

École Supérieure de la Statistique et de l'Analyse de l'Information	Classe : 2ème Année
Année Universitaire : 2023-2024	Date : 15.05.2024
Examen de Techniques de prévision	Durée : 1h 30

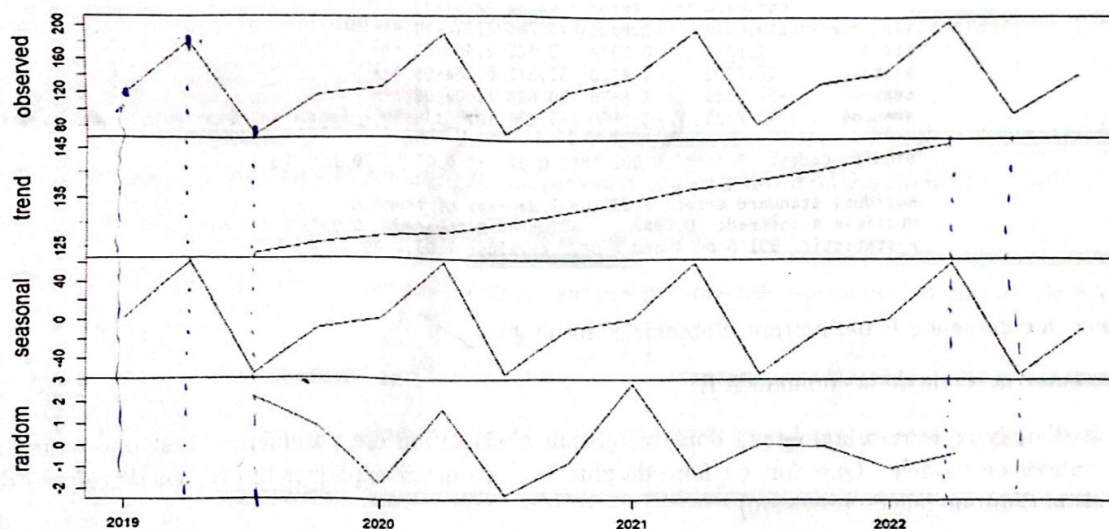
Exercice 1

Le tableau suivant indique l'évolution du chiffre d'affaires trimestriel (en DT) d'un magasin de 2019 à 2022.

tr_1^{19}	tr_2^{19}	tr_3^{19}	tr_4^{19}	tr_1^{20}	tr_2^{20}	tr_3^{20}	tr_4^{20}	tr_1^{21}	tr_2^{21}	tr_3^{21}	tr_4^{21}	tr_1^{22}	tr_2^{22}	tr_3^{22}	tr_4^{22}
120	181	71	119	128	190	73	124	140	196	84	133	145	206	96	142

Dans la suite on a décomposé la série en utilisant la commande `decompose` de R. Le résultat de la décomposition est présenté dans la figure suivante :

Decomposition of additive time series



1. A travers la représentation graphique de cette série (première figure en haut), pourquoi un modèle de décomposition additif semble-t-il le plus approprié?
2. Donner l'expression de la moyenne mobile utilisée pour extraire la partie saisonnière.
3. Retrouver les valeurs des composantes représentées dans les trois graphiques en bas. (Indication : vous devez calculer la série filtrée, séries sans tendance, coefficients saisonniers puis la série corrigée des variations saisonnières).
4. En examinant l'allure de la série (figure en haut), dites pourquoi une prévision basée sur un modèle de régression semble une technique appropriée ?

I-Prévision basée sur un modèle de régression : Dans la suite on suppose que la série contient une tendance linéaire et une composante saisonnière trimestrielle.

On décide d'estimer un modèle de régression de la forme $y = X\beta + \epsilon$. Dans ce modèle la composante saisonnière sera modélisée à travers les variables muettes.

Notons X la matrice du modèle, β le vecteur des coefficients et Y le vecteur des observations.

5. Rappeler l'expression de l'écriture matricielle de ce modèle de régression.
6. Identifier alors le vecteur β ainsi que les la matrice X et le vecteur Y .
7. Donner l'expression de l'estimateur obtenu par MCO de β .

