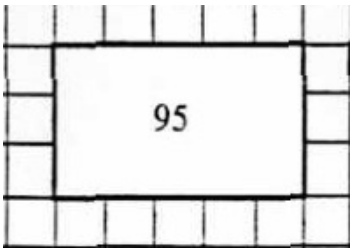
# 2.3 – Sampling Principles

***MDM4U***
***Jensen***

## Part 1: Random Rectangles Activity

1. a. Guess the average area of all rectangles on the page: **(guess)** _____

   b. Choose six rectangles (before you calculate any areas) that you think represent the entire population of rectangles well.




6 rectangles – subjective – "rectangle **expert**":

| rectangle number | area |
|---|---|
| _____ | _____ |
| _____ | _____ |
| _____ | _____ |
| _____ | _____ |
| _____ | _____ |
| _____ | _____ |
| average: | _____ |

2. a. After setting a new seed value on your calculator, use the randint function to choose six random rectangles for you.

6 rectangles – **random**:

| rectangle number | b. area |
|---|---|
| _____ | _____ |
| _____ | _____ |
| _____ | _____ |
| _____ | _____ |
| _____ | _____ |
| _____ | _____ |
| average: | _____ |

3. a. mean of sample averages:

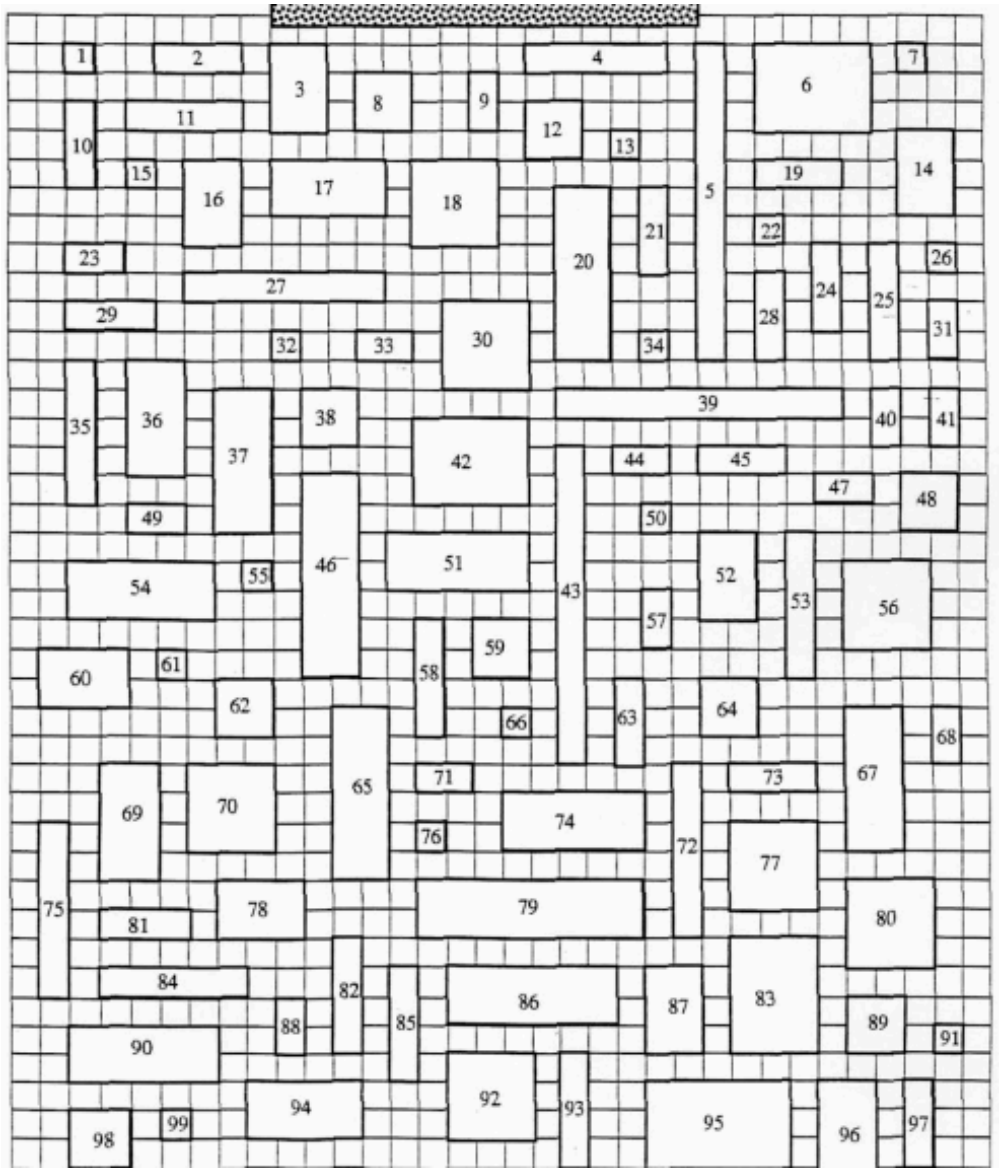
guesses _____

subjective (expert) _____

random _____

c. actual area of 100 rectangles (population): _____

Wrap-up (what have you learned?):



## Part 2: Random Sampling Methods

### 1. Simple Random Sampling

A sample is a _____ if it is selected so that:

- each member of the population is _____ likely to be chosen and the members of the sample are chosen independently of one other;

OR

- every set of n units has an _____ chance to be the sample actually selected.

**Example:** Put names in hat and draw until have desired sample size; more commonly, number names and use random number generator or other source of random numbers to select sample. Notice that some type of unbiased method must be used; haphazard ≠ random.

### 2. Systematic Random Sampling

A sample is a systematic random sample if you randomly choose some _____; then select every _____ element in the population, where $n$ is the sampling interval. This guarantees that the sample is taken from throughout the _____ but it requires an ordered list of everyone in the population.

