Т е м а 1. ТЕОРЕТИЧЕСКИЕ РАСПРЕДЕЛЕНИЯ. ПРОВЕРКА ДАННЫХ НА СООТВЕТСТВИЕ ЗАКОНУ НОРМАЛЬНОГО РАСПРЕДЕЛЕНИЯ

Объяснение. Различают эмпирические и теоретические распределения частот совокупности результатов наблюдений.

Эмпирическое распределение – распределение результатов измерений, полученных при изучении выборки, например распределение растений по высоте или массе, распределение поверхностных водоемов Могилевской области по содержанию нитратных форм азота и т.д. В основе эмпирического распределения лежат определенные математические закономерности, которые в генеральной совокупности, т.е. при очень большом числе наблюдений ($n \rightarrow \infty$), характеризуются некоторыми теоретическими распределениями.

На основе теоретических распределений построены статистические критерии, которые используются для проверки некоторых гипотез. Наиболее часто в исследовательской работе опираются на нормальное распределение или специальные распределения, получаемые из нормального для определенно поставленной задачи и при ограниченном числе степеней свободы (t, F, χ^2 – распределение).

Нормальное распределение. Нормальным, или гауссовым, называют распределение вероятностей непрерывной случайной величины X, которое описывается следующей функцией Y:

$$Y = \frac{1}{\sigma\sqrt{2\pi}}\ e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2}$$

где: Y – ордината кривой, или вероятность;

μ – генеральная средняя (математическое ожидание);

σ – стандартное отклонение генеральной совокупности ($n \rightarrow \infty$);

π и e – константы ($\pi \approx 3{,}14$; $e \approx 2{,}72$).

Положение и форма кривой нормального распределения полностью определяются двумя параметрами: генеральной средней μ, которая находится в центре распределения, и отклонением σ, которое измеряет вариацию отдельных наблюдений около средней. Максимум, или центр, нормального распределения лежит в точке $X = \mu$; точки перегиба кривой находятся при $X_1 = \mu - \sigma$ и $X_2 = \mu + \sigma$.

При $X \pm \infty$ кривая достигает нулевого значения (рис. 1)

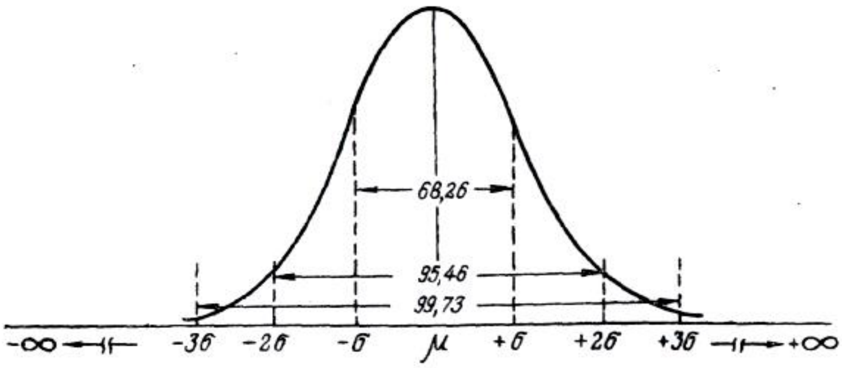

Рисунок 1 – Процент наблюдений (площадь), ограниченный кривой нормального распределения, для различных значений σ.

Для нормального распределения характерны следующие закономерности:

1. В области $\mu \pm \sigma$ лежит около 68% (68,26%) всех наблюдений;

2. Внутри пределов $\mu \pm 2\sigma$ находится около 95% (95,46%) всех значений случайной величины;

3. Интервал $\mu \pm 3\sigma$ охватывает более 99% (99,73%), следовательно, практически все значения.

Площадь под кривой, отграниченную от среднего на t стандартных отклонений, выраженную в процентах всей площади, называют статистической надежностью, или уровнем вероятности P, т.е. вероятностью появления значения признака, лежащего в области $\mu \pm t\sigma$. Вероятность того, что значение варьирующего признака находится вне указанных пределов, называется уровнем значимости P_1. Чем больше уровень вероятности, тем меньше уровень значимости, и наоборот.

