# Day 3 - Decomposition of time series

## Introduction

The purpose of today's lecture is to understand how to decompose a time series into its constituent components in R. To guide our exploration, we will return to the Vermont temperatures and Pan Am data sets.

```
# packages
library(tidyverse)

# load data and rename
## Pan Am
data(AirPassengers)
ap <- AirPassengers

## vt temps
vt_temps <- readr::read_csv("vt_temps.csv")
```

---

## Review: creating `ts` objects

Create a `ts` object, called `vt_ts`, for the monthly temperatures in Vermont that spans from 1970/06/01 to 2013/04/01. Plot the time series.

```r
# alternative using window()
vt_ts_long <- ts(
  vt_temps$AverageTemperature,
  start = c(1850, 1),
  end = c(2013, 9),
  freq = 12
)

vt_ts <- window(
  vt_ts_long,
  start = c(1970, 6),
  end = c(2013, 4)
)
```

– filtering

## Introducing definitions and notation

**Random variables**

A random variable, usually written $X$, is a variable whose possible values are the numerical outcomes of random phenomenon. There are two types of random variables, discrete and continuous.

Formally, a random variable is a mapping from a sample space $S$ to the real numbers.

**Discrete random variables:**

- # of dice landing on 3
- # of students coming to lecture

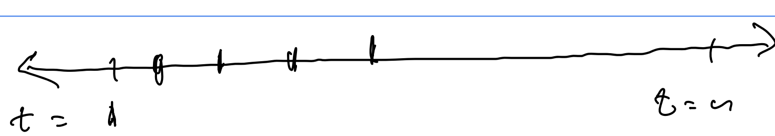
**Continuous random variables:**

- temp in VT

**Time series notation**

A discrete time series of length $n$ is a sequence of random variables, which we denote $\{X_t : t = 1, \dots, n\} = \{X_1, X_2, \dots, X_n\}$. When referring to an observed time series, we use lowercase letters, $\{x_t : t = 1, \dots, n\} = \{x_1, x_2, \dots, x_n\}$. If the length of the series $n$ does not need to be specified, we will often use the abbreviated notation $\{x_t\}$.

discrete ts:




continuous ts:

- $\bar{x} = \frac{\sum x_i}{n}$ "x-bar", mean

- $\hat{x}$ "x-hat", predicted value of $x$

  $\hat{y}_i = b_0 + b_1 x_i$

- $\hat{x}_{t+k|t}$ predicted value of x at time t+k, given time t

  conditional

  "forecast"

## Our first time series model

**Decomposition models**

An additive decomposition model is a simple model for a time series that estimates the trend, seasonal effect, and error.

$$x_t = m_t + s_t + z_t$$

$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

A multiplicative allows for the seasonal effect to increase as the trend increase.

$$x_t = m_t \cdot s_t \cdot z_t$$

If the time series is strictly positive, it may be easier to fit an additive model on the log scale than a multiplicative model on the original scale.

$$\log(x_t) = m_t + s_t + z_t$$

$$\log(x_t) = \log(m_t \cdot s_t \cdot z_t)$$

$$= \log(m_t) + \log(s_t) + \log(z_t)$$

## Estimating $m_t$, $s_t$, and $z_t$

How can we obtain an estimate of the trend effect?

Weekly:

$$\hat{m}_t = \frac{X_{t-3} + X_{t-2} + X_{t-1} + X_t + X_{t+1} + X_{t+2} + X_{t+3}}{7}$$

**Centered moving average**

For time series with a period of 12 (i.e. monthly data), the centered moving avg at time $t$ is given by

may ts in june

$$\hat{m}_t = \frac{\frac{1}{2}x_{t-6} + x_{t-5} + \cdots + x_{t-1} + x_t + x_{t+1} + \cdots + x_{t+5} + \frac{1}{2}x_{t+6}}{12}$$

where $t = 7, \ldots, n-6$

How can we obtain an estimate of the seasonal effect at each time $t$? How can we obtain an estimate of the overall seasonal effect associated with each month?

**Seasonal effects**

For an additive time series with a monthly frequency, the seasonal effect at time $t$ is estimated by

$$\hat{s}_t = x_t - \hat{m}_t$$

We can obtain a single estimate of the monthly effect by averaging the effect of each month.

$$\bar{s}_t = \frac{\sum \hat{s}_t}{T-1}$$

where $T$ denotes the number of years. Often times, the estimated seasonal effect is **centered** after calculation - more on this on Wednesday. If a time series is multiplicative, the seasonal effect is instead estimated by

$$\hat{s}_t = \frac{x_t}{\hat{m}_t}$$

How can we obtain an estimate of $z_t$?

