CO PTAKI MOGĄ POWIEDZIEĆ O ZANIECZYSZCZENIACH? ANALIZA STATYSTYCZNA DANYCH ŚRODOWISKOWYCH.

Łukasz J. Binkowski, Uniwersytet Komisji Edukacji Narodowej w Krakowie

Biomonitoring to jedna z metod oceny zanieczyszczenia środowiska, w której mierzy się zanieczyszczenie w próbkach pobranych od organizmów żywych (biomonitorów) [1]. Zmierzone w ten sposób stężenia niosą więcej informacji niż dane pochodzące bezpośrednio z komponentów środowiska (jak np. stężenia zanieczyszczeń w powietrzu, czy wodzie) bo uwzględniają tzw. biodostępność i wydajność barier, które każdy organizm posiada. Dane biomonitoringowe znajdują szerokie zastosowanie, ale wśród nich jedno z najważniejszych to wykorzystanie w ekotoksykologii. Charakteryzuje je kompleksowość i zależność od wielu czynników zewnętrznych. Oprócz informacji o środowisku i populacji są to najczęściej stężenia różnych substancji (jak np. metali czy pestycydów), które pochodzą z zaawansowanych protokołów laboratoryjnych, za którymi zawsze stoją ograniczenia techniczne [2]. Dlatego mimo to, że poszczególne zmienne w analizie statystycznej wielu projektów ekotoksykologicznych wydają się stosunkowo łatwe do zrozumienia to już sama ich analiza i wnioskowanie są dość specyficzne i dodatkowo wielowymiarowe, co stwarza tzw. wysoki próg wejścia dla początkujących badaczy. W praktyce są to jednak metody intuicyjne, które pozwalają wyciągać konkretne wnioski, więc warto je poznać.

Prześledźmy zatem drogę statystyczną wyników w projekcie ekotoksykologicznym i przeanalizujmy dylematy badaczy, które prawie zawsze w tego typu projektach się pojawiają. Przykładem niech będą badania zanieczyszczenia środowiska oparte o ptaki. To zwierzęta bardzo często badane w aspekcie zanieczyszczeń, które odegrały ogromną rolę

Copyright © StatSoft Polska 2025, info@StatSoft.pl

badaniach ekotoksykologicznych w czasach stosowania *DDT* (najbardziej znanego pestycydu na świecie wycofanego z użycia w krajach rozwiniętych kilkadziesiąt lat temu). To dzięki obserwacji ptaków zdiagnozowano problem wpływu pestycydów na tzw. *non-target species* (gatunki, w które nie celowano użycia tych związków). Była to jedna z największych katastrof ekologicznych, którą doskonale opisuje książka uznawana za jedną z najbardziej wpływowych pozycji popularno-naukowych XX wieku – *Silent Spring* [3].

Ptaki w ekotoksykologii

Ptaki, z uwagi na to, że są zwierzętami powszechnie występującymi, zajmującymi różne nisze środowiskowe, uznawane są za bardzo dobre biomonitory zanieczyszczenia środowiska. Zanieczyszczenia, które akumulują są efektem ich pobierania różnymi drogami np. pokarmową i oddechową. Łatwość wnikania oraz kumulowanie się wielu substancji u ptaków zagraża gatunkom znajdującym się na różnych poziomach łańcucha pokarmowego, choć na pewno najbardziej obciążone są gatunki drapieżne. Przykładowo Kucharska [4] i Binkowski [5] w swoich badaniach wykazali, że najwyższe stężenia rtęci (Hg; głównie w postaci formy organicznej - metylortęci) notuje się u typowych ptaków rybożernych. Najniższe stężenia notuje się zaś u gatunków, których dieta składa się przede wszystkim z materiału roślinnego. Warto dodać, że rtęć jest pierwiastkiem niezwykle interesującym dla ekotoksykologów ze względu na swój podwójny charakter – z jednej strony stanowi metal zanieczyszczający środowisko w formie nieorganicznej, z drugiej zaś, w postaci metylortęci, zachowuje się w organizmach jak związek organiczny.

