Типовой расчет

по разделу «Статистическое оценивание»

на тему «Статистический разведочный анализ данных»

Выполнил:

Люлин Дмитрий Антонович

Москва

2024 г.

В таблице ниже приведены данные по времени, которое было затрачено на аудиторскую проверку 100 предприятий отрасли.

Таблица 1

Время (час.), затрачиваемое на аудиторскую проверку 100 предприятий отрасли

108,3	108,4	103,2	106,4	103,9	101,1	102,7	101,9	103,7	102,0
105,3	104,4	106,2	102,0	105,2	103,3	102,3	103,1	103,6	101,2
103,6	102,8	101,8	102,8	101,0	101,2	104,3	100,5	102,6	101,3
105,6	103,8	102,9	101,6	101,2	105,2	103,5	105,7	104,3	100,8
104,9	103,9	104,9	103,9	104,6	103,3	104,6	102,6	106,2	105,5
105,3	106,2	104,2	107,1	102,2	109,6	103,8	101,8	101,2	104,6
103,5	104,7	106,7	105,2	102,0	105,4	105,3	105,3	102,8	103,4
101,5	104,2	103,3	104,9	106,9	102,3	100,2	101,5	101,9	105,9
103,7	107,8	102,5	101,8	102,7	100,8	102,7	102,3	101,8	104,7
102,8	103,1	105,3	105,1	104,3	104,8	102,8	103,9	101,5	103,1

1. Построение интервального вариационного ряда распределения

С помощью формулы Стерджеса сгруппируем выборку и получим интервальный ряд с минимальной нижней границей 99,5851 и шагом h = 1,2297. По этим данным можем сделать вывод, что наша выборка имеет бимодальное распределение, что уже является отклонением от нормального заокна, один из модальных интервалов при этом совпадает с медианным.

Таблица 2

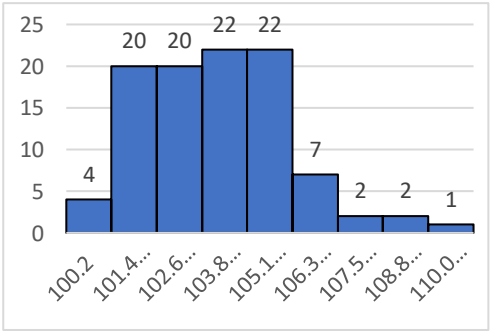
Границы интервалов					
нижняя граница a_i	верхняя граница b_i	частота m_i	накопленная частота m_i^H	относительная частота $w_i=m_i/n$	относительная накопленная частота $w_i^H=m_i^H/n$
99,58515	100,81485	4	4	0,04	0,04
100,81485	102,04455	20	24	0,2	0,24
102,04455	103,27425	20	44	0,2	0,44
103,27425	104,50395	22	66	0,22	0,66
104,50395	105,73365	22	88	0,22	0,88
105,73365	106,96335	7	95	0,07	0,95
106,96335	108,19305	2	97	0,02	0,97
108,19305	109,42275	2	99	0,02	0,99
109,42275	110,65245	1	100	0,01	1
		$n=\sum m_i=100$		$\sum w_i=1$	

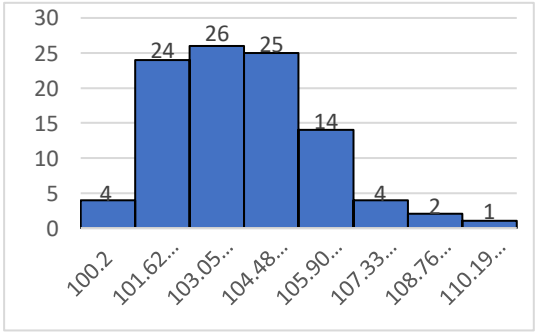
Далее проведем группировку по остальным формулам:

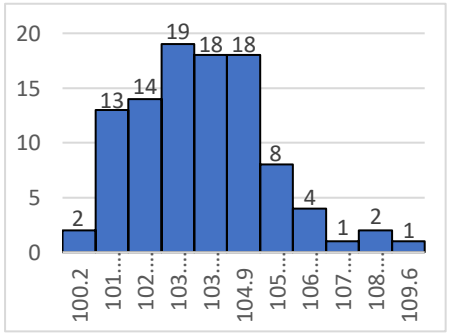
1) Формула Дэвида Скотта
2) Формула квадратного корня
3) Формула Фридмана - Диаконис

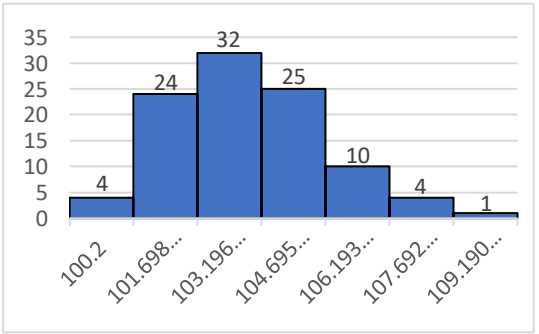
Таблица 3

Формула Стерджеса			Формула Скотта			Формула квадратного корня			Формула Фридмана-Диаконисса		
Шаг	h=1,2297	Формула: h = R/(1 + log(2; n))	Шаг	h=1,4274	Формула: h = 3,5 * S / n ^ (1/3)	Шаг	h=0,94	Формула: h = R / n^0,5	Шаг	h=1,4984	Формула: h = 2,6 * IQR / n ^ (1/3)
границы интервалов		Частота встречаемости m_i	границы интервалов		Частота встречаемости m_i	границы интервалов		Частота встречаемости m_i	границы интервалов		Частота встречаемости m_i
нижняя a_i	верхняя b_i		нижняя a_i	верхняя b_i		нижняя a_i	верхняя b_i		нижняя a_i	верхняя b_i	
99,58515	100,81485	4	99,48630698	100,913693	4	99,73	100,67	2	99,45079534	100,949205	4
100,81485	102,04455	20	100,913693	102,3410791	24	100,67	101,61	13	100,9492047	102,447614	24
102,04455	103,27425	20	102,3410791	103,7684651	26	101,61	102,55	14	102,447614	103,946023	32
103,27425	104,50395	22	103,7684651	105,1958512	25	102,55	103,49	19	103,9460233	105,444433	25
104,50395	105,73365	22	105,1958512	106,6232372	14	103,49	104,43	18	105,4444326	106,942842	10
105,73365	106,96335	7	106,6232372	108,0506233	4	104,43	105,37	18	106,942842	108,441251	4
106,96335	108,19305	2	108,0506233	109,4780093	2	105,37	106,31	8	108,4412513	109,939661	1
108,19305	109,42275	2	109,4780093	110,9053954	1	106,31	107,25	4			
109,42275	110,65245	1				107,25	108,19	1			
						108,19	109,13	2			
						109,13	110,07	1			









Как можно увидеть, формула Стерждеса дает наиболее правдоподобное представление о выборке по сравнению с другими формулами. Квадратный корень дает слишком большое количество интервалов, в том время как Скотт и Фридман-Диаконис слишком мало, в особенности второй.

Далее во всех расчетах будет использоваться интервальный ряд, рассчитанный по формуле Стерджеса.

2. Вычисление основных числовых характеристик по исходным данным и интервальному вариационному ряду

Ниже приведены вспомогательные таблицы для вычисления выборочных характеристик интервального ряда и не сгруппированных данных.

Таблица 4

xi	mi	xi*mi	Δi= xi - xcp	Δi·mi	Δi^2·mi	Δi^3·mi	Δi^4·mi
1	2	3	4	5	6	7	8
100,2	4	400,8	-3,48005	-13,9202	48,44302	-168,584	586,6815
101,4297	20	2028,594	-2,25035	-45,00702	101,2816	-227,919	512,898
102,6594	20	2053,188	-1,02065	-20,41302	20,83457	-21,2648	21,70396
103,8891	22	2285,56	0,209049	4,599078	0,961433	0,200987	0,042016
105,1188	22	2312,614	1,438749	31,652478	45,53997	65,52059	94,26768
106,3485	7	744,4395	2,668449	18,679143	49,84434	133,0071	354,9226
107,5782	2	215,1564	3,898149	7,796298	30,39113	118,4692	461,8104
108,8079	2	217,6158	5,127849	10,255698	52,58967	269,6719	1382,837
110,0376	1	110,0376	6,357549	6,357549	40,41843	256,9621	1633,649
∑	**100**	**10368,01**		**0,00**	**390,3042**	**426,0637**	**5048,812**

Таблица 5

x_i	x_i - xcp	$(x_i - xcp)^2$	$(x_i - xcp)^3$	$(x_i - xcp)^4$
108,3	4,605	21,206025	97,65374513	449,6954963
108,4	4,705	22,137025	104,1547026	490,0478759
103,2	-0,495	0,245025	-0,121287375	0,060037251
106,4	2,705	7,317025	19,79255263	53,53885485
103,9	0,205	0,042025	0,008615125	0,001766101
101,1	-2,595	6,734025	-17,47479487	45,3470927
.	.	.	.	.
.	.	.	.	.
Сумма				
10369,5	4,93*10^-12	358,3275	370,829925	3964,098949

После завершения всех расчетов в Excel получаем таблицу, приведенную ниже.

