# Анализ структуры совокупности

Contents

[Анализ структуры совокупности 1](#_Toc172908799)

[1) Доля. Структура совокупности 1](#_Toc172908800)

[2](#_Toc172908801)

[2) Квантили 3](#_Toc172908802)

[а) Медиана 4](#_Toc172908803)

[б) Квартили 6](#_Toc172908804)

[в) Децили 8](#_Toc172908805)

[г) Перцентили 10](#_Toc172908806)

[3) Коэффициенты, характеризующие концентрацию изучаемого признака в определенных группах 12](#_Toc172908807)

[4) Коэффициенты, характеризующие дифференциацию изучаемого признака в совокупности 13](#_Toc172908808)

[Децильный коэффициент дифференциации 13](#_Toc172908809)

[Квартильный коэффициент дифференциации 14](#_Toc172908810)

[Коэффициент фондов 14](#_Toc172908811)

[Кривая Лоренца и коэффициент Джинни 15](#_Toc172908812)

[5) Показатели, характеризующие изменение структуры во времени 18](#_Toc172908813)

## Доля. Структура совокупности

Структура совокупности – это строение изучаемой совокупности, состоящей из отдельных групп элементов.

В первую очередь структура характеризуется с помощью показателей удельного веса (показателей доли) отдельных элементов в общей численности совокупности.

Где - доля -ой группы в общей численности совокупности,

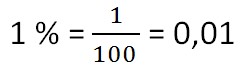
– число элементов в -ой группе,

– число элементов во всей совокупности.

Доли могут быть выражены в процентах и в десятичных дробях.

Обязательно должно соблюдаться правило:

Процент - это одна сотая часть от числа. Процент записывается с помощью знака %.

;

Чтобы перевести проценты в дробь, нужно убрать знак % и разделить числопроцентов на 100.

Чтобы перевести десятичную дробь в проценты, нужно дробь умножить на 100 и добавить знак %.

|  |  |
| --- | --- |
| *перевод процента в десятичную дробь* | *перевод дроби в проценты* |

**Структура совокупности – это таблица, в которой обязательно присутствуют два элемента: названия выделенных групп и доли этих групп в общей численности совокупности.**

Структура совокупности графически изображается обычно в виде круговой (секторной) диаграммы. Если требуется сравнить две структуры, то используется столбиковая диаграмма или столбиковая диаграмма с накоплением.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| бригада 1 | | Доля численности каждой группы рабочих в общей численности рабочих цеха, dj | | | |
| Заработная плата одного рабочего, тыс.руб;xi | число рабочих,fi, | в целых числах | | в процентах | |
| 15 | 20 | 20/75 = | 0,2667 | 0,2667 \* 100% = | 26,67% |
| 18 | 37 | 37/75 = | 0,4933 | 0,4933 \* 100% = | 49,33% |
| 20 | 14 | 14/75 = | 0,1867 | 0,1867 \* 100% = | 18,67% |
| 25 | 4 | 4/75 = | 0,0533 | 0,0533 \* 100% = | 5,33% |
| Итого: | 75 |  | 1,00 | 100,00% | |

Рабочих цеха по уровню заработной платы выглядит так.

А) В виде таблицы

Структура рабочих цеха по уровню заработной платы

|  |  |
| --- | --- |
| Заработная плата рабочего, тыс.руб; xi | Доля численности каждой группы рабочих в обшей численностиdj, в % |
| 15 | 26,67% |
| 18 | 49,33% |
| 20 | 18,67% |
| 25 | 5,33% |
| Итого: | 100,00% |

Б) Графическое изображение структуры в виде круговой диаграммы.

## 

В зависимости от цели анализа различают долю численности и долю изучаемого признака.

Например, когда требуется определить характер распределения единиц наблюдения по группам, например, какая часть людей находится в каждой выделенной группе по уровню зарплаты, то рассчитывают долю частоты (численности) группы - . Тогда говорят о структуре рабочих (какая часть рабочих в какой группе).

Если нужно оценить, как распределяется общая сумма изучаемого признака по группам, например, какую часть фонда оплаты труда получает каждая выделенная группа рабочих, то рассчитывают долю значения признака группы - . В этом случае говорят о структуре фонда оплаты труда (ФОТ) (суммы заработной платы всех рабочих), как ФОТ распределяется между группами рабочих.

