Unit 5 – Discrete Probability Distributions

Your Vacation Professor Colton

Unit 5 - Outline

Unit 5 – Discrete Probability Distributions

Intro

Discrete Probability Distributions

- Random Variables
- Discrete Probability Distributions
- Expected Value (Mean) and Standard Deviation

Random Variables

Random

- In statistics, the word **random** has a different meaning. Something is random when it varies by chance.
  - For example, when rolling a six-sided die there are six equally possible outcomes, the observed outcome on any one roll is random.

Random Variables

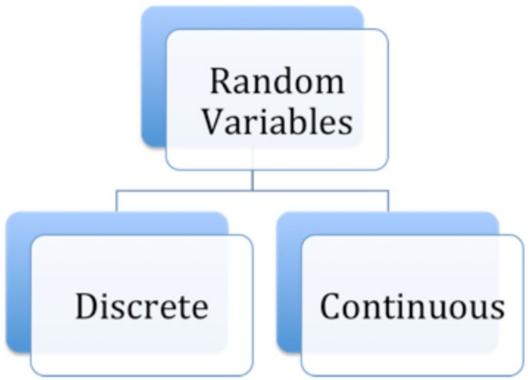
- A **random variable** is a numerical characteristic that takes on different values due to chance. Random variables are classified into two broad types: **discrete** and **continuous**.

Discrete Random Variable

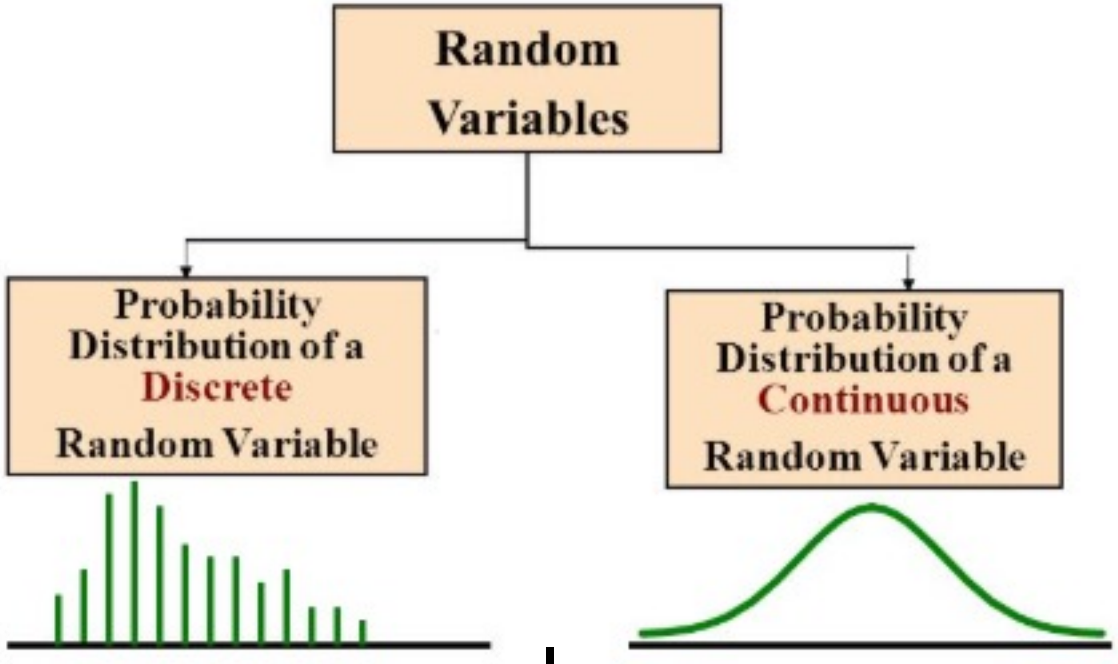
- Has a countable set of distinct possible values. Examples include:
  - Number of heads in 4 flips of a coin (possible outcomes are 0, 1, 2, 3, 4)
  - Number of classes missed last week (possible outcomes are 0, 1, 2, 3,..., up to the maximum number of classes)
  - Number of loans approved by a bank last week

Continuous Random Variable

- Any value (to any number of decimal places) within some interval is a possible value. Examples include:
  - Heights of individuals
  - Time to finish a test
  - Hours spent exercising last week



Discrete vs Continuous RVs



X	*P(X)*
Win = 3	0.3
Tie = 1	0.5
Loss = 0	0.2

- Can find probabilities for specific points, e.g. P(5) = 0.3, P(10) = 0.6
- Can find probabilities of multiple events, such as:
  - P(5 < X < 10) = P(6) + P(7) + P(8) + P(9)
  - Just adding up individual probabilities from the probability distribution table.

