МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО
ОБРАЗОВАНИЯ РФ
Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Томский государственный университет систем управления и
радиоэлектроники»
Кафедра автоматизированных систем управления

ПРИКЛАДНАЯ МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Лабораторный практикум

2022

УДК 519.23/25
ББК 22.172
П75. 2.

ПРИКЛАДНАЯ МАТЕМАТИЧЕСКАЯ СТАТИСТИКА
Лабораторный практикум

Составитель А.А. Мицель
Томск: Томский государственный университет систем управления и радиоэлектроники. – 2022. – 141 с.

В пособии приводится описание шести лабораторных работ по основным разделам прикладной математической статистики – генерация случайных чисел с заданным законом распределения, оценки параметров распределения вероятностей, оценка закона распределения на основе выборочных данных, дисперсионный анализ данных, корреляционный анализ случайных данных, линейная регрессия. Работы выполняются с помощью пакета EXCEL.

Пособие подготовлено для магистрантов направления 09.04.01. – «информатика и вычислительная техника» и для магистрантов направления 01.04.02 – «прикладная математика и информатика». Представляет интерес для инженеров, аспирантов, преподавателей, ученых, занимающихся вопросами обработки данных.

ОГЛАВЛЕНИЕ

ЛАБОРАТОРНАЯ РАБОТА №1

Тема. Генерация случайных чисел с заданным законом распределения

Цель работы:

1. Научиться использовать генератор случайных чисел пакета EXCEL для генерации случайных чисел с заданным законом распределения.
2. Познакомиться со способами представления выборочных данных.

1.1 Практическое задание 1

Задан закон распределения F *дискретной* случайной величины (приложение 1.1)

Требуется:

1) Сгенерировать средствами пакета EXCEL выборку из 100 значений случайной величины с законом F (для генерации случайных чисел распределенных по законам, которые отсутствуют в генераторе случайных чисел пакета EXCEL см. приложение 3).
2) Представить выборку в виде вариационного ряда.
3) Построить статистический ряд абсолютных частот, относительных частот и накопленных частот.
4) Построить полигон частот и сравнить его с многоугольником теоретического распределения *F*.
5) Найти основные выборочные характеристики:

1.5.1) среднее $\overline{X}$, дисперсию s^2, несмещённую дисперсию s_0^2, выборочный коэффициент асимметрии $A = \frac{\mu_3}{\mu_2^{3/2}} = \frac{\mu_3}{s^{3/2}}$, выборочный коэффициент эксцесса $E = \frac{\mu_4}{s^4} - 3$ и сравнить их с характеристиками теоретического распределения F;

1.5.2) Вычислить коэффициент вариации $V = \frac{s}{|\overline{x}|} 100\%$;

1.5.3) Вычислить квантили x_p порядка p (*p* = 0,25; 0,5; 0,5) по формуле

$x_p = x_l, \quad l = [np]+1.$

1.5.4) Вычислить выборочную моду распределения и сравнить с теоретической модой

1.2 Практическое задание 2

Задан закон распределения F *непрерывной* случайной величины (приложение 1). Требуется:

1) Сгенерировать средствами пакета EXCEL выборку из 100 значений случайной величины с законом F (для генерации случайных чисел распределенных по законам, которые отсутствуют в генераторе случайных чисел пакета EXCEL см. приложение 3).
2) Представить выборку в виде вариационного ряда.
3) Построить сгруппированный статистический ряд абсолютных частот, относительных частот и плотностей частот.
4) Построить гистограмму и сравнить ее с графиком плотности теоретического распределения F. Для корректного сопоставления гистограммы с графиком плотности теоретического распределения, следует помнить, что EXCEL при одновременном отображении графика и гистограммы, помещает точки графика в середину столбца гистограммы. Следовательно, значения плотности должны быть подсчитаны для середин столбцов гистограммы.
5) Построить график эмпирической функцию распределения и сравнить с графиком теоретического распределения F (для построения графиков использовать не менее 40 точек).
6) Найти основные выборочные характеристики:

2.6.1) среднее $\overline{x}$, дисперсию s^2, несмещённую дисперсию s_0^2, выборочный коэффициент асимметрии $A = \frac{\mu_3}{\mu_2^{3/2}} = \frac{\mu_3}{s^{3/2}}$, выборочный коэффициент эксцесса $E = \frac{\mu_4}{s^4} - 3$ и сравнить их с характеристиками теоретического распределения F;

2.6.2) Вычислить коэффициент вариации $V = \frac{s}{|\overline{x}|} 100\%$;

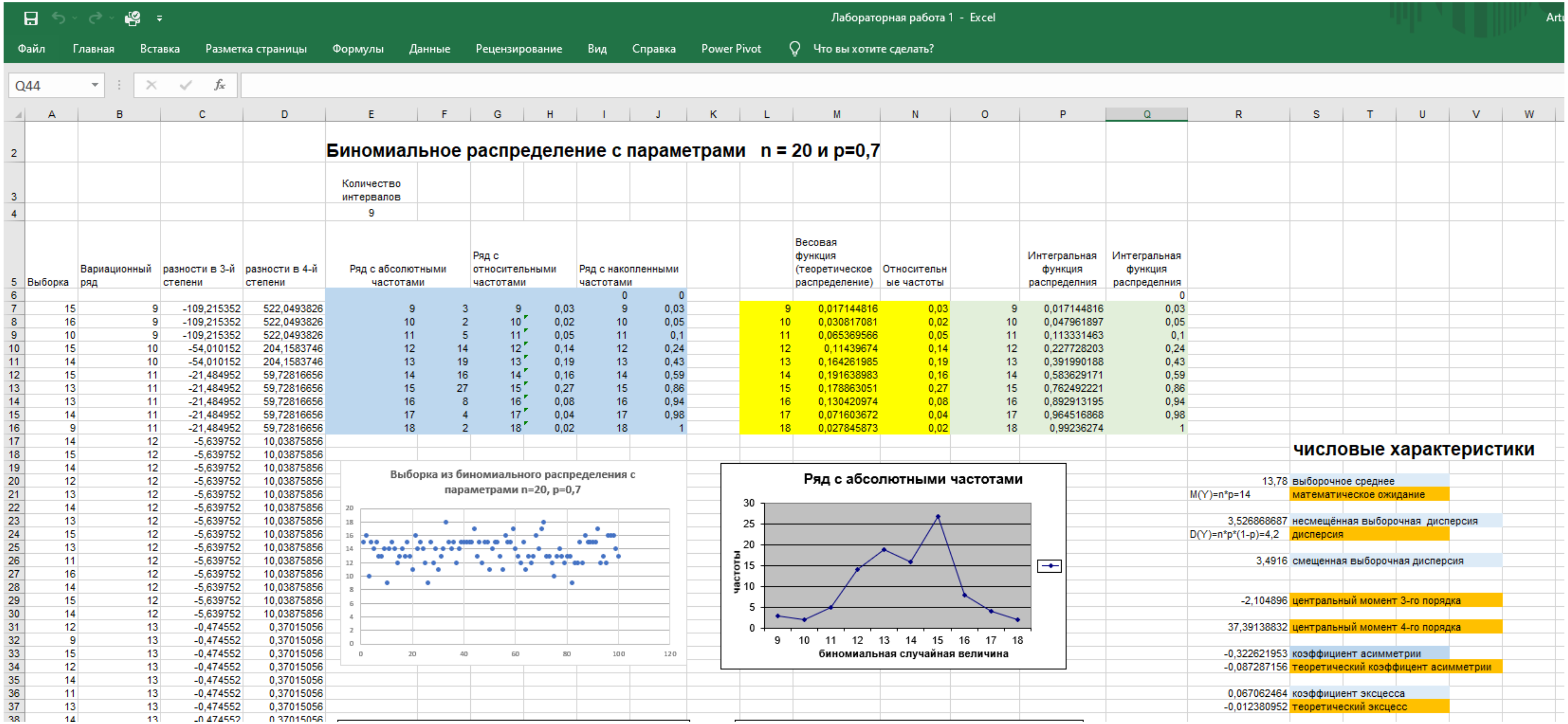
2.6.3) Вычислить квантили x_p порядка p (p = 0,25; 0,5; 0,5) по формуле

$$x_p = x_l, \quad l = [np] + 1.$$

2.6.4) Вычислить выборочную моду распределения и сравнить с теоретической модой

1.3. Excel-скриншоты заданий лабораторной работы №1

Задание 1

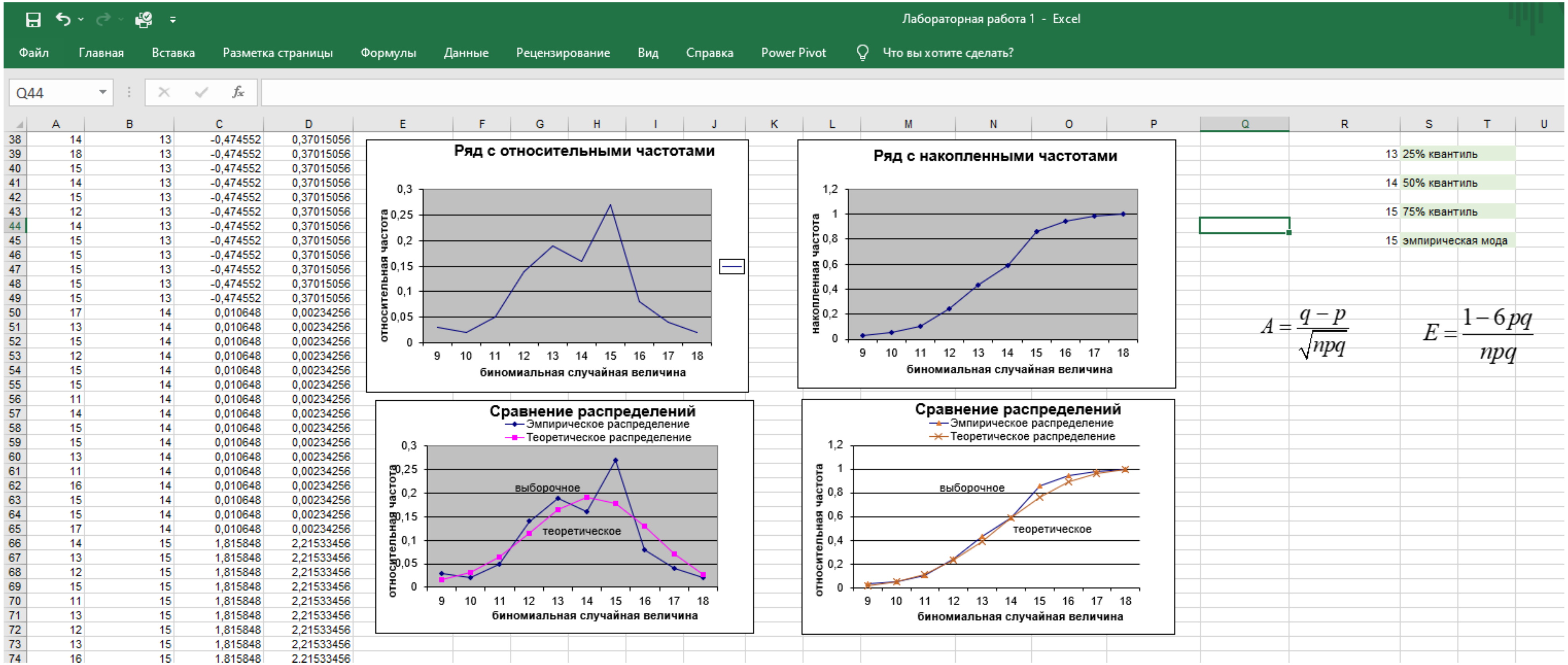


	Выборка	Вариационный ряд	разности в 3-й степени	разности в 4-й степени
7	15	9	-109,215352	522,0493826
8	16	9	-109,215352	522,0493826
9	10	9	-109,215352	522,0493826
10	15	10	-54,010152	204,1583746
11	14	10	-54,010152	204,1583746
12	15	11	-21,484952	59,72816656
13	13	11	-21,484952	59,72816656
14	13	11	-21,484952	59,72816656
15	14	11	-21,484952	59,72816656
16	9	11	-21,484952	59,72816656
17	14	12	-5,639752	10,03875856
18	15	12	-5,639752	10,03875856
19	14	12	-5,639752	10,03875856
20	12	12	-5,639752	10,03875856
21	13	12	-5,639752	10,03875856
22	14	12	-5,639752	10,03875856
23	13	12	-5,639752	10,03875856
24	15	12	-5,639752	10,03875856
25	13	12	-5,639752	10,03875856
26	11	12	-5,639752	10,03875856
27	16	12	-5,639752	10,03875856
28	14	12	-5,639752	10,03875856
29	15	12	-5,639752	10,03875856
30	14	12	-5,639752	10,03875856
31	12	13	-0,474552	0,37015056
32	9	13	-0,474552	0,37015056
33	15	13	-0,474552	0,37015056
34	12	13	-0,474552	0,37015056
35	14	13	-0,474552	0,37015056
36	11	13	-0,474552	0,37015056
37	13	13	-0,474552	0,37015056
38	14	13	-0,474552	0,37015056

	Ряд с абсолютными частотами		Ряд с относительными частотами		Ряд с накопленными частотами	
6					0	0
7	9	3	9	0,03	9	0,03
8	10	2	10	0,02	10	0,05
9	11	5	11	0,05	11	0,1
10	12	14	12	0,14	12	0,24
11	13	19	13	0,19	13	0,43
12	14	16	14	0,16	14	0,59
13	15	27	15	0,27	15	0,86
14	16	8	16	0,08	16	0,94
15	17	4	17	0,04	17	0,98
16	18	2	18	0,02	18	1

		Весовая функция (теоретическое распределение)	Относительные частоты		Интегральная функция распределния	Интегральная функция распределния
6						0
7	9	0,017144816	0,03	9	0,017144816	0,03
8	10	0,030817081	0,02	10	0,047961897	0,05
9	11	0,065369566	0,05	11	0,113331463	0,1
10	12	0,11439674	0,14	12	0,227728203	0,24
11	13	0,164261985	0,19	13	0,391990188	0,43
12	14	0,191638983	0,16	14	0,583629171	0,59
13	15	0,178863051	0,27	15	0,762492221	0,86
14	16	0,130420974	0,08	16	0,892913195	0,94
15	17	0,071603672	0,04	17	0,964516868	0,98
16	18	0,027845873	0,02	18	0,99236274	1

13,78	выборочное среднее
M(Y)=n*p=14	математическое ожидание
3,526868687	несмещённая выборочная дисперсия
D(Y)=n*p*(1-p)=4,2	дисперсия
3,4916	смещенная выборочная дисперсия
-2,104896	центральный момент 3-го порядка
37,39138832	центральный момент 4-го порядка
-0,322621953	коэффициент асимметрии
-0,087287156	теоретический коэффицент асимметрии
0,067062464	коэффициент эксцесса
-0,012380952	теоретический эксцесс



	A	B	C	D
38	14	13	-0,474552	0,37015056
39	18	13	-0,474552	0,37015056
40	15	13	-0,474552	0,37015056
41	14	13	-0,474552	0,37015056
42	15	13	-0,474552	0,37015056
43	12	13	-0,474552	0,37015056
44	14	13	-0,474552	0,37015056
45	15	13	-0,474552	0,37015056
46	15	13	-0,474552	0,37015056
47	15	13	-0,474552	0,37015056
48	15	13	-0,474552	0,37015056
49	15	13	-0,474552	0,37015056
50	17	14	0,010648	0,00234256
51	13	14	0,010648	0,00234256
52	15	14	0,010648	0,00234256
53	12	14	0,010648	0,00234256
54	15	14	0,010648	0,00234256
55	15	14	0,010648	0,00234256
56	11	14	0,010648	0,00234256
57	14	14	0,010648	0,00234256
58	15	14	0,010648	0,00234256
59	15	14	0,010648	0,00234256
60	13	14	0,010648	0,00234256
61	11	14	0,010648	0,00234256
62	16	14	0,010648	0,00234256
63	15	14	0,010648	0,00234256
64	15	14	0,010648	0,00234256
65	17	14	0,010648	0,00234256
66	14	15	1,815848	2,21533456
67	13	15	1,815848	2,21533456
68	12	15	1,815848	2,21533456
69	15	15	1,815848	2,21533456
70	11	15	1,815848	2,21533456
71	13	15	1,815848	2,21533456
72	12	15	1,815848	2,21533456
73	13	15	1,815848	2,21533456
74	16	15	1,815848	2,21533456

Задание 2

Лабораторная работа 1 - Excel   Artur Mitsel

Файл | Главная | Вставка | Разметка страницы | Формулы | Данные | Рецензирование | Вид | Справка | Power Pivot | Что вы хотите сделать?

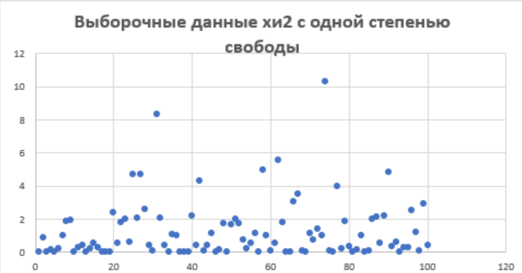
W50

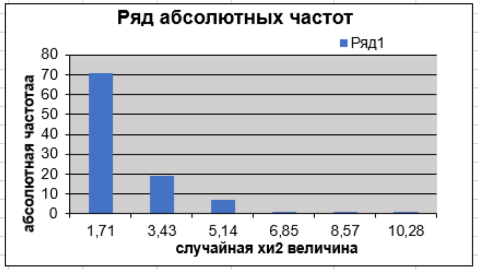
ХИ - квадрат распределение с одной степенью свободы

Количество интервалов	Шаг	минимальное значение	максимальное значение
7	1,713753596	0,00015	10,2827

	C: Вариационный ряд (Выборка)	D: разности в 3-й степени	E: разности в 4-й степени	G: Ряд с абсолютными частотами	H	I: Корректировка интервалов	J	K	L: Ряд с относительными частотами	M: Ряд с накопленными частотами
6										0
7	0,00015	-1,72129731	2,06288263	0,000147036	1	1,7139	71	1,71	0,71	0,71
8	0,00016	-0,05333438	0,02007596	1,713900632	70	3,42765	19	3,43	0,19	0,9
9	0,00026	-1,91709518	2,38154193	3,427654227	19	5,14141	7	5,14	0,07	0,97
10	0,00104	-1,30862307	1,43137281	5,141407823	7	6,85516	1	6,85	0,01	0,98
11	0,00214	-1,88198302	2,32356201	6,855161419	1	8,56892	1	8,57	0,01	0,99
12	0,00233	-1,12668587	1,17238557	8,568915015	1	10,2827	1	10,28	0,01	1
13	0,00237	-0,01275532	0,00298027	10,28266861	1					
14	0,00275	0,207493533	0,12283989							
15	0,00341	0,285549021	0,18803596							
16	0,00449	-1,77289154	2,14573568							
17	0,00655	-0,91833214	0,89261952							
18	0,00779	-0,58040019	0,484139							
19	0,00856	-1,78500931	2,16531282							
20	0,009	-1,15422246	1,21074503							
21	0,017	-0,30972904	0,20956041							
22	0,0238	-0,96492928	0,95351453							
23	0,02937	-1,84020038	2,25503604							
24	0,03213	-1,9021155	2,35676264							
25	0,03603	-1,90795379	2,3664126							
26	0,03612	1,399984046	1,56614072							
27	0,0374	-0,33503649	0,2326963							
28	0,04398	0,190124317	0,10932435							
29	0,06983	0,398768276	0,29351308							
30	0,07096	-0,26327219	0,16873557							
31	0,08272	41,98573007	145,926986							
32	0,08728	0,574664434	0,47777024							
33	0,10508	41,98573007	145,926986							
34	0,10974	2,350567654	3,12533811							

T	Теоретическая плотность	Эмпирическая плотность	Теоретическое распределение	Накопленные частоты
0,00			0,00	0,00
1,71	0,80901463	0,41429526	0,80901463	0,71
3,43	0,126963269	0,11086775	0,935977899	0,9
5,14	0,040641453	0,04084601	0,976619352	0,97
6,85	0,014516521	0,00583514	0,991135873	0,98
8,57	0,005446656	0,00583514	0,996582528	0,99
10,28	0,002072672	0,00583514	0,998655201	1

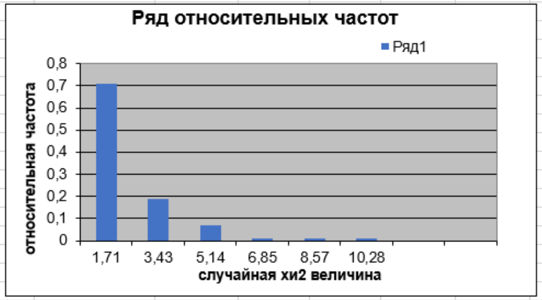


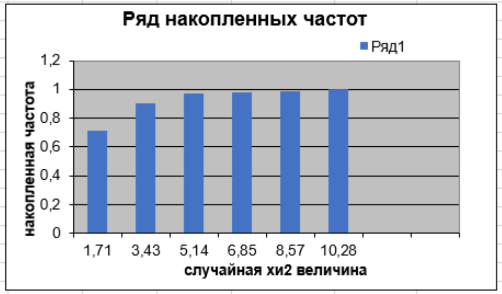


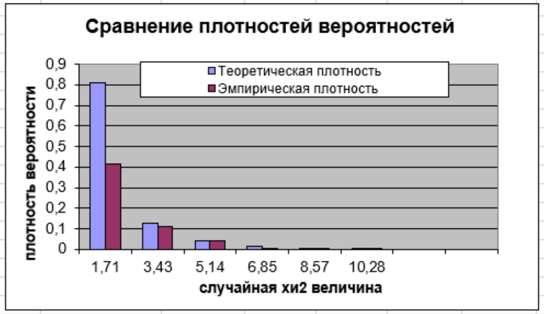
числовые характеристики

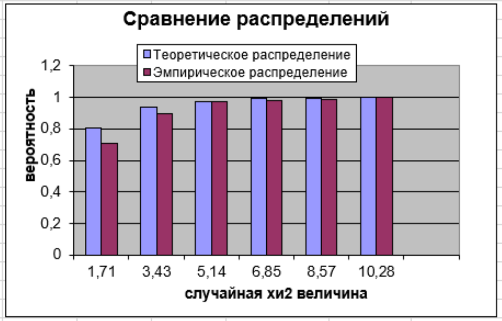
1,242428126	выборочное среднее
M(Y)=k=1	математическое ожидание
3,085808399	несмещённая выборочная дисперсия
D(Y)=2k=2	дисперсия
3,054950315	смещенная выборочная дисперсия
13,64792795	центральный момент 3-го порядка
104,5829065	центральный момент 4-го порядка
2,555997909	коэффициент асимметрии
2,828427125	теоретический коэффицент асимметрии

	C	D	E
34	0,10974	2,350567654	3,12533811
35	0,11179	-0,59545309	0,50095272
36	0,11473	-1,61231879	1,89060681
37	0,14304	355,5747177	2519,07539
38	0,14863	0,487023409	0,38317731
39	0,19161	-0,57714911	0,48052654
40	0,19346	-1,91303417	2,37481784
41	0,20187	-0,00403701	0,00064281
42	0,23837	-0,01558801	0,00389392
43	0,25426	-1,91664573	2,38079752
44	0,27043	-1,89711786	2,34851001
45	0,28619	-1,755785	2,11817472
46	0,30726	0,78370227	0,72255039
47	0,31194	-0,51486171	0,41265533
48	0,35775	27,86430727	84,4755447
49	0,40113	-1,47121046	1,67327147
50	0,40732	-0,58240288	0,48636766
51	0,40828	-0,00231762	0,00030671
52	0,40984	-1,90691394	2,36469313
53	0,41783	-1,32879814	1,46087149
54	0,44094	0,125347789	0,06273197
55	0,51886	-1,87645494	2,31446625
56	0,54299	0,081315923	0,03522904
57	0,54789	0,384871972	0,2799551
58	0,56584	0,096201425	0,04408004
59	0,56584	-0,11498123	0,05591169
60	0,60151	-1,16033253	1,21929827
61	0,61233	-0,30972904	0,20956041
62	0,72677	-0,00110219	0,00011385
63	0,75616	-1,90708518	2,36497628
64	0,86601	50,05011555	184,447786
65	0,99263	-0,01373471	0,00328922
66	1,00092	-1,55971589	1,80881435
67	1,00295	-0,37882372	0,27410454
68	1,00878	79,21523134	340,208317
69	1,02961	0,155883467	0,08389448
70	1,0832	-1,91716558	2,38165855
71	1,11009	-1,87849265	2,31781801
72	1,12879	5,795712182	10,4105997
73	1,13913	11,41637599	25,7061883
74	1,22971	-1,5413881	1,78053016
75	1,38816	-1,90512816	2,36174095
76	1,67566	-0,00146755	0,00016677









T	U
8,206046094	коэффициент эксцесса
12	теоретический эксцесс
0,087280916	25% квантиль
0,565835426	50% квантиль
1,88036318	75% квантиль
1,133696879	эмпирическая мода
0	xm-1
2,57077743	xm
0	pm-1
0,414295265	pm
0,110867747	pm+1

$$A = \frac{M(\xi - k)^3}{(2k)^{3/2}} = \sqrt{8/k}$$

$$E = \frac{M(\xi - \mu)^4}{\sigma^4} - 3 = 12/k$$

$$Mo = x_{m-1} + \frac{(x_m - x_{m-1})(\rho_m - \rho_{m-1})}{2\rho_m - \rho_{m-1} - \rho_{m+1}}$$

Приложение 1.1

Варианты заданий по лабораторной работе №1

Вариант 1.

1). F - биномиальное распределение с параметрами $n = 20$ и $p = 0{,}7$.

2). F - распределение χ^2 с одной степенью свободы.

Вариант 2.

1). F - биномиальное распределение с параметрами $n = 100$ и $p = 0{,}15$.

2). F - закон равномерной плотности на (-2; 5).

Вариант 3.

1) F - биномиальное распределение с $n = 50$ и $p = 0{,}42$.

2) F - показательный закон с параметром $\lambda = 0{,}04$.

Вариант 4.

1). F - закон Пуассона с параметром $\lambda = 8$.

2). F - распределение χ^2 с 2 степенями свободы.

Вариант 5..

1). F - биномиальное распределение с параметрами $n = 80$ и $p = 0{,}2$.

2). F - распределение Стьюдента с 3 степенями свободы.

Вариант 6.

1). F - закон Пуассона с параметром $\lambda = 12$.

2). F - нормальный закон с параметрами $a = -2$ и $\sigma = 3$.

3). .

Вариант 7.

1). F - биномиальное распределение с параметрами $n = 30$ и $p = 0{,}6$.

2). F - нормальный закон с параметрами $a = 0$ и $\sigma = 3$.

Вариант 8.

1). F - биномиальное распределение с параметрами $n = 40$ и $p = 0,6$.

2). F - нормальный закон с параметрами $a = 2$ и $\sigma = 2$.

Вариант 9.

1). F - закон Пуассона с параметром $\lambda = 10$.

2). F - показательный закон с параметром $\lambda = 0,1$.

Вариант 10.

1). F - биномиальное распределение с параметрами $n = 50$ и $p = 0,3$.

2). F - распределение χ^2 с одной степенью свободы.

Приложение 1.2

Генерация случайных чисел в EXCEL

Для генерации случайных чисел в пакете EXCEL 2003 выберите меню “Сервис”, “Пакет анализа”, ”Генерация случайных чисел”. Если в меню “Сервис” отсутствует подменю “Пакет анализа”, следует зайти в меню “Сервис”, “Надстройки” и подключить “Пакет анализа”. В EXCEL 2007 выберите меню “Данные”, “Анализ данных”. Если в меню “Данные” отсутствует подменю “Анализ данных”, то следует нажать кнопку “Office”, перейти в “Параметры Excel”, выбрать “Надстройки”, нажать “Перейти” и подключить “Пакет анализа”. Параметры генератора случайных чисел “число переменных” и ”число случайных чисел”. определяют соответственно число столбцов и строк для вывода случайных чисел. Можно использовать, например, “число переменных” для того, чтобы получить сразу несколько независимых выборок-столбцов объема, определяемого параметром “число случайных чисел”.

Приложение 1.3

Статистические функции пакета EXEL

1. Функции связанные с основными законами распределения случайных величин

- **БИНОМРАСП** (**число успехов**; **число испытаний**; **вероятность успеха**; **интегральная**)

Возвращает вероятности связанные с биномиальным распределением. Функция БИНОМРАСП используется для подсчета вероятностей числа успехов в испытаниях по схеме Бернулли.

Число успехов — количество успешных испытаний (m).

Число испытаний — общее число независимых испытаний (n).

Вероятность успеха — вероятность успеха в каждом испытании (p).

Интегральная — это логическое значение, определяющее форму функции. Если аргумент интегральная имеет значение ИСТИНА (1), то функция БИНОМРАСП возвращает вероятность того, что число успешных испытаний не более значения «число успехов»; если этот аргумент имеет значение ЛОЖЬ (0), то возвращается вероятность того, что число успешных испытаний в точности равно значению аргумента «число успехов».

Таким образом:

БИНОМРАСП (m; n; p; 0) = $P_n(m) = C_n^m p^m (1-p)^{n-m}$;

БИНОМРАСП (m; n; p; 1) = $\sum_{k=0}^{m} P_n(k) = \sum_{k=0}^{m} C_n^k p^k (1-p)^{n-k}$.

- **ПУАССОН**(x; **среднее**; **интегральная**)

Возвращает вероятности, связанные с распределением Пуассона (например, вероятности числа событий в простейшем потоке за некоторый промежуток времени, при известном среднем числе событий)

x — количество событий.

Среднее — среднее число событий (λ).

Интегральная — логическое значение, определяющее форму возвращаемого распределения вероятностей. Если аргумент «интегральная» имеет значение ИСТИНА (1), то функция ПУАССОН возвращает вероятность того, что число случайных событий будет от 0 до x включительно. Если этот аргумент имеет значение ЛОЖЬ (0), то возвращается вероятность того, что событий будет в точности x.

Таким образом**:** ПУАССОН (x; λ; 0) = $\lambda^x e^{-\lambda} / x!$: ПУАССОН (x; λ; 1) = $\sum_{k=0}^{x} \lambda^k e^{-\lambda} / k!$.

- **ГИПЕРГЕОМЕТ**(**число успехов в выборке**; **размер выборки**; **число успехов в совокупности**; **размер совокупности**)

Возвращает вероятности для гипергеометрического распределения. ГИПЕРГЕОМЕТ возвращает вероятность заданного количества успехов в выборке, если заданы размер выборки, количество успехов в генеральной совокупности и размер генеральной совокупности.

Число успехов в выборке — это количество успешных испытаний в выборке (m).

Размер выборки — размер выборки (n).

Число успехов в совокупности — количество успешных испытаний в генеральной совокупности (M).

Размер совокупности — размер генеральной совокупности (N).

Например, из генеральной совокупности, содержащей N шаров, среди которых M красных, выбирается наудачу n шаров. Тогда, вероятность того, что среди них ровно m красных равна: $P = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n} =$ ГИПЕРГЕОМЕТ(m; n; M; N).

- **НОРМРАСП**(**x**; **среднее**; **стандартное откл**; **интегральная**)

Возвращает вероятности, связанные с нормальным распределением.

x — значение, для которого определяется вероятность.

Среднее — математическое ожидание распределения (a).

Стандартное откл — среднеквадратическое отклонение распределения (σ).

Интегральная — логическое значение, определяющее форму функции. Если интегральная имеет значение ИСТИНА (1), то функция НОРМРАСП возвращает функцию распределения от аргумента x; если это аргумент имеет значение ЛОЖЬ (1), то возвращается плотности распределения от аргумента x.

Таким образом:

НОРМРАСП(x; a; σ; 0) $= \frac{1}{\sqrt{2\pi}\sigma} å^{-\frac{(x-a)^2}{2\sigma^2}}$;

НОРМРАСП(x; a; σ; 1) $= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} å^{-\frac{(t-a)^2}{2\sigma^2}} dt$.

- **НОРМСТРАСП**(**z**)

Возвращает функцию распределения стандартной нормальной величины, т.е. НОРМСТРАСП(z) = $P(\xi < z) = F_\xi(z)$, где $\xi \in N_{0,1}$, $F_\xi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$.

- **НОРМОБР**(**вероятность**; **среднее**; **стандартное откл**)

Возвращает квантиль нормального распределения для указанной вероятности, то есть НОРМСТОБР(α) возвращает значение τ_α, для которого $P(\xi < \tau_\alpha) = \alpha, \quad \xi \in N_{a,\sigma^2}$.

Вероятность — вероятность, соответствующая квантили.

Среднее — математическое ожидание распределения.

Стандартное откл — среднеквадратическое отклонение распределения.

- **НОРМСТОБР**(**вероятность**)

Возвращает квантиль стандартного нормального распределения для указанной вероятности, то есть НОРМСТОБР(α) возвращает значение τ_α, для которого $P(\xi < \tau_\alpha) = \alpha, \quad \xi \in N_{0,1}$

Вероятность — вероятность, соответствующая квантили.

- **СТЬЮДРАСП**(**x**; **степени свободы**; **хвосты**)

Возвращает вероятности, связанные с распределением Стьюдента.

x — численное значение, для которого требуется вычислить вероятности.

Степени свободы — число степеней свободы распределения.

Хвосты — число учитываемых хвостов распределения. Если хвосты = 1, то функция СТЬЮДРАСП возвращает вероятность того, что случайная величина, распределенная по закону Стьюдента, примет значение большее чем x. Т.е. СТЬЮДРАСП(x; n; 1) = $P(\xi > x) = 1 - F_\xi(x)$, где $\xi \in T_n$. Если хвосты = 2, то функция СТЬЮДРАСП возвращает вероятность того, что случайная величина, распределенная по закону Стьюдента, примет значение, большее, чем $|x|$. Т.е. СТЬЮДРАСП(x; n; 2) = $P(|\xi| > x)$, где $\xi \in T_n$.

- **СТЬЮДРАСПОБР**(**вероятность**; **степени свободы)** возвращает коэффициент Стьюдента t_α, соответствующий заданной вероятности α: т.е. значение t_α для которого $P(|\xi|>t_\alpha)=\alpha$, что тоже самое, что квантиль распределения Стьюдента уровня $1-\alpha/2$, то есть значение $\tau_{1-\alpha/2}$, для которого $P(\xi<\tau_{1-\alpha/2})=1-\alpha/2, \quad \xi\in T_n$.

Вероятность — вероятность, для которой находится значение коэффициента.

Степени свободы — число степеней свободы, характеризующее распределение.

- **ХИ2РАСП**(**x**; **степени свободы**)

Возвращает вероятность того, что случайная величина, распределенная по закону хи-квадрат примет значение, большее, чем x, т.е. ХИ2РАСП(x; n**)** = $P(\xi>x)=1-F_\xi(x)$, где $\xi\in\chi^2_n$.

x — это значение, для которого требуется вычислить вероятность.

Степени свободы — это число степеней свободы распределения хи-квадрат.

- **ХИ2ОБР**(**вероятность**; **степени свободы)**

возвращает критическую точку распределения хи-квадрат для заданной вероятности, то есть ХИ2ОБР(α, n)=t_α, где t_α значение, для которого $P(\xi>t_\alpha)=\alpha$, что тоже самое, что квантиль распределения хи-квадрат уровня $1-\alpha$.

Вероятность — вероятность, для которой находится критическая точка.

Степени свободы — число степеней свободы, характеризующее распределение.

- **FРАСП**(**x**; **степени свободы1**; **степени свободы2**)

Возвращает вероятность того, что случайная величина, распределенная по закону Фишера примет значение, большее, чем x, т.е. FРАСП(x; n1; n2) = $P(\xi>x)=1-F_\xi(x)$, где $\xi\in F_{n1,n2}$.

x — это значение, для которого требуется вычислить вероятность.

Степени свободы1, степени свободы2 — число степеней свободы, характеризующих распределение.

- **FРАСПОБР**(**вероятность**; **степени свободы1**; **степени свободы2**)

возвращает критическую точку распределения Фишера для заданной вероятности, то есть FРАСПОБР($\alpha, n1, n2$)=t_α, где t_α значение, для которого $P(\xi > t_\alpha) = \alpha$, что тоже самое, что квантиль распределения Фишера уровня $1-\alpha$.

Вероятность — вероятность, для которой находится критическая точка.

Степени свободы1, степени свободы2 — число степеней свободы, характеризующих распределение.

Приложение 1.4

Примеры использования функций EXCEL

Пример 1. Игральная кость подбрасывается 24 раза. Найти вероятность того, что 6 очков выпадут ровно 3 раза. Найти точное значение вероятности и приближенные, используя локальную формулу Муавра-Лапласа и формулу Пуассона.

Решение. Требуется найти вероятность того, что в n=24 испытаниях по схеме Бернулли с вероятностью успеха 1/6, число успехов будет равно 3. Для точного вычисления вероятности используем функцию **БИНОМРАСП.** Если параметр **интегральная** имеет значение **ЛОЖЬ (0)**, то функция **БИНОМРАСП** возвращает вероятность того, что число успешных испытаний в точности равно значению аргумента **число успехов**. Таким образом, для $n = 24$, $m = 3$, $p = 1/6$ находим:

$P_{24}(3)$ =БИНОМРАСП(3;24;1/6;0)=0,203681.

Найдем приближенное значение вероятности, используя локальную формулу Муавра-Лапласа. Согласно этой формуле, вероятность $P_n(m) \approx \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{npq}} e^{-\frac{(m-np)^2}{2npq}}$, т.е. приближенно равна плотности нормального распределения со средним np и среднеквадратичным отклонением $\sqrt{npq}$ в точке m. Значения плотности распределения нормальной величины возвращает функция **НОРМРАСП,** при значении параметра **интегральная** равном **ЛОЖЬ (0)**.

Таким образом: $P_{24}(3) \approx$ НОРМРАСП($3;\ 24 \cdot 1/6;\ \sqrt{24 \cdot 1/6 \cdot 5/6};\ 0$)=0,188073.

Найдем приближенное значение той же вероятности, используя формулу Пуассона. Согласно этой формуле, при малых p вероятность $P_n(m) \approx \frac{\lambda^m e^{-\lambda}}{m!}, \lambda = np$, т.е. приближенно равна вероятности пуассоновского распределения с параметром (средним значением) $\lambda = np$ в точке m. Вероятности отдельных значений для распределения Пуассона возвращает функция ПУАССОН при значении параметра **интегральная** равном **ЛОЖЬ (0)**. Таким образом: $P_{24}(3) \approx$ПУАССОН($3;\ 24 \cdot 1/6;\ 0$)=0,195367.

Заметим, что погрешность при использовании формулы Муавра-Лапласа составила $\approx$7,8%, а при использовании формулы Пуассона $\approx$4,1%.

Пример 2. Вероятность искажения одного символа при передачи сообщения равна 0,01. Какова вероятность, что сообщение, содержащее 200 символов, содержит не более 2-х искажений. Найти точное значение вероятности и приближенные, используя локальную формулу Муавра-Лапласа и формулу Пуассона.

Решение. Требуется найти вероятность того, что в n=200 испытаниях по схеме Бернулли с вероятностью успеха 0,01, число успехов будет не более 2. Для точного вычисления вероятности используем функцию **БИНОМРАСП.** Если параметр **интегральная** имеет значение **ИСТИНА (1)**, то функция **БИНОМРАСП** возвращает вероятность того, что число успешных испытаний лежит в пределах от 0 до значения, определяемого аргументом **число успехов**. Таким образом, для $n = 200$, $m = 2$, $p = 0{,}01$ находим: $P_{200}(0 \le m \le 2)$ =БИНОМРАСП(2;200;0,01;1)=0,676679.

Найдем приближенное значение вероятности, используя локальную формулу Муавра-Лапласа. Согласно этой формуле, вероятность $P_n(m)$, приближенно равна плотности нормального распределения в точке m со средним np и среднеквадратичным отклонением $\sqrt{npq}$. Значения плотности распределения нормальной величины возвращает функция **НОРМРАСП,** при значении параметра **интегральная** равном **ЛОЖЬ** **(0)**. Таким образом: $P_{200}(0 \le m \le 2) = P_{200}(0) + P_{200}(1) + P_{200}(2) \approx$

$\approx$НОРМРАСП($0;\ 2;\ \sqrt{1{,}98};\ 0$)+НОРМРАСП($1;\ 2;\ \sqrt{1{,}98};\ 0$)+НОРМРАСП($2;\ 2;\ \sqrt{1{,}98};\ 0$)=0,6 07013.

Найдем приближенное значение той же вероятности, используя формулу Пуассона. Согласно этой формуле, вероятность $P_n(m)$ при малых p приближенно равна вероятности пуассоновского распределения в точке m со средним значением $\lambda = np$. Значения вероятностей для распределения Пуассона возвращает функция **ПУАССОН**. Причем, если значение параметра **интегральная** равно **ИСТИНА (1)**, то функция **ПУАССОН** возвращает вероятность того, что случайная величина, имеющая распределение Пуассона примет значения в пределах от 0 до значения, определяемого аргументом **х.** Таким образом: $P_{200}(0 \le m \le 2) \approx$ПУАССОН($2;\ 200 \cdot 0{,}01;\ 1$)=0,676676.

Заметим, что погрешность, полученная при использования формулы Пуассона, в данном случае на порядок ниже, чем при использовании локальной формулы Муавра-Лапласа. Смысла использовать интегральную формулу Муавра-Лапласа в данном случае нет, поскольку интервал значений m мал $(0 \le m \le 2)$.

Найдем приближенное значение вероятности, используя теперь интегральную формулу Муавра-Лапласа. Согласно этой формуле, вероятность $P_n(m_1 \le m \le m_2) \approx \Phi\left(\frac{m_2 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{m_1 - np}{\sqrt{npq}}\right)$, т.е. приближенно равна вероятности попадания в интервал $(m_1; m_2)$ нормальной случайной величины со средним np и среднеквадратичным отклонением $\sqrt{npq}$. Следовательно, если $F_\xi(x)$ - функция распределения нормальной случайной величины с параметрами $a = np$ и $\sigma^2 = npq$, то $P_n(m_1 \le m \le m_2) \approx F_\xi(m_2) - F_\xi(m_1)$. Значения функции распределения нормальной величины возвращает функция **НОРМРАСП,** при значении параметра **интегральная** равном **ИСТИНА (1)**. Таким образом: $P_n(0 \le m \le 2) \approx$НОРМРАСП$(2; 2; \sqrt{1{,}98}; 1)$-НОРМРАСП$(0; 2; \sqrt{1{,}98}; 1)$=.

Пример 3. Монета подбрасывается 10000 раз. Найти вероятность того, что орел выпадет более 5100 раз. Найти точное значение вероятности и приближенное, используя интегральную формулу Муавра-Лапласа.

Решение. Требуется найти вероятность того, что в n=10000 испытаниях по схеме Бернулли с вероятностью успеха 1/2, число успехов будет более 5100. Для точного вычисления вероятности используем функцию **БИНОМРАСП.** Если параметр **интегральная** имеет значение **ИСТИНА (1)**, то функция **БИНОМРАСП** возвращает вероятность того, что число успешных испытаний не менее значения аргумента **число успехов**. Таким образом, находим: $P_{10000}(m > 5100) = 1 - P_{10000}(m \le 5100) =$ 1-БИНОМРАСП(5100;10000;1/2;1)=0,022213.

Найдем приближенное значение вероятности, используя интегральную формулу Муавра-Лапласа. Согласно этой формуле, вероятность $P_n(m_1 \le m \le m_2) \approx \Phi\left(\frac{m_2 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{m_1 - np}{\sqrt{npq}}\right)$, т.е. приближенно равна вероятности попадания в интервал $(m_1; m_2)$ нормальной случайной величины со средним np и среднеквадратичным отклонением $\sqrt{npq}$. Следовательно, если $F_\xi(x)$ - функция распределения нормальной случайной величины с параметрами $a = np$ и $\sigma^2 = npq$, то $P_n(m_1 \le m \le m_2) \approx F_\xi(m_2) - F_\xi(m_1)$.

Значения функции распределения нормальной величины возвращает функция **НОРМРАСП,** при значении параметра **интегральная** равном **ИСТИНА (1)**. Таким образом:

$P_{10000}(m > 5100) = 1 - P_{10000}(m \le 5100) \approx 1 - F_\xi(5100) =$ НОРМРАСП(

$5100; 10000 \cdot 1/2; \sqrt{10000 \cdot 1/2 \cdot 1/2}; 1)$)=0,02275.

Заметим, что более точной является приближенная формула $P_n(m_1 \le m \le m_2) \approx F_\xi(m_2 + 0{,}5) - F_\xi(m_1 - 0{,}5)$. Если использовать ее, то получим:

$P_{10000}(m > 5100) = 1 - P_{10000}(m \le 5100) \approx 1 - F_\xi(5100 + 1/2) =$

=НОРМРАСП($5100; 10000 \cdot 1/2; \sqrt{10000 \cdot 1/2 \cdot 1/2}; 1)$)=0,022216.

Пример 4. Случайная величина распределена по нормальному закону с параметрами $a = 3$ и $\sigma^2 = 7$. Найти:

а) вероятность того, что ξ примет значение в интервале $(1; 4)$;

б) квантиль распределения уровня $0{,}8$5;

в) критическую точку распределения уровня 0,07;

г) интервал, симметричный относительно математического ожидания, в котором с вероятностью 0,95 содержатся значения ξ.

Решение.

а) Вероятности, связанные с нормальным распределением, можно вычислять, используя функцию **НОРМРАСП**. Данная функция**,** при значении параметра **интегральная** равном **ИСТИНА(1)**, возвращает значения функции распределения $F_\xi(x)$ нормальной случайной величины. Поскольку $P(a < \xi < b) = F_\xi(b) - F_\xi(a)$, то

$P(1 < \xi < 4) = F_\xi(4) - F_\xi(1) =$

=НОРМРАСП($4; 3; \sqrt{7}; 1$) $-$ НОРМРАСП($1; 3; \sqrt{7}; 1$)=0,422426.

б) Квантили, критические точки и, вообще, различные значения, связанные с вероятностями для нормальной случайной величины, можно вычислять, используя функцию **НОРМОБР** или **НОРМСТОБР.** Функция **НОРМОБР** возвращает квантиль нормального распределения для указанной вероятности, то есть НОРМОБР($\beta;a;\sigma$) = τ_β, для которого $P(\xi<\tau_\beta)=\beta,\quad \xi\in N_{a,\sigma^2}$. Таким образом, квантиль уровня 0,85 равна $\tau_{0,85}=$ НОРМОБР($0,85;3;\sqrt{7}$)=5,742.

