STA457-Assignment 1*

Hyuk Jang

```r
# Code Space Setup
library(ggplot2)
library(tidyverse)
library(astsa)
library(forecast)
eccm <- read_csv("~/sta457assignment1/E-Commerce.csv")
AUS<- read_csv("~/sta457assignment1/AUS.csv")
```

Q2(a)

```r
# Q2(a)
# Create a quarterly time series
aus_ts <- ts(AUS$Value, start = c(2005, 1), frequency = 4)

# Fit a ETS model
aus_etsann <- ets(aus_ts, model = "ANN")

# Extract fitted values from the model
fitted_vals <- fitted(aus_etsann)


# Convert the time series and fitted values for a ggplot
time_vals <- time(aus_ts)
plot_df <- data.frame(
  Time = as.numeric(time_vals),
  Observed = as.numeric(aus_ts),
  Fitted = as.numeric(fitted_vals)
```

*https://github.com/anggimude/Time-Series

```
)

# Create the plot
graph1 <- ggplot(plot_df, aes(x = Time)) +
  geom_line(aes(y = Observed, color = "Observed"), size = 1) +
  geom_line(aes(y = Fitted, color = "Fitted"), size = 1) +
  scale_color_manual(values = c("Observed" = "black", "Fitted" = "red")) +
  labs(title = "Observed vs Fitted Values",
       x = "Year",
       y = "Value",
       color = "Legend") +
  theme_minimal()

print(graph1)
```

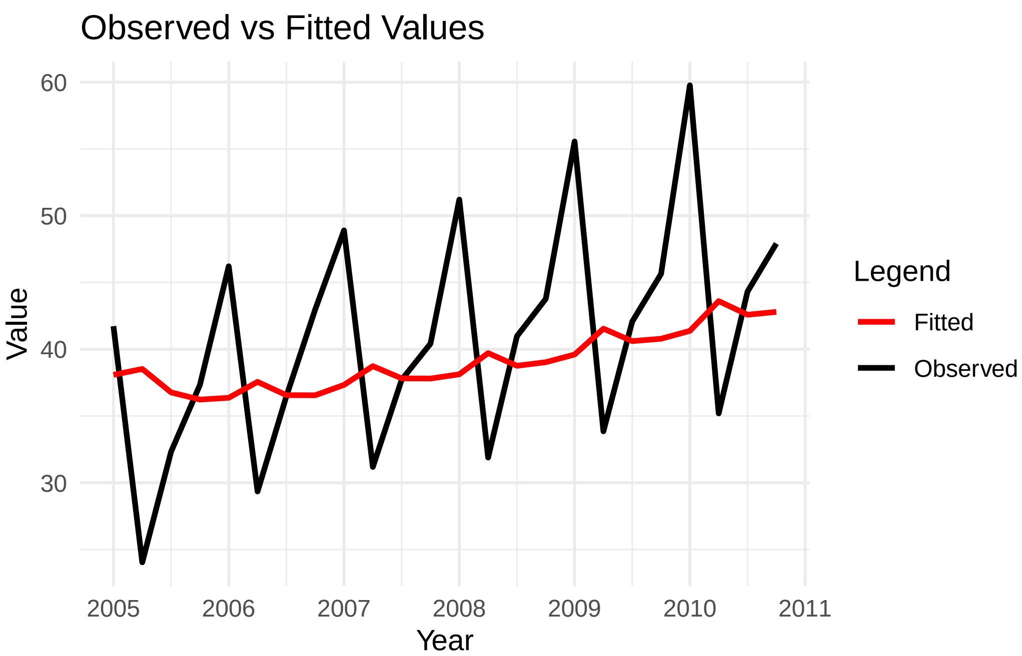


Figure 1: Observed and fitted plot of ETS model

Q2(b)

```r
# Q2(b)
# Calculate the residuals from the ETS model
aus_resid <- residuals(aus_etsann)

# Create a data frame for residuals
resid_df <- data.frame(
  Time = as.numeric(time(aus_resid)),
  Residuals = as.numeric(aus_resid)
)

# Plot the residuals
graph2 <- ggplot(resid_df, aes(x = Time, y = Residuals)) +
  geom_line(color = "blue", size = 1) +
  geom_hline(yintercept = 0, linetype = "dashed", color = "red", size = 1) +
  labs(title = "Residuals from ETS(A,N,N) Model",
       x = "Year",
       y = "Residuals") +
  theme_minimal()


print(graph2)
```

