5.1 **Cartel des prix de l'essences en Gaspésie** : Plusiers maires et préfets gaspésiens ont demandé à la Régie de l'énergie du Québec d'enquêter sur le prix de l'essence, beaucoup plus élevé en 2019 selon eux dans la région qu'ailleurs au Québec. Pour répliquer l'analyse de la Régie, les données suivantes ont été extraites du site de l'organisme gouvernemental pour la période 2014–2019. Les données `renergie` incluent les variables suivantes :

- `region` : région administrative, une parmi Bas-Saint-Laurent (1), Saguenay-Lac-Saint-Jean (2), Capitale-Nationale (3), Mauricie (4), Estrie (5), Montréal (6), Outaouais (7), Abitibi-Témiscamingue (8), Côte-Nord (9), Nord-du-Québec, excluant le Nunavik (10), Gaspésie-Îles-de-la-Madeleine (11), Chaudière-Appalaches (12), Laval (13), Lanaudière (14), Laurentides (15), Montérégie (16) et Centre-du-Québec (17).
- `date` : date hebdomadaire pour les prix minimum et moyens à la pompe en format `aaaa-mm-jj`.
- `pmin` : prix minimum (plancher) calculé par la Régie de l'énergie, incluant les taxes et les frais de transports.
- `pmoy` : prix moyen à la pompe affiché par les détaillants.

Faites une analyse des données longitudinales pour déterminer si la marge de profit des détaillants de la Gaspésie et des Îles-de-la-Madeleine est significativement plus élevée que partout ailleurs à l'aide d'une analyse de variance à un facteur qui prenne en compte la corrélation intra-région. *Indication : dans SAS, utilisez l'option `ddfm=satterth` pour le calcul des degrés de liberté avec la procédure `mixed`.*

(a) Tracez un graphique (a) du prix moyen et (b) de la différence entre le prix moyen et le prix minimum pour chaque région en fonction du temps. Commentez sur les différences observées entre ces deux graphiques.

Solution

Il est évident que le prix moyen de vente de l'essence est non-stationnaire et qu'il est soumis aux aléas des fluctuations des prix de l'essence, avec des changements saisonniers. En revanche, la marge de profit apparaît relativement stable (avec une baisse en 2019), même s'il y a d'importantes disparités entre régions.

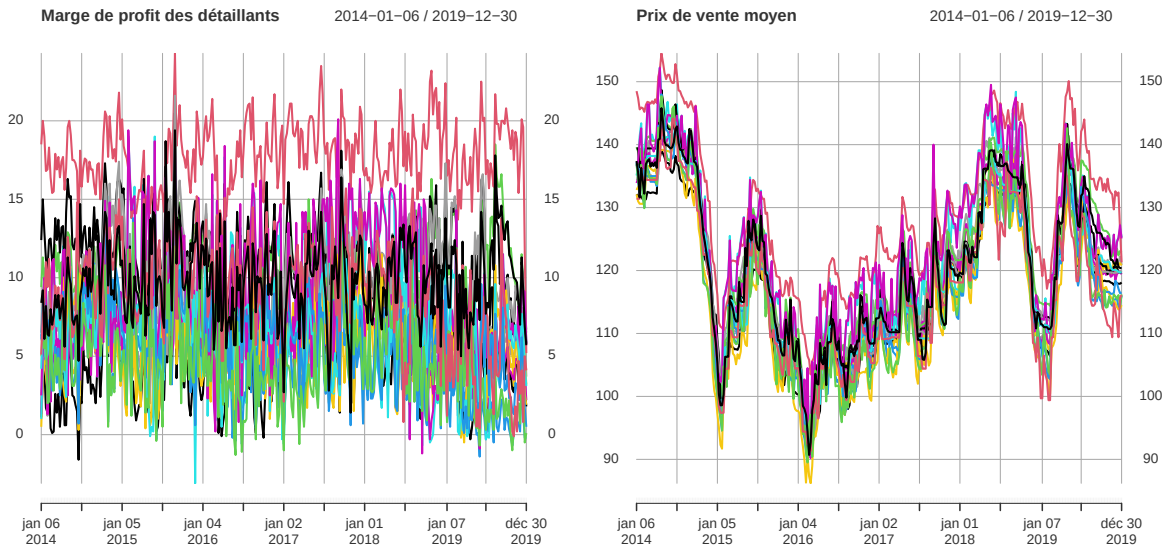

FIGURE 1 – Marge de profit hebdomadaire (en centimes CAD) et prix moyen au détail de l'essence ordinaire (en centimes CAD) calculés par la Régie de l'énergie.

