lecture6

Reading Assignment:
Lebreton et al. (1992) *Ecological Monograph*

Two good references on likelihood methods:
Azzalini, A. 1996. Statistical inference based on the likelihood. Chapman and Hall.

Royall, R. M. 1997. Statistical evidence: a likelihood paradigm. Chapman and Hall.

Relationship of PIMs, Design Matrix, and Estimates

The PIM matrices define the number of parameters that can be estimated. The number of parameters determines the number of rows in the design matrix. For an identity design matrix, the number of rows and the number of columns is identical, so that there is a one to one correspondence between the columns of the design matrix, and the estimates. However, the design matrix allows additional constraints to be placed on the parameters specified in the PIMs. Hence, the number of columns in the design matrix specifies the number of actual parameters that are estimated, which are the β's. The link function converts the β's of the desing matrix into the real parameter estimates.

The relationship looks something like this.

PIM $\Longrightarrow$ Design Matrix $\Longrightarrow$ Link Function $\Longrightarrow$ Real Parameter Estimates

Modeling Survival or Reporting Probabilities

We begin by considering model $\{S_t, r\}$. If one has banding and recovery data over, say, 15 years (or occasions) and has 15 MLEs of the conditional survival probabilities, questions arise as to what caused them to vary or, at least, what factors might have been associated with this variation. The investigator remembers that 1984 was a very bad winter and that 1992 and 1993 seemed very ideal. Thus, prior to data analysis one might hypothesize one or more covariates as been associated with conditional survival (or reporting) probabilities.

The first approach might be to plot the MLEs ($\hat{S}_j$) against the covariate (say, P = precipitation defined in some biologically meaningful way). Then, if not thinking very clearly, one might consider a least squares regression of the $\hat{S}_j$ against P_j. Therefore, the model

$$S_j = \beta_0 + \beta_1(P) \pm \epsilon_j$$

This is poor procedure as it violates most of the assumptions about "regression." For example, the ϵ_j (the residuals) are not independent, not normal (except asymptotically), and do not have constant variance. Dependence is established by the sampling variance-covariance matrix of the model parameters. The errors are binomial/multinomial; not particularly normal. The sampling variances are hardly constant. Thus, least squares "regression" will not provide estimates of the intercept (β_0) and slope (β_1) parameters because important assumptions have been violated. **In addition**, the predicted survival probability for an extreme value of precipitation might be >1! This is not desirable.

The solution to these issues is to embed the submodel

$$\beta_0 + \beta_1(P) \pm \epsilon_j$$

in the log-likelihood function. That is, substitute the expression $\beta_0 + \beta_1(P_1)$ everywhere in the log-likelihood function where S_1 is found. Then, substitute $\beta_0 + \beta_1(P_2)$ everywhere in the log-likelihood function were S_2 is found, etc., etc.

When this process has been completed, the log-likelihood has no conditional survival parameters left and only the parameters β_0 and β_1 remain (plus the reporting probability, r). Thus, $K = 3$ (instead of 16).

Thus, numerical methods can be used to get the MLEs $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{r}$; These are the values that maximize the log-likelihood function. Of course, their associated variance-covariance matrix can be gotten. The parameter estimate $\hat{\beta}_1$ provides information about the positive or negative slope between S_j and P_j in terms of the assumed linear relationship. Looking at the confidence interval around $\hat{\beta}_1$ provides some rough indication of the evidence for a relationship between these variables (or parameters). $QAIC_c$ will sort out the support for this model in relation to others in the set to be considered.

Most interesting is the fact that the estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ and the assumed linear structure can be used to derive estimates of the annual, conditional survival parameters (the S_j). Furthermore, these estimates will have a smaller sampling variance that those under the model $\{S_t, r\}$! What? Two reasons: (1) additional *information* has been provided in the form of the covariate P_j, and (2) a submodel has been imposed (assumed). Such estimates of survival probabilities are termed **derived parameters**. The basic model parameters are merely β_0, β_1 and r where $K = 3$. The 15 estimates of survival probability are **derived** from these.

There are two reasons for using such a submodel: (1) statistical reasons and the Principle of Parsimony and (2) biologists want to explore associations between estimated parameters and external covariates. Still, such an approach does not keep the derived survival probabilities ≤ 1. The solution to this issue is to employ a nonlinear **link function**.

Link Functions and Covariates

There are several choices of link functions that provide increased biological realism over a linear model and constrain the estimated model probabilities to the interval [0,1]. The workhorse for this class of models is the logistic (sub) model.

Logistic:

$$S_j = \frac{1}{1+e^{-[\beta_0+\beta_1(H_j)]}} = \frac{1}{1+\exp(-[\beta_0+\beta_1(H_j)])}$$

This submodel recognizes that the survival parameter is a probability, hence bounded between 0 and 1. Some other choices for link functions are summarized below:

Function	Link
Logistic: $S_j = \frac{1}{1+\exp(-[\beta_0+\beta_1(H_j)])}$	$\log_e\left(S_j/(1-S_j)\right) = \beta_0 + \beta_1(H_j)$
Log: $S_j = \exp(\{\beta_0+\beta_1(H_j)\})$	$\log_e(S_j) = \beta_0 + \beta_1(H_j)$
Log-log: $S_j = \exp[-\exp(\{\beta_0+\beta_1(H_j)\})]$	$\log_e\left(-\log_e(S_j)\right) = \beta_0 + \beta_1(H_j)$
Complementary log-log: $S_j = 1 - \exp[-\exp(\{\beta_0+\beta_1(H_j)\})]$	$\log_e\left(-\log_e(1-S_j)\right) = \beta_0 + \beta_1(H_j)$
Sin: $S_j = [\sin(\{\beta_0+\beta_1(H_j)\}) + 1]/2$	$\arcsin(2S_j-1) = \beta_0 + \beta_1(H_j)$
Identity: $S_j = \beta_0 + \beta_1(H_j)$	$S_j = \beta_0 + \beta_1(H_j)$

Either S or r can be modeled using a "link function." This is particularly useful when analyzing multiple groups. Note, the introductory material (above) dealt what we will now call the "identity link" in that no transformation is made. This has several disadvantages.

