

Statistical Analysis of Carbon Dating

The usual measure of the precision of a measurement is the standard deviation, which actually is a measure of probability based on what is often called the “bell curve” (Figure 1). The meaning of the term, standard deviation, is that the true value of the measurement will be within plus or minus one standard deviation of the mean (average) value 68% of the time, or within plus or minus two standard deviations 96% of the time, or within 3 standard deviations 99% of the time. So, a measurement of 6.1 ± 0.05 g means that the true value of the mass will be between 6.05 and 6.15 g 68% of the time. If you want to be more certain than your range brackets the true value, you need to make the interval wider, using a multiple of the standard deviation.

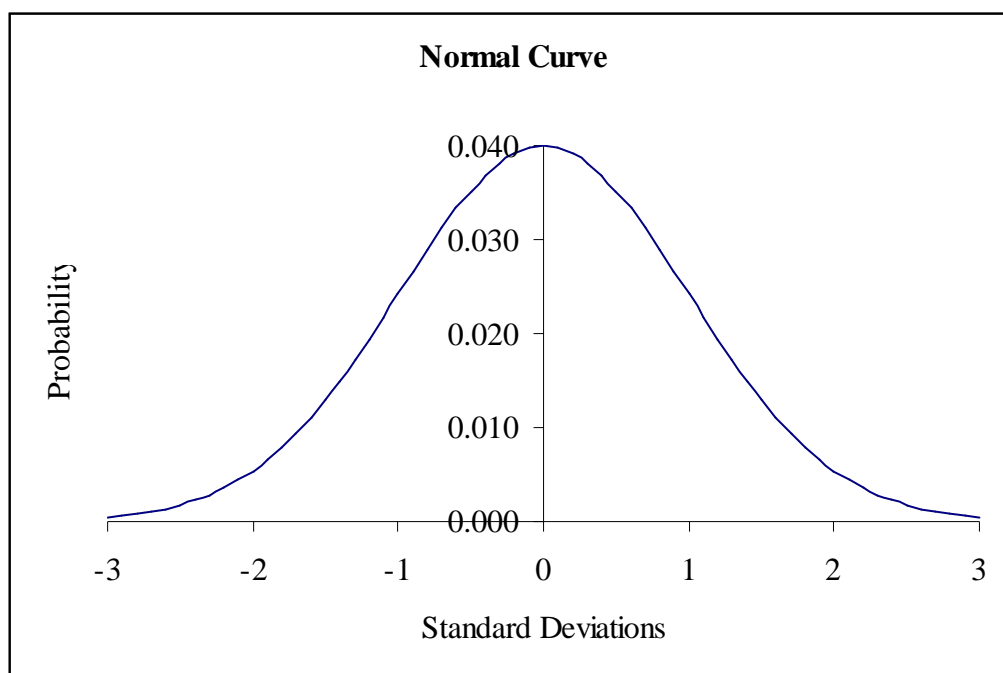


Figure 1. The normal curve

When more than one measurement is made and the measurements are combined mathematically to make a new number, the standard deviations must be combined, too. For example, if you measure the distance you traveled and the time it took to get there, and you wanted to report miles per hour, you would need to combine the standard deviations of the measurement of distance and the measurement of time in order to have a standard deviation for the value for speed.

The method to use for combining the standard deviations depends on the mathematical operation being performed. These methods of combination are based on probability, too. For example, if two lengths are to be added, the standard deviations are not added in the same way, because it is highly improbable that the true value of both measurements will be at the extreme edge of the standard deviation range. So, if two values are added, the standard deviations are combined as the square root of the sum of the squares. For example, 3.8 ± 0.5 ft plus 4.5 ± 0.7 ft = 7.7 ± 0.86 ft. ($\sqrt{(0.5^2 + 0.7^2)} = 0.86$) Notice that if the two values had been subtracted, the answer would be 0.7 ± 0.86 ft, which, in most cases, would be a meaningless value, since the

“window” for knowing the true value of the result even 68% of the time would be bigger than the value itself.

