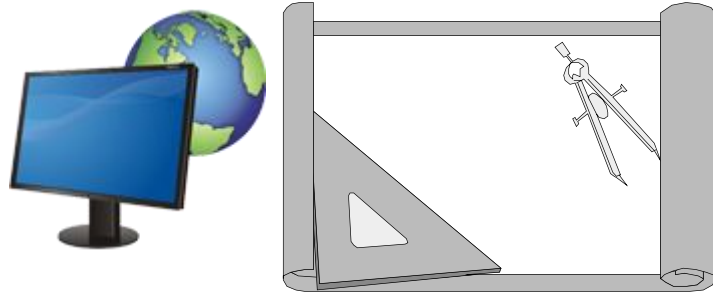




第二章 地理数据

——河北师范大学资环学院 胡引翠

计量地理学？



计量地理学？

计量地理学是将数学和电子计算机技术应用于地理学的一门新兴学科。

大量准确的数据对计量地理学研究很重要。

//



地理数据在建模分析中的作用：

- ①确定模型中的参数与初值；**
- ②检验模型的正确性、合理性和有效性。**

本章主要内容：关于数据的故事

地理数据类型

地理数据特征

采集与处理

数据库建立

地理数据类型

定性数据

定量数据

地理数据类型

空间数据

属性数据

本章主要内容：关于数据的故事

地理数据类型

地理数据特征

采集与处理

数据库建立

地理数据特征

形式化数量化

多时空尺度

多维性

不确定性

形式化、逻辑化与数量化，是所有地理数据的共同特征。

形式化：可以用数学的符号语言、图像语言和文字语言，表达自然现象和社会现象的空间结构和数量关系。

本章主要内容：关于数据的故事

地理数据类型

地理数据特征

采集与处理

数据库建立

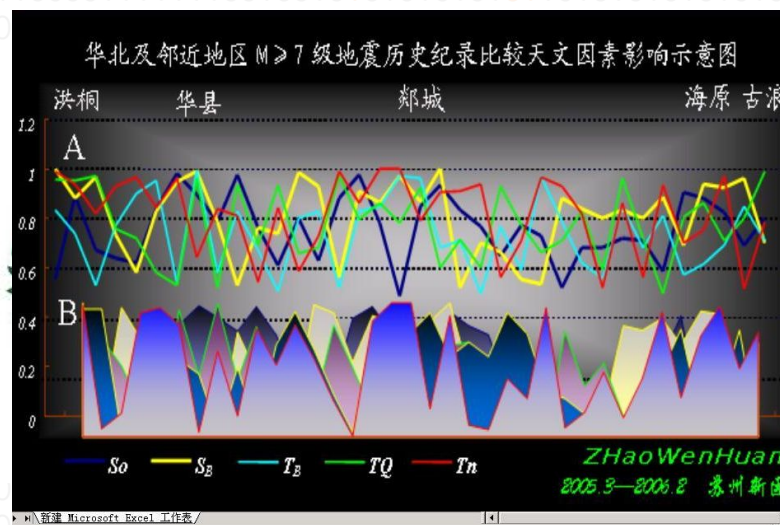
地理数据特征

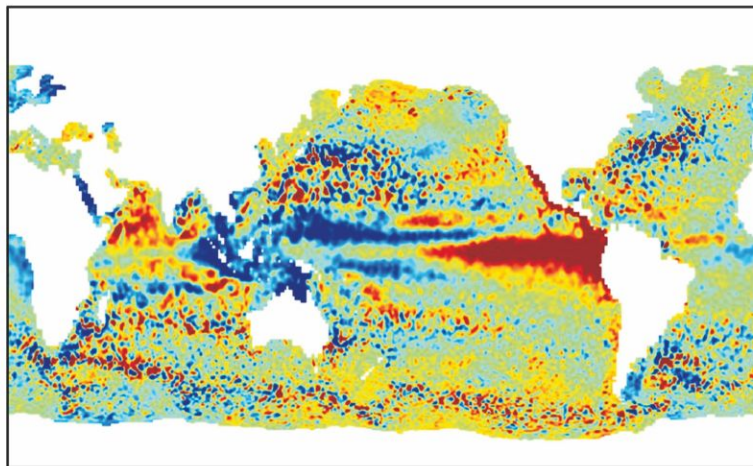
形式化数量化

多时空尺度

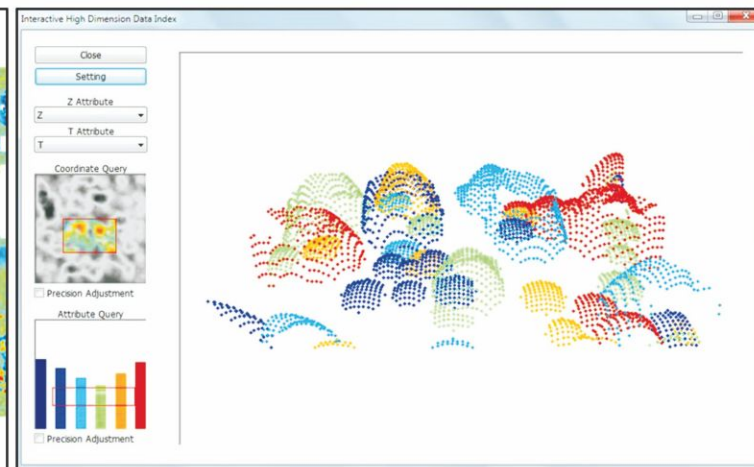
多维性

不确定性

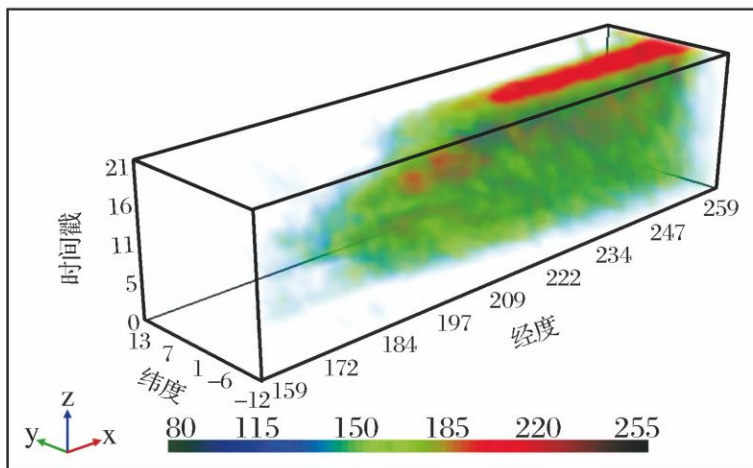




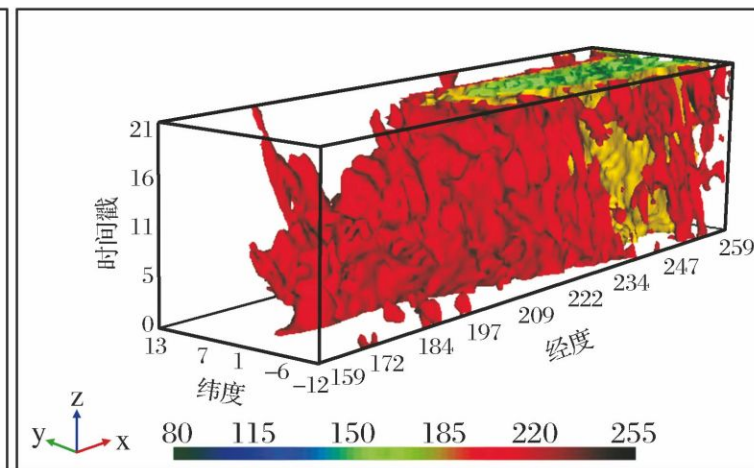
(a) 原始海面时空场数据



(b) 基于双层索引的时空场数据提取

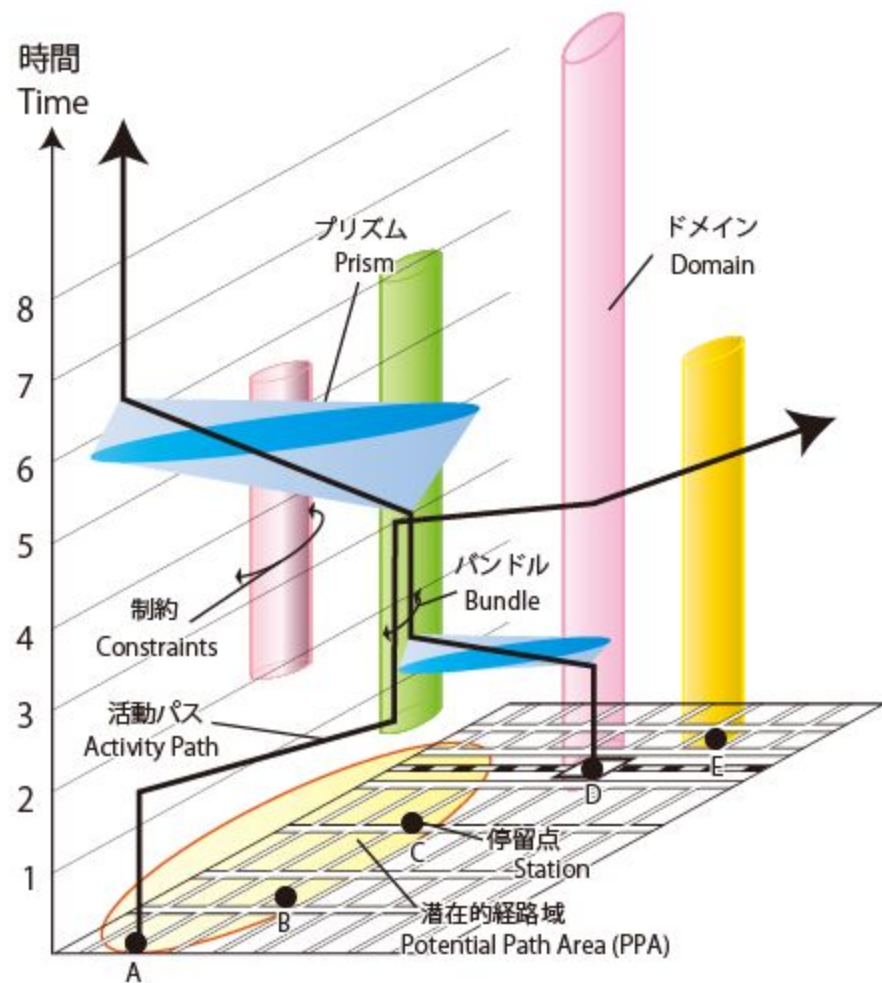


(c) 光线投射形式的时空立方体显示



(d) 等值面形式的时空立方体显示

