Statistical Science
2000, Vol. 15, No. 3, 254–278

Scientific Method, Statistical Method and the Speed of Light

R. J. MacKay and R. W. Oldford

Abstract. What is "statistical method"? Is it the same as "scientific method"? This paper answers the first question by specifying the elements and procedures common to all statistical investigations and organizing these into a single structure. This structure is illustrated by careful examination of the first scientific study on the speed of light carried out by A. A. Michelson in 1879. Our answer to the second question is negative. To understand this a history on the speed of light up to the time of Michelson's study is presented. The larger history and the details of a single study allow us to place the method of statistics within the larger context of science.

Key words and phrases: Statistical method, scientific method, speed of light, philosophy of science, history of science.

1. INTRODUCTION

> "The unity of science consists alone in its method, not in its material" (Karl Pearson, 1892 [43], page 12, his emphasis).

> "Statistics is the branch of scientific method which deals with the data obtained counting or measuring the properties of populations of natural phenomena. In this definition "natural phenomena" includes all the happenings of the external world, whether human or not" (M. G. Kendall, 1943 [30], page 2).

The view that statistics entails the quantitative expression of scientific method has been around since the birth of statistics as a discipline. Yet statisticians have shied away from articulating the relationship between statistics and scientific method, perhaps with good reason. For centuries great minds have debated what constitutes science and its method without resolution (e.g., see [36]). And in this century, historical examinations of scientific episodes (e.g., [32]) have cast doubt on method in scientific discovery. One radical position, established by examination of the works of Galileo, is that of the philosopher Paul Feyerabend, who writes of method in science:

> . . . *the events, procedures and results that constitute the sciences have no common structure*; there are no elements that occur in every scientific investigation but are missing elsewhere (Paul Feyerabend, 1988 [19], page 1, his emphasis).

Feyerabend then proposes, somewhat facetiously, that the only universal method to be found in science is "anything goes." Whether Feyerabend's view holds for science in general is debatable; that it does not hold for statistics is the primary thesis of this paper.

By examining in some detail one particular scientific study, namely A. A. Michelson's 1879 determination of the speed of light [37], we illustrate what we consider to be the common structure of statistics, what we propose to call *statistical method*.

There are several reasons for selecting Michelson's study. First, physical science is sometimes regarded as presenting a greater challenge to the explication of statistical method than, say, medical or social science where *populations of interest are well defined*. An early instance is Edgeworth's hesitation in 1884 to describe statistics as the "Science of Means in general (including physical observations)," preferring instead the less "philosophical" compromise that it is the science "of those Means which are presented by social phenomena" [18].

Second, the speed of light in vacuum is a fundamental constant whose value has become "known";

R. J. MacKay is Associate Professor, Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1. R. W. Oldford is Associate Dean of Computing, Faculty of Mathematics, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1.

in 1974, it was *defined* to be 299,792.458 km/s. So we are in the extremely rare inferential position of "knowing the answer." (By that time the determinations had so little variability that it was considered known to 1 part in 10^9, and the standard meter could not be measured to that great a precision. The second is similarly defined; it is the time taken for 9,192,631,770 periods of the radiation corresponding to the transition between two hyperfine levels of the atom cesium-133. Now the meter is defined to be the distance travelled by light through a vacuum in 1/299792458 second! See [9].)

Third, Michelson reported his study in an era when it was possible to publish a significant amount of detail, permitting others insight into the difficulties he faced and the solutions he found.

Fourth, the determination of the speed of light has been (and continues to be) important to science and to technology. Consequently its history is rich enough to provide a backdrop on which large scale questions of the nature of science and statistics can be discussed.

Fifth, the determinations are known in the statistical literature, first appearing in Stigler's paper [47] on robust estimates of location.

Finally, and most important, a historical study has the important characteristic of being based entirely on public material. Information gathered together into a single source is information that can be checked against common sources, that can be improved as new historical material becomes available and that can be a common test bed for others to use. To these ends, we have tried to present the history without reference to method.

These discussions require separate contexts of differing detail. A broad historical sweep is necessary to appreciate what can be meant by scientific method. It is provided in Section 2, where we give a history of the determination of the speed of light from antiquity to the late 1800s. The stage thus set, the optics, apparatus and method of Michelson's first determinations of the speed of light are described in Section 3. These provide the details necessary for discussion of statistical method. The structure which we propose is described in Section 4. Scientific method is examined in Section 5 and contrasted with statistical method in Section 6.

2. HISTORICAL BACKGROUND

The thought of Aristotle (384–322 BC) dominated western science for nearly two millenia. So powerful is his cosmology that it compels him to declare that "... light is due to the presence of something, but it is not a movement" ([5], 446^b25–447^a10). No movement, no speed. And if that were not enough, the argument for finite speed is easily dismissed:

> Empedocles (and with him all others who used the same forms of expression) was wrong in speaking of light as "travelling" or being at a given moment between the earth and its envelope, its movement being unobservable to us; that view is contrary both to the clear evidence of argument and to the observed facts; if the distance traversed were short, the movement might have been unobservable, but where the distance is from extreme East to extreme West, the strain upon our powers of belief is too great
> (*Aristotle* (384–322 BC), *On the Soul: Book II* [4], 418^b20–418^b27).

This view was echoed by many thinkers in western history: Augustine (ca. 354–430), John Pecham (ca. 1230–1292), Albert the Great (ca. 1200–1280), Thomas Aquinas (ca. 1225–1274) and Witelo (ca. 1230–ca. 1275) to name a few. So too, the opposite view was argued by some, notably Ibn Al-Haytham (ca. 965–1040) and Roger Bacon (ca. 1219-1292). But without empirical demonstration to the contrary, the case for instantaneous perception of the source could always be made. In the absence of data, arguments pro and con were forced to be based on the contemporary theory of light, or on interpretation of the conflicting views of ancient authorities, or on established religious doctrines, or on mathematical arguments that demonstrated the necessity or absurdity of one of the alternatives [34].

The debate continued into the beginning of the "scientific revolution" of the 17th century. Lindberg presents preliminary evidence of the debate in medieval Europe [34]. Such giants as Francis Bacon (1561–1626), Johannes Kepler (1571–1630) and René Descartes (1596–1650) believed the speed to be infinite. Bacon had doubts about the infinite speed when considering the great distances that light must travel from the stars to Earth but found such speed easier to swallow given the already fantastic speeds at which stars must travel in their daily orbit about the Earth! See Aphorism 46 of Book II of the Novum Organum, for example, [6].

Descartes, for example, likened the transmission of light to that of pushing on a stiff stick—the instant one end (the source) was pushed the other end (the perception) moved ([24], pages 258–259). The analogy is powerful; there is no perceptible movement anywhere along the stick, no matter how long a stick is used! Descartes strongly held this view; when his colleague and scientific mentor, Issac Beeckman (1588–1637), claimed to have performed an experiment which demonstrated the speed was finite, Descartes dismissed the claim saying that if it were true, then Descartes knows nothing of philosophy and his whole theory would be refuted! (see Note 1.) Beeckman and Descartes could not agree on an experiment to resolve the issue. (see Note 2.)

Among these giants, Galileo Galilei (1564–1642) stands alone in his disagreement; he wrote

> *Sagredo*:... I cannot believe that the action of light, however pure, can be without motion, and indeed the swiftest.
> *Salviati*: But what and how great should we take the speed of light to be? Is it instantaneous perhaps, and momentary? Or does it require time, like other movements? Could we assure ourselves by experiment which it may be?
> [*Galileo Galilei* (1564–1642), *Two New Sciences* (1638) [23], *page* 49].

In the same book, Galileo proposed a demonstration to determine whether light was instantaneous. It was essentially the same that Beeckman had proposed earlier and drew similar fire from Descartes. In a letter to the great experimental scientist Marin Mersenne (1588–1647), dated 11 October 1638, Descartes gave a scathing review (see Note 3) of Galileo's book. Of the proposed demonstration, Descartes wrote "His experiment to know if light is transmitted in an instant is useless, since eclipses of the moon, related so closely to calculations made of them, prove this incomparably better than anything that could be tested on earth" ([16], page 389). (This refutation appears to be based on the argument he gave to Beeckman as described in Note 2.) Nevertheless, the demonstration was tried in 1667 by members of the Florentine Academy, but without success [12]. Light's movement was either instantaneous or near enough so as to be too fast to measure successfully.

In 1676 the first empirical evidence of a finite speed was presented. The Danish astronomer Ole Römer (1644–1710), while investigating an entirely different matter, gathered data and found a discrepancy which led to the discovery. Interestingly, this important and purely scientific discovery came about while Römer was working on what we would today call a very applied problem.

NOTE 1 (From [14B], page 308): *"Contra ego, si quae talis mora sensu perciperetur, totam meam Philosophiam funditus eversam fore inquiebam."* A rough translation, due to our classically trained colleague G. W. Bennett, is "On the contrary, I would be worried that my entire Philosophy would be on the point of being completely overturned if any delay of this sort were to be perceived by the senses."

NOTE 2. It is doubtful that Beeckman's 1629 experiment [8] was successful. The experiment involved firing a mortar and observing its flash in a mirror situated some 1851.85 m away; the movement of a clock situated at the side of the mortar would measure the time elapsed. With today's value, the time for the flash to reach the mirror and return would be about 1/100,000 of a second! Descartes argues that even if Beeckman could detect a delay of 1/24 of a pulse beat (or about 1/24 of a second, yielding a speed of only around 89 km/s), then it should be possible to detect a delay between the occurrence and perception of a lunar eclipse of about one hour. The flaws in this argument are discussed in detail in [45].

NOTE 3. For example, "... his fashion of writing in dialogues, where he introduces three persons who do nothing but exalt each of his inventions in turn, greatly assists in [over]pricing his merchandise" ([16], page 388). The substantive criticisms are generally directed at Galileo's not having identified the causes of the phenomena he investigated. For most scientists at this time, and particularly for Descartes, that is the whole point of science.

2.1 Longitude

One of the great practical problems of that time was the determination of longitude, particularly at sea. The basis for the determination is the comparison of the local time at sea with the time at a fixed reference point—the prime meridian. If, for example, the local time is determined to be two hours earlier than the time at the prime meridian, the location must be $360 \times 2/24 = 30$ degrees longitude west of the prime meridian.

The times can be determined astronomically. For example, local time zero can be defined to be that time when some star, say Arcturus, is observed to cross the imaginary line of longitude running directly north–south through the local position; the corresponding standard time zero would be that time when the same star crosses the prime meridian. Stars are far enough away from us that these two crossings will occur at different moments of time. Carefully determined tables of prime meridian crossing times of various stars would allow navigators to set their local clock. To determine the difference between the local clock and the standard clock, closer astronomical events such as an eclipse or occultation of the moon or a planet can be used. These events are observed at essentially the same moment of time whatever the observer's location on Earth, and furthermore are predictable. So comparison of the local time of the close event with its tabulated standard time would give the time difference necessary to calculate longitude.

In 1609, after hearing Flemish reports of a spyglass constructed from two lenses that would enlarge the image of distant objects, Galileo set about the design and construction of the first astronomically useful telescope. [According to Stillman Drake ([15], page 29), Hans Lipperhey, a lens grinder from the Netherlands, is generally assigned credit for the telescope's invention and applied for its patent in 1608.] In March of the next year, Galileo reported his discovery of the four principal moons of Jupiter [21]. For the first time, here was an orbital system that was demonstrably not centred about the Earth. Galileo argued that this was compelling evidence against the the Ptolemaic system (all celestial bodies revolve around a fixed Earth) and in favour of the Copernican sun-centred system. His public support of the Copernican system as a true representation of the movement of the planets (as opposed to a convenient calculational model) brought Galileo into conflict with those who would interpret certain Biblical passages literally [22]. Some of these people wielded considerable influence within the Catholic church of Rome; by order of Pope Urban VIII he was banned from further publication and placed under house arrest from 1633 until his death in 1642. This did not prevent him from continuing his scientific work. (Today's visitor to Florence's Museum of Science can find a glass and ivory case displaying an ironic relic—Galileo's bony middle finger pointing heavenward.)

But this momentous scientific discovery also had commercial potential. King Philip III of Spain had offered a handsome prize to anyone who could come up with a practical method of determining a ship's position when out of sight of land. Galileo hit upon the idea of using the predicted times of the eclipses of Jupiter's moons to provide the common celestial clock necessary to determine longitude. In November of 1616 he began negotiations with Spain for navigational uses of his astronomical discoveries and in 1617 worked on developing a telescope for use at sea while continuing his negotiations with Spain [15]. Unfortunately the tables he produced were not accurate enough for their intended purpose—the theory at the time did not account for the perturbations of the moons due to their mutual interaction [13].

Although many writers advocated the use of telescopes at sea, those who appreciated the practical difficulty of directing a very long telescope at Jupiter while aboard a lively ship were skeptical and undoubtedly amused by the proposed method. It was never to become successful at sea. (The problem remained unsolved for more than 150 years until the development of accurate portable clocks by the English inventor John Harrison. For a popular account, see [46].) But on land, very accurate determinations of longitude could be obtained this way and resulted in a substantial reform of geography in the 17th and 18th centuries.

