


# Կենսավիճակագրության ներածություն

Ganna Sakanyan

June 4, 2022





# Բննարկվող թեմաներ

- Կենսավիճակագրության նշանակությունը և դերը առողջապահության ոլորտում
- Փոփոխականների տեսակները, չափման սանդղակները
- Նկարագրական վիճակագրություն
- Վիճակագրական եզրակացություններ
- Վիճակագրական մեթոդների ընտրություն

# Ի՞նչ է վիճակագրությունը

- ▶ Վիճակագրությունը գիտություն է, որն ուսումնասիրում է զանգվածային երևույթներն ու գործընթացները բնութագրող քանակական տվյալների հավաքման, մշակման, վերլուծության և մեկնաբանման մեթոդները

# Ի՞նչ նպատակով է օգտագործվում վիճակագրությունը առողջապահության և բժշկության ոլորտներում

- ▶ Բնակչության առողջությանը, բժշկական հաստատությունների գործունեությանը վերաբերվող տեղեկատվության մշակում և վերլուծություն
- ▶ Հանրային առողջապահության և բժշկության ոլորտներում գիտական հետազոտություններ կազմակերպման և դրանց արդյունքների վերլուծություն
- ▶ Մշակված առողջապահական ծրագրերի արդյունավետության գնահատում
- ▶ Կանխատեսումներ

# Ինչու՞ է պետք վերլուծել տեղեկատվությունը առողջապահության ոլորտում

- Հիմնախնդիրների լուծում և անհրաժեշտ որոշումների կայացում
  - բնակչության առողջության բարելավում
  - բուժօգնության որակի բարելավում
  - բուժօգնության կազմակերպում
  - առողջապահության պլանավորում
  - առողջապահության կառավարում
  - Առողջապահության ֆինանսավորում
  - քաղաքականության մշակում

# Տեղեկատվատեխնոլոգիայի տեսակները առողջապահության ոլորտում

- ▶ Բնակչության առողջությանը վերաբերվող
- ▶ Կլինիկական
- ▶ Վարչական
- ▶ Ֆինանսական

# Բնակչության առողջական վիճակը նկարագրող տվյալների հիմնական աղբյուրները

- ▶ էլեկտրոնային բազաների տվյալներ
- ▶ վիճակագրական հաշվետվություններ
- ▶ առաջնային բժշկական փաստաթղթեր
- ▶ նպատակային սկրինինգ հետազոտություններից ստացված տվյալներ



# Կյանքի պայմանների և կենսակերպի մասին տեղեկատվության ստացման եղանակները

- ▶ անկետավորում
- ▶ հարցազրույց
- ▶ անկետա-հարցազրույց
- ▶ ուղիղ դիտումներ







# Հիմնական վիճակագրական մեթոդներ

- ▶ Նկարագրական՝ թվային և վիզուալ
- ▶ Վերլուծական
- ▶ Կանխատեսական (մոդելավորման)

# Հուսալիություն և վավերականություն

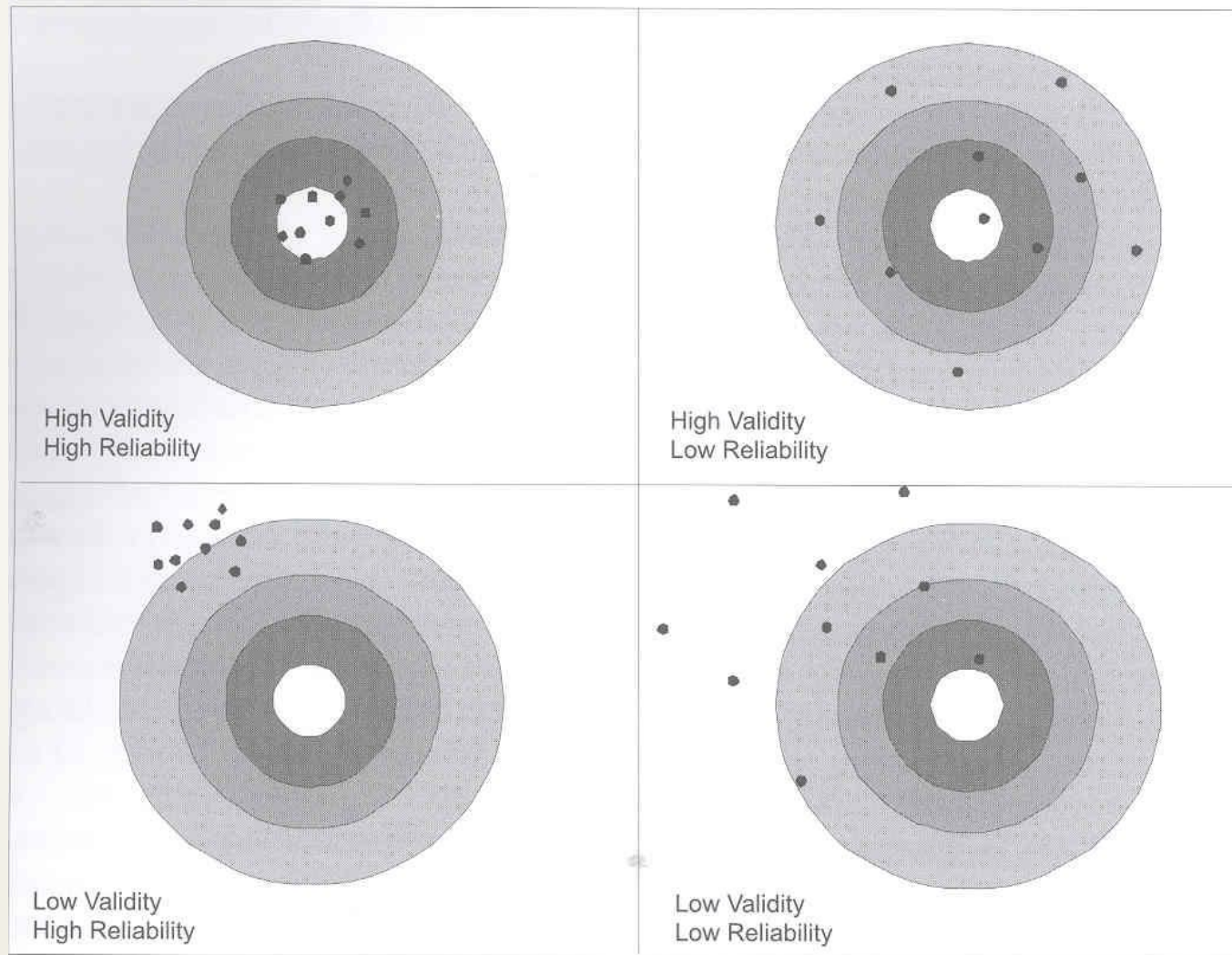
- ▶ Հուսալիություն՝ կայուն արդյունքներ տալու ունակություն
  - Բուն գործիքի տարատեսակություն
  - Բուն ցուցանիշի տատանումներ
  - Դիտորդի սխալներ
  - Սխալներ դիտորդների միջև
- ▶ Վավերականություն՝ ունակություն չափել (հայտնաբերել) հենց այն, ինչ պետք է որոշվի (ենթարկվում է սխտեմատիկ սխալների)

# Միավների առաջացման աղբյուրներ

- ▶ *Պատահական սխալ*
  - ▶ Չափումների փուլ
  - ▶ Տվյալների գրանցում
  - ▶ Տվյալների բազայի ստեղծման փուլում
- ▶ *Միատեմատիկ սխալ*
  - ▶ Հետազոտության տարբեր փուլերում (պլանավորման և վերլուծության ), կապված
    - ▶ Ընտրանքի ձևավորում
    - ▶ Չափում և գրանցում
    - ▶ Գործիքի ընտրություն

# Հուսալիություն և վավերականություն

12



# Փոփոխականների տեսակներ

## ➤ Որակական

- ✓ Դիխոտոմիկ
- ✓ Պոլիխոտոմիկ

## ➤ Քանակական

- ✓ Դիսկրետ
- ✓ Անըդհատ (շարունակական)

## ➤ Չափման սանդղակներ

- ✓ Անվանումների սանդղակ (նումինալների, պիտակների)
- ✓ Կարգային սանդղակ
- ✓ Միջակայքերի սանդղակ (ինտերվալների)
- ✓ Հարաբերությունների սանդղակ

# Տվյալների չափման մեթոդներ

- Հաճախ նույն տվյալները կարող են չափվել տարբեր սանդղակներում
  - ✓ Մարմնի զանգվածը (տարիքը) հարաբերությունների սանդղակում
  - ✓ Մարմնի զանգվածը (տարիքը) կարգային սանդղակում
  - ✓ Մարմնի զանգվածը (տարիքը) անվանական սանդղակում
- Պետք է կիրառվի զգուշությամբ
- Կարգերի սահմանները որոշվում են կամայականորեն, ինչը կարող է բերել խեղաթյուրումների

# Տվյալների մեթոդներ

## ➤ Տվյալների խմբավորում

- Հաճախականության բաշխման աղյուսակներ
  - բացարձակ հաճախականություն
  - հարաբերական հաճախականություն
  - կումուլյատիվ հաճախականություն

## ➤ Տվյալների ամփոփում

- Քանական փոփոխականներ
  - Կենտրոնական միտման ցուցանիշներ
  - Ցրվածության մեծություններ
- Որակական փոփոխականներ
  - Հարաբերական մեծություններ

## ➤ Ստացված արդյունքների պատկերում

- Գրաֆիկներ
- Աղյուսակներ

## Բացարձակ, հարաբերական և կումուլյատիվ հաճախականություններ

Խոլեստերինի մակարդակ (մգ/100 մլ)	Թվաքանակ	Հարաբերական հաճախականություն (%)	Կումուլյատիվ հաճախականություն (%)
80-119	13	1.2	1.2
120-159	150	14.1	15.3
160-199	442	41.4	56.7
200-239	299	28.0	84.7
240-279	115	10.8	95.5
280-319	34	3.2	98.7
320-359	9	0.8	99.5
360-399	5	0.5	100.0
Ընդամենը	1 067	100.0	



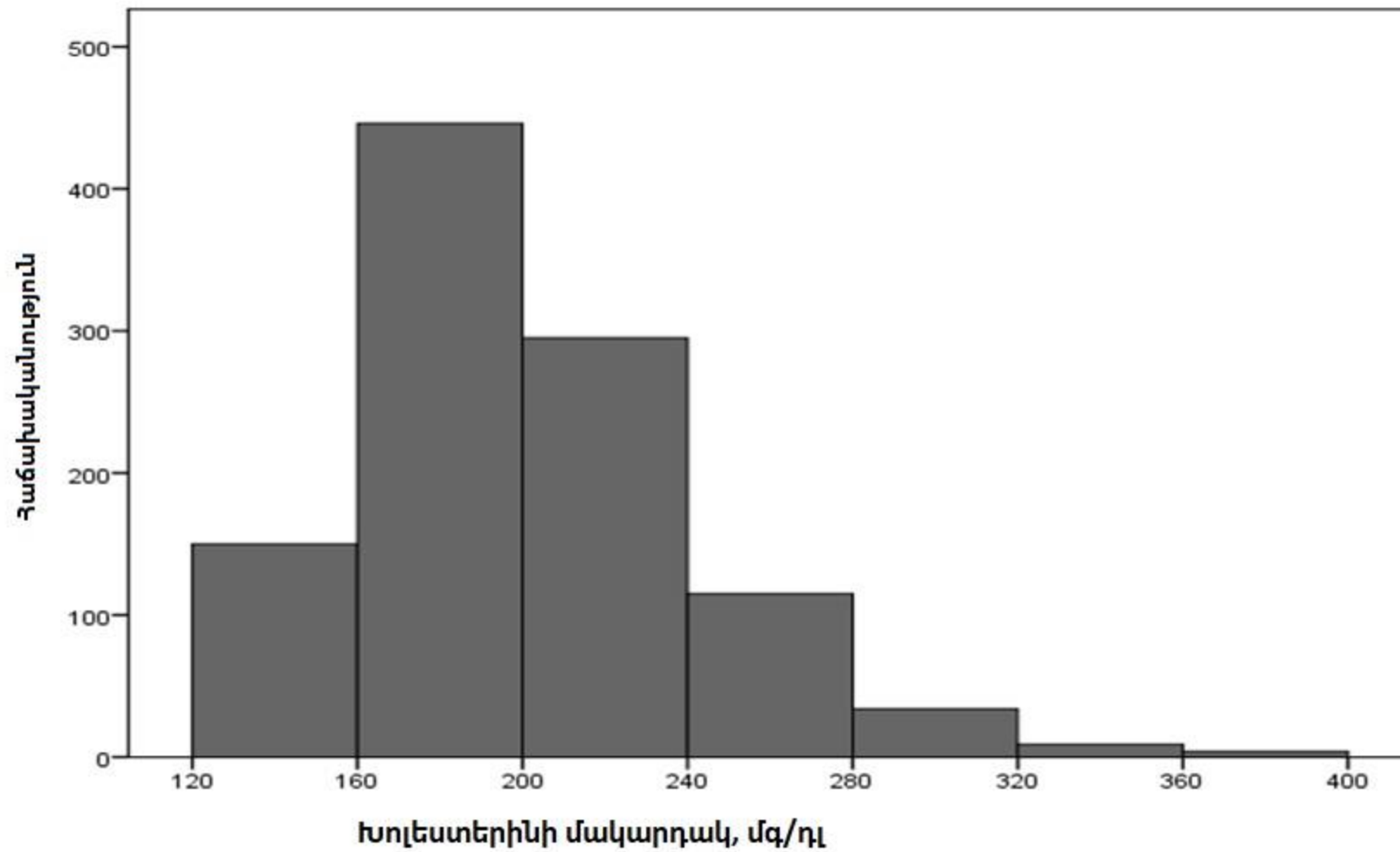
# Հարաբերական և կումուլյատիվ հաճախականություններ

Խոլեստերին (mg/100 ml)	Տարիք 25-34 տարեկան		Տարիք 55-64 տարեկան	
	Հարաբերական հաճախականություն (%)	Կումուլյատիվ Հաճախականություն (%)	Հարաբերական հաճախականություն(%)	Կումուլյատիվ Հաճախականություն
80-119	1.2	1.2	0.4	0.4
120-159	14.1	15.3	3.9	4.3
160-199	41.4	56.7	21.6	25.9
200-239	28.0	84.7	37.3	63.2
240-279	10.8	95.5	22.9	86.1
280-319	3.2	98.7	10.4	96.5
320-359	0.8	99.5	2.9	99.4
360-399	0.5	100	0.6	100

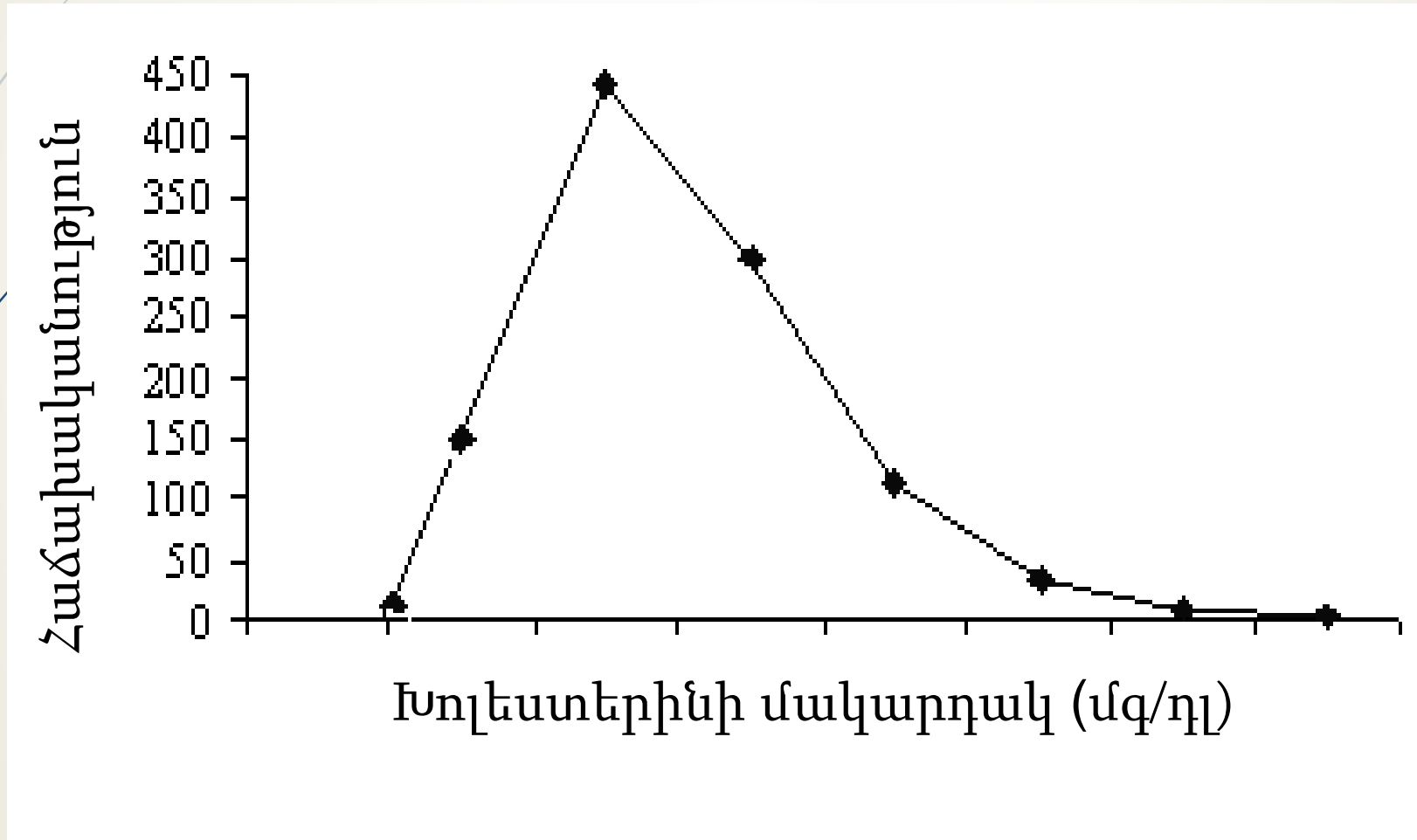
# Քանակական հատկանիշների գրաֆիկական պատկերում

