

Activity 6

Demonstrating Probability, Simulations, and Probability Distributions

This activity demonstrates the probability of an event happening with the simple example of simulating the tossing of a coin.

It is important that you understand the concept of “random seed” described in Topic 21 to fully understand all other topics that contain simulations.

Other topics explain how probabilities can be calculated or give a simulation that shows the reasonableness of the results found. Topics 29 and 30 may be of interest to those who seek ideas on using simulations to solve probability problems.

Topic 21—Coin Toss (The Law of Large Numbers)

If you toss four coins, you might get all tails, but in the long run, you would expect about half heads (.50) and half tails.

Setting a Random Seed

To simulate tossing of coins, you will set a random seed so that the results can be duplicated. If you have no need to duplicate these results, then skip this step. If you want to assure that your results are different from others, you can use the last four digits of your Social Security Number as your seed.

1. Type `321` `[STO]` `[MATH]` `<PRB>` `1:rand` `[ENTER]`, as shown in the first two lines in screen 1.
2. Put the numbers 1 to 150 in `L1` using `[2nd]` `[LIST]` `<OPS>` `5:seq(X, X, 1, 150)` `[STO]` `L1` `[ENTER]`.
3. Generate one (1) random coin toss at a time with a 0.5 chance of getting a head = $H = 1$. Do this 150 times.

Store the results in `L2` with `[MATH]` `<PRB>` `7:randBin(1, .5)` `[STO]` `L2` `[ENTER]`.

The results `{1 0 1 0 1 1 0 ...}` shown in screen 2 can be read as `{H T H T H H T...}`

```
321→rand
seq(X,X,1,150)→L1
{1 2 3 4 5 6 7 ...
```

(1)

```
randBin(1,.5,150)
→L2
{1 0 1 0 1 1 0 ...
cumSum(L2)→L3
{1 1 2 2 3 4 4 ...
L3/L1→L4
{1 .5 .6666666666...
```

(2)

Activity 6, Demonstrating Probability, Simulations, and Probability Distributions (cont.)

4. Press $\boxed{2\text{nd}}$ $\boxed{[LIST]}$ $\langle\text{OPS}\rangle$ $\boxed{6:\text{cumSum}}$ (L_2 $\boxed{[STO\rightarrow]}$ L_3) to store the cumulative sum of L_2 in L_3 . The results $\{1\ 1\ 2\ 2\ 3\ 4\ 4\ \dots\}$ indicate that the first head occurred on the first toss (1), the second head on the third toss ($1 + 0 + 1 = 2$), the third head at the fifth toss ($1 + 0 + 1 + 0 + 1 = 3$), and so on (see screen 2).

5. Press L_3 $\boxed{[]}$ L_1 $\boxed{[STO\rightarrow]}$ L_4 to get the proportion of heads. $\{1\ .5\ .6666\dots\}$ indicates 100 percent ($1/1 = 1$) heads on the first toss; 50 percent ($1/2$) heads after the second toss, because there is still just one head; and 66.67 percent heads, or two out of three heads, after the third toss. (See the last line of screen 2.)

In the short run, this is a lot of fluctuation.

To see what happens in the long run, set up and plot the results as follows.

1. Press $\boxed{2\text{nd}}$ $\boxed{[STAT PLOT]}$ $\boxed{1:\text{Plot1}}$ to set up for an **xyLine** plot, as shown in Topic 9 (see screen 3).
2. Press $\boxed{Y=}$ and set $Y_1 = 0.5$ to plot this horizontal line (see screen 4).
3. Set the **WINDOW** as shown in screen 5, and then press $\boxed{[TRACE]}$ for the plot of the first ten tosses ($X_{\text{max}} = 10$) as shown in screen 6.
4. Using $\boxed{[WINDOW]}$, change X_{max} to 100, leaving the other values as shown in screen 5.

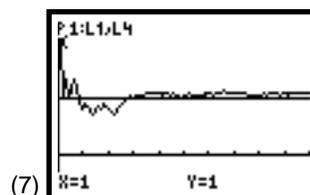
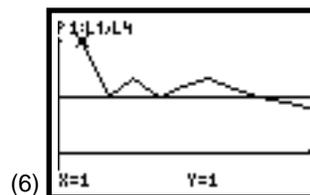
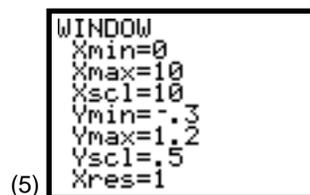
5. Press $\boxed{[TRACE]}$ for the plot in screen 7, which covers 100 tosses.

The first tic mark on the x-axis marks off the first 10 tosses, and the plot to that point is the same as shown in screen 6 but condensed by the change of scale. Notice how the plotted points hover about the $Y_1 = 0.5$ line.

6. Press $\boxed{[STAT]}$ $\boxed{1:\text{Edit}}$, and use $\boxed{[]}$ to move to the tenth row of proportions in L_4 with **0.4** (see screen 8).

L_3 has the number of heads, or four out of ten. You can investigate this from the home screen with L_3 $\boxed{[]}$ $\boxed{10}$ $\boxed{[]}$, as shown in screen 9. Screen 9 also shows that after 100 tosses, there are 55 heads and 45 tails (so the proportion of heads is $55/100 = 0.55$), and the 150th toss gives 76 heads and 74 tails ($76/150 = 0.506667$).

In the long run, the proportion of heads is approximately 0.5.



(8)

L_2	L_3	L_4	4
0	2	.5	
1	3	.6	
1	4	.66667	
0	4	.57143	
0	4	.5	
0	4	.44444	
0	4	.4	
$\text{L}_4(10) = .4$			

(9)

$\text{L}_3(10)$	
$\text{L}_3(100)$	4
$\text{L}_3(150)$	55
	76

Topic 22—Factorials, Permutations, and Combinations

The following functions have many uses, including calculating probabilities, as you will see in later topics.

Using Factorials

- A. How many ways can you pick a president and vice-president from two people, Ann and James?

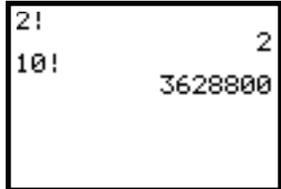
There are two choices for president, Ann and James. For each of these choices, there is only one choice left for vice-president. Therefore, $2 * 1 = 2! = 2$ ways:

Ann James and James Ann.

Press **2** **[MATH]** **<PRB>****4:!** **[ENTER]** for **2**, as shown in screen 10.

- B. How many ways can you set up a slate of 10 people for 10 positions?

$10 * 9 * 8 * 7 * 6 * 5 * 4 * 3 * 2 * 1 = 3,628,800$ possible slates.

