Akademia Górniczo-Hutnicza
im. Stanisława Staszica w Krakowie

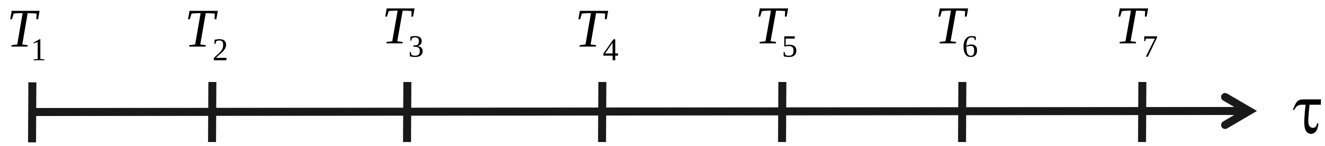
# Matematyczne metody opracowywania wyników

## Statystyka i rachunek niepewności

Paweł Żak
Wydział Odlewnictwa AGH
Katedra Inżynierii Procesów Odlewniczych

Kraków, 11 grudnia 2010

# Opracowanie krzywej stygnięcia




Formuły na przybliżanie pochodnej w punkcie $\tau_i$:

$$\frac{dT}{d\tau}(\tau_i) = \frac{T_{i+1} - T_i}{\Delta\tau} \quad ; \quad \frac{dT}{d\tau}(\tau_4) = \frac{T_5 - T_4}{\Delta\tau}$$

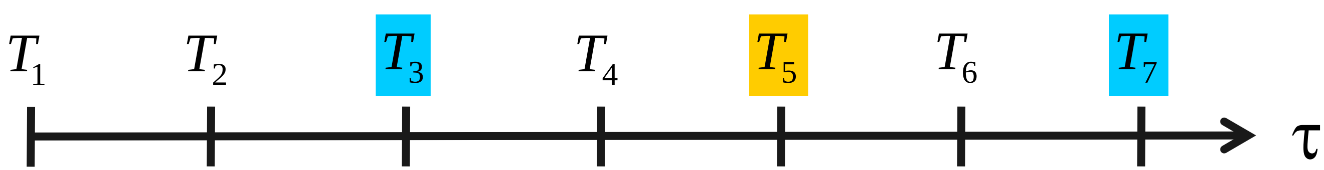
$$\frac{dT}{d\tau}(\tau_i) = \frac{T_{i+k} - T_{i-k}}{2k\Delta\tau} \quad ;$$

$$\frac{dT}{d\tau}(\tau_3) = \frac{T_5 - T_1}{4\Delta\tau}; \; \frac{dT}{d\tau}(\tau_4) = \frac{T_6 - T_2}{4\Delta\tau}; \frac{dT}{d\tau}(\tau_5) = \frac{T_7 - T_3}{4\Delta\tau};$$

# Opracowanie krzywej stygnięcia


$$\frac{dT}{d\tau}\left(\tau_3\right)=\frac{T_5-T_1}{4\Delta\tau}$$


$$\frac{dT}{d\tau}\left(\tau_4\right)=\frac{T_6-T_2}{4\Delta\tau}$$




$$\frac{dT}{d\tau}\left(\tau_5\right)=\frac{T_7-T_3}{4\Delta\tau}$$

# Pomiar zmiany względnej procentowej

Jeżeli chcemy porównać zmiany dwóch wartości, możemy wyznaczyć współczynnik zmiany względnej procentowej. Zapis:

$$100\% \cdot \frac{|A-B|}{A}$$

rozumiemy: procentowa zmiana wartości, gdzie A jest wartością wcześniejszą, B zmienioną.

Powyższa formuła przyjmuje tylko wartości o znaku zgodnym ze znakiem A.
Poniższa formuła pokazuje także kierunek zmian wartości:

$$100\% \cdot \frac{(A-B)}{A}$$

# Elementy statystyki

**Statystyka** jest nauką traktującą o metodach ilościowych badania zjawisk masowych. Stanowi ona podstawę wszelkiej działalności badawczej.

**Zjawisko masowe** jest to takie zjawisko, które badane w dużej grupie zdarzeń, wykazuje właściwą sobie prawidłowość, jakiej nie bylibyśmy w stanie zaobserwować w pojedynczym przypadku.

**Cechy statystyczne**, to właściwości, którymi odznaczają się jednostki tworzące badaną zbiorowość. W dalszej części oznaczane będą przez $x_i$. W przypadku, gdy wartości są kwalifikowane przedziałami $x_i$, oznacza środek przedziału.

AGH

## Elementy statystyki

Dane statystyczne mówią nam o **rozkładzie danej cechy** w zbiorowości. *Rozkład* możemy przedstawić w postaci:

-histogramów,
-diagramów,
-tabel,
-wykresów w postaci krzywych,

Jednak nie zawsze wystarczy to do porównania dwóch zbiorowości, lub do wyciągnięcia jednoznacznych wniosków.

