EXAMEN FINAL, ACT6420, HIVER 2012

ARTHUR CHARPENTIER

Les calculatrices sont autorisées. Tous les documents sont, en revanche, interdits.

Dans les feuilles qui suivent, il y a

- 10 questions générales sur les modèles de régression sur données individuelles
- 10 questions portant sur la modélisation des matchs de basket-ball (à partir des sorties disponibles dans l'Annexe A, 9 pages)
- 10 questions générales sur les modèles de séries temporelles
- 10 questions portant sur la modélisation et la prévision (à partir des sorties disponibles dans l'Annexe B, 14 pages)

Pour chaque question, quatre réponses sont proposées, une seule est valide, et vous ne devez en retenir qu'une (au maximum), et reportez votre réponse sur la feuille jointe. Choisir l'affirmation la "*moins juste*" signifie que trois affirmations sont valides et qu'une est fausse. Il faut identifier la fausse. Choisir l'affirmation la "*plus juste*" signifie que trois affirmations sont fausses et qu'une est valide. Il faut identifier cette dernière.

Pour le décompte des points,

- vous gagnez 2.5 points par bonne réponse
- vous perdez ne perdez aucun point par mauvaise réponse

Aucune justification n'est demandée. Votre note finale est le total des points.

Formulaire: si Z suit une loi $\mathcal{N}(0,1)$, $\mathbb{P}(Z>0.25)=40\%$, $\mathbb{P}(Z>0.52)=30\%$, $\mathbb{P}(Z>0.84)=20\%$, $\mathbb{P}(Z>1.28)=10\%$, $\mathbb{P}(Z>1.64)=5\%$ et $\mathbb{P}(Z>1.96)=2.5\%$.

Vous avez 3 heures.

1. Régressions sur données individuelles

Considérons un modèle de régression, de la forme

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \varepsilon_i \tag{1.1}$$

où Y est le poids d'un individu, en kilogrammes, et X sa taille, en mètres. Ce modèle sera appelé (1.1) dans la suite.

Question 1. Le modèle de régression contient un terme d'erreur ε_i pour plusieurs raisons. Parmi les affirmations suivantes, une seule n'est **pas valide**. Laquelle ?

A. parce qu'un partie du comportement de Y n'a pas pu être capturé par X_1,
B. parce que le modèle linéaire n'est (généralement) qu'une approximation de la vraie relation liant Y à la variable X_1,
C. parce que l'estimateur par moindre carré ne donne qu'une approximation numérique de l'estimateur du maximum de vraisemblance, que l'on corrige à l'aide du terme d'erreur,
D. parce qu'on souhaite tenir compte d'une éventuelle marge d'erreur dans l'observation de la variable Y.

Pour estimer les coefficients β_0 et β_1, l'approche naturelle est de considérer l'estimateur des moindres carrés ordinaires.

Question 2. Mais auparavant, il convient de faire des hypothèses afin d'avoir construit ainsi les meilleurs estimateurs linéaires non biaisés de β_0 et β_1. Laquelle de ces hypothèses sur les résidus n'est *pas nécessaire* (et ne fait pas partie des *hypothèses de base* telles qu'énoncées dans le cours)

A. les résidus doivent être centrés,
B. les résidus sont de variance identique,
C. les résidus doivent suivre une loi normale,
D. les résidus doivent être non corrélés, et non corrélés avec X_1.

Une fois cette hypothèse faite, on calcule par la méthode des moindres carrés $\widehat{\beta}_0$ et $\widehat{\beta}_1$, à partir de n observations. On peut alors construire la séries des erreurs, $\widehat{\varepsilon}_i = Y_i - \widehat{Y}_i$ où $\widehat{Y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 X_{1,i}$,

Question 3. Laquelle des affirmations suivante n'est **pas vraie**

A. la droite de régression $y = \widehat{\beta}_0 + \widehat{\beta}_1 x$ passe forcément par le point $(\overline{X}_1, \overline{Y})$ où, classiquement, $\overline{X}_1 = \frac{1}{n}\sum_{i=1}^n X_{1,i}$ et $\overline{Y} = \frac{1}{n}\sum_{i=1}^n Y_i$,
B. la somme des résidus est nulle $\sum_{i=1}^n \widehat{\varepsilon}_i = 0$,
C. la somme des valeurs observées des Y_i et la somme des valeurs prédites $\widehat{Y}_i$ doivent être égales, $\sum_{i=1}^n \widehat{Y}_i = \sum_{i=1}^n Y_i$,
D. il y a autant de points au dessus que de points en dessous de la droite de régression.