(b) Sélectionnez un modèle de covariance adéquat pour modéliser la dépendence temporelle intra-région. Vous devez choisir un modèle de covariance parmi (a) indépendance (covariance diagonale), (b) AR(1), (c) équi-symmétrie et (d) non-structuré. Justifiez adéquatement votre choix.

Solution

Le modèle avec la covariance non-structurée n'est pas estimable, car il y a plus de paramètres (avec 313 réplications temporelles) que d'observations. Le modèle d'équisymmétrie (AIC = 26660,4 et BIC = 26662) ne semble pas différent du modèle d'indépendance. On peut ajuster un modèle autorégressif d'ordre un pour les erreurs ; la statistique du rapport de vraisemblance REML est 1021,81 pour un degré de liberté, ce qui est hautement significatif. C'est le modèle qui a les plus petites valeurs pour les critères d'information AIC (25638,6) et BIC (25640,2) et c'est également le seul qui est logique dans le contexte.

(c) Rapportez les erreurs-type de la marge de profit moyenne des détaillants de la région Gaspésie-Îles-de-la-Madeleine (GIM) pour le modèle de régression ordinaire (qui suppose l'indépendance entre observations) et le modèle autorégressif d'ordre 1. Laquelle est la plus grande ? expliquez pourquoi.

Solution

Les erreurs-type sont celles de l'ordonnée à l'origine (peu importe la région). Il est plus facile pour la suite de mettre la région GIM comme référence. On obtient 0,1672 pour le modèle de régression ordinaire, contre 0,2625 pour le modèle avec les erreurs autorégressives. Le rapport des variances est $(0,1672/0,2625)^2 = 0,4057$ et les observations corrélées contiennent une fraction de l'information contenue dans un échantillon avec des données indépendantes.

(d) Calculez la différence entre la marge de profit des détaillants de la Gaspésie-Îles-de-la-Madeleine et des autres régions en prenant en compte la corrélation intra-région. Quelles différences sont statistiquement significatives ?

Solution

Sur 16 comparaisons, 15 indiquent une différence de marge de profit moyenne significativement différence de celle de la région Gaspésie-Îles-de-la-Madeleine. La seule région pour laquelle la marge de profit est comparable est la Côte-Nord, pour laquelle la valeur-p associée est 0,1457. La région Nord-du-Québec affiche une marge de profit moyenne plus élevée qu'en Gaspésie, mais cette marge est inférieure pour toutes les autres régions (presque 5 centimes avec les grands centres).

5.2 **Enseignement de la lecture** : les données sont tirées de

> J. Baumann, N. Seifert-Kessell, L. Jones (1992), *Effect of Think-Aloud Instruction on Elementary Students' Comprehension Monitoring Abilities*, Journal of Reading Behavior, **24** (2), pp. 143–172.

Ces chercheurs ont fait une étude pour déterminer l'efficacité relative de méthodes d'apprentissage de la lecture. L'échantillon de 66 élèves de quatrième année primaire comporte 32 filles et 34 garçons qui ont été alloués de façon aléatoire aux trois groupes. On s'intéresse à l'amélioration des capacités de lecture par rapport à une méthode d'enseignement traditionnelle (DR). Deux tests ont été administrés avant et pendant l'expérience pour mesurer l'efficacité des méthodes ; afin de rendre les résultats comparables, ils ont été repondérés de telle sorte à ce que une note parfaite vaille 1.