в) Критическая точка уровня β по определению есть значение t_β, для которого $P(\xi>t_\beta)=\beta$. Критическая точка уровня β совпадает с квантилью уровня $1-\beta$. Поэтому, критическая точка уровня 0,07 есть $t_{0,07}=\tau_{0,93}=$ НОРМОБР($0,07;3;\sqrt{7}$)=6,905.

г) Требуется найти такое значение δ, для которого $P(|\xi-M(\xi)|<\delta)=0,95$. Удобнее в данном случае воспользоваться функцией **НОРМСТОБР,** которая возвращает квантиль стандартного нормального распределения для указанной вероятности. Имеем: $P(|\xi-M(\xi)|<\delta)=2\Phi(\delta/\sigma)=2F_{0,1}(\delta/\sigma)-1=0,95$, где $F_{0,1}(x)$ - функция распределения стандартной нормальной величины. Или $F_{0,1}(\delta/\sigma)=1,95/2=0,975$. Используя **НОРМСТОБР,** находим квантиль для стандартной нормальной величины уровня 0,975: $\tau_{0,975}=$ НОРМСТОБР($0,975$)=1,96. Тогда $\delta/\sigma=1,96$, откуда $\delta=\sigma\cdot 1,96=5,186$. Таким образом, искомый интервал имеет вид: $(a-\delta;a+\delta)=(-2,186;8,186)$.

ЛАБОРАТОРНАЯ РАБОТА №2

Тема. Оценки параметров распределений вероятностей

Цель работы:

Оценка параметров распределения генеральной совокупности на основе выборочных данных.

2.1 Необходимые теоретические сведения

Различают два вида оценок параметров *точечные* и *интервальные.* Предположим, оценке подлежит параметр θ некоторого распределения вероятностей по выборочным данным $x_1, x_2, ..., x_n$ некоторой случайной величины X. Точечной оценкой параметра θ по выборочным данным является некоторый функционал $\theta^* = \varphi(x_1, x_2, ..., x_n)$, позволяющий получить наилучшую оценку в принятых критериях.

В качестве критериев, характеризующих пригодность оценки параметра распределения, используются такие ее свойства, как *состоятельность, несмещенность, эффективность* и *достаточность.*

Оценка θ^* параметра θ является:

– состоятельной, если при $n \to \infty, \quad \theta^* \to \theta$;

– несмещенной, если $M(\theta^*) = \theta$ (т. Е. математическое ожидание оценки совпадает с ее истинным значением);

– достаточной, если оценка θ^* извлекает максимальную информацию из выборки;

– эффективной, если $D(\theta^*) = \min$ (т.е. когда дисперсия оценки минимальна).

Под интервальной оценкой параметра θ понимается интервал, границы которого $a_н^*$ и $a_в^*$ являются функционалами от выборочных значений случайной величины, и который с заданной вероятностью α содержит оцениваемый параметр: $P\{a_н^* < \theta < a_в^*\} = \alpha$. Вероятность α называется *доверительной вероятностью,* а оценки $a_н^*$ и $a_в^*$ – соответственно *нижней* и *верхней доверительными границами.* Интервал $[a_н^*, a_в^*]$ называется *доверительным, интервалом.* Если длина доверительного интервала $l(\alpha) = a_в^* - a_н^* = const$, то для состоятельных и несмещенных оценок $\alpha \to 1$ при $n \to \infty$. При фиксированном объеме выборки n, α будет тем больше, чем больше l.

Различают два вида интервальных оценок: *одно-* и *двусторонние.* При *двусторонней* оценке задаются обе границы доверительного интервала, так что

$$P\{a_н^* < \theta < a_в^*\} = \alpha \text{ и } \quad P\{\theta < a_н^*\} = \alpha'; \; P\{\theta > a_в^*\} = \alpha'',$$

где $\alpha' + \alpha'' = 1 - \alpha$. Если $\alpha' = \alpha'' = \dfrac{1-\alpha}{2}$, то двусторонний доверительный интервал называется *симметричным.* Для него справедливы соотношения

$$P\{\theta < a_{в}^{*}\} = \frac{1+\alpha}{2} \text{ и } P\{a > a_{н}^{*}\} = \frac{1+\alpha}{2}.$$

При односторонних доверительных интервалах границы интервалов задаются так, чтобы

$$P\{a < a_{в}^{*}\} = \alpha \text{ или } P\{\theta > a_{н}^{*}\} = \alpha$$

Величина $q = (1-\alpha)$ — дополнение .доверительной вероятности до единицы называется *уровнем значимости.* Этим термином обозначается вероятность появления события, которую исследователь связывает с *неслучайным (значимым) событием.* Очевидно, что двусторонний интервал для симметричных распределений аналогичен одностороннему при удвоенном уровне значимости.

2.2 Методы оценки параметров распределения вероятностей

2.2.1 Метод моментов оценки параметров распределения

Идея этого метода заключается в приравнивании теоретических и эмпирических моментов.

Пусть $X = (x_1, x_2, ..., x_n)$ – независимая выборка из распределения P_θ, зависящего от неизвестного параметра $\theta = (\theta_1, \theta_2, ..., \theta_k) \in \Theta \subset R^k$. Моментом i -го порядка называется функция

$$\mu_i(\theta_1, ..., \theta_k) = E[x^i] = \begin{cases} \int x^i f(x, \theta_1, ..., \theta_k) dx, & \textit{если } x \textit{ непрерывная величина} \\ \sum\limits_j x_j^i p(x_j, \theta_1, ..., \theta_k), & \textit{если } x \textit{ дискретная величина} \end{cases}$$

где $f(x, \theta)$ – плотность распределения непрерывной случайной величины x,

$p(x_j, \theta)$ – вероятность дискретной случайной величины. Теоретический момент является функцией неизвестных параметров $\theta = (\theta_1, \theta_2, ..., \theta_k)$.

Выборочным (эмпирическим) моментом i -го порядка называется величина

$$m_i(x_1, x_2, ..., x_n) = \frac{1}{n}\sum_{j=1}^{n} x_j^i \ .$$

Отметим, что по своему определению эмпирические моменты являются функциями от выборки.

Для нахождения неизвестных параметров (будем обозначать их $\theta = (\theta_1, ..., \theta_k)$) составим систему уравнений

$$\mu_1(\theta_1, ..., \theta_k) = m_1,$$
$$\mu_2(\theta_1, ..., \theta_k) = m_2,$$
$$..........................$$
$$\mu_k(\theta_1, ..., \theta_k) = m_k.$$

Далее решаем систему относительно параметров $\theta = (\theta_1, ..., \theta_k)$. В результате получим

$$\theta_1 = \theta_1(x_1, ..., x_n),$$
$$\theta_2 = \theta_2(x_1, ..., x_n),$$
$$......................,$$
$$\theta_k = \theta_k(x_1, ..., x_n).$$

Найденные параметры зависят от выборки $X = (x_1, x_2, ..., x_n)$.

Пример 2.1. Пусть $X \in \Pi_\alpha$, где Π_α – показательный закон распределения с параметром α. Найти оценку параметра α.

Решение.

$\mu_1(\alpha) = E[x] = \int_0^\infty x \cdot \alpha e^{-\alpha x} dx = \frac{1}{\alpha}$. Приравниваем к $m_1 = \frac{1}{n}\sum_{j=1}^{n} x_j$. Отсюда получим:

$$\alpha = \frac{1}{m_1} = \frac{1}{\frac{1}{n}\sum_{j=1}^{n} x_j}.$$

Пример 2.2. Пусть $X \in U_{\alpha,\beta}$, где $U_{\alpha,\beta}$ – равномерный закон распределения с параметрами α, β. Найти оценки параметров α и β.

Решение.

$$\mu_1(\alpha,\beta) = E[x] = \frac{1}{\beta-\alpha}\int_\alpha^\beta x dx = \frac{\alpha+\beta}{2},\ \mu_2(\alpha,\beta) = E[x^2] = \frac{1}{\beta-\alpha}\int_\alpha^\beta x^2 dx = \frac{\beta^2+\alpha\beta+\alpha^2}{3}.$$

Получим систему

$$\frac{\alpha+\beta}{2} = m_1,$$

$$\frac{\beta^2+\alpha\beta+\alpha^2}{3} = m_2.$$

Решение системы: $\beta = m_1 + \sqrt{3}\sqrt{m_2 - m_1^2} = m_1 + \sqrt{3}\sigma$, $\alpha = m_1 - \sqrt{3}\sigma$. Здесь $\sigma^2 = m_2 - m_1^2$ – дисперсия выборочного распределения.

2.2.2. Метод максимального правдоподобия

Пусть $X = (x_1, x_2, ..., x_n)$ – независимая выборка из распределения P_θ, зависящего от неизвестного параметра $\theta = (\theta_1, \theta_2, ..., \theta_k) \in \Theta \subset R^k$. Функцией правдоподобия $L(\theta, x) = L(\theta_1, ..., \theta_k; x_1, ..., x_n)$ называют функцию

$$L(\theta,x) = \begin{cases} \prod\limits_{j=1}^{n} f(x_j,\theta_1,...,\theta_k), & \textit{если x непрерывная величина} \\ \prod\limits_{j=1}^{n} p(x_j,\theta_1,...,\theta_k), & \textit{если x дискретная величина} \end{cases}$$

В качестве оценки параметров $\theta = (\theta_1, ..., \theta_k)$ примем значения этих параметров, при которых функция правдоподобия принимает максимальное значение, т.е. $\theta = (\theta_1, ..., \theta_k) = \arg\left(\max\limits_\theta L(\theta, x)\right)$. Если функция $L(\theta, x) = L(\theta_1, ..., \theta_k; x_1, ..., x_n)$ является дифференцируемой по переменным $\theta_1, \theta_2, ..., \theta_k$, оценки параметров удовлетворяют системе уравнений:

$$\frac{dL(\theta_1,...,\theta_k;x)}{d\theta_i} = 0,\ i = 1, 2, ..., k.$$

Пример 2.3. Пусть $X \in \Pi_\alpha$, где Π_α – показательный закон распределения с параметром α . Найти оценку параметра α методом максимального правдоподобия.
Решение. Запишем функцию правдоподобия

$L(\alpha, x) = \prod_{j=1}^{n} \alpha e^{-\alpha x_j} = \alpha^n e^{-\alpha \sum_{j=1}^{n} x_j}$. Из условие максимума $\dfrac{dL(\alpha; x)}{d\alpha} = 0$ получим следующее уравнение: $n\alpha^{n-1} e^{-\alpha \sum_{j=1}^{n} x_j} - \alpha^n \sum_{j=1}^{n} x_j e^{-\alpha \sum_{j=1}^{n} x_j} = 0$. Отсюда следует

$\alpha = \dfrac{1}{\frac{1}{n}\sum_{j=1}^{n} x_j}$. Таким образом, эта оценка совпала с оценкой, полученной методом моментов (см. пример 1).

Пример 2.4. Пусть $X \in U_{\alpha,\beta}$, где $U_{\alpha,\beta}$ – равномерный закон распределения с параметрами α, β. Найти оценки параметров α и β методом максимального правдоподобия.

Решение.

Функция правдоподобия равна

$$L(\alpha, \beta) = \begin{cases} \prod_{j=1}^{n} \dfrac{1}{(\beta - \alpha)}, & \alpha \le x_1 \le, x_2 \le \ldots \le x_n \le \beta \\ 0, & X \in [\alpha, \beta] \end{cases}$$ или

$$L(\alpha, \beta) = \begin{cases} \dfrac{1}{(\beta - \alpha)^n}, & \alpha \le x_1 \le, x_2 \le \ldots \le x_n \le \beta \\ 0, & X \in [\alpha, \beta] \end{cases}$$, где $X = (x_1, \ldots, x_n)$ – упорядоченная выборка (вариационный ряд).

Проанализируем неравенства $\alpha \le x_1 \le, x_2 \le \ldots \le x_n \le \theta$. Если взять $\beta < x_n$, т.е. меньше максимального значения выборки, то $L(\alpha, \beta)$ обратится в ноль, и только при $\beta = x_n$ функция $L(\alpha, \beta)$ будет отлична от нуля и равна $\dfrac{1}{(\beta - \alpha)^n}$. Этот результат будет тем более верен, если $\beta < x_i,\ i < n$. Поэтому получим $\beta = x_n$. Проводя аналогичные рассуждения относительно левой границы интервала, получим $\alpha = x_1$.

Таким образом, оценки параметров α, β, полученные методом максимального правдоподобия, отличаются от оценок этих параметров, полученных методом моментов (см. пример 2).

2.2.3. Метод наименьших квадратов

Пусть дана табличная функция $y(x)$

x	x_1	x_2	…	x_n
y	y_1	y_2	…	y_n

Необходимо аппроксимировать эти данные некоторой параметрической функцией $f(\theta_1,...,\theta_k;x)$, т.е. заменить функцию $y(x)$ функцией $f(\theta_1,...,\theta_k;x)$:
$y(x) \approx f(\theta_1,...,\theta_k;x)$.

Параметры $\theta_1,...,\theta_k$ будем подбирать таким образом, чтобы расхождение табличной функции с функцией $f(\theta_1,...,\theta_k;x)$ было минимальным. Для этого построим функционал $F(\theta_1,...,\theta_k) = \sum_{j=1}^{n}\left(y_i - f(\theta_1,...,\theta_k;x_j)\right)^2$ и найдем его минимум.

Необходимое условие минимума имеет вид

$$\frac{dF(\theta_1,...,\theta_k)}{d\theta_i} = 0,\ i = 1,...,k.$$

Решаем эту систему уравнений и получаем значения параметров $\theta = (\theta_1,...,\theta_k)$.

Пример 2.5. Пусть дана независимая выборка из $X = (x_1, x_2, ..., x_n)$ из распределения P_θ. Разобьем весь диапазон данных $[x_{\min}, x_{\max}]$ на m интервалов и построим гистограмму. Обозначим середины интервалов $x_1, x_2, ..., x_m$. Тогда относительные частоты (высоты столбиков гистограммы) будут значениями табличной функции y. Таким образом, мы получили табличную функцию

x	x_1	x_2	$\cdots$	x_n
y	y_1	y_2	$\cdots$	y_n

Поставим следующую задачу. Подобрать параметры известного закона непрерывного распределения $f(\theta_1,...,\theta_k;x)$ так, чтобы расхождение между гистограммой и функцией $f(\theta,x)$ было минимально. В результате мы приходим к методу наименьших квадратов.

2.3 Вычисление точечных оценок

2.3.1. Точечная оценка параметров нормального распределения

Пусть случайная величина X имеет нормальное распределение $N(\mu,\sigma^2)$, где $\sigma>0;\ \mu\in R$. Плотность распределения имеет вид:

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}},\ x\in(-\infty,\infty).$$

Оценки среднего и дисперсии

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i,\ s^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2.$$

Оценка среднего – состоятельная, несмещенная, эффективная, достаточная и распределена как случайная величина тоже нормально со средним $M(\bar{x})=\mu$ и дисперсией $D(\bar{x})=\frac{\sigma^2}{n}$.

Оценка дисперсии – состоятельная, несмещенная, эффективная, достаточная.

2.3.2. Точечная оценка параметров показательного закона распределения

Пусть $X\in\Pi_\lambda$, где Π_λ – показательный закон распределения с параметром λ с плотностью распределения $f(x)=\lambda e^{-\lambda x},\ x\in[0,\infty)$.

Точечная оценка параметра λ^*, полученная методом максимального правдоподобия и методом моментов, совпадает и оценивается по формуле

$$\lambda^* = \frac{1}{\frac{1}{n}\sum_{j=1}^{n} x_j}.$$

2.3.3. Точечная оценка параметров равномерного закона распределения

Пусть $X \in U_{a,b}$, где $U_{a,b}$ – равномерный закон распределения с параметрами a,b. Оценки параметров a^* и b^*, полученные методом моментов, вычисляются по формулам

$$a^* = m_1 - \sqrt{3}\cdot s,\; b^* = m_1 + \sqrt{3}\cdot s.$$

Здесь $m_1 = \frac{1}{n}\sum_{i=1}^{n} x_i$ и $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ – среднее и дисперсия выборочного распределения.

2.3.4. Точечная оценка параметров биномиального закона распределения

Биномиальное распределение — распределение количества «успехов» в последовательности из n независимых случайных экспериментов, таких что вероятность «успеха» в каждом из них постоянна и равна p.

Пусть $X = (x_1, x_2, ..., x_n)$ – выборка из биномиального распределения, т.е. $X \in B_{n,p}$, где $B_{n,p}$ – биномиальный закон распределения с параметрами n, p.

Распределения вероятностей задаётся формулой:

$$f(x; n, p) = \binom{n}{x}\cdot p^x \cdot q^{n-x},\; k = 0,1,...,n,$$

где $\binom{n}{x} = \frac{n!}{(n-x)!x!}$ — биномиальный коэффициент

Здесь x – количество появлений события в серии из n испытаний, при условии, что в единичном испытании вероятность его появления равна p.

Тогда оценки параметров n^*, p^*, найденные методом моментов, вычисляются по формулам

$$p^* = 1 - \frac{s^2}{m_1};\; n^* = \frac{m_1}{p}.$$

Здесь $m_1 = \frac{1}{n}\sum_{j=1}^{n} x_j$, $s^2 = \frac{1}{n}\sum_{j=1}^{n}\left(x_j - m_1\right)^2$.

2.4 Планирование экспериментов для оценки параметров распределений

2.4.1. Нормальное распределение

Оценка среднего при известной дисперсии

Объем выборки, необходимый для оценки среднего μ с заданной предельной абсолютной ошибкой ε и доверительной вероятностью α при известной дисперсии σ^2 определяется соотношением

$$n = \left(\frac{u_\alpha \sigma}{\varepsilon}\right)^2 .$$

Пример 2.1. Напряжение зажигания газоразрядного прибора распределено нормально со стандартным, отклонением $\sigma = 50$В. Найти объем выборки, позволяющий оценить среднее значение напряжения зажигания с предельной абсолютной ошибкой $\varepsilon = 20$В при доверительной вероятности $\alpha = 0{,}95$.

Решение. Имеем $n = \left(1{,}645 \cdot \frac{50}{29}\right)^2 = 16{,}9 \approx 17$. Следовательно, желаемая точность оценки с вероятностью $\geq$ 0,95 достигается при объеме выборки $n \geq 17$.►

Оценка среднего при неизвестной дисперсии

Необходимый объем выборки определяется из соотношения

$$\frac{\varepsilon}{s} = \frac{t_\alpha(n)}{\sqrt{n}} ,$$

где $t_\alpha(n)$ – α -квантиль распределения Стьюдента при $v = n$ степенях свободы; s оценка стандартного отклонения.

Пример 2.2. Определить необходимый объем выборки для оценки среднего значения с предельной абсолютной ошибкой $\varepsilon = 0{,}8$ при доверительной вероятности $\alpha = 0{,}95$, если предполагаемое значение стандартного отклонения равно $s = 2$.

Решение. Имеем $\frac{t_{0,95}}{\sqrt{n}} = \varepsilon / s = 0{,}4$. Тогда из табл. 3 (см. приложение 2.2) для $\alpha = 0{,}95$ непосредственно находим $n = 26$.►

2.4.2. Экспоненциальное распределение

Предположим, что в течение некоторого времени t_u испытывается n приборов и при испытаниях обнаруживается r отказов. Необходимо определить значения

n и r, обеспечивающие оценку интенсивности отказов λ_0 с заданной относительной предельной ошибкой δ при доверительной вероятности α.

При испытаниях невосстанавливаемых приборов требуемый объем выборки равен

$$n = \frac{r}{\lambda_0 t_u a(r,\alpha)}.$$

Значения коэффициента $a(r,\alpha)$ приведены в табл. 5 (см. табл. 5, приложение 2.3).

Значения r находятся из соотношения $b(r,\alpha) = \frac{1}{1+\delta}$, где $b(r,\alpha)$ – коэффициент, зависящий от r и α (см. табл. 6, статистические таблицы, приложение 2.2.). По заданным α и δ сначала определяем $b(r,\alpha)$, затем по заданному значению α и вычисленному $b(r,\alpha)$ из табл. 6 находим r. Далее, для найденного значения r и заданного α по табл. 5 определяем значение $a(r,\alpha)$, и по заданному t_u и λ_0 вычисляем требуемый объем выборки n.

Пример 2.3. Найти требуемый объем испытаний для оценки интенсивности отказов невосстанавливаемого прибора, если заданы время испытаний $t_u = 1000$ ч., предельная относительная ошибка $\delta = 0,2$, предполагаемое значение интенсивности отказов $\lambda_0 = 10^{-3}$, доверительная вероятность $\alpha = 0,95$.

Решение. Находим $\frac{1}{1+\delta} = \frac{1}{1+0,2} = 0,833$. Из табл. 6 (см. приложение 2.1) для $b(r,\alpha) = 0,833$ и $\alpha = 0,95$ находим $r = 80$. Из табл. 5 для $r = 80$ и $\alpha = 0,95$ находим $a(r,\alpha) = 0,84$. Тогда искомый объем выборки $n = \frac{80}{10^{-3} \cdot 1000 \cdot 0,84} = 95$. ►

2.4.3 Биномиальное распределение

Предположим, что задано некоторое значение параметра биномиального распределения – p_0. Тогда, наименьший объем выборки, необходимый для того, чтобы подтвердить с вероятностью α, что $p \le p_0$ равен

$$n = \frac{\ln(1-\alpha)}{\ln(1-p_0)}.$$

Если среди n испытанных приборов не будет ни одного отказа, то с вероятностью α можно утверждать, что $p \le p_0$.

Пример 2.4. Найти объем выборки, позволяющий с достоверностью $\alpha = 0,90$ установить, что доля дефектных изделий в партии не превышает заданную величину $p_0 = 0,05$.

Решение. Имеем $n = \frac{\ln(1-\alpha)}{\ln(1-p_0)} = \frac{\ln 0,1}{\ln 0,95} = 45$. ►

2.5 Интервальные оценки параметров распределений

2.5.1. Оценка параметров нормального распределения

Пусть случайная величина x имеет нормальное распределение $N(\mu,\sigma^2)$, где $\sigma > 0$, $\mu \in R$ с плотностью $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $x \in (-\infty,\infty)$.

Оценка μ при известной дисперсии σ^2

Интервальные оценки с доверительной вероятностью α имеют вид:

$$\mu^{н}(\alpha) = \bar{x} - u_\gamma \frac{\sigma}{\sqrt{n}}, \quad \mu^{в}(\alpha) = \bar{x} + u_\gamma \frac{\sigma}{\sqrt{n}},$$

где u_γ – γ-квантиль стандартного нормального распределения; $\gamma = \dfrac{1+\alpha}{2}$ для двусторонней оценки, $\gamma = \alpha$ для односторонней оценки.

В табл. 1 (см. статистические таблицы) приведены квантили стандартного нормального распределения.

Для практического применения для уровней достоверности $\alpha = 0{,}9;\ 0{,}95;\ 0{,}99$ значения u_γ приведены в табл. 1

Таблица 1

α	Односторонние границы				Двусторонние границы			
	γ'	γ''	$u_{\gamma'}$	$u_{\gamma''}$	γ'	γ''	$u_{\gamma'}$	$u_{\gamma''}$
0,90	0,90	0,10	1,281	-1,281	0,950	0,050	1.645	-1.645
0,95	0,95	0,05	1,645	-1,645	0,975	0,025	1,960	-1,960
0,99	0,99	0,01	2,326	-2,326	0,995	0,005	2.576	- 2.576

Оценка μ при неизвестной дисперсии

В этом случае интервальная оценка с доверительной вероятностью α имеет вид:

$$\mu^{н}(\alpha) = \bar{x} - t_\gamma \frac{s}{\sqrt{n}}, \quad \mu^{в}(\alpha) = \bar{x} + t_\gamma \frac{s}{\sqrt{n}},$$

где t_γ – γ-квантиль распределения Стьюдента с $v = n-1$ степенями свободы; $\gamma = \dfrac{1+\alpha}{2}$ для двусторонней оценки, $\gamma = \alpha$ для односторонней оценки.

В табл. 2 (см. статистические таблицы) приведены критические точки распределения Стьюдента.

Оценка дисперсии σ^2

Интервальные оценки при доверительной вероятностью α равны

$$(s^2)^н = \frac{1}{\chi^2_{\gamma'}} \sum_{i=1}^{n} (x_i - \bar{x})^2, \quad (s^2)^в = \frac{1}{\chi^2_{\gamma''}} \sum_{i=1}^{n} (x_i - \bar{x})^2 ,$$

где χ^2_γ – γ-квантиль распределения χ^2 с $v = n-1$ степенями свободы (если параметр μ известен, $v = n$); $\gamma' = \frac{1+\alpha}{2}$ для двусторонней оценки и $\gamma' = \alpha$ для односторонней оценки; $\gamma'' = \frac{1-\alpha}{2}$ для двусторонней оценки и $\gamma'' = 1-\alpha$ для односторонней оценки.

Интервальная оценка σ может быть рассчитана также по формулам

$$s^н = \sqrt{\frac{n-1}{\chi^2_{\gamma'}}} \cdot s, \quad s^в = \sqrt{\frac{n-1}{\chi^2_{\gamma''}}} \cdot s ,$$

где $s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2}$.

Критические точки распределения χ^2 приведены в работе [2] на стр. 312, а также в таблице 4 (см. статистические таблицы).

2.5.2. Оценка параметров показательного распределения

Пусть $X \in \Pi_\lambda$, где Π_λ – показательный закон распределения с параметром λ с плотностью распределения $f(x) = \lambda e^{-\lambda x}, \ x \in [0, \infty)$.

Интервальная оценка параметра λ при доверительной вероятности α рассчитывается по формулам

$$\lambda^н = \frac{\chi^2_{\gamma'}}{2\sum_{i=1}^{n} x_i}, \quad \lambda^в = \frac{\chi^2_{\gamma''}}{2\sum_{i=1}^{n} x_i} ,$$

где χ^2_γ – γ-квантиль распределения хи-квадрат с $v = 2n$ степенями свободы;

$\gamma' = \frac{1+\alpha}{2}, \quad \gamma'' = \frac{1-\alpha}{2}$ для двусторонней оценки и $\gamma' = \alpha, \quad \gamma'' = 1-\alpha$ для односторонней оценки.

2.5.3. Оценка параметров биномиального распределения

Интервальные оценки параметра p с доверительной вероятностью α являются решениями уравнений Клоппера-Пирсона

$$\sum_{x=m}^{n} C_n^x p_н^x (1-p_н)^{(n-x)} = \frac{1-\alpha}{2}, \quad \sum_{x=0}^{m} C_n^x p_в^x (1-p_в)^{(n-x)} = \frac{\alpha}{2} .$$

На практике широко используется аппроксимация нормальным распределением [1]. В этом случае нижняя $p_н$ и верхняя $p_в$ границы равны

$$p_н = \frac{x + \frac{1}{2}u_\gamma^2 - u_\gamma \sqrt{\frac{x}{n}(n-x) + \frac{1}{4}u_\gamma^2}}{n + u_\gamma^2}, \quad p_в = \frac{x + \frac{1}{2}u_\gamma^2 + u_\gamma \sqrt{\frac{x}{n}(n-x) + \frac{1}{4}u_\gamma^2}}{n + u_\gamma^2},$$

где u_γ – γ-квантиль стандартного нормального распределения; $\gamma = \frac{1+\alpha}{2}$ для двусторонней оценки, $\gamma = \alpha$ для односторонней оценки; $x = [m_1]$ – целая часть среднего $m_1 = \frac{1}{n}\sum_{j=1}^{n} x_j$.

Эта аппроксимация рекомендуется при $x \geq 4$ и $n - x \geq 4$.
Интервальные оценки параметра n равны

$$n_н = \left[\frac{m_1}{p_в}\right], \; n_в = \left[\frac{m_1}{p_н}\right],$$

где $[\cdot]$– целая часть числа.

Примеры интервальных оценок

Пример 2.5. Требуется определить, какое количество книг N по некоторой теме должен иметь продавец, чтобы удовлетворить по возможности всех покупателей, если за четыре дня по этой теме было продано: 18, 12, 13, 15 книг.
Решение. На основании этих данных находим среднее и дисперсию

$$\bar{x} = \frac{18+12+13+15}{4} = 14,5;$$

$$s^2 = \frac{1}{3}\Big((18-14,4)^2 + (12-14,4)^2 + (13-14,4)^2 + (15-14,4)^2\Big) = 7,0;$$

$$s = \sqrt{7} = 2,65; \quad v = n - 1 = 3.$$

Примем доверительную вероятность $\alpha = 0,95$. Тогда для односторонней критической области имеем $\gamma = \alpha$, $q = (1-\alpha) = 0,05$, $2q = 0,1$ и по таблице распределения Стьюдента получим $t_{0,1} = 2,35$. Верхняя граница математического ожидания равна $\mu^в = 14,5 + \frac{2,65}{\sqrt{4}} 2,35 = 17,61$.

Следовательно, максимальное возможное количество книг, которое необходимо иметь продавцу, $N = 18$.►

Пример 2.6. В канцелярии офиса работают три секретаря. Время подготовки одного документа каждым секретарем в среднем составляет 16,3; 15,5 и 17,2 мин. Требуется оценить ориентировочное время и возможное отклонение во времени оформления документа, сданного в канцелярию.

Решение. Рассчитываем выборочное среднее значение и выборочную дисперсию $\overline{x}=16,33;\quad s^2=0,72$.

Приняв доверительную вероятность $\alpha=0,9$, получим

$\gamma'=\dfrac{1+\alpha}{2}=0,95;\quad \gamma''=\dfrac{1-\alpha}{2}=0,05$. При $v=2$ по таблицам χ^2-распределения, находим: $\chi^2_{0,95}=5,99;\quad \chi^2_{0,05}=0,103$. Тогда двусторонняя доверительная оценка дисперсии σ^2 равна:

$(s^2)^н=\dfrac{1}{\chi^2_{\gamma'}}\sum\limits_{i=1}^{n}(x_i-\overline{x})^2=\dfrac{2\cdot 0,72}{5,99}=0,24$; $(s^2)^в=\dfrac{1}{\chi^2_{\gamma''}}\sum\limits_{i=1}^{n}(x_i-\overline{x})^2=\dfrac{2\cdot 0,72}{0,103}=14,1$.

В результате получим $0,24\le\sigma^2\le 14,1$. После извлечения квадратного корня $0,49\le\sigma\le 3,61$.►

2.6 Примеры выполнения задания

2.6.1. Показательное распределение

Наблюдения интервалов времени между поступающими заявками в систему массового обслуживания заданы следующей выборкой (данные наблюдений приведены в условных единицах времени)

12,35	3,72	2,96	4,57	1,01	7,08	10,24	1,66	3,87	20,32
0,89	3,81	0,95	7,62	23,47	2,56	19,15	0,10	8,38	0,67
15,56	0,30	4,83	10,32	13,91	12,25	11,34	13,68	9,42	3,12
11,49	4,11	41,99	15,87	18,78	6,99	17,25	4,27	16,06	5,78
15,69	10,35	11,22	4,24	36,67	22,37	0,22	8,61	16,05	18,06
0,21	20,82	12,36	10,04	2,36	0,44	7,36	10,99	6,36	22,54
28,67	18,77	5,95	17,91	2,59	7,76	1,92	21,88	5,54	11,59
0,71	8,41	2,18	2,50	0,65	4,67	11,17	4,76	0,60	3,39
3,88	39,28	5,70	1,46	3,25	3,57	17,85	1,18	1,34	9,14
25,64	4,07	8,95	25,71	4,94	15,65	1,25	3,01	18,10	6,52

Известно, что интервалы времени между поступающими заявками подчиняются показательному закону. Выполнить следующие расчёты.

1) Рассчитать точечную оценку интенсивности поступления заявок λ^*

   $\lambda^*=\dfrac{1}{\overline{x}}=\dfrac{1}{9,718}=0,103$, т.е. в среднем в единицу времени (условную единицу времени) поступает 0,103 заявки, или другими словами, за 9,718 условных единиц времени поступает одна заявка.

2) Рассчитать интервальные оценки параметра λ при доверительной вероятности $\alpha=0,95$.

Расчёт проводим по формулам

$$\lambda^{н} = \frac{\chi^2_{\gamma'}}{2\sum_{i=1}^{n} x_i}, \quad \lambda^{в} = \frac{\chi^2_{\gamma''}}{2\sum_{i=1}^{n} x_i}.$$

Сумма выборочных данных равна $\sum_{i=1}^{n} x_i$ =972. Значения $\gamma' = 0,025$, $\gamma'' = 0,975$, количество степеней свободы $v = 2n = 200$. Квантили χ^2- распределения вычисляем с помощью пакета Excel: $\chi^2_{\gamma'}$=163, $\chi^2_{\gamma''}$=241. Подставим найденные значения в формулы для нижней и верхней границы доверительного интервала, получим: $\lambda^{н}$=0,084; $\lambda^{в}$=0,124.

3) Предположим, что в течение некоторого времени $t_u = 100$ усл. ед. времени проверяется регистрация n заявок и при испытаниях обнаруживается r незарегистрированных заявок. Необходимо определить значения n и r, обеспечивающие оценку предполагаемой интенсивности незарегистрированных заявок $\lambda_0 = 0,01$ с заданной относительной предельной ошибкой $\delta = 0,1$ при доверительной вероятности $\alpha = 0,95$.

Расчёт.

$\frac{1}{1+\delta} = \frac{1}{1+0,1} = 0,909$. Из табл. 6 (см. приложение 2.3)для $b(r,\alpha) = 0,909$ и $\alpha = 0,95$ находим $r = 300$. Из табл. 5 (см. приложение 2.3) для $r = 300$ и $\alpha = 0,95$ находим $a(r,\alpha) = 0,91$. Тогда искомый объем выборки $n = \frac{300}{10^{-2} \cdot 100 \cdot 0,91} = 330$.

2.6.2. Биномиальное распределение

По наблюдениям за посетителями магазина в различные дни была получена следующая выборка данных о количестве посетителей, купивших товары (т.е. покупателей)

14	15	16	11	13	13	17	9	13	15
12	14	16	14	16	16	14	17	12	12
12	12	18	15	11	17	18	13	14	15
15	18	15	16	11	14	13	15	16	15
14	14	10	15	16	15	11	14	13	14
15	15	15	15	12	13	17	14	16	10
16	13	15	18	15	14	13	14	14	13
16	16	15	14	16	14	11	15	16	16
14	15	15	15	15	15	15	15	16	13

14 16 11 11 16 16 11 15 15 14

Известно, что данные таких наблюдений подчиняются биномиальному закону. Выполнить следующие расчёты.

1) Рассчитать точечные оценки параметров p и n. Среднее и дисперсия выборки равны $m_1 = \frac{1}{n}\sum_{j=1}^{n} x_j$ = 14,71; $s^2 = \frac{1}{n-1}\sum_{j=1}^{n}\left(x_j - m_1\right)^2$ = 3,57.

   Оценки параметров n^*, p^* вычисляем по формулам

$$p^* = 1 - \frac{s^2}{m_1} = 0,75;\; n^* = \left[\frac{m_1}{p}\right] = 20.$$

2) Вычислить интервальные оценки параметров распределения для доверительной вероятности α =0,9.

Значение параметра $\gamma = \frac{1+\alpha}{2}$ =0,95. Квантиль стандартного нормального распределения u_γ вычисляем с помощью пакета Excel. Имеем u_γ =1,64.

Выполняем расчеты границ доверительного интервала для параметра p по формулам

$$p_н = \frac{[x] + \frac{1}{2}u_\gamma^2 - u_\gamma\sqrt{\frac{[x]}{n}(n-[x]) + \frac{1}{4}u_\gamma^2}}{n + u_\gamma^2} = 0,58;$$

$$p_в = \frac{[x] + \frac{1}{2}u_\gamma^2 + u_\gamma\sqrt{\frac{[x]}{n}(n-[x]) + \frac{1}{4}u_\gamma^2}}{n + u_\gamma^2} = 0,90.$$

Здесь $[x] = [m_1] = 14$.

Выполняем расчеты границ доверительного интервала для параметра n по формулам

$$n_н = \left[\frac{m_1}{p_в}\right] = 16;\; n_в = \left[\frac{m_1}{p_н}\right] = 25.$$

3) Найти объем выборки, позволяющий с достоверностью $\alpha = 0,90$ установить, что доля посетителей магазина, не ставших покупателями (ничего не купивших в магазине) не превышает заданную величину p_0 = 0,07.

Решение. Имеем $n = \dfrac{\ln(1-\alpha)}{\ln(1-p_0)} = \dfrac{\ln 0{,}1}{\ln 0{,}93}$ = 32.

2.6.3. Нормальное распределение

Измерения срока службы 100 деталей с точностью до одной десятой дали следующие результаты:

5.4	9.6	9.5	10.5	12.3	13.1	12.5	12.1	8.1	8.4
11.6	8.3	10	12.3	9.9	5.7	8.8	10.8	9.6	10.6
10.3	13.6	7.3	12.8	10.3	10.2	10.3	9.3	11.2	9.3
9.5	5.7	13.6	7.3	7.4	10.5	6.3	13	4.4	11.1
11.1	9.4	11.3	11.1	10.8	11.1	9	10.5	10	10.8
10.9	9	9.6	8.7	7	12.9	9.1	13.4	12.7	7.3
6.3	9.4	10	7	12.4	11.2	8.9	12.9	10.9	12.1
7.6	7.8	11	10.8	6.6	7.4	10.2	8.5	7.9	12
11.3	10.6	6.7	10.5	11	12.1	10	8.9	11.9	11
10.9	10.7	11.9	5.7	10.2	7.2	10.1	9.1	8.4	8

Известно, эти данные подчиняются нормальному распределению.

Выполнить следующие расчёты.

1) Рассчитать точечные оценки параметров μ и σ.

$$\mu^* = \frac{1}{n}\sum_{j=1}^{n} x_j = 9{,}9;\quad \sigma^* = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}\left(x_j - m_1\right)^2} = 2{,}1.$$

2) Вычислить интервальные оценки параметра μ и σ с доверительной вероятностью $\alpha = 0{,}9$.

Интервальная оценка параметра μ.

$$\mu^{н}(\alpha) = \overline{x} - t_\gamma \frac{s}{\sqrt{n}},\quad \mu^{в}(\alpha) = \overline{x} + t_\gamma \frac{s}{\sqrt{n}},$$

где t_γ – γ-квантиль распределения Стьюдента с $v = n-1$ степенями свободы; $\gamma = \dfrac{1+\alpha}{2}$ для двусторонней оценки. Значение параметра $\gamma = 0{,}95$; γ-квантиль распределения Стьюдента с $v = n-1 = 99$ вычисляем с помощью пакета Excel t_γ = 1,98. В результате получим

$$\mu^{н}(\alpha) = 9{,}9 - 1{,}98\frac{2{,}1}{\sqrt{100}} = 9{,}4;\ \mu^{н}(\alpha) = 9{,}9 + 1{,}98\frac{2{,}1}{\sqrt{100}} = 10{,}3.$$

Интервальная оценка параметра σ .

$$s^{н}=\sqrt{\frac{n-1}{\chi^2_{\gamma'}}}\cdot s,\quad s^{в}=\sqrt{\frac{n-1}{\chi^2_{\gamma''}}}\cdot s\,.$$

Значения $\gamma'=0,95$, $\gamma''=0,05$. Квантили хи-квадрат распределения вычисляем с помощью Excel: $\chi^2_{0,95}$=123,22; $\chi^2_{0,05}$=77,05.

В результате получим следующие значения границ доверительного интервала

$$s^{н}=\sqrt{\frac{99}{123,22}}\cdot 2,1=1,6;\quad s^{в}=\sqrt{\frac{99}{77,05}}\cdot 2,1=2,3\,.$$

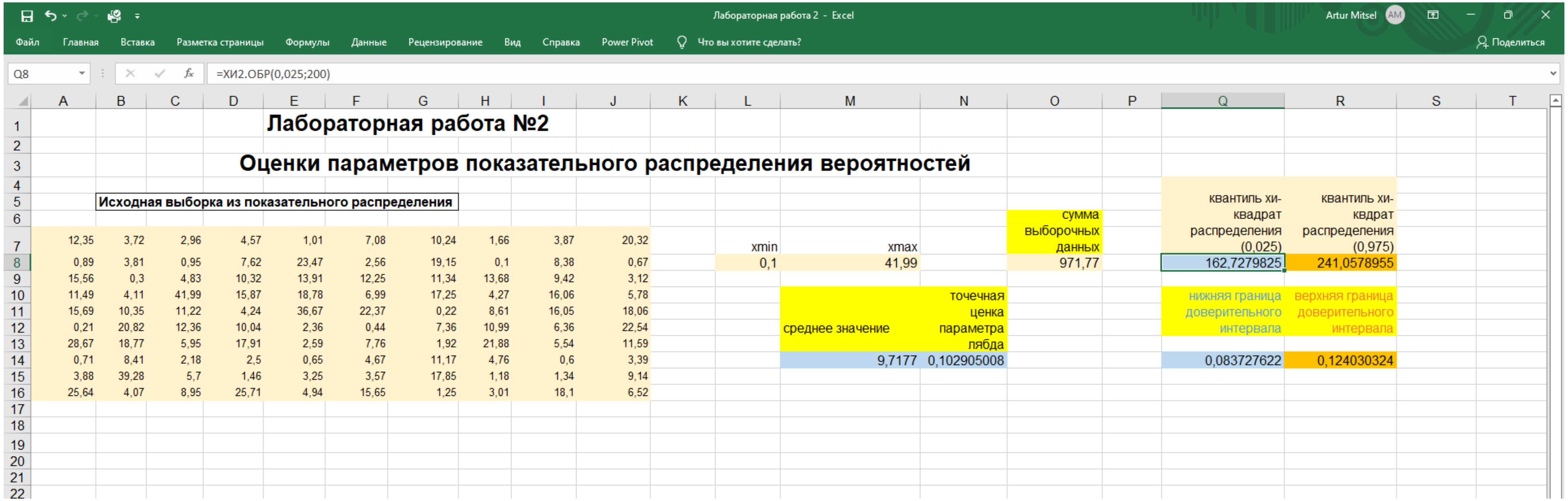
3) Найти объем выборки, необходимый для оценки среднего μ с заданной предельной относительно погрешностью $\delta=0,1$ и доверительной вероятностью $\alpha=0,9$.

Решение. Значение $s=2,1$, $\bar{x}=9,9$. Абсолютная погрешность среднего равна $\varepsilon=0,99$.

Имеем $\dfrac{t_{0,9}}{\sqrt{n}}=\varepsilon/s=0,471$. Тогда из табл. 3 (см. приложение 2.2) для $\alpha=0,90$ непосредственно находим $n=14$.

