

УДК: 625.144.6

СТАТИСТИЧЕСКАЯ ОБРАБОТКА РЕЗУЛЬТАТОВ ИССЛЕДОВАНИЯ
КОНКУРСНЫХ ЗАЯВОК ПО УДАЛЕНИЮ НЕЖЕЛАТЕЛЬНОЙ
РАСТИТЕЛЬНОСТИ С ТЕРРИТОРИИ ИНФРАСТРУКТУРНЫХ ОБЪЕКТОВ

Платонов А.А.

Федеральное государственное бюджетное образовательное учреждение высшего
образования «Ростовский государственный университет путей сообщения»

Email: paa7@rambler.ru

Аннотация: Для надлежащего содержания ряда инфраструктурных объектов необходимо выполнять работы по удалению с их территории нежелательной растительности. В статье приводятся сведения о математико-статистической обработке данных, полученных в результате исследования конкурсных заявок удаления нежелательной растительности с территории различных инфраструктурных объектов, формулируются выводы и рекомендации организациям-заказчикам работ.

Ключевые слова: инфраструктура, нежелательная растительность, конкурсная заявка, обработка данных, статистика.

STATISTICAL PROCESSING OF RESULTS OF RESEARCH
OF COMPETITION APPLICATIONS FOR REMOVING UNWANTED
VEGETATION FROM THE TERRITORY OF INFRASTRUCTURAL OBJECTS

Platonov A.A.

Federal State Budgetary Educational Institution of Higher Education

«Rostov State University of Railway Engineering»

email: paa7@rambler.ru

Abstract: For proper maintenance of a number of infrastructure facilities, it is necessary to carry out work to remove unwanted vegetation from their territory. The article provides information on the mathematical and statistical processing of data obtained as a result of the study of competitive applications for the removal of unwanted vegetation from the territory of various infrastructure facilities, formulates conclusions and recommendations to organizations-customers of works.

Keywords: infrastructure, unwanted vegetation, tender, data processing, statistics.

Введение

В августе-сентябре 2019 г. нами был проанализирован определённый объём информации, посвящённой необходимости удаления нежелательной древесно-кустарниковой растительности (НДКР) с территории различных инфраструктурных объектов [1] и размещённой на специализированном сайте сферы закупок [2], при этом целью поиска являлось формирование объективной информации о наиболее востребованных технологических операциях при выполнении работ по удалению НДКР [3].

Для реализации цели поиска нами было проанализировано более 8100 конкурсных заявок (тендеров) со следующими установленными для поиска параметрами:

– ключевое слово поиска: «растительность» (с учётом всех возможных форм указанного слова);

– дата размещения конкурсных заявки: с 2012 года (получено автоматически Официальным сайтом ЕИС).

Как было отмечено в работе [4] одним из показателей добротности выполняемой научной работы является наличие статистической обработки данных, полученных в результате соответствующих научных исследований, а первостепенной задачей для исследователя является доказательство факта достоверности полученных результатов, при этом сами статистические данные вовсе не обязаны выглядеть идеально [5].

Цель исследования

Учитывая это, целью настоящей работы является детальное описание и представление математико-статистической обработки данных, полученных в результате исследования конкурсных заявок по удалению нежелательной растительности с территории различных инфраструктурных объектов.

Для выполнения поставленной цели работы нам необходимо было выполнить следующие задачи:

1 Провести первичную статистическую обработку данных с выполнением их группировки и определением основных статистических показателей распределения.

2 Установить показатели вариации распределения, вычислить доверительные интервалы для генерального среднего и дисперсии.

3 Выполнить проверку ряда гипотез о виде исследуемого распределения.

4 Сформулировать выводы по результатам статистической обработки,

дать практические рекомендации производству.

Материалы и методы исследования.

При принятии / непринятии в анализ той или иной вышеуказанной конкурсной заявки нами учитывался целый ряд разработанных нами критериев, а именно:

- объект воздействия;
- начальная цена лота, р.;
- объём выполняемых работ, га;
- место выполнения работ.

В данной работе нами рассматривалась статистическая обработка объёма выполняемых работ (га) по удалению НДКР с территории таких объектов инфраструктуры транспорта, как полосы отвода авто- и железных дорог, газо- и нефтепроводов и высоковольтных линий (рис. 1), при этом объём выборки составлял 1470 единиц:



а

б

а – в охранной зоне трассы высоковольтной линии;

б – в полосе отвода автомобильной дороги

Рисунок 1 – Визуализация работ по удалению НДКР

Камеральная обработка полученных данных выполнялась в соответствии с общепризнанными методиками определения статистических характеристик, действующими межгосударственными и национальными стандартами, а также инструкциями.

Результаты исследования и их обсуждение

Как уже отмечалось выше, собранный нами в процессе анализа конкурсных заявок (в том числе, текстов технических заявок и смет на выполнение соответствующих работ, правил и проектов на выполнение работ по удалению растительности, иной документации) материал нуждался в определённой стати-

стической обработке и сведения полученных разрозненных данных воедино.

