Design and Analysis of Occupancy Studies Part 1b

Carl James Schwarz

Department of Statistics and Actuarial Science
Simon Fraser University
Burnaby, BC, Canada
cschwarz @ stat.sfu.ca

Outline I

Single Species; Single-Season - Covariates

Covariates

Single Species; Single-Season - Covariates

Covariates can be used for:

- Compare occupancy probabilities among habitat types. For example, is the occupancy in spruce sites different than the occupancy in lodgepole pine units.
- Reduce heterogeneity. For example, detectability may differ between spruce sites and lodgepole sites.

Single Species; Single-Season - Covariates I

Classes of covariates:

- Site covariates that don't change over a season. For example, habitat type or the area would be measured on each site. No missing values allowed for site-level covariates (why?)
- Visit (survey) covariates (typically external to the study) that affect all sites simultaneously on a particular visit. For example, rain on a visit may reduce detectability even though effort is same at all visits. No missing values allowed for external covariates (why?).

Single Species; Single-Season - Covariates II

- Sampling (visit $\times$ site) covariate measured at each visit at each site. For example, the observer who worked that site at this visit Missing values allowed ONLY if a site is not surveyed on a visit. If a site is measured on a visit, the sampling covariates must be available.

The covariates operate on the column (external/visit covariates), rows (site covariates), or cell (sampling covariates) in the detection matrix.

Single Species; Single-Season - Covariates

Two types of covariates:

- Continuous, e.g. occupancy is a function of stem density. Use the value directly as the covariate.
- Discrete, e.g. occupancy in spruce or lodgepole pine.
  - Create categorical variable for category membership (Recommended if possible, e.g. in *RPresence* or *RMark*.)
  - Create indicator variables (0,1) for category membership (necessary in *PRESENCE* or *MARK*).
    If a covariate has K categories, you will need to create $K-1$ indicator variables. (Not recommended in *RPresence* or *RMark*)

Single Species; Single-Season - Covariates

Modeling covariate effects is done using the *logit* (also known as the *log-odds*) link.
If we model $\psi = \beta_0 + \beta_1 x$, then as x varies, the predicted probability can be < 0 or > 1 which is non-sensical.
Consequently, we model the effects of covariates on the *logit* scale

$$logit \psi = \log \frac{\psi}{1-\psi} = \beta_0 + \beta_1 x$$

where log() is the natural (to the base e) logarithm.
The inverse transform is:

$$\psi = \frac{1}{1 + e^{-logit}}$$

Single Species; Single-Season - Covariates

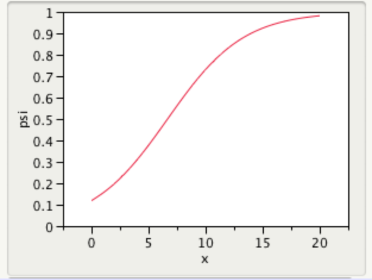
Probability	logit()
.01	-4.59
.10	-2.20
.4	-0.41
.5	0
.6	0.41
.9	2.20
.99	4.59

Notice the symmetry between a probability of 0.01 and 0.99, etc.

Single Species; Single-Season - Covariates

Example of relationship

$$logit(\psi) = -2 + 0.3x$$



The fitted curve never goes below 0 or above 1. It is approximately linear in the central range of the covariate. This relationship can be applied to detectability or occupancy and can include more than one covariate.

Covariates and Design Matrices

Every covariate creates a design matrix that links the covariate values to the parameter.

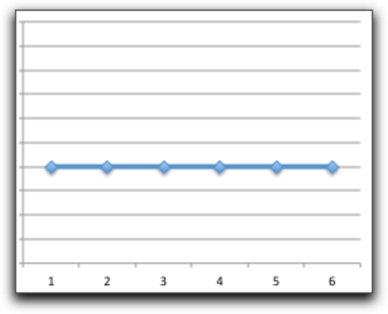
- Intercept (typically the first column) is the baseline.
- Categorical covariates create columns of $1/0$ with $K-1$ columns for K levels.
- Continuous covariates create columns with the covariate value.

Alternate design matrices are possible.

The design matrices are typically hidden from the user when using *RPresence* or *RMark*.

Different Design Matrices

H = indicator of habitat type (2 levels, 0 or 1); modelling detectability over multiple surveys.

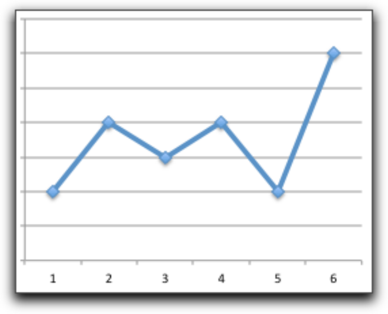


Model $p(*)$.

Index	Design matrix
1	1
2	1
3	1
4	1
5	1
6	1

Different Design Matrices

H = indicator of habitat type (2 levels, 0 or 1); modelling detectability over multiple surveys.

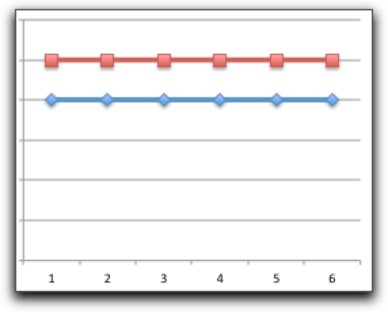


Model $p(t)$.

Index	Design matrix					
1	1	0	0	0	0	0
2	0	1	0	0	0	0
3	0	0	1	0	0	0
4	0	0	0	1	0	0
5	0	0	0	0	1	0
6	0	0	0	0	0	1

Different Design Matrices

H = indicator of habitat type (2 levels, 0 or 1); modelling detectability over multiple surveys.

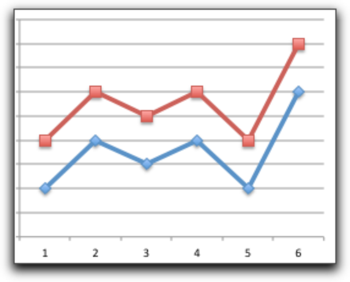


Model $p(H)$.

Index	Design matrix	
1	1	H
2	1	H
3	1	H
4	1	H
5	1	H
6	1	H

Different Design Matrices

H = indicator of habitat type (2 levels, 0 or 1); modelling detectability over multiple surveys.

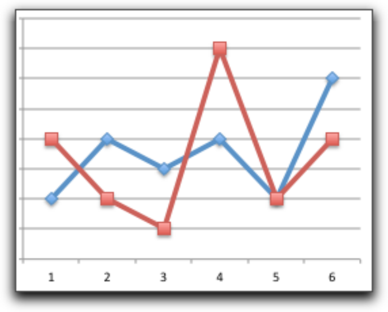


Model $p(t + H)$.

Index	Design matrix					
1	1	0	0	0	0	0	H
2	0	1	0	0	0	0	H
3	0	0	1	0	0	0	H
4	0	0	0	1	0	0	H
5	0	0	0	0	1	0	H
6	0	0	0	0	0	1	H

Different Design Matrices

H = indicator of habitat type (2 levels, 0 or 1); modelling detectability over multiple surveys.

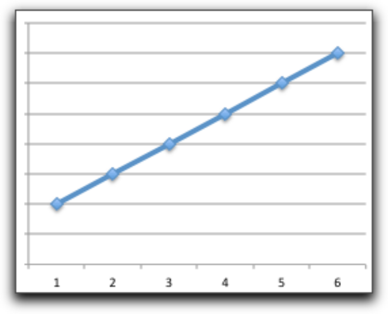


Model $p(t * H)$.

Index	Design matrix							
1	1	0	0	0	0	0	H	0
2	0	1	0	0	0	0	0	H
3	0	0	1	0	0	0	0	0
4	0	0	0	1	0	0	0	0
5	0	0	0	0	1	0	0	0
6	0	0	0	0	0	1	0	0

Different Design Matrices

H = indicator of habitat type (2 levels, 0 or 1); modelling detectability over multiple surveys.

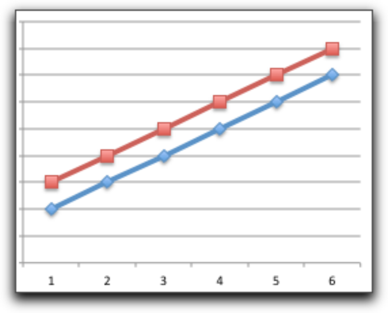


Model $p(Linear)$.

Index	Design matrix
1	1
2	2
3	3
4	4
5	5
6	6

Different Design Matrices

H = indicator of habitat type (2 levels, 0 or 1); modelling detectability over multiple surveys.

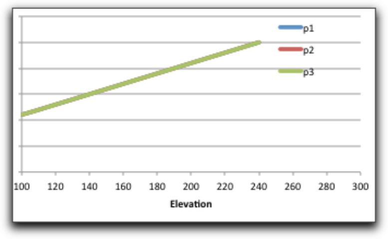


Model $p(Linear + H)$.

Index	Design matrix	
1	1	H
2	2	H
3	3	H
4	4	H
5	5	H
6	6	H

Different Design Matrices

EL =continuous covariate (elevation); modelling detectability as elevation changes

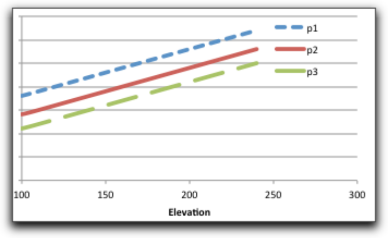


Index	Design matrix	
1	1	EL_1
2	1	EL_2
3	1	EL_3
4	1	EL_4
5	1	EL_5
6	1	EL_6

Model $p(Elev)$ – Lines are co-incident.

Different Design Matrices

EL =continuous covariate (elevation); modelling detectability as elevation changes

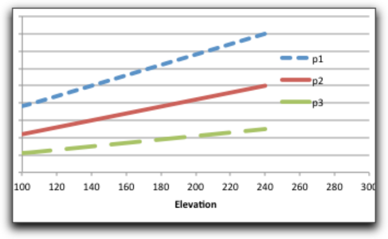


Model $p(t + ELEV)$.

Index	Design matrix						
1	1	0	0	0	0	0	EL_1
2	0	1	0	0	0	0	EL_2
3	0	0	1	0	0	0	EL_3
4	0	0	0	1	0	0	EL_4
5	0	0	0	0	1	0	EL_5
6	0	0	0	0	0	1	EL_6

Different Design Matrices

EL =continuous covariate (elevation); modelling detectability as elevation changes



Model $p(t * Elev)$.

Index	Design matrix					
1	1	0	0	0	EL_1	0
2	0	1	0	0	0	EL_2
3	0	0	1	0	0	0
4	0	0	0	1	0	0

Single Species; Single-Season

Single-Species Single-Season Occupancy Studies

Covariates and *PRESENCE*

Single Species; Single-Season - Covariates - PRESENCE

Mahoenui giant weta (*Deinacrida mahoenui*) is endemic to New Zealand and under stress from rats and other predators.
72 circular plots (3 m radius, primarily prickly gorse plants) were surveyed for weta.
Each plot surveyed 3-5 times.
Covariates to be considered:

- Observer. Three different observers and not every plot surveyed by each observer.
- Browse. Was each site browsed by goats, yes or no.

Single Species; Single-Season - Covariates - PRESENCE

Open the Weta file and look at how the data have been entered. Detection histories include many missing values. Are these MCAR?

0	0	0	0	-
0	0	0	0	-
0	0	0	1	-
0	0	0	0	-
0	0	0	0	-
0	0	0	0	-
0	0	0	0	-
0	0	0	0	-
0	0	0	0	-
1	1	0	0	-
0	0		1	0

Single Species; Single-Season - Covariates - PRESENCE

Browse covariate entered as two indicator variables (this has implications later on). No missing values allowed in site-level covariate.

Browsed	Unbrowsed
1	0
1	0
1	0
0	1
1	0
0	1
0	1
0	1

Single Species; Single-Season - Covariates - PRESENCE

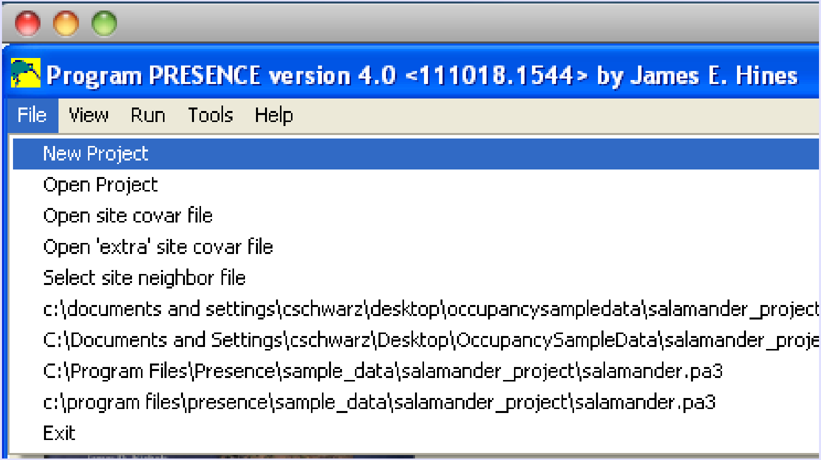
Observer covariate entered as three sets of indicator variables that span all five sampling occasions (why?), and has many missing values (why?) (this has implications later on).

