

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РФ
Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования
Томский государственный университет систем управления и
радиоэлектроники

Факультет систем управления
Кафедра автоматизированных систем (АСУ)

Е.Б. Грибанова

Статистика

**Методические указания по выполнению практических работ для
студентов направления подготовки
09.03.03 – Прикладная информатика (бакалавриат)**

2016

Е.Б. Грибанова

Статистика

Методические указания по выполнению практических работ для студентов направления подготовки 09.03.03 – Прикладная информатика (профиль прикладная информатика в экономике). – Томск: ТУСУР, 2016 (электр. ресурс). – 24 с.

В пособии приводится описание практических работ по дисциплине «Статистика» и представлены примеры выполнения заданий. Пособие подготовлено для студентов, обучающихся по направлению Прикладная информатика (профиль прикладная информатика в экономике).

СОДЕРЖАНИЕ

| | |
|--|----|
| 1. Практические задания | 4 |
| 1.1 Практическая работа 1 | 4 |
| Группировка данных. Построение гистограммы распределения | 4 |
| 1.2 Практическая работа 2 | 7 |
| Расчет коэффициента корреляции Пирсона | 7 |
| 1.3 Практическая работа №3..... | 9 |
| Мода и медиана..... | 9 |
| 1.4 Практическая работа № 4..... | 12 |
| Выборочное наблюдение | 12 |
| 1.5 Практическая работа № 5..... | 14 |
| Статистические индексы..... | 14 |
| 1.6 Практическая работа 6 | 16 |
| Метод k-средних | 16 |
| 1.7 Практическая работа № 7..... | 18 |
| Расчет коэффициента кластеризации | 18 |
| 2. Примеры выполнения заданий | 19 |
| Список литературы..... | 24 |

1. Практические задания

Задания выполняются в Microsoft Excel.

1.1 Практическая работа 1

Группировка данных. Построение гистограммы распределения

При использовании персональных компьютеров для обработки статистических данных группировка единиц объекта проводится с помощью стандартных процедур. Одна из таких процедур основана на использовании следующей формулы Стерджесса для определения оптимального числа групп:

$$n = 1 + 3,322 \lg N ,$$

где n - число групп;

N -объем совокупности.

После определения числа групп решается задача определения интервалов группировки. Интервал группировки - это интервал значений варьирующего признака, лежащих в пределах определенной группы. Каждый интервал имеет свою ширину, верхнюю и нижнюю границы или хотя бы одну, из них. Нижней границей интервала называется наименьшее значение признака в интервале, а верхней границей - наибольшее значение признака в нем. Ширина интервала (ее еще часто называют интервальной разностью) представляет собой разность между верхней и нижней границами интервала. Интервалы группировки, в зависимости от их величины, бывают равные и неравные. Последние делятся на прогрессивно возрастающие, прогрессивно убывающие, произвольные и специализированные. Если вариация признака проявляется в сравнительно узких границах и распределение носит более или менее равномерный характер, то строят группировку с равными интервалами. Величина равного интервала определяется по следующей формуле:

$$h = (X \max - X \min) / n ,$$

где h - величина равного интервала (шаг интервала);

X_{\max}, X_{\min} - максимальное и минимальное значение признака в совокупности

n - число групп.

Гистограмма применяется для изображения интервального вариационного ряда, который представляют столбики с основаниями, равными ширине интервалов, и высотой, соответствующей частоте.

Для таблицы 1.1 гистограмма приведена на рис.1.1.

Таблица 1.1 Интервальный ряд распределения

| №п/п | Размер жилой площади, приходящейся на одного человека | Число семей с данным размером жилой площади | Накопленное число семей |
|-------|---|---|-------------------------|
| 1 | 3-5 | 10 | 10 |
| 2 | 5-7 | 20 | 30 |
| 3 | 7-9 | 40 | 70 |
| 4 | 9-11 | 30 | 100 |
| 5 | 11-13 | 15 | 115 |
| Всего | | 115 | - |

