## INFO ENTRY - QUESTION INFO

## Topic Info

|  |  |
| --- | --- |
| info\_id | mod\_rai |

## Note banner

<!--

:::{hint}

replace me with text

:::

-->

## Overview

```{include} include/00\_coming\_soon.md

```

## In-depth

```{include} include/note\_adapted\_clarke\_et\_al\_2023.md

```

In ecology, relative abundance (RA) is any count of animals or animal sign (e.g., number of deer sighted, number of bird vocalizations per unit time, number of moose tracks per kilometer of transect) that is assumed to correlate with absolute abundance ({{ rtxt\_obrien\_2011 }}). RA is a controversial index for two reasons: 1) there is often no documented relationship between the number of animals or signs observed and population size (i.e., index validation), and 2) detection probability is assumed to be constant between the areas, times or species being compared ({{ rtxt\_obrien\_2011 }}; {{ rtxt\_thompson\_et\_al\_1998 }}).

To the first point: the relationship between the number of animals or signs and abundance is rarely established ({{ rtxt\_burton\_et\_al\_2015 }}). Researchers often assume that counts and population size scale linearly – but many other kinds of relationships are possible. When the assumed relationship between counts and abundance diverges from the actual relationship, inferences from RA are not very meaningful ({{ rtxt\_thompson\_et\_al\_1998 }}). Validating a count-abundance relationship requires comparison with a robust, accurate estimate of absolute density (e.g., {{ rtxt\_krebs\_et\_al\_1987}}; {{ rtxt\_rovero\_marshall\_2009 }}; {{ rtxt\_villette\_et\_al\_2016 }}).

To the second point: consider the canonical equation,

```{figure} ../03\_images/03\_image\_files/clarke\_et\_al\_2023\_eqn\_rai1.png

:width: 80px

:align: center

```

where \*𝑁\* is population size, \*𝐶\* is the count of animals or signs and \*𝑝\* is detection probability ({{ rtxt\_anderson\_2001 }}; {{ rtxt\_brennan\_2019 }}). This equation underlies many estimators of abundance, including capture-recapture (CR; see {bdg-link-primary-line}`Capture-recapture (CR) / Capture-mark-recapture (CMR)<https://ab-rcsc.github.io/rc-decision-support-tool\_concept-library/02\_dialog-boxes/03\_10\_mod\_cr\_cmr.html>`) and distance sampling (DS; see {bdg-link-primary-line}`Distance sampling<https://ab-rcsc.github.io/rc-decision-support-tool\_concept-library/02\_dialog-boxes/03\_20\_mod\_ds.html>`) methods ({{ rtxt\_obrien\_2011 }}). RA comparisons assume that detection probability \*𝑝\* is constant across space, time or species, and can therefore be ignored ({{ rtxt\_anderson\_2001 }}; {{ rtxt\_obrien\_2011 }}; {{ rtxt\_sollmann\_et\_al\_2013b }}), such that:

```{figure} ../03\_images/03\_image\_files/clarke\_et\_al\_2023\_eqn\_rai2.png

:width: 80px

:align: center

```

so count essentially becomes a surrogate for population size.

Assuming constant detection probability \*𝑝\* is problematic, since the likelihood an animal or sign is counted during a survey will vary with observational, environmental, and habitat- and species-specific factors, which in turn can vary with time ({{ rtxt\_anderson\_2001 }}). For example: at site A, animals may be difficult to spot in dense vegetation, while at site B, animals may be easy to spot in open grassland; and the effects of vegetation on observability may differ seasonally. If the effects of vegetation on detectability are not accounted for, how can we be sure that differences in animal counts at site A and B are due to true differences in abundance, and not simply artefacts of detection bias ({{ rtxt\_sollmann\_et\_al\_2013b }})?

In a camera trapping context, RA is the comparison of detection rates across space, time or species – where detection rates are typically reported as the number of images per 100 trap days, but can also be reported in terms of the total number of detections, other units of effort (e.g., camera trap hours), proportion of stations with detections, etc. ({{ rtxt\_burton\_et\_al\_2015 }}). As with other kinds of RA surveys, comparisons of camera trap detection rates can confound abundance with animal behaviour and observability ({{ rtxt\_anderson\_2001 }}; {{ rtxt\_burton\_et\_al\_2015 }}).

RA has been criticized as an abundance estimator. Anderson (2001) condemned the index as "unprofessional," while O’Brien (2011) called it a "metric of last resort." Sollmann et al. (2013b) used simulations to determine that camera trap RA analyses did not detect changes in big cat density, and called use of the index for wildlife management "alarming." Nevertheless, some researchers have had success with the method and/or have argued for its conceptual and practical advantages (e.g., {{ rtxt\_rowcliffe\_carbone\_2008 }}; {{ rtxt\_johnson\_2008 }}; {{ rtxt\_palmer\_et\_al\_2018 }}; {{ rtxt\_rovero\_marshall\_2009 }}). Broadley et al. (2019) used simulations to show that RA could be sensitive to density-dependent movement, but generally tracked abundance well. Banks-Leite (2014) emphasized the importance of careful sampling design and protocols to control for variation in detectability, arguing that researchers should not solely rely on statistical corrections.

Ultimately, there is no "silver bullet" and researchers must carefully consider their inferential objectives and potential sources of sampling and estimation bias when choosing response variables and modelling frameworks for camera trap data.

