

Ряды распределения и их количественные характеристики

Романцева Юлия Николаевна,
кандидат экономических наук, доцент
кафедры статистики и кибернетики
ФГБОУ ВО РГАУ-МСХА
им. К.А. Тимирязева



Вопросы:

- Цифровые ресурсы статистической информации.
- Подходы и методы упорядочивания массовых данных. Ряды распределения.
- Количественные характеристики рядов распределения.
- Статистические гипотезы, методика проверки гипотез относительно распределения единиц в совокупности и основных параметров.



Источники статистических данных

Первичные статистические данные

документированная информация по формам федерального статистического наблюдения, получаемая от респондентов, или информация, документируемая непосредственно в ходе наблюдения

Административные данные

используемая при формировании официальной статистической информации документированная информация, получаемая органами государственной власти на всех уровнях управления в связи с осуществлением ими разрешительных, регистрационных, контрольно-надзорных и других административных функций



Источники статистических данных

Государственная статистика:

- Официальный сайт Росстата. URL: <https://rosstat.gov.ru/>
- ЕМИСС Государственная статистика. <https://www.fedstat.ru/>
- Официальный сайт Министерства сельского хозяйства РФ. <https://mcx.gov.ru/>
- Официальный сайт Министерства финансов РФ. URL: [http:// www.minfin.gov.ru](http://www.minfin.gov.ru)
- Официальный сайт Центрального Банка России. URL: <http://www.cbr.ru>
- Московская международная валютная биржа. <http://www.micex.ru>

Международная статистика:

- Официальный сайт Евростата. <https://ec.europa.eu/eurostat>
- Официальный сайт Статистического отдела ООН. <https://unstats.un.org/UNSDWebsite/>
- Основные обзоры и доклады ООН в экономической и социальной областях. URL: <http://www.un.org/ru/development/surveys/>
- Официальный сайт Всемирного банка. URL: [http:// www.worldbank.org](http://www.worldbank.org)
- Официальный сайт Всемирной торговой организации. URL: <http://www.wto.org>
- Официальный сайт Европейского банка реконструкции и развития – URL: <http://www.ebrd.com>
- Официальный сайт Международного валютного фонда. URL: [http:// www.imf.org](http://www.imf.org)
- Официальный сайт Международной организации труда. URL: [http:// www.ilo.org](http://www.ilo.org)
- Системы национальных счетов (System of National Accounts 2008). URL: <http://unstats.un.org/unsd/nationalaccount/sna2008.asp>
- Официальный сайт Национального бюро экономических исследований США. URL: [http:// www.nber.org](http://www.nber.org)
- Официальный сайт Бюро экономического анализа (Bureau of Economic Analysis). URL: <http://www.bea.gov>
- Официальный сайт Национальной службы сельскохозяйственной статистики (NASS – National Agricultural Statistics Service). URL: www.nass.usda.gov
- Официальный сайт ОЭСР. URL: <http://www.oecd-ilibrary.org/statistics;jsessionid=3ddci6tti4o90.delta>



Анкетирование – метод исследования и первичного сбора информации у отдельного человека или группы людей, объединённых по определённому признаку. Проводится с использованием специальных вопросников (анкет), рассылаемых определенному кругу лиц или публикующиеся в периодической печати.

Источники данных анкетирования:

«Левада-центр», опросы и аналитика. <https://www.levada.ru/>
Единый архив статистических и эконометрических данных ВШЭ.
http://sophist.hse.ru/data_access.shtml
Фонд общественного мнения <https://fom.ru/>
Опросы ТАСС. <https://tass.ru/oprosy-obschestvennogo-mneniya>



Большие данные (Big Data) — это структурированные или неструктурированные массивы данных большого объема, обработка которых предусматривает использование специальных автоматизированных инструментов для дальнейшего использования в статистике, анализе, прогнозах и принятии решений.

Источники Big Data:

Корпоративные данные:

- Базы данных организации
- ERP
- CRM
- Системы электронного документооборота
- Архивы
- другие

Внешние данные

- Интернет: соцсети, блоги, СМИ, форумы, сайты
- Внешние наборы данных (из разных источников)
- Другие

Данные устройств

- Показания устройств (датчики, приборы, данные сотовой связи, метеорологические данные и др.)
- Данные мобильных устройств (частота использования приложений, геолокация и др.)
- Данные IoT-устройств

Подходы и методы упорядочивания массовых данных.

Ряды распределения



№ по ранжиру	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Заработная плата (тыс. руб.)	25	28	31	35	35	38	41	44	45	47	53	56	57	63	72

Пример ранжированного ряда распределения
работников по уровню заработной платы

Подходы и методы упорядочивания массовых данных.

Ряды распределения



Ранжированный ряд распределения позволяет:

- оценить характер изменения признака
- оценить интенсивность изменения признака
- выявить резко выделяющиеся значения

Подходы и методы упорядочивания массовых данных.

