JMP Academic Case Study 063

Caught on Camera: Analyzing Camera Trap Data to Inform Conservation

Contingency Analysis, Multiple Comparisons,
False Discovery Rate Control

Produced by

Jennifer L. Verdolin, PhD
verdolin.jennifer@gmail.com

JMP STATISTICAL DISCOVERY

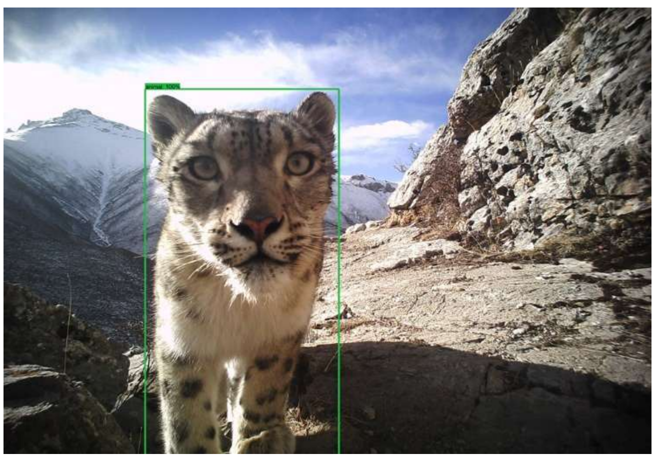
Caught on Camera: Analyzing Camera Trap Data to Inform Conservation
Contingency analysis, multiple comparisons, false discovery rate control

Key Ideas

This case study presents an environmental conservation scenario in which a large family of contingency analyses are conducted. Family-wise error rate is explained, and false discovery rate (FDR) control is applied to restrict the proportion of false positives (i.e., Type I errors) within the group of positive test results. The impact of FDR control on the number of statistically significant results is illustrated.

Background

Monitoring wildlife is critical for managing wilderness areas and guiding conservation efforts. Traditionally, wildlife monitoring involved manually surveying individual animals along a transect of some width and length. Manual surveying is costly and time-consuming, especially when conducted over large areas, such as those of the U.S. National Parks. In addition, manual surveying often misses rare, hard-to-observe species. For example, snow leopards are notoriously difficult to observe in the wild, making it hard to estimate the population size of this elusive and endangered feline.



Technological advances have fueled innovations in wildlife monitoring. The use of *camera traps* – motion-sensitive, remotely triggered cameras – has increased substantially, improving wildlife biologists' ability to monitor remote areas and rare species. A camera trap can be placed in a remote location, and any time an animal moves in view of the camera, a still image is captured. Researchers can then manually identify the species observed, or machine learning algorithms can attempt to classify the image automatically. Camera traps have the benefit of being cost-effective, minimally invasive, deployable across a range of habitats, and able to generate information on multiple species.

Camera traps can do more than reveal which and how many species live in a given area. Camera trap data can also be used to evaluate how species influence each other, whether human recreational activity is impacting species, how species are distributed across habitat types, and if living close to the edge of a park boundary presents challenges to certain species. Sometimes it can reveal surprising interactions between species, like this video of a badger and coyote on the move together or this video of a mountain lion and male deer walking together in the forest at night.

Individual camera trap observations such as these can yield important insights in themselves. Going further, statistical analysis of large-scale camera trap data sets can reveal important patterns in the distribution of species and interactions or dependencies among them, which in turn can guide further research and conservation efforts.

The Task

We will use camera trap data to analyze co-occurrence patterns among 21 species of animals in the mountains of southwestern Arizona in the United States of America. Which types of animals co-occur together in the wild? Which do not?

For the purposes of this case study, we will focus on black bears (Ursus Americanus). Specifically, we will identify types of animals that are observed more or less frequently when black bears are also observed by the same camera within the same calendar month. This will provide insight into how the presence of black bears is associated with increased or decreased presence of other species in the same location and during the same general time of year. For example, black bears are generalist predators and represent a significant threat to white-tailed deer fawns. We may find that evidence that deer attempt to avoid territory currently occupied by black bears.

