**Classwork 12**

**Probability Distributions**

Introduction: Take out your TI calculator and enter the following datasets in L1 and L2, respectively,

|  |  |
| --- | --- |
| **X =** *x* | p(**X**  = *x*) |
| *0* | *0.53* |
| *1* | *0.19* |
| *2* | *0.09* |
| *3* | *0.06* |
| *4* | *0.03* |
| *5* | *0.03* |
| *10* | *0.06* |

This table represents the probability distribution of a random variable **X**, the number of minutes late to class a randomly selected student in MAT 135 will be to class. In other words, the left column lists the possible outcomes for a randomly selected student (we assume students are not allowed into class after a 10 minute grace period; a policy I luckily don't adopt.) and the right column lists the corresponding probability of observing a student that is late by that much. For example, the probability of a student being 4 minutes late is 0.03 = 3%.

In the L3 and L4 lists enter the following datasets,

|  |  |
| --- | --- |
| **X =** *x* | p(**X**  = *x*) |
| *-$5* | *0.98* |
| *$4995* | *0.02* |

This table represents the probability distribution for a random variable **X**, the amount of money won on a game of chance that costs $5 to play with a chance of winning $5000. In other words, you win lost $5 with a probability of 0.98 and a win $4995 with a probability of 0.02.

Definitions

*Random Variable*: **X**

A random variable is a way of assigning a numerical value to any outcome in any experiment. For example, if you a flipping a coin, then the outcomes are heads or tails. A random variable for this experiment would assign a value of 0 to heads and a value of 1 to tails. The technical term for a random variable with two outcomes, 0 and 1, is known as a *Bernoulli random variable.*

*Probability Distribution: p*( **X** = *x* ) or *p* ( *x* ) for short.

A probability distribution assigns a probability to every value the random variable can take on for a given experiment. In essence, a probability distribution is a relative frequency distribution with one distinct difference: a probability distribution represents the *distribution of the population.* In other words, if we are given a probability distribution, we are given information about the *set of all observable outcomes for a given experiment.*

For example, if we have a random variable that measures the outcome of flipping a fair coin, (fair means each outcome is equally likely) assigning a value of 0 to heads and a value of 1 to tails, then its probability distribution would look like,

|  |  |
| --- | --- |
| **X =** *x* | p(**X**  = *x*) |
| *0* | *0.5* |
| *1* | *0.5* |

*Expected Value:*

The expected value is the most likely value to observe from a given population. Note, the formula for expected value is very similar to the formula for the sample mean stated in terms of frequencies,

In fact, the sample mean and the expected value are basically the same quantity. That is because we can interpret the relative frequency as a sample estimate of the population probability,

In other words, the formula for the sample mean can be restated as,

Which shows the sample mean is the sample estimator of the population's expected value. For this reason, we call the expected value of the population the *population mean*.

As an example, the expected value of a random variable that measures coin flips with a 0 or 1 would be,

Note, the expectation does not have to be an observable quantity. It merely represents the "center" of likely outcomes.

Sometimes the *expected value* is also called the *expectation*.

The First Law of Probability (Stated With Probability Distributions)

If a random variable can be assigned a value for every event in a sample space, then the sum of the random variable's probabilities should be 1. In other words, *something has to happen with a 100% probability.*

**Problems**

1. Verify the sum of probabilities for each of the given distribution is 1. Write down the sequence of instructions you need to perform this operation on your calculator:

a. Does each distribution represent a probability distribution?

2. Recall the probability distribution stored in L1:L2 represents the probability of a student being late to class. What is the expected amount of time the average student in MAT 135 will be late to class?

*Hint*: You will need to create a new list in L3 by using a formula to multiply the elements of L1 and L2. Then sum up the elements of the new list. In other words, apply the formula for an expected value!

a. What is the probability a student in MAT 135 is less than three minutes late to class?

b. What is the probability a student in MAT 135 is at least 4 minutes late to class?

3. The second probability distribution stored in L3:L4 represents the probability distribution for a game of chance. To play you must pay $5. You are asked to pick a number between 1 – 50. A number is then randomly selected from a random number generator. If you chose the number selected, you win $5000 dollars, minus the $5 you paid to play.

a. If you play this game a large number of times, how much on average can you expect to win or lose?

b. Suppose you own a casino which has a game like this. How much should you charge to break even in the long run?

*Note*: You will need to set this one up by hand to solve it!

Another Definition

*Variance*:

The variance of a probability distribution quantifies the variation of a population. The standard deviation of the population is related to the variance through the familiar formula,

The sample standard deviation is the sample estimator of the population standard deviation. Note, the formula for population variance is essentially the formula for sample variance, but stated in terms of a probability distribution. The quantity,

is simply the squared deviations of each observation from the population. Recall the expected value is just the population mean. By subtracting it from the value x, we are calculated a deviation. The formula then weights each deviation by its chance of occurring. The variance is just the expected value of the squared deviations!

**Problems**

4. Calculate the population standard deviation for the number of minutes late to class a randomly selected student in MAT 135 will be on a given class day.

You will need to subtract the expected value calculated in #2 from each element stored in the L1 list (i.e. the population) and then square it and multiply by the corresponding probability. In other words, you will need to create a new list and enter in the formula for population variance using the appropriate list variables and the **sum** list operation.

The formula for the new list should read,

= (L1 – sum(L3))^2) \* L2

where L1 is the list storing the population values, L2 is the list storing their corresponding probabilities of being observed and L3 is the list created in #2 for calculating the expected value.

After you have created this new column in L4, take the square root of its sum to find the population standard. Write its value in the space provided below: