

# An Active Tutorial on Distance Sampling

Alice Richardson  
University of Canberra

*Journal of Statistics Education* Volume 15, Number 1 (2007),  
[www.amstat.org/publications/jse/v15n1/richardson.html](http://www.amstat.org/publications/jse/v15n1/richardson.html)

Copyright © 2007 by Alice Richardson all rights reserved. This text may be freely shared among individuals, but it may not be republished in any medium without express written consent from the authors and advance notification of the editor.

---

**Key Words:** Estimation; Proportions; Sampling distribution; Statistical education.

## Abstract

The technique of distance sampling is widely used to monitor biological populations. This paper documents an in-class activity to introduce students to the concepts and the mechanics of distance sampling in a simple situation that is relevant to their own experiences. Preparation details are described. Variations and extensions to the activity are also suggested.

## 1. Introduction

Distance sampling is a widely used technique for estimating the size of a population. To use distance sampling to estimate the size of a population of objects in an area, the area of interest is first measured and sketched. A line (also known as a transect) that crosses the area is chosen at random. Observers then travel the length of the transect. The distance to objects visible from the transect is recorded and an estimate of population size is constructed. [Buckland, Anderson, Burnham and Laake \(1993\)](#) set out the state of the art in theory and applications of distance sampling. The statistical literature from then contains a steady stream of papers on the technique, including applications ranging from estimating the size of schools of fish ([Chen and Cowling \(2001\)](#)) to under-enumeration in censuses ([Welsh \(2002\)](#)).

Distance sampling can be seen as a more complex version of transect sampling, where only objects seen on the transect line are counted. The main advantage of distance sampling over the simpler process of transect sampling is that it allows data from beyond the selected line itself to still be used in the estimation process. Distance sampling is also frequently preferred over random sampling because it is easier to implement in rough terrain. The disadvantages of distance sampling are associated with the assumptions required to make it work. These centre on the distribution and visibility of objects in the area and the behaviour of the objects (particularly if they are animals that move around or travel in groups.)

Bishop (1998) produced a tutorial comparing transect sampling to random sampling using an aerial photograph. In this paper, we describe an activity that extends data collection to beyond the transect i.e. carries out distance sampling, and effort is spent on physically observing objects and measuring distances. [Otto and Pollock \(1990\)](#) carried out a similar experiment using beer cans, which we used as

the inspiration for developing this activity.

The activity follows the format of [Spurrier, Edwards and Thombs \(1995\)](#), whose activities cover a range of statistical concepts to a similar depth. The first time we used it, the activity followed a lecture by Dr. Ann Cowling on her experiences in applying distance sampling to fish populations in the Great Australian Bight. The lecture and activity appeared in a quantitative literacy course (see [Richardson \(2000\)](#)) as one of a range of statistical estimation techniques that included population size estimation using capture-recapture methods ([Scheaffer, Gnandesikan, Watkins and Witmer \(1996, p.126\)](#)), and estimating the maximum of a set of data ([Scheaffer, et al. \(1996, p.148\)](#)).

Distance sampling is not often presented in introductory statistics courses. We envisage this activity being used in courses where students learn about specialised methods of data collection. For example, the activity could be used in a survey methods course, as an unusual approach to the problem of census undercount. It would also fit in a general quantitative research methods course that covers a variety of survey and experimental designs for data collection. It could also be used in a more specific research methods for biologists course that focuses on the techniques used in biological research, such as quadrat sampling, capture-recapture methods and point transects. Finally, we have successfully included the activity in a very general quantitative literacy course. With the references provided within this paper, the activity is intended to be comprehensible without the need for supporting lectures.

## 2. Teaching Goals

The primary goal of this activity is to introduce students to the distance sampling method of estimation of population size in a biological context. Students also explore sources of variability in the method, in particular variability between observers and variability between transects.

A secondary teaching goal is to bring students face-to-face with the issues involved with collecting data including the appropriate degree of accuracy of recording measurements, the time taken to collect data, and working effectively in groups.

## 3. Logistics

To assist others in running this activity we have prepared details about the materials required and also guidelines for tutors or teaching assistants. They should be read in conjunction with the student notes shown in the [Appendix](#) at the end of this paper.