2.2 The First Evidence

In 1671 Römer went to Hven, an island community near Copenhagen, to help redetermine the longitude of the observatory located there. With others, he began observing a series of eclipses of Io, Jupiter's largest moon. In the end they had eight months of observations or, since Io makes one revolution of Jupiter in 42 hours, timings on about 140 eclipses over 2/3 of the year. The time intervals between these eclipses were not regular but appeared related to where the Earth was in its orbit. The length of the interval became shorter as the Earth approached Jupiter and longer as it moved away; the mathematically predicted time of an eclipse was too early if the Earth was near Jupiter and too late if the Earth was far from Jupiter. This systematic lack of fit allowed Römer to announce in Paris in September 1676 that the eclipse predicted for November 9 that year would actually occur 10 minutes later. Observation bore him out and Römer argued that the discrepancy was due to the finite speed of light. The light takes longer to reach us the farther we are from its source.

From his observations, Römer estimated that light takes about 22 minutes to cross the full diameter of Earth's orbit or about 11 minutes for light from the Sun to reach us on Earth. On this basis, he estimated light's speed to be about 214,000 km/s. (For more on Römer see [31]. For more detail on this study see [12].)

Römer's "proof" was not immediately accepted by all. Alternative explanations were provided by Gian Domenico Cassini (1625–1712) then an astronomer at the newly formed Academie des Sciences in Paris. In 1666 Cassini had published tables on the eclipses of the satellites of Jupiter from which work he also noticed inequalities in time intervals of eclipses that depended on the location of Jupiter in its own elliptical orbit. He had briefly considered a finite speed of light in 1675 but soon rejected it for a more traditional explanation. Cassini, and later his nephew Giacomo Filippo Maraldi (1665–1729), suggested that Jupiter's orbit and the motion of its satellites might explain the observed inequalities ([50], [42] and [31]). Many astronomers continued to hold the view that light's movement was instantaneous.

It was not until a study by James Bradley (1693–1762) (see [1] and [31]) was reported in 1729 that nearly all agreed that the speed is finite. Bradley had been studying the parallax of the stars and discovered an annual variation in the position of stars that could not be explained by the parallax effect. However, it could be explained by the motion of the Earth if light's speed were finite. Based on careful observations, Bradley estimated that light took 8 minutes and 12 seconds to reach the Earth from the Sun, resulting in a value for light's speed of 301,000 km/s.

In 1809, based on observations on the eclipses of Jupiter's moons for 150 years, Jean-Baptiste Joseph Delambre (1749–1822) estimated the time taken by light to travel from the Sun to the Earth to be 8 minutes and 13.2 seconds, resulting in a speed of about $300{,}267.64 \approx 300{,}300$ km/s. The time here is as reported in [42]. To calculate the speed, the distance between the Earth and the Sun must be known. In the estimate reported here, the distance used was 148,092,000 km as derived from Bradley's figures above.

The results of these early astronomical estimates are summarized in Table 1.

Unfortunately, measurements of the speed made in this way depended on the astronomical theory and observations used. Simon Newcomb (1835–1909) tells of an inaugural dissertation in 1875 by Glasenapp whereby observations of the eclipses of Io from 1848 to 1870 show that widely ranging values for the speed "could be obtained from different classes of these observations by different hypotheses" ([42], page 114). It was shown that values for the Sun-to-Earth time could be produced between 496 and 501 seconds, resulting in speeds between $295{,}592.8 \approx 295{,}600$ and $298{,}572.6 \approx 298{,}600$ km/s (again using Bradley's Earth-to-Sun distance).

Better determinations of the speed might be made if both source and observer were terrestrial. Because all would then be accessible, greater control could be exerted over the study and hence the observations. But this brings us back to the age-old problem: how could the speed of light be measured terrestrially?

TABLE 1

Studies based on astronomical observation

Year	Authors	Observational source	Speed (km/s)
1676	Römer	Jupiter satellites	214,000
1726	Bradley	Aberration of stars	301,000
1809	Delambre	Jupiter satellites	300,300

2.3 Terrestrial Determinations

Imagine two people standing at either end of a very long track. The first uncovers a powerful light source at an appointed time and the second records the time at which the light is seen. The length of the track divided by the difference between the start time and the time the light is perceived gives a measurement of the speed of light. (This is essentially the experiment proposed by Isaac Beeckman to Descartes in 1629. See Note 2.) The trouble, of course, is that light is so fast that the distance must either be very large or the time taken very small. Extremely large distances and extremely short time intervals are very difficult to measure directly.

Matters can be improved if both observers have light sources which they cover with a screen. Time measurement begins when the first observer removes the screen sending light to the second. The second light source is uncovered when the second observer sees the first. Now when the first observer sees the second light source he again screens his source. The time between uncovering and covering the first light source is a measure of the time light takes to travel twice the distance between the two observers. The improvements are obvious. The distance is doubled and a single clock has replaced two supposedly synchronized clocks. Here was Galileo's proposed study of 1638; nearly 200 years would pass before it was improved sufficiently to produce results.

The necessary innovations were introduced by Hippolyte Fizeau (1819–1896). One innovation was to replace the second person by a fixed flat mirror whose surface is perpendicular to the beam of light from the source. When this was done, the light beam was reflected directly back at its origin and one human source of variation was completely removed from the system. The second innovation was to automate the covering and uncovering of the source, thereby further reducing the variation from the first human source.

Together, these advances allowed Fizeau to replace the direct measurement of time with an indirect measurement of speed. Rather than measure time between uncovering and covering, Fizeau could measure the minimum speed that the screen must travel to cover the source at the exact time the light returns. The trick was to use an accurately machined toothed wheel placed spinning in front of the source to act as the moving screen. The teeth screen the source while the gaps uncover it and so the wheel acted just as Galileo's observer. Any light returning to the source strikes either a tooth or a gap. If the wheel was set spinning fast enough that every beam sent out struck a tooth on its way back, no image of the source is observed. Twice this speed produces a full image as the beam sent out returns through the next available gap. Three times the speed produces no image, and so on. The speed of rotation, coupled with the distance travelled (twice 8,633 m in Fizeau's setup), could be transformed into a measure of the speed of light. In this way, Fizeau produced the first terrestrial determination of the speed of light in 1849.

Others were quick to build on this monumental achievement. Only two years later Leon Foucault (1819–1868), a former collaborator of Fizeau, produced more accurate measurements based on a rotating mirror rather than a toothed wheel.

3. MICHELSON'S 1879 DETERMINATIONS OF THE SPEED OF LIGHT

In November of 1877 Albert Abraham Michelson (1852–1931), then a 24-year-old ensign in the U.S. Navy and an instructor in physics at the U.S. Naval Academy in Annapolis, Maryland, hit upon the means to improve Foucault's rotating mirror approach. Even then, he needed to conduct many preliminary studies before being confident of an improved value for the speed of light. In his own words ([37], page 115) "Between this time and March of the following year a number of preliminary experiments were performed in order to familiarize myself with the optical arrangements. Thus far the only apparatus used was such as could be adapted from the apparatus in the laboratory of the Naval Academy."

In April 1878, he initiated contact with Professor Simon Newcomb (1835–1909) of the U.S. Navy ([49], page 38), who was then superintendent of the navy's *Nautical Almanac* and renown in the navy and the scientific community as an astronomer. Michelson discussed his work and methods with Newcomb. At this point, however, Michelson was still an unknown who would not be funded by the U.S. Navy for such specialized research. Fortunately, having married Margaret McLean Heminway in the spring of 1877, he could turn to a wealthy father-in-law for financial support. His father-in-law (referred to in [37] only as a "private gentleman") had become deeply interested in Michelson's preliminary results and in July of 1878 provided him the $2000 necessary to purchase the fine optical instruments to carry out his measurements. So began a lifelong quest to determine the speed of light.

3.1 Optical Theory

One of the difficulties with having great distances between the source and the mirror in Fizeau's scheme is that the intensity of the light will decrease with distance. The image is brightened by placing a lens between the source and the mirror. If, as in Figure 1, the source S and the mirror, M are placed so that a point-source light from one is focussed precisely on the other, then the return image will be as bright and as crisp as possible.

Note that the distance between L and M is not equal to that between L and S. As M moves farther from the lens, S will need to be moved closer for both points to remain at the focus of the other's point source. This is true provided both points are beyond the focal length of the lens (that point where beams of light parallel on one side of the lens would meet on the other side).

By moving S and M farther apart, all the while keeping each at the other's point focus, we increase the distance the light must travel and therefore the time it will take. Even so, the time taken is exceedingly short and difficult to measure.

Instead of Fizeau's wheel, Foucault used a rotating mirror interposed between S and L as in Figure 2. [According to Newcomb ([42], page 117) this had been suggested much earlier by Charles Wheatstone (1802–1875) and tried without success by Dominique Francois Jean Arago (1786–1853) in 1838.] Light rays from the source that strike R and proceed through the lens L will strike M and return to the source S. If, after the light beam first strikes R outbound from S, R can be rotated before it is struck again by the beam returning from M, then the returning beam will no longer return exactly

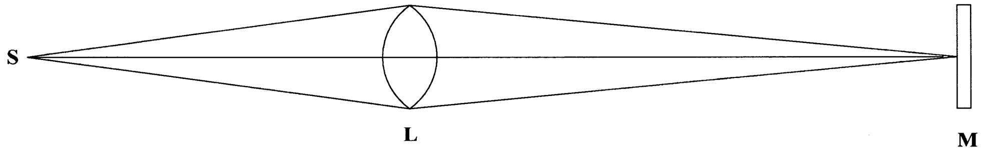

FIG. 1. *S and M are placed at the point-source focus of each other.*

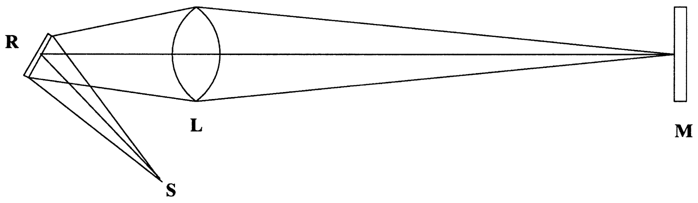

FIG. 2. *Interposing a mirror R between the source S and the lens L.*

to the source S but will instead be deflected away from S in the direction of the rotation (Figure 3).

By rotating the mirror at a constant speed, the amount of deflection will be the same for all light beams which go through L, strike M and return. Then, for a continuous beam of light from S and a constant high speed of rotation of R, an image of the source will appear beside S instead of coincident upon it (as shown in Figure 4). The faster R rotates or the longer is $|RS|$, the farther the returned image I will be displaced from the source S and the easier it will be to measure the deflection.

By carefully measuring the amount of displacement from S to I (see Figure 4), and the distance from S to R, the angle of deflection can be determined. Together with the known, fixed speed of rotation, this angle can be used to determine the time it took light to travel the distance from R to M and back. Dividing distance by time gives a determination of the speed of light.

Let θ denote the angle of deflection. Then the angle through which the mirror has rotated is easily shown to be $\theta/2$. The angle θ in degrees is $\arctan(|IS|/|IR|)$. If the speed of rotation is n measured in cycles per second, then the time taken for the light beam to travel from R to M and back is

$$\frac{1}{n} \times \frac{\theta/2}{360} \quad \text{seconds.}$$

The speed of light transmitted under the conditions of the study is therefore

$$2\frac{360n}{\arctan(|IS|/|SR|)} \times 2|RM|.$$

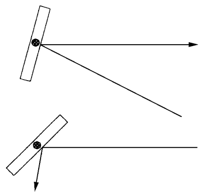

FIG. 3. *Rotating the mirror R causes the returning beam to be deflected.*

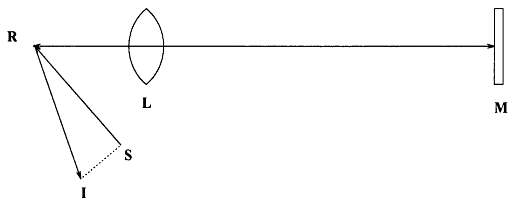

FIG. 4. *The return image I is displaced from the source S by the rotating mirror R.*

In this arrangement, the distances $|IS|$ and $|SR|$ should be as large as possible to reduce the error in measuring θ. The distance $|IS|$ is maximized by maximizing the speed of rotation of R and the distance $|RM|$. Michelson's principal innovation in Foucault's design allowed $|RM|$ to be very large. In Foucault's setup, M was spherical with center at R. The greatest distance $|RM|$ achieved by Foucault was 20 m ([37], page 117), which produced a displacement $|IS|$ of only 0.7 mm ([42], page 118). Michelson chose to place the rotating mirror at the focal point of the lens which allowed him to use a flat mirror for M. That is, R should be placed at that point where *parallel* light beams passing through the lens from M meet on the other side as in Figure 5.

Then if the diameter of M was as large as that of L any single beam passing from R through L would *necessarily* strike M *and return* through L to R *whatever the distance between L and M*. This permitted M to be placed very far away. The only difficulty is that the farther away M is from L, the closer the point-source focus S will be to the focal point R, which conflicts with maximizing the distance between S and R. This can be remedied somewhat by using a lens of large focal length.

