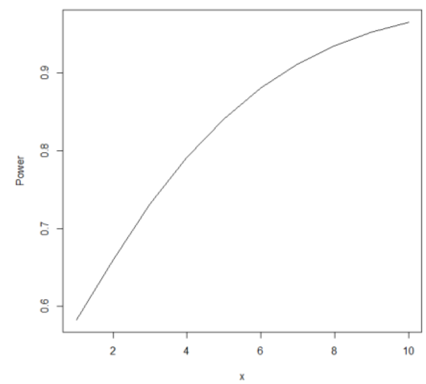
Power Analysis – Estimating Power to detect difference in estimated occupancy.

Project Description and Context



In this module, we'll use GENPRES to generate data for two groups of sites, which have different rates of occupancy. We specify a sampling design in a single season framework (number of visits to sites, per visit detection probability) and a value for test significance level, and we then compute the corresponding power of a likelihood-ratio test of the null hypothesis of no difference in occupancy rates for the 2 groups of sites. Power is defined as the probability that our test statistic indicates a "significant" difference, given that the alternative hypothesis is true and given the characteristics of the sampling design and test. This kind of analysis is very useful as you design a study where the objective is to be able to compare occupancy between sites, or across time, and you need to determine how much power your design will have to detect these differences if they exist. This can be done in two ways: 1) Generate expected value data and compute the power using the difference in log-likelihood values of the two hypothesized models (H0: psi(.),p(.), vs Ha: psi(grp),p(.)) as a parameter of the non-central chi-square distribution, and 2) Generate simulated data and compute the proportion of simulations where the difference in log-likelihood is greater than the threshold for a 'significant' chi-square statistic. For large sample sizes, the approaches should yield very similar answers, but for small samples, the simulations should usually provide more reliable results.

{FYI: Expected value data are data which are generated by computing the probability (p) of each outcome (detection history), and multiplying it by the number of sites (N). For example, the expected number of heads when flipping a fair coin 10 times is 5 (p X N = 0.5 X 10 = 5). Simulated data are generated by drawing random numbers between zero and one, and comparing them to model parameters, (occupancy or detection) to determine if each site is occupied or species detected. Simulating the process of flipping a fair coin 10 times might yield 4 heads, or 6 heads, or any pair of numbers between that sum to ten. If the simulation is repeated many times, the average number of heads per ten flips would be 5.}

In addition to programs PRESENCE and GENPRES, the program, POWER, will be needed. This program can be run in a web-browser window.

