

Thierry Hai Son ADAM

Relation entre consommation d'alcool et taux d'homicides en Europe

Une analyse multidimensionnelle

Préambule	3
Problématique	3
Hypothèse préalable	3
Définition des variables	3
Collecte des données	3
Méthodologie	3
Analyses Statistiques	4
Organisation et visualisation des données	4
Analyse des données collectées	4
Tendance centrale	4
Variation	5
Forme de la distribution	5
Corrélation et causalité	6
Matrice de corrélation de Pearson	6
Corrélations non paramétriques	7
Analyses de régression	7
Régression linéaire simple	7
Régression multiple	7
Analyse des outliers	8
Régression linéaire sur données normalisées et sans outliers	8
Modèles non linéaires	8
Régression polynomiale	8
Régression avec splines cubiques restreintes	9
Transformation log-log	9
Régression linéaire par morceaux (Piecewise Linear Regression)	9
Modèles d'apprentissage automatique	9
Analyses par pays et par année	9
Variations temporelles	10
Analyses régionales	10
Analyses par pays	10
Discussion	10
Interprétation des résultats	10
Limites méthodologiques	11
Comparaison avec les études antérieures	11
Recommandations pour les recherches futures	11
Conclusion	12
Annexes	13
Bibliographie	21

Préambule

Cette étude examine la relation entre la consommation d'alcool par habitant et le taux d'homicides (pour 100 000 habitants) dans 41 pays européens au cours des années 2000, 2005, 2010 et 2019. La méthodologie suit une approche structurée de définition des variables, collecte de données, organisation, visualisation et analyse (DCOVA). Des analyses non linéaires ont été utilisées pour détecter des relations complexes entre les variables.

Problématique

Cette étude cherche à déterminer s'il existe une relation statistique significative entre la consommation d'alcool par habitant et le taux d'homicides dans 41 pays européens pour les années 2000, 2005, 2010 et 2019. L'objectif principal est de voir si une association existe, qu'elle soit causale ou simplement corrélative, en étudiant à la fois les relations linéaires et non linéaires.

Hypothèse préalable

D'après la littérature scientifique et les rapports de santé publique, l'alcool est souvent associé à des comportements violents. L'hypothèse testée ici est qu'une augmentation de la consommation d'alcool pourrait correspondre à une hausse des taux d'homicides. Toutefois, il est possible que cette relation ne soit pas simplement linéaire et qu'elle dépende également de facteurs socio-économiques.

Définition des variables

L'analyse s'appuie sur quatre variables principales :

- **Consommation d'alcool** (variable indépendante) : Quantité totale d'alcool pur consommé par habitant, exprimée en litres par an pour les personnes âgées de plus de 15 ans.
- **Taux d'homicides** (variable dépendante) : Nombre d'homicides pour 100 000 habitants.
- **Variables de contrôle** :
 - PIB par habitant (en dollars américains)
 - Taux de chômage (en pourcentage)

Ces variables permettent d'examiner la relation potentielle entre la consommation d'alcool et les niveaux de violence, tout en prenant en compte d'autres facteurs économiques qui pourraient influencer sur les résultats.

Collecte des données

Les données proviennent de la base Our World in Data, couvrant les années 2000, 2005, 2010 et 2019 pour 41 pays européens. Le jeu de données comprend 157 observations complètes. Les valeurs manquantes ont été exclues de l'analyse lorsqu'elles étaient indisponibles pour une année donnée. Cette méthode assure une certaine cohérence temporelle tout en préservant la validité des résultats.

Méthodologie

Pour évaluer la relation entre la consommation d'alcool par habitant et les taux d'homicides, nous avons utilisé plusieurs méthodes statistiques, allant des corrélations simples aux modèles non linéaires avancés.

Analyses Statistiques

Les analyses statistiques ont été réalisées à l'aide de deux outils principaux :

- **Excel** : Utilisé pour des statistiques descriptives, des régressions linéaires, la normalisation des données et les tests de Chi².
- **Python** : Employé pour vérifier les résultats des régressions linéaires et développer des modèles non linéaires avancés, en s'appuyant sur des bibliothèques telles que pandas, numpy, scikit-learn et statsmodels.

Les modèles testés incluent :

1. Régressions linéaires simples et multiples
2. Régression polynomiale (degrés 2, 3 et 4)
3. Modèles utilisant des splines cubiques
4. Régression LOWESS (régression locale)
5. Transformations logarithmiques
6. Support Vector Regression (SVR)
7. Random Forest Regression
8. K-Nearest Neighbors (KNN) Regression
9. Régression linéaire par morceaux (Piecewise Linear Regression)

Organisation et visualisation des données

Les données ont été organisées par année, par pays, et par type de variable (consommation d'alcool, taux d'homicides, PIB par habitant, taux de chômage). Des tableaux et graphiques ont été utilisés pour faciliter la compréhension des tendances et des écarts observés.

La présence de données manquantes pour certaines années ou certains pays, notamment ceux d'Europe de l'Est, représente une limite importante. Cette hétérogénéité pourrait introduire un biais et doit être prise en compte lors de l'interprétation des résultats.

Analyse des données collectées

L'analyse a été structurée autour de trois aspects : la tendance centrale, la dispersion, et la forme des distributions. Nous avons calculé des moyennes, médianes, écarts-types, ainsi que des mesures d'asymétrie et d'aplatissement pour évaluer les distributions.

L'étape suivante consistera à reformuler la section sur l'analyse statistique.

Tendance centrale

Les mesures de tendance centrale permettent de déterminer la valeur typique autour de laquelle les données se regroupent :

Variable	Moyenne	Médiane	Interprétation
Consommation d'alcool (litres/an)	10.16	10.68	La moyenne est affectée par des valeurs extrêmes (notamment en Russie). La médiane légèrement supérieure indique une distribution penchée à gauche.
Taux d'homicides (pour 100 000 habitants)	2.78	1.39	Moyenne influencée par des valeurs extrêmes (Russie, Kazakhstan, Lituanie). La médiane plus faible montre une distribution penchée à droite.
PIB par habitant (\$)	39043.20	36964.26	Distribution relativement équilibrée avec quelques valeurs élevées.
Taux de chômage (%)	9.14	7.56	Distribution asymétrique vers la droite, avec quelques pays présentant des taux de chômage très élevés.

Variation

Mesures évaluées : Étendue, Variance, Écart-type.

Variable	Étendue	Variance	Écart-type	Interprétation
Consommation d'alcool	17.48 (19.05 - 1.57)	10.96	3.31	Forte disparité dans la consommation d'alcool entre les pays européens. Dispersion modérée autour de la moyenne.
Taux d'homicides	27.70 (27.98 - 0.28)	14.42	3.80	Variabilité importante entre les pays, accentuée par des valeurs proches de zéro.
PIB par habitant	129516.63	594,597,891	24380.21	Très grande disparité économique entre les pays européens.
Taux de chômage	31.63	30.81	5.55	Écarts considérables dans les situations du marché du travail.

Forme de la distribution

L'analyse de la skewness (asymétrie) et de la kurtosis (aplatissement) permet d'évaluer la répartition des données par rapport à la moyenne et leur déviation par rapport à une distribution normale.

Variable	Skewness	Interprétation	Kurtosis	Interprétation
Consommation d'alcool (litres/an)	-0.49	Légère asymétrie à gauche, indiquant une distribution globalement symétrique.	0.49	Distribution légèrement mésokurtique, proche de la normale avec des queues légèrement plus lourdes.
Taux d'homicides (pour 100 000 hab.)	3.81	Asymétrie forte à droite, due à des pays présentant des taux exceptionnellement élevés.	19.26	Distribution hautement leptokurtique : concentration autour de la moyenne avec des valeurs extrêmes très marquées.
PIB par habitant	1.56	Asymétrie à droite, avec quelques pays très riches tirant la distribution.	2.24	Distribution légèrement leptokurtique.
Taux de chômage	1.53	Asymétrie à droite, avec quelques pays à chômage élevé.	2.52	Distribution modérément leptokurtique.

