EXPLORATIONS

Chapter 7

Probability as Relative Frequency

Topic 15 begins with a simulation to justify the concept of the Law of Large Numbers. This concept states that as you take larger and larger samples, the relative frequency of which an event happens approaches the probability of that event happening. Topics 16 and 17 cover the binomial and geometric probability functions of the TI-89, along with simulations to understand the results, and calculations to find the mean and standard deviations.

Topic 15—Law of Large Numbers Simulation

Simple Random Samples

Example: A baseball player has a true .333 batting average (has a probability of 1/3 of getting a hit each time at bat). This could also be thought of as 1/3 of the population favors your candidate in a primary election (for those who do not like baseball). The baseball player might get four hits in a row, or miss five times in a row, but in the long run you expect him to get 1/3, or 33 1/3% hits. (The percent of those polled that favor your candidate will fluctuate widely for small samples, but will approach the true 33 1/3% as sample size approaches the population size.)

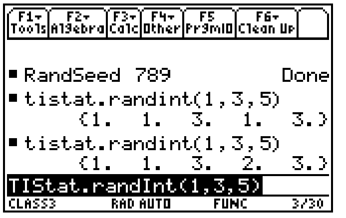
You will simulate each at-bat as a toss of a die, but will only consider 1, 2, or 3 spots — you will ignore 4, 5, or 6. If you were actually tossing die by hand it would be more efficient to use 1 and 6 (for example) as a hit, and 2, 3, 4, or 5 as no hit.

For this chapter, use folder **CLASS3**. To change folders:

1. Press [MODE].
2. Press ▽ to select Current Folder, and then press ▷.
3. Highlight **CLASS3** and press [ENTER].

From the Home screen:

1. Press [CATALOG] and select **RandSeed**.
2. Press [ENTER] to paste **RandSeed** to the status line.
3. Type **789** and press [ENTER] (top of screen 1).



(1)

From the Home screen:

1. Press [CATALOG], and then press [F3] **Flash Apps**.
2. Select **randint(…tistat** and paste it in the status line.
3. Type **1,3,5)**.
4. Press [ENTER] [ENTER] for results of **{1 1 3 1 3}** and **{1 1 3 2 3}** which can be considered as:

1	**1**	**3**	**1**	**3**
S	**S**	**F**	**S**	**F**

and

1	**1**	**3**	**2**	**3**
S	**S**	**F**	**F**	**F**

where **S** represents a success or hit and **F** represents a failure or miss.

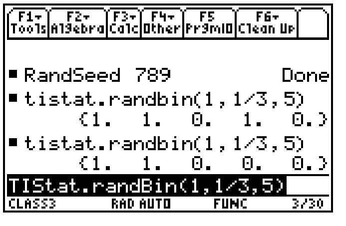
5. Repeat **RandSeed 789**. From the Flash Apps menu, enter **randbin(…tistat**, and then type **(1,1/3,5)**, and press [ENTER] [ENTER] for:

1	**1**	**0**	**1**	**0**
S	**S**	**F**	**S**	**F**

and

1	**1**	**0**	**0**	**0**
S	**S**	**F**	**F**	**F**

(See screen 2.)



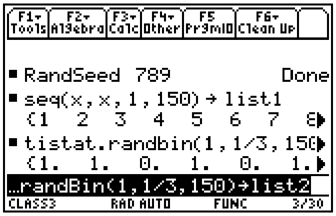
(2)

You see that both simulations give the same results, with **1** for successes and **0** for failures. Notice that after two at-bats, the baseball player is batting 100%, after three at-bats, 67%, and after four at-bats, 75%.

There is a lot of variability, so observe what happens in the long run.

