

ETF3231/5231: Business forecasting

Week 2: Time series graphics

<https://bf.numbat.space/>



Outline

- 1 Time series in R
- 2 Time plots
- 3 Time series patterns
- 4 Seasonal and seasonal subseries plots
- 5 Lag plots and autocorrelation
- 6 White noise

Outline

- 1 Time series in R
- 2 Time plots
- 3 Time series patterns
- 4 Seasonal and seasonal subseries plots
- 5 Lag plots and autocorrelation
- 6 White noise

Time series in R

Included in week 1:

- `tsibble` objects
- The `tsibble` index

Outline

- 1 Time series in R
- 2 Time plots
- 3 Time series patterns
- 4 Seasonal and seasonal subseries plots
- 5 Lag plots and autocorrelation
- 6 White noise

- First in any modelling/forecasting task should be to plot your data.
- Plots allow us to identify:
 - ▶ Patterns;
 - ▶ Unusual observations;
 - ▶ Changes over time;
 - ▶ Relationships between variables.

- First in any modelling/forecasting task should be to plot your data.
- Plots allow us to identify:
 - ▶ Patterns;
 - ▶ Unusual observations;
 - ▶ Changes over time;
 - ▶ Relationships between variables.

Patterns:

- trend
- seasonal
- cycles

Outline

- 1 Time series in R
- 2 Time plots
- 3 Time series patterns**
- 4 Seasonal and seasonal subseries plots
- 5 Lag plots and autocorrelation
- 6 White noise

Time series patterns

Trend pattern exists when there is a **long-term** increase or decrease in the data.

Seasonal pattern exists when a series is influenced by **seasonal factors** (e.g., the quarter of the year, the month, or day of the week).

Cyclic pattern exists when data exhibit rises and falls that are **not of fixed period** (duration usually of at least 2 years).

Seasonal or cyclic?

Differences between seasonal and cyclic patterns:

- seasonal pattern **constant length**; cyclic pattern **variable length**
- **average length** of cycle longer than length of seasonal pattern
- **magnitude** of cycle more variable than magnitude of seasonal pattern

Seasonal or cyclic?

Differences between seasonal and cyclic patterns:

- seasonal pattern **constant length**; cyclic pattern **variable length**
- **average length** of cycle longer than length of seasonal pattern
- **magnitude** of cycle more variable than magnitude of seasonal pattern

The timing of peaks and troughs is predictable with seasonal data, but unpredictable in the long term with cyclic data.

Outline

- 1 Time series in R
- 2 Time plots
- 3 Time series patterns
- 4 Seasonal and seasonal subseries plots**
- 5 Lag plots and autocorrelation
- 6 White noise

Seasonal plots

- Data plotted against the individual "seasons" in which the data were observed. (In this case a "season" is a month.)
- Something like a time plot except that the data from each season are overlapped.
- Enables the underlying seasonal pattern to be seen more clearly, and also allows any substantial departures from the seasonal pattern to be easily identified.
- In R: `gg_season()`

Seasonal subseries plots

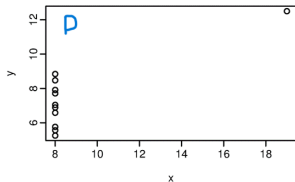
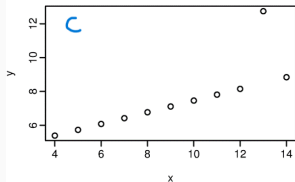
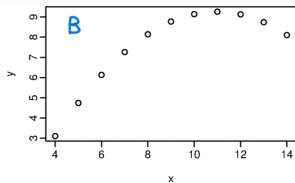
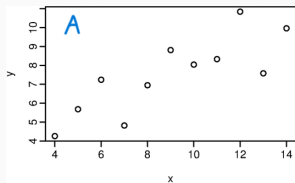
- Data for each season collected together in time plot as separate time series.
- Enables the underlying seasonal pattern to be seen clearly, and changes in seasonality over time to be visualized.
- In R: `gg_subseries()`

Outline

- 1 Time series in R
- 2 Time plots
- 3 Time series patterns
- 4 Seasonal and seasonal subseries plots
- 5 Lag plots and autocorrelation
- 6 White noise

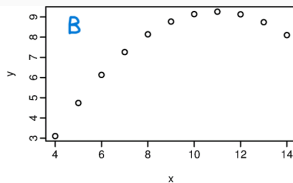
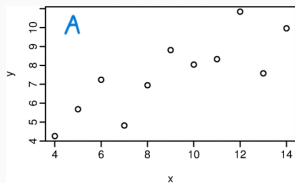
Correlation coefficient

■ Which one has the highest correlation?

