Disuguaglianza di Cauchy - Schwarz

$u_1 \dots u_n \,,\, v_1 \dots v_n \in \mathbb{R}$

$$\left(\sum_{i=1}^{n} u_i v_i\right)^2 \leq \left(\sum_{i=1}^{n} u_i^2\right)\left(\sum_{i=1}^{n} v_i^2\right)$$

$$0 \leq \sum_{i=1}^{n}\sum_{j=1}^{n}(u_i v_j - u_j v_i)^2 = \sum_i \sum_j \left(u_i^2 v_j^2 + u_j^2 v_i^2 - 2 u_i u_j v_i v_j\right) =$$
$$= \sum_i u_i^2 \sum_j v_j^2 + \sum_j u_j^2 \sum_i v_i^2 - 2\left(\sum_i u_i v_i\right)\left(\sum_j u_j v_j\right) =$$
$$= 2\left(\sum_i u_i^2\right)\left(\sum_j v_j^2\right) - 2\left(\sum_i u_i v_i\right)^2$$

Note sugli Integrali

$$\left(\int_a^b f(x)\, g(x)\, dx\right)^2 \leq \int_a^b f^2(x)\, dx \int_a^b g^2(x)\, dx$$

A patto che tutto CONVERGA

Dimostrazione di 2.2.1

Abbiamo il minimo
$Q(a,b) = \sum_{i=1}^{n} (y_i - a x_i - b)^2$ $\longrightarrow$ Troviamo i punti (a,b) stazionari di Q (DERIVANDO)

Nota 2.2.1

$$r(x,y)^2 = 1 - \underbrace{\frac{\overbrace{\sum (y_i - a^* x_i - b^*)^2}^{\text{explained variance}}}{\sum (y_i - \bar{y})^2}}_{\text{unexplained variance}}$$

$$\begin{cases} \emptyset = \dfrac{\partial Q}{\partial a} = \sum_{i=1}^{n} 2(y_i - a x_i - b)(-x_i) \\ \emptyset = \dfrac{\partial Q}{\partial b} = \sum_{i=1}^{n} 2(y_i - a x_i - b)(-1) \end{cases}$$

$$\begin{cases} \emptyset = \sum y_i x_i - a \sum x_i^2 - b \sum x_i \\ \emptyset = \sum y_i - a \sum x_i - n b \end{cases}$$

N.B. $\bar{x} = \frac{1}{n} \sum x_i$

$$\begin{cases} \bar{y} = a \bar{x} + b \\ \emptyset = \sum x_i y_i - a \sum x_i^2 - b \bar{x} n \end{cases}$$

. . . ("Fidatevi, non ho voglia", cit. il prof.)

$$a = \frac{\operatorname{cov}(x,y)}{\operatorname{var}(x)} \qquad b = \bar{y} - a\bar{x}$$

Esercizio 1

Mostrare che $var(x) = \frac{n}{n-1} \sum_{i=1}^{n} \left(\frac{x_i^2}{n} - \bar{x}^2 \right)$

Sfruttiamo la definizione di VARIANZA $\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i^2 + \bar{x}^2 - 2 x_i \bar{x}) =$

$= \frac{1}{n-1} \left(\sum x_i^2 + n \bar{x}^2 - 2\bar{x} \sum x_i \right) = \frac{n}{n-1} \left(\frac{1}{n} \sum x_i^2 - \bar{x}^2 \right)$

Facciamo la stessa cosa con la COVARIANZA $cov(x,y) = \frac{1}{n-1} \sum^{n} (x_i y_i - \bar{x} y_i - x_i \bar{y} + \bar{x}\bar{y}) = \frac{n}{n-1} \left(\left(\sum \frac{x_i y_i}{n} \right) - \bar{x}\bar{y} - \bar{x}\bar{y} + \bar{x}\bar{y} \right) =$

$= \frac{n}{n-1} \left(\sum \frac{x_i y_i}{n} - \bar{x}\bar{y} \right)$

N.B. Il caso della COVARIANZA è più generale di quello della VARIANZA (basta porre $\bar{x} = \bar{y}$)

Esercizio 2

Se $z_i = x_i + c$ $(c \in \mathbb{R})$ allora $\sigma(z) = \sigma(x)$

(dati traslati di uno stesso numero)

$$\sigma(z)^2 = \frac{1}{n-1} \sum (z_i - \bar{z})^2 = \frac{1}{n-1} \sum (x_i + c - \bar{x} - c)^2 = \sigma(x)^2$$

$$\bar{z} = \frac{1}{n} \sum z_i = \frac{1}{n} \sum (x_i + c) = \bar{x} + c$$

Se $w_i = a x_i$ $(a \neq 0)$ allora $\sigma(w) = a \cdot \sigma(x)$

$$\sigma(w) = \frac{1}{n-1} \sum (w_i - \bar{w})^2 = \frac{1}{n-1} \sum (a x_i - a\bar{x})^2 = a^2 \sigma(x)^2$$

