MODULE 11

PROBABILITY INTRODUCTION

Objectives:

1. Identify the two major assumptions for computing basic probabilities.
2. Calculate basic probabilities in discrete item cases.
3. Calculate basic probabilities for continuous variables that follow a normal distribution.

PROBABILITY is the "language" used by statisticians 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 13 and 14). Only a minimal understanding of probability is required to understand most basic inferential methods, including all of those in this course. Thus, only a very 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 "box of balls was thoroughly mixed" and more realistic examples will require randomization.[2]

> ◇ **Individuals must be randomly selected from the population or samples must be produced randomly for the concept of probability to work accurately.**

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 $\frac{15}{25} = 0.6$

[1]A deeper understanding of probability will be required to understand more complex inferential methods beyond those in this course.

[2]See sections in Module 3 for methods of selecting or allocating random individuals.

(i.e., 15 individuals ("balls") in the event ("red") divided by the total number of individuals ("all balls in the box"). Similarly, the probability of randomly selecting a woman from a room containing 20 women and 30 men is 0.4 ($=\frac{20}{50}$). In both of these examples, the calculation can be considered a probability because (i) individuals were randomly selected and (ii) a proportion of a total was computed.

◇ If every item has the same chance of being selected, then the probability of observing an item with a certain characteristic is the proportion of items in the entire population that have that characteristic.

The two previous examples are rather simple examples where the selection is made from a small, discrete number of items. Probabilities can also be computed for continuous variables if the distribution of that variable for the entire population is known. 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. Of course, information about the population is typically difficult to know. However, in many situations, a normal distribution may be used as a model of a population distribution. For example, as shown in an example in Module 7, 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.[3] This result can be considered a probability because the proportion of all individuals of interest in the entire population was found and the individual was randomly selected.

◇ The calculations from the normal distribution made in Module 7 are probability calculations as long as the individuals are randomly selected.

A theory that explains the distribution of statistics computed from all possible samples from a population will be developed in Module 12. This distribution will be used to compute the probability of observing a particular range of statistics in a random sample of individuals. This technique is the basis for making statistical inferences in Modules 13 and 14.

Review Exercises

11.1 A coin purse contains 17 nickels and 15 dimes. Use this to answer the questions below. *Answer*

(a) What is the probability of randomly selecting a nickel from this purse?
(b) What is the probability of randomly selecting a dime from this purse?
(c) What is the probability of randomly selecting a dime from this purse assuming that two nickels and three dimes have already been removed?

11.2 A very small green house contains 10 tomato, 12 pea, and 8 cauliflower plants. Use this to answer the questions below. *Answer*

(a) What is the probability of randomly selecting a tomato plant from this greenhouse?
(b) What is the probability of randomly selecting a cauliflower plant from this greenhouse?
(c) What is the probability of randomly selecting a pea plant from this greenhouse assuming that all tomato plants had died and were removed from the greenhouse?

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

11.3 Suppose that the length of all needles on a particularly large pine tree is known to be normally distributed with a mean of 75 mm and a standard deviation of 8 mm. Use this to answer the questions below. *Answer*

(a) What is the probability that a randomly selected needle is between 70 and 80 mm long?
(b) What is the probability that a randomly selected needle is longer than 90 mm?
(c) What is the probability that a randomly selected needle is less than 50 mm long?