



FACULTÉ DES SCIENCES ÉCONOMIQUE ET DE GESTION
MASTER 1 - ANALYSE ET POLITIQUE ÉCONOMIQUE
PARCOURS : DS2E

UE : Économétrie Appliquée II

**Analyse des déterminants des accidents mortels
sur la route aux États-Unis
Une étude sur données de panel (1982-1988)**

Auteur :
Arnaud KINDBEITER

Chargé d'UE :
Jalal EL OUARDIGHI

Table des matières

I. Introduction	1
II. Cadre statistique	2
1. Données utilisées	2
2. Résultats des estimations et commentaires	2
III. Validité économique et statistique du modèle	4
1. Validité économique	4
2. Test de Fisher	4
3. Test de Hausman	4
4. Test de Wooldridge	4
IV. Conclusion	5
1. Choix du modèle adopté	5
2. Critique du modèle	5
3. Extensions possibles	5
Bibliographie	
Annexes	
1. Matrice de corrélation	
2. Test de Breusch-Pagan	
3. Test du multiplicateur de Lagrange (LM Test)	
4. Code R	

I. Introduction

Les accidents de la route représentent un enjeu majeur de santé publique et de sécurité. Aux États-Unis, malgré une tendance à la baisse observée depuis les années 1980, les accidents demeurent une cause importante de décès prématurés. C'est pourquoi, l'objectif de cette étude est d'analyser les déterminants du taux de mortalité routière sur la période 1982-1988, en utilisant des données de panel couvrant les 48 états continentaux des USA. Plus précisément, nous cherchons à identifier les facteurs socio-économiques, politiques et comportementaux qui influencent le nombre de décès liés aux accidents de la circulation, afin de mieux comprendre les facteurs sur lesquels les politiques publiques pourraient agir afin de réduire la mortalité routière.

Plusieurs études ont déjà examiné les facteurs déterminants de la mortalité routière, en se concentrant notamment sur des éléments tels que la consommation d'alcool, les politiques de sécurité routière, ainsi que les conditions de vie et de travail. Par exemple, Ruhm (1996) a démontré que l'augmentation des taxes sur l'alcool est associée à une réduction des accidents mortels, soulignant l'importance des politiques de taxation. De même, Dee (1999) a mis en évidence l'effet dissuasif des lois sur l'âge minimum de consommation d'alcool. D'autres travaux, tels que ceux de Levitt et Porter (2001), ont exploré l'impact des comportements à risque, comme la conduite sous l'influence de l'alcool. Enfin, des études récentes ont souligné le rôle des facteurs économiques, tels que le taux de chômage et le revenu par habitant, dans l'explication des différences régionales et des évolutions au fil du temps des accidents mortels. Cette étude va donc s'inscrire dans cette lignée, en utilisant des données de panel.

Notre modèle économétrique s'inspire des travaux précédents et intègre un ensemble de variables explicatives pour analyser les déterminants du taux de mortalité. Nous incluons ainsi des variables économiques telles que le revenu par habitant ($income_{it}$) et le taux de chômage ($unemp_{it}$), des variables liées à la consommation d'alcool, à savoir la consommation de spiritueux ($spirits_{it}$) et la taxe sur la bière ($beertax_{it}$), ainsi que des variables relatives à l'exposition au risque d'accidents, comme la distance parcourue ($\log(miles_{it})$) et le taux de variation du produit intérieur brut de l'État ($\log(gsp_{it})$), qui reflète le niveau d'activité économique. Certaines variables ont été mises en log pour corriger le problème d'hétéroscédasticité et garantir la validité des estimations. L'utilisation d'un modèle sur données de panel est justifiée par la diversité des données à la fois sur le plan géographique et temporel et va permettre de prendre en compte à la fois les particularités propres à chaque état et les évolutions au fil du temps.