We begin by considering a logistic (sub)model of survival probabilities as a function of precipitation P;

$$S_j = \frac{1}{1+e^{-[\beta_0 + \beta_1(P_j)]}} \quad .$$

Conceptually, this expression is substituted in the log-likelihood for all values of S_j. In terms of modeling, it is most convenient to consider the logistic transform = the link function,

$$\text{logit}(S) = \log_e\left(\frac{S}{1-S}\right) = \beta_0 + \beta_1(P).$$

This is the linear model which people seem to be familiar. This approach can be extended. For example, perhaps one hypothesizes that conditional survival probabilities are associated with precipitation but varies by gender (G); then,

$$\text{logit}(S) = \log_e\left(\frac{S}{1-S}\right) = \beta_0 + \beta_1(P) + \beta_2(G) \,,$$

where, now P is a continuous covariate while G is an instrumental (or dummy) variable taking values of 1 or males and 0 or females (or vice versa, if one wants). This submodel allows a linear relationship between logit(S) and P with males and females being parallel; the males being higher by an amount $\hat{\beta}_2$ (or lower if this estimated parameter is negative). Already, considerable flexibility can be achieved in modeling survival (or reporting) probabilities.

Consider the further extensions,

$$\text{logit}(S) = \log_e\left(\frac{S}{1-S}\right) = \beta_0 + \beta_1(P) + \beta_2(G) + \beta_3(W) + \beta_4(G*W),$$

where, p = precipitation (continuous)
G = gender (1 = male, 0 = female) (2 "groups")
W = winter weather (0 = above average, 1 = below average) (2 "groups")
$G*W$ = interaction term between gender and winter weather.

Note, the variables G and W are discrete (0 or 1) variables while P is a continuous covariate. In this case, one might want to consider the model without the interaction term.

Similar modeling can be done with the other link functions and allows great flexibility in modeling.

Consider 2 groups, males and females:

Males

		m_{ij}			
2000	30	70	114	43	15
2000		80	97	55	19
2000			169	46	10
2000				72	24
2000					34

Females

		m_{ij}			
2200	28	66	111	38	11
2200		71	91	50	12
2200			149	41	9
2200				69	20
2200					31

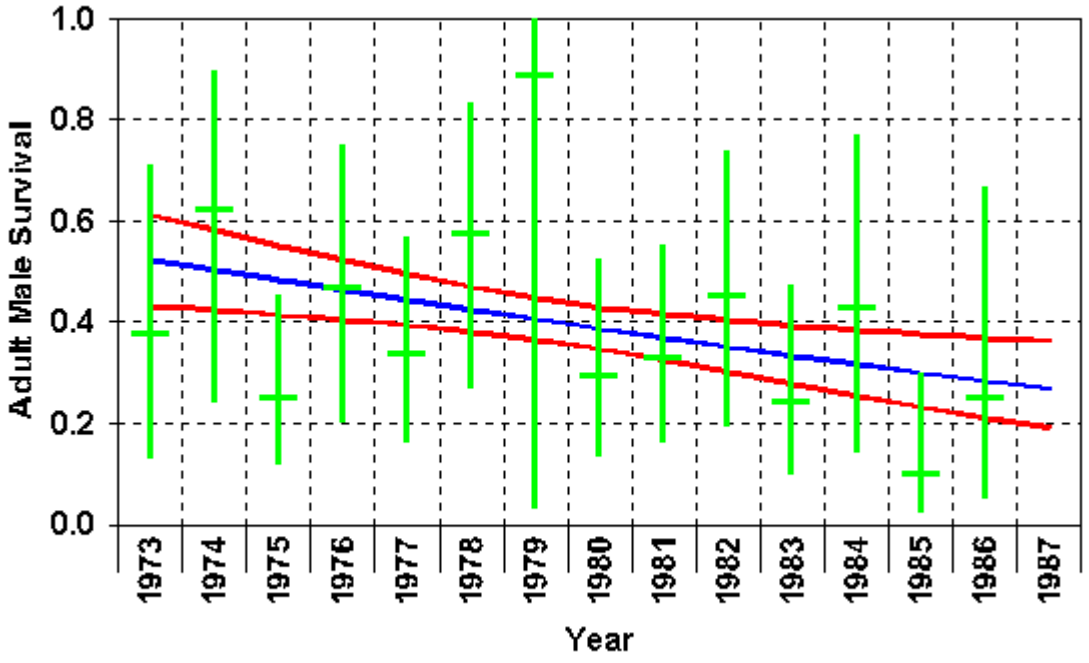
A number of models could be considered for the parsimonious analysis of these data using link functions. Other examples could include tag recovery data for 3 areas, or k age classes, or combinations of age, gender, and area.

Models for formal treatment vs. control experiments represent a particluarly important class of models. Here some of the animals are given a treatment, with the other animals serving as a control. Of course, the assignment of animals to the two groups must be randomized. The AFS Monograph #5 by Burnham et al. (1987) treats these cases.

Link functions and a continuous covariate can be used to model long term trends in survival probabilities. Here the logistic link model is

$$\log_e\left(S_j/(1\text{-}S_j)\right) = \beta_0 + \beta_1(T_j)$$

where T is year, 1, 2, 3, ..., 15 in the plot shown below. The trend is linear on logit (S), and nearly linear (over a narrow range) on S.



