

Data Analysis and Probability Workshop
Brenda Harlow
Tuesday August 20, 2002

Objective: The participants will use the graphing calculator to enhance the learning of data analysis and probability.

Activity 1 is to be used in the training of teachers to use the TI-83 graphing calculator. Because of the association to gambling, I would not use it in my classroom. However, my co-workers are all adults and they understand that this is an activity for the purpose of learning the calculator and in no way promoting gambling. As a fun part of this activity, I purchased a \$10 gift certificate to our “favorite” local restaurant (the math teachers at my school usually always go to this restaurant as a group for lunch when we have days of school without students) in case someone was “lucky” enough to win this powerball drawing.

Activities 2 – 4 are reproduced, with permission from instructor James Lang, from the Data Analysis and Probability Institute that I attended this summer.

The following list of websites might be helpful in teaching statistics in the classroom. They contain ideas and statistics that can be used in the classroom.

<http://archive.ncsa.uiuc.edu/edu/RSE/RSEyellow/lotto.html>

<http://www.mste.uiuc.edu/hill/dstat/dstatintro.html>

www.calcugator.com

<http://censtats.census.gov/data/WV>

www.as.wvu.edu

www.mste.uiuc.edu/stat/stat.html

Activity 1 – Winning the Lottery

This activity is to be used in the training of teachers to use the TI-83 graphing calculator. Because of the association to gambling, I would not use it in my classroom. However, my co-workers are all adults and they understand that this is an activity for the purpose of learning the calculator and in no way promoting gambling. As a fun part of this activity, I purchased a \$10 gift certificate to our “favorite” local restaurant (the math teachers at my school usually always go to this restaurant as a group for lunch when we have days of school without students) in case someone was “lucky” enough to win this powerball drawing.

This activity uses the following features of the TI-83 graphing calculator:

- random integers
- fractions
- factorial
- combinations (${}_nC_r$)

If the program listed at the end of the activity is discussed, the following features of the calculator are also used.

- store
- sequence
- sort
- dimensions
- lists
- display

Materials needed:

POWERBALL sheet for each participant

TI-83 graphing calculator for each participant

Links of interest associated with this activity:

<http://www.musl.com/powerball.shtm>

A. Complete 2 sections on the *POWERBALL* sheets by picking your “lucky numbers”.

B. Complete 2 sections on the *POWERBALL* sheet using the calculator to choose your numbers. Use the following keystrokes to choose the random integers

- To pick the 1st five numbers → math, PRB, 5 (randInt)
- randInt(will appear on the home screen.
- randInt(1, 49, 5) → we use these limits since the possible choices are 1 to 49 and we want to choose 5 of them. However, this method has “slight” possibility of selecting repeat numbers. If this happens, repeat the process until it selects 5 different numbers.
- To pick the powerball → math, PRB, 5
- randInt(1, 42) → It is not necessary to specify how many to choose, since we only want one number.

C. Complete the last section on the *POWERBALL* sheet using which method you think gives you the best chance to win the “lottery”.

D. Determine the probability of winning the “lottery”

$$\left(\frac{5}{49} \cdot \frac{4}{48} \cdot \frac{3}{47} \cdot \frac{2}{46} \cdot \frac{1}{45} \right) \left(\frac{1}{42} \right)$$

Method 1 → 1.248608925 E -8 → discuss why these fractions
0.00000001248608925

Method 2 → use factorial (!), which is in the Math > PRB menu

$$\frac{(5!)(44!)}{49!} \cdot \frac{1}{42}$$

1.248608925 E -8 → discuss why these numbers
0.00000001248608925

We can see that this is a very small number, thus a very slim chance of winning. Let’s calculate this in terms that are easier to understand.

