

# Sampling and Experimental Design

David Ferris, Noblesville High School

## **Some questions that can be answered by collecting data:**

Are coffee drinkers more likely to be female? Are females more likely to drink coffee than males? (What is the difference between these two questions?)

Does the new ebola vaccine work?

Do males have a better reaction time than females?

Do cell phones cause cancer?

How often do people wash their hands after using the restroom?

Can index finger lengths predict risk of prostate cancer?

Is autism onset related to childhood vaccinations?

Do diets work?

Will people drink blue soda?

Do people who text while driving have more accidents?

Is a larger snood in male turkeys related to perceived “toughness” by other male turkeys?

How many birds do cats kill each year?

How great is the placebo effect with depression drugs?

## Random Rectangles Activity:

1. Your  $w$  \_\_\_\_\_  $g$  \_\_\_\_\_ of the average area: \_\_\_\_\_

mean of class: \_\_\_\_\_ standard deviation of class: \_\_\_\_\_

2. P \_\_\_\_\_ f \_\_\_\_\_, then find the average area: \_\_\_\_\_

mean of class: \_\_\_\_\_ standard deviation of class: \_\_\_\_\_

3. R \_\_\_\_\_, then find the average area: \_\_\_\_\_

mean of class: \_\_\_\_\_ standard deviation of class: \_\_\_\_\_

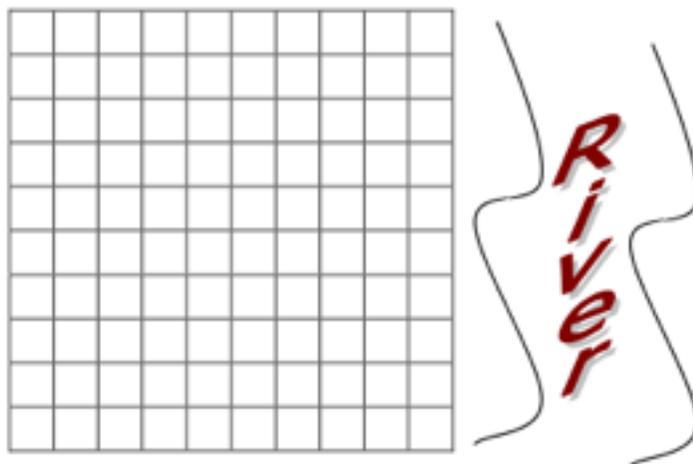
(A similar activity is “Jelly Blubbers.” See [noblestatman.com](http://noblestatman.com))

## An Exercise in Sampling: Rolling Down the River

A farmer has just cleared a new field for corn. It is a unique plot of land in that a river runs along one side. The corn looks good in some areas of the field but not others. The farmer is not sure that harvesting the field is worth the expense. He has decided to harvest 10 plots and use this information to estimate the total yield. Based on this estimate, he will decide whether to harvest the remaining plots.

### A. Method Number One: Convenience Sample

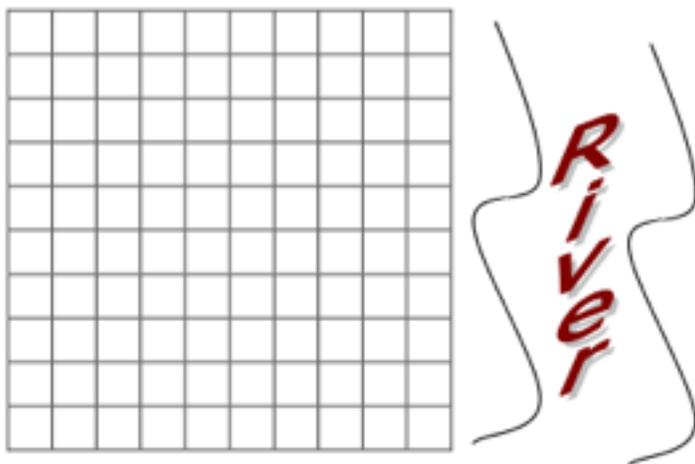
The farmer began by sampling plots easy to access. He drove his tractor out of the barn and sampled 10 contiguous plots, driving only horizontally or vertically. He may have driven over previously sampled plots, but he did not travel through a plot without sampling it. Mark 10 plots the farmer may have used.