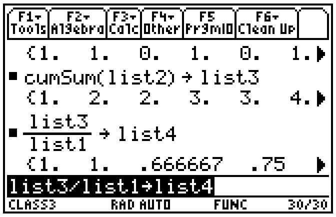
6. Set **RandSeed 789** (top of screen 3).
7. Enter the integers **1** to **150** in **list1**, with **seq (x,x,1,150)→list1** as in Topic 1.
8. Store the results of 150 at-bats with **tistat.randbin(1,1/3,150)→list 2**, which starts with **{1 1 0 1 0 1}** (bottom of screen 3).

(3)



9. Type **cumsum(list2)→list3** and press [ENTER] for **{1 2 2 3 3 4 . . .}** cumulative sums, which indicates you started with two consecutive hits = 1 + 1, but still only had two hits after the third time at bat, or 2 = 1 + 1 + 0, but picked up the third hit on the fourth time at bat with 3 = 1 + 1 + 0 + 1 (screen 4).

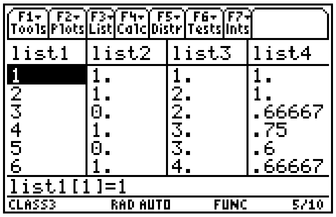
(4)



10. **list3/list1→list4** gives **{1 1 .666667 .75 . . .}**, indicating that you are hitting 100% for the first two times at bat, but 2 of 3 = 67% after 3 at-bats, and 3 of 4 = 75% after four at-bats (screen 4).

See the summary in the Stats/List Editor (screen 5).

(5)

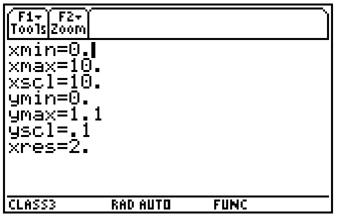


To view this graphically from the Stats/List Editor:

1. Set up and define **Plot 1** as Plot Type: **xyline**, Mark: **Dot**, X List: **list1**, Y List: **list4**, and Use Freq and Categories?: **NO**.
2. Set up ◆ [Y=] with **y1 = 1/3**. Select [F6] **Style**, **2:Dot**.
3. Deselect other plots and functions.

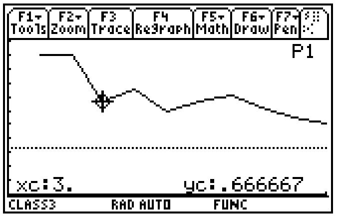
4. Set up the window using ◆ [WINDOW] with the following entries:
   - **xmin = 0**
   - **xmax = 10**
   - **xscl = 10**
   - **ymin = 0**
   - **ymax = 1.1**
   - **yscl = .1**
   - **xres = 2**

   (See screen 6.)



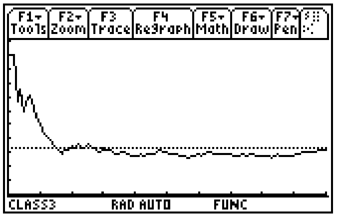
(6)

5. Press ◆ [GRAPH], [F3] **Trace**, and ▶ ▶ for the first 10 at-bats (screen 7).



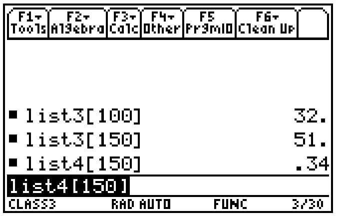
(7)

6. Change the **xmax** value to **100** (◆ [WINDOW]) and then press ◆ [GRAPH] (screen 8). Notice the first tenth of screen 8 duplicates the whole of screen 7 compressed. After 100 at-bats, the line is quite close to 1/3.



(8)

In screen 9, notice that after 100 at-bats there are 32 hits or 32% hits, **list3[100] = 32**, and 1 out of 100 off the true average of .33. After 150 at-bats, there are **list3[150] = 51**. You are only one over the 50 out of 150 that would give 33 1/3%.