0010010101010100010111101
11111000111000011001010101
1100000001110101010000001
11010100001110100011000101
1010101000010100101010101
1010001100111000001010100
0010010101010100010111101
11111000111000011001010101
1100000001110101010000001
11010100001110100011000101
1010101000010100101010101
1010001100111000001010100
0010010101010100010111101
11111000111000011001010101
1100000001110101010000001
11010100001110100011000101
1010101000010100101010101
1010001100111000001010100
0010010101010100010111101
11111000111000011001010101
1100000001110101010000001
11010100001110100011000101
1010101000010100101010101
1010001100111000001010100
0010010101010100010111101
11111000111000011001010101
1100000001110101010000001
11010100001110100011000101
1010101000010100101010101
1010001100111000001010100
0010010101010100010111101
11111000111000011001010101
1100000001110101010000001
11010100001110100011000101
1010101000010100101010101
1010001100111000001010100
0010010101010100010111101
11111000111000011001010101
1100000001110101010000001



000111111100101110000001
0011111111001011100000101
001101101001011101100010
001011111011001100001010
001111001011100010010
001110111001010100001010
100001100001111111001011
001110000110001100000110
100001100000000111001001
000000001110010000011100
110000001010101010000110
101101010101010101100111
101101010101010101111100
101101010101010101100011
110101010101010101001010
000001101010001100001111
000011000000001110101010
001010000000011111000001
000111111100101110000001
001111111001011100000101
001101101001011101100010
001011111011001100001010
001111011001011100010010
001110111001010100001010
100001100001111111001011
001110000110001100000110
100001100000000111001001

