International Journal of Applied Mathematics

Volume 31 No. 3 2018, 359-370

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

doi: http://dx.doi.org/10.12732/ijam.v31i3.5

ON THE MEAN OF SAMPLE-STANDARD-DEVIATION

Gregory L. Light

Department of Finance
Providence College
1 Cunningham Square, Providence, Rhode Island 02918, USA

Abstract: This paper studies the mean of the probability distribution of the sample-standard-deviation for all populations, numerically by examples in the main, but followed by analytical summary observations. We find that in applications for sample sizes greater than 20 one is likely to achieve ((E(s))/sigma) greater 0.9 and that the deciding factor of ((E(s))/sigma) is the concentration of mass in the population: a higher concentration of data in the population leads to a smaller ((E(s))/sigma) than otherwise.

AMS Subject Classification: 62F35, 62F10, 62F35

Key Words: biasedness of sample standard deviation, population and sample standard deviation, variance of sample standard deviation, mean of transformed distribution

1. Introduction

This paper is motivated by our teaching statistical inferences about the population standard deviation σ for undergraduate students (cf. [1], [4], [5]). The common rendition is to use the sample-variance s^2 to estimate σ^2 by the fact that the mathematical expectation $E\left(s^2\right) = \sigma^2$, thereby s to estimate σ, an approach that is so universally adopted that few studies exclusively on s have been found by our extensive literature research (see [4]). While we the instruc-

Received: March 19, 2018

tors may state that the distribution of s^2, being an unbiased estimator of σ^2, is centered at σ^2 for mathematical correctness, we find it expedient to say simply that s is centered around σ. The convenience here arises from the fact that in applications more often than not the intended estimation target is σ, not σ^2, which can carry units with no practical meanings. Nevertheless $E(s) \neq \sigma$ in general. In fact,

$$Var(s) + (E(s))^2 \equiv E\left(s^2\right) = \sigma^2 \equiv Var(X), \tag{1}$$

implying that

$$E(s) \leq \sigma,$$

with the two limiting cases:

$$(1) \lim_{n\to\infty} \frac{E(s)}{\sigma} = 1,$$

where $n \equiv$ sample size with replacement (assumed throughout this paper unless explicitly stated otherwise) for the simple reason that $\lim_{n\to\infty} s = \sigma$;

$$(2) \frac{E(s)}{\sigma} \to 0,$$

as demonstrated by the following simple example: Let $P(X=0) = \frac{m-1}{m} = 1 - P(X=2)$, $m \in \mathbb{N} - \{1\}$; then $E(x) = \frac{2}{m}$ and $\sigma^2 \equiv Var(X) = \frac{4(m-1)}{m^2}$, so that assuming $n = 2$ we have $P(s=0) = \left(\frac{m-1}{m}\right)^2 + \frac{1}{m^2}$ and $P\left(s = \sqrt{2}\right) = \frac{2(m-1)}{m^2}$; hence $E(s) = \frac{2\sqrt{2}(m-1)}{m^2}$ and $\frac{E(s)}{\sigma} = \frac{\sqrt{2(m-1)}}{m} \to 0$ as $m \to \infty$.

In the next Section 2 we will examine the ratio $\frac{E(s)}{\sigma}$ as distinguished by two cases: normally and non-normally distributed populations. Finally, in Section 3 we will conclude with some observations.

2. Analysis

2.1. Population Normally Distributed

In this case the ratio $\frac{(n-1)s^2}{\sigma^2}$ is well-known to have the chi-square χ^2 distribution with degrees of freedom $d.f. = n-1$. We consider a division of the area under the χ^2 density curve into ten equal parts of 0.1 each. Next we obtain the following ten percentiles: $5\%, 15\%, \cdots, 95\%$. The idea here is to represent each of these ten areas by its respective representative point on the χ^2 axis;

ideally we ought to identify the center of mass of each of these ten parts by the mean, but for expediency we will use the ten medians. We understand that *mean* $<$ *median* if the density curve is rising and *median* $<$ *mean* if the density curve is falling. Accordingly for low degrees of freedom, where the χ^2 density curves are highly right-skewed, our procedure of using the median will under-approximate $E(s)$. Otherwise as the degrees of freedom increase, the χ^2 density curves become more and more symmetric so that the under-approximations and the over-approximations of using the ten medians for the ten means in our procedure tend to cancel out.

