EXERCISE 9: DESIGN OF SINGLE SEASON OCCUPANCY STUDIES

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

TABLE OF CONTENTS

DESIGN OF SINGLE SEASON OCCUPANCY STUDIES SPREADSHEET EXERCISE

OBJECTIVES

- To review the standard, single-species, single-season occupancy model and the multinomial log likelihood.
- To use Solver to find the maximum likelihood estimates for the probability of detection and the probability of site occupancy under the standard design.
- To derive the variance of the occupancy estimate and p^* (overall detection probability).
- To learn how to simulate data for a standard design occupancy model.
- To evaluate the sensitivity of the standard design occupancy model to varying numbers of sites and surveys, given a species with known p and ψ (using simulated and expected data).
- To evaluate the sensitivity of the standard design occupancy model for a variety of species, given that the number of sites and surveys is known.

INTRODUCTION

In this exercise, we'll return to the single-species, single-season occupancy model. You might recall that we were introduced to this model in Chapter 3, and in Chapters 4 and 5 you learned how to include covariates in the analysis. Here, our focus is on learning how to design an optimal single-species, single-season occupancy model. This topic is discussed in detail in chapter 6 in the book, "Occupancy Estimation and Modeling," where the authors consider issues related to how to define a site, how to select a site, how to define a season, how to

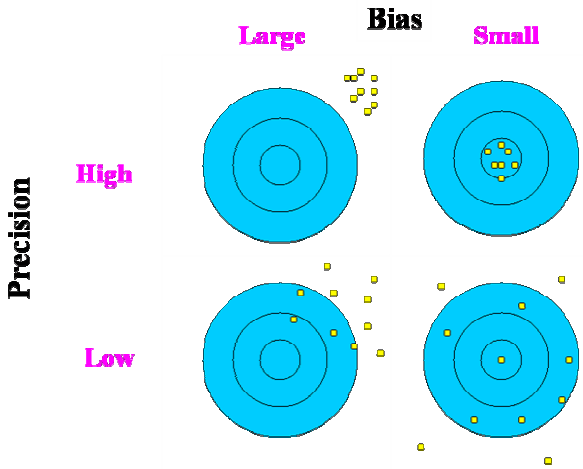
incorporate repeated surveys, and how to allocate effort. If you have not read that chapter yet, it certainly would be beneficial to do so now.

In this exercise, we will assume that you already have settled on what constitutes a site and how to select them. Our focus will be on understanding the trade-offs between increasing the number of sites that are surveyed versus increasing the number of times each site is visited (Section 6.5). For example, consider the following two study designs. The first involves 50 sites that are each surveyed four times, while the second involves 100 sites that are each surveyed twice. In both cases, the total number of surveys is 200, but which design is better suited for a species of interest? The key is to remember that the primary goal of modeling is to derive estimates of p and ψ, and ideally these estimates will be both unbiased and precise.

PRECISION AND BIAS IN MODELING

Precision and bias are both very important concepts in the modeling process. Model bias is the extent to which the model truly represents the population parameters. An unbiased model will produce accurate parameter estimates that reflect those of the true population. For example, if occupancy rate (ψ) is truly 0.7 but a model estimates this rate as 0.6, the model is biased. If the model estimated ψ = 0.7, it would provide an unbiased estimate of ψ. In contrast to bias, a model with high precision will have low standard error rates associated with estimated parameters (in this case, detection probability and occupancy). For example, if ψ = 0.7 with a standard error of 0.05 is far more precise than a model with a standard error of 0.20. You have much more confidence in precisely estimated parameters compared to imprecisely estimated parameters. But keep in

mind that just because a model is precise does not mean that it is unbiased. The figure below gives a visual representation of precision and bias.



In the upper right-hand quadrant, we find parameter estimates that have small bias (the estimate the true rate well) and high precision. This is ideal. In the upper left-hand quadrant, we find parameter estimates that are precise, but biased. In the lower left-hand quadrant, we find parameter estimates that are biased and unprecise, and in the lower right-hand quadrant we find parameter estimates that are unbiased but have low precision. In this exercise, we will explore how a single-species, single-season occupancy studies can be optimized to provide precise and unbiased parameter estimates while limiting the number of repeat surveys per site. Let's get started.

REVIEW OF SINGLE-SPECIES, SINGLE-SEASON OCCUPANCY MODEL

Let's start with a very quick review of standard occupancy model. For occupancy models, the goal is to determine the probability that a site is occupied, given that organisms are imperfectly detected. Let's assume that you are interested in developing an occupancy model for a butterfly species. You select 100 study sites,

and set out to survey each site three times in quick succession (or quick enough to assume that the site does not change in occupancy status between surveys). Or, alternatively, if the site is large enough, the site could have been sampled in three different locations in the same time period. In any event, the species is recorded as being present (1) or absent (0) from the site based on your survey efforts. Let's assume that on site 1 the butterfly was detected on the first survey, but not detected on the second or third survey. This is called an encounter history for site 1, and would be recorded as 100. In the same way, the encounter history for each site is determined. A history of 110 means that a species of interest was detected on the first and second samples, but not on sample 3. A history of 010 means that the species was detected on the second sampling occasion, but not on the first or third occasion. A history of 000 means the species was not detected on any of the 3 sampling occasions. This does not mean the site was unoccupied; only that the species was not detected there. How many possible histories are there? Because a particular sample occasion has only two outcomes (0 or 1), and there are 3 sampling occasions, there are $2^3 = 8$ kinds of histories in total.

The single-season occupancy analysis focuses on the different kinds of encounter histories (which are like the plant phenotypes in the multinomial exercise #2), their frequencies, and the probability that each frequency is realized. In a nutshell, multinomial maximum likelihood procedures estimate the key parameter of interest, psi (ψ), which is the probability that a site is occupied, as well as p_i (the probability of detecting the species on survey i, given the species is present on the site). For instance, the probability of observing a 100 history is $\psi*p_1*(1-p_2)*(1-p_3)$, and the probability of observing a 000 history is $\psi*(1-p_1)*(1-p_2)*(1-p_3) + (1-\psi)$. If this is all new to you, complete Chapter 3 before going any further.

MODEL ASSUMPTIONS

As explained in MacKenzie *et al.*, the assumptions of this model are as follows: 1) The system is demographically closed to changes in the occupancy status of site during the sampling period. At the species level, this means that a species cannot colonize/immigrate to a site, or go locally extinct/emigrate from that site during the course of the study. 2) Species are not falsely detected. 3) Detection at a site is independent of detection at other sites. This means that your sites should be far enough apart to be biologically independent. The multi-season occupancy model handles violations to demographic closure, and the misidentification model handles violations to false identification; both are discussed later in the book.

SINGLE-SEASON DESIGN SPREADSHEET SETUP

If you haven't already done so, click on the sheet labeled "Single-Season Design" and we'll get started. At the top of the sheet, you'll see a section labeled Inputs and Outputs:

	G	H	I	J	K	L	M	N	O	P
2	**Inputs**			**Outputs**						
3	S	J	R	K	Log_eL	$-2Log_eL$	$-2Log_eL$ (sat)	Deviance	c-hat	AIC
4	50	3	8	2	-50.14720	100.29440	-41.72562	-142.02003	-23.67000	104.29440

Don't worry about the output in row 4 at the moment; we'll go through these cells in a while. First, let's look at cells G4:I4 (the section labeled "Inputs"). Be aware, however, that even though these cells are labeled "inputs," you won't actually enter anything in these cells! Cell G4 contains the total number of sites, or S. Cell H4 contains the maximum number of surveys per site, or J. In this spreadsheet exercise, J can range from 2 to 5. Note that cells G4:H4 are connected to cells Y5 and Z5, which are located in the "Simulate Data" section…we'll cover this later too. Cell I4 is also computed, and is calculated as 2^J. This is the number of possible unique histories for the dataset, and we call this value R. If there are 2 possible

surveys, then there are 2^2 = 4 kinds of unique histories. If there are 3 possible surveys, then there are 2^3 = 8 kinds of unique histories. If there are 4 possible surveys, then there are 2^4 = 16 kinds of unique histories. If there are 5 possible surveys, then there are 2^5 = 32 kinds of unique histories. Now, let's skip down and look at the histories themselves.

	A	B	C	D	E	F	G
7	History						
8	1	2	3	4	5	History	Frequency
9	1	1	1	1	1	11111	0
10	1	1	1	1	0	11110	0
11	1	1	1	0	1	11101	0
12	1	1	1	0	0	11100	0
13	1	1	0	1	1	11011	0
14	1	1	0	1	0	11010	0
15	1	1	0	0	1	11001	0
16	1	1	0	0	0	11000	0
17	1	0	1	1	1	10111	0
18	1	0	1	1	0	10110	0
19	1	0	1	0	1	10101	0
20	1	0	1	0	0	10100	0
21	1	0	0	1	1	10011	0
22	1	0	0	1	0	10010	0
23	1	0	0	0	1	10001	0
24	1	0	0	0	0	10000	0
25	0	1	1	1	1	01111	0
26	0	1	1	1	0	01110	0
27	0	1	1	0	1	01101	0
28	0	1	1	0	0	01100	0
29	0	1	0	1	1	01011	0
30	0	1	0	1	0	01010	0
31	0	1	0	0	1	01001	0
32	0	1	0	0	0	01000	0
33	0	0	1	1	1	00111	0
34	0	0	1	1	0	00110	0
35	0	0	1	0	1	00101	0
36	0	0	1	0	0	00100	0
37	0	0	0	1	1	00011	0
38	0	0	0	1	0	00010	0
39	0	0	0	0	1	00001	0
40	0	0	0	0	0	00000	0
41	1	1	1	1		1111	0
42	1	1	1	0		1110	0
43	1	1	0	1		1101	0
44	1	1	0	0		1100	0
45	1	0	1	1		1011	0
46	1	0	1	0		1010	0
47	1	0	0	1		1001	0
48	1	0	0	0		1000	0
49	0	1	1	1		0111	0
50	0	1	1	0		0110	0
51	0	1	0	1		0101	0
52	0	1	0	0		0100	0
53	0	0	1	1		0011	0
54	0	0	1	0		0010	0
55	0	0	0	1		0001	0
56	0	0	0	0		0000	0
57	1	1	1			111	3.43
58	1	1	0			110	1.47
59	1	0	1			101	1.47
60	1	0	0			100	0.63
61	0	1	1			011	1.47
62	0	1	0			010	0.63
63	0	0	1			001	0.63
64	0	0	0			000	40.27
65	1	1				11	0
66	1	0				10	0
67	0	1				01	0
68	0	0				00	0