本章主要内容：关于数据的故事

地理数据类型

地理数据特征

采集与处理

数据库建立

地理数据特征

形式化数量化

多时空尺度

多维性

不确定性

不确定性是地理数据的基本特征之一。

地理数据不确定性的来源：

- 1、位置不确定性：数据源不确定性、数据处理
- 2、属性不确定性：定义模糊、数据源不确定性
- 3、时域不确定性：语义表达不确定性
- 4、模型不确定性：参数不确定性、模型传播

本章主要内容：关于数据的故事

地理数据类型

地理数据特征

采集与处理

数据库建立

采集与处理

采集

处理

渠道来源

应注意的问题

统计处理

空间分布处理

本章主要内容：关于数据的故事

地理数据类型

地理数据特征

采集与处理

数据库建立

1 地理数据统计处理

统计整理：统计分组、计算各组统计值、制作分布图。

2 地理数据空间分布

进行空间处理：格式转换、投影变换、空间分析等。

1、地理数据统计处理

	A	B	C	D	E	F	G	H	I	J
1	组别	姓名	成绩	分值		组别	姓名	成绩	分值	
2	A	A001	5	7		A	A001	5	7	7
3	A	A002	6	5		A	A002	6	5	5
4	E	A003	8	4.5		A	A010	6	5	5
5	B	A004	12	3		A	A012	6	5	5
6	C	A005	10	2		A	A015	12	2	2
7	C	A006	1	9		A	A018	0	10	10
8	E	A007	8	4.5		A	A020	1	9	9
9	C	A008	8	4.5		A	A025	7	3	3
10	C	A009	4	7.5		A	A026	3	8	8
11	A	A010	6	5		A	A028	14	1	1
12	B	A011	13	2		B	A003	8	4.5	4.5
13	A	A012	6	5		B	A004	12	3	3
14	B	A013	3	8		B	A007	8	4.5	4.5
15	B	A014	1	10		B	A011	13	2	2
16	A	A015	12	2		B	A013	3	8	8
17	C	A016	10	2		B	A014	1	10	10
18	B	A017	2	9		B	A017	2	9	9
19	A	A018	0	10		B	A022	7	6	6
20	C	A019	10	2		B	A023	6	7	7
21	A	A020	1	9		B	A030	14	1	1
22	C	A021	8	4.5		C	A005	10	2	2