In the case of ^{14}C dating, several measurements are combined to obtain an estimate of the age of an article, and several assumptions are made. The tolerances inherent in the measurements must be combined in certain ways to determine the “window” in which the true value resides. Further, if any of these underlying assumptions are found to be unreliable, the entire calculation method becomes meaningless.

^{14}C is formed in the upper atmosphere by neutron bombardment of ^{14}N . The neutrons are a product of the nuclear reactions in the sun. The ^{14}C diffuses through the rest of the atmosphere and becomes incorporated into living things. In today’s world, a gram of pure carbon will experience approximately 5 disintegrations of ^{14}C per minute. If an object containing carbon is sequestered in some way so as to prevent it from interacting with the environment freely, then the atoms of carbon cannot be exchanged with the environment any longer. ^{14}C will continue to disintegrate but will not be replenished from the surroundings. Therefore, the age of an object may be estimated from the amount of ^{14}C remaining in a sample as compared to the amount that would be there if it were free to equilibrate with the environment.

This process depends on several assumptions.

- **Assumption 1:** The rate at which ^{14}C is produced in the upper atmosphere is assumed to have been constant during the timescale of the experiment. So, if something is dated at 10,000 years, the assumption is made that the abundance of ^{14}C was the same then as it is now. This assumes that the solar flux of neutrons was the same then as it is now. This certainly is not true, since the sun has been diminishing in size as it uses up fuel, and the reaction rates of the nuclear events in the sun have changed measurably several times in the last century.
- **Assumption 2:** The rate decay of ^{14}C has been constant over that time. In the scientific world today, radioactive decay is assumed to occur at a constant rate governed by probability. These probabilities have been developed from measurements, so they are called empirical formulas. No one really knows why a nucleus decides to disintegrate when it does, or if that rate can be changed by as-yet-unknown forces. Many people have tried to change the decay rate with an external stimulus. Nothing has changed it so far. However, a reasonable possibility does exist that some stimulus will be found that causes a change. For example, the decay rate could depend on the proximity of other radioactive materials or could be stimulated by a “direct hit” by one of the many sub-atomic particles that pass through us every second, so their flux would influence the rate of decay. Our current model says that a radioactive nucleus decays at a random moment, the impetus for which we do not know. If we ever find what triggers it, then we will be able to alter the rate artificially.
- **Assumption 3:** It is further assumed that these constant decay rates are constant according to the formulas we have developed. These rates have been measured only for a little more than half a century. As an example, the decay rate for ^{14}C in published literature has changed by 0.3% over the last 40 years (from a half-life of 5730 years to a half-life of 5715 years). Another reputable reference gave the half-life of ^{14}C as 5568 years. This could be due to better measurement techniques in some places as compared to others. Or, it could be that the decay rates do not obey the model we have created for

them. A change of 0.3% in 40 years seems insignificant. But, over a time span of 40,000 years, that rate could be very much different.

- **Assumption 4:** The method assumes that the various isotopes of carbon (^{12}C , ^{13}C and ^{14}C) are incorporated into natural materials in proportions that match the atmospheric abundance. However, ^{14}C is 17% heavier than ^{12}C . So, it moves more slowly. In some natural tissues, the abundance of ^{14}C is significantly lower than in the environment.

Getting back to standard deviations, if we accept that all the assumptions are valid, ordinary statistical analysis sets a limit on the values that ^{14}C dating can produce. The formula for that dating is $\ln(N/N_0) = -\lambda t_a$. N_0 is the predicted original amount of radioactive material, in this case, ^{14}C . N is the amount of ^{14}C that remains. λ is the decay constant for ^{14}C , which is assumed to be constant over all time spans. And t_a is time (or the age of the object, hence the variable name t_a) in years.

The predicted original amount of radioactive material, N_0 , is not a very precise value. It is based on the mass of the carbon sample and the generally accepted natural abundance of ^{14}C . The generally accepted value of the abundance is based on the generally accepted measurement of 5 disintegrations of ^{14}C per minute per gram of pure carbon. In the professional literature, I have found values for the natural abundance of ^{14}C that differ by more than a factor of 2. Of course, if the relative standard deviation for that natural abundance were 100%, all future calculations would be meaningless. (The relative standard deviation is the standard deviation divided by the value with which it is associated, expressed as a percent. For example, if a value of 5 has a standard deviation of 0.5, then the relative standard deviation is 10%.)