Observer 1 covariate values (other observers similar):

1	0	0	0	-
1	0	0	0	-
1	0	0	0	-
1	0	0	0	-
1	0	0	0	-
1	0	0	0	-
1	0	0	0	-
1	0	0	0	-
1	0	0	0	-

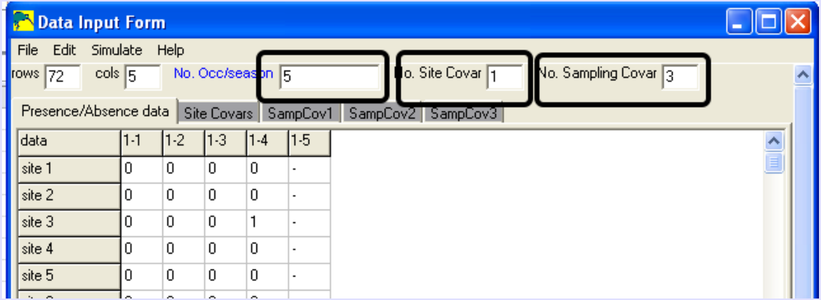
Single Species; Single-Season - Covariates - PRESENCE

Start PRESENCE and start a new project:



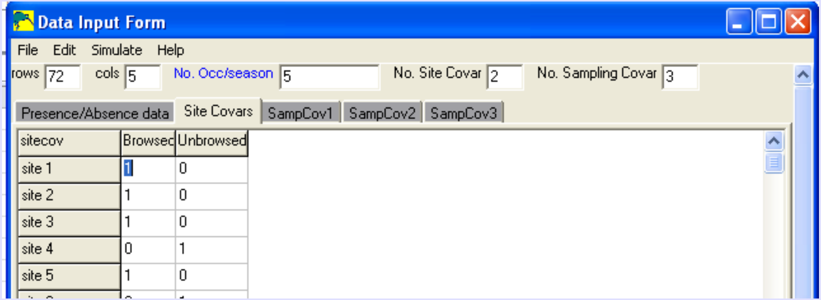
Single Species; Single-Season - Covariates - PRESENCE

Cut and paste the detection histories and adjust the number of occasions/seasons, number of site covariates, and number of sampling covariates:



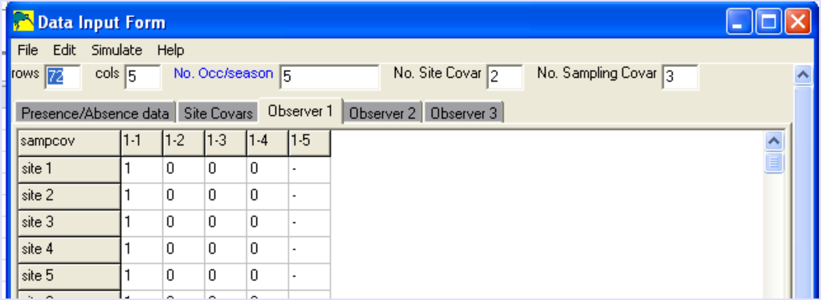
Single Species; Single-Season - Covariates - PRESENCE

Cut and paste the site covariate. Because these are FIXED for the study, there is only one tab.



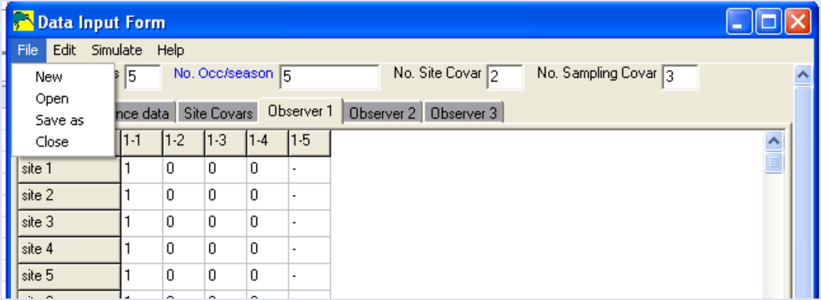
Single Species; Single-Season - Covariates - PRESENCE

Cut and paste the three observer covariate sets. Because these vary by occassion and by site, you need a separate covariate set for each observer. Rename the covariate (look under the Edit menu item).



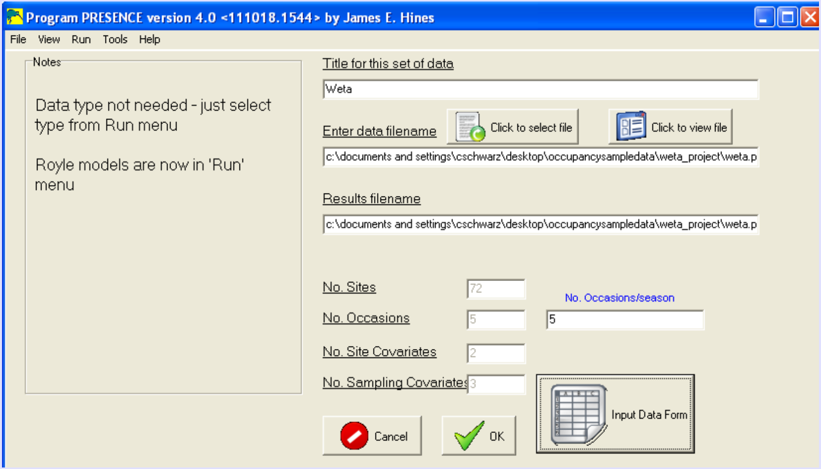
Single Species; Single-Season - Covariates - PRESENCE

Create a new folder, and save the detection histories and covarariates into *.pao files



Single Species; Single-Season - Covariates - PRESENCE

Check that you are ready to go:



Single Species; Single-Season - Covariates - PRESENCE

Fit the $\psi(*), p(*)$ model and look at estimates.

PRESENCE:c:\documents and settings\cschwarz\desktop\occupancysampledata\weta_project\weta_project\weta.pa3

File View Run Tools Help

Model	AIC	deltaAIC	AIC wgt	Model Likelih	no.Par.	-2*LogLike
1 group, Constant P	265.79	0.00	1.0000	1.0000	2	261.79

```
Proportion of sites occupied     (Psi)    = 0.6166 (0.0885)

Detection probabilities           (p):
  grp    srvy        p              se(p)
  ---    ----    ----------     ------------
   1       1      0.349365      ( 0.053728)
```

Single Species; Single-Season - Covariates - PRESENCE

Fitting a model where occupancy varies by browse. [Doesn't make sense to model occupancy as a function of observer - why?]
There are several (equivalent) ways to do this.

- Cell-effects approach.

$$logit(\psi) = \alpha_1 + \alpha_2(browse)$$

This requires a design matrix with an initial column of 1's (for α_1) and a second column of 0/1's to indicate if browsed (for α_2).
α_1 is interpreted as the logit(occupancy) for the (baseline) of unbrowsed and α_2 is the difference in logits between unbrowsed and browsed.

Single Species; Single-Season - Covariates - PRESENCE

Models where occupancy varies by browse.

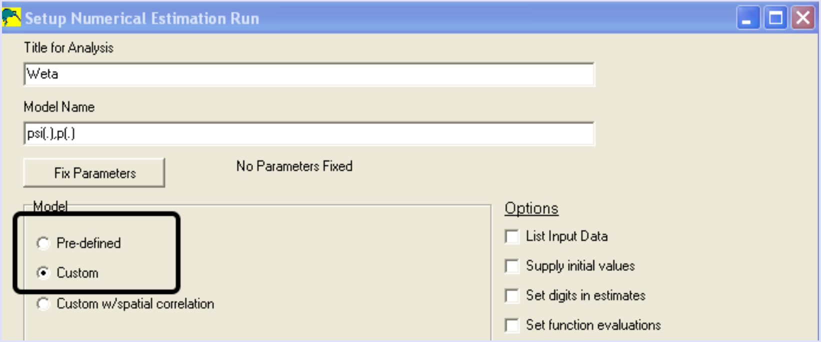
- Cell-means approach.

$$logit(\psi) = \alpha_1(unbrowsed) + \alpha_2(browsed)$$

  This requires a design matrix with initial columns of 0/1's to indicate UNbrowsed, and second column of 0/1's to indicate Browsed.
  α_1 is interpreted as the logit (occupancy) for the unbrowsed sites and α_2 is interpreted as the logit(occupancy) for the browsed sites directly.

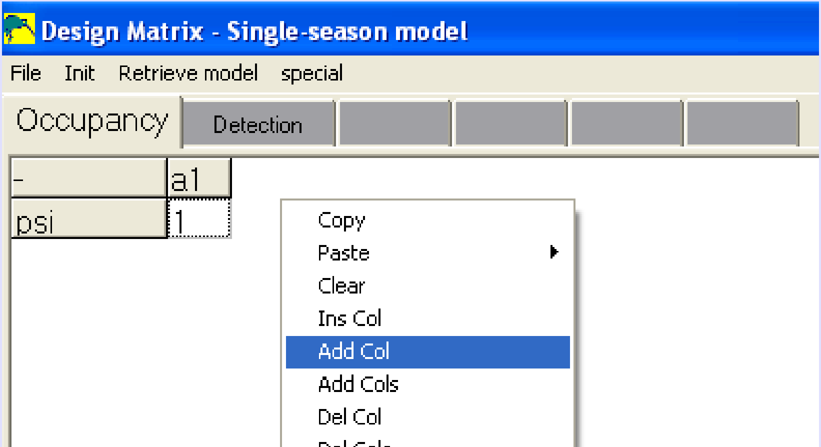
Single Species; Single-Season - Covariates - PRESENCE

Cell-effects approach: Fit Custom Model



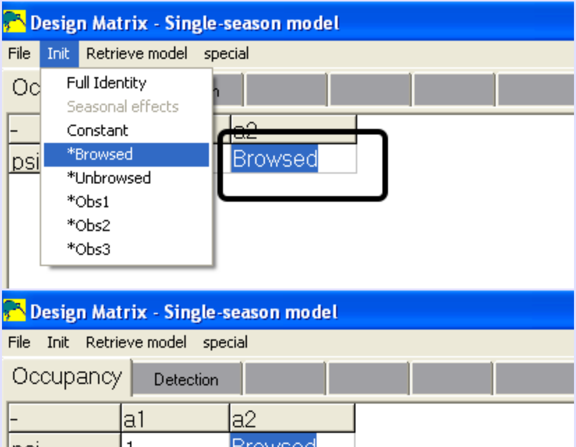
Single Species; Single-Season - Covariates - PRESENCE

Cell-effects approach: Add a second column to design matrix for occupancy



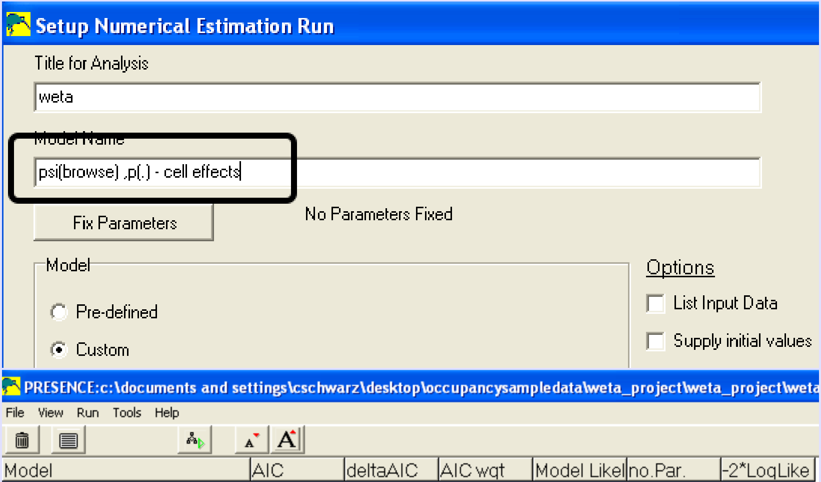
Single Species; Single-Season - Covariates - PRESENCE

Cell-effects approach: Place the *Browse* covariate in second column of design matrix.