Les données contiennent des renseignements sur

- `groupe` : unité expérimentale, une parmi lecture-experimental group, une parmi *directed reading-thinking activity* (`DRTA`), méthode de la pensée à voix haute (`TA`) et lecture dirigée (`DR`).
- `mpre` : moyenne du score de prédiction pré-intervention (standardisé) pour les tests de détection d'erreurs et decompréhension.
- `mpost` : même que `mpre`, mais pour les évaluations post-intervention.

Nous sommes intéressés tout d'abord par l'amélioration pour chacune des méthodes à l'aide de deux modèles.

(a) Dans leur article, Baumann *et al.* font une analyse de variance à un facteur pour les scores pré-intervention (`mpre`) avec le facteur `groupe`. Expliquez quel est l'utilité d'un tel test dans le contexte de l'étude.

Solution
Une analyse de variance à un facteur de chaque groupe sert à s'assurer qu'aucune groupe n'est plus fort/faible que les autres avant le début de l'expérience.

(b) Soit `dpp` = `mpost` − `mpre` la différence entre résultats standardisés post- et pré-intervention. Ajustez une analyse de variance à un facteur pour `dpp` avec le facteur `groupe` (modèle 1.1).

Solution
Le changement dans les habiletés de lecture pré- et post-intervention est différent selon les méthodes d'apprentissage. La statistique du test-F (effets de type 3) pour test l'hypothèse que le changement est identique vaut 15.986, à comparer à une loi $\mathsf{F}(2,63)$ distribution ; this yields a negligible p-value.

(c) Écrivez l'équation du modèle ajusté en termes de scores pré- et post-intervention et montrez que le modèle est un cas spécial d'un modèle de régression linéaire pour `mpost` avec un terme de décalage. Utilisez ce fait pour comparer le modèle d'analyse de variance à un facteur pour `dpp` avec `groupe` (modèle 1.1) à un modèle linéaire ayant `mpost` comme variable réponse et `mpre` et `groupe` comme variables explicatives (modèle 1.2). Au vu de l'ajustement de ce dernier, est-ce que le modèle d'analyse de variance est adéquat ? Justifiez votre réponse.

Solution
L'équation du modèle est

$$\texttt{mpost} = \beta_0 + \beta_1 \texttt{DRTA} + \beta_2 \texttt{TA} + \texttt{mpre} + \varepsilon \tag{5.2.1}$$

$$\texttt{mpost} = \beta_0 + \beta_1 \texttt{DRTA} + \beta_2 \texttt{TA} + \beta_3 \texttt{mpre} + \varepsilon \tag{5.2.2}$$

Si on fixe $\beta_3 = 1$ dans le modèle 5.2.1, on recouvre le modèle 5.2.2.
L'intervalle de confiance de Wald à 95% pour le coefficient β_3 est $0,6017 \pm 1,96 \times 0,08327 = [0,438; 0,765]$. Pour tester $\mathscr{H}_0 : \beta_3 = 1$ à niveau 5%, il suffit de regarder si cette valeur est dans l'intervalle de confiance. On rejette donc l'hypothèse nulle que l'analyse de variance à un facteur du modèle 5.2.1 est une simplification adéquate du modèle de régression linéaire 5.2.2.

Transformez les données de format court à format long ; ce dernier est plus convenable pour l'analyse de données longitudinales. En plus de `groupe`, vos données devraient contenir les colonnes suivantes

— `id` : identifiant de l'élève.
— `score` : moyenne pour l'évaluation.
— `test` : variable catégorielle, une de `mpost` ou `mpre`, qui renseigne sur le score correspond à la moyenne pré-intervention ou post-intervention.

Le Tableau 1 présente le format final que vous devriez obtenir.
On traite désormais les observations comme des données longitudinales et on considère deux modèles pour `score` en fonction du `groupe` et de `test`, en incluant un terme d'interaction entre les deux, mais avec des modèles de covariance intra-individus différents :

— modèle 5.2.3 avec une structure d'équicorrélation pour les erreurs ;
— modèle 5.2.4 avec une covariance non structurée.

