

Introduction to distance sampling

Workshop, 21-23 August 2019

Centre for Research into Ecological and Environmental Modelling

Exercise 7. Analysis of stratified data

For this exercise, we use data from a survey of Antarctic minke whales. The study region was divided into two strata (North and South) the two strata were surveyed by different vessels at the same time. The whales tend to be found in high densities against the ice edge, where they feed, and so the density in the southern stratum is typically higher than in the northern stratum (Figure 1). This is the primary reason for using a stratified survey design. It is also the reason for covering the southern stratum more intensely: in the southern stratum the transect length per unit area is more than 2.5 times that of the northern stratum.

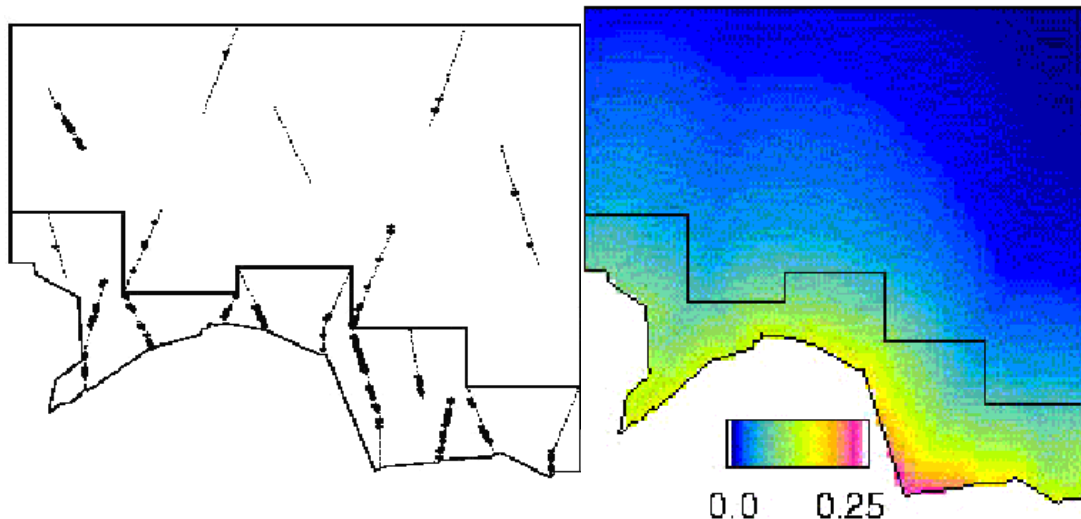


Figure 1. An example of the sort of survey design used and a typical whale density gradient. The irregular bottom border is the ice-edge. The ‘stepped’ black line defines the boundary between the strata; dotted lines are transects and dots are detections.

1 Objectives

The objectives of this exercise are to:

1. Create subsets of the data
2. Decide whether to fit separate detection functions or a pooled detection function
3. Specify different stratification options using the `dht2` function.

2 Getting started

Begin by reading in the data and in the commands below the data is renamed to avoid overwriting the original dataset. Distances are in kilometers and a truncation distance

of 1.5km is specified and used in the following detection function fitting. Perpendicular distances, transect lengths and study area size are all measured in kilometers; hence `convert.units` argument to `ds` is 1 and has been omitted. To keep things simple, a hazard rate detection function with no adjustments is used for all detection functions.

```
library(Distance)
data(minke)
# Rename data
whale <- minke
# Check data
head(whale)
# Specify truncation distance
whale.trunc <- 1.5
```

You will see that these data contain a column called 'Region.Label': this contains values 'North' or 'South'.

3 Full geographical stratification

First, we want to fit encounter rate and detection function separately in each strata. This is easily performed by splitting the data by region and using `ds` on each subset. The commands below do this for the southern region (note, there are alternative ways to select a subset of data).

```
# Create dataset for South
whale.S <- whale[whale$Region.Label=="South", ]
# Fit df to south
whale.df.S.strat <- ds(whale.S, key="hr", adjustment=NULL,
                      truncation=whale.trunc)
# Print results
summary(whale.df.S.strat)
```

Make a note of the AIC. Perform a similar commands to obtain estimates for the northern region. What is the total AIC?