(9)

The Law of Large Numbers is sometimes given in terms of the mean, indicating that as the sample size increases, the mean of the sample will approach the population mean. With **1** being a success and **0** a failure, the mean of this sample of **1**'s and **0**'s is the number of hits out of the number at-bats, or a proportion, as in this example:

$$\frac{32}{100} = .32 \quad \text{or} \quad \frac{51}{150} = .34.$$

Topic 16—Binomial Distribution

Example: A hitter has a probability of 1/3 of getting a hit each time at bat, with each at-bat independent of other at-bats. (A population has 33 1/3% of the people favoring a primary candidate and people are picked at random from the population.)

In the next five times at bat (or in a random sample of size 5):

1. What is the probability of getting exactly three hits?

   a. From the Stats/List Editor, press [F5] **Distr** for **B:Binomial Pdf**.

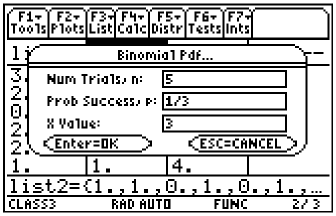
   b. Enter these settings:

      Num Trials, n: **5**

      Prob Success, p: **1/3**

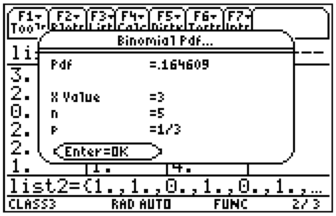
      X Value: **3**

   (See screen 10.)



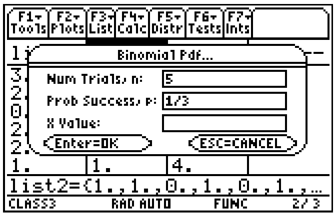
(10)

   c. Press [ENTER] to display Pdf = **.164609** = **P(3)** (screen 11). The hitter has about a 16% chance of getting three hits.

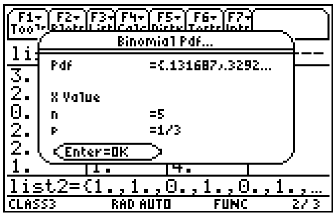


(11)

   If the X Value is left blank as in screen 12, then the entire probability distribution is partially shown in a list called **Pdf** (screen 13).

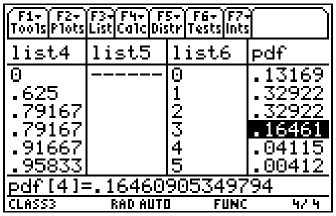


(12)



(13)

d. Press ENTER to display screen 14 with list **Pdf** pasted in the last column of the Stats/List Editor. The fourth value is **Pdf[3] = P(3) = .16461** as before.

(14)


2. What is the probability of getting at least two hits?

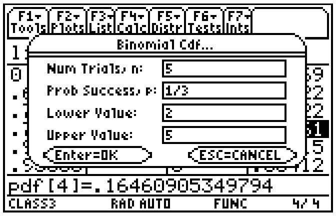
   a. You want **P(2) + P(3) + P(4) + P(5)** which you could add from screen 14, or press F5 **Distr** and select **C:Binomial Cdf** to display screen 15 with these settings:

      Num Trials, n: **5**

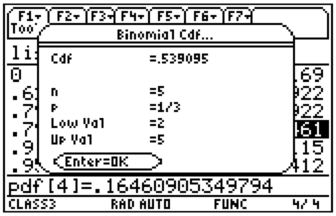
      Prob Success, p: **1/3**

      Lower Value: **2**

      Upper Value: **5**

(15)


   b. Press ENTER for a result of **P(2) +P(3) +…+P(5) = 0.539095**, or over a 50% chance of getting at least two hits in five times at bat (screen 16).

(16)


The Mean and Standard Deviation

For a binomial distribution with n **= 5**, p **= 1/3**:

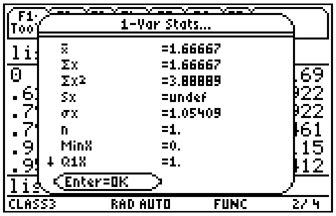
1. Set up a new list, **list6**, with values **0**, **1**, **2**, **3**, **4**, and **5**. These values represent each time at bat. (See screen 14.)