Ниже приведены скриншоты примеров выполнения заданий

2.7 Примеры выполнения заданий лабораторной работы №2


Лабораторная работа 2 - Excel

Q8 =ХИ2.ОБР(0,025;200)

Лабораторная работа №2

Оценки параметров показательного распределения вероятностей

Исходная выборка из показательного распределения

A	B	C	D	E	F	G	H	I	J
12,35	3,72	2,96	4,57	1,01	7,08	10,24	1,66	3,87	20,32
0,89	3,81	0,95	7,62	23,47	2,56	19,15	0,1	8,38	0,67
15,56	0,3	4,83	10,32	13,91	12,25	11,34	13,68	9,42	3,12
11,49	4,11	41,99	15,87	18,78	6,99	17,25	4,27	16,06	5,78
15,69	10,35	11,22	4,24	36,67	22,37	0,22	8,61	16,05	18,06
0,21	20,82	12,36	10,04	2,36	0,44	7,36	10,99	6,36	22,54
28,67	18,77	5,95	17,91	2,59	7,76	1,92	21,88	5,54	11,59
0,71	8,41	2,18	2,5	0,65	4,67	11,17	4,76	0,6	3,39
3,88	39,28	5,7	1,46	3,25	3,57	17,85	1,18	1,34	9,14
25,64	4,07	8,95	25,71	4,94	15,65	1,25	3,01	18,1	6,52

xmin	xmax	сумма выборочных данных	квантиль хи-квадрат распределения (0,025)	квантиль хи-квадрат распределения (0,975)
0,1	41,99	971,77	162,7279825	241,0578955

среднее значение	точечная оценка параметра лябда	нижняя граница доверительного интервала	верхняя граница доверительного интервала
9,7177	0,102905008	0,083727622	0,124030324

Лабораторная работа 2 - Excel

P37

Лабораторная работа №2

Оценки параметров биномиального распределения вероятностей

Исходная выборка из биномиального распределения

A	B	C	D	E	F	G	H	I	J
14	15	16	11	13	13	17	9	13	15
12	14	16	14	16	16	14	17	12	12
12	12	18	15	11	17	18	13	14	15
15	18	15	16	11	14	13	15	16	15
14	14	10	15	16	15	11	14	13	14
15	15	15	15	12	13	17	14	16	10
16	13	15	18	15	14	13	14	14	13
16	16	15	14	16	14	11	15	16	16
14	15	15	15	15	15	15	15	16	13
14	16	11	11	16	16	11	15	15	14

xmin	xmax	сумма выборочн
9	18	1431

среднее значение	дисперсия
14,31	3,56959596
14	

точечная оценка параметра p	точечная оценка параметра n
0,75	20

параметр p0	доверительная вероятность
0,07	0,9

наименьший объём выборки для p0 и a: 32

квантиль нормального распределения для гамма (доверительная вероятность 0.95): 1,644853627

нижняя граница доверительного интервала параметра p	верхняя граница доверительного интервала параметра p
0,58	0,9

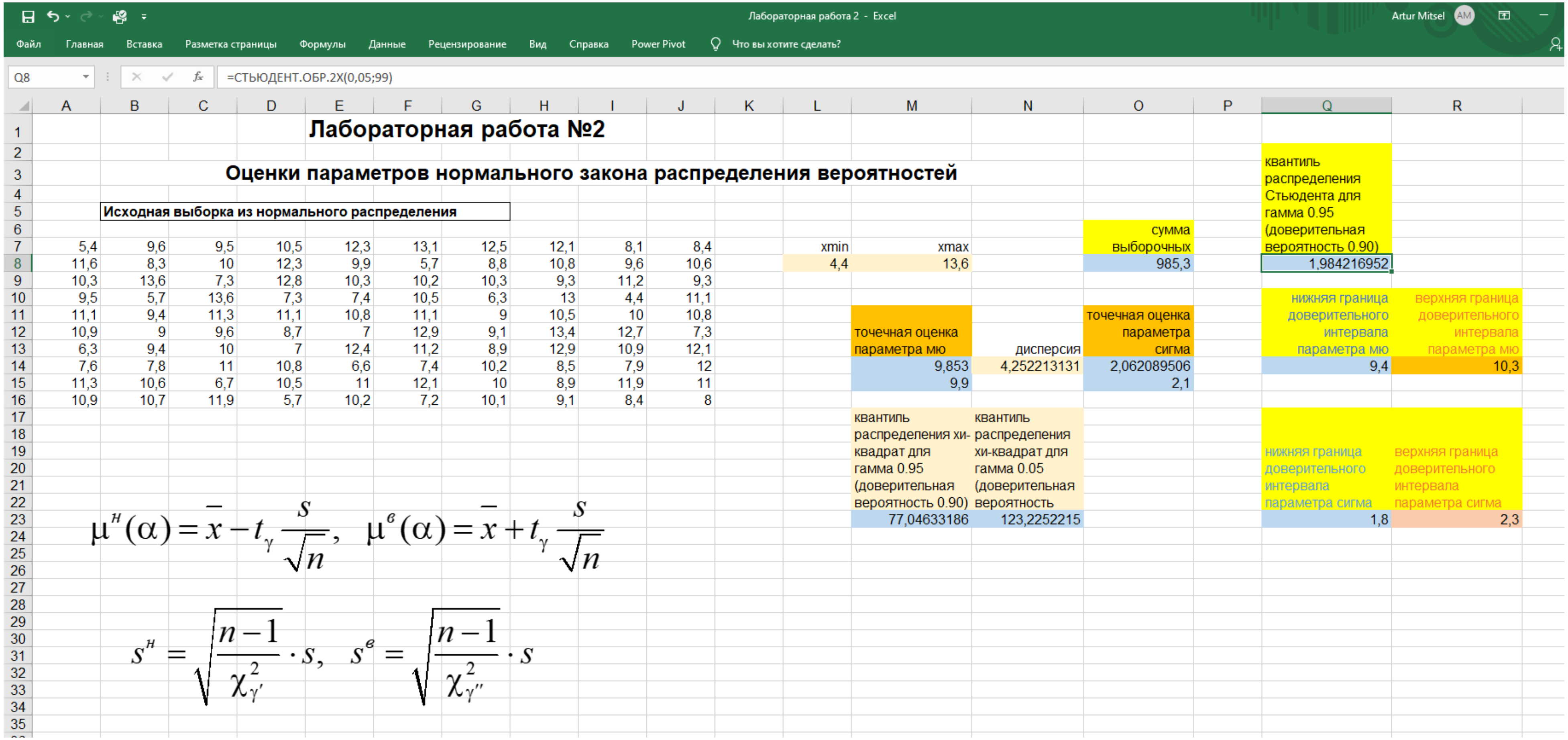
нижняя граница доверительного интервала параметра n	верхняя граница доверительного интервала параметра n
16	25

$$p_н = \frac{x + \frac{1}{2}u_\gamma^2 - u_\gamma\sqrt{\frac{x}{n}(n-x) + \frac{1}{4}u_\gamma^2}}{n + u_\gamma^2}$$

$$x = [m_1] = \left[\frac{1}{n}\sum_{j=1}^{n} x_j\right]$$

$$p_в = \frac{x + \frac{1}{2}u_\gamma^2 + u_\gamma\sqrt{\frac{x}{n}(n-x) + \frac{1}{4}u_\gamma^2}}{n + u_\gamma^2}$$

$$n = \frac{\ln(1-\alpha)}{\ln(1-p_0)}$$



Лабораторная работа 2 - Excel

Q8 =СТЬЮДЕНТ.ОБР.2Х(0,05;99)

Лабораторная работа №2

Оценки параметров нормального закона распределения вероятностей

Исходная выборка из нормального распределения

A	B	C	D	E	F	G	H	I	J
5,4	9,6	9,5	10,5	12,3	13,1	12,5	12,1	8,1	8,4
11,6	8,3	10	12,3	9,9	5,7	8,8	10,8	9,6	10,6
10,3	13,6	7,3	12,8	10,3	10,2	10,3	9,3	11,2	9,3
9,5	5,7	13,6	7,3	7,4	10,5	6,3	13	4,4	11,1
11,1	9,4	11,3	11,1	10,8	11,1	9	10,5	10	10,8
10,9	9	9,6	8,7	7	12,9	9,1	13,4	12,7	7,3
6,3	9,4	10	7	12,4	11,2	8,9	12,9	10,9	12,1
7,6	7,8	11	10,8	6,6	7,4	10,2	8,5	7,9	12
11,3	10,6	6,7	10,5	11	12,1	10	8,9	11,9	11
10,9	10,7	11,9	5,7	10,2	7,2	10,1	9,1	8,4	8

xmin: 4,4; xmax: 13,6

сумма выборочных: 985,3

квантиль распределения Стьюдента для гамма 0.95 (доверительная вероятность 0.90): 1,984216952

точечная оценка параметра мю: 9,853; 9,9

дисперсия: 4,252213131

точечная оценка параметра сигма: 2,062089506; 2,1

нижняя граница доверительного интервала параметра мю: 9,4

верхняя граница доверительного интервала параметра мю: 10,3

квантиль распределения хи-квадрат для гамма 0.95 (доверительная вероятность 0.90): 77,04633186

квантиль распределения хи-квадрат для гамма 0.05 (доверительная вероятность: 123,2252215

нижняя граница доверительного интервала параметра сигма: 1,8

верхняя граница доверительного интервала параметра сигма: 2,3

$$\mu^{н}(\alpha)=\bar{x}-t_{\gamma}\frac{s}{\sqrt{n}},\quad \mu^{в}(\alpha)=\bar{x}+t_{\gamma}\frac{s}{\sqrt{n}}$$

$$s^{н}=\sqrt{\frac{n-1}{\chi^{2}_{\gamma'}}}\cdot s,\quad s^{в}=\sqrt{\frac{n-1}{\chi^{2}_{\gamma''}}}\cdot s$$

2.8 Практическое задание

Имеется выборка объемом $n = 100$ из известного распределения F (см. приложение 2.1). При этом F может быть одним из следующих распределений:

1) F - показательное распределение с плотностью $f(x) = \lambda e^{-\lambda x}$, $x > 0$, где параметр $\lambda > 0$ - неизвестен.

2) F - биномиальное распределение с плотностью

$$f(x;n,p) = \binom{n}{x} \cdot p^x \cdot q^{n-x},\ k = 0,1,...,n,$$

где $\binom{n}{x} = \dfrac{n!}{(n-x)!x!}$ — биномиальный коэффициент.

3) F - нормальное распределение с плотностью $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $x \in R$, где параметры μ и $\sigma > 0$ - неизвестны.

Требуется вычислить следующие характеристики:

1. Точечные оценки параметров распределений.
2. Интервальные оценки параметров распределения для доверительной вероятности α =0,9; α =0,95; α =0,99.

3.

3.1. В предположении, что в течение некоторого времени $t_u = 100$ усл. ед. времени проверяется n событий и при испытаниях обнаруживается r незарегистрированных событий, необходимо определить значения n и r, обеспечивающие оценку предполагаемой интенсивности незарегистрированных событий $\lambda_0 = 0{,}01$ с заданной относительной предельной ошибкой $\delta = 0{,}1$, $\delta = 0{,}2$ и $\delta = 0{,}4$ при доверительной вероятности $\alpha = 0{,}90; \alpha = 0{,}95\ \alpha = 0{,}99$. Привести примеры использования показательного распределения.

3.2. Для биномиального распределения найти объем выборки, позволяющий с достоверностью $\alpha = 0{,}90; \alpha = 0{,}95\ \alpha = 0{,}99$ установить, что доля незарегистрированных событий не превышает заданную величину $p_0 = 0{,}03$; $p_0 = 0{,}05$; $p_0 = 0{,}07$. Привести примеры использования биномиального распределения.

3.3. Для нормального распределения найти объем выборки, необходимый для оценки среднего μ с заданной предельной абсолютной погрешностью $\delta = 0{,}05$; $\delta = 0{,}1$; $\delta = 0{,}15$ и доверительной вероятностью $\alpha = 0{,}90$; $\alpha = 0{,}95$ $\alpha = 0{,}99$. Привести примеры использования нормального распределения.

Приложение 2.1
Варианты заданий по лабораторной работе №2

Вариант 1.

Показательное распределение $f(x)=\lambda e^{-\lambda x},\ x\in[0,\infty)$ с параметром $\lambda=0,05$

2,5	80,6	2,2	19,6	27,4	34,1	8,8	30,5	57,7	57,2
18,3	40,7	6	5	4,5	9,2	56,3	7,9	11,1	4,3
22,7	8,7	52	0,4	0,4	14	2,1	3	2,5	22,6
24,2	13,5	1,7	6,7	25,4	43,3	20	4,3	79,1	1,7
6,4	1,5	11,9	29,1	24,2	0,8	22,7	1,5	22,2	50
10,4	21,6	1,1	69,7	26,2	1,2	6,3	11,3	43,7	5,7
41,5	24,9	0,5	3,4	10,9	17,3	0,3	3,1	10,9	4,4
25,2	1	17,1	8,5	24,6	49,7	66,2	22,2	46,7	72,4
12,8	6,9	1,5	76,1	2,9	21,5	23,4	14,5	20,2	58
2,8	8,1	21	37	19,4	7,7	7	2,6	66,4	5

Биномиальное распределение $f(x;n,p)=\binom{n}{x}\cdot p^x\cdot q^{n-x},\ k=0,1,...,n$ с параметрами $n=20;\ p=0,4$

9	9	6	6	9	9	6	9	8	4
7	4	10	9	8	8	11	5	9	6
9	4	10	6	9	5	9	7	7	10
4	10	8	6	11	6	12	11	8	10
6	12	8	6	7	8	8	9	3	9
8	9	8	7	10	6	11	7	9	6
10	12	6	6	9	5	8	10	7	5
8	6	11	6	8	6	9	9	8	7
8	11	7	16	7	10	7	6	6	11
7	6	10	11	4	11	8	4	11	9

Нормальное распределение $f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}},\ x\in(-\infty,\infty)$ с параметрами $\mu=5,\ \sigma=1$

5,1	4,07	4,32	5,45	4,4	6,43	6,07	4,31	4,97	3,76
3,21	6,14	3,19	4,82	4,83	6,14	5,41	3,75	5,11	5,82
3,78	3,93	3,38	7,55	5,25	5,35	3,03	5,51	5,49	5,48
4,54	4,64	4,55	6,27	4,08	5,3	4,12	3,58	4,43	5,18
6,3	3,48	4,71	5,51	4,82	4,5	5,49	4,13	3,95	4,06
3,53	3,2	6,16	5,06	4,86	4,92	4,9	5,32	4,15	4,43

3,33	6,05	3,97	5,41	5,8	4,51	6,12	5,74	3,12	5,94
5,82	3,26	5,56	4,68	3,57	4,43	3,85	4,93	5,06	4,94
5,35	5,19	5,15	5,02	5,3	7,15	5,26	5,38	3,9	4,51
5,81	3,13	5,44	3,37	4,84	6,36	4,75	4,29	3,73	3,04

Вариант 2.

Показательное распределение $f(x) = \lambda e^{-\lambda x}, \; x \in [0, \infty)$ с параметром $\lambda = 0{,}1$

11,4	10,2	40,5	14,1	1	5,1	0,7	1,5	16,9	3,4
3,7	11	8,2	12,5	1,6	6,2	2,5	11,9	44,6	6,5
21,1	10,7	1,2	11,9	9,7	16,3	30,4	1,9	11,3	10,1
26,1	5,4	16,3	11,4	9	14,7	4,6	0,8	3,9	3,5
1,6	0,5	2,4	24,6	3,7	24,4	6	1,6	6,3	11,4
20,4	4,8	5	21	22,1	19,4	3,1	10,3	8,7	5,3
33,7	12,7	0,6	1,5	4,6	5,9	26,2	30,5	1,4	3,2
74	3,1	22,6	0,4	0,1	14,5	5,8	5,3	7,3	7,7
10,7	13,5	15,4	5	29,6	19,9	13	14,6	21,2	27,2
16	0,3	10,7	24,1	26	7,6	4,2	14,1	2,9	2,8

Биномиальное распределение $f(x; n, p) = \binom{n}{x} \cdot p^x \cdot q^{n-x}, \; k = 0, 1, ..., n$ с параметрами

$n = 20; \; p = 0{,}45$

14	10	6	8	10	8	8	8	12	7
8	7	8	10	8	10	7	10	14	8
13	13	13	6	11	5	11	11	8	11
10	12	9	7	6	8	11	7	8	9
6	10	9	11	9	7	7	8	9	9
2	9	10	14	6	11	8	7	11	11
12	7	9	9	8	7	10	11	10	11
9	12	8	10	10	13	8	7	9	8
9	9	11	4	8	9	8	7	8	5
14	10	10	9	12	9	7	10	9	3

Нормальное распределение $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \; x \in (-\infty, \infty)$ с параметрами

$\mu = 10, \; \sigma = 2$

9,1	9,7	9	7,6	11,1	10,8	11,4	11,5	10,1	6,9
8,4	11,4	12,3	10,2	12,2	7,7	11	9,2	9,7	13,1
12,1	9,2	10	11,2	11,3	11,8	5,5	8,8	6,7	13,6
8,3	7	11,2	10,1	11,4	12,9	12,8	10	10,8	9,6
6,5	11,1	11,1	10,9	10,3	6,7	3,7	11,3	11	13,7

9,7	10,9	10,7	9,1	9,3	8,1	10,7	6,6	11,1	11,9
8,7	10,9	9,8	8,6	9,3	8,8	6,5	10,9	7,7	12,3
8,9	10,7	10,1	11,1	10,6	7,1	11	10,3	7,4	10,5
10,7	11	7,7	9,4	10,1	7,8	8,7	4,8	7,6	9
12,2	8,9	11,3	9,9	9,7	13,5	7,9	8,5	12,4	9,3

Вариант 3.

Показательное распределение $f(x) = \lambda e^{-\lambda x},\ x \in [0, \infty)$ с параметром $\lambda = 0{,}15$

8,6	11,3	0,8	0,8	0,1	2,1	1,5	2,7	7,9	1,2
0,8	6,9	39,2	1,8	10	3,8	15,7	7,9	7,3	5,1
8,9	10	25,2	0,3	3,4	3	8,1	6,7	7,4	6
7	8,1	1,9	19,4	0,4	28,2	0,5	9,7	16,3	0,2
6,3	8,4	3,8	20,1	5,7	0,7	0,3	0,5	10,3	5
1,2	9,4	6,9	15,6	1,5	0,2	0,1	1,7	34,3	3,3
10,9	0,6	1	2,5	1,6	4,3	5,8	14	12,5	7,4
6,7	0,3	6,5	17	0	18,5	13	6,9	17,3	12,4
1,8	10,1	0,2	0,1	1,1	1,7	21,3	0,9	1,3	9
3,2	7,1	0,2	4,1	5	6,6	17,3	2,2	9,1	2,6

Биномиальное распределение $f(x; n, p) = \binom{n}{x} \cdot p^x \cdot q^{n-x},\ k = 0, 1, \ldots, n$ с параметрами

$n = 20;\ p = 0{,}5$

13	12	7	10	13	6	11	8	11	9
10	10	9	11	10	5	10	9	11	9
13	8	10	9	11	8	13	9	11	11
10	9	12	8	10	11	9	10	12	11
7	10	8	9	12	11	12	14	9	10
12	7	7	10	9	9	13	12	10	10
9	8	12	12	11	9	8	10	10	10
14	14	9	12	10	8	10	13	10	9
12	9	11	11	3	12	9	13	6	13
11	10	8	14	10	9	8	8	9	12

Нормальное распределение $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},\ x \in (-\infty, \infty)$ с параметрами

$\mu = 15,\ \sigma = 3$

16,8	9,2	18,6	15,7	14,1	15,3	12,2	12,1	12,9	13,3
17,2	17,3	18,5	10,6	18,7	8,5	12,2	17,7	18,1	14,6
14,5	11,4	16,5	17,1	14,9	12,1	13,1	17,8	13	12,6
13,8	12,3	17,3	15,2	13,7	13,3	13,4	14,2	12	19,1
13,8	17	15,6	12,8	15,1	13	19,5	17,9	16,9	14,2

13,9	11,7	17,7	11,7	14,1	16,6	10,4	10,5	12,3	16,1
16,3	14,3	12,8	17,8	18,2	16,7	15,9	18,6	12,9	18,6
19,8	19,4	15,5	11,4	17,8	18,1	12,9	14,4	11	10,9
16,8	15,2	16,5	18,9	13,9	15,3	13,3	17,7	16	14,1
16,1	14,6	13,7	13,1	17,6	13,2	17,6	12,6	10,6	16,3

Вариант 4.

Показательное распределение $f(x) = \lambda e^{-\lambda x},\ x \in [0, \infty)$ с параметром $\lambda = 0,04$

4,2	4,7	3	7,4	4,9	17,7	4,9	12,1	10,4	1,9
0,8	1,2	6,7	1,8	6,1	0,4	0,3	8,4	0,3	4,7
0	3,3	1,5	1,5	10,9	6,9	2,7	9,8	0,6	1,7
1,3	3,7	1,1	3	6,1	10,1	0,5	3,1	1,1	0,3
9,2	0,2	3,9	0,5	5,7	2,3	0,3	10,6	4,5	7,4
2,2	2,9	5,3	1,2	6,3	12,5	9,4	15,5	14,3	4
18,5	3,3	3,9	3,3	1,2	0,2	13,8	2,9	9,7	8,6
8,4	0,7	0,1	1,3	0,7	3,7	5,3	1,7	15,4	7,9
5,7	2,7	5,5	0,1	5,7	15,8	5,5	3,4	10,6	2,9
0,7	7,2	9,6	16	15,2	1,2	6,4	9,2	3,1	1,8

Биномиальное распределение $f(x; n, p) = \binom{n}{x} \cdot p^x \cdot q^{n-x},\ k = 0, 1, \ldots, n$ с параметрами $n = 40;\ p = 0,6$

26	24	26	28	23	22	19	21	26	26
24	20	20	25	26	18	21	27	25	20
29	24	23	23	22	22	28	22	24	24
23	23	20	26	18	24	27	23	26	21
23	23	25	26	30	17	30	26	19	22
24	23	26	25	20	22	27	28	22	28
26	26	20	22	21	25	19	17	20	30
27	24	31	22	21	17	25	24	23	19
26	22	24	26	24	22	26	26	24	24
25	32	26	26	26	22	32	23	24	23

Нормальное распределение $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},\ x \in (-\infty, \infty)$ с параметрами $\mu = 6,\ \sigma = 1$

4,6	5,3	4,9	7,2	5,7	4,3	7,7	7	6,8	5,6
5,8	6,4	5,1	5,4	5,6	5,9	7,3	4,3	8	5,6
6,4	5,4	6,2	6,9	5,9	6,1	5,8	8,2	5,1	6,3
5,5	6,6	4,7	7,5	5,8	7,2	5,8	5,9	3,7	7,7
6,2	5,8	7	6,9	6,7	5,9	5,6	6	5,2	5,7

3,6	5,4	6,2	5,1	7,5	7,5	6,3	6,1	5,1	7,6
6,5	4,8	8,1	7,3	6	6	5,2	5,3	7,8	6,2
5,5	6,5	6	5,3	3,7	4,2	4,7	6,7	4,8	4,6
5,1	6,6	6,4	5,5	5,5	8	5,9	4,2	5,9	6,4
2,4	5,4	6,2	4,3	4,6	7	6	5,9	4,9	5,4

Вариант 5.

Показательное распределение $f(x)=\lambda e^{-\lambda x},\ x\in[0,\infty)$ с параметром $\lambda=0,03$

33,4	66,2	32,7	38,8	61,7	38	47,8	43,8	40	3,6
36,8	1,5	120,1	38,4	29,1	44,4	16,2	32,5	12,9	30,4
12,7	0,2	19,7	0	0,1	29,8	48,2	46,9	55,3	16
24	90,5	36,3	36,5	2,8	3	10,3	51,7	30,4	62,3
5,5	13,6	9,2	16,2	7,6	115	30,4	38,8	40,5	33,3
99,6	3,8	25,8	13,5	50,1	6,6	32,9	8,6	16,2	6,1
6,2	9	92,9	11,7	59	73,1	53,3	5	18	33,4
20,8	5,5	7,1	13,1	189,9	53,1	14,5	3,7	9,1	8,9
60,2	17	40,1	2,1	172,7	9,8	19	1,1	10,1	42,5
48,4	11,8	16,4	23,1	6,9	1	14,7	31	12	0,5

Биномиальное распределение $f(x;n,p)=\binom{n}{x}\cdot p^x\cdot q^{n-x},\ k=0,1,...,n$ с параметрами

$n=40;\ \ p=0,65$

22	25	25	24	25	23	28	27	23	31
31	29	25	24	27	20	26	24	32	27
28	30	26	31	28	26	31	25	27	32
25	24	25	29	26	25	26	29	24	22
32	26	29	20	24	22	29	27	26	26
29	24	22	27	25	28	27	25	27	29
30	22	29	29	22	29	24	33	20	29
25	27	27	26	23	29	30	24	23	28
23	28	25	33	24	27	27	25	25	26
29	25	30	26	22	28	26	25	26	25

Нормальное распределение $f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}},\ x\in(-\infty,\infty)$ с параметрами

$\mu=10,\ \sigma=1,5$

8,5	12,8	10	11,5	8,1	10,6	6,3	10,4	8,9	7,6
7,3	10,5	7,5	9,6	10,2	7,1	10,4	10,5	10,5	11,8
9,1	9,5	10,9	10,3	9	10,3	8,5	9,4	9,9	10,7
13,3	6,5	9,4	11,1	10,5	12,4	7,4	11,3	9,9	10,9
9,9	10,4	10,8	9,4	7,9	11,7	11	9,5	9,4	10,5

10,5	11,9	13	10,9	8,9	11,5	8,8	9,3	8,3	9,7
10,4	9,4	8	8,5	5,2	12,1	10	9,7	8,8	9,1
4,8	10,7	11,5	10,3	10,9	7,3	11,7	8,5	9,3	9
9,8	9,4	12,7	10,4	10,3	10,1	12,4	10,5	12,9	10,4
10,9	11,6	12,6	10,3	12,1	12	10,2	9,3	10,9	10,5

Вариант 6.

Показательное распределение $f(x) = \lambda e^{-\lambda x}, \; x \in [0, \infty)$ с параметром $\lambda = 0{,}02$

51,2	103,8	43,9	35,9	10,2	26,2	60,7	23,4	89,8	6,2
37,4	15,3	28,1	33,4	5,1	3,7	72,6	14,3	22,7	0,7
32,9	5,1	121,7	93,3	20,3	13,9	9,7	39,9	153,8	35,6
0,9	1,7	56,9	12,4	166,1	67,9	41,4	39,7	316	60,4
7,2	46,4	11,4	131,4	1,2	133	7,5	60,3	2,2	52,3
15,2	51,8	81,3	195,5	15,9	5,4	31,8	20,4	57,9	71,9
72	63,7	36,2	0,5	72,6	53,2	23,4	179,4	55,2	20,5
102,3	65,9	0,9	15	48,1	53,9	42,8	0,8	62,2	21,9
23,7	33,4	1,1	8,3	4,8	8,8	7,1	55	15	73,1
17,3	20,8	36,4	210,7	10,2	31,6	26,3	25,5	119,9	55,1

Биномиальное распределение $f(x; n, p) = \binom{n}{x} \cdot p^x \cdot q^{n-x}, \; k = 0, 1, \ldots, n$ с параметрами

$n = 40; \;\; p = 0{,}7$

31	30	26	27	27	30	26	32	28	31
26	28	30	28	26	30	29	26	34	31
24	27	25	33	31	27	32	25	26	27
33	29	29	25	28	26	29	29	28	31
32	30	29	25	27	25	28	31	29	32
31	24	28	33	32	26	31	33	27	33
30	26	31	28	22	22	24	24	27	29
26	31	24	26	30	27	30	31	29	30
26	35	26	27	27	25	28	34	33	31
29	33	27	24	27	31	33	25	30	22

Нормальное распределение $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \; x \in (-\infty, \infty)$ с параметрами

$\mu = 6, \;\; \sigma = 0{,}5$

6,5	6,2	5,6	5,7	5,1	6,3	5,5	5,5	6,6	6,7
6,4	5,9	5,7	6,4	6,6	6,3	6,2	5,8	6,8	6
7,4	6,1	6,1	6,1	5,8	6	6,4	5,7	6,1	5,8

5,8	6,9	6,6	6,2	5,6	5,6	5,6	6,2	6,1	6,1
5,7	6,2	6,5	6,7	6,2	5,5	6,6	6	6	6,6
6,2	6,3	6,5	5,2	6,9	5,9	5,5	5,4	5,8	5,6
5,7	5,4	6,8	5,9	4,9	4,9	5,7	6,3	6,2	6,3
5,8	6,9	6	5,7	5,4	6,2	5,8	5	6,2	5,4
6	5,1	6,3	4,9	6,2	4,9	6,3	6,5	5,9	6,1
5,5	6,3	5,7	5,2	5,8	5,7	6,8	6,7	5,5	6,4

Вариант 7.

Показательное распределение $f(x) = \lambda e^{-\lambda x},\ x \in [0, \infty)$ с параметром $\lambda = 0{,}01$

182,9	9,3	10,4	27,6	61,1	242,9	104,3	52,7	244,1	59,3
358,9	129,6	180,2	170,1	20,4	260,4	39	38,1	12,8	180,2
47,6	527,9	352	46,7	32,8	52,6	49,5	8,6	45,4	15,4
39,7	75,8	22,6	19,7	31,9	56,4	93,1	129,5	17,1	171,7
63,2	88,6	95,9	15,7	216	111,7	182,9	90	75,6	133,7
126,1	73,4	143,8	302	72,1	36,2	53,9	48,8	152,3	67,1
144,2	1,7	131	79,1	91,4	40,3	148,3	11	268,9	15
64,7	398,5	27,5	112,7	96,1	220,5	145,8	91,1	86,8	12,7
169,6	317,4	161,3	7,9	75,8	142	66,8	549,2	11,6	18,6
39,9	120,4	270,7	2,2	53,3	123,9	87	102,2	7,3	89,4

Биномиальное распределение $f(x; n, p) = \binom{n}{x} \cdot p^x \cdot q^{n-x},\ k = 0, 1, ..., n$ с параметрами

$n = 60;\ \ p = 0{,}75$

43	48	47	48	44	44	48	44	44	44
45	46	45	41	50	52	48	40	50	41
44	48	45	47	52	46	47	47	39	43
43	48	49	47	46	50	42	39	44	48
44	46	46	44	42	47	38	43	41	45
48	49	38	47	48	48	48	41	44	39
46	41	44	44	44	47	51	49	47	42
46	52	46	49	45	44	43	42	46	42
41	44	40	45	41	46	47	43	42	47
50	49	50	48	48	43	42	46	49	45

Нормальное распределение $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},\ x \in (-\infty, \infty)$ с параметрами

$\mu = 20,\ \sigma = 4$

15,1	16,8	16,6	17	16,9	20,1	22	23,5	20	23,5

23,8	14,1	18,5	17,4	19,2	17,4	19,8	15	19,6	22,8
15,2	23,9	17,5	23	22,1	19,7	17,1	15,8	13,3	15,6
23,3	22	24,8	20,7	20	13,1	12,8	22,1	23,9	18
29,3	13,8	22,9	20,3	16,7	28,3	21	18	24,7	19,6
15,1	19,8	19,6	22,3	20,3	18,3	22,3	14,2	18,8	16,1
21,3	19,7	22,5	21,7	24,9	13,9	25,2	29,3	17,5	17,5
13,9	28,5	21,6	22	25,1	18,5	13,7	18,7	23	18,4
18,2	21,5	19,5	21,6	24,2	16,2	24,7	16,6	15,6	19,5
16	18,8	18,3	16,7	14,1	20,6	22,4	15,5	18,8	18,5

Вариант 8.

Показательное распределение $f(x) = \lambda e^{-\lambda x},\ x \in [0, \infty)$ с параметром $\lambda = 0{,}11$

19,3	5,1	13,7	2,8	1,4	2,7	21,6	12,8	0,3	6,7
2,5	9,6	19,2	12,7	0,9	12,2	7	2,3	10,5	24,2
2,9	9,8	13,8	6	27,2	0,2	2,9	17,1	4,5	18,3
0,5	3,8	2,1	1,1	4,6	4,2	0,2	24,3	4,7	5,2
3,3	4,8	9,7	9	14,8	13,3	8,4	15,6	15,8	20,1
14,2	3,1	2,1	11	8	0	4,3	4,5	4,8	2,5
3,9	21,1	1,2	16,9	18,8	10,6	9,5	11	13,8	1,9
4,6	16,6	10,9	0,1	9,9	10,1	0,6	2,4	5,5	13,2
12,8	20,9	0,3	13,4	2,4	0,7	6,8	11,6	4,5	5,8
64,9	5,9	1,9	1,1	2,4	6,5	1,7	0,2	9,4	10,8

Биномиальное распределение $f(x; n, p) = \binom{n}{x} \cdot p^x \cdot q^{n-x},\ k = 0,1,...,n$ с параметрами

$n = 60;\ p = 0{,}80$

49	51	53	50	48	50	51	54	51	45
50	48	47	49	46	51	45	48	49	45
47	48	50	53	46	48	52	54	42	41
50	50	50	48	49	51	48	47	54	40
50	46	52	43	48	48	50	44	47	51
49	52	51	45	50	49	48	45	46	45
52	50	46	47	48	54	50	51	51	49
43	46	48	47	45	51	45	45	50	42
49	48	54	40	45	52	50	42	53	44
45	51	48	47	40	51	48	52	45	49

Нормальное распределение $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},\ x \in (-\infty, \infty)$ с параметрами

$\mu = 25,\ \sigma = 5$

20,6	27,1	12,6	20,9	18,8	27,5	21,5	14,6	14,9	29
22,6	18,7	27	6,5	22,8	19,6	24,1	26	18,5	12,7
19,2	20,8	19,7	19,6	28,4	22,1	25,7	17,9	24,2	24,7
10,4	21	10,2	15,6	17,3	18,9	21,1	19,5	15,9	22,4
13,3	26,7	16,2	23,7	22,2	24,6	28,2	10,1	17,7	21,4
18,6	13,6	19,7	16,9	17,1	15,2	11,8	17	29,5	17,6
24,6	21,6	28,7	14,3	18,3	14,8	20,7	23,7	18,3	23,7
23	21,2	19,7	25,4	22,4	20,4	23,1	17,9	15,4	17,5
20,6	22,4	13,1	24,5	16,3	20,5	17,1	20,5	17,1	28,5
19	20,5	15,6	14,4	19,9	22,8	12,8	20,9	24,5	32,4

Вариант 9.

Показательное распределение $f(x) = \lambda e^{-\lambda x},\ x \in [0, \infty)$ с параметром $\lambda = 0{,}12$

2,9	5	9,1	7,2	15,5	8	2,8	9	15,5	6
28,4	4,7	4,9	31,4	11,3	6,7	11,9	14	4	12,3
34	10,5	4,4	1,2	5,6	4,5	7,4	0,4	9,3	5,6
3,5	1,7	25,6	12,8	4,2	1,6	10	25,5	45,1	1,1
5,7	0,4	0,2	11,8	6,9	9,4	4,6	4,8	8,9	7,9
1,8	9,3	3,2	5,1	3,8	5	14,5	1,5	2	0,3
8,9	0,2	1,8	5,5	14,2	0,3	13	6,8	20,9	0,6
11	3,6	2,6	24,5	15,3	2,1	22	9,9	25,4	3,1
23,7	1,3	0,9	4,4	10,9	2	6,7	2,9	17,2	17,6
9,2	3,6	1,8	3,5	4,4	7,7	4,9	2,8	3,5	30,1

Биномиальное распределение $f(x; n, p) = \begin{pmatrix} n \\ x \end{pmatrix} \cdot p^x \cdot q^{n-x},\ k = 0,1,\ldots,n$ с параметрами

$n = 60;\ \ p = 0{,}85$

50	51	54	46	45	46	51	51	43	53
50	54	56	52	47	55	54	53	54	51
50	57	46	47	54	49	50	49	53	45
51	51	49	51	50	51	45	51	47	55
53	48	47	47	54	50	53	53	50	52
50	53	48	50	51	52	52	50	50	54
51	53	49	50	52	46	50	50	53	49
55	50	53	49	47	51	46	48	51	49
57	44	53	49	50	53	53	53	53	51
50	52	54	52	53	49	52	48	50	50

Нормальное распределение $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $x \in (-\infty, \infty)$ с параметрами

$\mu = 30,\ \sigma = 6$

28,1	14,4	45,9	21,7	18,4	37,8	33	42	38,4	38,9
25,5	27,8	29	35,1	28,5	29,2	28,7	35,3	25,9	36,6
27,8	15,1	32,8	27,1	28,8	32,3	23,8	24,1	34,6	35,3
19	41,3	25	27,9	35,7	29,8	24,7	21,6	37,1	23,2
32,2	25,9	34,1	29,7	30,2	29,4	37,9	36,4	24,7	40,5
36,6	29,7	24,8	25,7	30	25,2	42,9	25,3	31,5	29,9
27,3	35,1	27,9	29,4	22,5	20,3	18,5	33,7	47,1	34,3
24,9	45,9	38,7	34,1	23,1	35,5	32,7	32,2	30,5	35,5
31,5	28	24,2	31,7	34,8	26,1	33,3	32,7	32,7	37,7
25,9	31,7	28,2	32,5	38,5	35,2	36	19,8	38,6	21,5

Вариант 10.

Показательное распределение $f(x) = \lambda e^{-\lambda x},\ x \in [0, \infty)$ с параметром $\lambda = 0{,}13$

7,3	2,7	1,6	21,6	1,7	0	0,6	1,8	7,3	12,3
1,7	2,3	4,6	16,1	9,6	18	9,7	1,5	4,5	1,8
1,1	3,1	4,2	10,1	8,3	1,1	31,4	0,3	3,3	8,5
5,1	5	3,2	2,7	0,1	17,7	2	4,7	1,3	2,6
3,9	4,4	1,6	0,1	9,8	2,1	9,9	9,8	1,5	6,3
2,7	6	10,1	0,5	12,5	6,8	0,7	19	2,8	4,7
7,9	16,7	1,2	10,5	2,5	13,2	13,5	1,9	1,9	16,9
9,8	9,9	11	4,1	17,6	3,6	18,9	7,7	2,2	2,1
16,6	2,2	1,5	26	4,9	0,8	0,2	6,6	1,1	11,8
5,3	17,5	34,3	7,4	1,9	15	3,7	10	12,6	0,9

Биномиальное распределение $f(x; n, p) = \binom{n}{x} \cdot p^x \cdot q^{n-x},\ k = 0, 1, ..., n$ с параметрами

$n = 60;\ p = 0{,}85$

30	27	36	39	32	34	34	36	38	36
35	35	36	37	32	41	38	31	37	34
35	35	34	35	37	35	35	29	34	32
37	29	35	41	35	42	36	37	30	38
35	39	37	35	40	33	36	39	33	36
37	39	37	28	28	44	35	33	39	32
36	29	35	35	37	41	40	37	40	31
33	31	33	41	35	32	43	32	37	35
41	33	29	34	40	36	31	38	32	31
35	35	42	30	34	31	26	39	33	36

Нормальное распределение $f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $x\in(-\infty,\infty)$ с параметрами

$\mu=5,\ \sigma=1$

4,9	5,1	7,2	4,3	5,3	5,2	3,5	5,1	3,7	4,9
6,1	5,7	5	5,3	5,8	5,2	4,9	5	6,6	4,6
5,2	5,8	5,2	3,6	3,9	3,6	5,1	5,8	5,7	5,4
5,1	3,7	5,1	5,5	4,4	4,1	3,3	4,7	4,7	5,5
4,7	4,7	3,4	7,2	5,4	6,2	4,4	6,1	5,7	5,8
3,4	6,4	6,2	4,9	5,5	5,2	6	4	6,4	4,3
5,4	4,3	5	4,1	7,4	4,9	4,1	3,6	5,2	2,5
6,1	5,2	6	5,3	5	5,6	5,2	6,7	6	5,8
5,7	2,9	5	4,4	4,8	5,4	4,9	5,3	7,7	6,3
4,2	4,7	5,1	4	4,7	4,7	5,1	3,8	3,6	5,9

Приложение 2.2

Т а б л и ц а 3. Значения $t_\alpha(n)/\sqrt{n}$ ([1], стр. 198)

n	α			n	α			n	α		
	0,90	0,95	0,99		0,90	0,95	0,99		0,90	0,95	0,99
5	0,899	1,150	1,800	25	0,342	0,412	0,558	45	0,250	0,300	0,401
6	0,739	1,000	1,510	26	0,334	0,403	0,545	46	0748	0,297	0,396
7	0,715	0,890	1,320	27	0,328	0,394	0,532	47	0,245	0,294	0,392
8	0,657	0,816	1,190	28	0,322	0,387	0,521	48	0742	0,290	0, 388
9	0,611	0,754	1,080	29	0,316	0,380	0,513	49	0740	0,287	0783
10	0,574	0,706	1,000	30	0,310	0,372	0,502	50	0737	0,284	0779
11	0,541	0,663	0,936	31	0,304	0,366	0,492	55	0,226	0,270	0,360
12	0,515	0,630	0,881	32	0,295	0,354	0,475	60	0,216	0,258	0,343
13	0,491	0,598	0,833	33	0,290	0,349	0,468	65	0,207	0,248	0,329
14	0,471	0,572	0,797	34	0,288	0,347	0,463	70	0,199	0,238	0,316
15	0,453	0,550	0,762	35	0,286	0,346	0,461	75	0,192	0,230	0,305
16	0,436	0,530	0,730	36	0,282	0,344	0,459	80	0,186	0,222	0,295
17	0,422	0,512	0,704	37	0,278	0,333	0,447	90	0,175	0,209	0,277
18	0,410	0,495	0,679	38	0,274	0,329	0,441	100	0,166	0,198	0,263
19	0,396	0,479	0,655	39	0,270	0,324	0,434	120	0,151	0,181	0,239
20	0,386	0,466	0,637	40	0,266	0,320	0,428	150	0,135	0,161	0,213
21	0,376	0,454	0,618	41	0,264	0,316	0,422	200	0,117	0,139	0,184
22	0,366	0,442	0,601	42	0,260	0,312	0,417	250	0,104	0,124	0,164
23	0,357	0,431	0,585	43	0,256	0,308	0,411	300	0,095	0,114	0,150
24	0,349	0,421	0,571	44	0,253	0,304	0,406	400	0,082	0,098	0,129

- Кобзарь А.И. Прикладная математическая статистика. Для инженеров и научных работников. – М.: ФИЗМАТЛИТ, 2012. – 816с.

Приложение 2.3

Таблица 5. Значения $a(r,\alpha)$ ([1], стр. 200)

r	α			r	α			r	α		
	0,90	0,95	0,99		0,90	0,95	0,99		0,90	0,95	0,99
1	0,43	0,33	0,22	20	0,77	0,72	0,63	200	0,92	0,89	0,86
2	0,51	0,42	0,30	25	0,79	0,74	0,66	250	0,92	0,90	0,87
3	0,57	0,48	0,36	30	0,80	0,76	0,68	300	0,93	0,91	0,88
4	0,60	0,52	0,40	40	0,83	0,78	0,71	400	0,94	0,92	0,89
5	0,62	0,55	0,43	50	0,84	0,80	0,74	500	0,94	0,93	0,90
6	0,65	0,57	0,46	60	0,86	0,82	0,76	600	0,95	0,94	0,91
8	0,68	0,61	0,50	80	0,87	0,84	0,78	800	0,96	0,94	0,92
10	0,70	0,64	0,53	100	0,88	0,86	0,80	1000	0,96	0,95	0,93
15	0,74	0,68	0,59	150	0,90	0,88	0,84				

Таблица 6. Значения $b(r,\alpha)$ ([1], стр. 201)

r	α			r	α			r	α		
	0,90	0,95	0,99		0,90	0,95	0,99		0,90	0,95	0,99
1	0,26	0,21	0,15	11	0,66	0,60	0,51	60	0,84	0,81	0,75
2	0,38	0,32	0,24	12	0,67	0,62	0,53	80	0,86	0,83	0,78
3	0,45	0,39	0,30	13	0,68	0,63	0,5-4	100	0,88	0,85	0,80
4	0,50	0,44	0,35	14	0,69	0,64	0,55	150	0,90	0,87	0,83
5	0,54	0,48	0,38	15	0,70	0,65	0,56	200	0,91	0,89	0,85
6	0,57	0,51	0,40	20	0,74	0,69	0,60	250	0,92	0,90	0,86
7	0,59	0,53	0,44	25	0,76	0,72	0,64	300	0,93	0,91	0,88
8	0,62	0,55	0,46	30	0,78	0,74	0,66	400	0,94	0,92	0,89
9	0,63	0,57	0,48	40	0,81	0,77	0,70	500	0,94	0,93	0,90
10	0,65	0,59	0,50	50	0,83	0,79	0,73	600	0,95	0,94	0,91

- Кобзарь А.И. Прикладная математическая статистика. Для инженеров и научных работников. – М.: ФИЗМАТЛИТ, 2012. – 816с.

ЛАБОРАТОРНАЯ РАБОТА №3

Тема. Оценка закона распределения на основе выборочных данных

Цель работы:

Оценка закона распределения генеральной совокупности на основе выборочных данных.

3.1. Необходимые теоретические сведения

3.1.1 Критерии, основанные на сравнении теоретической плотности распределения и эмпирической гистограммы

Критерий χ^2 (Пирсона) для простой гипотезы

Пусть $\{x_1, x_2, \ldots, x_n\}$ выборка из генеральной совокупности F. Проверяется гипотеза $H_0 : F = F_1$ против альтернативы $H_1 : F \neq F_1$.

Представим выборку в виде группированного ряда, разбив предполагаемую область значений случайной величины на $m = 1 + 3{,}322 \cdot \lg n$ интервалов. Пусть n_i - число элементов выборки попавших в i-ый интервал, а p_i - теоретическая вероятность попадания в этот интервал при условии истинности H_0. Составим статистику $\rho(x) = \sum_{i=1}^{m} \frac{(n_i - np_i)^2}{np_i}$, которая характеризует сумму квадратов отклонения наблюдаемых значений n_i от ожидаемых np_i по всем интервалам группирования.

Теорема Пирсона. Если H_0 верна, то при фиксированном m и $n \to \infty$

$$\rho(x) = \sum_{i=1}^{m} \frac{(n_i - np_i)^2}{np_i} \Rightarrow \chi^2(m-1). \qquad (3.1)$$

Таким образом, статистику $\rho(x)$ можно использовать в качестве статистики критерия согласия для проверки гипотезы о виде закона распределения, который будет иметь вид:

$$F_n(x) = \begin{cases} H_0, & \rho(x) < \chi^2_{1-\alpha}(m-1) \\ H_1, & \rho(x) \ge \chi^2_{1-\alpha}(m-1) \end{cases}, \ \rho(x) = \sum_{i=1}^{m} \frac{\left(n_i - np_i\right)^2}{np_i}, \tag{3.2}$$

где $\chi^2_{1-\alpha}(m-1)$-квантиль распределения $\chi^2(m-1)$ с ($m-1$) степенями свободы. Данный критерий называется критерием χ^2 или критерием согласия Пирсона.

Замечание. Критерий не состоятелен для альтернатив, для которых $\widetilde{p}_i = p_i$ для всех $i \in \{1,2,\ldots,m\}$. Поэтому, следует стремиться к как можно большему числу интервалов группирования. Однако, с другой стороны, сходимость к χ^2 величины $\frac{(n_i - np_i)^2}{np_i}$ обеспечивается ЦПТ, то есть ожидаемое значение np_i для каждой ячейки не должно быть слишком мало. Поэтому обычно число интервалов выбирают таким образом, чтобы $np_i \ge 5$.

Критерий χ^2 (Пирсона) для сложной гипотезы

Пусть $\{x_1, x_2, \ldots, x_n\}$ выборка из генеральной совокупности F. Проверяется сложная гипотеза $H_0 : F = F_\theta$, где θ - неизвестный параметр распределения F (или вектор параметров), против альтернативы $H_1 : F \ne F_\theta$.

Пусть выборка по прежнему представлена в виде группированного ряда и n_i - число элементов выборки попавших в i-ый интервал, $i \in \{1,2,\ldots,m\}$. Статистику (2.1) мы не можем в этом случае использовать для построения критерия Пирсона, так как не можем вычислить теоретические значения вероятностей p_i, которые зависят от неизвестного параметра θ. Пусть θ^* - оценка параметра θ, а $p_i^*(\theta^*)$ - соответствующие ей оценки вероятностей p_i. Составим статистику $\rho(x) = \sum_{i=1}^{m} \frac{\left(n_i - np_i^*\right)^2}{np_i^*}$.

Теорема Пирсона. Если H_0 верна, и l - число компонент вектора θ (число неизвестных параметров распределения), то при фиксированном m и $n \to \infty$

$$\rho(x) = \sum_{i=1}^{m} \frac{\left(n_i - np_i^*\right)^2}{np_i^*} \Rightarrow \chi^2_{m-l-1}. \tag{3.3}$$

Таким образом, критерий Пирсона для параметрической гипотезы будет иметь вид:

$$F_n(x)=\begin{cases} H_0, & \rho(x)<\chi^2_{1-\alpha}(m-l-1) \\ H_1, & \rho(x)\ge\chi^2_{1-\alpha}(m-l-1) \end{cases}, \tag{3.4}$$

где $\chi^2_{1-\alpha}(m-l-1)$ - квантиль распределения $\chi^2(m-l-1)$ с $(m-l-1)$ степенями свободы.