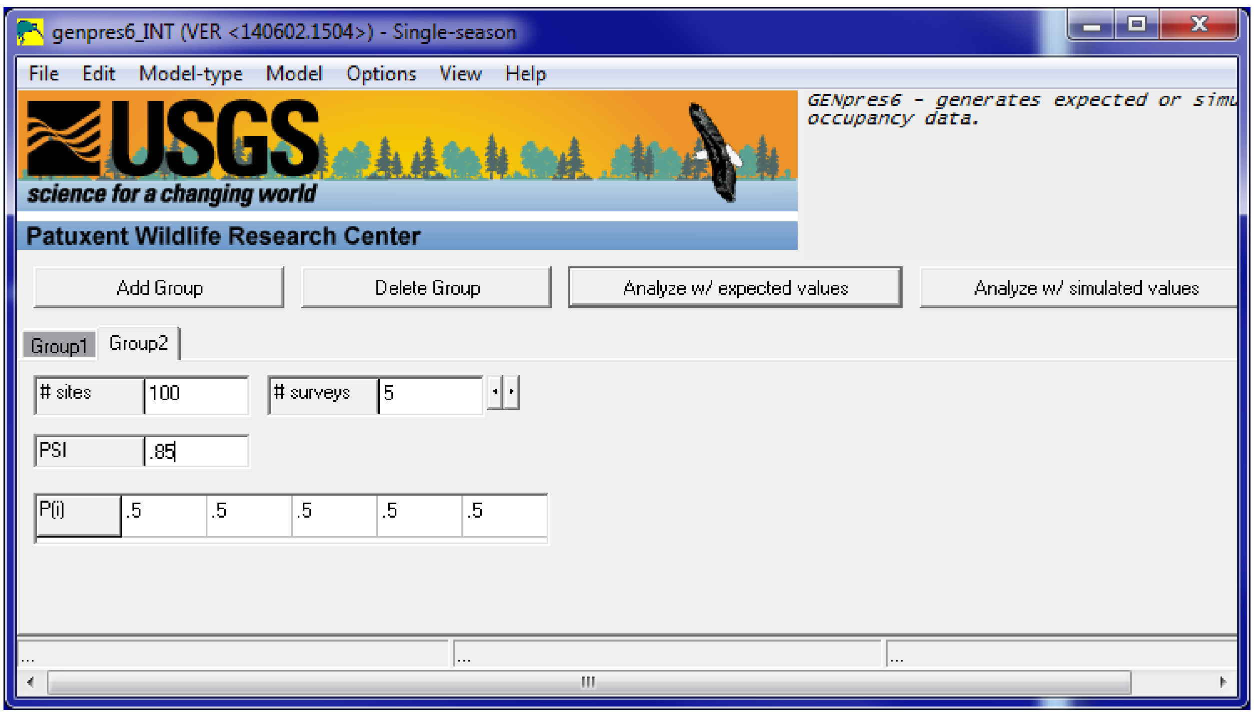
Exercise Objectives

- Learn how to create expected-value and simulated data using GENPRES
- Learn to compute power using both methods

Power Analysis – Estimating Power to detect difference in estimated occupancy.

