Concepts behind distance sampling

It's rarely practical to count all the objects of interest in a large park or reserve. In practice, we **sample**, counting the objects in a few small areas (often called 'plots' or 'quadrats') and calculating the density (D) from the number of objects recorded (N) and the area (A) of the plots ($D = N/A$). If we have a large sample of randomly-located plots, we will get a good estimate of the average density over the whole area and we can calculate the precision of our estimate (e.g., confidence interval, standard error, or coefficient of variance).

For plants and objects such as nests or dung-piles, demarcating a plot and searching it carefully to count the objects in it works well. But it doesn't work for animals, which tend to flee as soon as you start searching. Transect walks work better.

Instead of randomly placing plots in the area of interest, randomly-placed lines or '**transects'** are used. You move along the transect, recording the animals detected either side. One approach is to decide how far from the line you can be certain of seeing all the animals which are there, and only record animals which are within that distance of the line. This is called a **'fixed-width transect'** or **'strip transect'** and is really just a long, thin plot. The problem with this is the width of the strip: if it's too wide, you will not detect all the animals in it, and your estimate will be too low; if it's too narrow, you will have a smaller sample for a given survey effort, and small samples mean less precise estimates.

An alternative is to make the strip very wide, too wide to be sure of detecting all the animals, but to estimate what proportion of animals we do detect. The key to this is the **distance** of the animals from the transect line. We assume that we see all the animals on or very close to the line, and that the proportion detected decreases further away from the line. This is the concept behind 'distance sampling'.

Muntjac in Batang Ai NP

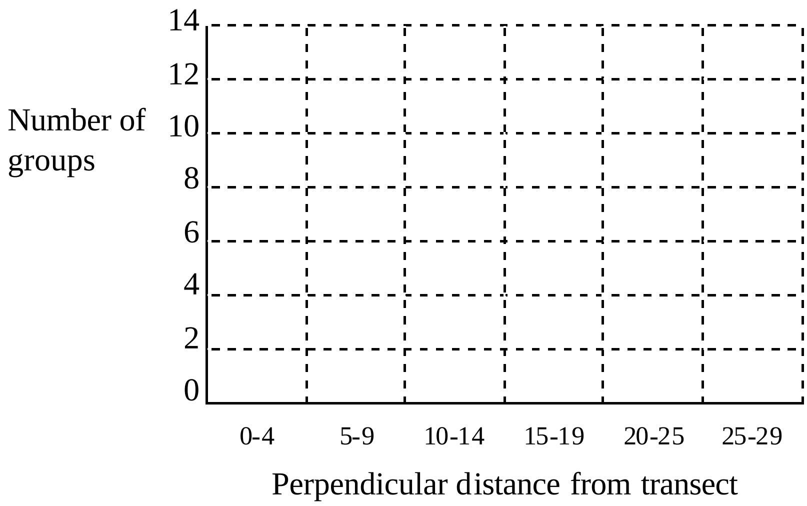
During transect surveys in Batang Ai National Park in Sarawak, Malaysia, in 1992, we saw 31 groups of muntjac (barking deer). The perpendicular distances from the transect line to the groups of animals when they were first spotted were:

5, 25, 4, 0, 0, 0, 2, 6, 4, 13, 8, 6, 5, 5, 8, 0, 15, 6, 20, 10, 4, 2, 6, 4, 8, 18, 6, 4, 1, 5, 5m

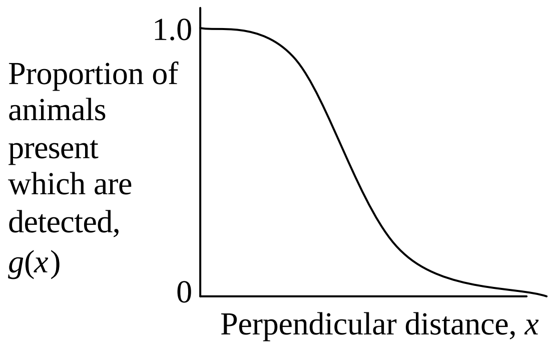
Now count the number of groups for different distances from the transect and fill in the table below:

Distance from rentis:	0-4m	5-9m	10-14m	15-19m	20-24m	25-30m
Number of groups:						

Then plot a histogram with the results:



