Metody matematyczne w transporcie

Testowanie hipotez statystycznych
Testy zgodności

Marzena Nowakowska
Katedra Technologii Informatycznych
Wydział Zarządzania i Modelowania Komputerowego

Testowane hipotez: pojęcia podstawowe

Hipoteza statystyczna - każde przypuszczenie o populacji generalnej wydane bez przeprowadzenia wyczerpującego badania, dające się zweryfikować metodami statystycznymi na podstawie wyników badań próby.

Do weryfikacji hipotez statystycznych służy specjalna metoda postępowania zwana testem statystycznym, która z ustalonym prawdopodobieństwem dla danej próby przyporządkowuje decyzję o odrzuceniu lub przyjęciu sprawdzanej (weryfikowanej) hipotezy.

Rodzaje hipotez statystycznych

Hipoteza statystyczna

Parametryczna

Testy istotności
sprawdzają hipotezę, że dany parametr jest statystycznie istotny

Nieparametryczna

Testy zgodności
sprawdzają hipotezę, że populacja ma określony typ rozkładu

Testy sprawdzające,
czy 2 próby pochodzą z jednej populacji

. . .

Hipoteza **parametryczna** – dotyczy parametru rozkładu cechy populacji (założenie o rozkładzie populacji) - precyzuje wartość parametru w rozkładzie populacji gen.; test służący do jej zweryfikowania nazywa się testem parametrycznym.

Hipoteza **nieparametryczna** – nie odnosi się do parametrów rozkładu cechy; weryfikują ją testy nieparametryczne

Schemat podejmowania decyzji dotyczącej weryfikacji H0

Sytuacja $\Rightarrow$ **Decyzja** $\Downarrow$	H0 prawdziwa	H0 fałszywa
H0 przyjąć	decyzja słuszna	błąd II rodzaju , prawdopodobieństwo jego popełnienia = β
H0 odrzucić	błąd I rodzaju , prawdopodobieństwo jego popełnienia = α	decyzja słuszna

Najczęściej wybierane wartości α (poziomu istotności): 0,01; 0,05; 0,1
Zazwyczaj stosowana wartość β: 0,2

Błędy w testowaniu hipotez

Błąd pierwszego rodzaju (α) to odrzucenie hipotezy zerowej, mimo, że jest prawdziwa. Prawdopodobieństwo popełnienia tego błędu to **poziom istotności** α. Informuje on, na jak mały błąd „zgadzamy się" przy weryfikacji hipotezy zerowej. Poziom istotności określa dopuszczalną częstość wystąpienia wyników niezgodnych z przyjętymi założeniami na skutek losowego charakteru próby.

Błąd drugiego rodzaju (β) to przyjęcie hipotezy zerowej, gdy jest ona w rzeczywistości **fałszywa**. Prawdopodobieństwo popełnienia tego błędu jest oznaczone przez β. Wyznaczona przez β wartość **1-β** to **moc testu (prawdopodobieństwo uniknięcia błędu II rodzaju). Jest** to prawdopodobieństwo odrzucenia hipotezy zerowej, gdy jest ona **fałszywa**, a hipoteza alternatywna jest prawdziwa. W praktyce 1-β oznacza prawdopodobieństwo przyjęcia decyzji o istnieniu efektu, gdy on rzeczywiście istnieje. Testem najmocniejszym jest ten, dla którego, przy ustalonym poziomie istotności α**,** wartość β jest najmniejsza (1- β największa; zazwyczaj 0,8).

Moc testu zależy od :

1. wielkości spodziewanego efektu (testy parametryczne) – im większy efekt tym większa moc,
2. wielkości odchylenia standardowego cechy w populacji (zmienności pomiarów) – im mniejsze odchylenie (mniejsza zmienność pomiarów), tym większa moc,
3. liczebności próby – im liczniejsza próba, tym większa moc,
4. od poziomu istotności testu – im niższy poziom istotności, tym mniejsza moc testu.

