ეკონომიკისა და ბიზნესის სტატისტიკა

თემა 3

მონაცემთა გაბნევის (გაფანტულობის) მახასიათებლები

გაფანტულობის საზომები

გაფანტულობა

- **გაბნევის დიაპაზონი**
- **კვარტილური დიაპაზონი**
- **სტანდარტ. გადახრა**
- **ვარიაციის კოეფიციენტი**

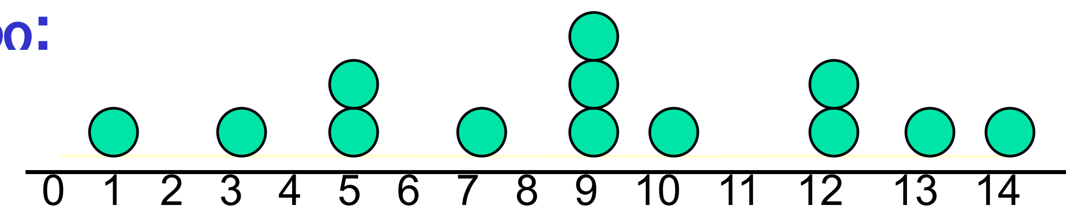
- გაფანტულობის საზომები გვაწვდის ინფორმაციას მონაცემთა მნიშვნელობების **განსხვავებებზე** ან **ცვალებადობაზე**

გაბნევის დიაპაზონი (Range)

- გაფანტულობის უმარტივესი საზომი
- სხვაობა უდიდეს და უმცირეს დაკვირვებებს შორის:

$$\text{დიაპაზონი} = X_{\text{უდიდესი}} - X_{\text{უმცირესი}}$$

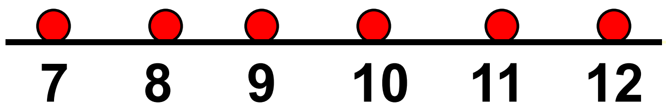
მაგალითი:



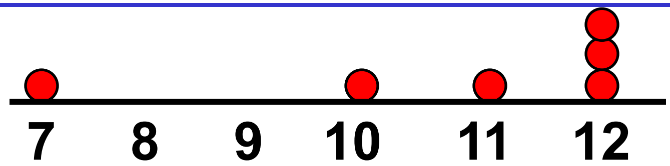
დიაპაზონი = 14 - 1 = 13

გაბნევის დიაპაზონის ნაკლი

- უგულებელყოფს მონაცემთა განაწილებას



დიაპაზონი = 12 - 7 = 5



დიაპაზონი = 12 - 7 = 5

- მგრძნობიარე ამოვარდნილი მნიშვნელობების მიმართ

1,1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,3,3,3,3,4,5

დიაპაზონი = 5 - 1 = 4

1,1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,3,3,3,3,4,120

დიაპაზონი = 120 - 1 = 119

კვარტილური დიაპაზონი

- ამოვარდნილი მნიშვნელობების პრობლემის დაძლევა გარკვეულწილად შესაძლებელია **კვარტილური დიაპაზონის (interquartile range) მეშვეობით**

