bar{Y}$  de dépenses par vacancier et par semaine. Que pensez-vous de la précision de l'estimateur de la moyenne des dépenses ?

### Saison 2017 : Sondage aléatoire simple stratifié

À l'issue de l'enquête sur la saison 2016, Hadrien et Tahar souhaitent améliorer leur enquête sans augmenter la taille de l'échantillon. Pour ce faire, ils envisagent de mener un sondage aléatoire simple stratifié :

- Hadrien propose de répartir les vacanciers en deux groupes selon que le jour de leur départ est pair ou impair ;
- Tahar propose de répartir les vacanciers en deux groupes selon que leur location est de catégorie « Standard » ou « Grand luxe ».

**Question 2.3** (1 point) Tahar parvient en définitive à convaincre Hadrien que sa stratification est la plus pertinente eu égard aux objectifs de l'enquête. Expliquez son raisonnement.

Grâce à leur fichier de gestion et aux données de l'enquête menée en 2016, Hadrien et Tahar disposent des informations suivantes :

Strate	$N_{2016}$	$n_{2016}$	$p_{2016}$	$\bar{y}_{2016}$	$s_{2016}^2$	$N_{2017}$
Standard	9 800	480	0,13	150	110	9 800
Grand luxe	200	20	0,05	2 983	150 266	200

À noter que le nombre de total de vacanciers et leur répartition en strate ne varie pas entre 2016 et 2017 ( $N_{2016} = N_{2017}$ ). Hadrien et Tahar proposent d'allouer l'échantillon de deux façons différentes :

- Hadrien propose d'intégrer 2 % de clients « Grand luxe » dans l'échantillon (soit 10 clients sur 500) ;
- Tahar propose d'intégrer 100 clients « Grand luxe » dans l'échantillon.

**Question 2.4** (3 points) Comment désigne-t-on la méthode d'allocation proposée par Hadrien ? En supposant qu'on obtienne les mêmes valeurs pour  $s^2$  qu'en 2016, calculez l'estimateur d'Horvitz-Thompson de la moyenne du montant de dépenses par vacancier et par semaine ainsi que sa variance en 2017.

**Question 2.5** (3 points) En supposant qu'on obtienne les mêmes valeurs pour  $s^2$  qu'en 2016, calculez l'estimateur d'Horvitz-Thompson de la moyenne du montant de dépenses par vacancier et par semaine et sa variance avec l'allocation proposée par Tahar en 2017. Cet estimateur est-il plus ou moins précis que ceux proposés précédemment ? Pourquoi ?

**Question 2.6** (2 points) Toujours en repartant des données de 2016 et pour conclure, quelle allocation proposeriez-vous à Hadrien et Tahar pour maximiser la précision de l'estimateur de la moyenne du montant de dépenses par vacancier et par semaine ? Citez le nom de la méthode utilisée et déterminez les allocations correspondantes (il n'est pas demandé de calculer la variance de l'estimateur obtenu).

### 3 Enquête de la mairie d'Avignon sur la location entre particuliers (5 points)

La mairie d'Avignon cherche à estimer l'importance de la location entre particuliers (du type Airbnb) pour l'accueil des touristes sur le territoire de la commune, en particulier au mois de juillet pendant le festival de création théâtrale.

La responsable statistique de l'agglomération mène une enquête par sondage aléatoire simple de 100 propriétaires avignonnais sur les 10 000 que compte la commune. La principale question de cette enquête porte sur le nombre de nuits du mois de juillet 2017 pendant lesquelles leur logement a été loué à des touristes (on parle de « nuitées touristiques ») par le biais d'une plate-forme du type Airbnb.

La responsable fait cependant face à un niveau élevé de non-réponse : sur 100 questionnaires envoyés, seuls 50 sont retournés par les propriétaires. 40 propriétaires indiquent ne pas avoir loué leur logement de particulier à particulier, les 10 autres pris ensemble totalisent 40 nuitées touristiques au mois de juillet 2017.