**Residual error series**

The residual error series, also called noise, is the raw time series adjusted for the trend and seasonal effects. On average, this series should have a mean of 0. For an additive decomposition model, the residual error series is

$$\hat{z}_t = x_t - \hat{m}_t - \bar{s}_t$$

For a multiplicative decomposition model, the residual error series is

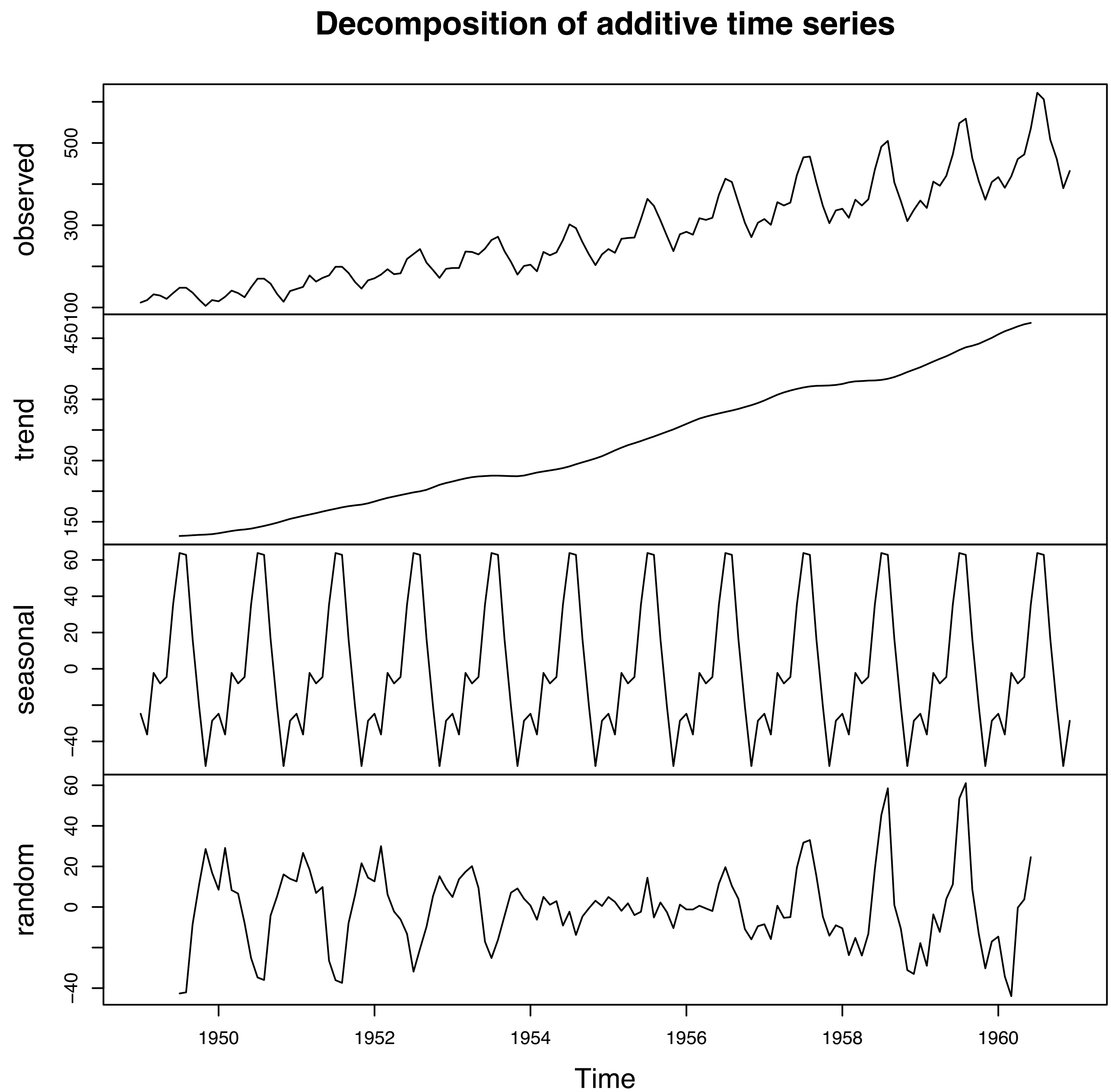
$$\hat{z}_t = \frac{x_t}{\hat{m}_t \cdot \bar{s}_t}$$