Corrélation et causalité

Cette section examine la relation entre la consommation d'alcool et le taux d'homicides en utilisant plusieurs mesures de corrélation.

Matrice de corrélation de Pearson

Variable	Consommation d'alcool	Taux d'homicide	PIB par habitant	Taux de chômage
Consommation d'alcool	1	0.106	0.198	-0.285
Taux d'homicide	0.106	1	-0.381	0.151
PIB par habitant	0.198	-0.381	1	-0.510
Taux de chômage	-0.285	0.151	-0.510	1

Interprétation :

- Corrélation très faible (0.106) entre consommation d'alcool et taux d'homicide
- Corrélation négative modérée (-0.381) entre PIB par habitant et taux d'homicide
- Corrélation positive faible (0.151) entre taux de chômage et taux d'homicide
- Corrélation positive faible (0.198) entre consommation d'alcool et PIB par habitant

Ces résultats indiquent que les facteurs socio-économiques (PIB et chômage) semblent davantage liés aux taux d'homicides que la consommation d'alcool (cf. Annexe 1)

Corrélations non paramétriques

Pour tenir compte des distributions non normales, des mesures de corrélation non paramétriques ont été calculées :

- **Corrélation de Spearman** : 0.054 (p-value = 0.499)
- **Tau de Kendall** : 0.037 (p-value = 0.493)

Ces tests confirment qu'il n'y a pas de relation significative entre la consommation d'alcool et le taux d'homicides (cf. Annexe 2)

Test du Chi2

Le test du Chi² a donné une statistique de 2.68 avec 4 degrés de liberté et une p-value de 0.613, indiquant l'absence de relation statistiquement significative entre la consommation d'alcool et le taux d'homicides (cf. Annexe 3 et 4).

Analyses de régression

Régression linéaire simple

L'analyse initiale a consisté à appliquer une régression linéaire simple entre la consommation d'alcool par habitant et le taux d'homicides :

- Coefficient (pente) : 0.122
- Ordonnée à l'origine : 1.545
- R^2 : 0.011
- p-value : < 0.001

Ces résultats montrent que la consommation d'alcool explique une très faible proportion de la variance du taux d'homicides (1,1 %), ce qui est négligeable (voir Annexe 5)

Régression multiple

Une régression multiple incluant le PIB par habitant et le taux de chômage comme variables de contrôle a été réalisée pour améliorer la qualité du modèle :

- R^2 : 0.180
- R^2 ajusté : 0.164
- F-statistique : 11.18 (p-value : 1.13e-06)

Coefficients:

- Consommation d'alcool : 0.214 (p-value : 0.016)
- PIB par habitant : -0.000066 (p-value : < 0.001)
- Taux de chômage : -0.009 (p-value : 0.886)

L'ajout de variables socio-économiques améliore significativement le modèle. Le PIB par habitant est la variable la plus significative, montrant une relation négative avec le taux d'homicides. En revanche, le taux de chômage n'a pas d'effet significatif.

Analyse des outliers

Un examen des valeurs extrêmes a identifié trois observations avec des z-scores supérieurs à 3 :

Pays	Année	Consommation d'alcool	Taux d'homicide
Kazakhstan	2000	6.64	15.26
Russie	2005	14.50	24.78
Russie	2000	13.86	27.98

L'analyse de régression sans ces valeurs extrêmes a donné un R^2 encore plus faible (0.0024), ce qui confirme l'absence de lien linéaire significatif.

Régression linéaire sur données normalisées et sans outliers

Les données normalisées et sans outliers montrent une légère amélioration dans la distribution des taux d'homicides (skewness réduite à 1.23, kurtosis réduite à 0.89). Toutefois, la consommation d'alcool normalisée présente une asymétrie accrue (skewness de -1.70). Malgré un modèle graphiquement plus cohérent, les résultats statistiques demeurent non significatifs (p-value = 0.44, R^2 = 0.0038). Aucune amélioration n'a été obtenue après normalisation ou suppression des outliers (cf. Annexes 6).

Il est important de noter que la violation de l'hypothèse de normalité n'affecte généralement pas les inférences sur de grands échantillons^{(Zeileis et al., 2008, 27(8), 1–25); Schmidt & Finan, 2018, 146–51)}, ce qui corrobore davantage l'insignifiance statistique observée dans nos résultats.

Modèles non linéaires

Plusieurs modèles non linéaires ont été testés pour identifier des relations complexes entre la consommation d'alcool et le taux d'homicides.

Régression polynomiale

Degré	R ²	RMSE	Amélioration significative
1 (linéaire)	0.0113	3.7644	-
2 (quadratique)	0.0229	3.7422	Non (p=0.1781)
3 (cubique)	0.0387	3.7119	Non (p=0.1153)
4 (quartique)	0.0696	3.6518	Oui (p=0.0260)

Le modèle polynomial de degré 4 montre une amélioration significative par rapport au modèle cubique (modèle suivant), ce qui suggère une relation plus complexe (voir Annexe 7).

Régression avec splines cubiques restreintes

L'utilisation des splines cubiques restreintes a donné un R² de 0.041 (p-value : 0.175). Cela montre une légère amélioration par rapport au modèle linéaire, mais la relation reste non significative (cf. Annexe 8).

Régression LOWESS

L'application de la régression LOWESS avec des fractions de 0.25, 0.5, et 0.75 a produit des Pseudo-R² de -0.0647, -0.1038, et -0.1144 respectivement. Ces valeurs négatives indiquent qu'aucun des modèles LOWESS n'améliore l'ajustement par rapport à la moyenne constante, confirmant l'absence de relation significative (cf. Annexe 9).

Transformation log-log

La transformation logarithmique des deux variables a donné un R² de 0.001 (p-value : 0.656), confirmant l'absence de relation log-linéaire ou log-log (cf. Annexe 10).

Régression linéaire par morceaux (Piecewise Linear Regression)

L'analyse par morceaux a identifié un seuil de consommation d'alcool à 10.74 litres par habitant :

- Pente avant le seuil : -0.0187
- Pente après le seuil : 0.4777
- R² global : 0.0270

Ces résultats suggèrent que la relation entre la consommation d'alcool et le taux d'homicides pourrait changer pour des niveaux élevés de consommation d'alcool (voir Annexe 11)

Modèles d'apprentissage automatique

Modèle	R ²	RMSE
KNN	0.3651	3.6065
Régression linéaire multiple	0.1607	-
SVR	0.1379	4.2024
Random Forest	-0.0768	4.6968

Le modèle KNN présente les meilleures performances, expliquant environ 36.5% de la variance du taux d'homicides (voir Annexe 12)

L'analyse d'importance des variables dans le modèle Random Forest montre que les facteurs socio-économiques jouent un rôle majeur :

- PIB par habitant : 60.3%
- Consommation d'alcool : 25.4%
- Taux de chômage : 14.3%

Analyses par pays et par année

Variations temporelles

L'analyse par année n'a pas révélé de relation significative entre la consommation d'alcool et le taux d'homicides pour les périodes étudiées :

- 2000 : R² = 0.0020, Coefficient = 0.0670
- 2005 : R² = 0.0363, Coefficient = 0.2786
- 2010 : R² = 0.0100, Coefficient = 0.0699
- 2019 : R² = 0.0004, Coefficient = 0.0090

Ces résultats montrent qu'il n'y a pas de relation stable dans le temps.

Analyses régionales

Des différences régionales sont observées :

- Europe de l'Est : R² = 0.0243, Coefficient = 0.1961
- Europe de l'Ouest : R² = 0.1151, Coefficient = -0.0917

Ces résultats suggèrent que la relation entre consommation d'alcool et homicides varie selon les contextes culturels et socio-économiques (voir Annexe 13 et 14).

