



**МИНИСТЕРСТВО ОБРАЗОВАНИЯ
РЕСПУБЛИКИ БЕЛАРУСЬ**

**Белорусский национальный
технический университет**

Кафедра «Горные машины»

ОСНОВЫ НАУЧНЫХ ИССЛЕДОВАНИЙ И ИННОВАЦИОННОЙ ДЕЯТЕЛЬНОСТИ

**Методические указания к практической работе
«Предварительная обработка экспериментальных данных»**

**Минск
БНТУ
2013**

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
Белорусский национальный технический университет

Кафедра «Горные машины»

ОСНОВЫ НАУЧНЫХ ИССЛЕДОВАНИЙ И ИННОВАЦИОННОЙ ДЕЯТЕЛЬНОСТИ

Методические указания к практической работе
«Предварительная обработка экспериментальных данных»
для студентов специальностей

1-36 10 01 «Горные машины и оборудование (по направлениям)»,
1-36 13 01 «Технология и оборудование торфяного производства»

Минск
БНТУ
2013

УДК 001.891+001.895(076.5)(075.8)

ББК 72я7

О-75

Составители:

Е. К. Костюкевич, Н. И. Березовский

Рецензенты:

Ф. Г. Халявкин, В. К. Мелешко

Методические указания содержат методику предварительной обработки экспериментальных данных и алгоритм решения задач статистического анализа экспериментальных данных средствами MS Excel.

© Белорусский национальный
технический университет, 2013

ВВЕДЕНИЕ

Изучение материала студент должен производить последовательно, раздел за разделом. В случае затруднений или при углубленном изучении материала следует обратиться к источникам информации, рекомендуемым преподавателем.

Отчет о работе рекомендуется выполнять в соответствии с ГОСТ 7.32–2001 «Отчет о научно-исследовательской работе». Общие требования к правилам оформления: четкость построения, логическая последовательность изложения материала, убедительность аргументов, краткость и точность формулировок, конкретность изложения результатов расчета, доказательность выводов и обоснованность рекомендаций.

Построение отчета рекомендуется делать в соответствии с нижеизложенной структурой:

1. Титульный лист.
2. Введение, в котором кратко дается состояние проблемы, выявляется необходимость и цель проведения работы.
3. Краткий анализ возможных подходов к решению поставленной задачи, выбор путей и средств достижения целей.
4. Методика исследования, описание эксперимента с указанием его цели, изложение сущности эксперимента и последовательности его проведения.
5. Оценка достоверности полученных результатов. (Производится в соответствии с методическими указаниями.)
6. Краткое заключение об основных результатах, о соответствии результатов и поставленной цели работы; возможность применения полученных результатов либо обоснование нецелесообразности продолжения исследований.
7. Список использованных источников. (Список использованных источников формируется в порядке появления ссылок на источники в тексте отчета. Нумерация источников производится арабскими цифрами (ссылки на источники указываются в тексте в квадратных скобках). Библиографические описания источников приводятся в соответствии с ГОСТ 7.1 и ГОСТ 7.82.)

Практическая работа

ПРЕДВАРИТЕЛЬНАЯ ОБРАБОТКА ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

Цель работы: изучение методики предварительной обработки экспериментальных данных, проверки соответствия распределения результатов измерения закону нормального распределения; изучение возможностей пакета MS Excel при решении задач статистической обработки экспериментальных данных.

Общие положения

Объект исследования – это объект любого характера, который изучается экспериментальным путем.

Эксперимент – это специальным образом спланированная и организованная процедура изучения некоторого объекта исследования, при которой на этот объект оказывают запланированные воздействия и регистрируют его реакции на эти воздействия.

Экспериментальные данные – все исходные и выходные числовые данные эксперимента, сведенные в таблицу экспериментальных данных.

Обработка экспериментальных данных – различные методы построения математической модели объекта по таблице экспериментальных данных.

Основным «рабочим инструментом» эксперимента и обработки экспериментальных данных является численное значение факторов воздействия и откликов объекта исследования, т. е. число.

Числа при экспериментировании получают тремя способами:

- подсчетом;
- измерением;
- методом экспертных оценок.

Предварительная обработка результатов измерений и наблюдений необходима для того, чтобы в дальнейшем, при построении эмпирических зависимостей, эффективно использовать статистические методы и корректно анализировать полученные результаты.

Содержание предварительной обработки в основном состоит в *отсеивании грубых погрешностей* измерения или погрешностей, неизбежно имеющих место при переписывании цифрового материала или при вводе на электронный носитель информации.

Другим важным моментом предварительной обработки данных является *проверка соответствия* распределения результатов измерения *закону нормального распределения*.

Если эта гипотеза неприемлема, то следует определить, какому закону распределения подчиняются опытные данные, и, если это возможно, преобразовать данное распределение к нормальному. Только после выполнения перечисленных выше операций можно перейти к построению эмпирических формул, применяя, например, метод наименьших квадратов.

1. Генеральная совокупность и выборка.

Генеральной называют совокупность всех мыслимых наблюдений, которые могли бы быть сделаны при данном комплексе условий.

Генеральная совокупность может быть *конечной* и *бесконечной*. Данное выше определение генеральной совокупности можно считать строго обоснованным только для случаев конечных генеральных совокупностей.

Понятие бесконечной генеральной совокупности – математическая абстракция, как и представление о том, что измерить случайную величину можно бесконечное число раз. Приблизительно бесконечную генеральную совокупность можно истолковать как предельный случай конечной генеральной совокупности. В распоряжении исследователя, никогда нет генеральной совокупности, он может изучать только ее часть – выборку, причем всегда ограниченного объема. Результаты ограниченного ряда наблюдений x_1, x_2, \dots, x_n случайной величины можно рассматривать как выборку из данной генеральной совокупности.

Выборка – любое конечное подмножество генеральной совокупности, предназначенное для непосредственных исследований,

Объем – количество единиц в выборке.

Относительной частотой случайного события, называется отношение числа появлений этого события к общему числу произведенных испытаний. Мера объективной возможности случайного события называется *вероятностью случайных событий*. Относи-

тельные частоты можно истолковать как *выборочные значения* вероятностей случайных событий.

Характеристики теоретических распределений можно рассматривать как характеристики, существующие в генеральной совокупности, а характеристики эмпирических распределений – как *выборочные характеристики*.

Можно встретить и другую терминологию. Характеристики распределения вероятностей в генеральной совокупности называют *параметрами*, а выборочные (эмпирические) значения характеристик – *оценками* или *статистиками*.

Параметры обозначаются буквами греческого алфавита, а оценки – соответствующими буквами латинского алфавита.

Исходными данными при оценивании, как и при проверке любых предположений (статистических гипотез), касающихся неизвестного распределения случайной величины, могут быть лишь только те результаты наблюдений, которые были получены в ходе проведения опытов (на выборке ограниченного объема). Причем предварительная обработка экспериментальных данных обычно начинается с подсчета тех или иных функций от результатов наблюдений (статистик).

Оценивание – определение приближенного значения неизвестного параметра генеральной совокупности по результатам наблюдений.

К оценкам предъявляются требования состоятельности, несмещенности, эффективности.

Состоятельная оценка – оценка, сходящаяся по вероятности к значению оцениваемого параметра при безграничном возрастании объема выборки.

Несмещенная оценка – оценка, математическое ожидание которой равно значению оцениваемого параметра.

Оценка параметра называется **эффективной**, если среди прочих оценок того же параметра она обладает наименьшей дисперсией.

2. Вычисление характеристик эмпирических распределений (выборочных характеристик).

Здесь и в дальнейшем речь идет только о непрерывно распределенных случайных величинах.

Пусть имеется ограниченный ряд наблюдений x_1, x_2, \dots, x_n случайной величины. Среднее значение наблюдаемого признака определяется по формуле

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad (1)$$

где n – количество x_i значений выборки или объем выборки;

x_i – результат измерения i -й единицы.

Таким образом, \bar{x} представляет собой эмпирическое или выборочное среднее. Если вычислено среднее, то легко найти отклонение каждого наблюдения от среднего:

$$d_i = x_i - \bar{x}. \quad (2)$$

Величину

$$S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (3)$$

называют **дисперсией** или вторым центральным моментом эмпирического распределения $S^2 = m_2$.

В случае одномерного эмпирического распределения произвольным моментом порядка k называется сумма k -х степеней отклонений результатов наблюдений от произвольного числа c , деленная на объем выборки n :

$$m_k = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^k, \quad (4)$$

где k может принимать любые значения натурального ряда чисел.

