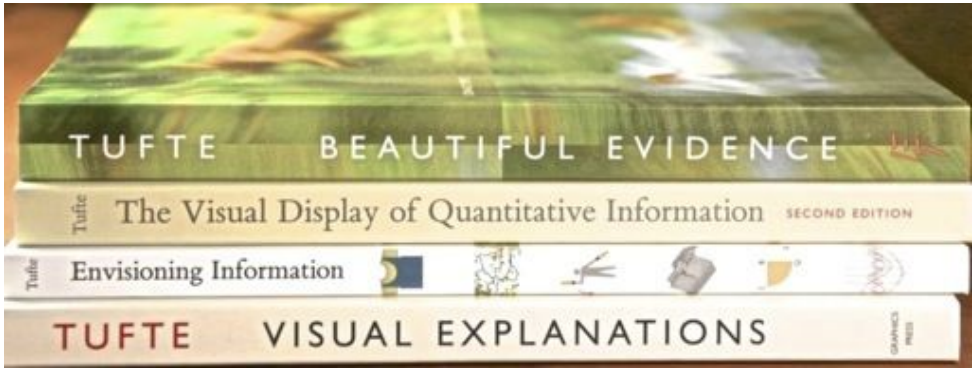
Significant Digits (in DS)

Objectives

- Motivation
- What are significant digits
- Determining significant digits by inspection
- Number of significant digits as a result of a computation
- Application to data science
- Rules-of-thumb

Motivation



The number of significant digits depends on data underlying the calculations. … Most social sciences are probably one or two digit sciences, at least on those days that we get the sign right.

— Edward Tufte

Motivation

Leading article

Too many digits: the presentation of numerical data

T J Cole

Emperor Joseph II: My dear young man, don't take it too hard. Your work is ingenious. It's quality work. And there are simply too many notes, that's all. Just cut a few and it will be perfect.

Mozart: Which few did you have in mind, Majesty?

Emperor Joseph II: Well, there it is.

Quotation from the film *Amadeus* (1984)

As a statistical reviewer for *Archives* and *BMJ* I am interested in the presentation of numerical data.[1] It concerns me that numbers are often reported to excessive precision, because too many digits can swamp the reader, overcomplicate the story and obscure the message.

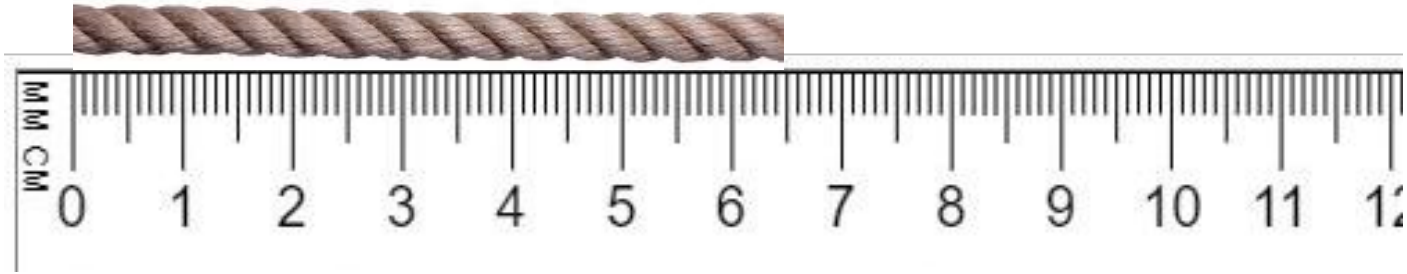
What are significant digits

The significant figures (also known as the significant digits ...) of a number are digits that carry meaning contributing to its measurement resolution.

-- Wikipedia

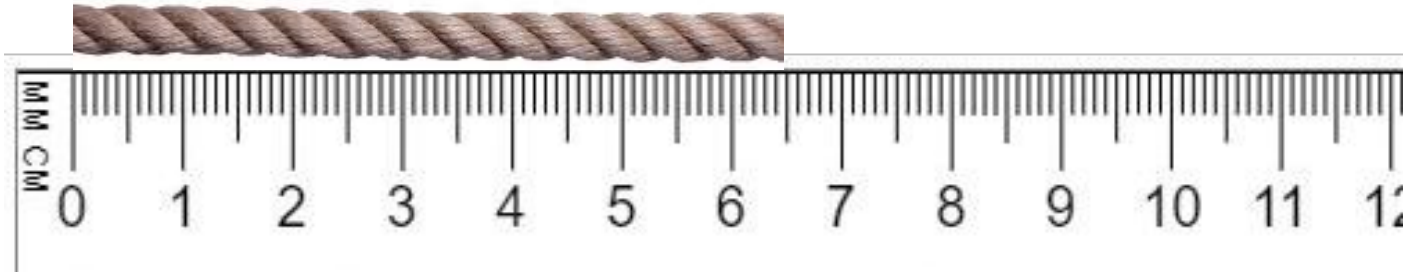
SD - the number of digits that carry meaning

How long is this rope?



