Федеральное агентство связи

Государственное федеральное образовательное учреждение высшего профессионального образования

ПОВОЛЖСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ТЕЛЕКОММУНИКАЦИЙ И ИНФОРМАТИКИ

ЭЛЕКТРОННАЯ БИБЛИОТЕЧНАЯ СИСТЕМА

Самара

ФЕДЕРАЛЬНОЕ АГЕНСТВО СВЯЗИ

Государственное образовательное учреждение высшего профессионального образования

ПОВОЛЖСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ТЕЛЕКОММУНИКАЦИЙ И ИНФОРМАТИКИ

Кафедра высшей математики

Богданова М.Г., Старожилова О.В.

Методические указания к практическим занятиям «Основные понятия статистики и выборочный метод»

ПО УЧЕБНОЙ ДИСЦИПЛИНЕ
Теория вероятностей и математическая статистика

Самара, 2011

УДК 519.2

Богданова М.Г., Старожилова О.В. Методические указания к практическим занятиям «Основные понятия статистики и выборочный метод».- Самара: ГОУВПО ПГУТИ, 2011-26с.

Методические указания к практическим занятиям «Основные понятия статистики и выборочный метод» помогут студентам проверить теоретическое освоение курса «Теория вероятностей и математическая статистика», содержат теоретические сведения основ статистического распределения, варианты заданий .

Рецензент:
Асташкин С.В. – д.ф.м.н., проф., зав.кафедрой Самарского государственного университета
Государственное образовательное учреждение высшего профессионального образования

Поволжский государственный университет телекоммуникаций и информатики

ОСНОВНЫЕ ПОНЯТИЯ СТАТИСТИКИ И ВЫБОРОЧНЫЙ МЕТОД

Цель работы: приобретение навыков группирования и обработки первичной статистической информации в интерактивной среде Excel.

Теоретическая часть

Математическая статистика изучает различные методы сбора, обработки результатов многократно повторяемых случайных событий. Для процесса построения и применения моделей характерно, чем больше данных, тем точнее, адекватнее модель.

Двумя основными задачами математической статистики являются:

- - определение способов сбора и группировки этих статистических данных;
- - разработка методов анализа полученных данных в зависимости от целей исследования, к которым относятся:

Для решения этих задач необходимо выбрать из большой совокупности однородных объектов ограниченное количество объектов, по результатам изучения которых можно сделать прогноз относительно исследуемого признака этих объектов.

Совокупность объектов исследуется относительно некоторого качественного или количественного признака, характеризующего эти объекты. Определим основные понятия математической статистики.

➢ **Определение** *Генеральная совокупность* – полное множество некоторых единиц, которые обладают теми или иными общими свойствами, существенными для их характеристики.

➢ **Определение** *Выборочная совокупность,* или просто *выборка* - совокупность случайно отобранных объектов.

➢ **Определение** *Объем* генеральной совокупности N и объем выборки n – число объектов в рассматриваемой совокупности.

Сущность выборочного метода состоит в том, чтобы по некоторой части генеральной совокупности выносить суждение о свойствах в целом.

Исследуемый признак генеральной совокупности является *дискретным,* если он принимает отдельные, изолированные возможные значения с определёнными вероятностями. Исследуемый признак генеральной совокупности является *непрерывным,* если он может принимать все значения из некоторого конечного или бесконечного промежутка.

Случайная выборка строится таким образом, что

- каждый объект генеральной совокупности имеет одинаковую вероятность быть отобранным;
- объекты выбирают независимо друг от друга.
- случайность гарантирует надежность.

Виды выборки

Пусть случайная величина X принимает в выборке значение x_1 - n_1 раз, x_2 - n_2 раз, …, x_n - n_n раз, и

$$\sum_{i=1}^{k} n_k = n, \qquad \text{где } n \text{ – объем выборки.}$$

➢ **Определение** *Варианты* - наблюдаемые значения x_1, x_2,…, x_n случайной величины X, принимаемые в выборке n_1 раз, n_2 раз, …, n_n раз.

➢ **Определение** *Частоты* $n_1, n_2, \ldots, n_k$ –. величины, показывающие, сколько раз встречается то или иное значение признака

➢ **Определение** *Относительные частоты* - отношение частот к объему $w_i = \dfrac{n_i}{n}$.

➢ **Определение** *Вариационный ряд* - последовательность вариант, записанных в порядке возрастания.

➢ **Определение** С*татистический ряд*- последовательность частот или относительных частот, записанных в порядке возрастания.

Различные значения признака X называются *вариантами.*

x_i	x_1	x_2	…	x_k
n_i	n_1	n_2	…	x_k
w_i	w_1	w_2	…	w_k

✓ **Замечание**: Под распределениями понимают в математической статистике — соответствие между наблюдаемыми вариантами и их частотами.

➢ **Определение** Расположение, упорядочение вариантов в порядке возрастания (убывания) называется *ранжированием* вариантов ряда.

При составлении выборки можно поступать двояко: после того, как объект отобран и над ним произведено наблюдение, он может быть возвращен, либо не возвращен в генеральную совокупность.

В соответствии со сказанным, выборки подразделяют на повторные и бесповторные

➢ **Определение** *Повторная выборка* – каждый отобранный объект перед выбором следующего возвращается в генеральную совокупность;

➢ **Определение** *Бесповторная выборка* – отобранный объект в генеральную совокупность не возвращается.

На практике обычно пользуются бесповторным случайным отбором.

✓ **Замечание** Для того, чтобы по исследованию выборки можно было сделать выводы о поведении интересующего признака генеральной совокупности, нужно, чтобы выборка правильно представляла пропорции генеральной совокупности, то есть была *репрезентативной* (представительной).