Citazione del professore: un'equazione è quando un cazzo è uguale ad un altro cazzo

UniPi - 2024

Esercizio 3

Se $a, b, c, d \in \mathbb{R}$ con $a, c \neq 0$ allora $r(ax+b, cy+d) = r(x,y)$

$$r(x,y) = \frac{cov(x,y)}{\sigma(x)\,\sigma(y)}$$

N.B. Dall'esercizio 2 $\sigma(ax+b) = a\,\sigma(x)$
$\sigma(cy+d) = c\,\sigma(y)$

$$cov(x,y) = \frac{1}{n-1}\sum (ax_i + b - a\bar{x} - b)(cy + d - c\bar{y} - d) = \frac{ac}{n-1}\sum (x_i - \bar{x})(y_i - \bar{y}) = ac \cdot cov(x\,y)$$

quindi ... $r(ax+b, cy+d) = \frac{\cancel{ac}\; cov(x,y)}{\cancel{a}\,\sigma(x)\,\cancel{c}\,\sigma(y)} = r(x,y)$

Esercizio 4

I dati $x_1 \dots x_n$ assumono valori $a_1 \dots a_m$ con frequenze $p(a_j, x) = \frac{\#\{i : x_i = a_j\}}{n}$

e.g.

$x_1 = 6$	$a_1 = 1$
$x_2 = 2$	$a_2 = 2$
$x_3 = 3$	$a_3 = 1$
$x_4 = 2$	$a_4 = 0$
$x_5 = 1$	$a_5 = 1$
$x_6 = 5$	$a_6 = 1$

mostrare che $\bar{x} = \sum_{j=1}^{m} a_j\, p(a_j, x)$ ①

e $var(x) = \frac{n}{n-1}\left(\left(\sum_{j=1}^{m} a_j^2\, p(a_j, x)\right) - \bar{x}^2\right)$ ②

① $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\left(\sum_{i : x_i = a_1} x_i + \sum_{i : x_i = a_2} x_i + \dots\right) =$

$= \frac{1}{n}\left(\sum_{i : x_i = a_1} a_1 + \dots\right) = \frac{1}{n}\left(a_1 \cdot \#\{i : x_i = a_1\} + a_2 \cdot \# \dots\right)$

$= a_1\, p(a_1, x) + a_2\, p(a_2, x) + \dots$

② Sappiamo che $var(x) = \frac{n}{n-1}\left(\sum \frac{x_i^2}{n} - \bar{x}^2\right)$
e posso applicare lo stesso ragionamento di ①

Esercizio in regressione lineare

$$\sum_{i=1}^{n} (y_i - a^* x_i - b) = \emptyset$$

TODO

Esercizio 5

Consideriamo $\delta = \frac{1}{n} \sum_{i=1}^{n} |x_i - \bar{x}|$ (MAD - Mean Absolute Deviation)

N.B. Il prof ci tiene a specificare che
MAD ≠ INCAZZATO

Mostrare che $\underbrace{\frac{\#\{i : |x_i - \bar{x}| > d\}}{n} \leq \frac{\delta}{d}}_{1}$ e $\underbrace{\delta \leq \sqrt{\frac{n}{n-1}}\, \sigma}_{2}$

N.B. Sappiamo già la DISUGUAGLIANZA di CHEBYSHEV (ЧЕБЫШЁВ)

$$\frac{\#\{i : |x_i - \bar{x}| > d\}}{n-1} \leq \frac{var(x)}{d^2}$$

1) $\delta(x) = \frac{1}{n} \sum_{i=1}^{n} |x_i - \bar{x}| \geq \frac{1}{n} \sum_{i : |x_i - \bar{x}| > d} d = \frac{d}{n} \#\{i : |x_i - \bar{x}| > d\}$

2) È una conseguenza delle DISUGUAGLIANZE TRA MEDIE

se $a_1 \dots a_n \geq 0$ allora

$$\sqrt[n]{a_1 \dots a_n} \leq \frac{1}{n} \sum_{i=1}^{n} a_i \leq \sqrt{\frac{1}{n} \sum_{i=1}^{n} a_i^2} \leq \sqrt[3]{\frac{1}{n} \sum_{i=1}^{n} a_i^3} \leq \dots \leq \max_{i=1 \dots n} a_i$$

Cercatevi la dimostrazione su WIKIPEDIA