### 3.1 Materials

After much thought about an appropriate object to use in this activity, we purchased and used wooden beads approximately 2.5 cm (1 inch) long. These are quite visible enough in mown grass over a restricted area. For example, on a 12 metre long transect students spotted over 80 beads up to 6 m away (there were 120 beads spread out in an area 19 m by 14 m.) Other objects that could be collected over a period of weeks prior to their use in this activity include the plastic tags on bags of bread, or corks from wine bottles.

Care is required in placing the beads in the field. Truly random allocation, perhaps by using a map of the area to be used, would take a very long time to arrange. Moving through the field and throwing handfuls of objects is easier but the objects tend to land in clusters. This makes data collection more of a burden and violates a key assumption of the distance sampling method, e.g. independence. We therefore recommend shifting individual beads after an initial scattering, so that not too many beads are too close to each other. A discussion of this method, and whether or not it is truly random, can form part of the discussion towards the end of the activity.

The amount of space allocated to the activity compared to the number of students involved is also a consideration. A large enough area should be chosen so as to avoid crowding. We found that with 15 to 20 students, an area approximately 20 metres by 20 metres works well.

At the end of the activity, students are asked to fan out across the field and slowly move across it in an attempt to pick up all the beads (a strategy known as a emu parade in Australia). In our experience about 10% of the beads were still lost. Hence it is important to have a ready and cheap supply of the objects used in this activity. It is also important to use objects that will not be a nuisance in the grass later, if any do get left behind.

### 3.2 Guidelines for Tutors

The activity is designed to last two hours with data collection along three transects taking about an hour.

It is important to spend some time at the start explaining what is involved in the activity, as students are often keen to leap into data collection without any clear picture of what they are doing and why they are doing it.

The main goal of this activity is to introduce students to the method of distance sampling. Students will also study sources of variability associated with the method, particularly variability between observers and variability between transects.

As background reading for this activity, tutors could be given copies of [Otto and Pollock \(1990\)](#) and [Welsh \(2002\)](#).

In the [Introduction](#), tutors are expected to lead a discussion on methods for estimating population size. As an initial idea tutors can suggest taking a census, followed by the idea of taking a census in a representative small area and multiplying that value up to become an estimate of the total population size.

The likely loss of beads during the clean-up of the activity also gives tutors an opportunity to discuss the accuracy of a census. The lost beads can be regarded as an example of undercount in a census.

In the [Background](#), tutors are expected to lead a discussion on possible methods for the selection of transects. Tutors can suggest firstly, that a ruler can be dropped at random on a map of the field. That placement can then be transferred to the field. This also shows that transects need not be parallel to one side of the field. Secondly, a numbered grid could be drawn on a map of the field and random numbers could be used to locate a starting point on a random side of the field and an ending point on another. These random starting and ending points are then transferred to the field. An optional point that tutors can explain is that the existence of several methods of choosing random transects is related to the problem of choosing random chords on a circle. Tutors may also be interested to note that this problem is a form of Bertrand's paradox: see [Holbrook and Kim \(2000\)](#).

On a practical level, it is useful to have string attached to pegs (pencils are a suitable substitute for pegs). These can be used to mark out an area approximately 20 m by 20 m for the activity to take place in, and they can also be hammered into the ground at the beginning and end of a transect.

When we ran this activity, the students decided to peg out one transect for each group of three and leave them fixed, rather than peg out random transects each time. In a class of 15 – 20 students, this usually results in between five and seven transects. Three of these transects were then selected and used by all the other groups in their data collection. The advantages of this selection of fixed transects is that (a) it speeds up the data collection process; (b) it eases crowding in the area of data collection; and (c) it is easy to comment on between-observer variability because each transect is travelled by at least one group.

## 4. Student Reaction

In 2001, twenty-two University of Canberra students completed an evaluation of this activity. The activity was rated out of 5, with 1 equating to poor and 5 equating to excellent. The median rating was 4, and only one student rated the activity at 2. In 2003, only five students were available to complete the evaluation and so the results are not presented here. It is difficult to measure the long-term effect of this activity on student learning because there is no formal assessment at the end of the course. Instead, each activity is assessed by means of a report completed on the day the activity is run. In 2005, a group of ten students gained a mean score of 75% in the short answer questions attached to this activity. We take this as partial evidence that the activity does have some immediate impact upon students' understanding of the process of distance sampling.

