Time Series Analysis

With application to Statistical Process Control

By Charles Widger, 9/30/2025, Ver 6

Introduction

Time Series in General

- Real number data are collected and ordered by time.
  - It may be from a single source. Examples:
    - Seasonal snowfall at Snoqualmie Pass.
    - Number of errors found by Peer Review for software under development.
      - Applies to requirements, design, source code
      - Also applies to verification requirements, design, test procedures, application, results
  - It may be from multiple sources. Examples:
    - Deaths from turberuclosis by year, for USA as aggregated from each State and Territory.
    - Thickness of a sheet of metal being manufactured and measured by multiple sensors across the sheet.
    - Sunspot observations as recorded at various source(s). Sources may disagree over observations made during the same time period.
- Data analyzed for identifying and dealing with trends and seasonal variation, and forming models for prediction.

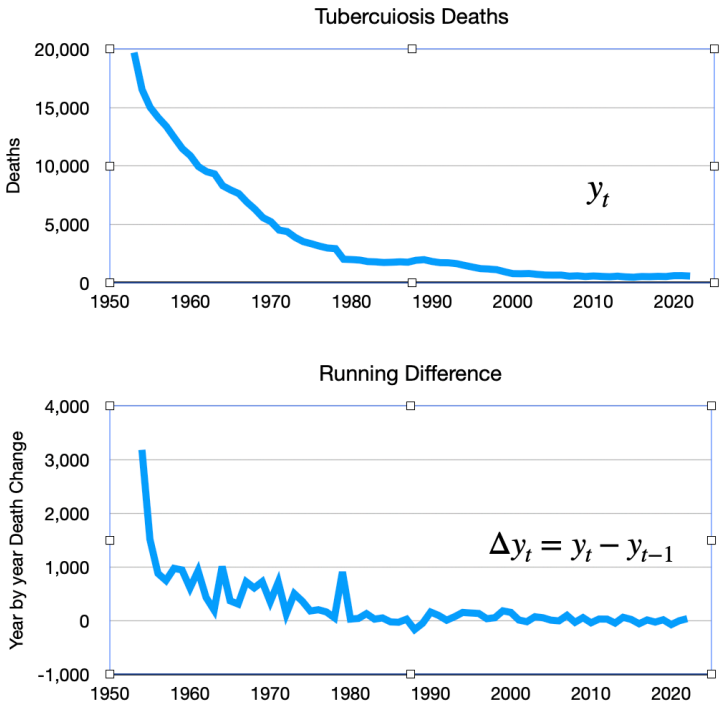
References

- **The Analysis of Time Series: An Introduction**, 6th Edition, Chris Chatfield, Chapman & Hall, CRC, 2004, ISBN 1-584488-317-0.
- **Economic Control of Quality of Manufactured Product**, 50th Anniversary, W. A. Shewhart (1891-1967), Quality Press, 1980, ISBN 0-87389-076-0. Initially published by Van Nostrand, 1931.
- **Understanding Statistical Process Control, 2nd Edition**, Donald Wheeler and David S. Chambers, SPC Press, 1992, ISBN 0-945320-13-2. Currently at the 3rd Edition. I have the 2nd Edition, which I use when making references.

Deaths from TB

[Source Data]

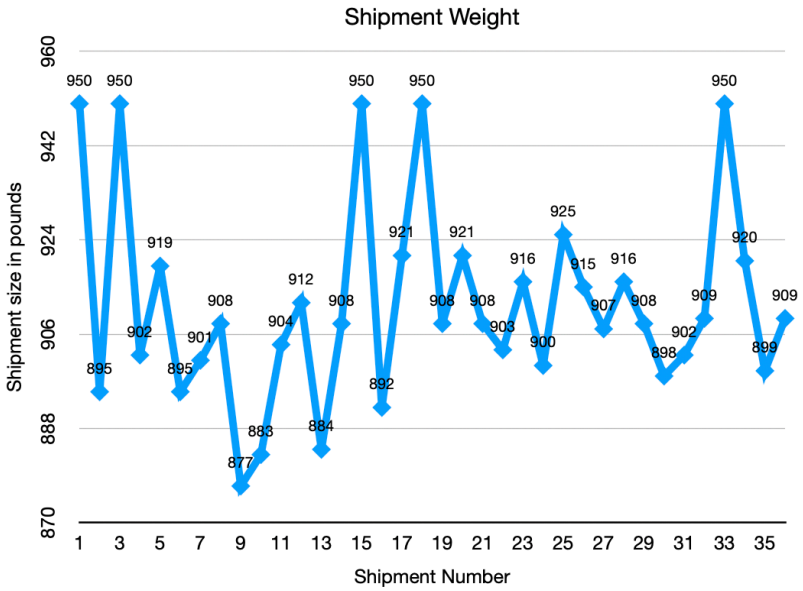
- The CDC produces an annual report with these data. The CDC version of the Running Difference graph is different in that it uses columns instead of a line graph.
- The deaths decreased in 1974 due to a change in data collected. Deaths from some causes, late term consquences of TB, were no longer attributed to TB.