Since only 1 “set” of numbers will be chosen as the winning numbers, we have a 1 in something chance of winning. To calculate this use one of the following methods:

$$\text{Method 1} \rightarrow \frac{1}{\frac{5}{49} \cdot \frac{4}{48} \cdot \frac{3}{47} \cdot \frac{2}{46} \cdot \frac{1}{45} \cdot \frac{1}{42}} = 80089128 \text{ hint: put the denominator in ()}$$

$$\text{Method 2} \rightarrow \frac{49 \cdot 48 \cdot 47 \cdot 46 \cdot 45 \cdot 42}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 \cdot 1} = 80089128 \text{ hint: put the denominator in ()}$$

$$\text{Method 3} \rightarrow \frac{(49!)(42)}{(5!)(44!)} = 80089128 \text{ hint: put the denominator in ()}$$

Method 4 → Since this is a combination of 49 numbers taken 5 at a time, we can use ${}_n C_r$ from the math > PRB menu.

$$\frac{{}_{49} C_5 \cdot 42}{80089128}$$

Using any of the methods above we see that a probability of 0.00000001248608925 means we have a 1 in 80089128 chance of winning.

E. Now, using a program* that I put into my calculator, I will let my calculator determine the winning numbers. The program is as follows:

```

: seq (X, X, 1, 49, 1) → L1    remember, the arrow is the store (sto) key
: rand (49) → L2
: SortA (L2, L1)
: 5 → dim (L1)
: seq (X, X, 1, 42, 1) → L3
: rand (42) → L4
: SortA (L4, L3)
: 1 → dim (L3)
: Disp “Here are the winning numbers”
: Disp L1
: Disp “with Powerball”
: Disp L3

```

*James Lang, instructor at the Data Analysis and Probability Institute, helped me with this program.

Activity 2 – How long is a minute

Form pairs. One person records time, in seconds, as the other person (the subject) tries to judge when a minute has elapsed. The timekeeper tells the subject when to start and then the subject tells the timekeeper when a minute has gone by. The subject records the time on his/her paper as try 1. The subject and the timekeeper switch positions and repeat the process. Since this is a small group, do the experiment twice and record this time as try 2 on the same piece of paper. If the group is large, do the experiment once and list the times as judge 1 and judge 2.

The papers are then collected and the times are put on the board in columns listed as try 1 and try 2. (Collecting the data in this manner eliminates the fear a student may have of being embarrassed by having to say his/her time aloud.)

Put the data for try 1 into L_1 and the data for try 2 into L_2 . Next use the LIST command augment (under OPS menu) to put the times together into L_3 . The syntax is $\text{augment}(L_1, L_2) \rightarrow L_3$.

Make a histogram for each set of time data.

- Use Xlist: L_1 , L_2 , or L_3
- Use Freq: 1
- Set the window so the data is visible, leaving space at the bottom so the trace values will not cover the histogram
 - a. x-values \rightarrow interval of data
 - b. x-scl \rightarrow bin width
 - c. y-values \rightarrow frequency of data (how many in each bin)
 - d. y-scl and x-res \rightarrow not important for a histogram
- *** Try changing the window a few times to see what affect this has on the looks of the histogram
- Trace the histogram

Activity 2 (con't)

Make side-by-side boxplots for the first and second tries. Compare the distributions.

- For plot 1 → use L_1
- For plot 2 → use L_2
- Trace each boxplot to find the 5-number summary (minimum, lower quartile, median, upper quartile, and maximum)

Use the LIST > Math commands to find the mean, median, and standard deviation of L_3 .

Use the STAT > Calc command 1-var stats to find all of the above information on L_3 .

Activity 3 – Estimating the mean and standard deviation

The following is a class activity to help the student develop an ability to estimate the mean and standard deviation. The class should already have an understanding of mean and standard deviation prior to doing this exercise.

- Divide the participants into groups of size 4. Each person should record his/her height, in inches. Then the group should calculate the mean and standard deviation of the 4 heights.
- Next, have a group stand up. Everyone else should estimate the mean and SD for the group and record it paper. The first group sits down and another group stands up. Again everyone else estimates the mean and SD for the group and records it. Repeat this until all groups have stood.
- Then have each group report the actual mean and SD. Record this in 2 columns on the board. The participants can check their estimates as these are recorded on the board.