Например,

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Заработная плата одного рабочего, тыс.руб. (xi) | для структуры рабочих цеха | | | для структуры фонда оплаты труда | | | |
| fi, число рабочих | dfj, доля  частоты | | сумма зарплаты  по группе | | dxj, доля  признака | |
| 15 | 20 | 20:75 = | 0,2667 | 15×20 = | 300 | 300:1346 = | 0,2229 |
| 18 | 37 | 37:75 = | 0,4933 | 18×37 = | 666 | 666:1346 = | 0,4948 |
| 20 | 14 | 14:75 = | 0,1867 | 20×14 = | 280 | 280:1346 = | 0,208 |
| 25 | 4 | 4:75 = | 0,0533 | 25×4 = | 100 | 100:1346 = | 0,0743 |
| Итого: | 75 |  | 1 |  | 1346 |  | 1 |

Тогда

Структура рабочих цеха по уровню заработной платы

|  |  |
| --- | --- |
| Заработная плата рабочего, тыс.руб; xi | Доля численности каждой группы рабочих в обшей численности dxj, в % |
| 15 | 26,67% |
| 18 | 49,33% |
| 20 | 18,67% |
| 25 | 5,33% |
| Итого: | 100,00% |

Структура фонда заработной платы по уровню заработной платы

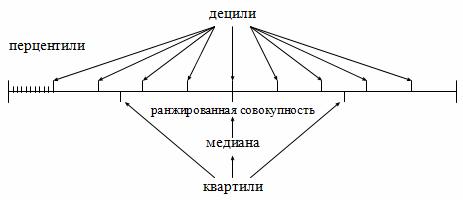
|  |  |
| --- | --- |
| Заработная плата рабочего, тыс.руб; xi | Доля суммы заработной платы каждой группы рабочих в обшей ФОТ, dxj, в % |
| 15 | 22,29% |
| 18 | 49,48% |
| 20 | 20,80% |
| 25 | 7,43% |
| Итого: | 100,00% |

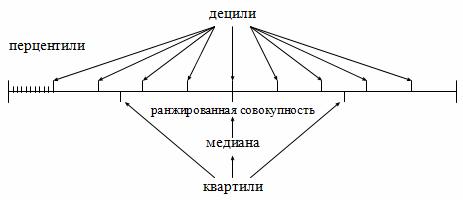
## Квантили

Также важной характеристикой структуры совокупности являются квантили.

Квантили - величины, разделяющие совокупность на определенное количество равных по численности элементов частей.

Наиболее известные – медиана, квартили, децили, перцентили.





Квантили (квартили, децили и перцентили чаще всего применяются для интервальных рядов, т.к. в с их помощью удобно анализировать большой объем данных.

#### Общие формулы для квантилей

1. Рассчитать номер квантиля
2. Если ряд дискретный, то рассчитать значение квантиля по формуле

+ (

1. Если ряд интервальный, то рассчитать значение квантиля по формуле

Где – квантиль, – номер квантиля, j – порядковый номер квантиля, - сумма всех частот (количество элементов в совокупности), – размерность квантиля (на сколько частей эти квантили делят совокупность), - нижняя границаквантильного интервала, – ширина квантильного интервала, - накопленная частота предквантильного интервала, - частота квантильного интервала.

### Медиана

Самый известный квантиль – **медиана,** делящая совокупность на две равные части.

#### Медиана для дискретного ряда.

Для определения медианы в дискретном ряду  сначала ***порядковый номер медианы* по формуле:**, а затем определяют, какое значение признака обладает накопленной частотой, равной номеру медианы.

Если ряд содержит **четное**число элементов, то номер медианы будет нецелым числом и **медиана будет равна средней из двух значений признака, находящихся в середине.** Номер первого из этих признаков – целая часть номера медианы, для второго -  номер медианы, округленный до целого числа.

*Пример 1. Найти медиану*

| Заработная плата  рабочего, тыс.руб; | бригада 1 | | бригада 2 | |
| --- | --- | --- | --- | --- |
| ,  число рабочих | ,  накопленная  частота | ,  число рабочих | ,  накопленная частота |
| 15 | 20 | 20 | 20 | 20 |
| 18 | 35 | 55 | 34 | 54 |
| 20 | 14 | 69 | 14 | 68 |
| 25 | 6 | 75 | 6 | 74 |
| Итого: | 75 |  | 74 |  |

1. Определяем номер медианы

для первой бригады ;для второй бригады ;

2. Для первой бригады номер медианы – целое число. Следовательно, нужно найти элемент совокупности, для которого накопленная частота S равна номеру медианы. Для этогоопределяем самую первую накопленную частоту, которая больше или равна номеру медианы. Это накопленная частота второго значения признака Хi=18, Si=55.

Определяем значение медианы. Для первой бригады Ме = х38= 18 тыс.руб.,

3. Номер медианы для второй бригады - нецелое число. Для определения медианы нужны значения двух элементов – х37 и х38. Определяем их значение с помощью самой первой накопленной частоты большей или равной порядковым номерам элементов (37 и 38). Х37= 18, х38= 18. Теперь определяем значение медианы

18 тыс. руб.

#### Медиана для интервального ряда

При вычислении медианы **для интервального вариационного ряда** сначала определяют медианный интервал, в пределах которого находится медиана.

Для этого:

1. определяется номер медианы по формуле: ,

2. затем по накопленной частоте определяется интервал, в который входит элемент с таким номером.Это первый интервал, для которого накопленная частота будет больше или равна номеру медианы.

