École Polytechnique de Montréal

Marc Bourdeau

Technologies statistiques

Variations on variation

________________ September MMVIII ________________

What is the length of 6.35cm Nails?

Measurements and Uncertainties

There is no true value of anything.
W.E. Deming[1] [1900 – 1993]

Summary

This exercise starts by using a very simple example to present the problem of evaluating a measurement system, and of taking an exact measurement. We then define an experimental protocol, using a very simple example that serves as a paradigm: the measurement of nail lengths. As always, we start by identifying the factors influencing the measurement (generally, the answer we seek to obtain, the computer model, etc.). In this case, (1) operators (measurement takers); (2) measuring instruments; (3) objects to measure. The experimental data has already been collected. In the first part of the data analysis, we submit the data to very rudimentary observation techniques. This reveals that we can eliminate the measurements taken by one of the operators (measurement takers) in order to insure operator homogeneity, and to verify that the measuring instruments do not measure in an identical manner. In the second part, we use more elaborate graphic techniques to show that in the end we cannot be sure of the answer.

Key words. Data description, data observation techniques, experimental protocol, influencing factors, means, standard deviations, standard errors, graphs, Tukey diagrams, histograms, Quantile Quantile Plots, scientific data analysis.

---

1 From W.E Deming's preface to *Statistical Method from the Viewpoint of Quality Control* (1938) by Walter Andrew Shewhart [1891 – 1967], reedited with this new preface by Dover Publications, New York NY.

Part One

Measures & Sources of Error

Fig. 1 The cross–staff or Jacob's staff (etching by Van Keulen, 1760) was used to determine the latitude of a position at sea.

From the 16th century onward, measuring a position at sea became an absolute necessity to avoid being lost at sea and to enable vessels to reach their intended destination once they left charted costal areas.

The cross-staff or Jacob's staff was one of the earliest instruments allowing such measurements. The operator of the cross-staff measured the angle between the sliding vane (the cross–piece) focused on the horizon and the sun at its summit, or noon, that he had to be able to determine. The date was also important, requiring determination of the altitude of Polaris.

To obtain this angle, the measurement taker had to push the sliding vane along the graduated main staff (the other piece); this allowed him to determine the altitude in solar degrees (or Polaris degrees) and to convert it to latitude using astronomical charts.[2] This instrument was invented using trigonometric calculations (already an advanced technology!), probably at the beginning of the 14th century, by Levi Ben Gerson [1288 – 1344] of Provence and of Languedoc. The cross staff underwent improvements over several centuries, and was at the origin of sextants, which were used approximately until the century. It is not difficult to imagine how inaccurate these instruments must have been, even in their most improved versions.

Inaccuracy factors. Taking measurements on a moving vessel in changing weather conditions (if nothing else, a clear sky was a must !) must have been a delicate operation: sighting in blinding sunlight ,or at night for the Polaris – an even more difficult feat. The operator was not always skilled at determining the angle on the graduated main staff; as for solar measurements, how accurate were local noon readings obtained using spring and pendulum clocks subject to variable weather conditions, to pitch and roll, etc.? The instrument itself was imprecisely calibrated and, moreover, the astronomical charts used to convert an angle into a latitude certainly lacked precision, since they were calculated by hand with only a certain degree of accuracy, and based on not very exact theories. In short, sources of error and factors of inaccuracy were plentiful!

---

[2] For the solar measurement only; the Polaris measurement was a direct reading.

But we might think that one day technological improvements of measuring systems will eliminate any errors, making it possible to obtain a true measurement. In fact, this will never happen.

This example is used to illustrate the approach that should be taken to establish the value of a measurement: what are the factors influencing this value ? What are the factors that can influence accuracy? This is the fundamental question underlying scientific inquiry. A measurement is a chain and each of its links is a source of error.

A scientific investigator must always keep this in mind. In truth, there is no science without a measuring system. Therefore, a student of statistics must understand, first of all, that *variation* or *uncertainty* is omnipresent and inevitable. Consequently, techniques to control it or reduce it must be developed. Be it in terms of measurement taking, measurement recording or measurement use, no certainty is ultimately possible and precision remains illusive. This is one of the factors explaining the need for statistical technology!

The Experiment

There is nothing like an experiment in a very simple setting to confirm the difficulty of obtaining precise measurements.

In a classroom, students measured 5 nails 6.35 cm (2.5 in.) in length.[3] The nails were measured by 19 students (designated here as "operators") who used two metric rulers ("measuring instruments"). Each operator used the first ruler twice on each nail.