Potrzebujemy scharakteryzować badaną zbiorowość wartościami **liczbowymi** lub, jeżeli jest to możliwe, opisać (estymować) rozkład przy pomocy funkcji matematycznych.

## Klasyfikacja danych pod względem ilości

W praktyce pomiarowej dysponujemy jedynie pewną skończoną, i zwykle niewielką, ilością pomiarów. Tę grupę pomiarów nazywa się najczęściej **próbą**.

Ze względu na wielkość próby klasyfikujemy w następujący sposób:

a) Próba bardzo mała $n < 10$
b) Próba mała $10 < n < 30$
c) Próba duża $n > 30$

## Często spotykane rozkłady
## Rozkład normalny

Gęstość rozkładu normalnego jest postaci:

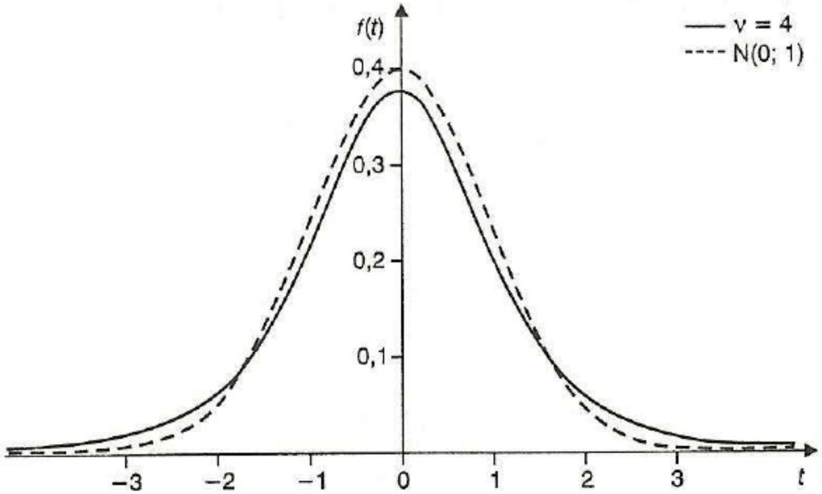
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left( -\frac{(x-\mu)^2}{2\sigma^2} \right)$$

ze względu na możliwość korzystania ze specjalnych tablic rozkładu często stosuje się rozkład normalny zestandaryzowany:

$$\varphi(u) = \frac{1}{\sqrt{2\pi}} \exp\left( -\frac{u^2}{2} \right)$$

# Często spotykane rozkłady
# Rozkład t-Studenta

$$f(t)=\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{v}{2}\right)\sqrt{v\pi}}\left(1+\frac{t^2}{v}\right)^{-\frac{n}{2}}$$




*Zastosowania*

Rozkład t Studenta stosowany jest w estymacji przedziałowej, w testach parametrów w szczególności dla średnich, wariancji oraz w testach istotności parametrów statystycznych – gdy mamy do czynienia z **małymi próbami** (najczęściej przyjmuje się że próba jest mała gdy jej liczebność n $\leq$ 30). Ponadto rozkład Studenta znajduje zastosowanie przy weryfikacji niektórych hipotez dotyczących średniej, gdy dysponuje się mała próbą, czli wtedy gdy nie można wykorzystać rozkładu normalnego, oraz przy sprawdzaniu wątpliwych wyników obserwacji.

# Często spotykane rozkłady
## Rozkład logonormalny

Gęstość rozkładu logarytmicznie normalnego jest postaci:

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left( -\frac{\left(\ln(x) - \mu\right)^2}{2\sigma^2} \right)$$

Rozkład ten obrazuje dobrze rozkład cech, gdy interesujące są dla nas relacje między wartościami.

Dobrze nadaje się do przybliżenia rozkładu cząstek w skali mikro (ziarna piasku, modyfikator, cząstki zbrojące)

# Często spotykane rozkłady
# Rozkład wykładniczy

Gęstość rozkładu wykładniczego ma postać:

$$\begin{cases} f(x) = \frac{1}{c}\exp\left(-\frac{x}{c}\right), & x \geq 0 \\ f(x) = 0, & x < 0 \end{cases}$$

Jest to rozkład interpretowany jako czas oczekiwania na sukces (im dłużej czekam na autobus tym mniejsza jest szansa na to, że w ciągu następnej sekundy przyjedzie)

# Charakterystyki liczbowe rozkładu jednej cechy w zbiorowości

Miary położenia

wartość przeciętna

Miary zmienności

miara rozproszenia, zróżnicowania dyspersji

Miary asymetrii lub skośności

Miary spłaszczenia i koncentracji

# Miary położenia:
## *Średnia arytmetyczna*

Średnia arytmetyczna – wzór prosty

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

Średnia arytmetyczna – wzór z wagami

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} f_i x_i \qquad n = \sum_{i=1}^{k} f_i$$

Średnia arytmetyczna informuje o przeciętnym poziomie badanej cechy w całej zbiorowości.