Układ doświadczalny

W badanich ekotoksykologicznych z udzialem ptaków, ale też i innych zwierząt, zazwyczaj mamy do czynienia z czterema grupami zmiennych. Pierwsza to oczywiście zmienne główne, jak stężenia wybranych substancji w organizmie (w języku statystycznym zwane zmiennymi zależnymi). W danym projekcie może być ich kilka albo zaledwie jedna, ale

Copyright © StatSoft Polska 2025, info@StatSoft.pl

nawet, gdy jest ich kilka to zazwyczaj w początkowej części analiz traktuje się je osobno, jedną po drugiej, i dopiero na dalszych etapach analizy szuka się pomiędzy nimi powiązania. Oczywiście nie jest to reguła, bo czasami od samego początku prowadzi się analizy wielowymiarowe, które agregują jednocześnie informacje z kilku zmiennych. Druga grupa to czynniki ustalone, które klasyfikują badane obiekty. Zaliczamy do nich np. gatunek, płeć, czy teren badawczy. Warto napomknąć, że poziomy tych czynników (np. dla płci poziomy to samce i samice) są skończone i je znamy. Trzecią grupę stanowią dodatkowe zmienne ciągłe (podobne w charakterze do zmiennych zależnych), które mogą posłużyć np. w modelowaniu statystycznym do wytłumaczenia (opisania) zmienności zmiennej zależnej. Dobrym przykładem jest też tutaj zmienna towarzysząca (ang. *covariate*) czy zmienne używane w analizie korelacji. Jest jeszcze czwarta grupa zmiennych, która nie występuje zawsze (lub nie zawsze można pozyskać dla niej dane) – czynniki losowe, które zazwyczaj wpływają na zmienne zależne, ale nie jesteśmy w stanie przewidzieć ich wszystkich poziomów albo nie zależy nam na poznaniu wpływu takiej zmiennej, ale chcemy, żeby model statystyczny go ujmował i dzięki temu prezentował dokładniejsze oszacowanie. Do tej grupy należą m.in. informacje dotyczące par wiązanych, zagnieżdżenia czynników, czy nawet proste informacje mówiące o tym, który badacz dokonywał danego pomiaru. To właśnie czynniki losowe są odpowiedzialne za potrzebę stosowania w wielu badaniach środowiskowych tzw. modeli mieszanych, które są rozwinięciem uogólnionych modeli liniowych (z ang. *General Linear Models* - *GLM*) – utartego standardu w dzisiejszych analizach statystycznych wyników środowiskowych.

Biorąc pod uwagę mnogość możliwości, które dają nam wspomniane typy zmiennych, łatwo się domyślić, że badania ekotoksykologiczne ptaków zazwyczaj różnią się od siebie w zakresie, zmiennych, czynników, pytań i hipotez. Dlatego w dalszej częśći posłużymy się przykładem projektu dotyczącego badań metali u różnych gatunków mew z kanału La Manche. Badania prowadzono na dwóch odrębnych terenach różniących się znacząco stopniem industrializacji, badano wiele pierwiastków, w tym m.in. rtęć. Dzięki zmiennym dodatkowym, jak np. poziomy stabilnych izotopów węgla, azotu i siarki we wnioskowaniu wprowadzono aspekty troficzne, żeby lepiej wytłumaczyć stwierdzone stężeniach [5, 6].

Copyright © StatSoft Polska 2025, info@StatSoft.pl

Od czego zacząć analizę

W każdym projekcie dla osoby prowadzącej analizę statystyczną zrozumienie układu doświadczalnego i poznanie poszczególnych zmiennych jest kluczowe. Zawsze należy przejrzeć dane, poznać poziomy czynników, czy ogólne zakresy zmiennych, zanim przystąpi się do właściej analizy. W poznaniu i analizie układu doświadczalnego świetnie sprawdzają się tabele przekrojowe (z ang. *cross tables*), a w poznaniu zakresu wartości prawdopodobnie najlepsze będą wykresy pudełkowe (z ang. *box-plots*).

Wykresy pudełkowe domyślnie prezentują wszystkie kwartyle z podzialem na: wąsy (kwartyl 0 i 100, czyli wartości minimalna i maksymalna), pudełko (dolny i górny kwartyl) oraz punkt środkowy (mediana). Warto dodatkowo, szczególnie w przypadku mniej licznych zbiorowości, umieścić na wykresie dane surowe. Dzięki temu łatwo zauważyć wartości odstające oraz charakter rozkładu danych.

