

EXAMPLES on IMPLEMENTING SAMPLING

QUESTION 1 A party box of donuts consists of 15 different donuts – see below for picture (!). A donut maker has an order for 100 boxes for an office-party. **Read that again to observe the premise of the problem...**

Because of the large order and wishing to insure that there are no major complaints, before dispatch the donut maker chooses to take a random sample of donuts to examine their specifications [size, weight].

a) Before arranging the 15 different donuts into the boxes, naturally, they are baked and laid out separately. Describe the implementation of a stratified sampling design to select 75 donuts, focusing on the randomization. Write a general introduction explaining the process and what constitutes a stratum, followed by the implementation using index cards. **Tip!** Before you start answer these Qs: how many donuts are there in all? How many types are there? How many donuts of each type are there? How many strata are there? How many do we need from each stratum?

b) Describe the implementation of a cluster sampling design to select 75 donuts, focusing on the randomization. Write a general introduction explaining the process and what constitutes a cluster, followed by the implementation using index cards.

Tip! Before you start answer these Qs: what is a cluster? How many clusters are there? How many clusters need to be selected?

Solution.

a) Observe the detail! Since there are 1500 donuts of 15 types ~ 15 homogeneous strata each with 100 donuts, to select 75 donuts, we shall take a SRS of 5 from each stratum and pool them together. **For each variety**, number each donut from 1-100 [**Note:** we'd use 00-99 for a random digit table!] and write 1-100 on identical index cards and shuffle them. Take 5 cards randomly without replacement and choose the corresponding donuts. Repeat process for each of the 15 strata.

b) Observe the detail! Since there are 100 donuts in each box – each box comprising a heterogeneous group of donuts ~ a cluster, to obtain 75 donuts, we shall take a SRS of 5 clusters from the 100, and choose all donuts in them to constitute the sample. For this, number the boxes from 1-100, and [**Note:** we'd use 00-99 for a random digit table!] and write 1-100 on identical index cards and shuffle them. Take 5 cards randomly without replacement and select all the donuts from the corresponding boxes to get 75 donuts.

QUESTION 2 An organization wishes to determine if there are systematic

differences in the education levels of its 1000 salespeople based on the volume of business they generate. The company examines its records and classifies its personnel into 700 low-performing salespeople, 250 average performers and 50 top performers. It plans to select a sample of [total] 40 salespeople.

- a) Which sampling scheme should the company employ? Why? Write ~3 sentences explaining the reasoning behind your choice – **Tip!** Read the 1st line – and its benefits.
- b) How many salespeople should the company select in each category?
- c) Explain – in 1 sentence – how the company shall go about sampling [an introduction sentence describing the general **logic** of the process] and then describe the randomization process.

Solution.

- a) A Stratified Random sampling scheme is preferable since we suspect that salespeople performance is associated with their education levels **and** those in different performance categories i.e. strata, systematically differ in their education while those in the same categories i.e. strata, are relatively homogeneous. Via Stratification, we shall be able to determine if indeed there are systematic differences in education levels between salespeople in different categories as well as insure that no subset of the population – especially the smallest top-performing members – arent excluded even by chance. Finally, since the responses are likely homogeneous within the strata, the estimates would be more precise due to lower S.E.(estimate)
- b) The company should choose $40/1000*700 = 28$ from 700 low-performing salespeople, $40/1000*250 = 10$ amongst the 250 average performers and $40/1000*50 = 2$ from the 50 top performers.
- c) The company should choose 28 salespeople randomly from the 700 low-performing salespeople, 10 randomly amongst the 250 average performers and 2 randomly from the 50 top performers. For this number the 700 low-performers from 1-700. Write 1-700 on identical slips of paper, and choose 28 slips without replacement. Select the corresponding low-performers. Repeat the process for the average and top-performers.

QUESTION 3 A Guidance Counselor at a small newly opened school wishes to use a stratified sample to obtain information GPA of students and is wondering how to create her strata: on the basis of year [Freshman, Sophomore, Junior, Senior] or by Difficulty of Class [Regular, Honours, AP].

- a) Which of the 2 factors should she employ for her strata? Explain your choice in 2-3 sentences, in detail, lucidly.

Be thoughtful! Make sure you address why she should pick one and not the other.

