Cross Validated

Intuitive explanation for dividing by $n-1$ when calculating standard deviation?

Asked 9 years ago Active 3 months ago Viewed 115k times

137

110

I was asked today in class why you divide the sum of square error by $n-1$ instead of with n, when calculating the standard deviation.

I said I am not going to answer it in class (since I didn't wanna go into unbiased estimators), but later I wondered - **is there** an intuitive explanation for this?!

standard-error intuition teaching bessels-correction

edited Jul 18 at 9:31
kjetil b halvorsen
38.2k 9 92 297

asked Oct 23 '10 at 22:04
Tal Galili
10.1k 26 119 180

29 I'd like to quote this zinger from the book *Numerical Recipes*: "...if the difference between n and $n-1$ ever matters to you, then you are probably up to no good anyway - e.g., trying to substantiate a questionable hypothesis with marginal data." – J. M. is not a statistician Oct 25 '10 at 14:09

11 a really elegant, intuitive explanation is presented here (below the proof) en.wikipedia.org/wiki/... The basic idea is that your observations are, naturally, going to be closer to the sample mean than the population mean. – WetlabStudent May 26 '14 at 16:23

12 @Tal, This is why schools suck. You ask them "why *this*?", and they reply "just memorize it". – Pacerier Jun 3 '15 at 11:51

1 If you are looking for an intuitive explanation, you should see the reason for yourself by actually taking samples! Watch this, it precisely answers you question. youtube.com/watch?v=xslIhnquFoE – Sahil Chaudhary Sep 24 '15 at 23:36

tl;dr: (from top answer:) "...the standard deviation which is calculated using deviations from the sample mean underestimates the desired standard deviation of the population..." See also: en.wikipedia.org/wiki/... So, unless you feel like calculating something somewhat complex, just use n-1 if it's from a sample. – Andrew Aug 26 '17 at 22:58

15 Answers

99

The standard deviation calculated with a divisor of $n-1$ is a standard deviation calculated from the sample as an estimate of the standard deviation of the population from which the sample was drawn. Because the observed values fall, on average, closer to the sample mean than to the population mean, the standard deviation which is calculated using deviations from the sample mean underestimates the desired standard deviation of the population. Using $n-1$ instead of n as the divisor corrects for that by making the result a little bit bigger.

When the sample is the whole population we use the standard deviation with n as the divisor because the sample mean *is* population mean.

(I note parenthetically that nothing that starts with "second moment recentered around a known, definite mean" is going to fulfil the questioner's request for an intuitive explanation.)

edited Aug 25 '17 at 16:45
Mooncrater
357 4 16

answered Oct 24 '10 at 3:46
Michael Lew
8,116 1 24 39

13 Let's not confuse "intuitive" with "nontechnical". – whuber ♦ Oct 24 '10 at 15:38

32 @Michael, This doesn't explain Why do we use `n−1` instead of `n−2` (or even `n−3`)? – Pacerier Jun 3 '15 at 11:52

1 @Pacerier Have a look at Whuber's answer below for detail on that point. In essence, the correction is n-1 rather than n-2 etc because the n-1 correction gives results that are very close to what we need. More exact corrections are shown here: en.wikipedia.org/wiki/Unbiased_estimation_of_standard_deviation – Michael Lew Jun 3 '15 at 21:37

1 Hi @Michael, so why deviation calculated from sample mean tends to be smaller than population mean? – Allen Nov 15 '16 at 3:57

1 "Because the observed values fall, on average, closer to the sample mean than to the population mean, the standard deviation which is calculated using deviations from the sample mean underestimates the desired standard deviation of the population." Why the sample mean always underestimates? What if it overestimates? – Bora M. Alper Nov 20 '16 at 8:03

55

A common one is that the definition of variance (of a distribution) is the second moment recentered around a *known, definite* mean, whereas the estimator uses an *estimated* mean. This loss of a degree of freedom (given the mean, you can reconstitute the dataset with knowledge of just $n-1$ of the data values) requires the use of $n-1$ rather than n to "adjust" the result.

Such an explanation is consistent with the estimated variances in ANOVA and variance components analysis. It's really just a special case.

