The road to figuring out the mean

Why adding up and dividing is not obvious

Damian Pavlyshyn

September 14 2021

Trial of the Pyx

Statistical Questions

- How (and how many) coins should be assayed? Larger samples were generally assumed to be better.

Statistical Questions

- How (and how many) coins should be assayed? Larger samples were generally assumed to be better.

Statistical Questions

- How (and how many) coins should be assayed? Larger samples were generally assumed to be better.
  Great pains were taken to collect an unbiased sample:

Statistical Questions

- How (and how many) coins should be assayed? Larger samples were generally assumed to be better.
  Great pains were taken to collect an unbiased sample:
  - At various times

Statistical Questions

- How (and how many) coins should be assayed? Larger samples were generally assumed to be better.
  Great pains were taken to collect an unbiased sample:
  - At various times
  - From diverse sources

Statistical Questions

- How (and how many) coins should be assayed? Larger samples were generally assumed to be better.
  Great pains were taken to collect an unbiased sample:
  - At various times
  - From diverse sources
  - Correlation was understood and avoided

Statistical Questions

- How (and how many) coins should be assayed? Larger samples were generally assumed to be better.
  Great pains were taken to collect an unbiased sample:
  - At various times
  - From diverse sources
  - Correlation was understood and avoided
- What is an acceptable irregularity in the weighing?

Edward I and the remedy (1279)

Remedy: tolerance allowed for the weight of gold measured in the Pyx.

Edward I and the remedy (1279)

Remedy: tolerance allowed for the weight of gold measured in the Pyx.

- In 1279: remedy is set at 40 grains per pound

Edward I and the remedy (1279)

Remedy: tolerance allowed for the weight of gold measured in the Pyx.

- In 1279: remedy is set at 40 grains per pound
- In 1848: remedy is set at 12 grains per pound

Edward I and the remedy (1279)

Remedy: tolerance allowed for the weight of gold measured in the Pyx.

- In 1279: remedy is set at 40 grains per pound
- In 1848: remedy is set at 12 grains per pound

Edward I and the remedy (1279)

Remedy: tolerance allowed for the weight of gold measured in the Pyx.

- In 1279: remedy is set at 40 grains per pound
- In 1848: remedy is set at 12 grains per pound

Linear! Masters of the mint could skim a little of each coin and not fear a large sample.

Edward I and the remedy (1279)

Remedy: tolerance allowed for the weight of gold measured in the Pyx.

- In 1279: remedy is set at 40 grains per pound
- In 1848: remedy is set at 12 grains per pound

Linear! Masters of the mint could skim a little of each coin and not fear a large sample.

Report from the Select Committee of the House of Commons on the Royal Mint, 1837:

Edward I and the remedy (1279)

Remedy: tolerance allowed for the weight of gold measured in the Pyx.

- In 1279: remedy is set at 40 grains per pound
- In 1848: remedy is set at 12 grains per pound

Linear! Masters of the mint could skim a little of each coin and not fear a large sample.

Report from the Select Committee of the House of Commons on the Royal Mint, 1837: *"Have not the French contractors an interest in coining money as near to the worse side of the remedy as they can [...]?*

Edward I and the remedy (1279)

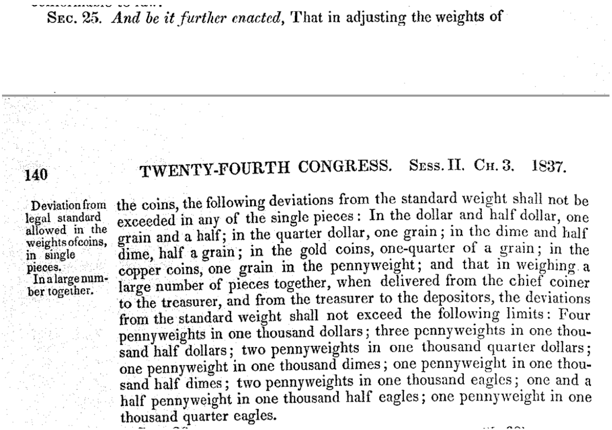
Remedy: tolerance allowed for the weight of gold measured in the Pyx.

- In 1279: remedy is set at 40 grains per pound
- In 1848: remedy is set at 12 grains per pound

Linear! Masters of the mint could skim a little of each coin and not fear a large sample.

