

# Re: Another Error in Afonso's Confidence Interval Code

---

*Source:* <http://sci.tech-archive.net/Archive/sci.stat.math/2007-04/msg00069.html>

---

- *From:* Jack Tomskey <[jtomsky@xxxxxxxxxxxxxxx](mailto:jtomsky@xxxxxxxxxxxxxxx)>
  - *Date:* Mon, 02 Apr 2007 16:03:37 EDT
- 

Let's try again to get Afonso to explain his Basic code for simulating confidence intervals. The complete code copied directly from an earlier post is posted down below the line of &&&&&&&&&.

The error I have cited (which he refuses to address) is shown here. I have added <--- with my comments. This is copied directly from Afonso's code (see the entire code posted further down.)

```
FOR i = 1 TO n
a = SQR(-2 * LOG(RND))
x(i) = a * COS(2 * pi * RND) <----- 1
md = md + x(i) / n <----- 2
soma2 = soma2 + x(i) * x(i) <----- 3
NEXT i
sqd = soma2 - n * (nmd) ^ 2 <----- 4
: sd = SQR(sqd / n) <----- 5
FOR ii = 1 TO n
x(ii) = (x(ii) - md) / sd <----- 6
NEXT ii
```

1. x(i) is random normal numbers; simulated data. OK
  2. md is the average of n random normal numbers. OK
  3. soma2 is the raw sum of squares of those numbers (square, then sum) OK
  4. This seems to be an attempt to go from the raw sum of squares to the usual corrected sum of squares. However, since nmd was never defined anywhere above then nmd is always "0". Thus n \* (nmd)^2 is "zero" for each and every sample of n data. This is the error
- I have tried repeatedly to get Afonso to address, but he has not.

## Re: Another Error in Afonso's Confidence Interval Code

5.  $sd = \text{SQR}(sqd/n)$  is an attempt to calculate the usual sample standard deviation. But since  $sqd$  is the raw sum of squares, this is not a meaningful calculation. In addition, there is a second error here. (I hoped Afonso would find this and correct it, but he has not.) He should divide the corrected sum of squares by  $n - 1$ , not by  $n$ , to get an unbiased estimate of sigma.

So now we have two errors for Afonso to address. Based on the cockroach theory (when we find one cockroach, that means there are more) I'll bet there are more errors beyond these two.

6. Here he is "standardizing" the data by dividing the individual data by subtracting the average and dividing by  $sd$ . But  $sd$  is not the usual unbiased estimate of sigma.

The failure to reduce the raw sum of squares to the corrected sum of squares is a major coding error. (Or is it a conceptual error? Afonso will tell us soon.) So  $sd$  is simply incorrect throughout. Hence the standardized values of  $x(i)$  are also incorrect.

Dividing by  $n$  instead of  $n - 1$  to obtain an unbiased estimate of sigma is another error. This one seems to be conceptual, but perhaps we'll learn more later.

Here's yet another problem...

Attempts to get this software code to produce the numerical results that Afonso cited earlier have failed.

This is an opportunity for Afonso to explain his software code in detail... step-by-step... so we can all follow what he is



## Re: Another Error in Afonso's Confidence Interval Code

```
IF x(jj) < u THEN w = w + 1
NEXT jj: xx(w) = u
NEXT ii
FOR tt = 1 TO n: z = xx(tt)
REM calcula FI(z)
IF z > 0 THEN kw = 0
IF z < 0 THEN kw = 1
zu = ABS(z): s = c * zu: antes = c * zu
FOR j = 0 TO 1000
xx = antes * fng(zu, j)
s = s + xx
antes = xx
IF ABS(xx) < .00005 THEN GOTO 20
NEXT j
20 IF kw = 0 THEN ff = .5 + s
IF kw = 1 THEN ff = .5 - s
b = ABS(ff - F(tt - 1))
bb = ABS(F(tt) - ff)
maior = b
IF bb > b THEN maior = bb
IF maior > mmaior THEN mmaior = maior
NEXT tt
mm = INT(1000 * mmaior + .5)
IF mm > 9000 THEN mm = 9000
w(mm) = w(mm) + 1
fff = INT(k / 50000): ff = k / 50000
IF ff <> fff THEN GOTO 1000
cc(1) = .95 * k: cc(2) = .99 * k
FOR iji = 1 TO 2
ciji = cc(iji): s = 0
FOR iij = 0 TO 9000
s = s + w(iij)
IF s > cijj THEN GOTO 100
NEXT iij
100 PRINT USING "###.### #.#### ";
iij / 1000; s / k
NEXT iji
1000 NEXT k
END
```

OMU, what Afonso is doing is generating a large number of samples of  $n$   $X_i$  from  $N(0,1)$ . He then tries to replicate what a user of Lilliefors would do. He calculates the sample mean and standard deviation, fits a normal cdf, and then finds the distance between the empirical cdf and the fitted normal cdf as the maximum absolute difference.

Now, you've found an error in Afonso's code in calculating the sample standard deviation. Apparently he meant  $md$  instead of  $nmd$ . Along with dividing by  $n$  instead of  $n-1$ , the effect is that his formula for the variance becomes  $\text{Sumsq}(X_i)/n$  rather than  $\text{Sumsq}(X_i - \bar{x})/(n-1)$ .

Re: Another Error in Afonso's Confidence Interval Code

His actual  $\text{Sumsq}(X_i)/n$  is an unbiased estimate of the population  $\sigma^2$ , as is also  $\text{Sumsq}(X_i - \bar{x})/(n-1)$ . The difference is that he has  $n$  degrees of freedom instead of  $n-1$  degrees of freedom for the estimated standard deviation. The user would not know the population mean in advance and thus has only  $n-1$  degrees of freedom in the estimate.

This would have a very small effect in his tables. His tables should give slightly smaller erroneous critical values for the Lilliefors statistic, maybe in the third or fourth decimal. However, since he is comparing his simulation results with those of 40 years ago, we are looking at those small differences.

Jack

.