2012-10-12 高等教育統計

標準分數 Z

位移

設資料 $x_1, x_2, x_3 \cdots x_n$ 將之標準化

$$z_i \triangleq \frac{\boxed{x_i - \bar{x}}}{\boxed{s}} \quad i=1,2\cdots n$$

改變衡量的單位

[若母體資料為 $\frac{x_i - \mu}{\sigma}$]

稱之為標準分數

性質1: $\bar{z} = \frac{z_1 + z_2 \cdots z_n}{n} = 0$

$\because \bar{x} = 0$

性質2: $s_z^2 \triangleq \frac{\sum_{i=1} z_i^2}{n-1} = n$

<pf> 1. $z_1 + \cdots + z_n$

$$= \frac{1}{s}[(x_1 - \bar{x}) + (x_2 - \bar{x}) \cdots + (x_n - \bar{x})]$$

$$= \frac{1}{s}\left(\sum_{i=1}^{n} x_i - n\bar{x}\right) = 0$$

pf2

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

$$= n-1$$

$$\therefore \frac{\sum_{i=1}^{n} z_i^2}{n-1} = 1$$

很少超過 3, −3

註：通常 $|z_i| \le 3$

例：數學平均為 $\bar{x}=64.5$，$s=8.65$，甲得85，乙得58，他們的 z 分數各為多少？

解：

甲：$z = \dfrac{85-64.5}{8.65} = 2.37$

乙：$z = \dfrac{58-64.5}{8.65} = -0.75$

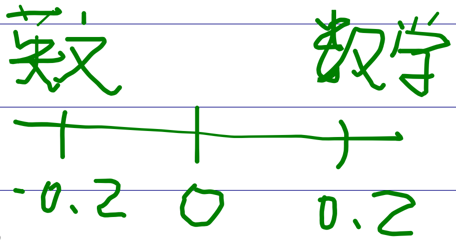
例:

数学平均 $\bar{x}=50,\ S=10$

英文平均 $\bar{y}=60,\ S=10$

甲 $x=52,\quad y=58$

则 $z_x=\frac{52-50}{10}=0.2$



$$z_y=\frac{58-60}{10}=-0.2$$

原本是 $x_1, x_2 \cdots\cdots x_n$

若令 $z_i=az_i+b \quad i=1,2,\cdots\cdots n$

将调整成

平均数是 b

标准差 a

则 $\bar{z}=\frac{z_1+z_2+\cdots+z_n}{n}$

$$=\frac{a(z_1+z_2\cdots+z_n)+nb}{n}$$

$$S^2=\frac{\sum_{i=1}^{n}(z_i-\bar{z})^2}{n-1}$$

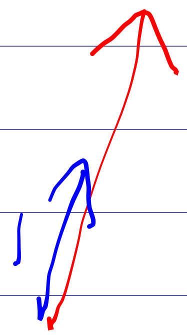
$$=\frac{\sum_{i=1}^{n}(az_i+b-b)^2}{n-1}$$

$$=\frac{a^2\sum_{i=1}^{n}z_i}{n-1}=a^2$$

① 標準化這個動作，也可以用向量來說明。

$$\vec{a} = \langle a_1, a_2 \cdots a_n \rangle$$

$\vec{a}$



長度為 $|\vec{a}|$

$$\therefore \hat{a} = \frac{\vec{a}}{|\vec{a}|}$$

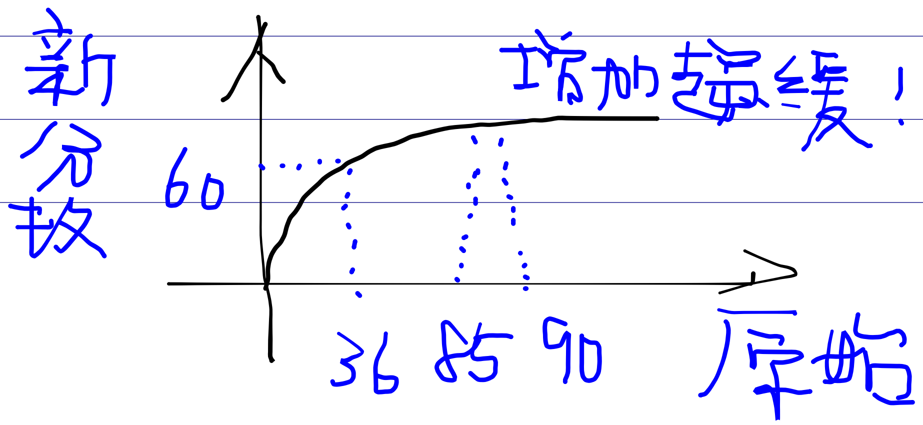
② 標準化之後，單位消失，變成純量

③ 調整分數的方法

$$z_i = a z_i + b$$

或 同加、同減

或 $\sqrt{x} \times 10$ (懲罰好學生)



Q:

$x_1 \cdots\cdots x_m \qquad \sigma_x \qquad \bar{x}$

$y_1 \cdots\cdots y_n \qquad \sigma_y \qquad \bar{y}$

將資料合併後，
有沒有可能合併之後

新的 $\sigma_Z < \min(\sigma_x, \sigma_Y)$?

請証明：(Hint：

用變異數來解)

$X_1, X_2, X_3, \cdots\cdots X_n$

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

C 是 $1 \cdots\cdots n$

任一個數字

CH6 常態分配

1. 在自然的情形下，只要資料夠大通常資料會呈常態分配，如智商、體重、身高之分配。

2. 常態分配之特性

a. 分配圖形對稱中心點 μ，往兩邊遞減且無限延伸

b. 資料落在 $[\mu-\sigma, \mu+\sigma]$ 約 68%

$[\mu-2\sigma, \mu+2\sigma]$ 約 95%

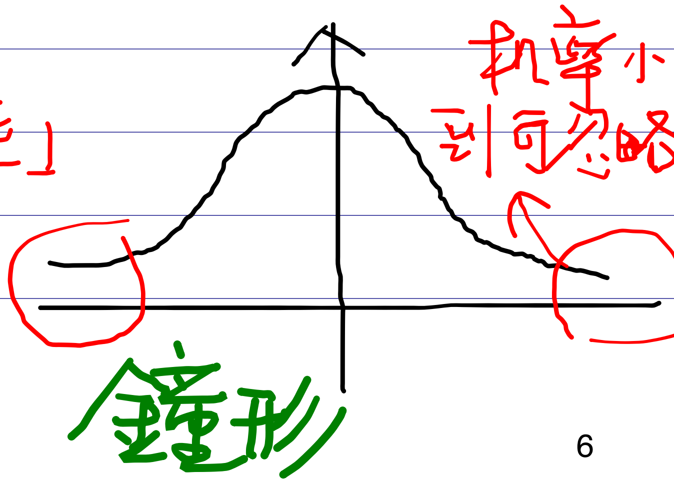
$[\mu-3\sigma, \mu+3\sigma]$ 約 99%

因此 $|Z_i| \le 3$

$|Z_i| > 3$ 幾乎不可能!!

為了「完備性」

※ 機率總和 = 1



Why 要學「常態分配」

中央極限定理

因為有它！

最重要的定理。

資料有 離散 / 連續