3.1.2 Критерии, основанные на сравнении теоретической и эмпирической функций распределения вероятностей

Пусть $F_n(x)$ – эмпирическая функция распределения случайной величины x, представленной выборкой $x_1 \le x_2 \le .... \le x_n$:

$$F_n(x)=\begin{cases} 0, & x<x_1; \\ \dfrac{i}{n}, & x_i \le x \le x_{i+1},\ 1\le i\le n-1; \\ 1, & x\ge x_n. \end{cases} \tag{3.5}$$

Для проверки нулевой гипотезы $H_0 : F_n(x)=F(x)$, где $F(x)$ – полностью определенная (с точностью до параметров) теоретическая функция распределения, рассматривается расстояние между эмпирической и теоретической функциями распределения

Критерий Колмогорова-Смирнова

Для практического применения используются формулы

$$D_n^+=\max_{1\le i\le n}\left(\frac{i}{n}-F(x_i)\right);\ D_n^-=\max_{1\le i\le n}\left(F(x_i)-\frac{i-1}{n}\right);\ D_n=\max\left(D_n^+,D_n^-\right). \tag{3.6}$$

Здесь $F(x_i)$ – теоретическая функция распределения.
Критические значения статистики Колмогорова-Смирнова для уровня значимости $0,01\le\alpha\le 0,2$ вычисляются по формуле

$$D_n(\alpha)=\left(\frac{-\ln(\alpha/2)}{2n}\right)^{1/2}-\frac{1}{6n}. \tag{3.7}$$

Критерий Колмогорова

Статистика Колмогорова вычисляется по формуле

$$D_n=\max\left|F_n(x)-F(x)\right|. \tag{3.8}$$

Критическое значение $D_n(\alpha)$ для статистики $D_n\sqrt{n}$ приведены в таблице 3.1

Таблица 3.1

α	0.1	0.05	0.02	0.01	0.001
$D_n(\alpha)$	1.224	1.358	1.52	1.627	1.95

Критерий Смирнова-Крамера-фон Мизеса

Статистика критерия имеет вид

$$w^2 = \frac{1}{12n} + \sum_{i=1}^{n} \left\{ F(x_i) - \frac{2i-1}{2n} \right\}^2, \qquad (3.9)$$

где $F(x)$ – теоретическая функция распределения.

При объеме выборки $n > 40$ можно использовать приведенные в табл. 3.2 квантили распределения w^2, которые следуют из его предельного распределения (α – уровень значимости, принятый для проверки H_0).

Таблица 3.2. Квантили распределения w^2 ([1], стр. 217)

α	0,900	0,950	0,990	0,995	0,999
$w^2(\alpha)$	0,3473	0,4614	0,7435	0,8694	1,1679

3.1.3 Специальные критерии проверки нормальности распределения

Модифицированный критерий χ^2

Статистика критерия подсчитывается по формуле

$$\chi^2 = \frac{m}{n} \sum_{i=1}^{m} n_i^2 - n, \qquad (3.10)$$

где n – объем выборки; n_i – количество членов выборки, попавшие в i-й интервал. Границы интервалов определяются как

$$a_i = \bar{x} + c_i s \ \ (i = 0, 1, ..., m). \qquad (3.11)$$

Здесь $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$, $s = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2}$; $m = 1 + 3{,}322 \lg n$ – количество интервалов.

Значения коэффициентов c_i для $m = 7$ приведены в табл. 3.3.

Таблица 3.3. Значения коэффициентов c_i модифицированного χ^2-критерия нормальности для $m=7$ ([1], стр. 231)

m	c_0	c_1	c_2	c_3	c_4	c_5	c_6	c_7
7	$-\infty$	-1,0676	-0,5659	-0,1800	0,1800	0,5659	1,0676	∞

Если $\chi^2 > d_m(\alpha)$, где $d_m(\alpha)$ – критическое значение статистики критерия на уровне значимости α, то гипотеза нормальности отклоняется. Критические значения $d_m(\alpha)$ для $m=7$ приведены в табл. 3.4

Таблица 3.4. Критические значения $d_m(\alpha)$ модифицированного χ^2 - критерия нормальности для $m=7$ ([1], стр. 232)

m	α		
	0,10	0,05	0,01
7	8,322	10,038	13,837

Критерий типа Колмогорова – Смирнова

Модифицированная статистика для проверки нормальности распределения в ситуации, когда оба его параметра оцениваются по выборке рассчитывается по формуле

$$D_n^{н} = D_n\left(\sqrt{n} - 0,01 + \frac{0,85}{\sqrt{n}}\right). \quad (3.12)$$

Критические значения $D_n^{н}(\alpha)$ (α – уровень значимости) приведены в табл. 3.5

Таблица 3.5. Критические значения статистики Колмогорова – Смирнова, модифицированной для проверки нормальности распределения ([1], стр. 233).

α	0,15	0,10	0,05	0,03	0,01
$D_n^{н}(\alpha)$	0,775	0,819	0,895	0,955	1,035

Критерий Фроцини

Статистика Фроцини рассчитывается по формуле

$$B_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left| F^*(z_i) - \frac{2i-1}{2n} \right|, \quad (3.13)$$

где $z_i = \dfrac{x_i - \overline{x}}{s}$; $\overline{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$, $s = \sqrt{\dfrac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}$; $F^*(z) = \dfrac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-\frac{1}{2}t^2}\,dt$.

Критические значения статистики B_n приведены в таблице 3.6.

Таблица 3.6 Критические значения статистики Фроцини B_n для проверки нормальности распределения. Здесь α – уровень значимости ([1], стр. 235)

n	α		
	0,9	0,95	0,99
20	0,2556	0,2839	0,3363
∞	0,2560	0,2840	0,3410

3.1.4 Специальные критерий проверки экспоненциальности распределения

Критерии типа Колмогорова –Смирнова

Пусть дана выборка $x_1 \le x_2 \le .... \le x_n$. Предположим, имеет место гипотетически закон распределения вероятностей $F(x) = 1 - \exp\left(-\lambda \cdot x\right)$, где λ – неизвестный параметр, определяемый из выборочных данных по формуле

$$\lambda = \frac{1}{\overline{x}};\ \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i\ . \qquad (3.13)$$

Обозначим $z_i = \lambda \cdot x_i$ и перейдем к нормированному экспоненциальному распределению $F^*(z_i) = 1 - \exp(-z_i)$, для которого можно применить следующие критерии согласия

$$D_n^+ = \max\left[\frac{i}{n} - F^*(z_i)\right];\ D_n^- = \max\left[F^*(z_i) - \frac{i-1}{n}\right];\ D_n \max\left(D_n^+, D_n^-\right). \qquad (3.14)$$

Критерий Купера

$$V = D_n^+ + D_n^-\ . \qquad (3.15)$$

Критерий Смирнова-Крамера-фон Мизеса

$$w^2 = \sum_{i=1}^{n}\left(F^*(z_i) - \frac{2i-1}{2n}\right)^2 + \frac{1}{12n}\ . \qquad (3.16)$$

Критерий Ватсона

$$U^2 = w^2 - n\left(\overline{F} - 0,5\right)^2, \text{ где } \overline{F} = \frac{1}{n}\sum_{i=1}^{n} F^*(z_i). \qquad (3.17)$$

Критерий Андерсона-Дарлинга

$$A^2 = -\frac{1}{n}\sum_{i=1}^{n}\Big((2i-1)\cdot\Big(\ln(F^*(z_i) + \ln(1 - F^*(z_{n+1-i}))\Big)\Big) - n. \qquad (3.18)$$

Критические значения для различных уровней значимости для объёма выборки n = 100 приведены в табл. 3.7.

Таблица 3.7
Критические значения статистик критериев согласии типа Колмогорова-Смирнова для проверки экспоненциальности распределения с неизвестными параметрами n = 100 ([1], стр. 283)

Уровень значимости α (верхние процентные точки) Объем выборки $n = 100$					
0,25	0,15	0,10	0,05	0,025	0,01
Статистика $\sqrt{n}D_n$					
0,824	0,911	0,972	1,072	1,171	1,278
Статистика V					
1,310	1,419	1,502	1,647	1,740	1,897
Статистика w^2					
0,113	0,144	0,170	0,215	0,263	0,328
Статистика U^2					
0,089	0,110	0,126	0,155	0,184	0,229
Статистика A^2					
0,710	0,875	1,008	1,250	1,510	1,865

Критерий Фроцини

$$B_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left|1 - EXP(-\lambda\cdot x_i) - \frac{2i-1}{2n}\right|, \qquad (3.19)$$

В таблице 3.8 приведены критические значения статистики Фроници для $n = \infty$. Полная таблица приведена в работе ([1], стр. 281)

Таблица 3.8
Критические значения статистики Фроцини $B_n(\alpha)$ для проверки экспоненциальности распределения. Здесь α – уровень значимости ([1], стр. 281)

n	α		
	0,9	0,95	0,99
∞	0,3380	0,3840	0,4760

3.1.5 Критерии согласия для равномерного распределения

Критерий Колмогорова-Смирнова

Дана выборка $U_1 < U_2 < ... < U_n$ из равномерного закона распределения.
Рассчитываются статистики

$$D^+ = \max_i \left(U_i - \frac{i}{n+1} \right); \ D^- = \max_i \left(\frac{i}{n+1} - U_i \right); \ D = \max\left(D^+, D^- \right). \quad (3.20)$$

Распределения указанных статистик быстро сходятся к предельному, если использовать их модификации:

$$D = \left(D + \frac{0,4}{n} \right) \cdot \left(\sqrt{n} + 0,2 + \frac{0,68}{\sqrt{n}} \right). \quad (3.21)$$

Критические значения для модифицированной статистики приведены в табл. 3.9.

Критерий Купера

$$V = D_n^+ + D_n^- . \quad (3.22)$$

$$V = \left(V - \frac{1}{n-1} \right) \cdot \left(\sqrt{n+1} + 0,1555 + \frac{0,24}{\sqrt{n+1}} \right) \quad (3.23)$$

Таблица 3.9. Критические значения D и V критерия равномерности ([1], стр. 330)

	Уровень значимости α				
	0,01	0,025	0,05	0,1	0,15
D	1,628	1,480	1,358	1,224	1,138
V	2,001	1,852	1,47	1,620	1,537

Критерий Фроцини

Дана выборка из равномерного распределения $U_1 < U_2 < U_3 ... U_n$ на отрезке [0; 1].
Статистика Фроцини рассчитывается по формуле

$$B_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left| U_i - \frac{2i-1}{2n} \right| . \quad (3.24)$$

Если $B_n < B_n(\alpha)$, то нулевая гипотеза о равномерном распределении принимается.

Критические значения статистики Фроцини приведены в таблице 3.10.

Таблица 3.10 Критические значения статистики Фроцини $B_n(\alpha)$ для проверки равномерности распределения. Здесь α – уровень значимости ([1], стр. 331)

n	α		
	0,9	0,95	0,99
15	0,4961	0,5730	0,7418
∞	0,4970	0,5780	0,7440

3.2. Практическое задание

Имеется выборка объемом $n=100$ из известного распределения F (см. приложение 3.1). Предполагается, что F может быть одним из следующих распределений:

1) F - нормальное распределение с плотностью $f(x)=\dfrac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}$, $x\in R$, где a и $\sigma>0$ – параметры;

2) F - показательное распределение с плотностью $f(x)=\lambda e^{-\lambda x}$, $x>0$, где $\lambda>0$ – параметр;

3) F - равномерное распределение на отрезке $[a,b]$, где a и b – параметры.

Требуется:

1) Представить выборку в виде группированного статистического ряда. Количество интервалов равно $m=1+3{,}322\cdot\lg(n)=7$. При разбивке на интервалы следует следить за тем, чтобы частоты n_i для всех интервалов были одного порядка, причем количество выборочных значений n_i попавших в каждый интервал должно быть не меньше 5 ($n_i\geq 5\quad\forall i=\overline{1,m}$). Допускается $1<n_i<5\quad\forall i=\overline{1,m}$ не более, чем в 2-х интервалах. В противном случае следует объединять соседние интервалы, добиваясь относительно равномерного распределения частот по интервалам.

2) Найти числовые характеристики выборки: выборочное среднее $\overline{x}=\dfrac{1}{n}\sum\limits_{i=1}^{n}x_i$, выборочную дисперсию $s^2=\dfrac{1}{n-1}\sum\limits_{i=1}^{n}(x_i-\overline{x})^2$, выборочный коэффициент асимметрии $\overline{A}=\dfrac{1}{n\cdot s^3}\sum\limits_{i=1}^{n}\left(x_i-\overline{x}\right)^3$ и выборочный коэффициент эксцесса $\overline{E}=\dfrac{1}{n\cdot s^4}\sum\limits_{i=1}^{n}\left(x_i-\overline{x}\right)^4-3$.

3) Построить гистограмму и сравнить ее с гистограммой плотности теоретического распределения. При построении кривых плотностей параметры

теоретических распределений заменить оценками, найденными по методу максимального правдоподобия или по методу моментов.

4) Выдвинуть гипотезу H_0 о виде закона распределения (на основе сравнения гистограммы с графиком плотности теоретического распределения).

5) Используя критерий Пирсона для сложной гипотезы, проверить гипотезу H_0 для уровня значимости $\alpha = 0,05$.

6) Проверить гипотезу H_0 на основе критериев сравнения интегральных функций распределения вероятностей. При этом теоретические распределения рассчитываются с использованием точных значений параметров.

7) Проверить гипотезу H_0 на основе специальных критериев сравнения функций распределения вероятностей. При этом теоретические распределения рассчитываются с использованием выборочных значений параметров.

8) Сделать выводы

3.3 Excel-скриншоты лабораторной работы №3

Лабораторная работа 3 - Excel

Лабораторная работа №3

Оценка закона распределения на основе выборочных данных

Критерии, основанные на сравнении теретической плотности распределения и эмпирической гистограммы

Исходная выборка (мю = 3; сигма =1) нормальное распределение

A	B	C	D	E	F	G	H	I	J
2,278887	2,205481	4,52891516	2,321785	0,223422	3,714581	3,37065092	2,57703722	2,4266727	3,746852
3,320703	3,43126	3,12112423	4,825038	3,841337	3,575221	2,01989895	2,64071549	2,4069049	2,096772
2,928605	1,927956	2,60043169	1,897424	4,602925	2,572516	1,7875708	2,24574695	2,1499112	4,250974
3,951802	4,028748	1,83335704	2,773831	2,794737	3,418282	4,66206519	1,37340965	2,3502979	2,693119
3,045724	4,444405	3,61809033	2,260508	2,955502	2,994913	3,23339453	0,32652213	3,9545738	2,282256
3,406787	1,612898	2,76794015	1,870326	4,177896	2,645119	3,77608775	2,28611032	2,5146095	2,922238
2,12998	4,533608	0,73279342	2,144188	4,337012	0,531013	2,72694923	3,07914423	3,6357823	2,213323
1,681824	2,223187	4,9957497	3,185692	4,060296	2,27292	3,28330874	4,23474365	2,8947351	1,689805
0,462655	3,895661	2,32226501	2,771554	4,241328	5,120823	2,44157707	3,15054184	4,5249861	2,932438
3,371633	4,564849	3,36852043	2,089964	4,145747	3,396587	4,37341885	2,48224968	2,1859227	1,423859

xmin	xmax	m	шаг	xmid
0,223421815	5,12082341	7	0,8162336	2,92465504

среднее	стандартное отклонение s	асимметрия	эксцесс
2,924655	1,06782599	-0,17008523	-0,139864178

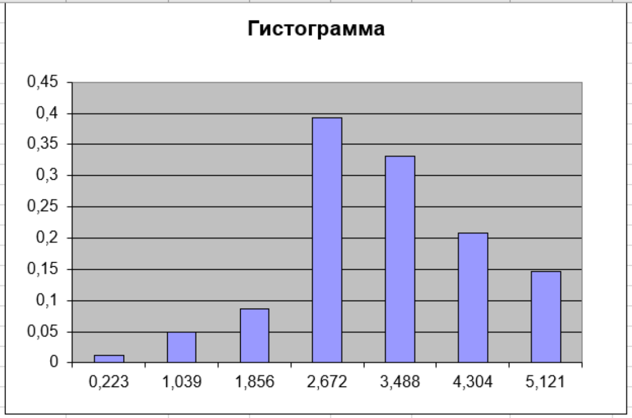
теоретическое значение мю	3	1	теоретическое значение сигма

диапазон	частоты	исправленный диапазон	частоты	середина интервала	относительная частота	середина интервала	плотность частоты	теретическая плотность показательного распределения	вероятность попадания в интервал pi*	значения npi*	(ni-npi*)^2/npi*	теретическое интегральное показательное распределение для простой гипотезы	Эмпирическое интегральное распределение	Отклонение между теоретическим и эмпирическим распределением для простой гипотезы
												0	0	0
0,223422	1	0,223422	1	0,22342181	0,01	0,223	0,012251	0,01522001	0,00261991	0,261991	2,07891627	0,002746721	0,01	0,007253279
1,039655	4	1,039655	4	1,03965541	0,04	1,039	0,049006	0,078572501	0,033005513	3,300551	0,14822632	0,024977764	0,05	0,025022236
1,855889	7	1,855889	7	1,85588901	0,07	1,856	0,08576	0,226425339	0,119759667	11,97597	2,06749445	0,126288806	0,12	0,006288806
2,672123	32	2,672123	32	2,67212261	0,32	2,672	0,392045	0,363289677	0,248013404	24,80134	2,0894314	0,371502186	0,44	0,068497814
3,488356	27	3,488356	27	3,48835621	0,27	3,488	0,330788	0,325066847	0,294617621	29,46176	0,2056996	0,687351224	0,71	0,022648776
4,30459	17	4,30459	17	4,30458981	0,17	4,304	0,208274	0,16221234	0,200676516	20,06765	0,46893811	0,90398372	0,88	0,02398372
5,120823	12	5,120823	12	5,12082341	0,12	5,121	0,147017	0,045055763	0,078374082	7,837408	2,21082916	0,983031667	1	0,016968333

9,269535	статистика наблюдаемая
9,487729	статистика критическая

Вывод: та как 9,269 < 9,488, то гипотеза H0 принимается, выборочные данные распределены по нормальному закону

Лабораторная работа 3 - Excel



Критерии, основанные на сравнении теоретической и эмпирической функции распределения вероятностей

Критерий Колмогорова - Смирнова

0,05 уровень значимости

0,0720577

0,0440501

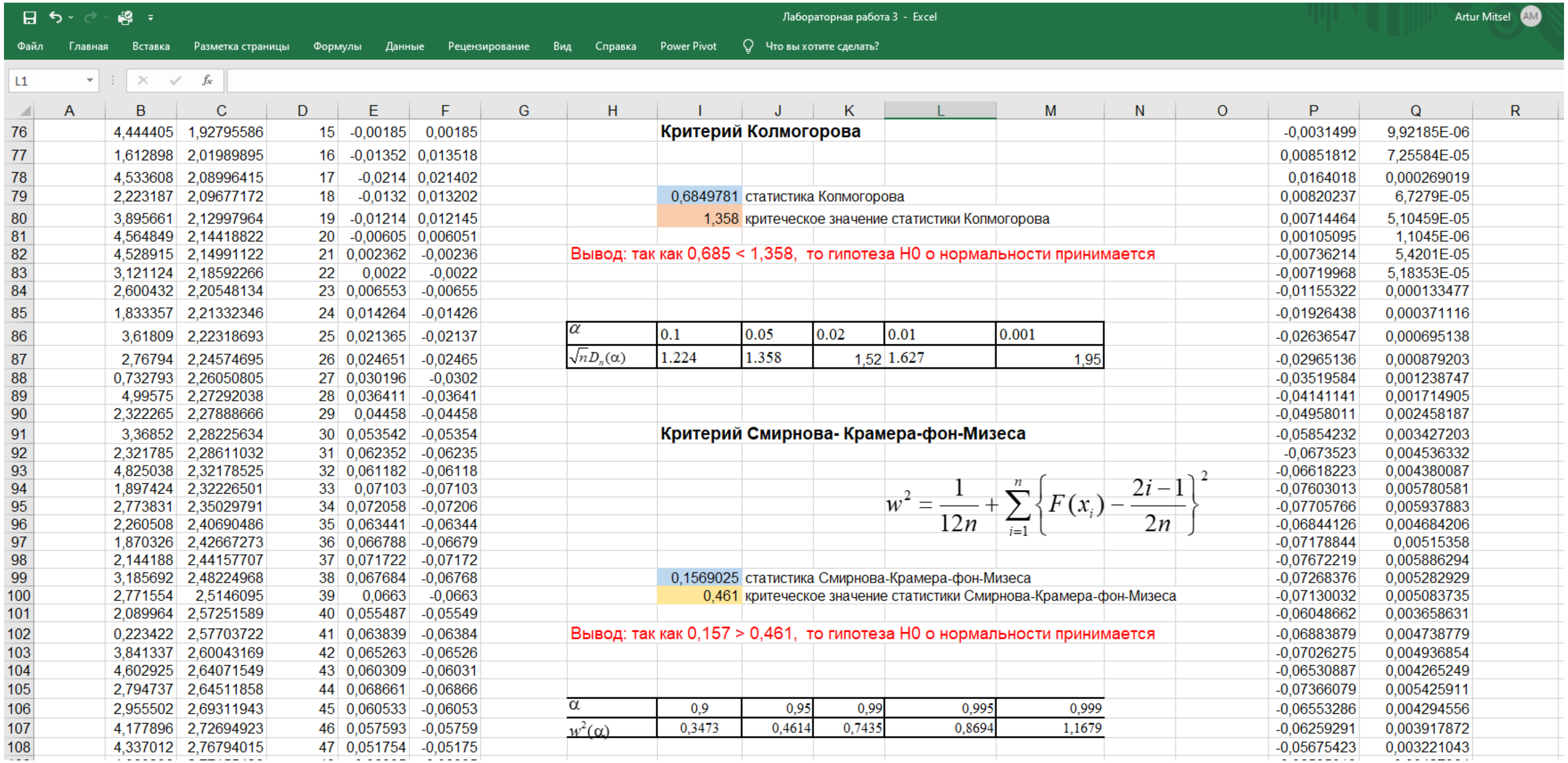
$$D_n(\alpha) = \left(\frac{-\ln(\alpha / 2)}{2n} \right)^{1/2} - \frac{1}{6n}$$

0,0720577 статистика Колмогорова-Смирнова

0,1374768 критеческое значение статистики Колмогорова-Смирнова

Вывод: так как 0,072 < 0,137, то гипотеза Н0 о нормальности принимается

Исходня выборка	Ранжирован ная выборка					
2,278887	0,22342181	1	-0,00275	0,002747	-0,00225328	5,07727E-06
3,320703	0,32652213	2	0,006247	-0,00625	-0,01124654	0,000126485
2,928605	0,4626553	3	0,014415	-0,01442	-0,01941515	0,000376948
3,951802	0,53101292	4	0,023225	-0,02323	-0,02822519	0,000796662
3,045724	0,73279342	5	0,028311	-0,02831	-0,0333112	0,001109636
3,406787	1,37340965	6	-0,00191	0,001912	-0,00308793	9,53534E-06
2,12998	1,42385932	7	0,002503	-0,0025	-0,00750331	5,62996E-05
1,681824	1,6128978	8	-0,01271	0,012705	0,0077053	5,93717E-05
0,462655	1,68182385	9	-0,01372	0,013722	0,00872234	7,60793E-05
3,371633	1,68980467	10	-0,00506	0,005065	6,4883E-05	4,20975E-09
2,205481	1,7875708	11	-0,01267	0,012674	0,00767407	5,88913E-05
3,43126	1,83335704	12	-0,01168	0,011677	0,00667729	4,45862E-05
1,927956	1,87032573	13	-0,00931	0,009307	0,00430675	1,85481E-05



	B	C	D	E	F	P	Q
76	4,444405	1,92795586	15	-0,00185	0,00185	-0,0031499	9,92185E-06
77	1,612898	2,01989895	16	-0,01352	0,013518	0,00851812	7,25584E-05
78	4,533608	2,08996415	17	-0,0214	0,021402	0,0164018	0,000269019
79	2,223187	2,09677172	18	-0,0132	0,013202	0,00820237	6,7279E-05
80	3,895661	2,12997964	19	-0,01214	0,012145	0,00714464	5,10459E-05
81	4,564849	2,14418822	20	-0,00605	0,006051	0,00105095	1,1045E-06
82	4,528915	2,14991122	21	0,002362	-0,00236	-0,00736214	5,4201E-05
83	3,121124	2,18592266	22	0,0022	-0,0022	-0,00719968	5,18353E-05
84	2,600432	2,20548134	23	0,006553	-0,00655	-0,01155322	0,000133477
85	1,833357	2,21332346	24	0,014264	-0,01426	-0,01926438	0,000371116
86	3,61809	2,22318693	25	0,021365	-0,02137	-0,02636547	0,000695138
87	2,76794	2,24574695	26	0,024651	-0,02465	-0,02965136	0,000879203
88	0,732793	2,26050805	27	0,030196	-0,0302	-0,03519584	0,001238747
89	4,99575	2,27292038	28	0,036411	-0,03641	-0,04141141	0,001714905
90	2,322265	2,27888666	29	0,04458	-0,04458	-0,04958011	0,002458187
91	3,36852	2,28225634	30	0,053542	-0,05354	-0,05854232	0,003427203
92	2,321785	2,28611032	31	0,062352	-0,06235	-0,0673523	0,004536332
93	4,825038	2,32178525	32	0,061182	-0,06118	-0,06618223	0,004380087
94	1,897424	2,32226501	33	0,07103	-0,07103	-0,07603013	0,005780581
95	2,773831	2,35029791	34	0,072058	-0,07206	-0,07705766	0,005937883
96	2,260508	2,40690486	35	0,063441	-0,06344	-0,06844126	0,004684206
97	1,870326	2,42667273	36	0,066788	-0,06679	-0,07178844	0,00515358
98	2,144188	2,44157707	37	0,071722	-0,07172	-0,07672219	0,005886294
99	3,185692	2,48224968	38	0,067684	-0,06768	-0,07268376	0,005282929
100	2,771554	2,5146095	39	0,0663	-0,0663	-0,07130032	0,005083735
101	2,089964	2,57251589	40	0,055487	-0,05549	-0,06048662	0,003658631
102	0,223422	2,57703722	41	0,063839	-0,06384	-0,06883879	0,004738779
103	3,841337	2,60043169	42	0,065263	-0,06526	-0,07026275	0,004936854
104	4,602925	2,64071549	43	0,060309	-0,06031	-0,06530887	0,004265249
105	2,794737	2,64511858	44	0,068661	-0,06866	-0,07366079	0,005425911
106	2,955502	2,69311943	45	0,060533	-0,06053	-0,06553286	0,004294556
107	4,177896	2,72694923	46	0,057593	-0,05759	-0,06259291	0,003917872
108	4,337012	2,76794015	47	0,051754	-0,05175	-0,05675423	0,003221043

Критерий Колмогорова

0,6849781 статистика Колмогорова

1,358 критеческое значение статистики Колмогорова

Вывод: так как 0,685 < 1,358, то гипотеза H0 о нормальности принимается

α	0.1	0.05	0.02	0.01	0.001
$\sqrt{n}D_n(\alpha)$	1.224	1.358	1,52	1.627	1,95

Критерий Смирнова- Крамера-фон-Мизеса

$$w^2 = \frac{1}{12n} + \sum_{i=1}^{n} \left\{ F(x_i) - \frac{2i-1}{2n} \right\}^2$$

0,1569025 статистика Смирнова-Крамера-фон-Мизеса

0,461 критеческое значение статистики Смирнова-Крамера-фон-Мизеса

Вывод: так как 0,157 > 0,461, то гипотеза H0 о нормальности принимается

α	0,9	0,95	0,99	0,995	0,999
$w^2(\alpha)$	0,3473	0,4614	0,7435	0,8694	1,1679

Лабораторная работа 3 - Excel   Artur Mitsel

Файл Главная Вставка Разметка страницы Формулы Данные Рецензирование Вид Справка Power Pivot Что вы хотите сделать? Поделиться

L1

Специальные критерии проверки гипотезы о нормальном распределении

	B	C	D	E	F	R
163						
164						
165						
166	Исходня выборка	Ранжированная выборка				
167	2,278887	0,22342181	1	0,004291	0,005709	0,0007087
168	3,320703	0,32652213	2	0,012515	-0,00252	0,007515
169	2,928605	0,4626553	3	0,019434	-0,00943	0,014434
170	3,951802	0,53101292	4	0,027506	-0,01751	0,0225065
171	3,045724	0,73279342	5	0,029946	-0,01995	0,0249462
172	3,406787	1,37340965	6	-0,01315	0,023152	0,0181516
173	2,12998	1,42385932	7	-0,00994	0,019941	0,014941
174	1,681824	1,6128978	8	-0,02964	0,039641	0,0346414
175	0,462655	1,68182385	9	-0,03223	0,042234	0,0372345
176	3,371633	1,68980467	10	-0,02376	0,033756	0,0287557
177	2,205481	1,7875708	11	-0,03347	0,04347	0,0384698
178	3,43126	1,83335704	12	-0,03339	0,043395	0,0383949
179	1,927956	1,87032573	13	-0,03173	0,041733	0,0367329
180	4,028748	1,89742412	14	-0,02803	0,038029	0,033029
181	4,444405	1,92795586	15	-0,02531	0,035309	0,0303091
182	1,612898	2,01989895	16	-0,03842	0,048417	0,0434173
183	4,533608	2,08996415	17	-0,0472	0,057203	0,0522033
184	2,223187	2,09677172	18	-0,03908	0,049082	0,0440818
185	3,895661	2,12997964	19	-0,03838	0,048378	0,0433779
186	4,564849	2,14418822	20	-0,03242	0,042422	0,0374222
187	4,528915	2,14991122	21	-0,02406	0,034062	0,0290623
188	3,121124	2,18592266	22	-0,02453	0,034528	0,0295284
189	2,600432	2,20548134	23	-0,02032	0,030317	0,0253167
190	1,833357	2,21332346	24	-0,01266	0,022658	0,0176578
191	3,61809	2,22318693	25	-0,00562	0,015619	0,0106186
192	2,76794	2,24574695	26	-0,00246	0,012458	0,0074582
193	0,732793	2,26050805	27	0,003017	0,006983	0,0019835
194	4,99575	2,27292038	28	0,009181	0,000819	0,004181
195	2,322265	2,27888666	29	0,017328	-0,00733	0,0123276
196	3,36852	2,28225634	30	0,026278	-0,01628	0,0212781
197	2,321785	2,28611032	31	0,035075	-0,02508	0,0300753
198	4,825038	2,32178525	32	0,033819	-0,02382	0,0288192
199	1,897424	2,32226501	33	0,043666	-0,03367	0,0386663
200	2,773831	2,35029791	34	0,044668	-0,03467	0,0396684
201	2,260508	2,40690486	35	0,036114	-0,02611	0,0311135

Критерий типа Колмогорова – Смирнова

0,05 уровень значимости

$$D_n^{\text{м}} = D_n\left(\sqrt{n} - 0{,}01 + \frac{0{,}85}{\sqrt{n}}\right)$$

0,0484181

0,0572033

0,0572033

0,5768386 статистика Колмогорова-Смирнова

0,895 кретеческое значение статистики Колмогорова-Смирнова

Вывод: так как 0,577 < 0,895, то гипотеза H0 о нормальности принимается

α	0,15	0,1	0,05	0,03	0,01
$\boxed{D_n^{\text{м}}}$	0,775	0,819	0,895	0,955	1,035

Модифицированный критерий χ^2

2,92465504 среднее значение

1,06782599 стандартное отклонение

$$\chi^2 = \frac{m}{n}\sum_{i=1}^{m} n_i^2 - n$$

$$a_i = \bar{x} + c_i s \quad (i = 0, 1, \ldots, m)$$

коэффициенты c_i	границы интервалов	частота	
-10	-7,7536049	0	0
-1,076	1,7756743	10	100
-0,5659	2,3203723	21	441
-0,18	2,7324464	15	225
0,18	3,1168637	12	144
0,5659	3,5289378	13	169
1,0676	4,0646661	12	144
10	13,602915	17	289
			1512

Лабораторная работа 3 - Excel

Файл Главная Вставка Разметка страницы Формулы Данные Рецензирование Вид Справка Power Pivot Что вы хотите сделать?

	A	B	C	D	E	F	R
202		1,870326	2,42667273	36	0,039518	-0,02952	0,0345181
203		2,144188	2,44157707	37	0,044507	-0,03451	0,0395074
204		3,185692	2,48224968	38	0,040675	-0,03067	0,0356745
205		2,771554	2,5146095	39	0,039511	-0,02951	0,0345109
206		2,089964	2,57251589	40	0,029214	-0,01921	0,0242139
207		0,223422	2,57703722	41	0,037613	-0,02761	0,032613
208		3,841337	2,60043169	42	0,039295	-0,02929	0,0342949
209		4,602925	2,64071549	43	0,034844	-0,02484	0,0298436
210		2,794737	2,64511858	44	0,043255	-0,03325	0,0382548
211		2,955502	2,69311943	45	0,035829	-0,02583	0,0308292
212		4,177896	2,72694923	46	0,033444	-0,02344	0,0284435
213		4,337012	2,76794015	47	0,02834	-0,01834	0,0233395
214		4,060296	2,77155426	48	0,037003	-0,027	0,0320034
215		4,241328	2,77383141	49	0,046161	-0,03616	0,0411613
216		4,145747	2,79473728	50	0,048418	-0,03842	0,0434181
217		3,714581	2,89473508	51	0,021177	-0,01118	0,0161767
218		3,575221	2,9222382	52	0,020903	-0,0109	0,0159029
219		2,572516	2,92860467	53	0,028524	-0,01852	0,0235244
220		3,418282	2,93243819	54	0,037092	-0,02709	0,0320922
221		2,994913	2,95550183	55	0,038477	-0,02848	0,0334772
222		2,645119	2,99491251	56	0,033771	-0,02377	0,0287706
223		0,531013	3,04572371	57	0,024865	-0,01487	0,0198652
224		2,27292	3,07914423	58	0,022483	-0,01248	0,0174832
225		5,120823	3,12112423	59	0,017011	-0,00701	0,0120107
226		3,396587	3,15054184	60	0,016233	-0,00623	0,0112334
227		3,370651	3,18569153	61	0,013439	-0,00344	0,0084388
228		2,019899	3,23339453	62	0,006241	0,003759	0,0012413
229		1,787571	3,28330874	63	-0,00152	0,011517	0,0065166
230		4,662065	3,32070261	64	-0,00464	0,014641	0,0096408
231		3,233395	3,36852043	65	-0,01117	0,021175	0,016175
232		3,776088	3,37065092	66	-0,0019	0,011905	0,0069048
233		2,726949	3,37163318	67	0,007759	0,002241	0,002759
234		3,283309	3,39658744	68	0,00926	0,00074	0,0042604
235		2,441577	3,40678742	69	0,015812	-0,00581	0,0108116
236		4,373419	3,4182823	70	0,021943	-0,01194	0,0169427
237		2,577037	3,43125965	71	0,027598	-0,0176	0,022598
238		2,640715	3,57522129	72	-0,00882	0,018819	0,0138185

5,84 модифицированная статистика хи-квадрат

10,038 критическое значение

Вывод: так как 5,84 < 10,038, то гипотеза H0 о нормальности принимается

Значения коэффициентов модифицированного хи-квадрат критерия нормальности для m=7

m	c_0	c_1	c_2	c_3	c_4	c_5	c_6	c_7
7	$-\infty$	-1,0676	-0,5659	-0,18	0,18	0,5659	1,0676	∞

Критические значения модифицированного хи-квадрат критерия нормальности для m=7

$m=7$	α		
	0,1	0,05	0,01

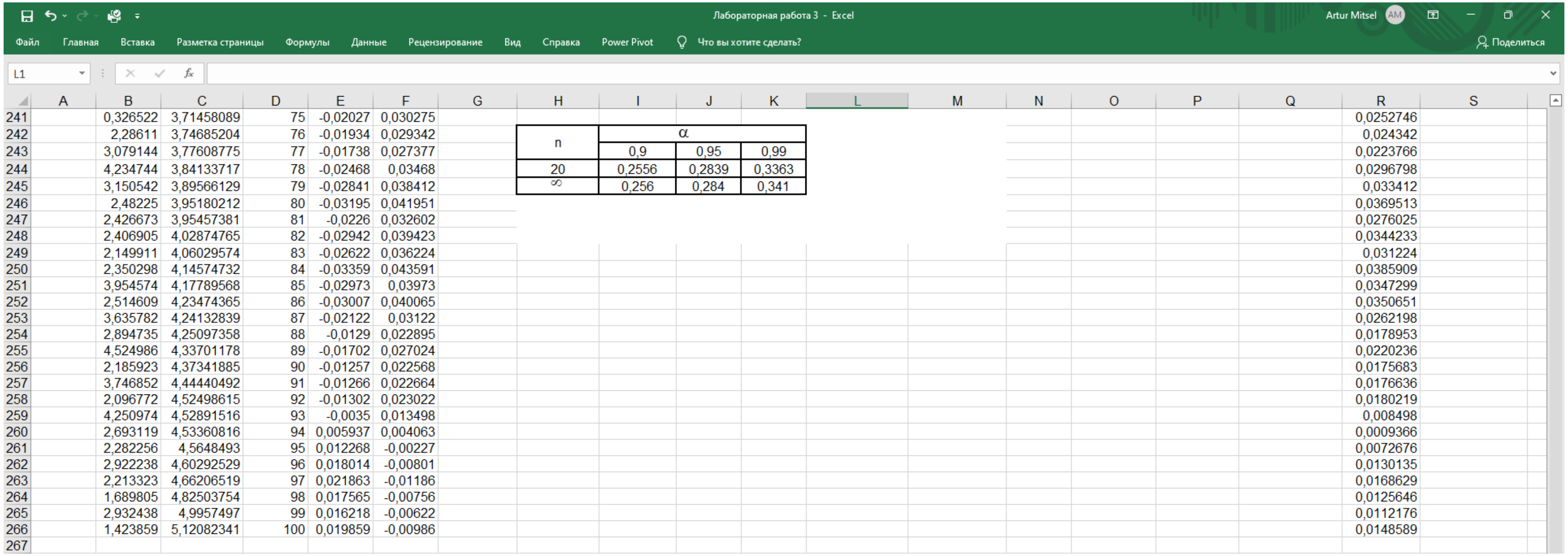
Критерий Фроцини

$$B_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left| F^*(z_i) - \frac{2i-1}{2n} \right|$$

0,23698202 статистика Фроцини

0,284 критическое значение статистики

Вывод: так как 0,237 < 0,284, то гипотеза H0 о нормальности принимается



	A	B	C	D	E	F	R
241		0,326522	3,71458089	75	-0,02027	0,030275	0,0252746
242		2,28611	3,74685204	76	-0,01934	0,029342	0,024342
243		3,079144	3,77608775	77	-0,01738	0,027377	0,0223766
244		4,234744	3,84133717	78	-0,02468	0,03468	0,0296798
245		3,150542	3,89566129	79	-0,02841	0,038412	0,033412
246		2,48225	3,95180212	80	-0,03195	0,041951	0,0369513
247		2,426673	3,95457381	81	-0,0226	0,032602	0,0276025
248		2,406905	4,02874765	82	-0,02942	0,039423	0,0344233
249		2,149911	4,06029574	83	-0,02622	0,036224	0,031224
250		2,350298	4,14574732	84	-0,03359	0,043591	0,0385909
251		3,954574	4,17789568	85	-0,02973	0,03973	0,0347299
252		2,514609	4,23474365	86	-0,03007	0,040065	0,0350651
253		3,635782	4,24132839	87	-0,02122	0,03122	0,0262198
254		2,894735	4,25097358	88	-0,0129	0,022895	0,0178953
255		4,524986	4,33701178	89	-0,01702	0,027024	0,0220236
256		2,185923	4,37341885	90	-0,01257	0,022568	0,0175683
257		3,746852	4,44440492	91	-0,01266	0,022664	0,0176636
258		2,096772	4,52498615	92	-0,01302	0,023022	0,0180219
259		4,250974	4,52891516	93	-0,0035	0,013498	0,008498
260		2,693119	4,53360816	94	0,005937	0,004063	0,0009366
261		2,282256	4,5648493	95	0,012268	-0,00227	0,0072676
262		2,922238	4,60292529	96	0,018014	-0,00801	0,0130135
263		2,213323	4,66206519	97	0,021863	-0,01186	0,0168629
264		1,689805	4,82503754	98	0,017565	-0,00756	0,0125646
265		2,932438	4,9957497	99	0,016218	-0,00622	0,0112176
266		1,423859	5,12082341	100	0,019859	-0,00986	0,0148589
267							

n	α		
	0,9	0,95	0,99
20	0,2556	0,2839	0,3363
∞	0,256	0,284	0,341

Лабораторная работа 3 - Excel

Лабораторная работа №3

Оценка закона распределения на основе выборочных данных

Критерии, основанные на сравнении теоретической плотности распределения и эмпирической гистограммы

Исходная выборка (лямбда = 0,1) экспоненциальное распределение

3,03648	0,705156	19,52125	8,397073	8,708589	6,173839	7,250544	5,5015417	7,558989	0,807856
2,946339	6,657837	4,136777	1,452183	2,366472	7,058314	16,26272	9,4036744	19,22265	4,6170659
20,02152	4,790966	27,48437	1,564677	18,81185	4,947902	4,973965	2,8797713	3,980613	23,700271
25,08843	8,926758	1,133714	6,73654	6,018875	0,011909	2,805156	14,103554	3,003869	18,944887
10,1322	7,384399	19,92034	1,415901	1,023573	5,848733	7,390788	24,537494	5,29265	12,598377
6,513966	6,08147	4,070533	33,04603	12,02763	20,10776	0,062145	20,326691	3,951574	9,9845036
13,43491	11,47232	8,007584	19,62279	10,82207	1,828649	1,498517	8,765619	1,93585	1,0120846
3,847007	36,6021	12,35922	21,82712	2,9447	15,79139	12,46056	6,3117168	17,15657	29,540988
4,914422	35,05551	22,58905	6,32492	1,209199	4,159881	23,67093	15,146467	0,22468	6,3272185
17,005	8,767086	12,59192	25,44738	3,679825	2,407155	8,716612	20,803881	4,162195	16,145475

xmin	xmax	m	шаг	xmid
0,011909	36,60210232	7	6	10,300223

параметр лямбда	дисперсия s2	асимметрия	эксцесс
0,097085279	74,89857507	1,0464038	0,428364801

0,1 теоретическое значение лямбда

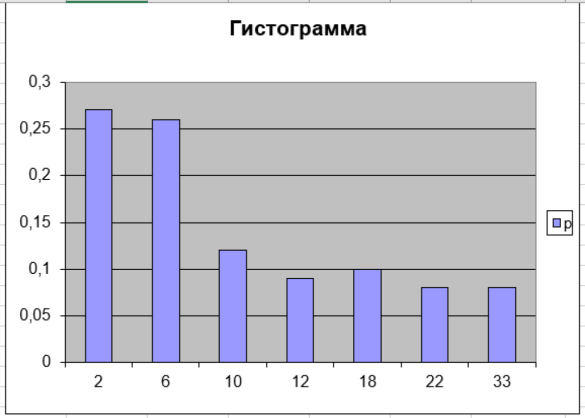
диапазон	частоты	исправленный диапазон	частоты	середина интервала	относительная частота	середина интервала	плотность частоты	теретическая плотность показательного распреде	вероятность попадания в интервал pi*	значения npi*	(ni-npi*)^2/npi*	теретическое интегральное показательное распределение для простой гипотезы	Эмпирическое интегральное распределение	Отклонение между теоретическим и эмпирическим распределением для простой гипотезы
												0	0	0
0	0	4	27	2	0,27	2	0,0675	0,079951	0,321819035	32,18190354	0,83438583	0,329679954	0,27	0,059679954
6	39	8	26	6	0,26	6	0,065	0,054222	0,218251544	21,82515439	0,798589352	0,550671036	0,53	0,020671036
12	26	12	12	10	0,12	10	0,03	0,036772	0,148014043	14,80140425	0,530210895	0,698805788	0,65	0,048805788
18	13	16	9	12	0,09	12	0,0225	0,030282	0,100380306	10,03803061	0,107342525	0,798103482	0,74	0,058103482
24	14	20	10	18	0,1	18	0,025	0,016913	0,068076013	6,807601284	1,497063229	0,864664717	0,84	0,024664717
30	5	24	8	22	0,08	22	0,02	0,01147	0,046167856	4,616785606	2,479244352	0,909282047	0,92	0,010717953
36	2	42	8	33	0,08	33	0,004444	0,003942	0,080342801	8,034280123	0,000146264	0,985004423	1	0,014995577
42	1													

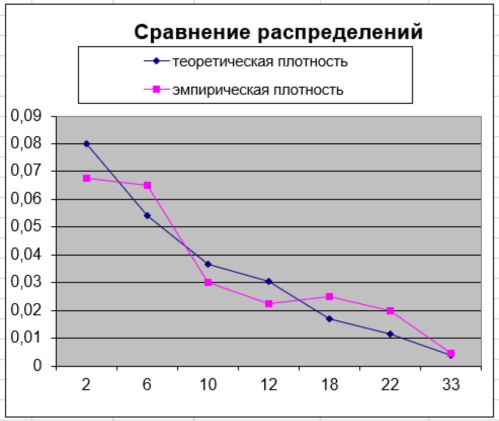
6,246982448 статистика наблюдаемая

11,07049769 статистика критическая

Вывод: гипотеза H0 принимается, выборочные данные распределены по показательному закону

Лабораторная работа 3 - Excel





Критерии, основанные на сравнении теоретической и эмпирической функции распределения вероятностей

Критерий Колмогорова - Смирнова

Исходня выборка	Ранжированная выборка							
12,35	0	1	0	0			-0,005	0,000025
0,89	0,21	2	-0,01078	0,010181	0,05	уровень значимости	0,005781	3,34204E-05
15,56	0,22	3	-0,00176	0,001132			-0,0032402	1,04991E-05
11,49	0,3	4	0,000446	-0,00129	0,026459		-0,0054455	2,96538E-05
15,69	0,44	5	-0,00305	0,001818	0,039233		-0,001954	3,81795E-06
0,21	0,6	6	-0,00824	0,006587			0,0032355	1,04682E-05
28,67	0,65	7	-0,00293	0,001156			-0,0020675	4,2744E-06
0,71	0,67	8	0,005195	-0,00702	0,039233	статистика Колмогорова-Смирнова	-0,0101952	0,000103942
3,88	0,71	9	0,011462	-0,01339	0,137477	критеческое значение статистики Колмогорова-Смирнова	-0,0164619	0,000270994
25,64	0,89	10	0,004846	-0,00722			-0,0098456	9,69353E-05
3,72	0,95	11	0,009373	-0,01189			-0,0143729	0,000206581
3,81	1,01	12	0,013933	-0,0166			-0,018933	0,00035846
0,3	1,18	13	0,008696	-0,01176			-0,0136961	0,000187582

$$D_n(\alpha)=\left(\frac{-\ln(\alpha/2)}{2n}\right)^{1/2}-\frac{1}{6n}$$

Вывод: так как 0,0392 < 0,1375, то гипотеза H0 принимается

Лабораторная работа 3 - Excel

Файл Главная Вставка Разметка страницы Формулы Данные Рецензирование Вид Справка Power Pivot Что вы хотите сделать?