(10) 

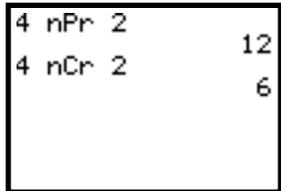
Using Permutations

How many ways can you pick a slate of two (president and vice-president) from four (Ann, James, Carl, and Doris)?

There are four choices for president. For each of these four choices, there are three choices for vice-president for a total of $4 * 3 = (4 * 3 * 2 * 1) / (2 * 1) = 4!/2! = 12$ possible slates:

Ann James	James Ann	Carl Ann	Doris Ann
Ann Carl	James Carl	Carl James	Doris James
Ann Doris	James Doris	Carl Doris	Doris Carl

Press **4** **[MATH]** **<PRB>** **2:nPr** **2** **[ENTER]** for **12**, as in screen 11.

(11) 

Using Combinations

How many ways can you make a committee of two from the four people above?

Because order is not important in setting up a committee (Ann James is a different slate than James Ann, but they make up the same committee); in fact, there are $2!$ too many values above. Therefore, $4!/(2!2!) = 6$ ways: Ann James, Ann Carl, Ann Doris, James Carl, James Doris, and Carl Doris.

Press **4** **[MATH]** **<PRB>****3:nCr** **2** **[ENTER]** for **6**, as shown in screen 11.

Activity 6, Demonstrating Probability, Simulations, and Probability Distributions (cont.)

Topic 23—Binomial Distribution

The production of an electronic component has a large 20 percent defective rate. If a random selection of six components is taken,

- What is the probability of getting exactly two defectives?
- What is the probability of getting at most two defectives?
- What is the probability of getting at least two defectives?
- What is the probability of getting from two to four defectives?

Generate the complete distribution by pressing $\boxed{2\text{nd}} \boxed{[DISTR]}$ **0:binompdf**(6 $\boxed{,}$ $\boxed{.}$ 2 \boxed{STO} $\boxed{L1}$ \boxed{ENTER} for the first two lines of screen 12, and then pressing \boxed{STAT} **1:Edit** to reveal the probabilities in L_1 . The values of 0 to 6 have been keyed into L_2 , as shown in screen 13.

- $P(1) = .39322$ is in the second row of the spreadsheet under L_1 (see screen 13), but you can calculate it directly using $\boxed{2\text{nd}} \boxed{[DISTR]}$ **0:binompdf**(6 $\boxed{,}$ $\boxed{.}$ 2 $\boxed{,}$ 1 \boxed{ENTER} for the results shown in the last two lines of screen 12.
- $P(0) + P(1) + P(2) = .90112$ is the sum of the first three values of L_1 , but you can also calculate it directly using $\boxed{2\text{nd}} \boxed{[DISTR]}$ **A:binomcdf**(6 $\boxed{,}$ $\boxed{.}$ 2 $\boxed{,}$ 2 \boxed{ENTER} for the results shown in the first two lines of screen 14. (The remainder of the screen shows an alternate method.)
- $P(2) + P(3) + P(4) + P(5) + P(6) = 1 - (P(0) + P(1)) = .34464$, or $1 - \boxed{2\text{nd}} \boxed{[DISTR]}$ **A:binomcdf**(6 $\boxed{,}$ $\boxed{.}$ 2 $\boxed{,}$ 1 \boxed{ENTER} for the results shown in screen 15.
- $P(2) + P(3) + P(4) = 0.34304 = (P(0) + (P(1) + P(2) + P(3) + P(4)) - (P(0) + P(1)))$, or $\boxed{2\text{nd}} \boxed{[DISTR]}$ **A:binomcdf**(6 $\boxed{,}$ $\boxed{.}$ 2 $\boxed{,}$ 4 $\boxed{)} - \boxed{2\text{nd}} \boxed{[DISTR]}$ **A:binomcdf**(6 $\boxed{,}$ $\boxed{.}$ 2 $\boxed{,}$ 1 $\boxed{)} \boxed{ENTER}$ as shown in screen 16. (Note that **sum(binompdf(6 $\boxed{,}$ $\boxed{.}$ 2 $\boxed{,}$ { 2 $\boxed{,}$ 3 $\boxed{,}$ 4 } $\boxed{)} \boxed{ENTER}$ also gives the same results.)**

```
binompdf(6,.2,L1
(.262144 .39321...
binompdf(6,.2,1
.393216
```

(12)

Note: pdf stands for probability density function.

L1	L2	L3	2
.26214	0		-----
.39322	1		
.24578	2		
.08192	3		
.01536	4		
.00154	5		
6.4E-5	6		

L2()=0

(13)

```
binomcdf(6,.2,2
.90112
binompdf(6,.2,(0
,1,2)
(.262144 .39321...
sum(Ans
.90112
```

(14)

Note: cdf stands for cumulative density function.

```
1-binomcdf(6,.2,
1)
.34464
```

(15)

```
binomcdf(6,.2,4)
-binomcdf(6,.2,1
)
.34304
sum(binompdf(6,.
2,{2,3,4})
.34304
```

(16)

Constructing a Probability Histogram

1. Set up **Plot1** and the **WINDOW** as shown in screens 17 and 18 and as discussed in Topic 2 on **Histograms**.
2. Press **TRACE** for the probability **Histogram** in screen 19 with $P(0) = n = 0.262144$.

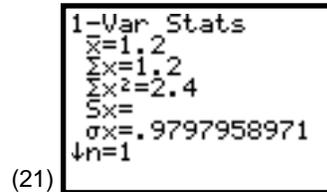
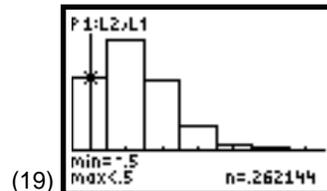
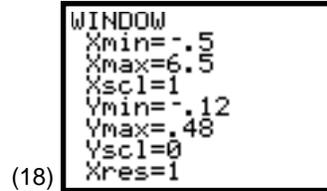
Displaying Mean and Standard Deviation

With X in L_2 and probabilities in L_1 , pressing **STAT** **<CALC>** **1:1-Var Stats** L_2 \square L_1 displays screen 20, and then pressing **ENTER** gives screen 21.

Note that $\mu = \bar{x} = 1.2 = n * p = 6 * 0.2 = 1.2$ and $\sigma = \sigma_x = 0.9797958971 = \sqrt{(n * p(1 - p))} = \sqrt{(6 * 0.2 * 0.8)}$.