Black bears are the focus here, but the exercises will ask you to choose another animal for investigation. The data contain 21 types of animals for which we can investigate patterns of co-occurrence, creating great potential to confirm existing understandings of interactions or dependencies as well as to provide clues to interactions or dependencies not yet known.

To complete our task, we will:

- Perform contingency analyses to uncover associations between observation of black bears with observation of the other 20 species in the data.
- Apply false discovery rate control to account for conducting a family of 20 hypothesis tests.
- Visualize species co-occurrences using an interactive geographic map.

The Data **camera-trap.jmp**

These real-world data were collected by the University of Arizona Jaguar and Ocelot Monitoring Project. The data set contains camera trap observations for 21 species of animals across two mountain ranges in southern Arizona, the Chiricahua and the Dos Cabezas. Each row represents all observations made by a single camera trap within a calendar month, with the data set spanning February 2018 to September 2021. There is one column per animal type, with each value indicating whether that animal was observed by the given camera trap during the given month.

The variables in the data set include:

Year	The year in which the observations were made.
Month	The month in which the observations were made.
Site	The mountain range in which the observations were made. (CHI = Chiricahua, DOS = Dos Cabezas)
Site Number	ID of the unique camera trap that made the observations.
Latitude	Latitude of the camera trap location.
Longitude	Longitude of the camera trap location.
[21 Species Cols]	Binary indicators of whether each species was observed (1 = yes, 0 = no).

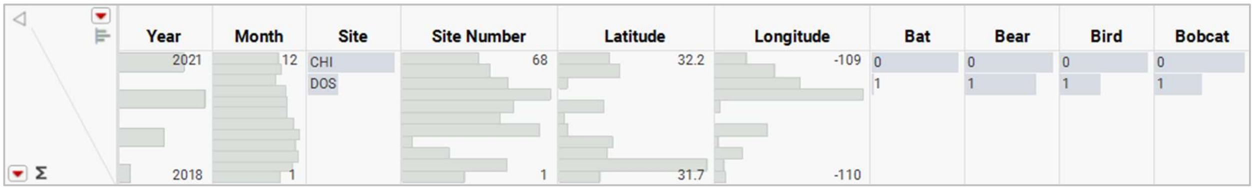
For example, Row 1 contains data from February 2018 for camera trap 27 located in the Chiricahua Mountains. The columns Coati, Cow, Coyote, Deer, Fox, and Puma have values of 1, indicating these were each observed by camera trap 27 one or more times during February 2018.

Note that many of the camera traps were not continuously operational during the entire 3.5-year period covered by the data. Therefore, certain physical locations are overrepresented relative to others.

Analysis

We'll begin by visualizing the data. First, we look to the header graphs inside the data table to get a quick idea of which species were observed more or less frequently. A subset of header graphs is shown in Exhibit 1. We can see, for example, that bats were far less frequently observed than bears. Other infrequently observed species include javelina, raccoon, raptor, ringtail, and rodent. Deer are the most frequently observed species in the data. The header graph for Site shows us that we have many more rows of data for the Chiricahua mountains than the Dos Cabezas mountains. This is due to more camera traps being set in the Chiricahua mountains.