These innovations produced a displacement of more than 100 mm. Such a large displacement solved another difficulty. Originally the eyepiece to observe the displaced image at S was offset using an inclined plate of silvered glass to avoid interference between the observer and the outgoing beam of light. Once the the displacement exceeded 40 mm, it was possible to remove the inclined plate and observe the displaced image directly. Michelson ([37], page 116) noted "Thus the eye-piece is much simplified and many possible sources of error are removed."

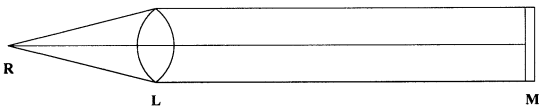

FIG. 5. *R at the focal point of L.*

3.2 Physical Apparatus

The following quotations and details are taken from Michelson's description of his study ([37], pages 118–124).

"The study would take place on a clear, almost level, stretch along the north sea-wall of the Naval Academy. A frame building was erected at the western end of the line, a plan of which is represented in Figure 3 [see our Figure 6, which reproduces Michelson's Figure 3].

"The building was 45 feet long and 14 feet wide, and raised so that the line along which the light travelled was about 11 feet above the ground. A heliostat at H reflected the sun's rays through the slit at S to the revolving mirror R, thence through a hole in the shutter, through the lens, and to the distant mirror."

The heliostat is an instrument used to focus the sun's rays and direct them in a narrow beam. This then was the source of light. Because it is easier than the heliostat to adjust, a small mirror F directs the beam from the heliostat to the slit.

"The lens was mounted in a wooden frame, which was placed on a support moving on a slide, about 16 feet long, placed about 80 feet from the building. ... The fixed mirror was ... about 7 inches in diameter, mounted in a brass frame capable of adjustment in a vertical and horizontal plane by screw motion. ...To facilitate adjustment, a small telescope furnished with cross-hairs was attached to the mirror by a universal joint. The heavy frame was mounted on a brick pier, and the whole surrounded by a wooden case to protect it from the sun."

Unlike Foucault, a flat mirror was used as the fixed mirror and a lens of long focal length focussed the light (an 8-in. non-achromatic lens with a 150-ft focus). The lens was placed in position about 80 ft from the building and the fixed mirror a distance of about 1920 ft from the building. Both the mirror M and the lens L needed to be placed perpendicular to a common central axis as in Figure 2.

Michelson gives no account of how the lens came to be positioned but he does describe the positioning of the mirror in some detail. First it was placed in position with the reflective surface facing the hole in the building.

"A theodolite [a land surveying instrument used to measure angles] was placed at about 100 feet in front of the mirror, and the latter was moved about by the screws till the observer at the theodolite saw the image of his telescope reflected in the center of the mirror. Then the telescope attached to the mirror was pointed (without moving the mirror itself) at a mark on a piece of card-board attached to the theodolite."

In this way the telescope atop the mirror was placed at right angles to its reflective surface.

"The theodolite was then moved to 1,000 feet, and, if found necessary, the adjustment [to the telescope] repeated."

With the telescope thus placed, the mirror was moved until its telescope pointed at the hole in the building. A final adjustment was made by having someone focus a spyglass at the fixed mirror from inside the building. The mirror was then moved using the screws until the observer saw the image of his spyglass reflected centrally in the mirror.

The rotating mirror was a 1.25-in. circular disc (0.2 in. thick) silvered on one side. It was held on a vertical spindle that was in turn held in a cast iron frame. This frame could be tilted side to side and forward and backward by means of small cords. The spindle had pointed ends which pivoted in conical sockets in the frame; these were the only contact points between the frame and the spindle. The top part of the spindle passed through the center of a small wheel inside a circular enclosure attached to the frame. This wheel held the spindle by fric-

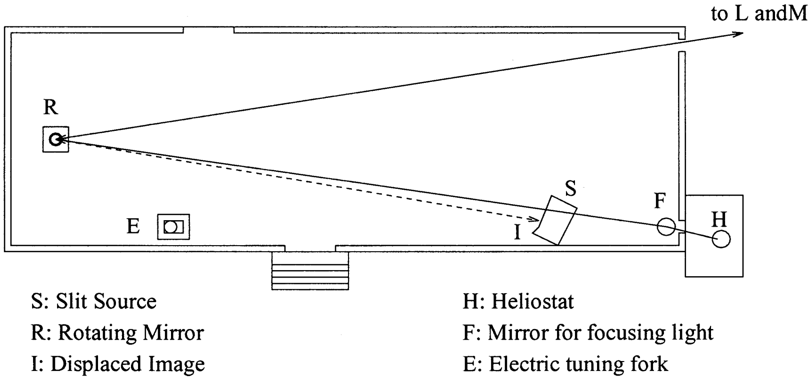

FIG. 6. *Room showing experimental setup.*

tion. Forcing air into the enclosure, over the surface of the wheel, and out again in a circular fashion would cause the wheel, and hence the spindle, to turn. The spindle would have to be carefully balanced so that it turned smoothly without wobbling. The air to power this small turbine came from a steam-powered pump located in the basement of the building. A tube connected the pump to the turbine. Because the mirror's rotational speed remains constant only while the pressure from the pump is constant, a system of regulators, valves and feed back control ([37], Figures 11 and 12, page 124) was installed to adjust the pressure and hence the speed. Michelson notes that the system could hold the speed of rotation constant for three or four seconds, which was sufficient to make a measurement.

To further increase the distance $|SR|$, the rotating mirror was placed slightly closer to the lens than at the focal point of the lens (i.e., its parallel beam focus). This would make for a slightly less clear image than having R at the focus as fewer rays strike and are returned from M.

"A limit is soon reached, however, for the quantity of light received diminishes rapidly as the revolving mirror approaches the lens."

This limit is about 15 ft closer to L than is its focal point. Michelson's previous studies showed that if R rotates at about 258 revolutions per second and the distance $|SR|$, or *radius*, is about 28.6 ft, then the deflection should be around 115 mm. (Names of variates, like "radius," whose values Michelson recorded are italicized here when first mentioned.)

3.3 Measurement Equipment

Michelson made use of several pieces of measurement equipment.

Distances $|SR|$ and $|RM|$ were measured using a steel tape, nominally 100 ft long.

The *displacement* $|IS|$ was measured using of a calibrated micrometer as shown in Figure 7. The source of the light was a narrow vertical slit that was fixed in place on the micrometer. The micrometer had a small telescope that could be moved left to right using a dial at the right. Each turn of the screw would move the telescope some small known amount. In Figure 7, the horizontal scale shown marks the amount turned. At the focus of the telescope lens (about 2 in.), and in nearly the same plane as the slit S, was a single vertical silk fibre that served as a vertical crosshair for alignment purposes. By turning the screw, the telescope could be positioned so that this fibre was centred on the returning image of the slit at I. The amount the telescope had to be moved from its initial position at the slit to the position of the image would be the displacement $|IS|$.

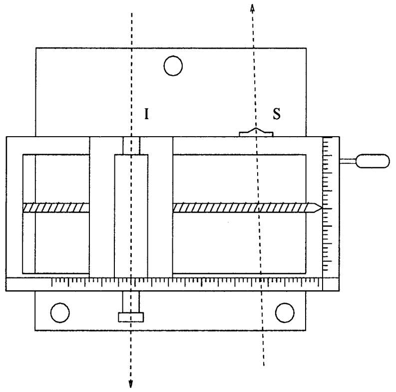

FIG. 7. *Micrometer measures the displacement* $|IS|$.

The speed of rotation n, *number of revolutions per second*, of the revolving mirror was set using an electric tuning fork which vibrated at about 128 cps. The valve from the pump was opened to rotate the mirror R and make its speed in revolutions per second match the frequency of the electric tuning fork in vibrations per second. The speed and frequency were matched by having a small mirror attached to one arm of the tuning fork placed so that some light reflected from the revolving mirror was in turn reflected by the tuning fork's mirror to produce an image of the disk of the revolving mirror on a piece of plane glass located near the lens of the eyepiece of the micrometer. If the tuning fork frequency and the speed of the revolving mirror were the same, then the final image appearing on the glass would be distinct. In most of Michelson's determinations, the frequency of the fork was half that of the revolving mirror, so that two distinct images were produced ([37], Figure 13, page 124).

The frequency of the electric tuning fork, called Vt_2, was measured by counting the *beats per second* between it and a standard tuning fork Vt_3 with known frequency 256.070 cps at 65°F. A 60-s count period was used. The *temperature* was recorded to correct the frequency of the standard fork for temperature. The frequency of the electric fork is thus one-half of the sum of 256.070, the number of beats per second and the correction for temperature.

The final result for the speed of the revolving mirror in revolutions per second is determined from the frequency of the electric tuning fork and the number of distinct images on the glass plate.

3.4 Producing One Determination of the Speed of Light

1. The distance $|RM|$ from the rotating mirror to the fixed mirror was measured five times, each time allowing for temperature, and the average

used as the “true distance” between the mirrors for all determinations.

2. The fire for the pump was started about a half hour before measurement began. After this time, there was sufficient pressure to begin the determinations.
3. The fixed mirror M was adjusted as described above and the heliostat placed and adjusted so that the Sun’s image was directed at the slit.
4. The revolving mirror was adjusted on two different axes. First it was inclined to the right or left so that the direct reflection of the light from the slit fell above or below the eyepiece of the micrometer. Michelson found that he had to tilt the revolving mirror as “Otherwise this light would overpower that which forms the image to be observed” [37]. The revolving mirror was then adjusted by being moved about, and inclined forward and backward, till the light was seen reflected back from the distant mirror” ([37], page 124). Some adjustment in the calculations was made for the tilting of the mirror.
5. The distance $|SR|$ from the revolving mirror to the crosshair of the eyepiece was measured using the steel tape.
6. The vertical crosshair of the eyepiece of the micrometer was centred on the slit and its position recorded in terms of the position of the screw.
7. The electric tuning fork was started. The frequency of the fork was measured two or three times for each set of observations.
8. The temperature was recorded.
9. The revolving mirror was started. The eyepiece was set approximately to capture the displaced image. If the image did not appear in the eyepiece, the mirror was inclined forward or back until it came into sight.
10. The speed of rotation of the mirror was adjusted until the image of the revolving mirror came to rest.
11. The micrometer eyepiece was moved by turning the screw until its vertical crosshair was centred on the return image of the slit. The number of turns of the screw was recorded. The displacement is the difference in the two positions. To express this as the distance $|IS|$ in millimetres the measured number of turns was multiplied by the calibrated number of millimetres. per turn of the screw.
12. Steps 10 and 11 were repeated until 10 measurements of the displacement $|IS|$ were made.
13. The rotating mirror was stopped, the temperature noted and the frequency of the electric fork was determined again.

4. STATISTICAL METHOD AND MICHELSON’S 1879 STUDY

Statistical method can be usefully represented as a series of five stages: *Problem, Plan, Data, Analysis, Conclusion*. We use the acronym PPDAC to refer to this series. Each stage of statistical method comes with its own issues to be understood and addressed (summarized in the table of Figure 8).

One stage leads to the next and is dependent on previous stages. Looking back, this means that each

Problem	- Units & Target Population (Process) - Response Variate(s) - Explanatory Variates - Population Attribute(s) - Problem Aspect(s) – causative, descriptive, predictive
Plan	- Study Population (Process) (Units, Variates, Attributes) - Selecting the response variate(s) - Dealing with explanatory variates - Sampling Protocol - Measuring processes - Data Collection Protocol
Data	- Execute the Plan and record all departures - Data Monitoring - Data Examination for internal consistency - Data storage
Analysis	- Data Summary numerical and graphical - Model construction build, fit, criticize cycle - Formal analysis
Conclusion	- Synthesis plain language, effective presentation graphics - Limitations of study discussion of potential errors

FIG. 8. *The statistical method.*

stage is carried out and legitimized (or not) in the context of the stages which precede it (e.g., there is little value in a Plan that does not address the Problem; in such a case, one of the two stages must be modified). Looking ahead at any stage, choices can be made that will simplify actions taken in a later stage (e.g., a well-designed Plan can simplify the Analysis). Bouncing back and forth between stages is common in the development of the complete PPDAC structure.

A structure for statistical method is useful in two ways: first to provide a template for actively using empirical investigation and, second, to critically review completed studies. The structure of all empirical studies, either implicitly or explicitly, can be represented by the five-stage model.

In this section, we expand on the key concepts and tasks of each stage, introducing new terminology as needed. Michelson's 1879 investigation will be used as illustration. As pointed out in the first section, in many ways this investigation is not typical of a statistical one and we urge the readers to test the proposed structure and language on other applications.

4.1 The Problem

Understanding what is to be learned from an investigation is so important that it is surprising that it is rarely, if ever, treated in any introduction to statistics. In a cursory review, we could find no elementary statistics text that provided a structure to understand the problem. For example, the popular and well-regarded book by Moore and McCabe [40] makes no mention of the role of statistics in problem formulation.