Then to recover $\frac{s}{\sigma}$ from $\chi^2 = \frac{(n-1)s^2}{\sigma^2}$, we invert to have

$$\frac{s}{\sigma} = \sqrt{\frac{\chi^2}{n-1}};$$

obtaining the above-said ten medians for χ^2, for $n = 2$ we have their square-roots: {0.0627, 0.1891, 0.3186, 0.4538, 0.5978, 0.7554, 0.9346, 1.1503, 1.4395, 1.9600}, each representing a probability of 0.1; thus

$$\frac{E(s)}{\sigma} \gtrsim \frac{0.0627 + 0.1891 + \cdots + 1.9600}{10} = 0.7862.$$

We now present the following table of $\frac{E(s)}{\sigma}$ as a function of *d.f.*:

$d.f.$	$\frac{E(s)}{\sigma} \approx$
1	0.7862
2	0.8799
3	0.9172
4	0.9370
5	0.9492
6	0.9575
7	0.9634
8	0.9679
9	**0.9714**
10	0.9743
11	0.9766
12	0.9785
13	0.9802
14	0.9816
15	0.9828
16	0.9838
17	0.9848
18	0.9856
19	0.9864
20	0.9871
100	0.9974

From the above table we see that for a population that is normally distributed: (1) the smallest ratio of $\frac{E(s)}{\sigma} \approx 0.8$ at $n = 2$, (2) once $n \geq 6$, $\frac{E(s)}{\sigma} \gtrsim 0.95$, and (3) $\forall n \geq 100$, $E(s) \lesssim \sigma$.

It is interesting to cite a real classroom experiment that gathered 75 sample-standard-deviations with sample size $n = 10$ without replacement from $N \approx 4000$ students' heights on our campus composed of men and women, with the original data (in inches) provided below:

3.5	4.6	4.4	3.8	2.8
4.5	1.3	3.6	3.5	2.4
2.4	4.1	1.6	2.2	5.1
3.9	4.2	4.8	4.2	3.4
3.8	4.7	3.5	4.6	3.6
5.2	4.2	4.8	3.3	4.6
3.5	3.7	5.0	4.8	3.3
4.1	3.4	4.5	3.3	5.1
3.2	3.9	3.5	3.5	3.4
2.7	1.2	3.2	3.7	2.7
3.4	4.3	4.5	3.7	4.7
4.4	4.2	2.7	1.6	5.2
3.3	2.7	4.3	3.2	3.5
3.1	3.3	4.6	3.6	4.6
3.9	4.1	3.4	3.2	3.5

with frequency distribution

bin	*frequency*
$(0.5, 1.5]$	2
$(1.5, 2.5]$	5
$(2.5, 3.5]$	28
$(3.5, 4.5]$	25
$(4.5, 5.5]$	15

The mean of these 75 sample-standard-deviations is 3.7, and the mean of the 75 sample-variances is 14.46 that has its square-root equal 3.8, so that

$$\frac{E(s)}{\sigma} \approx \frac{3.7}{3.8} = 0.97,$$

in perfect (coincidental) agreement with the above tabulated result for $\frac{E(s)}{\sigma} \approx 0.9714$ for $d.f. = 9$.

2.2. Population Non-Normally Distributed

We will examine this case by

(1) a discrete uniform distribution with $P(X = -1) = P(X = 1) = 0.5$,

(2) a continuous uniform distribution X over $[0, 1] \subset \mathbb{R}$, and

(3) same as (1) but with $P(X = 1) = 0.9 = 1 - P(X = -1)$.

We seek insight into the effect on $\frac{E(s)}{\sigma}$ of, correspondingly:

(1) decentralizing a distribution,

(2) changing a discrete distribution to its continuous analog, and

(3) changing a symmetric distribution to an asymmetric distribution.

For all the above three subcases our demonstrated calculation will be $n = 2$, as presented in the following three examples.