**Example:** If we wanted to get a systematic random sample of 10% of the students from King's which has approximately 600 students…

- Calculate number of students required for sample: $600 \times 0.10 = 60$
- Calculate the sampling interval: $sampling\ interval = \frac{population\ size}{sample\ size} = \frac{600}{60} = 10$
- Choose a random starting point using a random number generator
- Include every $10^{th}$ student from the randomly chosen starting point in your sample

### 3. Stratified Random Sampling

When using a stratified random sample, the population is divided into _____called _____ (e.g. age, geographical areas, grade, etc.)

A _____ of the members of _____ stratum is then taken. The size of the sample for each stratum is _____ to the stratum's size (you must survey the same _____ of people from each stratum).

**Example:** If we want a stratified random sample of 10% of the 600 King's students, we can divide the population in to four groups based on grade (9, 10, 11, 12) and then take a simple random sample of 10% of the students in each grade.

**4. Cluster Random Sampling**

When using a cluster random sampling method, divide the population into _____ or _____; randomly select a few of those groups and then sample _____members from the selected groups.

**Example:** _____ select 5 block C classes—survey _____ students in each class selected.

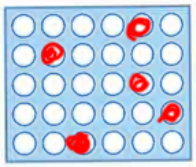
**5. Multi-Stage Random Sampling**

When using multi-stage random sampling, the population is organized in to groups, a simple random sample of groups is chosen, and then a simple random sample of people within the chosen groups is taken.

Example: _____ select 5 block C classes—survey _____ of the students in each class selected.

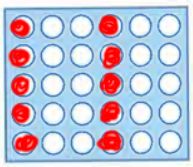
**Review of Different Random Sampling Techniques:**

***Simple Random***



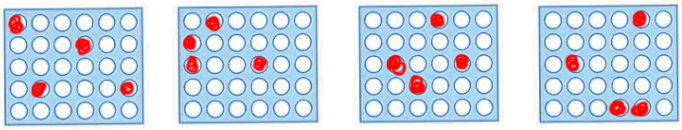
- all selections are equally likely

***Systematic Random***



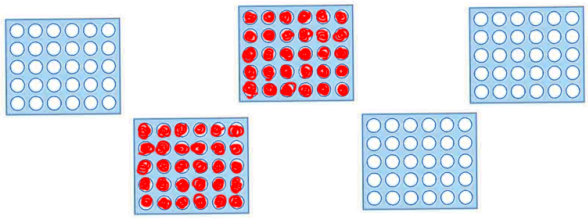
- random starting point, choose individuals at interval (every *n*th pe