2. Press F4 **Calc**, **1:1-Var Stats**, with List: **list6**, and Freq: **statvars\pdf**.

3. Press ENTER to display screen 17, with

   $\bar{x}$ **= 1.66667 =** μ **=** np **= 5 ∗ (1/3)**, or the expected number of hits, in the long run, for five times at bat. You could also verify from the table of values that the standard deviation

   $s_x = 1.05409 = \sigma_x = \sqrt{npq} = \sqrt{5*(1/3)*(2/3)} = 1.05409.$

(17)


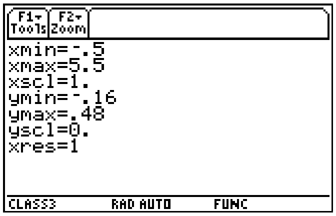
Probability Histogram

1. Set up and define **Plot 1** as Plot Type: **Histogram**, X List: **list6** (from the **class3** folder), Hist. Bucket Width: **1**, Use Freq and Categories?: **YES**, and Freq: **Statvars\pdf**.

2. Set up the window using ◆ [WINDOW] with the following entries:

   - **xmin = -.5**
   - **xmax = 5.5**
   - **xscl = 1**
   - **ymin = -.16**
   - **ymax = .48**
   - **yscl = 0**
   - **xres = 1**

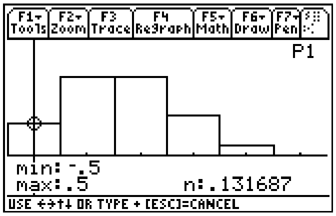
   (See screen 18.)



(18)

3. Press ◆ [GRAPH], and then press F3 **Trace** (screen 19).

Notice the center and spread. What does the histogram tell you about the hitter?

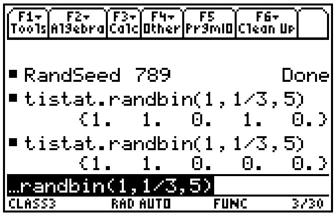


(19)

Simulation

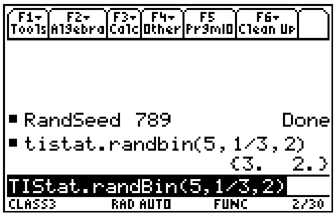
From Topic 15, you could toss a die repeatedly in groups of five tosses at a time (screen 20). To repeat the experiments:

1. Set **RandSeed 789** (top of screen 20).



(20)

2. Enter **tistat.randbin(5,1/3,2)** and press ENTER (screen 21).

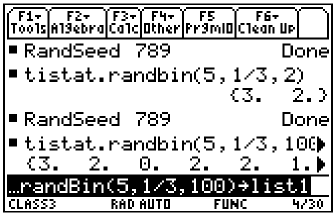


(21)

To simulate 100 experiments of five at-bats:

1. Set **RandSeed 789** (see Topic 15).

2. Calculate and store the number of successes for each of the 100 experiments in **list1** with **tistat.randbin (5,1/3,100)→list1** (screen 22).

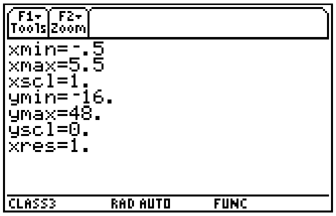
(22) 

3. Set up and define **Plot 1** as Plot Type: **Histogram**, X List: **list1**, Hist. Bucket Width: **1**, and Use Freq and Categories?: **NO**.

4. Set up the window using ◆ [WINDOW] with the following entries:

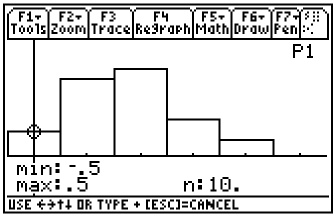
   - **xmin = -.5**
   - **xmax = 5.5**
   - **xscl = 1**
   - **ymin = -16**
   - **ymax = 48**
   - **yscl = 0**
   - **xres = 1**

   (See screen 23.)