Опираясь на сведения, приведённые в специализированной литературе, посвящённой статистической обработке материалов научного исследования [6-8], выполним статистическую сводку полученных результатов исследования с предварительной разработкой соответствующих плана и программы статистического наблюдения, а также программы сведения данных, включающей в себя определение статистических групп и подгрупп.

Применив формулу Стёрджесса выполним приближённое определение числа групп:

$$k = 1 + 3,322 \cdot \log n = 1 + 3,322 \cdot \log 1470 = 11,52 \quad (1)$$

Примем число групп исследуемого распределения объёма выполняемых работ $k = 11$. При этом ширина интервала каждой из групп составит:

$$h = \frac{x_{\max} - x_{\min}}{k} = \frac{1278 - 2}{11} = 116 \quad (2)$$

где x_{\max} и x_{\min} – соответственно максимальное и минимальное значения группировочного признака (объёма выполняемых работ) в совокупности.

Определив границы групп и отсортировав полученный ряд по возрастанию, установим статистические параметры распределений в группах (табл. 1).

Для оценки ряда распределения (рис. 2) установим и выполним определение следующих показателей.

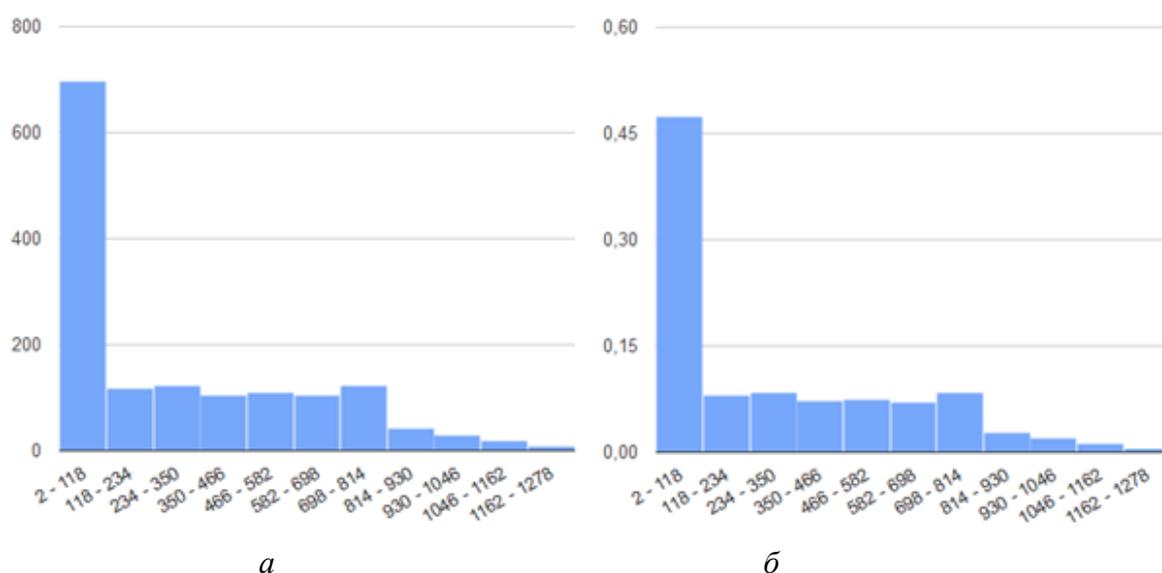


Рисунок 2 – Гистограммы частот (а) и относительных частот (б) распределений в группах объёма выполняемых работ

Таблица 1 – Статистические параметры распределений в группах

Группы ($x_i \dots$ x_{i+1})	Сере- дина интер- вала, $x_{центр}$	Кол- во, f_i	$x_i \cdot f_i$	Накоп- ленная частота, S	$ x_i - x_{cp} \cdot f_i$	$(x_i - x_{cp})^2 \cdot f_i$	Отно- сительная частота, f_i/f
2... 118	60	696	41760	696	171138,35	42080941,365	0,473
118... 234	176	116	20416	812	15067,05	1957036,655	0,0789
234... 350	292	123	35916	935	1708,27	23725,302	0,0837
350... 466	408	105	42840	1040	10721,71	1094811,021	0,0714
466... 582	524	109	57116	1149	23774,16	5185419,354	0,0741
582... 698	640	104	66560	1253	34747,6	11609575,912	0,0707
698... 814	756	122	92232	1375	54913,61	24717251,314	0,083
814... 930	872	41	35752	1416	23210,57	13139774,448	0,0279
930... 1046	988	29	28652	1445	19781,23	13493009,411	0,0197
1046... 1162	1104	18	19872	1463	14366,01	11465677,253	0,0122
1162... 1278	1220	7	8540	1470	6398,78	5849199,668	0,00476
Итого		1470	449656	-	375827,37	130616421,703	1

1. Показатели центра распределения.

