

**Méthode de détection et de
traitement des individus
atypiques**

—

Cas de l'enquête EVE

24 mars 2009

JMS – Session 8

**Claire Jacod
Julien Valentino**

1. Les recettes de la ligne voyages

- Évaluées à partir de l'**Enquête auprès des Voyageurs Étrangers (EVE)** :
 - trimestrielle
 - réalisée en partenariat avec l'**Insee** et sous-traitée à **TNS-Sofres**
 - cible les **visiteurs non-résidents** quittant la France quelle que soit la durée de leur séjour
 - cette population n'est **pas connue** : **estimée**
données de trafic
part des non-résidents dans le trafic

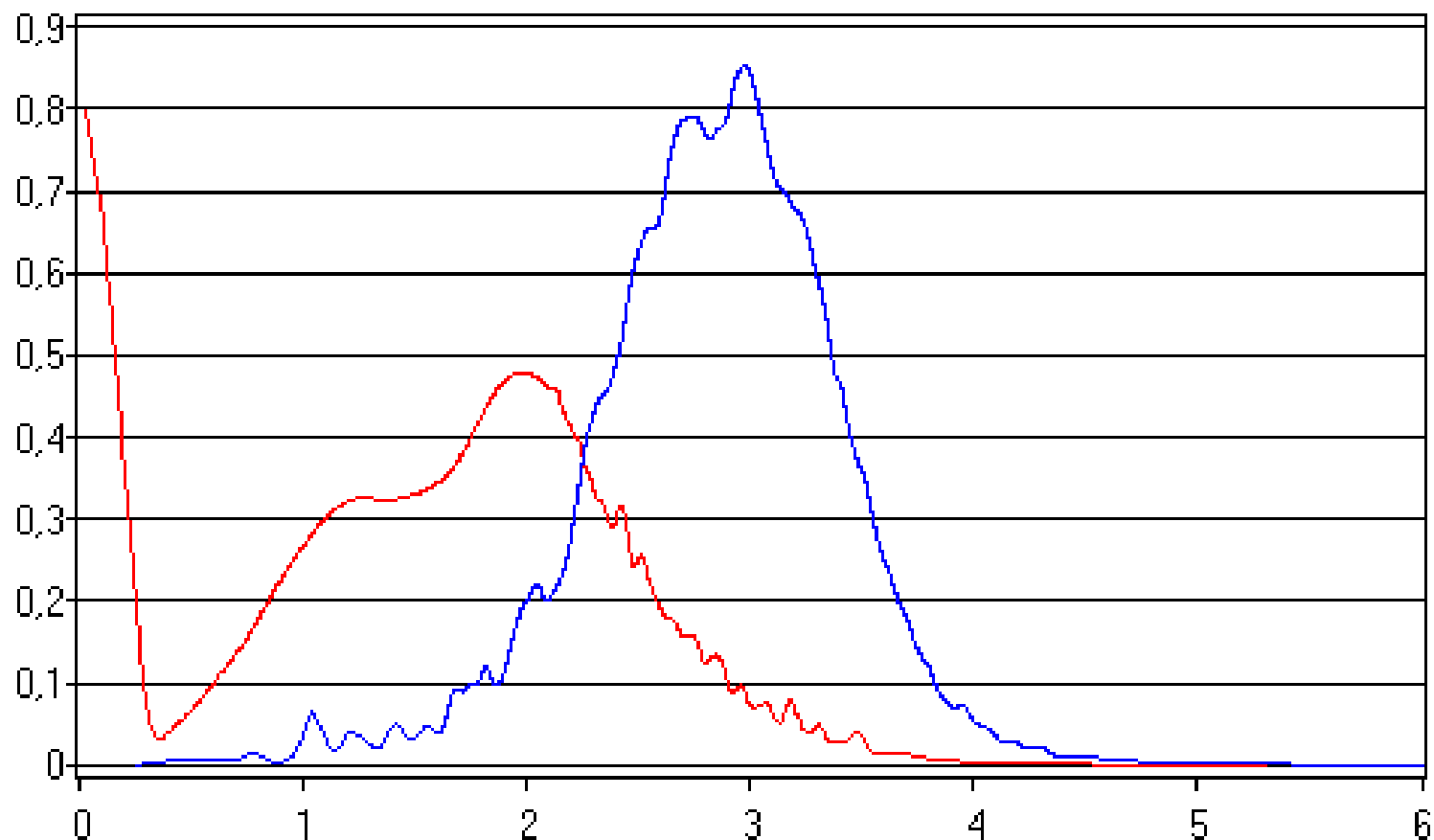
2. Rappels sur EVE

- Interrogation à différents points de **sortie du territoire**, selon le **mode de transport** (air, rail, route, mer)
- Le questionnaire interroge sur :
 - les **dépenses**
 - la **durée** de leur **séjour**
 - le **nombre de personnes** participant au séjour