Large Sample Domain Errors

If the sample size is too large (≥ 1000), a domain error will occur, as demonstrated in screens 22 and 23 ($n = 999$ works fine, but $n = 1000$ gives an error). You can approximate many large sample binomial distribution problems by using the *normal distribution* as explained at the end of Topic 24.



Activity 6, Demonstrating Probability, Simulations, and Probability Distributions (cont.)

Topic 24—Normal Distribution

Assume you are given a normal population of heights with $\mu = 68$ inches and $\sigma = 2.5$ inches. You can plot the normal probability density function (screen 25) with a **WINDOW**, as shown in screens 26, by pressing $\overline{Y=}$, and then with the cursor after $Y1=$, pressing $\overline{2nd}$ [DISTR] **1:normalpdf(X** (see screen 24) and then \overline{GRAPH} .

A. What proportion of the population is between 65.5 and 70.5 inches tall or between the z values of -1 and 1 ? This also could be stated as, what is the probability that a person picked at random from the population has a height between 65.5 and 70.5 inches?

1. Press $\overline{2nd}$ [DISTR] **2:normalcdf(-1** $\overline{,}$ **1** \overline{ENTER} for **.6827**, as shown in the first two lines in screen 27.
2. Enter the heights directly or shade an area for the answer.

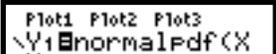
To enter the heights directly, add μ and σ , as shown in screen 27, using $\overline{2nd}$ [DISTR] **2:normalcdf(65** $\overline{,}$ **5** $\overline{,}$ **70** $\overline{,}$ **5** $\overline{,}$ **68** $\overline{,}$ **2** $\overline{,}$ **5** \overline{ENTER} for **0.6827**.

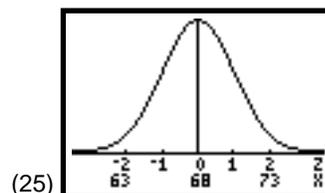
To shade an area for the answer, with the **WINDOW** as given in screen 26 and with all plots off, press $\overline{2nd}$ [DISTR] **<DRAW> 1:ShadeNorm(-1** $\overline{,}$ **1** \overline{ENTER} for the shaded plot, as shown in screen 29, where **Area = .6827** is given below the plot.

B. What proportion of the population is greater than 73 inches or a z value greater than 2?

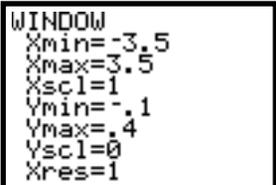
1. Press $\overline{2nd}$ [DISTR] **2:normalcdf(2** $\overline{,}$ **E99** \overline{ENTER} (obtain the **E** by pressing $\overline{2nd}$ [EE] above $\overline{,}$) for **.02275** with **E99 = 10^{99}** , a very large number or “infinity.”

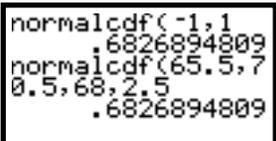
Note that the original height (73 inches) could have been used, as shown in screen 30, if we had used **normalcdf(73** $\overline{,}$ **E99** $\overline{,}$ **68** $\overline{,}$ **2** $\overline{,}$ **5** \overline{ENTER} for **0.02275**.

(24) 

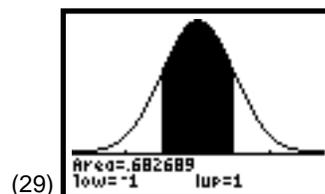


Note: The z values $-2, -1, 0, 1, 2$, and some of the corresponding heights of 63, 68, and 73 inches have been added to the plot.

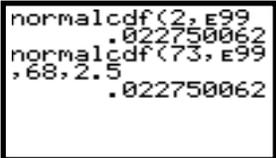
(26) 

(27) 

(28) 



Note: It may be necessary to clear a previous plot as in step 2.

(30) 

2. To shade the desired area after working problem A,
- Clear the previous drawing or shading using $\boxed{2\text{nd}} \boxed{\text{DRAW}} \boxed{1:\text{ClrDraw}} \boxed{\text{ENTER}}$ for Done on the home screen, as shown in screen 31.
 - Press $\boxed{2\text{nd}} \boxed{\text{DISTR}} \boxed{\text{<DRAW>}} \boxed{1:\text{ShadeNorm}} \boxed{(} \boxed{2} \boxed{,}$
 $\boxed{\text{E99}} \boxed{)} \boxed{\text{ENTER}}$ for the shaded graph in screen 32.
- C. What proportion of the population is less than 73 inches tall?

From problem B, this would be $1 - 0.02275 = 0.97725$, but you can also do this as shown in screens 33, 34, and 35 with negative infinity given as $-\text{E99}$.

Finding a Value From a Given Proportion

- D. What height separated the tallest 5 percent of the population from the other 95 percent?

Press $\boxed{2\text{nd}} \boxed{\text{DISTR}} \boxed{3:\text{invNorm}} \boxed{(} \boxed{0} \boxed{.} \boxed{95} \boxed{\text{ENTER}}$ for a z value of **1.645**. From this value, you can calculate the height as $\mu + z\sigma$, or $68 + 1.645 * 2.5 = 72.112$. (See screen 36.)

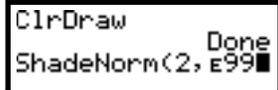
The height itself is returned if you add μ and σ using $\boxed{2\text{nd}} \boxed{\text{DISTR}} \boxed{3:\text{invNorm}} \boxed{(} \boxed{0} \boxed{.} \boxed{95} \boxed{,}$
 $\boxed{68} \boxed{,}$
 $\boxed{2} \boxed{,}$
 $\boxed{5} \boxed{\text{ENTER}}$ for **72.112** inches. (See the last lines of screen 36.)

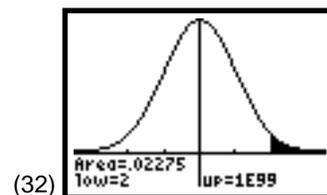
Using Normal Approximation to the Binomial

At the end of Topic 23, you saw that $n = 999$ was the largest sample size that the binomial functions could handle. If n is larger than 999, the following method (which is justified in Topic 25) can be used if $n * p \geq 10$ (some texts say 5) and $n * (1 - p) \geq 10$. If $n = 999$ and $p = 0.1$, you have $999 * 0.1 = 99.9$ and $999 * 0.9 = 899.1$, which are both ≥ 10 .

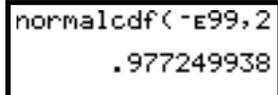
We will use $n = 999$ and $p = 0.1$ to show how well the method works.