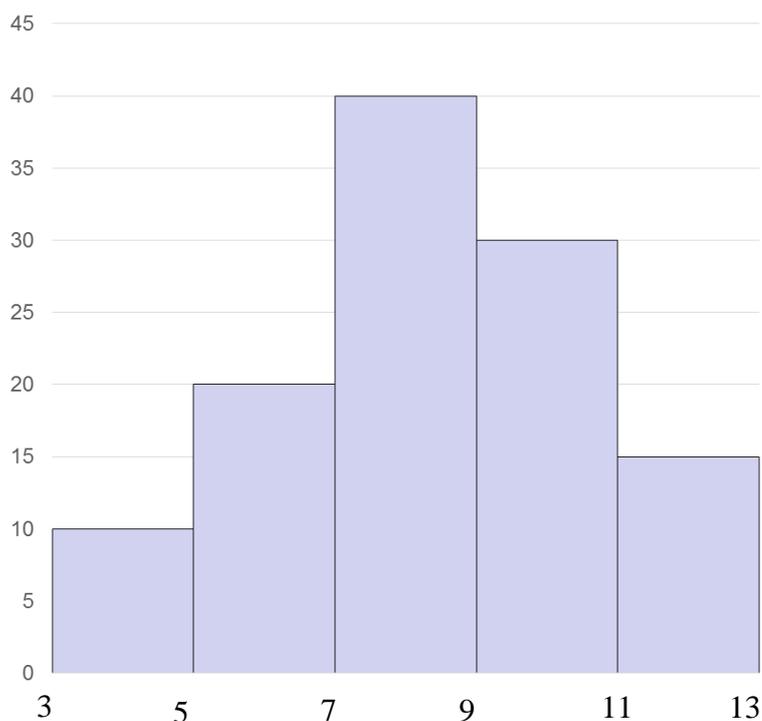


Рис.1.1 Гистограмма распределения

Задание

Выполните в Excel генерацию 100 случайных чисел. Осуществите группировку с равными интервалами. Постройте гистограмму распределения. Для генерации случайных чисел используйте три датчика:

1) $X=5+z(10-5)$

2) $X=-5\text{Ln}(z)$

3) $X=(\sum_{i=1}^{12} z_i - 6) \cdot 5 + 10$

где z - случайная величина, равномерно распределенная на $(0;1)$ (в Excel функция «=СЛЧИС()»).

По гистограмме распределения сделайте вывод о законе распределения случайных величин в трех случаях.

1.2 Практическая работа 2

Расчет коэффициента корреляции Пирсона

Коэффициент корреляции Пирсона вычисляется по следующим формулам:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$
$$r = \frac{n \sum_{i=1}^n x_i y_i - \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right)}{\sqrt{\left(n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 \right) \left(n \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i \right)^2 \right)}}$$

Коэффициент корреляции изменяется на отрезке от -1 до $+1$. Если между переменными существует сильная положительная связь, то значение r будет близко к $+1$. Если между переменными существует сильная отрицательная связь, то значение r будет близко к -1 . Когда между переменными нет линейной связи или она очень слабая, значение r будет близко к 0 .

Интерпретация коэффициента корреляции может быть выполнена в соответствии со следующей шкалой (табл.1.2).

Таблица 1.2 Шкала коэффициента корреляции

| Значение r | Уровень связи между переменными |
|---------------|---------------------------------|
| 0,75 – 1,00 | Очень высокая положительная |
| 0,50 – 0,74 | Высокая положительная |
| 0,25 – 0,49 | Средняя положительная |
| 0,00 – 0,24 | Слабая положительная |
| 0,00 – -0,24 | Слабая отрицательная |
| -0,25 – -0,49 | Средняя отрицательная |
| -0,50 – -0,74 | Высокая отрицательная |
| -0,75 – -1,00 | Очень высокая отрицательная |

Задание

Вычислите коэффициент корреляции Пирсона двух величин: успеваемость студента выбранной группы и пол студента. Выполните интерпретацию полученного результата. Для расчета коэффициента используйте данные официального сайта ТУСУР.

1.3 Практическая работа №3.

Мода и медиана

Модой распределения (M_o) называется такая величина изучаемого признака, которая в данной совокупности встречается наиболее часто, т.е. один из вариантов признака повторяется чаще, чем все другие.

Медиана – значение признака, приходящееся на середину ранжированной совокупности.

Рассмотрим определение моды и медианы по дискретному ряду распределения. В таблице 1.3 представлена информация об успеваемости студентов.

Таблица 1.3 Сведения об успеваемости студентов

| Оценка | Численность студентов |
|--------|-----------------------|
| 2 | 12 |
| 3 | 48 |
| 4 | 56 |
| 5 | 60 |
| Всего | 176 |

По данным табл.1.3 наибольшую частоту (60 чел.) имеет оценка «5», следовательно, это значение и является модальным ($M_o = 5$).

Положение медианы в ряду распределения определяется ее номером:

$$N_{Me} = \frac{n+1}{2},$$

где n - число единиц совокупности.

Для примера таблицы 1.3:

$$N_{Me} = \frac{n+1}{2} = \frac{176+1}{2} = 88,5.$$

Полученное значение указывает, что середина ряда приходится на 88 и 89 студентов. Необходимо определить, какую оценку получили студенты с данными номерами. Это можно сделать, рассчитав накопленные частоты (табл.1.4).

