

OD AUTORA

W literaturze można znaleźć informację, że w latach trzydziestych ubiegłego wieku na terenie Prus Wschodnich wykryto istotny związek między liczbą urodzin dzieci i liczbą bocianów. Mniej więcej w tym samym okresie w północnej Kanadzie liczba urodzin dzieci korelowała w sposób istotny ze średnim spożyciem alkoholu na głowę mieszkańca. Na podstawie tych anegdotycznych faktów mogłoby się wydawać, że – tak jak to nam tłumaczono w przedszkolu – bocian odgrywa u nas jakąś istotną rolę w przychodzeniu na świat nowych pokoleń, natomiast w Kanadzie, gdzie bocianów nie ma, ludzie widać radzą sobie jakoś inaczej. Zdrowy rozsądek podpowiada jednak, że z tą statystką chyba coś jest nie tak, skoro stosując naukową przecież metodę, jaką jest analiza korelacji, otrzymaliśmy powyższe wyniki. Przykładów podobnych korelacji można znaleźć więcej. Wielkość strat materialnych spowodowanych pożarami dodatnio koreluje z liczbą strażaków uczestniczących w ich gaszeniu. Nie powinniśmy jednak z tego wyciągać wniosku, że to strażacy powodują te straty! Przytoczone przykłady doskonale ilustrują pojęcie tzw. związków pozornych i pokazują, że istnienie korelacji nie zawsze świadczy o istnieniu związku przyczynowego, a stosowanie metod statystycznych nie może zastąpić zdrowego rozsądku. Z takiego typu pułapkami możemy się spotkać na każdym kroku: od planowania badań, poprzez zbieranie danych, aż do ich analizy i interpretacji. Dlatego jednym z zadań przedmiotu **metody statystyczne w biologii** jest nauka doboru odpowiednich narzędzi do statystycznej analizy danych oraz nauka prawidłowej interpretacji uzyskanych wyników. Trzeba pamiętać, że nawet najlepszy program komputerowy i najlepszy przewodnik do ćwiczeń nie zastąpi kreatywnego myślenia, a źle zaplanowanych badań statystyka nie naprawi.

Drugie wydanie tego przewodnika zostało poszerzone m.in. o rozdziały poświęcone analizom wielowymiarowym, które szczególnie w nau-

kach biologicznych znajdują szerokie zastosowanie. Większość zjawisk w przyrodzie ma złożony charakter i sytuacje, w których możemy wyodrębnić tylko jeden czynnik oddziałujący na badaną przez nas zmienną występują rzadko. Stąd też nawet na poziomie pracy magisterskiej coraz częściej zachodzi potrzeba wykonania np. analizy kowariancji, która jest znakomitym wprowadzeniem do bardziej zaawansowanych technik, takich jak ogólne i uogólnione modele liniowe.

Przewodnik ten ma przede wszystkim za zadanie ułatwić pracę na ćwiczeniach z przedmiotu metody statystyczne w biologii nie tylko studentom, ale i prowadzącym ćwiczenia. Nie pretenduje on do miana podręcznika ani nie jest dokładnym spisem treści wykładów prowadzonych na Wydziale Biologii Uniwersytetu Gdańskiego, choć zawarte w nim w większości fikcyjne przykłady mogą stać się inspiracją do przeprowadzenia własnych badań. Mam nadzieję, że dzięki temu przewodnikowi poziom stresu związanego z ćwiczeniami, kolokwiami i końcowym egzaminem zostanie istotnie zmniejszony, a także, że przedstawione przykłady staną się pomocne w przygotowaniu prac magisterskich.

Na zakończenie chciałbym przeprosić, że choć przewodnik adresowany jest do studentów różnych specjalności biologicznych, to w większości zawiera przykłady z dziedziny ekologii. Zawiedzionych mogę zapewnić, że z matematycznego punktu widzenia liczba jaj w gnieździe ptaka i liczba mutacji w chromosomie Y należą do tej samej grupy zmiennych skokowych i podlegają tym samym metodom analizy statystycznej.

Gdańsk 7.06.2014

Włodzimierz Meissner

WSTĘP

Wykłady i ćwiczenia z przedmiotu **metody statystyczne w biologii** mają za zadanie zaznajomienie studentów z podstawowymi metodami opisu statystycznego i analizy danych. Dobór zagadnień podyktowany jest z jednej strony zakresem minimum programowego dla tego przedmiotu, a z drugiej – chęcią przedstawienia metod, które najczęściej bywają stosowane w pracach magisterskich na kierunkach biologicznych.

Niniejszy przewodnik zawiera przede wszystkim szczegółowy opis przebiegu dwunastu ćwiczeń, jakie studenci odbywają na sali komputerowej, posługując się programami *MS Excel* firmy Microsoft i *STATISTICA* firmy StatSoft. Przedstawiono w nim 38 przykładów, opartych w większości na fikcyjnych danych, które zostały tak dobrane, aby pokazać różne, zwykle najczęściej spotykane, sytuacje, z jakimi można się spotkać w analizie statystycznej. Dodatkowo przygotowano kilkanaście zadań do samodzielnego wykonania. Pliki zawierające dane do ćwiczeń są dostępne na stronie internetowej Katedry Ekologii i Zoologii Kręgowców Uniwersytetu Gdańskiego: <http://www.kezk.bio.univ.gda.pl> (dział Dla studentów/do pobrania) oraz na stronie Wydawnictwa Uniwersytetu Gdańskiego: <http://www.kiw.ug.pl/biologia/11-meissner.html>. Opisy ćwiczeń bazujących na programie *STATISTICA* dotyczą wersji powyżej 8.0. We wcześniejszych wersjach sposób wykonywania niektórych analiz może się nieco różnić, a część z nich jest niedostępna. Problem ten nie dotyczy programu *Excel*, ponieważ stosowane podczas ćwiczeń funkcje istnieją we wcześniejszych jego wersjach.