(23) 

5. Press ◆ [GRAPH], and then press [F3] **Trace** (screen 24).

6. Compare with screen 19 and the following table. The P(X) values are from list **Pdf** in screen 14. For example, $x = 0$ occurs 10 times in the simulation, while you would expect it to occur about $100 * .1317 \approx 13.17$ times.

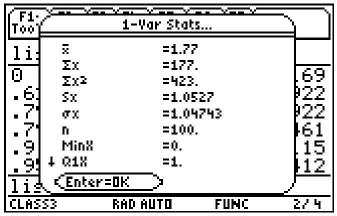
(24) 

X	Freq	P(X)
0	10	.1317
1	32	.3292
2	36	.3292
3	15	.1646
4	7	.0412
5	0	.0041
Total	100	1.0000

These comparisons show a reasonable simulation for using only n = **100**.

7. From the Stats/List Editor, press [F4] **Calc**, **1:1-VarStats**, with List: **list1**, Freq: **1** and press [ENTER] [ENTER] (screen 25).

8. Compare $\bar{x}$ **= 1.77** and s_x **= 1.0527** (screen 25) with μ **= 1.67** and σ_x **= 1.054** (screen 17).

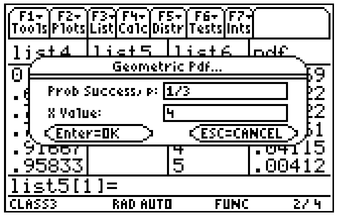
(25)


Topic 17—Geometric Distribution

Example: If the probability of getting a hit is 1/3:

1. What is the probability that the first hit will occur on the fourth time at bat? (If 1/3 of the population prefers a particular brand product, what is the probability that the fourth person randomly selected is the first to prefer this product?)

   a. From the Stats/List Editor, press [F5] **Distr**, **F:Geometric Pdf**, with Prob Success, p: **1/3**, and X Value: **4** (screen 26).

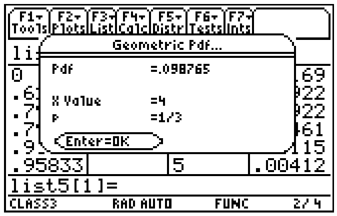
(26)


   b. Press [ENTER] for Pdf **= 0.098765 = P(4)**, or about a 10% chance for a first hit on the fourth at-bat (screen 27).

   c. From the Home screen this would be calculated by

$$\left(\frac{2}{3}\right)\left(\frac{2}{3}\right)\left(\frac{2}{3}\right)\left(\frac{1}{3}\right)=\left(\frac{2}{3}\right)^3*\left(\frac{1}{3}\right)=\frac{8}{81}=.098765.$$

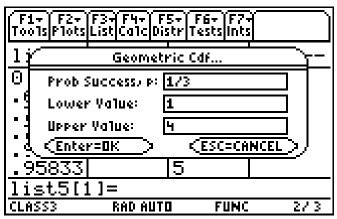
   There is a 2/3 probability that the batter does not get a hit for each of his first three at-bats and a 1/3 probability of getting a hit on his fourth at-bat.

(27)


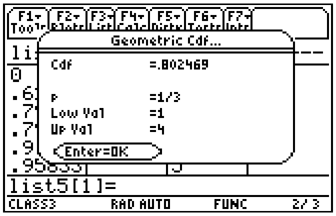
***Note:** This answer is the fourth entry in the* list **Pdf** *shown in screen 32.*

2. What is the probability that a hit will occur in one of the first four trips to the plate, or that the product is favored by one of the first four people sampled?