Anecdotal evidence suggests that students generally enjoy the chance to get out of the classroom, generate some data of their own and then analyse it using statistical software. Some students have commented adversely on the amount of time taken up by data collection. We counter this by pointing out that an introduction to the nature of primary data collection is one of the teaching goals of the activity. Alternatively, a shorter version of the activity is suggested in the next section.

## 5. Extension

A time saving variation on this activity would involve pegging out two lines, one metre on each side of each transect line. Each student could then walk the transect and count all the objects within one metre of the transect. This would also allow for students to contribute values individually for each transect, rather than in groups, and compute their own estimates of population size. The formula for population estimation still requires a value, to the nearest metre, for the distance to the furthest visible bead, so this information needs to be collected as well.

Another variation that would also save time involves spreading the work of walking along transects among groups as well as only recording data within 1 metre of certain transects. For example, if there are six groups, two groups could collect full data from each of the three transects. Then the groups that did transect 1 could split up and one group go to transect 2 and the other to transect 3 to simply count the number of objects within 1 metre. The same applies to the other two transects. This would result in two complete data collections and two data collections within 1 m for each transect, which still allows for variability between groups and transects and data collection methods to be assessed.

Based on the assumptions that the beads are randomly distributed in the field and that every bead in the field is seen, a histogram of distances to beads should be uniform, with every bar the same height as the first one. The probability of seeing a bead at further distances from the transect line is then the ratio of the height of each bar of the histogram to the height of the first bar. For students in advanced classes, these ratios can be used to carry out a logistic regression of the probability of seeing a bead on the distance of the bead from the transect line. However these students would need to undertake the full data collection along transects 2 and 3 in order to carry out such an analysis.

[Bishop \(1998\)](#) describes extensions associated with the use of parallel transects that could be used here as well.

---

## Appendix: Student Notes

The original version of the student notes gave detailed instructions on how to use SPSS to construct the graphs the students needed. These instructions have been omitted here, and instructions for any statistical computing package could be inserted. Tables 2 and 3 would be identical to [Table 1](#) except that

their headings refer to the second and third transects, so only [Table 1](#) has been included here.

## A.1 Introduction

Naturalists often want estimates of population sizes that are difficult to measure directly. For instance, how would you estimate the number of weeds in a field, or birds' nests in a forest? Just going out and wandering around the field or forest, counting all the animals or plants that you see, is too haphazard an approach to have any useful statistical properties. The purpose of this activity is to introduce ways of measuring population size indirectly. Your tutor will start this session by leading a discussion of possible methods for measuring population size. Tutors: see the [Guidelines](#).

### A.1.1 Materials Needed

For each group of three, a trundle wheel or 10m tapemeasure and string to mark transects.

### A.1.2 The Setting

You are a biologist, and your “animal” of interest is the wooden bead. You wish to know how many wooden beads live in the field near your classroom.

### A.1.3 Background

This activity will focus on the method of population size estimation known as distance sampling. Distance sampling has been used by scientists in a wide range of fields to estimate, for example, the number of weeds in a field, the number of birds' nests on an island and the number of schools of fish in the Great Australian Bight.

To use distance sampling to estimate the size of a population of objects in an area, the area of interest is first measured and sketched. A line (also known as a transect) that crosses the area is chosen at random. Your tutor will lead a discussion of how a transect could be chosen at random, and the implications of non-random selection of transects. Tutors: see the [Guidelines](#). Observers travel the length of a transect recording distances to objects visible from the transect. An estimate of population size is constructed based on the distances. We will now examine this method in the activity.