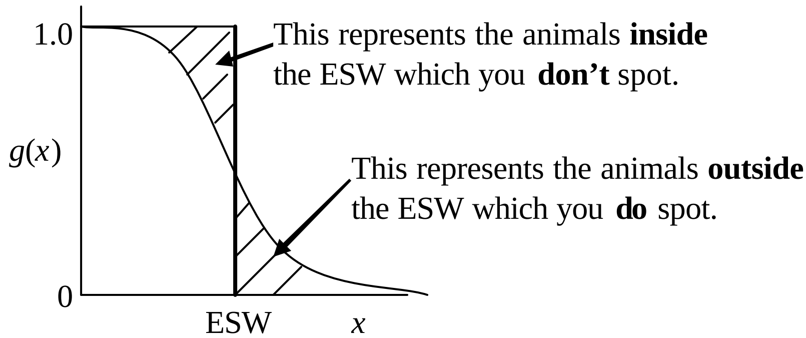
As you can see from the histogram, the number of animals we saw gets fewer farther from the transect. We assume that we see all the animals which are very close to the transect, and only a small proportion of those further away, as shown below, which should look rather like your histogram.



The curve in the figure is the '**detection function**', symbolized by $g(x)$. The crux of distance sampling is to find the equation for $g(x)$ which best fits the data. This involves fitting models to the data and finding the best model using likelihood and AIC; the DISTANCE software package does this for us.

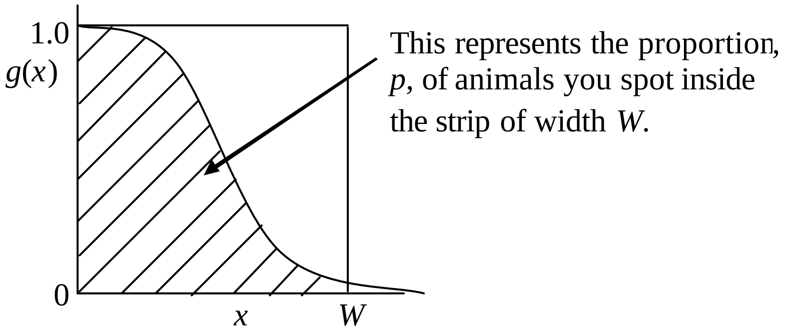
Once we have the detection function, we can proceed in two ways:

1. We can calculate the **effective strip width**, ESW, so that the number of animals detected **outside** the ESW exactly equals the number of animals missed **inside** the ESW. The calculation is then similar to that for fixed-width surveys, with the area surveyed being A = 2 x ESW x L, where L is the length of the transect. ("Effective strip width" is a bit of a misnomer: it's really the 'effective strip *half*-width'.) Distance sampling is sometimes referred to as '**variable-width sampling**'.



2. We can use the maximum distance from the line that we recorded animals, W, as the strip (half-) width, and calculate p, the **probability of observing an animal** which is present inside that strip. The

actual number of animals in the strip is $N = n / p$, where n is the number of animals seen, and we use N to calculate the density as for a fixed-width strip of area $A = 2 \times W \times L$.



The two approaches are equivalent, since ESW = $p \times W$. This said, the ESW concept may be easier to use, but p is analogous to detection probability in PRESENCE and MARK, and is important theoretically. DISTANCE calculates values for both.

We need one further concept. The histogram you drew from the muntjac data shows how many animals were seen at different distances, while $g(x)$ tells us the detection probability. To get probable numbers seen, we need to multiply $g(x)$ by the density of animals present (D): this is the '**density function**', $f(x)$. For line transects (but *not* point counts), $f(x)$ is the same shape as $g(x)$. Note that right on the transect, where $x = 0$, $g(0) = 1$ and $f(0) = D$.

Main points

- Distance sampling is based on the same ideas of plot or quadrat sampling, in that density is estimated by surveying a (spatial) **sample** of the area of interest. In plots or strip transects, we ensure we detect all the animals in the sample area.
- In distance sampling, we do not expect to see all the animals present. We assume that **probability of detection** is 1 for animals on the transect, and decreases away from the transect line.
- We measure the **perpendicular distance** from the transect to each animal detected, and fit a model (a mathematical equation) for the detection probability to the data, using likelihood and AIC.
- The model allows us to calculate the **effective strip width**, and we use this to calculate the density.
- The **DISTANCE** software package will do the calculations for us.