БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ФАКУЛЬТЕТ РАДИОФИЗИКИ И КОМПЬЮТЕРНЫХ ТЕХНОЛОГИЙ
Кафедра системного анализа и компьютерного моделирования

А. В. Дигрис, В. М. Молофеев

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Методические указания к лабораторным работам по разделу «математическая статистика»

Для студентов специальностей
1-31 04 02 «Радиофизика»,
1-31 04 04 «Аэрокосмические радиоэлектронные и информационные системы и технологии»
1-98 01 01 «Компьютерная безопасность»,
1-31 03 07 «Прикладная информатика»

МИНСК
2023

УДК 519.2(075.8)
ББК 22.17

Рекомендовано советом факультета
радиофизики и компьютерных технологий
_______________ 2023 г., протокол № ____

Дигрис, А. В., Молофеев, В. М.
Теория вероятностей и математическая статистика: метод. указания к лабораторным работам / А.В. Дигрис, В. М. Молофеев – Минск: БГУ, 2023. – 37 с.

Методические указания предназначены для проведения лабораторных работ по разделу «Математическая статистика» дисциплины «Теория вероятностей и математическая статистика». Содержат учебный материал по практическому применению методов статистического оценивания параметров распределения и проверки статистических гипотез.

Предназначено для студентов факультета радиофизики и компьютерных технологий БГУ.

ВВЕДЕНИЕ

Настоящее издание является учебно-методическим пособием для выполнения лабораторных работ по разделу «Математическая статистика» дисциплины «Теория вероятностей и математическая статистика» для студентов факультета радиофизики и компьютерных технологий.

Цель данного пособия состоит в обучении студентов практическим навыкам по статистической обработке экспериментальных данных, методам статистического оценивания параметров распределения и проверки статистических гипотез.

Для выполнения лабораторных работ предлагается использовать пакет «Анализ данных» электронных таблиц MS Excel.

Приведенные в издании теоретические сведения содержат достаточный объем информации, позволяющий выполнять задания лабораторных работ, не обращаясь к другим литературным источникам. Наличие значительного набора вариантов заданий к каждой лабораторной работе позволяет индивидуально оценить уровень знаний каждого отдельного студента.

ТЕМА 1. МЕТОДЫ ПЕРВИЧНОЙ ОБРАБОТКИ СТАТИСТИЧЕСКИХ ДАННЫХ

Цель работы. Изучение алгоритмов первичной обработки статистических данных, построение эмпирического закона распределения, построение доверительных интервалов для математического ожидания и дисперсии.

Краткие сведения. Для работы со статистическими данными может использоваться известное приложение MS Excel, являющееся частью пакета Microsoft Office. С целью генерации выборок случайных значений, а также использования готового набора функций для статистической обработки данных, в MS Excel применяется надстройка «Пакет анализа».

Настройка доступа к пакету для анализа данных MS Excel. Для получения доступа к пакету анализа данных в MS Excel необходимо перейти в меню «Файл» и в нижней части данного меню выбрать пункт «Параметры». В открывшемся диалоговом окне, в списке слева, необходимо выбрать пункт «Надстройки» (рисунок 1.1), после чего в правой нижней части этого же окна нажать на кнопку «Перейти».

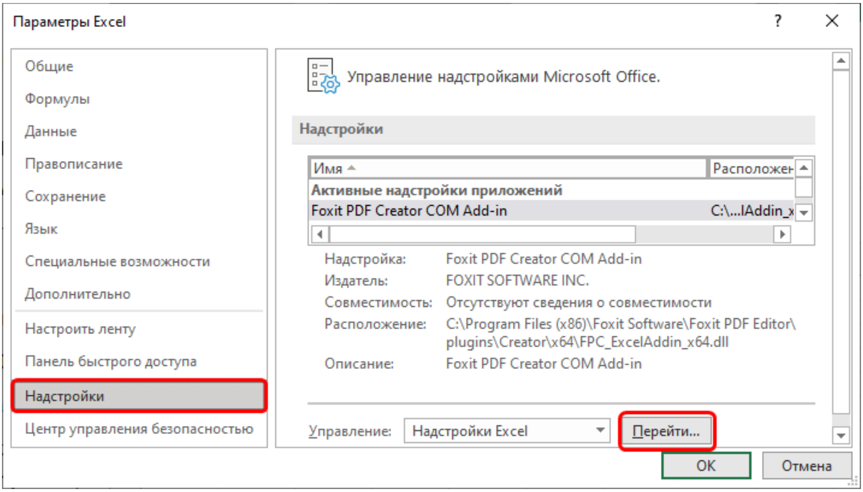

Рис. 1.1. Диалоговое окно настройки параметров MS Excel

В открывшемся дополнительном диалоговом окне для выбора надстроек (рисунок 1.2) необходимо отметить флажок «Пакет анализа» и далее закрыть оба диалоговых окна путем нажатия кнопки «ОК».

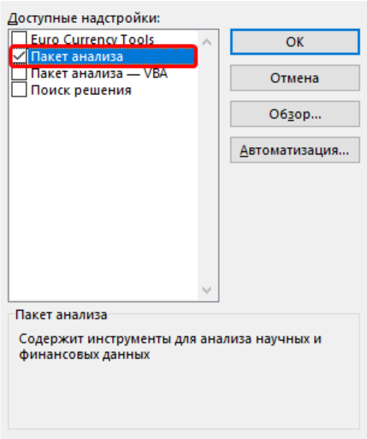

Рис. 1.2. Диалоговое окно для выбора надстроек

После описанных выше действий в разделе меню «Данные» MS Excel, справа, появится кнопка «Анализ данных», которая позволяет получить доступ к инструментам надстройки «Пакет анализа» (рисунок 1.3).

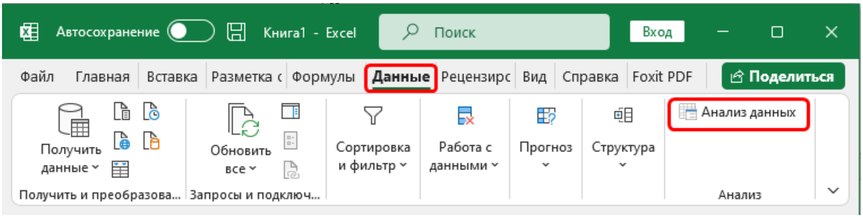

Рис. 1.3. Пункт меню для доступа к надстройке «Пакет анализа»

Моделирование статистических данных с использованием надстройки «Пакет анализа» MS Excel. Для моделирования средствами пакета для анализа данных MS Excel последовательности реализаций случайной величины с некоторым законом распределения необходимо, используя пункт меню «Анализ данных» (рисунок 1.3), открыть окно выбора инструментов для пакета анализа данных (рисунок 1.4). В данном окне надо в списке выбрать инструмент «Генерация случайных чисел», после чего нажать кнопку «OK», что приведет к переходу в окно настроек вышеуказанного инструмента.

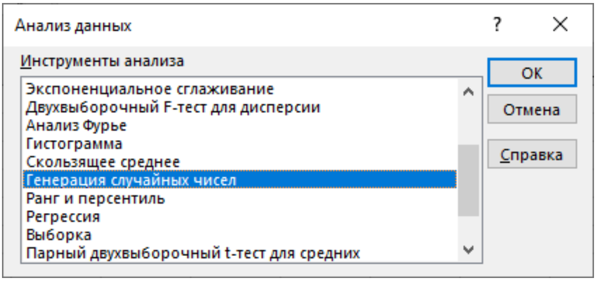

Рис. 1.4. Окно выбора инструментов из пакета для анализа данных

Для моделирования выборки непрерывной случайной величины с ***нормальным распределением $N(m, \sigma)$*** в окне настроек инструмента «Генерация случайных чисел» необходимо указать следующие значения:

Число переменных: 1

Число случайных чисел: в соответствии с длиной выборки

Распределение: Нормальное

Параметры

Среднее: значение математического ожидания (параметр m)

Стандартное отклонение: параметр σ

Параметры вывода

Выходной интервал: ячейка, начиная с которой будет выведен столбец сгенерированных значений.

Пример настроек инструмента «Генерация случайных чисел» для моделирования выборки из 1000 реализаций случайной величины, подчиняющейся нормальному распределению $N(0,1)$ приведен на рисунке 1.5. Значения будут выведены на текущий лист книги MS Excel, начиная с ячейки в первой строке колонки A.

Моделирование выборки для непрерывной случайной величины с ***экспоненциальным распределением $E(\lambda)$*** осуществляется с использованием метода обратных функций и реализуется в два этапа.