## A.2 The Activity

### A.2.1 Collecting the Data

Go outside to a field indicated by your tutor. Measure the area of the field in metres - we will call this area  $A$ . Write  $A$  here:  $A =$

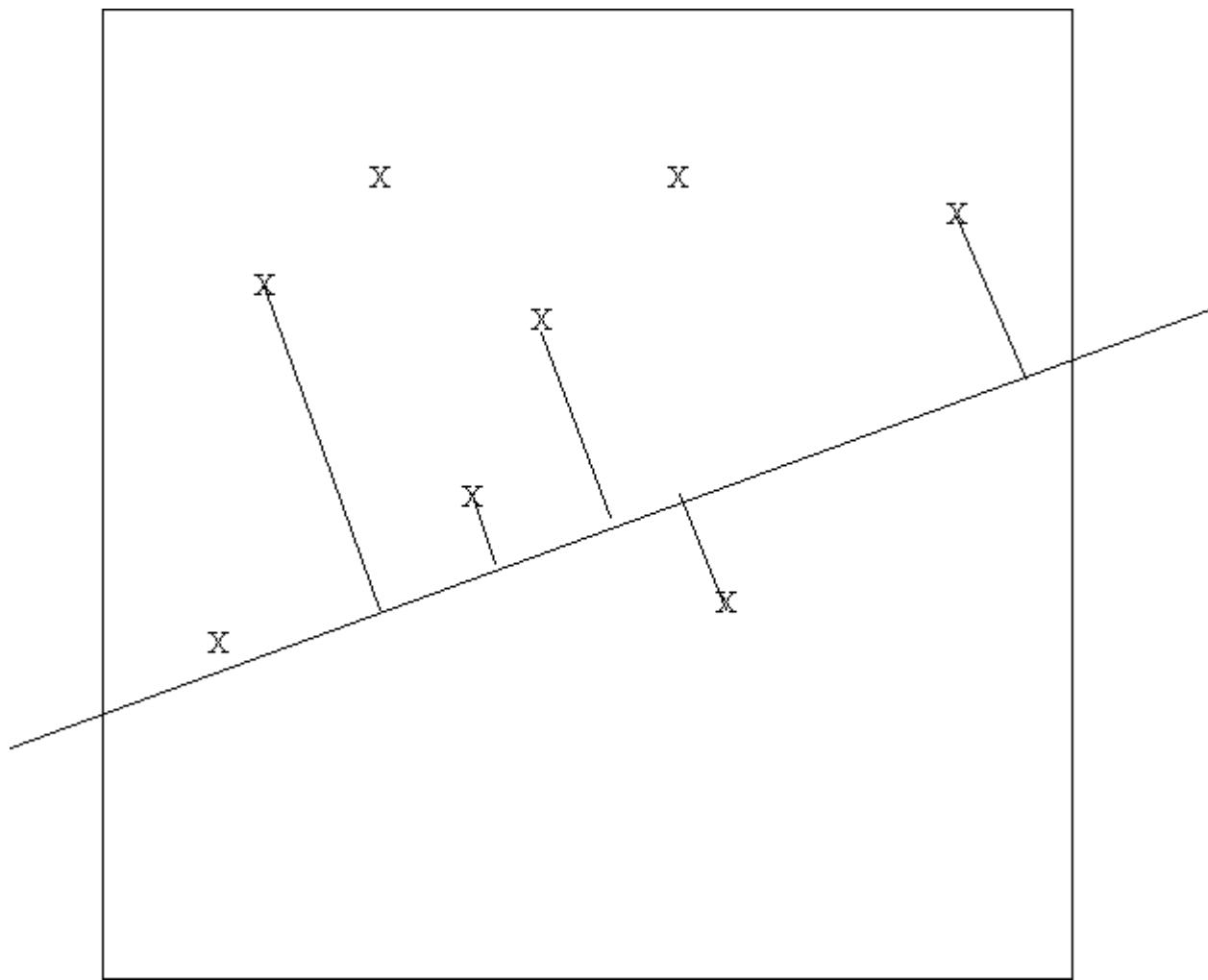
Each group of three should peg out a randomly selected transect. Number the transects as you will need to be able to tell them apart later on. Next, each member of the group of three has a specific task in the data collection along the transect: one (the “eyes”) will spot beads, one (the “ruler”) will measure distances and the third (the “writer”) will record the data. The “eyes” should walk along the transect indicated. Whenever he or she sees a bead on either side, he or she should stop and the “ruler” should measure how far the bead is from the transect. The “writer” records the distance measured by the “ruler”. The “ruler” should measure a perpendicular distance i.e. the distance, in metres, from the transect to the beads when the “eyes” are at right angles to the bead. It is not the job in this activity of the “writer” or the “ruler” to spot beads, even though they may see ones that the “eyes” miss. Record your measurements in [Table 1](#). If you need more room, continue on a blank sheet of paper.