Also make a note of the abundance in each region. What is the total abundance in the study region?

4 Fitting a pooled detection function

We want to compare the total AIC found previously with the AIC from fitting a detection function to all data combined. This is easy to obtain:

```
# Fit to pooled data
whale.df.all <- ds(whale, truncation=whale.trunc, key="hr", adjustment=NULL)
summary(whale.df.all)
```

Given the AIC value for the detection function from the pooled data, would you fit a separate detection function in each strata or not?

5 Stratification options using dht2

The command `summary(whale.df.all)` will provide the abundance estimates for each region and the total and for this simple example, this is sufficient. However, if we want to consider different stratification options, then the `dht2` function is useful.

After fitting a detection function, the `dht2` function, allows abundance estimates to be computed over some specified regions. In the command below, the pooled detection function is used to obtain estimates in each strata and over all (like the `summary` function previously used).

```
dht2(ddf=whale.df.all, flatfile=whales, strat_formula=~Region.Label,  
      stratification="geographical")
```

The arguments are:

- `ddf`
 - the detection function (fitted by `ds`)
- `flatfile`
 - the data object containing all the necessary information ¹
- `strat_formula=~Region.Label`
 - formula (hence the `~`) giving the stratification structure
- `stratification="geographical"`
 - in this example, we specify that each strata (specified in `strat_formula`) represents a geographical region.

Make a note the total abundance in the study region.

What happens if we were to ignore the regions and treat the data as though it came from one large study region? This can be done by changing the stratification formula, as shown below.

```
dht2(ddf=whale.df.all, flatfile=whales, strat_formula=~1,  
      stratification="geographical")
```

Has this changed the abundance estimate?

¹Data is referred to as being in a 'flatfile' format if it contains information on region, transects and observations. An alternative is to use a hierarchical structure and have region, transect and observation information in separate data files with links between them to ensure that transects are mapped to the relevant region and observations to the relevant transect.

Solution 7. Analysis of stratified data

Reading in the data:

```
library(Distance)
# Load data
data(minke)
# Rename data
whale <- minke
# Check data
head(whale, n=3)
```

```
##   Region.Label Area Sample.Label Effort distance
## 1      South 84734           1  86.75      0.10
## 2      South 84734           1  86.75      0.22
## 3      South 84734           1  86.75      0.16
```

```
# Specify truncation distance
whale.trunc <- 1.5
```

Fit detection function and encounter rate separately in each strata.

```
## Fit to each region separately - full geographical stratification
# Create dataset for South
whale.S <- whale[whale$Region.Label=="South", ]
# Fit detection function to South data
whale.df.S.strat <- ds(whale.S, truncation=whale.trunc, key="hr",
                      adjustment=NULL)
summary(whale.df.S.strat)
```

```
##
## Summary for distance analysis
## Number of observations : 39
## Distance range       : 0 - 1.5
##
## Model : Hazard-rate key function
## AIC   : 8.617404
##
## Detection function parameters
## Scale coefficient(s):
##           estimate      se
## (Intercept) -0.5102605 0.1921476
##
## Shape coefficient(s):
##           estimate      se
## (Intercept) 1.242147 0.3770297
##
##           Estimate      SE      CV
```

```

## Average p          0.4956472  0.06629151 0.1337474
## N in covered region 78.6849968 13.81375134 0.1755576
##
## Summary statistics:
##   Region Area CoveredArea Effort  n  k          ER          se.ER          cv.ER
## 1  South 84734          1453.23 484.41 39 13 0.08051031 0.01809954 0.2248102
##
## Abundance:
##   Label Estimate          se          cv          lcl          ucl          df
## 1 Total 4587.914 1200.141 0.2615874 2687.511 7832.138 21.13917
##
## Density:
##   Label Estimate          se          cv          lcl          ucl          df
## 1 Total 0.0541449 0.01416363 0.2615874 0.03171703 0.09243206 21.13917