Single Species; Single-Season - Covariates - PRESENCE

Cell-effects approach: Run the model (give a sensible model name)



Single Species; Single-Season - Covariates - PRESENCE

Cell-effects approach: Look at the estimates – α_2 is the difference in logit(occupancy) between the two browse classes.

```
Untransformed Estimates of coefficients for covariates (Beta's)
=================================================================
                                              estimate     std.error
A1    psi                                 :  -0.076044    0.432172
A2    psi.Browsed                         :   1.225278    0.720562
B1    p1                                  :  -0.622255    0.236644
```

$$
\begin{aligned}
logit(\psi_{unbrowsed}) =& \alpha_1(1) + \alpha_2(0) = -0.076 \\
\psi_{unbrowsed} =& 1/(1 + exp(-(-0.076)) = 0.4810 \\
logit(\psi_{browsed} =& \alpha_1(1) + \alpha_2(1) = -0.076 + 1.23 = 1.1493 \\
\psi_{browsed} =& 1/(1 + exp(-(1.1493)) = 0.7594.
\end{aligned}
$$

Single Species; Single-Season - Covariates - PRESENCE

Cell-effects approach: Look at the estimates for each site. Site 1=browsed, Site 4=unbrowsed.

```
-------------------------------------------------------------

   Individual Site estimates of <psi>
                 Site                   estimate   Std.err    95% conf. interval
psi              1 site 1             :  0.7594    0.1198     0.4660 - 0.9194
psi              2 site 2             :  0.7594    0.1198     0.4660 - 0.9194
psi              3 site 3             :  0.7594    0.1198     0.4660 - 0.9194
psi              4 site 4             :  0.4810    0.1079     0.2843 - 0.6837
psi              5 site 5             :  0.7594    0.1198     0.4660 - 0.9194
```

Single Species; Single-Season - Covariates - PRESENCE

Cell-effects approach: α_2 is the difference in logit(occupancy) between the two browse classes or $\exp \alpha_2$ is the ODDS-RATIO of occupancy in browsed vs. unbrowsed sites.

The odds of occupancy in browsed sites are $\exp(1.23) = 3.42$ times larger than in unbrowsed sites.

The 95% c.i. for the odds ratio is found as $(\exp(1.24 - 2 \times 0.72)$, $\exp(1.24 + 2 \times 0.72) = (0.82, 14.6)$. Note that this covers the value of 1, so there isn't very strong evidence of a browse effect.

The ΔAIC is also within 2 units of the model with no browse effect, so the evidence for a browse effect is minimal.

Single Species; Single-Season - Covariates - PRESENCE

Cell-means approach: Fit a custom model, create two columns in the design matrix for ψ but now use the browse/unbrowsed indicator variables directly:

Single Species; Single-Season - Covariates - PRESENCE

Cell-means approach: Fit the model, notice that the AIC results are identical between the two approaches.

PRESENCE:c:\documents and settings\cschwarz\desktop\occupancysampledata\weta_project\weta_project\weta_p...

File View Run Tools Help

Model	AIC	deltaAIC	AIC wgt	Model Likel	no.Par.	-2*LogLike
psi(browse) ,p(.) - cell effects	264.26	0.00	0.4056	1.0000	3	258.26
psi(browse/unbrowse),p(.) - cell means	264.26	0.00	0.4056	1.0000	3	258.26
psi(*) p(*)	265.79	1.53	0.1888	0.4653	2	261.79

Single Species; Single-Season - Covariates - PRESENCE

Cell-means approach: Now the estimates for each browse class are given directly:

$$
\begin{aligned}
logit(\psi_{unbrowsed}) =& \alpha_1(0) + \alpha_2(1) = -0.076 \\
\psi_{unbrowsed} =& 1/(1 + exp(-(-0.076)) = 0.4810 \\
logit(\psi_{browsed} =& \alpha_1(1) + \alpha_2(0) = 1.1493 \\
\psi_{browsed} =& 1/(1 + exp(-(1.1493)) = 0.7594.
\end{aligned}
$$

```
Untransformed Estimates of coefficients for covariates (Beta's)
======================================================================
                                          estimate     std.error
A1    psi.Browsed                     :   1.149233     0.655750
A2    psi.Unbrowsed                   :  -0.076044     0.432165
B1    p1                              :  -0.622255     0.236643
```

Single Species; Single-Season - Covariates - PRESENCE

Which approach is better?

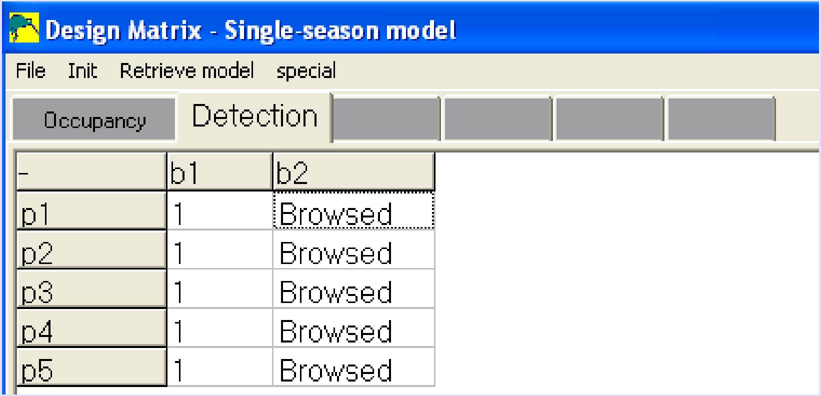
- For a single covariate, it makes no difference.
- For more than one covariate, use the cell effects approach where a factor with m levels has $m-1$ indicator variables and columns in the design matrix. Otherwise you can end up with a design matrix that is not full rank.

Single Species; Single-Season - Covariates - PRESENCE

Try fitting a model where detectability also depends on browse status of the site. i.e. $\psi(\textit{browse}), p(\textit{browse})$. Hint: You can retrieve the design matrices from models already done.

Single Species; Single-Season - Covariates - PRESENCE

Try fitting a model where detectability also depends on browse status of the site



Single Species; Single-Season - Covariates - PRESENCE

Try fitting a model where detectability also depends on browse status of the site. What does ΔAIC tell you?

PRESENCE:c:\documents and settings\cschwarz\desktop\occupancysampledata\weta_project\weta_project\weta_p...

File View Run Tools Help

Model	AIC	deltaAIC	AIC wgt	Model Likel	no.Par.	-2*LogLike
psi(browse) ,p(.) - cell effects	264.26	0.00	0.5455	1.0000	3	258.26
psi(*) p(*)	265.79	1.53	0.2538	0.4653	2	261.79
psi(browse) ,p(browse)	266.26	2.00	0.2007	0.3679	4	258.26

Single Species; Single-Season - Covariates - PRESENCE

Try fitting a model where detectability also depends on browse status of the site. Can you estimate the actual occupancy and detection probabilities in browsed/unbrowsed sites?

```
Untransformed Estimates of coefficients for covariates (Beta's)
===================================================================
                                          estimate    std.error
A1    psi                           :   -0.064383    0.476809
A2    psi.Browsed                   :    1.198036    0.841598
B1    p1                            :   -0.640355    0.380522
B2    p1.Browsed                    :    0.029622    0.485917
```

Single Species; Single-Season - Covariates - PRESENCE

Try fitting a model where detectability also depends on browse status of the site. Estimates of occupancy and detection in the two types of sites.

```
   Individual Site estimates of <psi>
                    Site                 estimate   Std.err    95% conf. interval
psi                 1 site 1           :  0.7565    0.1277      0.4438 - 0.9236
psi                 2 site 2           :  0.7565    0.1277      0.4438 - 0.9236
psi                 3 site 3           :  0.7565    0.1277      0.4438 - 0.9236
psi                 4 site 4           :  0.4839    0.1191      0.2692 - 0.7048
```

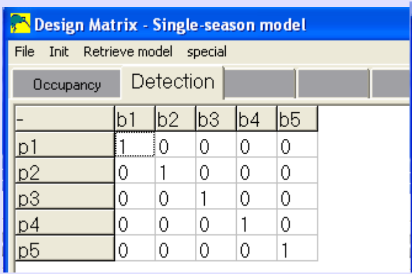
```
   Individual Site estimates of <p1>
                    Site                 estimate   Std.err    95% conf. interval
p1                  1 site 1           :  0.3519    0.0689      0.2309 - 0.4954
p1                  2 site 2           :  0.3519    0.0689      0.2309 - 0.4954
p1                  3 site 3           :  0.3519    0.0689      0.2309 - 0.4954
p1                  4 site 4           :  0.3452    0.0860      0.2000 - 0.5263
```

Single Species; Single-Season - Covariates - PRESENCE

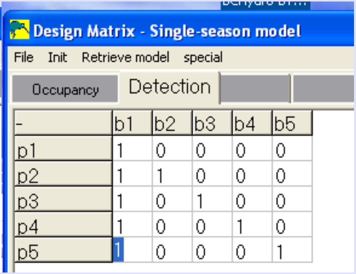
Try fitting a model where detectability depends on the visit $\psi(browse), p(t)$.

Single Species; Single-Season - Covariates - PRESENCE

Try fitting a model where detectability depends on the visit $\psi(browse), p(t)$.



-	b1	b2	b3	b4	b5
p1	1	0	0	0	0
p2	0	1	0	0	0
p3	0	0	1	0	0
p4	0	0	0	1	0
p5	0	0	0	0	1



-	b1	b2	b3	b4	b5
p1	1	0	0	0	0
p2	1	1	0	0	0
p3	1	0	1	0	0
p4	1	0	0	1	0
p5	1	0	0	0	1

Either design matrix works (why?) but have different interpretation of estimates (why?).

Single Species; Single-Season - Covariates - PRESENCE

Try fitting a model where detectability depends on the visit $\psi(browse), p(t)$. Estimate temporal effects.

PRESENCE:c:\documents and settings\cschwarz\desktop\occupancysampledata\weta_project\weta_project\weta_p...

File View Run Tools Help

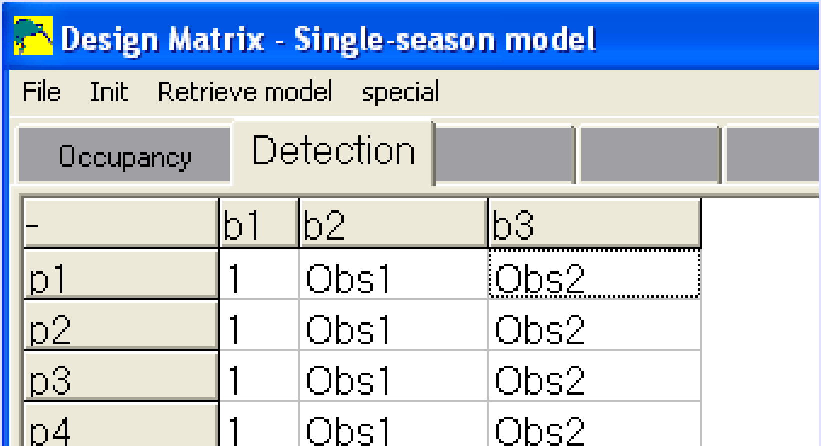
Model	AIC	deltaAIC	AIC wgt	Model Likel	no.Par.	-2*LogLike
psi(browse) ,p(t)	259.44	0.00	0.8586	1.0000	7	245.44
psi(browse) ,p(.) - cell effects	264.26	4.82	0.0771	0.0898	3	258.26
psi(*) p(*)	265.79	6.35	0.0359	0.0418	2	261.79
psi(browse) ,p(browse)	266.26	6.82	0.0284	0.0330	4	258.26

Single Species; Single-Season - Covariates - PRESENCE

Try fitting a model where detectability depends on the observer, but NOT on time $\psi(browse), p(observer)$. Hint: 3 observers need 2 NEW indicator columns. What does the intercept now mean?

Single Species; Single-Season - Covariates - PRESENCE

Try fitting a model where detectability depends on the observer, but NOT on time $\psi(browse), p(observer)$.



-	b1	b2	b3
p1	1	Obs1	Obs2
p2	1	Obs1	Obs2
p3	1	Obs1	Obs2
p4	1	Obs1	Obs2

Single Species; Single-Season - Covariates - PRESENCE

Try fitting a model where detectability depends on the observer, but NOT on time $\psi(browse), p(observer)$. Interpret the estimates.

PRESENCE:c:\documents and settings\cschwarz\desktop\occupancysampledata\weta_project\weta_project\weta_p...

