Questions and Answers (Q&A)

with Professor Haim Shore about his Research Findings regarding the Bible and Biblical Hebrew

As shown in:

"The Torah – Math Unveils the Truth"

(Movie initiated and produced by Mr. Oren Evron)

February, 2015

Comment: This document is divided into three sets of questions, each indicated by a heading (of same color as this comment), which precedes each set.

Questions relating to methodology, plots and statistical analysis:

Q1. Why had "simple *gematria*" been selected to assign numerical values to Hebrew words (denoted HNV in the movie), ignoring more complex traditional methods, for example, assigning different numerical values to certain Hebrew final letters (appearing at the end of the word while written differently, like final *Mem* or final *Nun*)?

A1. The essence of scientific research is looking for research findings that look at first glance meaningful and then testing, via scientific and statistical formal analysis, whether these findings are indeed significant (namely, accept or reject what superficially and intuitively seems to be of significance). Once I realized that I had obtained, via "simple *gematria*", significant statistical results no further attempt was made at alternative "*gematria* methods". Perhaps the comment hidden in this question has merit, and extending my research to other forms of *gematria* might deliver significant results as well.

Q2. The plots that are shown in the movie are impressive. However, it often seems that although a claim is made that the points indicate a linear relationship they are not really aligned exactly on a straight line. Why do we witness such points-configuration (arrangement), with deviations from a straight line?

A2. This is an excellent question that has been raised often and probably reflects a certain confusion that occasionally exists between mathematical modeling and statistical modeling (based on empirical observations).

With mathematical modeling we may have an equation of a straight line, for example: Y=3X+7, and we know that if we have inserted into the equation a numerical value, for example: X=10, then the value of the second variable in the equation is: Y=37. Furthermore, if we plot a graph of this equation, the selected point, and indeed any other point that we similarly calculate from the equation, would reside exactly on the plotted graph.

With statistical modeling, based on empirical observations, namely, observations taken from a natural phenomenon, the collected points would never align exactly on the plot, even if it is clear that the two variables, X and Y, maintain a certain relationship between them (of a straight line or otherwise).

In nature we always have a certain degree of randomness, besides measurement errors, that would cause the points to deviate from the "theoretical" mathematical/statistical model.

Are the random deviations small enough to determine that the statistical model is valid and describe well the relationship between X and Y? This is not a question that is answered in a subjective fashion! This question is answered only via an objective statistical analysis, using formal criteria so that no one would have doubts about the validity of the conclusions reached. Furthermore, any other individual that would perform same statistical analysis on the same set of data would arrive at the same conclusions.

To demonstrate all these with a numerical example (a fictitious example), suppose that I do not know whether there is a relationship between human weight and height. I collect a large number of observations of weight and height of people of various ages (including children), and then plot these observations in a graph (height (in centimeter) on the horizontal axis; weight (in kilograms) on the vertical axis).

If the scatterplot of the points would look like this (be reminded that all observations are fictitious):
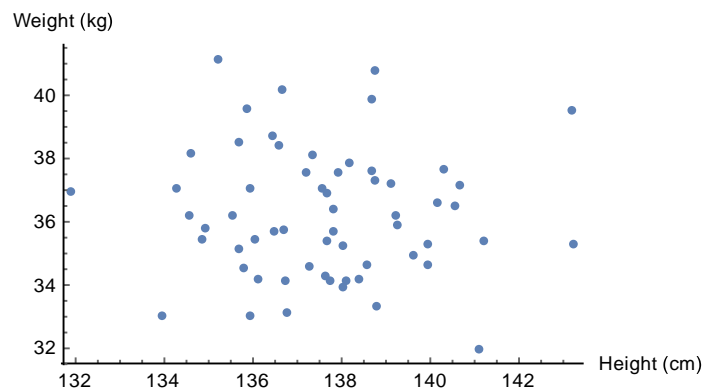

then obviously there is no relationship between weight and height.