The need to make *some* adjustment that inflates the variance can, I think, be made intuitively clear with a valid argument that isn't just *ex post facto* hand-waving. (I recollect that Student may have made such an argument in his 1908 paper on the t-test.) Why the adjustment to the variance should be *exactly* a factor of $n/(n-1)$ is harder to justify, especially when you consider that the adjusted SD is *not* an unbiased estimator. (It is merely the square root of an unbiased estimator of the variance. Being unbiased usually does not survive a nonlinear transformation.) So, in fact, the correct adjustment to the SD to remove its bias is *not* a factor of $\sqrt{n/(n-1)}$ at all!

Some introductory textbooks don't even bother introducing the adjusted sd: they teach one formula (divide by n). I first reacted negatively to that when teaching from such a book but grew to appreciate the wisdom: to focus on the concepts and applications, the authors strip out all inessential mathematical niceties. It turns out that nothing is hurt and nobody is misled.

answered Oct 23 '10 at 22:21

1 Thank you Whuber. I have to teach the students with the n-1 correction, so dividing in n alone is not an option. As written before me, to mention the connection to the second moment is not an option. Although to mention how the mean was already estimated thereby leaving us with less "data" for the sd - that's important. Regarding the bias of the sd - I remembered encountering it - thanks for driving that point home. Best, Tal – Tal Galili Oct 24 '10 at 7:15

3 @Tal I was writing in your language, not that of your students, because I am confident you are fully capable of translating it into whatever you know will reach them. In other words, I interpreted "intuitive" in your question to mean intuitive to *you*. – whuber ♦ Oct 24 '10 at 15:40

1 Hi Whuber. Thank you for the vote of confidence :). The loose of the degree of freedom for the estimation of the expectancy is one that I was thinking of using in class. The problem is that the concept of "degrees of freedom" by itself is one that needs knowledge/intuition. But combining it with some of the other answers given in this thread will be useful (to me, and I hope others in the future). Best, Tal – Tal Galili Oct 24 '10 at 21:12

For large n, there isn't typically much difference between dividing by n or $n-1$, so it would be acceptable to introduce the uncorrected formula provided it was intended to apply to large samples, no? – PatrickT Sep 16 '17 at 14:45

1 @Patrick You might be reading too much into my answer, because it *is* explicit about the reasons: they are pedagogical and have nothing to do with whether n is large or not. – whuber ♦ Sep 18 '17 at 12:58

50

By definition, variance is calculated by taking the sum of squared differences from the mean and dividing by the size. We have the general formula

$\sigma^2 = \frac{\sum_i^N (X_i-\mu)^2}{N}$ where μ is the mean and N is the size of the population.

According to this definition, variance of the a sample (e.g. sample t) must also be calculated in this way.

$\sigma_t^2 = \frac{\sum_i^n (X_i-\overline{X})^2}{n}$ where $\overline{X}$ is the mean and n is the size of this small sample.

However, by sample variance S^2, we mean an estimator of the population variance σ^2. How can we estimate σ^2 only by using the values from the sample?

According to the formulas above, the random variable X deviates from sample mean $\overline{X}$ with variance σ_t^2. The sample mean $\overline{X}$ also deviates from μ with variance $\frac{\sigma^2}{n}$ because sample mean gets different values from sample to sample and it is a random variable with mean μ and variance $\frac{\sigma^2}{n}$. (One can prove easily.)

Therefore, roughly, X should deviate from μ with a variance that involves two variances so add up these two and get $\sigma^2 = \sigma_t^2 + \frac{\sigma^2}{n}$. By solving this, we get $\sigma^2 = \sigma_t^2 \times \frac{n}{n-1}$. Replacing σ_t^2 gives our estimator for population variance:

$$S^2 = \frac{\sum_i^n (X_i-\overline{X})^2}{n-1}.$$

Dror Atariah
179 1 12

sevenkul
601 5 4

I hope this isn't too trivial: is it the fact that the sample mean converges to ND(μ, $\frac{\sigma}{\sqrt{n}}$) as n gets arbitrarily large the reason why sample mean deviates from the real mean with variance $\frac{\sigma^2}{n}$? – RexYuan Sep 12 '17 at 10:39

6 This is a better explanation than the others because it shows the equations and derivations instead of simply going yagga yagga with statistical terms. – Nav Jun 21 '18 at 6:19

1 @sevenkul can we some how view this visually? when you say, X should deviate from μ with that net variance, I am lost in visualizing that – Parthiban Rajendran Nov 22 '18 at 12:37

17

This is a total intuition, but the simplest answer is that is a correction made to make standard deviation of one-element sample undefined rather than 0.

