Crash Course in Forecasting - Quiz Questions

Instructions

- Each question has 4 options (A, B, C, D)
- Select the best answer for each question
- Answers and explanations are provided at the end
- Time limit: 20 minutes
- Passing score: 70% (7 out of 10 correct)

Questions

Question 1

What does the 'I' (Integrated) component in ARIMA represent?

A) The number of autoregressive terms in the model
B) The order of differencing needed to achieve stationarity
C) The number of moving average terms in the model
D) The integration of external variables into the model

Question 2

Which statistical test is commonly used to check for stationarity in time series data?

A) Chi-square test
B) T-test
C) Augmented Dickey-Fuller (ADF) test
D) F-test

Question 3

In a SARIMA(1,1,1)(1,1,1,12) model, what does the '12' represent?

A) The number of observations in the training set
B) The number of parameters to estimate
C) The seasonal period (12 months for monthly data)
D) The forecast horizon

Question 4

What is the primary purpose of examining ACF (Autocorrelation Function) plots?

A) To identify the AR (autoregressive) order (p)
B) To identify the MA (moving average) order (q)
C) To test for stationarity
D) To calculate forecast accuracy

Question 5

When should you use SARIMAX instead of SARIMA?

A) When you have missing data in your time series
B) When you want to incorporate external variables (like temperature or promotions)
C) When your data has multiple seasonal patterns
D) When you need faster computation time

Question 6

Which of the following indicates that a time series is NOT stationary?

A) Constant mean over time
B) Constant variance over time
C) Trend or systematic change in mean over time
D) No autocorrelation in residuals

Question 7

What does MAPE stand for, and what does a MAPE of 5% indicate?

A) Mean Absolute Prediction Error; forecasts are 5% away from actual on average
B) Mean Absolute Percentage Error; forecasts are off by 5% on average

C) Maximum Allowed Prediction Error; maximum error is 5%
D) Median Average Percentage Error; median error is 5%

Question 8

In the Ljung-Box test for residual diagnostics, what does a p-value > 0.05 suggest?

A) The residuals are autocorrelated (bad)
B) The model is overfitting
C) The residuals are NOT significantly autocorrelated (good)
D) The model needs more parameters

Question 9

Why is it important to apply log transformation to time series data with increasing variance?

A) To make the data stationary by removing trend
B) To stabilize variance and make seasonal fluctuations consistent
C) To remove outliers from the dataset
D) To speed up model computation

Question 10

What is the main difference between correlation and causation in forecasting?

A) There is no difference; they are the same concept
B) Correlation means two variables move together; causation means one causes the other
C) Correlation is stronger than causation
D) Causation can be tested with correlation coefficients

Answer Key

Question 1: B

Explanation: The 'I' (Integrated) in ARIMA represents the order of differencing (d) needed to transform a non-stationary time series into a stationary one. For example, if d=1, you take first differences: y'*t* = *y_t* - *y*{t-1}.

Question 2: C

Explanation: The Augmented Dickey-Fuller (ADF) test is the most commonly used statistical test for checking stationarity. If the p-value < 0.05, we reject the null hypothesis of non-stationarity, meaning the series is stationary.

Question 3: C

Explanation: In SARIMA(p,d,q)(P,D,Q,m), the 'm' parameter represents the seasonal period. For monthly data with yearly seasonality, m=12. For daily data with weekly seasonality, m=7.

Question 4: B

Explanation: The ACF (Autocorrelation Function) plot helps identify the MA (moving average) order (q). A sharp cutoff at lag q suggests an MA(q) model. The PACF (Partial Autocorrelation Function) is used to identify the AR order (p).

Question 5: B

Explanation: SARIMAX extends SARIMA by allowing eXogenous variables (external factors) to be included in the model. For example, including temperature when forecasting energy consumption, or including promotional data when forecasting sales.

Question 6: C

Explanation: A stationary time series has constant statistical properties over time. A trend (systematic change in mean) or changing variance indicates non-stationarity. Differencing or transformation is needed to achieve stationarity.

Question 7: B

Explanation: MAPE (Mean Absolute Percentage Error) measures forecast accuracy as a percentage. MAPE = mean(|Actual - Forecast| / |Actual|) × 100%. A MAPE of 5% means forecasts are typically within 5% of actual values.

Question 8: C

Explanation: The Ljung-Box test checks if residuals are autocorrelated. A p-value > 0.05 suggests we cannot reject the null hypothesis of no autocorrelation, meaning residuals behave like white noise (which is good - it means the model captured all patterns).

Question 9: B

Explanation: Log transformation stabilizes variance when seasonal fluctuations grow proportionally with the level of the series (multiplicative seasonality). It converts multiplicative patterns to additive ones, making the series easier to model.

Question 10: B

Explanation: Correlation means two variables move together (e.g., ice cream sales and drowning incidents both increase in summer). Causation means one variable directly causes changes in another. In forecasting, we often use correlations, but understanding causation leads to more robust predictions.