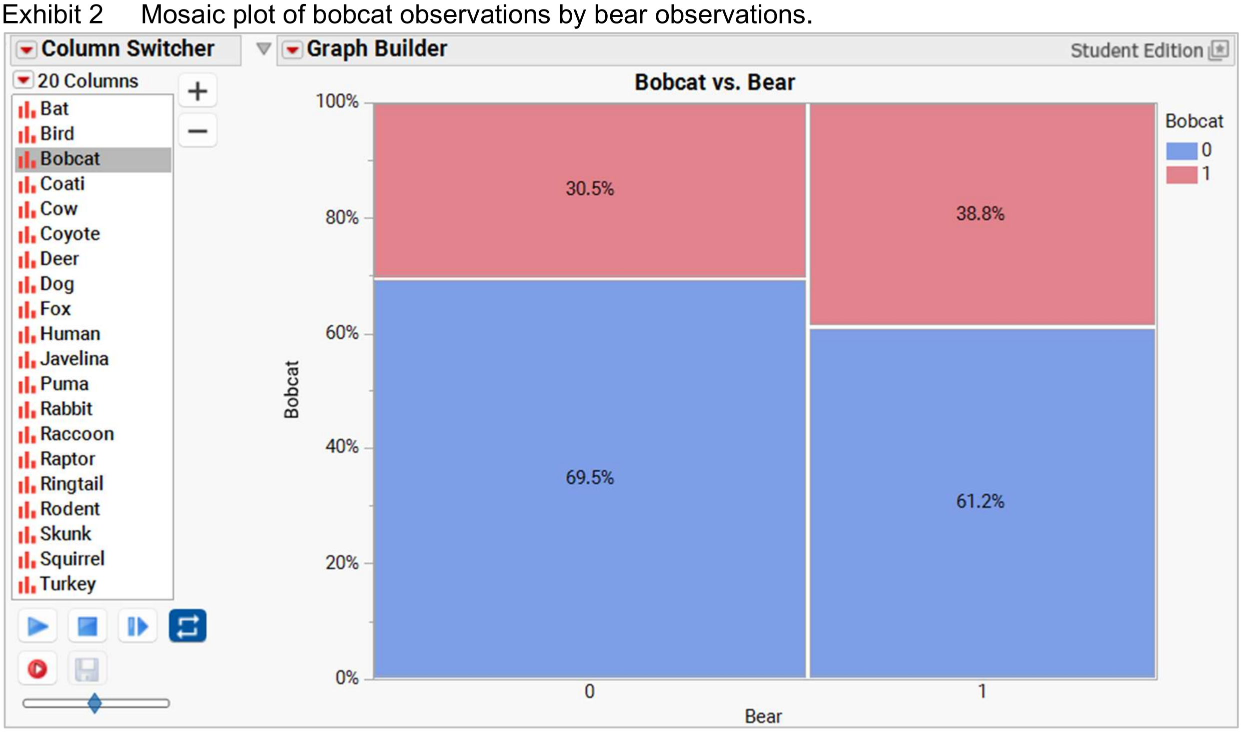
Exhibit 1 Header graphs showing relative proportions of the observations of different species



(Click the histogram icon in the top left of the data table to reveal the header graphs.)

We next visualize associations between observations of black bears and other animals. The mosaic plot in Exhibit 2 depicts the association between black bears and bobcats. The left half of the plot represents rows in which black bears were not observed, and in the red area we see that bobcats were observed in 30.5% of these cases. On the right half of the plot, representing when black bears were observed, bobcats were observed 38.8% of the time, an 8.3 percentage point increase. This suggests that observation of a black bear is associated with an increased probability of also observing a bobcat. This association could be due to habitat overlap or diet, though this may be unlikely because, unlike black bears, bobcats do not consume much plant material. Using the Column Switcher, we cycle through other species in place of bobcats (not depicted in Exhibit 2) and discover an even larger association between black bears and skunks, as well as an opposite association with ringtails, a member of the raccoon family, in which observing a black bear is associated with a decreased probability of observing a ringtail.

Exhibit 2 Mosaic plot of bobcat observations by bear observations.



(Go to Graph > Graph Builder. Drag Bear to the x-axis and Bobcat to the y-axis, and select the mosaic element in the top ribbon of elements. In the Mosaic options in the bottom left, set Cell Labeling to Label by Percent. To invoke the Column Switcher, go to the Graph Builder red triangle and select Redo > Column Switcher. Select Bobcat in the top list, and then select all available animal columns in the bottom list. Click OK to add the Column Switcher to the graph, and then click on any animal column in the list to update the graph automatically.)