File View Run Tools Help

Model	AIC	deltaAIC	AIC wgt	Model Likel	no.Par.	-2*LogLike
psi(browse) ,p(t)	259.44	0.00	0.6958	1.0000	7	245.44
psi(browse) ,p(observer)	262.04	2.60	0.1896	0.2725	5	252.04
psi(browse) ,p(.) - cell effects	264.26	4.82	0.0625	0.0898	3	258.26
psi(*) p(*)	265.79	6.35	0.0291	0.0418	2	261.79
psi(browse) ,p(browse)	266.26	6.82	0.0230	0.0330	4	258.26

```
Untransformed Estimates of coefficients for covariates (Beta's)
=======================================================================
                                            estimate     std.error
A1    psi                                :  -0.061333    0.434131
A2    psi.Browsed                        :   1.179205    0.701735
B1    p1                                 :  -0.215842    0.327006
B2    p1.obs1                            :  -1.029421    0.432426
B3    p1.obs2                            :  -0.279612    0.403365
```

Single Species; Single-Season - Covariates - PRESENCE

Joint effects of covariates.
Suppose that detectability depended both on occasion effects and observer effects. There are two types of models:

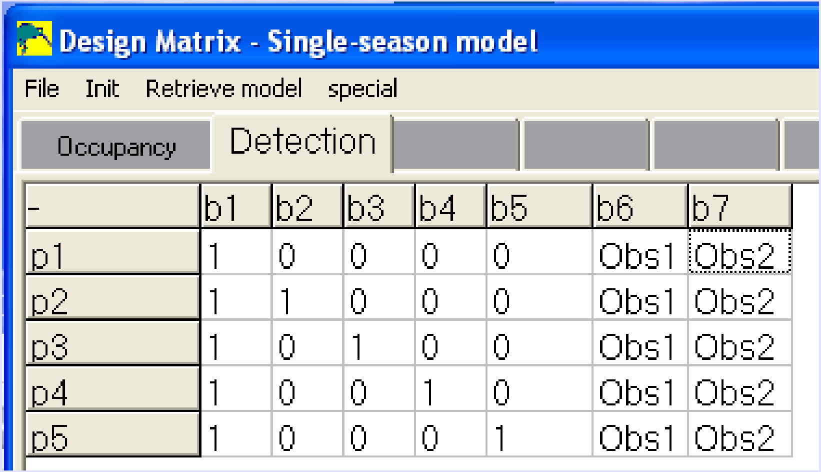
- Additive models. Observers vary among themselves, but are consistent among occasions. For example, one observer has a lower (and consistent) detectability in all occasions even though the detectability varies over occasions. Notation is $p(t\ +\ obs)$. Append columns for each covariate.
- Interaction models. Observers are not consistent over occasions. In some days, observer 1 is worst; on other days observer 2 is worst, etc. Notation is $p(t\ *\ obs)$. Append columns and then append multiplication of columns.

Single Species; Single-Season - Covariates - PRESENCE

Fit the model: $\psi(browse), p(observer\ +\ time)$.

Single Species; Single-Season - Covariates - PRESENCE

Model: $\psi(browse), p(observer + time)$.



-	b1	b2	b3	b4	b5	b6	b7
p1	1	0	0	0	0	Obs1	Obs2
p2	1	1	0	0	0	Obs1	Obs2
p3	1	0	1	0	0	Obs1	Obs2
p4	1	0	0	1	0	Obs1	Obs2
p5	1	0	0	0	1	Obs1	Obs2

Single Species; Single-Season - Covariates - PRESENCE

Model: $\psi(browse), p(observer\ +\ time)$.

PRESENCE:c:\documents and settings\cschwarz\desktop\occupancysampledata\weta_project\weta_project\weta_p...

File View Run Tools Help

Model	AIC	deltaAIC	AIC wgt	Model Likel	no.Par.	-2*LogLike
psi(browse) ,p(t + obs)	257.60	0.00	0.6358	1.0000	9	239.60
psi(browse) ,p(t)	259.44	1.84	0.2534	0.3985	7	245.44
psi(browse) ,p(observer)	262.04	4.44	0.0691	0.1086	5	252.04
psi(browse) ,p(.) - cell effects	264.26	6.66	0.0228	0.0358	3	258.26
psi(*) p(*)	265.79	8.19	0.0106	0.0167	2	261.79
psi(browse) ,p(browse)	266.26	8.66	0.0084	0.0132	4	258.26

Examine and interpret the estimates.

Single Species; Single-Season - Covariates - PRESENCE

Fit the model: $\psi(browse), p(observer * time)$.

Single Species; Single-Season - Covariates - PRESENCE

Fit the model: $\psi(browse), p(observer * time)$. Notice how "columns are multiplied" after the intercept column.

Design Matrix - Single-season model

File Init Retrieve model special

Occupancy | Detection

-	b1	b2	b3	b4	b5	b6	b7	b8	b9	b10	b11	b12	b13	b14	b15
p1	1	0	0	0	0	Obs1	Obs2	0	0	0	0	0	0	0	0
p2	1	1	0	0	0	Obs1	Obs2	Obs1	Obs2	0	0	0	0	0	0
p3	1	0	1	0	0	Obs1	Obs2	0	0	Obs1	Obs2	0	0	0	0
p4	1	0	0	1	0	Obs1	Obs2	0	0	0	0	Obs1	Obs2	0	0
p5	1	0	0	0	1	Obs1	Obs2	0	0	0	0	0	0	Obs1	Obs2

Single Species; Single-Season - Covariates - PRESENCE

Fit the model: $\psi(\mathit{browse}), p(\mathit{observer} * \mathit{time})$.

PRESENCE:c:\documents and settings\cschwarz\desktop\occupancysampledata\weta_project\weta_project\weta_p...

File View Run Tools Help

Model	AIC	deltaAIC	AIC wgt	Model Likel	no.Par.	-2*LogLike
psi(browse) ,p(t + obs)	257.60	0.00	0.3955	1.0000	9	239.60
psi(browse) ,p(t * obs)	257.69	0.09	0.3781	0.9560	17	223.69
psi(browse) ,p(t)	259.44	1.84	0.1576	0.3985	7	245.44
psi(browse) ,p(observer)	262.04	4.44	0.0429	0.1086	5	252.04
psi(browse) ,p(.) - cell effects	264.26	6.66	0.0142	0.0358	3	258.26
psi(*) p(*)	265.79	8.19	0.0066	0.0167	2	261.79
psi(browse) ,p(browse)	266.26	8.66	0.0052	0.0132	4	258.26

This model has same support as additive model, so don't include in final model table.

Single Species; Single-Season - Covariates - PRESENCE

Fit the following models:

- $\psi(browse), p(observer \ + \ time \ + \ browse)$.
- $\psi(*), p(observer \ + \ time \ + \ browse)$.
- $\psi(*), p(observer \ + \ time)$.

Single Species; Single-Season - Covariates - PRESENCE

What do you conclude?

PRESENCE:c:\documents and settings\cschwarz\desktop\occupancysampledata\weta_project\weta_project\weta_p...

File View Run Tools Help

Model	AIC	deltaAIC	AIC wgt	Model Likel	no.Par.	-2*LogLike
psi(browse) ,p(t + obs)	257.60	0.00	0.3370	1.0000	9	239.60
psi(*) ,p(t + obs)	258.55	0.95	0.2096	0.6219	8	242.55
psi(*) ,p(t + obs + browse)	259.41	1.81	0.1363	0.4045	9	241.41
psi(browse) ,p(t)	259.44	1.84	0.1343	0.3985	7	245.44
psi(browse) ,p(t + obs + browse)	259.60	2.00	0.1240	0.3679	10	239.60
psi(browse) ,p(observer)	262.04	4.44	0.0366	0.1086	5	252.04
psi(browse) ,p(.) - cell effects	264.26	6.66	0.0121	0.0358	3	258.26
psi(*) p(*)	265.79	8.19	0.0056	0.0167	2	261.79
psi(browse) ,p(browse)	266.26	8.66	0.0044	0.0132	4	258.26

Single Species; Single-Season

Single-Species Single-Season Occupancy Studies

Covariates and *RPresence*

Single Species; Single-Season - Covariates - *RPresence*

Site-level covariates (both continuous and categorical)

- Create a data frame with site level covariates. E.g.

```
>head(site.covar)
Site Elevation. Aspect
 1.      140         E
 2.      150         S
 3.      110         N
   ....
```

Categorical covariates should be alpha-numeric and not simple numerical codes.

Single Species; Single-Season - Covariates - *RPresence*

Site-level covariates.

- Specify site-level covariates when creating the *.pao object with the *unitcov* argument.

```
study.pao <- RPresence::createPao(input.history,
                                  unitcov=site.covar,
                                  title='blah blah')
```

Single Species; Single-Season - Covariates - *RPresence*

Model building. Continuous or categorical variables specified the same way.

```
fit1 <- RPresence::occMod(model=list(psi~aspect, p~1),
            type="so", data=...)
fit2 <- RPresence::occMod(model=list(psi~elevation, p~1
            type="so", data=...)
```

- AIC table is constructed in the usual way:

```
model.set <-RPresence::createAicTable(... list of model
```

Single Species; Single-Season - Covariates - *RPresence*

Model averaging takes place in the usual way.

- For each site for each model, the covariates for that site are used to make a prediction.

```
psi.ma <- RPresence::modAvg(model.set, param="psi")
```

- Append the site covariates to the predictions

```
psi.ma <- cbind(psi.ma, site.covar)
```

- Mske a plot

```
ggplot(data=psi.ma, aes(x=..., y=estimate)+
  geom_point()
```

Single Species; Single-Season - Covariates - *RPresence* I

Visit (Survey)-level covariates. The value of the covariate is applicable to ALL sites on this visit.
Visit $\times$ Site (Sampling) covariates. The value of the covariate is specific to that site on that visit.

Both are entered in the same way by creating a data.frame for every combination of site and visit.
This data frame must be sorted by visit and then by site within visit, i.e. all sites for visit 1, then all sites for visit 2, etc.
Add covariate values (continuous or categorical. If a site is not visited on a survey, you can set the covariate to *NA*

Add to the *pao object

Single Species; Single-Season - Covariates - *RPresence* II

```
weta.pao <- RPresence::createPao(input.history,
                                 unitcov=site_covar,
                                 survcov=survey.cov,
                                 title='weta SSSS')
```

Single Species; Single-Season - Covariates - *RPresence*

Example of survey covariates for observers

```
    visit site  Obs
1       1    1    01
2       1    2    01
3       1    3    01
...
15      1   15   01
16      1   16 <NA>
17      1   17 <NA>
18      1   18 <NA>
19      1   19 <NA>
20      1   20   01
21      1   21   01
22      1   22    01
```

Single Species; Single-Season - Covariates - *RPresence*

Mahoenui giant weta (*Deinacrida mahoenui*) is endemic to New Zealand and under stress from rats and other predators.
72 circular plots (3 m radius, primarily prickly gorse plants) were surveyed for weta.
Each plot surveyed 3-5 times.
Covariates to be considered:

- Observer. Three different observers and not every plot surveyed by each observer.
- Browse. Was each site browsed by goats, yes or no.

Single Species; Single-Season - Covariates - *RPresence*

Getting the data into *RPresence*.

1. Capture History

Get $n_{sites} \times n_{visit}$ data.frame (or matrix) of 1, 0, or NAs

```
input.history <- readxl::read_excel("Weta_pg116.xls",
                            sheet="detection_historie
                            na="-",
                            col_names=FALSE)   # notic
```

Detection histories include many missing values. Are these MCAR?

Single Species; Single-Season - Covariates - *RPresence*

Getting the data into *RPresence*.

2. Site Covariates

Get $n_{sites} \times n_{site-covariates}$ data.frame of site covariates.

- Continuous covariates occupy 1 column
- Categorical covariates can either be alpha-numeric code or a set of indicator variables.

With modern software, the former is preferred.

Single Species; Single-Season - Covariates - *RPresence*

Getting the data into *RPresence*.

2. Site Covariates

```
site_covar <- readxl::read_excel("Weta_pg116.xls",
                                  sheet="site_covar",
                                  na="-",
                                  col_names=TRUE)  # notice

# Create an alternate site level covariate that is a categ
# than indicator variables
site_covar$BrowCat <- site_covar$BrowseCat
xtabs(~BrowCat, data=site_covar,exclude=NULL, na.action=na
head(site_covar)
```

Single Species; Single-Season - Covariates - *RPresence*

Getting the data into *RPresence*.

2. Site Covariates

```
> head(site_covar)
  Browsed Unbrowsed BrowCat
1     1.00        0     B
2     1.00        0     B
3     1.00        0     B
4     0           1.00 N
5     1.00        0     B
6     0           1.00 N
```

Browse covariate can be entered as two indicator variables (this has implications later on) or as a categorical variable (preferred). No missing values allowed in site-level covariates.

Single Species; Single-Season - Covariates - *RPresence*

Getting the data into *RPresence*.

3. Visit Covariates

Get $n_{visit-covariates}$ sets of $n_{sites} \times n_{visits}$ data.frame of visit covariates.

- Continuous covariates have values in each cell
- Categorical covariates can either be alpha-numeric codes or a set of indicator variables.

With modern software, the former is preferred.

Single Species; Single-Season - Covariates - *RPresence*

Getting the data into *RPresence*.

3. Visit Covariates

```
# Get the individual covariates.
obs1 <- readxl::read_excel("Weta_pg116.xls",
                                  sheet="Obs1",
                                  na="-",
                                  col_names=FALSE)
....

# Create an alternate site level covariate that is a categ
# than indicator variables
obs <- obs1*1 + obs2*2 + obs3*3
```

Notice how I created a single categorical covariate for observer number (preferred)

Single Species; Single-Season - Covariates - *RPresence*

Getting the data into *RPresence*.

