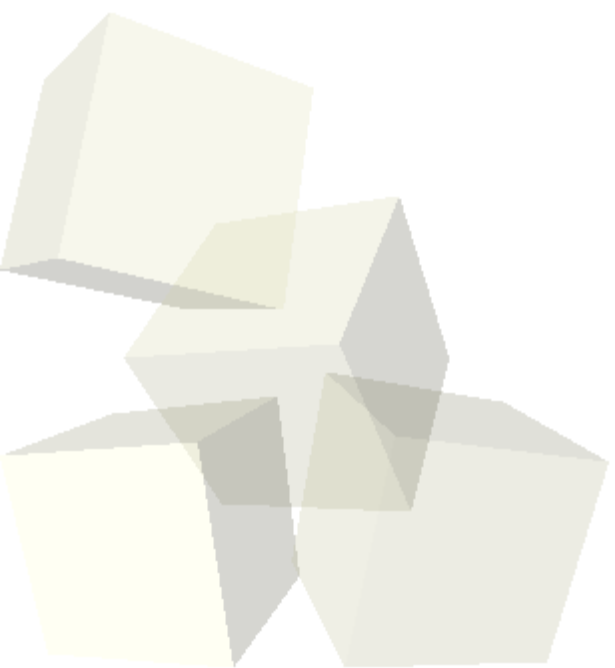




## Statystyka pomiarów – podejście od strony rachunku prawdopodobieństwa

*Ewa Pawelec*





# Analiza grupy pomiarów

- Mamy  $N$  pomiarów –  $x_1$  do  $x_N$  – jak zinterpretować? Jaka wielkość opisze nam wielkość najbliższą prawdziwej, a jaka najlepiej opisze rozrzut?
  - Wielkość oddająca najlepiej wartość prawdziwą, dla pomiarów obarczonych błędem przypadkowym – średnia (wartość oczekiwana):

$$\bar{x} = \frac{\sum_{i=1}^N x_i}{N}$$

- Wiele wielkości mogłoby oddawać rozrzut (np. suma wartości bezwzględnych różnic pomiędzy wynikiem pomiaru a średnią), ale zazwyczaj stosowane jest odchylenie standardowe, czyli pierwiastek z sumy kwadratów różnic (pierwiastek z wariancji):

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N}}$$

# Różne rodzaje liczenia średniej

- Mamy przykładowy zbiór pomiarów jednej wartości:  
 $x_i = 31, 29, 31, 33, 28, 29, 30, 29, 31, 30$  ( $i$  od 1 do  $N=10$ )
- Ten sam zbiór pomiarów można przedstawić jako:

$k$	1	2	3	4	5	6
$x_k$	28	29	30	31	32	33
$n_k$	1	3	2	3	0	1

dla  $k$  od 1 do  $M=6$ , gdzie  $n_k$  oznacza częstość występowania jakiejś liczby.

- Średnią z takiego zbioru liczb można wyliczyć na dwa sposoby:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_N}{N} = \frac{\sum_{i=1}^N x_i}{N}$$
$$\bar{x} = \frac{x_1 + 3 \cdot x_2 + 2 \cdot x_3 + \dots + x_M}{N} = \frac{\sum_{k=1}^M x_k \cdot n_k}{N}$$



- Obliczenie a) będzie to w takim razie suma po **wszystkich** pomiarach, natomiast b) – suma po **wszystkich różnych** pomiarach:

$$\text{a) } \bar{x} = \frac{\sum_{i=1}^N x_i}{N} \quad \text{b) } \bar{x} = \frac{\sum_{k=1}^M x_k \cdot n_k}{N} = \frac{\sum_{k=1}^M x_k \cdot n_k}{\sum_{k=1}^M n_k}, \quad \text{bo } \sum_{k=1}^M n_k = N$$

- Czasami może być lepsze używanie zamiast liczebności wyniku  $n_k$  tak zwanej częstości wyniku  $F_k$ :

$$F_k = \frac{n_k}{N}$$

która określa **rozkład** wyników.