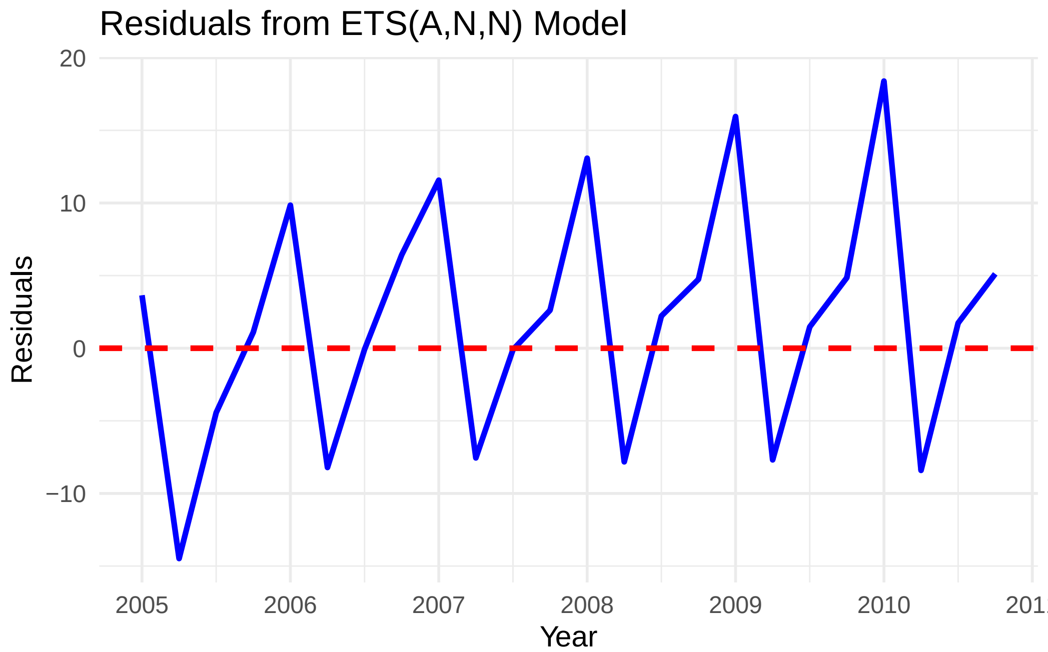


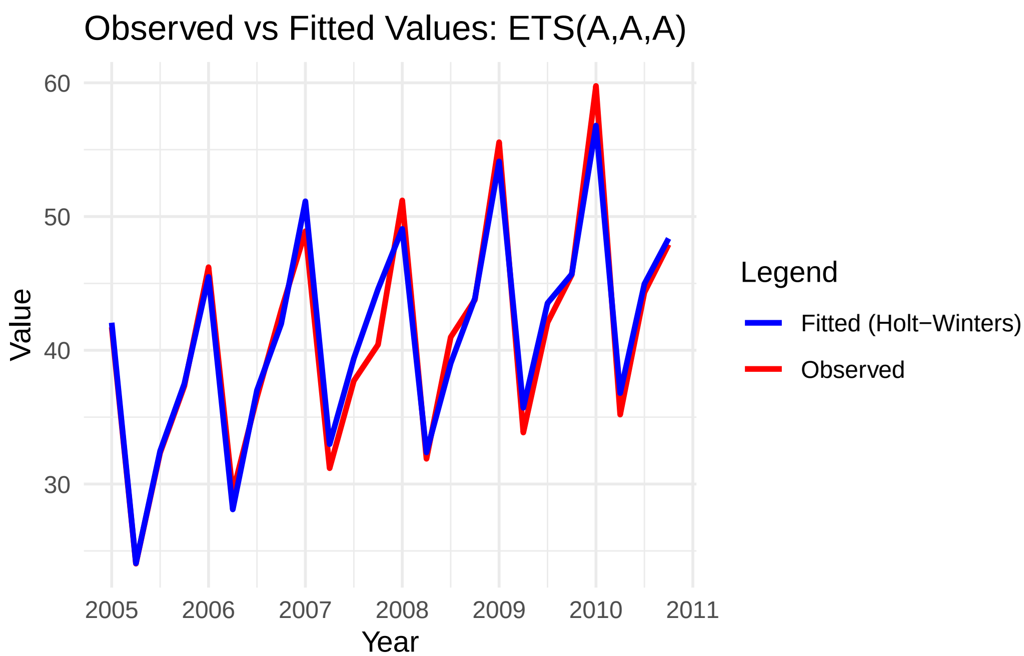
Figure 2: Plot of Residuals

Figure 2 and Figure 1 suggest that the model has some flaws in fitting the model. A well-fitted model generally exhibits randomly scattered residuals around zero, and no obvious trend. However, as we are able to observe in Figure 2, our data shows a clear periodic pattern and is not randomly scattered implying it may not be the best model to fit this data.

Q2(c)

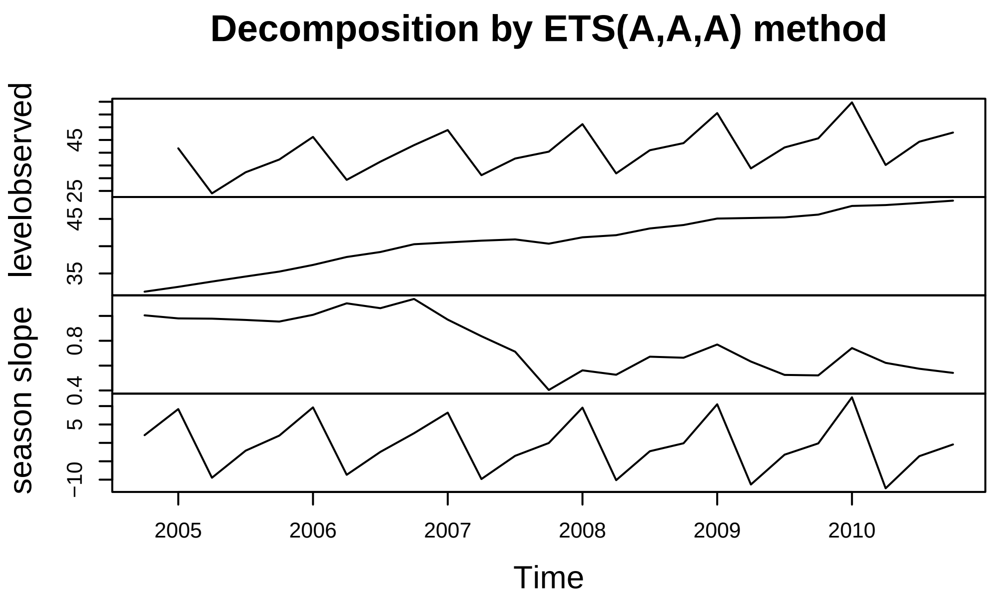
```
# Q2(c)
aus_etsaaa <- ets(aus_ts, model = "AAA")

# Extract fitted values
fitted_vals_aaa <- fitted(aus_etsaaa)

# Create a data frame for plotting
plot_df_aaa <- data.frame(
  Time = as.numeric(time(aus_ts)),
  Observed = as.numeric(aus_ts),
  Fitted   = as.numeric(fitted_vals_aaa)
```

```
)

# Plot observed values and fitted values
graph3 <- ggplot(plot_df_aaa, aes(x = Time)) +
  geom_line(aes(y = Observed, color = "Observed"), size = 1) +
  geom_line(aes(y = Fitted, color = "Fitted (Holt-Winters)"), size = 1) +
  scale_color_manual(values = c("Observed" = "red", "Fitted (Holt-Winters)" = "blue")) +
  labs(title = "Observed vs Fitted Values: ETS(A,A,A)",
       x = "Year",
       y = "Value",
       color = "Legend") +
  theme_minimal()

print(graph3)
```



