

# *Thống kê và vẽ đồ thị trong R*



---

## *Lời mở đầu*

---

### Tác giả

Duc Nguyen | [tuhocr.com](https://tuhocr.com)

Nội dung cuốn sách này điếm qua hầu hết các chủ đề thống kê và vẽ đồ thị thường gặp, bao gồm các trích dẫn đến tài liệu toàn văn để thuận tiện cho người đọc dễ tra cứu.

Cách tiếp cận đi từ làm rõ định nghĩa, thuật ngữ, kể đến là công thức, thuật toán, bài tập ví dụ và lời giải, sau cùng là tình huống cụ thể.

### Trích dẫn

Duc Nguyen (2025). "Thống kê và vẽ đồ thị trong R". TUHOCR. <https://thongkevavedothi.com>

```
@Book{Nguyen2025,  
  author    = {Duc Nguyen},  
  publisher = {TUHOCR},  
  title     = {Thống kê và vẽ đồ thị trong {R}},  
  year      = {2025},  
  url       = {https://thongkevavedothi.com},  
}
```



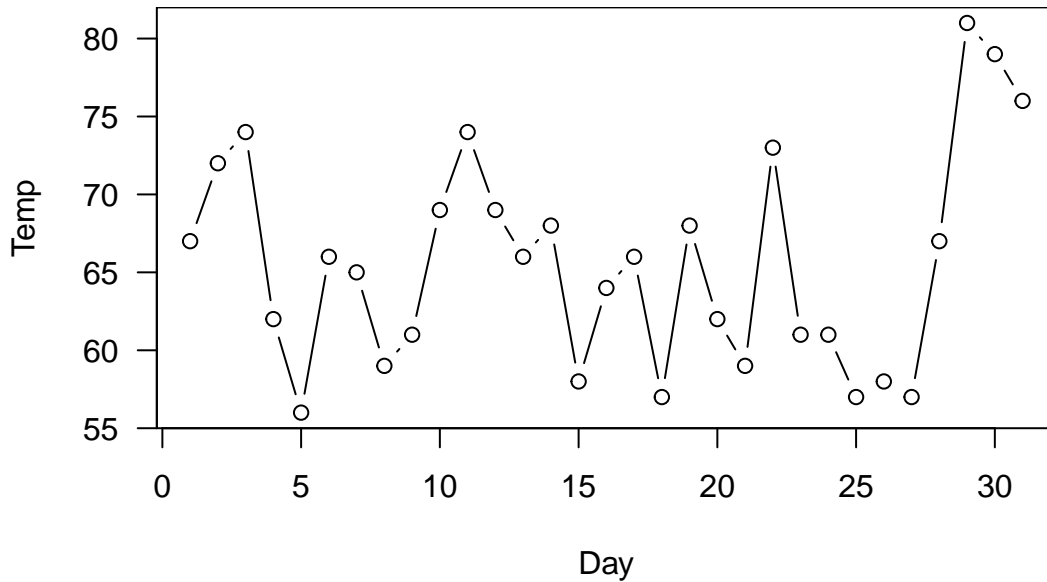
# 1

## *Phân tích chuỗi thời gian*

Time series là một chủ đề khá phức tạp vì liên quan đến nhiều khái niệm tương đối mới so với linear regression. Ta sẽ cần tiếp cận từ những khái niệm căn bản nhất về định nghĩa thuật ngữ (Shumway 2019).

Giả sử chúng ta lấy dữ liệu nhiệt độ theo ngày thì trục hoành sẽ là thời gian (ngày), trục tung sẽ là nhiệt độ. Về mặt ký hiệu ta sẽ xem biến nhiệt độ là biến ngẫu nhiên với các giá trị  $x_1, x_2, x_3, \dots, x_n$  tương ứng ở các mốc thời gian  $x_{t_1}, x_{t_2}, x_{t_3}, \dots, x_{t_n}$

Như vậy, ở một thời điểm  $x_{t=1}$  ta sẽ có 1 giá trị  $x_1$  tương ứng. Ví dụ ở thời điểm ngày 5 ( $x_{t=5}$ ) thì giá trị nhiệt độ là  $x_5 = 56$



Day	Temp
1	67
2	72
3	74
4	62
5	56
6	66
7	65
8	59
9	61
10	69

11	74
12	69
13	66
14	68
15	58
16	64
17	66
18	57
19	68
20	62
21	59
22	73
23	61
24	61
25	57
26	58
27	57
28	67
29	81
30	79
31	76

## 1.1 Chuyển đổi qua đối tượng `ts`

Trước khi thao tác trên dữ liệu time series, ta cần làm quen với object `ts` trong R.

