



Méthodologie d'analyse comparative des prix des pièces automobiles par Ovoko

La méthodologie suivante décrit l'approche systématique utilisée pour analyser la différence de prix entre les composants automobiles neufs et d'occasion disponibles sur la plateforme **Ovoko** et les marchés comparables de pièces neuves en France.

L'analyse couvre à la fois les pièces d'origine et les pièces de rechange, la TVA étant incluse et les frais d'expédition exclus de tous les calculs de prix. La méthodologie précise également la manière dont ces facteurs ont été traités afin de garantir la cohérence et la transparence du processus de comparaison.

1. Collecte initiale des données et sélection de l'échantillon

L'ensemble de données de base, **D**, comprenait un total de $N = 5,9 \times 10^6$ codes de pièces uniques. Une approche d'échantillonnage stratifié, basée sur la fréquence des ventes, a été utilisée pour sélectionner un échantillon représentatif des pièces fréquemment commercialisées.

- **Taille de la population** : $N = 5,872,117$ (total des codes de pièces).
- **Taille de l'échantillon** : $n = 700$ (codes de pièces les plus fréquemment achetés).

Les critères de sélection de l'échantillon $\mathbf{S} = \{s_1, s_2, \dots, s_{700}\}$ ont été définis par le **Volume des Ventes** (V_i) pour chaque code de pièce i au cours des 12 derniers mois sur la plateforme Ovoko. L'échantillon **S** comprend les 751 pièces ayant le volume le plus élevé V_i .

La taille de l'échantillon de $n = 700$ est suffisante car, pour une population aussi importante que $N = 5,872,117$, c'est la **taille absolue de l'échantillon**, et non la proportion (ex. $700/5.87$ million), qui détermine principalement la **marge d'erreur (MOE)**. En supposant un scénario pessimiste conservateur dans lequel la caractéristique étudiée présente une variabilité maximale ($p = 0.5$), une taille d'échantillon de $n = 700$ donne une **marge d'erreur d'environ**

$\pm 3.7\%$ à un niveau standard de confiance de 95%. Cette marge est calculée à l'aide de la formule suivante :

$$\text{MOE} = Z \cdot \sqrt{\frac{p(1-p)}{n}}$$

Étant donné que l'échantillonnage est **stratifié** pour cibler les pièces avec le V_i le plus haut, la taille réelle de l'échantillon requise pour le sous-ensemble de la population cible est probablement *inférieure* à ce que suggère ce calcul général, car la stratification des ventes élevées réduit la variabilité globale (p est probablement loin de 0.5) pour les pièces les plus critiques, ce qui donne une MOE effective encore plus faible.

2. Acquisition des données relatives aux prix des pièces neuves

Pour chaque pièce $s_j \in \mathbf{S}$, les données relatives au prix des pièces neuves ont été obtenues auprès des principales places de marché françaises en ligne et à partir des résultats des moteurs de recherche les mieux classés pour les mots-clés français pertinents. Pour chaque pièce s_j , un ensemble de k_j d'observations de prix individuels a été collecté, désigné par $\mathbf{P}_{\text{new},j} = \{p_{j,1}, p_{j,2}, \dots, p_{j,k_j}\}$.

Remarque : Pour obtenir des informations détaillées sur les plateformes incluses dans cette comparaison de prix, veuillez contacter research@ovoko.com.

3. Nettoyage des données et calcul des prix moyens fiables des pièces neuves

Afin de garantir la fiabilité du prix moyen des pièces neuves, un processus en deux étapes intégrant la détection des valeurs aberrantes via la méthode de **l'écart type (σ)** a été utilisé.

3.1. Calcul du prix moyen initial

Le **prix moyen initial de la pièce** s_j , noté $\bar{p}_{j,\text{init}}$, a été calculé à partir de toutes les observations de prix collectées $\mathbf{P}_{\text{new},j}$:

$$\bar{p}_{j,\text{init}} = \frac{1}{k_j} \sum_{i=1}^{k_j} p_{j,i}$$

3.2. Détection et filtrage des valeurs aberrantes

Les valeurs aberrantes ont été définies comme des prix dépassant **deux écarts-types** par rapport à la moyenne initiale. L'**écart-type** $\sigma_{j,\text{init}}$ de l'ensemble de prix initial $\mathbf{P}_{\text{new},j}$ a été

calculé comme suit :

$$\sigma_{j,\text{init}} = \sqrt{\frac{1}{k_j-1} \sum_{i=1}^{k_j} (p_{j,i} - \bar{p}_{j,\text{init}})^2}$$

Les prix $p_{j,i}$ ont été supprimés s'ils remplissaient la condition suivante :

$$|p_{j,i} - \bar{p}_{j,\text{init}}| > 2\sigma_{j,\text{init}}$$

L'ensemble des prix nettoyés est noté $\mathbf{P}'_{\text{new},j}$. Soit k'_j la taille de l'ensemble nettoyé.