Assume, however, that another sample of observations have been gathered, and the scatterplot looks like this:
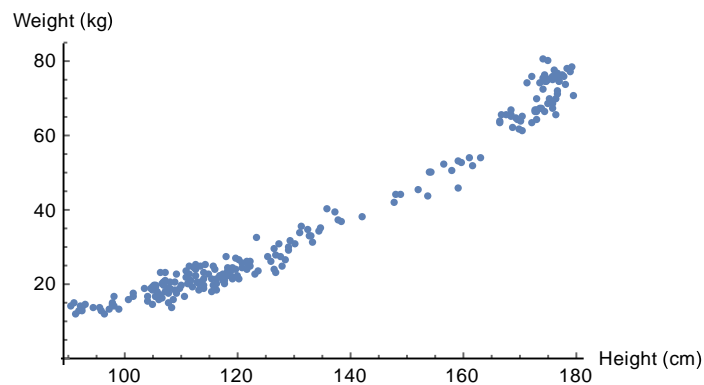
In this case, apparently there is a (statistical) relationship between weight and height. Therefore, we fit a statistical model (this time a nonlinear model), and we obtain the following plot of the model, along with the observations in the sample:
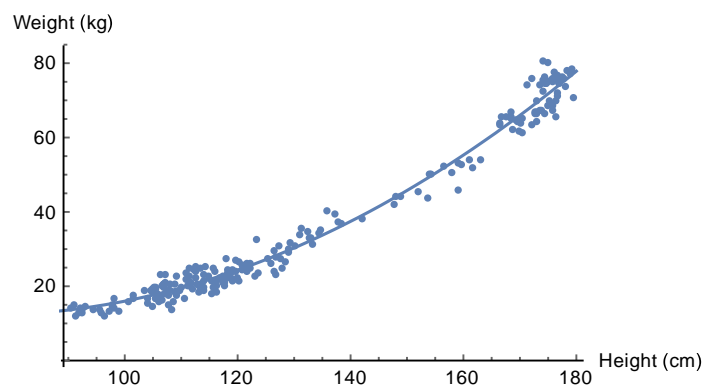
It is highly likely, according to the observed scatter of the points and the plotted model, that statistical modeling, and statistical analyses that accompany the fitted model, would lead us to the conclusion that there is a relationship between weight and height, and that the fitted model probably represents well this relationship.

And all of that without a single point actually aligning on the graph that represents the fitted statistical model!!

By similar vein, the straight-line models fitted to points in various plots of the movie are valid models that underwent formal statistical analysis to establish their statistical validity.

Q3. What are the dashed-lines in the plots (in red) and how were they generated?

A3. The dashed-lines indicate "Confidence intervals with 95% confidence". Described in a simplified and (statistically) not-highly-accurate fashion, if the fitted statistical model is valid then the dashed curves specify dispersion intervals, where points are expected to be found (with confidence of 95%).

The need to determine such intervals stems, as explained earlier, from the fact that empirical observations are always associated with a degree of uncertainty, generated by both natural randomness that exists in nature and from measurement errors. The confidence intervals quantify the influence of this uncertainty.

We note that if we chose a lower level of confidence, for example 90% (instead of the current 95%), the dashed lines would be even further apart from the curve (which describes the fitted statistical model), as well as from each other.

Q4. The use of log (e) delivers impressive results. However, it is not clear to me how the use of Log is justified. It looks to me like a kind of deception, or manipulation! Furthermore, the key-point for many of the analyses shown in the movie is that, as a result of the use of Log for the data on the vertical axis, points are aligned on a straight line. Would the points form a linear relationship even if Log was not employed for the data plotted on the vertical axis?

A4. We all know that there are several options to represent numerical values. For example, if I chose to display numbers in a binary format (of "0" and "1"), the number 77 would be written like this: 01001101.

As explained in the movie, using a log-scale constitutes a different approach to representing data, just as the binary method represents a decimal number using only {0,1}.