Report from the Select Committee of the House of Commons on the Royal Mint, 1837: *"Have not the French contractors an interest in coining money as near to the worse side of the remedy as they can [...]?Decidedly, and they do take advantage of that."*

United States Congress (1837)



SEC. 25. *And be it further enacted*, That in adjusting the weights of

140 TWENTY-FOURTH CONGRESS. SESS. II. CH. 3. 1837.

Deviation from legal standard allowed in the weights of coins, in single pieces.

In a large number together.

the coins, the following deviations from the standard weight shall not be exceeded in any of the single pieces: In the dollar and half dollar, one grain and a half; in the quarter dollar, one grain; in the dime and half dime, half a grain; in the gold coins, one-quarter of a grain; in the copper coins, one grain in the pennyweight; and that in weighing a large number of pieces together, when delivered from the chief coiner to the treasurer, and from the treasurer to the depositors, the deviations from the standard weight shall not exceed the following limits: Four pennyweights in one thousand dollars; three pennyweights in one thousand half dollars; two pennyweights in one thousand quarter dollars; one pennyweight in one thousand dimes; one pennyweight in one thousand half dimes; two pennyweights in one thousand eagles; one and a half pennyweight in one thousand half eagles; one pennyweight in one thousand quarter eagles.

United States Congress (1837)

SEC. 25. *And be it further enacted*, That in adjusting the weights of

Deviation from legal standard allowed in the weights of coins, in single pieces.

In a large number together.

the coins, the following deviations from the standard weight shall not be exceeded in any of the single pieces: In the dollar and half dollar, one grain and a half; in the quarter dollar, one grain; in the dime and half dime, half a grain; in the gold coins, one-quarter of a grain; in the copper coins, one grain in the pennyweight; and that in weighing a large number of pieces together, when delivered from the chief coiner to the treasurer, and from the treasurer to the depositors, the deviations from the standard weight shall not exceed the following limits: Four pennyweights in one thousand dollars; three pennyweights in one thousand half dollars; two pennyweights in one thousand quarter dollars; one pennyweight in one thousand dimes; one pennyweight in one thousand half dimes; two pennyweights in one thousand eagles; one and a half pennyweight in one thousand half eagles; one pennyweight in one thousand quarter eagles.

United States Congress (1837)

Coin	Per coin	Per 1000 coins	Per coin $\times\sqrt{1000}$
Dollar	1.5	96	47.4
Half-dollar	1.5	72	47.4
Quarter	1	48	31.6
(Half-)Dime	0.5	24	15.8
Eagle ($10)	0.25	48	7.9

Table: Allowable error in grains

Statistics to the rescue!

Statistics to the rescue!

Unfortunately not.

Statistics to the rescue!

Unfortunately not.
Paper money and non-precious coins made the process obsolete before it could be reformed.

Statistics to the rescue!

Unfortunately not.
Paper money and non-precious coins made the process obsolete before it could be reformed.
People were concerned with quantifying and controlling error long before they had the tools and vocabulary.

How to make a point estimate?

How to make a point estimate?

Not at all new or obvious question

How to make a point estimate?

Not at all new or obvious question
Thucydides (historian), c. 450 BCE:

How to make a point estimate?

Not at all new or obvious question

Thucydides (historian), c. 450 BCE: *"[The problem was for the Athenians] to force their way over the enemy's surrounding wall [...] Their method was as follows: they constructed ladders to reach the top of the enemy's wall, and they did this by calculating the height of the wall from the number of layers of bricks at a point which was facing in their direction and had not been plastered. The layers were counted by a lot of people at the same time, and though some were likely to get the figure wrong, the majority would get it right [...]. Thus, guessing what the thickness of a single brick was, they calculated how long their ladders would have to be"*

al-Biruni (astronomer), 1025:

al-Biruni (astronomer), 1025: *"As to the halving of the interval between the two times, it is a rule of procedure which has been adopted by calculators for the purpose of minimizing errors of observation so that the time calculated will be between the upper and lower bounds"*

al-Biruni (astronomer), 1025: *"As to the halving of the interval between the two times, it is a rule of procedure which has been adopted by calculators for the purpose of minimizing errors of observation so that the time calculated will be between the upper and lower bounds"*

Thucydides (historian), c. 450 BCE

al-Biruni (astronomer), 1025: *"As to the halving of the interval between the two times, it is a rule of procedure which has been adopted by calculators for the purpose of minimizing errors of observation so that the time calculated will be between the upper and lower bounds"*