Таблица 1.4 Накопленные частоты

| Оценка | Численность студентов | Накопленные частоты |
|--------|-----------------------|---------------------|
| 2 | 12 | 12 |
| 3 | 48 | 60 |
| 4 | 56 | 116 |
| 5 | 60 | 176 |
| Всего | 176 | |

Очевидно, что студентов с таким номером нет в первой группе, где всего 12 человек, нет их и во второй группе (12 + 48). 88 и 89-й номера студентов находятся в третьей группе (12 + 48 + 56 = 116), следовательно, медианным значением является оценка «4».

Модальный интервал (т.е. содержащий моду) в случае интервального распределения с равными интервалами определяется по наибольшей частоте; с неравными интервалами - по наибольшей плотности, а определение моды требует проведения расчетов на основе следующих формул:

$$M_o = x_0 + i \cdot \frac{(f_{M_o} - f_{M_o-1})}{(f_{M_o} - f_{M_o-1}) + (f_{M_o} - f_{M_o+1})},$$

где x_0 - нижняя граница модального интервала;

i - величина модального интервала;

f_{M_o} - частота модального интервала;

f_{M_o-1} - частота интервала, предшествующего модальному;

f_{M_o+1} - частота интервала, следующего за модальным.

Медиана в случае интервального распределения вычисляется по формуле:

$$M_e = x_0 + i \frac{\frac{1}{2} \sum f_i - S_{M_e-1}}{f_{M_e}},$$

где x_0 - нижняя граница медианного интервала;

i - величина медианного интервала;

f_{M_e} - частота медианного интервала;

S_{Me-1} - накопленная частота интервала, предшествующего медианному.

Для установления медианного интервала необходимо определить накопленную частоту каждого последующего интервала до тех пор, пока она не превысит половины суммы накопленных частот.

Задание:

1. Используя данные об успеваемости студентов (практическая работа №2), выполните расчет моды и медианы.

2. Используя интервальный ряд распределения, полученный в работе №1 (датчик №3), выполните расчет моды и медианы. Постройте графики и определите моду и медиану графическим способом.

1.4 Практическая работа № 4

Выборочное наблюдение

Собственно-случайная выборка заключается в отборе единиц из генеральной совокупности в целом, без деления ее на группы, подгруппы или серии отдельных единиц. При этом единицы отбираются в случайном порядке, не зависящем ни от последовательности расположения единиц в совокупности, ни от значений их признаков.

После проведения отбора с использованием какого-либо алгоритма, реализующего принцип случайности, или на основе таблицы случайных чисел, необходимо определить границы генеральных характеристик. Для этого рассчитываются средняя и предельная ошибки выборки.

Средняя ошибка повторной собственно-случайной выборки определяется по формуле:

$$\mu_{\tilde{x}} = \frac{\sigma_{\tilde{x}}}{\sqrt{n}}.$$

С учетом выбранного уровня вероятности и соответствующего ему значения t предельная ошибка выборки составит:

$$\Delta_{\tilde{x}} = t\mu.$$

Тогда можно утверждать, что при заданной вероятности генеральная средняя будет находиться в следующих границах:

$$\tilde{x} - \Delta_{\tilde{x}} \leq \bar{x} \leq \tilde{x} + \Delta_{\tilde{x}}.$$

Для бесповторной выборки средняя ошибка вычисляется по формуле:

$$\mu = \sqrt{\frac{\sigma^2}{n} \left(1 - \frac{n}{N}\right)}.$$

Задание

Используя данные практической работы №2, выполните 30% случайный отбор студентов группы и определите границы генеральной

средней (успеваемости). Рассмотрите повторный и бесповторный отбор.

1.5 Практическая работа № 5

Статистические индексы

Индекс – обобщенный показатель, сконструированный специальным образом и применяемый для наблюдения за количественными характеристиками социальных объектов, явлений или событий

Формирования интегрального показателя осуществляется с помощью следующих этапов:

- нормирование индикаторов;
- расчет интегрального показателя.

Нормирование может осуществляться различными способами.

В методе эталонного значения в исходных данных по каждому показателю определяется максимальный элемент, который принимается за единицу. Затем все показатели (a_{ij}) делятся на максимальный элемент предприятия-эталона ($\max a_{ij}$). В результате создается матрица стандартизованных коэффициентов:

$$x_{ij} = \frac{a_{ij}}{\max a_{ij}}$$

Эталонное предприятие формируется обычно из совокупности однородных объектов, принадлежащих к одной отрасли. Однако это не исключает возможности выбора предприятия-эталона из совокупности предприятий, принадлежащих к различным отраслям деятельности, так как многие финансовые показатели сопоставимы и для разнородных субъектов хозяйствования.

