

IX Sesja Paleolimnologiczna

15-16 marca 2018 r., Kraków

WARSZTATY STATYSTYCZNE

prof. dr hab. inż. Adam Walanus
walanus@geol.agh.edu.pl

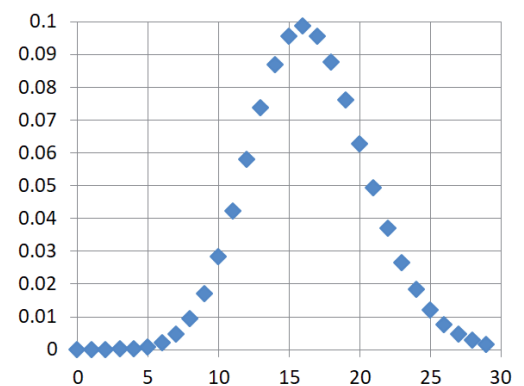
Wydział Geologii, Geofizyki i Ochrony Środowiska
Akademia Górniczo-Hutnicza w Krakowie



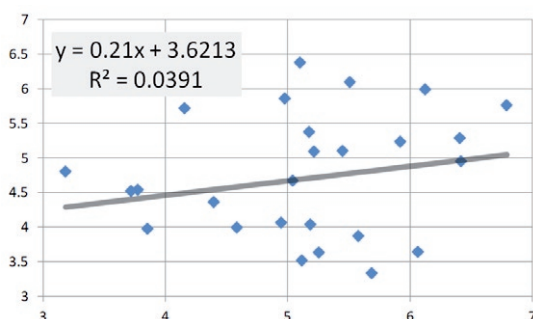
Statystyka to wydobywanie istotnej informacji z tła informacji nieważnej (nieważnej dla nas, w tym momencie). Tę nieważną informację można nazwać szumem. Następujący przykład jest prosty ale nie banalny. W centymetrze sześciennym osadu jeziornego, na głębokości 30 cm znaleziono 18 sztuk owada X, a na głębokości 50 cm „jedynie” 15. Nie interesuje nas gęstość, czy jest ich tam 15, czy 150 cm⁻³. Co ciekawsze, właściwie nie interesują nas liczby 18 i 15, a interesuje nas jedynie, czy 15 jest mniejsze od 18. Czy jest ISTOTNIE mniejsze? Interesują nas preferencje X-ów co do głębokości. Oczywiście, nawet 17 jest mniejsze od 18. Ale wnioskowanie, na takiej podstawie, że owe robaki wolą płytczną jest zupełnie nieuprawnione (osoby, które się z tym zupełnie nie zgadzają otrzymują zwrot wpisowego).

Tak więc: czy 15 jest istotnie mniejsze od 18?

Aby odpowiedzieć na to pytanie musimy wiedzieć, że „zmienna losowa”, którą się tu zajmujemy „podlega” rozkładowi prawdopodobieństwa Poissona. Jeżeli to wiemy i posłużymy się programem MS Excel, to otrzymamy następujący wykres:



We wzorze na rozkład Poissona, jako jego jedyny parametr wpisano średnią z 15 i 18. Jak widać obie te wartości, w tej sytuacji są bardzo prawdopodobne. Być może więc średnio owadów jest tam ok. 16.5 cm⁻³, a w takiej sytuacji oba wyniki są bardzo prawdopodobne i żaden nie przeczy hipotezie o braku preferencji głębokościowych. Powyższe wnioskowanie statystyczne jest nie całkiem ortodoksyjne. Jego celem było tylko zwrócenie uwagi, że 15 niekoniecznie jest mniejsze od 18.



Na podobnej zasadzie **NIE WOLNO** rysować linii regresji i wypisywać jej równania, tam gdzie żadnej zależności nie ma, na co wskazuje bardzo niska wartość R² (i zdroworozsądkowa obserwacja wykresu).



Program MS Excel jest bardzo dobrym narzędziem statystycznym i narzędziem porządkowania danych. Jest dobrym kursem porządku (kursem na porządek). Poza obszernym zestawem statystycznych formuł arkusza, Excel ma osobny moduł Analiza danych, gdzie podstawowe metody statystyczne (ale również transformacja Fouriera) dają wynik w pełnej postaci, łącznie z decydującą o wszystkim p-wartością.

Dla odważnych jest tam też tzw. Visual Basic, za pomocą którego **ZROBIMY WSZYSTKO**, głównie z dziedziny organizowania i przetwarzania danych, automatyzacji pracy itp.

Dla tych samych osób jest też R, darmowy program mający wszystko.

Tak więc: Excel, R i, jakoś tak na końcu STATISTICA (lub coś podobnego) – drogi program komercyjny, który dość łatwo da nam wyczerpujący wynik, odpowiadający niekoniecznie na nasze pytanie (celem programu komercyjnego jest zadowolenie klienta, a nie osiągnięcie przez niego prawdy - skrajnie niekomercyjny jest darmowy R – zero ułatwień = prawda w czystej postaci).