

TP1 STATISTIQUE DESCRIPTIVE

Frédéric Lavancier

Avant propos

- Ouvrir l'application R
- Saisir dans la console **library(Rcmdr)** pour lancer R Commander

Des tutoriels sur l'utilisation de R Commander, réalisés par François Husson (AgroCampus Rennes), sont disponibles sous format vidéo à l'adresse :

<http://www.screencast.com/users/agrocampus/folders/FilmR>

On y trouve en particulier des tutoriels sur :

- Installation de packages dans R
- Importation de données dans R
- Calcul de statistiques descriptives
- Construction de graphiques avec R

1 Importer et modifier des données avec R Commander

1.1 Changer le répertoire de travail

Le répertoire de travail se choisit grâce au menu *Fichier* → *Changer le répertoire de travail*

Cela permet de se trouver directement dans ce répertoire lorsqu'on souhaite importer un jeu de données. De même, cela permet d'enregistrer directement dans ce répertoire les sorties produites par R Commander, les graphiques, etc.

1.2 Importer un jeu de données

Tous les jeux de données suivants sont disponibles à l'adresse

<http://www.math.sciences.univ-nantes.fr/~lavancie/enseignement.html>

1. Ouvrir le fichier **EmissionsGES.xls** avec un tableur. Ce jeu de données correspond aux émissions de gaz à effet de serre en 2003, pour plusieurs pays, ainsi que l'objectif pour 2010 suite aux accords de Kyoto.

Importer ce jeu de données sous R Commander avec le menu *Données* → *Importer des données*.

- Choisir un nom pour le jeu de données (le nom est interne à R, il ne modifie pas le nom du fichier).

- Sélectionner la feuille du fichier Excel que l'on souhaite importer
- Vérifier que le jeu de données s'est bien importé en le visualisant (bouton *Visualiser*).

Remarque : Toute modification dans R de ce jeu de données ne modifiera pas le fichier initial.

Quelles sont les individus et les variables dans ce jeu de données ? Donner le type de chacune des variables.

2. Ouvrir le fichier **NO2trafic.txt** avec un éditeur de texte ou un tableur (Notepad ou Open Office par exemple). Ce jeu de données contient la mesure de dioxyde d'azote (NO2) dans des véhicules lors du trajet domicile-travail en région parisienne, en fonction du type de route empruntée ("A" : Autoroute, "P" : Périurbain, "T" : Tunnel, "U" : Urbain, "V" : Voie rapide urbaine) et en fonction du trafic (de "1" : fluide à "4" : congestionné).

Comment se présente le jeu de données ? En particulier, comment sont séparées les variables ? Quelle est la marque des décimales (un point ou une virgule) ? A quoi correspond la première ligne ?

Importer ce jeu de données sous R Commander en lui donnant un nom différent du précédent. Visualiser le résultat. Quelles sont les individus et les variables dans ce jeu de données ? Donner le type de chacune des variables.

1.3 Modifier les données importées

On peut basculer d'un jeu de données à l'autre pour le rendre actif en le sélectionnant au niveau du bouton *Données*.

Différentes modifications du jeu de données actif sont possibles dans R Commander via le bouton *Editer* (pour modifier la valeur d'une cellule ou changer le nom ou le type d'une variable), ou via les menus *Données* → *Jeu de données actif* et *Données* → *Gérer les variables dans le jeu de données actif*.

1. Dans le jeu de données EmissionsGES, éliminer la ligne ne contenant que des valeurs manquantes. Vérifier le résultat en visualisant le nouveau jeu de données.

2. Résumer le jeu de données NO2Trafic grâce au menu *Statistiques* → *Résumés* → *Jeu de données actif*... Que penser des résumés des variables qualitatives ? Convertir ces variables en variables numériques si R les considère comme des variables numériques.

1.4 Sauvegarder un jeu de données

Il y a deux sauvegardes possibles du jeu de données

- soit au format R, ce qui crée un ".RData" : dans ce cas, seul R pourra relire le jeu de données. On enregistre sous ce format depuis le menu *Données* → *Jeu de données actif* → *Sauver le jeu de données actif*... A la prochaine ouverture de R, il suffira de lancer le jeu de données avec *Données* → *Charger un jeu de données*... sans passer par l'étape d'importation.
- soit au format ".txt", ce qui permet d'ouvrir le jeu de données avec d'autres logiciels (comme un tableur). On enregistre sous ce format depuis le menu *Données* → *Jeu de données actif* → *Exporter le jeu de données actif*... Pour rouvrir ce jeu de données sous R, il faudra l'importer.