Первый центральный момент

$$m_1 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}). \quad (5)$$

Второй центральный момент

$$m_2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 . \quad (6)$$

Несмещенную оценку для \bar{S}^2 (или σ^2 - дисперсия теоретического распределения) можно найти по формуле

$$\bar{S}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 . \quad (7)$$

Выборочные *среднеквадратические отклонения* соответственно могут быть найдены по формулам:

$$S = \sqrt{S^2}; \bar{S} = \sqrt{\bar{S}^2} . \quad (8)$$

Из других моментов чаще всего используют моменты третьего и четвертого порядка:

$$m_3 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3 , \quad (9)$$

$$m_4 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4 . \quad (10)$$

Выборочное значение коэффициента вариации V , являющееся мерой относительной изменчивости наблюдаемой случайной величины в %, определяют по формуле

$$V = \frac{\bar{S}}{\bar{X}} \cdot 100\% . \quad (11)$$

Для нормальных и близких к нормальному распределений показатель V служит индикатором однородности выборочных наблюде-

ний: принято считать, что при выполнимости неравенства $V \leq 33\%$ выборка является количественно однородной по данному признаку.

Выборочные значения характеристик распределения имеет смысл вычислять только в случае, если выборка является случайной. Обычно на практике наблюдаемые значения x_1, x_2, \dots, x_n величины случайные и отклонения их от среднего значения обусловлены погрешностями измерения и т. д. В свою очередь, погрешности – результат действия многих факторов.

Если имеет место такой редкий случай, когда в распоряжении исследователя имеется вся генеральная совокупность и необходимо сделать из нее выборку, то используют один из методов рандомизации (случайного выбора).

3. Отсев грубых погрешностей.

Можно встретить большое количество различных рекомендаций для проведения отсева грубых погрешностей наблюдения (аномальных значений). Рассмотрим наиболее простой метод отсева грубых погрешностей. Если в распоряжении экспериментатора имеется выборка небольшого объема, то можно воспользоваться методом вычисления *максимального относительного отклонения*:

$$\tau = \frac{|x_{\min(\max)} - \bar{x}|}{\bar{S}} \leq \tau_{1-\alpha}, \quad (12)$$

где $x_{\min(\max)}$ - крайний (наибольший или наименьший) элемент выборки, по которой подчитывается \bar{x} , \bar{S} и τ , вычисленной при доверительной вероятности $p = 1 - \alpha$.

Таким образом, для выделения аномального значения вычисляют m , которое затем сравнивают с табличным значением $\tau_{1-\alpha}$.

Если это неравенство $\tau < \tau_{1-\alpha}$ соблюдается, то наблюдение не отсеивают, если не соблюдается, то наблюдение исключают. После исключения того или иного наблюдения или нескольких наблюдений характеристики эмпирического распределения должны быть пересчитаны по данным сокращенной выборки. Квантили распределения статистики τ при уровнях значимости $\alpha = 0,10$, $\alpha = 0,05$, $\alpha = 0,025$, $\alpha = 0,01$ или доверительной вероятности $p = 1 - \alpha = 0,90$;

0,95; 0,975; 0,99 даны в приложении А (табл. А1). На практике обычно используют уровень значимости $\alpha = 0,05$ (результат получается с 95%-й доверительной вероятностью).

Процедуру отсева нужно повторить и для следующего по абсолютной величине максимального относительного отклонения, но предварительно необходимо пересчитать \bar{x} и \bar{S} для выборки нового объема ($n-1$).

4. Полигон и гистограмма частот распределения.

Если полученные экспериментальные данные разделить на классы, то можно построить полигон и гистограмму частот. Разбиение на классы можно выполнить по правилу Штюргеса с округлением полученного значения до ближайшего целого числа. Число классов определяется по формуле

$$k \approx 1 + 3,32 \cdot \lg(n). \quad (13)$$

Далее определяют размах варьирования:

$$R = x_{\max} - x_{\min}. \quad (14)$$

Определяют ширину интервала

$$h = \frac{R}{k}. \quad (15)$$

Затем устанавливают границы интервалов и подсчитывают число попаданий случайной величины в каждый из выбранных интервалов (абсолютные частоты B_j), для этого значения экспериментальных данных просматривают по порядку от первой до последней строчки, и при чтении каждого результата соответствующую метку (точку или черточку) заносят в тот класс, к которому относится данное наблюдение. Каждая метка соответствует одному значению из выборки.

Затем определяют относительные частоты попаданий в j -й интервал (класс) как (B_j/n) и относительные накопленные частоты как $\Sigma(B_j/n)$.

Для проверки, сумма V_j равна количеству экспериментальных данных (опытов) n .

Гистограмма и полигон распределений являются графическим отображением частот, которые, в свою очередь, представляют собой оценки плотностей вероятностей.

Кумулятивная линия – график накопленных частот, в свою очередь оценивающих функцию распределения $F(x)$ в точке x . Многие наблюдения в природе при такой обработке дают колоколообразные полигоны распределения.

Если распределение случайной величины подчиняется определенному закону и может быть хотя бы приближенно описано кривой $y = ae^{-bx^2}$, то такое распределение называют **нормальным**. Так как к коэффициентам a и b предъявляется только одно требование, а именно: $a, b > 0$, то можно говорить о семействе кривых нормального распределения. С увеличением коэффициента a кривая «вытягивается» в высоту; при увеличении коэффициента b кривая «сплющивается».

Нормальное распределение обладает и другими важными свойствами, которые позволяют считать это распределение основой математической статистики. Рассмотрим эти свойства.

1. Ордината y , которая определяет высоту кривой для каждой точки оси Ox (абсциссы), представляет собой плотность вероятности некоторого значения переменной x и определяется следующей формулой:

$$y = f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad (16)$$

$$-\infty < x < +\infty, \sigma > 0,$$

где σ – среднеквадратическое отклонение теоретического распределения;

μ – среднее значение (математическое ожидание) теоретического распределения.

Из формулы (16) следует, что нормальное распределение полностью определяется величинами μ и σ ($\pi = 3,141593\dots$ и $e = 2,718282\dots$ – математические постоянные).

Математическое ожидание μ определяет положение кривой распределения относительно оси Ox .

Среднеквадратическое отклонение σ определяет форму кривой. Чем больше σ (разброс данных), тем кривая становится более полой (ее основание более широкое).

2. Кривая нормального распределения *симметрична* относительно среднего значения.

3. Максимум ординаты кривой

$$y_{\max} = \frac{1}{\sigma\sqrt{2\pi}}, \quad (17)$$

что при $\sigma = 1$ составляет примерно 0,4. Если $x \rightarrow \pm \infty$, то $y \rightarrow 0$ (асимптотически). Другими словами, очень большие и очень малые значения переменной x маловероятны.

4. Примерно 2/3 всех наблюдений лежит в площади, отсекаемой перпендикулярами к оси Ox ($\mu \pm \sigma$). При большом объеме выборки примерно 90 % всех наблюдений лежит между $-1,64\sigma$ и $+1,64\sigma$. Границы $-0,675\sigma$ и $+0,675\sigma$ называют *вероятными отклонениями*: в этом интервале находится около 50 % всех наблюдений.

Для нормального распределения среднее, мода и медиана совпадают.

Медианой выборки является среднее значение из всего упорядоченного набора значений.

Модой выборки называется значение, которое встречается большее число раз в выборке.

Для статистических методов построения эмпирических зависимостей очень важно, чтобы результаты наблюдений подчинялись нормальному закону распределения, поэтому проверка нормальности распределения – основное содержание предварительной обработки результатов наблюдений.

5. Проверка гипотезы нормальности распределения.

1. Среднее абсолютное отклонение.

Для небольших выборок ($n < 120$) можно найти простые рекомендации по проверке нормальности распределения. Для этого необходимо вычислить среднее абсолютное отклонение (CAO) по формуле

$$\bar{N}\bar{A}\bar{I} = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|. \quad (18)$$

Для выборки, имеющей приближенно нормальный закон распределения, должно быть справедливо выражение

$$\left| \frac{\bar{N}\bar{A}\bar{I}}{\bar{S}} - 0,7979 \right| < \frac{0,4}{\sqrt{n}}. \quad (19)$$

Пользуясь CAO, можно также с 95%-й доверительной вероятностью оценить μ (среднее значение теоретического распределения) по \bar{x} :

$$\mu = \bar{x} \pm (0,71 \div 0,6) \cdot \text{CAO}. \quad (20)$$

Коэффициент $(0,71 \div 0,6)$ зависит от величины выборки n (в данном случае $n = 15 \div 20$) и $1 - \alpha = 0,95$. Коэффициенты для определения 95%-х доверительных границ для среднего значения по CAO приведены в приложении А (табл. А2).

2. Размах варьирования R .

Быструю проверку гипотезы нормальности распределения для сравнительно широкого класса выборок $3 < n < 1000$ можно выполнить, используя размах варьирования R . Подсчитываем отношение

$$\frac{R}{S}$$

и сопоставляем с критическими верхними и нижними границами этого отношения, приведенными в табл. А3 приложения А).

$$k_i \leq \frac{R}{S} \leq k_a. \quad (21)$$

Если $\frac{R}{S}$ меньше нижней или больше верхней границы, то нормального распределения нет. Особенно важно, чтобы это условие соблюдалось при $\alpha = 0,10$ (10%-й уровень значимости).

3. Показатели асимметрии и эксцесса.