## Decomposition in R

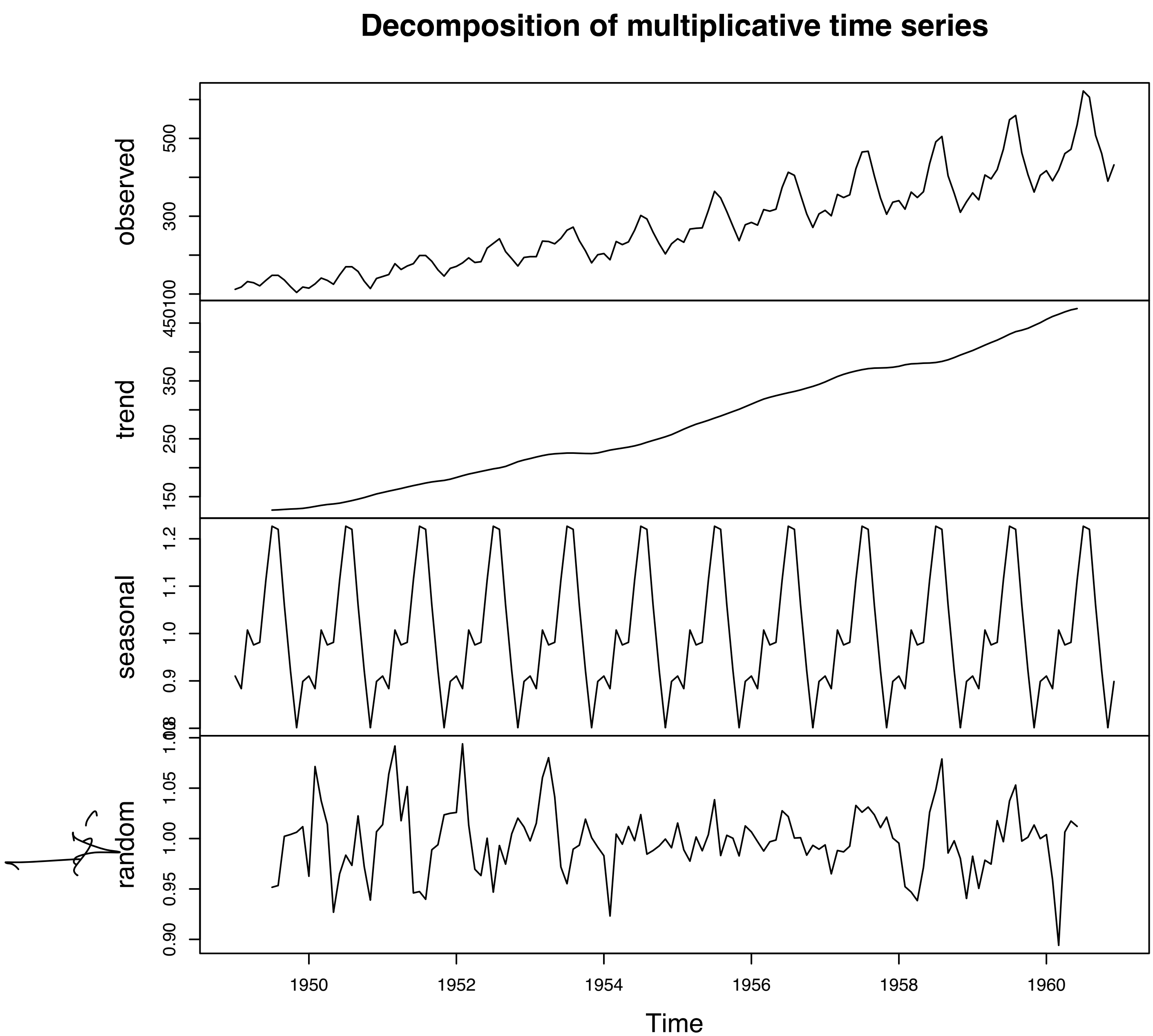
**Note**

The `decompose` function may be used in R to obtain a decomposition of a time series object.

```
plot(decompose(ap))
```



```
plot(decompose(ap, type = "multiplicative"))
```



```
plot(decompose(log(ap)))
```