Question 4. Le R^2 (appelé coefficient d'ajustement) mesure une quantité parmi les quatre suivantes. Laquelle ?

A. la corrélation (empirique) entre les $X_{1,i}$ et Y_i
B. la covariance (empirique) entre $X_{1,i}$ et Y_i
C. le ratio de la somme des $\widehat{\varepsilon}_i^2$ sur la somme des Y_i^2
D. le ratio de la somme des $\widehat{Y}_i^2$ sur la somme des Y_i^2

Supposons que l'on change d'unité pour la mesure de Y, en donnant désormais le poids en livres (pour rappels, une livre vaut 0,4535 kilogramme). On notera $\widetilde{Y}$ la nouvelle variable. On a désormais un modèle $\widetilde{Y}_i = \widetilde{\beta}_0 + \widetilde{\beta}_1 X_{1,i} + \widetilde{\varepsilon}_i$.

Question 5. Tous les 'coefficients" vont changer. Sauf un, lequel ?

A. la constante de la régression, i.e. $\widetilde{\beta}_0 = \beta_0$
B. la pente de la régression, i.e. $\widetilde{\beta}_1 = \beta_1$
C. le R^2, coefficient d'ajustement $\widetilde{R}^2 = R^2$
D. la variance des résidus, $\widetilde{\sigma}^2 = \sigma^2$

À partir d'un jeu de données, on a estimé le modèle suivant $Y = -100 + 100X_1 + \varepsilon$, avec $\text{Var}(\varepsilon) = 25$.

Question 6. Une de ces affirmations n'est pas correcte,

A. un individu qui pèse 70 kg pour 180 cm est léger, relativement à sa taille,
B. un individu qui pèse 75 kg pour 170 cm est lourd, relativement à sa taille,
C. aucune personne de moins d'un mètre ne faisait partie de l'échantillon
D. en moyenne, un individu de 2 mètres pèse 100 kg

Mais avant d'en tirer des conclusions, on nous suggère de faire un test de Student.

Question 7. Le test de Student peut être utilisé (une seule affirmation est **correcte**)

A. pour tester si la constante β_0 et la pente β_1 sont de signe contraire
B. pour tester si la constante β_0 et la pente β_1 sont opposés l'un de l'autre
C. pour tester si la constante β_0 est nulle ou pas,
D. pour tester si la variance des résidus σ^2 est nulle ou pas

Au lieu de changer l'unité de Y, on va décaller les données. On va noter désormais $\widetilde{X}_1$ la taille au delà d'un mètre, i.e. $\tilde{X}_1 = X_1 - 1$. On a désormais un modèle

$$Y_i = \widetilde{\beta}_0 + \widetilde{\beta}_1 \widetilde{X}_{1,i} + \widetilde{\varepsilon}_i.$$

Question 8. Tous les "coefficients" seront identiques, sauf un. Lequel ?

A. la constante de la régression, i.e. $\widetilde{\beta}_0 \neq \beta_0$
B. la pente de la régression, i.e. $\widetilde{\beta}_1 \neq \beta_1$
C. le R^2, coefficient d'ajustement $\widetilde{R}^2 \neq R^2$
D. la variance des résidus, $\widetilde{\sigma}^2 \neq \sigma^2$

En fait, le modèle a été estimé sur 59 personnes On a ainsi obtenu

$$\widehat{\beta}_0 = -96.63 \text{ et } \widehat{\mathrm{V}}(\widehat{\beta}_0) = 582.74 \quad \widehat{\beta}_1 = 98.10 \text{ et } \widehat{\mathrm{V}}(\widehat{\beta}_1) = 165.89 \quad \widehat{\sigma} = 6.085$$

Question 9. Laquelle des affirmations suivant est correcte (une seule l'est)

A. β_0 et β_1 ne sont pas signficativement non-nuls
B. β_0 n'est pas signficativement non-nul, alors que β_1 est signficativement non-nul
C. β_0 est signficativement non-nul, alors que β_1 n'est pas signficativement non-nul
D. β_0 et β_1 sont signficativement non-nuls

Une fois estimé le modèle, on s'aperçoit qu'une information supplémentaire est disponible: le sexe de l'individu. On a alors

$$Y_i = \alpha_0 + \alpha_1 X_{1,i} + \alpha_2 X_{2,i} + \eta_i,$$

où $X_{2,i}$ est une variable qui prend la valeur 1 si l'individu i est une femme, et 0 si l'individu i est un homme.

