

Using a spreadsheet to discover formulas for the mean and standard deviation of the binomial distribution

By: Mike Iannone

Activity Overview

In Problem #1 students will learn to use a number of features of a spreadsheet to compute the mean and standard deviation of a general discrete probability distribution. In Problem #2 students will use the techniques of Problem #1 to solve the problem given. In problem #3 students will discover an alternative method for computing the mean and standard deviation of a binomial distribution. In Problem #4 students will continue to vary the parameters of the binomial distribution and view a normal curve which overlays the scatter plot for the distribution.

Concepts included in Activity

- Discrete probability distribution
- Mean and standard deviation of a probability distribution
- Binomial distribution probabilities
- Normal distribution

Teacher Preparation

- All instructions for students to complete this activity appear on the student worksheet and on this form
- The activity is meant to be teacher led so that the teacher can make certain that students have the necessary prerequisite knowledge
- The expected screen shots for the completed Problem #2 are given on page 4 of this form
- Specific recommendations for the teacher to implement during the activity are included throughout this document

The Classroom.

- Students will need to download the activity to the handheld and have a printed copy of the worksheet
- Many commands utilized can be entered by using the alphabet keys on the handheld or by using built-in commands listed in the library of commands – students need to be made aware of this

The Activity

Problem #1

The following discrete probability distribution is given on the student worksheet.

x	0	1	2	3	4	5
p(x)	.04	.16	.26	.38	.12	.04

On page 1.2 have students carry out steps a, c and d

- Use the NavPad to scroll about the spreadsheet to see that each of these values are entered in columns C and D (be certain students view row 5 which doesn't show on the adjoining screen shot)
- Review the formula for computing the mean μ of a discrete probability distribution

i.e.
$$\mu = \sum x_i \cdot p(x_i)$$

- On the command line for column E (in the gray area above cell E1) enter the expression $= c * d$
- In cell B1 sum the resulting column E by entering $= sum(e1 : e6)$. At this point the teacher may wish to point out that the *sum* command can be found in the catalog of commands accessed by pressing the key. The teacher should also note that the calculated number that appears is the mean of the distribution. Students should place the value of the mean on their worksheet. (note: $\mu = 2.5$)

Have students advance to page 1.3 which should be identical to recalculated page 1.2. Then have students carry out steps b-e.

- Review the formula for computing the variance σ^2 and standard deviation σ of a discrete probability distribution:

i.e.
$$\sigma^2 = \sum (x_i - \mu)^2 \cdot p(x_i) \text{ and}$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{\sum (x_i - \mu)^2 \cdot p(x_i)}$$

- On the command line for column F enter $= (c - b1)^2 * d$
- In cell B3 enter the command: $= sum(f1 : f6)$

- The calculated number that appears is the variance of the distribution (note: $\sigma^2 = 1.29$)

- e. In cell B4 students enter the command: $=\text{sqrt}(b3)$. The number that appears is the standard deviation of the distribution. Students should place the value of the standard deviation on their worksheet. (note: $\sigma=1.13578$)
- f. In cells A1, A3, and A4 enter the words mean, variance and st dev respectively by enclosing these words in double quotes. This will help to remember what cells B1, B3 and B4 denote.
- g. In the title lines (above the command line) for columns C and D enter column names *num_val* and *prob_val* respectively which will be used to create a scatter plot of the distribution (note the underscore is found in the Symbols Palette accessed by pressing the $\text{ctrl} + \text{[}$ keys).

Have students advance to page 1.4 which should be identical to recalculated page 1.3

	A	B	C num_val	D prob_val
1	"mean"	2.5	0	0.04
2			1	0.16
3	"variance"	1.29	2	0.26
4	"st dev"	1.13578	3	0.38
5			4	0.12

Columns A-D

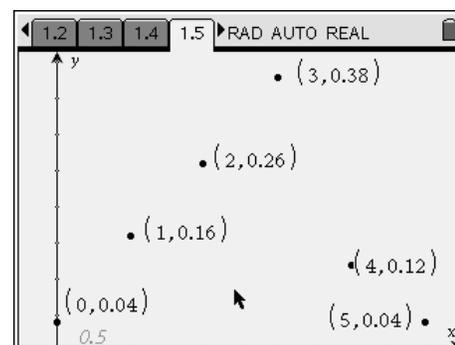
	C num_val	D prob_val	E	F
			=c[*]d[*	=(c[*]-b4)^
1	0	0.04	0.	0.25
2	1	0.16	0.16	0.36
3	2	0.26	0.52	0.065
4	3	0.38	1.14	0.095
5	4	0.12	0.48	0.27

Columns C-F

This page is provided simply for students to make certain their recalculated page 1.3 is correct. Again students will need to scroll about the page to view the entire spreadsheet.