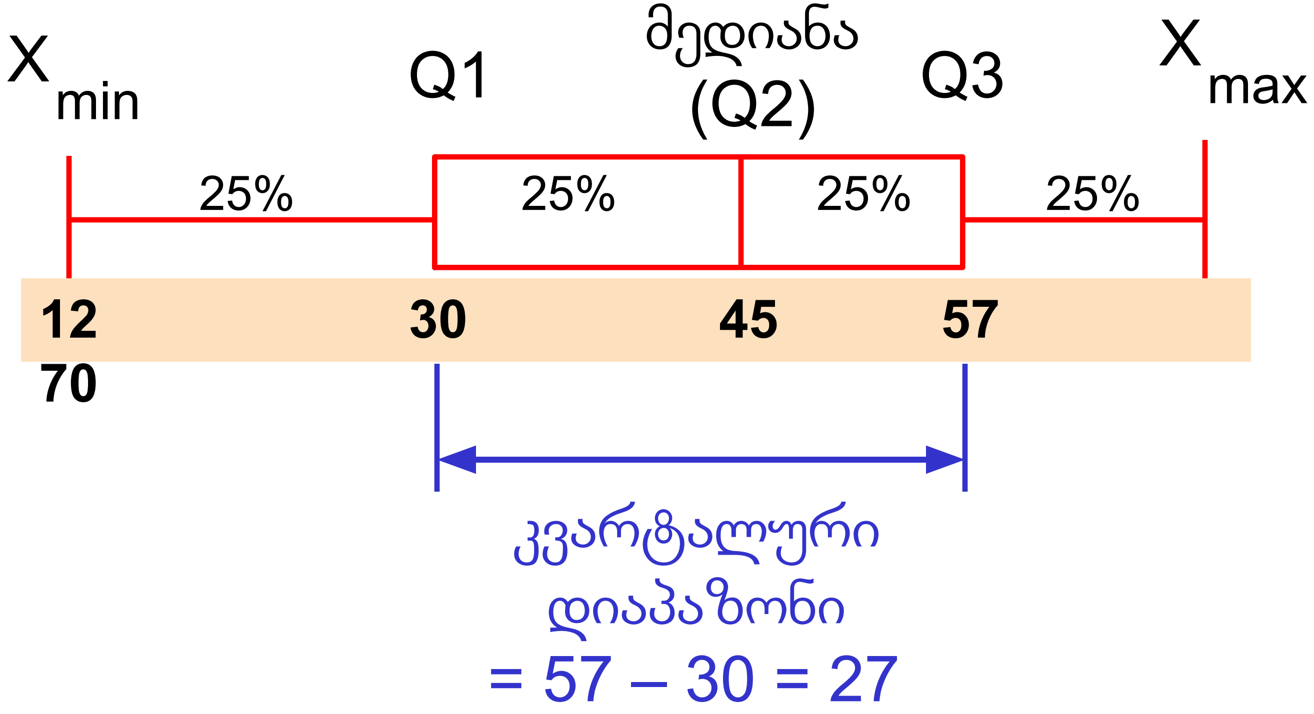
- ხდება მაღალი და დაბალი მნიშვნელობების ამოყრა და გაბნევის დიაპაზონის გამოთვლა მონაცემთა შუა 50%–ისათვის

- კვ. დიაპაზონი = მე-3 კვარტილი – 1-ელი კვარტილი

$$IQR = Q_3 - Q_1$$

კვარტილური დიაპაზონი

მაგალითი:



პოპულაციის ვარიაცია (დისპერსია)

- პოპულაციის საშუალოდან მნიშვნელობათა კვადრატული გადახრების საშუალო:

- პოპულაციის ვარიაცია:

$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}$$

სადაც

μ = პოპულაციის საშუალო

N = პოპულაციის ზომა

x_i = x ცვლადის i–ური მნიშვნელობა

შერჩევის ვარიაცია (დისპერსია)

- საშუალოდან მნიშვნელობათა კვადრატული გადახრების (მიახლოებითი) საშუალო
  - შერჩევის ვარიაცია:

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

სადაც $\overline{X}$ = არითმეტიკული საშუალო

n = შერჩევის ზომა

x_i = x ცვლადის i–ური მნიშვნელობა

პოპულაციის სტანდარტული გადახრა

- გაფანტულობის გაზომვის ყველაზე ხშირად გამოყენებადი საზომი
- გვიჩვენებს ვარიაციას საშუალოს მიმართ
  - პოპულაციის სტანდარტული გადახრა:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}}$$

შერჩევის სტანდარტული გადახრა

- გაფანტულობის გაზომვის ყველაზე ხშირად გამოყენებადი საზომი
- გვიჩვენებს ვარიაციას საშუალოს მიმართ
  - შერჩევის სტანდარტული გადახრა :

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

მაგალითი: შერჩევის სტანდარტული გადახრა

შერჩევის მონაცემები(x_i)