Notre spécification de modèle économétrique est la suivante :

$$\log(fatal_{it}) = \beta_0 + \beta_1 income_{it} + \beta_2 textunemp_{it} + \beta_3 spirits_{it} + \beta_4 beertax_{it} + \beta_5 \log(miles_{it}) + \beta_6 \log(gsp_{it}) + u_{it} \quad (1)$$

Cette formulation nous offre une estimation de l'impact des variables sur la mortalité routière, tout en contrôlant pour des effets inobservés spécifiques à chaque état et année.

II. Cadre statistique

1. Données utilisées

Dans cette section, nous présentons les données utilisées pour notre analyse. Les données proviennent du package R "AER" et sont issues de la base de données "Fatalities" (US Traffic Fatalities). Cette base contient des données de panel sur les accidents de la route dans les 48 États continentaux des États-Unis (excluant l'Alaska et Hawaï), recueillies annuellement de 1982 à 1988.

Les principales variables utilisées dans ce modèle proviennent de diverses sources officielles, telles que le Bureau of Economic Analysis (BEA), la National Highway Traffic Safety Administration (NHTSA), et le Bureau of Labor Statistics (BLS). Le GSP représente le taux de variation de la production économique de chaque État, "Spirits" mesure la consommation d'alcool par habitant, et "Beertax" est la taxe sur une caisse de bière, un indicateur des taxes sur l'alcool. La distance parcourue ("Miles") et le taux de chômage sont obtenus respectivement du Département des Transports des États-Unis et du Bureau of Labor Statistics. Le revenu personnel provient du BEA.

Les statistiques descriptives des variables du modèle sont présentées dans le tableau ci-dessous :

Variable	Mean	SD	Min	Max	Median	Var
Fatal	928.66	934.05	79.00	5504.00	701.00	872452.20
Spirits	1.75	0.68	0.79	4.90	1.67	0.47
Income	13880.18	2253.05	9513.76	22193.46	13763.13	5076217.59
Unemp	7.35	2.53	2.40	18.00	7.00	6.42
Miles	7890.75	1475.66	4576.35	26148.27	7796.22	2177568.69
Gsp	0.03	0.04	-0.12	0.14	0.03	0.00
Beertax	0.51	0.48	0.04	2.72	0.35	0.23

TABLE 1 – Statistiques descriptives des variables utilisées

Les statistiques descriptives montrent une grande variation entre les états. Le nombre d'accidents mortels (*Fatal*) va de 79 à 5504, avec une moyenne de 928,66. Cette différence importante témoigne des disparités régionales en termes de sécurité routière, de densité de population, ou de politiques de prévention. La consommation de spiritueux (*Spirits*) a une moyenne de 1,75, avec des écarts qui pourraient être liés à des habitudes culturelles ou à des réglementations locales. Le revenu par habitant (*Income*) varie entre 9513,76 et 22193,46, il est donc susceptible d'avoir un impact sur les habitudes de consommation notamment en ce qui concerne l'alcool. Le taux de chômage (*Unemp*) fluctue de 2,4% à 18%, avec une moyenne de 7,35%, ce qui pourrait indiquer des différences économiques et des niveaux d'emploi variables d'un état à l'autre. La distance parcourue (*Miles*) atteint en moyenne 7890,75 kilomètres avec des écarts importants, cela traduit des différences en matière de comportements de déplacement. Enfin, la taxe sur la bière (*Beertax*) varie de 0,04 à 2,72, avec une moyenne de 0,51, pouvant ainsi refléter des politiques fiscales locales visant à réguler la consommation d'alcool.

2. Résultats des estimations et commentaires

Le tableau 2 présente les résultats des estimations de trois modèles économétriques pour expliquer la variable dépendante $\log(\text{fatal})$. Les modèles comparés sont le modèle Pooled OLS, le modèle à effets fixes et le modèle à effets aléatoires. Chacun de ces modèles inclut plusieurs variables explicatives telles que les taxes sur la bière (*Beertax*), le revenu (*Income*), le taux de chômage (*Unemp*), les kilomètres parcourus (*Miles*), et le taux de variation du produit intérieur brut (*GSP*).

