Hist Stat 4 Fitting models to data – The path to Least Squares. Adrien-Marie Legendre (1752-1833)

The end of the 18th century was a time of great ferment all over the globe. The French Revolution began in 1789, just a few years after the American Revolution. Zeal for new ideas affected every enterprise from politics to science. In France a whole new set of scientific units was proposed – the metric system. In the old French system, a basic unit of length was called the *toise*; it was about 6 feet long, similar to a fathom in the English system. Now a new unit of length was required, to be called the ***meter*.**

How would the meter be defined? How long was one meter? It was decided that it would be *one ten millionth of the distance from the equator to the North Pol*e. Cool. Except that meant that someone had to know the distance from the equator to the North Pole.

Of course, it's not possible to walk due north with a ruler from the equator to the North pole, so it's necessary to choose to measure land-based *sections* of some particular meridian. In 1792, a French commission decided to measure along the particular global meridian arc that went from a mountain in Barcelona, Spain, through Paris, to a tower in Dunkirk at the northern tip of France. [Figure 1]

Even that section had to be measured in smaller pieces. Furthermore, it was impossible to measure everything perfectly. The collection of imprecise partial lengths had to be used to estimate the true total length.

We include such a scientific project in the history of statistics because it illustrates a fundamental statistical task – ***the aggregation of many pieces of data into a few informative values***.

Figure 1

Many scientists had been working to get an improved sense of the shape of globe, which they called "The Figure of the Earth." Projects like the French one had also been going on under the auspices of other powers. In America, in 1763 the pre-eminent surveyors Mason and Dixon got a commission from England to do similar work while establishing the almost perfectly north-south border between Maryland and Delaware.

The final estimate of the length of a meridian arc had to be as precise as possible, which meant that it had to agree with a model of the curved shape of the earth. That shape was an "oblate spheroid," based on work by Newton in 1687. Then it was necessary to somehow combine all the measurement data to get the best set of coefficients to specify the particular instance of that shape. The statistical question was "What do you mean by *best?"*

The meridian model-fitting problem is similar to finding the best fitting *line* in a scatter plot. In that case you start with an ideal model of a line, $y = mx + b$, and then find the specific values for *m* and *b* that make the line fit the data "best." The data are a bunch of observed (x, y) pairs. The "unknowns" in the model are *m* and *b*.

With *two* data points in a scatter plot, say (2,5) and (3,8), you can write two linear equations in two unknowns.

$$5 = m(2) + b$$
$$8 = m(3) + b$$

This pair of equations has a unique solution, which will give the formula for the specific line that fits those two points. Difficulty in a fitting problem arises when there are more equations than unknowns. For example, if you have three points in the scatter plot then there may be no straight line that touches all of them. You will have three equations but only two unknowns. We say the model is over-determined.

Suppose we add the third point (4,9). Then we get three equations.

$$5 = m(2) + b$$
$$8 = m(3) + b$$
$$9 = m(4) + b$$

This system of three equations in two unknowns has no solution. There are no choices for *m* and *b* that will satisfy all three equations. The statistical solution then is to somehow find values for *m* and *b* that do a good job of estimating the “true” solution, the line we would have gotten if there had been no errors of measurement.

The problem of fitting models to imperfect data was apparent in two fields of research in the 18th century, **astronomy** (determining the orbits of planets and comets), and **geodesy** (determining the shape of the earth). From about 1750 to 1800 a succession of approaches for model fitting appeared, leading to the one we most often use now, fitting of models by ***the method of least squares***. Each method had advantages.

Mayer’s Method

In mid-18th century the German astronomer, **Tobias Mayer** (1723-1762), studied the moon’s orbit. It was extremely valuable to know the orbit accurately because navigation of ships was often determined by reference to the position of the moon. In 1707 the British Royal Navy had lost four warships and 1,550 sailors in a naval disaster off the Isles of Scilly because of navigational error. In 1714 the British Government began offering prize money for techniques or inventions that would allow for accurate determination of the longitude of a ship at sea.

Mayer was known for the precision of his observations of the moon, which he published in lunar tables. In 1755 some of his lunar tables were used by the Royal Astronomer of England to improve the accuracy of longitude determination. Mayer died in 1762 at age 39, and three years later his widow was awarded ₤3000, one of the first longitude prizes (about $600,000 in today’s money).

