EXERCISE 10: SINGLE SEASON REMOVAL DESIGN

In collaboration with Sarah J. Frey

Rubenstein School of Environment and Natural Resources, University of Vermont

TABLE OF CONTENTS

REMOVAL MODEL SPREADSHEET EXERCISE

OBJECTIVES:

- To learn and understand the removal design occupancy model, how it differs from the standard design, and how it fits into a multinomial maximum likelihood analysis.
- To use Solver to find the maximum likelihood estimates for the probability of detection and the probability of site occupancy under the removal design.
- To derive the variance of the occupancy estimate and p^* (overall detection probability on occupied sites).
- To learn how to simulate data for a removal design occupancy model.
- To test the sensitivity of the removal design occupancy model to varying numbers of sites, surveys, p, and ψ (using simulated and expected data).

BACKGROUND: SPECIES OCCURRENCE AND DISTRIBUTION

Understanding factors that shape species occurrence and distribution is a fundamental concept in ecology. Mathematical modeling promotes an analysis of species' range borders and patterns (Holt and Keitt 2005), metapopulation dynamics (Hanski 1998), habitat-species relationships, and population response to environmental change (MacKenzie et al. 2003, MacKenzie et al. 2006). In the past, researchers used presence-absence of species within specified study areas to develop mathematical models that describe the spatial and temporal distribution of species (MacKenzie 2006); detection probability of the target species was often not considered in many of these models. However, as we've seen in previous chapters, analyses that do not take detection probability into account can result in biased outcome measures, usually producing figures that overestimate true

occurrence or relative abundance. The occupancy models described by MacKenzie et al. 2006 provide a robust method for describing distribution patterns across time and space. Once you've concluded that occupancy modeling is suitable for your own research, a central question becomes how many sites and how many surveys should be conducted to provide robust conclusions?

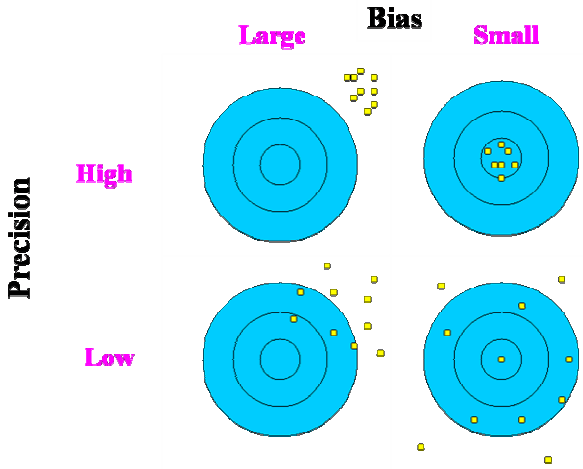
BALANCING SURVEY EFFORT WITH PRECISION

The planning process of biological research involves balancing limited resources (time, funds, and personnel) with the collection of high-quality data. It is often the goal of researchers to do just what it takes to reach a specified level of precision and no more, so that resources can be used to their maximum capabilities and provide information for managing the target species. Whether the goal is to minimize effort in order to obtain a desired level of precision or to minimize uncertainty in the occupancy estimate for a fixed level of effort (MacKenzie and Royle 2005), the researcher must identify the most efficient way to collect field data. In occupancy modeling, the key trade-offs are whether to increase the number of study sites, or whether to increase the number of surveys that are conducted at each site. Given *a priori* knowledge of your study organism, the goal is to determine what level of effort is necessary in the field that will result in good data.

PRECISION AND BIAS IN MODELING

Precision and bias are both very important concepts in the modeling process. An ideal mathematical model will have parameter estimates that are both precise and unbiased. 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 $\psi = 0.7$, 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 $\psi = 0.7$ 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 (they 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 an occupancy "removal design" can be used to provide precise and unbiased parameter estimates while limiting the number of repeat surveys per site.

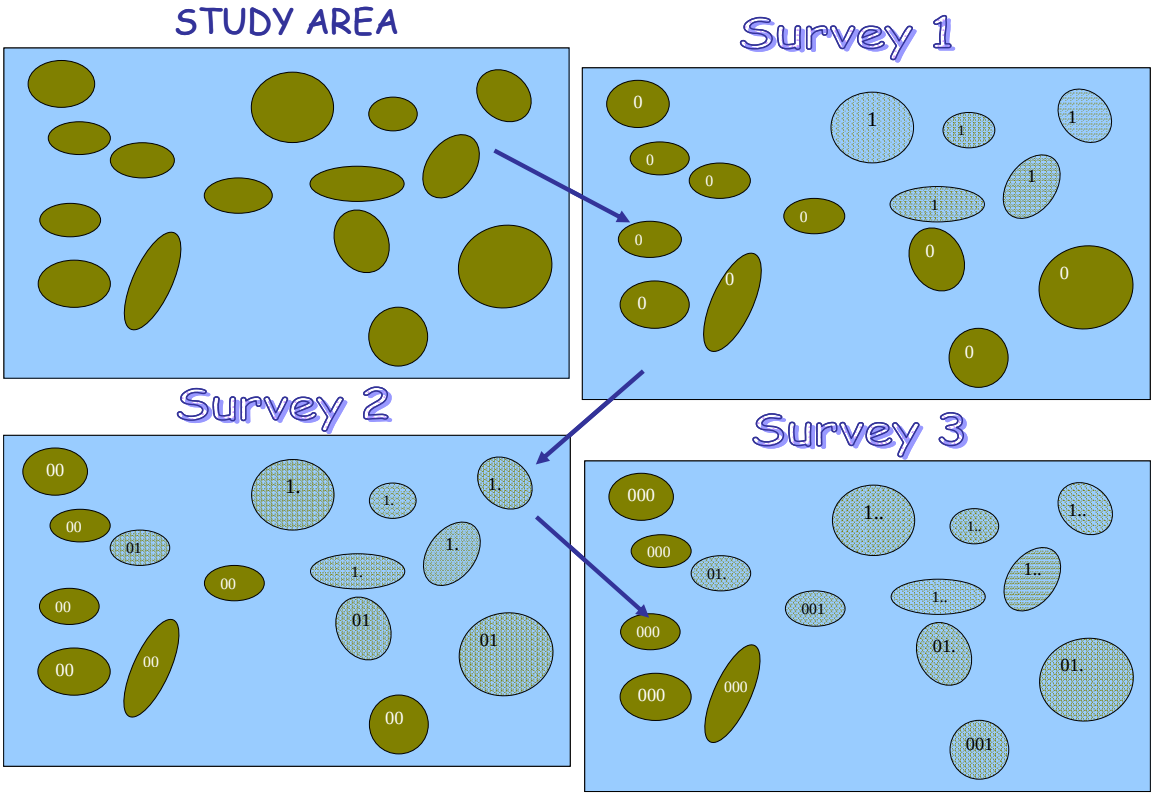
REMOVAL DESIGN BACKGROUND

A complete description of the removal design occupancy model along with comparison to the standard and double sampling designs can be found in MacKenzie and Royle (2005) and in chapter 6 of MacKenzie et al. (2006).

The removal design occupancy model is very similar to the standard design. The main difference is that the number of times a site is visited within a season may vary depending on whether a species was previously detected or not. The basic idea of the removal design is that a site is surveyed a maximum of J times within a season, but once the target species has been detected at a site, that site is no longer visited. If a species is detected at a site, any additional surveys at the site will provide further information about detection probability, but these additional surveys will not provide any additional information about occupancy. Thus, the removal method will allow the researcher to visit more sites in less time by not having to visit all of the sites the maximum number of surveys. Instead of directing effort to re-sampling the same sites regardless of detection, more sites can be added without increasing survey effort while increasing spatial replication (MacKenzie et al. 2006).

The following figure displays how a 3-survey, 16-site removal design could pan out. When a species is detected at a site, that site is removed from future surveys, which is represented by the lighter shade of green. As you can see, the pool of sites drops from 16, to 10, to 7, to 5 as the season goes on. This is a total of 35

(6*1 + 4*2 + 7*3) surveys with the removal design compared to 48 surveys (16*3) if the standard design had been used. This is a 27% decrease in survey effort from the standard design to sample the same number of sites and in less time, although less exhaustively.



MODEL ASSUMPTIONS

Like the standard design occupancy model, the population is assumed to be closed between surveys. Additionally, the species detection is assumed to be less than 1 and the target species is not falsely detected. Sites are independent in regards to the detection of the target species. An assumption specific to the removal design is that detection probabilities are assumed to be constant across surveys. This may be limiting in the analysis phase if it is suspected that detection probability is

related to survey-specific factors. Thus, the underlying model for detection must be a p(.) or p(covariate) model, such as p(date, habitat type, patch size). You cannot model p1, p2, p3, or p4 independently. In other words, the intercepts and slope effects of covariates must be equal for all survey periods. We'll come back to this important concept a bit later.

ENCOUNTER HISTORIES

The encounter histories are very similar to those of the standard design, except that after the target species is detected, the remaining surveys are not conducted. Thus, for a two-survey study (J = 2), there are three kinds of encounter histories: 1., or 01, or 00. A "1." history indicates that the species was detected on the first survey, and the site was then "removed" from further sampling. The "dot" after the 1 indicates that the site was not surveyed on survey 2. A "01" history indicates that the species was not detected on the first survey, but was detected on the second survey. A "00" history indicates that the species was not detected on either survey. For a three-survey study (J = 3), there are four kinds of encounter histories: 1.., 01., 001, and 000. In a four-survey study (J = 4), there are 5 possible histories: 1..., 01.., 001., 0001, and 0000. Remember, after the species is detected, the site is no longer surveyed. Thus, in removal design, the number of possible, unique histories is J + 1.

REMOVAL DESIGN SPREADSHEET OVERVIEW

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

	D	E	F	G	H	I	J	K	L	M
2	**Inputs**			**Outputs**						
3	S	J	R	K	Log_eL	AIC	p	p*	ψ	var(ψ)
4	50	4	5	2	-58.10508	120.21017	0.30000	0.7599	0.50000	0.163839

Don't worry about the output in row 4 for the moment; we'll go through these cells in a while. First, let's look at cells D4:F4 (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 D4 contains the total number of sites, or S. Cell E4 contains the maximum number of surveys per site, or J. In this spreadsheet exercise, J can range from 2 to 5. Note that cells D4:E4 are connected to cells X5 and Y5, which are located in the "Simulate Data" section…we'll cover this later too. Cell F4 is also computed, and is calculated as J+1. If there are 5 possible surveys, then there are 5 + 1 = 6 kinds of unique histories: 1...., 01..., 001.., 0001., 00001, and 00000. Now, let's skip down and look at the histories themselves.

