一、一維數據分析

1.算術平均數、加權平均數、中位數、眾數、全距、四分位差→略

2.變異數➪離均差平方的平均

平方的平均減平均的平方(平方bar減bar平方)→比第1個好用

3.標準差➪$\sqrt{變異數}$

其中:1.以算術平均數為中心的標準差，較任何其它平均數為中心的標準差小

2.若標準差為0→表所有資料均同質

3.標準差愈大資料愈分散

4.不受平移影響，但受伸縮影響

4.數據標準化➪$\frac{x-\overline{x}}{標準差}$

其中:標準化數據的平均數=0，標準差=1

二、二為數據分析

1.相關係數

(1)公式整理

有標準差σ_x、σ_y時➪$\frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{n.\sigma_x.\sigma_y}$

沒標準差時➪$\dfrac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})^2}\cdot\sqrt{\sum_{i=1}^{n}(y_i-\overline{y})^2}}$

最簡捷算法➪$\dfrac{\overline{xy}-\overline{x}\,\overline{y}}{\sigma_x\cdot\sigma_y}$

(2)性質

相關係數與單位無關

易受少數極端數據影響

不受平移伸縮影響，但受對稱影響

2.回歸直線

(1)最小平方法➪

對給定有限多個數對

要求出一個線性函數g(x)=a+bx，使得誤差的平方和

E=$(g(x_1)-y_1)^2+(g(x_2)-y_2)^2+\ldots+(g(x_n)-y_n)^2$為最小

(2)其他表示方法

$(y-\overline{y})=r.\dfrac{\sigma_x}{\sigma_y}(x-\overline{x})$

其中:斜率為$r.\frac{\sigma_x}{\sigma_y}=\frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sum_{i=1}^{n}(x_i-\overline{x})^{2}}$

必過中心點$(\overline{x},\ \overline{y})$

(3) $(y-\overline{y})=r.\frac{\sigma_x}{\sigma_y}(x-\overline{x})$ ⇨ $\frac{y-\overline{y}}{\sigma_y}=r.\frac{x-\overline{x}}{\sigma_x}$

其中:標準化資料的最佳直線過原點且斜率等於相關係數