But, since the best values for the abundance of ^{14}C in natural carbon are based on the measurement of about 5 disintegrations per minute per gram of carbon, the natural abundance can be assumed to be $4.34 \times 10^{-13} \pm 10\%$. The relative standard deviation of 10% came from the fact that the original calculation started with 5, which is an approximation meaning that the true value is between 4.5 and 5.5. For subsequent calculations, the 10% relative standard deviation will be used, although this value may be much higher and certainly is not lower.

The next variable to address is N , the amount of ^{14}C that remains. The abundance of ^{14}C in the sample when it was removed from free exchange with the environment thousands of years ago is assumed to have been at the same abundance as it is today, so that value is obtained by multiplying the natural abundance times the mass of carbon in the sample. The amount of ^{14}C that is present in the sample today (the value of N) is expected to be much less than the original amount (N_0). The value of N cannot be measured directly; rather, it is measured indirectly based on the number of ^{14}C that disintegrate during a test period that is relatively short. For the sake of convenience, we will call that value N_d (number of disintegrations). The value of N can be derived from the value of N_d using the same equation as before: $\ln(N_x/N) = -\lambda t_i$. N_x is the number of ^{14}C atoms that remain after the laboratory counting period, t_i , the testing time. Since $N = N_x + N_d$ (the total number of ^{14}C atoms at the beginning equals the number that remain plus the number that decayed), the formula can be rearranged to include the value that can be measured (N_d) in order to find the value that is needed (N):

$$\ln(N_x/N) = -\lambda t_i$$

becomes

$$\ln((N - N_d)/N) = -\lambda t_i$$

becomes

$$\ln(1 - N_d/N) = -\lambda t_i$$

N_d is the number of disintegrations measured by a beta spectrometer. Only whole numbers are possible. In other words, half of one disintegration cannot happen. It either disintegrates or it doesn't. So, N_d is called a counting value. By definition, its standard deviation is $\sqrt{N_d}$. Its relative standard deviation is $1/\sqrt{N_d}$, or $(\sqrt{N_d})/N_d$. This can be a very significant percentage or an inconsequential percentage, depending on the number of counts collected. For example, if only 4 counts are collected, the standard deviation is 2, and the relative standard deviation is 50%. But, if 100 counts are collected, the standard deviation is 10 and the relative standard deviation is 10%. If 10,000 counts are collected, the relative standard deviation is only 1%.

The number of collected counts depends on the amount of sample and the length of time over which the sample is collected, since disintegrations can be expected at the rate of about 5 counts per gram of pure carbon if the sample has the normal natural abundance of ^{14}C . The older the sample, the less ^{14}C , and the fewer counts can be expected over a given period of time. So, the oldest samples have the highest relative standard deviation. At a certain combination of age, sample size, and counting time, the relative standard deviation becomes so large that the answer is meaningless.

In addition, many samples are, by nature, very small. Generally, the sample must be digested and chemically treated in order to isolate just the carbon, so the original sample is destroyed. Museums are not pleased to give up substantial portions of important artifacts for dating purposes. So they give as little as they can and still obtain relevant data.

Most samples do not start out as pure carbon. Wood, which has a lot of carbon, is still only about 46% carbon. This ratio is close to the carbon fraction of most plants and fibers. And, the chemical technique for isolating the carbon is not perfect, so some is lost. The quality of the laboratory generally determines the loss rate. The very best will lose less than 5%. Those in universities generally lose over half because the work is done by students. This loss rate does not contribute to the standard deviation, because all that matters is the purity of the final sample. But it does impact the required size of the original sample, which may be limited.

