Uncertainties02A.wxmx

TABLE OF CONTENTS

1 Preface

In Uncertainties02A.wxmx we discuss a heuristic introduction to the standard deviation of the mean (SDOM).

Edwin L. (Ted) Woollett
https://home.csulb.edu/~woollett/
July 19, 2024

2 References

Joon Pahk, Error Analysis: An Introduction, Harvard Univ. Phys. 15a/16, Fall 2012 pdf,
https://studylib.net/doc/27026320/error-analysis---joon-pahk

Clemson Univ. Physics Tutorial: Measurement Uncertainty
https://science.clemson.edu/physics/labs/tutorials/uncertain/index.html

Harrison, David M., Error Analysis in Experimental Physical Science., Dept. of Physics, University of Toronto, 2004,
https://faraday.physics.utoronto.ca/PVB/Harrison/ErrorAnalysis/

[HH] Ifan Hughes, Thomas Hase, Uncertainties: A Practical guide to Modern Error Analysis, 2010, Oxford Univ Press.

R. J. Barlow, Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences (Manchester Physics Series), 1993, Wiley

Gerhard Bohm, Günter Zech,
Introduction to Statistics and Data Analysis for Physicists, 3rd revised edition, 2017,
available from CERN webpage:
https://s3.cern.ch/inspire-prod-files-d/da9d786a06bf64d703e5c6665929ca01

Frederick James, 'Statistical Methods in Experimental Physics', 2nd ed., 2006, World Scientific.

Louis Lyons, Statistics for Nuclear and Particle Physics, 1986, Cambridge Univ. Press,

Luca Lista, 'Statistical Methods for Data Analysis in Particle Physics', Lecture Notes in Physics 909, 2016, Springer-Verlag,

Glen Cowan , Statistical Data Analysis, Clarendon Press, Oxford, 1998

```
(%i6) load(draw)$
set_draw_defaults(line_width=2, grid = [2,2], point_type = filled_circle,
        head_type = 'nofilled, head_angle = 20, head_length = 0.5,
        background_color = light_gray, draw_realpart=false)$
load (descriptive)$  load (distrib)$
 fpprintprec : 5$ ratprint : false$
```

Homemade functions fll, head, tail, Lsum are useful for looking at long lists.

(%i10)
```
fll ( aL) := [ first (aL), last (aL), length (aL) ]$
head(L) := if listp (L) then rest (L, - (length (L) - 3) ) else
     error("Input to 'head' must be a list of expressions ")$
tail (L) := if listp (L) then rest (L, length (L) - 3 )  else
      error("Input to 'tail' must be a list of expressions ")$
Lsum (aList) := apply ("+", aList)$
```

3 fracData (List, a, b)

fracData (alist, a, b) calculates the fraction of the list numbers which lie in the interval [a, b].

(%i11)
```
fracData (myData, xx1, xx2) :=
block ([ ccnt : 0 ],
   for j thru length (myData) do
     if myData[j] >= xx1 and myData[j] <= xx2 then ccnt : ccnt + 1,
   float (ccnt/ length (myData)) )$
```

4 confidence (q, m, s)

With q a number in the interval 0 < q < 1, and with m the mean and s the standard deviation of a Normal distribution, confidence (q, m, s) prints out the values dx, m - dx, m + dx, and outputs a list [m - dx, m + dx] which allows one to have 100*q % confidence a random value of x will lie within that interval, ie., within m +/- dx.

(%i12)
```
confidence (qq, mm, ss) :=
block ([ddx],
   ddx :
     float ( (quantile_normal (qq + (1 - qq)/2, mm, ss) -
                  quantile_normal ( (1 - qq)/2  , mm, ss))/2 ),
   print ( "delx = ", ddx ),
   print ( " x1 = ", mm - ddx, ",  x2 = ", mm + ddx ),
   [mm - ddx, mm + ddx])$
```

5 What is Error?

Quoting Joon Pahk:

"First, what do we even mean when we talk about error in a lab setting? Let's begin to answer that by stating what we don't mean by "error": that the experimenter made a mistake. While that's always a possibility, the technical use of error in this course really refers to the inevitable uncertainty that accompanies any measurement. You cannot eliminate error simply by being careful. That's a key idea, so we'll put it in a big highlighted box:"

"Error refers to the inevitable uncertainty that accompanies any measurement."