На первом этапе при помощи инструмента «Генерация случайных чисел» моделируется выборка значений с равномерным распределением на отрезке [0;1] (пример настроек приведен на рисунке 1.6)

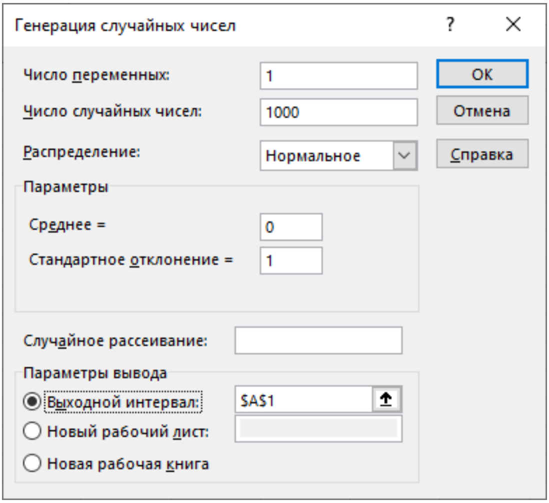

Рис. 1.5. Настройки инструмента «Генерация случайных чисел» для моделирования выборки из 1000 реализаций нормальной случайной величины $N(0,1)$

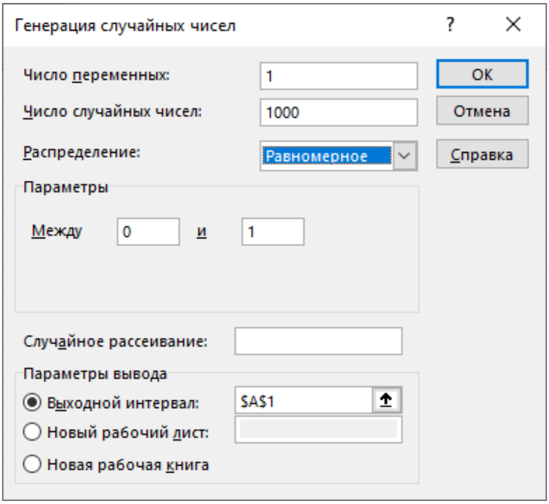

Рис. 1.6. Настройки инструмента «Генерация случайных чисел» для моделирования выборки из 1000 реализаций непрерывной случайной величины с равномерным распределением на отрезке [0;1]

На втором этапе с использованием полученной выборки для равномерного распределения вычисляют реализации экспоненциально распределенной случайной величины y_i по формуле:

$$y_i = -\frac{\ln(\alpha_i)}{\lambda} \quad i = 1, \dots, N \tag{1.1}$$

где α_i – реализация случайной величины с равномерным распределением на отрезке [0;1], λ – заданный параметр экспоненциального распределения, N – объем выборки.

Числовые характеристики случайных величин. Первичная обработка статистических данных, как правило, начинается с оценки числовых характеристик исследуемой случайной величины.

Оценки числовых характеристик дают информацию о центре рассеяния (оценки математического ожидания, медианы и моды), о диапазоне рассеяния (оценки дисперсии и среднеквадратического отклонения), о характере рассеяния случайной величины (оценки коэффициентов асимметрии и эксцесса).

Значение числовой характеристики случайной величины, рассчитанное по её выборке, называются *выборочным* или *эмпирическим*.

Для вычисления выборочных значений числовых характеристик случайной величины можно использовать инструмент пакета анализа данных MS Excel «Описательная статистика». Для доступа к настройкам данного инструмента необходимо открыть окно выбора инструментов для пакета анализа данных (рисунок 1.4), в списке инструментов выбрать инструмент «Описательная статистика» и нажать кнопку «OK».

Внешний вид окна настроек инструмента «Описательная статистика» приведен на рисунке 1.7. Для получения выборочных значений числовых характеристик случайной величины необходимо указать диапазон ячеек, где расположена сгенерированная выборка (поле «Входной интервал»), левую верхнюю ячейку области, куда будут выведены результаты вычислений (поле «Выходной интервал») и отметить флажок «Итоговая статистика», после чего нажать кнопку «OK». Пример результата, получаемого после применения инструмента «Описательная статистика» приведен на рисунке 1.8. В левой колонке отображены названия рассчитанных выборочных числовых характеристик (жирным шрифтом выделены те, из них, с которыми необходимо работать при выполнении задания), а в правой колонке показаны рассчитанные для них значения.

Взаимосвязь числовых характеристик рассматриваемых законов распределения с их параметрами приведена в таблице 1.1.

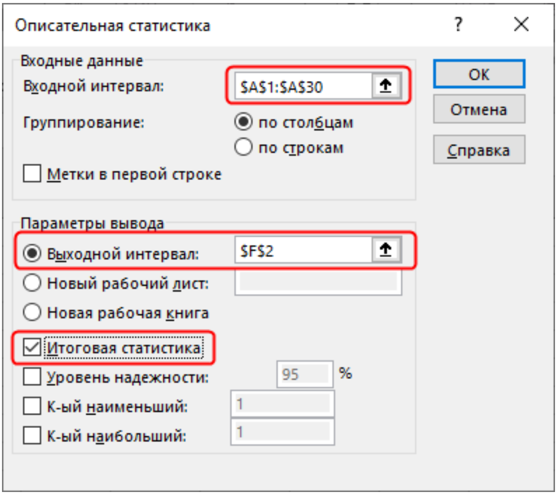

Рис. 1.7. Настройки инструмента «Описательная статистика»

Столбец1	
Среднее	0,01541
Стандартная ошибка	0,03208
Медиана	0,000536
Мода	-0,0589
Стандартное отклонение	1,014463
Дисперсия выборки	1,029136
Эксцесс	0,155272
Асимметричность	0,040931
Интервал	7,324561
Минимум	-4,00934
Максимум	3,315217
Сумма	15,4101
Счет	1000
	0

Рис. 1.8. Результат вычисления выборочных значений числовых характеристик случайной величины с распределением $N(0,1)$ при помощи инструмента «Описательная статистика»

Таблица 1.1

Теоретические значения числовых характеристик для непрерывных случайных величин с нормальным и экспоненциальным распределениями

	Нормальное распределение $N(m,\sigma)$	Экспоненциальное распределение $E(\lambda)$
Математическое ожидание	m	$1/\lambda$
Среднеквадратическое отклонение	σ	$1/\lambda$
Дисперсия	σ^2	$1/\lambda^2$
Медиана	m	$\ln{(2)}/\lambda$
Мода	m	0
Эксцесс	0	6
Асимметричность	0	2

Построение графиков распределений случайной величины. Для построения эмпирического закона распределения по экспериментальной выборке можно использовать инструмент «Гистограмма» пакета анализа данных MS Excel, который позволяет построить диаграмму *частот* (*абсолютных частот*) и эмпирическую *функцию относительной накопленной частоты (эмпирическая функция распределения)*. Для вызова данного инструмента необходимо выбрать его в окне выбора инструментов пакета анализа данных (рисунок 1.4) и нажать кнопку «ОК». Внешний вид окна настроек инструмента «Гистограмма» показан на рисунке 1.9.

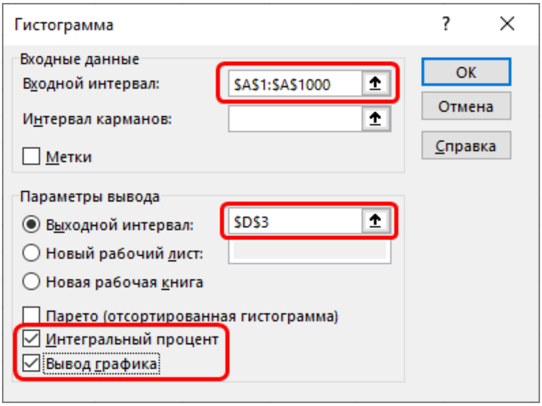

Рис. 1.9. Настройки инструмента «Гистограмма»

Для расчета значений *диаграммы частот* и *эмпирической функции распределения*, а также построения соответствующих им графиков, необходимо указать диапазон ячеек, где расположена рассматриваемая выборка (поле «Входной интервал»), левую верхнюю ячейку области, куда будут выведены результаты вычислений (поле «Выходной интервал»), а также отметить флажки «Интегральный процент» (добавляет вычисление *эмпирической функции распределения*) и «Вывод графика». Для получения результата работы инструмента надо нажать кнопку «ОК». Пример результатов работы инструмента «Гистограмма» приведен на рисунке 1.10.