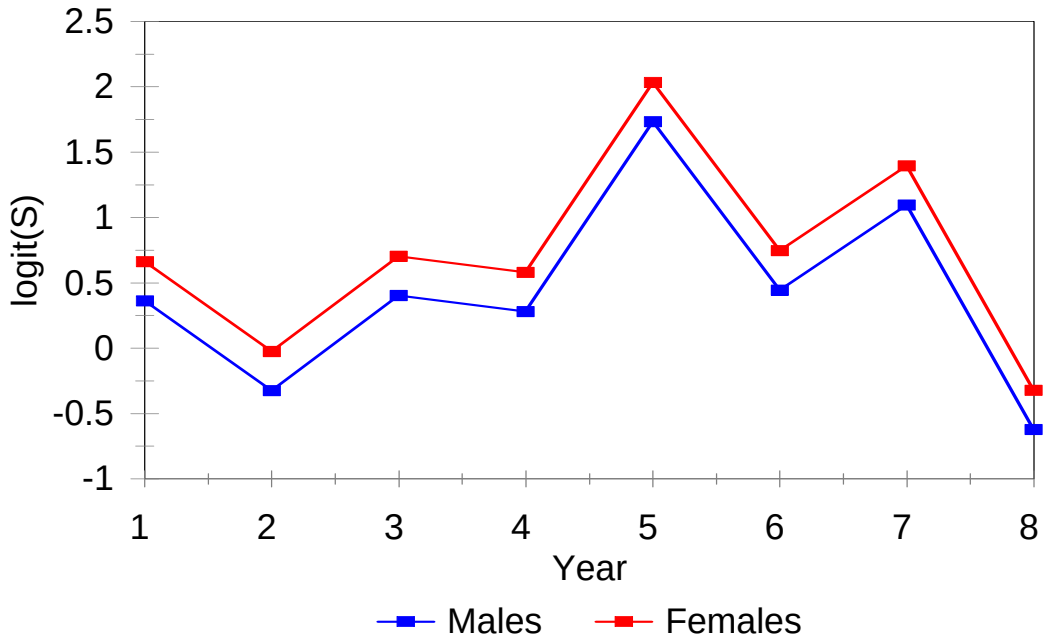
Also shown are the estimates from the year-specific model.
Question: Which of these two models is better?
Why? in what sense “better"?

Often there is a hypothesized “parallelism" in survival of the sexes, across time. This is easy to model,

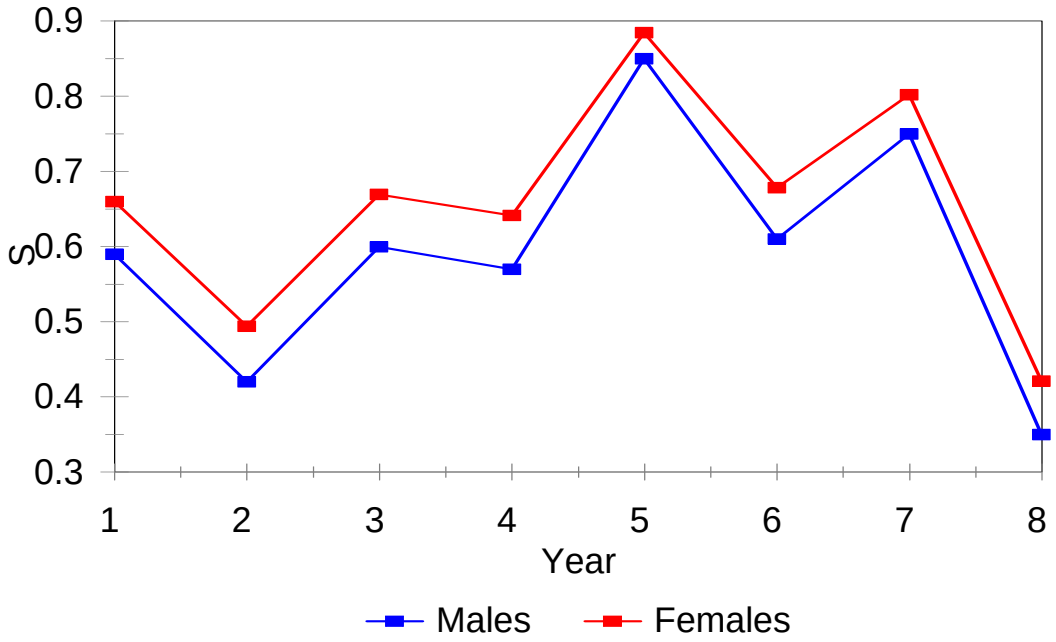
$$\log_e\left(S_j/(1\text{-}S_j)\right) = \beta_0 + \beta_1(G) + \beta_2(t_1) + \beta_3(t_2) \ldots$$

where G is a dummy variable (say, 0=male and 1=female) and the t_i is a set of dummy variables to index year. (Design matrix)

Model is $\{S_{g+t}, r\}$, to indicate that gender is an additive component to the year-specific variation in survival across years. On the logit scale, the estimates are parrallel for the 2 sexes.



On the back-transformed survival rates, the values of S are not parrallel



Modeling the reporting (sampling) probabilities as r_t (i.e., year-specific) requires many parameters. If there are 11 years of recovery data, one would need to estimate about 11 parameters for just the recovery probabilities,

$$r_1,\ r_2, r_3, \ldots r_{11}.$$

This is poor practice if some data on tagging effort is available. e. g., number of tagging days, number of person hours, number of nets used, etc. If such data on effort (E) is available, one can model the recovery probabilities as a function of effort as

$$\text{logit}(r) = \log_e\left(r/(1\text{-}r)\right) = \beta_0 + \beta_1(E).$$

Thus, only 2 parameters are required, instead of 11.

Question: is this approach better, given a particular set of data? Which model *should* be used for inference?

Assume a data set, $t = 4$ release occasions and 4 recovery years; data might be

R(i)	*m*(i,j)			
2448	112	65	90	68
2545		81	142	96
1205			63	55
1511				92

For model $\{S_t, r_t\}$, the (real) parameters are

$S_1,\ S_2,\ S_3,\ S_4,$

and

$r_1,\ r_2,\ r_3,\ r_4$

(with only the product $(1 - S_4)\, r_4$ estimable).

MARK fits all models via a *design matrix* linking these "real" parameters to regression-type "beta" parameters. For example, with a 1-1 link function between these two types of parameters, if we wanted a reduced parameter model with $S_1 = S_2$, $S_3 = S_4$, r_1, $r_2 = r_3 = r_4$ we could specific this model via the design matrix approach using equations

$$
\begin{aligned}
S_1 = \beta_1 \quad &= 1 \cdot \beta_1 + 0 \cdot \beta_2 + 0 \cdot \beta_3 + 0 \cdot \beta_4 \\
S_2 = \beta_1 \quad &= 1 \cdot \beta_1 + 0 \cdot \beta_2 + 0 \cdot \beta_3 + 0 \cdot \beta_4 \\
S_3 = \beta_2 \quad &= 0 \cdot \beta_1 + 1 \cdot \beta_2 + 0 \cdot \beta_3 + 0 \cdot \beta_4 \\
S_4 = \beta_2 \quad &= 0 \cdot \beta_1 + 1 \cdot \beta_2 + 0 \cdot \beta_3 + 0 \cdot \beta_4 \\
r_1 = \beta_3 \quad &= 0 \cdot \beta_1 + 0 \cdot \beta_2 + 1 \cdot \beta_3 + 0 \cdot \beta_4 \\
r_2 = \beta_4 \quad &= 0 \cdot \beta_1 + 0 \cdot \beta_2 + 0 \cdot \beta_3 + 1 \cdot \beta_4 \\
r_3 = \beta_4 \quad &= 0 \cdot \beta_1 + 0 \cdot \beta_2 + 0 \cdot \beta_3 + 1 \cdot \beta_4 \\
r_4 = \beta_4 \quad &= 0 \cdot \beta_1 + 0 \cdot \beta_2 + 0 \cdot \beta_3 + 1 \cdot \beta_4
\end{aligned}
$$