"Why is this uncertainty inevitable? The basic reason is because any real, physical measuring device has limitations on its precision. In a theoretical context, we can say things like "the stick is 75 centimeters long" and understand that to mean that the stick has a length of exactly 75 cm; no more, no less. But in an experimental setting, the point is to make measurements of physical quantities. We can measure the length of the stick with a ruler, but maybe the ruler is only marked to the nearest centimeter. "But," you object, "we can just get a better ruler." Fine, we'll get a ruler marked to the nearest millimeter. But that merely lessens the problem; it does not eliminate it. We can improve the measuring device as much as we want (there exist laser interferometers which can measure lengths to incredible precisions), but no matter what we do, we can never achieve infinite precision in a real experimental setting."

"So instead, we do the next-best thing: figure out how much precision we can achieve, and keep track of it. If you understand errors properly, you can determine how much precision you need and how to obtain that precision. That way, you can avoid the embarrassment of being error-ignorant. Imagine a researcher writing a grant proposal for a super-expensive laser interferometer to measure the length of a stick! Depending on what experiment is being done with the stick, it may be entirely sufficient to know its length to the nearest millimeter, or centimeter, or even "well, it's a bit less than a meter."

6 Truth? What is Truth?

Quoting Joon Pahk:

"Throughout the last section, we have been nonchalantly referring to a quantity we call 'the length of the stick.' But the entire point of the section was that there is no way to actually pin down the length of the stick; some amount of uncertainty in the length is absolutely inevitable. You can only make closer and closer approximations as you improve the quality of your measurement apparatus. So does it even make sense to talk about the stick as having a single, well-defined length?"

"The answer is a qualified “yes.” In general, we make the assumption that any quantity we are measuring does have a “true” value, and that our measurements represent guesses as to what this true value might be. This view is largely validated by the results of countless real experiments, which have shown that if performed carefully, the results of a physical experiment are at least repeatable; that is, they give consistent results if performed again, by other experimenters in other locations. So even though it is an assumption, it’s quite probably a very good one, and one that we’ll make henceforth without batting an eyelash. The stick does have a true length, and we can perform experiments to try to measure it"

"One caveat to the assumption of repeatability is that if your measuring apparatus is really precise, you may run into problems of definition. For instance, if you did spring for that fancy-schmancy laser interferometer to measure the length of the stick, you’d find that upon closer examination, the ends of the stick are bumpy, rather than sharply cut; and the length of the stick changes slightly with temperature, and whether you hold it vertically or horizontally; and so on. It gets harder and harder to specify exactly what we mean by “the length of the stick” under these conditions. In these cases, the precision of the measurement can be limited by the definition of the quantity being measured, rather than by the resolution of the measuring apparatus."

"Another important exception will arise when you study quantum mechanics, where you’ll discover that it is technically impossible to make any measurement of a quantity in a physical system without fundamentally altering the system itself. But we won’t have to worry about that for quite some time."

"While we’re on the subject of truth, we should point out that an experiment is correct if it has been performed and analyzed correctly. Even though we assume that there is a 'true' value to a quantity we are measuring, the error in an experimentally measured quantity is never found by comparing the measured value to some number found in a book or other reference. If we tell you to measure the gravitational acceleration at the Earth’s surface, and your result does not agree with the 9.8 m/s^2 that you are expecting to get, this discrepancy does not mean that your result is wrong. (Of course, it could be wrong; maybe you made a mistake somewhere. But you should not assume it is wrong merely because it is unexpected.)"

7 Uncertainty of a Measuring Device: Rule of Thumb

Quoting Clemson Univ. Prelab.

"Each instrument has an inherent amount of uncertainty in its measurement. Even the most precise measuring device cannot give the actual value because to do so would require an infinitely precise instrument. A measure of the precision of an instrument is given by its uncertainty. As a good rule of thumb, the uncertainty of a measuring device is 20% of the least count. Recall that the least count is the smallest subdivision given on the measuring device. The uncertainty of the measurement should be given with the actual measurement, for example, 41.64 ± 0.02 cm."

