SD Prisms 2D & 3D: VISUALIZING THE STANDARD DEVIATION

Hieu Nguyen, Tom Nguyen, Dr. Mamunur Rashid, DePauw University, Greencastle, IN
Summer 2022

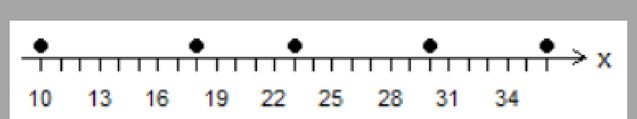
Abstract

SD Prism is a graphical R package used for visualizing the standard deviation of a data set. Given a raw data set, the standard deviation (SD) is defined as the square root of the sample variance. Sarkar and Rashid (2017) interpret the sample SD as the square root of twice the mean square of all pairwise half deviations between any two sample observations. This interpretation leads to a geometric visualization of the sample SD and a more elementary explanation as to why the denominator in the sample variance is one less than the sample size. In this article, we will explain step by step how to understand it mathematically and how the package implements the methodology to visualize the SD

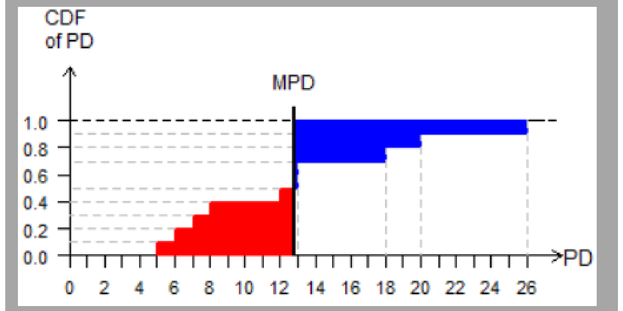
Mathematical expression & SP Prism 2D

There ar e 4 steps to construct the sample SD geometrically

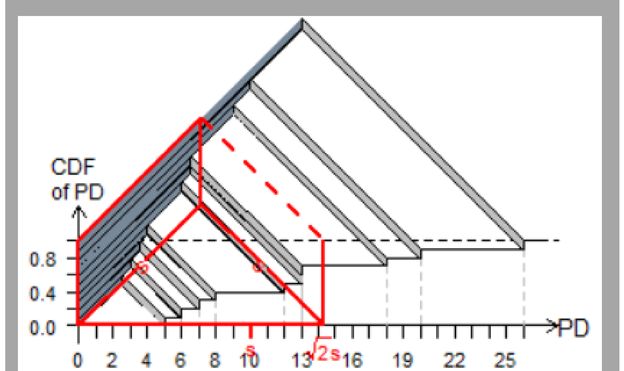
Step 1 (Dot plot of PD): Projecting the marked points (surrounded by tiny circles) vertically up, we construct the dot plot of all PD's, which we depict just above the given dot plot of x. Note that even though the marked points are distinct, some of them may be vertically aligned. Hence, these $\binom{n}{2}$ PD's need not be distinct.



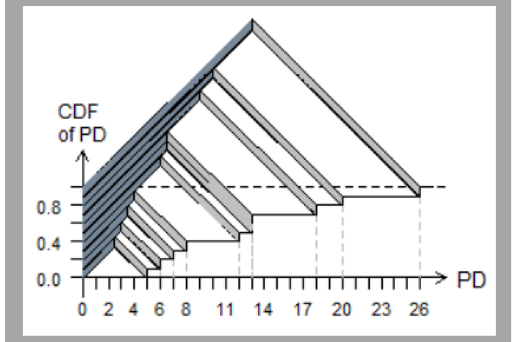
Step 2 (CDF of PD): Suppose that the distinct values of the PD's are $0 < d_1 < d_2 < \ldots < d_k$ with associated frequencies $f_1, f_2, .. f_k$. Clearly then $f_1 + f_2 + .. + f_k = \binom{n}{2}$. Draw the cumulative distribution function (CDF) of the PD's: It is a step function of the form $y = G(d)$, which begins at height $y = 0$ for $d \leq 0$, has jumps of magnitude $f_k/\binom{n}{2}$ occurring at d_k for $k = 1, 2, \ldots, K$ and ends at height $y = 1$ for $d \geq d_K$.