SPREADSHEET ENCOUNTER HISTORIES

Now take a look at cells A8:G68. In row 8, you'll see the numbers 1 through 5, which label the survey number. Columns A:E contain a binary 0 or 1, indicating whether the species was detected on a survey or not. Column F gives the final history, and column G gives the frequency of each history. Notice that the spreadsheet is currently tiered so that you can evaluate various kinds of designs, from J = 2 to J = 5. Cells F9:G40 (shaded yellow) given the encounter histories and frequencies for a study in which J = 5. Cells F41:G56 (shaded blue) given the encounter histories and frequencies for a study in which J = 4. Cells F57:G64 (shaded orange) given the encounter histories and

frequencies for a study in which J = 3, and cells F65:F68 (shaded green) given the encounter histories and frequencies for a study in which J = 2. As you can see, the current sheet shows encounter histories for a study in which J = 3. That is, 50 sites were surveyed with the standard single-season design, and the frequency of each of the possible encounter histories for J = 3 are shown. The 0's in the other tiers indicate that the encounter histories for J = 2, J = 4, or J = 5 are not possible. Now, you might be wondering, "How can 3.43 sites have a 111 history?" Well, you might have guessed that the encounter history frequencies were generated based on expectation, rather than with stochasticity. Don't let that throw you….we can easily paste encounter histories that are created with stochasticity (and in fact, we will show you how to create "stochastic data" in little while!).

Now take a look at the parameters that can be estimated for the standard single-season occupancy model:

	J	K	L	M	N
8	**Parameter**	**Estimate?**	**Betas**	**MLE**	**SE (MLE)**
9	p1	1		0.50000	
10	p2	0	0	0.50000	
11	p3	0	0	0.50000	
12	p4	0	0	0.50000	
13	p5	0	0	0.50000	
14	ψ	1		0.50000	0.07319
15	p* =			0.96875	

Cells J9:J14 list the possible parameters that can be estimated: p_1, p_2, p_3, p_4, p_5, and ψ. Of course, if J < 5, you will not estimate as many p_i's. As with other spreadsheets, cells K9:K14 are labeled "Estimate?" and you enter a 1 in a cell if the parameter will be estimated and a 0 if it will not be estimated or will be forced to be equal to another parameter. You might remember that there are two basic

models you can run: psi, p(.) and psi p(t). For this exercise, we are going to focus on the p(.) model only. Thus, and this is very important, that you can only estimate one p estimate, and the remaining p estimates must be forced to be equal to the first p estimate (i.e., the p(.) model). For example, below we can see that the beta for p_2, p_3, p_4, and p_5 (cells L10:L13) are forced to equal the beta for p_1 (cell L9) with the equation =L9. This model can be run to analyze data from a study where J = 5. If J = 4, you would clear cell L13, and then force $p_5 = 0$ within Solver itself. If J = 3, you would clear cells L12:L13, and then force $p_4 = 0$ and $p_5 = 0$ within Solver itself. And if J = 2, you would clear cells L11:L13, and then force $p_3 = 0$, $p_4 = 0$, and $p_5 = 0$ within Solver itself.

	J	K	L	M
8	**Parameter**	**Estimate?**	**Betas**	**MLE**
9	p1	1		=(SIN(L9)+1)/2
10	p2	0	=L9	=(SIN(L10)+1)/2
11	p3	0	=L9	=(SIN(L11)+1)/2
12	p4	0	=L9	=(SIN(L12)+1)/2
13	p5	0	=L9	=(SIN(L13)+1)/2
14	ψ	1		=(SIN(L14)+1)/2
15	p* =			=1-((1-M9)*(1-M10)*(1-M11)*(1-M12)*(1-M13))

Thus, ALL of the models we will explore in this exercise will estimate only two parameters: p and ψ. We do this only because the standard error for psi is very easy to compute when p is constant, and is more tedious when p is survey specific (we may add that in at some point, and it is very easy to run the survey-specific models in GENPRES). The betas estimates are converted to probabilities with a sin link. Click on cell M9 and you'll see the formula =(SIN(L9)+1)/2. We used the sin link in the previous exercise, so won't go into a lengthy explanation here.

PROBABILITY OF EACH HISTORY

The history probabilities are computed in cells O9:O68. Rather than entering an equation for each of the histories separately, we entered a single formula in cell

O9 to generate the probability of observing a 11111 history, and then copied it down (with slight modifications) for the other history types. Click on cell O9 and you'll see the formula =((A9*$M$9+NOT(A9)*(1-$M$9))*(B9*$M$10+NOT(B9)*(1-$M$10))*(C9*$M$11+NOT(C9)*(1-$M$11))*(D9*$M$12+NOT(D9)*(1-$M$12))*(E9*$M$13+NOT(E9)*(1-$M$13)))*$M$14. To obtain a 11111 history, the site must have been occupied and was detected on all 5 surveys ($\psi * p_1 * p_2 * p_3 * p_4 * p_5$). But rather than type that out, the formula, we can take advantage of the NOT function so we can copy the formula down for other histories too. We've color-coded the equation above so that you can clearly see 6 terms (representing the pi's and ψ). The first term is =((A9*$M$9+NOT(A9)*(1-$M$9)). This equation multiplies the outcome of survey 1 (in cell A9) by p_1, and then adds the opposite of the outcome of survey NOT(A9) by $(1-p_1)$. For history 11111, the first survey's outcome was a detection (1). So we can rewrite this as $1*p_1 + 0\ (1-p_1) = 1*p_1$. The next 4 terms are similarly constructed, and the final term is the ψ term (cell M14).

This equation makes more sense for histories in which the species was not detected on at least one survey, so let's look at how this equation works for history 11110. Click on cell O10 and you'll see the equation =((A10*$M$9+NOT(A10)*(1-$M$9))*(B10*$M$10+NOT(B10)*(1-$M$10))*(C10*$M$11+NOT(C10)*(1-$M$11))*(D10*$M$12+NOT(D10)*(1-$M$12))*(E10*$M$13+NOT(E10)*(1-$M$13)))*$M$14. The equation results in the following: $(1*p_1+0*(1-p_1)) * (1*p_2+0*(1-p_2)) * (1*p_3+0*(1-p_3)) * (1*p_4+0*(1-p_4)) * (0*p_5+1*(1-p_5))$, which reduces to $(1*p_1) * (1*p_2) * (1*p_3) * (1*p_4) * (1*(1-p_5))$, or simply $p_1*p_2*p_3*p_4*(1-p_5)$, which is how we normally would have entered the history

probability. This method of computing encounter history probabilities is more similar to the matrix methods that are used in MARK and PRESENCE.

The natural log of the history probabilities is computed in cells P9:P68 with the LN function.

THE MULTINOMIAL LOG LIKELIHOOD

Just as in the other models, we need to compute the model multinomial log likelihood.

	J	K	L	M	N	O	P
2	**Outputs**						
3	K	Log_eL	$-2Log_eL$	$-2Log_eL$ (sat)	Deviance	c-hat	AIC
4	2	-50.14720	100.29440	83.45125	16.84316	2.80719	104.29440

So now let's return to the top of the spreadsheet under the section labeled "Outputs." Cell J4 counts the number of parameters that will be estimated by a model. Remember that for a standard model with no covariates and constant detection probability, there can only be two parameters that are estimated: ψ and p(.). The Log_eL is computed in cell K4 with the equation =SUMPRODUCT(G9:G68,P9:P68). Note that this formula combines the encounter histories and history probabilities across all tiers. This won't affect our calculation because, once you've established J, the frequencies of the histories for different J's are 0's, and so those terms essentially drop out of the SUMPRODUCT calculation. The $-2Log_eL$ is computed in cell L4. The $-2Log_eL$ for the saturated model is computed in cell M4 with the equation =-2*SUMPRODUCT(G9:G68,I9:I68). Hopefully you remember how to compute this value – it's based on the raw frequencies and does not consider the occupancy

model parameters. Deviance is computed in cell N4 as the difference between the occupancy model's -2Log$_e$L and the saturated model's -2Log$_e$L. C-hat is computed in cell O4 with the equation =N4/(I4-J4), which is deviance divided by the model degrees of freedom. The model shown is for J = 3. Thus, there are 8 probabilities in the multinomial likelihood, and the occupancy model estimates only 2. Thus, this particular model has 8 - 2 = 6 degrees of freedom, and deviance divided by model degrees of freedom provides an estimate of c-hat (over-dispersion). AIC is computed in cell P4 as -2*Log$_e$L + 2K. Cell N14 is also a key output for this exercise, and we'll cover this output in a moment.