"Here are some typical uncertainties of various laboratory instruments:

Meter stick: ± 0.02cm
Vernier caliper: ± 0.01cm
Triple-beam balance: ± 0.02g
Graduated cylinder: 20% of the least count"

"Here's an example. The uncertainty of all measurements made with a meter stick whose smallest division (or least count) is one millimeter is 20% of 1mm or 0.02cm. Say you use that meter stick to measure a metal rod and find that the rod is between 10.2 cm and 10.3cm. You may think that the rod is closer to 10.2cm than it is to 10.3cm, so you make your best guess that the rod is 10.23cm in length. Since the uncertainty in the measurement is 0.02cm, you would report the length of the metal rod to be 10.23 ± 0.02cm (0.1023 ± 0.0002 m)."

8 *Sample and Parent Distribution*

[HH] Sec. 2.6

"In the theory of statistics, the parent distribution refers to the number of possible measured values, ξi ; the parent population might consist of an infinite number of values. Two independent parameters, the mean, μ, and a standard deviation, σ_parent, characterize the parent distribution, and are related thus:"

$$\sigma_{parent} = [\Sigma (\xi_i - \mu)^2 / N_{parent}]^{(1/2)}$$

"In practice when we take a series of measurements x_i in an experiment, we take a selection, or sample, from this parent distribution which results in a distribution called the sample distribution. This distribution is centred on the mean of the data set, <x>, and has a standard deviation:"

$$\sigma_{sample} = [\Sigma (x_i - \langle x \rangle)^2 / (N - 1)]^{(1/2)}$$

"The (N − 1) is required in the denominator because the mean, <x>, is also determined from the same data set and is thus no longer independently determined and the number of degrees of freedom is one fewer."

"As we repeatedly sample the parent distribution we slowly build up a distribution of values centred on the mean of the sample distribution,

<x> = (Σx_i) / N

which becomes an increasingly better approximation of μ as N increases. As all the measurements sample the same parent distribution they are all determined with the same precision as the parent distribution,

σ_sample ≈ σ_parent.

As more data are recorded the standard deviation of the data does not change, it simply becomes better defined."

8.1 Standard Error of the Mean σE, SDOM, Standard Error

We use Maxima to show the evolution of the mean and standard deviation of a sample distribution with sample size, N. The parent distribution was randomly generated from a Gaussian distribution with mean μ = 10, and standard deviation σ_parent = 1.

```
(%i14) print (" N", "<x>", "σ_sample")$
for N in [5, 10, 50, 100, 1000, 2500] do (
   sample : random_normal (10, 1, N),
   print (N, mean (sample), std1 (sample) ) )$
```

```
 N  <x>  σ_sample
5  10.175  1.5157
10  10.468  0.72664
50  10.223  0.97827
100  9.9793  1.0907
1000  9.9704  1.0116
2500  10.039  1.0108
```

From this example we see that as N gets larger the mean of the sample becomes better defined, but the standard deviation of the sample hardly changes.

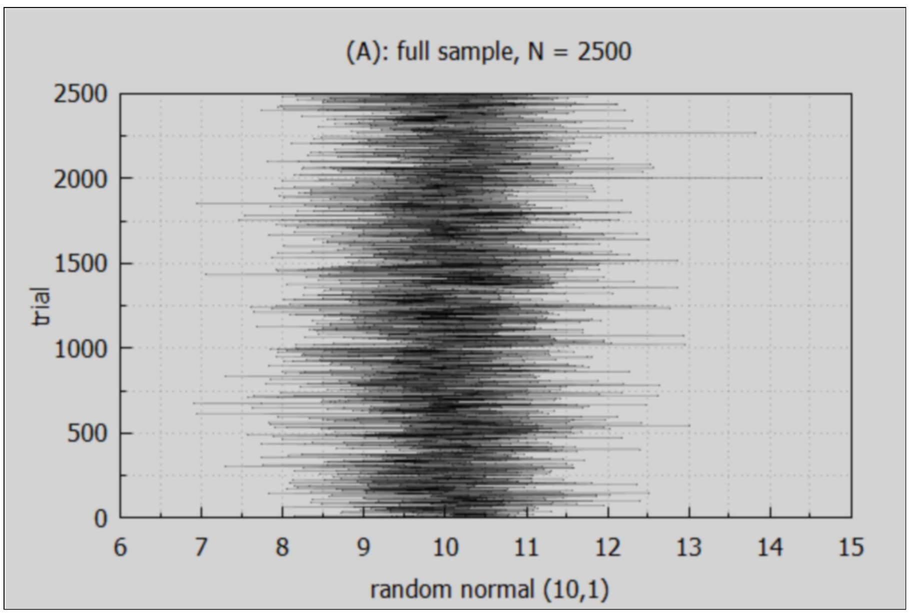
As we increase the sample size, the standard deviation of the sample does not get smaller; "thus the standard deviation of the sample is not a good measure of the error in the estimation of the mean of the parent population. We can clearly determine the position of the mean to a better precision than the standard deviation of the sample population. The important concept here is that of signal-to-noise: the precision with which we can determine the mean depends on the number of samples of the parent distribution."