C4

	A	B	C	D	E	F	P	Q
76		10,35	1,34	15	0,01459	-0,01801	-0,0195901	0,000383771
77		20,82	1,46	16	0,014158	-0,01784	-0,0191577	0,000367018
78		18,77	1,66	17	0,007046	-0,01115	-0,0120462	0,000145112
79		8,41	1,92	18	-0,00469	6,15E-05	-0,0003069	9,41683E-08
80		39,28	2,18	19	-0,01587	0,010749	0,0108746	0,000118256
81		4,07	2,36	20	-0,02022	0,014768	0,0152193	0,000231628
82		2,96	2,5	21	-0,0212	0,015504	0,0161992	0,000262415
83		0,95	2,56	22	-0,01586	0,01006	0,010858	0,000117897
84		4,83	2,59	23	-0,00818	0,002328	0,003177	1,00932E-05
85		41,99	2,96	24	-0,02621	0,019768	0,0212126	0,000449973
86		11,22	3,01	25	-0,01992	0,013401	0,0149222	0,000222673
87		12,36	3,12	26	-0,01802	0,011332	0,0130185	0,000169481
88		5,95	3,25	27	-0,01747	0,010596	0,0124726	0,000155567
89		2,18	3,39	28	-0,01752	0,010443	0,0125176	0,000156689
90		5,7	3,57	29	-0,02023	0,012908	0,0152275	0,000231877
91		8,95	3,72	30	-0,02065	0,013131	0,0156458	0,00024479
92		4,57	3,81	31	-0,01682	0,009193	0,0118221	0,000139762
93		7,62	3,87	32	-0,01091	0,003205	0,0059089	3,49152E-05
94		10,32	3,88	33	-0,00159	-0,00613	-0,0034123	1,16441E-05
95		15,87	4,07	34	-0,00436	-0,00359	-0,0006442	4,14981E-07
96		4,24	4,11	35	0,002987	-0,01098	-0,0079869	6,37911E-05
97		10,04	4,24	36	0,004424	-0,01256	-0,0094239	8,88096E-05
98		17,91	4,27	37	0,012464	-0,02063	-0,0174636	0,000304976
99		2,5	4,57	38	0,00318	-0,01167	-0,0081803	6,6918E-05
100		1,46	4,67	39	0,00688	-0,01547	-0,0118801	0,000141137
101		25,71	4,76	40	0,011263	-0,01994	-0,0162635	0,000264501
102		1,01	4,83	41	0,01693	-0,02568	-0,0219298	0,000480917
103		23,47	4,94	42	0,020181	-0,02903	-0,0251808	0,000634072
104		13,91	5,54	43	-0,00535	-0,004	0,0003534	1,24877E-07
105		18,78	5,7	44	-0,00447	-0,005	-0,0005254	2,76086E-07
106		36,67	5,78	45	0,001019	-0,01055	-0,0060193	3,62318E-05
107		2,36	5,95	46	0,001563	-0,01121	-0,0065626	4,30673E-05
108		2,59	6,36	47	-0,01059	0,000689	0,0055942	3,12949E-05
109		0,65	6,52	48	-0,009	-0,001	0,0039973	1,59782E-05

Критерий Колмогорова

0,5968 статистика Колмогорова

1,358 критеческое значение статистики Колмогорова

Вывод: так как 0,596 < 1,358, то гипотеза H0 принимается

α	0.1	0.05	0.02	0.01	0.001
$\sqrt{n}D_n(\alpha)$	1.224	1.358	1,52	1.627	1,95

Критерий Смирнова- Крамера-фон-Мизеса

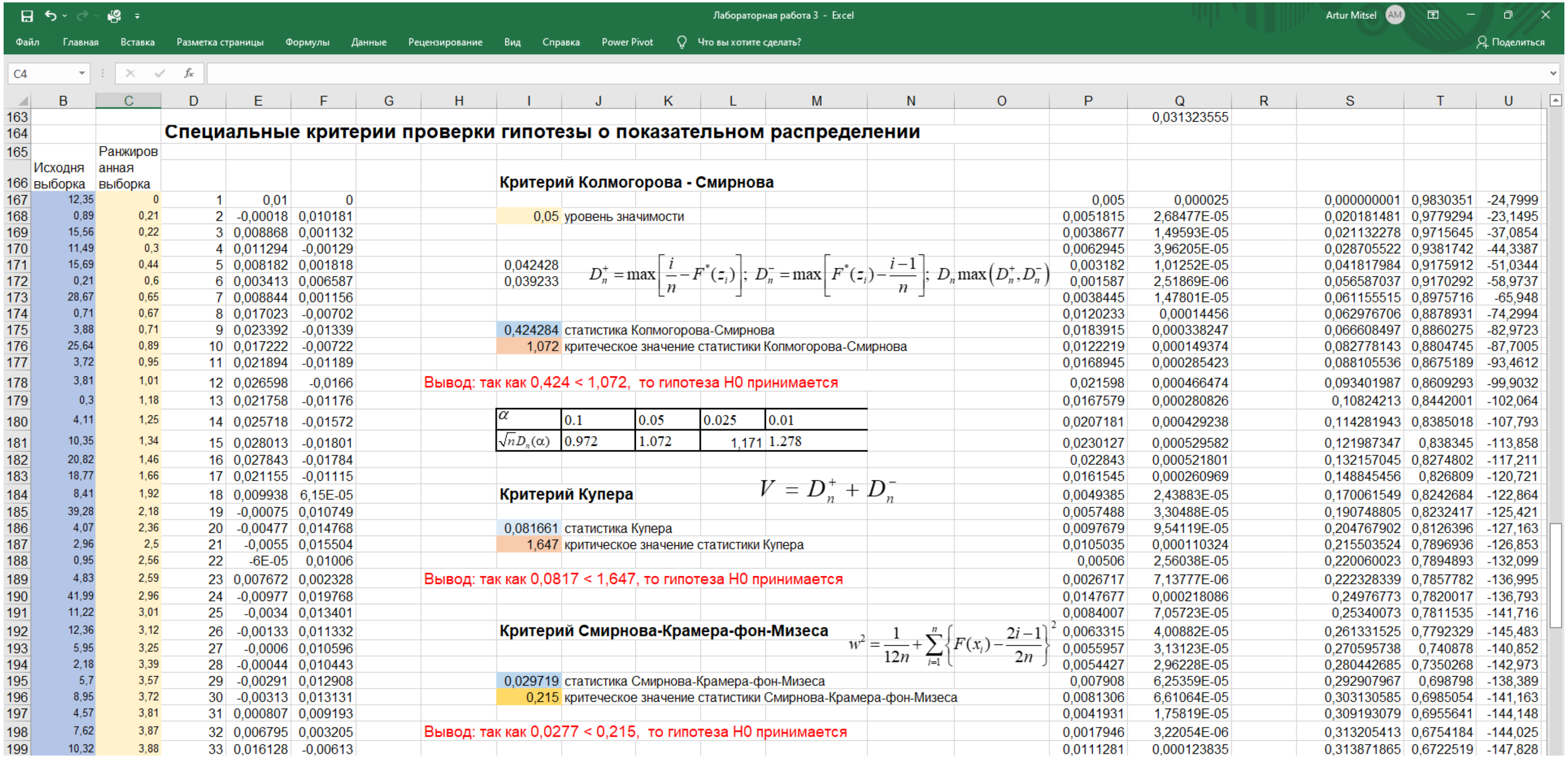
$$w^2 = \frac{1}{12n} + \sum_{i=1}^{n} \left\{ F(x_i) - \frac{2i-1}{2n} \right\}^2$$

0,032157 статистика Смирнова-Крамера-фон-Мизеса

0,461 критеческое значение статистики Смирнова-Крамера-фон-Мизеса

Вывод: так как 0,0322 < 0,461, то гипотеза H0 принимается

α	0,9	0,95	0,99	0,995	0,999
$w^2(\alpha)$	0,3473	0,4614	0,7435	0,8694	1,1679



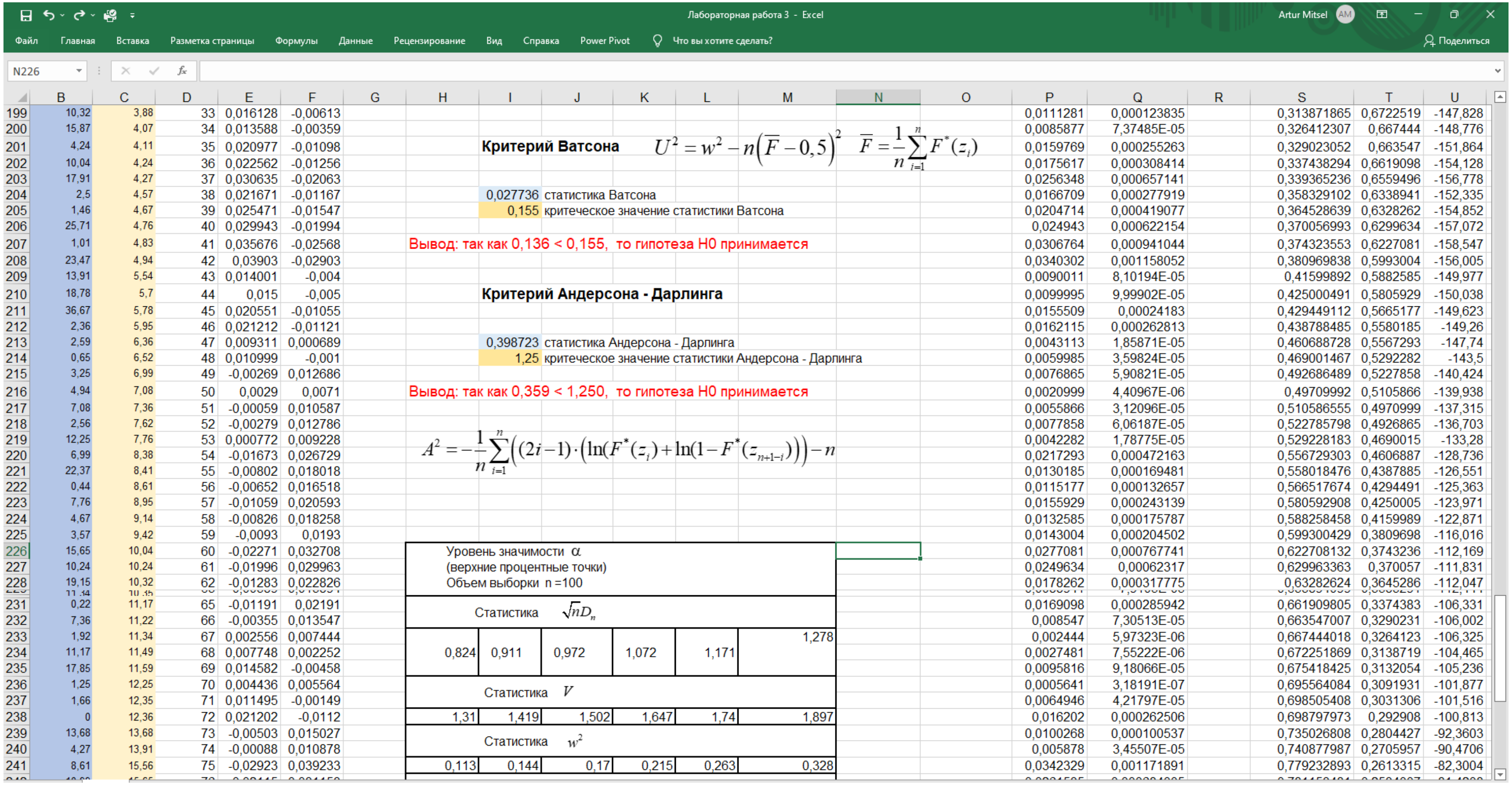
Специальные критерии проверки гипотезы о показательном распределении

	B	C	D	E	F	I–O	P	Q	S	T	U
163								0,031323555			
164			Специальные критерии проверки гипотезы о показательном распределении								
165		Ранжиров									
166	Исходня выборка	анная выборка				Критерий Колмогорова - Смирнова					
167	12,35	0	1	0,01	0		0,005	0,000025	0,000000001	0,9830351	-24,7999
168	0,89	0,21	2	-0,00018	0,010181	0,05 уровень значимости	0,0051815	2,68477E-05	0,020181481	0,9779294	-23,1495
169	15,56	0,22	3	0,008868	0,001132		0,0038677	1,49593E-05	0,021132278	0,9715645	-37,0854
170	11,49	0,3	4	0,011294	-0,00129		0,0062945	3,96205E-05	0,028705522	0,9381742	-44,3387
171	15,69	0,44	5	0,008182	0,001818	0,042428	0,003182	1,01252E-05	0,041817984	0,9175912	-51,0344
172	0,21	0,6	6	0,003413	0,006587	0,039233	0,001587	2,51869E-06	0,056587037	0,9170292	-58,9737
173	28,67	0,65	7	0,008844	0,001156		0,0038445	1,47801E-05	0,061155515	0,8975716	-65,948
174	0,71	0,67	8	0,017023	-0,00702		0,0120233	0,00014456	0,062976706	0,8878931	-74,2994
175	3,88	0,71	9	0,023392	-0,01339	0,424284 статистика Колмогорова-Смирнова	0,0183915	0,000338247	0,066608497	0,8860275	-82,9723
176	25,64	0,89	10	0,017222	-0,00722	1,072 критеческое значение статистики Колмогорова-Смирнова	0,0122219	0,000149374	0,082778143	0,8804745	-87,7005
177	3,72	0,95	11	0,021894	-0,01189		0,0168945	0,000285423	0,088105536	0,8675189	-93,4612
178	3,81	1,01	12	0,026598	-0,0166	Вывод: так как 0,424 < 1,072, то гипотеза H0 принимается	0,021598	0,000466474	0,093401987	0,8609293	-99,9032
179	0,3	1,18	13	0,021758	-0,01176		0,0167579	0,000280826	0,10824213	0,8442001	-102,064
180	4,11	1,25	14	0,025718	-0,01572	α | 0.1 | 0.05 | 0.025 | 0.01	0,0207181	0,000429238	0,114281943	0,8385018	-107,793
181	10,35	1,34	15	0,028013	-0,01801	$\sqrt{n}D_n(\alpha)$ | 0.972 | 1.072 | 1,171 | 1.278	0,0230127	0,000529582	0,121987347	0,838345	-113,858
182	20,82	1,46	16	0,027843	-0,01784		0,022843	0,000521801	0,132157045	0,8274802	-117,211
183	18,77	1,66	17	0,021155	-0,01115		0,0161545	0,000260969	0,148845456	0,826809	-120,721
184	8,41	1,92	18	0,009938	6,15E-05	Критерий Купера	0,0049385	2,43883E-05	0,170061549	0,8242684	-122,864
185	39,28	2,18	19	-0,00075	0,010749		0,0057488	3,30488E-05	0,190748805	0,8232417	-125,421
186	4,07	2,36	20	-0,00477	0,014768	0,081661 статистика Купера	0,0097679	9,54119E-05	0,204767902	0,8126396	-127,163
187	2,96	2,5	21	-0,0055	0,015504	1,647 критическое значение статистики Купера	0,0105035	0,000110324	0,215503524	0,7896936	-126,853
188	0,95	2,56	22	-6E-05	0,01006		0,00506	2,56038E-05	0,220060023	0,7894893	-132,099
189	4,83	2,59	23	0,007672	0,002328	Вывод: так как 0,0817 < 1,647, то гипотеза H0 принимается	0,0026717	7,13777E-06	0,222328339	0,7857782	-136,995
190	41,99	2,96	24	-0,00977	0,019768		0,0147677	0,000218086	0,24976773	0,7820017	-136,793
191	11,22	3,01	25	-0,0034	0,013401		0,0084007	7,05723E-05	0,25340073	0,7811535	-141,716
192	12,36	3,12	26	-0,00133	0,011332	Критерий Смирнова-Крамера-фон-Мизеса	0,0063315	4,00882E-05	0,261331525	0,7792329	-145,483
193	5,95	3,25	27	-0,0006	0,010596		0,0055957	3,13123E-05	0,270595738	0,740878	-140,852
194	2,18	3,39	28	-0,00044	0,010443		0,0054427	2,96228E-05	0,280442685	0,7350268	-142,973
195	5,7	3,57	29	-0,00291	0,012908	0,029719 статистика Смирнова-Крамера-фон-Мизеса	0,007908	6,25359E-05	0,292907967	0,698798	-138,389
196	8,95	3,72	30	-0,00313	0,013131	0,215 критеческое значение статистики Смирнова-Крамера-фон-Мизеса	0,0081306	6,61064E-05	0,303130585	0,6985054	-141,163
197	4,57	3,81	31	0,000807	0,009193		0,0041931	1,75819E-05	0,309193079	0,6955641	-144,148
198	7,62	3,87	32	0,006795	0,003205	Вывод: так как 0,0277 < 0,215, то гипотеза H0 принимается	0,0017946	3,22054E-06	0,313205413	0,6754184	-144,025
199	10,32	3,88	33	0,016128	-0,00613		0,0111281	0,000123835	0,313871865	0,6722519	-147,828

$$D_n^+ = \max\left[\frac{i}{n} - F^*(z_i)\right];\; D_n^- = \max\left[F^*(z_i) - \frac{i-1}{n}\right];\; D_n \max\left(D_n^+, D_n^-\right)$$

$$V = D_n^+ + D_n^-$$

$$w^2 = \frac{1}{12n} + \sum_{i=1}^{n}\left\{F(x_i) - \frac{2i-1}{2n}\right\}^2$$



Уровень значимости α (верхние процентные точки) Объем выборки n =100					
Статистика $\sqrt{n}D_n$					
0,824	0,911	0,972	1,072	1,171	1,278
Статистика V					
1,31	1,419	1,502	1,647	1,74	1,897
Статистика w^2					
0,113	0,144	0,17	0,215	0,263	0,328

	B	C	D	E	F	P	Q	S	T	U
199	10,32	3,88	33	0,016128	-0,00613	0,0111281	0,000123835	0,313871865	0,6722519	-147,828
200	15,87	4,07	34	0,013588	-0,00359	0,0085877	7,37485E-05	0,326412307	0,667444	-148,776
201	4,24	4,11	35	0,020977	-0,01098	0,0159769	0,000255263	0,329023052	0,663547	-151,864
202	10,04	4,24	36	0,022562	-0,01256	0,0175617	0,000308414	0,337438294	0,6619098	-154,128
203	17,91	4,27	37	0,030635	-0,02063	0,0256348	0,000657141	0,339365236	0,6559496	-156,778
204	2,5	4,57	38	0,021671	-0,01167	0,0166709	0,000277919	0,358329102	0,6338941	-152,335
205	1,46	4,67	39	0,025471	-0,01547	0,0204714	0,000419077	0,364528639	0,6328262	-154,852
206	25,71	4,76	40	0,029943	-0,01994	0,024943	0,000622154	0,370056993	0,6299634	-157,072
207	1,01	4,83	41	0,035676	-0,02568	0,0306764	0,000941044	0,374323553	0,6227081	-158,547
208	23,47	4,94	42	0,03903	-0,02903	0,0340302	0,001158052	0,380969838	0,5993004	-156,005
209	13,91	5,54	43	0,014001	-0,004	0,0090011	8,10194E-05	0,41599892	0,5882585	-149,977
210	18,78	5,7	44	0,015	-0,005	0,0099995	9,99902E-05	0,425000491	0,5805929	-150,038
211	36,67	5,78	45	0,020551	-0,01055	0,0155509	0,00024183	0,429449112	0,5665177	-149,623
212	2,36	5,95	46	0,021212	-0,01121	0,0162115	0,000262813	0,438788485	0,5580185	-149,26
213	2,59	6,36	47	0,009311	0,000689	0,0043113	1,85871E-05	0,460688728	0,5567293	-147,74
214	0,65	6,52	48	0,010999	-0,001	0,0059985	3,59824E-05	0,469001467	0,5292282	-143,5
215	3,25	6,99	49	-0,00269	0,012686	0,0076865	5,90821E-05	0,492686489	0,5227858	-140,424
216	4,94	7,08	50	0,0029	0,0071	0,0020999	4,40967E-06	0,49709992	0,5105866	-139,938
217	7,08	7,36	51	-0,00059	0,010587	0,0055866	3,12096E-05	0,510586555	0,4970999	-137,315
218	2,56	7,62	52	-0,00279	0,012786	0,0077858	6,06187E-05	0,522785798	0,4926865	-136,703
219	12,25	7,76	53	0,000772	0,009228	0,0042282	1,78775E-05	0,529228183	0,4690015	-133,28
220	6,99	8,38	54	-0,01673	0,026729	0,0217293	0,000472163	0,556729303	0,4606887	-128,736
221	22,37	8,41	55	-0,00802	0,018018	0,0130185	0,000169481	0,558018476	0,4387885	-126,551
222	0,44	8,61	56	-0,00652	0,016518	0,0115177	0,000132657	0,566517674	0,4294491	-125,363
223	7,76	8,95	57	-0,01059	0,020593	0,0155929	0,000243139	0,580592908	0,4250005	-123,971
224	4,67	9,14	58	-0,00826	0,018258	0,0132585	0,000175787	0,588258458	0,4159989	-122,871
225	3,57	9,42	59	-0,0093	0,0193	0,0143004	0,000204502	0,599300429	0,3809698	-116,016
226	15,65	10,04	60	-0,02271	0,032708	0,0277081	0,000767741	0,622708132	0,3743236	-112,169
227	10,24	10,24	61	-0,01996	0,029963	0,0249634	0,00062317	0,629963363	0,370057	-111,831
228	19,15	10,32	62	-0,01283	0,022826	0,0178262	0,000317775	0,63282624	0,3645286	-112,047
231	0,22	11,17	65	-0,01191	0,02191	0,0169098	0,000285942	0,661909805	0,3374383	-106,331
232	7,36	11,22	66	-0,00355	0,013547	0,008547	7,30513E-05	0,663547007	0,3290231	-106,002
233	1,92	11,34	67	0,002556	0,007444	0,002444	5,97323E-06	0,667444018	0,3264123	-106,325
234	11,17	11,49	68	0,007748	0,002252	0,0027481	7,55222E-06	0,672251869	0,3138719	-104,465
235	17,85	11,59	69	0,014582	-0,00458	0,0095816	9,18066E-05	0,675418425	0,3132054	-105,236
236	1,25	12,25	70	0,004436	0,005564	0,0005641	3,18191E-07	0,695564084	0,3091931	-101,877
237	1,66	12,35	71	0,011495	-0,00149	0,0064946	4,21797E-05	0,698505408	0,3031306	-101,516
238	0	12,36	72	0,021202	-0,0112	0,016202	0,000262506	0,698797973	0,292908	-100,813
239	13,68	13,68	73	-0,00503	0,015027	0,0100268	0,000100537	0,735026808	0,2804427	-92,3603
240	4,27	13,91	74	-0,00088	0,010878	0,005878	3,45507E-05	0,740877987	0,2705957	-90,4706
241	8,61	15,56	75	-0,02923	0,039233	0,0342329	0,001171891	0,779232893	0,2613315	-82,3004

242	10,99	15,65	76	-0,02115	0,031153	0,0261535	0,000684005	0,781153481	0,2534007	-81,4208
243	21,88	15,69	77	-0,012	0,022002	0,0170017	0,000289058	0,782001704	0,2497677	-81,5904
244	4,76	15,87	78	-0,00578	0,015778	0,0107782	0,00011617	0,785778207	0,2223283	-76,3424
245	1,18	16,05	79	0,000511	0,009489	0,0044893	2,01537E-05	0,789489287	0,22006	-76,1305
246	3,01	16,06	80	0,010306	-0,00031	0,0053064	2,81583E-05	0,789693562	0,2155035	-76,1329
247	3,87	17,25	81	-0,00264	0,01264	0,0076396	5,8363E-05	0,812639567	0,2047679	-70,2908
248	8,38	17,85	82	-0,00324	0,013242	0,0082417	6,79263E-05	0,823241739	0,1907488	-66,2027
249	9,42	17,91	83	0,005732	0,004268	0,0007316	5,35263E-07	0,824268383	0,1700615	-62,6444
250	16,06	18,06	84	0,013191	-0,00319	0,008191	6,70928E-05	0,826808982	0,1488455	-58,6743
251	16,05	18,1	85	0,02252	-0,01252	0,0175198	0,000306942	0,82748025	0,132157	-55,9584
252	6,36	18,77	86	0,021655	-0,01166	0,016655	0,00027739	0,838344975	0,1219873	-52,3978
253	5,54	18,78	87	0,031498	-0,0215	0,0264982	0,000702152	0,838501843	0,1142819	-51,4667
254	0,6	19,15	88	0,0358	-0,0258	0,0307999	0,000948631	0,844200149	0,1082421	-49,6871
255	1,34	20,32	89	0,029071	-0,01907	0,0240707	0,000579398	0,860929305	0,093402	-43,8604
256	18,1	20,82	90	0,032481	-0,02248	0,0274811	0,000755209	0,86751893	0,0881055	-41,9485
257	20,32	21,88	91	0,029525	-0,01953	0,0245255	0,000601498	0,880474544	0,0827781	-38,6797
258	0,67	22,37	92	0,033973	-0,02397	0,0289725	0,000839408	0,886027456	0,0666085	-34,7586
259	3,12	22,54	93	0,042107	-0,03211	0,0371069	0,001376924	0,887893077	0,0629767	-34,031
260	5,78	23,47	94	0,042428	-0,03243	0,0374284	0,001400883	0,897571625	0,0611555	-32,0084
261	18,06	25,64	95	0,032971	-0,02297	0,0279708	0,000782366	0,917029202	0,056587	-27,3799
262	22,54	25,71	96	0,042409	-0,03241	0,0374088	0,001399422	0,917591157	0,041818	-24,5857
263	11,59	28,67	97	0,031826	-0,02183	0,0268258	0,000719622	0,938174227	0,0287055	-17,9384
264	3,39	36,67	98	0,008435	0,001565	0,0034355	1,18026E-05	0,971564508	0,0211323	-9,79024
265	9,14	39,28	99	0,012071	-0,00207	0,0070706	4,99931E-05	0,97792942	0,0201815	-8,41302
266	6,52	41,99	100	0,016965	-0,00696	0,0119649	0,000143158	0,983035136	0,00001	-3,40696
267										
268						1,3826232	0,028885758	0,495547189		-100,399
269										

Статистика U^2					
0,089	0,11	0,126	0,155	0,184	0,229
Статистика A^2					
0,71	0,875	1,008	1,25	1,51	1,865

Критерий Фроцини

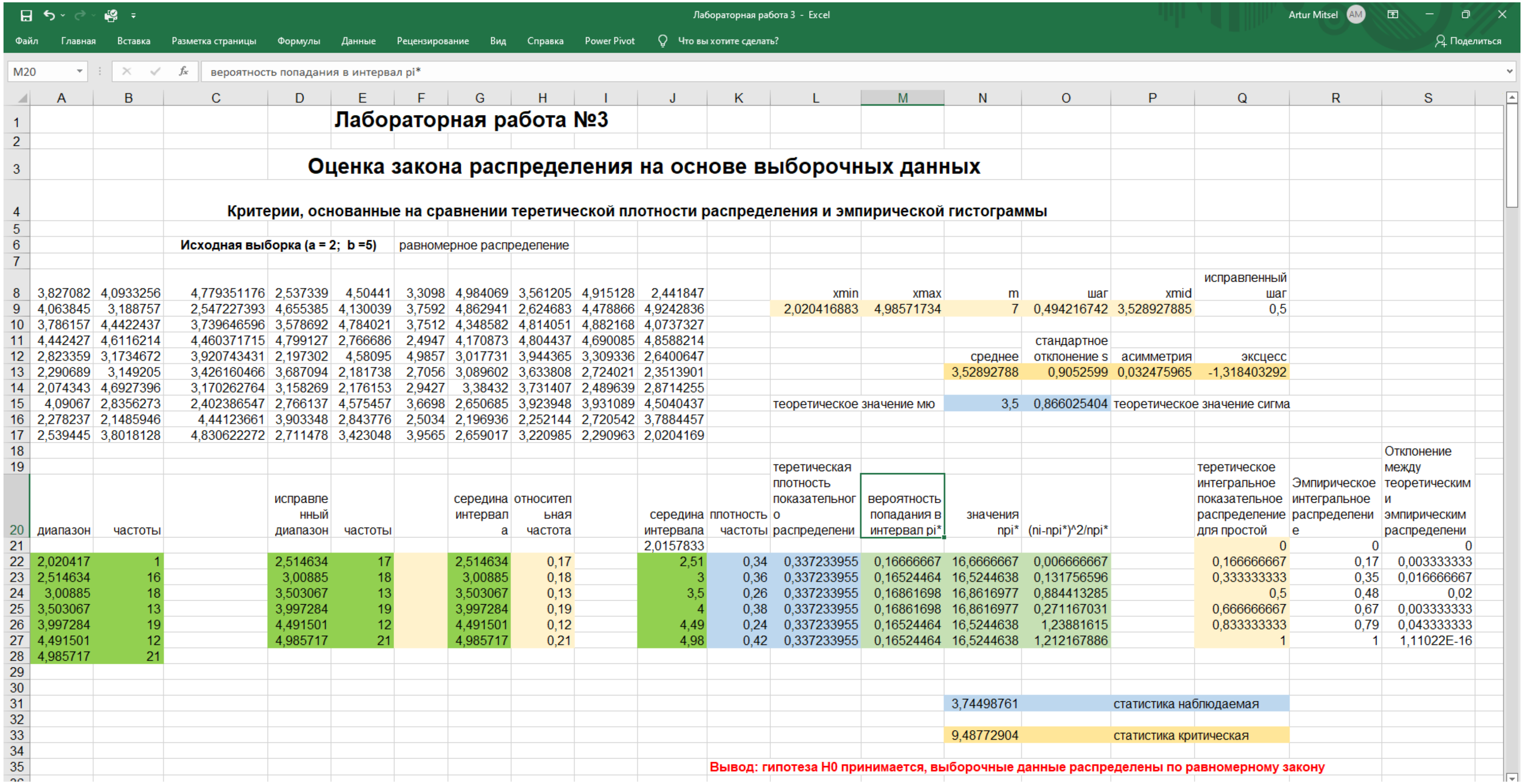
$$B_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left| 1 - EXP(-\hat{\lambda} \cdot x_i) - \frac{2i-1}{2} \right|$$

0,138262 статистика Фроцини

0,384 критеческое значение статистики Фроцини

Вывод: так как 0,138 < 0,384, то гипотеза H0 принимается

n	α		
	0,9	0,95	0,99
∞	0,338	0,384	0,476



Лабораторная работа №3

Оценка закона распределения на основе выборочных данных

Критерии, основанные на сравнении теретической плотности распределения и эмпирической гистограммы

Исходная выборка (a = 2; b =5) равномерное распределение

A	B	C	D	E	F	G	H	I	J
3,827082	4,0933256	4,779351176	2,537339	4,50441	3,3098	4,984069	3,561205	4,915128	2,441847
4,063845	3,188757	2,547227393	4,655385	4,130039	3,7592	4,862941	2,624683	4,478866	4,9242836
3,786157	4,4422437	3,739646596	3,578692	4,784021	3,7512	4,348582	4,814051	4,882168	4,0737327
4,442427	4,6116214	4,460371715	4,799127	2,766686	2,4947	4,170873	4,804437	4,690085	4,8588214
2,823359	3,1734672	3,920743431	2,197302	4,58095	4,9857	3,017731	3,944365	3,309336	2,6400647
2,290689	3,149205	3,426160466	3,687094	2,181738	2,7056	3,089602	3,633808	2,724021	2,3513901
2,074343	4,6927396	3,170262764	3,158269	2,176153	2,9427	3,38432	3,731407	2,489639	2,8714255
4,09067	2,8356273	2,402386547	2,766137	4,575457	3,6698	2,650685	3,923948	3,931089	4,5040437
2,278237	2,1485946	4,44123661	3,903348	2,843776	2,5034	2,196936	2,252144	2,720542	3,7884457
2,539445	3,8018128	4,830622272	2,711478	3,423048	3,9565	2,659017	3,220985	2,290963	2,0204169

xmin	xmax	m	шаг	xmid	исправленный шаг
2,020416883	4,98571734	7	0,494216742	3,528927885	0,5

среднее	стандартное отклонение s	асимметрия	эксцесс
3,52892788	0,9052599	0,032475965	-1,318403292

теоретическое значение мю 3,5 0,866025404 теоретическое значение сигма

диапазон	частоты	исправленный диапазон	частоты	середина интервала	относительная частота
2,020417	1	2,514634	17	2,514634	0,17
2,514634	16	3,00885	18	3,00885	0,18
3,00885	18	3,503067	13	3,503067	0,13
3,503067	13	3,997284	19	3,997284	0,19
3,997284	19	4,491501	12	4,491501	0,12
4,491501	12	4,985717	21	4,985717	0,21
4,985717	21				

середина интервала	плотность частоты	теретическая плотность показательного распределени	вероятность попадания в интервал pi*	значения npi*	(ni-npi*)^2/npi*	теретическое интегральное показательное распределение для простой	Эмпирическое интегральное распределение	Отклонение между теоретическим и эмпирическим распределени
2,0157833						0	0	0
2,51	0,34	0,337233955	0,16666667	16,6666667	0,006666667	0,166666667	0,17	0,003333333
3	0,36	0,337233955	0,16524464	16,5244638	0,131756596	0,333333333	0,35	0,016666667
3,5	0,26	0,337233955	0,16861698	16,8616977	0,884413285	0,5	0,48	0,02
4	0,38	0,337233955	0,16861698	16,8616977	0,271167031	0,666666667	0,67	0,003333333
4,49	0,24	0,337233955	0,16524464	16,5244638	1,23881615	0,833333333	0,79	0,043333333
4,98	0,42	0,337233955	0,16524464	16,5244638	1,212167886	1	1	1,11022E-16

3,74498761 статистика наблюдаемая

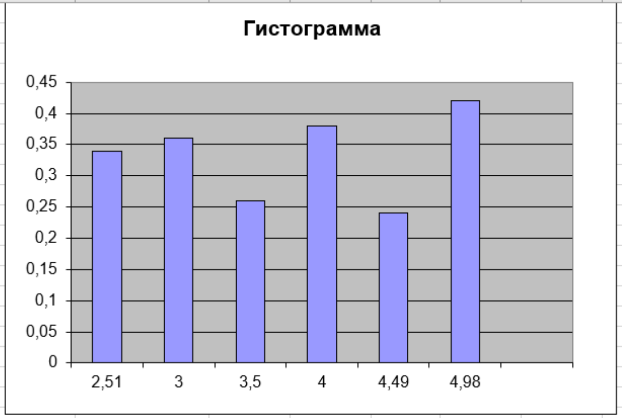
9,48772904 статистика критическая

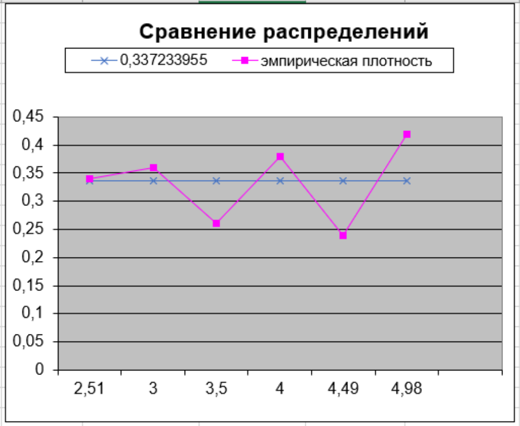
Вывод: гипотеза H0 принимается, выборочные данные распределены по равномерному закону

Лабораторная работа 3 - Excel

Файл Главная Вставка Разметка страницы Формулы Данные Рецензирование Вид Справка Power Pivot Что вы хотите сделать?

M20 | вероятность попадания в интервал pi*





Критерии, основанные на сравнении теоретической и эмпирической функции распределения вероятностей

Исходня выборка	Ранжированная выборка					
3,8270821	2,020416883	1	-0,00681	0,0068	0,001805628	3,26029E-06
4,0638447	2,074343089	2	-0,01478	0,0148	0,00978103	9,56685E-05
3,7861568	2,148594623	3	-0,02953	0,0295	0,024531541	0,000601796
4,4424268	2,176152837	4	-0,02872	0,0287	0,023717612	0,000562525
2,8233589	2,181737724	5	-0,02058	0,0206	0,015579241	0,000242713
2,2906888	2,196935942	6	-0,01565	0,0156	0,010645314	0,000113323
2,0743431	2,197302164	7	-0,00577	0,0058	0,000767388	5,88884E-07
4,0906705	2,252143925	8	-0,01405	0,014	0,009047975	8,18659E-05
2,2782373	2,278237251	9	-0,01275	0,0127	0,00774575	5,99966E-05
2,5394452	2,290688803	10	-0,0069	0,0069	0,001896268	3,59583E-06
4,0933256	2,290963469	11	0,003012	-0,003	-0,00801218	6,4195E-05
3,188757	2,351390118	12	-0,00713	0,0071	0,002130039	4,53707E-06
4,4422437	2,402386547	13	-0,01413	0,0141	0,009128849	8,33359E-05
4,6116214	2,44184698	14	-0,01728	0,0173	0,012282327	0,000150856

Критерий Колмогорова - Смирнова

0,05 уровень значимости

0,039525

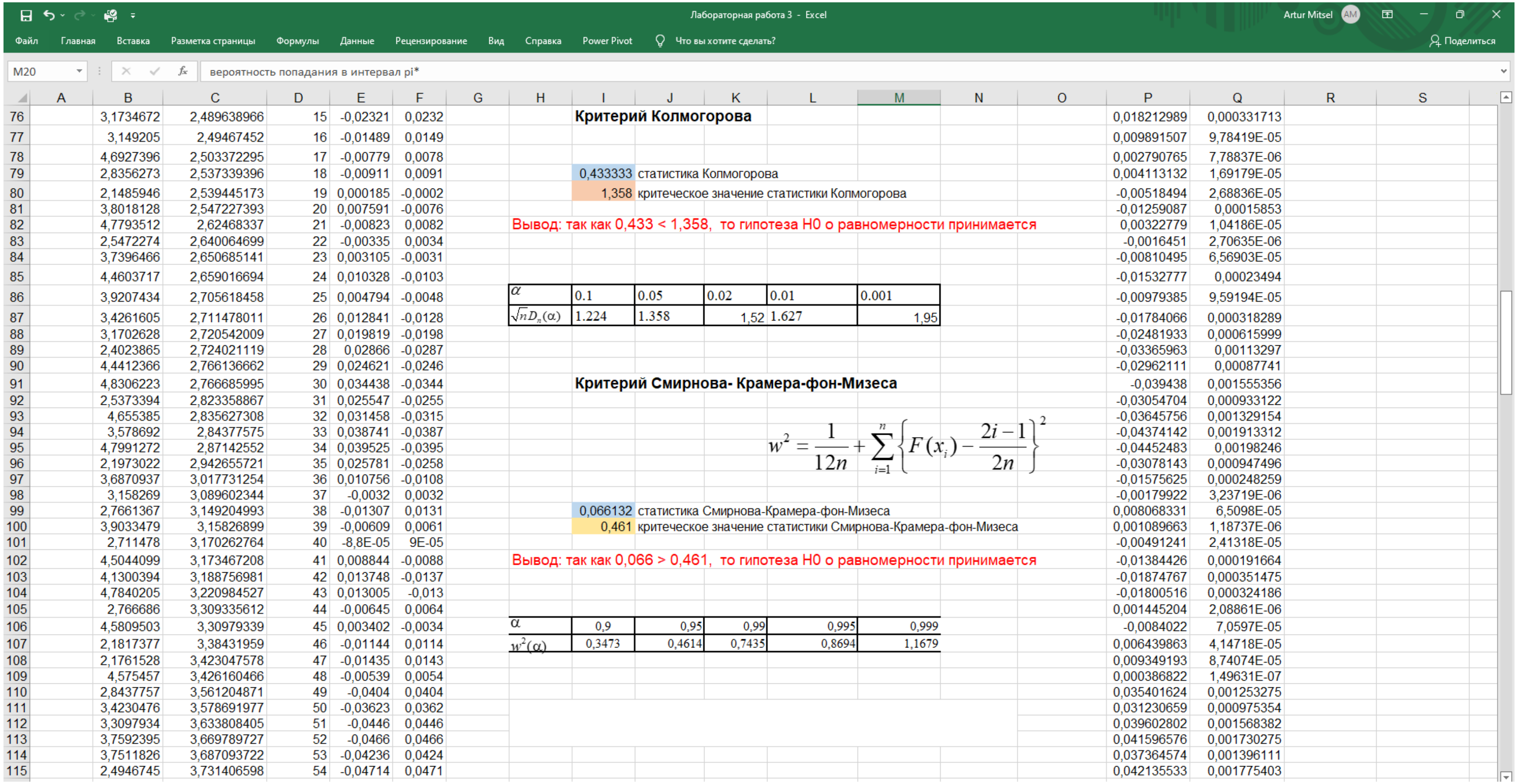
0,073746

$$D_n(\alpha) = \left(\frac{-\ln(\alpha/2)}{2n} \right)^{1/2} - \frac{1}{6n}$$

0,073746 статистика Колмогорова-Смирнова

0,137477 критеческое значение статистики Колмогорова-Смирнова

Вывод: так как 0,074 < 0,137, то гипотеза H0 о равномерности принимается



Лабораторная работа 3 - Excel

Файл Главная Вставка Разметка страницы Формулы Данные Рецензирование Вид Справка Power Pivot Что вы хотите сделать?