В соответствии с [6-8] к показателям центра распределения относятся средняя взвешенная x (выборочная средняя), мода M_o (определяющая наиболее часто встречающееся значение признака у единиц исследуемой совокупности), медиана M_e (делящая принятую в исследование выборку на две части, а именно: половина вариант меньше медианы, половина – больше), квартили Q (разделяющие значения признака на четверти) и децили D (децентили; разделяющие значения признака на неравномерные части).

С учётом этого, определим величину *средней взвешенной (выборочной средней)*, га:

$$x = \frac{\sum x_i \cdot f_i}{\sum f_i} = \frac{449656}{1470} = 306. \quad (3)$$

Величина *моды* определится с учётом величины начала модального ин-

тервала x_o , частоты f_2 (соответствующей модальному интервалу), предмодальной частоты f_1 и послемодальной частоты f_3 :

$$M_o = x_o + h \cdot \frac{f_2 - f_1}{(f_2 - f_1) + (f_2 - f_3)}. \quad (4)$$

Приняв за начало интервала величину 2 (на интервал от 2 до 118 га приходится наибольшее количество зафиксированных нами единиц объёмов выполняемых работ удаления растительности) путём проведения вычислений получаем наиболее часто встречающееся значение объёма выполняемых работ $M_o = 65$ га.

Для установления асимметричности распределения данных об объёмах выполняемых работ удаления растительности вычислим медиану как более устойчивую к воздействию отклоняющихся данных характеристику, в том числе при возможном наличии в принятой нами в исследование выборке нехарактерных «выбросов».

Для принятого нами интервального ряда распределения медианным является интервал от 2 до 118 га, так как именно в этом интервале накопленная частота $S = 696$ больше медианного (а именно: первого интервала, накопленная частота которого превышает половину общей суммы частот) номера. Тогда, га:

$$M_e = x_o + \frac{h}{f_{me}} \cdot \left(\frac{\sum f_i}{2} - S_{me-1} \right) = 118 + \frac{116}{116} \cdot \left(\frac{1470}{2} - 696 \right) = 157. \quad (5)$$

Расчёт выявил, что 50 % единиц совокупности объёмов выполняемых работ удаления НДКР по величине будут меньше 157 га.

Установим величину *квартилей*, определяющих такие значения исследуемого признака в ранжированном ряду распределения, при котором 25 % единиц совокупности (объёмов выполняемых работ удаления растительности) будут меньше некоторой рассчитанной величины Q_1 , следующие 25 % будут заключены между Q_1 и некоторой величиной Q_2 , а ещё 25 % – между Q_2 и некоторой рассчитанной величиной Q_3 . Остальные 25 % единиц совокупности будут превосходить Q_3 . С учётом этого, га:

$$Q_1 = x_o + \frac{h}{f_{me}} \cdot \left(\frac{\sum f_i}{4} - S_{me-1} \right) = 2 + \frac{116}{696} \cdot \left(\frac{1470}{4} - 0 \right) = 63. \quad (6)$$

Таким образом, 25 % единиц совокупности объёмов выполняемых работ будут

по величине меньше 63 га. Величина Q_2 совпадает с медианой ($Q_2 = M_e = 157$ га).

$$Q_3 = x_o + \frac{h}{f_{me}} \cdot \left(\frac{3 \cdot \sum f_i}{4} - S_{me-1} \right) = 466 + \frac{116}{109} \cdot \left(\frac{3 \cdot 1470}{4} - 1040 \right) = 533. \quad (7)$$

Расчёты выявили, что ещё 25 % единиц совокупности объёмов выполняемых работ превосходят значение 533 га.

В качестве последнего показателя центра распределения рассчитаем величину *децилей (децентилей)*, разделяющие значения единиц совокупности объёмов выполняемых работ в соотношениях 10%-80%-10%, а именно, га:

$$D_1 = x_o + \frac{h}{f_{me}} \cdot \left(\frac{\sum f_i}{10} - S_{me-1} \right) = 2 + \frac{116}{696} \cdot \left(\frac{1470}{10} - 0 \right) = 27. \quad (8)$$

Расчёты выявили, что 10 % единиц совокупности объёмов выполняемых работ удаления НДСКР будут по величине меньше 27 га.

$$D_9 = x_o + \frac{h}{f_{me}} \cdot \left(\frac{9 \cdot \sum f_i}{10} - S_{me-1} \right) = 698 + \frac{116}{122} \cdot \left(\frac{9 \cdot 1470}{10} - 1253 \right) = 765. \quad (10)$$

Таким образом, ещё 10 % единиц совокупности объёмов выполняемых работ будут превосходить значение 765 га.

2. Показатели вариации.

В соответствии с [6-8] показатели вариации можно разделить на абсолютные (к которым относятся размах вариации, среднее линейное отклонение, дисперсию, несмещённую оценку дисперсии, среднее квадратическое отклонение и её оценку), а также относительные (к которым относятся коэффициент вариации, линейный коэффициент вариации и коэффициент осцилляции).

2.1 Абсолютные показатели вариации.