Below we show "histograms of a simulation where 2500 points are chosen from a normal parent distribution with mean μ = 10 and standard deviation σ_parent = 1. Histogram (A) of the measurements is normal, centred as expected on 10, and has a standard deviation of σ_sample = 1. The upper plot of the sample points shows that most measurements are within two standard deviations of the mean of the parent distribution, with very few points with a deviation larger in magnitude than two standard deviations."

```
(%i18) sample : random_normal (10, 1, 2500)$
print ("mean = ",mean (sample),"  std1 = ",std1 (sample) )$
xypoints : makelist ([sample[j], j ], j, 1, 2500)$
wxdraw2d ( xrange = [6,15], yrange = [0, 2500], color = black,
    xlabel = " random normal (10,1)", ylabel = "trial",
   title = " (A): full sample, N = 2500",point_size = 0.1, line_width = 0.1,
    points_joined = true,  points (xypoints)  )$
```

mean = 10.022 *std1* = 0.99334

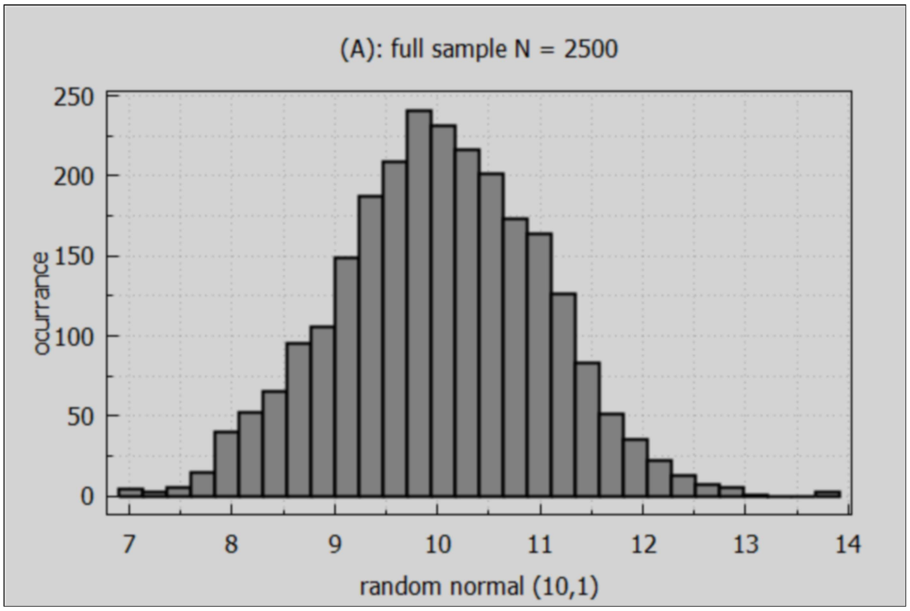
(%t18)



(%i20)
```
print ("mean = ",mean (sample)," std1 = ",std1 (sample) )$
wxhistogram (sample, xlabel = "random normal (10,1)", ylabel = "ocurrance",
 title = " (A): full sample N = 2500 ", fill_color = black,
 fill_density = 0.5 /* , frequency = density */ , nclasses = 30 )$
```

mean = 10.022 *std1* = 0.99334

(%t20)



"For histogram (B) the same data set was partitioned differently. The mean of every five points was calculated, yielding 500 mean values; histogram (B) shows the distribution of these means. It is evident from the width of the histogram that the distribution of these 500 means is significantly narrower than the distribution of the unaveraged data. Averaging five data points greatly reduces the statistical fluctuations in the distribution of the means."

