

Test de Préparation: Séries Temporelles et Prévisions

Informations

- **Durée estimée:** 45 minutes
 - **Nombre de questions:** 25
 - **Niveau:** Intermédiaire-Avancé
 - **Thèmes:** Composantes, stationnarité, modèles de prévision, applications bancaires
-

Section 1: Concepts Fondamentaux (5 questions)

Question 1

Quelles sont les quatre composantes principales d'une série temporelle?

- A) Moyenne, Variance, Écart-type, Médiane
- B) Tendance, Saisonnalité, Cycle, Irrégulier
- C) ACF, PACF, AIC, BIC
- D) AR, MA, ARIMA, SARIMA

Réponse

B) Tendance, Saisonnalité, Cycle, Irrégulier

Mnémotechnique: **TSCI** - Tendance: Direction long terme (hausse/baisse) - Saisonnalité: Pattern répétitif à période fixe (mensuel, trimestriel) - Cycle: Fluctuations à long terme (cycles économiques) - Irrégulier: Variations aléatoires imprévisibles

Formule additive: $Y_t = T_t + S_t + C_t + I_t$

Question 2

Quelle est la différence entre un modèle additif et un modèle multiplicatif?

- A) Additif pour les grandes séries, multiplicatif pour les petites
- B) Additif si l'amplitude saisonnière est constante, multiplicatif si elle varie avec le niveau
- C) Additif pour les tendances, multiplicatif pour les cycles
- D) Il n'y a pas de différence

Réponse

B) Additif si l'amplitude saisonnière est constante, multiplicatif si elle varie avec le niveau

Modèle	Formule	Quand utiliser
Additif	$Y = T + S + I$	Amplitude saisonnière constante
Multiplicatif	$Y = T \times S \times I$	Amplitude proportionnelle au niveau

Exemple bancaire: - Additif: Dépôts fluctuent de $\pm 10M$ HTG chaque mois - Multiplicatif: Dépôts fluctuent de $\pm 5\%$ chaque mois (donc plus si niveau élevé)

Question 3

Qu'est-ce qu'une série temporelle stationnaire?

- A) Une série qui ne change jamais
- B) Une série dont les propriétés statistiques (moyenne, variance) sont constantes dans le temps
- C) Une série sans tendance
- D) Une série avec seulement du bruit

Réponse

B) Une série dont les propriétés statistiques (moyenne, variance) sont constantes dans le temps

Conditions de stationnarité faible: 1. $E[Y_t] = \mu$ (moyenne constante) 2. $\text{Var}(Y_t) = \sigma^2$ (variance constante) 3. $\text{Cov}(Y_t, Y_{t+k}) = \gamma_k$ (covariance ne dépend que du lag k)

Importance: La plupart des modèles (ARIMA) requièrent la stationnarité.

Question 4

Quel test permet de vérifier la stationnarité d'une série?

- A) Test de Shapiro-Wilk
- B) Test ADF (Augmented Dickey-Fuller)
- C) Test de Durbin-Watson
- D) Test du Chi-carré

Réponse

B) Test ADF (Augmented Dickey-Fuller)

Test ADF: - H_0 : La série a une racine unitaire (NON stationnaire) - H_1 : La série est stationnaire

Interprétation: - $p < 0.05 \rightarrow$ Rejeter $H_0 \rightarrow$ Série stationnaire - $p \geq 0.05 \rightarrow$ Série NON stationnaire
→ Différencier

Autre test: KPSS (hypothèses inversées)

Question 5

Comment rendre une série non stationnaire en série stationnaire?

- A) Ajouter plus de données
- B) Appliquer une différenciation (calculer $Y_t - Y_{t-1}$)
- C) Calculer la moyenne mobile
- D) Supprimer les outliers

Réponse

B) Appliquer une différenciation (calculer $Y_t - Y_{t-1}$)

Transformations pour stationnarité: 1. **Différenciation simple:** $\Delta Y_t = Y_t - Y_{t-1}$ (retire la tendance) 2. **Différenciation saisonnière:** $Y_t - Y_{t-s}$ (retire la saisonnalité, s=période) 3. **Transformation log:** $\ln(Y_t)$ (stabilise la variance)

```
# Python
df['diff1'] = df['y'].diff()      # Différence d'ordre 1
df['diff12'] = df['y'].diff(12)    # Différence saisonnière
```

Section 2: ACF, PACF et ARIMA (5 questions)

Question 6

Que mesure l'ACF (Autocorrelation Function)?