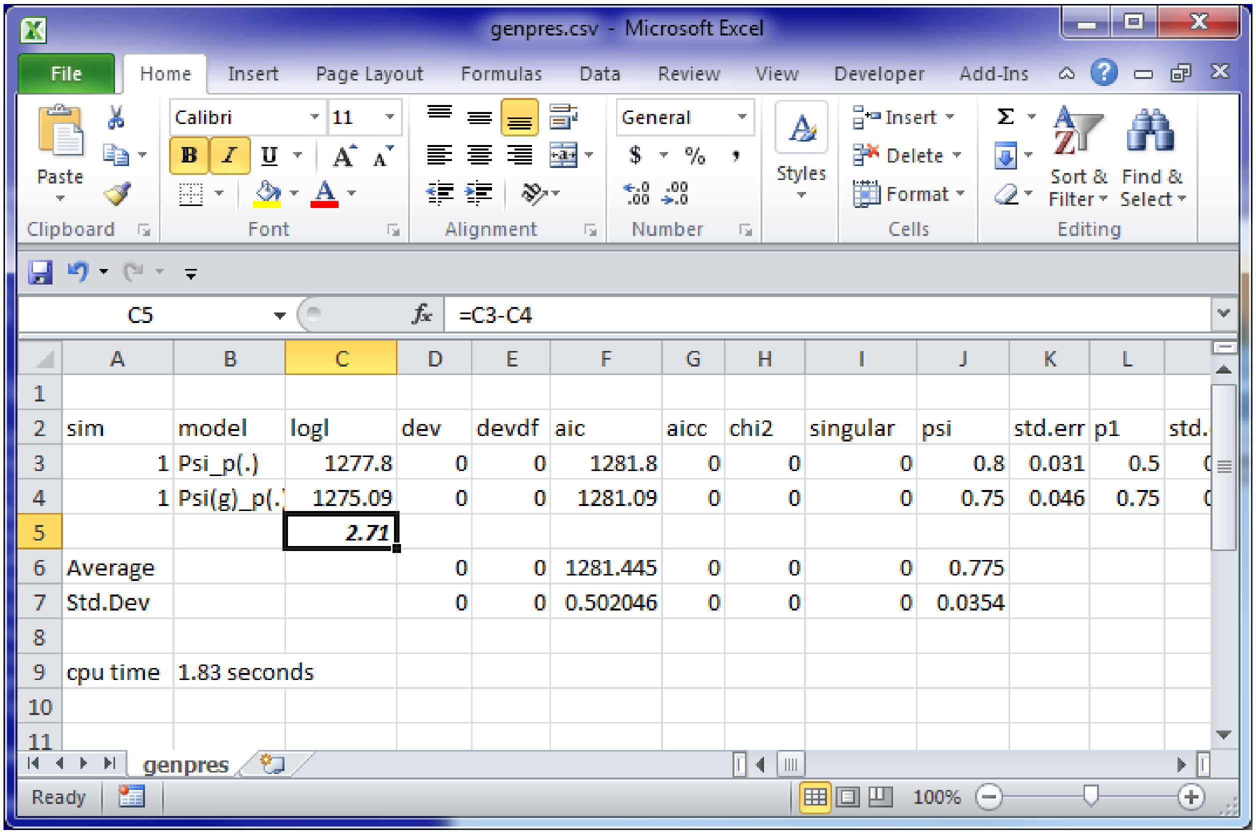
INSTRUCTIONS

Step 1 – Compute power using expected-value method: Begin GENPRES and create a 2^{nd} group of sites by clicking the 'Add Group' button. Click on the tab for the 2^{nd} group, then change the true probability of occupancy (PSI) for those sites to 0.85 (the default for Group 1 should be 0.75). Next, click the 'Model' menu, then click 'User-Defined', then 'Define model by name'. In the entry box, type 'Psi(g),p(.)' (without quotes), then click 'OK'. Generate expected-value data by clicking the 'Analyze w/ expected values' button, then click 'Y' to overwrite any previously generated csv files. GENPRES will generate data, and build the necessary input design-matrix files to run two models: 1) Psi(.),p(.) (which was selected by default), and 2) Psi(g),p(.). After running those two models, GENPRES will scan the output for the log-likelihood values and parameter estimates and save in a spreadsheet (csv) file. GENPRES will attempt to open the csv file with Excel, but sometimes fails to find where Excel is located on your computer. In that case, you can specify a location in the 'Options' menu. (Note: Version 6 or later of GENPRES is required to do this exercise.)



When the spreadsheet opens in Excel, calculate the difference in log-likelihood values by entering '=c3-c4' in cell C5. (Actually, the log-likelihood values in the spreadsheet are really -2*log-likelihood values, so the quantity will be twice the difference in log-likelihood.)

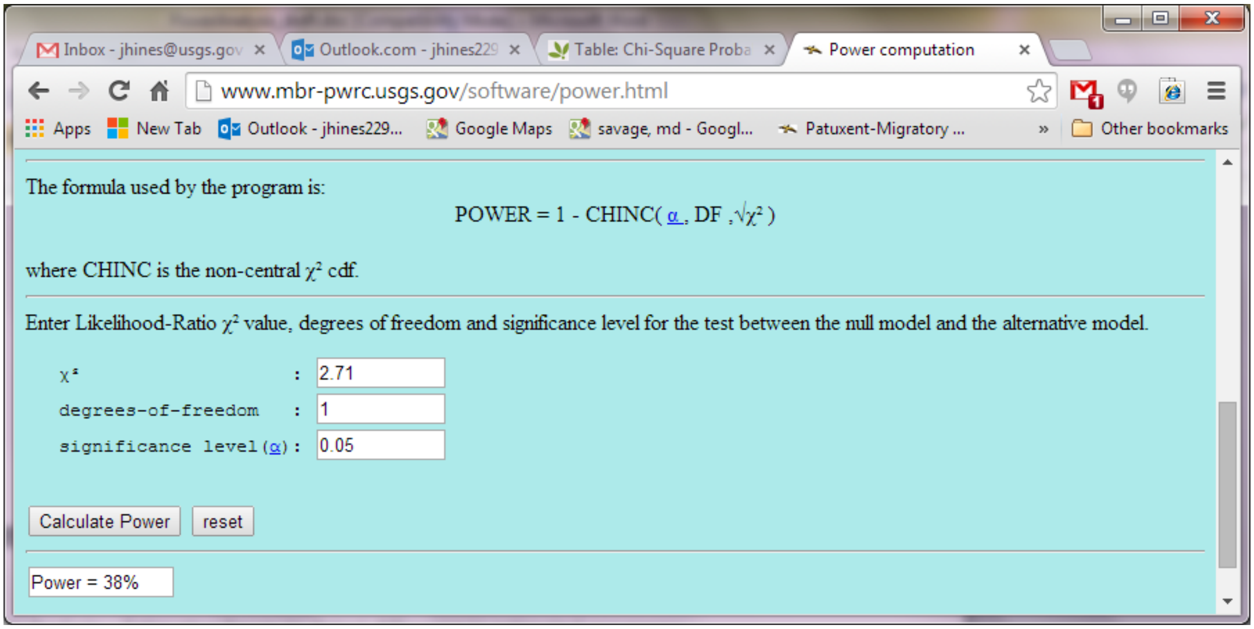
Power Analysis – Estimating Power to detect difference in estimated occupancy.