### 1.1.1 Tạo đối tượng `ts`

```
# Bước nhảy 1 tháng
y <- ts(1:36,
        frequency = 12,
        start = c(2025, 1))

y
```

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2025	1	2	3	4	5	6	7	8	9	10	11	12
2026	13	14	15	16	17	18	19	20	21	22	23	24
2027	25	26	27	28	29	30	31	32	33	34	35	36

```
# Bước nhảy 1 quý
y <- ts(1:36,
        frequency = 4,
        start = c(2025, 1))

y
```

	Qtr1	Qtr2	Qtr3	Qtr4
2025	1	2	3	4

2026	5	6	7	8
2027	9	10	11	12
2028	13	14	15	16
2029	17	18	19	20
2030	21	22	23	24
2031	25	26	27	28
2032	29	30	31	32
2033	33	34	35	36

```
# Bước nhảy 1 năm
y <- ts(1:36,
        frequency = 1,
        start = c(2025, 1))

y
```

Time Series:

Start = 2025

End = 2060

Frequency = 1

```
[1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
[26] 26 27 28 29 30 31 32 33 34 35 36
```

### 1.1.2 Chuyển đổi `data.frame` về `ts`

```
month_5 <- airquality[airquality$Month == 5, c("Day", "Temp")]

month_5 <- month_5[, "Temp", drop = FALSE]

# Tạo bước nhảy frequency theo row.names
month_5_ts <- as.ts(month_5)

month_5_ts
```

Time Series:

Start = 1

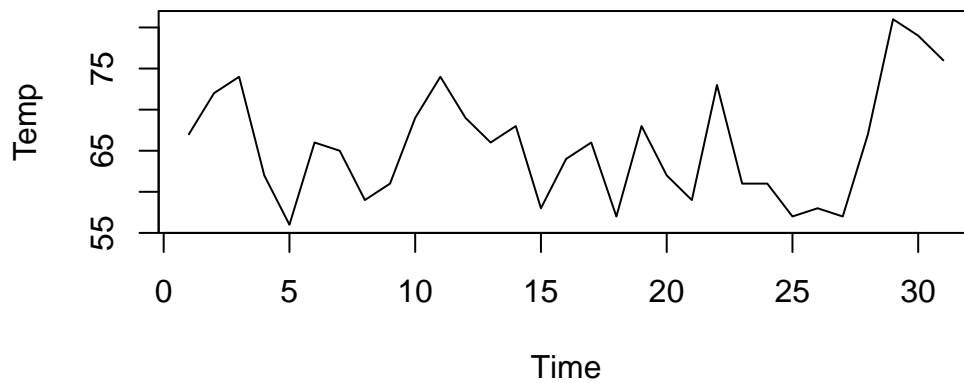
End = 31

Frequency = 1

```
Temp
[1,] 67
[2,] 72
[3,] 74
[4,] 62
[5,] 56
[6,] 66
[7,] 65
[8,] 59
[9,] 61
[10,] 69
[11,] 74
[12,] 69
```

```
[13,] 66
[14,] 68
[15,] 58
[16,] 64
[17,] 66
[18,] 57
[19,] 68
[20,] 62
[21,] 59
[22,] 73
[23,] 61
[24,] 61
[25,] 57
[26,] 58
[27,] 57
[28,] 67
[29,] 81
[30,] 79
[31,] 76
```

```
plot(month_5_ts)
```



Tạo ts theo cột Date trong data.frame

```
temp_ok <- airquality[ , c("Day", "Month", "Temp")]
temp_ok$Year <- 2025
temp_ok <- temp_ok[ , c(1,2,4,3)]
temp_ok$Date_ok <- paste0(temp_ok$Year,"-",temp_ok$Month,"-",temp_ok$Day)
temp_ok$Date_ok <- as.Date(temp_ok$Date_ok, format = "%Y-%m-%d")
temp_ok <- temp_ok[ , c(5, 4)]
library(dplyr)
```



```
temp_ok |> dplyr::arrange(Date_ok) -> temp_ok
temp_ok
```

	Date_ok	Temp
1	2025-05-01	67
2	2025-05-02	72
3	2025-05-03	74
4	2025-05-04	62
5	2025-05-05	56
6	2025-05-06	66
7	2025-05-07	65
8	2025-05-08	59
9	2025-05-09	61
10	2025-05-10	69
11	2025-05-11	74
12	2025-05-12	69
13	2025-05-13	66
14	2025-05-14	68
15	2025-05-15	58
16	2025-05-16	64
17	2025-05-17	66
18	2025-05-18	57
19	2025-05-19	68
20	2025-05-20	62
21	2025-05-21	59
22	2025-05-22	73
23	2025-05-23	61
24	2025-05-24	61
25	2025-05-25	57
26	2025-05-26	58
27	2025-05-27	57
28	2025-05-28	67
29	2025-05-29	81
30	2025-05-30	79
31	2025-05-31	76
32	2025-06-01	78
33	2025-06-02	74
34	2025-06-03	67
35	2025-06-04	84
36	2025-06-05	85
37	2025-06-06	79
38	2025-06-07	82
39	2025-06-08	87
40	2025-06-09	90
41	2025-06-10	87
42	2025-06-11	93
43	2025-06-12	92
44	2025-06-13	82
45	2025-06-14	80
46	2025-06-15	79