(d) Expliquez quel est la différence fondamentale entre les modèles 5.2.3-.4 et 5.2.2.

Solution
Dans le modèle 5.2.2, on conditionne sur `mpre` tandis qu'elle est aléatoire dans l'autre ; les scores pré- et post-interventions servent à estimer la variance pour les données en format long. On tient aussi compte de la corrélation entre les résultats d'un(e) même élève.

Écrivez la matrice de covariance des erreurs pour le modèle 5.2.3 et rapportez les corrélations estimées entre pré-interventions et post-interventions pour un(e) élève.

groupe	test	score	id
DR	mpre	0,23	1
DR	mpost	0,27	1
DR	mpre	0,35	2
DR	mpost	0,42	2
DR	mpre	0,41	3
DR	mpost	0,24	3
DR	mpre	0,57	4
DR	mpost	0,39	4
DR	mpre	0,67	5
DR	mpost	0,56	5

Tableau 1 – Premières 10 lignes des données de Baumann en format long

Solution
La covariance entre `mpre` et `mpost` est

$$\mathsf{Cov}\,(\boldsymbol{\varepsilon}_i) = \begin{pmatrix} \sigma^2+\tau & \tau \\ \tau & \sigma^2+\tau \end{pmatrix}, \qquad \widehat{\mathsf{Cor}}(\boldsymbol{\varepsilon}_i) = \begin{pmatrix} 1 & \frac{\widehat{\tau}}{\widehat{\sigma}^2+\widehat{\tau}} \\ \frac{\widehat{\tau}}{\widehat{\sigma}^2+\widehat{\tau}} & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0.6715 \\ 0.6715 & 1 \end{pmatrix}.$$

(e) À l'aide des sorties des modèles 5.2.3 et 5.2.4, testez si la variance des scores moyens pré-intervention et post-intervention sont les mêmes. Écrivez le nom du test que vous utilisez, la valeur numérique de la statistique et calculez la valeur-p avant de conclure dans le contexte du problème.

Solution
- On utilise un test du rapport de vraisemblance pour la comparaison, puisque le modèle d'équisymmétrie et le modèle de covariance non-structurée sont emboîtés.
- Deux fois la log-vraisemblance du modèle 5.2.3 est $-174,1$, tandis qu'on obtient $-175,7$ pour le modèle 5.2.4, soit une différence de 1,6.
- La loi nulle asymptotique du test est χ^2_1; on rejette l'hypothèse nulle d'homoscédasticité intra-individu si $D > 3.84$, c'est-à-dire si la statistique du rapport de vraisemblance excède le 95% percentile de la loi nulle (la valeur-p est 0,206).

(f) Puisqu'on a affaire à des données longitudinales, il serait logique de considérer en plus des deux modèles de covariance, un modèle autorégressif d'ordre 1, ou modèle $\mathsf{AR}(1)$, pour les erreurs. Est-ce que ça serait utile dans ce cas? Justifiez votre réponse.

Solution
Non, puisqu'il y a une seule période de temps et un seul paramètre de covariance pour la matrice intra-individu 2×2 - on obtiendrait le même modèle qu'avec le modèle d'équisymmétrie 5.2.3.

(g) Jusqu'à présent, on a supposé que les résultats pré- et post-intervention de tous les élèves avaient la même matrice de covariance. On pourrait cependant supposer que les paramètres de cette matrice de covariance diffèrent d'une méthode d'apprentissage à une autre. Est-ce que les données corroborent cette hypothèse?

Solution
Non. On peut ajuster un modèle d'équisymmétrie avec des paramètres différents pour chaque groupe. Le modèle d'équisymmétrie 5.2.3 sera emboîté puisqu'on le recouvre en fixant $\mathscr{H}_0 : \tau_{\texttt{DR}} = \tau_{\texttt{TA}} = \tau_{\texttt{DRTA}}$ et $\sigma^2_{\texttt{DR}} = \sigma^2_{\texttt{TA}} = \sigma^2_{\texttt{DRTA}}$. La log-vraisemblance du modèle sous l'alternative est $-175,7$ et la statistique du rapport de vraisemblance est $1,6$, à comparer avec une loi χ^2_4. On ne rejette pas l'hypothèse nulle à effet de quoi les paramètres

du modèle d'équisymmétrie sont les mêmes pour toutes les méthodes d'apprentissage, ce qui implique que la complexité additionnelle n'améliore pas significativement l'ajustement du modèle.