- Can NOT find probabilities for specific points, e.g. P(5) = 0, P(10) = 0
- Have to find probabilities for intervals:
  - P(5 ≤ X ≤ 10) or like in the Empirical Rule
  - This is finding the area under the curve between the end points.

https://medium.com/@seema.singh/probability-distributions-discrete-and-continuous-7a94ede66dc0

Discrete Probability Distributions

Discrete Probability Distributions

- Describes a **Discrete** random variable
- A collection of all possible values the RV can assume and all the associated probabilities
- A probability distribution is a theoretical description of a population

Outcome, X:	1	2	3	4	5	6	*Total*
Probability, P(X):	1/6	1/6	1/6	1/6	1/6	1/6	1

Properties

1. Each outcome has a probability between 0 and 1 (0% and 100%), $0 \leq P(x) \leq 1$.
   - This probability denotes the chance of that event occurring.
   - Using interval notation P(x) = [0,1] (inclusive)

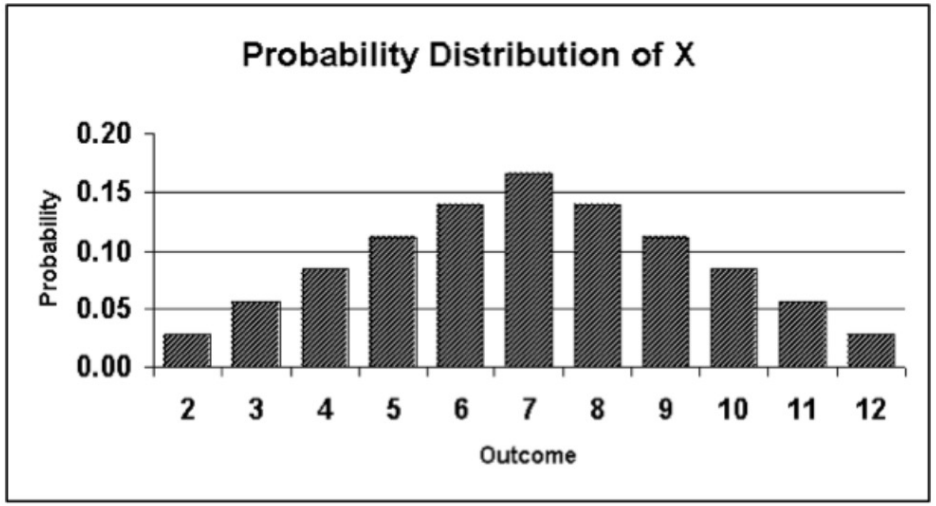
2. The sum of all outcome's probabilities is equal to 1, $\sum P(x) = 1$

Plotting a Discrete Probability Distribution

- To graph the probability distribution of a discrete random variable, construct a probability histogram:

Example: Let *X* represent the sum of two dice. Then the probability distribution of X is as follows:

x	2	3	4	5	6	7	8	9	10	11	12
P(x)	1/36	2/36	3/36	4/26	5/36	6/36	5/36	4/36	3/36	2/36	1/36



Expected Value Motivation

Lets say we were calculating the mean like usual of the following numbers: 1, 2, 3

$\bar{x} = \frac{1+2+3}{3} = 2$ we can rewrite this as $\bar{x} = \frac{1}{3}(1 + 2 + 3) = \frac{1}{3}(1) + \frac{1}{3}(2) + \frac{1}{3}(3) = 2$

Written like this, we can think of the the 1/3 as a probability and the numbers as our X. Then we have a probability distribution!

Still the mean $\bar{x}$ = 2.

X	1	2	3
P(X = x)	1/3	1/3	1/3

Now what if we said that 3 is more likely than the other numbers, so our new probability distribution is:

X	1	2	3
P(X = x)	1/6	1/6	2/3

Is the new mean going to increase, decrease or stay the same from previously??