Two notable exceptions are the paper by Hand [28] and the book by Chatfield [11]. Hand's aim was "to stimulate debate about the need to formulate research questions sufficiently precisely that they may be unambiguously and correctly matched with statistical techniques." He suggests five principles to aid in this matching but no structure or language. Chatfield provides excellent advice to get a clear understanding of the physical background to the situation under study, to clarify the objectives and to formulate the problem in statistical terms.

The purpose of the problem stage in statistical method is to provide a clear statement of what is to be learned. A well-defined structure and clear terminology will help translate the contextual problem into a form that can guide the design and implementation of the subsequent stages.

4.1.1 *Units and target population.* The *target population* is the collective of *units* about which we would like to draw conclusions. Care needs to be taken in specifying both.

In 1879, Michelson was keen to determine the speed of white light as it travels between any two relatively stationary points in a vacuum. A unit, then, is one transmission of such light between a source and a destination, both located in a vacuum. The target population is all such transmissions, before, during and after 1879.

For some investigations it may be easier to define the units or the collective in terms of a process which generates them. An example is a manufacturing process producing units under specified conditions. In such cases it might be more convenient to refer to the *target process* rather than the *target population*.

4.1.2 *Variates. Variates* are characteristics of each unit in the population and can take numerical or categorical values. The values of variates typically differ from unit to unit.

The primary variate of interest, which we call the *response variate*, is the speed of the light associated with each such transmission. There are many other variates, which we call *explanatory variates* attached to each unit such as the distance between the two points, the motion of the points with respect to each other, properties of the source and so on. In Michelson's problem, he has no direct interest in these other variates.

4.1.3 *Population attributes.* Population *attributes* are summaries describing characteristics of the population. Formally an attribute is a function applied to the entire population and determined through the variate values on individual units.

The attribute of interest is the average speed of light across all units in the target population. This example is unusual in that it was believed that the speed of white light is constant in a vacuum and so there is no variation in the value of the response variate from unit to unit in this target population.

Attributes can be numerical or graphical. For example, a scatterplot constructed using all units in the target population is an attribute. The coefficients of the least squares line fitted to this scatterplot and the residual variation around the line are numerical attributes.

A clear specification of the attributes of interest can resolve many issues. Lord's paradox, as presented by [28], is easily resolved by noting that it involves two different attributes. See our discussion in Hand [28].

4.1.4 *Problem aspect.* The *aspect* defines the basic nature of the problem and is *causative*, *predictive* or *descriptive*.

A problem with a causative aspect corresponds to one where interest lies in investigating the nature of a causative relationship between an explanatory variate and a response variate. The preceding language allows us to be more precise about what is meant by "causative relationship." By this we mean that a change in the value of the explanatory variate (while holding all other explanatory variates fixed) for all units in the population results in a change in the value of an attribute of interest.

A problem has a predictive aspect if the object is to predict the values of variates on one or more units in the target population. A problem has a descriptive aspect if the object is to estimate or describe one or more attributes of the population.

The problem aspect here is descriptive; the aim is to estimate a population attribute, the average speed of light. Had Michelson been attempting to show that the speed of light can be changed by, for example, having the destination move with respect to the source, then the problem has a causative aspect (as in the famous Michelson and Morley experiment [38]). Michelson's work does not easily lend itself to illustrating a predictive aspect. A more familiar example is forecasting future sales from past information.

It is important to decide the aspect at the problem stage because of the special requirements it can impose on the plan.

4.2 The Plan

The purpose of this stage is develop a plan for the collection and analysis of the data. We propose to break the planning into several substages, some of which inevitably overlap. In an active use of PPDAC, some iteration may be required within the stage and between stages before a satisfactory plan is developed.

4.2.1 *Specifying the study units and study population.* The *study population* is the collective of *study units* for which the values of the variates of interest could possibly be determined. This notion corresponds directly to the frame in sample survey literature. The difference between the attributes of interest in the study population and the corresponding attributes in the target population is called the *study error*. This is a simple quantitative assessment for numerical attributes but can be challenging to define for graphical ones.

The study units may or may not be part of the target population, as is the case in Michelson's study. Because the distances required to measure the speed of light were so large, it was not practical to have the light travel through even a partial vacuum. (Even as he was dying, Michelson directed a study to measure the speed of light in a mile-long tube that was evacuated to a near vacuum [39].) All of the units in Michelson's study involved the transmission of light through air at a particular location over a specified time period. The source and destination were a fixed distance apart and both remained stationary over the course of the study. Michelson decided to look at transmission of light at one hour before sunset or one hour after sunrise during a few days in June 1879. Within these constraints, he was free to choose the units on which he would determine the speed of light.

The study population and the study units were very different from the target in this instance. Michelson recognized that measuring the speed of light in air would result in a study error. He planned to correct the error by using a factor based on the refractive index of air. Note that this correction is outside the purview of statistical method. It requires contextual knowledge.

The statistical method ensures consideration of the relevance of the study population to the target population by forcing investigators to deal directly with the study error. Criteria beyond the study error such as cost, convenience and ethics will also be important in determining the study population.

4.2.2 *Selection of the response variates to be measured.* The Plan must include a step in which we decide what variates we will measure on each unit to be selected in the sample. Response variates, corresponding as much as possible to those used to define attributes of interest in the target population, must be clearly defined.

Michelson could not measure the speed of light on a unit directly with his apparatus. Instead, for each determination, he measured the following response variates to calculate the speed of light:

1. the displacement d of the image in the slit—this was measured on each unit;
2. the radius r, the distance between the crosshairs of the slit and the front face of the rotating mirror—this value was not always determined for units measured in the same time period but was measured each morning or evening when units were sampled;
3. the number of beats B per second between the electric Vt_2 fork and the standard Vt_3—this variate was determined once for each set of 10 determinations of d;

4. the temperature T—measured once for each set of 10 determinations of d.

The values of the response variates were combined with several constants according to the previous formulae ([37], page 133) to produce a value for the speed of light in air at temperature T.

4.2.3 *Dealing with explanatory variates.* It is useful at this point to list all possible explanatory variates which might explain variation in the response and to organize them in some fashion. One useful organization is the fishbone diagram, shown in Figure 9 for Michelson's study.

It is important to decide how explanatory variates will be dealt with during the planning stage. There are three choices. First, an explanatory variate can be held fixed or restricted to a range of values so as to restrict the study population. Second, once a unit is in a sample the value of an explanatory variate could be set deliberately or measured for later use in the analysis. Finally, the explanatory variate can be ignored completely. The third course of action is taken if it is known in advance that the explanatory variate is unimportant (e.g., it does not explain variation in the response variates) or out of ignorance, not recognizing the presence or importance of the variate.

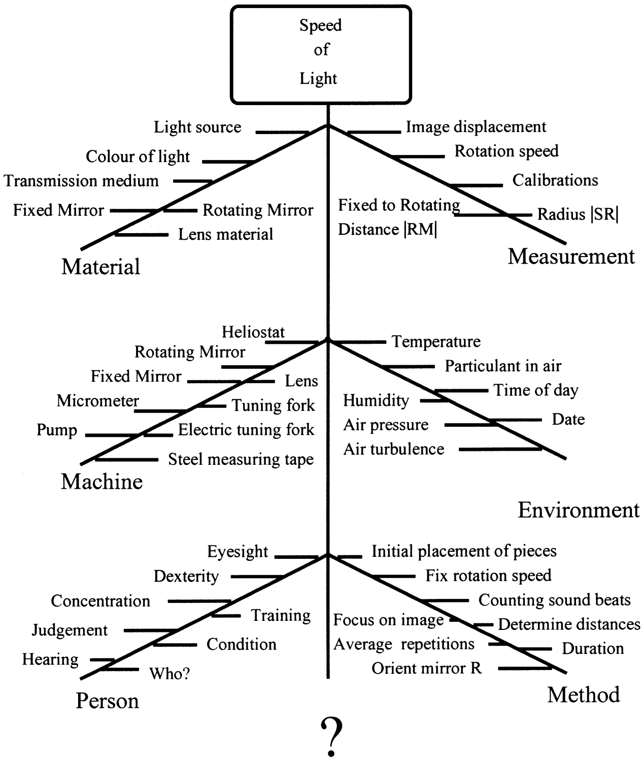

FIG. 9. *Fishbone diagram.*

Reviewing Michelson's apparatus and proposed method, there are many explanatory variates in the study population that may explain why the speed of light as determined from the measured response variates varies from unit to unit. Michelson recognized that it was important to consider these variates and in his Plan dealt with them in all three ways. For example, he fixed the distance from the rotating to the fixed mirror, thus further refining the study population. He also deliberately varied the angle of inclination of the plane of rotation of the revolving mirror from arctan(0.02) in the early determinations to arctan(0.015) in the final 12 sets. He measured a large number of explanatory variates such as the observer, the day, the quality of the image and so on. He ignored barometric pressure because ([37], page 141) "...error due to neglecting barometric height is exceedingly small."

The primary difference between *experimental* and *observational* Plans is highlighted at this stage. In an experimental Plan, values of explanatory variates corresponding to factors of interest are set by the experimenter and assigned to units in the sample. Traditional experimental design provides details on the assignment. In an observational Plan, the explanatory variates are not deliberately manipulated, except perhaps by restricting the study population or the sampling protocol. Their measured values are used in the analysis.

4.2.4 *The measuring processes.* A key element of the Plan is to decide how to measure the selected response and explanatory variates on the units in the sample. To determine the value of any variate on a unit, we call the measuring devices, methods and individuals involved the *measuring process*. Once a measuring process is specified, it is important to understand its properties. We call *measurement error* the difference between the value of the variate determined by the measuring process and the "true" value. Measurement error is propogated through the Analysis and hence to the Conclusion.

In many applications, a separate smaller PPDAC cycle is carried out to investigate the attributes of the measuring process within the overall study. We define the properties of the measuring process in terms of repeatedly measuring the same study unit. Two concepts are *measuring bias*, an attribute of the (target) measuring process describing systematic measurement error, and *measuring variability*, an attribute of the (target) measuring process describing the change in the measurement error from one determination to the next.

Michelson paid careful attention to the measuring processes he had specified for his study and discussed at great length investigations he undertook to ensure that there was little measuring bias and variability. Consider, for example, the measurement of the distance between the two mirrors ([37], page 125). To avoid bias, he calibrated a steel tape against a Wurdeman copy of the standard yard. The calibration used a comparator with two microscopes, one fixed and one that could be moved toward or away from the fixed microscope by turning a screw. The distance between the microscopes was set to 1 standard yard. Then the tape was placed in the comparator so that 0.1 ft corresponded to the crosshairs of the fixed microscope and the length of the first yard of the tape was determined by rotating the screw until the crosshairs of the movable microscope corresponded to 3.1 ft on the tape. This procedure was repeated 33 times to determine the cumulative number of turns of the screw corresponding to the length of the tape from 0.1 to 99.1 ft. The temperature was recorded so that an adjustment (unexplained) could be made.

Next, he carried out a separate study to determine the distance corresponding to 1 turn of the screw of the movable microscope. This was accomplished by measuring 20 times the number of turns that correspond to 1 mm and then averaging. It is clear that Michelson appreciated the power of averaging to reduce variability in measurement. Combining the results of the two studies and adjusting for temperature, the corrected length of the 100-ft steel tape was 100.006 ft.

To measure the distance between the two mirrors (approximately 2,000 ft), the plan was to place lead markers along the ground and use the tape to measure the distance from one to the next following a carefully defined standard procedure. The tape was to be placed along the (nearly) level ground and stretched using a constant weight of 10 lbs. This led Michelson to investigate the stretch of the tape.

To adjust for stretch, another small study was conducted in which the tape was stretched using a 15-lb force and the stretch in millimetres at 20-ft intervals was measured. The data are shown in Table 2.

TABLE 2

Length	Amount of stretch
100	8.0
80	5.0
60	5.0
40	3.5
20	1.5

The correction, in millimetres, for stretch in the tape to measure the distance between the mirrors is then

$$\text{correction} = \frac{8.0 + 5.0 + 5.0 + 3.5 + 1.5}{300} \times 100 \times \frac{10}{15}.$$

Converted to feet and multiplied by 20, the overall correction for stretch was +0.33 ft.

In the language we have introduced, for this small study the study population using a 15-lb force is different from the target population which requires a 10-lb stretching force. Note also the curious weighted average for estimating the amount of stretch per foot of tape.

The goal of introducing the corrections for stretch and length of the tape was to reduce bias in the final measuring of the distance between the two mirrors. To reduce the variability, the procedure was repeated five times (with corrections for temperature on each). The temperature-corrected measurements varied from 1,984.93 to 1,985.17 ft. Michelson used the average of the five determinations and then corrected for stretch and bias in the tape to get his final measure of distance between the two mirrors.

The case study is an excellent example of a careful scientist reducing measurement error from his measuring processes using two different approaches. Based on empirical studies, he reduced bias by calibration and correction, and he reduced variability by averaging. At the conclusion of his paper, Michelson provided a detailed discussion of the effects of possible measuring bias on his estimate of the speed of light. It is alarming to realize how often modern data are produced and analyzed with little consideration for the properties of the measuring process. (And no wonder since so little attention is paid to the measuring process in the teaching of statistics. Consider the advice of Moore and McCabe [40], page 223: “But, by and large, questions of measurement belong to the substantive fields of science, not the methodological field of statistics. We will therefore take for granted that all variables we work with have specific definitions and are satisfactorily measured.” Two useful references are Youden [54] and Wheeler and Lyday [52].)