(%i21)
```
fll (sample);
```

(%o21) [10.404,9.256,2500]

We are taking a running average, five values at a time.

(%i22)
```
for j thru 10 do print (j, sample[j])$
```

1 10.404
2 12.225
3 11.308
4 9.8532
5 10.143
6 11.842
7 9.0699
8 9.3968
9 9.6041
10 9.49

(%i25)
```
ss5 : []$
for j thru 500 do (
   ss : [],
  for k: 5*j - 4 thru 5*j do ss : cons (sample[k], ss),
   ss5 : cons (mean (ss), ss5) )$
ss5 : reverse (ss5)$
```

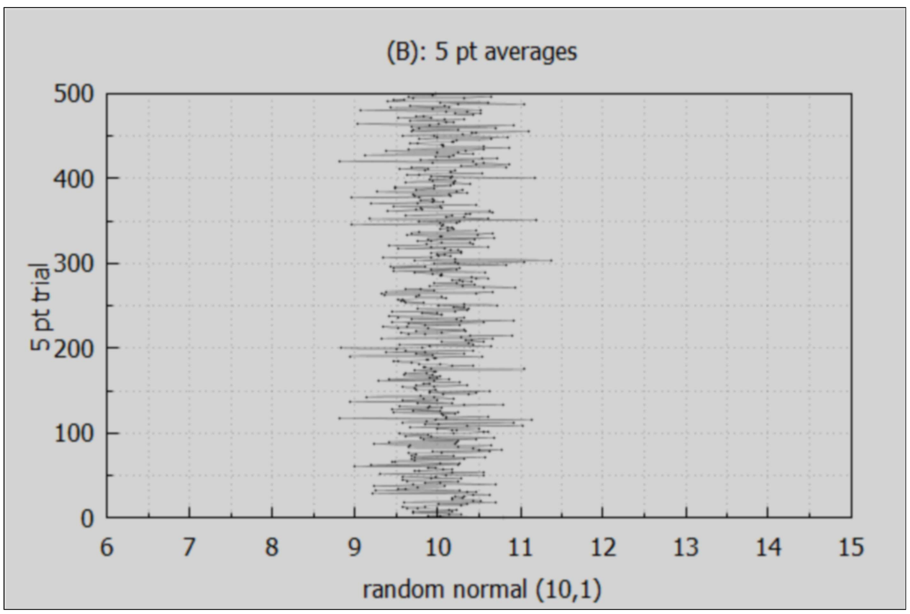
(%i26)
```
fll (ss5);
```

(%o26) [10.787,9.9756,500]

(%i29)
```
print ("mean = ",mean (ss5),"  std1 = ",std1 (ss5) )$
xypoints : makelist ([ss5[j], j ], j, 1, 500)$
wxdraw2d ( xrange = [6,15], yrange = [0, 500], color = black,
     xlabel = " random normal (10,1)", ylabel = "5 pt trial",
   title = " (B): 5 pt averages",point_size = 0.2, line_width = 0.2,
    points_joined = true,  points (xypoints)  )$
```

mean = 10.022 *std1* = 0.42926

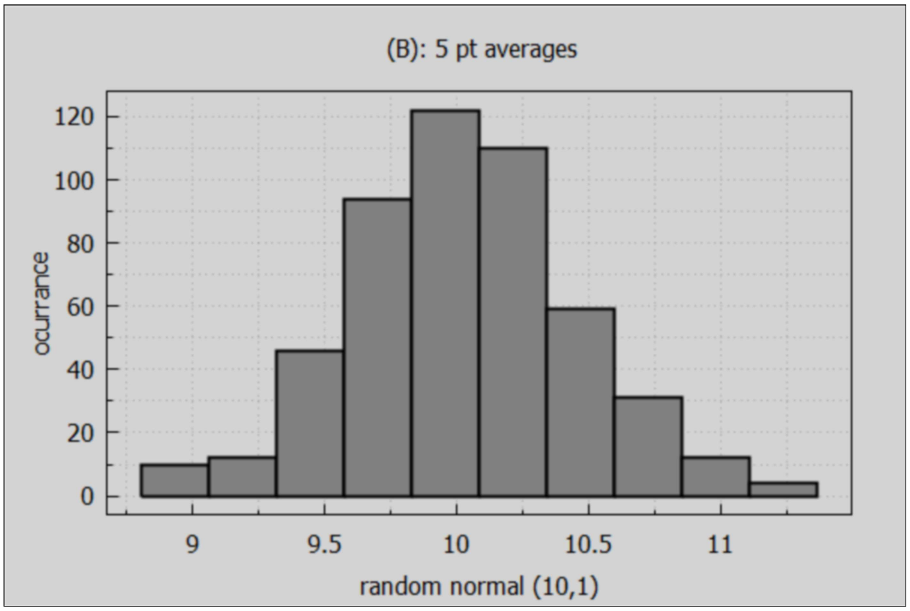
(%t29)