Q2(d)

```
# Q2(d)-1
# Plot the ETS decomposition
plot(aus_etsaaa)
```




```
# Q2(d)-2
# Extract the fitted smoothing parameters
params <- aus_etsaaa$par
cat("Estimated Smoothing Parameters (alpha, beta, gamma):\n")
```

```
Estimated Smoothing Parameters (alpha, beta, gamma):
```

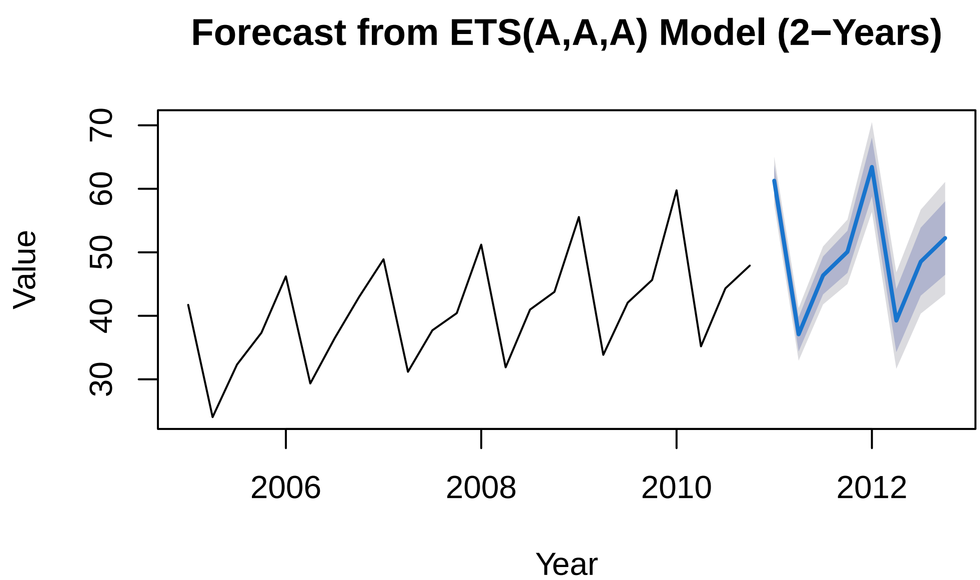
```
print(params)
```

```
      alpha        beta       gamma           l           b          s0
 0.35988288  0.07421998  0.64003047 31.66187756  1.00485461  2.08875079
         s1          s2
-2.02677753 -9.45531280
```

$\alpha = 0.35988$, $\beta = 0.074$, $\gamma = 0.64$.

Q2(e)

```r
# Q2(e)
# Forecast two years ahead
aus_etsaaafc <- forecast(aus_etsaaa, h = 8)

# Plot the forecast along with the original data and the 80%/95% confidence intervals
plot(aus_etsaaafc,
     main = "Forecast from ETS(A,A,A) Model (2-Years)",
     xlab = "Year",
     ylab = "Value")
```



Q3(a)

```r
# Q3(a)
# Load varve
data(varve)

# Determine the length of the series
n <- length(varve)

# Split the data
first_half <- varve[1:(n/2)]
second_half <- varve[(n/2 + 1):n]

# Compute the sample variances
var_first <- var(first_half)
var_second <- var(second_half)

# Print the results
cat("Sample variance of the first half:", var_first, "\n")
```

```
Sample variance of the first half: 133.4574
```

```r
cat("Sample variance of the second half:", var_second, "\n")
```