Некоторое представление о близости эмпирического распределения к нормальному может дать анализ показателей асимметрии и эксцесса. Показатель асимметрии можно определять по формуле

$$g_1 = \frac{m_3}{m_2^{3/2}}. \quad (22)$$

Для симметричных распределений $m_3 = 0$ и $g_1 = 0$. Для нормального распределения $m_4 / m_2^2 = 3$.

Для удобства сравнения эмпирического распределения и нормального в качестве показателя эксцесса принимают величину

$$g_2 = \frac{m_4}{m_2^2} - 3. \quad (23)$$

Несмещенные оценки для показателей асимметрии G_1 и эксцесса G_2 определяют соответственно по формулам:

$$G_1 = \frac{\sqrt{n(n-1)}}{n-2} g_1; \quad (24)$$

$$G_2 = \frac{n-1}{(n-2)(n-3)} [(n+1)g_2 + 6]. \quad (25)$$

Для проверки гипотезы нормальности распределения следует также вычислить среднеквадратические отклонения для показателей асимметрии и эксцесса соответственно:

$$S_{G_1} = \sqrt{\frac{6n(n-1)}{(n-2)(n+1)(n+3)}}; \quad (26)$$

$$S_{G_2} = \sqrt{\frac{24n(n-1)^2}{(n-3)(n-2)(n+3)(n+5)}}. \quad (27)$$

Если выполняются условия $|G_1| \leq 3S_{G_1}, |G_2| \leq 5S_{G_2}$, то гипотеза нормальности исследуемого распределения может быть принята.

4. По критерию χ^2 (хи-квадрат)

Рассмотрим методику проверки гипотезы нормальности распределения по χ^2 критерию. Применение критерия χ^2 предполагает также использование свойств так называемого *стандартного нормального распределения*, которое имеет вид:

$$y = f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \cong 0,4 e^{-\frac{z^2}{2}}. \quad (28)$$

Расчеты выполняют в табличной форме. Значения χ^2 определяют по формуле

$$\chi^2 = \sum_{j=1}^{n_{\text{кк}}} \frac{(B_j - E_j)^2}{E_j}, \quad (29)$$

где B_j – наблюдаемая частота;

E_j – ожидаемая по стандартному нормальному распределению частота.

$$E_j = f(z_j) \cdot k', \quad (30)$$

$$k' = \frac{n \cdot n_{\text{кк}}}{S_{\text{табл}}}, \quad (31)$$

где $f(z_j)$ – уравнение кривой стандартного нормального распределения:

$$f(z_j) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z_j^2}{2}}; \quad (32)$$

z_j – степень функции кривой нормального распределения:

$$z_j = \frac{|x_j - \bar{x}_{\text{ожид}}|}{\bar{s}_{\text{ожид}}}; \quad (33)$$

$\bar{x}_{\text{ожид}}$ – ожидаемое среднее значение наблюдаемого признака:

$$\bar{x}_{\text{ожид}} = \frac{1}{n} \sum_{j=1}^{n_{\text{кл}}} B_j x_j; \quad (34)$$

$\bar{s}_{\text{ожид}}$ – ожидаемая дисперсия:

$$\bar{s}_{\text{ожид}} = \sqrt{\frac{\sum_{j=1}^{n_{\text{кл}}} B_j x_j^2 - \frac{(\sum_{j=1}^{n_{\text{кл}}} B_j x_j)^2}{n}}{n-1}}; \quad (35)$$

$n_{\text{кл}}$ - число классов (интервалов).

Полученное значение χ^2 сравнивают с табличным или критическим значением $\chi^2_{(n_{\text{св}}, \alpha)} = \chi^2_{\text{св}}$ (приложение А, табл. А4, где значения уровней значимости приведены в %).

Число степеней свободы ν определяют по формуле

$$\nu = n_{\text{кл}} - 1 - r, \quad (36)$$

где r – число параметров распределения (для нормального распределения $r = 2$, так как оцениваются два параметра \bar{x}, \bar{S}).

Гипотеза нормальности распределения принимается в случае выполнения условия $\chi^2 \leq \chi_{кр}^2$.

Методика проверки нормальности распределения по показателям асимметрии и эксцесса очень хорошо иллюстрирует использование моментов, а также удобна при использовании компьютерных технологий.

Для практического применения (особенно при расчетах с использованием компьютерных технологий) рекомендуются в основном две методики: по размаху варьирования и по χ^2 -критерию, причем первая служит для быстрой «прикидочной» проверки, а вторая – для основательной проверки нормальности распределения.

В настоящее время обработке экспериментальных данных существенно облегчают современные компьютерные технологии, современное программное обеспечение. Например, электронные таблицы MS Excel.

Особенности использования средств инструмента «Описательная статистика» в надстройке «Пакет анализа» MS Excel

В состав MS Excel входит надстройка «Пакет анализа», которая содержит 19 статистических процедур и около 50 функций.

Для анализа данных с помощью средств этого пакета следует указать входные данные и выбрать параметры; в итоге расчет будет выполнен с помощью подходящей статистической или инженерной макروفункции, а результат будет помещен в выходной диапазон. Для доступа к этим инструментам необходимо в меню «Данные» нажать кнопку «Анализ данных». Если кнопка «Анализ данных» недоступна, необходимо загрузить надстройку «Пакет анализа».

Инструмент «Описательная статистика» (вместе с инструментом «Гистограмма», алгоритм использования которого будет описан далее) является наиболее часто используемым из всего «Пакета анализа», поскольку быстро и просто вычисляет основные статистические характеристики одномерных выборок.

В «Пакете анализа» инструмент «Описательная статистика» используется для генерации одномерного статистического отчета, который включает ряд показателей положения, вариации и формы распределения признаков выборочной и генеральной совокупностей, а также среднюю и предельную ошибки выборки для средней.

После выбора инструмента «Описательная статистика» в появившемся диалоговом окне инструмента (рис. 1) задаются следующие параметры:

1. Поле *Входной интервал* – вводится ссылка на диапазон ячеек, содержащих значения анализируемого признака или признаков. В качестве входного интервала может быть указан диапазон, который содержит ряды значений сразу нескольких анализируемых признаков. В таком случае показатели «Описательной статистики» будут рассчитаны для каждого ряда и представлены в единой таблице в виде отдельных столбцов.

2. Переключатель *Группирование: по столбцам/строкам* – устанавливается в положение по столбцам или по строкам в зависимости от того, в каком направлении располагаются анализируемые данные во входном диапазоне – вертикальном (по столбцам) или горизонтальном (по строкам).

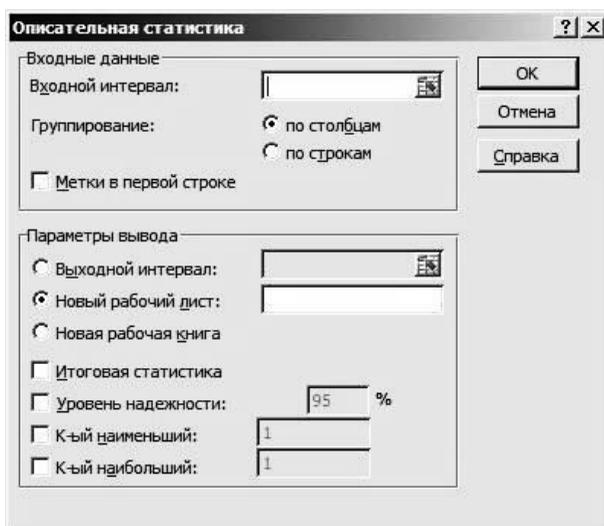


Рис. 1. Диалоговое окно инструмента «Описательная статистика»

3. Флажок *Метки* в первой строке – устанавливается в активное состояние, если первая строка во входном диапазоне содержит заголовки. Если заголовки отсутствуют, поле не активизируется. В этом случае будут автоматически созданы стандартные названия для данных выходного диапазона.

4. Поле *Выходной интервал* – вводится ссылка на ячейку заголовка первого столбца выходной результативной таблицы. Размер выходного диапазона ячеек определяется автоматически. В случае возможного наложения выходного диапазона на другие данные на экране появится соответствующее сообщение.

5. Переключатели *Новый рабочий лист* и *Новая рабочая книга* – устанавливаются в активное положение при необходимости открытия соответственно нового листа или новой книги. В новом листе результаты анализа располагаются начиная с ячейки A1, в новой книге – на первом листе начиная с ячейки A1.

6. Флажок *Итоговая статистика* – устанавливается в активное состояние, если для данных входного диапазона необходимо произвести расчет основных показателей.

7. Флажок *Уровень надежности* – устанавливается в активное состояние, если в результативную таблицу необходимо включить строку для оценки предельной ошибки выборки с заданной доверительной вероятностью. Значение уровня надежности выражается в процентах и задается в поле напротив флажка *Уровень надежности*. Уровень надежности 95,0 % (что равносильно доверительной вероятности $p = 0,95$ или же уровню значимости $\alpha = 0,05$) фиксируется в поле автоматически.

