МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ

БЕЛОРУССКИЙ НАЦИОНАЛЬНЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
КАФЕДРА ИНЖЕНЕРНОЙ ГЕОДЕЗИИ

Методические указания к лабораторным работам

по дисциплине «ТМОГИ»

для студентов 2 курса специальности «Геодезия»

Составил: **Ст.пр.каф.инж.геод. Будо А.Ю.**

Минск
2016

Содержание

Порядок выполнения лабораторных работ

Работы выполняются последовательно, начиная с первой. Исходные данные выбираются по вариантам, выдаваемым преподавателем. Также преподавателем устанавливаются крайние сроки сдачи работ на проверку (deadline). Правила оформления работ приведены в **Приложение А**.

Оценка $Est(Estimate)$ за каждую лабораторную работу вычисляется по формуле

$$Est = 10 \cdot T \cdot D/D_{max},$$

где $T(Test)$ – балл, полученный в ходе автоматизированной проверки;

$D(Defense)$ – балл, полученный по результатам защиты работы;

D_{max} – максимально возможный балл при защите работы.

Величина T рассчитывается на основе алгоритма Эвина Вильсона (1927г), в котором используется формула для нижней границы доверительного интервала

$$T = \frac{p + \frac{z_{\alpha/2}^2}{2 \cdot n} - z_{\alpha/2} \cdot \sqrt{\frac{p(1-p) + \frac{z_{\alpha/2}^2}{4 \cdot n}}{n}}}{1 + \frac{z_{\alpha/2}^2}{n}},$$

где $n = pos + WN$ – сумма правильных ответов и номера недели от начала года (зависит от даты проверки работы);

$p = \mathrm{pos/n}$ – доля правильных ответов;

$z_{\alpha/2}$ – квантиль[1] стандартного нормального распределения для вероятности (1–α/2). Для доверительного уровня в 0.95 значение квантиля равно 1.96.

Величина D вычисляется на основе следующей таблицы

Таблица – Критерии оценки при защите работы

Критерий	Оформленная работа сдана до дедлайна? (1/0)	Оформление уникальное, соответствует правилам оформления? (2/0)	На защите получены ответы на все вопросы - 2; Частично неправильные или неполные ответы - 1; Неправильные ответы или работа не защищалась - 0	Сумма
Max	1	2	2	5
Min	0	0	0	0

[1] *Квантиль* в математической статистике — значение, которое заданная случайная величина не превышает с фиксированной вероятностью.

Лабораторная работа 1. Классическая обработка многократных измерений одной величины

Цель: обработать результаты равноточных и неравноточных измерений, выполнить задачу эталонирования.

Исходными данными для работы является превышение **h** между двумя точками, измеренное **N** раз при разном количестве штативов **n** в каждом измерении. Значение вероятности принимается равным **P** = 1 – q, где уровень значимости **q** = №варианта / 100. При вычислениях удерживается на один десятичный знак больше, чем в исходных данных.

1.1. Равноточные измерения одной величины

Предполагая, что условия теоремы Гаусса-Маркова и центральной предельной теоремы Ляпунова соблюдены, т.е. измерения независимы и их математическое ожидание и дисперсия постоянны, а закон распределения погрешностей измерений является нормальным, тогда последовательность обработки будет следующей:

Вычисляется среднее арифметическое в качестве оценки математического ожидания измеренных превышений

$$\bar{h} = \frac{\sum_{i=1}^{N} h_i}{N}. \tag{1.1}$$

Вычисляется оценка стандарта в виде средней квадратической погрешности (СКП) по формуле Бесселя

$$m = \sqrt{\frac{\sum_{i=1}^{N} \left(h_i - \bar{h}\right)^2}{N-1}}. \tag{1.2}$$

Вычисляется оценка среднего арифметического (СКП ср. арифм.)

$$M = \frac{m}{\sqrt{N}}. \tag{1.3}$$

Рассмотренный выше способ оценки называется точечным. Более совершенным является так называемый способ доверительных интервалов. При нём определяется значение квантиля **t** распределения Стьюдента для доверительной вероятности **P** при количестве степеней свободы **r** = N – 1. Для этого можно воспользоваться статистическими таблицами, например, (Приложение Б). Например, если дана выборка из 29 элементов (28 степеней свободы), то для вероятности P=0.95 и r=28 выбирается значение t=1.7011. Также для нахождения квантиля t можно воспользоваться специализированными программами, например, в Microsoft Excel 2007 формула для вычисления квантиля при q=1–P=0.05 и r=28 будет иметь вид

=СТЬЮДРАСПОБР(0.05*2;28) (1.4)

Поскольку для нахождения истинного значения измеряемой величины используются интервальные оценки, то для вычисления нижней границы интервала истинного значения превышения необходимо воспользоваться формулой

$$h_{min} = \bar{h} - t \cdot M. \tag{1.5}$$

а для верхней границы соответственно

$$h_{max} = \bar{h} + t \cdot M, \tag{1.6}$$

и это свидетельствует о том, что с 90% уверенностью (двусторонняя критическая область) мы находим истинное значение, лежащее на интервале (h_{min}, h_{max}).

Для оценки теоретического значения стандарта по статистическим таблицам, например, (Приложение В) определяется величина χ^2. Для нижнего интервала χ_1^2 при P_1= q/2; и для верхнего χ_2^2 при P_2= 1–q/2. Для примера, когда r=28 и q=0.05 формула в Excel 2007 для вычисления нижнего χ_1^2 имеет вид

=ХИ2ОБР(1–0.05/2;28), (1.7)

и верхнего χ_2^2

=ХИ2ОБР(0.05/2;28), (1.8)

Полученные значения соответственно χ_1^2=15.3079 и χ_2^2=44.4608.

После этого вычисляют коэффициенты

$$v_1 = \sqrt{\frac{r}{\chi_2^2}}. \tag{1.9}$$

и

$$v_2 = \sqrt{\frac{r}{\chi_1^2}}. \tag{1.10}$$

получая значения v_1= 0.7936 и v_2=1.3525. После этого для истинного значения стандарта можно вычислить нижнюю границу

$$m_{min} = v_1 \cdot m \tag{1.11}$$

и

$$m_{max} = v_2 \cdot m. \tag{1.12}$$

Аналогичным образом рассчитывается доверительный интервал для СКП среднего арифметического. Нижняя граница

$$M_{min} = v_1 \cdot M \tag{1.13}$$

и

$$M_{max} = v_2 \cdot M. \tag{1.14}$$

Для более глубокого понимания процесса обработки ряда равноточных измерений одной и той же величины можно воспользоваться дополнительной литературой [1, с.97].

1.2. Неравноточные измерения одной величины

Наилучшей оценкой математического ожидания для неравноточных измерений является среднее взвешенное или общая арифметическая середина, которую можно вычислить по формуле

$$\bar{h}_2 = \frac{[p \cdot h]}{[p]} = \frac{e^T \cdot P \cdot h}{e^T \cdot P \cdot e}. \tag{1.15}$$

где e – вектор-столбец, состоящий из **N** единиц

P – диагональная матрица **N*N** со значениями весов измерений на диагонали. Вес измерения при нивелировании может быть задан как $p_i = \frac{1}{n_i}$, где n_i – количество штативов в i-той секции.

СКП единицы веса может быть определена по формуле Бесселя

$$\mu = \sqrt{\frac{\sum_{i=1}^{N} p_i (h_i - \bar{h}_2)^2}{N-1}} = \sqrt{\frac{V^T P V}{N-1}}. \tag{1.16}$$

Величина $\bar{h}_2$ определяется с точностью (СКП средневзвешенного)

$$m_{h_2} = \sqrt{\frac{\sum_{i=1}^{N} p_i (h_i - \bar{h}_2)^2}{(N-1) \cdot \sum_{i=1}^{N} p_i}} = \sqrt{\frac{V^T \cdot P \cdot V}{(N-1) \cdot (e^T \cdot P \cdot e)}}. \tag{1.17}$$

где V – вектор, состоящий из поправок, вычисляемых по формуле

$$V_i = h_i - \bar{h}_2. \tag{1.18}$$

Истинные значения h_2 могут быть найдены для нижней границы как

$$h_{2\,min} = \bar{h}_2 - t \cdot m_{h_2}, \tag{1.19}$$

а для верхней границы соответственно

$$h_{2\,max} = \bar{h}_2 + t \cdot m_{h_2}. \tag{1.20}$$

Нижняя граница истинного значения $\mu_{\text{ист}}$ (СКП единицы веса) может быть найдена для нижней границы как

$$m_{\mu,min} = v_1 \cdot \mu, \tag{1.21}$$

а для верхней границы

$$m_{\mu,max} = v_2 \cdot \mu, \tag{1.22}$$

где значения v_1 и v_2 вычисляются по формулам (1.9), (1.10).

Нижняя граница истинного значения $m_{h_{2,\text{ист}}}$ (СКП средневзвешенного) может быть найдена как

$$m_{h_2 min} = v_1 \cdot m_{h_2}, \tag{1.23}$$

и верхняя граница

$$m_{h_2 max} = v_2 \cdot m_{h_2}, \tag{1.24}$$

где значения v_1 и v_2 вычисляются по формулам (1.9), (1.10).

Для более глубокого понимания процесса обработки ряда неравноточных измерений одной величины можно воспользоваться дополнительной литературой [1, с.127].

1.3. Задача эталонирования

В случаях, когда необходимо определить точность прибора и есть эталон измеряемой величины (компаратор), производят N измерений эталона и вычисляют истинные погрешности, считая значение эталона равным истинному значению.

$$\Delta_i = h_i - h_{\text{эт}}, \tag{1.25}$$

где $h_{\text{эт}}$ – принятое за истинное значение измеряемой величины. В работе в качестве эталонного принять значение равное медиане[2] исходного ряда превышений.

Для оценки точности прибора используется формула Гаусса

$$m = \sqrt{\frac{\sum_{i=1}^{N} \Delta_i^2}{N}}. \tag{1.26}$$

Полученная СКП характеризует точность прибора, если в измерениях отсутствовали значимые грубые и систематические погрешности.

Значимость грубых ошибок может быть оценена по правилу трёх сигм (правило Райта): все измерения выходящие за интервал

$$(h_{\text{эт}} - 3 \cdot m) \le h_i \le (h_{\text{эт}} + 3 \cdot m) \tag{1.27}$$

считаются грубыми и не включаются в дальнейшую обработку.

После удаления измерений с грубыми ошибками из ряда, проверяют наличие значимого систематического влияния.

Систематическое влияние считается значимым при невыполнении неравенства

$$-t \cdot m_{\bar{\Delta}} \le \bar{\Delta} \le t \cdot m_{\bar{\Delta}} \tag{1.28}$$

где t – квантиля распределения Стьюдента для доверительной вероятности P (для каждого варианта своя) при количестве степеней свободы r = N – 1.

$\bar{\Delta}$ – среднее арифметическое из истинных погрешностей;

[2] *Медианой ряда чисел* называется число, стоящее посередине упорядоченного по возрастанию ряда чисел (в случае, если количество чисел нечётное). Если же количество чисел в ряду чётно, то медианой ряда является полусумма двух стоящих посередине чисел упорядоченного по возрастанию ряда.

$m_{\bar{\Delta}}$ – оценка среднего арифметического $\bar{\Delta}$, вычисляемая по формуле

$$m_{\bar{\Delta}} = \frac{m}{\sqrt{N}}, \tag{1.29}$$

где m – СКП, вычисленная по формуле (1.26) для ряда без грубых ошибок.

При невыполнении неравенства (1.28) вычисляют новый ряд, свободный от систематического влияния $\Delta'_i = \Delta_i - \bar{\Delta}$, а оценку точности выполняют по формуле Бесселя

$$m = \sqrt{\frac{\sum_{i=1}^{N} {\Delta'_i}^2}{N-1}}. \tag{1.30}$$

В работе выполнить расчёт по формуле (1.30) даже при выполнении условия (1.28)

Лабораторная работа 2. Альтернативная обработка многократных измерений одной величины

Цель: выполнить обработку непараметрическими методами (закон распределения неизвестен) многократно измеренного превышения, установить наличие систематического влияния и грубых ошибок в измеренном ряду данных.

