ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Biased sampling

Activity Summary:

This activity introduces people to the concept of sampling – the process of collecting observations to analyse – and when it can give the 'wrong' answer

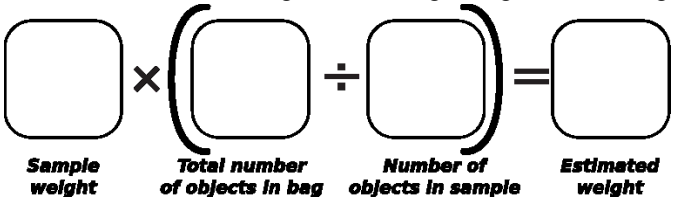
Activity Learning Outcomes:	Suggested Resources:
• Understand sampling variability • Understand random sampling and when it may give an 'incorrect' answer • Understand a representative sample	• 20 small and 5 large objects (e.g. 20 small sweets, 5 full size sweets) • Weighing scales • Cloth bag • Calculator (image shows oversized calculator)

How to run the activity:

- Prior to each participant's engagement with the activity, place all 25 objects in the cloth bag. (Ensure the participants do not see the mix of small/large objects)

- Explain that we want to calculate the weight of all objects in the bag. (Do not mention there are different sized objects in the bag)
- It is known that there are 25 objects in the bag. (It is important that this is fixed and known)
- Explain that the obvious solution is to weigh all 25 objects. However, in many real-life situations it is not possible to measure every object – it may cost a lot of money to take measurements or it might take too much time. Hence is it often not feasible to measure every object.
- Instead, we can select some objects into a **sample**, measure the weight of the sample, and then estimate the weight of all the objects.

- The activity:
  a) Get the participant to draw five (5) objects from the cloth bag and weigh them
  b) Calculate an estimate of the weight of the bag using the following formula



  c) Put the objects back in the bag and repeat by drawing another five (5) objects. For groups, get each participant to play the game. (Do not repeat too many times, max. 3?)
  d) Finally, measure the true weight of all objects and compare with the estimates.

Exploring the activity:

- Apart from checking against the true total weight, can we tell if the sample estimate of the total weight is 'wrong'? Does the sampling variability help?
- Repeat the activity drawing samples of three (3), seven (7) and ten (10) objects.

What's going on?

- The act of drawing objects is called **sampling**. This creates a **sample** which we use to calculate an estimate.
- Repeating the game several times will show that we can get **different** samples, each with a different estimate of the weight. This highlights the concept of **sampling variability** – which estimate is right? They all are, they are estimates based on each sample.
- The closest estimates will come from **representative samples**, samples that represent the numbers of small/large objects in the bag. In real-life studies, it is important to have representative samples otherwise it is not clear if any wider conclusions.

- Unknown to the participant the bag contains small and large objects in unequal numbers. It is likely that participants will draw multiple large objects from the bag. In a sample of five objects, we should have one large and four small
- There is an unconscious bias towards picking the larger objects (they take up more space in the bag and feel important). This creates a **biased sample** since not all objects are equally likely to be drawn from the bag.
- Even if we could perform a **random sample** – where each object was equally likely to be drawn, it is highly likely that a given sample of size five will not be representative, and so will give a 'wrong' estimate.

Video demonstration:

A video demonstrating this activity is available on the RSS website at www.rss.org.uk/hands-on

Risk assessment:

There are no risks associated with this activity
(If using sweets, as shown in the example photograph, the sweets are not consumed so there are no food-related risks. If in doubt use alternatives, for example rocks/pebbles, large/small building blocks, etc.)

Additional information and taking it further:

Wikipedia explanation of sampling: https://en.wikipedia.org/wiki/Sampling_(statistics)

Credits:

Dr Simon R White (Medical Research Council Biostatistics Unit, University of Cambridge).

Version P1 Published 2017-08-31