

Travail Pratique I
Modèles linéaires en actuariat
ACT-2003

Marie-Pier CÔTÉ
AUTOMNE 2018

1 Consignes générales

- Le travail doit être effectué en équipe de 3 à 4 personnes.
- Date limite pour la remise : mercredi le 7 novembre 2018 à 23h59.
- La remise du rapport doit se faire sur la boîte de dépôt du site de cours. Aucun rapport en format papier ne sera accepté.
- L'analyse de données peut être effectuée en R ou en Excel (ou les deux). Vous n'avez pas à remettre le code informatique.
- Le rapport, incluant les formules mathématiques, doit être tapé avec un logiciel approprié (Word, L^AT_EX, etc.) Un gabarit L^AT_EX est disponible sur le site de cours si vous souhaitez l'utiliser.
- Le rapport doit être rédigé de façon structurée. Les graphiques et tableaux doivent être expliqués dans le corps du texte.
- 10% des points sont accordés pour la qualité du français, à raison de 1/2 point par faute d'orthographe ou de grammaire.
- 10% des points sont accordés pour la qualité de la présentation, et pour la présence d'une page titre adéquate, d'une table des matières et d'un sommaire exécutif.

2 Description du projet

Les réclamations pour blessures corporelles liées à des accidents automobiles représentent un risque important pour la compagnie d'assurance pour laquelle vous travaillez.

Vous avez accès à la base de données `AutoBodyInjury.csv` provenant de l'*Insurance Research Council* (IRC), une division de l'*American Institute for Chartered Property Casualty Underwriters* et de l'*Insurance Institute of America*. Ces données ont été recueillies en 2002 aux États-Unis. L'échantillon contient 1336 observations, provenant toutes d'un même état.

Les variables suivantes sont contenues dans la base de données `AutoBodyInjury.csv`, disponible sur le site de cours :

Variable	Description
CASENUM	Numéro d'identification de la réclamation
ATTORNEY	Indique si le réclamant ou la réclamante est représenté(e) par un avocat (=1 si oui et =2 sinon)
CLMSEX	Sexe du réclamant ou de la réclamante ("M" pour homme ou "F" pour femme)
MARITAL	Statut matrimonial du réclamant ou de la réclamante ("married" pour marié(e), "single" pour célibataire, "widowed" pour veuf ou veuve, ou "divorced" pour divorcé(e))
CLMINSUR	Indique si la conductrice ou le conducteur du véhicule est assuré(e) (=1 si non, =2 si oui, et =3 si non applicable)
SEATBELT	Indique si le réclamant ou la réclamante portait une ceinture de sécurité ou était attaché(e) dans un siège pour enfant (=1 si oui, =2 si non, et =3 si non applicable)
CLMAGE	Âge du réclamant ou de la réclamante
LOSS	Perte économique totale du réclamant ou de la réclamante en milliers de dollars

Vous êtes des membres de l'équipe actuariat corporatif de la compagnie Assurancetourix, qui offre un produit d'assurance automobile couvrant les blessures corporelles dans cet état. Votre directeur vous demande de jeter un coup d'oeil aux données recueillies pour déterminer s'il existe un lien entre la perte économique et les variables disponibles. Si ce lien existe, vous devez proposer un modèle de régression linéaire, qui pourra être utilisé aux fins de prévisions.

En fait, au moment où un accident survient, la compagnie Assurancetourix doit établir une réserve pour couvrir le montant lié à la réclamation. Pour l'instant, le modèle utilisé est simpliste ; la réserve est toujours établie à 6000 \$. Votre mandat est d'élaborer un modèle qui aura une meilleure puissance prédictive.

3 Contenu du rapport

Votre directeur connaît les rudiments de la régression linéaire multiple. Il n'est donc pas nécessaire d'expliquer la théorie dans votre rapport. Par contre, il s'attend à lire une analyse complète des données et des résultats, qui comprendra les éléments suivants :

Une page titre

Un sommaire exécutif : Seule cette partie du rapport sera lue par la vice-présidente de l'actuariat corporatif. Elle doit donc être brève (maximum 250 mots) tout en contenant les informations importantes sur votre analyse et vos conclusions.