Enregistrer au format ".RData" et ".txt" les trois jeux de données précédents modifiés. Visualiser leur nouvelle version dans un tableur.

2 Analyse univariée

2.1 Jeu de données EmissionsGES

1. Calculer la moyenne et la médiane de la variable `Emissions2003`. Combien de pays ont une émission en 2003 supérieure à la médiane ? supérieure à la moyenne ? Commenter.

2. Pour quantifier au mieux la variabilité des émissions en 2003 au sein des pays, vaut-il mieux calculer l'écart-type ou l'écart inter-quartiles ? Que valent-ils ?

3. Représenter la répartition des émissions en 2003 sous forme d'un histogramme, puis d'un boxplot. Créer un nouveau tableau sans les USA (utiliser le menu *Données* → *Jeu de données actif* → *Sous ensemble* ou *Données* → *Jeu de données actif* → *Éliminer des lignes* en enregistrant le nouveau tableau sous un autre nom) et refaire ces représentations graphiques sans les USA.

4. Calculer la moyenne et la médiane des émissions en 2003 sans les USA et comparer avec la première question.

2.2 Jeu de données NO2trafic

1. Quel est l'effectif total ?

Quelles sont les fréquences de chacune des modalités des variables qualitatives ?

Représenter la répartition des variables qualitatives à l'aide d'un graphe en barres, puis d'un camembert. Pour la variable "type", réordonner les modalités dans l'ordre décroissant de leur fréquence, puis refaire les représentations graphiques.

Complément : pour représenter un nouveau graphique dans une nouvelle fenêtre, sans écraser le graphique précédent, il faut saisir `windows()` dans la fenêtre de commande de R avant de lancer le graphique. Cela ouvre une nouvelle fenêtre graphique vierge. Sous linux, la commande équivalente est `x11()` et sous mac `quartz()`.

2. Représenter la répartition de la variable `NO2` sous forme d'histogramme, puis sous forme de boxplot. Calculer la moyenne \bar{x}_n , l'écart-type S , le minimum, le maximum et les quartiles de cette variable. Que vaut l'intervalle $\mathcal{I} = [\bar{x}_n - 2S; \bar{x}_n + 2S]$?

Compléter la phrase suivante : au moins ...% des observations se trouvent dans l'intervalle \mathcal{I} .

Compléter la phrase suivante : à peu près ...% des observations se trouvent entre le premier et le troisième quartile.

Compléter la phrase suivante : moins de 5% des observations ont leurs valeurs strictement supérieures au quantile $Q(p)$ où $p = \dots$. Que vaut $Q(p)$ pour la variable `NO2` ?

2.3 Vers les lignes de commandes

Ecrire un script de lignes de commandes qui permet d'effectuer les analyses précédentes. L'intérêt est de n'avoir pas à recommencer les mêmes manipulations avec R Commander lors de la prochaine étude d'un jeu de données, mais uniquement de lancer le script correspondant.

Pour ce faire, aller dans le menu R : *Fichier* → *Nouveau script*. Les lignes de commandes se lancent dans la "R Console" en les copiant directement dans la console depuis le script, ou en appuyant sur "CTRL+R" au niveau de la ligne de commande dans le script.

On pourra notamment tester différentes options graphiques comme les couleurs, les légendes, les titres, etc, ce qui n'est pas possible avec R Commander.

On pourra s'inspirer des commandes apparaissant dans la "fenêtre de script" de R Commander. Quelques fonctions utiles sont données ci-dessous.

Quelques fonctions utiles : supposons que le tableau de données se nomme "Dataset".

Accéder aux données du tableau :

`Dataset` affiche le jeu de données complet

`Dataset$var` affiche la variable "var" du jeu de donnée

`Dataset[10,4]` permet d'accéder au 10ème élément de la 4ème colonne de "Dataset"

`Dataset[,4]` permet d'accéder à toutes les valeurs de la 4ème colonne de "Dataset". Equivalent à `Dataset$var` si "var" est le nom de la 4ème colonne.

`Dataset[,-4]` renvoie le tableau sans la 4ème colonne et nomme ce nouveau tableau "NewDataset"

`NewDataset=Dataset[,-4]` permet de nommer le tableau sans la 4ème colonne "NewDataset"

`Dataset$var[10]` permet d'accéder à l'élément 10 de "var"

`Dataset$var[1:10]` permet d'accéder aux éléments de 1 à 10 de "var"

`Dataset$var[-10]` renvoie toutes les valeurs de "var" sauf la 10ème.