A fundamental contribution of Mayer’s work was his method of combining observational data to fit a model of the moon’s motion. He first published this approach in 1750. From telescope sightings Mayer had 27 data points, which were assumed to be not on a flat line but on an arc traced out by the moon. His model therefore was not $y = mx + b$, but was one that involved trigonometry to deal with the curvature. It had this form: $\beta - x = y\alpha - z\alpha\sin\theta$. His observed data were the values of *x, y,* and *z,* and his unknowns were α, β, and θ. By algebra he could solve three such equations in three unknowns, so he cleverly reduced his set of 27 equations to a set of three, by splitting them into three groups of 9 equations. He then used the sums of the coefficients sums to collapse each set of nine in to one equation. [Figure 2] Before Mayer’s publication, other astronomers had averaged observations, but no one had added or averaged coefficients of equations. His approach for reducing the number of equations gained wide usage and became known as “Mayer’s method.”

Fig. 2

Nachdem dieses geschehen war, formierte ich, nach der vorigen allgemeinen Gleichung, aus jeder Beobachtung eine besondere; und erhielte also 27 Gleichungen, die nach der Ordnung also stehen:

I) $\beta - 13°.\ 10' = +0{,}8836\alpha - 0{,}4682\alpha\sin\theta$
II) $\beta - 13.\ 8 = +0{,}9996\alpha - 0{,}0282\alpha\sin\theta$
III) $\beta - 13.\ 12 = +0{,}9899\alpha + 0{,}1421\alpha\sin\theta$
IV) $\beta - 14.\ 15 = +0{,}2221\alpha + 0{,}9750\alpha\sin\theta$
V) $\beta - 14.\ 42 = +0{,}0006\alpha + 1{,}0000\alpha\sin\theta$
VI) $\beta - 13.\ 1 = +0{,}9308\alpha - 0{,}3654\alpha\sin\theta$
VII) $\beta - 14.\ 31 = +0{,}0602\alpha - 0{,}9982\alpha\sin\theta$
VIII) $\beta - 14.\ 57 = -0{,}1570\alpha + 0{,}9876\alpha\sin\theta$
IX) $\beta - 13.\ 5 = +0{,}9097\alpha - 0{,}4152\alpha\sin\theta$
X) $\beta - 13.\ 2 = +1{,}0000\alpha + 0{,}0055\alpha\sin\theta$
XI) $\beta - 13.\ 12 = +0{,}9689\alpha + 0{,}2476\alpha\sin\theta$
XII) $\beta - 13.\ 11 = +0{,}8878\alpha + 0{,}4602\alpha\sin\theta$
XIII) $\beta - 13.\ 34 = +0{,}7549\alpha + 0{,}6558\alpha\sin\theta$
XIV) $\beta - 13.\ 53 = +0{,}5755\alpha + 0{,}8178\alpha\sin\theta$
XV) $\beta - 13.\ 58 = +0{,}3608\alpha + 0{,}9326\alpha\sin\theta$
XVI) $\beta - 14.\ 14 = +0{,}1302\alpha + 0{,}9915\alpha\sin\theta$
XVII) $\beta - 14.\ 56 = -0{,}1068\alpha + 0{,}9943\alpha\sin\theta$
XVIII) $\beta - 14.\ 47 = -0{,}3363\alpha + 0{,}9418\alpha\sin\theta$
XIX) $\beta - 15.\ 56 = -0{,}8560\alpha + 0{,}5170\alpha\sin\theta$
XX) $\beta - 13.\ 29 = +0{,}8002\alpha + 0{,}5997\alpha\sin\theta$
XXI) $\beta - 15.\ 55 = -0{,}9952\alpha - 0{,}0982\alpha\sin\theta$
XXII) $\beta - 15.\ 39 = -0{,}8409\alpha + 0{,}5412\alpha\sin\theta$
XXIII) $\beta - 16.\ 9 = -0{,}9429\alpha + 0{,}3330\alpha\sin\theta$
XXIV) $\beta - 16.\ 22 = -0{,}9768\alpha + 0{,}2141\alpha\sin\theta$
XXV) $\beta - 15.\ 38 = -0{,}6262\alpha - 0{,}7797\alpha\sin\theta$
XXVI) $\beta - 14.\ 54 = -0{,}4091\alpha - 0{,}9125\alpha\sin\theta$
XXVII) $\beta - 13.\ 7 = +0{,}9284\alpha - 0{,}3746\alpha\sin\theta$

Group	Equation
I	$9\beta - 118°8' = +8.4987\alpha - 0.7932\alpha\sin\theta$
II	$9\beta - 140°17' = -6.1404\alpha + 1.7443\alpha\sin\theta$
III	$9\beta - 127°32' = +2.7977\alpha + 7.9649\alpha\sin\theta$

The three equations were derived by adding equations 1, 2, 3, 6, 9, 10, 11, 12. and 27 in group I, equations 8, 18, 19, 21, 22, 23, 24, 25, and 26 in group II, and the rest in group III.