3. Caractéristiques des dépenses

Densité des dépenses totales des visiteurs - période 2007T4-2008T3 - avant traitement

Densité estimée



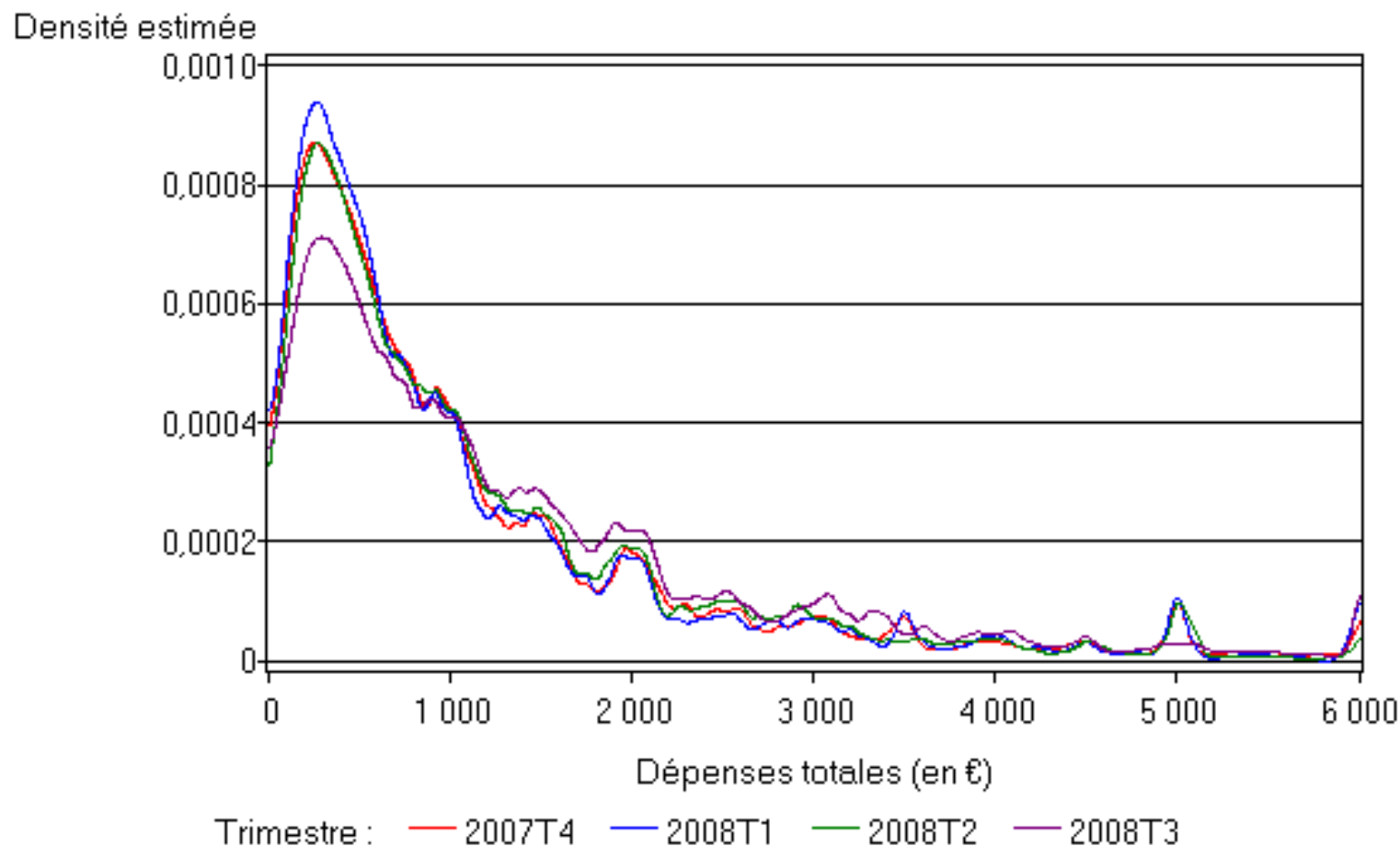
Dépenses totales (+ 1€) - échelle logarithmique (base 10)

Type de visiteur : — Excursionnistes — Touristes

3. Caractéristiques des dépenses

- Pics sur les valeurs rondes

Densité des dépenses totales des touristes - avant traitement



- Quelques déclarations de dépenses très élevées

4. Problèmes d'individus atypiques

- Échec des méthodes « classiques », car :
 - **caractéristiques des distributions** (multimodales, pics dans des valeurs très élevées)
 - population d'intérêt **mal connue** et très **hétérogène**
 - **pas d'informations sociodémographiques** sur les visiteurs