Таблица 6

№ п/п	Характеристика			По исходным данным		По интервальному ряду
				по формуле	Excel	
Характеристики центра группирования						
1	средняя арифметическая			103,695	103,695	103,680051
2	выборочная медиана			103,55	103,55	103,6096227
3	выборочная мода			105,3; 102,8	105,3; 102,8	бимодальное распределение
Показатели вариации						
4	выборочная дисперсия			3,5833	3,5833	3,90304157
5	выборочное среднее квадратическое отклонение			1,893	-	1,975611695
6	Коэффициент вариации (в %)			1,8255	-	1,905488738
7	Выборочный центральный момент 1-ого порядка			$4{,}9316 * 10^{-14}$	-	$-1{,}27898*10^{-14}$
8	Выборочный центральный момент 2-ого порядка			3,583275	-	3,90304157
9	Выборочный центральный момент 3-ого порядка			3,70829925	-	4,260637114
10	Выборочный центральный момент 4-ого порядка			39,64098949	-	50,48812452
11	Выборочный коэффициент асимметрии			0,546706911	0,555067866	0,552547707
12	Выборочный коэффициент эксцесса			0,087338206	0,154354589	0,314231834

Анализируя полученные результаты, видим, что характеристики центра группирования по исходным данным совпадают при расчетах по формулам и в Excel. Более того, среднее значение и медиана по интервальному ряду также незначительно отличаются от реальных показателей по выборке, что еще раз доказывает эффективность использования группировки по формуле Стерджеса.

Показатели вариации имеют уже более значимые различия в случае анализа по исходным данным и по интервальному ряду. Выборочные моменты 3 и 4 порядков, а также коэффициент эксцесса отличаются крайне сильно. Коэффициент асимметрии составляет чуть более 0,5 по модулю, что говорит о правосторонней асимметрии и является отклонением от нормального распределения. Коэффициент эксцесса по сгруппированным данным составляет 0,3142, а по расчетам выборки 0,0873 или 0,1544, что означает незначительную плосковершинность нашего распределения, но не является весомым аргументов против нормального распределения.

Как итог, мы имеем два признака отсутствия у генеральной совокупности нормального распределения: бимодальность и относительно высокую асимметрию (>0,5).

3. Графическое представление одномерных количественных данных

На этом этапе приведем несколько графиков, которые являются аппроксимацией плотности распределения и функции распределения генеральной совокупности.

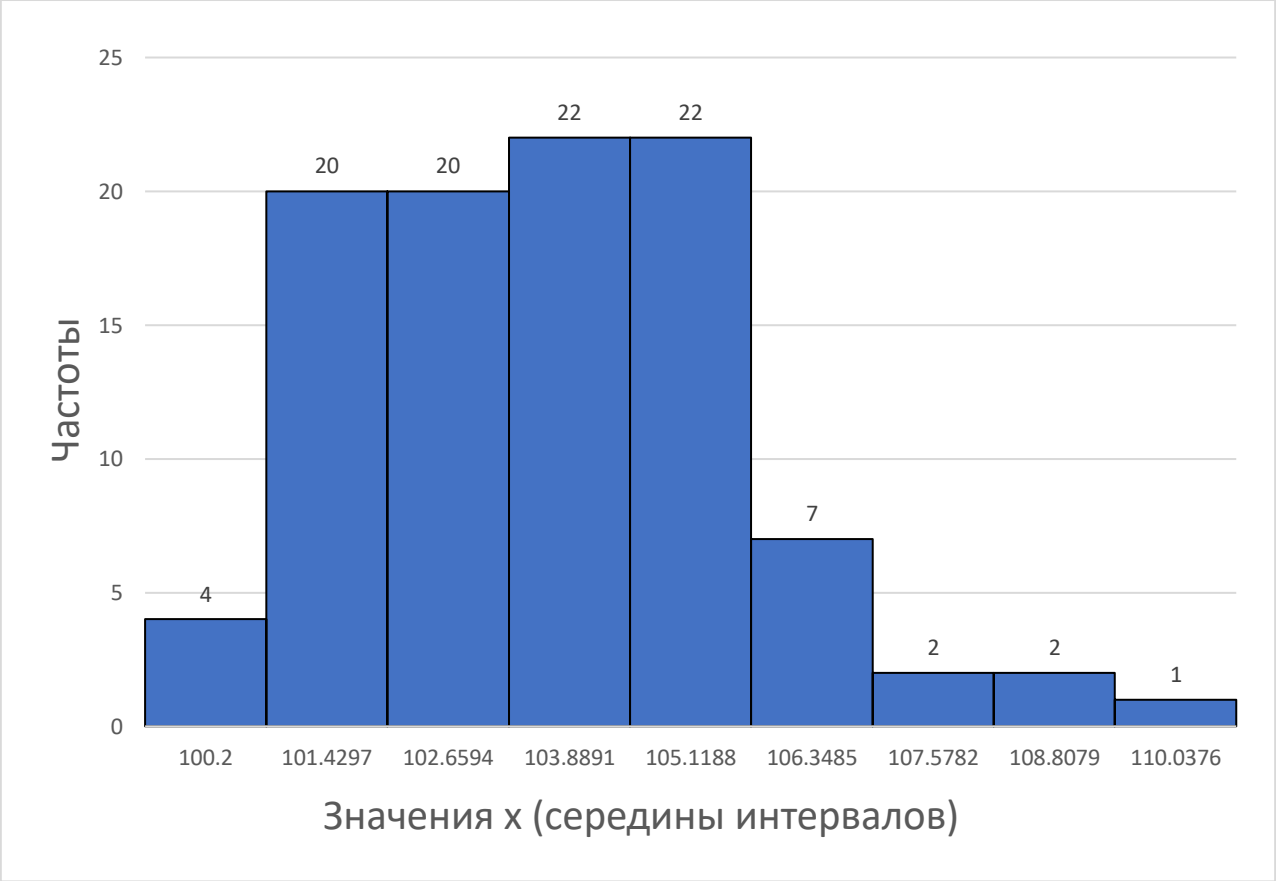


Рис.1.Гистограмма частот интервального ряда распределения времени, затрачиваемого на аудиторскую проверку 100 предприятий отрасли

Несмотря на то, что самая большая частота совпадает с медианным интервалом, мы визуально наблюдаем бимодальность и существенную правостороннюю асимметрию, так

как большинство частот сосредоточено в левой области. Эта гистограмма является первой наглядной визуализацией асимметрии нашей выборки.

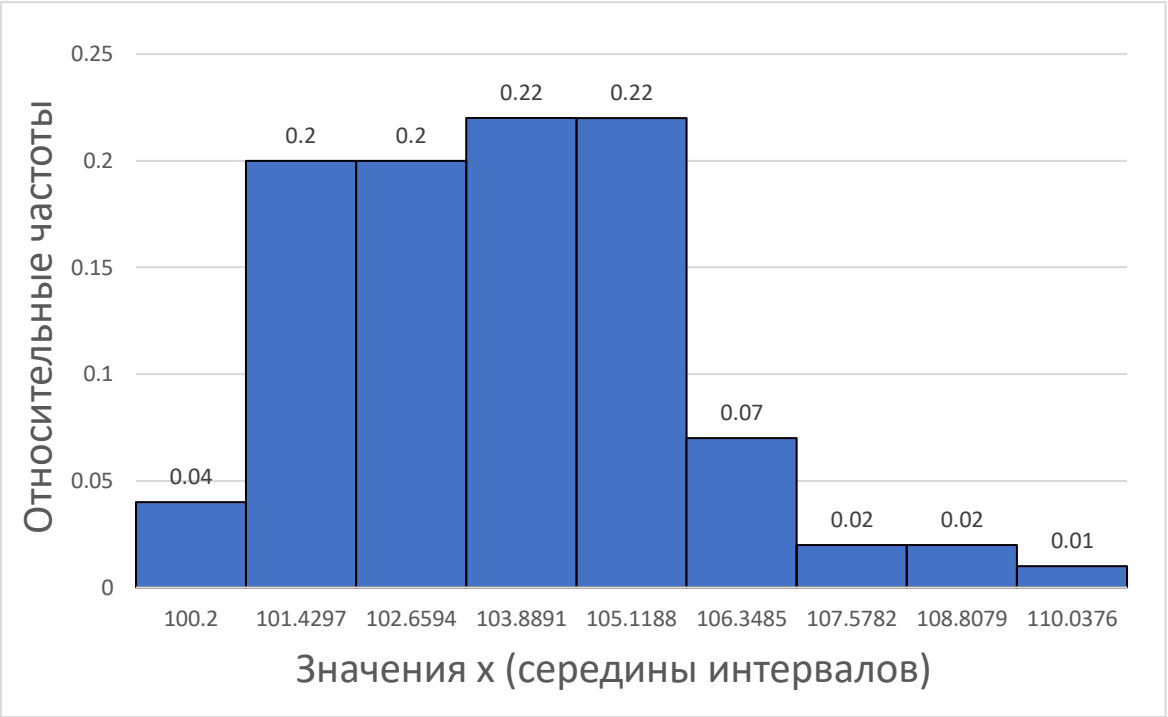


Рис.2.Гистограмма относительных частот интервального ряда распределения времени, затрачиваемого на аудиторскую проверку 100 предприятий отрасли