3. Visit Covariates

These must be stacked into a vector of length $n_{sites} \times n_{visits}$ where values for visit 1 appear first (for all sites), then for visit 2, etc

```
survey.cov <- data.frame(
    site=rep(1:nrow(input.history) , ncol(input.history)),
    visit=rep(1:ncol(input.history), each=nrow(input.histor
    obs1 =as.vector(unlist(obs1)),
    obs2 =as.vector(unlist(obs2)),
    obs3 =as.vector(unlist(obs3)),
    Obs  =paste("O",as.vector(unlist(Obs)),sep=""),     # n
survey.cov$Obs[ grepl("NA",survey.cov$Obs)] <- NA
head(survey.cov)
```

Notice how *obs* was forced to be alphanumeric so *RPresence* will not treat it as a continuous variable.

Single Species; Single-Season - Covariates - *RPresence*

Getting the data into *RPresence*.

3. Visit Covariates

These must be stacked into a vector of length $n_{sites} \times n_{visits}$ where values for visit 1 appear first (for all sites), then for visit 2, etc

```
 > head(survey.cov)
  site visit obs1 obs2 obs3 obs
1    1     1    1    0    0   1
2    2     1    1    0    0   1
3    3     1    1    0    0   1
4    4     1    1    0    0   1
....
```

Final column is a categorical covariate with values of "O1", "O2","O3" (be sure that these are alpha-number codes, otherwise *RPresence* will treat as continuous variables.;

Single Species; Single-Season - Covariates - *RPresence*

Getting the data into *RPresence*.
Finally, create the **.pao* object.

```
weta.pao <- RPresence::createPao(input.history,
                                 unitcov=site_covar,
                                 survcov=survey.cov,
                                 title='weta SSSS')
weta.pao
```

Single Species; Single-Season - Covariates - *RPresence*

Fit the $\psi(*)$, $p(*)$ model and look at estimates.

```
mod.pdot <- RPresence::occMod(
           model=list(psi~1, p~1),
           type="so", data=weta.pao)
summary(mod.pdot)

# look at estimated occupancy probability. RPresence gives
mod.pdot.psi <-mod.pdot$real$psi[1,]  # occupancy probabil
mod.pdot.psi

# look at the estimated probability of detection. It gives
mod.pdot.p    <- mod.pdot$real$p[seq(1, by=nrow(input.histo
mod.pdot.p

# alternatively
RPresence::print_one_site_estimates(mod.pdot, site = 1)
```

Single Species; Single-Season - Covariates - *RPresence*

Fit the $\psi(*), p(*)$ model and look at estimates.

```
AIC=265.7872
-2*log-likelihood=261.7872        num. par=2
>
> mod.pdot.psi
            est          se lower_0.95 upper_0.95
unit1 0.6165958 0.08854195  0.4356219   0.770157
>
> mod.pdot.p
                est          se lower_0.95 upper_0.95
unit1_1-1 0.3493651 0.05372869  0.2525413  0.4604435
...
```

Single Species; Single-Season - Covariates - *RPresence*

Fitting a model where occupancy varies by browse.
It doesn't make sense to model occupancy as a function of observer - why?

There are several (equivalent) ways to do this.

Single Species; Single-Season - Covariates - *RPresence*

Models where occupancy varies by browse.

- Cell-means approach.

$$logit(\psi) = \alpha_1(unbrowsed) + \alpha_2(browsed)$$

  This requires a design matrix with initial columns of 0/1's to indicate UNbrowsed, and second column of 0/1's to indicate Browsed.
  α_1 is interpreted as the logit (occupancy) for the unbrowsed sites and α_2 is interpreted as the logit(occupancy) for the browsed sites directly.

Only useful for models with a single categorical covariate for a parameter.

Single Species; Single-Season - Covariates - *RPresence*

Models where occupancy varies by browse.

- Cell-effects approach.

$$logit(\psi) = \alpha_1 + \alpha_2(browse)$$

This requires a design matrix with an initial column of 1's (for α_1) and a second column of 0/1's to indicate if browsed (for α_2).

α_1 is interpreted as the logit(occupancy) for the (baseline) of unbrowsed and α_2 is the difference in logits between unbrowsed and browsed.

Can be used with any number of categorical covariates. Trick is figuring out which is the reference class used (corresponding to the α_1 term)

Single Species; Single-Season - Covariates - *RPresence*

Cell-means approach:

```
mod.pdot.psiB.1 <- RPresence::occMod(
      model=list(psi~-1+Browsed+Unbrowsed, p~1),
      type="so", data=weta.pao)
```

or (prefered)

```
mod.pdot.psiB.4 <- RPresence::occMod(
      model=list(psi~-1+BrowCat, p~1),
      type="so", data=weta.pao)
```

No need for you to define indicator variables with the latter model.

Single Species; Single-Season - Covariates - *RPresence*

Cell-means approach: Results (same for ALL models)

```
> summary(mod.pdot.psiB.1)
AIC=264.2643
-2*log-likelihood=258.2643
num. par=3
>
> mod.pdot.psiB.1.psi[1:5,]
            est        se lower_0.95 upper_0.95
unit1 0.7593708 0.1198231  0.4660490  0.9194190
...
unit4 0.4809982 0.1078853  0.2843322  0.6837339
...
```

Probability of Occupancy for Browsed areas is 0.75; that of unbrowwed areas 0.48.

Single Species; Single-Season - Covariates - *RPresence*

Cell-means approach:

Look at design matrix

```
# This is the logit occupancy for each browse category
mod.pdot.psiB.1$dmat$psi
mod.pdot.psiB.1$beta$psi
```

```
> mod.pdot.psiB.1$dmat$psi
    a1            a2
psi "psi.Browsed" "psi.Unbrowsed"
>
> mod.pdot.psiB.1$beta$psi
  psi.coeff
1  1.149233
2 -0.076044
```

Single Species; Single-Season - Covariates - *RPresence*

Cell-means approach:

$$\begin{aligned} logit(\psi_{unbrowsed}) &= \alpha_1(0) + \alpha_2(1) = -0.076 \\ \psi_{unbrowsed} &= 1/(1 + exp(-(-0.076)) = 0.4810 \\ logit(\psi_{browsed} &= \alpha_1(1) + \alpha_2(0) = 1.1493 \\ \psi_{browsed} &= 1/(1 + exp(-(1.1493)) = 0.7594. \end{aligned}$$

Single Species; Single-Season - Covariates - *RPresence*

Cell-means approach: Estimate the odds ratio

```
mod.pdot.psiB.1.oddsratio.browse <-
        exp( sum(c(1,-1)*mod.pdot.psiB.1$beta$psi))
mod.pdot.psiB.1.oddsratio.browse
```

```
> mod.pdot.psiB.1.oddsratio.browse
[1] 3.405109
```

Odds of Occupancy for Browsed areas is 3.41x that of unbrowsed areas.

Single Species; Single-Season - Covariates - *RPresence*

Cell-effects approach:

```
mod.pdot.psiB.2 <- RPresence::occMod(
        model=list(psi~Browsed, p~1),
        type="so", data=weta.pao)
```

or (prefered)

```
mod.pdot.psiB.3 <- RPresence::occMod(
        model=list(psi~BrowCat, p~1),
        type="so", data=weta.pao)
```

No need for you to define indicator variables with the latter model.

Single Species; Single-Season - Covariates - *RPresence*

Cell-effects approach:

Look at design matrix

```
# This is the logit occupancy for each browse category
mod.pdot.psiB.3$dmat$psi
mod.pdot.psiB.3$beta$psi
```

```
> mod.pdot.psiB.1$dmat$psi
    a1            a2
psi "psi.Browsed" "psi.Unbrowsed"
>
> mod.pdot.psiB.3$dmat$psi
    a1  a2
psi "1" "psi.BrowCatN"
> mod.pdot.psiB.3$beta$psi
  psi coeff
```

Single Species; Single-Season - Covariates - *RPresence*

Cell-effects approach:

$$\begin{aligned}
logit(\psi_{unbrowsed}) =& \alpha_1(1) + \alpha_2(0) = -0.076 \\
\psi_{unbrowsed} =& 1/(1 + exp(-(-0.076)) = 0.4810 \\
logit(\psi_{browsed} =& \alpha_1(1) + \alpha_2(1) = -0.076 + 1.23 = 1.1493 \\
\psi_{browsed} =& 1/(1 + exp(-(1.1493)) = 0.7594.
\end{aligned}$$

Single Species; Single-Season - Covariates - *RPresence*

Cell-effects approach: Estimate the odds ratio

```
mod.pdot.psiB.3.oddsratio.browse <-
        exp( sum(c(0,-1)*mod.pdot.psiB.1$beta$psi))
mod.pdot.psiB.3.oddsratio.browse
```

```
> mod.pdot.psiB.3.oddsratio.browse
[1] 3.405113
```

Odds of Occupancy for Browsed areas is 3.41x that of unbrowsed areas.

Single Species; Single-Season - Covariates - *RPresence*

Odds of Occupancy for Browsed areas is 3.41x that of unbrowsed areas.
Standard errors for the log(odds ratio) can also be found (contact me) and is 0.72.
Then the 95% c.i. for the odds ratio is found as
$(\exp(1.24 - 2 \times 0.72), \exp(1.24 + 2 \times 0.72) = (0.82, 14.6)$.
Note that this covers the value of 1, so there isn't very strong evidence of a browse effect.

The ΔAIC is also within 2 units of the model with no browse effect, so the evidence for a browse effect is minimal.

Single Species; Single-Season - Covariates - *RPresence*

Which approach is better?

- For a single covariate, it makes no difference.
- For more than one covariate, use the cell effects approach where a factor with m levels has $m-1$ indicator variables and columns in the design matrix. Otherwise you can end up with a design matrix that is not full rank.

Single Species; Single-Season - Covariates - *RPresence*

Try fitting a model where detectability also depends on browse status of the site. i.e. $\psi(browse), p(browse)$.

Single Species; Single-Season - Covariates - *RPresence*

Try fitting a model where detectability also depends on browse status of the site. i.e. $\psi(browse), p(browse)$.

```
mod.pB.psiB <- RPresence::occMod(
      model=list(psi~BrowCat, p~BrowCat),
      type="so", data=weta.pao)
```

Single Species; Single-Season - Covariates - *RPresence*

Model $\psi(browse), p(browse)$.
Estimated occupancy by browse:

```
mod.pB.psiB.psi <-mod.pB.psiB$real$psi  # occupancy probab
mod.pB.psiB.psi[1:5,]
```

```
> # This is the logit occupancy for each browse category
> mod.pB.psiB.psi[1:5,]
            est        se lower_0.95 upper_0.95
unit1 0.7565126 0.1276939  0.4439733  0.9236045
...
unit4 0.4839101 0.1190770  0.2691584  0.7047733
```

Single Species; Single-Season - Covariates - *RPresence*

Model $\psi(browse), p(browse)$.

Estimated detection by browse:

```
mod.pB.psiB.p <-mod.pB.psiB$real$p  # detection probabilit
mod.pB.psiB.p[1:5,]
```

```
> # This is the logit occupancy for each browse category
> mod.pB.psiB.p[1:5,]
                est         se lower_0.95 upper_0.95
unit1_1-1 0.3518920 0.06891346  0.2309473  0.4953761
...
unit4_1-1 0.3451663 0.08600807  0.2000208  0.5263392
...
```

Single Species; Single-Season - Covariates - *RPresence*

Try fitting a model where detectability depends on the visit $\psi(browse), p(t)$.

Single Species; Single-Season - Covariates - *RPresence*

Model $\psi(browse), p(t)$.

```
mod.pt.psiB <- RPresence::occMod(
       model=list(psi~BrowCat, p~factor(visit)),
       type="so", data=weta.pao)

mod.pt.psiB$real$psi[1:5,]
mod.pt.psiB$real$p[seq(1, by=nrow(input.history), length.ou
```

Not clear why SURVEY doesn't work now (?) but we have a *visit* covariate that we defined earlier. Be sure to declare it as a factor so that it is not treated as a continuous covariates.

Single Species; Single-Season - Covariates - *RPresence*

Model $\psi(browse), p(t)$.

```
> mod.pt.psiB$real$psi[1:5,]
            est        se lower_0.95 upper_0.95
unit1 0.7699149 0.1232454  0.4611357  0.9290003
...
unit4 0.4931709 0.1116897  0.2884113  0.7002475
...
> mod.pt.psiB$real$p[seq(1, by=nrow(input.history),
      length.out=ncol(input.history)),]
                est         se lower_0.95 upper_0.95
unit1_1-1 0.3520562 0.09836946 0.18920071  0.5585270
unit1_1-2 0.3175294 0.08921332 0.17192709  0.5104317
unit1_1-3 0.1694830 0.06707655 0.07424149  0.3417958
unit1_1-4 0.3115815 0.08815339 0.16822898  0.5031897
```

Single Species; Single-Season - Covariates - *RPresence*

Try fitting a model where detectability depends on the observer, but NOT on time $\psi(\textit{browse}), p(\textit{observer})$. Hint: 3 observers need 2 NEW indicator columns. What does the intercept now mean?