- Wtedy równanie na średnią można zapisać jako

$$\bar{x} = \sum_{k=1}^M x_k \cdot F_k$$



# Częstość wyniku - własności

- Warto pamiętać, że częstość ma następującą własność:

$$\sum_{k=1}^M F_k = \frac{\sum_{k=1}^M n_k}{N} = \frac{\sum_{k=1}^M n_k}{N} = \frac{N}{N} = 1$$

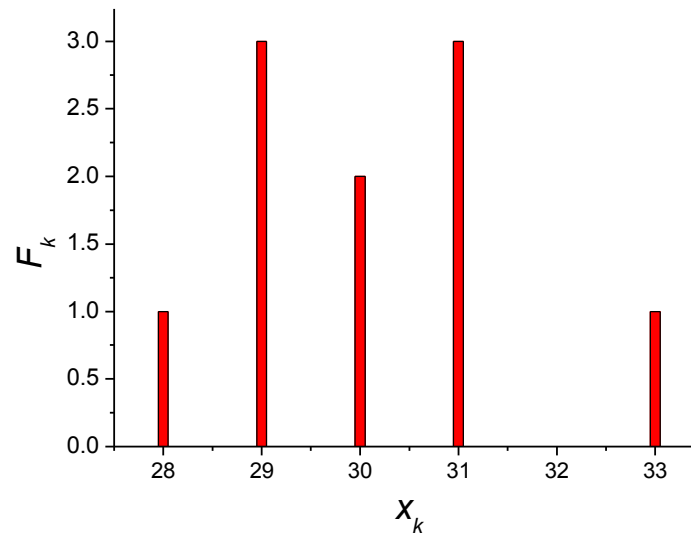
jest to logiczne, ponieważ częstość podaje „jaka część wyników pomiaru jest równa  $x_k$ ”, a części powinny się sumować do jedności.

- Zapis średniej wykorzystujący częstość występowania wyniku zwany jest **średnią ważoną**, gdzie jako waga wyniku występuje jego częstość.
- Używanie częstości wyniku do opisu eksperymentu ma tę zasadniczą zaletę, że można w ten sposób porównywać wyniki pomiarów różnolicznych (np. jeden pomiar gdzie wielkość zmierzono 10 razy, a drugi – 25 razy).



# Wykresy wyników pomiarów

- Wyniki pomiarów można przedstawiać różnie. Jedną z możliwości jest tak zwany histogram (wykres słupkowy):



- Taki histogram można utworzyć wtedy, kiedy mamy wiele wyników pomiaru o tych samych wartościach, a możliwe wartości są wyraźnie różne (np. rzut kostką lub liczba punktów zdobyta podczas testu).



# Przedziały wyników

- Jeżeli natomiast mamy wyniki pomiarów, które są dokładniejsze i bardziej „rozmyte”, np.:  
 $x_i = 31,4; 28,9; 30,1; 29,6; 27,7; 28,8; 30,2; 28,9; 30,3; 30,4$   
to ich analizę można przeprowadzić dzieląc na **przedziały**:

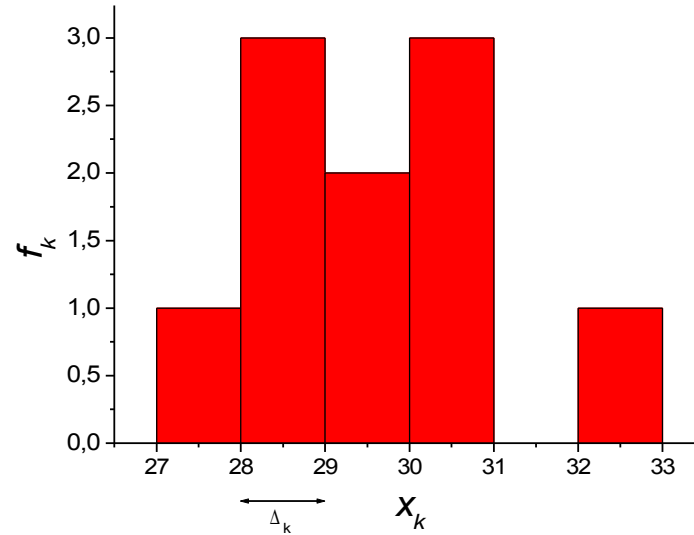
$k$	1	2	3	4	5	6
<i>przedział</i>	27-28	28-29	29-30	30-31	31-32	32-33
<i>liczba pomiarów</i>	1	3	2	3	0	1