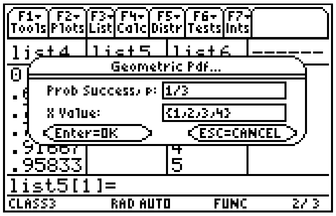
   a. From the Stats/List Editor, press [F5] **Distr**, **G:Geometric Cdf**, with Lower Value: **1**, and Upper Value: **4** (screen 28).

(28)


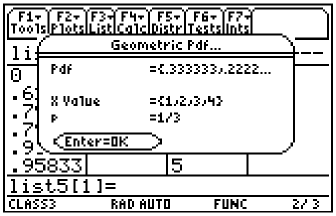
b. Press [ENTER] to display
Cdf **= .802469 = P(1) + P(2) + P(3) + P(4)** (screen 29).

(29) 

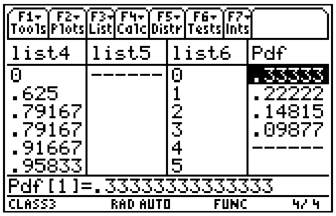
c. The **Geometric Pdf** ([F5] **Distr**, **F:**) could have been used with Prob Success, p: **1/3**, and X Value: **{1,2,3,4}** (screen 30).

(30) 

d. Press [ENTER] to display **P(1)** and **P(2)** (screen 31).

(31) 

e. Press [ENTER] and the last list in the Stats/List Editor is named **Pdf**. This list has individual values for **P(1)**, **P(2)**, **P(3)**, and **P(4)** (screen 32).

(32) 

3. What is the probability that it will take more than four at-bats to get a hit?

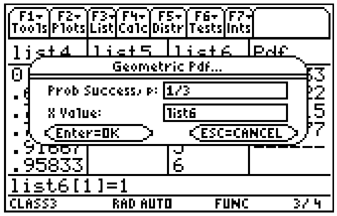
P(5) + P(6) + P(7) + …. = 1 - [P(1) + P(2) + P(3) + P(4)] = 1 - .802469 = 0.197531, or about a 20% chance.

The Mean of a Geometric Distribution

Example: Estimate the mean of a geometric distribution with p = **1/3**.

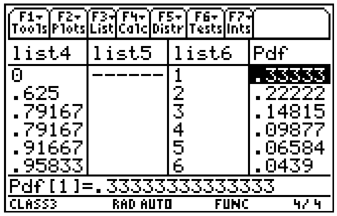
1. Store the integers from 1 to 50 in **list6** with **seq(x,x,1,50)** as in Topic 1.

2. Store the geometric distribution probabilities (at least the first 50) in list **Pdf** by selecting [F5] **Distr**, **F:Geometric Pdf**, with Prob Success, p: **1/3**, and X Value: **list6** (screen 33).



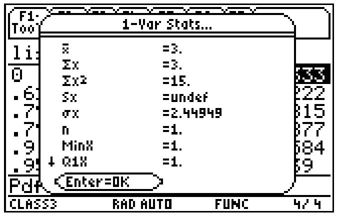
(33)

3. Press [ENTER] to return to the Stats/List Editor (screen 34).



(34)

4. Press [F4] **Calc**, **1:1-VarStats** on List: **list6**, Freq: **statvars\pdf**, and press [ENTER] for $\bar{x}$ **= 3** (screen 35). This confirms what theory tells you, that is, $\mu = 1/p = 1 \div 1/3 = 3$.



(35)

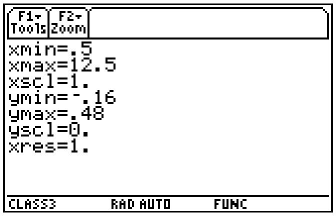
Probability Histogram

1. Set up and define **Plot 1** as Plot Type: **Histogram**, X List: **list6**, Hist. Bucket Width: **1**, Use Freq and Categories?: **YES**, and Freq: **statvars\pdf**.

