STAT 443: Homework 4

Aronn Grant Laurel (21232475)

3 April, 2025

Question 1

(a & b)

```r
knitr::include_graphics("q1.jpg")
```

1. Consider the AR(2) process from Assignment 2 (Q1):

$$X_t = 0.1X_{t-1} + 0.2X_{t-2} + Z_t, \quad \{Z_t\}_{t\in\mathbb{Z}} \sim WN(0, \sigma^2),$$

and recall that its autocorrelation function has the form

$$\rho(h) = \frac{15}{36}(-0.4)^{|h|} + \frac{21}{36}0.5^{|h|}, \quad h \in \mathbb{Z}.$$

(a) Derive the normalized spectral density function of $\{X_t\}_{t\in\mathbb{Z}}$.

(b) Write down the (power) spectral density function of $\{X_t\}_{t\in\mathbb{Z}}$.

(c) Plot the spectral density and comment on its behaviour.

$$\sum_{h=1}^{\infty} r^h \cos(hw) = \frac{r\cos(w) - r^2}{1 - 2r\cos(w) + r^2}$$

a) Given $f^*(w) = \frac{1}{\pi}(1 + 2\sum_{k=1}^{\infty} \rho(k)\cos(wk))$

$$= \frac{1}{\pi}\left[(1 + 2\sum_{k=1}^{\infty}\left(\frac{15}{36}(-0.4)^{|k|} + \frac{21}{36}(0.5)^{|k|}\right)\cos(wk)\right]$$

$$= \frac{1}{\pi}\left(1 + \frac{5}{6}\sum_{k=1}^{\infty}(-0.4)^{|k|}\cos(wk) + \frac{7}{6}\sum_{k=1}^{\infty}(0.5)^{|k|}\cos(wk)\right)$$

$$= \frac{1}{\pi}\left(1 + \frac{5}{6}\left(\frac{-0.4\cos(w) - 0.16}{1 - 0.8\cos(w) + 0.16}\right) + \frac{7}{6}\left(\frac{0.5\cos(w) - 0.25}{1 - \cos(w) + 0.25}\right)\right)$$

b) $\sigma_X^2 = Var(X_t)$

Using Yule-Walker's Equation:

$\gamma(0) = 0.1\gamma(1) + 0.2\gamma(2) + \sigma^2$ — (1)

$\gamma(1) = 0.1\gamma(0) + 0.2\gamma(1)$ — (2)

$\gamma(2) = 0.1\gamma(1) + 0.2\gamma(0)$ — (3)

In (2): $0.8\gamma(1) = 0.1\gamma(0)$

$\gamma(1) = 0.125\gamma(0)$

In (3) ∧ (2): $\gamma(2) = 0.1(0.125\gamma(0)) + 0.2\gamma(0)$

$\gamma(2) = 0.2125\gamma(0)$

In (1), $\gamma(0) = 0.1(0.125\gamma(0)) + 0.2(0.2125\gamma(0)) + \sigma^2$

$\gamma(0) = \sigma^2/(1 - 0.1\times0.125 + 0.2\times0.2125) = \frac{\sigma^2}{0.945}$

$\therefore f^*(w) = \frac{f(w)}{\sigma_X^2}$

$$f(w) = \sigma_X^2 f^*(w) = \frac{\sigma_X^2}{\pi}\left(1 + \frac{5}{6}\left(\frac{-0.4\cos(w) + 0.16}{-0.8\cos(w) + 1.16}\right) + \frac{7}{6}\left(\frac{0.5\cos(w) - 0.25}{-\cos(w) + 1.25}\right)\right)$$

$$= \frac{1}{0.95\pi}\left(1 + \frac{5}{6}\left(\frac{-0.4\cos(w) + 0.16}{-0.8\cos(w) + 1.16}\right) + \frac{7}{6}\left(\frac{0.5\cos(w) - 0.25}{-\cos(w) + 1.25}\right)\right)$$

(c)

```r
# Calculate the spectral density
spectral_density <- function(omega) {
  term2 <- (15/18) * ((-0.4*cos(omega) - 0.16) / (1.16 + 0.8 * cos(omega)))
  term3 <- (21/18) * ((0.5*cos(omega) - 0.25) / (1.25 - cos(omega)))

  return((1/(0.95*pi)) * (1 + term2 + term3))
}

# Create a sequence of frequencies from 0 to pi
omega <- seq(0, pi, length.out = 1000)
```

```r
# Calculate the spectral density for each frequency
spec_values <- sapply(omega, spectral_density)

# Create a data frame for plotting
spec_df <- data.frame(
  omega = omega,
  frequency = omega/(2*pi),
  spectral_density = spec_values
)


# Plot using ggplot2
ggplot(spec_df, aes(x = omega, y = spec_values)) +
  geom_line() +
  labs(title = "Spectral Density : AR(2) Process",
       x = "Frequency",
       y = "f(w)") +
  theme_minimal()
```

