

សម្រាប់លេខ

15 N.A. 2564

នាសកាជិរាង

លក្ខណីលេខ 623020762-6

Data

1
2
1
0
-1
1

1 12 2 5
2 11 7 2
1 15 9 3
0 10 1 -3
-1 20 12 -2
1 19 6 -5

2D

1 12 2 5
1 2 11 7 2
2 1 15 9 3
1 0 10 1 -3
0 -1 20 12 -2
1 19 6 -5
1 19 6 -5
1 19 6 -5
1 19 6 -5

3D

- ផែតិតនាក់ដឹងទៅអំពី 3 ថែរ 100 គោលរាល់ និងការបញ្ចូនការប្រើប្រាស់ទិន្នន័យ

Data

1 12 2 5
1 2 11 7 2
2 1 15 9 3
1 0 10 1 -3
0 -1 20 12 -2
1 19 6 -5
1 19 6 -5
1 19 6 -5

1 12 2 5
1 2 11 7 2
2 1 15 9 3
1 0 10 1 -3
0 -1 20 12 -2
1 19 6 -5
1 19 6 -5
1 19 6 -5

1 12 2 5
1 2 11 7 2
2 1 15 9 3
1 0 10 1 -3
0 -1 20 12 -2
1 19 6 -5
1 19 6 -5
1 19 6 -5

4D

- ធម្មតាណុយការណីកិច្ច ដែលបានបង្ហាញ និងការប្រើប្រាស់

Data

	Attribute 1	Attribute 2	Attribute 3	Attribute 4
Record 1	1	12	2	5
Record 2	2	11	7	2
Record 3	1	15	9	3
Record 4	0	10	1	-3
Record 5	-1	20	12	-2
Record 6	1	19	6	-5

2 ថែរ

ឯកសារ \rightarrow Record
ឯកសារ \downarrow Attribute

- ការកើតឡើ Data \Rightarrow Attribute និងការ
បង្ហាញនូវព័ត៌មាន Data ដែលបាន
បង្ហាញជា Attribute

Types of Data Sets: (1) Record Data

ຂໍ້ມູນສາມາດແກ່ໄວ້ການຈົບປັດໄດ້

- Relational records
 - Relational tables, highly structured
- Data matrix, e.g., numerical matrix, crosstabs

	China	England	France	Japan	USA	Total
Active Outdoors Crochet Glove	12.00	4.00	1.00	240.00	257.00	
Active Outdoors Lyra Glove	10.00	6.00		323.00	339.00	
InFlux Crochet Glove	3.00	6.00	8.00		132.00	149.00
InFlux Lyra Glove		2.00			143.00	145.00
Triumph Pro Helmet	3.00	1.00	7.00		333.00	344.00
Triumph Vertigo Helmet	3.00	22.00		474.00	499.00	
Xtreme Adult Helmet	8.00	8.00	7.00	2.00	251.00	276.00
Xtreme Youth Helmet		1.00			76.00	77.00
Total	14.00	43.00	54.00	3.00	1,972.00	2,086.00

Pers_ID	Surname	First_Name	City
0	Miller	Paul	London
1	Ortega	Alvaro	Valencia
2	Huber	Urs	Zurich
3	Blanc	Gaston	Paris
4	Bertolini	Fabrizio	Rom

Car_ID	Model	Year	Value	Pers_ID
101	Bentley	1973	100000	0
102	Rolls Royce	1965	330000	0
103	Peugeot	1993	500	3
104	Ferrari	2005	150000	4
105	Renault	1998	2000	3
106	Renault	2001	7000	3
107	Smart	1999	2000	2

- Transaction data

TID	Items
1	Bread, Coke, Milk
2	Beer, Bread
3	Beer, Coke, Diaper, Milk
4	Beer, Bread, Diaper, Milk
5	Coke, Diaper, Milk

	Item	Count	Y	Pla	Bl	Score	W	M	I	IsI	IsM	Session
Document 1	3	0	5	0	2	6	0	2	0	2	0	1
Document 2	0	7	0	2	1	0	0	0	3	0	0	2
Document 3	0	1	0	0	1	2	2	0	3	0	0	3