Table 1. Distances, in metres, to object recorded along first transect.

Object No.	Distance						
1		21		41		61	
2		22		42		62	
3		23		43		63	
4		24		44		64	
5		25		45		65	
6		26		46		66	
7		27		47		67	
8		28		48		68	
9		29		49		69	
10		30		50		70	
11		31		51		71	
12		32		52		72	
13		33		53		73	
14		34		54		74	
15		35		55		75	
16		36		56		76	
17		37		57		77	
18		38		58		78	
19		39		59		79	
20		30		60		80	

Choose two other transects pegged out by two other groups. Write down which transects you have selected at the top of Tables 2 and 3. If your tutor says you have time, walk along these transects as before, this time recording your measurements in Tables 2 and 3. Otherwise, simply record the number of objects seen within 1 m of the transects.

An example of a transect and distances to observed objects is shown in Figure 1.



[Figure 1](#)

Figure 1. Example transect and distances to observed objects

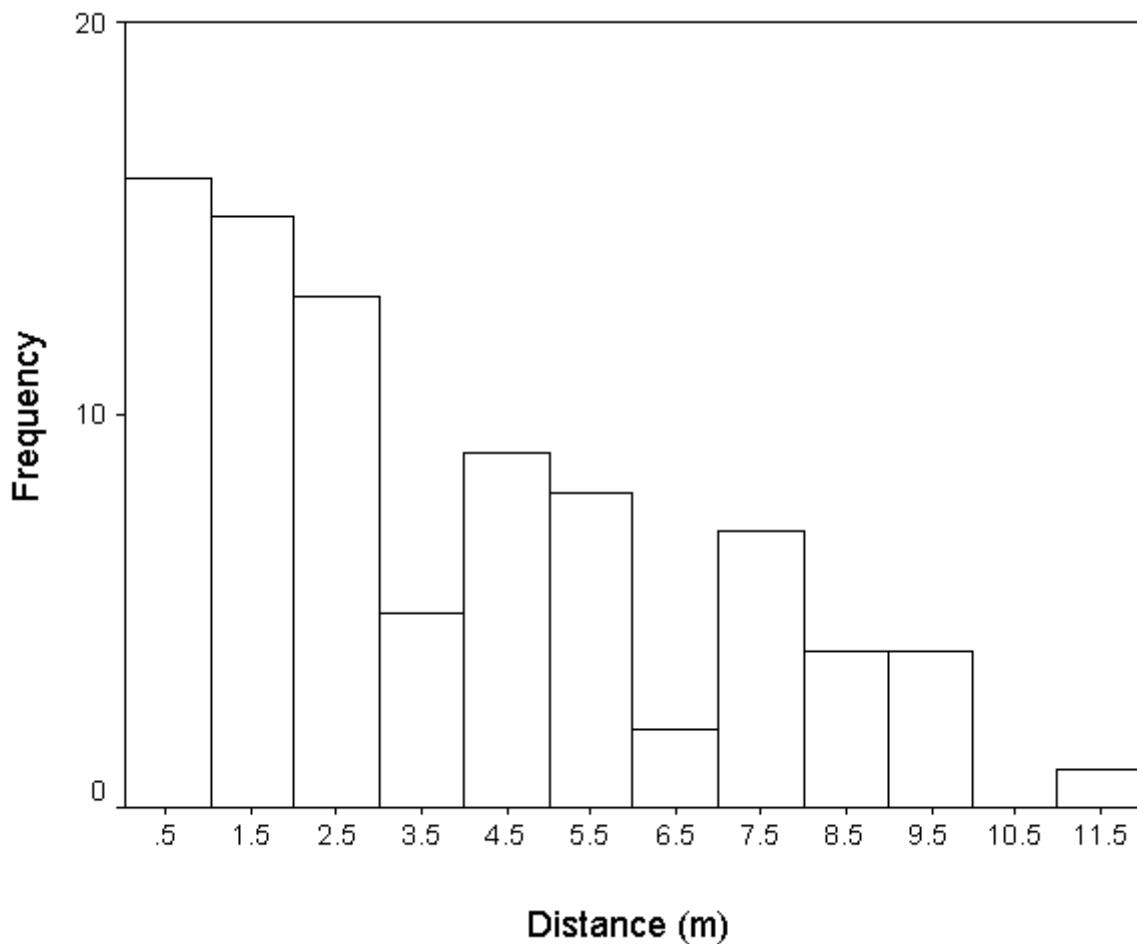
---

