

1. The Standard Normal Distribution

Used to approximate or describe histograms of many (but not every) types of data. Properties are:

- a. Symmetric, bell-shaped, the "bell curve", see page A-105 of your textbook.
- b. Mean 0, SD 1
- c. The median is where 50% (half) of the observations are on either side. In this distribution, the mean is equal to the median. The values on the horizontal axis are called "Z SCORES" or "STANDARD UNITS". Values of Z above the average are positive, values of Z below the average are negative.
- d. Area under the curve is equal to 100% when expressed as a percentage. The shaded area under the curve represents the percentages of the observations in your data between given values of Z.
- e. 68%-95%-almost 100% rule (see p. 64) About 68% fall within plus or minus 1 SD of the mean About 95% fall within plus or minus 2 SD of the mean Nearly 100% (99.7%) fall within plus or minus 3 SD
- f. The curve never crosses the horizontal axis, it gets very close at the extremes though. It extends to negative and positive infinity.

2. Standard (Deviation) Units

A score z is in STANDARD UNITS if it tells how many SD's the original score is above or below the average. For example, if $z=1.3$, then the original score was 1.3 SD's above average; if $z = -0.55$, then the original score was 0.55 SD's BELOW average. The formula for converting data from original units to Z scores is:

$$z = \frac{(\text{value of interest} - \text{average of all the values})}{\text{standard deviation of all the values}}$$

you can call this a "normal calculation"

3. Examples of the use of Standard Units

See handout on heights and weights of models. The average height of US Women over age 18 is 65 inches (about 5'5") with a standard deviation of about 3 inches. Elle McPherson is a fashion model, she is listed at 6' tall (72") using standard units we can compare her to the average US woman

$$z = \frac{(72 - 65)}{3} = \frac{7}{3} = 2.33 \text{ we can round this to the nearest } z \text{ score or } 2.35$$

or she has a Z score of 2.35. She is 2.35 standard deviations above average in height. She is taller than (98.12% + .94%) or 99.06% of all adult women. See Table A 105, look up 2.35, ignore the 2.52 next to it and focus on the 98.12. The 98.12 is the shaded area or the total area (percentage) between a z-score of -2.35 and +2.35. By symmetry, we know that (100-98.12=1.88) so 1.88% remain in the two unshaded tails, the lower tail is 1.88/2 or .94%, add it to the 98.12 to get 99.06

4. Converting Standard Units back to original values

Idea: suppose you are told that the minimum height for a career in fashion modeling is $Z=1.15$, how tall is that? - $1.15 = (\text{value of interest} - 65)/3 = \text{about } 68.45 \text{ inches or about } 5' 8.5'' \text{ tall.}$

5. Why bother with Standard Units?

Standard Units allow quick comparisons across variables with different units of measure. For example, suppose all the test scores in a class are normally distributed. The first test was worth 45 points, the mean was 33, the standard deviation was 5. A student received a 40 on that test and was told it was an A-, her Z score was

$$Z = (40 - 33) / 5 = 1.40$$

The next test had a mean of 58 and standard deviation of 8. If she were to do as well on the second test as she did on the first (that is get an A-), what would her new score need to be?

$$Z = 1.40 = (\text{new score} - 58) / 8$$

or she would need to score a 69.2 or something around a 69 to 70. The thing to remember is this: the second test has more points, a different mean, and a different standard deviation – it's different, but if we convert the "raw" scores to standard scores (Z), comparisons are easy. In a way, it lets us compare apples and oranges.

6. Assessing Normality

A. Common sense: if the normal curve implies nonsense results (for example, that people have negative incomes, or that some women have a negative number of children), the normal curve doesn't apply and using the normal curve will give the wrong answer.

B. Construct a histogram: if the data look like a normal curve, the normal curve probably applies; otherwise, it does not.

C. Do the data fall in a 68-95-99.7% pattern? If yes, normality is probably being met.