Question 10. Laquelle des affirmations suivantes est correcte (une seule l'est)

A. à taille identique, une femme pèse α_2 kilogrammes de plus qu'un homme
B. à taille identique, une femme pèse α_2 kilogrammes de moins qu'un homme
C. à taille identique, une femme pèse $\alpha_0 + \alpha_2$ kilogrammes de plus qu'un homme
D. à taille identique, une femme pèse $\alpha_0 + \alpha_2$ kilogrammes de moins qu'un homme

2. Régressions sur données individuelles et matchs de basket

Les questions suivantes portent sur les sorties données dans l'Annexe A.

Question 11. Dans la sortie de régression **R1**, à quoi corrrespond la valeur `131.3` (valeur encadrée)

A. à la valeur de la pente dans la régression linéaire
B. à la statistique de Student permettant de tester si la pente de la droite de régression est nulle, ou pas
C. à la valeur t qui est (en ‰) la probabilité que le modèle soit linéaire
D. au score maximum à la fin d'un match, prédit par le modèle

Question 12. A la lecture des sorties **R2**, la quelle de ces affirmations vous semble la **moins valide** ?

A. les résidus suivent une loi normale centrée
B. avec ce modèle, on a environ 5% de chance de faire une erreur de prévision supérieure à 20 points
C. il y a de l'hétéroscédasticité, car la variance des résidus n'est pas constante avec la différence de points à la mi-temps
D. en moyenne, la différence de points à la fin du match est deux fois plus grande qu'à la mi-temps

Question 13. À l'aide de la sortie **R3** prédire la différence de points à la fin du match en faveur d'une équipe qui joue à domicile et qui mêne de 13 points à la mi-temps.

A. environ 20 points d'avance
B. environ 17 points d'avance
C. environ 12 points d'avance
D. environ 6 points d'avance

Question 14. À l'aide de la sortie **R4** laquelle des affirmations suivante semble la **moins juste** ? (une seule l'est)

A. les résidus sont nuls
B. l'écart-type des résidus est nul
C. il manque une constante dans la régression pour que le modèle soit valide
D. le R^2 de la régression vaut 1

Question 15. À l'aide de la sortie **R5** quelle affirmation serait la **moins juste** ?

A. les modèles `reg4` et `reg5` sont équivalents
B. le modèle `reg5` est meilleur que le modèle `reg6` au sens du R^2 ajusté, et du critère d'Akaike
C. le modèle `reg5` est meilleur que le modèle `reg2` au sens du R^2 ajusté, et du critère d'Akaike
D. Ces modèles ne sont pas comparables, car ls n'utilisent pas les mêmes variables explicatives.

Dans les questions suivantes (16-18), on va se demander si ces estimations donnent les mêmes résultats pour toutes les équipes, et on va se restreindre ‘a des sous-bases.

Question 16. Pour le modèle présenté dans la série des sorties **R6**, seuls les matchs joués par 6 équipes ont été retenus. Quelle interprétation du nombre `0.707` (encadré) retiendriez vous ?

A. l'équipe de basket de `COLUMBIA` a 70.7% de chance de gagner un match
B. l'équipe de basket de `COLUMBIA` a 70.7% de chance de gagner un match, si elle gagne à la mi-temps
C. le coefficient est non-significatif, ce qui veut dire que pour les matchs de l'équipe `COLUMBIA`, s'il y avait x points de différence à la mi-temps, la meilleure prédiction possible est qu'il y aura toujours x points de différence à la fin du match.
D. le coefficient est non-significatif, ce qui veut dire que l'équipe de basket de `COLUMBIA` a autant de chance de gagner contre toutes les équipes

Question 17. À l'aide des sorties **R6** laquelle affirmation, parmi les 4 proposées vous semble la **plus juste** ?

A. si l'équipe de `PERDUE` perd de 4 points à la mi-temps, l'équipe a plus de chance de gagner le match que de le perdre
B. si l'équipe d'`UCLA` gagne de 4 points à la mi-temps, l'équipe a plus de chance de perdre le perdre que de le gagner
C. si l'équipe de `COLUMBIA` perd de 4 points à la mi-temps, l'équipe a plus de chance de gagner le match que de le perdre
D. si l'équipe de `NORTHWESTERN` gagne de 4 points à la mi-temps, l'équipe a plus de chance de perdre le match que de le gagner

On va ici supposer que l'équipe qui nous intéresse perd fortement à la mi-temps (de plus de 10 poins d'écart).