SPREADSHEET ENCOUNTER HISTORIES

	E	F
6	Histories	Frequency
7	1.	0.00
8	01	0.00
9	00	0.00
10	1..	0.00
11	01.	0.00
12	001	0.00
13	000	0.00
14	1...	7.50
15	01..	5.25
16	001.	3.68
17	0001	2.57
18	0000	31.00
19	1....	0.00
20	01...	0.00
21	001..	0.00
22	0001.	0.00
23	00001	0.00
24	00000	0.00

Now take a look at cells E7:F24. You'll see that the spreadsheet is currently tiered so that you can evaluate various kinds of designs, from J = 2 to J = 5. Cells E7:F9 (shaded orange) given the encounter histories and frequencies for a study in which J = 2. For J = 2, there are three kinds of histories: 1., 01, and 00. Cells E10:F13 (shaded yellow) given the encounter histories and frequencies for a study in which J = 3. For J = 3, there are three kinds of histories: 1.., 01., 001, and 000. Cells E14:F18 (shaded green) given the encounter histories and frequencies for a study in which J = 4, and cells

E19:F24 (shaded blue) given the encounter histories and frequencies for a study in which J = 5. As you can see, the current sheet shows encounter histories for a study in which J = 4. That is, 50 sites were surveyed with the removal design, and the frequency of each of the possible encounter histories for J = 4 are shown. The 0's in the other tiers indicate that the encounter histories for J = 2, J = 3, or J = 5 are not possible. Now, you might be wondering, "How can 7.5 sites have a 1… 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 J = 2 to J = 5 (cells G7:G24):

	E	F	G	H	I
6	Histories	Frequency	Parameters	Estimated?	Betas
7	1.	0.00	p_1		
8	01	0.00	p_2		
9	00	0.00	ψ		
10	1..	0.00	p_1		
11	01.	0.00	p_2		
12	001	0.00	p_3		
13	000	0.00	ψ		
14	1...	7.50	p_1	1	-0.41152
15	01..	5.25	p_2	0	-0.41152
16	001.	3.68	p_3	0	-0.41152
17	0001	2.57	p_4	0	-0.41152
18	0000	31.00	ψ	1	0.00000
19	1....	0.00	p_1		
20	01...	0.00	p_2		
21	001..	0.00	p_3		
22	0001.	0.00	p_4		
23	00001	0.00	p_5		
24	00000	0.00	ψ		

You can see that when J = 2 (two surveys maximum), you can estimate 3 parameters: p_1, p_2, and ψ (cells G7:G9). And when J = 3, you can estimate 4 parameters: p_1, p_2, p_3, and ψ (cells G10:G13). And when J = 4, you can estimate 5 parameters: p_1, p_2, p_3, p_4, and ψ (cells G14:G18), etc.

As with other spreadsheets, cells H7:H24 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. As you can see, you only need to enter 1's and 0's for the tier that describes the data (in this case shown below, J = 4). Also note, 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 the spreadsheet is set up to

estimate p_1 for a study in which J = 4, and the beta for p_2, p_3, and p_4 (cells I15:I17) are forced to equal the beta for p_1 (cell I14) with the equation =I14.

	E	F	G	H	I
14	1...	7.5	p_1	1	
15	01..	5.25	p_2	0	=I14
16	001.	3.675	p_3	0	=I14
17	0001	2.5725	p_4	0	=I14
18	0000	31.0025	ψ	1	

Thus, ALL of the removal models will estimate only two parameters: p and ψ (as long as there are no covariates associated with these two parameters, that is). The beta estimates are converted to probabilities with a sin link. Click on cell J7 and you'll see the formula =(SIN(I7)+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 using the 2 model parameters (p and ψ) in a way that reflects the standard design. The only difference is that where there is a dot "." in the encounter history, no parameters are estimated. Let's start with the history "1." for J=2 surveys (cell K7).

	D	E	F	G	H	I	J	K	L
6		Histories	Frequency	Parameters	Estimated?	Betas	MLE	P(history)	Ln Probability
7		1.	0.00	p_1			0.5000	0.2500	-1.38629
8		01	0.00	p_2			0.5000	0.1250	-2.07944
9	0	00	0.00	ψ			0.5000	0.6250	-0.47000
10		1..	0.00	p_1			0.5000	0.2500	-1.38629
11		01.	0.00	p_2			0.5000	0.1250	-2.07944
12		001	0.00	p_3			0.5000	0.0625	-2.77259
13	0	000	0.00	ψ			0.5000	0.5625	-0.57536
14		1...	7.50	p_1	1		0.5000	0.2500	-1.38629
15		01..	5.25	p_2	0	0.00000	0.5000	0.1250	-2.07944
16		001.	3.68	p_3	0	0.00000	0.5000	0.0625	-2.77259
17		0001	2.57	p_4	0	0.00000	0.5000	0.0313	-3.46574
18	50	0000	31.00	ψ	1		0.5000	0.5313	-0.63252
19		1....	0.00	p_1			0.5000	0.2500	-1.38629
20		01...	0.00	p_2			0.5000	0.1250	-2.07944
21		001..	0.00	p_3			0.5000	0.0625	-2.77259
22		0001.	0.00	p_4			0.5000	0.0313	-3.46574
23		00001	0.00	p_5			0.5000	0.0156	-4.15888
24	0	00000	0.00	ψ			0.5000	0.5156	-0.66238

Since the species was detected on the first survey, you know that the site was occupied. The probability of getting this history given the data = $\psi*p_1$, or =J9*J7. The next history, "01" (cell K8) states that the site is occupied, it was not detected on the first survey, but it was detected in the second survey. The probability of getting this survey = $\psi*(1-p_1)*p_2$, or =J9*(1-J7)*J8. Following along to history "00" (cell K9), the probability of getting this history = $\psi*(1-p_1)*(1-p_2)+(1-\psi)$, or =J9*(1-J7)*(1-J8) + (1-J9). Note that the sum of the probabilities for any given tier must sum to be 1.

Now let's look at the histories for a study in which J = 3. Cell K10 has the equation, =J13*J10, which is the probability of observing a "1.." history and is simply $\psi*p1$. Cell K11 has the equation, =J13*(1-J10)*J11, which is the probability of observing a "01." history and is simply $\psi*(1-p_1)*p_2$. Cell K12 has the equation, =J13*(1-J10)*(1-J11)*J12, which is the probability of observing a "001" history and is simply $\psi*(1-p_1)*(1-p_2)*p_3$. The last history, 000 is a tad trickier because you have to account for the chance that the site was occupied, but you did not detect

the species of interest. Cell K13 has the equation =J13*(1-J10)*(1-J11)*(1-J12)+(1-J13), which is = ψ*(1-p_1)*(1-p_2)*(1-p_3) + (1-ψ) which translates to the site was occupied and you did not detect it on any of the three surveys OR the site was truly unoccupied.

The history probabilities for J = 4 and J = 5 are constructed in a similar manner. Click on cells K14:K24 to make sure you understand how the probabilities are computed. The natural log of the history probabilities is computed in cells L7:L24 with the LN function.

THE MULTINOMIAL LOG LIKELIHOOD

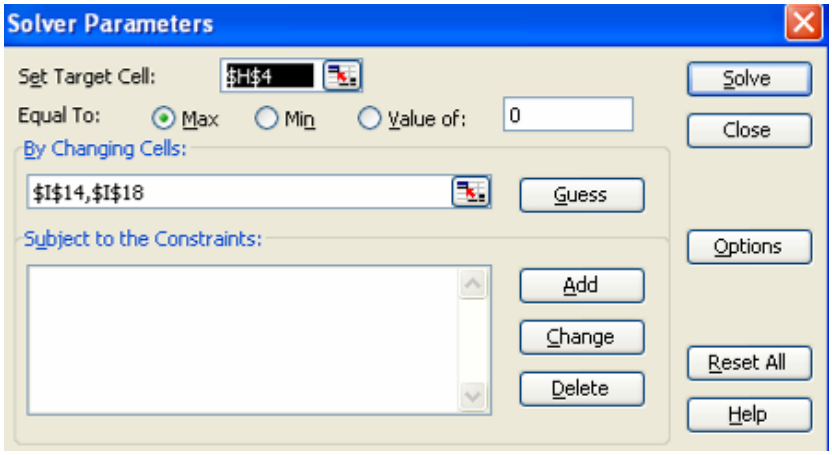
Just as in the other models, we need to compute the model multinomial log likelihood.

	D	E	F	G	H	I	J	K	L	M
2	**Inputs**			**Outputs**						
3	S	J	R	K	Log_eL	AIC	p	p*	ψ	SE (ψ)
4	50	4	5	2	-58.10508	120.21017	0.30000	0.7599	0.50000	0.163839

So now let's return to the top of the spreadsheet under the section labeled "Outputs." Cell G4 counts the number of parameters (K) that will be estimated by a model. Remember that for a removal model with no covariates, there can only be two parameters that are estimated: ψ and p(.). The Log_eL is computed in cell H4 with the equation =SUMPRODUCT(F7:F24,L7:L24). 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. AIC is computed in cell I4 as -2*Log_eL + 2K. Cells J4:M4 are the key outputs for the model, and we'll revisit these in a moment.

MAXIMIZING THE LOG LIKELIHOOD

Let's continue with our example where J = 4. In order to maximize the log likelihood, we will use the Solver tool. Open Solver, and set target cell H4 to a maximum by changing cells I14,I18. Don't forget that cells I15:I17 MUST be set to equal cell I14 to enforce the p(.) model.



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

	E	F	G	H	I	J	K	L
6	Histories	Frequency	Parameters	Estimated?	Betas	MLE	P(history)	Ln Probability
7	1.	0.00	p_1			0.5000	0.2500	-1.38629
8	01	0.00	p_2			0.5000	0.1250	-2.07944
9	00	0.00	ψ			0.5000	0.6250	-0.47000
10	1..	0.00	p_1			0.5000	0.2500	-1.38629
11	01.	0.00	p_2			0.5000	0.1250	-2.07944
12	001	0.00	p_3			0.5000	0.0625	-2.77259
13	000	0.00	ψ			0.5000	0.5625	-0.57536
14	1...	7.50	p_1	1	-0.41152	0.3000	0.1500	-1.89712
15	01..	5.25	p_2	0	-0.41152	0.3000	0.1050	-2.25379
16	001.	3.68	p_3	0	-0.41152	0.3000	0.0735	-2.61047
17	0001	2.57	p_4	0	-0.41152	0.3000	0.0515	-2.96714
18	0000	31.00	ψ	1	0.00000	0.5000	0.6200	-0.47796
19	1....	0.00	p_1			0.5000	0.2500	-1.38629
20	01...	0.00	p_2			0.5000	0.1250	-2.07944
21	001..	0.00	p_3			0.5000	0.0625	-2.77259
22	0001.	0.00	p_4			0.5000	0.0313	-3.46574
23	00001	0.00	p_5			0.5000	0.0156	-4.15888
24	00000	0.00	ψ			0.5000	0.5156	-0.66238

You can see that Solver estimated beta for p_1 as -0.41152, which corresponds to p = 0.3000, and that Solver estimated beta for ψ as 0.000, which corresponds to ψ = 0.5. These data happened to be generated by expectation with the following entries, so Solver found the correct solution.

	R	S	T	U	V	W	X	Y
1	**SIMULATE DATA**							
2								
3	**Inputs**							
4	**ψ**	**p1**	**p2**	**p3**	**p4**	**p5**	**J**	**S**
5	0.5	0.3	0.3	0.3	0.3	0.3	4	50

Now, let's think about this removal design in terms of number of sites that were visited. For the data shown above, 7.5 sites were visited once, 5.25 sites were visited twice, 3.68 sites were visited three times, and 2.57 + 31 sites were visited 4 times. This makes a total of 163 visits using the removal design. If we used the standard design, where all sites were visited 4 times, there would be a total of 200 visits. Assuming you can make a maximum of 200 visits to sites in one season, you

could use those 200-163 = 37 extra visits to survey new sites, increasing your spatial replication. The trade-off, however, is that detection probability, p, must be constant across surveys (i.e., they cannot be survey-specific).

