

14. Statistics

Designing an effective survey

Your computer science class has made an app for your school. Some of the features of the app include:

- maps for getting around the site
- information about the teachers
- results from the school sports teams' recent games
- access to your homework platform
- a calendar that shows your class timetable

You are on the sub-team that is responsible for setting up a group of student testers to provide feedback on the app, and for collecting and analysing the feedback. You are keen to use the right sampling techniques so that the feedback is representative of the broader population of future users of the app.

Task one: Sampling

1. You consider taking a systematic sample of 40 people from your total school population of 840. What sampling interval would you use?

Sampling interval = population / desired sample size
 $840 / 40 = 21$

2. You consider standing outside the canteen and inviting the first 40 students who walk past you to join your tester group. Explain why this sampling method may not be representative of the whole school population.

Any sensible answer is acceptable. For example:

- Year groups might have different lunchtimes, so the first 40 people may all be from one year.
- People who have packed lunches and don't use the canteen may be excluded.
- This method may include groups of friends that have similar views to one another.

3. You want to ensure you have a fair representation of different age groups in your group of testers. You consider inviting a certain number of students per year to join the tester group based on how large each year is in proportion to the total school population. What is the name for this sampling method?

Stratified sampling as the population has been divided into smaller sub-groups (strata) based on shared characteristics (year group).

4. Another option would be to assign all students a number then have a computer programme randomly generate numbers within that range. What is the name of this sampling method and what are the pros and cons of choosing this option?

Random sampling. Any sensible answers acceptable, such as:

Pros:

- As testers would be chosen at random, there would be less chance of bias.
- Simple sampling method so would be quick to select the participants.

Cons:

- The sample may not be representative of the school population.
- More participants may be required under this sample method in order to try to get a representative sample size.

5. You decide to take a stratified sample of 40 students from the following school population data. How many students from each year should you invite to join your tester group?

Year group	Number of students
7	197
8	204
9	189
10	112
11	138

Year group	Number of students	Workings	Sample size from each year
7	197	$40 \times (197/480) = 9.4$	9
8	204	$40 \times (204/480) = 9.7$	10
9	189	$40 \times (189/480) = 9$	9
10	112	$40 \times (112/480) = 5.3$	5
11	138	$40 \times (138/480) = 6.6$	7
Total	840		40

Task two: Questionnaire design

Below is an extract from the draft questionnaire that will be shared with the tester group. Suggest improved questions and possible answers, and write down your criticisms of the current survey.

Questionnaire	
Question	Answer options
How often did you use the app over the week?	Very often
	Occasionally
	Not very often
	Never
What did you think of the timetable feature?	Very good
	Good
	Average
How much time did you spend on the app over the week?	10-15 minutes
	15-20 minutes
	20-30 minutes
	30-60 minutes

Any sensible criticisms and suggested improvements are acceptable.

Question: How often did you use the app over the week?

Comment: As people may interpret answers such as 'very often' or 'occasionally' differently, the question may not be answered consistently.

Improvement: How many times did you use the app in the last week? 12 times or more; 8-11 times; 4-7 times; 1-3 times; or I did not use the app.

Question: What did you think of the timetable feature?

Comment: There was no option to give a negative answer and no option to provide feedback on what worked well or not so well with this feature was gathered.

Improvement: This question could split into two open-ended questions with free text answers, such as 'what did you like about the timetable feature' and 'what could be improved with the timetable feature.'

Question: How much time did you spend on the app over the week?

Comment: There are no options for less than 10 minutes or more than one hour. The time intervals for each option are inconsistent and overlap.

Improvement: Change the answers to: 0-15 minutes; 16-30 minutes; 31-45 minutes; and 46 minutes or more.