Obtenir des résumés numériques d'une variable du jeu de données :

`mean(Dataset$var)` donne la moyenne de la variable "var"

`median(Dataset$var)` donne la médiane de la variable "var"

`var(Dataset$var)` calcule la variance de la variable "var"

`sd(Dataset$var)` calcule l'écart-type de la variable "var"

`quantile(Dataset$var, 0.9)` calcule le quantile d'ordre 0.9 de la variable "var"

`quantile(Dataset$var, c(0.25,0.75,1))` calcule les quantiles d'ordre 0.25, 0.75 et 1 (le maximum) de la variable "var"

`order(Dataset$var)` donne la position de la plus petite valeur, puis de la seconde plus petite valeur, etc, jusqu'à la position de la plus grande valeur de "var".

`sort(Dataset$var)` classe les valeurs de la variable "var" de la plus petite à la plus grande.

Faire quelques représentations graphiques :

`hist(Dataset$var, col="red")` représente l'histogramme de la variable "var" en rouge

`boxplot(Dataset$var, col="blue")` représente la boîte de dispersion de la variable "var" en bleu

`table(Dataset$var)` donne le tableau d'effectifs de "var" si celle-ci est un facteur

`barplot(table(Dataset$var), legend=TRUE, col="cyan")` représente le diagramme en barres des effectifs de "var" (si celle-ci est un facteur) avec une légende et en cyan.

Obtenir de l'aide sur une fonction et ses options possibles :

`help(fun)` ou `?fun` ouvre l'aide de la fonction "fun". On y trouve toutes les options possibles de la fonction et des exemples.

TP2 STATISTIQUE DESCRIPTIVE

Frédéric Lavancier

L'objectif de ce TP est d'analyser les données du fichier **Desbois.csv**.

Ces données sont issues de l'article de Dominique Desbois "Introduction to Scoring Methods : Financial Problems of Farm Holdings", disponible sur internet. Il contient différentes informations sur des exploitations agricoles en France observées entre 1988 et 1994 :

- CNTY : département de l'exploitation
- DIFF : défaut de paiement de la part de l'exploitant (1 : non, 2 : oui)
- Status : statut légal (1 : individuel, 0 : entreprise)
- Hectare : surface de l'exploitation
- ToF : type d'exploitation (de 1 à 6 : élevage, céréales, mix, etc.)
- OWNLAND : l'exploitant est propriétaire (1) ou non (0)
- AGE : âge de l'exploitant
- Harvest : année d'observation de l'exploitation
- des variables micro-économiques :
 - *Capitalisation* : r1 (total debt / total assets); r2 (stockholders' equity / invested capital); r3 (short term debt / total debt); r4 (short term debt / total assets); r5 (long and medium term debt / total assets);
 - *Poids de la dette* : r6 (total debt / gross product); r7 (long and medium term debt / gross product); r8 (short term debt / gross product);
 - *Liquidité* : r11 (working capital / gross product); r12 (working capital / (real inputs - financial expenses)); r14 (short term debt / circulating assets);
 - *Service de la dette* (avec EBITDA = Earnings Before Interest, Taxes, Depreciation and Amortization = "résultat net") : r17 (financial expenses / total debt); r18 (financial expenses / gross product); r19 (financial expenses + refunding of long and medium term capital) / gross product; r21 (financial expenses / EBITDA); r22 (financial expenses + refunding of long and medium term capital)/EBITDA;
 - *Rentabilité des capitaux* : r24 (EBITDA /total assets);
 - *Revenus* : r28 (EBITDA / gross product); r30 (available income / gross product); r32 (EBITDA - financial expenses) / gross product;
 - *Activité de production* : r36 (immobilized assets / gross product); r37 (gross product / total assets).

Importation du jeu de données.

1. Importer ce jeu de données dans R Commander. Visualiser le résultat.
2. Quelles sont les individus et les variables dans ce jeu de données? Donner le type de chacune des

variables. Vérifier que R interprète bien chaque variable qualitative en facteur, corriger si besoin.

Analyse des exploitations étudiées

3. Quels départements français ont été étudiés dans ce jeu de données ?

4. Quel est le nombre d'exploitants étudiés dans l'échantillon ? Quel pourcentage sont en défaut de paiement ? Quel pourcentage sont propriétaires de leur exploitation ?