We find several notable numerical associations in these mosaic plots, but it is possible that any of these associations are due to random chance and do not reflect evidence of true associations in the population of animals living in these mountains. For example, the numerical association between black bears and bobcats may simply be due to random chance in which animals happened to wander by certain camera traps at certain times during the study, and across the entire populations of black bears and bobcats there is no association. Ultimately, we need to determine if the numerical associations we see constitute evidence of real association in the population, and for this, we must conduct statistical tests.

Contingency Analysis

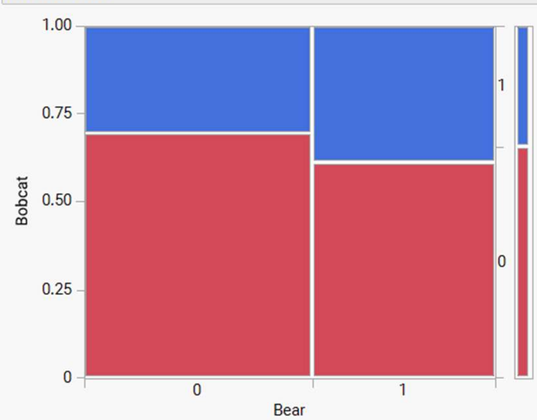
We are interested in inferring whether, in the population of animals in the region of our study, the observation of black bears is associated with increased or decreased observation of other types of animals. To address this goal, we'll perform a series of *contingency analyses* between our nominal black bear variable and the other nominal animal variables in the data set. The analysis will produce summary statistics, such as the percentages in our previous mosaic plot, as well as statistical tests that help us determine whether any numerical associations constitute evidence of a true effect in the population. In these tests, the null hypothesis is that, in the full population, there is no association between the two variables (here, no association between the observation of black bears and some other animal). A p-value below a pre-specified cutoff (*alpha*, typically 0.05) is taken as evidence that the null hypothesis is false and that the population does exhibit an association.

We'll start by analyzing the association between black bears and bobcats. Exhibit 3 presents the output of the contingency analysis. It includes a mosaic plot and a contingency table summarizing the data numerically, with a key for interpreting the values in each cell provided in the top left corner. We see the Row % values corresponding to the percentage values overlaid in the mosaic plot in Exhibit 2.

Exhibit 3 Results of the bobcat-by-bear contingency analysis.

Contingency Analysis of Bobcat By Bear Student Edition

Mosaic Plot



Tests

N	DF	-LogLike	RSquare (U)
632	1	2.3869677	0.0059

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	4.774	0.0289*
Pearson	4.786	0.0287*

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	0.9884	Prob(Bobcat=1) is greater for Bear=0 than 1
Right	0.0178*	Prob(Bobcat=1) is greater for Bear=1 than 0
2-Tail	0.0348*	Prob(Bobcat=1) is different across Bear

Risk Difference

Description	Risk Difference	Lower 95%	Upper 95%
P(1|1)-P(1|0)	0.083057	0.008457	0.157657

Contingency Table

Bear	Count / Total % / Col % / Row %	Bobcat 0	Bobcat 1	Total
0	Count	244	107	351
	Total %	38.61	16.93	55.54
	Col %	58.65	49.54	
	Row %	69.52	30.48	
1	Count	172	109	281
	Total %	27.22	17.25	44.46
	Col %	41.35	50.46	
	Row %	61.21	38.79	
Total	Count	416	216	632
	Total %	65.82	34.18	

(Go to Analyze > Fit Y by X. Place Bobcat in the Y, Response role and Bear in the X, Factor role and click OK. Under the red triangle, select Risk Difference.)

The report also provides results of three statistical tests of association between the two variables: the Likelihood Ratio and Pearson chi square tests and Fisher's exact test. Each produces a p-value below 0.05, highlighted in red. (We're considering only the 2-tail p-value for Fisher's Exact Test.) This result provides strong enough statistical evidence to reject the hypothesis that observations of bears and bobcats are not associated.