Установим *размах вариации*, определяемый как разность между максимальным и минимальным значениями единиц совокупности объёмов выполняемых работ, га:

$$R = x_{\max} - x_{\min} = 1278 - 2 = 1276. \quad (9)$$

Для возможности учёта различия всех единиц исследуемой совокупности определим *среднее линейное отклонение*, га:

$$d = \frac{\sum |x_i - x_{cp}| \cdot f_i}{\sum f_i} = \frac{375827,374}{1470} = 256. \quad (10)$$

Таким образом, нами было выявлено, что каждое значение ряда отличается от другого в среднем на 256 га объёмов выполняемых работ.

Определим величину меры рассеивания единиц исследуемой совокупности (а именно, *дисперсию*):

$$D = \frac{\sum (x_i - x_{cp})^2 \cdot f_i}{\sum f_i} = \frac{130616421,703}{1470} = 88854,709 . \quad (11)$$

Величина исправленной (*несмещённой*) оценки дисперсии определится как:

$$S^2 = \frac{\sum (x_i - x_{cp})^2 \cdot f_i}{\sum f_i - 1} = \frac{130616421,703}{1469} = 88915,195 . \quad (12)$$

С учётом установленного нами *среднего квадратического отклонения*, га:

$$\sigma = \sqrt{D} = \sqrt{88854,709} = 298,085 , \quad (13)$$

отметим, что каждое значение ряда объёмов выполняемых работ отличается от среднего значения $x = 306$ га в среднем на 298,085 га при следующей величине *оценки среднеквадратического отклонения*, га:

$$s = \sqrt{S^2} = \sqrt{88915,195} = 298,187 . \quad (14)$$

2.2 Относительные показатели вариации.

Определим величину меры относительного разброса значений совокупности объёмов выполняемых работ удаления растительности, показывающей, какую долю среднего значения этой величины составляет её средний разброс (*коэффициент вариации*), %:

$$v = \frac{\sigma}{x_{cp}} \cdot 100\% = \frac{298,085}{306} \cdot 100\% = 97,41. \quad (15)$$

Учитывая, что полученная нами величина коэффициента вариации $v > 70$ %, необходимо признать совокупность объёма выполняемых работ как приближающуюся к грани неоднородности, а вариацию исследуемого признака – как сильную.

Установим долю усреднённого значения признака абсолютных отклонений от средней величины (*линейный коэффициент вариации* или *относительное линейное отклонение*), %:

$$K_d = \frac{d}{x_{cp}} \cdot 100\% = \frac{256}{306} \cdot 100\% = 83,66. \quad (16)$$

Выявим относительную колеблемость крайних значений совокупности объёмов выполняемых работ удаления растительности вокруг средней (*коэффициент осцилляции*), %:

$$K_r = \frac{R}{x_{cp}} \cdot 100\% = \frac{1276}{306} \cdot 100\% = 416,99. \quad (17)$$

3. Показатели формы распределения.

В соответствии с [6-8] к показателям формы распределения относятся:

– моментный коэффициент асимметрии A_s (положительная величина которого указывает на наличие правосторонней асимметрии);

– средняя квадратическая ошибка коэффициента асимметрии S_{A_s} (с помощью которой оценивается существенность показателя асимметрии: при $|A_s|/S_{A_s} < 3$ асимметрия принимается как несущественная, а её наличие объясняется влиянием различных случайных обстоятельств);

– показатель эксцесса (островершинности) E_x (определяющий положение вершины исследуемого распределения объёмов выполняемых работ удаления растительности относительно вершины кривой нормального распределения, при этом $E_x > 0$ указывает на более островершинное по сравнению с нормальным распределение);

– средняя квадратическая ошибка коэффициента эксцесса S_{E_x} (оценивающая отклонение исследуемого распределения от нормального; при $E_x/S_{E_x} > 3$ отклонение считается существенным).

В целом симметричным является распределение, в котором частоты любых двух вариантов, равностоящих в обе стороны от центра распределения, равны между собой. Расчёт статистических параметров формы распределения проводим в таблице 2.

С учётом вышеизложенного, оценим существенность показателя асимметрии, определив моментный коэффициент асимметрии A_s :

$$A_s = \frac{\sum (x_i - x_{cp})^3 \cdot f_i}{f_i \cdot \sigma^3} = \frac{36785490966,693}{1470 \cdot 298,085^3} = 0,945, \quad (18)$$

а также среднюю квадратическую ошибку коэффициента асимметрии S_{As} :

$$S_{As} = \sqrt{\frac{6 \cdot (k-2)}{(k+1) \cdot (k+3)}} = \sqrt{\frac{6 \cdot (11-2)}{(11+1) \cdot (11+3)}} = 0,567 . \quad (19)$$

