

第十一章时间序列分析方法

授课教师：吴翔
wuhsiang@hust.edu.cn

OCT 15 - 18, 2019

- 1 时间序列分析概述 (2 个课时)
- 2 时间序列经典分析方法 (3 个课时)
- 3 时间序列案例分析 (1 个课时)
- 4 时间序列分析实习 (2 个课时)

时间序列分析概述 (2 个课时)

课程存储地址

- 课程存储地址: <https://github.com/wuhsiang/Courses>
- 资源: 课件、案例数据及代码



图 1: 课程存储地址

参考教材

- James D. Hamilton 著. 时间序列分析 (2 册) . 北京: 人民卫生出版社. 2015.
- Jonathan D. Cryer & Kung-Sik Chan 著. 时间序列分析及应用 (R 语言) (原书第 2 版) . 北京: 机械工业出版社. 2011.
- Robert I. Kabacoff 著. R 语言实战 (第二版) . 北京: 人民邮电出版社. 2016.
- David Salsburg 著. 女士品茶: 统计学如何变革了科学和生活. 南昌: 江西人民出版社. 2016.

本节知识点

- 时间序列分析方法起源
- 时间序列基本概念
- 时间序列分析要素
- 时间序列分析建模

11.1 时间序列分析方法起源

时间序列分析的方法，起源于英国统计学家 Ronald Fisher 在英国洛桑试验站 (1919-1933) 的**农作物收成变动研究**。这一研究产生了统计史上的三个重要方法：

- 时间序列分析思想
- 随机对照实验
- 方差分析

我们通过回顾这一研究，以深入理解时间序列分析的思想。

Fisher 与洛桑试验站

英国**洛桑试验站 (Rothamsted Experimental Station)**，现为洛桑研究所 (Rothamsted Research)，在漫长的历史中 (1843-1919) 积累了大量的“实验数据”和其它记录，包括：

- 降水量和温度的每日精确记录
- 施肥量与土壤检测数据的每周记录
- 农作物收成的每年记录
- 人造肥料和不同农作物（小麦、黑麦、大麦、马铃薯等）的组合实验方案

洛桑试验站的农作物



图 2: 洛桑试验站

收成变动研究

Fisher 考虑特定年份 (t) 的特定田地 (i) 上的农作物产量 (Y_{it}), 并试图回答以下问题:

- **降水量**对小麦产量有何影响?
- 不同**肥料**对不同品种的马铃薯产量有何影响?

课堂讨论: 第一个问题 (5min)

如何预测小麦产量?

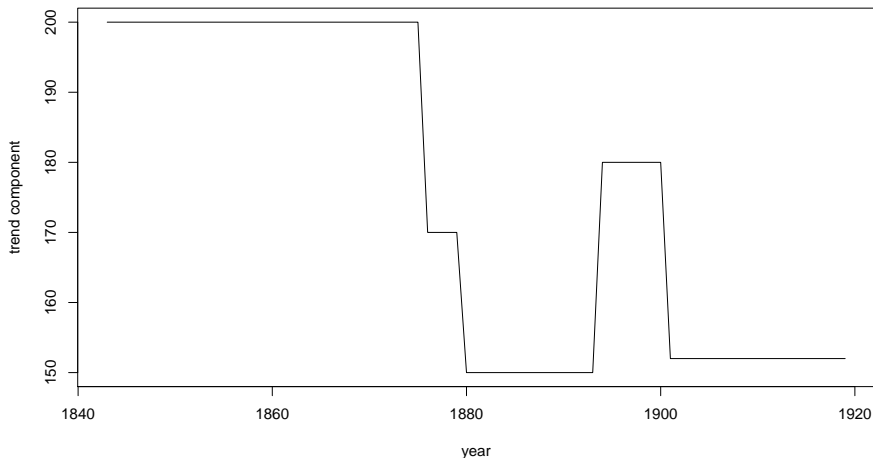
选定一块田地 (名曰: “宽埂”), 这块田地**只使用过**动物粪便作为肥料, 考虑**降水量如何影响小麦产量**。

- 降水量和小麦均是**时变**的, 即考虑不同年份 (t) 的降水量 (X_t) 和小麦产量 (Y_t)。
- 探讨降水量 X_t 与小麦产量 Y_t 的关系, 即属于**时间序列分析**范畴。
- 问题演变为, 如何**预测**小麦产量 Y_t ?