(h) Utilisez le modèle 5.2.4 pour déterminer si les résultats pour les méthodes d'enseignement `DRTA` et `TA` sont significativement meilleures que la méthode standard `DR`.

Solution

Cette hypothèse revient à avoir le même surenchérissement, soit à regarder si le terme d'interaction est null. Selon le modèle, l'interaction `test*groupe` est statistiquement significative. L'amélioration pour la méthode `DRTA` versus `DR` pour le score post-intervention est de 0,1457 (0,032) et celle de la méthode `TA` (par rapport à la méthode `DR`) est de 0,1656 (0,032). Les scores prédits pour`DR`, `DRTA` et `TA` pour les tests post-intervention sont 0,3629 ; 0,4786 et 0,475. Cela implique une baisse de résultat pour la méthode `DR` de 0,141, tandis que les résultats des autres méthodes se maintiennent.

5.3 **Tolérance d'adolescents face à la délinquance** : Les données proviennent du *American National Longitudinal Survey of Youth*, une étude longitudinale qui a démarré en 1997 et qui suit une cohorte de jeunes Américains nés entre 1980 et 1984. Un total de 8984 participants âgés de 12 à 17 sont inclus pour la première fois en 1997 et a été suivie à 15 reprises jusqu'à maintenant.

On considère 16 individus qui ont répondu aux cinq premières vagues d'entrevue entre l'âge de 11 et 15 ans, avec un suivi annuel. , et en particulier à certaines variables mesurant la tolérance des jeunes face aux comportements délinquants. Les données disponibles nous permettent de suivre l'évolution de 16 jeunes chaque année entre l'âge de 11 ans et 15 ans (donc 5 mesures pour chaque individu). Chaque année au début de l'étude puis tous les 2 ans ensuite, les participants ont rempli un questionnaire permettant d'évaluer leur tolérance face aux comportements délinquants. À l'aide d'une échelle à 4 points où les choix sont « très mal » (1), « mal » (2), « un peu mal » (3) et « tout à fait acceptable » (4), les jeunes indiquaient s'ils qualifiaient de « mal » pour quelqu'un de leur âge de (a) tricher à un examen, (b) détruire le bien d'autrui à dessein, (c) fumer de la marijuana, (d) voler quelque chose d'une valeur de moins de 5 dollars, (e) frapper ou menacer quelqu'un sans raison, (f) consommer de l'alcool, (g) entrer par effraction dans un bâtiment ou véhicule afin de voler, (h) vendre des drogues dures et (i) voler quelque chose d'une valeur de plus de 50$. Chaque score a été mesuré sur une échelle de Likert allant de très mal (1) à complètement acceptable (4). Les données `tolerance` incluent les variables suivantes :

- `id` : identifiant du participant.
- `age` : âge du participant lors du suivi
- `tolerance` : moyenne du score pour les neuf questions sur la tolérance à la délinquence.
- `sexe` : indicateur binaire, un pour les hommes, zéro pour les femmes.
- `exposition` : score moyen de l'exposition du participant à 11 ans aux comportements délinquants dans son entourage. Cette variable est une estimation du participant de la proportion de ses ami(e)s qui ont été impliqué(e)s dans chacune des activités (a) à (i) décrites plus haut.

(a) Présentez et interprétez les statistiques descriptives des variables `tolerance`, `sexe` and `exposition`.

Solution

Étant donné que `exposition` et `sexe` prennent des valeurs fixes à tous les pas de temps, leurs statistiques descriptives devraient être calculées pour un âge fixe.

