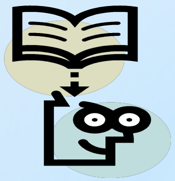
Анализ вариационных рядов распределения

(показатели дифференциации и концентрации)



Вопросы:

1. **Показатели дифференциации:**
   - **Децильный коэффициент**
   - **Фондовый коэффициент**
2. **Показатели концентрации:**
   - **Коэффициент Герфиндаля**
   - **Коэффициент Джини**
   - **Коэффициент Лоренца**

При необходимости более подробного изучения структуры вариационного ряда, рассчитываются показатели, аналогичные медиане.
Для этого вариационный ряд делится на большее число групп, равных по числу единиц в каждой группе.

Квартильные группы –

4 группы по 25% единиц в каждой группе

Квинтильные группы –

5 групп по 20 % единиц в каждой группе

Децильные группы –

10 групп по 10% единиц в каждой группе

Децили **– это максимальные значения признака в соответствующих децильных группах:**

$Д_1$ **– первый дециль (значение признака у 10%-ой единицы)**

$Д_2$ **– второй дециль (значение признака у 20%-ой единицы)**

…

…

…

$Д_9$ **– девятый дециль (значение признака у 90%-ой единицы)**

$Д_{10}$ **– десятый дециль (значение признака у 100%-ой единицы), максимальное значение признака в данном вариационном ряду.**

Децильный коэффициент дифференциации (Кд)

$$К_{Д} = Д_{9} / Д_{1}$$

1. По накопленной частоте или накопленной частости определяются интервалы, в которые попадают 10%-я и 90%-я единицы ранжированного ряда

среднедушевой доход в месяц, руб.	длина интервала	частота, млн. чел.	частость, %	накопленная	
				частота	частость
		m_i	w_i	F_i $F_i=m_1+m_2+\ldots+m_i$	p_i $p_i=w_1+w_2+\ldots+w_i$
0 - 3500	3500	3,2	2,2	3,2	2,2
3500 - 5000	1500	5,3	3,7	8,5	5,9
5000 - 7000	2000	10,0	7,0	18,5	12,9
7000 - 10000	3000	17,3	12,1	35,8	25,0
10000 - 15000	5000	26,9	18,8	62,7	43,8
15000 - 25000	10000	36,2	25,3	98,9	69,1
25000 - 35000	10000	19,1	13,3	118,0	82,4
35000 - 45000	10000	25,2	17,6	143,2	100,0

2. Рассчитывается $Д_9$ – девятый дециль (значение признака у 90%-й единицы)

$$Д_9 = x_{k-1} + \Delta x_k \frac{9/10 * \sum m_i - F_{k-1}}{m_k} =$$

$$= x_{k-1} + \Delta x_k \frac{9/10 * \sum w_i - p_{k-1}}{w_k} =$$

$$= 35000 + 10000 \frac{90 - 82{,}4}{17{,}6} = 39317{,}46\ руб.$$

где x_{k-1} – нижняя граница интервала, содержащего девятый дециль;

Δx_k – длина интервала, содержащего девятый дециль;

F_{k-1} (p_{k-1}) – частота (частость), накопленная до интервала, содержащего девятый дециль;

m_k (w_k) – частота (частость) интервала, содержащего девятый дециль.

3. Рассчитывается $Д_1$ – первый дециль (значение признака у 10%-й единицы)