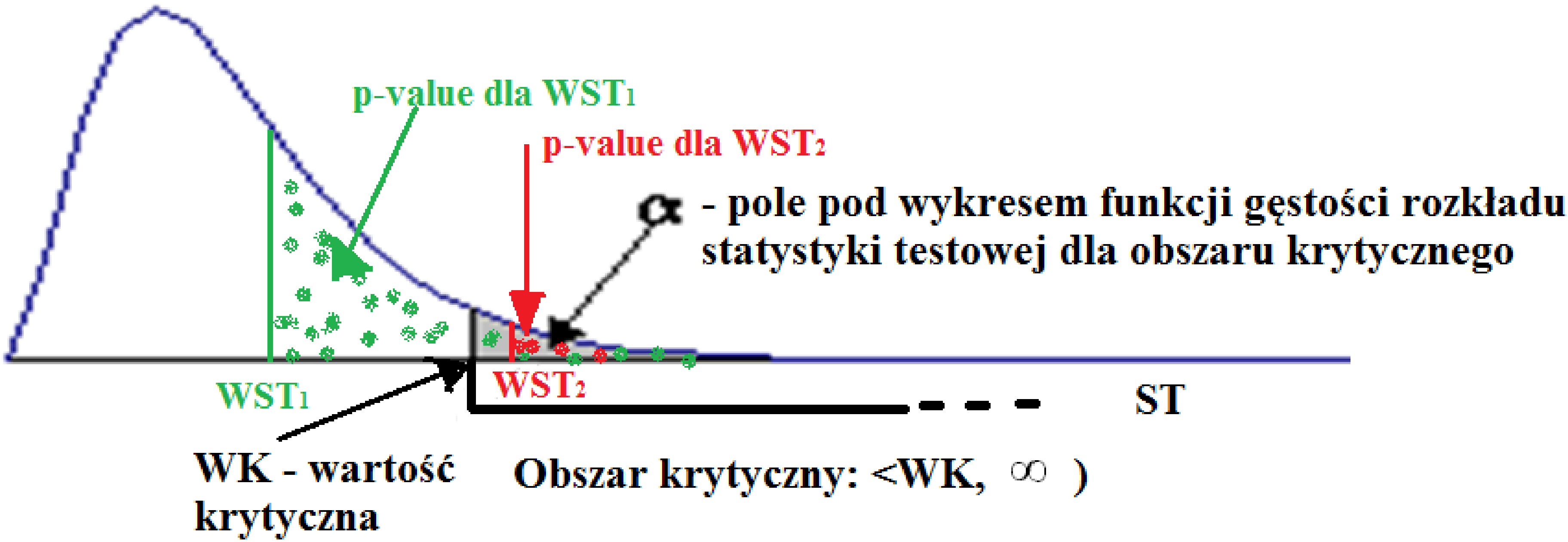
Procedura postępowania przy weryfikowaniu hipotez statystycznych

1. Należy podać treści hipotez – zerowej i alternatywnej:
   H0: *tu treść hipotezy zerowej*
   H1: *tu treść hipotezy alternatywnej*
2. Poziom istotności, który jest wybrany do weryfikacji H0
3. Informacja o tym, za pomocą jakiego testu jest weryfikowana H0
   Statystyka testowa jest zmienną losową:
   - rozkład statystyki testowej
   - obszar krytyczny i poziom istotności testu
   - wartość statystyki testowej *WST* dla danych z próby jest jedną z realizacji ww. statystyki testowej

   W trakcie procesu weryfikacji hipotezy zerowej ocenia się wartość statystyki testowej badając jej położenie względem obszaru krytycznego.
4. Wyznaczona wartość statystyki testowej oraz odpowiadające jej prawdopodobieństwo testowe (p-wartość, p-value)
5. Wniosek z badania obszaru krytycznego lub porównania p-value z α

Najczęściej spotykane rozkłady, którym podlegają statystyki testowe: normalny, t-Studenta, χ^2 (chi-kwadrat), F (Fishera).