- Document data: Term-frequency vector (matrix) of text documents

- Normalization ເພື່ອມາຍືນວ່າທີ່ມີຂໍ້ມູນທີ່ມີຄວາມຈົບປັດທີ່ກຳນົດໄດ້ ໃຫ້ມີຄວາມບຸນດຸນໃຫຍ່
- ມີການ crosstabs ແລ້ວ Attribute ຈົດສົ່ງ

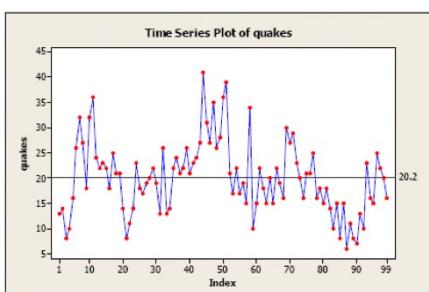
ຂໍ້ມູນສົດສົ່ງ (ໃຫຍ່) Types of Data Sets: (3) Ordered Data

ຮູບພາບຢັນໂນມູນຕົວສຳເນົາທີ່ມີເວລີກ, time series

- Video data: sequence of images

ສຳເນົາໃຫຍ່ ມີຈົດລົງທຶນ 1, 2, 3 ແລ້ວສົ່ງໄວ້

- Temporal data: time-series



- Sequential Data: transaction sequences

- DNA

- Genetic sequence data

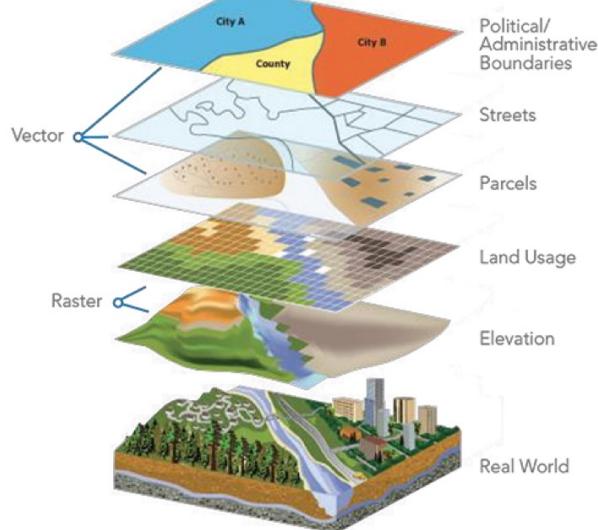
- ສູງລາວ, ໄລຍາຊີ

Human	Chimpanzee	Macaque	Start
GTTTGAGG	-- ATGTTCAACAAATGCCTTTCATTCTCTATTTCAGACCTGCCCA		GTTTGAGG
GTTTGAGG	-- ATGTCATAAATAGCTCTTCACTCCTGTATTTCAGACCCGCCA		GTTTGAGG
Human	Chimpanzee	Macaque	GACAAATTCTGCTAGCAGCCTTTGTCATTATCCTGCTCATTTACACTGCCCA
GATCTGGAGACTAACTCTGAAATAAAAGCTGATAATTATTTTCTCAAACAA			GATCTGGAGACTAACTCTGAAATAAAAGCTGATAATTATTTTCTCAAACAA
Human	Chimpanzee	Macaque	TATCTGGAGACTAAACTCTGAAATAAAAGCTGATAATTATTTTCTCAAACAA
CAGAACATCGATTAGCAAATTACTCTTAAGATAATTAACTTACATTTCTGCTTA			CAGAACATCGATTAGCAAATTACTCTTAAGATAATTAACTTACATTTCTGCTTA
Human	Chimpanzee	Macaque	CAGAACATCGATTAGCAAATTACTCTTAAGATAATTAACTTACATTTCTGCTTA
CCCCTGAGTTGATGTTGAGCGTATGTCAGTTCTAAAGCCAGGTATAC			CCCCTGAGTTGATGTTGAGCGTATGTCAGTTCTAAAGCCAGGTATAC
Human	Chimpanzee	Macaque	TTATACATTATTCAGTTCTGCAAGTATGTCAGTTCTGCAAGTATAC
GACAGGTAAATAAAAAGATATTATTCAGTTCTGCAAGTATGTCAGTTCTGCA			GACAGGTAAATAAAAAGATATTATTCAGTTCTGCAAGTATGTCAGTTCTGCA
H I Y S F L S K			H I Y S F L S K
AAC TGT TCGCGGTGTTGCTAA	-- TGT AAAACAAACTCAAGTACA		AAC TGT TCGCGGTGTTGCTAA -- TGT AAAACAAACTCAAGTACA
AAC TGT TCGCGGTGTTGCTAA	-- TGT AAAACAAACTCAAGTACA		AAC TGT TCGCGGTGTTGCTAA -- CBT AAAACAAACTCAAGTACA