answered Oct 24 '10 at 10:28
user88

11 Why not, then, use $\frac{n}{n^2-1}$ or even $\frac{1}{\exp(1)-\exp(1/n)}$ as corrections? :-) – whuber ♦ Sep 22 '11 at 6:25

1 @whuber Parsimony (-; – user88 Sep 22 '11 at 10:17

4 $\frac{1}{n-1}$ is even more "parsimonious". :-) – whuber ♦ Sep 22 '11 at 15:02

2 @mbq, Regarding your answer ~"it's a correction made to make standard deviation of one-element sample undefined rather than 0", is that *really* the reason why, or is this a joke answer? You know non-mathers like us can't tell. – Pacerier Jun 3 '15 at 11:56

4 Formally, it is a consequence than reason, but, as i wrote, I find it to be a good intution to memorize it. – user88 Jun 3 '15 at 17:45

14

You can gain a deeper understanding of the $n-1$ term through geometry alone, not just why it's not n but why it takes exactly this form, but you may first need to build up your intuition cope with n-dimensional geometry. From there, however, it's a small step to a deeper understanding of degrees of freedom in linear models (i.e. model df & residual df). I think there's little doubt that Fisher thought this way. Here's a book that builds it up gradually:

Saville DJ, Wood GR. *Statistical methods: the geometric approach*. 3rd edition. New York: Springer-Verlag; 1991. 560 pages. 9780387975177

(Yes, 560 pages. I did say gradually.)

edited Aug 25 '17 at 17:17
Mooncrater
357 4 16

answered Oct 24 '10 at 11:01
onestop
16.3k 2 46 77

Thanks onestop - I didn't think there would be an answer from that direction. Any way to sum-up the intuition, or is that not likely to be possible? Cheers, Tal – Tal Galili Oct 24 '10 at 11:20

Can you summarise a bit of the intuition rather than just a book reference? – oliversm Sep 30 at 18:34

12

The estimator of the population variance is biased when applied on a sample of the population. In order to adjust for that bias on needs to divide by n-1 instead of n. One can show mathematically that the estimator of the sample variance is unbiased when we divide by n-1 instead of n. A formal proof is provided here:

https://economictheoryblog.com/2012/06/28/latexlatexs2/

Initially it was the mathematical correctness that led to the formula, I suppose. However, if one wants to add intuition to a formula the already mentioned suggestions appear reasonable.

First, observations of a sample are on average closer to the sample mean than to the population mean. The variance estimator makes use of the sample mean and as a consequence underestimates the true variance of the population. Dividing by n-1 instead of n corrects for that bias.

Furthermore, dividing by n-1 make the variance of a one-element sample undefined rather than zero.

answered Sep 2 '16 at 20:08

Richard Hansen
141 1 3

12

Why divide by $n-1$ rather than n? Because it is customary, and results in an unbiased estimate of the variance. However, it results in a biased (low) estimate of the standard deviation, as can be seen by applying Jensen's inequality to the concave function, square root.

So what's so great about having an unbiased estimator? It does not necessarily minimize mean square error. The MLE for a Normal distribution is to divide by n rather than $n-1$. Teach your students to think, rather than to regurgitate and mindlessly apply antiquated notions from a century ago.

edited Aug 25 '17 at 17:16

Mooncrater
357 4 16

answered Aug 28 '15 at 15:28

Mark L. Stone
11.4k 1 21 47

8 (+1) The more I think about this situation (and I've given it some real thought, to the extent of researching the earlier papers such as Student's 1908 Biometrica contribution to try to track down when and why $n-1$ made its appearance), the more I think that "because it's customary" is the only possible correct answer. I am unhappy to see the downvotes and can only guess that they are responding to the last sentence, which could easily be seen as attacking the O.P., even though I doubt that was your intention. – whuber ♦ Aug 28 '15 at 17:45

1 My last sentence was friendly advice to all concerned, as opposed to an attack on the OP. – Mark L. Stone Aug 29 '15 at 1:50

In much use it will not matter, when used in tests or for confidence intervals one would have to adjust other parts of the procedure and in the end obtain the same result! – kjetil b halvorsen Aug 25 '17 at 16:55

8

$z = -\frac{\beta}{\alpha}$. This shows that, for any given n real numbers $x_1, x_2, \ldots, x_n$, the quantity

$$G(a) = \sum_{i=1}^{n}(x_i - a)^2 = \left(\sum_{i=1}^{n} x_i^2\right) - 2a\left(\sum_{i=1}^{n} x_i\right) + na^2,$$

has minimum value when $a = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$.