Tip for the wise! We have discussed the choice of the stratification factor in class + in Solutions *repeatedly*.

b) The Counselor decides to stratify also by Gender. Describe in 1-2 sentences the implementation of the design. **Tip!** How many strata are there *then*?

c) What 3 advantages does a stratified sampling scheme have over an SRS? Write 2-3 sentences, in context, **comparing** the 2 designs.

Solution.

a) The Counselor should stratify by Difficulty of Class since it is reasonable to suspect that it is **associated** with GPA so that students in the same stratum would respond similarly / be relatively homogeneous re GPAs, while those in different strata / difficulty-levels would systematically differ in their responses i.e. GPAs. On the other hand [**this is a comparison Q!**], student in different Years are unlikely to be homogeneous in their GPAs **OR** Years is likely not associated with GPA, so shouldn't be chosen as the stratification factor.

b) For this, the counselor should divide the student population into homogeneous non-overlapping strata by Difficulty level of Class [Regular, Honours, AP] **and** Gender [M, F] i.e. form **six strata**, and then take an SRS of proportional size from each of the 6 strata.

c) **Observe the detail!**

1. Via stratification, the counselor can compare and draw inferences about the GPAs of students from the 6 strata, and determine if there are any systematic differences, something not possible using a SRS.

2. Further, especially since the school small newly opened school, stratification shall insure that none of the key subsets of the population is excluded even by chance, a possibility for an SRS!

3. Finally, through stratification, especially **if GPA is strongly associated with Gender and Difficulty level**, we can reduce the variability of the responses within each stratum thereby reducing the S.E.(estimate) and increasing the Power of the test, none of which is possible for a SRS.

QUESTION 4 A charter school has a total enrollment of 800 students, of whom 200 each are in the Freshman, Sophomore, Junior and Senior grade-levels. Their auditorium seats all 800 students.

In general, the 12th graders get priority seating and occupy seats in the front numbered from 1-200. Next come the Juniors in seats 201-400, then the Sophomores in 401-600, and finally, the Freshmen in seats 601-800. The student council wishes to conduct a survey in the 1st of an all-school assembly in the auditorium and seeks to take a sample of 80 students from the 800.

- a. Describe the implementation of a SRS scheme to select 80 students.
- b. Describe – in detail and clearly – the implementation of a Stratified Random Sampling scheme to select 80 students, especially the randomization process. [Write an introduction explaining the *logic* or *reasoning*, followed by the implementation.]
Things to ALWAYS ponder for a Stratified Sample: How many Strata are there? How many individuals are there in each stratum? How many students are, ergo, needed from each stratum?
- c. Describe – in detail and clearly – the implementation of a Cluster Random Sampling scheme to select 80 students, especially the randomization process. [Write an introduction explaining the *logic* or *reasoning*, followed by the implementation.]
Things to ALWAYS ponder for a Cluster Sample: How many clusters are there? How many students are in each cluster? How many clusters are, ergo, needed?

Solution.

- a. Do this yourselves!
- b. Assuming that 4 grade levels are homogeneous in their responses or characteristics to the question of interest, we shall choose 20 students from each of Freshmen, Sophomores, Juniors and Seniors. For this, number the Freshman from 1-200, and write 1-200 on identical slips and shuffle. Choose 20 slips without replacement and select the corresponding students. Repeat the process for Sophomores, Juniors and Seniors and combine to form the stratified sample.
- c. Since each *column* is heterogeneous consisting of Freshmen, Sophomores, Juniors and Seniors, and there are 20 columns ~ 20 clusters each with 40 students = 800 students in all, and we need 80 students, we need to randomly select 2 clusters out of 20. For this, number the columns from 1-20. Write 1-20 on identical slips of paper, shuffle and draw 2 slip at random and select *all students* in the corresponding clusters.

QUESTION 5 It is estimated that that about 340 shoppers shop at a grocery store.

- a) The store wants to perform a systematic random sample of 20 shoppers on the amount they spent. Describe – in detail and clearly – how this would be implemented, especially the randomization process. [Write an introduction explaining the *logic* or *reasoning*, followed by the implementation.]
- b) In the situation described in *a*, one of the managers decides to perform a cluster sample of 20 shoppers out of the 340 shoppers. For this, he suggests that out of the 8 and $\frac{1}{2}$ hours that the store is open, each half-hour period be regarded as a “cluster” of customers. Describe -- in detail and clearly -- how this would be implemented, especially the randomization process. Write an introduction explaining

the *logic* or *reasoning*, followed by the implementation. **Things to ALWAYS ponder for a Cluster Sample:** How many clusters are there? How many shoppers are in each cluster? How many clusters are, ergo, needed?