$$Д_1 = x_{k-1} + \Delta x_k \frac{1/10 * \sum m_i - F_{k-1}}{m_k} =$$

$$= x_{k-1} + \Delta x_k \frac{1/10 * \sum w_i - p_{k-1}}{w_k} =$$

$$= 5000 + 2000 \frac{10 - 5,9}{7,0} = 6164,00\, руб.$$

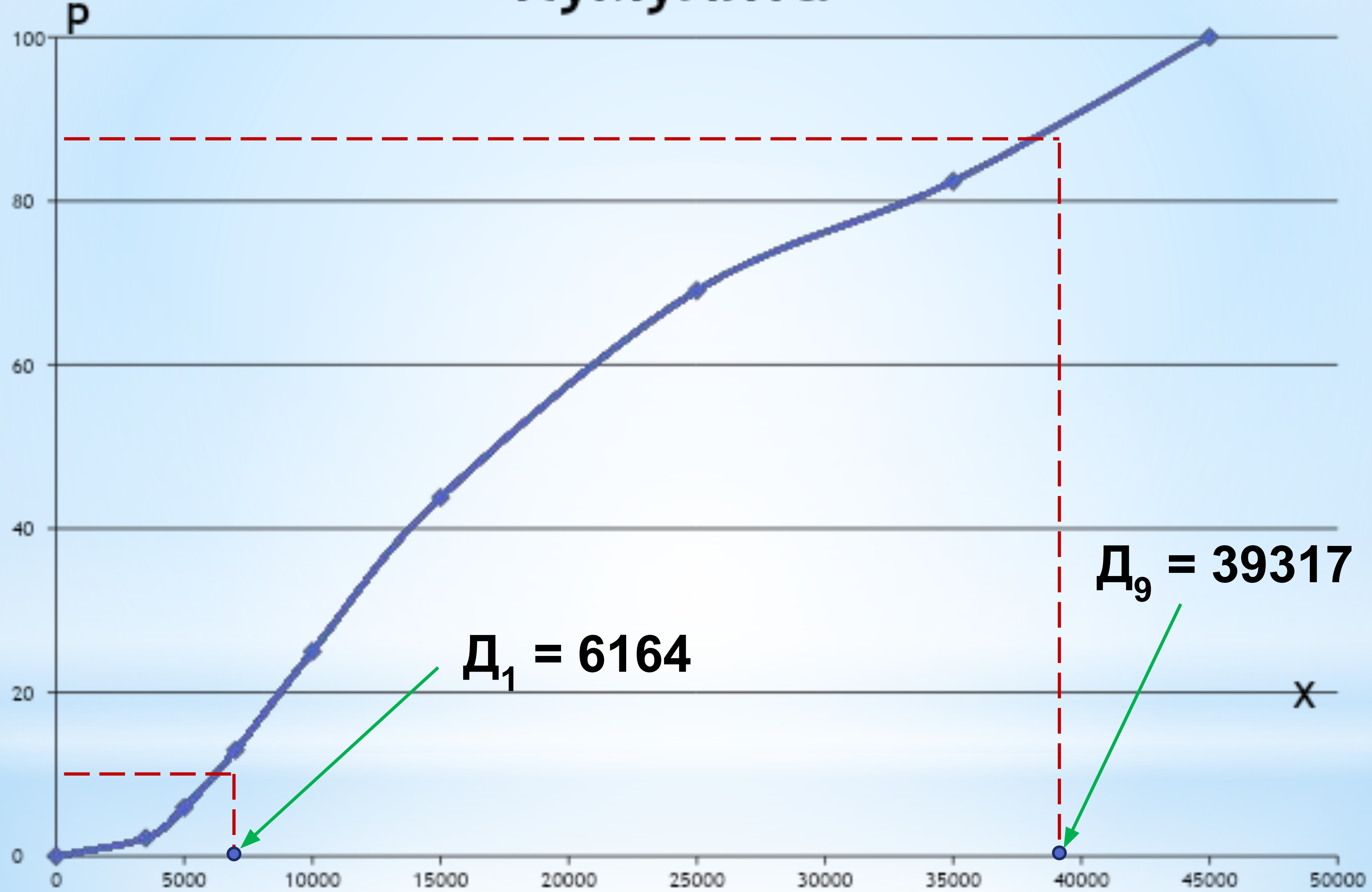
где x_{k-1} – нижняя граница интервала, содержащего первый дециль;

Δx_k – длина интервала, содержащего первый дециль;

F_{k-1} (p_{k-1}) – частота (частость), накопленная до интервала, содержащего первый дециль;

m_k (w_k) – частота (частость) интервала, содержащего первый дециль.

Кумулята



P
100
80
60
40
20
0
X
0
5000
10000
15000
20000
25000
30000
35000
40000
45000
50000
Д₁ = 6164
Д₉ = 39317

$$K_Д = Д_9 / Д_1 =$$

$$= 39317{,}46 / 6164{,}00 = 6{,}4\,раза$$

Показывает, во сколько раз значение признака 90%-й единицы (девятый дециль - *$Д_9$*) больше значения признака 10%-й единицы (первого дециля *$Д_1$*)

Фондовый коэффициент дифференциации:

$$K_Ф = \frac{\overline{x}_{10}}{\overline{x}_1}$$

$\overline{x}_{10}$ и $\overline{x}_1$ – среднедушевой доход населения в 10-й и 1-й децильных группах, соответственно

среднедушевой доход в месяц, руб.	середина интервала	частость, %	
	x_i	w_i	$x_i w_i$
0 - 3500	1750	2,2	3911
3500 - 5000	4250	3,7	15730
5000 - 6164	5582	4,1	22886

$$\overline{x}_1 = \frac{\sum x_i w_i}{\sum w_i} = \frac{42527}{10} = 4252{,}7\ руб.$$