M20 | вероятность попадания в интервал pi*

	A	B	C	D	E	F	P	Q
76		3,1734672	2,489638966	15	-0,02321	0,0232	0,018212989	0,000331713
77		3,149205	2,49467452	16	-0,01489	0,0149	0,009891507	9,78419E-05
78		4,6927396	2,503372295	17	-0,00779	0,0078	0,002790765	7,78837E-06
79		2,8356273	2,537339396	18	-0,00911	0,0091	0,004113132	1,69179E-05
80		2,1485946	2,539445173	19	0,000185	-0,0002	-0,00518494	2,68836E-05
81		3,8018128	2,547227393	20	0,007591	-0,0076	-0,01259087	0,00015853
82		4,7793512	2,62468337	21	-0,00823	0,0082	0,00322779	1,04186E-05
83		2,5472274	2,640064699	22	-0,00335	0,0034	-0,0016451	2,70635E-06
84		3,7396466	2,650685141	23	0,003105	-0,0031	-0,00810495	6,56903E-05
85		4,4603717	2,659016694	24	0,010328	-0,0103	-0,01532777	0,00023494
86		3,9207434	2,705618458	25	0,004794	-0,0048	-0,00979385	9,59194E-05
87		3,4261605	2,711478011	26	0,012841	-0,0128	-0,01784066	0,000318289
88		3,1702628	2,720542009	27	0,019819	-0,0198	-0,02481933	0,000615999
89		2,4023865	2,724021119	28	0,02866	-0,0287	-0,03365963	0,00113297
90		4,4412366	2,766136662	29	0,024621	-0,0246	-0,02962111	0,00087741
91		4,8306223	2,766685995	30	0,034438	-0,0344	-0,039438	0,001555356
92		2,5373394	2,823358867	31	0,025547	-0,0255	-0,03054704	0,000933122
93		4,655385	2,835627308	32	0,031458	-0,0315	-0,03645756	0,001329154
94		3,578692	2,84377575	33	0,038741	-0,0387	-0,04374142	0,001913312
95		4,7991272	2,87142552	34	0,039525	-0,0395	-0,04452483	0,00198246
96		2,1973022	2,942655721	35	0,025781	-0,0258	-0,03078143	0,000947496
97		3,6870937	3,017731254	36	0,010756	-0,0108	-0,01575625	0,000248259
98		3,158269	3,089602344	37	-0,0032	0,0032	-0,00179922	3,23719E-06
99		2,7661367	3,149204993	38	-0,01307	0,0131	0,008068331	6,5098E-05
100		3,9033479	3,15826899	39	-0,00609	0,0061	0,001089663	1,18737E-06
101		2,711478	3,170262764	40	-8,8E-05	9E-05	-0,00491241	2,41318E-05
102		4,5044099	3,173467208	41	0,008844	-0,0088	-0,01384426	0,000191664
103		4,1300394	3,188756981	42	0,013748	-0,0137	-0,01874767	0,000351475
104		4,7840205	3,220984527	43	0,013005	-0,013	-0,01800516	0,000324186
105		2,766686	3,309335612	44	-0,00645	0,0064	0,001445204	2,08861E-06
106		4,5809503	3,30979339	45	0,003402	-0,0034	-0,0084022	7,0597E-05
107		2,1817377	3,38431959	46	-0,01144	0,0114	0,006439863	4,14718E-05
108		2,1761528	3,423047578	47	-0,01435	0,0143	0,009349193	8,74074E-05
109		4,575457	3,426160466	48	-0,00539	0,0054	0,000386822	1,49631E-07
110		2,8437757	3,561204871	49	-0,0404	0,0404	0,035401624	0,001253275
111		3,4230476	3,578691977	50	-0,03623	0,0362	0,031230659	0,000975354
112		3,3097934	3,633808405	51	-0,0446	0,0446	0,039602802	0,001568382
113		3,7592395	3,669789727	52	-0,0466	0,0466	0,041596576	0,001730275
114		3,7511826	3,687093722	53	-0,04236	0,0424	0,037364574	0,001396111
115		2,4946745	3,731406598	54	-0,04714	0,0471	0,042135533	0,001775403

Критерий Колмогорова

0,433333 статистика Колмогорова

1,358 критеческое значение статистики Колмогорова

Вывод: так как 0,433 < 1,358, то гипотеза H0 о равномерности принимается

α	0.1	0.05	0.02	0.01	0.001
$\sqrt{n}D_n(\alpha)$	1.224	1.358	1,52	1.627	1,95

Критерий Смирнова- Крамера-фон-Мизеса

$$w^2 = \frac{1}{12n} + \sum_{i=1}^{n}\left\{F(x_i) - \frac{2i-1}{2n}\right\}^2$$

0,066132 статистика Смирнова-Крамера-фон-Мизеса

0,461 критеческое значение статистики Смирнова-Крамера-фон-Мизеса

Вывод: так как 0,066 > 0,461, то гипотеза H0 о равномерности принимается

α	0,9	0,95	0,99	0,995	0,999
$w^2(\alpha)$	0,3473	0,4614	0,7435	0,8694	1,1679

Лабораторная работа 3 - Excel

M20 — вероятность попадания в интервал pi*

Специальные критерии проверки гипотезы о равномерном распределении

	Исходня выборка	Ранжированная выборка				
167	3,8270821	2,020416883	1	0,01	0	0,013185748
168	4,0638447	2,074343089	2	0,001814	0,0082	0,028225886
169	3,7861568	2,148594623	3	-0,01323	0,0232	0,027519452
170	4,4424268	2,176152837	4	-0,01252	0,0225	0,019402865
171	2,8233589	2,181737724	5	-0,0044	0,0144	0,01452822
172	2,2906888	2,196935942	6	0,000472	0,0095	0,004651723
173	2,0743431	2,197302164	7	0,010348	-0,0003	0,013146227
174	4,0906705	2,252143925	8	0,001854	0,0081	0,011945782
175	2,2782373	2,278237251	9	0,003054	0,0069	0,006144868
176	2,5394452	2,290688803	10	0,008855	0,0011	0,003762505
177	4,0933256	2,290963469	11	0,018763	-0,0088	0,006615413
178	3,188757	2,351390118	12	0,008385	0,0016	0,013813141
179	4,4422437	2,402386547	13	0,001187	0,0088	0,017120538
180	4,6116214	2,44184698	14	-0,00212	0,0121	0,023237619
181	3,1734672	2,489638966	15	-0,00824	0,0182	0,014935779
182	3,149205	2,49467452	16	6,42E-05	0,0099	0,007868964
183	4,6927396	2,503372295	17	0,007131	0,0029	0,009323824
184	2,8356273	2,537339396	18	0,005676	0,0043	3,39632E-05
185	2,1485946	2,539445173	19	0,014966	-0,005	0,007341608
186	3,8018128	2,547227393	20	0,022342	-0,0123	0,008779177
187	4,7793512	2,62468337	21	0,006221	0,0038	0,003966284
188	2,5472274	2,640064699	22	0,011034	-0,001	0,002452143
189	3,7396466	2,650685141	23	0,017452	-0,0075	0,00964246
190	4,4603717	2,659016694	24	0,024642	-0,0146	0,003926763
191	3,9207434	2,705618458	25	0,018927	-0,0089	0,011950722
192	3,4261605	2,711478011	26	0,026951	-0,017	0,018894035
193	3,1702628	2,720542009	27	0,033894	-0,0239	0,027720761
194	2,4023865	2,724021119	28	0,042721	-0,0327	0,02351797
195	4,4412366	2,766136662	29	0,038518	-0,0285	0,033332716
196	4,8306223	2,766685995	30	0,048333	-0,0383	0,024220699
197	2,5373394	2,823358867	31	0,039221	-0,0292	0,030083364
198	4,655385	2,835627308	32	0,045083	-0,0351	0,037335433
199	3,578692	2,84377575	33	0,052335	-0,0423	0,038010992

Критерий типа Колмогорова – Смирнова

0,05 уровень значимости

$$D^{+}=\max_i\left(U_i-\frac{i}{n+1}\right);\ D^{-}=\max_i\left(\frac{i}{n+1}-U_i\right) \qquad D=\max\left(D^{+},D^{-}\right)$$

0,053011
0,076383
0,076383

$$\widetilde{D}=\left(D+\frac{0,4}{n}\right)\cdot\left(\sqrt{n}+0,2+\frac{0,68}{\sqrt{n}}\right)$$

0,825369 статистика Колмогорова-Смирнова
1,358 критеческое значение статистики Колмогорова-Смирнова

Вывод: так как 0,825 < 1,358, то гипотеза H0 о равномерности принимается

	Уровень значимости α				
	0,01	0,025	0,05	0,1	0,15
$\widetilde{D}$	1,628	1,48	1,358	1,224	1,138
$\widetilde{V}$	2,001	1,852	1,47	1,62	1,537

Критерий Купера

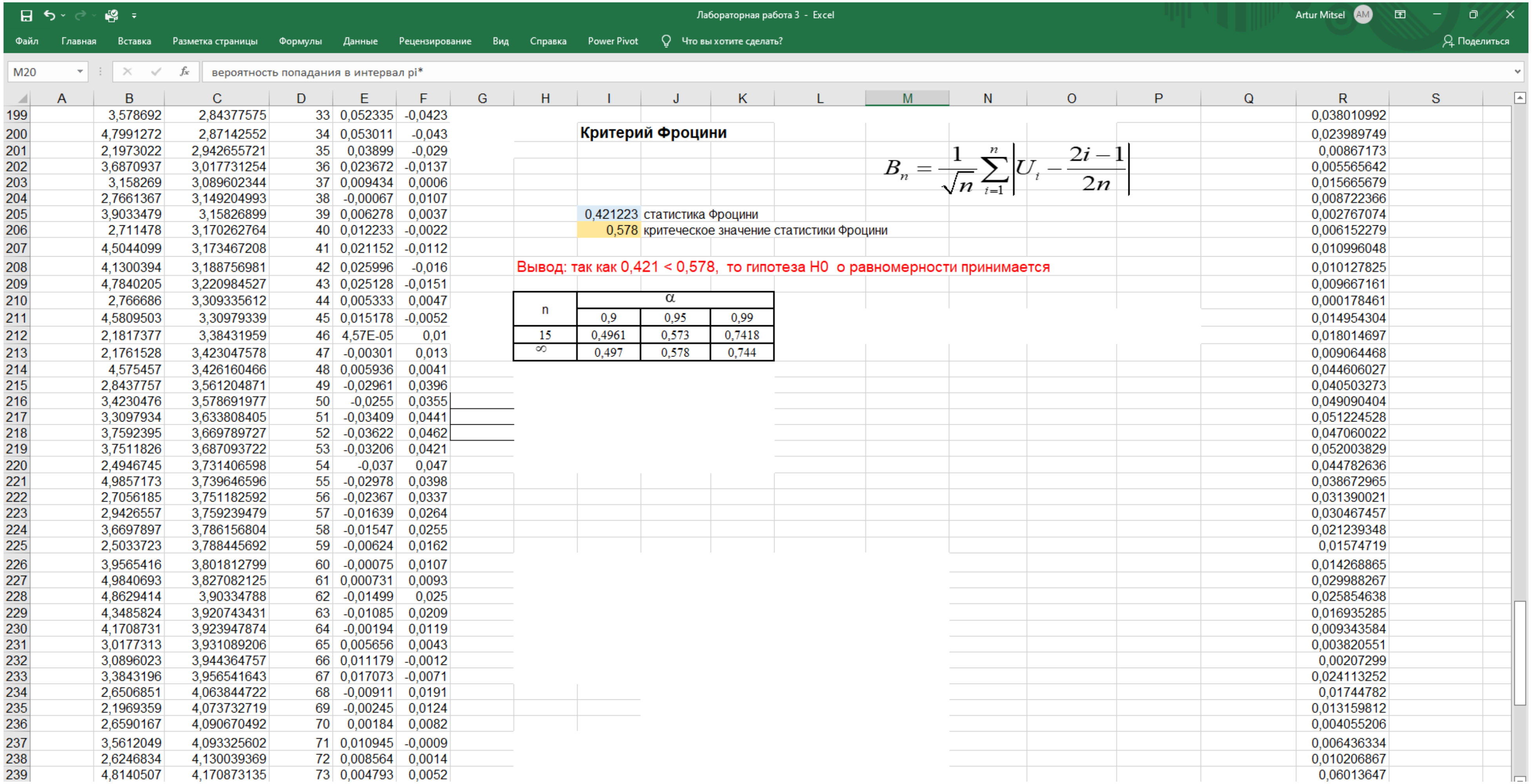
$$V=D_n^{+}+D_n^{-}$$

$$\widetilde{V}=\left(V-\frac{1}{n-1}\right)\cdot\left(\sqrt{n+1}+0,1555+\frac{0,24}{\sqrt{n+1}}\right)$$

0,129394

1,220275 статистика Купера
1,47 критеческое значение статистики Купера

Вывод: так как 1,220 < 1,470, то гипотеза H0 о равномерности принимается



	A	B	C	D	E	F	R
199		3,578692	2,84377575	33	0,052335	-0,0423	0,038010992
200		4,7991272	2,87142552	34	0,053011	-0,043	0,023989749
201		2,1973022	2,942655721	35	0,03899	-0,029	0,00867173
202		3,6870937	3,017731254	36	0,023672	-0,0137	0,005565642
203		3,158269	3,089602344	37	0,009434	0,0006	0,015665679
204		2,7661367	3,149204993	38	-0,00067	0,0107	0,008722366
205		3,9033479	3,15826899	39	0,006278	0,0037	0,002767074
206		2,711478	3,170262764	40	0,012233	-0,0022	0,006152279
207		4,5044099	3,173467208	41	0,021152	-0,0112	0,010996048
208		4,1300394	3,188756981	42	0,025996	-0,016	0,010127825
209		4,7840205	3,220984527	43	0,025128	-0,0151	0,009667161
210		2,766686	3,309335612	44	0,005333	0,0047	0,000178461
211		4,5809503	3,30979339	45	0,015178	-0,0052	0,014954304
212		2,1817377	3,38431959	46	4,57E-05	0,01	0,018014697
213		2,1761528	3,423047578	47	-0,00301	0,013	0,009064468
214		4,575457	3,426160466	48	0,005936	0,0041	0,044606027
215		2,8437757	3,561204871	49	-0,02961	0,0396	0,040503273
216		3,4230476	3,578691977	50	-0,0255	0,0355	0,049090404
217		3,3097934	3,633808405	51	-0,03409	0,0441	0,051224528
218		3,7592395	3,669789727	52	-0,03622	0,0462	0,047060022
219		3,7511826	3,687093722	53	-0,03206	0,0421	0,052003829
220		2,4946745	3,731406598	54	-0,037	0,047	0,044782636
221		4,9857173	3,739646596	55	-0,02978	0,0398	0,038672965
222		2,7056185	3,751182592	56	-0,02367	0,0337	0,031390021
223		2,9426557	3,759239479	57	-0,01639	0,0264	0,030467457
224		3,6697897	3,786156804	58	-0,01547	0,0255	0,021239348
225		2,5033723	3,788445692	59	-0,00624	0,0162	0,01574719
226		3,9565416	3,801812799	60	-0,00075	0,0107	0,014268865
227		4,9840693	3,827082125	61	0,000731	0,0093	0,029988267
228		4,8629414	3,90334788	62	-0,01499	0,025	0,025854638
229		4,3485824	3,920743431	63	-0,01085	0,0209	0,016935285
230		4,1708731	3,923947874	64	-0,00194	0,0119	0,009343584
231		3,0177313	3,931089206	65	0,005656	0,0043	0,003820551
232		3,0896023	3,944364757	66	0,011179	-0,0012	0,00207299
233		3,3843196	3,956541643	67	0,017073	-0,0071	0,024113252
234		2,6506851	4,063844722	68	-0,00911	0,0191	0,01744782
235		2,1969359	4,073732719	69	-0,00245	0,0124	0,013159812
236		2,6590167	4,090670492	70	0,00184	0,0082	0,004055206
237		3,5612049	4,093325602	71	0,010945	-0,0009	0,006436334
238		2,6246834	4,130039369	72	0,008564	0,0014	0,010206867
239		4,8140507	4,170873135	73	0,004793	0,0052	0,06013647

Критерий Фроцини

$$B_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left| U_i - \frac{2i-1}{2n} \right|$$

0,421223 статистика Фроцини

0,578 критеческое значение статистики Фроцини

Вывод: так как 0,421 < 0,578, то гипотеза H0 о равномерности принимается

n	α		
	0,9	0,95	0,99
15	0,4961	0,573	0,7418
∞	0,497	0,578	0,744

Приложение 3.1

Варианты заданий по лабораторной работе №3

Вариант 1. Нормальное распределение $f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}},\ x\in(-\infty,\infty)$ с параметрами $\mu=2,\ \sigma=5$

5,905	2,089	-0,54	-1,76	-0,32	2,653	2,591	-0,77	-8,72	-6,09
2,752	8,509	3,658	3,583	-0,91	0,96	-3,7	6,593	-2,09	0,058
1,013	-9,79	0,862	1,67	4,889	1,696	6,856	-6,17	1,606	4,305
-2,78	-0,23	1,763	-6,66	-5,1	12,8	-2,02	-1,87	-5,89	4,256
5,554	-2,16	4,859	-0,39	3,725	-2,08	5,382	-0,66	5,4	5,062
0,673	-1,57	5,876	4,409	3,775	1,657	-0,52	-7,42	11,83	9,073
7,574	5,853	12,65	1,74	0,566	6,627	2,047	3,326	-6,31	1,257
4,689	-1,38	-3,85	9,836	6,439	6,108	6,025	-0,55	6,035	-3,47
2,015	-1,23	5,246	3,273	7,755	4,351	-0,19	6,378	4,582	5,176
13	-2,34	-7,11	4,14	-2,56	-1,05	17,58	3,33	2,854	0,502

Вариант 2. Нормальное распределение $f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}},\ x\in(-\infty,\infty)$ с параметрами $\mu=5,\ \sigma=5$

11,57	16,66	-1,8	2,817	8,19	3,297	6,599	9,933	4,618	5,749
4,484	3,538	9,919	15,14	16,08	3,546	1,867	10,39	4,792	5,92
3,517	2,766	-4,37	9,782	7,036	-0,67	4,359	-5,5	-3,58	2,481
3,621	2,673	-3,18	6,431	1,579	0,016	-0,03	11,86	-5,38	0,475
12,69	3,897	9,73	3,518	10,26	11,75	0,76	10,46	-2,09	13,61
8,939	2,018	2,174	-1,26	7,532	4,902	9,171	-1,37	-2,85	4,996
18,58	1,85	-5,37	-1,97	3,929	10,1	8,714	2,423	3,798	6,924
9,802	2,864	9,058	7,484	0,482	6,848	4,001	0,367	4,394	-1,69
-7,31	6,267	8,759	13,07	0,468	0,8	5,101	3,599	0,057	6,368
4,374	4,74	12,03	12,15	7,708	0,165	2,382	-4,08	10,82	3,183
11,57	16,66	-1,8	2,817	8,19	3,297	6,599	9,933	4,618	5,749

Вариант 3. Нормальное распределение $f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}},\ x\in(-\infty,\infty)$ с параметрами $\mu=5,\ \sigma=1$

5,834	4,553	5,137	4,795	5,551	3,882	5,396	4,171	4,32	4,763
4,848	4,049	6,429	2,418	5,556	4,58	6,551	3,297	4,812	4,764
6,645	5,743	5,584	5,988	4,938	4,836	6,421	6,496	4,378	4,514
3,64	4,595	4,757	5,152	3,684	6,371	4,729	2,867	5,954	4,37

3,503	5,389	4,972	7,599	5,121	5,801	5,066	6,875	3,533	5,73
4,452	5,072	6,532	4,174	6,16	3,592	4,709	5,009	4,174	6,608
5,418	5,76	6,265	6,254	3,489	5,932	6,627	4,303	4,342	4,605
5,625	6,619	5,25	2,918	4,729	5,234	5,999	5,277	3,761	5,45
6,173	5,783	4,515	3,626	6,41	5,683	4,688	4,226	5,862	4,82
5,722	5,224	5,777	4,264	4,273	4,646	5,003	4,203	3,643	5,644

Вариант 4. Показательное распределение $f(x)=\lambda e^{-\lambda\cdot x},\ x\in[0,\infty)$ с параметром $\lambda=5$

0,392	0,162	0,318	0,601	0,301	0,033	0,303	0,05	0,234	0,007
0,435	0,347	0,162	0,028	0,095	0,081	0,006	0,352	0,196	0,178
0,062	0,018	0,388	0,006	0,013	0,291	0,084	0,329	0,181	0,514
0,193	0,243	0,322	0,089	0,606	0,105	0,244	0,053	0,213	0,029
0,221	0,634	0,126	0,034	0,494	0,059	0,185	0,061	0,062	0,154
0,209	0,237	0,141	0,137	0,705	0,162	0,051	0,024	0,294	0,059
0,201	0,524	0,123	0,004	0,04	0,297	0,144	0,666	0,101	0,378
0,112	0,345	0,486	0,081	0,944	0,308	1,122	0,088	0,126	0,162
0,167	0,197	0,44	0,019	0,01	0,117	0,023	0,022	0,004	0,109
0,266	0,369	2E-04	0,247	0,217	0,034	0,091	0,557	0,07	0,109

Вариант 5. Показательное распределение $f(x)=\lambda e^{-\lambda\cdot x},\ x\in[0,\infty)$ с параметром $\lambda=2$

0,166	0,02	0,574	0,996	1,349	0,408	0,571	0,204	0,705	0,332
0,241	0,678	0,05	0,331	0,135	1,406	0,169	0,213	1,438	0,044
0,455	0,431	0,459	0,419	0,172	0,395	0,051	0,119	1,378	1,568
0,459	0,098	0,383	0,875	0,331	0,438	0,426	0,627	0,365	0,151
0,441	0,258	0,607	0,551	0,197	0,711	0,41	0,242	0,073	2,426
0,129	0,101	0,487	0,014	0,127	0,072	0,113	0,12	0,109	0,64
0,521	0,041	0,368	0,277	0,11	0,227	0,043	0,696	0,213	0,026
0,045	0,098	1,089	0,432	0,137	1,872	0,191	0,006	0,085	0,241
0,113	0,212	0,189	0,047	0,419	0,551	0,276	0,377	1,478	0,272
0,391	1,135	0,084	0,298	0,745	0,271	0,236	0,048	1,253	0,81

Вариант 6. Показательное распределение $f(x)=\lambda e^{-\lambda\cdot x},\ x\in[0,\infty)$ с параметром $\lambda=3$

1,109	0,542	0,449	0,129	0,337	0,12	0,315	0,687	0,398	0,145
0,337	0,319	0,964	0,168	0,232	1,603	0,018	0,837	0,402	0,047
0,106	0,459	0,267	0,076	0,104	0,693	0,295	0,292	0,217	0,513
0,195	0,292	0,546	0,484	0,559	1,03	1,318	0,055	0,291	0,698
0,217	1,505	0,418	0,356	0,113	0,026	0,001	0,641	0,179	0,64
0,104	0,019	0,186	0,041	0,039	0,005	0,851	0,004	0,518	0,155
0,087	0,113	0,136	0,383	0,08	0,107	0,235	1,081	0,235	0,267
0,335	0,211	0,006	0,509	0,004	1,081	0,029	0,385	0,698	0,039
0,018	0,289	0,296	0,131	0,426	0,257	0,041	0,388	0,162	0,255
0,171	0,083	0,826	0,342	0,418	0,08	0,364	0,357	0,288	0,553

Вариант 7. Равномерное распределение с параметрами: $a=0,\ b=1$

0,078	0,598	0,272	0,104	0,874	0,773	0,937	0,911	0,944	0,413
0,502	0,659	0,666	0,447	0,479	0,417	0,531	0,828	0,257	0,785
0,935	0,259	0,541	0,701	0,399	0,673	0,539	0,012	0,479	0,624
0,421	0,213	0,16	0,238	0,087	0,281	0,999	0,886	0,77	0,976
0,828	0,234	0,509	0,553	0,255	0,829	0,029	0,26	0,792	0,871
0,036	0,549	0,852	0,331	0,498	0,391	0,415	0,07	0,897	0,615
0,65	0,124	0,026	0,668	0,253	0,789	0,327	0,706	0,802	0,695
0,328	0,419	0,905	0,486	0,194	0,502	0,41	0,304	0,927	0,174
0,484	0,696	0,254	0,649	0,812	0,128	0,719	0,97	0,719	0,742
0,028	0,115	0,491	0,529	0,163	0,77	0,672	0,563	0,308	0,685

Вариант 8. Равномерное распределение с параметрами: $a=1,\ b=5$

4,333	1,705	2,346	3,06	4,273	2,781	2,187	1,06	1,524	1,81
3,137	4,953	3,069	1,172	2,067	2,092	2,929	4,541	4,744	2,805
1,248	4,128	2,611	2,551	4,294	1,177	1,785	3,414	2,498	4,351
2,418	4,532	1,659	4,613	4,53	1,1	4,435	3,257	2,44	1,569
1,25	3,682	3,929	2,662	3,743	3,557	4,25	3,837	3,873	1,99
3,119	4,826	3,819	2,333	2,804	3,855	3,796	1,241	3,443	4,203
4,906	4,618	1,71	2,69	2,15	1,064	4,965	3,742	4,739	1,469
2,993	4,657	2,81	3,344	1,486	3,68	1,685	4,553	3,36	4,529
2,794	1,045	2,06	4,996	4,384	1,219	4,965	3,233	1,148	3,547
3,01	1,366	3,886	2,08	4,675	4,055	2,338	2,464	2,439	3,615

Вариант 9. Равномерное распределение с параметрами: $a=3,\ b=6$

4,05	5,313	3,397	5,516	5,578	4,053	5,391	4,34	5,359	4,94
5,071	4,194	3,379	5,338	4,87	3,692	5,558	4,715	4,224	3,524
5,325	4,771	5,544	5,422	5,102	3,888	4,249	4,924	5,39	4,731
5,344	3,518	4,697	4,343	5,503	4,282	4,015	3,587	4,119	4,585
5,07	5,269	3,309	3,969	5,065	4,795	3,321	4,793	4,124	5,699
4,326	5,727	4,176	3,885	3,062	4,706	3,468	5,933	4,293	5,425
4,465	4,311	5,553	4,974	3,315	4,827	5,153	4,703	5,624	5,185
3,905	3,914	3,76	4,963	5,545	3,477	5,38	5,883	4,317	5,953
3,435	4,692	3,681	3,331	4,338	3,522	3,635	4,959	4,536	4,05
3,588	3,776	5,196	4,319	3,284	5,432	3,223	3,554	3,727	3,387

Вариант 10. Показательное распределение $f(x)=\lambda e^{-\lambda\cdot x},\ x\in[0,\infty)$ с параметром $\lambda=1$

0,848	0,432	2,474	0,804	0,039	1,246	1,783	2,775	0,069	0,392
0,176	1,156	2,503	0,866	1,183	0,023	0,979	1,345	0,075	0,263
0,231	1,056	0,731	1,913	2,57	0,482	1,687	0,115	2,242	1,276
1,053	0,284	2,706	0,047	2,468	0,034	3,08	0,694	0,036	0,112
1,461	3,226	3,154	0,546	1,014	1,397	1,915	0,679	0,541	0,928
0,172	1,089	1,016	1,52	0,031	0,11	0,439	2,192	1,694	1,525

0,148	0,362	0,997	0,111	0,012	1,239	1,013	0,609	1,282	1,193
0,122	0,43	0,151	2,238	2,66	0,547	0,195	0,962	0,642	3,38
0,038	0,676	1,158	0,29	0,167	0,001	0,281	0,39	1,066	0,642
0,74	0,713	1,239	0,56	0,506	0,462	0,866	1,533	0,719	0,319

ЛАБОРАТОРНАЯ РАБОТА №4

Тема. Дисперсионный анализ данных

Цель работы:

Выполнить однофакторный и двухфакторный анализ данных.

4.1. Однофакторный параметрический анализ

1) Сгенерировать средствами пакета EXCEL 8 выборок из 10 значений случайной величины с нормальным законом распределения $N(\mu, \sigma^2)$. Варианты значений параметров μ, σ^2 приведены в приложении 3. Здесь $k = 8$ – количество уровней фактора A, $n_1 = n_2 = ... n_8 = n = 10$ – объемы выборок.

2) Выполнить следующие расчеты:

- вычислить средние значения по уровням $\bar{x}_i = \frac{1}{n}\sum_{j=1}^{n} x_{ji}$;
- вычислить общее среднее для всей выборки $\bar{x} = \frac{1}{k \cdot n}\sum_{i=1}^{k}\sum_{j=1}^{n} x_{ji}$;
- вычислить общую выборочную дисперсию $s_0^2 = \frac{1}{N-1}\sum_{i=1}^{k}\sum_{j=1}^{n}(x_{ji} - \bar{x})^2,\ N = k \cdot n$;
- вычислить выборочные дисперсии по уровням $s_i^2 = \frac{1}{n-1}\sum_{j=1}^{n}(x_{ji} - \bar{x}_i)^2$;
- проверить однородность дисперсий s_i^2, используя статистику Кохрана (см. [1], стр. 418) $g = \frac{\max\limits_{1 \le i \le k} s_i^2}{\sum_{i=1}^{k} s_i^2},\ k = 8$. Гипотеза об однородности дисперсий принимается, если выполняется неравенство $g \le g_\alpha(k, n);\ k = 8,\ n = 10$, где α – доверительная вероятность (положить $\alpha = 0{,}95$). Значения $g_\alpha(k, n)$ приведены в таблице 8 (см. статистические таблицы).

- вычислить дисперсию, характеризующую фактор случайности $s_{сл}^2 = \frac{1}{k}\sum_{i=1}^{k} s_i^2$, где k – количество уровней фактора A ($k = 8$);
- вычислить дисперсию фактора A: $s_A^2 = \frac{n}{n-1}\sum_{i=1}^{k}(\overline{x}_i - \overline{x})^2$;

3) Проверить нулевую гипотезу равенства средних значений на различных уровнях фактора A:

$$H_0 : m_1 = m_2 = ... = m_k = m,$$

используя критерий $F = \frac{s_A^2}{s_{сл}^2}$. Если $F = \frac{s_A^2}{s_{сл}^2} \le F_{0,05}(7;\ 72)$, то гипотеза H_0 принимается, следовательно, номер уровня фактора A не влияет на исследуемый процесс X.

4.2. Однофакторный непараметрический анализ

4.2.1 Критерий Краскела Уоллеса (произвольные альтернативы)

1) Сгенерировать средствами пакета EXCEL 8 выборок из 10 значений случайной величины с биномиальным законом распределения $B(n,\ p)$. Варианты значений параметров $n,\ p$ приведены в приложении 3. Здесь $k = 8$ – количество уровней фактора A, $n_1 = n_2 = ...n_8 = n = 10$ – объемы выборок.
2) Заменить данные $\{x_{ji}\}$ (наблюдения) их рангами $\{r_{ji}\}$ путем упорядочивания табличных данных $\{x_{ji}\}$ в порядке возрастания.
3) Для каждой обработки i (уровня фактора, столбца таблицы) вычислить:
   - суммарный и средний ранги $R_i = \sum_{j=1}^{n_i} r_{ji}$ и $\overline{R}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} r_{ji}$;
   - средний по всей совокупности ранг $R = \frac{N+1}{2}$, где $N = \sum_{i=1}^{k} n_i$ – общее число наблюдений ($N = 80$);

- вычислить статистику *Краскела-Уоллеса*

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{n} n_i \left(\overline{R}_i - \frac{N+1}{2} \right)^2, \; n_1 = n_2 = ... n_8 = 10.$$

4) Проверить, если ли среди x_{ji} совпадающие значения. Если имеются совпадающие значения, то при ранжировании и переходе к r_{ji} надо использовать средние ранги (например, если 2 значения (5 и 5) занимают ранги 11, 12, то средний ранг (11,5) надо присвоить им обоим).
5) Вычислить модифицированную статистику H':

   $$H' = \frac{H}{1 - \left(\sum_{j=1}^{m} T_j / (N^3 - N) \right)},$$ где m — число групп совпадающих наблюдений;

   $T_j = t_j^3 - t_j$ (t_j — число совпадающих наблюдений в группе j).
6) Проверить гипотезу H_0 о том, что расхождение наблюдений в сериях опытов для различных уровней факторов можно объяснить только случайными причинами. На статистическом языке это предположение означает, что все данные таблицы x_{ji} принадлежат одному и тому же распределению. Если $H \le \chi^2_\alpha(v = 7)$ (или $H' \le \chi^2_\alpha(v = 7)$), то гипотеза H_0 принимается.

4.2.2 Критерий Джонкхиера (альтернативы с упорядочиваем)

1) Для табличных данных, полученных в задании 3.2.1 для каждой пары уровней u и v, где $1 \le u < v \le k$, вычислить по выборкам с номерами u и v статистики Манна-Уитни: $U(u,v) = \sum_{i=1}^{n} \sum_{i=1}^{n} \varphi(x_i, y_j)$, где x и y – два сравниваемых столбца таблицы анализируемых данных.

   Здесь $\varphi(x_i, y_j) = \begin{cases} 0, & если\, x_i > y_j; \\ \frac{1}{2}, & если\, x_i = y_j; \\ 1, & если\, x_i < y_j. \end{cases}$
2) Вычислить статистику Джонкхиера $I = \sum U(u,v)$ для $1 \le u < v \le k$.
3) Вычислить среднее и дисперсию статистики Джонкхиера

$$MI=\frac{1}{4}\left(N^2-\sum_{i=1}^{k}n_j\right);\ DI=\frac{1}{72}\left[N^2(2N+3)-\sum_{j=1}^{k}n_j^2(2n_j+3)\right]$$, где n_j – количество наблюдений в каждом уровне ($n_1=n_2=...n_8=10$), а затем нормированную величину $I^*=\dfrac{I-MI}{\sqrt{DI}}$.

4) Проверить гипотезу H_0 о том, что расхождение наблюдений в сериях опытов для различных уровней факторов можно объяснить только случайными причинами. Если $\left|I^*\right|\le u_\alpha,\ \alpha=0{,}975$, то гипотеза H_0 принимается.

4.3. Двухфакторный параметрический анализ

1) Сгенерировать средствами пакета EXCEL 10 выборок из 30 значений случайной величины с нормальным законом распределения $N(\mu_m,\sigma_m^2),\ m=1,...,10$. Варианты значений параметров μ_m,σ_m^2 приведены в приложении 3. Из этих данных подготовить первый столбец таблицы, составленный из средних выборочных значений: $x_1=\dfrac{1}{30}\sum_{i=1}^{30}x_i^{(1)}$ – по первой выборке; $x_2=\dfrac{1}{30}\sum_{i=1}^{30}x_i^{(2)}$ – по второй выборке и т.д. $x_{10}=\dfrac{1}{30}\sum_{i=1}^{30}x_i^{(10)}$ – по десятой выборке. Этот столбец будет соответствовать первому уровню фактора A. Каждое значение $x_i,\ i=1,...,10$ в этом столбце соответствует фактору B.

Аналогично формируются 2-й, 3-й и т.д. 8-й столбцы таблицы, которые будут соответствовать 2-му, 3-му и т.д. 8-му уровням фактора A. В результате мы получим таблицу, подобную таблице задания №1, в которой $k=8$ – количество уровней фактора A, $n_1=n_2=...n_8=10$ – уровни фактора B.

2) Выполнить следующие расчеты:

- вычислить средние значения по уровням фактора A $\bar{x}_i=\dfrac{1}{n}\sum_{j=1}^{n}x_{ji}$;
- вычислить средние значения по уровням фактора B $\bar{x}_j=\dfrac{1}{n}\sum_{i=1}^{k}x_{ji}$;
- вычислить общее среднее для всей выборки $\bar{x}=\dfrac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n}x_{ji}^2$;

- вычислить выборочные дисперсии по уровням $s_i^2 = \frac{1}{n-1}\sum_{j=1}^{n}(x_{ji} - \overline{x}_i)^2$;
- вычислить дисперсию, характеризующую фактор случайности $s_{сл}^2 = \frac{1}{(n-1)(k-1)}\sum_{i=1}^{k}\sum_{j=1}^{n}\left(x_{ji} - \overline{x}_i - \overline{x}_j + \overline{x}\right)^2$, где k – количество уровней фактора A ($k = 8$);
- вычислить дисперсию фактора A: $s_A^2 = \frac{n}{n-1}\sum_{i=1}^{k}(\overline{x}_i - \overline{x})^2$;

3) Проверить нулевую гипотезу равенства средних значений на различных уровнях фактора A:

$$H_0 : \alpha_1 = \alpha_2 = ... = \alpha_k ,$$

используя критерий $F = \frac{s_A^2}{s_{сл}^2}$. Если $F = \frac{s_A^2}{s_{сл}^2} \le F_{0,05}(7;\, 72)$, то гипотеза H_0 принимается, следовательно, фактор A не влияет на исследуемый процесс X.

4.4. Двухфакторный непараметрический анализ

4.4.1. Критерий Фридмана (произвольные альтернативы)

1) Заменить данные $\{x_{ji}\}$ (наблюдения), полученные в задании 3.3 их рангами $\{r_{ji}\}$ путем упорядочивания табличных данных $\{x_{ji}\}$ в порядке возрастания. При этом ранги подсчитываются по каждой строке отдельно (т.е. по каждому фактору B_j)

1) Для каждого столбца таблицы вычислить:

- суммарный и средний ранги $R_i = \sum_{j=1}^{n_i} r_{ji}$ и $\overline{R}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} r_{ji}$; $n_i = k = 8$.
- $\overline{R} = \frac{k+1}{2}$
- вычислить статистику *Фридмана* $S = \frac{12n}{k(k+1)}\sum_{i=1}^{k}\left(\overline{R}_i - \overline{R}\right)^2$.

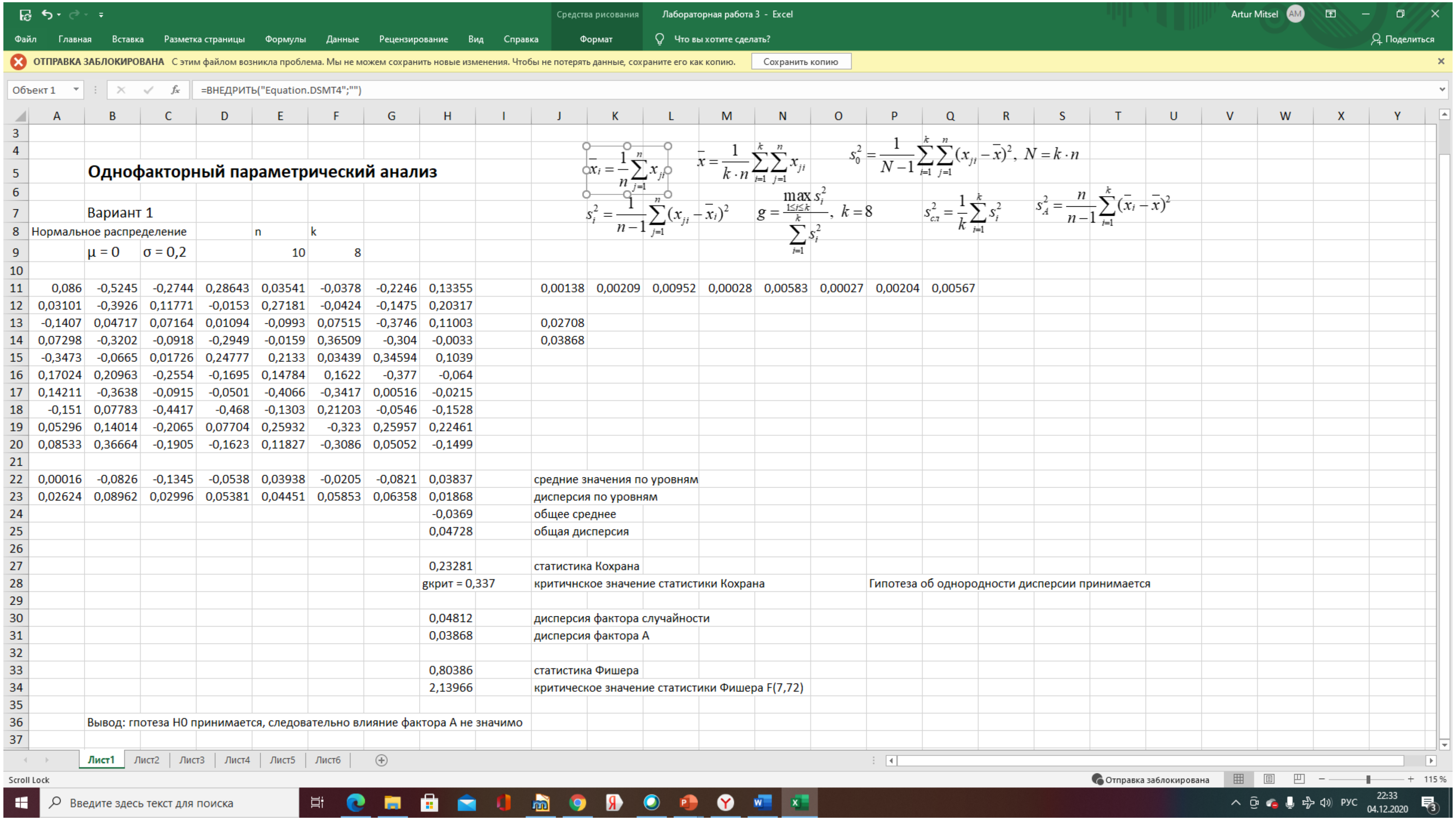
3) Проверить гипотезу H_0 о том, что расхождение наблюдений в сериях опытов для различных уровней факторов можно объяснить только случайными причинами. Если $S \le \chi^2_{0,05}(7)$, то гипотеза H_0 принимается.

4.4.2. Критерий Пейджа (альтернативы с упорядочиванием)

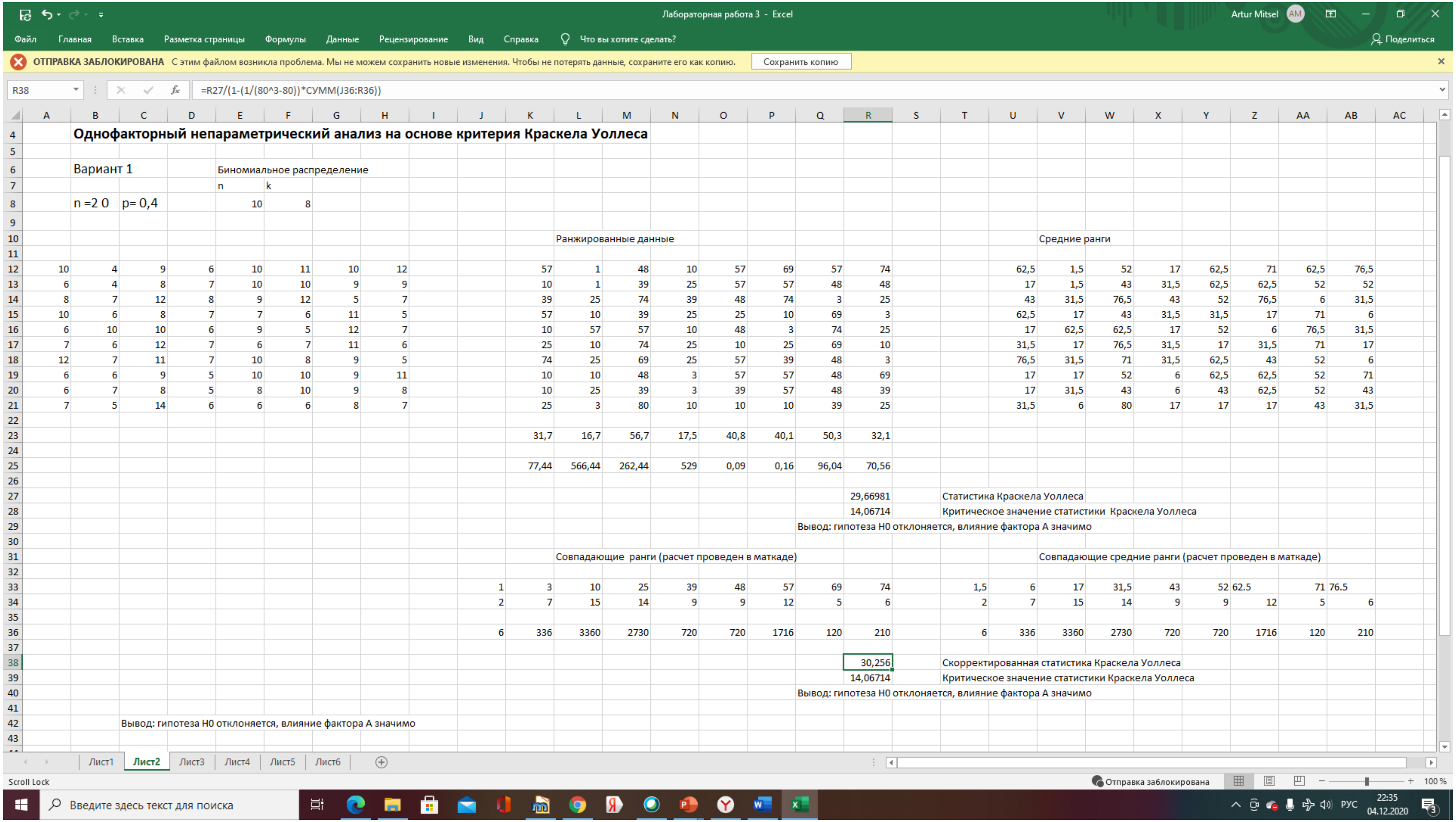
1) Упорядочить по возрастанию суммарные ранги $R_i = \sum_{j=1}^{n_i} r_{ji}$, полученные в задании 3.4.1.
2) Вычислить статистику Пейджа $L = \sum_{i=1}^{k} i \cdot R_i = 1 \cdot R_1 + 2 \cdot R_2 + 3 \cdot R_3 + ... + k \cdot R_k$.
3) Проверить гипотезу H_0 о том, что расхождение наблюдений в сериях опытов для различных уровней факторов можно объяснить только случайными причинами. Если $L_{набл} \le L_q(k,n)$, то гипотеза H_0 принимается.

