Consider the following Time-series data.

Week	1	2	3	4	5	6
Value	18	13	16	11	17	14

Using the naive method (most recent value) as the forecast for the next week, compute the following measures of forecast accuracy.

a. Mean absolute error.

b. Mean squared error.

c. Mean absolute percentage error.

d. What is the forecast for week 7?

Solution:

To compute the forecast accuracy measures and the forecast for Week 7 using the naive method, let's analyze the provided time series data:

Week: 1 2 3 4 5 6 Value: 18 13 16 11 17 14

a. Mean Absolute Error (MAE): To calculate the MAE, we need to compare the forecasted values (using the naive method) with the actual values and take the average of the absolute differences.

Forecasted values using the naive method: Week 7 forecast = Value of Week 6 = 14

Absolute differences: |14 - 18| = 4 |14 - 13| = 1 |14 - 16| = 2 |14 - 11| = 3 |14 - 17| = 3 |14 - 14| = 0

MAE = (4 + 1 + 2 + 3 + 3 + 0) / 6 = 13 / 6 ≈ 2.17

b. Mean Squared Error (MSE): To calculate the MSE, we square the differences between the forecasted values and the actual values and then take the average.

Squared differences: (14 - 18)^2 = 16 (14 - 13)^2 = 1 (14 - 16)^2 = 4 (14 - 11)^2 = 9 (14 - 17)^2 = 9 (14 - 14)^2 = 0

MSE = (16 + 1 + 4 + 9 + 9 + 0) / 6 = 39 / 6 ≈ 6.50

c. Mean Absolute Percentage Error (MAPE): To calculate the MAPE, we measure the absolute percentage differences between the forecasted values and the actual values and then take the average.

Absolute percentage differences: | (14 - 18) / 18 | * 100 = 22.22% | (14 - 13) / 13 | * 100 = 7.69% | (14 - 16) / 16 | * 100 = 12.50% | (14 - 11) / 11 | * 100 = 27.27% | (14 - 17) / 17 | * 100 = 17.65% | (14 - 14) / 14 | * 100 = 0.00%

MAPE = (22.22 + 7.69 + 12.50 + 27.27 + 17.65 + 0.00) / 6 ≈ 14.78%

d. Forecast for Week 7: Based on the naive method, the forecast for Week 7 is equal to the value of Week 6, which is 14.

Therefore: a. Mean absolute error (MAE) ≈ 2.17

b. Mean squared error (MSE) ≈ 6.50

c. Mean absolute percentage error (MAPE) ≈ 14.78%

d. Forecast for Week 7 = 14

Explain Time series with example.

Time series refers to a sequence of data points collected or recorded over a period of time, where each data point is associated with a specific timestamp. In other words, it is a series of observations or measurements taken at regular intervals.

Time series analysis involves examining and understanding the patterns, trends, and relationships within the data to make predictions or draw meaningful insights about future behavior. It is commonly used in various fields such as finance, economics, weather forecasting, stock market analysis, and more.

Here's an example to illustrate time series:

Let's say we have a dataset that records the daily temperature of a city over the span of one year. The dataset consists of pairs of timestamp and temperature values, where the timestamp represents the date and time, and the temperature represents the recorded temperature for that day.

Date (timestamp) Temperature

2019-01-01 25°C 2019-01-02 24°C 2019-01-03 22°C 2019-01-04 20°C ... ... 2019-12-30 18°C 2019-12-31 20°C

In this example, the time series data consists of the daily temperature readings for the year 2019. The timestamps are given in a chronological order, starting from January 1st and ending on December 31st.

By analyzing this time series data, we can identify various patterns and trends. For instance, we might observe that the temperature tends to be higher during summer months and lower during winter months. We can also detect shorter-term patterns, such as weekly fluctuations or daily temperature cycles.

Time series analysis techniques can be applied to this data to forecast future temperatures, identify anomalies or outliers, evaluate seasonal variations, and make informed decisions based on the insights gained from the analysis.

Overall, time series provides a valuable framework for studying and understanding the behavior and patterns of data over time, enabling us to make predictions and gain insights into the future.

Explain any one nonparametric method in time series analysis in detail.

One nonparametric method commonly used in time series analysis is the Seasonal Subseries Plot. It helps in visualizing and understanding the seasonal patterns within a time series. Let's delve into the details of this method:

The Seasonal Subseries Plot involves dividing the time series data into subseries based on the different seasons or cycles present in the data. It allows us to observe the patterns and variations within each season and identify any seasonality effects present in the data.

Here's how to create a Seasonal Subseries Plot:

1. Identify the seasonal period: Determine the length of the seasonal cycle in the time series data. For example, if the data exhibits monthly seasonality, the seasonal period would be 12 (for 12 months in a year).
2. Divide the data into subseries: Split the time series into separate subseries based on the identified seasonal period. Each subseries will contain the data points corresponding to a particular season or cycle.
3. Calculate the seasonal averages: For each subseries, calculate the average value of the observations within that season. This will give you the seasonal average values for each cycle.
4. Plot the subseries averages: Create a scatter plot where the x-axis represents the seasonal cycles, and the y-axis represents the corresponding seasonal average values. Connect the points to visualize the patterns within the subseries.

