SKEWNESS

1. Definition: Skewness is a statistical measure that quantifies the degree of asymmetry of a distribution of a real-valued random variable about its mean. It can be positive, zero, negative, or undefined. Skewness indicates whether the data is concentrated more on one side of the mean than the other.

The formula for calculating skewness, known as the Fisher-Pearson coefficient of skewness, is written as:

Skewness = 3(μ−ν)

σ

Where: μ is the mean, ν is the median and σ is the standard deviation

2. Description: Skewness helps us understand how a data is distributed around the mean, indicating whether the distribution is symmetric or skewed to one side. A distribution with zero skewness is known as symmetrical distribution, while positive skewness indicates a longer tail on the right side and negative skewness indicates a longer tail on the left side of the distribution

3. Demonstration: Let us consider a dataset of daily stock returns for two groups of stocks, A and B, with the following return values:

Stock A Returns: [2, 5, 8, 10, 12, 15, 18, 20, 50]

Stock B Returns: [95, 88, 85, 80, 78, 75, 72, 68, 45, 22]

To calculate the skewness for each set of stock returns:

Calculate Mean (μ):

Stock A: Mean = (2 + 5 + 8 + 10 + 12 + 15 + 18 + 20 + 50) / 9 = 15.6

Stock B: Mean = (95+88+85+80+78+75+72+68+45+22) / 10 = 70.8

Calculate Median (ν):

Stock A: Median = 12

Stock B: Median = 76.5

Calculate Standard Deviation (σ):

Stock A: Standard Deviation = √[Σ(xi - μ)² / (n-1)] ≈ 13.368

Stock B: Standard Deviation = √[Σ(xi - μ)² / (n-1)] ≈20.721

Calculate Skewness:

Stock A: Skewness = 1.7

Stock B: Skewness = -1.23

The calculated skewness values for Stock A and Stock B indicate the degree of asymmetry in their return distributions:

Stock A has a positive skewness (1.7), indicating a longer tail on the right side.

Stock B has a negative skewness (-1.23), indicating a longer tail on the left side.

4. Diagram: Below is the frequency polygon graphs that describe both positive and negative skewness.

import matplotlib.pyplot as plt

import seaborn as sns

# Data for Stock A Returns (positively skewed)

stock\_a\_returns = [2, 5, 8, 10, 12, 15, 18, 20, 50]

# Data for Stock B Returns (negatively skewed)

stock\_b\_returns = [95, 88, 85, 80, 78, 75, 72, 68, 45, 22]

# Create subplots for visualization

fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(12, 6))

# Distribution plot for Stock A Returns (positively skewed)

sns.distplot(stock\_a\_returns, ax=ax1, kde=True)

ax1.set\_title('Stock A Returns (Positively Skewed)')

# Distribution plot for Stock B Returns (negatively skewed)

sns.distplot(stock\_b\_returns, ax=ax2, kde=True)

ax2.set\_title('Stock B Returns (Negatively Skewed)')

# Rotate x-axis labels for better readability

plt.setp(ax1.xaxis.get\_majorticklabels(), rotation=45)

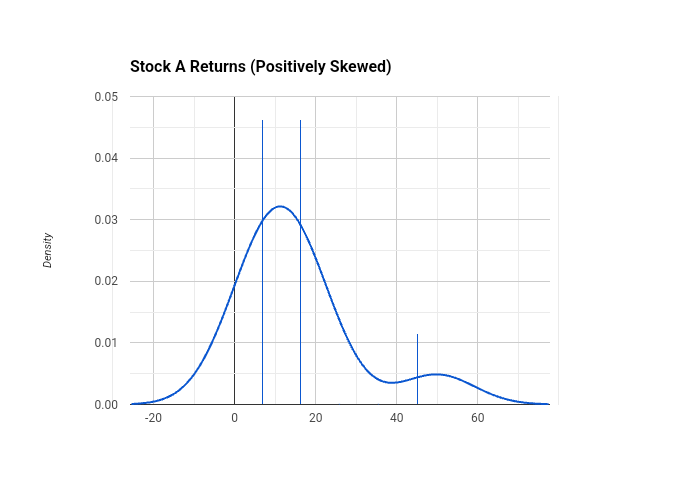
plt.setp(ax2.xaxis.get\_majorticklabels(), rotation=45)

