which forms the image to be observed."[24] 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."[25]. Some adjustment in the calculations was made for the tilting of the mirror.

5. The distance |SR| from the revolving mirror to the cross-hair of the eyepiece was measured using the steel tape.
6. The vertical cross-hair 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 cross-hair 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 mm. per turn of the screw.
12. Steps 10 and 11 were repeated until ten 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

Using the apparatus and methods described above, Michelson began the first of his many studies to determine the speed of light. The study was conducted in 1879 and the results were published in 1880. Here we examine the detail of the study to illustrate what we mean by *statistical method.*

Statistical method, unlike Scientific Method as we will see in the next section, 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. One stage leads to the next and is dependent on previous stages. Looking back, this means that each 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, decisions can be made that will simplify actions taken in a later stage (e.g. a well specified Problem can be addressed by a simple Plan).

The 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. Below, we use the structure in the second manner to examine Michelson's 1879 study.

Each stage of statistical method comes with its own issues to be understood and addressed. In the context of Michelson's study, we introduce language appropriate to describe these issues. As noted in the Introduction, we chose Michelson's study for several reasons. In many ways, it is not typical of a statistical investigation. We urge the readers to test the proposed structure and language on other applications.

[24] *Ibid.*
[25] *Ibid*, page 124

4.1 The Problem

The purpose of this stage is to provide a clear statement of what is to be learned from the study. To do so, it is important to translate the contextual problem under study into a language that can guide the design and implementation of the subsequent stages of Statistical Method.

Understanding what is to be learned from the study 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 [38] makes no mention of the role of statistics in problem formulation. Two exceptions are the paper by Hand [27] and Chatfield's book, [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.

One of our goals is to provide appropriate language to formulate the problem and a structured approach to this formulation.

4.1.1 Units and Target Population

- these specify the collective to which we are interested in applying our learning or conclusion.

To illustrate the terminology, 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 is one transmission of such light between a source and destination, both located in a vacuum. The target population is all such transmissions, before, during and after 1879.

4.1.2 Variates

- numerical or categorical values attached to every unit in the target population (values of variates may 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, there is no direct interest in these other variates.

4.1.3 Population Attributes

- functions that apply to the entire population calculated through the variate values on individual units.

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

4.1.4 Problem Aspect

- either *causative*, *predictive* or *descriptive*. A problem with a causative aspect corresponds to one where interest lies in investigating the nature of a causative relationship involving two or more variates in the target population. A problem has a predicitve 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 here is descriptive. The aim is to estimate a population attribute. If Michelson had 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. It is important to decide the aspect at the problem stage because of the special requirements of the Plan needed to establish causation.

4.1.5 Discussion

This language can be used to clarify many statistical issues. Consider, for example, Deming's notion of an enumerative versus analytic study [14], [26]. If the target population can be listed so that a probabilistic sampling protocol giving every unit a positive inclusion probability can be used, the study is enumerative. Otherwise it is analytic. Deming was particularly interested in contrasting the use of formal statistical procedures in sample surveys to studies of industrial processes [26], where the target population included units not yet produced. All studies on such processes are analytic.

In applying Exploratory Data Analysis, the target population is the sample, the units that are included in the study. The object is to find interesting relationships among the usually many variates available. The relationships are defined as attributes in the target population.

Attributes can be numerical or graphical. For example, a scatter plot constructed in our imaginations using all units in the targt population is an attribute. The coefficients of the least squares line fitted to this scatter plot and the residual variation around the line are numerical attributes. A clear specification of the attribute of interest can resolve many issues. Lord's paradox, as presented by [27], is easily resolved by noting that it involves two different attributes. See our discussion to Hand.

Problems with a predictive aspect pose a serious challenge to the language introduced here. Suppose the value of a response variate is measured yearly for a number of years and the object is to predict the value for a subsequent year. What is the target population? Are there attributes that can be defined to specify the Problem? We return to this issue in Section 6 when the language for all stages of PPDAC is available.

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 sub-stages, 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. For numerical attributes, the difference between the attributes of interest in the study population and the corresponding attributes in the target population is called *study error*.

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.[27] 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

[26] We have found it is often useful to specify the target population by describing the process that generates the units. This process is called, naturally, the target process

[27] 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 [37]

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.

There are several competing criteria in specifying the study population. Ideally, the attributes of interest, defined for the study population should equal those in the target population to eliminate study error. This can be guaranteed only if the target and study populations are the same, a very rare circumstance. In most cases, we need to rely on the judgement of the investigators to assess how different these attributes are. Applying statistical method ensures that this comparison is considered before the Plan is executed. Since all units in the study population by definition must be available for inclusion in the study, this population is at best a subset of the target population. For example, in process studies, the study population is limited to those units to be produced over a selected time frame. In many cases, such as the use of animals to study human health issues, for economic or ethical reasons, the study units are fundamentally different from the units in the target population.

This language can again be used to discuss many important issues. For example, in meta-analysis , one major issue is the inclusion or exclusion of studies from the analysis. One aspect of this issue can be discussed by comparing the study population to the target for each study considered for inclusion.

It is our contention that all study populations are finite. Within the specified time frame, the limited financial resources available, and the time taken to make a determination, Michelson's study population could have been enumerated. Of course, this was not done; the point is that it was possible. Accepting the contention, it follows that all studies are potentially enumerative if conclusions are applied to the study population. Extrapolating these conclusions to the target population requires induction and is context dependent.