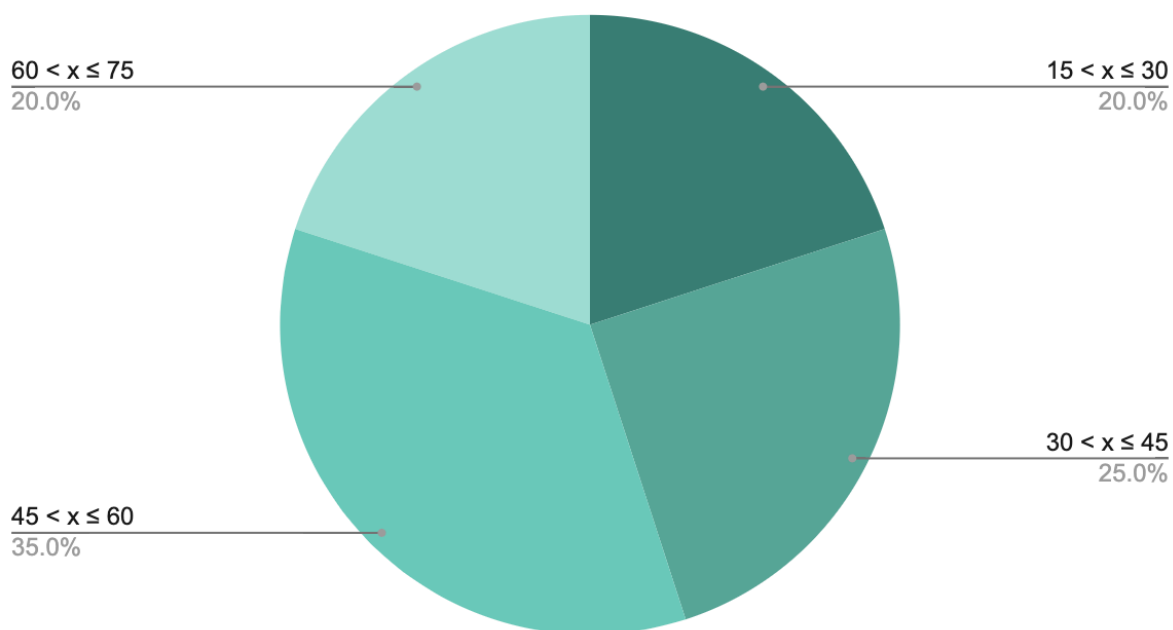
Task three: Analysing the results

- The below frequency table shows how long each member of the test group spent on the app over the course of a week. Select an appropriate way to visually present this data and explain why you have selected this.

Time spent on the app (minutes)	Number of students
$0 < x \leq 15$	0
$15 < x \leq 30$	8
$30 < x \leq 45$	10
$45 < x \leq 60$	14
$60 < x \leq 75$	8

Possible visualisations include a bar chart, pie chart or frequency polygon. Example:

Time spent on the app (minutes)



I selected a pie chart because it gives the audience of the data a general sense of how long people spent on the app. It also easily allows the audience to see the answers relative to each other. This makes it easy for the audience to see the most and least popular responses for example.

2. You produce a summary of your findings to share with the class and make the following statements. Comment on the accuracy of these statements:

a. 55% of the school will spend more than 45 minutes on the app each week.

55% of the testers spent more than 45 minutes on the app. However, this does not mean that this will be true for the broader population. The testers were specifically invited to provide feedback, so it is reasonable to assume that they may have spent longer exploring the app than an average user would.

b. The school is 49% male and 51% female overall. Our tester group was 60% female but we do not believe that this would have impacted our results.

The tester group was not representative of the gender split of the school population. It is not possible to state that this had no impact on survey results. Females and males may have had differing views on the app which may have skewed the survey results.

c. Overall, the school felt that the app was very good and would continue to use it.

As only a sample of students were included in the tester group, it is not possible to conclude that the whole school liked the app. It would be better to conclude: 'Overall, the tester group felt that the app was very good and would continue to use it. We are therefore confident that this will be well received by the broader school population.'