Since then, the farmer had second thoughts about this selection and has decided to come to you (knowing that you are an AP statistics student, somewhat knowledgeable, but far cheaper than a professional statistician) to determine the approximate yield of the field. You will still be allowed to pick 10 plots to harvest early. Your job is to determine which of the following methods is the best one to use – and to decide if this is an improvement over the farmer's original plan.

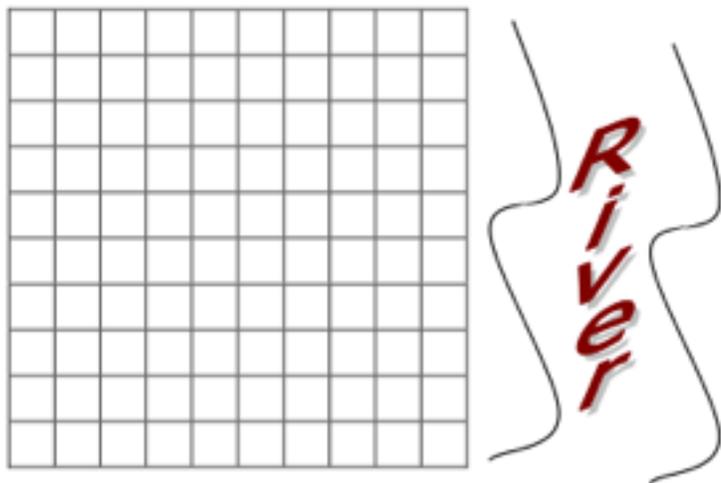
### B. Method Number Two: Simple Random Sample

Use your calculator or a random number table to choose 10 plots to harvest. Mark them on the grid below, and describe your method of selection.



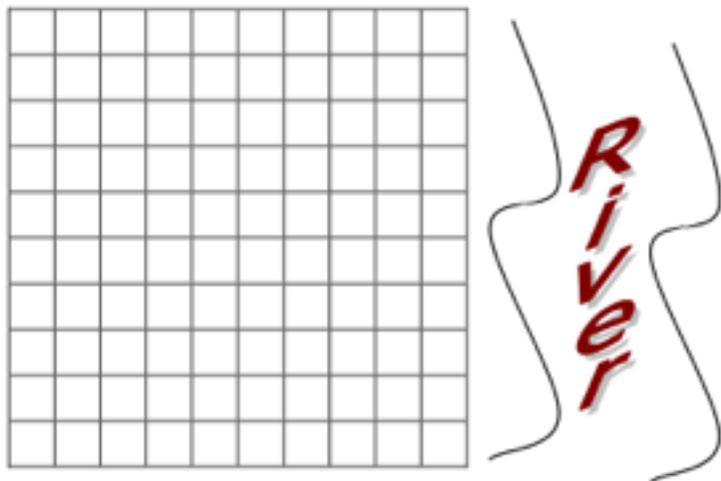
**C. Method Number Three: Stratified Sample**

Consider the field as grouped in vertical columns (called strata). Using your calculator or a random number table, randomly choose one plot from each vertical column and mark these plots on the grid.



**D. Method Number Four: Stratified Sample**

Consider the field as grouped in horizontal rows (also called strata). Using your calculator or a random number table, randomly choose one plot from each horizontal row and mark these plots on the grid.



**OK, the crop is ready.** Your teacher will show you the actual yields in the field. Estimate the average yield per plot based on each of the four sampling techniques.

Sampling Method	Mean yield per plot	Estimate of total yield
Convenience Sample (farmer's)		
Simple Random Sample		
Vertical Strata		
Horizontal Strata		

Observations:

- 1) You have looked at four different methods of choosing plots. Is there a reason, other than convenience, to choose one method over another?
- 2) How did your estimates vary according to the different sampling methods you used?
- 3) Which sampling method should you use? Why do you think this method is best?
- 4) How could the farmer take a **cluster sample** in this activity?