- Հիստոգրամ
- Հաճախականությունների բաշխման  
բազմանկյուն (պոլիգոն)
- Կումուլյատիվ հաճախականությունների  
բաշխման կորագիծ

# Հիստոգրամ



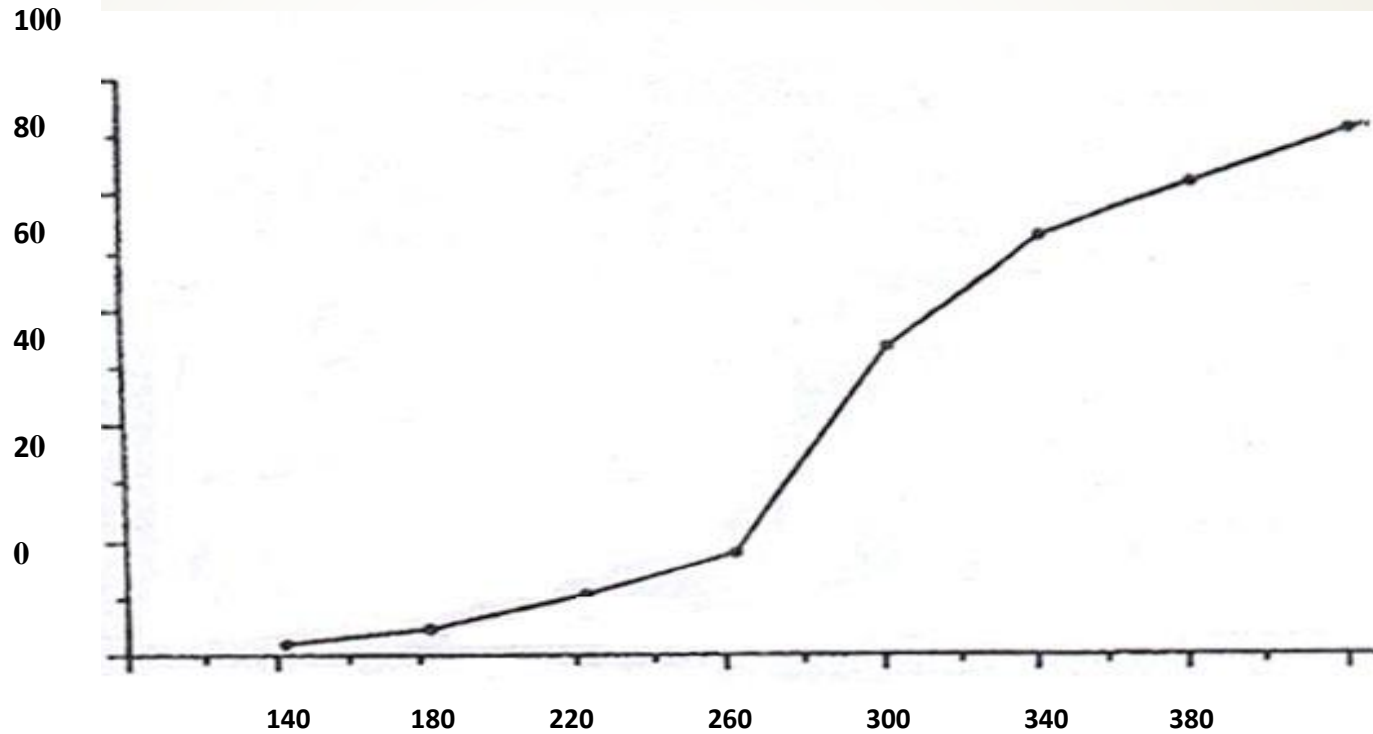
# Հաճախականությունների քաշիման բազմանկյուն (սյուիգոն)



# Կումուլյատիվ հաճախականությունների բաշխման կորագիծ

21

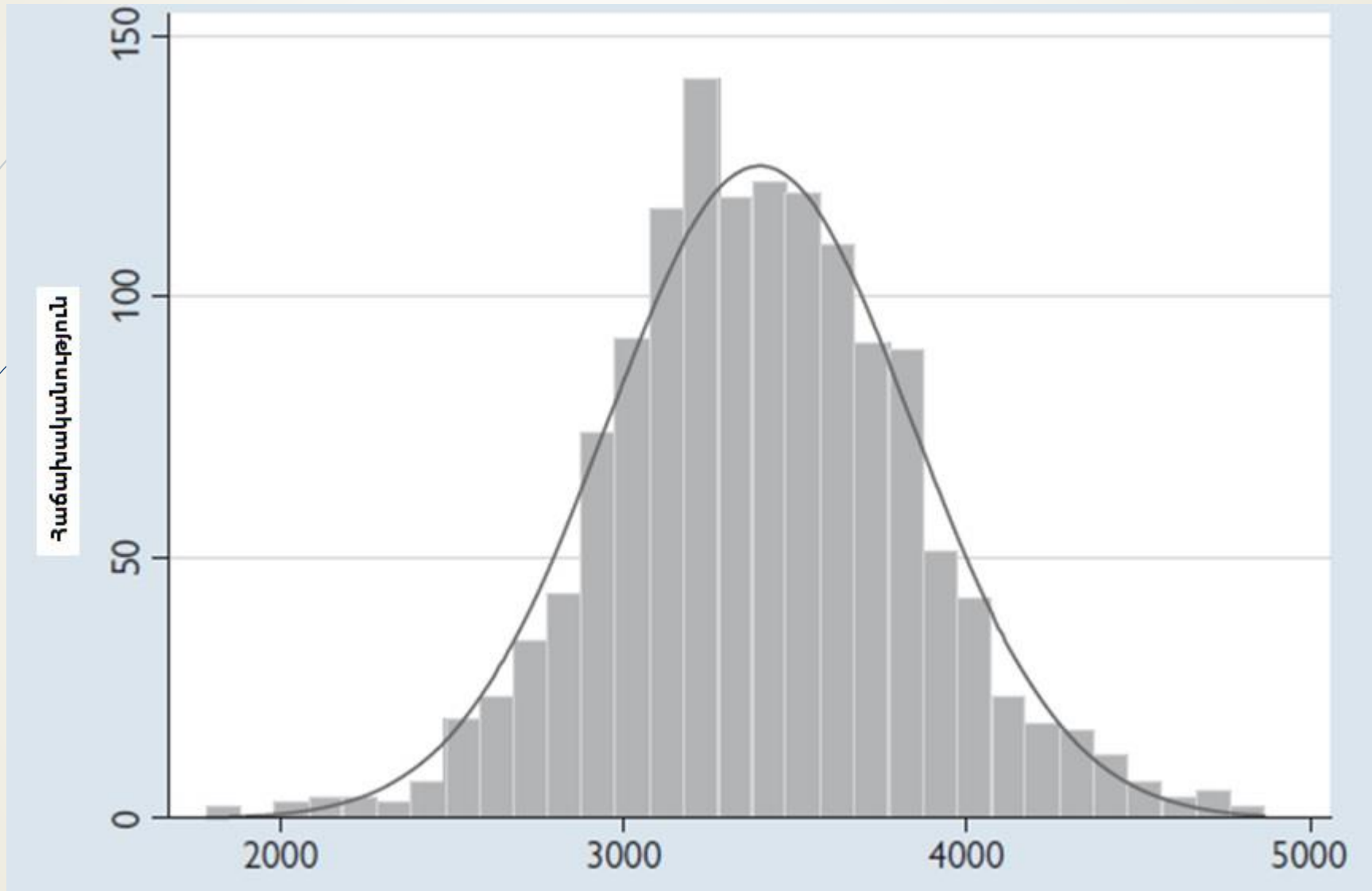
Կումուլյատիվ  
հաճախականություն (%)



Խտլեստերինի մակարդակ, մգ/դլ

# Նորմալ (սիմետրիկ) բաշխում

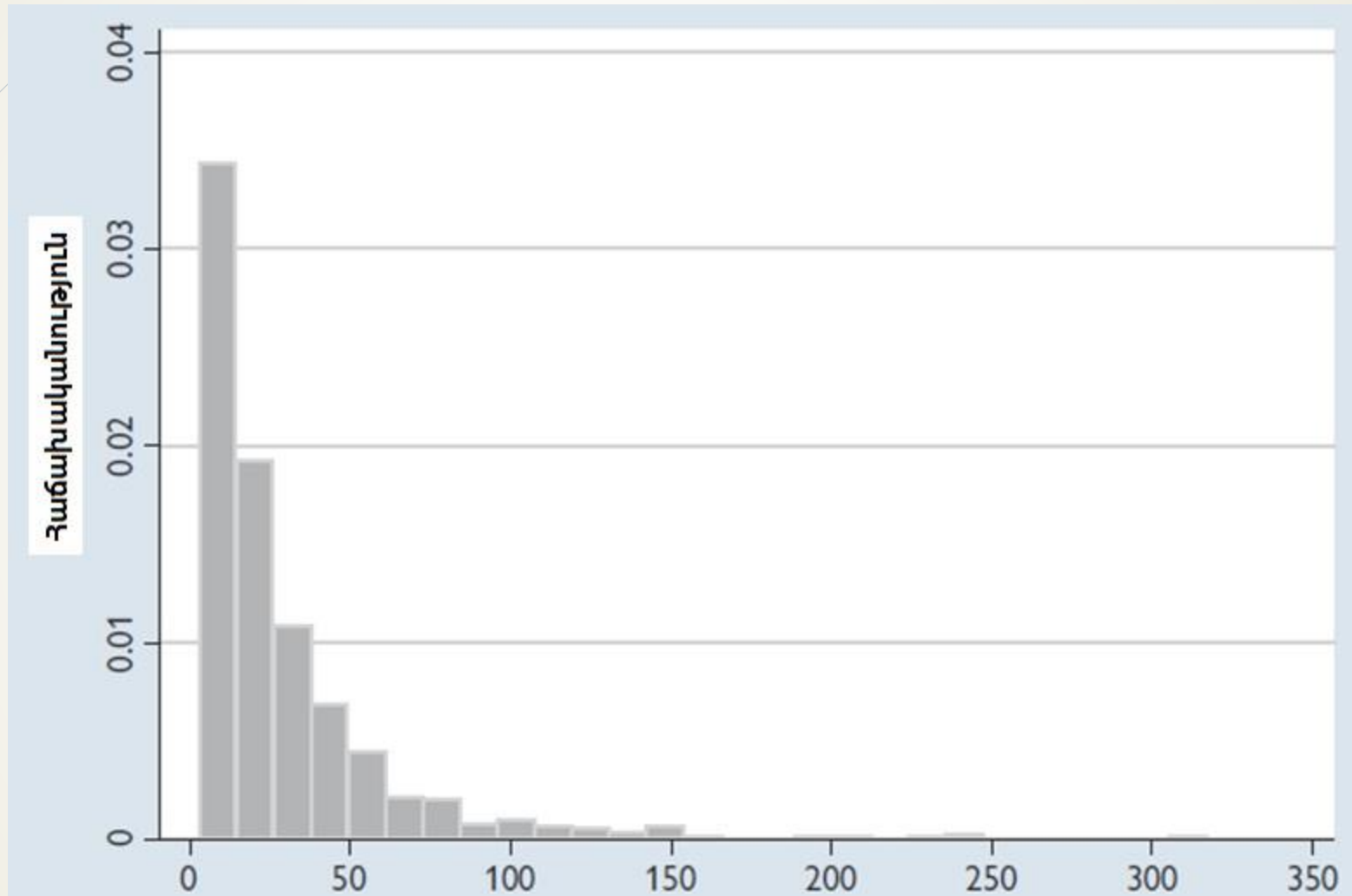
22



Նորածին երեխաների քաշի արժեքները (գր)

# Հաճախականությունների բաշխման ասիմետրիկ տեսակներ Դրական ասիմետրիա

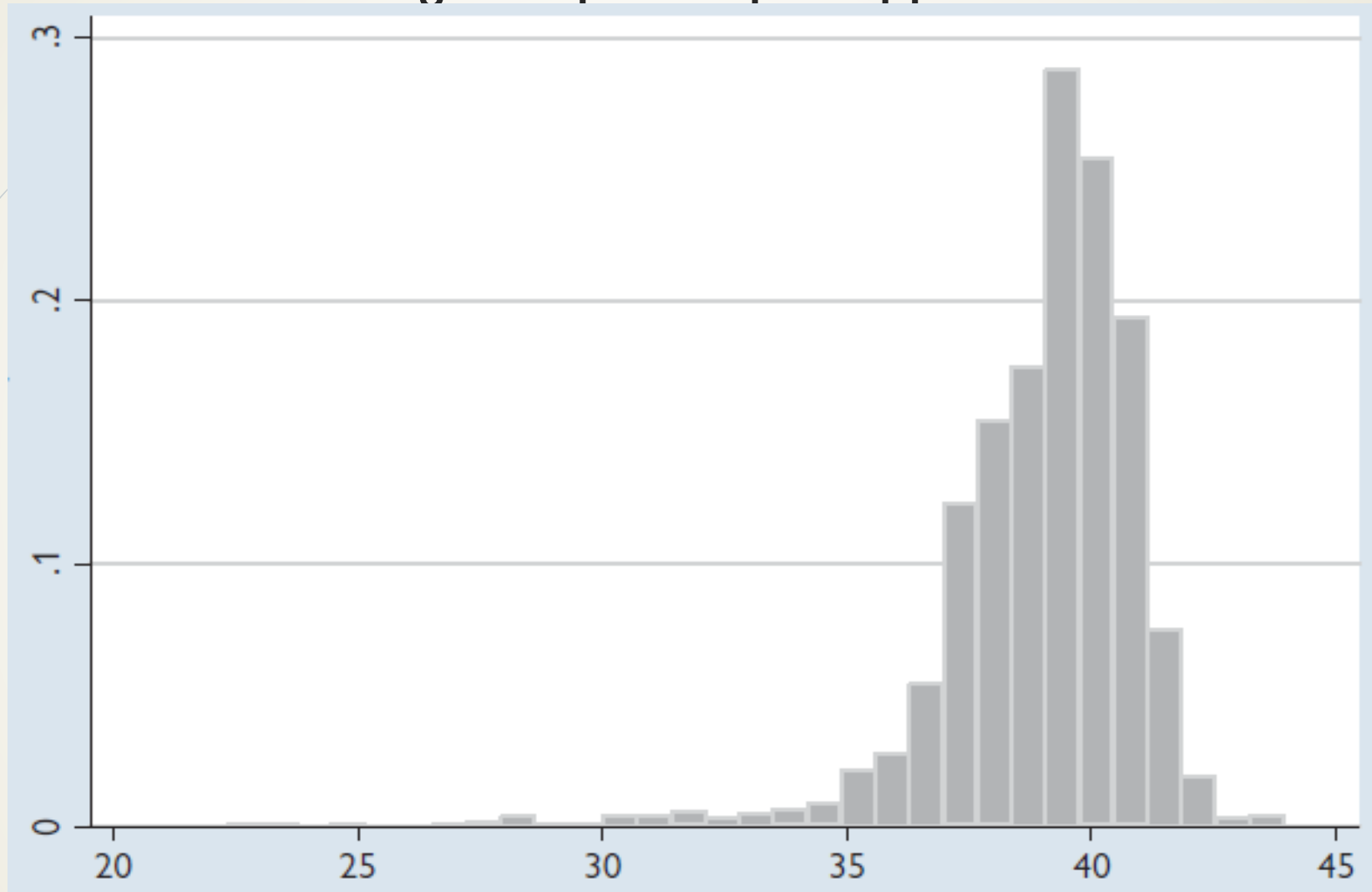
23



Ալկոհոլի օգտագործումը անցած շաբաթվա ընթացքում (գրամ)

