Question R02.1

\*\*The code part that is the main work is in the attached .rmd file. Only the textual part is described and summarized here. \*\*

Given that we need to forecast the temperature for the year 2040, we will initially attempt to predict the target temperature using the methods provided in the workshop (R code Part 1).

Taking June 2040 as an example, the predictions given by the two models are 0.6240408 2.824960 respectively. It is not difficult to see that there is a big difference between the two.

My forecast uses the ARIMA method in packages("forecast"), This is an iterative, comprehensive forecasting method with three different variables for forecasting and smoothing respectively. I used it to directly forecast 25 years of data (R code Part 2).

Then we use Cross-Validation to test its stability, the average MSE is 0.194546, Compared with the previous model (workshop2.3), there is no obvious improvement (R code Part3).

Limitations:

Linearity Assumption: ARIMA assumes linearity in the relationships within time series data. It may not capture non-linear trends and relationships, potentially leading to poor performance when dealing with highly non-linear data. This is also the reason why it shows low accuracy.

Sensitivity: ARIMA models are sensitive to outliers and anomalies, which can lead to unstable model performance. Preprocessing for outlier handling and robustness is often necessary. There are usually more outliers in temperature changes, which in my impression is due to other reasons (such as strong climate changes in a short period of time and other factors such as El Niño, etc.)

Long-Term Forecasting: ARIMA models may not perform well in long-term forecasting, as their performance tends to degrade, particularly when capturing trends or seasonal patterns that are hard to model.

Advantages:

Good for Short-to-Medium-Term Forecasting: Most of the time we only need short-term climate data (for example, about 15 days), ARIMA models are often effective for short-to-medium-term forecasting tasks where data exhibit clear patterns and are not subject to drastic structural changes.

Model Selection: ARIMA models offer automated model selection through functions like auto.Arima I use, which can determine the best orders for the ARIMA model based on model fit and statistical criteria.

Adaptability: SARIMA models can handle a wide range of time series data, including data with trends, seasonality, and autocorrelation. They are flexible and adaptable to various data patterns.

standard forecasting methods:

If I were making this prediction, I would divide climate predictions into short-term and long-term, In the short term (such as the 15-day data that can be seen on the weather forecast), I will choose to use ARIMA to predict. Since it is analysed based on past data, it has greater stability in the short term.

For long-term forecasting, however, more complex models are needed. The SARIMA model set for seasons has greater flexibility and accuracy and can adjust the model more accurately, but the parameters require manual intervention, which means it is not fully automated. The forecast results showed obvious seasonal cyclical changes (R code part 4). It is still very rough, as sufficient time is necessary to fit a proper model.