При количестве измерений меньшем пятидесяти сложно выявить закон распределения погрешностей, поэтому в таких случаях находят несколько альтернативных оценок, которые сравниваются между собой. В данной работе используются те же исходные данные, что и в Лабораторная работа 1.

2.1. Выявление мешающих параметров непараметрическими методами

Наличие *систематического влияния* в измерениях можно определить, построив линию тренда, т.е. аппроксимировав исходные данные функций вида

$$h = a \cdot i + b \tag{2.1}$$

где i – номер измерения по порядку от 1 до N (переставлять результаты измерений нельзя);

a – показатель систематического влияния;

b – оценка наиболее надёжного значения.

Найти неизвестные коэффициенты линии a, b можно решив систему уравнений, которая в матричном виде может быть представлена как

$$\binom{a}{b} = N^{-1} \cdot B \tag{2.2}$$

где

$$N = \begin{pmatrix} \sum(i^2) & \sum i \\ \sum i & N \end{pmatrix} \tag{2.3}$$

$$B = \begin{pmatrix} \sum(h_i \cdot i) \\ \sum(h_i) \end{pmatrix} \tag{2.4}$$

После вычисления коэффициентов прямой линии необходимо её построить вместе с графиком функции $h = f(i)$.

Далее необходимо рассчитать погрешность модели по формуле

$$\mu = \sqrt{\frac{\sum(h^{\text{выч}}{}_i - h_i)^2}{N - k}} \tag{2.5}$$

где $h^{\text{выч}}{}_i = a \cdot i + b$;

N – количество измерений;

k – количество неизвестных параметров, в нашем случае k = 2.

Систематическое влияние на измерения можно определить по правилу «трёх сигм» или правилу Райта, согласно которому систематическое влияние считается значимым, если не выполняется условие

$$-3 \cdot m_a \leq a \leq 3 \cdot m_a \tag{2.6}$$

где погрешность m_a может быть определена по формуле

$$m_a = \mu \cdot \sqrt{(N^{-1})_{1,1}} \tag{2.7}$$

в которой $(N^{-1})_{1,1}$ – первый диагональный элемент обратной нормальной матрицы.

При более точном подходе необходимо вычислить t-статистику Стьюдента

$$t = \frac{a}{m_a} \tag{2.8}$$

Если модуль вычисленного значения окажется меньше табличного, определённого для вероятности P и числа степеней свободы r = N – k, то делается вывод об отсутствии значимого систематического влияния с этой вероятностью. В работе принять P = 95%.

Для нахождения измерений с *грубыми ошибками* может быть использован **критерий Хэмпэла**, согласно которому грубым считается измерение, лежащее вне интервала

$$AMO\_low \leq h \leq AMO\_high \tag{2.9}$$

где $med(h)$ – медиана, вычисляемая из вариационного ряда измерений h_i;

АМО – абсолютное медианное отклонение, вычисляемое по формуле

$$AMO = med(|h_i - med(h)|) \tag{2.10}$$

нижняя граница которого вычисляется по формуле

$$AMO\_low = med(h) - 5.2 \cdot AMO \tag{2.11}$$

верхняя граница

$$AMO\_high = med(h) + 5.2 \cdot AMO \tag{2.12}$$

2.2. Альтернативные оценки результатов измерений

Перед получением альтернативных оценок должны быть найдены среднее арифметическое

$$\bar{h} = \frac{\sum_{i=1}^{N} h_i}{N}, \tag{2.13}$$

средняя квадратическая погрешность

$$m = \sqrt{\frac{\sum_{i=1}^{N}\left(h_i - \bar{h}\right)^2}{N-1}} \tag{2.14}$$

$med(h)$ – медиана;
средняя абсолютная погрешность

$$v = \frac{\sum_{i=1}^{N}\left|h_i - \bar{h}\right|}{N-1} \tag{2.15}$$

Данные величины являются оценками математического ожидания и стандарта для двух крайних законов распределения (закон Гаусса и закон Лапласа). При этом первая пара оценок весьма чувствительна к отклонению результатов измерений от нормальности и к влиянию мешающих параметров. Вторая пара оценок нечувствительна к этим отклонениям (робастна). Поэтому, степень отличия среднего арифметического от медианы может сказать о значимости посторонних влияний. Если отличия не значимы, то используется первая пара оценок, если значима, то вторая.

Другой подход в определенной выше ситуации заключается в вычислении непараметрических оценок, которые по определению свободны от закона распределения. Наиболее распространенные оценки такого рода – это **L-оценки** и **R-оценки**.

В работе предлагается вычислить следующие наиболее часто встречающиеся **L-оценки** (оценки в линейных комбинациях):

1. **Усеченное среднее** (α-усеченное среднее). Для её нахождения в вариационном ряду необходимо отбросить с левой и правой стороны α% значений, а из оставшихся взять обычное среднее арифметическое;

2. **Винзоризованное среднее** (α-винзоризованное среднее). Для его нахождения необходимо в вариационном ряду α% крайних значений присвоить значения: слева – α+1 значение, а справа – (n – α – 1) значение. Другими словами, необходимо $k=(N\cdot\alpha)$ последним значениям вариационного ряда присвоить значение предыдущего для них элемента, а первым $k=(N\cdot\alpha)$ значениям присвоить значение следующего после них элемента. Из преобразованного ряда берется обычное среднее арифметическое.

Пусть α = 20% и имеется набор данных (отсортированных по возрастанию):

2, 3, 4, 5, 7, 9, 10, 12, 14, 30

Расчет 20% винзоризованного среднего в нашем примере предполагает перед вычислением среднего арифметического замену первых двух и последних двух значений в ряду данных (2, 3 и 14, 30):

4, 4, 4, 5, 7, 9, 10, 12, 12, 12.

После замены и расчета среднего результат = 7.9.

Сравнение полученных оценок с обычным средним арифметическим также может сказать (по определенному выше правилу) какую величину взять в качестве конечной.

В работе для вычисления усечённого и винзоризованного среднего значение α принять равным 10%, количество крайних значений в вариационном ряду округлять в большую сторону.

Из **R-оценок** (оценки в ранговых критериях) предлагается вычислить следующие :

1. **Оценка Бикела-Ходжеса**. Находится как медиана из ряда, полученного из средних арифметических двух значений из вариационного ряда: первое – последнее, второе – предпоследнее и т.д.;

$$\theta_{\text{Б}-\text{Х}} = med\left(\frac{h_{i:n} + h_{n+1-i:n}}{2}\right) \tag{2.16}$$

2. **Оценка Лемана-Хождеса**. Её получают как медиану из всех возможных пар средних в ряду измерений. В работе можно использовать упрощенную оценку, когда в комбинациях для формирования средних значений номер первого слагаемого j всегда меньше номера второго слагаемого k.

$$\bar{h} = \theta_{\text{Л}-\text{Х}} = med\left(\frac{h_{(j)} + h_{(k)}}{2}\right) \tag{2.17}$$

Наряду с этими оценками большое распространение в условиях неопределенности и малом количестве измерений получила **адаптивная оценка Хогга**, когда по величине индикатора выбирается та, или иная формула вычисления оценки. Для её получения используется следующий подход:

$$\bar{h} = \begin{cases} S(0.25; N), k < 2; \\ C_t(0; N), 2 < k < 4; \\ C_t(0.25; N), 4 < k < 5.5; \\ C_t(0.5; N), 5.5 < k. \end{cases} \tag{2.18}$$

где $S(0.25; N)$ – среднее из первых 25% и последних 25% значений вариационного ряда;

$C_t(\alpha; N)$ – α-урезанное среднее. Если α=0, то получают стандартное среднее арифметическое;
при α=0.25 из вариационного ряда удаляется 25% наименьших и 25% наибольших значений, а из оставшихся берётся среднее арифметическое;
при α=0.5 удаляется по 50% слева и справа – стандартная медиана

Для оценки коэффициента k используется два подхода.

1. В качестве индикатора k берётся значение оценки не центрированного эксцесса

$$k = E = \frac{\sum_{i=1}^{N}\left(h_i - \bar{h}\right)^4}{N \cdot m^4} \tag{2.19}$$

2. Значение коэффициента, обозначенного t_N, вычисляют по формуле

$$k = t_n = \frac{a_N(0.05) - b_N(0.05)}{a_N(0.5) - b_N(0.5)}. \tag{2.20}$$

где $a_N(\beta)$, $b_N(\beta)$ – среднее по $(100 \cdot \beta)\%$ наибольших и наименьших элементов вариационного ряда соответственно. В случае дробных значений $a_N(\beta)$ и $b_N(\beta)$ округляются в большую сторону.

Вычисления выполнить при коэффициенте k, который рассчитан с использованием первого и второго подхода. Сделать выводы.

Лабораторная работа 3. Методы выявления дополнительных мешающих параметров

Цель: выполнить обработку многократно измеренного превышения, выявить наличие *гетероскедастичности* и *автокорреляции* в измеренном ряду данных.

Как известно, к **основным мешающим параметрам** относят значимые грубые и систематические погрешности, незнание влияния которых может испортить эффективность используемой оценки. Способы их выявления делят на 1) параметрические (закон распределения известен или может быть определён); 2) непараметрические (закон распределения определить невозможно).

К **дополнительным мешающим параметрам** можно отнести эффект *гетероскедастичности* (неравноточности результатов измерений) и эффект *автокорреляции* (зависимости элементов в одном ряду между собой).

В данной работе используются те же исходные данные, что и в Лабораторная работа 1.

3.1. Выявление эффектов гетероскедастичности

Наиболее простым тестом выявления степени неравноточности групп результатов измерений является критерий *ранговой корреляции Спирмена*. Критерий выявляет корреляцию между номером измерения i и поправкой

$$v_i = h_i - \bar{h}\,, \qquad (3.1)$$

которая при отсутствии неравенства дисперсий измерений должна быть статистически не значимой. Для этого находят ранги исследуемого ряда следующим образом:

1. Присваивают каждому измерению номер i от 1 до N;
2. Вычисляют среднее арифметическое и отклонение от него v_i для всех элементов ряда;
3. Выстраивают ряд отклонений в вариационный ряд (по возрастанию);
4. Получают ранг n_i отклонения v_i , который равен номеру i каждого элемента исходного ряда отклонений в вариационном ряду. Т.е., если элемент в исходном ряду измерений имел порядковый №6, а в построенном вариационном ряду данный элемент получил №15, то принимаются следующие значения: i = 6; n_i = 15.
5. И так далее для каждого элемента ряда.

Коэффициент ранговой корреляции Спирмена вычисляется по формуле:

$$r_{i,v} = 1 - 6 \cdot \frac{\sum_{i=1}^{N}((n_i - i)^2)}{N \cdot (N^2 - 1)}, \qquad (3.2)$$

Полученный коэффициент корреляции исследуется на значимость при помощи квантиля **t** распределения Стьюдента, который сравнивается с рассчитанным значением

$$t = \frac{|r_{i,v}| \cdot \sqrt{N-2}}{\sqrt{1 - r_{i,v}^2}}, \tag{3.3}$$

Полученное значение сравнивается с эталонным, выбираемым из статистических таблиц по модифицированной вероятности для двухстороннего интервала (1+P)/2 и числу степеней свободы N–2. При выполнении неравенства $t > t_{эт}$ исходная гипотеза об отсутствии гетероскедастичности отвергается с вероятностью ошибки (1–P)/2. При вычислениях принять значение P=0.95. Тогда, например, при числе измерений N = 20, число степеней свободы N–2=18 и значение квантиля для модифицированной вероятности (1+P)/2 будет равно $t_{эт}$=2.10.