# Հաճախականությունների բաշխման ասիմետրիկ տեսակներ Բացասական ասիմետրիա

24



Հոլության տևողությունը (ըստ շաբաթների)



# Կենտրոնական միտման մեծություններ

25

- **Միջին թվաբանական** ( $\bar{X}$  կամ  $\mu$ )՝ մեկ թվով ամփոփում է տվյալների ամբողջ շարքը:  $\bar{X}$  -ով նշում են ընտրանքի միջինը,  $\mu$ -ով՝ պոպուլյացիայի միջին թվաբանականը
- **Մեդիան /Կիսորդ** (**Me**)՝ այն արժեքն է, որը բաժանում է տվյալների կարգավորված շարքը 2 հավասար մասերի
- **Մոդ** (**Mo**)՝ տվյալների շարքում ամենահաճախակի հանդիպող արժեքն է: Օգտագործվում է հազվադեպ

# Միջին թվաբանական

26

- ▶ Առավել հաճախ կիրառվող մեծությունն է
- ▶ Արձագանքում է շարքում յուրաքանչյուր փոփոխմանը
- ▶ Շատ զգայուն է բաշխման ծայրային արժեքների ազդեցությանը
- ▶ Ասիմետրիկ բաշխման դեպքում միջին թվաբանականի փոխարեն օգտագործվում է մեդիանը

# Ցրվածության մեծություններ

27

- Լայնք (ամպլիտուդ)
- Ցրվածք (դիսպերսիա)
- Ստանդարտ շեղում
- Տարափոխման (վարիացիայի) գործակից
- Միջքվարթիլային ցրվածություն

# Հայնք

- ▶ Շարքի ամենամեծ և ամենափոքր արժեքների տարբերությունն է

$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$
121	110	114	100	160	130	130

$$A = 160 - 100 = 60$$

# Ցրվածք (դիսպերսիա)

29

- ▶ Շարքում եղած բոլոր արժեքների միջին շեղումն է սվյալ շարքի համար որոշած միջին թվաբանականից

$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$
121	110	114	100	160	130	130

$$\begin{aligned} S^2 &= \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} = \frac{(X_1 - \bar{X})^2 + \dots + (X_7 - \bar{X})^2}{n-1} = \frac{(121 - 123.6)^2 + \dots + (130 - 123.6)^2}{7-1} = \\ &= \frac{2247.72}{6} = 374.62 \approx 374.6 \end{aligned}$$

# Մտանդարտ շեղում

- ▶ Ցրվածության մեծություններից ամենահաճախակի օգտագործվող ցուցանիշն է
- ▶ Ցույց է տալիս հատկանիշի ցրվածության աստիճանը, արտահայտված տվյալ հատկանիշի չափման միավորով
- ▶ Ընտրովի հետազոտությունից ստացված ստանդարտ շեղումը նշվում է որպես «SD», իսկ պոպուլացիոն ստանդարտ շեղումը՝ որպես « $\sigma$ »
- ▶ Հաշվարկվում է հետևյալ բանաձևով՝

$$SD = \sqrt{s^2} = \sqrt{\frac{\sum (X - \bar{X})^2}{n-1}}$$

# Ստանդարտ շեղում

31

$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$
121	110	114	100	160	130	130

$$\begin{aligned} S^2 &= \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} = \frac{(X_1 - \bar{X})^2 + \dots + (X_7 - \bar{X})^2}{n-1} = \frac{(121 - 123.6)^2 + \dots + (130 - 123.6)^2}{7-1} = \\ &= \frac{2247.72}{6} = 374.62 \approx 374.6 \end{aligned}$$

$$SD = \sqrt{374.6} \approx 19.35$$

# Տարափոխման (վարիացիայի) գործակից

32

- Ցրվածության հարաբերական ցուցանիշ է
- Օգտագործվում է այն դեպքերում, երբ անհրաժեշտ է լինում համեմատել տարբեր չափման միավոր ունեցող հատկանիշների ցրվածության աստիճանը
- Հաշվարկի բանաձևն է՝

$$C_v = \frac{SD}{X} * 100\%$$

- Գնահատման սանդղակ  
մինչև 10% - թույլ տարափոխում  
10-20% - միջին  
20%-ից բարձր՝ ուժեղ



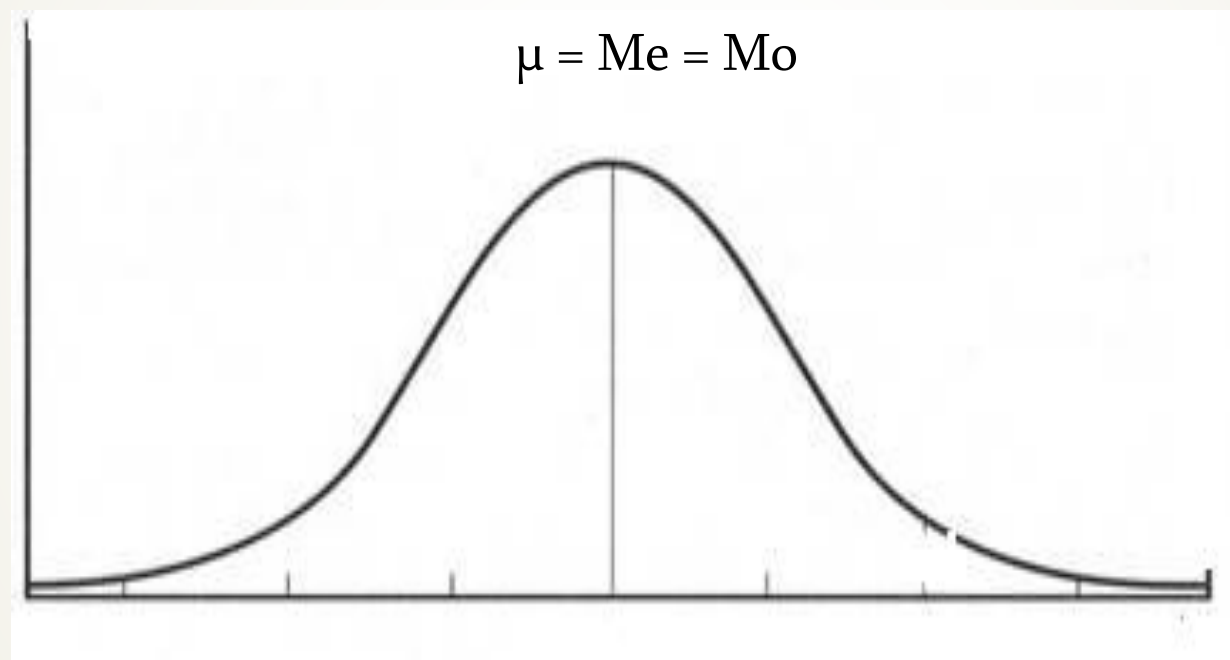
# Օրինակ

- ▶ Արյան գլյուկոզայի մակարդակի նորմայի սահմանները կազմում են 3.3- 5.5 մմոլ/լ, միջին թվաբանականը = 4.4 մմոլ/լ, ստանդարտ շեղումը = 0.55 մմոլ/լ
  - ▶  $C_v = 0.55 * 100\% / 4.4 = 12.5\%$
- ▶ Արյան pH-ի նորմայի սահմանները կազմում են 7.35- 7.45, միջին թվաբանականը = 7.4, ստանդարտ շեղումը = 0.025
  - ▶  $C_v = 0.025 * 100\% / 7.4 = 0.34\%$

# Օրինակ

- ▶ Արյան գլյուկոզայի մակարդակի նորմայի սահմանները կազմում են 3.3- 5.5 մմոլ/լ, միջին թվաբանականը = 4.4 մմոլ/լ, ստանդարտ շեղումը = 0.55 մմոլ/լ
  - ▶  $C_v = 0.55 * 100\% / 4.4 = 12.5\%$
- ▶ Արյան pH-ի նորմայի սահմանները կազմում են 7.35- 7.45, միջին թվաբանականը = 7.4, ստանդարտ շեղումը = 0.025
  - ▶  $C_v = 0.025 * 100\% / 7.4 = 0.34\%$

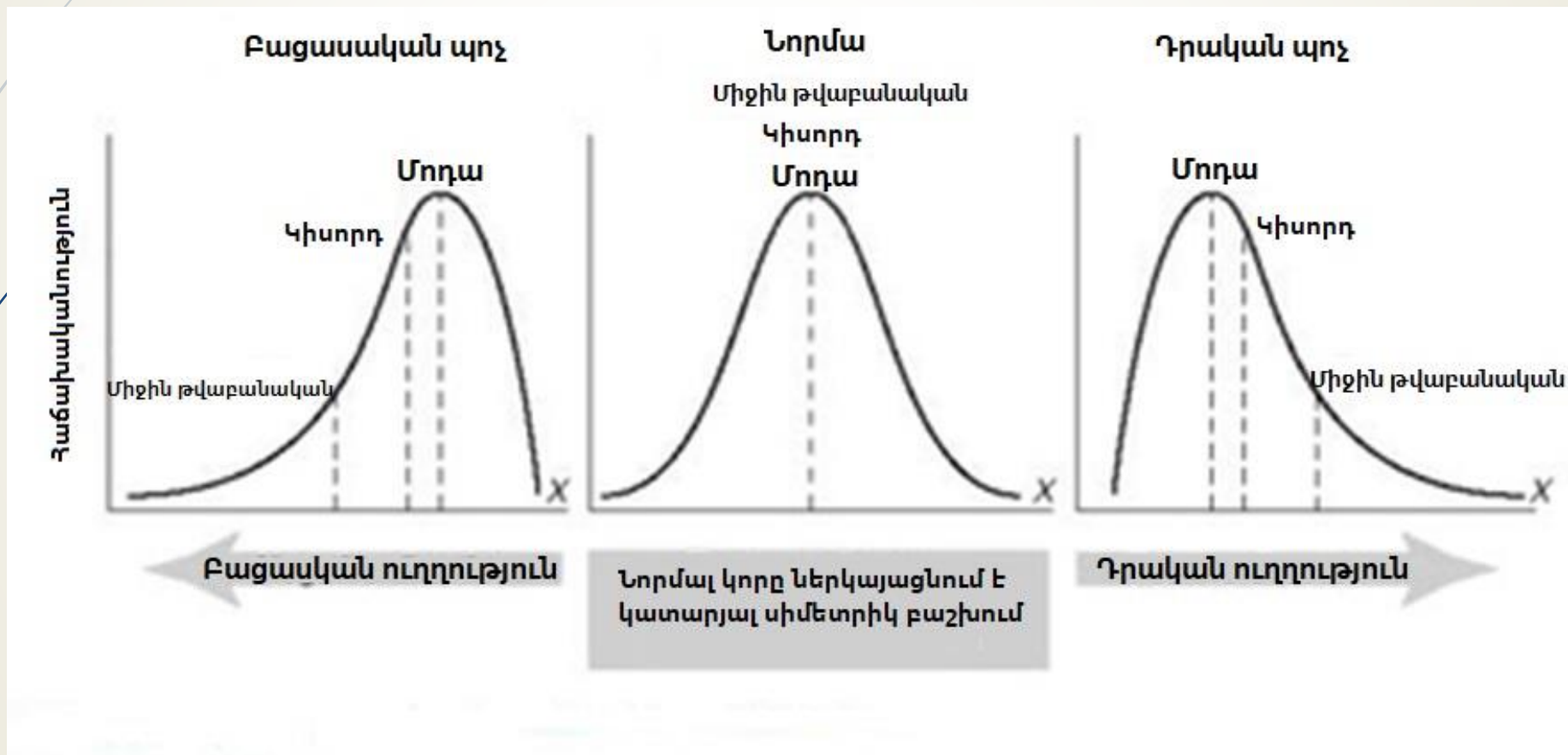
# Հաճախականությունների նորմալ բաշխում



## Նորմալ քաշխաման առանձնահատկություններ

- ▶ Վիճակագրության մեջ ամենահայտնի քաշխումն է
- ▶ Կարող է ընդունել արժեքներ  $-\infty$  մինչև  $+\infty$
- ▶ Նորմալ քաշխաման դեպքում  $\mu=Me=Mo$
- ▶ Նորմալ քաշխաման կորը սիմետրիկ է
- ▶ Կորի տակ գտնվող մակերեսը հավասար է 100%

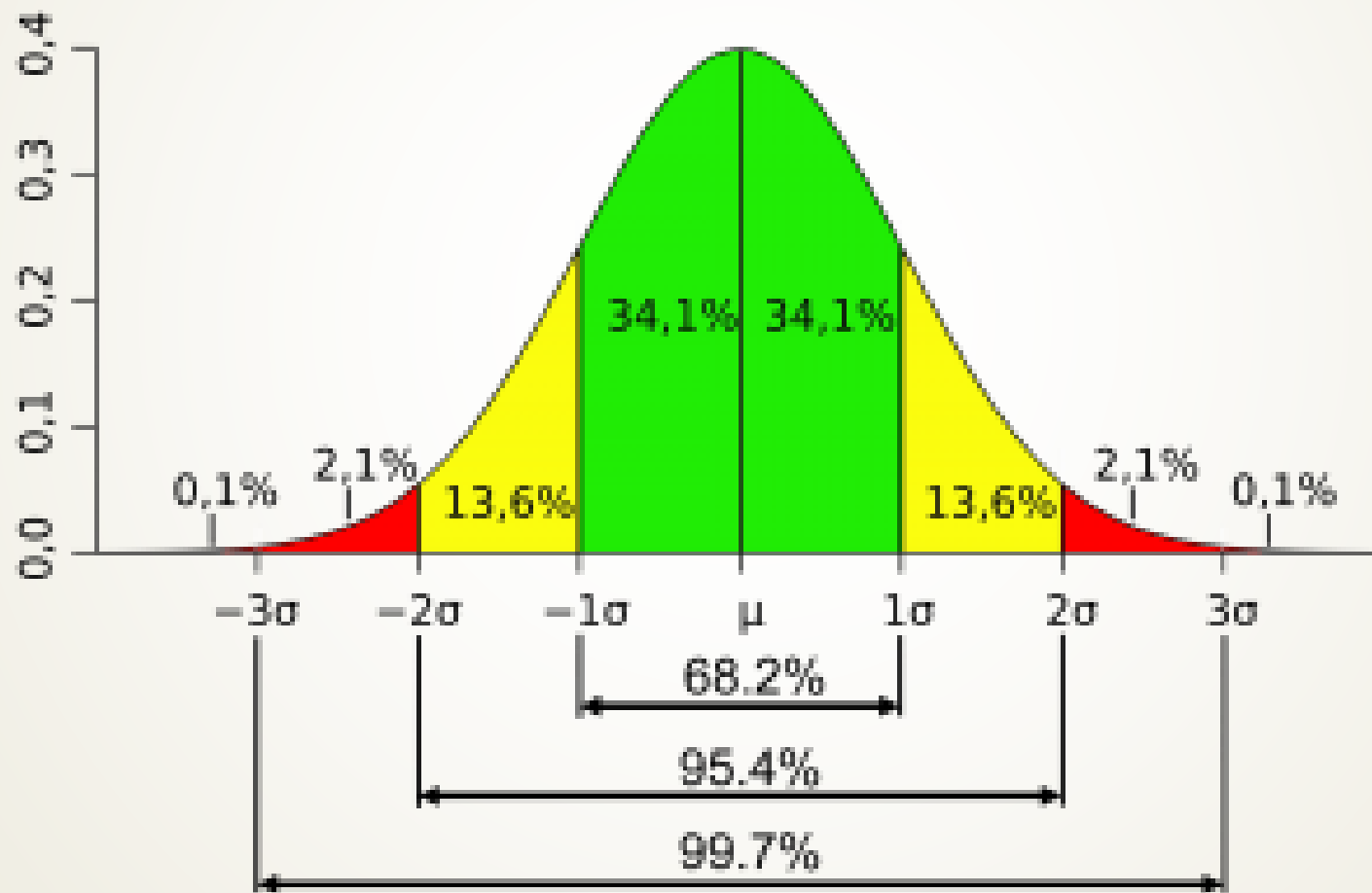
# Հաճախականությունների բաշխման ոչ նորմալ տեսակները



# Մտանդարտ շեղման 3 կանոններ

1. Նորմալ բաշխման դեպքում պոպուլյացիայի շուրջ 68.3% գտնվում է  $\mu \pm 1\sigma$  սահմաններում
2. Նորմալ բաշխման դեպքում պոպուլյացիայի շուրջ 95% գտնվում է  $\mu \pm 2\sigma$  սահմաններում
3. Նորմալ բաշխման դեպքում պոպուլյացիայի շուրջ 99.7% գտնվում է  $\mu \pm 3\sigma$  սահմաններում