3. затем определяем значение медианы по формуле:

где:

Ме — искомая медиана

 — нижняя граница интервала, который содержит медиану

 — ширина интервала (верхняя граница интервала – нижняя граница)

 — сумма частот или число элементов в группе

- накопленная частота интервала, предшествующего медианному

 — частота медианного интервала

**Пример**. Найти медиану для интервального ряда.

|  |  |  |
| --- | --- | --- |
| Возрастные группы | Число студентов | Накопленная частота |
| До 20 лет | 346 | 346 |
| 20 — 25 | 872 | 1218 |
| 25 — 30 | 1054 | 2272 |
| 30 — 35 | 781 | 3053 |
| 35 — 40 | 212 | 3265 |
| 40 — 45 | 121 | 3386 |
| 45 лет и более | 76 | 3462 |
| Итого | 3462 |  |

Решение:

1. Определим номер медианы (Σfi/2 = 3462/2 = 1731),
2. Медианный интервал находится в возрастной группе 25-30 лет, так как это первый интервал, для которого накопленная частота больше (или равна) номера медианы (1218<1731, 2272 > 1731).
3. Далее подставляем в формулу необходимые числовые данные и получаем значение медианы:

Это значит, что одна половина студентов находится в возрасте до 27,4 года.

### Квартили

**Квартили**представляют собой значения признака, делящие упорядоченную по возрастанию совокупность **на четыре равные по количеству элементов части**.

Различают квартиль первого порядка (нижний квартиль) **,**квартиль второго порядка (совпадает с медианой) **,**квартиль третьего порядка (верхний квартиль). Первый (нижний) квартиль отсекает от совокупности ¼ часть единиц с минимальными значениями, а третий (верхний) отсекает ¼ часть единиц с максимальными значениями, второй квартиль является медианой.

Для расчёта квартилейможно поделить вариационный ряд медианой на две равные части, а затем в каждой из них найти медиану.

К примеру, если выборка состоит из 6 элементов, тогда за начальную квартиль выборки принимается второй элемент, а за нижнюю квартиль пятый элемент.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 1 квартиль |  |  | 2 квартиль |  |

медиана

В случае, если вариационный ряд состоит к примеру, из 9 элементов, тогда за верхнюю квартиль принимают арифметическое среднее 2-го и 3-го элементов, а за нижнюю арифметическоесреднее 7-го и 8-го элементов.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  | медиана |  |  |  |  |

1 квартиль 3 квартиль

#### Расчет квартилей для дискретного ряда:

1. В дискретном ряду сначала определяют номера квартилей:

номер 1-го квартиля

позиция 3-го квартиля

2. Если номер квартиля– целое число, то значение квартиля будет равно величине элемента ряда, которое обладает накопленной частотой равной номеру квартиля. Например, номер квартиля равен 20, его значение будет равно значению признака с S =20 (накопленной частотой равной 20).

Если номер квартиля – нецелое число, то квартилем будет условное число между двумя наблюдениями. Значением квартиля будет сумма, состоящая из значения элемента, для которого накопленная частота равна целому значению номера квартиля, и указанной части (нецелая часть номера квартиля) разности между значением этого элемента и значением следующего элемента.

Например, если номер квартиля равна 20,25, квартиль попадает между 20-м и 21-м наблюдениями, и его значение будет равно значению 20-го наблюдения плюс 1/4 (0,25) разности между значением 20-го и 21-го наблюдений.

*Пример.  Найти третий квартиль*

| Заработная плата  рабочего, тыс.руб; | бригада 1 | | бригада 2 | |
| --- | --- | --- | --- | --- |
| ,  число рабочих | ,  накопленная  частота | ,  число рабочих | ,  накопленная частота |
| 15 | 20 | 20 | 20 | 20 |
| 18 | 37 | 57 | 37 | 57 |
| 20 | 14 | 71 | 14 | 71 |
| 25 | 4 | 75 | 5 | 76 |
| Итого: | 75 |  | 76 |  |

1. Определяем номер 3-го квартиля

для первой бригады ;

для второй бригады ,75;

1. Для первой бригады номер квартиля – целое число. Следовательно, нужно найти элемент совокупности, для которого накопленная частота S равна номеру квартиля. Для этого определяем самую первую накопленную частоту, которая больше или равна номеру квартиля. Это накопленная частота второго значения признака Хi=18, Si=57.  
   Определяем значение третьего квартиля для первой бригады :Q3 = x57 = 18 тыс.руб.

Это значит, что заработная плата 75% рабочих 1-ой бригады менее 18 тыс.руб

1. Номер квартиля для второй бригады - нецелое число. Для определения квартиля нужны значения двух элементов – х57 и х58. Значение квартиля находится между ними.   
   Определяем их значение с помощью самой первой накопленной частоты большей или равной порядковым номерам элементов (57 и 58). Х57= 18, х58= 20.   
   Теперь определяем значение 3-го квартиля для второй бригады:  
   Q3= х57 + (х58 – х57)×0,75 = 18 + (20– 18)×0,75 = 19,5 тыс. руб.

