Analysis of Weekly Appointment Data

Anandika Aggarwal, Avishek Paudel, Paramdeep Nijjer

STAT 370 (AS01): Time Series Analysis

MacEwan University

2024/12/08

I. Introduction

The MacEwan University Career Office provides the dataset for this project. The dataset contains 52 data points representing weekly appointment records from the MacEwan University career office, covering the period from September 2023 to September 2024. The data captures the number of Absent, Attended, Available, and Cancelled each week. The university career office is interested in analyzing trends in student engagement with their services, as indicated by attendance and cancellation counts. Specifically, they aim to understand:

<u>Attendance Trends:</u> The number of appointments successfully attended each week, providing insights into student demand and engagement.

<u>Cancellation Trends:</u> The number of appointments cancelled each week may indicate issues like scheduling conflicts or changing student priorities.

The dataset allows us to examine weekly patterns, assess whether attendance or cancellations follow seasonal trends, and identify any periods of significant changes in engagement. This analysis can inform strategic decisions to improve scheduling, resource allocation, and student support.

We excluded Absent and Available data in our analysis as the career office's primary concern is managing resources efficiently—such as allocating staff and scheduling appointments. Attendance and cancellations directly impact resource utilization, making them more relevant metrics than absences or availability. Cancellations are proactive decisions students make, whereas absences might reflect a failure to show up without notice. Cancellations provide more structured data, as they indicate students' intentions, while absences could be due to diverse, unclear reasons. Attendance and cancellation data often show patterns and trends over time, which can be analyzed to predict future behaviour. Absences and availability, on the other hand, may not provide as much predictive value since they are indirect measures. Attendance measures positive engagement with the career office, while cancellations can highlight areas for improvement. They provide a balanced view of student interaction, helping the office improve its outreach and operations.

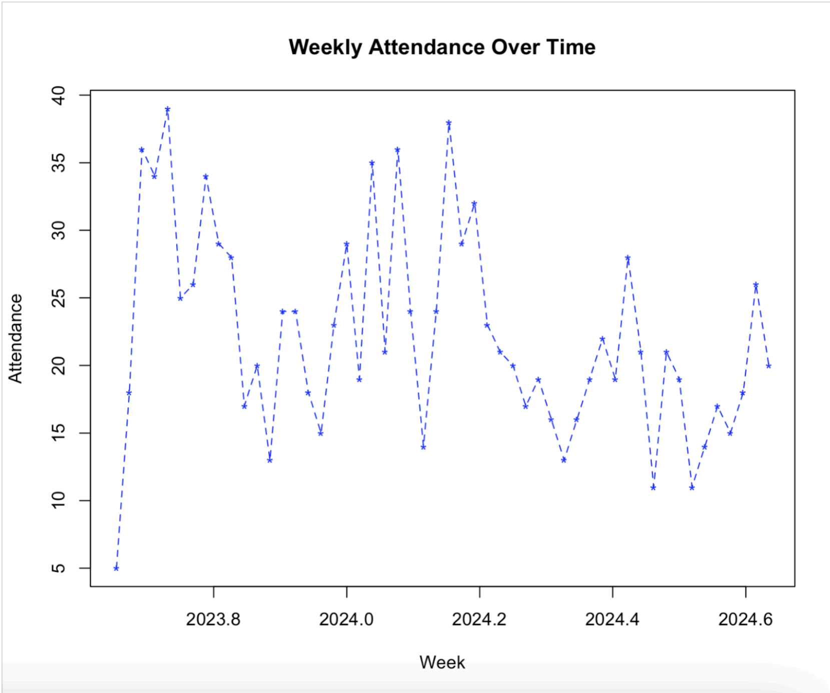
II. Data Cleaning

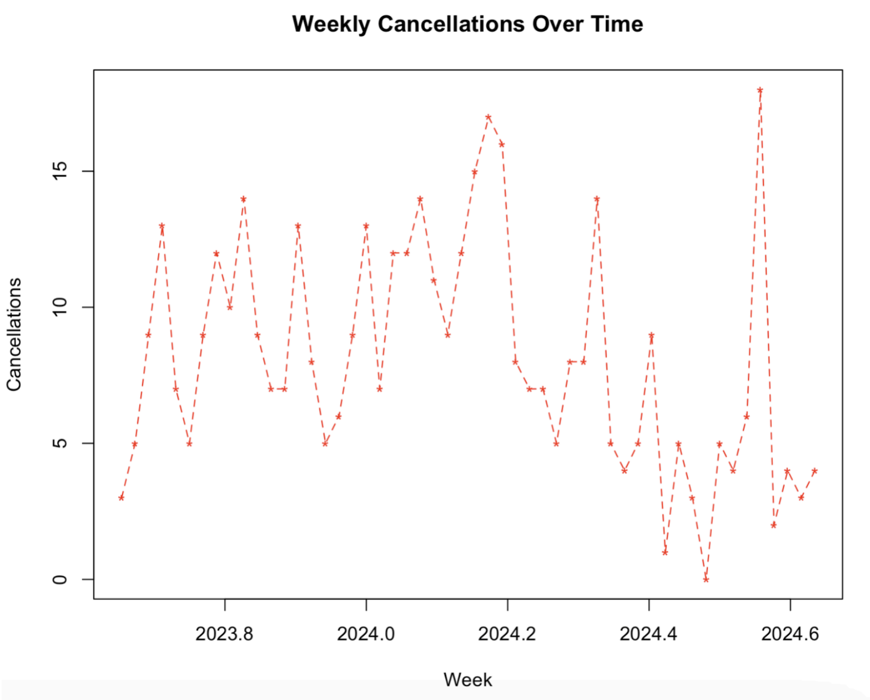
The dataset begins on September 1, 2023, which falls within Week 35 of the year. Week numbers were assigned based on this starting point and incremented sequentially through the year.

For each week (Monday to Friday), we aggregated the data related to attendance and cancellations. Any appointments scheduled during this time frame were grouped and labelled under the respective week number. We filtered out irrelevant data from the initial dataset, focusing specifically on attendance (students who showed up) and cancellations (appointments cancelled in advance). Data on absences (no-shows) and availability (total slots) was excluded to avoid noise and maintain focus on actionable trends. After grouping by week, the total number of attended and cancelled appointments was calculated each week. This summary allowed us to create a clean, structured dataset for further analysis.

III. Result

Plot 1: Line plot of Weekly Attendance and Cancellation over time





The line graphs for both cancellation and attendance provide a comprehensive view of appointment behaviour, highlighting weeks with high attendance and high cancellations.

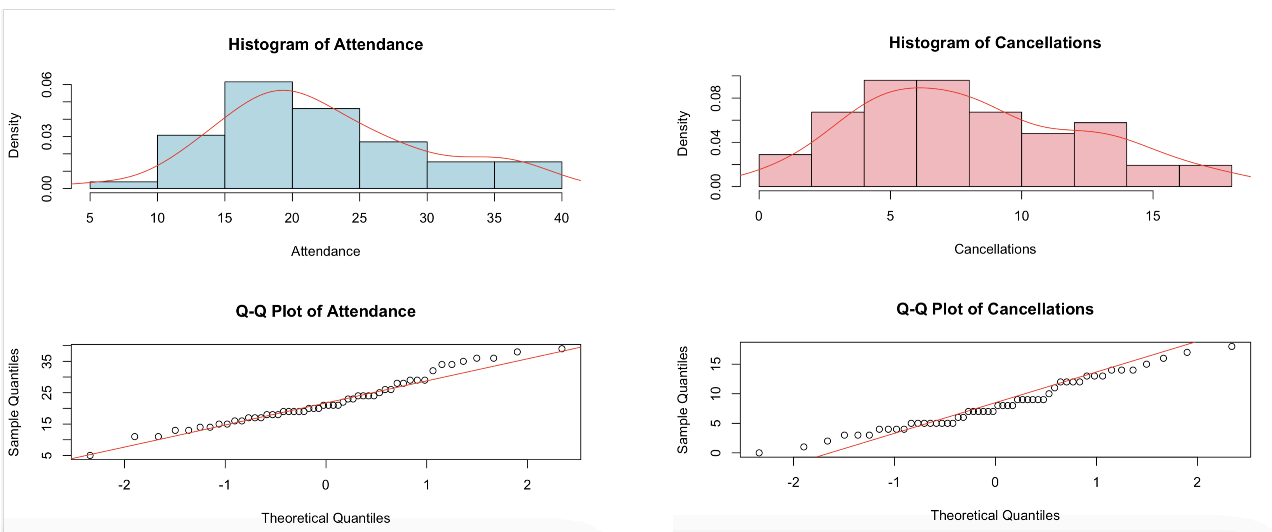
From the graphs, attendance peaked in August of 2023 and then again in the first month of 2024 and cancellation peaked on Mid and beginning of 2024.

Table 1: Numerical summary of weekly Attendance and Cancellation

Variables	Minimum	Maximum	Mean	Median	Standard Deviation
Attendance	5.00	39.00	22.21	21.00	7.60
Cancellation	0.00	18.00	8.154	7.500	4.27

Attendance has a higher mean and variability than cancellations. Cancellation data shows there are weeks with no cancellations (minimum = 0), while there is always attendance every week (minimum = 5).

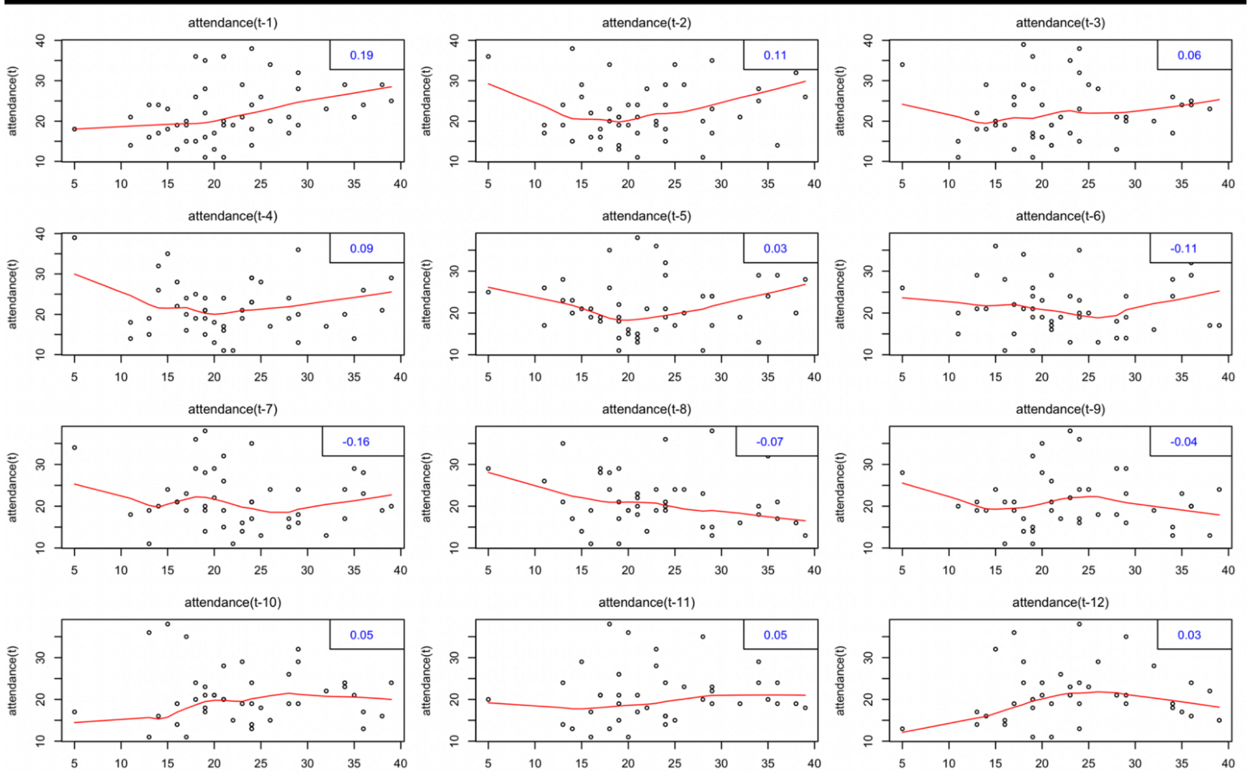
Plot 2: Histogram and Q-Q plot of attendance and Cancellation



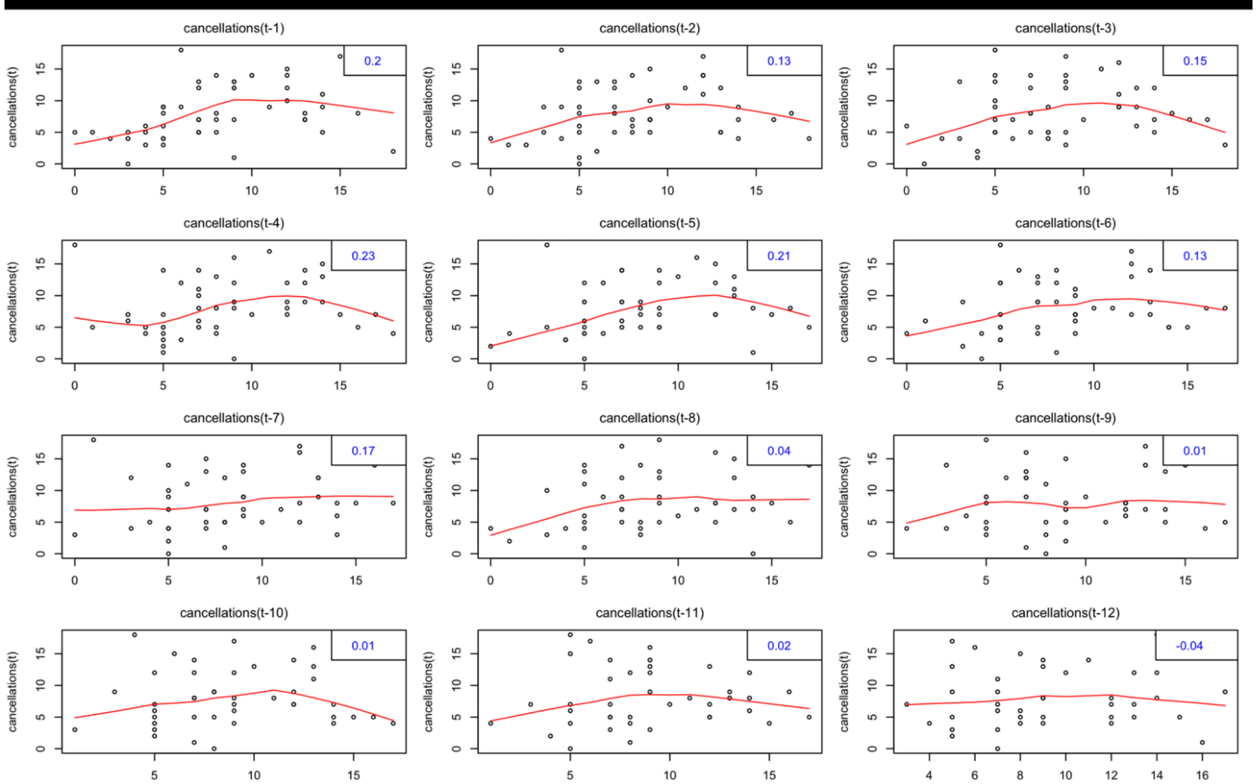
Both attendance and cancellations show a slight right skew, but attendance is closer to normal than cancellations. Q-Q plots confirm deviations from normality, especially in the tails. We will transform our cancellation data.

