STAT 443: Lab 9

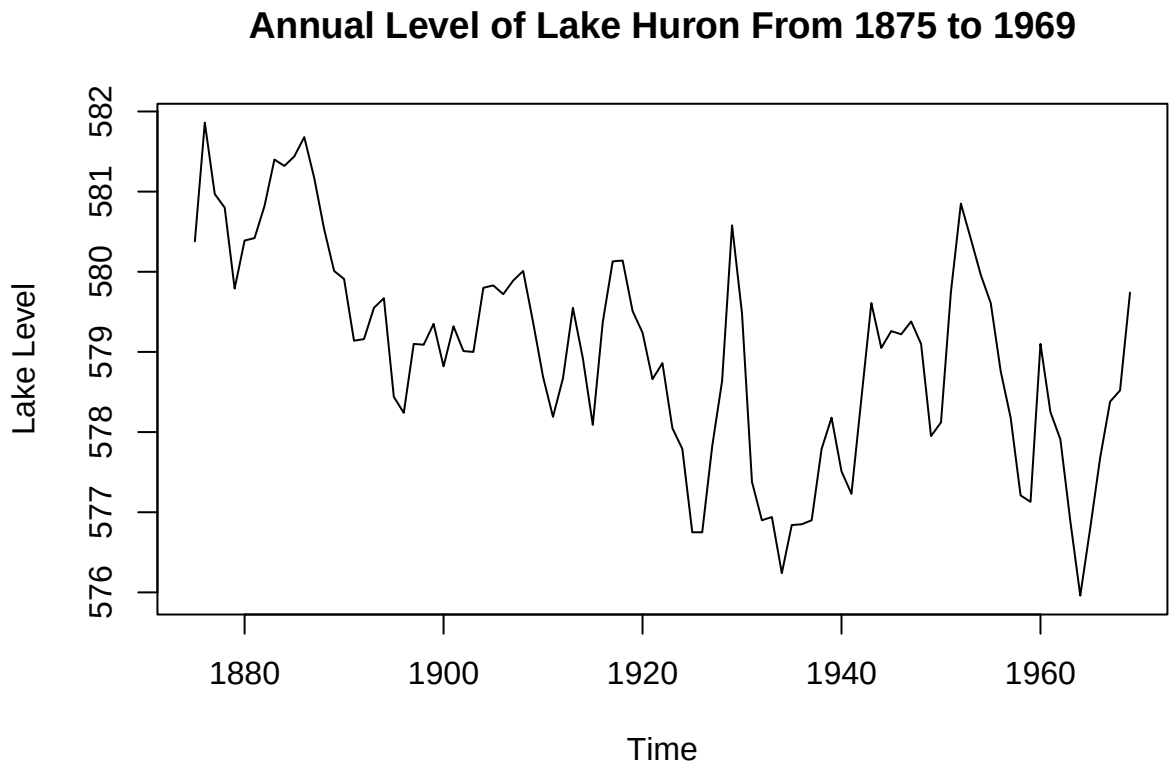
Wenxuan Zan (61336194)