5. Présentation de la méthode

- Pour chaque **type de visiteur** et chaque trimestre, **2 seuils** :
 - **dépenses totales**
 - **dépenses par nuitée ou par personne**
- Questionnaire **atypique** : **dépenses supérieures à au moins un des seuils**

5. Présentation de la méthode

- Détecter la **1^e plage de dépenses** suffisamment large pour laquelle il n'y a **pas de questionnaire correspondant**
- Seuil = **1^e valeur d'annulation de la densité** des dépenses totales ou par nuitée ou personne

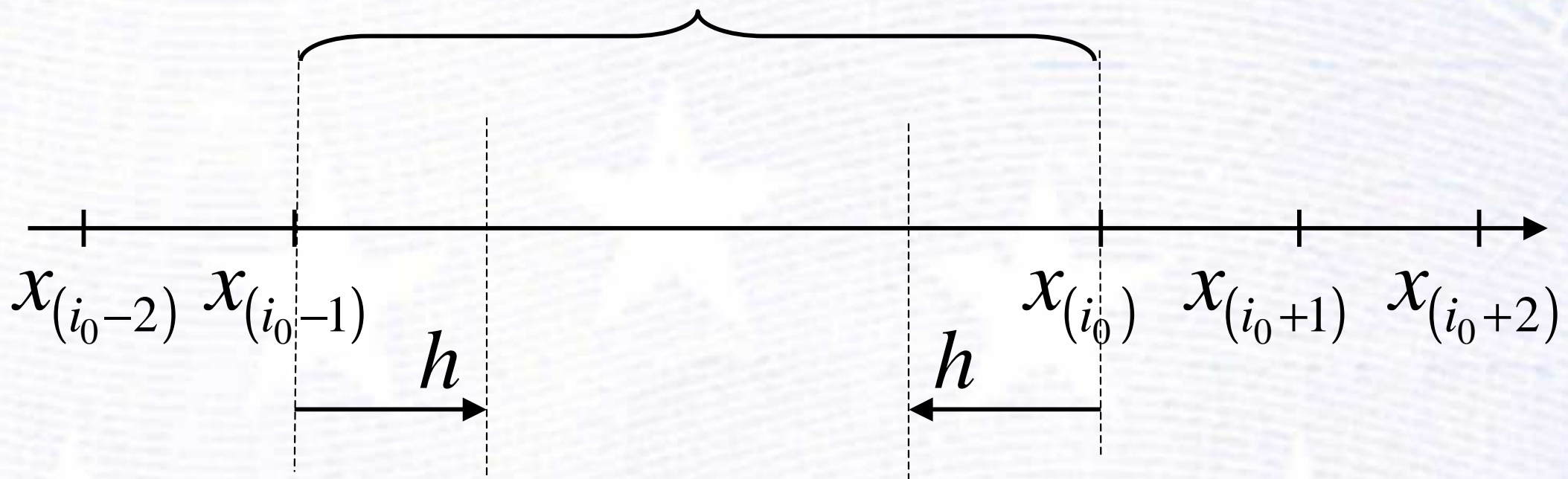
6. Enchaînement des calculs - phase 1

- On trie les dépenses par **ordre croissant**
- 1^{er} point d'annulation de la densité : entre les 2 premières observations **significativement espacées**



