**Transformacja danych:**

-Przekształcenie danych mających rozkład inny niż normalny do rozkładu normalnego

Przekształcenia liniowe - czyli mnożenie przez jedną liczbę i dodawanie kolejnej,
np. a x zmienna + b, gdzie *a*=3 a *b*=2.

Często stosowane przekształcenia:

- Arc sin $\sqrt{x}$

-Transformacja Boxa-Coxa $\frac{x^{λ}-1}{λ}$

- logarytmowanie, potęgowanie, pierwiastkowanie itp.

**Arc sin (tzw. transformacja Blissa)**

Stosuje się zazwyczaj dla danych mających rozkład dwumianowy wyrażonych w procentach,
przyjmujących najczęściej wartości w przedziale (0-20% lub 80-100%)

**Transformacja Boxa-Coxa**

Jest to często stosowana transformacja, w przypadku rozkładów asymetrycznych (lewostronnie lub prawostronnie skośnych lub też „uciętych” rozkładów normalnych)

Gdy dane mają rozkład Poissona można również zastosować transformację $\sqrt{x }$ lub $\sqrt{x+0,5}$
Zmienna transformowana może być np. liczbą bakterii w polu widzenia. Pierwiastkowanie przeprowadza się dla wartości zmiennych w zakresie 10<x<50, a stałą dodaje się dla zmiennych wyrażonych w wartościach poniżej 10.

**Logarytmowanie**

Stosujemy zazwyczaj w przypadku, gdy wraz ze wzrostem wartości średniej zwiększa się wariancja
(a tym samym odchylenie standardowe), czyli występuje korelacja między średnią a wariancją.

Stosowanie transformacji log(x) może nie być możliwe, np. w takim przypadku, jeśli zmienna przyjmuje wartość 0, wtedy można zastosować transformację log(x+1)

W przypadku rozkładów prawoskośnych dane można przekształcić poprzez log naturalny lub dziesiętny o podstawie 10.

W przypadku rozkładów lewoskośnych: log (a-x); gdzie a jest dobraną stałą większą od maksymalnej wartości z x a>max(xi)

**Pierwiastkowanie**

Stosuje się w przypadku rozkładów zbliżonych do rozkładu Poissona, tzn. w rozkładach prawostronnie skośnych, w których wartość średnia jest zbliżona do wariancji. Podobnie jak w przypadku transformacji log(x) może występować problem, jeśli zmienna przyjmuje wartość 0 (lub wartości ujemne). Można zastosować wtedy transformację $\sqrt{x+a}$ ; gdzie ***a*** jest określoną wartością np. a=0,5

**Problemy związane z transformacją**

1. Brak możliwości transformowania niektórych rozkładów do rozkładu normalnego,
np. nie da się przekształcić zmiennej skokowej do zmiennej ciągłej, tak więc w przypadku jeśli zmienna jest zmienną skokową (dyskretną), która przyjmuje niewielką liczbę wartości
(np. 1, 2, 3, 4 i 5) to niemożliwe jest zastosowanie transformacji, tak aby rozkład tej zmiennej był rozkładem normalnym
2. Trudności w interpretacji wyników. Ze względu na to, że po transformacji wartości parametrów (np. wartość średnia) ulegają zmianie, to nie można wnioskować np. o procentowej różnicy między średnimi na podstawie parametrów obliczonych na zmiennej transformowanej.

Polecenia:

Wykresy- wykresy tradycyjne- histogram- wyświetl krzywą normalną

Analiza- opis statystyczny- eksploracja- łącznie- wykresy- histogram- wykresy normalności z testami

Analiza- opis statystyczny- wykresy K-K (lub P-P)- przekształcenia

Analiza- testy nieparametryczne- jedna próba- Automatyczne porównanie danych empirycznych z hipotetycznymi- zmienne- ustawienia- Kołmogorow-Smirnoff

Przekształcenia- Kategoryzacja wizualna