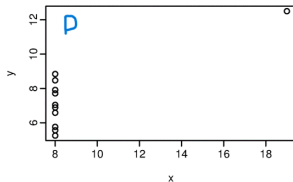
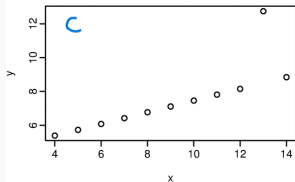


Correlation coefficient

■ Which one has the highest correlation?

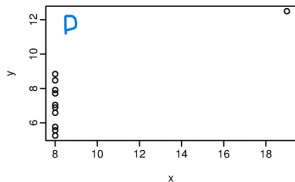
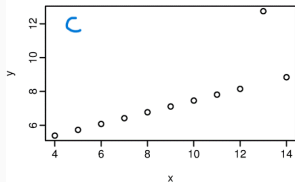
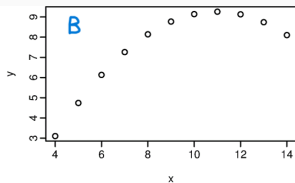
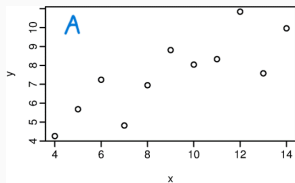


<https://PollEv.com/georgeathana023>



Correlation coefficient

■ Which one has the highest correlation?



<https://PollEv.com/georgeathana023>



All these have $r = 0.82$. Hence importance of plots.

Autocorrelation

Autocovariance (c_k) and autocorrelation (r_k): measure linear relationship between lagged values of a time series y .

Autocorrelation

Autocovariance (c_k) and autocorrelation (r_k): measure linear relationship between lagged values of a time series y .

We measure the relationship between:

- y_t and y_{t-1}
- y_t and y_{t-2}
- y_t and y_{t-3}
- ...
- y_t and y_{t-k}
- etc.

Trend and seasonality in ACF plots

- When data have a **trend**, the autocorrelations for small lags tend to be large and positive.
- When data are **seasonal**, the autocorrelations will be larger at the seasonal lags (i.e., at multiples of the seasonal frequency)
- When data are **trended and seasonal**, you see a combination of these effects.

Outline

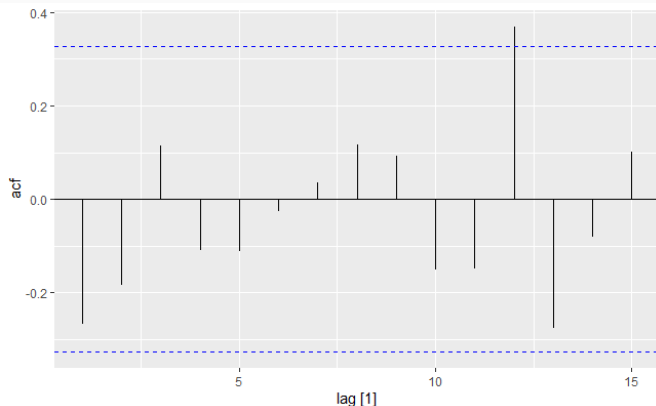
- 1 Time series in R
- 2 Time plots
- 3 Time series patterns
- 4 Seasonal and seasonal subseries plots
- 5 Lag plots and autocorrelation
- 6 White noise

Example: White noise autocorrelation

Example:

$T = 36$ and so critical values at $\pm 1.96/\sqrt{36} = \pm 0.327$.

All autocorrelations lie within these limits, confirming that the data are white noise. (More precisely, the data cannot be distinguished from white noise.)



Note: 5% chance to be outside the critical values (Type I error). You want to see spikes a long way out or many of them. Don't get too excited for 1 just outside especially at longer lags.