The Risk Difference table shows us that observation of a black bear is associated with an increase of 0.083 in the probability of observing a bobcat, with a confidence interval of 0.008457 to 0.157657. Note that 0.083 corresponds to the percentage point difference from Exhibit 2. (Also note that further methods of quantifying the size of this association, including relative risk and odds ratios, are available under the red triangle.) This result suggests that black bears and bobcats co-occur more than we would expect from random chance alone.

Multiple Comparisons and False Discovery Rate Control

After testing the association between black bears and bobcats, we have 19 additional tests to conduct, making 20 tests in total. Challenges arise when conducting a large group or *family* of hypothesis tests, (aka *multiple comparisons*). Put simply, conducting many hypothesis tests increases the probability of obtaining false positive results (i.e., Type I errors), believing we have evidence of an association in the entire population when no such association actually exists. False positives can have major consequences; for example, if we come to believe, falsely, that the presence of black bears is associated with a decrease in the presence of a protected species, we may undertake further research to identify and address the cause of this (nonexistent) association, which would be a waste of resources and potentially cause ecological harm with no commensurate benefit. More generally, if we arrive at incorrect conclusions, this can impact wildlife management decisions and to policies that are not truly supported by the data.

When conducting a single hypothesis test, the value of alpha – our p-value cutoff – is the probability of obtaining a false positive result when the null hypothesis (i.e., no association) is actually true. When conducting a family of tests, the *family wise error rate* (FWER) is conceptually like alpha: it is the probability, assuming the null hypothesis is true in all tests, of obtaining one or more false positive results in the family of tests. The equation for calculating FWER is:

$$FWER = 1 - (1 - \alpha)^m$$

- α is the alpha level for individual tests
- m is the number of tests conducted

For a family of 20 tests with an alpha of 0.05, the FWER is 0.64, meaning that if black bears actually are not associated with any of the other animals, we nevertheless have a probability of 0.64 of obtaining one or more false positive results across the 20 tests. This risk of false positives is unacceptably high, so we must take some measure to reduce it.

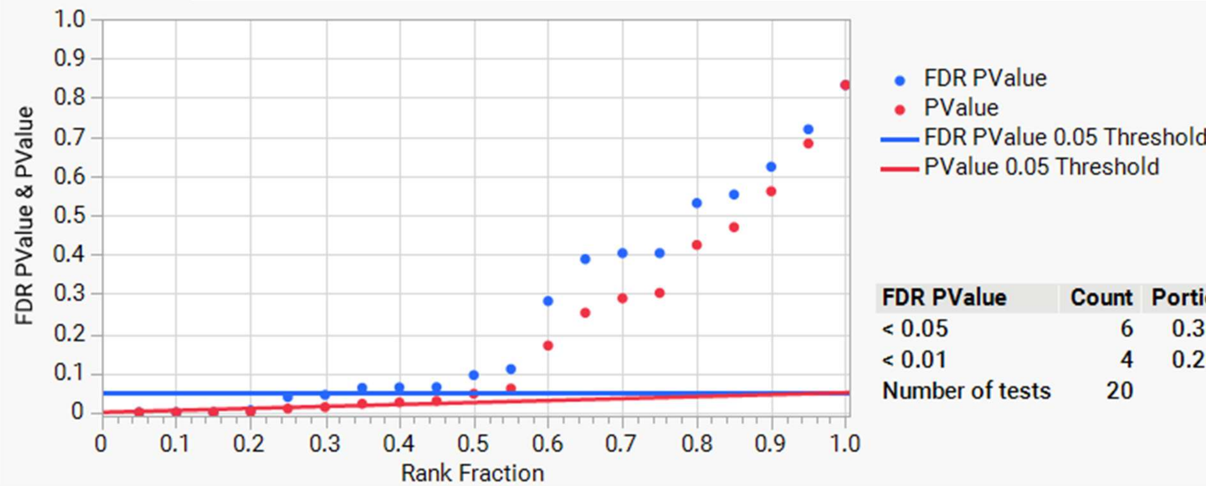
There are numerous methods for guarding against false positives. Here, we use Benjamini-Hochberg False Discovery Rate (FDR) control, which adjusts p-values upward to limit the proportion of false positive results among all positive tests. For example, if we control FDR at the 0.05 level and obtain 20 positive test results, we expect that only one of those positive results (0.05 x 20 = 1) to be a false positive. More conservative methods exist for reducing false positives, but these generally come at the cost of increasing the rate of false negative results (i.e., Type II errors). FDR control represents a tradeoff between guarding against both false positives and false negatives.