- A) La corrélation entre Y et X
- B) La corrélation entre Y_t et Y_{t-k} pour différents lags k
- C) La variance de la série
- D) La tendance de la série

Réponse

B) La corrélation entre Y_t et Y_{t-k} pour différents lags k

ACF (Autocorrélation): - Mesure la corrélation entre une observation et ses valeurs passées - Inclut les effets directs ET indirects - Utilisée pour identifier l'ordre q du modèle MA

PACF (Autocorrélation partielle): - Corrélation directe uniquement (contrôlée pour les lags intermédiaires) - Utilisée pour identifier l'ordre p du modèle AR

Question 7

Dans la notation ARIMA(p, d, q), que représentent p, d et q?

- A) Probabilité, Densité, Quantile
- B) Ordre autorégressif, ordre de différenciation, ordre de moyenne mobile
- C) Paramètre, Dimension, Qualité
- D) Passé, Différence, Question

Réponse

B) Ordre autorégressif, ordre de différenciation, ordre de moyenne mobile

ARIMA(p, d, q): - **p**: Ordre AR (AutoRegressive) - combien de lags de Y utiliser - **d**: Ordre I (Integrated) - combien de fois différencier - **q**: Ordre MA (Moving Average) - combien de lags d'erreur utiliser

Exemple: ARIMA(1, 1, 1) - AR(1): Y_t dépend de Y_{t-1} - I(1): Une différenciation - MA(1): ε_t dépend de ε_{t-1}

Question 8

L'ACF montre une décroissance exponentielle et le PACF une coupure nette au lag 2. Quel modèle suggère ce pattern?

- A) MA(2)
- B) AR(2)
- C) ARIMA(2,1,0)
- D) ARIMA(0,1,2)

Réponse

B) AR(2)

Règles de lecture ACF/PACF:

Pattern ACF	Pattern PACF	Modèle
Décroissance exponentielle	Coupure au lag p	AR(p)
Coupure au lag q	Décroissance exponentielle	MA(q)
Décroissance des deux	Décroissance des deux	ARMA(p,q)

Ici: ACF décroît exponentiellement + PACF coupe au lag 2 → AR(2)

Question 9

Quelle est la différence entre ARIMA et SARIMA?

- A) SARIMA est plus rapide
- B) SARIMA inclut une composante saisonnière
- C) ARIMA est pour les grandes séries, SARIMA pour les petites
- D) Il n'y a pas de différence

Réponse

B) SARIMA inclut une composante saisonnière

SARIMA(p, d, q)(P, D, Q, s): - (p, d, q): Partie non saisonnière - (P, D, Q, s): Partie saisonnière - s : Période saisonnière (12 pour mensuel, 4 pour trimestriel)

Exemple: SARIMA(1, 1, 1)(1, 1, 1, 12) - Modèle avec tendance ET saisonnalité mensuelle - Idéal pour les dépôts bancaires avec pattern annuel

Question 10

Comment choisir entre plusieurs modèles ARIMA?

- A) Choisir celui avec le plus grand R^2
- B) Choisir celui avec le plus petit AIC ou BIC
- C) Choisir le modèle le plus complexe
- D) Choisir aléatoirement

Réponse

B) Choisir celui avec le plus petit AIC ou BIC

Critères de sélection: - **AIC (Akaike):** $AIC = 2k - 2\ln(L)$ - **BIC (Bayesian):** $BIC = k \times \ln(n) - 2\ln(L)$

Où k = nombre de paramètres, L = vraisemblance

Plus bas = Meilleur (pénalise la complexité)

```
# Sélection automatique
from pmtdarima import auto_arima
model = auto_arima(y, seasonal=True, m=12)
print(model.aic())
```

Section 3: Modèles de Lissage (5 questions)

Question 11

Quelle est la différence entre une moyenne mobile simple et une moyenne mobile exponentielle?

- A) La MMS est plus rapide à calculer
- B) La MME donne plus de poids aux observations récentes
- C) La MMS est pour les séries courtes, MME pour les longues
- D) Il n'y a pas de différence

Réponse

B) La MME donne plus de poids aux observations récentes

Type	Pondération	Formule
MMS	Égale	$MM = (Y_1 + Y_2 + \dots + Y_k) / k$
MME	Exponentielle	$EMA = \alpha \times Y_t + (1-\alpha) \times EMA_{t-1}$

La MME réagit plus vite aux changements récents grâce au paramètre α (smoothing factor).

```
df['MMS_12'] = df['y'].rolling(window=12).mean()
df['MME'] = df['y'].ewm(span=12).mean()
```

Question 12

Que modélise le lissage exponentiel de Holt-Winters?

