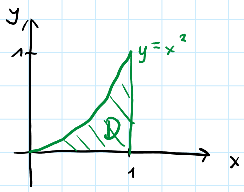
PRZYKŁAD DO PONIŻSZEGO

$$f(x,y) = \begin{cases} 12xy, & 0 \leqslant x \leqslant 1 \\ & 0 \leqslant y \leqslant x^2 \\ 0 & \text{poza} \end{cases}$$



$$f_x(x) = \begin{cases} 6x^5 & x \in \langle 0,1 \rangle \\ 0 & \text{poza} \end{cases}$$

KRZYWA REGRESJI

- (X,Y) - X,Y zmienne losowe niezależne
- $f(X,Y)$ - gęstość łączna
- gęstością warunkową zmiennej losowej Y gdy $X = x$ nazywamy funkcję

$$f_{(Y|X)}(y|x) = \frac{f(x,y)}{f_x(x)}$$

- WARUNKOWĄ WARTOŚCIĄ OCZEKIWANĄ zmiennej losowej Y względem X, gdy $X = x$ nazywamy liczbę

$$E(Y|X=x) = \int_{-\infty}^{\infty} y f_{(Y|X)}(y|x)\,dy$$

zależną od X

- FUNKCJĄ REGRESJI (zmiennej losowej Y względem X) nazywamy funkcję

$$m_y(x) = E(Y|X=x)$$

  - jeżeli X, Y są niezależne, to $E(Y|X=x) = EY$ dla każdego x
  - dla dowolnej funkcji $h(x)$ zachodzi

$$\min E(Y - h(X))^2 = E(Y - E(Y|X))^2$$

- PRZYKŁAD: Wyznaczyć funkcję regresji zmiennej losowej Y względem zmiennej losowej X dla danych z ostatniego przykładu

$$f_{(Y|X)}(y|x) = \frac{f(x,y)}{f_x(x)} = \frac{12xy}{6x^5} = \frac{2y}{x^4} \qquad (x,y) \in D \qquad x \neq 0$$

$$E(Y|X=x) = \int_{-\infty}^{\infty} y f_{(Y|X)}(y|x)\,dy = \int_0^{x^2} y\frac{2y}{x^4}\,dy = \frac{2}{x^4}\cdot \left.\frac{y^3}{3}\right|_{y=0}^{x^2} = \frac{2}{x^4}\left(\frac{x^6}{3} - 0\right) = \frac{2}{3}x^2$$

$$m_y(x) = \begin{cases} \frac{2}{3}x^2 & x \in (0,1) \\ 0 & \text{...} \end{cases}$$

$$m_y(x) = \begin{cases} \frac{2}{3}x^2 & x \in (0,1) \\ 0 & \text{poza} \end{cases}$$

np. dla $x = \frac{1}{2}$ $\rightarrow$ $m_y(x) = m_y\left(\frac{1}{2}\right) = \frac{2}{3}\left(\frac{1}{2}\right)^2 = \frac{1}{6}$

WSTĘPNA ANALIZA DANYCH STATYSTYCZNYCH

- PRÓBA PROSTA LOSOWA długości n $(n \in \mathbb{N})$ to układ n niezależnych zmiennych losowych o jednakowym rozkładzie

$(X_1, X_2 \dots X_n)$, X_k $k=1,\dots,n$ – taki sam rozkład prawdopodobieństwa

$(x_1, x_2 \dots x_n)$ x_n – obserwacja (realizacja) próby losowej $(X_1 \dots X_n)$

- PRZYKŁAD: w jednakowych 27 próbkach dielektryka określono zawartość tytanianu baru i otrzymano następujące wyniki (w mg):

x_1 8.31, 8.33, 8.36, 8.41, 8.42, 8.42, 8.44, 8.46, 8.47,
8.48, 8.48, 8.48 8.48, 8.49, 8.50, 8.50, 8.50 8.51,
8.55, 8.56, 8.58, 8.58, 8.62, 8.62, 8.65, 8.70, 8.70, x_{27}