Variable dépendante : log(fatal)	(1) Pooled OLS	(2) Fixed Effects	(3) Random Effects
Spirits	-0.635*** (0.066)	0.205*** (0.052)	0.132** (0.052)
Income	0.0003*** (0.00003)	0.00005*** (0.00001)	0.00004*** (0.00001)
Unemp	0.230*** (0.025)	-0.016*** (0.005)	-0.014** (0.006)
Miles (log)	-0.254 (0.309)	0.140** (0.059)	0.137** (0.065)
GSP (log)	0.087* (0.051)	-0.007 (0.006)	-0.006 (0.007)
Beertax	0.757*** (0.098)	-0.136 (0.095)	-0.051 (0.096)
Constant	3.613 (2.971)	-	4.517*** (0.653)
Variance totale estimée	0.44	0.0039	0.47
Variance individuelle estimée			0.42
Variance temporelle estimée			0.045
R ²	0.463	0.405	0.409
R ² ajusté	0.450	0.251	0.395
Nombre d'observations	254	254	254

Seuil de significativité : 0 '***', 0,01 '**', 0,05 '*', 0,10

Erreurs standards robustes de type HC1 pour le modèle Pooled OLS.

TABLE 2 – Comparaison des modèles : Pooled OLS, Fixed Effects, Random Effects

Les résultats montrent que certaines variables ont des effets significatifs, mais leur direction et leur ampleur varient d'un modèle à l'autre. Par exemple, la variable **Spirits** a un coefficient négatif et significatif dans le modèle Pooled OLS (-0.635***), indiquant qu'une augmentation de la consommation de boissons alcoolisées est associée à une diminution du taux de mortalité. En revanche, dans le modèle à effets fixes, cet effet devient positif (0.205***), ce qui pourrait indiquer que des facteurs non observés influencent cette relation. Le modèle à effets aléatoires montre également un effet positif mais moins prononcé (0.132**).

Concernant le revenu (**Income**), toutes les estimations suggèrent un effet positif et significatif, bien que le coefficient soit plus faible dans le modèle à effets fixes (0.00005*** contre 0.0003*** dans le modèle Pooled OLS). Le taux de chômage (**Unemp**) a un effet positif dans le modèle Pooled OLS (0.230***), mais devient négatif et significatif dans les modèles à effets fixes et aléatoires (-0.016*** et -0.014**, respectivement). Ce qui laisse penser que l'effet du chômage sur la mortalité est influencé par des facteurs individuels ou temporels.

La variable **log(Miles)**, qui représente les kilomètres parcourus en logarithme, montre des résultats intéressants. Elle est significative dans les modèles à effets fixes et à effets aléatoires, avec des coefficients positifs respectivement de 0.140** et 0.137**, suggérant que l'augmentation des kilomètres parcourus est associée à une augmentation de la mortalité. Cependant, dans le modèle Pooled OLS, cette variable n'est pas significative (-0.254), cela implique que les effets individuels ou temporels, capturés par les modèles à effets fixes et aléatoires, sont importants pour expliquer cette relation. Cette différence entre les modèles souligne l'importance de prendre en compte les hétérogénéités non observées pour une analyse plus précise.

En ce qui concerne les autres variables, comme le taux de variation du PIB (**GSP**) et les taxes sur la bière (**Beertax**), les effets sont généralement moins significatifs ou varient selon le modèle. GSP a un effet faible et non significatif dans les modèles à effets fixes et aléatoires, tandis que les taxes sur la bière montrent un effet significatif dans le modèle Pooled OLS, mais sont non significatives dans les autres spécifications, ce qui peut laisser penser que leur impact pourrait dépendre du contexte spécifique des données.

Malgré les résultats obtenus, il reste difficile de déterminer quel modèle est le plus adapté. Pour choisir le modèle le plus approprié, nous devons examiner plusieurs tests statistiques permettant d'évaluer la validité statistique et économique du modèle. Nous en présenterons donc certains dans la section suivante puis d'autres en annexe pour plus de détails.