Single Species; Single-Season - Covariates - *RPresence*

Model $\psi(browse), p(observer)$.

```
mod.pO.psiB <- RPresence::occMod(
        model=list(psi~BrowCat, p~obs),
        type="so", data=weta.pao)
```

Needed to insert non-missing values when no visit occur.

Single Species; Single-Season - Covariates - *RPresence*

Model $\psi(browse), p(observer)$.

```
> mod.p0.psiB$real$psi[1:5,]
            est        se lower_0.95 upper_0.95
unit1 0.7535936 0.1163803  0.4723969  0.9126370
...
unit4 0.4846713 0.1084314  0.2865457  0.6877352
...> mod.p0.psiB$real$p[seq(1, by=nrow(input.history), leng
                est         se lower_0.95 upper_0.95
unit1_1-1 0.2235210 0.06272749  0.1241580  0.3689096
unit1_1-2 0.4462478 0.08080860  0.2980129  0.6047011
unit1_1-3 0.3786094 0.07998710  0.2383368  0.5426237
unit1_1-4 0.4462478 0.08080860  0.2980129  0.6047011
unit1_1-5 0.2235210        NaN        NaN        NaN
```

Visit 5 at site 1 was missed, so the estimate at this time point is

Single Species; Single-Season - Covariates - *RPresence*

Joint effects of covariates.
Suppose that detectability depended both on occasion effects and observer effects. There are two types of models:

- Additive models. Observers vary among themselves, but are consistent among occasions. For example, one observer has a lower (and consistent) detectability in all occasions even though the detectability varies over occasions. Notation is $p(t + obs)$. Append columns for each covariate.
- Interaction models. Observers are not consistent over occasions. In some days, observer 1 is worst; on other days observer 2 is worst, etc. Notation is $p(t * obs)$. Append columns and then append multiplication of columns.

This is easily done in *RPresence* without having to physically create the extra columns using standard modelling notation of *R*.

Single Species; Single-Season - Covariates - *RPresence*

Fit the model: $\psi(browse), p(observer\ +\ time)$.

Single Species; Single-Season - Covariates - *RPresence*

Model: $\psi(browse), p(observer + time)$.

```
mod.pOpV.psiB <- RPresence::occMod(
      model=list(psi~BrowCat, p~obs+factor(visit)),
      type="so", data=weta.pao)
```

Needed to insert non-missing values when no visit occur.
Note use of *factor()* function for the visit.

Single Species; Single-Season - Covariates - *RPresence*

Model: $\psi(browse), p(observer + time)$.

```
> mod.pOpV.psiB$real$psi[1:5,]
            est        se lower_0.95 upper_0.95
unit1 0.7673332 0.1206151  0.4673420  0.9253559
...
unit4 0.5059982 0.1149003  0.2938128  0.7160446
..
> mod.pOpV.psiB$real$p[seq(1, by=nrow(input.history), lengt
                est         se   lower_0.95 upper_0.95
unit1_1-1 0.2142123 0.08831391 8.882740e-02  0.4325632
unit1_1-2 0.4050194 0.11960192 2.046700e-01  0.6429452
unit1_1-3 0.1799850 0.08137467 6.932630e-02  0.3927389
unit1_1-4 0.4254702 0.12515624 2.135160e-01  0.6688850
unit1_1-5 0.6247542 6.55569910 2.622312e-24  1.0000000
```

Single Species; Single-Season - Covariates - *RPresence*

Fit the following models:

- $\psi(browse), p(observer\ +\ time\ +\ browse)$.
- $\psi(*), p(observer\ +\ time\ +\ browse)$.
- $\psi(*), p(observer\ +\ time)$.

Single Species; Single-Season - Covariates - *RPresence*

Construct the AIC table.

```
models<-list(mod.pdot,
             mod.pdot.psiB.1,
             mod.pB.psiB,
             mod.pt.psiB,
             mod.p0.psiB,
             mod.p0pV.psiB,
             mod.p0pVpB.psiB,
             mod.p0pVpB.psi.,
             mod.p0pVpB.psi.
             )
results<-RPresence::createAicTable(models)
summary(results)
```

Single Species; Single-Season - Covariates - *RPresence*

What do you conclude from the AIC table?

```
> summary(results)
                                          Model DAIC    wgt
1                 psi(BrowCat)p(factor(visit)) 0.00 0.2289
2           psi(BrowCat)p(obs P factor(visit)) 0.16 0.2110
3           psi(BrowCat)p(obs P factor(visit)) 0.16 0.2110
4           psi(BrowCat)p(obs P factor(visit)) 0.16 0.2110
5 psi(BrowCat)p(obs P factor(visit) P BrowCat) 2.16 0.0776
6                           psi(BrowCat)p(obs) 4.61 0.0229
7             psi(-1 P Browsed P Unbrowsed)p() 4.83 0.0205
8                                     psi()p() 6.35 0.0096
9                       psi(BrowCat)p(BrowCat) 6.82 0.0075
```

Single Species; Single-Season - Covariates - *RPresence*

Model averaging of the ψ values.

```
RPresence::modAvg(results, param="psi")[1:5,]
```

```
               est        se lower_0.95 upper_0.95
unit1 0.7659572 0.1224557  0.4617713  0.9258388
unit2 0.7659572 0.1224557  0.4617713  0.9258388
unit3 0.7659572 0.1224557  0.4617713  0.9258388
unit4 0.5029083 0.1155986  0.2901462  0.7146223
unit5 0.7659572 0.1224557  0.4617713  0.9258388
...
```

Single Species; Single-Season - Covariates - *MARK*

Single-Species Single-Season Occupancy Studies

Covariates and *MARK*

Single Species; Single-Season - Covariates - *MARK*

Using *MARK* software is similar except design matrix is for ALL parameters (ψ and p)

You will need to construct the *.inp file with the following format:

hhhhh 1 br o1-1 o1-2 o1-3 o1-4 o1-5 o2-1 o2-2 o2-3 o2-4 o2-5 ;

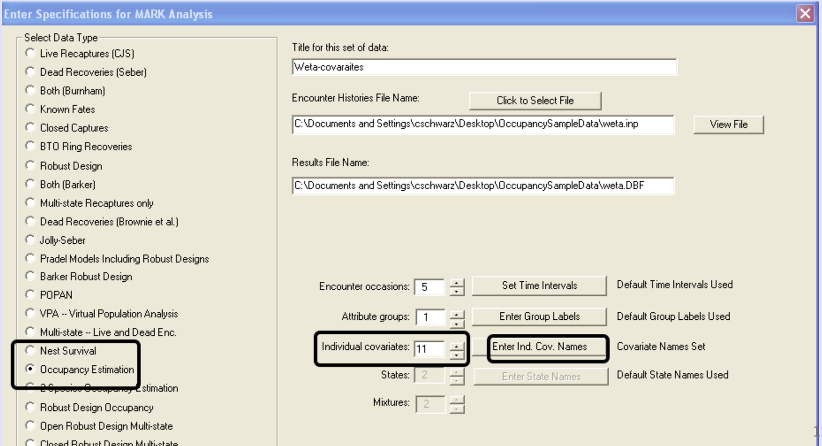
where

- hhhhh is the detection history; Use "." if missing data.
- 1 indicates a count of 1 site with this detection history;
- br is 1/0 if browsed or not-browsed;
- o1-1 . . . o1-5 is 1/0 if observer 1 took a reading at site at time i; Use 0.0 if missing data in history (ignored);
- 02-1 . . . o2-5 is 1/0 if observer 2 took a reading at site at time i; Use 0.0 if missing data in history (ignored);
- ";" terminates the input line.

Look at the *weta.inp* file in the *OccupancySampleData* folder.

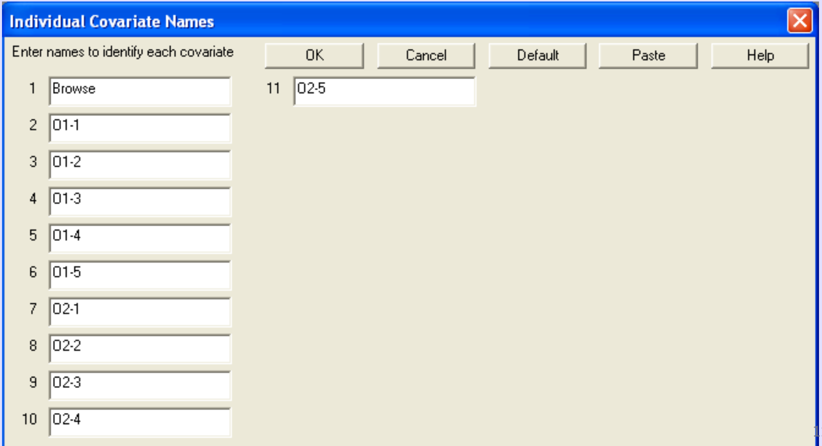
Single Species; Single-Season - Covariates - *MARK*

Start *MARK*; Select Occupancy Modeling; Select the data file; Specify the number of covariates (11 in this case) and name them:



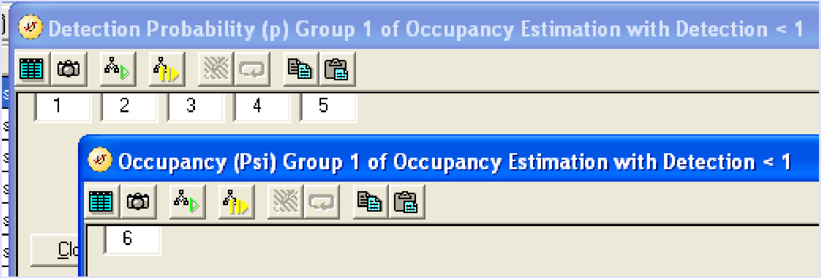
Single Species; Single-Season - Covariates - *MARK*

Start *MARK*; Select Occupancy Modeling; Select the data file; Specify the number of covariates (11 in this case) and name them:



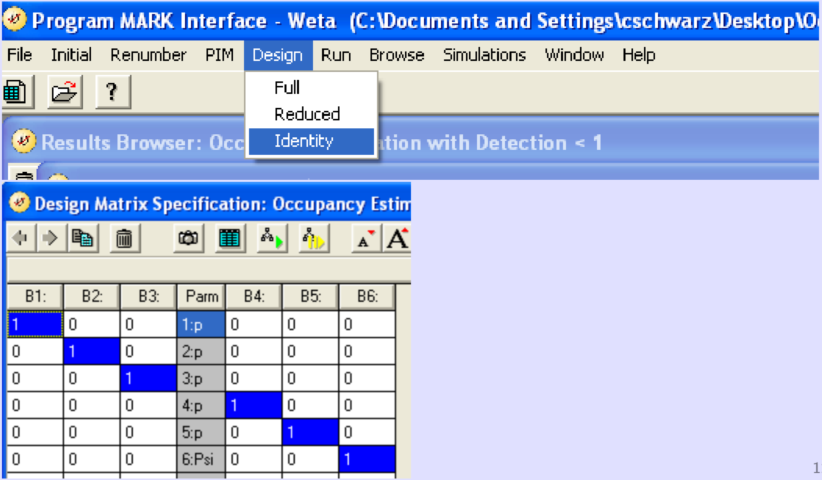
Single Species; Single-Season - Covariates - *MARK*

Model $\psi(*), p(*)$ using a DESIGN matrix.
Set PIMS to separate numbers for each parameter (default).



