EXERCISE 13: SINGLE-SPECIES, MULTIPLE-SEASON OCCUPANCY MODELS

Please cite this work as: Donovan, T. M. and J. Hines. 2007. Exercises in occupancy modeling and estimation.
<http://www.uvm.edu/envnr/vtcfwru/spreadsheets/occupancy.htm>

SINGLE-SPECIES, MULTIPLE-SEASON OCCUPANCY MODELS

SPREADSHEET EXERCISE

OBJECTIVES

- To learn and understand the multi-season occupancy model, and how it fits into a multinomial maximum likelihood analysis.
- To use Solver to find the maximum likelihood estimates for the probability of detection, the probability of site occupancy, the probability of site extinction, and the probability of site colonization.
- To assess deviance of the saturated model.
- To reinforce concepts of model fit.
- To learn how to simulate robust occupancy data.

BASIC INFORMATION

The multi-season occupancy model is a natural extension of the single season occupancy model, and is rapidly becoming the analytical framework of choice for many monitoring programs and for estimating metapopulation parameters. This is another very creative model introduced to us by Darryl MacKenzie, Jim Nichols, Jim Hines, Melinda Knutson, and Alan Franklin. The model is described in chapter 7 of the book, Occupancy Modeling and Estimation. Click on the worksheet labeled "Multi-Season and let's get started.

BACKGROUND

This worksheet is set up similarly to the Single-Season Occupancy worksheet, with a few major additions. Think back on the basic occupancy model. The data are collected across study sites, and the sites are repeatedly surveyed.

History	Frequency	Parameter
100	22	p1
111	73	p2
101	25	p3
110	41	ψ
000	55	
011	15	
001	5	
010	14	
# Sites =	250	
# Histories =	8	
Naïve Estimate =	0.78	

The assumption of the single-season model is that the sites are closed to changes of the occupancy state during the sampling interval. That is, if we survey sites three times, the assumption is that site does not go extinct (if it is occupied), and does not get colonized (if it is empty) between the first sampling session and the third sampling session.

However, in many cases we are very interested in <u>changes</u> in occupancy patterns. For instance, metapopulation theory is based on changes in occupancy patterns of patches over time. In this case, we may want to estimate the probability that an empty site will be colonized, and probability that an occupied site goes extinct. The MacKenzie et al. 2003 model (hereafter, the Multi-SeasonOccupancy Model) specifically deals with these issues.

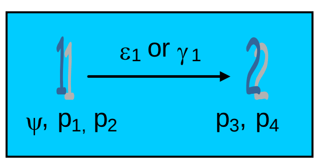
There are two major differences between the multi-season worksheet and the single-season worksheet (Exercise 3). The first is the histories, and the second is the addition of two new parameters, ε (extinction probability) and γ (colonization probability). To keep things simple, this spreadsheet does not evaluate covariates (although you certainly could add them if you wanted to…. the process is conceptually the same as in the single-season occupancy model).

THE SINGLE SPECIES, MULTIPLE-SEASON MODEL

In the multi-season occupancy model, there are primary periods and secondary sampling sessions. Let's assume that we survey a site in May as we've done in the past. The site is visited and a 1 is recorded if a species was detected, otherwise a 0 is recorded. The site is then visited again under the assumption of closure. So, two surveys were conducted in the first primary period. From the data collected in that period, we can estimate the parameters ψ, p_1, and p_2 – just like the standard occupancy model.

Now let's assume that we repeat this process 1 month later (June), under the assumption that the occupancy pattern of the sites may have changed between months. Again, we survey the site two times and record a 1 if an animal was detected and a 0 if it was not. So we have two primary periods (May and June), and <u>within</u> each primary sampling period the site was surveyed twice and the population is assumed to be closed. However, we now assume that the population is open <u>between</u> the primary periods. That is, the species can go locally extinct, and the species can colonize previously unoccupied patches/sites between the first primary period and the second primary period. These processes, together with the initial occupancy estimates for the first sampling period, determine whether the species occurs on a site during the second primary period. Therefore, with the second sampling session, we can derive estimates for p_3 and p_4, as well as extinction and colonization rates.

The diagram below outlines the flow of the study and highlights the parameters that can be estimated in a two season occupancy model. In season 1, we estimate ψ, p_1, p_2. Between seasons, occupied sites can either go extinct with a probability of ε_1, or can remain occupied with a probability of $1-\varepsilon_1$. Between seasons, unoccupied sites can either become colonized with a probability of γ_1, or can remain unoccupied with a probability of $1-\gamma_1$. In season 2, surveys are completed and we estimate p_3 and p_4.



Note that we don't estimate ψ for primary period 2 because occupancy is determined by the state of the site in primary period 1, and whether the site went extinct, was colonized, remained occupied, or remained empty. This model could be extended for 3, 4, 5, or 100 seasons, which is one reason why it is very useful from a long-term monitoring perspective.

SPREADSHEET INPUTS

Let's get oriented to the inputs, which are the encounter histories and their frequencies. The encounter histories are divided into the primary and secondary periods. Capture histories for the first primary period (May) are given in cells C6:C21, and capture histories for the second primary period (June) are given in cells D6:D21. Note that all combinations of the histories are given (there are 16 histories, which is 2^4 - four sampling sessions, each with two possible outcomes: detected or not detected. This means that you can estimate up to 15 parameters for this model). The frequencies of each history are given in cells E6:E21. The total number of sites surveyed is summed in cell E22, and the total number of unique histories is given in cell E23. The naïve occupancy estimate is the number of sites out of the total in which the species was detected.

	C	D	E
5	History		Freq
6	11	11	11
7	10	11	9
8	01	11	7
9	00	11	13
10	11	10	16
11	10	10	10
12	01	10	12
13	00	10	28
14	11	01	16
15	10	01	7
16	01	01	10
17	00	01	20
18	11	00	60
19	10	00	63
20	01	00	44
21	00	00	174
22	Total sites =		500
23	Total histories =		16
24	Naïve estimate =		0.652

THE SATURATED MODEL

You might recall from the general (single-season) occupancy model that we have all the information we need (the frequencies of each history) to compute the saturated model's $-2\text{Log}_e\text{L}$. Do you remember how we did it? The probability of getting a particular history is simply the frequency of that history divided by the total number of sites. The results give the

MLE's for the raw data – they are simply the proportion of sites showing a particular history, and are computed in cells M6:M21. The sum of these cells is 1.0. The natural log of these probabilities is computed in cells N6:N21. To get the saturated model's Log_eL, we simply multiply the frequency of each history by the natural log of its probability, and then add the results up across histories. This result is multiplied by -2 to give the saturated model's $-2Log_eL$ (cell N22), which you might remember is the basis for computing deviance. The formula in cell N22 is =-2*SUMPRODUCT(E6:E21,N6:N21). You can try plugging in different probabilities to your heart's content, but you won't find a lower $-2Log_eL$ than the one given in cell N22 because the probabilities fit the observed field data perfectly.

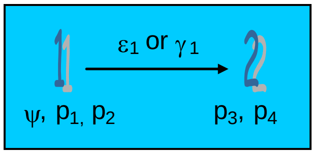
	M	N
4	Saturated Model	
5	Probability	Ln (Prob)
6	0.022	-3.816713
7	0.018	-4.017384
8	0.014	-4.268698
9	0.026	-3.649659
10	0.032	-3.442019
11	0.02	-3.912023
12	0.024	-3.729701
13	0.056	-2.882404
14	0.032	-3.442019
15	0.014	-4.268698
16	0.02	-3.912023
17	0.04	-3.218876
18	0.12	-2.120264
19	0.126	-2.071473
20	0.088	-2.430418
21	0.348	-1.055553
22	$-2Log_eL$(sat) =	2223.794

THE MULTI-SEASON OCCUPANCY PARAMETERS

Now let's focus on estimating the probability of each history, not from the raw data, but from the multi-season occupancy model parameters. The parameters for this model are listed in cells F6:F12, and there are 7 of them for this exercise: p_1 and p_2 estimate the probability of detection for surveys conducted in the first primary period, and p_3 and p_4 estimate the probability of detection for surveys conducted in the second primary period.

	F	G	H	I
5	Parameter	Estimate?	Betas	MLE
6	p1	1	-3	0.42944
7	p2	1	-2	0.04535
8	p3	1	-1	0.07926
9	p4	1	0	0.50000
10	ψ	1	1	0.92074
11	ε1	1	2	0.95465
12	γ1	1	3	0.57056

What about psi, ε_1 (epsilon) and γ_1 (gamma)? To get a handle on what these estimates are, look at the blue figure on the spreadsheet.



The large number 1 in the diagram indicates the first primary period. You should see that the parameters estimated for this session are psi, p_1, and p_2.....exactly like we did in the single-season model. Now notice the arrow pointing to the number 2, which indicates the second primary period. If a site is occupied in the first primary period, it can either go extinct with a probability of ε or can remain occupied with a probability of $(1-\varepsilon)$. If a site is not occupied in the first primary period, it can become recolonized with a probability of γ, or it can remain unoccupied with a probability of $(1-\gamma)$. Whether the species is detected or not in the second primary period depends on the site's occupancy state and on p_3 and p_4. Note that we don't estimate ψ for primary period 2 because occupancy is determined by the state of the site in primary period 1, and whether the site went extinct, was colonized, remained occupied, or remained empty.

As we've done in previous worksheets, we enter a 1 in cells G6:G12 to indicate that all of the parameters will be uniquely estimated, and enter some beta estimates in cells H6:H12.

	F	G	H	I
5	Parameter	Estimate?	Betas	MLE
6	p1	1	-3	0.04743
7	p2	1	-2	0.11920
8	p3	1	-1	0.26894
9	p4	1	0	0.50000
10	ψ	1	1	0.73106
11	ε1	1	2	0.88080
12	γ1	1	3	0.95257

Ultimately, our goal will be to find the combination of betas that maximizes the multinomial likelihood function. The model probabilities are linked to the betas once again with a logit link in cells I6:I12. These cells are named pe1, pe2, pe3, pe4, psi1, eps1, and gam1. Click on one of the cells (I6:I12) and you should see the name of the cell appear to the left of the formula bar.

PROBABILITY OF EACH HISTORY

Now we're ready to write out the probability of obtaining a particular history (cells J6:J21). Let's start with the history 11 11 (cells C6:D6), which indicates that an animal was detected in both the primary and second period, and was detected in each survey. The probability of getting this history (cell J6) is =psi1*pe1*pe2*(1-eps1)*pe3*pe4. The site must have been occupied in the first period (ψ), WAS detected in the first sampling session (p_1), and was detected again in the second sampling session (p_2). It then failed to go extinct between the first and second periods (1-ε), and was detected in the third sampling session (p_3) and again in the fourth sampling session (p_4).

How about the history 10 11 (cells C7:D7)? The probability of getting this history (cell J7) is =psi1*pe1*(1-pe2)*(1-eps1)*pe3*pe4. The site must have been occupied in the first period (ψ), was detected in the first sampling session (p_1), but was missed in the second sampling session ($1-p_2$). The site then failed to go extinct ($1-\varepsilon$), and was then detected in sampling sessions 3 (p_3) and 4 (p_4).

Let's try a trickier one, history 00 01. See if you can write out the probability equation in the space below:

The equation in cell J17 is =psi1*(1-pe1)*(1-pe2)*(1-eps1)*(1-pe3)*pe4 + (1-psi1)*gam1*(1-pe3)*pe4. The site could have been occupied in the first primary period (ψ), missed on survey 1 ($1-p_1$), missed on survey 2 ($1-p_2$), failed to go extinct ($1-\varepsilon$), was missed on survey 3 ($1-p_3$), and was detected on survey 4 (p_4), OR it could have been unoccupied in primary session 1 ($1-\psi$), was colonized in period 2 (γ), was missed on survey 3 ($1-p_3$), and detected on survey 4 (p_4).

Now try history 00 00! (Check your answer with the formula in cell J21.)

It's pretty straight-forward but we encourage you to work your way through the remaining histories (cells J6:J21). After you've convinced yourself that the equations are correct, notice that the sum of the histories, calculated in cell J22) is 1.

THE MULTINOMIAL LOG LIKELIHOOD

You probably can guess where we're headed….

We need to write out the multinomial log likelihood. The equation is given in cell H14:

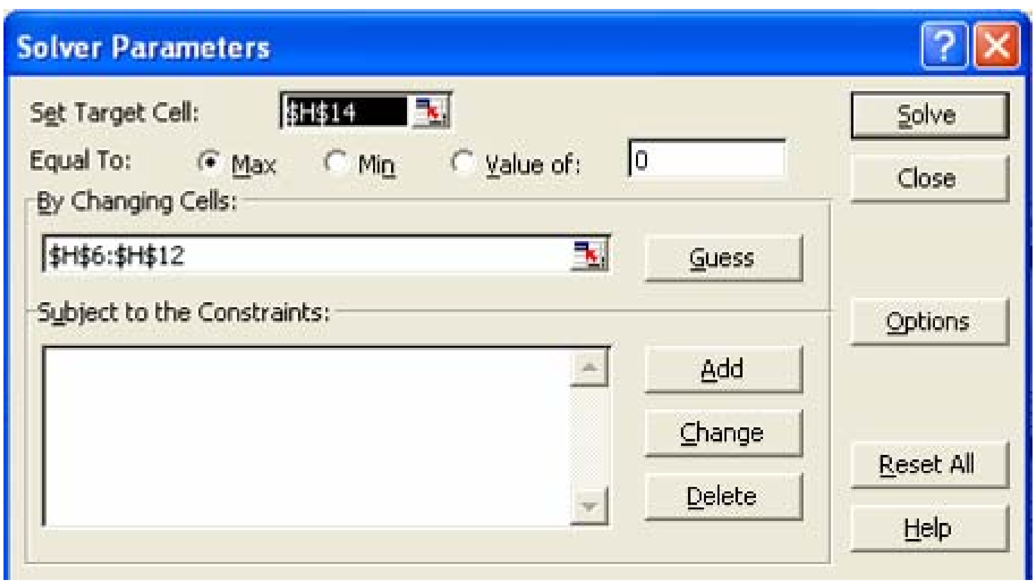
=E6*LN(J6)+E7*LN(J7)+E8*LN(J8)+E9*LN(J9)+E10*LN(J10)+E11*LN(J11)+E12*LN(J12)+E13*LN(J13)+E14*LN(J14)+E15*LN(J15)+E16*LN(J16)+E17*LN(J17)+E18*LN(J18)+E19*LN(J19)+E20*LN(J20)+E21*LN(J21). Click on this cell, then click somewhere in the formula bar and you should see the cells used in the equation "light up" - this often helps to see the pattern more clearly. This formula is exactly what we used in previous worksheets for the Log Likelihood calculations. This formula could have been substantially shorted by using the SUMPRODUCT function, but then you can't see the symmetry of the lit up equation.

MAXIMIMIZING THE LOG LIKELIHOOD

The goal now is to maximize the multinomial log likelihood by changing the beta values in cells H6:H12 (and hence parameter MLE's).

	F	G	H	I
5	Parameter	Estimate?	Betas	MLE
6	p1	1		0.50000
7	p2	1		0.50000
8	p3	1		0.50000
9	p4	1		0.50000
10	ψ	1		0.50000
11	ε1	1		0.50000
12	γ1	1		0.50000

Go to Tools | Solver, and set target cell H14 to a maximum by changing cells H6:H12.



Press Solve and Solver will attempt to converge on an answer. Once Solver has found the betas, the MLE's that are linked to the betas are the maximum likelihood parameter estimates for p1, p2, p3, p4, ψ, ε, and γ. Here are the results we got:

	F	G	H	I
5	Parameter	Estimate?	Betas	MLE
6	p1	1	0.344269485	0.58523
7	p2	1	0.146092624	0.53646
8	p3	1	-0.281413068	0.43011
9	p4	1	-0.500775969	0.37736
10	ψ	1	0.646207271	0.65616
11	ε1	1	-0.294950958	0.42679
12	γ1	1	-0.664854695	0.33965

Remember, when you run the model in MARK or PRESENCE, the programs report the estimate as well as the standard errors, and both are important. This spreadsheet does not include the standard errors for the various estimates. Let's look at these estimates more closely. In season 1, 65.6% of the sites were occupied. In this season, detection probability was 0.58 and 0.53. Between season 1 and season 2, 42.6% of the occupied sites went extinct, and 33.9% of the unoccupied sites were colonized. This means that occupied sites in season 1 had a 1 - 42.6% = 57.4% chance of remaining occupied, and that unoccupied sites in season 1 had a 1-33.9% = 66.1% chance of staying empty in season 2. That's quite a turnover. These estimates are the foundation of metapopulation theory, so they can be very useful. From these estimates, you can derive ψ for season 2 as:

$$\psi_{t+1} = \psi_1 (1-\varepsilon_1) + (1-\psi_1)\gamma_1,$$

which is simply the fraction of sites occupied in season 1 that failed to go extinct, $\psi_1 (1-\varepsilon_1)$, plus the fraction of unoccupied sites in season 1 that were colonized, $(1-\psi_1)\gamma_1$.

In our example, ψ_2 = 0.656*(1-0.426) + (1-0.656)*0.339 = 0.49316. So, with the estimates from this model, the occupancy rate declined quite substantially from season 1 to season 2. We will revisit this topic when in the PRESENCE exercise.

MODEL OUTPUT

Now let's look at the remaining model output:

	G	H
13	**OUTPUTS**	
14	Log_eL =	-1114.85
15	$-2Log_eL$ =	2229.70
16	Deviance =	5.906
17	Model DF =	9
18	C hat =	0.66
19	Effective N	500
20	K =	7
21	AIC =	2243.70
22	AICc =	2243.93

The model's Log_eL is computed in cell H14 (this is the cell you maximized), and the model's $-2Log_eL$ is computed in cell H15. Deviance for this model is computed as the difference between this model's $-2Log_eL$ and the saturated model's $-2Log_eL$ with the equation =H15-N22. The model's degree of freedom is reported in cell H17 with the equation =E23-H20. It is simply the number of kinds of histories minus K (where K is the number of parameters estimated in cell H19). This computation goes back to the basic multinomial equation: we have 16 terms in the multinomial equation, and thus we estimate 16 probabilities in a saturated model. The model we just ran estimated those 16 probabilities indirectly by estimating 7 parameters: p_1, p_2, p_3, p_4, ψ, ε, and γ. That leaves us with 16-7 = 9 parameters that are left over. AIC, AICc, and N are computed with the same equations as the general occupancy model.

ASSESSING FIT

	K	L
4	Chi Square	
5	Expected	$(O-E)^2/E$
6	9.583	0.2096753
7	8.280	0.0625994
8	6.792	0.0064007
9	15.346	0.3586158
10	15.811	0.0022549
11	13.662	0.9816133
12	11.206	0.0562611
13	25.321	0.2834968
14	12.697	0.8593283
15	10.971	1.4373615
16	8.999	0.111406
17	20.333	0.0054644
18	64.909	0.3713234
19	56.087	0.8521186
20	46.004	0.0872777
21	174.000	4.114E-12
22	500.0	5.6852
23	p =	0.7709646

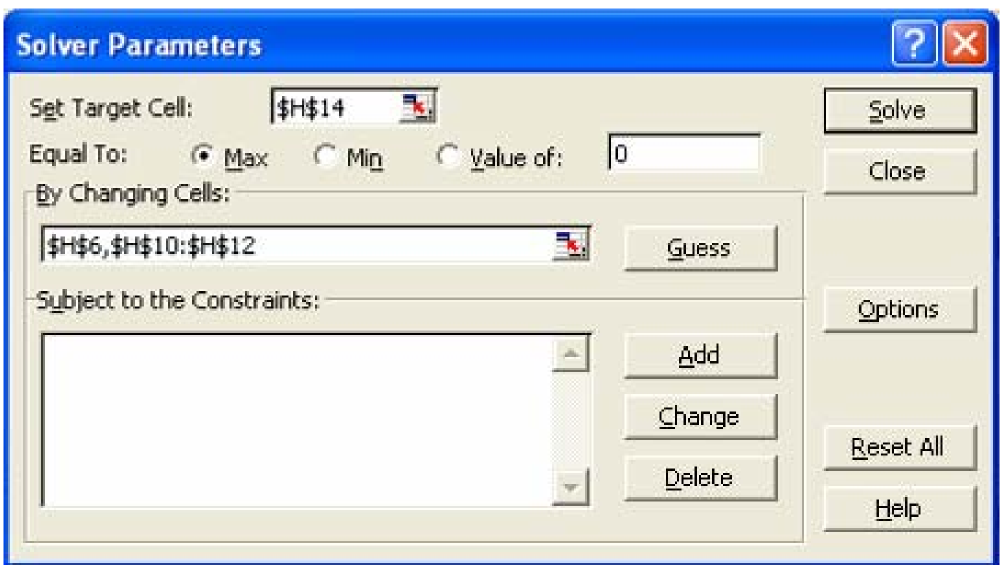
How does this model fit? Well, we've entered formula in cells K4:L22 to compute the Pearson Chi-Square. However, note that this approach has not been formally tested for the multi-season occupancy model, so it's still unclear whether this is a good measure of fit. Given this model's MLE's, the number of sites expected to have each encounter history is computed in cells K6:K21. This is exactly the same procedure we used to compute Chi-Square in the general (single season) occupancy model: the probability of each history is multiplied by the total number of sites. The $(O-E)^2/E$ is computed in cells L6:L21. Remember, here is where you look for cell values greater than 3.8 – those with high values don't fit. The sum of $(O-E)^2/E$ is computed in cell L22 and is the model's Chi-Square value. The p value associated with the Chi-Square value is computed in cell L23 with a CHIDIST function. This particular model appears to fit the data, as indicated by a p value much greater than 0.05. You could also run the MacKenzie and Bailey GOF but you'd have to modify your spreadsheet.

COMPARING MODELS

As we've done in the past, you can run different models and compare model results with model selection procedures. For instance, you might run a model where $p_1 = p_2 = p_3 = p_4$ if you believed that the detection probabilities could have been the same in every sampling effort. To set up this model, you'd enter a 0 in cells G7:G9 to indicate that p_2, p_3, and p_4 won't be estimated – instead they will be constrained to equal p_1. And you'd enter =H6 in cells H7:H9 to force their betas to be equal to the beta for p_1.

	F	G	H
5	Parameter	Estimate?	Betas
6	p1	1	
7	p2	0	=H6
8	p3	0	=H6
9	p4	0	=H6
10	ψ	1	
11	ε1	1	
12	γ1	1	

Then, run Solver. Set cell H14 to a maximum by changing cells H6, H10:H12.



Here are the results we got:

	F	G	H	I
5	Parameter	Estimate?	Betas	MLE
6	p1	1	0.017637275	0.50441
7	p2	0	0.017637275	0.50441
8	p3	0	0.017637275	0.50441
9	p4	0	0.017637275	0.50441
10	ψ	1	0.859492002	0.70255
11	ε1	1	0.039154811	0.50979
12	γ1	1	-1.049538461	0.25931
13		**OUTPUTS**		
14		Log_eL =	-1120.63	
15		$-2Log_eL$ =	2241.26	
16		Deviance =	17.466	
17		Model DF =	12	
18		C hat =	1.46	
19		Effective N	500	
20		K =	4	
21		AIC =	2249.26	
22		AICc =	2249.34	

How does this model compare to the first model? The first model had an AICc score of 2243.93, while this one had an AICc score of 2249.34, a difference of around 5 AICc units; the first model has more support. We'll run these models in MARK and PRESENCE too to verify the results. Remember that you can easily incorporate covariates into the multi-season model by specifying linear constraints. We haven't added them to this exercise, but the principles outlines in Chapters 4 and 5 of this e-book apply to this model as well.

SIMULATING DATA BY EXPECTATION AND WITH STOCHASTICITY

OK, we're getting close to the end. The last thing we need to cover in the spreadsheet is how to simulate multi-season occupancy data. All of this should be familiar to you by now.

	Q	R	S	T	U	V	W	X
3	**Parameter Inputs**							
4	p1 =	p2 =	p3 =	p4 =	ψ =	ε1 =	γ1 =	N =
5	0.5	0.5	0.5	0.5	0.7	0.5	0.3	500
6								
7	Individual	rand ψ	rand p1	rand p2	rand ε	rand γ	rand p3	rand p4
8	1	0.47219	0.78183	0.50905	0.80125	0.47699	0.63875	0.86771272
9	2	0.15148	0.14435	0.51854	0.86443	0.10579	0.30224	0.510548333
10	3	0.30298	0.61971	0.86676	0.36856	0.22125	0.99151	0.862516479
11	4	0.76775	0.14252	0.41342	0.40829	0.49829	0.95904	0.395187085
12	5	0.74089	0.61386	0.47887	0.77415	0.91296	0.44294	0.16313406

First, start by entering real parameter estimates in cells Q5:W5, and also enter N (the number of sites you want to simulate). As with the other spreadsheets, we'll simulate data in two ways.

	Y	Z	AA	AB	AC
6	T1		T2		
7	p1	p2	p3	p4	History
8	0	0	0	0	00 00
9	0	0	0	0	00 00
10	1	1	0	0	11 00
11	0	0	0	0	00 00
12	0	0	0	0	00 00

The first way involves a random (stochastic) component to assigning encounter histories for a site. The sites are listed out in column Q. Then, for each site, we assign a random ψ, p_1, p_2, ε, γ, p_3, and p_4 as shown in columns R:X. The histories for each site will be based on these random numbers, and also the parameter estimates you specified previously. The outcome (0 or 1) for visit 1 is provided in column Y, the outcome for visit 2 is provided in column Z, the outcome for visit 3 is provided in column AA, and the outcome for visit 4 is provided in column AB. The site's history is simply the concatenation of these four outcomes. So, let's look at how the outcomes are determined. Cell Y8 has the formula =IF(AND(R8<$U$5,S8<$Q$5),1,0) and gives the outcome for visit 1 for site 1. IF cell R8 (the random ψ) is less than cell U5 (the specified ψ), AND cell S8 (the random p_1) is less than cell Q5 (the specified p_1), then the site was occupied and the species was detected, and a 1 is returned. If either of those conditions is not true, a 0 is returned. The same logic goes for determining the outcome for visit 2 (cell Z8 has the equation =IF(AND(R8<$U$5,T8<$R$5),1,0).

Things get more complicated for visits 3 and 4 because the outcome depends on the outcomes from the first primary period. Cell AA8 has the equation =IF(OR(AND(R8<$U$5,U8>$V$5,W8<$S$5),AND(R8>$U$5,V8<$W$5,W8<$S$5)),1,0), and generates an outcome for visit 3 for site 1. Cell AB8 has the equation =IF(OR(AND(R8<$U$5,U8>$V$5,X8<$T$5),AND(R8>$U$5,V8<$W$5,X8<$T$5)),1,0) and generates an outcome for visit 4 in site 1. See if you can work your way through the logic behind this formula.

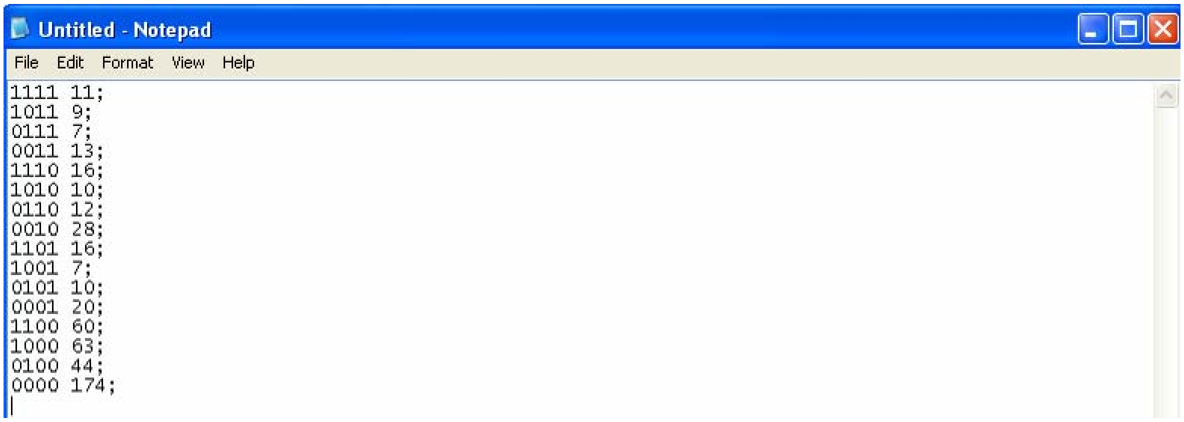
The second method of generating data is based on expectation. There's nothing fancy about this at all: just enter the same old encounter history probabilities that we did earlier for each type of encounter history, and multiply by N. To make things easier, we copied the formula for each history probability and pasted it in red type in column AG. These are the equations entered into cells AF8:AF23. Note the sum of the expected frequencies must equal N (whatever you specified in cell X5).

	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO	AP	AQ
6	Expected											
7	Frequency											
8	10.9375	psi1*pe1*pe2*(1-eps1)*pe3*pe4										
9	10.9375	psi1*pe1*(1-pe2)*(1-eps1)*pe3*pe4										
10	10.9375	psi1*(1-pe1)*pe2*(1-eps1)*pe3*pe4										
11	22.1875	psi1*(1-pe1)*(1-pe2)*(1-eps1)*pe3*pe4+(1-psi1)*gam1*pe3*pe4										
12	10.9375	psi1*pe1*pe2*(1-eps1)*pe3*(1-pe4)										
13	10.9375	psi1*pe1*(1-pe2)*(1-eps1)*pe3*(1-pe4)										
14	10.9375	psi1*(1-pe1)*pe2*(1-eps1)*pe3*(1-pe4)										
15	22.1875	psi1*(1-pe1)*(1-pe2)*(1-eps1)*pe3*(1-pe4)+(1-psi1)*gam1*pe3*(1-pe4)										
16	10.9375	psi1*pe1*pe2*(1-eps1)*(1-pe3)*pe4										
17	10.9375	psi1*pe1*(1-pe2)*(1-eps1)*(1-pe3)*pe4										
18	10.9375	psi1*(1-pe1)*pe2*(1-eps1)*(1-pe3)*pe4										
19	22.1875	psi1*(1-pe1)*(1-pe2)*(1-eps1)*(1-pe3)*pe4+(1-psi1)*gam1*(1-pe3)*pe4										
20	54.6875	psi1*pe1*pe2*eps1+psi1*pe1*pe2*(1-eps1)*(1-pe3)*(1-pe4)										
21	54.6875	psi1*pe1*(1-pe2)*eps1+psi1*pe1*(1-pe2)*(1-eps1)*(1-pe3)*(1-pe4)										
22	54.6875	psi1*(1-pe1)*pe2*eps1+psi1*(1-pe1)*pe2*(1-eps1)*(1-pe3)*(1-pe4)										
23	170.9375	psi1*(1-pe1)*(1-pe2)*(1-eps1)*(1-pe3)*(1-pe4)+psi1*(1-pe1)*(1-pe2)*eps1+(1-psi1)*(1-gam1)+(1-psi1)*gam1*(1-pe3)*(1-pe4)										
24	500.0000											

Why would you generate data based on expectation? Well, it happens to be a GREAT way to conduct a power analysis when you are in the design phase of your study. If you have some idea of what the parameter estimates might be, you can simulate data with varying sample sizes, run the analysis through MARK or PRESENCE, and then critically evaluate how sample size affects the estimates in terms of bias (how far off where the estimates from MARK or PRESENCE compared to values entered in cells Q5:W5) and precision (by examining the standard errors around each estimate). It's relatively easy to do, and is an important step in planning a study.

CREATING MARK AND PRESENCE INPUT FILES

The final step is to compare these results with those estimated in MARK or PRESENCE. To create a MARK input file, copy cells O6:O21, and paste them into WordPad.




```
1111 11;
1011 9;
0111 7;
0011 13;
1110 16;
1010 10;
0110 12;
0010 28;
1101 16;
1001 7;
0101 10;
0001 20;
1100 60;
1000 63;
0100 44;
0000 174;
```

Then save the file as "Multi-Season Occupancy.Inp". This is the file you'll import into MARK.

To create PRESENCE input files, we used the "tally" trick in column A to create histories for individual sites. In cells B6:B21, we list the 16 possible histories. Then, in cells A5:A21, we use the SUM function to track a running tally of the history frequencies. For instance, the 1111 history is our first history, and the running tally is 11. The second history listed is the 1011 history, and the running tally sums the frequency of that history, plus the frequency of the 1111 history, and so on.

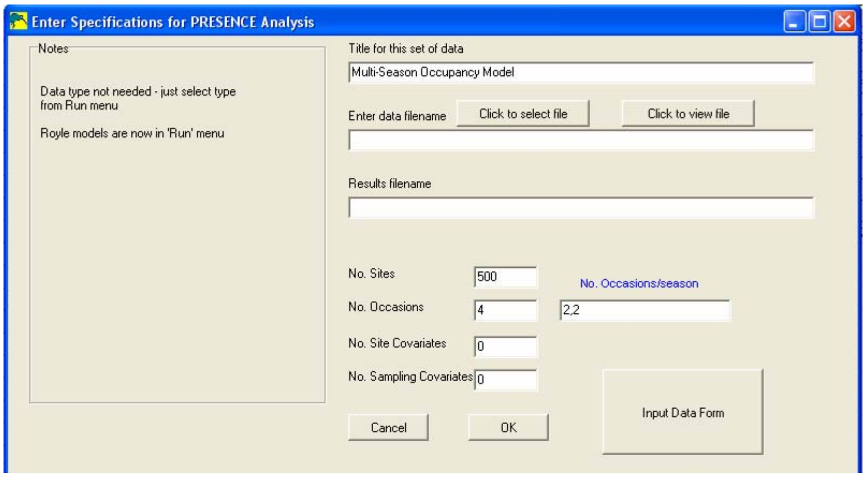
	A	B
4	Tally	
5	0	
6	11	1111
7	20	1011
8	27	0111
9	40	0011
10	56	1110
11	66	1010
12	78	0110
13	106	0010
14	122	1101
15	129	1001
16	139	0101
17	159	0001
18	219	1100
19	282	1000
20	326	0100
21	500	0000

The next step is to use this tally to assign encounter histories to each of the 500 sites. First, the site numbers are listed in column A from 1 to 500. Then, the history to each site is assigned by using the LOOKUP function. Click on cell B30 and you'll see the formula =LOOKUP(A30-1,$A$5:$A$21,$B$6:$B$21). This formula looks up the site number (minus 1), and finds that number in the tally (cells A5:A21). Because the tally is sorted in ascending order, the lookup function does not need a perfect match. It simply finds the category that does not exceed the lookup number, and returns the associated history listed in cells B6:B21. Columns C:F simply "break" the history apart into survey-specific results. You'll copy cells C30:F529 when it's time to run the analysis in PRESENCE.

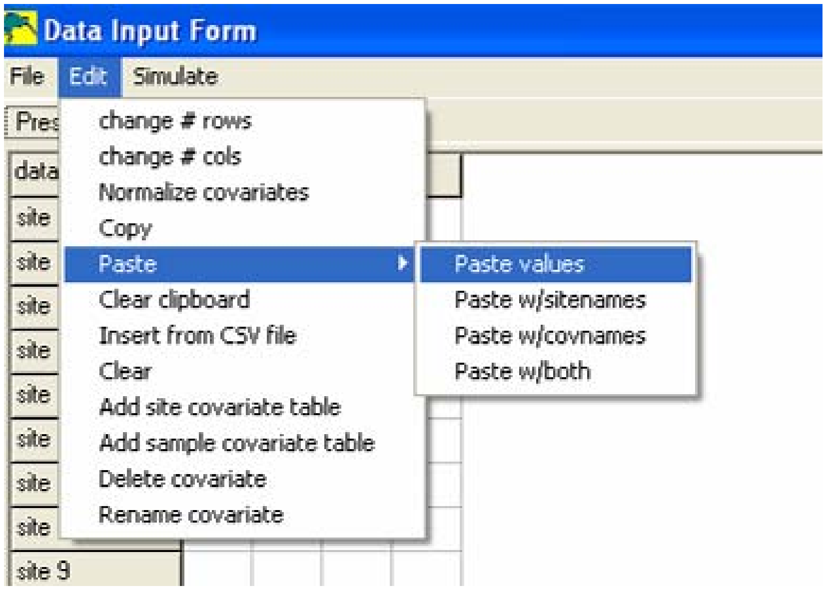
MULTIPLE-SEASON OCCUPANCY ANALYSIS IN PROGRAM PRESENCE

GETTING STARTED

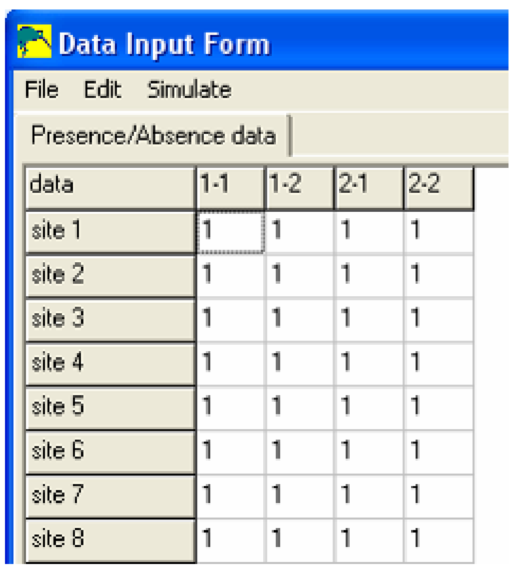
Open PRESENCE, and click File | New Project to begin a brand new data analysis.



Enter a title for this set of data (e.g., Multi-Season Occupancy Model). Enter 500 for the No. Sites, and 4 for the No. Occasions. Importantly, enter "2,2" in the No. Occasions/Season text box. This notation lets PRESENCE know that it is a multi-season model, where the first number indicates the number of primary seasons, followed by a comma, and the second number indicates the number of secondary seasons within a primary season. Then click on the Input Data Form button. Copy cells C30:F529, and then click on the first, blank data cell and go to Edit | Paste | Paste Values:

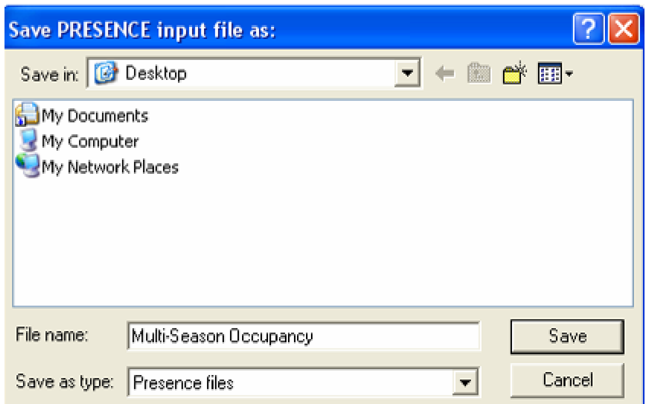


Your data should look like this:

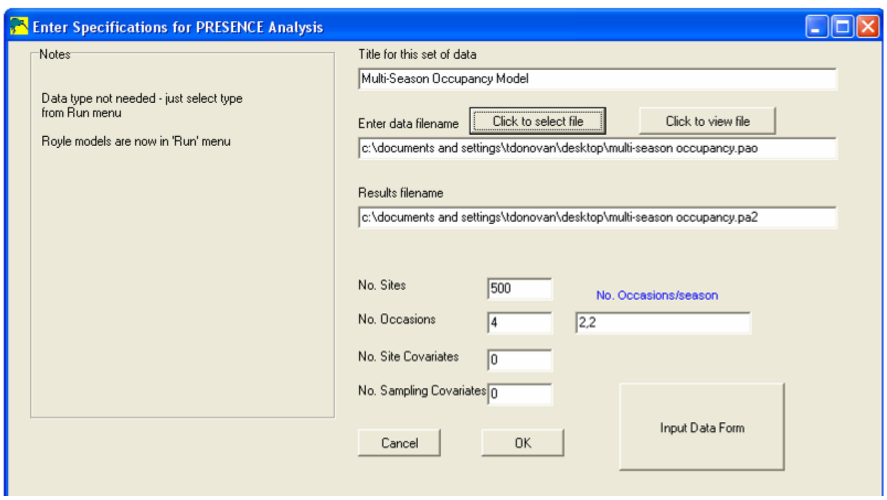


Note that PRESENCE automatically labels each survey by its primary and secondary period. For instance, 1-1 indicates the survey occurred in primary season 1 and was the first survey, while 1-2 indicates the survey occurred in primary season 2 and was the second survey.

Then go to File | Save As and enter a file name for your new PRESENCE input file, and store it somewhere where you can retrieve it easily:



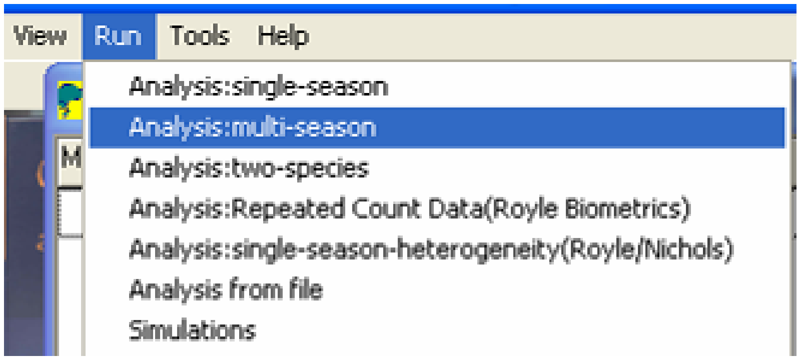
Now, return to the Enter Specifications form, click the button labeled "Click to Select File" and browse to your freshly created input file:



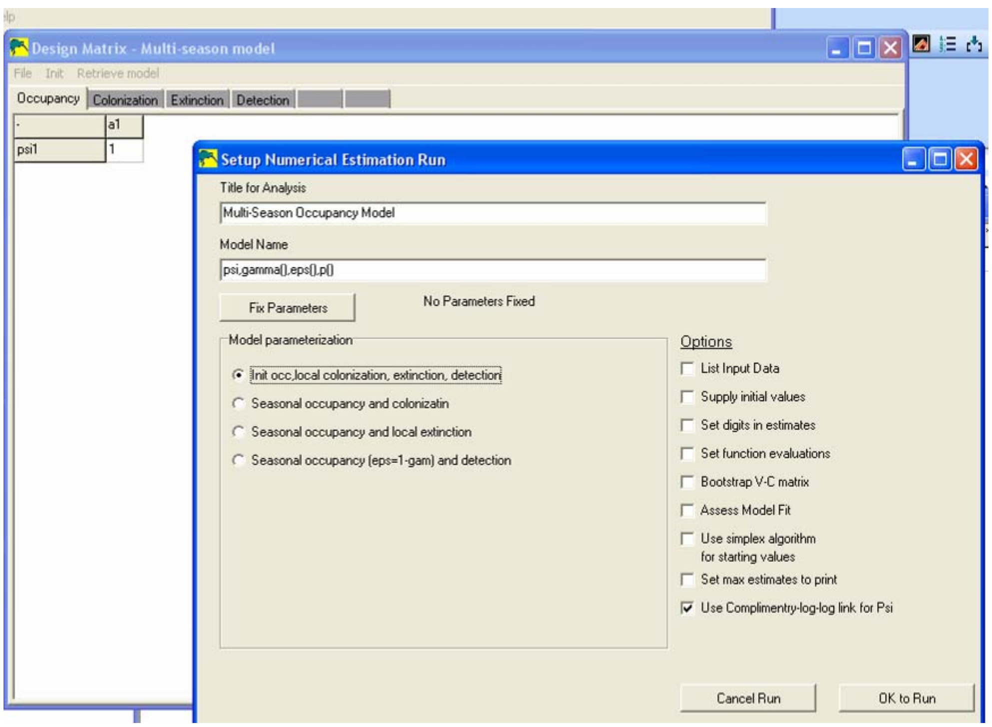
Press OK and you're ready to run your first model. As with the spreadsheet exercise, we will run only two models: Model $\psi(.)\varepsilon(.)\gamma(.)p(t)$ and model $\psi(.)\varepsilon(.)\gamma(.)p(.)$. Then, we'll talk about the different ways you can run these same models in PRESENCE with a different parameterization scheme.

MODEL $\psi(.)\varepsilon(.)\gamma(.)p(t)$

To run a multi-season analysis, go to Run | Analysis: multi-season.



You'll then see the following form (the Setup Numerical Estimation Run form):

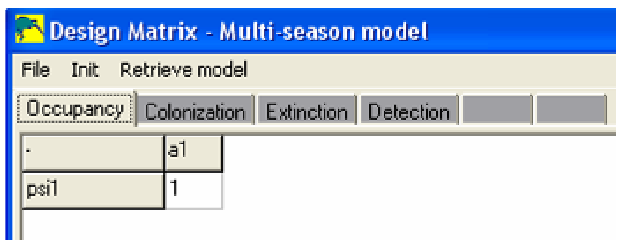


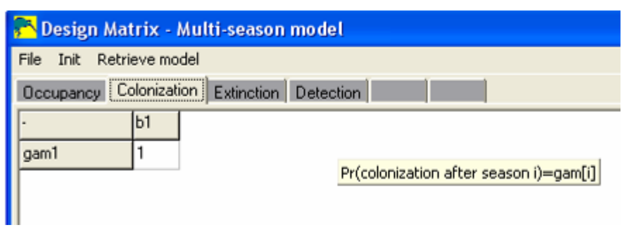
For multi-season models, PRESENCE allows you to analyze the same data with different model parameterizations. Look at the option box labeled Model parameterization, and you'll see four options:

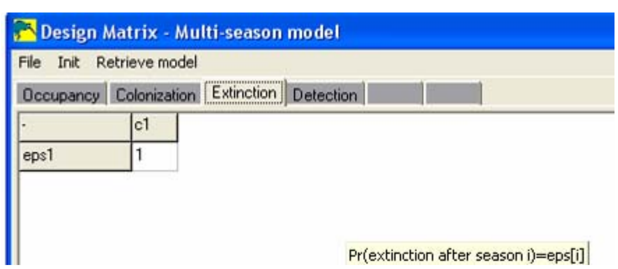
1. Init occ, local colonization, extinction, detection

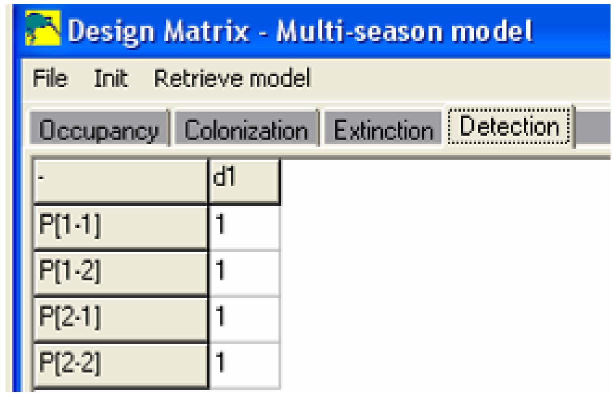
2. Seasonal occupancy and colonization
3. Seasonal occupancy and local extinction
4. Season occupancy (eps = 1-gam) and detection

Notice that the Design Matrix (DM) in PRESENCE opens automatically, and that the DM changes when you select different options. By default, option 1 is selected, so we'll start with this option. Make sure this option is selected, and then take a look at the DM for option 1 by clicking on the four different tabs (Occupancy, Colonization, Extinction, and Detection):



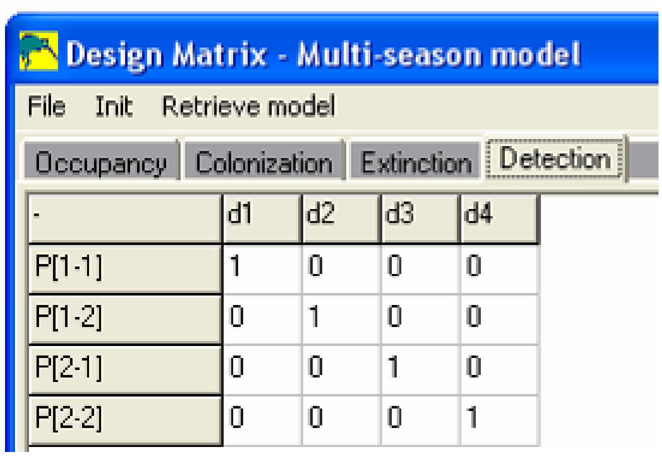






-	d1
P[1-1]	1
P[1-2]	1
P[2-1]	1
P[2-2]	1

With option 1, PRESENCE will estimate the following parameters: ψ_1 (psi1 with the beta called a1), γ_1 (gam1 with the beta called b1), ε_1 (eps1 with the beta called c1), and p., where $p_{1\text{-}1}$, $p_{1\text{-}2}$, $p_{2\text{-}1}$, and $p_{2\text{-}2}$ are equal with the beta called d1. This option is parameterized in the same way that our spreadsheet was parameterized, except it defaults to p(.) model. We want our first model to be p(t), so click somewhere in Detection tab and then go to Init | Full Identity, and you'll see that PRESENCE now assigns a unique beta for each of the survey periods.

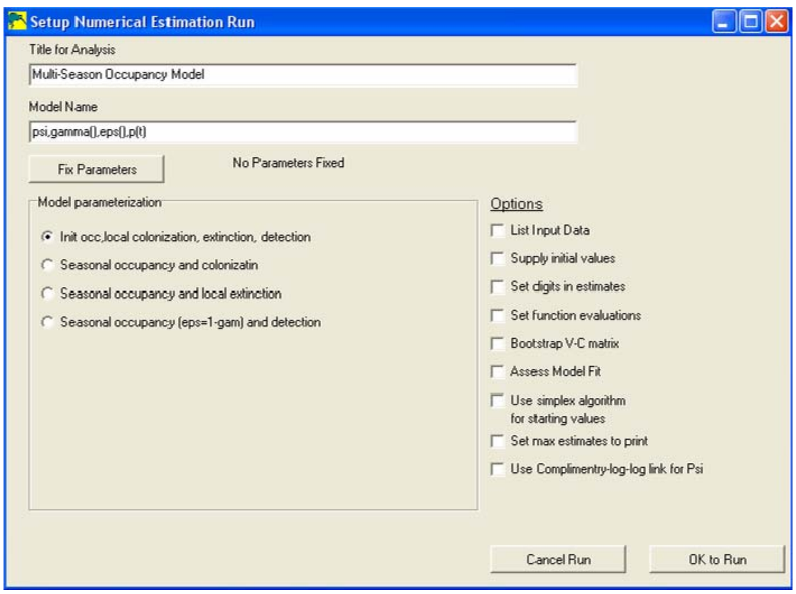


-	d1	d2	d3	d4
P[1-1]	1	0	0	0
P[1-2]	0	1	0	0
P[2-1]	0	0	1	0
P[2-2]	0	0	0	1

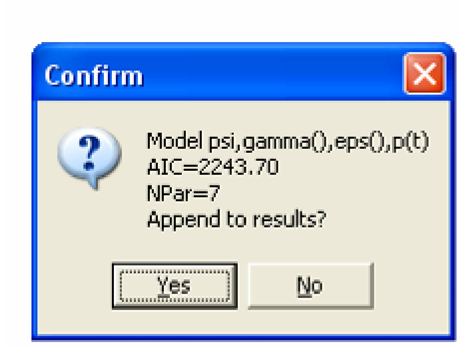
In this case, PRESENCE will call the p_1 beta "d1", the p_2 beta "d2", the p_3 beta (or the first survey in primary period 2) "d3", and the p_4 beta (or the second survey in primary period 2) "d4". Make sense? In the spreadsheet, we set this model up as:

	F	G	H	I
5	Parameter	Estimate?	Betas	MLE
6	p1	1		0.50000
7	p2	1		0.50000
8	p3	1		0.50000
9	p4	1		0.50000
10	ψ	1		0.50000
11	ε1	1		0.50000
12	γ1	1		0.50000

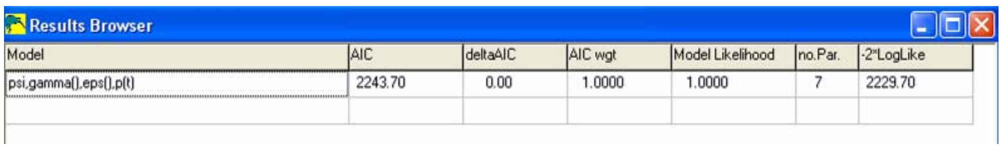
When you are satisfied that the DM is set up correctly, return to the Set-up Numerical Estimation Run. You'll want to change the title for this model to psi, gamma(),esp(),p(t). Then press the button labeled "OK to Run."



PRESENCE then finds the solution, and presents the following dialogue box:

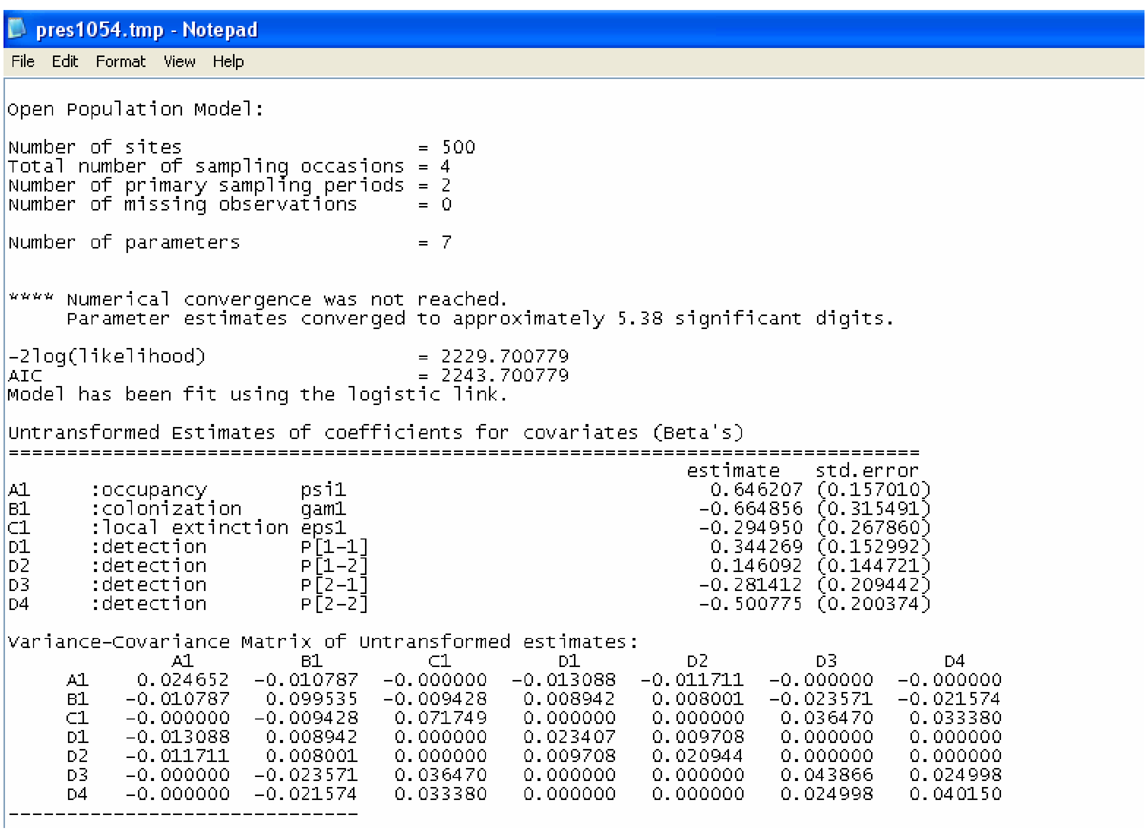


Click Yes, and the results are appended to the Results Browser:


Results Browser

Model	AIC	deltaAIC	AIC wgt	Model Likelihood	no.Par.	-2*LogLike
psi,gamma(),eps(),p(t)	2243.70	0.00	1.0000	1.0000	7	2229.70

Let's look at the full output for this model. Right-click on the model name in the Results Browser, and select View Model Output. Then scroll down to get to the meat of the output:


pres1054.tmp - Notepad

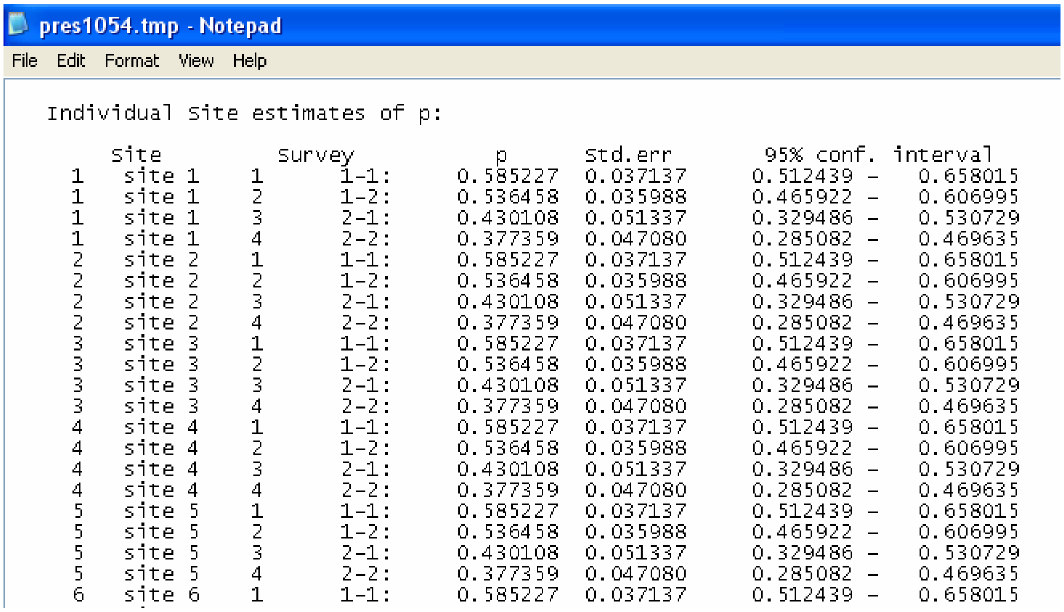
File Edit Format View Help

```
Open Population Model:

Number of sites                       = 500
Total number of sampling occasions    = 4
Number of primary sampling periods    = 2
Number of missing observations        = 0

Number of parameters                  = 7


**** Numerical convergence was not reached.
     Parameter estimates converged to approximately 5.38 significant digits.

-2log(likelihood)                     = 2229.700779
AIC                                   = 2243.700779
Model has been fit using the logistic link.

Untransformed Estimates of coefficients for covariates (Beta's)
==============================================================================
                                                          estimate   std.error
A1     :occupancy        psi1                              0.646207 (0.157010)
B1     :colonization     gam1                             -0.664856 (0.315491)
C1     :local extinction eps1                             -0.294950 (0.267860)
D1     :detection        P[1-1]                            0.344269 (0.152992)
D2     :detection        P[1-2]                            0.146092 (0.144721)
D3     :detection        P[2-1]                           -0.281412 (0.209442)
D4     :detection        P[2-2]                           -0.500775 (0.200374)

Variance-Covariance Matrix of Untransformed estimates:
                 A1         B1         C1         D1         D2         D3         D4
     A1    0.024652  -0.010787  -0.000000  -0.013088  -0.011711  -0.000000  -0.000000
     B1   -0.010787   0.099535  -0.009428   0.008942   0.008001  -0.023571  -0.021574
     C1   -0.000000  -0.009428   0.071749   0.000000   0.000000   0.036470   0.033380
     D1   -0.013088   0.008942   0.000000   0.023407   0.009708   0.000000   0.000000
     D2   -0.011711   0.008001   0.000000   0.009708   0.020944   0.000000   0.000000
     D3   -0.000000  -0.023571   0.036470   0.000000   0.000000   0.043866   0.024998
     D4   -0.000000  -0.021574   0.033380   0.000000   0.000000   0.024998   0.040150
------------------------------
```

Note that the PRESENCE output specifies the total number of sampling occasions is 4, and the number of primary sampling occasions is 2. This model estimates 7 parameters. It is not over-parameterized because there are 16 possible histories, so we can run a model with up to 15 parameters legitimately. For comparison, the spreadsheet results are shown, and the results match.

	F	G	H	I
5	Parameter	Estimate?	Betas	MLE
6	p1	1	0.344269481	0.58523
7	p2	1	0.146092594	0.53646
8	p3	1	-0.281413076	0.43011
9	p4	1	-0.500775891	0.37736
10	ψ	1	0.646207396	0.65616
11	ε1	1	-0.294950494	0.42679
12	γ1	1	-0.664854542	0.33965
13		**OUTPUTS**		
14		Log_eL =	-1114.85	
15		$-2Log_eL$ =	2229.70	
16		Deviance =	5.906	
17		Model DF =	9	
18		C hat =	0.66	
19		Effective N	500	
20		K =	7	
21		AIC =	2243.70	
22		AICc =	2243.93	

Scroll down a bit further, and you'll see the detection estimates for each site for each survey listed:


pres1054.tmp - Notepad

File Edit Format View Help

```
  Individual Site estimates of p:

       Site          Survey            p       Std.err        95% conf. interval
   1     site 1     1    1-1:     0.585227  0.037137      0.512439 -    0.658015
   1     site 1     2    1-2:     0.536458  0.035988      0.465922 -    0.606995
   1     site 1     3    2-1:     0.430108  0.051337      0.329486 -    0.530729
   1     site 1     4    2-2:     0.377359  0.047080      0.285082 -    0.469635
   2     site 2     1    1-1:     0.585227  0.037137      0.512439 -    0.658015
   2     site 2     2    1-2:     0.536458  0.035988      0.465922 -    0.606995
   2     site 2     3    2-1:     0.430108  0.051337      0.329486 -    0.530729
   2     site 2     4    2-2:     0.377359  0.047080      0.285082 -    0.469635
   3     site 3     1    1-1:     0.585227  0.037137      0.512439 -    0.658015
   3     site 3     2    1-2:     0.536458  0.035988      0.465922 -    0.606995
   3     site 3     3    2-1:     0.430108  0.051337      0.329486 -    0.530729
   3     site 3     4    2-2:     0.377359  0.047080      0.285082 -    0.469635
   4     site 4     1    1-1:     0.585227  0.037137      0.512439 -    0.658015
   4     site 4     2    1-2:     0.536458  0.035988      0.465922 -    0.606995
   4     site 4     3    2-1:     0.430108  0.051337      0.329486 -    0.530729
   4     site 4     4    2-2:     0.377359  0.047080      0.285082 -    0.469635
   5     site 5     1    1-1:     0.585227  0.037137      0.512439 -    0.658015
   5     site 5     2    1-2:     0.536458  0.035988      0.465922 -    0.606995
   5     site 5     3    2-1:     0.430108  0.051337      0.329486 -    0.530729
   5     site 5     4    2-2:     0.377359  0.047080      0.285082 -    0.469635
   6     site 6     1    1-1:     0.585227  0.037137      0.512439 -    0.658015
```

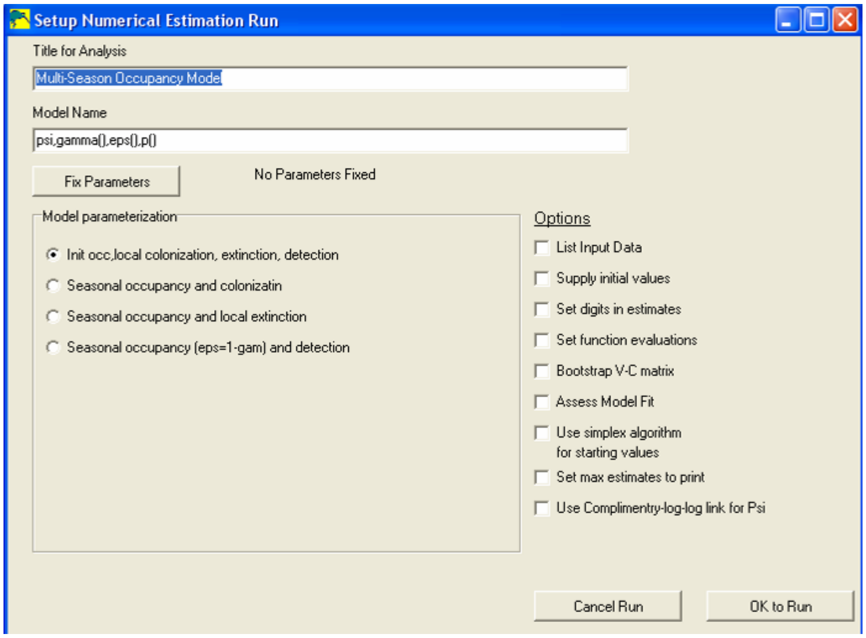
OK, now let's run our second model.

MODEL $\psi(.)\varepsilon(.)\gamma(.)p(.)$

In this model, we estimate ψ, ε_1, γ_1, and set p as constant across all occasions. In the spreadsheet, this model was set up as:

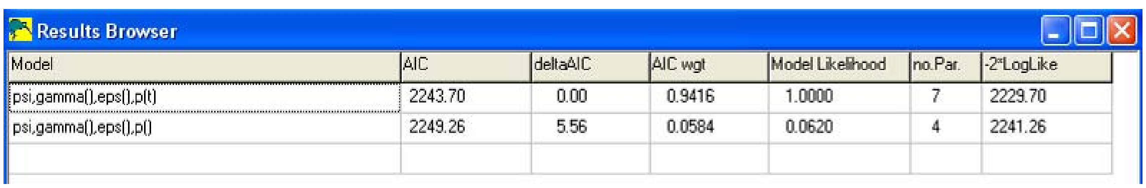
	F	G	H	I
5	Parameter	Estimate?	Betas	MLE
6	p1	1		=EXP(H6)/(1+EXP(H6))
7	p2	0	=H6	=EXP(H7)/(1+EXP(H7))
8	p3	0	=H6	=EXP(H8)/(1+EXP(H8))
9	p4	0	=H6	=EXP(H9)/(1+EXP(H9))
10	ψ	1		=EXP(H10)/(1+EXP(H10))
11	ε1	1		=EXP(H11)/(1+EXP(H11))
12	γ1	1		=EXP(H12)/(1+EXP(H12))

In PRESENCE, this model is the default model for the first option in the model parameterization option box (Init occ, local extinction, detection):



So all you need to do to run this model is press the "OK to Run" button.

Append the results to the Results Browser:


Results Browser

Model	AIC	deltaAIC	AIC wgt	Model Likelihood	no.Par.	-2*LogLike
psi,gamma(),eps(),p(t)	2243.70	0.00	0.9416	1.0000	7	2229.70
psi,gamma(),eps(),p()	2249.26	5.56	0.0584	0.0620	4	2241.26

Then right-click on the model name to bring up the results:

pres9131.tmp - Notepad

File Edit Format View Help

```
P[2-1]  1
P[2-2]  1
========================


Open Population Model:

Number of sites                        = 500
Total number of sampling occasions     = 4
Number of primary sampling periods     = 2
Number of missing observations         = 0

Number of parameters                   = 4


**** Numerical convergence was not reached.
     Parameter estimates converged to approximately 5.81 significant digits.

-2log(likelihood)                      = 2241.260481
AIC                                    = 2249.260481
Model has been fit using the logistic link.

Untransformed Estimates of coefficients for covariates (Beta's)
==============================================================================
                                                          estimate    std.error
A1      :occupancy        psi1                            0.859493 (0.181411)
B1      :colonization     gam1                           -1.049538 (0.343199)
C1      :local extinction eps1                            0.039155 (0.170654)
D1      :detection        P[1-1]                          0.017637 (0.102718)

Variance-Covariance Matrix of Untransformed estimates:
              A1          B1          C1          D1
     A1    0.032910   -0.014543   -0.007505   -0.011650
     B1   -0.014543    0.117786    0.022000    0.005159
     C1   -0.007505    0.022000    0.029123    0.006798
     D1   -0.011650    0.005159    0.006798    0.010551
```

The spreadsheet answers also match:

	F	G	H	I
5	Parameter	Estimate?	Betas	MLE
6	p1	1	0.017637275	0.50441
7	p2	0	0.017637275	0.50441
8	p3	0	0.017637275	0.50441
9	p4	0	0.017637275	0.50441
10	ψ	1	0.859492002	0.70255
11	$\varepsilon 1$	1	0.039154811	0.50979
12	$\gamma 1$	1	-1.049538461	0.25931
13		**OUTPUTS**		
14		Log_eL =	-1120.63	
15		$-2Log_eL$ =	2241.26	
16		Deviance =	17.466	
17		Model DF =	12	
18		C hat =	1.46	
19		Effective N	500	
20		K =	4	
21		AIC =	2249.26	
22		AICc =	2249.34	

That's all the models we ran in the spreadsheet. You can easily add covariates to the multi-season model by following the same steps that were described in Exercises 4 and 5. In selecting option 1 for the model run, you can constrain ψ_1, γ_1, ε_1, or p to be functions of specified covariates. Remember, you can derive occupancy for season 2 (ψ_2) by knowing the estimates of γ_1 and ε_1. The formula is:

$$\psi_{t+1} = \psi_t(1-\varepsilon_t) + (1-\psi_t)\gamma_t$$

In words, the proportion of sites that are occupied in season 2 is the proportion of sites occupied in season 1 that failed to go extinct, plus the proportion of empty sites in season 1 that became occupied. Using the estimates from the current model,

$$\psi_2 = (0.70255)*(1-0.50979)+(1-0.70255)*0.25931) = 0.42153.$$

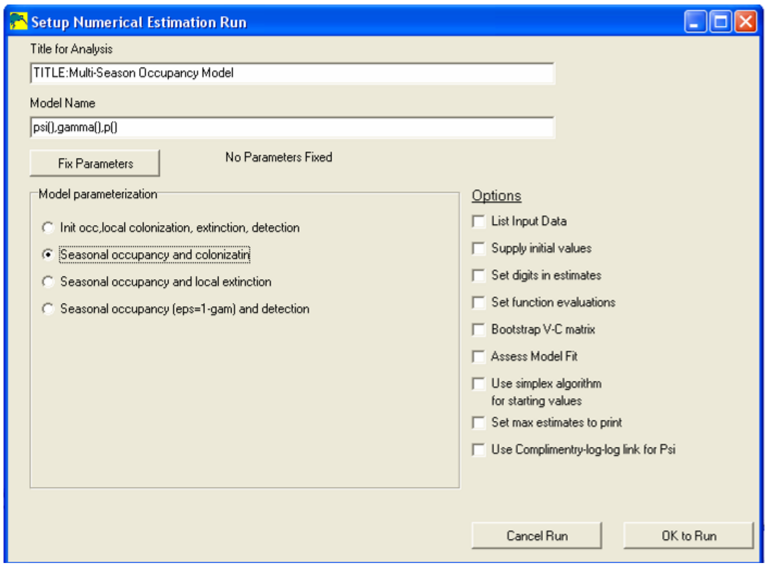
To get these estimates, you need to back-transform the beta estimates to probabilities. Now, what if you want to model ψ_2 directly? The benefit to running the model in this way is that you can model ψ_2 as a function of covariates directly. Fortunately, PRESENCE allows you to parameterize this model in multiple ways, and in some cases you'll want to use these alternatives. Let's explore these options next.

ALTERNATIVE PARAMETERIZATION OPTION 2

We'll now run model $\psi(.)\varepsilon(.)\gamma(.)p(.)$ using three different parameterizations in PRESENCE, starting with Alternative Parameterization Option 2. Click on the main form in PRESENCE (with the book picture), and go to Run | Analysis: Multi-Season. Look at the option box labeled Model parameterization, and you'll see four options:

1. Init occ, local colonization, extinction, detection
2. Seasonal occupancy and colonization
3. Seasonal occupancy and local extinction
4. Season occupancy (eps = 1-gam) and detection

Choose Option 2:



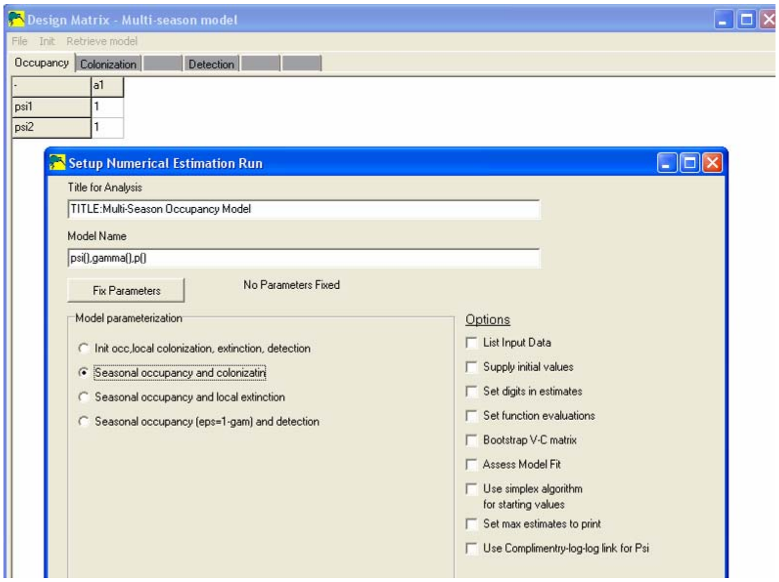
Remember, in option 1 we estimated ψ_1, ε_1, and γ_1, and then derived ψ_2 as:

$$\psi_{t+1} = \psi_t(1-\varepsilon_t) + (1-\psi_t)\gamma_t$$

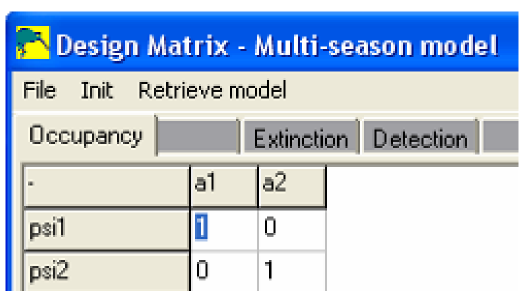
In option 2, we estimate ψ_2 directly, and derive extinction (ε_1). This is done by simply re-arranging the above equation as:

$$\varepsilon_1 = 1 - \frac{\psi_{t+1} - (1-\psi_t)\gamma_t}{\psi_t}$$

When you click on this option, notice that the DM now shows only three tabs (Occupancy, Colonization, and Detection) and that there is no Extinction tab…that's because extinction will be derived. Under the Occupancy tab, we can now model both ψ_1 and ψ_2 (the default is that the two estimates are equal, with the corresponding beta called a1).

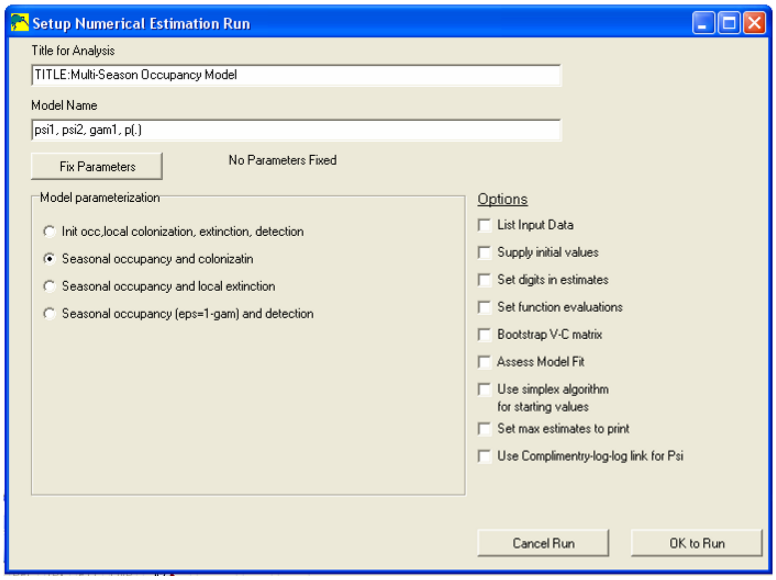


To run model $\psi(.)\varepsilon(.)\gamma(.)p(.)$ with this new parameterization, we need to estimate ψ_1 distinctly from ψ_2, so click in the occupancy tab and go to Init | Full Identity, and you'll see that ψ_1 will be estimated uniquely (with a corresponding beta called a1) and that ψ_2 will be estimated uniquely (with a corresponding beta called a2):



Click on the Extinction tab and you'll see that the model is set up to estimate ε_1 (with a corresponding beta called c1). Then click on the Occupancy tab and keep the default option, which is that all four p's are constrained to be equal (with a corresponding beta called d1).

Let's run this model, but let's change the title to "psi1, psi2, gam1, p(.)". Remember, the specified model is exactly the same model as $\psi(.)\varepsilon(.)\gamma(.)p(.)$, but with a different parameterization:



Press the "OK to Run" button and add the results to the Results Browser:


Results Browser

Model	AIC	deltaAIC	AIC wgt	Model Likelihood	no.Par.	-2*LogLike
psi,gamma(),eps(),p(t)	2243.70	0.00	0.8896	1.0000	7	2229.700779
psi,gamma(),eps(),p()	2249.26	5.56	0.0552	0.0620	4	2241.260481
psi1, psi2, gam1, p(.)	2249.26	5.56	0.0552	0.0620	4	2241.26

You should see that both the p(.) models have identical AIC results, as they should – they are the exact same models but with different parameterization. View the output of this new model by right-clicking on the model title:

```
pres6225.tmp - Notepad
File  Edit  Format  View  Help
     Parameter estimates converged to approximately 5.94 significant digits.

-2log(likelihood)                   = 2241.260481
AIC                                 = 2249.260481
Model has been fit using the logistic link.

Untransformed Estimates of coefficients for covariates (Beta's)
==============================================================================
                                                          estimate    std.error
A1      :occupancy        psi1                            0.859493  (0.181414)
A2      :occupancy        psi2                           -0.316484  (0.127361)
B1      :colonization     gam1                           -1.049539  (0.343206)
D1      :detection        P[1-1]                          0.017637  (0.102720)

Variance-Covariance Matrix of Untransformed estimates:
              A1          A2          B1          D1
     A1    0.032911    0.008510   -0.014544   -0.011651
     A2    0.008510    0.016221    0.008880   -0.005991
     B1   -0.014544    0.008880    0.117790    0.005159
     D1   -0.011651   -0.005991    0.005159    0.010551
------------------------------

   Individual Site estimates of Psi:

        Site          Survey          Psi       Std.err        95% conf. interval
     1   site 1    1      1-1:     0.702555   0.037910      0.628250 -    0.776859
     1   site 1    2      1-2:     0.421533   0.031056      0.360663 -    0.482403
     2   site 2    1      1-1:     0.702555   0.037910      0.628250 -    0.776859
     2   site 2    2      1-2:     0.421533   0.031056      0.360663 -    0.482403
```

Notice that PRESENCE provides the beta estimates for psi1, psi2, gam1, and p, and then towards the bottom of the output lists psi1 and psi2 on a site by site basis. Because no covariates were specified in this model, all sites have a ψ_1 estimate of 0.70255, and all sites have a ψ_2 estimate of 0.421533. The estimates of epsilon are not shown, but we can derive the estimate as

$$\varepsilon_1 = 1 - \frac{\psi_{t+1} - (1-\psi_t)\gamma_t}{\psi_t}$$

Given the parameter estimates from this model, first let's back-transform the beta estimates to obtain γ_1:

γ_1 = exp(-1.049)/(1+exp(-1.049) = 0.25931. Now, ε_1 is derived as:

$$1 - \frac{0.421533 - (1-0.70255)*0.25931}{0.70255} = 0.5097$$

Hopefully this makes sense. It's just a different way of obtaining estimates from the model, but in a way that you can model the parameters you select for more detailed covariate analysis.

ALTERNATIVE PARAMETERIZATION OPTION 3

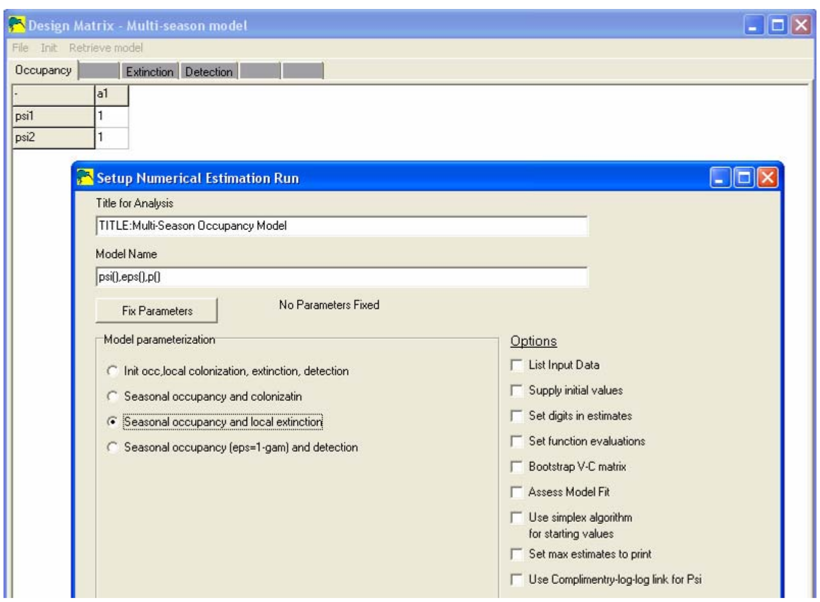
Remember, in option 1 we estimated ψ, ε_1, and γ_1, and then derived ψ_2 as:

$$\psi_{t+1} = \psi_t(1-\varepsilon_t) + (1-\psi_t)\gamma_t$$

In option 2, we estimate γ_1 and ψ_2 directly and derive extinction (ε_1). In option 3, we estimate ε_1 and ψ_2 directly, and derive colonization (γ_1). This is done by simply re-arranging the above equation as:

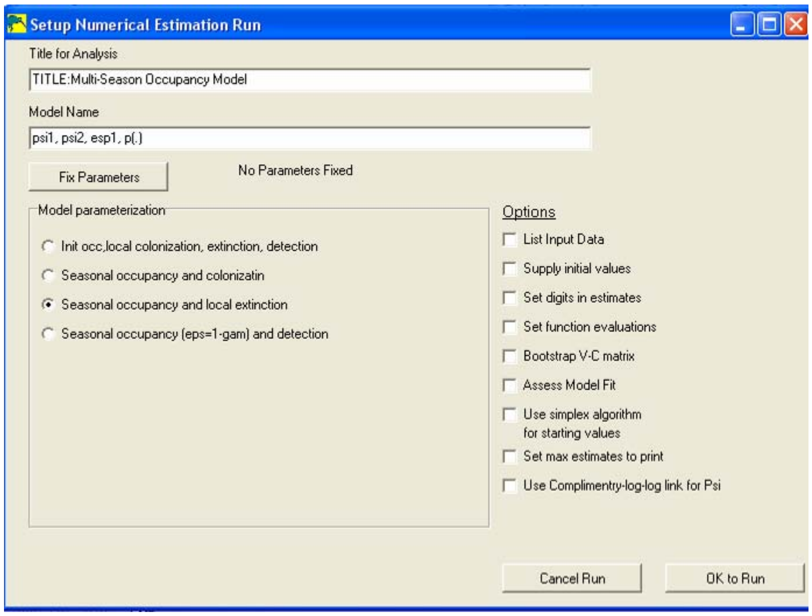
$$\gamma_1 = \frac{\psi_{t+1} - \psi_t(1-\varepsilon_t)}{(1-\psi_t)}$$

Go to Run | Analysis: Multi-Season, and select Option 3:



Note that there is no Colonization tab in the DM for this option. Again, we'll need to tell PRESENCE to estimate ψ1, and ψ2 uniquely so that we are matching model $\psi(.)\varepsilon(.)\gamma(.)p(.)$. Click on that tab and then go to Init | Full Identity.

Now we can run this model, but again let's rename it "psi1, psi2, esp1, p(.)" as shown:



Add the results to the Results Browser, and you'll see the same AIC values and other outputs as the previous two options:

Results Browser

Model	AIC	deltaAIC	AIC wgt	Model Likelihood	no.Par.	-2*LogLike
psi,gamma(),eps(),p(t)	2243.70	0.00	0.8431	1.0000	7	2229.700779
psi,gamma(),eps(),p()	2249.26	5.56	0.0523	0.0620	4	2241.260481
psi1, psi2, gam1, p(.)	2249.26	5.56	0.0523	0.0620	4	2241.26
psi1, psi2, esp1, p(.)	2249.26	5.56	0.0523	0.0620	4	2241.26

Go ahead and study the PRESENCE output for this model, and see if you can derive the estimate for colonization.

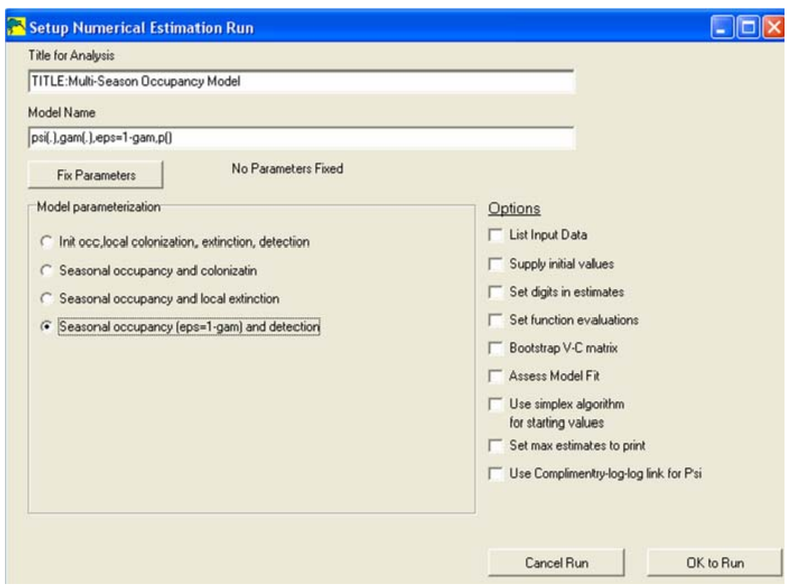
ALTERNATIVE PARAMETERIZATION OPTION 4

Remember, in option 1 we estimated ψ, ε_1, and γ_1, and then derived ψ_2 as:

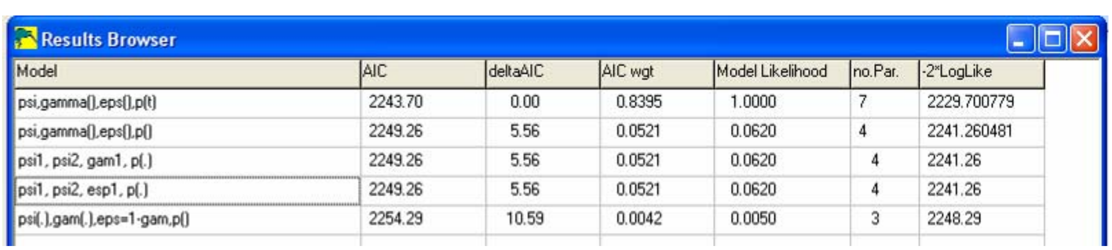
$$\psi_{t+1} = \psi_t(1-\varepsilon_t) + (1-\psi_t)\gamma_t$$

In option 2, we estimate ψ_2 directly, and derive extinction (ε_1). In option 3, we estimate ψ_2 directly and derive colonization (γ_1). Option 4 is more similar to option 1 in that we estimate ψ_1 only, but in this model we force extinction to be equal to 1 minus colonization. Thus, $\varepsilon_1 = 1-\gamma_1$. Thus, if extinction probability is 0.2, then probability of an empty site NOT being colonized is also 0.2. What is the significance of this model? Actually, I'm not sure why this model is significant…I thought the Levins model balanced extinction with recolonization. …

To run this model, simply select option 4 and press the "OK to Run" button:



Add the results to the Results Browser:



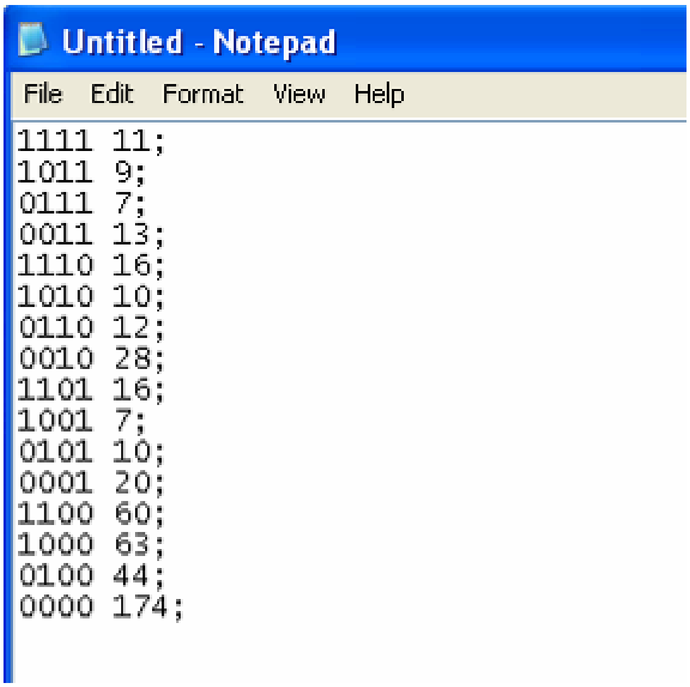
Model	AIC	deltaAIC	AIC wgt	Model Likelihood	no.Par.	-2*LogLike
psi,gamma(),eps(),p(t)	2243.70	0.00	0.8395	1.0000	7	2229.700779
psi,gamma(),eps(),p()	2249.26	5.56	0.0521	0.0620	4	2241.260481
psi1, psi2, gam1, p(.)	2249.26	5.56	0.0521	0.0620	4	2241.26
psi1, psi2, esp1, p(.)	2249.26	5.56	0.0521	0.0620	4	2241.26
psi(.),gam(.),eps=1-gam,p()	2254.29	10.59	0.0042	0.0050	3	2248.29

You can see that this model differs from the previous options…it is simply a model where you force the extinction rate to be equal to 1-colonization rate.

MULTIPLE-SEASON OCCUPANCY ANALYSIS IN PROGRAM MARK

GETTING STARTED

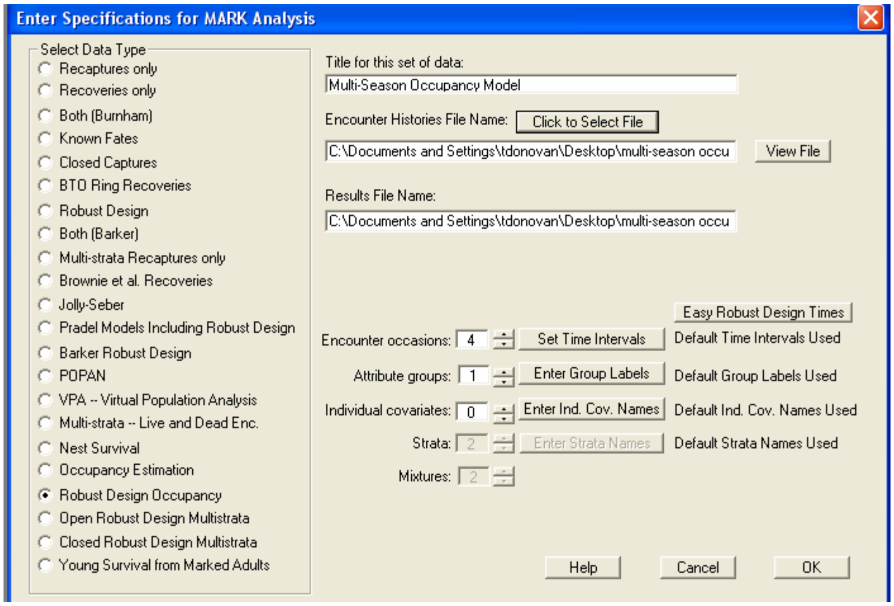
If you haven't already done so, copy cells O6:O21 into NotePad, and save the file as "Multi-Season Occupancy.Inp". This is the file you'll import into MARK.



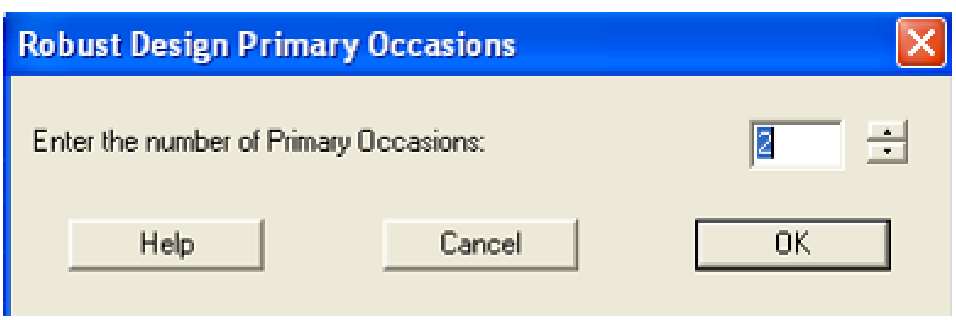

```
1111 11;
1011 9;
0111 7;
0011 13;
1110 16;
1010 10;
0110 12;
0010 28;
1101 16;
1001 7;
0101 10;
0001 20;
1100 60;
1000 63;
0100 44;
0000 174;
```

Now, open MARK and start a new file. Enter a title for the analysis (E.g., Multi-Season Occupancy Model), and browse to your input file. Choose the radio button for the Robust Design Occupancy, and enter 4 as the number of occasions (because there were 2 primary periods, each consisting of 2 samples per site). There are no covariates in this exercise, and the data are not grouped in any way.

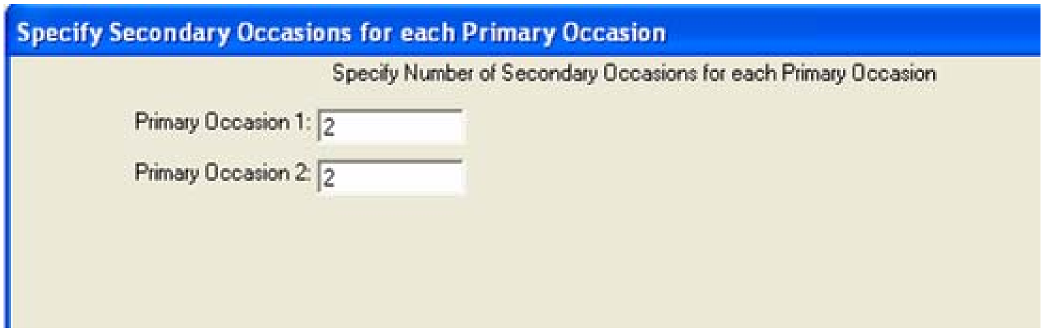
Before clicking OK, we need to tell MARK how the 4 encounter occasions are divided between primary and secondary sessions.



Click on the button labeled Easy Robust Design Times, and enter 2 for the number of primary occasions:



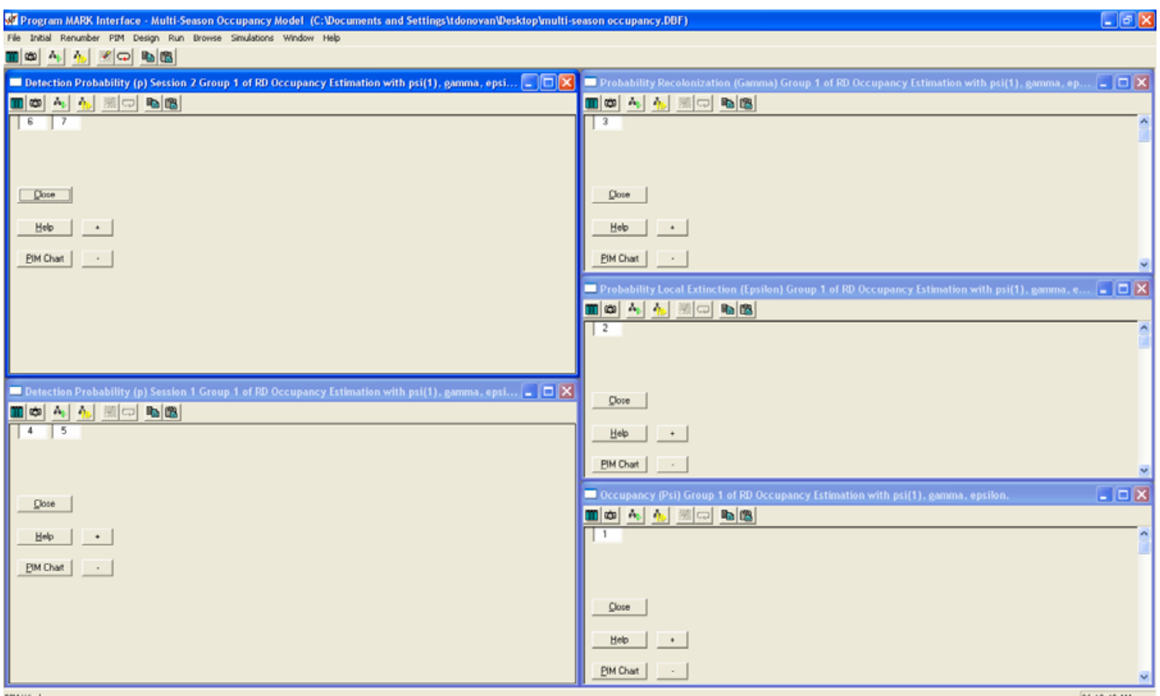
Then enter 2 for each primary occasion to indicate that each primary session consisted of 2 samples per site:



Now you can press OK, and get started with some analyses.

MARK PIM

Let's start by looking at all of the PIMs available to us in the Multi-Season (Robust) Occupancy model. Go to PIM | Open Parameter Index Matrix | Select All | OK, and then tile the windows by selecting Windows | Tile, and you should see all the PIMs as shown:

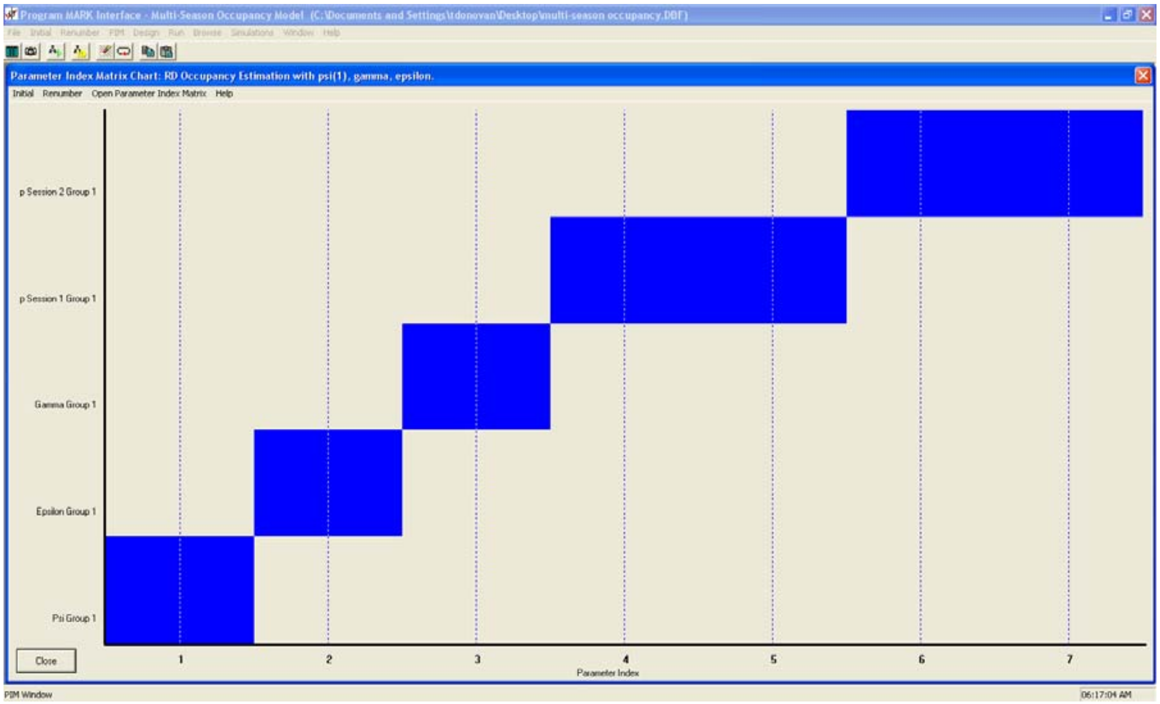


The default model is one in which ψ, ε, γ, p_1, p_2, p_3, and p_4 are separately estimated, so K = 7 for this model. In the spreadsheet, we set up this model as follows, and let Solver find values in cells H6:H12 that maximized the multinomial log likelihood.

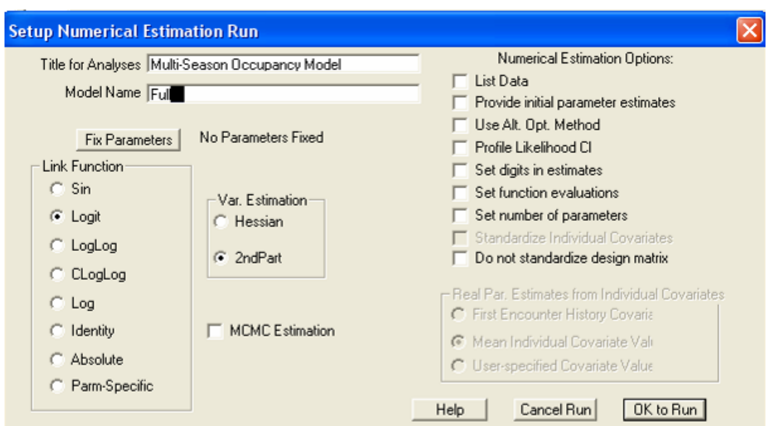
	F	G	H	I
5	Parameter	Estimate?	Betas	MLE
6	p1	1		0.50000
7	p2	1		0.50000
8	p3	1		0.50000
9	p4	1		0.50000
10	ψ	1		0.50000
11	ε1	1		0.50000
12	γ1	1		0.50000

The PIMs also indicate that 7 parameters will be estimated, and the number system tells MARK that ψ will be indexed as Parameter 1, ε will be indexed as Parameter 2, γ will be indexed as Parameter 3, p_1 and p_2 will be indexed as Parameter 4 and Parameter 5, and p_3 and p_4 will be indexed as Parameter 6 and Parameter 7.