- Tutaj także możemy określić częstość, ale tym razem wyników z określonego przedziału
- Przedział (komórkę) definiujemy przez jego szerokość  $\Delta_k$ , a zamiast  $F_k$  przydatna będzie wielkość  $f_k$ , określająca jak wiele z naszych wyników trafiło do komórki o numerze  $k$  i szerokości  $\Delta_k$ . Liczba wyników w komórce jest równa  $\Delta_k \cdot f_k$
- Uwaga –  $\Delta_k$  wybieramy sami! Wcale nie musi być równe 1.

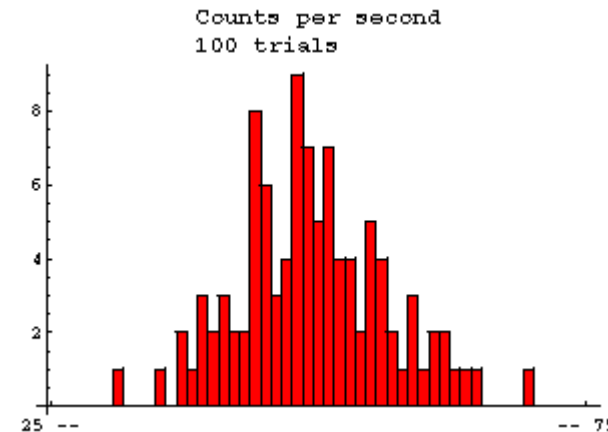
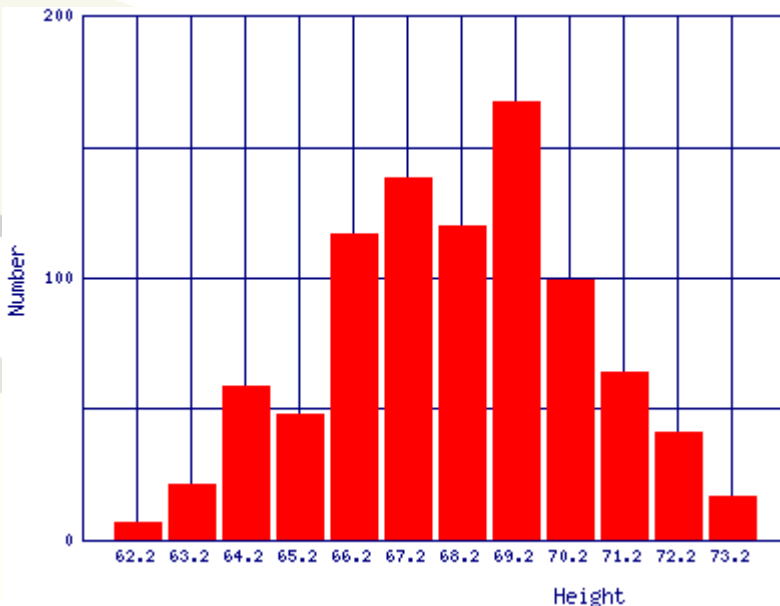


# Histogramy cd

- Taki wykres nazywany jest histogramem komórkowym:



- Przykładowe histogramy dla większej liczby pomiarów:

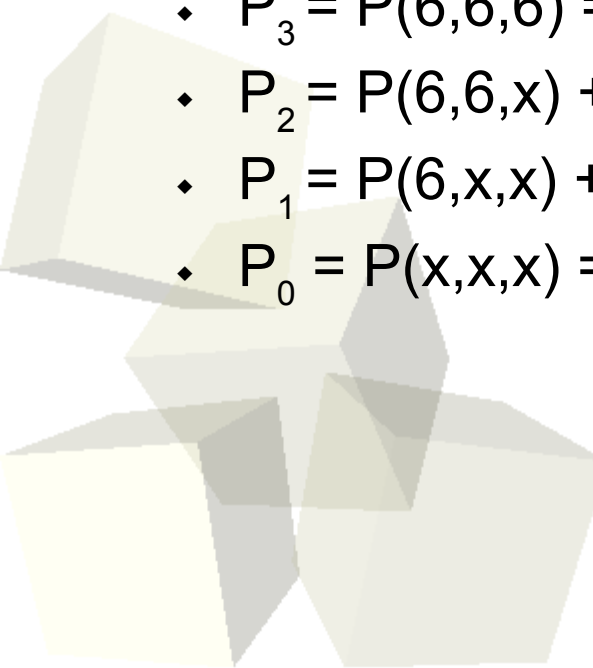






# Rozkłady szczególne

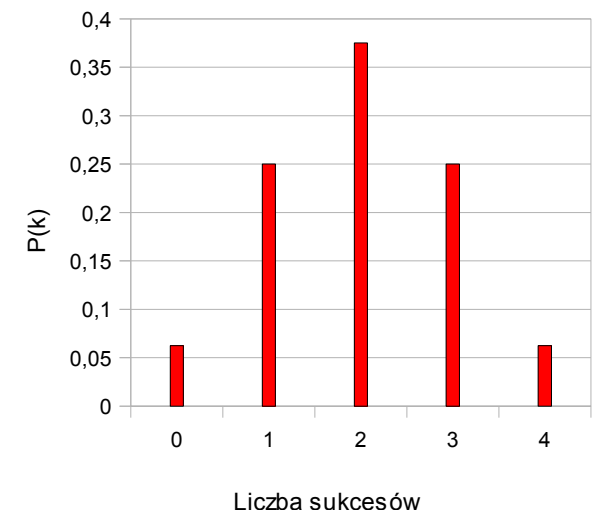
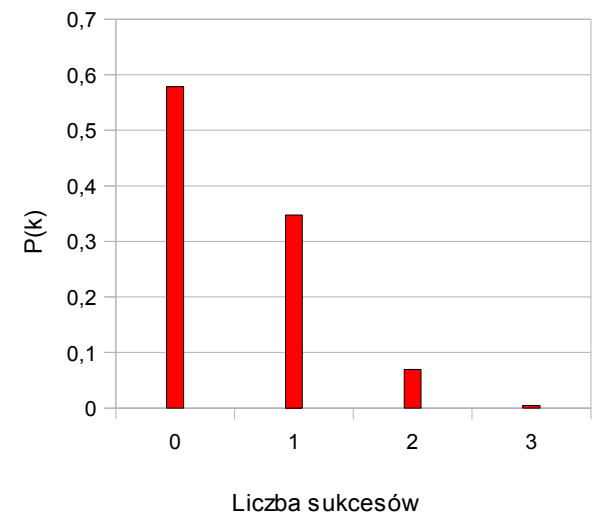
- Zamiast częstości pomiarów można rozważać prawdopodobieństwo pomiaru – efekt jest ten sam
- Rozkład wyników rzutu trzema kostkami, dla którego „sukces” to wypadnięcie szóstki. Możliwe efekty:  
(6,6,6), (6,6,x), (6,x,6), (x,6,6), (6,x,x), (x,6,x), (x,x,6), (x,x,x)  
gdzie x – „nie szóstka”
- Prawdopodobieństwo zera, jednego, dwóch i trzech sukcesów w jednej próbie (jednym rzucie):
  - $P_3 = P(6,6,6) = (1/6)^3 = 0,46\%$
  - $P_2 = P(6,6,x) + P(6,x,6) + P(x,6,6) = 3 \cdot (1/6)^2 \cdot 5/6 = 6,94\%$
  - $P_1 = P(6,x,x) + P(x,6,x) + P(x,x,6) = 3 \cdot (5/6)^2 \cdot 1/6 = 34,72\%$
  - $P_0 = P(x,x,x) = (5/6)^3 = 57,87\%$