Now, suppose that the x_i are a sample of size n from a distribution with unknown mean μ and unknown variance σ^2. We can estimate μ as $\frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$ which is easy enough to calculate, but an attempt to estimate σ^2 as $\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2 = n^{-1}G(\mu)$ encounters the problem that we don't know μ. We can, of course, readily compute $G(\bar{x})$ and we know that $G(\mu) \geq G(\bar{x})$, but how much larger is $G(\mu)$? The answer is that $G(\mu)$ is larger than $G(\bar{x})$ by a *factor* of approximately $\frac{n}{n-1}$, that is,

$$G(\mu) \approx \frac{n}{n-1}G(\bar{x}) \tag{1}$$

and so the *estimate* $n^{-1}G(\mu) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$ for the variance of the distribution can be approximated by $\frac{1}{n-1}G(\bar{x}) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$.

So, what is an *intuitive* explanation of (1)? Well, we have that

$$\begin{aligned}
G(\mu) &= \sum_{i=1}^{n}(x_i - \mu)^2 \\
&= \sum_{i=1}^{n}(x_i - \bar{x} + \bar{x} - \mu)^2 \\
&= \sum_{i=1}^{n}\left((x_i - \bar{x})^2 + (\bar{x} - \mu)^2 + 2(x_i - \bar{x})(\bar{x} - \mu)\right) \\
&= G(\bar{x}) + n(\bar{x} - \mu)^2 + (\bar{x} - \mu)\sum_{i=1}^{n}(x_i - \bar{x}) \\
&= G(\bar{x}) + n(\bar{x} - \mu)^2 \qquad (2)
\end{aligned}$$

since $\sum_{i=1}^{n}(x_i - \bar{x}) = n\bar{x} - n\bar{x} = 0$. Now,

$$\begin{aligned}
n(\bar{x} - \mu)^2 &= n\frac{1}{n^2}\left(\sum_{i=1}^{n}(x_i - \mu)\right)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2 + \frac{2}{n}\sum_{i=1}^{n}\sum_{j=i+1}^{n}(x_i - \mu)(x_j - \mu)
\end{aligned}$$

Except when we have an extraordinarily unusual sample in which all the x_i are larger than μ (or they are all smaller than μ), the summands $(x_i - \mu)(x_j - \mu)$ in the double sum on the right side of (3) take on positive as well as negative values and thus a lot of cancellations occur. Thus, the double sum can be expected to have *small* absolute value, and we simply ignore it in comparison to the $\frac{1}{n}G(\mu)$ term on the right side of (3). Thus, (2) becomes

$$G(\mu) \approx G(\bar{x}) + \frac{1}{n}G(\mu) \implies G(\mu) \approx \frac{n}{n-1}G(\bar{x})$$

as claimed in (1).

answered Sep 1 '15 at 4:17
Dilip Sarwate
33k 3 61 157

8 Only on this stack exchange would this ever be considered an intuitive answer. – Joseph Garvin Mar 13 '17 at 3:15

6

Sample variance can be thought of to be the exact mean of the pairwise "energy" $(x_i - x_j)^2/2$ between all sample points. The definition of sample variance then becomes

$$s^2 = \frac{2}{n(n-1)} \sum_{i<j} \frac{(x_i - x_j)^2}{2} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2.$$

This also agrees with defining variance of a random variable as the expectation of the pairwise energy, i.e. let X and Y be independent random variables with the same distribution, then

$$V(X) = E\left(\frac{(X-Y)^2}{2}\right) = E((X - E(X))^2).$$

To go from the random variable defintion of variance to the defintion of sample variance is a matter of estimating a expectation by a mean which is can be justified by the philosophical principle of typicality: The sample is a typical representation the distribution. (Note, this is related to, but not the same as estimation by moments.)

answered Oct 25 '10 at 9:51
B Student

2 I couldn't quite follow you at the last paragraph. Isn't mathematical fact that $V(X) = E\left(\frac{(X-Y)^2}{2}\right) = E((X - E(X))^2)$? Even though the equation is interesting, I don't get how it could be used to teach n-1 intuitively? – KH Kim Jun 17 '12 at 1:59