Boscovich’s method of least absolute deviations

Mayer’s method made sense and was fairly easy to execute, but there was no over-riding theory behind how to select the equations for the various groups, and no criteria for judging how good the final fit was. In contrast, the next development in model fitting was based on the general principle to ***minimize total error*** between the model and the data. One way to accomplish this is to minimizes ***the sum of the absolute values of the vertical deviations*** of the data from the model. The Jesuit priest and mathematician, **Roger Boscovich** (1711-1787) developed this approach. Boscovich was born in Croatia but was educated and worked in Italy, often on projects for Pope Benedict.

In 1755, he had published an analysis of data from five locations on the globe to test Newton's hypothesis that the earth is an oblate sphere (flattened at the poles and bulging at the equator). He had concluded that Newton *might* be wrong.

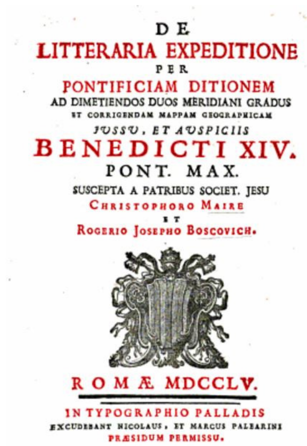
> "Thus it is evident that the determination … cannot be reconciled with the ellipse of Newton…" (1755)

Nonetheless, he continued working with these data. In 1760 he republished the 1755 data explaining how he now used a new method of analysis. This time his results were more in accordance with Newton's hypothesis. The significant contribution of Boscovich's new method was that, unlike Mayer's intuitive approach, *it was based on a set of principles* that a model fit should satisfy.

> … these three conditions are complied with: the first, that their differences shall be proportional to the differences between the versed sines of twice their latitudes; the second, that ***the sum of the positive corrections shall be equal to the sum of the negative ones***; the third, that ***the sum of all the corrections, positive as well as negative, shall be the least possible***… [Boscovich 1760, as translated from the French by Stigler]

The first condition about "versed sines" reflected the nature of his particular model, that it was measuring a curved arc. (The versed sine of an angle is 1 minus the cosine. It was a common entry in trig tables of the 18th century.) The second condition is that the negative deviations from the final model balance the positive ones. And the third is the one that calls for minimizing the sum of the absolute values of the deviations.

The data that Boscovich used in 1755 were measurements taken at five spots on the globe. At each location researchers had measured the length (in toise units) of one degree of latitude. For a perfectly spherical earth all these lengths would be equal. For an oblate sphere the length of a degree of latitude would be slightly longer at the poles. Boscovich's revised analysis based on his three principles showed that the earth was essentially an oblate sphere. In the tables you can see the arc lengths at the five locations. (57,000 toise is about 69 miles.) Note that Quito is near the equator, and Lapland is near the North Pole.


DE
LITTERARIA EXPEDITIONE
PER
PONTIFICIAM DITIONEM
AD DIMETIENDOS DUOS MERIDIANI GRADUS
ET CORRIGENDAM MAPPAM GEOGRAPHICAM
JUSSU, ET AUSPICIIS
BENEDICTI XIV.
PONT. MAX.
SUSCEPTA A PATRIBUS SOCIET. JESU
CHRISTOPHORO MAIRE
ET
ROGERIO JOSEPHO BOSCOVICH.
ROMÆ MDCCLV.
IN TYPOGRAPHIO PALLADIS
EXCUDEBANT NICOLAUS, ET MARCUS PALEARINI
PRÆSIDUM PERMISSU.

Cover of Boscovich's 1755 publication

Gradus	Latitudo ° ′	$\frac{1}{2}$ lin.v. ad rad. 10000	Hexapedæ	Diff. a primo obſerv	Diff. computata	Error
Quitenſis	0, 0	0	56751	0	0	0
Prom.B.S.	33,18	2987	57037	286	240	—46
Romanus	42,59	4648	56979	228	372	144
Pariſien.	49,23	5762	57074	323	461	138
Lapponic.	66,19	8386	57422	671	671	0

Location	Latitude (θ)	Arc length (toises)	Boscovich's $\sin^2\theta \times 10^4$
(1) Quito	0°0′	56,751	0
(2) Cape of Good Hope	33°18′	57,037	2,987
(3) Rome	42°59′	56,979	4,648
(4) Paris	49°23′	57,074	5,762
(5) Lapland	66°19′	57,422	8,386

Source: Boscovich and Maire (1755, p. 500). Reprinted in Boscovich and Maire (1770, p. 482).

Translation as shown in Stigler, History of Statistics

The basic algebraic model had the form $a = z + y\sin^2\theta$, where a and θ were observed data values while z and y were the unknowns. Adhering to his three principles, Boscovich, following Newton's techniques, developed a geometrically based algorithm to solve this system of five equations in two unknowns. His approach was later formalized algebraically and made popular by the great French mathematician, **Pierre-Simon Laplace**, who illustrated it in his major work, *Mécanique céleste (Celestial Mechanics)* in 1799.