Гистограмма частот является подобием функции плотности истинного распределения. Как мы видим, гистограмма частот и гистограмма относительных частот отличаются друг от друга лишь значениями ординаты. Графики подтвердили выводы из прошлого пункта о бимодальности и асимметричности распределения, наглядно видно, что частота выборки относительно времени значительно смещена относительно центра.

Из гистограммы по тем же точкам построим полигоны частот, для еще более наглядной визуализации предполагаемой функции плотности распределения.

Рис.3.Полигон частот интервального ряда распределения времени, затрачиваемого на аудиторскую проверку 100 предприятий отрасли

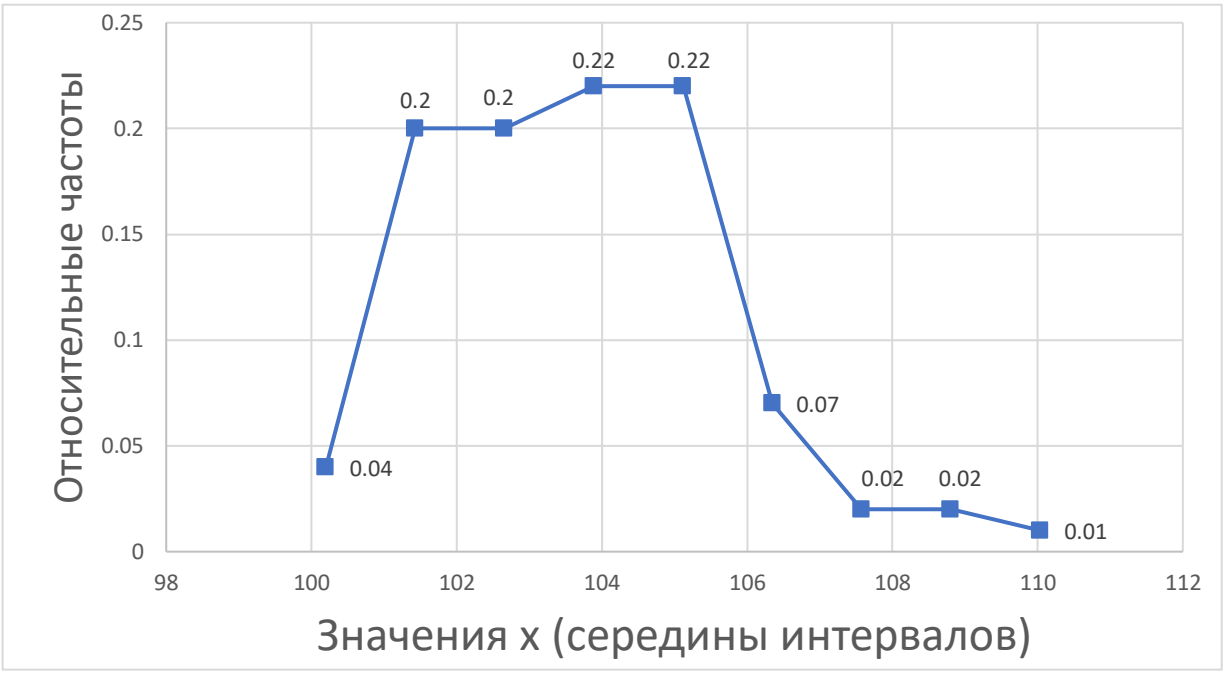


Рис.4.Гистограмма относительных частот интервального ряда распределения времени, затрачиваемого на аудиторскую проверку 100 предприятий отрасли

Как мы можем увидеть, и здесь относительные и обычные частоты в виде ординаты не изменяют графики, а лишь отражают долю и количество соответственно.

Последующие два графика – кумуляты частот и относительных частот интервального ряда, имитирующие функцию распределения генеральной совокупности и представляющие накопленные интервальные частоты.

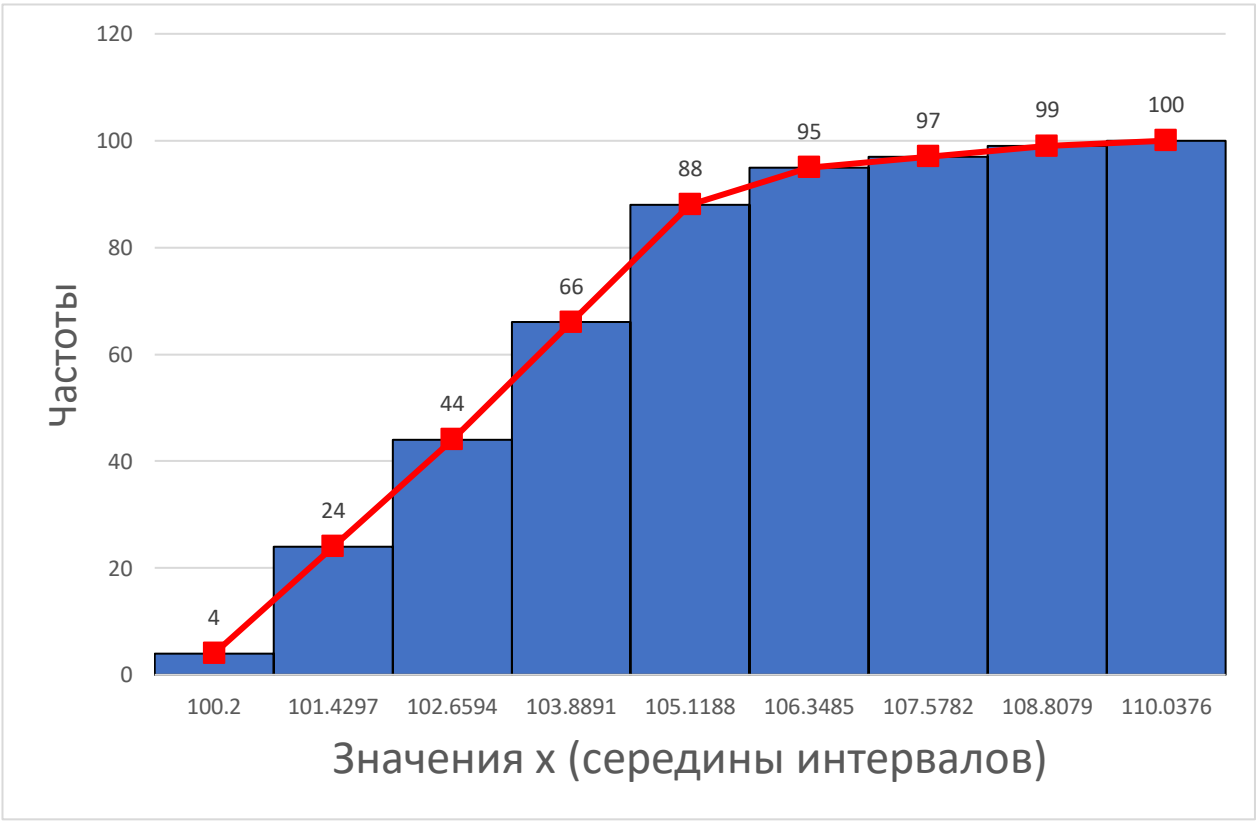


Рис. 5. Кумулята накопленных частот интервального ряда распределения времени, затрачиваемого на аудиторскую проверку 100 предприятий отрасли

Рис. 6. Кумулята накопленных относительных частот интервального ряда распределения времени, затрачиваемого на аудиторскую проверку 100 предприятий отрасли

По кумуляте мы видим, что самые большие накопления частот начинаются с 2 интервала и длятся до середины всех интервалов, то есть от значения 101,4297 до 105,1188. Для очередного подтверждения асимметрии заметим, что в первых четырех интервалах совокупный прирост - 66 % всех наблюдений, а в последних четырех – 12%.