Of course, the objective is to establish the true length of each nail.

Inaccuracy factors Theoretically, we are measuring nails of the same size. The three factors of inaccuracy must then be:

1 The nails or objects to measure ;
2 The rulers or measuring instruments ;
3 The operators of these instruments.

For any measurement, it is essential to identify the factors influencing its value, and possible sources of error. The technique available for doing this consists of Ishikawa diagrams, or cause-and-effect diagrams, discussed in detail in class and illustrated using several examples.

Experimental protocol It is important to plan experimental conditions with care. If need be, we can start with a pilot experiment to insure in advance that no detail has been forgotten. A posteriori, a poorly designed experiment can be carried out again, if it has proven to be impossible to correct as it went along. It is also important to select system dimensions for the processing of the data obtained and recorded by the operators. This is usually done using a spreadsheet program. Dimensions are selected based on experimental objectives.

1 *Use of the instrument.* First, operators are shown how to use the measuring instrument. In our case, the ruler is placed along the nail at random (not with the zero at one end),

---

[3] The exact value of the old unit of measurement "inch" is based on the IS, as is the case for most units of measurement of the old *British Imperial* system or of the *United States Customary measure.* The inch is equivalent to 2.54 cm since 1959. It should be noted that in this age of globalization this old unit of measurement remains in use in only three countries: Liberia, founded at the beginning of the 19th century by freed (or affranchised) American slaves who, under the aegis of the American Colonization Society (sic) went back to Africa much to the regret of the native populations; Myanmar, former Burma; and finally the United States: all of them countries that cannot be said to be particularly progressive. Canada, given its closeness to the United States, has a kind of commercially mixed system, not entirely American, but modeled to some extent on the *British Imperial* system. This explains the existence of 1.89 l juice containers, equal to 2 American liquid *quarts*, not to be confused with dry quarts of the American system, that are slightly larger (1101.22 cm3 instead of 946.35 ml for *liquid quarts*), or with quarts of the Imperial system (1136.52 cm3), knowing that in Britain there is only one kind of quart, that must also exist in parts of Canada. Similarly, the 355 ml beer bottles found everywhere in Canada are the equivalent of 12 *fluid ounces* in the United States, a much simpler figure, surely…Be that as it may, some of these numbers must be rounded off. Should they be those in *fluid ounces* or those based on IS values? The reader is referred to **Ken Adler's** exciting book **The Measure of All Things: The Seven-Year Odyssey that Transformed the World, London UK: Little Brown, 2002**. The book is particularly interesting as regards scientific, political and commercial stakes.

allowing for greater precision. In fact, the measurement taker (who must know how to add…) benefits from two millimeter fractions, one at each end of the nail, to determine the length of the nail.

2. A certain number of nails is chosen randomly from the same box of nails – preferably an odd number, since this provides the measurements with a natural "center", i.e. a mean[4]
3. Long enough time intervals are left between the measurements performed by an operator, so that they don't influence each-other. At each measuring session, the objects to measure are presented in a different order. This prevents interfacing due to the order of the objects to measure. However, a record of the different orders must be kept, so that, the measurements for each nail can be grouped together without error.
4. The objects to measure (5 nails) are presented glued to a piece of cardboard, along with the ruler (measuring instrument) and a sheet of paper to record the measurements obtained, to each operator (measurement taker) in turn. It is important that the measurements of an operator not be known by the other measurement takers, to insure that operators are as independent of each-other as possible. (Click on the icon on the right for some illustrations.).
5. The operator records on the data sheet the measurements he obtained, and signs the sheet. It is important to know the identity of the operators simply for *traceability* purposes: to retrace the sources of error. This principle applies to all systems of production.
6. Since data processing system dimensions were set in advance, the data are entered, taking care to check the figures entered. Transcription errors are always possible and, as early as this stage, incorrect data can be identified and corrected.

Processing of the Data

Objectives What are the objectives of our experiment? First, to determine whether the characteristics of the measuring system allow us to obtain the length of the nails accurately. In effect, we are creating a metrology setting, frequently found in industrial environments.

Whether or not we can find the « true value » for the lengths of each nail will be deduced only indirectly.

It goes without saying that metrology is an important, essentially statistical, aspect of engineering sciences, and that here we merely refer to it in passing.

Click on the right-hand icon to see the experimental data, as well as the first data processing results.