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

	D	E	F	G	H	I	J	K	L	M
2	**Inputs**			**Outputs**						
3	S	J	R	K	Log_eL	AIC	p	p*	ψ	SE (ψ)
4	50	4	5	2	-58.10508	120.21017	0.30000	0.7599	0.50000	0.163839

Cell J4 provides the estimate of p from the given model. The equation in cell J4 is =IF(E4=2,J7,IF(E4=3,J10,IF(E4=4,J14,p_1))), which is a nested IF function that steps through the various tiers and returns the appropriate p. The function starts out by evaluating if cell E4 = 2. If this is true, then J = 2 and the spreadsheet returns the p in cell J7. If this is not true, then the spreadsheet moves to the next IF function. This next IF function starts out by evaluating if cell E4 = 3. If this is true, then J = 3 and the spreadsheet returns the p in cell J10. If it is not true, then the spreadsheet moves to the next IF function: IF(E4=4,J14,p_1). The function starts out by evaluating if cell E4 = 4. If this is true, then J = 4 and the spreadsheet returns the p in cell J14. If this is not true, then J must equal 5 and the spreadsheet returns the p in cell J19 (note that this cell is named p_1 in the spreadsheet.).

Cell K4 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-J4)^E4), which is $1-(1-p)^J$. If J = 4, then the chance of MISSING the species on all four surveys is $(1-p)^4$, or (1-p) * (1-p) * (1-p) * (1-p).

One minus this result is the chance of OBSERVING the species at least once across the four surveys, or p*.

Cell L4 provides the estimate of ψ for the model, and has the equation =IF(E4=2,J9,IF(E4=3,J13,IF(E4=4,J18,psi))). Work your way through this equation, which is similar to the equation in cell J4.

VARIANCE OF THE OCCUPANCY ESTIMATE

The last cell of the output is cell M4, 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 removal design model as follows:

$$\mathrm{var}(\hat{\psi}) = \frac{\psi}{s}\left[(1-\psi) + \frac{p^*(1-p^*)}{(p^*)^2 - J^2 p^2 (1-p)^{J-1}}\right]$$

Where p* = 1-((1-p_1)*(1-p_2)*(1-p_3)...*(1-p_J)), S = total number of sites, J = the maximum number of surveys, and p and ψ are the MLE's from the model (equation 6.5 in the book, Occupancy Estimation and Modeling). In the spreadsheet, variance is computed as L4/D4*((1-L4)+(K4*(1-K4))/(K4^2-E4^2*J4^2*(1-J4)^(E4-1))). Note that there are two terms (components) in this computation. The first component deals with the binomial variation of the estimate, and the second component is related to the parameter uncertainty due to imperfect detection (p<1.0). The square root of the variance yields the standard error of the estimate.

The standard error is calculated in cell M4 with the equation =SQRT(L4/D4*((1-L4)+(K4*(1-K4))/(K4^2-E4^2*J4^2*(1-J4)^(E4-1)))).

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.

	R	S	T	U	V	W	X	Y
1	**SIMULATE DATA**							
2								
3	**Inputs**							
4	**ψ**	**p1**	**p2**	**p3**	**p4**	**p5**	**J**	**S**
5	0.5	0.3	0.3	0.3	0.3	0.3	4	50

Occupancy (ψ), detection probability (p), number of sites (S), and number of surveys (J) are set by the user (you!). In cell R5, enter an occupancy rate. The above diagram shows ψ = 0.5, which indicates that the species is fairly common. If

you wanted to model a rare species, you would let ψ be < ~0.2. In cell S5, enter a detection rate. The diagram above shows p = 0.3, so the species is very elusive. If you wanted to model a species that is easily detected, you would let p be > ~0.7. Cells T5:W5 are grayed out, and don't enter anything there. Remember, in a removal model the assumption is that detection probability is constant across surveys, so each of those cells has the equation =S5 in it. In cell X4, enter J, or the maximum number of surveys in the study. J is currently set to 4, but can be as low as 2 or as high as 5 (in this spreadsheet). 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.

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.

	W	X	Y
12	**Expected Data**		
13	Histories	Frequency	Sum
14	1.	0.00	
15	01	0.00	
16	00	0.00	0
17	1..	0.00	
18	01.	0.00	
19	001	0.00	
20	000	0.00	0
21	1...	7.50	
22	01..	5.25	
23	001.	3.68	
24	0001	2.57	
25	0000	31.00	50
26	1....	0.00	
27	01...	0.00	
28	001..	0.00	
29	0001.	0.00	
30	00001	0.00	
31	00000	0.00	0

Cell X14 gives the expected number of sites that should have a "1." History given that J = 2. The equation in that cell is =IF(X5=2,R5*S5*Y5,0). This simple IF function evaluates if cell X5 = 2. If X5 = 2, then J = 2 and the formula returns R5*S5*Y5, which is ψ*p1*S. If X5 does not equal 2, then the formula returns a 0. All of the other cells (cells X15:X31) have very similar equations. Let's review just one more. Click on cell X25 and you'll see the equation =IF(X5=4,R5*(1-S5)*(1-T5)*(1-U5)*(1-V5)*Y5+(1-R5)*Y5,0). This

equation provides the number of sites that are expected to have a 0000 history, given R = 5. The equation first evaluates if cell X5 = 4. If so, it returns R5*(1-S5)*(1-T5)*(1-U5)*(1-V5)*Y5+(1-R5)*Y5, which is $\psi*(1-p_1)*(1-p_2)*(1-p_3)*(1-p_4)*S+(1-\psi)*S$. If cell X5 does not equal 4, the formula returns a 0. You've worked with expected frequencies in previous exercises. Note that the sum of the frequencies is provided for each level of J (cell Y16 for J = 2; cell Y20 for J = 3, cell Y25 for J = 4; cell Y31 for J = 5). You can enter values for ψ, p_1, J and S in the inputs section, and then click the button labeled "Paste Expected Data" -- cells X14:X31 will be pasted into cells F7:F24 for analysis.

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.

	Q	R	S	T	U	V	W	X	Y	Z
34	Site	Surveyed?	Occupied?	Survey 1	Survey 2	Survey 3	Survey 4	Survey 5	Potential History	Actual History
35	**1**	1	0	0	0	0	0	0	00000	0000;
36	**2**	1	1	1	-	-	-	-	1----	1---;
37	**3**	1	0	0	0	0	0	0	00000	0000;
38	**4**	1	1	0	0	0	1	-	0001-	0001;
39	**5**	1	1	0	0	0	0	1	00001	0000;

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). Click on cell R35 and you'll see the formula =IF(Q35<=Y5,1,0). This equation evaluates whether the site number listed in cell Q35 is less than or equal to S provided in cell Y5. If so, the formula returns a 1 indicating that the site was surveyed, and if not, the formula returns a 0 indicating that the site was not surveyed. Now click on cell S35 and you'll see the equation =IF(RAND()<R5, 1,

0). This equation determines whether the site is occupied or not by our species of interest. If a random number is less than the ψ specified in cell R5, the site is occupied and a 1 is returned, otherwise the site is unoccupied and a 0 is returned. Notice that a site can be occupied, but not surveyed.

OK, now we know which sites are occupied and which are unoccupied, and we know which sites were surveyed and which weren't. Now we need to determine the outcome of each survey, and we'll do this for J = 5. Click on cell T35 and you'll see the formula =IF(AND(S35=1,RAND()<S5),1,0). This is an IF function, with a nested AND function in it. This function basically says, "If the site was occupied (S35=1) AND if a random number is less than p specified in cell S5, then the species was detected and a 1 is returned; otherwise the species was not detected and a 0 is returned. The formula for surveys 2 through 5 are roughly the same, but add a little "removal" twist. Click on cell U35 and you'll see the equation =IF(AND(S35=1,T35=1),"-",IF(AND(S35=1,RAND()<T5),1,0)). This function contains two IF functions. The first IF function evaluates if S35 is 1 (the site is occupied) AND the species was detected in the first survey (T35 = 1). If so, then a "-" is returned, indicating that the second survey was skipped because the animal was detected in survey 1 (per the removal design). If both of these conditions are not true, the formula moves to the next IF function. This function evaluates whether the site is occupied (S35=1) AND if a random number is less than the specified p_2 in cell T5. If BOTH of these conditions are true, the formula returns a 1 and the species was detected on survey 2; otherwise the species was not detected and a 0 is returned. Work your way though the equations in cells V35:X35.

Now click on cell Y35 and you'll see the equation =IF(R35=1,T35&U35&V35&W35&X35,"FALSE"). This cell gives the "potential history" for the site – it is the history for J = 5 but the last surveys will be lopped off when J < 5. The equation evaluates if cell R35 is 1. If so, the site was surveyed and the equation returns the concatenation of cells T35 & U35 & V35 & W35 & X35. If the site was not surveyed, the formula returns the word "FALSE". In this way, the potential histories are provided only for sites 1 through S.

Cell Z35 gives the actual history for site 1, depending on what J is. The equation is =IF(Y35="FALSE","FALSE",IF(X5=5,Y35&";",IF(X5=4,LEFT(Y35,4)&";",IF(X5=3,LEFT(Y35,3)&";",LEFT(Y35,2)&";")))). It's a bit long but is again some nested IF functions, which evaluate what J is in cell Y35 and then returns only the survey results for survey 1 to survey J.

	R	S	T
12	**Stochastic Data**		
13	Histories	Frequency	Sum
14	1.	0	
15	01	0	
16	00	0	0
17	1..	0	
18	01.	0	
19	001	0	
20	000	0	0
21	1...	8	
22	01..	9	
23	001.	4	
24	0001	2	
25	0000	27	50
26	1....	0	
27	01...	0	
28	001..	0	
29	0001.	0	
30	00001	0	
31	00000	0	0

The stochastic data are summarized in cells R14:T31. These cells use the SUMIF 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.

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 removal 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 surveyed (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 removal model under varying conditions of S and J. Let's let J range from 2 to 5, and let's let S range from 25 to 200 in

increments of 25. For each combination of J and S, we'll simulate data, analyze the data with Solver, and store the final estimates of p*, ψ, and SE(ψ). We'll be filling in the following table:

	C	D	E	F	G	H
32	**Exercise 1: Maximizing J and S for a species where y = 0.5 and p = 0.5**					
33	J	S	p*	ψ	SE (ψ)	
34	2	25				
35	2	50				
36	2	75				
37	2	100				
38	2	125				
39	2	150				
40	2	175				
41	2	200				
42	3	25				
43	3	50				
44	3	75				
45	3	100				
46	3	125				
47	3	150				
48	3	175				
49	3	200				
50	4	25				
51	4	50				
52	4	75				
53	4	100				
54	4	125				
55	4	150				
56	4	175				
57	4	200				
58	5	25				
59	5	50				
60	5	75				
61	5	100				
62	5	125				
63	5	150				
64	5	175				
65	5	200				

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

	R	S	T	U	V	W	X	Y
3	**Inputs**							
4	**ψ**	**p1**	**p2**	**p3**	**p4**	**p5**	**J**	**S**
5	0.5	0.5	0.5	0.5	0.5	0.5	2	25

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!).