Второй распространенный тест на наличие гетероскедастичности в результатах измерений называется *тестом Голдфелда-Квандта*. Его суть: если в вариационном ряду группа первых результатов и группа последних имеет достаточно похожую меру рассеивания, то эффект неравноточности результатов измерений незначителен. Для практической реализации теста поступают следующим образом:

1) делят вариационный ряд на три примерно равных части (меньшая в середине, равные по краям). Два крайних подряда аппроксимируют по методу наименьших квадратов в зависимости от номера *i* (см. Выявление мешающих параметров непараметрическими методами – задача выявления тренда).

2) вычисляют суммы квадратов отклонений для первой $[v^2]_1$ и второй $[v^2]_2$ регрессий, а также F–статистику Фишера по формуле

$$F = \frac{[v^2]_2}{[v^2]_1}. \tag{3.4}$$

При этом в числителе должна быть большая величина. Критическое значение $F_{кр}$ выбираются из таблиц распределения Фишера по уровню значимости $\alpha = 1 - P$ и числу степеней свободы $r_1 = r_2 = \mathrm{k} - t - 1$, где k – число элементов в крайнем ряду, t – число неизвестных параметров (объясняющих переменные) в приятой модели регрессии, т.е. t = 2. Если выполняется неравенство $F < F_{кр}$, то с вероятностью *P* гипотеза об отсутствии гетероскедастичности принимается.

При вычислениях принять значение P=0.95. Тогда, например, при числе степеней свободы $r_1 = r_2 = 5$ значение квантиля *F*-распределения вероятности будет равно $F_{0.95;5;5}$=5.05. Для нахождения квантиля можно воспользоваться статистическими таблицами [Приложение Г] либо функцией Excel.

=FРАСПОБР(0.05;5;5) (3.5)

3.2. Выявление систематического влияния непараметрическими способами

Наиболее часто используемые критерии выявления систематических влияний в результатах измерений при условии, что закон распределения неизвестен и количество измерений невелико, это *критерий серий* и *критерий «восходящих» и «нисходящих» серий*.

Критерии серий относят к критериям, выявляющим значимость систематического влияния только монотонного характера, (сдвиг или тренд) на основе проверки вероятностной независимости среди элементов исследуемого ряда. Для этого производят вычисления следующим образом

1. Строят вариационный ряд и находят медиану med(h)
2. Формируют знаковый ряд из плюсов и минусов по правилу: если значение исходного (невариационного) ряда больше медианы, то вместо i-го числа записывают знак «+», если меньше, то знак «–». Элементы ряда равные med(h) пропускают.
3. Находят количество серий (последовательностей подряд идущих знаков) $v(N)$. Один плюс или один минус тоже будет считаться серией.
4. Находят число элементов в наибольшей серии $\tau(N)$

Для стохастической независимости и, следовательно, отсутствия значимого систематического влияния монотонного характера должны одновременно выполняться два неравенства

$$v(N) > 0.5 \cdot \left(N + 1 - 1.96\sqrt{N-1}\right), \tag{3.6}$$

$$\tau(N) > 3.3 \cdot \log_{10}(N+1). \tag{3.7}$$

Если хотя бы одно из неравенств не выполняется, то в измерениях присутствует значимое систематическое влияние. При этом гипотеза об отсутствии систематического влияния в исходных измерениях отвергается с вероятностью ошибки, заключённой в пределах от 0.05 до 0.0975, т.е. с доверительной вероятностью 0.9025 – 0.95.

В отличие от *критерия серий* рассматриваемый далее *критерий «восходящих» и «нисходящих» серий* выявляет смещение среднего значения не только монотонного характера (тренд или сдвиг), но и более общего, например, периодического характера. В нём также исследуется последовательность знаков, но закон её построения следующий: на месте значения h_i исходного ряда ставится «+», если

$$h_{i+1} - h_i > 0 \tag{3.8}$$

и соответственно знак «–» при выполнении неравенства

$$h_{i+1} - h_i < 0. \tag{3.9}$$

Если несколько последовательных измерений равны, то используется только одно из них. Гипотеза об отсутствии систематического влияния принимается в случае выполнения неравенств

$$v(N) > \frac{1}{3} \cdot (2 \cdot N - 1) - 1.96\sqrt{\frac{16 \cdot N - 29}{90}}, \tag{3.10}$$

и

$$\tau(N) > \tau_0(N), \tag{3.11}$$

где $\tau_0(N) = 5,$ при $N \leq 26$; $\tau_0(N) = 6,$ при $26 < N \leq 153$; $\tau_0(N) = 7,$ при $153 < N \leq 1170$.

В случае невыполнения одного из неравенств (3.10) – (3.11), гипотеза об отсутствии систематического влияния отвергается с уровнем значимости (вероятностью ошибки первого рода) от 0.05 до 0.0975.

3.3. Выявление эффектов автокорреляции

Под автокорреляцией ряда принято понимать тесноту связи между элементами одного ряда. Чтобы упорядочить эти связи, используют понятие лага – величины сдвига между исследуемыми элементами. Наиболее часто встречается автокорреляция лага (сдвига) 1, т.е. между рядом стоящими элементами в ряду: 1 и 2, 2 и 3, 3 и 4 и т.д. Самым известным и используемым тестом на исследование такого рода зависимости является критерий Дарбина-Уотсона, когда по статистике DW, вычисленной по величинам остатков V после аппроксимации ряда линейной функцией, делается вывод о виде и значимости автокорреляции. Эта статистика тесно связана с выборочным коэффициентом корреляции между рядом стоящими остатками V_{i-1} и V_i.

$$r = 1 - \frac{DW}{2}, \tag{3.12}$$

$$DW = \frac{\sum_{i=2}^{N}((V_i - V_{i-1})^2)}{\sum_{i=1}^{N}(V_i{}^2)}, \tag{3.13}$$

Тогда из (3.12) – (3.13) имеем:

– если $DW \approx 2$, то $r \approx 0$ (отсутствие автокорреляции);

– если $DW \approx 0$, то $r \approx 1$ (положительная автокорреляции);

– если $DW \approx 4$, то $r \approx -1$ (отрицательная автокорреляции).

Общая схема критерия Дарбина-Уотсона следующая:

1. Строят эмпирическое уравнение регрессии, например, h_i от i и находят остатки

$$V_i = h_i - \hat{h}_i, \tag{3.14}$$

например, как в тесте на основе ранговой корреляции Спирмена (см. Выявление эффектов гетероскедастичности)

2. Рассчитывают по формуле (3.13) статистику DW и при приближенном оценивании, по изложенному выше правилу смотрят к какому числу из 0, 2 или 4 находится ближе вычисленное значение статистики.

3. Исходя из расчетов делают приближенный вывод о возможности того или иного исхода. Можно считать, что если $1.5 < DW < 2.5$, то автокорреляция отсутствует, при $-0.5 < DW < +0.5$ имеем положительную автокорреляцию, т.е. остатки все время возрастают, а для $3.5 < DW < 4.5$ – отрицательную автокорреляцию (остатки все время убывают). Результаты тем надежней, чем ближе статистика к ключевым точкам.

Расчётно-графическая работа. Исследование ряда погрешностей на соответствие нормальному закону распределения

Цель: освоить различные методики исследования погрешностей результатов измерений на соответствие нормальному закону распределения с предварительным анализом на систематические и грубые ошибки; получить основные вероятностно-статистические характеристики многократно измеренной величины.

Порядок выполнения работы

1. Для ряда данных, соответствующего варианту, выданному преподавателем, выполнить оценку центра распределения и исключить из ряда промахи.

Предварительные вычисления:

2. Исследовать полученный ряд на наличие систематической составляющей и при ее наличии исключить из элементов, получив новые, преобразованные величины.
3. Исследовать выборку на наличие в ней грубых погрешностей и, если они обнаружатся, исключить их из ряда.
4. Исследовать по соответствующим формулам F-критерия Фишера ряд на степень неоднородности.

В пунктах 2-4 принять доверительную вероятность $\beta = 0.95$.

Приближенные исследования:

5. Вычислить оценки основных параметров нормального распределения.
6. Выполнить приближенное исследование ряда погрешностей на соответствие нормальному закону, вычислив среднюю абсолютную и вероятную ошибки и их отношение к средней квадратической ошибке ряда и дополнительные характеристики статистической совокупности в виде асимметрии и эксцесса, установить их значимость.
7. Сделать выводы о соответствии закону Гаусса по приближенным критериям.

Графические исследования:

8. Рассчитать размер интервалов необходимых для построения гистограммы. Интервалы должны накрывать все исследуемые значения.
9. Сосчитать количество элементов, попадающих в каждый интервал. Откладывая на горизонтальной оси величины интервалов, а на вертикальной – высоты прямоугольников построить гистограмму.
10. Построить на гистограмме кривую распределения нормального закона по вычисленным в п.5 основным характеристикам.
11. Визуально оценить степень приближения эмпирического распределения, представленного в виде гистограммы, к теоретическому распределению, представленному в виде непрерывной кривой.

Вероятностные исследования:

12. Оценить, верна ли с вероятностью β принятая гипотеза о законе распределения погрешностей по χ^2 - критерию Пирсона.
13. Произвести вычисление вероятности точного выполнения гипотезы.

14. Сделать общий вывод о соответствии исследуемого ряда погрешностей нормальному закону распределения. Заключительный анализ результатов исследований выполняется в виде пояснительной записки, включающей текстовый материал, со ссылками на формулы и источники литературы.

Теоретические основы выполнения исследований.

Обработка результатов измерений имеет место всегда, когда одна из определяемых величин получена несколько раз с отличными друг от друга значениями. При этом корректная оценка результатов возможна только при известных правилах, определяющих поведение погрешностей измерений Δ. К таким правилам относят законы поведения погрешностей в дифференциальной $f(\Delta)$ и интегральной $F(\Delta)$ формах, их основные численные характеристики и представления законов в графическом виде. Интегральная форма называется *функцией распределения погрешностей* Δ_i , дифференциальная форма – *функцией плотности распределения погрешностей* Δ_i . К основным характеристикам законов относят наиболее вероятное значение определяемой величины, называемое *математическим ожиданием* и обозначаемое $MO(\Delta)$ или $M(\Delta)$, μ; меру рассеивания измерений вокруг математического ожидания, называемую *дисперсией* и обозначаемую $\sigma^2(\Delta)$ или $D(\Delta)$ (чаще используют просто величину $\sigma(\Delta)$, называемую *стандартом,* так как он не имеет квадратичной размерности как у дисперсии). К дополнительным характеристикам законов относят меру скошенности относительно вертикальной оси симметрии, называемую *асимметрией* и обозначаемую A или S_k и меру крутости, называемую *эксцессом* и обозначаемую E.

Множество теоретических и практических исследований показывают, что результаты геодезических измерений подчиняются *нормальному закону распределения (закону Гаусса)* и имеют вид

$$F(X) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot \int_{-\infty}^{x} e^{\frac{-(t-\mu)^2}{2\cdot\sigma^2}} \cdot dt\,, \qquad (1)$$

с основными характеристиками, представленными выше. Следует твердо усвоить, что математическое ожидание и стандарт являются теоретическими, заведомо неизвестными значениями. Получить эти значения можно только при числе измерений $n = \infty$. Поэтому, в качестве приближений к этим параметрам используются их оценки. Например, оценка математического ожидания μ в виде *среднего весового* (общей *арифметической середины*) $\bar{X}$ или $X_{ср}$. Или оценка стандарта σ в виде *средней квадратической погрешности* m. Эти оценки вычисляются достаточно просто и имеют минимальное отклонение от своих теоретических значений, если поведение погрешностей подчинено нормальному закону Гаусса. Необходимо помнить, что теоретическое значение математического ожидания из ряда случайных погрешностей для закона Гаусса $MO(\Delta) = 0$. В упрощенном виде под законами распределения случайной величины будем понимать совокупность частот попадания ее реализации в заданные интервалы.

Таким образом, для правильного выбора алгоритма обработки результатов измерений необходимо установить, что с достаточной вероятностью β ошибки подчинены тому или иному закону распределения. Для геодезических измерений часто требуется исследование на соответствие погрешностей нормальному распределению Гаусса. Тогда в качестве наиболее надежных оценок результатов измерений можно использовать среднее взвешенное (или среднее арифметическое) и среднюю квадратическую погрешность.