8. Флажки *K-й наименьший* и *K-й наибольший* – активизируются, если в результативную таблицу необходимо включить строку соответственно для k -го наименьшего (начиная с минимума χ_{\min}) и k -го наибольшего (начиная с максимума χ_{\max}) значений элементов в выборке. В этом случае в поле, расположенном напротив каждого флажка, вводится число k . При $k = 1$ выходные строки будут содержать соответственно χ_{\min} и χ_{\max} .

Между терминологией инструмента «Описательная статистика» и терминами, принятыми в отечественной статистике, имеется ряд расхождений. Согласование терминологии приводится в табл. 1.

Вычисленные значения всех вышеперечисленных показателей представляются в единой результативной таблице на рабочем листе

Excel. При этом показатели могут рассчитываться сразу для нескольких рядов данных в соответствии с заданным входным диапазоном ячеек.

Следует обратить внимание на то, что расчет параметров в режиме «Описательная статистика» имеет ряд важных особенностей:

1. В качестве значений параметров: *Стандартное отклонение, Дисперсия выборки, Эксцесс, Асимметричность* – MS Excel генерирует оценки соответствующих параметров для генеральной совокупности, а не для выборки.

2. Для применения «Описательной статистики» предварительное ранжирование исходных данных не требуется: при вычислении показателей ранжирование выполняется автоматически.

Таблица 1

Статистическая интерпретация параметров описательной статистики

Параметр инструмента «Описательная статистика»	Статистический показатель	Обозначение
Среднее	Средняя арифметическая величина признака в выборке, вычисленная по несгруппированным данным	\bar{x}
Стандартная ошибка	Средняя ошибка выборки – среднее квадратическое отклонение выборочной средней \bar{x} от математического ожидания генеральной средней \bar{x}	μ_x
Медиана	Значение признака, приходящееся на середину ранжированного ряда выборочных данных	Me

Параметр инструмента «Описательная статистика»	Статистический показатель	Обозначение
Мода	Значение признака, повторяющееся в выборке с наибольшей частотой	M_o
Стандартное отклонение	Генеральное среднее квадратическое отклонение, оцененное по выборке	\bar{S}
Дисперсия выборки	Генеральная дисперсия, оцененная по выборке	\bar{S}^2
Эксцесс	Коэффициент эксцесса, оценивающий по выборке значение эксцесса в генеральной совокупности	G_2
Асимметричность	Коэффициент асимметрии, оценивающий по выборке величину асимметрии в генеральной совокупности	G_1
Интервал	Размах вариации в выборке	R
Минимум	Минимальное значение признака в выборке	X_{\min}
Максимум	Максимальное значение признака в выборке	X_{\max}
Сумма	Суммарное значение элементов выборки	ΣX_i
Счет	Объем выборки	n
Уровень надежности (95,0 %)	Предельная ошибка выборки, оцененная с заданным уровнем надежности	Δ_x

3. Появление в ячейке *Мода* индикатора ошибки #Н/Д указывает на то, что в анализируемых данных нет одинаковых значений признака. В этом случае в качестве моды *Мо* выбирается то значение признака, которое соответствует максимальной ординате теоретической кривой распределения.

4. Индикатор ошибки # ДЕЛ/0! в ячейке *Эксцесс* и/или *Асимметричность* означает, что в результативной таблице стандартное отклонение является нулевым или же заданный входной диапазон данных содержит менее четырех элементов данных.

Особенности использования средств инструмента «Гистограмма» в надстройке «Пакет анализа» MS Excel

В надстройке Excel «Пакет анализа» инструмент «Гистограмма» используется для генерации интервального вариационного ряда с равными по величине интервалами, а также для построения гистограммы и кумуляты сформированного ряда распределения.

Инструмент «Гистограмма» производит следующие действия:

- рассчитывает число интервалов по формуле (13);
- определяет величину интервала h по формуле несколько от-

личной от формулы (15): $h = \frac{R}{k-1}$;

- определяет нижние границы интервалов;
- формирует интервальный вариационный ряд в соответствии с величинами k, h ;
- рассчитывает частоты и накопленные частоты интервалов, определяя число попаданий данных в сформированные интервалы;
- строит столбиковую диаграмму частот (которая может быть преобразована в гистограмму) и кумуляту накопленных частот для полученного ряда распределения;
- генерирует для вариационного ряда выходную таблицу.

Между терминологией, генерируемой в режиме «Гистограмма» выходной таблицы, и терминами, принятыми для вариационных рядов, имеются расхождения. Согласование терминологии приводится в табл. 2.

Статистическая интерпретация терминологии
инструмента «Гистограмма»

Термин инструмента «Гистограмма»	Термин, принятый в статистике
Карманы	Интервалы вариационного ряда
Интервал карманов	Диапазон ячеек, содержащих в возрастающем порядке верхние границы интервалов
Интегральный процент	Накопленная частота, выраженная в процентах

Инструмент «Гистограмма» имеет два режима работы:

- режим автоматического формирования интервалов вариационного ряда, имеющих равную величину h ;
- режим формирования интервалов ряда в соответствии с границами, заданными пользователем. Если при этом заданные интервалы будут не равны между собой, то в сгенерированной столбиковой диаграмме частоты попадания в интервал не будут связаны с размером интервала, что не позволит правильно оценить характер распределения единиц изучаемой совокупности.

Запуск инструмента «Гистограмма» осуществляется аналогично инструменту «Описательная статистика» надстройки «Пакет анализа».

В появившемся диалоговом окне инструмента «Гистограмма» (рис. 2) задаются следующие параметры:

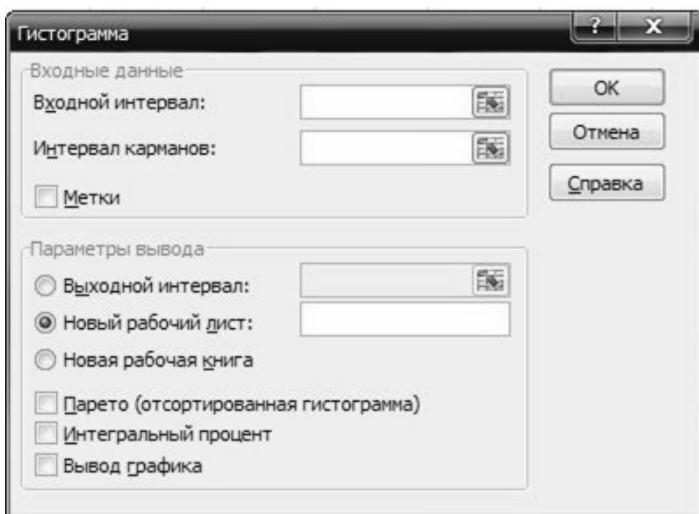


Рис. 2. Диалоговое окно инструмента «Гистограмма»

1. Поле *Входной интервал* – вводится ссылка на диапазон ячеек, содержащих значения анализируемого признака.
2. *Интервал карманов* (необязательный параметр) – вводится ссылка на диапазон ячеек, в которых задаются верхние границы интервалов. Если такой диапазон не указан, Excel осуществляет расчет нижних границ интервалов автоматически.
3. Флажок *Метки* не активизируется.
4. Поле *Выходной интервал* – вводится ссылка на ячейку заголовка первого столбца формируемой таблицы интервального вариационного ряда.
5. Переключатель *Новый рабочий лист/Новая рабочая книга* – открывает Новый рабочий лист/Новую рабочую книгу.
6. Флажок *Парето* (отсортированная гистограмма) – устанавливается в активное состояние при необходимости представить данные в порядке убывания частоты. Если флажок снят, то данные в выходном диапазоне будут приведены в порядке следования интервалов.
7. Флажок *Интегральный процент* – устанавливается в активное состояние, если необходимо рассчитать накопленные частоты (выраженные в процентах) и построить график кумуляты.

8. Флажок *Вывод графика* – устанавливается в активное состояние при необходимости автоматического построения столбиковой диаграммы.

Необходимо отметить, что инструменты «Пакет анализа» имеют определенные ограничения и иногда удобнее воспользоваться статистическими функциями или другими средствами MS Excel. Преимуществом функций перед данными средствами является то, что функции автоматически пересчитываются при любых изменениях, сделанных в выборке, тогда как эти средства необходимо выполнять заново, если выборка изменилась.

В приложении Б представлен перечень в алфавитном порядке некоторых функций MS Excel, позволяющих реализовывать обработку экспериментальных данных, непосредственно на листе электронной таблицы.

Демонстрационный пример выполнения

Исходные данные демонстрационного примера:

Данные наблюдения роста группы двадцатилетних юношей-студентов-третьекурсников (табл. 3).

1. Определяем среднее значение выборки \bar{x} по (1):

$$\bar{x} = \frac{1}{20} \cdot 3629 = 181,45.$$

2. Определяем дисперсию по (7):

$$\bar{S}^2 = \frac{1}{19} \cdot 1005695 = 529,313.$$

3. Среднеквадратичное отклонение по (8):

$$\bar{S} = 23,007.$$

4. Определяем отсев грубых погрешностей.

Для отсева погрешностей используем метод максимального относительного отклонения.