Выборка будет репрезентативной, если её осуществить случайно, когда все объекты генеральной совокупности имеют одинаковую вероятность быть отображенными и отбор одного объекта не влияет на вероятность отбора другого объекта совокупности.

Учитывая закон больших чисел, можно утверждать, что это условие выполняется, *если каждый объект выбран случайно, причем для любого объекта вероятность попасть в выборку одинакова.*

Способы отбора

На практике применяются различные способы отбора. Принципиально эти способы можно подразделить на два вида:

1. Отбор, не требующий расчленения генеральной совокупности на части, сюда относятся:
   - простой случайный бесповторный отбор;
   - простой случайный повторный отбор.
2. Отбор, при котором генеральная совокупность разбивается на части, сюда относятся:
   - типический отбор;
   - механический отбор;
   - серийный отбор.

➢ **Определение** *Простой случайный отбор* - отбор, при котором объекты извлекают по одному из всей генеральной совокупности.

➢ **Определение** *Типический отбор* - отбор, при котором объекты отбираются не из всей генеральной совокупности, а из каждой ее «типической» части.

Типическим отбором пользуются тогда, когда обследуемый признак заметно колеблется в различных типических частях генеральной совокупности.

➢ **Определение** *Механический отбор* - отбор, при котором генеральная совокупность «механически» делится на столько групп, сколько объектов должно войти в выборку, и из каждой группы отбирается один объект.

➢ **Определение** *Серийный отбор* - отбор, при котором объекты отбирают из генеральной совокупности не по одному, а «сериями», которые подвергаются сплошному обследованию.

На практике часто применяется комбинированный отбор, при котором сочетаются указанные выше способы.

Табличное представление статистических данных

После того, как данные собраны, выполняется их обработка, при этом необходимо обеспечить наглядность представления данных, позволяющую получить какие-то первоначальные представления об их закономерности. Эта наглядность достигаются путем построения таблиц и графиков

Если исследуется некоторый непрерывный признак, то вариационный ряд может состоять из очень большого количества чисел. В этом случае удобнее использовать *группированную выборку*.

Для ее получения интервал, в котором заключены все наблюдаемые значения признака, разбивают на несколько равных частичных интервалов длиной h, а затем находят для каждого частичного интервала n_i – сумму частот вариант, попавших в i-й интервал. Составленная по этим результатам таблица называется *группированным статистическим рядом*:

➢ **Определение** Числа, показывающие, сколько раз встречаются варианты из данного интервала, называются *частотами*, а отношение их к общему числу наблюдений – *относительными частотами*.

Частоты и относительные частоты называют *весами.*

Номера интервалов	1	2	…	k
Границы интервалов	$(a, a + h)$	$(a + h, a + 2h)$	…	$(b - h, b)$
Сумма частот вариант, попавших в интервал	n_1	n_2	…	n_k

➢ **Определение** *Вариационный ряд* – ранжированный ряд вариантов с соответствующими весами.

Частота признака m_x - величина, показывающая, сколько раз встречается то или иное значение признака. Относительная частота w_x - отношение частоты m_x к общему объему выборки n:

$$\omega_x = \frac{m_x}{n} = \frac{m_x}{\sum m_x}$$

Наряду с понятиями частоты и относительной частоты, в математической статистике рассматриваются понятия *накопленной частоты* $m_x^{нак}$ и *накопленной относительной частоты* $\omega_x^{нак}$ которые показывают, во скольких наблюдениях признак принял значения не больше заданного значения x:

$$m_x^{нак} = \sum m_x, \qquad \omega_x^{нак} = \frac{m_x^{нак}}{n}$$

В случае непрерывной случайной величины рассматривают не дискретные значения признака, а их значения в пределах определенного интервала. В качестве частоты при таком подходе выступает количество случаев, в которых признак принял значения, входящие в некоторый интервал.

Такую величину называют *интервальной частотой* и обозначают m_h (соответственно рассматривается также и *интервальная относительная частота* w_h*)*.

Полученный таким образом ряд называют *интервальным вариационным рядом***.**

Для построения интервального ряда необходимо установить величину интервала h. Она должна быть такой, чтобы ряд был не слишком громоздким и не отвлекал внимание на частности, и, в то же время, обеспечивал выявление характерных черт и закономерностей исследуемых явлений.

➢ **Определение** *Вариационный размах* – разность между наибольшим и наименьшим вариантами ряда

$$R = x_{\max} - x_{\min}$$

➢ **Определение** *Группировка* - разбиение вариантов на различные интервалы.

Для определения величины интервала h можно использовать *формулу Стэрджесса*:

$$h = \frac{R_B}{1 + 3.3221 lg(n)}$$

где R_B - вариационный размах и является мерой разброса данных; $x_{\max}$ и $x_{\min}$ - соответственно наибольшее и наименьшее значение признака в выборке.

Ширина интервала

$$k = \frac{x_{\max} - x_{\min}}{h}.$$

Когда величина интервала h выбрана, строят шкалу интервалов. При этом за верхнюю границу первого интервала принимают обычно величину

$$a_1 = x_{min} + h/2$$

а верхняя граница каждого последующего интервала определяется добавлением к верхней границе предыдущего значения интервала h

$$a_j = a_{j-1} + h, \quad (j = 2,3...)$$

до тех пор, пока начало очередного интервала не окажется больше $x_{\max}$.

Затем все значения признака, входящие в выборку, распределяются между соответствующими интервалами, и строится интервальный вариационный ряд.

Графическое представление статистических данных

Наиболее часто используют следующие виды графического представления характеристик выборки: полигон, гистограмма и кумулятивная кривая. Гистограмма и полигон позволяют выявить преобладающие значения признака и характер распределения частот и относительных частот.