20 March, 2023

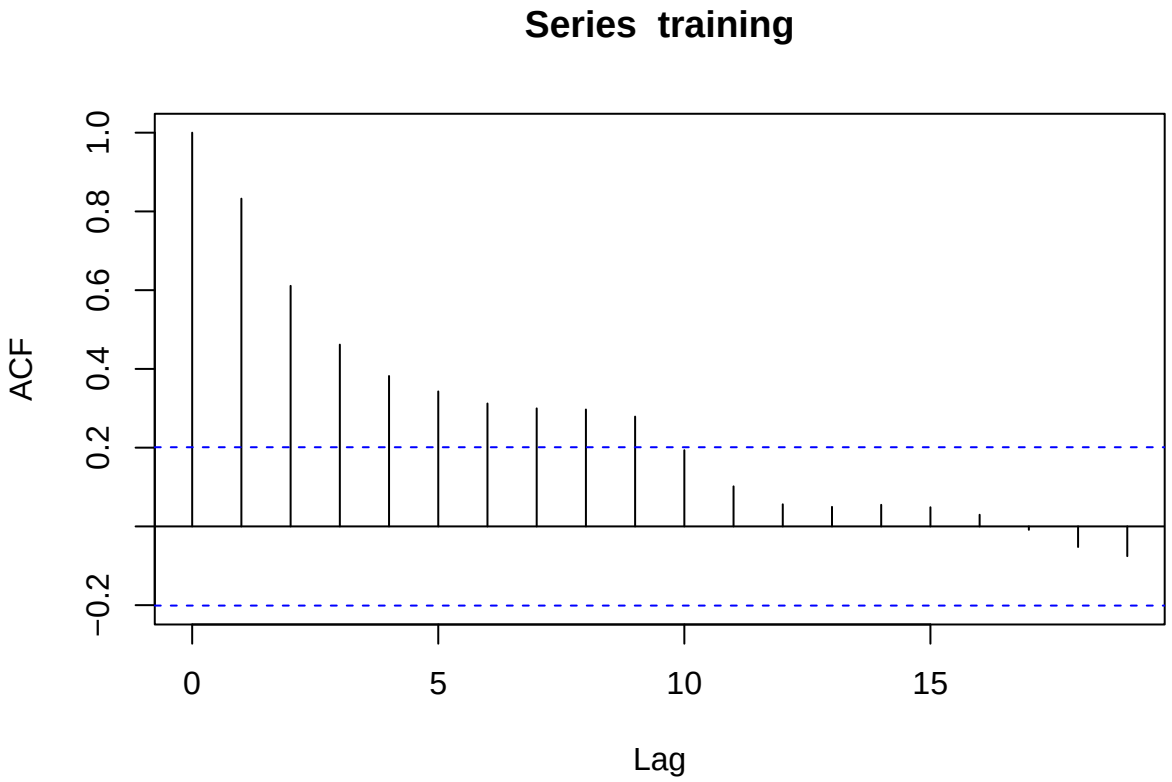
```
data("LakeHuron")
training <- window(LakeHuron, start = c(1875), end = c(1969))
testing <- window(LakeHuron, start = c(1970), end = c(1972))
```

Question 1

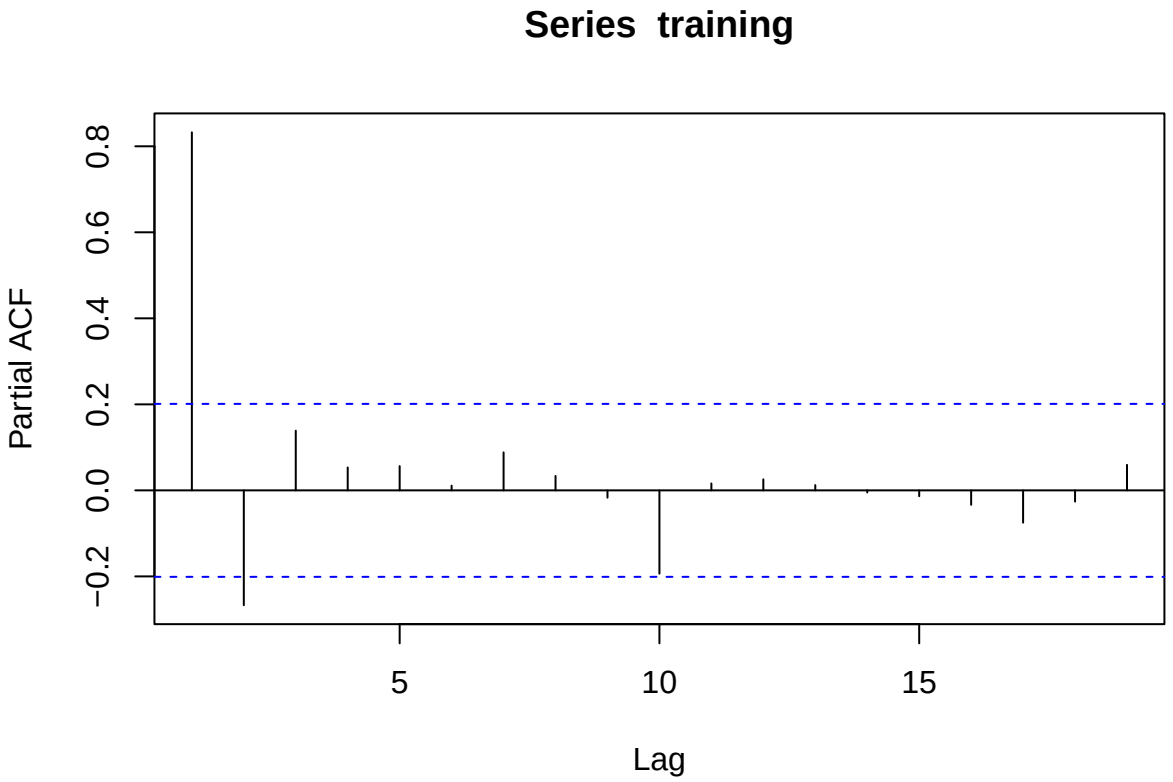
```
plot(training,
     ylab = "Lake Level",
     main = "Annual Level of Lake Huron From 1875 to 1969")
```




```
acf(training)
```




```
pacf(training)
```