Таблица 2 – Статистические параметры расчёта центральных моментов

Группы ($x_i \dots x_{i+1}$)	Середина интервала, $x_{центр}$	Кол-во, f_i	$(x_i - x_{cp})^3 \cdot f_i$	$(x_i - x_{cp})^4 \cdot f_i$
2-118	60	696	-10347216831,292	2544260957123,3
118-234	176	116	-254196429,04853	33017176446,809
234-350	292	123	-329507,32867648	4576341,239632
350-466	408	105	111792866,34289	11415344496,281
466-582	524	109	1130999928,5346	246684164004,4
582-698	640	104	3878893572,6478	1295983200574,5
698-814	756	122	11125520662,38	5007725512621,2
814-930	872	41	7438578271,4342	4211065183833,5
930-1046	988	29	9203737761,0487	6277975964594,9
1046-1162	1104	18	9150889611,5767	7303430825613,9
1162-1278	1220	7	5346821060,3967	4887590965293,7
Итого		1470	36785490966,693	31819153870944

Расчёты выявили, что в анализируемом ряду распределения объёмов выполняемых работ удаления растительности наблюдается правосторонняя ($A_s = 0,945 > 0$) и несущественная ($0,945 / 0,567 = 1,67 < 3$) асимметрия.

Установим величину показателя эксцесса (островершинности) E_x :

$$E_x = \frac{\sum (x_i - x_{cp})^4 \cdot f_i}{f_i \cdot \sigma^4} - 3 = \frac{31819153870944}{1470 \cdot 298,085^4} = -0,258 , \quad (20)$$

а также среднюю квадратическую ошибку коэффициента эксцесса S_{Ex} :

$$S_{Ex} = \sqrt{\frac{24 \cdot k \cdot (k-2) \cdot (k-3)}{(k+1)^2 \cdot (k+3) \cdot (k+5)}} = \sqrt{\frac{24 \cdot 11 \cdot (11-2) \cdot (11-3)}{(11+1)^2 \cdot (11+3) \cdot (11+5)}} = 0,767 . \quad (21)$$

Расчёты выявили, что анализируемый ряд распределения объёмов выполняемых работ удаления растительности является плосковершинным ($E_x = -0,258 < 0$, т. е. линия распределения вариант данного ряда проходит ниже кривой нормального распределения) при не существенном ($S_{Ex} < 3$) отклонении от нормального распределения.

4. Показатели доверительного интервала

Для нахождения диапазона значений оцениваемой величины установим *доверительный интервал для генерального среднего*, представляющий собой диапазон, для которого можно утверждать с заданной вероятностью, что он будет содержать оцениваемый параметр:

$$\left(x_{cp} - t_{кр} \cdot \frac{s}{\sqrt{n}}; x_{cp} + t_{кр} \cdot \frac{s}{\sqrt{n}} \right). \quad (22)$$

Учитывая, что в соответствии с [6-8] фактор надёжности для 95%-го доверительного интервала $t_{кр}=1.96$, получим:

$$\left(306 - 1,96 \cdot \frac{298,187}{\sqrt{1470}}; 306 + 1,96 \cdot \frac{298,187}{\sqrt{1470}} \right). \quad (23)$$

Таким образом, с вероятностью 0,95 можно утверждать, что среднее значение оцениваемой величины (объёма выполняемых работ удаления растительности) при выборке большего объёма не выйдет за пределы следующего доверительного интервала: 290,75 ... 321,24 га.

5. Проверка гипотез о виде распределения.

5.1 Применим *критерий согласия Пирсона* для проверки гипотезы о нормальном распределении величин объёма выполняемых работ удаления растительности:

$$K = \sum \frac{(f_i - f_{p_i})^2}{f_{p_i}}, \quad (24)$$

где p_i – вероятность попадания в i -й интервал случайной величины, распределённой по гипотетическому закону, для вычисления которых применим формулу и таблицу функции Лапласа (табл. 3):

Таблица 3 – Статистические параметры расчёта критерия согласия Пирсона

Группы ($x_i \dots$ x_{i+1})	f_i	$x_1 =$ ($x_i -$ x_{cp})/ s	$x_2 =$ ($x_{i+1} -$ x_{cp})/ s	$\Phi(x_1)$	$\Phi(x_2)$	$p_i =$ $\Phi(x_2) -$ $\Phi(x_1)$	Ожидае- мая частота, $1470 \cdot p_i$	Слагае- мые стати- стики Пирсона, K_i
2- 118	696	-1,0191	-0,6301	-0,346	-0,238	0,1072	157,584	1839,601
118- 234	116	-0,6301	-0,2411	-0,238	-0,098	0,1402	206,094	39,3846

Продолжение таблицы 3

234-350	123	-0,2411	0,1479	-0,098	0,0596	0,1583	232,701	51,7158
350-466	105	0,1479	0,537	0,0596	0,2054	0,1458	214,326	55,7663
466-582	109	0,537	0,926	0,2054	0,3238	0,1184	174,048	24,3108
582-698	104	0,926	1,315	0,3238	0,4066	0,0828	121,716	2,5786
698-814	122	1,315	1,704	0,4066	0,4564	0,0498	73,206	32,5227
814-930	41	1,704	2,093	0,4564	0,4821	0,0257	37,779	0,2746
930-1046	29	2,093	2,482	0,4821	0,4938	0,0117	17,199	8,0972
1046-1162	18	2,482	2,8711	0,4938	0,498	0,0042	6,174	22,6521
1162-1278	7	2,8711	3,2601	0,498	0,4996	0,0016	2,4402	8,5205
Итого:	1470	-	-	-	-	-	-	2085,424