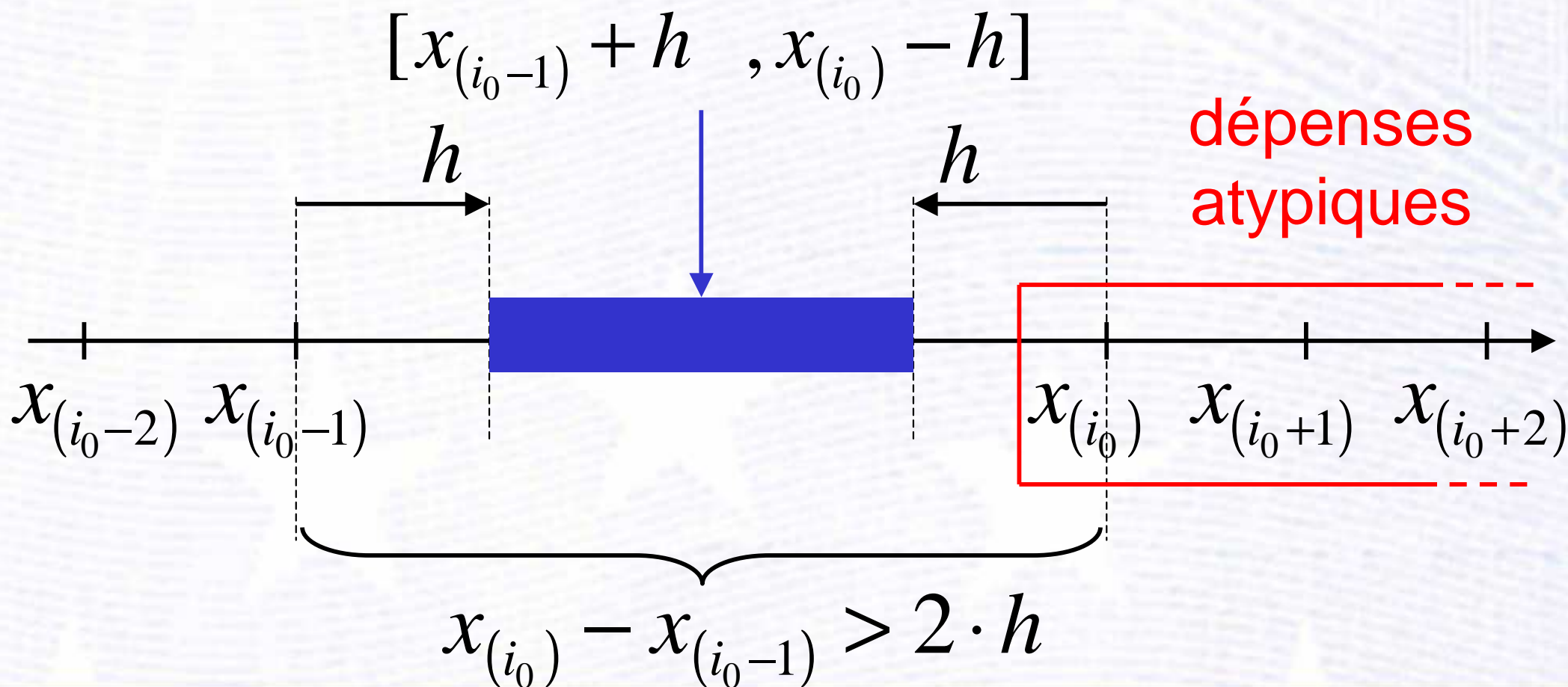
6. Enchaînement des calculs - phase 1

- Fenêtre : $h = 0,75 \cdot \left(\frac{n}{100}\right)^{-1/5} \cdot \min_{i=1..[\frac{n}{2}]} \left(x_{\left(i+\left[\frac{n}{2}\right]\right)} - x_{(i)} \right)$
- Si $x_{(i_0)} - x_{(i_0-1)} > 2 \cdot h$