Карман	*Частота*	*Интегральный %*
-3,84171	1	0,10%
-3,59385	0	0,10%
-3,346	1	0,20%
-3,09815	0	0,20%
-2,8503	0	0,20%
-2,60245	1	0,30%
-2,35459	4	0,70%
-2,10674	8	1,50%
-1,85889	16	3,10%
-1,61104	23	5,40%
-1,36319	28	8,20%
-1,11533	41	12,30%
-0,86748	63	18,60%
-0,61963	55	24,10%
-0,37178	88	32,90%
-0,12393	97	42,60%
0,123926	118	54,40%
0,371778	91	63,50%
0,61963	89	72,40%
0,867482	78	80,20%
1,115334	63	86,50%
1,363186	37	90,20%
1,611038	34	93,60%
1,85889	26	96,20%
2,106742	21	98,30%
2,354594	8	99,10%
2,602446	4	99,50%
2,850298	2	99,70%
3,09815	1	99,80%
3,346002	0	99,80%
3,593854	0	99,80%
Еще	2	100,00%

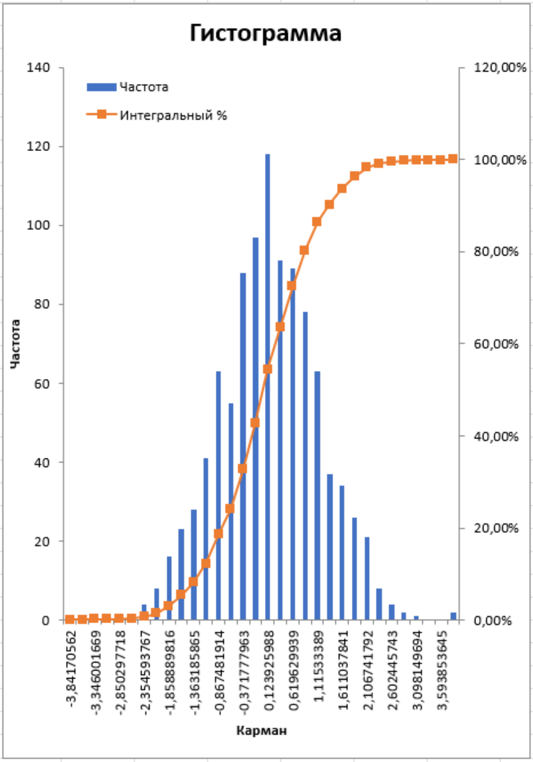

Рис. 1.10. Результат работы инструмента «Гистограмма»

Данные, показанные в примере на рисунке 1.10, в колонке «Карман» содержат правые границы интервалов (*карманов*), для которых вычисляется частота попадания в соответствующий интервал (*карман*). Значения в колонке «Интегральный %» показывают (в процентах) значение относительной накопленной частоты, вычисленное до правой границы соответствующего интервала (*кармана*).

Вычисление *теоретических* значений функции распределения случайных величин, рассматриваемых в лабораторной работе, можно выполнить с использованием следующих встроенных функций MS Excel:

- **НОРМ.РАСП($\boldsymbol{x;m;\sigma;intg}$)** вычисляет значения закона распределения нормальной случайной величины, где x – значение случайной величины, m и σ – значения математического ожидания и средне-квадратического отклонения, соответственно, $intg$ – логическое значение (ИСТИНА для получения значений функции распределения или ЛОЖЬ для вычисления значений плотности распределения);
- **ЭКСП.РАСП($\boldsymbol{x;\lambda;intg}$)** вычисляет значения закона распределения экспоненциальной случайной величины, где x – значение случайной величины, λ – значение параметра экспоненциального распределения, $intg$ – логическое значение (ИСТИНА для получения значений функции распределения или ЛОЖЬ для вычисления значений плотности распределения).

При сравнении значений теоретической функции распределения, рассчитанных при помощи описанных выше функций, с эмпирическими значениями, рассчитанными при помощи инструмента «Гистограмма», в качестве аргумента x необходимо использовать значения из колонки «Карман» (рисунок 1.10).

Построение интервальных оценок числовых характеристик случайных величин. Интервальная оценка (доверительный интервал) для некоторой числовой характеристики случайной величины позволяют по имеющейся выборке определить диапазон, в котором с заранее заданной доверительной вероятностью β лежит истинное (теоретическое) значение данной числовой характеристики.

Доверительная вероятность (доверительный уровень) $\boldsymbol{\beta}$ для вычисления интервальной оценки числовой характеристики выбирается заранее (до вычисления границ интервальной оценки) и показывает степень уверенности, с которой можно утверждать, что истинное значение числовой характеристики будет находиться внутри рассчитанного интервала. Часто при вычислении интервальных оценок вместо величины β используется ***уровень значимости $\boldsymbol{\alpha = 1 - \beta}$***.

Выражение, которое необходимо использовать для вычисления интервальной оценки числовой характеристики случайной величины, может зависеть от:

1. типа рассматриваемой числовой характеристики (математическое ожидание, дисперсия, т.д.);
2. вида распределения, которому подчиняется случайная величина;
3. теоретической информации, которая заранее известна о случайной величине;
4. объема выборки.

Далее будут рассмотрены выражения для вычисления интервальных оценок некоторых числовых характеристик случайных величин с нормальным и экспоненциальным законами распределения.

Интервальные оценки числовых характеристик для нормального закона распределения. Интервальная оценка уровня значимости α для математического ожидания случайной величины с нормальным распределением при известной генеральной дисперсии вычисляется по формуле:

$$m^* - \frac{z_{1-\alpha/2}\sigma}{\sqrt{N}} < m < m^* + \frac{z_{1-\alpha/2}\sigma}{\sqrt{N}} \qquad (1.2)$$

где: m^* − среднее значение по выборке; σ – теоретически известное среднеквадратическое отклонение; N − объем выборки; $z_{1-\alpha/2}$ – квантиль распределения $N(0,1)$ порядка $1-\alpha/2$.

Для вычисления интервальной оценки по формуле (1.2) в MS Excel можно воспользоваться следующими функциями:

- **ДОВЕРИТ.НОРМ(α; σ; N)** вычисляет полуширину доверительного интервала (выражение $\frac{z_{1-\alpha/2}\sigma}{\sqrt{N}}$); α – уровень значимости, σ – теоретически известное среднеквадратическое отклонение, N − объем выборки.
- **НОРМ.СТ.ОБР(p)** возвращает обратное значение интегральной функции распределения нормальной случайной величины $N(0,1)$. Если в качестве аргумента p передать величину $1-\alpha/2$, то функция вернет значение квантиля $z_{1-\alpha/2}$ распределения $N(0,1)$.

Значение m^* можно взять из результатов работы инструмента «Описательная статистика» (значение в строке «Среднее» на рисунке 1.8).

Выражение (1.2) может также использоваться для построения доверительного интервала для математического ожидания и в случае произвольного закона распределения для выборок объемом более 50.

Интервальная оценка уровня значимости α для математического ожидания случайной величины с нормальным распределением при неизвестной генеральной дисперсии вычисляется по формуле:

$$m^* - \frac{t_{1-\alpha/2,N-1}\sigma^*}{\sqrt{N}} < m < m^* + \frac{t_{1-\alpha/2,N-1}\sigma^*}{\sqrt{N}} \qquad (1.3)$$

где: m^* – среднее значение по выборке; σ^* – оценка среднеквадратического отклонения по выборке; N – объем выборки; $t_{1-\alpha/2,N-1}$ – квантиль порядка $1-\alpha/2$ для распределения Стьюдента с $N-1$ степенью свободы.

Для вычисления интервальной оценки по формуле (1.3) в MS Excel можно воспользоваться следующими функциями:

- **ДОВЕРИТ.СТЬЮДЕНТ(α; σ^*; N)** вычисляет полуширину доверительного интервала (выражение $\frac{t_{1-\alpha/2,N-1}\sigma^*}{\sqrt{N}}$); α – уровень значимости, σ^* – оценка среднеквадратического отклонения по выборке, N – объем выборки.
- **СТЬЮДЕНТ.ОБР(*p*, *n*)** возвращает обратное значение интегральной функции распределения для распределения Стьюдента с n степенями свободы. Если в качестве аргументов p и n передать, соответственно, величины $1-\alpha/2$ и $N-1$, то функция вернет значение квантиля $t_{1-\alpha/2,N-1}$ распределения Стьюдента.

Значения m^* и σ^* можно взять из результатов работы инструмента «Описательная статистика» (значения в строках «Среднее» и «Стандартное отклонение» на рисунке 1.8).

