



ეკონომიკისა და ბიზნესის სტატისტიკა

თემა 2

ცენტრალური ტენდენციის
მახასიათებლები



მონაცემთა ტიპები

მონაცემები

კატეგორიული

რაოდენობრივი

დისკრეტული

უწყვეტი

მაგალითები:

- ქორწინების სტატუსი
- გაქვთ თუ არა ხმის მიცემის უფლება
- თვალების ფერი

მაგალითები:

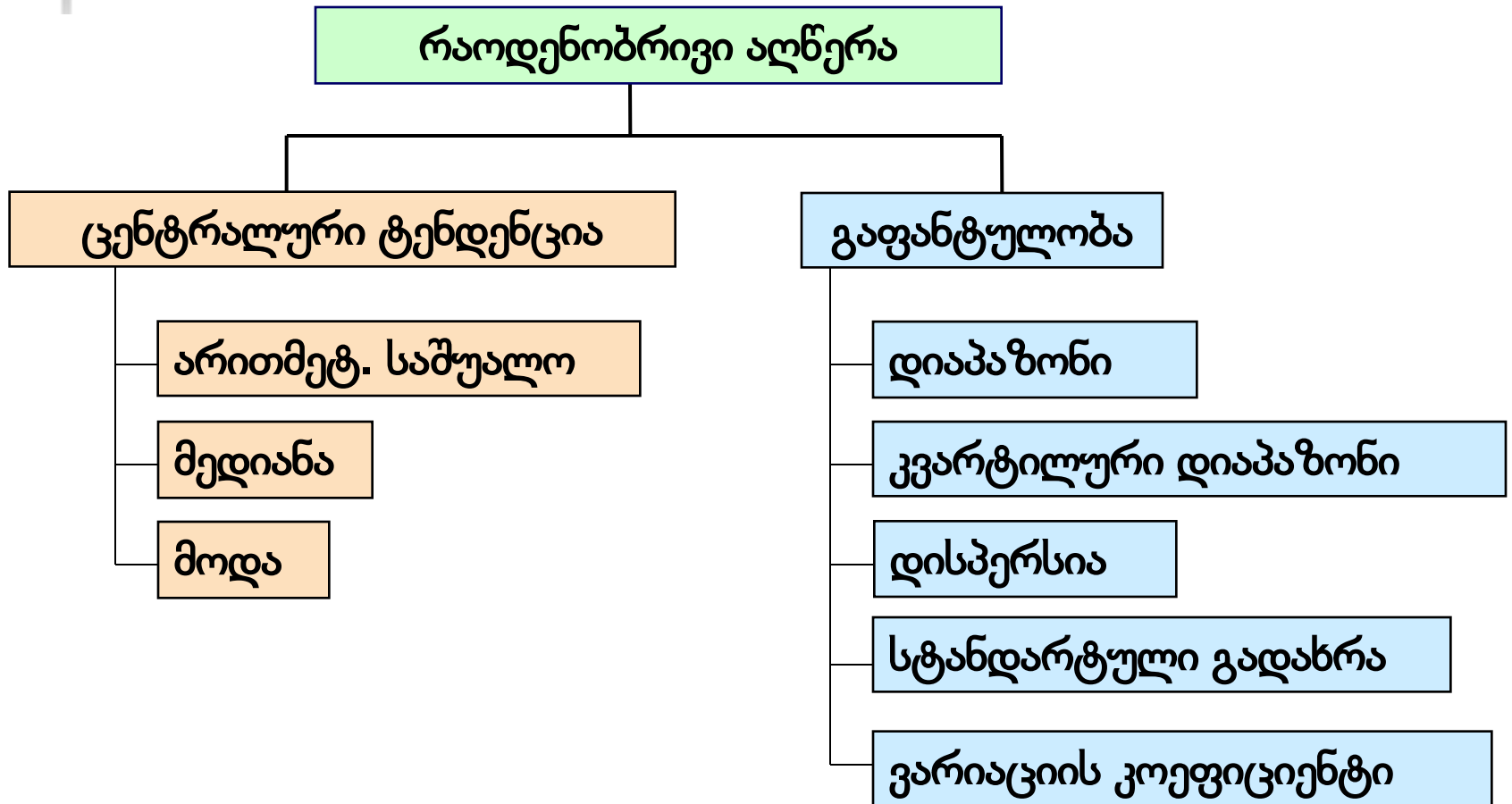
- შვილთა რაოდენობა
- შეცდომები 1 საათში (თვლადი ობიექტები)