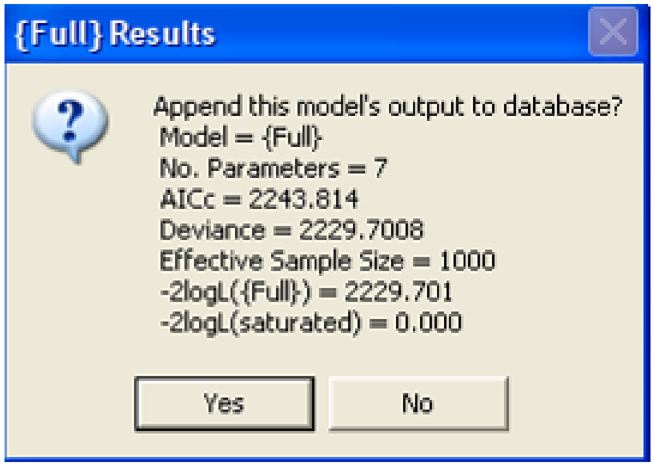
The PIM Chart for this model looks like this:



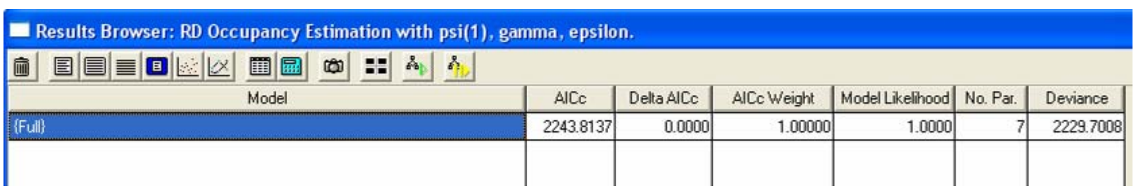
OK, now go to Run | Current Model, and enter a title for this model (e.g., full). Use the logit link since that the link we used in the spreadsheet.



Click OK to run, and use the default identity link. MARK then shows the following dialogue box:

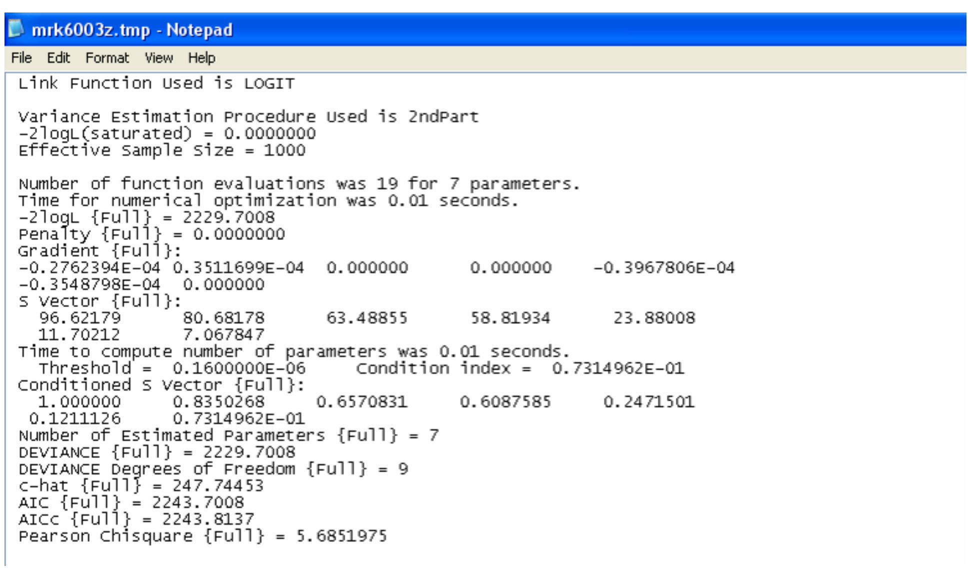


Click Yes, and the results will be added to the Results Browser:



Model	AICc	Delta AICc	AICc Weight	Model Likelihood	No. Par.	Deviance
{Full}	2243.8137	0.0000	1.00000	1.0000	7	2229.7008

By now, you should be a pro at reading and interpreting MARK's full output. Let's study the results:




```
Link Function Used is LOGIT

Variance Estimation Procedure Used is 2ndPart
-2logL(saturated) = 0.0000000
Effective Sample Size = 1000

Number of function evaluations was 19 for 7 parameters.
Time for numerical optimization was 0.01 seconds.
-2logL {Full} = 2229.7008
Penalty {Full} = 0.0000000
Gradient {Full}:
-0.2762394E-04 0.3511699E-04  0.000000      0.000000     -0.3967806E-04
-0.3548798E-04  0.000000
S Vector {Full}:
  96.62179      80.68178      63.48855      58.81934      23.88008
  11.70212      7.067847
Time to compute number of parameters was 0.01 seconds.
  Threshold =  0.1600000E-06     Condition index =  0.7314962E-01
Conditioned S Vector {Full}:
  1.000000     0.8350268      0.6570831      0.6087585      0.2471501
 0.1211126     0.7314962E-01
Number of Estimated Parameters {Full} = 7
DEVIANCE {Full} = 2229.7008
DEVIANCE Degrees of Freedom {Full} = 9
c-hat {Full} = 247.74453
AIC {Full} = 2243.7008
AICc {Full} = 2243.8137
Pearson Chisquare {Full} = 5.6851975
```

	G	H
13	**OUTPUTS**	
14	Log_eL =	-1114.85
15	-2Log_eL =	2229.70
16	Deviance =	5.906
17	Model DF =	9
18	C hat =	0.66
19	Effective N	500
20	K =	7
21	AIC =	2243.70
22	AICc =	2243.93

We can see that the -2Log_eL's are the same for MARK and the spreadsheet. However, MARK assumes that -2Log_eL for the saturated model is 0, and therefore reports Deviance as this model's -2Log_eL – 0 = -2Log_eL. The spreadsheet computes Deviance differently, and because of this, c-hat is different as well. This model estimates 7 parameters and has a model DF = 9. How was that computed? Well, there are 16 histories possible, so there are 16 terms in the multinomial log likelihood, in which we need to estimate a probability of realizing each history. We used the multi-season occupancy model to derive these probabilities, and used 7 parameters to generate them (ψ, ε, γ, p_1, p_2, p_3, p_4). That leaves us with 16 – 7 = 9 "leftover" parameters, which is the model DF. The AIC and AICc computations also match.

PEARSON CHI-SQUARE

What about the Pearson Chi-Square? We see that the MARK output provides the same result as cell L22 in the spreadsheet. Do you recall how to look at the observed and expected values in MARK? Go to Output | Specific Model Output | Residuals | List Observed and Expected in NotePad:

	C	D	E	K	L
4				Chi Square	
5	History		Freq	Expected	$(O-E)^2/E$
6	11	11	11	9.583	0.2096754
7	10	11	9	8.280	0.0625994
8	01	11	7	6.792	0.0064008
9	00	11	13	15.346	0.358616
10	11	10	16	15.811	0.002255
11	10	10	10	13.662	0.9816133
12	01	10	12	11.206	0.0562613
13	00	10	28	25.321	0.2834967
14	11	01	16	12.697	0.859329
15	10	01	7	10.971	1.4373614
16	01	01	10	8.999	0.1114062
17	00	01	20	20.333	0.0054644
18	11	00	60	64.909	0.3713234
19	10	00	63	56.087	0.8521174
20	01	00	44	46.004	0.0872777
21	00	00	174	174.000	4.068E-12
22	total sites =		500	500.0	5.6852
23	total histories =		16	p =	0.7709646

mrk9048z.tmp - Notepad

```
Robust Occupancy Model

Model: {Full}

Encounter          Group 1            Group 1
Histories          Observed           Expected
1111        11.0          9.5825
1110        16.0         15.8112
1101        16.0         12.6969
1100        60.0         64.9094
1011         9.0          8.2800
1010        10.0         13.6621
1001         7.0         10.9711
1000        63.0         56.0868
0111         7.0          6.7915
0110        12.0         11.2060
0101        10.0          8.9987
0100        44.0         46.0038
0011        13.0         15.3459
0010        28.0         25.3208
0001        20.0         20.3333
0000       174.0        174.0000
```

The results match (although the histories are sorted a bit differently). Remember, by knowing the expected frequencies and total number of sites, you can derive the probability of getting any history by dividing the expected history by the total number of sites.

BETA ESTIMATES AND PARAMETER ESTIMATES

OK, what about the estimates from MARK?

