ROYAL STATISTICAL SOCIETY

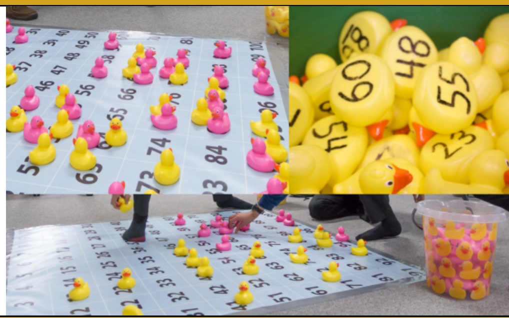
DATA | EVIDENCE | DECISIONS

Capture Recapture

Activity Summary:

This activity illustrates how to estimate the size of a population.

It explores a core idea of statistical inference and is very interactive.



Activity Learning Outcomes:

- Understand the words “estimate”, “sample” and “population”.
- Understand how the population size is estimated.
- Understand capture-recapture studies.

Suggested Resources:

- 50-60 uniquely identifiable objects (ideally numbered or labelled).
- Opaque container for objects.
- (optional) mat to place objects on.
- (optional) second set of objects, see below.
- Alternatively, a deck of playing cards (52 unique objects, no container needed).

How to run the activity:

Setup:

- Place objects in the container. If using a deck of cards, remove 1-5 cards beforehand.

Introduction:

- Explain that we want to know the total number of objects. We could simply count them all, but this is not viable for any large or difficult to reach **population,** such as fish in a river or birds. Define the technical term **population** as the whole group of interest. The objects are the **population** of interest.
- Get the participant to guess the size of the population. They have no real information to inform their guess. If using playing cards, be clear you have some playing cards (rather than a whole deck of playing cards) to avoid guesses of 52.
- Explain the participant is going to take a **(random) sample**, which is a random subset of the objects.

Activity (one set of objects in one container):

- Get the participant to take a **random sample** of objects from the container and record their labels (see example recording sheet). Take out a sample of 15-20 (out of the 50-60) objects.
- After this first sample, ask the participant again to estimate the size of the population. Although they may have a better idea, having drawn objects from the container, they really only know that the population has at least A objects (where A is the size of the first sample).
- Return the objects to the container and explain we will take a second sample. We call this a **capture-recapture study**. During the second sample, the **recapture**, record which objects are sampled. Some objects will not have been seen before, some will be seen for a second time. The second sample size

(B) should again be between 15-20 objects (but does not need to equal the size of the first capture).

- For the final time, ask the participant what is the population size? They should deduce that the population must be at least $a+b+c$ in size, where c is the number of objects in common, a and b are the number of objects only seen in the first ($A=a+c$) and second ($B=b+c$) sample respectively.
- A statistical **estimate** is a numerical characteristic of interest obtained from the sample, in this case the size of the population. An estimate of the population size is $(AB)/c$.
- Finally, reveal the true population size and compare to the participant's estimate. Discuss that we do not expect the estimate to exactly equal the true population size.

Activity (two sets of distinguishable objects in two containers, plus mat):

- The only difference with the above description is that the capture and recapture can be conducted at the same time (note, the two containers must be identical – have the same objects labelled the same way, but visually different).
- A mat (see example photograph) including spaces for objects as they are drawn becomes a visual representation of the capture and recapture. The counts a, b, and c can be seen from which objects are on the mat.

Exploring the activity:

- Repeat the activity several times (keeping the sample sizes the same) to explore the variability of the estimate.
- Do we tend to over or under estimate the population size?
- What happens if the samples are too small? What if $c=0$?
- We assume capturing has no effect on being recaptured, is this reasonable? When might it be invalid? For example, animals if captured may be more difficult to recapture.

What's going on?

- Let the population size be N, so that there are d unseen objects, i.e. $N=a+b+c+d$.
- Assuming two random samples and that objects are sampled with the same probability (equal catchability assumption), then the proportion of the population in the first sample, A/N, that will be marked will equal the proportion of the second sample that is recaptured (marked), c/B.

$$A/N = (a+c)/N = c/(b+c) = c/B \quad \rightarrow \quad N = (a+c)(b+c)/c = AB/c = ab/c + (a + b + c)$$

Video demonstration:

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

Risk assessment:

Depending on the size of the objects and mat (if used), be mindful of slip hazards and choke hazards.

Additional information and taking it further:

Lincoln–Petersen method: https://en.wikipedia.org/wiki/Mark_and_recapture

Credits:

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

Version 4.0 Published 2018/September/05