UNIVERSITY OF WATERLOO

Faculty of Mathematics
Waterloo, Ontario N2L 3G1

Centre for Education in
Mathematics and Computing

# Grade 7/8 Math Circles

February 17, 2021

## *Random Sampling - Problem Set Solutions*

1. For each scenario, identify the population and sample.

   (a) A toothpaste company randomly selects 10 dentists from a province's registry of dentists to ask if they would recommend their company's toothpaste.

   *Solution* : The population is all dentists on the province's registry, and the sample is the 10 randomly selected dentists.

   (b) A food safety inspector randomly checks the temperature of 20 meals that exit the restaurant kitchen during the hour from 6–7pm to see if they have been cooked to a safe temperature.

   *Solution* : The population is all the meals that exited that restaurant kitchen during the hour from 6–7pm, and the sample is the 20 randomly selected meals that the inspector checked during the hour from 6–7pm.

   (c) A fast food company randomly selects 10 of their own restaurants each month. They then interview all of the employees at that location about their work environment.

   *Solution* : The population is all of the employees at the company's restaurants, and the sample is all of the employees at the 10 randomly selected restaurants.

   (d) A hockey equipment manufacturer wants to learn about who is buying their equipment online, so with every tenth purchase, they ask the customer a few simple questions before checking out.

   *Solution* : The population is everyone who buys the manufacturer's equipment online, and the sample is every tenth purchaser.

2. For each scenario, identify the parameter and statistic.

   (a) During an election, journalists survey voters as they exit the polls to find that 40% of those people were voting for the Purple Party. However, on the news that

night, the final vote count declares that only 35% of voters voted for the Purple Party.

*Solution* : The parameter is 35%, and the statistic is 40%.

(b) A quality control inspector randomly selected 50 cans of soda from a factory production line and found that on average, they weighed 410 grams. However, while loading up a pallet for shipping, the factory worker noticed that they only weighed 404 grams on average.

*Solution* : The parameter is 404g, and the statistic is 410g.

(c) A party planner asked their client how many of the 500 people eating at their banquet were vegetarian. Since their guest list wasn't finalised, they told the party planner that 2 of their 10 special guests requested vegetarian meals. Based on this information, the party planner ordered 100 vegetarian meals and 400 non-vegetarian meals. However, at the end of the night, 75 guests informed the staff that they were disappointed that there wasn't a vegetarian meal prepared for them.

*Solution* : The parameter is $\frac{175}{500}$, and the statistic is $\frac{2}{10}$. The statistic was scaled by the total number of meals to get $\frac{2}{10} \times 500 = 100$, which is why the party planner was short 75 vegetarian meals.

3. For each scenario, identify what kind of random sampling method is being used.

(a) A new cereal brand divides a country up by postal code and selects 200 areas to test for their brand's popularity.

*Solution* : This is an example of cluster random sampling. The population is all consumers in the country, divided up by the company into clusters of consumers based on postal code. All consumers in each of the 200 areas is then part of the sample.

(b) For a live bingo game, balls with every possible number are placed in a tumbler to be scrambled up. At the beginning of each turn, the caller randomly chooses a ball and calls out the number.

*Solution* : This is an example of simple random sampling. Each numbered ball has an equal chance of being selected. Every turn, a sample size of one ball is chosen from the tumbler.

(c) Each week, one employee at a restaurant is randomly selected to be surveyed about their working environment.

*Solution* : This is an example of simple random sampling. Each employee has an equal chance of being selected. Each week, a sample size of one employee is chosen.

(d) A group of researchers wants to learn about the impact of climate change on farmers. They randomly select farmers from each state to survey proportionally based on how much land there is in the state, so that more farmers are surveyed in bigger states, and less in smaller states.

*Solution* : This is an example of stratified random sampling. The group of researchers has defined individual states as strata, and calculated how many farmers from each state should be included in the sample based on land area.

(e) A farmer wanted to check on their corn crops, so they divided their land up with a grid and randomly chose 5 areas to go to inspect their plants.

*Solution* : This is an example of cluster sampling. The population is all of the farmer's corn crops, divided up by the farmer into clusters based on sections of their land.

4. What are the strata being used? Calculate how many individual units from each of the strata should be included in a sample of size 200.

| age | # in population | # in sample |
|---|---|---|
| 11 | 90 | ? |
| 12 | 270 | ? |
| 13 | 480 | ? |
| 14 | 360 | ? |
| total | 1200 | 200 |

*Solution* : There are four strata defined by age.

| age | # in population | proportion in population | # in sample |
|---|---|---|---|
| 11 | 90 | $\frac{90}{1200} = \frac{3}{40}$ | $\frac{3}{40} \times 200 = 15$ |
| 12 | 270 | $\frac{270}{1200} = \frac{9}{40}$ | $\frac{9}{40} \times 200 = 45$ |
| 13 | 480 | $\frac{480}{1200} = \frac{2}{5}$ | $\frac{2}{5} \times 200 = 80$ |
| 14 | 360 | $\frac{360}{1200} = \frac{3}{10}$ | $\frac{3}{10} \times 200 = 60$ |
| total | 1200 | $\frac{1200}{1200} = 1$ | $1 \times 200 = 200$ |

5. What are the strata being used? If the number of individual units from each of the strata is included in the sample as indicated, how many individuals from each strata are in the population?

| shoe size | # in sample | # in proportion |
|---|---|---|
| 5 | 2 | ? |
| 6 | 6 | ? |
| 7 | 19 | ? |
| 8 | 13 | ? |
| 9 | 8 | ? |
| total | 48 | 1008 |

*Solution* : There are five strata defined by shoe size.