➢ **Определение** *Полигон* - ломаная линия с координатами (x_i, m_x) где x_i откладываются на оси абсцисс, а m_x, – на оси ординат.

Если на оси ординат откладывать не абсолютные, а относительные (w_i) частоты, то получим *полигон относительных частот.*

Полигон служит обычно для представления дискретного вариационного ряда. В системе координат (x, m_x) строятся точки, соответствующие значениям частот или относительных частот ряда, а затем эти точки соединяются прямыми линиями.

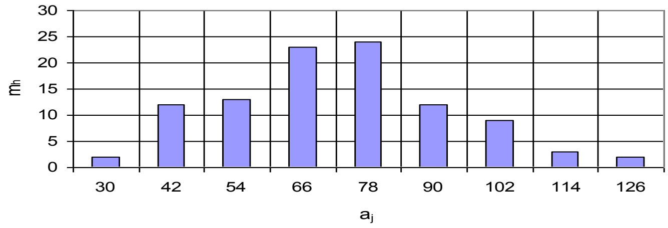
Для непрерывного признака графической иллюстрацией служит *гистограмма*, то есть ступенчатая фигура, состоящая из прямоугольников, основаниями которых служат частичные интервалы длиной h, а высотами – отрезки длиной n_i/h (гистограмма частот) или w_i/h (гистограмма относительных частот). В первом случае площадь гистограммы равна объему выборки, во втором – единице

➢ **Определение** *Гистограмма* – прямоугольники, с основаниями, равными интервалам значений признака и высотами, равными частотам.

Полигон (гистограмма) аналогичны кривой распределения, эмпирическая функция распределения – функции распределения случайной величины.

Гистограмма — это диаграмма, используемая, как правило, для представления интервального вариационного ряда.

Наиболее существенное отличие от полигона в том, что частота и относительная частота отображаются не точкой, а прямой, параллельной оси абсцисс на всем интервале.



Любой вариационный ряд можно изобразить графически в виде кривой накопленных частот — *кумуляты*.

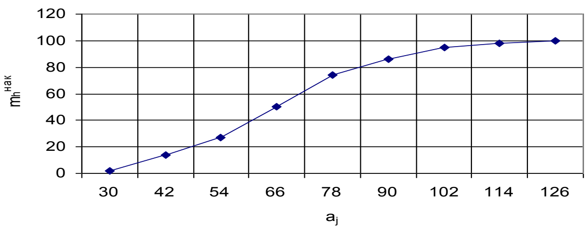
На оси абсцисс откладываются либо варианты, либо границы интервалов. На оси ординат — накопленная частота. Получают точки при пересечении каждой пары абсциссы и ординаты, которые соединяют плавной кривой.

➢ **Определение** *Кумулятивная кривая* (кривая сумм) — ломаная, составленная по последовательно суммированным, т.е. накопленным частотам или относительным частотам.

При построении кумулятивной кривой дискретного признака на ось абсцисс наносятся значения признака, а ординатами служат нарастающие итоги частот. Соединением вершин ординат прямыми линиями получают кумуляту.

При построении кумуляты интервального признака на ось абсцисс откладываются границы интервалов и верхним значениям присваивают накопленные частоты. Кумулятивную кривую называют полигоном накопленных частот.

Кумулятивная кривая строится для накопленных частот или накопленных относительных частот, причем по оси ординат откладывают верхнюю границу интервала соответствующего интервального ряда, так что последняя точка кумулятивной кривой всегда отвечает либо количеству наблюдений в выборке, либо равна единице.



Практическая часть

Задание

- Проранжировать первичный ряд данных,
- определить частоты, найти абсолютную и относительные плотности распределения,
- перегруппировать данные для сопоставления и анализа двух рядов,
- графически изобразить кривые (плотности) распределения рядов,
- представить данные в виде полигона частот, гистограмм, кумулятивных кривых по известным накопленным частотам.

Условие Имеются разрозненные данные по рентабельности активов банков с доходами от 50 до 100 млн. долл.:

1,51; 0,85; 1,37; 1,62; 0,80; 2,0; 1,49; 1,58; 1,75; 1,24; 1,28; 1,04; 1,98; 1,15; 1,66; 1,33; 1,73; 1,13; 1,36; 1,28.

Сравнить полученный сгруппированный ряд с известным интервальным рядом распределения по уровню рентабельности активов банков с доходами от 100 до 300 млн. долл.

Таблица. 1.

Группы банков с доходами от 100 до 300 млн. долл.	
Рентабельность активов	Количество банков в % (частости)
0,6 – 0,8	10
0,8 – 1,0	30
1,0 – 1,1	10
1,1 – 1,2	15
1,2 – 1,4	20
1,4 – 1,8	10
1,8 – 2,0	5
ИТОГО:	**100**

Выполнение задания. 1.

В Excel заполняется столбец исходных данных рис. 1.

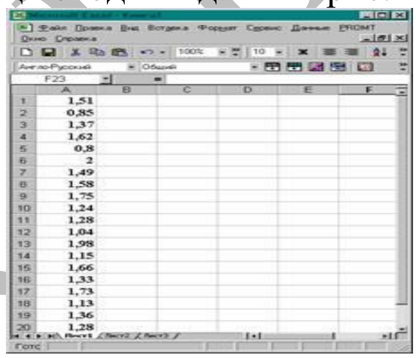


Рис. 1.

2. Выполняется сортировка столбца *A* - первичного ряда в порядке возрастания. В результате получен новый интервальный ранжированный ряд рис.2.