We can the spectral density being dominated by low frequencies, with the spectral density is at highest at frequency = 0 as it's peak. As we move further away from frequency 0, we observe a dip / decreasing spectral density and another smaller peaks at approximately frequency 3. Overall, this suggests that our AR(2) process has long term dependencies.

Question 2

```r
knitr::include_graphics("q2.jpg")
```

2. Given the power spectral density function

$$f(\omega) = \frac{1}{\pi}\Big(1.1 + 0.54\cos(\omega) - 0.2\cos(2\omega)\Big), \qquad \omega \in (0, \pi),$$

compute the autocovariance function $\gamma(k)$ of the underlying stochastic process for $k = 0, 1, 2$ and $k > 2$.

Given $f(w) = \frac{1}{\pi}\left(\gamma(0) + 2\sum_{i=1}^{\infty}\gamma(k)\cos(kw)\right) = \frac{1}{\pi}\left(\gamma(0) + 2\gamma(1)\cos(w) + 2\gamma(2)\cos(2w)\right)$

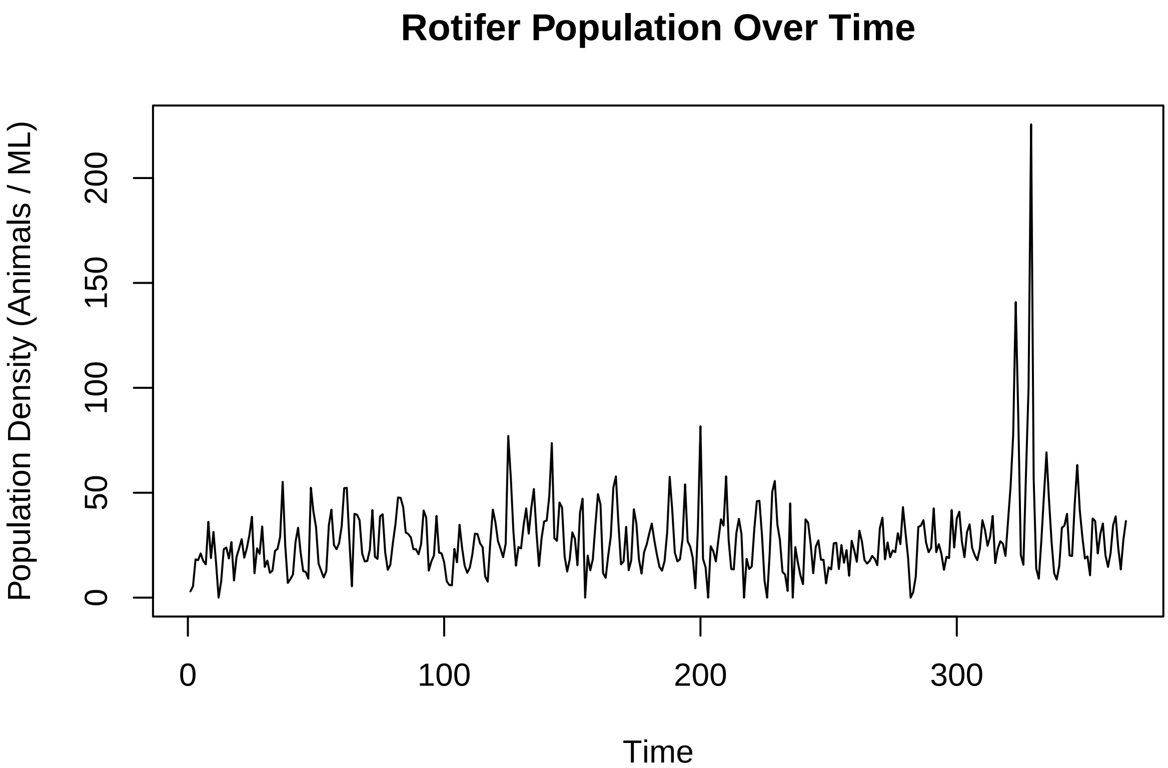
$\therefore$

$$\gamma(K) = \begin{cases} 1.1 & \text{if } k=0 \\ 0.27 & \text{if } k=1 \quad \longleftarrow \quad 2\gamma(1) = 0.54,\ \gamma(1) = 0.27 \\ -0.1 & \text{if } k=2 \quad \longleftarrow \quad 2\gamma(2) = -0.2,\ = -0.1 \\ 0 & \text{if } k>2 \end{cases}$$

Question 3

(a)

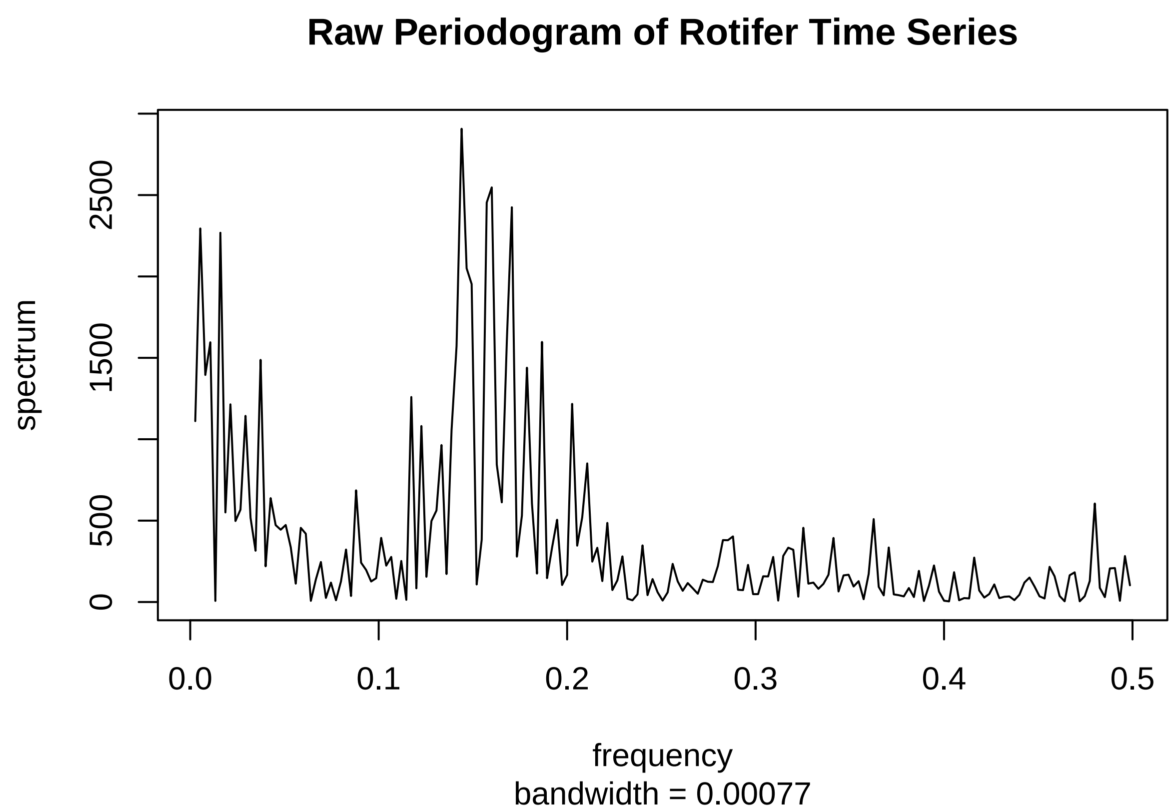
```r
predator<- read.csv("predator_prey.csv")

rot_ts <- ts(predator$rotifers..animals.ml.)

plot(rot_ts,
     main = "Rotifer Population Over Time",
     ylab = "Population Density (Animals / ML)"
     )
```



We observe a consistent fluctuation until at time ~350 where we observe a large spike / peak before dropping back to the consistent fluctuation.