Ilustracja graficzna podejmowania decyzji w testowaniu hipotez



Pole po wykresem funkcji gęstości rozkładu statystyki testowej jest prawdopodobieństwem (wartością z przedziału <0, 1>.
Wartości (przedziały) dla których są wyznaczone te prawdopodobieństwa to wartości statystyki testowej: wartość krytyczna, wartość statystyki testowej dla wybrane próby.

α – poziom istotności

p-value (p-wartość) – prawdopodobieństwo testowe

Podstawa decyzji w testowaniu hipotez – obszar krytyczny

Rozkład statystyki testowej *ST* jest wykorzystywany do wnioskowania o prawdziwości hipotezy zerowej. Na podstawie tego rozkładu i poziomu istotności α określa się przedziały wartości, które wyznaczają obszar odrzucenia lub nieodrzucenia hipotezy zerowej.

Obszar odrzucenia hipotezy statystycznej to taki przedział liczbowy, że jeżeli wyznaczona statystyka testowa *WST* przyjmie wartość należącą do niego, to hipotezę zerową należy odrzucić. Obszar odrzucenia określany jest jako **obszar krytyczny**.

Obszar nieodrzucenia (przyjęcia) hipotezy statystycznej to taki przedział liczbowy, że jeżeli wyznaczona statystyka testowa *WST* przyjmie wartość z tego przedziału, to nie ma podstaw do odrzucenia hipotezy zerowej.

Wartość krytyczna testu – wartość liczbowa wyznaczająca obszar krytyczny (koniec przedziału odrzucenia lub końce przedziałów dla testu dwustronnego).

Prawdopodobieństwo testowe i poziom istotności

Prawdopodobieństwo testowe (***p-wartość*** lub ***p-value***) jest prawdopodobieństwem wystąpienia, w powtarzanym próbkowaniu, wartości statystyki testowej takiej samej lub skrajniejszej (większej lub mniejszej – w kierunku obszaru krytycznego) jak wartość tej statystyki wyznaczona z badanej (w procesie weryfikacji hipotezy zerowej) próby.

Poziom istotności testu jest prawdopodobieństwem α odrzucenia hipotezy zerowej, mimo że jest prawdziwa.

Programy do analiz statystycznych, np. SAS (EG), wyznaczają *p-value* (*p wartość,* wartość p) Użytkownik wnioskuje o wyniku testowania porównując *p-value* i α.

Wnioskowanie

W przypadku testu istotności, do odrzucenia hipotezy zerowej wystarczy aby spełniony był jeden z podanych niżej warunków (oba są równoważne):

- **wartość statystyki testowej *WST* policzona z próby należy do obszaru krytycznego**

  Jeśli WST należy do obszaru krytycznego, to hipotezę zerową należy odrzucić, w przeciwnym przypadku nie ma podstaw do jej odrzucenia.

- **p-value** $\leq \alpha$

  α jest wartością przyjętą przez badacza

  Jeśli p-value $\leq \alpha$, to H0 należy odrzucić na korzyść H1. W przeciwnym przypadku formułuje się opinię, iż nie ma podstaw do odrzucenia H0.