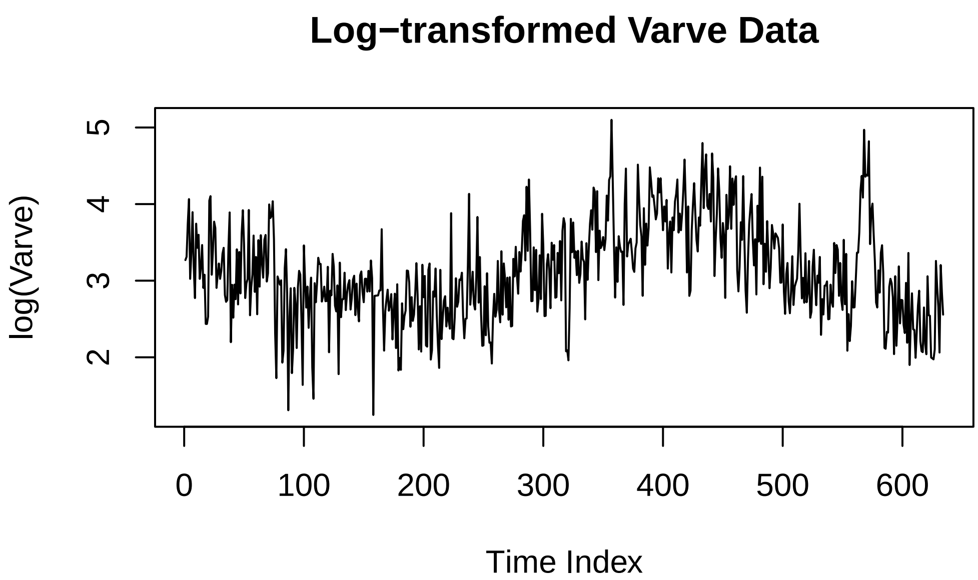
```
Sample variance of the second half: 594.4904
```

From the results of the code, we can observe that the first half has a variance of 133.5 while the second half has a variance of 594.5. This significant difference in variances implies that the time series varve exhibits heteroskedasticity.

Q3(b)

```r
# Q3(b)
# Do the log transformation
y <- log(varve)

# Plot the trajectory
```

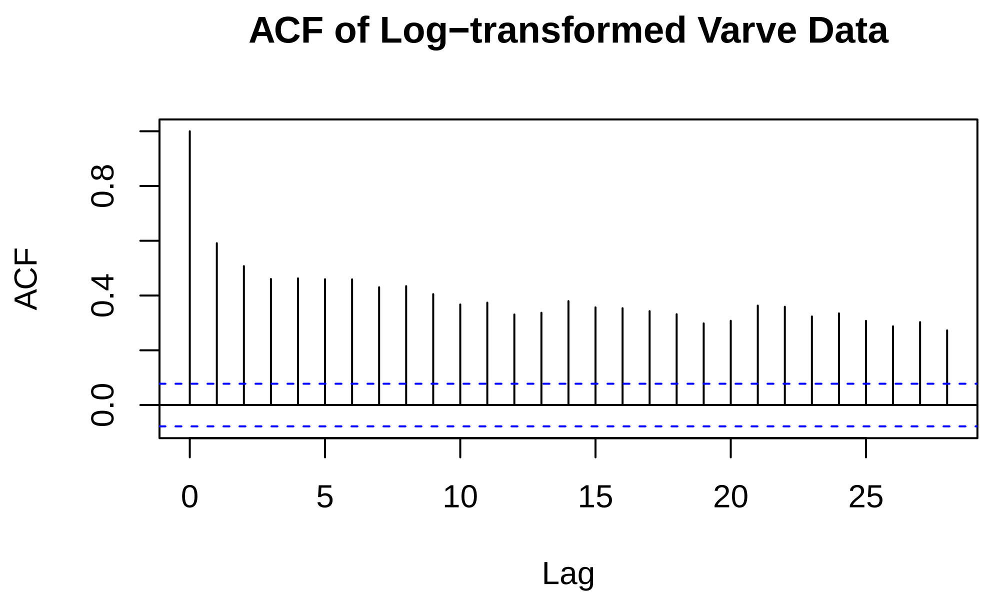
```r
plot(y, type = "l",
     main = "Log-transformed Varve Data",
     xlab = "Time Index",
     ylab = "log(Varve)")
```