# Introduction to Sampling

1. population vs. sample:

2. A good, representative sample...

Good sampling still has variation from sample to sample, called

\_\_\_\_\_.

3. \_\_\_\_\_ sampling... has SYSTEMATIC variation that misrepresents the population in some important way.

Like spoonfuls of soup from \_\_\_\_\_...

A biased sample cannot typically be \_\_\_\_\_ or \_\_\_\_\_ to extract useful information.

Usually, the best "fix" is to \_\_\_\_\_.

## **TYPES OF SAMPLING:**

4. Simple Random Sample (SRS):

5. systematic sampling

6. cluster sampling

7. stratified sampling:

8. multistage sampling

9. convenience sampling

**TYPES OF BIAS:**

10. response bias

11. nonresponse bias

12. voluntary response bias

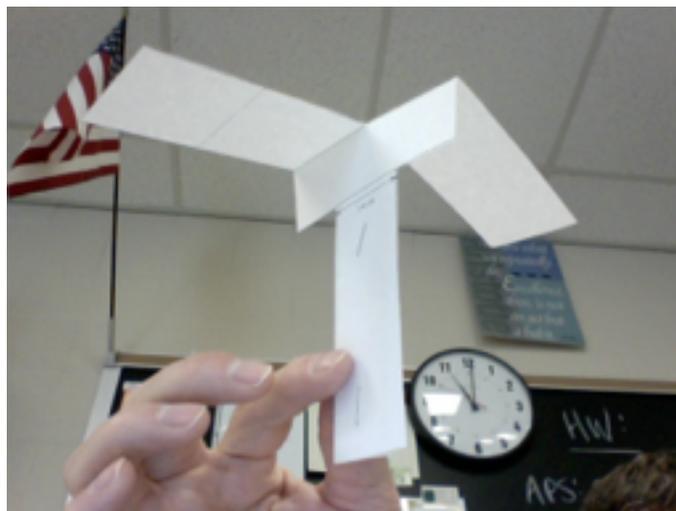
13. undercoverage

14. BE SURE TO KNOW: The difference between variability and bias

# Paper Helicopter Experiment

Name: \_\_\_\_\_

1. Make a paper helicopter according to the directions and the photo below. You will need scissors, two staples, and a template from which to cut out the helicopter.



Type	Height	Distance from target				Mean	SD
Long rotor	(Low)						
Short rotor	(Low)						
Long rotor	(Medium)						
Short rotor	(Medium)						
Long rotor	(High)						
Short rotor	(High)						

2. Using three different heights (low, medium and high—you choose the distances) and two different helicopters (short rotor and long rotor), launch helicopters and measure how far they land from a target point on the floor. Conduct four launches at each position for each type of helicopter. This will give you 24 total launches. Collect the data in the table below, and answer the questions on the next page.

(Long rotor helicopter pictured above.)

# Helicopter Activity Questions

1. The 24 distances you recorded were not all exactly the same—there was a lot of variation. What were some of the sources (causes) of this variation?
2. These sources of variation can be categorized into three broad types. What are they?
3. How could you have “accounted for” the two “non-expected” general sources of variation?
4. What are the factors in the experiment?
5. What are the levels of each factor?
6. How many total treatments did you have? \_\_\_\_\_
7. Name at least three possible sources of bias in this experiment (be sure to remember the definition of bias).

## Experimental Design: Learning to Manage Variability

Floyd Bullard and Dan Teague  
NC School of Science and Mathematics

In George Cobb's excellent text, *Introduction to the Design and Analysis of Experiments*, he describes the variability inherent in an experiment in the following way:

Any experiment is likely to involve three kinds of variability:

1. *Planned, systematic variability.* This is the kind we want since it includes the differences due to the treatments.
2. *Chance-like variability.* This is the kind our probability models allow us to live with. We can estimate the size of this variability if we plan our experiment correctly.
3. *Unplanned, systematic variability.* This kind Threatens Disaster! We deal with this variability in two ways, by randomization and by blocking. Randomization turns unplanned, systematic variation into planned, chance-like variation. Blocking turns unplanned, systematic variation into planned, systematic variation.