***Stratified Random***



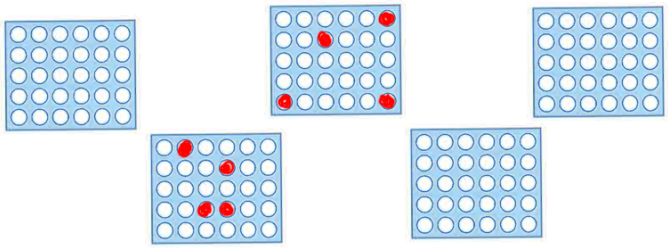
- Divide population into groups then survey an equal percentage of each group.

***Cluster Random***



- Divide the population into groups. Choose a random sample of groups and then survey **every member** of the groups chosen.

***Multi-Stage Random***



- Divide the population into groups. Choose a random sample of groups and then choose a random sample of members of the chosen groups.

## Part 3: Types of Non-Random Samples

### 1. Convenience sampling

The easiest way to obtain a sample is to choose it without any random mechanism (also called haphazard sampling). Choosing individuals from the population who are _____ results in a convenience sample. Convenience sampling often produces _____ data.

**Example:** Suppose we want to know how long students at a large high school spent doing homework last week. We might go to the school library and ask the first 30 students we see about their homework time.

### 2. Voluntary Response Sampling

A voluntary response sample consists of people who _____by responding to a general _____. Voluntary response samples attract people who feel strongly about an issue, and who often share the same opinion. This leads to _____.

**Example:** A radio host invites listeners to call in to give opinions on a new band.

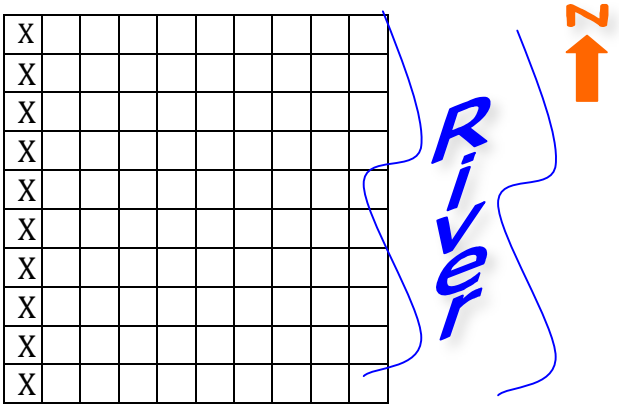
## Part 4: River Activity

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.

Part I.

A. Method Number One: Convenience Sample

The farmer began by choosing 10 plots that would be easy to harvest. They are marked on the grid below:

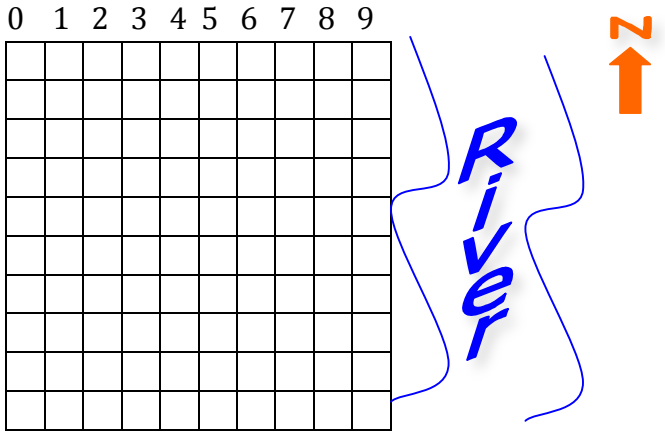


Since then, the farmer has had second thoughts about this selection and has decided to come to you (knowing that you are an AP statistics student, somewhat knowledgeable, but far cheaper than a professional statistician) to determine the approximate yield of the field.

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.