## Figures

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | image | file\_name | caption | ref\_id |
| rai | A diagram of a diagram  Description automatically generated | gilbert\_et\_al\_2020\_fig3.png | Modified from Gilbert et al. (2020) - Fig 3. | gilbert\_et\_al\_2020 |
| rai |  | denes\_et\_al\_2015\_fig1\_clipped.png | \*\*Dénes et al. (2015) - Fig. 1.\*\* Mechanisms that cause different types of zero observations in count surveys and how species rarity, detectability and sampling effort affect them.  :::{dropdown}  (a) False zeroes are due to either imperfect detection or temporary absence. True zeroes can occur when the sample unit is unoccupied by the species, due to demographic stochasticity or due to ecological mechanisms such as unsuitable habitat or interspecific competition. (b) For common and detectable species (lower right), the majority of zeroes can be expected to result from ecological processes. As species detectability decreases, new false zeroes arise due to detection error (lower left). Species rarity results in fewer detections (dark green bars), additional true zeroes arise from unoccupied sample units (white bars) and increased demographic stochasticity (beige bars). (c) When the area sampled and/or the time of visit are small/ short relative to the species home range or movements, individuals may not be available for detection during the survey, resulting in additional false zeroes and fewer non-zero observations.  ::: | denes\_et\_al\_2015 |
| rai |  | blasco\_moreno\_2019\_fig1.png | \*\*Blasco-Moreno et al. (2019) - Fig 1.\*\* Different sources of zeros that could emerge in count data.  :::{dropdown}  The example shows the presence (>0) or absence (0) of herbivores on a plant species. Zeros due to the lack of experience of the observer (a–b) or resulting from a poor experimental design (c–h) are called False Zeros and should be minimized when performing the experiment. Structural Zeros, that is, zeros related to the ecological system under study (i–k), and Random Zeros emerging from the sampling variability (l) are known as True Zeros. Classifying a zero as a design error or structural zero depends on whether the event is part of the hypotheses tested. Only when the study includes the possibility of a zero value as part of the hypotheses (e.g. the study aims to test whether the interaction is occurring) the resulting zeros would be structural and should be included in the statistical analysis. The following text explains different scenarios that would result in a zero value, and, in brackets, how errors due to false zeros can be minimized: (a) the insects or the damage exerted are so small that the observer cannot detect them [sample when the insects are expected to be well developed]; (b) the observer does not see the herbivore (e.g. it is mistaken for a seed) or the damage is associated to other causes not related to herbivory (e.g. mechanical damage during sampling, pathogens, etc.) [the observer should be trained properly]; (c) the distributional areas of herbivores and plants are not coincident [know the species distribution before sampling]; (d) a herbivore is not present in a certain location within its distributional area, for example due to the microclimatic conditions [sample in habitats with adequate environmental conditions for a herbivore, or perform replicate surveys in different areas]; (e) a single survey is conducted, and is not coincident with the herbivore phenology [know the herbivore life cycle or perform long‐term surveys]; (f) a long‐term survey is conducted, but the low sampling frequency does not enable capture of the presence of the herbivore [sample on a more frequent basis]; (g) herbivores are not found because they are absent at the time of sampling [record plant damage instead of the presence of insects]; (h) herbivores are so infrequent that the design cannot capture their presence [perform extensive sampling with a high number of replicates]; (i) phenology of plants and herbivores are not completely coincident at a temporal level; (j) herbivores do not recognize a plant as a potential host; (k) herbivores recognize a plant as a host but prefer to feed on another species and (l) the herbivore population is not large enough to saturate the available plant resources.  ::: | blasco\_moreno\_et\_al\_2019 |
| rai |  | denes\_et\_al\_2015\_fig2\_clipped.png | \*\* Denes et al. (2015) - Fig. 2\*\* Summary of the main modelling approaches for estimating abundance of unmarked animal populations described in the text.  :::{dropdown}  Red boxes represent important model assumptions (in bold) and sampling design requirements (in italic), green boxes represent the types of input data used by each model, lilac and orange ellipses represent established and emerging methods, respectively, and blue diamonds represent additional parameters estimated. w indicates models that estimate potential occupancy probability, / indicates models that estimate probability of temporary emigration from the sample unit, and q indicates models that account for correlation in detection of individuals. p is site-level detection probability, c and x are arrival rate and survival probability parameters, respectively, r is the spatial correlation in counts, and Ω is the probability that a species is present in the supercommunity.  ::: | denes\_et\_al\_2015 |
| rai | TheWorldOfZeroInflatedModels | zi\_models\_mindmap.png | https://www.mdpi.com/2673-4591/39/1/38 | figure5\_ref\_id |
|  |  | clarke\_et\_al\_2023\_eqn\_rai1.png |  | clarke\_et\_al\_2023 |
|  |  | clarke\_et\_al\_2023\_eqn\_rai2.png |  | clarke\_et\_al\_2023 |
|  |  | denes\_et\_al\_2015\_fig2\_clipped.png | figure8\_caption | denes\_et\_al\_2015 |
|  |  | figure9\_filename.png | figure9\_caption | figure9\_ref\_id |
|  |  | figure10\_filename.png | figure10\_caption | figure10\_ref\_id |
|  |  | figure11\_filename.png | figure11\_caption | figure11\_ref\_id |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  | Engineering Proceedings | Free Full-Text | Modelling of Leishmaniasis  Infection Dynamics: A Comparative Time Series Analysis with VAR, VECM,  Generalized Linear and Markov Switching Models | zi\_models\_unknown1.png | figure4\_caption |  |

## Video

|  |  |  |  |
| --- | --- | --- | --- |
|  | caption | URL (no < / > before/after URL | ref\_id |
|  | Using Hurdle Models to Analyze Zero-Inflated Count Data | https://www.youtube.com/embed/CvM6j8hE8lE?si=E\_kNQm9YYwgUECM3 | dectre\_accel\_2016 |
|  | Hurdle models | https://www.youtube.com/embed/q2NRQBcihQY?si=r9NeRpegoj47uVn- | cao\_2021 |
| zip | Zero-inflated Poisson (ZIP) regression | https://www.youtube.com/embed/ztNQvAabgtU?si=9rY7DVbBWN\_ByPvf | tilestats\_2021 |
|  | Poisson Regression Review | https://www.youtube.com/embed/A8H6gc9Eq0w?si=Ade-D5-J5JtCZwil | marinstats\_2020a |
|  | Poisson Regression: Zero Inflation (Excessive Zeros) | https://www.youtube.com/embed/eIY--zc5f24?si=dgtSF\_vfwUxuG1Mi | marinstats\_2020b |
|  | Fitting Poisson and zero-inflated Poisson models. | vid6\_url | russel\_2020 |
|  | vid7\_caption | vid7\_url | vid7\_ref\_id |
|  | vid8\_caption | vid8\_url | vid8\_ref\_id |
|  | vid9\_caption | vid9\_url | vid9\_ref\_id |
|  | vid10\_caption | vid10\_url | vid10\_ref\_id |
|  | vid11\_caption | vid11\_url | vid11\_ref\_id |
|  | vid12\_caption | vid12\_url | vid12\_ref\_id |

