MAKING AN UNBIASED ESTIMATE OF THE STANDARD DEVIATION OF A POPULATION THAT IS DISTRIBUTED NORMALLY

Robert W. Hayden

If we take a sample of size n,

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

is an unbiased estimator of the population variance, σ^2 (1, p.227). From this we might suppose that s is an unbiased estimator of σ, but such is not the case. The supposition is based on the appealing but false idea that $\sqrt{E(X^2)} = E(X)$. As a counterexample to this idea, note that if z has the standard normal distribution,

$$E(z) = \mu = 0$$

but

$$\sqrt{E(z^2)} = \sqrt{1} = 1.$$

Unfortunately, this error still appears in print from time to time (2, p.1-51).

Since s is a biased estimator of σ, we might seek a correction factor, c, such that cs is an unbiased estimator of σ. This is a standard problem for an undergraduate mathematical statistics course (1, p.228, Problem 7.5-5), and the result is that such a c exists for $n > 1$, and is given by

$$c(n) = \frac{\Gamma(\frac{n-1}{2})\sqrt{n-1}}{\Gamma(\frac{n}{2})\sqrt{2}}$$

where Γ is the gamma function.

This form for $c(n)$ is fine for theoretical discussions, but the transcendence of the gamma function makes implementation on a computer or hand-held calculator difficult if numerical values of c are required. For this purpose,

$$c(n) = \begin{cases} \dfrac{2^{n-2.5}\sqrt{n-1}}{(n-2)\sqrt{\pi}\binom{n-3}{\frac{n-3}{2}}} & n \text{ odd} \\ \binom{n-3}{\frac{n-2}{2}} \dfrac{\sqrt{\pi(n-1)}}{2^{n-2.5}} & n \text{ even}, n \geq 4 \end{cases}$$

is a much more useful form. From this, the following values of $c(n)$ are easily computed on a hand-held calculator such as the TI-54:

n	c	c'
2	1.253	1.250
3	1.128	1.125
4	1.085	1.083
5	1.064	1.062
6	1.051	1.050
8	1.0362	1.0357
10	1.0281	1.0278
20	1.0132	1.0132
40	1.0064	1.0064
100	1.0025	1.0025
200	1.0013	1.0013

The amount by which c differs from 1 is a measure of the bias in using s to estimate σ. This bias is significant for very small samples, but less than 1% when $n > 26$. The quantity

$$c' = \frac{n-0.75}{n-1}$$

is a simple approximation to c. It is accurate to within one quarter of 1% for all n.

References

1. Hogg, R.V., and Tanis, E.A., *Probability and Statistical Inference*, New York, Macmillan, 1977.

2. Kelly, K.A. et al., *Scientific Calculator Sourcebook*, Dallas, Texas Instruments, 1981.