

CURRENT RESEARCH IN SOCIAL PSYCHOLOGY

Volume 3, Number 6

Submitted: October 12, 1998

Resubmitted: November 3, 1998

Accepted: November 4, 1998

Publication date: November 12, 1998

A SIMPLE APPROXIMATION FOR THE CALCULATION OF PROBABILITIES IN NORMAL DISTRIBUTIONS

Michael K. Ponton

The George Washington University

[69]

[70]

INTRODUCTION

The frequencies of measured occurrences are often compared to theoretically expected distributions. A common theoretical distribution is represented by the bell-shaped, symmetric normal curve. Many experimental measurements associated with research in the social sciences (Popham & Sirotnik, 1991) and the physical sciences (Anderson, Ball, Murphy & Associates, 1975) have been found to coincide with this normal distribution where the probability of occurrence is given by the area under the curve. The center of symmetry for the normal curve coincides to the mean, median, and mode of the occurrences.

The normal distribution is often presented as a function of standardized scores plotted on the abscissa. One such standard score is the z-score where $z = (x - m)/s$ (Popham & Sirotnik, 1991, p. 32); x , m , and s represent the raw score, mean score, and standard deviation, respectively. The corresponding normal curve is centered about the z-score of zero and all probabilities can be determined by evaluating the area under the curve from zero to the absolute value of the desired z-score making use of the fact that the entire area represents a probability of unity.

DISCUSSION

The computation of the area under the normal curve requires a numerical solution of the error function and can represent hundreds of lines of computer code. To facilitate the computation of this area for probability determinations when a computer or tabulated values are not available, the following approximation can be used with very good results to calculate the area from a z-score of zero to the absolute value of the desired z-score z :

$$\text{Area} = K * \text{sqrt}[1 - e^{-(z^2)/2}], \quad (1)$$

where

$$K = 0.5 + [1/\sqrt{\pi} - 0.5] * [e^{-(z^2)/\sqrt{2\pi}}]. \quad (2)$$

Note that e and Pi are irrational numbers where e represents the base of the system of natural logarithms (i.e., approximately 2.7182818...) and Pi denotes the ratio of a circle's circumference to its diameter (i.e., approximately 3.1415926...). Additionally, the square root operation is designated as "sqrt" and exponentiation is represented by ^. The desired z-score is given by z. Note that equations 1 and 2 can be easily evaluated using most calculators.

[70]

[71]

A comparison between the numerical solution presented by Press, Flannery, Teukolsky and Vetterling (1986) with equation 1 were computed via digital computer producing the following comparisons for area computations:

---- Area Computations ----			
z	Press et al.	Equation 1	Error (%)
----	-----	-----	-----
0.01	0.00398936	0.00398974	-0.0095
0.50	0.19146249	0.19130836	0.0805
1.00	0.34134474	0.34065416	0.2023
1.50	0.43319273	0.43239561	0.1840
2.00	0.47724986	0.47703871	0.0442
2.50	0.49379033	0.49407849	-0.0584
3.00	0.49865010	0.49897581	-0.0653
3.50	0.49976736	0.49993652	-0.0338

As indicated by the data in the percent error column, the agreement between the numerical solution of Press et al. and equation 1 is very good.

CONCLUDING REMARKS

Presented is a simple approximation that can be used, in lieu of a complicated numerical solution, to determine the probability of occurrences for values associated with the normal distribution. The usefulness of this approximation resides in situations where laborious programming is not desired or when tabulated values may not be available. This may occur in purely academic milieus (e.g., classroom settings) or in situations where only simple equations can be programmed on calculators or within spreadsheet software algorithms. Another potential use is in distribution simulations where occurrences coincide with the normal distribution.

Note that the tabulated errors are for the area from zero to some finite z-score. The error in area in the tail region of the normal distribution (i.e., from a finite z-score to infinity) would be extremely large as z gets very large. This is due to the fact that small errors in the approximation represent progressively larger deviations in the small area region of the distribution tail. This tail region is especially important for social scientists in determining whether to reject or fail-to-reject the null hypothesis. Note the following probability differences between calculations (via

calculator) using equation 1 and published values (Beyer, 1981) for z-scores of typical interest in social science research:

[71]

[72]

z	Equation 1	Beyer
1.65	0.0501	0.0495
2.33	0.0097	0.0099
2.57	0.0048	0.0051
3.10	0.0007	0.0010

The calculated probabilities are found by subtracting the area computed using equation 1 from 0.5. Thus, one should be careful in using the approximate solution with a low probability criterion for rejecting the null hypothesis.

To simplify the presented approximation even further, equation 2 may be replaced by equation 3:

$$K = 0.5 + 0.064 * [e^{(-0.4 * (z^2))}]. \quad (3)$$

To reduce the percent error for the range $1 < z < 2$, 0.064 can be replaced with 0.066.

REFERENCES

Anderson, S. B., Ball, S., Murphy, R. T., & Associates (1975). *Encyclopedia of Education Evaluation: Concepts and Techniques for Evaluating Education and Training Programs*. San Francisco: Jossey-Bass Publishers.

Beyer, W. H. (Ed.) (1981). *CRC Standard Mathematical Tables (26th Ed.)*. Boca Raton, FL: CRC Press, Inc.

Popham, W. J., & Sirotnik, K. A. (1991). *Understanding Statistics in Education*. Itasca, IL: F. E. Peacock Publishers, Inc.

Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (1986). *Numerical Recipes*. Cambridge: Cambridge University Press.

[72]

[73]

AUTHOR BIOGRAPHY

Michael K. Ponton is a doctoral candidate in higher education administration within the Graduate School of Education and Human Development at The George Washington University. His doctoral research is focused on developing a psychological instrument that measures an adult's level of personal initiative to engage in autonomous learning. Mr. Ponton is also an adjunct instructor in the Educational Technology Leadership graduate program at GWU in which he co-

teaches a Web-based course on managing computer applications. His e-mail address is ponton@prodigy.net.