

NOM/MATRICULE : _____

EXAMEN 1

Consignes :

L'examen 1 compte pour une valeur de 27.5% de la note finale. Vous avez de 8h30 à 10h20 pour compléter l'examen. Cet examen est divisé en trois parties : 1) partie "papier", à répondre directement sur l'examen, 2) partie SAS à réaliser sur ordinateur dans un programme SAS et 3) une partie R à réaliser dans sur ordinateur dans un programme R.

Si vous terminez l'examen avant la fin, vous pouvez m'en aviser en levant la main afin que vous puissiez copier vos programmes sur une clé USB. Sinon, vous aurez entre 10h20 et 10h30 pour me remettre vos programmes informatiques par courrier électronique à denis.talbot@mat.ulaval.ca.

CONSEILS :

Sauvegardez régulièrement vos programmes informatiques afin de ne pas les perdre!

Si vous ne parvenez pas à faire la lecture des données, continuez à écrire les programmes comme si la lecture avait fonctionné.

PARTIE 1 : Partie papier

Numéro 1 : 5 points

L'ONU pense à mettre en place un programme pour améliorer la propreté dans les villages d'une certaine région du Mali. Il existe 600 villages qui seraient admissibles au programme. De ces 600 villages, l'ONU en choisit aléatoirement 150 pour mener une expérience. La moitié des villages choisis participeront au programme et l'autre moitié n'y participera pas. L'objectif est de déterminer si le programme permet de diminuer la mortalité infantile dans les villages.

Identifiez :

- a) La population :

- b) L'individu ou l'unité statistique :

- c) L'unité expérimentale :

- d) L'unité d'échantillonnage :

- e) L'échantillon :

Numéro 2 : 5 points

Chacun des ménages des villages participant à l'étude menée par l'ONU doit remplir un questionnaire. Voici un extrait du questionnaire :

Combien d'individus forment le ménage : _____

Revenu annuel du ménage : Moins de 1000\$US
 1000\$US - 2000\$US
 2000\$US - 3000\$US
 Plus de 3000\$US

Votre ménage dispose-t-il d'une latrine : Oui
 Non

À quelle distance de votre foyer se situe l'établissement de soins de santé le plus près : _____ km

Quel est le plus haut niveau d'étude atteint par un membre de votre foyer :

- Études primaires commencées, mais non terminées
 - Études primaires complétées
 - Études secondaire commencées, mais non terminées
 - Études secondaire terminées
 - Études post-secondaires commencées, mais non terminées
 - Études post-secondaires terminées
-

De quels types sont ces variables?