Solution.

a. Since $k = 340/20 = 17$, informally form 20 groups of 17 shoppers to select 1 shopper from each group. For this, choose a 2-digit random number from 01-17. Choose the corresponding shopper and every 17th shopper beyond as they exit the store! **OR** Write 1-17 on identical slips of paper, shuffle and draw 1 slip at random and select the corresponding shopper in the morning...and every 17th shopper beyond.

b. Since there are 17 half-hour periods ~ 17 clusters, each with ~20 shoppers, we shall randomly select 1 cluster out of the 17. For this, write 1-17 on identical slips of paper, shuffle and draw 1 slip at random and select *all* customers in the corresponding time period / cluster.

QUESTION 6 Suppose a classroom has desks arranged in 8 rows [horizontal] and 4 columns [vertical] for its 32 students – **sketch a rough figure to assist you!**

No, seriously, you scoundrel! – and the students sit by order of height, from the 1st row [shortest] to the last [tallest]. ← Understand what's going on, carefully! There's a cluster and a strata in there...if you don't illustrate the problem, you will be lost...

a) Describe an SRS of 8 shall be taken from the 32 students.

b) If 8 students are to be chosen at random via a cluster sample, describe the process, in general [**1 introductory sentence explaining logic**] and elaborate upon the randomization process [**2 sentences**] – in detail – *using slips of paper*. **Tip!** Which are the clusters ~ heterogeneous groups? The Rows or the Columns?

c) If 8 students are to be chosen at random via a stratified sample, describe the process, in general [**1 introductory sentence explaining logic**] and elaborate upon the randomization process [**2 sentences**] – in detail – *using slips of paper*. **Tip!** Which are the strata ~ homogeneous groups? The Rows or the Columns?

d) Suppose a systematic sample of 8 students is to be chosen. describe the process, in general [**1 introductory sentence explaining logic**] and elaborate upon the randomization process [**2 sentences**] – in detail – *using slips of paper*.

Solution.

Note: if there are 8 rows, then each should have 4 students = 32. If there are 4 columns, then each should have 8 students = 32. **Think about it!**

a) Number the 32 students from 1-32. Write 1-32 on identical slips and shuffle them.

Choose 8 slips without replacement, and select the corresponding students to constitute the sample.

b) **INTRODUCTION** For a cluster sample, for 8 students, we shall randomly select 1 of the 4 columns – each with 8 students of different heights – since each column would constitute a cluster broadly reflecting the variability of heights in the class.

IMPLEMENTATION For this label the 4 columns, 1-4, and write 1-4 on identical slips of paper and shuffle them; choose 1 slip at random...[**OR** roll a tetrahedral die]...and select all students from the corresponding column.

c) **INTRODUCTION** For a stratified sample of 8 students, we shall randomly select 1 student from each of the 8 rows, since each row likely has students of similar heights.

IMPLEMENTATION For the 1st row, label the 4 students, 1-4, and write 1-4 on identical slips of paper and shuffle them; choose 1 slip at random...[**OR** roll a tetrahedral die]...and select the corresponding student. Repeat the process for each of the rows and pool the students to constitute the sample.

d) **INTRODUCTION** For a systematic sample, for 8 students, we shall form 8 lists or groups of $32/8 = 4$ students and select 1 from each group.

IMPLEMENTATION Label the students from 1-32 – **this is needed to identify the OTHER students!** – and for the 1st group only, write 1-4 on identical slips of paper, and shuffle them; choose 1 slip at random...[**OR** roll a tetrahedral die]...and select the corresponding student from the 1st group and every 4th student – **OR** choose the corresponding student in each list – beyond to constitute the sample!