- A) Niveau uniquement
- B) Niveau et tendance
- C) Niveau, tendance et saisonnalité
- D) Saisonnalité uniquement

Réponse

C) Niveau, tendance et saisonnalité

Hiérarchie des lissages exponentiels:

Modèle	Composantes	Paramètres
Simple (SES)	Niveau	α
Holt	Niveau + Tendance	α, β
Holt-Winters	Niveau + Tendance + Saison	α, β, γ

```
from statsmodels.tsa.holtwinters import ExponentialSmoothing
model = ExponentialSmoothing(
    y, trend='add', seasonal='add', seasonal_periods=12
).fit()
```

Question 13

Le paramètre α dans le lissage exponentiel simple est de 0.9. Que signifie cela?

- A) 90% de poids sur les observations anciennes
- B) 90% de poids sur l'observation la plus récente
- C) Lissage très fort
- D) Prévision constante

Réponse

B) 90% de poids sur l'observation la plus récente

Formule SES: $\hat{Y}_{t+1} = \alpha \times Y_t + (1-\alpha) \times \hat{Y}_t$

Interprétation de α : - α proche de 0: Lissage fort (poids aux anciennes observations) - α proche de 1: Peu de lissage (réactif aux changements récents)

$\alpha = 0.9$ signifie: - 90% de poids sur la dernière observation - 10% sur l'historique lissé - Prévision très réactive (peut être bruitée)

Question 14

Prophet (Facebook) est particulièrement utile pour:

- A) Les très courtes séries (< 10 observations)
- B) Les séries avec jours fériés et patterns complexes
- C) Les séries sans saisonnalité
- D) Les séries stationnaires uniquement

Réponse

B) Les séries avec jours fériés et patterns complexes

Avantages de Prophet: - Gestion automatique des jours fériés - Saisonnalités multiples (quotidienne, hebdomadaire, annuelle) - Robuste aux données manquantes - Facile à utiliser (pas besoin de spécifier p, d, q) - Gestion des changements de tendance

```
from prophet import Prophet
model = Prophet(yearly_seasonality=True)
model.add_country_holidays(country_name='HT')  # Fériés Haïti
model.fit(df)
```

Question 15

Quelle méthode de validation est appropriée pour les séries temporelles?

- A) K-fold cross-validation standard
- B) Train-test split aléatoire
- C) Time Series Split (validation temporelle)
- D) Leave-one-out

Réponse

C) Time Series Split (validation temporelle)

Important: Ne jamais mélanger les données temporelles!

Time Series Split:

```

Fold 1: Train [1-100], Test [101-120]
Fold 2: Train [1-120], Test [121-140]
Fold 3: Train [1-140], Test [141-160]
...
from sklearn.model_selection import TimeSeriesSplit
tscv = TimeSeriesSplit(n_splits=5)
for train_idx, test_idx in tscv.split(y):
    train, test = y[train_idx], y[test_idx]
    # Entrainer et évaluer

```

Raison: Prédire le futur à partir du passé, pas l'inverse!

Section 4: Métriques et Diagnostics (5 questions)

Question 16

Quelle métrique utilise des pourcentages pour mesurer l'erreur de prévision?

- A) MAE (Mean Absolute Error)
- B) MSE (Mean Squared Error)
- C) MAPE (Mean Absolute Percentage Error)
- D) RMSE (Root Mean Squared Error)

Réponse

C) MAPE (Mean Absolute Percentage Error)

Formules:

Métrique	Formule	Unité
MAE	$\sum Y_i - \hat{Y}_i / n$	Même que Y
MSE	$\sum (Y_i - \hat{Y}_i)^2 / n$	Y^2
RMSE	\sqrt{MSE}	Même que Y
MAPE	$\sum (Y_i - \hat{Y}_i)/Y_i \times 100 / n$	%

MAPE avantage: Interprétable en pourcentage, comparable entre séries
Problème si Y_i proche de 0

Question 17

Un modèle a un MAPE de 8%. Comment interpréter ce résultat?

