

Lista zadań nr 1
Statystyka Matematyczna
II rok Inżynierii Bezpieczeństwa / Inżynierii Środowiska / Gospodarki Wodnej
(semestr 3, studia stacjonarne 1-go stopnia)

Zadania dotyczące statystyki opisowej

I. W ramach systematycznych badań hydrologicznych zarejestrowano maksymalne roczne stany wód [w cm] rzeki Odry w przekroju wodowskazowym Miedonia. Dane pochodzą z lat 1901-1940 (zmienna **wmax1901_40**) i z lat 1941-1983 (zmienna **wmax1941_83**) (źródło danych: Instytut Inżynierii Środowiska, AR, Wrocław, rok 1992). Dane zawarte są dodatkowo w plikach MAXMIED.xls oraz MAXMIED.sta.

lp_rok	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
wmax1901_40	450	1094	1175	447	318	331	640	510	719	785	901	487	665	386	1014	353	489	597	541	756
wmax1941_83	1078	529	350	411	296	398	473	342	760	220	836	412	443	595	396	244	306	604	590	1005

lp_rok	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
wmax1901_40	357	633	400	494	785	706	403	218	510	888	888	409	247	367	345	506	876	561	1304	1078
wmax1941_83	314	582	557	274	885	915	388	812	385	836	330	1045	276	513	493	519	960	263	492	499
lp_rok	41	42	43																	
wmax1941_83	401	419	281																	

Dla tych danych dokonaj analizy statystycznej poprzez elementy statystyki opisowej:

a) dokonaj graficznej prezentacji materiału empirycznego poprzez:

1. histogram częstości wg następującego algorytmu:

1.1. uporządkuj dane rosnąco

1.2. wyznacz: **min**, **max**, **R=max-min**

1.3. wyznacz liczbę klas (kategorii, przedziałów zmienności) **k**: powinna być ona funkcją rozmiaru próby tj. liczby obserwacji **n**. Dobrą propozycją jest następująca klasyczna reguła, określona przez poniższe nierówności:

$$\frac{3}{4}\sqrt{n} \leq k = f(n) \leq \sqrt{n}$$

1.4. wyznacz długość klasy (przedziału) **b= R/k** (lub **b=(R+2ε)/k**, dla przedziału zmienności (**min-ε**; **max+ε**), gdzie ε wiąże się z dokładnością pomiarów; jest to zalecane w przypadkach mało wygodnej wartości liczbowej **b**

1.5. dla danej i-tej klasy (tj. przedziału $[x_i, x_{i+1})$) oblicz **n_i** – liczbę obserwacji, która wpada do tej klasy

1.6 dla tak uzyskanego szeregu rozdzielczego :

dla **i=1,...,k**

i	$[x_i, x_{i+1})$	n _i
---	------------------	----------------

sporządź wykres jako układ prostokątów, którego podstawą jest przedział $[x_i, x_{i+1}]$, a wysokość odpowiada wartości **n_i**

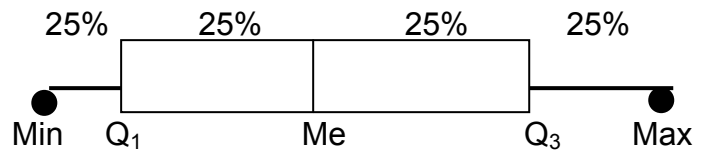
2. histogram częstości, przyjmując przedział zmienności [50 ; 100] i jednakową rozpiętość przedziałów klasowych równą 10.

3. wykres pudełkowy w oparciu o statystyki pozycyjne (rangowe)

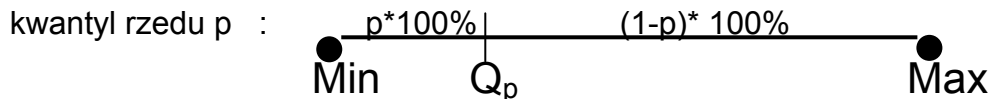
Miary pozycyjne (kwantyle)

Kwantyle - definiuje się jako wartości cechy badanej zbiorowości, przedstawionej w postaci szeregu statystycznego, które dzielą zbiorowość na określone części pod względem liczby jednostek, części te pozostają do siebie w określonych proporcjach.

- kwartył pierwszy Q_1
- mediana (kwartył drugi) Q_2 (Me)
- kwartył trzeci Q_3
- decyle ($P_{10}, P_{20}, \dots, P_{90}$)



rys. 1. Kwartylowy wykres pudełkowy



$$Me = \begin{cases} x_{(n+1)/2} & \text{gdy } n \text{ jest nieparzyste} \\ \frac{x_{n/2} + x_{(n/2)+1}}{2} & \text{gdy } n \text{ jest parzyste} \end{cases}$$

Dla szeregu szczegółowego (uporządkowanej próby statystycznej rozmiaru n)

- b) wyznacz podstawowe charakterystyki liczbowe
1. miary położenia
 2. miary rozproszenia
 3. charakterystyki rozkładu z próby

Patrz np. pozycja:

Rachunek prawdopodobieństwa i statystyka matematyczna w zadaniach.

Część II. Statystyka matematyczna, Krysicki W., Bartos J., Dyczka W., Królikowska K., Wasilewski M., PWN, Warszawa, 1993, Wyd.III. **strony 1-31**

- c) zbierz wyniki i podaj ich interpretację.

II. Sformułuj wnioski i zaproponuj hipotezy robocze będące przedmiotem dalszego wnioskowania statystycznego

Opracował dr hab. Andrzej Michalski