5. Représenter la répartition de l'âge des exploitants. Quel est l'âge moyen ? l'âge médian ? l'âge minimum ? l'âge maximum ? Donner l'intervalle interquartile. Quelle proportion approximative d'exploitants ont leur âge dans cet intervalle ?

6. Représenter la répartition des surfaces d'exploitation. Quelle est la surface moyenne, la surface maximale et la surface correspondant au quantile à 90% ? A quels département appartiennent les 10 exploitations les plus importantes ? Quelle est la taille minimale des ces 10 exploitations ? Sont-elles en défaut de paiement ?

7. Quel type d'exploitation est la plus présente dans l'échantillon ?

Quelques analyses croisées

8. Représenter la répartition des surfaces par département.

9. Dans quel département les surfaces sont les plus importantes en moyenne ? Que vaut la surface moyenne dans ce département ? Dans quel département la dispersion des surfaces est-elle la plus petite ? la plus grande ? Quantifier ces dispersions pour chaque département.

10. Représenter la répartition des surfaces d'exploitation selon le statut de l'exploitation (variable Status). Quelle est la surface moyenne dans chaque cas ? Rappeler le nombre d'exploitations dans chaque statut.

11. Construire le tableau croisé de la variable CNTY et de la variable ToF. Trouve-t-on le même type d'exploitation dans chaque département ? Justifier. Représenter la distribution conditionnelle de ToF en fonction du département.

12. Les propriétaires (variable OWNLAND) ont-ils plutôt de grandes exploitations ? Sont-ils plutôt plus âgés que les autres ? Pour quels types d'exploitation y-a-t-il une sous-représentation d'exploitants propriétaires ? Justifier les affirmations.

Analyse des exploitants en défaut de paiement

13. Représenter la répartition de la taille des exploitations en fonction de la variable DIFF.

14. Construire le tableau croisé de la variable DIFF et de la variable CNTY. En déduire la représentation graphique de la répartition conditionnelle de la variable DIFF en fonction du département. Y-a-t-il des départements où les exploitants en défaut de paiement sont plus fréquents dans l'échantillon ?

15. De la même manière, étudier le lien éventuel, graphique et numérique, entre la variable DIFF et les autres variables du jeu de données, hormis les variables micro-économiques de r1 à r37. Finalement, quel profil d'exploitations semble avoir le plus de chances d'être en défaut de paiement ?

16. Etudier le lien entre la variable DIFF et les variables micro-économiques. Selon vous, à l'aide de quels indices micro-économiques du jeu de données pourrait-on anticiper le plus efficacement un défaut de paiement ?

TP3 STATISTIQUE DESCRIPTIVE

Frédéric Lavancier

1 Emissions de Gaz à Effet de Serre

Reprendre le jeu de données EmissionsGES (importé depuis le fichier **EmissionsGES.xlsx**).

1. Calculer la corrélation et représenter le nuage de points entre la variable Emissions2003 et la variable Population.
2. Ajuster la droite des moindres carrés à ce nuage de points. Quelle est son équation? Que vaut le coefficient de détermination R^2 ? Superposer la droite des moindres carrés au nuage de points.
3. Refaire les deux questions précédents sans les USA.
4. On souhaite prédire les émissions de gaz à effet de serre en 2003 pour un pays de 50 millions d'habitants. Quelle prévision proposer?

2 Ozone

Importer le fichier **Ozone.txt** dans R Commander. Il contient la concentration d'ozone maximale, la température maximale et la vitesse du vent, mesurées chaque jour pendant 3 mois d'été à Rennes.

1. Résumer les variables du jeu de données (moyenne, écart-type, etc.)
2. Représenter la boîte de dispersion de la variable Ozone.
3. Représenter le nuage de points de la variable Ozone en fonction de la Température. Un lien semble-t-il présent?
4. Faire de même entre la variable Ozone et la vitesse du vent, puis entre la Température et la vitesse du vent.
5. Calculer les corrélations entre toutes les variables du jeu de données. Les valeurs sont-elles en accord avec les nuages de points?
6. Estimer la ligne des moindres des carrés expliquant l'ozone en fonction de la température. Que vaut le coefficient de détermination R^2 ?
7. Estimer la ligne des moindres des carrés expliquant l'ozone en fonction de la vitesse du vent. Que vaut le coefficient de détermination R^2 ?
8. Représenter le nuage de points 3D de l'ozone en fonction de la température et de la vitesse du vent. Estimer le modèle de régression linéaire expliquant l'ozone en fonction des 2 variables Température et Vent. Que vaut le R^2 ?
9. Les prévisions météo pour demain indiquent une température de 30 degrés et un vent quasi nul. Quelle concentration d'ozone peut-on attendre?