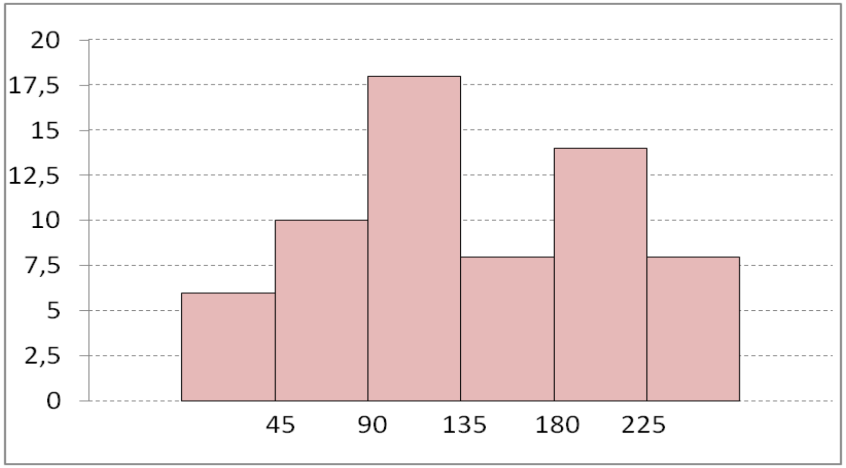
Test zgodności

Test zgodności jest wykorzystywany do sprawdzenia rozkładu cechy. Badana jest hipoteza H0, że cecha X ma rozkład zgodny z funkcją określonej klasy (zgodny z określonym rozkładem teoretycznym). Hipoteza alternatywna H1 głosi, że rozkład cechy X nie jest zgodny w ww. rozkładem teoretycznym, co jest równoznaczne ze stwierdzeniem, że dystrybuanta cechy (prawdopodobieństwo skumulowane) X nie należy do danej klasy dystrybuant.

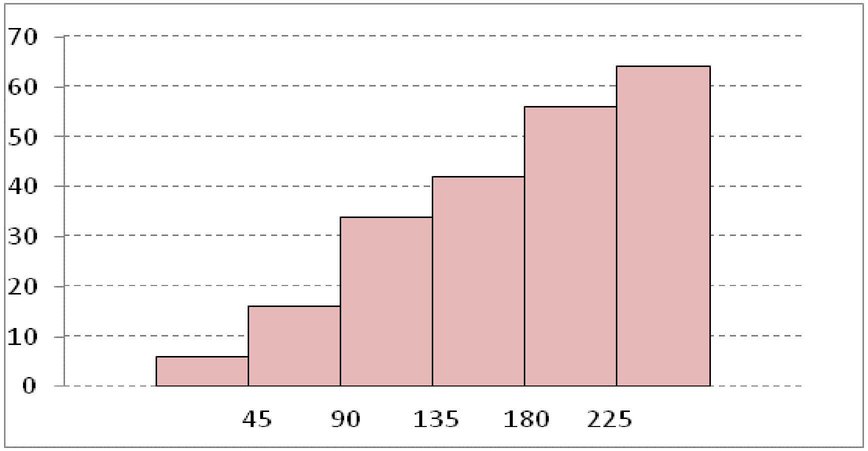
Jeśli hipoteza zerowa jest słuszna, to różnice między rozkładem testowanym (teoretycznym) i empirycznym nie powinny być znaczące (istotne).
Najpopularniejsze testy zgodności:
- test $\chi 2$
- test Kołmogorowa-Smirnowa
- test normalności: Shapiro-Wilka

Parametry badanego rozkładu teoretycznego zazwyczaj są estymowane z próby (estymatory próby lub MLE - max likelihood estimates)