3.3. Recalcul du prix moyen fiable

Le **prix moyen fiable de la pièce neuve** s_j , noté $\bar{p}_{j,\text{new}}$, a été calculé à partir de l'ensemble filtré $\mathbf{P}'_{\text{new},j}$:

$$\bar{p}_{j,\text{new}} = \frac{1}{k'_j} \sum_{p \in \mathbf{P}'_{\text{new},j}} p$$

4. Calcul du prix global

Cette étape consistait à calculer le prix moyen représentatif unique pour l'ensemble de l'échantillon \mathbf{S} , tant pour les pièces neuves que pour les pièces d'occasion.

4.1. Prix moyen global des pièces neuves

Le **prix moyen global des pièces neuves** (\bar{P}_{new}) a été calculé comme étant la moyenne des prix moyens fiables de toutes les pièces $n = 700$ de l'échantillon :

$$\bar{P}_{\text{new}} = \frac{1}{n} \sum_{j=1}^{700} \bar{p}_{j,\text{new}}$$

4.2. Prix moyen global des pièces d'occasion

Un processus analogue de collecte, de nettoyage et de calcul de la moyenne des données a été appliqué aux prix des pièces d'occasion provenant directement de la plateforme Ovoko pour le même échantillon \mathbf{S} , ce qui a permis d'obtenir un ensemble de prix moyens fiables $\bar{p}_{j,\text{used}}$ pour chaque pièce d'occasion s_j .

Le **prix moyen global des pièces d'occasion** (\bar{P}_{used}) a été calculé comme suit :

$$\bar{P}_{\text{used}} = \frac{1}{n} \sum_{j=1}^{700} \bar{p}_{j,\text{used}}$$

Exemple de calcul (illustratif pour une pièce, s_1) :

Supposons que pour une seule pièce s_1 , les nouveaux prix collectés soient :

$\mathbf{P}_{\text{new},1} = \{100, 110, 105, 120, 300\}$ (Euros).

1. **Moyenne initiale** : $\bar{p}_{1,\text{init}} = (100 + 110 + 105 + 120 + 300)/5 = 735/5 = 147$
2. **Initial σ** : $\sigma_{1,\text{init}} \approx 81.3$
3. **Seuil d'aberration** : $2 \times \sigma_{1,\text{init}} \approx 162.6$.
 - $|300 - 147| = 153$. Puisque $153 < 162.6$, la valeur 300 ne serait **pas** supprimée si l'on utilisait 2σ de manière stricte. *Cependant, afin d'illustrer clairement le processus de nettoyage, supposons qu'un seuil légèrement plus strict ait été appliqué ou que la valeur ait été encore plus élevée.*
 - Si la valeur était : 400: $|400 - 147| = 253 > 162.6$. 400 serait supprimée.
 - Supposons que la valeur aberrante 300 **soit supprimée** pour clarifier les étapes restantes. $\mathbf{P}'_{\text{new},1} = \{100, 110, 105, 120\}$.
4. **Prix moyen fiable** : $\bar{p}_{1,\text{new}} = (100 + 110 + 105 + 120)/4 = 435/4 = 108.75$

5. Comparaison des prix et analyse quantitative

La dernière étape consistait en une comparaison quantitative directe des prix moyens globaux calculés.

La **différence de prix absolue** (ΔP_{abs}) entre les pièces neuves et d'occasion est la suivante :

$$\Delta P_{\text{abs}} = \bar{P}_{\text{new}} - \bar{P}_{\text{used}}$$

La **différence de prix relative (pourcentage d'économie)** ($\Delta P_{\%}$) quantifie l'économie potentielle moyenne réalisée en achetant une pièce d'occasion :

$$\Delta P_{\%} = \left(1 - \frac{\bar{P}_{\text{used}}}{\bar{P}_{\text{new}}}\right) \times 100\%$$

Cette analyse fournit une mesure statistiquement fiable de la différence de prix moyenne sur la base d'un échantillon pondéré des ventes sur le marché des pièces automobiles.