10	**12**	**14**	**15**	**17**	**18**	**18**	**24**

n = 8 საშუალო = $\bar{x}$ = 16

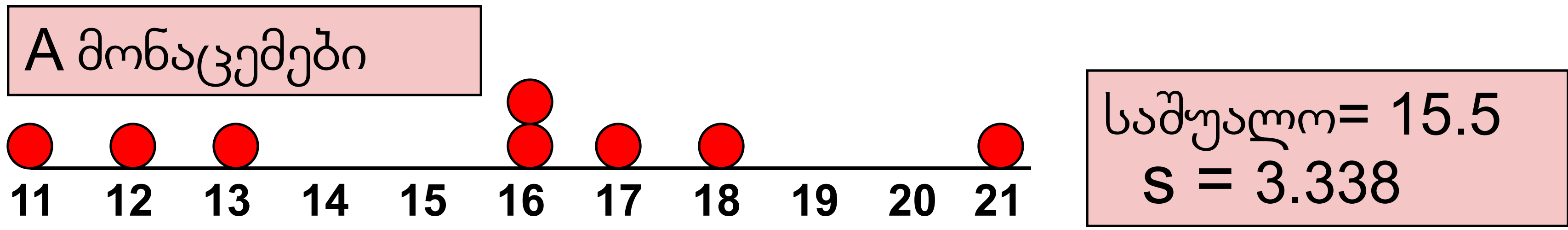
$$s = \sqrt{\frac{(10-\bar{X})^2 + (12-\bar{x})^2 + (14-\bar{x})^2 + \cdots + (24-\bar{x})^2}{n-1}}$$

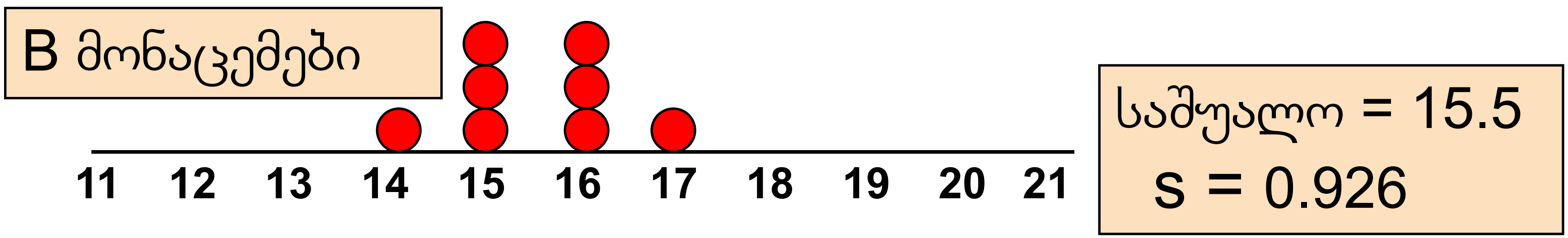
$$= \sqrt{\frac{(10-16)^2 + (12-16)^2 + (14-16)^2 + \cdots + (24-16)^2}{8-1}}$$

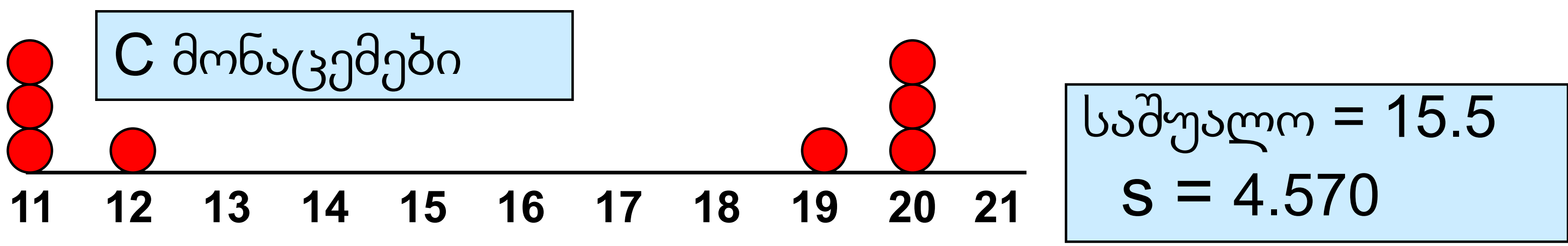
$$= \sqrt{\frac{126}{7}} = \boxed{4.2426}$$