47	2025-06-16	77
48	2025-06-17	72
49	2025-06-18	65
50	2025-06-19	73
51	2025-06-20	76
52	2025-06-21	77
53	2025-06-22	76
54	2025-06-23	76
55	2025-06-24	76
56	2025-06-25	75
57	2025-06-26	78
58	2025-06-27	73
59	2025-06-28	80
60	2025-06-29	77
61	2025-06-30	83
62	2025-07-01	84
63	2025-07-02	85
64	2025-07-03	81
65	2025-07-04	84
66	2025-07-05	83
67	2025-07-06	83
68	2025-07-07	88
69	2025-07-08	92
70	2025-07-09	92
71	2025-07-10	89
72	2025-07-11	82
73	2025-07-12	73
74	2025-07-13	81
75	2025-07-14	91
76	2025-07-15	80
77	2025-07-16	81
78	2025-07-17	82
79	2025-07-18	84
80	2025-07-19	87
81	2025-07-20	85
82	2025-07-21	74
83	2025-07-22	81
84	2025-07-23	82
85	2025-07-24	86
86	2025-07-25	85
87	2025-07-26	82
88	2025-07-27	86
89	2025-07-28	88
90	2025-07-29	86
91	2025-07-30	83
92	2025-07-31	81
93	2025-08-01	81
94	2025-08-02	81
95	2025-08-03	82
96	2025-08-04	86
97	2025-08-05	85

98	2025-08-06	87
99	2025-08-07	89
100	2025-08-08	90
101	2025-08-09	90
102	2025-08-10	92
103	2025-08-11	86
104	2025-08-12	86
105	2025-08-13	82
106	2025-08-14	80
107	2025-08-15	79
108	2025-08-16	77
109	2025-08-17	79
110	2025-08-18	76
111	2025-08-19	78
112	2025-08-20	78
113	2025-08-21	77
114	2025-08-22	72
115	2025-08-23	75
116	2025-08-24	79
117	2025-08-25	81
118	2025-08-26	86
119	2025-08-27	88
120	2025-08-28	97
121	2025-08-29	94
122	2025-08-30	96
123	2025-08-31	94
124	2025-09-01	91
125	2025-09-02	92
126	2025-09-03	93
127	2025-09-04	93
128	2025-09-05	87
129	2025-09-06	84
130	2025-09-07	80
131	2025-09-08	78
132	2025-09-09	75
133	2025-09-10	73
134	2025-09-11	81
135	2025-09-12	76
136	2025-09-13	77
137	2025-09-14	71
138	2025-09-15	71
139	2025-09-16	78
140	2025-09-17	67
141	2025-09-18	76
142	2025-09-19	68
143	2025-09-20	82
144	2025-09-21	64
145	2025-09-22	71
146	2025-09-23	81
147	2025-09-24	69
148	2025-09-25	63

```

149 2025-09-26    70
150 2025-09-27    77
151 2025-09-28    75
152 2025-09-29    76
153 2025-09-30    68

```

```

df_ts <- stats::ts(temp_ok[, 2],
                   start = temp_ok[1, 1],
                   end = temp_ok[nrow(temp_ok), 1])

df_ts

```

Time Series:

Start = 20209

End = 20361

Frequency = 1

```

[1] 67 72 74 62 56 66 65 59 61 69 74 69 66 68 58 64 66 57 68 62 59 73 61 61 57
[26] 58 57 67 81 79 76 78 74 67 84 85 79 82 87 90 87 93 92 82 80 79 77 72 65 73
[51] 76 77 76 76 76 75 78 73 80 77 83 84 85 81 84 83 83 88 92 92 89 82 73 81 91
[76] 80 81 82 84 87 85 74 81 82 86 85 82 86 88 86 83 81 81 81 82 86 85 87 89 90
[101] 90 92 86 86 82 80 79 77 79 76 78 78 77 72 75 79 81 86 88 97 94 96 94 91 92
[126] 93 93 87 84 80 78 75 73 81 76 77 71 71 78 67 76 68 82 64 71 81 69 63 70 77
[151] 75 76 68

```

```

# as.numeric(temp_ok[1, 1])

# as.numeric(as.Date("1970-01-01"))

attributes(df_ts)$tsp[1]