The trajectory of the transformed series seems to be more stable than the original series because we are able to observe the less extreme fluctuations; the data is more well-fit with its trend. Though there exists a large spike; compared to the heteroskedasticity observed from part(a), it is much smaller implying the transformation has stabilized the variance.

Q3(c)

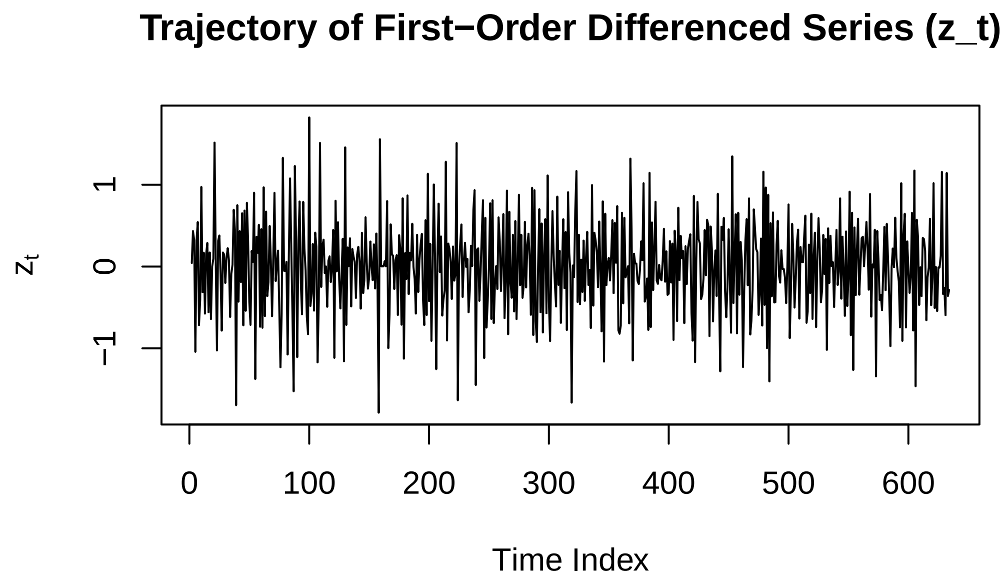
```r
# Q3(c)
# Plot the sample ACF
acf(y, main = "ACF of Log-transformed Varve Data")
```



The plot suggests that the ACF drops gradually conveying there is a moderate level of autocorrelation between the observations. In addition, the decay of the ACF is not rapid and most lags have an autocorrelation above the threshold. These characteristics of the ACF plot is similar to a long-memory process. Thus, the log transformation seems to have stabilized the variance over time but there still persists long-term dependencies, not fully eliminating autocorrelations.