### A.2.2 Data Analysis: Single Transect, Different Observers

Enter the distances recorded from the first transect into a single column a statistical computing package data table. We will begin our data analysis by looking at a histogram of these distances, with classes or bins of width 1 m. This means that the frequencies on the vertical axis can also be interpreted directly as frequencies per metre. If you choose a narrower bin, the first bar of the histogram may not be the tallest, particularly if you happen to miss beads that are “under your feet”. If this occurs, widen the bins until the first bar is the tallest.

What shape does the histogram follow? A typical histogram of distances is shown in [Figure 2](#). Notice how the frequency of observations reduces as the distance to the beads increases. If the beads are distributed at random in the field, then this drop-off is not because there are fewer beads at greater distances from the line; it is because the beads are harder to see at those distances.

---



[Figure 2](#)

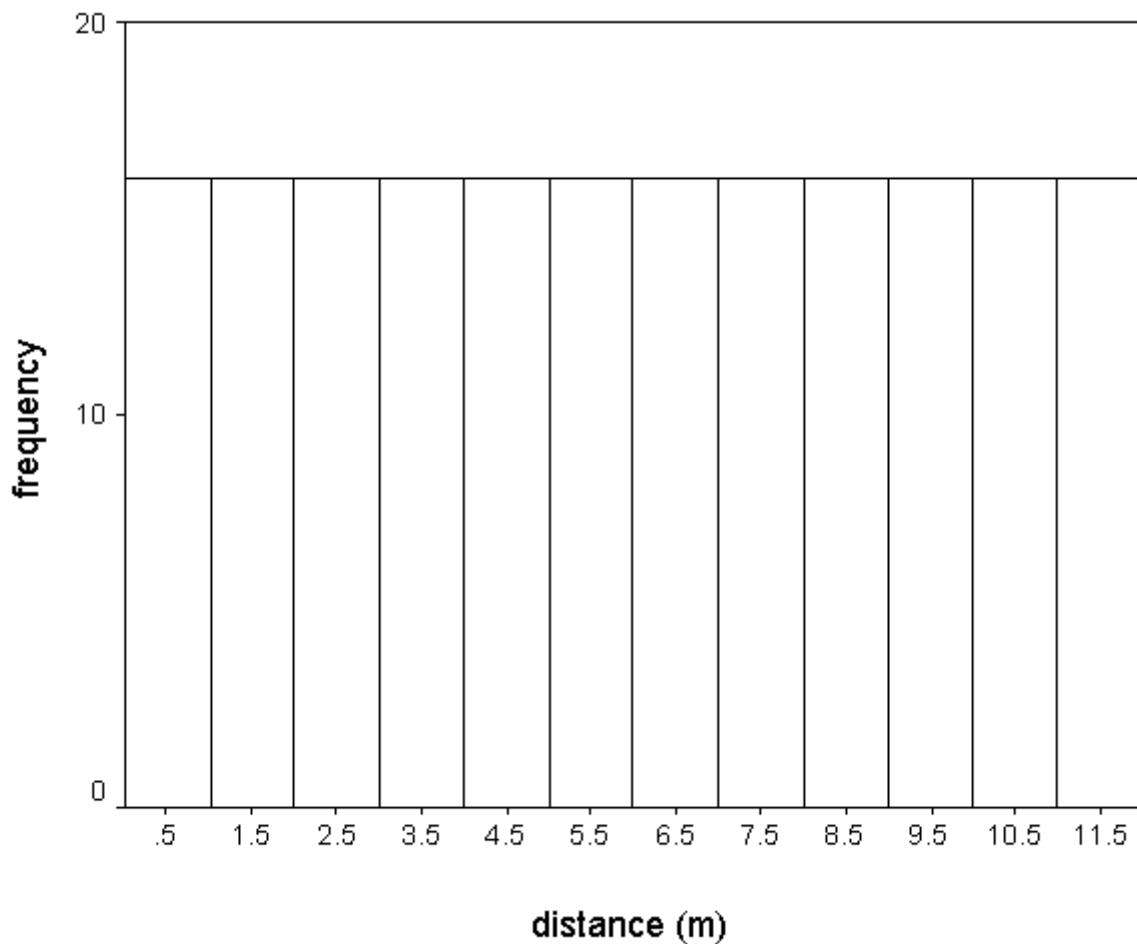
Figure 2. Histogram of distance to objects

