Local flight

This time series analysis focuses on local flight volume data, employing various methods including time series decomposition, stationarity tests, SARIMAX model fitting, and performance evaluation. Below is a detailed analysis of the results and potential improvements.

**1. Time Series Decomposition**

The original data was decomposed into trend, seasonal, and residual components.

* **Trend Component**: The data initially exhibited an upward trend, followed by stabilization between time points 30 and 40, likely due to supply constraints or market saturation.
* **Seasonal Component**: The seasonal component revealed periodic fluctuations in flight volume, such as peaks between time points 10 and 20, possibly related to holiday travel demand.
* **Residual Component**: Significant residual deviations were observed between time points 20 and 30, suggesting that the model did not fully capture the data dynamics, potentially indicating unconsidered exogenous variables or nonlinear characteristics.

**2. Stationarity Test**

**ADF** and **KPSS** tests indicated that the original data was non-stationary, requiring differencing. After first-order differencing, stationarity improved significantly, though residuals still exhibited some structural bias, suggesting the consideration of higher-order differencing or other transformation methods.

**3. SARIMAX Model Evaluation**

The **SARIMAX** model was used to fit the data, employing an ARIMA(1,1,1) structure for feature capture.

* **Autoregressive Term (AR)**: The high p-value suggests that the AR term is not significant. It may be advisable to simplify the model by removing this component to reduce complexity and prevent overfitting.
* **Moving Average Term (MA)**: The MA term had a significant p-value, indicating its effectiveness in reducing short-term noise and improving prediction reliability.
* **Model Selection Metrics (AIC and BIC)**: The AIC and BIC values were relatively high, suggesting that parameter optimization via grid search could further improve model fitting and prediction performance.

**4. Model Prediction Performance Analysis**

**4.1 Comparison of Prediction Results with Actual Data**

**Test set performance was suboptimal**, particularly between time points 50 and 60, where the model failed to capture rapid data fluctuations.

* **Prediction Curve Smoothness**: To improve responsiveness to short-term fluctuations, it may be worthwhile to increase model flexibility by incorporating additional MA or AR terms or including exogenous variables.
* **Impact of External Variables**: Flight volume is influenced by factors such as fuel prices, weather, and policy changes. Including these factors to construct a multivariate time series model could enhance predictive accuracy.

**4.2 Directions for Improving Prediction Performance**

* **More Complex Models**: Consider employing higher-order ARIMA or nonlinear models such as LSTM to address high volatility.
* **Feature Engineering**: Extract additional features related to seasonality and external variables to enhance prediction accuracy.

**5. Residual Analysis**

**5.1 Residual Distribution**

Ideally, residuals should be randomly distributed with a mean close to zero. However, at specific time points (e.g., 55 and 65), residuals significantly deviated from zero.

**5.2 Autocorrelation Analysis**

* **ACF and PACF**: Significant autocorrelations were observed at multiple lags, suggesting the addition of lag terms or the inclusion of more explanatory variables.

Monthly death

This analysis focuses on the time series of monthly death counts, using methods such as time series decomposition, stationarity testing, SARIMAX model fitting, and prediction performance evaluation. The following sections present a concise evaluation of the results and potential improvements.

**1. Time Series Decomposition**

The monthly death data was decomposed into trend, seasonal, and residual components.

* **Trend Component**: The trend shows an increase in deaths, peaking around time point 30, then declining beyond time point 40, reflecting the epidemic's progression and possible interventions.
* **Seasonal Component**: Seasonal fluctuations indicate periods with consistently higher death counts, likely linked to seasonal factors like increased winter vulnerabilities.
* **Residual Component**: Significant residual fluctuations suggest exogenous factors or noise not captured by the trend and seasonal components, particularly between time points 20 to 30.

**2. Stationarity Test**

Stationarity was tested using **ADF** and **KPSS** tests:

* **ADF Test**: The ADF test indicates non-stationarity.
* **KPSS Test**: The KPSS test supports this. First-order differencing improved stationarity, though higher-order differencing or alternative transformations might help further.

**3. SARIMAX Model Evaluation**

A **SARIMAX** model (ARIMA(1,1,1)) was fitted:

* **Autoregressive Term (AR)**: The AR term's high p-value suggests it may not be significant, and removing it could reduce overfitting.
* **Moving Average Term (MA)**: The MA term is significant and helps capture short-term irregularities.
* **Model Selection Metrics (AIC and BIC)**: The high AIC and BIC values suggest room for optimization, possibly through grid search.

**4. Model Prediction Performance Analysis**

**4.1 Comparison of Prediction Results with Actual Data**

The **model's predictive performance is suboptimal**, particularly during the testing phase, where predictions are too flat and fail to capture actual volatility.

* **Prediction Curve Smoothness**: The lack of sensitivity to fluctuations suggests adding more AR or MA terms or including exogenous regressors.
* **External Variables Impact**: Including variables like healthcare capacity and policy interventions could improve accuracy, especially during extreme changes.

**4.2 Directions for Improving Prediction Performance**

* **Model Complexity**: Advanced models like higher-order ARIMA or LSTM could capture complex temporal patterns better.
* **Feature Engineering**: Include features such as temperature, healthcare utilization, and population movement.

**5. Residual Analysis**

**5.1 Residual Distribution**

Residuals show systematic deviations between time points 54 and 64, indicating that the model has not fully captured underlying patterns.

**5.2 Autocorrelation Analysis**

* **ACF and PACF**: Significant autocorrelations suggest adding lagged variables or using higher-order differencing.