Cross Validated

what is the sample correlation between the mean and the standard deviation under normality assumption

Asked 5 years, 2 months ago Modified 5 years, 2 months ago Viewed 2k times

2

It is known that for the normal distribution, the mean and the standard deviation are independent; it is also the case for samples as sample size tends to infinity.

However, for small samples, the sample mean and the sample standard deviation are not independent. Indeed, the sample mean is used to compute the sample standard deviation.

As an example, simulating 100,000 samples of size 1000, and computing for each means and standard deviations, I get a correlation of 0.0005 between the 2 statistics. For 1,000,000 samples of size 3, I get a negative correlation of -0.0003.

Is there a way to find the expected correlation as a function of sample size? Its distribution? Or is it zero on average?

correlation descriptive-statistics expected-value small-sample

Share Cite Edit Follow

asked Nov 11, 2017 at 20:01
Denis Cousineau
170 1 15

1 In what sense are you attempting to distinguish a "normal distribution" from samples? Can you help us understand what it might mean for two *parameters* of a distribution, such as its mean and SD, to be "independent"? Could you explain why you think the independence of sample mean and SD depends on sample size? At what size do these statistics become independent? What are the standard errors on the means from your simulations--that is, do you have any significant evidence at all that the correlation differs from zero? – whuber ♦ Nov 11, 2017 at 20:19

1 Answer

Sorted by: Highest score (default)

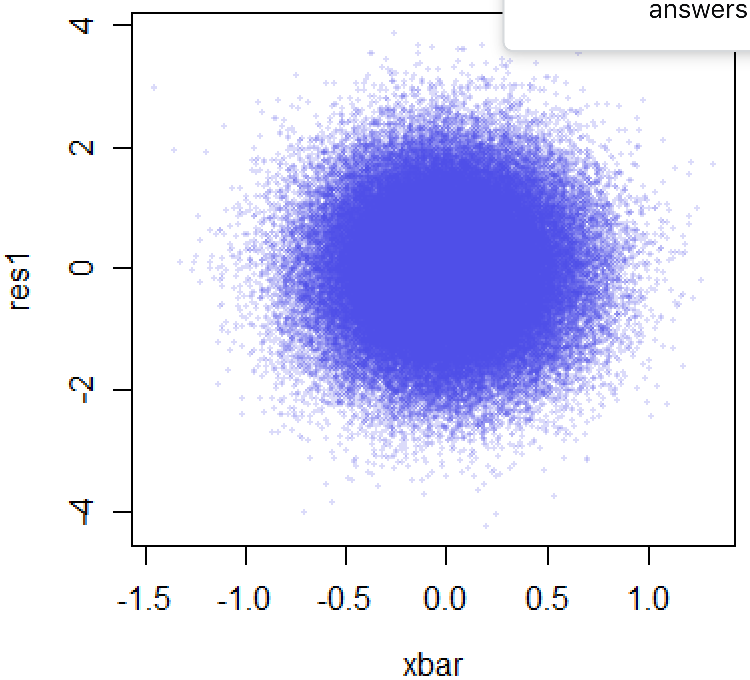
2

While the (usual) formula for sample standard deviation contains the sample mean (so you might say that it depends on the mean algebraically), that doesn't impact their independence in the statistical sense.

The population correlation of the sample mean and the sample standard deviation is zero because the random variables (sample mean, sample standard deviation) are independent.

Indeed $\bar{X}$ and $X_i - \bar{X}$ are independent and the independence of the sample variance (and similarly standard deviation) from $\bar{X}$ follows from that.

Below is a plot of $x_1 - \bar{x}$ vs $\bar{x}$ for sam mal distribution; it's possible to show these are independent (and inde



The *sample* correlation between the random variables won't be exactly zero -- it, too, is a random variable. If you gather m samples of size n and compute the mean and sd in each sample of size n and calculate their correlation over the m samples, that will have some deviation from 0 each time, but the population quantity it estimates is 0. Imagine choosing say 50 points in the plot above for example; the sample correlation of those 50 points won't be exactly 0.

Share Cite Edit Follow

edited Nov 11, 2017 at 22:14

answered Nov 11, 2017 at 22:02

Glen_b

268k 34 582 977

Ok, thanks. I should have looked at the standard error. It was indeed of the same magnitude as the estimates I obtained. – Denis Cousineau Nov 12, 2017 at 1:47