**Chapter 7 – Random Variables**

A \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ is a numerical variable whose value depends on the outcome of a chance experiment.

A random variable is \_\_\_\_\_\_\_\_\_\_\_\_\_ if its set of possible values is a collection of isolated points on the number line.

A random variable is \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ if its set of possible values includes an entire interval on the number line.

Example: shoe size vs. foot length

Example: age vs. number of birthdays

A \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_, or \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_, of a discrete random variable X lists the possible values of X and their probabilities. This is usually done in a table, probability histogram, or formula.

Let X = number of heads in two flips of a coin. Then, the probability distribution of X is:

 table: histogram:

The probabilities in a distribution give information about the \_\_\_\_\_\_\_\_\_\_\_\_\_\_ behavior of the variable. That is, probabilities are \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_. For example, if P(X = 2) = .25, then after many observations of X, the value X = 2 will occur about 25% of the time, on average.

Properties of a discrete random variable X:

 1. 0 ≤ P(X) ≤ 1

 2. 

Let X = the number of shots I make in 3 attempts.

 

Calculate the probability that

* I make all 3 shots: *P(X = 3) =*
* I make less than half of my shots *P(X < 1.5) =*
* I make at least one shot *P(X ≥ 1) =*

Note: It is important that you always define the variable and restate the question in terms of the variable.

The \_\_\_\_\_\_\_\_ of a random variable X, denoted , describes where the probability distribution of X is centered.

Let X = number of license plates on a randomly selected car. The probability distribution of X is shown in the table below. Find the mean value of X.

|  |  |  |  |
| --- | --- | --- | --- |
| X | 0 | 1 | 2 |
| P(X) | .05 | .25 | .7 |

Thus, to find the mean of a discrete random variable X, multiply each X value times its probability and add them all together.

 

The term \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_, E(x) is sometimes used in place of mean value, .

The expected (mean) value will not necessarily be equal to one of the values in the distribution. Instead, we interpret the expected value as a long term average. For example, if we were to select many cars, we would expect there to be 1.65 license plates per car, on average.

Suppose I ask you to play a betting game with me. You bet $5 and if you draw an ace of hearts, I will pay you $100. If you get any other ace, I will pay you $10 and if you get any other heart, I will give you your money back. If you get any other card, I keep your money.

1. Create a probability model for the amount you will win.
2. Should you play this game?

Now that we can find out where the distribution of a random variable is centered, we would also like to know how far the values are from the center, on average. This, of course, is the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_!

Using the license plate distribution from earlier, if we had a sample of 100 cars, we would expect 5 0’s, 25 1’s, and 70 2’s.



On average, the number of license plates on a car will be 0.57 from the mean value of 1.65.

Note: We divide by n instead of n-1 and use the symbol  instead of  since we are not calculating the standard deviation of a sample. There is no uncertainty about the values in the distribution or their probabilities.

Note: It is calculated in much the same way as the SD of a data set, except that we weight each value with its probability, since the values that show up more often should have more influence in the calculation.

Thus, the standard deviation of a *discrete* random variable is:



The VARIANCE of a *discrete* random variable is the square of the standard deviation:

 

Find and interpret the SD for the betting game from earlier:

In a certain class, there are 15 girls and 10 boys. Suppose that you are randomly selecting 2 students for a committee. Create a probability model for the number of girls in the sample. Then, calculate and interpret the mean and SD of the number of girls.

**Random Variables**

Now that we know how to find the mean and standard deviation of a random variable, we would also like to know the mean and standard deviation of a *function* of that random variable.

If x = the length of a taxi ride in miles, suppose that  = 5.2 miles and  = 2.8 miles. Two taxi companies have different fare-schedules: company 1 charges $2.50 per mile and company 2 charges $2 per mile plus an initial fee of $5. Thus, if we are interested in the total fare of a taxi ride (y), the functions are:

****

How can we find the mean and standard deviation for y in each case?

Recall from chapter 6 that adding a constant to every value in a data set changes the measures of position (mean, median, min, max, Q1, Q3, etc.) BUT NOT the measures of spread (standard deviation, range, IQR).

However, multiplying every value in a data set by a constant changes both the measures of position AND the measures of spread.

In addition to paying the fare, the passenger usually gives the driver a tip (t). Assume that  = 1.50,  = .70, and that the tip does not depend on the fare (that is, t and y are independent).

What are the mean and standard deviation of the new variable c, the total cost of a trip, including tip?

The mean and variance of a combination of random variables:

If x and y are random variables with means  and standard deviations , then:

 

Note: the mean formulas work regardless of whether x and y are independent.

  thus 

Note: the variance (SD) formulas work ONLY when x and y are independent.

Note: Always remember that the variances add, not the standard deviations.

Note: Even when we are subtracting the random variables, we add their variances. More variables always means more variability!

Note: All 4 of these rules work for combining more than 2 random variables as well.

Thus,  for company 1:

Find  for company 2.

A company selling vegetable seeds in packets of 20 estimates that the number of Basil seeds that will actually grow has a mean of 15 with a SD of 2 and the number of Cilantro seeds that will actually grow has a mean of 14 with a SD of 3.

1. If you buy one pack of each, how many seeds should you expect will grow? What is the SD?
2. If you buy one pack of each, what is the expected value of the difference in the number of seeds that will grow? What is the SD?
3. If you buy 3 packs of Basil seeds, how many seeds should you expect will grow? What is the SD?
4. If you buy 3 packs of Basil and 2 packs of Cilantro, how many seeds should you expect will grow? What is the SD?

**Combining Normal Random Variables**

Any linear combination of normally distributed random variables is also normally distributed.

For example, if x and y are normally distributed then (x + y), (x – y), (2x + 3y), etc. are also normally distributed.

If men’s weights (in pounds) are N(190, 40), women’s weights are N(145, 30), and a man and woman are selected at random, what is the probability that the sum of their weights is at least 300 pounds? What assumptions must you make to do this problem?

What is the probability that the woman is heavier than the man?

If you were to select 3 men at random, what is the probability that the sum of their weights is at most 500 pounds?

**Using the TI-84 to find the Mean and SD of a discrete RV:**

Finding the mean and standard deviation of a discrete random variable (x) on the TI-83:

* Enter the values of x in list 1 and their probabilities in list 2.
* Stat: Calc: 1-Var Stats L1, L2 (Note: this is the process we use to find the mean and standard deviation of a frequency distribution. In this case, we are using probabilities as relative frequencies.)

Suppose the grades for Mrs. Hineman’s students on the AP Stats exam (x) have the following probability distribution. Find the mean and standard deviation of x.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| x | 1 | 2 | 3 | 4 | 5 |
| p(x) | .04 | .16 | .40 | .21 | .19 |

What proportion of people score within 1 standard deviation of the mean?

Within 2 SDs?

If I were to randomly select 3 statistics students, what is the mean and standard deviation of the sum of their scores?

If I were to randomly select 2 students, what are the mean and standard deviation of the difference in their scores?

If one randomly selected student took 2 different versions of the AP Exam, find the mean and SD of the sum of his two scores.

*Note: You must show some work to get credit for expected value calculations on the AP Exam.*