(b)

```
raw_spec <- spectrum(rot_ts,
                     log = "no",
                     main = "Raw Periodogram of Rotifer Time Series")
```




```
frequencies <- raw_spec$freq
spectrum_values <- raw_spec$spec

# Get the max
max_index <- which.max(spectrum_values)

dominant_freq <- frequencies[max_index]
dominant_freq
```

```
## [1] 0.144
```

```
angular_freq <- 2 * pi * dominant_freq
angular_freq
```

```
## [1] 0.9047787
```

```
wavelength_days <- 1 / dominant_freq
wavelength_days
```

```
## [1] 6.944444
```

For our dominating frequency 0.144, we get wavelength (days) ~ 6.944 and angular frequency ~ 0.905

(c)

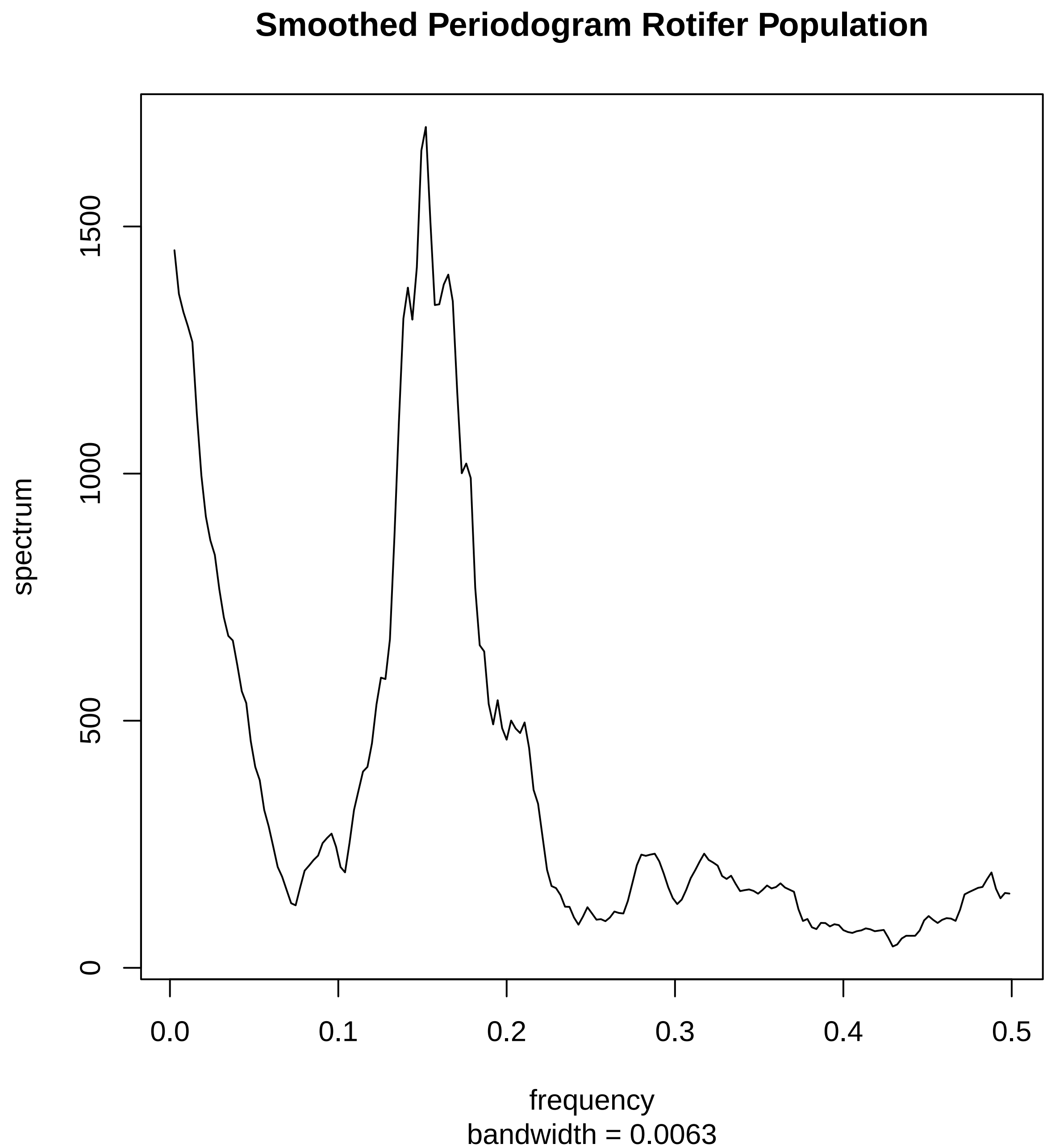
```r
N <- length(rot_ts)

num_cycles <- N / wavelength_days
num_cycles
```

```
## [1] 52.704
```

On a cycle length of approximately 7 days, we get the predator prey cycles to be approximately 52.7

(d)

```r
spec_smooth <- spec.pgram(rot_ts,
                          spans = c(366 / 40), # As per piazza response @116
                          log = "no",
                          main = "Smoothed Periodogram Rotifer Population")
```



We may want to change the raw periodogram since we know it is not a consistent estimator of the spectral density function and change it to an estimator that is more consistent. As per piazza response @116, I chose N / 40 over 2 *sqrt(N) for my span smoothing because I believe it includes several peaks and dips relatively better.

(e)

```
N <- length(rot_ts) # 366

omega_59 <- (2 * pi * 59) / N
omega_59
```

```
## [1] 1.012863
```

Our Fourier Frequency for p = 59 is 1.012863.

(f)

```r
N <- length(rot_ts)  # 366

# Create an empty data frame to store significant frequencies
significant_freqs <- data.frame(
  p = integer(),
  freq = numeric(),
  angular_freq = numeric(),
  wavelength = numeric(),
  F_stat = numeric(),
  p_value = numeric()
)

# Loop through potential frequencies (Fourier components)
for (p in 1:(N/2)) {
  t <- 1:N  # Time vector
  omega_p <- 2 * pi * p / N  # angular frequency

  # Fit the linear model
  model <- lm(rot_ts ~ cos(omega_p * t) + sin(omega_p * t))

  # Extract F-statistic and p-value from model summary
  model_summary <- summary(model)
  f_stat <- model_summary$fstatistic[1]
  p_value <- pf(f_stat,
              model_summary$fstatistic[2],
              model_summary$fstatistic[3],
              lower.tail = FALSE)

  # Store results if significant at 99% confidence level (p < 0.01)
  if (p_value < 0.01) {
    significant_freqs <- rbind(significant_freqs, data.frame(
      p = p,
      freq = p / N,
      angular_freq = omega_p,
      wavelength = N / p,
      F_stat = f_stat,
      p_value = p_value
    ))
  }
}

# Sort significant frequencies by F-statistic (DESC)
significant_freqs <- significant_freqs[order(-significant_freqs$F_stat), ]

print(significant_freqs)
```

```
##          p        freq angular_freq wavelength   F_stat      p_value
```

```
## value   2 0.005464481   0.03433435 183.000000 9.261036 0.0001194628
## value5 59 0.161202186   1.01286321   6.203390 8.744184 0.0001954750
## value2  4 0.010928962   0.06866869  91.500000 6.947354 0.0010942468
## value4 54 0.147540984   0.92702734   6.777778 6.495157 0.0016923460
## value3 52 0.142076503   0.89269299   7.038462 6.255040 0.0021341862
## value6 63 0.172131148   1.08153190   5.809524 6.178262 0.0022986471
## value1  3 0.008196721   0.05150152 122.000000 5.921789 0.0029462120
```

(g)