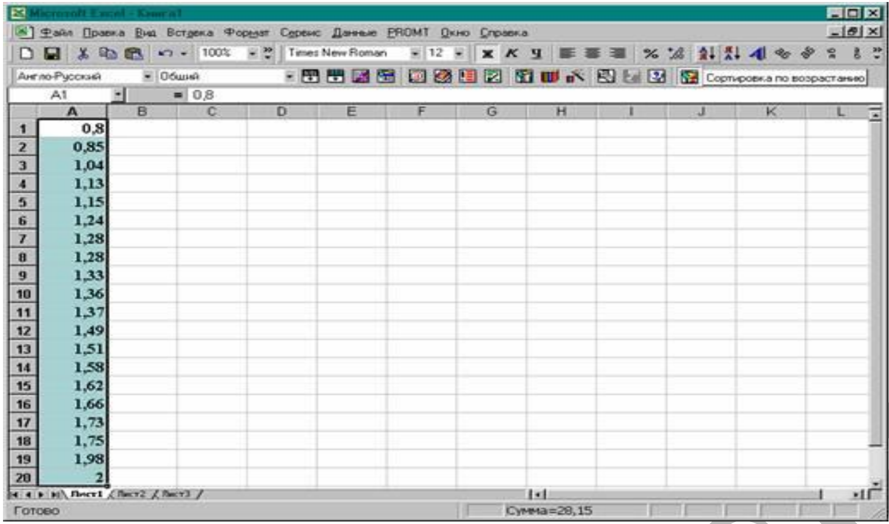


Рис. 2.

3. Определяются частоты нового ряда.

Для этого используется данные об объеме совокупности исследуемых банков $N = 20$.

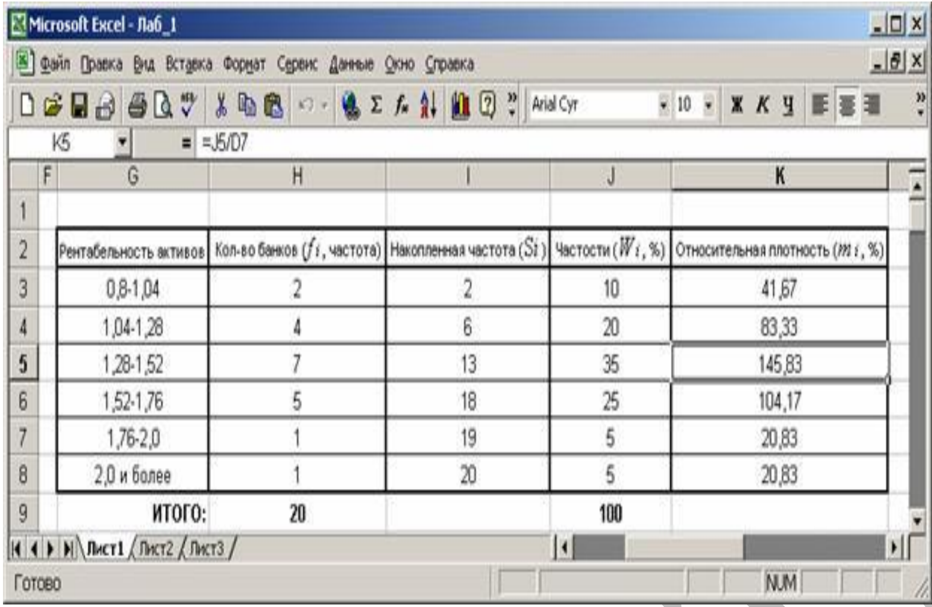
Дискретный вариационный ряд разбивается на интервалы, число которых подсчитывается по формуле Стержесса

$$k = \quad 1 + 3{,}322 \lg N \quad , \qquad (1)$$

в которой квадратные скобки означают округление числа 5,32 , тогда $k = 5$. Длина частичного интервала определяется:
$x_{\max} = 2{,}0$, $x_{\min} = 0{,}8$, $h = 0{,}24$. Тогда границы интервалов будут такими:

x0=	$x_{\min} = 0{,}8$;
x1=	$x_{\min} + h = 1{,}04$;
x2=	$x_{\min} + 2h = 1{,}28$;
x3=	$x_{\min} + 3h = 1{,}52$;
x4=	$x_{\min} + 4h = 1{,}76$;
x5=	$x_{\min} + 5h = 2$.

Подсчитывается количество банков принадлежащих каждому из интервалов.
Вычисляется накопленная частота и процентное отношение частоты к общему объему всей совокупности $N = 20$ или частость.
Для сопоставления полученных данных интервального вариационного ряда с данными другого вариационного ряда с неравными интервалами необходимо рассчитать относительную плотность распределения



Рентабельность активов	Кол-во банков (f_i, частота)	Накопленная частота (S_i)	Частости (W_i, %)	Относительная плотность (m_i, %)
0,8-1,04	2	2	10	41,67
1,04-1,28	4	6	20	83,33
1,28-1,52	7	13	35	145,83
1,52-1,76	5	18	25	104,17
1,76-2,0	1	19	5	20,83
2,0 и более	1	20	5	20,83
ИТОГО:	20		100	

Рис. 3.

4. Необходимо перегруппировать данные исследуемого интервального

Группы банков с доходами от 100 до 300 млн. долл.		Группы банков с доходами от 50 до 100 млн. долл.
Рентабельность активов	Количество банков в % (частости)	Количество банков в % (частости)
0,6 – 0,8	10	-
0,8 – 1,0	30	8,33 = 0,2m_1
1,0 – 1,1	10	6,67
1,1 – 1,2	15	8,33
1,2 – 1,4	20	24,17
1,4 – 1,8	10	43,33 =(1,52-1,4)m_3+0,24 m_4+(1,8-1,76) m_5
1,8 – 2,0	5	4,17
2,0 и более	-	5,00
ИТОГО:	**100**	**100,00**

вариационного ряда для сопоставления и анализа двух рядов.