(%i31)
```
print ("mean = ",mean (ss5),"  std1 = ",std1 (ss5) )$
wxhistogram (ss5, xlabel = "random normal (10,1)", ylabel = "ocurrance",
 title = " (B): 5 pt averages", fill_color = black,
 fill_density = 0.5  /* , frequency = density, nclasses = 30 */  )$
```

mean =   10.022   *std1* =   0.42926

(%t31)



"Averaging five data points greatly reduces the statistical fluctuations in the distribution of the means. This trend continues in case (C) where the data set was partitioned into 250 measurements of means of 10 points per measurement."

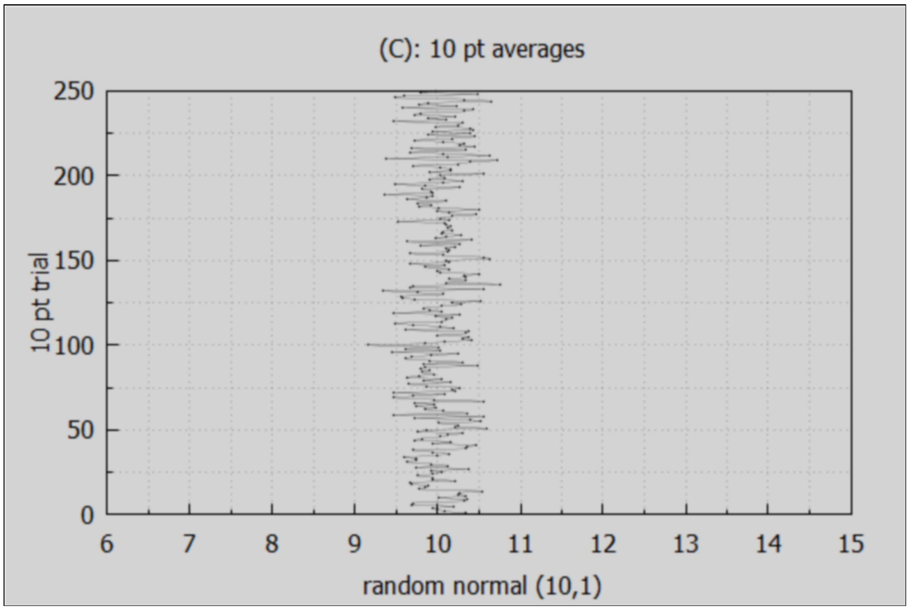
(%i35)
```
ss10 : []$
for j thru 250 do (
   ss : [],
  for k: 10*j - 9 thru 10*j do ss : cons (sample[k], ss),
   ss10 : cons (mean (ss), ss10) )$
ss10 : reverse (ss10)$
fll (ss10);
```

(%o35) [10.334,9.9704,250]

(%i38)
```
print ("mean = ",mean (ss10),"  std1 = ",std1 (ss10) )$
xypoints : makelist ([ss10[j], j ], j, 1, 250)$
wxdraw2d ( xrange = [6,15], yrange = [0, 250], color = black,
     xlabel = " random normal (10,1)", ylabel = "10 pt trial",
   title = " (C): 10 pt averages",point_size = 0.2, line_width = 0.2,
    points_joined = true,  points (xypoints)  )$
```