Analyses par pays

L'étude a identifié 23 pays pour lesquels il existe une relation potentielle entre la consommation d'alcool et le taux d'homicides ($R^2 > 0.5$). Les coefficients de régression varient considérablement, allant de -3.3047 (Macédoine du Nord) à 6.9524 (Turquie), ce qui suggère des dynamiques nationales spécifiques.

Exemples notables :

- Russie : $R^2 = 0.5534$, Coefficient = 3.9912
- Slovaquie : $R^2 = 0.9718$, Coefficient = 1.5275
- Islande : $R^2 = 0.8848$, Coefficient = -1.2875
- Azerbaïdjan : $R^2 = 0.9789$, Coefficient = -2.7533

Des coefficients négatifs dans plusieurs pays montrent que la relation peut être inverse dans certains contextes nationaux.

Discussion

L'analyse globale n'a pas permis de démontrer une relation statistiquement significative entre la consommation d'alcool et le taux d'homicides en Europe dans les modèles linéaires simples. Cependant, les modèles non linéaires et les analyses spécifiques par pays révèlent des nuances importantes.

Interprétation des résultats

1. **Relation non linéaire complexe** : Les modèles polynomial (degré 4) et KNN suggèrent qu'il pourrait exister une relation non linéaire entre la consommation d'alcool et les homicides, que les modèles simples ne détectent pas.
2. **Importance des facteurs socio-économiques** : Le PIB par habitant est un prédicteur plus fiable du taux d'homicides que la consommation d'alcool, ce qui souligne le rôle clé des conditions économiques.
3. **Variations par pays** : Les analyses montrent des relations très variables selon les pays, positives ou négatives, ce qui indique que des facteurs culturels, historiques et sociaux peuvent influencer la relation entre alcool et violence.
4. **Effet de seuil potentiel** : L'analyse par morceaux suggère qu'une consommation d'alcool élevée (au-delà de 10.74 litres par habitant) pourrait avoir une relation plus forte avec les taux d'homicides.

Limites méthodologiques

1. **Agrégation des données** : L'utilisation de données au niveau national peut masquer des relations qui existent au niveau individuel ou local.
2. **Manque de distinction des types de consommation** : La mesure globale de la consommation d'alcool ne distingue pas entre une consommation modérée, excessive ou épisodique.
3. **Facteur non pris en compte** : Des éléments comme l'accès aux armes, l'efficacité judiciaire, les inégalités sociales ou les normes culturelles peuvent influencer les résultats.
4. **Données incomplètes** : Le manque de données pour certains pays et certaines années réduit la représentativité de l'échantillon.

Comparaison avec les études antérieures

Nos résultats apportent des nuances par rapport aux recherches précédentes :

- **Langlade (2011)** : Cette étude confirme que la consommation d'alcool seule n'est pas un bon prédicteur du taux d'homicides. Cependant, nos analyses suggèrent que des modèles non linéaires pourraient révéler des relations plus complexes.
- **Rossow (2001)** : Les variations observées d'un pays à l'autre dans notre étude corroborent les conclusions de Rossow sur l'importance des contextes nationaux dans l'association entre alcool et violence.
- **Pridemore (2002)** : Nos résultats montrent une relation plus forte en Russie, ce qui rejoint l'idée que des facteurs culturels spécifiques influencent le lien entre alcool et homicides.

Recommandations pour les recherches futures

1. **Approches non linéaires** : Explorer davantage les modèles non linéaires et les interactions entre variables pour mieux comprendre la relation complexe entre alcool et homicides.
2. **Indicateurs de consommation problématique** : Différencier la consommation modérée de l'abus d'alcool (ex. indicateurs de consommation excessive épisodique, taux de maladies liées à l'alcool).
3. **Études longitudinales** : Analyser l'évolution de cette relation dans chaque pays sur de longues périodes pour détecter d'éventuels changements ou tendances.
4. **Facteurs contextuels** : Prendre en compte des variables culturelles, des indicateurs d'inégalité (indice de Gini) et des politiques publiques sur l'alcool qui peuvent influencer les résultats.
5. **Analyses régionales et locales** : Examiner les variations infranationales pour identifier des relations qui peuvent être masquées au niveau national.

Conclusion

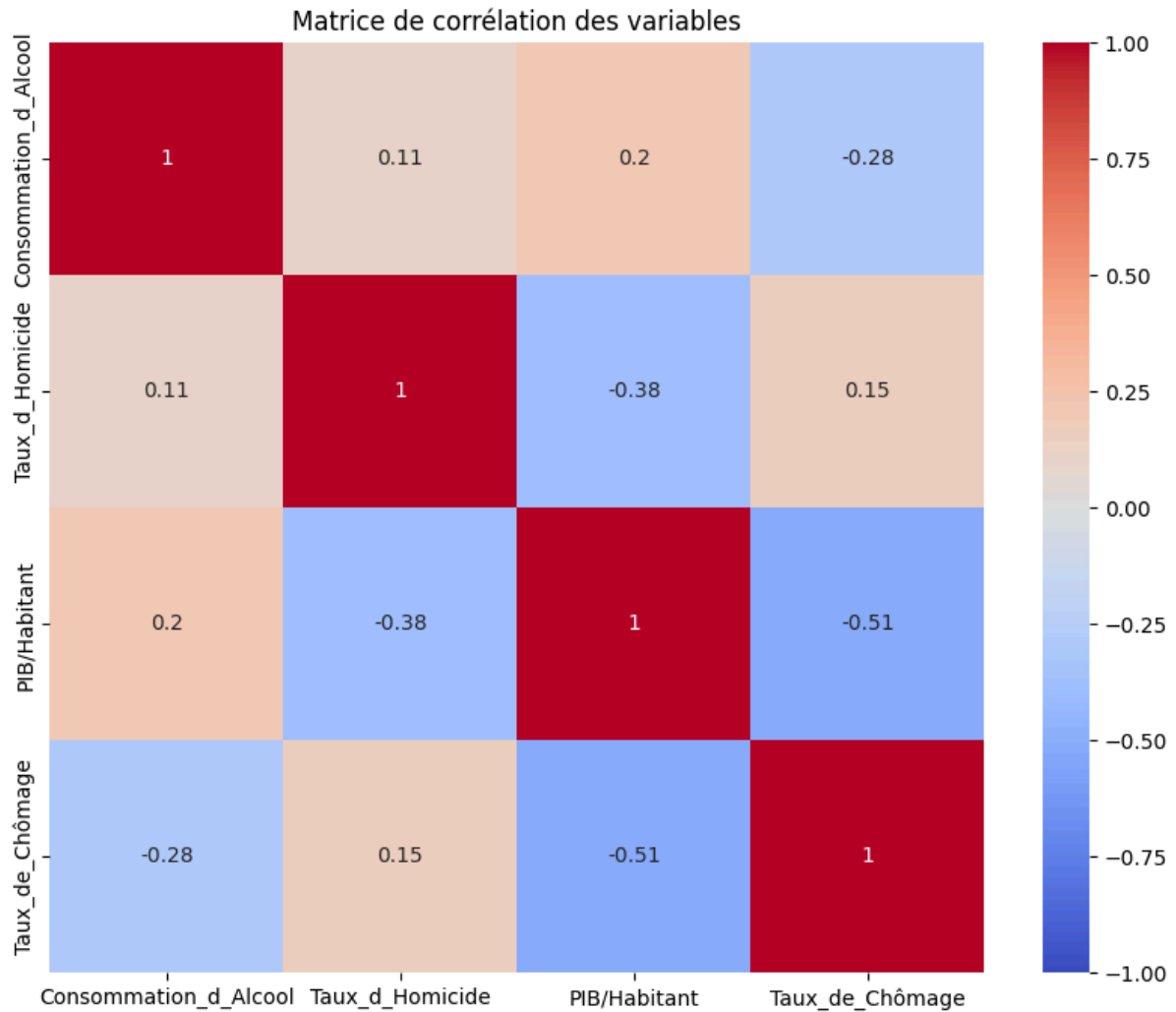
Cette étude montre que la relation entre la consommation d'alcool et le taux d'homicides en Europe est complexe et non linéaire. Les analyses linéaires simples n'ont pas révélé de lien significatif, mais les modèles avancés comme le KNN indiquent des relations plus subtiles, expliquant jusqu'à 36,5 % de la variance.