The equipment used to measure the decay of ^{14}C has an upper limit for the sample size, too. The largest sample that can be measured in conventional equipment is about 8 g of pure carbon. Actually, the carbon first is isolated from the sample. Then, it is hydrogenated to make benzene, which is a liquid and easy to handle. To this is added a combination of chemicals called a scintillation cocktail. When ^{14}C decays, it emits a beta particle (an electron). This electron hits the scintillation cocktail compounds and makes a small flash of light. The flashes of light are counted. Actually, any nuclear event will be counted. But, since the only radioactive material in the bottle is ^{14}C (everything else having been removed in the purification steps), all the flashes of light are attributed to ^{14}C . Unfortunately, nuclear events are always happening around us. One example is cosmic rays. These events also cause flashes, so there is a background of counts that will happen if any ^{14}C is present or not. An excellent background is 0.2 counts per minute (cpm). The typical background is 0.4 cpm. Unless disintegrations from ^{14}C occur significantly more frequently than these background events, the ^{14}C data is lost in the noise of the instrument. Using the typical value used by the EPA for environmental measurements, a value must be more than five times the background to be significant. So, a sample must produce at least 0.8 count per minute (a total of 1 count per minute) to be "countable." Other scientists have used lower values. Some reputable researchers have used a value of three times the noise as the detection limit, or a ^{14}C disintegration rate of 0.4 counts

perm inute for a total of 0.6 counts per minute. This idea will be applied to real values after a few more ideas are introduced.

The duration of counting (t_i) is limited by the drift in the electronics of the beta spectrometer. The best instruments can collect counts for 24 hours before drifting significantly. The older or less expensive models usually cannot go more than about an hour. But, assuming that the best instrument is used, at 5 disintegrations per minute, 7200 disintegrations per day would be collected per gram of pure carbon. For a 1-gram sample of pure carbon, the relative standard deviation of the count rate for fresh (not old) carbon would be 1.18%. This is a pretty good value, especially compared to the 10% relative standard deviation for N_0 (the number of atoms of ^{14}C when the sample was new, thousands of years ago). However, new carbon is not being counted. Rather, old samples are counted. If the sample were, for example, about 5,000 years old, it would be expected to yield only 2.5 disintegrations per minute, for a total of 3600 counts, with a relative standard deviation of 1.67%. This still is pretty good. But if the article is, for example, supposed to be 40,000 years old, the relative standard deviation climbs to 13.2% for a 1-gram sample. But, the relative standard deviation drops back to 4.7% if the sample size is increased to the maximum of 8 g. All of these numbers assume that the best instrument is used, and that it was able to count for a whole day without electronic drift. Further, it assumes that the instrument is 100% efficient. Most are about 80% efficient. The efficiency affects both the standard deviation and the point at which the counts from ^{14}C start to get lost in the noise. As N_d gets smaller, its relative standard deviation ($1/\sqrt{N_d}$) gets larger.

Going back to the original formula, $\ln(N/N_0) = -\lambda t_a$, and combining it with the laboratory formula, $\ln(1 - N_d/N) = -\lambda t_i$, a single formula can be obtained that allows measurable values to produce the desired result, the age of the material. The formulas must be combined to avoid including the standard deviation of a value more often than needed, which would produce an unwarranted increase in relative standard deviation that depended only on mathematical manipulation and not on measurement. Starting with

$$\ln(N/N_0) = -\lambda t_a$$

rearranging

$$t_a = -\ln(N/N_0)/\lambda$$

taking the negative inside the natural log

$$t_a = \ln(N_0/N)/\lambda$$

Switching to the other formula

$$\ln(1 - N_d/N) = -\lambda t_i$$

rearranging to isolate N in preparation for substitution into the first formula

$$1 - N_d/N = \exp(-\lambda t_i)$$

$\exp()$ is a convenient way of writing “e to the power of.”

This prevents having to have subscripts inside of superscripts.