Granice raportowania

Dane dotyczące stężęń pozyskuje się zazwyczaj z użyciem instrumentalnych technik laboratoryjnych, które najczęściej należą do tzw. metod względnych. Kluczowe znaczenie dla nich ma kalibracja, czyli ustalenie zależności pomiędzy sygnałem detektora aparatu a stężeniem danej substancji. Najczęściej do kalibracji stosuje się krzywą kalibracyjną, z którą wiążą się dodatkowe parametry — przede wszystkim granice wykrywalności (ang. *limit of detection, LoD*) oraz oznaczalności (ang. *limit of quantification, LoQ*). Określają one minimalne stężenie, które można odpowiednio wykryć lub oznaczyć. Zgodnie z zasadami systemów jakości obowiązujących w laboratoriach, wyniki poniżej tych granic należy raportować jako wartości niższe od odpowiedniego progu (np. <LoD lub <0,1 µg/L). Jednak to, co dla laborantów wydaje się oczywiste, dla badaczy jest problematyczne — zwłaszcza w kontekście analiz statystycznych, bo nie da się np. wyliczyć konwencjonalnie średniej z wyników zawierających wartość <LoD.

Copyright © StatSoft Polska 2025, info@StatSoft.pl

Ponieważ w ekotoksykologii dane zawierające wartości poniżej granic są powszechne, na przestrzeni lat powstało kilka sposobów radzenia sobie z tym problemem. Najczęściej stosowaną metodą jest tzw. substytucja, czyli zastąpienie wyników poniżej granicy arbitralnie przyjętą lub losowo wygenerowaną liczbą. Choć metoda ta ma pewne zalety, jest często krytykowana za wprowadzanie sztucznego trendu do danych. Obecnie za lepsze podejścia uznaje się dwie grupy rozwiązań: proste analizy nieparametryczne oraz metody zaadaptowane z analiz przeżycia. Te drugie dają znacznie większe możliwości analizy, ale niewiele popularnych programów statystycznych ma wbudowane narzędzia do tego typu analiz. Znakomicie sprawdza się tu język R, który oferuje dedykowane pakiety, takie jak np. NADA [7]. W programie Statistica, również można przeprowadzić część takich analiz, jednak zwykle wymaga to wcześniejszego przekształcenia danych i dodatkowych obliczeń.

Którejkolwiek metody postępowania z danymi poniżej granicy raprotowania nie użyjemy, musimy o tym pamiętać w dalszej analizie. Będzie to determinowało czasami wybór metod statystycznych albo przynajmniej będzie musiało być uwzględnione w opisie wyników.

Rozkład danych

Znajomość rozkładu danych jest również bardzo ważna, bo nie tylko determinuje możliwe do zastosowania testy statystyczne, ale również, o czym wiele osób zapomina, określa, które statystyki opisowe najlepiej przedstawią zmienność i tendencje centralne danej zbiorowości. W przypadku stężeń pierwiastków, poza nielicznymi wyjątkami, mamy prawie zawsze do czynienia z rozkładem log-normalnym (to taki, który po zlogarytmowaniu staje się rozkładem normalnym). W takim przypadku średnia arytmetyczna, która nie jest odporna na wartości skrajne, będzie sztucznie podwyższona i znacznie lepiej sprawdzą się średnia geometryczna, czy mediana.

Warto tutaj nadmienić, że czasami w analizie, szczególnie gdy liczebność jest niewielka (mniej niż kilkadziesiąt obserwacji) prawidłowa identyfikacja rozkładu na podstawie badanej populacji jest trudna i często prowadzi do błędnych wniosków. Wynika to wyłącznie