Les résultats soulignent surtout l'influence prépondérante des facteurs socio-économiques, notamment le PIB par habitant, dans l'explication des taux d'homicides. Ils révèlent aussi une grande diversité des relations selon les pays, avec des associations positives dans certains cas, négatives dans d'autres.

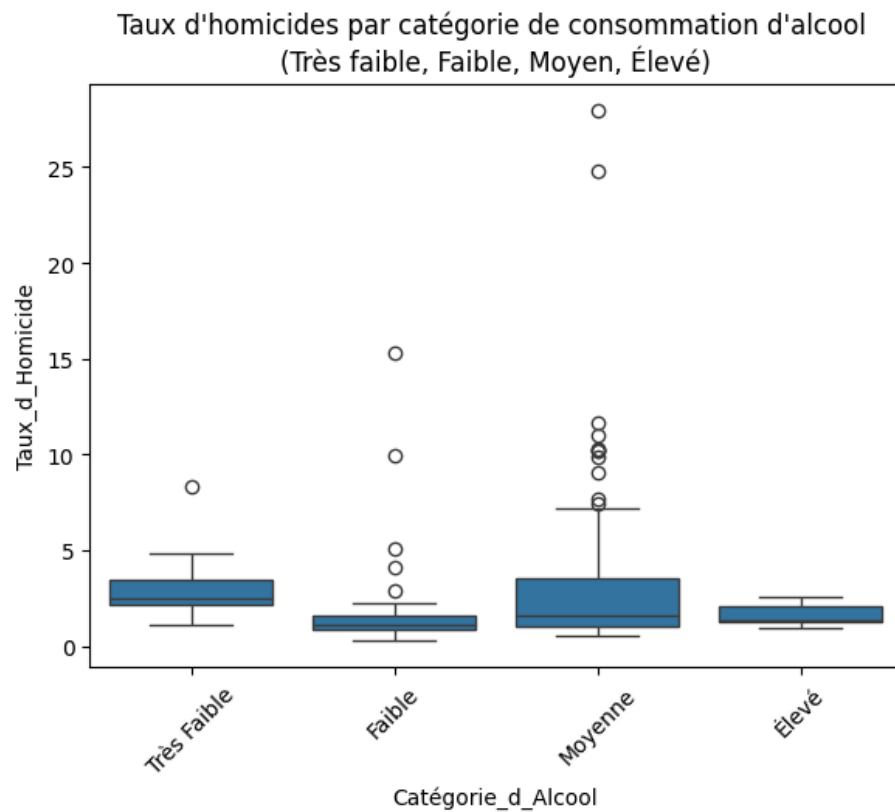
Ces observations suggèrent que comprendre le lien entre alcool et violence nécessite des méthodes plus sophistiquées, capables de capturer les relations non linéaires et les spécificités culturelles et socio-économiques. Il est essentiel de dépasser les analyses simplistes pour explorer comment la consommation d'alcool peut affecter les comportements violents dans des contextes variés.

Annexes

Annexe 1:



Annexe 2 :

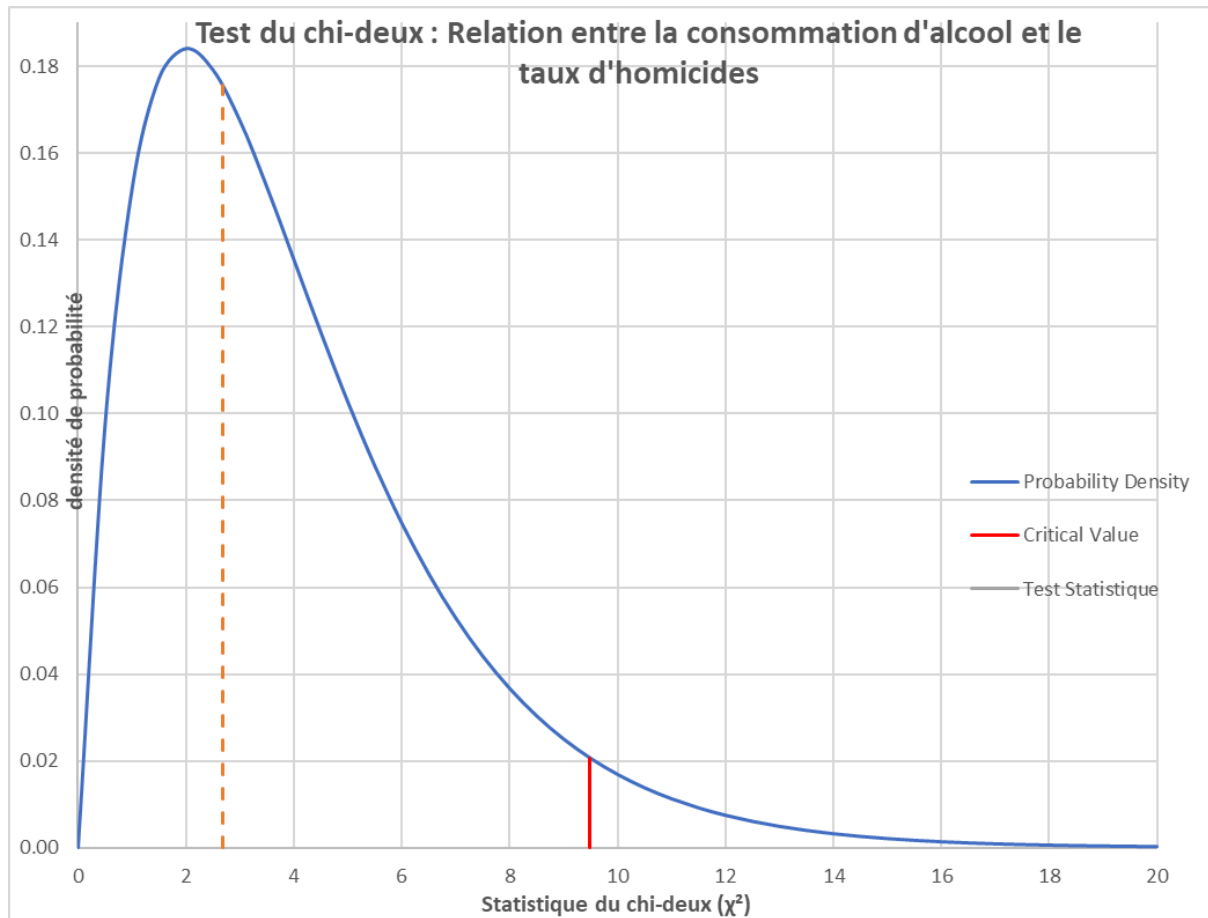


Annexe 3 :Fréquences observés et attendues (Chi²)

