

ETC3550/ETC5550 Applied forecasting

Week 1: Introduction to forecasting and R

af.numbat.space



Contact details

Chief Examiner: Professor Rob Hyndman

✉ rob.hyndman@monash.edu

🏠 robjhyndman.com

Tutors

- **Mitchell O'Hara-Wild**
- Elena Sanina
- Zhixiang (Elvis) Yang
- Jarryd Chapman
- Xiefei (Sapphire) Li
- Xiaoqian Wang

Brief bio

- Professor of Statistics, Monash University
- Co-author of most popular forecasting textbook in the world
- Developer of most popular forecasting software in the world

How my forecasting methodology is used:

- Pharmaceutical Benefits Scheme
- Electricity demand
- Australian tourism demand
- Ageing population
- COVID-19 cases
- TAC large claims

Unit objectives

- 1 To obtain an understanding of common statistical methods used in business and economic forecasting.
- 2 To develop the computer skills required to forecast business and economic time series data;
- 3 To gain insights into the problems of implementing and operating large scale forecasting systems for use in business.

Unit objectives

- 1 To obtain an understanding of common statistical methods used in business and economic forecasting.
- 2 To develop the computer skills required to forecast business and economic time series data;
- 3 To gain insights into the problems of implementing and operating large scale forecasting systems for use in business.

Teaching and learning approach

- Approximately one hour of online videos each week.
- One 90 minute in-person tutorial each week.
- One 50 minute in-person seminar each Friday.
- One tutorial will be recorded each week and posted online.

Key reference

Hyndman, R. J. & Athanasopoulos, G. (2021) *Forecasting: principles and practice*, 3rd edition

Key reference

Hyndman, R. J. & Athanasopoulos, G. (2021) *Forecasting: principles and practice*, 3rd edition

[OTexts.com/fpp3/](https://otexts.com/fpp3/)

Key reference

Hyndman, R. J. & Athanasopoulos, G. (2021) *Forecasting: principles and practice*, 3rd edition

[OTexts.com/fpp3/](https://otexts.com/fpp3/)

- Free and online
- Data sets in associated R packages
- R code for examples
- Embedded online lectures

Outline

Week	Topic	Chapter
1	Introduction to forecasting and R	1
2	Time series graphics	2
3	Time series decomposition	3
4	The forecaster's toolbox	5
5-6	Exponential smoothing	8
7-9	ARIMA models	9
10	Multiple regression and forecasting	7
11-12	Dynamic regression	10

Assessment

- Four assignments and one larger project: 40%
- Exam (2 hours): 60%.

Assessment

- Four assignments and one larger project: 40%
- Exam (2 hours): 60%.

Task	Due Date	Value
Assignment 1	Fri 8 Mar	2%
Assignment 2	Sun 24 Mar	6%
Assignment 3	Sun 14 Apr	6%
Assignment 4	Sun 5 May	6%
Retail Project	Fri 24 May	20%
Final Exam	Official exam period	60%

Assessment

- Four assignments and one larger project: 40%
- Exam (2 hours): 60%.

Task	Due Date	Value
Assignment 1	Fri 8 Mar	2%
Assignment 2	Sun 24 Mar	6%
Assignment 3	Sun 14 Apr	6%
Assignment 4	Sun 5 May	6%
Retail Project	Fri 24 May	20%
Final Exam	Official exam period	60%

- Need at least 45% for exam, and 50% for total.
- **ETC5550 students:** Extra exam question.

af.numbat.space

- Includes all course materials
- Links for assignment submissions
- Link to discussion forum.

Please don't send emails. Use the forum.

International Institute of Forecasters



- The IIF provides a prize to the top student in this subject each year.
- US\$100 plus one year membership.