Types of Data Sets: (4) Spatial, image and multimedia Data

សម្រាប់ការបង្កើតរបស់ពីរបៀបក្នុងផ្ទាល់ការគ្រប់គ្រង X, Y

□ Spatial data: maps



□ Image data:

□ Video data: Spatio-temporal

សម្រាប់ការបង្កើតរបស់ពីរបៀបក្នុងផ្ទាល់ការគ្រប់គ្រង X, Y

សម្រាប់ Maps

សម្រាប់ការបង្កើតរបស់ពីរបៀប Record

Important Characteristics of Structured Data

- ↳ Dimensionality ដែលវិភាគនូវលក្ខណៈ 2, 3, 4, 5 ទៅការបង្កើតរបស់ពីរបៀប Dimension នៃ ឯកសារទឹកចំណាំ
 - ↳ Sparsity ដែលស្ថិតនៅក្នុងប្រព័ន្ធដែលមិនមែនចំណាត់ថ្នាក់ឡើយ
 - ↳ Resolution ដែលពិនិត្យនូវពិន្ទុផ្លូវ
 - ↳ Distribution ដែលស្ថិតនៅក្នុងប្រព័ន្ធ
- Sparsity

	China	England	France	Japan	USA	Total
Active Outdoors Crochet Glove		12.00	4.00	1.00	240.00	257.00
Active Outdoors Lycra Glove		10.00	6.00		323.00	339.00
Influx Crochet Glove	3.00	6.00	8.00		132.00	149.00
Influx Lycra Glove		2.00				143.00
Triumph Pro Helmet	3.00	1.00	7.00		333.00	344.00
Triumph Vertigo Helmet		3.00	22.00		474.00	499.00
Xtreme Adult Helmet	8.00	8.00	7.00	2.00	251.00	276.00
Xtreme Youth Helmet			1.00			76.00
Total	14.00	43.00	54.00	3.00	1,972.00	2,086.00

Data Objects

- ក្នុងប្រព័ន្ធដែលបានបង្កើតឡើង data អាមេរិក
- ធនធានជាតិគឺជាប្រព័ន្ធ

- sales database: customers, store items, sales
- medical database: patients, treatments
- university database: students, professors, courses
- Also called samples, examples, instances, data points, objects, tuples
- Data objects are described by attributes សម្រាប់ការបង្កើតរបស់ពីរបៀប

Databases Then Record

Attributes

□ Attribute (or dimensions, features, variables)

- A data field, representing a characteristic or feature of a data object.

□ E.g., *customer_ID, name, address*

□ Types:

□ Nominal (e.g., red, blue) *ສົດໝາຍແລະ ສົດໝາຍອັນດຸ*

□ Binary (e.g., {true, false}) *ຕະຫຼາມຕະຫຼາມ*

□ Ordinal (e.g., {freshman, sophomore, junior, senior}) *ບົນລວມຕະຫຼາມສິລະນະ*

□ Numeric: quantitative *ໜ້າຈົກ 11000 21100*

□ Interval-scaled: 100°C is interval scales

□ Ratio-scaled: 100°K is ratio scaled since it is twice as high as 50°K

□ Q1: Is *student ID* a nominal, ordinal, or interval-scaled data?

□ Q2: What about *eye color*? Or color in the color spectrum of physics?