среднедушевой доход в месяц, руб.	середина интервала	частость, %	
	x_i	w_i	$x_i w_i$
39317 - 45000	42158,5	10,0	421585

$$\bar{x}_{10} = \frac{\sum x_i w_i}{\sum w_i} = \frac{421585}{10} = 42158{,}5\,руб.$$

$$K_{Ф} = \frac{\bar{x}_{10}}{\bar{x}_1} = \frac{42158{,}5}{4252{,}7} = 9{,}9$$

Фондовый коэффициент показывает во сколько раз среднее значение признака в десятой децильной группе больше среднего значения признака в первой децильной группе

ПОКАЗАТЕЛИ КОНЦЕНТРАЦИИ

Коэффициент Герфиндаля:

$$H = \sum_{i=1}^{n} {d_i}^2$$

Определяется как сумма квадратов долей изучаемого признака выделенных групп в общей сумме признака

x_i	m_i	$x_i m_i$	w_i	$x_i w_i$		
1750	3,2	5600	2,2	3911	0,002	0,000
4250	5,3	22525	3,7	15730	0,008	0,000
6000	10,0	60000	7,0	41899	0,021	0,000
8500	17,3	147050	12,1	102689	0,051	0,003
12500	26,9	336250	18,8	234811	0,117	0,014
20000	36,2	724000	25,3	505587	0,252	0,063
30000	19,1	573000	13,3	400140	0,199	0,040
40000	25,2	1008000	17,6	703911	0,350	0,123
Σ	143,2	2876425	100,0	2008677	1,000	0,243

$$d_i = \frac{x_i m_i}{\sum x_i m_i}$$

$$d_i = \frac{x_i w_i}{\sum x_i w_i}$$

$$H = \sum(d_i^2) = \sum\left(\frac{x_i m_i}{\sum x_i m_i}\right)^2 = \sum\left(\frac{x_i w_i}{\sum x_i w_i}\right)^2 \approx 0,243$$

Рассмотрим случай равномерного распределения признака по группам ($d_i = Const$):

i		
1		
2	…	…
3	…	…
…	…	…
…	…	…
…	…	…
…	…	…
n		
Σ	1,000	

$$d_i = \frac{\sum d_i}{n} = \frac{1}{n}$$

$$\sum(d_i^2) = \frac{1}{n^2} + \cdots + \frac{1}{n^2} =$$
$$= n \times \frac{1}{n^2} = \frac{1}{n} = H_{min}$$

$$H_{min} = \frac{1}{8} = 0,125$$

Рассмотрим случай крайне неравномерного распределения признака по группам:

i		
1		
2		
3		
…	…	…
…	…	…
…	…	…
n-1		
n		

$$H_{max} \to 1$$

1	0,002
2	0,008
3	0,021
4	0,051
5	0,117
6	0,252
7	0,199
8	0,350

Пределы изменения коэффициента Герфиндаля

$$H_{min} = 0,125$$

$$H_{max} \to 1$$

$$\frac{1}{n} \leq H < 1$$

$$H = 0,243$$

Отличие фактического значения коэффициента Герфиндаля (Н) от возможного в данном ряду минимального значения ($Н_{min}$) указывает на наличие *доминирующих групп* в этом ряду.