This set of equations provides a model wherein the 8 apparent *S* and *r* parameters are in fact produced by only 4 ($= K$) parameters

An easier to "read" symbolic representation of this model-form as imposed on the otherwise time-specific *S* and *r* parameters is

	β_1	β_2	β_3	β_4
S_1	= 1	0	0	0
S_2	= 1	0	0	0
S_3	= 0	1	0	0
S_4	= 0	1	0	0
r_1	= 0	0	1	0
r_2	= 0	0	0	1
r_3	= 0	0	0	1
r_4	= 0	0	0	1

The 8 (rows) by 4 (columns) matrix of known constants, ones and zeros here, is the *design matrix* for the reduced model starting from the full {$S(t)$, $r(t)$} model. In matrix notation: parameter vector $= \mathrm{X}\underline{\beta}$. The number of columns of the design matrix corresponds exactly to the number of β parameters. As a matrix in this example, the design matrix is simple

$$\mathrm{X} = \begin{matrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{matrix}$$

So, to specify a reduced (i.e., simpler) model to *MARK* based on a given easily-specified general model (in terms of PIMs), it suffices to “tell" *MARK* the desired design matrix, i.e., the known coefficients in some generalized linear relationship between each real parameter, S or r, and the beta parameters. For example, if a logit link is used then implicitly we are using relationships like

$$\log\left(\frac{S_1}{1-S_1}\right) = 1\cdot\beta_1 + 0\cdot\beta_2 + 0\cdot\beta_3 + 0\cdot\beta_4 \equiv \beta_1$$

or

$$\log\left(\frac{r_3}{1-r_3}\right) = 0\cdot\beta_1 + 0\cdot\beta_2 + 0\cdot\beta_3 + 1\cdot\beta_4 \equiv \beta_4$$

Hence, in the likelihood everywhere S_1 appears we would replace it by

$$S_1 = \frac{1}{1+\mathrm{e}^{-(1\cdot\beta_1+0\cdot\beta_2+0\cdot\beta_3+0\cdot\beta_4)}}$$

or

$$r_3 = \frac{1}{1+\mathrm{e}^{-(0\cdot\beta_1+0\cdot\beta_2+0\cdot\beta_3+1\cdot\beta_4)}}$$

The link function and the design matrix can (and are) specified independently. You can have the same design matrix with any link function in *MARK*.

In the above example the “effects" on the *S* and *r* are qualitative (like an ANOVA). We can have more regression-like models if we have more typical covariates, such as harvest pressure, or environmental variables that might affect survival, or sampling effort that might affect reporting or capture rates. For example, let x_i be an annual index of harvest pressure, such as $x_1 = 3.5, \quad x_2 = 4.2, \quad x_3 = 7.1, \quad x_4 = 6.5$. Now we can use the regression-like model

$$\log\left(\frac{S_i}{1-S_i}\right) = \beta_1 + x_i \cdot \beta_2$$

Keeping the same model for the r_i as above, the design matrix is

	β_1	β_2	β_3	β_4
S_1	= 1	3.5	0	0
S_2	= 1	4.2	0	0
S_3	= 1	7.6	0	0
S_4	= 1	6.5	0	0
r_1	= 0	0	1	0
r_2	= 0	0	0	1
r_3	= 0	0	0	1
r_4	= 0	0	0	1

If we had a covariate y_i linearly related to as r_i as

$$\log\left(\frac{r_i}{1-r_i}\right) = \beta_3 + y_i \cdot \beta_4$$

we could use a model, via the design matrix, below:

1	x_1	0	0
1	x_2	0	0
1	x_3	0	0
1	x_4	0	0
0	0	1	y_1
0	0	1	y_2
0	0	1	y_3
0	0	1	y_4

The first reduced model above ($K = 4$, qualitative effects) can be specified to *MARK* using PIMs. The PIMs for *S* and *r* would look like (respectively)

```
1  1  2  2          3  4  4  4
   1  2  2             4  4  4
      2  2                4  4
         2                   4
```

There are now 4 real parameters and *MARK* would use, by default, an identity design matrix. But there would be a design matrix – there is always a design matrix.

We could take for this example data the default PIMs below

```
1  2  3  4          5  6  7  8
   2  3  4             6  7  8
      3  4                7  8
         4                   8
```

Then we specify to *MARK* the 8 by 4 design matrix for the reduced “ANOVA" like model (i.e., $S_1 = S_2$, etc.). However, here the regression like model must be specified by starting with the general model and then creating the underlying design matrix for *MARK* to use.

For covariate-based models you must use the design matrix option to specify the model.

There is another type of time varying “covariate" model: smooth trends or patterns in the temporal sequence of survival (or reporting rate) parameters. You might think the annual survival rates are trending down (or up) in a linear fashion. The model to use is then

$$\log\left(\frac{S_i}{1-S_i}\right) = \beta_1 + i \cdot \beta_2$$

(or an identity link). (This is a {*S*(T)} model). For this trend model and for unrestricted r_i the design matrix is

```
1  1  0  0  0  0
1  2  0  0  0  0
1  3  0  0  0  0
1  4  0  0  0  0
0  0  1  0  0  0
0  0  0  1  0  0
0  0  0  0  1  0
0  0  0  0  0  1
```

If we also want to have a model with a linear trend on the r_i, hence

$$\log\left(\frac{r_i}{1-r_i}\right) = \beta_3 + \mathrm{i}\cdot\beta_4$$

the design matrix must be

1	1	0	0
1	2	0	0
1	3	0	0
1	4	0	0
0	0	1	1
0	0	1	2
0	0	1	3
0	0	1	4

This is model {*S*(T), *r*(T)}.