Thucydides (historian), c. 450 BCE *"Homer gives the number of ships as 1,200 and says that the crew of each Boetian ship numbered 120, and the crews of Philoctetes were fifty men for each ship. By this, I imagine, he means to express the maximum and minimum of the various ships' companies [...] If, therefore, we reckon the number by taking an average of the biggest and smallest ships [...]"*

Edward Wright (navigator), 1599:

Edward Wright (navigator), 1599:

ror herein. Exact trueth amongſt the vnconſtant waues of the ſea is not to bee looked for, though good inſtruments bee neuer ſo well applyed. Yet with heedfull diligence we may come ſo neare the trueth as the nature of the ſea, our ſight and inſtruments will ſuffer vs. Neither if there be diſagreement betwixt obſeruations, are they all by & by to be reiected? but as when many arrows are ſhot at a marke, and the marke afterwards taken away, hee may bee thought to worke according to reaſon, who to find out the place where the marke ſtood, ſhall ſeeke out the middle place amongſt all the arrowes : ſo amongſt many different obſeruations, the middlemoſt is likeſt to come neareſt the truth.

Edward Wright (navigator), 1599:

ror herein. Exact trueth amongſt the vnconſtant waues of the ſea is not to bee looked for, though good inſtruments bee neuer ſo well applyed. Yet with heedfull diligence we may come ſo neare the trueth as the nature of the ſea, our ſight and inſtruments will ſuffer vs. Neither if there be diſagreement betwixt obſeruations, are they all by & by to be reiected? but as when many arrows are ſhot at a marke, and the marke afterwards taken away, hee may bee thought to worke according to reaſon, who to find out the place where the marke ſtood, ſhall ſeeke out the middle place amongſt all the arrowes : ſo amongſt many different obſeruations, the middlemoſt is likeſt to come neareſt the truth.

Johannes Kepler (astronomer), early 17th century

Johannes Kepler (astronomer), early 17th century

On 1600 January 12/23 at 11h 50m the right ascension of Mars was:

	deg	min	sec
using the bright foot of Gemini	134	23	39
using Cor Leonis	134	27	37
using Pollux	134	23	18
at 12h 17m, using the third in the wing of Virgo	134	29	48
The mean, treating observations impartially	134	24	33

What's even the point of repeated sampling?

Robert Boyle, 1660:

What's even the point of repeated sampling?

Robert Boyle, 1660: *"Experiments ought to be estimated by their value, not their number; [...] a single experiment [...] may as well deserve an entire treatise. [...] As one of those large and orient pearls [...] may outvalue a very great number of those little [...] pearls, that are to be bought by the ounce"*

What's even the point of repeated sampling?

Robert Boyle, 1660: *"Experiments ought to be estimated by their value, not their number; [...] a single experiment [...] may as well deserve an entire treatise. [...] As one of those large and orient pearls [...] may outvalue a very great number of those little [...] pearls, that are to be bought by the ounce"*

Without a proper understanding of errors, this is not at all unreasonable.

What's even the point of repeated sampling?

Robert Boyle, 1660: *"Experiments ought to be estimated by their value, not their number; [...] a single experiment [...] may as well deserve an entire treatise. [...] As one of those large and orient pearls [...] may outvalue a very great number of those little [...] pearls, that are to be bought by the ounce"*

Without a proper understanding of errors, this is not at all unreasonable.

Thomas Simpson, 1756:

What's even the point of repeated sampling?

Robert Boyle, 1660: *"Experiments ought to be estimated by their value, not their number; [...] a single experiment [...] may as well deserve an entire treatise. [...] As one of those large and orient pearls [...] may outvalue a very great number of those little [...] pearls, that are to be bought by the ounce"*

Without a proper understanding of errors, this is not at all unreasonable.