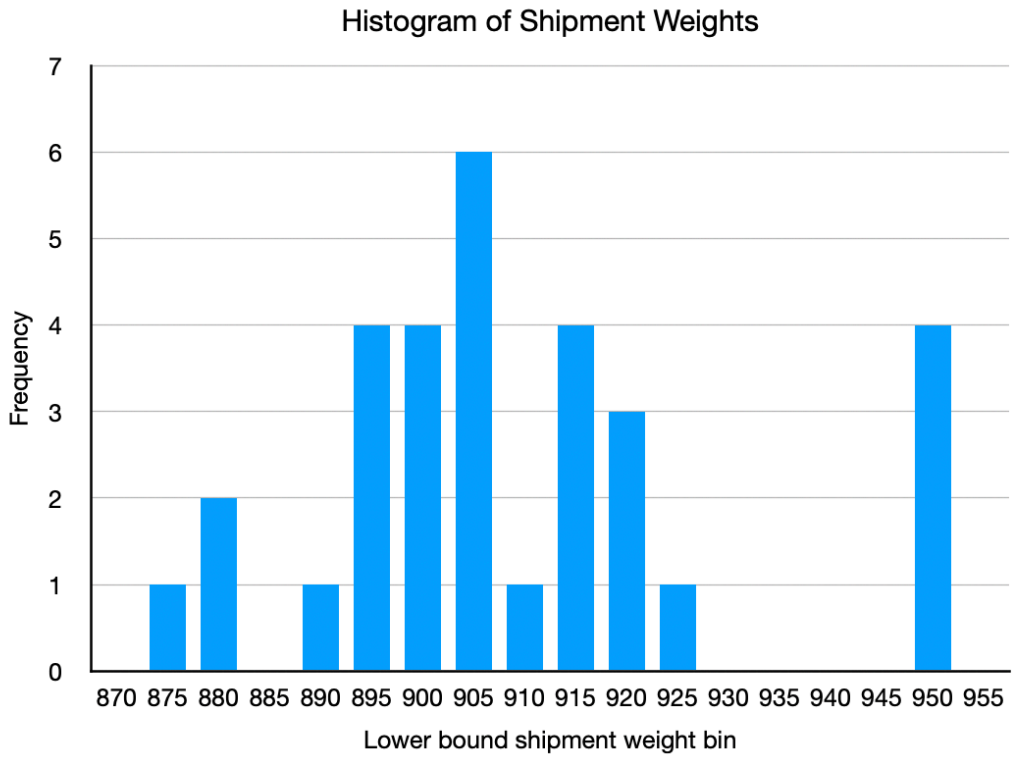
Alternative Data Displays

Useful for analysis of data distribution

- Consider a time series of 36 shipment weights in pounds. The minimum is 877 and the maximum is 950. The graph displays these.
- A *Histogram* of this time series bins these into 16 parts of 5 lbs each. It displays the frequency of weights occuring in each bin. Graph on next slide.
- A *Stem-and-Leaf-Plot*, also known as a *Stemplot*, divides the data into a selected Stem and Leafs. For this example, the stem is the 100s and the 10s digits of the weight, and the leaf is the units digit. Graph on slide after next.