```

```

# Combine selection and df fitting for North
whale.df.N.strat <- ds(whale[whale$Region.Label=="North", ],
                      truncation=whale.trunc, key="hr", adjustment=NULL)
summary(whale.df.N.strat)

```

```

##
## Summary for distance analysis
## Number of observations : 49
## Distance range       : 0 - 1.5
##
## Model : Hazard-rate key function
## AIC   : 37.27825
##
## Detection function parameters
## Scale coefficient(s):
##           estimate          se
## (Intercept) -0.01081049 0.2203317
##
## Shape coefficient(s):
##           estimate          se
## (Intercept) 1.022021 0.6908073
##
##           Estimate          SE          CV
## Average p          0.7592228 0.09979405 0.1314424
## N in covered region 64.5396850 9.61422927 0.1489662
##
## Summary statistics:
##   Region Area CoveredArea Effort  n  k          ER          se.ER          cv.ER
## 1  North 630582          4075.14 1358.38 49 12 0.03607238 0.01317937 0.3653591
##
## Abundance:
##   Label Estimate          se          cv          lcl          ucl          df
## 1 Total 9986.789 3877.709 0.3882839 4470.104 22311.77 13.97691
##

```

```
## Density:
##   Label   Estimate          se          cv          lcl          ucl          df
## 1 Total 0.01583742 0.006149413 0.3882839 0.007088855 0.03538283 13.97691
```

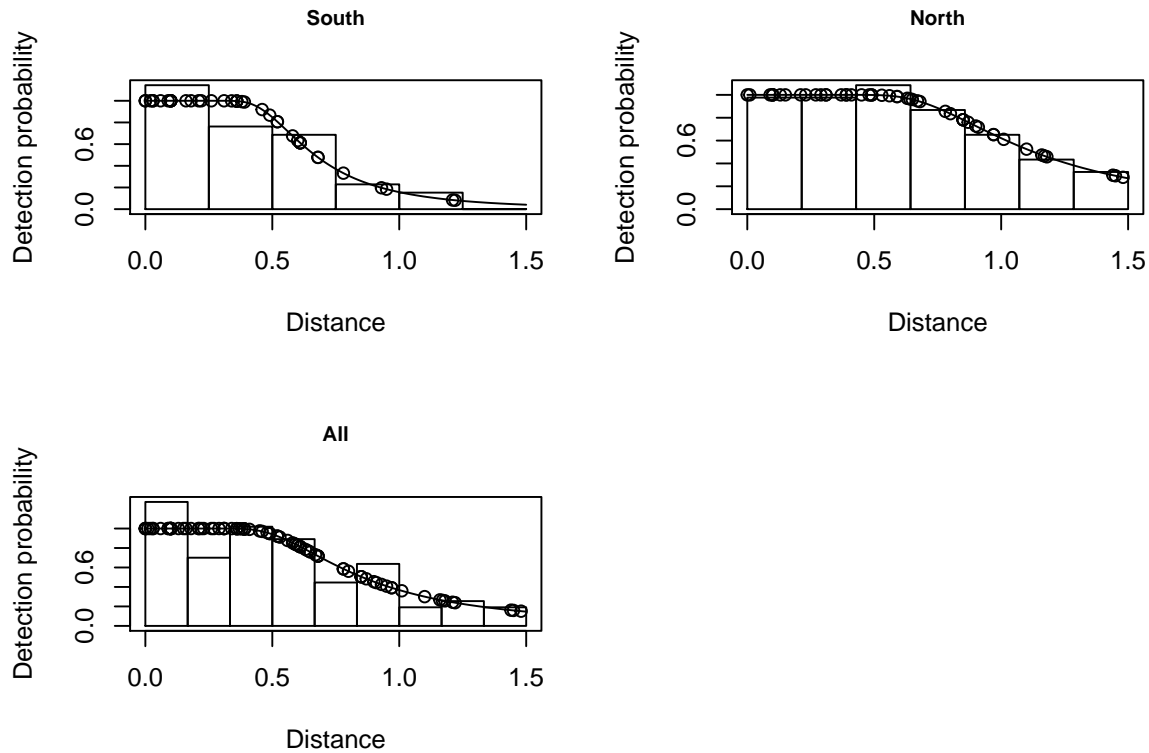
Next we fitted a pooled detection function.

```
## Fit to pooled data
whale.df.all <- ds(whale, truncation=whale.trunc, key="hr", adjustment=NULL)
summary(whale.df.all)
```

```
##
## Summary for distance analysis
## Number of observations : 88
## Distance range          : 0 - 1.5
##
## Model : Hazard-rate key function
## AIC    : 48.63688
##
## Detection function parameters
## Scale coefficient(s):
##           estimate          se
## (Intercept) -0.2967912 0.176565
##
## Shape coefficient(s):
##           estimate          se
## (Intercept) 0.964833 0.3605099
##
##           Estimate          SE          CV
## Average p          0.6224451 0.06679541 0.1073113
## N in covered region 141.3779400 17.77437207 0.1257224
##
## Summary statistics:
##   Region   Area CoveredArea  Effort  n  k          ER      se.ER      cv.ER
## 1 North 630582      4075.14 1358.38 49 12 0.03607238 0.01317937 0.3653591
## 2 South 84734      1453.23  484.41 39 13 0.08051031 0.01809954 0.2248102
## 3 Total 715316      5528.37 1842.79 88 25 0.04775368 0.01129627 0.2365529
##
## Abundance:
##   Label   Estimate          se          cv          lcl          ucl          df
## 1 North 12181.313 4638.5533 0.3807926 5499.920 26979.371 12.96742
## 2 South 3653.313  910.0737 0.2491091 2181.589  6117.879 17.96142
## 3 Total 15834.626 4834.1866 0.3052921 8388.389 29890.769 15.25427
##
## Density:
##   Label   Estimate          se          cv          lcl          ucl          df
## 1 North 0.01931757 0.007355987 0.3807926 0.008721974 0.04278487 12.96742
## 2 South 0.04311508 0.010740360 0.2491091 0.025746327 0.07220099 17.96142
## 3 Total 0.02213655 0.006758113 0.3052921 0.011726829 0.04178680 15.25427
```