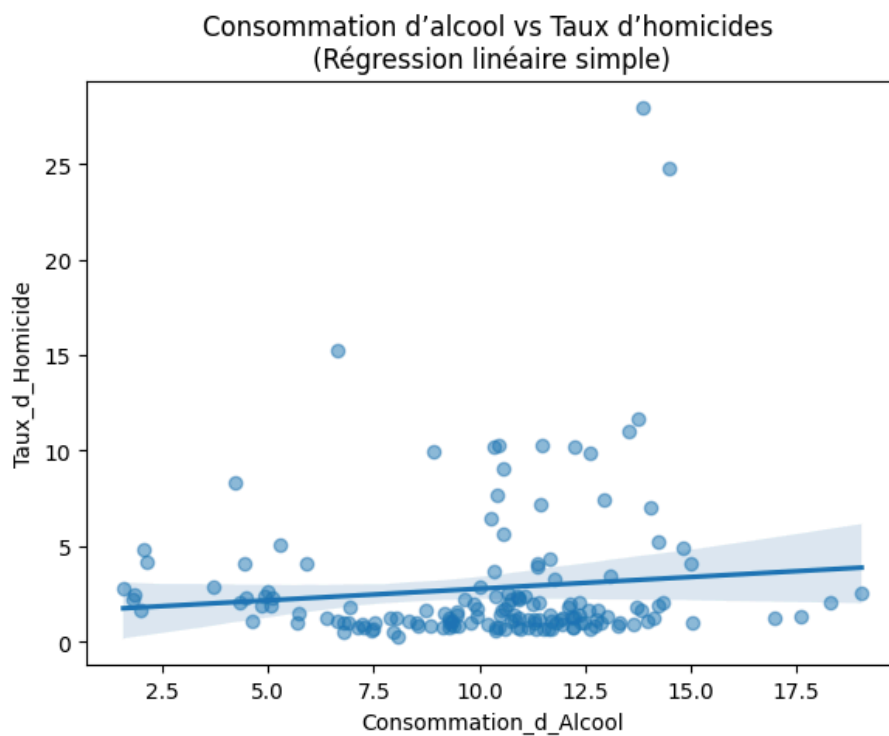
	Fréquences observées			
	Niveau du taux d'homicide			Total
Niveau d'alcool	Faible	Moyen	élevé	
Faible	5	5	5	15
Moyen	5	4	7	16
Élevé	2	7	6	15
Total	12	16	18	46

Fréquence attendues			
Niveau d'alcool	Niveau du taux d'homicide		
	Faible	Moyen	Élevé
Faible	3.91	5.22	5.87
Moyen	4.17	5.57	6.26
Élevé	3.91	5.22	5.87

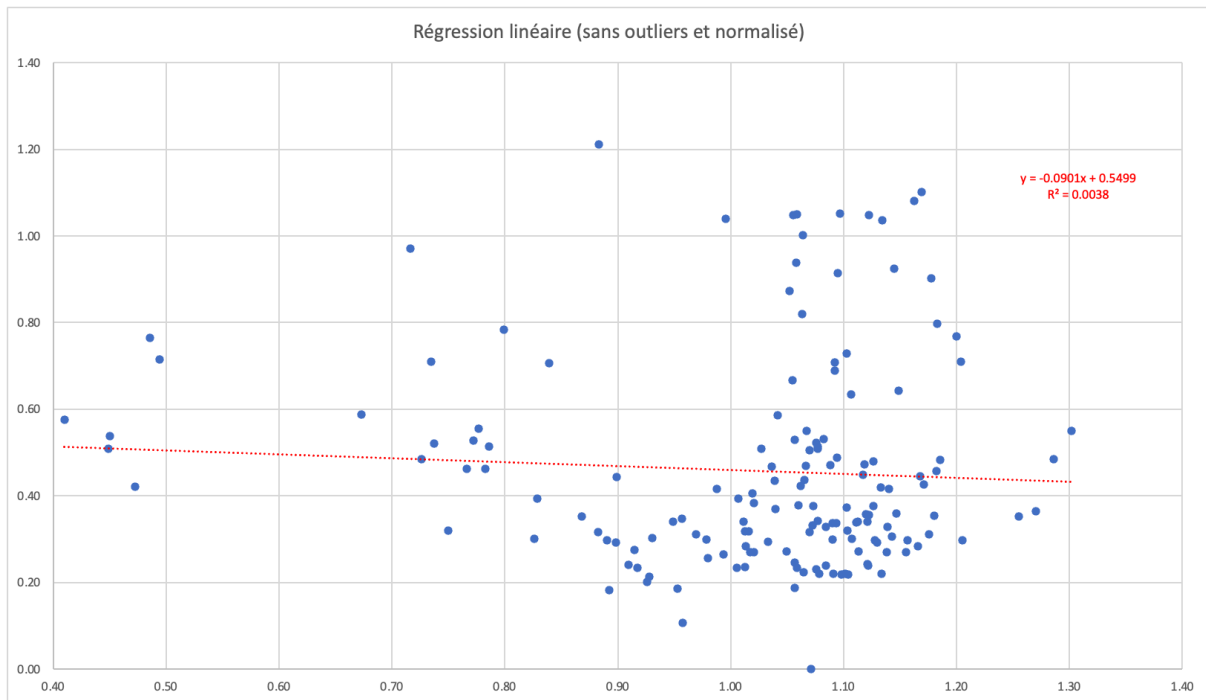
Annexe 4:



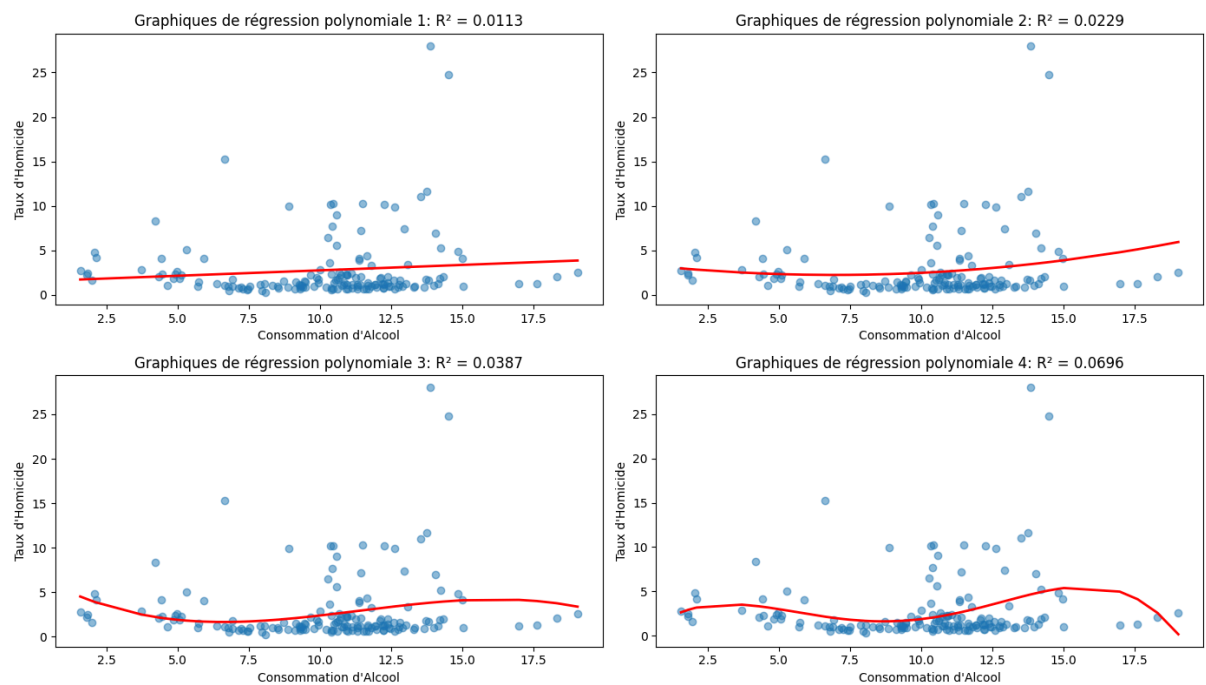
Annexe 5 :