Shipment Weight Histogram



Stem & Leaf Plot for Shipment Weights

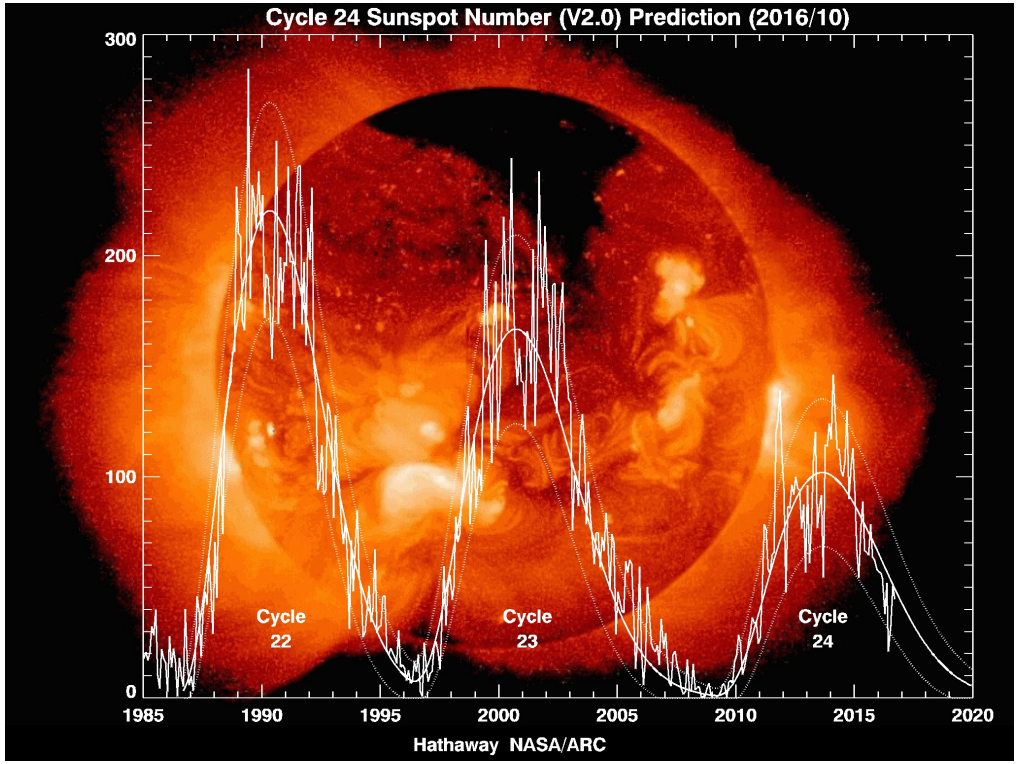
Stem & Leafs for Shipment Weights

Stem	Leaf										
87	7										
88	3	4									
89	5	5	2	8	9						
90	2	1	8	4	8	8	0	7	2	9	9
91	9	2	6	5	6						
92	1	1	5	0							
93											
94											
95	0	0	0	0							
96											
97											

Sunspot Cycle

Cycles last about 11.5 years

- The generally cyclic behavior of sunspots is known as Spörer's Law. Gustav Spörer (1822-1895) refined work done around 1861 by Richard Christopher Carrington (1826-1875). Also credited are the Maunder couple, Edgar Walter Maunder (1851-1928) and his wife Annie Scott Dill neé Russell (1868-1947), whom he married in 1895 as a widower.
- Cycle 1 begins around 1760. The period 1645 - 1715 is known as the Maunder Minimum because so very few sunspots were observed.
- Sunspots have magnetic polarities that change at the top of cycle.
- Sunspots contribute to solar wind. More sunspots result in auroras and can affect electronic equipment.
- The Solar Cycle remains an area of active research.

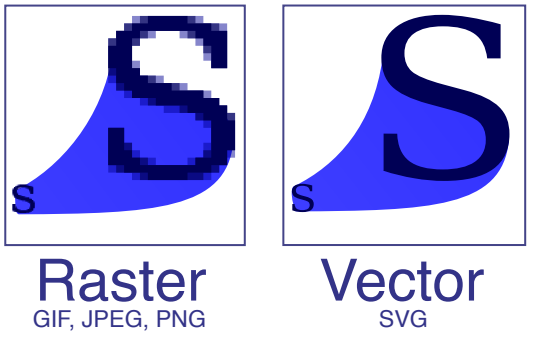


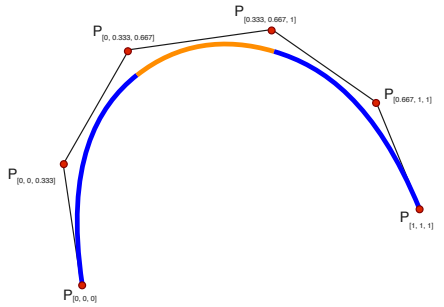
Graph shows time plot with fitted smooth curves
Note the curves are cyclic, but not periodic

Splines

Creating Piecewise C^2 Smooth Curves

- One may be familiar with the term “spline” from computer graphics software. These programs are of two basic types: vector based and raster based. Each type has advantages and disadvantages.
- The term spline is borrowed from ship building by the British in making full-sized drawings of aircraft in World War II. The spline is a wooden drafting tool. These drawings needed large spaces, for which a loft of a building was typically used. “Lofting” is still a term used for fuselage and wing design.
- The curves are polynomial functions, typically of low degree for easy calculations.
- The Karl Weierstraß (1815-1897) Approximation Theorem says for every continuous function $f(x)$ on the closed interval $[a, b]$, there is a set of polynomial functions $P_n(x)$ of degree at most $n = 0,1,2,\ldots$, such that $\lim_{n\to\infty} \left(\sup_{a\le x\le b} |f(x) - P_n(x)| \right) = 0$ with uniform convergence.