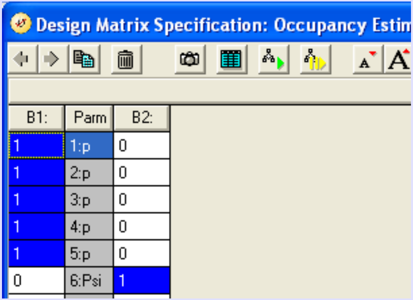
Single Species; Single-Season - Covariates - *MARK*

Select DESIGN → Identity Matrix:



Single Species; Single-Season - Covariates - *MARK*

For model $\psi(*), p(*)$ change DESIGN matrix to the following (why?):



B1:	Parm	B2:
1	1:p	0
1	2:p	0
1	3:p	0
1	4:p	0
1	5:p	0
0	6:Psi	1

Single Species; Single-Season - Covariates - *MARK*

Select RUN; specify model name; add to results table in usual way; look at estimates;

Results Browser: Occupancy Estimation with Detection < 1

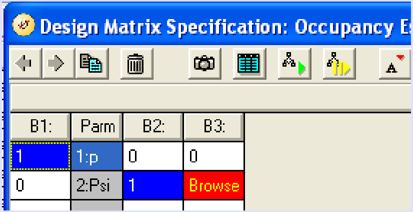
Model	AICc	Delta AICc	AICc Weight	Model Likelihood	No. Par.	Deviance	-2Log(L)
{psi(browse), p(t + observer)}	260.5027	0.0000	0.28995	1.0000	9	239.5994	239.5994
{psi(*), p(t + observer)}	260.8323	0.3296	0.24589	0.8481	8	242.5466	242.5466
{psi(browse), p(t)}	261.1868	0.6841	0.20595	0.7103	7	245.4368	245.4368
{psi(*), p(t + observer + browse)}	262.3180	1.8153	0.11699	0.4035	9	241.4147	241.4147
{psi(browse), p(observer)}	262.9512	2.4485	0.08524	0.2940	5	252.0421	252.0421
{psi(browse) p(*)}	264.6172	4.1145	0.03706	0.1278	3	258.2643	258.2643
{psi(*) p(*)}	265.9611	5.4584	0.01893	0.0653	2	261.7872	261.7872

```
                                  weta

                   Real Function Parameters of {psi(*) p(*)}
                                                          95% Confidence Interval
Parameter                   Estimate       Standard Error      Lower          Upper
-------------------------  --------------  --------------  --------------  ----------
   1:p                     0.3493651       0.0537299       0.2525377       0.4604481
   2:Psi                   0.6165958       0.0885440       0.4356143       0.7701624
```

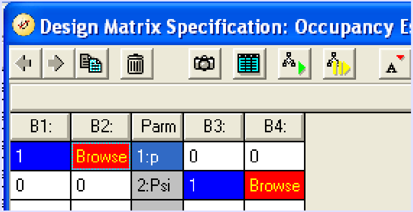
Single Species; Single-Season - Covariates - *MARK*

For model $\psi(browse), p(*)$ change DESIGN matrix to the following (why?):



Single Species; Single-Season - Covariates - *MARK*

For model $\psi(browse), p(browse)$ change DESIGN matrix to the following (why?):



B1:	B2:	Parm	B3:	B4:
1	Browse	1:p	0	0
0	0	2:Psi	1	Browse

Single Species; Single-Season - Covariates - *MARK*

For model $\psi(browse), p(t)$ change DESIGN matrix to the following (why?):

B1: p Int	B2: p t1	B3: p t2	B4: p t3	B5: p t4	Parm	B6: Psi Int	B7:
1	1	0	0	0	1:p	0	0
1	0	1	0	0	2:p	0	0
1	0	0	1	0	3:p	0	0
1	0	0	0	1	4:p	0	0
1	0	0	0	0	5:p	0	0
0	0	0	0	0	6:Psi	1	Browse

Single Species; Single-Season - Covariates - *MARK*

For model $\psi(browse), p(observer)$ change DESIGN matrix to the following (why?):

B1: p Int	B2: p t1	B3: p t2	Parm	B4: Psi Int	B5:
1	01-1	02-1	1:p	0	0
1	01-2	02-2	2:p	0	0
1	01-3	02-3	3:p	0	0
1	01-4	02-4	4:p	0	0
1	01-5	02-5	5:p	0	0
0	0	0	6:Psi	1	Browse

Single Species; Single-Season - Covariates - *MARK*

For model $\psi(browse), p(t + observer)$ change DESIGN matrix to the following (why?):

B1: p Int	B2:	B3:	B4:	B5:	B6: p t1	B7: p t2	Parm	B8: Psi Int	B9:
1	1	0	0	0	01-1	02-1	1:p	0	0
1	0	1	0	0	01-2	02-2	2:p	0	0
1	0	0	1	0	01-3	02-3	3:p	0	0
1	0	0	0	1	01-4	02-4	4:p	0	0
1	0	0	0	0	01-5	02-5	5:p	0	0
0	0	0	0	0	0	0	6:Psi	1	Browse

Single Species; Single-Season - Covariates - *MARK*

Fit $\psi(browse), p(t\ +\ observer\ +\ browser)$:

Single Species; Single-Season - Covariates - *MARK*

Model $\psi(*), p(t + observer + browser)$ DESIGN matrix is (why?) :

B1: p Int	B2:	B3:	B4:	B5:	B6: p t1	B7: p t2	B8:	Parm	B9: Psi Int
1	1	0	0	0	01-1	02-1	Browse	1:p	0
1	0	1	0	0	01-2	02-2	Browse	2:p	0
1	0	0	1	0	01-3	02-3	Browse	3:p	0
1	0	0	0	1	01-4	02-4	Browse	4:p	0
1	0	0	0	0	01-5	02-5	Browse	5:p	0
0	0	0	0	0	0	0	0	6:Psi	1

Single Species; Single-Season - Covariates - *MARK*

Results window similar to PRESENCE. AIC is corrected for small sample sizes. Perform model averaging, etc.

Results Browser: Occupancy Estimation with Detection < 1

Model	AICc	Delta AICc	AICc Weight	Model Likelihood	No. Par.	Deviance	-2Log(L)
{psi(browse), p(t + observer)}	260.5027	0.0000	0.28648	1.0000	9	239.5994	239.5994
{psi(*), p(t + observer)}	260.8323	0.3296	0.24296	0.8481	8	242.5466	242.5466
{psi(browse), p(t)}	261.1868	0.6841	0.20349	0.7103	7	245.4368	245.4368
{psi(*), p(t + observer + browse)}	262.3180	1.8153	0.11559	0.4035	9	241.4147	241.4147
{psi(browse), p(observer)}	262.9512	2.4485	0.08422	0.2940	5	252.0421	252.0421
{psi(browse) p(*)}	264.6172	4.1145	0.03661	0.1278	3	258.2643	258.2643
{psi(*) p(*)}	265.9611	5.4584	0.01870	0.0653	2	261.7872	261.7872
{psi(browse) p(browse)}	266.8576	6.3549	0.01194	0.0417	4	258.2606	258.2606

Single Species; Single-Season - Covariates - *RMark*

Single-Species Single-Season Occupancy Studies

Covariates and *RMark*

Single Species; Single-Season - Covariates - *RMark*

Site-level covariates.

- Continuous covariates
  - Enter as a numeric columns in the history data frame.
  - Specify variable name in formula, e.g. $Psi = list(formula = \sim Area)$.
- Categorical covariates
  - Enter as a alphanumeric columns in the history data frame and declare as a factor.
  - Specify the categorical covariates in the *groups=* option in the *process.data()* function.
  - Specify variable name in formula, e.g. $Psi = list(formula = \sim Browse)$.

Single Species; Single-Season - Covariates - *RMark*

Visit (Survey)-level covariates. The value of the covariate is applicable to ALL sites on this visit.

- Continuous covariates
  - Enter as a numeric columns in the *ddl* for *p*
  - Specify variable name in formula, e.g. $p = list(formula = \sim temperature)$.
- Categorical covariates
  - Enter as a alphanumeric columns in the *ddl* for *p*
  - Specify variable name in formula, e.g. $p = list(formula = \sim snowing)$ where *snowing* is coded as y/n.

```
my.ddl <- make.design(my.data)
my.ddl$p$snowing <- .....
my.ddl$p$temperature <- ....

RMark::mark(my.data, ddl=my.ddl, .....)
```

Single Species; Single-Season - Covariates - *RMark* I

Visit × Site (Sampling)-level covariates. The value of the covariate is applicable to a particular visit at a particular site.
THIS IS MESSY!

- Continuous covariates
  - Need a separate column for each visit names as *CovName1 CovName2 CovName3 ...* in the *input.history*
  - Specify variable name in formula, e.g. $p = list(formula = \sim CovName)$.
- Categorical covariates (MESSY)
  - You must create $K - 1$ indicator variables for the K categories.

Single Species; Single-Season - Covariates - *RMark* II

- Need $K - 1 \times n.visits$ columns labelled as *Cov21 Cov22 Cov22 Cov23* .. where CovIV represents the indicator variable I for visit V for the covariates and I ranges from $2, \ldots, K$ and V ranges from $1, \ldots, n.visits$.
- Specify variable name in formula, e.g. $p = list(formula = \sim Cov2 + Cov3 + Cov4 ...)$.

The *make.time.factor()* is helpful here.

Single Species; Single-Season - Covariates - *RMark*

Mahoenui giant weta (*Deinacrida mahoenui*) is endemic to New Zealand and under stress from rats and other predators.
72 circular plots (3 m radius, primarily prickly gorse plants) were surveyed for weta.
Each plot surveyed 3-5 times.
Covariates to be considered:

- Observer. Three different observers and not every plot surveyed by each observer.
- Browse. Was each site browsed by goats, yes or no.

Single Species; Single-Season - Covariates - *RMark*

Getting the data into *RMark*.

1. Capture History

Get $n_{sites} \times n_{visit}$ data.frame (or matrix) of 1, 0, or NAs

```
input.data <- readxl::read_excel(file.path("..","weta.xls")
                                        sheet="detection_histo
                                        na="-",
                                        col_names=FALSE)   # no
input.history <- data.frame(freq=1,
                            ch=apply(input.data[,1:5],1,pas
input.history$ch <- gsub("NA",".", input.history$ch, fixed=
head(input.history)
```

Detection histories include many missing values. Are these MCAR? Check the detection histories carefully.

Single Species; Single-Season - Covariates - *RMark*

Getting the data into *RMark*.

2. Site Covariates

Get $n_{sites} \times n_{site-covariates}$ data.frame of site covariates.

- Continuous covariates occupy 1 column
- Categorical covariates can either be alpha-numeric code or a set of indicator variables.

With modern software, the former is preferred.

Single Species; Single-Season - Covariates - *RMark*

Getting the data into *RMark*.

2a. Site Covariates categorical and/or continuous) are added to the input.history data frame

```
site_covar <- readxl::read_excel("Weta_pg116.xls",
                                  sheet="site_covar",
                                  na="-",
                                  col_names=TRUE)  # notice

# Create an alternate site level covariate that is a catego
input.history$BrowCat <-factor( site.covar$BrowsCat)
xtabs(~BrowCat, data=input.history, exclude=NULL, na.actio
```

NO MISSING Values allowed for site-level convariates

Single Species; Single-Season - Covariates - *RMark*

Getting the data into *RMark*.

2b. Visit level covariates (same for all sites) entered using *ddl* as seen earlier.

Single Species; Single-Season - Covariates - *RMark*

Getting the data into *RMark*.

3. Visit x Site Covariates - HARD

```
# Get the individual covariates.
obs1 <- readxl::read_excel("Weta_pg116.xls",
                                  sheet="Obs1",
                                  na="-",
                                  col_names=FALSE)
....
```

Single Species; Single-Season - Covariates - *RMark*

Getting the data into *RMark*.
3. Visit x Site Covariates - HARD
Must create (# levels-1) x (# times) covariates in data frame.
E.g. Observer21 Observer31 Observer22 Observer32 Observer23 Observer33 Observer24 Obsever34 Observer25 Observer35
See code in *weta.R*

Single Species; Single-Season - Covariates - *RMark*

Getting the data into *RMark*.
Finally, process the data in the usual way.
NOTE: categorical site-level covariates (e.g. browse category) must be entered as a "group" variable.
Visit level covariates are entered using *ddl*'s. (not shown here).
SitexVisit covariates are tedious.

```
weta.data <- process.data(data=input.history,
                          group="BrowCat",
                          model="Occupancy")
summary(weta.data)
```

Single Species; Single-Season - Covariates - *RMark*

Fit the $\psi(*), p(*)$ model and look at estimates.

```
mod.fit1 <-  RMark::mark(weta.data,
                           model="Occupancy",
                           model.parameters=list(
                             Psi   =list(formula=~1),
                             p      =list(formula=~1)
                           )
)

# look at estimated occupancy probability.
get.real(mod.fit1, "Psi", se=TRUE)

# look at the estimated probability of detection. It gives
get.real(mod.fit1, "p",   se=TRUE)
```

Single Species; Single-Season - Covariates - *RMark*

Fit the $\psi(*)$, $p(*)$ model and look at estimates.

```
# Name : p(~1)Psi(~1)

Npar :  2
-2lnL:  261.7871
AICc :  265.9611

>get.real(mod.fit1, "Psi", se=TRUE)
            all.diff.index par.index  estimate       se
Psi gB a0 t1            11         2 0.6165958 0.088544 0
Psi gN a0 t1            12         2 0.6165958 0.088544 0
                   ucl fixed    note group age time Age Ti
Psi gB a0 t1 0.7701624                    B   0    1   0
Psi gN a0 t1 0.7701624                    N   0    1   0
```

Single Species; Single-Season - Covariates - *RMark*

Fit the $\psi(*), p(*)$ model and look at estimates.

```
> get.real(mod.fit1, "p",   se=TRUE)
              all.diff.index par.index  estimate         se
p gB a0 t1                 1         1 0.3493651 0.0537299 0.2
...
p gB a4 t5                 5         1 0.3493651 0.0537299 0.2
p gN a0 t1                 6         1 0.3493651 0.0537299 0.2
...
p gN a4 t5                10         1 0.3493651 0.0537299 0.2
                    ucl fixed    note group age time Age Time
p gB a0 t1 0.4604481                      B   0    1   0    0
...
p gB a4 t5 0.4604481                      B   4    5   4    4
p gN a0 t1 0.4604481                      N   0    1   0    0
```

Single Species; Single-Season - Covariates - *RMark*

Fitting a model where occupancy varies by browse.
It doesn't make sense to model occupancy as a function of observer - why?