按某一标志划分为若干性质不同但又有联系的几个部分称“统计分组”。

统计分组根据分组标志的性质，分为按品质标志分组和按数量标志分组。

1、地理数据统计处理

几种常用的统计指标和参数

集中趋势

平均值

中位数

众数

离散程度

极差

离差

离差
平方和

方差
标准差

变异系数

分布特征

偏度系数

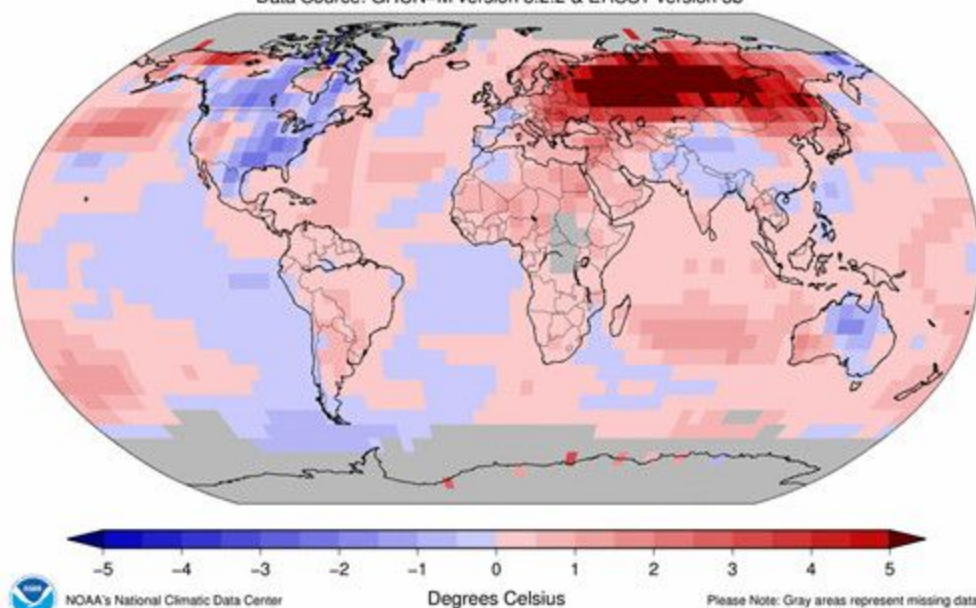
峰度系数

2013年11月 陆地和海洋温度异常值 (以1981-2010年数据为基准)

Land & Ocean Temperature Anomalies Nov 2013

(with respect to a 1981-2010 base period)