მაგალითები:

- წონა
- ძაბვა (ზომადი მახასიათებლები)



მონაცემთა რაოდენობრივი აღწერა





ცენტრალური ტენდენციის საზომები

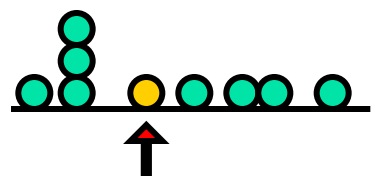
ცენტრალური ტენდენცია

საშუალო

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

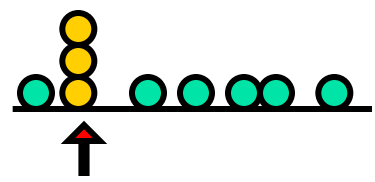
არითმეტიკული
საშუალო

მედიანა



დალაგებულ
მონაცემთა შუა
წერტილი

მოდა



ყველაზე ხშირი
დაკვირვება



(არითმეტიკული) საშუალო (Mean)

- ცენტრალური ტენდენციის ყველაზე ხშირად გამოყენებადი საზომი
 - N მნიშვნელობათა პოპულაციისათვის:

$$\mu = \frac{\sum_{i=1}^N x_i}{N} = \frac{x_1 + x_2 + \square + x_N}{N}$$

პოპულაციის მნიშვნელობები

პოპულაციის ზომა

- n ზომის შერჩევისათვის:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{x_1 + x_2 + \square + x_n}{n}$$

დაკვირვების ობიექტები

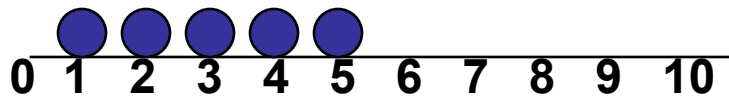
შერჩევის ზომა



არითმეტიკული საშუალო

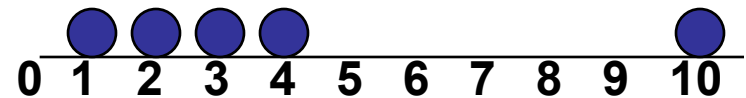
(გაგრძ.)

- საშუალო = მნიშვნელობათა ჯამი გაყოფილი მნიშვნელობათა რაოდენობაზე
- იცვლება ექსტრემალური (ამოვარდნილი) მნიშვნელობების (outliers) გამო



საშუალო = 3

$$\frac{1+2+3+4+5}{5} = \frac{15}{5} = 3$$



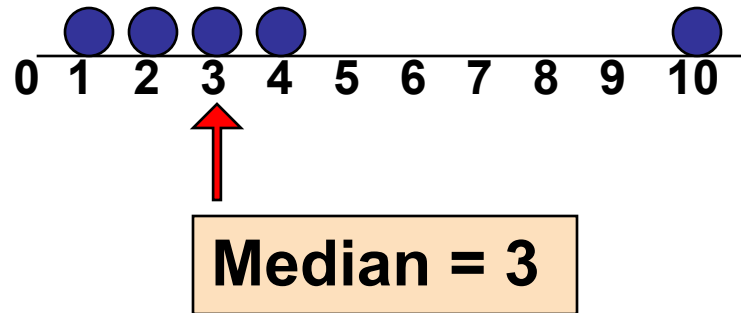
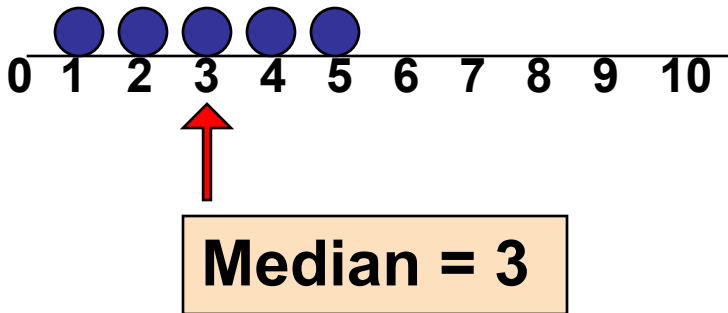
საშუალო = 4

$$\frac{1+2+3+4+10}{5} = \frac{20}{5} = 4$$



მედიანა

- დალაგებულ (ზრდადობით ან კლებადობით) მნიშვნელობებში მედიანა არის “შუა” რიცხვი (50% მის ზემოთ და 50% ქვემოთ)