Из теории погрешностей известно, что источниками элементарных погрешностей измерений является множество факторов разного характера, учесть которые практически невозможно. Поэтому суммарные ошибки, вызванные этими факторами, носят случайный характер и подчиняются законам теории вероятностей. Практическую реализацию формул теории вероятностей при конечном числе наблюдений рассматривает математическая статистика.

Одна из основных теорем теории вероятностей, называемая *центральной предельной теоремой Ляпунова,* гласит, что если элементарные составляющие общей ошибки измерения примерно одного порядка, невелики по величине, не зависимы между собой и не имеют систематической составляющей, то общая ошибка распределена по закону, очень близкому нормальному. Вместе с тем, в процессе измерений часто имеет место наличие *грубых погрешностей* (погрешностей, которые больше заданного допуска) или промахов (т.е. очень сильно отличающихся от других). Кроме этого, в результатах наблюдений могут содержаться и какие-либо постоянные составляющие, называемые *систематическими ошибками.* Не редки случаи включения в обработку измерений с разной точностью (неоднородных измерений).

Методики производства геодезических измерений рассчитываются так, чтобы по возможности полнее исключить систематические влияния на результаты, которые к тому же должны быть свободны от грубых погрешностей и однородны. Наличие как грубых, систематических погрешностей так и неоднородности нарушает условия теоремы Ляпунова и делает проблематичным использование для обработки измерений алгоритмов, полученных на основе закона распределения Гаусса.

Вычислительные схемы, основанные на нормальном законе (например, разные формы *метода наименьших квадратов* (*МНК*)), имеют самый простой вид и дают вполне надежные результаты при условии выполнения приведенных выше условий. Из всего сказанного совершенно очевидно, что для использования оценок, соответствующих закону Гаусса, необходимо выполнить исследование на степень соответствия ряда погрешностей нормальному закону. Но для достоверности результатов исследований необходимо провести предварительные вычисления и выявить наличие грубых погрешностей, значимость систематической составляющей и неоднородности. После этого закон распределения устанавливается на основе имеющейся статистической совокупности погрешностей. При этом результаты будут тем более надежны, чем больше используется в исследованиях величин погрешностей.

Основными задачами, решаемыми вероятностно-статистическими методами в контексте исследования закона распределения случайных величин для общего случая, являются:

1. Определение вида закона распределения ряда погрешностей – задача *сглаживания или выравнивания статистического ряда.*

2. Определение степени согласования опытных данных с гипотезой о том, что случайная величина подчинена полученному в п. 1 закону распределения с вероятностью β. Для этого используются различные «критерии согласия» – задача *проверки правдоподобия гипотез*.

На основе перечисленных пунктов производится определение «наилучших» оценок неизвестных параметров распределения в виде оценок математического ожидания $MO(\Delta)$ как центра распределения и дисперсии $D(\Delta)$ как меры рассеивания – задача *наилучшего оценивания.*

Эти же задачи характерны в целом и для всей математической статистики. Обычно используют три группы алгоритмов исследования на закон распределения случайных величин:

1. Приближенные – вычисляется ряд характеристик распределения и сравнивается с теоретическими значениями;

2. Графические – по результатам измерений строится какое-либо графическое представление закона в виде многоугольника распределения, полигона распределения, гистограммы и др.;

3. Вероятностные – выдвигается гипотеза о соответствии ряда погрешностей закону распределения с вероятностью β и проверяется ее правильность или вычисляется величина-критерий, по которой выносят суждение о вероятности соответствия результатов измерений выдвинутой гипотезе.

Очевидно, что перед началом работы по определению соответствия ряда нормальному закону, следует убедиться, что он достаточно случаен, т.е. достаточно неплохо выполняются условия теоремы Ляпунова. Следует иметь в виду, что невыполнение этих условий полностью обесценит результаты исследований. Для этого производятся предварительные вычисления.

Оценки центра распределения

Координата центра распределения определяет положение случайной величины на числовой оси. Дать однозначное определение этого понятия невозможно. Центр распределения может быть найден несколькими способами:

- как медиана распределения,
- как мода распределения,
- как математическое ожидание.

По возможности наиболее точная оценка центра распределения по выборке случайных величин исключительно важна, так как центр распределения используется в формулах для вычисления дисперсии, среднеквадратичного отклонения, коэффициента асимметрии и эксцесса распределения. Некорректное определение центра влечет за собой ошибки в определении всех этих величин.

Оценку центра распределения по выборке можно проводить различными способами. Не зная априорно закона распределения случайной величины, невозможно заранее указать наиболее приемлемый способ. К тому же, некоторые из этих оценок чувствительны к наличию аномальных значений в выборке (промахов).

Для нахождения центра распределения можно использовать альтернативные L, R-оценки, адаптивную оценку Хогга и др., рассмотренные в Лабораторная работа 2.

Но в данной работе для корректной оценки центра распределения $X_{центр}$ мы будем вычислять его пятью различными способами:

$X_{медиана}$, $X_{центр\_сгибов}$, $\bar{X}$, $\bar{X}_{50\%}$, $X_{центр\_размаха}$.

После этого пять полученных оценок упорядочим по возрастанию и выберем из них в качестве центра распределения серединное, то есть третье по счету, значение.

1) $X_{медиана}$ – обычная медиана.

2) $\bar{X}$ – среднее арифметическое.

3) $X_{центр\_размаха}$ – среднее между максимальным и минимальным значением в выборке.

4) $X_{центр\_сгибов}$ – Центр 50%-ного интерквантильного промежутка (нечувствительна к промахам в выборке). Перед вычислением упорядочивают выборку по возрастанию. Вычисляют четвертую часть от объема выборки, то есть.

$$M=ЦЕЛОЕ(N/4) \tag{2}$$

Тогда центр сгибов определяется по формуле:

$$\text{Х}_{\text{центр\_сгибов}} = \frac{X_{M+1} + X_{N-M}}{2}, \tag{3}$$

5) $\bar{X}_{50\%}$ –среднее арифметическое по 50%-му интерквантильному промежутку вычисляется по формуле

$$\text{Х}_{\text{центр\_сгибов}} = \frac{1}{N-2M} \cdot \sum_{k=M+1}^{N-M} x_k , \tag{4}$$

где M вычисляется по формуле (2).

Исключение промахов из выборки

Промахами в выборке случайных величин будем называть аномально отклоняющиеся от центра распределения значения по сравнению с основной массой данных. Исключать промахи из выборки необходимо, т.к. они могут существенно исказить оценку параметров распределения. Для исключения промахов введём понятие коэффициента цензурирования – безразмерной величины G, при которой промахами считаются все значения из выборки, лежащие за пределами интервала

$$\text{Х}_{\text{центр}} - G \cdot \bar{\sigma} \le x \le \text{Х}_{\text{центр}} + G \cdot \bar{\sigma} , \tag{5}$$

где

$$\bar{\sigma} = \sqrt{\frac{1}{N-1} \cdot \sum_{i=1}^{N} \left(x_i - \text{Х}_{\text{центр}}\right)^2}\,, \tag{6}$$

Интуитивно понятно, что коэффициент цензурирования должен зависеть от объема выборки и рассчитанного по выборке значения эксцесса. Действительно, такое отклонение от центра, которое является промахом для средневершинного (а тем более плосковершинного) распределения, для островершинного распределения с его длинными "тяжелыми" спадами может безусловно принадлежать выборке.

Эмпирическая формула для коэффициента цензурирования как функции от объема выборки N и эксцесса E, пригодная к применению для широкого класса распределений следующая:

$$G = 1.55 + 0.8 \cdot lg(N/10) \cdot \sqrt{E-1}\,, \tag{7}$$

где

$$E = \frac{1}{\bar{\sigma}^4} \frac{N^2 - 2N + 3}{(N-1)(N-2)(N-3)} \sum_{i=1}^{N} \left(x_i - \text{Х}_{\text{центр}}\right)^4 - \frac{3(2N-3)(N-1)}{N(N-2)(N-3)}\,, \tag{8}$$

После удаления промахов нужно пересчитать параметры распределения. При этом в качестве центра распределения уже можно использовать среднее арифметическое, как состоятельную и несмещенную оценку математического ожидания.

Предварительные вычисления для исследования

В предварительных вычислениях ряд исследуется на наличие значимых систематических и грубых погрешностей, а также меры однородности результатов по точности на основе каких-либо критериев.

Определение значимости систематического влияния. Следует иметь в виду, что систематические влияния в рядах присутствуют всегда, но они могут быть значимы и не значимы. При определении наличия значимых систематических погрешностей в ряду имеют место два случая:

1) известно истинное значение определяемой величины $X_{\text{ист}}$ и произведено N ее измерений x_i. В этом случае пользуются зависимостью

$$\frac{|\Delta_i|}{M_{\bar{X}}} \le z_q\,, \tag{9}$$

где

$$\Delta_i = x_i - X_{\text{ист}}\,, \tag{10}$$

средняя квадратическая погрешность среднего арифметического

$$M_{\bar{X}} = \frac{m}{\sqrt{N}}, \tag{11}$$

m – средняя квадратическая погрешность (*СКП*) одной величины,
N – число элементов в ряду.
z_q – квантиль t-распределения Стьюдента определяется по *уровню значимости q* (или вероятности β) и числу *избыточных измерений* (*числу степеней свободы*) $k = N - 1$ и выбирается из статистических таблиц [Приложение Б] или вычисляется, например, в Excel (1.4). Если неравенство (9) выполняется, то с вероятностью $\beta = 1 - q$ считаем, что значимые систематические погрешности в ряду измерений отсутствуют;

2) истинное значение величины неизвестно. Тогда наличие в результатах наблюдений постоянной составляющей может быть выяснено по наиболее распространенному в геодезии *критерию Аббе*. Для этого выдвигается гипотезу, что с вероятностью β в предложенном ряду отсутствует значимое систематическое влияние. По исследуемым величинам вычисляется практическая величина

$$\delta = \frac{\tilde{m}^2}{m^2}, \tag{12}$$

которая является отношением двух оценок дисперсий, средние квадратические ошибки которых получены как

$$m = \sqrt{\frac{\sum_{i=1}^{N} v_i^2}{N-1}} \tag{13}$$

и

$$\tilde{m} = \sqrt{\frac{\sum_{i=1}^{N} d_i^2}{2 \cdot (N-1)}}, \tag{14}$$

где уклонение *i* - той величины от среднего

$$v_i = x_i - \bar{X} \tag{15}$$

последовательные разности

$$d_i = x_{i+1} - x_i\,. \tag{16}$$

В (14) суммирование производится по $(N - 1)$ элементам. Для сравнения, по заданной вероятности β (или уровню значимости *q*), числу степеней свободы *r* и с использованием статистических таблиц *критерия Аббе* [Приложение Д] получают контрольную величину δ_q. Тогда, при $\delta > \delta_q$ принимается гипотеза об отсутствии систематической ошибки с вероятностью $\beta = 1 - q$. В противном случае $(\delta < \delta_q)$ следует принять гипотезу о постоянной составляющей в статистической совокупности и для корректной оценки исследуемых параметров ее необходимо исключить из ряда измерений. Для этого получают усредненную величину систематического влияния, равную среднему арифметическому из всех элементов. Данную величину исключают из измерений, получая новый ряд $X_{исп}$ с уменьшенной по сравнению с исходным рядом систематической составляющей.

$$x_{ист_i} = x_i - \bar{X} \ . \tag{17}$$

При N > 60 величина δ_q может быть вычислена по формуле

$$\delta_q = 1 + \frac{n_q}{\sqrt{N + 0.5 \cdot \left(1 + n_q^2\right)}}, \tag{18}$$

где n_q – квантиль нормированного нормального распределения. Для его нахождения в Microsoft Excel 2007 используется формула

= НОРМСТОБР(q) (19)

Например, при q=0.05 будет получено значение $n_q = -1.64485$

Анализ формул (12) – (14) говорит о том, что постоянная часть будет значима в измерениях, когда величины уклонений от среднего минимальны, а последовательные разности максимальны.