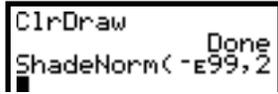
Use a normal distribution with $\mu = n * p = 99.9$ and $\sigma = \sqrt{(n * p * (1 - p))} = \sqrt{(999 * 0.1 * 0.9)} = 9.482$. (See screen 37.)

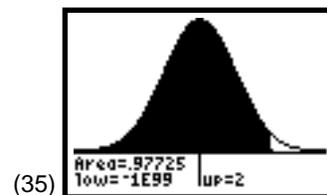
(31) 

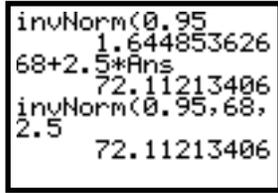


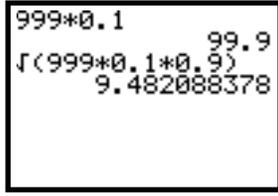
Note: $\text{ShadeNorm}(73, \text{E99}, 68, 2.5)$ also works but the window must be adjusted accordingly for each change in μ or σ while the window in screen 26 works for all standard normal distributions or z values.

(33) 

(34) 



(36) 

(37) 

Activity 6, Demonstrating Probability, Simulations, and Probability Distributions (cont.)

A. What is the probability of exactly 100 successes?

The discrete binomial **Histogram** has a rectangle of base 1 and the height of the probability. For $X = 100$, the base goes from 99.5 to 100.5 with area under the normal curve of **0.0421** compared to the exact binomial answer of **0.0420**. (See screen 38.)

```
(38) normalcdf(99.5,100.5,99.9,9.482)
      .0420519511
      binompdf(999,.1,100)
      .0420167909
```

B. What is the probability of less than 100 successes?

Normal approx. = **0.483**. Exact binomial = **0.489**. (See screen 39.)

```
(39) normalcdf(-E99,99.5,99.9,9.482)
      .4831754689
      binomcdf(999,.1,99)
      .4887840163
```

C. What is the probability of from 100 to 200 successes?

Normal approx. = **0.517**. Exact binomial = **0.511**. (See screen 40.)

```
(40) normalcdf(99.5,200.5,99.9,9.482)
      .5168245311
      binomcdf(999,.1,200)-binomcdf(999,.1,99)
      .5112159837
```

Topic 25—Distribution of Sample Proportions (Simulation)

The distribution of sample proportions will be simulated. For small sample sizes, a skewed distribution results because the population proportion used is not 0.5. As the sample size increases, the distribution becomes more normal, reinforcing the normal approximating given at the end of Topic 23.

The following simulation is of a sampling from a very large population with 67 percent whites and 33 percent other races. (Topic 47 shows that the binomial simulator **randBin** works very well.) Start with samples of size 10.

p = 0.67, n = 10

1. Set a random seed as explained in Topic 21 and given in the first two lines in screen 41.

```
(41) 321→rand
      randBin(1,.67,10)
```

2. Press **MATH** <**PRB**>**7:randBin(1** **▢** **0** **▢** **67** **▢** **10** representing *one* pick at a time with a *0.67* chance of a success and repeating the selection *ten* times.

```
(42) (1,0,1,1,1,1,0,1,0,0)
```

3. Press **ENTER** **ENTER** **ENTER** for the three sets of 0s and 1s, as shown in screens 42, 43, and 44 (your screen will not give the complete list, but using **▸** will reveal the last values).

```
(43) (1,1,0,1,1,1,0,1,1,0)
```

One (1) represents a “white,” thus with 67 percent, or approximately 70 percent whites, you would expect *about* seven 1s and three 0s. The first three samples reveal six, seven, and nine 1s, respectively, out of 10, or 0.6, 0.7, and 0.9 proportion of successes.

```
(44) (1,0,1,1,1,1,1,1,1,1)
```

4. Reset the seed, and calculate the sum of ten values generated at a time from the population with $p = 0.67$ for 100 samples.

Divide each sum by 10 for the proportion of success in each sample, and store in L_1 . (See screen 45.)

Notice in screen 45 that the first three values of **.6**, **.7**, and **.9** are from the first three samples above with sums of 6, 7, and 10.

5. Set up **Plot1** for a **Histogram** as shown in screen 46, and set the **WINDOW** as shown in screen 47.
6. Press **TRACE**, and use **▸** to highlight the class with the largest frequency (26), which is also the class that contains the value of 0.67, as shown in screen 48.

The distribution is somewhat skewed to the left because with $p = 0.67$, we would not expect many samples to have no whites or even only one or two whites. The frequency distribution from the previous **Histogram** is given in the following table.

p:	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Frequency:	0	0	0	2	3	12	23	26	24	10	0

The measure of the spread for a binomial distribution is $\sigma = \sqrt{p(1-p)/n} = \sqrt{(0.67 * .33/10)} = .1487 = 0.15$, as shown in the first two lines in screen 49.

7. Press **STAT** <**CALC**> **1:1-Var Stats L1** **ENTER**.

This gives $Sx = .1385 = 0.14$ (screen 50) from the simulation, which is approximately the same as above. The mean of **0.68** is close to the theoretical value of 0.67.

```
321→rand
randBin(10,0.67,
100)/10→L1
(.6 .7 .9 .7 .6...
```

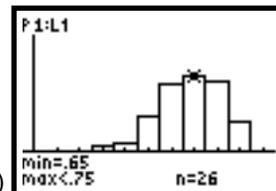
(45)

```
Plot1 Plot2 Plot3
Off Off
Type: L1 L2 L3
Xlist:L1
Freq:1
```

(46)

```
WINDOW
Xmin=-.05
Xmax=1.05
Xscl=.1
Ymin=-11
Ymax=46
Yscl=0
Xres=1
```

(47)



(48)

```
√(.67*.33/10)
.1486943173
1-Var Stats L1
```

(49)

```
1-Var Stats
x̄=.68
Σx=68
Σx²=48.14
Sx=.1385349024
σx=.1378404875
↓n=100
```

(50)

Activity 6, Demonstrating Probability, Simulations, and Probability Distributions (cont.)

$p = 0.67$, $n = 30$

Repeat the steps on the previous page with a larger sample size of $n = 30$.

The distribution becomes more symmetric and more closely centered on $p = 0.67$. Note that the first sample gives $(6 + 7 + 9)/30 = .73333333$. (See screen 51.) Recall that 6, 7, and 9 were the totals for the first three batches of ten values simulated. (See screen 45.)

As you can see in screen 52, as the sample size increases, the distribution of sample proportions is squeezing more closely to the population value of $p = 0.67$.