Dans la suite, on décompose l'historique en deux périodes :

- d'apprentissage : 2019, 2020 et 2021.
- de test : 2022.

On vous donne les résultats de l'estimation du modèle sur la période d'apprentissage :

```
Call:
tslm(formula = ca_train ~ trend + season)

Residuals:
    Min       1Q   Median       3Q      Max
-3.0000 -1.3333 -0.1667  1.1042  2.9167

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 119.6458    1.6219   73.768 2.21e-11 ***
trend         1.9375     0.1972    9.826 2.40e-05 ***
season2       57.7292     1.8320   31.511 8.38e-09 ***
season3      -57.2083     1.8636  -30.698 1.00e-08 ***
season4      -9.8125     1.9150   -5.124 0.00136 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.231 on 7 degrees of freedom
Multiple R-squared:  0.9982,    Adjusted R-squared:  0.9972
F-statistic: 991.6 on 4 and 7 DF,  p-value: 1.031e-09
```

8. Donner la commande R permettant d'obtenir ce résultat.
9. Commenter la sortie de la commande R.
10. Le fait d'analyser le résultat obtenu dans ce tableau d'estimation des paramètres n'est plus suffisant pour valider ce modèle. Que doit on faire de plus ? (vous devez expliquer brièvement les étapes de validation d'un modèle de prévision).
11. Utiliser le modèle estimé pour construire des prévisions du nombre de ventes pour la période du test.
12. Calculer les deux indicateurs MAPE et RMSE sur les prévisions obtenues sur la période de la prévision test (année 2022). Commenter.
13. On souhaite prédire la série pour l'année 2023. On vous donne maintenant le résultat de l'estimation du modèle sur tout l'historique. Faites des prévisions pour l'année 2023.
14. Donner la commande R permettant d'obtenir des prévisions pour l'année 2023.

II-Prévision par lissage exponentielle :

15. Calculer la série lissée sur la période d'apprentissage de 2019 à 2021 en utilisant un lissage exponentiel simple avec $\lambda = 0.3$ et [pour le point initial vous prenez y_1^{LES} la moyenne sur les 5 premières observations de la série].

```

call:
tslm(formula = ca_ts ~ trend + season)

Residuals:
    Min       1Q   Median       3Q      Max
-3.850 -1.113 -0.300  1.163  2.600

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 118.7250    1.3082  90.753 < 2e-16 ***
trend         2.0750     0.1152  18.015 1.64e-09 ***
season2       57.9250     1.4615  39.634 3.20e-13 ***
season3      -56.4000     1.4751 -38.236 4.74e-13 ***
season4      -9.9750     1.4974  -6.662 3.55e-05 ***
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.06 on 11 degrees of freedom
Multiple R-squared:  0.9983,    Adjusted R-squared:  0.9976
F-statistic: 1575 on 4 and 11 DF, p-value: 4.366e-15

```

16. En déduire alors des prévisions tests (avec $\lambda = 0.3$) de la série pour l'année 2022.
17. Recalculer les deux indicateurs proposés RMSE et MAPE sur la base des prévisions tests de l'année 2022 obtenues par lissage exponentiel simple (avec $\lambda = 0.3$).
18. Comparer le pouvoir prédictif de ces deux modèles. Interpréter.
19. Proposer une approche permettant d'améliorer la qualité de la prévision obtenue par le lissage exponentiel.
20. Donner la commande pour une bonne utilisation du lissage exponentiel dans ce contexte.

III-Prévision par les méthodes SARIMA et TBATS :

21. Rappeler les étapes de Box et Jenkins pour la prévision à partir des modèles SARIMA.
22. Peut-on utiliser les modèles SARIMA pour prédire cette série?
23. Donner une commande R permettant la recherche des paramètres "optimaux" d'un modèle SARIMA pour cette série.
24. Rappeler les étapes de la méthode TBATS.
25. La méthode TBATS peut-elle être utilisée dans ce contexte ? Si oui préciser les paramètres de cette méthode.