MAXIMIZING THE LOG LIKELIHOOD

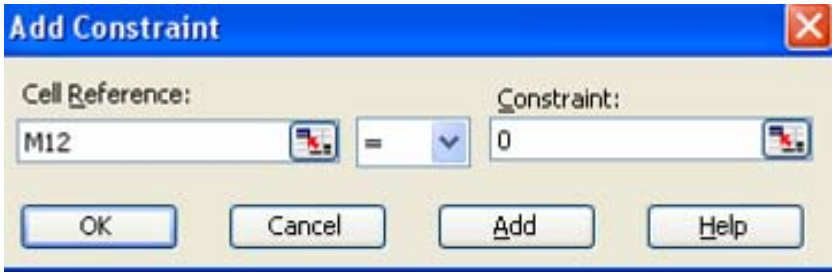
Just as a quick refresher, we'll be analyzing the encounter histories and frequencies for a study in which J = 3 and S = 50 (cells F57:G64). The frequencies should look like this:

	F	G
57	111	3.43
58	110	1.47
59	101	1.47
60	100	0.63
61	011	1.47
62	010	0.63
63	001	0.63
64	000	40.27

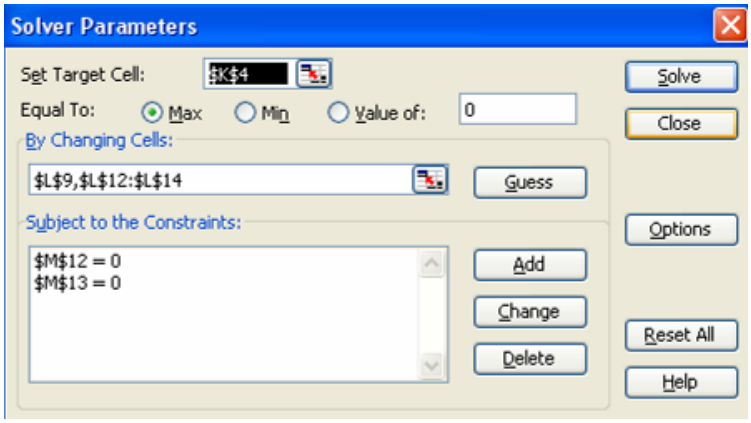
For this model, we will estimate p_1 and ψ, and will force the betas for p_2 and p_3 to be equal to the beta for p_1. We do this by constraining the betas in the usual way. The set-up looks like this:

	J	K	L
8	**Parameter**	**Estimate?**	**Betas**
9	p1	1	
10	p2	0	=L9
11	p3	0	=L9
12	p4	0	
13	p5	0	
14	ψ	1	
15	p* =		

In order to maximize the log likelihood, we will use the Solver tool. Open Solver, and set target cell K4 to a maximum by changing cells L9,L12:L14. Now, although you will let Solver work on the betas for p_4 and p_5, we need to add two constraints within Solver: M12 = 0, and M13 = 0. To add a constraint, just click on the Add button and enter in the constraints:



Press OK, and then add the second constraint. When you're finished, your Solver dialogue box should look like this:



Then press Solve, and keep the Solver solution. Here are our results:

	J	K	L	M	N
8	**Parameter**	**Estimate?**	**Betas**	**MLE**	**SE (MLE)**
9	p1	1	0.411517	0.70000	
10	p2	0	0.411517	0.70000	
11	p3	0	0.411517	0.70000	
12	p4	0	-1.570795	0.00000	
13	p5	0	-1.570795	0.00000	
14	ψ	1	-0.643501	0.20000	0.05777
15	p* =			0.97300	

You can see that Solver estimated beta for p_1 as 0.411517, which corresponds to p = 0.7000, and that Solver estimated beta for ψ as -0.643501, which corresponds to ψ = 0.2. These data happened to be generated by expectation with the following entries, so Solver found the correct solution.

	S	T	U	V	W	X	Y	Z	AA
4	ψ	p1	p2	p3	p4	p5	S	J	Histories
5	0.2	0.7	0.7	0.7	0.7	0.7	50	3	8

VARIANCE OF THE OCCUPANCY ESTIMATE

Now let's discuss two pieces of critical output that we have not yet examined. Cell M15 estimates p*, which is the probability of detecting the species <u>at least once</u> in the surveys; this estimate is critical in determining the variance of ψ. The equation in cell K4 is =1-((1-M9)*(1-M10)*(1-M11)*(1-M12)*(1-M13)), which is 1-[(1-p_1)*(1-p_2)*(1-p_3)*(1-p_4)*(1-p_5)]. If J = 4, then the chance of MISSING the species on all four surveys is (1-p_1) * (1-p_2) * (1-p_3) * (1-p_4). One minus this result is the chance of OBSERVING the species at least once across the four surveys, or p*.

The last cell of the output is cell N14, and is one we're very interested in from a study-design perspective. Variance estimates are very useful in that they will give

us an idea of how precisely ψ was estimated. Variance should decrease with increased surveys and sites. The variance of the occupancy estimate, var(ψ) can be computed for the standard, single-season design model with constant detection probability as follows:

$$\mathrm{var}(\hat{\psi}) = \frac{\psi(1-\psi)}{s} + \frac{\psi(1-p^*)}{Sp^*} + \frac{\psi(1-p^*)Jp(1-p^*)}{Sp^*[p^*(1-p)-Jp(1-p^*)]}$$

where p* = 1-((1-p_1)*(1-p_2)*(1-p_3)...*(1-p_J)), S = total number of sites, J is the maximum number of surveys, and p and ψ are the MLE's from the model (equation 4.8 in the book, Occupancy Estimation and Modeling). In the spreadsheet, variance is computed as (M14*(1-M14))/G4+(M14*(1-M15))/(G4*M15)+(M14*(1-M15)*H4*M9*(1-M15))/(G4*M15*(M15*(1-M9)-H4*M9*(1-M15))). Note that there are three terms (components) in this computation. In the words of MacKenzie et al., "The first component reflects the binomial variance. The second component is the uncertainty in the number of occupied sites, assuming that p is known, and the third component is the contribution of Var(ψ_{MLE}) from having to estimate p from the data simultaneously." The square root of the variance yields the standard error of the estimate. The standard error is calculated in cell N14 with the equation =SQRT((M14*(1-M14))/G4+(M14*(1-M15))/(G4*M15)+(M14*(1-M15)*H4*M9*(1-M15))/(G4*M15*(M15*(1-M9)-H4*M9*(1-M15)))).

Now, let's return to the Output portion of the spreadsheet.

	J	K	L	M	N	O	P
2	**Outputs**						
3	K	Log_eL	$-2Log_eL$	$-2Log_eL$ (sat)	Deviance	c-hat	AIC
4	2	-41.72562	83.45125	83.45125	0.00000	0.00000	87.45125

Because the data were simulated based on expectation, the deviance is 0. This won't happen with stochastically generated data. That's basically all there is in terms of analysis.

The remainder of the exercise will focus on deriving estimates of p, ψ, and SE(ψ) under different scenarios of S and J, and under different scenarios of p and ψ but with fixed S and J. That is, for a given p and ψ for a species, we will attempt to determine the optimal number of sites (S) and optimal number of surveys (J) that will provide precise and unbiased estimates of ψ. And then, we'll determine what kinds of species (in terms of detectability and occupancy) are best studied for a study design where S and J are known *a priori*. The first step is to learn how to simulate data, so we'll do that now.

SIMULATING DATA

On the right side of the spreadsheet you will see the simulated data. At the top are the inputs -- the parameters that we will set in order to test the model under varying conditions.

	S	T	U	V	W	X	Y	Z	AA
2	**SIMULATE DATA**								
3									
4	ψ	p1	p2	p3	p4	p5	S	J	Histories
5	0.2	0.7	0.7	0.7	0.7	0.7	50	3	8

Occupancy (ψ), detection probability (p), number of sites (S), and number of surveys (J) are set by the user (you!). In cell SR5, enter an occupancy rate. The above diagram shows ψ = 0.2, which indicates that the species is fairly rare. If you

wanted to model a common species, you would let ψ be > ~0.7. In cell T5, enter a detection rate. The diagram above shows p = 0.7, so the species is easily detectable. If you wanted to model a species that is elusive, you would let p be < ~0.2. Cells U5:X5 are grayed out, and don't enter anything there. Remember, for this spreadsheet exercise we will only analyze the psi, p(.) model so each of those cells has the equation =T5 in it. You certainly could run the psi, p(t) model, but you'd have to enter a new formulae in cell N14 to compute the standard error of ψ in a survey-specific model. S (cell Y5) is the total number of study sites that are surveyed. This spreadsheet is set up to analyze a maximum of 200 sites, but this can easily be expanded. In cell Z5, enter J, or the maximum number of surveys in the study. J is currently set to 3, but can be as low as 2 or as high as 5 (in this spreadsheet).

Based on these inputs, there are two ways to create encounter history frequencies. The first is by expectation, and the second is with stochasticity. Let's start with the expected values.

	AK	AL	AM	AN	AO	AP	AQ	AR	AS
3		**1**	**2**	**3**	**4**	**5**	**psi**	**S**	**J**
4	p_i	0.7	0.7	0.7	0.7	0.7	0.2	50	3
5	$(1-p_i)$	0.3	0.3	0.3	0.3	0.3			
6									
7							**EXPECTATION**		
8		**History**					**History**	**Probability**	**Frequency**
9		1	1	1	1	1	11111	0.033614	0
10		1	1	1	1	0	11110	0.014406	0
11		1	1	1	0	1	11101	0.014406	0

Scroll over to columns AK:AS. Note that the parameter estimates are completely grayed out here, and that they mirror the entries you made in cells S5:Z5. Don't enter anything in this section at all! As we've seen before, the histories are entered on a survey by survey level in columns AL:AP, and the concatenated history

is given in column AQ. Cell AR9 gives the expected number of sites that should have a 11111 history. The formula in that cell is =$AQ$4*(($AL$4*AL9+$AL$5*NOT(AL9))*($AM$4*AM9+$AM$5*NOT(AM9))*($AN$4*AN9+$AN$5*NOT(AN9))*($AO$4*AO9+$AO$5*NOT(AO9))*($AP$4*AP9+$AP$5*NOT(AP9))), which uses the same NOT trick we used earlier in the exercise. Take some time to make sure you understand this equation. It is copied down for the other histories…the only modification occurs for histories 00000, 0000, 000, and 00, where the term +(1-ψ) is included.

Given the parameters and history probabilities, we now can compute the number of sites that are expected to have each history by multiplying the probability by the total number of sites. Click on cell AS9 and you'll see the equation =IF($AS$4=5,AR9*$AR$4,0). This IF function first evaluates whether cell AS4 = 5. If it does, then each site is surveyed 5 times, and the expected frequency is computed as the probability (AR9) * the total number of sites (cell AR4). If AS4 does NOT equal 5, the spreadsheet returns a 0, indicating that 0 sites have that history. Make sense?

Now, given the parameter values you entered, you can paste the expected frequencies into cells G9:G68 by pressing the button labeled "Paste Data from Expected Values."

The second method for creating encounter history frequencies is with stochasticity. This method occurs in two steps. In the first step, we create encounter histories for each site on a site-by-site basis. In the second step, we sum the site-by-site data to create encounter history frequencies for analysis.