Thomas Simpson, 1756: *"It is well known to your Lordship, that the method practiced by astronomers, in order to diminish the errors arising from the imperfections of instruments, and of the organs of sense, by taking the Mean of several observations, has not been generally received, but that some persons, of considerable note, have been of opinion, and even publickly maintained, that one single observation, taken with due care, was as much to be relied on as the Mean of a great number."*

Roger Cotes, 1722:

Roger Cotes, 1722: *"Let p be the place of some object defined by observation, q, r, s, the places of the same object from subsequent observations. Let there also be weights P, Q, R, S reciprocally proportional to the displacements which may arise from the errors in the single observations, and which are given from the given limits of error; and the weights P, Q, R, S are conceived as being placed at p, q, r, s, and their center of gravity Z is found: I say the point Z is the most probable place of the object, and may be safely had for its true place."*

Daniel Bernoulli, 1777:

Daniel Bernoulli, 1777: *"Astronomers as a class are men of the most scrupulous sagacity; it is to them therefore that I choose to propound these doubts that I have sometimes entertained about the universally accepted rule for handling several slightly discrepant observations of the same event. By these rules the observations are added together and the sum divided by the number of observations; the quotient is then accepted as the true value of the required quantity, until better and more certain information is obtained. In this way, if the several observations can be considered as having, as it were, the same weight, the center of gravity is accepted as the true position of the object under investigation."*

Gauss and the location of Ceres (1801)

Gauss's assumptions

Gauss and the location of Ceres (1801)

Gauss's assumptions

1. Small errors are more likely than large errors.

Gauss and the location of Ceres (1801)

Gauss's assumptions

1. Small errors are more likely than large errors.
2. The error distribution is symmetric around zero.

Gauss and the location of Ceres (1801)

Gauss's assumptions

1. Small errors are more likely than large errors.
2. The error distribution is symmetric around zero.
3. In the presence of several measurements of the same quantity, the most likely value of the quantity being measured is their average.

Gauss and the location of Ceres (1801)

Gauss's assumptions

1. Small errors are more likely than large errors.
2. The error distribution is symmetric around zero.
3. In the presence of several measurements of the same quantity, the most likely value of the quantity being measured is their average.

Gauss and the location of Ceres (1801)

Gauss's assumptions

1. Small errors are more likely than large errors.
2. The error distribution is symmetric around zero.
3. In the presence of several measurements of the same quantity, the most likely value of the quantity being measured is their average.

The last assumption is that the MLE of a sample is its mean.

Gauss, 1809: *"It has been customary certainly to regard as an axiom the hypothesis that if any quantity has been determined by several direct observations, made under the same circumstances and with equal care, the arithmetical mean of the observed values affords the most probable value, if not rigorously, yet very nearly at least, so that it is always most safe to adhere to it."*

Conclusion: The normal distribution seals the mean

- Lagrange develops the method of least-squares without probability and reached the same conclusion as Gauss.

Conclusion: The normal distribution seals the mean

- Lagrange develops the method of least-squares without probability and reached the same conclusion as Gauss.
- Adolphe Quetelet makes diverse observations and finds the normal distribution ubiquitous.

Conclusion: The normal distribution seals the mean

- Lagrange develops the method of least-squares without probability and reached the same conclusion as Gauss.
- Adolphe Quetelet makes diverse observations and finds the normal distribution ubiquitous.
- Galton demonstrates normal distribution by experiment.

Conclusion: The normal distribution seals the mean

- Lagrange develops the method of least-squares without probability and reached the same conclusion as Gauss.
- Adolphe Quetelet makes diverse observations and finds the normal distribution ubiquitous.
- Galton demonstrates normal distribution by experiment.
- Central limit theorem is proved in greater and greater generality.

Conclusion: The normal distribution seals the mean

- Lagrange develops the method of least-squares without probability and reached the same conclusion as Gauss.
- Adolphe Quetelet makes diverse observations and finds the normal distribution ubiquitous.
- Galton demonstrates normal distribution by experiment.
- Central limit theorem is proved in greater and greater generality.
- "Standard law of the facility of errors"

A. Bakker.
The early history of average values.
Journal of Statistics Education, 11, 2003.

C. Eisenhart.
The development of the concept of the best mean of a set of measurements from antiquity to the present day.
American Statistical Association Presidential Address, 1971.