4.5. Excel-скриншоты заданий лабораторной работы №4

Задание 4.1



Задание 4.2.1



Лабораторная работа 3 - Excel

R38 =R27/(1-(1/(80^3-80))*СУММ(J36:R36))

Однофакторный непараметрический анализ на основе критерия Краскела Уоллеса

Вариант 1

n =2 0 p= 0,4

Биномиальное распределение

n	k
10	8

A	B	C	D	E	F	G	H
10	4	9	6	10	11	10	12
6	4	8	7	10	10	9	9
8	7	12	8	9	12	5	7
10	6	8	7	7	6	11	5
6	10	10	6	9	5	12	7
7	6	12	7	6	7	11	6
12	7	11	7	10	8	9	5
6	6	9	5	10	10	9	11
6	7	8	5	8	10	9	8
7	5	14	6	6	6	8	7

Ранжированные данные

K	L	M	N	O	P	Q	R
57	1	48	10	57	69	57	74
10	1	39	25	57	57	48	48
39	25	74	39	48	74	3	25
57	10	39	25	25	10	69	3
10	57	57	10	48	3	74	25
25	10	74	25	10	25	69	10
74	25	69	25	57	39	48	3
10	10	48	3	57	57	48	69
10	25	39	3	39	57	48	39
25	3	80	10	10	10	39	25
31,7	16,7	56,7	17,5	40,8	40,1	50,3	32,1
77,44	566,44	262,44	529	0,09	0,16	96,04	70,56

Средние ранги

U	V	W	X	Y	Z	AA	AB
62,5	1,5	52	17	62,5	71	62,5	76,5
17	1,5	43	31,5	62,5	62,5	52	52
43	31,5	76,5	43	52	76,5	6	31,5
62,5	17	43	31,5	31,5	17	71	6
17	62,5	62,5	17	52	6	76,5	31,5
31,5	17	76,5	31,5	17	31,5	71	17
76,5	31,5	71	31,5	62,5	43	52	6
17	17	52	6	62,5	62,5	52	71
17	31,5	43	6	43	62,5	52	43
31,5	6	80	17	17	17	43	31,5

29,66981 Статистика Краскела Уоллеса

14,06714 Критическое значение статистики Краскела Уоллеса

Вывод: гипотеза H0 отклоняется, влияние фактора А значимо

Совпадающие ранги (расчет проведен в маткаде)

J	K	L	M	N	O	P	Q	R
1	3	10	25	39	48	57	69	74
2	7	15	14	9	9	12	5	6
6	336	3360	2730	720	720	1716	120	210

Совпадающие средние ранги (расчет проведен в маткаде)

T	U	V	W	X	Y	Z	AA	AB
1,5	6	17	31,5	43	52	62.5	71	76.5
2	7	15	14	9	9	12	5	6
6	336	3360	2730	720	720	1716	120	210

30,256 Скорректированная статистика Краскела Уоллеса

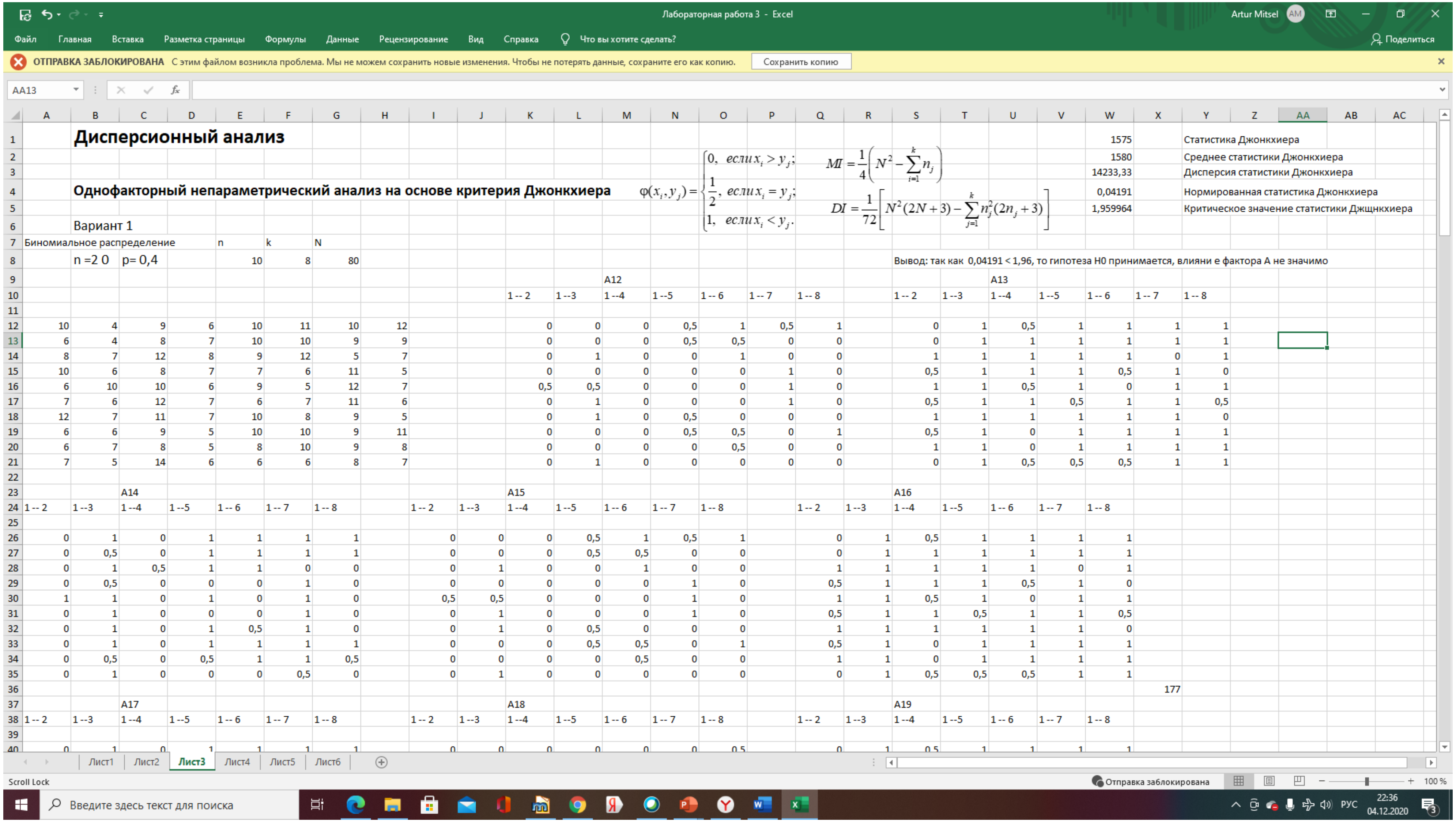
14,06714 Критическое значение статистики Краскела Уоллеса

Вывод: гипотеза H0 отклоняется, влияние фактора А значимо

Вывод: гипотеза H0 отклоняется, влияние фактора А значимо

Лист1 Лист2 Лист3 Лист4 Лист5 Лист6

Задание 4.2.2



Дисперсионный анализ

Однофакторный непараметрический анализ на основе критерия Джонкхиера

Вариант 1

$$\varphi(x_i, y_j) = \begin{cases} 0, & если\, x_i > y_j; \\ \frac{1}{2}, & если\, x_i = y_j; \\ 1, & если\, x_i < y_j. \end{cases}$$

$$MI = \frac{1}{4}\left(N^2 - \sum_{i=1}^{k} n_j\right)$$

$$DI = \frac{1}{72}\left[N^2(2N+3) - \sum_{j=1}^{k} n_j^2(2n_j+3)\right]$$

W	Y
1575	Статистика Джонкхиера
1580	Среднее статистики Джонкхиера
14233,33	Дисперсия статистики Джонкхиера
0,04191	Нормированная статистика Джонкхиера
1,959964	Критическое значение статистики Джщнхкиера

Биномиальное распределение

	n	k	N
n =2 0 p= 0,4	10	8	80

Вывод: так как 0,04191 < 1,96, то гипотеза H0 принимается, влияни е фактора A не значимо

	A	B	C	D	E	F	G	H
12	10	4	9	6	10	11	10	12
13	6	4	8	7	10	10	9	9
14	8	7	12	8	9	12	5	7
15	10	6	8	7	7	6	11	5
16	6	10	10	6	9	5	12	7
17	7	6	12	7	6	7	11	6
18	12	7	11	7	10	8	9	5
19	6	6	9	5	10	10	9	11
20	6	7	8	5	8	10	9	8
21	7	5	14	6	6	6	8	7

A12

	1 -- 2	1 --3	1 --4	1 --5	1 -- 6	1 -- 7	1 -- 8
12	0	0	0	0,5	1	0,5	1
13	0	0	0	0,5	0,5	0	0
14	0	1	0	0	1	0	0
15	0	0	0	0	0	1	0
16	0,5	0,5	0	0	0	1	0
17	0	1	0	0	0	1	0
18	0	1	0	0,5	0	0	0
19	0	0	0	0,5	0,5	0	1
20	0	0	0	0	0,5	0	0
21	0	1	0	0	0	0	0

A13

	1 -- 2	1 --3	1 --4	1 --5	1 -- 6	1 -- 7	1 -- 8
12	0	1	0,5	1	1	1	1
13	0	1	1	1	1	1	1
14	1	1	1	1	1	0	1
15	0,5	1	1	1	0,5	1	0
16	1	1	0,5	1	0	1	1
17	0,5	1	1	0,5	1	1	0,5
18	1	1	1	1	1	1	0
19	0,5	1	0	1	1	1	1
20	1	1	0	1	1	1	1
21	0	1	0,5	0,5	0,5	1	1

A14

	1 -- 2	1 --3	1 --4	1 --5	1 -- 6	1 -- 7	1 -- 8
26	0	1	0	1	1	1	1
27	0	0,5	0	1	1	1	1
28	0	1	0,5	1	1	0	0
29	0	0,5	0	0	0	1	0
30	1	1	0	1	0	1	0
31	0	1	0	0	0	1	0
32	0	1	0	1	0,5	1	0
33	0	1	0	1	1	1	1
34	0	0,5	0	0,5	1	1	0,5
35	0	1	0	0	0	0,5	0

A15

	1 -- 2	1 --3	1 --4	1 --5	1 -- 6	1 -- 7	1 -- 8
26	0	0	0	0,5	1	0,5	1
27	0	0	0	0,5	0,5	0	0
28	0	1	0	0	1	0	0
29	0	0	0	0	0	1	0
30	0,5	0,5	0	0	0	1	0
31	0	1	0	0	0	1	0
32	0	1	0	0,5	0	0	0
33	0	0	0	0,5	0,5	0	1
34	0	0	0	0	0,5	0	0
35	0	1	0	0	0	0	0

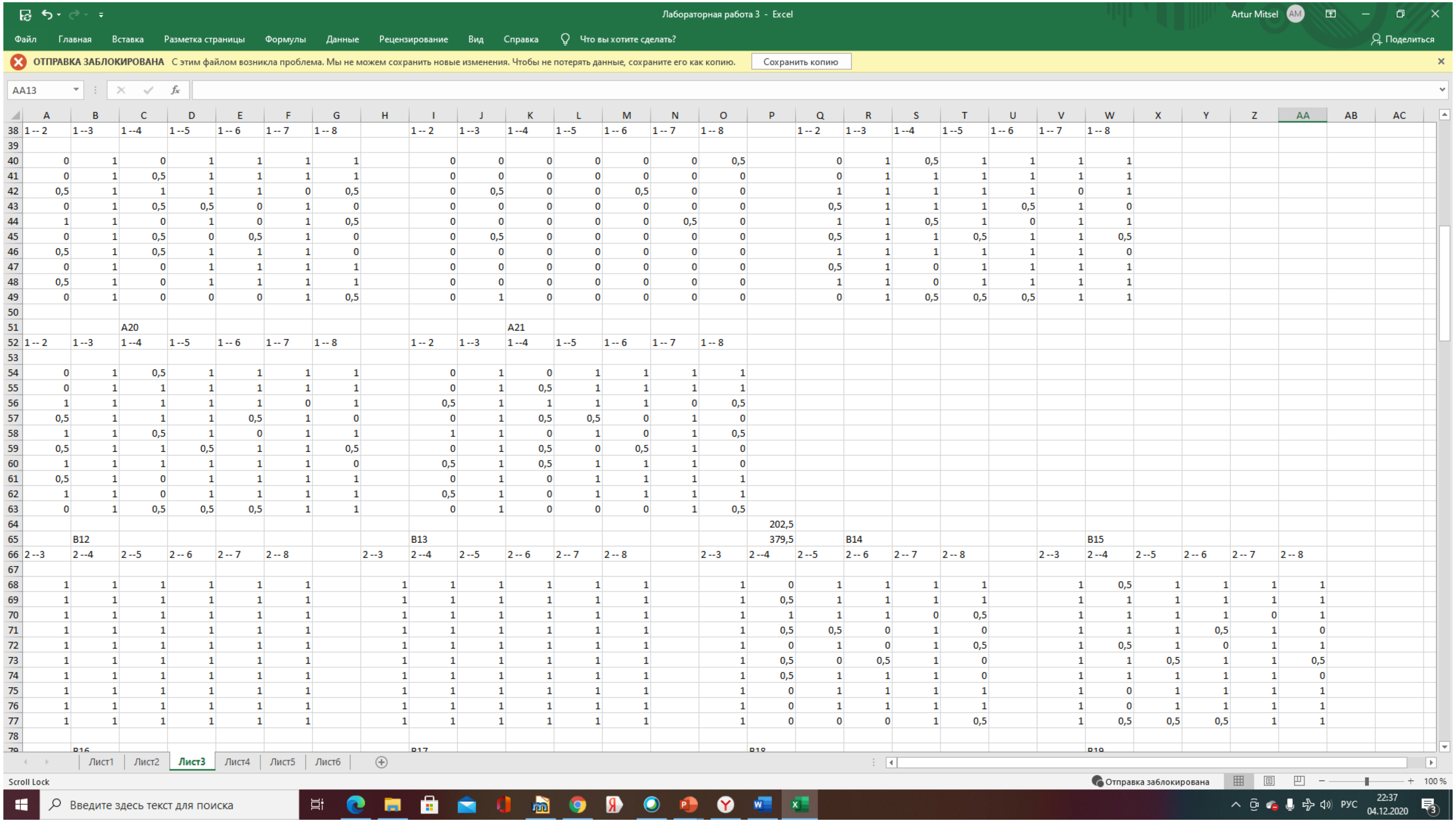
A16

	1 -- 2	1 --3	1 --4	1 --5	1 -- 6	1 -- 7	1 -- 8
26	0	1	0,5	1	1	1	1
27	0	1	1	1	1	1	1
28	1	1	1	1	1	0	1
29	0,5	1	1	1	0,5	1	0
30	1	1	0,5	1	0	1	1
31	0,5	1	1	0,5	1	1	0,5
32	1	1	1	1	1	1	0
33	0,5	1	0	1	1	1	1
34	1	1	0	1	1	1	1
35	0	1	0,5	0,5	0,5	1	1

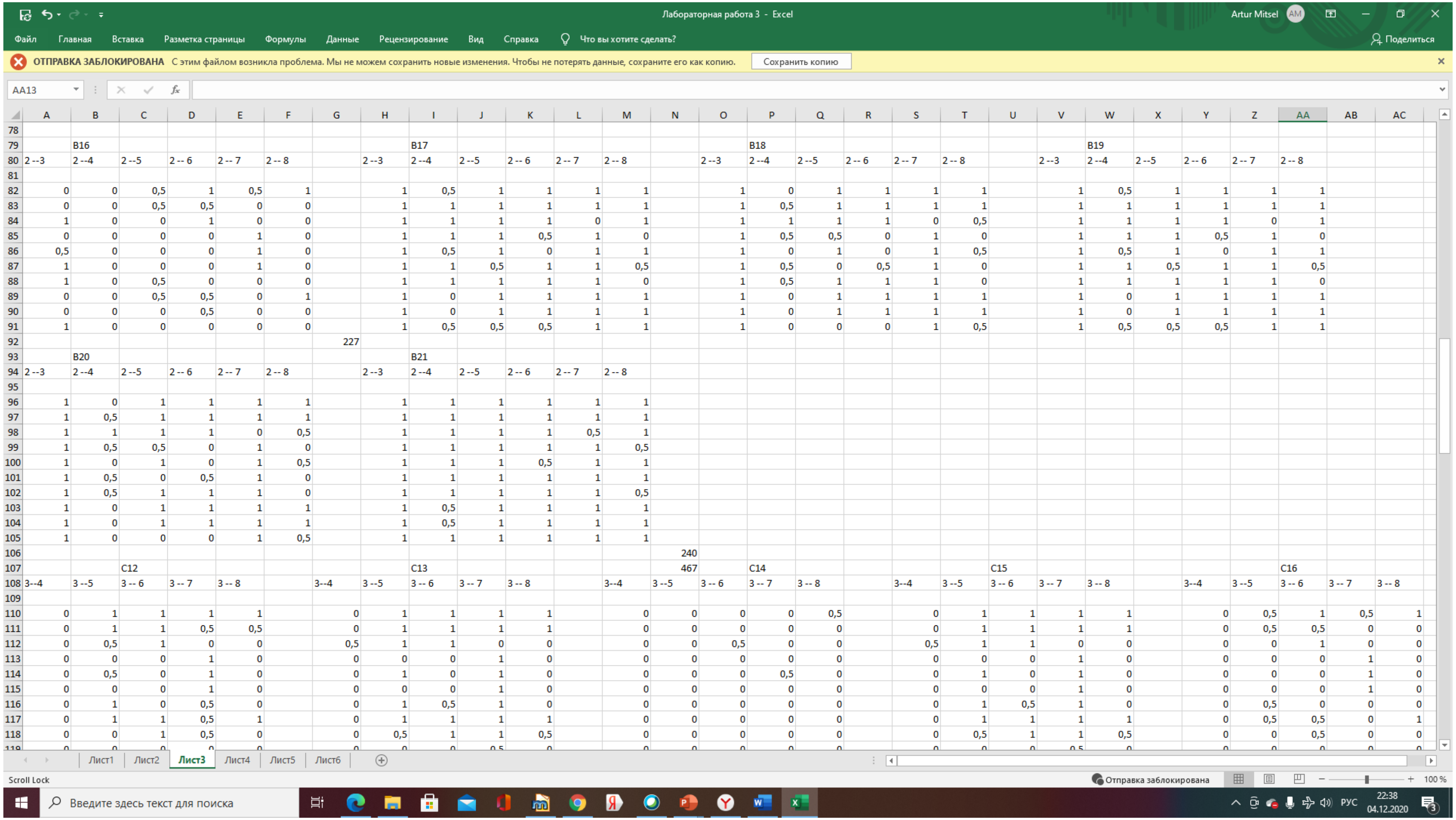
177

A17 A18 A19

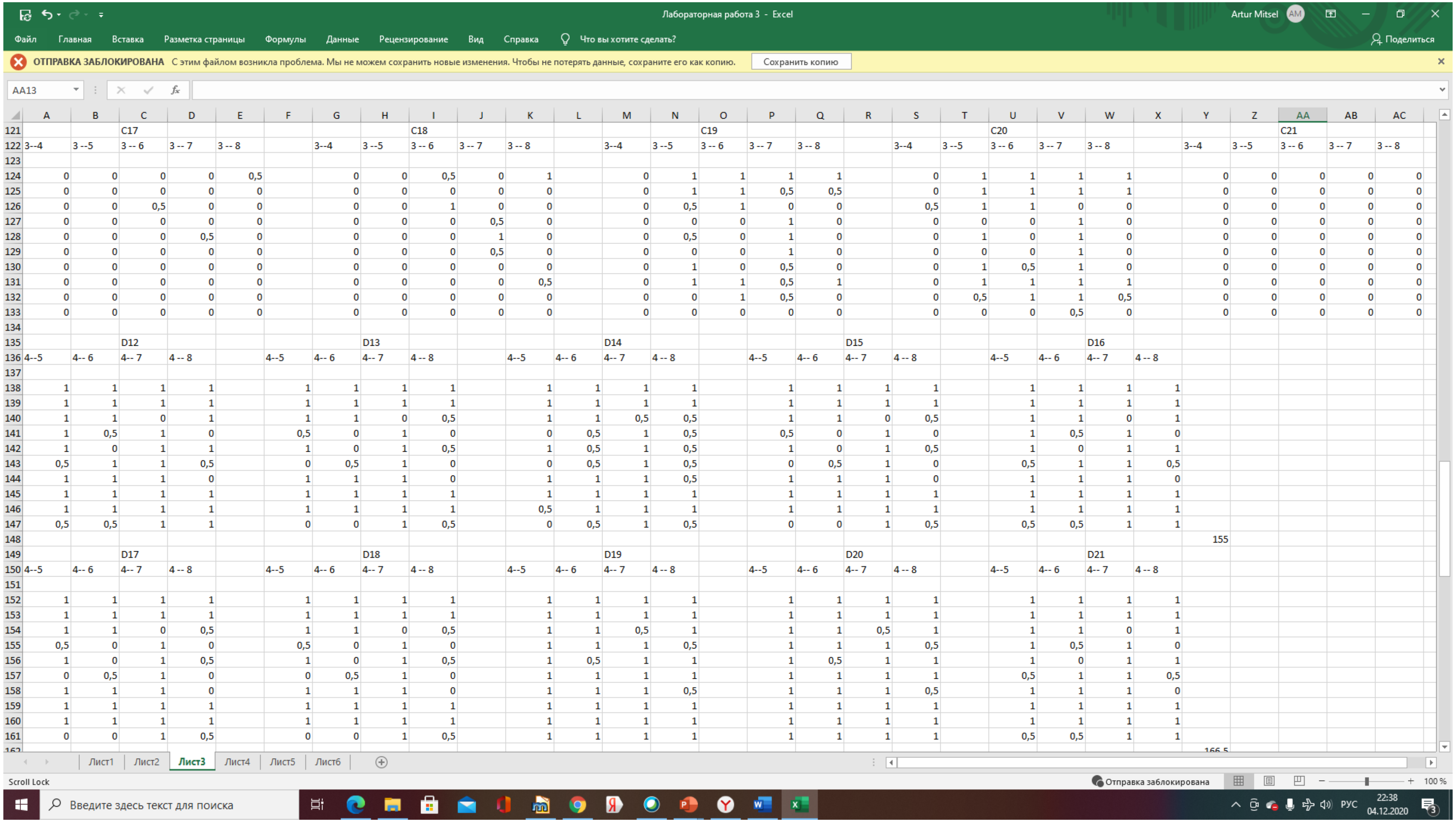
	1 -- 2	1 --3	1 --4	1 --5	1 -- 6	1 -- 7	1 -- 8



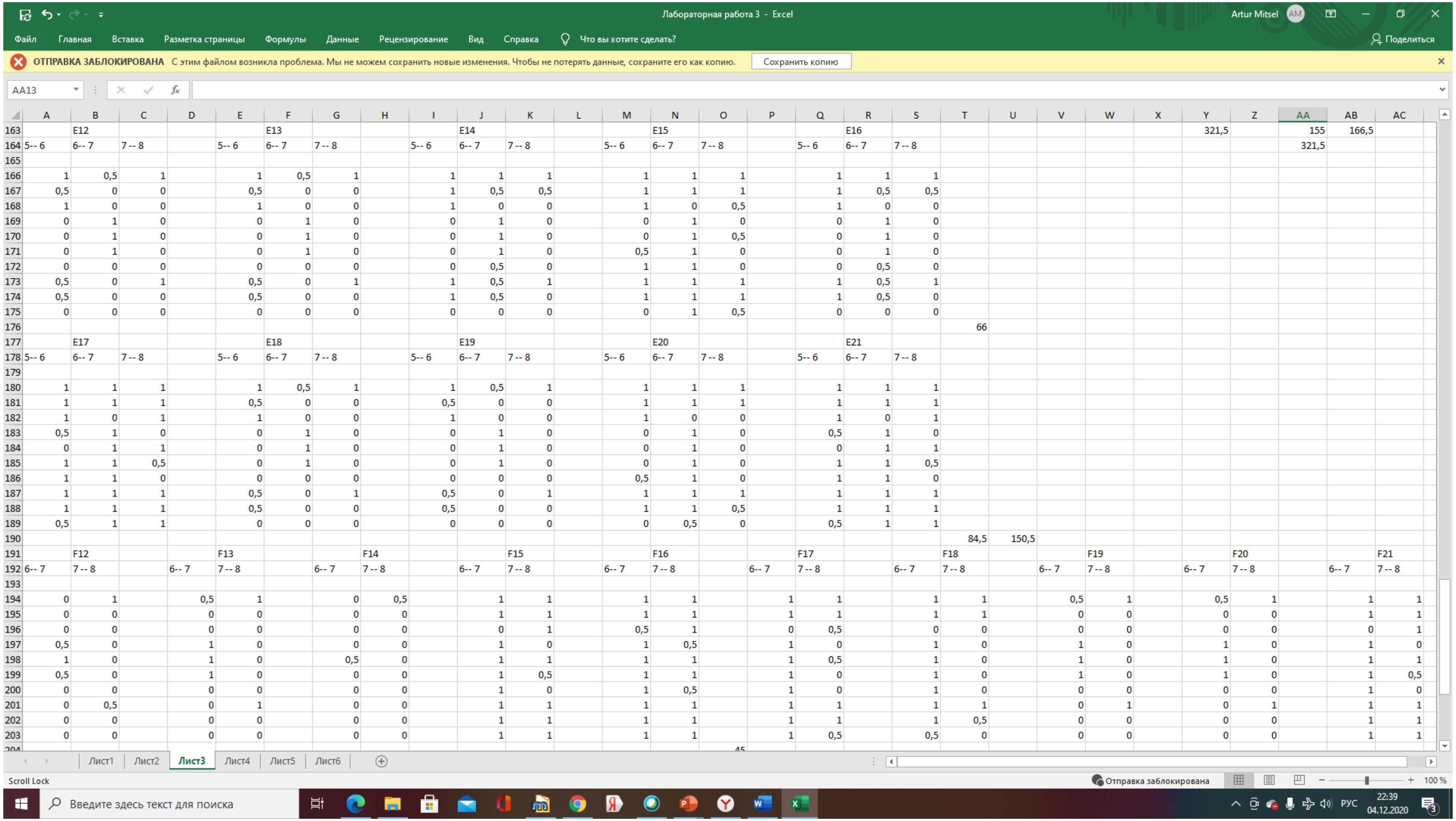
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA
38	1 -- 2	1 --3	1 --4	1 --5	1 -- 6	1 -- 7	1 -- 8		1 -- 2	1 --3	1 --4	1 --5	1 -- 6	1 -- 7	1 -- 8		1 -- 2	1 --3	1 --4	1 --5	1 -- 6	1 -- 7	1 -- 8				
39																											
40	0	1	0	1	1	1	1		0	0	0	0	0	0	0,5		0	1	0,5	1	1	1	1				
41	0	1	0,5	1	1	1	1		0	0	0	0	0	0	0		0	1	1	1	1	1	1				
42	0,5	1	1	1	1	0	0,5		0	0,5	0	0	0,5	0	0		1	1	1	1	1	0	1				
43	0	1	0,5	0,5	0	1	0		0	0	0	0	0	0	0		0,5	1	1	1	0,5	1	0				
44	1	1	0	1	0	1	0,5		0	0	0	0	0	0,5	0		1	1	0,5	1	0	1	1				
45	0	1	0,5	0	0,5	1	0		0	0,5	0	0	0	0	0		0,5	1	1	0,5	1	1	0,5				
46	0,5	1	0,5	1	1	1	0		0	0	0	0	0	0	0		1	1	1	1	1	1	0				
47	0	1	0	1	1	1	1		0	0	0	0	0	0	0		0,5	1	0	1	1	1	1				
48	0,5	1	0	1	1	1	1		0	0	0	0	0	0	0		1	1	0	1	1	1	1				
49	0	1	0	0	0	1	0,5		0	1	0	0	0	0	0		0	1	0,5	0,5	0,5	1	1				
50																											
51			A20								A21																
52	1 -- 2	1 --3	1 --4	1 --5	1 -- 6	1 -- 7	1 -- 8		1 -- 2	1 --3	1 --4	1 --5	1 -- 6	1 -- 7	1 -- 8												
53																											
54	0	1	0,5	1	1	1	1		0	1	0	1	1	1	1												
55	0	1	1	1	1	1	1		0	1	0,5	1	1	1	1												
56	1	1	1	1	1	0	1		0,5	1	1	1	1	0	0,5												
57	0,5	1	1	1	0,5	1	0		0	1	0,5	0,5	0	1	0												
58	1	1	0,5	1	0	1	1		1	1	0	1	0	1	0,5												
59	0,5	1	1	0,5	1	1	0,5		0	1	0,5	0	0,5	1	0												
60	1	1	1	1	1	1	0		0,5	1	0,5	1	1	1	0												
61	0,5	1	0	1	1	1	1		0	1	0	1	1	1	1												
62	1	1	0	1	1	1	1		0,5	1	0	1	1	1	1												
63	0	1	0,5	0,5	0,5	1	1		0	1	0	0	0	1	0,5												
64																202,5											
65		B12							B13							379,5		B14					B15				
66	2 --3	2 --4	2 --5	2 -- 6	2 -- 7	2 -- 8		2 --3	2 --4	2 --5	2 -- 6	2 -- 7	2 -- 8		2 --3	2 --4	2 --5	2 -- 6	2 -- 7	2 -- 8		2 --3	2 --4	2 --5	2 -- 6	2 -- 7	2 -- 8
67																											
68	1	1	1	1	1	1		1	1	1	1	1	1		1	0	1	1	1	1		1	0,5	1	1	1	1
69	1	1	1	1	1	1		1	1	1	1	1	1		1	0,5	1	1	1	1		1	1	1	1	1	1
70	1	1	1	1	1	1		1	1	1	1	1	1		1	1	1	1	0	0,5		1	1	1	1	0	1
71	1	1	1	1	1	1		1	1	1	1	1	1		1	0,5	0,5	0	1	0		1	1	1	0,5	1	0
72	1	1	1	1	1	1		1	1	1	1	1	1		1	0	1	0	1	0,5		1	0,5	1	0	1	1
73	1	1	1	1	1	1		1	1	1	1	1	1		1	0,5	0	0,5	1	0		1	1	0,5	1	1	0,5
74	1	1	1	1	1	1		1	1	1	1	1	1		1	0,5	1	1	1	0		1	1	1	1	1	0
75	1	1	1	1	1	1		1	1	1	1	1	1		1	0	1	1	1	1		1	0	1	1	1	1
76	1	1	1	1	1	1		1	1	1	1	1	1		1	0	1	1	1	1		1	0	1	1	1	1
77	1	1	1	1	1	1		1	1	1	1	1	1		1	0	0	0	1	0,5		1	0,5	0,5	0,5	1	1
78																											



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC
78																													
79		B16							B17							B18							B19						
80	2 --3	2 --4	2 --5	2 -- 6	2 -- 7	2 -- 8		2 --3	2 --4	2 --5	2 -- 6	2 -- 7	2 -- 8		2 --3	2 --4	2 --5	2 -- 6	2 -- 7	2 -- 8		2 --3	2 --4	2 --5	2 -- 6	2 -- 7	2 -- 8		
81																													
82	0	0	0,5	1	0,5	1		1	0,5	1	1	1	1		1	0	1	1	1	1		1	0,5	1	1	1	1		
83	0	0	0,5	0,5	0	0		1	1	1	1	1	1		1	0,5	1	1	1	1		1	1	1	1	1	1		
84	1	0	0	1	0	0		1	1	1	1	0	1		1	1	1	1	0	0,5		1	1	1	1	0	1		
85	0	0	0	0	1	0		1	1	1	0,5	1	0		1	0,5	0,5	0	1	0		1	1	1	0,5	1	0		
86	0,5	0	0	0	1	0		1	0,5	1	0	1	1		1	0	1	0	1	0,5		1	0,5	1	0	1	1		
87	1	0	0	0	1	0		1	1	0,5	1	1	0,5		1	0,5	0	0,5	1	0		1	1	0,5	1	1	0,5		
88	1	0	0,5	0	0	0		1	1	1	1	1	0		1	0,5	1	1	1	0		1	1	1	1	1	0		
89	0	0	0,5	0,5	0	1		1	0	1	1	1	1		1	0	1	1	1	1		1	0	1	1	1	1		
90	0	0	0	0,5	0	0		1	0	1	1	1	1		1	0	1	1	1	1		1	0	1	1	1	1		
91	1	0	0	0	0	0		1	0,5	0,5	0,5	1	1		1	0	0	0	1	0,5		1	0,5	0,5	0,5	1	1		
92							227																						
93		B20							B21																				
94	2 --3	2 --4	2 --5	2 -- 6	2 -- 7	2 -- 8		2 --3	2 --4	2 --5	2 -- 6	2 -- 7	2 -- 8																
95																													
96	1	0	1	1	1	1		1	1	1	1	1	1																
97	1	0,5	1	1	1	1		1	1	1	1	1	1																
98	1	1	1	1	0	0,5		1	1	1	1	0,5	1																
99	1	0,5	0,5	0	1	0		1	1	1	1	1	0,5																
100	1	0	1	0	1	0,5		1	1	1	0,5	1	1																
101	1	0,5	0	0,5	1	0		1	1	1	1	1	1																
102	1	0,5	1	1	1	0		1	1	1	1	1	0,5																
103	1	0	1	1	1	1		1	0,5	1	1	1	1																
104	1	0	1	1	1	1		1	0,5	1	1	1	1																
105	1	0	0	0	1	0,5		1	1	1	1	1	1																
106														240															
107			C12						C13					467		C14					C15						C16		
108	3--4	3 --5	3 -- 6	3 -- 7	3 -- 8		3--4	3 --5	3 -- 6	3 -- 7	3 -- 8		3--4	3 --5	3 -- 6	3 -- 7	3 -- 8		3--4	3 --5	3 -- 6	3 -- 7	3 -- 8		3--4	3 --5	3 -- 6	3 -- 7	3 -- 8
109																													
110	0	1	1	1	1		0	1	1	1	1		0	0	0	0	0,5		0	1	1	1	1		0	0,5	1	0,5	1
111	0	1	1	0,5	0,5		0	1	1	1	1		0	0	0	0	0		0	1	1	1	1		0	0,5	0,5	0	0
112	0	0,5	1	0	0		0,5	1	1	0	0		0	0	0,5	0	0		0,5	1	1	0	0		0	0	1	0	0
113	0	0	0	1	0		0	0	0	1	0		0	0	0	0	0		0	0	0	1	0		0	0	0	1	0
114	0	0,5	0	1	0		0	1	0	1	0		0	0	0	0,5	0		0	1	0	1	0		0	0	0	1	0
115	0	0	0	1	0		0	0	0	1	0		0	0	0	0	0		0	0	0	1	0		0	0	0	1	0
116	0	1	0	0,5	0		0	1	0,5	1	0		0	0	0	0	0		0	1	0,5	1	0		0	0,5	0	0	0
117	0	1	1	0,5	1		0	1	1	1	1		0	0	0	0	0		0	1	1	1	1		0	0,5	0,5	0	1
118	0	0	1	0,5	0		0	0,5	1	1	0,5		0	0	0	0	0		0	0,5	1	1	0,5		0	0	0,5	0	0
119	0	0	0	0	0		0	0	0	0,5	0		0	0	0	0	0		0	0	0	0,5	0		0	0	0	0	0

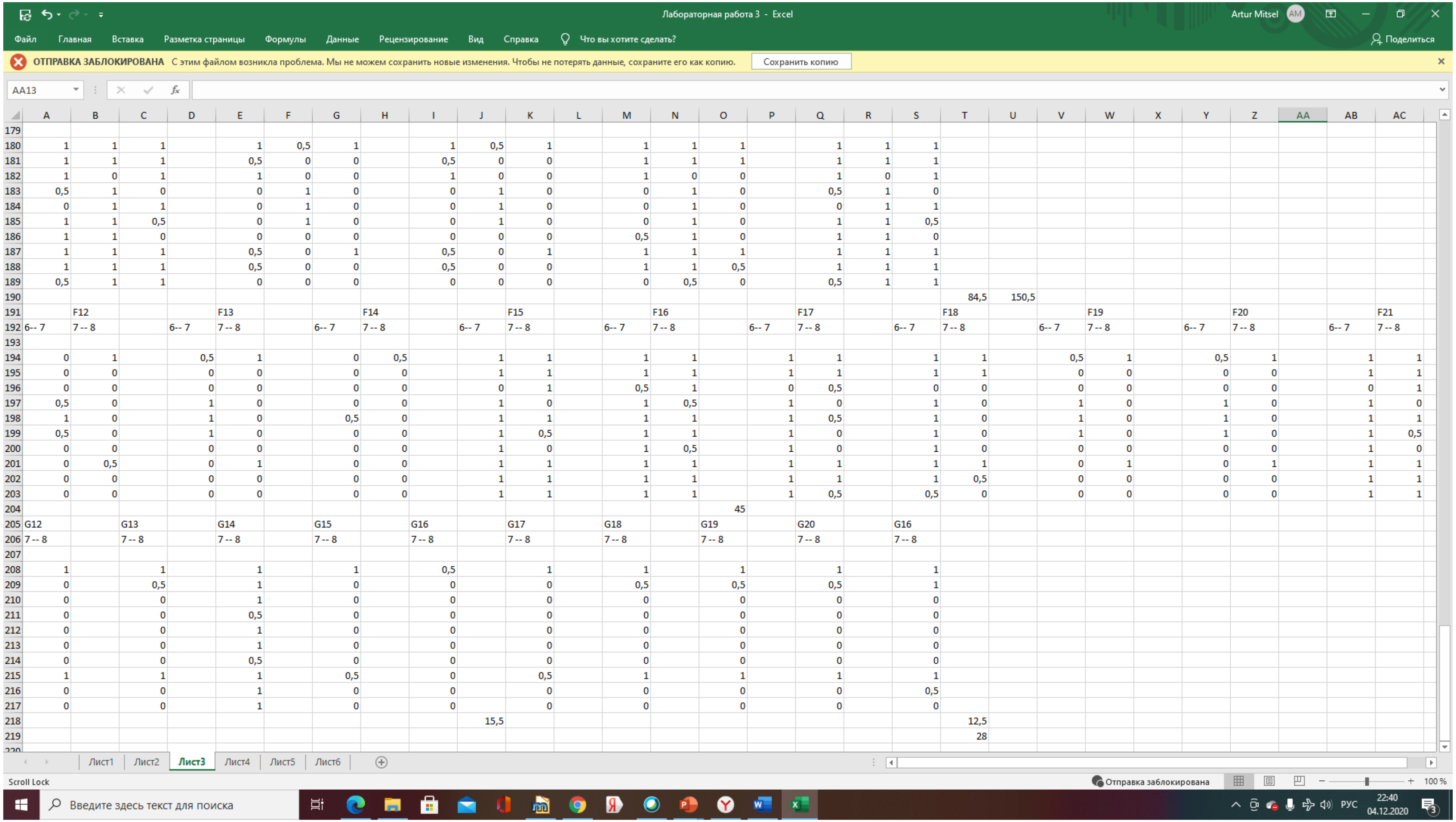


	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC
121			C17						C18						C19						C20						C21		
122	3--4	3 --5	3 -- 6	3 -- 7	3 -- 8		3--4	3 --5	3 -- 6	3 -- 7	3 -- 8		3--4	3 --5	3 -- 6	3 -- 7	3 -- 8		3--4	3 --5	3 -- 6	3 -- 7	3 -- 8		3--4	3 --5	3 -- 6	3 -- 7	3 -- 8
123																													
124	0	0	0	0	0,5		0	0	0,5	0	1		0	1	1	1	1		0	1	1	1	1		0	0	0	0	0
125	0	0	0	0	0		0	0	0	0	0		0	1	1	0,5	0,5		0	1	1	1	1		0	0	0	0	0
126	0	0	0,5	0	0		0	0	1	0	0		0	0,5	1	0	0		0,5	1	1	0	0		0	0	0	0	0
127	0	0	0	0	0		0	0	0	0,5	0		0	0	0	1	0		0	0	0	1	0		0	0	0	0	0
128	0	0	0	0,5	0		0	0	0	1	0		0	0,5	0	1	0		0	1	0	1	0		0	0	0	0	0
129	0	0	0	0	0		0	0	0	0,5	0		0	0	0	1	0		0	0	0	1	0		0	0	0	0	0
130	0	0	0	0	0		0	0	0	0	0		0	1	0	0,5	0		0	1	0,5	1	0		0	0	0	0	0
131	0	0	0	0	0		0	0	0	0	0,5		0	1	1	0,5	1		0	1	1	1	1		0	0	0	0	0
132	0	0	0	0	0		0	0	0	0	0		0	0	1	0,5	0		0	0,5	1	1	0,5		0	0	0	0	0
133	0	0	0	0	0		0	0	0	0	0		0	0	0	0	0		0	0	0	0,5	0		0	0	0	0	0
134																													
135			D12					D13					D14					D15					D16						
136	4--5	4-- 6	4-- 7	4 -- 8		4--5	4-- 6	4-- 7	4 -- 8		4--5	4-- 6	4-- 7	4 -- 8		4--5	4-- 6	4-- 7	4 -- 8		4--5	4-- 6	4-- 7	4 -- 8					
137																													
138	1	1	1	1		1	1	1	1		1	1	1	1		1	1	1	1		1	1	1	1					
139	1	1	1	1		1	1	1	1		1	1	1	1		1	1	1	1		1	1	1	1					
140	1	1	0	1		1	1	0	0,5		1	1	0,5	0,5		1	1	0	0,5		1	1	0	1					
141	1	0,5	1	0		0,5	0	1	0		0	0,5	1	0,5		0,5	0	1	0		1	0,5	1	0					
142	1	0	1	1		1	0	1	0,5		1	0,5	1	0,5		1	0	1	0,5		1	0	1	1					
143	0,5	1	1	0,5		0	0,5	1	0		0	0,5	1	0,5		0	0,5	1	0		0,5	1	1	0,5					
144	1	1	1	0		1	1	1	0		1	1	1	0,5		1	1	1	0		1	1	1	0					
145	1	1	1	1		1	1	1	1		1	1	1	1		1	1	1	1		1	1	1	1					
146	1	1	1	1		1	1	1	1		0,5	1	1	1		1	1	1	1		1	1	1	1					
147	0,5	0,5	1	1		0	0	1	0,5		0	0,5	1	0,5		0	0	1	0,5		0,5	0,5	1	1					
148																									155				
149			D17					D18					D19					D20					D21						
150	4--5	4-- 6	4-- 7	4 -- 8		4--5	4-- 6	4-- 7	4 -- 8		4--5	4-- 6	4-- 7	4 -- 8		4--5	4-- 6	4-- 7	4 -- 8		4--5	4-- 6	4-- 7	4 -- 8					
151																													
152	1	1	1	1		1	1	1	1		1	1	1	1		1	1	1	1		1	1	1	1					
153	1	1	1	1		1	1	1	1		1	1	1	1		1	1	1	1		1	1	1	1					
154	1	1	0	0,5		1	1	0	0,5		1	1	0,5	1		1	1	0,5	1		1	1	0	1					
155	0,5	0	1	0		0,5	0	1	0		1	1	1	0,5		1	1	1	0,5		1	0,5	1	0					
156	1	0	1	0,5		1	0	1	0,5		1	0,5	1	1		1	0,5	1	1		1	0	1	1					
157	0	0,5	1	0		0	0,5	1	0		1	1	1	1		1	1	1	1		0,5	1	1	0,5					
158	1	1	1	0		1	1	1	0		1	1	1	0,5		1	1	1	0,5		1	1	1	0					
159	1	1	1	1		1	1	1	1		1	1	1	1		1	1	1	1		1	1	1	1					
160	1	1	1	1		1	1	1	1		1	1	1	1		1	1	1	1		1	1	1	1					
161	0	0	1	0,5		0	0	1	0,5		1	1	1	1		1	1	1	1		0,5	0,5	1	1					
162																									166,5				



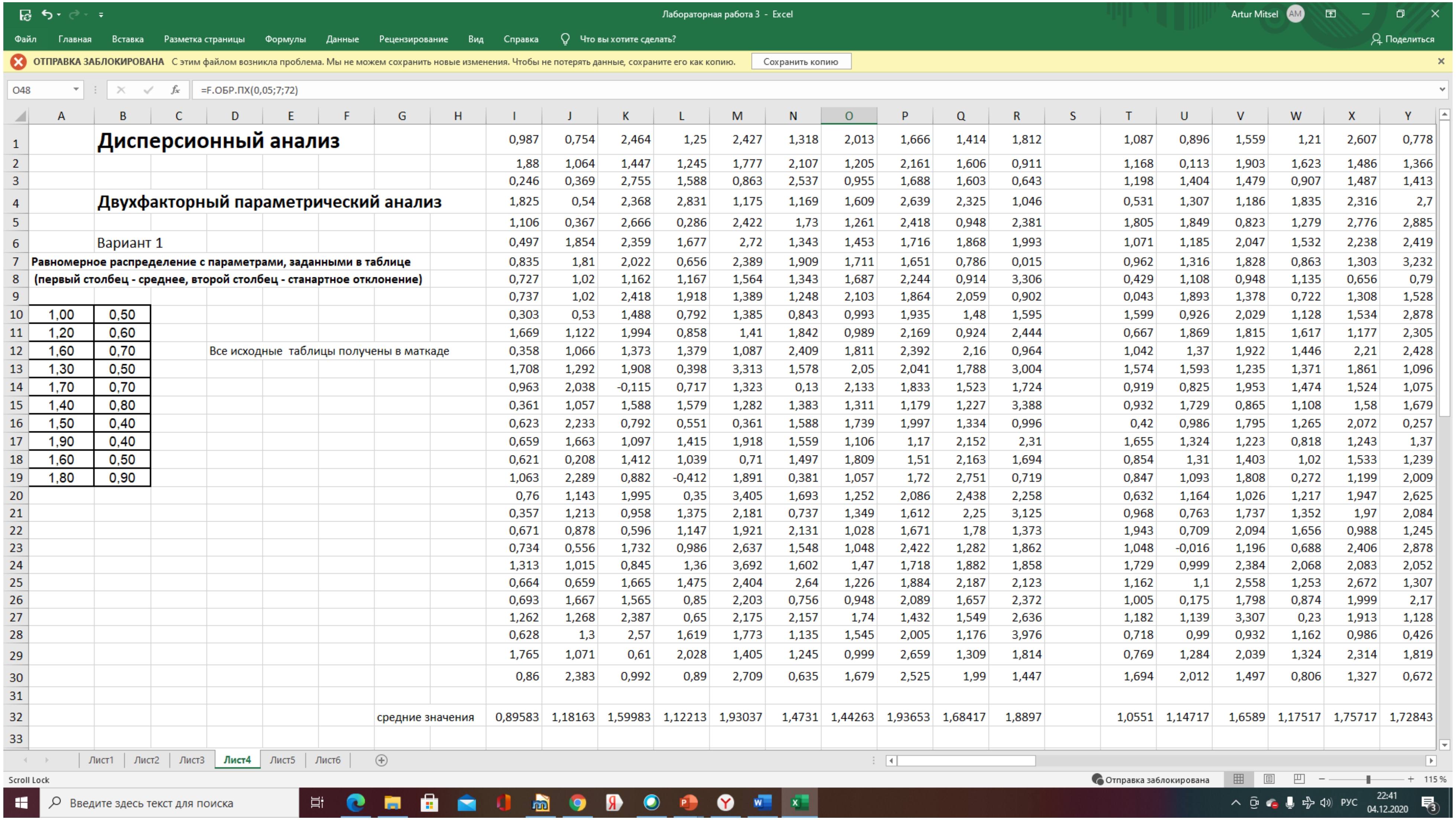
	A	B	C	E	F	G	I	J	K	M	N	O	Q	R	S	T	U	Y	AA	AB
163		E12			E13			E14			E15			E16				321,5	155	166,5
164	5-- 6	6-- 7	7 -- 8	5-- 6	6-- 7	7 -- 8	5-- 6	6-- 7	7 -- 8	5-- 6	6-- 7	7 -- 8	5-- 6	6-- 7	7 -- 8				321,5	
165																				
166	1	0,5	1	1	0,5	1	1	1	1	1	1	1	1	1	1					
167	0,5	0	0	0,5	0	0	1	0,5	0,5	1	1	1	1	0,5	0,5					
168	1	0	0	1	0	0	1	0	0	1	0	0,5	1	0	0					
169	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0					
170	0	1	0	0	1	0	0	1	0	0	1	0,5	0	1	0					
171	0	1	0	0	1	0	0	1	0	0,5	1	0	0	1	0					
172	0	0	0	0	0	0	0	0,5	0	1	1	0	0	0,5	0					
173	0,5	0	1	0,5	0	1	1	0,5	1	1	1	1	1	0,5	1					
174	0,5	0	0	0,5	0	0	1	0,5	0	1	1	1	1	0,5	0					
175	0	0	0	0	0	0	0	0	0	0	1	0,5	0	0	0					
176																66				
177		E17			E18			E19			E20			E21						
178	5-- 6	6-- 7	7 -- 8	5-- 6	6-- 7	7 -- 8	5-- 6	6-- 7	7 -- 8	5-- 6	6-- 7	7 -- 8	5-- 6	6-- 7	7 -- 8					
179																				
180	1	1	1	1	0,5	1	1	0,5	1	1	1	1	1	1	1					
181	1	1	1	0,5	0	0	0,5	0	0	1	1	1	1	1	1					
182	1	0	1	1	0	0	1	0	0	1	0	0	1	0	1					
183	0,5	1	0	0	1	0	0	1	0	0	1	0	0,5	1	0					
184	0	1	1	0	1	0	0	1	0	0	1	0	0	1	1					
185	1	1	0,5	0	1	0	0	1	0	0	1	0	1	1	0,5					
186	1	1	0	0	0	0	0	0	0	0,5	1	0	1	1	0					
187	1	1	1	0,5	0	1	0,5	0	1	1	1	1	1	1	1					
188	1	1	1	0,5	0	0	0,5	0	0	1	1	0,5	1	1	1					
189	0,5	1	1	0	0	0	0	0	0	0	0,5	0	0,5	1	1					
190																84,5	150,5			

	A	B	D	E	G	H	J	K	M	N	O	P	Q	S	T	V	W	Y	Z	AB	AC
191		F12		F13		F14		F15		F16			F17		F18		F19		F20		F21
192	6-- 7	7 -- 8	6-- 7	7 -- 8	6-- 7	7 -- 8	6-- 7	7 -- 8	6-- 7	7 -- 8		6-- 7	7 -- 8	6-- 7	7 -- 8	6-- 7	7 -- 8	6-- 7	7 -- 8	6-- 7	7 -- 8
193																					
194	0	1	0,5	1	0	0,5	1	1	1	1		1	1	1	1	0,5	1	0,5	1	1	1
195	0	0	0	0	0	0	1	1	1	1		1	1	1	1	0	0	0	0	1	1
196	0	0	0	0	0	0	0	1	0,5	1		0	0,5	0	0	0	0	0	0	0	1
197	0,5	0	1	0	0	0	1	0	1	0,5		1	0	1	0	1	0	1	0	1	0
198	1	0	1	0	0,5	0	1	1	1	1		1	0,5	1	0	1	0	1	0	1	1
199	0,5	0	1	0	0	0	1	0,5	1	1		1	0	1	0	1	0	1	0	1	0,5
200	0	0	0	0	0	0	1	0	1	0,5		1	0	1	0	0	0	0	0	1	0
201	0	0,5	0	1	0	0	1	1	1	1		1	1	1	1	0	1	0	1	1	1
202	0	0	0	0	0	0	1	1	1	1		1	1	1	0,5	0	0	0	0	1	1
203	0	0	0	0	0	0	1	1	1	1		1	0,5	0,5	0	0	0	0	0	1	1
204											45										