Data Source: GHCN-M version 3.2.2 & ERSST version 3b



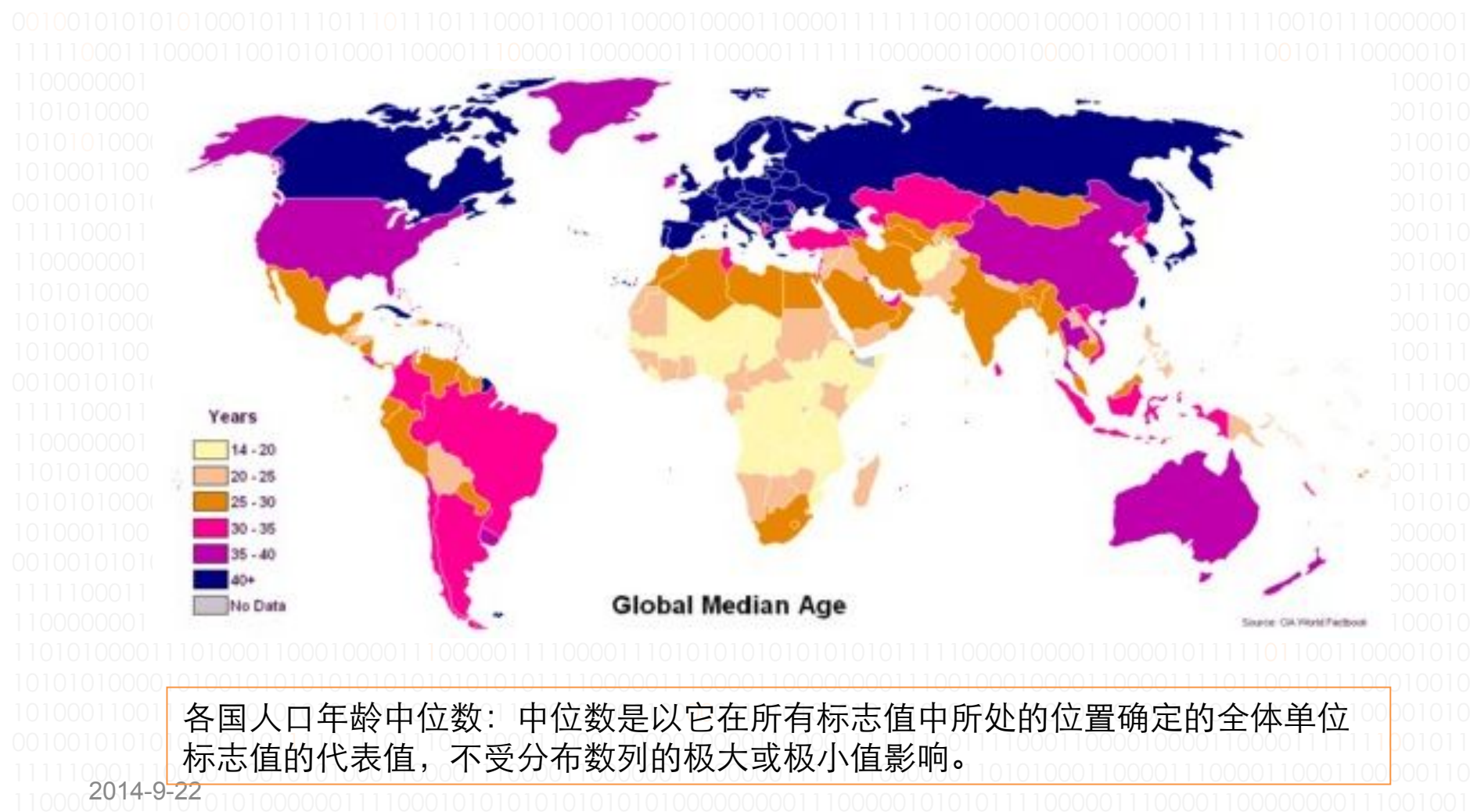
统计平均：受极大极小值影响强烈

2014-9-22

AD:



“咱村有个张千万，其他九个穷光蛋，若是拿来一平均，个个都是张百万”