小麦产量的影响因素分解

- 土壤退化导致的产量总体稳步减小
- 长期缓慢变化，每个变化周期为期数年
- 不同年份的气候因素（例如降水量）导致的变化

长期缓慢的变化



理解长期而缓慢的变化

- 难以解释的长期缓慢变化：产量从 1876 年开始剧烈下降，从 1880 年开始下降更加剧烈，至 1894 年产量开始改善，从 1901 年开始又剧烈下降
- 小麦田地中杂草的生长情况与之相反
- 最终解释：1876 年前，人们雇佣小男孩到田里除草 -> 1876 年《教育法》规定适龄儿童必须上学 -> 1880 年法律处罚不让适龄儿童上学的家庭 -> 1894 年洛桑附近的女子寄宿学校校长认为，高强度户外运动有助于儿童健康 -> 1901 年校长去世

肥料使用对马铃薯产量的影响研究

- 洛桑试验站的早期方案：在不同片田地上（或者 Fisher 的田地分块），针对不同马铃薯品种，使用不同化肥，并记录其产量
- 潜在问题
 - **田地**相关的混淆因素，例如土壤、排水方式、营养物质、杂草等
 - **年份**相关的混淆因素，例如气候变化等

Fisher 开创的统计方法

- 随机对照实验
 - 每片田地分块之后，进一步把每块田地分为若干排
 - 采用**随机化**方案，对每块地的每一排实施不同的处理
- 方差分析

分解思想

Fisher 在洛桑试验站研究农作物收成变动时, 采用了**分解 (decomposition)** 的思想:

- 时间序列中的**要素分解**
- 方差分析中的**效应分解**

11.1.2 时间序列基本概念

● 时间序列 (time series)

- 定义：一组在特定时刻的观测值 Y_t ，例如特定田地上的小麦收成
- 领域：广泛存在于宏观经济、金融财务以及医疗领域

● 时间序列分析 (time series analysis)

- 数据：时间序列数据，与横截面数据、面板数据，为三类主要的观测数据类型
- 分析方法：通常基于宏观经济学理论建模，并被视作宏观计量经济学的主要方法

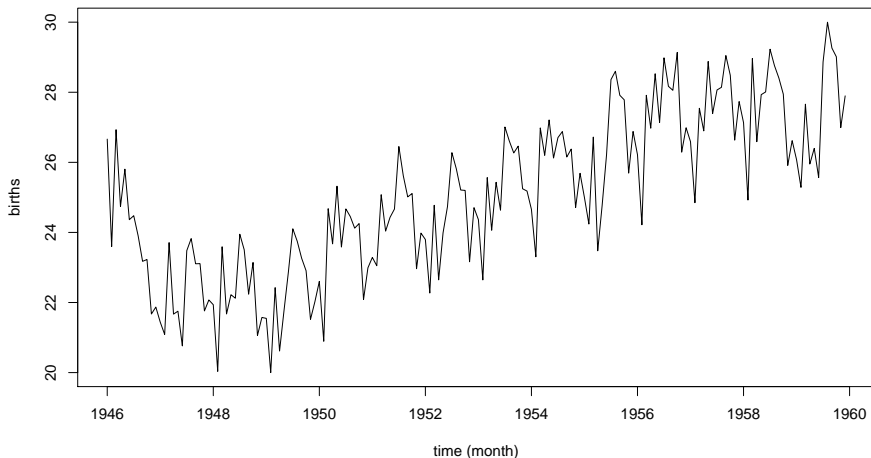
横截面 vs 时间序列数据

- 横截面数据 (cross sectional data)
 - 定义：不同**研究对象**在某一**时点**的变量观测数据
 - 例子：通常的问卷调查
- 时间序列数据 (time series data)
 - 定义：也称为纵剖面 (longitudinal sectional) 数据，是某一研究对象在不同**时点**的变量观测数据
 - 例子：股票价格数据

面板数据

- 定义：不同研究对象在不同时点的变量观测数据
- 特征：具有“横截面”和“时间序列”两个维度
- 例子：中国健康与养老追踪调查 (China Health and Retirement Longitudinal Study, CHARLS)、中国健康与营养调查 (China Health and Nutrition Survey, CHNS) 等

时间序列数据



11.1.3 时间序列分析要素

影响时间序列观测值的因素，可以分为以下几类：

- ① **趋势因素 (trend component)**：观测值的长期的趋势，通常是非线性的
- ② **循环因素 (cyclical component)**：非季节因素引起的波动，通常也被归入趋势因素中
- ③ **季节因素 (seasonal component)**：在一定时期内呈现的规律变化，例如一年内随着自然季节的更替而发生的变化
- ④ **不规则因素 (irregular component)**：随机因素