4.2.5 *The sampling protocol.* The *sampling protocol* is the procedure used to select units from the study population to be measured. The goal of the sampling protocol is to select units that are representative of the study population with respect to the attribute(s) of interest. The sampling protocol deals with how and when the units are selected and how many units are selected.

Michelson decided to sample a number of units one hour after sunrise and one hour before sunset for a number of days between June 5 and July 2.

The units were selected in groups of 10 with from one to six groups taken per time period. Units were selected by Michelson and, on two occasions, by his assistants Lieutenant Nazro and Mr. Clason. In all, 1,000 units were sampled. Over the course of the sampling, other explanatory variates were manipulated (speed of rotation of the mirror, the angle of inclination of the rotating mirror etc.) Michelson recognized the importance of selecting units with different values for these explanatory variates so that he could verify that they did not affect the measured velocity of light. Consider, for example, his discussion of observer bias in the final section of the paper. To deal with this issue, additional sets of measurements were taken by another observer who was blind to Michelson's results. There was no systematic difference in the two sets of values.

We call *sample error* the difference between the attribute of interest in the study population and the corresponding attribute in the sample. As with measuring processes, there may be bias and variability associated with the sampling protocol. These are properties of the protocol and not of any particular sample of units. As with the measuring process, *sampling bias* and *sampling variability* are defined in terms of the properties of the sample error when repeatedly applying the sampling protocol to the study population. These replications are always hypothetical, which means that we can describe sampling bias and variability only through a model of the sampling protocol. We postpone discussion of this model to the Analysis section although in the active use of PPDAC, mathematical models for the potential sampling protocol (and measuring processes) are used to help with issues such as sample size determination.

4.2.6 *The data collection protocol.* The *data collection protocol* is the procedure for executing the above steps of the Plan to collect and record the data. It deals with management and adminstrative issues such as who does what and when. It also includes a plan for monitoring the data as they are collected to ensure quality.

Michelson gives us no indication of how he planned to record and monitor his data. However, the meticulous care he showed elsewhere in the planning of his study suggests that he would have been especially careful to ensure that the data were recorded as measured.

In today's context, among other issues, this step will include consideration of data entry, file structures, analysis software and so on, especially for Plans in which a large amount of data is to be accumulated.

4.3 The Data

The purpose of the Data stage is to execute the Plan and assure the quality of the data in preparation for the analysis.

4.3.1 *Execute the plan.* As far as we can tell, Michelson used all of the measurements on the 1,000 units that he collected. Unfortunately, he did not report all 1,000 data points but instead gave the average value of the displacement d for the 10 determinations in each set. All recorded explanatory variates were treated as constant over the set.

4.3.2 *Data monitoring.* By the end of the plan stage, some sense of clearly aberrant values for the variates would be known. Monitoring the recorded values of the data as they occur is important to assure their quality and to make changes to procedures as needed.

Although Michelson claims to have spent two months working with the apparatus it is curious that his first recorded set of measurements are with electric light at night. He then abandons this method in favour of natural light after observing that "the image was no more distinct at sunset and the [electric] light was not steady" ([37], page 124). This suggests that some monitoring of the data occurred. He describes checking for other sources of error and making changes to his plan as he goes.

Had Michelson access to today's computational resources, it is likely that he would have at least monitored the speed determinations as they came in each day. Figure 10 is a plot of the recorded values for the speed of light in air versus the day of collection. Because so many values were recorded as identical, the plotted values have uniform random noise in the range from −4 to 4 added; this has

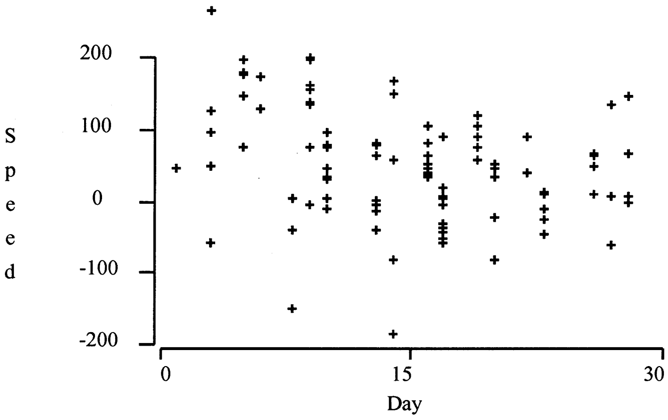

FIG. 10. *Adjusted speed of light (jittered) versus day.*

the desired visual effect of spreading the points out in the plot.

There is an apparent decreasing relationship that is only stronger if the three outlying values are ignored. The noticeable exceptions to this relationship appear to be the values obtained on the last three days. Checking with the data we see that on the third last day Michelson inverted the rotating mirror R. After two days in this position, he inverted it again to get the original position. Arguably, these changes affected the process and prior to that time the study process seemed to be drifting downward. Michelson does not seem to have noticed this.

4.3.3 *Data examination.* Here the *internal consistency* of the data as a whole is assessed, again with the intention of assuring the quality of the data for subsequent analysis. The data is examined for patterns and unexpected features.

With so many variates recorded, there are many possible plots that might be displayed which show interesting patterns in the data. Besides the trend identified in Figure 10, a cursory examination reveals many more. For example, the three-dimensional scatterplot of day, temperature and jittered speed can be rotated into the position shown in Figure 11 revealing three distinct clusters.

Once patterns have been identified, three decisions are possible: ignore them; redo the Plan and Data stages; or most likely pass the information on to be handled in the Analysis stage.

Michelson did not question the internal consistency of his data in the paper.

4.3.4 *Data storage for subsequent analysis.* The values for the measured speed of light in air for each set and the associated response and explanatory variates are given in Tables 3 and 4. Table 5 explains the columns in the tables. Nowadays sometimes much consideration needs to be given to the choice of media and the definition and arrangement of data structures used to store the data.

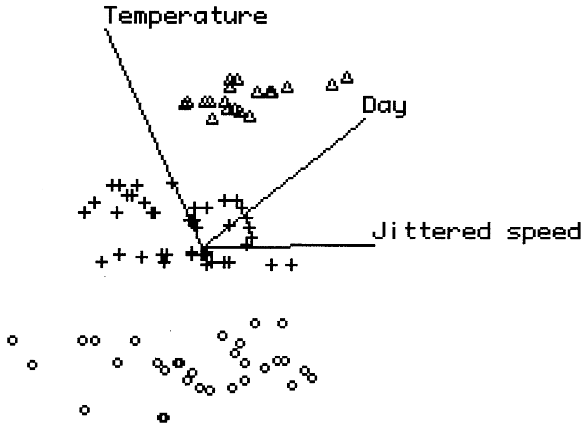

FIG. 11. *Three clusters in three-dimensional space.*

4.4 The Analysis

The purpose of the Analysis stage is to use the collected data and information from the Plan to deal with the questions formulated in the Problem step. The form and formality of the Analysis depends on many things, including the following: the complexity of the Problem and Plan; the skill of the analyst; the amount of variability induced by the Plan; and the intended audience of the study. We propose the following general breakdown of the stage:

- build a model for the Plan and Data;
- fit and assess the model;
- use the final model to address the Problem.

A statistical model describes the behavior of the measured response variates for the units included in the sample if we repeatedly executed the Data step according to the Plan. The model reflects properties of the study population, the sampling protocol and the measurement systems used. The model also includes the influence of measured explanatory variates on the response variate.

Once an initial model is postulated, fitting and model assessment tools can be used to suggest refinements to the model. This iterative process continues until the model is consistent with the internal structure of the collected data and known information about the sampling protocol and measuring systems. The final model is used to estimate attributes of interest in the study population and to assess the uncertainty due to sampling and measuring errors.

Michelson limited his analysis to the calculation of the average of the 100 measured velocities in air, a numerical summary and an estimate of possible error, a formal procedure. The error is based on a worse-case scenario, combining probable errors based on the estimated standard deviations of replicate determinations and maximal systematic error, based on Michelson's knowledge of his apparatus and the functions used to calculate the speed of light from the measured response variates. For more discussion on the use of probable error, see Stigler [48].

After making a small adjustment for temperature (in air) based on the effects of temperature change on the systems used to determine ϕ, the angle of deflection, and correcting to a vacuum, Michelson concludes his analysis by reporting the speed of light in vacuo (kilometres per second) to be

$$299{,}944 \pm 51.$$

TABLE 3

Michelson's data: first 50 observations

Speed	Beat	Cor	Day	Diff	Qual	Disp	Image	Radius	Revs	Screw	Slit	Tday	Temp	Remarks
50	1.423	−0.132	1	0.17	3	114.55	114.85	28.672	257.36	0.99614	0.300	Night	76	1
−60	1.533	−0.084	3	0.10	2	114.56	114.64	28.655	257.52	0.99614	0.074	PM	72	2
100	1.533	−0.084	3	0.08	2	114.50	114.58	28.647	257.52	0.99614	0.074	PM	72	2
270	1.533	−0.084	3	0.12	2	85.84	85.91	28.647	193.14	0.99598	0.074	PM	72	2
130	1.533	−0.084	3	0.07	2	85.89	85.97	28.650	193.14	0.99598	0.074	PM	72	2
50	1.533	−0.084	3	0.07	2	114.53	114.61	28.650	257.42	0.99614	0.074	PM	72	2
150	1.533	−0.216	5	0.07	3	114.47	114.54	28.658	257.39	0.99614	0.074	PM	83	2
180	1.533	−0.216	5	0.10	3	114.46	114.54	28.658	257.39	0.99614	0.074	PM	83	2
180	1.533	−0.216	5	0.08	3	114.47	114.57	28.662	257.39	0.99614	0.074	PM	83	2
80	1.533	−0.216	5	0.06	3	114.50	114.57	28.660	257.39	0.99614	0.074	PM	83	2
200	1.533	−0.216	5	0.13	2	114.53	114.61	28.678	257.39	0.99614	0.074	PM	83	2
180	1.517	−0.300	6	0.11	2	114.52	114.60	28.685	257.29	0.99614	0.074	PM	90	
130	1.517	−0.300	6	0.08	2	114.54	114.62	28.685	257.29	0.99614	0.074	PM	90	
−150	1.450	−0.072	8	0.09	2	114.74	114.81	28.690	257.45	0.99614	0.074	AM	71	
−40	1.450	−0.072	8	0.05	2	114.70	114.78	28.690	257.45	0.99614	0.074	AM	71	
10	1.450	−0.072	8	0.09	1	114.68	114.76	28.690	257.45	0.99614	0.074	AM	71	
200	1.500	−0.084	9	0.09	3	112.56	112.64	28.172	257.49	0.99614	0.074	AM	72	
200	1.500	−0.084	9	0.10	3	112.56	112.63	28.172	257.49	0.99614	0.074	AM	72	
160	1.500	−0.084	9	0.08	2	112.57	112.65	28.172	257.49	0.99614	0.074	AM	72	
160	1.517	−0.168	9	0.06	3	112.56	112.82	28.178	257.42	0.99614	0.260	PM	79	
160	1.517	−0.168	9	0.13	3	112.56	112.82	28.178	257.42	0.99614	0.260	PM	79	
140	1.517	−0.168	9	0.07	3	112.57	112.83	28.178	257.42	0.99614	0.260	PM	79	
160	1.517	−0.168	9	0.06	3	112.56	112.82	28.178	257.42	0.99614	0.260	PM	79	
140	1.517	−0.168	9	0.11	3	112.57	112.83	28.178	257.42	0.99614	0.260	PM	79	
80	1.517	−0.168	9	11	3	113.15	113.41	28.152	258.70	0.99614	0.260	PM	79	3
0	1.517	−0.168	9	6	3	111.88	112.14	28.152	255.69	0.99614	0.260	?	79	4
50	1.500	0.012	10	0.12	1	112.57	112.83	28.152	257.58	0.99614	0.260	AM	64	
80	1.517	0.012	10	0.05	1	112.57	112.83	28.152	257.60	0.99614	0.260	AM	64	
100	1.517	0.000	10	0.11	1	112.55	112.81	28.152	257.59	0.99614	0.260	AM	65	
40	1.517	−0.012	10	0.09	1	112.57	112.83	28.152	257.57	0.99614	0.260	AM	66	
30	1.517	−0.024	10	0.12	1	112.57	112.83	28.152	257.56	0.99614	0.260	AM	67	
−10	1.517	−0.228	10	0.06	1	112.52	112.78	28.159	257.36	0.99614	0.260	PM	84	5
10	1.500	−0.240	10	0.08	1	112.50	112.76	28.159	257.33	0.99614	0.260	PM	85	5
80	1.483	−0.228	10	0.08	1	112.46	112.72	28.159	257.32	0.99614	0.260	PM	84	5
80	1.483	−0.228	10	0.09	1	112.47	112.73	28.159	257.32	0.99614	0.260	PM	84	
30	1.483	−0.228	10	0.09	1	112.49	112.75	28.159	257.32	0.99614	0.260	PM	84	
0	1.517	0.036	13	0.09	2	112.59	112.85	28.149	257.62	0.99614	0.260	AM	62	
−10	1.500	0.024	13	0.06	2	112.58	112.84	28.149	257.59	0.99614	0.260	AM	63	
−40	1.500	0.012	13	0.07	1	112.59	112.85	28.149	257.58	0.99614	0.260	AM	64	
0	1.500	−0.144	13	0.07	3	112.54	112.80	28.157	257.43	0.99614	0.260	PM	77	6
80	1.500	−0.144	13	0.08	3	112.51	112.77	28.157	257.43	0.99614	0.260	PM	77	6
80	1.500	−0.144	13	0.11	3	112.51	112.77	28.157	257.43	0.99614	0.260	PM	77	6
80	1.500	−0.144	13	0.09	3	112.51	112.77	28.157	257.43	0.99614	0.260	PM	77	6
60	1.500	−0.144	13	0.08	3	112.52	112.78	28.157	257.43	0.99614	0.260	PM	77	6
−80	1.500	0.084	14	0.07	1	112.64	112.90	28.150	257.65	0.99614	0.265	AM	58	
−80	1.500	0.084	14	0.10	1	112.64	112.90	28.150	257.65	0.99614	0.265	AM	58	
−180	1.483	0.072	14	0.07	1	112.66	112.92	28.150	257.62	0.99614	0.265	AM	59	
60	1.483	−0.120	14	0.09	2	112.52	112.79	28.158	257.43	0.99614	0.265	PM	75	
170	1.483	−0.120	14	0.10	2	112.48	112.75	28.158	257.43	0.99614	0.265	PM	75	
150	1.483	−0.120	14	0.08	2	112.49	112.76	28.158	257.43	0.99614	0.265	PM	75	