Исследование на наличие грубых погрешностей. При отсутствии систематической составляющей, или после их исключения по (17) проводим исследование на наличие в ряду грубых погрешностей. В зависимости от требований задачи существует масса критериев, решающих поставленную задачу: критерий Граббса, Диксона, Шарлье, Шовенэ и др. В работе для выявления грубых погрешностей предлагается использовать критерий Граббса. Критерий дает вероятность выполнения выдвинутой гипотезы о том, что максимальное, или минимальное значение из ряда не являются грубыми погрешностями. Для этого по экстремальным значениям выборки X_{max} и X_{min}, среднему арифметическому и средней квадратической погрешности, вычисляют значения

$$\left.\begin{aligned} z_{max} &= \frac{X_{max} - \bar{X}}{m} \\ z_{min} &= \frac{\bar{X} - X_{min}}{m} \end{aligned}\right\} \tag{20}$$

Если $z_{выч} \le z_q$ для максимального и минимального значения, то следует принять гипотезу об отсутствии в ряду грубых погрешностей, так как экстремальные значения не являются грубыми. Значения теоретической величины критерия z_q получают по заданному аргументу q и числу элементов в выборке N по специальным статистическим таблицам критерия Граббса для z_q (Приложение Ж), взятым из [7, табл.IX] Если же $z_{выч} > z_q$, тогда или наибольшее или наименьшее значение ряда из дальнейшей обработки следует исключить, а к новым крайним значениям еще раз применить критерий *Граббса.*

Для N > 25 значения z_q можно с приемлемой точностью рассчитать по уравнениям, указанным в таблице.

Таблица – Вычисление критических значений z_q

q	z_q при N > 25
0.1	0.3053·Ln(N) + 1.6513
0.05	0.2849·Ln(N) + 1.9517
0.01	0.2648·Ln(N) + 2.4839

Выявление неоднородности в исследуемом ряду. Для выявления степени *неоднородности* (неравноточности) результатов измерений используем самый простой *F-критерий Фишера*. Для этого разбивают исследуемую выборку на две одинаковые подвыборки и для каждой вычисляются средние квадратические погрешности m_1 и m_2. Тогда, если вычисленное значение статистики Фишера $F_{выч} < F_{таб}$ (табличному значению), то принимается гипотеза с вероятностью β об отсутствии неравноточности (разнородности) в ряду исследуемых величин. Здесь величина $F_{выч} = m^2_1 / m^2_2$, причем в числителе должна быть большая дисперсия. Значение $F_{таб}$ выбирают из таблиц распределения *Фишера* по n_1 (для числителя) и n_2 (для знаменателя) степеням свободы и вероятности (1+β)/2 [Приложение Г]. Также расчёт можно произвести в Excel по формуле (3.5).

Таким образом, ряд исследуемых случайных величин будет подготовлен в соответствии с *центральной предельной теоремой Ляпунова* для корректного исследования на соответствие нормальному закону распределения.

Вычисление основных характеристик ряда

Как было показано выше, к основным характеристикам закона распределения относят *математическое ожидание* $M(X)$ (так называемый *начальный момент первого порядка*) и стандарт σ (корень из *центрального момента второго порядка*). Эти величины теоретические и могут быть точно определены только при числе элементов в ряду N = ∞. В практической деятельности используют какой-либо вид их оценок, которые вычислимы при конечном N. Для нормального закона это соответственно *среднее арифметическое* $\overline{X} = X_{ср.}$ и *средняя квадратическая погрешность m.*

Оценки математического ожидания, дисперсии и стандарта получим по следующим формулам

$$M(X) = \frac{1}{N} \cdot \sum_{i=1}^{N} x_i \, , \qquad (21)$$

$$D(X) = \frac{1}{N-1} \cdot \sum_{i=1}^{N} (x_i - \bar{X})^2 \, , \qquad (22)$$

$$\sigma = \sqrt{D(X)} \approx m\,, \tag{23}$$

Если в качестве исследуемых величин берутся истинные ошибки (например, *невязки*), то сумму в формулах (21) – (22) для усреднения делят на количество элементов N (форма Гаусса); если в качестве погрешностей используют уклонения от оценки (например, от среднего арифметического), то в качестве делителя берут число степеней свободы N – 1 (формула Бесселя).

Приближенные критерии исследования

Приближенные критерии исследования ряда погрешностей на соответствие нормальному закону распределения используют сравнение некоторых известных теоретических характеристик нормального закона и их вычисленных по результатам измерений аналогов. Степень отличия теоретических величин от вычисленных оценок говорит о степени приближения исследуемого ряда к нормальному закону.

Кроме наиболее распространенной средней квадратической погрешности *m* используют *средние абсолютные* v и *вероятные* (*срединные*) ошибки *r*, которые являются оценками теоретических *абсолютных центральных моментов первого порядка* и 0.5 (или 2) – *квантили закона распределения Гаусса* соответственно. Между тремя ошибками m, v и r для нормального закона распределения величин имеются теоретические строгие соотношения

$$\left.\begin{aligned} k_1 &= \frac{m}{v} = 1.25 \\ k_2 &= \frac{m}{r} = 1.48 \\ k_3 &= \frac{v}{r} = 1.18 \end{aligned}\right\} \tag{24}$$

По величине отклонения вычисленных значений от их теоретических аналогов можно судить о степени приближения ряда к закону Гаусса, а также возможно использовать (24) для вычисления любых других двух погрешностей, если известна одна из трех. В качестве меры значимости отличия вычисленной величины от ее теоретического значения можно использовать, например, "критерий ничтожных влияний", гласящий, что величина считается неизменной, если ее вариация составляет не более 11 % от самой величины. Например, для первой формулы из (24) имеем

$$\left|\frac{m}{v} - 1.25\right| \le 0.11 \cdot k_1 \tag{25}$$

Тогда отклонения коэффициентов считаются допустимыми, если их абсолютная разность не более 0.138, 0.163 и 0.130 соответственно.

Для вычисления *средней абсолютной ошибки* пользуются формулами

$$v = \frac{1}{N} \cdot \sum_{i=1}^{N} |x_i| \,, \tag{26}$$

и

$$v = \frac{1}{N-1} \cdot \sum_{i=1}^{N} |x_i - \bar{X}| \,, \tag{27}$$

При этом формула (26) используется, если ошибки x_i – истинные и оценка математического ожидания равна нулю.

Чтобы определить *вероятную ошибку,* величины располагают в так называемый абсолютный вариационный ряд (по возрастанию их абсолютных значений). Если в ряду нечетное число элементов, то искомая ошибка равна значению величины, находящейся точно по середине (величина с номером (N+1)/2 в построенном вариационном ряду). Если число исследуемых величин четное, то значение ошибки находится как среднее арифметическое из двух чисел, стоящих в середине ряда (среднее между элементами с номерами N/2 и (N+1)/2).

К дополнительным характеристикам распределения случайных величин, составляющих статистическую совокупность, называемым характеристиками формы, относят *эксцесс E* – меру “крутости” и *асимметрию A* – меру “скошенности”, которые получаются с использованием моментов более высокого порядка. При этом следует помнить, что для нормального закона распределения теоретические значения *асимметрии* и *эксцесса* равны нулю. Тогда значения эксцесса и асимметрии можно считать несущественными при условиях

$$|A| \le t \cdot m_A \,, \tag{28}$$

и

$$|E| \le t \cdot m_E \,, \tag{29}$$

где

$$A = \frac{1}{N \cdot m^3} \cdot \sum_{i=1}^{N} (x_i - \bar{X})^3 \,, \tag{30}$$

$$E = \frac{1}{N \cdot m^4} \cdot \left(\sum_{i=1}^{N} (x_i - \bar{X})^4 \right) - 3 \,, \tag{31}$$

Значение вероятностного коэффициента t в (28), (29) выбирают в зависимости от вероятности изменяя её как (1+ β)/2, так как интервал двухсторонний и используя таблицы нормального распределения. Для наиболее часто используемой вероятности 0.95 коэффициент для двухстороннего интервала будет равен 1.96.

Эмпирические значения средних квадратических погрешностей дополнительных характеристик распределения могут быть вычислены по сокращенным формулам

$$m_A = \sqrt{\frac{24}{N}}, \tag{32}$$

и

$$m_A = \sqrt{\frac{6}{N}}, \tag{33}$$

Следует иметь ввиду, что если при вычислении асимметрии A её значение окажется больше нуля, то кривая эмпирического распределения будет скошена влево, а когда $A < 0$ – то вправо. При величине эксцесса $E > 0$ эмпирическое распределение "высоковершинное" (его вершина выше вершины теоретической кривой нормального распределения). В противном случае ($E < 0$) распределение "низковершинное".

Вычисления оценок параметров нормального распределения исследуемого ряда величин (среднего арифметического и средней квадратической погрешности), а также ее дополнительных характеристик (коэффициентов k, асимметрии и эксцесса) позволяют сделать только предварительное заключение о соответствии эмпирического распределения теоретическому нормальному закону Гаусса и только по близости соответствующих теоретических характеристик распределения их вычисленным аналогам.

Графический критерий исследования

Для дальнейших исследований погрешностей на соответствие их нормальному закону распределения строят для ряда одно из его графических представлений, например, в виде гистограммы или многоугольника распределения, с нанесенной поверх её теоретической кривой закона Гаусса с параметрами $\bar{X}$ и m, называемой *огивой*.

В данной работе предлагается использовать *гистограмму*. Построение гистограммы начинают с разбиения ряда погрешностей на интервалы. Число интервалов k зависит от точности измерений, количества элементов в выборке и является в некотором смысле произвольным. Основное требование к количеству и величине интервалов заключается в том, чтобы полученный на их основе график был наглядным и правдоподобным. Длину интервала Q можно получить, например, используя следующие формулы

$$Q = \frac{(X_{max} - X_{min})}{k}, \qquad (34)$$

если известно число интервалов k, и

$$Q = \frac{(X_{max} - X_{min})}{1 + \log_2 N}, \qquad (35)$$

если используется количество элементов N в ряду (формула *Г. Стерджесса*). Полученное значение длины интервала Q округляют в сторону уменьшения до удобного для дальнейших вычислений числа. При этом значения крайних элементов ряда (максимальное и минимальное) должны обязательно лежать в пределах вычисленного интервала. Сами длины интервалов принято выражать в долях средней квадратической погрешности m соответствующих им единицах произведенных измерений (секунды, метры и т.д.).

В геодезии чаще всего в такого рода исследованиях ряд делят на 12 интервалов, каждый из которых должен иметь размер 0.5 m.

Далее необходимо подсчитать число N_j элементов ряда, принадлежащих j-му интервалу, и вычислить практические оценки неизвестных вероятностей (частоты Q_j) по формуле

$$Q = \frac{N_j}{N}, \qquad (36)$$

При этом необходимо проследить, чтобы сумма частот N_j по всем интервалам равнялась числу всех элементов в выборке, а сумма частот Q_j была равна единице в пределах ошибки округления.

Для построения гистограммы исследуемых случайных величин значение j-той частоты делят на длину принятого интервала, обычно равного половине средней квадратической ошибки. Таким образом, получают вертикальные составляющие гистограммы, называемые высотами прямоугольников

$$h_j = \frac{Q_j}{0.5 \cdot m}, \qquad (37)$$

Результаты вычислений целесообразно представлять в таблице с соответствующими контролями по числу элементов и частоте в интервалах.