Это значит, что заработная плата 75% рабочих второй бригады менее 19,5 тыс.руб.

#### Расчет квартилей для интервального ряда:

Для расчета квартилей для интервального ряда

1. Определяем номер квартиля по тем же формулам, что и для дискретного ряда,
2. Определяем квартильный интервал по накопленной частоте. Это первый интервал, для которого накопленная частота будет больше или равна номеру квартиля.
3. Рассчитываем квартиль по формуле:

Где:

J – номер квартиля,

 - нижняя граница интервала, содержащего квартиль. Интервал определяется по накопленной частоте интервалов,

 - ширина интервала, содержащего квартиль,

 - накопленная частота интервала, предшествующего интервалу, содержащему квартиль,

 - частота интервала, содержащего квартиль.

**Пример**. Найти первый квартиль для интервального ряда.

|  |  |  |
| --- | --- | --- |
| Возрастные группы | Число студентовf | Накопленная частота S |
| До 20 лет | 346 | 346 |
| 20 — 25 | 872 | 1218 |
| 25 — 30 | 1054 | 2272 |
| 30 — 35 | 781 | 3053 |
| 35 — 40 | 212 | 3265 |
| 40 — 45 | 121 | 3386 |
| 45 лет и более | 76 | 3462 |
| Итого | 3462 |  |

Решение:

1. Определяем номер первого квартиля по формуле
2. Первый квартиль находится в возрастной группе 20-25 лет, так как это первый интервал, для которого накопленная частота больше (или равна) номера квартиля (346<865,75; 1218>865,75).
3. Определяем первый квартиль по формуле

Это значит, что четверть студентов младше 22,98 лет.

### Децили

**Децили – значения признака, делящие ранжированный ряд на десять равных частей.**

Первый дециль отсекает 1/10 часть совокупности, а девятый дециль отсекает 9/10 частей. Таким образом, различают 9 децилей.

Рассчитываются децили аналогично квартилям.

#### Расчет децилей для дискретного ряда

1. Определяем номер дециля по формуле: ,
2. Если номер дециля – целое число, то значение дециля будет равно величине элемента ряда, которое обладает накопленной частотой равной номеру дециля. Например, если номер дециля равен 20, его значение будет равно значению признака с S =20 (накопленной частотой равной 20).

Если номер дециля – нецелое число, то дециль попадает между двумя наблюдениями. Значением дециля будет сумма, состоящая из значения элемента, для которого накопленная частота равна целому значению номера дециля, и указанной части (нецелая часть номера дециля) разности между значением этого элемента и значением следующего элемента.

Например, если номер дециля равна 20,25, дециль попадает между 20-м и 21-м наблюдениями, и его значение будет равно значению 20-го наблюдения плюс 1/4 разности между значением 20-го и 21-го наблюдений.

#### Расчет децилей для интервального ряда

1. Определяем номер дециля по формуле: ,
2. Определяем децильный интервал. Это первый интервал, для которого накопленная частота будет больше или равна номеру дециля.
3. Рассчитываем дециль по формуле:

где – значение j-го дециля,

- нижняя граница децильного интервала;

- ширина децильного интервала;

– сумма всех частот,



-накопленная частота интервала, предшествующего децильному;



- частота децильного интервала.



*Пример.  Найти 9-ый дециль D9*

| Заработная плата  рабочего, тыс.руб; | бригада 1 | |
| --- | --- | --- |
| ,  число рабочих | ,  накопленная  частота |
| 15 | 20 | 20 |
| 18 | 37 | 57 |
| 20 | 14 | 71 |
| 25 | 4 | 75 |
| Итого: | 75 |  |

1. Определяем номер 9-го дециля

для первой бригады ;

1. Номер дециля - нецелое число. Для определения дециля нужны значения двух элементов – х68 и х69. Значение дециля находится между ними.   
   Определяем их значение с помощью самой первой накопленной частоты большей или равной порядковым номерам элементов (68 и 69). Х68= 20, х69= 20.   
   Теперь определяем значение 9-го дециля:  
   *D7* = x68+ (х69 – х68)×0,4=20 + (20 – 20)×0,2 =20тыс.руб.

Это значит, что заработная плата90% рабочих бригады не превышает 18 тыс.руб.