Если с экономической стороны лучшим является минимальное значение показателя (например, затраты на рубль продукции), то надо изменить шкалу расчета так, чтобы наименьшему результату соответствовала наибольшая сумма показателя:

$$x_{ij} = \frac{\min a_{ij}}{a_{ij}}$$

Если задача решается с учетом разного веса показателей, полученные квадраты умножаются на величину соответствующих весовых коэффициентов (K), установленных экспертным путем, после чего результаты складываются по строкам. Таким образом, интегральный показатель вычисляется по формуле:

$$R_j = K_1 x_{j1}^2 + K_2 x_{j2}^2 + \dots + K_n x_{jn}^2.$$

R – интегральный показатель j -го объекта

K – коэффициент важности.

Задание

Выбрав 5 показателей деятельности региона и 7 регионов, составьте рейтинг регионов РФ методом эталонного значения. Данные нужно взять на сайте Федеральной службы государственной статистики (<http://www.gks.ru/>) (не ранее 2014 года).

1.6 Практическая работа 6

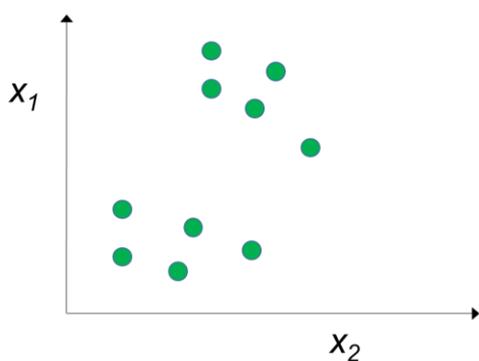
Метод k-средних

Пусть имеется набор данных: $x^{(1)}, x^{(2)}, x^{(3)} \dots x^{(n)}$ (рис.1.2 а), каждый элемент которого характеризуется двумя показателями: x_1, x_2 . Необходимо объединить данные в две группы (кластера). Для этого нужно выполнить следующие шаги:

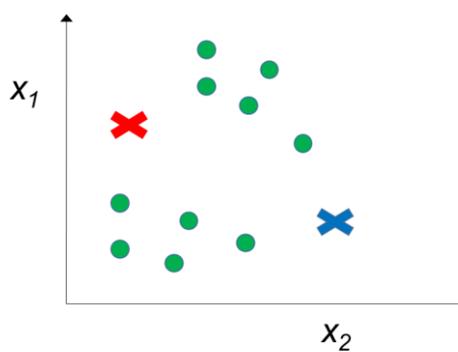
1. случайно сгенерировать центры кластеров (рис.1.2 б);
2. произвести обход всех элементов. В зависимости от того, к какому центру ближе точка, она относится к первому или второму кластеру (рис.1.2 в).

3. определяются новые центры кластеров как среднее значение всех элементов, относящихся к конкретному кластеру (рис.1.2 г). Возврат на шаг 2.

Данные шаги повторяются до тех пор, пока центры кластеров не будут значительно изменяться.



а)



б)