Расчетная схема представлена ниже.

Необходимо вычислить новые величины интервалов, используя заготовленный рисунок 4:

$$h_{01} = y_0 - x_0\,,\ h_{11} = x_1 - y_0\,, h_{12} = y_1 - x_1\,, h_{21} = x_2 - y_1\,, h_{22} = x_3 - x_2\,,$$

$$h_{23} = x_4 - x_3\,, h_{24} = y_2 - x_4\,, h_{31} = x_5 - y_2\,, h_{32} = x_6 - x_5\,,$$

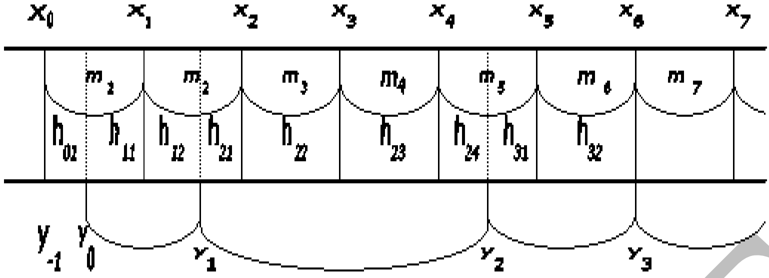


Рис. 4.

Здесь – узловые точки x_i рентабельности активов, отстоящие друг от друга на один и тот же равный шаг h , y_i – узловые точки интервального ряда рис. 1.

На промежутках $[y_{-1},\ y_0]$ U $[y_0,\ y_1]$ U $[y_1,\ y_2]$ U $[y_2,\ y_3]$ частости W_i перераспределяются следующим образом:

$[y_{-1}, y_0]$ — $W_0 = h_{01}m_1$,

$[y_0, y_1]$ — $W_1 = h_{11}m_1 + h_{12}m_2$,

$[y_1, y_2]$ — $W_2 = h_{21}m_2 + h_{22}m_3 + h_{23}m_4 + h_{24}m_5$,

$[y_2, y_3]$ — $W_3 = h_{31}m_5 + h_{32}m_6$,

полученными новыми значениями W_i заполняется третий столбец таблицы 2.

5. Графическое представление кривой (ненормированной плотности) распределения исходного ряда рис. 3.

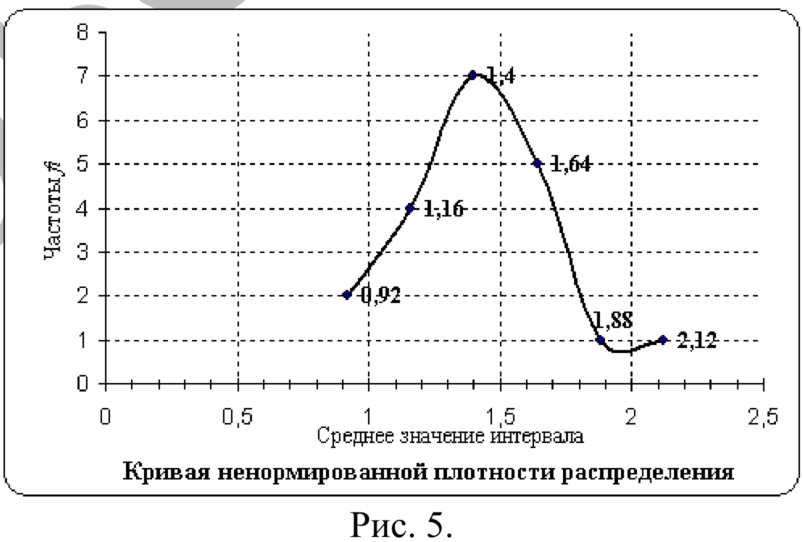


Рис. 5.

6. Полигон частот рис. 6.

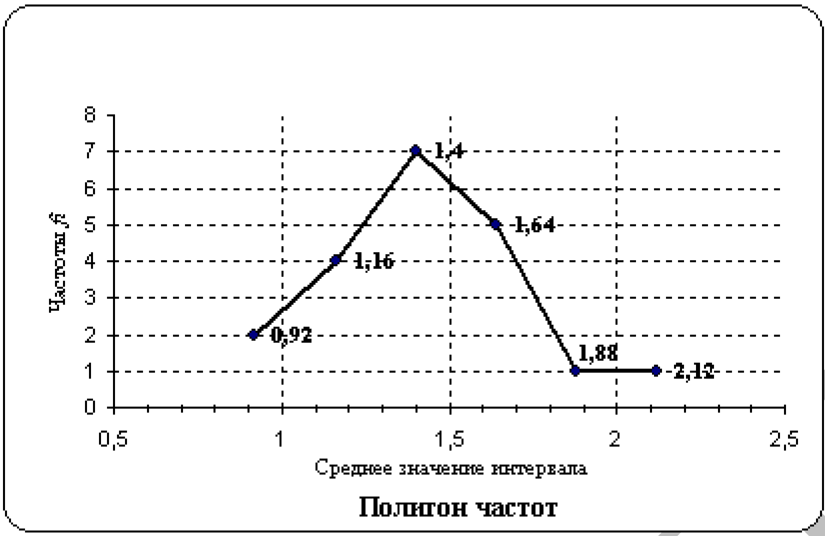


Рис. 6.

7. Гистограмма рис. 7.