***Пример****. Найти седьмой дециль D7 для интервального ряда.*

|  |  |  |
| --- | --- | --- |
| Возрастные группы | Число студентовf | Накопленная частота S |
| До 20 лет | 346 | 346 |
| 20 — 25 | 872 | 1218 |
| 25 — 30 | 1054 | 2272 |
| 30 — 35 | 781 | 3053 |
| 35 — 40 | 212 | 3265 |
| 40 — 45 | 121 | 3386 |
| 45 лет и более | 76 | 3462 |
| Итого | 3462 |  |

Решение:

1. Определяем номер седьмого дециля по формуле 7×
2. Седьмой дециль находится в возрастной группе 30-35 лет, так как это первый интервал, для которого накопленная частота больше (или равна) номера дециля (2272<2424,1; 3053>2424,1).
3. Определяем седьмой дециль по формуле

Это значит, что 70% студентов младше 30,97 лет.

### Перцентили

**Перцентили –значения признака, делящие ранжированный ряд на 100 равных частей.**

Различают 99 перцентилей.

#### Расчет перцентилей для дискретного ряда

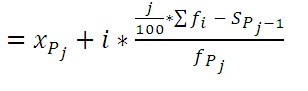
1. Определяем номер перцентиля по формуле: ,
2. Если номер перцентиля – целое число, то значение перцентиля будет равно величине элемента ряда, которое обладает накопленной частотой равной номеру перцентиля. Например, если номер перцентиля равен 20, его значение будет равно значению признака с S =20 (накопленной частотой равной 20).

Если номер перцентиля – нецелое число, то перцентиль попадает между двумя наблюдениями. Значением перцентиля будет сумма, состоящая из значения элемента, для которого накопленная частота равна целому значению номера перцентиля, и указанной части (нецелая часть номера перцентиля) разности между значением этого элемента и значением следующего элемента.

Например, если номер перцентиля равна 20,25, перцентиль попадает между 20-м и 21-м наблюдениями, и его значение будет равно значению 20-го наблюдения плюс 1/4 разности между значением 20-го и 21-го наблюдений.

#### Расчет перцентилей для интервального ряда

1. Определяем номер перцентиля по формуле: ,
2. Определяем перцентильныйинтервал по накопленной частоте. Это первый интервал, для которого накопленная частота будет больше или равна номеру перцентиля.
3. Определяем значение перцентиля по формуле



– значение j-го перцентиля,

- нижняя граница перцентильного интервала;

- ширина перцентильного интервала;

– сумма всех частот,

-накопленная частота интервала, предшествующего перцентильному;



- частота перцентильного интервала.



*Пример.  Найти 10-ой перцентиль P10*

| Заработная плата  рабочего, тыс.руб; | бригада 1 | |
| --- | --- | --- |
| ,  число рабочих | ,  накопленная  частота |
| 15 | 20 | 20 |
| 18 | 37 | 57 |
| 20 | 14 | 71 |
| 25 | 4 | 75 |
| Итого: | 75 |  |

1. Определяем номер 10-го перцентиля

;

1. Номер перцентиля - нецелое число. Для определения перцентиля нужны значения двух элементов – х7 и х8. Значение перцентиля находится между ними.   
   Определяем их значение с помощью самой первой накопленной частоты большей или равной порядковым номерам элементов (7 и 8). Х7= 15, х8= 15.   
   Теперь определяем значение 67-го перцентиля:  
    = x7+ (х8 – х7)×0,6= 15 + (15 – 15)×0,6 = 15тыс.руб.

Это значит, что заработная плата 10% рабочих бригады менее 15тыс.руб.

***Пример****. Найти 95-ый перцентильP95для интервального ряда.*

|  |  |  |
| --- | --- | --- |
| Возрастные группы | Число студентовf | Накопленная частота S |
| До 20 лет | 346 | 346 |
| 20 — 25 | 872 | 1218 |
| 25 — 30 | 1054 | 2272 |
| 30 — 35 | 781 | 3053 |
| 35 — 40 | 212 | 3265 |
| 40 — 45 | 121 | 3386 |
| 45 лет и более | 76 | 3462 |
| Итого | 3462 |  |

Решение:

1. Определяем номер 95-го перцентиляпо формуле 7×
2. 95-ый перцентильнаходится в возрастной группе 40-45 лет, так как это первый интервал, для которого накопленная частота больше (или равна) номера дециля (3265<3289,85; 3386>3289,85).
3. Определяем седьмой дециль по формуле

Это значит, что 95% студентов младше 40,99 лет.

## Коэффициенты, характеризующие концентрацию изучаемого признака в определенных группах

#### Коэффициент Герфиндаля

Обобщенная оценка степени структуризации явления в целом обычно выполняется по формуле *уровня концентрации* (или *коэффициент Герфиндаля*), который более чувствителен к изменению долей групп с наибольшим удельным весом в итоге, определяемый по формуле:

где – доля *j*-ой группы в общем итоге изучаемого значения признака; m – количество выделенных в совокупности групп.

Максимальное значение коэффициента – 10 000 (если доли в процентах) или 1(если доли в числах). Чем ближе значение коэффициента к максимальному, тем более высокая концентрация изучаемого признака в одной или нескольких группах. Минимальное значение коэффициента – 100 (если доля в процентах) или 0,01 (если доли в числах).