In science and engineering, it is quite common to statistically model on a log-scale when data that we wish to model stretch over many orders of magnitude; for example, the moon diameter (3,475 km) and the sun diameter (1,390,000 km). By transition to log-scale, one can model all data expressed on a scale of a single order of magnitude; for example, the moon diameter on a log-scale (8.15) and the sun diameter on a log-scale (14.1).

In other words: the scale of the data is now in the range {0-100} instead of the former original scale (of the data) of {0 —10,000,000}.

The question of why a straight line is formed when the data of the vertical axis are expressed by their log (rather than in the original scale) is equivalent to asking: "Why does the sun rise in the east and not in the west?"

Answer: Because this is how nature is. This is a finding of our observing of nature!! No explanation is feasible!!

(of course explaining this in terms of the rotation of Earth on its axis does not deliver an explanation, only formulate same question in different terms; still no explanation can be provided).

Similarly, one cannot explain why data that appear in various tables in the movie occasionally are aligned on a straight line, using the original scale, and in other times only when a log-scale is pursued. This question will always remain unanswered. The only logical response is: This is the empirical finding!!

The truly relevant and important question is not why we have obtained that finding and not another. This can never be answered. The real question is this:

"Is the alignment of the points around a straight line coincidental?"

(Namely, have a high probability of occurring randomly).

Having found this probability to be extremely small, we could produce the appropriate scientifically-valid conclusion, using same line of logic as pursued in any other discipline of science.

Finally, we note that for all examples that relate to celestial objects, it is always modeling on log-scale that produces meaningful results. As related to in the movie, this fact alone is miraculous because it enables any geometric property that characterizes a sphere (as all celestial objects are) to be linearly correlated with the respective Hebrew Numerical Values (HNV).

Q5. How many attempts have you made before you realized that the log-scale is the one that really work (in the sense of providing a linear relationship with HNV)?

A5. As one may already deduce from earlier answers, representation of data on a log-scale **is not** one of numerous possible alternatives!!

In statistical modeling it is **common practice** to model data either on the original scale or on a log-scale (for reasons described before). Therefore, there are only two commonly pursued options for statistical modeling of data. As is shown in the movie, in certain cases a statistically significant linear relationship is obtained for data represented on the original scale and for some on the log-scale.

WHY? Because that is the observed finding!!!...

Another question is why Log has base "e". Answer: For various reasons, some of which have been addressed in the movie. This is the common base adopted in scientific and engineering modeling (as Hertz is the common unit to represent measurements of cyclic phenomena). It should be noted, however, that even if a different base was selected, for example "2", the points in the plots would still be aligned about a straight line (though the values of the parameters that appear in the linear equation, obtained from the statistical modeling, would be somewhat different).

Q6. The reported goodness-of-fit for each plot that appears in the movie is impressive. However, comments posted on the Internet imply that people have calculated deviations of the points from the linear equations to obtain large deviations, implying that the points are not really aligned on a straight line; this result supposedly invalidates the linear-relationship claim. Would you care to explain this issue?

A6. Empirical observations, based on measurements, would never "conduct themselves" to fit exactly a mathematical model. There is a component of randomness associated with any observation of a natural phenomenon. This component reflects both natural random variation that exists in nature, like in the weight-height example (A2), and measurement error, which is added to the natural (random) variation. This implies that there is no sense in calculating deviations of points from the fitted mathematical model.

What scientists usually do is calculate the intensity of the relationship between the two variables modeled, like that between human's weight and height, and occasionally even among more than two variables.

If that relationship is expected to be linear we measure the intensity of this association by the linear correlation, which ranges from -1 to +1.

If the association between the two variables is purely mathematical and it represents a line that is increasing (like that between Celsius scale and Fahrenheit scale) the correlation would be +1. If the association between the two variables is purely mathematical but the line is decreasing (has a negative slope) the correlation would be -1. If there is no correlation (the points are scattered in a random cloud, like in the first plot above), the correlation would be zero.

Other scatter configurations will reflect intermediate degrees of intensity of association that will be reflected in the values of the calculated correlation.

