**EPICB325 - Statistiques pour Sciences humaines II**

**TP8 (Partie pratique d'un examen)**

*Cet "examen d'entrainement" est plus long qu'un vrai examen.*

***Conseils préliminaires***

* Lisez attentivement les questions et répondez dans ce fichier (directement après chaque question).
* **Sauvez très régulièrement votre fichier Word!**
* Quand vous avez répondu à toutes les questions, collez les codes R que vous avez utilisés à la fin de ce document, dans la section ***Appendice: code R***.
* Dans les exercices, il faut parfois ajouter une figure. Vous pouvez exporter un graphique à partir de Rstudio en cliquant sur *"Export…"* au-dessus du graphique, puis *"Copy to clipboard…".* Vous pouvez changer la largeur et/ou hauteur du graphique avant de cliquer sur *"Copy Plot"*. Ensuite vous le collez directement dans le fichier Word.
* A la fin, sauvegardez ce fichier comme pdf (*Fichier -> Enregistrer sous…*, puis dans *Format du fichier*, choisissez pdf). Pour le nom, suivez la convention *filière-nom-prénom*, où *filière* peut être *pol* ou *commu*. Par exemple, *commu-kiriliouk-anna.pdf.*
* Chargez le fichier pdf ainsi obtenu dans l'espace dédié sur webcampus (*Examen janvier 2022: remise).*

***Description des données***

Les données concernent les salaires annuels des professeurs universitaires aux États-Unis en 2008-2009. Ces données ont été collectées dans le but de discerner une potentielle différence entre les salaires des hommes et des femmes.

Le fichier salaires contient n = 397 observations des variables suivantes:

* rank : indique le grade d'un professeur; assistant (premier grade), associate (deuxième grade), full (troisième grade).
* discipline : indique la discipline d'un professeur; A (département théorique) ou B (département appliqué).
* yrs.since.phd: nombre d'années depuis l'obtention du doctorat.
* yrs.service: nombre d'années de travail en tant que professeur.
* sex: le sexe d'un professeur, female (femme) ou male (homme).
* salary: le salaire d'un professeur, en milliers de dollars par an.

Récupérez le fichier salaires.RDS sur webcampus, chargez-le, et vérifiez que tout est en ordre avec str avant de procéder:

* data <- readRDS("salaires.RDS")
* str(data)

'data.frame': 397 obs. of 6 variables:

$ rank : Factor w/ 3 levels "assistant","associate",..: 3 3 1 3 3 2 3 3 3 3 ...

$ discipline : Factor w/ 2 levels "A","B": 2 2 2 2 2 2 2 2 2 2 ...

$ yrs.since.phd: int 19 20 4 45 40 6 30 45 21 18 ...

$ yrs.service : int 18 16 3 39 41 6 23 45 20 18 ...

$ sex : Factor w/ 2 levels "female","male": 2 2 2 2 2 2 2 2 2 1 ...

$ salary : num 139.8 173.2 79.8 115 141.5 ...

***Questions***

Veuillez répondre directement en dessous de chaque question.

1. Quel est le salaire moyen des professeurs? Quel est le salaire moyen des professeurs féminins? Quelle est la proportion des professeurs ayant le grade associate ?
2. Est-ce que les hommes sont plus représentés dans les départements théoriques ou appliqués, comparé aux femmes? Si vous voulez, vous pouvez montrer un graphique qui illustre votre réponse.
3. Ajustez un modèle de régression linéaire (appelez-le m1) pour prédire le salaire d'un professeur en fonction du nombre d'années depuis l'obtention du doctorat. Interprétez le coefficient de détermination de ce modèle.
4. Écrivez l'équation de la droite de régression estimée à la question 3). À quel salaire un professeur pourrait s'attendre 10 ans après l'obtention de son doctorat?
5. Est-ce que l'hypothèse d'une variance constante des résidus est raisonnable pour ce modèle?
6. Ajustez un modèle de régression linéaire (appelez-le m2) pour prédire le salaire d'un professeur en fonction du nombre d'années depuis l'obtention du doctorat et en fonction du sexe. Faites un nuage de points de yrs.since.phd contre salary et ajoutez les deux droites de régression estimées, une pour les hommes, et une autre pour les femmes. N'oubliez pas d'ajouter une légende. Que voyez-vous?
7. Interprétez **soigneusement** les coefficients du modèle m2.
8. Donnez un intervalle de confiance pour le coefficient de régression de la variable yrs.since.phd et interprétez-le. Utilisez un niveau .
9. Considérons le test d'hypothèses contre , où est le coefficient de la variable sex. Comment peut-on interpréter cette hypothèse alternative? Donnez la p-valeur associée à ce test d'hypothèse et expliquez si la variable sexest significative au niveau
10. Ajustez un modèle de régression linéaire (appelez-le m3) pour prédire le salaire d'un professeur en fonction de yrs.since.phd, rank, discipline, et sex. À quel salaire un professeur féminin de grade associate d'un département appliqué pourrait s'attendre 10 ans après l'obtention de son doctorat?
11. Dans le modèle m3, arrive-t-on à la même conclusion concernant la significativité de la variable sexqu’avec le modèle m2 ? Pourquoi (pas)? Même question pour la variable yrs.since.phd? Avez-vous une idée pourquoi c'est le cas?
12. Supposons maintenant que l'on veuille utiliser la variable salary en tant que variable explicative. Pourrait-on prédire le rang d'un professeur en fonction de son salaire avec une régression logistique? Pourquoi (pas)?
13. Ajustez un modèle de régression logistique pour prédire le sexe d'un professeur en fonction de son salaire. Interprétez le coefficient associé . Que se passe-t-il quand on ajoute rank comme deuxième variable explicative?