SEE - <https://www.youtube.com/@marinstatlectures>

**\*\*Intensity of use\*\*:** the expected number of use events of a specific resource unit during a unit of time' (i.e., "how frequently a particular resource unit is used") (Keim et al., 2019). "Intensity of use differs from probability of occupancy, selection or use, which can remain constant even when the intensity of use varies' (Keim, DeWitt, & Lele, 2011; Lele et al., 2013).

**\*\*Probability of use\*\*:** The probability of at least one, use event of that resource unit during a unit of time' (i.e., would a particular resource unit be used at least once) (Keim et al., 2019).

## Shiny

Shiny name = Microbiome Analysis: Relative Abundance Boxplots

Shiny caption = A **Shiny app** allows you to visualize data by using R scripts without having to interact with the R script itself. This Shiny app will allow you to plot your **Relative Abundance** microbiome data in an easy-to-view format. If this is your first time utilizing this Shiny app, follow the step below to start visualising your data now!

Shiny URL = https://guthub.org/shiny/sample-apps/absboxplot/

Shiny name = shiny\_name2

Shiny caption =shiny\_caption2

Shiny URL = shiny\_url2

## Analytical tools & resources

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| note | Type | Name | Note | URL | ref\_id |
|  | R resource | abmi.camera.extras: Animal Density from Camera Data > Probabilistic gaps |  | Main resource page: <https://mabecker89.github.io/abmi.camera.extras/index.html>;<br>[Probabilistic gaps]<https://mabecker89.github.io/abmi.camera.extras/articles/gaps.html> | becker\_et\_al\_2020 |
|  | App/Program | Introduction to Camera Trap Data Management and Analysis in R > Chapter 12 Activity |  | <https://bookdown.org/c\_w\_beirne/wildCo-Data-Analysis/activity.html> | wildco\_lab\_2021d |
|  | App/Program | R package "activity" | Provides functions to express clock time data relative to anchor points (typically solar); fit kernel density functions to animal activity time data; plot activity distributions; quantify overall levels of activity; statistically compare activity metrics through bootstrapping; evaluate variation in linear variables with time (or other circular variables). | <https://cran.r-project.org/web/packages/activity/index.html> | rowcliffe\_2023 |
|  | R package | R package "overlap" | Estimates of Coefficient of Overlapping for Animal Activity Patterns | <https://cran.r-project.org/web/packages/overlap/index.html> | campbell\_2024 |
|  | Tutorial | Chapter 6 Modeling Relative Abundance |  | <https://cornelllabofornithology.github.io/ebird-best-practices/abundance.html> | strimasmackey\_et\_al\_2023 |
| zip | R package | glmmTMB: Generalized Linear Mixed Models using Template Model Builder | resource6\_note | <https://cran.r-project.org/web/packages/glmmTMB/index.html> | brooks\_et\_al\_2017  note: later, include under ZIP |
| zip | R package | R package "zicounts" | Counts data models: zero-inflation as well as interval icensored | <https://github.com/cran/zicounts> | resource7\_ref\_id  note: later, include under ZIP under ZIP |
| zip | R package | R package "DHARMa" | Can be used to assess goodness-of-fit of a mixed effect model via quantile–quantile (Q–Q) plots of standardized residuals  DHarma  we have made extensive simulations, which have shown that the various tests have certain advantages and disadvantages. The basic results are that:   * The most powerful and reliable test is option 3, but this costs a lot of time and is not available for all regression packages, as it requires that Pearson residuals are available * Option 2, the parametric Pearson-chi2 is fast if Pearson residuals are available, but based on a naive expectation of df (counts RE as 1 df) and the test statistic is thus biased towards underdispersion for mixed models. Similar to the df approximation, Bias increasing with the number of RE levels. When testing only for overdispersion (alternative = "greater"), this makes the test more conservative, but it also costs power. * The DHARMa default option 1 is fast, nearly unbiased (i.e. you can test under and overdispersion), and only slightly less powerful as test 3, PROVIDED that simulations are made conditional on the fitted REs. Note that the latter is not the DHARMa default, so you have to actively request conditional simulations, e.g. for lme4 by specifying re.form = NULL. Power compared to the parametric Pearson-chi2 test depends on the number of RE levels, it will be more powerful for typical number of RE levels. | <https://CRAN.R-project.org/package=DHARMa> | hartig\_2019 |
| hurdle | R package | R package "Pscl" | resource9\_note | < https://cran.r-project.org/web/packages/pscl/index.html> | jackman\_2024 |
| hurdle | R package | R package "countreg" | Can be used to assess goodness-of-fit of a mixed effect hurdle model via rootograms ({{ rtxt\_kleiber\_zeileis\_2016 }}) | <https://rdrr.io/rforge/countreg/><br>  https://rdrr.io/rforge/countreg/f/inst/doc/countreg.pdf><br><https://www.zeileis.org/papers/Kleiber+Zeileis-2016.pdf> | zeileis\_et\_al\_2008 |
| hurdle | resource11\_type | A guide to modeling outcomes that have lots of zeros with Bayesian hurdle lognormal and hurdle Gaussian regression models | resource11\_note  but it cannot model correlation within individuals (sites) if they are sampled repeatedly (random effects) | <https://www.andrewheiss.com/blog/2022/05/09/hurdle-lognormal-gaussian-brms> | resource11\_ref\_id |
|  | resource12\_type | resource12\_name | resource12\_note | resource12\_url | resource12\_ref\_id |
|  | resource13\_type | resource13\_name | resource13\_note | resource13\_url | resource13\_ref\_id |
|  | resource14\_type | resource14\_name | resource14\_note | resource14\_url | resource14\_ref\_id |
|  | resource15\_type | resource15\_name | resource15\_note | resource15\_url | resource15\_ref\_id |
|  | resource16\_type | resource16\_name | resource16\_note | resource16\_url | resource16\_ref\_id |
|  | resource17\_type | resource17\_name | resource17\_note | resource17\_url | resource17\_ref\_id |
|  | resource18\_type | resource18\_name | resource18\_note | resource18\_url | resource18\_ref\_id |
|  | resource19\_type | resource19\_name | resource19\_note | resource19\_url | resource19\_ref\_id |
|  | resource20\_type | resource20\_name | resource20\_note | resource20\_url | resource20\_ref\_id |
|  | resource21\_type | resource21\_name | resource21\_note | resource21\_url | resource21\_ref\_id |
|  | resource22\_type | resource22\_name | resource22\_note | resource22\_url | resource22\_ref\_id |
|  | resource23\_type | resource23\_name | resource23\_note | resource23\_url | resource23\_ref\_id |
|  | resource24\_type | resource24\_name | resource24\_note | resource24\_url | resource24\_ref\_id |
|  | resource25\_type | resource25\_name | resource25\_note | resource25\_url | resource25\_ref\_id |