## Miary położenia:
## *Średnia geometryczna*

Średnia geometryczna – wzór prosty

$$G = \sqrt[n]{x_1 x_2 \ldots x_n} = \sqrt[n]{\prod_{i=1}^{n} x_i}$$

Średnia geometryczna – wzór z wagami

$$G = \sqrt[n]{x_1^{f_1} x_2^{f_2} \ldots x_k^{f_k}} = \sqrt[n]{\prod_{i=1}^{k} x_i^{f_i}} \qquad n = \sum_{i=1}^{k} f_i$$

Średnia geometryczna
postaci po zlogarytmowaniu

$$\ln(G) = \frac{1}{n} \sum_{i=1}^{n} \ln(x_i)$$

$$\ln(G) = \frac{1}{n} \sum_{i=1}^{k} f_i \ln(x_i)$$

Średnia geometryczna znajduje przede wszystkim zastosowanie przy badaniu średniego tempa zmian analizowanych zjawisk.

Średnia geometryczna reaguje znacznie słabiej na wartości ekstremalne niż średnia arytmetyczna.

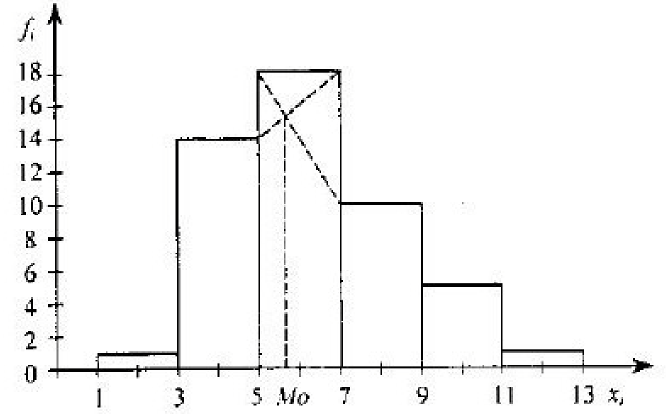
# Miary położenia: *Modalna*

**Modalna (dominanta)** to wartość cechy, która pojawia się największą ilość razy. W przypadku, gdy mamy do czynienia z przedziałami można ją wyznaczyć ze wzoru:

$$Mo = x_m + \frac{f_m - f_{m-1}}{\left(f_m - f_{m-1}\right) + \left(f_m + f_{m+1}\right)} h$$

$x_m$ – dolna granica przedziału najbardziej licznego (modalnej)
$f_m$ – liczebność przedziału modalnej
$f_{m-1}$ – liczebność przedziału poprzedzającego
$f_{m+1}$ – liczebność przedziału następującego
h – interwał
(rozpiętość przedziału modalnej)

Jest odpowiednikiem średniej odczytywana z histogramu. Jej interpretacja może być znacznie zaburzona ze względu na przesunięcie mierzonych wartości




Rys. 2.5. Wyznaczenie modalnej na podstawie histogramu

# Miary zmienności

Określenie przeciętnego poziomu badanej cechy nie wystarcza do scharakteryzowania struktury zbiorowości statystycznej.

Konieczne jest uzupełnienie wiedzy na temat rozkładu o informację mówiącą o stopniu zróżnicowania poszczególnych jednostek statystycznych.

Służą do tego miary zmienności:
-rozstęp
-odchylenie przeciętne
-wariancja
-odchylenie standardowe

# Miary zmienności
## *Rozstęp*

**Rozstęp** jest różnicą między wartością największą i najmniejszą obserwowaną w danej zbiorowości:

$$R = \max_i \{x_i\} - \min_i \{x_i\}$$

Rozstęp jest stosowany głównie w sytuacjach, gdy konieczne jest szybkie określenie obszaru zmienności danej zbiorowości, o niewielkiej liczbie jednostek

Rozstęp mówi nam o rozproszeniu wartości. Im rozstęp jest mniejszy, tym rozproszenie mniejsze, a wartości bardziej skupione są wokół średniej.

# Miary zmienności
## *Odchylenie przeciętne*

**Odchylenie przeciętne** jest równe średniej arytmetycznej bezwzględnych wartości odchyleń zmiennej od jej średniej:

$$D = \frac{1}{n}\sum_{i=1}^{n}\left|x_i - \bar{x}\right|$$

lub w postaci z wagami:

$$D = \frac{1}{n}\sum_{i=1}^{k} f_i\left|x_i - \bar{x}\right|$$

$f_i$ są wagami odpowiadającymi poszczególnym wariantom $x_i$

Odchylenie przeciętne określa, o ile wszystkie jednostki danej zbiorowości różnią się średnio ze względu na wartość od średniej arytmetycznej tej zbiorowości.