Question 18. À l'aide des sorties **R7**, laquelle des affirmations suivantes vous semble la **moins juste** ?

A. si l'équipe de `PERDUE` perd de 10 points à la mi-temps, la probabilité de l'emporter à la fin du match est supérieure à 25%
B. si l'équipe d'`UCLA` perd de 10 points à la mi-temps,, la probabilité de l'emporter à la fin du match est supérieure à 25%
C. si l'équipe de `NORTHWESTERN` perd de 10 points à la mi-temps, la probabilité de l'emporter à la fin du match est supérieure à 25%
D. si l'équipe de `MICHIGAN STATE` perd de 10 points à la mi-temps, la probabilité de l'emporter à la fin du match est supérieure à 25%

Dans les questions suivantes (19-20), on va se demander ce qui se passe si les équipes sont ex-aequo, ou presque. Pour cela, on se limite aux matchs où à la mi-temps, les équipes étaient ex-aequo (différence de points nulle) ou avec un point d'écart.

Question 19. À l'aide des sorties **R8** et **R3**, si une équipe perd d'*un point* à la mi-temps, et joue à domicile, quelle serait la différence de prédiction entre les deux modèles. Une seule des affirmations suivantes est **fausse**.

A. avec le modèle de la sortie **R8**, peu importe que l'équipe mène ou perde d'un point, la prédiction sera significativement la même
B. avec le modèle de la sortie **R8**, peu importe que l'équipe qui perd joue à domicile, ou pas, la prédiction sera significativement la même
C. les deux modèles prévoient une victoire (avec un nombre de points d'écart à la fin du match positif, en moyenne)
D. les deux modèles `reg2` et `reg10` prévoient une victoire, mais avec un point de *moins* (environ) avec le modèle `reg2` qu'avec le modèle `reg10`

Question 20. À l'aide de la sortie **R8**, si une équipe perd d'un point à la mi-temps, en jouant à l'extérieur, quelle serait la probabilité qu'elle gagne le match ?

A. 0%
B. 20%
C. 40%
D. 80%

3. Modélisation de séries temporelles

Dans les questions 21-26, on considére un modèle autorégressif de la forme

$$X_t = \phi X_{t-1} + \varepsilon_t, \tag{3.1}$$

où (ε_t) est un bruit blanc. On supposera que $\phi \neq 0$, et on supposera dans toute la section que la série (X_t) est stationnaire.

Question 21. Laquelle des affirmations suivant n'est **pas correcte**

A. ε_t et X_{t-1} sont non-corrélés
B. X_t et ε_{t-1} sont non-corrélés
C. ε_t et ε_{t-1} sont non-corrélés
D. X_t et ε_t sont corrélés

Question 22. Laquelle des affirmations suivant est **correcte**

A. $\text{cor}(X_t, X_{t-1}) = \phi$
B. $\text{cor}(X_t, X_{t-2}) = 0$
C. $\text{cor}(X_t, \varepsilon_t) = 0$
D. $\text{cor}(\varepsilon_t, \varepsilon_{t-1}) = \phi$

On suppose ici que le bruit (ε_t) est de variance $\sigma^2 = 1$.

Question 23. Que vaut la variance de (X_t) ?

A. $\text{Var}(X_t) = 1$
B. $\text{Var}(X_t) = (1 - \phi^2)$
C. $\text{Var}(X_t) = \dfrac{1}{1 - \phi^2}$
D. $\text{Var}(X_t) = 0$

On suppose que l'on travaille sur des données journalières. On a observé -1 hier, et $+1$ aujourd'hui

Question 24. Quelle prévision feriez vous pour demain, à l'aide du modèle (3.1) ?

A. $\widehat{X} = \phi$
B. $\widehat{X} = -\phi$
C. $\widehat{X} = 0$
D. $\widehat{X} = \phi + \varepsilon$

Question 25. Quel serait la variance de l'erreur de prévision à deux jours (pour après demain) ?

A. ϕ
B. 1
C. $1 + \phi^2$
D. ϕ^2

Question 26. Une seule des affirmations suivantes s est correcte. Laquelle ?

