1 Wstęp:

1.1 Cechy X:

- mierzalne (ilościowe) - właściwości można zmierzyć i wyrazić za pomocą odpowiednich jednostek
- porządkowe (podtyp ilościowych) - badają natężenie badanej właściwości przedstawionej w sposób opisowy - np. oceny
- niemierzalne (jakościowe) - zwykle określane słownie nie mają jednoznacznych wartości liczbowych np. (płeć firma)

1.2 Rozstęp:

różnica między wartością maksymalną a minimalną cechy.
rozstęp między-kwantylowy (IQR). $Q_3 - Q_1$

1.3 Kwantyle Q:

Dzieli zbiorowość na części od względem ilości cech.

- Q_1 - dzieli na dwie części - w taki sposób że 25% ma wartości niższe a 75% wyższe.
- Q_2 połowa cech ma wartości mniejsze a połowa większe - mediana - wartość środkowa.
- Q_3 - 75% mniejszych cech, 25% większych.

1.4 Średnia:

Przybliżenie wartości oczekiwanej. Suma kwadratów odległości od średniej jest minimalna.

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

x_i - wartość i
n - ilość wartości

1.5 Wariancja s^2:

Średnia z wartości odchyleń od średniej arytmetycznej.

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

x_i - wartość i
n - ilość wartości

1.6 Odchylenie Standardowe σ:

Pierwiastek kwadratowy z wariancji - Stanowi miarę zróżnicowania, przeciętne zróżnicowanie wartości cechy od średniej arytmetycznej.

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

x_i - wartość i
n - ilość wartości
Około $\frac{2}{3}$ wszystkich jednostek cechy znajduje się w przedziale: $(\bar{x} - s, \bar{x} + s)$
Około 99% wszystkich wartości znajduje się w przedziale: $(\bar{x} - 3s, \bar{x} + 3s)$

2 Szeregi:

2.1 Szereg Szczegółowy:

uporządkowany ciąg wartości cechy statystycznej

2.2 Szereg Rozdzielczy:

uzyskuje się dzieląc dane statystyczne na pewne kategorie i podając liczebność danych w każdej kategorii - reprezentowane za pomocą histogramu. często zakłada się że $k = \sqrt{n}$
k - liczba przedziałów,
n - liczebność

2.2.1 Wariancja:

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

$\bar{x_i}$ -średnia wartości w przedziale i
n_i - ilość wartości w przedziale i

3 Prawdopodobieństwo:

$$P(B) = \sum_i P(A_i)P(B|A_i)$$

$P(B|A_i)$- prawdopodobieństwo B pod warunkiem A_i
$P(A_i)$ - prawdopodobieństwo A_i - rozdzielne warunki
$P(B)$ - prawdopodobieństwo B

Warunek unormowania prawdopodobieństw:

$$\sum_{k=0} P_k = 1$$

3.1 Wzór Bayesa:

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)}$$

$P(A|B)$ - prawdopodobieństwo A pod warunkiem B
$P(B|A)$- prawdopodobieństwo B pod warunkiem A
$P(A)$ - prawdopodobieństwo A
$P(B)$ - prawdopodobieństwo B

4 Zmienna Dyskretna:

Takie zmienne, które przyjmują skończony przeliczalny zbiór możliwych wartości. Właściwości: suma prawdopodobieństw wszystkich zmiennych $\sum_{i=1} p_i = 1$.