There is another type of covariate we can include in our models of survival and reporting rates: individual covariates, such as weight on release, or some continuous index of fitness. (The above examples are temporally varying covariates that effect all animals equally). Individual covariates included in the data are input with individual encounter histories, such as

```
1  2  3  4
LD LD LD LD      size  sex  site
10 00 01 00  1    30    1    1
00 11 00 00  1    18    0    2
00 10 00 01  1    23    1    1
00 00 10 00  1    15    0    2
```

If we now fit a model with full time effects and one individual covariate applied to both *S* and *r* there would still be the 8 “real" parameters (S_1, S_2, S_3, S_4, r_1, r_2, r_3, r_4) implicitly in the likelihoods (i.e., models). However, the full specification of the model takes 10 actual “beta" parameters: β_1 to β_{10}. The design matrix as specified to *MARK* looks like

	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}
S_1	1	0	0	0	0	0	0	0	size	0
S_2	0	1	0	0	0	0	0	0	size	0
S_3	0	0	1	0	0	0	0	0	size	0
S_4	0	0	0	1	0	0	0	0	size	0
r_1	0	0	0	0	1	0	0	0	0	size
r_2	0	0	0	0	0	1	0	0	0	size
r_3	0	0	0	0	0	0	1	0	0	size
r_4	0	0	0	0	0	0	0	1	0	size

This is only one possible ordering of the design matrix columns. There is no intrinsically required ordering.

You do not build a row in the design matrix for every animal. Instead, with models that include individual covariates *MARK* is told it is such a model type. Also *MARK* is told the names of the covariates that are with the individual encounter histories. So MARK knows to build the likelihoods history-by-history. Ignoring sex and site as factors in the model here, for the first animal the likelihood contribution is

$$\mathcal{L}_{\text{animal(1)}} = S_1 S_2 (1 - S_3)\, r_3$$

with (for logit link) relationships

$$\log\left(\frac{S_1}{1-S_1}\right) = \beta_1 + 30 \cdot \beta_9$$

$$\log\left(\frac{S_2}{1-S_2}\right) = \beta_2 + 30 \cdot \beta_9$$

$$\log\left(\frac{S_3}{1-S_3}\right) = \beta_3 + 30 \cdot \beta_9$$

$$\log\left(\frac{r_3}{1-r_3}\right) = \beta_7 + 30 \cdot \beta_{10}$$

The real strength of certain types of parsimonious models that have to be expressed in design matrix form arise when there are multiple data sets or what is essentially the same thing, multiple factors possible effecting survival and reporting rate parameters. Hence a context of

Multiple Factors and/or Data Sets

time
age
gender
area (capture sites) ← group effects
control/treatment
sampling effort
environmental factors
regulations

We will now focus on the interplay of time and group factors in building models. There are models such as

{*S*(g*t), r(g*t)} ≡ a separate {S(t), r(t)} model for each group
{*S*(t), r(g)}
{*S*(g), r(t)}
{*S*(g+t), r(g+t)}
{*S*(g+t), r(g)}
{*S*(t), r(t)}
{*S*(g), r(T)}
{*S*(g+T), r(g+T)}
{*S*(g+T), r(t)}
{*S*(T), r(T)}
{*S*(g), r(g)}
{*S*(.), r(.)}

We focus on the additive effects models, such as {*S*(g+t), r(g + t)}, which is definable to *MARK* by its design matrix. We must now have at least two groups; I will assume two groups and first focus on the part of the design matrix for survival rate.

S_{m1}	1	1	0	0	1	0	0
S_{m2}	1	0	1	0	1	0	0
S_{m3}	1	0	0	1	1	0	0
S_{m4}	1	0	0	0	1	0	0
S_{f1}	1	1	0	0	0	0	0
S_{f2}	1	0	1	0	0	0	0
S_{f3}	1	0	0	1	0	0	0
S_{f4}	1	0	0	0	0	0	0
r_{m1}	0	0	0	0	0	1	1
r_{m2}	0	0	0	0	0	1	1
r_{m3}	0	0	0	0	0	1	1
r_{m4}	0	0	0	0	0	1	1
r_{f1}	0	0	0	0	0	1	0
r_{f2}	0	0	0	0	0	1	0
r_{f3}	0	0	0	0	0	1	0
r_{f4}	0	0	0	0	0	1	0

The above design matrix is for model {*S*(g + t), r(g)}, for two groups (female and male at one tagging site for four years).

This model has $K = 7$ intrinsic parameters (the β) even though it has (nominally) 16 real parameters. The most general model here would have $K = 14$ estimable parameters. If the reduce model is suitable we should prefer it better represents the information in the data and because it will provide more precise estimates of the survival and reporting rate parameters.

These additive effects models can be a lot more parsimonious than the effects-saturated model ({*S*(g*t), *r*(g*t)}). Consider the comparison in terms of *K* for $t = 8$ and $g = 4$:

model {*S*(g*t), *r*(g*t)} then $K = 60$
model {*S*(g+t), *r*(g+t)} then $K = 21$
model {*S*(g+t), *r*(g)} then $K = 15$
model {*S*(g+T), *r*(g)} then $K = 9$

Consider the design matrix for model {*S*(g+T), *r*(g)} for 3 groups and 5 times. The model for the survival rates is like an analysis of covariance where there is one common slope but each group has its own intercept. The design matrix can be given as

S(1,1)	1	0	0	1	0	0	0
S(1,2)	1	0	0	2	0	0	0
S(1,3)	1	0	0	3	0	0	0
S(1,4)	1	0	0	4	0	0	0
S(1,5)	1	0	0	5	0	0	0
S(2,1)	0	1	0	1	0	0	0
S(2,2)	0	1	0	2	0	0	0
S(2,3)	0	1	0	3	0	0	0
S(2,4)	0	1	0	4	0	0	0
S(2,5)	0	1	0	5	0	0	0
S(3,1)	0	0	1	1	0	0	0
S(3,2)	0	0	1	2	0	0	0
S(3,3)	0	0	1	3	0	0	0
S(3,4)	0	0	1	4	0	0	0
S(3,5)	0	0	1	5	0	0	0
r(1,1)	0	0	0	0	1	0	0
r(1,2)	0	0	0	0	1	0	0
r(1,3)	0	0	0	0	1	0	0
r(1,4)	0	0	0	0	1	0	0
r(1,5)	0	0	0	0	1	0	0
r(1,1)	0	0	0	0	0	1	0
r(1,2)	0	0	0	0	0	1	0
r(1,3)	0	0	0	0	0	1	0
r(1,4)	0	0	0	0	0	1	0
r(1,5)	0	0	0	0	0	1	0
r(1,1)	0	0	0	0	0	0	1
r(1,2)	0	0	0	0	0	0	1
r(1,3)	0	0	0	0	0	0	1
r(1,4)	0	0	0	0	0	0	1
r(1,5)	0	0	0	0	0	0	1

