

1. Rozkład prawdopodobieństwa:

W formie tabeli, każdy element wpisujemy tylko raz i wpisujemy prawdopodobieństwo jego wystąpienia, tzn:

i	i+1	i+2	i+3	...	i+n
$\frac{K_i}{L}$	$\frac{K_{i+1}}{L}$	itd	itd	itd	itd

Gdzie:

K_i - ilość wystąpień elementu

L - łączna ilość elementów

$$\sum_{i=1}^n \frac{K_i}{L} = 1 \text{ - Suma wszystkich } K_i/L \text{ ma się równać 1.}$$

2. Mediana:

$Me = \frac{(n+1)}{2}$ - Dla nieparzystej ilości elementów, czyli po prostu środkowy wyraz
 n - ilość elementów

Jak ilość elementów jest parzysta, to liczymy średnią arytmetyczną z dwóch środkowych elementów i otrzymujemy medianę.

3. Odchylenie standardowe:

Najpierw potrzebna jest wartość oczekiwana:

$$EX = \sum_{i=1}^n x_i * p_i \quad x_i - \text{element } i\text{-ty}$$

p_i - prawdopodobieństwo wystąpienia elementu x_i (bierzemy z rozkładu)

Następnie, samo odchylenie liczymy ze wzoru:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - u)^2}{n}} = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n} - u^2}$$

Gdzie:

σ - odchylenie standardowe

n - liczba obserwacji, czyli liczba elementów;

x_i - dany element;

u - wartość oczekiwana

4. Kowariancja:

$$\text{cov}(X,Y) = E(x*y)-(E(x)*E(y))$$

Gdzie:

$E(x*y)$ - średnia z iloczynu zmiennych (właściwie to wartość oczekiwana, ale w tym wypadku będzie to średnia arytmetyczna;

$E(x)$ - średnia zmiennej x ;

$E(y)$ - średnia zmiennej y ;

5. Korelacja:

$$r(x,y) = \frac{\text{cov}(X,Y)}{\sigma_x \sigma_y}$$

Gdzie:

$\text{cov}(X,Y)$ - kowariancja zmiennych X i Y

σ_x - odchylenie standardowe zmiennej x

σ_y - odchylenie standardowe zmiennej y

Wynik zawsze powinien być od -1 do 1. Jeśli jest inaczej to źle liczysz ;)

Interpretacja korelacji:

Jeżeli wyjdzie z przedziału od -0,5 do 0 lub od 0 do 0,5 to jest to słaba korelacja;

Jeżeli wyjdzie z przedziału od -1 do -0,5 lub od 0,5 do 1 to jest to silna korelacja;

Proponuję założyć, że mamy w excelu tabelę powiedzmy 50 osób. Dla każdej osoby dany jest wiek (zmienna x) i waga (zmienna y);

Nie mamy narzucone nic, a wydaje mi się, że taka tabela będzie prosta do ogarnięcia i wszystkie te obliczenia będą b. Proste.

Istotność:

Zakładamy poziom istotności. Przeważnie $\alpha = 0,05$

Następnie wykonujemy test t-studenta, na podstawie wzoru:

$$T = \frac{x_1 - x_2}{S_{x_1 - x_2}}$$

Gdzie:

X_1 - średnia arytmetyczna pierwszej grupy

X_2 - średnia arytmetyczna drugiej grupy

$$S_{x_1 - x_2} = \sqrt{\frac{(n_1-1)*s_1^2 + (n_2-1)*s_2^2}{n_1+n_2-2}} * \left(\frac{1}{n_1} + \frac{1}{n_2}\right)$$

Gdzie:

N_1 - liczebność pierwszej grupy

N_2 - liczebność drugiej grupy

S_1 - liczebność pierwszej grupy

S_2 - liczebność drugiej grupy

Wynik porównujemy z tabelą rozkładu t-studenta. Jeżeli wynik będzie równy bądź większy niż w tablicy to znaczy, że wynik jest istotny statystycznie. Jeżeli mniejszy, to będzie nieistotny statystycznie.