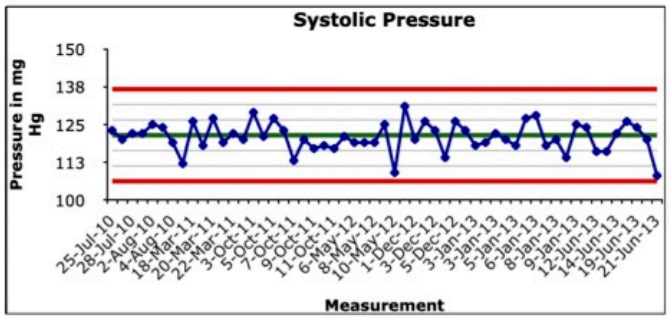


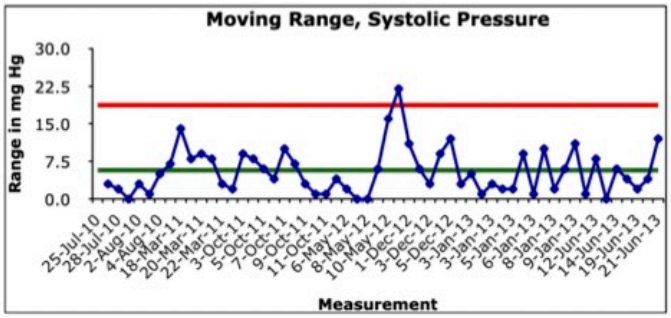


Statistical Process Control

Shewhart Chart

- The Shewhart Chart is a basic tool for Statistical Process Control (SPC). These were invented by Walter A. Shewhart (1891-1967) in the early 1920s.
- The chart consists of two graphs, one a running record of a time series of interest along with the average of these values. Also displayed are upper and lower control limits. The other is a graph of the dispersion this running record, again with its average and a upper control limit.
- The control limits show how the process is expected to behave in future applications.
  - If any of the control limits are exceeded, the process is considered out of control.
  - It may happen that the dispersion is too great for the bounds to be useable. The process needs to be improved until it yields controlled results.
- Shewhart recognized two kinds of sources for producing non-compling product: those with assignable cause, and those that do not, rather they have an unknown chance cause.
  - Sources of assignable causes include such things as changes in input material, machinery out of calibration, skill of workers, working environment, focus of management. Pareto diagrams provide a means to determining these.





Development of SPC

Short History

- Shewhart conceived his method in working with Western Electric, which produced all the telephone equipment for Bell Telephone, then a monopoly for telephone service. The process produced switching systems, these were tested, and if out of compliance, either scrapped or restored to acceptable performance.
- Applying his method during the manufacturing process focussed changes on reducing sources of variation. It did this by providing a means of measuring improvement (or not). This had an economic effect of improving productivity and reducing scrap. Hence the name of his book, **Economic Control of Quality ...** .
- World War II provided opportunities for production improvements. W. Edwards Deming (1900-1993) provided instuction on the use of SPC for factory workers. He was disappointed that upper management rejected interest in it. Management typically thought that meeting specifications was good enough. This gives no basis for seeing whether any changes in process really make any difference.
- Following WW II, use of SPC waned due to high demand for US goods. Deming was sent to Japan to assist with the 1951 census there. Deming eventually met Ichiro Ishikawa, a cofounder of the Union of Japanese Scientists and Engineers (JUSE) (in Japanese: 日本科学技術連盟). Its purpose was to bring together leaders and technical experts from all major Japanese industries to revitalize Japan's economy and eliminate waste by improving quality. Deming made presentations, which were acted upon. As a thank you, the JUSE now awards a Deming Prise.
- Interest in SPC in the USA becan to renew in 1981, when Ford hired Deming as a consultant for quality. Between 1979 and 1982, Ford sales were down, and the company recorded $3 B is losses. Deming claimed 85% of Ford's problems were caused by management actions.
- In the 1990s, Boeing sent upper management to Toyota to learn about its practices regarding quality and its management. Upon return, management began to implement changes.