There is an easier way to built this model for *MARK*. Let the PIMs for *S* be indexed by default as below

```
 1  2  3  4  5      6  7  8  9 10     11 12 13 14 15
    2  3  4  5         7  8  9 10        12 13 14 15
       3  4  5            8  9 10           13 14 15
          4  5               9 10              14 15
             5                  10                 15
```

Then initialize the three PIMs for r each to a constant:

```
16 16 16 16 16   17 17 17 17 17   18 18 18 18 18
   16 16 16 16      17 17 17 17      18 18 18 18
      16 16 16         17 17 17         18 18 18
         16 16            17 17            18 18
            16               17               18
```

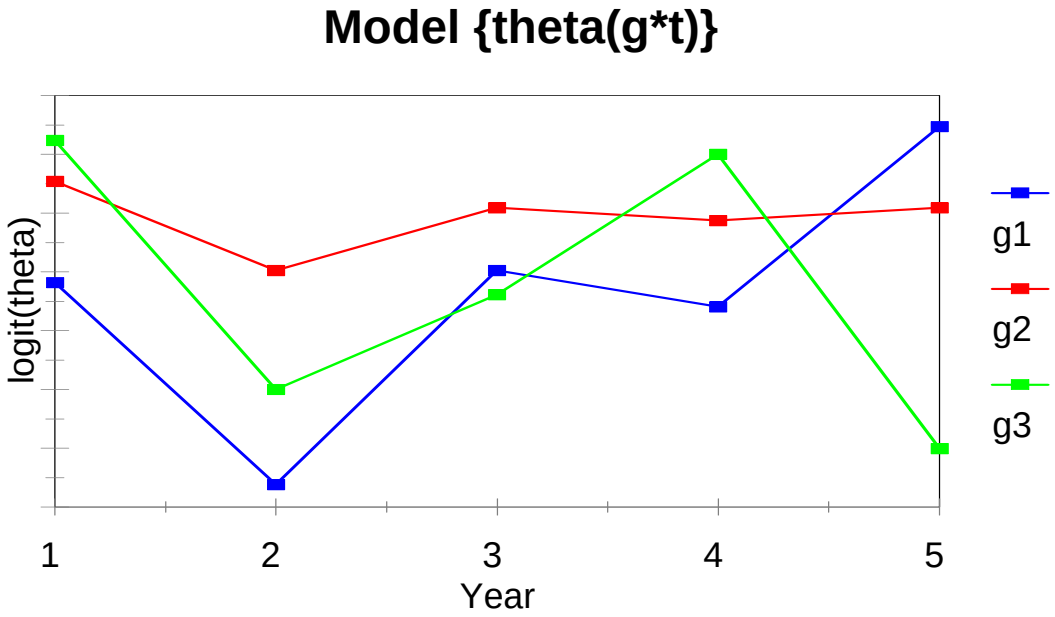
Now *MARK* “knows" there are only three reporting rates and the design matrix *for the same model* {*S*(g+T), *r*(g)} is as below:

S(1,1)	1	0	0	1	0	0	0
S(1,2)	1	0	0	2	0	0	0
S(1,3)	1	0	0	3	0	0	0
S(1,4)	1	0	0	4	0	0	0
S(1,5)	1	0	0	5	0	0	0
S(2,1)	0	1	0	1	0	0	0
S(2,2)	0	1	0	2	0	0	0
S(2,3)	0	1	0	3	0	0	0
S(2,4)	0	1	0	4	0	0	0
S(2,5)	0	1	0	5	0	0	0
S(3,1)	0	0	1	1	0	0	0
S(3,2)	0	0	1	2	0	0	0
S(3,3)	0	0	1	3	0	0	0
S(3,4)	0	0	1	4	0	0	0
S(3,5)	0	0	1	5	0	0	0
r(1)	0	0	0	0	1	0	0
r(2)	0	0	0	0	0	1	0
r(3)	0	0	0	0	0	0	1

For many reduced parameter group and time models there is no unique way to represent the design matrix to *MARK*, even though the fitted model is unique (i.e., the $\hat{S}$'s and $\hat{r}$'s you get will be unique for the data).

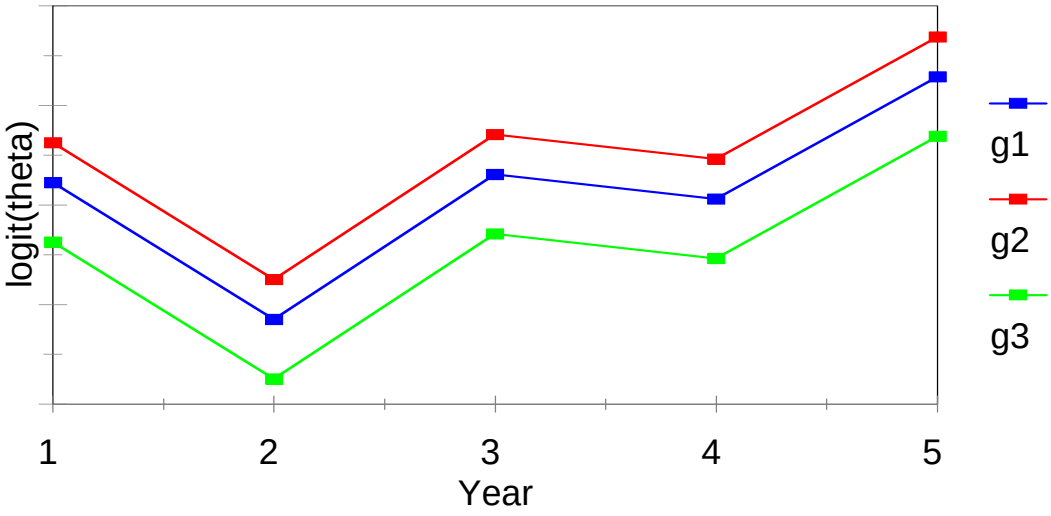
Now, a few summary statements:
1. *MARK* always bases the fitted model on a design matrix.
2. The default is an identity design matrix given the parameters established by the PIMs.
3. Otherwise, you first set up the PIMs, then further constrain, or define, the model by using a design matrix.
4. Models fit by *MARK* are unique as regards the set of identifiable *S* and *r*, but the underlying design matrix need not be unique (this makes it hard to teach principles of constructing design matrices).
5. Additive effects models (say g + t effects) are basically setting interactions (g × t) to zero; this is often the basis of a good *model* with limited data (ala parsimony) (a full effects model is group effects + time effects + all interactions ≡ a g*t model)