Интервальная оценка уровня значимости α для дисперсии случайной величины с нормальным распределением вычисляется по формуле:

$$\frac{(N-1)D^*}{\chi^2_{1-\alpha/2,N-1}} < D < \frac{(N-1)D^*}{\chi^2_{\alpha/2,N-1}} \qquad (1.4)$$

где: D^* – оценка дисперсии по выборке; N – объем выборки; $\chi^2_{1-\alpha/2,N-1}$ и $\chi^2_{\alpha/2,N-1}$ – квантили распределения χ^2 с $N-1$ степенью свободы уровня $1-\alpha/2$ и $\alpha/2$, соответственно.

Для вычисления интервальной оценки по формуле (1.4) в MS Excel можно воспользоваться функцией **ХИ2.ОБР(*p*, *n*)**, которая возвращает обратное значение функции распределения χ^2 с n степенями свободы. Если в качестве аргументов p и n передать, соответственно, порядок квантиля (например, величину $1-\alpha/2$ или $\alpha/2$) и $N-1$, то функция вернет значение квантиля (например, $\chi^2_{1-\alpha/2,N-1}$ или $\chi^2_{\alpha/2,N-1}$) распределения χ^2 с $N-1$ степенью свободы.

Значение D^* можно взять из результатов работы инструмента «Описательная статистика» (значение в строке «Дисперсия выборки» на рисунке 1.8).

Интервальные оценки числовых характеристик для экспоненциального распределения. Интервальная оценка для параметра λ экспоненциального распределения уровня значимости α вычисляется по формуле:

$$\frac{\chi^2_{\alpha/2,2N}}{2Nm^*} < \lambda < \frac{\chi^2_{1-\alpha/2,2N}}{2Nm^*} \quad (1.5)$$

где: m^* – среднее значение по выборке; N – объем выборки; $\chi^2_{\alpha/2,2N}$ и $\chi^2_{1-\alpha/2,2N}$ – квантили распределения χ^2 с $2N$ степенями свободы порядков $\alpha/2$ и $1-\alpha/2$, соответственно.

Для вычисления интервальной оценки по формуле (1.5) в MS Excel можно воспользоваться функцией **ХИ2.ОБР**, описанной ранее. Значение m^* можно взять из результатов работы инструмента «Описательная статистика» (значение в строке «Среднее» на рисунке 1.8).

Интервальная оценка уровня значимости α для математического ожидания экспоненциально распределенной случайной величины с параметром λ вычисляется по формуле:

$$^1/_{\lambda_{right}} < m < {^1/_{\lambda_{left}}} \quad (1.6)$$

где λ_{left} и λ_{right} – левая и правая границы доверительного интервала для параметра λ, рассчитанные по формуле (1.5).

Порядок выполнения лабораторной работы № 1.

1. Используя инструмент «Генерация случайных чисел» пакета анализа данных в соответствии с вариантом задания (смотри таблицу 1.2) сформировать случайные экспериментальные данные из пяти выборок различного объема: X_1, X_2, X_3, X_4, X_5 для нормального закона распределения $N(m, \sigma)$ и Y_1, Y_2, Y_3, Y_4, Y_5 для экспоненциального закона распределения $E(\lambda)$. Объемы выборок в соответствии с их индексами: $N_1 = 20 \div 30$, $N_2 = 50 \div 70$, $N_3 = 150 \div 170$, $N_4 = 250 \div 300$, $N_5 = 1000$.

2. Используя инструмент «Описательная статистика» пакета анализа данных, для каждой выборки рассчитать выборочные числовые характеристики (математическое ожидание, медиана, мода, среднеквадратичное отклонение, дисперсия, коэффициент эксцесса, коэффициент асимметрии); проанализировать сходимость выборочных числовых характеристик к их теоретическим значениям.

3. Используя инструмент «Гистограмма» пакета анализа данных, построить эмпирический закон распределения (диаграмма частот, функция накопленной частоты) для выборок X_4, X_5, Y_4, Y_5. Для этих же интервалов (карманов) вычислить и добавить на графики теоретическую функцию распределения, сопоставить ее с эмпирической кривой.

4. Для выборок X_1, X_2, X_3 рассчитать интервальные оценки для математического ожидания (рассмотреть случай известной и неизвестной дисперсии) и дисперсии при уровне значимости *1%, 5%, 10%.* Для выборок Y_1, Y_2, Y_3 рассчитать интервальные оценки для параметра λ экспоненциального закона распределения и математического ожидания при уровне значимости *1%, 5%, 10%.* Проанализировать динамику изменения ширины доверительного интервала в зависимости от размера выборки и уровня значимости.

Форма отчета: электронная книга таблицы MS Excel с полученными результатами.

Таблица 1.2

Параметры для законов распределения исследуемых случайных величин

Номер варианта	Случайная величина X нормально распределенная		Случайная величина Y, распределенная по экспоненциальному закону
	m	σ	λ
1	20	2	0,0500
2	30	4	0,0333
3	40	5	0,0250
4	50	6	0,0200
5	60	7	0,0167
6	70	8	0,0143
7	80	10	0,0125
8	90	11	0,0111
9	100	12	0,0100
10	110	14	0,0091
11	120	15	0,0083
12	130	16	0,0077
13	140	17	0,0071
14	150	18	0,0067
15	160	20	0,0063
16	170	21	0,0059

Окончание таблицы 1.2

Номер варианта	Случайная величина X нормально распределенная		Случайная величина Y, распределенная по экспоненциальному закону
	m	σ	λ
17	180	22	0,0056
18	190	23	0,0053
19	200	24	0,0050
20	210	25	0,0048
21	220	26	0,0045
22	230	27	0,0043
23	240	28	0,0042
24	250	29	0,0040
25	260	30	0,0038
26	270	31	0,0037
27	280	32	0,0036
28	290	33	0,0034
29	300	34	0,0033
30	310	35	0,0032

ТЕМА 2. ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

Цель работы. Изучение методов проверки статистических гипотез о числовых характеристиках; критериев проверки статистической однородности наборов случайных данных, подчиняющихся нормальному, либо экспоненциальному закону распределения.

Краткие сведения. Два набора случайных значений могут считаться однородными, если:

1. они подчиняются одному и тому же закону распределения;
2. значения параметров закона распределения, которому подчиняются обе выборки, равны.

Проверка выборки на соответствие нормальному закону распределения (малые выборки). Для проверки выборки случайных значений на соответствие нормальному закону распределения, в случае малых выборок, можно использовать статистический критерий, основанный на вычислении оценок коэффициентов асимметрии и эксцесса.

Для осуществления вышеуказанной проверки необходимо выполнить следующие действия:

1. вычислить для тестируемой выборки оценки коэффициентов асимметрии S^* и эксцесса E^*, например, с использованием инструмента «Описательная статистика» пакета анализа данных MS Excel;
2. вычислить для тестируемой выборки дисперсии оценок коэффициентов асимметрии D_S^* и эксцесса D_E^* по следующим формулам:

$$D_S^* = \frac{6N(N-1)}{(N-2)(N+1)(N+3)}$$
$$D_E^* = \frac{24N(N-1)^2}{(N-3)(N-2)(N+3)(N+5)} \qquad (2.1)$$

где N – объем выборки.

3. считается, что выборка близка к нормальному закону распределения, если выполняются следующие неравенства:

$$|S^*| \le 3 * \sqrt{D_S^*}$$
$$|E^*| \le 5 * \sqrt{D_E^*}. \qquad (2.2)$$

Проверка равенства параметров нормального закона распределения для тестируемых выборок. Предположение о равенстве для двух нормальных выборок математического ожидания и дисперсии можно сделать рассчитав для них доверительные интервалы с использованием выражений (1.3) и (1.4), соответственно. Если доверительные интервалы для математического ожидания двух выборок перекрываются незначительно (менее 10%), то, скорее всего, математические ожидания для этих выборок различны. Иначе, можно сделать предположение о равенстве математических ожиданий. Аналогично делается предположение о равенстве дисперсий.

Для проверки равенства дисперсий двух независимых выборок, подчиняющихся нормальному закону распределения, используется критерий Фишера. При использовании данного критерия проверяется гипотеза:

$$H_0: D_1 = D_2,$$

при альтернативной гипотезе:

$$H_1: D_1 \neq D_2,$$

где: D_1 и D_2, дисперсии первой и второй выборок, соответственно.

В данном случае, в качестве статистического критерия выступает отношение двух выборочных дисперсий:

$$F = \frac{D_1^*}{D_2^*}.$$

Гипотеза о равенстве дисперсий принимается, если расчетное значение F будет меньше критического для правостороннего критерия и больше критического для левостороннего. В качестве критического значения выбирается α-процентная точка распределения Фишера $F_{N_1-1,N_2-1,\alpha}$ с числом степеней свободы $N_1 - 1, N_2 - 1$.

