1

Seasonality

Setup

Given a time series x_t , we conceptually split it into two parts:

Decomposition: $x_t = f(t) + e_t$

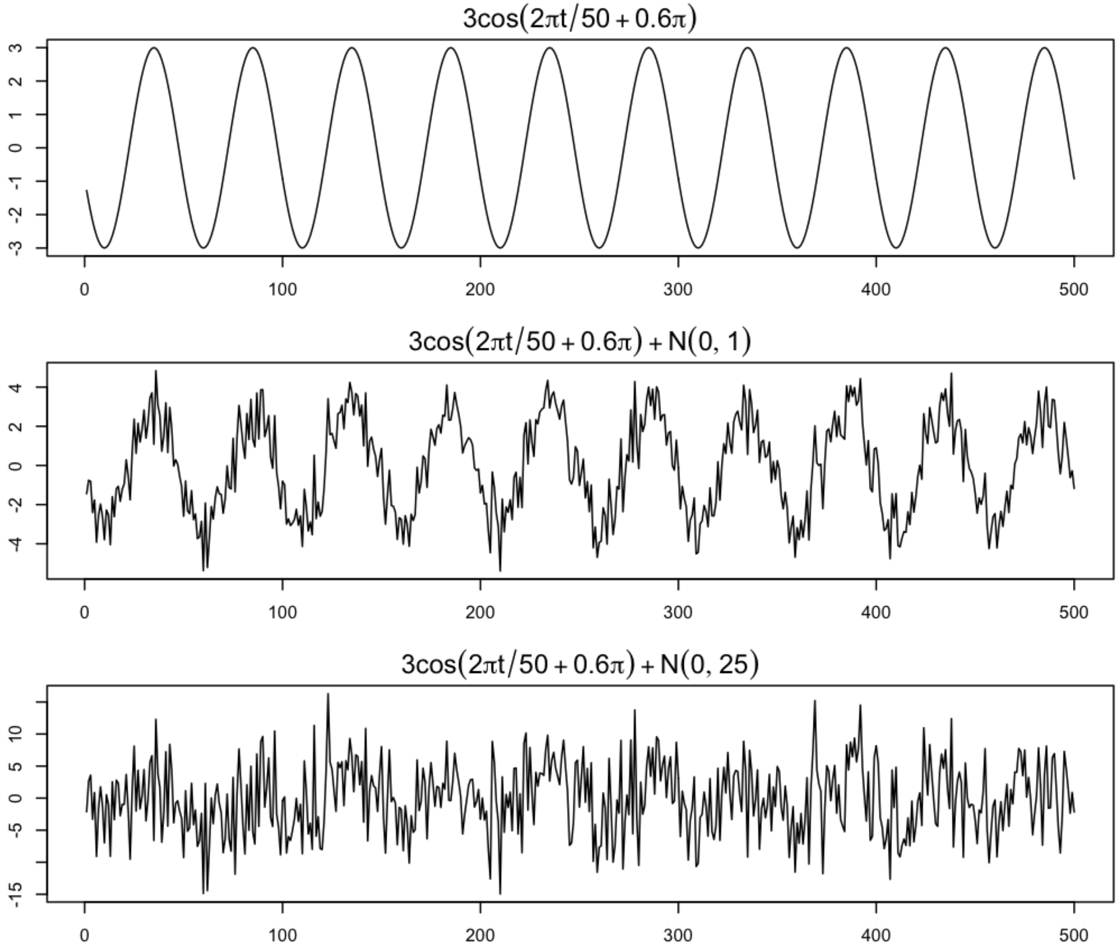
- "Signal": $f(t)$
  - The "interesting part"
- "Noise": e_t
  - "Everything else"
  - Usually, want e_t to be white noise

Purpose of many time series models is to decompose the time series to understand the underlying trend.