A. La série $Y_t = t + X_t$ est stationnaire
B. La série $Y_t = X_t - X_{t-1}$ est stationnaire
C. La série Y_t vérifiant la relation $Y_t = Y_{t-1} + X_t$ est stationnaire
D. La série Y_t vérifiant la relation $Y_t = \phi^{-1} Y_{t-1} + \varepsilon_t$ est stationnaire

En regardant de plus près, on se rend compte que la série (ε_t) ne peut être considérée comme étant un bruit blanc. On suppose maintenant que la série (X_t) suit une processus $ARMA(1,1)$

$$X_t = \phi X_{t-1} + \eta_t + \theta \eta_{t-1} \tag{3.2}$$

où (η_t) est un bruit blanc de variance σ^2.

Question 27. Quelle est la condition nécessaire et suffisante, sur ϕ, θ et σ, pour que (X_t) soit stationnaire

A. $|\phi| < 1$, $|\theta| < 1$ et $\sigma^2 < 1$,
B. $|\phi| < 1$ et $|\theta| < 1$
C. $|\theta| < 1$
D. $|\phi| < 1$

Question 28. Parmi les affirmations suivantes, une seule est toujours vraie. Laquelle ? (un seul choix est possible)

A. Toutes les prédictions faites avec le modèle (3.2) seront identiques avec celles faites à l'aide du modèle (3.1)
B. Toutes les prédictions faites avec le modèle (3.2) auront une variance (conditionnelle à l'information disponible aujourd'hui) identiques à celles du modèle (3.1)
C. L'autocorrélation à l'ordre 1 est toujours non-nulle $\rho_X(1) \neq 0$
D. À partir, d'un moment, on retrouve la propriété vérifiée par les $AR(1)$ sur les autocorrélations, à savoir $\rho_X(h) = \phi\rho_X(h-1)$ (pour $h \geq 1$).

Question 29. Un voisin suggère que dans le modèle (3.2), on pourrait avoir $\phi+\theta = 0$. Dans ce cas, quelle affirmation parmi les suivantes serait **correcte**

A. (X_t) suit une marche aléatoire
B. (X_t) suit un bruit blanc
C. (X_t) est, en fait, un processus $AR(1)$
D. (X_t) est, en fait, un processus $MA(1)$

On suppose désormais que la série (X_t) suit une processus $AR(p)$ stationnaire,

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + u_t, \tag{3.3}$$

où (u_t) est un bruit blanc. On supposera que $\phi_p \neq 0$.

Question 30. Parmi les affirmations suivantes, une seule est (de manière générale) **vraie**. Laquelle ?

A. la série des autocorrélations $\rho(h) = \text{cor}(X_t, Xt - h)$ est nulle pour $h > p$

B. la série des autocorrélations partielles $\psi(h) = \text{cor}(X_t - \text{EL}(X_t | X_{t-1}, \cdots, X_{t-h+1}, X_{t-h}) - \text{EL}(X_{t-h} | X_{t-1}, \cdots, X_{t-h+1}))$ est nulle pour $h > p$

C. l'autocorrélation d'ordre p est non nulle, $\rho(p) \neq 0$

D. l'autocorrélation d'ordre p vaut 1, $\rho(p) = 1$

4. Modélisation de séries temporelles

On cherche ici à modéliser, puis à proposer une prévision pour une série de ventes, dont les sorties sont proposées dans l'Annexe B.

On notera ρ_X la fonction d'autocorrélations de la série (X_t) et ψ_X sa fonction d'autocorrélations partielles. Pour un processus $ARMA(p,q)$, on utilisera la notation suivante,

$$X_t = \alpha + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q},$$

où (ε_t) est un bruit blanc de variance σ^2, et où $\phi_p \neq 0$ et $\theta_q \neq 0$.

Question 31. À l'aide de la sortie **T1**, parmi les affirmations suivantes, laquelle n'est **pas correcte**

A. $\rho(12)$ est significativement non-nulle
B. $\rho(6)$ est significativement non-nulle
C. $\rho(3)$ est significativement non-nulle
D. $\psi(12)$ est significativement non-nulle

Question 32. À l'aide des sorties **T1** et **T2**, parmi les affirmations suivantes, laquelle est **correcte** (avec un seuil d'acceptation à 95%)

A. la série (X_t) est stationnaire
B. la série (X_t) n'est pas stationnaire
C. la série (X_t) est un bruit blanc
D. la série (X_t) est une marche aléatoire

Question 33. À l'aide des sorties **T3**, sur la modélisation par un processus $AR(18)$ pour (X_t), laquelle de ces affirmations suivantes vous semble **la plus juste**,

A. ϕ_{18} est significativement non-nulle
B. la plus grande valeur p pour laquelle ϕ_p est significativement non-nulle est 12
C. la constante est significative
D. le bruit associé (ε_t) n'est sûrement pas un bruit blanc