## References / Glossary

|  |  |
| --- | --- |
| ref\_id |  |
| {{ rbib\_anderson\_2001 }}  {{ rbib\_banksleite\_2014 }}  {{ rbib\_blasco\_moreno\_et\_al\_2019 }}  {{ rbib\_brennan\_2019 }}  {{ rbib\_broadley\_et\_al\_2019 }}  {{ rbib\_burton\_et\_al\_2015 }}  {{ rbib\_carbone\_et\_al\_2001 }}  {{ rbib\_cao\_2021 }}  {{ rbib\_clark\_et\_al\_2003 }}  {{ rbib\_dectre\_accel\_2016 }}  {{ rbib\_denes\_et\_al\_2015 }}  {{ rbib\_gilbert\_et\_al\_2020 }}  {{ rbib\_hartig\_2019 }}  {{ rbib\_heilbron\_1994 }}  {{ rbib\_kleiber\_zeileis\_2016 }}  {{ rbib\_krebs\_et\_al\_1987 }}  {{ rbib\_johnson\_2008 }}  {{ rbib\_marinstats\_2020a }}  {{ rbib\_marinstats\_2020b }}  {{ rbib\_markle\_et\_al\_2020 }}  {{ rbib\_martin\_et\_al\_2005 }}  {{ rbib\_mullahy\_1986 }}  {{ rbib\_obrien\_2011 }}  {{ rbib\_palmer\_et\_al\_2018 }}  {{ rbib\_rovero\_marshall\_2009 }}  {{ rbib\_russel\_2020 }}  {{ rbib\_sollmann\_et\_al\_2013b }}  {{ rbib\_thompson\_et\_al\_1998 }}  {{ rbib\_tilestats\_2021 }}  {{ rbib\_villette\_et\_al\_2016 }}  {{ rbib\_welsh\_et\_al\_2000 }}  {{ rbib\_zeileis\_et\_al\_2008 }}  +check others |  |

## Notes

|  |  |  |  |
| --- | --- | --- | --- |
|  | Assumptions | Pros | Cons |
| hurdle | Independence of observations, error distributions, and linearity (Markle\_et\_al\_2020) |  |  |

### Zero-inflation / Overdispersion

* Martin et al. 2005 Zero Tolerance Ecology Improving Ecological Inference by Modelling the Source of Zero Observations.pdf
* BlascoMoreno et al. 2019 What Does a Zero Mean Understanding False, Random and Structural Zeros in Ecology.pdf
* Campbell et al. 2021 The Consequences of Checking for Zero Inflation and Overdispersion in the Analysis of Count Data.pdf

(#i\_mod\_rai\_poisson)=

### # Poisson model

(#i\_mod\_rai\_nb)=

### # Negative binomial (NB) model

(#i\_mod\_rai\_zip)=

### # Zero-inflated poisson (ZIP)

* "Zero-inflated Poisson models (hereafter ZIP; Lambert 1992) fit two linear models simultaneously to address the sources of zero-inflation (probability of a zero value; inversely, probability of occurrence) separately from zeros that belong to the count distribution (Lambert 1992; Blasco‐Moreno\_et\_al\_2019). The zero-inflation model-part is a logistic regression, and since counts are repeatedly measured over time, the binary response serves as a detection history (Dénes\_et\_al\_2015). Thus, the zero-inflation model-part can include predictors of "false" zeros (e.g., imperfect detection) and "true" zeros that relate to occurrence (Lambert 1992; Blasco-Moreno\_et\_al\_2019). Additionally, this model-part can also mitigate problematic variability from other "false" zeros (e.g., not present but generally occurs, or uses the feature, but not at that site). The second model-part is a Poisson model used to assess predictors of the counts (relative abundance given presence [Lambert 1992; Wenger and Freeman 2008])."