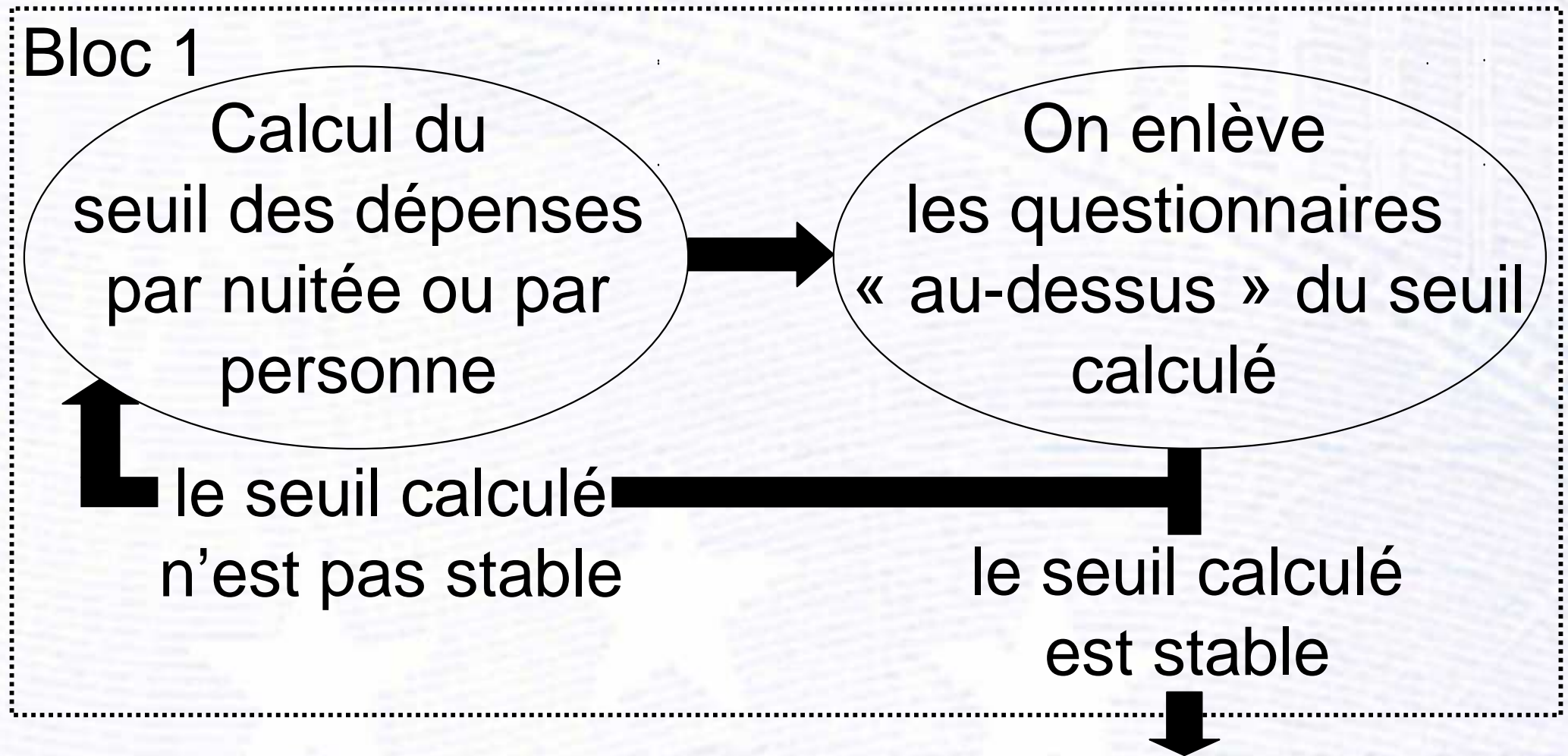


6. Enchaînement des calculs - phase 1

- Alors l'estimation de la densité est nulle sur l'intervalle

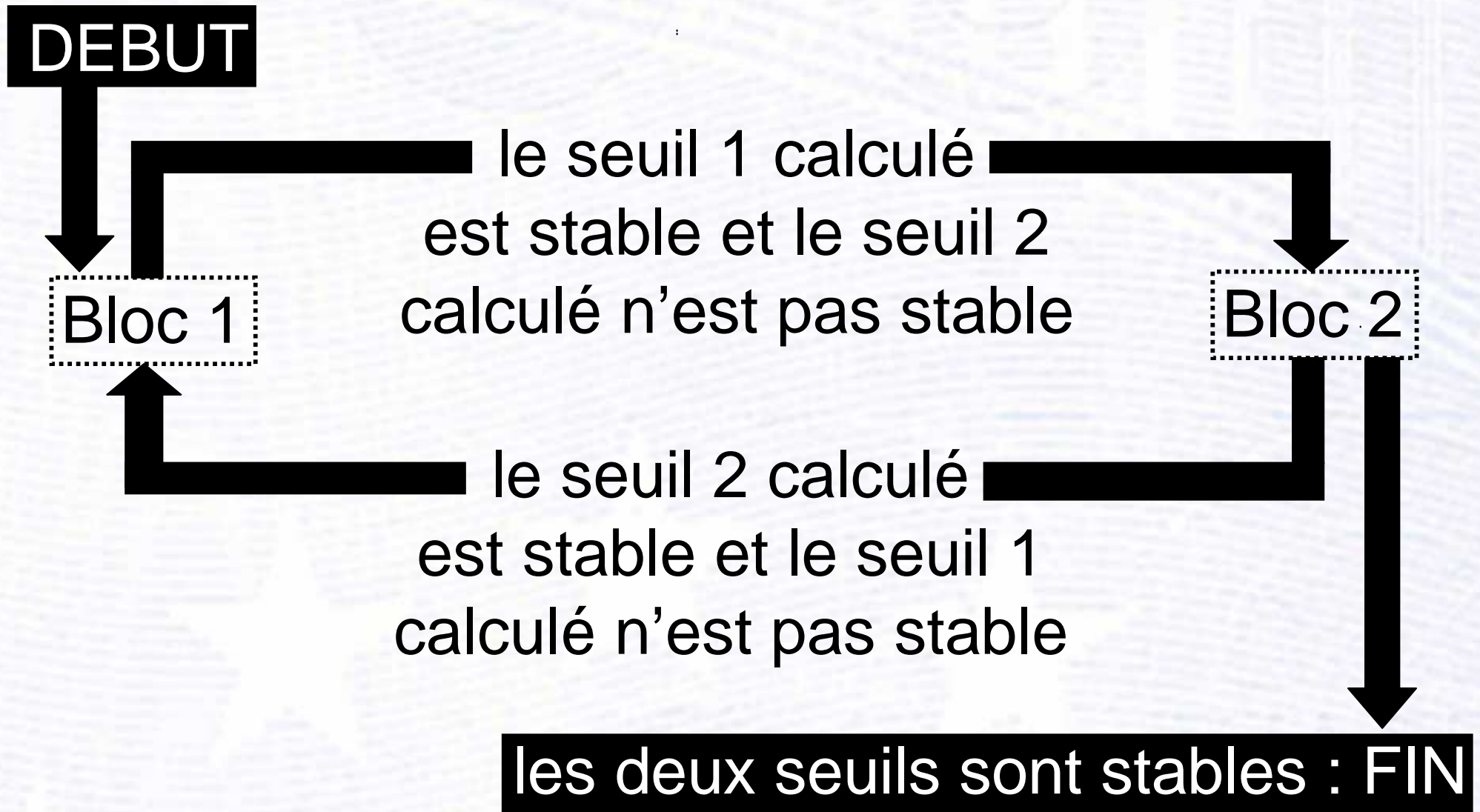


6. Enchaînement des calculs - phase 2