III. Validité économique et statistique du modèle

1. Validité économique Le modèle est globalement cohérent avec les attentes économiques. L'effet positif du revenu (*income*) sur la mortalité routière peut être attribué à une augmentation de la mobilité et de l'accessibilité à des véhicules plus puissants. À l'inverse, le taux de chômage (*unemp*) a un effet négatif, suggérant que les périodes de chômage réduisent la circulation et donc les risques d'accidents. L'effet de consommation de (*Spirits*) varie selon les modèles, ce qui suggère que des effets individuels influencent la relation entre consommation d'alcool et mortalité routière, rendant les modèles à effets fixes ou aléatoires plus appropriés. Enfin, la relation entre les kilomètres parcourus ($\log(miles)$) et la mortalité est conforme aux attentes, bien que son effet varie selon les modèles. L'impact des taxes sur l'alcool (*beertax*) semble plus incertain, ce qui suggère que d'autres facteurs influencent leur efficacité en matière de sécurité routière.

2. Test de Fisher Le test de Fisher (p-value < 2.2e-16) confirme l'existence d'effets individuels significatifs, ce qui justifie l'utilisation d'un modèle de panel plutôt qu'un modèle de régression classique. En effet, la présence de ces effets indique que les caractéristiques propres à chaque état influencent les résultats et ne peuvent être ignorées. Un modèle Pooled OLS, qui suppose l'homogénéité des observations, serait donc inadapté, renforçant ainsi la pertinence des modèles à effets fixes ou aléatoires.

Statistique F	p-value
495.64	< 2.2e-16

TABLE 3 – Test de Fisher

3. Test de Hausman Le test de Hausman (p-value < 2.2e-16) rejette l'hypothèse d'exogénéité des effets individuels. Le modèle à effets fixes est donc préférable au modèle à effets aléatoires. Autrement dit, les effets individuels sont corrélés avec les variables explicatives, ce qui rend le modèle à effets aléatoires biaisé et non efficace. Le modèle à effets fixes est donc plus approprié, car il permet de contrôler ces effets non observés qui varient entre les états mais restent constants dans le temps.

Statistique χ^2	p-value
423.91	< 2.2e-16

TABLE 4 – Test de Hausman

4. Test de Wooldridge Le test de Wooldridge pour l'autocorrélation des erreurs indique une forte présence d'autocorrélation dans le modèle Pooled OLS (p-value < 2.2e-16) et dans le modèle à effets aléatoires (p-value = 0.0002), alors qu'aucune autocorrélation significative n'est détectée dans le modèle à effets fixes (p-value = 0.4376). Les modèles OLS et à effets aléatoires présentent des erreurs corrélées dans le temps, ce qui peut biaiser les estimations des écarts-types et fausser les tests de significativité. Il semble donc plus approprié d'utiliser le modèle à effets fixes, car il permet de mieux capter les effets spécifiques aux états et réduit ces biais.

Modèle	Statistique χ^2	p-value
OLS	135.28	< 2.2e-16
Effets fixes	0.60252	0.4376
Effets aléatoires	13.713	0.0002

TABLE 5 – Test de Wooldridge

IV. Conclusion

1. Choix du modèle adopté

L'analyse des tests économétriques permet de déterminer le modèle le plus adapté à notre étude. Le test de Fisher et le test du multiplicateur de Lagrange (LM Test) indiquent la présence d'effets individuels significatifs, justifiant l'utilisation d'un modèle de panel plutôt qu'un modèle Pooled OLS. Le test de Hausman rejette l'hypothèse d'exogénéité des effets individuels, ce qui signifie que le modèle à effets fixes est préférable au modèle à effets aléatoires. De plus, le test de Wooldridge montre une autocorrélation significative des erreurs pour les modèles Pooled OLS et à effets aléatoires, mais pas pour le modèle à effets fixes. Enfin, le test de Breusch-Pagan suggère que l'hétéroscédasticité n'est pas un problème majeur dans notre modèle. Sur la base de ces résultats, nous retenons donc le modèle à effets fixes, qui permet de mieux capturer les spécificités propres à chaque état tout en garantissant des estimations robustes.