Таблица – Вычисление значений для гистограммы эмпирического распределения

Интервал в долях m	Интервал в ед. измерения		Абсолютная частота N_j	Относительная частота $Q_j=N_j/N$	Сторона прямоуг. $h_j=Q_j/(0.5\cdot m)$
[-3.0m ; -2.5m]	-2.7499	-2.2916	0	0.00	0.00
(-2.5m ; -2.0m]	-2.2916	-1.8332	2	0.04	0.09
(-2.0m ; -1.5m]	-1.8332	-1.3749	0	0.00	0.00
(-1.5m ; -1.0m]	-1.3749	-0.9166	3	0.06	0.13

(-1.0m ; -0.5m]	-0.9166	-0.4583	8	0.16	0.35
(-0.5m ; 0m]	-0.4583	0.0000	7	0.14	0.31
(0m ; 0.5m]	0.0000	0.4583	10	0.20	0.44
(0.5m ; 1.0m]	0.4583	0.9166	10	0.20	0.44
(1.0m ; 1.5m]	0.9166	1.3749	7	0.14	0.31
(1.5m ; 2.0m]	1.3749	1.8332	2	0.04	0.09
(2.0m ; 2.5m]	1.8332	2.2916	1	0.02	0.04
(2.5m ; 3.0m]	2.2916	2.7499	0	0.00	0.00
		Σ =	50	1	

При таком представлении гистограммы площадь построенного прямоугольника будет равна величине соответствующей частоты, а общая площадь примерно единице. Построение эмпирического распределения производят по значениям длин интервалов (ось абсцисс) и высот прямоугольников (ось ординат). На выбор масштабов накладывается лишь условие наглядности. Вид гистограммы дает возможность предположить о мере соответствия исследуемых величин нормальному закону распределения Гаусса.

На этом же графике необходимо построить теоретическую кривую, соответствующую нормальному закону, которая наилучшим образом сглаживает (выравнивает) данное эмпирическое распределение. Кривая строится на основе формулы плотности вероятности для закона Гаусса

$$\varphi(X) = \frac{1}{m \cdot \sqrt{2\pi}} \cdot e^{\frac{-(X-\mu)^2}{2 \cdot m^2}}, \qquad (38)$$

Учитывая, что $t = (X - \mu)/m$ – это нормированные границы интервалов, введём обозначение

$$K =\cdot \frac{1}{m \cdot \sqrt{2\pi}}, \qquad (39)$$

тогда формулу (38) можно представить как

$$\varphi(X) = K \cdot e^{\frac{-t^2}{2}}, \qquad (40)$$

Обычно величина t изменяется от –3 до 3 через 0.5, так что вычисления не представляют трудности. Необходимо учитывать, что функция симметричная, т.е., $f(-x) = -f(x)$. Значения функции (40) с $m = 1$ приведены в книгах по статистике или обработке измерений в виде таблиц и также могут быть использованы при вычислениях.

По значениям t и функции (40) на рисунок гистограммы наносится ряд точек, которые соединяются плавной сплошной линией (*огивой*). Эта линия и будет соответствовать теоретической кривой нормального закона распределения со стандартными характеристиками, которая визуально сравнивается с практическим представлением распределения в виде гистограммы (Рисунок 1).

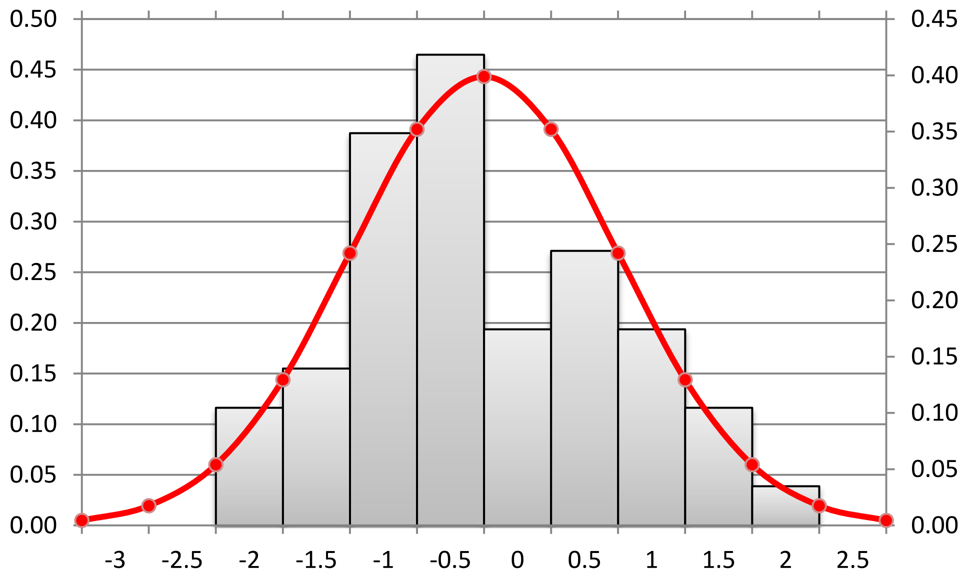

Рисунок 1 – Гистограмма и огива.

Точные критерии исследования

Строгую оценку степени соответствия эмпирического распределения теоретическому выполним на основе статистического аппарата *проверки гипотез*. Для этого задается уровень значимости q (вероятность невыполнения гипотезы) или β (вероятность выполнения гипотезы) и выдвигается гипотеза, что ряд исследуемых величин подчинен нормальному распределению Гаусса с вероятностью β. При этом, естественно, $\beta + q = 1$.

Для оценки степени приближения к нормальному закону по критерию χ^2 *Пирсона* следует вычислить величину

$$\chi^2 = \sum_{i=1}^{N} \frac{(N_i - N \cdot p_i)^2}{N \cdot p_i} \qquad (41)$$

Здесь p_i – теоретическая вероятность попадания случайной величины в соответствующий интервал, Данная величина вычисляется по значениям интеграла вероятности $\Phi(t_i)$, которые выбираются из статистических таблиц по величинам границ интервалов t_i

$$p_i = \frac{1}{2} \cdot \left(\Phi(t_{i+1}) - \Phi(t_i)\right) \qquad (42)$$

и не деля на два, если используются таблицы *нормированной функции Лапласа* $\Phi(t_i)$ (Приложение З). Для вычислений можно использовать также любой статистический программный пакет для расчета вероятности попадания величин в соответствующий интервал.

Степень согласованности эмпирического распределения с теоретическим можно провести двумя способами. Во-первых, задать вероятность β, по которой по числу степеней свободы r из статистических таблиц χ^2 Пирсона (Приложение В) или из какой-либо программы (например, Excel) получают эталонное значение $\chi^2_{эт}$. Если вычисленное значение меньше эталонного, то с вероятностью β можно принять выдвинутую гипотезу о соответствии ряда нормальному закону распределения.

При втором подходе по вычисленной величине χ^2 и числу степеней свободы из статистических таблиц выбирают вероятность β того, что выдвинутая гипотеза принимается. При этом для целей геодезии можно использовать следующие характеристики для вероятностей:

$\beta > 0.5$ – отличное согласование с выдвинутой гипотезой о нормальности;

$0.3 < \beta < 0.5$ – согласие считают хорошим;

$0.1 < \beta < 0.3$ – согласие удовлетворительное;

$\beta < 0.1$ – согласие считается неудовлетворительным.

Число степеней свободы r определяется как $r = k - 1 - s$, где k – число интервалов, s – число определяемых для характеристики закона распределения параметров (чаще всего определяют математическое ожидание и дисперсию, тогда $s = 2$).

Все вычисления целесообразно свести в таблицу

Таблица – Вычисление критерия χ^2 Пирсона

Интервал в долях m		Границы вероятностной ф-ии $\frac{1}{2}\cdot\Phi(t_i)$	Вероятность попадания в интервал p_i	Практическая абсолютная частота N_i	Теоретическая абсолютная частота $N\cdot p_i$	$N_i - N\cdot p_i$	χ^2
-3m	-2,5m	-0.499	0.0049	1	0.24	0.76	2.358
-2,5m	-2m	-0.494	0.0165	0	0.83	-0.83	0.827
-2m	-1,5m	-0.477	0.0441	1	2.20	-1.20	0.657
-1,5m	-1m	-0.433	0.0919	5	4.59	0.41	0.036
-1m	-0,5m	-0.341	0.1499	8	7.49	0.51	0.034
-0,5m	0m	-0.191	0.1915	11	9.57	1.43	0.213
0m	0,5m	0.000	0.1915	9	9.57	-0.57	0.034
0,5m	1m	0.191	0.1499	8	7.49	0.51	0.034
1m	1,5m	0.341	0.0919	4	4.59	-0.59	0.076
1,5m	2m	0.433	0.0441	2	2.20	-0.20	0.019
2m	2,5m	0.477	0.0165	0	0.83	-0.83	0.827
2,5m	3m	0.494	0.0049	1	0.24	0.76	2.358
3m		0.499					
			0.9973	50	49.87		7.474

Список литературы

1. Большаков В.Д. Теория ошибок наблюдений: Учебник для вузов. – 2-е изд., перераб. и доп. М., Недра, 1983. – 223 с.
2. Большаков В.Д., Маркузе Ю.И. Практикум по теории математической обработки геодезических измерений: Учебное пособие для вузов. – М.: Недра, 1984. – 352 с.
3. Чеботарёв А.С. Способ наименьших квадратов с основами теории вероятностей. – Издательство геодезической литературы, 1958. – 610 с.
4. Leick A. Adjustment Computations. – Department of Spatial Information Science and Engineering. University of Maine, 1980. – 245 p.
5. Leick A., Humphrey D. Adjustments with examples. – University of Maine, 1986. – 450 p.
6. Дегтярёв А.М. Вероятностно-статистические методы в геодезии. Конспект лекций. – Новополоцк: ПГУ, 2005. – 208 с.
7. Михелев, Д.Ш. Геодезические измерения при изучении деформаций крупных инженерных сооружений / Д.Ш. Михелев, И.В. Рунов, А.И. Голубцов. – М., «Недра», 1977, 152 с.

Приложения

Приложение А. Правила оформления лабораторных работ

Лабораторная работа должна быть выполнена на стандартной белой бумаге формата А4 по ГОСТ 2.301 с одной стороны листа.

Должны быть установлены стандартные поля по СТБ 6.38:

- левое поле – 30 мм;
- правое поле – 10 мм
- верхнее и нижнее поля – 20 мм.

Лабораторная работа должна быть оформлена в соответствии с ГОСТ 2.105 одним из следующих способов:

1. С применением печатающих и графических устройств вывода Персонального Компьютера (ПК) – ГОСТ 2.004 – шрифтом Times New Roman чёрного цвета с высотой 14 пт, через полтора интервала, в обычном начертании, выравнивая по ширине (пункт – единица, принятая в полиграфии: 1пт=1/72"=0.352мм).
2. Рукописным – чертёжным шрифтом по ГОСТ 2.304 с высотой не менее 2.5 мм, чёрными чернилами (пастой, тушью) – чётким почерком.

Абзацы в тексте начинают отступом 1.25 см, одинаковым по всему тексту.

Вписывать в отпечатанный текст отдельные слова, формулы, условные знаки, а также выполнять иллюстрации следует чёрными чернилами (пастой, тушью). Для выполнения иллюстраций разрешается использовать графические редакторы, фотографии, ксерокопии и т.п.

Для оформления формул используется встроенный в Microsoft Word стандартный текстовый редактор формул либо Microsoft Euqation (Вставка – Объект – Microsoft Euqation 3.0). Формулы и уравнения в тексте следует оформлять в соответствии с ГОСТ 2.105, раздел 4. Формула должна располагаться по центру страницы. Номер формулы указывается арабскими цифрами в круглых скобках в той же строке с выравниванием по правому краю. В формулах в качестве символов следует применять обозначения, установленные соответствующими государственными стандартами. Пояснения символов и числовых коэффициентов, входящих в формулу, если они не пояснены ранее в тексте, должны быть приведены непосредственно под формулой, при этом после формулы должна присутствовать запятая. В том случае, когда пояснения символов и численных коэффициентов, входящих в формулу пояснены ранее в тексте, после формулы должна ставиться точка. Пояснения каждого символа следует давать с новой строки в той последовательности, в которой символы приведены в формуле. Первая строка пояснения должна начинаться с новой строки (без абзацного отступа) со слова «где» и без двоеточия после него. Формулы, следующие одна за другой и не разделённые текстом, разделяют точкой с запятой. Ссылки в тексте на порядковые номера формул дают в скобках, например, «… в формуле (1.3)».