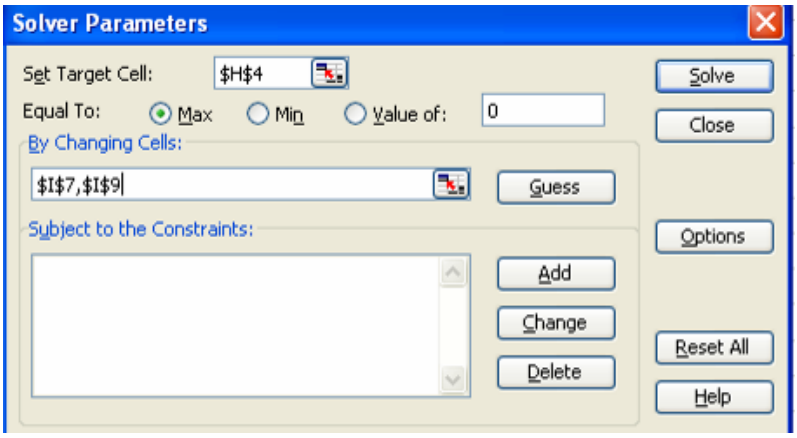
Once, you've entered the inputs, you'll see the expected frequencies provided in cells X14:X31.

	W	X	Y
12	**Expected Data**		
13	Histories	Frequency	Sum
14	1.	6.25	
15	01	3.13	
16	00	15.63	25
17	1..	0.00	
18	01.	0.00	
19	001	0.00	
20	000	0.00	0
21	1...	0.00	
22	01..	0.00	
23	001.	0.00	
24	0001	0.00	
25	0000	0.00	0
26	1....	0.00	
27	01...	0.00	
28	001..	0.00	
29	0001.	0.00	
30	00001	0.00	
31	00000	0.00	0

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

	E	F	G	H	I
6	Histories	Frequency	Parameters	Estimated?	Betas
7	1.	6.25	p_1	1	
8	01	3.125	p_2		=I7
9	00	15.625	ψ	1	
10	1..	0	p_1		
11	01.	0	p_2		
12	001	0	p_3		
13	000	0	ψ		
14	1...	0	p_1		
15	01..	0	p_2		
16	001.	0	p_3		
17	0001	0	p_4		
18	0000	0	ψ		
19	1....	0	p_1		
20	01...	0	p_2		
21	001..	0	p_3		
22	0001.	0	p_4		
23	00001	0	p_5		
24	00000	0	ψ		

In this first run, J = 2, so we will be working with the orange cells and will be estimating the betas for p_1 and ψ that are associated with J = 2 (cells I7,I9). Don't forget that the beta for p_2 must be set to equal the beta for p_1. Now you're ready to find the MLE's. Open Solver, and set target cell H4 to a maximum by changing cells I7,I9. Press Solve.



You should get the following results:

	E	F	G	H	I	J	K	L
6	Histories	Frequency	Parameters	Estimated?	Betas	MLE	P(history)	Ln Probability
7	1.	6.25	p_1	1	0.00000	0.5000	0.2500	-1.38629
8	01	3.13	p_2		0.00000	0.5000	0.1250	-2.07944
9	00	15.63	ψ	1	0.00000	0.5000	0.6250	-0.47000

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 the output section:

	G	H	I	J	K	L	M
2	**Outputs**						
3	K	Log_eL	AIC	p	p*	ψ	SE (ψ)
4	2	-22.50640	49.01280	0.50000	0.75	0.50000	0.264575

Ouch! The standard error for ψ is 0.264575. The lower 95% confidence interval can be computed as 0.5 – 1.96*0.264575 = 0, and the upper 95% confidence interval can be computed as 0.5 + 1.96*0.264575 =1. Although our estimate is unbiased, the precision is very low – useless in fact. Now, select cells K4:M4, and paste the values only into cells E34:G34.

	C	D	E	F	G	H
32	**Exercise 1: Maximizing J and S for a species where y = 0.5 and p = 0.5**					
33	J	S	p*	ψ	SE (ψ)	
34	2	25	0.75	0.5	0.26457513	
35	2	50				

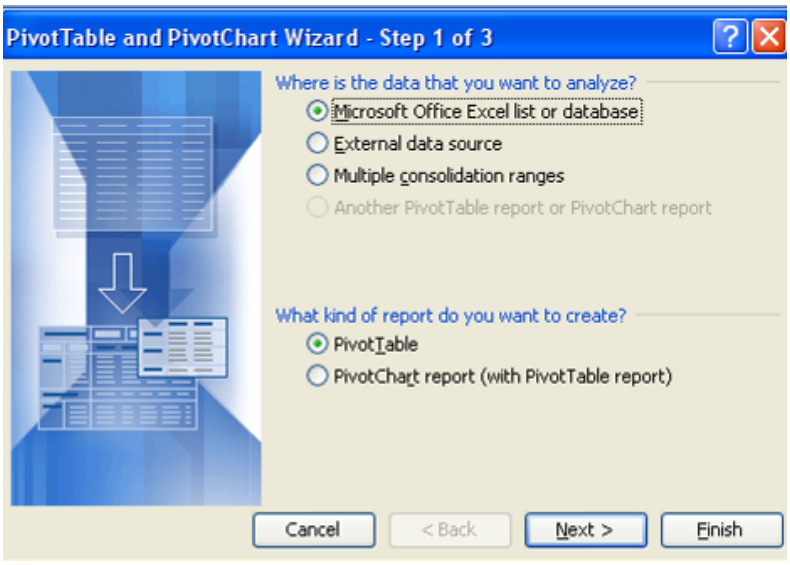
OK, one simulation down, 31 to go. Now you're ready to run the next scenario: J = 2, S = 50. Basically, you'd repeat this process until the entire table is filled. This is fairly repetitive work, so we created a macro to do everything for us. If you choose to run the macro, first press the button labeled "Clear Data". Then, press the button labeled "Run Analysis # 1" and the spreadsheet should do the rest.

The final dataset should look like this:

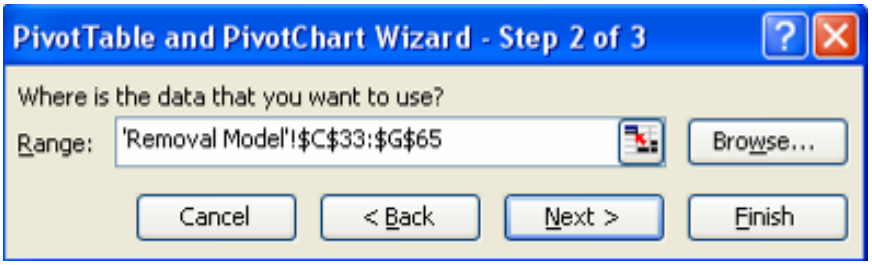
	C	D	E	F	G
32	**Exercise 1: Maximizing J and S for a species where y = 0.5 and p = 0.**				
33	J	S	p*	ψ	SE (ψ)
34	2	25	0.75	0.5	0.26457513
35	2	50	0.75	0.5	0.18708287
36	2	75	0.75	0.5	0.15275252
37	2	100	0.75	0.5	0.13228757
38	2	125	0.75	0.5	0.1183216
39	2	150	0.75	0.5	0.10801234
40	2	175	0.75	0.5	0.1
41	2	200	0.75	0.5	0.09354143
42	3	25	0.875	0.5	0.14411534
43	3	50	0.875	0.5	0.10190493
44	3	75	0.875	0.5	0.08320503
45	3	100	0.875	0.5	0.07205767
46	3	125	0.875	0.5	0.06445034
47	3	150	0.875	0.5	0.05883484
48	3	175	0.875	0.5	0.05447048
49	3	200	0.875	0.5	0.05095247
50	4	25	0.9375	0.5	0.1144237
51	4	50	0.9375	0.5	0.08090978
52	4	75	0.9375	0.5	0.06606255
53	4	100	0.9375	0.5	0.05721185
54	4	125	0.9375	0.5	0.05117184
55	4	150	0.9375	0.5	0.04671328
56	4	175	0.9375	0.5	0.04324809
57	4	200	0.9375	0.5	0.04045489
58	5	25	0.96875	0.5	0.10538107
59	5	50	0.96875	0.5	0.07451567
60	5	75	0.96875	0.5	0.06084179
61	5	100	0.96875	0.5	0.05269053
62	5	125	0.96875	0.5	0.04712785
63	5	150	0.96875	0.5	0.04302164
64	5	175	0.96875	0.5	0.0398303
65	5	200	0.96875	0.5	0.03725783

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.

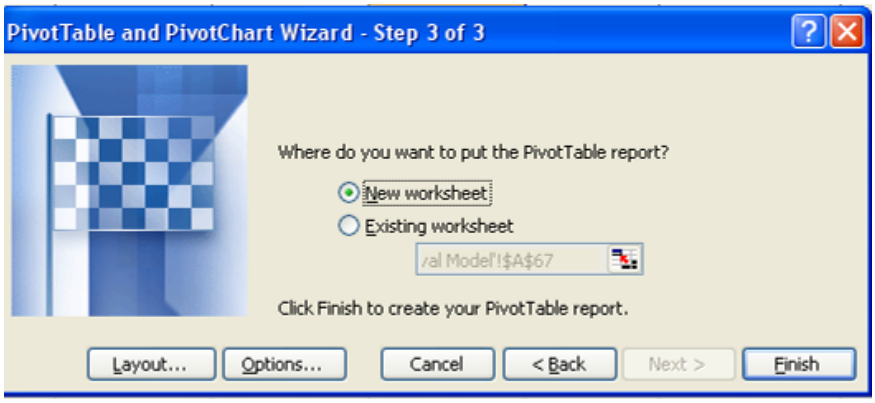
A good way to visualize these results is to display them in a Surface Graph, where the x axis is J, the y axis is S, and the Z axis is the standard error of ψ. The lower the standard error, the better. We can use the Pivot Table option in Excel organize our data so that we can make this graph. The pivot table results are listed on the sheet labeled "Pivot Table Data 1". If you want to re-create this pivot table, go to Data | Pivot Table, and you'll see the following dialogue box:



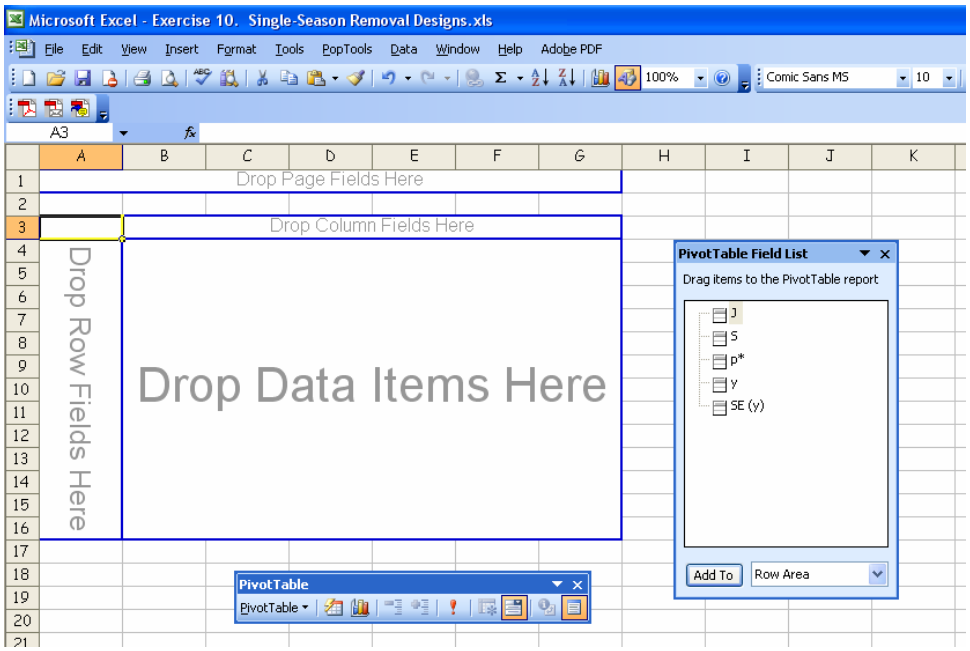
We want to analyze the data on the table we just created, so click the first radio button and also click on the PivotTable radio button. Press Next, and you'll be asked to specify where the data are located. Use your mouse to highlight cells C33:G65, and then press Next.