The experiment involved two measuring instruments (rulers), the first used twice by each operator and for each of the 5 nails, and the second used only once. Under normal circumstances, another measuring session with the second measuring instrument should have been held. But practical considerations related to optimal use of the students' time shortened the experiment somewhat. Finally we did not keep all the data; only data from students who participated in all three parts of the experiment were entered into the file. Of a total of approximately thirty students at the start, natural attrition[5] reduced our sample of operators to 19…

The data (which can be imported by cliking this link) were modified at the start as soon as some thinking went into the process. At first, the data were entered by the operators using dimensions similar to spreadsheets; this eliminates one source of error. But this stage could have been avoided thanks to some reflection.

The modified data are presented in a form closer to a statistical data file, that is, a matrix

[4] In a production quality control situation, nails could have been chosen at random from different boxes. Here, in order to show the imprecision of our system of measurement (or of any system of this kind), we are trying to have the most homogenous nails possible. But sometimes, on the contrary, the objective is to maximize the variability of the objects to measure.

[5] Absenteeism, the fact that it was the beginning of the year, etc.; some students did not sign their data sheets or had illegible handwriting.

subject x variables

The lines are "subjects" or statistical units, columns are "variables" or descriptions of statistical units. These columns are the characteristics of statistical units or of the measurements themselves.

This first transformation allows us to arrive at a number of observations, specifically the means of the sizes of the 5 nails measured by each subject and using each instrument, for each use of the measuring instrument (two times for the first ruler and one time only for the second).

But these medians do not necessarily make sense in themselves; we still need to know if the measurements differ from one time to another; if the two rulers are identical, i.e. measure the same way; if the nails are "identical" or of the same length.

1. (2 points) Can you give another reason that would make it illegitimate to group together all the nail measurements and calculate the mean length?

Throughout the course, statistical technologies making it possible to answer these questions will be presented. For the moment, we will stay with descriptive observations (not strictly speaking statistical), that will allow us to draw conclusions related to our objectives[6]

An initial permutation of the column of the previous matrix given in the file, followed by an ascending sort (called "personalized" in Excel) facilitates answering the questions that follow. The new matrix is also included in the file.

2. Performances with the first measuring instrument.

First, we evaluate the *quality* of the operators by the consistency of measurements with the first ruler. Specific matrices must be designed, with the operators entered on the lines, and with measures of the quality of measurements in the columns. Thus we first used a variation measure, and we chose the standard deviation (we could also have chosen the variation between the maximum and minimum measurements).

(a) (3 points) Explain why the quality of an operator can be quantified by his standard deviation in the repeated use of an instrument on an object to measure. Why use multiple objects to measure in these repeated measurements? Once we have selected only the operators with similar qualities, how do we measure the quality of measuring instruments, if we use more than one? Finally, how can we make sure that the objects to measure are homogenous, i.e. of similar lengths? How then can we be certain of the measure of an object? (We do not expect perfectly complete and masterful answers.)

(b) (2 points) For the first measuring instrument, calculate the standard deviation for each operator and for each nail. You must design a new data matrix with the 19 operators entered on the lines, and with six columns showing the standard deviation between the two measurements made with the first instrument on the six nails by each operator. You will add two extra columns showing the means of standard deviation for each operator, as well as the standard deviations of these 6 standard deviations.

For the 19 operators, create graphic representations of the qualities of their measurements with the first instrument, first for several nails one at a time, then for several pairs of nails, and finally for all the nails at the same time. You will also represent, for each operator, standard deviation means and their standard deviation.

---

[6] It should be noted that any data analyst must be able to handle data using the computer programs available to him, in order to obtain information. Analysts know, of course, that all programs are not equally flexible: some operations and graphic designs – very useful in data analysis, since it is of foremost importance to *see* the data, √are simple using one program but can become complicated with another program.

(c) (2 points) Follow the same procedure for the means of the two measurements for each operator and for each nail. Design a matrix similar to the one you created at the preceding step (2b). Create a graphic representation of this sequence of means for each operator, for several nails taken one by one, as in the previous step. Can you still identify operators different than the rest? Do you have the impression that the nails are all more or less the same, or are there some that stand out from the rest? Give the reasons for your opinion in a few words. Are all the graphics you created equally useful?

(d) (2 points) After having calculated mean, minimum and maximum values for all operators of nail lengths, for each nail, and after comparing graphics of nail length means for each operator, for several pairs of nails, then for several triplets and finally for the totality of nails when using the first instrument, do you have the impression that some operators stand out? Give the reasons for your opinion in a few words.

