1 Section 7.4: Mean

When doing some experiment, there is usually some set of objects we want to know something about. For example, all people who live in the United States, all cars made by Ford in the last 10 years, or all students at UVA. We call these sets a **Population**. Often times the population is far too large to handle directly, so we find some subset of the population called a **Sample**. We will use the samepele to try to learn some information about the population.

One common question we might ask is "what is the average of some characteristic for the population?" For example, what is the average height of a person in the U.S., how long will the average Ford last, or how many students does the average UVA student take per semester? If we don't have access to the whole population data, then we can take a sample and compute the **sample mean**. The sample mean or average of n numbers is given by

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

Similarly, the **population mean** is given by

$$\mu = \frac{x_1 + x_2 + \cdots + x_N}{N}$$

where there are N elements in the population.

These things may seem the same but there is a subtle difference between them. The Population mean will never change, it is some fixed constant. On the other hand, the sample size may change depending on what sample we pick. To this end we have the following definitions. A numerical discriptor made on the Population is called a **Parameter**. A numberical discriptor made on a sample is called a **statistic**. So sample mean is a statistic and population mean is parameter. We use statistics to try to learn about parameters.

What if our data was organized so that each oucome was grouped together. For example, say we asked people how many movies they had seen in the last wekk. Instead of keeping track of everyone's individual response, we could just keep a tally of how many people saw $0, 1, 2, \ldots$ and so on. So we would get results $x_1, \ldots, x_r$ and each response would have a frequency $f_1, \ldots, f_r$. Then we could find the mean by computing

$$\bar{x} = \frac{x_1 f_1 + \cdots + x_r f_r}{n} = x_1 \left(\frac{f_1}{n}\right) + \cdots + x_r \left(\frac{f_r}{n}\right)$$

so notice that we multiply each outcome by it's relative frequency.

So far we have been dealing with emperical data where we are used to the ideas of average. But what do we mean by the "average" outcome of a theoretical experiment or a random variable. Suppose X is a random variable with probability distribution given by

k	$P(X = k)$
x_1	p_1
$\vdots$	$\vdots$
x_r	p_r

Then we define the **Expected Value** of X to be

$$E(X) = x_1 p_1 + x_2 p_2 + \cdots + x_r p_r$$

So we weight each outcome with how likely it is. Notice the similarity to the sample mean when we used the relative frequencies. Also note that if all outcomes are equally likely, then the expected value is also the mean of the outcomes.

What does expected value mean? First of all, it is not the most likely outcome, because it may not even be an outcome. Expected value represents the number which should be the average value of the outcomes if repeated many times. It is a measure of what will happen on average.

Example: Let X be the outcome of a fair 8 sided dice roll. Find $E(X)$.

$$E(X) = 1\frac{1}{8} + 2\frac{1}{8} + \cdots + 8\frac{1}{8} = 4.5$$

Example: A Virginia match 3 lottery ticket costs \$1. There is a 1/1000 chance of guessing correctly. If you do guess the number correctly then you win \$500. Find the expected value of playing the lottery.

Let X be the amount of money you win from playing. Then the possibilites for X are -1 and 499 with associated probabilities $999/1000$ and $1/1000$. So we get

$$E(X) = -1\frac{999}{1000} + 500\frac{1}{1000} = -\frac{499}{1000}$$

so you expect to lose about $0.50 every time you play. (not the best value)

The VA state lottery posts all of this information for all of its games and you can compute the expected value for each of them. None of them will be positive.

Some properties of expected value: If X and Y are independent random variables then

1. $E(X + Y) = E(X) + E(Y)$

2. $E(aX) = aE(X)$

3. $E(XY) = E(X)E(Y)$

We can use this (or direct computation) to see that if X is a Binomial random variable with parameters p and n then $E(X) = np$. This makes sense since we expect to succeed p of the time.

We say a game is "fair" if the expected value is 0. Determine if the following games are fair:

1. Flip 2 coins, if they are different you win a dollar and if they are the same you lose a dollar.

2. Roll a 4 sided dice. If you roll a 1 you lose $6. If you roll a 2 or 3 you win $1. If you roll a 4 you get $3.

3. Bet any amount of money you want then draw two cards. If you get a pair then you double your money. Otherwise you lose your bet.