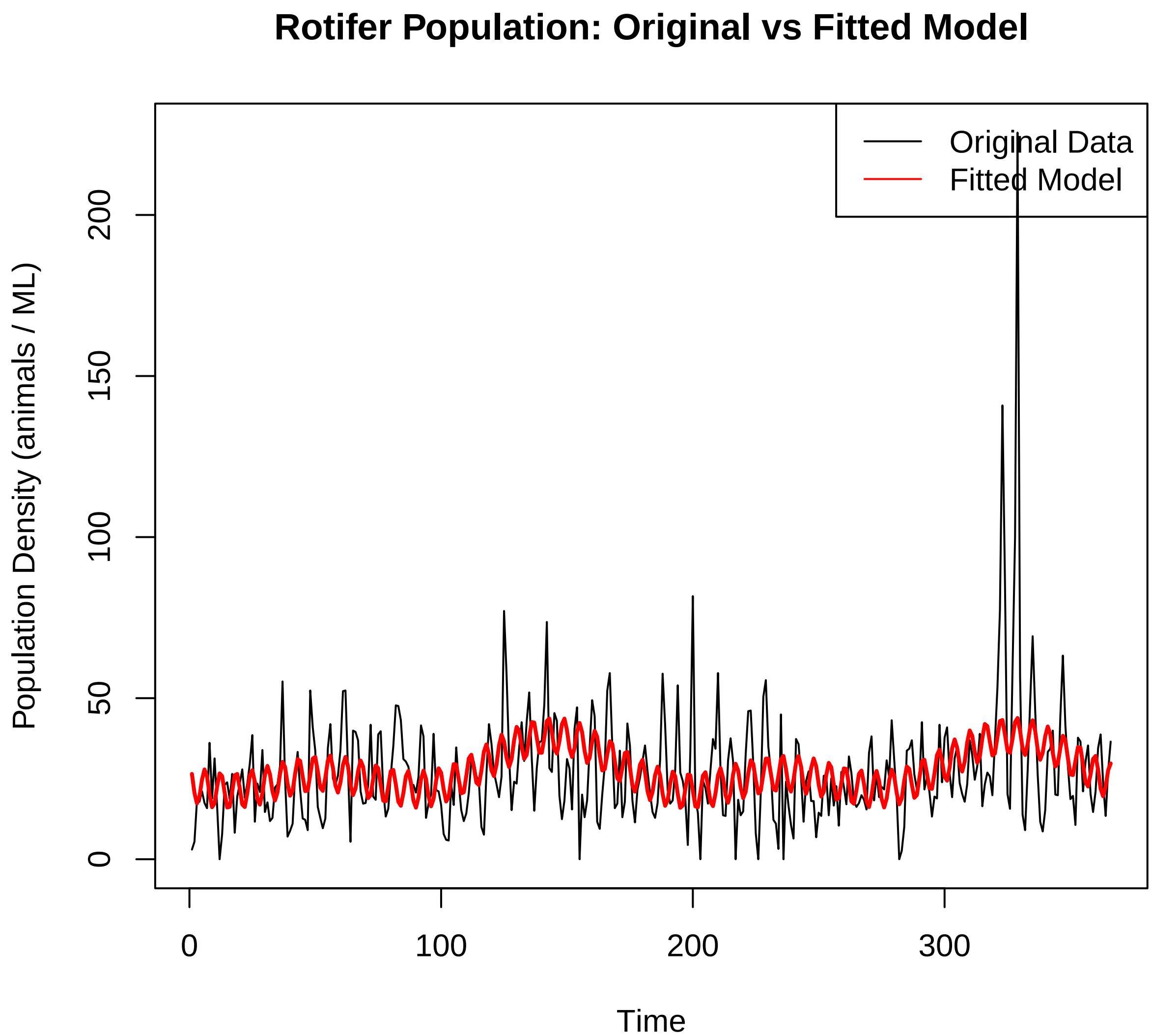
```r
top_freqs <- head(significant_freqs, 3) # top 3 significant frequencies
N <- length(rot_ts)  # 366
t <- 1:N  # Time index

# Start building the formula as a string
formula_terms <- c()

# Loop through each top frequency and construct sin-cos terms
for (i in 1:nrow(top_freqs)) {
  p <- top_freqs$p[i]
  omega_p <- 2 * pi * p / N  # angular frequency

  formula_terms <- c(
    formula_terms,
    paste0("cos(", omega_p, " * t)"),
    paste0("sin(", omega_p, " * t)")
  )
}


formula_str <- paste("rot_ts ~", paste(formula_terms, collapse = " + "))
data_frame <- data.frame(t = t, rot_ts = rot_ts)

# Fit linear model
model_fit <- lm(as.formula(formula_str), data = data_frame)

summary(model_fit)
```

```
##
## Call:
## lm(formula = as.formula(formula_str), data = data_frame)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -42.302 -10.218  -1.337   7.060 181.717
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   27.3898     0.9274  29.533  < 2e-16 ***
## cos(0.0343343459408721 * t)    1.1511     1.3116   0.878 0.380713
## sin(0.0343343459408721 * t)   -5.7612     1.3116  -4.393 1.48e-05 ***
## cos(1.01286320525573 * t)      5.7107     1.3116   4.354 1.75e-05 ***
## sin(1.01286320525573 * t)     -0.2595     1.3116  -0.198 0.843252
```

```
## cos(0.0686686918817441 * t)  -4.5184     1.3116  -3.445 0.000639 ***
## sin(0.0686686918817441 * t)  -2.4074     1.3116  -1.835 0.067263 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 17.74 on 359 degrees of freedom
## Multiple R-squared:  0.1314, Adjusted R-squared:  0.1169
## F-statistic:  9.05 on 6 and 359 DF,  p-value: 3.187e-09
```

(h)

```r
# Prediction
pred_vals <- predict(model_fit)

plot(rot_ts,
     ylab="Population Density (animals / ML)",
     main="Rotifer Population: Original vs Fitted Model")


lines(pred_vals, col="red", lwd=2)
legend("topright", legend=c("Original Data", "Fitted Model"),
col=c("black", "red"), lty=1)
```


Rotifer Population: Original vs Fitted Model
Original Data
Fitted Model
Population Density (animals / ML)
0
50
100
150
200
0
100
200
300
Time