The management of these three sources of variation is the essence of experimental design. To focus the discussion, we will consider the variation inherent in the following experimental setting.

**Example Experiment:** Compare two kinds of rabbit food on weight gain (in ounces) from the age of 2 weeks to the age of 6 weeks of life. We want to know if the rabbits will gain more weight on one diet than on the other. We have space to house 8 rabbits for this experiment.

### Sources of Variation

The most obvious is, perhaps, that the rabbits are all different rabbits, and so they will all grow at different rates. If different breeds of rabbit are used, then we will have an additional source of variation. Young California rabbits do not grow at the same rate as young Florida White rabbits. The environment in which the rabbits live will not be exactly the same. They all cannot live in the same location, some will be in slightly warmer areas while others will be in areas with more light. They will not all have exactly the same amount of exercise or sleep. The food will be carefully weighed before it is given to the rabbits, but there will inevitably be measurement error in the amount of food given to each rabbit. Similarly, the rabbits will be weighed before the experiment begins and after the experiment ends. There will be (hopefully small) measurement error in both these weighings.

All of the aspects of the experimental setting mentioned so far can be considered natural chance-like variation. There is another source of essential variation: the systematic difference in the rate of growth that is a result of the different diets. This is a variation that we want to investigate. One way to think about the experiment is that we want to know if the variation that is a result of the diet is larger than the variation that is due to all the natural variation inherent in rabbit growth. In designing our experiment, we want to accentuate this planned, systematic variation, while reducing the natural chance-like variation.

For rest of article, see file: expdesign\_rabbits.pdf

For more good articles about design, see: [http://courses.ncssm.edu/math/Stat\\_Inst/Notes.htm](http://courses.ncssm.edu/math/Stat_Inst/Notes.htm)

Cobb's book (mentioned above) is an even deeper discussion of experimental design.

# Unique Concepts/Terms in Experiments

Blocking

Confounding

Generalizability of studies

Ramsey and Schafer (2002) :

		Assignment of Units to Groups		
		By Randomization	Not by Randomization	
Selection of units	At random	A random sample is selected from one population; units are then randomly assigned to different treatment groups.	Random samples are selected from existing distinct populations.	Inferences to the populations can be drawn.
	Not at random	A group of study units is found; units are then randomly assigned to treatment groups.	Collections of available units from distinct groups are examined.	
		Causal inferences can be drawn.		

The entire document can be found at AP Central



## Placebo and placebo effect

(See YouTube video from 60 minutes: Treating Depression: Is there a placebo effect?)