# Miary zmienności
## *Wariancja*

**Wariancja** jest to średnia arytmetyczna z kwadratów odchyleń poszczególnych wartości cechy od ich średniej arytmetycznej:

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$$

przypadek z wagami:

$$s^2 = \frac{1}{n}\sum_{i=1}^{k} f_i\left(x_i - \bar{x}\right)^2$$

Wariancja jest wartością bardzo trudną do interpretacji. Jej wyznaczenie jest jednak konieczne do wyznaczenia najistotniejszego parametru określającą dyspersję rozkładu: odchylenia standardowego.

## Miary zmienności
## *Korekty wartości wariancji*

Jeżeli  wariancja liczona jest z małej próby i przeznaczona jest do wnioskowania statystycznego o populacji, z której próba ta pochodzi, stosujemy tzw. poprawkę Bessela:

$$s^2 = \frac{1}{(n-1)} \sum_{i=1}^{k} f_i \left(x_i - \bar{x}\right)^2$$

podobnie dla postaci prostej.

Wariancja wyznaczona ze średnich wartości przedziałów może być obarczona błędem grupowania. W szczególności, gdy liczba przedziałów (k) jest nieduża w stosunku do przeciętnej długości przedziału (h). Stwarza to możliwość przeszacowania wariancji.
W sytuacjach takich stosujemy poprawkę Shepparda:

$$s^2 = \frac{1}{n} \sum_{i=1}^{k} f_i \left(x_i - \bar{x}\right)^2 - \frac{h^2}{12}$$

# Miary zmienności
## *Odchylenie standardowe*

**Odchylenie standardowe** jest pierwiastkiem kwadratowym z wariancji:

$$s = \sqrt{s^2}$$

informuje nas ono jak jest średnia wartość odchyleń od wartości średniej wyznaczonej dla zbiorowości. W przypadku analizy polegającej na pomiarze konkretnej wielkości fizycznej, jest ono interpretowane jako błąd pomiaru. W tej sytuacji jest ona zarazem parametrem sigma rozkładu normalnego opisującego prawdopodobieństwo napotkania na konkretną wartość podczas pomiaru.

# Miary zmienności
## *Współczynnik względnej zmienności*

**Współczynnik względnej zmienności** jest względną miarą dyspersji. Podczas analizy korzystamy z różnych dotąd poznanych miar. Najczęściej spotykane są:

$$V_s = \frac{s}{\left|\overline{x}\right|} \cdot 100\%$$

$$V_D = \frac{D}{\left|\overline{x}\right|} \cdot 100\%$$

Używamy go, gdy zaistnieje potrzeba dokonania porównań zbiorowości pod względem cech o różnych mianach. Pozwala na uzyskanie „dyspersji procentowej”.

# Miary asymetrii

W przypadku rozkładów symetrycznych średnia arytmetyczna oraz modalna równają się sobie. Są one położone centralnie i to wokół nich skupiają się pozostałe wartości.
Do analizy asymetryczności rozkładu można wykorzystać tę informację. Wprowadzamy tzw. **współczynniki asymetrii**:

$$A_s = \frac{\bar{x} - Mo}{s}$$

$$A_s = \frac{\bar{x} - Mo}{D}$$

Współczynnik ten określa kierunek oraz siłę asymetrii. Jeżeli As < 0 asymetria jest lewostronna, As = 0 rozkład jest symetryczny, As > 0 asymetria jest prawostronna.

# Miary spłaszczenia i koncentracji

Koncentracja zbiorowości wokół wartości średniej nosi nazwę **kurtozy**, inaczej **spłaszczenia**. Miarą skupienia poszczególnych obserwacji wokół średniej jest moment centralny rzędu czwartego (wzór z wagą):

$$M_4 = \frac{1}{n}\sum_{i=1}^{k} f_k\left(x_i - \bar{x}\right)^4$$

względną miarą koncentracji otrzymamy ze wzoru: $K = \frac{M_4}{(s)^4}$

K dla rozkładu normalnego N(0,1) wynosi 3, zatem by zdefiniować kurtozę, w sposób opisujący kształt krzywej w sposób bardziej przemawiający do wyobraźni (ze względu na odniesienie do rozkładu normalnego) wprowadzono następującą formę **współczynnika koncentracji**:

$$Ku = \frac{M_4}{(s)^4} - 3$$

Dla Ku < 0 rozkład jest mniej smukły niż rozkład normalny

Ku = 0, opisuje kształt rozkładu normalnego

Ku > 0 kształt rozkładu będzie bardziej smukły niż w przypadku rozkładu normalnego

# Kontynuujemy w przyszłym semestrze