Vous devez remettre un rapport d'au plus **8 pages** (excluant la page de titre, la table des matières et la bibliographie). Le projet compte pour 20% de la note finale seulement si la note combinée des évaluations individuelles est supérieure à 50%.

Instructions

1. Vous devez travailler en équipes de trois ou quatre.
2. **Date de remise**: au plus tard le mercredi 25 mars 2020 à 16h00. Les projets remis en retard seront pénalisés et sujets à une déduction cumulative de 20% de la note maximale atteignable pour chaque jour de retard.
3. Vous devez remettre une copie papier du rapport final en classe ou au secrétariat du département des Sciences de la décision (local 4.632, CSC).
4. Vous devez remettre les documents suivants en ligne dans Remise des travaux sur *ZoneCours*
   (a) Un document PDF contenant le rapport du projet (un par équipe).
   (b) Un fichier `.sas` ou `.R` (encodage `UTF8`) contenant le code **nettoyé** et **commenté** de vos analyses. Assurez-vous de la reproductibilité de vos résultats indépendamment de l'ordinateur ou du système d'exploitation.

Choisissez une base de données et **faites** une analyse complète de ces données à l'aide de la régression linéaire afin de **répondre** aux questions de recherche.
Trois bases de données vous sont proposées. Vous pouvez choisir l'une d'entre elles ou trouver une base de données publique de votre choix. Les équipes qui décident de travailler avec d'autres données que celles fournies doivent **contacter** la chargée de cours par courriel et lui **envoyer** une description complète de ces dernières à des fins de validation.

Structure du rapport

Introduction Décrivez brièvement les objectifs de votre analyse, le type de données analysées, ainsi que les questions clés qui méritent une attention particulière.

Analyse Décrivez votre jeu de données (incluant une analyse exploratoire) ainsi que les méthodes d'analyse statistique utilisées en vos propres termes. Rapportez seulement les éléments clés de votre analyse. Votre rapport doit avoir la forme d'un texte et donc ne doit pas contenir de code `SAS` ou `R` ni de sorties directes de ces logiciels.

Discussion approfondie des résultats: n'inclure que les figures et tableaux les plus pertinents pour votre analyse et décrire toute conclusion que vous en tirez. Toutes les tableaux et figures doivent être formatées et bien expliquées (titre, unités sur les axes, description…) Assurez-vous que les tableaux, les axes des diagrammes ainsi que les légendes soient lisibles et informatifs. Finalement, assurez-vous de faire référence dans le texte à chaque figure et/ou tableau que vous incluez.

Conclusion Discutez des éléments pertinents de votre analyse et du message à retenir. Vous devez convaincre le lecteur de vos aptitudes à mener cette analyse de données, tout en mettant en valeur les éléments de votre analyse à améliorer.

Éléments de discussion

Analyse exploratoire des données type des variables explicative, relation entre variables explicatives et réponse, présence de valeurs aberrantes ou de valeurs manquantes, l'utilité (ou pas) d'une potentielle transformation de la variable réponse et/ou des variables explicatives.

Modélisation choix des outils statistiques en relation avec les données, sélection de modèle et des variables importantes, résumé des modèles ajustés, diagnostics graphiques les plus pertinents, détection de potentiels problèmes de colinéarité, discussion des critères de validité du modèle et de la robustesse de vos résultats face à la violation potentielle de ces critères.

Compiled 19/02/2020 at 14:29

Interprétation Tests statistiques et diagnostics graphiques pour vérifier la validité des hypothèses du modèle. Interprétation des coefficients du modèle final que vous aurez sélectionné sur une échelle adéquate.

Discussion : Discussion portant sur les questions de recherche formulées, en vous basant sur le modèle ajusté. Discussions des limitations de l'analyse.

Commentaires additionnels

1. Évitez les fautes d'orthographes, les erreurs typographiques et les incohérences. Veuillez lire attentivement votre rapport avant de le remettre et utilisez un correcteur d'ortographe au besoin.
2. Bien que votre rapport ne doive contenir que les résultats les plus pertinents, votre code peut inclure des analyses additionnelles (analyse exploratoire supplémentaire, modèles et diagnostics).
3. Votre code doit être commenté et annoté, hormis les commandes explicites.
4. Le niveau technique de votre rapport devrait correspondre à un document pouvant être compris par des étudiants à la maîtrise en gestion à HEC Montréal. En lisant la partie méthode de votre rapport, un lecteur indépendant devrait être en mesure de pouvoir reproduire exactement votre analyse.