(#i\_mod\_rai\_zinb)=

### # Zero-inflated negative binomial (ZINB)

(#i\_mod\_rai\_hurdle)=

### # Hurdle model

* Hurdle models differ from traditional linear models in that two linear models are fitted simultaneously to count data.---(Markle\_et\_al\_2020)
  + Hurdle models are used to address overdispersion in count data by fitting two linear models that assess the sources of abundant zeros and positive counts, simultaneously (Mullahy, 1986)(McCullagh & Nelder 1989; Hinde & Deme´trio 1998; Poortema 1999).---(Martin, 2005)
* "Hurdle models suggest a two-part process. The first part induces an event, and once the hurdle to the first event has been cleared, the second part determines the number of subsequent events."---(Martin, 2005)
  + "The first of these was a logistic linear model (can be used to model prevalence) and the second was a truncated Poisson or negative binomial linear model." ---(Markle\_et\_al\_2020)
    - "The first model is a logistic regression used to evaluate what factors influence whether a zero value occurs (i.e. for a given sample camera month, does a visit or not (Mullahy, 1986)."---(Martin, 2005)
  + If a positive value does occur, the ‘hurdle’ is cleared, and those non-zero values are considered in the second linear model as a truncated Poisson to determine what factors influence the level of positive values {Mullahy, 1986 #900}.---(Martin, 2005)
* The distribution of a hurdle model is best characterized by the zero-truncated "version of the parent distribution" {Mullahy, 1986 #900}(rather than positive values ).---(Mullahy, 1986)
* Similar to most linear models, hurdle models have assumptions regarding independence of observations, error distributions, and linearity.---(Markle\_et\_al\_2020)

### Model evaluation & fit

* You can evaluate autocorrelation in a hurdle model by evaluating residual plots of model fit via the R package pscl (Zeileis\_et\_al\_2008)---(Markle\_et\_al\_2020)
* If autocorrelation is present, you might consider including random effects (mixed-effects hurdle model) (Molenberghs and Verbeke, 2005)---(Markle\_et\_al\_2020)
* Goodness-of-fit of a mixed effect hurdle model can be assessed using rootograms (Kleiber and Zeileis, 2016) implemented in R package countreg (Zeileis\_et\_al\_2008), and quantile–quantile (Q–Q) plots of standardized residuals using R package DHARMa (Hartig).

### Hurdle models vs Zero-inflated Poisson models (ZIP) regression models

* "Hurdle models are very similar to zero-inflated Poisson models (ZIP) regression models, except that in hurdle models, the distribution is dependant only on the truncated values (Mullahy, 1986), whereas, in a standard ZIP, the distribution of the positive values also relates to the probability of a zero outcome (Mullahy, 1986)." ---(Martin, 2005)
* "Standard zero-inflated Poisson models are useful to differentiate between zeros that occur because of some other process or as a result of sampling error." ---(Martin, 2005)

"Whereas hurdle models can more useful when only true sampling zeros are present and….In that case, the intersection of zero-values and non-zero values represent an inflection point between two processes." ---(Martin, 2005)

* **ECOL papers using hurdle refs:**The ecological literature has seen a recent upsurge of interest in techniques for dealing with excess zero values. Zero-inflated models have been applied in a range of ecological scenarios, including data sets with zero inflation caused by true zero (Welsh\_et\_al\_1996, 2000; Barry & Welsh 2002; Podlich\_et\_al\_2002; Kuhnert\_et\_al\_2005; Martin\_et\_al\_2005) and false-zero observations (Kery 2002; MacKenzie\_et\_al\_2002, 2003, 2004; Tyre\_et\_al\_2003; Wintle\_et\_al\_2004).
* …. As such, hurdle models often perform better than standard ZIP models when the ….{Martin, 2005 #913} and should be used when… {Lambert, 1992 #919}{Welsh, 2000 #920}{Martin, 2005 #913}

Compare assumption, pro, con

|  |  |  |  |
| --- | --- | --- | --- |
|  | Assumptions | Pros | Cons |
| Hurdle model | Independence of observations, error distributions, and linearity (Markle\_et\_al\_2020) |  |  |

##### POPULATE - MOD

File from = 00\_tools\00\_00\_template-master.docx

**---**jupytext:  
[[space2]]formats: md:myst

[[space2]]text\_representation:

[[space4]]extension: .md

[[space4]]format\_name: myst

[[space4]]format\_version: 0.17.2 <!--0.13-->

[[space4]]jupytext\_version: 1.16.4 <!-- 6.5.4-->  
kernelspec:  
[[space2]]display\_name: Python 3

[[space2]]language: python

[[space2]]name: python3  
editor\_options:   
[[space2]]markdown:

[[space4]]wrap: none  
---

**(i\_**mod\_rai)=  
# {{ title\_i\_mod\_rai }}

<!--

:::{hint}

replace me with text

:::

-->

**```{include} pro\_con\_**assump/mod\_rai\_apc.md  
```

**:::::::{tab-set}**

**::::::{tab-item} Overview**  
```{include} include/00\_coming\_soon.md

```  
::::::

**::::::{tab-item} In-depth**  
```{include} include/note\_adapted\_clarke\_et\_al\_2023.md

```

In ecology, relative abundance (RA) is any count of animals or animal sign (e.g., number of deer sighted, number of bird vocalizations per unit time, number of moose tracks per kilometer of transect) that is assumed to correlate with absolute abundance ({{ rtxt\_obrien\_2011 }}). RA is a controversial index for two reasons: 1) there is often no documented relationship between the number of animals or signs observed and population size (i.e., index validation), and 2) detection probability is assumed to be constant between the areas, times or species being compared ({{ rtxt\_obrien\_2011 }}; {{ rtxt\_thompson\_et\_al\_1998 }}).

To the first point: the relationship between the number of animals or signs and abundance is rarely established ({{ rtxt\_burton\_et\_al\_2015 }}). Researchers often assume that counts and population size scale linearly – but many other kinds of relationships are possible. When the assumed relationship between counts and abundance diverges from the actual relationship, inferences from RA are not very meaningful ({{ rtxt\_thompson\_et\_al\_1998 }}). Validating a count-abundance relationship requires comparison with a robust, accurate estimate of absolute density (e.g., {{ rtxt\_krebs\_et\_al\_1987}}; {{ rtxt\_rovero\_marshall\_2009 }}; {{ rtxt\_villette\_et\_al\_2016 }}).

To the second point: consider the canonical equation,