Заметим, что средняя μ, дисперсия σ^2 и стандартное отклонение σ – параметры генеральной совокупности, когда $n \rightarrow \infty$. Выборочные наблюдения позволяют получать оценки этих параметров. Так, выборочная средняя $\bar{x}$ является оценкой генеральной средней μ, выборочная дисперсия S^2 - оценкой σ^2 (дисперсии генеральной совокупности), выборочное стандартное отклонение S – оценкой σ. Для достаточно больших выборок (n >20 – 30 и особенно n >100) закономерности нормального распределения, указанные выше для параметров генеральной совокупности, справедливы и для их оценок, а именно: в области $\bar{x} \pm S$ находится около 68%, внутри пределов $\bar{x} \pm 2S$ – около 95% и в интервале $\bar{x} \pm 3S$ – более 99% всех наблюдений.

Результаты различных наблюдений, полевых и вегетационных опытов чаще всего располагаются приблизительно в соответствии с симметричной кривой нормального распределения, когда частоты вариант, равно отстоящих от средней, равны между собой. Но нередко некоторые признаки растений и животных дают распределения, значительно отличающиеся от нормального, – *асимметричные*, или скошенные.

Асимметрия может быть положительной, или правосторонней, когда увеличиваются частоты правой части, и отрицательной, или левосторонней, когда увеличиваются частоты левой части вариационной кривой.

Причинами ассиметричных распределений могут быть:

1. Неправильно взятая выборка, когда в нее вошло непропорционально много (или мало) вариант с большим или меньшим их значением.

2. Действие определенных факторов, сдвигающих частоту варьирующего признака в ту или другую сторону от среднего значения.

Когда какие-либо причины благоприятствуют более частому появлению и средних и крайних значений признака, образуются так называемые положительные эксцессивные распределения, имеющие вид острой пирамиды с расширенным основанием, или отрицательные эксцессивные распределения, когда в центре их имеется не вершина, а впадина, и вариационная кривая становится двухвершинной.

Многовершинные и двухвершинные кривые в большинстве случаев указывают, что в выборку попали представители нескольких совокупностей с различными средними. Например, посеяна смесь сортов, имеются закономерные различия в плодородии почвы на отдельных частях земельного участка и т. п. В генетических работах двухвершинные и многовершинные кривые могут свидетельствовать о появлении объектов с новыми свойствами или признаками и указывать на результативность применяемого фактора.

Нормальное распределение – наиболее часто встречающийся в практике экспериментальной работы закон распределения случайной величины, т. е. величины, значение которой нельзя точно предсказать. Главная его особенность заключается в том, что он является предельным законом, к которому приближаются другие законы распределения.

t – распределение Стьюдента. Закон нормального распределения проявляется при $n > 20–30$. Если же эксперимент проводят при n=4–6, то естественно ожидать, что среди измерений различных параметров очень больших отклонений не будет. Поэтому стандартное отклонение S, подсчитанное по малой выборке, в большинстве случаев будет меньше, чем по всей генеральной совокупности σ. Следовательно, в этих случаях полагаться на критерии нормального распределения в своих выводах нельзя.

С начала XX в. в математической статистике стало разрабатываться новое направление, которое можно назвать статистикой малых выборок. Наибольшее практическое значение для экспериментальной работы имело открытое в 1908 г. английским статистиком и химиком В. Госсетом *t*-распределение, получившее название распределения Стьюдента (от англ. *student* – студент, псевдоним В. Госсета). Распределение t Стьюдента для выборочных средних определяется равенством:

$$t = \frac{\bar{x} - \mu}{S/_{\sqrt{n}}} = \frac{\bar{x} - \mu}{S_{\bar{x}}}$$

Числитель формулы означает отклонение выборочной средней от средней совокупности μ, а знаменатель является показателем, оценивающим стандартную ошибку средней генеральной совокупности. Таким образом, величина t измеряется отклонением выборочной средней $\bar{x}$ от средней совокупности μ, выраженным в долях ошибки выборки $S_{\bar{x}}$, принятой за единицу.

Максимумы частоты нормального и t-распределения совпадают, но форма t-распределения полностью зависит от числа степеней свободы. При очень малых значениях степеней свободы она принимает вид плосковершинной кривой, причем площадь, отграниченная кривой, больше, чем при нормальном распределении, а при увеличении числа наблюдений ($n > 30$) распределение t приближается к нормальному и переходит в него при $n \to \infty$.