- `sexe` : 43,75% des participants sont des hommes.
- `exposition` : moyenne (erreur-type) égale à 1,19 (0.08), valeurs variant entre 0,81 et 1,99.

La tolérance varie au fil du temps et on va donc utiliser les 80 observations pour calculer les statistiques descriptives de cette variable. La moyenne (erreur-type) de la tolérance est $1,62(0,055)$ et ses valeurs varient entre 1 et 3,46.

(b) Évaluez graphiquement la relation entre `tolerance` et les variables `sexe`, `exposition` et `age`. Résumez brièvement vos observations.

Solution

Les diagrammes sont présentés dans la Figure 2. On remarque peu de différences entre les hommes et les

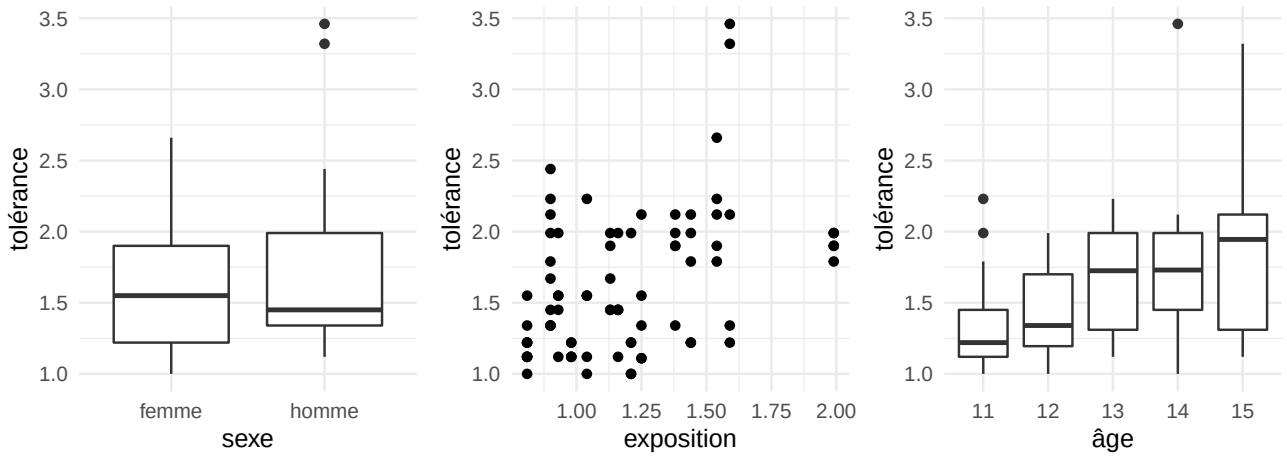

FIGURE 2 – Boîte-à-moustache et nuage de points de la variable réponse `tolerance` contre les variables explicatives.

femmes, même si en moyenne, les femmes sont plus tolérantes. On remarque deux valeurs aberrantes de la tolérance. La tolérance semble augmenter d'une manière linéaire avec l'âge. La tolérance semble également augmenter avec l'exposition, même si la relation entre ces deux variables est plus nette pour des expositions basses et élevées. Finalement, on remarque une plus grande hétérogénéité de la tolérance à 15 ans.

(c) Faites un graphique des trajectoires de la tolérance aux comportements délinquants en fonction de l'âge du participant et commentez.

Solution

Pour la plupart des participants, les trajectoires du score de tolérance dans la Figure 3 indiquent une décroissance au cours du temps. Le graphique en spaghetti montre que les deux scores de tolérance les plus élevés, qui se trouvent à l'extérieur des moustaches dans le graphique de gauche de la Figure 2, appartiennent à la même personne.