Looking at the ACF plot, the values of auto-correlation decrease exponentially and the partial ACF plot shows a cut-off at lag 2. I would suggest a ARMA(2,0) model given the above observations.

Question 2

```
model <- arima(training, order = c(2,0,0), include.mean = TRUE)
model
```

```
##
## Call:
## arima(x = training, order = c(2, 0, 0), include.mean = TRUE)
##
## Coefficients:
##          ar1      ar2  intercept
##       1.0617  -0.2707   579.0319
## s.e.  0.1006   0.1030     0.3339
##
## sigma^2 estimated as 0.484:  log likelihood = -101.01,  aic = 210.01
```

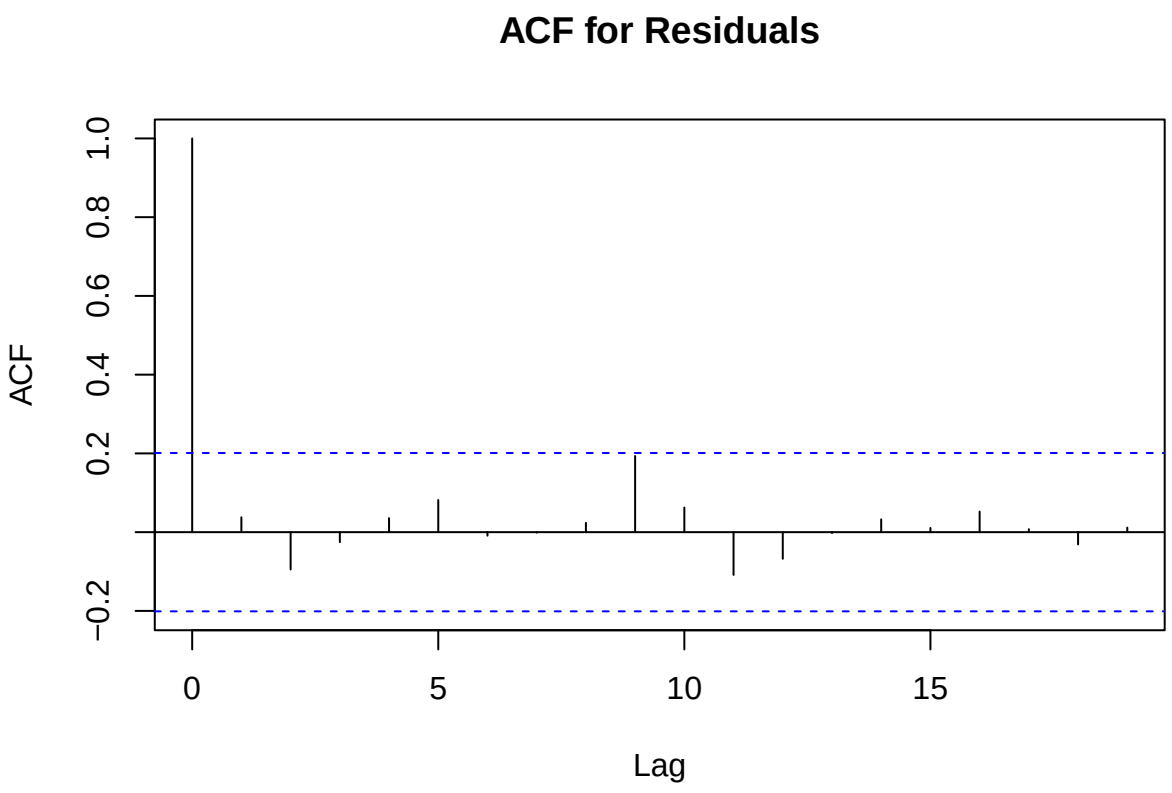
Fitted model:

$$X_t - \hat{\mu} = 1.0617(X_{t-1} - \hat{\mu}) - 0.2707(X_{t-2} - \hat{\mu}) + Z_t;$$

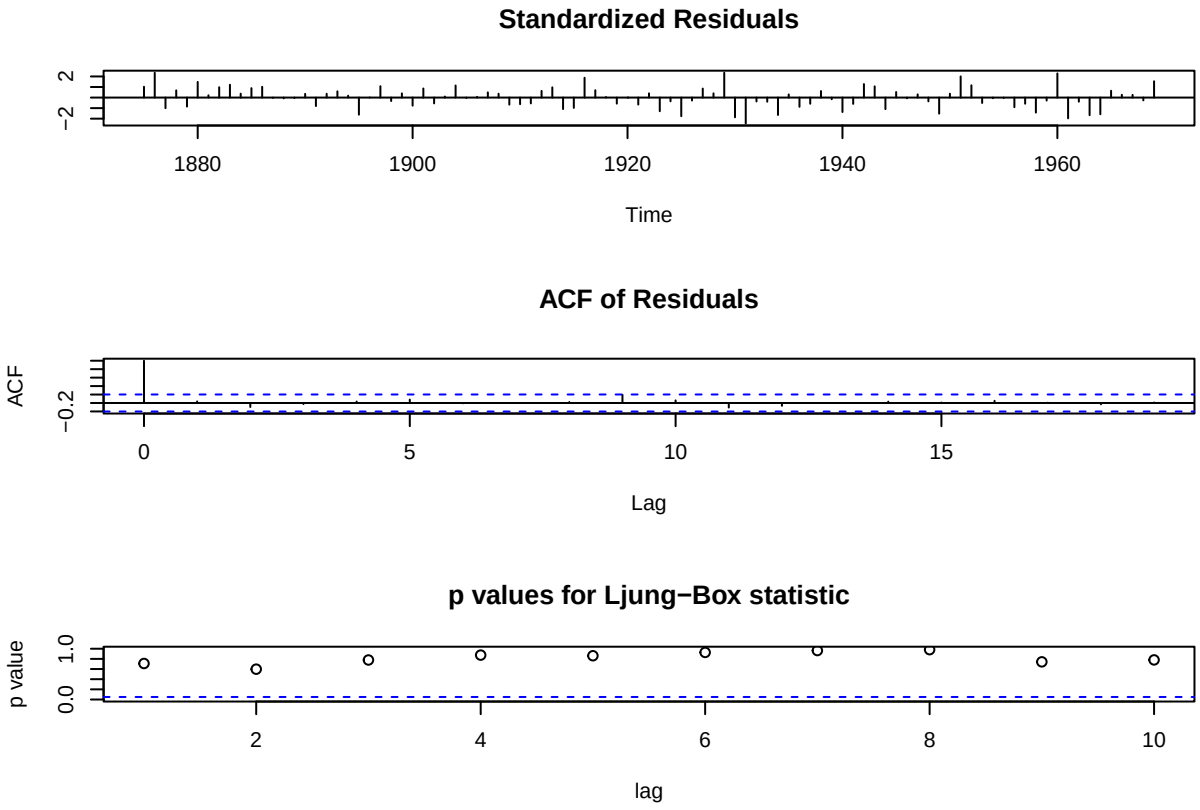
where $\hat{\mu} = 579.0319$ and $Z_t \sim WN(0, 0.484)$

Question 3

```
acf(model$residuals,main = "ACF for Residuals")
```



```
tsdiag(model)
```



i) The ACF plot for the residuals indicates no significant autocorrelations after lag 0.

ii) Almost all standardized residuals fall between the ± 2 range, and the p-values for Ljung-Box statistic are not significant even at lag $= 10$.