- Or course it's going to increase!
- 3 is more likely, it's going to have more "impact / effect" on our mean calculation. Lets do that:

$$\bar{x} = \frac{1}{6}(1) + \frac{1}{6}(2) + \frac{2}{3}(3) = 2.5 > 2$$

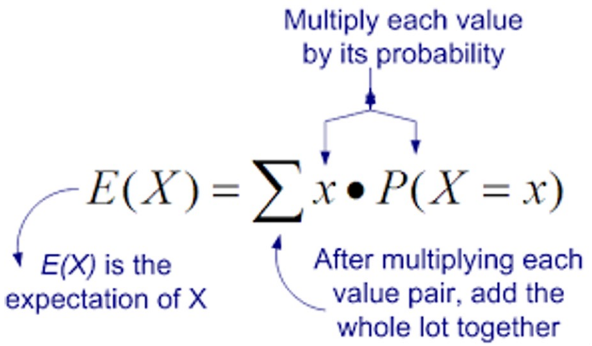
What we are actually calculating here is called the Expected Value, a new twist on an old concept!!

Expected Value and SD

- Just like how we calculated the mean and standard deviation when finding <u>summary (descriptive) statistics</u>, we can find the mean for probability distributions!

<u>Expected Value and SD</u>

- **Mean** for a probability distribution has a *different* <u>interpretation</u> than the mean we have seen before, which we thought of as the <u>center</u>.
- Now, the **mean** of a <u>probability distribution</u> is viewed in the *long run*, the value you would <u>expect</u> to see <u>on average</u>.
  - Think as a <u>weighted average</u> or average in the distribution sense, this is how we calculate it our calculator actually!
  - Can be a value you <u>don't necessarily observe</u>.
  - Ex. The average household size for the U.S. in 2017 is 2.6 people per household.
    - Won't see 2.6 people in any household, but would see the average of a city block be 2.6
  - Notation: Population mean = μ = E(X)



- **Standard Deviation** follows along the same lines.
  - It is the <u>average deviation</u> from the <u>expected value</u>.
  - Fancy formula, that we will never do by hand!

$$\sigma = \sqrt{\sigma^2} = \sqrt{\sum (x-\mu)^2 P(x)}$$

X	*P(X)*
Win = 3	0.3
Tie = 1	0.5
Loss = 0	0.2

E(X) = 3 x P(3) + 1 x P(1) + 0 x P(0)
= 3 (0.3) + 1 (0.5) + 0 (0.2)
= 1.4

https://www.texasgateway.org/resource/42-mean-or-expected-value-and-standard-deviation

Discrete Probability Distribution Example

In American Roulette, there are two zeroes and 36 non-zero numbers (18 red, 18 black and 2 green). If a player bets on red, his chance of winning is therefore 18/38 and his chance of losing is 20/38. Let's say it costs \$1 to play. If you win, you get \$2 and if you lose you get \$0.

What is the expected earnings?

Discrete Probability Distribution Example

In American Roulette, there are two zeroes and 36 non-zero numbers (18 red, 18 black and 2 green). If a player bets on red, his chance of winning is therefore 18/38 and his chance of losing is 20/38. Let's say it costs $1 to play. If you win, you get $2 and if you lose you get $0.

What is the expected earnings?

	Win	Lose
x		
P(X=x)		

Discrete Probability Distribution Example

In American Roulette, there are two zeroes and 36 non-zero numbers (18 red, 18 black and 2 green). If a player bets on red, his chance of winning is therefore 18/38 and his chance of losing is 20/38. Let's say it costs $1 to play. If you win, you get $2 and if you lose you get $0.

What is the expected earnings?

X = Amount won

$E(X) = \Sigma[x\ P(X=x)] = 2*(18/38) + 0*(20/38) = 36/38 = \$0.95$

So do you walk away with money? (Hint: Vegas)

	Win	Lose
x	2	0
P(X=x)	18/38	20/38

Discrete Probability Distribution Example

In American Roulette, there are two zeroes and 36 non-zero numbers (18 red, 18 black and 2 green). If a player bets on red, his chance of winning is therefore 18/38 and his chance of losing is 20/38. Let's say it costs \$1 to play. If you win, you get \$2 and if you lose you get \$0.

What is the expected earnings?

X = Amount won

$E(X) = \Sigma[x\ P(X=x)] = 2*(18/38) + 0*(20/38) = 36/38 = \$0.95$

So do you walk away with money? (Hint: Vegas)

No, remember it costs \$1 to play so \$0.95 - \$1 = -\$0.05.