Seasonality

Signal-to-noise ratio (SNR): Ratio of amplitude of the signal to error

The larger the SNR, the easier it is to detect the signal



Reference

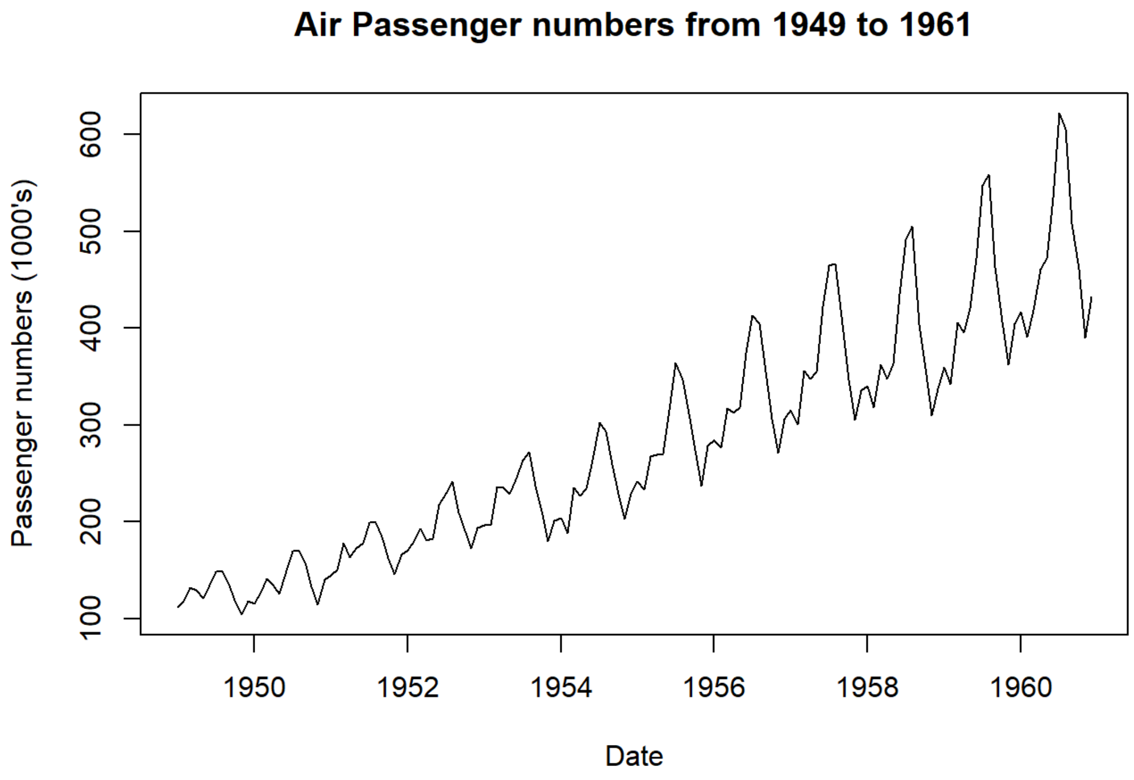
Seasonality

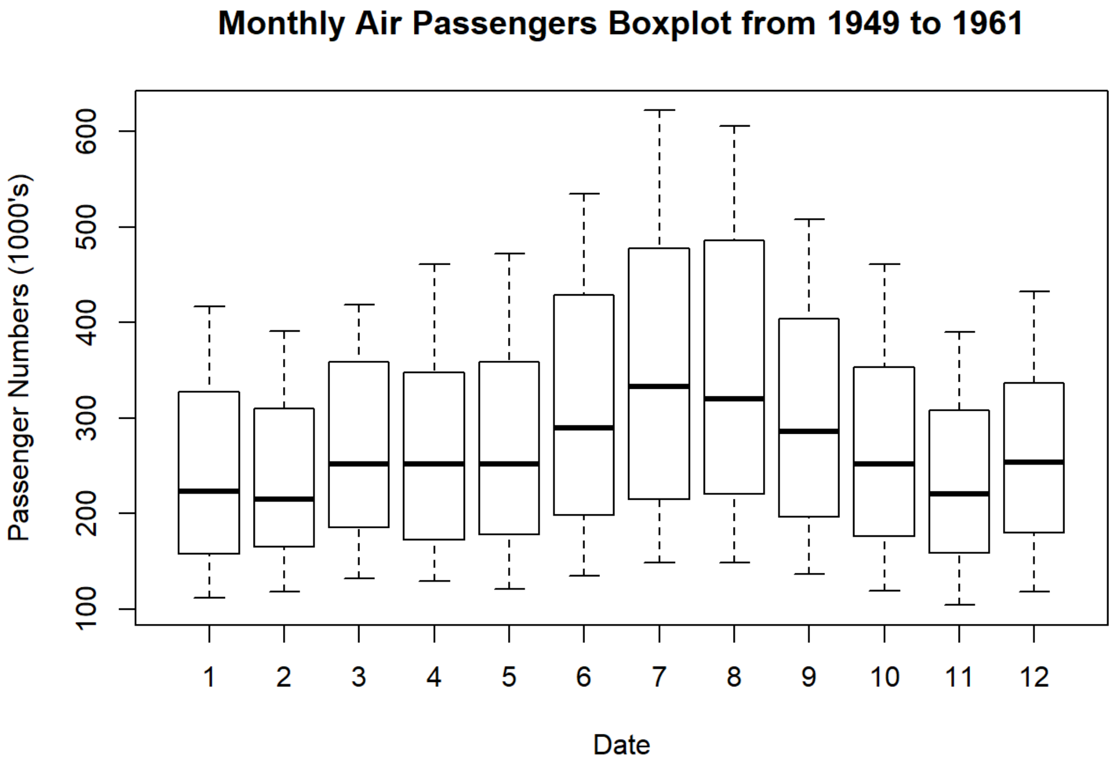
If the signal $f(t)$ repeats itself after some time d, we say it is seasonal with period d.

- Important: Signal might have both a non-seasonal and seasonal components:
  - $f(t) = m(t) + s(t)$ || signal = non-seasonal + seasonal
- Examples of seasonality:
  - Weather patterns
  - Sleep cycles
  - Sound waves

Seasonality (cont.)