Для вычисления расчетного значения F и соответствующего ему критического значения используется инструмент «Двухвыборочный F-тест для дисперсии» из пакета анализа данных MS Excel (рисунок 2.1).

Для настройки параметров указанного выше инструмента, необходимо указать диапазоны, где расположены значения для тестируемых выборок (поля «Интервал переменной 1» и «Интервал переменной 2»), уровень значимости α (поле «Альфа») и левую верхнюю ячейку, начиная с которой будут выведены результаты работы инструмента (поле «Выходной интервал»). Пример окна с выполненными настройками для инструмента «Двухвыборочный F-тест для дисперсии» приведен на рисунке 2.2.

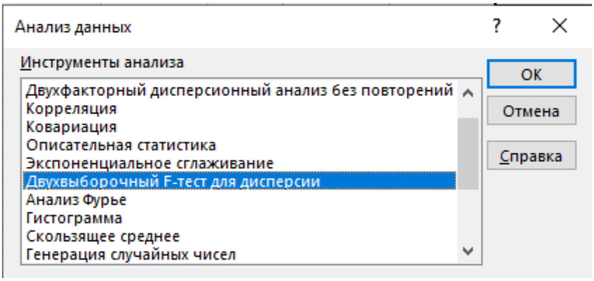

Рис. 2.1. Выбор инструмента «Двухвыборочный F-тест для дисперсии» из пакета для анализа данных MS Excel

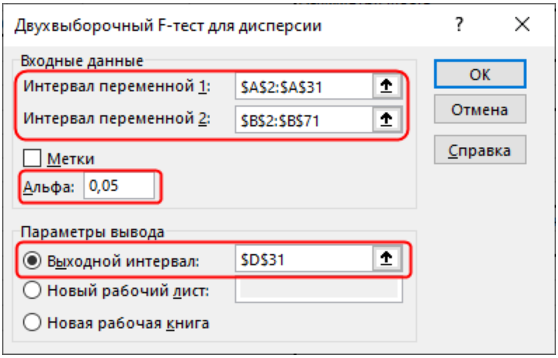

Рис. 2.2. Настройки инструмента «Двухвыборочный F-тест для дисперсии»

Пример результатов, полученных при помощи инструмента «Двухвыборочный F-тест для дисперсии» показан на рисунке 2.3.

Двухвыборочный F-тест для дисперсии

	Переменная 1	*Переменная 2*
Среднее	20,71202324	20,17854901
Дисперсия	14,5180813	17,16606338
Наблюдения	30	70
df	29	69
F	0,845743195	
P(F<=f) одностор	0,314405309	
F критическое о	0,574844501	

Рис. 2.3. Результаты работы инструмента «Двухвыборочный F-тест для дисперсии»

Для проверки гипотезы о равенстве дисперсий двух выборок из результатов работы инструмента «Двухвыборочный F-тест для дисперсии» необходимо рассмотреть значения «F» и «F критическое одностороннее». Если значение *«F»<1* и *«F»>«F критическое одностороннее»*, или *«F»>1* и *«F»<«F критическое одностороннее»* с вероятностью $1-\alpha$, то гипотеза о равенстве дисперсий принимается. При $F=1$ дисперсии абсолютно одинаковы, и проверка не требуется.

Для проверки гипотезы о равенстве математических ожиданий двух независимых выборок используется критерий Стьюдента. При использовании данного критерия проверяется гипотеза:

$$H_0\colon m_1 = m_2,$$

при альтернативной гипотезе

$$H_1\colon m_1 \neq m_2.$$

Рассчитываемый статистический критерий зависит от результата проведенной на предыдущем этапе проверки гипотезы о равенстве дисперсий.

При равенстве дисперсий в качестве статистического критерия принимается случайная величина:

$$T = \frac{m_1^* - m_2^*}{\sqrt{\widetilde{D}\left(\frac{1}{N_1} + \frac{1}{N_2}\right)}}$$

Имеющая, в случае справедливости нулевой гипотезы, распределение Стьюдента с числом степеней свободы $k = N_1 + N_2 - 2$.

Здесь величина $\widetilde{D}$ - "объединенная" дисперсия, вычисляемая следующим образом:

$$\widetilde{D} = \frac{\sum_{i=1}^{N_1}\left(x_{1,i}-m_1^*\right)^2+\sum_{i=1}^{N_2}\left(x_{2,i}-m_2^*\right)^2}{(N_1-1)+(N_2-1)} = \frac{(N_1-1)D_1^*+(N_2-1)D_2^*}{N_1+N_2-2}.$$

В рассматриваемом случае (альтернативная гипотеза в виде неравенства) для проверки нулевой гипотезы применяется двусторонняя критическая область, где критические точки находятся из условия:

$$|T| > t_{k,\alpha/2}$$

где $t_{k,\alpha/2}$ – $(\alpha/2)$-процентная точка распределения Стьюдента с k степенями свободы.

В пакете анализа данных MS Excel проверка гипотезы по критерию Стьюдента при условии, что для тестируемых выборок дисперсии неизвестны, но считаются равными (например, на основании результатов F-теста), используется инструмент «Двухвыборочный t-тест с одинаковыми дисперсиями» (рисунок 2.4).

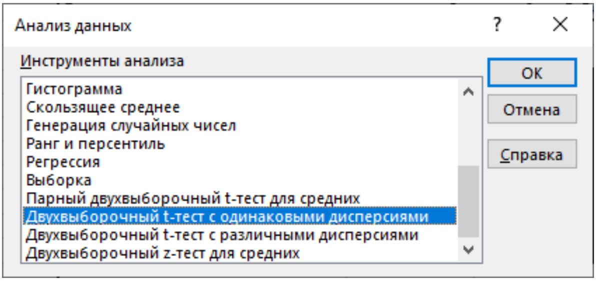

Рис. 2.4. Выбор инструмента «Двухвыборочный t-тест с одинаковыми дисперсиями» из пакета для анализа данных MS Excel

Диалоговое окно для настроек данного инструмента показано на рисунке 2.5.

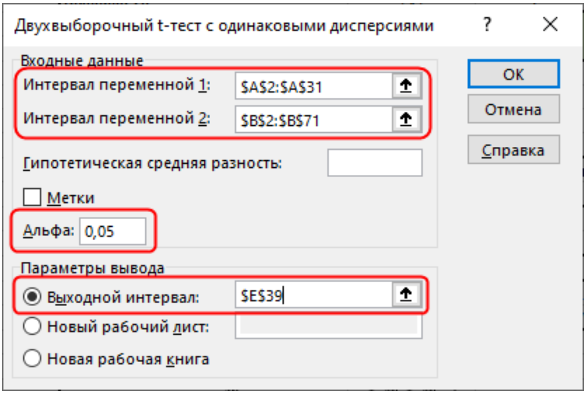

Рис. 2.5. Настройки инструмента «Двухвыборочный t-тест с одинаковыми дисперсиями»

Названия полей для выполнения необходимых настроек t-теста и их назначение аналогичны соответствующим полям настроек инструмента «Двухвыборочный F-тест для дисперсии», описанным выше.

Пример результата, полученного после выполнения t-теста, показан на рисунке 2.6. Для того, чтобы сделать вывод на основании результатов работы инструмента «Двухвыборочный t-тест с одинаковыми дисперсиями», необходимо сравнить значения «t-статистика» и «t критическое двухстороннее». Если значение «t-статистика» меньше значения «t критическое двухстороннее» с вероятностью $1 - \alpha$, то гипотеза о равенстве математических ожиданий двух выборок принимается.

Двухвыборочный t-тест с одинаковыми дисперсиями		
	Переменная 1	*Переменная 2*
Среднее	20,71202324	20,17854901
Дисперсия	14,5180813	17,16606338
Наблюдения	30	70
Объединенная дисперсия	16,38247685	
Гипотетическая разность	0	
df	98	
t-статистика	0,603994955	
P(T<=t) одностороннее	0,273621324	
t критическое односторо	1,660551217	
P(T<=t) двухстороннее	0,547242649	
t критическое двухсторон	1,984467455	

Рис. 2.6. Результаты работы инструмента «Двухвыборочный t-тест с одинаковыми дисперсиями»

Проверка выборки на соответствие экспоненциальному закону распределения. Для проверки выборки случайных значений на соответствие экспоненциальному закону распределения рекомендуется использовать критерий согласия Бартлетта для малой выборки.