Your turn

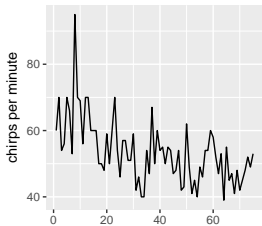
- 1 We have introduced various functions for time series graphics include `autoplot()`, `gg_season()`, `gg_subseries()`, `gg_lag()` and ACF. Use these functions to explore the quarterly tourism data for the Snowy Mountains.

```
snowy <- tourism |> filter(Region == "Snowy Mountains")
```

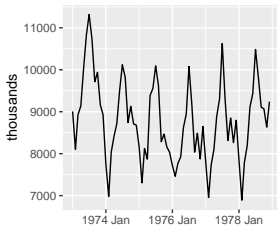
What do you learn?

Which is which?

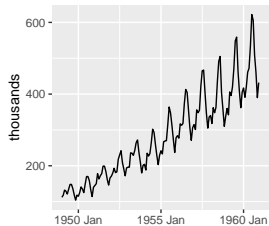
1. Daily temperature of cow



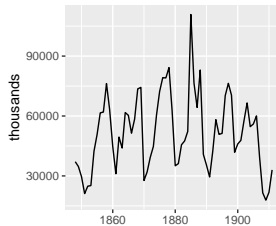
2. Monthly accidental deaths



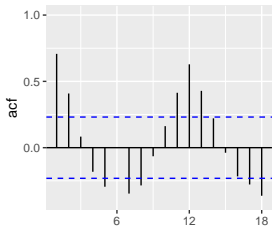
3. Monthly air passengers



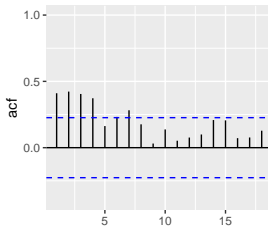
4. Annual mink trappings



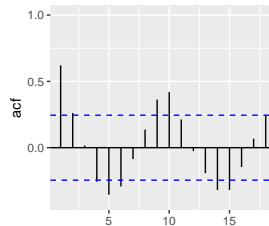
A



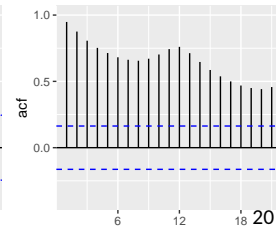
B



C

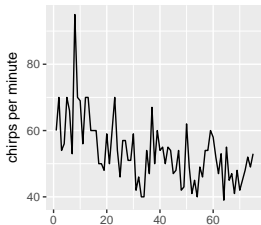


D

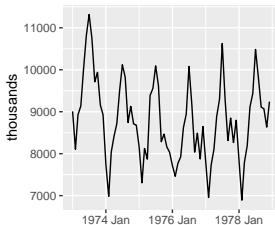


Which is which?

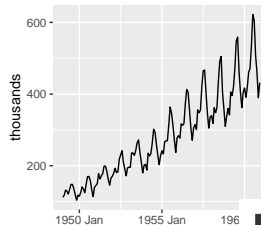
1. Daily temperature of cow



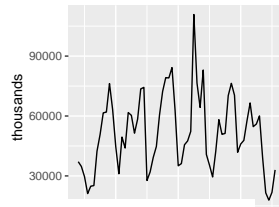
2. Monthly accidental deaths



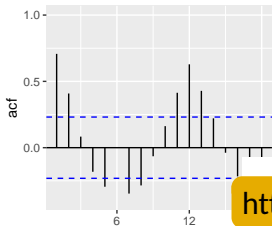
3. Monthly air passengers



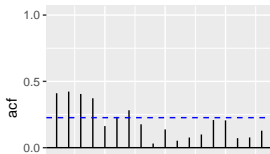
4. Annual mink trappings



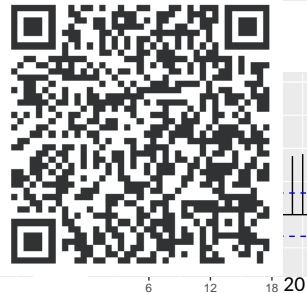
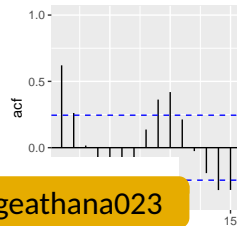
A



B



C



<https://PollEv.com/georgeathana023>

Your turn

- 3 You can compute the daily changes in the Google stock price in 2018 using

```
dgoog <- gafa_stock |>
  filter(Symbol == "GOOG", year(Date) >= 2018) |>
  mutate(trading_day = row_number()) |>
  update_tsibble(index=trading_day, regular=TRUE) |>
  mutate(diff = difference(Close))
```

Does `diff` look like white noise?