Copyright © StatSoft Polska 2025, info@StatSoft.pl

z liczebności, a żeby tego dowieść wystarczy, że czytelnik wylosuje z rozkładu normalnego kilka razy próbkę o niewielkich liczebnościach. Będziemy wiedzieć, że będą to próbki pochodzące z rozkładu normalnego, ale patrząc na ryciny będziemy mieć spore wątpliwości, czy na pewno jest to rozkład normalny. Wątpliwości te będą gasnąć wraz ze wzrostem liczebności badanej populacji. Dlatego w ekotoksykologii często rozkładu się uparcie nie identyfikuje (właśnie przy niewielkich populacjach) a zakłada, bazując na danych literaturowych bądź doświadczeniu badacza. Oczywiście nie zwalnia to analityka z potrzeby przyjrzenia się rozkładkowi i próbie dopasowania rozkładu teporetycznego. Robi się to wielotorowo, czasami nawet oceniając zgodność między średnią a medianą (która jest zachowana tylko w rozkładach symetrycznych), czy analizując kurtozę i skośność. Niezmiennie jednak najlepszym podejściem będzie tzw. inspekcja wizualna. Tradycyjnie stosuje się do tego histogramy, ale program Statistica ma tutaj wiele innych pomocnych narzędzi, tak jak i środowisko R, w którym godnym polecenia pakietem jest fitdistrplus [8]. Pozwoli on nie tylko ocenić zgodność rozkładu z teoretycznym rozkładem normalnym, ale również z innymi typami rozkładów, jak log-normalny, gamma, czy jednostajny.

Główne wnioskowanie

Po ujęciu granic raportowania i identyfiakcji rozkładu, a przede wszystkim po ustaleniu celu projektu, przychodzi pora na główną analizę statystyczną. Jeszcze 10-15 lat temu większość raportów z badań prezentowała statystyki opisowe, wyniki testów istotności różnic i analizę korelacji. Było to poprawne podejście, ale uległo zmianie z uwagi na powszechny dostęp do zaawansowanego oprogramowania statystycznego. Dzisiaj testy istotności różnic raczej nie są już stosowane jako główne narzędzie do wnioskowania w projektach ekotoksykologicznych. Nadal często się z nich korzysta, ale głównie do celów pomocniczych, np. przy łączeniu grup na potrzeby dalszych analiz. Ich podstawową rolę przejęło modelowanie statystyczne z *GLM* na czele, które nie spłaszcza wnioskowania i daje możliwość bardziej szczegółowego opisu danych. To jednak, co stanowi podstawę tych rozwiązań w pierwszym kontakcie nie podoba się osobom, które pracowały wcześniej

Copyright © StatSoft Polska 2025, info@StatSoft.pl

wyłącznie z testami istatoności różnić. Mianowicie, w wynikach *GLM* to nie wartość prawdopodobieństwa p jest najważniejsza. Co więcej, w zasadzie zaleca się aby zupełnie odchodzić od jej użycia we wnioskowaniu (to bardziej ogólny trend nie skierowany w samą wartość p, a istotę zero-jedynkowego charakteru oceny istotności różnic) [9]. Wtedy nie mówimy, czy np. płeć wpłynęła istotnie na poziom Hg w krwi ptaków, ale jesteśmy w stanie oszacować jakie będzie stężenie Hg u samic, a jakie u samców – to tzw. szacowanie (modelowanie) zmiennej. Takie podejście na pewno wymaga więcej pracy przy interpretacji wyników, ale daje znacznie większe możliwości wyciągania wniosków.