Далее рассмотрим графики не сгруппированным данных.

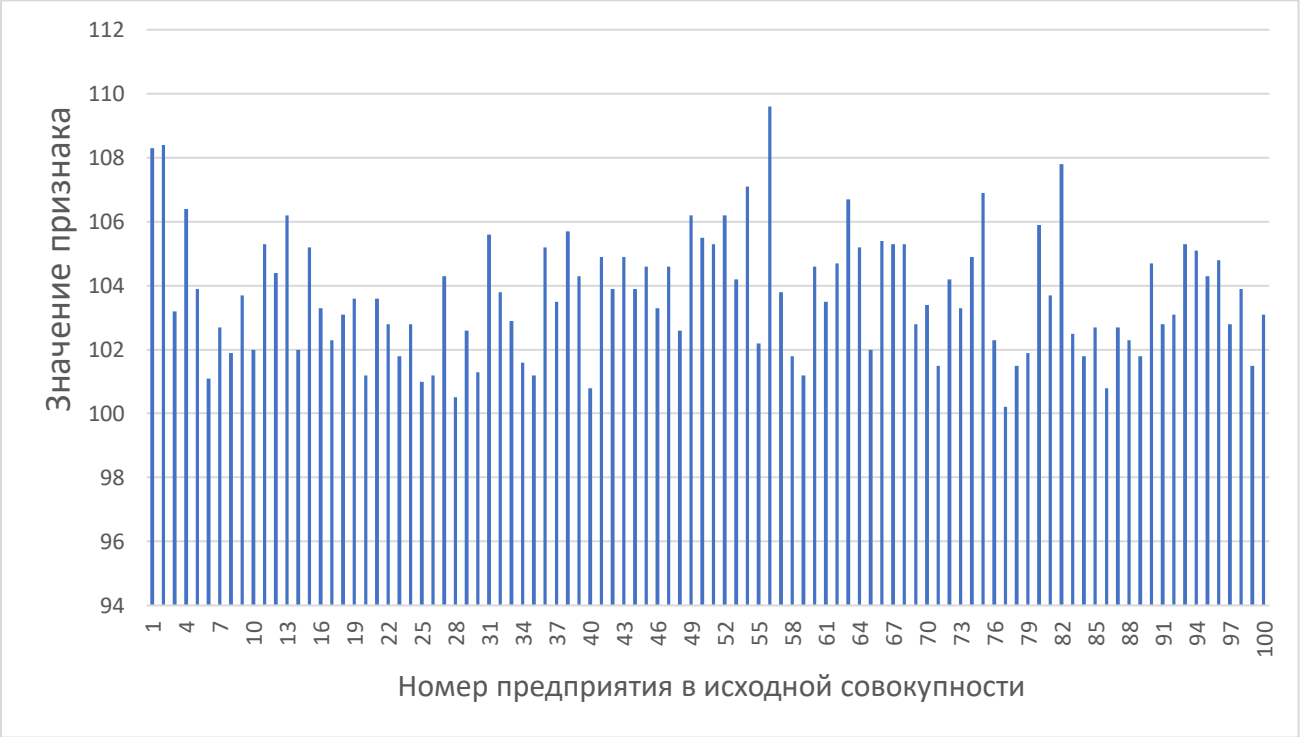


Рис. 7. Столбиковая диаграмма времени, затрачиваемого на аудиторскую проверку 100 предприятий отрасли в исходном неранжированном порядке

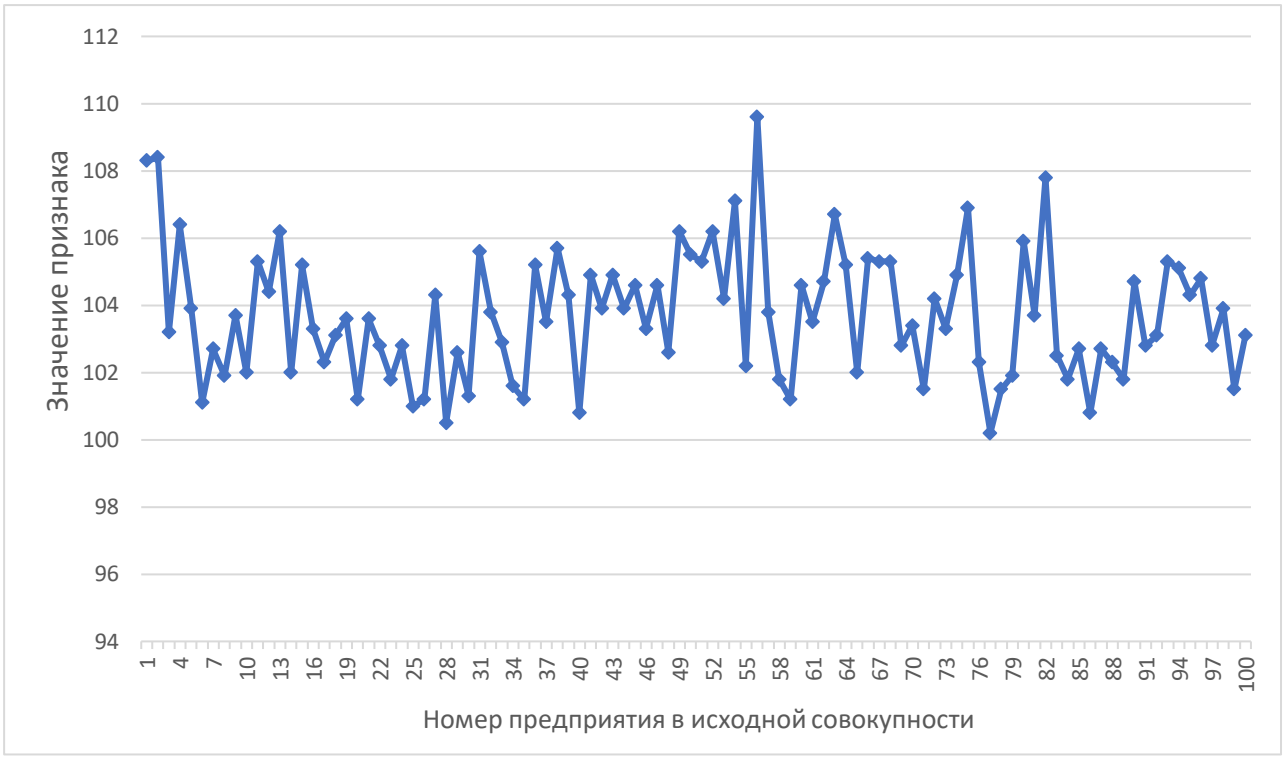


Рис. 8. Точечный график времени, затрачиваемого на аудиторскую проверку 100 предприятий отрасли в исходном неранжированном порядке

По неранжированным данным можем увидеть примерный разброс значений, в нашем случае от 100 до 110, и явные выбросы, как, например, наблюдение 56, которое явно имеет самое большое значение из всей выборки и область, где сосредоточено наибольшее количество значений, от 101 до 106. Также можем заметить, что сильных выбросов вниз нет, в то время как наверх имеется как минимум три (1, 2 и 56 наблюдения).

Проведем ранжирование данных.