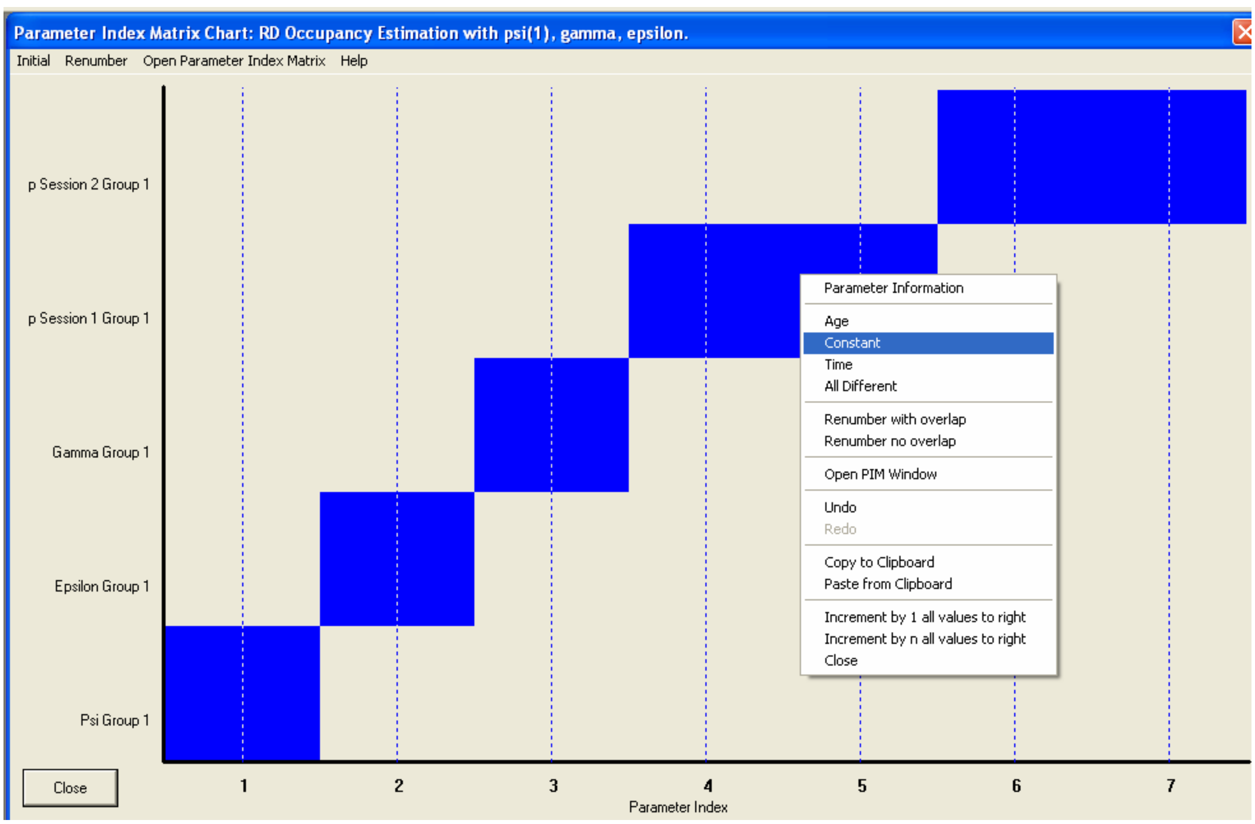
```
mrk6003z.tmp - Notepad
File  Edit  Format  View  Help
Pearson Chisquare {Full} = 5.6851975

                        LOGIT Link Function Parameters of {Full}
                                                     95% Confidence Interval
Parameter                 Beta         Standard Error      Lower           Upper
-------------------------  --------------  --------------  --------------  --------------
   1:Psi                  0.6462073       0.1570126       0.3384626       0.9539520
   2:Epsilon             -0.2949499       0.2678653      -0.8199659       0.2300662
   3:Gamma               -0.6648540       0.3154939      -1.2832221      -0.0464859
   4:p Session 1          0.3442698       0.1529946       0.0444004       0.6441392
   5:p Session 1          0.1460928       0.1447227      -0.1375637       0.4297493
   6:p Session 2         -0.2814126       0.2094464      -0.6919275       0.1291024
   7:p Session 2         -0.5007755       0.2003781      -0.8935166      -0.1080343

                        Real Function Parameters of {Full}
                                                     95% Confidence Interval
Parameter                 Estimate       Standard Error      Lower           Upper
-------------------------  --------------  --------------  --------------  --------------
□  Program  MARK  - Survival Rate Estimation with Capture-Recapture Data
   Compaq Version 4.4(Win32) May        21-Dec-2006 06:20:17    Page  004
   Multi-Season Occupancy Model
 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

   1:Psi                  0.6561553       0.0354245       0.5838170       0.7219093
   2:Epsilon              0.4267925       0.0655307       0.3057709       0.5572642
   3:Gamma                0.3396501       0.0707615       0.2170022       0.4883806
   4:p Session 1          0.5852273       0.0371373       0.5110983       0.6556885
   5:p Session 1          0.5364584       0.0359883       0.4656632       0.6058138
   6:p Session 2          0.4301075       0.0513385       0.3336044       0.5322308
   7:p Session 2          0.3773585       0.0470807       0.2903847       0.4730177

                    Estimates of Derived Parameters

                        Psi Estimates of {Full}
                                        95% Confidence Interval
Grp. Str. Per.      Psi-hat        Standard Error      Lower           Upper
---- ----  --------------  --------------  --------------  --------------
  1     1   0.6561553       0.0354245       0.5867233       0.7255873
  1     2   0.4929000       0.0515276       0.3919059       0.5938942

    Time in seconds for last procedure was 0.05

□  Program  MARK  - Survival Rate Estimation with Capture-Recapture Data
   Compaq Version 4.4(Win32) May        21-Dec-2006 06:20:17    Page  005
   Multi-Season Occupancy Model
 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

	F	G	H	I
5	Parameter	Estimate?	Betas	MLE
6	p1	1	0.344269485	0.58523
7	p2	1	0.146092624	0.53646
8	p3	1	-0.281413068	0.43011
9	p4	1	-0.500775969	0.37736
10	ψ	1	0.646207271	0.65616
11	ε1	1	-0.294950958	0.42679
12	γ1	1	-0.664854695	0.33965

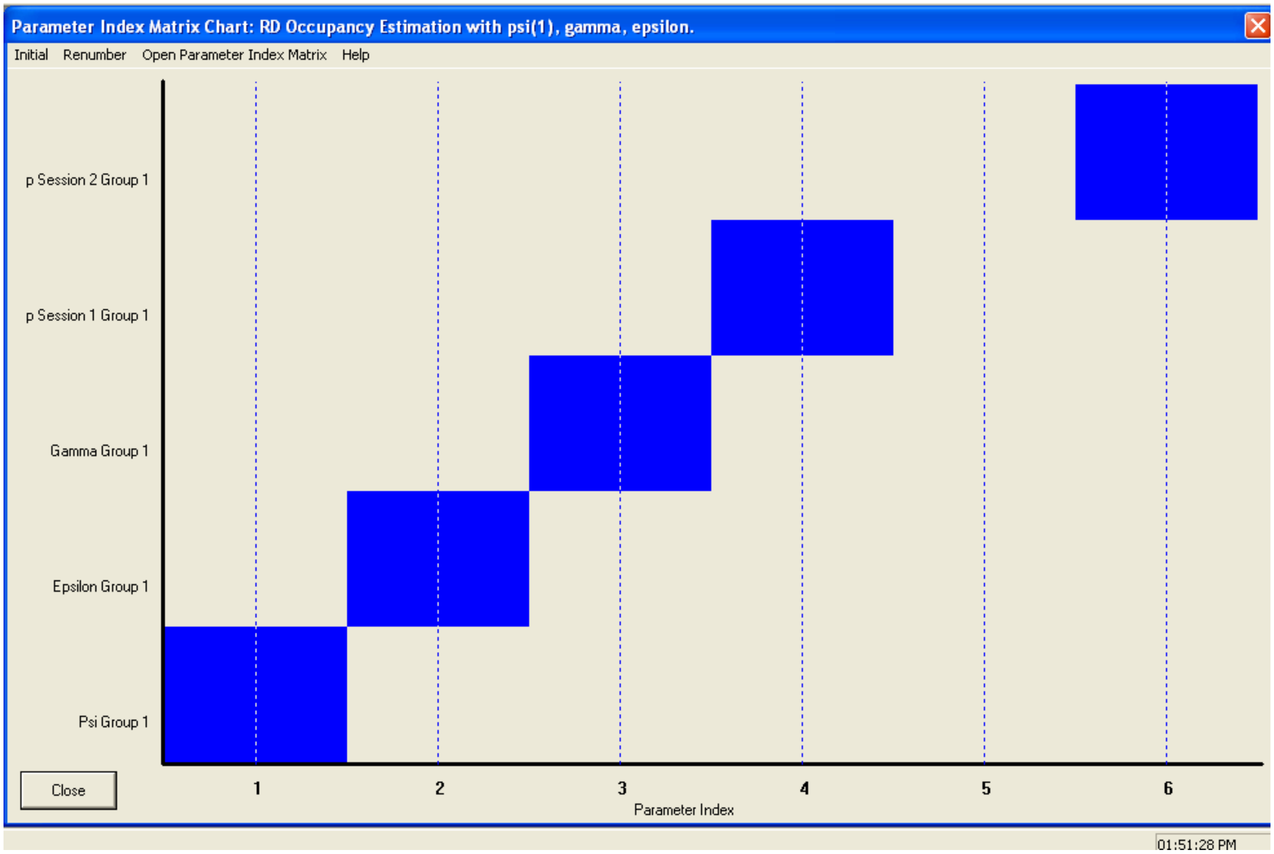
You can see that the estimates match, though now you should very closely inspect the standard errors and confidence limits to determine if MARK had any trouble with the estimation procedure. Note that MARK provides derived estimates of y2 towards the bottom of the output. Recall that y2 was derived based on estimates of ψ_1, ε_1, and γ_1 as follows: $\psi_{t+1} = \psi_1(1-\varepsilon_1) + (1-\psi_1)\gamma_1$. You can add this equation to the spreadsheet if you'd like.

MODEL $\psi(.)\varepsilon(.)\gamma(.)p(.)$

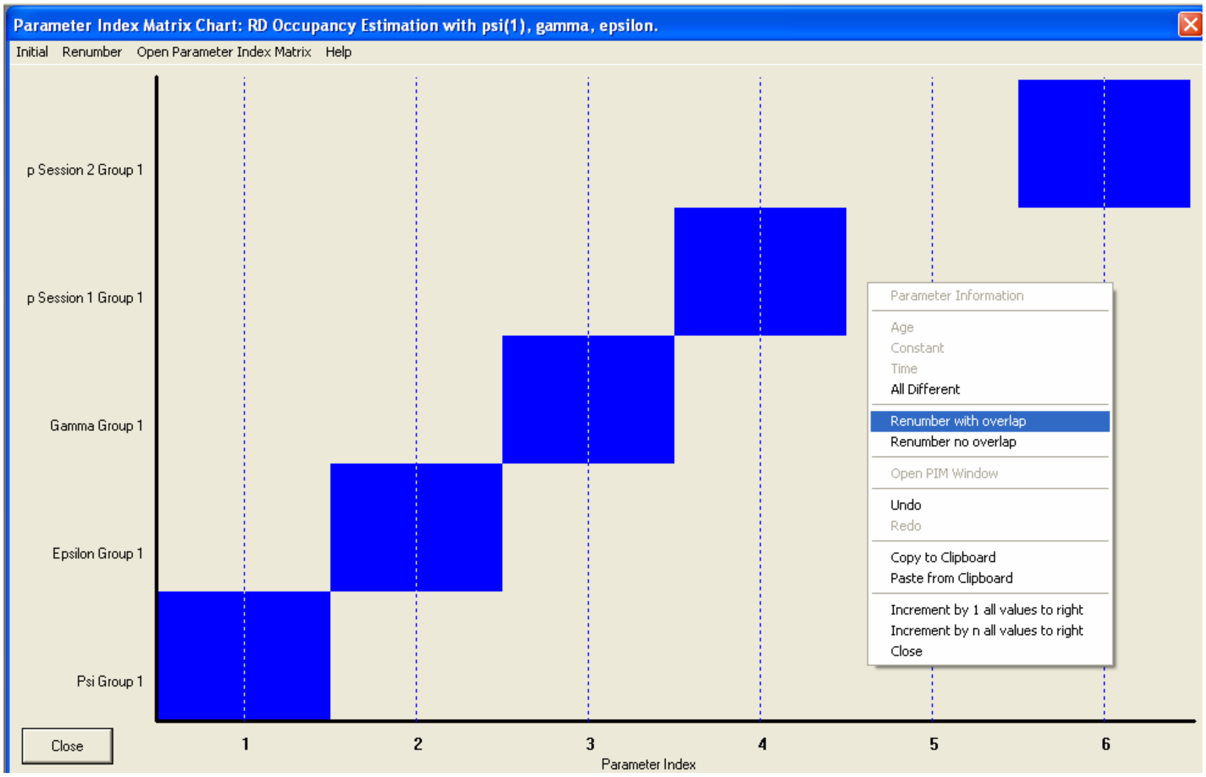
Let's run just one more model, one in which all four p's are constrained to be equal to each other. To run this model, you can alter the PIMs so that the parameter indices for p are all the same number. Or you can alter the PIM chart. Go to PIM | PIM Chart, and the chart from the last model retrieved will appear. Now, all we need to do is force p1 = p2 = p3, = p4. We can do that easily by right-clicking on one of the p's, and selecting the "make constant" option.



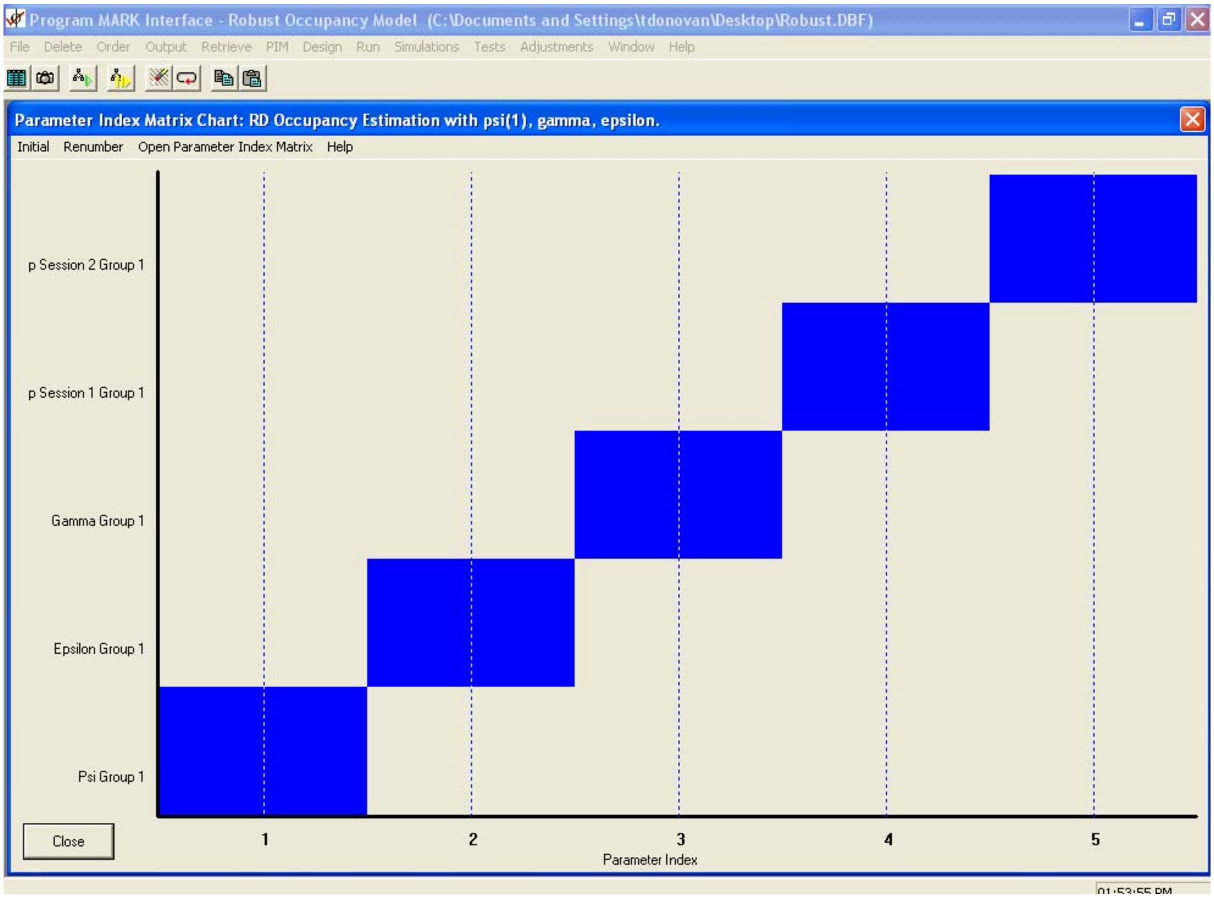
Repeat this step for the p's in session 2, and your PIM chart now looks as shown:



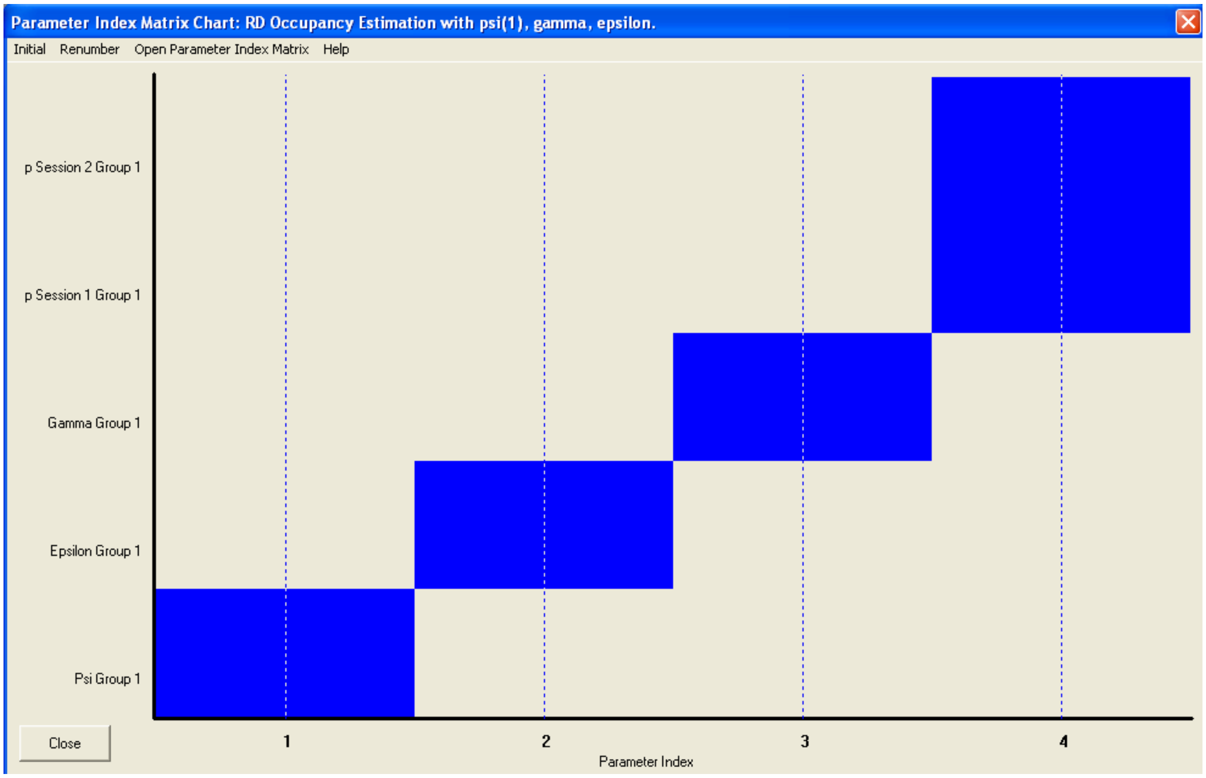
While there's nothing "wrong" with this chart, the parameter indices go from 1 to 4, then over to 6. You can renumber the PIMs by right-clicking on the chart, and selecting "renumber with overlap".



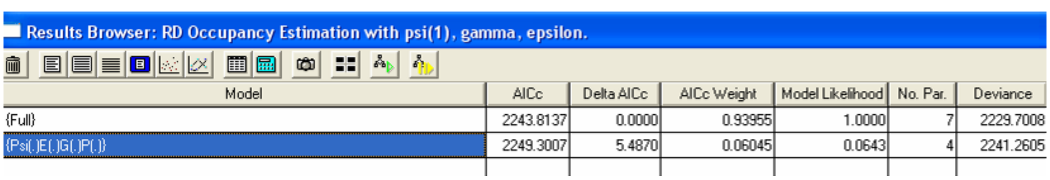
And you'll get a new PIM chart that looks more reasonable:



Now, you'll want to select the top right p box, and drag it so that it overlays the other p box (forcing the p's to be equal).



Go ahead and run this model, and call it Psi(.)E(.)G(.)P(.), and add the results to the Results Browser:


Results Browser: RD Occupancy Estimation with psi(1), gamma, epsilon.

Model	AICc	Delta AICc	AICc Weight	Model Likelihood	No. Par.	Deviance
{Full}	2243.8137	0.0000	0.93955	1.0000	7	2229.7008
{Psi(.)E(.)G(.)P(.)}	2249.3007	5.4870	0.06045	0.0643	4	2241.2605

Now let's look at this model's output:

```
mrk9747z.tmp - Notepad
File  Edit  Format  View  Help

Time for numerical optimization was 0.01 seconds.
-2logL {Psi(.)E(.)G(.)P(.)} = 2241.2605
Penalty {Psi(.)E(.)G(.)P(.)} = 0.0000000
Gradient {Psi(.)E(.)G(.)P(.)}:
 0.2445545E-04 0.4376128E-04  0.000000     0.4468660E-04
S Vector {Psi(.)E(.)G(.)P(.)}:
  203.0218      45.32020      26.96506      7.917554
Time to compute number of parameters was 0.01 seconds.
  Threshold =  0.1000000E-06     Condition index =  0.3899854E-01
Conditioned S Vector {Psi(.)E(.)G(.)P(.)}:
  1.000000      0.2232282     0.1328185     0.3899854E-01
Number of Estimated Parameters {Psi(.)E(.)G(.)P(.)} = 4
DEVIANCE {Psi(.)E(.)G(.)P(.)} = 2241.2605
DEVIANCE Degrees of Freedom {Psi(.)E(.)G(.)P(.)} = 12
c-hat {Psi(.)E(.)G(.)P(.)} = 186.77171
AIC {Psi(.)E(.)G(.)P(.)} = 2249.2605
AICc {Psi(.)E(.)G(.)P(.)} = 2249.3007
Pearson Chisquare {Psi(.)E(.)G(.)P(.)} = 17.407024


                    LOGIT Link Function Parameters of {Psi(.)E(.)G(.)P(.)}
                                                            95% Confidence Interval
Parameter                    Beta         Standard Error      Lower           Upper
-------------------------  --------------  --------------  --------------  --------------
   1:Psi                   0.8594930       0.1814174       0.5039149       1.2150710
   2:Epsilon               0.0391545       0.1706569       -0.2953329      0.3736420
   3:Gamma                 -1.0495387      0.3432050       -1.7222206      -0.3768568
   4:p Session 1           0.0176370       0.1027216       -0.1836973      0.2189714


                    Real Function Parameters of {Psi(.)E(.)G(.)P(.)}
                                                            95% Confidence Interval
Parameter                  Estimate       Standard Error      Lower           Upper
-------------------------  --------------  --------------  --------------  --------------
   1:Psi                   0.7025547       0.0379111       0.6233789       0.7711950
   2:Epsilon               0.5097874       0.0426479       0.4266988       0.5923387
   3:Gamma                 0.2593137       0.0659194       0.1515854       0.4068852
   4:p Session 1           0.5044091       0.0256784       0.4542044       0.5545252


                     Estimates of Derived Parameters


                    Psi Estimates of {Psi(.)E(.)G(.)P(.)}
                                             95% Confidence Interval
Grp. Str. Per.     Psi-hat          Standard Error      Lower            Upper
---- ----   --------------   --------------  --------------  --------------
  1     1   0.7025547        0.0379111       0.6282490       0.7768604
  1     2   0.4215328        0.0310552       0.3606645       0.4824011

    Time in seconds for last procedure was 0.02

  Program  MARK  - Survival Rate Estimation with Capture-Recapture Data
  Compaq Version 4.4(Win32) May        21-Dec-2006 06:58:19    Page  004
  Multi-Season Occupancy Model
 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

Work your way through this output, checking the spreadsheet result as you go. Make sure you completely understand how each output is computed! Now let's look at the spreadsheet estimates….

	F	G	H	I
5	Parameter	Estimate?	Betas	MLE
6	p1	1	0.017637275	0.50441
7	p2	0	0.017637275	0.50441
8	p3	0	0.017637275	0.50441
9	p4	0	0.017637275	0.50441
10	ψ	1	0.859492002	0.70255
11	ε1	1	0.039154811	0.50979
12	γ1	1	-1.049538461	0.25931
13		**OUTPUTS**		
14		Log_eL =	-1120.63	
15		-2Log_eL =	2241.26	
16		Deviance =	17.466	
17		Model DF =	12	
18		C hat =	1.46	
19		Effective N	500	
20		K =	4	
21		AIC =	2249.26	
22		AICc =	2249.34	

In the words of Evan Cooch, "Nice when things work out, eh?" That's wraps up this model. Of course, you could make this model much more interesting by adding covariates. You'd simply use the Design Matrix and the same steps we used in the covariate exercises, but this time apply the covariate constraints to p, ψ, ε, or γ.