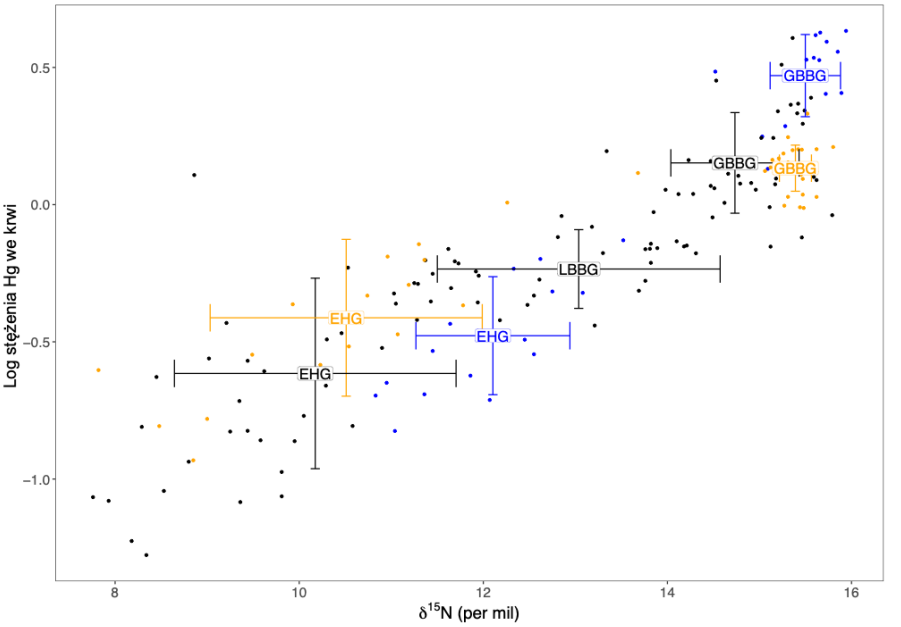
Jak wcześniej wspomnianio, analizy *GLM* można, a czasami nawet trzeba rozbudować do modeli mieszanych (*GLMM*), gdzie dochodzą czynniki losowe. Dodatkowo w analizach *GLM* należy sprawdzić, czy pomiędzy zmiennymi, które ujmujemy w modelu nie zachodzą silne związki (tzw. współliniowość), jaki jest rozkład danych, itp. Praktycznie nie ma tutaj więcej komplikacji niż przy klasycznej analizie wariancji, która *nota bene* powinna być traktowana jako uproszony model *GLM*. Dlatego jeżeli czytelnik jeszcze nie miał okazji pracować z *GLM*, nie powiniem dłużej czekać, bo jest to relatywnie proste w użyciu, a diametralnie zmienia wymiar wnioskowania i jego możliwości.

Oczywiście obok zaawansowanego modelowania czasami w pracach należy pokazać najprostsze analizy. Dobrym przykładem jest analiza korelacji i to nie w postaci współczynników korelacji, a zwykłego wykresu rozrzutu. Na rycinie 1 pochodzącej z naszego przykładowego projektu, oparto jeden z jego najważniejszych wniosków – stwierdzenie wzrostu stężenia rtęci wzdłuż łańcucha troficznego, czyli tzw. biomagnifikację. Czasami więc siłą wniosków tkwi w prostej analizie i zwykłej obserwacji zmiennych i zależności pomiędzyu nimi.

Obecnie w badaniach ekotoksykologicznych mierzy się wiele zmiennych, które potem albo możemy analizować jedna po drugiej, albo zastosować podejście wielowymiarowe. Tutaj prawie zawsze doskonale sprawdza się analiza składowych głównych (z ang. *Principal Component Analysis – PCA*). Możemy jej użyć do badania zależności pomiędzy zmiennymi albo do tzw. redukcji wymiarów stosując składową głowną jako reprezentację wielu

Copyright © StatSoft Polska 2025, info@StatSoft.pl

zmiennych. To, w połączeniu z analizami *GLM* pozwala ująć w modelowaniu wiele zmiennych, które w klasycznym podejściu byłyby niemożliwe do zastosowania. Podobnie jak w przypadku *GLM*, *PCA* nie jest to trudne, a czasami pozwala uprościć prezentację zestawu wykresów i współczynników korelacji do jednej ryciny.



Rycina 1. Zależność między dwiema zmiennymi prezentująca tendencję centralną, rozrzut oraz dane surowe (etykiety tekstowe prezentują różne gatunki, a kolory różne tereny badawcze). W aspekcie merytorycznym rycina przedstawie biomagnifikację rtęci u ptaków morskich.

Podsumowanie

Jak przestawiono w artykule powszechnie stosowane metody pracy w projektach ekotoksykologicznych są logiczne i przyjemne w stosowaniu. W związku z rozwojem i łatwym dostępem do odpowiedniego oprogramowania (Statistica i język R) każdy z nas może wdrożyć te metody do swojego warsztatu badawczego, co pozytywnie odbije się na wyciąganych wnioskach i zakresach projektów badawczych. A teraz podsumowując wróćmy do naszego przykładu badań rtęci i innych pierwiastków u ptaków z wybrzeży Francji.