Для выполнения вышеуказанной проверки необходимо выполнить следующие действия:

1. вычислить экспериментальное значение для критерия Бартлетта по формуле:

$$B_r = \frac{2N[ln\left(\frac{t_N}{N}\right) - \frac{1}{N}(\sum_{i=1}^{N} ln\, x_i)]}{1+(N+1)/6N}, \qquad (2.3)$$

где N – объем выборки, $t_N = \sum_{i=1}^{N}(x_i)$, x_i – i-й элемент выборки;

2. при допущении об экспоненциальном распределении величина B_r имеет распределение χ^2 с $N-1$ степенью свободы. Соответственно, гипотеза об экспоненциальном законе распределения принимается, если с вероятностью $1-\alpha$ выполняется условие:

$$\chi^2_{\alpha/2,N-1} < B_r < \chi^2_{1-\alpha/2,N-1}, \qquad (2.4)$$

где N – объем выборки, α – уровень значимости, $\chi^2_{\alpha/2,N-1}$ и $\chi^2_{1-\alpha/2,N-1}$ – квантили уровня $\alpha/2$ и $1-\alpha/2$ для распределения χ^2 с $N-1$ степенью свободы.

Проверка равенства параметров экспоненциального закона распределения для тестируемых выборок. Если для двух выборок случайных значений было установлено, что они подчиняются экспоненциальному закону распределения, то их можно будет считать однородными, если параметр экспоненциального распределения λ для обеих выборок может считаться одинаковым. Поскольку параметр $\lambda = 1/m$, где m – математическое ожидание экспоненциально распределенной случайной величины, то проверка равенства параметров λ для распределений двух выборок может быть сведена к проверке равенства их математических ожиданий.

Предположение о равенстве для двух выборок, подчиняющихся экспоненциальному распределению, их параметров λ можно сделать рассчитав для этих величин доверительные интервалы с использованием выражения (1.5). Если соответствующие доверительные интервалы двух выборок перекрываются незначительно, то, скорее всего, величины, для которых они были построены, различны. В противном случае можно сделать предположение о равенстве этих величин и следовательно, об однородности двух выборок, подчиняющихся экспоненциальному закону распределения.

Если объем выборок достаточно велик (более 50 значений), для проверки гипотезы о равенстве математических ожиданий двух выборок, подчиняющихся произвольному закону распределения, можно использовать двухвыборочный z-критерий для средних значений. Вышеуказанное утверждение справедливо, поскольку закон распределения выборочного среднего при больших объемах выборки асимптотически сходится к нормальному закону.

При использовании данного критерия проверяется гипотеза:

$$H_0: m_1 = m_2,$$

при альтернативной гипотезе

$$H_1: m_1 \neq m_2.$$

Расчетное значение критерия вычисляется следующим образом:

$$Z = \frac{m_1^* - m_2^*}{\sqrt{\frac{D_1}{N_1} + \frac{D_2}{N_2}}}$$

где в качестве D_1 и D_2 используются известные теоретические значения дисперсий. В то же время, при больших объемах выборок закон распределения выборочной дисперсии стремится к нормальному закону. При этом погрешность вычисления выборочной дисперсии с ростом объема выборки будет уменьшаться, а ее среднее значение будет равно теоретической дисперсии исходной случайной величины. На этом сновании, при вычислении z-критерия для больших выборок, для которых генеральная дисперсия не известна, вместо теоретических значений дисперсии можно использовать их выборочные значения, т.е.: $D_1 \approx D_1^*$; $D_2 \approx D_2^*$.

Критические точки для проверяемой гипотезы находятся из условия:

$$|Z| > z_{\alpha/2},$$

где $z_{\alpha/2}$ – $(\alpha/2)$-процентная точка стандартного нормального распределения.

Данный критерий реализуется в MS Excel инструментом «Двухвыборочный z-тест для средних» (рисунок 2.7).

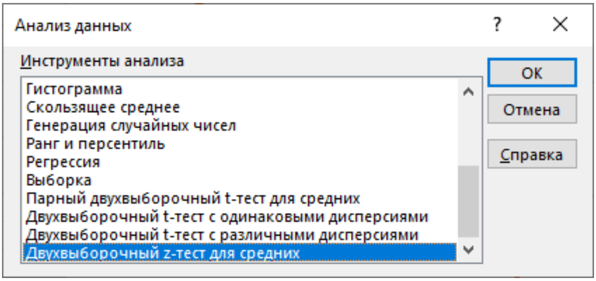

Рис. 2.7. Выбор инструмента «Двухвыборочный z-тест для средних» из пакета анализа данных MS Excel

Диалоговое окно для настроек инструмента «Двухвыборочный z-тест для средних» показано на рисунке 2.8.

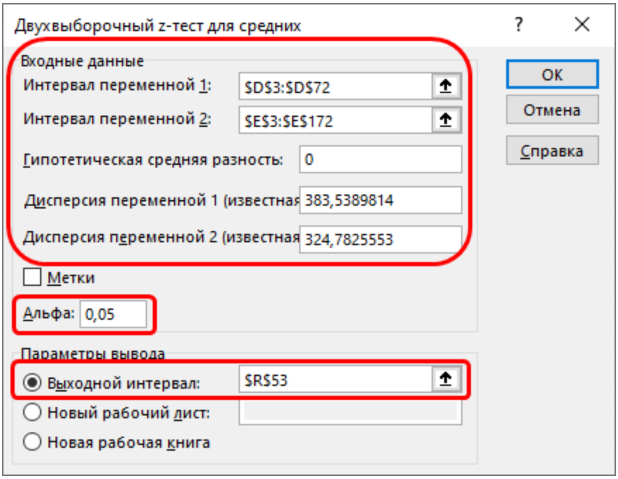

Рис. 2.8. Настройки инструмента «Двухвыборочный z-тест для средних»

Как видно из рисунка 2.8 перед выполнением z-теста необходимо указать диапазоны ячеек, где расположены тестируемые выборки (поля «Интервал переменной 1» и «Интервал переменной 2»), выборочные дисперсии (поля «Дисперсия переменной 1» и «Дисперсия переменной 2») (вычисляются по тестируемым выборкам), уровень значимости α (поле «Альфа») и левую верхнюю ячейку, начиная с которой будут выведены результаты работы инструмента (поле «Выходной интервал»).

Пример результата, полученного после выполнения z-теста, показан на рисунке 2.9.

Двухвыборочный z-тест для средних		
	Переменная	*Переменная 2*
Среднее	20,3229269	19,16034315
Известная дисперсия	383,538981	324,7825553
Наблюдения	70	170
Гипотетическая разность средних	0	
z	0,42767452	
P(Z<=z) одностороннее	0,33444405	
z критическое одностороннее	1,64485363	
P(Z<=z) двухстороннее	0,6688881	
z критическое двухстороннее	1,95996398	

Рис. 2.9. Результаты работы инструмента «Двухвыборочный z-тест для средних»

Для того, чтобы на основании результатов работы инструмента «Двухвыборочный z-тест для средних», сделать вывод о том, могут ли математические ожидания выборок считаться одинаковыми, необходимо сравнить значения «z» и «z критическое двухстороннее». Если значение «z» меньше значения «z критическое двухстороннее» с вероятностью $1-\alpha$, то гипотеза о равенстве математических ожиданий двух выборок принимается.

Порядок выполнения лабораторной работы № 2.

1. Проверить статистическую однородность выборок X_1 и X_2, сформированных в лабораторной работе № 1. Для этого:

- проверить обе выборки на соответствие нормальному закону распределения, используя оценки коэффициентов асимметрии и эксцесса;
- выдвинуть предположение о статистической однородности рассматриваемых выборок, сопоставляя интервальные оценки для математического ожидания и дисперсии;
- сравнить дисперсии D_1 и D_2 для выборок, используя F-тест при уровнях значимости α = 1%, 5% и 10% для проверки гипотезы $\boldsymbol{H_0}: \boldsymbol{D_1} = \boldsymbol{D_2}$ при альтернативной гипотезе $\boldsymbol{H_1}: \boldsymbol{D_1} \neq \boldsymbol{D_2}$;

- сравнить математические ожидания m_1 и m_2 для выборок, используя t-тест при уровнях значимости α = 1%, 5% и 10% для проверки гипотезы $\boldsymbol{H_0}: \boldsymbol{m_1 = m_2}$ при альтернативной гипотезе $\boldsymbol{H_1}: \boldsymbol{m_1 \neq m_2}$.

2. Проверить статистическую однородность выборок Y_2 и Y_3, сформированных в лабораторной работе № 1. Для этого:

- проверить обе выборки на соответствие экспоненциальному закону распределения, используя критерий Бартлетта при уровне значимости 5%;
- выдвинуть предположение о статистической однородности рассматриваемых выборок, сопоставляя интервальные оценки для параметра λ экспоненциального закона распределения;
- сравнить математические ожидания m_2 и m_3 для выборок, используя z-тест при уровне значимости 5% для проверки гипотезы $\boldsymbol{H_0}: \boldsymbol{m_2 = m_3}$ при альтернативной гипотезе $\boldsymbol{H_1}: \boldsymbol{m_2 \neq m_3}$.