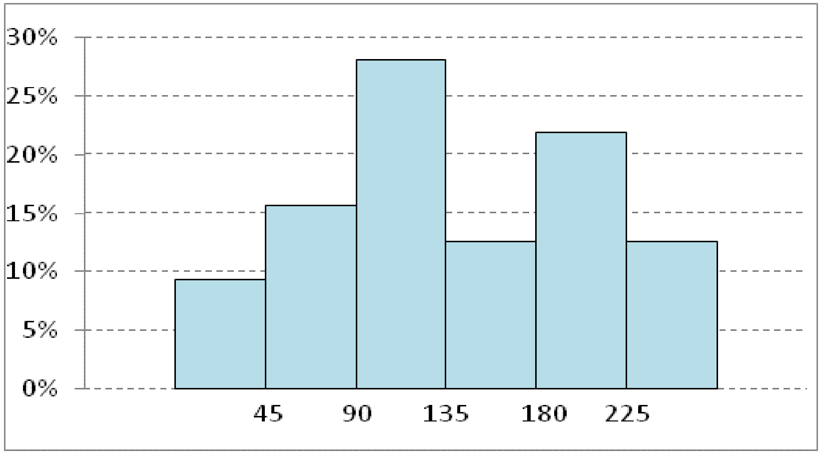
Histogram zmiennej losowej



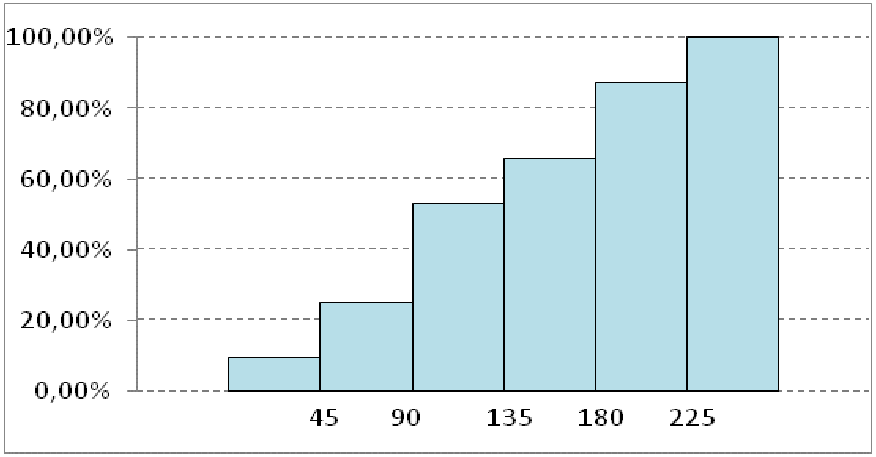
Histogram częstości (frequency)



Histogram częstości skumulowanej



Histogram gęstości (density)



Histogram gęstości skumulowanej

Obserwowana funkcja gęstości **Obserwowana funkcja gęstości skumulowanej**

Test χ^2

Porównuje się liczebności empiryczne i teoretyczne cechy w kolejnych przedziałach szeregu rozdzielczego. Analizowaną statyką testu jest zmienna, która ma graniczny rozkład chi-kwadrat χ^2 o $\nu = k\text{-}r\text{-}1$ stopniach swobody (gdy prawdziwa jest H0), gdzie *k* jest liczbą klas szeregu rozdzielczego, a *r* jest liczbą szacowanych parametrów.

Jeśli dla zadanego poziomu istotności statystyka testowa przekroczy wartość krytyczną $\chi^2(\alpha,\nu)$ (czyli znajdzie się w zbiorze krytycznym) hipotezę H0 należy odrzucić.

Wymaga się, by liczebności elementów próby w poszczególnych przedziałach klasowych przekraczały wartość 5.

Test Kołmogorowa-Smirnowa

Porównuje się wartości dystrybuant: empirycznej i teoretycznej, wyznaczonych dla kolejnych wartości cechy. Cecha musi być ciągła.

Analizowaną statyką testu jest zmienna, która ma rozkład λ Kołmogorowa, jeśli prawdziwa jest H0.

Jeśli dla zadanego poziomu istotności statystyka testowa przekroczy wartość krytyczną D_α (czyli znajdzie się w obszarze krytycznym) hipotezę H0 należy odrzucić.

Parametry rozkładu hipotetycznego powinny być znane (nie estymowane z próby). Gdy ten warunek nie jest spełniony, ten test można stosować, gdy próba jest duża (działa prawo wielkich liczb).

Prawo wielkich liczb: gdy liczebność próby rośnie nieograniczenie, to empiryczna dystrybuanta zmierza jednostajnie do dystrybuanty teoretycznej.

Test normalności Shapiro-Wilka

Umożliwia wyłącznie testowanie hipotezy zerowej, że **badana cecha ma rozkład normalny**. Funkcja gęstości takiego rozkładu dana jest wzorem :

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

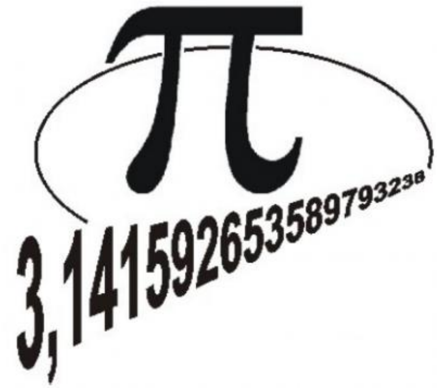
Gdzie: *m* i σ są parametrami rozkładu, odpowiednio: średnią i odchyleniem standardowym.