As the scatter of the points around the line becomes smaller (points are aligned nearer to the line) the **absolute value** of the correlation would be larger (getting nearer to the limiting value of one).

In the movie, the correlation is related to as the "Percentage Match". This is a popular approximation to what the correlation really represent. A more exact (and more complicated..) scientific formulation would state that the linear correlation squared represents the percentage of the explained variation in Y (relative to the total observed variability of Y), that may be attributed to variation in X and its association with Y (as the latter is described by the fitted statistical model).

Q7. What is the difference between how we refer to purely mathematical data and how we treat data collected via empirical observation of nature?

A7. As explained earlier, mathematical "data" may be obtained at home, as you sit opposite a computer screen, introducing a certain value for X (the variable represented on the horizontal axis) and calculate from the mathematical equation the corresponding Y value. Points thus calculated will always "fall" exactly on the curve that describes the mathematical model.

6

By contrast, data obtained via empirical observation of nature include a component of randomness attributable both to natural (unexplained) variation and to the measurement error inherent in any measuring of a natural phenomenon. An example for the latter is when one measures repeatedly same temperature and repeatedly obtain different values, even though the "real temperature" stayed the same!! This is due to measurement error, which reflects the limited accuracy of any measurement device.

Conclusion: Unlike with purely mathematical "data", derived from calculations, empirical data always store a component of randomness; therefore one cannot expect plotted data to be aligned exactly on the curve that describes the fitted statistical model.

Q8. What is the chance of obtaining by random all of the results shown in the movie?

A8. There are several methods to calculate the requested overall probability, once you have probabilities of obtaining by chance each separate finding, for example, via Bonferroni inequalities. Since these methods are complex and at best approximate, we found it of no merit to engage in such calculation of probabilities.

Generally speaking, one could say that if all examples in the movie can be assumed to be statistically independent then the overall probability would be equal to the product of the individual probabilities (namely, extremely small).

Since we cannot make such an assumption (for example, the Hebrew *Eretz* (Earth) takes part in several analyses), the question of the overall probability is left open, even though it can justifiably be assumed to be very near zero.

Q9. How does the computer simulation work? Can anyone repeat the simulation results and conclusions?

A9. As succinctly explicated in the movie, the computer simulation comprises generating thousands of trios of words randomly, where the probability of sampling a certain letter (to be included in any "word") is equal to its occurrence rate (relative frequency) in the Hebrew Jewish Bible. Furthermore, the number of letters in each word of every simulated trio is the same as in the original Hebrew word (which the simulated "word" is supposed to replace). The net result of these two restrictions is **increasing** the probability of a trio of words to be aligned on a straight line (in other words, if these two restrictions were removed, the small probabilities reported in the movie would have been even smaller!!!). Once a trio of words had been simulated by the computer, HNV values were calculated for each word in the trio and a ratio was calculated between two slopes: the slope of the line that connects the first point with the second, and the slope of the line that connects the second point with the third (again please be reminded that on the vertical axis we register values of the physical property, and on the horizontal axis the HNV values).

If the trio of words are aligned exactly on a straight line (irrespective of whether the line is increasing or decreasing), the slopes-ratio is exactly one.

In Chapter 21 of my book I show histograms of the slopes-ratio obtained for various analyses. One may easily realize that the probability of obtaining a slopes-ratio of one or thereabouts is extremely small. Assuming that data represented by the histogram derive from a normal (Gaussian) distribution, estimates of means and standard deviations of the slopes-ratio were calculated and these served to calculate probabilities (as explained in detail in my book and summarized in Table 21.2 therein).

Regarding the second part of the question: Anyone equipped with a computer and a suitable software (I have used Wolfram's *Mathematica* to conduct the simulation study) can arrive at the same conclusions as appearing in my book. Note that simulation results may differ somewhat form one simulation study to the next (due to the computerized random sampling); however the statistical conclusions are expected to remain the same.

