

**Exercise 1:** We will slowly go through a distance analysis by hand:

Suppose we walked a single transect that was 500 m long and counted every squirrel we see. Our truncation distance is 30 m and the transect was randomly positioned in a region of 1 square Kilometer. We obtained the following distances (in meters) to the squirrels sighted, perpendicular distance to the transect (call these Dists from now on):

17.9, 0.3, 7.3, 16.0, 3.6, 33.7, 19.4, 11.0, 1.3, 2.6, 0.8, 25.9, 3.7, 1.1, 4.1,  
1.3, 0.0, 8.0, 21.0, 13.7, 6.7, 14.4, 9.4, 10.7, 2.0, 8.3, 7.6, 6.5, 3.1, 12.3

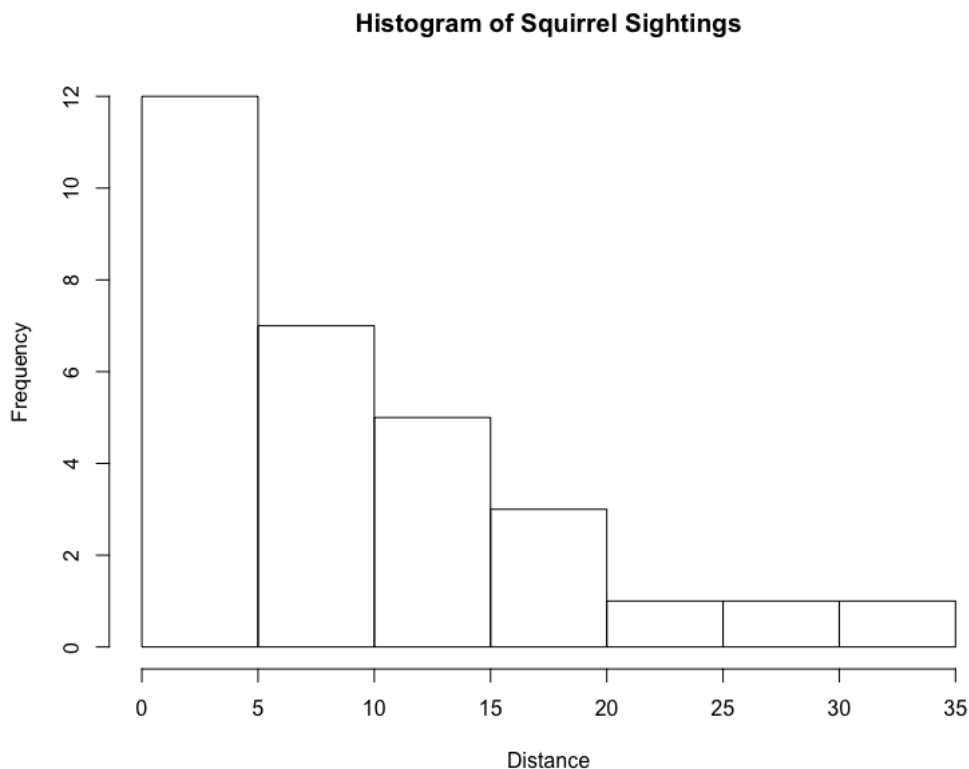
or, as R data:

```
c(17.9, 0.3, 7.3, 16, 3.6, 33.7, 19.4, 11, 1.3, 2.6, 0.8, 25.9, 3.7, 1.1, 4.1,  
1.3, 0, 8, 21, 13.7, 6.7, 14.4, 9.4, 10.7, 2, 8.3, 7.6, 6.5, 3.1, 12.3)
```

- a. Make a histogram plot showing the proportion in 0 to 5m, 5 to 10m, 10 to 20m, etc. Of the models: Uniform (flat), half-normal and (negative) exponential, which seems to be a better fit?

**Solution:**

Looking at the histogram I would suggest that negative exponential looks to be the best fit, but it might be worthwhile to experiment with the half-normal as well.



- b. Let's try the exponential model. It has the probability function  $f(x) = \frac{1}{b} e^{\frac{-x}{b}}$ , where  $b$  is the parameter we need to estimate. In this case, the maximum likelihood estimate of  $b$  is just the mean of the  $x$  values. Compute the fitted density function (that is, plug the estimated  $b$  into the probability function).