- A) Très mauvais modèle
- B) Modèle acceptable - erreur moyenne de 8% par rapport aux valeurs réelles
- C) Le modèle explique 8% de la variance
- D) 8% des observations sont mal prédites

Réponse

B) Modèle acceptable - erreur moyenne de 8% par rapport aux valeurs réelles

Interprétation du MAPE:

MAPE	Qualité
< 10%	Très bon
10-20%	Bon
20-50%	Acceptable
> 50%	Mauvais

Un MAPE de 8% signifie que les prévisions sont en moyenne à $\pm 8\%$ des valeurs réelles.

Exemple: Si les dépôts réels sont de 100M HTG, la prévision est typiquement entre 92M et 108M HTG.

Question 18

Les résidus d'un modèle ARIMA montrent une autocorrélation significative. Que conclure?

- A) Le modèle est parfait
- B) Le modèle ne capture pas toute l'information - il reste un pattern
- C) Les données sont incorrectes
- D) Il faut ajouter des variables exogènes

Réponse

B) Le modèle ne capture pas toute l'information - il reste un pattern

Diagnostic des résidus d'un bon modèle: 1. Résidus non autocorrélés (ACF proche de 0) 2. Moyenne proche de 0 3. Variance constante 4. Distribution normale

Si autocorrélation dans les résidus: - Augmenter l'ordre p ou q - Ajouter une composante saisonnière - Considérer un modèle plus complexe

Test: Ljung-Box pour l'absence d'autocorrélation

Question 19

Qu'indique le test de Ljung-Box avec p-value = 0.35?

- A) Les résidus sont autocorrélés
- B) Les résidus ne sont pas significativement autocorrélés
- C) Le modèle est mauvais
- D) La série n'est pas stationnaire

Réponse

B) Les résidus ne sont pas significativement autocorrélés

Test de Ljung-Box: - H_0 : Pas d'autocorrélation dans les résidus - H_1 : Présence d'autocorrélation

Interprétation: - $p > 0.05 \rightarrow$ Ne pas rejeter $H_0 \rightarrow$ Résidus OK ✓ - $p < 0.05 \rightarrow$ Rejeter $H_0 \rightarrow$ Autocorrélation présente

Avec $p = 0.35 > 0.05$, les résidus sont acceptables (pas de pattern non capturé).

Question 20

Comment interpréter une prévision avec intervalle de confiance [85, 115] pour la valeur centrale 100?

- A) La vraie valeur sera exactement 100
- B) On est 95% confiant que la vraie valeur sera entre 85 et 115
- C) L'erreur est de 15%
- D) Le modèle n'est pas fiable

Réponse

B) On est 95% confiant que la vraie valeur sera entre 85 et 115

L'intervalle de confiance (généralement à 95%) indique la plage dans laquelle la vraie valeur devrait tomber.

Utilisation en banque: - Prévision dépôts: 100M HTG [85M - 115M] - Planification prudente: Prévoir pour 85M (borne basse) - Gestion de la liquidité: Capacité pour 115M (borne haute)

Plus l'intervalle est large, plus l'incertitude est grande.

Section 5: Applications Bancaires (5 questions)

Question 21

Une banque veut prévoir les dépôts mensuels qui montrent une saisonnalité claire (plus élevés en décembre). Quel modèle recommandez-vous?

- A) ARIMA(1,1,1) sans saisonnalité
- B) SARIMA ou Holt-Winters avec saisonnalité
- C) Moyenne mobile simple
- D) Régression linéaire

Réponse

B) SARIMA ou Holt-Winters avec saisonnalité

Pour une série avec saisonnalité: - **SARIMA(p,d,q)(P,D,Q,12)**: Modèle autorégressif avec composante saisonnière - **Holt-Winters**: Lissage avec niveau, tendance et saisonnalité

```
# Option 1: SARIMA
from statsmodels.tsa.statespace.sarimax import SARIMAX
model = SARIMAX(y, order=(1,1,1), seasonal_order=(1,1,1,12))

# Option 2: Holt-Winters
from statsmodels.tsa.holtwinters import ExponentialSmoothing
model = ExponentialSmoothing(y, trend='add', seasonal='add', seasonal_periods=12)
```

Ne pas utiliser ARIMA simple car il ignorerait la saisonnalité.

Question 22

La prévision du taux de défaut pour le prochain trimestre est de 4.5% avec IC [3.8%, 5.2%]. Quelle provision recommander?