Have students advance to page 1.5

Explain to the students that this is a scatter plot of the given probability distribution and that on geometry page 1.6 they will replicate page 1.5



Have students advance to page 1.6 which is a blank graph which students will use to generate the graph on page 1.5 by following the steps below:

- a. From the Graph Type menu select Scatter Plot. (menu,3:Graph Type,4:Scatter Plot)
- b. In the Entry Line at the bottom that appears select *num_val* in the box for *x* and *prob_val* in the box for *y* (click on boxes to view variables)
- c. In the Window menu select Zoom-Data (menu,4:Window,9:Zoom-Data)
- d. In the View menu, select the Hide Entry Line (menu,2:View,6:Hide Entry Line)

- e. In the Points & Lines menu, select Point On (menu,6:Points&Lines,2:Point On) and click on each point to enter coordinates (place pointer over point and hit the Enter button)

Finally for this problem, have students draw the scatter plot showing on the handheld in the space provided on the student worksheet. If done correctly, page 1.6 on the student handheld should be a replica of Page 1.5.

Problem #2

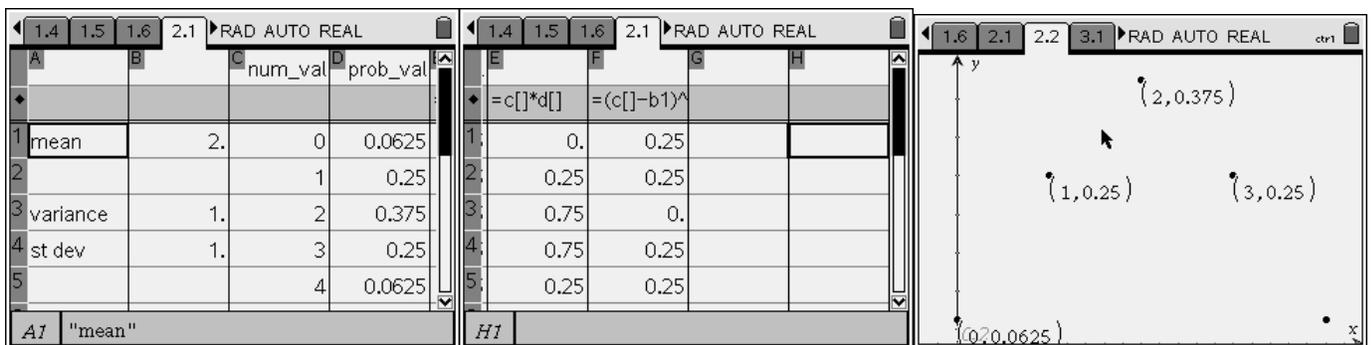
The following discrete probability distribution is given on the student worksheet.

A number is selected randomly from the set {1,2,3,4,5,6,7,8}. Then a second number from the same set, then a third, and a fourth, all from the same given set. Let x denote the number of even numbers drawn in the 4 selections. The discrete probability distribution for x is given below:

x	0	1	2	3	4
$p(x)$.0625	.25	.375	.25	.0625

Using the process for Problem #1 have students use the blank Spreadsheet page 2.1 and Graphs & Geometry page 2 to compute the mean and the standard deviation of the distribution and construct the graph of the scatter plot. Then enter each in the space provided on the student worksheet.