Exhibit 4 presents the results of performing our 20 contingency analyses while implementing FDR control. This report was produced using JMP's Response Screening platform. The FDR PValue Plot presents two Likelihood Ratio p-values for each test: red values are unadjusted (as in Fit Y by X), and blue values have had FDR adjustment applied. Any point that falls below the line of the same color indicates a statistically significant result at the 0.05 level after FDR control. The table to the right indicates that 6 tests are below an FDR p-value of 0.05. The table below presents numerical results of each test, and we see that the 6 with FDR p-values below 0.05 are squirrel, bird, skunk, ringtail, coati, and raptor. Notably, several tests show unadjusted p-values below 0.05 but FDR p-values above 0.05: bat, human, bobcat, and puma. FDR control therefore have eliminated 4 tests that had positive results before the control was applied. To guard against false positives, we conclude that we do not have sufficient evidence of association between black bears and these 4 animals.

Exhibit 4 Response Screening report of 20 contingency analyses with FDR control.

Response Screening

FDR PValue Plot | FDR Logworth By Effect Size



FDR PValue	Count	Portion
< 0.05	6	0.300
< 0.01	4	0.200
Number of tests	20	

Result Table | 2 by M Results

X=Bear

Y	Count	PValue	Logworth	FDR PValue	FDR Logworth	Effect Size
Squirrel	632	2.02e-8	7.69503	4.036e-7	6.39400	0.2231
Bird	632	2.07e-6	5.68383	2.071e-5	4.68383	0.18892
Skunk	632	3.34e-6	5.47634	2.226e-5	4.65243	0.18481
Ringtail	632	0.00112	2.95237	0.00558	2.25340	0.12609
Coati	632	0.0097	2.01313	0.038809	1.41107	0.1031
Raptor	632	0.01333	1.87532	0.044418	1.35245	0.09817
Bat	632	0.02179	1.66172	0.062261	1.20578	0.09059
Human	632	0.02542	1.59480	0.063553	1.19686	0.08886
Bobcat	632	0.02889	1.53920	0.064208	1.19241	0.08702
Puma	632	0.04737	1.32453	0.094733	1.02350	0.07905
Turkey	632	0.06059	1.21758	0.110169	0.95794	0.07478
Rabbit	632	0.16999	0.76957	0.28332	0.54772	0.05469
Cow	632	0.25356	0.59593	0.390087	0.40884	0.04546
Deer	632	0.30402	0.51709	0.405364	0.39216	0.04096
Fox	632	0.29037	0.53705	0.405364	0.39216	0.04207
Coyote	632	0.42614	0.37044	0.532679	0.27353	0.0317
Dog	632	0.47136	0.32665	0.554539	0.25607	0.0286
Raccoon	632	0.56285	0.24961	0.62539	0.20385	0.02308
Javelina	632	0.68451	0.16462	0.720537	0.14234	0.01613
Rodent	632	0.83343	0.07913	0.833428	0.07913	0.00835

(Go to Analyze > Screening > Response Screening. Enter Bear as the X and all other species as the Y, Response. Click OK.)