The AIC value for the detection function in the South was 8.617 and the AIC for the

North was 37.28. This gives a total AIC of 45.9. The AIC value for the pooled detection function was 48.64. Because 48.64 is greater than 45.9, estimation of separate detection functions in each stratum is preferable.



6 Stratification options

In the full geographical stratification, both encounter rate and detection function were estimated separately for each region (or strata). This resulted in the following abundances:

Table 1: Abundance estimates using full geographical stratification.

Label	Estimate
North	9987
South	4588
Total	14575

Next, the distances were combined to fit a pooled detection function but encounter rate was obtained for each region. This resulted in the following abundances:

Table 2: Abundance estimates calculating encounter rate by strata and a pooled detection function.

Label	Estimate
North	12181
South	3653
Total	15835

An equivalent result could be produced using the `dht2` function, which does not require

```
# Geographical stratification with pooled detection function
newdht <- dht2(ddf=whale.df.all, flatfile=whale, strat_formula=~Region.Label,
               stratification="geographical")
print(newdht, report="abundance")
```

```
## Summary statistics:
## Region.Label Area CoveredArea Effort n k ER se.ER cv.ER
## North 630582 4075.14 1358.38 49 12 0.036 0.013 0.365
## South 84734 1453.23 484.41 39 13 0.081 0.018 0.225
## Total 715316 5528.37 1842.79 88 25 0.048 0.022 0.469
##
## Abundance estimates:
## Region.Label Estimate se cv LCI UCI df
## North 12181 4638.553 0.381 5500 26979 12.967
## South 3653 910.074 0.249 2182 6118 17.961
## Total 15835 4834.187 0.305 8388 29891 15.254
##
## Component percentages of variance:
## Region.Label Detection ER
## North 7.94 92.06
## South 18.56 81.44
## Total 4.98 95.02
```

If geographic stratification were ignored, the abundance estimate of would be 18,293 whales. This estimate is substantially larger than the estimates above. The reason is that the survey design was geographically stratified with less survey effort in the north and more in the south and this is being ignored in the unstratified analysis.