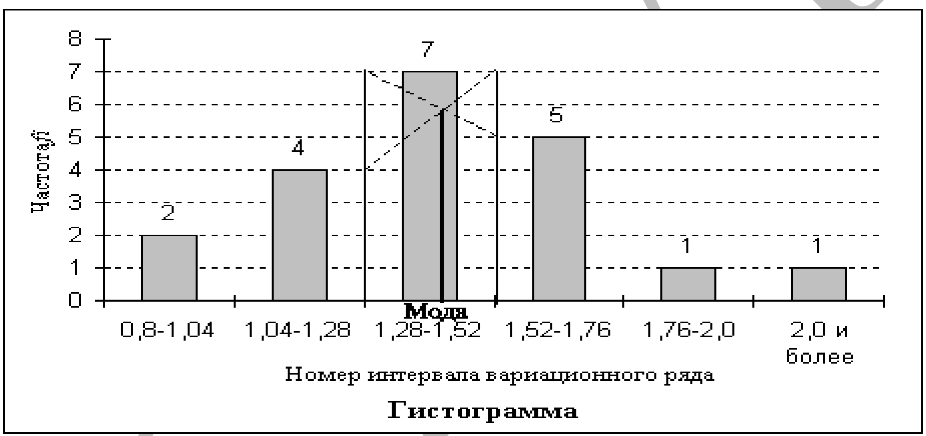


Рис. 7.

8. Кумулятивная кривая накопленных частот.

Рис. 8.

9. Сравнительная оценка частот интервальных вариационных рядов по уровню рентабельности активов банков с доходами от 100 до 300 млн. долл. и от 50 до 100 млн. долл. (табл. 2).

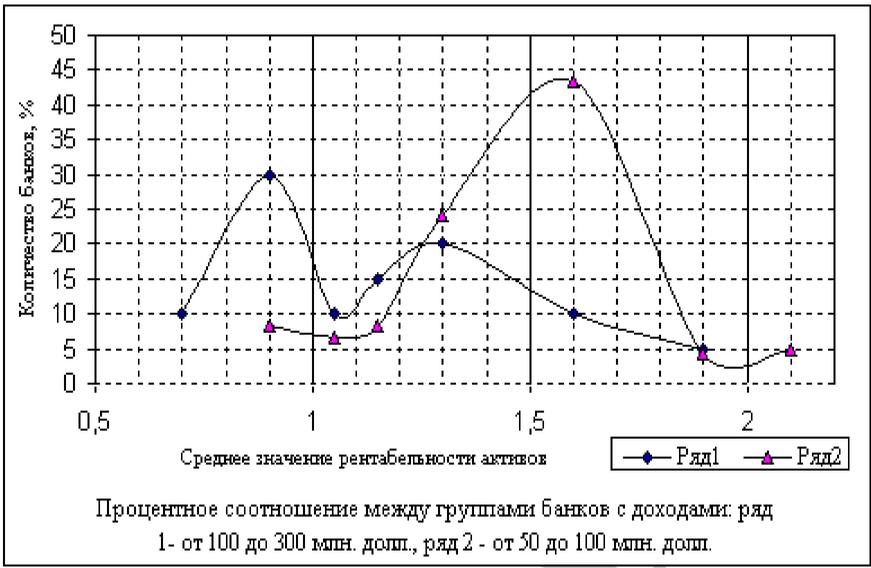


Рис. 9.

Выводы. Количество банков с доходами от 100 до 300 млн. долл. до уровня рентабельности 1,25 больше, чем банков с доходами от 50 до 100 млн. долл. После уровня рентабельности 1,25 банков с доходами от 50 до 100 млн. долл. значительно больше, чем банков с доходами от 100 до 300 млн. долл. Экстремальные значения при рентабельности активов 0,9 имеют 30 % банков с доходами от 100 до 300 млн. долл. и при рентабельности активов 1,6 имеют 43,33 % банков с доходами от 50 до 100 млн. долл.

Представленная графическая зависимость позволяет сделать вывод о том, что наиболее рентабельными являются банки ряда 2 (рис. 9).

Варианты заданий. Варианты указаны римскими цифрами. Сгруппированный ряд сравнивать с рядом заданным в табл. 1.

Таблица 3.