Распределение t Стьюдента имеет важное значение при работе с малыми выборками: позволяет определить доверительный интервал, накрывающий среднюю совокупности μ, и проверить ту или иную гипотезу относительно генеральной совокупности. При этом нет необходимости знать параметры совокупности μ и σ, достаточно иметь их оценки $\bar{x}$ и S для определенного объема выборки n.

χ^2-распределение. Распределение χ^2 (хи-квадрат) впервые было рассмотрено Р. Хельмертом (1876) и К. Пирсоном (1900). Если $x_1, x_2, \ldots, X_n$ − независимые случайные величины, распределенные по стандартному нормальному закону, то распределение случайной величины
$Y= x_1^2+ x_2^2+ \ldots+ x_n^2$ имеет распределение χ^2 с n степенями свободы. Распределение χ^2 зависит от одного параметра, который называется степенью свободы. Число степеней свободы равно n-1, где n - размер выборки. Плотность распределения χ^2 выражается формулой

$$f(y) = \frac{1}{2^{\frac{n}{2}} \times \Gamma \times (\frac{n}{2})} \times y^{\frac{n}{2}-1} \times e^{-y/2} \quad \text{, где}$$

n - число степеней свободы

e - основание натурального логарифма (2.71...)

Γ - гамма-функция

F-распределение Фишера. Если из нормально распределенной совокупности взять две независимые выборки объемом n_1 и n_2 и подсчитать дисперсии S_1^2 и S_2^2 со степенями свободы $\nu_1 = n_1 - 1$ и $\nu_2 = n_2 - 1$, то можно определить отношение дисперсий:

$$F = \frac{S_1^2}{S_2^2}$$

Отношение дисперсий берут таким, чтобы в числителе была большая дисперсия, и поэтому $F \geq 1$.

Распределение F зависит только от числа степеней свободы ν_1 и ν_2.

Кривые, полученные из функции распределения для всех возможных значений F, особенно при небольшом числе наблюдений, имеют асимметричную форму – длинный «хвост» больших значений и большую концентрацию малых величин F.

Статистические методы проверки гипотез. Любая разность между средними арифметическими двух генеральных совокупностей всегда существенна. В опытной работе обычно сравнивают средние величины не генеральных, а выборочных совокупностей, поэтому приходится доказывать методами математической статистики статистическую значимость различий в действии изучаемых факторов и на основании этого делать выводы и предложения для производства.

При этом исследователь не в состоянии исключить возможность ошибки: он может отклонить истинную гипотезу или принять ошибочную. В статистическом анализе говорят, он может совершить ошибку первого или второго рода.

Вероятность совершения ошибки первого рода (то есть отклонения истинной нулевой гипотезы) называют уровнем значимости (α).

Вероятность избежать ошибки второго рода (то есть принятия ошибочной нулевой гипотезы) называют мощностью анализа ($1-\beta$).

Уровень значимости $\alpha=0,05$ означает, что в 5 случаях из 100 мы ошибочно отклоним верную нулевую гипотезу. Чем меньше значение α, тем больше должно быть отклонение выборочной оценки признака от истинного (наблюдающегося в генеральной совокупности) значения для того, чтобы привести нас к ошибочному выводу.

Использование статистических методов или критериев проверки гипотез – основа принятия решений при изучении явлений, обусловленных случайной вариацией.

Статистическая гипотеза – это научное предположение о тех или иных законах распределения рассматриваемых случайных величин, которое может быть проверено на основе выборки. Почти всегда задача заключается в проверке нулевой гипотезы (H_0) об отсутствии реального различия между фактическими и ожидаемыми наблюдениями.

Принятие H_0 гипотезы означает, что данные не противоречат предложению об отсутствии различий между фактическими и теоретическими или двумя рядами фактических распределений. Отбрасывание гипотезы означает, что эмпирические данные несовместимы с H_0 и верна другая, альтернативная гипотеза.

Справедливость H_0 проверяется вычислением статистических критериев проверки для определенного уровня значимости. Для проверки H_0 используют критерии двух видов: параметрические и непараметрические.