Иллюстрации следует располагать в работе непосредственно на странице с текстом после абзаца, в котором они упоминаются впервые, или отдельно на следующей странице. Иллюстрации обозначают словом «Рисунок» или «Рис.» и нумеруют последовательно. Иллюстрации нумеруют последовательно в пределах работы, например:

«Рисунок 1. Общеземной эллипсоид» без кавычек. Слово «Рисунок», его номер и наименование печатают обычным шрифтом под рисунком посередине строки размером шрифта 14 пт.

На все иллюстрации и таблицы должны быть даны ссылки в тексте работы.

Опечатки и описки допускается исправлять подчисткой или закрашиванием белой краской и нанесением на том же месте исправлений машинным или рукописным способом чёрными чернилами (пастой, тушью). Повреждение листов, помарки и следы прежнего текста не допускаются. Допускается не более трёх исправлений на одной странице.

В тексте работы не допускается применять сокращения слов (кроме установленных правилами орфографии и соответствующими государственными стандартами).

Приложение Б. Односторонние и двусторонние критические значения коэффициента Стьюдента (t-критерий)

Односторонний	P=0.90	0.95	0.975	0.99	0.995	0.9975	0.999	0.9995
Двусторонний	0.80	0.90	0.95	0.98	0.99	0.995	0.998	0.999
r = 1	3.0770	6.3130	12.7060	31.820	63.656	127.656	318.306	636.619
2	1.8850	2.9200	4.3020	6.964	9.924	14.089	22.327	31.599
3	1.6377	2.35340	3.182	4.540	5.840	7.458	10.214	12.924
4	1.5332	2.13180	2.776	3.746	4.604	5.597	7.173	8.610
5	1.4759	2.01500	2.570	3.649	4.0321	4.773	5.893	6.863
6	1.4390	1.943	2.4460	3.1420	3.7070	4.316	5.2070	5.958
7	1.4149	1.8946	2.3646	2.998	3.4995	4.2293	4.785	5.4079
8	1.3968	1.8596	2.3060	2.8965	3.3554	3.832	4.5008	5.0413
9	1.3830	1.8331	2.2622	2.8214	3.2498	3.6897	4.2968	4.780
10	1.3720	1.8125	2.2281	2.7638	3.1693	3.5814	4.1437	4.5869
11	1.363	1.795	2.201	2.718	3.105	3.496	4.024	4.437
12	1.3562	1.7823	2.1788	2.6810	3.0845	3.4284	3.929	4.178
13	1.3502	1.7709	2.1604	2.6503	3.1123	3.3725	3.852	4.220
14	1.3450	1.7613	2.1448	2.6245	2.976	3.3257	3.787	4.140
15	1.3406	1.7530	2.1314	2.6025	2.9467	3.2860	3.732	4.072
16	1.3360	1.7450	2.1190	2.5830	2.9200	3.2520	3.6860	4.0150
17	1.3334	1.7396	2.1098	2.5668	2.8982	3.2224	3.6458	3.965
18	1.3304	1.7341	2.1009	2.5514	2.8784	3.1966	3.6105	3.9216
19	1.3277	1.7291	2.0930	2.5395	2.8609	3.1737	3.5794	3.8834
20	1.3253	1.7247	2.08600	2.5280	2.8453	3.1534	3.5518	3.8495
21	1.3230	1.7200	2.2.0790	2.5170	2.8310	3.1350	3.5270	3.8190
22	1.3212	1.7117	2.0739	2.5083	2.8188	3.1188	3.5050	3.7921
23	1.3195	1.7139	2.0687	2.4999	2.8073	3.1040	3.4850	3.7676
24	1.3178	1.7109	2.0639	2.4922	2.7969	3.0905	3.4668	3.7454
25	1.3163	1.7081	2.0595	2.4851	2.7874	3.0782	3.4502	3.7251
26	1.315	1.705	2.059	2.478	2.778	3.0660	3.4360	3.7060
27	1.3137	1.7033	2.0518	2.4727	2.7707	3.0565	3.4210	3.6896
28	1.3125	1.7011	2.0484	2.4671	2.7633	3.0469	3.4082	3.6739
29	1.3114	1.6991	2.0452	2.4620	2.7564	3.0360	3.3962	3.8494
30	1.3104	1.6973	2.0423	2.4573	2.7500	3.0298	3.3852	3.6460
32	1.3080	1.6930	2.0360	2.4480	2.7380	3.0140	3.3650	3.6210
34	1.3070	1.6909	2.0322	2.4411	2.7284	3.9520	3.3479	3.6007
36	1.3050	1.6883	2.0281	2.4345	2.7195	9.490	3.3326	3.5821
38	1.3042	1.6860	2.0244	2.4286	2.7116	3.9808	3.3190	3.5657
40	1.303	1.6839	2.0211	2.4233	2.7045	3.9712	3.3069	3.5510
42	1.320	1.682	2.018	2.418	2.6980	2.6930	3.2960	3.5370
44	1.301	1.6802	2.0154	2.4141	2.6923	3.9555	3.2861	3.5258
46	1.300	1.6767	2.0129	2.4102	2.6870	3.9488	3.2771	3.5150
48	1.299	1.6772	2.0106	2.4056	2.6822	3.9426	3.2689	3.5051
50	1.298	1.6759	2.0086	2.4033	2.6778	3.9370	3.2614	3.4060
55	1.2997	1.673	2.0040	2.3960	2.6680	2.9240	3.2560	3.4760
60	1.2958	1.6706	2.0003	2.3901	2.6603	3.9146	3.2317	3.4602
65	1.2947	1.6686	1.997	2.3851	2.6536	3.9060	3.2204	3.4466
70	1.2938	1.6689	1.9944	2.3808	2.6479	3.8987	3.2108	3.4350
80	1.2820	1.6640	1.9900	2.3730	2.6380	2.8870	3.1950	3.4160
90	1.2910	1.6620	1.9867	2.3885	2.6316	2.8779	3.1833	3.4019
100	1.2901	1.6602	1.9840	2.3642	2.6259	2.8707	3.1737	3.3905
120	1.2888	1.6577	1.9719	2.3578	2.6174	2.8598	3.1595	3.3735
150	1.2872	1.6551	1.9759	2.3515	2.6090	2.8482	3.1455	3.3566
200	1.2858	1.6525	1.9719	2.3451	2.6006	2.8385	3.1315	3.3398
250	1.2849	1.6510	1.9695	2.3414	2.5966	2.8222	3.1232	3.3299
300	1.2844	1.6499	1.9679	2.3388	2.5923	2.8279	3.1176	3.3233

Приложение В. Квантили распределения χ^2 для различной доверительной вероятности P и числа степеней свободы r

	P														
	0,99	0,975	0,95	0,9	0,8	0,7	0,6	0,5	0,4	0,3	0,2	0,1	0,05	0,025	0,01
1	0,0002	0,0010	0,0039	0,0158	0,0642	0,1485	0,2750	0,4549	0,7083	1,0742	1,6424	2,7055	3,8415	5,0239	6,6349
2	0,0201	0,0506	0,1026	0,2107	0,4463	0,7133	1,0217	1,3863	1,8326	2,4079	3,2189	4,6052	5,9915	7,3778	9,2103
3	0,1148	0,2158	0,3518	0,5844	1,0052	1,4237	1,8692	2,3660	2,9462	3,6649	4,6416	6,2514	7,8147	9,3484	11,3449
4	0,2971	0,4844	0,7107	1,0636	1,6488	2,1947	2,7528	3,3567	4,0446	4,8784	5,9886	7,7794	9,4877	11,1433	13,2767
5	0,5543	0,8312	1,1455	1,6103	2,3425	2,9999	3,6555	4,3515	5,1319	6,0644	7,2893	9,2364	11,0705	12,8325	15,0863
6	0,8721	1,2373	1,6354	2,2041	3,0701	3,8276	4,5702	5,3481	6,2108	7,2311	8,5581	10,6446	12,5916	14,4494	16,8119
7	1,2390	1,6899	2,1673	2,8331	3,8223	4,6713	5,4932	6,3458	7,2832	8,3834	9,8032	12,0170	14,0671	16,0128	18,4753
8	1,6465	2,1797	2,7326	3,4895	4,5936	5,5274	6,4226	7,3441	8,3505	9,5245	11,0301	13,3616	15,5073	17,5345	20,0902
9	2,0879	2,7004	3,3251	4,1682	5,3801	6,3933	7,3570	8,3428	9,4136	10,6564	12,2421	14,6837	16,9190	19,0228	21,6660
10	2,5582	3,2470	3,9403	4,8652	6,1791	7,2672	8,2955	9,3418	10,4732	11,7807	13,4420	15,9872	18,3070	20,4832	23,2093
11	3,0535	3,8157	4,5748	5,5778	6,9887	8,1479	9,2373	10,3410	11,5298	12,8987	14,6314	17,2750	19,6751	21,9200	24,7250
12	3,5706	4,4038	5,2260	6,3038	7,8073	9,0343	10,1820	11,3403	12,5838	14,0111	15,8120	18,5493	21,0261	23,3367	26,2170
13	4,1069	5,0088	5,8919	7,0415	8,6339	9,9257	11,1291	12,3398	13,6356	15,1187	16,9848	19,8119	22,3620	24,7356	27,6882
14	4,6604	5,6287	6,5706	7,7895	9,4673	10,8215	12,0785	13,3393	14,6853	16,2221	18,1508	21,0641	23,6848	26,1189	29,1412
15	5,2293	6,2621	7,2609	8,5468	10,3070	11,7212	13,0297	14,3389	15,7332	17,3217	19,3107	22,3071	24,9958	27,4884	30,5779
16	5,8122	6,9077	7,9616	9,3122	11,1521	12,6243	13,9827	15,3385	16,7795	18,4179	20,4651	23,5418	26,2962	28,8454	31,9999
17	6,4078	7,5642	8,6718	10,0852	12,0023	13,5307	14,9373	16,3382	17,8244	19,5110	21,6146	24,7690	27,5871	30,1910	33,4087
18	7,0149	8,2307	9,3905	10,8649	12,8570	14,4399	15,8932	17,3379	18,8679	20,6014	22,7595	25,9894	28,8693	31,5264	34,8053
19	7,6327	8,9065	10,1170	11,6509	13,7158	15,3517	16,8504	18,3377	19,9102	21,6891	23,9004	27,2036	30,1435	32,8523	36,1909
20	8,2604	9,5908	10,8508	12,4426	14,5784	16,2659	17,8088	19,3374	20,9514	22,7745	25,0375	28,4120	31,4104	34,1696	37,5662
21	8,8972	10,2829	11,5913	13,2396	15,4446	17,1823	18,7683	20,3372	21,9915	23,8578	26,1711	29,6151	32,6706	35,4789	38,9322
22	9,5425	10,9823	12,3380	14,0415	16,3140	18,1007	19,7288	21,3370	23,0307	24,9390	27,3015	30,8133	33,9244	36,7807	40,2894
23	10,1957	11,6886	13,0905	14,8480	17,1865	19,0211	20,6902	22,3369	24,0689	26,0184	28,4288	32,0069	35,1725	38,0756	41,6384
24	10,8564	12,4012	13,8484	15,6587	18,0618	19,9432	21,6525	23,3367	25,1063	27,0960	29,5533	33,1962	36,4150	39,3641	42,9798
25	11,5240	13,1197	14,6114	16,4734	18,9398	20,8670	22,6156	24,3366	26,1430	28,1719	30,6752	34,3816	37,6525	40,6465	44,3141
26	12,1981	13,8439	15,3792	17,2919	19,8202	21,7924	23,5794	25,3365	27,1789	29,2463	31,7946	35,5632	38,8851	41,9232	45,6417
27	12,8785	14,5734	16,1514	18,1139	20,7030	22,7192	24,5440	26,3363	28,2141	30,3193	32,9117	36,7412	40,1133	43,1945	46,9629
28	13,5647	15,3079	16,9279	18,9392	21,5880	23,6475	25,5093	27,3362	29,2486	31,3909	34,0266	37,9159	41,3371	44,4608	48,2782
29	14,2565	16,0471	17,7084	19,7677	22,4751	24,5770	26,4751	28,3361	30,2825	32,4612	35,1394	39,0875	42,5570	45,7223	49,5879
30	14,9535	16,7908	18,4927	20,5992	23,3641	25,5078	27,4416	29,3360	31,3159	33,5302	36,2502	40,2560	43,7730	46,9792	50,8922
31	15,6555	17,5387	19,2806	21,4336	24,2551	26,4397	28,4087	30,3359	32,3486	34,5981	37,3591	41,4217	44,9853	48,2319	52,1914
32	16,3622	18,2908	20,0719	22,2706	25,1478	27,3728	29,3763	31,3359	33,3809	35,6649	38,4663	42,5847	46,1943	49,4804	53,4858
33	17,0735	19,0467	20,8665	23,1102	26,0422	28,3069	30,3444	32,3358	34,4126	36,7307	39,5718	43,7452	47,3999	50,7251	54,7755
34	17,7891	19,8063	21,6643	23,9523	26,9383	29,2421	31,3130	33,3357	35,4438	37,7954	40,6756	44,9032	48,6024	51,9660	56,0609
35	18,5089	20,5694	22,4650	24,7967	27,8359	30,1782	32,2821	34,3356	36,4746	38,8591	41,7780	46,0588	49,8018	53,2033	57,3421
36	19,2327	21,3359	23,2686	25,6433	28,7350	31,1152	33,2517	35,3356	37,5049	39,9220	42,8788	47,2122	50,9985	54,4373	58,6192
37	19,9602	22,1056	24,0749	26,4921	29,6355	32,0532	34,2216	36,3355	38,5348	40,9839	43,9782	48,3634	52,1923	55,6680	59,8925
38	20,6914	22,8785	24,8839	27,3430	30,5373	32,9919	35,1920	37,3355	39,5643	42,0451	45,0763	49,5126	53,3835	56,8955	61,1621
39	21,4262	23,6543	25,6954	28,1958	31,4405	33,9315	36,1628	38,3354	40,5935	43,1053	46,1730	50,6598	54,5722	58,1201	62,4281
40	22,1643	24,4330	26,5093	29,0505	32,3450	34,8719	37,1340	39,3353	41,6222	44,1649	47,2685	51,8051	55,7585	59,3417	63,6907
41	22,9056	25,2145	27,3256	29,9071	33,2506	35,8131	38,1055	40,3353	42,6506	45,2236	48,3628	52,9485	56,9424	60,5606	64,9501
42	23,6501	25,9987	28,1440	30,7654	34,1574	36,7550	39,0774	41,3352	43,6786	46,2817	49,4560	54,0902	58,1240	61,7768	66,2062
43	24,3976	26,7854	28,9647	31,6255	35,0653	37,6975	40,0496	42,3352	44,7063	47,3390	50,5480	55,2302	59,3035	62,9904	67,4593
44	25,1480	27,5746	29,7875	32,4871	35,9743	38,6408	41,0222	43,3352	45,7336	48,3957	51,6389	56,3685	60,4809	64,2015	68,7095
45	25,9013	28,3662	30,6123	33,3504	36,8844	39,5847	41,9950	44,3351	46,7607	49,4517	52,7288	57,5053	61,6562	65,4102	69,9568
46	26,6572	29,1601	31,4390	34,2152	37,7955	40,5292	42,9682	45,3351	47,7874	50,5071	53,8177	58,6405	62,8296	66,6165	71,2014
47	27,4158	29,9562	32,2676	35,0814	38,7075	41,4744	43,9417	46,3350	48,8139	51,5619	54,9056	59,7743	64,0011	67,8206	72,4433
48	28,1770	30,7545	33,0981	35,9491	39,6205	42,4201	44,9154	47,3350	49,8401	52,6161	55,9926	60,9066	65,1708	69,0226	73,6826
49	28,9406	31,5549	33,9303	36,8182	40,5344	43,3664	45,8895	48,3350	50,8660	53,6697	57,0786	62,0375	66,3386	70,2224	74,9195
50	29,7067	32,3574	34,7643	37,6886	41,4492	44,3133	46,8638	49,3349	51,8916	54,7228	58,1638	63,1671	67,5048	71,4202	76,1539