Available for download from CRAN:

<https://cran.r-project.org>



Available for download from RStudio:

<https://www.rstudio.com/products/rstudio/download/>

Main packages



Main packages

```
# Install required packages (do once)
install.packages(c("tidyverse", "fpp3", "GGally"), dependencies = TRUE)
```

Main packages

```
# Install required packages (do once)
install.packages(c("tidyverse", "fpp3", "GGally), dependencies = TRUE)
```

```
# At the start of each session
library(fpp3)
```

Exercises Week 1

- Make sure you are familiar with R, RStudio and the tidyverse packages.
- Do first five chapters of `learnr.numbat.space`.
- Assignment 1

Assignment 1: forecast the following series

- 1 Google closing stock price on 20 March 2024
- 2 Maximum temperature at Melbourne airport on 10 April 2024
- 3 The difference in points (Collingwood minus Essendon) scored in the AFL match between Collingwood and Essendon for the Anzac Day clash. 25 April 2024
- 4 The seasonally adjusted estimate of total employment for April 2024. ABS CAT 6202, to be released around mid May 2024
- 5 Google closing stock price on 22 May 2024

Due Friday 8 March

For each of these, give a point forecast and an 80% prediction interval.

Assignment 1: forecast the following series

- 1 Google closing stock price on 20 March 2024
- 2 Maximum temperature at Melbourne airport on 10 April 2024
- 3 The difference in points (Collingwood minus Essendon) scored in the AFL match between Collingwood and Essendon for the Anzac Day clash. 25 April 2024
- 4 The seasonally adjusted estimate of total employment for April 2024. ABS CAT 6202, to be released around mid May 2024
- 5 Google closing stock price on 22 May 2024

Due Friday 8 March

For each of these, give a point forecast and an 80% prediction interval.

Prize: \$50 cash prize

Assignment 1: scoring

Y = actual, F = point forecast, $[L, U]$ = prediction interval

Point forecasts:

$$\text{Absolute Error} = |Y - F|$$

- Rank results for all students in class
- Add ranks across all five items

Prediction intervals:

$$\text{Interval Score} = (U - L) + 10(L - Y)_+ + 10(Y - U)_+$$

- $u_+ = \max(u, 0)$
- Rank results for all students
- Add ranks across all five items

tsibble objects

```
global_economy
```

```
# A tsibble: 15,150 x 6 [1Y]
```

```
# Key:      Country [263]
```

	Year	Country	GDP	Imports	Exports	Population
	<dbl>	<fct>	<dbl>	<dbl>	<dbl>	<dbl>
1	1960	Afghanistan	537777811.	7.02	4.13	8996351
2	1961	Afghanistan	548888896.	8.10	4.45	9166764
3	1962	Afghanistan	546666678.	9.35	4.88	9345868
4	1963	Afghanistan	751111191.	16.9	9.17	9533954
5	1964	Afghanistan	800000044.	18.1	8.89	9731361
6	1965	Afghanistan	1006666638.	21.4	11.3	9938414
7	1966	Afghanistan	1399999967.	18.6	8.57	10152331
8	1967	Afghanistan	1673333418.	14.2	6.77	10372630
9	1968	Afghanistan	1373333367.	15.2	8.90	10604346
10	1969	Afghanistan	1408888922.	15.0	10.1	10854428

```
# i 15,140 more rows
```

tsibble objects

```
global_economy
```

```
# A tsibble: 15,150 x 6 [1Y]
```

```
# Key:           Country [263]
```

	Index	Country	GDP	Imports	Exports	Population
		<fct>	<dbl>	<dbl>	<dbl>	<dbl>
1	1960	Afghanistan	5377777811.	7.02	4.13	8996351
2	1961	Afghanistan	5488888896.	8.10	4.45	9166764
3	1962	Afghanistan	5466666678.	9.35	4.88	9345868
4	1963	Afghanistan	7511111191.	16.9	9.17	9533954
5	1964	Afghanistan	8000000044.	18.1	8.89	9731361
6	1965	Afghanistan	10066666638.	21.4	11.3	9938414
7	1966	Afghanistan	13999999967.	18.6	8.57	10152331
8	1967	Afghanistan	16733333418.	14.2	6.77	10372630
9	1968	Afghanistan	13733333367.	15.2	8.90	10604346
10	1969	Afghanistan	14088888922.	15.0	10.1	10854428

```
# i 15,140 more rows
```


tsibble objects

```
global_economy
```

```
# A tsibble: 15,150 x 6 [1Y]
```

```
# Key:      Country [263]
```

