Lab 9 Assignment — Distance Sampling

Due before your next lab

The purpose of this lab is to learn how to use program DISTANCE or the R package 'Distance' to estimate abundance and density. You will do this by analyzing fake data on Mongolian gazelle (*Procapra gutturosa*). The data are fake, but they are based loosely on the study that we discussed in lecture. Note that gazelles were detected in groups (ie, herds), and the distance data are distances to group centers.

Put your answers in a Word file and upload it to ELC. Name the file something like "Chandler-lab9.docx". Due before your next lab.

Program DISTANCE

Part I: Format data and import into DISTANCE

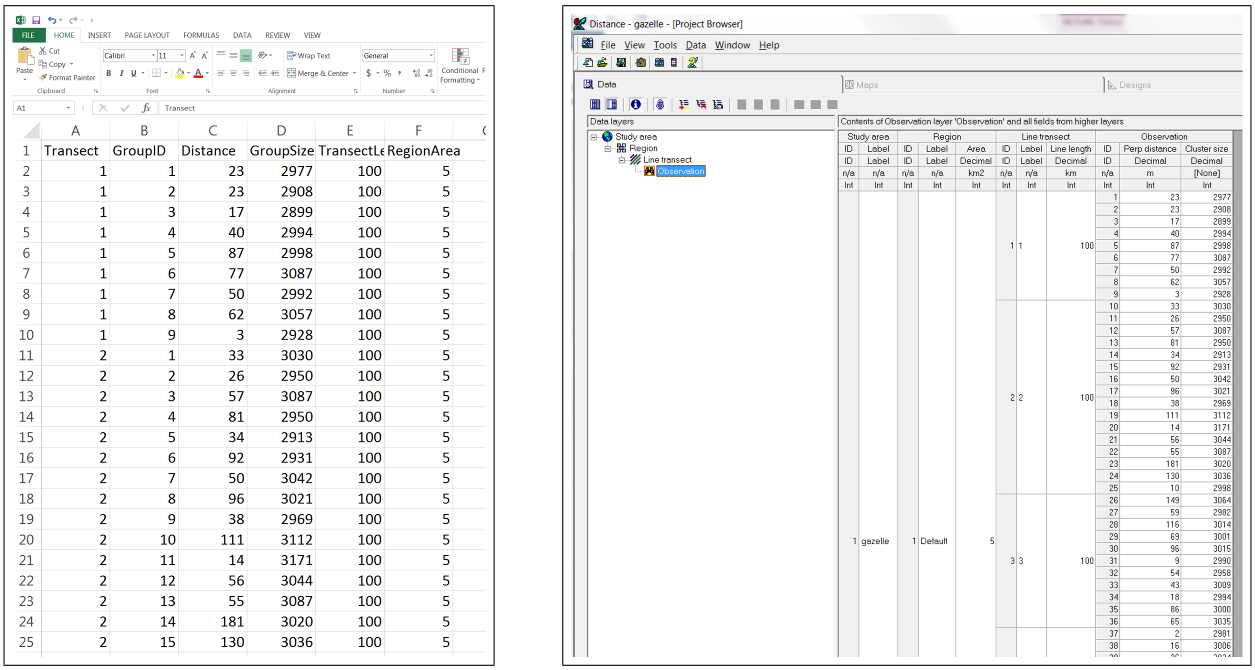
1. Open DISTANCE and create a new project by clicking `File > New Project ...`. Name the project something like "Gazelle.dst"
2. Use the "New Project Setup Wizard" and select the following options:
   (a) Analyze a survey that has been completed
   (b) Line transect, single observer, perpendicular distance, clusters of objects
   (c) Distance = Meter, transect length = Kilometers, area = Square kilometer
   (d) Do not add any multipliers
   (e) Proceed to Data Import Wizard
3. Import Data Wizard
   (a) Choose the "GazelleFakeData.txt" file.
   (b) Lowest layer = observation, Highest layer = region. Accept other defaults on Step 3 window and then hit "Next".
   (c) On next window, tell it to ignore first row of data by checking the box labeled 'Do not import first row'. Hit Next.
   (d) On the "Step 5" window, you need to label columns ("layer") by clicking on the grey boxes above each column.
      (i) Label the Transect column as: line transect, label, label
      (ii) Tell it to ignore the GroupID column
      (iii) Label distance column as: observation, perpendicular distance, decimal
      (iv) Label the group size column as: observation, cluster size, decimal

(v) Label transect length column as: line transect, line length, decimal

(vi) Label area column as: region, area, decimal

(e) Hit Next, and then choose "Overwrite existing data".

4. Hit Next again, and then click on "Observation" on the left to display the data. You should see data like that shown in Fig. 1 below. Area refers to the area covered by the 10 transects, which were 100 km long and 250 m wide. Cluster size is the number of gazelle detected in each group.



	A	B	C	D	E	F
1	Transect	GroupID	Distance	GroupSize	TransectLe	RegionArea
2	1	1	23	2977	100	5
3	1	2	23	2908	100	5
4	1	3	17	2899	100	5
5	1	4	40	2994	100	5
6	1	5	87	2998	100	5
7	1	6	77	3087	100	5
8	1	7	50	2992	100	5
9	1	8	62	3057	100	5
10	1	9	3	2928	100	5
11	2	1	33	3030	100	5
12	2	2	26	2950	100	5
13	2	3	57	3087	100	5
14	2	4	81	2950	100	5
15	2	5	34	2913	100	5
16	2	6	92	2931	100	5
17	2	7	50	3042	100	5
18	2	8	96	3021	100	5
19	2	9	38	2969	100	5
20	2	10	111	3112	100	5
21	2	11	14	3171	100	5
22	2	12	56	3044	100	5
23	2	13	55	3087	100	5
24	2	14	181	3020	100	5
25	2	15	130	3036	100	5

Figure 1: Data formatted in Excel (left) and the same data in program DISTANCE.