Установление границы критической области $K_{кр}$ при уровне значимости $\alpha = 0,05$ осуществляем по таблицам распределения χ^2 при числе групп $k = 11$ и параметре $r = 2$:

$$K_{кр} = \chi^2(k - r - 1; 0,05) = \chi^2(11 - 2 - 1; 0,05) = 15,50731. \quad (25)$$

Таким образом, полученная в таблице 3 величина критерия согласия Пирсона попадает в критическую область ($K_{набл} = 2085,424 > K_{кр} = 15,50731$), с учётом чего нами отвергается гипотеза о нормальном распределении величин объёма выполняемых работ удаления растительности (рис. 3, а).

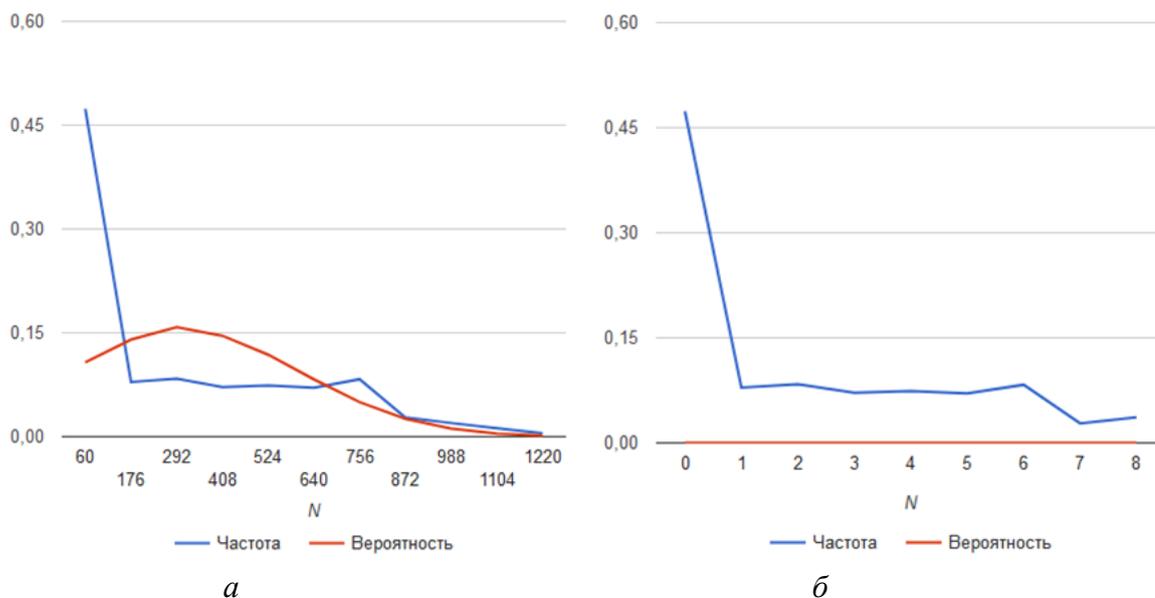
5.2 Применим *правило 3-х сигм* для проверки гипотезы о нормальном распределении величин объёма выполняемых работ удаления растительности, учитывая, что при нормальном распределении все значения случайных величин попадают в интервал:

$$(x_{ср} - 3 \cdot \sigma; x_{ср} + 3 \cdot \sigma). \quad (26)$$

Для рассматриваемого случая распределения объёма выполняемых работ этот интервал составит:

$$(306 - 3 \cdot 298,085; 306 + 3 \cdot 298,085) = (-588,255; 1200,255). \quad (27)$$

Учитывая максимальную величину группировочного признака (объёма выполняемых работ удаления растительности) $x_{max} = 1278$ га, отвергаем гипотезу о нормальном распределении величин указанного признака.



a – для нормального распределения; *б* – для распределения Пуассона

Рисунок 3 – Полигон эмпирических частот и вероятностей

5.3 Применим закон Пуассона для проверки гипотезы о нормальном распределении величин объёма выполняемых работ удаления растительности с учётом вероятности p_i попадания в i -й интервал случайной величины, распределённой по гипотетическому закону:

$$p_i = \frac{\lambda^i}{i!} e^{-\lambda} . \quad (28)$$

В качестве оценки параметра λ закона Пуассона принимаем ранее определённую нами выборочную среднюю $x_{cp} = 306$ га. В этом случае применяемый закон Пуассона будет иметь следующий вид:

$$p_i = \frac{306^i}{i!} e^{-306} . \quad (29)$$

Установление границы критической области $K_{кр} = 16,8945$ при уровне значимости $\alpha = 0,05$ осуществляем по таблицам распределения χ^2 при числе групп $k = 11$ и параметре $r = 1$ (параметр λ).