TABLE 4
Michelson's data: *last* 50 *observations*

Speed	Beat	Cor	Day	Diff	Qual	Disp	Image	Radius	Revs	Screw	Slit	Tday	Temp	Remarks
80	1.517	0.063	16	0.07	3	112.67	112.94	28.172	257.65	0.99614	0.265	AM	60	
110	1.517	0.048	16	0.09	3	112.65	112.92	28.172	257.63	0.99614	0.265	AM	61	
50	1.517	0.036	16	0.07	2	112.67	112.94	28.172	257.62	0.99614	0.265	AM	62	
70	1.517	0.024	16	0.03	2	112.66	112.93	28.172	257.61	0.99614	0.265	AM	63	
40	1.450	−0.156	16	0.13	2	133.21	133.48	33.345	257.36	0.99627	0.265	PM	78	
40	1.500	−0.168	16	0.09	2	133.23	133.49	33.345	257.40	0.99627	0.265	PM	79	
50	1.500	−0.180	16	0.07	2	133.22	133.49	33.345	257.39	0.99627	0.265	PM	80	
40	1.483	−0.168	16	0.13	2	133.24	133.50	33.345	257.39	0.99627	0.265	PM	79	
40	1.483	−0.168	16	0.06	2	133.22	133.49	33.345	257.38	0.99627	0.265	PM	79	
40	1.483	−0.168	16	0.10	2	133.22	133.49	33.345	257.38	0.99627	0.265	PM	79	
90	1.533	0.048	17	0.12	2	133.29	133.56	33.332	257.65	0.99627	0.265	AM	61	
10	1.533	0.036	17	0.08	2	133.31	133.58	33.332	257.64	0.99627	0.265	AM	62	
10	1.533	0.024	17	0.09	2	133.31	133.57	33.332	257.63	0.99627	0.265	AM	63	
20	1.533	0.012	17	0.11	2	133.30	133.57	33.332	257.61	0.99627	0.265	AM	64	
0	1.533	0.000	17	0.13	2	133.30	133.56	33.332	257.60	0.99627	0.265	AM	65	
−30	1.533	−0.180	17	0.06	3	133.21	133.48	33.330	257.42	0.99627	0.265	PM	80	
−40	1.500	−0.192	17	0.10	3	133.19	133.46	33.330	257.38	0.99627	0.265	PM	81	
−60	1.500	−0.204	17	0.05	3	133.20	133.46	33.330	257.37	0.99627	0.265	PM	82	
−50	1.517	−0.204	17	0.08	3	133.20	133.46	33.330	257.38	0.99627	0.265	PM	82	
−40	1.500	−0.192	17	0.08	3	133.19	133.46	33.330	257.38	0.99627	0.265	PM	81	
110	1.542	−0.288	19	0.08	3	133.16	133.43	33.345	257.32	0.99627	0.265	PM	89	
120	1.550	−0.288	19	0.06	3	133.15	133.42	33.345	257.33	0.99627	0.265	PM	89	
90	1.550	−0.300	19	0.09	3	133.17	133.43	33.345	257.32	0.99627	0.265	PM	90	
60	1.533	−0.300	19	0.07	3	133.16	133.43	33.345	257.30	0.99627	0.265	PM	90	
80	1.517	−0.300	19	0.07	3	133.16	133.42	33.345	257.29	0.99627	0.265	PM	90	
−80	1.517	−0.084	20	0.15	3	133.20	133.47	33.319	257.50	0.99627	0.265	AM	72	
40	1.517	−0.096	20	0.04	3	133.17	133.44	33.319	257.49	0.99627	0.265	AM	73	
50	1.517	−0.108	20	0.11	3	133.16	133.42	33.319	257.48	0.99627	0.265	AM	74	
50	1.517	−0.120	20	0.06	3	133.16	133.42	33.319	257.47	0.99627	0.265	AM	75	
−20	1.517	−0.132	20	0.10	3	133.18	133.44	33.319	257.45	0.99627	0.265	AM	76	
90	1.508	−0.252	22	0.05	2	133.15	133.42	33.339	257.33	0.99627	0.265	PM	86	
40	1.508	−0.252	22	0.08	2	133.17	133.44	33.339	257.33	0.99627	0.265	PM	86	
−20	1.483	−0.096	23	0.11	3	133.22	133.49	33.328	257.46	0.99627	0.265	AM	73	
10	1.483	−0.108	23	0.06	3	133.20	133.47	33.328	257.44	0.99627	0.265	AM	74	
−40	1.483	−0.120	23	0.09	3	133.21	133.47	33.328	257.43	0.99627	0.265	AM	75	
10	1.467	−0.120	23	0.09	3	133.19	133.45	33.328	257.42	0.99627	0.265	AM	75	
−10	1.483	−0.132	23	0.08	3	133.20	133.47	33.328	257.42	0.99627	0.265	AM	76	
10	1.483	−0.132	23	0.10	3	133.19	133.45	33.328	257.42	0.99627	0.265	AM	76	
10	1.500	−0.240	26	0.05	2	99.68	35.32	33.274	193.00	0.99645	135.000	PM	85	7
50	1.508	−0.252	26	0.06	2	99.67	35.34	33.274	193.00	0.99645	135.000	PM	86	7
70	1.508	−0.252	26	0.10	2	99.66	35.34	33.274	193.00	0.99645	135.000	PM	86	7
70	1.517	−0.252	26	0.09	2	99.66	35.34	33.274	193.00	0.99645	135.000	PM	86	7
10	1.500	−0.216	27	0.07	2	132.98	2.17	33.282	257.35	0.99627	135.145	PM	83	7
−60	1.500	−0.228	27	0.09	2	133.00	2.15	33.282	257.34	0.99627	135.145	PM	84	7
10	1.467	−0.252	27	0.06	2	133.01	2.14	33.311	257.28	0.99627	135.145	PM	86	7
140	1.467	−0.252	27	0.08	2	133.00	2.14	33.311	257.28	0.99627	135.145	PM	86	7
150	1.450	−0.252	28	0.05	3	99.45	99.85	33.205	192.95	0.99606	0.400	PM	86	8
0	1.450	−0.252	28	0.03	3	66.34	66.74	33.205	128.63	0.99586	0.400	PM	86	8
10	1.467	−0.252	28	0.07	3	47.96	50.16	33.205	96.48	0.99580	0.400	PM	86	8
70	1.450	−0.240	28	0.06	3	33.17	33.57	33.205	64.32	0.99574	0.400	PM	85	8

TABLE 5
Michelson's data: key to variates

Speed	Encoded speed of light in air; add 299,800 to get kilometre-per-second scale
Beat	Number of beats per second between forks
Cor	Correction for temperature to standard fork in beats per second
Day	Day of experiment in progress (June 5 = 1)
Diff	Difference between greatest and least values of revolutions
Qual	Quality of the image I; the more distinct it is, the higher the quality (3 = good, 1 = poor)
Disp	Displacement of image 1 from slit *S* in micrometer divisions
Image	Micrometer position of deflected image
Radius	radius of measurement, in feet
Revs	number of revolutions per second
Screw	value of one turn of the screw in millimetres
Slit	micrometer position of slit
Tday	time of day observation recorded (AM = 1 hour after sunrise, PM = 1 hour before sunset)
Temp	Air temperature in degrees Fahrenheit
Remarks	1. Electric light 2. Frame inclined at various angles 3. Set micrometer and counted oscillations 4. Oscillations of image of revolving mirror, probably PM 5. Readings taken by Lieut. Nazro 6. Readings taken by Mr. Clason 7. Mirror inverted 8. Mirror erect

Although Michelson did not formally propose a model, he carried out numerous checks that are equivalent to aspects of model assessment ([37], page 139). For example, to see if the measured speed of light was systematically influenced by the distinctness of the image, an explanatory variate, he calculated and compared the average velocities stratified by distinctness of image. This checking was repeated for many other explanatory variates.

Today, we can use corresponding graphical methods. Perhaps the speed depends on some of the explanatory variates that are not part of its calculation. For example, has the effect of temperature been successfully removed from the determinations? A plot of speed versus temperature is shown in Figure 12. A fairly weak increasing trend is discernible in the plot. However, even this trend depends heavily on the three points in the lower left corner and so is not likely to alter the result significantly. Again the values have been jittered to resolve the overplotting of identical values.

Curiously, in his comparisons of group averages, Michelson did not compare morning and evening measurements nor attempt to relate the measurement to the date, as we explored in the Data stage. There are other interesting relationships to be found in this data; we leave further exploration to the reader.

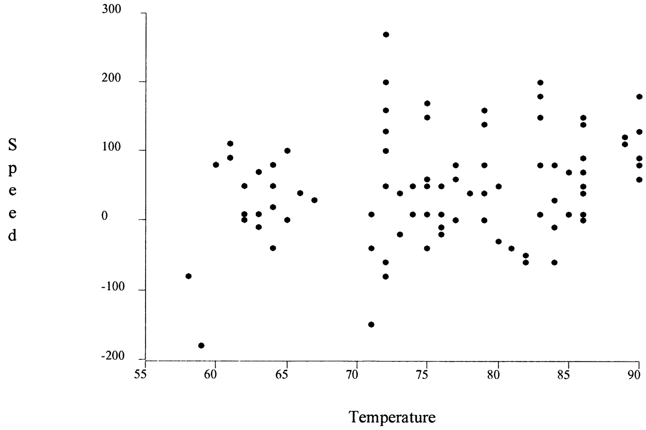

FIG. 12. *Adjusted speed of light (jittered) versus temperature.*

Note that there is often not a clear distinction between the checks for internal consistency in the Data stage and these model checks in the Analysis stage. The same plots or summaries may appear in either.

Today, we can contemplate any number of ways to summarize, model and analyze the data. For example, we might construct a histogram and calculate a 5-number summary of the 100 reported values. Based on a Gaussian model, which appears to fit the data well, a 95% confidence interval for the mean is

$$299{,}852.3 \pm 15.7.$$

Correcting for temperature, following Michelson, and converting to a vacuum, a 95% confidence interval for the speed of light (kilometres per second) in vacuo is

$$299{,}944.3 \pm 15.7.$$

Note that the confidence interval is much shorter than that reported by Michelson, who included both variability and possible bias in his calculation. Other more complex modelling, analyses and model assessment can be made. The above is used to demonstrate the substages within the Analysis stage of PPDAC. Again it is evidence of Michelson's precision as a scientist that his analysis so carefully parallels what can be done today.

Another output of this stage are interesting observations (such as the possible outliers mentioned in discussing Figures 10 and 12) that may well direct future investigations.

4.5 The Conclusion

The purpose of the Conclusion stage is to report the results of the study in the language of the Problem. Concise numerical summaries and presentation graphics should be used to clarify the discussion. Statistical jargon should be avoided. As well, the Conclusion provides an opportunity to discuss the strengths and weaknesses of the Plan, Data and Analysis especially in regards to possible errors that may have arisen. The error classification that we have developed provides a structure for this discussion.

In Michelson's study, he concludes by reporting the speed of light (kilometres per second) in vacuo as $299{,}944 \pm 51$. He then discusses possible "Objections" including, among others not mentioned above, uncertainty of the laws of reflection and refraction in media in rapid rotation, retardation caused by reflection, imperfections in the lens, periodic variation in friction at the pivots of the rotating mirror and change of speed of rotation. In each case, he refers back to the Plan and the model assessment to demonstrate that the objection would have little effect on the estimate of the speed of light.