Plot 3: Lag Plot (1-12 lags) of Attendance and Cancellation

[redacted]



[redacted]

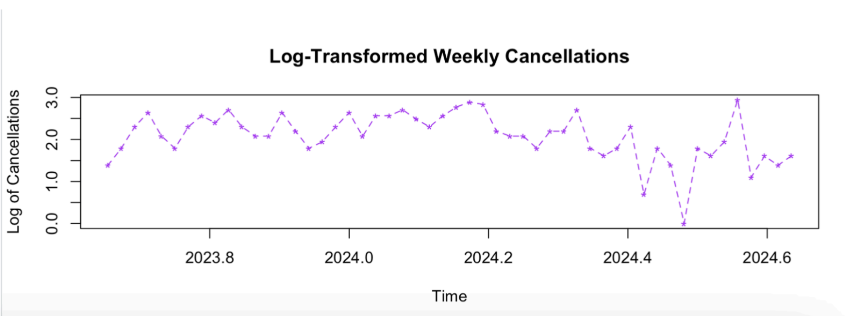


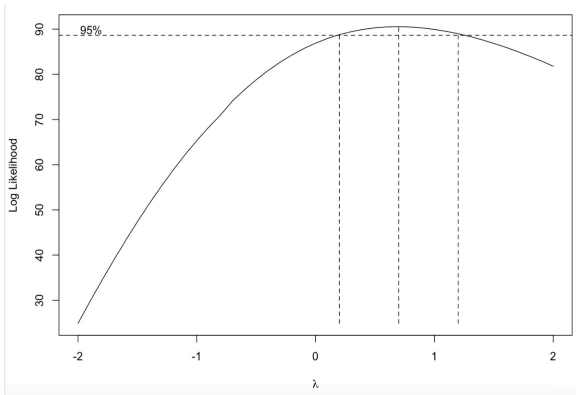
From the lag plot, both attendance and cancellation show weak autocorrelation, which suggests that it is only slightly influenced by values from previous weeks. As the lag increases, the correlation weakens and eventually becomes negligible.

The relatively higher correlation at lag 4 suggests some periodicity in the data, although it is very weak. We will now transform our cancellation data and analyze ACF and PACF to study periodic behaviour across other lags.

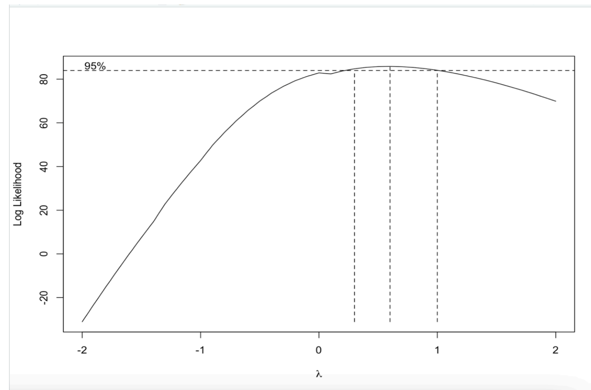
Below is the output of Box Cox:

```
> boxcox_attendance <- BoxCox.ar(attendance, method="yule-walker")
> boxcox_attendance$ci  # Confidence interval for lambda
[1] 0.2 1.2
> boxcox_attendance$mle  # Maximum likelihood estimate (MLE) for lambda
[1] 0.7
> # Apply Box-Cox transformation for Cancellations
> cancellations_shifted <- cancellations+1
> boxcox_cancellations <- BoxCox.ar(cancellations_shifted, method="yule-walker")
> boxcox_cancellations$ci  # Confidence interval for lambda
[1] 0.3 1.0
> boxcox_cancellations$mle  # Maximum likelihood estimate (MLE) for lambda
[1] 0.6
```





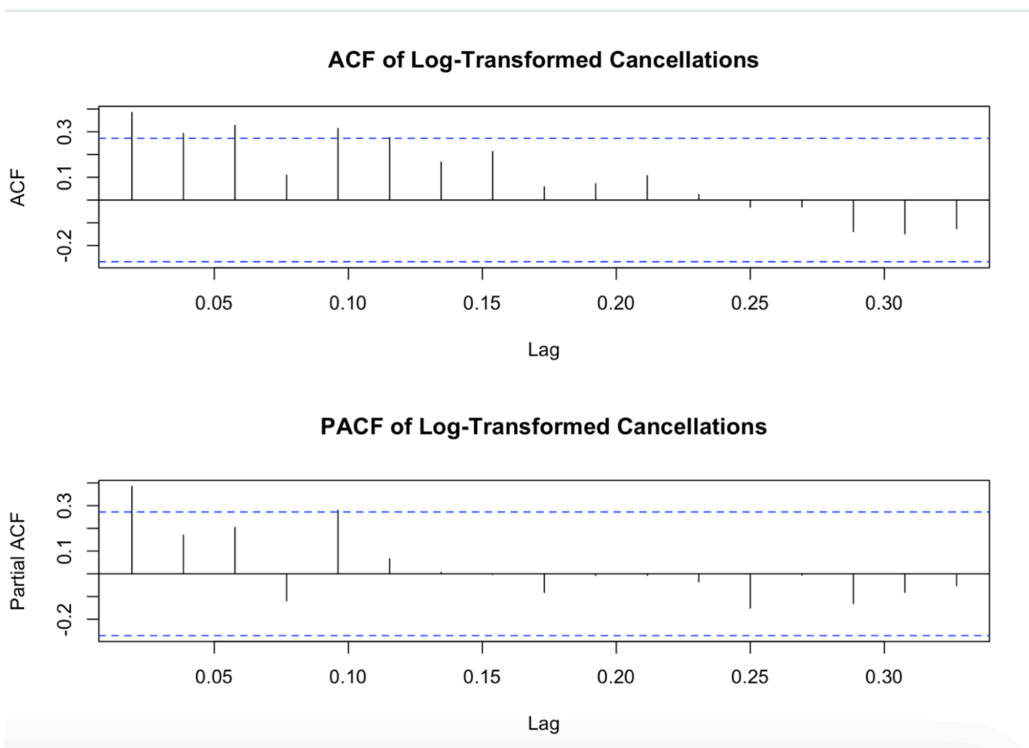
Attendance

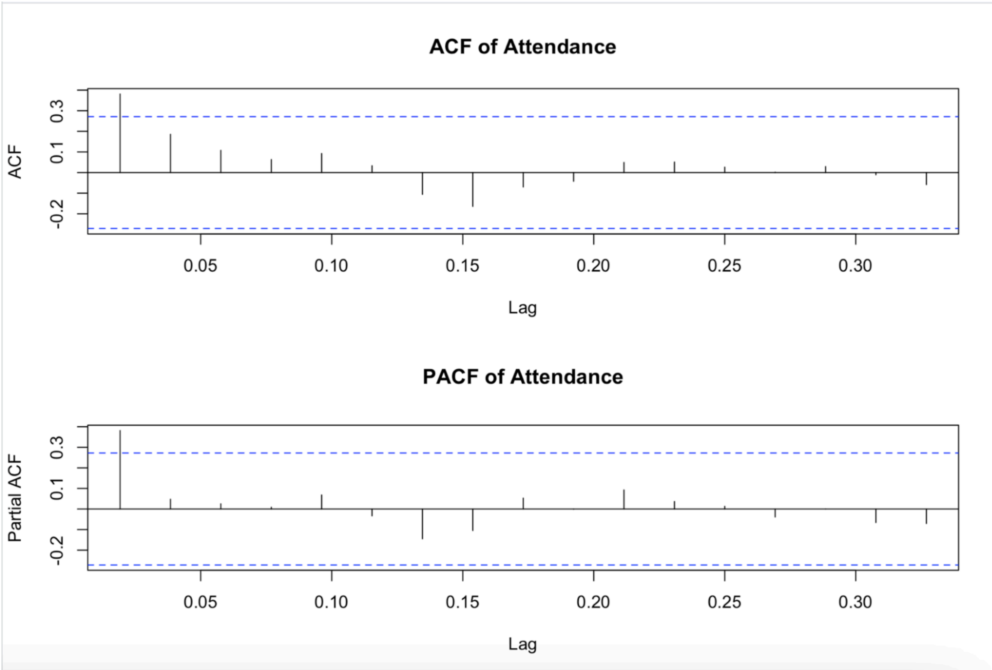


Cancellations

Further, We generate acf, pacf and eacf of attendance, and log cancellation data.

Plot 4: ACF and PACF of Attendance and Cancellation





Both ACF and PACF cuts off after lag 1 for attendance and

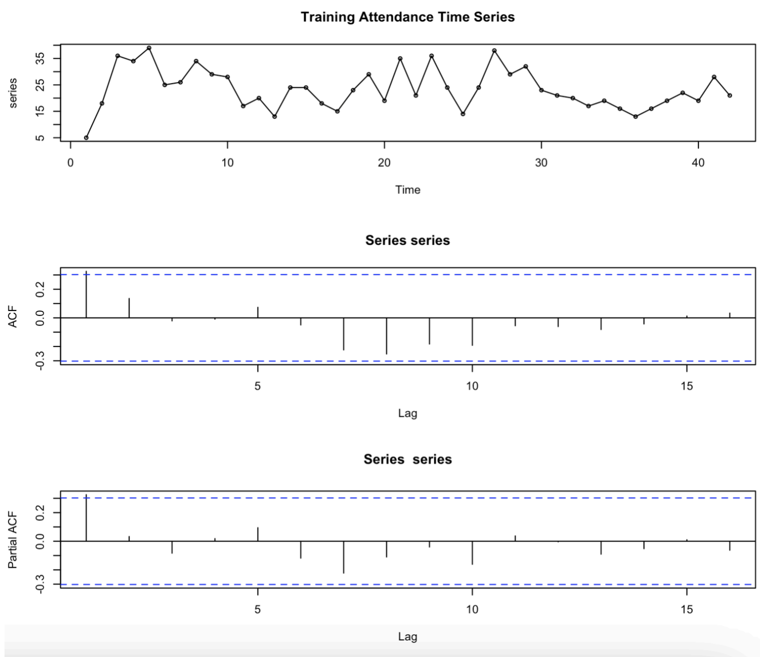
R output of eacf of attendance and Cancellation are below:

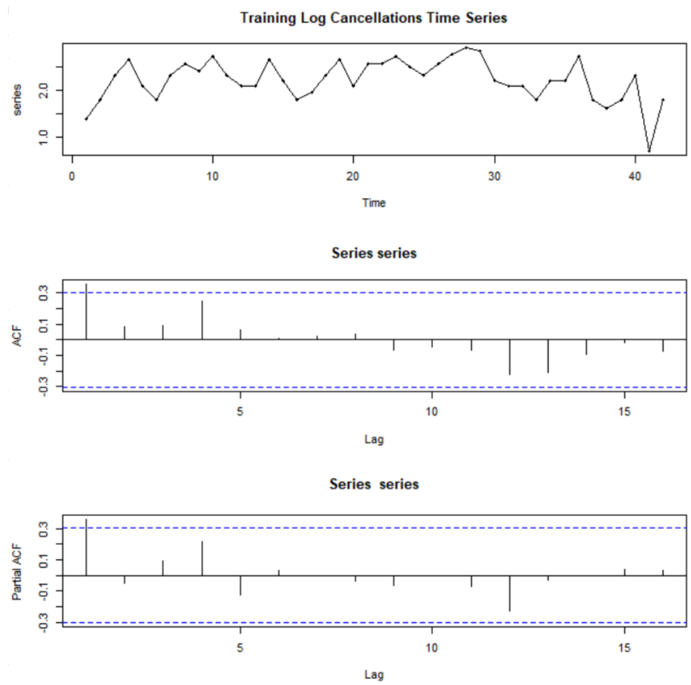
```
> eacf(attendance)
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 x o o o o o o o o o o  o  o  o
1 o o o o o o o o o o o  o  o  o
2 x o o o o o o o o o o  o  o  o
3 x o o o o o o o o o o  o  o  o
4 x x o o o o o o o o o  o  o  o
5 x o o o o o o o o o o  o  o  o
6 o o o o o o o o o o o  o  o  o
7 x o o x o o o o o o o  o  o  o
> eacf(log_cancellations)
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 x x x o x o o o o o o  o  o  o
1 x o x o o o o o o o o  o  o  o
2 x o o o o o o o o o o  o  o  o
3 x o o o o o o o o o o  o  o  o
4 x x x x o o o o o o o  o  o  o
5 o o o o o o o o o o o  o  o  o
6 o o o o o o o o o o o  o  o  o
7 o o o x o o o o o o o  o  o  o
```

The eacf suggested AR(1), MA(1), ARMA(1,1) model for Attendance and ARMA(1,1), ARMA(2,1), MA(3) for Cancellation.

