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 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 = (value of interest - average all the values)

standard deviation of all the values

you can also think of this as a "normal calculation"

3. Examples of the use of Standard Units

See handout on heights and weights of various people. The average height of an adult woman in the United States is about 64 inches (5'4") with a standard deviation of 2.5 inches. Cameron Diaz is an actress and is 5'9" tall. Her height in standard units is 2.0 or she has a Z score of 2.0. She is 2 standard deviations above average in height. She is taller than (95.45% + 2.275%) 97.725% of all women in the U.S. See Table A 105, look up 2.0, ignore the 5.40 next to it and focus on the 95.45. The 95.45 is the shaded area or the total area (percentage) between a z-score of -2.0 and +2.0. By symmetry, we know that (100-95.45=4.55) 4.55% are left in the two unshaded tails, the lower tail is 4.55/2 or 2.275%, add it to the 95.45 to get 97.725

4. Converting Standard Units back to original values

Idea: suppose you are told that a woman's standardized height is Z=-1.56, how tall is she? -1.56 = (value of interest - 64)/2.5 = 60.1 or about 5 feet tall.

5. Why bother with Standard Units?

Standard Units allow quick comparisons across variables with different units of measure. Z scores or standard units have no units, everything you convert is standardized. For example, Cameron Diaz 118lbs, that's a Z score of -.6 or she is lighter than 45.15 + 27.425 = 72.575% of all women. If she was as heavy as she is tall, in other words if she were to weigh more than 97.725% of all women she would need to be:

2.0 = (value of interest - 136) / 30

or the value of interest = 196 lbs. That is, she'd need to gain 78lbs to be as heavy as she is tall. Not that we particularly care about her weight, but the normal table is a tool that can be used with a variety of datasets.

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.

Example: in 1980, the average number of children born per woman was 1.95, with an SD of 1.91. Does the normal curve apply? Try calculating how many children a woman would have if she is 2 standard deviations BELOW the mean.

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.