Compiled 19/02/2020 at 14:29

Grille d'évaluation

1. Profondeur de l'analyse	Vaste palette d'outils statistiques analyse approfondie	3	2	Peu d'idées, choix limité des outils statistiques 1 0
2. Analyse exploratoire des données	Analyse approfondie, correcte	3	2	Analyse superficielle, incorrecte 1 0
3. Adéquation et justesse de la méthode statistique utilisée	Convenable, utilisation adéquate des outils statistiques	3	2	Incorrecte, plusieurs erreurs 1 0
4. Description de la méthode statistique	Détaillée et claire	3	2	Pas reproductible, pas claire 1 0
5. Présentation claire des résultats (tableaux, figures)	Adéquats, bien formatés et expliqués titre adéquat et bonne discussion, nombre adéquat de décimales	3	2	Inadéquate, pas de discussion ou discussion non motivée et sorties pas formatées 1 0
6. Interprétation des résultats	Correcte et appropriée	3	2	Peu ou pas d'interprétation 1 0
7. Conclusion, discussion et critique	Claires et réfléchies	3	2	Banales et inadéquates 1 0
8. Questions de recherche	Formulées d'une manière claire Réponse détaillée	3	2	Objectifs confus Réponse superficielle ou inexistante 1 0
9. Qualité de rédaction	Grammaire et ponctuation correctes, formules mathématiques formatées	3	2	Erreurs de grammaire, syntaxe et de ponctuation 1 0
10. Références et reproductibilité	Références complètes, appropriées et détaillées Code commenté et analyse reproductible	3	2	Références inadéquates code peu ou pas commenté pas reproductible 1 0

Total (max 30)

Compiled 19/02/2020 at 14:29

Projet 1: salaires aux É.U. d'Amérique

Les données `wages85` sont extraites du livre

> E. R. Berndt. *The practice of econometrics: classic and contemporary*, 1991, Addison-Wesley Pub. Co., 702 p.

Voici une traduction libre de la description de cette base de données

> *Cette base de données contient les informations de 534 travailleurs en emploi, sélectionnés de manière aléatoire lors du recensement de mai 1985 de la population active des États-Unis, mené par le département américain du commerce. Ce recensement mensuel de plus de 50,000 ménages sert aux statistiques sur l'emploi et le chômage. Les données collectées incluent certaines charactéristiques socio-démographiques individuelles ainsi que le statut d'emploi.*

Le jeu de données contient les variables suivantes:

- `ED`: années d'éducation
- `SOUTH`: 1 si la personne réside au Sud
- `NONWH`: 0 si la personne est Caucassienne
- `HISP`: 1 si la personne est hispanique
- `FE`: 1 si la personne est une femme
- `MARR`: 1 si marié et conjoint présent (dans le même ménage)
- `MARRFE`: 1 si la personne est une femme mariée et que son époux est présent (dans le même ménage)
- `EX`: années d'expérience sur le marché du travail (`AGE-ED`−6) (minimum = 0 imposé *ex post*)
- `EXSQ`: années d'expérience sur le marché du travail, au carré
- `UNION`: 1 si l'emploi est syndiqué
- `LNWAGE`: logarithme naturel des gains horaires moyens (en USD)
- `AGE`: âge (en années)
- `MANUF`: 1 si la personne travaille dans l'industrie manufacturière
- `CONSTR`: 1 si la personne travaille dans l'industrie de la construction
- `MANAG`: occupation, 1 si la personne est gestionnaire ou occupe un poste administratif
- `SALES`: occupation, 1 si la personne travaille dans le secteur des ventes
- `CLER`: occupation, 1 si employé de bureau
- `SERV`: occupation, 1 si la personne travaille dans le secteur des services
- `PROF`: occupation, 1 si technicien ou professionnel

Discutez dans votre rapport des facteurs qui influencent les gains horaires et leur variabilité à l'aide d'un modèle linéaire parsimonieux. Votre rapport devra aussi répondre aux questions suivantes:

1. Quel modèles est le plus approprié pour analyser les données, un modèle multiplicatif pour `LNWAGE` ou un modèle additif pour exp(`LNWAGE`)?
2. Quantifiez les effets nets sur les gains horaires des années d'éducation, des années d'expérience sur le marché du travail, ainsi que de la syndicalisation.
3. Évaluez l'existence d'inéquités salariales entre hommes et des femmes et tester sa significativité, en tenant compte de la présence de facteurs confondants et de l'effet des autres variables explicatives.