Graphical examples of various models:



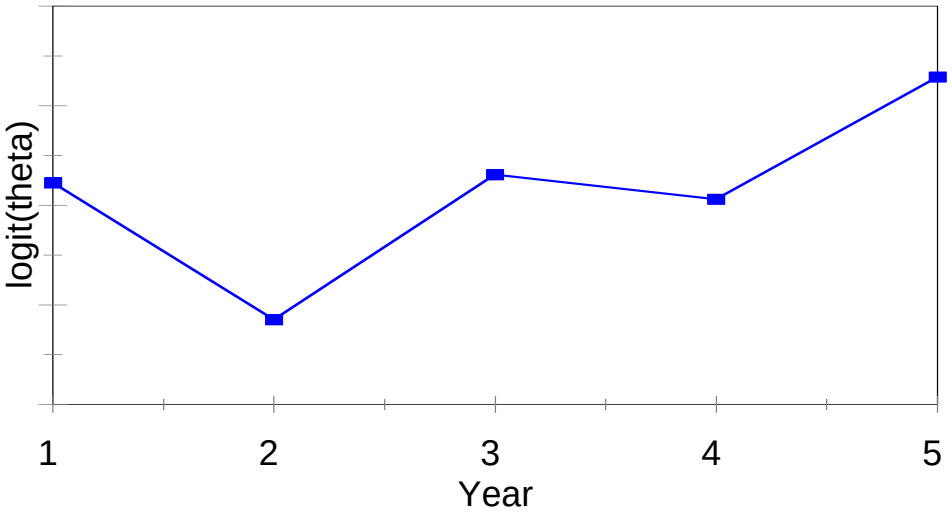
Variation in theta by time and by group

Model {theta(g+t)}



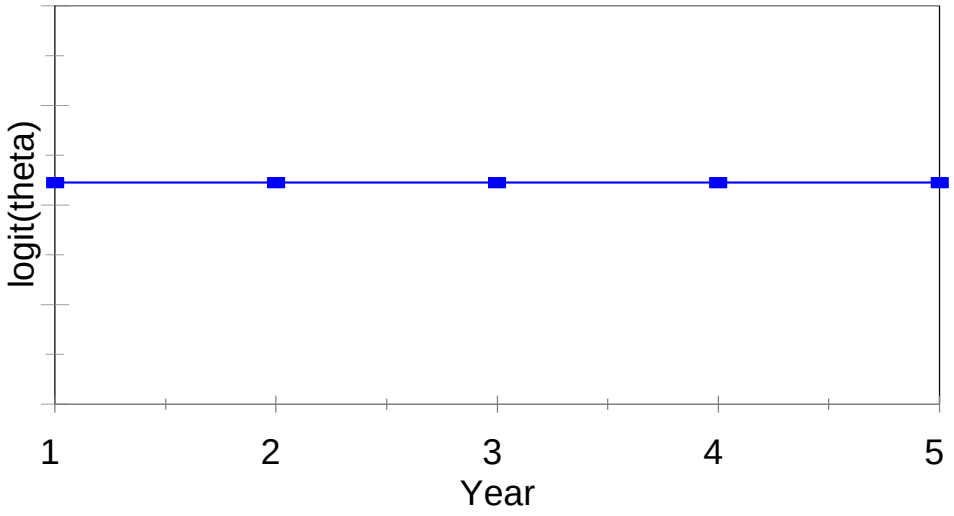
Variation in theta by time and by group
Additive model with no interactions

Model {theta(t)}



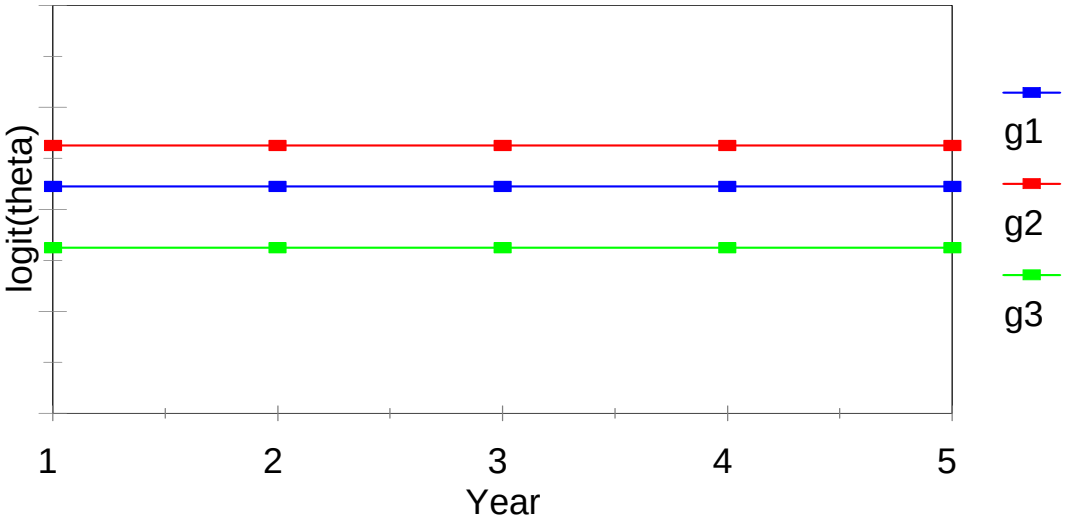
Variation in theta by time

Model {theta(.)}



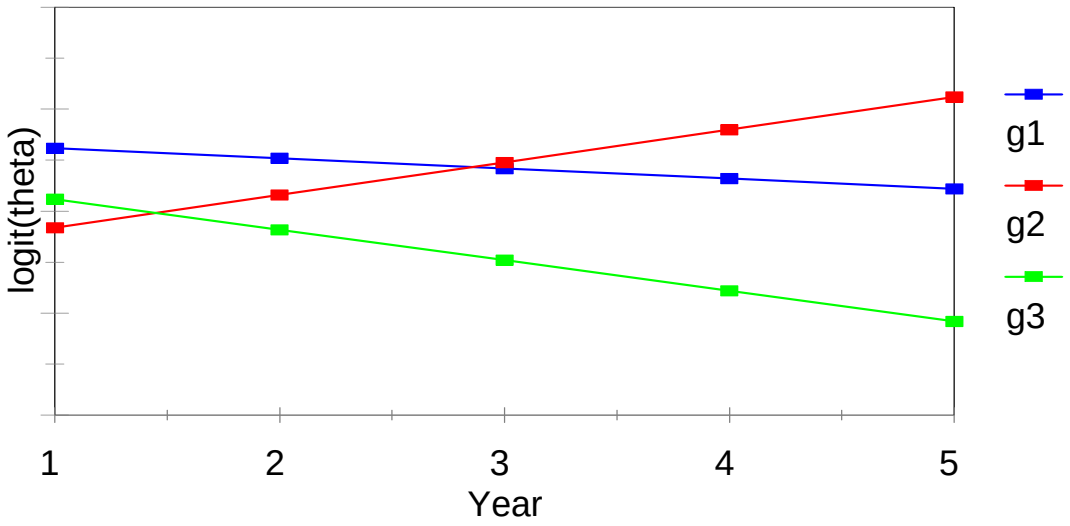
No variation in theta by time or groups

Model {theta(g)}



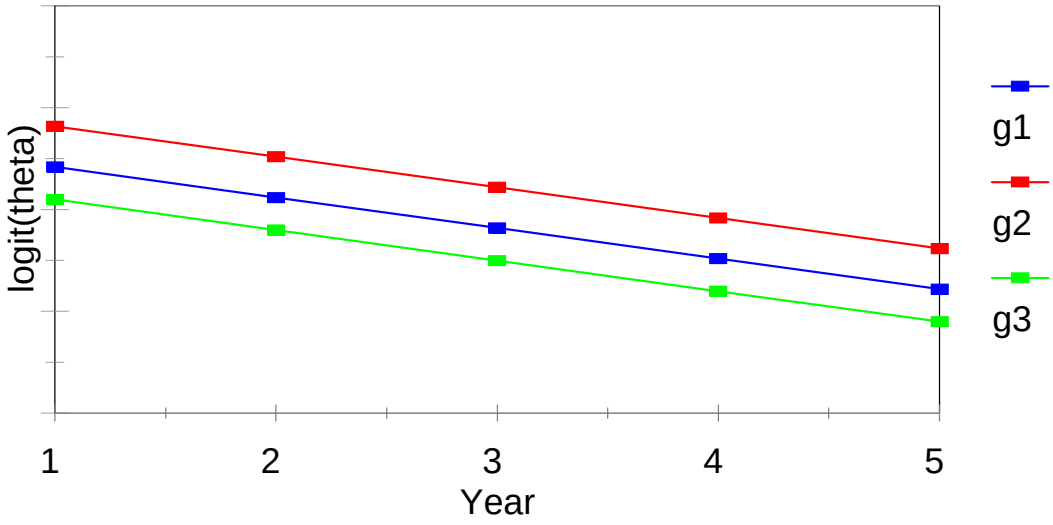
Variation in theta by groups

Model {theta(g*T)}

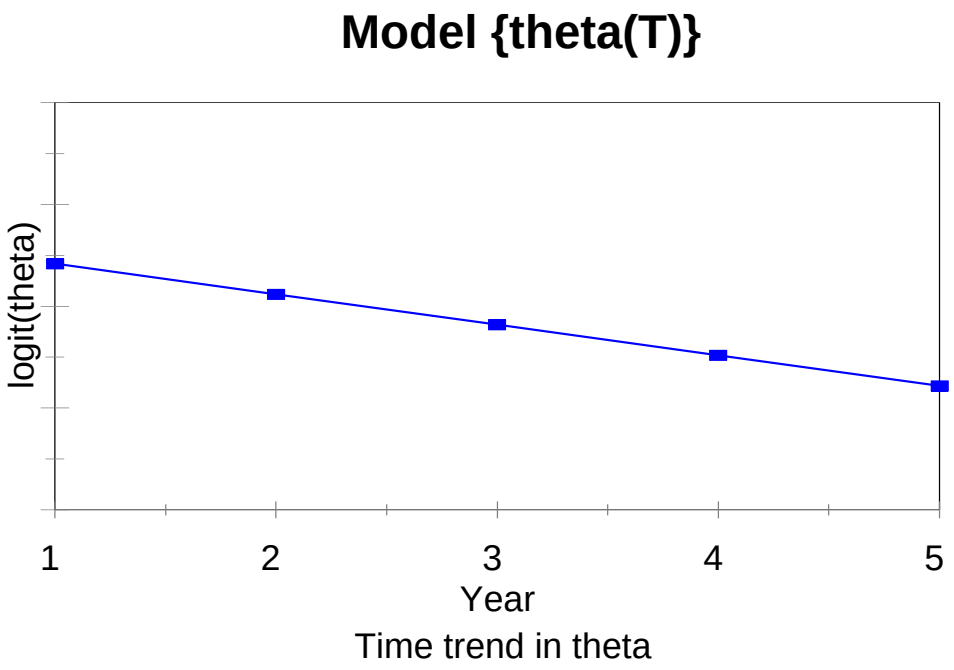


Time trend in theta by groups
Additive model with interactions

Model {theta(g+T)}



Time trend in theta by groups
Additive model with no interactions


Model {theta(T)}
logit(theta)
1
2
3
4
5
Year
Time trend in theta