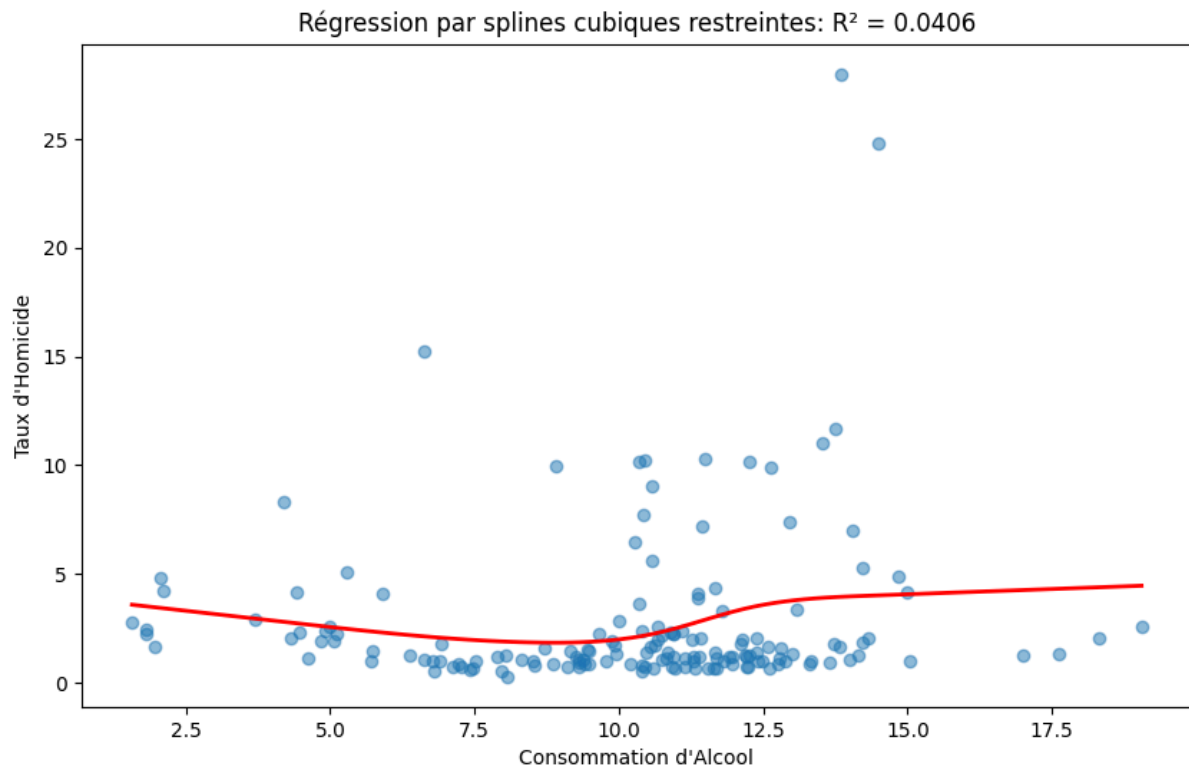
Annexe 6 :



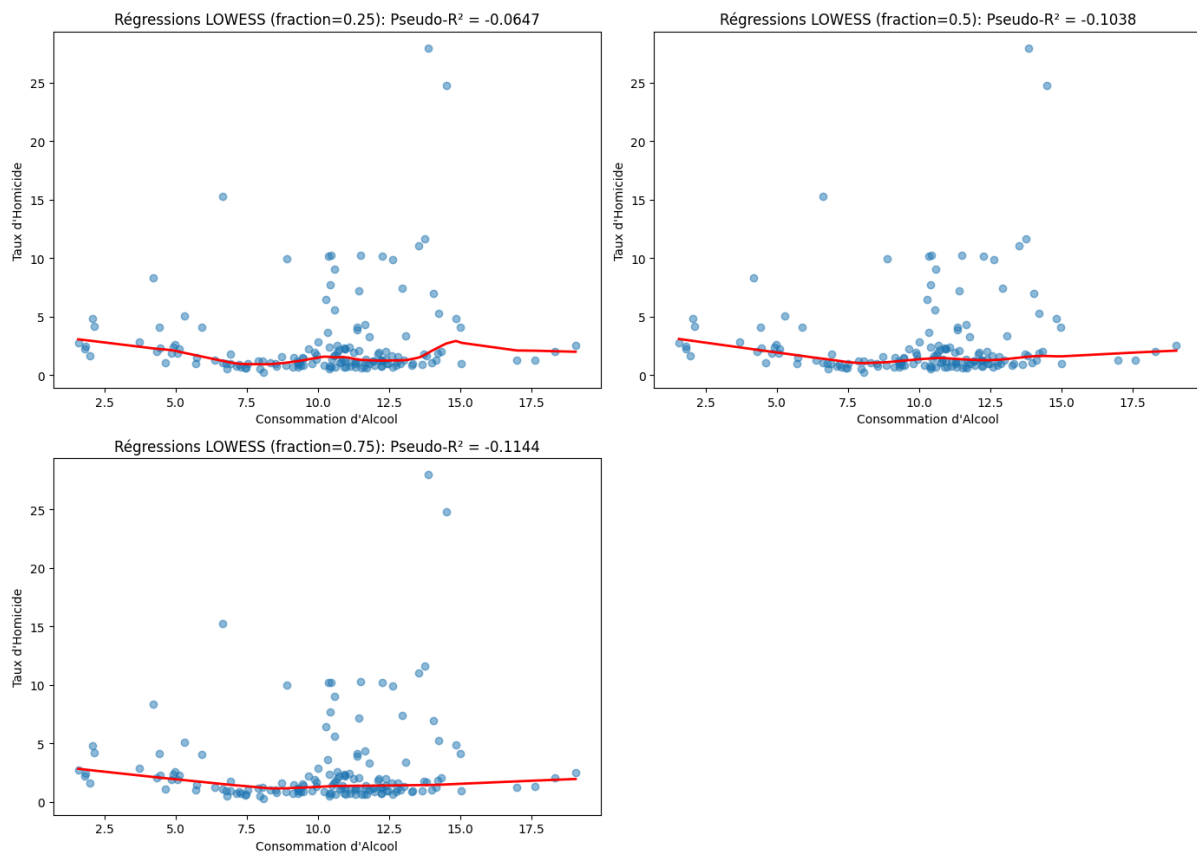
Annexe 7 :