Условие отсева

$$\tau = \frac{|x_{\min(\max)} - \bar{x}|}{\bar{s}} \leq \tau_{1-\alpha}.$$

Таблица 3

Исходные данные и данные для вычисления \bar{x} , \bar{s}^2 , \bar{s}

№ п/п	x_i , см	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	183	1,6	2,403
2	170	-11,5	131,103
3	176	-5,4	29,702
4	178	-3,4	11,902
5	176	-5,4	29,702
6	180	-1,4	2,102
7	176	-5,4	29,702
8	185	3,6	12,603
9	184	2,6	6,503
10	174	-7,4	55,502
11	168	-13,5	180,903
12	174	-7,4	55,502
13	189	7,6	57,003
14	172	-9,4	89,302
15	175	-6,4	41,602
16	167	-14,5	208,803
17	179	-2,4	6,002
18	276	94,6	8939,703
19	169	-12,5	155,003
20	178	-3,4	11,902
Σ	3629,00	0	10056,95

$$x_{\max} = 276$$

$$|x_{\max} - \bar{x}| = 94,55$$

$$x_{\min} = 167$$

$$|x_{\min} - \bar{x}| = 14,45$$

Выбираем наибольший $|x_{\max/\min} - \bar{x}|$

$$\tau = \frac{94,55}{23,007}.$$

Сравниваем с квантилью распределения, зависящей от количества наблюдений (приложение А, табл. А1), при $n = 20$ и $\alpha = 0,05$

$$\tau_{1-\alpha} = 2,62;$$

$$4,11 > 2,62.$$

Следовательно условие отсева не соблюдается, *производим отсев значения 276.*

Поэтому необходимо повторить вычисления, начиная с п. 1, производя проверку присутствия грубых погрешностей, после каждого отсева значений. Результаты приведены в табл. 4.

Таблица 4

Повторные данные для вычисления \bar{x} , \bar{s}^2 , \bar{S}

№ п/п	x_i , см	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^3$	$(x_i - \bar{x})^4$
1	183	6,5	42,593	277,974	1814,146
2	170	-6,5	41,909	-271,303	1756,330
3	176	-0,5	0,224	-0,106	0,050
4	178	1,5	2,330	3,556	5,427
5	176	-0,5	0,224	-0,106	0,050
6	180	3,5	12,435	43,849	154,627
7	176	-0,5	0,224	-0,106	0,050
8	185	8,5	72,698	619,847	5285,008
9	184	7,5	56,645	426,331	3208,705
10	174	-2,5	6,119	-15,137	37,444
11	168	-8,5	71,803	-608,439	5155,717
12	174	-2,5	6,119	-15,137	37,444
13	189	12,5	156,909	1965,487	24620,305
14	172	-4,5	20,014	-89,536	400,554

№ п/п	x_i , см	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^3$	$(x_i - \bar{x})^4$
15	175	-1,5	2,172	-3,200	4,716
16	167	-9,5	89,751	-850,270	8055,187
17	179	2,5	6,382	16,124	40,733
18	169	-7,5	55,856	-417,450	3119,888
19	178	1,5	2,330	3,555766147	5,427222013
Σ	3353,00	0	646,74	1085,93	53701,81
Σ_{mod}		88,42			

Примечание: $\Sigma_{\text{mod}} = \sum_{i=1}^n |x_i - \bar{x}|$.

5. Определяем другие статистические характеристики:
– коэффициент вариации по формуле (11)

$$V = \frac{5,944}{176,47} \cdot 100\% = 3,4\%;$$

– коэффициент асимметрии по формуле (22). Для начала определяем второй центральный момент и момент третьего порядка по формуле (4):

$$m_2 = \frac{1}{18} \cdot 646,74 = 35,93;$$

$$m_3 = \frac{1}{18} \cdot 1085,93 = 60,33;$$

$$g_1 = \frac{6,33}{35,93^{3/2}} = 0,28 \neq 0.$$

Следовательно, некоторая асимметрия имеет место.

– коэффициент эксцесса по формуле (23). Для начала определяем момент четвертого порядка по формуле (4)

$$m_4 = \frac{1}{18} \cdot 53701,81 = 2983,43;$$

$$g_2 = \frac{2983,43}{35,93^2} - 3 = -0,689 < 0.$$

Имеется также и небольшой эксцесс.

Результаты вычисления выборочных характеристик, упомянутых, выше сведены в табл. 5.

Таблица 5

Выборочные характеристики распределения

\bar{x}	\bar{s}	$V, \%$	m_2	m_3	m_4	g_1	g_2
176,47	5,994	3,4	35,93	60,33	2983,43	0,28	-0,689

6. Полигон и гистограмма частот распределения.

Число классов k приблизительно можно вычислить по формуле (15)

$$k \approx 1 + 3,32 \lg n = 1 + 3,32 \lg 19 = 5,2 \approx 5.$$

Размах варьирования по формуле (14)

$$R = x_{\max} - x_{\min} = 189 - 167 = 22 \text{ см.}$$

Ширина интервалов по формуле (15)

$$h = \frac{22}{5} = 4,4.$$

В табл. 6 приведены результаты разбивки массива исходных данных на классы, вычисления частот.

Разбивка массива исходных данных на классы,
вычисление частот

№ п/п	Классы	Средины интервалов	Подсчет частот	Частоты		
				Абсолютные, B_j	Относительные, B_j/n	Относительные накопленные, $\Sigma(B_j/n)$
1	От 167 до 171,4 исключит.	169,2	****	4	0,21	0,211
2	[171,4...175,8]	173,6	****	4	0,21	0,421
3	[175,8...180,2]	178,0	*****	7	0,37	0,789
4	[180,2...184,6]	182,4	**	2	0,11	0,895
5	От 184,6 до 189 включит.	186,8	**	2	0,11	1,00
	Проверка			$\sum_{j=1}^{n_{\text{кл}}} B_j = n = 19$	$\sum_{j=1}^{n_{\text{кл}}} (B_j / n) = 1$	

Кумулятивная линия, гистограмма и полигон распределений, построенные по данным табл.6, представлены на диаграммах (рис. 3).

7. Выполняем проверку выборки на нормальность распределения по следующим критериям:

– по среднему абсолютному отклонению CAO согласно (18)

$$\tilde{N}\tilde{A}\tilde{I} = \frac{1}{19} \cdot 88,42 = 4,654;$$

– условие соответствия по (19):

$$\left| \frac{4,654}{5,994} - 0,7979 \right| < \frac{0,4}{\sqrt{19}}.$$

$$0,022 < 0,092.$$

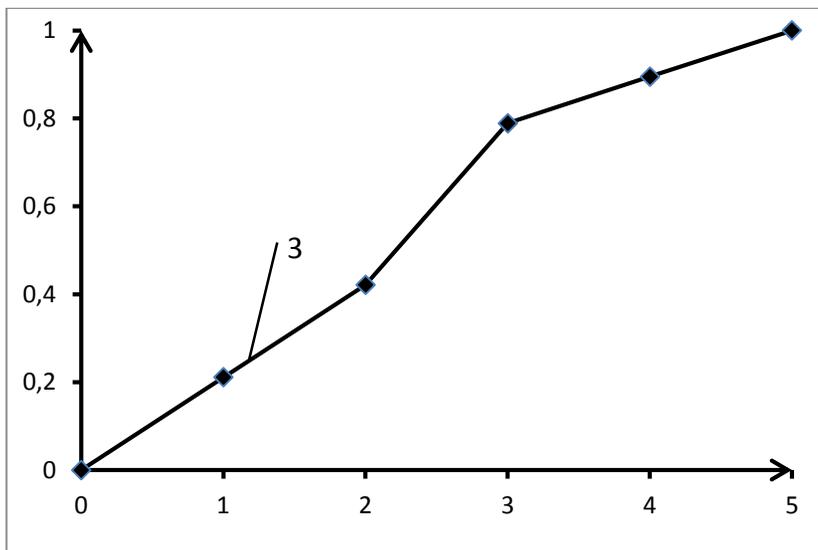
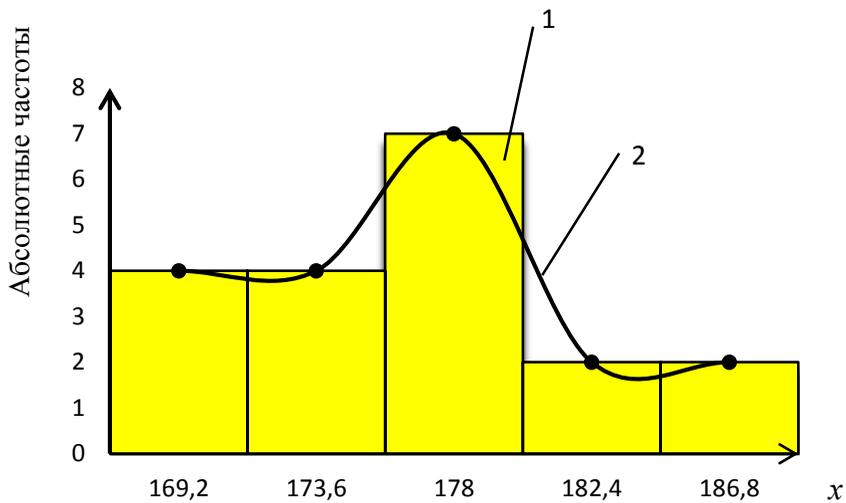


Рис. 3. Графическое изображение распределения частот:
 1 – полигон распределения; 2 – гистограмма распределения;
 3 – кумулятивная линия

Условие соответствия выполняется, следовательно, гипотеза нормальности распределения выборки данных, приведенных в табл. 4, подтверждается.