X_k – zawartość tytanianu baru w k-tej próbce, $k = 1 \dots 27$

- SZEREG ROZDZIELCZY

x_{min}, x_{max} $\qquad$ $x_{min} = 8{,}31$ $\qquad$ $x_{max} = 8{,}70$

$r = x_{max} - x_{min}$ $\qquad$ $r = 0{,}39$ $\qquad$ $h = \frac{0{,}39}{5} \approx 0{,}08$

k – liczba klas dla $n \leq 50$:
$\frac{\sqrt{n}}{2} \leq k \leq \sqrt{n}$

k powinno być ≥ 5 i ≤ 30

h – szerokość klasy

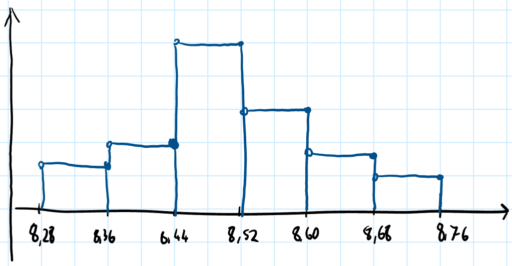
$h = \frac{r}{k}$

dla $n = 27$, $k = 5$ ← w rzeczywistości 1 klasa więcej, bo rozszerzyliśmy przedziały szersze niż zakładane

chcemy, żeby 1. obserwacja była gdzieś kawałek za środkiem klasy, nie na samym początku

klasa (i)	(8.28, 8.36]	(8,36, 8.44]	(8.44, 8,52]	(8,52, 8.60]	(8.60, 8,68]	(8.68, 8.76]
liczebność klasy n_i	3	4	11	4	3	2
częstość klasy $\frac{n_i}{n}$	$\frac{3}{27}$	$\frac{4}{27}$	$\frac{11}{27}$	$\frac{4}{27}$	$\frac{3}{27}$	$\frac{2}{27}$

- HISTOGRAM – graficzne przedstawienie szeregu rozdzielczego



- funkcję określoną na próbie losowej $(X_1, X_2 \dots X_n)$ nazywamy STATYSTYKĄ
- WAŻNIEJSZE STATYSTYKI:

① Średnia z próby

$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ $\qquad \bar{x} =$

② Wariancja z próby

$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right) - (\bar{X})^2$

$S_1^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

} zależne od autora pakietu

③ Dyspersja z próby (odchylenie standardowe)

$S = \sqrt{S^2} \qquad S_1 = \sqrt{S_1^2}$

④ Moment rzędu k

$M_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$

moment centralny rzędu k: $C_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k$

⑤ Współczynnik zmienności

$v = \frac{S}{\bar{X}} \cdot 100\%$

⑥ Mediana z próby

ustawiamy próbę $(X_1, X_2 \ldots X_n)$ w ciąg niemalejący $X_n \leq X_{n+1} \leq X_{n+2} \ldots$

$$Me = \begin{cases} X_{\left(\frac{n+1}{2}\right)} & n \text{ nieparzyste} \\ \frac{X_{\frac{n}{2}} + X_{\left(\frac{n}{2}\right)+1}}{2} & n \text{ parzyste} \end{cases}$$

⑦ Kwartyl dolny Q_1

Q_1 – mediana z obserwacji $\leq Me$

⑧ Kwartyl górny Q_3

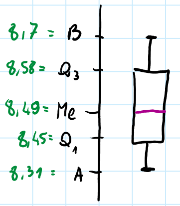
Q_3 – środkowa (mediana) z obserwacji $> Me$

w przedziale $[Q_1, Q_3]$ jest około połowy obserwacji

⑨ Odstęp międzykwartylowy

$IRQ = Q_3 - Q_1$

⑩ Wykres ramkowy



$B = \min\{X_{max}, Q_3 + \frac{3}{2} IRQ\}$

$A = \max\{X_{min}, Q_1 - \frac{3}{2} IRQ\}$

Obserwacje odstające – obserwacje poza przedziałem $[A, B]$

• obliczenia dla przykładu:

$\bar{x} = \frac{1}{27}(8{,}31 \ldots + 8{,}70) = 8.503$

- obliczenia dla przykładu:

$\bar{x} = \frac{1}{27}(8{,}31 \ldots + 8{,}70) = 8.503$

$S^2 = \frac{1}{27}\left((8{,}31)^2 + \ldots + (8{,}70)^2\right) - (8.503)^2 = 0{,}011$

$S = 0{,}104$

$V = \frac{S}{\bar{x}} \cdot 100\% = \frac{0{,}104}{8{,}508} \cdot 100\% = 1{,}2\%$

$Me = x_{14} = 8.49$

$Q_1 = \frac{x_7 + x_8}{2} = 8{,}45$

$Q_3 = x_{21} = 8{,}58$

$A = \max\{8.31,\ 8.45 - \frac{3}{2} \cdot 0{,}13\} = 8{,}31$

$B = \min\{8.7,\ 8.58 + \frac{3}{2} \cdot 0{,}13\} = 8{,}7$