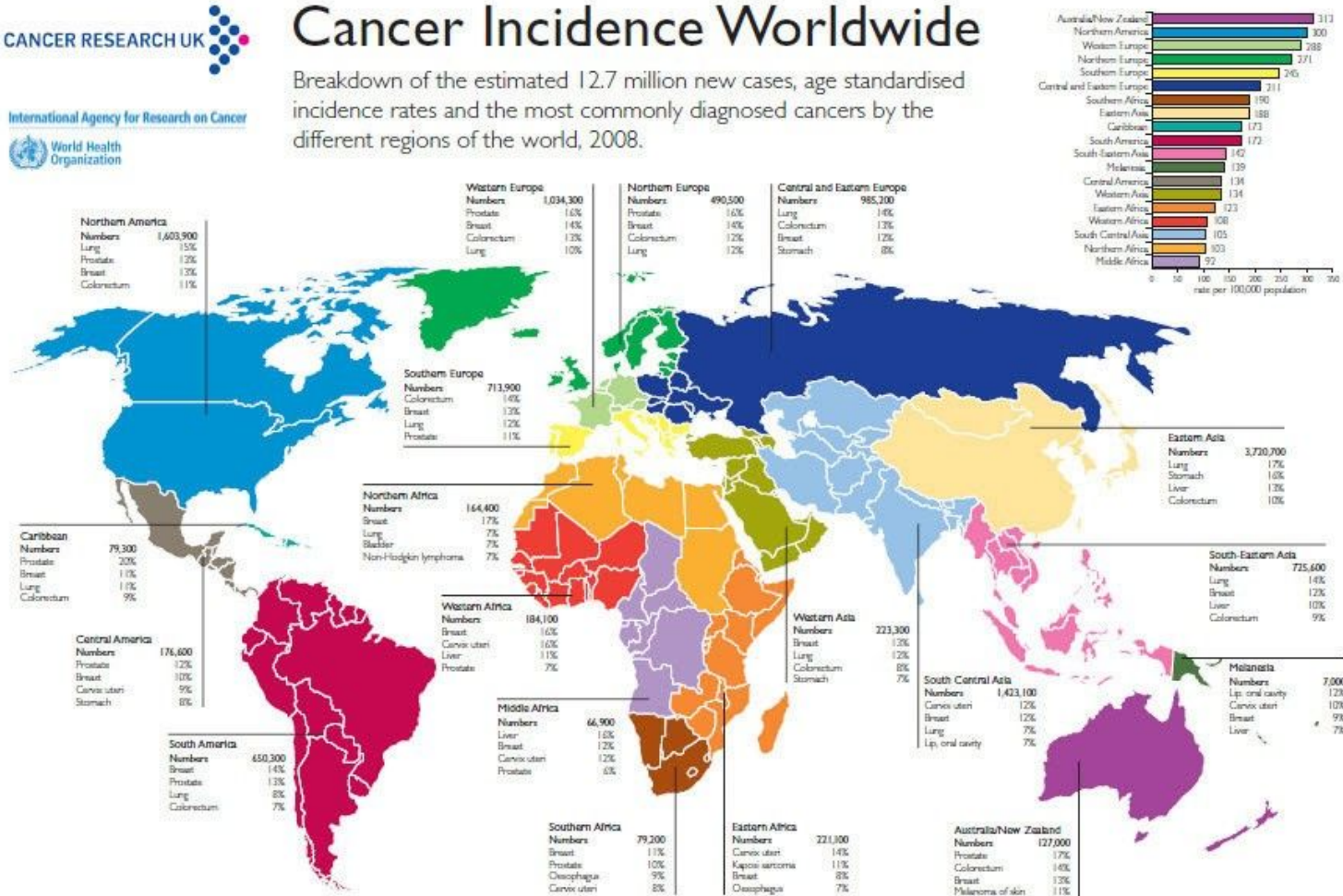


我国处于全球中位数



Cancer Incidence Worldwide

Breakdown of the estimated 12.7 million new cases, age standardised incidence rates and the most commonly diagnosed cancers by the different regions of the world, 2008.



2014-9-2

Source: GLOBOCAN 2008, Cancer Incidence and Mortality Worldwide. IARC, 2010 (<http://globocan.iarc.fr>)

<http://info.cancerresearchuk.org/cancerstats/>

© Cancer Research UK
Registered Charity no 239121

全国地市州盟相对富裕程度地图



中郡县域经济研究所 地区相对富裕程度监测评价中心绘制 2010.12

中国经济网
www.ce.cn

全国地市州盟的相对富裕程度的相对差距有所缩小，但绝对差距还非常大，大部分地市州盟的相对富裕程度在众数线85附近，低于平均线93，并且低于全国平均线100，全国地市州盟的相对富裕程度的“平均数不代表大多数”。

众数是唯一的？

极差 $R = \max_i \{x_i\} - \min_i \{x_i\}$

离差 $d_i = x_i - \bar{x}$

离差平方和 $d^2 = \sum_{i=1}^n (x_i - \bar{x})^2$

方差

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

标准差

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

如果以样本方差对标准差进行无偏估计，则计算公式为：

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

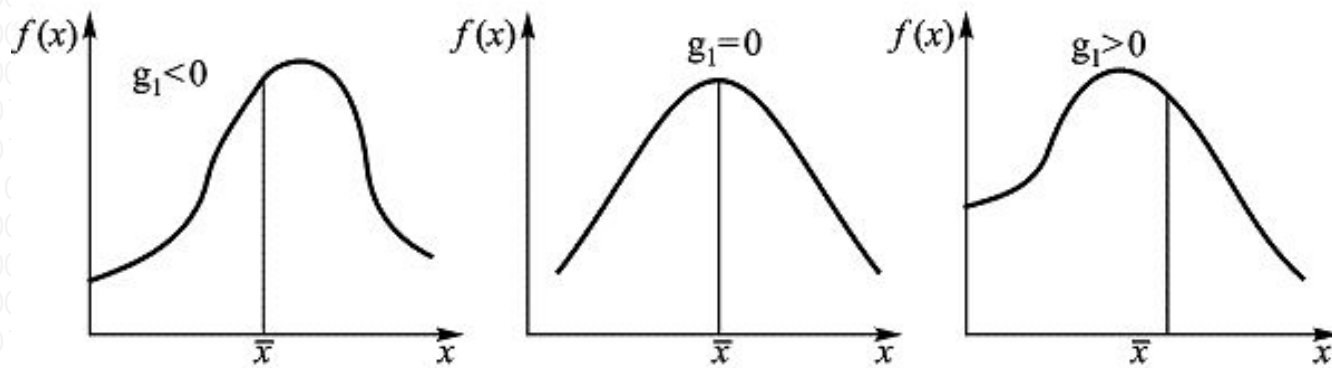
变异系数：

$$C_v = \frac{S}{\bar{x}} \times 100\% = \frac{1}{\bar{x}} \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \times 100\%$$