Rearranging

$$N_d/N = 1 - \exp(-\lambda t_i)$$

$$N = N_d/(1 - \exp(-\lambda t_i))$$

One last expansion, $N_0 = mA$

(m = mass of carbon sample; A = abundance of ^{14}C in the carbon sample)

Substituting them all together:

$$t_a = \ln((1 - \exp(-\lambda t_i))mA/N_d)/\lambda$$

Having derived the formula that relates all the measurements, all the standard deviations may be compiled to determine the precision of the derived value. As can be seen in Table 1, most of the relative standard deviations are negligible.

Description	Variable	Value	Relative Standard Deviation
Number of disintegrations	N_d	25 to 2500	$1/\sqrt{N_d}$ (20% to 2.5%)
Decay constant	λ	$1.21 \times 10^{-4}/\text{yr}$	< 0.01%
Testing time	t_t	24 hr	< 0.01%
Mass of carbon	m	< 8 g	< 0.01%
Abundance	A	4.32×10^{-13}	> 10%

Table 1. Relative standard deviations of measured values

The relative standard deviations of λ , t_t , and m are negligible compared to the other values. However, it is important to remember that the relative standard deviation of λ is based on the rather large assumption that the decay rate has been constant for tens of thousands of years. These three relative standard deviations will be ignored through the rest of the discussion. If they were included, the math would get much messier, but the final value would not be noticeably different.

The relative standard deviation of the value of the abundance of ^{14}C (A) was 10% or worse. The relative standard deviation of N_d depended on the sample size, the counting time, and the age of the object. This may range from 1% to 25%, depending on age and sample size. When N_d and A are combined mathematically, the resulting relative standard deviation (the sum of the squares of the relative standard deviations, square-rooted) will be between 10% and 27%.

The last step in calculating the relative standard deviation of the age is to take the combined relative standard deviation of N_d and A through the natural logarithm function. This process is entirely different than the methods used for adding, subtracting, multiplying, or dividing. For adding or subtracting, the new standard deviation is the sum of the squares of the standard deviations, square-rooted. For multiplying or dividing, the new relative standard deviation is the sum of the squares of the relative standard deviations, square-rooted. For natural logs, it goes like this; the standard deviation of the new value is the relative standard deviation of the starting value. For example, if the combined relative standard deviation of N_d and A is 10%, then the standard deviation of the natural logarithm is 0.1.

The important point of that relationship is that you start with a relative standard deviation and end up with a regular standard deviation. To get it back into a relative standard deviation, you must divide by the original value. This can be very significant or not very significant, depending on the values. For example, consider the arbitrary equation $y = \ln(x)$ in which x has a relative standard deviation of 10%. The relative standard deviation of y would be $0.1/y$ (relative standard deviation is the standard deviation divided by the value which it modifies). If y is greater than 1, then the relative standard deviation will be less than 10%, an improvement over the relative standard deviation of x . But if y were less than 1, the resulting relative standard deviation would be greater than 10%.

In this case, the value inside the natural logarithm parentheses was, originally N/N_0 . This value is always less than 1 since the number of ^{14}C that remain (N) will always be less than the number at the beginning (N_0). The older the sample is, the smaller N/N_0 become. The natural logarithm of 1 is zero. So, the standard deviation of the age of a sample in which all the

expected ^{14}C was still present would be infinite. If the sample were fairly young (less than 10,000 years), the absolute value of $\ln(N/N_0)$ would be less than 1. (In standard deviations, everything is squared and then square-rooted, so all the negative signs go away. Plus, as was done in the derivation above, the value inside the natural logarithm can be changed to N_0/N by taking the negative sign inside the natural logarithm. But, this does not change the absolute value of the natural logarithm.) But when N/N_0 passes about 13000 years, the absolute value of the natural logarithm becomes greater than 1, so the relative standard deviation of the result gets smaller.

