

# Introduction to Distance Sampling

## Analysis of Clustered Data Outline Solutions

Example analyses, which were used in getting these solutions, and which are referred to below, are in the project file "Cluster solutions.dst".

1. Relevant results can be found in the analysis "E(s) by ln(s)\_g(x)".

- (a) No, the mean observed cluster size is 2.25 (se=0.229) the regression estimate of E(s) is 1.89 (se=0.139). The regression method not only corrects for size bias, but has also given a smaller standard error.
- (b) 7.2% of the variance of the density estimate comes from mean cluster size estimation – so a small amount compared to the variance caused by encounter rate and estimating the detection function.

2. Relevant results can be found in the analysis "truncation E(s)".

The detection function shoulder extends out to about the end of the second distance interval, so all data beyond this were discarded for estimation of cluster size (NB: this truncation does not affect the estimation of the detection function or encounter rate).

- (a) It is different because in this analysis all data beyond the second distance interval have been discarded in an attempt to eliminate any size bias in the data. Compare this result (1.85, se=0.183) with the mean of the observed clusters using the untruncated data (2.25, se=0.229) - note how this result is much closer to the estimate of E(s) using regression (1.89, se=0.139).
- (b) It is largely because the observed mean cluster size is based on only 41 observations, while the regression estimate in analysis 1 above is based on 88 observations – you pay for discarding data with increased variance.

3. Relevant results can be found in the analyses "Post-stratified E(s) using mean", "Post-stratified E(s)\_pooled f(0)\_regr" and "Post-stratified E(s)\_strat f(0)\_regr".

- (a) Mean cluster size is not relevant for strata 1 & 2 as the clusters in each were all the same size. The mean cluster size for the final stratum was 5.81 (se = 0.748).
- (b) In the analysis "Post-stratified E(s)\_pooled f(0)\_regr", cluster size strata are pooled for estimation of the detection function and the regression method has been used within cluster size strata. The regression estimate for the third stratum is 4.36 (se = 0.563) which is lower than the mean cluster size, suggesting that size bias is present in this stratum. So, for these data, it would not have been correct to assume that the effect of size bias had been eliminated by post-stratifying (and then using the mean of the observed cluster sizes in each stratum).
- (c) In the analysis "Post-stratified E(s)\_strat f(0)\_regr", the detection function has been estimated separately in each cluster size stratum. The detection functions are different from each other - it looks like cluster sizes 3 and above are detected with certainty almost all the way out to 1.2nm. It is questionable whether this is actually possible. In addition, the sample size for the third stratum is very small (only 16 observations) – less than is usually recommended for modelling a detection function.

### Overall conclusion: considering all the analyses

- There are questions raised about all the analyses using post stratification. Post stratification and using the mean cluster size per stratum did not eliminate size bias, as we discovered when we checked by using post-stratification with regression. Using a pooled detection function was not ideal, as we suspected that the detection functions would be different for different strata. This was confirmed when detection functions were estimated per stratum. However, the sample size in the third stratum was too small to have enough evidence to believe that the fitted detection function was plausible.
- That leaves the regression and the truncation method to choose between. There were no problems with the regression method, and although the truncation method gave a similar estimate of cluster size, data were thrown away, resulting in a larger standard error.

Overall conclusion: the regression method is the preferred method.