In the long run, you walk away owing \$0.05

	Win	Lose
x	2	0
P(X=x)	18/38	20/38

Calculator Fun Sessss: Expected Value and SD

<u>GOAL</u>: Find the Expected Value!

X	*P(X)*
0	0.1
1	0.15
2	0.05
3	0.2
4	0.2
5	0.25
6	0.05

1. Check to see if we have a valid probability distribution!
2. Enter data.
   a) X values go in L_1.
   b) Probabilities (our "weights") go in L_2.
3. 1-Var Stats
   a) List is L_1 (X values).
   b) FreqList is L_2 (Probabilities).
   c) Calculate!

<u>Interpret results</u>:

- Let's say this data is about the number of touchdowns the Bengals had in games this season.
- **Find** the <u>expected value</u>. **Show** you work. **Interpret** in <u>context</u>.

Calculator Fun Sessss: Expected Value and SD

GOAL: Find the Expected Value and SD!

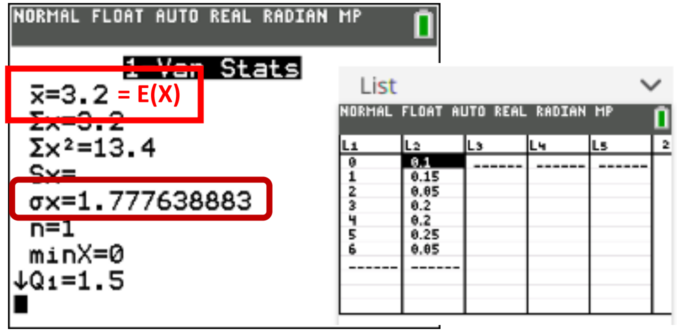
X	*P(X)*
0	0.1
1	0.15
2	0.05
3	0.2
4	0.2
5	0.25
6	0.05

$\Sigma P(X) = 1$

1. Check to see if we have a valid probability distribution!
   - *Yes!! Sum of all Probabilities = 1 and all individual P(X) are in between [0,1]*
2. Enter data.
   a) X values go in L_1.
   b) Probabilities (our "weights") go in L_2.
3. 1-Var Stats
   a) List is L_1 (X values).
   b) FreqList is L_2 (Probabilities).
   c) Calculate!

NOTE
- *This is the same way we calculated a weighted mean when we had a frequency table*
- *But now we have probabilities in L_2 instead of counts*

E(X) = μ = $\bar{x}$ in calc = 3.2
SD = $\boldsymbol{\sigma}$ = 1.777



Interpret results:
- Let's say this data is about the number of touchdowns the Bengals had in games this season.
- **Find** the expected value. **Show** you work. **Interpret** in context.

E(X) = 3.2, X = L1 and P(X) = L2), 1-VarStat(List = L1, FreqList = L2)

So we can expect the Bengals to score 3.2 touchdowns per game (in the long run).

Minor Note
- Sx (sample standard deviation) should be blank when we have probabilities in L_2
- This is because with probabilities, we are thinking of this as a population, so all individuals from the population have these probabilities
  - Probability distributions describe populations!

LCQ: Expected Value

A typical three-reel mechanical slot machine has different payoffs determined by the number and position of various pictures. Suppose the payoff (in dollars) is a discrete random variable with probability distribution given in the table below.

Center Pay line	3 7's	3 bars	3 plums	3 bells	3 oranges	3 cherries	2 cherries	1 cherry	
x	500	100	50	20	10	5	2	1	0
P(x)	$\frac{1}{8000}$	$\frac{1}{8000}$	$\frac{9}{8000}$	$\frac{48}{8000}$	$\frac{64}{8000}$	$\frac{30}{8000}$	$\frac{530}{8000}$	$\frac{3120}{8000}$	$\frac{4197}{8000}$

a) Find the expected payoff.
b) Interpret the mean. Then if it costs $1.00 to play, what happens in the long run?