	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE
2		SIMULATE DATA												
3														
4		ψ	p1	p2	p3	p4	p5	S	J	Histories				
5		0.2	0.7	0.7	0.7	0.7	0.7	50	3	8				
6														
7		Random Numbers						Survey					Potential	Actual
8	Site	ψ	p1	p2	p3	p4	p5	1	2	3	4	5	History	History
9	1	0.60528	0.99642	0.88389	0.5354	0.93852	0.92364	0	0	0	0	0	00000	000

Let's start by looking at the site-by-site simulation, focusing on the equations used for site 1 (the formulae for site 1 are simply copied down for the other sites). Cells S8:X9 are simply random numbers, with the equation =RAND(). We'll use these random numbers to generate an encounter history for site 1. Click on cell Y9 and you'll see the formula =IF(AND(S9<$S$5,T9<$T$5),1,0). This formula essentially says, if cell S9 (the random ψ) is less than cell S5 (the specified ψ), AND cell T9 (the random p1) is less than cell T5 (the specified p1), then return a 1 (indicating the species was detected); otherwise return a 0. Cells Z9:AC9 have similar formula, except that the p's are tied to the particular survey of interest. Cell AD9 concatenates the results of all 5 surveys with the equation =Y9&Z9&AA9&AB9&AC9. This cell is listed as the "potential history". The actual history is provided in cell AE9, which has the equation =IF(AND(R9<=$Y$5,$Z$5=2),LEFT(AD9,2),IF(AND(R9<=$Y$5,$Z$5=3),LEFT(AD9,3),IF(AND(R9<=$Y$5,$Z$5=4),LEFT(AD9,4),IF(AND(R9<=$Y$5,$Z$5=5),LEFT(AD9,5),"FALSE")))). This equation features several nested IF functions. The first IF function evaluates whether cell R9 <= Y5 AND cell Z5 = 2. If so, then the site number is less than S (the number of study sites) and J = 2, and the left two surveys in cell AD9 are returned. If either of the conditions mentioned are not true, then the function steps into the next IF function: (AND(R9<=$Y$5,$Z$5=3),LEFT(AD9,3). The second IF function evaluates whether cell R9 <= Y5 AND cell Z5 = 3. If so, then the site number is less than S

(the number of study sites) and J = 3, and the left three surveys in cell AD9 are returned. If either of the conditions mentioned are not true, then the function steps into the next IF function, and so on.

The stochastic data are summarized in cells AH9:AH68. These cells use the COUNTIF function to count the number of each kind of history. Work your way through the equations if you wish. If you press F9, the calculate key, Excel will draw new random numbers, and hence new survey results for each site, and the new results will be summarized in this table. Press F9 5 times and you will have simulated 5 datasets for the ψ, p, J, and S specified in the Inputs section. Press F9 100 times and you will have simulated 100 datasets for the ψ, p, J, and S specified. If you wish to paste in the data for analysis, simply press the button labeled "Paste Stochastically Generated Data."

EXERCISE 1. MAXIMIZING J AND S FOR A SPECIES WHERE Ψ AND P ARE KNOWN

OK, now that you know how to simulate data, we'll put those cells to work. In this first exercise, we will explore the optimal single-season occupancy design for a species in which ψ = 0.5 and in which p = 0.5. This species is fairly common (occurring on half the sites surveyed) but isn't easily detected (if a site is occupied, there is only a 50% chance of detecting the species). We've selected this hypothetical species arbitrarily, but you could plug in different values for ψ and p to suite your own study organism.

Now, given that ψ = 0.5 and p = 0.5, what is the optimal survey design? Well, before you start, it's a good idea to specify the level of precision you wish to

achieve in your parameter estimates. This could be something like "95% confidence limits of an estimate are within +/- 0.1 of the estimate" or something like "standard errors of an estimate are less than 0.05." Before you start your study, you should have some basic idea of what level of precision will be acceptable. Most of us want high precision in the estimates, but these come at a cost in terms of J and S.

Let's assume that, *a priori*, you know that you can sample up to 200 study sites in a single season, and that you can repeatedly sample each site up to 5 times (J = 5). The question is: for the species of interest, is this the appropriate study design, or could you achieve the level of precision that you desire with fewer sites or fewer repeat visits?

To answer this question, we'll run our occupancy model under varying conditions of S and J. Let's let S range from 25 to 200 in increments of 25, and let's let J range from 2 to 5. For each combination of S and J, we'll simulate data, analyze the data with Solver, and store the final estimates of ψ and SE(ψ). We'll be filling in the following table, which is located on the sheet labeled "Exercise 1":

	A	B	C	D	E
1	**Exercise 1: Maximizing J and S for a species where ψ = 0.5 and p = 0.5**				
2					
3					
4					
5	Simulation	S	J	ψ hat	SE (ψ)
6	1	25	2		
7	2	50	2		
8	3	75	2		
9	4	100	2		
10	5	125	2		
11	6	150	2		
12	7	175	2		
13	8	200	2		
14	9	25	3		
15	10	50	3		
16	11	75	3		
17	12	100	3		
18	13	125	3		
19	14	150	3		
20	15	175	3		
21	16	200	3		
22	17	25	4		
23	18	50	4		
24	19	75	4		
25	20	100	4		
26	21	125	4		
27	22	150	4		
28	23	175	4		
29	24	200	4		
30	25	25	5		
31	26	50	5		
32	27	75	5		
33	28	100	5		
34	29	125	5		
35	30	150	5		
36	31	175	5		
37	32	200	5		

In our first analysis, S = 25 and J = 2. So our inputs should look like this:

	S	T	U	V	W	X	Y	Z	AA
2	**SIMULATE DATA**								
3									
4	ψ	p1	p2	p3	p4	p5	S	J	Histories
5	0.5	0.5	0.5	0.5	0.5	0.5	25	2	4

Given those inputs, we can either evaluate data created by expectation, or we can evaluate data created with stochasticity.....your choice. To keep things simple, we're going to evaluate the data created by expectation (only because if you analyze stochastic data, you'd want to repeat our little experiment here about 1,000 to get an idea of the range of the potential results…feel free to do this later if you want!).

Once, you've entered the inputs, you'll see the expected frequencies for J = 2 provided in cells AS65:AS68.

	AQ	AR	AS
65	11	0.125	3.125
66	10	0.125	3.125
67	01	0.125	3.125
68	00	0.625	15.625

Just click on the button labeled "Paste Data from Expected Values", and these data will be ready for analysis.

	F	G
65	11	3.125
66	10	3.125
67	01	3.125
68	00	15.625

In this first run, J = 2, so we will be estimating the betas associated with p_1 and ψ, and will constrain the beta for p_2 to equal the beta for p_1. We'll constrain $p_3 = p_4 = p_5 = 0$ within the Solver dialogue box.

	J	K	L
8	**Parameter**	**Estimate?**	**Betas**
9	p1	1	
10	p2	0	=L9
11	p3	0	
12	p4	0	
13	p5	0	
14	ψ	1	

Now you're ready to find the MLE's. Open Solver, and set target cell K4 to a maximum by changing cells L9,L11:L14, but add the constraints that M11 = 0, M12 = 0, and M13 = 0. Press Solve.

You should get the following results:

	J	K	L	M	N
8	**Parameter**	**Estimate?**	**Betas**	**MLE**	**SE (MLE)**
9	p1	1	-2.28E-13	0.50000	
10	p2	0	-2.28E-13	0.50000	
11	p3	0	-1.570795	0.00000	
12	p4	0	-1.570795	0.00000	
13	p5	0	-1.570795	0.00000	
14	ψ	1	-2.38E-13	0.50000	0.17321

Solver found the unbiased estimates for p_1 and ψ, but what we're really interested in is the SE(ψ). So take a look at the estimates shown in cell N14. Ouch! The standard error for ψ is 0.17321. The 95% confidence intervals for the standard error can be computed as 0.50 +/- 1.96*0.17321, or 0.161 - 0.839. Although our

estimate is unbiased, the precision is very low. Now, select cells M14:N14, and paste the values into cells D6:E6 on the sheet "Exercise 1".

	A	B	C	D	E
5	Simulation	S	J	ψ hat	SE (ψ)
6	1	25	2	0.5	0.173205081
7	2	50	2		
8	3	75	2		
9	4	100	2		
10	5	125	2		

OK, one simulation down, 31 to go. Now you're ready to run the next scenario: S = 50, J = 2. Basically, you'd repeat this process until the entire table is filled. This is fairly repetitive work, so we created a macro to do everything for us. If you choose to run the macro, first press the button labeled "Clear Data". Second, open Solver and clear out any constraints that may be in there…they will foul up our macro. Then, press the button labeled "Run Analysis # 1" and the spreadsheet should do the rest.

The final dataset should look like this:

	B	C	D	E
5	S	J	ψ hat	SE (ψ)
6	25	2	0.50000	0.17321
7	50	2	0.50000	0.12247
8	75	2	0.50000	0.10000
9	100	2	0.50000	0.08660
10	125	2	0.50000	0.07746
11	150	2	0.50000	0.07071
12	175	2	0.50000	0.06547
13	200	2	0.50000	0.06124
14	25	3	0.50000	0.12247
15	50	3	0.50000	0.08660
16	75	3	0.50000	0.07071
17	100	3	0.50000	0.06124
18	125	3	0.50000	0.05477
19	150	3	0.50000	0.05000
20	175	3	0.50000	0.04629
21	200	3	0.50000	0.04330
22	25	4	0.50000	0.10871
23	50	4	0.50000	0.07687
24	75	4	0.50000	0.06276
25	100	4	0.50000	0.05436
26	125	4	0.50000	0.04862
27	150	4	0.50000	0.04438
28	175	4	0.50000	0.04109
29	200	4	0.50000	0.03844
30	25	5	0.50000	0.10377
31	50	5	0.50000	0.07338
32	75	5	0.50000	0.05991
33	100	5	0.50000	0.05189
34	125	5	0.50000	0.04641
35	150	5	0.50000	0.04237
36	175	5	0.50000	0.03922
37	200	5	0.50000	0.03669