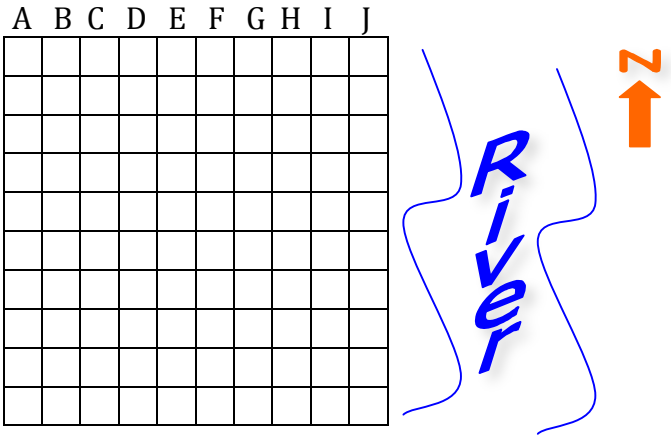
## B. Method Number Two: Simple Random Sample

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.



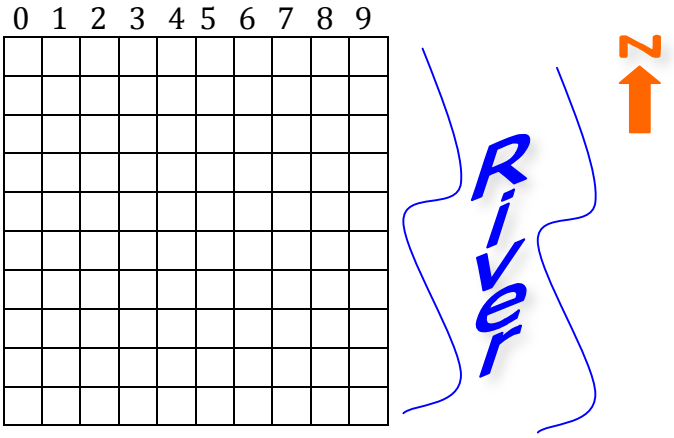

## C. Method Number Three: Stratified Sample

You and the farmer think the river might have a strong influence on corn production so you decide to consider the field as grouped in vertical columns (called strata—remember you can only stratify data your sample when you think a factor will have a strong influence on the outcome.). Using your random number table, randomly choose one plot from each vertical column and mark on the grid. (Label your columns A through J, rows 0 through 9.)



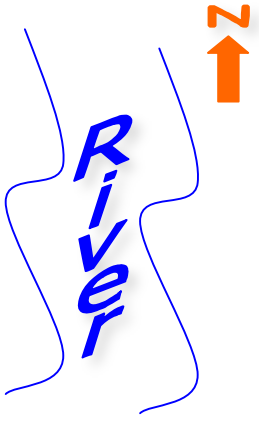
## D. Method Number Four: Stratified Sample

You and the farmer rethink the plan and decide that direction (north—south) may have a strong influence on corn production. You decide to consider the field as grouped in horizontal rows (also called strata). Using your random number table, randomly choose one plot from each horizontal row and mark them on the grid. (Label your rows A through J, columns 0 through 9.)




**OK, the crop is ready!** Below is a grid with the yield per plot. Estimate the average yield per plot based on each of the four sampling techniques.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 6 | 17 | 20 | 38 | 47 | 55 | 69 | 76 | 82 | 97 |
| 7 | 14 | 23 | 34 | 43 | 56 | 63 | 75 | 81 | 92 |
| 2 | 14 | 28 | 30 | 50 | 50 | 62 | 80 | 85 | 96 |
| 9 | 15 | 27 | 34 | 43 | 51 | 65 | 72 | 88 | 91 |
| 4 | 15 | 28 | 32 | 44 | 50 | 64 | 76 | 82 | 97 |
| 5 | 16 | 27 | 31 | 48 | 59 | 69 | 72 | 86 | 99 |
| 5 | 18 | 28 | 34 | 50 | 60 | 62 | 75 | 90 | 90 |
| 8 | 15 | 20 | 38 | 40 | 54 | 62 | 77 | 88 | 93 |
| 7 | 17 | 29 | 39 | 44 | 53 | 61 | 77 | 80 | 90 |
| 7 | 19 | 22 | 33 | 49 | 53 | 67 | 76 | 86 | 97 |




| 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)** Compare your results to someone else in the class. Were your results similar?

**4)** When we compare the class boxplots for each sampling method. What do you see?

**5)** Which sampling method should you use? Why do you think this method is best?

**6)** What was the actual yield of the farmer's field? How did the boxplots relate to this actual value?