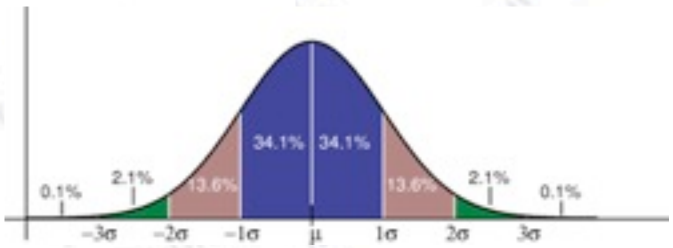
Applied Statistics

Troels C. Petersen (NBI)





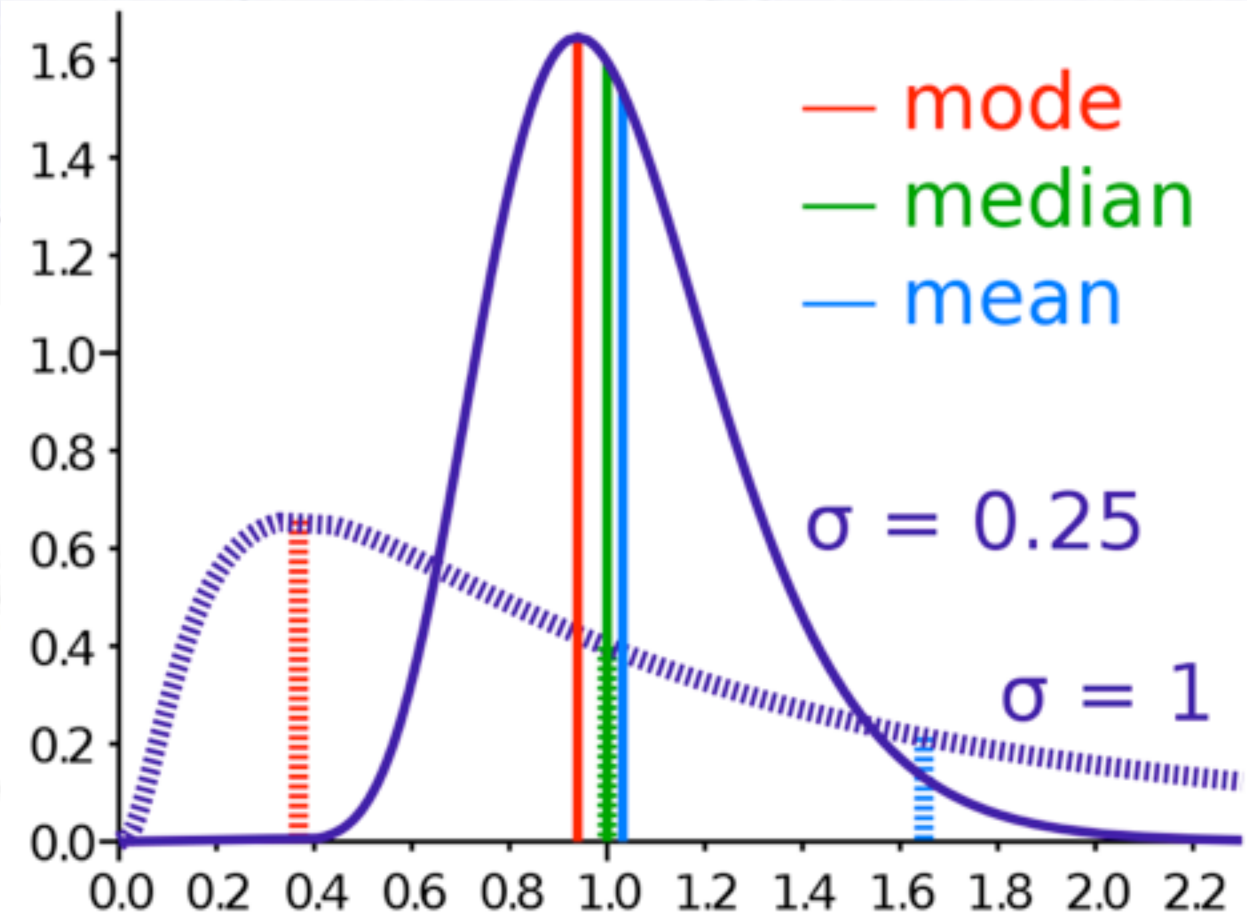
"Statistics is merely a quantization of common sense"

Defining the mean

There are several ways of defining "a typical" value from a dataset:

a) Arithmetic mean   b) Mode (most probably)   c) Median (half below, half above)

d) Geometric mean   e) Harmonic mean



Mean and width

It turns out, that the best estimator for the **mean** is (as you all know):

$$\hat{\mu} = \frac{1}{N}\sum x_i = \bar{x}$$

For the **width** of the distribution (a.k.a. **standard deviation** or **RMS**) it is:

$$\hat{\sigma} = \sqrt{\frac{1}{N}\sum (x_i - \mu)^2}$$

Note the "hat", which means "estimator". It is sometimes dropped...