# Մտանդարտ շեղման 3 կանոնները



# Z-ԱՐԺԵՔՆԵՐ (ՆՈՐՄԱՎՈՐԿԱԾ ՇԵՂՈՒՄ)

- ▶ Z-արժեքը ցույց է տալիս նորմալ բաշխման դեպքում ցանկացած մեծության հեռավորությունը միջին թվաբանականից
- ▶ Հաշվարկվում է բանաձևով՝

$$z_i = \frac{X_i - \mu}{\sigma}$$



# Օրինակ

- ▶ Նորածին երեխաների քաշի արժեքները ունեն նորմալ քաշիվածություն,  $\mu = 3000$  գ,  $\sigma = 1000$  գ  
Ո՞ր %-ն են կազմում այդ պոպուլյացիայում 4000 գրամից ավելի քաշով նորածինները

$$Z = (4000 - 3000) / 1000 = +1$$

Ըստ z-արժեքների աղյուսակի  $z=1$  արժեքին համապատասխանում է 15.9%:

- ▶ Ո՞ր %-ն են կազմում 2000-4000 գ քաշ ունեցող նորածինները
- ▶ Ո՞ր %-ն են կազմում  $>5000$  գ քաշ ունեցող նորածինները
- ▶ Ո՞ր %-ն են կազմում  $<3500$  գ քաշ ունեցող նորածինները

# Հ ԲԱՇԽՄԱՆ ԱՂՅՈՒՍԱԿ

$x$	Մասկերևույթ $-x$ և $+x$ արժեքների սկզբն	Մասկերևույթ կտրու արժեքում ( $-x$ և $+x$ )	Մասկերևույթ մեկ արժեքում ( $-x$ կամ $+x$ )
0,00	0,000	1,000	0,500
0,05	0,040	0,960	0,480
0,10	0,080	0,920	0,460
0,15	0,119	0,881	0,440
0,20	0,159	0,841	0,421
0,25	0,197	0,803	0,401
0,30	0,236	0,764	0,382
0,35	0,274	0,726	0,363
0,40	0,311	0,689	0,345
0,45	0,347	0,653	0,326
0,50	0,383	0,617	0,309
0,55	0,418	0,582	0,291
0,60	0,451	0,549	0,274
0,65	0,484	0,516	0,258
0,70	0,516	0,484	0,242
0,75	0,547	0,453	0,227
0,80	0,576	0,424	0,212
0,85	0,605	0,395	0,198
0,90	0,632	0,368	0,184
0,95	0,658	0,342	0,171
1,00	0,683	0,317	0,159
1,05	0,706	0,294	0,147
1,10	0,729	0,271	0,136
1,15	0,750	0,250	0,125
1,20	0,770	0,230	0,115
1,25	0,789	0,211	0,106
1,28	0,800	0,200	0,100
1,30	0,806	0,194	0,097
1,35	0,823	0,177	0,089
1,40	0,838	0,162	0,081
1,45	0,853	0,147	0,074
1,50	0,866	0,134	0,067
1,55	0,879	0,121	0,061
1,60	0,890	0,110	0,055
1,645	0,900	0,100	0,050
1,65	0,901	0,099	0,049
1,70	0,911	0,089	0,045
1,75	0,920	0,080	0,040
1,80	0,928	0,072	0,036
1,85	0,936	0,064	0,032
1,90	0,943	0,057	0,029
1,95	0,949	0,051	0,026
1,96	0,950	0,050	0,025
2,00	0,954	0,046	0,023

## Որակական փոփոխականները բնութագրող ցուցանիշներ

- ▶ Էքստենսիվ ցուցանիշ
- ▶ Ինտենսիվ ցուցանիշ
- ▶ Հարաբերակցության ցուցանիշ

# Էքստենսիվ ցուցանիշ

- Հանդիսանում է կառուցվածքային ցուցանիշ
- Ցույց է տալիս բաղադրամասի տեղը ամբողջի մեջ
- Բնութագրում է տվյալ միջավայրը՝ տվյալ ժամանակահատվածում
- Չի կարող օգտագործվել համեմատությունների համար
- Հաշվարկվում է հետևյալ բանաձևով՝

$$\frac{\text{Բաղադրամասն արտահայտող թիվ} * 100}{\text{Ամբողջն արտահայտող թիվ}}$$

Օրինակ՝ կարմրախտի տեսակարար կշիռը բոլոր ինֆեկցիոն հիվանդությունների մեջ

# Օրինակ

45

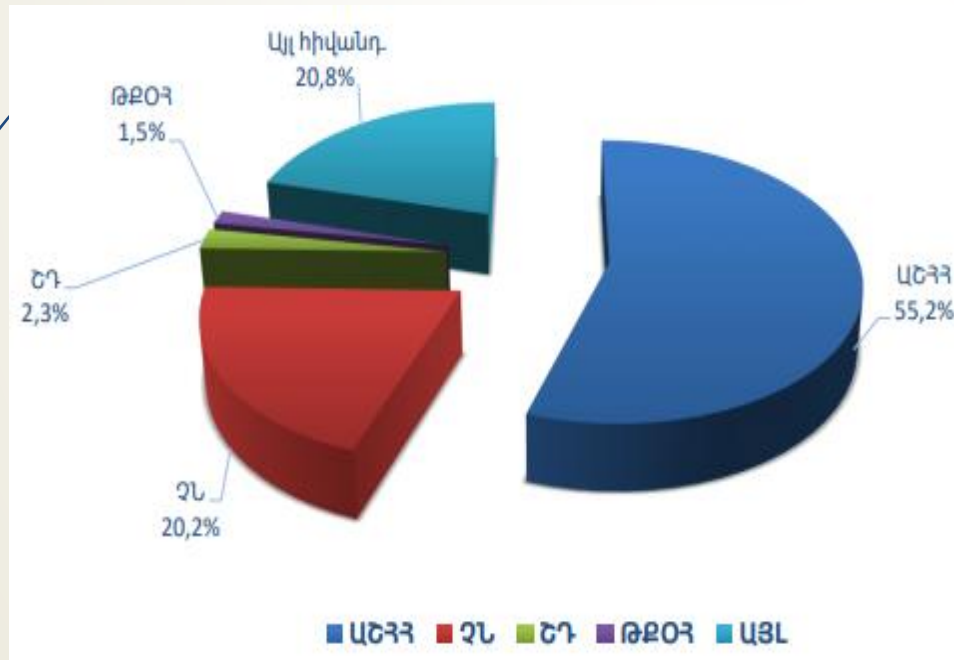
Հիվանդության անվանումը	N, %
սուր աղիքային վարակիչ հիվանդություններ	1000 դեպք (20%)
սուր շնչառական հիվանդություններ	1000 դեպք (20%)
շաքապային դիաբետ	100 դեպք (2%)
խալիպ	100 դեպք (2%)
այլ հիվանդություններ	2800 դեպք (56%)
Ընդամենը	5000 դեպք (100%)

Հիվանդության անվանումը	N, %
սուր աղիքային վարակիչ հիվանդություններ	500 դեպք (11.11%)
սուր շնչառական հիվանդություններ	1000 դեպք (22.22%)
շաքապային դիաբետ	100 դեպք (2,22%)
խալիպ	100 դեպք (2.22%)
այլ հիվանդություններ	2800 դեպք (62.22%)
Ընդամենը	4500 դեպք (100%)

# Էքստենսիվ ցուցանիշ՝ գրաֆիկական պատկերում

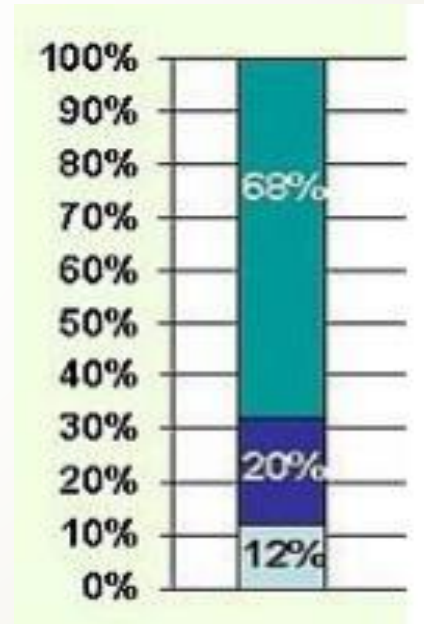
## Սեկտորային դիագրամ

ՀՀ բնակչության մահացության կշիռն (%) ըստ առավել տարածված հիվանդությունների, 2018



## Ներայունակային դիագրամ

Ա. քաղաքում ինֆեկցիոն հիվանդությունների կառուցվածքը տոկոսներով, 2018

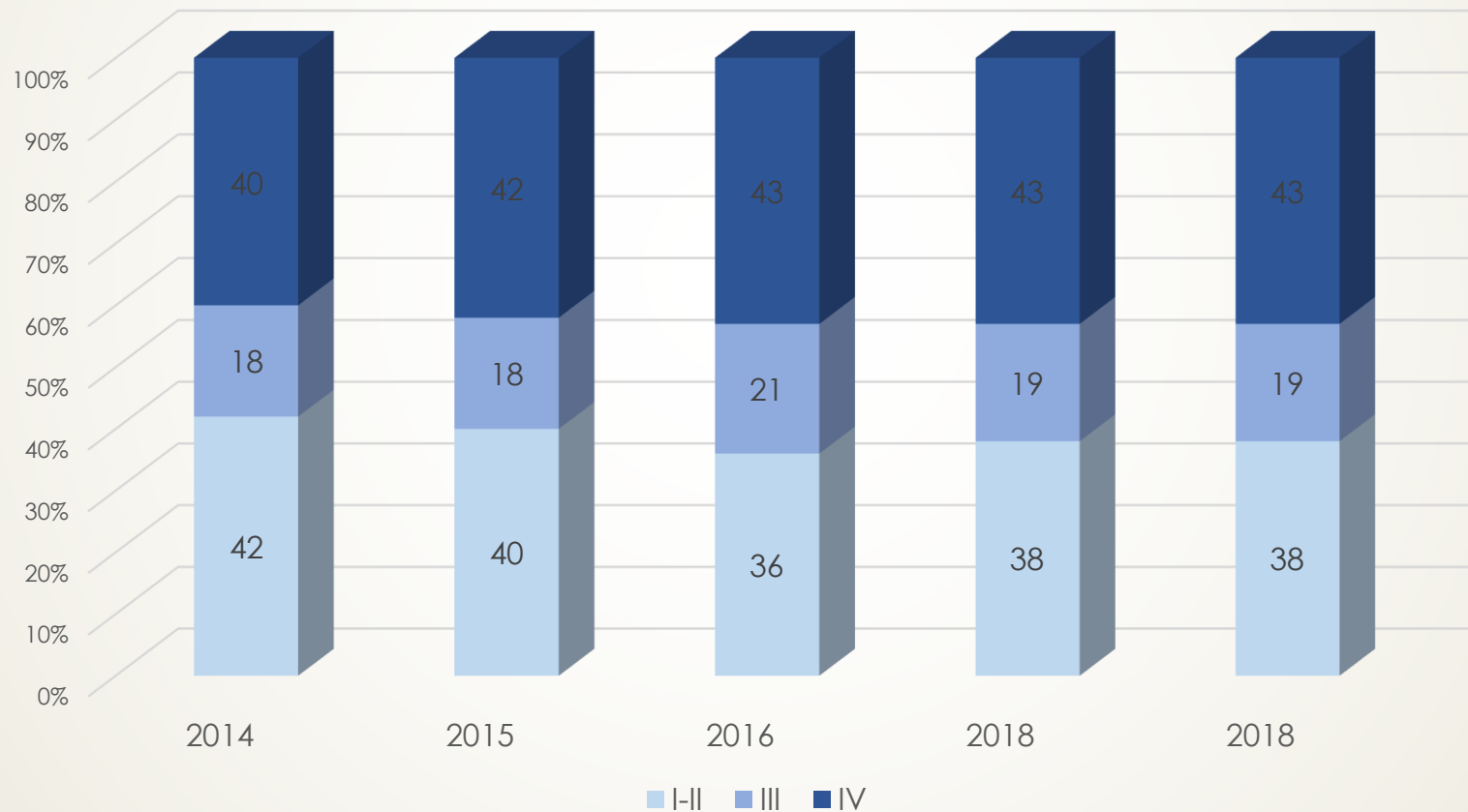


- Կարմրուկ
- Էպիդեմիկ պարոտիտ
- Այլ ինֆ. հիվանդություններ

# Օրինակ

47

Չարորակ նորագոյացությունների հայտնաբերումը ըստ հիվանդության փուլերի, 2014-2018 թթ.



# Ինտենսիվ ցուցանիշ

48

- ▶ Ցույց է տալիս երևույթի տարածվածությունը կամ հանդիպման հաճախականությունը տվյալ միջավայրում
- ▶ Հաշվարկի բանաձև՝

$$\frac{\text{Երևույթն արտահայտող բացարձակ թիվ} * 100 (1000 \text{ կամ } 10.000)}{\text{Միջավայրն արտահայտող թիվ}}$$

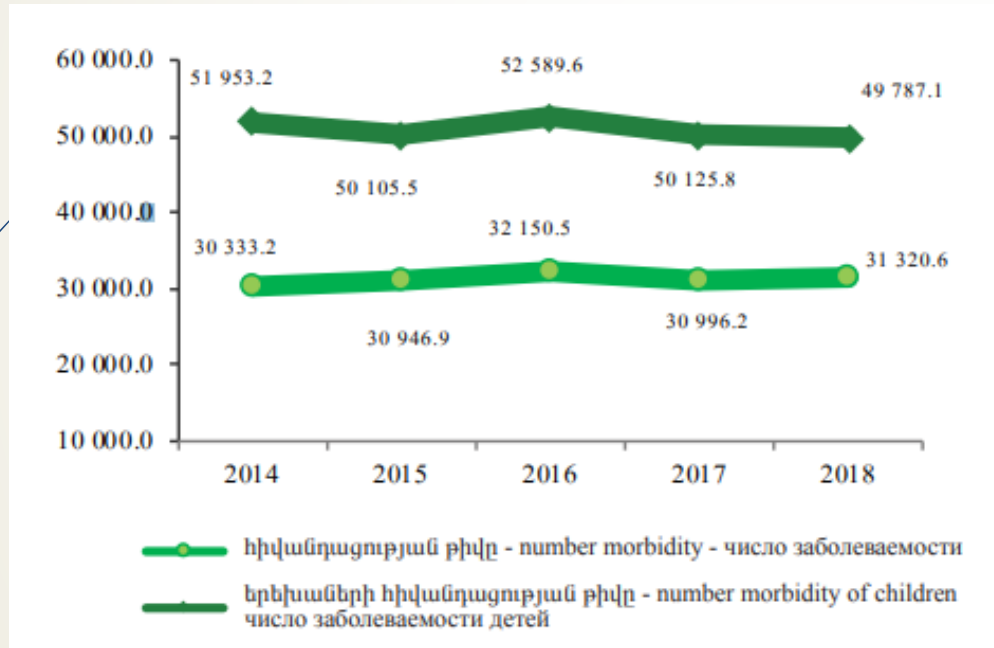
Օրինակ՝ ՄԱՀ-ի տարածվածությունը Հայաստանում



# Ինտենսիվ ցուցանիշ՝ գրաֆիկական պատկերում

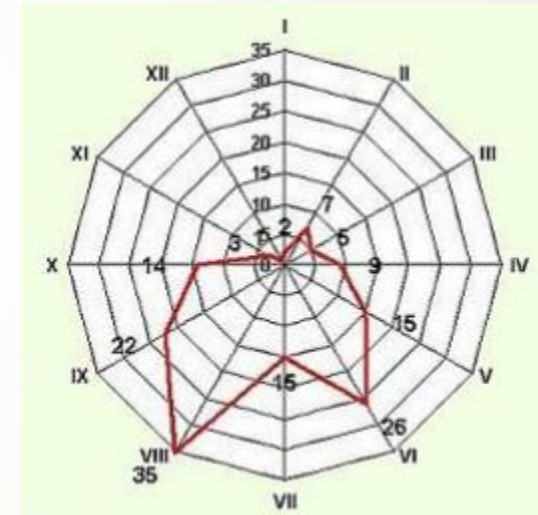
գծային դիագրամ

Հանդիպման հաճախականությունը



Բնակչության և 0-14 տարեկան երեխաների առաջնային հիվանդացությունը (100 000 բնակչի հաշվով)