**Solution:**

Computing the mean of the distances we get,  $b = 9.123$ . Therefore our probability function looks like,

$$p(x) = \frac{1}{9.123} e^{\frac{-x}{9.123}}$$

- c. Now plug each distance into the fitted probability function to get 30 probability values. Multiply these all together. This is the likelihood. The AIC is then  $-2\ln(\text{likelihood}) + 2(\text{number of parameters})$ . In this case, the number of parameters is 1 (only  $b$ ). What is the AIC of this model?

**Solution:**

Computing the likelihood and AIC in R we get an AIC of 194.6501,

**Code:**

```
> y = (1/mean(P11))*exp((-P11)/mean(P11))
> prod(y)
> -2*log(prod(y)) + 2
[1] 194.6501
```

- d. Finally, compute  $f(0)$ , which you get by plugging zero into the density function you obtained in part (b). Plug this value into the line transect density formula to estimate the density of squirrels per square meter. Convert this to an estimate of the density per square Kilometer.

**Solution:**

Recall that the line transect density estimator formula is given by,

$$\hat{\lambda} = \frac{n\hat{f}(0)}{2L}.$$

Computing the estimator in R we get the following,

**Code:**

```
> density = (length(P11)*(1/mean(P11))*exp(0/mean(P11)))/(2*500)
> density*1000000
[1] 3288.272
```

**Exercise 2:** Now repeat the above analysis for the case where you use a half-normal probability function.

- a. In this case the formula for the probability function is  $f(x) = \frac{\sqrt{2}}{\sqrt{\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$ . The maximum likelihood estimator of  $\sigma$  is rather simple: first, square all the distances, then average them, then take the square root of this number. What is the detection function (after you plug in your estimated  $\sigma$ )?

**Solution:**

Computing the maximum likelihood estimator of  $\sigma$  in r we get,  $\sigma = 12.2075$ . So our half-normal probability function comes out to be,

$$f(x) = \frac{\sqrt{2}}{\sqrt{\pi}(12.2075)} e^{-\frac{x^2}{2(12.2075)^2}}$$

- b. This was a maximum likelihood estimator of  $\sigma$ . In general, how do you get maximum likelihood estimators? (i.e. what **\*\*IS\*\*** a maximum likelihood estimator)?

**Solution:**

As we described in the previous homework likelihood is used to describe the joint probability of achieving our sampled data given certain model parameters. The maximum likelihood estimator is the model parameters which maximizes that joint probability. In general deriving the estimator involves forming the likelihood function and then taking the derivative with respect to the model parameters and solving for when it equals zero, like any other optimization problem. Note that since the likelihood function is often a very large product (depending on size of data), usually we compute the log-likelihood function in order to avoid overflow errors.

- c. Next, plug each distance into the probability function to get 30 probabilities. Multiply these together to get the likelihood. Then compute the AIC. How does it compare to the AIC for the exponential model? Which model fits best?

**Solution:**

Computing the AIC in r we get a value of 195.6705, recall that the exponential model achieved 194.6501 which is slightly better.

**Code:**

```

> f <- (sqrt(2)/(sqrt(pi)*(12.2075)))*exp(-(P11^2/(2*(12.2075)^2)))
0.022306507 0.065340460 0.054659092 0.027687635 0.062579028
0.001446957 0.018488627 0.043551144 0.064990633 0.063894441
0.065219995 0.006884148 0.062425942 0.065095384 0.061775864
0.064990633 0.065360193 0.052729841 0.014883848 0.034819537
0.056221494 0.032595757 0.048591529 0.044512864 0.064488871
0.051871768 0.053845449 0.056721698 0.063286376 0.039342580

> prod(f)
[1] 8.810186e-43

> -2*log(prod(f)) + 2
[1] 195.6705

```

- d. Finally, compute  $f(0)$ , then compute the density of squirrels under the half-normal model.

**Solution:**

Computing the density in `r` we get that the half-normal model gives us 1960.806 squirrels per square kilometer,

**Code:**

```

> density = (length(P11)*
             (sqrt(2)/(sqrt(pi)*(12.2075)))*exp(-(0^2/(2*(12.2075)^2))
             /(2*500))
> density
[1] 0.001960806
> density*1000000
[1] 1960.806

```

**Exercise 3:** There is a good reason for truncation. Suppose you were using the exponential model, but the last squirrel you saw wasn't at 12.3 meters: due to an amazing circumstance you managed to see a squirrel at a distance of 180 m, across a small clearing. How much does your estimated density change when you use this new observation? ( You don't have to calculate the AIC, but the effect on the AIC is very severe, also, which puts the entire model in doubt.)