## Blinding (single and double)

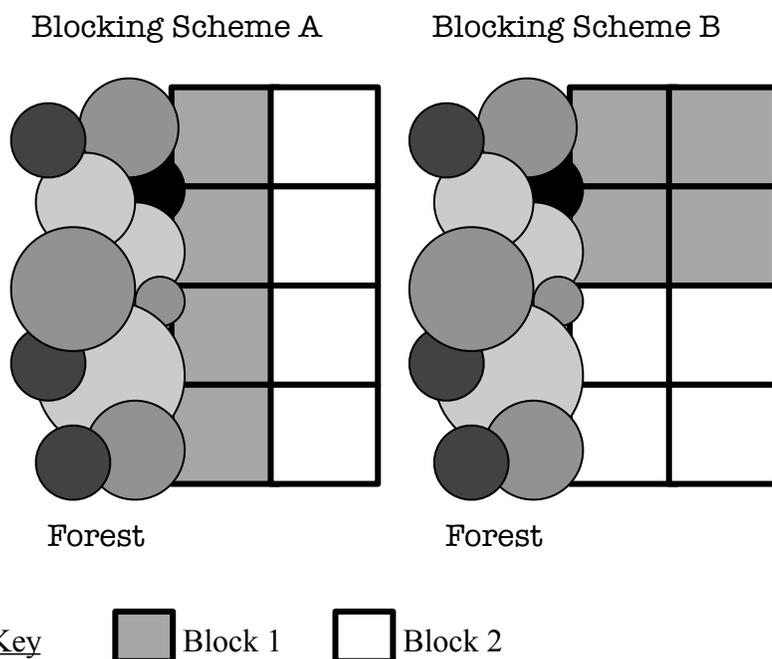
## Three Types of Experiments

## Experimental Design:

### Blocking: Fruit Trees

- Students are designing an experiment to compare the productivity of two varieties of dwarf fruit trees. The site for the experiment is a field that is bordered by a densely forested area on the west side. The field has been divided into eight plots of approximately the same area. The students have decided that the test plots should be blocked. Four trees, each of two varieties, will be assigned at random to the four plots within each block, with one tree planted in each plot.

The two blocking schemes shown below are under consideration. For each scheme, one block is identified by the white region and the other block indicated by the grey region in the figures.



- Which of the blocking schemes, A or B, is better for this experiment? Explain.
- Even though the students have decided to block, they must randomly assign the varieties of the trees to the plots within each block. What is the purpose of this randomization in the context of the experiment?

## **Blocking Activity: Can you see the trees for the forest?**

This activity is a simulation of problem 4 on the 2001 AP Statistics exam. The simulation will ask you to assign trees of two varieties, a new hybrid and a standard variety, to the eight available plots in the field. You'll have the opportunity to assign the trees to the plots in three different ways: a design with no blocking, a design that uses blocking scheme A from the problem, and a design that uses blocking scheme B from the problem. Together with your classmates, your goal is to find the experimental design that has the best chance of correctly identifying which variety of tree performs best, on average.

Let's assume that the new variety really is better. In our simulation, trees of the new variety will have an average production of 53 pounds of fruit, while the old variety typically produces 47 pounds of fruit.

Ideally, designs that have a better chance of detecting the superiority of the new trees are preferred.

We'll also assume that trees planted close to the forest grow more poorly than those away. If a tree is planted toward the forest, its production is reduced by 10 pounds, and for those away the production will be increased by 10 pounds.

Even within a variety, there is variability in the production from tree to tree. Also, individual plots in the field will have differences due to water, fertility, or other factors, adding variability in the production from plot to plot. In this simulation the variability of trees and plots will be accounted for by adding or subtracting a random quantity for each tree. The total productivity will be calculated as a sum:

*Variety (47 or 53) + Proximity to forest (-10 or +10) + Variation (Random)*

Here's one way to determine the random amount to add or subtract, using a coin and a die. First flip the coin: if it's heads, the tree is a little stronger than average, tails a little weaker. How much stronger or weaker is determined by rolling the die. For example, H 3 means add 3 to the productivity of that tree, T 5 means subtract 5.

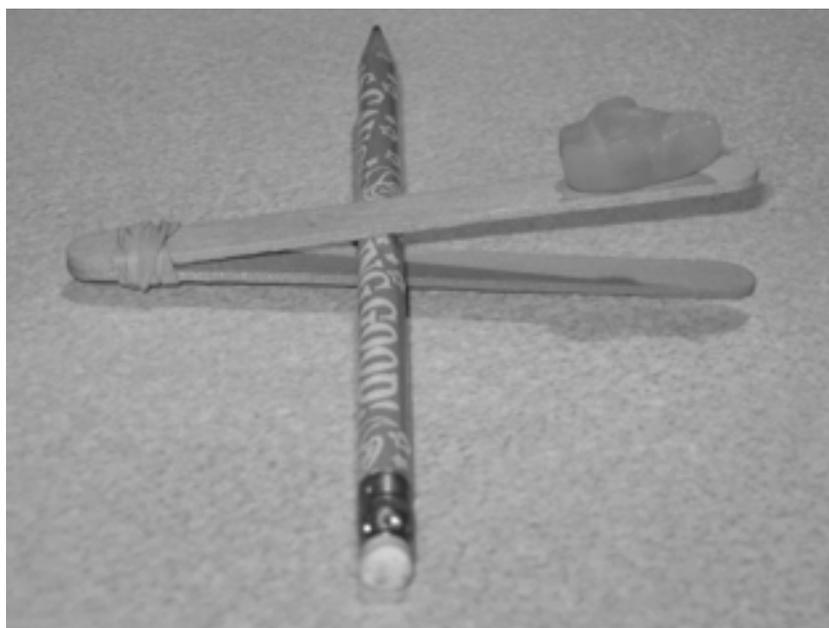
There are other methods of accounting for the differences of individual trees, including using a table of random digits or a graphing calculator. Your teacher may want you to use one of these other strategies.