С учётом того, что полученное (путём расчёта по формуле Пуассона вероятностей p_i появления ровно i событий в n случаях) наблюдаемое значение статистики Пирсона попадает в установленную нами критическую область

($K_{набл} = 2,30993858E+35 > K_{кр} = 16,8945$), отвергаем принятую нами основную гипотезу, а именно: величины объёма выполняемых работ удаления растительности распределены не по закону Пуассона (рис. 3, б).

5.4 Проверка гипотезы о *показательном распределении* генеральной совокупности.

Принимаем в качестве оценки объёма выполняемых работ показательного распределения величину, обратную выборочной средней:

$$\lambda = 1/x_{cp} = 1/306 = 0,00327 . \quad (30)$$

С учётом полученного значения λ , плотность предполагаемого показательного распределения имеет вид:

$$f(x) = 0,00327 \cdot e^{-0,00327 \cdot x}, \quad x \geq 0 . \quad (31)$$

Вероятности p_i попадания величин объёма выполняемых работ в каждый из интервалов определяются по формуле:

$$p_i = p(x_i < X < x_{i+1}) = e_i^{-\lambda \cdot x} - e_{i+1}^{-\lambda \cdot x} . \quad (32)$$

Расчёт статистических параметров проверки распределения по показательному закону проводим в таблице 4:

Таблица 4 – Статистические параметры проверки распределения по показательному закону

Группы ($x_i \dots x_{i+1}$)	f_i	f_{p_i}	$f_i - f_{p_i}$	$(f_i - f_{p_i})^2$	Слагаемые статистики Пирсона, K_i
2-118	696	460,919	235,081	55263,0578	119,898
118-234	116	315,4497	-199,4497	39780,1754	126,106
234-350	123	215,8915	-92,8915	8628,8302	39,968
350-466	105	147,7546	-42,7546	1827,955	12,372
466-582	109	101,1222	7,8778	62,06	0,614
582-698	104	69,2073	34,7927	1210,5324	17,491
698-814	122	47,365	74,635	5570,3869	117,606
814-930	41	32,4162	8,5838	73,6808	2,273
930-1046	29	22,1854	6,8146	46,4381	2,093
1046-1162	18	15,1836	2,8164	7,9323	0,522
1162-1278	7	10,3915	-3,3915	11,5024	1,107
Итого:	1470	–	–	–	440,05

Установление границы критической области $K_{кр} = 16,8945$ при уровне

значимости $\alpha = 0,05$ осуществляем по таблицам распределения χ^2 при числе групп $k = 11$ и параметре $r = 1$ (параметр λ).

Учитывая, что полученная в табл. 4 величина критерия согласия Пирсона попадает в критическую область ($K_{набл} = 440,05 > K_{кр} = 16,8945$), отвергаем гипотезу о распределении величин объёма выполняемых работ удаления растительности по показательному закону.

5.5 Выполним проверку гипотезы о *равномерном распределении* генеральной совокупности исследуемых величин объёма выполняемых работ удаления растительности. Для реализации указанной цели установим величину оценочных параметров равномерного распределения a^* и b^* являющиеся концами интервала, в котором наблюдались возможные значения принятых в исследовании величин, по зависимостям:

$$\begin{aligned} a^* &= x_{ср} - \sqrt{3} \cdot \sigma = 306 - \sqrt{3} \cdot 298,085 = -210,29 \\ b^* &= x_{ср} + \sqrt{3} \cdot \sigma = 306 + \sqrt{3} \cdot 298,085 = 822,29 \end{aligned} \quad (33)$$

При этом плотность предполагаемого равномерного распределения определится как:

$$f(x) = 1/(b^* - a^*) = 1/(822,29 - (-210,29)) = 0,000968. \quad (34)$$

Расчёт статистических параметров проверки равномерности распределения генеральной совокупности проводим в таблице 5.

Установление границы критической области $K_{кр} = 15,50731$ при уровне значимости $\alpha = 0,05$ осуществляем по таблицам распределения χ^2 при числе групп $k = 11$ и параметре $r = 2$ (параметры a и b).

Так как наблюдаемое значение статистики Пирсона попадает в критическую область ($K_{набл} = 548,5596 > K_{кр} = 15,50731$), то есть основания отвергнуть основную гипотезу о равномерности распределения генеральной совокупности исследуемых величин объёма выполняемых работ по удалению нежелательной древесно-кустарниковой растительности (НДКР) с территории различных инфраструктурных объектов.

С учётом вышеизложенных рассуждений по проверке гипотез о виде распределения исследуемого объёма выполняемых работ можно утверждать, что наиболее близко данное распределение соответствует показательному закону, что подтверждается наименьшим наблюдаемым значением статистики Пирсона $K_{набл} = 440,05$, а также полученными нами гистограммами частот и относительных частот (рис. 2) распределений в группах объёма выполняемых работ.

