Letter to the Editor

Robin de Nijs
Copenhagen University Hospital, Rigshospitalet

Journal of Statistics Education Volume 23, Number 1 (2015),
www.amstat.org/publications/jse/v23n1/de_Nijs_letter.pdf

Letter in response to: Doane D.P., and Seward L.E. (2011), "Measuring Skewness: A Forgotten Statistic?," *Journal of Statistics Education*, 19(2), 1-28
http://www.amstat.org/publications/jse/v19n2/doane.pdf

Critical values for Pearson's second coefficient of skewness

With interest, I have read the article by Doane and Seward (2011) about measuring skewness. The article contains a definition for skewness Sk_2 based on the difference between mean $\bar{x}$ and median m divided by the sample standard deviation s. Depending on its origin, the skewness measure is multiplied by 3, yielding Pearson's second coefficient of skewness (Pearson 1895) or 2 (Bonferroni 1930), or just 1 (Hotelling and Solomons 1932, Yule 1911).

Doane and Seward (2011) used the original definition by Pearson:

$$Sk_2 = 3\frac{\bar{x}-m}{s}.$$

It is written in the article that no tables of critical values could be found. This is probably true for Pearson's second coefficient of skewness, but critical values for related statistics can be found. Cabilio and Masaro (1996) used a test statistic S_K equal to Pearson's second coefficient, where the "3" is replaced by the square root of the number of samples n:

$$S_K = \sqrt{n}\frac{\bar{x}-m}{s}.$$

Cabilio and Masaro calculated critical values for this, which can be used to obtain the critical values in Doane and Seward by multiplying by $3/\sqrt{n}$.

de Nijs and Klausen (2013) derived an equation for the expected squared difference between mean and median related to the sample standard deviation. The expectation value is denoted by $\langle \quad \rangle$ and the variance by var. For distributions with $\langle \bar{x} \rangle = \langle m \rangle$ this equation is given by

$$\langle (\bar{x} - m)^2 \rangle = \mathrm{var}(\bar{x} - m) = \beta_n^2 \frac{s^2}{n},$$

where β_n is a constant depending on the underlying distribution and the number of samples n. For normal distributions $\beta_\infty = \sqrt{\pi/2 - 1} \approx 0.7555$. de Nijs and Klausen (2013) published values for β_n with a four digit precision for normal and uniform distributions. The critical values for $\langle (\bar{x} - m)^2 \rangle$ can be used to calculate critical values for Pearson's second coefficient of skewness by multiplying β_n by $3\Phi^{-1}(0.95)/\sqrt{n}$, where $\Phi^{-1}(0.95)$ the critical value for a singlesided 5% confidence level of a normal distribution, which is approximately 1.644854.

In Table 1, a comparison is made for the tabulated values in the three articles by calculating critical values for Pearson's second coefficient of skewness as in Doane and Seward. The values in the last column of Table 1 have a precision of four digits. The values in Cabilio and Masaro are intervals because they round to the nearest hundredth. To achieve four digit accuracy, the Cabilio and Masaro values are adjusted to include the smallest possible value and the largest possible value that would round to the given value. For example, the Cabilio and Masaro value for n = 10 is 1.01; this value could be as low as 1.005 or as high as 1.015.

The small deviations between the columns can be explained by the limited amount of samples in the simulations. Doane and Seward used 40,000 samples and Cabilio & Masaro 10^6 samples.

Table 1. The 5% upper percentile values for Pearson's second coefficient of skewness for a normal distribution depending on the number of samples *n*.

n	Table 3, 5% upper in Doane and Seward	Table 2 in Cabilio and Masaro	Table 1 in de Nijs and Klausen
10	0.9629	0.9534-0.9629	0.9661
20	0.7617	0.7480-0.7547	0.7554
30	0.6433	0.6326-0.6381	-
100	0.3669	-	0.3656
Equation:	-	$3P_{95}/\sqrt{n}$	$3\Phi^{-1}(0.95)\beta_n/\sqrt{n}$

References

Bonferroni, C. E. (1930), "Elementi di statistica generale." Seeber, Firenze.

Cabilio P., and Masaro J. (1996), "A simple test of symmetry about an anknown median," *Canadian Journal of Statistics*, 24(3), 349-361.
http://onlinelibrary.wiley.com/doi/10.2307/3315744/pdf

Doane D. P., and Seward L. E. (2011), "Measuring Skewness: A Forgotten Statistic?," *Journal of Statistics Education*, 19(2), 1-28
http://www.amstat.org/publications/jse/v19n2/doane.pdf

Hotelling, H., and Solomons, L. M. (1932), "The Limits of a Measure of Skewness," *The Annals of Mathematical Statistics*, 3, 141-142.

de Nijs R., and Klausen T. L. (2013), "On the expected difference between mean and median," *Electronic Journal of Applied Statistical Analysis*, 6(1), 110-117.
http://siba-ese.unisalento.it/index.php/ejasa/article/view/11468/11411

Pearson, K. (1895), "Contributions to the Mathematical Theory of Evolution, II: Skew Variation in Homogeneous Material," *Transactions of the Royal Philosophical Society, Series A*, 186, 343-414.

Yule, G. U. (1911), "Introduction to the Theory of Statistics," Griffin, London. (Thirteenth Edition, 1944).

Robin de Nijs
Copenhagen University Hospital, Rigshospitalet
Department of Clinical Physiology, Nuclear Medicine & PET
Section 4.01.2
Blegdamsvej 9
DK-2100 Copenhagen, Denmark
Phone: +45-35454011
Fax: +45-35454015
Email: robin.de.nijs@regionh.dk