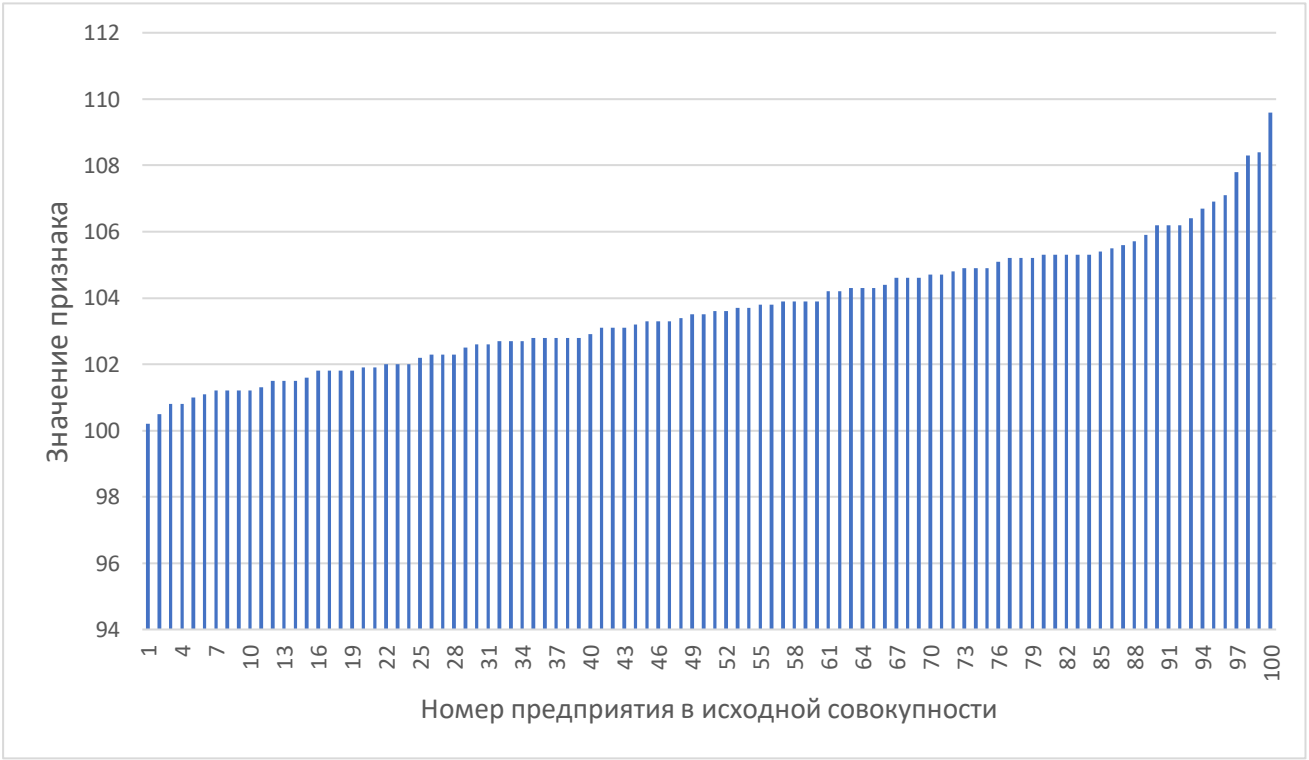


Рис. 9. Столбиковая диаграмма времени, затрачиваемого на аудиторскую проверку 100 предприятий отрасли в исходном ранжированном порядке

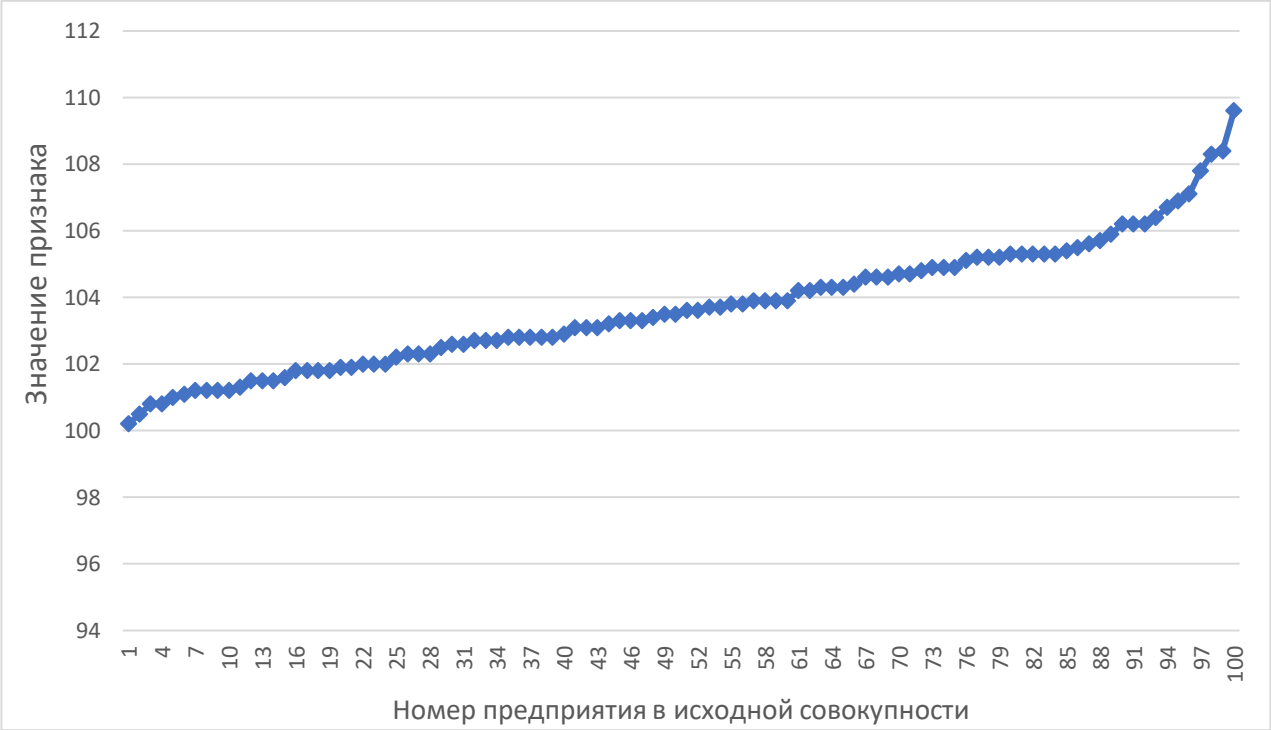


Рис. 10. Точечный график времени, затрачиваемого на аудиторскую проверку 100 предприятий отрасли в исходном ранжированном порядке

По ранжированным исходным данным можем увидеть, что время увеличивается достаточно равномерно на протяжении всех 100 наблюдений, кроме тех самых выбросов вверх, о которых я писал ранее.

По графикам исходных данных можно понять, для анализа данных важно группировать выборку, формируя интервальный вариационный ряд – сгруппированные данные визуально эффективнее показывают структуру данных, асимметрию, модальные значения и т.д., при этом не сильно искажая выборочные характеристики. С помощью этого первичного анализа я могу выдвинуть гипотезу, что распределение времени аудиторской проверки компаний не является нормальным распределением. На следующий этапах анализа мы углубимся в оценку выборки с целью подтверждения или опровержения моей гипотезы.

4. Расчет теоретической нормальной кривой распределения

Следующим этапом будет расчет теоретических частот в предположении того, что распределение является нормальным. Для этого воспользуемся таблицей:

Таблица 7

Интервалы		**m_i**	**t_{1i}**	**t_{2i}**	**$\Phi(t_{1i})$**	**$\Phi(t_{2i})$**	**p_i**	**$n*p_i$**	**m_i^T**	**p_i Excel**
a_i	b_i									
99,58515	100,8149	4	-5	-1,45029	-1	-0,8529	0,07355	7,355	7	0,073489
100,8149	102,0446	20	-1,45029	-0,82785	-0,8529	-0,5935	0,1297	12,97	13	0,13039
102,0446	103,2743	20	-0,82785	-0,20541	-0,5935	-0,1663	0,2136	21,36	21	0,214749
103,2743	104,504	22	-0,20541	0,417035	-0,1663	0,3255	0,2459	24,59	25	0,243046
104,504	105,7337	22	0,417035	1,039475	0,3255	0,7017	0,1881	18,81	19	0,189035
105,7337	106,9634	7	1,039475	1,661915	0,7017	0,9031	0,1007	10,07	10	0,101027
106,9634	108,1931	2	1,661915	2,284355	0,9031	0,9774	0,03715	3,715	4	0,03709
108,1931	109,4228	2	2,284355	2,906795	0,9774	0,9964	0,0095	0,95	1	0,00935
109,4228	110,6525	1	2,906795	5	0,9964	1	0,0018	0,18	0	0,001617
сумма		**100**				**сумма**	**1**	**100**	**100**	**1,000**

Для расчета t_i воспользуемся формулой $t_{1i} = (a_i - x(cp)) / S$ и $t_{2i} = (b_i - x(cp)) / S$, причем граничные t целенаправленно сделаем большими (для табличных значений хватит «5 + 0») для того, чтобы захватить «хвосты» нормального распределения. Заметим, что вероятности по табличным значениям и по формуле Excel несущественно отличаются, это происходит, потому что Excel не округляет значения.

По полученным результатам наблюдаются отклонения по всем интервалам, более того отклонения существенны и могут быть как положительные (интервал 2), так и отрицательные (интервал 1 и 4). Это позволяет сделать вывод, что наша выборка неоднородна, ее форма может быть скошена и антисимметрична, и скорее всего не является нормальным распределением.

Для визуального восприятия обратимся к графику.

На нем также видно существенное различие между теоретической нормальной кривой распределения и гистограммой частот по интервальному ряду.