Mean and width

It turns out, that the best estimator for the **mean** is (as you all know):

$$\hat{\mu} = \frac{1}{N}\sum x_i = \bar{x}$$

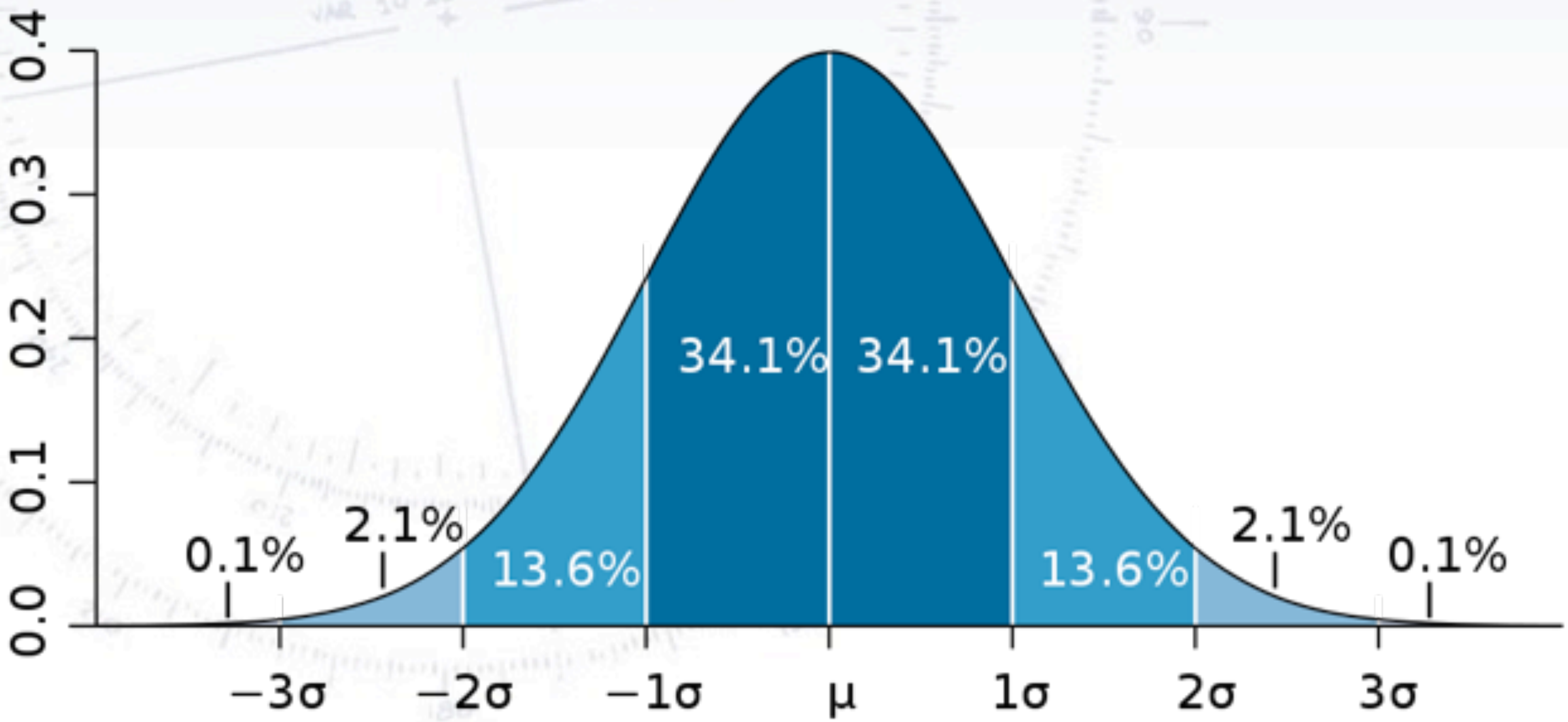
For the **width** of the distribution (a.k.a. **standard deviation** or **RMS**) it is:

$$\hat{\sigma} = \sqrt{\frac{1}{N-1}\sum (x_i - \bar{x})^2}$$

Note the "hat", which means "estimator". It is sometimes dropped...

Relation between RMS and Gaussian width...

When a distribution is Gaussian, the RMS corresponds to the Gaussian width σ:



Mean and width

What is the **uncertainty on the mean?** And how quickly does it improve with more data?

$$\hat{\sigma}_{\mu} = \hat{\sigma}/\sqrt{N}$$

<u>Example:</u>

Cavendish Experiment

(measurement of Earth's density)

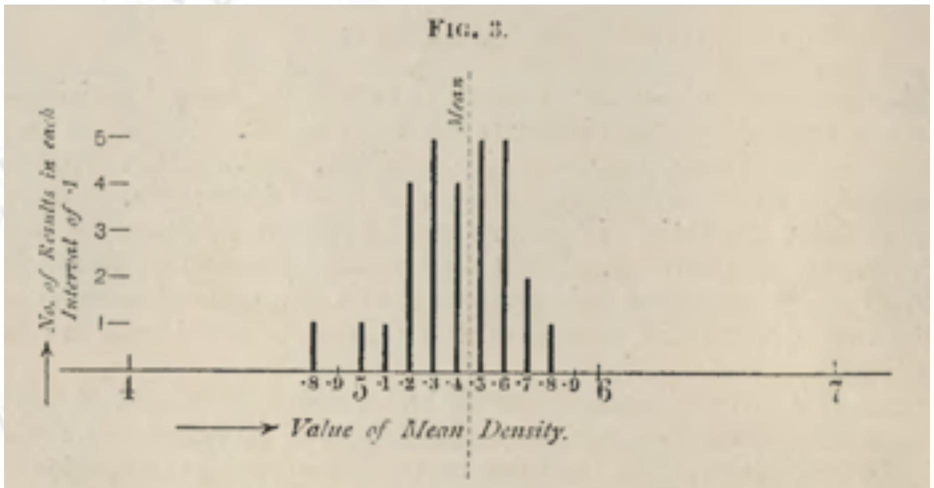
N = 29

mu = 5.42

sigma = 0.333

sigma(mu) = 0.06

Earth density = 5.42 ± 0.06



Weighted mean and width

What if we are given data, which has different uncertainties?
How to average these, and what is the uncertainty on the average?

$$\hat{\lambda} = \frac{\sum y_i/\sigma_i^2}{\sum 1/\sigma_j^2}$$

$$V[\hat{\lambda}] = \frac{1}{\sum 1/\sigma_j^2}$$