საშუალო მნიშვნელობიდან “საშუალო” გადახრა

სტანდარტული გადახრების შედარება







დისპერსიის და სტანდარტული გადახრის უპირატესობები

- მონაცემთა სიმრავლის თითოეული მნიშვნელობა მონაწილეობს გათვლებში

- საშუალოდან შედარებით უფრო დაშორებულ მნიშვნელობებს მეტი წონა აქვთ (რადგან საშუალოდან გადახრები კვადრატში აიყვანება)

ვარიაციის კოეფიციენტი

- ზომავს ფარდობით განფენილობას
- ყოველთვის აისახება პროცენტულად (%)
- გვიჩვენებს გაფანტულობას საშუალოსთან მიმართებაში
- შესაძლებელია მისი გამოყენება მონაცემთა ორი ან მეტი სიმრავლის შემთხვევაში (რომლებიც ასახულია ზომის განსხვავებულ ერთეულებში)

$$CV = \left(\frac{s}{\bar{x}} \right) \cdot 100\%$$

ვარიაციის კოეფიციენტის შედარება

- A დასახელების აქცია:
  - საშუალო ფასი გასულ წელს = $50
  - სტანდარტული გადახრა = $5

$$CV_A = \left(\frac{s}{\bar{X}}\right) \cdot 100\% = \frac{\$5}{\$50} \cdot 100\% = 10\%$$

- B დასახელების აქცია :
  - საშუალო ფასი გასულ წელს = $100
  - სტანდარტული გადახრა = $5

$$CV_B = \left(\frac{s}{\bar{X}}\right) \cdot 100\% = \frac{\$5}{\$100} \cdot 100\% = 5\%$$

ორივე აქციას აქვს თანაბარი სტანდარტული გადახრა, მაგრამ B აქცია ნაკლებად ცვალებადია

კოვარიაცია

- კოვარიაცია ზომავს **ორ ცვლადს შორის** წრფივი დამოკიდებულების მიმართულებას

- პოპულაციის კოვარიაცია:

$$\boxed{\mathrm{Cov}(x,y)=\sigma_{xy}=\frac{\sum_{i=1}^{N}(x_i-\mu_x)(y_i-\mu_y)}{N}}$$

- შერჩევის კოვარიაცია:

$$\boxed{\mathrm{Cov}(x,y)=s_{xy}=\frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{n-1}}$$

  - ზომავს მხოლოდ დამოკიდებულების მიმართულებას
  - არ გულისხმობს მიზეზ–შედეგობრივ კავშირებს

კოვარიაციის ინტერპრეტაცია

- **კოვარიაცია** ორ ცვლადს შორის:

$Cov(x,y) > 0$ $\rightarrow$ x და y მოძრაობენ ერთსა და იმავე მიმართულებით

$Cov(x,y) < 0$ $\rightarrow$ x და y მოძრაობენ სხვადასხვა მიმართულებით

$Cov(x,y) = 0$ $\rightarrow$ x და y დამოუკიდებლები არიან

კორელაციის კოეფიციენტი

- ზომავს **ორ ცვლადს შორის** წრფივი დამოკიდებულების **ფარდობით** სიძლიერეს
- პოპულაციის კორელაციის კოეფიციენტი:

$$\boxed{\rho = \frac{Cov(x,y)}{\sigma_X \sigma_Y}}$$

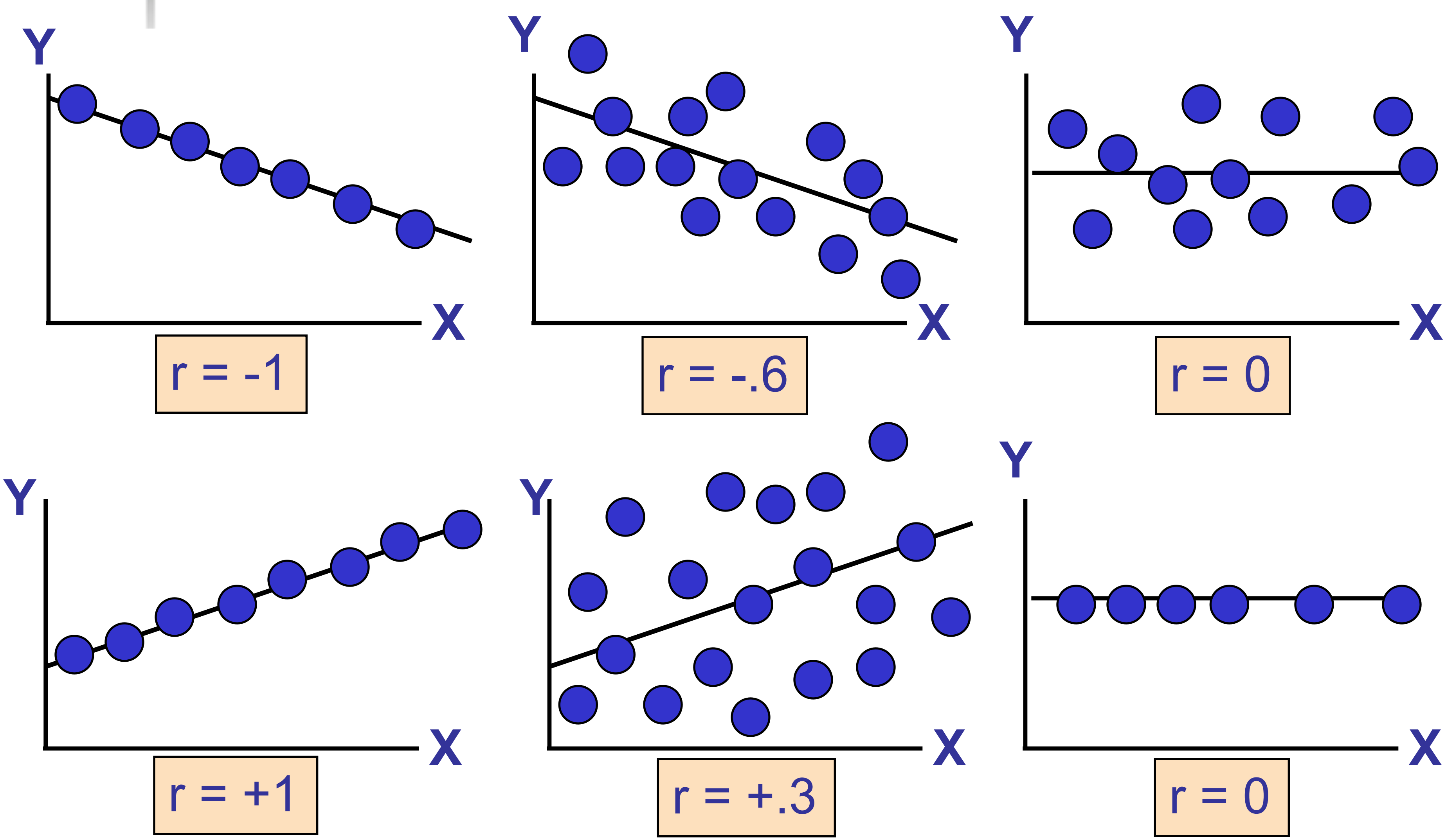
- შერჩევის კორელაციის კოეფიციენტი:

$$\boxed{r = \frac{Cov(x,y)}{s_X s_Y}}$$

r კორელაციის კოეფიციენტის თვისებები

- არ არის გამოსახული აბსოლუტურ ერთეულებში
- იცვლება –1-სა და1-ს შორის
- რაც უფრო ახლოსაა –1-თან, მით უფრო ძლიერია უარყოფითი წრფივი დამოკიდებულება
- რაც უფრო ახლოსაა 1-თან, მით უფრო ძლიერია დადებითი წრფივი დამოკიდებულება
- 0-თან სიახლოვეში, ზოგადად წრფივი დამოკიდებულება სუსტდება

მონაცემთა წერტილოვანი დიაგრამები სხვადასხვა r-ით



წრფივი კავშირის ძიება

- წრფივი განტოლება „უმცირეს კვადრატთა“ მეთოდის გამოყენებით

$$\hat{y} = b_0 + b_1 x$$

- $\hat{y}$-ს ეწოდება *დამოკიდებული* ცვლადი
- x-ს კი ეწოდება - *დამოუკიდებელი* ცვლადი

$$b_1 = \frac{Cov(x,y)}{s_x^2}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$