The Boscovich-Laplace approach of fitting by least absolute deviations is still sometimes used today, particularly in cases where data are not from a normal distribution.

Legendre and the method of least squares

In this approach the "best" model is the one that minimizes the ***sum of the squares** of the vertical deviations*. This method is now the most commonly used to establish goodness of fit. For example, Excel uses it to get the regression line on a scatter plot. The first person to give a thorough explanation and justification of this approach was **Adrien-Marie Legendre**, another French mathematician who at the end of the 18th century, like Laplace, was working on problems both in astronomy and geodesy.

Legendre was one of the people calculating the length of the meridian through Paris described in Figure 1. In 1798 he had described reducing a set of too many equations in these data and had mentioned the need to "balance the errors," but he had arrived at a solution in a fairly arbitrary way. Then, in 1805, while he was writing a paper, *New Methods for determining the orbits of comets*, he saw that he could improve his solution to the too-many-equations problem by minimizing the *sum of the squares of the errors*.

Importantly, he realized that his new technique was equally applicable in observations on the shape of the earth. So he added an appendix, *Sur la Methode des Moindres Quarres*, (*On the Method of Least Squares*). Perhaps to stress the generalizability of his method, he gave a clearly worked out illustration using the data from the meridian study, even though the memoir was about the orbits of comets. Here is part of his introduction.

On the Method of least squares

…

If there are the same number of equations as unknowns *x*, *y*, *z*, &c., there is no difficulty in determining the unknowns, and the error *E* can be made absolutely zero. But more often the number of equations is greater than that of the unknowns, and it is impossible to do away with all the errors.

In a situation of this sort, which is the usual thing in physical and astronomical problems, where there is an attempt to determine certain important components, a degree of arbitrariness necessarily enters in the distribution of the errors, and it is not to be expected that all the hypotheses shall lead to exactly the same results; but it is particularly important to proceed in such a way that extreme errors, whether positive or negative, shall be confined within as narrow limits as possible.

Of all the principles which can be proposed for that purpose, I think there is none more general, more exact, and more easy of application, than that of which we made use in the preceding researches, and which consists of rendering the sum of squares of the errors a minimum. By this means, there is established among the errors a sort of equilibrium which, preventing the extremes from exerting an undue influence, is very well fitted to reveal that state of the system which most nearly approaches the truth.

Sur la Méthode des moindres quarrés.

Si l'on a autant d'équations que d'inconnues *x*, *y*, *z*, &c., il n'y a aucune difficulté pour la détermination de ces inconnues, et on peut rendre les erreurs E absolument nulles. Mais le plus souvent, le nombre des équations est supérieur à celui des inconnues, et il est impossible d'anéantir toutes les erreurs.

Dans cette circonstance, qui est celle de la plupart des problêmes physiques et astronomiques, où l'on cherche à déterminer quelques élémens importans, il entre nécessairement de l'arbitraire dans la distribution des erreurs, et on ne doit pas s'attendre que toutes les hypothèses conduiront exactement aux mêmes résultats; mais il faut sur-tout faire en sorte que les erreurs extrêmes, sans avoir égard à leurs signes, soient renfermées dans les limites les plus étroites qu'il est possible.

De tous les principes qu'on peut proposer pour cet objet, je pense qu'il n'en est pas de plus général, de plus exact, ni d'une application plus facile que celui dont nous avons fait usage dans les recherches précédentes, et qui consiste à rendre *minimum* la somme des quarrés des erreurs. Par ce moyen, il s'établit entre les erreurs une sorte d'équilibre qui empêchant les extrêmes de prévaloir, est très-propre à faire connoître l'état du système le plus proche de la vérité.

The next major development in model fitting was the work of **Laplace** and **Gauss** to attach probabilities to the final estimates found by least squares. That work took shape over the next twenty years.

Sources:

For the image in Figure 1. Alder, Ken *The Measure of all Things* Free Press 2003

For the table in Figure 2. Mayer, Tobias, in Kosmographische Nachrichten und Sammlungen auf das Jahr 1748

For general background and English version of Boscovich. Stigler, Stephen, *The History of Statistics*, Harvard U. Press. 1986

For Latin version of Boscovich, *De Litteraria Expeditione.* https://books.google.com/books?id=UDigAAAAMAAJ

Legendre, A. M. *Nouvelles méthodes pour la détermination des orbites des comètes*, Didot, 1805, available online.

For English translation of Legendre. Smith, D. E. *A Source Book in Mathematics,* Dover Publications, 1959