---

There is no reason to expect that there are fewer objects far away from the observer compared to close by. Thus the skewed appearance of the histogram indicates that the further away an object is, the less likely an observer is to see it.

Assuming that the beads really are distributed at random in the field, then there should be the same number of them within 1 m of the transect line, between 1 and 2 m of the transect line and so on. So if you, the observer had seen every bead up to, say 12 m away, theoretically the histogram above should look like [Figure 3](#).

---



[Figure 3](#)

Figure 3. Histogram of expected distance to objects

---

You can now use the results of your distance sampling to estimate the population size in the following way.

1. Count up the number of objects you saw i.e. the number of objects represented in the histogram in [Figure 2](#). Call this number  $n$ .
2. Because of the assumption of a uniform distribution of objects, the number of unseen objects is the difference in the number of objects represented in the histogram in [Figure 3](#), and the number of objects represented in the histogram in [Figure 2](#). Count up the number of unseen objects and call this number  $u$ .
3. Call the length of the transect line  $L$  and the largest upper class limit in the histogram  $w$ .
4. The estimated density of objects i.e. the estimated number of objects per unit area is

$$\text{density estimate} = \frac{\text{number of objects (seen and unseen)}}{\text{area}} = \frac{n+u}{2wL}.$$

5. To estimate the number of objects in the whole area, multiply the density by the total area to obtain

$$\text{population size estimate} = \text{density estimate} * \text{total area}.$$

Apply the formula above to your value of  $w$ ,  $L$  and  $n$ , and arrive at an estimate of the population size. In the example,  $L = 20$ ,  $w = 12$  and  $n = 84$ , and 16 beads were seen within 1 m of the transect line. Therefore we estimate that there were  $(12 * 16) - 84 = 192 - 84 = 108$  unseen beads. Thus

$$\text{density estimate} = \frac{84 + 108}{2 \times 12 \times 20} = 0.4 \text{ beads per square metre.}$$

Finally, since in the example the total area is 20 m long and 20 m wide with a total area of 400 m<sup>2</sup>,

$$\text{population size estimate} = 0.4 * 400 = 160 \text{ beads.}$$

If you had a short time for this activity and only recorded the number of beads seen within 1 m of certain transects, this value is used to compute  $n + u$  as follows. If you saw 19 beads within 1 m of the transect line, we estimate that there were  $12 * 19 = 228$  beads (seen and unseen). Thus

$$\text{density estimate} = \frac{228}{2 \times 12 \times 20} = 0.475 \text{ beads per square metre}$$

and

$$\text{population size estimate} = 0.475 * 400 = 190 \text{ beads.}$$

We will now compare the population size estimates from the whole class. Your tutor will ask each group to read out their transect numbers, and the corresponding estimates of population size. Record the transects numbers and estimates in a spreadsheet. Calculate descriptive statistics for these estimates separately for each transect. The amount of variation in any one of these sets of estimates reflects the variation between observers. Your tutor (who laid out the beads) will also now reveal the true population size.

### A.2.3 Data Analysis: Several Transects, Single Observer

Enter the data from Tables 2 and 3 into the statistical analysis data table. Produce two separate histograms of the distances in these two columns. Estimate the number of objects in the whole area, using the same formula as before. Calculate the range and standard deviation of the estimates of population size you obtained from your three randomly selected transects. The amount of variation in these three estimates reflects the variation between transects.

## A.3 Conclusion

Distance sampling is not the only way to estimate a population size that it is difficult to obtain directly.

You could conduct a census in a number of small areas within the study area - these small areas are known as quadrats. The population size is then estimated by scaling up the count in the quadrats, in a

similar fashion to what is done with the distance sampling estimate we calculated. Point sampling is also sometimes used in preference to distance sampling. An observer stands at a fixed point and measures the distance to all the objects he or she can see from that point. Capture-recapture methods are also used to estimate population size, where a number of objects are caught, tagged and released back into the population. A second capture is carried out, and the proportion of tagged objects in the second capture is used to estimate the total population size.