变异系数表示地理数据的相对变化（波动）程度

偏度系数：

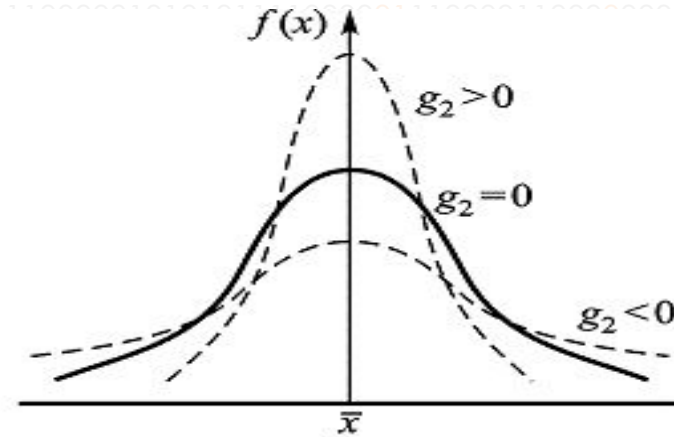
$$g_1 = \sum_{i=1}^n \frac{1}{n} \left(\frac{x_i - \bar{x}}{\sigma} \right)^3$$



测度地理数据分布的不对称性情况，刻画以平均值为中心的偏向情况

峰度系数：

$$g_2 = \sum_{i=1}^n \frac{1}{n} \left(\frac{x_i - \bar{x}}{\sigma} \right)^4 - 3$$



它测度了地理数据在均值附近的集中程度

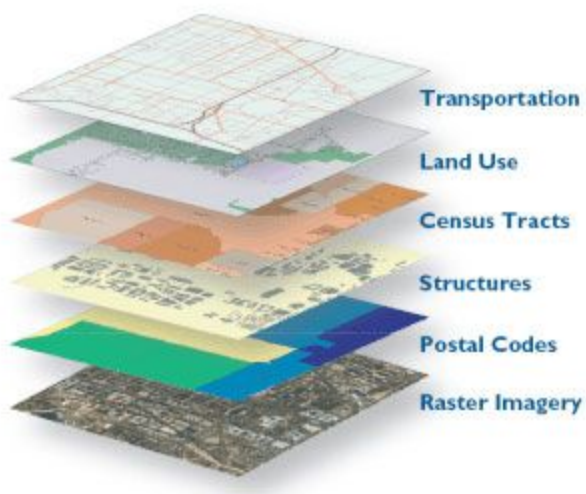
本章主要内容：关于数据的故事

地理数据类型

地理数据特征

采集与处理

数据库建立



any questions?

2014-9-22

例如：对于表中的数据，分别计算极差、离差、离差平方和、方差、标准差、标准差的无偏估计，以及变异系数。

某农场各农田地块的面积

地块编号	1	2	3	4	5	6	7	8	9	10	11	12	平均值	中位数	众数
面积/hm ²	12	83	50	35	55	50	72	40	85	29	65	75	54.25	52.5	50

步骤：

(1) 计算极差

$$R = 85 - 12 = 73$$

(2) 计算离差

序号	1	2	3	4	5	6	7	8	9	10	11	12
面积	12	83	50	35	55	50	72	40	85	29	65	75
离差	-42.25	28.75	-4.25	-19.25	0.75	-4.25	17.75	-14.25	30.75	-25.25	10.75	20.75

✓ (3) 按照公式 (2.4.9) 计算离差平方和

$$d^2 = \sum_{i=1}^{12} d_i = 5\,666.25$$

✓ (4) 按照公式 (2.4.10) 计算方差

$$\sigma^2 = \frac{1}{12} d^2 = 472.19$$

✓ (5) 按照公式 (2.4.11) 计算标准差

$$\sigma = \sqrt{\sigma^2} = \sqrt{472.19} = 21.73$$

✓ (6) 按照公式 (2.4.12) 计算标准差的无偏估计

$$S = \sqrt{\frac{d^2}{12 - 1}} = 21.729\,9$$

✓ (7) 按照公式 (2.4.13) 计算变异系数

$$C_v = \frac{S}{\bar{x}} \times 100\% = 0.418\,4$$



2012年8月26日，北极海冰覆盖面积下降至417万平方公里以下，打破了2007年417万平方公里的最低纪录。

图中，海冰呈现为白色或淡蓝色，其中白色代表海冰密度最高的区域。无冰水面呈现为蓝色，土地呈现为灰色。

你怎么看？

2014-9-22