շառավղային դիագրամ

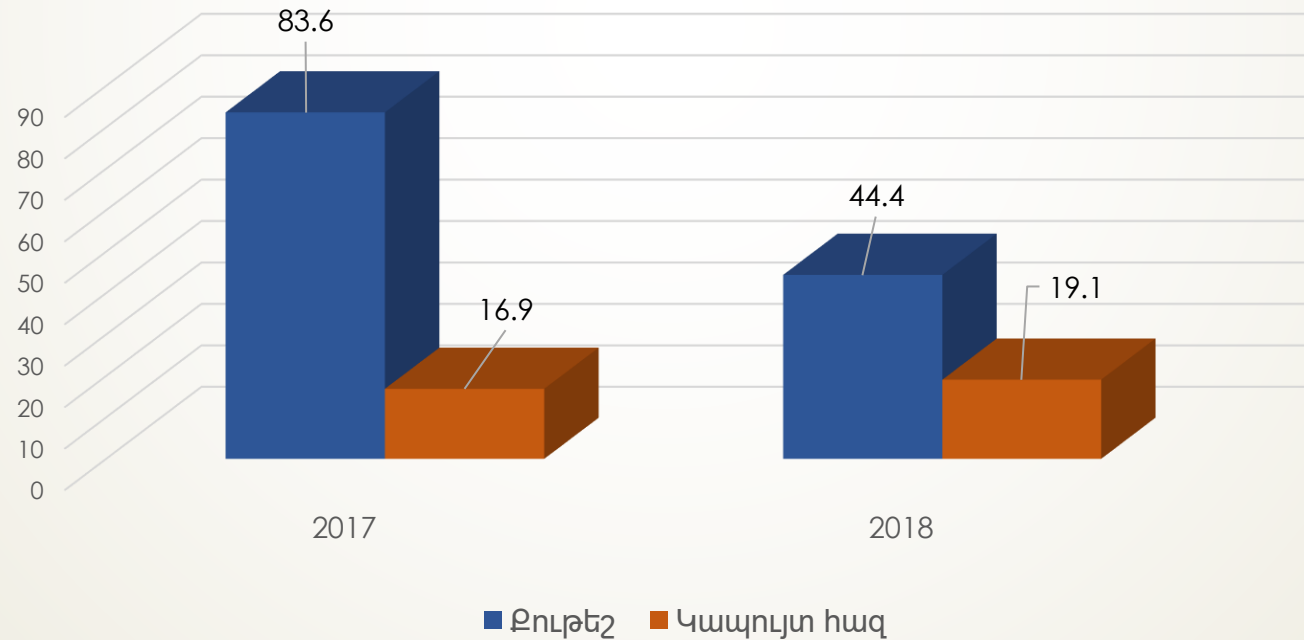


Դիզենտերիայի դեպքերի սեզոնային փոփոխությունները Ա. քաղաքում տվյալ տարում

# Ինտենսիվ ցուցանիշ՝ գրաֆիկական պատկերում

## Մյունակային դիագրամ

Ա. քաղաքի բնակչության հիվանդացությունը քուրեշով և  
կապույտ հազով 2017 և 2018 թթ. ընթացքում /100 000  
բնակչի հաշվով/



# Հարաբերակցության ցուցանիշ

51

- ▶ Երկու ինքնուրույն, իրար հետ կապ չունեցող վիճակագրական ամբողջությունների փոխհարաբերությունն է

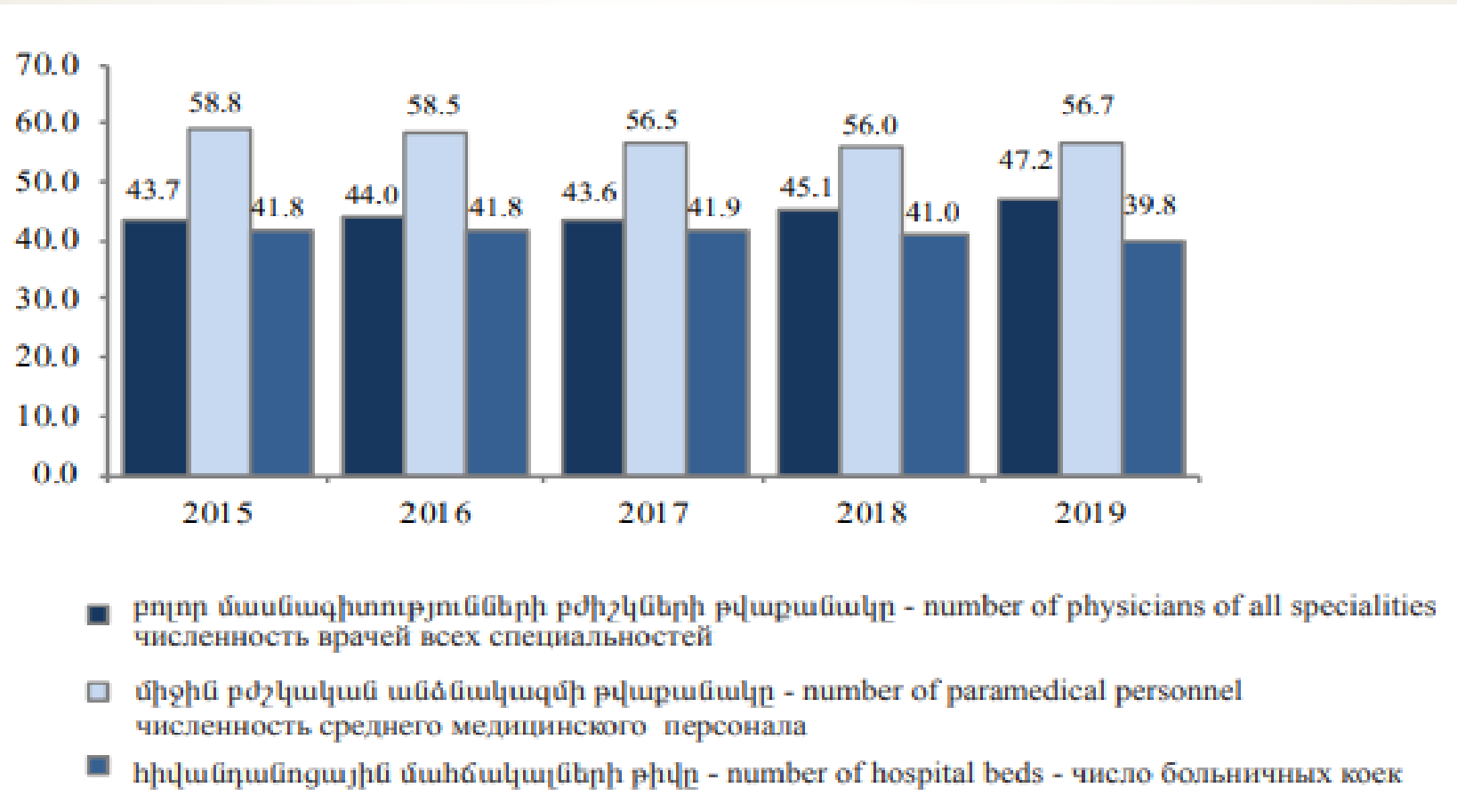
- ▶ Հաշվարկում է հետևյալ բանաձևով՝

$$\frac{\text{Երկույթն արտահայտող բացարձակ թիվ} \times 100 (1000 \text{ կամ } 10.000)}{\text{Միջավայրն արտահայտող թիվ}}$$

Օրինակ՝ բնակչության ապահովվածությունը բժիշկներով, հիվանդանոցային մահճակալներով

# Հարաբերակցության ցուցանիշ՝ գրաֆիկական պատկերում

Բնակչության ապահովվածությունը բժիշկներով,  
միջին բժշկական անձնակազմով և հիվանդանոցների  
մահճակալներով (10 000 բնակչության հաշվով)



## Զուգորդման աղյուսակներ

- ▶ Հասարակ են ու հարմար *որակական* հատկանիշների միջև փոխադարձ կապի հաստատման համար
- ▶ Դրանց օգտագործումը հնարավոր է նաև քանական հանկանիշների համար, եթե դրանք չափված են նոմինալների սանդակով

# Քառադաշտ աղյուսակներ (2x2)

54

a	b	a+b
c	d	c+d
a+c	b+d	

Յուլի գործակից

$$Q = \frac{ad - cb}{ad + cb}$$

$$\Phi = \frac{ad - cb}{\sqrt{(a+b)(c+d)(b+d)(a+c)}}$$

Կոնսինգենցիայի գործակից

# 55 Տվյալների ներկայացում – հիմնական հասկացություններ

- ▶ Համընդհանուր ամբողջություն /պոպուլյացիա/ դիտարկման միավորների հանրագումարն է: Ամբողջությունը բնութագրող տվյալները կոչվում են պարամետրեր
- ▶ Ընտրանք՝ վիճակագրական ամբողջության մասն է, բնութագրող թվային տվյալները կոչվում են վիճականի

Ցուցանիշ	Պարամետր	Վիճականի
Միջին թվաքանական	$\mu$	$\bar{X}$
Հարաբերություն	$\pi$	p
Ստանդարտ շեղում	$\sigma$	s
Դիսպերսիա	$\sigma^2$	$s^2$

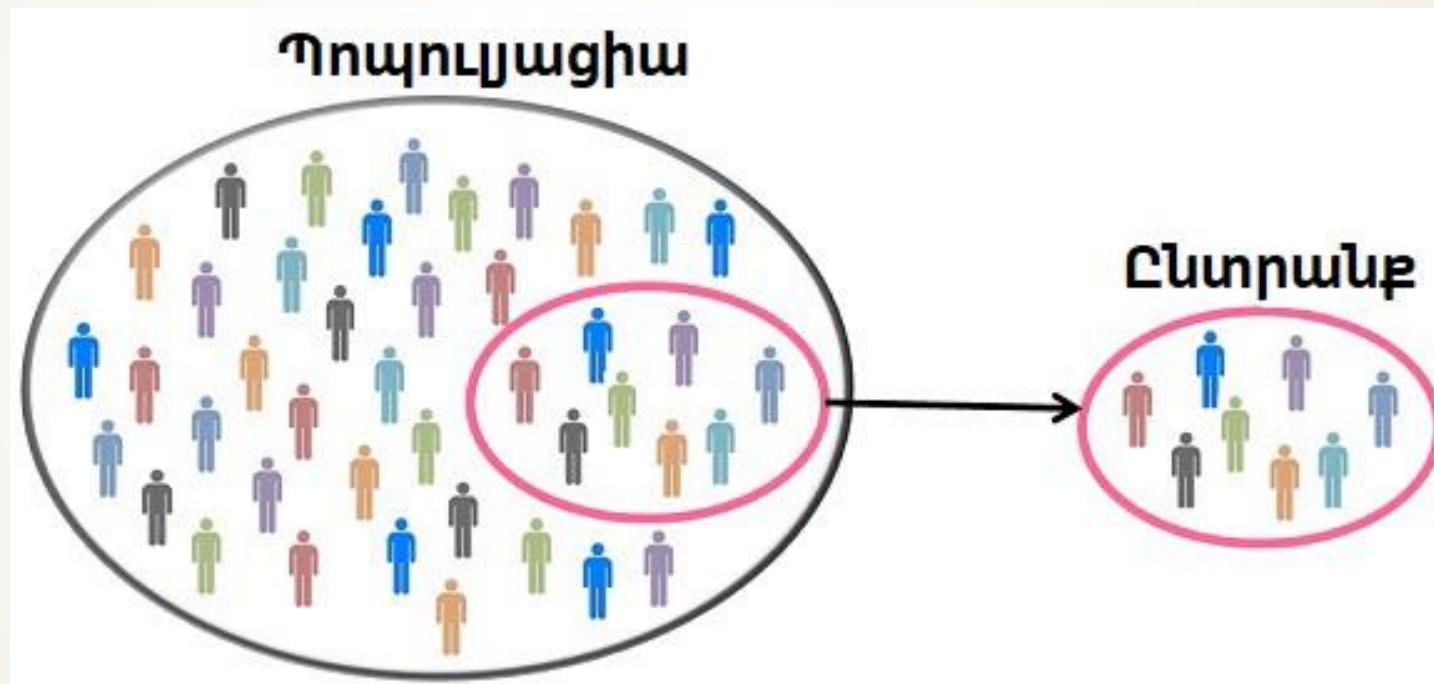
## Հնտրանքի ներկայացուցչականությունը, նրան ներկայացվող պահանջները

- ▶ Քանազական ներկայացուցչականություն
- ▶ Որակական ներկայացուցչականություն





# Պոպուլյացիա և ընտրանք



# Պատահական ընտրանքի ձևավորման եղանակները

- ▶ Պարզ պատահական ընտրանք
- ▶ Շերտավորված (ստրատիֆիկացված) պատահական ընտրանք
- ▶ Խմբային (կլաստերային) ընտրանք
- ▶ Սիստեմատիկ (մեխանիկական) ընտրանք