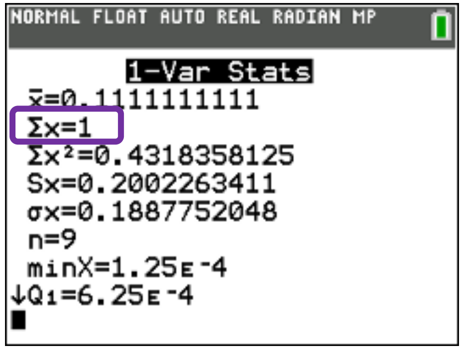
LCQ: Expected Value

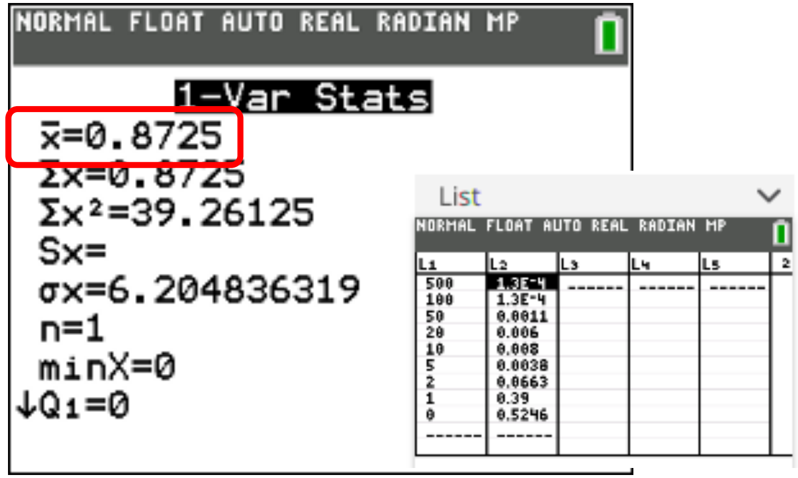
A typical three-reel mechanical slot machine has different payoffs determined by the number and position of various pictures. Suppose the payoff (in dollars) is a discrete random variable with probability distribution given in the table below.

Center Pay line	3 7's	3 bars	3 plums	3 bells	3 oranges	3 cherries	2 cherries	1 cherry	
x	500	100	50	20	10	5	2	1	0
P(x)	$\frac{1}{8000}$	$\frac{1}{8000}$	$\frac{9}{8000}$	$\frac{48}{8000}$	$\frac{64}{8000}$	$\frac{30}{8000}$	$\frac{530}{8000}$	$\frac{3120}{8000}$	$\frac{4197}{8000}$

a) Find the expected payoff.
b) Interpret the mean. Then if it costs $1.00 to play, what happens in the long run?

Check to make sure P(X) were entered correctly (1-Var Stat on P(X):





a) In calc $\bar{x}$ = E(X) = 0.8725, 1-Var Stat on X and P(X).

b) Over the long run, you will make 87 cents per game.

However, since it costs one dollar to play, you will actually end up losing 13 cents per game on average (–\$1 + \$0.87 = –\$0.13)

Summary LCQ!

Setup: Suppose 20 students are surveyed to determine how many siblings they have. The results are shown below:

X = # of siblings	*Frequency*
0	2
1	8
2	7
3	3

X = # of siblings	*P(X)*
0	
1	
2	
3	

1) Construct a probability distribution from the data above.

2) Calculate the Expected Value

3) Calculate the Standard Deviation

Summary LCQ!

Setup: Suppose 20 students are surveyed to determine how many siblings they have. The results are shown below:

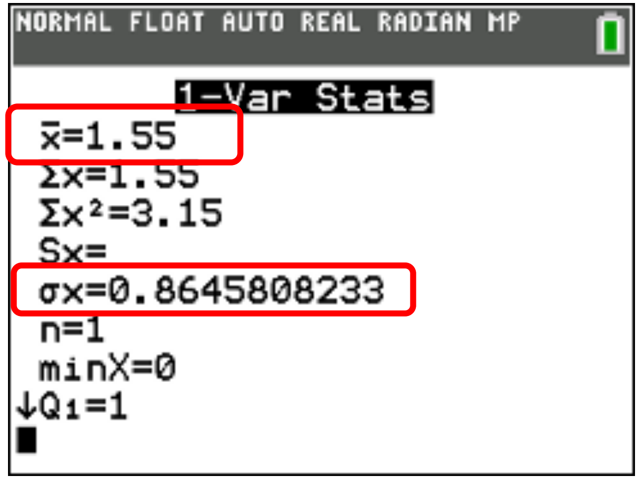
X = # of siblings	*Frequency*
0	2
1	8
2	7
3	3

X = # of siblings	*P(X)*
0	2/20 = 0.1
1	8/20 = 0.4
2	0.35
3	0.15

1) Construct a probability distribution from the data above.