Форма отчета. Электронные таблицы MS Excel, реализующие задание.

ТЕМА 3. КРИТЕРИИ СОГЛАСИЯ

Цель работы. Получить практические навыки использования критериев Пирсона и Колмогорова-Смирнова для проверки статистических гипотез о виде закона распределения

Краткие сведения. Для проверки соответствия выборки случайных значений некоторому закону распределения могут использоваться критерии согласия. Среди часто используемых на практике критериев согласия можно выделить критерий Колмогорова-Смирнова и критерий Пирсона (хи-квадрат).

Критерий Колмогорова-Смирнова. Проверка по критерию Колмогорова-Смирнова основана на сравнении с некоторой критической величиной максимального абсолютного отклонения значений эмпирической и теоретической интегральных функций распределения, которые строятся на основе тестируемой выборки и некоторого закона распределения непрерывной случайной величины, соответственно.

В случае если известны все параметры рассматриваемого закона распределения, преимуществом данного критерия является его точная достоверность для любого объема выборки.

При использовании критерия Колмогорова-Смирнова исходная ***выборка случайных значений объемом N в обязательном порядке должна быть преобразована в вариационный ряд*** $\{x_i,\ i = 1, 2, \dots, N\}$ путем сортировки ее значений по возрастанию.

После получения вариационного ряда, в каждой его точке вычисляют значения теоретической интегральной функции $F(x_i)$ для выбранного закона распределения и ступенчатой эмпирической функции по формуле:

$$F_N(x_i) = \frac{i}{N},\ i = 1,2, \dots N, \tag{3.1}$$

где x_i – i-е значение вариационного ряда, N – объем выборки.

На основе $F(x_i)$ и $F_N(x_i)$, $i = 1,2, \dots N$ вычисляют следующие две вспомогательные величины:

$$D_{\mathrm{N}}^{+} = \max_{1 \le i \le N}\{F_N(x_i) - F(x_i)\}, \quad D_n^{-} = \max_{1 \le i \le N}\{F(x_i) - F_N(x_{i-1})\}, \tag{3.2}$$

где $F_N(x_0) \equiv 0$.

Статистикой распределения Колмогорова-Смирнова D_N будет являться наибольшее (вертикальное) расстояние между F и F_N (рисунок 3.1):

$$D_N = max\{D_N^{+}, D_N^{-}\}. \tag{3.3}$$

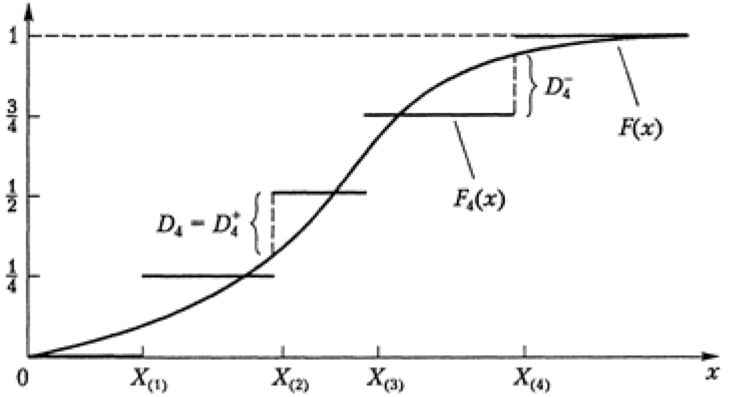

Рис. 3.1. Графическое представление вычисления статистики Колмогорова-Смирнова D_N при $N = 4$.

Большое значение D_N указывает на неудовлетворительное согласие, и гипотеза о соответствии выборки предполагаемому закону распределения случайной величины отвергается.

Для проверки величины D_N можно использовать приближение Стефенса, в соответствии с которым строится неравенство, аналитический вид которого зависит от того, оценивались ли параметры используемого теоретического закона распределения по исследуемой выборке.

Если используемая в тесте теоретическая интегральная функция распределения F непрерывна и ни один из ее параметров не оценивался по тестируемой выборке, то распределение величины D_N, не зависит от вида F и, согласно приближению Стефенса, проверочное неравенство будет иметь вид:

$$\left(\sqrt{N} + 0.12 + \frac{0.11}{\sqrt{N}}\right) D_N \leq c_{1-\alpha}, \tag{3.4}$$

где критическое значение $c_{1-\alpha}$ зависит только от α. Значения $c_{1-\alpha}$, предназначенные для использования в неравенстве (3.4), для различных значений α приведены в таблице 3.1.

Таблица 3.1

Критические значения для проверки по критерию Колмогорова-Смирнова в соответствии с приближением Стефенса (3.4)

α	0.1	0.05	0.025	0.01
$c_{1-\alpha}$	1.224	1.358	1.480	1.626

В случае, когда выборка тестируется на соответствие нормальному закону распределения $N(m,\sigma)$ и при этом для вычисления теоретической интегральной функции F в качестве значений параметров m и σ используются их оценки, рассчитанные по тестируемой выборке, то для проверки, согласно приближению Стефенса, применяется неравенство вида:

$$\left(\sqrt{N}\ -\ 0.01+\frac{0.85}{\sqrt{N}}\right)D_N\le c_{1-\alpha}, \tag{3.5}$$

Критические значения $c_{1-\alpha}$ для использования в выражении (3.5) представлены в таблице 3.2.

Таблица 3.2

Критические значения для проверки по критерию Колмогорова-Смирнова в соответствии с приближением Стефенса (3.5)

α	0.1	0.05	0.025	0.01
$c_{1-\alpha}$	0.819	0.895	0.955	1.035

При тестировании выборки на соответствие экспоненциальному закону распределения $E(\lambda)$, при условии, что для вычисления его теоретической интегральной функции F в качестве значения параметра λ использовалась его оценка по тестируемой выборке, проверочное неравенство, соответствующее приближению Стефенса, будет иметь вид:

$$\left(\sqrt{N}\ +0.26+\frac{0.5}{\sqrt{N}}\right)\left(D_N\ -\ \frac{0.2}{N}\right)\le c_{1-\alpha}, \tag{3.6}$$

Критические значения $c_{1-\alpha}$ для формулы (3.6) приведены в таблице 3.3.

Таблица 3.3

Критические значения для проверки по критерию Колмогорова-Смирнова в соответствии с приближением Стефенса (3.6)

α	0.1	0.05	0.025	0.01
$c_{1-\alpha}$	0. 990	1.094	1.190	1.308

Если неравенство, выбранное для проверки по критерию Колмогорова-Смирнова среди выражений (3.4) – (3.6), выполняется с вероятностью $1-\alpha$, то гипотеза о соответствии выборки предполагаемому закону распределения принимается.

При реализации проверки по критерию Колмогорова-Смирнова в MS Excel для вычисления теоретических функций распределения случайных величин с нормальным и экспоненциальным распределениями можно использовать функции **НОРМ.РАСП** и **ЭКСП.РАСП**, соответственно.

Критерий Пирсона (хи-квадрат). Проверка по критерию согласия Пирсона выполняется путем сопоставления эмпирических и теоретических частот попадания значений случайной величины в некоторые интервалы из ее области определения, на основании которых строится статистика вида:

$$\chi^2 = \sum_{i=1}^{k} \frac{(\nu_i - Np_i)^2}{Np_i}, \tag{3.7}$$

где: χ^2 – расчётное значение критерия Пирсона, ν_i – количество значений из выборки, попавших в i-й интервал (абсолютная частота), p_i – теоретическая вероятность попадания в i-й интервал, k – количество интервалов разбиения, N – объем экспериментальной выборки.

При использовании критерия Пирсона объем тестируемой выборки должен быть достаточно большим (N > 250).

Набор непересекающихся интервалов, используемый при вычислении величины χ^2, должен удовлетворять следующим требованиям:

1. набор интервалов должен полностью покрывать область определения случайной величины, на соответствие которой тестируется выборка;
2. теоретически ожидаемое число попаданий в каждый интервал (величина Np_i) не должно быть слишком мало (на практике должно быть не меньше 5).