**Solution:**

Recomputing the density with a new last encounter and the exponential model we get a value of 2038.967 squirrels per kilometer. This is a dramatically different from the previous calculation.

**Code:**

```
> P11 <- c(17.9, 0.3, 7.3, 16, 3.6, 33.7, 19.4, 11, 1.3,
           2.6, 0.8, 25.9, 3.7, 1.1, 4.1, 1.3, 0, 8, 21,
           13.7, 6.7, 14.4, 9.4, 10.7, 2, 8.3, 7.6, 6.5,
           3.1, 180)

> density = (length(P11)*(1/mean(P11))*exp(0/mean(P11)))/(2*500)
> density
[1] 0.002038967

> density*1000000
[1] 2038.967
```

**Exercise 4:** I have attached two papers. The first paper (Harbor Porpoises) discusses a line transect survey. Give a brief description of the survey. In particular, why did they stratify the study?

**Solution:**

The goal of the survey was to record the distribution, abundance, and trends of Harbor Porpoise in the Glacier Bay National Park and the Icy Strait. From 1991 to 1993 line transect surveys were conducted three times a year to record baseline abundance and distribution patterns as well as population estimates for the total number of Harbor Porpoises. Between 94 to 05 the surveys were continued, Harbor Porpoise presence data was collected(not via line-transect survey) as a by product of a 12 year killer whale survey. This data showed fewer Porpoise encounters, however the study cites this as being likely due to poorer data(no effort data, and differences in methodology.) As a result, in order to model the porpoise population 4 years of line transect surveys were conducted with a similar methodology to the early 90's surveys. The region of interest was divided into 5 stratum, and when a sighting was made sighting angle, radar distance, species, counts, and direction of travel were recorded. Search effort was recorded by transect length, weather conditions, visibility index, and observer position. The stratum were decided because previous studies identified that there is plenty of porpoise movement inside of these 'sub regions' but not outside, and by stratifying we can take advantage of this behavior to reduce our variance.

**Exercise 5:** The other paper discusses a study of birds. In this case the authors were worried that one of the main assumptions of line transect studies was violated. What assumption was that? How did they deal with this problem?

**Solution:**

The second paper is similar to the first in the sense that the survey was conducted using line transect sampling. The subject of the survey was Murrelets in Glacier Bay National Park. The authors of the survey were very concerned with the assumption of 100 percent detection near the center line of the transect, since murrelets are 'small, cryptic, and highly mobile'. To relax this assumption an extra term was added to the density estimator to adjust the detection function (in the paper it is referred to as  $P_c$ ).

**Exercise 6:** Suppose we wish to monitor changes in the bird population in a 3 square-Kilometer park, over several years. We also would like to get a good estimate of the total population. Our options are:

- \* Randomly place 10 point transects. Return to the same transects each year.
- \* Randomly place 10 point transects. Each year re-randomize the locations.
- \* Randomly place 5 point transects where we will return to each year, and each year place an additional 5 that are just sampled one year.

- a. What are the pros and cons of each sampling plan? Which would you recommend?

**Solution:**

By selecting the same point transects every year we are allowing more bias to get a better picture of how the populations might change, since we have directly comparable measurements year to year. By re-randomizing the locations we accept more variance in our estimator since our samples are not directly comparable year over year, however we encounter less bias. A mix of both could do a good job at capturing the change year over year while still containing enough variability to be generalizable. I would go with the third option.

- b. I suppose we have to also define what randomized means. Some designs include SRS or systematic (grid). Which design (could be one other than SRS or systematic) would you recommend?

**Solution:**

From what we've seen in class and in the text, most often a systematic grid sample

used when sampling data in the field. As it offers the greatest chance to capture the most variability in the population. A srs over a space could result in a clustered sample which is not representative of the population.

- c. In the analysis, would you want a single detection curve for all transects, or would you allow for different detection curves for each transect? (explain)

**Solution:**

I would allow for different detection curves, as terrain and conditions can vary within time and space and it is important to account for those changes in our estimator. For example one point-transect could be located in a portion of the park with thick brush and forestry that obscures detectability, while another could be out in an open field. These transects must have different detectability curves.

**Exercise [:** 7] Using the `ds` function in R, get a population DENSITY estimate with a 95 percent confidence interval, based on the following data, which is assumed to be based on a single transect.

Transect length = 100 m

Units of distance from transect: meters

Area of surveyed region = 10,000 square meters.