Параметрические критерии – это критерии, основанные на предположении, что распределение признака в совокупности подчиняется известному закону (например, закону нормального распределения). К таким показателям относятся критерии t и F, использование которых предполагает вычисление оценок параметров нормального распределения.

Непараметрические критерии – это такие, использование которых не требует вычисления оценок неизвестных параметров распределения и приближенного значения распределения признака. Этими критериями пользуются в предварительных исследованиях и в случаях, если данные не подчиняются закону нормального распределения.

Статистические показатели выборочной совокупности (выборки) являются приближенными оценками показателей (параметров) генеральной совокупности. Оценка может быть точечная (представлена одним числом) или интервальная.

Через выборочную среднюю $\bar{x}$ и ее ошибку $S_{\bar{x}}$ можно дать точечную оценку генеральной средней в виде $\bar{x} \pm S_{\bar{x}}$. Это означает, что $\bar{x}$ – оценка генеральной средней μ с ошибкой, равной $S_{\bar{x}}$.

Интервальная оценка характеризуется двумя числами – концами интервала, покрывающего оцениваемый параметр.

Доверительный интервал – это такой интервал, который с заданной вероятностью покрывает оцениваемый параметр. Центр такого интервала – выборочная оценка параметра, а пределы, или доверительные границы, определяются средней ошибкой оценки и уровнем вероятностей.

Статистические методы проверки гипотез основываются на использовании: t-критерия для оценки существенности разности выборочных средних для независимых и сопряженных выборок; τ-критерия (греч. тау) для проверки гипотезы о принадлежности «сомнительной» варианты к совокупности; критерия F для оценки существенности различий в степени вариации признаков; критерия χ^2 для оценки степени соответствия эмпирических данных теоретическим ожиданиям.

Проверка данных на соответствие закону нормального распределения

В данном разделе приведена краткая обобщенная информация о нормальном распределении данных и наиболее простых способах проверки соответствия данных закону нормального распределения (в идеализированном случае). Раздел позволяет получить только наиболее общие представления «о нормальности». Более подробно с законом нормального распределения можно ознакомиться, применив специализированную учебную литературу [Полякова, Шаброва, 2015; Елисеева, 2011; Калинина, 2013; Гмурман, 2013; Лысенко, Дмитриева, 2013 и др.].

В природе все находится в гармонии. Те или иные признаки каких-либо объектов природы, как правило, проявляются согласно следующему закону: типичными (средними) значениями признака характеризуется большая часть объектов, по мере отклонения от среднего значения признака – количество объектов, характеризующихся такими значениями признаками, постепенно сокращается. Чем больше характеристика объектов отклонена от нормы (от среднего) тем более он уникален. Тем меньше вероятность среди всего многообразия найти данный объект.

Простой пример: распределение роста у человека. Большинство людей обладают средним ростом. В то время как люди небольшого роста и высокие люди встречаются значительно реже.

Подобным образом обстоит дело и с большинством других признаков у объектов живой и неживой природы. Такое «типичное» распределение данных получило название «нормального».

Множество статистических методов основаны на использовании особенностей нормального распределения данных. Такие методы принято называть «параметрическими». К этой группе статистических методов из рассматриваемых в данном пособии относятся: расчет tкритериев Стьюдента и определение значений коэффициента линейной корреляции Пирсона. Для того, чтобы использовать данные методы сначала нужно проверить распределение значений исходных характеристик. В случае, если закон распределения не близок «нормальному», то параметрические методы не могут быть использованы.

Методы, для которых соблюдение закона нормального распределения является необязательным, получили название «непараметрические». Непараметрическими методами из рассматриваемых в учебнике являются: t-критерий Уилкоксона; Uкритерий Манна-Уитни; коэффициент ранговой корреляции Спирмена и методы кластерного анализа.

Как же осуществить проверку данных на соответствие (близость) закону нормального распределения? Существуют различные методы. В целом они могут быть подразделены на 2 группы: графические и расчетные. С использованием современного программного обеспечения и стандартных офисных компьютеров проверка нормальности распределения не составляет большого труда и занимает считанные секунды. В Excel, Statistica, SPSS и других программах существует широкий инструментарий для проверки распределения исследуемых данных: построение графиков распределения (гистограммы, нормальные вероятностные графики, ящичные диаграммы с усами), вычисление формальных критериев (Колмогорова-Смирнова, Шапиро-Уилка) и прочих характеристик распределения. В данном разделе будут рассмотрены наиболее простые методы проверки на нормальность, которые можно применить без использования специализированного программного обеспечения.