## Experimental Design Vocabulary:

random	undercoverage
pseudorandom	nonresponse bias
simulation	response bias
sample	observational study
survey	retrospective study
bias	prospective
population	experiment
probability sample	factor
representative sample	subjects
census	experimental unit
statistic	levels (of a factor)
parameter	treatment
simple random sample (SRS)	control, randomization, replication, (blocking)
sampling frame	statistically significant
sampling variability	control group
stratified sample	single-blind
cluster sample	double-blind
multistage sampling	placebo
systematic sampling	placebo effect
voluntary response	confounding
convenience sampling	lurking variable(s)

See [noblestatman.com](http://noblestatman.com) for document summarizing several vocabulary games. Also a Brainscape flashcard set is here: <https://www.brainscape.com/packs/4327749/invitation?referrer=714956>

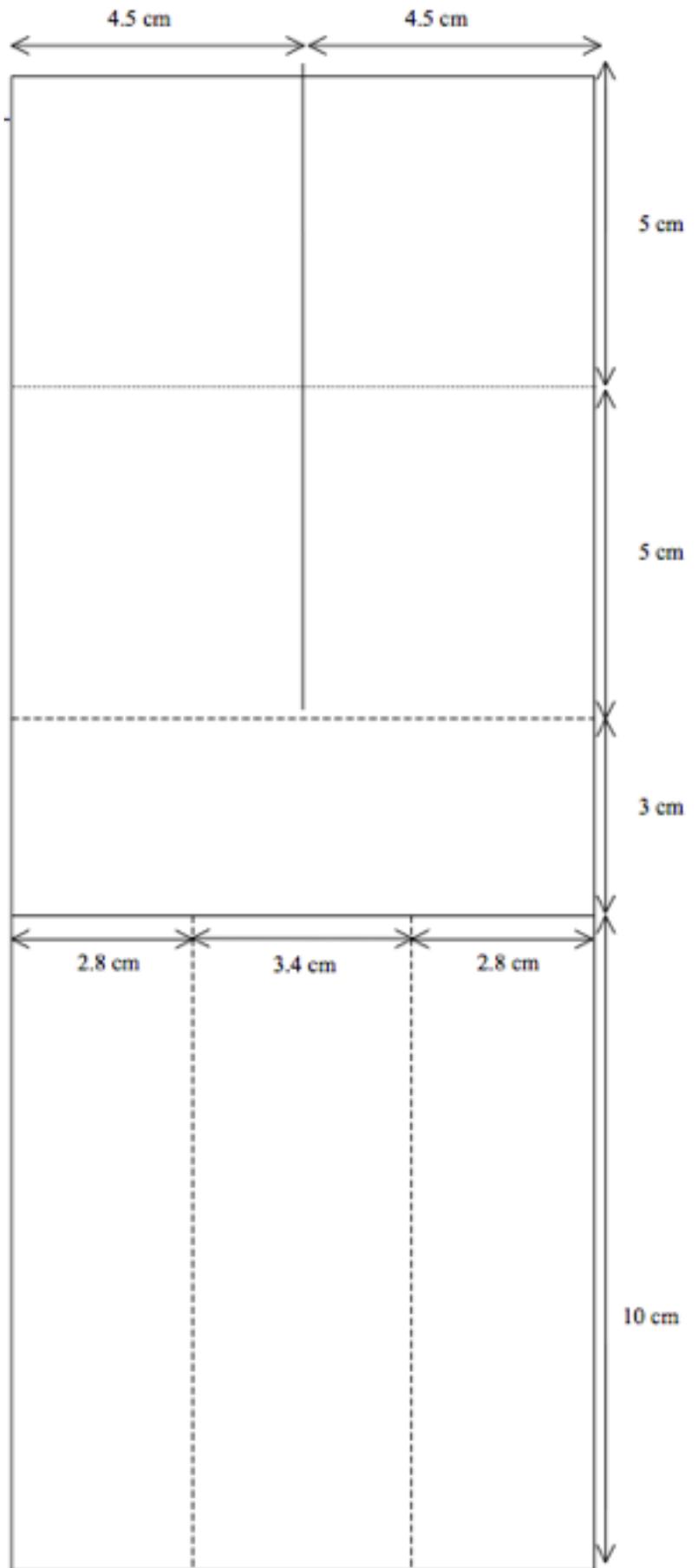
## Gummi Launcher:

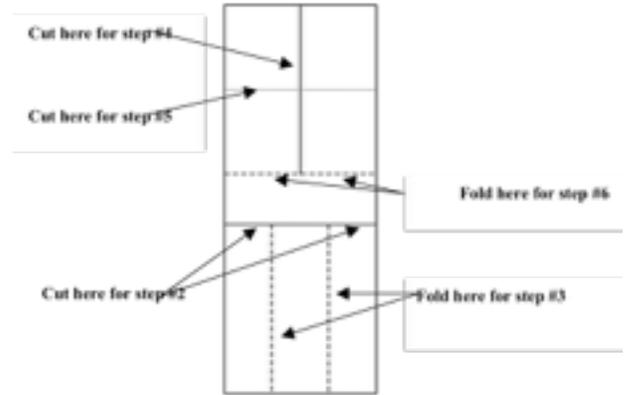


(see Activity-Based Statistics, “Gummi Bears in Space”)

### Helicopter Activity: Possible answers

- 1 & 2: **“Time-based:”** Helicopter fatigue, launcher-person fatigue, increased/decreased drop skill, increased/decreased measuring skill, AC/air turned on, etc.  
**“Lack of Control-based:”** Launch height, measuring technique, construction accuracy, drop technique, different sizes of targets, etc.  
**“Planned:”** rotor length, drop height
3. For “time-based,” we could have randomized the order of the launches using a die. Let “1” stand for “Long, Low,” “2” stand for “Short, Low,” etc. Then any “time-based” variation would be spread out randomly among all six treatment groups and essentially “cancel out,” leaving only the variation caused by the treatments. For “lack of control-based,” we could have built some sort of helicopter launcher that would give us the exact same launch position. We could have had strict quality control on the construction: One large paper clip, cut exactly on the lines, two staples, etc. We could have defined exactly how to measure the distance of the helicopter from the target. We could have given each group a dime to use as a target.
4. There were two factors, rotor length and drop height.
5. There were two levels of rotor length: short and long. There were three levels of drop height: low, medium and high.
6. There were six treatments: Long/Low, Short/Low, Long/Medium, Short/Medium, Long/High, Short/High.
7. There could have been measuring bias. Perhaps the person measuring the distances believes the higher launch heights are surely less accurate, so they tend to round those distances up?  
Perhaps the launcher-person is right handed and left-eyed, so they line up the drop position 2cm to the right of the target for each drop?  
Perhaps there is a draft from the air vent that is causing all drops to be “off” in a certain direction?





1. Cut out the rectangular shape of the helicopter on the solid lines.
2. Cut one-third of the way in from each side of the helicopter to the vertical dashed lines on the solid line.
3. Fold both sides toward the center creating the base. The base can be stapled at the top and bottom. Try to be consistent about where the staples are placed. Use a paper clip to add some weight to the body.
4. For long-rotor helicopters, cut down from the top along the solid center line to the horizontal dashed line.
5. For short-rotor helicopters, proceed as in step 4, but cut the rotors off along the horizontal line marked.
6. Fold the rotors in opposite directions.