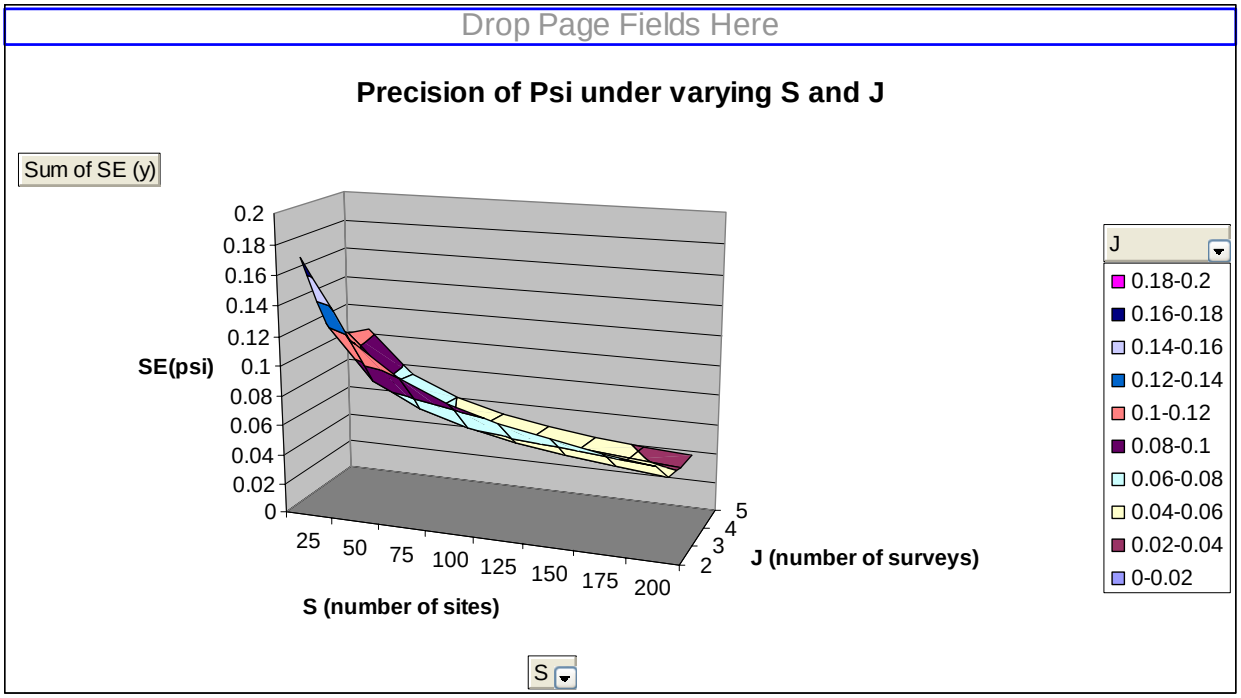
You can see that the estimates of ψ are unbiased for all scenarios. Now you can compare the standard errors for ψ across a variety of J and S scenarios. As expected, the lowest SE occurs when J = 5 and S = 200. But you might be able to live with less precision.

A good way to visualize these results is to display them in a Surface Graph, where the x axis is J, the y axis is S, and the Z axis is the standard error of ψ. The lower the standard error, the better. We can use the Pivot Table option in Excel

organize our data so that we can make this graph. We've taken the liberty of adding the Pivot table directly on the sheet "Exercise 1." (How to create pivot tables is described in detail in Chapter 10). Here are the data:

	G	H	I	J	K	L
3	**PIVOT TABLE OF DATA**					
4						
5	Sum of SE (y)	J				
6	S	2	3	4	5	Grand Total
7	25	0.173205081	0.122474487	0.108711461	0.103774904	0.508165934
8	50	0.122474487	0.08660254	0.076870611	0.073379939	0.359327578
9	75	0.1	0.070710678	0.062764592	0.059914469	0.293389739
10	100	0.08660254	0.061237244	0.054355731	0.051887452	0.254082967
11	125	0.077459667	0.054772256	0.048617243	0.046409548	0.227258714
12	150	0.070710678	0.05	0.044381268	0.042365927	0.207457874
13	175	0.065465367	0.046291005	0.04108907	0.039223227	0.192068669
14	200	0.061237244	0.04330127	0.038435306	0.036689969	0.179663789
15	Grand Total	0.757155064	0.53538948	0.475225283	0.453645436	2.221415263

A pivot table is simply one way to reorganize your data. In this case, S is given in cells G7:G14, and J is given in cells H6:K6. The data are the standard error estimates of psi. We've also taken the liberty of plotting these results as a surface graph:

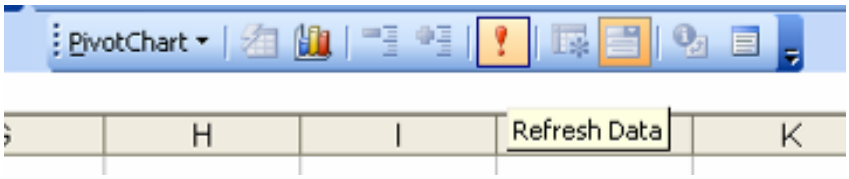


This graph nicely displays the range of standard errors for psi under different levels of both S and p. So, for our species where ψ = 0.5 and where p = 0.5, both S and J are important. Note: if your graph is empty, go to View | Toolbars, and make sure the Pivot Table toolbar is selected. Then, click on your pivot graph and press the exclamation point on the pivot table toolbar to update your graph.

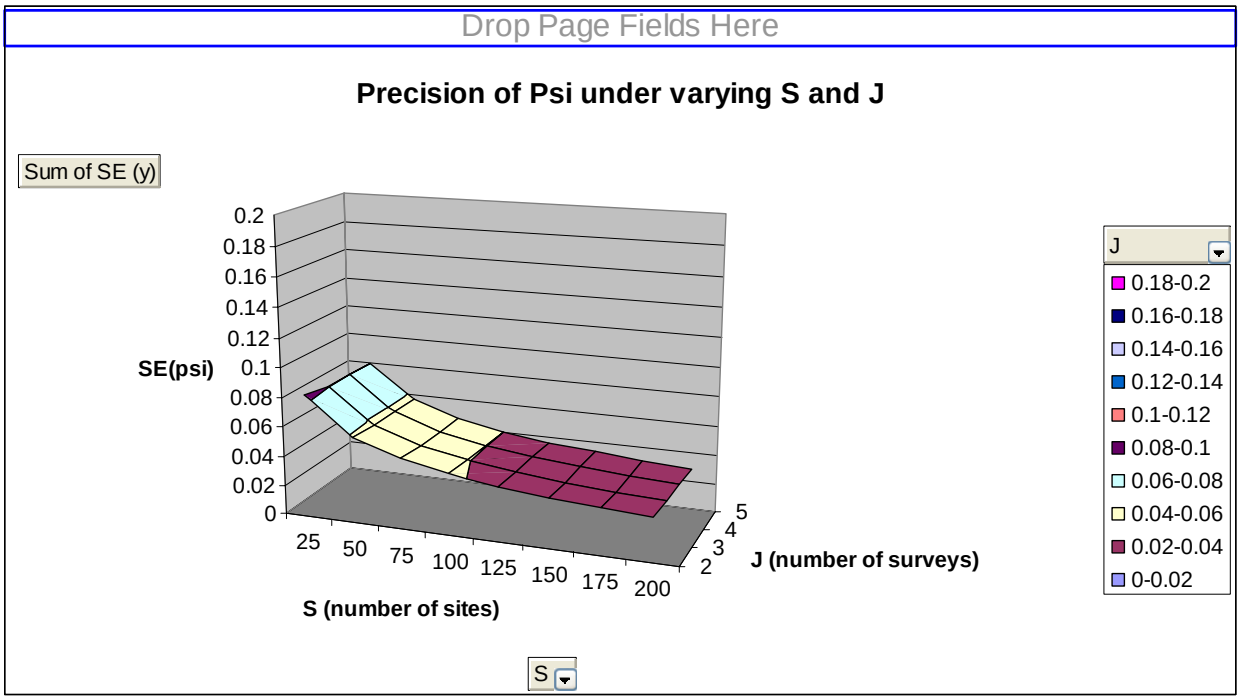
It's very easy to analyze different "kinds" of species. Let's try an analysis where ψ = 0.2 and p = 0.8. Hence, the species is rare, but readily detectable. First, set up your inputs as follows:

	S	T	U	V	W	X
4	ψ	p1	p2	p3	p4	p5
5	0.2	0.8	0.8	0.8	0.8	0.8

Then, click on the sheet labeled "Exercise 1" and clear the data. Then press the button labeled "Run Analysis # 1." Then, click somewhere on the surface chart, and then select the button with the exclamation point on it that is located on the pivot table toolbar:



If for some reason this toolbar is not open, go to View | Toolbars | Pivot Table. When you press the "!" button, the data in the pivot table are automatically refreshed, as is the surface graph. Here are the results for a species where ψ = 0.2 and p = 0.8:

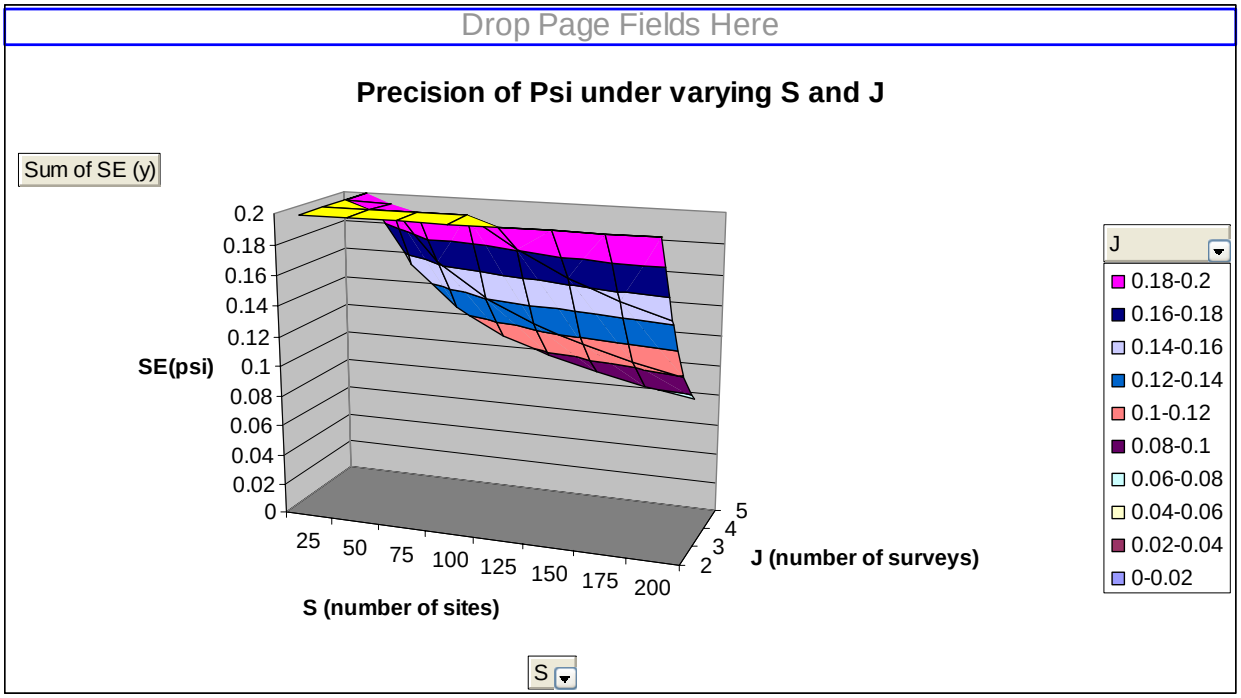


You can see that, if you were designing a study for this species, the precision of ψ is not very sensitive to changes in J (because it is highly detectable), but you certainly would need to consider how many sites you need to study.