Jeśli wartość statystyki testowej dla określonego poziomu istotności α i liczebności próby *n* jest mniejsza od wartości krytycznej, H0 należy odrzucić.

Rozkład normalny

Zwany też rozkładem Gausa, jest rozkładem o największym znaczeniu w analizach statystycznych.

Funkcja gęstości jest dana zależnością:



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

Parametry:

μ (alternatywny symbol: *m*) - średnia,

σ - odchylenie standardowe.

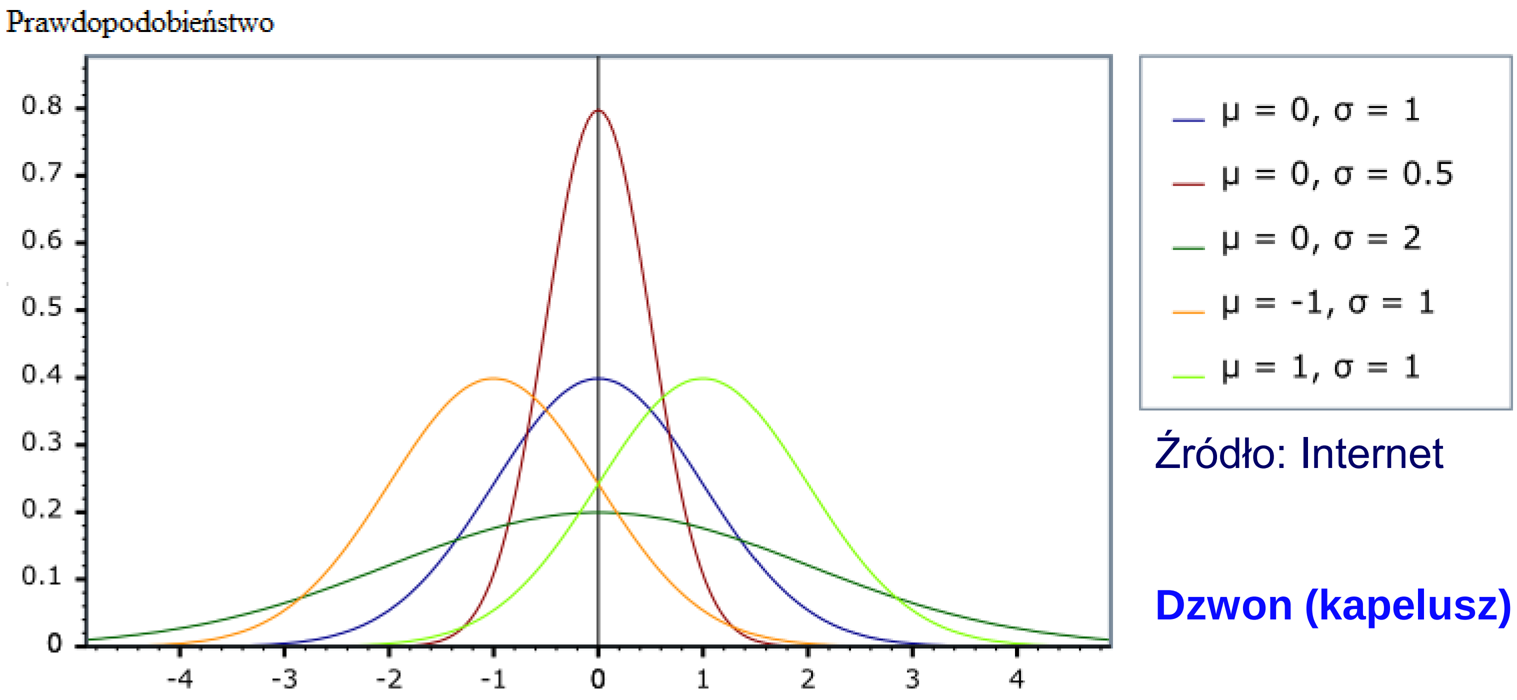
π =3,14159… (stosunek obwodu koła do średnicy)