Среднее значение теоретического распределения по (20)

$$\mu = 176,47 \pm 0,62 \cdot 4,654 \text{ (см).}$$

По этим данным средний рост двадцатилетних юношей-студентов генеральной совокупности может колебаться от 173,58 до 179,36 см.

– по размаху варьирования $\frac{R}{S}$ (21)

$$\frac{R}{S} = \frac{22}{5,994} = 3,67.$$

При $n = 19$ и $\alpha = 0,10$ нижняя и верхняя границы по табл. А3 приложения А соответственно равны 3,25 и 4,27, т. е.

$$3,25 < 3,67 < 4,27.$$

Следовательно, гипотеза нормальности распределения подтверждается и по этому критерию.

– по коэффициентам асимметрии и эксцесса.

Условия:

$$1) |G_1| \leq 3S_{G_1},$$

где G_1 - несмещенная оценка для показателя асимметрии (24).

$$G_1 = \frac{\sqrt{19 \cdot (19-1)}}{19-2} \cdot 0,28 = 0,305.$$

$$2) |G_2| \leq 5S_{G_2},$$

где G_2 - несмещенная оценка для показателя эксцесса (25).

$$G_2 = \frac{19-1}{(19-2)(19-3)} [(19+1)(-0,689) + 6] = -0,515.$$

S_{G_1} , S_{G_2} - среднеквадратическое отклонения для показателей асимметрии и эксцесса соответственно (26)-(27).

$$S_{G1} = \sqrt{\frac{6 \cdot 19 \cdot (19-1)}{(19-2)(19+1)(19+3)}} = 0,524.;$$

$$0,305 < 1,572.$$

$$S_{G2} = \sqrt{\frac{24 \cdot 19 \cdot (19-1)^2}{(19-3)(19-2)(19+3)(19+5)}} = 1,014.$$

$$0,515 > 5,071.$$

Выполнение указанных условий свидетельствует, что гипотеза нормальности распределения может быть принята.

- по критерию χ^2 (свойства кривой нормального распределения) должно выполняться условие

$$\chi^2 = \sum_{j=1}^{n_{\text{кл}}} \frac{(B_j - E_j)^2}{E_j} \leq \chi_{\text{крит}}^2,$$

где B_j — абсолютная частота в классе (табл. 6);

E_j — ожидаемая частота по кривой нормального распределения (30).

Процедура вычисления приведена χ^2 в табл. 7.

$\chi_{\text{крит}}^2$ зависит от числа классов и уровня значимости.

Ожидаемое среднее значение наблюдаемого признака $\bar{x}_{\text{ожд}}$ определяется по формуле (34):

$$\bar{x}_{\text{ожд}} = \frac{33556}{19} = 176,611.$$

Ожидаемая дисперсия $\bar{s}_{\text{ожд}}$ определяется по формуле (35):

$$\bar{s}_{\text{ожд}} = \sqrt{\frac{5931784 - \frac{33556^2}{19}}{19-1}} = 5,498.$$

Процедура вычисления χ^2

№ п/п	Средины интервалов x_j	Частоты B_j	x_j^2	$B x_j$	$B x_j^2$	$x_j - \bar{x}_j$	z_j	$f(z_j)$	E_j	$B_j - E_j$	$(B_j - E_j)^2$	$\frac{(B_j - E_j)^2}{E_j}$
1	169,2	4	28628,64	676,80	114514,56	-7,41	1,35	0,16	2,78	1,22	1,49	0,54
2	173,6	4	30136,96	694,40	120547,84	-3,01	0,55	0,34	5,93	-1,93	3,74	0,63
3	178,0	7	31684,00	1246,00	221788,00	1,39	0,25	0,39	6,68	0,32	0,10	0,02
4	182,4	2	33269,76	364,80	66539,52	5,79	1,05	0,23	3,96	-1,96	3,84	0,97
5	186,8	2	34894,24	373,60	69788,48	10,19	1,85	0,07	1,24	0,76	0,58	0,47
Σ				3355,60	593178,40							2,62

k' определяется по (36)

$$k' = \frac{19 \cdot 5}{5,498} = 17,279.$$

z_j определяется по (33).

Вычисляется вектор-столбец $f(z_j)$ (32).

Вычисляется для всех классов ожидаемую частоту по кривой нормального распределения E_j по формуле (30) и т. п.

В результате расчетов полностью заполняется табл. 7 и по формуле (29) определяется расчетное значение критерия Пирсона χ^2

В табл. 7 критерий $\chi^2 = 2,62$. По табл. А4 приложения А находят табличное значение:

$$\chi_{(2;10)}^2 = 4,06 > 2,62 = \chi_{\text{ёё}}^2.$$

Таким образом, гипотеза о том, что наблюдаемые частоты распределены нормально, принимается на 10%-м уровне.

Вывод: так как условия соответствия на нормальность распределения выполняются, то распределение величин идет по нормальному закону. Гипотеза нормального распределения на достаточно «жестком» 10%-м уровне принимается. 95 % всех значений выборки варьируется в пределах от 173,58 до 179,36.

ЗАДАНИЕ

1. Изучить методику предварительной обработки экспериментальных данных.
2. Получить исходные данные у преподавателя. Исходными данными для выполнения работы являются значения в выбранном столбце (от 17 до 20 значений) соответствующей задачи.
3. Оформить отчет в соответствии со структурой указанной во введении.

Задача 1

В лаборатории ОАО «Белгорхимпрома» определено процентное содержание частиц галитовых отходов фракции крупнее 5 мм. Результаты исследования гранулометрического состава приведены в таблице ниже. Требуется определить статистические характеристики результатов определения процентного содержания частиц галитовых отходов фракции крупнее 5 мм.

Номер варианта									
1	2	3	4	5	6	7	8	9	10
28	11	21	10	29	24	29	23	26	11
18	22	22	29	18	25	28	19	18	23
15	23	13	28	26	14	28	21	14	18
23	14	24	16	25	12	19	20	23	29
11	25	25	25	25	28	24	17	29	16
22	15	25	14	14	29	13	21	15	23
33	22	26	34	13	19	23	18	24	20
36	18	28	23	23	14	20	26	11	21
11	29	19	22	22	21	25	25	23	29
20	10	10	21	21	10	23	19	21	18
22	18	29	17	23	23	25	18	35	19
22	20	19	18	25	20	23	14	22	16
19	22	21	25	18	15	21	16	18	36
20	18	23	18	23	25	20	28	7	23
14	21	22	25	7	35	9	18	25	21
22	22	23	18	25	19	15	15	21	16
18	19	21	20	25	18	20	17	21	20
-	20	22	-	23	19	-	21	17	19
-	-	25	-	19	-	-	22	-	23
-	-	-	-	21	-	-	18	-	29

Задача 2

При исследовании инженерно-геологических характеристик смеси галитовых и шламовых отходов в лаборатории ОАО «Белгорхимпрома» определена их пористость, %. Результаты лабораторных исследований приведены в таблице ниже. Требуется определить статистические характеристики полученных в лаборатории результатов исследований.

Номер варианта									
1	2	3	4	5	6	7	8	9	10
44	53	38	32	52	39	40	37	52	28
32	44	50	39	48	36	36	40	48	48
45	51	52	36	35	42	51	38	50	46
38	47	37	47	56	48	44	56	42	50
50	43	46	51	38	39	40	43	32	38
48	37	33	44	43	50	38	52	53	44
33	36	51	43	40	44	54	47	41	33
42	41	35	38	42	45	46	32	33	52
33	38	41	47	33	52	35	41	51	41
44	36	41	38	50	48	52	42	33	51
53	44	51	47	43	37	29	41	44	39
42	50	48	37	46	33	51	38	41	36
37	39	53	47	51	44	43	50	42	45
52	48	45	56	38	43	40	42	33	39
52	34	42	48	39	50	44	45	51	35
39	43	51	44	40	38	54	46	40	37
40	37	38	56	43	33	52	38	47	32
39	-	48	46	-	39	41	-	50	49
44	-	-	48	-	36	45	-	52	51
39	-	-	36	-	42	40	-	48	-

Задача 3

Лабораторией ОАО «Белгорхимпрома» проводились исследования фильтрационных свойств солеотвальных техногенных грунтов, были определены значения пористости в верхней части солеотвалов, %. Данные исследований приведены в таблице ниже. Требуется определить статистические характеристики полученных в лаборатории результатов определения пористости в верхней части солеотвалов.