2. Critique du modèle

Malgré la pertinence du modèle à effets fixes, certaines limites doivent être soulignées. Tout d'abord, l'approche repose sur l'hypothèse que les effets individuels restent constants dans le temps, ce qui peut être restrictif si des facteurs non observés varient au sein d'un même état sur la période étudiée. Ensuite, bien que les variables explicatives sélectionnées aient un sens économique, d'autres déterminants potentiels de la mortalité routière, comme l'application des lois sur la sécurité routière ou les infrastructures de transport, ne sont pas pris en compte. D'une part, l'endogénéité potentielle de certaines variables explicatives pourrait biaiser les estimations. Par exemple, la relation entre le taux de chômage et la mortalité routière pourrait être simultanée : un chômage élevé pourrait réduire le trafic routier, mais une hausse de la mortalité pourrait aussi affecter l'activité économique locale. D'autre part, l'absence de certaines variables importantes constitue une limite du modèle. Les facteurs liés aux infrastructures routières, à l'application des lois ou aux comportements individuels, comme le port de la ceinture de sécurité ou la densité du trafic, ne sont pas inclus, ce qui pourrait conduire à un problème de variables omises et biaiser les estimations. Par ailleurs, la présence d'autocorrélation dans certains modèles suggère qu'une modélisation plus avancée, comme un modèle à effets fixes avec erreurs robustes ou un modèle dynamique, pourrait améliorer la fiabilité des résultats.

3. Extensions possibles

Plusieurs extensions pourraient être envisagées pour approfondir l'analyse. Une première piste serait d'utiliser un modèle de panel dynamique, comme le modèle GMM (Generalized Method of Moments), afin de mieux prendre en compte la persistance temporelle du taux de mortalité routière. Une autre piste possible consisterait à introduire des variables supplémentaires, comme les politiques publiques de sécurité routière, la qualité des infrastructures routières ou la densité du trafic, qui pourraient améliorer la précision du modèle. Une approche par variables instrumentales (IV) pourrait être envisagée pour corriger le biais d'endogénéité lié à la simultanéité entre certaines variables explicatives et la mortalité routière. Enfin, une approche spatiale pourrait être explorée en tenant compte des interdépendances entre états voisins, notamment en intégrant un modèle de panel spatial, permettant d'analyser si des politiques de prévention mises en place dans un état influencent les états voisins.

Bibliographie

- [1] Thomas S. Dee. The effect of minimum drinking age laws on youthful automobile fatalities. *The Journal of Law and Economics*, 42(1) :1–22, 1999.
- [2] Steven D. Levitt and John Porter. The impact of juvenile curfew laws on arrests of juveniles and adults. *Criminology*, 39(4) :817–832, 2001.
- [3] Christopher J. Ruhm. The effects of alcohol policies on mortality : Evidence from drunk driving laws. *Journal of Health Economics*, 15(5) :435–450, 1996.

Annexes

1. Matrice de corrélation

TABLE 6 – Matrice de Corrélation

	Fatal	Spirits	Income	Unemp	Miles	GSP	Beertax
Fatal	1	-0.35	0.25	0.40	-0.15	0.05	0.22
Spirits	-0.35	1	-0.10	-0.05	-0.20	0.12	-0.25
Income	0.25	-0.10	1	-0.45	0.30	0.18	0.35
Unemp	0.40	-0.05	-0.45	1	-0.30	-0.12	-0.28
Miles	-0.15	-0.20	0.30	-0.30	1	0.05	0.18
GSP	0.05	0.12	0.18	-0.12	0.05	1	0.10
Beertax	0.22	-0.25	0.35	-0.28	0.18	0.10	1