| age | # in sample | proportion in sample | # in population |
|---|---|---|---|
| 5 | 2 | $\frac{2}{48} = \frac{1}{24}$ | $\frac{1}{24} \times 1008 = 42$ |
| 6 | 6 | $\frac{6}{48} = \frac{1}{8}$ | $\frac{1}{8} \times 1008 = 126$ |
| 7 | 19 | $\frac{19}{48} = \frac{19}{48}$ | $\frac{19}{48} \times 1008 = 399$ |
| 8 | 13 | $\frac{13}{48} = \frac{13}{48}$ | $\frac{13}{48} \times 1008 = 273$ |
| 9 | 8 | $\frac{8}{48} = \frac{1}{6}$ | $\frac{1}{6} \times 1008 = 168$ |
| total | 1008 | $\frac{48}{48} = 1$ | $1 \times 1008 = 1008$ |

6. You've been tasked to collect a sample of soil from some farmland to test for nutrients and give the land a score based on those tests. A map of the farm has been given to you, with the land already divided into a 6×6 gride of 36 squares. To generate a score for the land, you test soil from the centre of 6 of the grid squares, and the sum of those 6 scores is the overall score for the land.

   By using the random number generator linked here: *https://www.calculatorsoup.com/calculators/statistics/random-number-generator.php*, find a score for the land using a sample created using:

   (a) simple random sampling, by selecting 6 squares to test.

   *Solution* : Answers may vary.

   Six numbers generated from 1–36: 5, 9, 33, 30, 26, 20

Score $= 4 + 8 + 7 + 4 + 8 + 7 = 38$

(b) stratified sampling, using the rows as strata and selecting a square from each stratum.

*Solution* : Answers may vary.

| row | randomly generated # from 1–6 | score |
|---|---|---|
| 1 | 5 | 4 |
| 2 | 6 | 1 |
| 3 | 4 | 5 |
| 4 | 5 | 4 |
| 5 | 4 | 4 |
| 6 | 4 | 5 |
| total score | | 23 |

(c) stratified sampling, using the columns as strata and selecting a square from each stratum.

*Solution* : Answers may vary.

| column | randomly generated # from 1–6 | score |
|---|---|---|
| 1 | 1 | 9 |
| 2 | 6 | 6 |
| 3 | 5 | 5 |
| 4 | 4 | 6 |
| 5 | 2 | 3 |
| 6 | 4 | 2 |
| total score | | 31 |

(d) cluster sampling, using the rows as clusters and selecting one cluster to test.

*Solution* : Answers may vary.

Number generated from 1–6: 2

Score $= 8 + 6 + 8 + 5 + 3 + 1 = 31$

(e) cluster sampling, using the columns as clusters and selecting one cluster to test.

*Solution* : Answers may vary.

Number generated from 1–6: 5

Score $= 4 + 3 + 4 + 4 + 2 + 3 = 20$

A diagram of the entire farmland with the scores of each grid square is below. Again, it's important to remember that in real life, you wouldn't actually have all of this information about the population—only what you learn from your sample!

north

| | *1* | *2* | *3* | *4* | *5* | *6* |
|---|---|---|---|---|---|---|
| *1* | 9 | 8 | 6 | 4 | 4 | 2 |
| *2* | 8 | 6 | 8 | 5 | 3 | 1 |
| *3* | 9 | 7 | 6 | 5 | 4 | 1 |
| *4* | 9 | 7 | 5 | 6 | 4 | 2 |
| *5* | 8 | 8 | 5 | 4 | 2 | 4 |
| *6* | 8 | 6 | 7 | 5 | 3 | 3 |

west east

south

(f) What you have calculated are statistics using different sampling methods. Compare them to a parameter score of 32. What do you notice about the scores you generated with the different methods from (a) to (e)? Are some of them closer or further away from 32? (We only know the parameter because we created the question!)

*Solution* : The scores are all different! The scores generated from (c) and (d) were quite close to 32. However, the scores from (a), (b), and (e) are not close to 32 at all.

(g) Remember the difference between when stratified sampling vs. cluster sampling works. (You can find the information at the end of the lesson.) After inspecting the diagram, do you notice anything that makes some of the methods more appropriate than others? Which methods make the most sense? Does this match up with your results from (f)?

*Solution* : We want strata to be "similar within, but different between," and we want clusters to be "different within, but similar between."

Notice that overall, as we move from west to east, the scores tend to get smaller, and each column has scores that are similar to each other.

In Column 1, we have scores from 8–9; in Column 2, we have scores from 6–8; in Column 3, 5–7; in Column 4, 4–6; in Column 5, 2–4; and in Column 6, 1–4. Thus, the columns would make suitable **strata**, and that's why (c) gave us a fairly accurate statistic! That's also why (e) gave us a very inaccurate one—each column is similar within, but different between, which makes for very unrepresentative clusters.

Each row appears to have a fairly representative group of scores, unlike the columns. They each contain one square from each column, which gives us scores that range from high to low, and the rows are fairly similar to each other. Each row is different within, but like between, making them perfect for **clusters**! This lines up with our findings: (d) gave us a very accurate statistic, while (b) was way off.

7. To figure out the best strategy to sell strawberries, you decide to estimate the mean weight of strawberries in your field by taking a random sample. What could be an appropriate way to create a random sample here? Why? What are potential problems to your method? Can you think of more methods?

   *Solution* : Answers may vary. Refer to Question 6 for some ideas!

   One way would be to use cluster sampling. We could do this by dividing up the field into rows, then randomly selecting a row to pick strawberries from.

   Ideally, the rows are different within, but like between, making them reasonably representative of the entire field. That way, cluster sampling makes sense, and we can simplify the physical process of picking our sample by only having to walk through one row to pick strawberries!

   A potential issue would be if the rows aren't representative of the field—this could happen, for instance, if some of the rows get less water each day. If this were the case, stratified sampling by rows or cluster sampling by columns would make more sense.