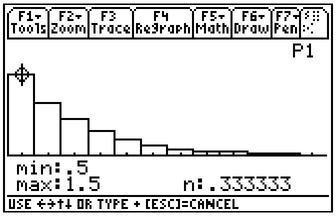
2. Set up the window using ◆ [WINDOW] with the following entries:

   - **xmin = .5**
   - **xmax = 12.5**
   - **xscl = 1**
   - **ymin = -.16**
   - **ymax = .48**
   - **yscl = 0**
   - **xres = 1**

   (See screen 36.)

(36)


3. Press ◆ [GRAPH], and then press [F3] **Trace** with a distribution skewed to the right and each probability rectangle 2/3 the height of the previous one (screen 37).

(37)


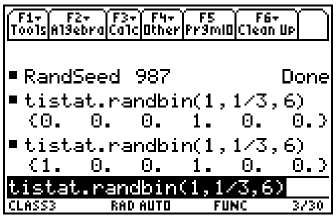
Simulation

From the Home screen:

Note: *This simulation could be done by tossing a die as in Topic 15.*

1. Set **RandSeed 987** (see Topic 15).
2. Enter **tistat.randbin(1,1/3,6)**.
3. Press [ENTER] [ENTER] with the first 10 values (screen 38).

0	**0**	**0**	**1**	**0**	**0**	**1**	**0**	**0**	**1**
F	**F**	**F**	**S**	**F**	**F**	**S**	**F**	**F**	**S**

(38)


The first success was on the fourth at-bat (**FFFS**). Since each of the results is independent of the others, you can start the experiment over and a hit on the third at-bat (**FFS**) and again the third at-bat. In summary, **4**, **3**, **3** for the first three experiments.

If you continue in this manner for 10 experiments, the following results are obtained: **4**, **3**, **3**, **3**, **1**, **1**, **3**, **1**, **1**, **5** and could be used to estimate **P(1) ≈ 4/10 = .40, P(2) ≈ 0, P(3) ≈ .40, P(4) ≈ .10, P(5) ≈ .10**.

To get better estimates, continue for 100 experiments and store the results in **list1**. The results, aided by running the program on the next page after setting **RandSeed 987**, are in the table with the actual probabilities also given.

X	Freq	P(X)
1	30	.333
2	26	.222
3	15	.148
4	9	.099
5	6	.066
6	4	.044
7	4	.029
8	3	.020
9	2	.013
10	1	.009
11	0	.006
Total	100	.989

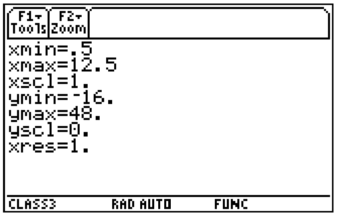
With the outcomes stored in **list1**:

4. Set up and define **Plot 1** as Plot Type: **Histogram**, X List: **list1**, Hist. Bucket Width: **1**, and Use Freq and Categories?: **NO**.

5. Set up the window using ◆ [WINDOW] with the following entries:

   - **xmin = .5**
   - **xmax = 12.5**
   - **xscl = 1**
   - **ymin = -16**
   - **ymax = 48**
   - **yscl = 0**
   - **xres = 1**

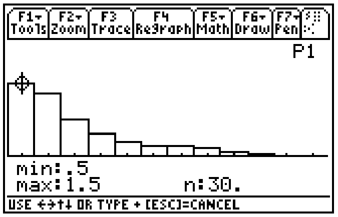
   (See screen 39.)



(39)

6. Press ◆ [GRAPH], and then press [F3] **Trace** (screen 40).

In addition to the relative frequencies being approximately correct in the table, the histogram in screen 40 is very similar to the one in screen 37.



(40)

```
PROGRAM
()
Prgm
TIStat.clrList(list1)
For i, 1, 100
1→count
lbl aaa
If TIStat.randBin(1,1/3)=0 Then
count+1→count
Goto aaa
Else
count→list1[i]
EndIf
EndFor
EndPrgm
```