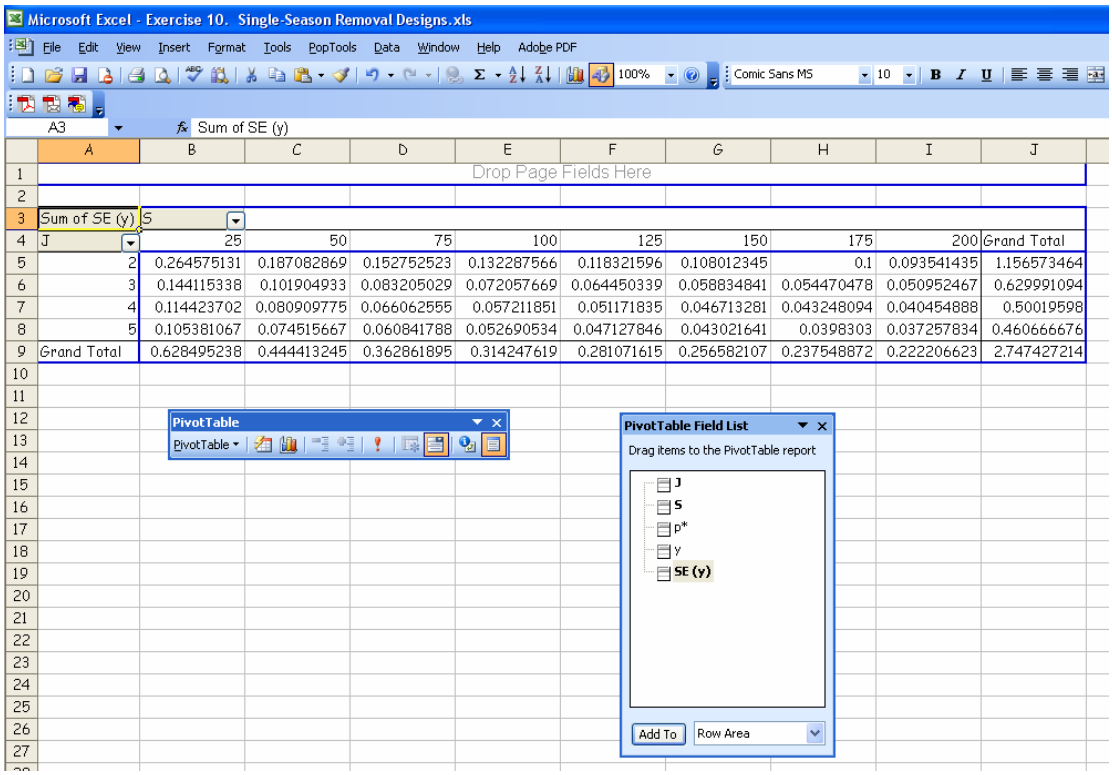
Now specify the location where you'd like the pivot table to go. We selected "New Sheet", but you can put it on the model page if you'd like (just make sure there is ample room so that you don't paste over any of the current spreadsheet cells!).



Then click "Finish", and you should see the following:

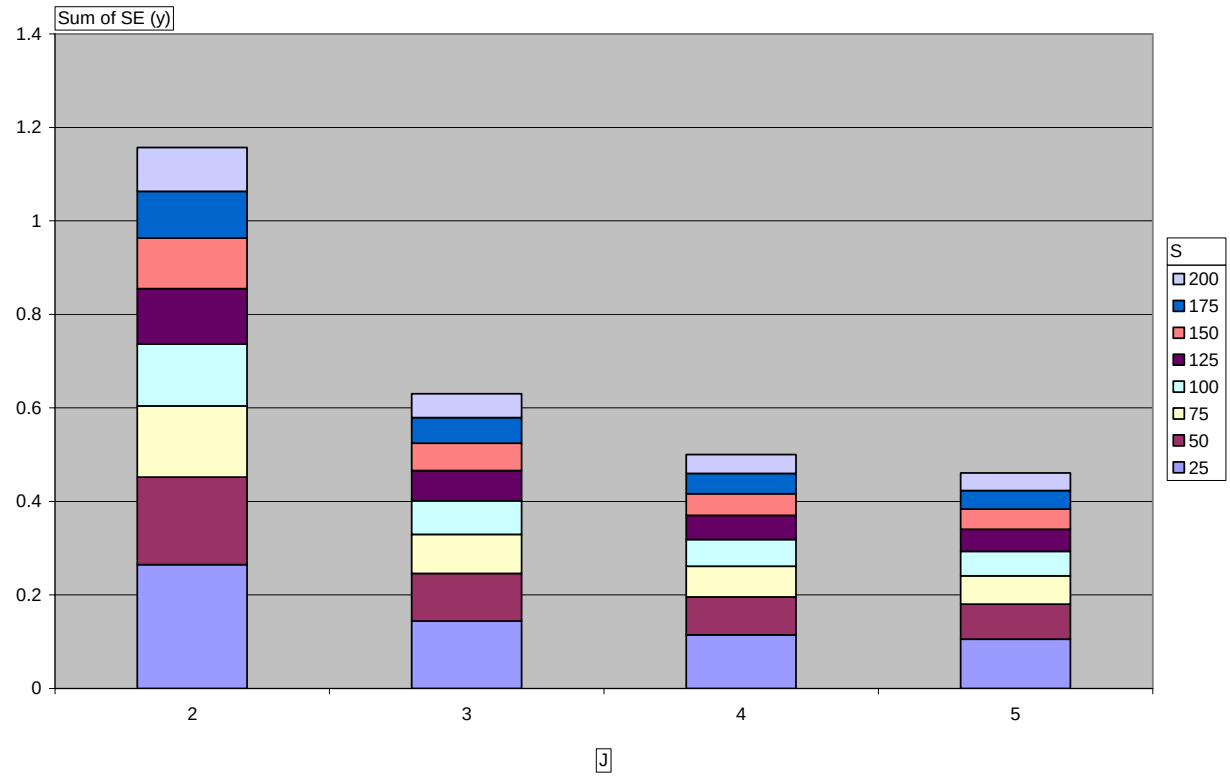


You want a table where the J's are the rows and the S's are the columns (or visa versa), and where the SE(ψ) are the data. So grab the icon labeled J in the box labeled "Pivot Table Field List", and drag it over to the Pivot Table where it reads "Drop Row Fields Here." (Click on the icon once, hold the mouse position, move it to the new location, and then release the mouse click). Then grab the icon labeled S in the box labeled "Pivot Table Field List", and drag it over to the Pivot Table where it reads "Drop Column Fields Here." Then grab the icon labeled SE(y) (which is SE(ψ), but the formatting was lost in the pivot table process) in the box labeled "Pivot Table Field List", and drag it over to the Pivot Table where it reads "Drop Data Items Here." Your pivot table should now look like this:

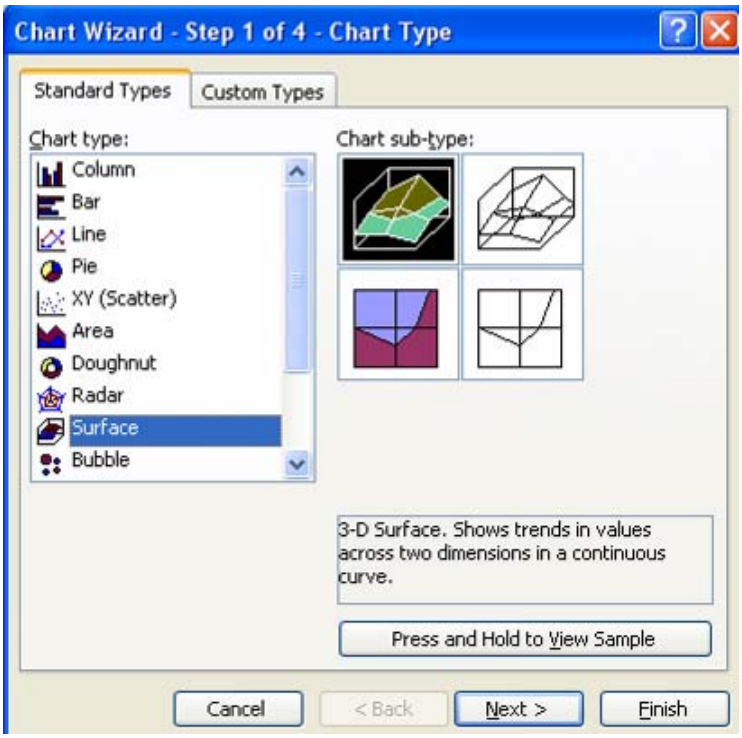


Our pivot table associated with exercise 1 is on a sheet labeled "Pivot Table Data 1." Notice that there is a Pivot Table toolbar, and on it is a graphing icon. If you

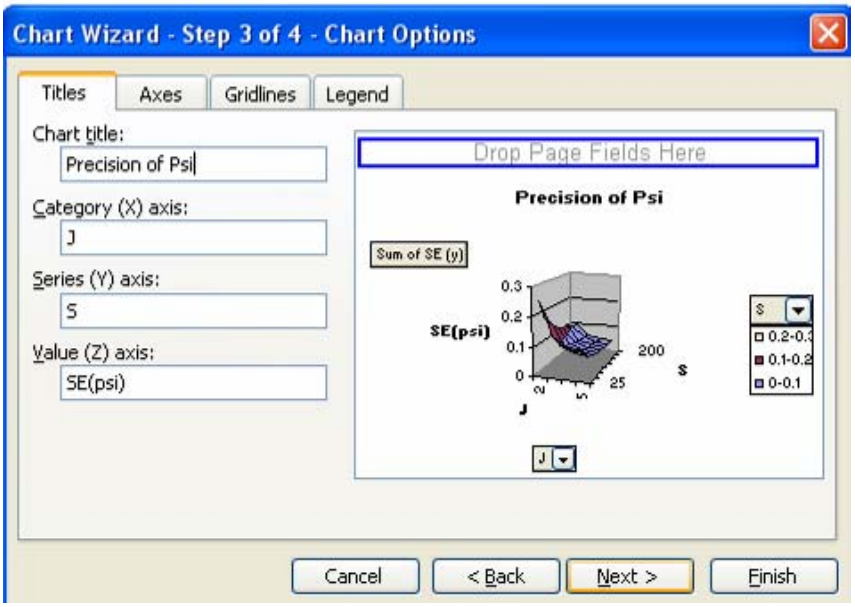
click on it, Excel will create new chart with the data displayed in columns by default:



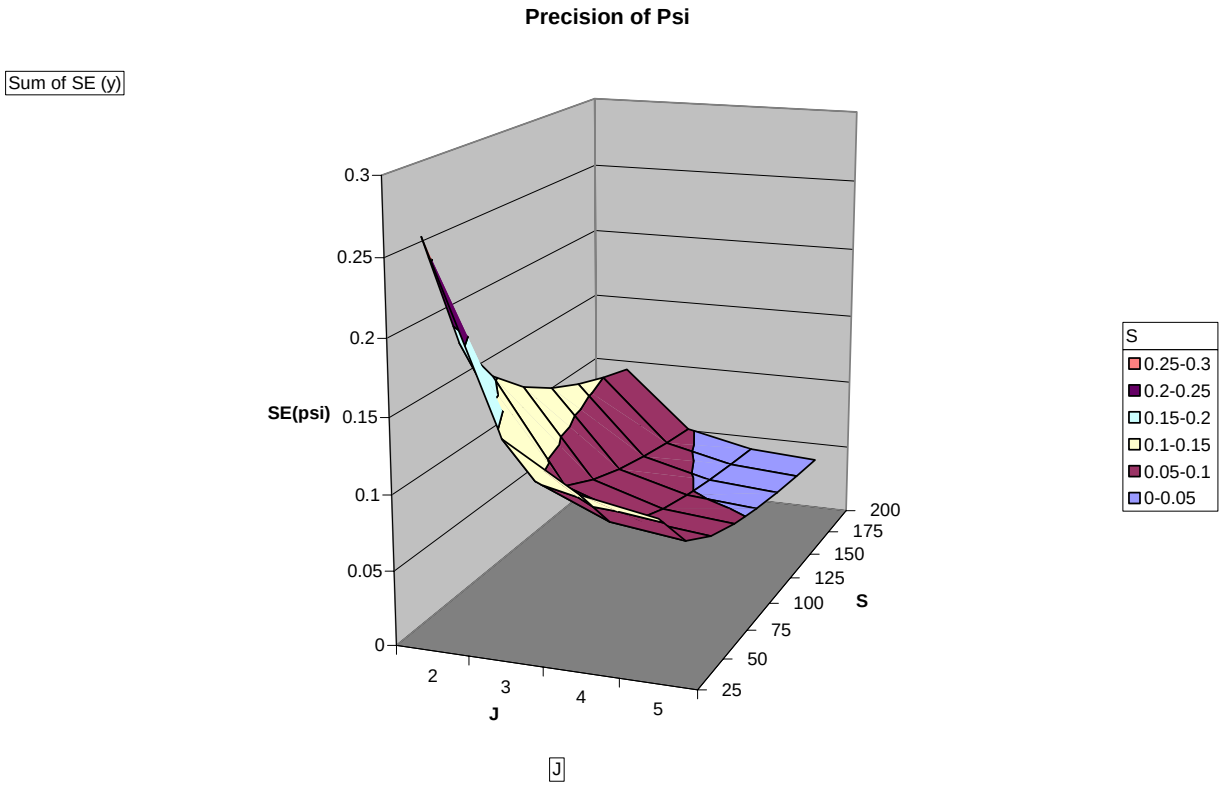
Our graph for exercise 1 is displayed on the sheet labeled "Chart 1". The bar graph isn't very intuitive. So click on the chart button again in the toolbar and now you can select Surface Graph as the charting option:



Click Next and fill in the axis titles:

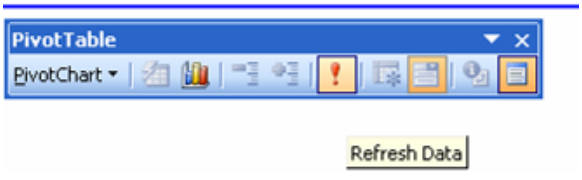


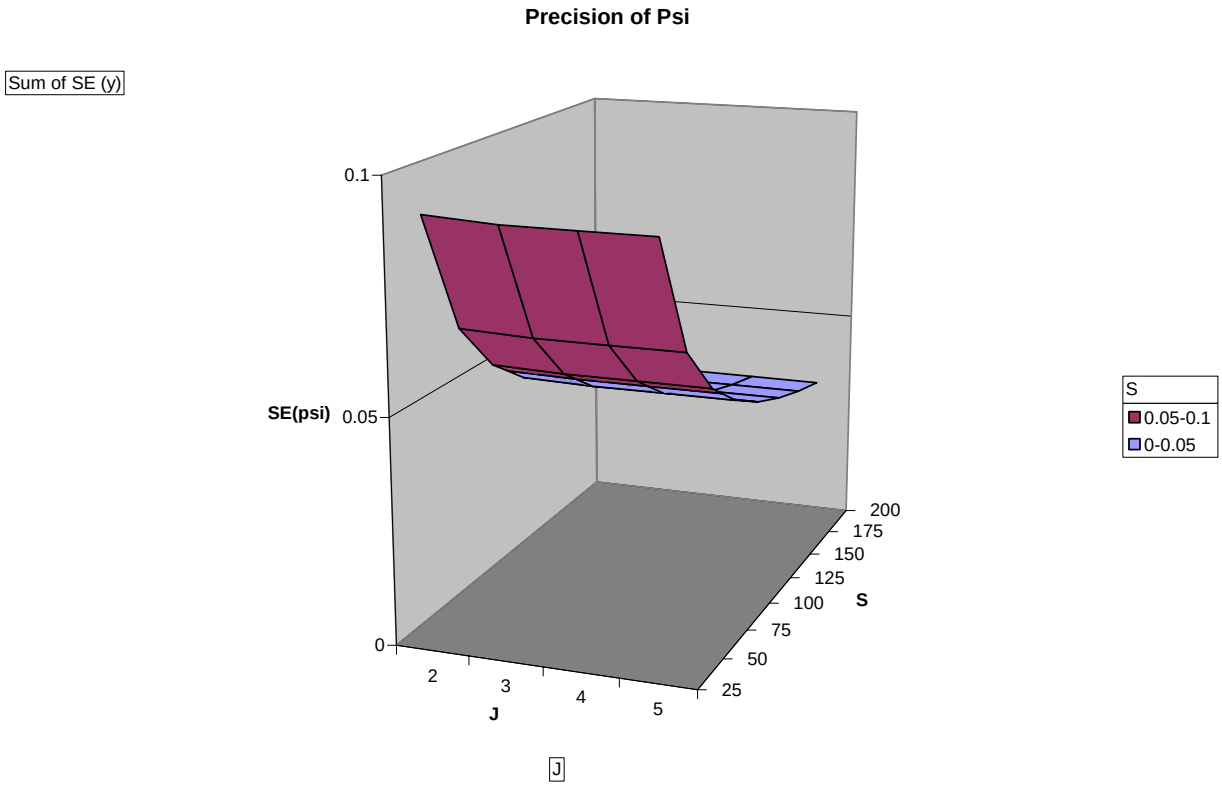
Then press "Finish".



Now, for our species where ψ = 0.5 and p = 0.5, we can clearly evaluate our various study design options. If you specified that SE(psi) should be less than 0.05, you could conduct a study where S is >= 150 and J >=3, and choose which option works best for you within this range.

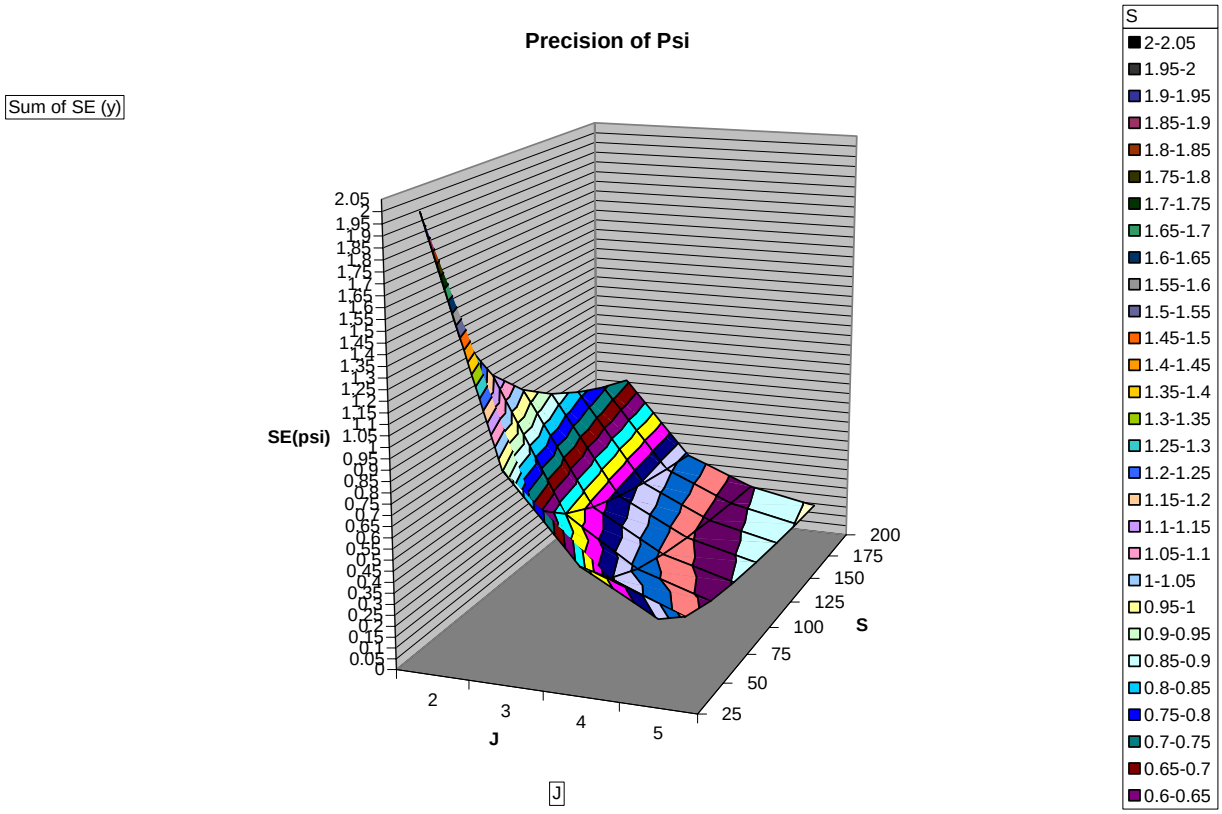
Now, suppose you wanted to evaluate a species where ψ = 0.3 and p = 0.9. Simply enter these parameters in the input section (cells R5:S5), clear out the old results, and run the simulation again. When you press the exclamation point on the Pivot Table toolbar, the table will be automatically refreshed to reflect the new results.





The results for this "new" species look quite different. For a rare species with high detectability, J is far less important than S in terms of optimizing your study design. But note that the Z axis has been automatically scaled to fit the data.

What would the results be if our species was common (ψ = 0.7) but very elusive (p = 0.2)? Here are our results:



For a common species that is elusive, J should be high, but S can be quite low. Again, note that the Z axis is automatically scaled to fit the data (so if you want to compare the graphs of two different kinds of species, make sure the Z axis is scaled the same way). Hopefully you are getting the idea that the optimal study design depends on the characteristics of your target species.

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

In this second exercise, we will explore the optimal removal 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.