Вар.	I	II	III	IV	V	VI	VII	VIII	IX	X
1	0,52	0,65	0,89	1,21	1,25	1,69	1,45	1,85	0,35	0,68
2	—	1,63	—	1,69	1,85	—	0,42	1,64	1,05	1,78
3	1,22	0,53	1,29	—	1,84	0,94	1,21	—	1,43	1
4	1,43	1,45	—	0,41	1,98	0,78	1,78	1,28	1,21	1,43
5	0,87	1,34	1,06	—	0,89	1,21	1,54	1,78	0,73	1,11
6	1,55	1,68	0,69	1,75	0,95	1,54	1,66	1,37	1,25	0,74
7	—	1,88	—	—	1,96	0,50	1,10	—	1,43	1,05
8	0,65	0,99	1,56	—	1,43	0,65	—	0,89	—	—
9	0,65	1,75	—	0,54	0,97	1,05	1,25	1,51	1,14	1,22
10	1,89	0,59	1,75	—	1,99	1,25	1,48	1,88	1,64	—
11	1,14	2,10	—	1,52	1,54	2,01	1,03	1,56	0,75	0,89
12	0,91	1,87	0,89	0,65	1,05	0,94	1,66	1,11	0,63	1,92
13	1,37	1,43	0,92	—	1,47	1,14	0,65	2	—	1,64
14	1,43	—	1,25	1,49	1,03	—	1,96	1,43	1,08	0,72
15	1,78	1,37	1,45	1,37	1,21	1,78	1,62	1,22	0,74	—
16	0,96	0,89	1,51	0,63	1,07	0,59	1,43	1,01	1,51	1,21
17	1,25	1,65	1,65	—	1,42	1,45	1,51	1,23	1,11	1,01
18	1,11	1,21	1,78	1,25	1,65	1,29	1,81	—	0,65	1,37
19	0,58	1,43	1,08	—	2	1,21	1,11	—	1,88	0,63
20	1,56	1,52	—	1,37	1,08	1,11	1,70	1,25	1,54	1,09
21	1,09	1,34	1,23	0,89	1,84	1,44	1,37	—	0,85	0,78
22	2	1,37	1,42	—	1,54	—	1,67	0,63	0,68	1,25
23	1,21	1,87	—	0,74	1,89	1,22	1,74	1,54	1,55	0,46
24	0,99	1,25	1,67	—	1,84	1,98	1,42	1,21	0,89	1,43
25	1,45	—	0,65	1,64	1,42	1,32	1,83	0,95	1,22	—
26	0,89	1,47	1,58	1,57	1,37	0,89	1,06	0,97	—	0,67
27	—	1,14	1,02	1,58	1,65	1,64	1,43	1,25	1,21	1,54
28	1,87	1,35	—	1,78	1,24	—	1,59	1,05	0,71	1,51
29	0,62	1,11	1,79	—	—	1,42	1,04	0,74	1,78	—
30	1,23	1,02	1,21	1,85	1,06	1,56	0,89	1,28	1,05	0,65

Контрольные вопросы

1. Сформулируйте основные задачи математической статистики.
2. Дайте определение генеральной и выборочной совокупностей.
3. Какие способы отбора выборки Вы знаете? Приведите примеры.
4. Что такое вариационный ряд.
5. Приведите пример статистического распределения выборки. Найдите объем выборки.
6. Что такое статистическая оценка неизвестного параметра генеральной совокупности?
7. Напишите формулы для нахождения выборочной средней и дисперсии выборки.
8. Запишите формулы для нахождения генеральной средней и генеральной дисперсии.
9. В чем различие между полигоном частот и полигоном относительных частот?
10. Чему равна площадь прямоугольника в гистограмме частот?
11. Как определить моду на полигоне частот?
12. Чему равна площадь одного прямоугольника в гистограмме частот?
13. Чему равна сумма площадей всех прямоугольников в гистограмме частот?
14. Какие факторы должны учитываться при выборе числа интервалов гистограммы?

Приложение

1. Сортировка и группировка исходных статистических данных.

Табл.1

№ банка	Рентабельность активов X_i
1	0,8
2	0,85
3	1,04
4	1,13
5	1,15
6	1,24
7	1,28
8	1,28
9	1,33
10	1,36
11	1,37
12	1,49
13	1,51
14	1,58
15	1,62
16	1,66
17	1,73
18	1,75
19	1,98
20	2

Исходные данные после сортировки по возрастанию

N	20
k	5
x min	0,8
x max	2
h	0,24

Объем выборки (общее кол-во банков)

Число интервалов. Вычисляется по формуле Стержесса: $k = 1 + 3{,}322 \lg N$. Для вычисления логарифма использоать ф-ю Excel LOG10

Шаг интервала: $h = (X_{max} - X_{min})/k$

x0	0
x1	0,24
x2	0,48
x3	0,72
x4	0,96
x5	1,2

Вычисляем границы интервалов: $X_i = X_{min} + l * h$

Сгруппируем исходные данные согласно вычисленным границам интервалов.

Кол-во банков попавших в интервал от X_i до X_{i+1}, правую границу не включающий

$f_i/N*100$ для каждого интервала

$m_i = W_i/h$

Ср. значением на интервале является ср. арифм. значение соответствующего промежутка: $X_{cp} = (X_i + X_{i+1})/2$

Табл.2

Рентабельность активов	Кол-во банков (f_i частота)	Накопленная частота (S_i)	Частости (W_i, %)	Относительная плотность (m_i, %)	Средняя рентабельность на интервале
0,8-1,04	2	2	10	41,67	0,92
1,04-1,28	4	6	20	83,33	1,16
1,28-1,52	7	13	35	145,83	1,4
1,52-1,76	5	18	25	104,17	1,64
1,76-2,0	1	19	5	20,83	1,88
2,0 и более	1	20	5	20,83	2,12
ИТОГО:	20		100		

2. Перегруппировка данных исследуемого интервального вариационного ряда для сопоставления и анализа двух рядов.

Табл.3

Группы банков с доходами от 100 до 300 млн. долл.		Группы банков с доходами от 50 до 100 млн. долл.	Средняя рентабельность на интервале
Рентабельность активов	Количество банков в % (частости)	Количество банков в % (частости)	
0,6 – 0,8	10	0	0,7
0,8 – 1,0	30	8,33	0,9
1,0 – 1,1	10	6,67	1,05
1,1 – 1,2	15	8,33	1,15
1,2 – 1,4	20	24,17	1,3
1,4 – 1,8	10	43,33	1,6
1,8 – 2,0	5	4,17	1,9
2,0 и более	0	5,00	2,2
ИТОГО:	**100**	**100,00**	

Данные 1-го и 2-го столбцов являются общими для всех вариантов.

Количество банков на интервале вычисляется как произведение шага интервала ряда, выбранного для сравнения (см. столбец 1), на относительную плотность m_i исходного ряда, соответстующую этому интервалу. Если интервалы не полностью перекрываются, следует разбить их на более мелкие, с границами как у исходного ряда. Так, например, для интервала 1,4 - !,8 частость будет вычисляться как (1,52 - 1,4)m3 + 0,24 m4 + (1,8 - 1,76)m5 = 43,33

При правильных расчетах суммарное количество банков должно сосставить 100% (см. пособие табл.2)

3. Графическое представление кривой (плотности) распределения исходного ряда.