	Index	Key	GDP	Imports	Exports	Population
			<dbl>	<dbl>	<dbl>	<dbl>
1	1960	Afghanistan	537777811.	7.02	4.13	8996351
2	1961	Afghanistan	548888896.	8.10	4.45	9166764
3	1962	Afghanistan	546666678.	9.35	4.88	9345868
4	1963	Afghanistan	751111191.	16.9	9.17	9533954
5	1964	Afghanistan	800000044.	18.1	8.89	9731361
6	1965	Afghanistan	1006666638.	21.4	11.3	9938414
7	1966	Afghanistan	1399999967.	18.6	8.57	10152331
8	1967	Afghanistan	1673333418.	14.2	6.77	10372630
9	1968	Afghanistan	1373333367.	15.2	8.90	10604346
10	1969	Afghanistan	1408888922.	15.0	10.1	10854428

```
# i 15, 140 more rows
```

tsibble objects

```
global_economy
```

```
# A tsibble: 15,150 x 6 [1Y]
```

```
# Key:      Country [263]
```

	Index	Key	Measured variables			
1	1960	Afghanistan	537777811.	7.02	4.13	8996351
2	1961	Afghanistan	548888896.	8.10	4.45	9166764
3	1962	Afghanistan	546666678.	9.35	4.88	9345868
4	1963	Afghanistan	751111191.	16.9	9.17	9533954
5	1964	Afghanistan	800000044.	18.1	8.89	9731361
6	1965	Afghanistan	1006666638.	21.4	11.3	9938414
7	1966	Afghanistan	1399999967.	18.6	8.57	10152331
8	1967	Afghanistan	1673333418.	14.2	6.77	10372630
9	1968	Afghanistan	1373333367.	15.2	8.90	10604346
10	1969	Afghanistan	1408888922.	15.0	10.1	10854428

```
# i 15, 140 more rows
```

tsibble objects

```
tourism
```

```
# A tsibble: 24,320 x 5 [1Q]
# Key:           Region, State, Purpose [304]
  Quarter Region  State Purpose  Trips
  <qtr> <chr>    <chr> <chr>    <dbl>
1 1998 Q1 Adelaide SA      Business 135.
2 1998 Q2 Adelaide SA      Business 110.
3 1998 Q3 Adelaide SA      Business 166.
4 1998 Q4 Adelaide SA      Business 127.
5 1999 Q1 Adelaide SA      Business 137.
6 1999 Q2 Adelaide SA      Business 200.
7 1999 Q3 Adelaide SA      Business 169.
8 1999 Q4 Adelaide SA      Business 134.
9 2000 Q1 Adelaide SA      Business 154.
10 2000 Q2 Adelaide SA      Business 169.
# i 24,310 more rows
```

tsibble objects

```
tourism
```

```
# A tsibble: 24,320 x 5 [1Q]
# Key:           Region, State, Purpose [304]
  Quarter Region  State Purpose  Trips
  <qtr> <chr>    <chr> <chr>    <dbl>
1 1998 Q1 Adelaide SA      Business 135.
2 1998 Q2 Adelaide SA      Business 110.
3 1998 Q3 Adelaide SA      Business 166.
4 1998 Q4 Adelaide SA      Business 127.
5 1999 Q1 Adelaide SA      Business 137.
6 1999 Q2 Adelaide SA      Business 200.
7 1999 Q3 Adelaide SA      Business 169.
8 1999 Q4 Adelaide SA      Business 134.
9 2000 Q1 Adelaide SA      Business 154.
10 2000 Q2 Adelaide SA      Business 169.
# i 24,310 more rows
```

