In a nutshell

In a nutshell

- SD is non-negative, and has the same units as the list

In a nutshell

- SD is non-negative, and has the same units as the list

Main property

No matter what the list, the vast majority of entries will be in the range **average $\pm$ a few SDs.**

In a nutshell

- SD is non-negative, and has the same units as the list

Main property

No matter what the list, the vast majority of entries will be in the range **average $\pm$ a few SDs.**

Pafnuty Lvovich Chebychev (1821-1894)

In a nutshell

- SD is non-negative, and has the same units as the list

Main property

No matter what the list, the vast majority of entries will be in the range **average $\pm$ a few SDs.**

Pafnuty Lvovich Chebychev (1821-1894) made this precise.

In a nutshell

- SD is non-negative, and has the same units as the list

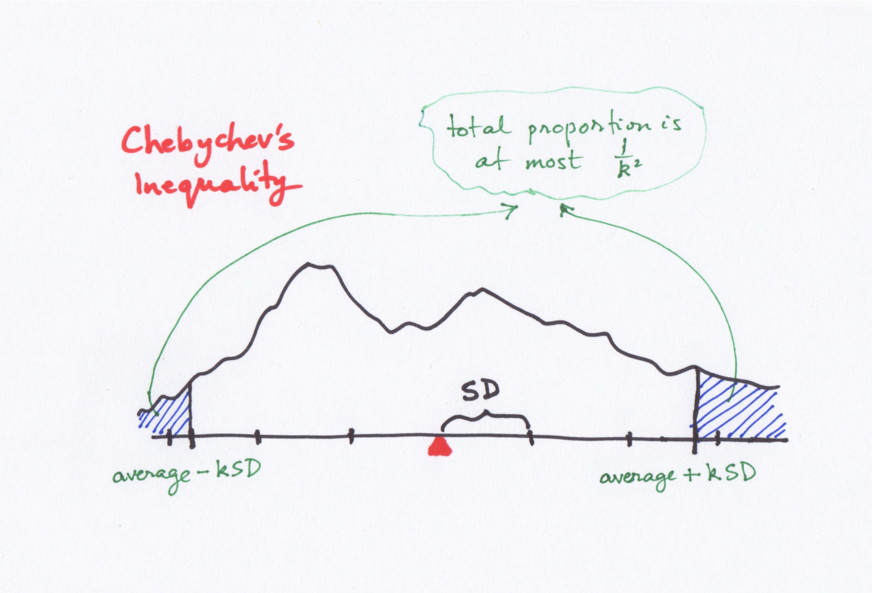
Main property

No matter what the list, the vast majority of entries will be in the range **average $\pm$ a few SDs.**

Pafnuty Lvovich Chebychev (1821-1894) made this precise.

(Tschebyscheff!!)

Chebychev's inequality



A formal statement

A formal statement

Let k be any positive number.

A formal statement

Let k be any positive number.

Chebychev's inequality

In any list, the proportion of entries that are k or more SDs away from the average is at most $1/k^2$.

A formal statement

Let k be any positive number.

Chebychev's inequality

In any list, the proportion of entries that are k or more SDs away from the average is at most $1/k^2$.

Example: In any list of numbers,

A formal statement

Let k be any positive number.

Chebychev's inequality

In any list, the proportion of entries that are k or more SDs away from the average is at most $1/k^2$.

Example: In any list of numbers,
the proportion of entries that are 2 or more SDs away from the average is at most $1/4$.

A formal statement

Let k be any positive number.

Chebychev's inequality

In any list, the proportion of entries that are k or more SDs away from the average is at most $1/k^2$.

Example: In any list of numbers,
the proportion of entries that are 2 or more SDs away from the average is at most $1/4$.

- **At most 25%** of the list is **outside** the interval **average** $\pm$ **2**$\times$**SD**.

A formal statement

Let k be any positive number.

Chebychev's inequality

In any list, the proportion of entries that are k or more SDs away from the average is at most $1/k^2$.

Example: In any list of numbers,
the proportion of entries that are 2 or more SDs away from the average is at most $1/4$.

- **At most 25%** of the list is **outside** the interval **average $\pm$ 2$\times$SD**.
- **At least 75%** of the list is **in** the interval **average $\pm$ 2$\times$SD**.

Why the SD is a good measure of spread

In any list, at most 1/9 of the entries are 3 or more SDs away from average.

Why the SD is a good measure of spread

In any list, at most 1/9 of the entries are 3 or more SDs away from average.

At least 8/9 of the entries are in the range average $\pm$ 3$\times$SD.

Why the SD is a good measure of spread

In any list, at most 1/9 of the entries are 3 or more SDs away from average.

At least 8/9 of the entries are in the range average $\pm$ 3$\times$SD.

Rough statement

No matter what the list, the vast majority of entries will be in the range **average $\pm$ a few SDs.**

Why the SD is a good measure of spread

In any list, at most $1/9$ of the entries are 3 or more SDs away from average.

At least 8/9 of the entries are in the range average $\pm$ 3$\times$SD.

Rough statement

No matter what the list, the vast majority of entries will be in the range **average $\pm$ a few SDs.**

Precise statement

No matter what the list, a proportion of at least $1 - 1/k^2$ of the entries will be in the range **average $\pm$ $k\times$SD.**

A bound on one tail

Chebychev says that the proportion **outside** the range **average** $\pm k \times$**SD** is **at most** $1/k^2$.