S. Stahl.
The evolution of the normal distribution.
Mathematics Magazine, 79:96–112, 2016.

S. M. Stigler.
Statistics on the Table.
Harvard University Press, 1999.

D. Teets and K. Whitehead.
The discovery of ceres: How gauss became famous.
Mathematics Magazine, 72(2):82–93, 1999.

Gauss's proof (in modern notation)

The likelihood of a true value of θ given observations $x_1, \ldots, x_n$ is

Gauss's proof (in modern notation)

The likelihood of a true value of θ given observations $x_1, \ldots, x_n$ is

$$L(\theta) = \prod_{i=1}^{n} \varphi(x_i - \theta),$$

$$L'(\theta) = -\sum_{i=1}^{n} \varphi'(x_i - \theta) \prod_{j \neq i} \varphi(x_j - \theta)$$

$$= -L(\theta)\Big(\sum_{i=1}^{n} \frac{\varphi'(x_i - \theta)}{\varphi(x_i - \theta)} \Big)$$

Setting $\theta = \bar{x}$, which we assume is the MLE, yields

Setting $\theta = \bar{x}$, which we assume is the MLE, yields

$$0 = -L(\bar{x})\left(\sum_{i=1}^{n} \frac{\varphi'(x_i - \bar{x})}{\varphi(x_i - \bar{x})}\right),$$

$$0 = \sum_{i=1}^{n} \frac{\varphi'(x_i - \bar{x})}{\varphi(x_i - \bar{x})}$$

$$:= \sum_{i=1}^{n} f(x_i - \bar{x}).$$

Setting $\theta = \bar{x}$, which we assume is the MLE, yields

$$
\begin{aligned}
0 &= -L(\bar{x})\Big(\sum_{i=1}^{n} \frac{\varphi'(x_i - \bar{x})}{\varphi(x_i - \bar{x})}\Big), \\
0 &= \sum_{i=1}^{n} \frac{\varphi'(x_i - \bar{x})}{\varphi(x_i - \bar{x})} \\
&:= \sum_{i=1}^{n} f(x_i - \bar{x}).
\end{aligned}
$$

By symmetry of φ, we have that $f(-x) = -f(x)$.

Since our assumptions must hold for all observations x, take in particular for some arbitrary α,

Since our assumptions must hold for all observations x, take in particular for some arbitrary α,

$$\begin{aligned} x_1 &= x_0, \\ x_2, \ldots, x_n &= x_0 - n\alpha \\ \bar{x} &= x_0 - (n-1)\alpha. \end{aligned}$$

Since our assumptions must hold for all observations x, take in particular for some arbitrary α,

$$\begin{aligned} x_1 &= x_0, \\ x_2, \dots, x_n &= x_0 - n\alpha \\ \bar{x} &= x_0 - (n-1)\alpha. \end{aligned}$$

Our previous equation is then

$$\begin{aligned} 0 &= f((n-1)\alpha) + (n-1)f(-\alpha) \\ &= f((n-1)\alpha) - (n-1)f(\alpha). \end{aligned}$$

Since our assumptions must hold for all observations x, take in particular for some arbitrary α,

$$\begin{aligned} x_1 &= x_0, \\ x_2, \ldots, x_n &= x_0 - n\alpha \\ \bar{x} &= x_0 - (n-1)\alpha. \end{aligned}$$

Our previous equation is then

$$\begin{aligned} 0 &= f((n-1)\alpha) + (n-1)f(-\alpha) \\ &= f((n-1)\alpha) - (n-1)f(\alpha). \end{aligned}$$

This holds for all n and α. Together with the fact that f is continuous, this implies that, for some k,

Since our assumptions must hold for all observations x, take in particular for some arbitrary α,

$$\begin{aligned} x_1 &= x_0, \\ x_2, \ldots, x_n &= x_0 - n\alpha \\ \bar{x} &= x_0 - (n-1)\alpha. \end{aligned}$$

Our previous equation is then

$$\begin{aligned} 0 &= f((n-1)\alpha) + (n-1)f(-\alpha) \\ &= f((n-1)\alpha) - (n-1)f(\alpha). \end{aligned}$$

This holds for all n and α. Together with the fact that f is continuous, this implies that, for some k,

$$f(x) = kx.$$

We thus have the differential equation

We thus have the differential equation

$$\begin{aligned}\frac{\varphi'(x)}{\varphi(x)} &= kx, \\ \varphi(x) &= Ae^{kx^2/2}, \\ &= \frac{h}{\sqrt{\pi}}e^{-h^2x^2},\end{aligned}$$

We thus have the differential equation

$$\begin{aligned}\frac{\varphi'(x)}{\varphi(x)} &= kx,\\ \varphi(x) &= Ae^{kx^2/2},\\ &= \frac{h}{\sqrt{\pi}}e^{-h^2x^2},\end{aligned}$$

where in the last step we use the fact that φ is maximised at 0.