**ARIMA Model Explained**

ARIMA, which stands for Autoregressive Integrated Moving Average, is a powerful statistical method used for time series analysis and forecasting. Here's a detailed explanation:

Components: ARIMA models consist of three key components:

Autoregressive (AR) component: Represents the relationship between the current observation and its lagged values.

Integrated (I) component: Denotes differencing the time series data to make it stationary, i.e., removing trends and seasonality.

Moving Average (MA) component: Depicts the relationship between the current observation and a residual error from a moving average model.

Order: ARIMA models are denoted as ARIMA(p, d, q), where:

p: Order of the autoregressive component.

d: Degree of differencing needed to make the series stationary.

q: Order of the moving average component.

A stationary time series is one whose properties do not depend on the time at which the series is observed.[15](https://otexts.com/fpp2/stationarity.html#fn15) Thus, time series with trends, or with seasonality, are not stationary — the trend and seasonality will affect the value of the time series at different times.

## What Is Seasonality?

Seasonality is a characteristic of a time series in which the data experiences regular and predictable changes that recur every calendar year. Any predictable fluctuation or pattern that recurs or repeats over a one-year period is said to be seasonal.

**The Dickey-Fuller test, also known as the Augmented Dickey-Fuller (ADF) test, is a statistical test commonly used to determine if a time series is stationary or not. Here's a breakdown:**

Purpose: It checks the null hypothesis that a unit root is present in a time series dataset, indicating non-stationarity. A rejection of the null hypothesis suggests stationarity.

Test Statistic: The test statistic from the ADF test is compared with critical values at certain confidence levels to make a decision about the stationarity of the time series.

Interpretation:

If the test statistic is less than the critical value, the null hypothesis is rejected, indicating stationarity.

If the test statistic is greater than the critical value, the null hypothesis cannot be rejected, suggesting non-stationarity.

Implementation: The test is often performed using libraries like StatsModels in Python, which provides the adfuller() function.

Output: The test results include the test statistic, p-value, number of lags used in the regression, number of observations used, and critical values for different confidence levels.

Usage: It's essential in time series analysis to ensure stationarity, as many forecasting methods assume it. ADF test helps identify whether differencing or other transformations are necessary.

**What is a Critical Value?**

A critical value is a threshold or boundary used in hypothesis testing to determine the significance of a test statistic. It's compared with the test statistic to decide whether to reject the null hypothesis. If the test statistic exceeds the critical value, it indicates evidence against the null hypothesis. Conversely, if the test statistic falls below the critical value, there's insufficient evidence to reject the null hypothesis.

Critical values are specific to each hypothesis test and are determined based on the chosen significance level (alpha) and the distribution of the test statistic under the null hypothesis. They define the rejection region of the null hypothesis.

Understanding critical values is crucial in hypothesis testing as they help in making decisions about accepting or rejecting hypotheses.

**what is p value in dicky fuller test**

🗒️ Answer

In the Dickey-Fuller test, the p-value indicates the probability of observing the test statistic (or a more extreme value) under the null hypothesis that the time series data has a unit root, implying non-stationarity [[2](https://support.minitab.com/en-us/minitab/help-and-how-to/statistical-modeling/time-series/how-to/augmented-dickey-fuller-test/interpret-the-results/all-statistics-and-graphs/)]. A lower p-value suggests stronger evidence against the null hypothesis of non-stationarity, indicating that the data is likely stationary. Generally, if the p-value is less than a chosen significance level (such as 0.05), the null hypothesis is rejected, implying stationarity [[5](https://www.machinelearningplus.com/time-series/augmented-dickey-fuller-test/)]. Conversely, if the p-value is higher than the significance level, there is insufficient evidence to reject the null hypothesis, suggesting non-stationarity [[6](https://support.minitab.com/en-us/minitab/help-and-how-to/statistical-modeling/time-series/how-to/augmented-dickey-fuller-test/interpret-the-results/key-results/)].

**Partial Autocorrelation Function (PACF) and Autocorrelation Function (ACF) are crucial tools in ARIMA modeling.**

ACF (Autocorrelation Function): ACF measures the correlation between a time series and its lagged values. It helps identify the order of the Moving Average (MA) component in ARIMA. If ACF shows a sharp cutoff after a certain lag, it suggests a non-seasonal MA term.

PACF (Partial Autocorrelation Function): PACF gives the correlation between a time series and its lagged values, removing the effects of intermediate lags. It assists in determining the order of the Autoregressive (AR) component in ARIMA. A sharp cutoff in PACF after a particular lag indicates a non-seasonal AR term [[4](https://towardsdatascience.com/interpreting-acf-and-pacf-plots-for-time-series-forecasting-af0d6db4061c)].

**Summary**

Detailed Explanation of SARIMAX Model Summary

Model Information:

Dependent Variable: The variable being modeled, often denoted as 'y'.

No. Observations: The number of observations used in the model, which is 4686 in this case.

Model: Specifies the type of model used, which is SARIMAX (Seasonal AutoRegressive Integrated Moving Average with eXogenous regressors) with parameters (p=3, d=1, q=2), indicating the order of autoregressive, differencing, and moving average terms, respectively.

Date and Time: Indicates when the model was run.

Parameter Estimates:

Coefficients (coef): Estimated coefficients for each parameter in the model.

Standard Errors (std err): Standard errors associated with the parameter estimates, indicating the uncertainty in the estimates.

Z-Scores (z): Z-scores for each coefficient, indicating the number of standard deviations away from the mean.

P-Values (P>|z|): P-values associated with the coefficients, used to test the null hypothesis that the coefficient is zero.

Information Criteria:

Log Likelihood: The log-likelihood of the model, a measure of how well the model fits the data.

AIC (Akaike Information Criterion): A metric used for model selection, balancing goodness of fit and model complexity. Lower values indicate a better fit.

BIC (Bayesian Information Criterion): Similar to AIC but penalizes models with more parameters more heavily.

Residuals and Diagnostic Tests:

Residual Standard Error (sigma2): The estimated variance of the residuals.

Ljung-Box (L1): A test for the absence of serial correlation in the residuals.

Jarque-Bera (JB): A test for normality of the residuals.

Heteroskedasticity (H): A test for the presence of heteroskedasticity (non-constant variance) in the residuals.

**Explanation of MAPE (Mean Absolute Percentage Error)**

🗒️ Answer

MAPE, or Mean Absolute Percentage Error, is a metric used to evaluate the accuracy of a forecasting model. It measures the average absolute percentage difference between the predicted values and the actual values. The formula for MAPE is:

MAPE=1𝑛∑𝑖=1𝑛∣𝐴𝑖−𝐹𝑖𝐴𝑖∣×100MAPE=n1​∑i=1n​∣∣​Ai​Ai​−Fi​​∣∣​×100

Where:

𝐴𝑖Ai​ is the actual value.

𝐹𝑖Fi​ is the forecasted or predicted value.

𝑛n is the total number of observations.

MAPE provides a percentage error for each observation, and averaging these errors gives an overall indication of the forecasting accuracy. A lower MAPE indicates better accuracy, with 0 indicating perfect predictions.