Compiled 19/02/2020 at 14:29

Projet 2: prix de voitures de marque Ford

Le jeu de données `ford` a été extrait du site web `cars.com` en décembre 2018, par Joseph Naberhaus. Les données fournies ont été légèrement modifiées. Le jeu de données contient les variables suivantes pour $n = 32\,556$ voitures:

- `price`: prix (en USD)
- `miles`: kilométrage (en miles)
- `fuel_type`: type de combustible, soit hybride électrique et essence (`hybrid`), diesel (`diesel`), essence (`gasoline`), ou mélange éthanol/essence (`E85FlexFuel`).
- `drivetrain`: type de corps de roue libre, voiture à quatre roues motrices `4WD`, voiture munie d'une traction avant `FWD` ou voiture munie de roues arrière motrices `RWD`
- `transmission`: type de transmission, soit transmission variable continue (`CVT`), soit manuelle (`Manual`) soit automatique (`Auto`).
- `engine`: variable catégorielle indiquant le type de moteur en V, soit `V3`, soit `V4`, soit `V5`, soit `V6`, soit `V8` soit `V10`.
- `diesel`: variable binaire égale à 1 si la voiture est à moteur diesel.
- `turbo`: variable binaire égale à 1 si la voiture a un turbocompresseur.
- `used`: variable binaire égale à 1 si la voiture est usagée.
- `year`: année du modèle de la voiture.
- `limited`: variable binaire égale à 1 si le modèle de la voiture fait partie d'une série limitée.
- `color`: couleur de la peinture extérieure, soit autre (`other`), soit blanc (`white`), soit noir (`black`) ou soit rouge (`red`).

Discutez dans votre rapport des facteurs qui influencent le prix des voitures Ford. Votre rapport devra aussi répondre aux questions suivantes:

1. S'il y a lieu, quantifiez les effets nets sur le prix d'une voiture de son kilométrage ainsi que le fait qu'elle soit usagée ou neuve.
2. Est-ce que le rabais pour une voiture usagée est le même pour toute année d'usure additionnelle? Pour ce faire, il peut être utilde de restreindre votre analyse aux 10 ou 15 dernières années.

Compiled 19/02/2020 at 14:29

Projet 3: prix de logements à Ames, Iowa

Les données `ameshousing` sur le logement proviennent de

> Dean De Cock (2011). *Ames, Iowa: Alternative to the Boston Housing Data as an End of Semester Regression Project*, Journal of Statistics Education, **19** (3).

La version fournie pour ce projet a été légèrement modifiée. Le jeu de données contient les variables suivantes pour $n = 1460$ maisons à Ames, Iowa.

- `SalePrice`: prix de vente (en USD);
- `LotArea`: surface du lot (en pieds carrés);
- `Bedrooms`: nombre de chambres à coucher (excluant le sous-sol);
- `GarageArea`: surface du garage (en pieds carrés);
- `OverallCond`: score reflétant l'état global du logement et variant de plus qu'excellent (10) à très mauvais (1);
- `OverallQual`: score reflétant la qualité globale des matériaux utilisés ainsi que la finition du logement, et variant de plus qu'excellent (10) à très mauvais (1);
- `HouseType`: type de logement, soit individuel unifamilial (`1FmCon`), soit logement unifamilial converti en bi-familial (`2FmCon`), soit duplex (`Duplx`), soit maison de ville (`Twnhs`);
- `YearBuilt`: année de construction;
- `YearRemodAdd`: année de rénovation (même que année de construction si aucune modification ou rénovation n'a été effectuée);
- `YrSold`: année de vente du logement;
- `HalfBath`: nombre de salles de bain partielles;
- `FullBath`: nombre de salles de bain complètes;
- `Garage`: variable binaire égale à 1 si le logement possède un garage.

Discutez dans votre rapport des facteurs qui influencent le prix des logements à Ames. Votre rapport devra aussi traiter des questions suivantes:

1. Quantifiez si l'effet du nombre de chambres à coucher sur le prix du logement est linéaire ou pas.
2. En partant du scénario hypothétique d'une famille qui cherche à acheter une maison de ville à Ames avec trois chambres à coucher, discutez des économies qu'elle doit faire si sa banque requiert un accompte de 20%.

Compiled 19/02/2020 at 14:29