Domestic visitor
nights in
thousands by
state/region and
purpose.

tsibble objects

```
tourism
```

```
# A tsibble: 24,320 x 5 [1Q]
# Key:           Region, State, Purpose [304]
  Index  Year  Quarter Region  State Purpose  Trips
  <dbl> <dbl> <chr>   <chr>  <chr> <chr>   <dbl>
1 1998   Q1 Adelaide SA      Business 135.
2 1998   Q2 Adelaide SA      Business 110.
3 1998   Q3 Adelaide SA      Business 166.
4 1998   Q4 Adelaide SA      Business 127.
5 1999   Q1 Adelaide SA      Business 137.
6 1999   Q2 Adelaide SA      Business 200.
7 1999   Q3 Adelaide SA      Business 169.
8 1999   Q4 Adelaide SA      Business 134.
9 2000   Q1 Adelaide SA      Business 154.
10 2000  Q2 Adelaide SA      Business 169.
# i 24,310 more rows
```

Domestic visitor
nights in
thousands by
state/region and
purpose.

tsibble objects

```
tourism
```

```
# A tsibble: 24,320 x 5 [1Q]  
# Key:           Region, State, Purpose [304]
```

	Year	Quarter	Region	State	Purpose	Trips
1	1998	Q1	Adelaide	SA	Business	135.
2	1998	Q2	Adelaide	SA	Business	110.
3	1998	Q3	Adelaide	SA	Business	166.
4	1998	Q4	Adelaide	SA	Business	127.
5	1999	Q1	Adelaide	SA	Business	137.
6	1999	Q2	Adelaide	SA	Business	200.
7	1999	Q3	Adelaide	SA	Business	169.
8	1999	Q4	Adelaide	SA	Business	134.
9	2000	Q1	Adelaide	SA	Business	154.
10	2000	Q2	Adelaide	SA	Business	169.

```
# i 24 310 more rows
```

Domestic visitor
nights in
thousands by
state/region and
purpose.

tsibble objects

```
tourism
```

```
# A tsibble: 24,320 x 5 [1Q]
```

```
# Key:           Region, State, Purpose [304]
```

Index	Keys	Measure
1	1998 Q1 Adelaide SA Business	135.
2	1998 Q2 Adelaide SA Business	110.
3	1998 Q3 Adelaide SA Business	166.
4	1998 Q4 Adelaide SA Business	127.
5	1999 Q1 Adelaide SA Business	137.
6	1999 Q2 Adelaide SA Business	200.
7	1999 Q3 Adelaide SA Business	169.
8	1999 Q4 Adelaide SA Business	134.
9	2000 Q1 Adelaide SA Business	154.
10	2000 Q2 Adelaide SA Business	169.

Domestic visitor
nights in
thousands by
state/region and
purpose.

```
# i 24 310 more rows
```

tsibble objects

- A `tsibble` allows storage and manipulation of multiple time series in R.
- It contains:
 - ▶ An index: time information about the observation
 - ▶ Measured variable(s): numbers of interest
 - ▶ Key variable(s): optional unique identifiers for each series
- It works with tidyverse functions.

The `tsibble` index

Time index variables can be created with these functions:

Frequency	Function
Annual	<code>start:end</code>
Quarterly	<code>yearquarter()</code>
Monthly	<code>yearmonth()</code>
Weekly	<code>yearweek()</code>
Daily	<code>as_date()</code> , <code>ymd()</code>
Sub-daily	<code>as_datetime()</code>

Your turn

- 1 Download `tourism.xlsx` from <http://robjhyndman.com/data/tourism.xlsx>, and read it into R using `read_excel()` from the `readxl` package.
- 2 Create a `tsibble` which is identical to the `tourism` `tsibble` from the `tsibble` package.
- 3 Find what combination of `Region` and `Purpose` had the maximum number of overnight trips on average.
- 4 Create a new `tsibble` which combines the `Purposes` and `Regions`, and just has total trips by `State`.