(e) (2 points) Partially redesign the previous graphics by removing these out-of-the-ordinary operators. Can you tell if the nails are more or less the same? Give the reasons for your opinion in a few words. Write your conclusion in view of all these observations.

3. Comparative performances of the two measuring instruments.

We introduce the statistical measures of data spread. At this stage, we include all the operators.

You are to create, by adding columns to those of the previous matrix, the rows of measurements for each nail made with the second measuring instrument. The rows of this matrix are identical to those of the previously created matrix.

Add the following elements in rows: minimum, maximum, average and mean values for the first and second instrument, for the 19 operators, for each nail. These are data location indicators. You are also to enter the following rows that are data spread indicators: maximum–minimum differences, standard deviations and standard errors, that is, values of standard deviations divided by $\sqrt{n}$ where n is the number of observations made by the operators. Here, $n = 19$.

We now have observations for each of the 19 subjects (in lines or rows of the matrix), and for each of the 6 nails, indicating the values of average lengths obtained by the two measurements using the first instrument, as well as the value recorded for nail length. In addition, for each nail and instrument, the matrix gives location indicators and indicators of the variation of nail lengths, presenting 12 columns in all.

(a) (2 points) Create graphics that compare, for a certain number of nails, the 19 measurements of the operators made with the first and second measuring instruments. Can you formulate a hypothesis about the properties of the two instruments? Explain your answer.

(b) (2 points) Are operators who seemed different when using the first instrument still different when using the second? Explain.

4. Sequence to the comparison of the two measuring instruments.

Here, you are asked to purge the data by eliminating the out-of-the-ordinary operators you identified in section 3.

(a) (4 points) You will now use the values, for the first instrument and for each nail, of the means for the 19 operators calculated in section 3, as well as the values of standard errors.

For each nail, calculate an interval based on the mean for the first instrument, 2 x « standard error ».

This interval is a good approximation of the mean *confidence interval* that indicates a range with about 95% probability (this will be explained in great detail later in the course) for locating the mean. All surveys reflect a concept of this type: the mean will be valid with this 19 out of 20 (or 95%) margin of error; that is, 19 times out of 20 it will be located in this interval, called 95% confidence interval for another sample.

Finally, for each nail, verify if the values for average lengths obtained with the second instrument are located in the so-called confidence interval.

In conclusion, comment on the properties of the second measuring instrument compared to the first.

(b) (4 points) Create the graphic based on the 5 nails showing average values measured with the first and second instrument. Keep in mind the confidence intervals for the first measuring instrument, imagining them around the averages for each nail. What do you observe regarding the properties of the two measuring instruments? Are your observations confirmed by the locations of the 5 averages for each of the two measuring instruments?

5. You now have 10 measures for average nail lengths, 5 for each measuring instrument. The remaining operators seem sufficiently homogenous in their measurements, as reflected in these average overall measurements for each of the 5 nails.

(a) (3 points) Now calculate, for each measuring instrument and based on these 5 observations of average lengths, their means (these are means of means) and standard errors. Determine the two confidence intervals of these means, one for each instrument.

The nominal value for nail length is 6.35cm. Is this nominal value within the two confidence intervals? Are you ready to agree that the nominal value given is exact?

(b) (2 points) What is the exact value for the length of the third nail ?

And so we arrive at the inevitable conclusion. No, the value for the length of a nail cannot be measured exactly. If it cannot be measured, exact value for lengths does not exist… Well, does not « exist » in a sense we will not specify…But in that case, what lies at the bottom of natural reality, the reality of material things? This is another metaphysical question we will leave unanswered!

Hopefully, we don't need to make a general demonstration of the statement "*There is no true value of anything"* (Deming). This does not mean that there can never be agreement on an exact, "real" value, but the probability remains unlikely. Experiments using the principles illustrated here could be applied to the navigational instruments of the end of the 20th century (Fig. 3 & 4), rather than to nail lengths; but we would still see sources of error and imprecise measurements!

We can approach exact measurements but we will never attain them.

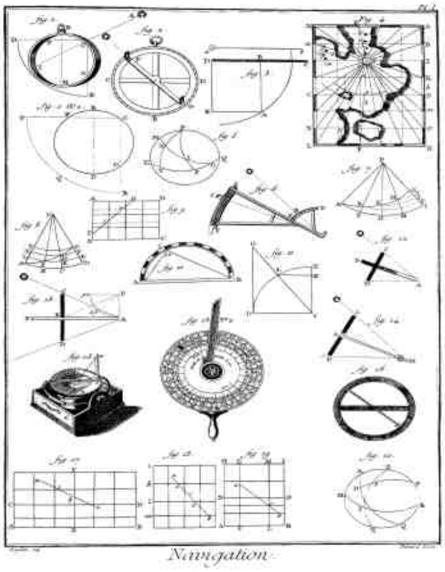

Fig. 2 – Navigational instruments in use at the end of the 18th century. Plate from the Encyclopedia of Aembert & Diderot.