- Bloc 2 : même chose pour les **dépenses totales**

6. Enchaînement des calculs - phase 2



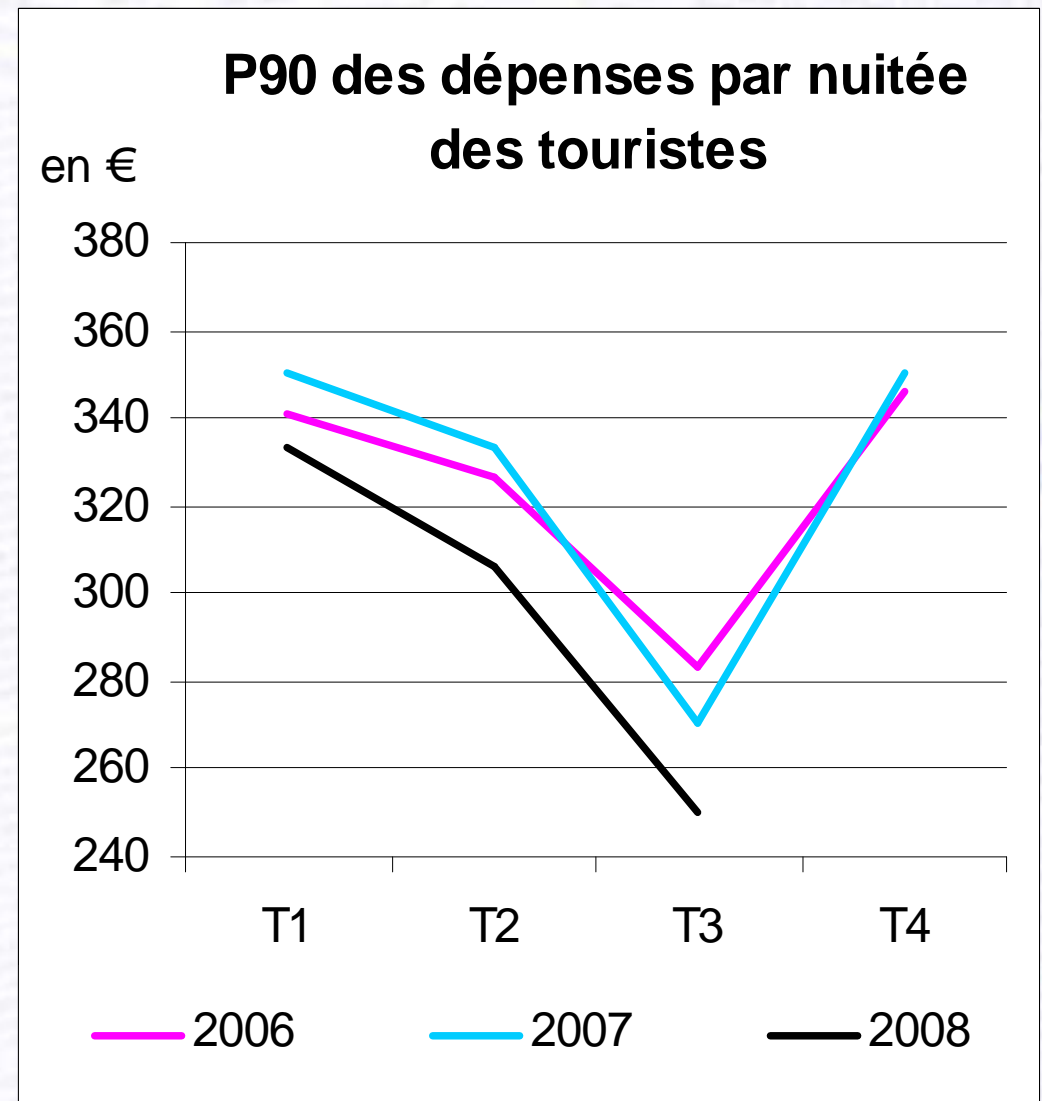
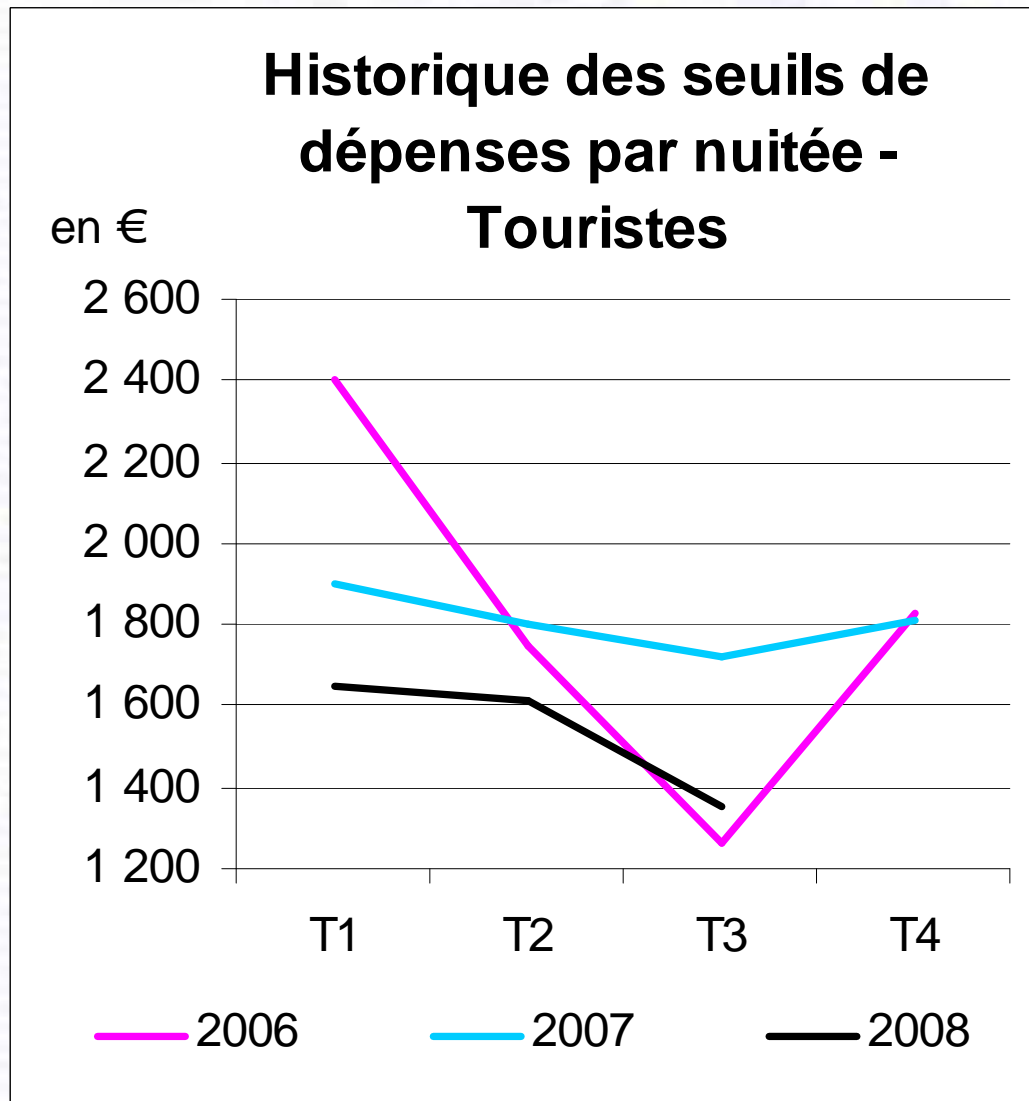
6. Enchaînement des calculs - phase 3

- Pour régler le **problème des valeurs rondes** : la **dépense totale** de chaque questionnaire est **multipliée par $(1+\varepsilon)$** , où :

$$\varepsilon \sim \mathcal{N}(0 ; 0,05)$$

- L'ensemble du processus est **itéré 31 fois**
- **Jeu de seuils** retenu : « **médiane** » des 31 itérations