e (exp) =2,71828… (stała Eulera)

x - dowolna wartość z przedziału $(-\infty, +\infty)$

Carl Friedrich Gauss (1777-1855), niemiecki matematyk, fizyk, astronom i geodeta, profesor uniwersytetu w Getyndze i dyrektor obserwatorium astronomicznego. Zaliczany do trójki największych matematyków świata, obok Archimedesa i Newtona.

Charakterystyka rozkładu normalnego



Źródło: Internet

Dzwon (kapelusz) Gaussa

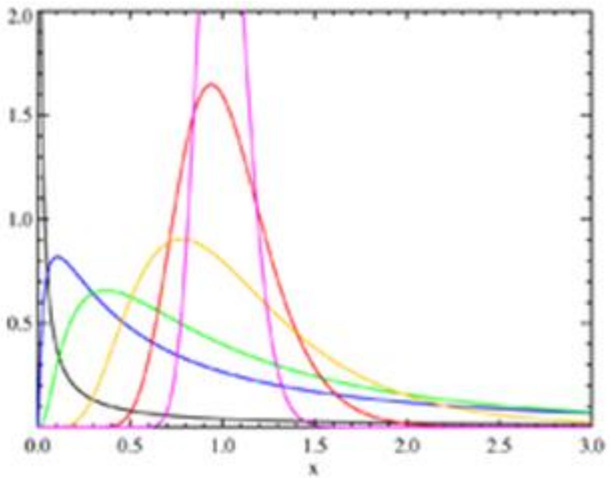
- Rozkład jest symetryczny; wartości cechy rozkładają się tak samo z lewej i prawej strony średniej; dominanta = mediana =średnia
- 68,27% wyników jest w przedziale (m-σ, m+σ)
- 95,45% wyników jest w przedziale (m-2σ, m+2σ)
- 99,73% wyników jest w przedziale (m-3σ, m+3σ)

Funkcje gęstości wybranych rozkładów teoretycznych dla zmiennej losowej ciągłej

Rozkład lognormalny

$$f(x)=\frac{1}{(x-\theta)\sigma\sqrt{2\pi}}exp\left(-0.5\left(\frac{log(x-\theta)-\zeta}{\sigma}\right)^2\right)$$

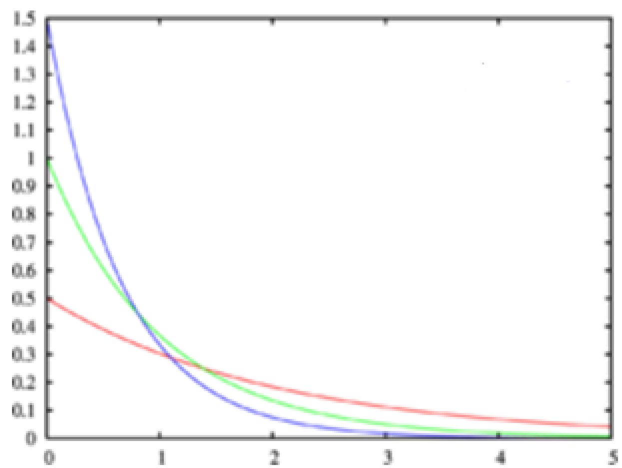
θ *Theta* wartość przesunięcia (progu)
ζ *Zetha* parametr kształtu
σ *Sigma* parametrem skali



Rozkład wykładniczy

$$f(x)=\frac{1}{\sigma}exp\left(-\frac{x-\theta}{\sigma}\right)$$

θ wartość przesunięcia (progu)
σ parametr skali.