SD - the number of digits that carry meaning

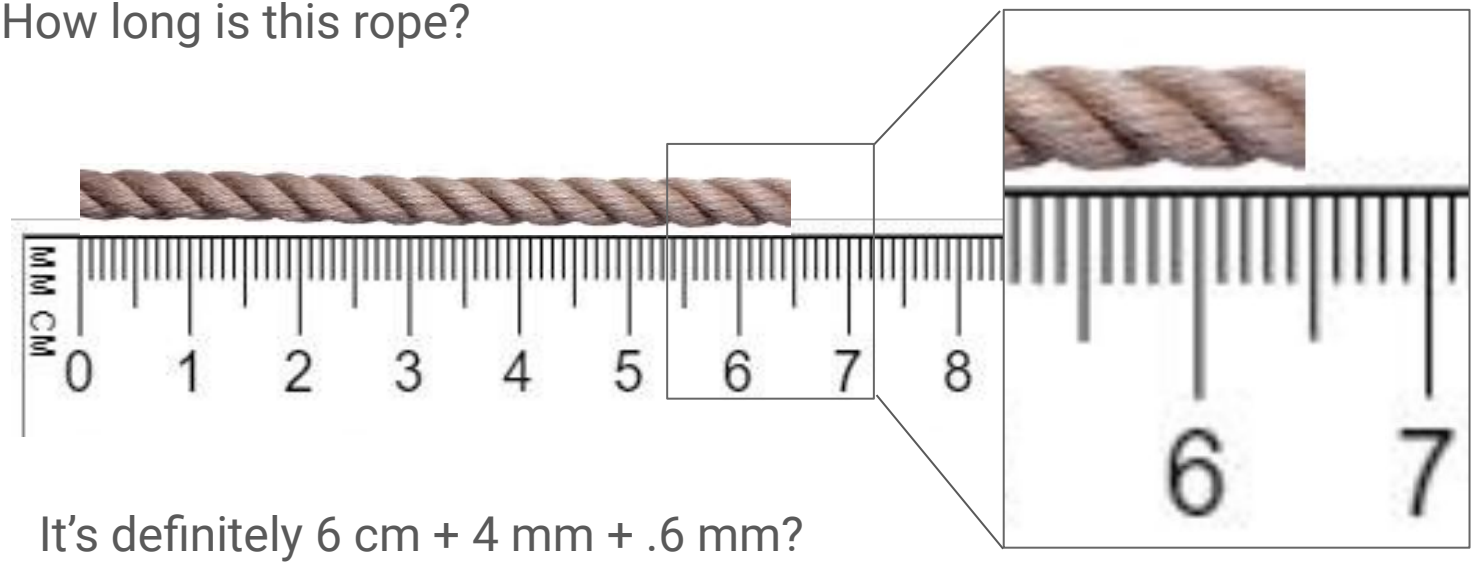
How long is this rope?



It’s definitely 6 cm + …?

SD - the number of digits that carry meaning

How long is this rope?



It's definitely 6 cm + 4 mm + .6 mm?

L = 6.46 mm

For sure

For sure

Some uncertainty, but our best guess

SD - the number of digits that carry meaning

- Significant digits are digits that carry meaning in interpreting value.
- It's useful to report the number of certain digits and the first digit where there is some uncertainty in the value.
- It is not useful (and shows lack of understanding) to report more digits after the first uncertain digit.

L = 6.46 mm <- 3 significant figures

L = 6.46180902834983 43

wat?

Identifying significant digits by inspection

- All non-zero digits are significant: 1, 2, 3, 4, 5, 6, 7, 8, 9.
- Zeros between non-zero digits are significant: 102, 2005, 50009.
- Leading zeros are never significant: 0.02, 001.887, 0.000515.
- In a number with or without a decimal point, trailing zeros (those to the right of the last non-zero digit) are significant provided they are justified by the precision of their derivation: 2.02000, 57.540.
    - Some are ambiguous, and you need to know the intent of the person recording the number: 100; 389,000

What are NOT significant:
Spurious digits introduced **by calculations carried out to greater precision than that of the original data** or measurements reported to a greater precision than the equipment supports.

— Wikipedia

Significant digits rules for computations

Addition and subtraction:

The last significant decimal place in the calculated result should be the same as the leftmost or largest decimal place of the last significant figure out of all the quantities in the calculation.

$$100.0 + 1.234 = 101.\overline{2}34\ldots \approx 101.2$$

Multiplication and division:

The calculated result should have as many significant figures as the measured number with the least number of significant figures.

$$1.234 \times 2.0 = 2.\overline{4}68\ldots \approx 2.5$$

— Wikipedia

Extrapolating these rules to data science….

How many significant digits in my linear regression coefficients and R^2?

https://davegiles.blogspot.com/2011/12/reported-accuracy-for-regression.html

Rules-of-thumb

Too_many_digits.pdf

Miscellaneous

Some numbers have infinite significant digits, so shouldn’t limit the significant digits of your output:

- pi
- a conversion (like 60 seconds per minute)
- a one-hot-encoded categorical
- categorical rankings