# Wykres wyników – histogram

- Wyniki przedstawione na wykresie pokazują, że w przypadku trzech kostek i liczenia szóstek wybitnie nie jest to wykres symetryczny
- Wykres symetryczny dostajemy jedynie, jeżeli prawdopodobieństwo sukcesu i porażki wynosi tyle samo (np. 4 rzuty monetą):
  - $P_4 = P_0 = (1/2)^4 = 6,25\%$
  - $P_1 = P_3 = 4 \cdot (1/2)^3 = 25\%$
  - $P_2 = 37,5\%$
- Ogólna nazwa na takie rozkłady: rozkład dwumianowy





# Rozkład dwumianowy

- Mamy  $N$  prób, w każdej możemy dostać sukces z prawdopodobieństwem  $p$ . Jakie jest prawdopodobieństwo, że dostaniemy dokładnie  $k$  sukcesów? Funkcja prawdopodobieństwa:

$$P(k) = \binom{N}{k} p^k (1-p)^{N-k} \qquad \binom{N}{k} = \frac{N!}{k!(N-k)!}$$

- Wartość oczekiwaną i wariancję (odpowiedniki średniej, oraz kwadratu odchylenia standardowego) dla takiego rozkładu można policzyć dokładnie, i będą one wynosiły (jeżeli naszą zmienną oznaczymy jako  $X$ )

$$E(X) = N \cdot p$$

$$\text{Var}(X) = N \cdot p \cdot (1-p)$$



# Rozkład Poissona

- Rozkład Poissona dostajemy jeżeli  $N \rightarrow \infty$  oraz  $p \rightarrow 0$  (a  $Np$  pozostaje ciągle całkiem duże, ale skończone). Rozkład taki otrzymujemy m.in. dla zliczania rozpadów jąder atomowych, wyświecania fotonów ze wzbudzonego ośrodka itp.
- Rozkład Poissona opisywany jest parametrem  $\lambda = Np$ , a prawdopodobieństwo  $k$  sukcesów (zliczeń) dane jest wzorem:

$$P(k, \lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$$

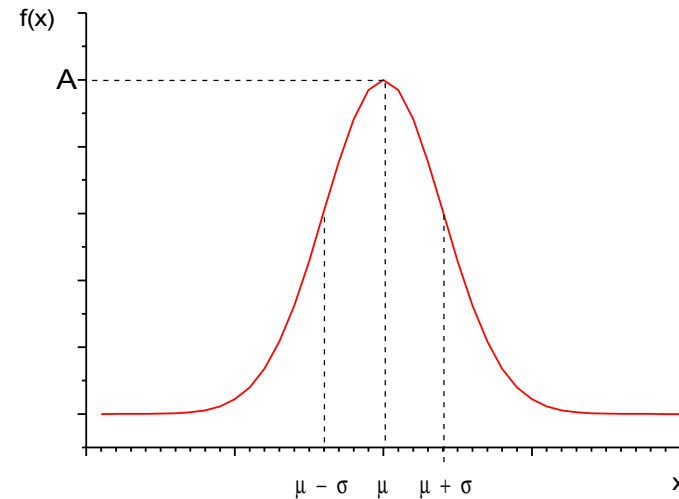
- Można sprawdzić, że wariancja dla rozkładu Poissona także wynosi  $\lambda$ , czyli odchylenie standardowe dla tego rozkładu (błąd licznika przy zliczeniach) wynosi  $\sqrt{\lambda}$ , czyli pierwiastek z liczby zliczeń.



# Rozkład normalny

- Jeżeli zwiększymy liczbę prób  $N$ , a  $p$  pozostanie stałe, będziemy z coraz lepszym przybliżeniem otrzymywać krzywą Gaussa (krzywą dzwonową):