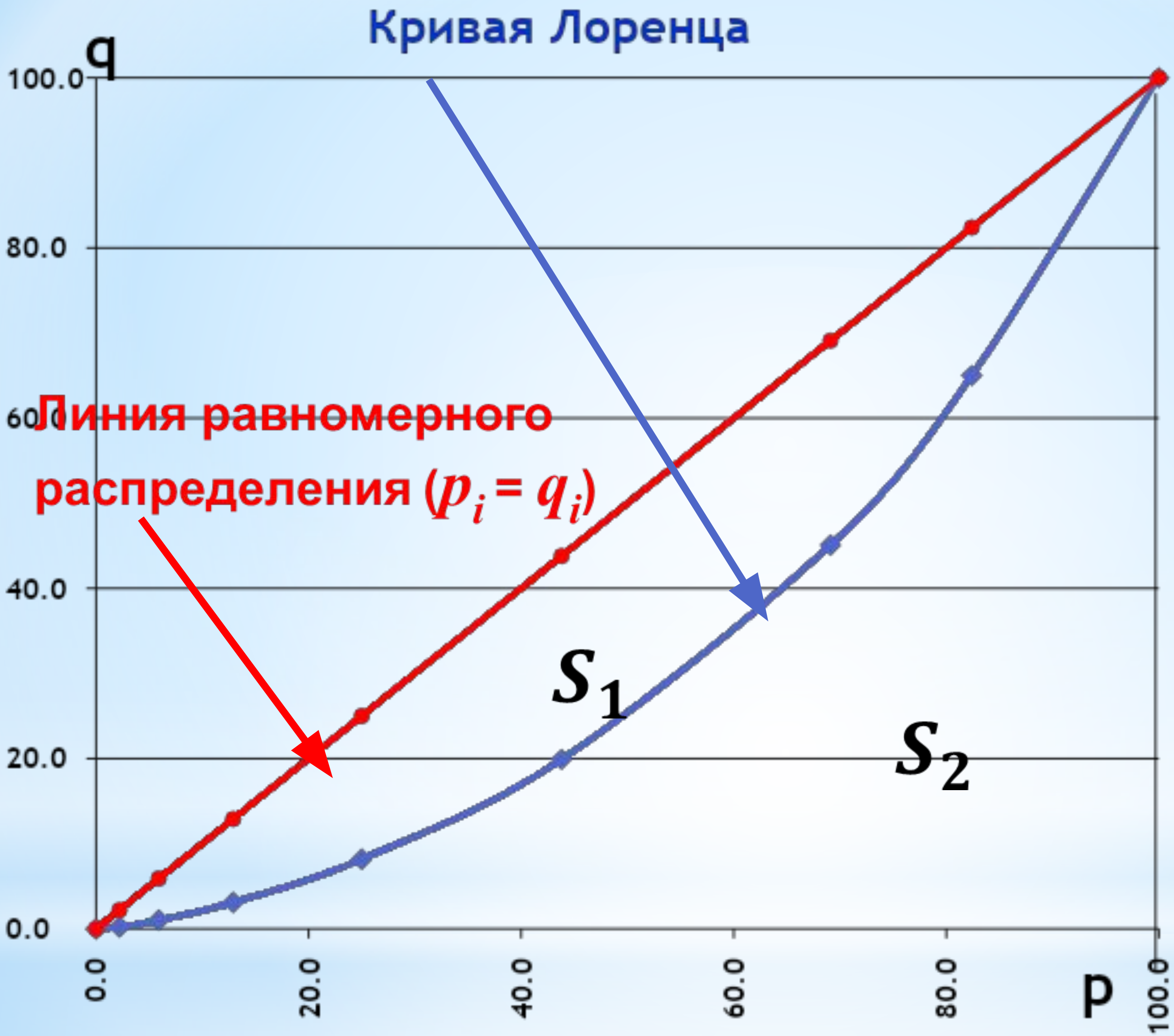
Близость коэффициента Герфиндаля (Н) к 1 характеризует неравномерность распределения

Коэффициент Джини:

p_i – накопленная частость первых i групп населения (в%)

q_i – накопленная доля признака первых i групп населения

w_i	p_i		q_i
2,2	2,2	0,002	0,002
3,7	5,9	0,008	0,010
7,0	12,9	0,021	0,031
12,1	25,0	0,051	0,082
18,8	43,8	0,117	0,199
25,3	69,1	0,252	0,451
13,3	82,4	0,199	0,650
17,6	100,0	0,350	1,000
100,0		1,000	



p_i	q_i
0,0	0,0
2,2	0,2
5,9	1,0
12,9	3,1
25,0	8,2
43,8	19,9
69,1	45,1
82,4	65,0
100,0	100,0