4.1 Dystrybuanta:

Funkcja przyporządkowująca skumulowaną wartość prawdopodobieństwa mniejszego od danego x

$$F(x) = P(X \leq x) = \sum_{x_i : x_i \leq x} p(x_i)$$

4.2 Wartość Oczekiwana:

Najbardziej prawdopodobna wartość, przyjmowana przez X

$$E(X) = \sum_{i=1}^{n} x_i p_i$$

p_i - prawdopodobieństwo i wartość
x_i - wartość i
n - ilość wartości

Własność wartości oczekiwanej:

$$E(aX + b) = aE(X) + b$$

$$E(X - E(X)) = 0$$

$$E(X + Y) = E(X) + E(Y)$$

$$E(XY) = E(X)E(Y)$$

4.3 Wariancja:

wartość przeciętna kwadratu odchylenia zmiennej losowej od jej wartości oczekiwanej.

$$V(X) = \sum_{i=1}^{N} (x_i - E(X))^2 p_i = E(X^2) - [E(X)]^2$$

$E(X)$ - wartość spodziewana
p_i - prawdopodobieństwo i wartość
x_i - wartość i
n - ilość wartości

Własności wariancji:

$$V(aX + b) = a^2 V(X)$$

$$V(X + Y) = V(X) + V(Y)$$

$$V(X - Y) = V(X) + V(Y)$$

4.4 Rozkład dwumianowy:

Prawdopodobieństwo k sukcesów przy N próbach.

$$P(X = k) = \binom{N}{k} p^k (1-p)^{N-k}$$

$$V(X) = Np(1-p)$$

$$E(X) = Np$$

$$F(X) = P(X \leq x) = \sum_{k \leq x} \binom{N}{k} p^k (1-p)^{N-k}$$

N - ilość prób
k - ilość sukcesów
p - prawdopodobieństwo sukcesu

4.5 Rozkład Poissona:

Prawdopodobieństwo k sukcesów przy dużej ilości prób.

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

k - ilość sukcesów
λ - wartość spodziewana

$$\lambda = E(x) = Np$$

5 Zmienna Ciągła:

Przyjmuje wartości rzeczywiste z określonego przedziału. Pole pod wykresem gęstości prawdopodobieństwa = 1

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

5.1 Dystrybuanta:

Prawdopodobieństwo przyjęcia wartości mniejszej od x jest pole od $-\infty$ do niej.

$$F(x_k) = \int_{-\infty}^{x_k} P(x)dx$$

$$\begin{aligned} &P(x) && \texttt{- prawdopodobienstwo wartości } x \\ &P(a < x \geq b) = F(b) - F(a) && \texttt{- prawdopodobienstwo wartości } x \end{aligned}$$

5.2 Wartość Oczekiwana μ :

$$E(x) = \int_{-\infty}^{\infty} xP(x)dx$$

$P(x)$ - prawdopodobieństwo przyjęcia wartości$x$

5.3 Wariancja:

$$V(X) = \int_{-\infty}^{\infty} [X - E(X)]^2 P(x)dx$$

5.4 Rozkład Normalny:

$$N(\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}} \left(\frac{x - \mu}{\sigma} \right)^2$$

$$\begin{aligned} N(\mu, \sigma) &\quad \texttt{- rozkład normalny} \\ N(0, 1) &\quad \texttt{- standardowy rozkład normalny} \\ \mu &\quad \texttt{- wartość oczekiwana} \\ \sigma &\quad \texttt{- odcylenie standardowe} \\ P(a < x \geq b) = F(b) - F(a) &\quad \texttt{- prawdopodobienstwo wartości } x \\ P(\mu - \sigma < x \geq \mu + \sigma) \approx 68\% &\quad \texttt{- prawdopodobienstwo wartości } x \\ P(\mu - 2\sigma < x \geq \mu + 2\sigma) \approx 95\% &\quad \texttt{- prawdopodobienstwo wartości } x \\ P(\mu - 3\sigma < x \geq \mu + 3\sigma) \approx 99\% &\quad \texttt{- prawdopodobienstwo wartości } x \end{aligned}$$

5.4.1 Standaryzacja Rozkładu normalnego

$X = N(\mu, \sigma)$- `Dane o podanym rozkładznie normalnym`

$Z = N(0, 1)$- `dane o normalnym rozkładzie standardowym`

$Z = \frac{X - \mu}{\sigma} = N(0, 1)$