4 I like this approach, but it omits a key idea: to compute the mean energy between *all* pairs of sample points, one would have to count the values $(x_i - x_i)^2$, even though they are all zero. Thus the numerator of s^2 remains the same but the denominator ought to be n, not $n-1$. This shows the sleight of hand that has occurred: somehow, you need to justify *not* including such self-pairs. (Because

4

Suppose that you have a random phenomenon. Suppose again that you only get one $N = 1$ sample, or realization, x. Without further assumptions, your "only" reasonable choice for a sample average is $\overline{m} = x$. If you do not subtract 1 from your denominator, the (uncorrect) sample variance would be

$$V = \frac{\sum_N (x_n - \overline{m})^2}{N}$$

, or:

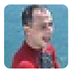
$$\overline{V} = \frac{(x - \overline{m})^2}{1} = 0\,.$$

Oddly, the variance would be null with only one sample. And having a second sample y would risk to increase your variance, if $x \neq y$. This makes no sense. Intuitively, an infinite variance would be a sounder result, and you can recovered it only by "dividing by $N - 1 = 0$".

Estimating a mean is fitting a polynomial with degree 0 to the data, having one degree of freedom (dof). This Bessel's correction applies to higher degrees of freedom models too: of course you can fit perfectly $d + 1$ points with a d degree polynomial, with $d + 1$ dofs. The illusion of a zero-squared-error can only be counterbalanced by dividing by the number of points minus the number of dofs. This issue is particularly sensitive when dealing with very small experimental datasets.

edited Jun 25 '17 at 21:08

answered Sep 2 '16 at 20:49



Laurent Duval
1,522 1 12 31

It is unclear why "an infinite variance would be a sounder result" than a zero variance. Indeed, you seem to use "sample variance" in the sense of a variance *estimator*, which is more confusing yet. – whuber ♦ Jun 26 '17 at 16:42

1 I understand. To answer an intuitive explanation between two options, I tried to suggest that one of the two is somehow unacceptable, based on the mundane rule that $0 < \infty$. A rephrasing is indeed necessary, and upcoming – Laurent Duval Jun 26 '17 at 18:33

4

At the suggestion of whuber, this answer has been copied over from another similar question.

Bessel's correction is adopted to correct for bias in using the sample variance as an estimator of the true variance. The bias in the uncorrected statistic occurs because the sample mean is closer to the middle of the observations than the true mean, and so the squared deviations around the sample mean systematically underestimates the squared deviations around the true mean.

To see this phenomenon algebraically, just derive the expected value of a sample variance without Bessel's correction and see what it looks like. Letting S_*^2 denote the uncorrected sample variance (using n as the denominator) we have:

$$
\begin{aligned}
S_*^2 &= \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2 \\
&= \frac{1}{n} \sum_{i=1}^{n} (X_i^2 - 2\bar{X}X_i + \bar{X}^2) \\
&= \frac{1}{n} \left(\sum_{i=1}^{n} X_i^2 - 2\bar{X} \sum_{i=1}^{n} X_i + n\bar{X}^2 \right) \\
&= \frac{1}{n} \left(\sum_{i=1}^{n} X_i^2 - 2n\bar{X}^2 + n\bar{X}^2 \right) \\
&= \frac{1}{n} \left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2 \right) \\
&= \frac{1}{n} \sum_{i=1}^{n} X_i^2 - \bar{X}^2.
\end{aligned}
$$

Taking expectations yields:

$$
\begin{aligned}
\mathbb{E}(S_*^2) &= \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}(X_i^2) - \mathbb{E}(\bar{X}^2) \\
&= \frac{1}{n} \sum_{i=1}^{n} (\mu^2 + \sigma^2) - (\mu^2 + \frac{\sigma^2}{n}) \\
&= (\mu^2 + \sigma^2) - (\mu^2 + \frac{\sigma^2}{n}) \\
&= \sigma^2 - \frac{\sigma^2}{n} \\
&= \frac{n-1}{n} \cdot \sigma^2
\end{aligned}
$$

So you can see that the uncorrected sample variance statistic underestimates the true variance σ^2. Bessel's correction replaces the denominator with $n-1$ which yields an unbiased estimator. In regression analysis this is extended to the more general case where the estimated mean is a linear function of multiple predictors, and in this latter case, the denominator is reduced further, for the lower number of degrees-of-freedom.

answered Apr 23 '18 at 0:04

Reinstate Monica
40.8k 2 52 176

0

Generally using "n" in the denominator gives smaller values than the population variance which is what we want to estimate. This especially happens if the small samples are taken. In the language of statistics, we say that the sample variance provides a “biased” estimate of the population variance and needs to be made "unbiased".