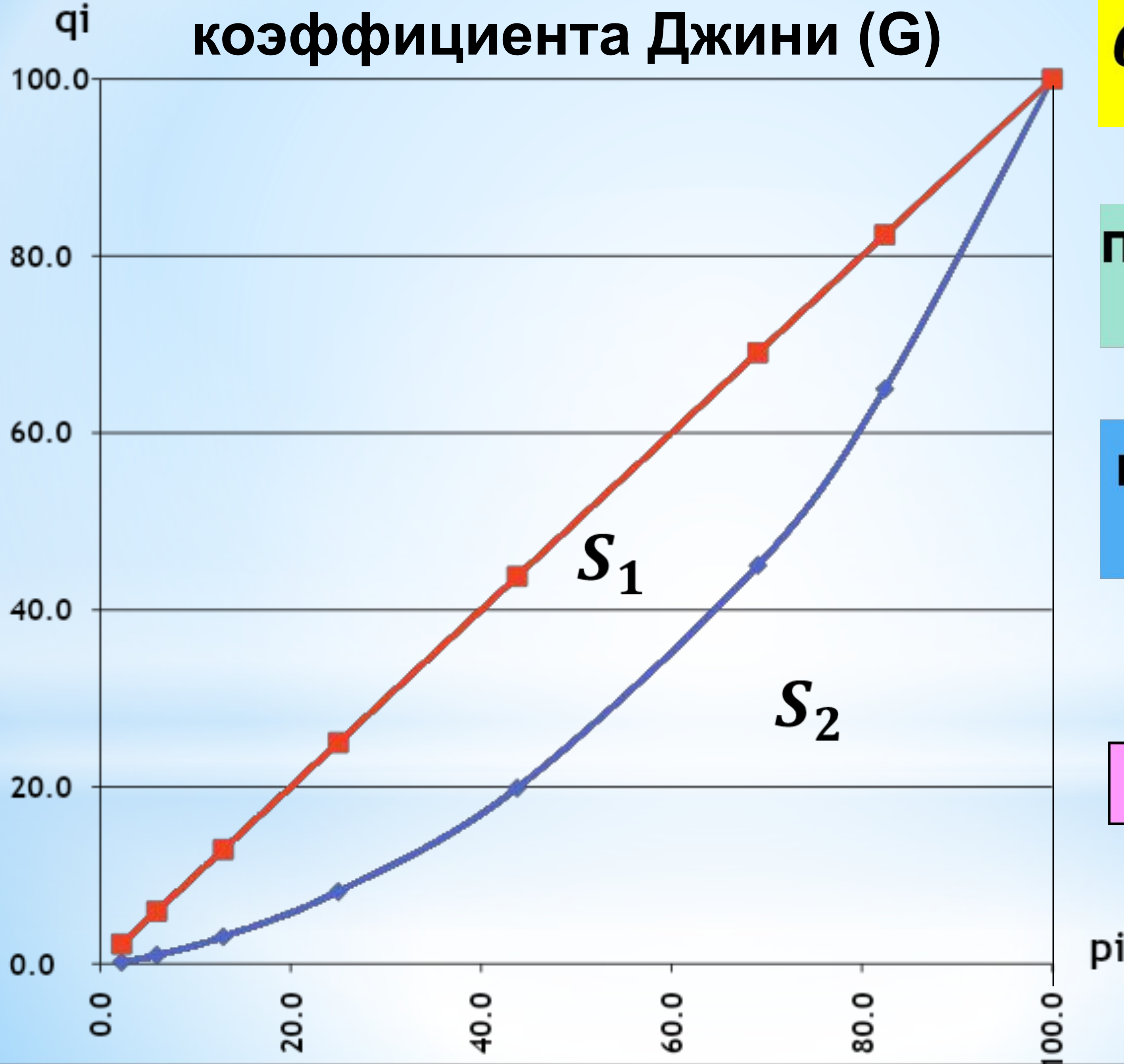
$$G = \frac{S_1}{S_1 + S_2}$$

$$G = \sum p_i q_{i+1} - \sum p_{i+1} q_i$$

p_i	q_i	$p_i q_{i+1}$	$p_{i+1} q_i$
0,022	0,002	0,000	0,000
0,059	0,010	0,002	0,001
0,129	0,031	0,011	0,008
0,250	0,082	0,050	0,036
0,438	0,199	0,197	0,137
0,691	0,451	0,449	0,371
0,824	0,650	0,824	0,650
1,000	1,000	-	-
		1,533	1,203

$$G = \sum p_i q_{i+1} - \sum p_{i+1} q_i = 1{,}533 - 1{,}203 \approx 0{,}330$$

Пределы изменения коэффициента Джини (G)



$$G = \frac{S_1}{S_1 + S_2}$$

при $S_1 \to 0$
G → 0

при $S_2 \to 0$
G → 1

$$0 \leq G < 1$$

$$G = \sum p_i q_{i+1} - \sum p_{i+1} q_i = 0{,}330$$

p \ q	Доли единицы	Проценты
Доли единицы	G	:100*
Проценты	:100*	:10000

***Допускается расчет коэффициента Джини в процентах**

Величина коэффициента Джини (G) характеризует степень концентрации изучаемого признака у единиц наблюдения в выделенных группах.

$\mathbf{G \leq 0,3}$ **– невысокая**

$\mathbf{0,3 < G \leq 0,5}$ **– средняя**

$\mathbf{0,5 < G \leq 0,8}$ **– высокая**

$\mathbf{G > 0,8}$ **– крайне высокая**

Коэффициент Лоренца

$$L = \frac{1}{2} \times \sum |w_i - d_i|$$

$$0 \leq L < 1$$

Допускается расчет коэффициента Лоренца в %

СПАСИБО ЗА ВНИМАНИЕ!