Графический способ. Среди всего множества методов наиболее распространенным и простым является метод (графический) построения диаграммы распределения данных. Суть метода заключается в том, чтобы имеющиеся значения исследуемых характеристик разбить на ряд равных диапазонов. Далее нужно посчитать количество объектов исследования вошедших в эти диапазоны и построить гистограмму, где по оси OX следует отложить указанные диапазоны значений (в порядке возрастания), а по оси OY количество объектов исследования. Если указанная диаграмма будет иметь колоколообразный вид и будет симметричной (или близкой к таковой), то считают, что распределение близко нормальному. При этом значение среднего арифметического и медианы должны быть также близкими (и моды в идеальных случаях). Такой графический способ применим в тех случаях, когда выборка представлена достаточно большим количеством объектов исследования (от нескольких десятков и больше).

Приведем пример графического метода определения соответствия распределения данных закону нормального распределения. Пусть имеются сведения о росте группы людей, состоящей из 30 человек (таблица 1).

Таблица 1 **Рост группы людей**

№	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Рост, см	147	151	152	156	157	157	161	162	162	163	164	165	166	166	167	167	167	168	169	171	171	172	172	173	176	177	178	182	183	191

Рост людей группы варьирует в диапазоне от 147 до 191 см. Произведем разбивку значений роста на удобное количество диапазонов. Указанных диапазонов должно быть достаточное количество, в противном случае диаграммы не дадут представления о характере распределения данных. Оптимальным количеством (на наш взгляд) является 9-15 диапазонов. В случае с таблицей 1 удобно использовать диапазон по 5 см. Таким образом, получится 9 диапазонов равной ширины (по 5 см).

Производим подсчет количеств объектов исследования (людей) вовлеченных в каждый диапазон. Отобразим полученный результат в виде таблицы 2.

Таблица 2 **Группы объектов исследования по значениям роста**

Группы	Частота (количество людей)
<150	1
[150;155)	2
[155;160)	3
[160;165)	5
[165;170)	8
[170;175)	5
[175;180)	3
[180;185)	2
≥185	1

Используя данные таблицы 2 построим гистограмму (рисунок 2).

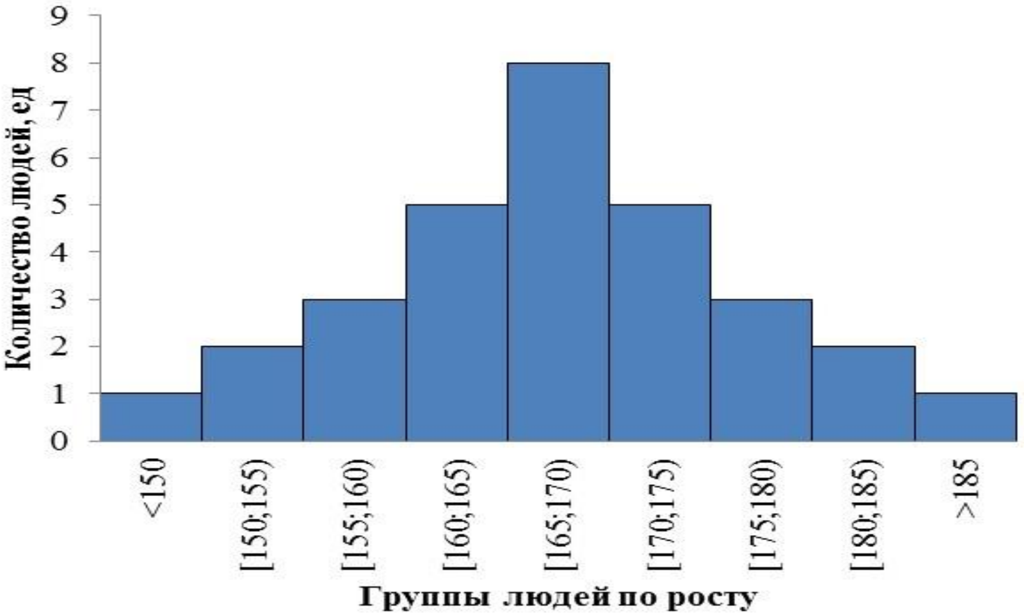

Рисунок 2. Гистограмма распределения объектов исследования

Гистограмма распределения объектов исследования близка колоколообразной форме, симметрична. Таким образом, распределение близко нормальному. Далее необходимо вычислить среднее арифметическое, медиану и моду.