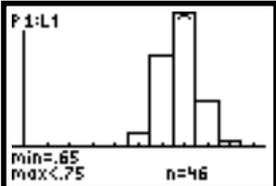
$Sx = .0831 = 0.083$ is very close to $\sigma = .0858 = 0.086$. The mean of 0.679 is close to the theoretical value of 0.670 . (See the calculations in screens 53 and 54.)

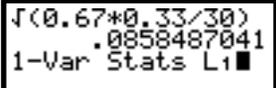
Because the mean of the sample proportions is the same as the population proportion, the sample proportion is said to be an *unbiased estimator* of the population proportion. (See Topic 39.)

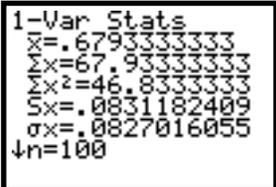
The property of getting a more normal-like distribution with a smaller standard deviation as sample size increases can be proven by the *Central Limit Theorem*. The *Central Limit Theorem* is usually discussed in terms of sample means (see Topic 26). If you consider a success a “1” and a failure a “0,” then a proportion is indeed a mean.

Note: It may be helpful for a class of students to first do the simulation on the previous page by tossing dice. Have each student toss a die ten times and record the number showing, for example, (6, 2, 3, 4, 3, 2, 5, 4, 6, 6). Count the number of times 3, 4, 5, or 6 appears (eight times in our example or a proportion of 0.80). If 25 students do this one time (or four times) each, store the 25 (or 100) proportions in L_1 , construct a **Histogram**, and then calculate the **1-Var Stats**. The results obtained will be similar to that above because $p = 4/6 = 0.666... = 0.67$.

(51) 

(52) 

(53) 

(54) 

Topic 26—Distribution of Sample Means (Central Limit Theorem) (Simulation)

For this topic, set the mode for two decimal places.
(Refer to “Setting Modes” under Do This First.)

The distribution of sample means will be simulated from a continuous uniform distribution with all possible values between 0 and 10 (see screen 55). This population has $\mu = 5.00$ and $\sigma = 2.89$.

As sample sizes from the above population increases, you will see that the distribution of the means of these samples becomes more normally distributed with $\mu = 5.00$ and $\sigma = 2.89/\sqrt{(n)}$.

Sample Size $n = 1$

1. Set a random seed as explained in Topic 21 and shown in the first two lines in screen 56.
2. Type **10** **[MATH]** **<PRB>****1:rand** **[]** **100** **[]** **[STO]** **L1** **[ENTER]**.

This gives a set of 100 random values between 0 and 10, starting with **3.25, 7.44, 3.41, 6.23, 4.28...4.13, 2.70**, as shown in the last line of screen 56. (Use **[]** to see the rest of the values.)

3. Set up **Plot1** for a **Histogram** (screen 57), and set up the **WINDOW** as shown in screen 58.
4. Press **[TRACE]** for the **Histogram** of the 100 values, as shown in screen 59.

Because there are ten classes, you would expect about ten values per class.

5. Press **[STAT]** **<CALC>** **1:1-Var Stats** **L1** **[ENTER]**.

This gives **Sx = 2.74**, or approximately the same as $\sigma = 2.89$, or $2.89/\sqrt{(1)}$. The mean of **4.88** is approximately the same as the theoretical value for all possible samples $\mu = 5.00$. (See screens 60 and 61.)



(56)

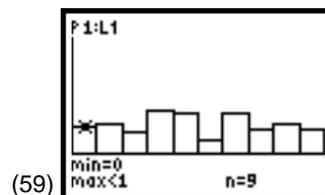
```
321→rand      321.00
10rand(100)→L1
(3.25 7.44 3.41...
```

(57)

```
Plot1 Plot2 Plot3
Off Off Off
Type: L1 L2 L3
Xlist:L1
Freq:1
```

(58)

```
WINDOW
Xmin=0
Xmax=10
Xscl=1
Ymin=-11
Ymax=46
Yscl=0
Xres=1
```



Note: There are 9 values in the first class with 10, 7, 15, 14, 5, 14, 8, 10, and 8 in the other classes. Topic 47 shows that these results are reasonable.

(60) **1-Var Stats L1**

(61)

```
1-Var Stats
x̄=4.88
Σx=488.32
Σx²=3125.98
Sx=2.74
σx=2.72
↓n=100.00
```


Central Limit Theorem

As sample sizes from a population increase, the distribution of the means of these samples becomes more normally distributed with the mean of all possible samples the same as the population mean μ . (The sample mean is an unbiased estimator of the population mean. See Topic 39.) The standard deviation for all possible samples, the population standard deviation, is divided by the square root of the sample size (σ/\sqrt{n}). (For a check on the normality of our samples, see Topic 27.)

Topic 27—Assessing Normality (Normal Probability Plots)

 For screens 70-78 in this topic, set the mode for two decimal places; for screens 79-85, return the mode to floating decimal. (Refer to “Setting Modes” under Do This First.)

A **Histogram** or a stem-and-leaf plot can give an indication if a group of data is normally distributed or not. A normal probability plot is better for assessing normality in that the eye can more easily check if the data points lie on a straight line.

Working with Sample Means

In Topic 26, 100 sample means were generated by generating samples of size 10 with a uniform random-number generator. A **Histogram** was constructed from the sample means (screens 66 and 67). This **Histogram** is duplicated in screens 70 and 71.

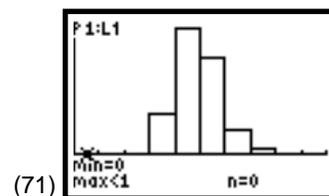
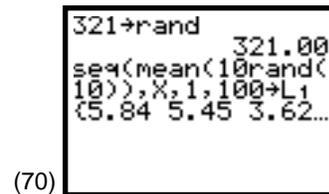
To plot a normal probability plot, turn on **Plot1** as shown in screen 72, and then press **ZOOM** **9:ZoomStat** **TRACE**.

The plot constructed (see screen 73) gives good evidence of normality because the line is fairly straight and the points are more dense in the center while thinning out in the tails.

Note: It may be helpful with a class of students to first do a simulation by tossing a die. Although these will be discrete rather than continuous distributions, the idea is still clear. For example, if a student tosses a die 10 times for the following results grouped into two sets of five values: [(6,2,3,4,3), (2,5,4,6,6)], there are no 1s, two 2s, two 3s, two 4s, one 5, and three 6s. These can be combined with other class members' results for a fairly uniform distribution of 1s, 2s, 3s, 4s, 5s, and 6s with a mean of 3.5 and a standard deviation of approximately 1.71.