Приложение Г. Значения критерия Фишера (F-критерия) для уровня значимости α = 0.05

r_1 - число степеней свободы большей дисперсии,
r_2 - число степеней свободы меньшей дисперсии.

	r_1										
r_2	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**15**
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	245.95
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.43
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.70
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.86
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.62
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	3.94
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.51
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.22
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.01
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.85
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.72
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.62
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.53
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.46
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.40
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.35
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.31
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.27
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.23
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.20

Приложение Д. Критерий Аббе

P / n	0,001	0,01	0,05	P / n	0,001	0,01	0,05
4	0,2949	0,3128	0,3902	32	0,4963	0,6089	0,7177
5	2080	2690	4102	33	5027	6141	7216
6	1817	2808	4451	34	5090	6193	7256
7	1848	3070	4680	35	5150	6242	7292
8	2018	3314	4912	36	5208	6290	7328
9	0,2210	0,3544	0,5121	37	0,5265	0,6337	0,7363
10	2408	3759	5311	38	5319	6381	7396
11	2598	3957	5482	39	5373	6425	7429
12	2778	4140	5638	40	5425	6467	7461
13	2949	4309	5778	41	5475	6508	7491
14	0,3112	0,4466	0,5908	42	0,5524	0,6548	0,7521
15	3266	4611	6027	43	5571	6587	7550
16	3413	4746	6137	44	5616	6622	7576
17	3552	4872	6237	45	5660	6659	7603
18	3684	4989	6330	46	5701	6693	7628
19	0,3809	0,5100	0,6417	47	0,5743	0,6727	0,7653
20	3926	5203	6498	48	5781	6757	7676
21	4037	5301	6574	49	5817	6787	7698
22	4142	5393	6645	50	5853	6814	7718
23	4241	5479	6713	51	5887	6842	7739
24	0,4334	0,5562	0,6776	52	0,5922	0,6869	0,7759
25	4423	5639	6836	53	5955	6896	7779
26	4509	5713	6893	54	5989	6924	7799
27	4591	5784	6946	55	6020	6949	7817
28	4670	5850	6996	56	6051	6974	7836
29	0,4748	0,5915	0,7046	57	0,6083	0,6999	0,7853
30	4822	5975	7091	58	6114	7024	7872
31	4895	6034	7136	59	6145	7049	7891
				60	6174	7071	7906

Приложение Ж. Критерий Граббса

Значения z_q, удовлетворяющие уравнению $P\left(\frac{|x_э-\bar{x}|}{s}>z_q\right)=2q$,

где $s=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}$ для распределения $r=\frac{|x_э-\bar{x}|}{s}$

№	0,2%	0,5%	1%	2%	5%	10%	20%
3	1,414	1,414	1,414	1.414	1,414	1,412	1,406
4	1,731	1,730	1,728	1,723	1,710	1,689	1,645
5	1,990	1,982	1,972	1,955	1,917	1,869	1,791
6	2,203	2,183	2,161	2,130	2,067	1 996	1,894
7	2,377	2,344	2,310	2,265	2,182	2,093	1,974
8	2,521	2,476	2,431	2,374	2,273	2,172	2,041
9	2,643	2,586	2,532	2,464	2,349	2.238	2,097
10	2,747	2,680	2,616	2,540	2.414	2.294	2,146
11	2,837	2,760	2 689	2,606	2,470	2,343	2,190
12	2,915	2,830	2,753	2,663	2.519	2,387	2,220
13	2,984	2,892	2,809	2,713	2,563	2,426	2,264
14	3,046	2,947	2,859	2,759	2,602	2 461	2,297
15	3,102	2,997	2.905	2,800	2,638	2,494	2,327
16	3,152	3,042	2,946	2.837	2,670	2.523	2,354
17	3,198	3,083	2,983	2,871	2,701	2,551	2,380
18	3,240	3,120	3,017	2,903	2,728	2,577	2,404
19	3,278	3,155	3,049	2,932	2,754	2,601	2,426
20	3,314	3,187	3,079	2,959	2,779	2,623	2,447
21	3,347	3,217	3,106	2,984	2,801	2,644	2,467
22	3,378	3,245	3,132	3,008	2,823	2,664	2,486
23	3,407	3,271	3,156	3,030	2,843	2,683	2,504
24	3,434	3,295	3,179	3,051	2,862	2,701	2,521
25	3,458	3,318	3,200	3,071	2,880	2,718	2,537
26	3,483	3,340	3,220	3,089	2,897	2,734	2,553
27	3,506	3,360	3,239	3,107	2,913	2,749	2,568
28	3,528	3,380	3,258	3,124	2,929	2,764	2,582
29	3,548	3,399	3,275	3,140	2,944	2,778	2,596
30	3,567	3,416	3,291	3,156	2,958	2,792	2,609
31	3,586	3,433	3.307	3,171	2,972	2,805	2,622
32	3,603	3,449	3,322	3,185	2,985	2,818	2,634
33	3,620	3,465	3,337	3,199	2,998	2,830	2,646
34	3,636	3,480	3,351	3,212	3,010	2,842	2,657
35	3,652	3,494	3,364	3,244	3,022	2,853	2,668
36	3,667	3,507	3,377	3,236	3,033	2,864	2,679
37	3,681	3,521	3,389	3,248	3,044	2,874	2,689
38	2,695	3,533	3,401	3,259	3,055	2,885	2,699
39	3,708	3,545	3,413	3,270	3,065	2,894	2,709
40	3,720	3,557	3,424	3,281	3,075	2,904	2,718

Приложение 3. Нормированная функция Лапласа

$$\Phi(t) = \frac{2}{\sqrt{2\pi}} \int_0^t e^{-\frac{1}{2}x^2} dx = \frac{2}{\sqrt{2\pi}} \left(t - \frac{t^3}{6} + \frac{t^5}{40} - \frac{t^7}{336} + \frac{t^9}{3456} - \dots \right)$$

t	$\Phi(t)$	t	$\Phi(t)$	t	$\Phi(t)$
0,00	0,00000	1,25	0,78870	2,50	0,98758
0,05	0,03988	1,30	0,80640	2,55	0,98922
0,10	0,07966	1,35	0,82298	2,60	0,99068
0,15	0,11924	1,40	0,83849	2,65	0,99195
0,20	0,15852	1,45	0,85294	2,70	0,99307
0,25	0,19741	1,50	0,86639	2,75	0,99404
0,30	0,23582	1,55	0,87886	2,80	0,99489
0,35	0,27366	1,60	0,89040	2,85	0,99563
0,40	0,31084	1,65	0 90106	2,90	0,99627
0,45	0,34729	1,70	0,91087	2,95	0,99682
0,50	0,38292	1,75	0,91988	3,00	0,99730
0,55	0,41768	1,80	0,92814	3,10	0,99806
0,60	0,45149	1,85	0,93569	3,20	0,99863
0,65	0,48431	1,90	0,94257	3,30	0,99903
0,70	0,51607	1,95	0,94882	3,40	0,99933
0,75	0,54675	2,00	0,95450	3,50	0,99953
0,80	0,57629	2,05	0,95964	3,60	0,99968
0,85	0,60468	2,10	0,96427	3,70	0,99978
0,90	0,63188	2,15	0,96844	3,80	0,99986
0,95	0,65789	2,20	0,97219	3,90	0,44440
1,00	0,68269	2,25	0,97555	4,00	0,99994
1,05	0,70628	2,30	0,97855	4,10	0,99996
1,10	0,72867	2,35	0,98123	4,20	0,99997
1,15	0,74986	2,40	0,98360	4,40	0,99999
1,20	0,76986	2,45	0,98571	4,50	0,599994