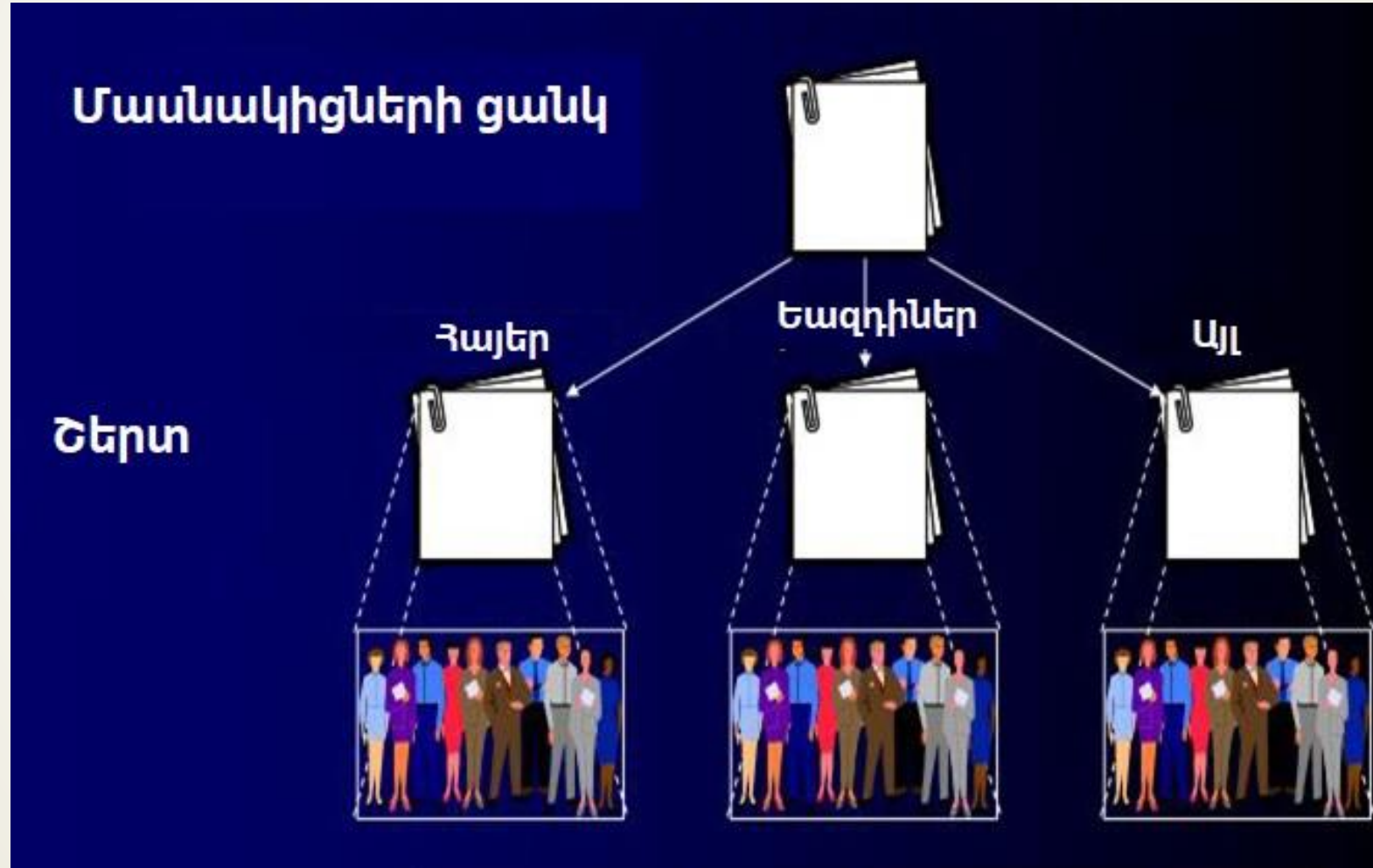
# Պարզ պատահական ընտրանք

Մասնակիցների ցանկ

Պատահական ենթախումբ

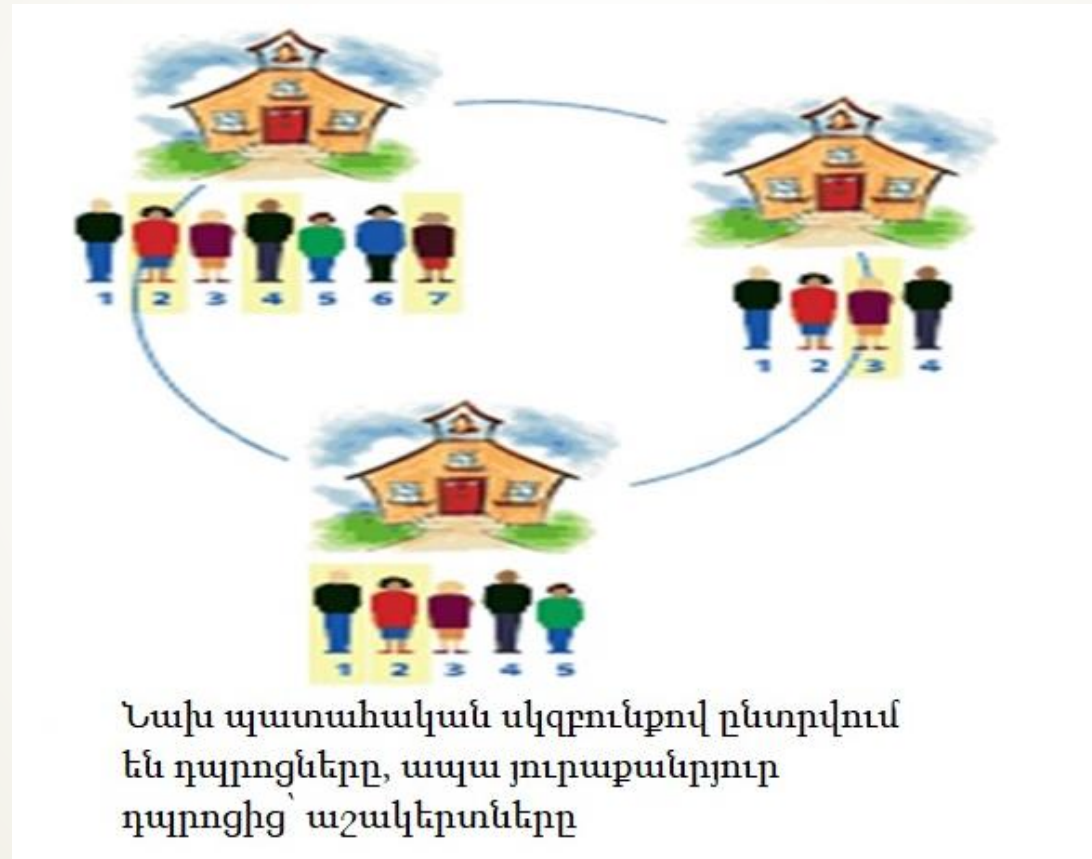


# Շերտավորված (ստրատիֆիկացված) պատահական ընտրանք



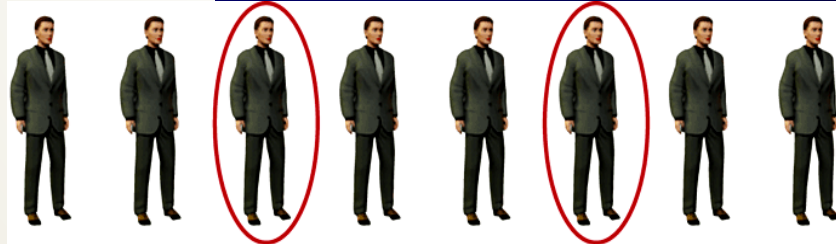
# Խմբային (կլաստերային) ընտրանք

Երկփուլային կլաստերային  
ընտրություն



# Միատեսակ (մեխանիկական) ընտրանք

$N = 100$



1	26	51	76
2	27	52	77
3	28	53	78
4	29	54	79
5	30	55	80
6	31	56	81
7	32	57	82
8	33	58	83
9	34	59	84
10	35	60	85
11	36	61	86
12	37	62	87
13	38	63	88
14	39	64	89
15	40	65	90
16	41	66	91
17	42	67	92
18	43	68	93
19	44	69	94
20	45	70	95
21	46	71	96
22	47	72	97
23	48	73	98
24	49	74	99
25	50	75	100

# Վիճակագրական եզրակացություն

Հետազոտողին հետաքրքրող համընդհանուր ամբողջության մասին եզրակացությունը հիմնված է ընտրանքային հետազոտությունից ստացված տվյալների վրա և նկարագրում է տվյալ եզրակացության անորոշության աստիճանը

# Վիճակագրական եզրակացություն

64

- ➔ **Գնահատում`** մեկ ընտրանքի արդյունքներից ստացված եզրակացություն է այն պոպուլյացիայի մասին, որից ձևավորվել տվյալ ընտրանքը
- ➔ **Վարկածի ստուգում`** համեմատվող երկու և ավելի խմբերում երևույթի մակարդակի տարբերության մասին եզրակացություն է



# Վիճակագրական եզրակացության մեթոդի ընտրություն

- ▶ Հետազոտության նպատակ, խնդիրներ
- ▶ Հետազոտության դիզայն
- ▶ Համեմատվող պոպուլյացիաների թվաքանակ
- ▶ Ուսումնասիրվող փոփոխականների տեսակ
- ▶ Արժեքների բաշխվածության տեսակ

# Վիճակագրական վերլուծական մեթոդները ըստ արժեքների բաշխվածության տեսակի

- ▶ Պարամետրիկ մեթոդներ (պահանջում են նորմալ բաշխվածություն)
  - ▶ Միջին թվաբանականի (ցուցանիշի) միջակայքի կառուցումը
  - ▶ t-տեստ անկախ ընտրանքների համար, զույգային ընտրանքների համար
  - ▶ Z-տեստ
  - ▶ Դիսպերսիոն վերլուծություն անկախ ընտրանքների համար (ANOVA)
  - ▶ Դիսպերսիոն վերլուծություն կրկնակի չափումների համար (Repeated Measures ANOVA)
- ▶ Ոչ պարամետրիկ մեթոդներ (չեն պահանջում նորմալ բաշխվածություն)
  - ▶ Ոչ կարգային տեստեր
    - ▶ Պիրսոնի համապատասխանության գործակից (Chi-square test)
    - ▶ Մակ-Նիմարի տեստ
  - ▶ Կարգային տեստեր
    - ▶ Մանս-Ուիլքոքսոնի՝ համապատասխանում է անկախ ընտրանքների t-թեստին
    - ▶ Ուիլքոքսոնի՝ համապատասխանում է զույգային ընտրանքների t-թեստին
    - ▶ Կռուսկալ-Ուոլլիսի թեստ՝ համապատասխանում է դիսպերսիոն վերլուծությանը անկախ ընտրանքների համար (independent ANOVA)
    - ▶ Ֆրիդմանի թեստ՝ համապատասխանում է դիսպերսիոն վերլուծությանը կրկնակի հետազոտությունների համար (repeated measures ANOVA)

# Վիճակագրական գնահատում

67

- Միջին ( $\mu$ ) և ստանդարտ շեղում ( $\sigma$ )՝ անհայտ պարամետրեր են, որոնք բնութագրում են պոպուլյացիան
- Միջին թվաքանակային  $\bar{X}$  ( $M$ ) և ստանդարտ շեղում  $SD$  ( $ST$ )՝ վիճականիներ են, որոնք հաշվարկվում են ընտրանքից և օգտագործվում են պարամետրերը գնահատելու համար
- Հիմնական հարցը, որին պատասխանում ենք՝ որքանո՞վ է ընտրանքային միջինը ճիշտ բնութագրում պոպուլյացիոն միջինը

# Ընտրանքային միջինների բաշխում

68

Օրինակ՝ ենթադրենք ամբողջ պոպուլյացիան կազմված է 10 օրդինատորից ( $N=10$ ), որոնց տարիքները հետևյալն են՝

27, 28, 31, 35, 35, 40, 42, 43, 50, 52 տարեկան

Պոպուլյացիոն միջին թվաբանականը և ստանդարտ շեղումը կլինեն հավասար՝

$\mu=38.3$  տարեկան և  $\sigma = 8.6$  տարեկան

Վերցնենք ընտրանքներ, որոնցում  $n=4$

Ինչպիսի՞ հնարավոր ընտրանքային միջիններ կարելի է ստանալ:

Հնարավորինս փոքրը կկազմի 30.25

Հնարավորինս մեծը կկազմի 46.75

Ամենաբնորոշը կլինի հավասար 38

# Պատահական ընտրանքների միջինների բաշխում

69

- Հնարավոր ընտրանքային միջինների տարատեսակությունը շատ ավելի քիչ է, քան ընտրանքի առանձին էլեմենտների
- Եթե պատկերենք բոլոր ընտրանքային միջինները հաճախականությունների բաշխման կորի տեսքով, ապա կպարզվի, որ ստացված բաշխումը շատ նման է նորմալին
- Եթե ընտրանքների թիվը մեծացվի և բոլոր ստացված ընտրանքային միջինները տեղադրվեն բաշխման կորի վրա, ապա բացահայտվում է, որ դրանց բաշխումը իսկապես նորմալ է

# Կենտրոնական սահմանային թեորեմ

- Ընտրանքային միջինների բաշխման կորը միշտ լինում է նորմալ, անկախ պոպուլյացիայում հատկանիշի բաշխման տեսակից
- Բոլոր ընտրանքային միջինների միջինը հավասար է ողջ պոպուլյացիայի իրական միջինին
- Բոլոր ընտրանքային միջինների ստանդարտ շեղումը կոչվում է ստանդարտ սխալ և կախված է ինչպես ընտրանքի ստանդարտ շեղումից, այնպես էլ ընտրանքի ծավալից

# Մտանդարտ սխալ

- ▶ Մտանդարտ սխալը (միջին թվաբանականի միջին սխալ՝  $SE, m$ ) – օգտագործվում է ընտրանքային միջինների տարատեսակության բնութագրման համար

$$SE(m) = SD / \sqrt{n}$$

- ▶ Որքան տարատեսակ են ընտրանքի միավորները, այնքան տարատեսակ են նրանցից ստացված միջինները

# Ընտրանքային միջինների բաշխում

72

- Պոպուլյացիայի օրդինատորների ( $N=10$ ) տարիքային բաշխում
  - տարիքների 68%-ը կգտնվի ( $\mu \pm 1\sigma$ )-ի միջակայքում
  - տարիքների 95%-ը կգտնվի ( $\mu \pm 2\sigma$ )-ի միջակայքում
  - տարիքների 99%-ը կգտնվի ( $\mu \pm 3\sigma$ )-ի միջակայքում
- Ընտրանքների ( $n=4$ ) տարիքային միջինների բաշխում
  - 68%-ը կգտնվի ( $\bar{X} \pm 1 SD/\sqrt{n}$ )-ի միջակայքում
  - 95%-ը կգտնվի ( $\bar{X} \pm 2 SD/\sqrt{n}$ )-ի միջակայքում
  - 99%-ը կգտնվի ( $\bar{X} \pm 3 SD/\sqrt{n}$ )-ի միջակայքում



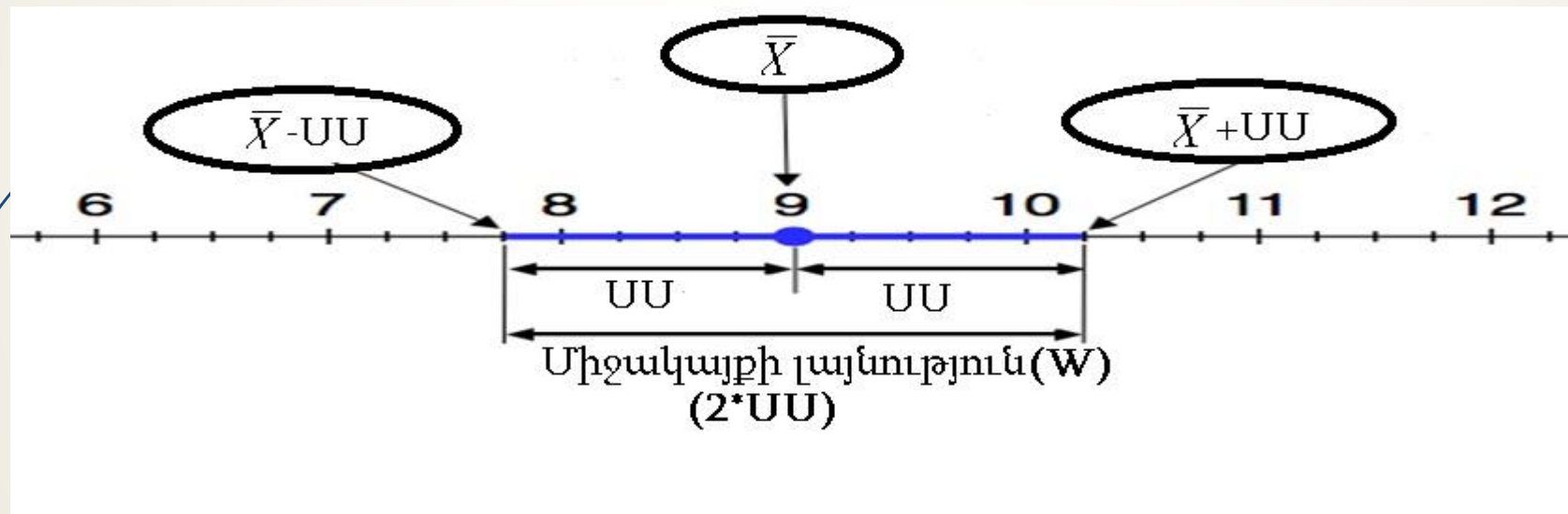
# Միջին թվաբանականի վստահության միջակայք

73

- ▶ **Վստահության միջակայք (CI)**՝ միջակայք է, որտեղ կգտնվի համընդհանուր ամբողջության իրական, բայց անհայտ միջին թվաբանականը
- ▶ Վստահության միջակայքի վերջնական արժեքները կոչվում են **ստորին և վերին վստահության սահմաններ** կամ **վստահության սահմաններ (CL)** և չեն մտնում միջակայքի մեջ

## Վստահության միջակայքի լայնություն/երկարություն

- Վստահության միջակայքի վերին և ստորին սահմանների միջև տարբերությունը կոչվում է լայնություն
- Միջակայքի լայնությունը թվային գծի ողջ երկարությունն է



Որտեղ,  $UU$ ՝ սահմանային սխալ,  $W$  (width) =  $2 \cdot UU = 2 \cdot t(z) \cdot SE$

# Միջին թվաբանականի վստահության միջակայքի հաշվարկ

- ❖ Բժշկական և կենսաբանական հետազոտություններում վստահության միջակայքի հաշվարկի վստահության հավանականության նվազագույն աստիճանը պետք է լինի 95%
- ❖ Վստահության միջակայքի հաշվարկի բանաձևն է՝
  - 95% ՎՄ  $= \bar{x} \pm z_{0.05/2} SE$
  - 99% ՎՄ  $= \bar{x} \pm z_{0.01/2} SE$ , որտեղ  
 $\bar{x}$ ՝ ընտրանքային միջինն է,  $SE$ ՝ ստանդարտ սխալը,  $z$ ՝ աղյուսակից վերցրած համապատասխան  $z$ -ի արժեքն է  
 Մեկնաբանում.
- Բոլոր հնարավոր ընտրանքային միջինների 95% (99%) հավանականությամբ կգտնվեն վերը նշված սահմաններում
- Մենք 95% (99%) հավանականությամբ վստահ ենք, որ իրական միջինը գտնվում է այդ միջակայքում

## տ քաշխում

տ քաշխում` այլընտրանքային քաշխում է, որն  
օգտագործվում է, երբ իրական պոպուլյացիոն  
ստանդարտ շեղումը ( $\sigma$ ) անհայտ է

# Վստահության միջակայքի որոշում ընտրանքային հետազոտության դեպքում

77

Վստահության միջակայքի որոշման բանաձևը այս դեպքում ընդունում է հետևյալ տեսքը՝

$$95\% CI = \bar{X} \pm t_{0.05/2} * SD / \sqrt{n}$$

Այսինքն, բանաձևում z-արժեքի փոխարեն օգտագործվում է t-արժեքը, պոպուլյացիոն  $\sigma$ -ի փոխարեն՝ ընտրանքային **SD**-ը, որը հանդիսանում է  $\sigma$ -ի լավագույն գնահատականը

### t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	<b>0.50</b>	<b>0.25</b>	<b>0.20</b>	<b>0.15</b>	<b>0.10</b>	<b>0.05</b>	<b>0.025</b>	<b>0.01</b>	<b>0.005</b>	<b>0.001</b>	<b>0.0005</b>
two-tails	<b>1.00</b>	<b>0.50</b>	<b>0.40</b>	<b>0.30</b>	<b>0.20</b>	<b>0.10</b>	<b>0.05</b>	<b>0.02</b>	<b>0.01</b>	<b>0.002</b>	<b>0.001</b>
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	<b>1.984</b>	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	<b>1.962</b>	2.330	2.581	3.098	3.300
<b>Z</b>	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	<b>Confidence Level</b>										

# տ քաշխման կիրառման օրինակ

79

- ▶ Ենթադրենք, արդեն նշված օրդինատորների պոպուլյացիայից ունենք ընտրանք  $n=5$  ծավալով, տարիքները 28, 35, 42, 43, 50

Այս դեպքում, օրդինատորների միջին տարիքը կկազմի  $(\bar{X}) = 39.6$ , իսկ  $SD = 8.4$

- ▶  $95\%$  ՎՄ  $= \bar{X} \pm t_{0.05/2} * SD / \sqrt{n}$

- ▶  $t$ -ի արժեքը որոշվում է  $t$ -արժեքների աղյուսակից, հաշվի առնելով վստահության միջակայքի որոշման հավանականության աստիճանը (95% կամ 99%) և ազատության աստիճանի արժեքը ( $df = n - 1 = 4$ ): Այս դեպքում այն կազմում է  $t = 2.776$