2) Calculate the Expected Value **= 1.55**

3) Calculate the Standard Deviation **= 0.865**



PROBLEM SESSION!!!!!!!!!!!

RV Example

The maintenance staff of a large office building regularly replaces fluorescent ceiling lights that have gone out. During a visit to a typical floor, the staff may have to replace several lights. The manager of this staff has given the following probabilities to the number of lights (identified by the random variable X) that need to be replaced on the floor.

X	0	1	2	3	4
P(X = x)	0.2	0.15	0.2	0.3	0.15

a) Find $P(X > 1)$
b) Find $P(X \leq 3)$

RV Example Solution

The maintenance staff of a large office building regularly replaces fluorescent ceiling lights that have gone out. During a visit to a typical floor, the staff may have to replace several lights. The manager of this staff has given the following probabilities to the number of lights (identified by the random variable X) that need to be replaced on the floor.

X	0	1	2	3	4
P(X = x)	0.2	0.15	0.2	0.3	0.15

a) $P(X > 1) = P(X = 2) + P(X = 3) + P(X = 4) = 0.2 + 0.3 + 0.15 =$ **0.65**

b) $P\,(X \leq 3) = 1 - P(X = 4) = 1 - 0.15 =$ **0.85**

RV Example, continued

The maintenance staff of a large office building regularly replaces fluorescent ceiling lights that have gone out. During a visit to a typical floor, the staff may have to replace several lights. The manager of this staff has given the following probabilities to the number of lights (identified by the random variable X) that need to be replaced on the floor.

X	0	1	2	3	4
P(X = x)	0.2	0.15	0.2	0.3	0.15

c) How many lights should the manager expect to replace on the floor?

d) What is the standard deviation of the number of lights on a floor that are replaced?

This shows it by hand, which we can do!
But we can also use the calc!!

RV Example, continued, Solution

X	0	1	2	3	4
P(X = x)	0.2	0.15	0.2	0.3	0.15

- Let X = the number of light that need to be replaced
- Find the mean number of lights that need to be replaced on a floor.

- $E(x) = \mu = \sum x \cdot P(x)$
- $E(x) = \mu = 0(0.2) + 1(0.15) + 2(0.2) + 3(0.3) + 4(0.15) = \mathbf{2.05}$

- On average, we would expect to replace 2.05 lights on a floor.

This shows it by hand, which we WON'T do! USE the calc!!

RV Example, continued, Solution, p. 2

X	0	1	2	3	4
P(X = x)	0.2	0.15	0.2	0.3	0.15

- Let X = the number of light that need to be replaced
- The mean number of lights that need to be replaced on a floor is 2.05.
- $\sigma = \sqrt{\sum[X - E(x)]^2 P(x)}$
- $\sigma = \sqrt{(0-2.05)^2 \cdot 0.2 + (1-2.05)^2 \cdot 0.15 + (2-2.05)^2 \cdot 0.2 + (3-2.05)^2 \cdot 0.3 + (4-2.05)^2 \cdot 0.15}$

  $\sigma = \sqrt{1.8475} = \mathbf{1.3592}$

- For every floor we would expect to replace between 2.05 ± 1.36, or between 0.69 and 3.41 lights.

Textbook Problems #3 & 5

Suppose the probabilities of a customer purchasing 0, 1, or 2 books at a book store are 0.5, 0.3, and 0.2, respectively.

a) Create the probability distribution
b) What is the expected number of books a customer will purchase?
c) Find the standard deviation of the number of books a customer will purchase.

This shows it by hand, but USE the calc!! Especially for SD

Textbook Problems #3 & 5 Solution

a)

X	0	1	2
P(X = x)	0.5	0.3	0.2

b) E(X) = 0(0.5) + 1(0.3) + 2(0.2) = 0.7 books

c) $\sigma^2 = \sum[X - E(x)]^2 P(x)$

$= [(0 - 0.7)^2(0.5)] + [(1 - 0.7)^2(0.3)] + [(2 - 0.7)^2(0.2)] = 0.61$

SD = $\sqrt{0.61} = 0.7810$ books