Przedstawiona na początku każdego ćwiczenia część teoretyczna stanowi podsumowanie najważniejszych zagadnień w zakresie niezbędnym do wykonania poszczególnych przykładów. Pełniejsze omówienie problematyki związanej z przedstawionymi metodami statystycznymi można znaleźć w ogólnodostępnych podręcznikach, z których dla początkujących można polecić dwa: autorstwa Zielińskiego (1999) *Jak po-*

kochać statystykę, czyli *STATISTICA do poduszki* i Łomnickiego (2007) *Wprowadzenie do statystyki dla przyrodników*. Dla osób nieco bardziej zaawansowanych przeznaczone jest trzypięciotomowe wydawnictwo zatytułowane *Przystępny kurs statystyki* (Stanisz 2006, 2007a, b) oraz dwa jednotomowe: *Biostatystyka* (Watała 2002) i *Analiza statystyczna w psychologii i pedagogice* (Ferguson & Takane 2008). Spośród książek napisanych w języku angielskim godny polecenia jest wielokrotnie wznawiany podręcznik *Biostatistical Analysis* (Zar 1996).

Przed rozpoczęciem ćwiczeń trzeba sprawdzić, czy sposób obliczania percentyli przez program *STATISTICA* jest ustawiony na **Dystrybucja empiryczna z interpolacją**.

Opcję tę ustawia się w menu **Narzędzia/Opcje**, w zakładce **Ogólne**. Włączenie innej opcji będzie miało wpływ m.in. na otrzymywanie wartości mediany lub na statystykę F w teście Browna–Forsythe’a.

ĆWICZENIE 1

Arkusz kalkulacyjny MS Excel

Zakres ćwiczenia:

- Układ menu, adresy komórek, formatowanie komórek, wpisywanie i kopiowanie formuł.
- Stworzenie bazy danych (np. baza danych osobowych grupy).
 - Przykłady skomplikowanych wzorów formuł (piętrowe ułamki, pierwiastkowanie, potęgowanie, logarytmowanie).
 - Posługiwanie się adresem komórki, znak \$.
 - Wykresy.
- Tabela przestawna.
- Kreator funkcji: ORAZ, FAŁSZ, PRAWDA, JEŻELI, ROK, DZIEŃ, MIESIĄC, MIN, MAX, ODCZ.STANDARDOWE, ŚREDNIA, MEDIANA, PEARSON, PIERWIASTEK, POTĘGA, SUMA.

Arkusze kalkulacyjne, takie jak *Microsoft Excel*, *Quattro Pro*, *OpenOffice Calc*, *Lotus Symphony Spreadsheets*, *iWork Numbers*, to programy komputerowe wchodzące w skład tzw. pakietów biurowych, które obecnie stały się podstawowym oprogramowaniem instalowanym w komputerach. Najczęściej używanym arkuszem kalkulacyjnym jest *Excel* i nim będziemy posługiwać się podczas ćwiczeń. Jednak wszystkie zaprezentowane zadania można również wykonać, używając innych programów tego typu. Warto wspomnieć, że dostępne są również darmowe programy, które nie wchodzą w skład pakietów biurowych, a w pełni spełniają wszystkie zadania potrzebne przy pisaniu pracy magisterskiej (np. *GS-Calc*, *Gnumeric*).

Arkusze kalkulacyjne pozwalają na wykonywanie różnych operacji na liczbach i datach oraz zestawienie danych w tabelach, w tym w postaci uporządkowanej bazy danych. Za ich pomocą można także tworzyć

różne typy wykresów. Dzięki swoim zaawansowanym funkcjom programy te mogą okazać się wystarczające do przeprowadzenia analiz i wykresów koniecznych do wykonania pracy magisterskiej, w której analiza statystyczna sprowadza się do statystyki opisowej. Podstawowe funkcje *Excels* zostały przedstawione na zajęciach z przedmiotu **technologie informacyjne** i dlatego nie będą tutaj szczegółowo omawiane. Celem ćwiczenia jest przypomnienie najważniejszych działań, jakie można wykonywać z pomocą arkusza kalkulacyjnego, oraz ich zastosowanie w praktyce.

Zadania do samodzielnego wykonania:

Zadanie 1.1. Na podstawie danych zawartych w poniższej tabeli należy obliczyć wartość następującego wyrażenia:

$$Wynik = \sqrt{\frac{(C - B)^2}{(A + B) \cdot (C + D)}} + (0,5 \cdot D).$$

A	B	C	D	Wynik
12	0,3	2	63	5,612808
14	0,2	4	62	5,569148
12	0,1	3	61	5,523664
15	0,6	5	64	5,658444
18	0,5	2	59	5,431574

Wynik widnieje w ostatniej kolumnie.

Zadanie 1.2. Na podstawie danych podanych w tabeli obliczyć wartość wyrażenia W:

$$W = \frac{\sqrt[3]{\sum_{i=1}^{i=6} (C^3 - C)}}{\frac{(\bar{x}_A - \bar{x}_B)^2}{s_A \cdot s_B}},$$

A	B	C
12	28	1,1
14	16	1,3
18	16	1,5
11	12	1,8
16	19	1,9
15	14	2

gdzie $\bar{x}_A$, $\bar{x}_B$, s_A , s_B oznaczają wartości średnie i odchylenia standardowe dla danych z kolumny A i B.