```{figure} ../03\_images/03\_image\_files/clarke\_et\_al\_2023\_eqn\_rai1.png

:width: 80px

:align: center

```

where \*𝑁\* is population size, \*𝐶\* is the count of animals or signs and \*𝑝\* is detection probability ({{ rtxt\_anderson\_2001 }}; {{ rtxt\_brennan\_2019 }}). This equation underlies many estimators of abundance, including capture-recapture (CR; see {bdg-link-primary-line}`Capture-recapture (CR) / Capture-mark-recapture (CMR)<https://ab-rcsc.github.io/rc-decision-support-tool\_concept-library/02\_dialog-boxes/03\_10\_mod\_cr\_cmr.html>`) and distance sampling (DS; see {bdg-link-primary-line}`Distance sampling<https://ab-rcsc.github.io/rc-decision-support-tool\_concept-library/02\_dialog-boxes/03\_20\_mod\_ds.html>`) methods ({{ rtxt\_obrien\_2011 }}). RA comparisons assume that detection probability \*𝑝\* is constant across space, time or species, and can therefore be ignored ({{ rtxt\_anderson\_2001 }}; {{ rtxt\_obrien\_2011 }}; {{ rtxt\_sollmann\_et\_al\_2013b }}), such that:

```{figure} ../03\_images/03\_image\_files/clarke\_et\_al\_2023\_eqn\_rai2.png

:width: 80px

:align: center

```

so count essentially becomes a surrogate for population size.

Assuming constant detection probability \*𝑝\* is problematic, since the likelihood an animal or sign is counted during a survey will vary with observational, environmental, and habitat- and species-specific factors, which in turn can vary with time ({{ rtxt\_anderson\_2001 }}). For example: at site A, animals may be difficult to spot in dense vegetation, while at site B, animals may be easy to spot in open grassland; and the effects of vegetation on observability may differ seasonally. If the effects of vegetation on detectability are not accounted for, how can we be sure that differences in animal counts at site A and B are due to true differences in abundance, and not simply artefacts of detection bias ({{ rtxt\_sollmann\_et\_al\_2013b }})?

In a camera trapping context, RA is the comparison of detection rates across space, time or species – where detection rates are typically reported as the number of images per 100 trap days, but can also be reported in terms of the total number of detections, other units of effort (e.g., camera trap hours), proportion of stations with detections, etc. ({{ rtxt\_burton\_et\_al\_2015 }}). As with other kinds of RA surveys, comparisons of camera trap detection rates can confound abundance with animal behaviour and observability ({{ rtxt\_anderson\_2001 }}; {{ rtxt\_burton\_et\_al\_2015 }}).

RA has been criticized as an abundance estimator. Anderson (2001) condemned the index as "unprofessional," while O’Brien (2011) called it a "metric of last resort." Sollmann et al. (2013b) used simulations to determine that camera trap RA analyses did not detect changes in big cat density, and called use of the index for wildlife management "alarming." Nevertheless, some researchers have had success with the method and/or have argued for its conceptual and practical advantages (e.g., {{ rtxt\_rowcliffe\_carbone\_2008 }}; {{ rtxt\_johnson\_2008 }}; {{ rtxt\_palmer\_et\_al\_2018 }}; {{ rtxt\_rovero\_marshall\_2009 }}). Broadley et al. (2019) used simulations to show that RA could be sensitive to density-dependent movement, but generally tracked abundance well. Banks-Leite (2014) emphasized the importance of careful sampling design and protocols to control for variation in detectability, arguing that researchers should not solely rely on statistical corrections.

Ultimately, there is no "silver bullet" and researchers must carefully consider their inferential objectives and potential sources of sampling and estimation bias when choosing response variables and modelling frameworks for camera trap data.

::::::

**::::::{tab-item} Visual** resources

:::::{grid} 3  
:class: wrapper

::::{grid-item-card} {{ rtxt\_gilbert\_et\_al\_2020 }}  
:::{figure} ../03\_images/03\_image\_files/gilbert\_et\_al\_2020\_fig3.png  
:class: img\_grid  
:::  
Modified from Gilbert et al. (2020) - Fig 3.  
::::

::::{grid-item-card} {{ rtxt\_denes\_et\_al\_2015 }}  
:::{figure} ../03\_images/03\_image\_files/denes\_et\_al\_2015\_fig1\_clipped.png   
:class: img\_grid  
:::  
\*\*Dénes et al. (2015) - Fig. 1.\*\* Mechanisms that cause different types of zero observations in count surveys and how species rarity, detectability and sampling effort affect them.

:::{dropdown}

(a) False zeroes are due to either imperfect detection or temporary absence. True zeroes can occur when the sample unit is unoccupied by the species, due to demographic stochasticity or due to ecological mechanisms such as unsuitable habitat or interspecific competition. (b) For common and detectable species (lower right), the majority of zeroes can be expected to result from ecological processes. As species detectability decreases, new false zeroes arise due to detection error (lower left). Species rarity results in fewer detections (dark green bars), additional true zeroes arise from unoccupied sample units (white bars) and increased demographic stochasticity (beige bars). (c) When the area sampled and/or the time of visit are small/ short relative to the species home range or movements, individuals may not be available for detection during the survey, resulting in additional false zeroes and fewer non-zero observations.

:::  
::::

::::{grid-item-card} {{ rtxt\_blasco\_moreno\_et\_al\_2019 }}  
:::{figure} ../03\_images/03\_image\_files/blasco\_moreno\_2019\_fig1.png   
:class: img\_grid  
:::  
\*\*Blasco-Moreno et al. (2019) - Fig 1.\*\* Different sources of zeros that could emerge in count data.

:::{dropdown}