a) Taille du ménage

b) Revenu annuel

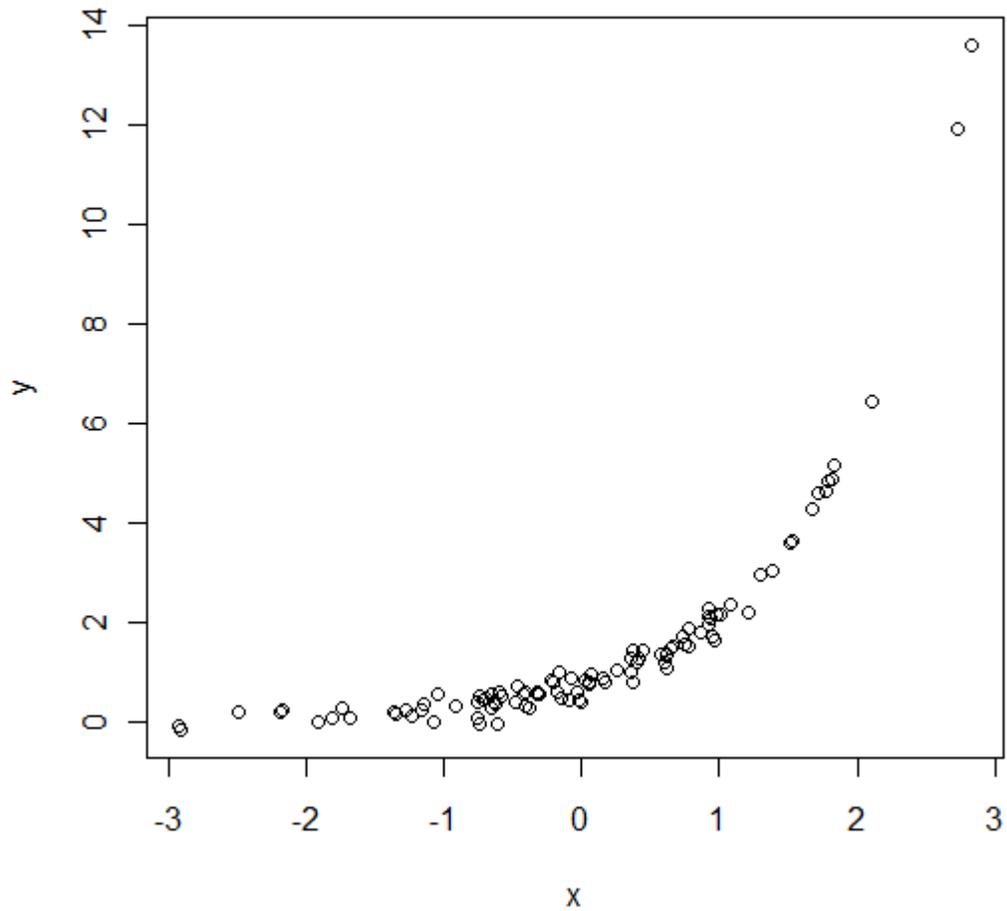
c) Présence d'une latrine

d) Distance par rapport à un établissement de santé

e) Niveau d'études

Numéro 3 : 10 points

Voici un graphique représentant les données pour deux variables continues.



Quelle statistique descriptive utiliseriez-vous pour mesurer l'association entre ces deux variables? Justifiez brièvement votre réponse.

Numéro 4 : 10 points

Dans le cadre de votre emploi pour la caisse Desjardins, vous avez calculé des statistiques descriptives concernant les profits mensuels réalisés, en dollars canadiens, par votre division :

Minimum	40 000\$
Moyenne	87 000\$
Q1	55 000\$
Médiane	72 000\$
Q3	107 000\$
Maximum	250 000\$
Écart-type	67 000\$
Déviatation absolue médiane (MAD)	53 000\$

Vous réalisez que vous avez oublié de retirer les frais fixes de vos calculs. Chacun des profits mensuels sont donc inférieurs de 12 000\$ à ce que vous aviez calculé initialement. Par exemple, si vous aviez entré un montant de 80 000\$, vous auriez dû entrer 68 000\$. Expliquez comment vous pouvez rapidement corriger les statistiques descriptives présentées dans le tableau. Vous n'avez pas à donner les valeurs corrigées.

PARTIE 2 : Partie SAS **35 points**

Le jeu de données suivant contient des informations sur les dates d'entrée en fonction de 2 employés :

```
Shion Uzuki 10/08/08  
Celes Cher 28/03/05
```

Nom	Description	Type	Colonnes
Nom	Nom de l'employé	Alphanumérique	1-11
Date_entree	Date d'entrée en fonction	Alphanumérique	13-20
Jour_entree	Jour d'entrée en fonction	Numérique	13-14
Mois_entree	Mois d'entrée en fonction	Numérique	16-17
Annee_entree	Année d'entrée en fonction	Numérique	19-20

a) Effectuez la lecture du jeu de données, en entrant les données tel qu'elles sont présentées et en créant le 5 variables énoncées dans le tableau.

Voici un second jeu de données portant sur les mêmes employés avec leur salaire annuel :

```
Celes Cher 37000  
Shion Uzuki 33000
```

b) Effectuez la lecture de ce second jeu de données en entrant les données tel qu'elles sont présentées. (Remarquez que l'ordre des noms est inversé!)

c) Effectuez la juxtaposition des deux jeux de données lus.

d) Exportez le jeu de données créé dans un fichier texte ou les valeurs sont séparées par un ;.

PARTIE 3 : Partie R 35 points

D'abord, faites rouler le code suivant afin d'avoir accès au jeu de données JohnsonJohnson sous format matrice en R :

```
jj.mat = matrix(JohnsonJohnson, ncol = 4, byrow = T);
```

Ce jeu de données contient les revenus trimestriels de l'entreprise Johnson & Johnson pour les années de 1960 à 1980.

- a) Renommez les 4 colonnes de cette matrice par Qtr1, Qtr2, Qtr3 et Qtr4 et renommez les lignes de cette matrice pour qu'ils correspondent aux années de 1960 à 1980
- b) Ajoutez une colonne à la matrice qui correspond au revenu total annuel, c'est-à-dire la somme des revenus pour les 4 trimestres.
- c) Calculez le revenu moyen par trimestre.
- d) À l'aide des résultats obtenus en b) et c), commentez brièvement ce qu'on peut observer dans les données.