Copyright © StatSoft Polska 2025, info@StatSoft.pl

Dzięki poprawnemu podeściu do rozkładu, zastosowaniu modelowania statystycznego, analiz korelacji i analiz wielowymiarowych Zespół, który prowadził te badania dowiedział się wielu ciekawych rzeczy, które nie tylko zostały opublikowane, ale posłużyły do zaplanowania kolejnych projektów. Wszystkie analizy, które zostały wykonane prowadzono w Statistica i R. A oto te najważniejsze wnioski, które wyniknęły z tych badań.

W badaniach przeanalizowano czynniki ekologiczne wpływające na akumulację pierwiastków śladowych, ze szczególnym uwzględnieniem rtęci, u piskląt trzech sympatrycznych gatunków mew gniazdujących na francuskim wybrzeżu Kanału La Manche. Najwyższe stężenia rtęci stwierdzono u gatunku rybożernego, a najniższe u oportunisty pokarmowego. Osobniki zajmujące wyższą pozycję troficzną (szacowaną na podstawie $\delta^{15}N$) i żerujące w środowisku morskim (szacowane na podstawie $\delta^{13}C$ i $\delta^{34}S$) wykazywały większą akumulację Hg, co wskazuje na biomagnifikację rtęci w łańcuchu pokarmowym. W przypadku szerszego zestawu pierwiastków stężenia układały się od najniższych dla Co do najwyższych dla Fe (Co < Cd < Ni < Mn ≤ Pb < Cr < Hg < Cu < Se < As < Zn < Fe); Ag i V zawsze były poniżej granicy oznaczalności, a dla Cd, Cr i Ni odsetek wyników nieoznaczonych przekraczał 30%. Wzorce stężeń różniły się między gatunkami i lokalizacjami; dodatkowo zaobserwowano m.in. biomagnifikację As, biodylucję Pb oraz podwyższone poziomy Pb w uprzemysłowionym estuarium Sekwany. U części osobników stosunek Se:Hg był niższy niż 4, co oznacza, że selen może nie zapewniać u nich skutecznej ochrony przed toksycznością rtęci. Wyniki podkreślają, że ekologia troficzna i wybór siedliska determinują narażenie piskląt mew na metale i metaloidy, a estuarium Sekwany pozostaje istotnym źródłem emisji rtęci i innych pierwiastków do środowiska morskiego.

Literatura

1. Hoffman D.J., Rattner B.A., Burton G.A. Jr., Cairns J. Jr. (2003) Handbook of Ecotoxicology, Lewis Publishers, Boca Raton.
2. Skoog D.A., Holler F.J. and Crouch S.R. (2007) Principles of Instrumental Analysis. Thomson Brooks/Cole, Boston.

Copyright © StatSoft Polska 2025, info@StatSoft.pl

3. Carlson R. (1962) Silent Spring, Houghton Mifflin, Boston.

4. Kucharska K., Binkowski Ł.J., Batoryna M., Dudzik K., Zaguła G., Stawarz R. (2019) Blood mercury levels in mute swans (*Cygnus olor*) are not related to sex, but are related to age, with no blood parameter implications. Environmental Pollution 252: 21-30.

5. Binkowski L.J., Fort J., Brault-Favrou M., Gallien F., Le Guillou G., Chastel O., Bustamante P. (2021) Foraging ecology drives mercury contamination in chick gulls from the English Channel. Chemosphere 267: 128622.

6. Binkowski L.J., Fort J., Churlaud C., Gallien F., Le Guillou G., Bustamante P. (2024) Levels of trace elements in the blood of chick gulls from the English Channel: Spatial and trophic implications. Science of the Total Environment 953: 175891.

7. Lee L. (2025) NADA: Nondetects and Data Analysis for Environmental Data. https://doi.org/10.32614/CRAN.package.NADA

8. Delignette-Muller M.L., Dutang C. (2015) fitdistrplus: An R Package for Fitting Distributions. Journal of Statistical Software 64(4): 1-34.

9. Erickson R., Rattner B.A. (2020) Moving Beyond $p < 0.05$ in Ecotoxicology: A Guide for Practitioners. Environmental Toxicology and Chemistry 39(9): 1657–1669.

Copyright © StatSoft Polska 2025, info@StatSoft.pl