	A	B	C	D	E	F	G	H	I	J	K	L	
1													
2	sim	model	logl	dev	devdf	aic	aicc	chi2	singular	psi	std.err	p1	std.
3	1	Psi_p(.)	1277.8	0	0	1281.8	0	0	0	0.8	0.031	0.5	(
4	1	Psi(g)_p(.	1275.09	0	0	1281.09	0	0	0	0.75	0.046	0.75	(
5			2.71										
6	Average			0	0	1281.445	0	0	0	0.775			
7	Std.Dev			0	0	0.502046	0	0	0	0.0354			
8													
9	cpu time	1.83 seconds											
10													
11													

Next, open a web-browser and type ‘http://www.mbr-pwrc.usgs.gov/software’ into the address bar. Scroll down to the ‘Power’ program link, click it, then click the ‘Run program online’ link. When the web-page appears, replace the chi-square value (4.23) with the value calculated in the spreadsheet. *The number of degrees of freedom should be the difference in number of parameters in the two models*. In this case, it is 1, so the default value on the web-page is OK. Click ‘Calculate power’ and observe the value reported.

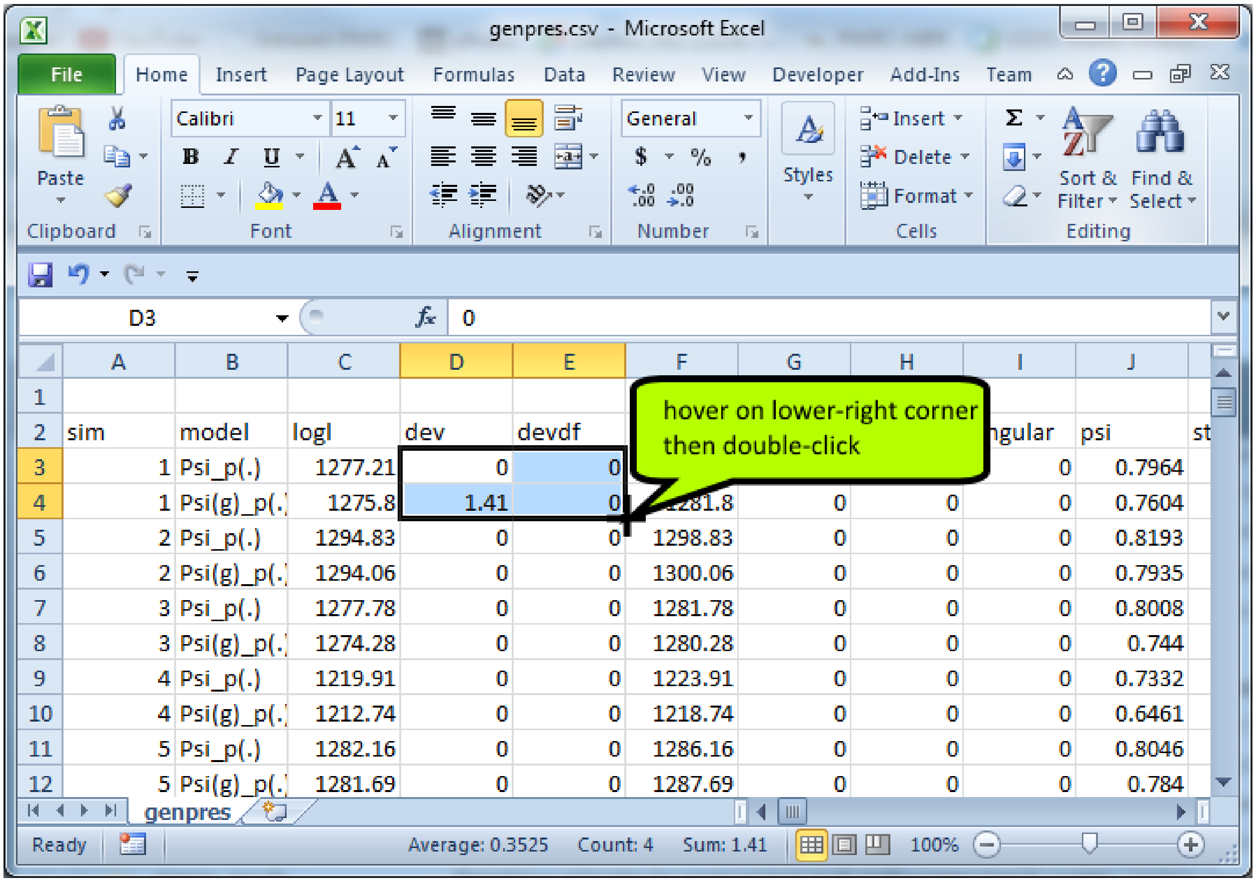
Power Analysis – Estimating Power to detect difference in estimated occupancy.