Single Species; Single-Season - Covariates - *RMark*

Fit the $\psi(Browse), p(*)$ model and look at estimates.

```
mod.fit2 <-  RMark::mark(weta.data,
                         model="Occupancy",
                         model.parameters=list(
                           Psi   =list(formula=~BrowCat),
                           p     =list(formula=~1)
                         )
)
```

No need for you to define indicator variables when using site-level categorical variables with *RMark*.

Single Species; Single-Season - Covariates - *RMark*

Fit the $\psi(Browse), p(*)$ model and look at estimates.

```
# Name : p(~1)Psi(~1)

Npar :  3
-2lnL:  258.2643
AICc :  264.6172

>get.real(mod.fit2, "Psi", se=TRUE)
            all.diff.index par.index  estimate        se
             all.diff.index par.index  estimate        se
Psi gB a0 t1             11         2 0.7593708 0.1198262 (
Psi gN a0 t1             12         3 0.4809980 0.1078863 (
                   ucl fixed    note group age time Age Ti
Psi gB a0 t1 0.9194233                    B   0    1   0
```

Single Species; Single-Season - Covariates - *RMark*

Try fitting a model where detectability also depends on browse status of the site. i.e. $\psi(\textit{browse}), p(\textit{browse})$.

```
mod.fit3 <-  RMark::mark(weta.data,
                         model="Occupancy",
                         model.parameters=list(
                           Psi    =list(formula=~BrowCat),
                           p      =list(formula=~BrowCat)
                         )
)
```

Single Species; Single-Season - Covariates - *RMark*

Model $\psi(browse), p(browse)$.
Estimated detection by browse:

```
# Npar :  4
-2lnL:  258.2606
AICc :  266.8576

>get.real(mod.fit3, "p",   se=TRUE)
           all.diff.index par.index  estimate        se
p gB a0 t1              1         1 0.3518919 0.0689239 0.2
...
p gB a4 t5              5         1 0.3518919 0.0689239 0.2
p gN a0 t1              6         2 0.3451663 0.0860159 0.2
...
p gN a4 t5             10         2 0.3451663 0.0860159 0.2
```

Single Species; Single-Season - Covariates - *RMark*

Fitting a model where detectability depends on the observer, but NOT on time $\psi(browse), p(observer)$. Hint: 3 observers need 2 NEW indicator columns. What does the intercept now mean?

Single Species; Single-Season - Covariates - *RMark*

Model $\psi(browse), p(observer)$.

```
mod.fit4 <-  RMark::mark(weta.data,
              model="Occupancy",
              model.parameters=list(
                Psi    =list(formula=~BrowCat),
                p      =list(formula=~Observer2+Observer3)
                        )
)
```

Single Species; Single-Season - Covariates - *RMark*

Model $\psi(browse), p(observer)$.

```
Npar :   5
-2lnL:   252.0421
AICc :   262.9512
```

Single Species; Single-Season - Covariates - *RMark*

Model $\psi(browse), p(observer)$.
What do we get if ask for estimates of p?

```
get.real(mod.fit4, "p", se=TRUE)
               all.diff.index par.index  estimate        se
p gB a0 t1                  1         1 0.3027271 0.0546476 0.2
...
p gB a4 t5                  5         5 0.2921418 0.0547807 0.1
p gN a0 t1                  6         1 0.3027271 0.0546476 0.2
...
p gN a4 t5                 10         5 0.2921418 0.0547807 0.1
                     ucl fixed    note group age time Age Time
p gB a0 t1 0.4189906                     B   0    1   0    0
...
p gB a4 t5 0.4095578                     B   4    5   4    4
```

Single Species; Single-Season - Covariates - *RMark*

Model $\psi(browse), p(observer)$.
We need to estimate the value of p at specified value of the covariates.
What are the parameter index numbers that refer to p?

```
>ddl = make.design.data(weta.data)
>ddl$p # see the index numbers
   par.index model.index group age time Age Time BrowCat
1          1           1     B   0    1   0    0       B
2          2           2     B   1    2   1    1       B
3          3           3     B   2    3   2    2       B
4          4           4     B   3    4   3    3       B
5          5           5     B   4    5   4    4       B
6          6           6     N   0    1   0    0       N
7          7           7     N   1    2   1    1       N
```

Single Species; Single-Season - Covariates - *RMark*

Model $\psi(browse), p(observer)$.
Create the value of the observer variables for which we want predictions

```
obs.df <-data.frame(Observer21=c(0,1,0),
                    Observer22=c(0,1,0),
                    Observer23=c(0,1,0),
                    Observer24=c(0,1,0),
                    Observer25=c(0,1,0),
                    Observer31=c(0,0,1),
                    Observer32=c(0,0,1),
                    Observer33=c(0,0,1),
                    Observer34=c(0,0,1),
                    Observer35=c(0,0,1))
```

Single Species; Single-Season - Covariates - *RMark*

Model $\psi(browse), p(observer)$.
Make predictions

```
>obs.p <- covariate.predictions(mod.fit4, indices=1:10, dat
>obs.p$estimates
   Observer35  estimate          se       lcl       ucl fixe
1           0 0.2235211 0.06273686 0.1241452 0.3689373
...
11          0 0.3786094 0.07998875 0.2383321 0.5426302
...
21          1 0.4462479 0.08081048 0.2980075 0.6047077

30          1 0.4462479 0.08081048 0.2980075 0.6047077
```

Single Species; Single-Season - Covariates - *RMark*

Joint effects of covariates.
Suppose that detectability depended both on occasion effects and observer effects. There are two types of models:

- Additive models. Observers vary among themselves, but are consistent among occasions. For example, one observer has a lower (and consistent) detectability in all occasions even though the detectability varies over occasions. Notation is $p(t + obs)$. Append columns for each covariate.
- Interaction models. Observers are not consistent over occasions. In some days, observer 1 is worst; on other days observer 2 is worst, etc. Notation is $p(t * obs)$. Append columns and then append multiplication of columns.

This is easily done in *RMark* without having to physically create the extra columns using standard modelling notation of *R*.

Single Species; Single-Season - Covariates - *RMark*

Fit the model: $\psi(browse), p(observer + time)$.

```
mod.fit4 <-  RMark::mark(weta.data,
              model="Occupancy",
              model.parameters=list(
                Psi   =list(formula=~BrowCat),
                p     =list(formula=~Observer2+Observer3)
                        )
```

Single Species; Single-Season - Covariates - *RMark*

Fit the following additional models:

- $\psi(browse), p(observer + time + browse)$.
- $\psi(*), p(observer + time + browse)$.
- $\psi(*), p(observer + time)$.

Single Species; Single-Season - Covariates - *RMark*

Construct the AIC table.

```
collect.model(type="Occupancy")
```

Single Species; Single-Season - Covariates - *RMark*

What do you conclude from the AIC table?

```
                                                    Model DAIC    wgt
                                                               model r
5           p(~time + Observer2 + Observer3)Psi(~BrowCat)
8                 p(~time + Observer2 + Observer3)Psi(~1)
7       p(~time + BrowCat + Observer2 + Observer3)Psi(~1)
4                   p(~Observer2 + Observer3)Psi(~BrowCat)
6 p(~time + BrowCat + Observer2 + Observer3)Psi(~BrowCat)
2                                     p(~1)Psi(~BrowCat)
1                                           p(~1)Psi(~1)
3                               p(~BrowCat)Psi(~BrowCat)
  DeltaAICc      weight  Deviance
5 0.0000000 0.32904560  239.5994
8 0.3296085 0.27905031  242.5466
```

Single Species; Single-Season - Covariates - Summary and Cautions

Creating data frames for covariates

- Covariates for ψ are straightforward.
- Covariates for p are tricky:
  - Different order among the packages. Documentation is weak on this point.
  - For all packages except *JAGS*, you can leave missed visits in the data.frame

Try not to have too many covariates with smallish datasets.

Single Species; Single-Season - Covariates - Summary and Cautions

Specifying models.

- *PRESENCE* and *MARK* require you create your own design matrices.
- *RPresence*, *RMark*, *unmarked* have a simple model syntax using categorical variables
- *JAGS* used *model.matrix()* function to help you.

May be difficult to understand what the β parameters mean.

Single Species; Single-Season - Covariates - Final Remarks

Time-varying covariates must be available at all measured sites at all times regardless of detection. Do NOT use covariates that depend on a detection, e.g. type of tree that birds was detected in because no detection implies no covariate.

Numerical difficulties arise if covariate values are too large and too dispersed.

- Subtract a large constant. e.g. use year-2000 as covariate.
- Divide by a large constant, e.g. use ha rather than m^2.
- Shrink the range, e.g. standardize by subtracting mean and dividing by std deviation

These actions do NOT affect AIC etc, but can make final interpretation a bit tricky as you need to unstandardized the final estimates.

Single Species; Single-Season - EXERCISE

Occupancy of American Toads. Extracted from

> *Darryl I. MacKenzie, et al. 2002.*
> *Estimating site occupancy rates when detection probabilities are less than one.*
> *Ecology 83:2248-2255.*
> *doi:10.1890/0012-9658(2002)083[2248:ESORWD]2.0.CO;2]*

29 sites with 82 sampling occasions in 2000.
Volunteers visited sites and recorded presence/absence of toads by calls.
Habitat (type of pond, permanent or ephemeral) and temperature at visit recorded.

Single Species; Single-Season - EXERCISE

29 sites with 82 sampling occasions in 2000.

Volunteers visited sites and recorded presence/absence of toads by calls.

Lots of missing data!

```
---------------------0---0-----------------0---0-----------0----------------0-----
-----0--0-------------1---------------------------0-------------------------0-----
-----------------------0-------------------0---0-----------0----------------0-----
---0------0------1------1------0---0--------0------0-------0---0--------0----0----
-----------------0--------------0------------0------0-------------00--------------
-------------------------1----------------------0--------------------------0------
10000001000000----101001-11110--0000---0000---1000000---010000000--00000--0011----
0--------------------00---------0-0--------------0-------------0---0------------0-
-0------------------0--------0-------------------------------0------------------0-
------0--------1----------1-------------------------0------0-------0--------------
------------------------------------------1-----------0---------0----------------0
----------------0-------0-----------0--------0----0-----0----0-----------0-------0
---------------------0-------------------0----------------------------------0-----
------0-------0--------------------0--0----------0--0----------0-----------------0
-----0--------0-0-----00---0---------0--0------0------0--0------------0-----------
--------0----------------------------------------------0-------------0------------
------0---0----0--------0----0-------------0-------0------------0-----------------
----------0-----0--0---0----------------------0----------0-------------0----------
-------------------------0--------------------0-----------------------0-----------
----------0--------------1-------------------------1--------0---------------0-----
---0---------------1----------------------0---------------------0------------0----
0-------------0-------------0--------------------0--------------0-------------0---
----------------0---------------------0---------0---------------0----------0------
```

Single Species; Single-Season - EXERCISE

Because of the spareness of the data, fit the following models:

- $\psi(Habitat), p(Temperature)$
- $\psi(*), p(Temperature)$
- $\psi(Habitat), p(*)$
- $\psi(*), p(*)$

CAUTION: How many columns do you need for each covariate (why?)

Single Species; Single-Season - EXERCISE

Final Results:

Model	AIC	deltaAIC	AIC wgt	Model Likeli	no.Par.	-2*LogLike
psi(Habitat),p(Temp)	183.16	0.00	0.3032	1.0000	4	175.16
psi(*),p(Temp)	183.58	0.42	0.2458	0.8106	3	177.58
psi(Habitat),p(*)	183.65	0.49	0.2373	0.7827	3	177.65
1 group, Constant P	183.86	0.70	0.2137	0.7047	2	179.86

What is total model weight for habitat effects? Temperature effects?

Single Species; Single-Season - EXERCISE

Brook trout: 77 streams, 3 segments/stream.
Collected via electrofishing three 50 m sections of streams at sites in the Upper Chattachochee River basin.

Covariates.

- elevation,
- cross sectional area each occasion

What effect (if any) of elevation on occupancy?
Do $p(t)$ (or equivalent) models make sense in this setting?

Single Species; Single-Season - Grossbeaks

An occupancy study was made on Blue Grosbeaks (*Guiraca caerulea*) on 41 old fields planted to longleaf pines (*Pinus palustris*) in southern Georgia, USA.

Surveys were 500 m transects across each field and were completed three times during the breeding season in 2001.

Columns in the file are:

- *field* - field number
- *v1, v2, v3* - detection histories for each site on each of 3 visit during the 2001 breeding season.
- *field.size* - size of the files
- *bqi* - Enrollment in bobwhite quail initiative; does occupancy increase if field belongs to this initiative?
- *crop.hist* - crop history
- *crop1, crop2* - indicator variables for the crop history

Single Species; Single-Season - Exercise

Fit models using the covariates given in the file.