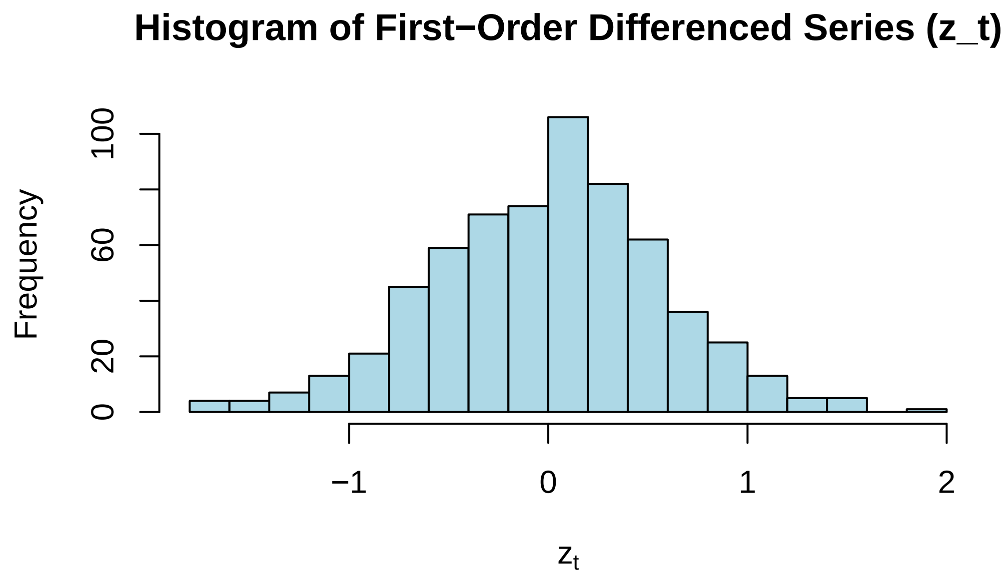
Q3(d)

```
# Q3(d)
# Compute the first-order differences
z <- diff(y)

# Plot z_t
plot(z, type = "l",
     main = "Trajectory of First-Order Differenced Series (z_t)",
     xlab = "Time Index",
     ylab = expression(z[t]))
```

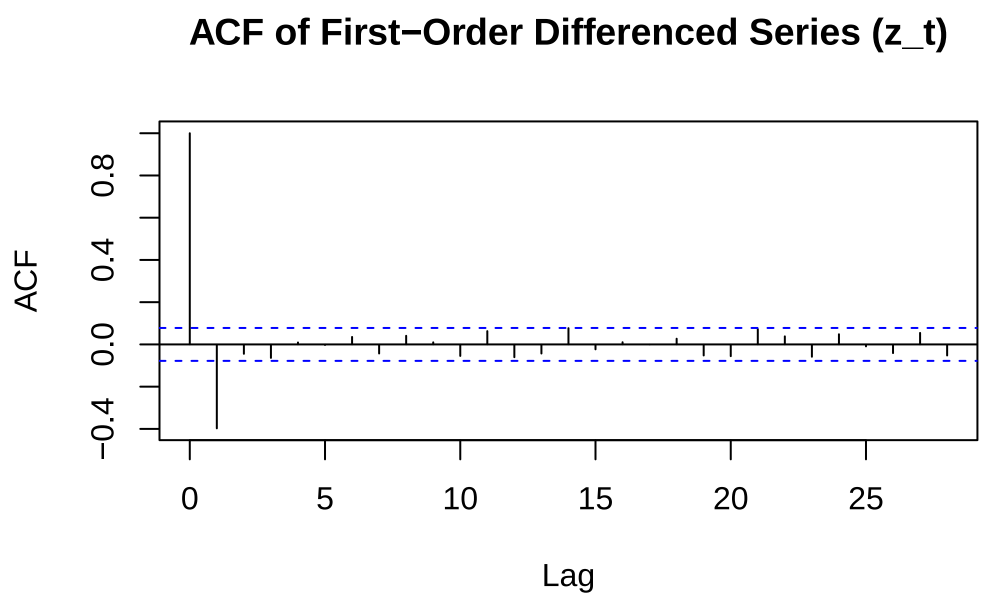


Q3(e)

```r
# Q3(e)
# Plot a histogram of z_t
hist(z, breaks = 20, col = "lightblue", border = "black",
     main = "Histogram of First-Order Differenced Series (z_t)",
     xlab = expression(z[t]))
```




```r
# Q3(e)-1
# Plot the sample ACF of z_t
acf(z, main = "ACF of First-Order Differenced Series (z_t)")
```



The histogram shows that z_t is normally distributed around the mean 0. In addition, the distribution is symmetric and with a relatively small spread, recommends z_t is stationary after differencing. The ACF plot shows a steep drop to near 0 after the first lag implying the dependencies in the original series have been removed which stabilizes the mean and variance. Thus, these characteristics indicates the differenced series z_t is stationary.