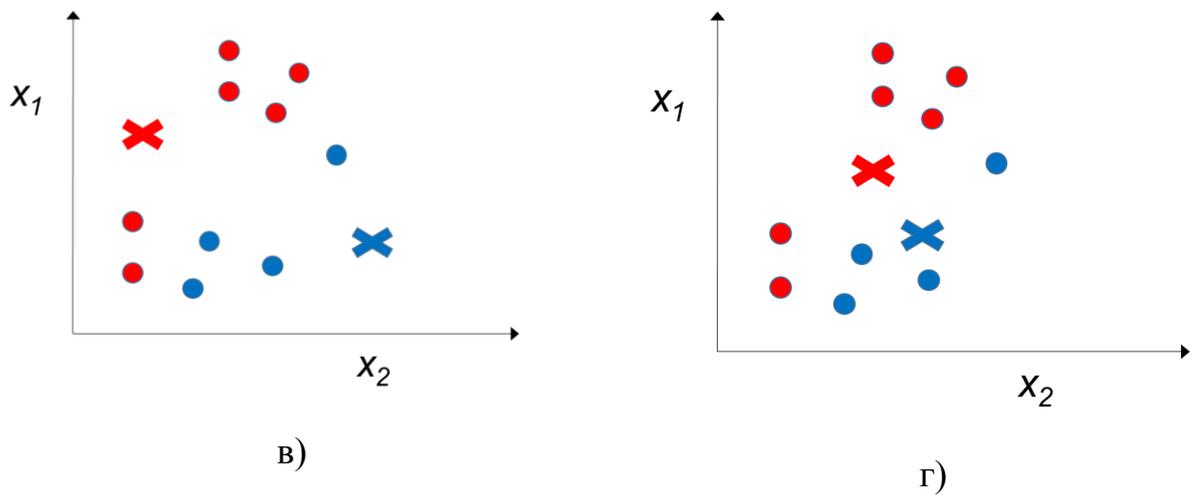


Рис.1.2 Шаги метода k-средних: а) исходный набор данных; б) определение центров кластеров; в) отнесение элементов к конкретному кластеру; г) определение новых центров и отнесение элементов к новым центрам

Таким образом, алгоритм метода можно представить в следующем виде:

Входные данные:

- K (число кластеров);
- Набор данных $x^{(1)}, x^{(2)}, x^{(3)} \dots x^{(n)}$;

Случайно генерируются K центров кластеров: m_1, \dots, m_k

Цикл {

for $i=1$ to n

$c^{(i)}$ =индекс (от 1 до K) центра кластера, ближайшего к $x^{(i)}$

for $j=1$ to K

m_j =average(среднее) точек закрепленных за кластером j

}

В качестве центров кластеров на начальном этапе можно выбрать наиболее удаленные друг от друга точки.

Задание

Выберите 7 объектов города Томска. Используя программу Дубль ГИС вычислите расстояние между ними. Данные представьте в виде матрицы. Напишите макрос осуществляющий разбиение объектов на кластеры методом k -средних. Рассмотрите случаи, когда число кластеров равно 2 и 3.

1.7 Практическая работа № 7

Расчет коэффициента кластеризации

Коэффициент кластеризации является показателем процентного соотношения соседних вершин связанных между собой (например, с помощью него можно определить, какой процент друзей конкретного человека в социальных сетях дружат между собой, рис.1.3). Формула расчета:

$$Cl(i) = \frac{k_{link}}{k},$$

где k -число пар соседних узлов.

k_{link} -число пар соседних узлов, связанных между собой.

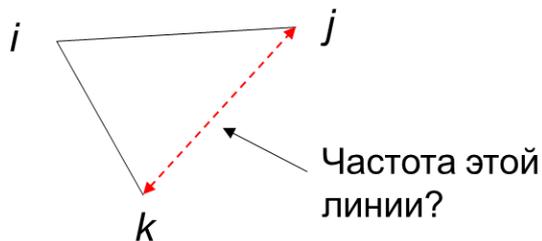


Рис.1.3 Кластеризация

Средняя кластеризация рассчитывается по формуле:

$$Cl_{avg} = \frac{\sum_i Cl(i)}{N}.$$

Например, коэффициент кластеризации вершины с номером пять (рис.1.4) будет равна 1/3. Узел «5» имеет три возможных пары соседей: 13, 12, 23. Из них только одна пара 13 связана между собой.

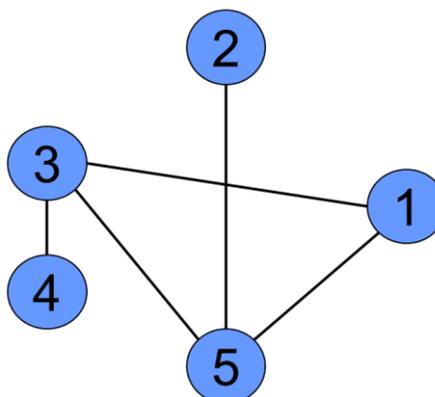


Рис.1.4 Пример графа

Задание

Используя данные своей страницы в социальных сетях, вычислите коэффициент кластеризации (для 15 друзей). Нарисуйте полученный граф.

2. Примеры выполнения заданий

1. Вычислите коэффициент корреляции Пирсона двух величин: успеваемость студента выбранной группы и пол студента. Выполните интерпретацию полученного результата. Для расчета коэффициента используйте данные официального сайта ТУСУР.

Для выполнения задания была использована информация об успеваемости группы 734 по дисциплине «физика» (на 03.02.2016). На рис. 2.1 представлены результаты вычисления коэффициента корреляции. Формулы вычисления ячеек следующие:

$$I3=(A24*E25-D25*C25)/(КОРЕНЬ(A24*F25-C25*C25)*КОРЕНЬ(A24*G25-D25*D25)).$$

Рассчитанное значение совпало со вычисленным с помощью стандартной функции: I4=КОРРЕЛ(C2:C24;D2:D24).