Distances (in meters): `d = c(14, 2, 16, 2, 26, 38, 13,`

`22, 5, 17, 16, 4, 28, 11, 5, 16, 6, 10, 14, 17, 14,`

`9, 7, 12, 38, 5, 2, 4, 1, 26)`

**Solution:**

Using the `ds` function in `r` we can check a whole host of key functions and adjustments to find the best model for detectability. Doing so we get that the uniform key function with `cos` adjustment fits best. Setting up the dataframe to compute the density with the `ds` function we get a density of 0.006734768 with a confidence interval of (0.008897834, 0.004571702).

**Code:**

```
## Seeing which keyfunction-adjustment pair provides the best fit
> AIC(tmp_hn, tmp_hr, tmp_unif, tmp_hn_herm,
      tmp_hn_null, tmp_hn_poly, tmp_hr_herm,
      tmp_hr_null, tmp_hr_poly, tmp_unif_herm,
      tmp_unif_poly)
```

df	AIC
----	-----

```

tmp_hn      1 212.3018
tmp_hr      2 213.2516
tmp_unif     1 211.1484
tmp_hn_herm  1 212.3018
tmp_hn_null  1 212.3018
tmp_hn_poly  1 212.3018
tmp_hr_herm  2 213.2516
tmp_hr_null  2 213.2516
tmp_hr_poly  2 213.2516
tmp_unif_herm 2 212.9460
tmp_unif_poly 2 213.0254

```

```

## We can see that it is the uniform function with cos adjustment
## Generating dataframe to compute abundance.

```

```

> D <- data.frame(distance=d,
                  Area = rep(10000,length(d)),
                  Region.Label=rep(1,length(d)),
                  Sample.Label = rep(1,length(d)),
                  Effort = rep(100, length(d)))
> D

```

	distance	Area	Region.Label	Sample.Label	Effort
1	14	10000	1	1	100
2	2	10000	1	1	100
3	16	10000	1	1	100
4	2	10000	1	1	100
5	26	10000	1	1	100
6	38	10000	1	1	100
7	13	10000	1	1	100
8	22	10000	1	1	100
9	5	10000	1	1	100
10	17	10000	1	1	100
11	16	10000	1	1	100
12	4	10000	1	1	100
13	28	10000	1	1	100
14	11	10000	1	1	100
15	5	10000	1	1	100
16	16	10000	1	1	100
17	6	10000	1	1	100
18	10	10000	1	1	100
19	14	10000	1	1	100
20	17	10000	1	1	100
21	14	10000	1	1	100
22	9	10000	1	1	100
23	7	10000	1	1	100



24	12	10000	1	1	100
25	38	10000	1	1	100
26	5	10000	1	1	100
27	2	10000	1	1	100
28	4	10000	1	1	100
29	1	10000	1	1	100
30	26	10000	1	1	100

```
## Refitting detectability model
```

```
> tmp_unif_cos <- ds(data = D, transect = "line", key = 'unif', adjustment = 1)
```

```
Starting AIC adjustment term selection.
```

```
Fitting uniform key function with cosine(1) adjustments
```

```
AIC= 211.148
```

```
Fitting uniform key function with cosine(1,2) adjustments
```

```
AIC= 212.702
```

```
Uniform key function with cosine(1) adjustments selected.
```

```
## Pulling density estimate from summary report.
```

```
> summary(tmp_unif_cos)
```

```
Summary for distance analysis
```

```
Number of observations : 30
```

```
Distance range : 0 - 38
```

```
Model : Uniform key function with cosine adjustment term of order 1
```

```
Strict monotonicity constraints were enforced.
```

```
AIC : 211.1484
```

```
Detection function parameters
```

```
Scale coefficient(s):
```

```
NULL
```

```
Adjustment term coefficient(s):
```

```

              estimate      se
cos, order 1 0.7061412 0.1784585
```

	Estimate	SE	CV
Average p	0.586118	0.06130662	0.1045977
N in covered region	51.184236	8.05021737	0.1572792

```
Summary statistics:
```

Region	Area Covered	Area	Effort	n	k	ER	se.ER	cv.ER
1	1	10000	7600	100	30	1	0.3	0

```
Abundance:
```

	Label	Estimate	se	cv	lcl	ucl	df
1	Total	67.34768	10.81533	0.1605896	40.37697	112.3341	2.961142

Density :

	Label	Estimate	se	cv	lcl	ucl	df
1	Total	0.006734768	0.001081533	0.1605896	0.004037697	0.01123341	2.961142

>