(d) En prenant en compte la corrélation intra-sujet (au cours des cinq années) et en assumant cette dernière fixe peu importe l'année, ajustez le modèle

$$\begin{aligned}\texttt{tolerance}_{ij} &= \beta_0 + \beta_1 \texttt{sexe}_{ij} + \beta_2 \texttt{exposition}_{ij} + \beta_3 \texttt{age}_{ij} + \varepsilon_{ij},\\ \boldsymbol{\varepsilon}_i &\overset{\text{iid}}{\sim} \mathcal{N}(\mathbf{0}_5, \Sigma_i), \boldsymbol{\Sigma}_i \sim \mathsf{CS}, i = 1,\ldots,16; j = 1,\ldots,5. \end{aligned} \tag{M$_1$}$$

i. Interprétez l'effet des variables du modèle et commentez sur les résultats.
ii. Calculez la corrélation entre deux valeurs de tolérances pour les mesures à l'âge 11 et 12 ans, de même qu'entre 11 et 15 ans.

Solution

Les estimés des coefficients sont $\widehat{\beta}_0 = -1,07(0,45)$, $\widehat{\beta}_1 = 0,23(0,13)$, $\widehat{\beta}_2 = 0,74(0,20)$, $\widehat{\beta}_3 = 0,13(0,03)$, $\widehat{\sigma}^2 = 0,127$ et $\widehat{\tau} = 0,035$.

— L'interprétation de l'ordonnée à l'origine ne fait aucun sens, car l'âge ne peut pas être nul.
— L'estimé du coefficient de `sexe` est $0,229$ et reflète un score plus élevé pour les hommes que pour les femmes, *ceteris paribus*. Cependant, cet effet n'est pas statistiquement significatif.

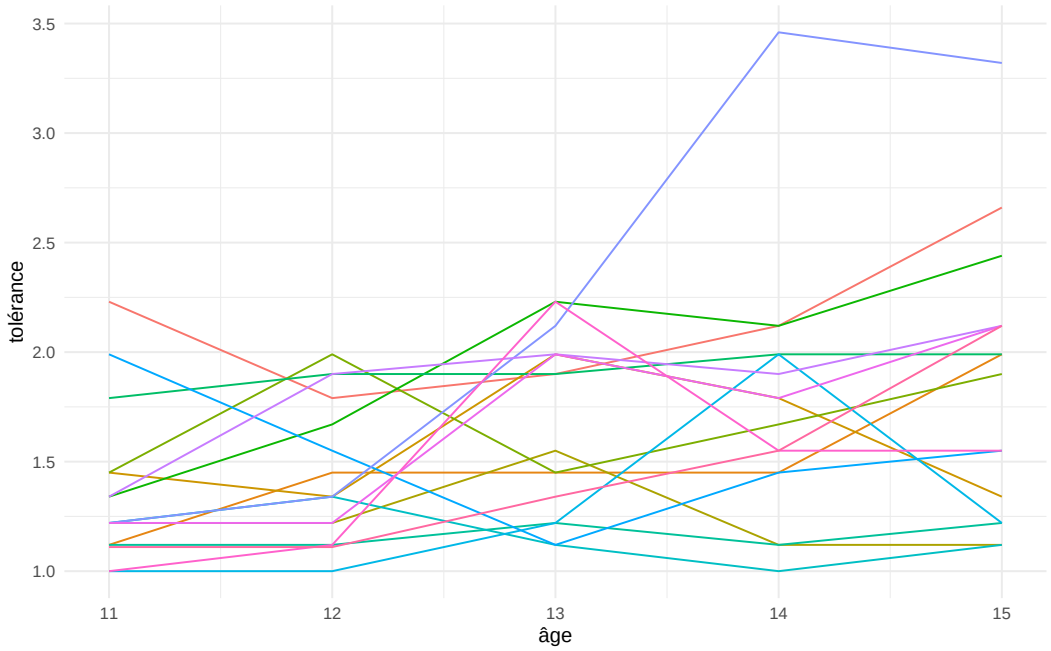

FIGURE 3 – Graphique en spaghetti pour les trajectoires du score de tolérance de 16 individus participant à l'étude.