Let's try one more. This time, let's consider a common species that is elusive, where ψ = 0.8 and p = 0.2. Set up your inputs as follows:

	S	T	U	V	W	X
4	ψ	p1	p2	p3	p4	p5
5	0.8	0.2	0.2	0.2	0.2	0.2

Then, click on the sheet labeled "Exercise 1" and clear the data. Then press the button labeled "Run Analysis # 1." Then, click somewhere on the surface chart, and then select the button with the exclamation point on it that is located on the pivot table toolbar to update the pivot table data and graph:



As you can see, the precision of ψ is much lower than for the other species, and in this case J is very important. If a species is very hard to detect on a site, you

should carefully examine your data collection protocols and do everything you can to maximize p (the probability of detecting the species, given it is present) when you are in the field.

EXERCISE 2: MAXIMIZING SE(ψ) FOR SPECIES WHERE J AND S ARE KNOWN.

In this second exercise, we will explore the optimal standard occupancy design for a species in which J and S are known. Let's say that, for logistical reasons, you already know the number of sites that can be studied, as well as the maximum number of visits to a site (perhaps you are limited by finances, or maybe you inherited a dataset and now wish to evaluate the strength of this design for different kinds of species). In this exercise, we'll let J = 3 and S = 50. We've selected this hypothetical scenario arbitrarily, but you could plug in different values for J and S to suite your own study organism. Your inputs should look like this:

	Y	Z
4	S	J
5	50	3

Now, given that J = 3 and S = 50, what kinds of species are well-surveyed with this design? Again we will be examining the precision of ψ.

This time, we'll run our model under varying conditions of ψ and p. Let's let ψ range from 0.2 to 0.8 in increments of 0.1, and we'll let p range from 0.2 to 0.8 in increments of 0.1. As before, for each combination of ψ and p, we'll simulate data, analyze the data with Solver, and store the final estimates of ψ and SE(ψ). We'll be filling in the following table that is on the sheet labeled "Exercise 2":

	A	B	C	D	E
6	Simulation	ψ	p	ψ hat	SE (ψ)
7	1	0.2	0.2		
8	2	0.2	0.3		
9	3	0.2	0.4		
10	4	0.2	0.5		
11	5	0.2	0.6		
12	6	0.2	0.7		
13	7	0.2	0.8		
14	8	0.3	0.2		
15	9	0.3	0.3		
16	10	0.3	0.4		
17	11	0.3	0.5		
18	12	0.3	0.6		
19	13	0.3	0.7		
20	14	0.3	0.8		
21	15	0.4	0.2		
22	16	0.4	0.3		
23	17	0.4	0.4		
24	18	0.4	0.5		
25	19	0.4	0.6		
26	20	0.4	0.7		
27	21	0.4	0.8		
28	22	0.5	0.2		
29	23	0.5	0.3		
30	24	0.5	0.4		
31	25	0.5	0.5		
32	26	0.5	0.6		
33	27	0.5	0.7		
34	28	0.5	0.8		
35	29	0.6	0.2		
36	30	0.6	0.3		
37	31	0.6	0.4		
38	32	0.6	0.5		
39	33	0.6	0.6		
40	34	0.6	0.7		
41	35	0.6	0.8		
42	36	0.7	0.2		
43	37	0.7	0.3		
44	38	0.7	0.4		
45	39	0.7	0.5		
46	40	0.7	0.6		
47	41	0.7	0.7		
48	42	0.7	0.8		
49	43	0.8	0.2		
50	44	0.8	0.3		
51	45	0.8	0.4		
52	46	0.8	0.5		
53	47	0.8	0.6		
54	48	0.8	0.7		
55	49	0.8	0.8		

The first scenario is one in which ψ = 0.2 and p = 0.2, so our inputs should look like this:

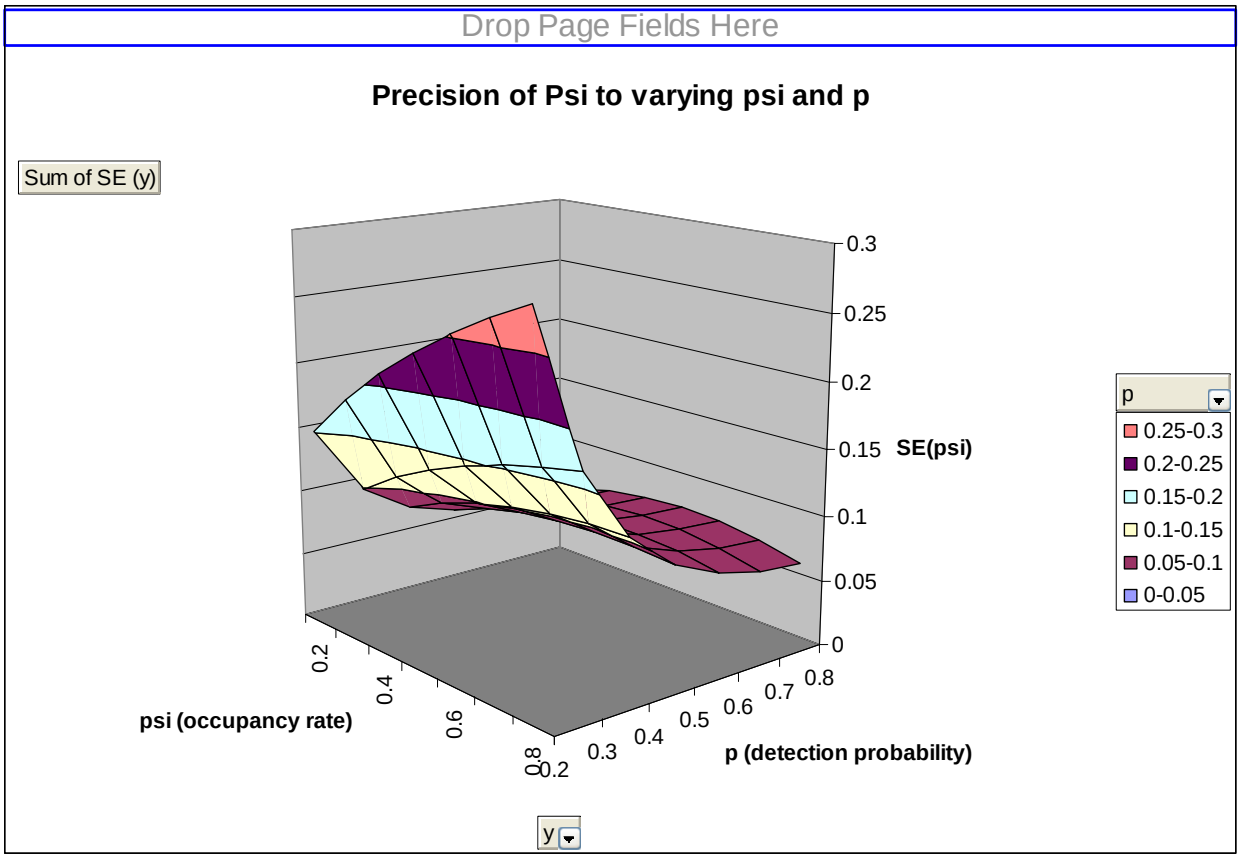
	S	T	U	V	W	X	Y	Z	AA
2	**SIMULATE DATA**								
3									
4	ψ	p1	p2	p3	p4	p5	S	J	Histories
5	0.2	0.2	0.2	0.2	0.2	0.2	50	3	8

Make sure that your spreadsheet inputs match those shown. Once again we'll be analyzing encounter histories from these inputs that are based on expectation (cells AS9:AS68), and these will be pasted into the model for analysis (cells G9:G68).

As before, we recorded a macro to run all 49 simulations for you. Just press the "Clear Data" button to clear out old results, enter data for S and J in cells Y5:Z5, and then press the button labeled "Run Analysis # 2" to run the simulation. Your results should look like this:

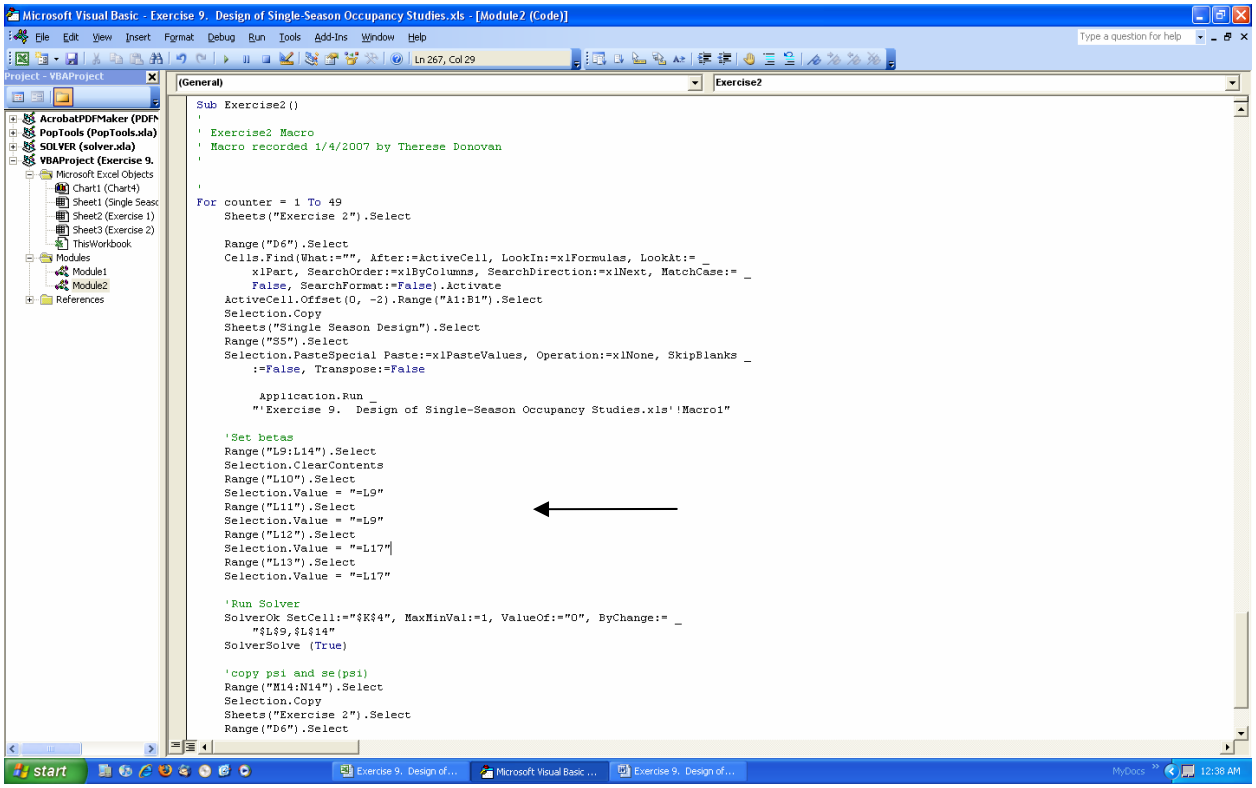
	B	C	D	E
6	ψ	p	ψ hat	SE (ψ)
7	0.2	0.2	0.2000	0.1513
8	0.2	0.3	0.2000	0.0977
9	0.2	0.4	0.2000	0.0752
10	0.2	0.5	0.2000	0.0648
11	0.2	0.6	0.2000	0.0600
12	0.2	0.7	0.2000	0.0578
13	0.2	0.8	0.2000	0.0569
14	0.3	0.2	0.3000	0.1837
15	0.3	0.3	0.3000	0.1172
16	0.3	0.4	0.3000	0.0888
17	0.3	0.5	0.3000	0.0755
18	0.3	0.6	0.3000	0.0692
19	0.3	0.7	0.3000	0.0664
20	0.3	0.8	0.3000	0.0652
21	0.4	0.2	0.4000	0.2102
22	0.4	0.3	0.4000	0.1323
23	0.4	0.4	0.4000	0.0985
24	0.4	0.5	0.4000	0.0825
25	0.4	0.6	0.4000	0.0748
26	0.4	0.7	0.4000	0.0712
27	0.4	0.8	0.4000	0.0698
28	0.5	0.2	0.5000	0.2329
29	0.5	0.3	0.5000	0.1445
30	0.5	0.4	0.5000	0.1055
31	0.5	0.5	0.5000	0.0866
32	0.5	0.6	0.5000	0.0774
33	0.5	0.7	0.5000	0.0731
34	0.5	0.8	0.5000	0.0713
35	0.6	0.2	0.6000	0.2527
36	0.6	0.3	0.6000	0.1545
37	0.6	0.4	0.6000	0.1103
38	0.6	0.5	0.6000	0.0883
39	0.6	0.6	0.6000	0.0774
40	0.6	0.7	0.6000	0.0722
41	0.6	0.8	0.6000	0.0701
42	0.7	0.2	0.7000	0.2704
43	0.7	0.3	0.7000	0.1626
44	0.7	0.4	0.7000	0.1131
45	0.7	0.5	0.7000	0.0877
46	0.7	0.6	0.7000	0.0747
47	0.7	0.7	0.7000	0.0684
48	0.7	0.8	0.7000	0.0658
49	0.8	0.2	0.8000	0.2863
50	0.8	0.3	0.8000	0.1691
51	0.8	0.4	0.8000	0.1141
52	0.8	0.5	0.8000	0.0849
53	0.8	0.6	0.8000	0.0691
54	0.8	0.7	0.8000	0.0612
55	0.8	0.8	0.8000	0.0578

We created a pivot table from these data and created a surface chart from the pivot data. Here are the results in graphical form:



These results may seem a bit counter-intuitive. If you can only sample 50 sites three times each, then you only want to survey species that have high detection probability. Common species that are elusive will yield high standard errors.

Now, running a new scenario will not be as straight-forward as it was in exercise 1. You can certainly change S while keeping J = 3 and re-run the analysis. But to change J, you'll need to go into the Visual Basic for Applications code and make some modifications. Go to Tools | Macros | Macro | Exercise2, and then click the Edit button. Look for the section labeled "Set Betas":



The model we just ran, where J = 2, had the following code entered:

```
Range("L10").Select
Selection.Value = "=L9"
Range("L11").Select
Selection.Value = "=L9"
Range("L12").Select
Selection.Value = "=L17"
Range("L13").Select
Selection.Value = "=L17"
```

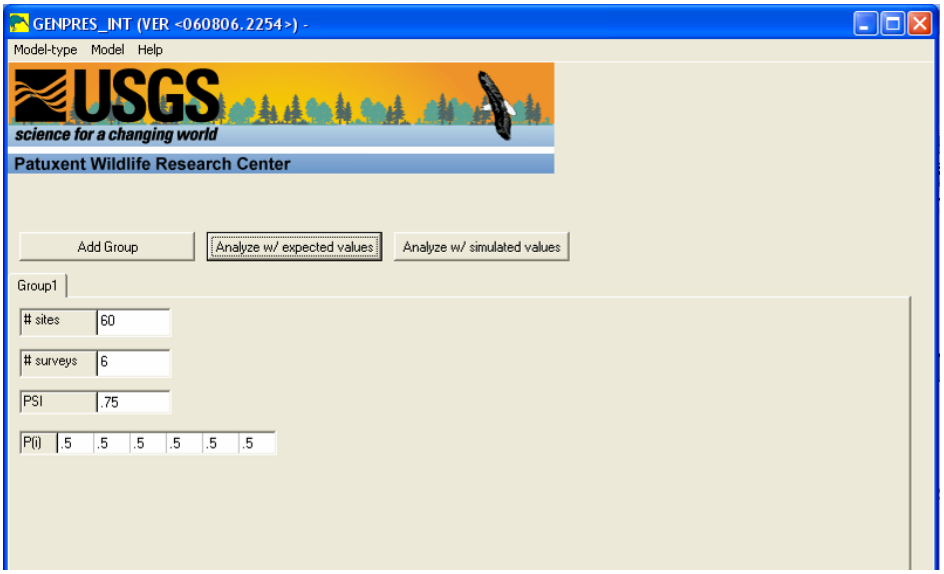
This section forces the beta for p_2 and p_3 in cell L10:L11 to equal the beta for p_1 in cell L9. It also forces the betas for p_4 and p_5 to equal the value in cell L17, which is -1.5707953066856. A beta of -1.5707953066856 corresponds to a parameter estimate of 0 when the sin link is used. In this way, we forced p_4 and p_5 = 0, and also forced $p_1 = p_2 = p_3$. You can use the current set up as long as J = 3 (i.e., change S). But if you want to run simulations where J is not equal to 3, you'll need to adjust the code a bit.

DESIGN OF SINGLE SEASON OCCUPANCY STUDIES IN GENPRES

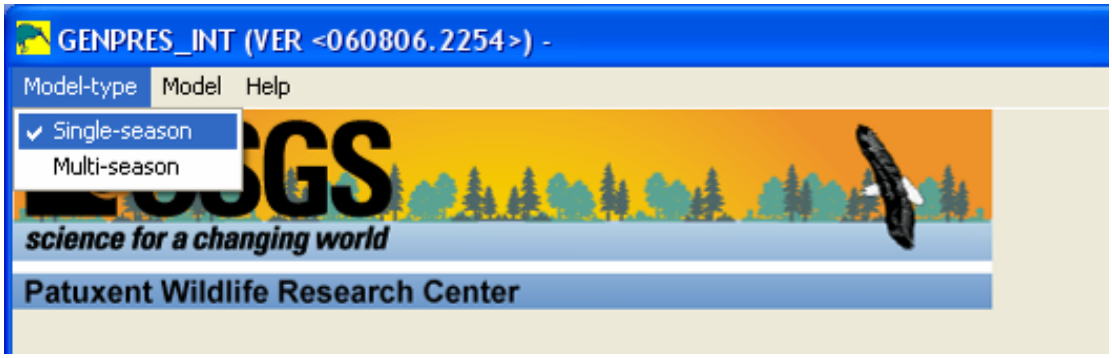
GETTING STARTED

GENPRES is a computer program that was developed to aid researchers in designing optimal occupancy studies. Download the Windows setup program (genpres_setup.exe), and execute it (double-click from windows explorer). Data can be generated with this program, however, to analyze the data, program MARK is required. Program MARK can be downloaded (for free) from http://www.cnr.colostate.edu/~gwhite/mark/mark.htm

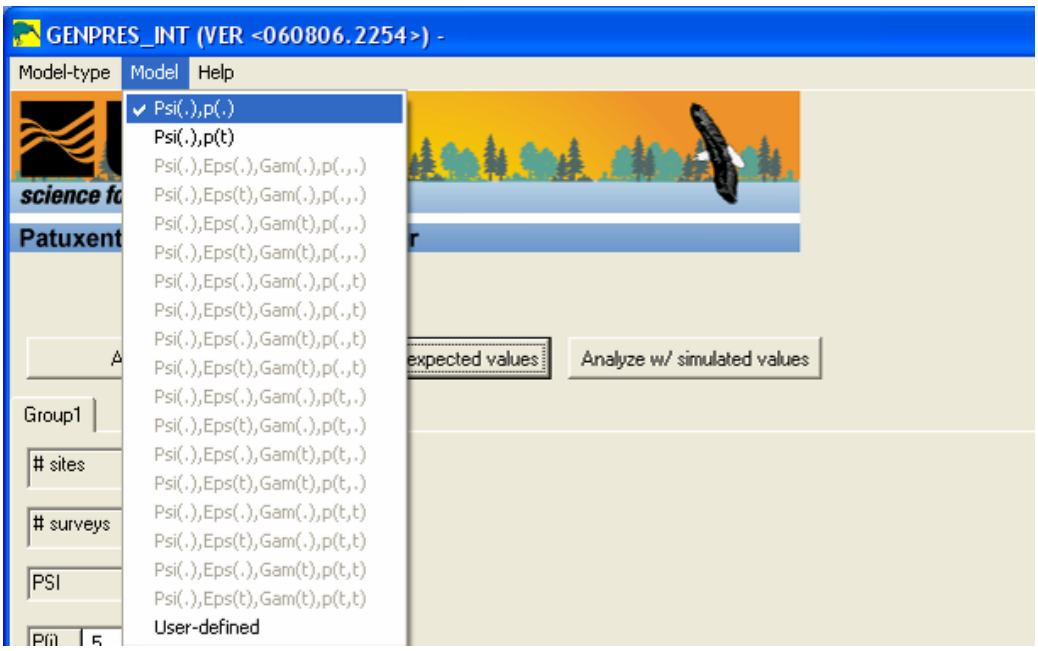
When you open GENPRES, you'll see the following screen:



The program is pretty straight-forward. First, on the toolbar, choose Model-type, and then select the Single-Season Model.