Step 2 – – Compute power using simulated-values method:

Close the spreadsheet, and click on the GENPRES window. Click 'Analyze- simulate by site' button and enter 20 in the box for the number of simulations to do. When the program finishes, compute the difference in log-likelihood for the 1st simulation by entering '=c3-c4' in cell D4. Enter '=if(D4>3.84,1,0)' in cell E4. (The "If" function causes the cell to contain either the value "1" or "0", depending on the result of the test, "D4>3.84". If cell D4 is greater than 3.84, the cell will contain "1". Otherwise the cell will contain "0". The 3.84 value in the equation represents the minimum 'significant' chi-square value at the 5% level when degrees of freedom = 1. This value was obtained from a chi-square lookup table [use web-search to find one].)

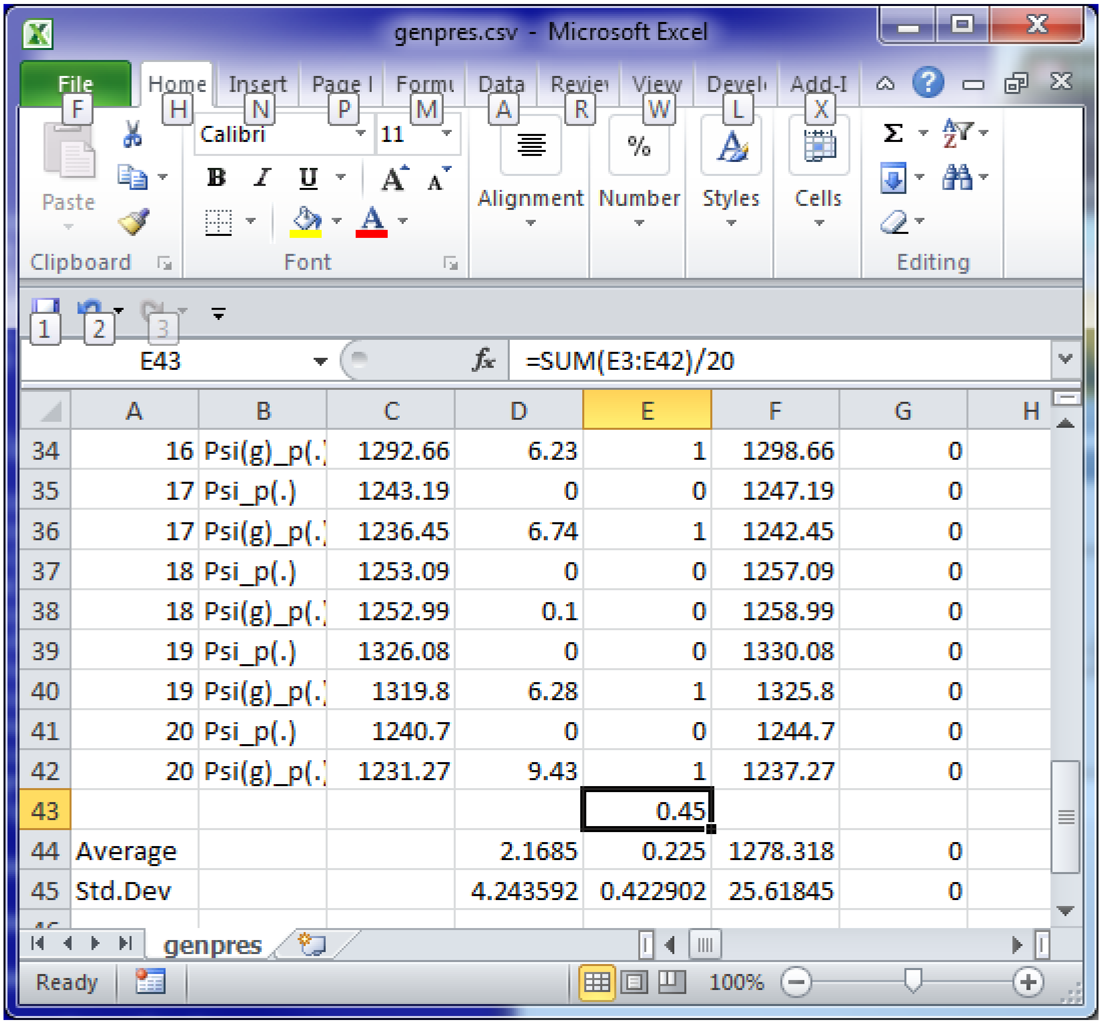
Power Analysis – Estimating Power to detect difference in estimated occupancy.