# t քաջիման կիրառման օրինակ

80

➔  $95\% \text{ ՎՄ} = X \pm t_{0.05/2} * SD / \sqrt{n} = 39.6 \pm 2.776 (8.4 / \sqrt{5}) =$   
 $= 39.6 \pm 10.41 = (29.19; 50.1)$

Վստահության ստորին սահմանն է 29.19, վերինը՝ 50.1

➔ 95%-ով վստահ ենք, որ իրական միջինը գտնվում է տվյալ միջակայքում՝ 29.19 վերև և 50.1 ցածր



# Ցուցանիշի վստահության միջակայքի հաշվարկ

81

- Հիմնված է ընտրովի հետազոտությունից ստացված ցուցանիշի արժեքի և նրա միջին սխալի վրա
- Հաշվարկի բանաձևն է՝

$$CI = \bar{p} \pm z * SE, \text{ որտեղ}$$

$\bar{p}$  ընտրովի հետազոտությունից ստացված ցուցանիշի արժեքն է,

$SE$ ՝ ցուցանիշի միջին սխալն է

$$SE(m) = \sqrt{\frac{\bar{p}(100 - \bar{p})}{n}}$$

# Օրինակ

82

- N-բնակավայրում հետազոտվել է 60 և ավելի բարձր տարիքի թվով 25 կին, որոնցից 8-ի մոտ հայտնաբերվել է հիպերտոնիա: Պահանջվում է պարզել հիպերտոնիայի տարածվածությունը նշված տարիքի կանանց մոտ և որոշել ստացված ցուցանիշի վստահության միջակայքը 95% հավանականությամբ
- 95% վՄ =  $p \pm z_{0.05/2} * SE$ , որտեղ  
 $p = \frac{8 * 100}{25} = 32/100$ ,  $SE = \sqrt{(32)(68) / 25}$ ,  
 $z = 1.96$   
 $95\% \text{ վՄ} = 32 \pm 1.96 * 9.3 = 32 \pm 18.2 = (13.3; 50.7)$
- Վստահության միջակայքի ստորին սահմանն է 13.3 և վերինը՝ 50.7
- 95%-ով վստահ ենք, որ իրական պոպուլյացիոն ցուցանիշը գտնվում է այդ միջակայքում

# Վարկածի ստուգում

83

- Չրոյական վարկած ( $H_0$ )՝ ենթադրություն անկախ և կախյալ փոփոխականների միջև կապի բացակայության վերաբերյալ
- Այլընտրանքային (հետազոտական) վարկած ( $H_A$ )՝ ենթադրություն անկախ և կախյալ փոփոխականների միջև կապի առկայության վերաբերյալ

# Ի՞նչ կարող է տեղի ունենալ վիճակագրական եզրակացություն կատարելիս

Իրական վիճակ

Որոշման ընդունում

$H_0$  ճիշտ է

$H_0$  սխալ է

$H_0$  ընդունվել է

Սխալ չկա

II տեսակի սխալ ( $\beta$ )

$H_0$  ժխտվել է

I տեսակի սխալ ( $\alpha$ )

Սխալ չկա

$H_0$ ընդունվել է	Սխալ չկա	II տեսակի սխալ ( $\beta$ )
$H_0$ ժխտվել է	I տեսակի սխալ ( $\alpha$ )	Սխալ չկա

# Միավորների տեսակները

85

- **I տեսակի սխալ ( $\alpha$ )**՝ հավանականություն է հերքել իրականում ճիշտ գրոյական վարկածը

Արտահայտվում է հավաստիության մակարդակով

Առավել հաճախակ ընդունված արժեքներն են՝ 0.05 կամ 0.01

- **II տեսակի սխալ ( $\beta$ )**՝ հավանականություն է ընդունել իրականում սխալ գրոյական վարկածը

Խնդիր է առաջացնում միայն այն դեպքում, եթե չի հաջողվում հերքել գրոյական վարկածը

Առավել հաճախակ ընդունված արժեքներն են՝ 0.1 կամ 0.2

# ρ- արժեքների ընդունված մեկնաբանությունները

**ρ-ի արժեքը**

**Մեկնաբանությունը**

$p > 0.05$

Վիճակագրորեն նշանակալի չէ

$0.01 < p < 0.05$

Վիճակագրորեն նշանակալի է

$p < 0.01$

Արդյունքը խիստ

վիճակագրորեն նշանակալի է

# Վարկածների ստուգում

87

- Վարկածների ստուգման երեք այլընտրանքային մեթոդներ
  - ▶ **t (z)** վիճակագրի հաշվարկ և համեմատություն կրիտիկական արժեքի հետ
  - ▶ **p**-արժեքի հաշվարկ
  - ▶ Արդյունքների տարբերության վստահության միջակայքի կառուցում
- Նշված երեք մեթոդների դեպքում ստացված արդյունքները չեն կարող հակասել միմյանց

# Վարկածների ստուգման ընդհանուր փուլերը

88

- Զրոյական և այլընտրանքային և վարկածների սահմանում

$$H_0: \mu_1 = \mu_2$$

$$H_A: \mu_1 \neq \mu_2$$

- Նշանակալիության մակարդակի ( $\alpha$ -արժեքի) որոշում
- Թեստի վիճականիի (t կամ z) հաշվարկ
- Կրիտիկական արժեքների, հերքման տիրույթի որոշում
- Եզրակացություն



# Վարկածների ստուգում

- ▶ Վարկածների ստուգումը երկու խմբերի միջինները համեմատելիս
  - ▶ t-թեստ
- ▶ Վարկածների ստուգումը երկու խմբերի ցուցանիշները համեմատելիս
  - ▶ z-թեստ

# Երկու միջինների համեմատումը անկախ ընտրանքների դեպքում

- ▶ Ընտրանքներ երկու նորմալ բաշխված անհայտ դիսպերսիաներով պոպուլյացիաներից՝ Տեստի վիճակագիրը՝

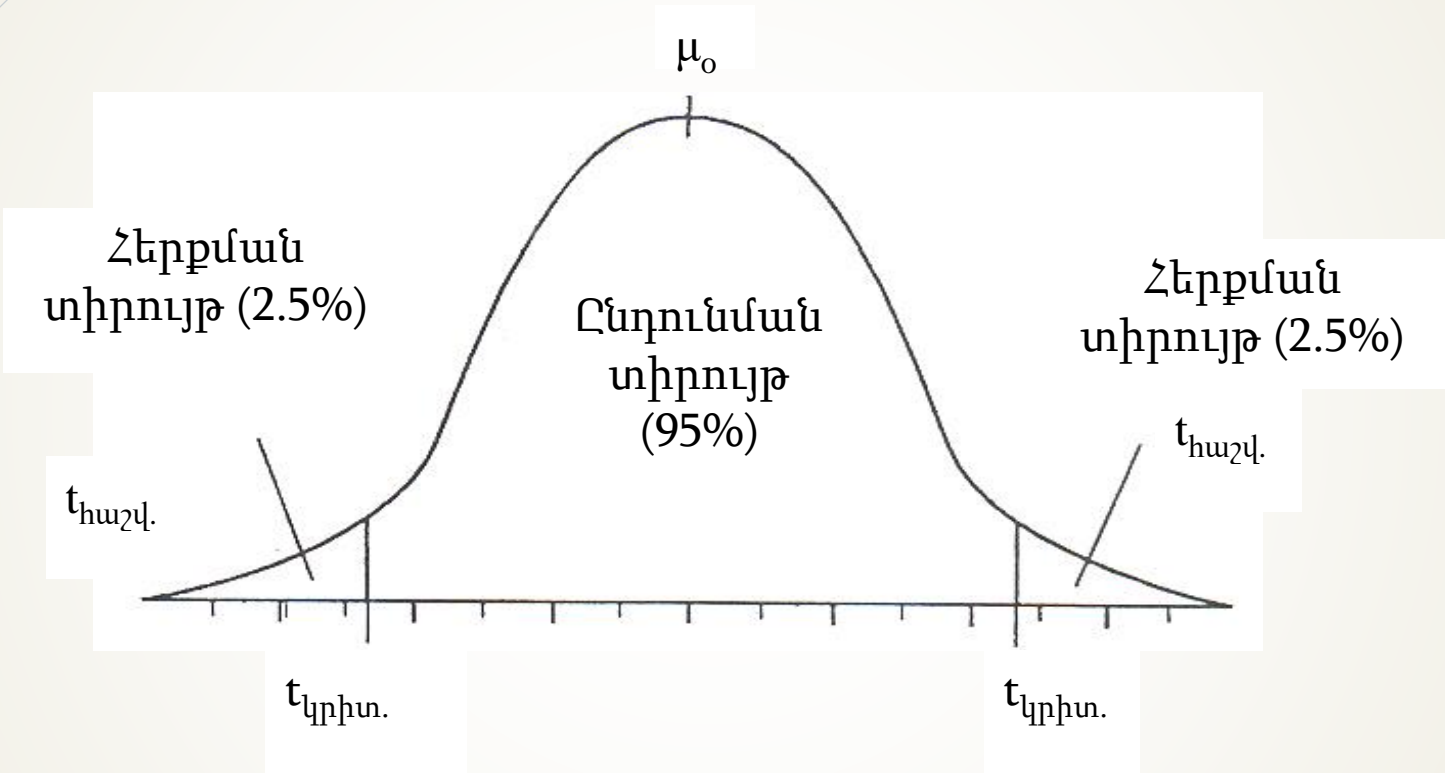
$$t = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{SE_1^2 + SE_2^2}}$$

որտեղ  $\bar{x}_1$  և  $\bar{x}_2$  երկու խմբերի միջիններն են  $SE_1$  և  $SE_2$  ստանդարտ շեղումները

## Վարկածները ստուգելու ժամանակ որոշման ընդունման կանոնները

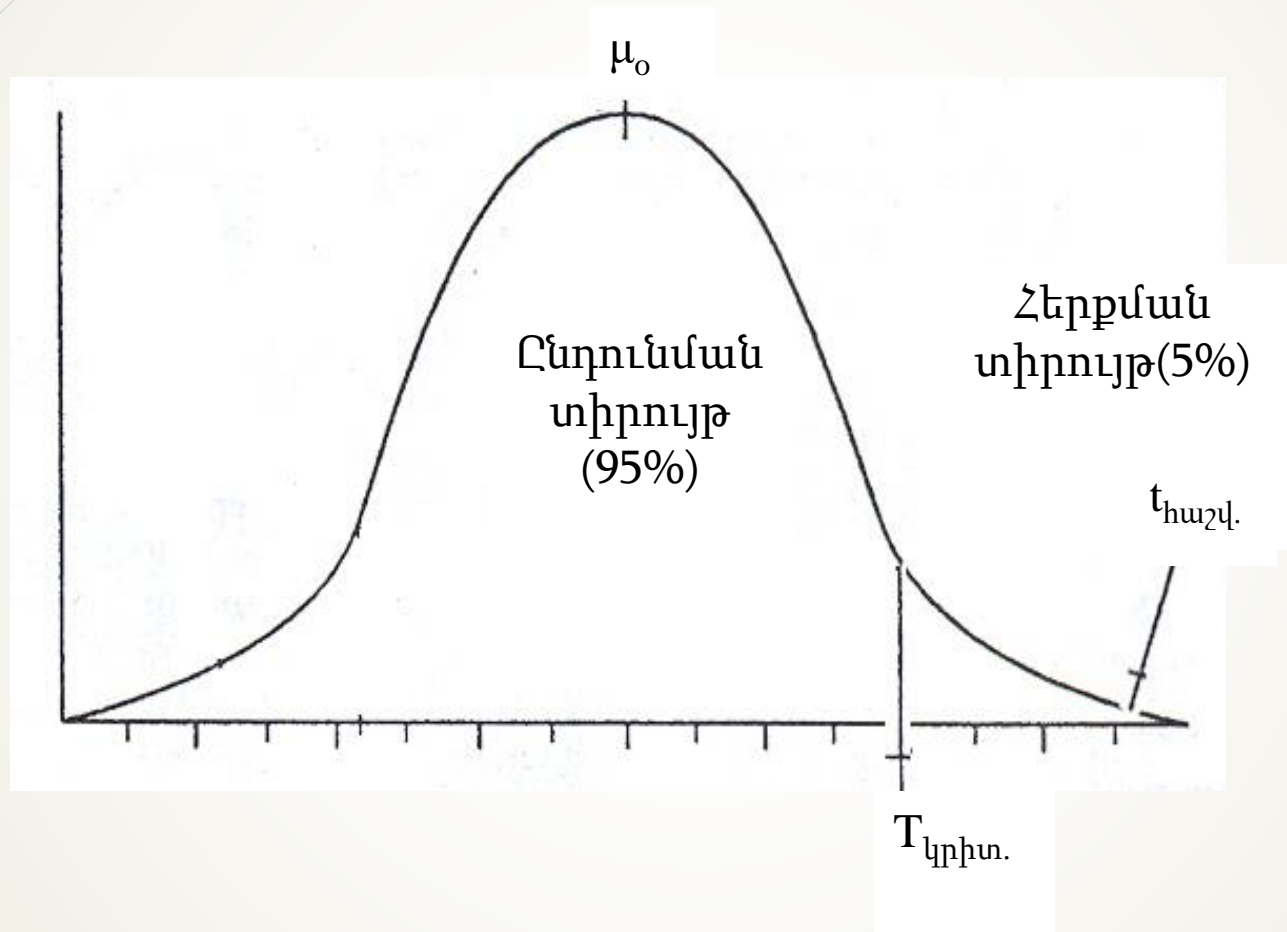
- ▶ Թեստի  $t$  հաշվարկայինը համեմատվում է աղյուսակային (կրիտիկական) արժեքի հետ
  - ▶ Կրիտիկական արժեքն ընտրելու համար որոշվում է նշանակալիության մակարդակը և հաշվարկվում ազատության աստիճանը  $df=(n_1+n_2)-2$
- ▶ Չրոյական վարկածը ( $H_0$ ) հերքվում է, եթե  $t$  ընկնում է հերքման տիրույթ
  - երկկողմանի թեստի դեպքում, եթե  $t > t_{\alpha/2}$
  - միակողմանի թեստի դեպքում
    - աջակողմյանի թեստ, եթե  $t > t_{\alpha}$
    - ձախակողմյանի թեստ, եթե  $t < -t_{\alpha}$
- ▶ Եթե  $t$  ընկնում են հերքման տիրույթ, ապա  $p < \alpha$

# Երկկողմանի t-թեստ



# Աջակողմյան t-թեստ

93



## Տարբերության վստահության միջակայքի կառուցումը (Confidence Interval, CI)

- ▶ Երբեմն ավելի տեղեկատվական է, քան  $p$ - արժեքը
- ▶ Հիմնված է բազմապատիկ ընտրանքների գաղափարի վրա
- ▶ Հիմնված է ընտրանքից ստացված տվյալների վրա
- ▶ Մշտապես երկկողմանի է
- ▶ Զրոյական վարկածը ընդունվում է, եթե 95% ՎՄ-ը ընդգրկում է “0” արժեքը

# Վստահության միջակայքի կառուցումը

95

- ▶ Նորմալ բաշխված պոպուլյացիաներ անհայտ դիսպերսիաներով.

$$CI = (\bar{X}_1 - \bar{X}_2) \pm t_{df} \sqrt{\frac{s^2_1}{n_1} + \frac{s^2_2}{n_2}}$$

# Օրինակ

96

24 հիպերտոնիկներ բաժանված են 2 խմբի՝ 12-ական մարդ, որոնք ստացել են տարբեր բուժում: Մեկ ամիս անց համեմատվել են նրանց դիաստոլիկ ճնշումները: Տվյալները ներկայացված են աղյուսակում:

	<b>n</b>	միջինը	ստանդարտ շեղումը
խումբ 1	12	117.0	21.7
խումբ 2	12	93.0	20.2



# Օրինակ /շարունակություն/

- ▶ Վարկածների ձևակերպում
- ▶ Նշանակալիության մակարդակի որոշում (սովորաբար,  $\alpha \leq 0.05$ )
- ▶  $t$  արժեքի հաշվարկ

$$\Rightarrow t = \frac{(117 - 93)}{\sqrt{\frac{470.89}{12} + \frac{408.04}{12}}} = 2.8$$

# t- արժեքների աղյուսակ

98

	<b>df</b>	<b>p = 0.10</b>	<b>0.05</b>	<b>0.01</b>	<b>0.001</b>
	1	6.314	12.706	63.657	636.62
	2	2.920	4.303	9.925	31.598
	3	2.353	3.182	5.841	12.924
	4	2.132	2.776	4.604	8.610
	5	2.015	2.571	4.032	6.869
	6	1.943	2.447	3.707	5.959
	7	1.895	2.365	3.499	5.408
	8	1.860	2.306	3.355	5.041
	9	1.833	2.262	3.250	4.781
	10	1.812	2.228	3.169	4.587
	11	1.796	2.201	3.106	4.437
	12	1.782	2.179	3.055	4.318
	13	1.771	2.160	3.012	4.221
	14	1.761	2.145	2.977	4.140
	15	1.753	2.131	2.947	4.073
	16	1.746	2.120	2.921	4.015
	17	1.740	2.110	2.898	3.965
	18	1.734	2.101	2.878	3.922
	19	1.729	2.093	2.861	3.883
	20	1.725	2.086	2.845	3.850
	21	1.721	2.080	2.831	3.819
	22	1.717	2.074	2.819	3.792
	23	1.714	2.069	2.807	3.767
	24	1.711	2.064	2.797	3.745
	25	1.708	2.060	2.787	3.725
	26	1.706	2.056	2.779	3.707
	27	1.703	2.052	2.771	3.690
	28	1.701	2.048	2.763	3.674
	29	1.699	2.045	2.756	3.659
	30	1.697	2.042	2.750	3.646
	40	1.684	2.021	2.704	3.551
	60	1.671	2.000	2.660	3.460
	120	1.658	1.980	2.617	3.373
(Normal distribution)	$\infty$	1.645	1.960	2.576	3.291

## Օրինակ /շարունակություն/

- ▶ Ազատության աստիճանի հաշվարկ՝  
 $df = n_1 + n_2 - 2$
- ▶ Հերքման տիրույթի որոշում
  - ▶  $H_0$  հերքվում է, քանի որ  $t = 2.8 > 2.07$ ,  
ընկնում է հերքման տիրույթ
- ▶ p-արժեքի որոշում  
 $p < 0.05$

# Օրինակ /շարունակություն/

100

## ▶ Վստահության միջակայքի կառուցում

- Չրոյական վարկածը ( $H_0$ ) հերքվում է, եթե ՎՄ-ը չի ընդգրկում “0” արժեքը

$$CI = (\bar{X}_1 - \bar{X}_2) \pm t_{df} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$CI = (117 - 93) \pm 2.07 \times 8.56 = 24 \pm 17.71 = (6.29; 41.51)$$

- ▶  $H_0$ -ն հերքվում է: 95 % հավանականությամբ կարելի է պնդել, որ իրական պոպուլյացիոն միջինների տարբերությունը առնվազն 6.29 մմ.ս.ս. է, առավելագույնը 41.51 մմ.ս.ս.