Next, go to the Model toolbar and select model Psi(.)p(.). Remember, this was the model we used in the spreadsheet. But you can see that GENPRES also will evaluate model Psi(.)p(t)….something to keep in mind.



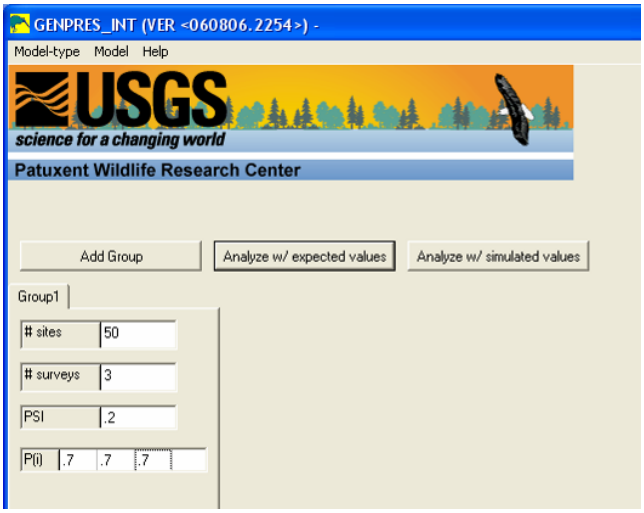
Now, the next thing you need to do is specify the number of sites (S), the number of surveys (J), psi, and p.

RUNNING A SIMULATION

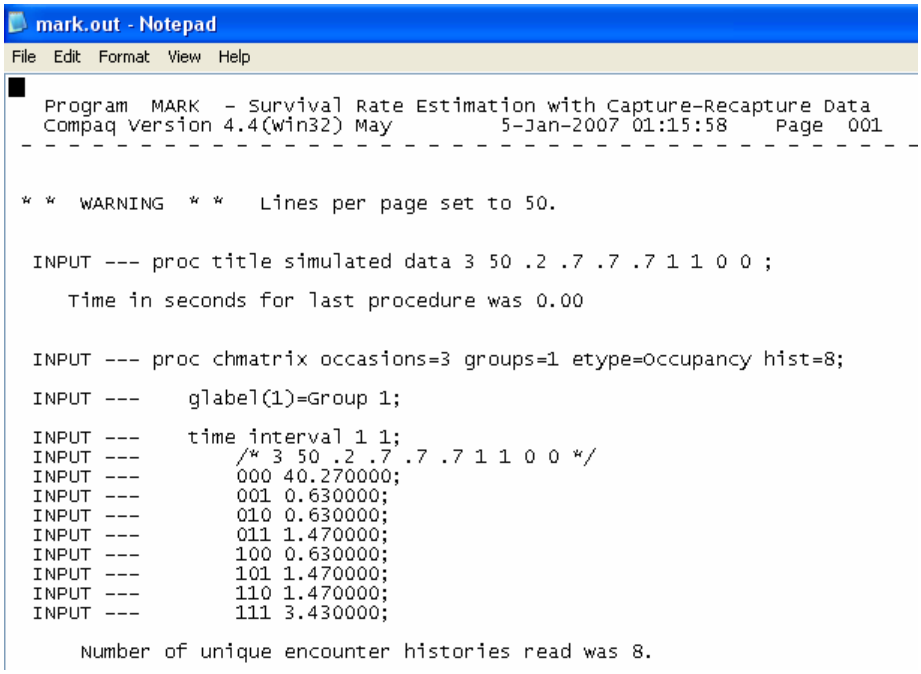
Let's do the very first run we did for the spreadsheet, and then we'll compare outputs.

	S	T	U	V	W	X	Y	Z	AA
4	ψ	p1	p2	p3	p4	p5	S	J	Histories
5	0.2	0.7	0.7	0.7	0.7	0.7	50	3	8

Here are the GENPRES inputs:



Notice that you can run the analysis with either expected values (frequencies) or with simulated values. Since we analyzed frequencies created by expectation, click on the button labeled "Analyze w/ expected values." That's all there is to it! GENPRES then presents the MARK output page. We'll work through the output section by section.



mark.out - Notepad

File Edit Format View Help

```
  Program  MARK  - Survival Rate Estimation with Capture-Recapture Data
  Compaq Version 4.4(Win32) May         5-Jan-2007 01:15:58    Page  001
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

* *  WARNING  * *   Lines per page set to 50.

 INPUT --- proc title simulated data 3 50 .2 .7 .7 .7 1 1 0 0 ;

    Time in seconds for last procedure was 0.00

 INPUT --- proc chmatrix occasions=3 groups=1 etype=Occupancy hist=8;

 INPUT ---    glabel(1)=Group 1;

 INPUT ---    time interval 1 1;
 INPUT ---        /* 3 50 .2 .7 .7 .7 1 1 0 0 */
 INPUT ---        000 40.270000;
 INPUT ---        001 0.630000;
 INPUT ---        010 0.630000;
 INPUT ---        011 1.470000;
 INPUT ---        100 0.630000;
 INPUT ---        101 1.470000;
 INPUT ---        110 1.470000;
 INPUT ---        111 3.430000;

     Number of unique encounter histories read was 8.
```

Towards the top of the output, the encounter histories and their expected frequencies are provided. These correspond to spreadsheet cells F57:G64.

	F	G
57	111	3.43
58	110	1.47
59	101	1.47
60	100	0.63
61	011	1.47
62	010	0.63
63	001	0.63
64	000	40.27

So far, so good. Now scroll down a bit more, and you'll see more key outputs. Notice that the default link function is the sin link (which is why we used this link in the spreadsheet). Below that, information about the model's $-2\text{Log}_e\text{L}$, the saturated model's $-2\text{Log}_e\text{L}$, AIC, etc. is provided.

mark.out - Notepad

File Edit Format View Help

```
 INPUT ---    rlabel(1)=Psi;

 INPUT ---    blabel(2)=p;

 INPUT ---    rlabel(2)=p;

Link Function Used is SIN

Variance Estimation Procedure Used is 2ndPart
-2logL(saturated) = 0.0000000
Effective Sample Size = 50

* *  WARNING  * *   Error number 2 from VA09AD optimization routine.

Number of function evaluations was 8 for 2 parameters.
Time for numerical optimization was 0.01 seconds.
-2logL {Psi(.),p(.)} = 83.451248
Penalty {Psi(.),p(.)} = 0.0000000
Gradient {Psi(.),p(.)}:
  0.000000      0.000000
S Vector {Psi(.),p(.)}:
  48.71161      23.98022
Time to compute number of parameters was 0.01 seconds.
  Threshold =  0.6000000E-07     Condition index =  0.4922897
Conditioned S Vector {Psi(.),p(.)}:
  1.000000     0.4922897
Number of Estimated Parameters {Psi(.),p(.)} = 2
DEVIANCE {Psi(.),p(.)} = 83.451248
DEVIANCE Degrees of Freedom {Psi(.),p(.)} = 6
c-hat {Psi(.),p(.)} = 13.908541
AIC {Psi(.),p(.)} = 87.451248
AICc {Psi(.),p(.)} = 87.706567
Pearson Chisquare {Psi(.),p(.)} = 0.1000260E-11
```

Here there are a few discrepancies between the program and the spreadsheet. Although the -2Log$_e$L and AIC values match the spreadsheet, MARK is reporting the -2LogeL for the saturated model = 0, whereas it was 83.45 in the spreadsheet. Because of this, the deviance and c-hat calculations also don't match. These really don't affect this exercise, so let's push on.

	G	H	I	J	K	L	M	N	O	P
2	**Inputs**			**Outputs**						
3	S	J	R	K	Log$_e$L	-2Log$_e$L	-2Log$_e$L (sat)	Deviance	c-hat	AIC
4	50	3	8	2	-41.72562	83.45125	83.45125	0.00000	0.00000	87.45125

Scroll down a bit further, and you'll see the beta estimates, real parameter estimates, and the standard errors:

SIN Link Function Parameters of {Psi(.),p(.)}

Parameter	Beta	Standard Error	95% Confidence Interval Lower	Upper
1:Psi	-0.6435011	0.1444333	-0.9265903	-0.3604119
2:p	0.4115168	0.2033938	0.0128649	0.8101687

Real Function Parameters of {Psi(.),p(.)}

Parameter	Estimate	Standard Error	95% Confidence Interval Lower	Upper
1:Psi	0.2000000	0.0577733	0.1096797	0.3365802
2:p	0.7000000	0.0932068	0.4943423	0.8477714

Here you see the betas and real estimates match those from the spreadsheet. The standard error provided for these data (0.577733) matches the result in cell N14.

	J	K	L	M	N
8	**Parameter**	**Estimate?**	**Betas**	**MLE**	**SE (MLE)**
9	p1	1	0.411517	0.70000	
10	p2	0	0.411517	0.70000	
11	p3	0	0.411517	0.70000	
12	p4	0	-1.570795	0.00000	
13	p5	0	-1.570795	0.00000	
14	ψ	1	-0.643501	0.20000	0.05777
15	p* =			0.97300	

The latest version of GENPRES asks after each simulation if you would like to delete or append to a spreadsheet (genpres.csv) file. You can click 'Yes' the first time, then click 'No' on the other runs to generate a table of results.