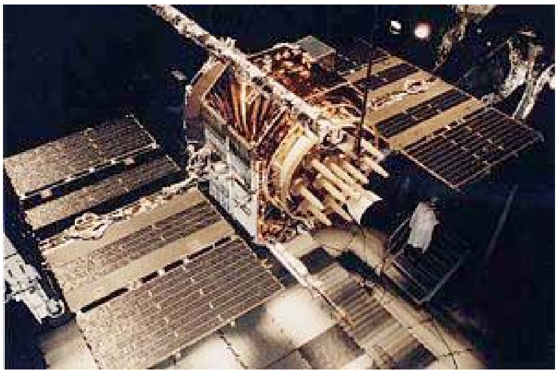

Fig. 3– NavStar satellite in the GPS family (end of 20th. Century)

Fig. 4 – Left: pocket GPS receivers ; right : vessel GPS serving to establish and control trajectories.

These illustrations are impressive to say the least...

Part Two

Over a hundred years ago, Karl Pearson proposed that all observations arise from probability distributions, and that the purpose of science is to estimate the parameters of these distributions.
David Salsburg[7]

Note: We are still working with the data pertaining to nail measurements. This time, we are adding a type of questioning that requires recourse to more complex observation techniques than those used in the previous analysis of the same data. We will now have to use a more complete statistical program than Excel.

The previous analyses make it possible to state that the data are coherent for all operators except the sixth. Therefore, the present analyses do not have to take into account the operator « variable » once the data of the sixth operator are eliminated.

Import into a statistical program the data from the first matrix, but with the data modified as described in Part One. In the following graphic representations, keep all the operators except the sixth.

The Experiment

In a classroom, we measured 5 nails of 6.35cm (or 2.5in) length. The nails were measured by 19 students (called 'operators' in this context), who used two metric rulers (called 'measuring instruments') ; the first ruler was used twice by each operator on each nail.

.

Supplemental Analysis

1 (5 points) Use the following observation techniques: histograms with an increasing number of categories (number of boxes): it is possible to have as many as twenty, if not more; Quantile Quantile Plots based on Gaussian law; Tukey diagrams, classified by measuring instrument. What do you observe? The observation techniques are complementary; explain in what manner.

2 (10 points) The *following observations will apply to the first instrument only*. Repeat the following analyses, using the common variable "nail". Identify measurements that are suspicious or clearly wrong. What are the most convenient graphic techniques for doing this? Without trying to explain these aberrant results, eliminate them from the analyses you are repeating. Are you ready to believe that the data conform to Guaussian law? Do you think that these principles differ from one nail to another?

3. (10 points) Finally, group together the data into a single undifferentiated block, still applying to the first measuring instrument and to all operators except the sixth; consider the observations as a whole.

(a) Construct Tukey diagrams and Quantile Quantile Plots based on Gaussian law. Observe the functioning of these techniques as you successively eliminate the odd elements. State you conclusions briefly, explaining your reasoning.

(b) Now construct histograms with increasing numbers of categories. Would you say that what you observe when the number of categories increases is due to the fineness of unit of measurement? In fact, measurements made with a ruler are limited by the precision of the instrument. Thus, when using a regular ruler, even if we use the most precise ruler measuring technique described in Part One of the description of these data, the measurement will show little precision within a

---

[7] David Salsburg, 2001, *The Lady Tasting Tea. How Statistics Revolutionized Science in the Twentieth Century*. New York NY : Henry Holt & Company, p.291

millimeter, values will differ very little[8] and boxes indicating small fractions will often remain empty. What do you conclude about the probabilistic model of measurement?

Brief Epistemology and History

Karl Pearson (1857 – 1936), who was one of the founders of statistical sciences, also played an important role in epistemology. In fact, he had an extensive philosophical background [9] He developed a series of ideas in support of the theory that ultimate reality about measurements is most closely related to their distribution. Since it is impossible to arrive at precise measurements of things, the only exact reality we can know about them is this: it is at the level of structural, and not natural, reality that probabilistic models can exist. But we can never be certain which models, any more than which parameters of the laws are involved, that is, specific models that do not offer any greater precision than the measurements themselves. Thus, ultimate reality is probabilistic… The certainties of the 19th century fly out the window!