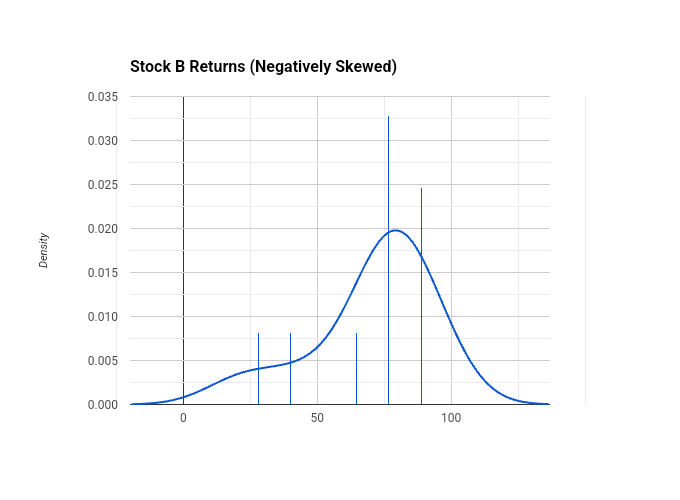
# Tight layout for better spacing

plt.tight\_layout()

# Display the visualizations

plt.show()





5. Diagnosis: Here are some methods to diagnose the presence of skewness:

Visual Inspection: Plotting a histogram or density plot of the data can visually reveal asymmetry in the distribution. Skewed data will show a longer tail on one side compared to the other.

Skewness Coefficient: Calculating the skewness coefficient using Fisher-Pearson coefficient of skewness formula can provide a numerical measure of asymmetry. A non-zero value indicates skewness.

Statistical Tests: Conducting formal statistical tests for skewness, like tests for skewness, kurtosis, and normality, can help confirm the presence of skewness in the data.

Comparison with Normal Distribution: Comparing the distribution of the data with a normal distribution can also help identify skewness. A symmetric distribution should resemble a normal distribution, while skewed data will deviate from it.

6. Damage: Skewed data can have detrimental effects on predictive models like artificial neural networks and deep learning models, impacting their accuracy and performance. The presence of skewness in data can lead to the following:

Reduced Prediction Accuracy: Skewed data can affect the training process of machine learning models, leading to reduced prediction accuracy and reliability

Misclassification Rates: Skewed data, especially imbalanced data where certain classes dominate, can result in high misclassification rates for minority classes. This can lead to poor prediction capability and biased outcomes in classification models.

Model Instability: Skewness can introduce instability in predictive models, making them less robust and reliable in handling real-world data with varying distributions.

Inaccurate Insights: Skewed data may distort the insights gained from the analysis, potentially leading to incorrect conclusions and decisions based on biased or incomplete information.

Challenges in Model Interpretation: Skewness can complicate the interpretation of model results, making it harder to understand the relationships between variables and the impact of different factors on outcomes.

Directions: To effectively address skewness and improve the performance of predictive models, several approaches and models can be considered:

Regression Models: Regression models are suitable for predicting specific numerical values in time series data. They can handle skewed data to some extent but may be influenced by significant skewness, potentially impacting prediction accuracy

Transformation Techniques: Utilize transformation methods like Log, Square Root, Box-Cox, Yeo-Johnson, and Quantile Transformations to mitigate positive skewness in the data and make it more suitable for modeling

Time Series Analysis: Time series analysis techniques, such as autoregressive integrated moving average (ARIMA) models or seasonal decomposition of time series (STL), can help capture patterns and trends in time series data while addressing skewness

Non-linear Models: Non-linear regression models like Tobit regression for skewed data can be considered to account for the distributional characteristics of the target variable and handle outliers effectively.

7. Conclusion

By incorporating these suggested models and techniques into your analysis of skewed time series data, you can enhance the robustness and accuracy of your predictive models, leading to more reliable insights and decision-making processes in financial forecasting or other applications requiring accurate modeling of time-dependent data.