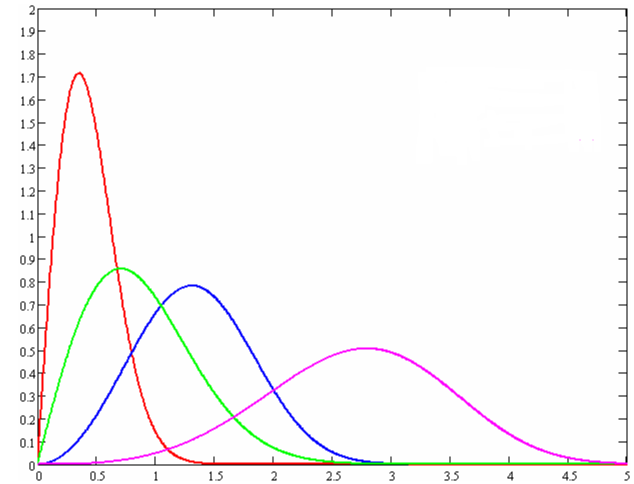


Funkcje gęstości wybranych rozkładów teoretycznych dla zmiennej losowej ciągłej

Rozkład Weibull'a

$$f(x) = \frac{c}{\sigma}\left(\frac{x-\theta}{\sigma}\right)^{c-1} exp\left(-\left(\frac{x-\theta}{\sigma}\right)^{c}\right)$$

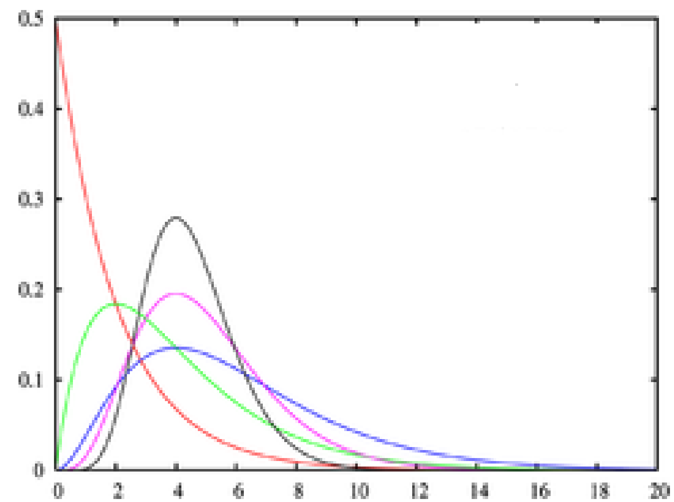
θ wartość przesunięcia (progu)
σ parametr skali
c parametr kształtu



Przesunięty rozkład gamma

$$f(x) = \frac{\lambda \cdot (\lambda \cdot (x-\theta))^{k-1} e^{-\lambda \cdot (x-\theta)}}{\Gamma(k)} \quad \text{dla} \quad x \geq \theta$$

Parametrami rozkładu są:
k – parametr kształtu,
λ - parametr skali
średnia = k/λ wariancja = k/λ^2
Dla rozkładu przesuniętego – dodatkowo:
θ - przesunięcie (próg);
przy braku przesunięcia $\theta = 0$



Funkcje gęstości wybranych rozkładów teoretycznych dla zmiennej losowej ciągłej

Rozkład beta

$$f(x) = \frac{1}{B} x^{r-1}(1-x)^{t-r-1} \quad \text{dla} \quad 0 \le x \le 1, \quad r > 0, \quad t - r > 0$$

B jest stałą normującą równą: $B = \Gamma(r) \cdot \Gamma(t-r)/\Gamma(t)$

średnia = r / t wariancja = $r \cdot (t-r)/[t^2 \cdot (t+1)]$

$\Gamma(k)$ jest funkcją gamma dana zależnością:

$$\Gamma(k) = \int_0^\infty e^{-u} u^{k-1} du$$

dla naturalnych k: $\Gamma(k) = (k-1)!$

Rozkład beta jest bardzo wygodnym narzędziem opisu danych doświadczalnych