Here's an actual time series that strongly exhibits seasonality





Reference

Indicator Variables

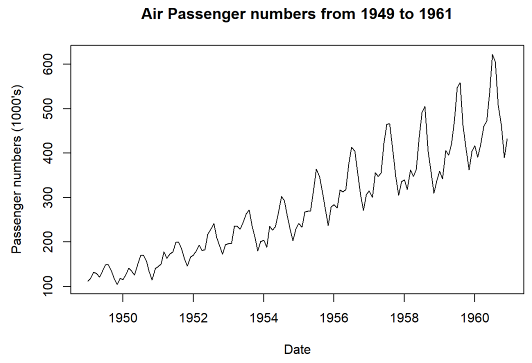
If our signal is (partly) seasonal, how to we incorporate this into our model?

- Fix: use indicator variables to capture seasonal effects
  - An indicator variable I_A is a binary variable associated with a statement A.
  - $I_A = 1$ when A is true
  - $I_A = 0$ when A is false
    - ex. If A = "today is August", then I_A=1 only on August and is 0 all other days.

Seasonality

Indicator Variable Example

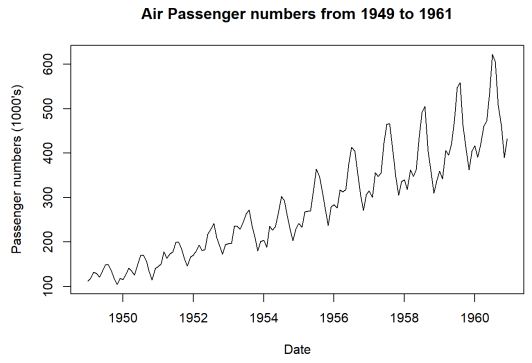
Ex. Modeling number of air passengers by month.