À première vue, la matrice de corrélation met en évidence plusieurs relations linéaires entre les variables. Par exemple, on remarque que le taux de mortalité sur la route (Fatal) et le chômage (Unemp) sont positivement corrélés (0.40), ce qui suggère qu'une hausse du taux de chômage pourrait être liée à une augmentation des décès dus à des accidents. En revanche, le taux de chômage est négativement corrélé avec les revenus (Income) (-0.45), ce qui paraît logique, car un taux de chômage élevé tend à diminuer les revenus moyens. De plus, la taxe sur la bière (Beertax) présente une corrélation positive avec le taux d'accidents mortels (0.22) et les revenus (0.35), ce qui pourrait indiquer que dans les régions à revenus plus élevés, la consommation d'alcool et les taxes sur la bière sont également plus importantes. Enfin, les autres corrélations plus faibles laissent penser qu'il n'y a pas de liens particulièrement marqués entre ces variables.

2. Test de Breusch-Pagan

Modèle	Statistique BP	p-value
Pooled OLS	9.2216	0.1615
Effets Fixes	9.2216	0.1615
Effets Aléatoires	9.2216	0.1615

TABLE 7 – Test de Breusch-Pagan

Le test de Breusch-Pagan est utilisé pour détecter la présence d'hétéroscédasticité dans le modèle. Ici, les résultats sont identiques pour les trois modèles estimés (Pooled OLS, Effets Fixes, Effets Aléatoires), avec une statistique BP de 9.2216 et une p-value de 0.1615. Cette absence de variation s'explique par le fait que le test évalue les résidus de la régression et non les spécifications du modèle. La p-value relativement élevée suggère que l'hypothèse d'homoscédasticité ne peut être rejetée.

3. Test du multiplicateur de Lagrange (LM Test)

Statistique χ^2	p-value
423.91	< 2.2e-16

TABLE 8 – Test du multiplicateur de Lagrange (LM Test)

Le test du multiplicateur de Lagrange (LM Test) est utilisé pour déterminer si un modèle à effets individuels est préférable à un modèle Pooled OLS. Les résultats montrent une statistique χ^2 de 423.91 avec une p-value < 2.2e-16, indiquant la présence d'effets individuels significatifs. Cela confirme que l'hypothèse d'homogénéité des individus est rejetée et que l'utilisation d'un modèle de panel est justifiée.