— L'effet de l'exposition est statistiquement significatif (avec une valeur-p de $0,0024$). Les personnes qui sont plus exposées aux comportements délinquants sont en moyenne plus tolérantes, avec une augmentation estimée à $0,745$ quand le score moyen d'exposition augmente de un et que l'âge et le sexe restent constants.
— Le score de tolérance moyen augmente de $0,13$ par année, *ceteris paribus*. Cet effet est statistiquement significatif.
— La corrélation entre individus, supposée constante pour chaque année, est $\widehat{\rho} = 0,22$ et est significativement différente de zéro (le test du rapport de vraisemblance teste $\mathscr{H}_0 : \tau = 0$). La valeur-p égale à $0,0257$ indique que le modèle avec une structure d'équicorrélation résulte en un ajustement qui est significativement meilleur que le modèle où on suppose l'indépendance des observations.

(e) En supposant une structure autorégressive d'ordre 1 pour les erreurs, ajustez le modèle

$$\begin{aligned}\texttt{tolerance}_{ij} &= \beta_0 + \beta_1 \texttt{sexe}_{ij} + \beta_2 \texttt{exposition}_{ij} + \beta_3 \texttt{age}_{ij} + \varepsilon_{ij},\\ \boldsymbol{\varepsilon}_i &\overset{\text{iid}}{\sim} \mathcal{N}(\mathbf{0}_5, \Sigma_i), \boldsymbol{\Sigma}_i \sim \mathsf{AR}_1, i = 1,\ldots,16; j = 1,\ldots,5. \end{aligned} \tag{M$_2$}$$

i. Identifiez tous les paramètres de la matrice de covariance.
ii. Calculez la corrélation entre deux valeurs de tolérances pour les mesures à l'âge 11 et 12 ans, de même qu'entre 11 et 15 ans et comparez ces corrélations avec celles obtenues pour le modèle d'équicorrélation.

Solution

La corrélation entre deux observations espacées par une seule période est estimée à $\widehat{\rho} = 0,54$ et la variance est estimée à $\widehat{\sigma}^2 = 0,17$. La corrélation décroit de manière géométrique dans le modèle $\mathsf{AR}(1)$, à l'opposé du modèle avec une structure d'équicorrélation où la corrélation reste constante. La corrélation entre deux mesures à l'âge de 11 et 12 ans est $\widehat{\rho} = 0,54$ et celle entre deux mesures à l'âge de 11 et 15 ans est $\widehat{\rho}^4 = 0,086$.

(f) En supposant l'indépendance des observations, ajustez le modèle

$$\begin{aligned} \texttt{tolerance}_{ij} &= \beta_0 + \beta_1 \texttt{sexe}_{ij} + \beta_2 \texttt{exposition}_{ij} + \beta_3 \texttt{age}_{ij} + \varepsilon_{ij}, \\ \boldsymbol{\varepsilon}_i &\overset{\text{iid}}{\sim} \mathcal{N}(\mathbf{0}_5, \sigma^2 \mathbf{I}_5), i = 1, \ldots, 16; j = 1, \ldots, 5, \end{aligned} \tag{M$_3$}$$

Lequel des modèles M_1, M_2 et M_3 choisiriez- vous? Justifiez votre choix.

Solution

Les deux modèles M_1 et M_2 mènent à de meilleurs ajustements que le modèle M_3 selon les tests du rapport de vraisemblance. Afin de comparer les modèles M_1 et M_2, qui ne sont pas emboîtés, on peut comparer leurs critères d'information. La valeur du AIC (BIC) pour le modèle M_1 est 88,9 (90,4) et celui du modèle M_2 est 74,5 (76,1), reflétant ainsi une préférence pour le modèle avec une structure autorégressive d'ordre 1. La seule différence entre ces trois modèles réside dans la valeur de la corrélation intra-groupe.

— M_1 suppose que les observations d'un même individu sont corrélées dans le temps et que cette corrélation, égale à $(\sigma^2 + \tau)$, est constante.
— M_2 suppose que les observations d'un même individu sont corrélées dans le temps et que cette corrélation suit un processus autorégressif d'ordre un, c'est-à-dire que la corrélation ρ^h pour deux observations séparées par h années décroît suivant une suite géométrique.
— M_3 suppose que les observations d'un même individu sont indépendantes d'une année à une autre.