#### Эффективное число групп

Обратная индексу Герфиндаля величина – это *эффективное число групп* в структуре, которое показывает количество групп без учета групп, имеющих ничтожно малые доли, определяется по формуле:

*E = 1/H.*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Выработка на человека | кол-во человек | среднее по группе (середина интервала) | выработка на группу | dxj | dxj2 |
| 10-16 | 5 | 13 | 65 | 0,0035 | 0,00001225 |
| 16-22 | 110 | 19 | 2090 | 0,1127 | 0,01270129 |
| 22-28 | 182 | 25 | 4550 | 0,2454 | 0,06022116 |
| 28-34 | 120 | 31 | 3720 | 0,2006 | 0,04024036 |
| 34-40 | 90 | 37 | 3330 | 0,1796 | 0,03225616 |
| 40-46 | 60 | 43 | 2580 | 0,1392 | 0,01937664 |
| 46-52 | 45 | 49 | 2205 | 0,119 | 0,014161 |
|  | 612 |  | 18540 | 1 | 0,17896886 |

Коэффициент Герфиндаля, что говорит о незначительной концентрации изучаемого признака (выработки) в одной или нескольких группах.

Количество эффективных групп E = 1/H = 1/0.179 = 5.6. Этот показатель показывает, что изучаемый признак сконцентрирован в 6 группах из 7 имеющихся. Что опять же подтверждает относительно равномерное распределение выработки среди групп рабочих.

## Коэффициенты, характеризующие дифференциацию изучаемого признака в совокупности

На основе квантилей рассчитываются различные коэффициенты дифференциации изучаемого признака: децильный (квартильный, перцентильный) коэффициент, коэффициент фондов, коэффициент Джинни. Эти коэффициенты характеризуют неравномерность распределения признака в изучаемой совокупности. Например, неравенство в распределении доходов (как сильно доходы наиболее обеспеченной группы отличаются от доходов наименее обеспеченной группы).

Децильный коэффициент дифференциации,

где

Этот коэффициент показывает во сколько раз самое маленькое значение признака среди 10% единиц наблюдения с самыми большими значениями признака (самый маленький доход 10% самых обеспеченных) больше чем самое большое значение признака среди 10% единиц наблюдения с самыми маленькими значениями признака (самый большой доход 10% самых малообеспеченных).

Квартильный коэффициент дифференциации,

где

Этот коэффициент показывает во сколько раз самое маленькое значение признака среди 25% единиц наблюдения с самыми большими значениями признака (самый маленький доход 25% самых обеспеченных) больше чем самое большое значение признака среди 25% единиц наблюдения с самыми маленькими значениями признака (самый большой доход 25% самых малообеспеченных).

Коэффициент фондов– соотношение среднего значения изучаемого признака в десятой децильной группе к среднему значению изучаемого признака в первой децильной группе

Этот коэффициент показывает во сколько раз среднее значение признака среди 10% единиц наблюдения с самыми большими значениями признака (средний доход 10% самых обеспеченных) больше чем среднее значение признака среди 10% единиц наблюдения с самыми маленькими значениями признака (средний доход 10% самых малообеспеченных).

Первая децильная группа – это интервал значений от минимального до первого дециля. Десятая децильная группа – это интервал значений от девятого дециля до максимального значения. Среднее для интервала значений – это середина интервала.

Следовательно,

;

;

Чем большее значение принимают эти коэффициенты, тем большее неравенство в распределении благ (обязанностей) между десятой децильной группой и первой децильной группой.

Рассмотрим эти коэффициенты на примере.

Сведения о выработке рабочих за октябрь

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Выработка одного рабочего | | | | кол-во человек | | накопленная частота | | | |
| 10-16 | | | | 5 | | 5 | | | |
| 16-22 | | | | 110 | | 115 | | | |
| 22-28 | | | | 182 | | 297 | | | |
| 28-34 | | | | 120 | | 417 | | | |
| 34-40 | | | | 90 | | 507 | | | |
| 40-46 | | | | 60 | | 567 | | | |
| 46-52 | | | | 45 | | 612 | | | |
|  | | | | 612 | |  | | | |
|  | | | |  | | |  | | |
|  | номер | | | | значение | | | | | |
| 1 дециль | 1×(612+1):10= | | 61,3 | | 16+(22-16)×((612/10)-5)/110= | | | | 19,07 | |
| 9 дециль | 9×(612+1):10= | | 551,7 | | 40+(46-40)×(9×(612/10)-567)/60= | | | | 44,38 | |
| 1 квартиль | 1×(612+1):4= | | 153,25 | | 22+(28-22)×((612/4)-115)/182= | | | | 23,25 | |
| 3 квартиль | 3×(612+1):4= | | 459,75 | | 34+(40-34)×(3×(612/4)-417)/90= | | | | 36,8 | |

|  |  |  |
| --- | --- | --- |
| Децильный коэффициент ; | 44,38 : 19,07 = | 2,3 |
| Квартильный коэффициент ; | 36,8 : 23,25 = | 1,6 |

Значение коэффициентов позволяют сделать следующие выводы: самая маленькая выработка 10% лучших рабочих превышает лучшую выработку 10 % самых непроизводительных рабочих в 2,3 раза.