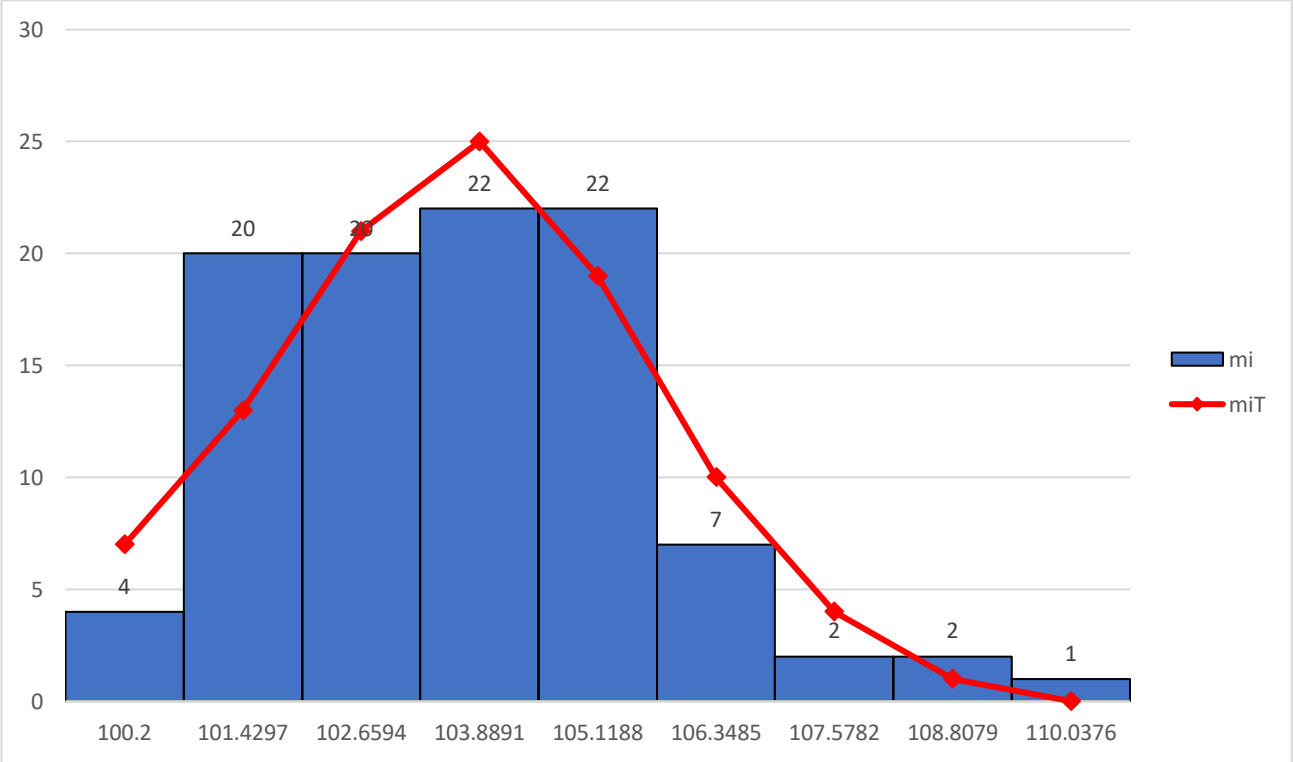


Рис. 11. Гистограмма частот интервального ряда распределения и теоретическая кривая Гаусса нормального распределения

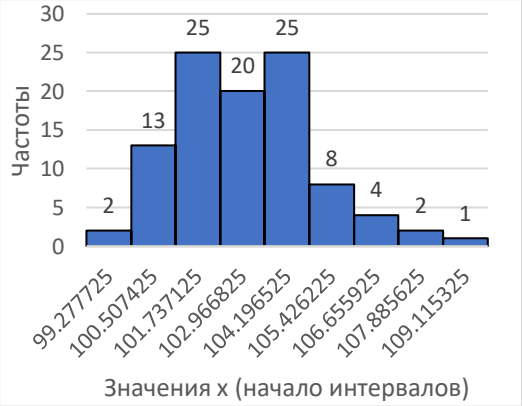
5. Построение альтернативных видов гистограмм – ASH (Average Shifted Histogram) – Усредненная Смещенная (по сдвигу) Гистограмма

Визуальный и числовой вид выборки сильно зависит не только от выбора ширины интервала, но и от определения нижней границы первого интервала. В предыдущий пунктах мы пользовались формулой (вне зависимости от метода расчета ширины интервала) определения первой нижней границы, как разности между минимальным значением выборки и половиной ширины интервала.

Для достижения наибольшей наглядности и увеличения эффективности анализа, рассчитаем интервальные ряды 4 методами, а далее усредним полученные гистограммы.

Таблица 8

$a_1 = X_{(1)} - h$			$a_1 = X_{(1)} - 3h/4$			$a_1 = X_{(1)} - h/2$			$a_1 = X_{(1)} - h/4$		
Границы интервалов		Частоты m_i	Границы интервалов		Частоты m_i	Границы интервалов		Частоты m_i	Границы интервалов		Частоты m_i
нижняя a_i	верхняя b_i		нижняя a_i	верхняя b_i		нижняя a_i	верхняя b_i		нижняя a_i	верхняя b_i	
98,9703	100,2	1	99,277725	100,507425	2	99,58515	100,81485	4	99,89257	101,12227	6
100,2	101,4297	10	100,50742	101,737125	13	100,81485	102,04455	20	101,1223	102,35197	22
101,4297	102,6594	20	101,73712	102,966825	25	102,04455	103,27425	20	102,3520	103,58167	22
102,6594	103,8891	25	102,96682	104,196525	20	103,27425	104,50395	22	103,5817	104,81137	22
103,8891	105,1188	20	104,19652	105,426225	25	104,50395	105,73365	22	104,8114	106,04107	17
105,1188	106,3485	16	105,42622	106,655925	8	105,73365	106,96335	7	106,0411	107,27077	7
106,3485	107,5782	4	106,65592	107,885625	4	106,96335	108,19305	2	107,2708	108,50047	3
107,5782	108,8079	3	107,88562	109,115325	2	108,19305	109,42275	2	108,5005	109,73017	1
108,8079	110,0376	1	109,11532	110,345025	1	109,42275	110,65245	1	109,7302	110,95987	0



После завершения вспомогательной таблицы мы уже можем увидеть достаточно интересные результаты. Гистограммы в зависимости от выбора первой нижней границы имеют абсолютно разный вид. Первые две отдаленно напоминают кривую плотности нормального распределения, в то время как третья имеет 3 модальных интервала и нормальной ее назвать точно нельзя.

Далее усредним значения и получим искомую гистограмму и частоты ASH.

Таблица 9

Расчет усредненной по сдвигу гистограммы ASH

Границы интервалов	**Частота встречаемости m_i**				**Частоты ASH**
	Ряд 1	**Ряд 2**	**Ряд 3**	**Ряд 4**	
98,9703	1	0	0	0	0,0625
99,27773	1	2	0	0	0,1875
99,58515	1	2	4	0	0,4375
99,89258	1	2	4	6	0,8125
100,2	10	2	4	6	1,375
100,5074	10	13	4	6	2,0625
100,8149	10	13	20	6	3,0625
101,1223	10	13	20	22	4,0625
101,4297	20	13	20	22	4,6875
101,7371	20	25	20	22	5,4375
102,0446	20	25	20	22	5,4375
102,352	20	25	20	22	5,4375
102,6594	25	25	20	22	5,75
102,9668	25	20	20	22	5,4375
103,2743	25	20	22	22	5,5625
103,5817	25	20	22	22	5,5625
103,8891	20	20	22	22	5,25
104,1965	20	25	22	22	5,5625
104,504	20	25	22	22	5,5625
104,8114	20	25	22	17	5,25
105,1188	16	25	22	17	5
105,4262	16	8	22	17	3,9375
105,7337	16	8	7	17	3
106,0411	16	8	7	7	2,375
106,3485	4	8	7	7	1,625
106,6559	4	4	7	7	1,375
106,9634	4	4	2	7	1,0625
107,2708	4	4	2	3	0,8125
107,5782	3	4	2	3	0,75
107,8856	3	2	2	3	0,625
108,1931	3	2	2	3	0,625
108,5005	3	2	2	1	0,5
108,8079	1	2	2	1	0,375
109,1153	1	1	2	1	0,3125
109,4228	1	1	1	1	0,25
109,7302	1	1	1	0	0,1875
110,0376	0	1	1	0	0,125
110,345	0	0	1	0	0,0625
110,6525	0	0	0	0	0
Сумма					
	400	400	400	400	100