- Setup:
  - $x_t = m(t) + s(t) + e_t$
  - $m(t)$: non-seasonal component. Includes effects from factors like increase in demand for flight travel, commercialization of airlines.
  - $s(t)$: seasonal component. Flight volume spikes around holidays and other seasonal events.
  - e_t: noise

Indicator Variable Example (cont.)

Model: $f(t) \approx \beta_0 + \beta_1 t + \alpha_{Feb} I_{Feb} + \ldots + \alpha_{Dec} I_{Dec}$

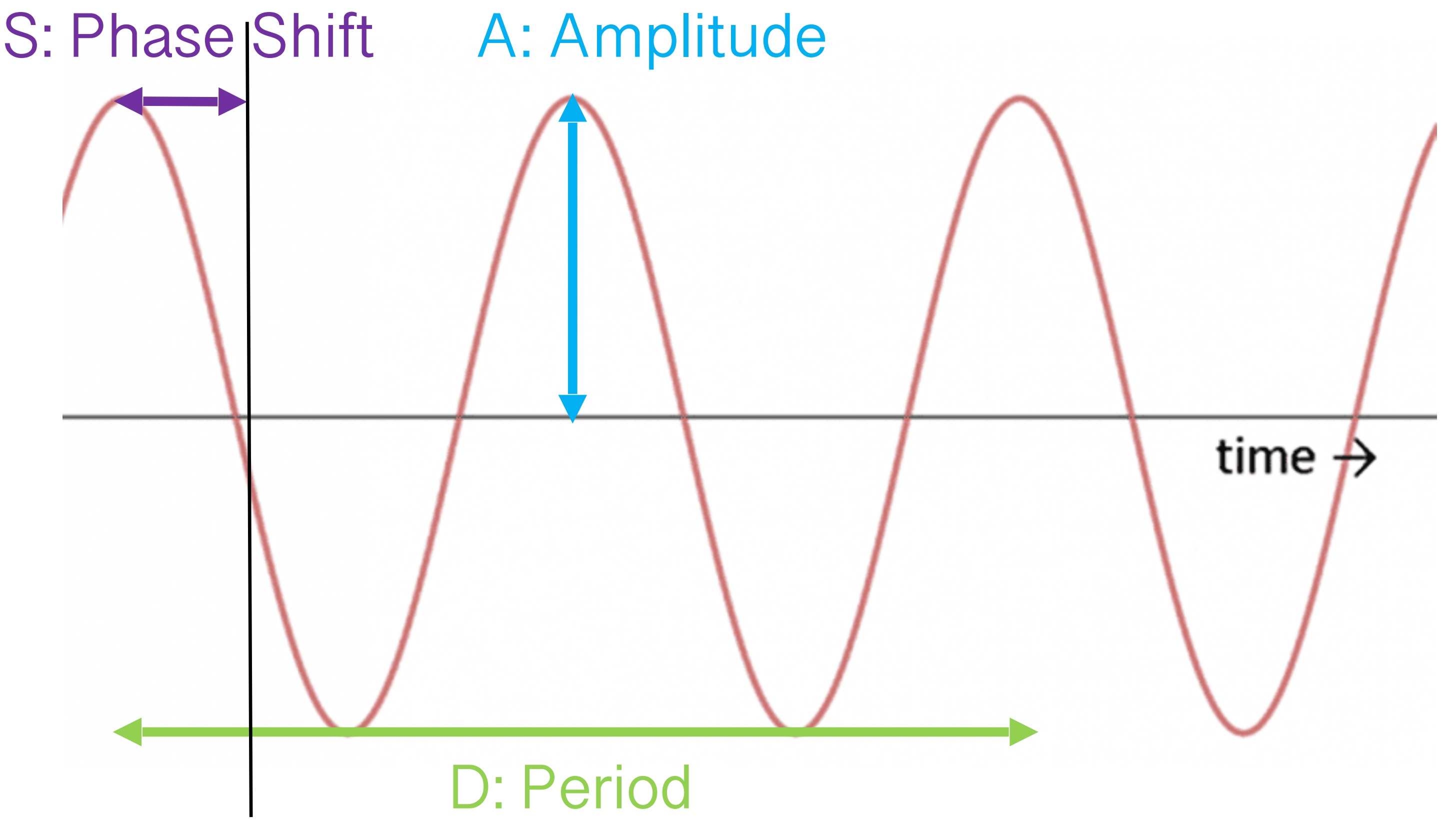


- Non-seasonal component: model this as a linear trend
- Seasonal component: linear function of the indicator variables.
    - Note that we can either:
        - Regress on all month indicators, omitting the intercept.
        - Regress on an intercept and the month indicators, omitting one.
    - α_{Feb} represents the difference in #passengers in February and β_0 (which can be interpreted as the seasonality in the omitted month).

2 Sinusoids