Таблица 5 – Статистические параметры проверки равномерности распределения генеральной совокупности

Группы ($x_i \dots x_{i+1}$)	f_i	f_{p_i}	$f_i - f_{p_i}$	$(f_i - f_{p_i})^2$	Слагаемые статисти- ки Пирсона, K_i
2-118	696	467,5229	228,4771	52201,7661	111,6561
118-234	116	165,137	-49,137	2414,448	14,6209
234-350	123	165,137	-42,137	1775,5295	10,7519
350-466	105	165,137	-60,137	3616,4627	21,8998
466-582	109	165,137	-56,137	3151,3665	19,0833
582-698	104	165,137	-61,137	3737,7368	22,6342
698-814	122	165,137	-43,137	1860,8036	11,2682
814-930	41	165,137	-124,137	15410,003	93,3165
930-1046	29	165,137	-136,137	18533,2918	112,2298
1046-1162	18	165,137	-147,137	21649,3065	131,099
1162-1278	7	0	7	49	-
Итого	1470	-	-	-	548,5596

Выводы и рекомендации.

Результаты математико-статистической обработки данных, полученных в результате исследования конкурсных заявок по удалению нежелательной растительности с территории различных инфраструктурных объектов позволили сделать следующие выводы.

Для исследованной выборки в 1470 единиц среднее значение требуемого объёма работ по удалению НДСР, указываемого в конкурсных заявках, составило 306 га при наиболее часто встречающемся значении объёма выполняемых работ в 65 га. Установлено, что около 50 % единиц совокупности будут меньше 157 га, при этом 25 % единиц совокупности будут по величине меньше 63 га, а ещё 25 % единиц совокупности превосходят значение 533 га. Выявлено приближение совокупности объёма выполняемых работ к грани неоднородности (коэффициент вариации $\nu > 70$ %) с признанием сильной вариации данного признака. Определена плосковершинность анализируемого ряда распределения с правосторонней (коэффициент асимметрии $A_s > 0$) несущественной асимметрией, наличие которой объясняется влиянием различных случайных обстоятельств. При оценке эксцесса было установлено не существенное отклонение исследуемой выборки объёма выполняемых работ от нормального распределения. В тоже время проверка гипотез вида распределения не подтвердила близость данной выборки к нормальному распределению, а также к целому ряду других, что можно объяснить недостаточным объёмом выборки и наличием нехарактерных показателей (в частности, размещением организациями-Заказчиками работ по удалению НДСР так называемых «перспективных» заявок с указанием объёмов работ на несколько лет вперёд). Вместе с тем следует

отметить, что увеличение объема генеральной совокупности и одновременное уменьшение интервала группировки полигон либо гистограмма распределения приблизится к некоторой кривой распределения, которая на наш взгляд может быть описана показательным законом. При этом с вероятностью 0,95 можно утверждать, что среднее значение оцениваемой величины (объема выполняемых работ) при выборке большего объема не выйдет за пределы следующего доверительного интервала: 290,75 ... 321,24 га. Для соответствия распределения объема выполняемых работ по удалению НДКР показательному закону целесообразно рекомендовать организациям – Заказчикам при формировании «перспективных» конкурсных заявок распределять общие указываемые объемы работ по отдельным годам.

СПИСОК ЛИТЕРАТУРЫ

1 Platonov, A. A. Visualization of volumes of works for removing unwanted vegetation from the territory of infrastructural objects / A. A. Platonov, L. N. Bogdanova // Colloquium-journal. 2020. – № 2-2 (54). – С. 143-148.

2 О системе ЕИС // Единая информационная система в сфере закупок [сайт] [2019]. – URL: <http://zakupki.gov.ru/epz/main/public/home.html> (дата обращения: 29.11.2019).

3 Минаков, Д. Е. Современные требования к текущему содержанию земельных участков полосы отвода и охранных зон транспортной инфраструктуры / Д. Е. Минаков, А. А. Платонов, Е. Ю. Минаков // В сборнике : ТРАНСПОРТ : НАУКА, ОБРАЗОВАНИЕ, ПРОИЗВОДСТВО Ростовский государственный университет путей сообщения. Воронеж, 2020. – С. 230-236.

4 Кружилин, С. Н. Методика статистической обработки результатов лесокультурных исследований с использованием ППП Mathcad / С. Н. Кружилин, Е. В. Барышникова, М. П. Мишенина // Лесотехнический журнал. 2019. – Т. 9. – № 3 (35). – С. 56-67.

5 Warne R., Lazo M., Ramos T., Ritter N. Statistical Methods Used in Gifted Education Journals, 2006-2010 // Gifted Child Quarterly. – 2012. – Vol. 56. – № 3. – pp. 134-149. doi: 10.1177/0016986212444122.

6 Гмурман, В. Е. Теория вероятностей и математическая статистика : учеб. пособие для вузов / В. Е. Гмурман. – М : Высш. шк., 2003. – 479 с.

7 Халафян, А. А. STATISTICA 6. Статистический анализ данных / А. А. Халафян. – М : Бином-Пресс, 2010. – 512 с.

8 Глинский, В. В. Статистический анализ / В. В. Глинский, В. Г. Ионин. – М: Инфра-М, 2002. – 240 с.