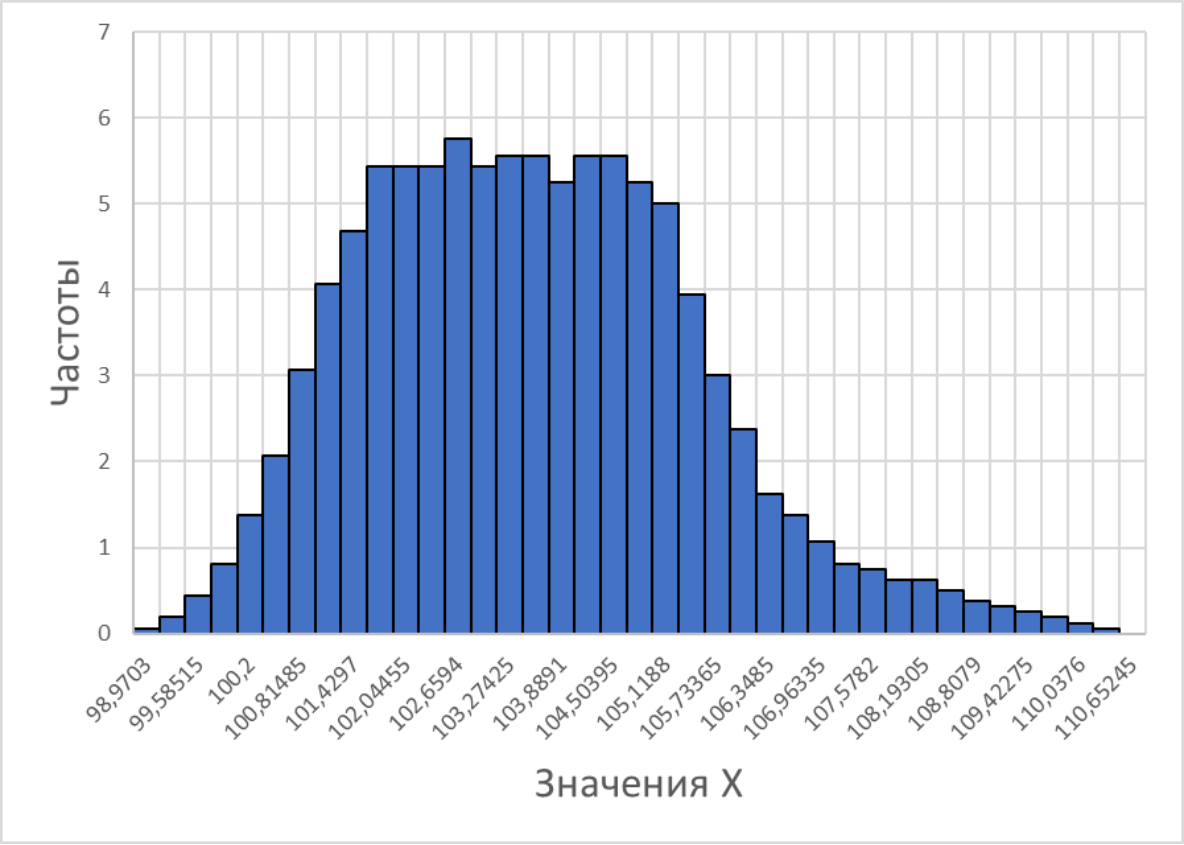


Рис.12. *Усредненная по сдвигу гистограмма ASH с количеством гистограмм усреднения 4 по данным времени, затрачиваемого на аудиторскую проверку 100 предприятий отрасли*

Построенная гистограмма дала достаточно интересные результаты. Очевидно, распределение не является нормальным, оно существенно смещено вправо, тем не менее сама форма основного объема данных визуально «под определенным углом» похожа на нормальное распределение с удлиненным правым хвостом. Делая промежуточные выводы, можно сказать, что асимметричность данных относительно центра дает нам повод полагать, что распределение не подчиняется нормальному закону.

6. Построение ядерных оценок плотности 3 методами

Перейдем к последнему этапу проверки гипотезы – ядерной оценке. Она позволяет провести анализ по выборке, не склоняясь к какому – либо фундаментальному распределению, например, нормальному или экспоненциальному. За шаг в диапазоне значений для не сгруппированных данных возьмем 0,0949 для того, чтобы длина диапазона совпадала с длиной выборки.

Таблица 10

N	Min	Max	Размер шага	IQR	S
100	100,2	109,6	0,094949495	2,675	1,892954

Ширину окна рассчитаем по формуле h = 0,9 * N^(-1/5) * min{IQR/1,34;S} и получим 0,6782. Для сгруппированных данных размер шага возьмем такой же, как и ширина окна. В зависимости от ширины окна мы будем получать более гладкие (в случае маленького окна) и более острые (в случае большого окна) графики.

Сначала проведем оценку с треугольным ядром.

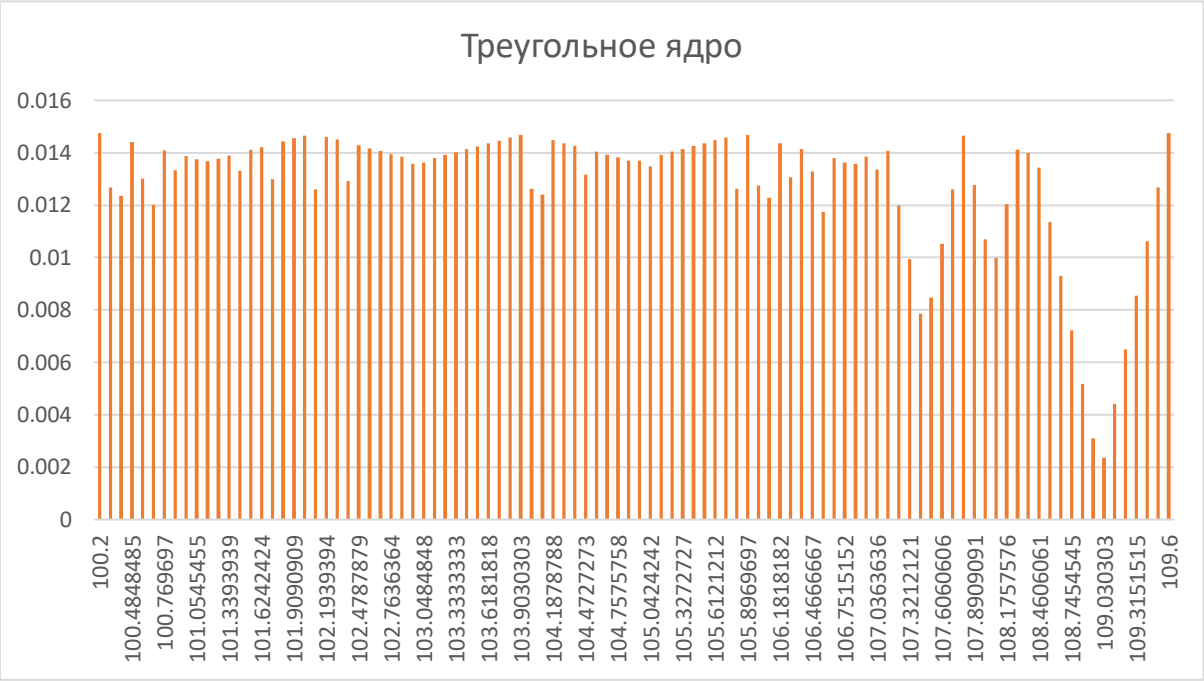


Рис.13. Гистограмма оценки плотности распределения выборки по треугольному ядру

Рис.14. Гистограмма оценки плотности распределения интервального ряда по треугольному ядру

Нормализованные значения треугольной оценки плотности показывают, что большинство значений сосредоточено в более низкой части диапазона, с несколькими выбросами на более высоких уровнях, что указывает на асимметричное распределение. Сумма нормализованных значений близка к единице, что говорит о корректной нормализации.

Далее проведем оценку по прямоугольному ядру.