Номер варианта									
1	2	3	4	5	6	7	8	9	10
30,2	25,9	29,2	26,9	23,7	21,3	26,7	25,2	22,1	22,2
27,6	21,4	23,5	35,7	26,8	28,3	33,5	24,3	29,3	28,5
29,1	21,3	21,4	26,5	25,2	33,5	25,7	26,8	34,1	27,6
28,0	27,3	28,5	28,2	21,9	25,8	32,4	28,7	36,8	29,7
31,7	29,9	24,1	26,9	21,4	22,2	32,3	36,3	35,5	24,8
29,8	25,5	26,6	27,7	38,8	32,9	36,1	31,1	32,2	33,3
27,4	28,6	28,3	27,7	25,8	27,6	26,3	32,3	31,2	25,2
28,6	29,4	29,2	29,4	36,3	30,9	29,6	27,6	27,6	28,9
26,2	28,4	35,3	29,5	28,6	29,8	33,8	26,4	32,6	31,4
27,3	35,4	30,5	21,5	28,1	30,4	28,7	33,09	31,3	30,4
29,7	36,4	35,3	36,8	35,3	37,8	22,8	24,1	30,6	27,6
31,2	33,8	27,3	29,4	26,3	31,6	29,3	33,3	34,8	28,9
27,6	22,4	27,5	28,7	26,8	28,3	31,5	37,3	37,3	28,5
28,6	24,4	23,2	29,4	26,3	30,9	27,6	27,6	25,6	28,9
26,2	24,4	31,3	21,5	28,6	29,8	24,8	26,4	31,6	31,4
29,7	27,4	25,3	27,8	24,3	27,8	32,1	36,1	29,6	27,6
30,2	25,9	24,2	26,9	27,7	24,3	26,7	22,2	22,1	25,2
27,6	28,1	39,6	12,3	36,5	36,1	21,4	29,7	27,5	28,2
42,1	23,9	27,2	28,7	-	-	23,7	22,2	21,5	36,1
-	36,8	22,2	-	-	-	26,8	-	23,5	-

Задача 4

Лабораторией ОАО «Белгорхимпрома» проводились исследования шламовых грунтов и определены значения естественного рас-
сосодержания, %. Результаты лабораторных исследований приведены в таблице ниже. Требуется определить статистические характеристики полученных результатов исследований.

Номер варианта									
1	2	3	4	5	6	7	8	9	10
44,07	43,41	42,9	49,1	28,6	59,80	37,14	49,1	41,8	42,12
43,02	53,3	39,9	59,5	35,13	62,80	37,63	58,2	44,5	55,2
54,31	52,9	47,5	49,1	32,34	61,65	38,13	54,5	52,6	46,1
45,19	55,29	45,1	50,5	39,55	60,10	38,94	57,95	42,5	45,9
43,41	51,06	50,5	48,5	32,88	59,00	39,60	54,4	50,4	42,9
45,3	48,18	50,12	45,2	36,1	62,00	40,13	60,7	53,34	46,2
57,9	57,89	47,1	58,2	36,13	59,63	40,34	58,8	64,7	50,3
54,29	48,26	46,1	49,9	44,3	59,80	40,55	58,2	61,13	63,2
41,06	49,46	44,2	59,5	37,41	61,80	40,88	52,5	54,25	45,3
58,18	56,19	55,1	46,9	31,46	62,00	41,00	57,95	63,3	40,9
58,89	51,77	54,1	62,2	45,46	60,00	41,13	57,4	54,2	54,29
42,26	52,88	47,5	62,5	41,41	59,80	41,30	50,25	41,77	55,25
54,46	56,44	53,1	57,68	40,32	60,12	41,41	50,8	42,88	40,3
58,19	56,41	50,5	58,02	36,19	61,32	41,46	65,1	53,44	46,28
51,77	51,3	43,1	58,18	34,25	60,25	41,46	61,35	52,41	43,29
38,88	52,9	42,99	58,34	39,13	60,70	41,41	65,1	61,3	54,1
43,44	39,29	46,2	48,63	40,35	58,80	41,32	58,2	52,9	50,2
39,07	51,06	49,1	58,89	34,34	58,20	41,19	48,5	44,41	53,9
51,42	-	48,4	49,12	35,9	58,50	37,65	-	53,32	42,1
40,31	-	-	-	30,5	-	40,30	-	64,32	-

Задача 5

Проводились исследования физико-механических характеристик фрезерного торфа. Результаты лабораторных исследований влажности (%) приведены в таблице ниже. Требуется определить статистические характеристики полученных в лаборатории результатов определения влажности торфа.

Номер варианта									
1	2	3	4	5	6	7	8	9	10
53	44	38	40	52	39	32	28	52	37
44	39	50	36	48	36	39	48	48	40
51	45	52	51	35	42	36	46	50	38
47	38	37	44	56	48	47	50	42	56
43	50	46	40	38	39	51	38	32	43
37	48	33	38	43	50	44	44	53	52
36	33	51	54	40	44	43	33	41	47
41	42	35	46	42	45	38	52	33	32
38	33	41	35	33	52	47	41	51	41
36	44	41	52	50	48	38	51	33	42
44	53	51	29	43	37	47	39	44	41
50	42	48	51	46	33	37	36	41	38
39	37	53	43	51	44	47	45	42	50
48	52	45	40	38	43	56	39	33	42
34	52	42	44	39	50	48	35	51	45
43	39	51	54	40	38	44	37	40	46
37	40	38	52	43	33	56	32	47	38
-	39	48	41	-	39	46	49	50	-
-	44	-	45	-	36	48	51	52	-
-	32	-	40	-	42	36	-	48	-

ПРИЛОЖЕНИЯ

ПРИЛОЖЕНИЕ А

Таблица А1

Квантили распределения максимального относительного отклонения $\tau_{1-\alpha}$

Объем выборки n	Уровни значимости			
	0,10	0,05	0,025	0,01
15	2,33	2,49	2,64	2,80
16	2,35	2,52	2,67	2,84
17	2,38	2,55	2,70	2,87
18	2,40	2,58	2,73	2,90
19	2,43	2,60	2,75	2,93
20	2,45	2,62	2,78	2,96
21	2,47	2,64	2,80	2,98
22	2,49	2,66	2,82	3,01
23	2,50	2,68	2,84	3,03
24	2,52	2,70	2,86	3,05
25	2,54	2,72	2,88	3,07

Таблица А2

Коэффициенты для определения 95%-х доверительных границ для среднего значения по САО

n	Коэффициент	n	Коэффициент	n	Коэффициент
2	12,71	9	1,00	20	0,60
3	3,45	10	0,93	25	0,53
4	2,16	11	0,87	30	0,48
5	1,66	12	0,82	40	0,41
6	1,40	13	0,78	60	0,33
7	1,21	14	0,75	120	0,23
8	1,09	15	0,71		

Таблица А3

Критические границы отношений R/S

Объем выборки n	Нижние границы k_n					Верхние границы k_b				
	Вероятность ошибки ρ									
	0,005	0,01	0,025	0,05	0,10	0,10	0,05	0,025	0,01	0,005
3	1,735	1,737	1,745	1,758	1,782	1,997	1,999	2,000	2,000	2,000
4	1,83	1,87	1,93	1,98	2,04	2,409	2,429	2,439	2,445	2,447
5	1,98	2,02	2,09	2,15	2,22	2,712	3,753	2,782	2,803	2,813
6	2,11	2,15	2,22	2,28	2,37	2,949	3,012	3,056	3,095	3,115
7	2,22	2,26	2,33	2,40	2,49	3,143	3,222	3,282	3,338	3,369
8	2,31	2,35	2,43	2,50	2,59	3,308	3,399	3,471	3,543	3,585
9	2,39	2,44	2,51	2,59	2,68	3,449	3,552	3,634	3,720	3,772
10	2,46	2,51	2,59	2,67	2,76	3,57	3,685	3,777	3,875	3,935
11	2,53	2,58	2,66	2,74	2,84	3,68	3,80	5,903	4,012	4,079
12	2,59	2,64	2,72	2,80	2,90	3,78	3,91	4,02	4,134	4,208
13	2,64	2,70	2,78	2,86	2,96	3,87	4,00	4,12	4,244	4,325
14	2,70	2,75	2,83	2,92	3,02	3,95	4,09	4,21	4,34	4,431
15	2,74	2,80	2,88	2,97	3,07	4,02	4,17	4,29	4,44	4,53
16	2,79	2,84	2,93	3,01	3,12	4,09	4,24	4,37	4,52	4,62
17	2,83	2,88	2,97	3,06	3,17	4,15	4,31	4,44	4,60	4,70
18	2,87	2,92	3,01	3,10	3,21	4,21	4,37	4,51	4,67	4,78
19	2,90	2,96	3,05	3,14	3,25	4,27	4,43	4,57	4,74	4,85
20	2,94	2,99	3,09	3,18	3,29	4,32	4,49	4,63	4,80	4,91
25	3,09	3,15	3,24	3,34	3,45	4,53	4,71	4,87	5,06	5,19
30	3,21	3,27	3,37	3,47	3,59	4,70	4,89	5,06	5,26	5,40
35	3,32	3,38	3,48	3,58	3,70	4,84	5,04	5,21	5,42	5,57
40	3,41	3,47	3,57	3,67	3,79	4,96	5,16	5,34	5,56	5,71
45	3,49	3,55	3,66	3,75	3,88	5,06	5,26	5,45	5,67	5,83
50	3,56	3,62	3,73	3,83	3,95	5,14	5,35	5,54	5,77	5,93
55	3,62	3,69	3,80	3,90	4,02	5,22	5,43	5,63	5,86	6,02
60	3,68	3,75	3,86	3,96	4,08	5,29	5,51	5,70	5,94	6,10
65	3,74	3,80	3,91	4,01	4,14	5,35	5,57	5,77	6,01	6,17
70	3,79	3,85	3,96	4,06	4,19	5,41	5,63	5,83	6,07	6,24
75	3,83	3,90	4,01	4,11	4,24	5,46	5,68	5,88	6,13	6,30
80	3,88	3,94	4,05	4,16	4,28	5,51	5,73	5,93	6,18	6,35
85	3,92	3,99	4,09	4,20	4,33	5,56	5,78	5,98	6,23	6,40
90	3,96	4,02	4,13	4,24	4,36	5,60	5,82	6,03	6,27	6,45