QUESTION 7 In the Qs below, identify the type of sampling used.

a) To estimate the percentage of defects in a recent manufacturing batch, a quality-control manager at Intel selects every 8th chip that comes off the assembly line starting with the 3rd until she obtains a sampling of 140 chips.

b) To determine the average IQ of ninth-grade students, a school psychologist obtains a list of all high schools in the local public school system. She randomly selects five of these schools and administers an IQ test to all ninth-grade students at the selected schools.

c) To determine customer opinion of their boarding policy, Southwest Airlines randomly selects 60 flights during a certain week and surveys all passengers on the flights.

d) A member of Congress wishes to determine her constituency's opinion regarding estate taxes. She divides her constituency into three income classes: low-income households, middle-income households, and upper-income households. She then takes a simple random sample of households from each income class.

e) In an effort to identify if an advertising campaign has been effective, a marketing firm conducts a nationwide poll by randomly selecting individuals from a list of

known users of the product.

f) A radio station asks its listeners to call in their opinion regarding the use of U.S. forces in peacekeeping missions.

g) A farmer divides his orchard into 50 subsections, randomly selects 4, and samples all the trees within the 4 subsections to approximate the yield of his orchard.

h) A school official divides the student population into five classes: freshman, sophomore, junior, senior, and graduate student. The official takes a simple random sample from each class and asks the members' opinions regarding student services.

i) A survey regarding download times on a certain Web site is administered on the Internet by market research firm to anyone who would like to take it.

j) A group of lobbyists has a list of the 100 senators of the United States. To determine the Senate's position regarding farm subsidies, they decide to talk with every seventh senator on the list, starting with the third.

k) A small-town newspaper reporter wants to get local reaction to a controversial new film. She waits outside the theater during an afternoon show and, starting with the second, asks every fifth patron leaving how much they liked the movie.

l) To determine his DSL Internet connection speed, Shawn divides up the day into four parts: morning, midday, evening, and late night. He then measured his Internet connection speed at 5 randomly selected times during each part of the day.

m) A statistics instructor with a large number of students' attempts to reduce time spent grading by only grading a portion of assigned homework problems. He randomly selects one of the first four problems in the assignment and then grades that problem and every fourth problem thereafter.

n) 24 Hour Fitness wants to administer a satisfaction survey to its current members. Using their membership roster, the club randomly selects 40 club members and asks them about their level of satisfaction with the club.

Solution. a) To estimate the percentage of defects in a recent manufacturing batch, a quality-control manager at Intel selects every 8th chip that comes off the assembly line starting with the 3rd until she obtains a sampling of 140 chips.

Systematic

b) To determine the average IQ of ninth-grade students, a school psychologist obtains a list of all high schools in the local public school system. She randomly selects five of these schools and administers an IQ test to all ninth-grade students at the selected schools. **Cluster**

c) To determine customer opinion of their boarding policy, Southwest Airlines randomly selects 60 flights during a certain week and surveys all passengers on the flights. **Cluster**

d) A member of Congress wishes to determine her constituency's opinion regarding estate taxes. She divides her constituency into three income classes: low-income households, middle-income households, and upper-income households. She then

takes a simple random sample of households from each income class. **Stratified**

e) In an effort to identify if an advertising campaign has been effective, a marketing firm conducts a nationwide poll by randomly selecting individuals from a list of known users of the product. **SRS**

f) A radio station asks its listeners to call in their opinion regarding the use of U.S. forces in peacekeeping missions. **Voluntary Response**

g) A farmer divides his orchard into 50 subsections, randomly selects 4, and samples all the trees within the 4 subsections to approximate the yield of his orchard.

Cluster

h) A school official divides the student population into five classes: freshman, sophomore, junior, senior, and graduate student. The official takes a simple random sample from each class and asks the members' opinions regarding student services.

Stratified

i) A survey regarding download times on a certain Web site is administered on the Internet by market research firm to anyone who would like to take it. **Voluntary Response**

Response

j) A group of lobbyists has a list of the 100 senators of the United States. To determine the Senate's position regarding farm subsidies, they decide to talk with every seventh senator on the list, starting with the third. **Systematic**

k) A small-town newspaper reporter wants to get local reaction to a controversial new film. She waits outside the theater during an afternoon show and, starting with the second, asks every fifth patron leaving how much they liked the movie.

Systematic

l) To determine his DSL Internet connection speed, Shawn divides up the day into four parts: morning, midday, evening, and late night. He then measured his Internet connection speed at 5 randomly selected times during each part of the day.

Stratified

m) A statistics instructor with a large number of students' attempts to reduce time spent grading by only grading a portion of assigned homework problems. He randomly selects one of the first four problems in the assignment and then grades that problem and every fourth problem thereafter. **Systematic**

n) 24 Hour Fitness wants to administer a satisfaction survey to its current members. Using their membership roster, the club randomly selects 40 club members and asks them about their level of satisfaction with the club. **SRS**