Analysis on the Training set

Plot 5: ACF, PACF and plot of Training data set





ACF and PACF cut off after lag1 in Attendance and log cancellation of training data which suggests the AR (1) model. Also, the apid decay of ACF and the single spike in PACF at lag 1 indicate no long-term autocorrelation or seasonality in the data.

Below is the output of fitting different models in attendance:

```
> ar1_attendance

Call:
arima(x = train_attendance, order = c(1, 0, 0))

Coefficients:
         ar1  intercept
      0.3705    23.1211
s.e.  0.1533     1.7335

sigma^2 estimated as 50.85:  log likelihood = -142.18,  aic = 288.36
> ma1_attendance = arima(train_attendance, order=c(0, 0, 1))
> ma1_attendance

Call:
arima(x = train_attendance, order = c(0, 0, 1))

Coefficients:
         ma1  intercept
      0.2942    23.2661
s.e.  0.1305     1.4374

sigma^2 estimated as 52.29:  log likelihood = -142.73,  aic = 289.47
> arma11_attendance = arima(train_attendance, order=c(1, 0, 1))
> arma11_attendance

Call:
arima(x = train_attendance, order = c(1, 0, 1))

Coefficients:
         ar1      ma1  intercept
      0.4154  -0.0513    23.1052
s.e.  0.3169   0.3286     1.7712

sigma^2 estimated as 50.82:  log likelihood = -142.17,  aic = 290.33
```

Looking at the AIC (lower the better), Sigma squared (Variance), and Log Likelihood (higher the better), AR(1) and ARMA (1,1) models are the most suitable for the training attendance data. MA(1) performs worse and does not fit the data.

```
> ma3_cancellations

Call:
arima(x = train_log_cancellations, order = c(0, 0, 3))

Coefficients:
         ma1     ma2      ma3  intercept
      0.4856  0.1360  -0.1906     2.2138
s.e.  0.1553  0.1778   0.1796     0.0878

sigma^2 estimated as 0.1548:  log likelihood = -20.64,  aic = 49.27
> arma11_cancellations = arima(train_log_cancellations, order=c(1, 0, 1))
> arma11_cancellations

Call:
arima(x = train_log_cancellations, order = c(1, 0, 1))

Coefficients:
         ar1     ma1  intercept
      0.3524  0.0478     2.1963
s.e.  0.5631  0.5986     0.1028

sigma^2 estimated as 0.1599:  log likelihood = -21.19,  aic = 48.38
> arma21_cancellations = arima(train_log_cancellations, order=c(2, 0, 1))
> arma21_cancellations

Call:
arima(x = train_log_cancellations, order = c(2, 0, 1))

Coefficients:
         ar1     ar2     ma1  intercept
      0.2599  0.0367  0.1404     2.1961
s.e.  1.5473  0.6376  1.5397     0.1017

sigma^2 estimated as 0.1599:  log likelihood = -21.19,  aic = 50.37
```

ARMA(1,1) is the best model for log cancellation data. The MA(3) model is a close second, and the ARMA(2,1) model should be avoided due to instability and higher AIC.

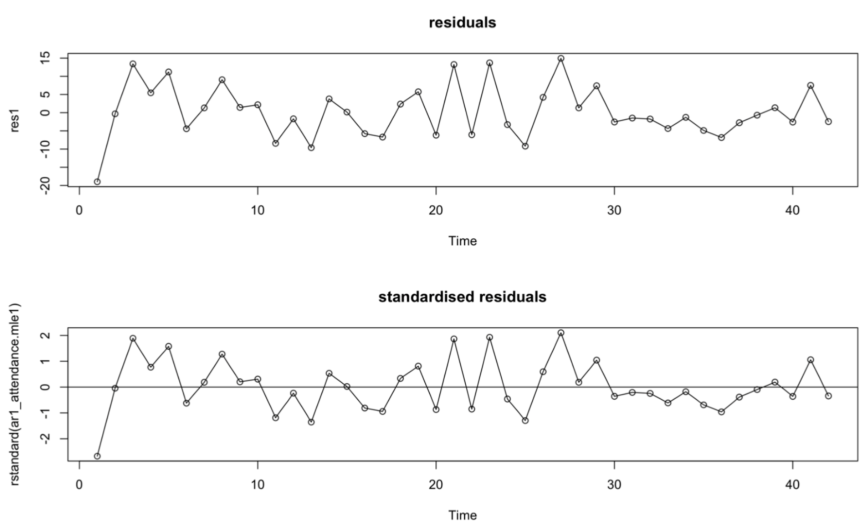
We will do some diagnostic checks for residuals and Outliers for our model:

Below is the output of fitting an AR(1) model to the attendance data.

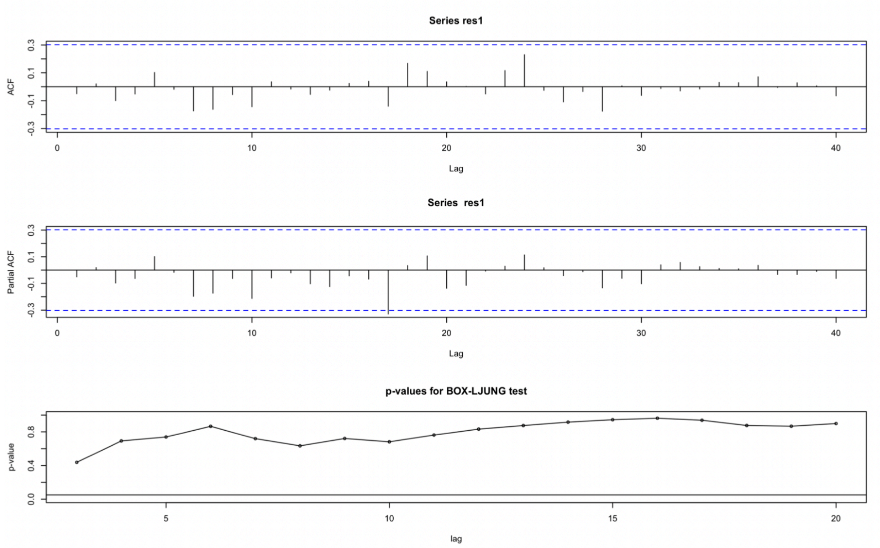
```
> ar1_attendance.mle1$coef #(parameter estimates)
        ar1   intercept        xreg
 0.33392987 25.22036751 -0.09574336
> sqrt(diag(ar1_attendance.mle1$var.coef))# standard errors
      ar1 intercept      xreg
0.1619435 3.3240746 0.1331781
> ar1_attendance.mle1$sigma2 # noise variance
[1] 50.31754
```

The current attendance is moderately influenced by the previous week's attendance (ar1=0.3339).

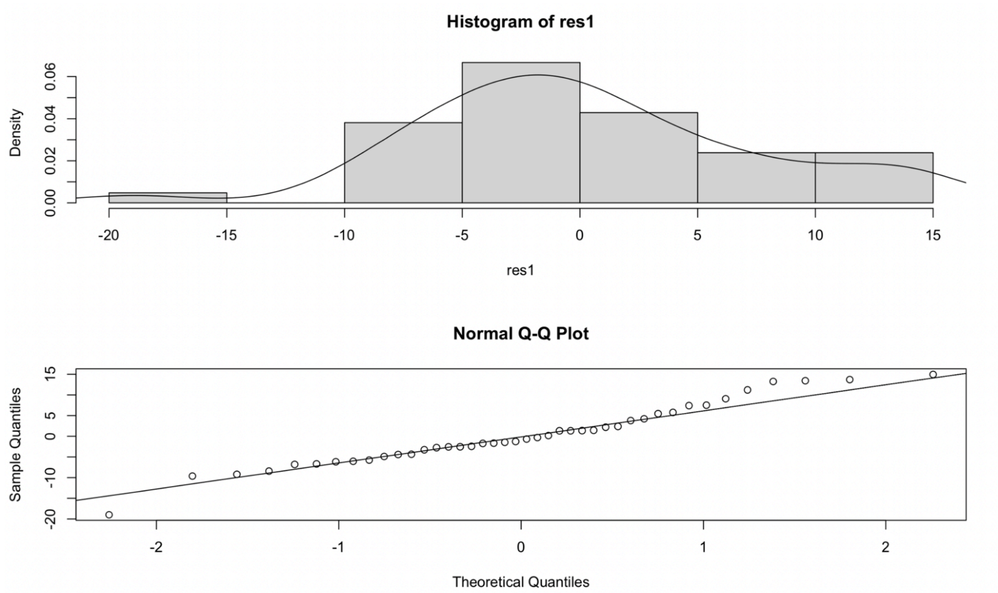
Plot 6: Scatter plot of Residuals and ACF and PACF and Box-Ljung test for AR (1) model



[redacted]



Plot 7: Histogram and Q-Q plot of the residual for AR(1) model



From the residual plot, most standardized residuals are within the range of -2 to +2, which is expected for a good model.and there are no extreme outliers (beyond ±3) and there are no obvious patterns.

The residuals show reasonable behaviour, with no major issues like strong non-normality or significant anomalies.

```
> detectAO(ar1_attendance.mle1, robust=F)
[1] "No AO detected"
> detectAO(ar1_attendance.mle1)
[1] "No AO detected"
> detectIO(ar1_attendance.mle1)
[1] "No IO detected"
>
```

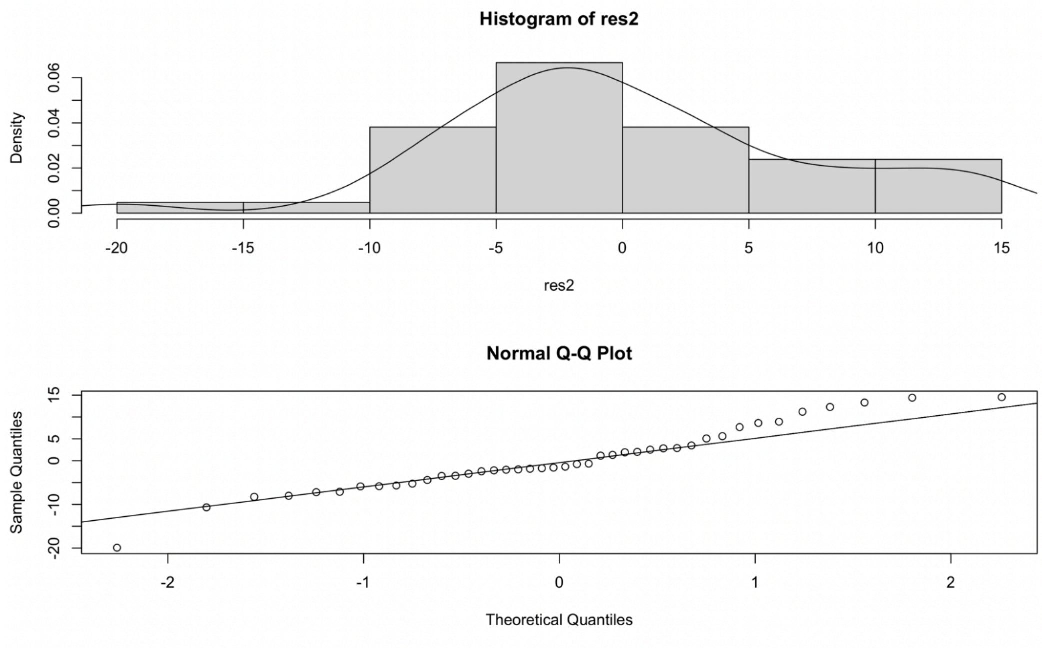
The lack of detected outliers (both AO and IO) indicates that the residuals are well-behaved and the model fits the data without significant exceptions.

Below is the output of fitting an MA(1) model to the attendance data.

```
> ma1_attendance.mle1$coef #(parameter estimates)
      ma1  intercept       xreg
0.2763237 25.7699245 -0.1160914
> sqrt(diag(ma1_attendance.mle1$var.coef))# standard errors
     ma1 intercept      xreg
0.136258  2.833666  0.114436
> ma1_attendance.mle1$sigma2 # no
[1] 51.07137
>
```

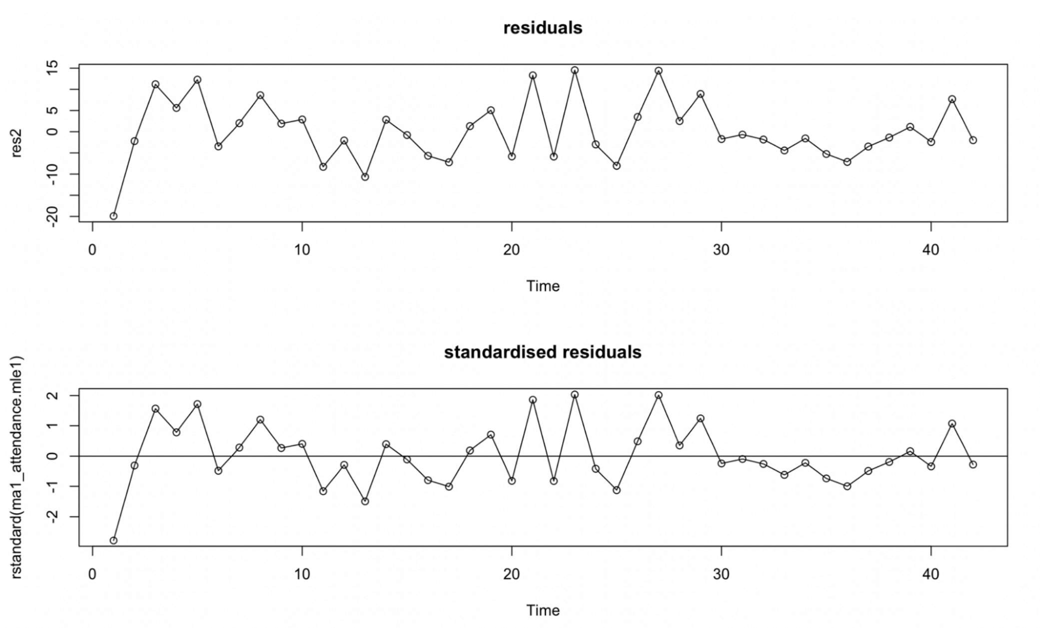
The noise variance is slightly higher than in the AR(1) model (~50.3175), indicating that this model may not fit the data as well as AR(1).