The completed student pages should appear as follows:



Problem #3

Six numbers are selected randomly from the set {1,2,3,4,5,6,7,8,9,10} with replacement (i.e. first a number is selected, then a second number then a third, a fourth, a fifth and a sixth - each time a number is selected it is replaced back into the original set before the next number is selected). Let x denote the number of primes drawn in the 6 selections (note: 1 is considered to be a unit number and not a prime). The students will determine the entire probability distribution of x by completing the table below (correct to 3 decimal places):

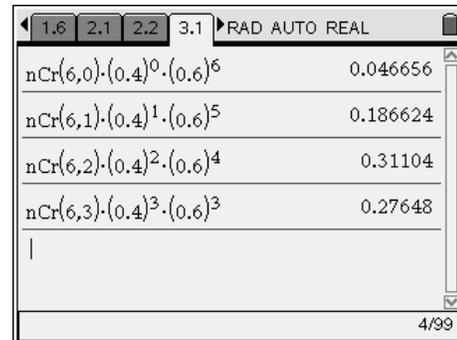
x	0	1	2	3	4	5	6
$p(x)$.047	.187	.311	.276			

They then need to determine the mean, variance, standard deviation and scatter plot for the distribution.

At this point the teacher should review that the distribution is binomial and that the probability for each x is computed using the formula $p(x) = nCr(6, x) * p^x * (1 - p)^{n-x}$ where n = number of trials and p = probability of success on each trial. Point out that in the problem above $n = 6$ and $p = 0.4$.

Use calculator page 3.1 for students to

- a. Compute the probabilities for $x = 0, 1, 2, 3, 4, 5, 6$ using the formula $p(x) = nCr(6, x) \cdot p^x (1 - p)^{n-x}$ for the probability of x . Again note that n denotes the number of trials and p denotes the probability of success. The first 4 values of x are computed for you.



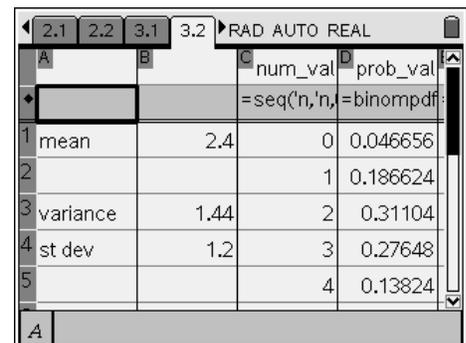
- b. Complete the table above. Students might want to check that the sum of the probabilities is 1. The built-in function $nCr()$ is found in the probability submenu (menu, 5:Probability, 3:Combinations) but can also be typed directly. The complete distribution will be as follows:

x	0	1	2	3	4	5	6
$p(x)$.047	.187	.311	.276	.138	.037	.004

- c. Compute the command $seq(n,n,0,6,1)$ (again this command can be typed directly or accessed from menu, 6:Statistics,4:List Operations,5:Sequence). Ask students what they observe (students will observe that this command provides the values of x in the table).
- d. Compute $binomPdf(6,0.4)$, this built-in function can be found in catalog of commands or typed directly. Have students compare the values to the completed table and ask students what they observe. (students will observe that these provide the values of $p(x)$ in the table).

Thus these last two commands give us a quick way to produce the probability distribution for this problem and are accurate to more decimal places.

On spreadsheet page 3.2 observe that these last two commands appear on the command line for columns C and D and that the mean and variance are computed via columns E and F as before. Students should enter the mean and standard deviation in the space provided on the student worksheet

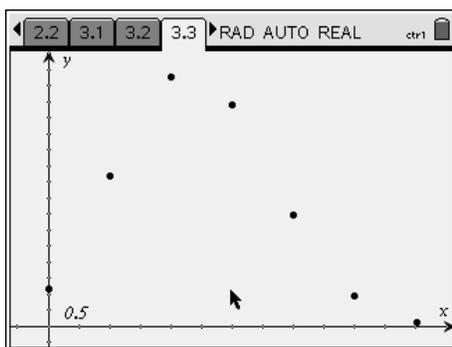


Students should then go back to Page 3.1 and define $n = 6$ (use command $n := 6$) and $p := 0.4$ (use command $p := 0.4$) on separate lines and then calculate the commands $n * p$ and $\sqrt{n * p * (1 - p)}$ separately and enter the results in the space provided on the student worksheet.

Ask students what they observe. Students should observe the first formula yields the mean and the second formula the standard deviation

Students should next advance to **Page 3.3** which is a scatter plot of the distribution and then advance to **Page 3.4** where the values for n and p are entered in cells B1 and B2 respectively and that the mean and standard deviation are now computed by their respective formulas $n * p$ and $\sqrt{n * p * (1 - p)}$ in cells B4 and B5 and again in B6 and B7 by definition via columns E and F. The current values of n and p are set respectively as 6 and 0.4

Students should complete the table that appears on the student worksheet. After observing table, students should write their observations and any conclusions reached. Students need to recalculate entire worksheet using CTRL-R each time an entry is changed.



