MODULE 12

PROBABILITY INTRODUCTION

PROBABILITY is the "language" used to describe the proportion of times that a random event will occur. The language of probability is at the center of statistical inference (see Modules 14 and 15). Only a minimal understanding of probability is required to understand most basic inferential methods, including all of those in this course. Thus, only a short, example-based, introduction to probability is provided here.[1]

The most basic forms of probability assume that items are selected randomly. In other words, simple probability calculations require that each item, whether that item is an individual or an entire sample, has the same chance of being selected. Thus, in simple intuitive examples it will be stated that the individuals were "thoroughly mixed" and more realistic examples will require randomization.[2]

If every individual has the same chance of being selected, then the probability of an event is equal to the proportion of items in the event out of the entire population. In other words, the probability is the number of items in the event divided by the total number of items in the population.

For example, the probability of selecting a red ball from a thoroughly mixed box containing 15 red and 10 blue balls is equal to $\frac{15}{25} = 0.6$ (i.e., 15 individuals ("balls") in the event ("red") divided by the total number of individuals ("all balls in the box"); Figure 12.1-Left). Similarly, the probability of randomly selecting a woman from a room with 20 women and 30 men is 0.4 ($= \frac{20}{50}$; Figure 12.1-Right). In both examples, the calculation can be considered a probability because (i) individuals were randomly selected and (ii) a proportion of a total was computed.

[1]A deeper understanding of probability is required to understand more complex inferential methods beyond those in this course.
[2]See Module 3 for methods to randomly select or allocate individuals.

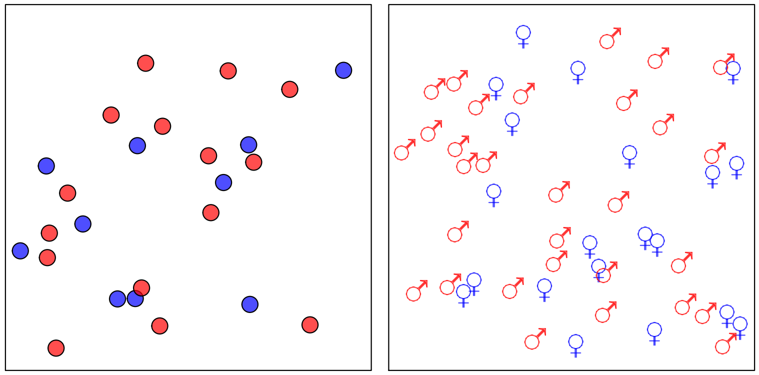

Figure 12.1. Depictions of a ‘box’ with 15 red balls and 10 blue balls (Left) and a ‘room’ with 30 men and 20 women (Right).

The two previous examples are simple because the selection is from a small, discrete set of items. Probabilities may be computed for a continuous variable if the distribution of that variable is known for the entire population. For example, the probability that a random individual is greater than 71 inches tall can be calculated if the distribution of heights for all individuals in the population is known (or reasonably approximated). For example, as shown in Module 8, if it can be assumed that heights is $N(66, 3)$, then the proportion of individuals in the population taller than 71 inches tall is 0.0478 (Figure 12.2).[3] This result is a probability because (i) the individual was randomly selected and (ii) the proportion of all individuals of interest in the entire population was found.

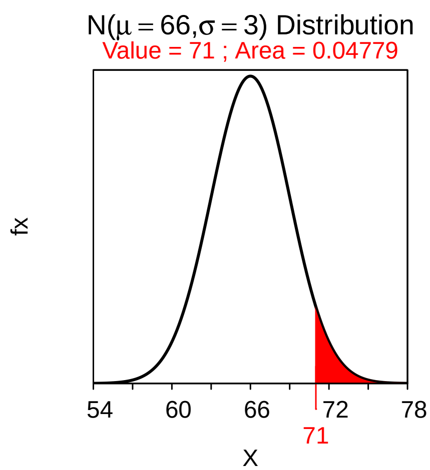

Figure 12.2. Calculation of the probability that a randomly selected individual from a $N(66, 3)$ population will have a height greater than 71 inches.

A theory that explains the distribution of statistics computed from all possible samples from a population will be developed in Module 13. This distribution will be used to compute the probability of observing a particular range of statistics from random samples. This technique is the basis for making statistical inferences in Modules 14 and 15.

[3]As computed with `distrib(71,mean=66,sd=3,lower.tail=FALSE)`.