Example 1.

Here we have $E(X) = 0$ and $Var(X) = 1$, with the following calculation of s for $n = 2$.

$\#successes\ W$ in Bernoulli 2 trials	P	$Sample - mean$	s	$s \cdot P$
0	0.25	-1	0	0
1	0.5	0	$\sqrt{2}$	$\frac{\sqrt{2}}{2}$
2	0.25	1	0	0
			$\sum s \cdot P = E\left(\frac{s}{\sigma}\right)$	$\frac{\sqrt{2}}{2} \approx 0.71$

Following the same procedure we now tabulate the results for $n = 2, 3, 4$ and 10 below:

n	$\frac{E(s)}{\sigma}$
2	0.71
3	0.87
4	0.93
10	0.9959

Example 2.

For a continuous analogue of Example 1, consider $f(x) \equiv 1 \forall x \in [0, 1]$; then we have $E(X) = \frac{1}{2}$ and $Var(X) = \int_0^1 x^2 dx - \left(\frac{1}{2}\right)^2 = \frac{1}{12}$ so that $\sigma(X) = \frac{1}{\sqrt{12}}$. For the probability distribution of s, again we demonstrate it for $n = 2$: $\{x_1, x_2\} \subset [0, 1]$. Then we have $\forall x_1 - x_2 \equiv d \in [0, 1]$, $s = \sqrt{\left(\frac{d}{2}\right)^2 + \left(\frac{d}{2}\right)^2} = \frac{d}{\sqrt{2}} \in \left[0, \frac{1}{\sqrt{2}}\right]$. Consider the right triangle formed by the three vertices: $(0, 0)$, $(1, 0)$, and $(1, 1)$ on the plane of $\{(x_1, x_2)\}$; then $A \equiv \{x_1 - x_2 = d \rfloor d \in [0, 1]\}$ is a family of similar triangles with the length of any particular edge of any triangle in proportion to its probability density; e.g., $x_1 - x_2 = 0$ for which $s = 0$ has the longest edges in A. As such, the probability density function of s is

$$\phi(s) = 2\sqrt{2} - 4s \quad \forall s \in \left[0, \frac{1}{\sqrt{2}}\right].$$

Then we have

$$E(s) = \int_0^{\frac{1}{\sqrt{2}}} s\left(2\sqrt{2} - 4s\right) ds = \frac{\sqrt{2}}{6}$$

so that

$$\frac{E(s)}{\sigma} = \frac{\sqrt{2}}{6} \cdot \sqrt{12} = \frac{\sqrt{6}}{3} \approx 0.82.$$

Comparing this Example 2 with the previous Example 1, we see that the ratio, as from Equation (1),

$$\frac{Var(s)}{Var(X)} \equiv 1 - \frac{(E(s))^2}{Var(X)} \tag{2}$$

decreases from $\left(1 - \left(\frac{\sqrt{2}}{2}\right)^2\right) = \frac{1}{2}$ to $\left(1 - \frac{\sqrt{6}}{3}^2\right) = \frac{1}{3}$, suggesting that the sampling distribution of s becomes more concentrated as X changes from discrete to continuous and at the same time $\frac{E(s)}{\sigma}$ increase from $\frac{\sqrt{2}}{2} \approx 0.71$ to $\frac{\sqrt{6}}{3} \approx 0.82$ (cf. [3] for $Var(s^2)$).

Example 3.

Altering the discrete uniform distribution in Example 1 to $P(X=1) = 0.9 = 1 - P(X=-1)$ and carrying out the same calculation, we present the results below, where $E(X) = 0.8$ and $\sigma(X) = 0.6$:

n	$\frac{E(s)}{\sigma} \approx$
2	0.42
3	0.52
4	0.59
5	0.64
6	0.68
7	0.72
8	0.75
9	0.77
10	0.80
11	0.81
12	0.83
13	0.85
14	0.86
15	0.87
16	0.88
17	0.89
18	0.90
19	0.91
20	0.91
100	0.99

The above suggests that for asymmetric population distributions one achieves $\frac{E(s)}{\sigma} \geq 0.9$ for $n \geq 20$.