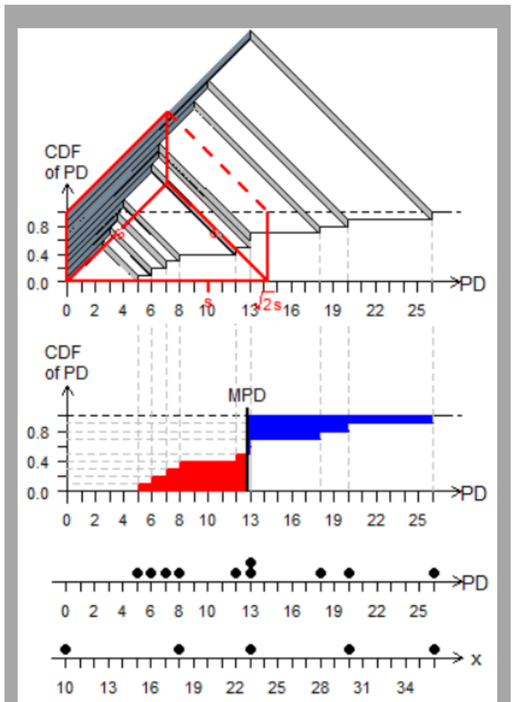


Step 4 (Search for a composite right prism) : Consider a single composite right prism having y-thickness 1 and a xz-cross-section in the shape of an RIT. Allow the size of this RIT to vary by changing the x-size until we find a suitable size such that the volume of the single composite right prism equals V_+. The intermediate value theorem guarantees the existence and uniqueness of this composite prism. Then each leg of the RIT of the composite prism equals the sample SD, and the hypotenuse of that RIT is $\sqrt{2}$ times the sample SD, or the RMS-PD.



Step 3 (Erect right prisms): If we superimpose on the graph of the CDF of PD horizontal lines $y = t/\binom{n}{2}$ for $t = 1, 2, \ldots, \binom{n}{2}$, stretching from l to the CDF, we see $\binom{n}{2}$ rectangles corresponding to the $\binom{n}{2}$ PD's. Each rectangle has width given by a PD and height $1/\binom{n}{2}$. Using each of these $\binom{n}{2}$ rectangles as the xy-face, we erect a right prism of y-thickness $1/\binom{n}{2}$ and xz-cross-section given by an RIT whose hypotenuse equals the width of the rectangle (or the corresponding PD). Then the total volume V_+ of all $\binom{n}{2}$ such right prisms is precisely the left hand side of expression (5), or the MS-PHD. Therefore, in view of Proposition 1, the total volume V_+ also equals half the sample variance. Optionally, rectangles that are equally wide (representing duplicate PD's) may be joined together by removing the internal horizontal lines.

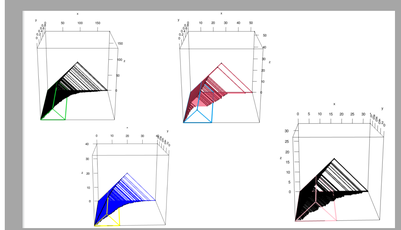


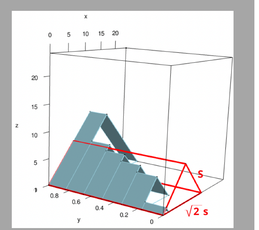


SD Prism 3D

The SDPrism3D function enhances the output of the SDPrism2D function and produces a 3D object that the user can rotate and visualize from any desired direction. It requires a third-party software called XQuart to generate the 3D object.
It also invokes two R packages rgl and plot3d.
The function takes three arguments, with only one mandatory argument.

- **Data**: the user must input the data set, usually a vector of values.
- **PriCol**: The color for the individual prisms. Users can use a number or a word to specify the color they desire. The default color is light blue.
- **sdCol:** The color for the composite prism whose size represents the Standard Deviation. Users can use a number or a word to specify the color they desire. The default color is red





These models are intractable as the users can change the color, rotate and zoom in/out. These are for user experience and increasing the package application

Acknowledgment

We give special thanks to J. William Asher and Melanie J. Norton Endowed Fund in the Sciences for sponsoring our project. Thanks to Science Research Fellow and Student-Faculty Summer Research collaborator for this opportunity. Thanks the faculty authors for their guidance and support.