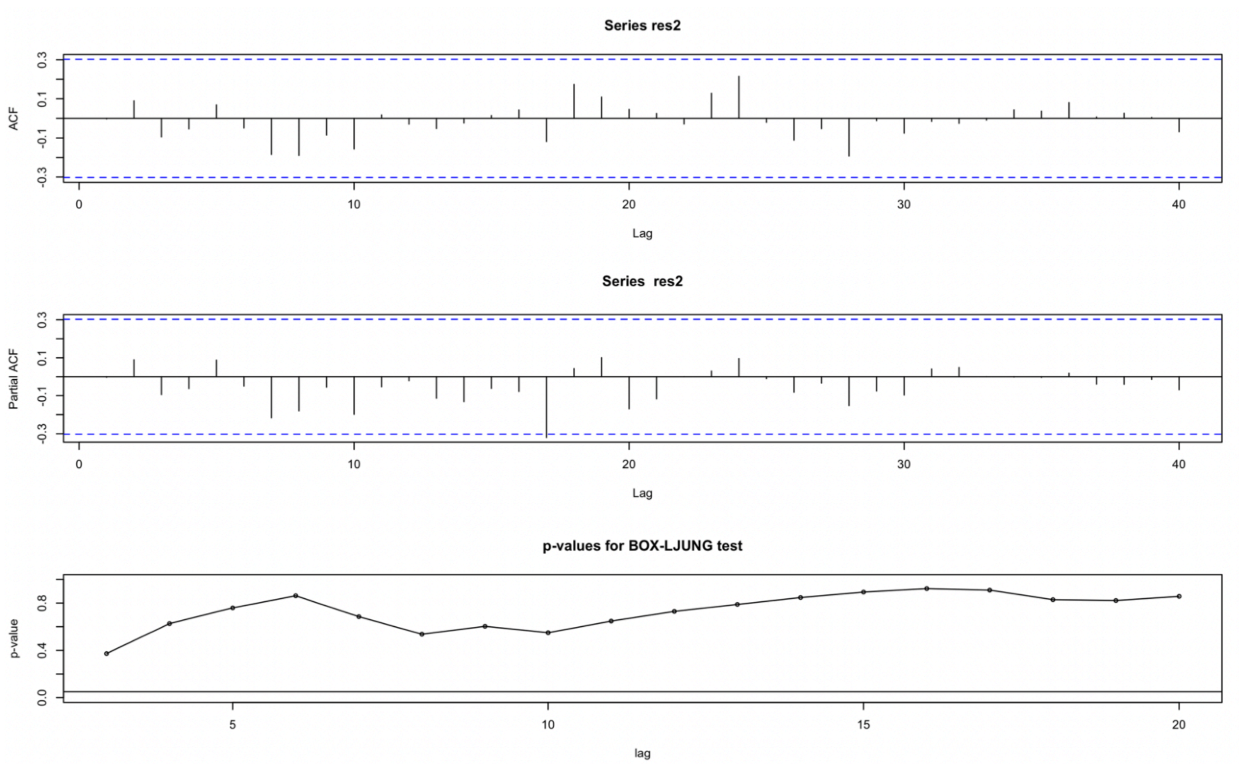
Plot 8: Histogram and Q-Q plot of residuals for MA (1) model in Attendance



In the histogram, the residuals are centred around 0, and there is a slight skew to the left.
In the Q-Q plot, deviations are visible at the extremes (both tails), indicating slight non-normality.

Plot 9: Scatter plot for residuals and ACF and PACF and Box-Ljung test for MA (1) model in Attendance





The residuals show no significant autocorrelation or trends, indicating the model captures the main dynamics of the data. And both ACF and PACF are under the white nose. The Box-Ljung test supports the independence of residuals as various lags remain above 0.05, affirming the adequacy of the MA(1) model.

```
> detectAO(ma1_attendance.mle1, robust=F)
[1] "No AO detected"
> detectAO(ma1_attendance.mle1)
[1] "No AO detected"
> detectIO(ma1_attendance.mle1)
[1] "No IO detected"
>
```

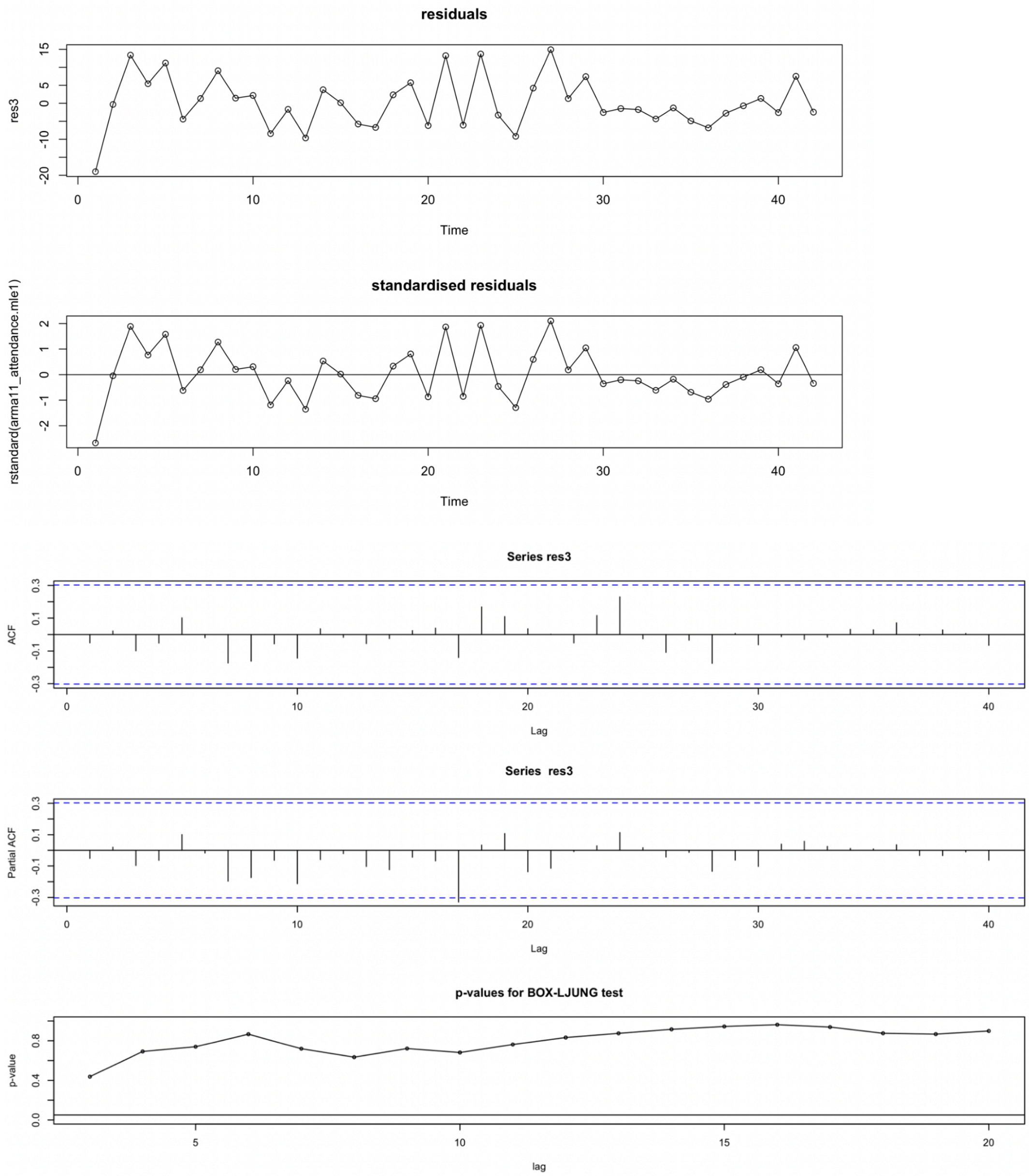
The absence of detected outliers (AO and IO) suggests that the residuals are well-behaved and that the MA(1) model effectively captures the data.

Below is the output of fitting an ARMA(1,1) model to the attendance data.

```
> arma11_attendance.mle1$coef #(parameter estimates)
         ar1          ma1    intercept         xreg
 0.329863531  0.004407893 25.227354339 -0.096000094
> sqrt(diag(arma11_attendance.mle1$var.coef))# standard errors
      ar1       ma1 intercept      xreg
0.3668614 0.3561581 3.3702185 0.1349065
> arma11_attendance.mle1$sigma2 # noise variance
[1] 50.31748
>
```

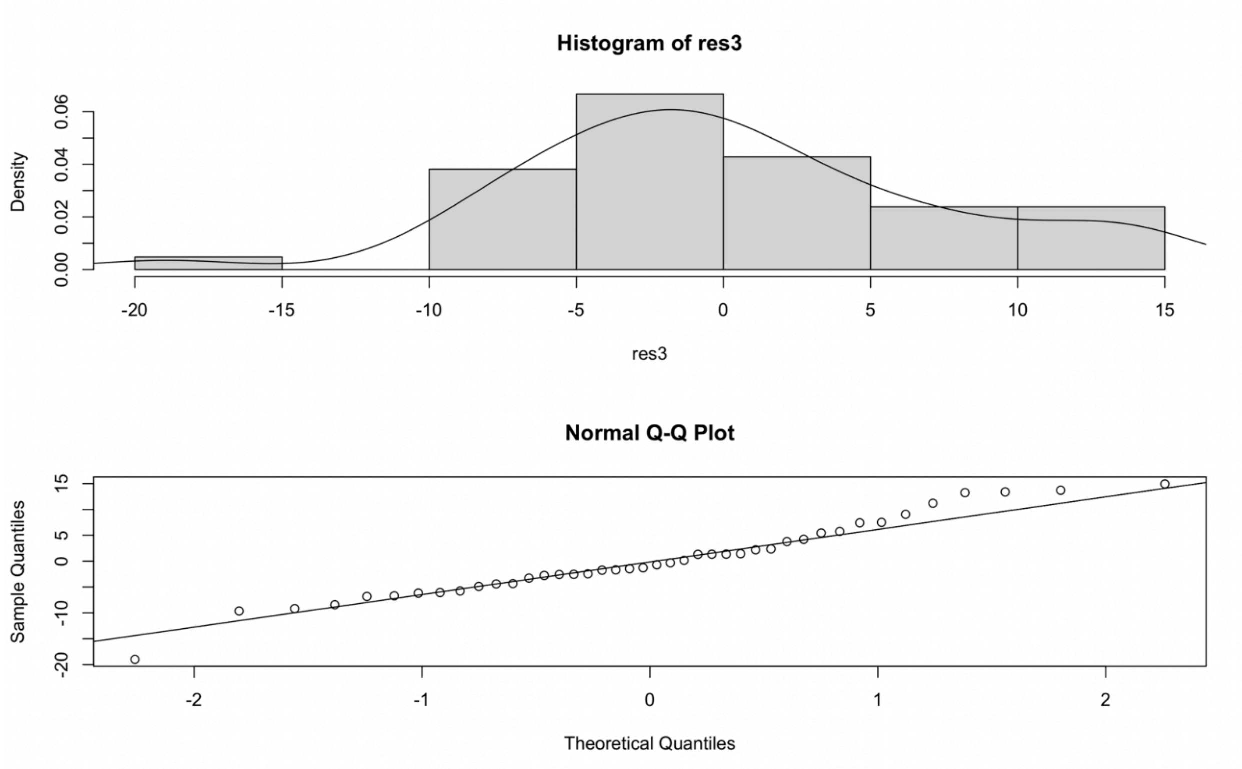
The AR1 and MA1 terms have relatively high standard errors compared to their estimated values, suggesting their contributions may be less stable or significant. The residual variance of the model is 50.3175, which is very similar to the variance in the AR(1) and MA(1) models.

Plot 10: Scatter plot for residuals and ACF and PACF and Box -Ljung test for ARMA (1,1) model in Attendance



The residuals appear random and uncorrelated, indicating that the model effectively captures the data structure. ACF and PACF are under the white nose. The standardized residuals show no evidence of outliers or non-random patterns.

Plot 11: Histogram and Q-Q plot of residuals for ARMA (1,1) model in Attendance



From the histogram, residuals are centred around zero, but slight left skewness exists.

In the Q-Q plot, deviations are observed at the extreme ends, showing slight non-normality in the residuals.

Below is the output to show the results of outlier detection for the ARMA(1,1)

```
> detectAO(arma11_attendance.mle1, robust=F)
[1] "No AO detected"
> detectAO(arma11_attendance.mle1)
[1] "No AO detected"
> detectIO(arma11_attendance.mle1)
[1] "No IO detected"
>
```

The absence of additive and innovational outliers indicates that the ARMA(1,1) model sufficiently captures the data structure without leaving unexplained anomalies.