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 cross-hairs 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 formulae (3) and (4) ([36] page 133) to produce a value for the speed of light in air at temperature T.

4.2.3 Dealing with explanatory variates

There are usually many other variates associated with each unit in the study population. We call these explanatory variates that can be used to explain differences in the response variates from unit to unit in the study population.

It is important to decide how explanatory variates will be dealt with during the planning stage. There are three choices. First, the study population can be redefined by holding an explanatory variate fixed or by restricting its range of values. Second, the explanatory variate can be measured or deliberately set for each unit included in the study and its value utilized in the analysis or third, 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 (i.e. it does not explain variation in the response variates) or out of ignorance, not recognizing the presence or importance of the variate.

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 defining 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 twelve 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 [36] (page 141) "... error due to neglecting barometric height is exceedingly small".

We have constructed a fishbone diagram (Figure 8), a useful tool for deciding how to deal with potentially important explanatory variates.

The primary difference between experimental and observational studies can be explained with this language. In an experimental study, values of explanatory variates corresponding to factors of interest are set by the experimenter and assigned to units in the sample. In an observational study, these explanatory variates are not controlled, except perhaps by 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 *measurement 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 Conclusions.

In many applications, an iteration of PPDAC is applied 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 *bias*, an attribute of the (target) measuring process describing systematic measurement error, and *variability*, an attribute of the (target) measuring process describing the change in the 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 bias and variability. Consider, for example, the measurement of the distance between the two mirrors [36](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 can be moved towards 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 .1 ft corresponded to the cross-hairs of the fixed microscope and the length of the first yard of the tape was determined by rotating the screw until the cross-hairs 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 .1 ft 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 2000 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 mm at 20 ft intervals was measured. The data are shown below.

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

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

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

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

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 measurement of the distance between the two mirrors. To reduce the variability of the distance measurement, the procedure was repeated 5 times (with corrections for temperature on each). The temperature corrected measurements varied from 1984.93 to 1985.17 ft. Michelson used the average of the 5 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 and reducing variability by averaging. At the conclusion of his paper, Michelson provided a detailed discussion of the effects of possible measurement 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. [28]

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 13 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, 1000 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 effect the measured velocity of light.

[28] 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 [38] 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 [51] and Wheeler and Lyday [50].

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, bias and 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 representation of the sampling protocol by a mathematical model. 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.

In experimental plans, once the units have been sampled, the value of one or more explanatory variates is deliberately manipulated on the sampled units before the response variate is measured. The *experimental design* is that part of the sampling protocol that deals with how this manipulation will be done.

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. The goal is to avoid mistakes.

Michelson gives us no indication of how he planned to record 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, amongst 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 primary purpose of the Data stage is to execute the Plan, monitoring any deviations or exceptional occurrences as they occur. Once the data are collected, processed (Michelson had to calculate the speed of light for each determination) and stored, we propose to search for anomalies and to cleanse the data set when appropriate. The goal is internal consistency. This is likely to be more profitable in an active use of PPDAC as questions about the validity of any particular value can be answered directly by the individuals making and recording the measurements.

4.3.1 Execute the plan

As far as we can tell, Michelson used all of the measurements on the 1000 units that he collected. Unfortunately, he did not report all 1000 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

Michelson did not question the internal consistency of his data in print. Given our computational resources, it is an easy task to examine a large number of plots of the response variates versus the order of collection or other explanatory variates not involved in the Problem. Figure 9 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 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. However, checking with the data as presented in the tables, 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 downwards.

4.3.3 Data Examination

This also holds for morning and evening measurements considered separately. Clearly, the date is acting as a surrogate for some other lurking variable. What that could be is not known.

4.3.4 Data Storage

The values for the measured speed of light in air for each set and the associated response and explanatory variates are given in Table 2 and 3. Table 4 explains the columns in the tables.

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 the complexity of the Problem and Plan, the skill of the analyst, the amount of variability induced by the Plan, and the audience for the documentation of the study. We propose the following general breakdown of the stage:

- construct graphical and numerical summaries selected to address the Problem directly.
- model the Plan and data
- fit and assess the model
- develop formal statistical procedures

All sub-stages are directed to addressing the Problem.

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 [46].

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

$$299944 \pm 51$$

Although Michelson did not formally propose a model, he carried out numerous checks that are equivalent to aspects of model assessment ([36] 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

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 2: Michelson's data: First 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 3: Michelson's data: Last 50 observations.

Speed	-	Encoded speed of light in air. Add 299800 to get km/sec 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 I 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.

Table 4: Michelson's data: Key to variates.

removed from the determinations? A plot of speed versus temperature is shown in Figure 10. 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 over-plotting 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.

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 gausssian model, which appears to fit the data well, a 95% confidence interval for the mean is

$$299852.3 \pm 15.7$$

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

$$299944.3 \pm 15.7$$

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 modeling, analyses and model assessment can be made. The above is used to demonstrate the sub-stages 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.

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 and graphical summaries can 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.

In Michelson's study, he concludes by reporting the speed of light (km/s) in vacuo as 299944 $\pm$ 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 bias in Michelson's final reported speed. Note that the defined true value is well outside both the confidence interval and Michelson's interval of plausible values.

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.[29] 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 can't 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 [31] 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^{b}19 - 36). 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

[29] John Losee's book [34] provides a reasonable starting point.