TP4 STATISTIQUE DESCRIPTIVE

Frédéric Lavancier

1 Représentation de séries temporelles

1. Enregistrer le fichier **Beer.txt**, donnant la production mensuelle de bière en Australie de 1956 à 1995, et indiquer à R le répertoire de travail contenant le fichier (dans *Fichier* → *Changer le répertoire courant...*).

2. Recopier le code ci-dessous et l'exécuter avant chaque ligne **####**. Expliquer le rôle de chaque commande.

```
beerdata=read.table('Beer.txt')
names(beerdata)
plot(beerdata$V1)
#####
plot(beerdata$V1,type='l')
#####
beer=ts(beerdata$V1,start=1956,freq=12)
#####
plot(beer)
#####
acf(beer,36)
#####
```

3. En vous inspirant du code précédent, importer le jeu de données **GDP.txt** résumant le PIB annuel aux Etats-Unis depuis 1929.

4. Définir la série temporelle d'intérêt au format "ts" (time series) sous R avec la fonction **ts**, en précisant la date de départ (option **start**). La représenter et décrire ses caractéristiques principales (tendance, saisonnalité, décomposition additive ou multiplicative).

5. Répéter les étapes précédentes pour les jeux de données suivants, en précisant la période de la série avec l'option **freq** dans la fonction **ts** (si la série présente une saisonnalité) :

- **TempDub.txt** : température mensuelles à Dubuque (Iowa) de 1964 à 1975,
- **CO2Hawaï.txt** : niveau mensuel de CO2 à Alert, Canada, de 1994 à 2004,
- **TraficAerien.txt** : trafic aérien mensuel international de 1949 à 1960,
- **GoogleReturn.txt** : rendement journalier de l'action Google du 20/08/2004 au 13/09/2006.

4. Rappeler la définition des auto-corrélations d'une série temporelle (ACF). Représenter les ACF des séries précédentes jusqu'au "lag" 30 à l'aide de la fonction `acf` et commenter les représentations.

5. On s'intéresse de plus près à la série `GoogleReturn`. Que vaut la corrélation entre le rendement un jour donné et le rendement de la veille? et de l'avant-veille? La série s'arrête le 13 septembre 2006, peut-on espérer l'exploiter pour prédire le rendement de l'action pour le 14 septembre 2006?

2 Estimation d'une tendance

1. Recopier le code ci-dessous et l'exécuter avant chaque ligne `####`. Expliquer le rôle de chaque commande.

```
temps=as.numeric(time(beer))
reg=lm(beer ~ temps)
reg
#####
plot(beer)
lines(temps,reg$fitted.values,col='red')
#####
beer5=filter(beer,c(1/5,1/5,1/5,1/5,1/5))
plot(beer)
lines(beer5,col='red')
#####
beer12=filter(beer,c(1/24,rep(1/12,11),1/24))
lines(beer12,col='green')
#####
```

2. On souhaite estimer la tendance de la série `GDP` de façon paramétrique

- Une tendance linéaire est-elle apparente?
- Estimer une tendance quadratique par moindres carrés en s'inspirant du code ci-dessus ou en se référant au code du cours.
- On souhaite estimer une tendance exponentielle. Définir une nouvelle série valant le logarithme de la série initiale. Si la tendance de `GDP` est exponentielle, quelle forme doit avoir la tendance de cette nouvelle série? Estimer cette tendance sur la nouvelle série. En déduire une estimation de la tendance exponentielle de la série `GDP`.
- Superposer à la représentation de la série `GDP` sa tendance quadratique et sa tendance exponentielle. Laquelle semble la plus adaptée?