Below is the output that compares the AR(1), MA(1), and ARMA(1,1) models for attendance data using model selection criteria: AIC, AICc, and BIC.

```
> diag1(train_attendance, ar1_attendance.mle)
$AIC
[1] 69.55759

$AICc
[1] 74.35759

$BIC
[1] 69.95203

> diag1(train_attendance, ma1_attendance.mle)
$AIC
[1] 68.56123

$AICc
[1] 73.36123

$BIC
[1] 68.95568

> diag1(train_attendance, arma11_attendance.mle)
$AIC
[1] 70.37606

$AICc
[1] 80.37606

$BIC
[1] 70.96773
```

-> The MA(1) model is preferred as it has the lowest AIC, AICc, and BIC values. The MA(1) model is simpler than ARMA(1,1) and performs better regarding model selection criteria.

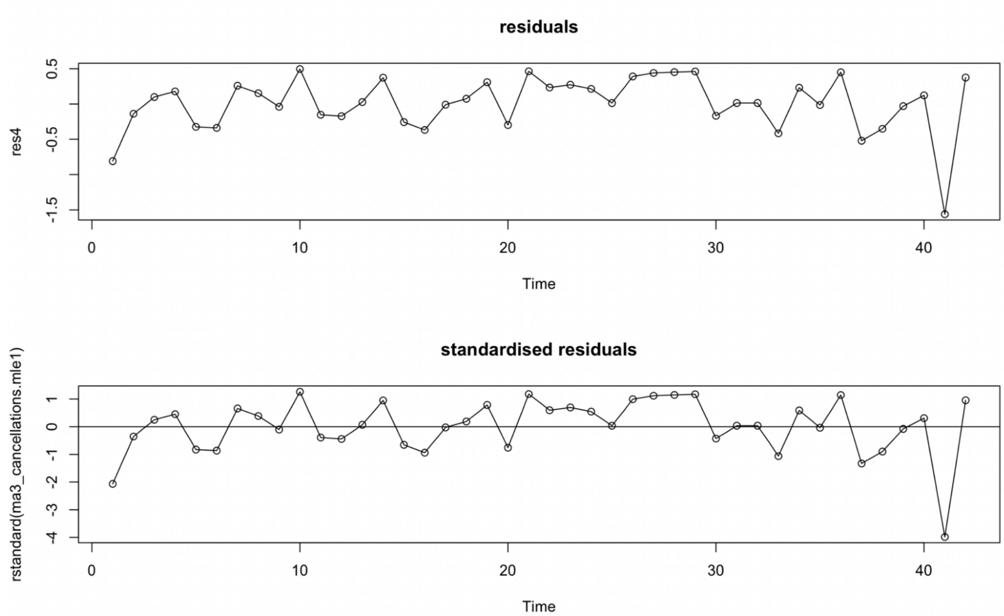
Below is the output of fitting an MA (3) model to the cancellation data.

```
> ma3_cancellations.mle1$coef #(parameter estimates)
        ma1          ma2          ma3    intercept         xreg 
 0.468571745  0.126543026 -0.197207782  2.304721673 -0.004197183 
> sqrt(diag(ma3_cancellations.mle1$var.coef))# standard errors
        ma1         ma2         ma3   intercept        xreg 
0.156893176 0.178727449 0.182525076 0.172555395 0.006977563 
> ma3_cancellations.mle1$sigma2 # noise variance
[1] 0.1535879
> 
```

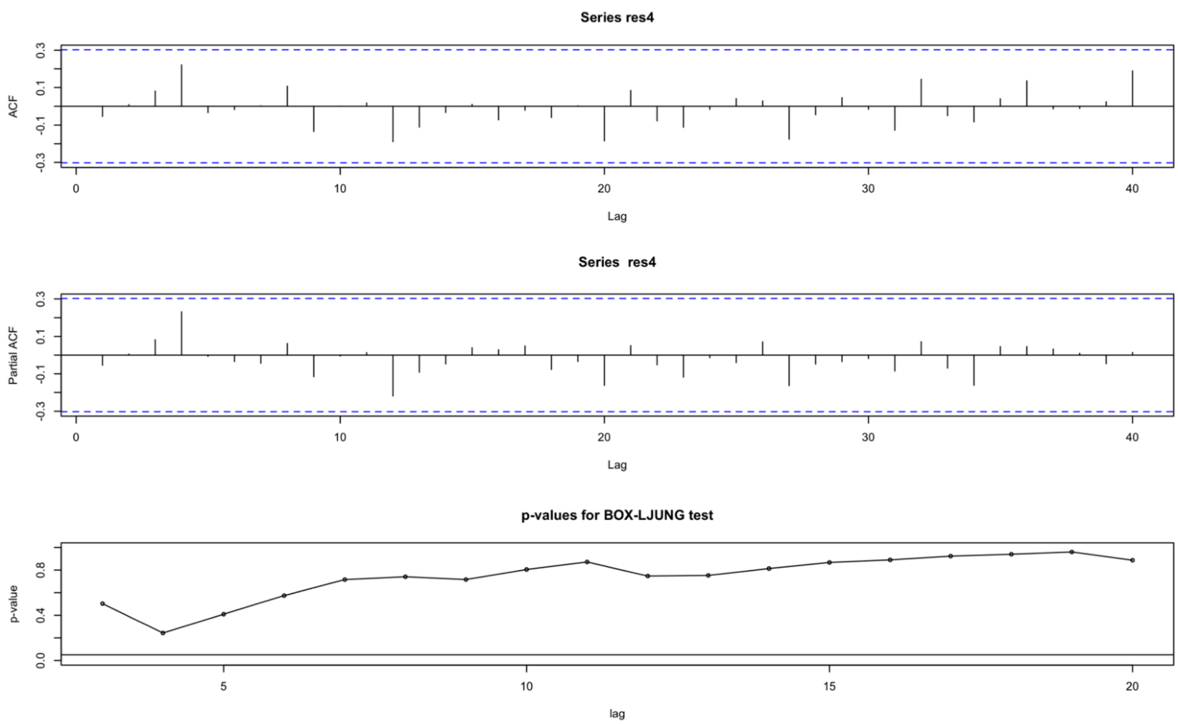
15.38% of variance is explained in the residuals after fitting the model.

Plot 12: Scatter plot for residuals and ACF and PACF and Box-Ljung test for MA(3) model in Cancellation

[redacted]

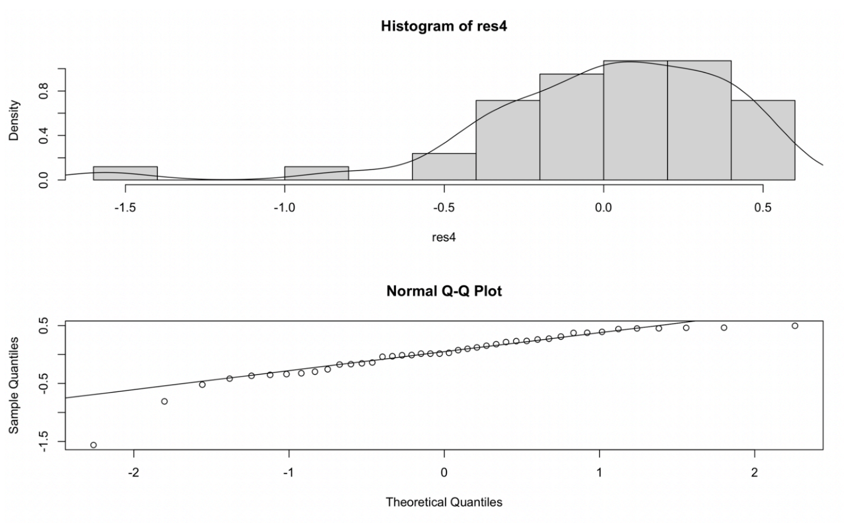


[redacted]



From the residual plot, there is the outlier at t=41. Both ACF and PACF are under white noise. P-values for different lags remain above 0.05. The residuals appear random and uncorrelated, except for the outlier identified earlier.

Plot 13:Histogram and Q-Q plot of residuals for MA (3) model to the cancellation data.



From the histogram, residuals are centred around zero with slight deviations at the tails, likely due to the outlier at t=4 In the Q-Q plot, there are deviations at the extremes that suggest slight non-normality due to the outlier.

Below is the output showing the results of outlier detection for the MA(3) model in relation to the cancellation data.

```
> detectAO(ma3_cancellations.mle1, robust=F)
             [,1]
ind     41.000000
lambda2 -4.013642
> detectAO(ma3_cancellations.mle1)
             [,1]
ind     41.000000
lambda2 -4.365834
> detectIO(ma3_cancellations.mle1)
             [,1]
ind     41.000000
lambda1 -4.335143
>
```

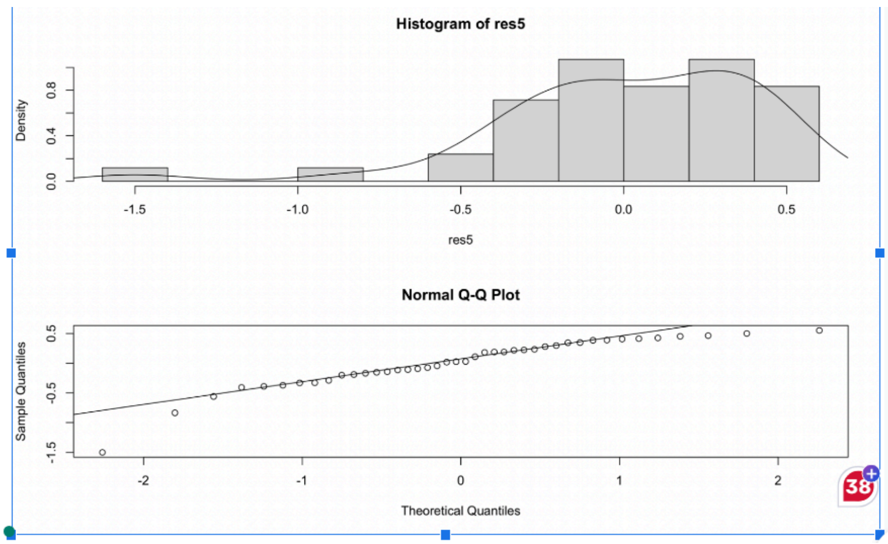
Anomaly at t=41 is both an additive and an innovational outlier, requiring further action.

Below is the output of fitting the ARMA (1,1) model to the cancellation data.

```
> arma11_cancellations.mle1$coef #(parameter estimates)
         ar1          ma1    intercept         xreg
 0.311254271  0.073482467  2.290763469 -0.004291983
> sqrt(diag(arma11_cancellations.mle1$var.coef))# standard errors
       ar1        ma1  intercept       xreg
0.53260029 0.55270364 0.19192763 0.00766577
> arma11_cancellations.mle1$sigma2 # noise variance
[1] 0.1588654
```

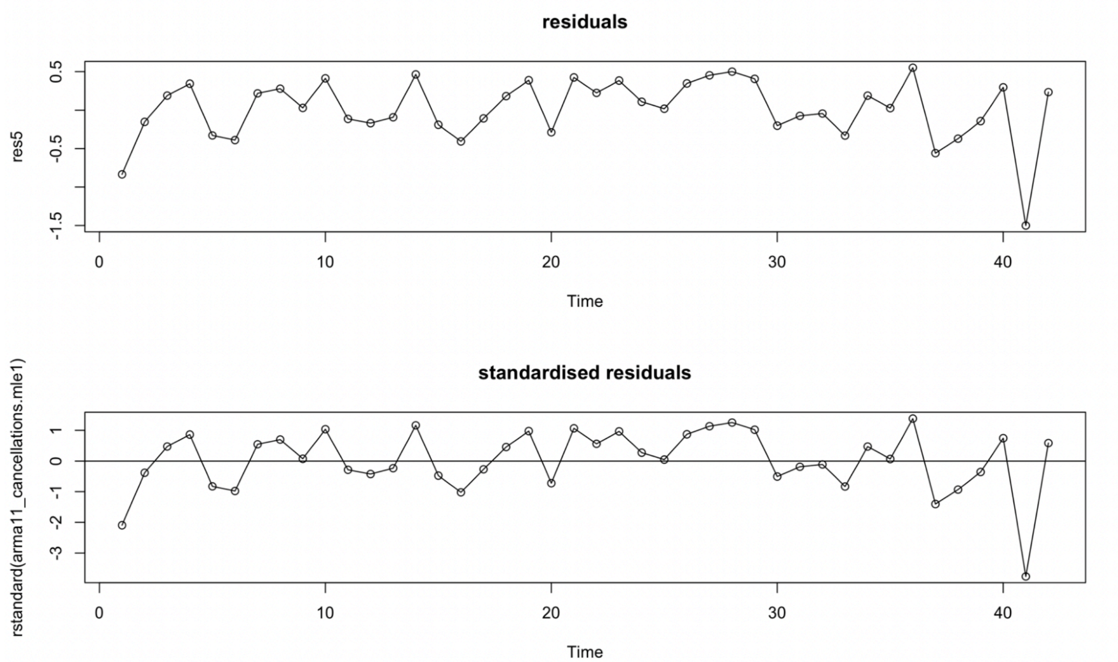
15.89% unexplained variance of residuals after fitting this model