Numeric

Attribute Types

□ Nominal: categories, states, or “names of things”

□ *Hair_color* = {auburn, black, blond, brown, grey, red, white} *ຕົວຢ່າງ*

□ marital status, occupation, ID numbers, zip codes

□ Binary

□ Nominal attribute with only 2 states (0 and 1) *ກົດໝາຍ 0 ແລະ 1, ຖະກົດວິທີ*

□ Symmetric binary: both outcomes equally important

□ e.g., *gender* Left/Right-handed, coke/Pepsi, hot/Cold

□ Asymmetric binary: outcomes not equally important. *ກົດໝາຍ 0 ໂດຍແມ່ນສຳເນົາກົດໝາຍ 1 ໃຫຍ້ກົດໝາຍ 1 ໂດຍແມ່ນສຳເນົາກົດໝາຍ 0*

□ e.g., medical test (positive vs. negative)

□ Convention: assign 1 to most important outcome (e.g., HIV positive)

□ Ordinal

□ Values have a meaningful order (ranking) but magnitude between successive values is not known

□ *Size* = {small, medium, large}, grades, army rankings

Numeric Attribute Types

គឺជាការបង្ហាញតម្លៃ

→ តម្លៃនៃការកំណត់វិធាន៍

→ ចំនួនការសែនសែន ឬ ចំនួនតម្លៃ

Discrete Attributes

តម្លៃត្រូវបានបញ្ជូន

→ រាយរាយដែលមានតម្លៃត្រូវបានបញ្ជូន
Ex. male, female, education level, gender

→ Zip codes

→ Binary

Continuous Attributes

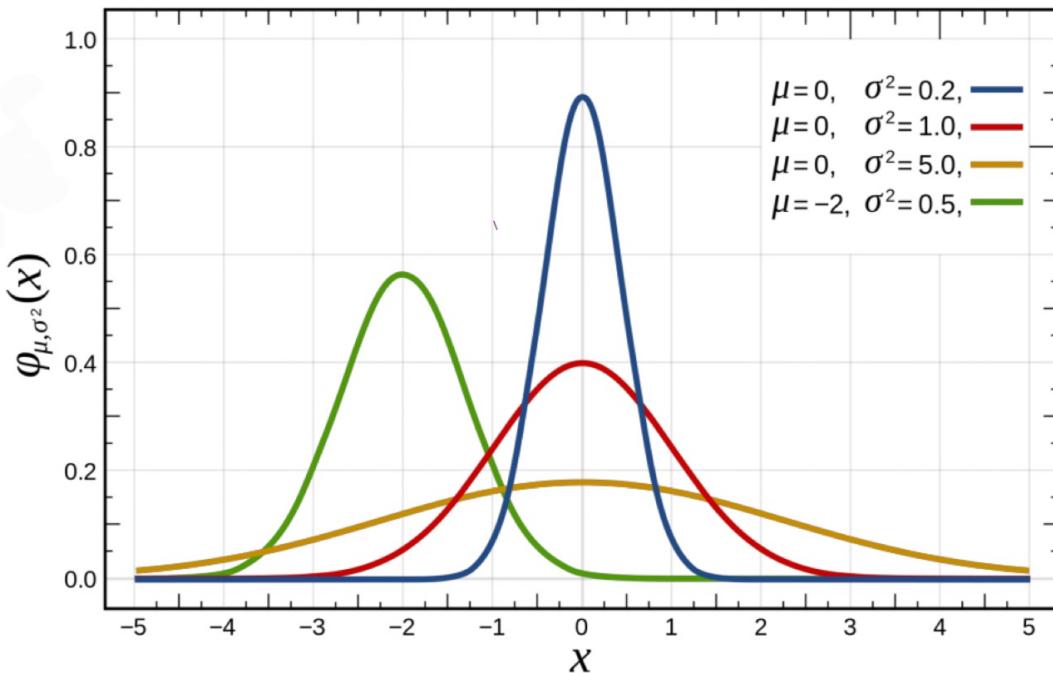
តម្លៃត្រូវបានបញ្ជូន

→ តម្លៃត្រូវបានបញ្ជូនដោយគ្មាន
Ex. age 180 ក្នុង 181 និងទៅលើនីមួយៗ 180.5

Basic Statistical Descriptions of Data

សំណើនូវការបង្ហាញនូវការបង្ហាញនៃព័ត៌មាន

- Variation តើមិនមែនភាពខ្សោយអ្នករាយការណ៍ទេ
- spread តើមិនមែនភាពខ្សោយអ្នករាយការណ៍ទេ
- Median តើមានមួយទេ



Ex. រាយរាយដែលមានតម្លៃត្រូវបានបញ្ជូន 20 ដុល្លារ
តុលាបន្ទាយ និង $\mu=20$