# Ստյուդենտի գույգային չափորոշիչ

101

$$S_d = \sqrt{\frac{\sum (d - \bar{d})^2}{n-1}}$$

$$S_{\bar{d}} = \frac{S_d}{\sqrt{n}}$$

$$t = \frac{\bar{d}}{S_{\bar{d}}}$$

- Հաշվարկել ցուցանիշի փոփոխությունը յուրաքանչյուրի մոտ
- Հաշվարկել այդ փոփոխությունների միջինը և ստանդարտ շեղումը
- Հաշվարկել Ստյուդենտի գործակցի մեծությունը
- Համեմատել այն կրիտիկական արժեքի հետ

Պահանջվում է փոփոխությունների բաշխման նորմալություն

## Երկու մասնաբաժինների (ցուցանիշների) տարբերության վարկածի ստուգում

- ▶ Զրոյական և այլընտրանքային վարկածների ձևակերպում

$$H_0: \pi_1 = \pi_2; H_A: \pi_1 \neq \pi_2,$$

- ▶ Թեստի վիճականու հաշվարկ

$$z = \frac{p_1 - p_2}{\sqrt{p * q * (\frac{1}{n_1} + \frac{1}{n_2})}}$$

որտեղ  $p_1$  և  $p_2$ ՝ ընտրանքների ցուցանիշներն են,  
 $p$ ՝ երկու խմբերի «միացյալ» մասն է ;  $q=1-p$

- ▶ Հերքման տիրույթի որոշում և եզրակացություն
- ▶ Պարտադիր պայման՝  $n * p > 5, n * (1-p) > 5$

# Օրինակ

103

Ավտովթարներից երեխաների մահացության հետազոտության ժամանակ ստացվել են հետևյալ տվյալները: Վթարի ենթարկված 123 երեխաներից, որոնք օգտվում էին ամրագոտիներից, 3-ը մահացել էին; 290-ից, որոնք չէին օգտվել ամրագոտիներից, մահացել էին 13-ը :

$$p_1 = 3/123=0.024, p_2 = 13/290=0.045$$

$$p = (3+13)/(123+290) = 0.039$$

# Օրինակ

104

- ▶  $H_0: \pi_1 = \pi_2$  ;  $H_A : \pi_1 \neq \pi_2$
- ▶ Թեստի վիճականու հաշվարկ

$$z = \frac{(0.024 - 0.045)}{\sqrt{(0.039)(1 - 0.039)\left(\frac{1}{123} + \frac{1}{290}\right)}} = -1.01$$

- ▶ Հերքման տիրույթի որոշում և եզրակացություն
- ▶  $p = 0.317$  - Ջրոյական վարկածը չի հերքվում, ընտրանքը չի ներկայացնում բավականաչափ վկայություններ, որ մահացությունը խմբերում տարբերվում է



## Երկու մասնաբաժինների (ցուցանիշների) տարբերության վստահության միջակայքի կառուցում

$$CI = p_1 - p_2 \pm z_{\alpha/2} \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

$$\begin{aligned} 95\%CI &= 0.024 - 0.045 \pm 1.96 \sqrt{\frac{0.024(1-0.024)}{123} + \frac{0.045(1-0.045)}{290}} = \\ &= (-0.057; 0.015) \end{aligned}$$

95% ՎՄ ընդգրկում է 0, ուստի զրոյական վարկածը չի հերքվում

# Երկուսից ավելի միջինների համեմատություն

106

t-տեստի բազմաթիվ օգտագործման թերություններ՝

- ▶ ժամանակատար է
- ▶ յուրաքանչյուր տեստի վիճակագրական զգայունությունը համեմատաբար ցածր է
- ▶ 1-ին տեսակի սխալի առաջացման հավանականությունը բարձր է

# Երկուսից ավելի միջինների համեմատություն: Դիսպերսիոն վերլուծություն (ANOVA):

- Զրոյական և այլընտրանքային վարկածների ձևակերպումը

$$H_0: \mu_1 = \mu_2 = \mu_3; H_A: \mu_1 \neq \mu_2 \neq \mu_3$$

- F -հարաբերակցության հաշվարկը

$$F = \frac{\text{միջխմբային դիսպերսիա}}{\text{ներխմբային դիսպերսիա}}$$

- Ըստ ընդունած  $\alpha$ -ի արժեքի F-ի կրիտիկական արժեքի որոշում և համեմատում հաշվարկված արժեքի հետ

- Զրոյական վարկածը  $H_0$  հերքվում է, եթե  $F_{\text{հաշվ.}} > F_{\text{կրիտ.}}$

- Հաջորդ քայլը.

- t-թեստ Բոնֆերոնիի ուղղումով
- t-թեստ ըստ Նյումեն-Քեյսլի (զույգ առ զույգ համեմատություններ)
- t-թեստ ըստ Թյուկիի (բոլոր զույգերի համեմատությունները)
- t-թեստ ըստ Դաննեթի (մեկ ստուգիչի հետ)

# Դիսպերսիոն վերլուծության տեսակները

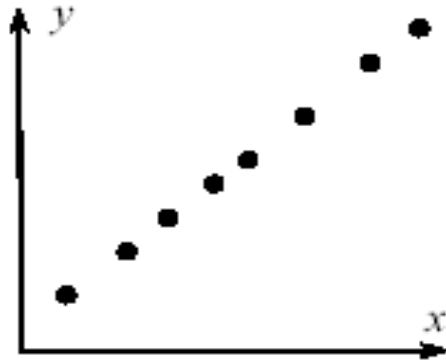
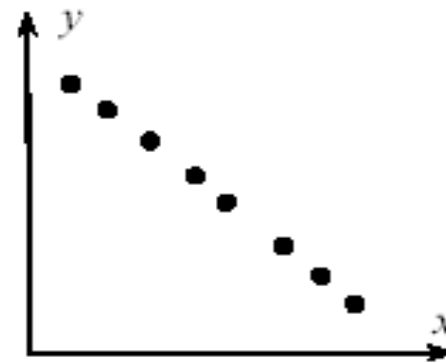
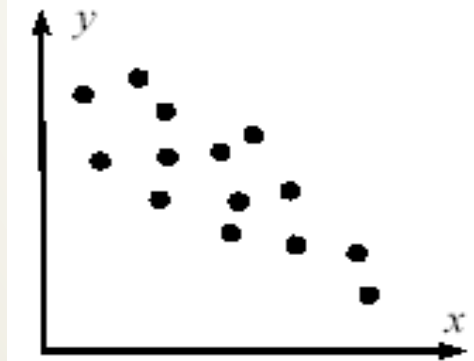
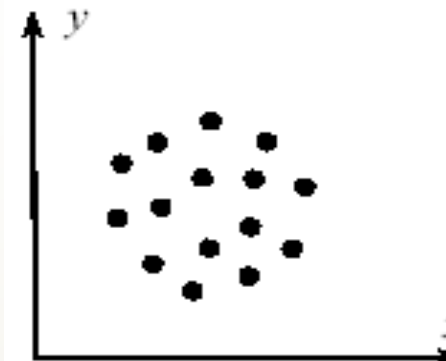
108

- ▶ Միաչափ դիսպերսիոն վերլուծություն (One-way ANOVA)- օգտագործվում, երբ համեմատվող խմբերը միմյանցից տարբերվում են միայն մեկ փոփոխականով
- ▶ Երկչափ դիսպերսիոն վերլուծություն (Two-way ANOVA) - եթե համեմատվող խմբերը տարբերվում են միաժամանակ երկու փոփոխականներով
- ▶ Բազմագործոնային դիսպերսիոն վերլուծություն (Multifactor ANOVA) - մի քանի փոփոխականներով տարբերվող խմբերի միջև տարբերությունների ուսումնասիրման անհրաժեշտության դեպքում

# Կորեյացիոն վերլուծության մեթոդներ

- Զույգ կորեյացիա
  - Պիրսոնի զույգ գծային
  - Սպիրմենի կարգային
- Մասնակի (պարզիալ)
- Բազմակի

# Կապի գրաֆիկական պատկերումը

A  $r=+1$ B  $r=-1$ C  $r=-0.7$ D  $r=0$ 

# Կորելյացիայի գործակցի արժեքի մեկնաբանությունը

- ▶ Կորելյացիայի գործակիցը ( $r$ ) բնութագրում է փոփոխականների միջև կապի ուժն ու բնույթը
- ▶ Կորելյացիայի գործակիցը կարող է կարող է ընդունել արժեքներ՝  $-1$  մինչև  $+1$
- ▶ Կորելյացիայի գործակցի նշանը ցույց է տալիս կապի ուղղությունը, իսկ թվային արժեքը՝ ուժգնությունը
- ▶ Կորելյացիոն վերլուծությունը սիմետրիկ է

# Պիրսոնի մեթոդ

- ▶ Քանակական փոփոխականների համար է
- ▶ Պարամետրիկ թեստ է, պահանջում է նորմալ բաշխում, գծայնություն

- ▶ Հաշվարկի բանաձևը՝

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}$$

- ▶ որտեղ  $(x - \bar{x})$  և  $(y - \bar{y})$  փոփոխականների շեղումներն են միջին թվաբանականներից



# Սպիրմենի (Կարգային) մեթոդ

113

- ▶ Կարգային սանդղակով չափվող փոփոխականների համար է
- ▶ Ոչ պարամետրիկ թեստ է, չի պահանջում է նորմալ բաշխում, գծայնություն
- ▶ Հաշվարկի բանաձևը՝

$$\rho = 1 - \frac{6 * \sum d^2}{n * (n^2 - 1)}$$

- ▶ Որտեղ  $d$  – կարգերի միջև տարբերությունն է

# Կորելյացիայի գործակցի նշանակալիություն գնահատում

- ▶ Բանաձևը կիրառելի է կորելյացիոն վերլուծության բոլոր եղանակների համար
- ▶  $n$  – զույգ տվյալների թիվն է
- ▶  $t$  – կրիտիկական արժեքների աղյուսակը նույն է

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

# Ռեգրեսիոն վերլուծության տեսակները

▶ Պարզ գծային ռեգրեսիոն վերլուծություն

$$\gamma = \alpha + \beta X$$

▶ Բազմակի ռեգրեսիոն վերլուծություն

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$$

▶ Լոգիստիկ ռեգրեսիոն վերլուծություն

$$\text{Prob(event)} = \frac{1}{1 + e^{-Z}}$$

$$Z = B_0 + B_1 * X_1 + B_2 * X_2 + \dots + B_n * X_n$$

## Պիրսոնի $\chi^2$ (համապատասխանության) գործակցի բնութագիրը

- Բոլոր ոչ պարամետրիկ տեստերից ամենահաճախ կիրառվողն է
- Կիրառվում է որակական հատկանիշների վերլուծության համար
- Հաշվարկում օգտագործում են բացարձակ թվեր
- Կարող է օգտագործվել երկու և ավելի խմբեր համեմատելու համար
- Հնարավորություն է տալիս ուսումնասիրել միայն փոփոխականների միջև եղած հավաստի կապի առկայությունը

# Ստուգում ըստ $\chi^2$ գործակցի՝ հիմնական փուլերը

- Զրոյական և այլընտրանքային վարկածների ձևակերպում
- Նշանակալիության մակարդակի որոշում
- Սպասվող հաճախականությունների հաշվարկ
- $\chi^2$  գործակցի հաշվարկ

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

որտեղ O- փաստացի տվյալներն են, որոնք ստացվել են հետազոտությունից, E-  $H_0$  վարկածի հիման վրա հաշված սպասվելիք տվյալներն են

- Ազատության աստիճանի որոշում՝  $df = (C-1) * (R-1)$
- Ստացված  $\chi^2$  գործակցի համեմատումը կրիտիկական արժեքի հետ  
Զրոյական վարկածը հերքվում է, եթե  
 $\chi^2_{\text{հաշվ.}} > \chi^2_{\text{կրիտ.}}$

# Մտուցում ըստ $\chi^2$ գործակցի՝ օրինակ

118

Բուժման մեթոդները	Հիվանդների ընդհանուր թիվը	Ցավի սինդրոմ /Փաստացի տվյալներ/			Ցավի սինդրոմ /սպասվող տվյալներ ըստ $\chi^2$ հիպոտեզի/		
		անհետ.	պակաս.	անփոփ.	անհետ.	պակաս.	անփոփ.
Բուժական համալիր միջոցները+ հեպարին	40	14	21	5	9.8	15.6	14.6
Համալիր բուժական միջոցներ	42	6	11	25	10.2	16.4	15.4
Ընդամենը	82 (100%)	20 (24.4%)	32 (39.0%)	30 (36.6%)			

# Մտուցում ըստ $\chi^2$ գործակցի՝ օրինակ

119

$$\begin{aligned}\chi^2 &= \sum \frac{(O-E)^2}{E} = \frac{(14-9.8)^2}{9.8} + \frac{(21-15.6)^2}{15.6} + \frac{(5-14.6)^2}{14.6} + \frac{(6-10.2)^2}{10.2} + \\ &+ \frac{(11-16.4)^2}{16.4} + \frac{(25-15.4)^2}{15.4} = 1.8 + 1.9 + 6.3 + 1.7 + 1.8 + 5.9 = 19.5\end{aligned}$$

➤  $df = (C-1)*(R-1) = (3-1)*(2-1) = 2$

➤  $\alpha = 0,05$ -ին համապատասխանող

$\chi^2$  կրիտիկական արժեքը = 5,99:

$\chi^2$  հաշվարկայինը = 19,5:

Քանի որ  $\chi^2_{\text{հաշվ.}} > \chi^2_{\text{կրիտ.}}$ , գրոյական վարկածը ժխտվում է և ընդունվում այլընտրանքայինը:

# Յետսի ուղղումն անընդհատության համար

- ▶ 2x2 աղյուսակներում կիրառվում է Յետսի ուղղումը՝ յուրաքանչյուր (O-E) տարբերությունից հանվում է 0.5

$$\chi^2 = \sum \frac{(|O - E| - 0.5)^2}{E}$$



# Ֆիշերի ճշգրիտ հավանականություն

- Հաշվարկվում է  $2 \times 2$  աղյուսակների համար երբ սպասվող թվերը որևէ մեկ վանդակում փոքր է 5:

# Ինչպես ընտրել վիճակագրական մեթոդը

Հատկանիշ	2 խումբ	2-ից ավելի	Մի խումբ “մինչև-հետո	Մի խումբ տարբեր մեթոդներ	Հատկանիշների միջև կապի գնահատում
Քանակական (նորմալ բաշխում)	Ստյուդենտի անկախ խմբերի	Դիսպերսիոն	Չոլգային t-տեստ	Դիսպերսիոն կրկնակի չափումների համար	Գծային ռեգրեսիա, կորելյացիա
Որակական	$\chi^2$	$\chi^2$	Մակ-Նիմարի	Կոկրեյն	Չոլգորդման
Կարգային	Մանս-Ուիլսոնի	Կռուսկալ-Ուոլլիս	Ուիլքոքսոն	Ֆրիդման	Սպիրմենի կարգային կորելյացիա