- ხდება ექსტრემალური მნიშვნელობების გავლენის თავიდან აცილება



მედიაანის პოვნა

მედიაანის მდებარეობა = დალაგებული მონაცემების

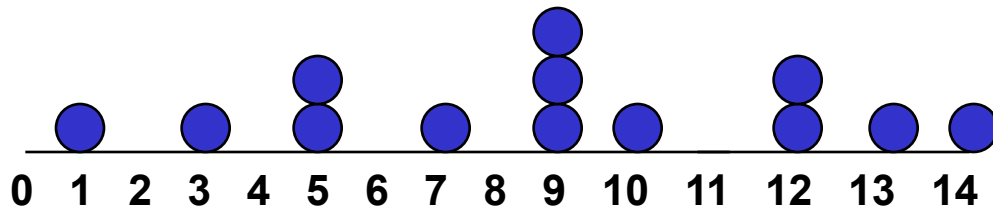
$$-\frac{n+1}{2}\text{-ე წევრი}$$

- თუ რიცხვთა ოდენობა კენტია, მედიანა შუა წევრია
- თუ რიცხვთა ოდენობა ლუნია, მედიანა არის ორი შუა წევრის საშუალო

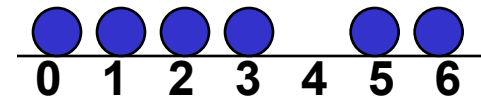


მოდა

- ცენტრალური ტენდენციის საზომი
- ყველაზე ხშირი მოვლენა
- ექსტრემალური მნიშვნელობების გავლენას ადგილი არ აქვს
- გამოიყენება რაოდენობრივი და კატეგორიული მონაცემებისათვის
- შესაძლებელია გვექონდეს რამდენიმე ან არც ერთი მოდა