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 removal 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 p*, ψ, and SE(ψ). We'll be filling in the following table:

	I	J	K	L	M
32	**Exercise 2: Maximizing SE(y) for species where J = 3 and S = 50.**				
33	ψ	p	p*	ψ	SE (ψ)
34	0.2	0.2			
35	0.2	0.3			
36	0.2	0.4			
37	0.2	0.5			
38	0.2	0.6			
39	0.2	0.7			
40	0.2	0.8			
41	0.3	0.2			
42	0.3	0.3			
43	0.3	0.4			
44	0.3	0.5			
45	0.3	0.6			
46	0.3	0.7			
47	0.3	0.8			
48	0.4	0.2			
49	0.4	0.3			
50	0.4	0.4			
51	0.4	0.5			
52	0.4	0.6			
53	0.4	0.7			
54	0.4	0.8			
55	0.5	0.2			
56	0.5	0.3			
57	0.5	0.4			
58	0.5	0.5			
59	0.5	0.6			
60	0.5	0.7			
61	0.5	0.8			
62	0.6	0.2			
63	0.6	0.3			
64	0.6	0.4			
65	0.6	0.5			
66	0.6	0.6			
67	0.6	0.7			
68	0.6	0.8			
69	0.7	0.2			
70	0.7	0.3			
71	0.7	0.4			
72	0.7	0.5			
73	0.7	0.6			
74	0.7	0.7			
75	0.7	0.8			
76	0.8	0.2			
77	0.8	0.3			
78	0.8	0.4			
79	0.8	0.5			
80	0.8	0.6			
81	0.8	0.7			
82	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:

	R	S	T	U	V	W	X	Y
3	**Inputs**							
4	**ψ**	**p1**	**p2**	**p3**	**p4**	**p5**	**J**	**S**
5	0.2	0.2	0.2	0.2	0.2	0.2	3	50

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 X14:X31), and these will be pasted into the model for analysis (cells F7:F24).

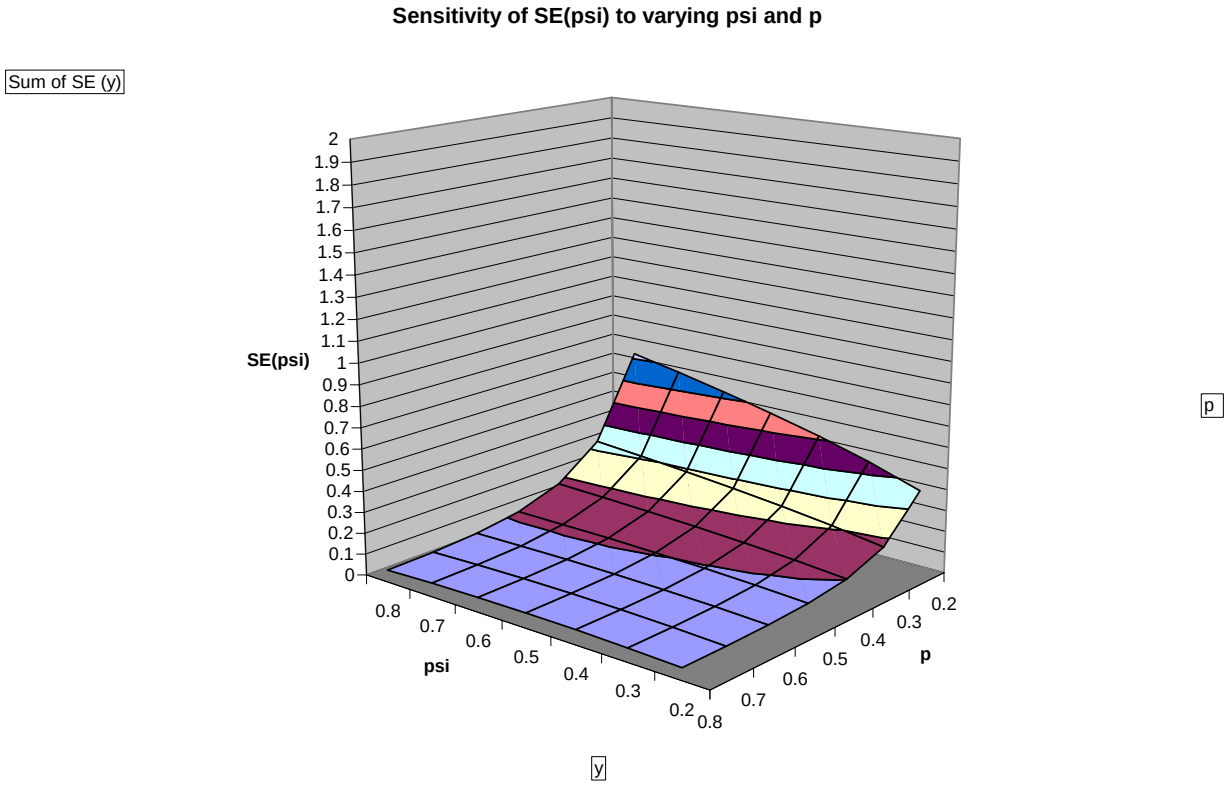
	W	X
13	Histories	Frequency
14	1.	0.00
15	01	0.00
16	00	0.00
17	1..	2.00
18	01.	1.60
19	001	1.28
20	000	45.12
21	1...	0.00
22	01..	0.00
23	001.	0.00
24	0001	0.00
25	0000	0.00
26	1....	0.00
27	01...	0.00
28	001..	0.00
29	0001.	0.00
30	00001	0.00
31	00000	0.00

	E	F	G	H
6	Histories	Frequency	Parameters	Estimated?
7	1.	0.00	p_1	
8	01	0.00	p_2	
9	00	0.00	ψ	
10	1..	2.00	p_1	1
11	01.	1.60	p_2	
12	001	1.28	p_3	
13	000	45.12	ψ	1
14	1...	0.00	p_1	
15	01..	0.00	p_2	
16	001.	0.00	p_3	
17	0001	0.00	p_4	
18	0000	0.00	ψ	
19	1....	0.00	p_1	
20	01...	0.00	p_2	
21	001..	0.00	p_3	
22	0001.	0.00	p_4	
23	00001	0.00	p_5	
24	00000	0.00	ψ	

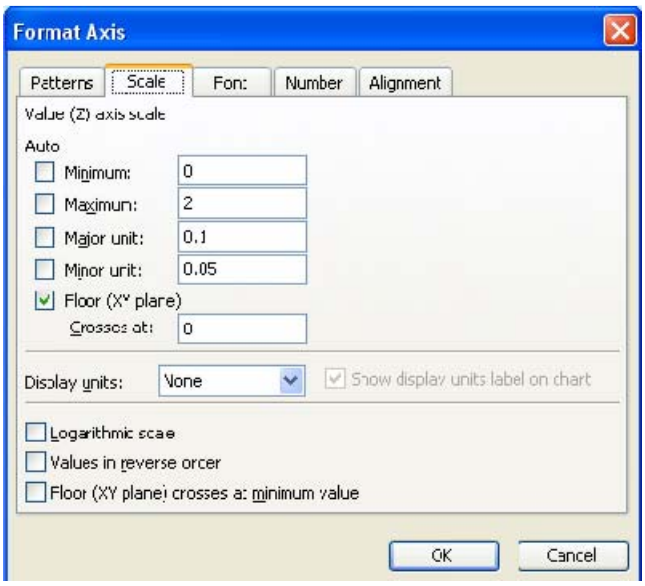
As before, we recorded a macro to run all 49 simulations for you. Just press the "Clear Data" button to clear out old results, enter data for J and S in cells X5:Y5, and then press the button labeled "Run Analysis # 2" to run the simulation. Your results should look like this:

	I	J	K	L	M
32	**Exercise 2: Maximizing SE(y) for species where J = 3 and S = 50.**				
33	ψ	p	p*	ψ	SE (ψ)
34	0.2	0.2	0.4880	0.2000	0.3637
35	0.2	0.3	0.6570	0.2000	0.1707
36	0.2	0.4	0.7840	0.2000	0.1012
37	0.2	0.5	0.8750	0.2000	0.0732
38	0.2	0.6	0.9360	0.2000	0.0622
39	0.2	0.7	0.9730	0.2000	0.0582
40	0.2	0.8	0.9920	0.2000	0.0569
41	0.3	0.2	0.4880	0.3000	0.4447
42	0.3	0.3	0.6570	0.3000	0.2076
43	0.3	0.4	0.7840	0.3000	0.1215
44	0.3	0.5	0.8750	0.3000	0.0862
45	0.3	0.6	0.9360	0.3000	0.0721
46	0.3	0.7	0.9730	0.3000	0.0670
47	0.3	0.8	0.9920	0.3000	0.0653
48	0.4	0.2	0.4880	0.4000	0.5128
49	0.4	0.3	0.6570	0.4000	0.2381
50	0.4	0.4	0.7840	0.4000	0.1374
51	0.4	0.5	0.8750	0.4000	0.0954
52	0.4	0.6	0.9360	0.4000	0.0784
53	0.4	0.7	0.9730	0.4000	0.0720
54	0.4	0.8	0.9920	0.4000	0.0699
55	0.5	0.2	0.4880	0.5000	0.5724
56	0.5	0.3	0.6570	0.5000	0.2643
57	0.5	0.4	0.7840	0.5000	0.1503
58	0.5	0.5	0.8750	0.5000	0.1019
59	0.5	0.6	0.9360	0.5000	0.0817
60	0.5	0.7	0.9730	0.5000	0.0740
61	0.5	0.8	0.9920	0.5000	0.0715
62	0.6	0.2	0.4880	0.6000	0.6261
63	0.6	0.3	0.6570	0.6000	0.2874
64	0.6	0.4	0.7840	0.6000	0.1610
65	0.6	0.5	0.8750	0.6000	0.1061
66	0.6	0.6	0.9360	0.6000	0.0825
67	0.6	0.7	0.9730	0.6000	0.0733
68	0.6	0.8	0.9920	0.6000	0.0702
69	0.7	0.2	0.4880	0.7000	0.6752
70	0.7	0.3	0.6570	0.7000	0.3082
71	0.7	0.4	0.7840	0.7000	0.1698
72	0.7	0.5	0.8750	0.7000	0.1083
73	0.7	0.6	0.9360	0.7000	0.0809
74	0.7	0.7	0.9730	0.7000	0.0698
75	0.7	0.8	0.9920	0.7000	0.0659
76	0.8	0.2	0.4880	0.8000	0.7207
77	0.8	0.3	0.6570	0.8000	0.3271
78	0.8	0.4	0.7840	0.8000	0.1771
79	0.8	0.5	0.8750	0.8000	0.1087
80	0.8	0.6	0.9360	0.8000	0.0767
81	0.8	0.7	0.9730	0.8000	0.0630
82	0.8	0.8	0.9920	0.8000	0.0580

We created a pivot table from these data (listed on the sheet labeled Pivot Table Data 2) and created a surface chart from the pivot data (provided on the sheet labeled Chart 2). Here are the results in graphical form:



