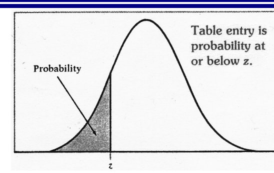
TAKE IT TO THE MAT

A NEWSLETTER ADDRESSING THE FINER POINTS OF MATHEMATICS INSTRUCTION

Regional Professional Development Program
February 25, 2002 — High School Edition

rpdp.ccsd.net

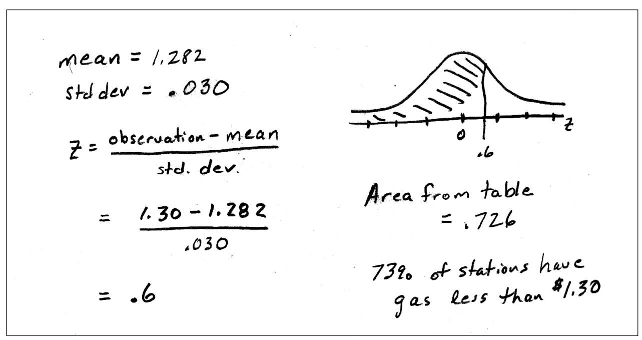


This issue of *Take It to the MAT* will look at the *normal distribution*, a topic present in many pre-calculus courses as well as statistics and probability courses. In particular, we will look at a common mistake that students make as a result of misconceptions about the normal distribution.

Consider the following exercise: A statewide study of service stations for the week of February 25, 2002 found that the mean per gallon price of gasoline was $1.282 with a standard deviation of $0.030. What percentage of service stations has gasoline prices below $1.30 per gallon?

The typical statistics student will believe that since the raw data is not given, some table or calculation may need to be used. A table for cumulative probabilities under the standard normal distribution is shown at the right. Each value in the second column is the probability an observation will be at or below the corresponding *z*-score. Most students will make use of this table. One student's solution to the exercise is shown below.

Z	Prob
–3.0	0.001
–2.9	0.002
–2.8	0.003
–2.7	0.003
–2.6	0.005
–2.5	0.006
–2.4	0.008
–2.3	0.011
–2.2	0.014
–2.1	0.018
–2.0	0.023
–1.9	0.029
–1.8	0.036
–1.7	0.045
–1.6	0.055
–1.5	0.067
–1.4	0.081
–1.3	0.097
–1.2	0.115
–1.1	0.136
–1.0	0.159
–0.9	0.184
–0.8	0.212
–0.7	0.242
–0.6	0.274
–0.5	0.309
–0.4	0.345
–0.3	0.382
–0.2	0.421
–0.1	0.460
0.0	0.500
0.1	0.540
0.2	0.579
0.3	0.618
0.4	0.655
0.5	0.691
0.6	0.726
0.7	0.758
0.8	0.788
0.9	0.816
1.0	0.841
1.1	0.864
1.2	0.885
1.3	0.903
1.4	0.919
1.5	0.933
1.6	0.945
1.7	0.955
1.8	0.964
1.9	0.971
2.0	0.977
2.1	0.982
2.2	0.986
2.3	0.989
2.4	0.992
2.5	0.994
2.6	0.995
2.7	0.997
2.8	0.997
2.9	0.998
3.0	0.999



The solution seems reasonable enough. The student correctly calculated the *z*-score for $1.30 to be 0.6, correctly read the table, and correctly wrote that about 73% of the area under the normal curve, and thus gas priced under $1.30, is to the left of $z = 0.6$. The problem is that it is wrong—for two reasons.

The first reason is that the student used the normal distribution. Nowhere are we told that the distribution of gas prices is normal or even approximately normal. But students do this quite often. They do a bevy of normal curve problems from their textbook and then assume everything in the universe is normally distributed.

The second reason is that the normal distribution is a *continuous* distribution. It is unlikely that our data could be considered continuous—it is pretty *discrete*. It tends to have only certain values characterized by "jumps" in the values, i.e. $1.199, $1.209, $1.219, etc. Even though the mean does not end with nine tenths of a cent, it is likely that all of the data do. Thus, we may have slightly over-estimated our percentage by using $1.30 instead of $1.299.

The second error is the lesser of the two. Using continuous distributions to analyze discrete data can lead to errors, but usually only minor ones in the approximation. But, the first error is deadly serious. If one doesn't know that data (or a population) is normally distributed, one cannot assume it is!