Близкое к нулю значение говорит об отсутствии связи между величинами «пол» и «успеваемость».

| | A | B | C | D | E | F | G | H | I | J |
|----|----|------------------|----------|-------|----|-----|-----|---|-------------------------------|----------|
| 1 | № | Студент | Оценка X | Пол Y | XУ | X^2 | Y^2 | | | |
| 2 | 1 | Боровков С.С. | 4 | 1 | 4 | 16 | 1 | | Коэфф. корреляции, r | |
| 3 | 2 | Бусыгин И.В. | 5 | 1 | 5 | 25 | 1 | | | -0,00395 |
| 4 | 3 | Ворожцов С.А. | 5 | 1 | 5 | 25 | 1 | | -0,003951119 | |
| 5 | 4 | Герасимов В.А. | 3 | 1 | 3 | 9 | 1 | | Связь слабая отрицательная | |
| 6 | 5 | Ерохов Е.В. | 3 | 1 | 3 | 9 | 1 | | | |
| 7 | 6 | Иванов А.С. | 3 | 1 | 3 | 9 | 1 | | | |
| 8 | 7 | Косаченко Т.С. | 4 | 0 | 0 | 16 | 0 | | | |
| 9 | 8 | Кравченко А.В. | 5 | 1 | 5 | 25 | 1 | | | |
| 10 | 9 | Кузнецова М.Г. | 4 | 0 | 0 | 16 | 0 | | | |
| 11 | 10 | Легкодимова А.С. | 2 | 0 | 0 | 4 | 0 | | | |
| 12 | 11 | Лозинг В.Р. | 4 | 0 | 0 | 16 | 0 | | | |
| 13 | 12 | Люгай Д.А. | 2 | 1 | 2 | 4 | 1 | | | |
| 14 | 13 | Мартынов Р.В. | 3 | 1 | 3 | 9 | 1 | | | |
| 15 | 14 | Машуков Н.А. | 5 | 1 | 5 | 25 | 1 | | | |
| 16 | 15 | Мухин А.Э. | 4 | 1 | 4 | 16 | 1 | | | |
| 17 | 16 | Мякишева Э.С. | 5 | 0 | 0 | 25 | 0 | | | |
| 18 | 17 | Оюн К.М. | 2 | 1 | 2 | 4 | 1 | | | |
| 19 | 18 | Парников Г.Г. | 2 | 1 | 2 | 4 | 1 | | | |
| 20 | 19 | Пехова А.В. | 4 | 0 | 0 | 16 | 0 | | | |
| 21 | 20 | Сапунов А.В. | 5 | 1 | 5 | 25 | 1 | | | |
| 22 | 21 | Федурин А.С. | 5 | 1 | 5 | 25 | 1 | | | |
| 23 | 22 | Черемных А.А. | 4 | 1 | 4 | 16 | 1 | | | |
| 24 | 23 | Шивцов Д.Д. | 5 | 1 | 5 | 25 | 1 | | | |
| 25 | | Итого: | 88 | 17 | 65 | 364 | 17 | | | |
| 26 | | | | | | | | | | |

Рис.2.1 Результаты расчетов коэффициента корреляции

2. Используя интервальный ряд распределения (табл.2.1), выполните расчет моды и медианы. Постройте графики и определите моду и медиану графическим способом.

Таблица 2.1 Интервальный ряд распределения

| Номер интервала | Нижняя граница | Верхняя граница |
|-----------------|----------------|-----------------|
| 1 | -0,219 | 1,034 |
| 2 | 1,034 | 2,288 |
| 3 | 2,288 | 3,541 |
| 4 | 3,541 | 4,794 |
| 5 | 4,794 | 6,047 |
| 6 | 6,047 | 7,301 |
| 7 | 7,301 | 8,554 |
| 8 | 8,554 | 9,807 |

Вычисление моды и медианы по формулам представлено на рис. 2.2. Модальным является интервал 3,541-4,794, т.к. его частота наибольшая. Медианным является этот же интервал 3,541-4,794, т.к. его накопленная частота первая превышает половину всей суммы частот.