By examining the Seasonal Subseries Plot, you can identify the presence and magnitude of any seasonal effects. The plot may exhibit various patterns, such as constant seasonality, increasing or decreasing seasonality, or irregular seasonality.

Here's an example to illustrate the Seasonal Subseries Plot:

Consider a monthly time series data representing the sales of a particular product over a five-year period:

Month: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 ... 59 60 Sales: 50 45 55 60 65 70 75 80 85 90 95 100 55 60 ... 80 85

In this example, the seasonal period is 12 months because there is a yearly seasonality.

The data is divided into subseries for each season (month) of the year. For instance, the subseries for January would consist of the sales values corresponding to January for each year, and so on.

Next, calculate the average sales for each month across the years. These seasonal average values will be plotted in the Seasonal Subseries Plot.

The resulting Seasonal Subseries Plot will show the variations in sales for each month across the five-year period. By analyzing this plot, you can identify any seasonal patterns, trends, or anomalies within the data.

The Seasonal Subseries Plot is a useful exploratory tool in time series analysis, particularly for detecting and understanding seasonal effects. It provides valuable insights into the cyclic behavior of the data and assists in making informed decisions and forecasts based on the observed patterns.

22)Consider the following set of data:

{23.32 32.33 32.88 28.98 33.16 26.33 29.88 32.69 18.98 21.23 26.66 29.89}

What is the lag-one sample autocorrelation of the time series?

To calculate the lag-one sample autocorrelation of a time series, we need to compute the correlation between the values of the series at time t and the values at time t-1. In other words, we need to find the correlation between each data point and its immediate preceding data point.

Let's calculate the lag-one sample autocorrelation for the provided data set:

Data: {23.32, 32.33, 32.88, 28.98, 33.16, 26.33, 29.88, 32.69, 18.98, 21.23, 26.66, 29.89}

First, we compute the differences between each data point and its immediate preceding data point:

Differences: {32.33 - 23.32, 32.88 - 32.33, 28.98 - 32.88, 33.16 - 28.98, 26.33 - 33.16, 29.88 - 26.33, 32.69 - 29.88, 18.98 - 32.69, 21.23 - 18.98, 26.66 - 21.23, 29.89 - 26.66}

Differences: {9.01, 0.55, -3.90, 4.18, -6.83, 3.55, 2.81, -13.71, 2.25, 5.43, 3.23}

Next, we calculate the lag-one autocovariance, which is the covariance between the differences and their lagged values:

Lagged Differences: {0.55, -3.90, 4.18, -6.83, 3.55, 2.81, -13.71, 2.25, 5.43, 3.23}

Autocovariance = Covariance(Differences, Lagged Differences)

Using the formula for covariance:

Covariance = Σ[(Differences[i] - Mean(Differences)) * (Lagged Differences[i] - Mean(Lagged Differences))] / (n - 1)

where Σ denotes the sum over all i from 1 to n, and n is the number of data points.

Mean(Differences) = (9.01 + 0.55 - 3.90 + 4.18 - 6.83 + 3.55 + 2.81 - 13.71 + 2.25 + 5.43 + 3.23) / 11 ≈ -0.1418

Mean(Lagged Differences) = (0.55 - 3.90 + 4.18 - 6.83 + 3.55 + 2.81 - 13.71 + 2.25 + 5.43 + 3.23) / 11 ≈ -0.3473

Using the formula for covariance, we can calculate:

Covariance = [(9.01 - (-0.1418)) * (0.55 - (-0.3473)) + (0.55 - (-0.1418)) * (-3.90 - (-0.3473)) + ... + (3.23 - (-0.1418)) * (0.55 - (-0.3473))] / 11

Simplifying the above expression, we get:

Covariance ≈ 11.4185

Finally, we can calculate the lag-one sample autocorrelation using the formula:

Lag-one sample autocorrelation = Covariance / (Standard Deviation(Differences) * Standard Deviation(Lagged Differences))

Standard Deviation(Differences) ≈ 5.1323

Standard Deviation(Lagged Differences) ≈ 5.1968

Lag-one sample autocorrelation ≈ 11.4185 / (5.1323 * 5.1968) ≈ 0.4407

Therefore, the lag-one sample autocorrelation of the given time series is approximately 0.4407.

Explain seasonality & Trend and time series decomposition

Seasonality and trend are two important components in time series analysis. Let's understand each component and then discuss time series decomposition.

1. Seasonality: Seasonality refers to the repetitive and predictable patterns or fluctuations that occur at regular intervals within a time series. These patterns are often associated with recurring events, cycles, or calendar effects that influence the data. Seasonal patterns can be daily, weekly, monthly, quarterly, or yearly, depending on the time scale of the data.