Then, select cells D3:E4 to the clipboard (shown above), and hover the cursor over the lower-right corner. When the cursor changes to a '+' symbol, double-click. This will copy those 4 cells down through the rest of the rows in the spreadsheet. In cell, E43, enter '=sum(E3:E42)/20'. This cell now contains the proportion of simulations which gave a 'significant' chi-square value. This proportion estimates the power to detect the difference in occupancy. The '20' in the formula represents the number of simulations. The power from the simulated data analysis (value in cell E43, 45% in my case) should be similar to the value computed in the expected-value analysis (38% in this case). However, with only 20 simulations, the value for power from the simulations could vary greatly. If you have more time and the motivation, you could do a few hundred simulations and the two methods should yield similar estimates of power.

Note that this same basic simulation approach can be used with model selection statistics as well. For example, one might compute the proportion of simulations in which the AICc value was smaller for the "group" model than for the "dot" model. Similarly, one might compute the proportion of simulations for which the AICc value for the dot model was at least 2 greater than that for the group model. The goal of these computations would be to provide an idea of how likely it will be to conclude, now based on a model selection approach instead of a hypothesis-testing approach, that the group model is preferable to the dot model.

Power Analysis – Estimating Power to detect difference in estimated occupancy.



	A	B	C	D	E	F	G	H
34	16	Psi(g)_p(.)	1292.66	6.23	1	1298.66	0	
35	17	Psi_p(.)	1243.19	0	0	1247.19	0	
36	17	Psi(g)_p(.)	1236.45	6.74	1	1242.45	0	
37	18	Psi_p(.)	1253.09	0	0	1257.09	0	
38	18	Psi(g)_p(.)	1252.99	0.1	0	1258.99	0	
39	19	Psi_p(.)	1326.08	0	0	1330.08	0	
40	19	Psi(g)_p(.)	1319.8	6.28	1	1325.8	0	
41	20	Psi_p(.)	1240.7	0	0	1244.7	0	
42	20	Psi(g)_p(.)	1231.27	9.43	1	1237.27	0	
43					0.45			
44	Average			2.1685	0.225	1278.318	0	
45	Std.Dev			4.243592	0.422902	25.61845	0	

Exercise:

1. Calculated power above, using a 5-survey single season design to detect a 10% difference in occupancy, was 38% using the large-sample approximation (expected values). Imagine a scenario where you are tasked with managing an endangered species, where your management objective is to detect a 25% change (decline) in occupancy, with at least 80% power and a significance level of 5% (assuming initial occupancy is 0.75, detection is 0.3, and 100 sites/group). How many surveys would you need at the default value of 100 sites per group? How many sites would you need to survey at the default value of 5 surveys?
2. Working in small groups, determine the power to detect a difference in detection probability over time (5 surveys), using the models: Psi(.),p(.) vs Psi(.),p(t). Assume 1 group, 100 sites, 5 surveys, occupancy is 75%, and detection probabilities are: 0.3, 0.35, 0.4, 0.45, 0.5.