The relative standard deviation of N/N_0 gets larger with age, but the value by which it is divided gets larger also. Figure 2 shows the relationship between age and the resulting relative standard deviation for a 1-gram and an 8-gram sample of pure carbon. The mass of the original object would be at least twice that mass, since the percentage of carbon is rarely more than 50%. The curves turn up sharply at the right end because the ^{14}C becomes so depleted that there are not enough disintegrations to count. The curves turn up on the left because the ratio of N and N_0 is not different enough from 1 to be statistically significant. For the curve to reach 10,000 years, the carbon sample must be at least 0.27 g. The sample must weigh at least 0.9 g to reach 20,000 years, 3 g for 30,000 years, and more than the maximum of 8 grams to reach 38,000 years. This is based on the very best of conditions and equipment: 24-hour counting period, 0.2 cpm background, 100% efficiency, and a detection limit that is only three times the noise. Based on this graph, but not on the assumptions that produced it, ^{14}C dating is useful only between 5000 years and the upper limit imposed by the size of the sample and counting time.

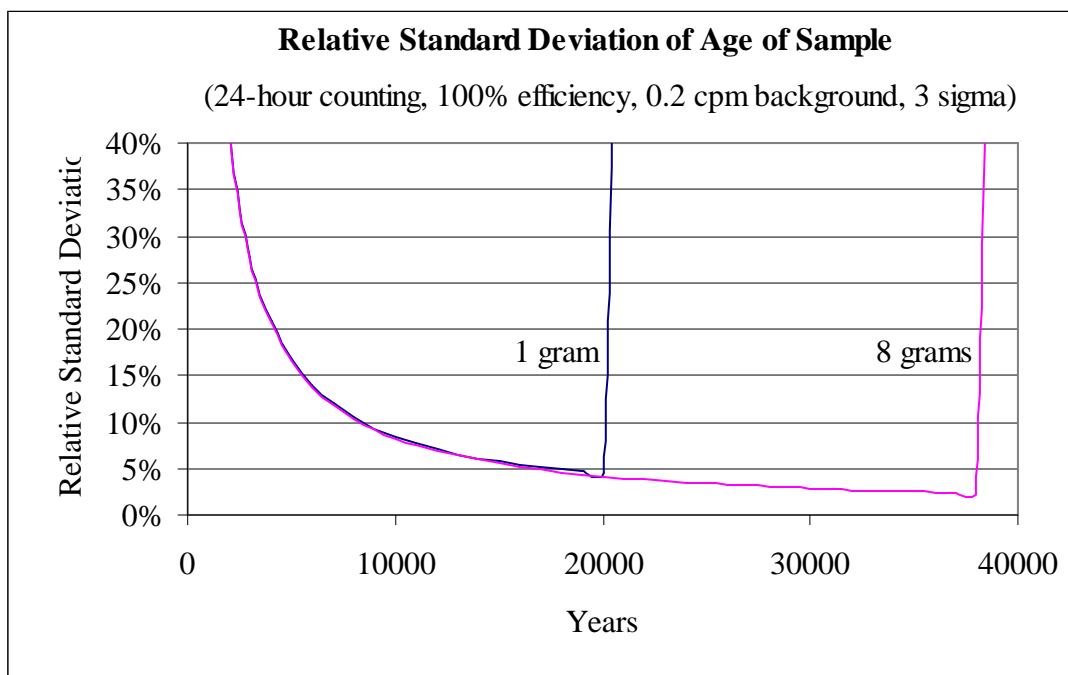


Figure 2. Relative standard deviation of the measurement of the age of 1 gram and 8 gram samples of pure carbon.

Figure 3 shows the same relative standard deviation comparison for potassium/argon data. Potassium is a common mineral in some rocks. A certain fraction of the potassium on earth is in the form of an unstable isotope, ^{40}K that decomposes to argon. The half-life is

1,260,000,000 years. The natural abundance is 0.117%. The method is not useful for ages that are less than 100 million years because N and N_0 are not sufficiently different. The same types of assumptions apply to potassium/argon dating that apply to ^{14}C dating.

- **Assumption 1:** The abundance of ^{40}K was the same during the period from 100 million to 600 million years ago as it is now.
- **Assumption 2:** The rate decay of ^{40}K has been constant over that time.
- **Assumption 3:** It is further assumed that these constant decay rates are constant according the formulas we have developed. These rates have been measured only for a little more than half a century.