Q10. Would computerized-simulation deliver more accurate estimates of probabilities than those obtained from Linear-Regression Analysis (based on the original empirical observations)? If affirmative, please explain why!

A10. Statisticians examining the conclusions derived from the linear-regression analyses may claim that too few observations took part in the statistical analysis to derive valid conclusions. I reject this claim altogether. Linear-regression results, as displayed by the computer output, already take account of the number of observations in the sample. Therefore the probability calculations are valid or the computer would not show them. Furthermore, correlation values obtained for the analyses shown in the movie (and those not shown but appear in my book..) are so extreme (nearing +1 or -1) that one can hardly attribute them (statistically speaking) to chance.

Personally, I believe that conclusions based on the computer-simulation results are more reliable (in terms of the accuracy of estimates of probabilities obtained). The reason is that the number of assumptions, which need to be made in order for conclusions derived from the simulation study to be valid, is by far smaller than the number of assumptions associated with linear-regression analysis.

Furthermore, probability calculations based on simulation results **directly address** the question of whether the Bible-research findings could be attained randomly. Relating to linear regression, the central question in such analysis is whether the observed variability of the Y data (the vertical axis) is significantly larger than natural **random variation** that exists in the data. From the response to this question, one may infer conclusions regarding the probability of the findings to occur by chance. However, as one may realize, the "Randomness Question" is not addressed directly as within the framework of the computer-simulation analysis.

Questions relating to Professor Haim Shore:

Q11. Are all quotations in the movie, attributed to the professor, supported and backed by him? How about the scientific findings claimed to be the output of the professor's scientific studies (plots, tables and the like)?

A11. The movie's producers had approached me of their own initiative and kindly suggested that I accompanied the movie production process.

I had responded positively, clarifying that my objective was to ensure that although the results of my scientific study would be displayed in a popular fashion (in order to render them accessible to the public at large), they would be shown in the most accurate way possible.

I was happy to learn that during the production of the Hebrew version and later the English version, the movie's producers, Mr. Oren Evron and his colleagues, adhered meticulously to their promise to take account of my comments regarding display of the scientific results.

I have not commented during the production process on how the movie producers related to the significance of the findings addressed in the movie. I considered it inappropriate that I would in any conceivable way limit the freedom of thought, freedom of creativity and liberty of expression of the movie's producers.

All of these were mutually agreed upon in advance and, from my perspective, were strictly adhered to by the producers throughout the production process.

Q12. What is the overall significance of the scientific findings of the professor's studies, as he perceives them?

A12. I was strict in the past, and remain so at present, to relate only to the scientific findings and not address in public their non-statistical significance. Both in terms of experience, know-how or knowledge, I have no relative advantage regarding interpreting the significance of the results. Anyone can draw her (or his) own conclusions, as they find fit, and these conclusions are personal and private and cannot (should not) be interfered with.

Q13. The title of the professor's book includes "Coincidences". In the professor's view, are his findings really "coincidences", with no significance to be attached to them? Are they to be perceived merely as "curious eventualities"?

A13. My book comprises two sets of "Coincidences":

* Coincidences for which no statistical analysis had been (or can be) implemented;

* "Coincidences" that have been subjected to statistical analysis, resulting in statistically significant outcomes obtained in accordance with commonly accepted principles of modern-day science.

"Coincidences" displayed in the movie belong exclusively to the second set.

According to modern-day scientific approach to the investigation of nature, if under the null "Randomness Hypothesis" the probability of a certain scientific finding, obtained via a formal statistical analysis of empirical observations, is extremely low– one cannot refer to this finding as random.

Therefore, I am allowed to state, within the confines of science, that the phenomena addressed in the movie, which had been subjected to statistical analysis and produced (statistically) significant results, cannot be related to as coincidental.

I do not find it appropriate to convey any further sentiment or thought about the statistical results for reasons expounded earlier.

Q14. Why are the findings described in the movie not published in scientific journals?