Sinusoid Parameters

Sine and cosine are the two sinusoidal functions we'll be considering.



$$g(t) = A\ sin\left(\frac{2\pi F(t - S)}{D}\right)$$

- A is the amplitude
- S is the phase shift
- D is the period
- F is the frequency
  - This sine wave will repeat after D time, its period. However, within a period it will repeat F times.

Sinusoid Application

If our signal is (partly) seasonal, how to we incorporate this into our model?

Fix: Use sinusoids to model the season part

$$s(t) \approx \sum_i a_i \sin\left(\frac{2\pi F_i t}{D}\right) + b_i \cos\left(\frac{2\pi F_i t}{D}\right)$$

Considerations

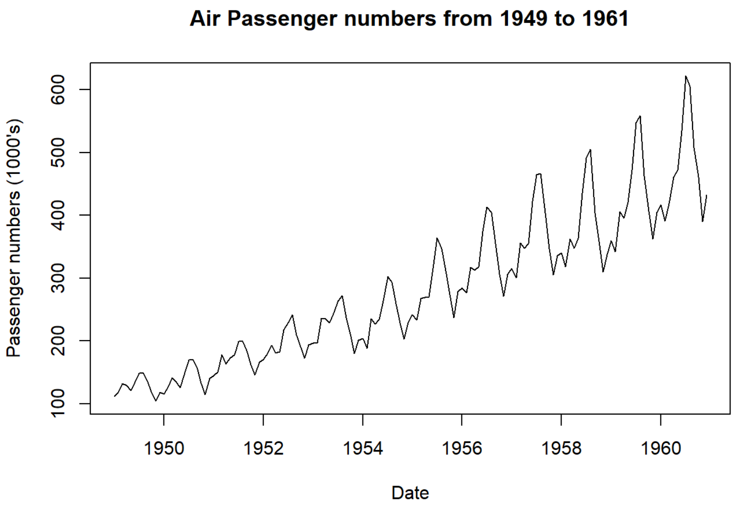
- Need to figure out the (approximately) correct D
- Which frequencies F_i to use?

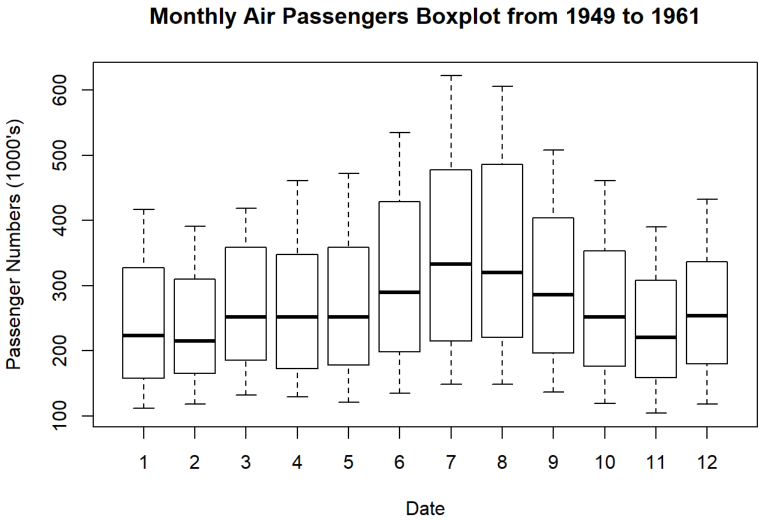
Seasonality

Combining Indicator Variable and Sinusoid

We can use both sinusoid and indicator variable to capture seasonality:

$$f(t) \approx \beta_0 + \beta_1 t + \sum_i a_i sin\left(\frac{2\pi F_i t}{D}\right) + b_i cos\left(\frac{2\pi F_i t}{D}\right) + \alpha_{Holiday} I_{Holiday}$$

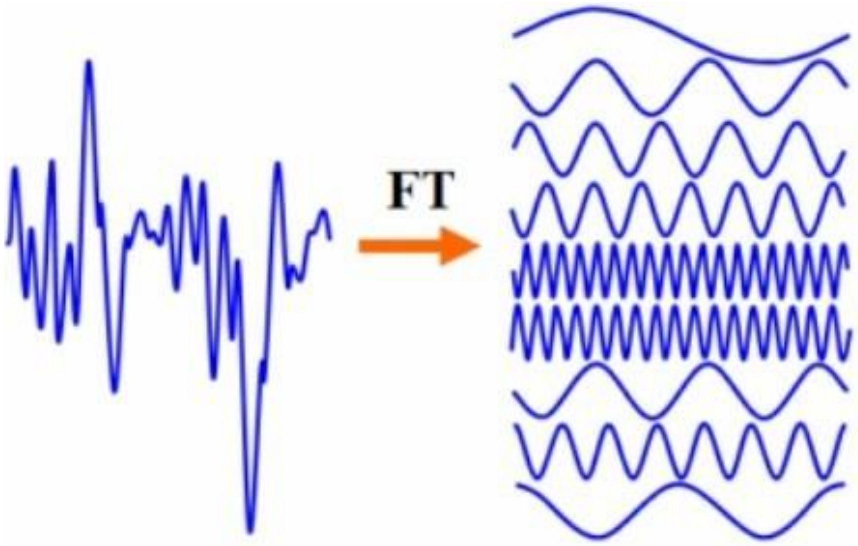




1 Periodogram

Discrete Fourier Transform

- **What is it?**: Loosely, decomposes the input into a weighted sum of different frequency waves



$$input = \alpha_0 Wave_0 + \cdots + \alpha_{n-1} Wave_{n-1}$$

*Note each wave has its own coefficient α_i

DFT

$$input = \alpha_0 Wave_0 + \cdots + \alpha_{n-1} Wave_{n-1}$$

- Is an invertible linear map: $C^n \to C^n$.
  - Can think of it as a orthogonal change of basis
  - Input: some vector in C^n
  - Output: the coefficients $\alpha_0, \cdots, \alpha_{n-1}$
- Fourier waves: $w_j(t) = e^{\frac{i2\pi jt}{n}}$ for $j = 0,1,\cdots,n-1$
  - Above formula gives the t^{th} entry for the j^{th} wave vector.
  - Using Euler's identity, $e^{ix} = \cos(x) + i\sin(x)$, can think of Fourier waves as ordinary sinusoids in the real/imaginary part

Properties

- If the input is dominated by a single (or a few) frequency, then the magnitude of its coefficient will be the largest relatively.
- The periodogram plots the magnitude of the Fourier coefficients
  - Hence, periodogram will "detect the dominant frequencies" in the input
- Periodogram: we stop plotting at frequency $\frac{n}{2}$ because, for real inputs, the Fourier coefficients will come in conjugate pairs
  - Conjugates have the same magnitude.
    - The full periodogram will be symmetric about $\frac{n}{2}$, so no need to plot past that