Another way to attack the distance sampling problem is to find a smooth curve e.g. half a normal distribution, that matches the shape of the histogram in Figure 2. Then use this function to estimate the probability of seeing objects at various distances from the line, and in turn use those probabilities to estimate the population size.

## A.4 Short Answer Writing Questions

1. What were your three estimates of the population size from each transect that you travelled? Show your work in each case. What was the true population size?
2. Calculate the mean, standard deviation and range of the estimates in question 1. How accurate are your estimates? How precise? The amount of variation in this set of estimates reflects the variation between transects. Is this small or large?
3. Select one of the transects. Write down the estimates of the population size from all the other groups who travelled that transect.
4. Calculate the mean, standard deviation and range of the estimates in question 3. How accurate are the class's estimates? How precise? The amount of variation in this set of estimates reflects the variation between observers. Is this small or large?
5. The estimation of population size on the basis of distances relies on the following assumptions.
  - a. Every object on the line is observed without fail.
  - b. Every object is detected at its initial location.
  - c. Measurements to the objects are exact.
  - d. Objects are distributed at random i.e. they do not form clusters.

Explain whether each assumption has been met in this activity.

6. Suppose a conservation group wishes to use distance sampling to estimate the number of whales of a particular species present in the Great Southern Ocean in spring. Explain whether the assumptions in question 5 are likely to be satisfied. What measures could you take to ensure that the assumptions are more likely to be satisfied?

---

## Acknowledgments

The author would like to thank Ann Cowling for recounting her experiences with distance sampling of southern blue-fin tuna in the Great Australian Bight; Alan Welsh for the general method and reference to Otto and Pollock; and Glenys Bishop for general discussions on in-class activities. She would also like to thank the students and tutors in the University of Canberra course The World of Chance, who since 2001 have cheerfully taken part in the distance sampling activity and suggested refinements to it. Finally, she thanks the referees and editor for their careful reading of the paper and their suggestions that have improved it.

---

## References

Bishop, G. (1998), "A series of tutorials for teaching statistical concepts in an introductory course I. Sampling from an aerial photograph," *Journal of Statistics Education* [On line], 6(2).  
[www.amstat.org/publications/jse/v6n2/bishop.html](http://www.amstat.org/publications/jse/v6n2/bishop.html)

Buckland, S.T., Anderson, D.R., Burnham, K.P, and Laake, J.L. (1993), *Distance Sampling: Estimating Abundance of Biological Populations*, London: Chapman and Hall.

Chen, S.-X. and Cowling, A. (2001), "Measurement errors in line transect surveys where detectability varies with distance and size," *Biometrics*, 57, 732 – 742.

Holbrook, J. and Kim, S.S. (2000), "Bertrand's paradox revisited," *Mathematical Intelligencer*, 22, 16 – 19.

Otto, M.C. and Pollock, K.H. (1990), "Size bias in line transect sampling: a field test," *Biometrics*, 46, 239 – 245.

Scheaffer, R.L., Gnanadesikan, M., Watkins, A. and Witmer, J.A. (1996), *Activity-Based Statistics*, New York: Springer.

Spurrier, J.D., Edwards, D. and Thombs, L.A. (1995), *Elementary Statistics Laboratory Manual*, Belmont, CA: Duxbury Press.

Welsh, A.H. (2002), "Incomplete enumeration in sample surveys: whither distance sampling?" *Australian and New Zealand Journal of Statistics* 44, 13 – 22.

---

Alice Richardson  
School of Information Sciences and Engineering  
University of Canberra  
Canberra ACT 2601  
Australia  
[Alice.Richardson@canberra.edu.au](mailto:Alice.Richardson@canberra.edu.au)

---

[Volume 15 \(2007\)](#) | [Archive](#) | [Index](#) | [Data Archive](#) | [Information Service](#) | [Editorial Board](#) | [Guidelines for Authors](#) | [Guidelines for Data Contributors](#) | [Home Page](#) | [Contact JSE](#) | [ASA Publications](#)