Vous pouvez consulter ce lien : [Conseils sur l'écriture d'un sommaire exécutif](#).

Une table des matières

Une analyse univariée des données : Dans cette section, votre directeur s'attend à lire un résumé clair et concis des données. Cela peut être fait à l'aide de tableaux de fréquence, d'histogrammes ou d'autres types de diagrammes. Les moyennes, écart-types, médiane, minimum et maximum sont également intéressants pour des variables continues. Expliquez brièvement les données. Faites les graphiques en nuage de points de la variable endogène en fonction de chacune des variables exogènes à considérer.

Note : Si vous identifiez des erreurs dans les données, elles doivent d'abord être corrigées. Mentionnez ces erreurs et la façon dont elles ont été corrigées dans l'annexe A de votre rapport.

Explication du modèle proposé : Dans cette section, présentez les éléments suivants

1. l'équation du modèle proposé ;
2. le traitement des variables qualitatives dans le modèle ;
3. l'explication des interactions entre les variables du modèle (s'il y a lieu) ;
4. l'interprétation des paramètres du modèle ;
5. des intervalles de confiance à 95% sur chacun des paramètres du modèle ;
6. le coefficient de détermination ajusté et son interprétation ;
7. le tableau ANOVA ainsi que des commentaires sur la validité globale de la régression.

Attention : votre directeur ne veut pas voir toutes les étapes que vous avez dû faire pour atteindre ce modèle que vous préférez. Vous devez donc sélectionner les variables qui sont significatives, ou effectuer les transformations sur les variables (si nécessaire) avant de proposer votre modèle final.

Analyse des résidus et test pour manque d'ajustement : Dans cette section, expliquez si les postulats du modèle de régression linéaire sont vérifiés. Présentez les graphiques et les analyses reliées à la vérification des postulats. Effectuez un test pour manque d'ajustement et commentez brièvement.

Prévisions : Votre directeur veut savoir quelle sera, selon votre modèle, la perte économique totale pour une réclamation qui s’est passée hier. Il s’agit d’un homme célibataire de 45 ans, qui portait sa ceinture de sécurité au moment de l’accident. Le conducteur du véhicule était assuré et le blessé est représenté par un avocat. Dans cette section, donnez un intervalle de confiance à 95% pour la “vraie” valeur de la perte économique totale en milliers de dollars. Comparez cet intervalle de confiance avec celui pour la *moyenne* de la perte économique totale pour un accident avec les mêmes caractéristiques. Expliquez pourquoi les deux intervalles sont si différents.

Donnez également les intervalles de confiance pour *la moyenne* de la perte économique totale pour des réclamations avec les caractéristiques suivantes, et interprétez brièvement les résultats.

CLMAGE	MARITAL	CLMSEX	SEATBELT	CLMINSUR	ATTORNEY
70	single	M	1	1	1
45	married	M	1	1	1
45	divorced	M	1	1	1
45	widowed	M	1	1	1
45	single	F	1	1	1
45	single	M	2	1	1
45	single	M	1	2	1
45	single	M	1	1	2
22	single	F	2	1	2

Recommandations : Pour conclure votre rapport, donnez vos recommandations à votre directeur (maximum une page). Le modèle que vous proposez est-il meilleur que celui qui est présentement utilisé ? Y a-t-il d’autres considérations ou ajustements dont vous devriez tenir compte ?

Annexe A : Listez les erreurs dans les données et les modifications apportées (maximum 1/2 page).

Annexe B : Si vous avez effectué une transformation, expliquez pourquoi et quelle méthode vous avez utilisé (maximum une page).

Annexe C : Expliquez comment vous avez sélectionné les variables (maximum une page). Si vous avez fait des tests F partiels, ou une sélection selon un critère en particulier (critère AIC, Cp de Mallows), présentez les dans cette section. Votre directeur doit pouvoir comprendre pourquoi certaines variables sont exclues du modèle alors que d’autres sont conservées.

Amusez-vous bien ! ☺