Процентные точки распределения χ^2 -критерия

m	Уровни значимости									
	40 %	30 %	20 %	10 %	5 %	2,5 %	1 %	0,5 %	0,1 %	0,05 %
1	0,708	1,074	1,642	2,706	3,841	5,024	6,635	7,879	10,828	12,116
2	1,833	2,408	3,219	4,605	5,991	7,378	9,210	10,597	13,816	15,202
3	2,946	3,665	4,642	6,251	7,815	9,348	11,345	12,838	16,266	17,730
4	4,045	4,878	5,989	7,779	9,488	11,143	13,277	14,860	18,467	19,997
5	5,132	6,064	7,289	9,236	11,070	12,832	15,086	16,750	20,515	22,105
6	6,211	7,231	8,558	10,645	12,592	14,449	16,812	18,548	22,458	24,103
7	7,283	8,383	9,803	12,017	14,067	16,013	18,475	20,278	24,322	26,018
8	8,351	9,524	11,030	13,362	15,507	17,535	20,090	21,955	26,125	27,868
9	9,414	10,656	12,242	14,684	16,919	19,023	21,666	23,589	27,877	29,666
10	10,473	11,781	13,442	15,987	18,307	20,483	23,209	25,188	29,588	31,420
11	11,530	12,899	14,631	17,275	19,675	21,920	24,725	26,757	31,264	33,136
12	12,584	14,011	15,812	18,549	21,026	23,336	26,217	28,300	32,909	34,821
13	13,636	15,119	16,985	19,812	22,362	24,736	27,688	29,819	34,528	36,478
14	14,685	16,222	18,151	21,064	23,685	26,119	29,141	31,319	36,123	38,109
15	15,733	17,322	19,311	22,307	24,996	27,488	30,578	32,801	37,697	39,719
16	16,780	18,418	20,465	23,542	26,296	28,845	32,000	34,267	39,252	41,308
17	17,824	19,511	21,615	24,769	27,587	30,191	33,409	35,718	40,790	42,879
18	18,868	20,601	22,760	25,989	28,869	31,526	34,805	37,156	42,312	44,434
19	19,910	21,689	23,900	27,204	30,144	32,852	36,191	38,582	43,820	45,973
20	20,951	22,775	25,038	28,412	31,410	34,170	37,566	39,997	45,315	47,498

ПРИЛОЖЕНИЕ Б

Некоторые функции MS Excel, позволяющие реализовывать
обработку экспериментальных данных

Функция	Описание
ДИСП	Оценивает дисперсию по выборке. ДИСП(число1;число2; ...)
КВАДРОТКЛ	Возвращает сумму квадратов отклонений. КВАДРОТКЛ(число1;число2;...)
МАКС	Возвращает наибольшее значение из набора значений. МАКС(число1 ;число2; ...)
МЕДИАНА	Возвращает медиану заданных чисел. МЕДИАНА(число1; число2;)
МИН	Возвращает наименьшее значение в списке аргументов. МИН(число1; число2; ...)
МОДА	Возвращает значение моды множества данных. Возвращает наиболее часто встречающееся или повторяющееся значение в массиве или интервале данных МОДА(число1; число2;...)
СКОС	Возвращает асимметрию распределения. СКОС(число1;число2; ...)
СРЗНАЧ	Возвращает среднее арифметическое аргументов. СРЗНАЧ(число1; число2; ...)
СРОТКЛ	Возвращает среднее арифметическое абсолютных значений отклонений точек данных от среднего. СРОТКЛ(число1; число2;...)
СТАНДОТКЛОН	Оценивает стандартное отклонение по выборке. СТАНДОТКЛОН(число1; число2; ...)
СЧЁТ	Подсчитывает количество чисел в списке аргументов. СЧЁТ(значение1; значение2; ...)

Функция	Описание
СЧЁТЕСЛИ	Подсчитывает количество ячеек в диапазоне, удовлетворяющих заданному условию. СЧЁТЕСЛИ(диапазон; критерий)
СЧЁТЗ	Подсчитывает количество значений в списке аргументов. СЧЁТЗ(значение1; значение2; ...)
ХИ2ОБР	Возвращает обратное значение односторонней вероятности распределения χ^2 -квадрат. ХИ2ОБР(вероятность; степени_свободы)
ЧАСТОТА	Возвращает распределение частот в виде вертикального массива. ЧАСТОТА(массив_данных; массив_интервалов)
ЭКСЦЕСС	Возвращает эксцесс множества данных ЭКСЦЕСС (число1; число2; ...)

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Львовский, Е. Н. Статистические методы построения эмпирических формул : учебное пособие для вузов / Е. Н. Львовский. – М. : Высшая школа, 1988. – 239 с.
2. Статистическая обработка данных : методические рекомендации по курсу «Основы научных исследований» / А. А. Павленко, С. А. Федоров. – Хабаровск : ТОГУ, 2008. – 24 с.
3. Обработка экспериментальных данных в MS Excel : методические указания к выполнению лабораторных работ для студентов дневной формы обучения / сост. Е. Г. Агапова, Е. А. Битехтина. – Хабаровск : ТОГУ, 2012. – 32 с.
4. Основы научных исследований : учебник для техн. вузов / В. И. Крутов [и др.] ; под. ред. В. И. Крутова. – М. : Высшая школа, 1988. – 400 с.
5. Первозванский, А. А. Математические модели в управлении производством / А. А.Первозванский. – М. : Наука, 1975. – 615 с.
6. Венцель, Е. С. Исследование операций / Е. С. Венцель. – М. : Наука, 1988. – 208 с.
7. Герасимович, А. И. Математическая статистика / А. И. Герасимович, Я. И. Матвеева. – М. : Высшая школа, 1978. – 279 с.
8. Кислов, Н. В. Учебное пособие по курсу «Основы научных исследований» для студентов специальности 0507 – «Торфяные машины и комплексы» / Н. В. Кислов, В. Т. Васильев; кол. авт. Белорусского политехнического института, кафедра «Торфяные машины». – Минск : БПИ, 1981. – 114 с.
9. Анализ статистической совокупности в программе MS Excel : методические указания и задание к лабораторной работе № 1 / сост. А. Н. Акжигитова, Н. С. Циндин, Н. Ф. Разуваева. – Пенза: Информационно-издательский центр ПГУ, 2007. – 52 с.
10. Вадзинский, Р. Статистические вычисления в среде Excel. Библиотека пользователя / Р. Вадзинский. – СПб. : Питер, 2008. – 608 с.
11. Веденева, Е. А. Функции и формулы. Excel 2007. Библиотека пользователя / Е. А. Веденева. – СПб. : Питер, 2008. – 384 с.

Учебное издание

ОСНОВЫ НАУЧНЫХ ИССЛЕДОВАНИЙ И ИННОВАЦИОННОЙ ДЕЯТЕЛЬНОСТИ

Методические указания к практической работе
«Предварительная обработка экспериментальных данных»
для студентов специальностей

1-36 10 01 «Горные машины и оборудование (по направлениям)»,
1-36 13 01 «Технология и оборудование торфяного производства»

Составители:

КОСТЮКЕВИЧ Елена Казимировна
БЕРЕЗОВСКИЙ Николай Иванович

Подписано в печать 21.03.2013. Формат 60×84 ¹/₁₆. Бумага офсетная. Ризография.
Усл. печ. л. 2,67. Уч.-изд. л. 2,09. Тираж 100. Заказ 1254.

Издатель и полиграфическое исполнение: Белорусский национальный технический университет. ЛИ № 02330/0494349 от 16.03.2009. Пр. Независимости, 65. 220013, г. Минск.