ROYAL
STATISTICAL
SOCIETY

DATA | EVIDENCE | DECISIONS

How many penguins?

Activity Summary:

This activity introduces people to the concepts of populations and samples.

Using information obtained from a sample of "local" penguins, we infer how many penguins live in a colony in Antarctica.

Activity Learning Outcomes:

- Understand what the words "population" and "sample" mean.
- Be able to use an estimate obtained from a sample to infer back to the population.
- Appreciate the usefulness of this inference in real life applications.

Suggested Resources:

- Aerial photograph of penguin guano on snow with appropriate scale (available from website)
- "Local" penguins – toys/images that can be lifted and replaced with ease
- Calculator (the bigger the better!)

How to run the activity:

- Prior to the event, decide on the annual amount of guano (poo) produced per "local" penguin either via a sticker on the base of each penguin, or on a sheet of paper which the penguins sit on (as in the photograph on page 2).

- Begin the activity by asking the participant if they can count how many penguins can be seen in the aerial photograph.
- Once the participant has established that it is impossible to count individual penguins in the photograph, help them to estimate the surface area of the guano using the provided scale (or specify it for younger participants).

- Encourage participants to think about where they have seen penguins more conveniently than in Antarctica (zoo, aquarium etc.).
- Introduce the participant to the "local" penguins and explain that an estimate of the amount of guano produced per year per penguin can be found underneath each one.
- Then ask them to select 3 penguins from the sample, and lift them up to reveal the estimate.
- Help the participant to calculate the mean amount of guano produced by the sampled "local" penguins using the formula:

$$\text{Mean amount of guano} = \frac{\text{Penguin 1 guano} + \text{Penguin 2 guano} + \text{Penguin 3 guano}}{3}$$

- Now help the participant to infer how many penguins are in the aerial photo, and hence in that colony within Antarctica using the formula:

$$\text{Number of penguins} = \frac{\text{Surface area of guano in the photo}}{\text{Mean amount of guano}}$$

- What happens if the participant selects a different sample of "local" penguins?
- What happens if a larger sample of "local" penguins is used?
- How certain are we in our count of penguins in the aerial photo?
- What happens if the sample of "local" penguins is not random?

What's going on?

- A **population** is a theoretical concept used to describe the entire group. The population of interest in this activity is all the penguins producing the guano patch visible in the aerial photograph. The secondary population of interest is all of the "local" penguins.
- A **sample** is taken from a population to provide estimates of population parameters. The sample of interest in this activity is the random sample of "local" penguins selected by the participant. Estimates obtained from the "local" penguins can be used to infer information about the population of penguins in the aerial photograph.
- Different random samples will lead to different estimates of the mean guano produced. These will vary around the true mean of the population. The larger the sample, the smaller this variability will be.
- This activity can be extended to investigate the effect of non-random sampling – the "local" penguins can be arranged such that those with green hats (in this example) produce smaller amounts of guano per year than those with red hats for example. In this case, taking non-random samples of just those penguins with green hats would lead to a smaller estimate of the number of penguins in the photograph. This might be equivalent to sampling Rockhopper penguins when those in the photograph are known to be Emperor penguins for example.

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.

Additional information and taking it further:

StatTrek guide to samples and populations: http://stattrek.com/sampling/populations-and-samples.aspx

Significance article describing an alternative way to estimate the penguin population: The Emperor strikes back, Peter Fretwell. 12 June 2012 (free access)

Credits:

Idea & penguin photograph by Laura Bonnett (University of Liverpool), inspired by a *Significance* article by Peter Fretwell.

Aerial photograph courtesy of DigitalGlobe