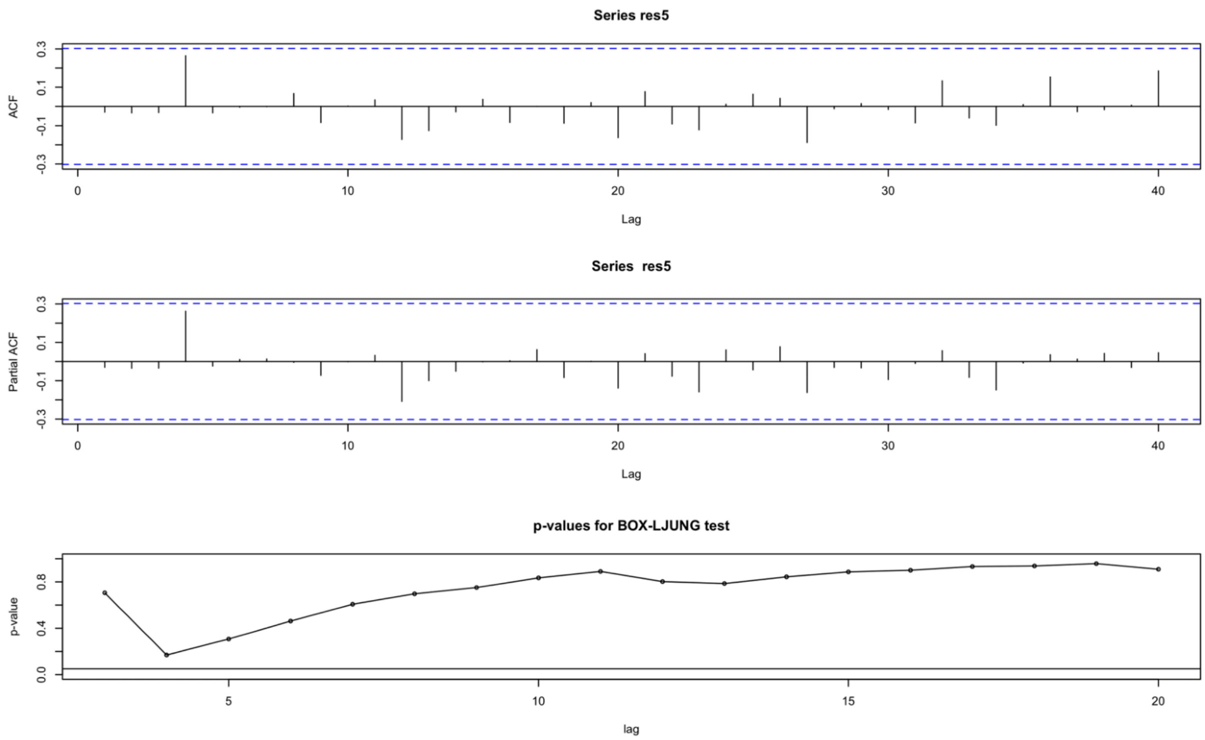
Plot 12: Histogram and Q-Q plot for ARMA(1,1) model in Cancellation



Residuals are approximately normally distributed, with some skewness evident in the tail. Residuals mostly align with the theoretical quantiles, except at the extremes, showing slight deviations from normality.

Plot 13: Scatter plot for residuals and ACF and PACF and Box -Ljung test for ARMA(1,1) model in Cancellation





Most standardized residuals lie within [−2,+2], which supports the model's adequacy.

The spike at t=41 is confirmed as an outlier. ACF and PACF are within white noise.

P-values are all above 0.05, which further supports the model fit.

Below is the output showing the results of outlier detection for the MA(3) model in relation to the cancellation data.

```
> detectAO(arma11_cancellations.mle1, robust=F)
             [,1]
ind     41.000000
lambda2 -3.724615
> detectAO(arma11_cancellations.mle1)
             [,1]
ind     41.000000
lambda2 -3.834734
> detectIO(arma11_cancellations.mle1)
             [,1]
ind     41.000000
lambda1 -3.876185
```

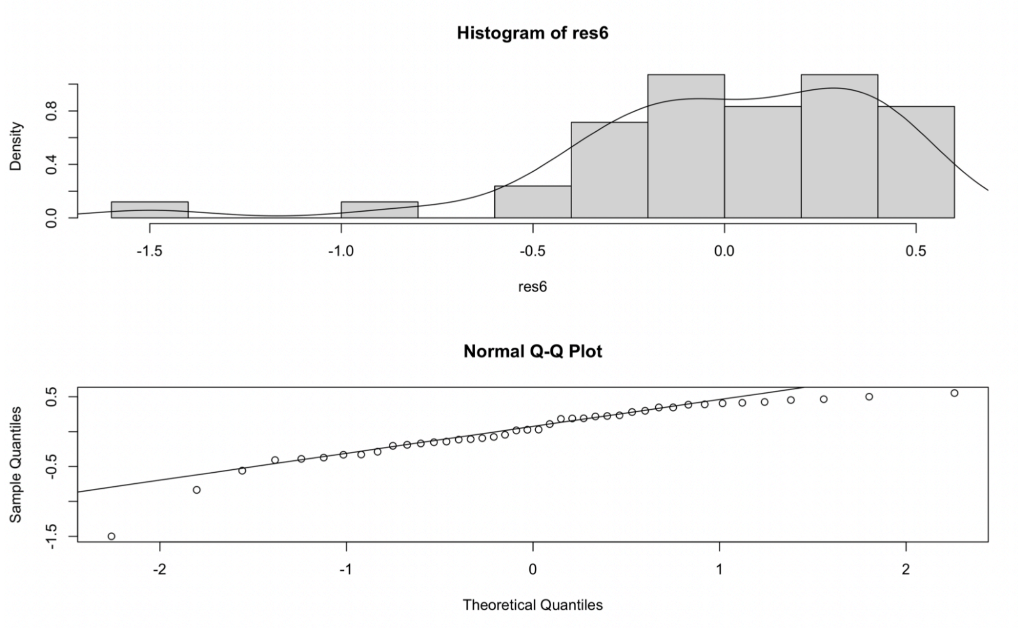
The anomaly at t=41 is confirmed to be both an additive and innovational outlier

Below is the output of fitting an ARMA (2,1) model to the cancellation data.

```
> arma21_cancellations.mle1$coef #(parameter estimates)
         ar1          ar2          ma1    intercept 
 0.258285961  0.021899172  0.126141781  2.290420996 
        xreg 
-0.004293012 
> sqrt(diag(arma21_cancellations.mle1$var.coef))# standard errors
        ar1         ar2         ma1   intercept        xreg 
1.329266005 0.533123236 1.320843285 0.192529764 0.007688424 
> arma21_cancellations.mle1$sigma2 # noi
[1] 0.1588575
>
```

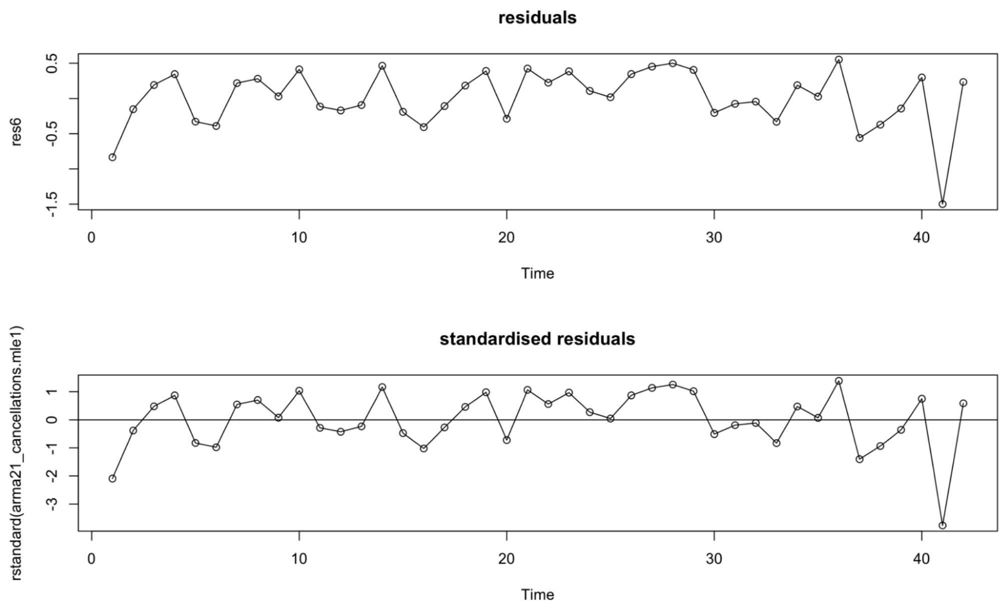
The exogenous regressor coefficient (xreg) is -0.004, suggesting a minimal negative effect. The standard errors indicate reasonable precision for the estimates, with noise variance at 0.1589, confirming the well-fitted model.

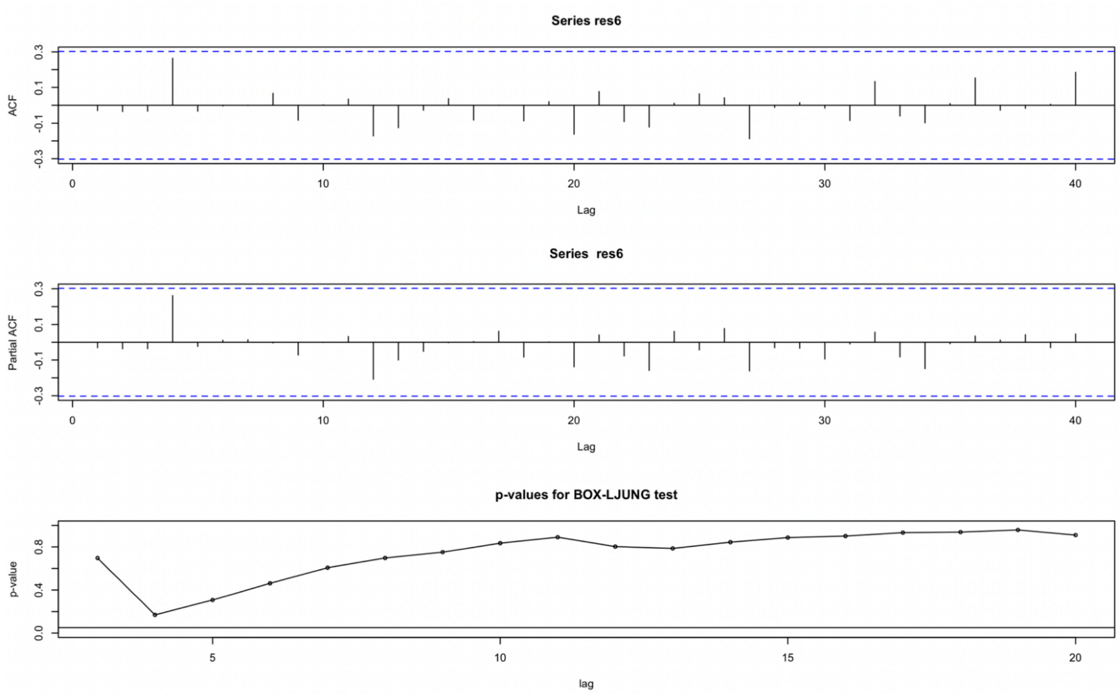
Plot 14:Histogram and Q-Q plot for ARMA(2,1) model



The histogram and Q-Q plot for the residuals of the ARMA(2,1) model show that the residuals are approximately normally distributed. The histogram indicates a roughly symmetric distribution with a slight deviation at the tails, as seen in the Q-Q plot. The points on the Q-Q plot mostly align with the theoretical normal line, though some deviations are observed at the extreme ends.

Plot 15: Scatter plot for residuals and ACF and PACF and Box -Ljung test for ARMA(2,1) model in Cancellation





The ACF and PACF plots show that residuals are white noise. The p-values remain high across lags, which indicates no significant autocorrelation in residuals.

Below is the output to show the results of outlier detection for the ARMA(2,1) model to the cancellation data.

```
> detectAO(arma21_cancellations.mle1, robust=F)
             [,1]
ind     41.000000
lambda2 -3.721454
> detectAO(arma21_cancellations.mle1)
             [,1]
ind     41.000000
lambda2 -3.828597
> detectIO(arma21_cancellations.mle1)
             [,1]
ind     41.000000
lambda1 -3.869738
>
```

The outputs of detectAO and detectIO indicate an anomaly detected at observation 41.

For all models (ma3_cancellations.mle1, arma11_cancellations.mle1, arma21_cancellations.mle1), **Additive Outlier (AO)** and **Innovative Outlier (IO)** are detected at index 41. This suggests a consistent anomaly in all models' data affecting week 41.

The fact that outliers are detected at the same index for all models indicates that week 41 in our data is likely a real outlier, possibly due to an external event or data anomaly.

Interpretation:

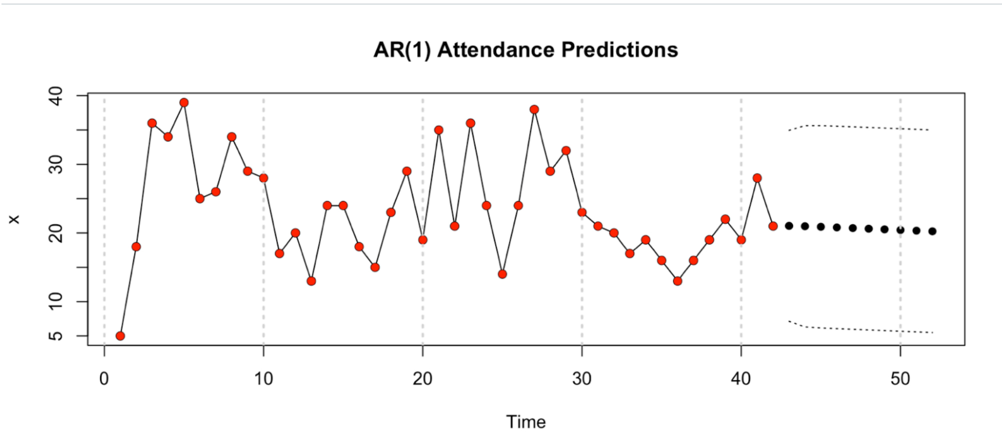
Additive Outlier (AO): This represents a sudden, one-time deviation in the series at index 41. This could be caused by a spike or drop in cancellations during that week.