mean = 10.022 *std1* = 0.301

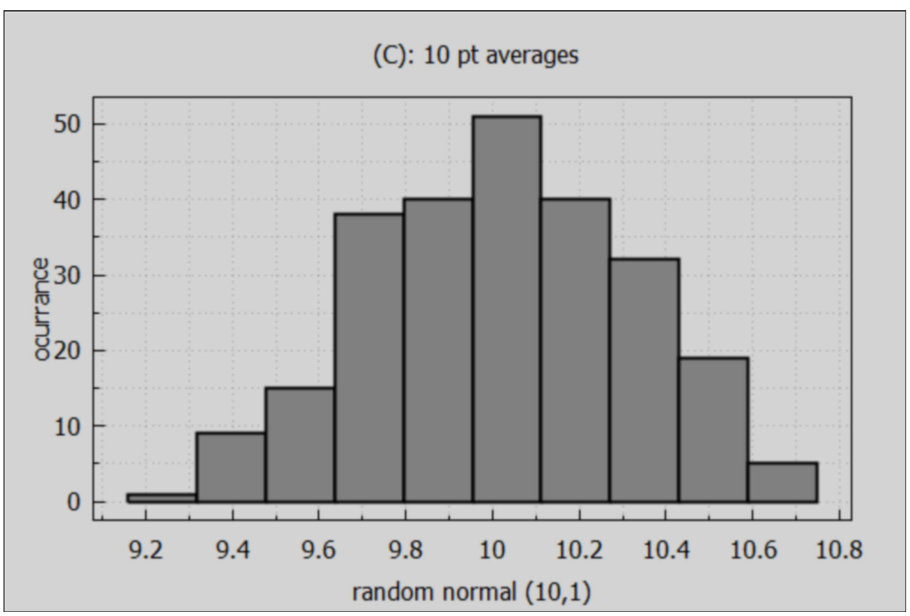
(%t38)



(%i40)
```
print ("mean = ",mean (ss10),"  std1 = ",std1 (ss10) )$
wxhistogram (ss10, xlabel = "random normal (10,1)", ylabel = "ocurrance",
 title = " (C): 10 pt averages", fill_color = black,
 fill_density = 0.5  /* , frequency = density, nclasses = 30 */  )$
```

mean = 10.022 *std1* = 0.301

(%t40)



In plots (D), 50 measurements of means obtained from 50 data points are plotted.

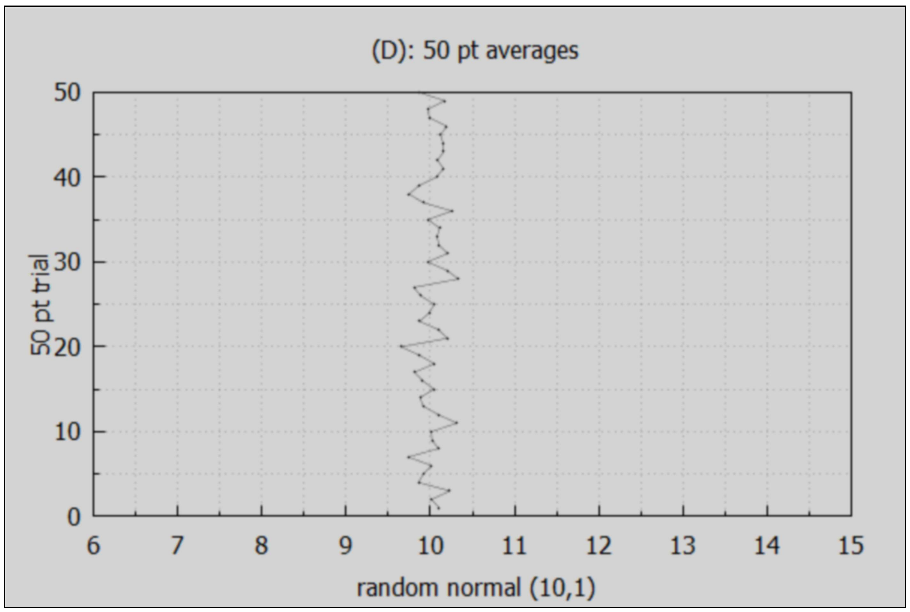
(%i44)
```
ss50 : []$
for j thru 50 do (
   ss : [],
  for k: 50*j - 49 thru 50*j do ss : cons (sample[k], ss),
   ss50 : cons (mean (ss), ss50) )$
ss50 : reverse (ss50)$
fll (ss50);
```