Computation of Control Limits

Answer in Wikipedia

- The referenced article speaks to "rational subgroups" of size $n, 1 < n \leq 10$. Actually, it gives tables of constants to use for $1 < n \leq 15$. Wheeler & Chambers provide methods to derive them for size $n, 15 < n \leq 100$. See *op. cit.* pp. 393 ff.
  - The Wikipedia article does not explain what a makes a subgroup rational. We will provide guidance from Wheeler & Chambers.
  - It does not speak to individual data, not in subgroups, such as the tuberculosis deaths. For these time series X, the X chart data limits are called "natural process limits." We have: $UNPL_X = \bar{X} + (2.660)\overline{mR}$, $LNPL_X = \bar{X} - (2.660)\overline{mR}$, $UCL_R = (3.268)\overline{mR}$. Note $mR_k = |X_k - X_{k-1}|$, $1 < k \leq n$, is the moving range. Bars above a symbol indicate averages.
  - It further notes that the constants are derived from the case that the data are independent normal distributed. While this is correct, Shewhart noted that these values are very useful even when the data are not so distributed. This matter is discussed at length in Wheeler & Cambers, pp. 65 - 82. I recall reading an earlier version of Shewhart's book in which he attributes an intuitive reason for this based on Pafnuty Chebyshev (1821-1894) inequality that applies to all distributions.

Injection Molding Setup

Wheeler & Cambers pp. 102 - 114

- A company bought a new molding press with updated features and which used different input materials. It decided to do some calibration runs to see whether their processes needed updating.
- The mold produced 4 units at a time by injecting material into each Cavity. The parts had an important globalar bulb. A measuring tool was fabricated to measure the thickness of the bulb by which it exceeded 12 mm.
- Samples were taken 5 times a day at suitable intervals. The collection times were labeled A - E. The collection for each complete day is called a cycle.
- Collection continued over 20 cycles. Three subsets were considered:
  - The columns of the form as shown: $n = 4$ elements. This looks for variation hour by hour between subgroups.
  - The rows of the form as shown: $n = 5$ elements. This looks for variation cavity to cavity within a subgroup.
  - The Cavity data across all cycles: $n = 20$ elements. This looks for variation cavity to cavity between subgroups.

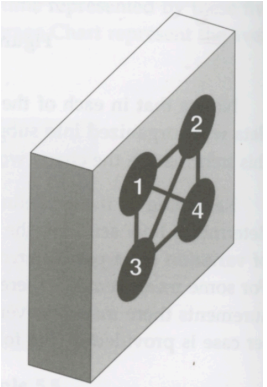


Figure 5.10: Socket Mold

	Cycle of Press				
(Hour of Collection)	*A*	*B*	*C*	*D*	*E*
Cavity I	15	16	17	16	18
Cavity II	10	13	11	10	10
Cavity III	7	8	10	7	10
Cavity IV	8	9	10	10	10

Figure 5.9: Hourly Socket Thickness Data

Summary of Results

Data analysis and process changes

- End of first set of cycles:
  - Parts from Cavity I were consistently much bigger than from the others and some exceeded specification. A shim was installed to reduce the size of the part. A second series of cycles was started.
- End of second set of cycles:
  - Parts from Cavity II and Cavity III had different results than from the first set of cycles. It was discovered the tool maker, who inserted the shim for Cavity I, noticed there was wax buildup in these cavities, so he cleaned them up. Regular cleaning was added to the process.
- Monitoring of subsequent sets of cycles:
  - Consistent, stable process noted.
- The discussion in Wheeler & Cambers speaks to the rationale for these conclusions and why the evidence appears in some of the charts, but not in others.

Principles of Subgrouping

Wheeler & Cambers pp. 112 - 114

- Never knowlingly subgroup unlike things together.
- Minimize variation within each subgroup.
- Minimize the opportunity for variation between subgroups.
- Average across noise, not across signals.
- Treat the chart in accordance with the use of the data.
- Establish operational definitions for the sampling procedure.

On Learning More

- Time Series generally is a statistics graduate level course. At the UW, one of the professors teaching the course has a side gig with the Fred Hutchinson Cancer Institute.
- SPC is a 3rd year undergraduate course in Industrial & Systems Engineering at the University of Washington
  - Industial Engineers layout factories, such as determining production flows, workstations, personnel duties, machine placement, numerical machine programming, use of robots, and containment and discard of materials such as paint overspray.
    - Recall Ford was contracted to build respirators early in the COVID-19 epidemic. The CEO at the time remarked that it took his engineers 1 day to setup a factory to do this. This was possible because Ford builds complex things from complex parts, so the engineers had plenty of practice in implementing the needed processes.
    - Ford also built the B-24 factory at Willow Run, Michigan during WW II. At its peak, the factory finished an airplane every 63 minutes, every day, 7 days a week, 24 hours a day.