Гипотеза о соответствии по критерию Пирсона тестируемой выборки выбранному закону распределения принимается, если с вероятностью $1 - \alpha$ выполняется неравенство:

$$\chi^2 \le \chi^2_{r,\alpha}, \tag{3.8}$$

где α – заранее выбранный уровень значимости критерия, $r = k - 1$ – число степеней свободы, $\chi^2_{r,\alpha}$ правая критическая точка уровня доверительной вероятности $1 - \alpha$ для распределения хи-квадрат с числом степеней свободы r.

Для вычисления величин ν_i в формуле (3.7) можно воспользоваться инструментом «Гистограмма» пакета анализа данных MS Excel. В отличии от первой лабораторной работы, при настройке данного инструмента нет необходимости отмечать флажок «Интегральный процент». После использования инструмента «Гистограмма» будет получено две колонки значений: колонка «Карман» со значениями правых границ интервалов разбиения, использованных для вычисления гистограммы, и колонка «Частота» со значениями ν_i (смотри рисунок 3.2, фрагмент таблицы слева). При этом начало первого интервала (первого кармана) будет соответствовать началу области определения случайной величины, на соответствие которой проверяется выборка, а правая граница последнего интервала (со словом «Еще» в колонке «Карман») правой границе вышеуказанной области.

Карман	*Частота*	*Np*
8,62	1	0,67
10,13	1	1,38
11,64	1	3,45
13,15	3	7,51
14,65	17	14,22
16,16	19	23,39
17,67	34	33,43
19,18	42	41,53
20,69	50	44,84
22,19	42	42,07
23,70	42	34,30
25,21	20	24,30
26,72	14	14,96
28,22	3	8,01
29,73	6	3,72
31,24	3	1,50
32,75	1	0,53
Еще	1	0,22

Частота	*Np*
3	5,49
3	7,51
17	14,22
19	23,39
34	33,43
42	41,53
50	44,84
42	42,07
42	34,30
20	24,30
14	14,96
3	8,01
11	5,97

Рис. 3.2. Порядок объединения интервалов с малыми значениями теоретически ожидаемых частот Np перед вычислением статистики критерия Пирсона.

Вероятности p_i в выражении (3.7) вычисляются как разность теоретических функций распределения на краях интервала:

$$p_i = F_i - F_{i-1}, \qquad (3.9)$$

где F_i значение теоретической функции распределения, вычисляемое на правой границе i –го интервала (кармана) при помощи **НОРМ.РАСП** для нормального и **ЭКСП.РАСП** для экспоненциального законов. При этом необходимо иметь ввиду, что $F_0 \equiv 0$, а на правой границе области определения случайной величины $F \equiv 1$.

После вычисления значений ν_i и Np_i на основе интервалов, полученных из инструмента «Гистограмма», для соблюдения описанных выше требований к теоретическим частотам следует, при необходимости, выполнить объединение соседних интервалов. Для этого значения в колонках ν_i и Np_i для тех строк, где $Np_i < 5$ суммируются с соответствующими значениями соседних строк так, чтобы в результирующем наборе все величины в колонке Np_i были больше 5. Пример выполнения процедуры объединения интервалов приведен на рисунке 3.2.

Использование экспоненциального распределения для оценки времени безотказной работы. Экспоненциальный закон распределения играет большую роль в теории надежности, где он описывает время работы до отказа в период нормального режима эксплуатации, когда отсутствуют ранние отказы и отказы, вызванные деградацией структурных элементов.

В этом случае надежность характеризуется вероятностью безотказной работы $P(t)$ и гамма-процентной наработкой t_γ, т.е. временем наработки, в течении которого вероятность безотказной работы системы, либо структурных элементов не становится ниже уровня 90, 95, 99%.

Вероятность безотказной работы рассчитывается как:

$$P(t) = 1 - F(t),$$

а гамма-процентная наработка t_γ определяется, как решение уравнения

$$P(t_\gamma) = \gamma,$$

где γ – уровень вероятности безотказной работы.

Порядок выполнения лабораторной работы № 3

1. Используя критерии согласия проверить статистическую гипотезу о нормальном законе распределения, для следующих выборок, сформированных в лабораторной работе № 1:
   - для выборки X_3, используя критерий Колмогорова-Смирнова;
   - для выборки X_4, используя критерий Пирсона.

   В ходе выполнения указанных выше проверок построить:
   - на одном графике функцию накопленных частот (эмпирическая функция распределения) и теоретическую функцию распределения при проверке по критерию Колмогорова-Смирнова;
   - на одном графике диаграмму абсолютных частот, рассчитанную по выборке, и теоретически ожидаемое число попаданий в интервалы (теоретические частоты) при проверке по критерию Пирсона.
2. Используя критерии согласия проверить статистическую гипотезу о экспоненциальном законе распределения, для следующих выборок, сформированных в лабораторной работе № 1:
   - для выборки Y_3, используя критерий Колмогорова-Смирнова;
   - для выборки Y_4, используя критерий Пирсона.

   В ходе выполнения указанных выше проверок построить:

- на одном графике функцию накопленных частот (экспериментальная функция распределения) и теоретическую функцию распределения при проверке по критерию Колмогорова-Смирнова;
- на одном графике диаграмму абсолютных частот, рассчитанную по выборке, и теоретически ожидаемое число попаданий в интервалы (теоретические частоты) при проверке по критерию Пирсона.

3. В предположении, что величина Y, носит характер времени работы до отказа, по экспериментальным данным выборок Y_3 и Y_4 построить графики вероятности безотказной работы и оценить время работы (гамма-процентную наработку) до достижения вероятности безотказной работы уровня 90, 95, 99%.

Форма отчета. Электронные таблицы MS Excel, реализующие задание.

КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Основные понятия математической статистики. Случайная выборка и ее описание. Статистическое распределение. Полигон частот, гистограмма, эмпирическая функция распределения. Выборочные оценки моментов.
2. Свойства основных законов распределения непрерывных случайных величин (равномерное распределение; экспоненциальное распределение; нормальное распределение).
3. Основные распределения случайных величин, используемые в математической статистике: χ^2-распределение, t-распределение Стьюдента, F - распределение Фишера.
4. Точечные оценки. Несмещенные, эффективные и состоятельные оценки. Свойства оценок математического ожидания и дисперсии.
5. Интервальное оценивание параметров статистического распределения. Доверительный интервал для математического ожидания, доверительный интервал для дисперсии.
6. Проверка гипотезы о нормальном законе распределения, используя оценки коэффициента асимметрии и эксцесса.
7. Проверка гипотезы об экспоненциальном законе распределения критерием Бартлетта.
8. Задача проверки статистических гипотез. Ошибки первого и второго рода. Критические точки.
9. Проверка гипотез о равенстве средних значений, проверка гипотез о равенстве дисперсий.
10. Проверка статистической однородности случайных данных.
11. Критерии согласия. Критерий Пирсона. Критерий Колмогорова-Смирнова.

ЛИТЕРАТУРА

1. Афифи А., Эйзен С. Статистический анализ: Подход с использованием ЭВМ. М.: Мир, 1982.-488 с.
2. Берк К., Кэйри П. Анализ данных с помощью Microsoft Excel. – М.: Издательский дом «Вильямс», 2005. – 560 с.
3. Гилевский, С.В. Теория вероятностей и математическая статистика: учеб. пособие / С.В. Гилевский, В.М. Молофеев. Мн.: БГУ, 2015. 175 с.
4. Гмурман, В.Е. Теория вероятностей и математическая статистика: учеб. пособие для студ. вузов / В.Е. Гмурман, 12-е изд., перераб. – Москва: Высшее образование, 2008. – 479 с.
5. Гмурман, В.Е. Руководство к решению задач по теории вероятностей и математической статистике: учеб. пособие для бакалавров: учебное пособие для студентов вузов / В. Е. Гмурман. - 11-е изд., перераб. и доп. - Москва: Юрайт, 2013. - 404 с.
6. Козлов А.Ю., Шишов В.Ф. Пакет анализа MS Excel в экономико-статистических расчетах: Учеб. пособие для вузов / Под. ред. проф. В.С.Мхитаряна. - М.: ЮНИТИ-ДАНА, 2003. - 139 с.
7. Левин Д.М., Стефан Д., Кребиль Т.С., Беренсон М.Л. Статистика для менеджеров с использованием Microsoft Excel. - М.: Издательский дом «Вильямс», 2004. – 1312 с.
8. Макарова Н.В., Трофимец В.Я. Статистика в Excel: Учебное пособие. – М.: Финансы и статистика, 2003. – 386 с.
9. Маталыцкий, М.А. Теория вероятностей и математическая статистика: учебник для студ. уво по физико-математическим спец. / М.А. Маталыцкий, Г.А. Хацкевич. - Минск: Вышэйшая школа, 2017. - 591 с.

СОДЕРЖАНИЕ