The example shows the presence (>0) or absence (0) of herbivores on a plant species. Zeros due to the lack of experience of the observer (a–b) or resulting from a poor experimental design (c–h) are called False Zeros and should be minimized when performing the experiment. Structural Zeros, that is, zeros related to the ecological system under study (i–k), and Random Zeros emerging from the sampling variability (l) are known as True Zeros. Classifying a zero as a design error or structural zero depends on whether the event is part of the hypotheses tested. Only when the study includes the possibility of a zero value as part of the hypotheses (e.g. the study aims to test whether the interaction is occurring) the resulting zeros would be structural and should be included in the statistical analysis. The following text explains different scenarios that would result in a zero value, and, in brackets, how errors due to false zeros can be minimized: (a) the insects or the damage exerted are so small that the observer cannot detect them [sample when the insects are expected to be well developed]; (b) the observer does not see the herbivore (e.g. it is mistaken for a seed) or the damage is associated to other causes not related to herbivory (e.g. mechanical damage during sampling, pathogens, etc.) [the observer should be trained properly]; (c) the distributional areas of herbivores and plants are not coincident [know the species distribution before sampling]; (d) a herbivore is not present in a certain location within its distributional area, for example due to the microclimatic conditions [sample in habitats with adequate environmental conditions for a herbivore, or perform replicate surveys in different areas]; (e) a single survey is conducted, and is not coincident with the herbivore phenology [know the herbivore life cycle or perform long‐term surveys]; (f) a long‐term survey is conducted, but the low sampling frequency does not enable capture of the presence of the herbivore [sample on a more frequent basis]; (g) herbivores are not found because they are absent at the time of sampling [record plant damage instead of the presence of insects]; (h) herbivores are so infrequent that the design cannot capture their presence [perform extensive sampling with a high number of replicates]; (i) phenology of plants and herbivores are not completely coincident at a temporal level; (j) herbivores do not recognize a plant as a potential host; (k) herbivores recognize a plant as a host but prefer to feed on another species and (l) the herbivore population is not large enough to saturate the available plant resources.

:::  
::::

**::::{grid-item-card} {{** rtxt\_denes\_et\_al\_2015 }}  
:::{figure} ../03\_images/03\_image\_files/denes\_et\_al\_2015\_fig2\_clipped.png   
:class: img\_grid  
:::  
\*\* Denes et al. (2015) - Fig. 2\*\* Summary of the main modelling approaches for estimating abundance of unmarked animal populations described in the text.

:::{dropdown}

Red boxes represent important model assumptions (in bold) and sampling design requirements (in italic), green boxes represent the types of input data used by each model, lilac and orange ellipses represent established and emerging methods, respectively, and blue diamonds represent additional parameters estimated. w indicates models that estimate potential occupancy probability, / indicates models that estimate probability of temporary emigration from the sample unit, and q indicates models that account for correlation in detection of individuals. p is site-level detection probability, c and x are arrival rate and survival probability parameters, respectively, r is the spatial correlation in counts, and Ω is the probability that a species is present in the supercommunity.

:::  
::::

::::{grid-item-card} {{ rtxt\_figure5\_ref\_id }}  
:::{figure} ../03\_images/03\_image\_files/zi\_models\_mindmap.png   
:class: img\_grid  
:::  
   
::::

::::{grid-item-card} {{ rtxt\_clarke\_et\_al\_2023 }}  
:::{figure} ../03\_images/03\_image\_files/clarke\_et\_al\_2023\_eqn\_rai1.png  
:class: img\_grid  
:::  
   
::::  
:::::

**::::{grid-item-card} {{** rtxt\_dectre\_accel\_2016 }}

<div class="iframe-container-vid"><iframe class="iframe-responsive-vid" src="https://www.youtube.com/embed/CvM6j8hE8lE?si=E\_kNQm9YYwgUECM3"></iframe></div>

Using Hurdle Models to Analyze Zero-Inflated Count Data  
::::

::::{grid-item-card} {{ rtxt\_cao\_2021 }}   
<div class="iframe-container-vid"><iframe class="iframe-responsive-vid" src="https://www.youtube.com/embed/q2NRQBcihQY?si=r9NeRpegoj47uVn-"></iframe></div>

Hurdle models  
::::

::::{grid-item-card} {{ rtxt\_tilestats\_2021 }}  
<div class="iframe-container-vid"><iframe class="iframe-responsive-vid" src="https://www.youtube.com/embed/ztNQvAabgtU?si=9rY7DVbBWN\_ByPvf"></iframe></div>

Zero-inflated Poisson (ZIP) regression  
::::

**::::{grid-item-card} {{** rtxt\_marinstats\_2020a }}   
<div class="iframe-container-vid"><iframe class="iframe-responsive-vid" src="https://www.youtube.com/embed/A8H6gc9Eq0w?si=Ade-D5-J5JtCZwil"></iframe></div>

Poisson Regression Review  
::::

::::{grid-item-card} {{ rtxt\_marinstats\_2020b }}  
<div class="iframe-container-vid"><iframe class="iframe-responsive-vid" src="https://www.youtube.com/embed/eIY--zc5f24?si=dgtSF\_vfwUxuG1Mi"></iframe></div>

Poisson Regression: Zero Inflation (Excessive Zeros)  
::::

::::{grid-item-card} {{ rtxt\_russel\_2020 }}

<div class="iframe-container-vid"><iframe class="iframe-responsive-vid" src="vid6\_url"></iframe></div>

Fitting Poisson and zero-inflated Poisson models.  
::::

**::::::**

::::::{tab-item} Shiny apps/Widgets  
:::::{card} Microbiome Analysis: Relative Abundance Boxplots  