通常将趋势因素和循环因素合并在一起考虑，成为趋势-循环因素 (trend-cycle)，或简称**趋势因素**。

课堂讨论

讨论以下情境中，趋势因素 (trend component) 和季节因素 (seasonal component) 的**含义**是什么

- 京东 & 淘宝上某一产品的销售量
- 某一医院在过去十年的每日门诊病人数量

11.1.4 时间序列建模

时间序列可以分解为趋势因素、季节因素和随机因素

$$Y_t = f(T_t, S_t, E_t)$$

常用的函数类型 $f(\cdot)$ 有两种：累加、累乘。

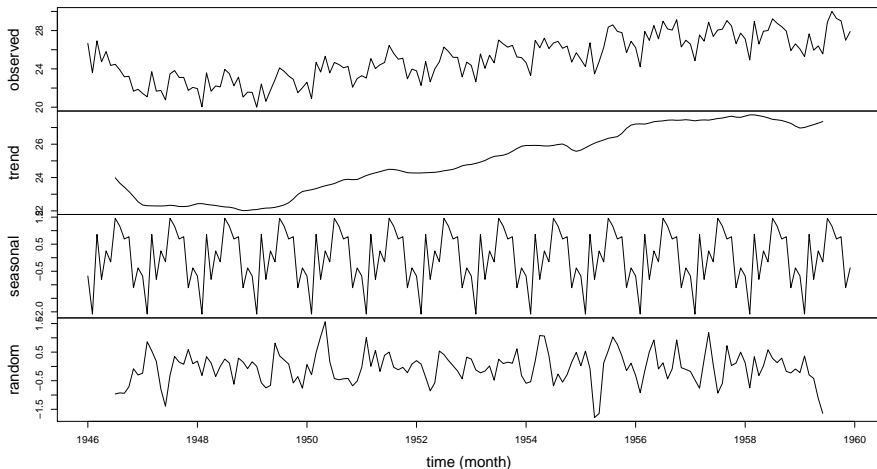
加法模型

假定趋势因素、季节因素和随机因素**相互独立**，则可以用**加法模型 (additive model)** 来分解各个因素

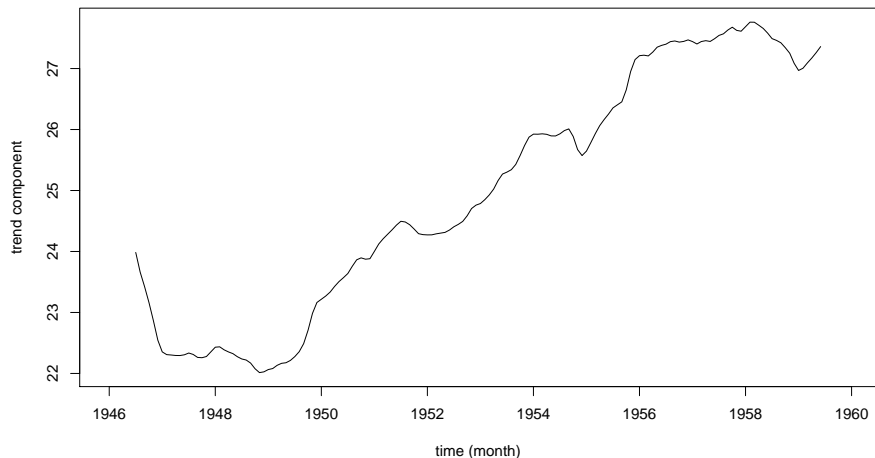
$$Y_t = T_t + S_t + E_t.$$

因素分解：加法模型

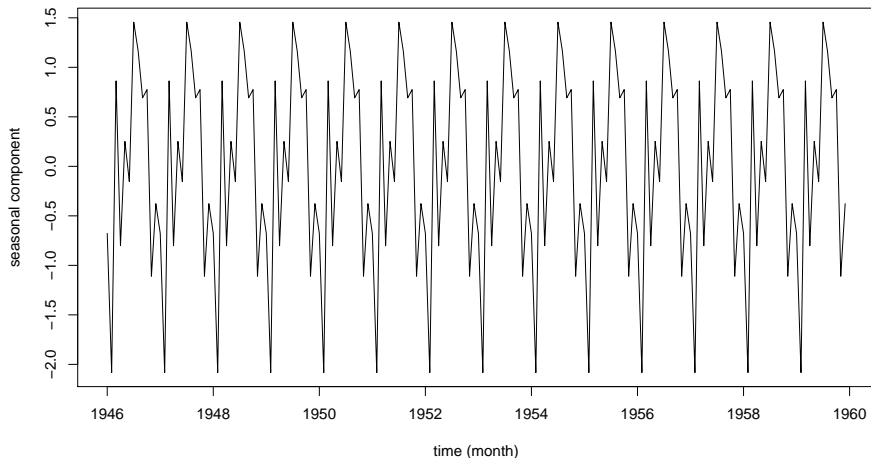
Decomposition of additive time series



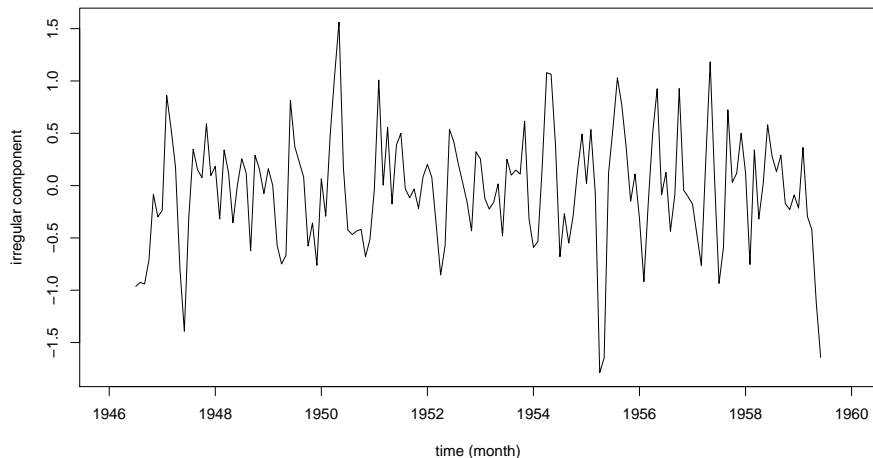
加法模型之趋势因素



加法模型之季节因素



加法模型之随机因素



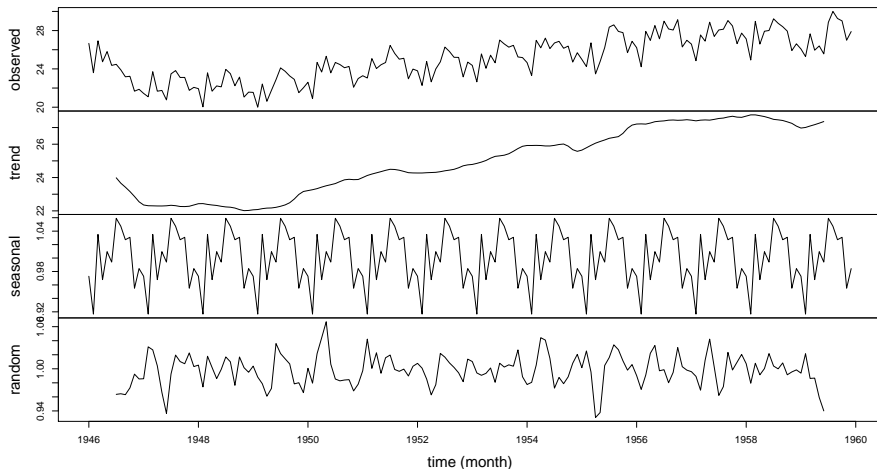
乘法模型