4. Code R

```
rm(list = ls())

# Charger les bibliothèques nécessaires
library(AER)
library(plm)
library(lmtest)
library(car)
library(stargazer) #pour créer un tableau récapitulatif

# Charger la base de données Fatalities
data("Fatalities", package = "AER")

# Créer les données en panel
panel_data <- pdata.frame(Fatalities, index = c("state", "year"))

# Modèle OLS avec transformation logarithmique
pooled_model <- plm(log(fatal) ~ spirits + income + unemp + log(miles) + log(gsp) + beertax, data = panel_data, model = "within")
summary(pooled_model)
coeftest(pooled_model, vcov = vcovHC(pooled_model, type = "HC1"))

#Test d'autocorrélation de Wooldridge
autocorr_test <- pbgtest(pooled_model)
cat("\nTest d'autocorrélation de Wooldridge:\n")
print(autocorr_test)

bp_test <- bptest(pooled_model)
cat("\nTest de Breusch-Pagan (OLS):\n")
print(bp_test)

# Modèle à effets fixes (GLS "within")
fixed_model <- plm(log(fatal) ~ spirits + income + unemp + log(miles) + log(gsp) + beertax, data = panel_data, model = "within")
summary(fixed_model)

bp_test <- bptest(fixed_model)
cat("\nTest de Breusch-Pagan (OLS):\n")
print(bp_test)

#Test d'autocorrélation de Wooldridge
autocorr_test <- pbgtest(fixed_model)
```

```

cat("\nTest d'autocorrélation de Wooldridge:\n")
print(autocorr_test)

# Modèle à effets aléatoires (GLS "random")
random_model <- plm(log(fatal) ~ spirits + income + unemp + log(miles) + log(gsp) + beertax, data)
summary(random_model)

bp_test <- bptest(random_model)
cat("\nTest de Breusch-Pagan (OLS):\n")
print(bp_test)

#Test d'autocorrélation de Wooldridge
autocorr_test <- pbgtest(random_model)
cat("\nTest d'autocorrélation de Wooldridge:\n")
print(autocorr_test)

#Test de Fischer (test d'existence des effets individuels)
fisher_test <- pFtest(fixed_model, pooled_model)
cat("Test de Fischer:\n")
print(fisher_test)

#Test du multiplicateur de Lagrange (LM Test)
lagrange_test <- plmtest(pooled_model, effect = "individual", type = "bp")
cat("\nTest du multiplicateur de Lagrange (LM Test):\n")
print(lagrange_test)

#Test de Hausman (pour choisir entre effets fixes et aléatoires)
hausman_test <- phptest(fixed_model, random_model)
cat("\nTest de Hausman:\n")
print(hausman_test)

# Supprimer les lignes avec des valeurs manquantes dans les variables sélectionnées
panel_data_clean <- na.omit(panel_data[c("fatal", "spirits", "income", "unemp", "miles", "gsp", "beertax")])

# Calcul des statistiques descriptives manuellement
descriptive_stats <- data.frame(
  Variable = c("fatal", "spirits", "income", "unemp", "miles", "gsp", "beertax"),
  Mean = round(sapply(panel_data_clean[c("fatal", "spirits", "income", "unemp", "miles", "gsp", "beertax"), ],
    function(x) mean(x))),
  SD = round(sapply(panel_data_clean[c("fatal", "spirits", "income", "unemp", "miles", "gsp", "beertax"), ],
    function(x) sd(x))),
  Min = round(sapply(panel_data_clean[c("fatal", "spirits", "income", "unemp", "miles", "gsp", "beertax"), ],
    function(x) min(x))),
  Max = round(sapply(panel_data_clean[c("fatal", "spirits", "income", "unemp", "miles", "gsp", "beertax"), ],
    function(x) max(x))),
  Median = round(sapply(panel_data_clean[c("fatal", "spirits", "income", "unemp", "miles", "gsp", "beertax"), ],
    function(x) median(x))),
  Var = round(sapply(panel_data_clean[c("fatal", "spirits", "income", "unemp", "miles", "gsp", "beertax"), ],
    function(x) var(x))),
  N = sapply(panel_data_clean[c("fatal", "spirits", "income", "unemp", "miles", "gsp", "beertax"), ],
    function(x) length(x)))

descriptive_stats <- descriptive_stats[, -1]

# Afficher les statistiques descriptives ajustées
print(descriptive_stats)

# Tableau comparatif des différents modèles
stargazer(pooled_model, fixed_model, random_model, type = "text",

```

```

title = "Comparaison des modèles : Pooled OLS, Fixed Effects, Random Effects",
align = TRUE, no.space = TRUE,
column.labels = c("Pooled OLS", "Fixed Effects", "Random Effects"),
covariate.labels = c("Spiritueux", "Revenu", "Chômage", "Miles (log)", "GSP (log)", "T",
dep.var.labels = "Log(Fatalités)",
model.numbers = FALSE,
omit.stat = c("f", "ser"), # Omitting F-stat and standard error
star.cutoffs = c(0.1, 0.05, 0.01),
notes = "HC1 robust standard errors for Pooled OLS.")

# Calculer la matrice de corrélation
cor_matrix <- cor(panel_data_clean)

# Enregistrer la matrice de corrélation en PDF
pdf("matrice_correlation.pdf") # Ouvrir un périphérique graphique pour le PDF

# Afficher la matrice de corrélation avec des carreaux (square)
corrplot(cor_matrix, method = "square", type = "upper", tl.col = "black", tl.srt = 45,
         title = "Matrice de Corrélation", mar = c(0,0,2,0))
dev.off() # Fermer le périphérique graphique et enregistrer le fichier

```