Расчет среднего ($\bar{X}$), медианы (*Me*) и моды (*Mo*):

$$\bar{X} = \frac{\begin{gathered}(147+151+152+156+157+161+162+162+163+164+165+166+166+ \\ +167+167+168+169+167+171+171+172+173+176+177+178+ \\ +182+183+191)\end{gathered}}{30} = 167{,}1.$$

$$Me = \frac{167+167}{2} = 167.$$

Mo =167 (наиболее часто повторяющееся число).

Из расчетов видно, среднее арифметическое, медиана и мода практически равны, что также является одним из косвенных признаков нормального распределения данных.

Расчет асимметрии и эксцесса. Для проверки распределения на нормальность также используют метод расчета асимметрии и эксцесса. Для нормального распределения асимметрия и эксцесс равны 0. В реальности имеются некоторые отклонения от 0. Если это отклонение невелико, то можно считать, что распределении близко нормальному. Рассчитывается асимметрия (*A*) по формуле (1):

$$A = \frac{\sum(x_i - \bar{X})^3}{n б^3}, \quad (1)$$

где А – асимметрия;

x_i – значения рассматриваемой характеристики в каждом конкретном случае; $\bar{X}$ – среднее арифметическое значение; σ – стандартное отклонение; n – количество объектов исследования.

Эксцесс (E) рассчитывается по формуле (2):

$$E = \frac{\sum(x_i - \bar{X})^4}{n\sigma^4} - 3 \qquad (2)$$

После расчетов асимметрии и эксцесса полученные их значения сравниваются с показателями, именуемыми ошибками репрезентативности асимметрии (m_A) и эксцесса (m_E), которые вычисляются согласно формулам (3) и (4).

$$m_A = \sqrt{\frac{6}{n}}. \qquad (4.3)$$

$$m_E = \sqrt{\frac{6}{n}}2. \qquad (4.4)$$

Если сравниваемые значения асимметрии и эксцесса по модулю (без учета знаков ±) меньше трехкратного значения соответствующих ошибок репрезентативности, то распределение близко нормальному. Если нет – то распределение не соответствует нормальному.

Произведем расчеты асимметрии и эксцесса для рассмотренного выше примера. Для использования формул асимметрии и эксцесса необходимо помимо среднего, которое уже рассчитано, вычислить стандартное отклонение (σ).

Формулы стандартного отклонения, асимметрии и эксцесса содержат общую часть $(x_i - \bar{X})$. Для того, чтобы облегчить и упорядочить вычисления всех этих параметров воспользуемся табличной формой записи результатов вычислительных процедур (таблица 3).