$$f(x) = A \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

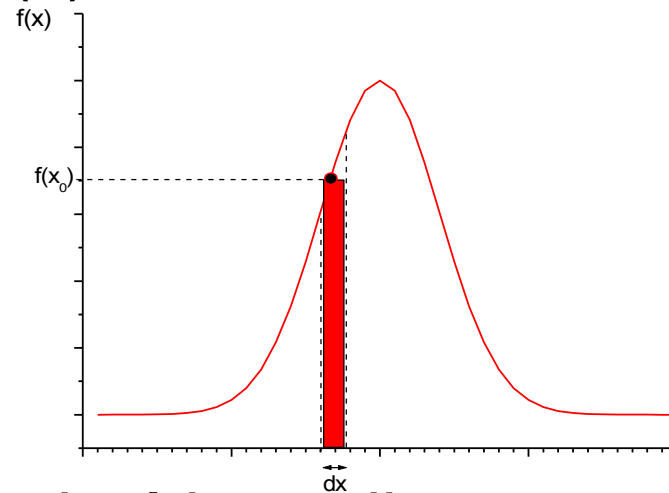


- Wielkość  $\mu$  określa środek krzywej,  $\sigma$  definiuje jej szerokość.
- Żeby sprawdzić ile mieliśmy wyników w jednej komórce histogramu komórkowego obliczaliśmy  $\Delta_k \cdot f_k$ . Żeby to policzyć dla rozkładu ciągłego należy **całkować**.

# Częstość wyników – rozkład ciągły

- Dla bardzo wąskich pasków (szerokość  $dx$ ) liczba wyników będzie w przybliżeniu równa  $f(x) \cdot dx$ :

$$N(x) \approx f(x) dx$$



- Dla większych pasków musi to być już całka:

$$N(a; b) = \int_a^b f(x) dx$$

- Operacja całkowania jest “uciągleniem” liczenia sumy.
- Częstość wyników odzwierciedla nam prawdopodobieństwo, że nowy wynik znajdzie się w tym zakresie, czyli zamiast  $N(x)$  możemy rozważać  $P(x)$ .

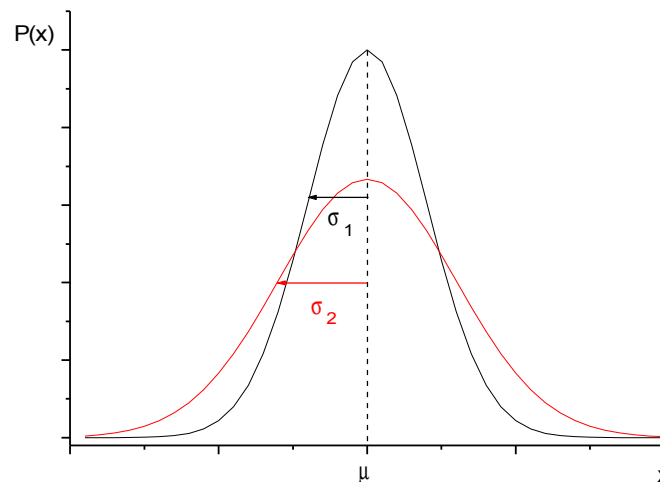


# Rozkład normalny – definicje

- Rozkład wyników opisany krzywą Gaussa nazywamy rozkładem normalnym.
- Rozkład normalny podaje nam prawdopodobieństwo że nasz wynik znajdzie się w określonym zakresie. Ponieważ prawdopodobieństwo, że nasz wynik znajdzie się w zakresie  $(-\infty; \infty)$  musi być równe 1, więc:

$$P(-\infty; \infty) = \int_{-\infty}^{\infty} f(x) dx = 1$$

- Tak więc dla różnych szerokości, rozkłady będą miały różną wysokość maksimum:



# Średnia i odchylenie standardowe

- Jeżeli mamy rozkład ciągły to zamiast sumy będziemy liczyć całkę. W takim razie:

$$\bar{x} = \sum_{k=1}^M x_k \cdot F_k \quad \rightarrow \quad \bar{x} = \int_{-\infty}^{\infty} x \cdot f(x) dx$$

- Tak samo wzór na odchylenie standardowe:

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N}} = \sqrt{\sum_{n=1}^M (x_k - \bar{x})^2 \cdot F(k)} \quad \rightarrow \quad \sigma = \sqrt{\int_{-\infty}^{\infty} (x - \bar{x})^2 \cdot f(x) dx}$$

- Jeżeli  $f(x)$  ma być dane rozkładem normalnym, to rozkład normalny musi mieć wzór (stała wyliczona z warunku  $\int_{-\infty}^{\infty} f(x) dx = 1$ ):

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} \cdot e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$