Question 34. À l'aide des sorties **T4**, sur la modélisation par un processus $AR(12)$ pour (X_t), laquelle de ces affirmations suivantes vous semble **la moins juste**,

A. ϕ_{12} est significativement non-nulle
B. le bruit associé (ε_t) peut être considére comme étant un bruit blanc
C. le bruit associé (ε_t) n'est pas un bruit blanc,
D. même si on n'a pas vérifier le bruit associé (ε_t) est centré, il l'est forcément car on a mis une constante dans le modèle

Question 35. À l'aide des sorties **T5**, sur la modélisation par un processus $AR(3)$ pour (X_t), laquelle de ces affirmations suivantes vous semble la **plus juste**,

A. le bruit associé (ε_t) peut être considére comme étant un bruit blanc
B. le bruit associé (ε_t) n'est probablement pas un bruit blanc,
C. on ne peut pas savoir si le bruit associé (ε_t) est un bruit blanc, car on n'a pas fait de test de racine unité sur ce dernier.
D. on ne peut pas savoir si le bruit associé (ε_t) est un bruit blanc, car on n'a pas vérifié que ce dernier était centré

Question 36. À l'aide des sorties **T6**, sur la modélisation par un processus $ARMA(3, 12)$ pour (X_t), laquelle de ces affirmations suivantes vous semble la **moins juste**,

A. ϕ_3 et θ_{18} sont significativement non-nulles
B. le bruit associé (ε_t) peut être considére comme étant un bruit blanc
C. le bruit associé (ε_t) n'est probablement pas un bruit blanc,
D. en utilisant le critère AIC, on ne retiendrait pas ce modèle car on en a trouvé des plus simples, et meilleurs auparavant.

Question 37. À l'aide des sorties **T7**, sur la modélisation par un processus $AR(9)$ pour (X_t), laquelle de ces affirmations suivantes vous semble la **moins juste**,

A. ϕ_9 est significativement non-nulle

B. comme $|\phi_6|$ est plus grand que $|\phi_9|$ on rejette un $AR(9)$ au profit d'un modèle $AR(6)$

C. le bruit associé (ε_t) n'est probablement pas un bruit blanc,

D. bien que le modèle est est suggéré (en premier choix) par la fonction `armaselect`, sur la base d'un critère SBC minimal parmi tous les $ARMA(p, q)$, on ne devrait pas le retenir.

Question 38. À l'aide des sorties **T8**, sur la modélisation par un processus $AR(6)$ pour (X_t), laquelle de ces affirmations suivantes vous semble la **plus juste**,

A. le modèle ne peut pas être correct car le bruit associé (ε_t) n'est probablement pas un bruit blanc,

B. le modèle ne peut pas être correct car comme $\rho(12)$ était significativement non-nulle, il faut forcément un modèle $ARMA(p, q)$ avec $p + q \geq 12$,

C. le modèle ne peut pas être correct car la variance du bruit associé (ε_t) est plus grande que 1, et donc le modèle n'est pas stationnaire,

D. le modèle ne peut pas être correct car $|\phi_1 + \phi_2 + \cdots + \phi_6| \approx 1$, et donc on est en présence d'une racine unité,

Question 39. À l'aide des sorties **T9** et **T10**, sur les modélisations par des processus $AR(10)$ et $ARMA(10,1)$ de la série (X_t), laquelle de ces affirmations suivante vous semble **la plus juste**,

A. le processus $ARMA(10,1)$ est préférable au processus $AR(10)$, car le bruit associé (ε_t) est de variance plus faible
B. le processus $ARMA(10)$ est préférable au processus $AR(10,1)$ car le coefficient θ_1 n'est pas significatif
C. le modèle $ARMA(10,1)$ ne peut être retenu car comme la somme des coefficients vaut $-\pi$, il faut un modèle avec une composante cyclique,
D. le modèle $ARMA(10)$ ne peut être retenu car $\rho_\varepsilon(12)$ était significativement non-nulle, il faudrait tester un modèle $MA(12)$ sur les résidus

Question 40. À l'aide des sorties **T11**, parmi les modèles suivants, lequel retriendriez-vous pour faire une prévision sur un an (pour les 12 prochaines observations) ? On retiendra le meilleur modèle sur les observations récentes,

A. un processus $AR(9)$
B. un processus $AR(10)$
C. un processus $AR(12)$
D. un processus $AR(12,3)$