**Question 3.1** (1 point) Quel(s) problème(s) cette non-réponse pose-t-elle pour l'estimation du nombre total de nuitées touristiques louées de particulier à particulier? Sans correction de la non-réponse, à combien s'élèverait l'estimateur du nombre total de nuitées touristiques louées de particulier à particulier à Avignon au moins de juillet 2017?

La responsable statistique se propose dans un premier temps de mener une correction par pondération globale (sans distinguer de groupes de réponses homogènes).

**Question 3.2** (1 point) Quel est le poids des logements répondants après cette correction de la non-réponse? Quel est l'estimateur du nombre total de nuitées touristiques louées de particulier à particulier à Avignon au mois de juillet 2017 après cette correction?

Des méthodes de *web-scraping* sont utilisées pour identifier quels logements de l'échantillon sont disponibles sur des plates-formes de location entre particuliers. On récupère par ce biais la localisation précise dans la ville (centre-ville ou périphérie) des logements disponibles sur des plates-formes de location entre particuliers. En revanche, le nombre de nuitées louées en juillet 2017 n'est pas une information directement disponible sur les sites.

**Question 3.3** (2 points) Décrivez précisément la ou les méthodes de correction de la non-réponse que la responsable statistique pourrait désormais mettre en œuvre.

À l'approche de l'édition 2018 du festival d'Avignon, les plates-formes de location entre particuliers publient le nombre total de nuitées réservées par leur intermédiaire au mois de juillet 2017. Ce nombre de nuitées est bien supérieur à celui obtenu par la responsable statistique de l'agglomération à partir de l'enquête.

**Question 3.4** (1 point) Comment expliquer ce phénomène? En particulier, à quel mécanisme de non-réponse semble correspondre la situation rencontrée dans cette enquête?

# Annexe : Présentation de l'enquête sur la fréquentation touristique dans l'hôtellerie de l'Insee

Les éléments suivants sont extraits de la page : <https://www.insee.fr/fr/metadonnees/source/s1242>.

## Présentation de l'opération

**Organisme producteur** : Insee - Pôle de compétence Tourisme - Direction régionale de Languedoc-Roussillon

**Type d'opération** : Enquête

**Objectifs** : L'enquête de fréquentation dans l'hôtellerie a pour objectif l'observation conjoncturelle de la fréquentation dans l'hôtellerie de tourisme et l'étude de la structure de la clientèle, notamment de son origine géographique. Les informations sont produites au plan national, régional, voire infra-régional.

**Thèmes abordés** : Le parc est mesuré au 31 décembre de chaque année et la fréquentation sur chaque mois. Les principaux thèmes abordés sont les suivants :

- caractéristiques de l'établissement ;
- capacité d'accueil ;
- occupation quotidienne des chambres ;
- arrivées et nuitées selon le pays de résidence des touristes ;
- durée moyenne de séjour ;
- proportion de tourisme d'affaires.

## Champ de l'opération

**Champ géographique** : L'enquête couvre la métropole et les départements d'outre-mer (DOM).

**Champ(s) statistique(s) couvert(s)** : L'enquête couvre les activités définies par les codes NAF rév. 2 : 55.10Z (hôtels et hébergement similaire).

## Caractéristiques techniques

**Unité statistique enquêtée** : Hôtels à vocation touristique, qu'ils soient classés ou non

**Périodicité de l'opération** : Mensuelle

**Période et mode de collecte** : La collecte est réalisée par voie postale.

**Autres spécifications** :

- L'enquête est réalisée auprès d'un échantillon national d'environ 6 700 hôtels (12 000 y compris les extensions régionales).
- Concernant l'extension aux départements d'outre-mer (DOM), deux questionnaires spécifiques ont été mis en place : un pour les trois DOM antillais et un pour la Réunion. Elle permet ainsi une observation statistique du tourisme sur l'ensemble du territoire français.
- Cette enquête rentre dans le cadre du règlement européen 692/2011 concernant la production d'informations statistiques sur le tourisme à Eurostat. Elle répond également aux besoins nationaux de la direction du tourisme et locaux pour les acteurs et décideurs régionaux du tourisme.