iii) Given the above two observations, the model fits is good

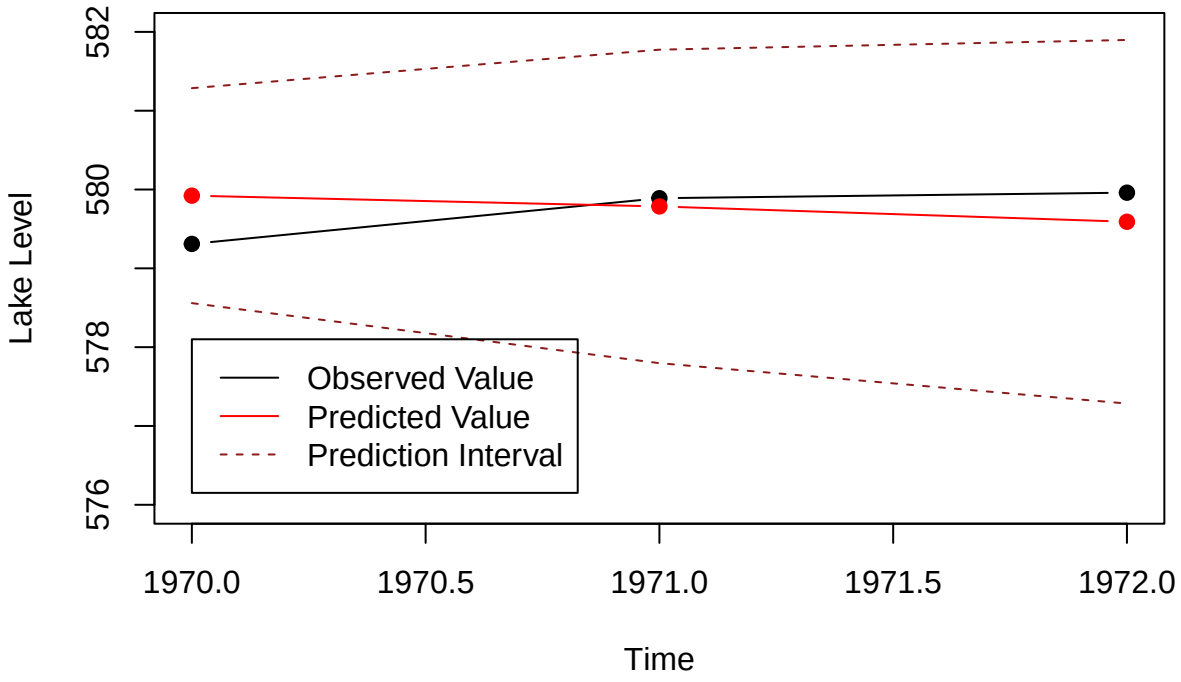
Question 4

```
predicted_levels <- predict(model,n.ahead = 3,prediction.interval = TRUE,level = 0.95)
q4result <- tibble(Time = c("1970", "1971", "1972"),prediction = predicted_levels$pred,
                   lower_bound = prediction - 1.96*predicted_levels$se,
                   upper_bound = prediction + 1.96*predicted_levels$se)
kable(q4result)
```

Time	prediction	lower_bound	upper_bound
1970	579.9223	578.5587	581.2859
1971	579.7856	577.7967	581.7744
1972	579.5911	577.2846	581.8976

Question 5

```
q4result[,"Observed"] <- testing
plot(testing,
     xlim = c(1970.0, 1972.0),
     ylim = c(576,582),
     type = "b",
     pch = 19,
     ylab = "Lake Level")
lines(q4result$prediction, col = "red", type = "b", pch = 19)
lines(q4result$lower_bound, col = "firebrick4", lty = "dashed")
lines(q4result$upper_bound, col = "firebrick4", lty = "dashed")
legend(1970.0,
       578.1,
       legend = c("Observed Value",
                  "Predicted Value",
                   "Prediction Interval"),
       lty = c("solid","solid","dashed"),
       col = c("black", "red","firebrick4"))
```



The prediction interval is wider for predictions that are further into the future. Although all observed values are within the 95% prediction interval, we should notice that the predicted values show a downward trend whereas the observed values showed a upward trend.