THE UNIVERSITY OF TEXAS AT AUSTIN
SCHOOL OF INFORMATION

MATHEMATICAL NOTES FOR LIS 397.1
INTRODUCTION TO RESEARCH IN
LIBRARY AND INFORMATION SCIENCE

Ronald E. Wyllys
Last revised: 2003 Jan 15

THE t-DEVIATION AND z-DEVIATION FORMULAS

Introduction

You can use the Gaussian distribution as an approximation to the distribution of some actual variable that has been observed in a sample or a population. When you do so, you use the pertinent one of the following formulas:

Equation 1. Deviation in a sample

$$t = \frac{X - \bar{x}}{s}$$

Equation 2. Deviation in a population

$$z = \frac{X - \mu}{\sigma}$$

In these formulas t and z represent the value to be used in entering a table of the Gaussian distribution, X represents an observed value (or, more generally, a value of the actual variable whose behavior is being investigated), $\bar{x}$ represents the sample mean (calculated from the observations), s represents the sample standard deviation (also calculated from the observations), and μ and σ represent the population mean and standard deviation, respectively.

An Example of the Use of the Gaussian Distribution

For example, suppose μ = 220 cm and σ = 25 cm. Then if X is an observed value such that X = 190 cm, X can be said to fall 1.2 standard deviations below the mean, for

$$z = \frac{X - \mu}{\sigma} = \frac{190 - 220}{25} = -\frac{30}{25} = -1.2$$

Equations 1 and 2 provide a way of transforming observed sample values into values of the Gaussian distribution, and vice versa. Through such transformations you can use the information about the Gaussian distribution that is embodied in a table of it to provide information about the distribution of the actual variable, by going from observed values to Gaussian-distribution values and then back to real-world values.

The rationale for the transformation involves two steps, which we shall discuss in terms of Equation 1 (analogous reasoning applies to Equation 2).

In the first step, which is embodied in the numerator of the formula, we move the mean from $\bar{x}$, the mean of the original variable, to 0, the mean desired for the new (i.e., the transformed) variable. (We want the new variable to have a mean of 0 in order to match the mean of the standardized Gaussian distribution.) We do this as follows: For each and every observed value X, we subtract the value of the sample mean $\bar{x}$ from X, so that the mean of the resultant set of differences is 0. What this step leaves in the numerator is the *signed* distance between the particular observed value and the sample mean, which can be thought of as a kind of center of the set of observed values. (The *signed* distance is the distance with a plus, +, or minus, -, sign attached to it. The plus sign is to be understood to be present if there is no sign marked.)

In the second step, we change the units in which the signed distance in the numerator is expressed. Specifically, we move from the unit of measurement that was used in the original observations (e.g., inches for height measurements) to a new unit, the standard deviation of the Gaussian distribution. We make the change by dividing the numerator by s, expressed in the original units. The transformation process is exactly the same as that involved when we change from one unit of physical measurement to another.

For example, suppose we had weighed a loaf of bread and found that its weight was 24 ounces. If we wanted to move from the old unit, ounces, to the new unit, pounds, to express the weight of the loaf, we would divide 24 by 16 to obtain the answer, 1.5 pounds. The key is to recognize that the transformation from old to new units involves dividing the observed number of old units by the number of old units in one new unit. In the example, we used the fact that in one pound (the new unit) there are 16 ounces (the old unit). The relationship may be written as

$$\text{measurement expressed in new units} = \frac{\text{measurement expressed in old units}}{\text{number of old units in one new unit}}$$

which should remind you of Equations 1 and 2. Thus, if we divide the signed difference in the numerator of Equation 1 by the number of old units in one standard deviation, we obtain a value of t that is expressed in the new unit, the standard deviation.

The overall result of Equation 1 is the transformation of the observations from the original units of measurement into new values, t, expressed in the new unit of measurement, the standard deviation, with the t-expressed observations centered on 0. Such t values, so centered, are what are needed to use tables of the Gaussian distribution (or, in the case of small samples [especially, samples of size 30 or smaller], to use tables of the Student's t distribution).

As we noted earlier, analogous reasoning applies to using Equation 2 to deal with observations of an entire population, as opposed to observations of a sample. Of course, most of the time in statistical work, we are dealing with observations of a sample and with using those observations to make suitable inferences about the nature of the population.