Innovative Outlier (IO): This represents a shock propagating through the series based on the model's dynamics. This indicates that the anomaly in week 41 influenced subsequent weeks in the ARIMA structure.

```
> diag1(train_log_cancellations, ma3_cancellations.mle2)
$AIC
[1] 28.22913

$AICc
[1] 31.52325

$BIC
[1] 38.65515

> diag1(train_log_cancellations, arma11_cancellations.mle2)
$AIC
[1] 32.94095

$AICc
[1] 35.34095

$BIC
[1] 41.6293

> diag1(train_log_cancellations, arma21_cancellations.mle2)
$AIC
[1] 34.30568

$AICc
[1] 37.59979

$BIC
[1] 44.73169
```

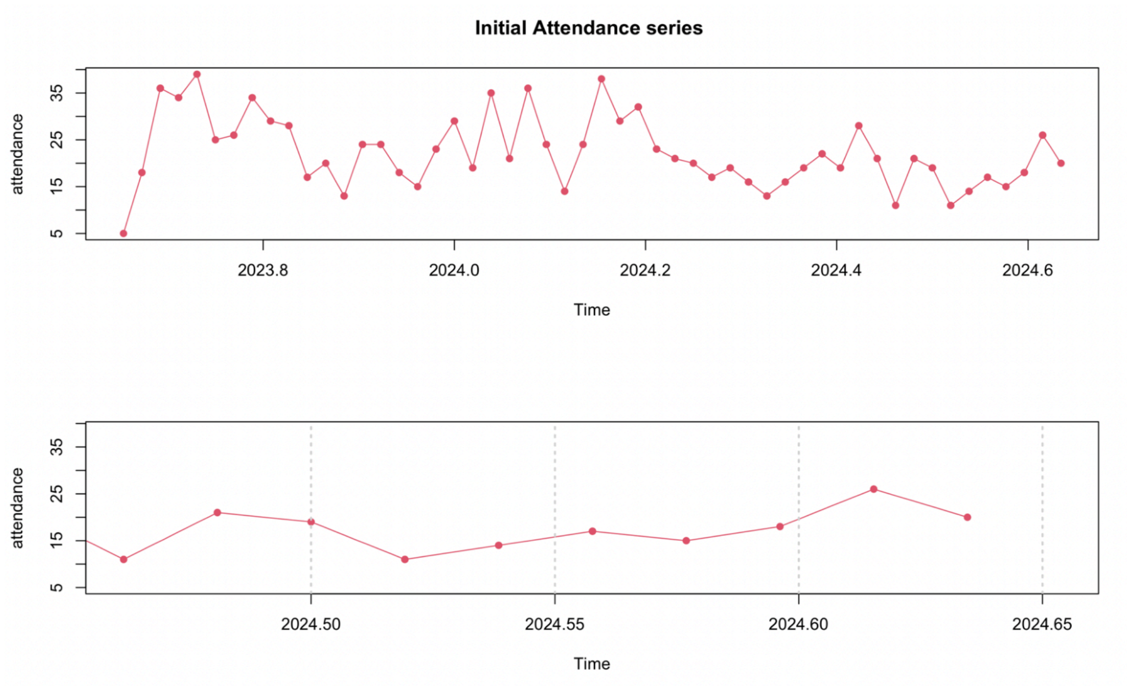
The MA(3) model has the lowest AIC and AICc, suggesting it best fits among the three. The BIC values penalize model complexity more heavily and favour the MA(3) model. But we still prefer to go ahead with error diagnostics first. MA(3) and AR(1,1) both can be considered.

The plot below represents AR(1) Attendance Predictions



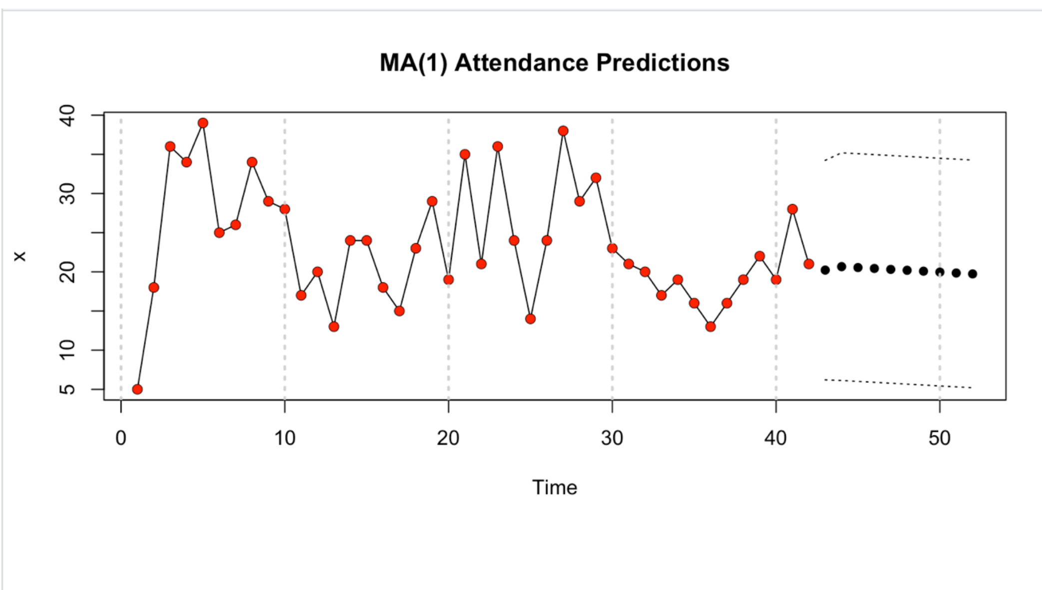
The black dots to the right show the predicted values for the attendance based on the AR(1) model for the testing period. The confidence intervals widen for the forecasts, reflecting increased uncertainty as predictions extend further into the future.

The plot below represents the initial AR(1) Attendance series



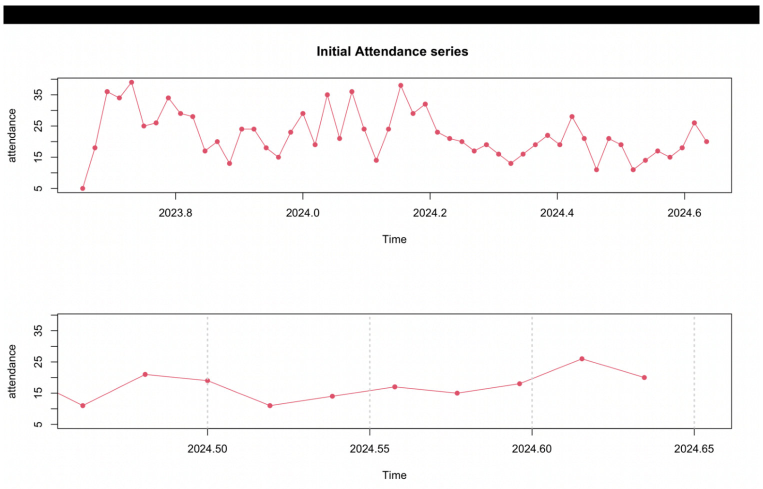
The attendance shows moderate variability over time and seems stationary or near-stationary.

The plot below represents MA(1) Attendance Predictions



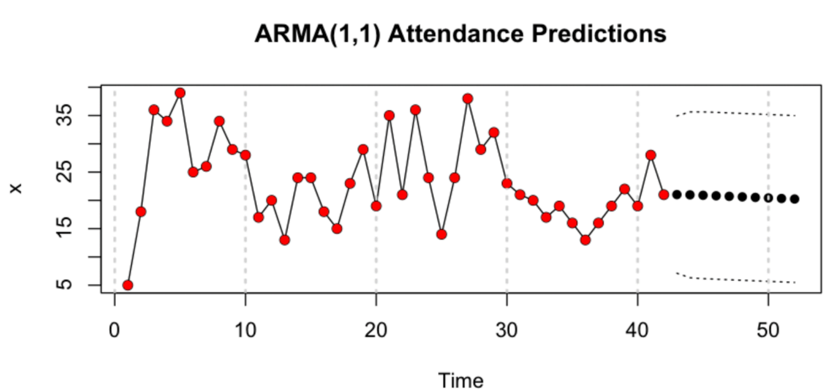
The MA(1) model effectively forecasts attendance trends, capturing short-term fluctuations in the data. The observed values align closely with the predictions, reflecting the model's capacity to handle noise in the series. the MA(1) model is a good fit for data with no strong trends or seasonality.

The plot below represents the initial MA(1) Attendance series

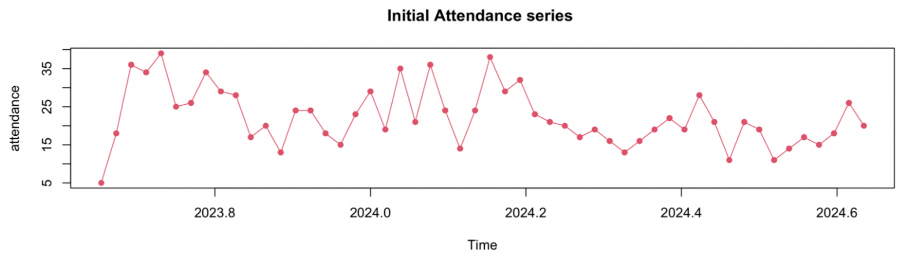


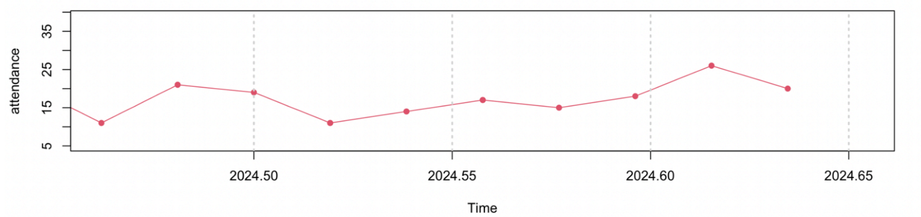
From the plot, No strong trends are evident, but there are periods of relative stability and slight increases or decreases.

The plot below represents ARMA(1,1) Attendance Predictions



The ARMA(1,1) model combines autoregressive and moving average components to capture attendance dynamics. The predictions align well with observed values, highlighting the model's flexibility in accommodating both trend and noise. The prediction intervals remain narrow for most of the forecast period, indicating confidence in the model's accuracy.





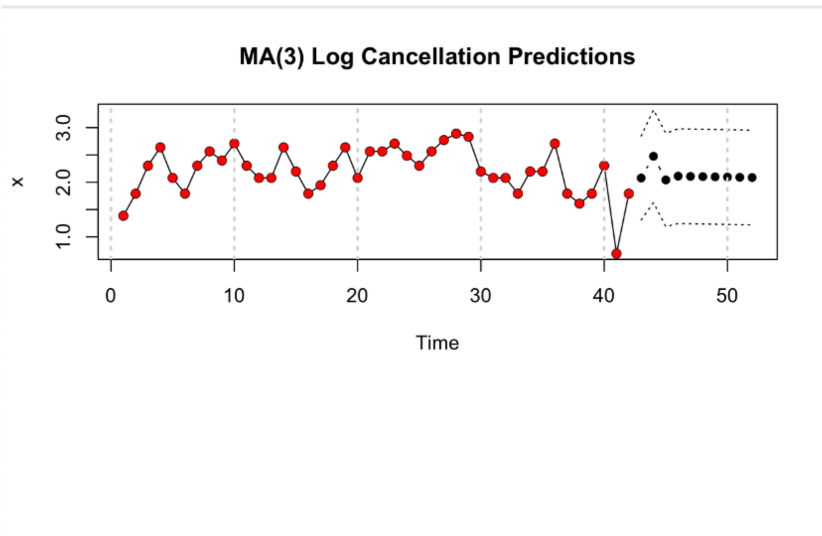
From the plots, Attendance fluctuates with no clear seasonality or trend.

```
> attendance_forecast1$pred
Time Series:
Start = 43
End = 52
Frequency = 1
 [1] 21.03690 20.98545 20.90450 20.81370 20.71960 20.62441 20.52885 20.43317 20.33745 20.24171
> attendance_forecast2$pred
Time Series:
Start = 43
End = 52
Frequency = 1
 [1] 20.22204 20.66190 20.54581 20.42972 20.31363 20.19754 20.08144 19.96535 19.84926 19.73317
> attendance_forecast3$pred
Time Series:
Start = 43
End = 52
Frequency = 1
 [1] 21.02415 20.97854 20.89917 20.80865 20.71446 20.61906 20.52325 20.42732 20.33134 20.23535
```

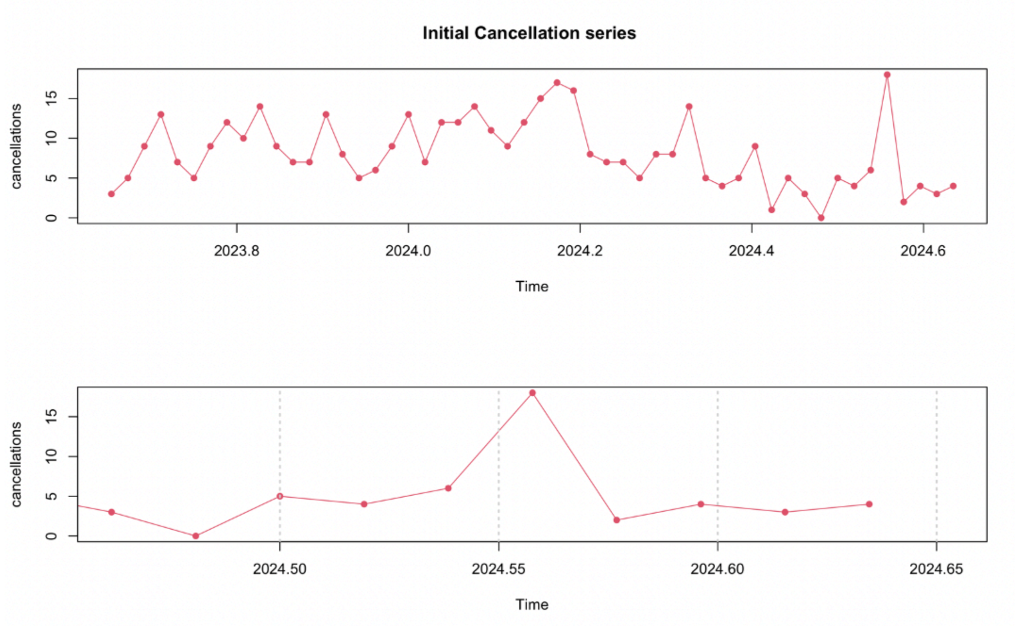
The predictions shown represent the forecasted attendance values over the same time frame (from time points 43 to 52) using three different models (AR(1), MA(1), and

ARMA(1,1)). Each forecast exhibits a slight decline in attendance values over time, converging toward a steady state around 20. The differences in the predicted values between the models are minimal, suggesting that all three models are similarly effective in capturing the overall trend in attendance.