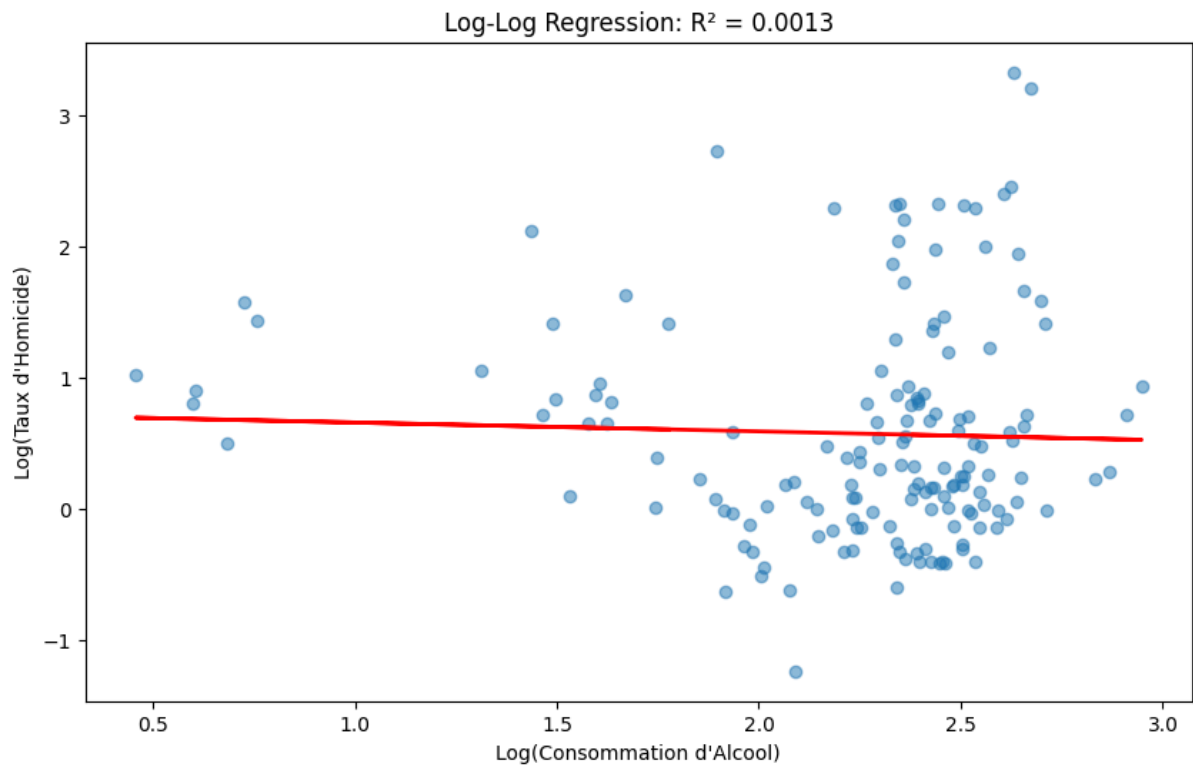
Annexe 8 :



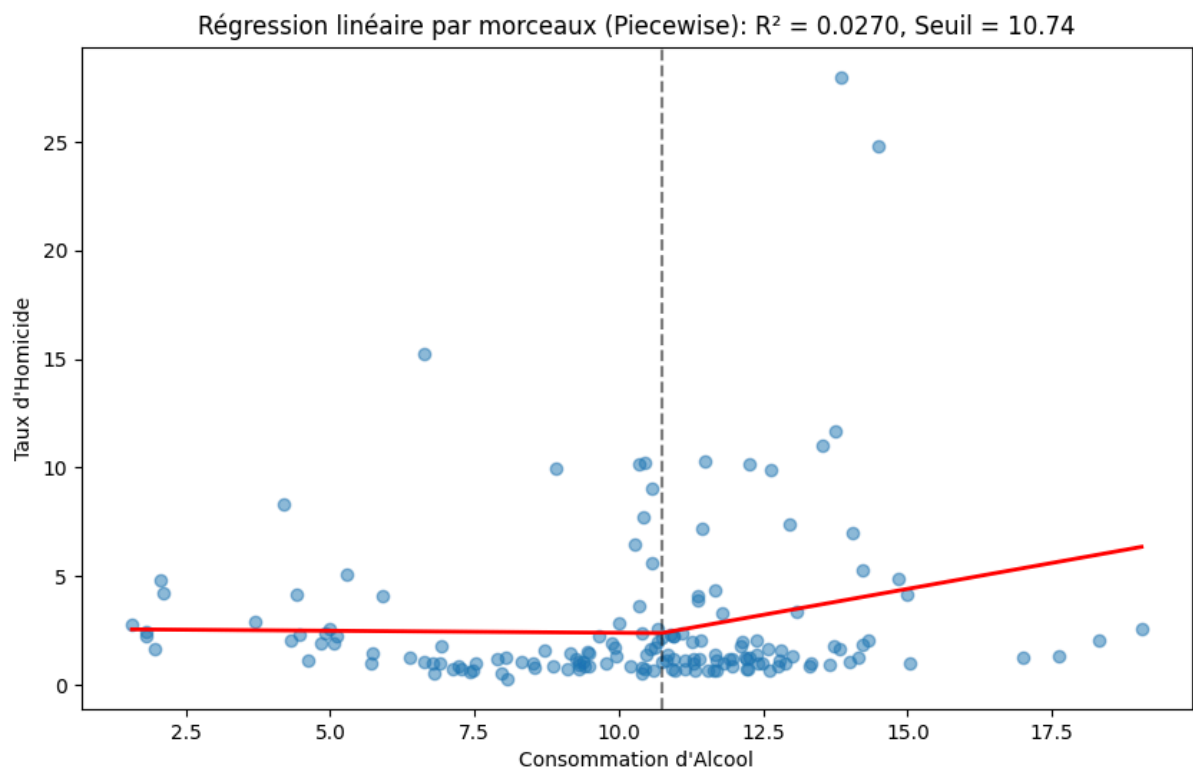
Annexe 9 :



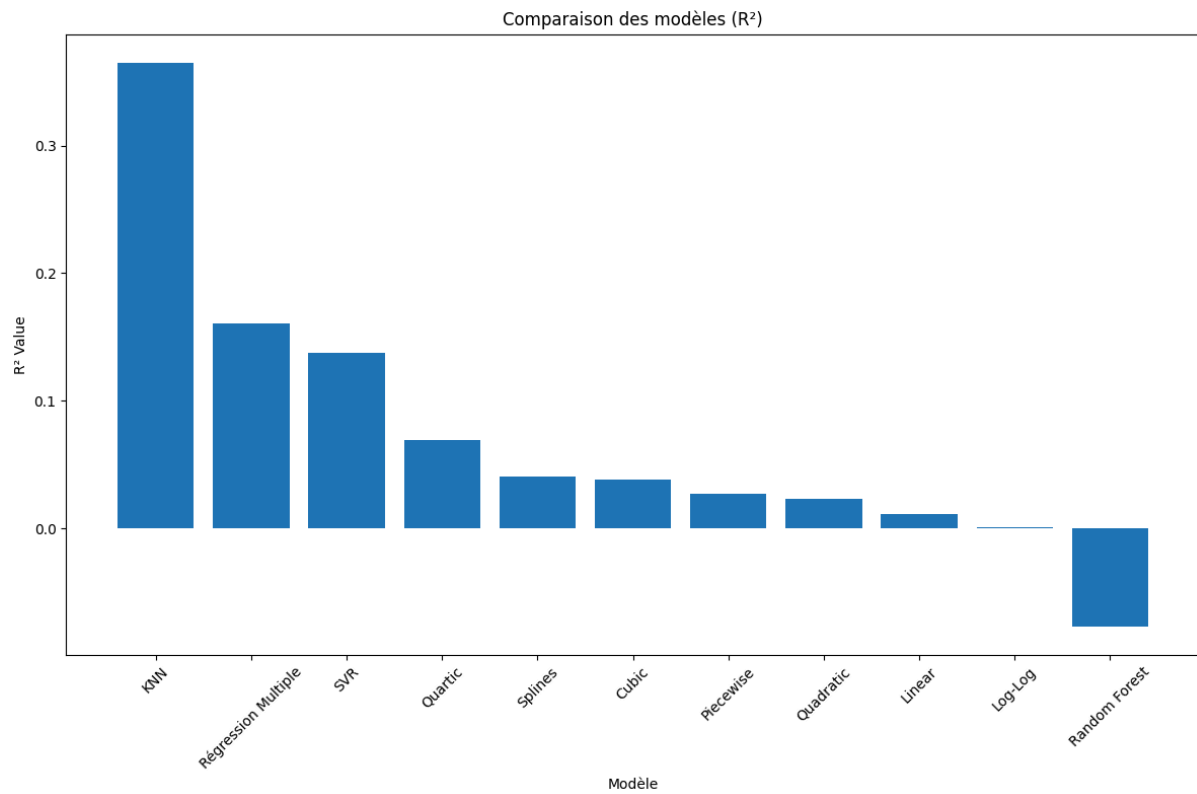
Annexe 10 :