If you are looking for an intuitive explanation, you should let your students see the reason for themselves by actually taking samples! Watch this, it precisely answers your question.

https://www.youtube.com/watch?v=xslIhnquFoE

edited Sep 25 '15 at 1:21

answered Sep 24 '15 at 23:36

Sahil Chaudhary
119 4

0

The sample mean is defined as $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, which is quite intuitive. But the sample variance is $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$. Where did the $n-1$ come from ?

To answer this question, we must go back to the definition of an unbiased estimator. An unbiased estimator is one whose expectation tends to the true expectation. The sample mean is an unbiased estimator. To see why:

$$E[\bar{X}] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{n}{n}\mu = \mu$$

Let us look at the expectation of the sample variance,

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i^2) - n\bar{X}^2$$

$$E[S^2] = \frac{1}{n-1}\left(nE[(X_i^2)] - nE[\bar{X}^2]\right).$$

Notice that $\bar{X}$ is a random variable and not a constant, so the expectation $E[\bar{X}^2]$ plays a role. **This is the reason behind the** $n-1$.

$$E[S^2] = \frac{1}{n-1}(n(\mu^2 + \sigma^2) - n(\mu^2 + Var(\bar{X}))).$$

$$Var(\bar{X}) = Var(\frac{1}{n}\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n}\frac{1}{n^2}Var(X_i) = \frac{\sigma^2}{n}$$

$$E[S^2] = \frac{1}{n-1}(n(\mu^2 + \sigma^2) - n(\mu^2 + \sigma^2/n)). = \frac{(n-1)\sigma^2}{n-1} = \sigma^2$$

As you can see, if we had the denominator as n instead of $n-1$, we would get a biased estimate for the variance! But with $n-1$ the estimator S^2 is an unbiased estimator.

edited Jan 16 '16 at 9:59

answered Jan 15 '16 at 17:18

3 But it doesn't follow that S is an unbiased estimator of the standard deviation. – Scortchi - Reinstate Monica ♦ Jan 15 '16 at 17:32

-1

I think it's worth pointing out the connection to Bayesian estimation. Suppose you assume your data is Gaussian, and so you measure the mean μ and variance σ^2 of a sample of n points. You want to draw conclusions about the population. The Bayesian approach would be to evaluate the posterior predictive distribution over the sample, which is a generalized Student's T distribution (the origin of the T-test). This distribution has mean μ, and variance

$$\sigma^2\left(\frac{n+1}{n-1}\right),$$

which is even larger than the typical correction. (It has $2n$ degrees of freedom.)

The generalized Student's T distribution has three parameters and makes use of all three of your statistics. If you decide to throw out some information, you can further approximate your data using a two-parameter normal distribution as described in your question.

From a Bayesian standpoint, you can imagine that uncertainty in the hyperparameters of the model (distributions over the mean and variance) cause the variance of the posterior predictive to be greater than the population variance.

edited Aug 28 '15 at 22:48

answered Aug 28 '15 at 15:16

Neil G
10.3k 2 35 74

-4

My goodness it's getting complicated! I thought the simple answer was... if you have all the data points you can use "n" but if you have a "sample" then, assuming it's a random sample, you've got more sample points from inside the standard deviation than from outside (the definition of standard deviation). You just don't have enough data outside to ensure you get all the data points you need randomly. The n-1 helps expand toward the "real" standard deviation.

answered Apr 7 '16 at 1:16

user111282
1

3 This doesn't make sense. More points from inside the SD than outside? If that means within 1 SD of the mean versus not within, whether that is true has nothing to do with taking a sample. For necessary constraints on fractions within intervals around the mean, see Chebyshev's inequality. To the main question here, "helps expand" doesn't explain $n-1$ at all, as even granting your argument $n-2$ might be better still, and so forth, as there is no algebra here, even implicitly. Unfortunately this adds nothing to other answers except a confused set of ideas, either incorrect or irrelevant. – Nick Cox Apr 7 '16 at 1:41

protected by Nick Cox Apr 7 '16 at 1:41

Thank you for your interest in this question. Because it has attracted low-quality or spam answers that had to

Would you like to answer one of these unanswered questions instead?