მოდა = 9



არც ერთი მოდა

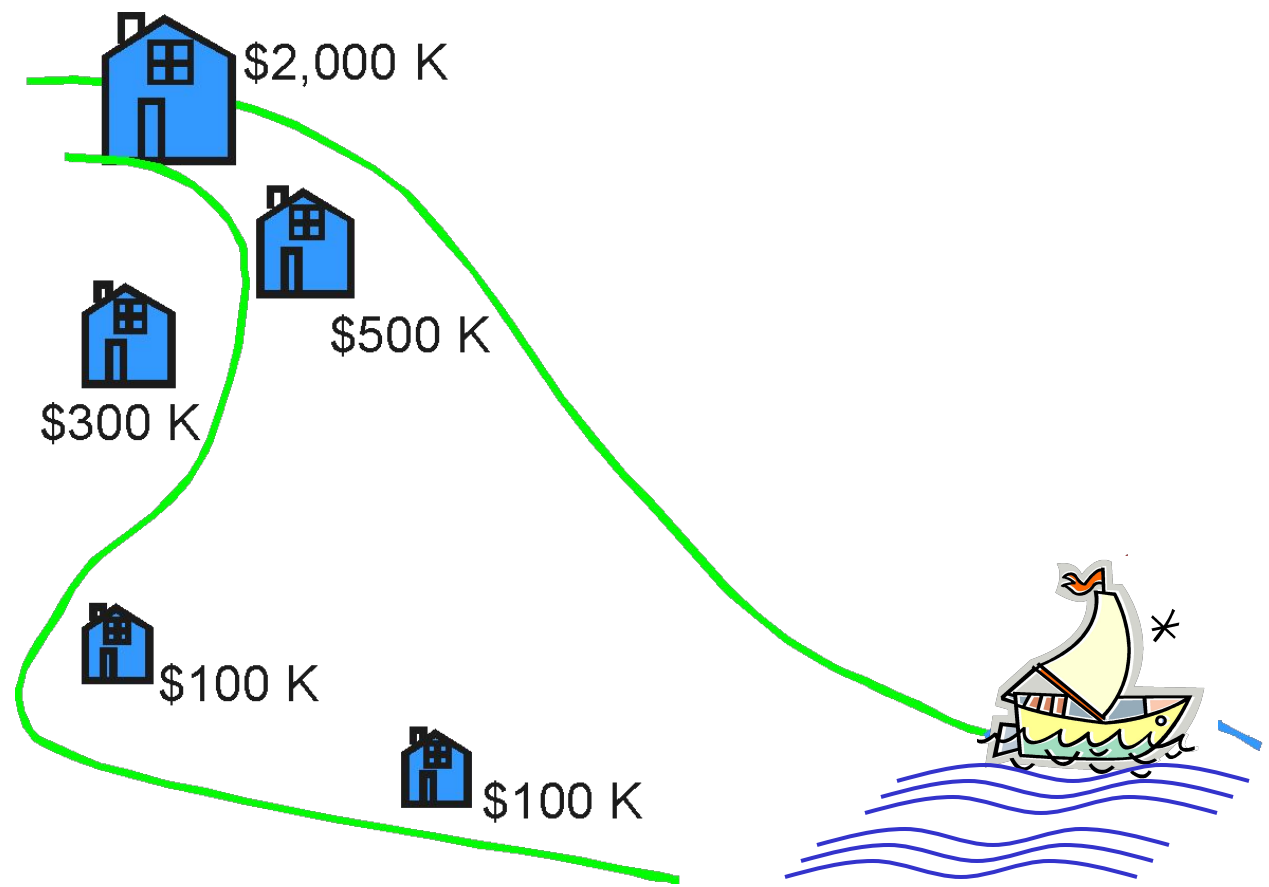


მაგალითი: გამეორება

- 5 სახლი ზღვის პირას

სახლის ფასები:

\$2,000,000
500,000
300,000
100,000
100,000





შემაჯამებელი სტატისტიკები

სახლის ფასები:

\$2,000,000

500,000

300,000

100,000

100,000

ჯამი 3,000,000

- საშუალო: $(\$3,000,000/5)$
= **\$600,000**

- მედიანა: დალაგებულ მონაცემთა შუა
წერტილი
= **\$300,000**

- მოდა: ყველაზე ხშირი მნიშვნელობა
= **\$100,000**



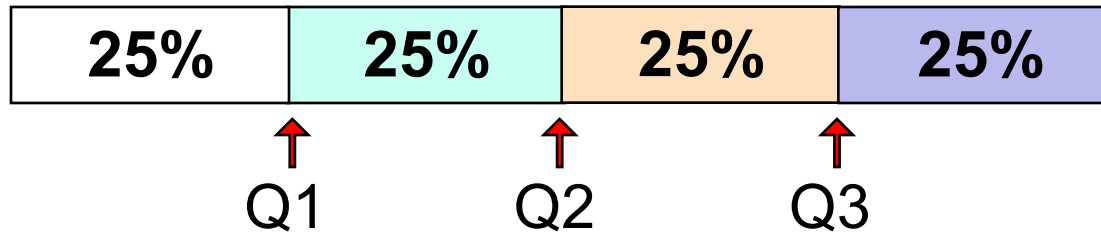
მდებარეობის რომელი საზომია “საუკეთესო”?

- ძირითადად გამოიყენება **საშუალო**, თუ არ არის ამოვარდნილი მნიშვნელობები
- ასეთ შემთხვევაში გამოიყენება **მედიანა**, რადგან ის არ არის მგრძობიარე ამოვარდნილი მნიშვნელობების მიმართ.
 - **მაგალითი:** სახლის მედიანური ფასების გამოყენება რეგიონისთვის – ნაკლებად მგრძობიარე ამოვარდნილი მნიშვნელობების მიმართ



კვარტილები

- კვარტილები ყოფენ ზრდადობით (კლებადობით) დალაგებულ მონაცემებს 4 სეგმენტად, რომლებიც შეიცავს მნიშვნელობათა თანაბარ რაოდენობებს.



- პირველი კვარტილი, Q_1 , არის მნიშვნელობა, რომელზეც დაკვირვებათა 25% უფრო ნაკლებია და 75% – უფრო მეტი.
- Q_2 არის იგივე რაც **მედიანა** (50% უფრო ნაკლები, 50% უფრო მეტი)
- მესამე კვარტილისთვის (Q_3) მხოლოდ დაკვირვებათა 25% არის უფრო მეტი



შენონილი საშუალო

- მონაცემთა სიმრავლის შენონილი საშუალო არის

$$\bar{x} = \frac{\sum_{i=1}^n w_i x_i}{\sum w} = \frac{w_1 x_1 + w_2 x_2 + \dots + w_n x_n}{\sum w_i}$$

- სადაც w_i არის i -ური დაკვირვების წონა

- გამოიყენება, როდესაც მონაცემები უკვე დაჯგუფებულია n კლასად, სადაც i -ურ კლასში მოქცეულია w_i ოდენობის მნიშვნელობები