A bound on one tail

Chebychev says that the proportion **outside** the range **average** $\pm k \times$**SD** is **at most** $1/k^2$.

This is an upper bound on the total proportion in the two tails:

A bound on one tail

Chebychev says that the proportion **outside** the range **average** $\pm k \times$**SD** is **at most** $1/k^2$.

This is an upper bound on the total proportion in the two tails:

$\leq$ **average** $-k \times$**SD** $\geq$ **average** $+k \times$**SD**

left tail right tail

A bound on one tail

Chebychev says that the proportion **outside** the range **average** $\pm k \times$ **SD** is **at most** $1/k^2$.

This is an upper bound on the total proportion in the two tails:

$\leq$ **average** $-k \times$ **SD**	$\geq$ **average** $+k \times$ **SD**
left tail	right tail

So it is also an upper bound on each of the two tails separately:

A bound on one tail

Chebychev says that the proportion **outside** the range **average** $\pm k \times$**SD** is **at most** $1/k^2$.

This is an upper bound on the total proportion in the two tails:

$\leq$ **average** $-k \times$**SD** — left tail

$\geq$ **average** $+k \times$**SD** — right tail

So it is also an upper bound on each of the two tails separately:

The proportion in the left tail is at most $1/k^2$.

The proportion in the right tail is at most $1/k^2$.

A bound on one tail

Chebychev says that the proportion **outside** the range **average** $\pm k \times$**SD** is **at most** $1/k^2$.

This is an upper bound on the total proportion in the two tails:

$\leq$ **average** $-k \times$**SD** — left tail

$\geq$ **average** $+k \times$**SD** — right tail

So it is also an upper bound on each of the two tails separately:

The proportion in the left tail is at most $1/k^2$.

The proportion in the right tail is at most $1/k^2$.

But you **cannot say** that the proportion in one tail is at most **half** of $1/k^2$, unless you know that the two tails are equal.

Typical use

Age: average 20 years, SD 5 years

What percent are more than 80 years old?

Typical use

Age: average 20 years, SD 5 years

What percent are more than 80 years old?

- **Markov's bound:** Age is non-negative;

Typical use

Age: average 20 years, SD 5 years

What percent are more than 80 years old?

- **Markov's bound:** Age is non-negative; $80 = 4$ times the average

Typical use

Age: average 20 years, SD 5 years

What percent are more than 80 years old?

• **Markov's bound:** Age is non-negative; $80 = 4$ times the average

Markov's inequality: the proportion is at most $1/4$, that is:

Markov: at most 25% are more than 80 years old.

Typical use

Age: average 20 years, SD 5 years

What percent are more than 80 years old?

• **Markov's bound:** Age is non-negative; $80 = 4$ times the average

Markov's inequality: the proportion is at most $1/4$, that is:

Markov: at most 25% are more than 80 years old.

Did not use the SD!

Typical use

Age: average 20 years, SD 5 years

What percent are more than 80 years old?

• **Markov's bound:** Age is non-negative; $80 = 4$ times the average
Markov's inequality: the proportion is at most $1/4$, that is:
Markov: at most 25% are more than 80 years old.
Did not use the SD!

• **Chebychev's bound:** 80 years is 60 years above average. That's 12 SDs above average.

Typical use

Age: average 20 years, SD 5 years

What percent are more than 80 years old?

- **Markov's bound:** Age is non-negative; $80 = 4$ times the average

Markov's inequality: the proportion is at most $1/4$, that is:

Markov: at most 25% are more than 80 years old.

Did not use the SD!

- **Chebychev's bound:** 80 years is 60 years above average. That's 12 SDs above average.

Chebychev's inequality, applied to the right tail:

the proportion $\geq$ **average** $+12\times$**SD** is at most $1/144 = 0.7\%$, roughly,

Typical use

Age: average 20 years, SD 5 years

What percent are more than 80 years old?

- **Markov's bound:** Age is non-negative; $80 = 4$ times the average

Markov's inequality: the proportion is at most $1/4$, that is:

Markov: at most 25% are more than 80 years old.

Did not use the SD!

- **Chebychev's bound:** 80 years is 60 years above average. That's 12 SDs above average.

Chebychev's inequality, applied to the right tail:

the proportion $\geq$ **average** $+12\times$**SD** is at most $1/144 = 0.7\%$, roughly, and so

Chebychev: at most 0.7% are more than 80 years old.

Typical use

Age: average 20 years, SD 5 years

What percent are more than 80 years old?

- **Markov's bound:** Age is non-negative; $80 = 4$ times the average

Markov's inequality: the proportion is at most $1/4$, that is:

Markov: at most 25% are more than 80 years old.

Did not use the SD!

- **Chebychev's bound:** 80 years is 60 years above average. That's 12 SDs above average.

Chebychev's inequality, applied to the right tail:

the proportion $\geq$ **average** $+12\times$**SD** is at most $1/144 = 0.7\%$, roughly, and so

Chebychev: at most 0.7% are more than 80 years old.

Both bounds are correct. But Chebychev's bound is much sharper, because it uses the SD, not just the average.