Ряды распределения



Число обработок ядохимикатами	Количество полей
x_i	n_i
1	3
2	5
3	8
4	4
5	1
Итого	21

Дискретный ряд распределения полей по числу обработок ядохимикатами

Подходы и методы упорядочивания массовых данных.

Ряды распределения



Дискретный ряд распределения позволяет:

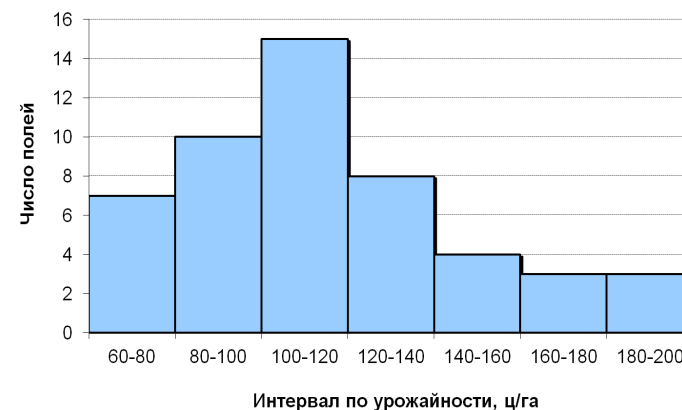
- установить минимальное и максимальное значение признака
- какие значения признака встречаются наиболее часто и редко
- полигон дискретного ряда дает возможность предположить близость эмпирического распределения нормальному закону

Подходы и методы упорядочивания массовых данных.

Ряды распределения



№ п/п	Интервал по урожайности, ц/га	Число полей (n_i)
1	60-80	7
2	80-100	10
3	100-120	15
4	120-140	8
5	140-160	4
6	160-180	3
7	180-200	3
Итого		50



Интервальный ряд распределения позволяет:

- интервалы с минимальными и максимальными значениями признака
- интервалы значений признака, наиболее часто встречающихся, редко встречающихся
- гистограмма распределения дает возможность установить близость формы эмпирического распределения нормальному закону распределения

Подходы и методы упорядочивания массовых данных.

Ряды распределения



Направление (x_i)	Число мест (n_i)
Государственное и муниципальное управление	5
Экономика	15
Менеджмент	5
Профессиональное обучение	75
Экономическая безопасность	10
Информационные системы и технологии	75
Прикладная информатика	52
Итого	237

Вариационный ряд распределения по качественному признаку

Распределение бюджетных мест в РГАУ-МСХА имени К.А. Тимирязева по направлениям подготовки



Количественные характеристики рядов распределения

Показатели центральной тенденции:

- параметрические
- непараметрические

Показатели вариации:

- абсолютные
- средние
- относительные

Показатели формы распределения:

- островершинности
- скошенности



Количественные характеристики рядов распределения

Степенные средние простые и взвешенные:

- средняя арифметическая
- средняя гармоническая

$$\bar{x} = \frac{\sum x_i}{N} \quad \bar{x} = \frac{\sum x_i f_i}{\sum f_i}$$

- средняя геометрическая
- средняя квадратичная

$$\bar{x} = \sqrt[n]{\prod_{i=1}^n K_i}$$

Структурные средние:

- модальное значение – наиболее часто встречающееся значение x_{mo}
- медианное значение – значение, принадлежащее единице в середине ранжированного ряда x_{me}



Количественные характеристики рядов распределения

Показатели вариации:

- Размах вариации $R = x_{max} - x_{min}$

- Объем вариации $W = \sum (x_i - \bar{x})^2$

- Среднее линейное отклонение $L = \frac{\sum |x_i - \bar{x}| f_i}{\sum f_i}$

- Дисперсия $\sigma^2 = \frac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i}$

- Среднеквадратическое отклонение $\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i}}$

- Коэффициент вариации $V = \frac{\sigma}{\bar{x}} \times 100$

Показатели формы распределения:

- Показатель асимметрии $A_s = \frac{\sum (x_i - \bar{x})^3 / \sum n_i}{\sigma^3}$

- Экссесс $E = \frac{\sum (x_i - \bar{x})^4 / \sum n_i}{\sigma^4} - 3$

Статистические гипотезы, методика проверки гипотез относительно распределения единиц в совокупности и основных параметров



Общая схема проверки гипотез:

- Формулировка двух гипотез: нулевой (рабочей) – H_0 и альтернативной – H_A
- Выбор статистического критерия
- Выбор уровня значимости
- Расчет фактического значения критерия по выборке
- Нахождение теоретического (критического) значения критерия по специальным математическим таблицам
- Сравнение фактического значения критерия с табличным

$Q_{\text{факт}} \leq Q_{\text{табл}}$ – критерий находится в области согласия с нулевой гипотезой

$Q_{\text{факт}} > Q_{\text{табл}}$ – критерий находится в зоне отказа от нулевой гипотезы

- Формулировка вывода о справедливости одной из выдвинутых гипотез



Спасибо за внимание!