A Shiny app allows you to visualize data by using R scripts without having to interact with the R script itself. This Shiny app will allow you to plot your **Relative Abundance** microbiome data in an easy-to-view format. If this is your first time utilizing this Shiny app, follow the step below to start visualising your data now!

<div class="iframe-container-shiny"><iframe class="iframe-responsive-shiny" src="https://guthub.org/shiny/sample-apps/absboxplot/ "></iframe></div>   
:::**::**  
::::::

**:::::{tab-item} Analytical** tools & Resources  
| Type | Name | Note | URL |Reference |  
|:----------------|:-------------------------------|:----------------------------------------------------------------|:----------------------|:----------------------------------------|  
| R resource | abmi.camera.extras: Animal Density from Camera Data > Probabilistic gaps | | Main resource page: <https://mabecker89.github.io/abmi.camera.extras/index.html>;<br>[Probabilistic gaps]<https://mabecker89.github.io/abmi.camera.extras/articles/gaps.html> | {{ rbib\_becker\_et\_al\_2020 }} |  
| App/Program | Introduction to Camera Trap Data Management and Analysis in R > Chapter 12 Activity | | <https://bookdown.org/c\_w\_beirne/wildCo-Data-Analysis/activity.html> | {{ rbib\_wildco\_lab\_2021d }} |  
| App/Program | R package "activity" | Provides functions to express clock time data relative to anchor points (typically solar); fit kernel density functions to animal activity time data; plot activity distributions; quantify overall levels of activity; statistically compare activity metrics through bootstrapping; evaluate variation in linear variables with time (or other circular variables). | <https://cran.r-project.org/web/packages/activity/index.html> | {{ rbib\_rowcliffe\_2023 }} |  
| R package | R package "overlap" | Estimates of Coefficient of Overlapping for Animal Activity Patterns | <https://cran.r-project.org/web/packages/overlap/index.html> | {{ rbib\_campbell\_2024 }} |  
| Tutorial | Chapter 6 Modeling Relative Abundance | | <https://cornelllabofornithology.github.io/ebird-best-practices/abundance.html> | {{ rbib\_strimasmackey\_et\_al\_2023 }} |  
| R package | glmmTMB: Generalized Linear Mixed Models using Template Model Builder | resource6\_note | <https://cran.r-project.org/web/packages/glmmTMB/index.html> | {{ rbib\_brooks\_et\_al\_2017 }} |  
| R package | R package "zicounts" | Counts data models: zero-inflation as well as interval icensored | <https://github.com/cran/zicounts> | {{ rbib\_resource7\_ref\_id }} |  
| R package | R package "DHARMa" | Can be used to assess goodness-of-fit of a mixed effect model via quantile–quantile (Q–Q) plots of standardized residuals | <https://CRAN.R-project.org/package=DHARMa> | {{ rbib\_hartig\_2019}} |  
| R package | R package "Pscl" | resource9\_note | < https://cran.r-project.org/web/packages/pscl/index.html> | {{ rbib\_jackman\_2024 }} |  
| R package | R package "countreg" | Can be used to assess goodness-of-fit of a mixed effect hurdle model via rootograms ({{ rtxt\_kleiber\_zeileis\_2016 }}) | <https://rdrr.io/rforge/countreg/><br>

https://rdrr.io/rforge/countreg/f/inst/doc/countreg.pdf><br><https://www.zeileis.org/papers/Kleiber+Zeileis-2016.pdf> | {{ rbib\_zeileis\_et\_al\_2008 }} |  
| resource11\_type | A guide to modeling outcomes that have lots of zeros with Bayesian hurdle lognormal and hurdle Gaussian regression models | resource11\_note | <https://www.andrewheiss.com/blog/2022/05/09/hurdle-lognormal-gaussian-brms> | {{ rbib\_resource11\_ref\_id }} |  
| resource12\_type | resource12\_name | resource12\_note | resource12\_url | {{ rbib\_resource12\_ref\_id }} |  
| resource13\_type | resource13\_name | resource13\_note | resource13\_url | {{ rbib\_resource13\_ref\_id }} |  
| resource14\_type | resource14\_name | resource14\_note | resource14\_url | {{ rbib\_resource14\_ref\_id }} |  
| resource15\_type | resource15\_name | resource15\_note | resource15\_url | {{ rbib\_resource15\_ref\_id }} |  
| resource16\_type | resource16\_name | resource16\_note | resource16\_url | {{ rbib\_resource16\_ref\_id }} |  
| resource17\_type | resource17\_name | resource17\_note | resource17\_url | {{ rbib\_resource17\_ref\_id }} |  
| resource18\_type | resource18\_name | resource18\_note | resource18\_url | {{ rbib\_resource18\_ref\_id }} |  
| resource19\_type | resource19\_name | resource19\_note | resource19\_url | {{ rbib\_resource19\_ref\_id }} |  
| resource20\_type | resource20\_name | resource20\_note | resource20\_url | {{ rbib\_resource20\_ref\_id }} |

<!-- END\_RESOURCE\_TABLE -->  
::::::

**::::::{tab-item} References**  
{{ rbib\_anderson\_2001 }}

{{ rbib\_banksleite\_2014 }}

{{ rbib\_blasco\_moreno\_et\_al\_2019 }}

{{ rbib\_brennan\_2019 }}

{{ rbib\_broadley\_et\_al\_2019 }}

{{ rbib\_burton\_et\_al\_2015 }}

{{ rbib\_carbone\_et\_al\_2001 }}

{{ rbib\_cao\_2021 }}

{{ rbib\_clark\_et\_al\_2003 }}

{{ rbib\_dectre\_accel\_2016 }}

{{ rbib\_denes\_et\_al\_2015 }}

{{ rbib\_gilbert\_et\_al\_2020 }}

{{ rbib\_hartig\_2019 }}

{{ rbib\_heilbron\_1994 }}

{{ rbib\_kleiber\_zeileis\_2016 }}

{{ rbib\_krebs\_et\_al\_1987 }}

{{ rbib\_johnson\_2008 }}

{{ rbib\_marinstats\_2020a }}

{{ rbib\_marinstats\_2020b }}

{{ rbib\_markle\_et\_al\_2020 }}

{{ rbib\_martin\_et\_al\_2005 }}

{{ rbib\_mullahy\_1986 }}

{{ rbib\_obrien\_2011 }}

{{ rbib\_palmer\_et\_al\_2018 }}

{{ rbib\_rovero\_marshall\_2009 }}

{{ rbib\_russel\_2020 }}

{{ rbib\_sollmann\_et\_al\_2013b }}

{{ rbib\_thompson\_et\_al\_1998 }}

{{ rbib\_tilestats\_2021 }}

{{ rbib\_villette\_et\_al\_2016 }}

{{ rbib\_welsh\_et\_al\_2000 }}

{{ rbib\_zeileis\_et\_al\_2008 }}

+check others   
::::::

**:::::::**