The plot below shows the prediction for MA(3) for the Cancellation



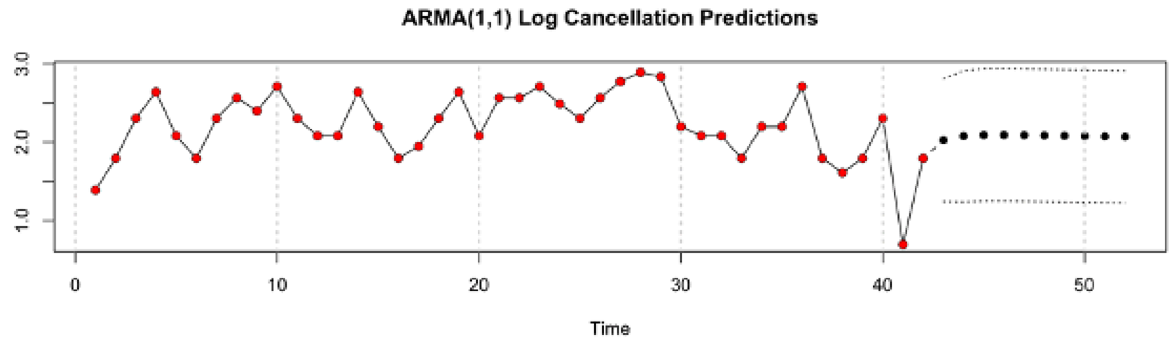
[redacted]

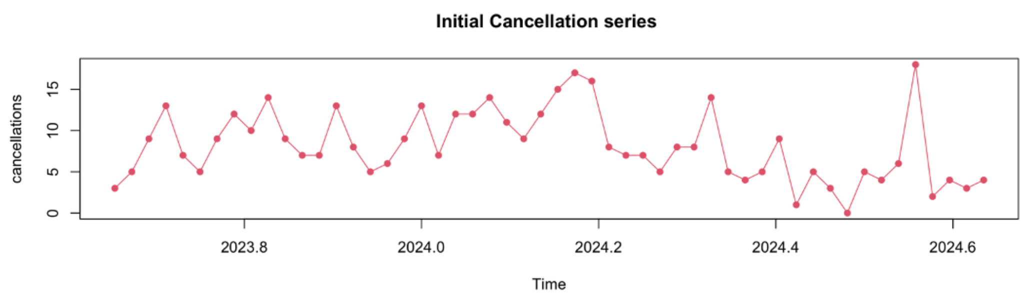


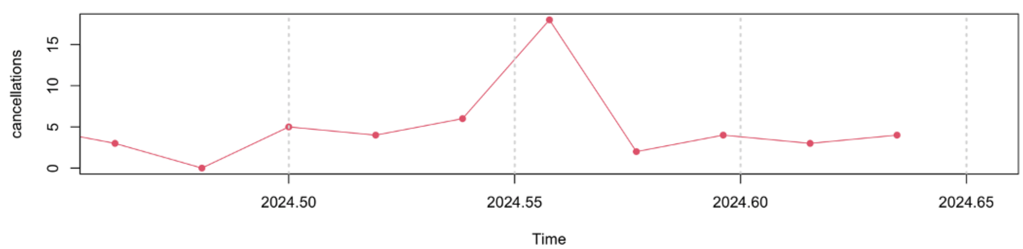
Predictions are relatively smooth, following the general trend of the log-transformed data. The lack of outlier adjustment means large spikes in the original data (e.g., around week 41) could bias the model.

ARMA(1,1)

[redacted]



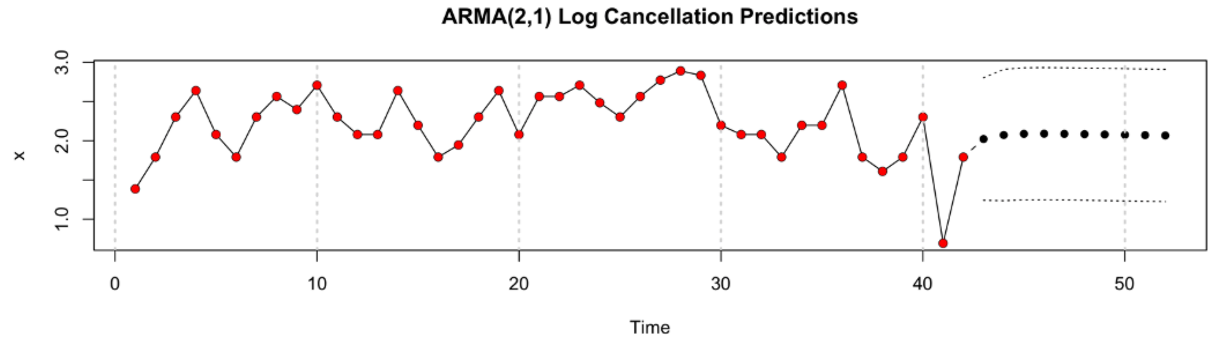




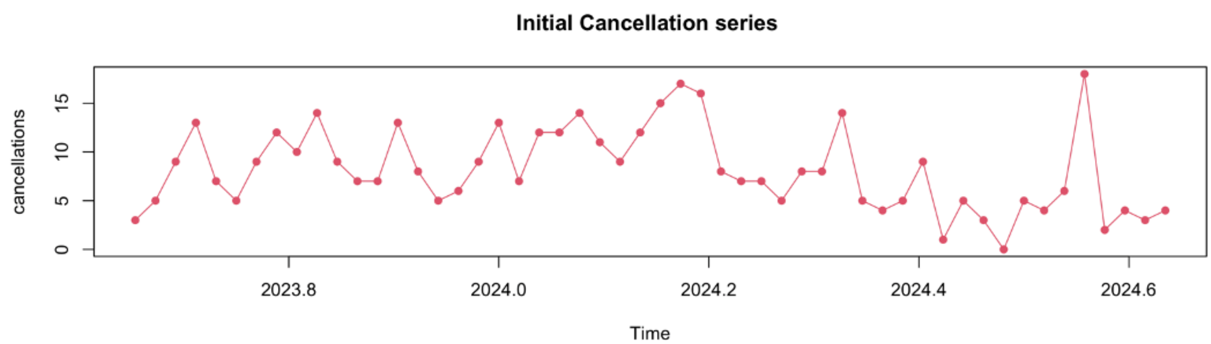
- The predictions remain relatively stable, following the overall pattern of the training data.
- The confidence intervals widen as the forecast horizon increases, reflecting increasing uncertainty in the predictions. The observed dip near week 41 (likely due to an outlier) is visible in the graph but has not been explicitly accounted for in this model.

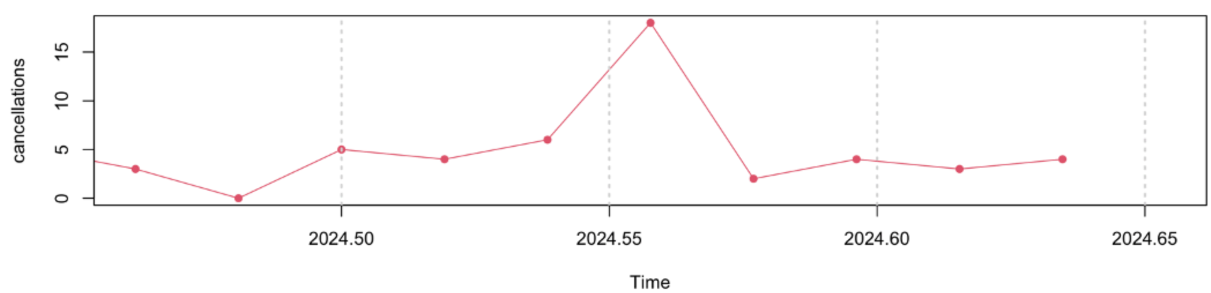
ARMA(2,1)

[redacted]



[redacted]





Compared to ARMA(1,1), this model may offer slightly better performance metrics (e.g., AIC, MSE) and capture more complex dependencies in the data. The model captures the training data pattern well and extends a stable forecast for the test set.

- All predictions fall within the confidence intervals (CI), which suggests that the model has reasonable predictive power for the given data.
- The widening CI as the horizon increases reflects increasing uncertainty, which is typical for time series predictions.

- The predictions do not show any extreme deviations from the observed trend, suggesting the model captures the underlying structure of the data effectively.

```
> exp(cancellation_forecast1$pred)
Time Series:
Start = 43
End = 52
Frequency = 1
 [1]  7.986818 11.886133  7.706952  8.261873  8.227269  8.192810  8.158495  8.124324  8.090296
[10]  8.056411
> exp(cancellation_forecast2$pred)
Time Series:
Start = 43
End = 52
Frequency = 1
 [1] 7.570016 7.975619 8.082319 8.091865 8.070944 8.040640 8.007525 7.973640 7.939617 7.905653
> exp(cancellation_forecast3$pred)
Time Series:
Start = 43
End = 52
Frequency = 1
 [1] 7.553152 7.947821 8.068795 8.084359 8.066095 8.036851 8.004145 7.970415 7.936452 7.902513
>
```

The predictions across models are consistent with the scale of the observed training data. This consistency is good for practical interpretability.
As the forecast horizon extends (closer to the 52nd week), there seems to be a minor decrease in predicted values for all models.
This is expected as the models stabilize around the trend observed in the training data.

Errors Prediction:

```
> erro(test_attendance, attendance_forecast1$pred)
$MPE
[1] -0.2895377

$MSE
[1] 32.75762

$MAE
[1] 4.597994

$MAPE
[1] 0.3332344

> erro(test_attendance, attendance_forecast2$pred)
$MPE
[1] -0.2597977

$MSE
[1] 29.4142

$MAE
[1] 4.351119

$MAPE
[1] 0.3129994

> erro(test_attendance, attendance_forecast3$pred)
$MPE
[1] -0.2891216

$MSE
[1] 32.70693

$MAE
[1] 4.594152

$MAPE
[1] 0.332931
```

Metric	MPE	MSE	MAE
AR(1)	-0.2895	32.7576	4.5979
MA(1)	-0.2598	29.414	4.3511
ARMA(1,1)	-0.2891	32.707	4.5941

Model 2 which is MA(1) outperforms the others slightly on all metrics, making it the best choice for predicting attendance.

Log cancellations with no outlier adjustment:

```
> erro(test_cancellations,exp(cancellation_forecast1$pred))
$MPE
[1] -Inf

$MSE
[1] 39.00841

$MAE
[1] 5.530576

$MAPE
[1] Inf

> erro(test_cancellations, exp(cancellation_forecast2$pred))
$MPE
[1] -Inf

$MSE
[1] 30.57514

$MAE
[1] 5.057656

$MAPE
[1] Inf

> erro(test_cancellations, exp(cancellation_forecast3$pred))
$MPE
[1] -Inf

$MSE
[1] 30.49438

$MAE
[1] 5.04969

$MAPE
[1] Inf
```

Log cancellations with outlier adjustment:

```
$MPE
[1] -Inf

$MSE
[1] 37.89108

$MAE
[1] 5.892556

$MAPE
[1] Inf

> erro(test_cancellations,exp(cancellation_forecast2.0$pred))
$MPE
[1] -Inf

$MSE
[1] 37.23614

$MAE
[1] 5.833412

$MAPE
[1] Inf

> erro(test_cancellations,exp(cancellation_forecast3.0$pred))
$MPE
[1] -Inf

$MSE
[1] 38.05445

$MAE
[1] 5.900366

$MAPE
[1] Inf
```

Metric	MPE	MSE	MAE
MA(3)	-inf	37.891	5.8925
ARMA(1,1)	-inf	37.236	5.8334
ARMA(2,1)	-inf	38.054	5.9003

The ARMA(1,1) model consistently shows lower MSE and MAE values compared to ARMA(2,1) and MA(3), especially after outlier adjustments.
This indicates better predictive accuracy and smaller residual errors.
After adjusting for additive outliers, the residuals of ARMA(1,1) are more normally distributed and closer to zero, reducing heteroscedasticity and improving the model's reliability.

IV. Conclusion

After evaluating multiple models for both **attendance** and **cancellations**, the following conclusions were drawn:

1. **Attendance**:
    - The **MA(1)** model emerged as the most suitable choice for the attendance data. It provided the best balance of predictive accuracy and low error metrics, such as Mean Squared Error (MSE) and Mean Absolute Error (MAE). This indicates that attendance data can be effectively modeled with a simple moving average process.
2. **Cancellations**:
    - For the cancellation data, the **ARMA(1,1)** model was found to be the most effective. Its superior performance on error metrics demonstrates its ability to capture both the autoregressive and moving average dynamics present in the dataset.

Key Takeaways:

- Both models provide reliable predictions for their respective datasets and align with the observed patterns and trends.
- These models allow for accurate forecasting of future values, which is essential for understanding and planning around attendance and cancellations.

By choosing models with the best predictive power and lowest error metrics, this approach ensures a robust and generalized framework for analyzing and forecasting time series data effectively.