3. Summary Remarks

From the above analysis we see that while in general

$$0 \leq \frac{E(s)}{\sigma} \leq 1,$$

in applications for $n \geq 20$ one is likely to achieve $\frac{E(s)}{\sigma} \geq 0.9$.

We also note that from our treatment of $Var(s)+(E(s))^2 \equiv E\left(s^2\right)$, we incidentally analyzed the general ratio of $(E(X))^2/E\left(X^2\right)$ or $|E(X)|/\sqrt{E(X^2)}$, relating the first moment of X to its second. Geometrically we consider an $N - sphere$ of radius R:

$$\begin{aligned} \frac{|E(X)|}{\sqrt{E(X^2)}} &= \frac{\left|\sum_{i=1}^N x_i\right|}{\sqrt{\sum_{i=1}^N x_i^2}} \cdot \frac{\sqrt{N}}{N} \\ &\equiv \frac{\left|\sum_{i=1}^N x_i\right|}{R} \cdot \frac{1}{\sqrt{N}} \\ &\leq 1; \end{aligned}$$

equality holds if and only if $x_i = a \; \forall i = 1, \cdots, N$, i.e., the triangle inequality or the Cauchy–Schwarz inequality with one vector equal to $\langle \pm 1, \cdots, \pm 1 \rangle$.

Finally while this paper addressed only the mean of the sample-standard-deviation s, we by the way presented a numerical procedure of arriving at the entire distribution of s for a normally distributed population for any sample size $n \geq 2$, i.e.,

$$s = \frac{\sigma}{\sqrt{n-1}} \sqrt{\chi^2},$$

where chi-square $\chi^2 \equiv y \in (0, \infty)$ has probability density

$$\phi(y) = \frac{0.5^{0.5(n-1)} y^{0.5(n-3)}}{\Gamma(0.5(n-1))} \cdot e^{-0.5y}.$$

Consider a change of variable

$$f^{-1} : y \mapsto x = \sqrt{y},$$

then

$$\begin{aligned} x &= \sqrt{n-1}\left(\frac{s}{\sigma}\right) \text{ has density} \\ \phi(y) &= \phi(f(x)) \equiv \hat{\phi}(x) \\ &= \frac{0.5^{0.5(n-1)} x^{n-3}}{\Gamma(0.5(n-1))} \cdot e^{-0.5x^2} \cdot 2x \\ \text{(where } dy &= 2xdx). \end{aligned}$$

As such for $n = 2$ we have

$$\begin{aligned} \hat{\phi}(x) &= \frac{2}{\sqrt{2\pi}} e^{-0.5x^2} \text{ so that} \\ \int_0^\infty \hat{\phi}(x)\, dx &= 1, \end{aligned}$$

evidencing the fact that for $d.f. = 1$ one has $\chi^2 = Z^2$ so that its square-root gives the standard normal distribution over $[0, \infty)$ with the density doubled (cf. [2] for a treatment on lognormal distributions). Since doubling the right half of the normal distribution does not affect the center of mass over $Z \in [0, \infty)$, $E\left(\hat{\phi}(x)\right)$ for $n = 2$ is simply the center of mass of the right half of the normal distribution.

We now make an overall comparison over the above four cases as based on $n = 2$. We see that $\frac{E(s)}{\sigma}$ increases from (a) 0.42, to (b) 0.71, to (c) 0.79, and to (d) 0.82 as the population changes from (a) the highly asymmetric, to (b) the uniform discrete, to (c) the normal, and to (d) the uniform continuous. An observation from (a) to (b) and (c) to (d) suggests that centralizing data as in (a) and (c) decreases $\frac{E(s)}{\sigma}$. As for from (b) to (c), we consider a discrete distribution X with $P(X = -1) = P(X = 1) = \frac{1}{m}$ and $P(X = 0) = 1 - \frac{2}{m}$, $\forall m \geq 2$. Then we have

$$\begin{aligned} \sigma &= \sqrt{\frac{2}{m}}, \text{ and for } n = 2: \\ P(s = 0) &= \frac{m^2 - 4m + 6}{m^2}, \\ P\left(s = \frac{\sqrt{2}}{2}\right) &= \frac{4(m-2)}{m^2}, \text{ and} \\ P\left(s = \sqrt{2}\right) &= \frac{2}{m^2}. \end{aligned}$$