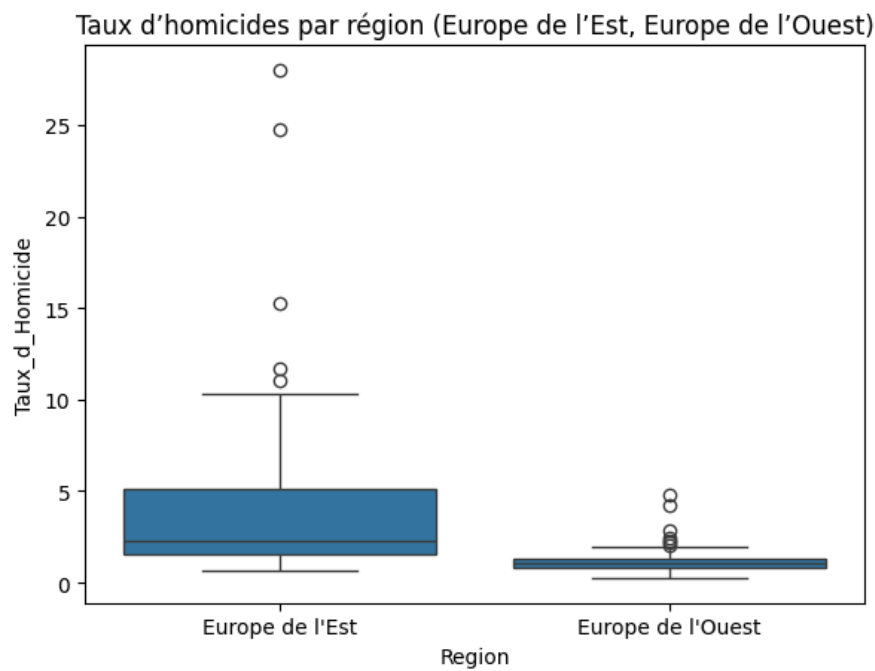
Annexe 11 :



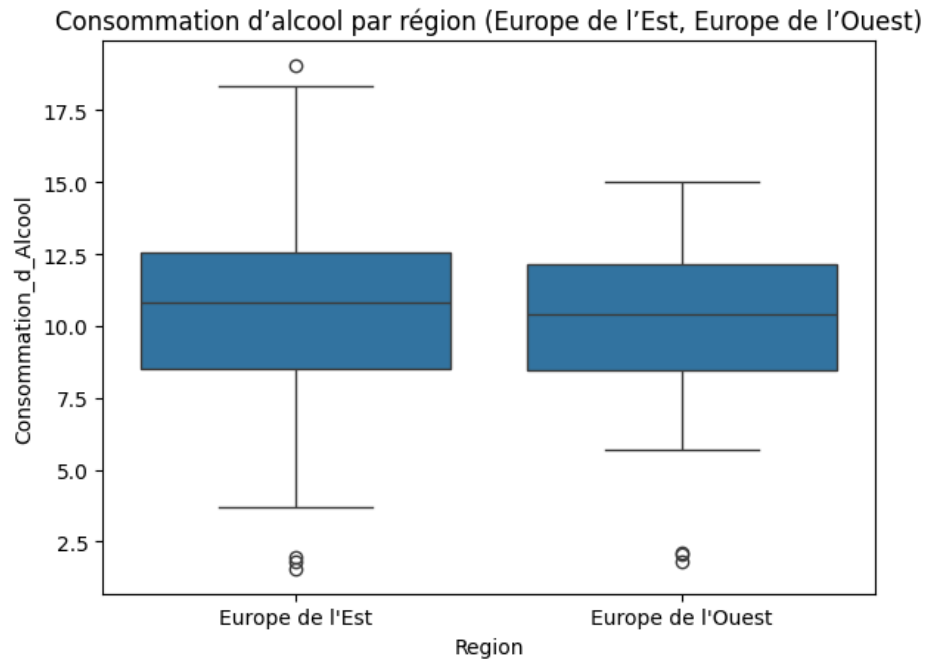
Annexe 12 :



Annexe 13 :



Annexe 14 :



Bibliographie

Achim, Z., Kleiber, C., & Jackman, S. (2008). *Regression Models for Count Data* in R. *Journal of Statistical Software*, 27(7), 1–25. <https://doi.org/10.18637/jss.v027.i08>

Langlade, A. (2011). *Analyse et explication de la variation du taux d'homicide en Europe* (Thèse de doctorat, Université de Montréal). Disponible sur : <https://umontreal.scholaris.ca/items/297c4cc3-8f6b-4b37-8840-7abeab8dfd21>

Nahhas, R. W. (2014). *Applied Linear Regression (4th ed.)*. Hoboken, NJ: Wiley. Disponible sur : <https://www.bookdown.org/rwnahhas/RMPH/mlr-normality.html#ref-weisberg2014>

Our World in Data. (s.d.). *Alcohol Consumption*. Disponible sur : <https://ourworldindata.org/grapher/total-alcohol-consumption-per-capita-litres-of-pure-alcohol>

Our World in Data. (s.d.). *Homicide Rate*. Disponible sur : <https://ourworldindata.org/grapher/homicide-rate-unodc?tab=table>

Planète Santé. (s.d.). *L'alcool est à l'origine d'un acte violent sur deux*. Disponible sur : <https://planetesante.ch/Magazine/Addictions/Alcool/L-alcool-est-a-l-origine-d-un-acte-violent-sur-deux>

Pvalue.io. (s.d.). *Analyses univariées et multivariées*. Disponible sur : <https://www.pvalue.io/fr/analyses-univariées-et-multivariées/>

PubMed. (2017). *Linear Regression and the Normality Assumption*. *Journal of Clinical Epidemiology*, 98, 146–151. <https://doi.org/10.1016/j.jclinepi.2017.12.006>

University of Virginia Library. (s.d.). *Normality Assumption*. Disponible sur : <https://library.virginia.edu/data/articles/normality-assumption>

Organisation mondiale de la santé. (2021). *L'alcool dans la Région européenne : consommation, méfaits et politiques*. Organisation mondiale de la santé. Disponible sur : <https://apps.who.int/iris/bitstream/handle/10665/347440/WHO-EURO-2001-3852-43611-61283-fre.pdf?sequence=1>

Office des Nations Unies contre la drogue et le crime. (2011). *Étude mondiale sur l'homicide (2011)*. Nations Unies. Disponible sur : https://www.unodc.org/documents/data-and-analysis/statistics/Homicide/BOOK_Etude_mondiale_sur_lhomicide_2011_ebook.pdf

Banque mondiale. (2025). *Indicateurs du développement mondial (PIB par habitant, taux de chômage)*. Banque mondiale. Disponible sur : <https://data.worldbank.org/indicator>

McKinney, W. (2010). *Data Structures for Statistical Computing in Python*. *Proceedings of the 9th Python in Science Conference*, 51-56. DOI : 10.25080/Majora-92bf1922-00a

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., ... & Oliphant, T. E. (2020). *Array programming with NumPy*. Nature, 585(7825), 357-362. DOI : 10.1038/s41586-020-2649-2

Hunter, J. D. (2007). *Matplotlib: A 2D Graphics Environment*. Computing in Science & Engineering, 9(3), 90-95. DOI : 10.1109/MCSE.2007.55

Waskom, M. L. (2021). *Seaborn: Statistical Data Visualization*. Journal of Open Source Software, 6(60), 3021. DOI : 10.21105/joss.03021

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). *Scikit-learn: Machine Learning in Python*. Journal of Machine Learning Research, 12, 2825-2830. Disponible sur : <http://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html>

Seabold, S., & Perktold, J. (2010). *Statsmodels: Econometric and statistical modeling with Python*. Proceedings of the 9th Python in Science Conference, 92-96. DOI : 10.25080/Majora-92bf1922-011