If you calculate the average of the two groups of five digits, you obtain, in our example, $(6 + 2 + 3 + 4 + 3)/5 = 3.6$ and $(2 + 5 + 4 + 6 + 6)/5 = 4.6$. These two values can be combined with other class members' results and a **Histogram**. Then **1-Var Stats** calculated to reveal a distribution centered around a mean of approximately 3.5 and a standard deviation of approximately $1.71/\sqrt{5} = 0.76$. You could also work with a sample size of 10 in a similar manner.

(Tossing five dice, of course, can be simulated with **MATH** **<PRB>** **5:randInt** **1** **6** **5** **ENTER**.)



Activity 6, Demonstrating Probability, Simulations, and Probability Distributions (cont.)

Viewing a Uniform Distribution

Compare the above data to the 100 randomly generated values between 0 and 10, also from Topic 26 (screens 56 and 59), with a **Histogram** that indicates the data could be from a uniformly distributed population (duplicated in screens 74 and 75).

Set up **Plot1** for a normal probability plot (see screen 72), and press **ZOOM** **9:ZoomStat**. (See screen 76.)

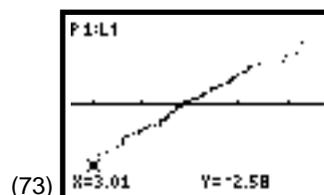
Note that the normal probability plot indicated a symmetric distribution but certainly not a straight line. The points are not more dense in the middle but are uniform throughout. For small x , the n scores (on the y-axis) are more negative than would be expected for a normal distribution. For large x , the n scores are larger than expected.

Viewing a Skewed Distribution

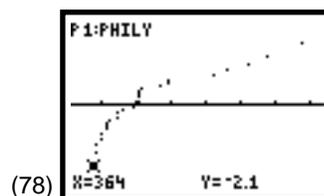
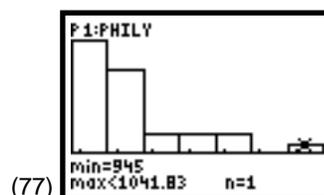
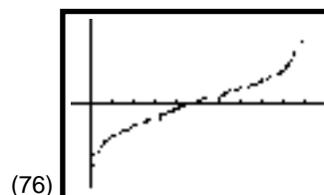
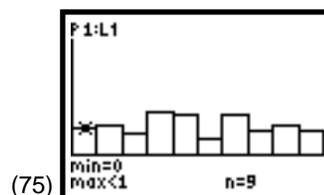
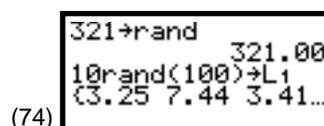
The heights of tall buildings in Philadelphia given in Topic 1 and stored in list **PHILY** have a distribution skewed to the right, as is clear from the **Histogram** repeated here from Topic 1, screen 2. (See screen 77.)

Set up **Plot1** for a normal probability plot with the data list **PHILY**, and then press **ZOOM** **9:ZoomStat** **TRACE** for the plot shown in screen 78.

The data points stretched off to the right in screen 78 are a clear indication of the skewness in that direction.



Note: The x-axis is the data axis. The n score of each data value (x -coordinate) becomes the y -coordinate for the points plotted. (See the reference at the end of this topic.)



Working with Sample Proportions

In Topic 25, the distribution of 100 sample proportions generated from a random binomial distribution with $n = 30$ and $p = 0.67$ was put in a **Histogram** (screens 51 and 52) and is duplicated in screens 79 and 80.

The normal probability plot (screen 81) is fairly straight, but there are groups of points. There are three 0.5s (or 15/30) at the lower left. (Press **TRACE**, and use **▸** to see the coordinates.) There are only two 0.533s (or 16/30), but eleven 0.567s (or 17/30) at the other end with one 0.833 (or 25/30) and two 0.867s (or 26/30).

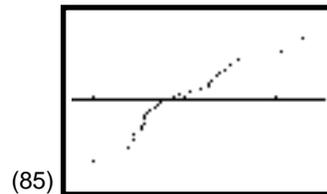
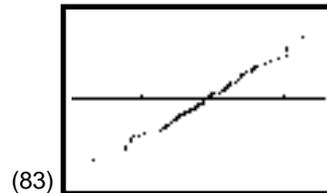
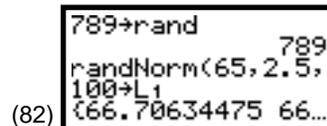
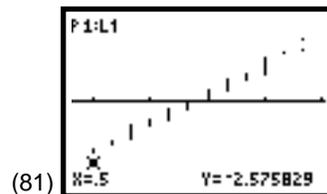
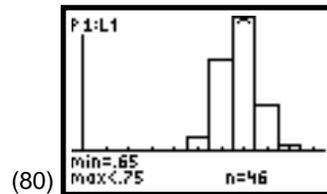
Having stacks of points is called *granularity*. In this case, granularity occurs because we only expect values close to $p = 0.7$ (or 21/30). Between 15/30 and 27/30, there are only 13 possibilities, so with 100 simulations, we are forced to have multiple values. The distribution becomes more normal-like as both $n * p$ and $n * (1 - p)$ become larger, assuring more possible values.

Using Randomly Generated Normal Values

Three different samples of sizes 100, 30, and 10 were generated randomly from a population with a mean of 65 and a standard deviation of 2.5 using **MATH** **<PRB>** **6:randNorm**. (See screens 82, 84, and 86.)

The normal probability plots are given in screens 83, 85, and 87. Notice that as the sample size gets smaller, it is less obvious that the samples are from a normal population. To show the ten points more clearly, note that boxes instead of dots are used in screen 87.

The *Introduction to the Practice of Statistics* by David S. Moore and George P. McCabe (W. H. Freeman and Company) gives a number of examples of using normal probability plots to check if sample data meet the normal distribution stipulations required for some statistical tests.



Activity 6, Demonstrating Probability, Simulations, and Probability Distributions (cont.)

Topic 28—Distribution of the Difference of Two Independent Sample Means (Simulation)

For this topic, set the mode for two decimal places. (Refer to “Setting Modes” under Do This First.)

To show that mean = $\mu_1 - \mu_2$ and standard deviation = $\sqrt{(\sigma_1^2/n_1 + \sigma_2^2/n_2)}$:

- Set the random seed as explained in Topic 21 and shown in the first two lines of screen 88.

Pick a random sample of size $n_1 = 3$ from a normal distribution with $\mu_1 = 85$ and $\sigma_1 = 7$ using [MATH] <PRB> 6:randNorm(85 [] 7 [] 3 [ENTER] for **88.18, 80.41, and 87.88** with a mean of **85.49**.
- Reset the seed, and then pick 100 such samples and store them in **L1** (as shown in screen 89) using [2nd] [LIST] <OPS> 5:seq([2nd] [LIST] <MATH> 3:mean([MATH] <PRB> 6:randNorm(85 [] 7 [] 3 [] [] [] X [] 1 [] 100 [] [STO] **L1**.