Здесь формулы расчета ячеек следующие:

$$I8=I2+I3*((I4-I5)/((I4-I5)+(I4-I6)))$$

$$I16=I11+I12*(((I15/2)-I14)/(I13))$$

| | A | B | C | D | E | F | G | H | I |
|----|--------|----------|-------|--------|---------------------|---|---|---------|----------|
| 1 | Группа | Диапазон | | Кол-во | Накопленные частоты | | | | |
| 2 | 1 | -0,219 | 1,034 | 4 | 4 | | | x0 = | 3,541 |
| 3 | 2 | 1,034 | 2,288 | 7 | 11 | | | l = | 1,253 |
| 4 | 3 | 2,288 | 3,541 | 15 | 26 | | | fMO = | 26 |
| 5 | 4 | 3,541 | 4,794 | 26 | 52 | | | fMO-1 = | 15 |
| 6 | 5 | 4,794 | 6,047 | 17 | 69 | | | fMO+1 = | 17 |
| 7 | 6 | 6,047 | 7,301 | 16 | 85 | | | | |
| 8 | 7 | 7,301 | 8,554 | 13 | 98 | | | Mo = | 4,23015 |
| 9 | 8 | 8,554 | 9,807 | 2 | 100 | | | | |
| 10 | Всего | | | 100 | | | | | |
| 11 | | | | | | | | x0 = | 3,541 |
| 12 | | | | | | | | l = | 1,253 |
| 13 | | 26 | | | | | | fMe = | 26 |
| 14 | | 1,253 | | | | | | SMe-1 = | 26 |
| 15 | | | | | | | | fi = | 100 |
| 16 | | | | | | | | Me = | 4,697615 |

Рис.2.2 Результаты расчетов моды и медианы

Графическое определение моды и медианы представлено на рис. 2.3.

Для графического определения медианы нужно построить график кумуляты – накопленных частот. Далее провести линию, соответствующую половине частот, до пересечения с графиком и опустить из полученной точки перпендикуляр. Для графического определения моды нужно найти интервал с наибольшей частотой. Соединить левый угол столбца с левым углом прямоугольника расположенного справа от него, а правый угол – с правым углом прямоугольника расположенного слева от него. Результаты, полученные с помощью формул и с помощью графика, совпали.

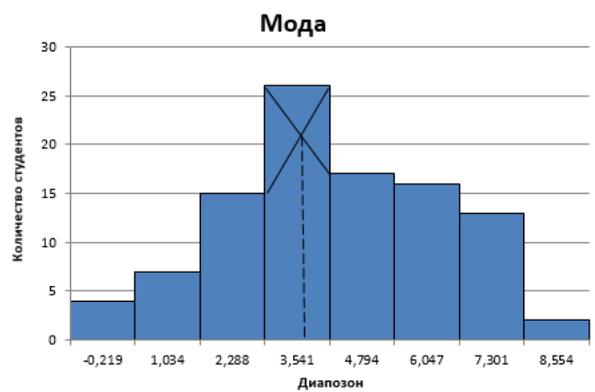
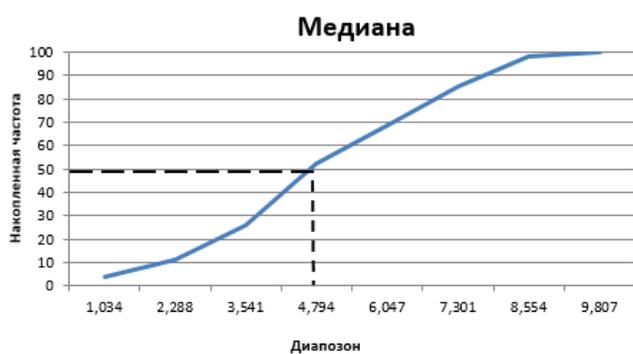


Рис.2.3 Графическое определение моды и медианы

3. Выбрав 5 показателей деятельности региона и 7 регионов, составьте рейтинг регионов РФ методом эталонного значения. Данные нужно взять на сайте Федеральной службы государственной статистики (<http://www.gks.ru/>) (не ранее 2014 года).

На рис. 2.4 представлена исходная информация по регионам. Были рассмотрены следующие регионы:

- Ивановская область;
- Московская область;
- Республика Ингушетия;
- Нижегородская область;
- Республика Хакассия;
- Новосибирская область;
- Хабаровский край.

Для оценки регионов были использованы показатели (в скобках указано значение коэффициента важности):

- Численность населения (0,4);
- Численность безработных (0,7);
- Число больничных организаций (0,8);
- Число зарегистрированных преступлений (0,6);
- Валовый региональный продукт (0,6).

| | A | B | C | D | E | F | G | H |
|---|--|--------------------|--------------------|----------------------|-----------------------|--------------------|-----------------------|------------------|
| 1 | Регионы | | | | | | | |
| 2 | | Ивановская область | Московская область | Республика Ингушетия | Нижегородская область | Республика Хакасия | Новосибирская область | Хабаровский край |
| 3 | Численность населения, тыс. чел | 1036,9 | 7231,1 | 463,9 | 3270,2 | 535,8 | 2746,8 | 1338,3 |
| 4 | Численность безработных, тыс. чел | 22,9 | 104,9 | 63,2 | 75,2 | 16,1 | 73,6 | 43,9 |
| 5 | Число больничных организаций | 50 | 193 | 13 | 133 | 28 | 107 | 69 |
| 6 | Число зарегистрированных преступлений: всего, тыс. | 14,9 | 91,1 | 1,6 | 38,6 | 13,9 | 46,5 | 29,6 |
| 7 | Валовой региональный продукт: всего, млн. руб | 157735 | 2551284 | 45171 | 925833 | 143534 | 821415 | 473695 |