(%o44) [10.107,9.8622,50]

(%i47)
```
print ("mean = ",mean (ss50),"  std1 = ",std1 (ss50) )$
xypoints : makelist ([ss50[j], j ], j, 1, 50)$
wxdraw2d ( xrange = [6,15], yrange = [0, 50], color = black,
     xlabel = " random normal (10,1)", ylabel = "50 pt trial",
   title = " (D): 50 pt averages",point_size = 0.2, line_width = 0.3,
    points_joined = true,  points (xypoints)  )$
```

mean = 10.022 *std1* = 0.15295

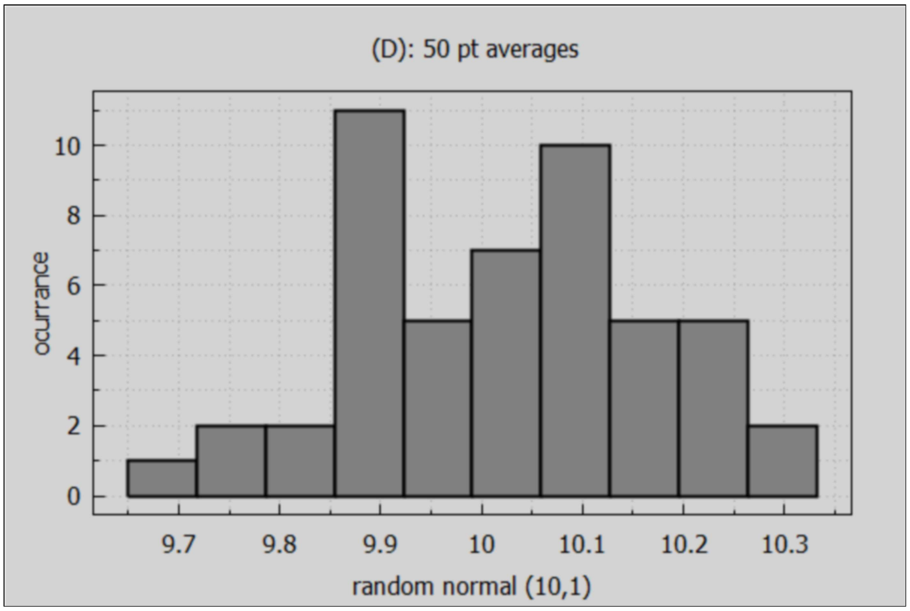
(%t47)



(%i49)
```
print ("mean = ",mean (ss50),"  std1 = ",std1 (ss50) )$
wxhistogram (ss50, xlabel = "random normal (10,1)", ylabel = "ocurrance",
 title = " (D): 50 pt averages", fill_color = black,
 fill_density = 0.5  /* , frequency = density, nclasses = 30 */  )$
```

mean = 10.022 *std1* = 0.15295

(%t49)



"The width of the histogram of means is a measure of the precision of the mean. It is clearly evident from the above plots that the width of the histogram of the means decreases as the size of the sample used to calculate the mean increases—this is a consequence of averaging over statistical fluctuations."

"The width of the histogram of means is the standard deviation of the mean, also known as the standard error, σE."

"When the number of measurements involved in calculating the mean increases, the means are better defined; consequently the histograms of the distributions of the means are narrower. Note that the precision with which the mean can be determined is related to the number of measurements used to calculate the mean."

"In practice one does not generate histograms of the means based on trials containing many measurements; rather one uses all N measurements x_i to calculate one value for the mean <x>. We expect the standard error (the standard deviation of the mean), σE, to decrease as the number of data points we collect, N, increases."

It can be shown "that the standard error is reduced by a factor of √N with respect to the sample standard deviation:"

σE = σ_sample / N^½

"A data set containing N multiple readings yields one value of the mean. Thus we should quote our findings as the mean ± the error of the mean, i.e."

<x> +/- σE == <x> +/- σ_sample / N^½

"In other words, we are saying that there is a two-thirds chance that the measured parameter is within the range [<x> - σE , <x> + σE]."

"One can interpret the standard error as being a standard deviation, not of the measurements, but rather the mean: this is why the standard error σE is also called the standard deviation of the mean (SDOM)."