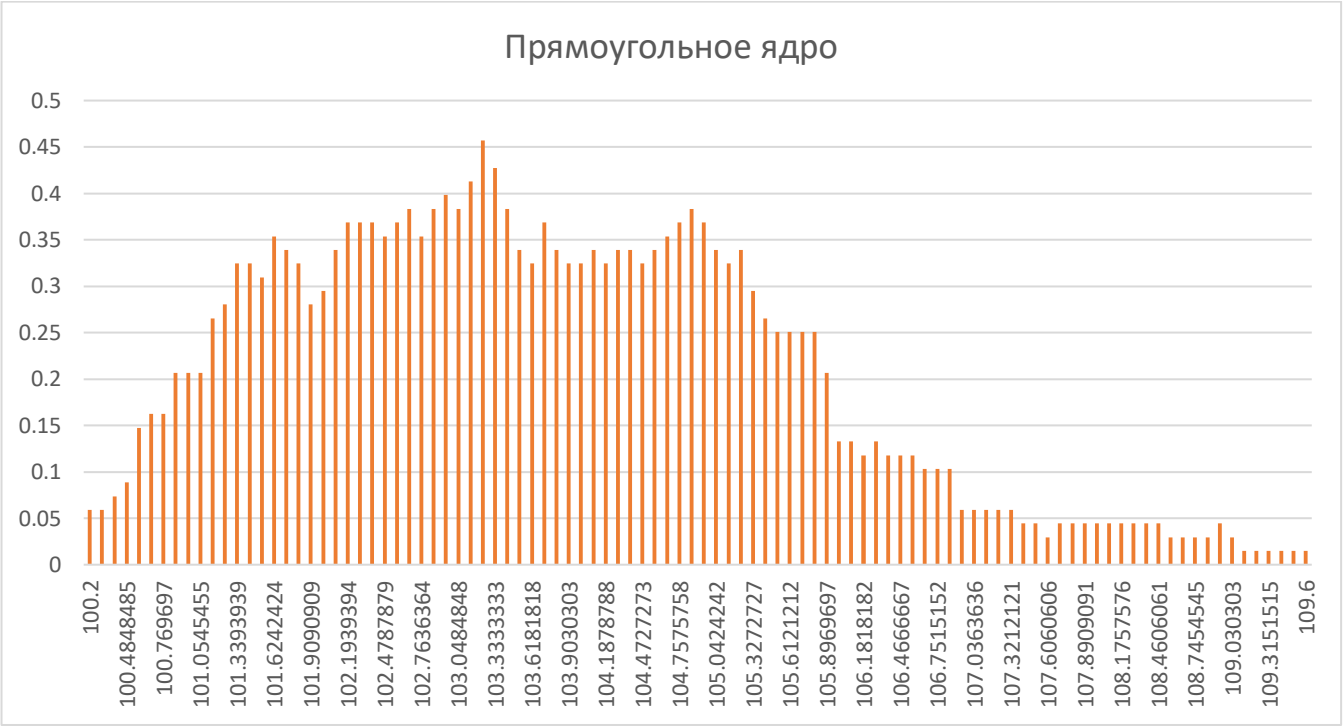


Рис.15. Гистограмма оценки плотности распределения выборки по прямоугольному ядру

Рис.16. Гистограмма оценки плотности распределения интервального ряда по прямоугольному ядру

На данном этапе мы можем увидеть, что обе гистограммы напоминают по своей форме гистограмму интервального ряда по Стерджесу, гистограмму ASH. Однако, оказалось, что для оценки плотности сгруппированных данных по прямоугольному ядру выбранный мной размер шага не подходит, в получившейся гистограмме много нулевых интервалов несмотря на то, что размах между значениями изначальной выборки относительно равномерный. В гистограмме по не сгруппированным данным можно наглядно увидеть моду и остальные выбросы частот.

Наконец, рассчитаем оценку по гауссовскому ядру.

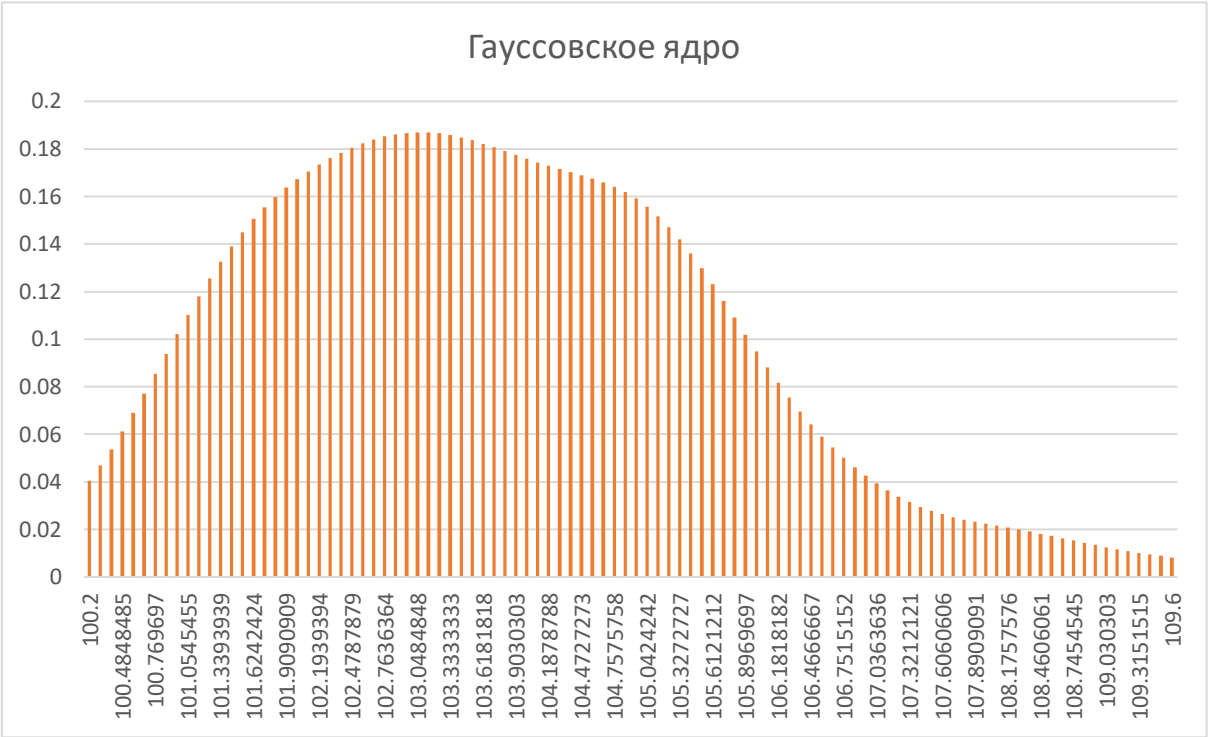


Рис.17. Гистограмма оценки плотности распределения выборки по гауссовскому ядру

Рис.18. Гистограмма оценки плотности распределения интервального ряда по гауссовскому ядру

В итоге мы видим, что оценки по гауссовскому ядру крайне похожи на гистограмму интервального ряда по формуле Стерджеса и ASH гистограмму. Размер шага для не сгруппированных данных был выбран вполне удачно, итоговая гистограмма представляет собой достаточно гладкий график, по которому можно сделать выводы и предположения о генеральной совокупности. Гауссовское распределение асимметрично относительно среднего значения, и анализ показывает, что наибольшая плотность вероятности находится в центральной части, что может указывать на наличие кластера данных. Рисунки 17-18 в очередной раз доказывают правостороннюю асимметрию нашей выборки.

7. Выводы по проделанному анализу

Проведя все этапы анализа, можно сделать выводы по методах, которые были применены. По не ранжированным данным невозможно провести эффективного анализа, лишь примерно прикинуть диапазон значений, увидеть выбросы и т.д. Формула Стерджеса действительно является самой эффективной среди остальных, показывая нам достоверную картину по выборке и не почти не искажая выборочные характеристики. Несмотря на это, выбор первой нижней границы также сильно влияет на результат, в чем мы убедились, строя усредненную по сдвигу гистограмму ASH. Но усреднив их, мы получаем очень подробную и эффективную группировку, которая помогает более уверенно делать различные выводы.

Говоря о распределении времени, несмотря на такие факты, как:

1) Низкий уровень коэффициента эксцесса
2) Близкое расположение относительно друг друга медианы, моды и среднего
3) Колоколообразный вид плотности, полученный при постройке гистограммы ASH, а также оценок плотности с помощью гауссовского и прямоугольного ядра

Мы можем с уверенностью сказать, что наша выборка отклоняется от нормального распределения, а значит выдвинутая гипотеза подтвердилась. Этому есть несколько причин:

1) Высокий коэффициент асимметрии
2) Бимодальность выборки и интервального ряда, причем рассчитанного с применением 4 разных формул для ширины интервала
3) Концентрация значений левее медианы (правосторонняя асимметрия), подтвержденный гистограммой по интервальному ряду, гистограммой ASH и ядерными оценками, а также длинный правый хвост

Рассматривая данные результаты в контексте того, что это измерения времени, можно сказать, что размах выборки в 9 минут при минимальном 100, а максимальном 109 (значения указаны примерно) не является существенным. Более того, правосторонняя асимметрия означает, что в среднем проверка документов занимает меньшее время, нежели могла бы судя по выборке, что нельзя интерпретировать, как отрицательный эффект. Чем эффективнее и быстрее выполняются задачи, тем лучше, следовательно в данном случае правосторонняя асимметрия является преимуществом, чем недостатком несмотря на то, что хуже поддается статистическим оценкам, ввиду своего несоответствия нормальному закону.