Рис. 2.4 Данные по регионам

Для нормирования значений был использован метод эталонного значения. Индикаторы, для которых лучшим значением является максимальное: численность населения, число больничных организаций, валовой региональный продукт. Индикаторы, для которых лучшим значением является минимальное: численность безработных, число зарегистрированных преступлений. Рассчитанное значение индикатора представлено на рис. 2.5. Интегральный показатель рассчитывается как сумма произведений коэффициента важности и квадрата нормированного значения:

$$B18 = \$J\$13 * B13 * B13 + \$J\$14 * B14 * B14 + \$J\$15 * B15 * B15 + \$J\$16 * B16 * B16 + \$J\$17 * B17 * B17$$

Согласно полученным данным первое место по выбранным показателям занимает Московская область, а последнее – Хабаровский край.

| | Ивановская область | Московская область | Республика Ингушетия | Нижегородская область | Республика Хакасия | Новосибирская область | Хабаровский край | Максимум и минимум | Коэф. важности | |
|----|---|--------------------|----------------------|-----------------------|--------------------|-----------------------|------------------|--------------------|----------------|-----|
| 12 | | | | | | | | | | |
| 13 | Численность населения, тыс. чел | 0,14 | 1,00 | 0,06 | 0,45 | 0,07 | 0,38 | 0,19 | 7231,1 | 0,4 |
| 14 | Численность безработных, тыс. чел | 0,70 | 0,15 | 0,25 | 0,21 | 1,00 | 0,22 | 0,37 | 16,1 | 0,7 |
| 15 | Число больничных организаций | 0,26 | 1,00 | 0,07 | 0,69 | 0,15 | 0,55 | 0,36 | 193 | 0,8 |
| 16 | Число зарегистрированных преступлений: всего | 0,11 | 0,02 | 1,00 | 0,04 | 0,12 | 0,03 | 0,05 | 1,6 | 0,6 |
| 17 | Валовой региональный продукт: всего, млн. руб | 0,06 | 1,00 | 0,02 | 0,36 | 0,06 | 0,32 | 0,19 | 2551284 | 0,6 |
| 18 | Интегральный показатель | 0,42 | 1,82 | 0,65 | 0,57 | 0,73 | 0,40 | 0,23 | | |
| 19 | Место | 5 | 1 | 3 | 4 | 2 | 6 | 7 | | |

Рис.2.5 Рейтинг регионов

Список литературы

1. Шмойлова Р. А., Минашкин В. Г., Садовникова Н. А., Шувалов Е. Б. Теория статистики. – М.: Финансы и статистика, 2004. – 656 с.
2. Лузина Л. И. Статистика: Учебное пособие. – Томск: ТМЦДО, 2009. – 141 с.
3. Светульников С. Г., Светульников И. С. Методы социально-экономического прогнозирования: Учебник для вузов. – СПб.: Изд-во СПбГУЭФ, 2009. – 147 с.
4. Иванов О. В. Статистика. Учебный курс для социологов и менеджеров. Часть 1. Описательная статистика. Теоретико-вероятностные основания статистического вывода. – М. 2005. – 187 с.
5. Jackson M. O. Social and economic networks: models and analysis. [Электронный ресурс]. – Режим доступа - <https://www.coursera.org/learn/social-economic-networks/home/welcome>.
6. Andrew Ng. Machine learning. [Электронный ресурс]. – Режим доступа - <https://www.coursera.org/learn/machine-learning/home/welcome>
7. Online social networks and media. [Электронный ресурс]. – Режим доступа - <http://www.cs.uoi.gr/~tsap/teaching/2013-cs-114/slides/L14-lecture5.pdf>.
8. Грибанова Е.Б. Информационная система рейтинговой оценки объектов экономики / Е.Б. Грибанова, А.Н. Алимханова, П.Э. Тугароол // Доклады Том. гос. ун-та систем управления и радиоэлектроники. – 2016. – № 2(19). – С. 51–55.
9. Карминский А.М., Полозов А.А. Энциклопедия рейтингов: экономика, общество, спорт. – М.: ИД «Форум», Инфра-М, 2016. – 36 с.