	A	B	C	D
			=seq('n',n,=binompdf)	
1	n =	6	0	0.046656
2	p =	0.4	1	0.186624
3			2	0.31104
4	n*p =	2.4	3	0.27648
5	sqrt(np(1...	1.2	4	0.13824
A7	"n = "			

Note to teacher: The student answers to problems 3 and 4 appear at the end of this document

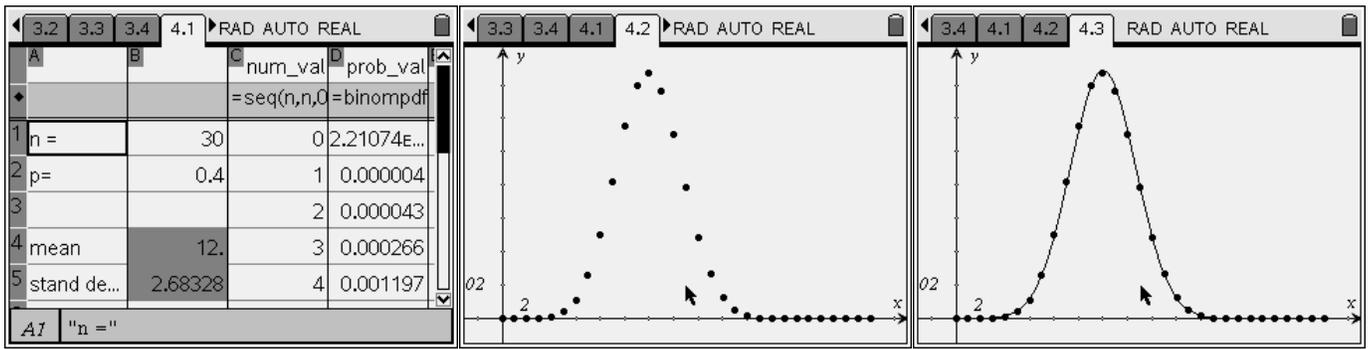
Problem #4

The spreadsheet on Page 4.1 simplifies the process of finding the mean and standard deviation of any binomial distribution for any value of n less than or equal to 100 and for any value of p . The current Page has an $n = 30$ and $p = .4$. The value for n has been entered in cell A1 and the value for p in cell A2.

Page 4.2 provides a scatter plot for the current values of n and p and **Page 4.3** extends our understanding of the binomial distribution as it overlays the normal curve with the mean and standard deviation for the binomial distribution defined by the current values of n and p stored in cells B1 and B2 respectively.

Students should complete the table that appears on the student worksheet for this problem. Each time they change the values of n and p they should view pages 4.2 and 4.3 as well. Point out that that every time the values of n and p change, the scatter plot needs to be resized by re-selecting Zoom-Data in the Window submenu.

Once again, have students write their observation about the comparison between the normal curve and the binomial distribution scatter plot in the space provided on the student worksheet.



Answers for student worksheet

Problem 3:

A. Mean 2.4 Standard Deviation 1.2

B. $n * p$ 2.4 $\sqrt{n * p * (1 - p)}$ 1.2

Compare the numbers you entered in lines A and B. What do you observe? Write your response below:

The values in A and B are equal

n (cell B1)	p (cell B2)	Mean (by formula) (cell B4)	Mean (by definition) (cell B7)	Stand Dev (by formula) (cell B5)	Stand Dev (by definition) (cell B8)
20	0.6	12	12	2.19	2.19
10	0.45	4.5	4.5	1.57	1.57
15	0.5	7.5	7.5	1.94	1.94
24	.75	18	18	2.12	2.12

After observing table, write your observation and any conclusions you reach below:

For binomial distribution while the mean and standard deviation may be completed by definition, the computational formulas $mean = n * p$ and $standard\ deviation = \sqrt{n * p * (1-p)}$ provide a more expeditious method of obtaining these values.

Problem 4:

n	p	mean	standard deviation
30	0.4	12	2.68
20	0.6	12	2.19
15	0.45	6.75	1.93
24	0.75	18	2.12