3. On s'intéresse à présent à la série `CO2Hawai`.

- Proposer une forme paramétrique pour la tendance, l'estimer et la superposer à la série.
- Calculer la moyenne mobile centrée d'ordre 6 de la série et la superposer à la série.
Indication : `filter(x, c(1/12,1/6,1/6,1/6,1/6,1/6,1/12))` calcule la moyenne mobile centrée de x associée aux coefficients $(1/12, 1/6, 1/6, 1/6, 1/6, 1/6, 1/12)$. Pour aller plus vite, on peut utiliser `rep(1/6,5)` qui est le vecteur contenant 5 fois de suite $1/6$.
- Refaire la même chose à l'ordre 12, 15 et 24. Quel ordre de lissage retenir?
- Superposer la tendance paramétrique et la tendance par moyenne mobile retenue.

3 Estimation d'une saisonnalité

1. On suppose disposer de la série `tempdub`, issue du fichier `TempDub.txt`. Recopier le code ci-dessous et l'exécuter avant chaque ligne `####`. Expliquer le rôle de chaque commande.

```
plot(tempdub)
#####
profil=aggregate(tempdub~cycle(tempdub),FUN=mean)
profil
profil$tempdub
plot(profil$tempdub)
#####
saison=rep(profil$tempdub,length.out=length(tempdub))
temps=as.numeric(time(tempdub))
plot(temps,saison,type='l')
#####
plot(tempdub)
lines(temps,saison,col='red')
#####
```

2. On souhaite estimer le profil saisonnier de la série `CO2Hawai`.

- Retrancher à la série `CO2Hawai` une estimation de sa tendance (effectuée à la partie précédente). On note \tilde{x} cette nouvelle série. Représenter la pour vérifier qu'elle est bien sans tendance apparente.
- Estimer le profil saisonnier de \tilde{x} en s'inspirant du code ci-dessus. Représenter ce profil et l'analyser.
- Construire la série saisonnière en prolongeant le profil saisonnier par périodicité jusqu'à atteindre la taille initiale de `CO2Hawai`.
- Constuire la série ajustée, valant la somme de la tendance estimée et de la série saisonnière. Superposer cette série ajustée à une représentation de la série `CO2Hawai`.

4 Décomposition complète d'une série

1. Recopier le code ci-dessous et l'exécuter avant chaque ligne `####`. Expliquer le rôle de chaque commande.

```
res=decompose(beer)
plot(res)
#####
res$trend
res$seasonal
res$figure
res$random
#####
plot(beer)
lines(res$trend+res$seasonal,col='red')
#####
plot(res$trend+res$random,col='red')
#####
```

2. On s'intéresse à la série CO2Hawaï.

- Appliquer une décomposition additive comme ci-dessus à la série CO2Hawaï et représenter le résultat.
- Comparer le profil saisonnier estimé par la décomposition précédente avec le profil saisonnier estimé dans la partie précédente.
- Comparer la série ajustée obtenue grâce à la décomposition précédente avec la série ajustée construite dans la partie précédente.
- Construire puis représenter la série CO2Hawaï corrigée des variations saisonnières.
- Utiliser la série CVS précédente pour effectuer une estimation paramétrique de la tendance.
- Utiliser la tendance précédente pour construire une série ajustée et la prolonger sur 12 valeurs pour en déduire une prévision de la concentration de CO2 à Hawaï l'année suivante.

3. On désire effectuer une décomposition de la série **TraficAerien.txt**, notée x ci-dessous. Executer les commandes ci-dessous, expliquer leur rôle et analyser le résultat. Quelle décomposition finale retenir ?

```
plot(x)
#####
res=decompose(x)
plot(res)
#####
plot(x)
lines(res$trend+res$seasonal, col='red')
#####
res2=decompose(x, 'multiplicative')
plot(res2)
#####
plot(x)
lines(res2$trend*res2$seasonal, col='red')
#####
```

5 Application à la série SNCF

1. Importer le fichier **snf.txt** contenant le trafic voyageur mensuel en france de 1985 à 1997 inclus. Définir la série au format "ts" sous R en indiquant la date de départ et la période. Représenter la série.

2. Quelle type de décomposition de la tendance et de la saisonnalité semble-t-elle la plus judicieuse, additive ou multiplicative ? Estimer la décomposition retenue. Représenter la tendance estimée et le profil saisonnier typique.

3. En déduire une série ajustée. La superposer à la série initiale à titre de comparaison.

4. Représenter la série scnf corrigée des variations saisonnières. Utiliser cette dernière pour estimer une tendance linéaire. Quelle est l'équation de la droite estimée (bien préciser les unités utilisées) ?

5. Construire la série ajustée basée sur la tendance linéaire estimée ci-dessus, et en déduire une prévision du trafic snf pour l'année 1998.