In our language, we would start with the reported speed of light based on the confidence interval. Other than the discussion given by Michelson, we would add the possible error due to the difference between the target and study population.

We can find no reason in the paper as to why there is such a relatively large error in Michelson's final reported speed. Note that the defined true value, 299,792.458 km/s, is well outside both the confidence interval and Michelson's interval of plausible values.

4.6 Discussion

Too often, statistics has been presented solely as a set of analysis tools. But as the above structure makes explicit, the analysis is but the fourth stage in a series of five which constitute the statistical method. The three stages which precede the analysis are critical to the enterprise—the entire structure forces the proper balance. Seen as a whole, statistical method is not only ubiquitous in empirical investigations but unavoidable.

Nowhere is the need for this balance more apparent than in the teaching of statistics. Over the past seven years we have taught a variety of courses at different levels using the PPDAC structure at the core of the course. Besides giving balance to method, we have found that the structure compels discussion of substantive problems which can be drawn from a wide variety of application areas—industrial, scientific, technological, social and commercial. The statistical method can be taught at almost any level of mathematical sophistication. Substantive and interesting problems can be addressed without resort to complex analysis tools, large data sets or even significant computational resources. What is required is a rich context for each example to describe the details within the structure; these examples tend to grow into case studies.

In our introductory courses we have found over time that the complexity of analysis methods has been reduced as more and more time is devoted to the stages other than Analysis. On final examinations, for example, only about one-third of the marks are assigned to questions directly related to the Analysis stage. The major goals of our introductory course are first to understand the universal need for empirical methods and second to understand and be able to use the statistical method in a variety of contexts.

The structure and language introduced can also be used to clarify some statistical issues which have provoked controversy in the past. Here we give three examples.

EXAMPLE 1. Deming [14] characterized studies as enumerative and analytic. Hahn and Meeker [27] described the concepts in detail. Deming was particularly interested in contrasting the use of formal statistical procedures in sample surveys to their use in studies of industrial processes which include units not yet produced. (Here is an instance where it is more natural to describe the process that generates the units rather than the collection of units of interest and so target *process* is preferred to target *population*). Deming claimed that standard statistical inference procedures (e.g., confidence intervals) would not apply in analytic studies.

In our language, a study is enumerative if the *target* population can be listed so that a probabilistic sampling protocol giving every unit a positive inclusion probability can be used. Otherwise it is analytic. Deming's concern is essentially the possibility of study error which is not captured by the uncertainty expressed by the formal statistical procedures.

EXAMPLE 2. Tukey [51] characterized analyses as either exploratory or confirmatory. Confirmatory analysis is the assessment of prespecified questions and is the traditional domain of inferential statistics. Tukey describes exploratory data analysis (EDA) more as an attitude and not as a bundle of techniques. According to Tukey, the five-stage

PPDAC method is well suited to confirmatory analysis but not to exploratory analysis (nor to science at large). (Tukey [51] names the stages as Question, Design, Collection, Analysis, Answer.) However, by fleshing out the stages as we have above, we can see where exploratory analysis fits in.

The attitude and tools of EDA are clearly important to meet the goals of the monitoring and examination tasks of the Data stage. These tasks amount to carrying out a small PPDAC investigation where the sample of the larger study is now regarded as identical to a target population within this smaller PPDAC. The Problem is to examine many attributes (typically graphical) looking for unexpected values of these attributes.

Alternatively EDA applies to those investigations where the sample is the entire study population. For example, when presented with a massive dataset the investigator is often interested in examining the attributes of that dataset as if it constitutes the entire population. In these instances the target population is still something different from the study population (however large that might be) and so the difficulty of study error remains, even for data miners.

EXAMPLE 3. Statistics is sometimes criticized as applying only to a single study whereas scientific progress demands replication. The statistical method described above would seem to reinforce that view. However, multiple studies can and should be examined within the PPDAC framework. There the difficulties inherent in "meta-analysis" are clarified. For example, one major issue is the inclusion or exclusion of studies from the analysis. One feature of this issue can be discussed by comparing the study population to the target for each investigation considered for inclusion. Alternatively the set of possible studies can be taken as the target population and the set of realized study taken as the study population. Then the sampling protocol determines which studies are included.

5. ON METHOD IN SCIENCE

When examining the writings of those who have thought long and hard about the nature of science one finds the same difficulties appearing again and again. (John Losee's book [35] provides a reasonable starting point.) There is, for the most part, a great enthusiasm that science is progressing in some sense, that we are learning ever more about the world around us, that we are continually solidifying that knowledge, that our increasingly sophisticated technology is testament to the power of science. Yet, when pressed, not only can we not agree on the method of science, we cannot quite agree on what science is, or even whether what it talks about is real! Looking over the history described in this paper we can get some inkling as to why this state of affairs persists.

The progress seems real enough, from the question of light's speed being meaningless, to discussion of whether it is finite or not, to increasing evidence for finite speed, to ever "better" estimates of its value. It might seem that scientific knowledge is the conjunction of the facts accumulated so far, that theories live or die according to their verification or falsification by these facts and that, eventually, the truth will be inferred from the collection of facts.

Kuhn's work [32] describes a framework for this progress: within a scientific "paradigm" normal science is pursued as a puzzle-solving activity, this eventually produces anomalies, anomalies accumulate until a crisis is reached, a new paradigm is somehow introduced, normal science proceeds again and so on. For example, normal science was pursued within a paradigm where light was without speed, astronomical anomalies began to appear, leading ultimately to a theory where light had a finite speed, whereupon normal science set about solving problems to establish its value. In a more elaborate history, many such Kuhnian cycles would have been detectable.

But what about method? Long ago Aristotle wrote that knowledge, being "a state of capacity to demonstrate," required the teaching of the principles of demonstration and so the teaching of science necessarily "...proceeds sometimes through induction and sometimes by deduction"([2], 1139^b19–1139^b36). But each is tricky to apply—Francis Bacon, that strongest of proponents of inductive method, allowed his perception of the incredible speed at which stars move in their orbit about the Earth to form his inductive base and so concluded that an infinite speed of light was reasonable; no lesser talents than Aristotle and Descartes by pure deduction demonstrated that light could not possibly have finite speed. Using induction and deduction in combination as in the hypothetico-deductive approach is no easier. It appears explicitly only twice in the above history—used once by Aristotle to dismiss the argument of Empedocles, and once by Descartes to dismiss that of Beeckman—and wrong in both cases! At various times each of these has been suggested as *the* method of science.

A slightly different tack is to take one such method and raise it to the status of a criterion to distinguish science from non-science. Karl Popper did this in 1934 with the hypothetico-deductive approach. Contemptuous of the widely held view that the use of inductive methods distinguished science from non-science, Popper proposed instead that "it must be possible for an empirical scientific system to be refuted by experience" ([44], page 41). That is, to merit the name scientific a theory must be falsifiable; a decisive experiment which refutes the theory is a crucial falsifying experiment. (In a paper meant to be a general resource [25], I. J. Good gives partial prior credit to R. A. Fisher since tests of significance [20] predate Popper. This credit seems misplaced: Popper uses falsifiability as a *demarcation criterion* for science; Fisher does nothing of the sort.) By this criterion, the geocentric theory of the universe is scientific, being falsifiable by any orbital system not centred about the Earth; Galileo's discovery of the moons of Jupiter refuted this theory. Similarly the scientific theories of light held by Aristotle and Descartes were refuted by Römer's determination of the speed of light. This criterion is turned into method by having scientists focus on trying to refute theory; theories are corroborated only by surviving the most stringent of testing.

But normal science is conservative. Crucial experiments are typically only recognized as such long after the fact—Cassini et al. showed at the time that Römer's observations could be accommodated by existing theory. (See [33], pages 71–90, for further examples and discussion.) If theories were thrown out when first refuted, the result would be chaos. Instead normal science motors along, sometimes fine-tuning its theory to accommodate the new information, sometimes patching the theory with auxiliary hypotheses and sometimes just tossing the information into the back seat where Popper's refutations become Kuhn's anomalies. As the anomalies accumulate, the ride gets rougher and some members of the scientific community become increasingly uneasy that a crisis is around the corner.

It is here that Kuhn's work is most interesting and most troublesome. Kuhn likens the transition from one paradigm to the next to that of a gestalt shift in visual perception. Like a gestalt shift, a paradigm shift is sudden and without reason. Unlike a gestalt shift, a paradigm shift does not allow the scientist to switch between paradigms; no neutral third viewpoint exists from which both paradigms can be seen—if there were, then this would be the new paradigm. This is not to say that the new paradigm cannot be reasoned about and justified to some satisfaction, but rather that it may not be possible to do so by comparing it to the old. For once the transition is complete, the convert's view of the field will have changed—its methods, its concepts, its questions, even its data—and the old paradigm can only be viewed from the perspective of the new. In a word, the two paradigms are incommensurate. Concepts, theory, methods and data that are meaningful according to one might not be according to the other.

Consider the concept of light. According to Aristotle, light required an intervening transparent substance (like air or water); it could not exist in a vacuum. Things are transparent, of course, only because they contain a "certain substance" which is "also found in the eternal upper body" (possibly aether? itself a concept Aristotle tells us he has changed from that of Anaxagoras [3], 270^b20–270^b25). "Of this substance, light is the activity." But it is not movement. Moreover, the visibility in the dark of bioluminescent plants and animals does *not* depend upon light! (See [4], 418^a26–419^b2, for most of the points made here.) From this Aristotle says he has explained light. Not only is Aristotle's concept different from ours, but to really understand what he means by light we would need to become immersed in his paradigm. Scientific concepts like light change in irreversible ways; some like aether disappear altogether—even after thousands of years of service.

Nor are concepts alone determined by the paradigm. So too are the "empirical facts"–Francis Bacon's data included fantastic speeds for the movement of the stars about the Earth; Glasek-napp demonstrated that different theory produced different "observed" speeds of light. Even relatively raw "sense data" can be dependent upon theory. Soon after Galileo announced the discovery of Jupiter's moons, he had others verify his observations using his telescopes. Many could not see the satellites; those who could see multiple lighted spots could not be certain that these were not artefacts of the new instrument. Only once the optics of telescopes was developed could there be confidence in the verity of the observations. (See [19], Chapter 9.) Modern instruments produce observations that are irrevocably "theory laden."

Paradigm shifts, incommensurability and theory-laden data have all contributed to what Ian Hacking [26] calls "a crisis in rationality"—at least for philosophers of science. Is there such a thing as scientific reasoning? Are the entities with which science deals real or are they human constructs? Does it make sense to think that there is in fact an ideal truth to which science might converge?

6. AND WHAT OF STATISTICS?

When statisticians look at the nature of science, they see reflected the nature of statistics.

(A notable exception is Pearson's *The Grammar of Science* [43].) Deduction becomes probability theory; induction, statistical theory (e.g., [7], pages 6, 7); scientific method is hypothetico-deductive (e.g., [10], [17], [41]), self-evident in statistics through formal hypothesis testing and model criticism; put it together and you have, reminiscent of Aristotle, what George Box has called "the advancement of learning" [10]. But, as the previous section has shown, science is not really like that. Neither should be our understanding of statistics. (Indeed, John Tukey's long battle for the legitimacy of exploratory data analysis might have been easier if there had been greater sympathy in the statistical research community for separate contexts for discovery and for justification in science. For example, see [51].)

Certainly statistical investigation meets with the same issues raised in the previous section but it can deal with them more easily. This is because it has a considerably more focussed domain of application. For example, consider the two old chestnuts of the philosophy of science—the realist–antirealist debate and the problem of induction.

The realist–antirealist debate concerns whether the entities of science are real or mere theoretical constructs. The primary entities of statistical investigation are the units of the *study* population and the values of variates measured on them. The units and their collective must be determined with sufficient care for it to be possible to select any individual from the collective. Sometimes considerable effort must be put into ensuring that measuring systems return reliable values of the variates they purport to measure. Within this context, statisticians become scientific realists in Hacking's sense: if we can select them and take measurements on them, they are real [26]; if we cannot, then statistical investigation ceases. Whether future scientific study shows the units to be composites of other more "fundamental" units or that the variates measured are to be interpreted differently is beside the point.

As regards induction, for statistics the problem can be neatly separated into two pieces (see Figure 13). Ultimately, interests lies in the *target* population, as it is nearest to the broad scientific concerns of the problem. This population may be infinite, possibly uncountably so, and its definition can involve phrases like "all units now and *in the future*." Drawing conclusions about this population will often require arguments that are extra-statistical for they will be based on the similarities of, and differences between, the *target* population and the *study* population. Such arguments may ultimately be unable to avoid assuming Hume's "uniformity of nature" principle ([29], page 89) and hence what philosophers mean by the "problem of induction."