Таблица 3 **Упорядоченное представление расчетов**

№	x_i	$(x_i-\bar{X})$	$(x_i-\bar{X})^2$	$(x_i-\bar{X})^3$	$(x_i-\bar{X})^4$
1	2	3	4	5	6
1	147	$(147-167{,}1)=-20$	404,0	–8120,6	163224,1
2	151	$(151-167{,}1)=-16$	259,2	–4173,3	67189,8
3	152	$(152-167{,}1)=-15$	228,0	–3443,0	51988,6
4	156	$(156-167{,}1)=-11$	123,2	–1367,6	15180,7
5	157	$(157-167{,}1)=-10$	102,0	–1030,3	10406,0
6	157	$(157-167{,}1)=-10$	102,0	–1030,3	10406,0
7	161	$(161-167{,}1)=-6{,}1$	37,2	–227,0	1384,6
8	162	$(162-167{,}1)=-5{,}1$	26,0	–132,7	676,5
9	162	$(162-167{,}1)=-5{,}1$	26,0	–132,7	676,5
10	163	$(163-167{,}1)=-4{,}1$	16,8	–68,9	282,6
11	164	$(164-167{,}1)=-3{,}1$	9,6	–29,8	92,4
12	165	$(165-167{,}1)=-2{,}1$	4,4	–9,3	19,4
13	166	$(166-167{,}1)=-1{,}1$	1,2	–1,3	1,5
14	166	$(166-167{,}1)=-1{,}1$	1,2	–1,3	1,5
15	167	$(167-167{,}1)=-0{,}1$	0,01	–0,001	0,0001
16	167	$(167-167{,}1)=-0{,}1$	0,01	–0,001	0,0001
17	167	$(167-167{,}1)=-0{,}1$	0,01	–0,001	0,0001
18	168	$(168-167{,}1)=0{,}9$	0,8	0,7	0,7
19	169	$(169-167{,}1)=1{,}9$	3,6	6,9	13,0
20	171	$(171-167{,}1)=3{,}9$	15,2	59,3	231,3
21	171	$(171-167{,}1)=3{,}9$	15,2	59,3	231,3
22	172	$(172-167{,}1)=4{,}9$	24,0	117,6	576,5
23	172	$(172-167{,}1)=4{,}9$	24,0	117,6	576,5
24	173	$(173-167{,}1)=5{,}9$	34,8	205,4	1211,7
25	176	$(176-167{,}1)=8{,}9$	79,2	705,0	6274,2
26	177	$(177-167{,}1)=9{,}9$	98,0	970,3	9606,0
27	178	$(178-167{,}1)=10{,}9$	118,8	1295,0	14115,8
28	182	$(182-167{,}1)=14{,}9$	222,0	3307,9	49288,4
29	183	$(183-167{,}1)=15{,}9$	252,8	4019,7	63912,9
30	191	$(191-167{,}1)=23{,}9$	571,2	13651,9	326280,9

Далее находим значения $\Sigma(x_i-\bar{X})^2$; $\Sigma(x_i-\bar{X})^3$ и $\Sigma(x_i-\bar{X})^4$, которые будут равны соответственно суммам всех значений столбиков 4, 5 и 6 таблицы 3.

$\sum(x_i - \bar{X})^2 = 404 + 259{,}2 + 228 + 123{,}2 + 102 + \cdots + 222 + 252{,}8 +$
$+571{,}2 = 2800{,}7.$
$\sum(x_i - \bar{X})^3 = -8120{,}6 + (-4173{,}3) + (-3443) + \cdots + 4019{,}7 +$
$+13651{,}9 = 4748{,}8.$
$\sum(x_i - \bar{X})^4 = 163224{,}1 + 67189{,}8 + 51988{,}6 + \cdots + 63912{,}9 +$
$+326280{,}9 = 793849{,}5.$

Далее определяем значение стандартного отклонения (σ):

$6 = \sqrt{\frac{2800{,}7}{30-1}}$=9,8.

Рассчитываем асимметрию (A) подставив все известные значения в формулу (4.1): $A = \frac{4748{,}8}{30*9{,}8^3}$=0,17.

Рассчитаем эксцесс (E) подставив все известные значения в формулу (4.2): $E = \frac{793849{,}5}{30*9{,}8^4} - 3 = -0{,}16$.

Далее вычислим ошибки репрезентативности асимметрии (m_A) и эксцесса (m_E) по формулам (4.3) и (4.4) соответственно:

$m_A = \sqrt{\frac{6}{30}} = 0{,}45$.

$m_E = \sqrt{\frac{6}{30}} * 2 = 0{,}9$.

Так как $|A| < 3 * m_A$ и $|E| < 3 * m_E$, то распределение близко нормальному.

В целом, проверка на соответствие закону нормального распределения требует достаточно большого количества вычислений, особенно если выборка состоит из десятков объектов исследования. Существующие компьютерные программы, о которых упоминалось ранее, позволяют проверить распределение данных за несколько минут. Поэтому проверка данных на соответствие закону нормального распределения должна осуществляться с большей надежностью не по одному критерию (графическому или расчетному), а по их совокупности.

Задание: произвести расчет асимметрии и эксцесса для данных из индивидуального задания. По результатам расчетов сделать вывод о соответствии распределения данных нормальному.