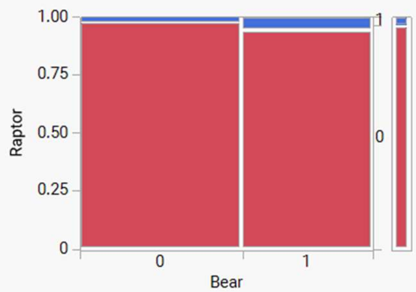
From here, we inspect the full details of the positive tests. To illustrate, exhibit 5 presents the contingency analysis results for observations of black bears and raptors. From the mosaic plot and Risk Difference table, we see that observation of a black bear is associated with an increase of 0.037 in the probability of observing a raptor. Raptors are not frequently observed in general – the blue areas in the mosaic plot are very small – so this seemingly minor increase may be considered practically important. (More advanced measures of the association that we have not covered here, such as the odds ratio, would confirm this.) The first exercise will ask you to interpret the result of another of the 5 remaining positive tests.

Exhibit 5 Results of raptor-by-bear contingency analysis.

Fit Selected Items

Contingency Analysis of Raptor By Bear Student Edition

Mosaic Plot



Contingency Table

Raptor

Count / Total % / Col % / Row %	Bear	0	1	Total
0	Count	344	7	351
	Total %	54.43	1.11	55.54
	Col %	56.49	30.43	
	Row %	98.01	1.99	
1	Count	265	16	281
	Total %	41.93	2.53	44.46
	Col %	43.51	69.57	
	Row %	94.31	5.69	
Total	Count	609	23	632
	Total %	96.36	3.64	

Tests

N	DF	-LogLike	RSquare (U)
632	1	3.0627032	0.0310

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	6.125	0.0133*
Pearson	6.091	0.0136*

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	0.9965	Prob(Raptor=1) is greater for Bear=0 than 1
Right	0.0121*	Prob(Raptor=1) is greater for Bear=1 than 0
2-Tail	0.0176*	Prob(Raptor=1) is different across Bear

Risk Difference

Description	Risk Difference	Lower 95%	Upper 95%
P(1|1)-P(1|0)	0.036996	0.006207	0.067786

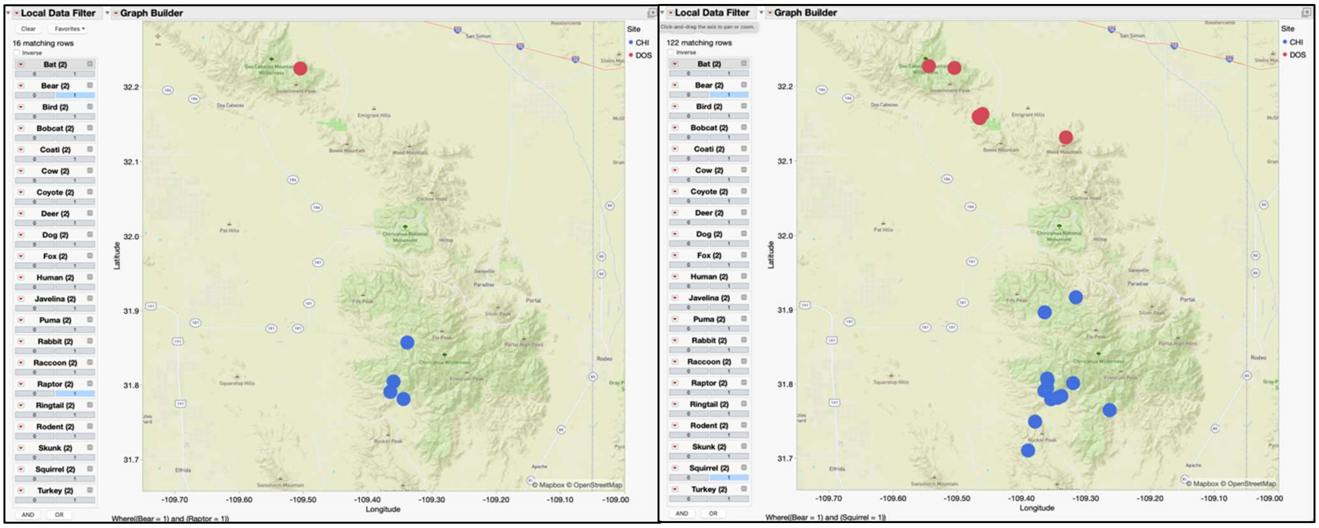
(Inside the Response Screening report, select the rows of the Result Table with FDR p-values less than 0.05. Under the red triangle, select Fit Selected.)