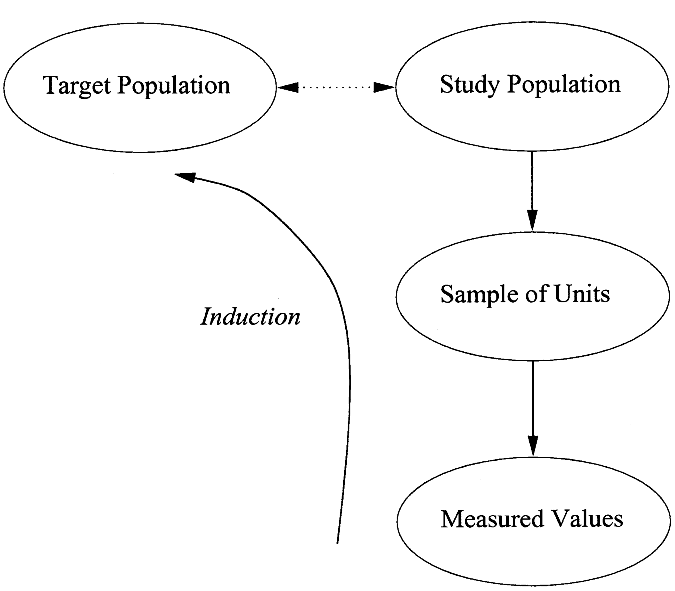

FIG. 13. *Induction from the set of measured values to the target population.*

Such weighty problems dissipate when focus shifts to drawing conclusions about the *study* population. Such is its definition that all study populations are finite in size and random selection of units to form a sample is possible. Random selection provides the strongest grounds for inductive inference. When, for whatever reason, random selection has not been employed then either the case that it has been near enough approximated, or that the sample is itself similar in its attributes of interest to the study (or target) population must be made. The latter is much like making the case for the transfer of conclusions from the *study* to the *target* population and so can be just as difficult. In either case, the arguments will to a large extent be extra-statistical.

The critical reader might suppose that the structure we propose is designed to relegate all the difficult problems to the realm of the "extra-statistical." But this is not sweeping them under the rug. Just the opposite. They are exposed as potentially weak links in the chain of inference about which statistics has nothing to say. [This does not preclude further statistical studies being carried out to address some of these problems (e.g., further investigation of study error).] The five-stage structure is a template for any statistical investigation and so its applicability could be regarded as a demarcation criterion for statistics. Post hoc, the structure allows us to identify the strengths and weaknesses in the statistical argument; in some investigations, even weak arguments may be all that are available. Ad hoc, it provides a useful strategy for finding out about populations and their attributes.

Many instances of PPDAC could occur within a scientific enquiry. Sometimes one PPDAC sequence will be nested within another as, for example, when investigating a measuring process or a sampling protocol within a larger study. Other times PPDAC sequences will occur one after the other or in parallel. The important point is that each PPDAC stands on its own as a linear structure from Problem to Conclusion. A cyclical representation as in [53] is misleading and confuses scientific enquiry with statistical method.

7. CONCLUSIONS

Statistics is not about the method of science with its paradigm shifts and incommensurability; it is about investigating phenomena as they relate to populations of units. The statistical method as we have described it is not the scientific method. (For those who wish to explore this point further, a confirmatory view can be found in [51].) As fascinating as the questions raised in Section 5 might be, they are not our questions. That is a good thing; the empirical evidence to date suggests that they may not be resolvable.

The five-stage PPDAC process with the associated language and substages provides a good framework for describing investigations such as Michelson's, especially for people learning the intricacies of statistics. More important, in actively planning and executing an empirical investigation, we believe that the framework is very valuable to ensure that important issues are at least considered—and this is the case for every statistical investigation. Although other organizations of the details are always possible, we believe that any such organization will be essentially isomorphic to the PPDAC structure and that this captures the method of statistics.

Karl Pearson had it almost right. Whatever the case for science, we can say that the unity of statistics consists alone in its method, not in its material. It is this method that should be given the broadest dissemination.

ACKNOWLEDGMENTS

Thanks are due to many people for many helpful discussions. They include our colleagues Greg Bennett and Winston Cherry of the Department of Statistics and Actuarial Science, astronomers Judith Irwin of Queen's University and Dieter Brookner of Kingston, who pointed out Cotter's book [13] to us, and Stephen Stigler of the University of Chicago for his helpful comments on early drafts of this paper.

All quantitative graphics were produced using the Quail statistical software environment available at www.stats.uwaterloo.ca/Quail. Research supported by NSERC, Canada.

REFERENCES

[1] ALEXANDER, A. F. O'D. (1970). James Bradley. In *Dictionary of Scientific Biography* (C. C. Gillispie, ed.) **2** 387–389. Scribner's, New York.

[2] ARISTOTLE. Nicomachean ethics. In *The Complete Works of Aristotle: The Revised Oxford Translation* (J. Barnes, ed.) **2** 1729–1867. Princeton Univ. Press (1984). (Translated by W. D. Ross with revision by J. O. Urmson.)

[3] ARISTOTLE. On the heavens. In *The Complete Works of Aristotle: The Revised Oxford Translation* (J. Barnes, ed.) **1** 447–511. Princeton Univ. Press (1984). (Translated by J. L. Stocks.)

[4] ARISTOTLE. On the soul. In *The Complete Works of Aristotle: The Revised Oxford Translation* (J. Barnes, ed.) **1** 641–692. Princeton Univ. Press (1984). (Translated by J. A. Smith.)

[5] ARISTOTLE. Sense and sensibilia. In *The Complete Works of Aristotle: The Revised Oxford Translation* (J. Barnes, ed.) **1** 693–713. Princeton Univ. Press (1984). (Translated by J. I. Beare.)

[6] BACON, F. (1620). Novum organum. In *Novum Organum: With Other Parts of the Great Instauration* (P. Urbach and J. Gibson, ed.) 33–293. Open Court, Chicago. (1994 translation.)

[7] BARNETT, V. (1982). *Comparative Statistical Inference*, 2nd ed. Wiley, New York.

[8] BEECKMAN, I. (1629). Folio 340, 19 March 1629. In *Journal Tenupar Isaac Beeckman de 1604 à 1634* (C. de Waard, Ed.) **3** 112. Nijoff, The Hague (1939–1953).

[9] BIPM, (1983). *Comptes Rendus des Séances de la 17^e CGPM*. BIPM, Sévres, France.

[10] BOX, G. E. P. (1976). Science and statistics. *J. Amer. Statist. Assoc.* **71** 791–799.

[11] CHATFIELD, C. (1995). *Problem Solving . A Statistican's Guide,* 2nd ed. Chapman and Hall, London.

[12] COHEN, I. B. (1940). Roemer and the first determination of the velocity of light. *Isis* **31** 327–379.

[13] COTTER, C. H. (1968). *A History of Nautical Astronomy*. North-Holland, Amsterdam.

[14] DEMING, W. E. (1953). On the distinction between enumerative and analytic surveys. *J. Amer. Statist. Assoc.* **48** 244–255.

[14B] DESCARTES, R. (1634). Letter to Beeckman, August 22 1634. In *Oeuvres de Descartes* **1** 307–312. (C. Adam and P. Tannery, eds.) 1969, Librairie Philosophique J. VRIN, Paris.

[15] DRAKE, S., ed. (1957). *Discoveries and Opinions of Galileo*. Anchor Books, Doubleday, New York. (Translated with an introduction and notes by Stillman Drake.)

[16] DRAKE, S. (1978). *Galileo at Work: His Scientific Biography*. Univ. Chicago Press.

[17] DURBIN, J. (1987). Statistics and statistical science. *J. Roy. Stat. Soc. Ser. A* **150** 177–191.

[18] EDGEWORTH, F. Y. (1885). Methods of statistics, In *Proceedings of the Silver Jubilee of the Royal Statistical Society* 181–217.

[19] Feyerabend, P. (1988). *Against Method*, rev. ed. Verso of New Left Books, New York.

[20] Fisher, R. A. (1925). *Statistical Methods for Research Workers*. Oliver & Boyd, Edinburgh. (13th ed. published 1958.)

[21] Galilei, G. (1610). The starry messenger. In *Discoveries and Opinions of Galileo* (G. S. Drake, ed.) 21–58.

[22] Galilei, G. (1615). Letter to Madame Christina of Lorraine Grand Duchess of Tuscany concerning the use of Biblical quotations in matters of Science. In *Discoveries and Opinions of Galileo* (S. Drake, ed.) 173–216.

[23] Galilei, G. (1638). *Two New Sciences Including Centers of Gravity and Force of Percussion*. Univ. Wisconsin Press, Madison (1974). (Translated, with introduction and notes, by Stillman Drake.)

[24] Gaukroger, S. (1995). *Descartes: An Intellectual Biography*. Clarendon, Oxford.

[25] Good, I. J. (1988). Scientific method and statistics. In *Encyclopedia of Statistical Sciences* **8** 291–304. Wiley, New York.

[26] Hacking, I. (1985). *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*. Cambridge Univ. Press.

[27] Hahn, G. J. and Meeker, W. Q. (1993). Assumptions for statistical inference. *Amer. Statist.* **47** 1–11.

[28] Hand, D. J. (1994). Deconstructing statistical questions. *J. Roy. Statist. Soc. Ser. A* **157** 317–356.

[29] Hume, D. (1739–1740). *A Treatise on Human Nature*. Oxford Univ. Press. [Analytical Index by L. A. Selby-Bigge (1888); 2nd ed. (1978), with text revised and notes by P. H. Nidditch.]

[30] Kendall, M. G. and Stuart, A. (1958). *The Advanced Theory of Statistics* **1**. *Distribution Theory*. Griffin, London. [First published in 1943; revised in 1958 with A. Stuart.]

[31] Kobal, Z. (1975). Ole Christensen Römer. In *Dictionary of Scientific Biography* **11** 525–527. Scribner's, New York.

[32] Kuhn, T. S. (1962). *The Structure of Scientific Revolutions*. Princeton Univ. Press. (2nd ed., 1970.)

[33] Lakatos, I. (1978). Falsification and the methodology of scientific research programmes. In *The Methodology of Scientific Research Programmes: Philosophical Papers* (J. Worrall and G. Currie, ed.) **1** 8–101. Cambridge Univ. Press. (First published in 1970.)

[34] Lindberg, D. C. (1978). Medieval Latin Theories of the Speed of Light. In *Roemer et la Vitesse de la Lumière* (R. Taton, ed.) 45–72. Librairie Philosophique J. VRIN, Paris.

[35] Losee, J. (1993). *A Historical Introduction to the Philosophy of Science*, 3d ed. Oxford Univ. Press.

[36] Madden, E. H., ed. (1966). *Theories of Scientific Method: The Renaissance through the Nineteenth Century*. Univ. Washington Press.

[37] Michelson, A. A. (1880). Experimental determination of the velocity of light made at the U. S. Naval Academy, Annapolis. *Astronomical Papers* **1** 109–145.

[38] Michelson, A. A. and Morley, E. W. (1887). The motion of the earth relative to the luminiferous ether. *Amer. J. Sci.* **34** 333ff.

[39] Michelson, A. A., Pease, F. G. and Pearson, J. (1935). Measurement of the velocity of light in a partial vacuum. *Astrophysics Journal* **82** 26–61.

[40] Moore, D. and McCabe, G. (1993). *Introduction to the Practice of Statistics*, 2nd ed. Freeman, New York.

[41] Nelder, J. A. (1986). Statistics, science and technology. *J. Roy. Statist. Ser. A* **149** 109–121.

[42] Newcomb, S. (1882). Measures of the velocity of light made under the direction of the Secretary of the Navy during the years 1880–82. *Astronomical Papers* **3** 107–239.

[43] Pearson, K. (1911). *The Grammar of Science*. Meridian Books, New York. [First published in 1892, revised 1900, final edition in 1911. Reprinted (1957) with introduction by Ernest Nagel.]

[44] Popper, K. R. (1959). *The Logic of Scientific Discovery*. Basic Books, New York. (First English edition; original German edition published 1934.)

[45] Sakellariadis, S. (1982). Descartes' experimental proof of the infinite velocity of light and Huygens' rejoinder. *Archive for History of Exact Sciences* **26** 1–12.

[46] Sobel, D. (1995). *Longitude: The True Story of a Lone Genius Who Solved the Greatest Scientific Problem of His Time*. Walker, New York.

[47] Stigler, S. M. (1977). Do robust estimators work with REAL data? (with discussion). *Ann. Statist.* **5** 1055–1098.

[48] Stigler, S. M. (1986). *The History of Statistics: The Measurement of Uncertainty before 1900*. Harvard Univ. Press.

[49] Swenson, L. S., Jr. (1972). *The Ethereal Aether: A History of the Michelson–Morley–Miller Aether-Drift Experiments* 1880–1930. Univ. Texas Press.

[50] Taton, R. (1971). Gian Domenico Cassini. In *Dictionary of Scientific Biography* **3** 100–104. Scribner's, New York.

[51] Tukey, J. W. (1980). We need both exploratory and confirmatory. *Amer. Statist.* **34** 23–25. [Reproduced in *Collected Works of John Tukey* **4**.]

[52] Wheeler, D. J. and Lyday, R. W. (1989). *Evaluating the Measurement Process*. SPC Press, Knoxville, TN.

[53] Wild, C. J. and Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *Internat. Statist. Rev.* **67** 223–248.

[54] Youden, W. J. (1985). *Experimentation and Measurement*. National Science Teachers' Association, Washington, DC.