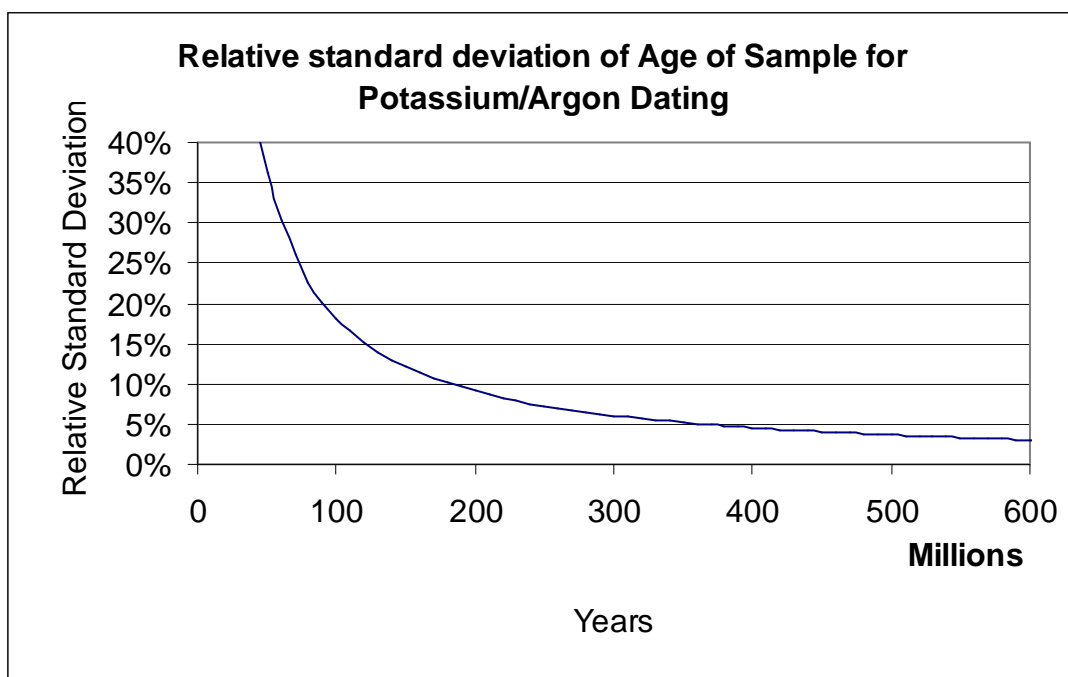


Figure 3. Relative standard deviations of the ages assigned by potassium/argon dating.

In summary, statistical analysis limits the range of ages that can be determined by current dating methods using the best equipment and the largest sample sizes for the longest counting times: 5000 to 38,000 years, and greater than 100 million years. Despite the limitations imposed by these methods, many ages of objects are published that fall outside those ranges.

The ^{14}C method assumes that the rate of neutron bombardment from the sun to the upper atmosphere has been the same for 60,000 years. But the rate of neutron bombardment sometimes changes on a scale of decades, so the assumption is not reasonable. If the quantity of neutrons has been increasing over time (as would be expected from a shrinking sun, with more neutrons able to escape from the nuclear reactions in the core because the outer layers of the sun are getting thinner), all ages determined by ^{14}C dating would yield results that indicated that the artifact was older than it really was.

The ^{40}K method assumes that the abundance of ^{40}K has been the same for more than 100 millions years and that the potassium in the rocks has not exchanged with the environment in that

same period of time. Yet, those who use such dating methods assume that mountain ranges are formed in similar time scales, and that continents rise from and sink into the molten interior of the earth on similar time scales. So, rocks cannot be assumed to have been isolated for that long. Interestingly, if time scales on earth were measurable across hundreds millions of years, the quantity of radioactive material on earth would be huge, based on the quantity that survives to this day.

Whenever an age based on ^{14}C is published, it should carry with it the size of the sample, the background rate, the efficiency, and the counting time. Without these data, the validity of the given age cannot be determined. Many samples are measured using only 0.1 g. With that size of sample and the very best equipment, the ^{14}C counts can never be separated from the background.