- A) Provisionner pour 4.5%
- B) Provisionner pour 3.8% (optimiste)

- C) Provisionner pour 5.2% (prudent)
- D) Ne pas provisionner

Réponse

C) Provisionner pour 5.2% (prudent)

Principe de prudence bancaire: - Utiliser la borne supérieure de l'IC pour les provisions - Éviter le sous-provisionnement (risque réglementaire)

Calcul: - Encours crédit: 10 milliards HTG - Provision recommandée: $10B \times 5.2\% = 520M$ HTG

Remarque: Les régulateurs (BRH) attendent généralement une approche conservatrice.

Question 23

Un modèle de prévision des transactions ATM montre des pics tous les 30 jours. Que reflète ce pattern?

- A) Bruit aléatoire
- B) Saisonnalité mensuelle (probablement liée aux salaires)
- C) Tendance haussière
- D) Problème de données

Réponse

B) Saisonnalité mensuelle (probablement liée aux salaires)

En Haïti et dans la plupart des pays: - Les salaires sont versés en fin de mois - Pic de retraits ATM après le versement - Pattern répétitif avec période de ~30 jours

Applications: - Dimensionner l'approvisionnement des ATM - Planifier les équipes de support - Prévoir la demande de liquidité

Modèle approprié: SARIMA avec s=30 (jours) ou Prophet avec saisonnalité hebdomadaire/mensuelle.

Question 24

Comment décomposer une série pour identifier la tendance sous-jacente des prêts?

- A) Calculer la moyenne
- B) Utiliser seasonal_decompose de statsmodels
- C) Supprimer les outliers
- D) Appliquer une régression

Réponse

B) Utiliser seasonal_decompose de statsmodels

```
from statsmodels.tsa.seasonal import seasonal_decompose

# Décomposition
decomposition = seasonal_decompose(
    df['prets'],
    model='additive', # ou 'multiplicative'
    period=12          # Période saisonnière
)
```

```

# Accéder aux composantes
tendance = decomposition.trend
saisonnalite = decomposition.seasonal
residus = decomposition.resid

# Visualiser
decomposition.plot()

```

Cela isole: - Tendance: Direction long terme des prêts - Saisonnalité: Patterns récurrents - Résidus: Variations inexplicées

Question 25

Une série de dépôts a un test ADF avec p-value = 0.45. Quelle est la prochaine étape avant de modéliser?

- A) Appliquer directement ARIMA
- B) Différencier la série et retester
- C) Utiliser Prophet sans modification
- D) Conclure que les données sont mauvaises

Réponse

B) Différencier la série et retester

Avec $p = 0.45 > 0.05$, le test ADF indique que la série est NON stationnaire.

Étapes: 1. Différencier: `y_diff = y.diff().dropna()` 2. Retester: `adfuller(y_diff)` 3. Si toujours non stationnaire, différencier à nouveau 4. Utiliser le nombre de différenciations comme ordre d dans ARIMA(p,d,q)

```

from statsmodels.tsa.stattools import adfuller

# Série originale non stationnaire
y_diff = df['depots'].diff().dropna()
result = adfuller(y_diff)
print(f"P-value après différenciation: {result[1]:.4f}")
# Si p < 0.05, série stationnaire → d=1 pour ARIMA

```

Résumé et Mnémotechniques

“TSCI” - Composantes

T - Tendance: Direction long terme
S - Saisonnalité: Pattern régulier
C - Cycle: Fluctuations économiques
I - Irrégulier: Bruit aléatoire

“SADIM” - Étapes de Modélisation

S - Stationnarité: Tester (ADF)
A - ACF/PACF: Analyser les lags
D - Déterminer: Choisir p, d, q
I - Implémenter: Ajuster le modèle

M - Mesurer: Évaluer (AIC, MAPE)

Seuils à Retenir

Test/Métrique	Seuil	Interprétation
ADF p-value	< 0.05	Stationnaire
MAPE	< 10%	Bon modèle
AIC/BIC	Plus bas	Meilleur modèle
Ljung-Box p	> 0.05	Résidus OK

Modèles Courants

Situation	Modèle Recommandé
Sans tendance ni saison	SES ou ARIMA
Avec tendance	Holt ou ARIMA avec d>0
Avec saison	Holt-Winters ou SARIMA
Patterns complexes	Prophet

Score et Auto-évaluation

Score	Niveau	Recommandation
23-25	Expert	Prêt pour l'examen
18-22	Avancé	Réviser les points faibles
13-17	Intermédiaire	Revoir le document complet
< 13	Débutant	Étude approfondie nécessaire

Test préparé pour l'examen Data Analyst - UniBank Haiti Thème: Séries Temporelles - Maîtriser la dimension temps