STAT 515

Statistical Methods I

Sections 2.4-2.5

The Variance and Standard Deviation

Brian Habing
Department of Statistics
University of South Carolina

Outline

- The logic behind the variance and standard deviation
- The Empirical Rule and Chebychev's Inequality
- Z Scores

Example

Consider the data set 2, 3, 3, 4, 8

Measures of Variability

A measure of variability describes how spread out the data is.

Three common measures of spread are the range, the standard deviation, and the inter-quartile range.

The Range

The Range is the largest observation minus the smallest observation.

Something more informative?

The Standard Deviation

Variance:

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

Standard Deviation:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

68-95-99.7

As we will see in Chapter 5, for a "normal distribution" roughly 68% of the observations fall within 1*s* of the mean, 95% within 2*s*, and 99.7 within 3*s*. (Text suggests range/4 to range/6 as an estimate of *s*.)

WSJ BLOGS

Real Time Economics

Economic insight and analysis from The Wall Street Journal.

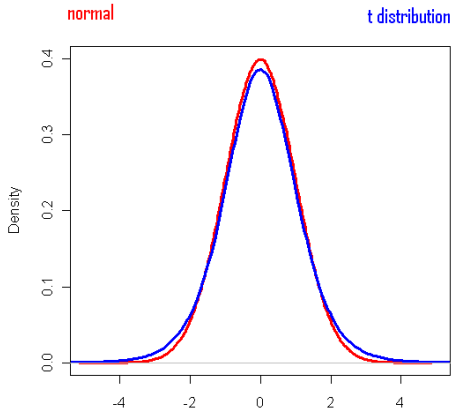
SEPTEMBER 23, 2010, 4:38 PM ET

Volcker Spares No One in Broad Critique

By Damian Paletta

Former **Federal Reserve** Chairman **Paul Volcker** scrapped a prepared speech he had planned to deliver at the **Federal Reserve Bank of Chicago** on Thursday, and instead delivered a blistering, off-the-cuff critique leveled at nearly every corner of the financial system.

10) Risk management — "Markets that are prone to excesses in one direction or another are not simply managed under the assumption that we can assume that everybody follows a normal distribution curve. Normal distribution curves — if I would submit to you — do not exist in financial markets. Its not that they are fat tails, they don't exist. I keep hearing about fat tails, and Jesus, it's only supposed to occur every 100 years, and it appears every 10 years."



Chebyshev's Inequality

Chebyshev's inequality says that at least $1-1/k^2$ of the data must be within k standard deviations of the mean for any distribution (that has a mean and sd).

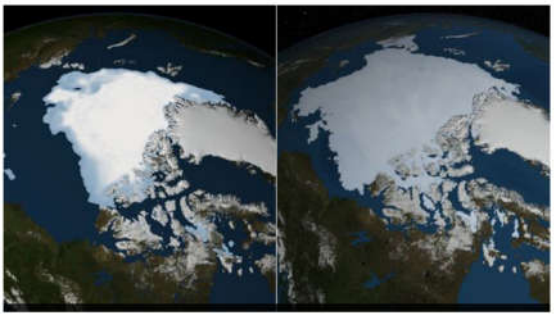
Equivalently that at most $1/k^2$ of the data can be further than k standard deviations away from the mean.

Example

Starting salaries in your field have a mean of \$31,000 and a standard deviation of \$5,000. What can we say about the percent with a starting salary between \$26,000 and \$36,000?

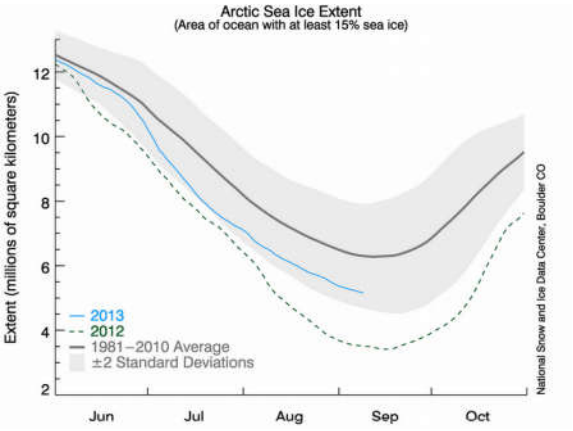
Arctic sea ice up 60 percent in 2013

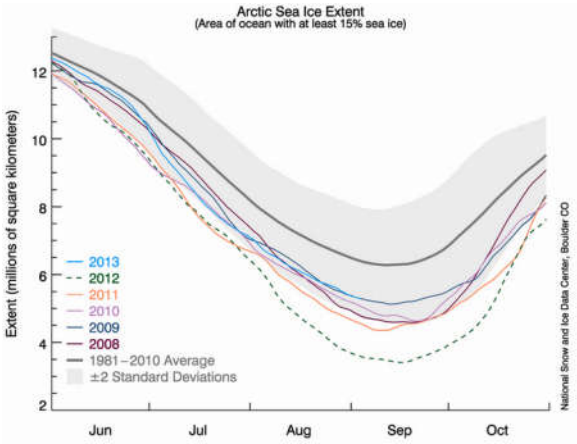
Published September 09, 2013 / FoxNews.com



Print
Email

About a million more square miles of ocean are covered in ice in 2013 than in 2012, a whopping 60 percent increase -- and a dramatic





Others…

A huge number of other bounds also exist (check out Chebyshev's inequality on Wikipedia for example).

The Vysochanskiï-Petunin inequality says that for a unimodal distribution that at most the, percent of observations more than k standard deviations from the mean is the maximum of $(4/(3k^2)$ -1/3 and 4 $/(9k^2)$)

Conditions such as continuity and symmetry can also affect the bounds.

Z-Scores

The 100 Greenest Companies in America

Inside NEWSWEEK's exclusive rankings of the world's most environmentally friendly companies

GREEN RANKINGS 2010

Our exclusive environmental ranking of the 500 largest U.S. companies and the 100 biggest global corporations.

RANKING THE COMPANIES

To calculate a company's overall ranking, the three component scores were standardized, combined using a weighted average, and mapped to a 100-point scale for publication.

The raw component scores were first converted to standardized values called Z scores, which reflect how individual companies performed in relation to the average for each of the three scores. These Z scores serve as a common metric, allowing environmental impact, green policies, and reputation—which were measured in very different ways—to be compared, much the way fractions must be converted to have a common denominator before they can be added together.