Part II – Estimate gazelle density and abundance

Half-normal detection function

1. Click the "Analyses" tab, then right click on the "New analysis" line, and then choose "Analysis details".
2. Name this analysis "Half-normal"
3. Under "data filters", click on "Properties" and name the data filter "truncate250". Click on the "Truncation" tab and tell it to discard all observations beyond 250m. Hit OK.
4. Under "model definitions" choose "Properties" and name it "HN" for half-normal. Then click the "Detection function" tab. Under "Models", specify a model with a half-normal "key function". Next, click on "Adjustment terms" and choose "Manual selection" with 0 adjustment terms. Click the "Cluster size" tab and choose "Use mean of observed clusters". Hit "OK".
5. Run the model.
6. On the "Results" tab on the right, scroll through the pages to the histograms of detection distances and the fitted detection function. Find the one that looks the best, in terms of the detection function fitting the histogram well, and then copy and paste it into your Word file. The easiest way to do this is to copy and paste into MS Paint and then save it as an image file. Then use `Insert > Picture` in Word. Or you can take a screen shot of the histogram and paste that in.
7. Add a figure caption below the histogram that explains the graph.
8. Create a table to report estimates of abundance (N), density (D), group density (DS), sigma (called A(1) in DISTANCE), and p. Include standard errors (SEs) and confidence intervals (CIs) in your table. Define each of the parameters in your table (one sentence per parameter).

Hazard-rate detection function

1. Run a second analysis in which you use a "hazard-rate" detection function instead of the half-normal detection function. Do this by closing the results window, and right-clicking under the "Analyses" tab to select "New Analysis…".
2. Next, right-click on the new line that appeared (should be highlighted in blue) and choose 'Analysis Details' again as you did before. Change the name of the analysis, and create a new 'Model definition' in which you change the key function to 'hazard-rate'. For all other options, use the same settings as before.
3. Repeat steps 6-7 above. Why do you think the results differ when you use different detection functions? Which of the two models is better (has the lower AIC)?

R package 'Distance'

Open R (or RStudio) and install the package using the following command:

```
install.packages("Distance")
```

Several other packages will be installed during the process. Now we can load the package like so:

```
library(Distance)
```

Before we can do any analysis, we need to import the Mongolian gazelle data. Make sure that the file `GazelleFakeData.txt` is in your working directory.

```
getwd()       ## Location on your computer where R will look for files

## [1] "C:/Users/rbchan/courses/applied-popdy/labs/distance"

list.files()  ## You should see 'GazelleFakeData.txt' here

## [1] "figs"                "GazelleFakeData.txt" "lab-distance.pdf"
## [4] "lab-distance.Rnw"    "lab-distance.tex"
```

If you need to change your working directory, you can use the dropdown menu options, or you can use a command like this:

```
setwd("C:/Users/RichardC/courses/") ## Change the path in quotes!
```

Once you have the data file in your working directory, you can import the fake gazelle data. The data file is a tab-separated text file instead of a comma-seperated text file like we used last week:

```
gazelleData <- read.table("GazelleFakeData.txt", header=TRUE, sep="\t")
```

The structure of the data can be assessed like this:

```
str(gazelleData)

## 'data.frame': 118 obs. of  6 variables:
##  $ Transect      : int  1 1 1 1 1 1 1 1 1 2 ...
##  $ GroupID       : int  1 2 3 4 5 6 7 8 9 1 ...
##  $ Distance      : int  23 23 17 40 87 77 50 62 3 33 ...
##  $ GroupSize     : int  2977 2908 2899 2994 2998 3087 2992 3057 2928 3030 ...
##  $ TransectLength: int  100 100 100 100 100 100 100 100 100 100 ...
##  $ RegionArea    : int  5 5 5 5 5 5 5 5 5 5 ...
```

We have to reformat the data to meet the requirements of the Distance package. The code below creates a new data frame, renames the columns, and converts the distance units to kilometers.

```
gazelleData2 <- gazelleData[,c("Transect", "GroupSize")]
colnames(gazelleData2) <- c("Sample.Label", "size")
```

```r
gazelleData2$distance <- gazelleData$Distance / 1000 ## Convert to km
gazelleData2$Region.Label <- 1
gazelleData2$Effort <- gazelleData$TransectLength
gazelleData2$Area <- gazelleData$RegionArea
```

The distance sampling model can be fitted to the data using the `ds` function. The `key` argument specifies the detection function. You could switch from the half-normal to the hazard-rate function using: `key="hr"`.

```r
model.hn <- ds(data=gazelleData2, key="hn",
               transect="line", truncation=250/1000,
               adjustment=NULL, quiet=TRUE)

## Fitting half-normal key function

## Key only model:  not constraining for monotonicity.

## AIC= -430.39
```

Once you have fitted the model, you can obtain the parameter estimates using the following command:

```r
summary(model.hn)
```

To view the estimated detection function, do this:

```r
plot(model.hn)
```

Answer all of the questions listed under “Part II” above.