A14. These findings have not been published in scientific journals **not because they were reviewed and rejected** (a common reviewing process is conducted for any article submitted to be published in a scientific journal).

I had approached several editors of prestigious scientific journals, explained the nature of my findings and requested that my article be subjected to a regular reviewing process (like any other scientific article submitted for publication). The response to my request had been uniformly negative.

In other words: The findings of my research have never been reviewed but by myself. Yet they have been exposed in their entirety to the public in my book (including detailed explanations of the scientific contents of the results).

In my book, I have addressed the publication issue (Section 22.5.3, carrying the heading: "Why not publish in recognized scientific journals?").

As referenced therein, I do not find fault with editors that have declined my request to let a paper about my findings undergo a regular reviewing process. My findings are so "impossible", so incomprehensible to grasp within a regular frame-of-mind, that one requires a major "diskette change" to agree to review such an article.

I resigned to this state-of-affairs (my book had been originally submitted to the publisher in the summer of 2006 and was published in January of 2007). I have concluded that perhaps conditions are not yet ripe and the time for a wide-spread circulation of my findings may not have yet arrived. I have assumed that this time would ultimately come, as indeed happened when Oren Evron (the movie producer) had approached me in mid-August of 2014.

(as of the time of writing this post (mid-February, 2015), nearly 400 e-mail communications have been exchanged between us regarding the production of the Hebrew and the English versions of this movie; Combined number of views of both versions of the movie, the Hebrew and the English, has by now surpassed 350,000 and it keeps growing).

Q15. "Is there a credible source that has reviewed your findings and is willing to publish them in his journal"?

A15. There is no such "animal" as "a source that is willing to publish the findings in his magazine". Scientific journals are commonly owned by prestigious international publishers, and each journal has its own editor and editorial board, mostly from academia. These are periodically being replaced as their tenure ends. I myself have been member in several such editorial boards and still serve as a member in some (though in fewer journals than before due to my desire to reduce such editorial activities, owing to my advanced age). Obviously, I myself have published in such scientific journals for many years, in fact over a hundred articles and additionally books and books' chapters (all accessible via my ResearchGate page).

I admit that after my initial restricted effort to publish my findings I have not persisted in pursuing this effort (my main academic research areas are quality and reliability engineering, statistical modeling and its engineering application). Naturally I will be forthcoming if an editor of a respectable scientific magazine had approached me with the intention of publishing (after a due reviewing process) an article outlining my research findings.

General questions:

Q16. Can one generate predictions based on findings displayed in the movie?

A16. There is a particular prediction derived from the statistical model linking Genesis creation timescale to the cosmological timescale. This prediction is outlined in the final part of the movie (based on a similar prediction detailed in my book). This prediction relates to the "creation time" of dark energy and dark matter, claiming that, according to the statistical model, these have preceded the Big Bang by about three billion years.

Q17. Are there other "issues" that may become known, based on HNVs of Hebrew words that appear in the Hebrew Jewish Bible?

A17. This question invokes the prediction issue, addressed in the earlier question. In principle, the option you outline in your question apparently does have merit and is worth exploring.

One such example (devoid of any statistical analysis) is detailed in my book. At the end of the nineteenth century, much research effort had been put to finding biological roots of dizziness and how this condition may be cured (both in animals and in humans). Ultimately, after much research efforts, spanning various groups of researchers over several countries, it was found that the source of the problem is damage to the balance mechanism that resides in the ear. Perhaps "consulting" biblical Hebrew would direct the researchers much earlier to the source of the problem: the words "Hearing, "Ear" and "Balance" in biblical Hebrew miraculously all derive from the same root!!

(find more details in Section 10.3.2 of my book).

Q18. Beyond your personal blog and this movie, are there additional sources where your findings are publicized?

A18. In my blog I have a post detailing, via provided links, sources where my research findings are addressed, or videos of my public appearances. These do not include Internet sites that relate to my research as these may be easily accessed by a simple search based on my name (in Hebrew or in English).

My personal blog address is: https://haimshore.wordpress.com