How to account for study day?

Scott Jennings

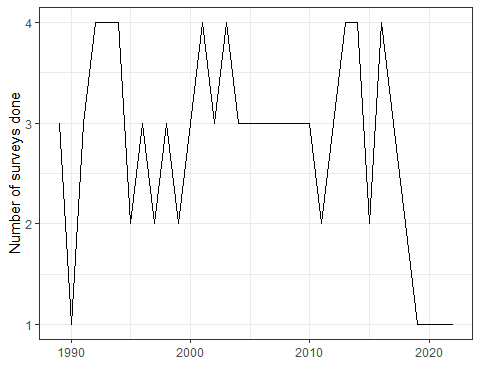
12/7/2021

We conducted 1-4 censuses per year. We are interested an estimate of the annual number of waterbirds using Tomales Bay. We have 2 options to handle study day to arrive at an annual estimate:

1. average (or otherwise summarize; e.g. median, 75th percentile) across all censuses in each year to derive an annual estimated mean number of birds, and use this value as the response in models
   * advantage: relatively simple math and yields simpler models with fewer covariates
   * disadvantage: within year variance in bird abundance is lost in our estimate of between year changes in abundance. number of censuses done is considered unimportant in our estimate of annual abundance (unless we include some weighting)
2. treat counted abundance on each census as the response in models and include study day directly as a model variable
   * advantage: between census variability and number of censuses done are included directly in the estimate of between year changes in abundance
   * disadvantage: more complicated model structures; need to decide how to model study day (I think quadratic would be a logical, flexible choice)

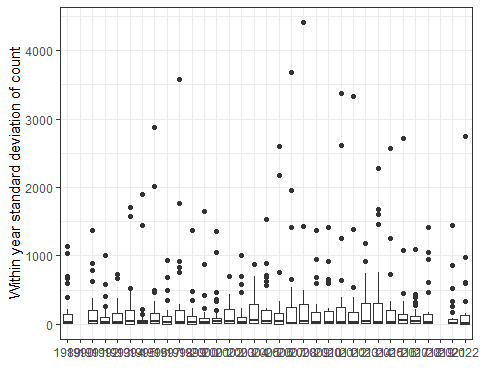
To help choose between these, we’ll do some basic data exploration and visualizations looking at the within year variability in abundance counts.

First, how many counts were done each year?



| number of censuses done | number of years with this many censuses |
| --- | --- |
| 1 | 4 |
| 2 | 6 |
| 3 | 15 |
| 4 | 8 |

Is the variation among counts in the same year constant across years?



There are some high variance year X species instances. Does this high variance lead to different inference if the mean, median, 75th percentile, or max value are used to estimate trends? The 75th percentile will allow high counts to have a greater effect on the estimate of annual abundance for a species when there is high variation among counts that year. It may be more important to consider individual counts when we are investigating ecological factors associated with variation in abundance.

It appears that the overall pattern in abundance is relatively insensitive to how we summarize across visits within a year. Mean and median generally yield similar estimates each year, but 75th percentile and especially maximum enhances the effect of some high counts (we probably wouldn’t really use max; it is included for reference). The result of this is that we may end up with different species meeting the definition of “concerning decline” (based on IUCN 20% over 10 years) if we use mean/median vs 75th percentile. For some species (MALL, CLGR, RTLO, BRAC, DCCO, AMWP) we may want to check mean and 75th percentile.