Activity 3.1 – Effects of Transforming data

Put the data for height into L_1 . Find the 5-number summary for the heights.

Suppose you express the heights in cm. How will the 5-number summary be changed? How does the mean change? How about the standard deviation? (TI tip: If your data is in L_1 , then on the home screen $2.54L_1 \rightarrow L_2$ creates and stores a list of the corresponding centimeter heights)

Express your heights as number of inches above five feet ($L_1 - 60 \rightarrow L_3$). How does this change the summary statistics?

Take the square root of the heights ($\sqrt{L_1} \rightarrow L_4$). How does this change the summary statistics?

Activity 4 – Finding a linear equation

Since we know that activity 3.1 (height to cm) is a linear regression, it is a good example to start discussing the line of best fit.

Have participants to make a scatterplot of height to cm on the calculator.

- Stat plot > 2nd, Y =
- Enter
- Turn this plot on
- Type: scatteplot
- Xlist: L_1
- Ylist: L_2
- 2nd, Quit
- Set the window \rightarrow a good method for this is to find the 5-number summary of L_1 and L_2 to find the minimum and maximum values and set the window a little larger than that.
- Graph

Activity 4 (con't)

To use the calculator to find the line of best fit, use the STAT > Calc menu and choose LinReg(ax + b) L₁, L₂ (enter). If you would like to see the correlation (r), we need to have the diagnostics on. To do this, go to the catalog (2nd > O), find diagnostics on and press enter twice.

If you want to find the line of best fit and store it in the Y = menu, the syntax is LinReg(ax + b) L₁, L₂, Y₁. (Y₁ is in the VARS > Y-Vars > Function menu). . Graphing will then draw the line.

Activity 4.1 – Using linear regression

A classified ad offered several used Toyota Corollas for sale. Listed below are the ages of the cars and the advertised prices.

Age (years)	Prices advertised
1	12995 , 10950
2	10495
3	10995 , 10995
4	6995 , 7990
5	8700 , 6995
6	5990 , 4995
9	3200 , 2250 , 3995
11	2900 , 2995
13	1750

- Make a scatterplot for these data.
- Describe the association between age and price of a used Corolla.
- Find the equation of the regression line.
- If you want to sell a 7-year old Corolla, what price seems appropriate?

***Remember:** If there is a chance a student might be embarrassed by the answer to a question (*NEVER attempt to do anything with personal weight*), a good method to use is collect the data on small pieces of paper so no one knows who the information belongs to.

Brenda

Examples Of Class Generated Data That Could Be Used

When demonstrating dot diagrams, you may want to collect from each student: the number of working televisions -not a redneck joke - in their home (usually skewed right), their height in inches (usually mounded), the number of years they have lived in Upshur County (usually clusters in 2 or 3 groups), or the number of hours sleep they got last night.

When demonstrating quartiles and box plots you may want to ask the students to count the number of pennies - definitely answer on small paper - they have in their pockets or purses (usually yields a skewed distribution). For a skewed left distribution, ask for the dates on those pennies.

For qualitative data you might collect things like eye color (small number of groups – good for bar graphs and pie charts); favorite ice cream flavor (lots of groups with frequencies of 1 – good for showing the value of limiting choices in questionnaire writing); state or county of birth (good ice-breaker – gets students interested in each other); last fast food restaurant visited (who's number 1 in your class? – Mickey D's or Burger King?); and birth-order (how many “first-born” or “only” are in your class?)

For bivariate data, collect “length of hand” versus “height” in either inches or centimeters. Students pair off and measure each other. This gives a scatterplot with a linear trend. You could also use “fist circumference” versus “foot length” to test the theory that if a sock can wrap around your fist, it will fit your foot.

Hopefully this gives you a few ideas on using probability/ data analysis in your classroom and the technology learned today will enhance the learning process. Thank you so much for attending this workshop.