Q4

Q4

(a) $z_t = \mu + w_t + \theta w_{t-1}$

$z_t - \mu = w_t + \theta w_{t-1}$

let $x_t = w_t + \theta w_{t-1}$

$\Rightarrow \quad r(h) = cov(x_t, x_{t+h}) = E[x_t . x_{t+h}]$

1. $h=0$, $r(0) = cov(x_t, x_t) = var(x_t) = var(w_t + \theta w_{t-1})$

$= var(w_t) + var(\theta w_{t-1}) = \sigma_w^2 + \theta^2 \sigma_w^2 = \sigma_w^2 (1+\theta^2)$, since

w_t is assumed to be independent with mean 0, variance σ_w^2.

2. $|h|=1$, $r(1) = cov(x_t, x_{t+1}) = cov(w_t + \theta w_{t-1}, w_{t+1} + \theta w_t)$

$= cov(w_t, w_{t+1}) + \theta cov(w_t, w_t) + \theta cov(w_{t-1}, w_{t+1}) + \theta^2 cov(w_{t-1}, w_t)$

$= 0 + \theta \sigma_w^2 + 0 + 0 = \theta \sigma_w^2$

3. $|h| > 1$, let $h=2$.

$r(2) = cov(x_t, x_{t+2}) = cov(w_t + \theta w_{t-1}, w_{t+2} + \theta w_{t+1}) = 0$ since

no w_t term appears twice making all $r(h) = 0 \ \forall |h| > 1$

as w_t is independent.

$\therefore \ r(0) = \sigma_w^2 (1+\theta^2)$, $r(1) = \theta \sigma_w^2$, $r(h) = 0 \ \forall |h| > 1$ //

(b) $\rho(h) = \frac{r(h)}{r(0)}$

$h=0$, $\rho(0) = \frac{r(0)}{r(0)} = 1.$

$|h|=1$, $\rho(1) = \frac{r(1)}{r(0)} = \frac{\theta \sigma_w^2}{\sigma_w^2(1+\theta^2)} = \frac{\theta}{1+\theta^2}$

$|h|>1$, $\rho(h) = \frac{r(h)}{r(0)} = \frac{0}{r(0)} = 0$ //

(c) $\hat{r}(0) = \frac{1}{T}\sum_{t=1}^{T}(z_t - \bar{z})^2 = 0.332$

$\hat{\rho}(1) = \frac{\hat{r}(1)}{\hat{r}(0)} = -0.397$

(d) from (a), (b) we have proven $\rho(1) = \frac{\theta}{1+\theta^2}$

Then we can write $\hat{\rho}(1) = \frac{\theta}{1+\theta^2}$, $(1+\theta^2)\hat{\rho}(1) = \theta$

$\Rightarrow \hat{\rho}(1)\theta^2 - \theta + \hat{\rho}(1) = 0$. $\theta = \frac{1 \pm \sqrt{1-4\hat{\rho}(1)^2}}{2\hat{\rho}(1)}$

$\hat{\rho}(1) = -0.397 \Rightarrow \theta = \frac{1 \pm \sqrt{0.3696}}{-0.794}$

$\Rightarrow \theta = -2.03$ or $\theta = -0.494$

Since the MA process needs to be invertible,

we want $|\theta| < 1 \Rightarrow \theta = -0.494$

Similarly $r(0) = (1+\theta^2)\sigma_w^2$, $\hat{r}(0) = (1+\theta^2)\sigma_w^2$

$\sigma_w^2 = \frac{\hat{r}(0)}{1+\theta^2} = \frac{0.332}{1+(-0.494)^2} = 0.267$

$\therefore \theta = -0.494$, $\sigma_w^2 = 0.267$ //