```

[1] 20209

```
attributes(df_ts)$tsp[2]
```

[1] 20361

```

date_begin <- as.Date("1970-01-01") + attributes(df_ts)$tsp[1]

date_end <- as.Date("1970-01-01") + attributes(df_ts)$tsp[2]

date_all_day <- seq.Date(from = date_begin,
                        to = date_end,
                        by = "day")

date_all_month <- seq.Date(from = date_begin,
                          to = date_end,
                          by = "month")

date_all_day[!(date_all_day %in% date_all_month)] <- NA

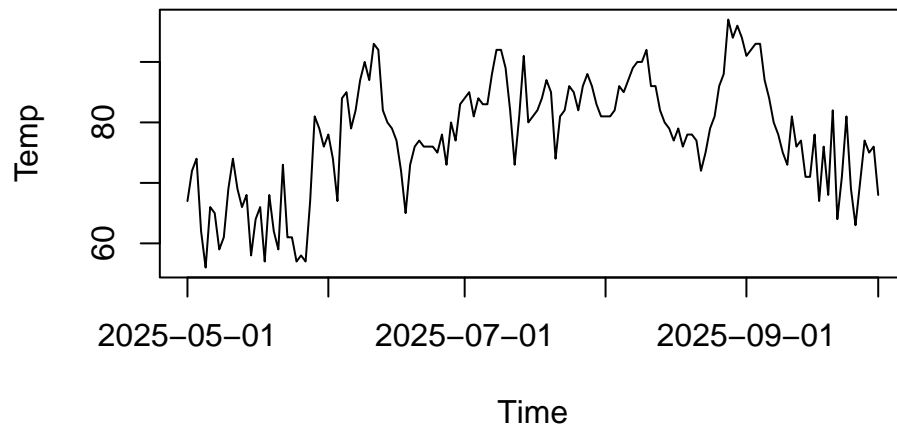
```

```
date_all_day[length(date_all_day)] <- date_end

par(mar = c(4,4,4,4))

plot(df_ts,
      ylab = "Temp",
      xaxt = "n")

axis(side = 1,
      at = date_all_day,
      labels = date_all_day)
```





## 2

---

# *Nhập môn kinh tế lượng*

---

### 2.1 Điểm danh tài liệu quan trọng

Dòng sách “Introductory Econometrics: A Modern Approach” được viết bởi Jeffrey M. Wooldridge (2019). Dataset của tài liệu này tham khảo ở [đây](#). Florian Heiss (2020) có viết lại cuốn này làm rõ hơn cách ráp code R ở [đây](#).





# 3

---

## Hồi quy tuyến tính

---

### 3.1 Điểm danh tài liệu quan trọng

Bài giảng của Prof. Kerby Shedden [[source](#) | [mirror](#)]

---

### 3.2 Model selection

Lecture note của Kerby Shedden [[link](#)]



---

## Tài liệu tham khảo

- Heiss, Florian. 2020. *Using r, Python and Julia for Introductory Econometrics*. <http://book.thuviencanhon.com:8033/results?query=&dir=tuhocr%2FEconometrics%2FFlorian+Heiss&after=&before=&sort=relevancyrating&ascending=0&page=1>.
- Shumway, Robert H. 2019. *Time Series: A Data Analysis Approach Using r*. Edited by David S. Stoffer. Boca Raton: CRC Press, Taylor & Francis Group,. [http://book.thuviencanhon.com:8033/results?query=filename:%22Time+Series\\_\\_+A+Data+Analysis+Approach+Using+R%22&dir=%3Call%3E&after=&before=&sort=relevancyrating&ascending=0&page=1](http://book.thuviencanhon.com:8033/results?query=filename:%22Time+Series__+A+Data+Analysis+Approach+Using+R%22&dir=%3Call%3E&after=&before=&sort=relevancyrating&ascending=0&page=1).
- Wooldridge, Jeffrey M. 2019. *Introductory Econometrics: A Modern Approach*. 7th ed. <http://book.thuviencanhon.com:8033/results?query=&dir=tuhocr/Econometrics/Jeffrey+M.+Wooldridge&after=&before=&sort=relevancyrating&ascending=0&page=1>.



# A

---

## *Quy cách trích dẫn*

---

### A.1 Trích dẫn theo họ tên

<https://www.bibtex.com/f/author-field/>

```
% The King of Pop: Michael Joseph Jackson
author = "Michael Joseph Jackson"
author = "Jackson, Michael Joseph"
author = "Jackson, Michael J"
author = "Jackson, M J"
```

```
% An example with a suffix
author = "Stoner, Jr, Winifred Sackville"
```

```
% An exmaple with a particle
author = "Ludwig van Beethoven"
author = "van Beethoven, Ludwig"
author = "van Beethoven, L"
```

```
% Corporate names or names of consortia
author = "{Barnes and Noble, Inc.}"
author = "{FCC H2020 Project}"
```