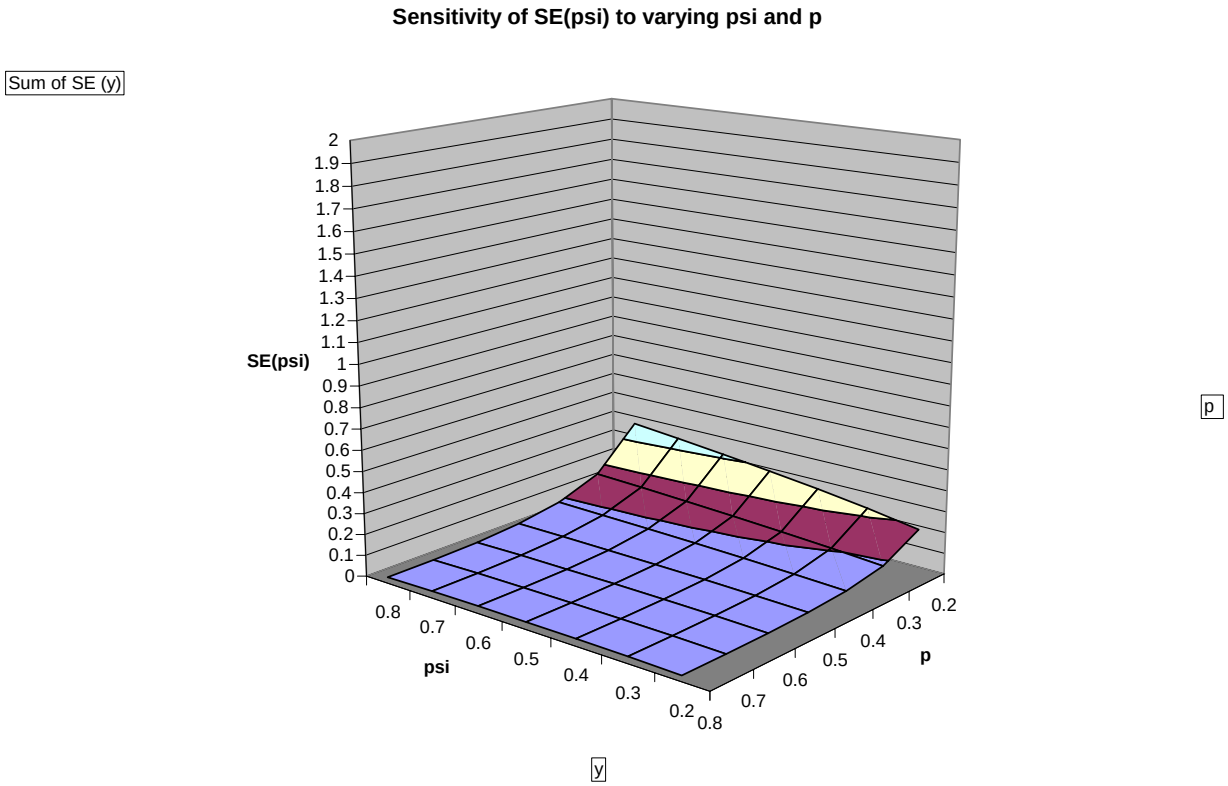
So, a study in which J = 3 and S = 50 is most appropriate for species in which p > 0.5. Keep in mind that the categories shown above show SE(ψ) in increments of 0.1. Note also that we scaled this chart so that SE(ψ) has a minimum value of 0 and a maximum value of 2 so that you can compare different runs more effectively. To set the scales, double-click on the axis for SE(ψ), and then click on the Scale tab, click off the check-boxes that automatically scale the minimum and maximum values, and then enter the minimum and maximum values you wish (see below).



We didn't choose this option for exercise 1 because doing so can produce some very funky-looking graphs!

Under this removal design, the study yields the highest precision for ψ for an organism where ψ = 0.2 and p = 0.8, but the standard error is still 0.0569. The variance is made up of 2 parts, one involving psi, the other involving p. The first part should be symmetrical (var(psi=0.2) = var(psi=0.8)) since you have psi*(1-psi) in part 1. The 2nd part should be smallest for the highest value of p. So, the smallest variance should be at psi close to 0 or close to 1, and p close to 1.

You can play around with different values of J and S. Just enter new values in cells X5:Y5, clear out the old results, and then run the simulation again. Don't forget to press the button with the exclamation point to update the pivot table and chart to view your results. For example, here are our results for J = 3, S = 200:

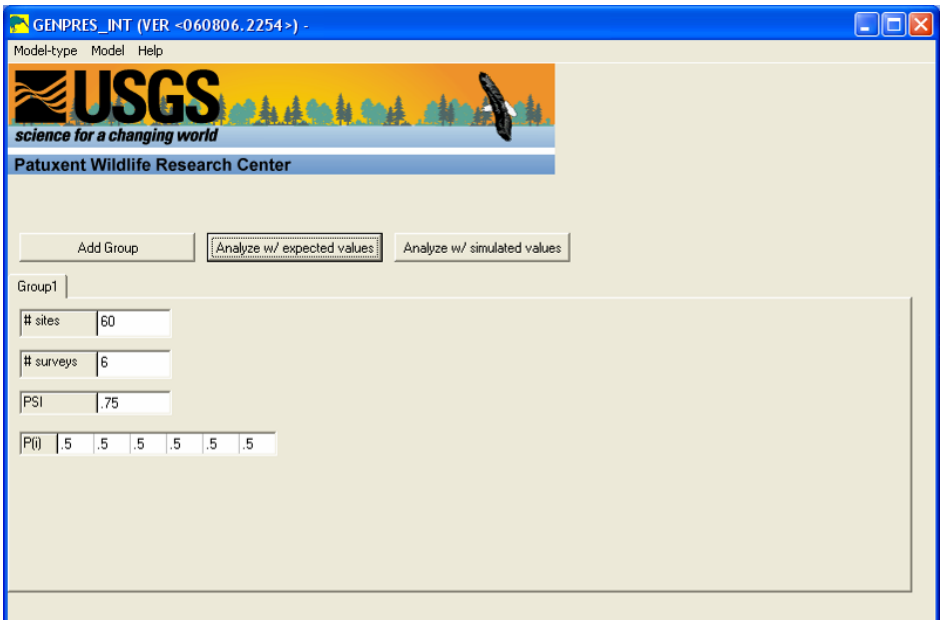


Feel free to explore a variety of options in the spreadsheet, either by setting ψ and p and running simulations where S and J vary (exercise 1), or by setting S and J and running simulations where ψ and p vary (exercise 2). We also suggest you read the article by MacKenzie and Royle, who present summarized results for removal model designs. Now, let's take a look at how you would use the program GENPRES to assess various removal designs.

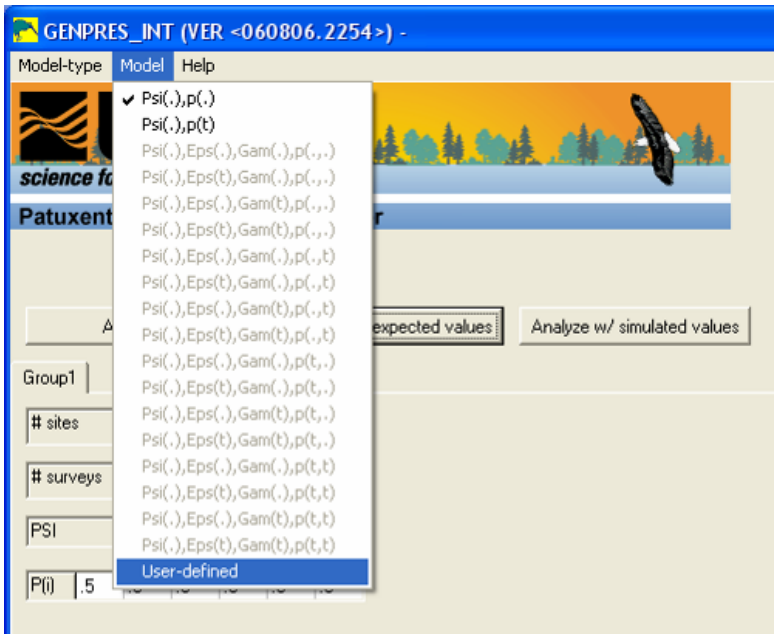
OPTIMAL REMOVAL DESIGNS IN PROGRAM GENPRES

GETTING STARTED WITH GENPRES

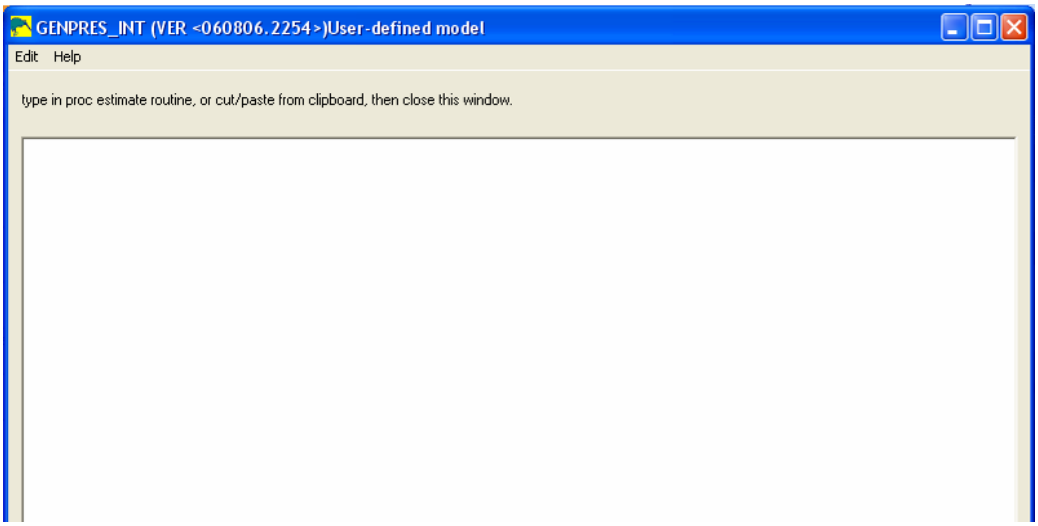
OK, now we will take a look at how to program a removal model design in program GENPRES. Open GENPRES and you'll see the main form:



The removal model is not a standard design, so go to Model | User-defined



You'll be brought to a form where you can type in code to analyze various removal models.



CONCLUSIONS

A priori knowledge of your study organism and clear research objectives are valuable in determining survey design and allocating survey effort, especially if resources are limited. When dealing with a rare and difficult to detect species (low p and ψ), your maximum number of surveys will most likely need to be higher to get unbiased, more precise model output. The large effect detection probability has on output variability and bias reinforces the importance of its incorporation into the modeling process. Also keep in mind that in some instances, the removal design may be less robust to violation of assumptions than the standard design. Careful consideration of study design is crucial for developing sound management and conservation efforts based on species distribution and occurrence across the landscape.

LITERATURE CITED

Hanski, I. 1998. Metapopulation dynamics. Nature **396**:41-49.

Holt, R., and T. Keitt. 2005. Species' borders: a unifying theme in ecology. OIKOS **108**:3-6.

MacKenzie, D. 2006. What are issues with presence-absence data for wildlife managers? Journal of Wildlife Management **69**:849-860.

MacKenzie, D., J. Nichols, J. Hines, M. Knutson, and A. Franklin. 2003. Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. Ecology **84**:2200-2207.

MacKenzie, D., J. Nichols, J. Royle, K. Pollock, L. Bailey, and J. Hines. 2006. Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence. Elsevier Inc., Oxford, UK.

MacKenzie, D., and J. Royle. 2005. Designing occupancy studies: general advice and allocating survey effort. Journal of Applied Ecology **42**:1105-1114.