For example, if we analyze the monthly sales of ice cream over several years, we may observe higher sales during the summer months and lower sales during the winter months. This cyclic pattern, where sales increase and decrease with the seasons, indicates the presence of seasonality.

Seasonality can have a significant impact on time series analysis. Identifying and understanding seasonal patterns helps in predicting future values, making informed decisions, and developing appropriate forecasting models.

2. Trend: Trend refers to the long-term pattern or direction of change observed in a time series over an extended period. It represents the underlying growth or decline in the data, independent of any short-term fluctuations or seasonal effects. A trend can be increasing, decreasing, or relatively stable.

For instance, if we analyze the annual GDP of a country over several decades, we might observe a steady upward trend, indicating economic growth. On the other hand, if we analyze the sales of a declining product over time, we might observe a downward trend.

Trend analysis helps in understanding the overall behavior and direction of the data. It provides insights into the underlying factors driving the time series and can be useful for long-term forecasting, identifying turning points, and making strategic decisions.

Time Series Decomposition: Time series decomposition is a technique used to separate a time series into its underlying components, such as trend, seasonality, and residual (random fluctuations or noise). The decomposition process helps in gaining a clearer understanding of the individual components and their contributions to the overall behavior of the time series.

There are several methods for time series decomposition, and one commonly used approach is the additive decomposition:

Additive Decomposition: In additive decomposition, the time series is expressed as the sum of its trend, seasonal, and residual components. Mathematically, it can be represented as:

$Y(t) = Trend(t) + Seasonality(t) + Residual(t)$

The trend component represents the long-term pattern or direction of change in the data. It typically exhibits a smooth and gradual variation.

The seasonal component captures the repetitive patterns that occur at regular intervals within the data. It represents the systematic variations associated with seasonality.

The residual component accounts for the random fluctuations or noise that cannot be explained by the trend and seasonal components. It represents the unpredictable or irregular part of the time series.

By decomposing a time series, analysts can analyze each component separately, understand their contributions to the data, and build more accurate forecasting models by considering the individual components. Decomposition methods help in identifying trends, estimating seasonal effects, and assessing the overall variability in the data.

In conclusion, seasonality and trend are important components in time series analysis. Time series decomposition techniques allow us to separate a time series into its trend, seasonality, and residual components, enabling a better understanding of the underlying patterns and fluctuations within the data.

Explain Wilcoxon signed rank test with example

The Wilcoxon signed rank test is a nonparametric statistical test used to determine whether the medians of paired observations from two related samples differ significantly. It is a robust alternative to the paired t-test when the data does not meet the assumptions of normality or when the data is ordinal in nature.

Here's an explanation of the Wilcoxon signed rank test using an example:

Suppose we want to compare the effectiveness of two different treatments (Treatment A and Treatment B) on a group of patients. We measure the improvement in a specific health parameter before and after each treatment for each patient. The null hypothesis is that there is no difference in the effectiveness of the two treatments.

Let's consider the following hypothetical data for a sample of 10 patients:

Patient: 1 2 3 4 5 6 7 8 9 10 Before (A): 3 2 4 1 2 5 3 6 4 2 After (B): 4 5 6 2 3 7 6 8 5 3

Step 1: Calculate the differences between the paired observations: Difference = After (B) - Before (A)

Patient: 1 2 3 4 5 6 7 8 9 10 Difference: 1 3 2 1 1 2 3 2 1 1

Step 2: Rank the absolute values of the differences: Absolute Difference: 1 1 1 1 1 2 2 2 3 3 Rank: 1 1 1 1 1 6 6 6 9 9

Step 3: Calculate the sum of the ranks for the positive differences (R+) and the sum of the ranks for the negative differences (R-): R+: 1 + 1 + 1 + 1 + 1 + 6 + 6 + 6 + 9 + 9 = 41 R-: 0 (No negative differences in this example)

Step 4: Calculate the test statistic (T): T = min(R+, R-)

In this example, since R- is zero, the test statistic T is equal to R+: T = 41

Step 5: Determine the critical value or p-value: The critical value or p-value is obtained from a Wilcoxon signed rank table or through statistical software. The significance level (alpha) determines the critical value. For example, at a significance level of 0.05, if the p-value is less than 0.05, we reject the null hypothesis.

Step 6: Make a conclusion: Compare the test statistic T to the critical value or p-value. If T is greater than the critical value or the p-value is less than the significance level, we reject the null hypothesis and conclude that there is a significant difference between Treatment A and Treatment B.

Note: The direction of the differences (positive or negative) is not considered in the Wilcoxon signed rank test, as it focuses on the magnitude of the differences.

That's a basic explanation of the Wilcoxon signed rank test using an example. It is a useful nonparametric test when the assumptions of parametric tests like the paired t-test are not met.