Самая низкая выработка 25% лучших рабочих превышаю лучшую выработку 25% самых непроизводительных рабочих в 1,6 раза.

Для расчета коэффициента фондов нужно найти средние значение 1-ой и 10-ой децильных групп.

= ;

Этот коэффициент позволяет сделать вывод, что средняя выработка лучших рабочих в 3.3 раза превышает выработку самых низкопроизводительных рабочих.

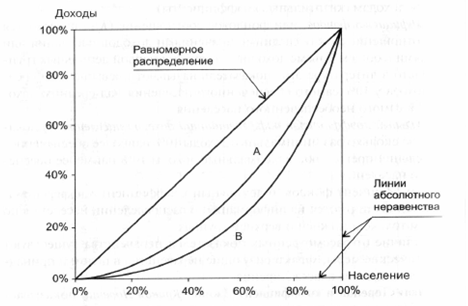
### Кривая Лоренца и коэффициент Джинни

Кривую Лоренца часто используют, чтобы графически показать распределение дохода среди населения. Однако её можно использовать и при оценке равномерности распределения любого блага или обязанностей.

Рассмотрим как построить кривую Лоренца для оценки неравномерности распределения дохода.

При построении кривой Лоренца по горизонтальной оси откладываются единицы наблюдения, упорядоченные по возрастанию дохода (процентные группы, получающие доход, начиная от самого минимального).

По вертикальной оси откладывается накопительная (кумулятивная) доля дохода для каждой процентной группы.

Отрезки, соединяющие полученные точки, образуют кривую Лоренца.

При абсолютно равномерном распределении дохода среди населения кривая Лоренца будет биссектрисой угла системы ординат. Чем ближе полученная кривая Лоренца к биссектрисе, тем более равномерно распределено благо среди населения. Чем больше кривая отклоняется (становится более вогнутой), тем больше неравенство в распределении.

На рисунке показано распределение дохода в двух странах. Кривая Лоренца для страны А намного ближе к диагонали, что свидетельствует о более равномерном распределении дохода в стране А, чем в стране В.

На основании данных, используемых при построении кривой Лоренца, рассчитывают коэффициент, отражающий степень неравномерности распределения – коэффициент Джинни.

Коэффициент Джинни определяется по следующй формуле:

,

где - накопительная доля дохода для i-ой группы,

- накопительная доля дохода для (i+1)-ой группы,

- накопительная доля единиц наблюдения (групп населения) для i-ой группы,

- накопительная доля единиц наблюдения (групп населения) для (i+1)-ой группы,

– количество групп населения.

Значение коэффициента Джинни наxодится в интервале от 0 до 1. При G=0 – абсолютное неравенство в распределении, при G=1 – абсолютное равенства. Чем ближе коэффициент к 1, тем более равномерно распределено изучаемое благо, чем ближе к 0 – тем более неравномерно.

Пример расчета коэффициента Джинни приведен на следующей странице.

= 0,1661 – коэффициент Джинни

Полученное небольшое значение коэффициента Джинни свидетельствует о достаточно равномерном распределении выработки среди рабочих.

Построим кривую Лоренца, откладывая значения dxjнакна оси y, а значенияdfjнак – на оси x.