Accordingly,

$$E(s) = \frac{2\sqrt{2}}{m^2}(m-1), \text{ so that}$$
$$\frac{E(s)}{\sigma} = 2(m-1)\,m^{-\frac{3}{2}}.$$

We tabulate the effect of m on $\frac{E(s)}{\sigma}$ as follows:

m	$\frac{E(s)}{\sigma}$
2	0.71
3	**0.77**
4	0.75
5	0.72
6	0.68
7	0.65
8	0.62
9	0.59
10	0.57
11	0.55
12	0.53
13	0.51
14	0.50
15	0.48
16	0.47
17	0.46
18	0.45
19	0.43
20	0.42
21	0.42
22	0.41
23	0.40
24	0.39
25	0.38

We observe that at $m = 3$, $\frac{E(s)}{\sigma} = 0.77$ reaches the maximum, when the uniform distribution is still maintained, indicating that: (1) in the trend of changing the uniform distribution from discrete to continuous, $\frac{E(s)}{\sigma}$ is to increase to the limit of $\frac{E(s)}{\sigma} = 0.82$, that of case (d), and (2) however, once $m = 4$

when concentration of mass occurs, $\frac{E(s)}{\sigma}$ decreases. Now since at $m = 22$, $\frac{E(s)}{\sigma} = 0.41 < 0.42$, which was the case of (a), we see that population being symmetric or not has only indirect effect on $\frac{E(s)}{\sigma}$; the deciding factor is concentration of masses in the population, to lead to a smaller $\frac{E(s)}{\sigma}$ than otherwise. As such we put forth a conjecture that the continuous uniform distribution achieves the maximum $\frac{E(s)}{\sigma}$ for any given sample size n among all probability distributions; in addition, any discrete distribution if altered into its frequency polygon to become a continuous distribution is to increase $\frac{E(s)}{\sigma}$ since points of positive probability measures represent concentration of masses as compared with probability density. As such, case (c) of the normal distribution is subject to two opposite influences, that of being continuous and that of having concentration of masses. More plainly, concentration of masses in a population yields more samples containing similar (or even the same) values so that $E(s)$ becomes smaller; of course σ also becomes smaller, but not as proportionately small by considering the fact that the range of a sample is in general smaller that of the population.

Finally a generalized formulation of our analysis in this paper that may be worth further pursuing is a comparison between

$$\int_{\Omega(y)} y\phi(y)\,dy = E(Y)$$

and

$$\begin{aligned} & f\left(\int_{\Omega(y)} f^{-1}(y)\,\phi(y)\,dy\right) \\ = \; & f\left(\int_{\Omega(x)} x\phi(f(x))\,f'(x)\,dx\right) \\ = \; & f\left(\int_{\Omega(x)} x\hat{\phi}(x)\,dx\right) \\ = \; & f(E(X)) \\ = \; & f\left(E\left(f^{-1}(Y)\right)\right), \end{aligned}$$

treated in this paper as $E\left(s^2\right)$ compared with $(E(s))^2$.

References

[1] S.A. Book, Why $n-1$ in the formula for the sample standard deviation? *The Two-Year College Math. J.*, **10** (1979), 330-333.

[2] H. Chen, Comparisons of lognormal population means, *Proc. Amer. Math. Soc.*, **121** (1994), 915-924.

[3] E. Cho, M.J. Cho, Variance of sample variance, *Amer. Stat. Assoc. Joint Statistical Meetings Section on Research Methods* (2008), 1291-1293.

[4] F.M. Hossain, A. Shams, M. Arif, A.H. Joarder, Shorter variation of standard deviation for small sample, *Int. J. Mathematical Education in Science and Technology*, **45** (2014), 311-316.

[5] N.H. Wasserman, S. Casey, J. Champion, M. Huey, Statistics as unbiased estimators: exploring the teaching of standard deviation, *Research in Mathematics Education*, **19** (2017), 236-256.