This gives a list of means with the first value of **85.49**, which repeats the results in screen 88.

- Generate 100 random samples of size $n_2 = 4$ from a normal distribution with $\mu_2 = 60$ and $\sigma_2 = 8$, and then store in **L2** the same way you did above, with the first mean of **59.85**, as shown in screen 90.
- Store the difference using **L1** [] **L2** [STO] **L3** with the first difference of **85.49 - 59.85 = 25.64** rounded a bit differently from the **25.63** shown. (See the last line in screen 90.)

The mean and the standard deviation of **L1** are **85.01** and **3.95**, which are close to the theoretical values of $\mu_1 = 85$ and $\sigma_1/\sqrt{n_1} = 7/\sqrt{3} = \sqrt{(7^2/3)} = 4.04$. (See Topic 26.)

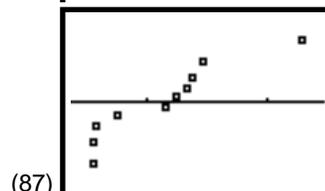
For the sample in **L2**, we have **60.41** and **3.97** compared to 60.00 and 4.00. (See screen 92.)

The mean and the standard deviation of the distribution of our 100 differences in **L3** are **24.60** and **5.54** compared to the theoretical values of $\mu_1 - \mu_2 = 85 - 60 = 25.00$ and $\sqrt{(\sigma_1^2/n_1 + \sigma_2^2/n_2)} = \sqrt{(7^2/3 + 8^2/4)} = 5.69 \neq 4.04 + 4.00$. (See screen 93.)

Note that the variances add and not the standard deviations.

The three stat plots are set up for **Histograms** of the data in **L3**, **L2**, and **L1** with the **WINDOW** and the resulting plot as shown in screens 94 and 95 (text labeling was added later).

(86) 789→rand
randNorm(65,2.5,
10→L1
(66.70634475 66...



(88) 321→rand
randNorm(85,7,3
(88.18 80.41 87...
mean(Ans
85.49

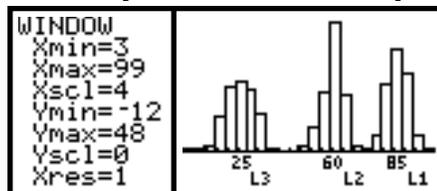
(89) 321→rand
seq(mean(randNor
m(85,7,3)),X,1,1
00)→L1
(85.49 86.19 80...

(90) seq(mean(randNor
m(60,8,4)),X,1,1
00)→L2
(59.85 66.52 59...
L1-L2→L3
(25.63 19.67 20...

(91) mean(L1)
stdDev(L1) 85.01
√(7²/3) 3.95
4.04

(92) mean(L2)
stdDev(L2) 60.41
3.97
√(8²/4) 4.00

(93) mean(L3)
stdDev(L3) 24.60
5.54
√(7²/3+8²/4) 5.69

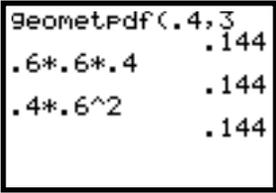


Topic 29—Geometric Distribution

Assume that 40 percent of a large lot of electrical components are from the ABA Company. If components are selected at random:

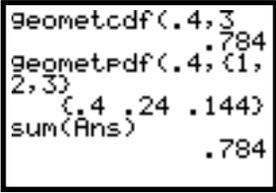
- A. What is the probability that a component from ABA will first be selected on the third pick?

Press $\boxed{2\text{nd}} \boxed{[DIST]} \text{D:geometpdf}(\boxed{.4} \boxed{.} \boxed{3} \boxed{[ENTER]}$ for **.144**, as shown in screen 96.

(96) 

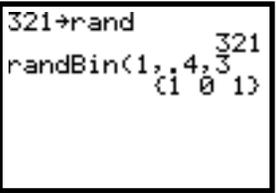
- B. What is the probability that an ABA component will be one of the first three components picked?

For $P(1) + P(2) + P(3)$, press $\boxed{2\text{nd}} \boxed{[DIST]} \text{E:geometcdf}(\boxed{.4} \boxed{.} \boxed{3} \boxed{[ENTER]}$ for **.784**, as shown on the first two lines in screen 97. (The remainder of the screen gives an alternate method.)

(97) 

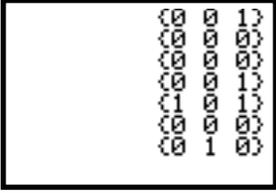
Determining the Probability: Simulation

- Set the seed as shown in Topic 21 using $321 \boxed{[STO\blacktriangleright]} \boxed{[MATH]} \boxed{<PRB>} \boxed{1:rand} \boxed{[ENTER]}$. (See the first two lines of screen 98.)
- Use $\boxed{[MATH]} \boxed{<PRB>} \boxed{7:randBin}(\boxed{1} \boxed{,} \boxed{.} \boxed{4} \boxed{,} \boxed{3} \boxed{[ENTER]}$, as discussed in Topic 21.

(98) 

The result is **{1 0 1}**, where the **1** represents selecting an ABA component and a **0** represents a component from another company. (See the last line of screen 98.)

For each of the three selections, there is a 40 percent chance of selecting an ABA component.

(99) 

- Clear the screen, and press $\boxed{[ENTER]}$ seven times for the results in screen 99.
 - The first simulation (screen 98) and the sixth simulation (fifth line of screen 99) had a component selected on the *first* pick.
 - The second and fifth simulations had a component selected for the first time on the *third* pick.
 - The third, fourth, and seventh simulations had *zero* components selected in three picks.
 - The eighth simulation had a component selected on the *second* pick.

Activity 6, Demonstrating Probability, Simulations, and Probability Distributions (cont.)

The approximate probabilities with only eight tries are thus $P(0) = 3/8 = 0.375$, $P(1) = 2/8 = 0.25$, $P(2) = 1/8 = 0.125$, and $P(3) = 2/8 = 0.25$.

If you repeat step 3 six more times and combine the results, you obtain the following results and comparison with theory.

Simulation	Theory
$P(0) = 11/50 = 0.220$	0.216
$P(1) = 26/50 = 0.520$	0.400
$P(2) = 7/50 = 0.140$	0.240
$P(3) = 6/50 = 0.120$	0.144

Not bad results for only 50 tries.