Кривая Лоренца

Пример расчета коэффициента Джинни.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Выработка на человека | кол-во человек | среднее по группе (середина интервала) | выработка на группу | dxj | dfj | dxj+dxj-1= | dxjнак | dfj+dfj-1= | dfjнак | (dxjнак+dxj+1нак) | | (dfj+1нак-dfjнак) | | (dxj+1нак+dxjнак)×(dfj+1нак-dfjнак) | |
| 10-16 | 5 | 13 | 65 | 0,0035 | 0,0082 | 0,0035 = | 0,0035 | 0,0082 = | 0,0082 | 0,0035 + 0,1162 = | 0,1197 | 0,1879-0,0082 = | 0,1797 | 0,1797×0,1197 = | 0,02151 |
| 16-22 | 110 | 19 | 2090 | 0,1127 | 0,1797 | 0,1127 + 0,0035 = | 0,1162 | 0,1797 + 0,0082 = | 0,1879 | 0,1162 + 0,3616 = | 0,4778 | 0,4853-0,1879 = | 0,2974 | 0,2974×0,4778 = | 0,142098 |
| 22-28 | 182 | 25 | 4550 | 0,2454 | 0,2974 | 0,2454 + 0,1162 = | 0,3616 | 0,2974 + 0,1879 = | 0,4853 | 0,3616 + 0,5622 = | 0,9238 | 0,6814-0,4853 = | 0,1961 | 0,1961×0,9238 = | 0,181157 |
| 28-34 | 120 | 31 | 3720 | 0,2006 | 0,1961 | 0,2006 + 0,3616 = | 0,5622 | 0,1961 + 0,4853 = | 0,6814 | 0,5622 + 0,7418 = | 1,304 | 0,8285-0,6814 = | 0,1471 | 0,1471×1,304 = | 0,191818 |
| 34-40 | 90 | 37 | 3330 | 0,1796 | 0,1471 | 0,1796 + 0,5622 = | 0,7418 | 0,1471 + 0,6814 = | 0,8285 | 0,7418 + 0,881 = | 1,6228 | 0,9265-0,8285 = | 0,0980 | 0,098×1,6228 = | 0,159034 |
| 40-46 | 60 | 43 | 2580 | 0,1392 | 0,098 | 0,1392 + 0,7418 = | 0,881 | 0,098 + 0,8285 = | 0,9265 | 0,881 + 1 = | 1,881 | 1-0,9265 = | 0,0735 | 0,0735×1,881 = | 0,138254 |
| 46-52 | 45 | 49 | 2205 | 0,119 | 0,0735 | 0,119 + 0,881 = | 1 |  | 1,0000 | - | | - | | - | |
|  | 612 |  | 18540 | 1 | 1 |  |  |  |  |  | |  | | Всего | 0,8339 |

.

## Показатели, характеризующие изменение структуры во времени

Для анализа изменения структуры совокупности во времени используют показатели структурных сдвигов.

#### Квадратический коэффициент абсолютных структурных сдвигов

= ,

Где - доля j-ой группы в анализируемом периоде (отчетном),

- доля j-ой группы в периоде, с которым сравнивают отчетный (базисный),

m – количество групп, выделенных в совокупности

Если никаких изменений в структуре не произошло, то коэффициент равен нулю. Коэффициент показывает на сколько процентов изменились в среднем доли групп в отчетном периоде по сравнению с базисным.

#### Индекс различий

Коэффициент может принимать значения от 0 до 1. В действительности, он никогда не бывает равен 1. Чем ближе значения коэффициента к 1, тем более сильно изменилась структура.

Рассчитаем

А) коэффициент структурных сдвигов

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| m | Выработка на человека | кол-во человек | | dfj | | dfjноя - dfjокт | | (dfjноя - dfjокт)2 |
| октябрь | ноябрь | октябрь | ноябрь |
| 1 | 10-16 | 5 | 3 | 0,0082 | 0,0049 | 0,0049 - 0,0082 = | -0,0033 | 0,00001089 |
| 2 | 16-22 | 110 | 130 | 0,1797 | 0,2124 | 0,2124 - 0,1797 = | 0,0327 | 0,00106929 |
| 3 | 22-28 | 182 | 170 | 0,2974 | 0,2778 | 0,2778 - 0,2974 = | -0,0196 | 0,00038416 |
| 4 | 28-34 | 120 | 133 | 0,1961 | 0,2173 | 0,2173 - 0,1961 = | 0,0212 | 0,00044944 |
| 5 | 34-40 | 90 | 50 | 0,1471 | 0,0817 | 0,0817 - 0,1471 = | -0,0654 | 0,00427716 |
| 6 | 40-46 | 60 | 80 | 0,098 | 0,1307 | 0,1307 - 0,098 = | 0,0327 | 0,00106929 |
| 7 | 46-52 | 45 | 46 | 0,0735 | 0,0752 | 0,0752 - 0,0735 = | 0,0017 | 0,00000289 |
|  | Всего | 612 | 612 | 1 | 1 |  |  | 0,00726312 |

Коэффициент структурных сдвигов= 0,0322,

Полученный коэффициент показывает, что в среднем доли групп изменились на 0,0322 или на 3.22%.

Б) индекс различий

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Выработка на человека | кол-во человек | | dfj | | |dfjноя - dfjокт| |
| октябрь | ноябрь | октябрь | ноябрь |
| 10-16 | 5 | 3 | 0,0082 | 0,0049 | 0,0033 |
| 16-22 | 110 | 130 | 0,1797 | 0,2124 | 0,0327 |
| 22-28 | 182 | 170 | 0,2974 | 0,2778 | 0,0196 |
| 28-34 | 120 | 133 | 0,1961 | 0,2173 | 0,0212 |
| 34-40 | 90 | 50 | 0,1471 | 0,0817 | 0,0654 |
| 40-46 | 60 | 80 | 0,098 | 0,1307 | 0,0327 |
| 46-52 | 45 | 46 | 0,0735 | 0,0752 | 0,0017 |
| Всего | 612 | 612 | 1 | 1 | 0,1766 |

= 0.5×0.1766 = 0,0883

Небольшое значение индекса различий говорит о незначительном изменении в структуре численности рабочих по объёму выработки на человека в ноябре по сравнению с октябрём.