Geographic Mapping of the Observations

We have identified associations between observation of black bears and 6 other species. The analysis aggregated data from many camera traps spread across two mountain ranges in Arizona, and we may be interested in where in those mountain ranges observations of species co-occurrence were made. For example, at which camera traps are black bears and raptors co-occurrences being observed? Or black bears and squirrels?

Exhibit 6 presents geographic maps showing the camera traps that observed both black bears and raptors (left) or black bears and squirrels (right) within the same calendar month. Chiricahua locations are in blue and Dos Cabezas in red. We see on the left that only 5 camera traps among the many in the data set observed both black bears and raptors, including just one in the Dos Cabezas mountains. This contrasts with the map on the right, which shows many more cameras having observed both black bears and squirrels. This is reasonable, given that observations of squirrels were much more frequent than observations of raptors in the data set.

Exhibit 6 Maps showing bear and raptor sightings (left) and bear and squirrel sightings (right)



(Go to Graph > Graph Builder. Drag Longitude to the x-axis, latitude to the y-axis, and Site to the color role. Deselect the smoother element in the top ribbon of elements. Right click inside the graph and select Graph > Background Map. Select Street Map Service, and in the URL drop-down menu, select Mapbox Outdoors. Go to the Graph Builder red triangle and select Local Data Filter. Select all 20 animal columns and click the plus sign. Select the value of 1 under an animal to filter down to only those camera traps that observed that animal. For example, select 1 under both Bear and Raptor to produce the left graph.)

Summary

Statistical insights

We conducted 20 hypothesis tests to determine which types of animals co-occur with black bears in the same place and at the same time of year. While unadjusted contingency analyses would have produced 10 animals with positive results, the probability of false positive results in a family of 20 tests was too great. We implemented false discovery rate (FDR) control to reduce the probability of false positive results, leading to positive results with respect to 6 animals: squirrel, bird, skunk, ringtail, coati, and raptor.

Implications

Identifying these associations provides clues as to which species have interactions or other interdependencies with black bears in the region under study. Further research would be necessary to establish exactly why these associations exist. The current study represents an exploratory first step.

JMP features and hints

Fit Y by X performs contingency analysis when categorical variables are entered in both the X and Y roles. The Response Screening platform performs families of hypothesis tests with FDR control. Graph Builder has geographic mapping capabilities. The Column Switcher and Local Data Filter provide quick was of automatically updating graphs and analyses.

Exercises

1. The analysis in Exhibit 4 identified 6 animals showing evidence of an association with black bears, and we examined the association between black bears and raptors in Exhibit 5. Using the discussion of Exhibit 5 as a guideline, follow up on the statistically significant result with respect to ringtails. Are ringtails more or less likely to be observed when a black bear is observed? What is the Risk Difference value, and what does it tell us?

2. If we performed a hypothesis test on every possible pair of animals in this data set, we would conduct 210 tests. What would the family-wise error rate be for this family of tests? Without FDR control, how likely would we be to obtain at least one false positive result in this family of 210 tests?

3. If we used FDR control at the 0.10 level and obtained 30 positive test results, how many of these results would we expect to be false positives? Is this higher or lower than the number of false positives we'd expect if we used FDR control at the 0.05 level?

4. We focused on black bears, but a researcher may focus on another animal. Choose another animal and perform a family of tests as we did with black bears, answering the following questions:

    a. How many tests show statistically significant results at the 0.05 level *before* FDR control?

    b. How many tests show statistically significant results at the 0.05 level *after* FDR control?

    c. For the test with the smallest FDR-adjusted p-value, follow up on the test as you did for Exercise 1 to explain the association that you've discovered.