Topic 30—Hypergeometric Distribution

In a bag of twelve apples, three apples have worms. If you randomly selected two apples (without replacement),

- A. What is the probability that neither apple will have worms?
- Type 0 $\text{[STO] X [ALPHA] [:]}$ (above the [.] key).
 - Type $\text{[] 3 [MATH] <PRB> 3:nCr}$ (as discussed under combinations in Topic 22).
 - Continue as shown in screen 100, storing the results in **A**.

This takes the number of ways you can pick none of the wormy apples and two of the good apples and divides by the number of ways you can pick any two apples for $P(0) = 0.545$.

- B. What is the probability that at least one apple has a worm in it?

In screen 100, use the [2nd] [ENTRY] feature to change the 0 to a 1 and the **A** to a **B** for $P(1) = .409$. In a similar manner, $P(2) = .045$ so that $P(1) + P(2) = 0.454$. (See screens 101 and 102.)

(100)

```

0→X:(3 nCr X)(9
nCr (2-X))/(12 n
Cr 2)→A
.5454545455
    
```

(101)

```

1→X:(3 nCr X)(9
nCr (2-X))/(12 n
Cr 2)→B
.4090909091
2→X:(3 nCr X)(9
nCr (2-X))/(12 n
Cr 2)→C
    
```

(102)

```

C
.0454545455
B+C
.4545454545
1-A
.4545454545
    
```

Determining the Probability: Simulation

1. Set the seed as shown in Topic 21 using **321** **[STO]** **[MATH]** **<PRB> 1:rand** **[ENTER]**. (See the first two lines of screen 103.)
2. Use **[MATH]** **<PRB> 5:randInt (1** **[,]** **12** **[,]** **2** **[ENTER]**, as discussed in Topic 19, for two random integers between 1 and 12. Let **1**, **2**, and **3** represent the apples with worms and **4** and **12** those without.

(103)

```

321→rand          321
randInt(1,12,2
                {4 9}
    
```

The result of **{4 9}** is given in the last line of screen 103.

3. Clear the screen, and press **[ENTER]** seven times for the results in screen 104.

(104)

```

                {5 8}
                {6 4}
                {9 8}
                {11 12}
                {6 8}
                {10 8}
                {1 3}
    
```

It is not until the eighth and last simulation of **{1 3}** that we pick a wormy apple, and then we get two.

The approximate probabilities with only eight tries are thus $P(0) = 7/8 = 0.875$, $P(1) = 0/8 = 0$, and $P(2) = 1/8 = 0.125$.

If you repeat step 3 six more times and combine the results, you obtain the following results and comparison with theory.

<u>Simulation</u>	<u>Theory</u>
$P(0) = 30/50 = 0.600$	0.545
$P(1) = 18/50 = 0.360$	0.409
$P(2) = 2/50 = 0.040$	0.045

Not bad results for only 50 tries.

Activity 6, Demonstrating Probability, Simulations, and Probability Distributions (cont.)

Topic 31—Poisson Distribution

If the mean number of accidents at a given intersection is 2.3 per month,

- A. What is the probability of having exactly two accidents for a given month?

Press $\boxed{2\text{nd}} \boxed{[DISTR]} \mathbf{B:poissonpdf(2 \ . \ 3 \ . \ 2 \ [ENTER]}$ for $P(2) = .265$. (See screen 105.)

```
PoissonPdf(2.3,2)
.2651846416
2.3^2*e^(-2.3)/2
.2651846416
```

- B. What is the probability of having at most two accidents?

Press $\boxed{2\text{nd}} \boxed{[DISTR]} \mathbf{C:poissoncdf(2 \ . \ 3 \ . \ 2 \ [ENTER]}$ for $P(0) + P(1) + P(2) = .596$. (See screens 106 and 107.)

```
Poissoncdf(2.3,2)
.5960388264
PoissonPdf(2.3,(
0,1,2)
(.1002588437 .2...
sum(Ans)
.5960388259
```

- C. What is the probability of having at least two accidents?

For $P(2) + P(3) + P(4) + P(5) + \dots = 1 - (P(0) + P(1)) = 1 - \mathbf{poissoncdf(2.3,1)} = .669$. (See screen 108.)

```
1-poissoncdf(2.3,1)
.6691458157
```

- D. What is the probability of having from two to four accidents?

For $P(2) + P(3) + P(4) = (P(0) + P(1) + P(2) + P(3) + P(4)) - (P(0) + P(1)) = \mathbf{poissoncdf(2.3,4)} - \mathbf{poissoncdf(2.3,1)} = .585$. (See screen 109.)

```
Poissoncdf(2.3,4)
)-Poissoncdf(2.3,1)
.5853950965
sum(PoissonPdf(2.3,(2,3,4))
.5853950964
```

Topic 32—Exponential Distribution

Assume that the shelf-life of cake is exponentially distributed with a mean of $\mu = 4$ days. What proportion of the cakes put on the shelf today would you expect to still be fresh 5 days from now?

$$P(a \leq x \leq b) = e^{-(a/\mu)} - e^{-(b/\mu)}$$

$$P(5 \leq x < \text{E99}) = \text{2nd} [e^x] (\ominus) 5 (\div) 4 \text{] } \text{2nd} [e^x] (\ominus) \text{E99} (\div) 4 \text{] } = .2865 \text{ (See screen 110.)}$$

The $\text{2nd} [e^x]$ key is in the first column above LN .

E99 represents infinity or a very large number, as given in Do This First in the beginning of this handbook.

Exponential pdf: $f(x) = (1/\mu) * e^{-(x/\mu)}$

$$f(0) = 1/\mu = 1/4 = 0.25 = \text{Ymax (approximation)}$$

Put the **pdf** in the **Y=** editor as shown in screen 111, set the **WINDOW** as shown in screen 112, and press TRACE for the plot shown in screen 113.

To shade the area of interest, press $\text{2nd} [\text{DRAW}] 7:\text{Shade}(0, \text{Y1}, 5, 20)$ Y1 5 20 ENTER (screen 114) for the shaded region shown in screen 115. **Y1** is pasted from **VARS <Y-VARS> 1:Function 1:Y1**.

```
e^(-5/4)-e^(-E99/4)
      .2865047969
```

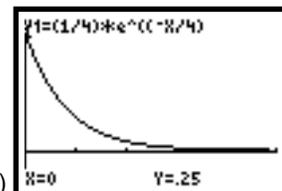
(110)

```
Plot1 Plot2 Plot3
Y1=(1/4)*e^(-X/4)
Y2=
Y3=
Y4=
Y5=
Y6=
```

(111)

```
WINDOW
Xmin=0
Xmax=25
Xscl=5
Ymin=0.07
Ymax=0.28
Yscl=0
Xres=1
```

(112)



```
Shade(0, Y1, 5, 20)
```

(114)