It is not at all certain that these considerations have stood the test of time. Phenomenology[10] has wrecked havoc with our ideas of what we can truly know.

K. Pearson's book, *The Grammar of Science* (1892, reedited by Cosmo classics, New York NY. 2007) greatly influenced the thinking of the times. For instance, Albert Einstein (1879-1955), before finding a job as office clerk in the Bern (Switzerland) patent Office, while struggling to make a living, placed an ad offering to give math lessons. A group of friends who gathered around him (earning not a penny), adopted the name Olympian Academy (1902). Quite a name! The group came together to read and discuss books. The first of these, suggested by Einstein, was *The Grammar of Science*.

It was followed by others, such as the ones by Ernst Mach[1838 – 1916] (*Beiträge zur Analyse der Empfindungen*, 1886 [Contributions to the Analysis of Sensations]; it is noteworthy that Mach was a renowned physicist…), Henri Poincaré (*La science et l'hypothèse)* [Science and Hypothesis], John Stuart Mill, David Hume, Baruch Spinoza, etc. Many of the ideas that came to fruition particularly in Einstein's miraculous year (1905) had their roots in these high-level discussions[11]. Discussion and exchanging ideas among friends are an important source of exploration and discovery…in any field.

Pearson's book deals at length with questions of space and time (chapter VI), geometry and the laws of movement (chapters VII and IX), as well as matter (chapter VIII). These are all epistemological questions. Mach, a great scientist, also contributed to the field of epistemology, as did the other philosophers, all of them very important for the development of these ideas.

Poincaré's book *La science et l'hypothèse* (1902), the book read by the Olympian Academy, brought into question ideas commonly accepted at the time: absolute space and time, the notion of either… This followed directly from Pearson's ideas, but the perspective was much richer on the mathematical level.

Most of the authors who were read by the members of the Olympian Academy are linked with the ideas that contributed to Einstein's fame; these ideas concern ultimate reality. Knowledge is not born of nothing.[12]

---

[8] Most operators, if not all, will use values that are exact to one tenth of a millimeter at most.

[9] It is interesting to note that the "C" of the original first name, Carl, became a "K" after Pearson read Karl Marx. Socialism was and still is in vogue in England.

[10] The Wikepedia entry on phenomenology in French is less complete, but perhaps complementary. In any case, contemporary philosophy, at least that rooted in German idealism, often becomes a kind of gibberish.

[11] In this respect, it is interesting that Henri Poincaré (1854-1912), one of the greatest mathematicians of all times, is often credited with the first formulations of the equations and theory of general and restricted relativity.

[12] We can make two remarks in this regard. The first is that most of the works mentioned here gave rise to numerous writings in subsequent times, and that the Olympian Academy concerned itself with works of great quality and influence, published very recently. The system of information transmission worked very well… at least at this level. The second point is that it is thanks to the fact that they were armed with extensive knowledge of the natural world that many of these authors ventured to explore the deep philosophical questions ensuing from this knowledge. Today, philosophers seem unaware that their field is the culmination of outstanding work

Einstein, modest like almost all great men, said it himself: knowledge « is standing upon the shoulders of giants", a quote he borrowed from Newton.

But the Academy had its lighter moments. The gathered company read Cervantes, and Einstein played the violin for his friends…

For those who might be interested in a fascinating book on the quest of a mathematician & probability theorist:

Marc Petit, 2003, *L'équation de Kolmogoroff*[13] , Paris : Gallimard, coll. Folio n° 4240.

The horror of the Second World War, the quest of a young mathematical genius forced into hiding…

---

in parallel fields; most of them tend to build ideas on nothing, not to say about nothing, in short, they follow in the footsteps of German idealists. Given the impenetrable nature of scholarly philosophy, scientists feel justifiably excluded. In any case, they are so busy trying to conduct their "down-to-earth research activities" and to survive on public funds, that they have no energy left to break through the barriers of science, like Olympic athletes. Knowledge remains divided in two separate camps.

[13] Andreï Nikolaïevich Kolmogoroff (Tambov 1903 – Moscow 1987) is the creator of axiom probability theory (1933), which is essentially the theory taught even at introductory levels. His name is associated with the stochastic process and his work has had significant applications in physics, particularly disorderly phenomena theory (turbulence) and dynamic systems.