Строится при помощи "Мастера диаграмм" тип "Точечная" со значениями, соединенными сглаживающими линиями. Исходные данные - табл.2 столбцы 2,6.

На вкладке "Ряд" установить курсор в поле "Значения X" и выделить данные столбца 6 табл.2; для поля "Значения Y" выделить данные 2-го столбца.

На вкладке "Параметры" открыть окно "Заголовки" и ввести названия диаграммы и осей. Отредактировать сетку.

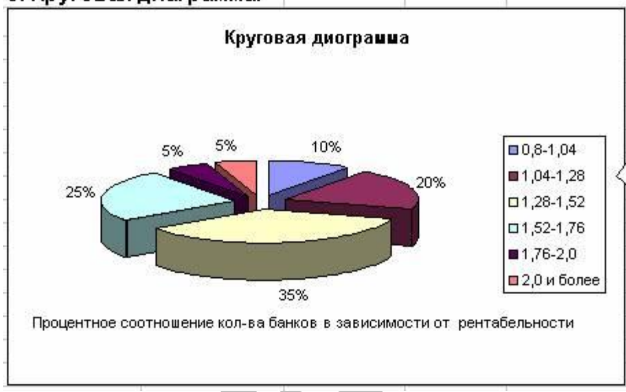
4. Полигон частот.

Строится по табл.2 столбцы 2,6. Тип "Точечная диаграмма, на которой значения соединены отрезками".

5. Гистограмма.

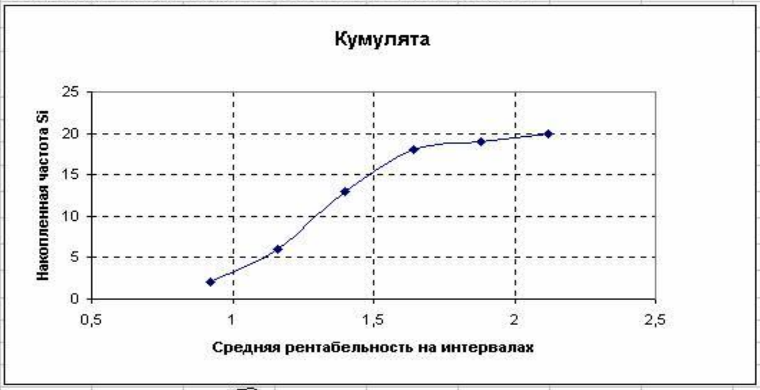
Исходные данные - 1и 2 столбцы табл.2. Тип диаграммы "Гистограмма с накоплением".

6. Круговая диаграмма.



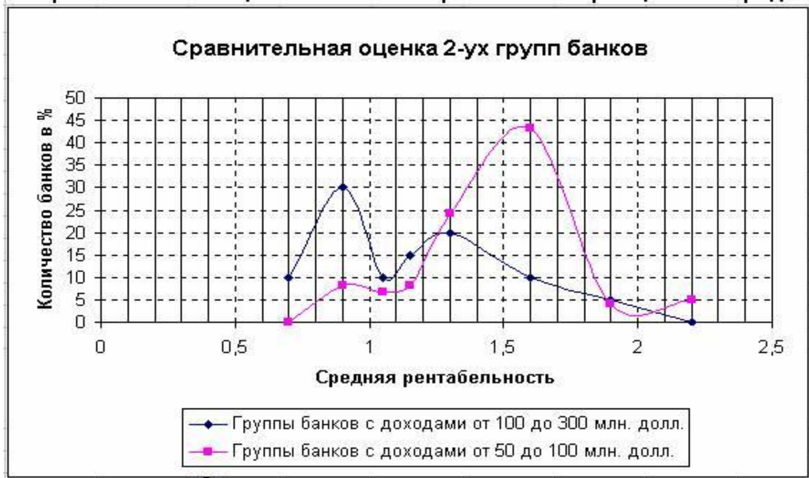
Исходные данные - 1и 2 столбцы табл.2. Тип диаграммы "Разрезанная круговая диаграмма".

7. Кумулятивная кривая накопленных частот.



Строится по табл.2 столбцы 3,6. Тип "Точечная диаграмма со значениями, соединенными сглаживающими линиями".

8. Сравнительная оценка частот интервальных вариационных рядов.



Тип графика "Точечная диаграмма со значениями, соединенными сглаживающими линиями". Исходные данные из табл.3. Использовать вкладку "Ряд" мастера диаграмм "Точечный график".
Для первой кривой выбрать "Значения X" 4-ый столбец, "Значения Y" 2-ой.
Затем нажать кнопку "Добавить ряд" и ввести для 2-ой кривой "Значения X" - 4-ый столбец, "Значения Y" - 3-ий.

Сделать выводы в письменной форме

Список литературы

1. Венцель Е.С. Теория вероятностей. М.: Академия, 2005
2. Гмурман В.Е. Теория вероятностей и математическая статистика: учеб пособие. - М.: Образование, 2007. - 479с.
3. Вентцель Е.С. Задачи и упражнения по теории вероятностей. – М.: Высшая школа, 2002. – 448 с.
4. Кремер Н.Ш. Теория вероятностей и математическая статистика: Учебник для вузов. – М.: ЮНИТИ-ДАНА, 2001.
5. Калинина В.Н., Панкин В.Ф. Математическая статистика. - М.: Высшая школа, 2001.
6. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике. - М. Высшая школа , 2001 -400с.