趋势因素、季节因素和随机因素不满足相互独立的条件，则可以用**乘法模型 (multiplicative model)** 来分解各个因素

$$Y_t = T_t \times S_t \times E_t.$$

因素分解：乘法模型

Decomposition of multiplicative time series



时间序列分析步骤

- 1 搜集数据, 绘制时间序列图
- 2 因素分解, 得到趋势因素、季节因素和随机因素
- 3 对特定情境建模, 分别预测趋势因素和季节因素的时间序列值
- 4 获得最终的预测模型

时间序列经典分析方法 (3 个课时)

本节知识点

- 移动平均法
- 指数平滑法
- 生长曲线法
- 灰色系统预测法 (略)

11.2.1 移动平均法

- 简单移动平均法
- 加权移动平均法
- 趋势移动平均法

简单移动平均法

加权移动平均法

趋势移动平均法

11.2.2 指数平滑法

- 一次指数平滑法
- 二次指数平滑法

一次指数平滑法

二次指数平滑法

11.2.3 生长曲线法

- 指数曲线模型
- Logistic 曲线模型
- Compertz 曲线模型

指数曲线模型

Logistic 曲线模型

Compertz 曲线模型

时间序列案例分析 (1 个课时)

本节知识点

- 时间序列分析建模与预测
- <https://github.com/wuhsiang/Courses/blob/master/healthinfo/cases/case-dhaka.Rmd>

时间序列分析实习 (2 个课时)