7. Résultats – seuils



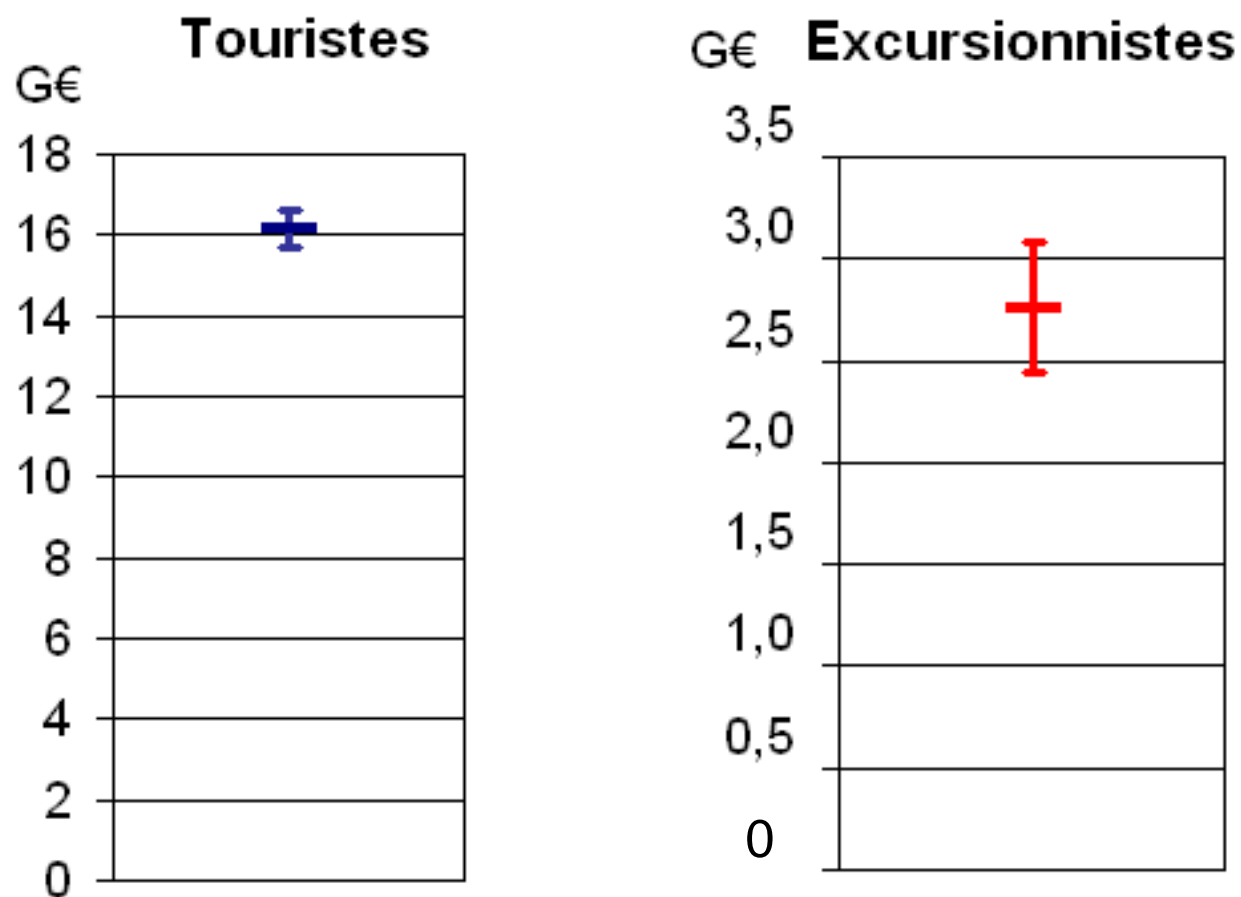
7. Résultats – dépenses

- **Questionnaires dépassant un des seuils :**
 - **imputés** par TNS-Sofres
 - **contrainte** de ne pas dépasser les seuils calculés
- Impact des seuils sur les dépenses totales : exemple du 3^e trimestre 2008

Type de visiteur	Dépenses totales extrapolées (G€)		Ecart (%)	Nombre de questionnaires atypiques (%)
	Avant seuils	Après seuils		
Excursionnistes	5,2	3,6	30	2,7
Touristes	51,0	40,3	21	1,0

7. Résultats – qualité des dépenses

- Après calcul des seuils, 1^{er} trimestre 2008 :
Dépenses totales – moyenne et écart-type sur 50 échantillons bootstrap



8. Propriétés de la méthode

- Propriété de la fenêtre : un questionnaire **rejeté** et un questionnaire **conservé** ne peuvent **pas avoir des dépenses proches**
1^{er} trimestre 2008 :

Type de visiteur	Ecart en euros
Excursionnistes	61
Touristes	312

- Les questionnaires qualifiés d'atypiques le restent après **transformation $aX+b$**

8. Propriétés de la méthode

- Les 2 seuils permettent de détecter les questionnaires aux **fortes dépenses** de façon « **absolue** » et « **relative** »
- Prise en compte de la **saisonnalité** et des **effets conjoncturels** sur la distribution des dépenses
- La méthode retenue **simple** à implémenter et à utiliser
- Compatible avec le travail d'imputation de TNS-Sofres

Merci de votre attention

-

Des questions ?