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC
179																													
180	1	1	1		1	0,5	1		1	0,5	1		1	1	1		1	1	1										
181	1	1	1		0,5	0	0		0,5	0	0		1	1	1		1	1	1										
182	1	0	1		1	0	0		1	0	0		1	0	0		1	0	1										
183	0,5	1	0		0	1	0		0	1	0		0	1	0		0,5	1	0										
184	0	1	1		0	1	0		0	1	0		0	1	0		0	1	1										
185	1	1	0,5		0	1	0		0	1	0		0	1	0		1	1	0,5										
186	1	1	0		0	0	0		0	0	0		0,5	1	0		1	1	0										
187	1	1	1		0,5	0	1		0,5	0	1		1	1	1		1	1	1										
188	1	1	1		0,5	0	0		0,5	0	0		1	1	0,5		1	1	1										
189	0,5	1	1		0	0	0		0	0	0		0	0,5	0		0,5	1	1										
190																				84,5	150,5								
191		F12			F13			F14			F15			F16			F17			F18			F19			F20			F21
192	6-- 7	7 -- 8		6-- 7	7 -- 8		6-- 7	7 -- 8		6-- 7	7 -- 8		6-- 7	7 -- 8		6-- 7	7 -- 8		6-- 7	7 -- 8		6-- 7	7 -- 8		6-- 7	7 -- 8		6-- 7	7 -- 8
193																													
194	0	1		0,5	1		0	0,5		1	1		1	1		1	1		1	1		0,5	1		0,5	1		1	1
195	0	0		0	0		0	0		1	1		1	1		1	1		1	1		0	0		0	0		1	1
196	0	0		0	0		0	0		0	1		0,5	1		0	0,5		0	0		0	0		0	0		0	1
197	0,5	0		1	0		0	0		1	0		1	0,5		1	0		1	0		1	0		1	0		1	0
198	1	0		1	0		0,5	0		1	1		1	1		1	0,5		1	0		1	0		1	0		1	1
199	0,5	0		1	0		0	0		1	0,5		1	1		1	0		1	0		1	0		1	0		1	0,5
200	0	0		0	0		0	0		1	0		1	0,5		1	0		1	0		0	0		0	0		1	0
201	0	0,5		0	1		0	0		1	1		1	1		1	1		1	1		0	1		0	1		1	1
202	0	0		0	0		0	0		1	1		1	1		1	1		1	0,5		0	0		0	0		1	1
203	0	0		0	0		0	0		1	1		1	1		1	0,5		0,5	0		0	0		0	0		1	1
204															45														
205	G12		G13		G14		G15		G16		G17		G18		G19		G20		G16										
206	7 -- 8		7 -- 8		7 -- 8		7 -- 8		7 -- 8		7 -- 8		7 -- 8		7 -- 8		7 -- 8		7 -- 8										
207																													
208	1		1		1		1		0,5		1		1		1		1		1										
209	0		0,5		1		0		0		0		0,5		0,5		0,5		1										
210	0		0		1		0		0		0		0		0		0		0										
211	0		0		0,5		0		0		0		0		0		0		0										
212	0		0		1		0		0		0		0		0		0		0										
213	0		0		1		0		0		0		0		0		0		0										
214	0		0		0,5		0		0		0		0		0		0		0										
215	1		1		1		0,5		0		0,5		1		1		1		1										
216	0		0		1		0		0		0		0		0		0		0,5										
217	0		0		1		0		0		0		0		0		0		0										
218										15,5										12,5									
219																				28									

Задание 4.3



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
1		**Дисперсионный анализ**							0,987	0,754	2,464	1,25	2,427	1,318	2,013	1,666	1,414	1,812		1,087	0,896	1,559	1,21	2,607	0,778
2									1,88	1,064	1,447	1,245	1,777	2,107	1,205	2,161	1,606	0,911		1,168	0,113	1,903	1,623	1,486	1,366
3									0,246	0,369	2,755	1,588	0,863	2,537	0,955	1,688	1,603	0,643		1,198	1,404	1,479	0,907	1,487	1,413
4		**Двухфакторный параметрический анализ**							1,825	0,54	2,368	2,831	1,175	1,169	1,609	2,639	2,325	1,046		0,531	1,307	1,186	1,835	2,316	2,7
5									1,106	0,367	2,666	0,286	2,422	1,73	1,261	2,418	0,948	2,381		1,805	1,849	0,823	1,279	2,776	2,885
6		Вариант 1							0,497	1,854	2,359	1,677	2,72	1,343	1,453	1,716	1,868	1,993		1,071	1,185	2,047	1,532	2,238	2,419
7	**Равномерное распределение с параметрами, заданными в таблице**								0,835	1,81	2,022	0,656	2,389	1,909	1,711	1,651	0,786	0,015		0,962	1,316	1,828	0,863	1,303	3,232
8	**(первый столбец - среднее, второй столбец - станартное отклонение)**								0,727	1,02	1,162	1,167	1,564	1,343	1,687	2,244	0,914	3,306		0,429	1,108	0,948	1,135	0,656	0,79
9									0,737	1,02	2,418	1,918	1,389	1,248	2,103	1,864	2,059	0,902		0,043	1,893	1,378	0,722	1,308	1,528
10	1,00	0,50							0,303	0,53	1,488	0,792	1,385	0,843	0,993	1,935	1,48	1,595		1,599	0,926	2,029	1,128	1,534	2,878
11	1,20	0,60							1,669	1,122	1,994	0,858	1,41	1,842	0,989	2,169	0,924	2,444		0,667	1,869	1,815	1,617	1,177	2,305
12	1,60	0,70		Все исходные таблицы получены в маткаде					0,358	1,066	1,373	1,379	1,087	2,409	1,811	2,392	2,16	0,964		1,042	1,37	1,922	1,446	2,21	2,428
13	1,30	0,50							1,708	1,292	1,908	0,398	3,313	1,578	2,05	2,041	1,788	3,004		1,574	1,593	1,235	1,371	1,861	1,096
14	1,70	0,70							0,963	2,038	-0,115	0,717	1,323	0,13	2,133	1,833	1,523	1,724		0,919	0,825	1,953	1,474	1,524	1,075
15	1,40	0,80							0,361	1,057	1,588	1,579	1,282	1,383	1,311	1,179	1,227	3,388		0,932	1,729	0,865	1,108	1,58	1,679
16	1,50	0,40							0,623	2,233	0,792	0,551	0,361	1,588	1,739	1,997	1,334	0,996		0,42	0,986	1,795	1,265	2,072	0,257
17	1,90	0,40							0,659	1,663	1,097	1,415	1,918	1,559	1,106	1,17	2,152	2,31		1,655	1,324	1,223	0,818	1,243	1,37
18	1,60	0,50							0,621	0,208	1,412	1,039	0,71	1,497	1,809	1,51	2,163	1,694		0,854	1,31	1,403	1,02	1,533	1,239
19	1,80	0,90							1,063	2,289	0,882	-0,412	1,891	0,381	1,057	1,72	2,751	0,719		0,847	1,093	1,808	0,272	1,199	2,009
20									0,76	1,143	1,995	0,35	3,405	1,693	1,252	2,086	2,438	2,258		0,632	1,164	1,026	1,217	1,947	2,625
21									0,357	1,213	0,958	1,375	2,181	0,737	1,349	1,612	2,25	3,125		0,968	0,763	1,737	1,352	1,97	2,084
22									0,671	0,878	0,596	1,147	1,921	2,131	1,028	1,671	1,78	1,373		1,943	0,709	2,094	1,656	0,988	1,245
23									0,734	0,556	1,732	0,986	2,637	1,548	1,048	2,422	1,282	1,862		1,048	-0,016	1,196	0,688	2,406	2,878
24									1,313	1,015	0,845	1,36	3,692	1,602	1,47	1,718	1,882	1,858		1,729	0,999	2,384	2,068	2,083	2,052
25									0,664	0,659	1,665	1,475	2,404	2,64	1,226	1,884	2,187	2,123		1,162	1,1	2,558	1,253	2,672	1,307
26									0,693	1,667	1,565	0,85	2,203	0,756	0,948	2,089	1,657	2,372		1,005	0,175	1,798	0,874	1,999	2,17
27									1,262	1,268	2,387	0,65	2,175	2,157	1,74	1,432	1,549	2,636		1,182	1,139	3,307	0,23	1,913	1,128
28									0,628	1,3	2,57	1,619	1,773	1,135	1,545	2,005	1,176	3,976		0,718	0,99	0,932	1,162	0,986	0,426
29									1,765	1,071	0,61	2,028	1,405	1,245	0,999	2,659	1,309	1,814		0,769	1,284	2,039	1,324	2,314	1,819
30									0,86	2,383	0,992	0,89	2,709	0,635	1,679	2,525	1,99	1,447		1,694	2,012	1,497	0,806	1,327	0,672
31																									
32							средние значения		0,89583	1,18163	1,59983	1,12213	1,93037	1,4731	1,44263	1,93653	1,68417	1,8897		1,0551	1,14717	1,6589	1,17517	1,75717	1,72843
33																									

Лабораторная работа 3 - Excel

ОТПРАВКА ЗАБЛОКИРОВАНА С этим файлом возникла проблема. Мы не можем сохранить новые изменения. Чтобы не потерять данные, сохраните его как копию. Сохранить копию

O48 =F.ОБР.ПХ(0,05;7;72)

	A	B	C	D	E	F	G	H	I	J
33										
34		**Имитированные данные наблюдений**								
35									сумма	средние
36	0,89583333	1,0551	0,970533	0,968967	0,945533	0,8687	0,9264	1,1199	7,750967	0,968871
37	1,18163333	1,147167	1,174367	1,0767	1,1843	1,297867	1,148267	1,3653	9,5756	1,19695
38	1,59983333	1,6589	1,625767	1,472533	1,696867	1,751933	1,758833	1,6697	13,23437	1,654296
39	1,12213333	1,175167	1,3128	1,3186	1,258033	1,424533	1,391367	1,390467	10,3931	1,299138
40	1,93036667	1,757167	1,9231	1,692333	1,717867	1,5707	1,6563	1,8352	14,08303	1,760379
41	1,4731	1,728433	1,385	1,611167	1,380667	1,5796	1,327833	1,508267	11,99407	1,499258
42	1,44263333	1,451333	1,510933	1,486867	1,6556	1,5428	1,5022	1,5654	12,15777	1,519721
43	1,93653333	2,0009	1,8152	1,9425	2,044933	1,760867	1,956233	1,919867	15,37703	1,922129
44	1,68416667	1,6149	1,6301	1,680867	1,7099	1,647933	1,4446	1,706633	13,1191	1,639888
45	1,8897	1,6747	2,012433	2,1637	1,8931	2,214667	1,909267	1,7843	15,54187	1,942733
46	15,1559333	15,26377	15,36023	15,41423	15,4868	15,6596	15,0213	15,86503	сумма	
47	1,51559333	1,526377	1,536023	1,541423	1,54868	1,56596	1,50213	1,586503	среднее	
48										
49								общее среднее		1,540336
50						дисперсия фактора случайности				0,012948
51						дисперсия фактора А				0,005716
52						дисперсия фактора В				1,010036
53	-0,0482946	0,100189	0,005975	-0,00099	-0,03168	-0,12579	-0,00426	0,104862		0,802517
54	0,00942625	-0,03582	-0,01827	-0,12134	-0,02099	0,075293	-0,01048	0,122183		1,396257
55	-0,0297196	0,018564	-0,02422	-0,18285	0,034227	0,072014	0,142744	-0,03076		2,559467
56	-0,1522612	-0,11001	0,017975	0,018375	-0,04945	0,099772	0,130435	0,045162		1,259183
57	0,19473042	0,010747	0,167034	-0,06913	-0,05086	-0,2153	-0,06587	0,028654		3,726315
58	-0,0014154	0,243135	-0,10995	0,110821	-0,12694	0,054718	-0,13322	-0,03716		2,170024
59	-0,0523446	-0,05443	-0,00447	-0,03394	0,127535	-0,00254	0,020685	-0,00049		2,081191
60	0,03914708	0,09273	-0,10262	0,019284	0,11446	-0,18689	0,07231	-0,04843		3,750161
61	0,06902208	-0,01103	-0,00547	0,039892	0,061669	-0,01758	-0,15708	0,020579		2,836417
62	-0,0282904	-0,25407	0,074013	0,21988	-0,05798	0,24631	0,00474	-0,2046		3,570966
63										
64	0,00233237	0,010038	3,57E-05	9,83E-07	0,001004	0,015824	1,82E-05	0,010996	0,326573	229,7023
65	8,8854E-05	0,001283	0,000334	0,014723	0,000441	0,005669	0,00011	0,014929	0,117914	
66	0,00088325	0,000345	0,000586	0,033434	0,001171	0,005186	0,020376	0,000946	0,012987	
67	0,02318349	0,012102	0,000323	0,000338	0,002445	0,009954	0,017013	0,00204	0,058177	
68	0,03791994	0,000115	0,0279	0,004779	0,002586	0,046355	0,004339	0,000821	0,048419	
69	2,0034E-06	0,059114	0,012088	0,012281	0,016113	0,002994	0,017747	0,001381	0,001687	
70	0,00273996	0,002962	2E-05	0,001152	0,016265	6,47E-06	0,000428	2,38E-07	0,000425	
71	0,00153249	0,008599	0,01053	0,000372	0,013101	0,034926	0,005229	0,002345	0,145766	
72	0,00476405	0,000122	3E-05	0,001591	0,003803	0,000309	0,024675	0,000423	0,00991	
73	0,00080035	0,064553	0,005478	0,048347	0,003361	0,060668	2,25E-05	0,041861	0,161923	

$$\bar{x}_i = \frac{1}{n}\sum_{j=1}^{n} x_{ji} \qquad \bar{x}_j = \frac{1}{k}\sum_{i=1}^{k} x_{ji} \qquad \bar{x} = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n} x_{ji}$$

$$s_{сл}^2 = \frac{1}{(n-1)(k-1)}\sum_{i=1}^{k}\sum_{j=1}^{n}\left(x_{ji} - \bar{x}_i - \bar{x}_j + \bar{x}\right)^2 \qquad s_A^2 = \frac{n}{n-1}\sum_{i=1}^{k}(\bar{x}_i - \bar{x})^2$$

$$s_B^2 = \frac{k}{k-1}\sum_{j=1}^{n}(\bar{x}_j - \bar{x})^2$$

$$Q_1 = \sum_{i=1}^{k}\sum_{j=1}^{n} x_{ji}^2 \qquad Q_2 = \frac{1}{n}\sum_{i=1}^{k}\left(\sum_{j=1}^{n} x_{ji}\right)^2 \qquad Q_3 = \frac{1}{k}\sum_{j=1}^{n}\left(\sum_{i=1}^{k} x_{ji}\right)^2 \qquad Q_4 = \frac{1}{nk}\left(\sum_{i=1}^{k}\sum_{j=1}^{n} x_{ji}\right)^2$$

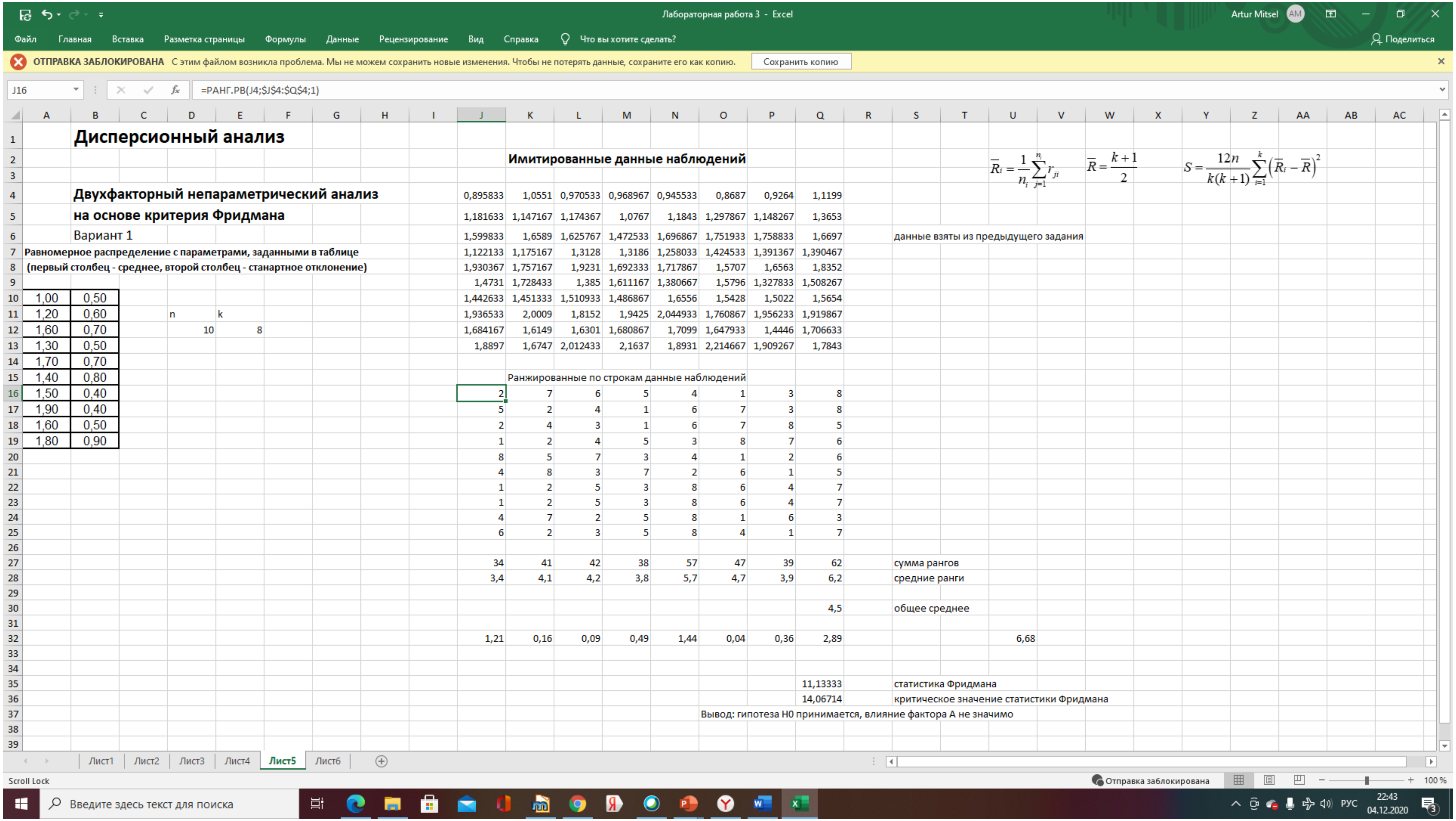
$$s_{сл}^2 = \frac{Q_1 + Q_4 - Q_2 - Q_3}{(k-1)(n-1)} \qquad s_A^2 = \frac{Q_2 - Q_4}{n-1} \qquad s_B^2 = \frac{Q_3 - Q_4}{k-1}$$

	K	L	M	N	O	P	Q	R	S
47					0,441447		статистика Фишера для фактора А		
48					2,139656		критическая статистика Фишера для фактора А		
49				Вывод: так как 0,441 < 2,1396, то гипотеза Н0 принимается,					
50		0,012948		влияние фактора А не значимо					
51		0,005716							
52		1,010036							
53	1,113236	0,941935	0,938896	0,894033	0,75464	0,858217	1,254176		60,07748
54	1,315991	1,379137	1,159283	1,402566	1,684458	1,318516	1,864044		91,69212
55	2,751949	2,643117	2,168354	2,879356	3,06927	3,093495	2,787898		175,1485
56	1,381017	1,723444	1,738706	1,582648	2,029295	1,935901	1,933398		108,0165
57	3,087635	3,698314	2,863992	2,951066	2,467098	2,74333	3,367959		198,3318
58	2,987482	1,918225	2,595858	1,90624	2,495136	1,763141	2,274868		143,8576
59	2,106368	2,28292	2,210772	2,741011	2,380232	2,256605	2,450477		147,8113
60	4,003601	3,294951	3,773306	4,181752	3,100651	3,826849	3,685888		236,4532
61	2,607902	2,657226	2,825313	2,923758	2,715684	2,086869	2,912597		172,1108
62	2,80462	4,049888	4,681598	3,583828	4,904748	3,645299	3,183726		241,5496
64	232,9826	235,9368	237,5986	239,841	245,2231	225,6395	251,6993		

	V	W	X	Y	Z	AA	AB
44	197,7483		189,8623		196,8811		189,8109
47		78,00849	статистика Фишера для фактора В				
48		2,012705	критическая статистика Фишера для фактора В				
49		Вывод: так как 78,00 > 2,01239, то гипотеза Н0 отклоняется,					
50		влияние фактора В значимо					

Лист1 Лист2 Лист3 **Лист4** Лист5 Лист6

Задание 4.4.1



Лабораторная работа 3 - Excel

J16 =РАНГ.РВ(J4;$J$4:$Q$4;1)

Дисперсионный анализ

Двухфакторный непараметрический анализ

на основе критерия Фридмана

Вариант 1

Равномерное распределение с параметрами, заданными в таблице

(первый столбец - среднее, второй столбец - станартное отклонение)

A	B
1,00	0,50
1,20	0,60
1,60	0,70
1,30	0,50
1,70	0,70
1,40	0,80
1,50	0,40
1,90	0,40
1,60	0,50
1,80	0,90

n	k
10	8

Имитированные данные наблюдений

J	K	L	M	N	O	P	Q
0,895833	1,0551	0,970533	0,968967	0,945533	0,8687	0,9264	1,1199
1,181633	1,147167	1,174367	1,0767	1,1843	1,297867	1,148267	1,3653
1,599833	1,6589	1,625767	1,472533	1,696867	1,751933	1,758833	1,6697
1,122133	1,175167	1,3128	1,3186	1,258033	1,424533	1,391367	1,390467
1,930367	1,757167	1,9231	1,692333	1,717867	1,5707	1,6563	1,8352
1,4731	1,728433	1,385	1,611167	1,380667	1,5796	1,327833	1,508267
1,442633	1,451333	1,510933	1,486867	1,6556	1,5428	1,5022	1,5654
1,936533	2,0009	1,8152	1,9425	2,044933	1,760867	1,956233	1,919867
1,684167	1,6149	1,6301	1,680867	1,7099	1,647933	1,4446	1,706633
1,8897	1,6747	2,012433	2,1637	1,8931	2,214667	1,909267	1,7843

данные взяты из предыдущего задания

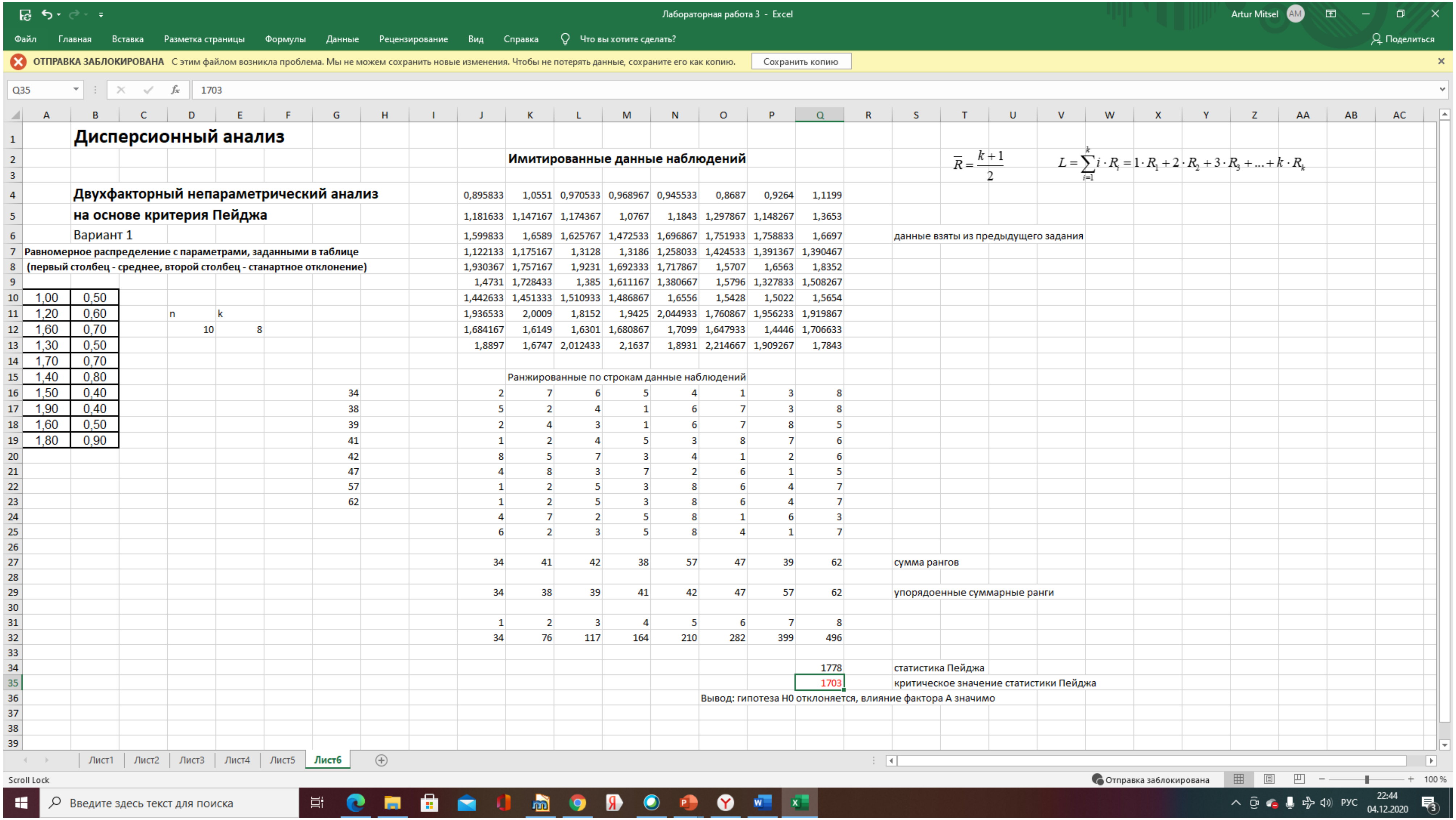
$$\overline{R}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} r_{ji} \qquad \overline{R} = \frac{k+1}{2} \qquad S = \frac{12n}{k(k+1)}\sum_{i=1}^{k}\left(\overline{R}_i - \overline{R}\right)^2$$

Ранжированные по строкам данные наблюдений

J	K	L	M	N	O	P	Q	
2	7	6	5	4	1	3	8	
5	2	4	1	6	7	3	8	
2	4	3	1	6	7	8	5	
1	2	4	5	3	8	7	6	
8	5	7	3	4	1	2	6	
4	8	3	7	2	6	1	5	
1	2	5	3	8	6	4	7	
1	2	5	3	8	6	4	7	
4	7	2	5	8	1	6	3	
6	2	3	5	8	4	1	7	
34	41	42	38	57	47	39	62	сумма рангов
3,4	4,1	4,2	3,8	5,7	4,7	3,9	6,2	средние ранги
							4,5	общее среднее
1,21	0,16	0,09	0,49	1,44	0,04	0,36	2,89	6,68
							11,13333	статистика Фридмана
							14,06714	критическое значение статистики Фридмана

Вывод: гипотеза H0 принимается, влияние фактора А не значимо

Задание 4.4.2



Дисперсионный анализ

$$\overline{R} = \frac{k+1}{2} \qquad L = \sum_{i=1}^{k} i \cdot R_i = 1 \cdot R_1 + 2 \cdot R_2 + 3 \cdot R_3 + \ldots + k \cdot R_k$$

Двухфакторный непараметрический анализ

на основе критерия Пейджа

Вариант 1

Равномерное распределение с параметрами, заданными в таблице

(первый столбец - среднее, второй столбец - станартное отклонение)

1,00	0,50
1,20	0,60
1,60	0,70
1,30	0,50
1,70	0,70
1,40	0,80
1,50	0,40
1,90	0,40
1,60	0,50
1,80	0,90

n	k
10	8

Имитированные данные наблюдений

0,895833	1,0551	0,970533	0,968967	0,945533	0,8687	0,9264	1,1199
1,181633	1,147167	1,174367	1,0767	1,1843	1,297867	1,148267	1,3653
1,599833	1,6589	1,625767	1,472533	1,696867	1,751933	1,758833	1,6697
1,122133	1,175167	1,3128	1,3186	1,258033	1,424533	1,391367	1,390467
1,930367	1,757167	1,9231	1,692333	1,717867	1,5707	1,6563	1,8352
1,4731	1,728433	1,385	1,611167	1,380667	1,5796	1,327833	1,508267
1,442633	1,451333	1,510933	1,486867	1,6556	1,5428	1,5022	1,5654
1,936533	2,0009	1,8152	1,9425	2,044933	1,760867	1,956233	1,919867
1,684167	1,6149	1,6301	1,680867	1,7099	1,647933	1,4446	1,706633
1,8897	1,6747	2,012433	2,1637	1,8931	2,214667	1,909267	1,7843

данные взяты из предыдущего задания

Ранжированные по строкам данные наблюдений

34	2	7	6	5	4	1	3	8
38	5	2	4	1	6	7	3	8
39	2	4	3	1	6	7	8	5
41	1	2	4	5	3	8	7	6
42	8	5	7	3	4	1	2	6
47	4	8	3	7	2	6	1	5
57	1	2	5	3	8	6	4	7
62	1	2	5	3	8	6	4	7
	4	7	2	5	8	1	6	3
	6	2	3	5	8	4	1	7

34	41	42	38	57	47	39	62	сумма рангов
34	38	39	41	42	47	57	62	упорядоенные суммарные ранги
1	2	3	4	5	6	7	8	
34	76	117	164	210	282	399	496	

1778 статистика Пейджа

1703 критическое значение статистики Пейджа

Вывод: гипотеза H0 отклоняется, влияние фактора А значимо

Приложение 4.1

Варианты заданий по лабораторной работе №4

Варианты задания 4.1

Номер варианта	1	2	3	4	5	6	7	8	9	10
μ	0	0,5	1,0	2,0	2,5	3,0	3,5	4,0	4,5	5,0
σ	0,2	0,2	0,2	0,5	0,5	0,5	1	1	1	1,5

Варианты задания 4.2

Номер варианта	1	2	3	4	5	6	7	8	9	10
n	20	30	40	35	25	100	90	80	70	50
p	0,4	0,5	0,7	0,4	0,8	0,3	0,6	0,7	0,1	0,5

Варианты задания 4.3

m	1		2		3		4		5	
	μ_m	σ_m	μ_m	σ_m	μ_m	σ_m	μ_m	σ_m	μ_m	σ_m
1	1	0,5	1,4	0,5	1	1,5	2	0,5	1	1,5
2	1,2	0,6	2,2	0,6	1,2	1,6	2,2	0,6	1,2	2,6
3	1,6	0,7	1,6	0,7	1,6	0,7	1,6	0,7	1,6	1,7
4	1,3	0,5	1,3	0,5	1,3	0,5	2,3	0,5	1,3	0,5
5	1,7	0,7	2,5	0,7	1,7	0,7	1,7	0,7	1,7	1,7
6	1.4	0,8	1.4	0,8	1.4	1,8	1.4	0,8	1.4	0,8
7	1,5	0,4	1,5	0,4	1,5	0,4	1,5	0,4	1,5	1,4
8	1,9	0,4	2,9	0,4	1,9	0,4	2,9	0,4	1,9	0,4
9	1,6	0,5	1,6	0,5	1,6	0,5	1,6	0,5	1,6	1,5
10	1,8	0,9	1,8	0,9	1,8	0,9	2,8	0,9	1,8	0,9
m	6		7		8		9		10	
	μ_m	σ_m	μ_m	σ_m	μ_m	σ_m	μ_m	σ_m	μ_m	σ_m
1	1	0,5	1,4	0,05	0	1,5	1	1,5	2	1,5
2	3,2	0,6	2,2	0,6	0,2	1,6	1,2	0,6	1,2	2,6
3	1,6	0,7	1,6	0,7	1,6	0,7	1,6	0,7	1,6	1,7

4	1,3	0,5	1,3	0,5	0,3	0,5	2,3	0,5	1,3	0,5
5	3,7	0,7	2,5	0,07	1,7	0,7	1,7	0,7	2,7	1,7
6	1.4	0,8	1.4	0,8	1.4	1,8	1.4	1,8	0.4	0,8
7	1,5	0,4	1,5	0,4	0,5	0,4	1,5	0,4	1,5	1,4
8	1,9	0,4	2,9	0,4	1,9	0,4	2,9	0,4	1,9	0,4
9	3,6	0,5	1,6	0,5	1,6	0,5	1,6	0,5	1,6	1,5
10	1,8	0,9	1,8	0,09	0,8	0,9	2,8	0,9	1,8	0,9

ЛАБОРАТОРНАЯ РАБОТА №5

Тема. Корреляционный анализ случайных данных

Цель работы:

Рассчитать параметрические и непараметрические коэффициенты корреляции.

5.1. Вычисление параметрических коэффициентов корреляции

1) Сгенерировать средствами пакета EXCEL 5 выборок из 10 значений случайной величины с нормальным законом $N(\mu, \sigma^2)$. Эти 5 выборок будем использовать в качестве независимых признаков $(x_1, x_2, x_3, x_4, x_5)$. Варианты значений параметров μ, σ^2 приведены в приложении 4.

2) Рассчитать зависимый признак

$y_i = a_0 + a_1 \cdot x_{1i} + a_2 \cdot x_{2i} + a_3 \cdot x_{3i} + a_4 \cdot x_{4i} + a_5 \cdot x_{5i} + \varepsilon_i$, $i = 1,...,10$,

где $a_0, a_1, ..., a_5$ – параметры (их значения приведены в приложении 4);

ε_i – случайная погрешность с нормальным законом распределения $\varepsilon_i \sim N(0, \sigma_i^2)$, где $\sigma_i = 0,2 \cdot M(y_i / x)$.

Здесь $M(y_i / x) = a_0 + a_1 \cdot x_{1i} + a_2 \cdot x_{2i} + a_3 \cdot x_{3i} + a_4 \cdot x_{4i} + a_5 \cdot x_{5i}$

5.1.1 Парные коэффициенты корреляции

3) Выполнить следующие расчеты:

- вычислить выборочные средние и дисперсии зависимого признака y и независимых признаков x_1, x_2, x_3, x_4, x_5, а также средние значения произведений $\overline{y \cdot x_j}$ по формулам:

- $s_y^2 = \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)^2$; $s_{x_j}^2 = \frac{1}{n}\sum_{i=1}^{n} x_{ij}^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_{ij}\right)^2$; $j = 1,...,5$;

- $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$; $\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$; $\overline{y \cdot x}_j = \frac{1}{n}\sum_{i=1}^{n} y_i \cdot x_{ij}$.

  Здесь x_{ij} – i-е значение случайной величины x из j-й выборки.

- вычислить парные коэффициенты корреляции между зависимым признаком y и независимыми признаками x_1, x_2, x_3, x_4, x_5 (коэффициенты корреляции Пирсона) по формулам: $r_{yx_j} = \dfrac{\overline{y \cdot x_j} - \bar{y} \cdot \bar{x}_j}{s_y \cdot s_{x_j}}$.
- скорректировать коэффициент корреляции (так как $n < 15$) по формуле $r^* = r\left(1 + \dfrac{1-r^2}{2(n-3)}\right)$.
- вычислить t-статистики $t_j = \sqrt{\dfrac{r_{yx_j}^2 (n-2)}{1 - r_{yx_j}^2}}$.

4) Проверить гипотезу о значимости коэффициентов корреляции. Если $t_j > t_{0,95}(n-2=8)$, то коэффициент r_{yx_j} значимый, и, следовательно, связь между y и x_j статистически значима.

5.1.2 Множественный коэффициент корреляции

5) Вычислить парные коэффициенты корреляции между зависимыми признаками $r_{x_i x_j} = \dfrac{\overline{x_i \cdot x_j} - \bar{x}_i \cdot \bar{x}_j}{s_{x_i} \cdot s_{x_j}};\ j > i,\ i = 1,...,5$, где $\overline{x_i \cdot x_j} = \dfrac{1}{n}\sum_{k=1}^{n} x_{ki} \cdot x_{kj};\ j > i;\ i = 1,...,5$

Вычислить множественный коэффициент корреляции между результирующим признаком Y и факторными признаками $x_1, x_1, ..., x_5$ по формуле

$$R_{y/x_1,x_2,...,x_m} = \sqrt{1 - \frac{|\rho|}{|\rho_1|}},$$

где $|\rho|$ – определитель матрицы парной корреляции

$$\rho = \begin{pmatrix} 1 & \rho_{yx_1} & \rho_{yx_2} & \rho_{yx_2} & \rho_{yx_4} & \rho_{yx_5} \\ \rho_{x_1y} & 1 & \rho_{x_1x_2} & \rho_{x_1x_3} & \rho_{x_1x_4} & \rho_{x_1x_5} \\ \rho_{x_2y} & \rho_{x_2x_1} & 1 & \rho_{x_2x_3} & \rho_{x_2x_3} & \rho_{x_2x_5} \\ \rho_{x_3y} & \rho_{x_3x_1} & \rho_{x_3x_2} & 1 & \rho_{x_3x_4} & \rho_{x_3x_5} \\ \rho_{x_4y} & \rho_{x_4x_1} & \rho_{x_4x_2} & \rho_{x_4x_3} & 1 & \rho_{x_4x_5} \\ \rho_{x_5y} & \rho_{x_5x_1} & \rho_{x_5x_2} & \rho_{x_5x_3} & \rho_{x_5x_4} & 1 \end{pmatrix};$$

$|\rho_1|$ – алгебраическое дополнение элемента ρ_{11}.

6) Вычислить скорректированный коэффициент корреляции:

$$R_{y/x_1,x_2,...,x_k} = \sqrt{1-(1-R^2)\frac{n-1}{n-k-1}},$$

7) Вычислить статистику Фишера $F = \dfrac{\frac{1}{2}R^2_{y/x_1,...,x_5}}{\frac{1}{n-6}\left(1-R^2_{y/x_1,...,x_5}\right)}$;

8) Проверить гипотезу о значимости множественного коэффициента корреляции. Если $F > F_{0,95}(v_1 = 5,\ v_2 = n-6)$, то множественный коэффициент корреляции считается значимым.

5.2. Вычисление непараметрических коэффициентов корреляции

1) Сгенерировать средствами пакета EXCEL 5 выборок из 10 значений случайной величины с биномиальным законом распределения $B(n,\ p)$. Эти 5 выборок будем использовать в качестве независимых признаков (x_1, x_2, x_3, x_4, x_5). Варианты значений параметров $n,\ p$ приведены в приложении 4.
2) Вычислить ранги элементов матрицы X.

5.2.1 Коэффициент ранговой корреляции Спирмена

3) Вычислить:

- величины $d_k^2(i,j) = \left(R_{k,x_i} - R_{k,x_j}\right)^2;\ i \neq j,\ i,j = 1,...,5;\ k = 1,...,10$;
- коэффициенты ранговой корреляции Спирмена $\rho_{x_i/y_j} = 1 - \dfrac{6\sum_{k=1}^{n} d_k^2(i,j)}{n(n^2-1)}$;
- t-статистики $t_{ij} = \left|\rho_{x_i/y_j}\right|\sqrt{\dfrac{n-2}{1-\rho^2_{x_i/y_j}}}$

4) Проверить значимость коэффициентов корреляции. Если $t_{ij} > t_{0,95}(n-2)$, то коэффициент ρ_{x_i/y_j} считается значимым.

5.2.2 Коэффициент ранговой корреляции Кендалла

5) Рассматриваются все комбинации пар столбцов исходной таблицы данных – (1;2), (1;3), (1;4), (1;5), (2;3), (2;4), (2;5), (3;4), (3;5), (4;5). Первый столбец обозначим X, второй – Y.
   - В каждой паре столбцов значения первого столбца X упорядочиваются по возрастанию, а значения второго столбца Y располагаются в порядке, соответствующем значениям X
   - для каждого ранга Y определяется число следующих за ним значений рангов, превышающих его по величине. Суммируя эти числа, определяем величину P (число последовательностей) — меру соответствия последовательностей рангов X и Y (см. пример в лекции);
   - для каждого ранга Y определяется число следующих за ним рангов, меньших его величины. Суммируя величины, получаем величину Q (число инверсий);
   - определяется разность по всем членам ряда $S = P - Q$ и вычисляется τ. Связь между признаками можно признать статистически значимой, если значение коэффициента корреляции $|\tau| > \tau_\alpha = u_\alpha \sqrt{\dfrac{2(2n+5)}{9n(n-1)}}$.

5.2.3 Коэффициент конкордации (множественный коэффициент ранговой корреляции)

6) Проранжировать столбцы исходной таблицы $\{x_{ji}\}$ (наблюдения) их рангами $\{r_{ji}\}$ не упорядочивая табличные данные.
7) Для каждой j -й строки таблицы вычислить:

   a. сумму рангов $R_j = \sum_{i=1}^{5} r_{ji}$ и квадрат суммы R_j^2;

b. сумму рангов по всей совокупности ранг $R = \sum_{j=1}^{10} R_j$ и $R^2 = \sum_{j=1}^{10} R_j^2$;

c. вычислить коэффициент конкордации

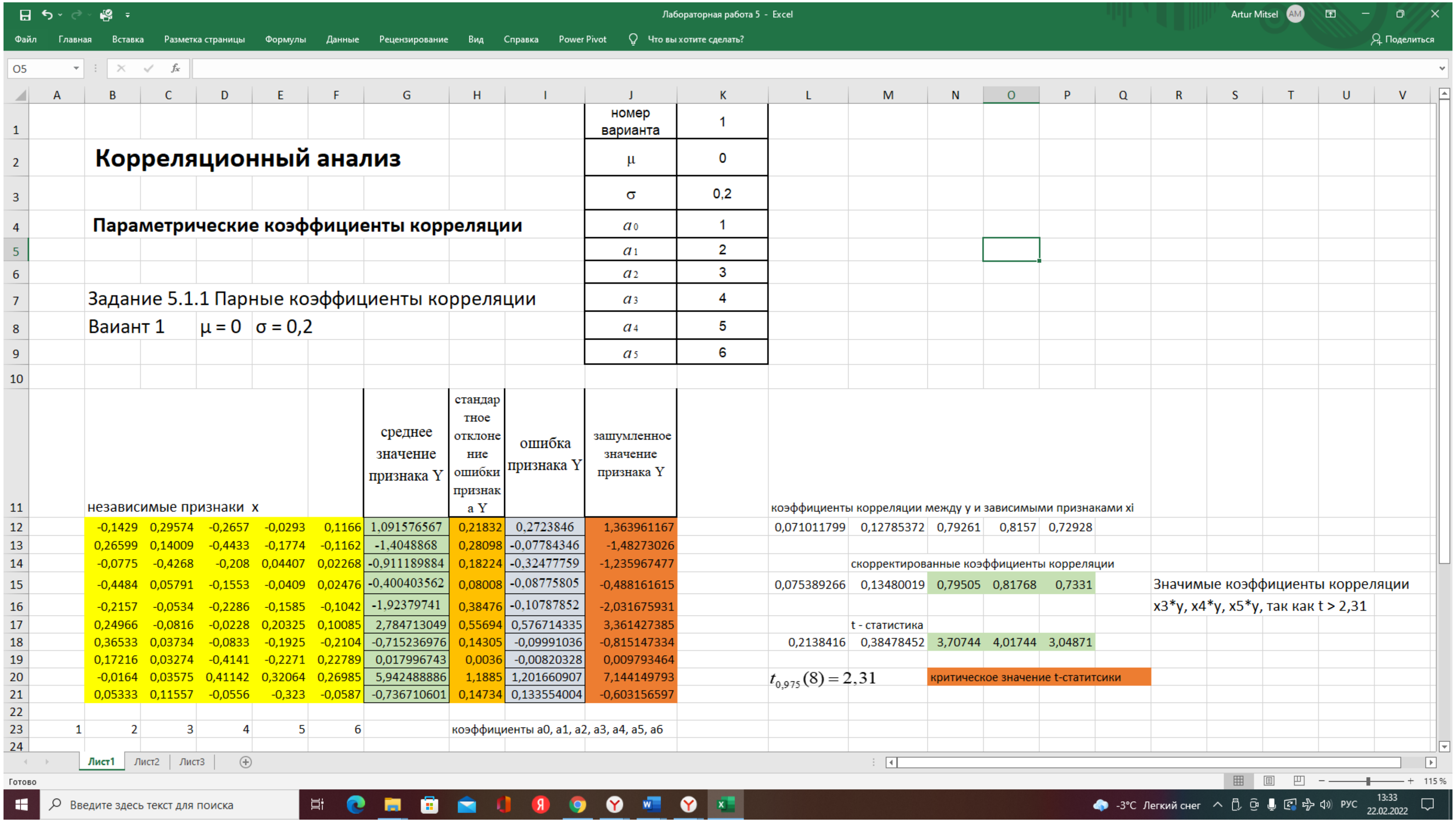
$$W = \frac{12\left(R^2 - \frac{R^2}{n}\right)}{m^2(n^3 - n)}, \; m = 5, \; n = 10.$$

8) Проверить значимость связи между признаками. Если $W > W_\alpha$, где $W_\alpha = \frac{1}{m(n-1)}\chi_\alpha^2(n-1)$, то с вероятностью α корреляция между признаками признается значимой. Если среди последовательностей рангов есть совпадения, то коэффициент конкордации следует вычислять по формуле

$$W = \frac{12\left(R^2 - \frac{R^2}{n}\right)}{m^2(n^3 - 1) - m\sum_{j=1}^{m} T_j},$$

где $T_j = \sum_{k=1}^{l}\left(t_k^3 - t_k\right)$, t_k – количество совпавших рангов в k-й последовательности; l – число связок (групп с одинаковыми значениями) в ранговой последовательности j-ой строки (эксперта). Совпавшим рангам присваиваются средние ранги.

5.3 Excel-скриншоты заданий лабораторной работы №5



Корреляционный анализ

Параметрические коэффициенты корреляции

номер варианта	1
μ	0
σ	0,2
a_0	1
a_1	2
a_2	3
a_3	4
a_4	5
a_5	6

Задание 5.1.1 Парные коэффициенты корреляции

Ваиант 1 μ = 0 σ = 0,2

независимые признаки x					среднее значение признака Y	стандартное отклонение ошибки признака Y	ошибка признака Y	зашумленное значение признака Y
-0,1429	0,29574	-0,2657	-0,0293	0,1166	1,091576567	0,21832	0,2723846	1,363961167
0,26599	0,14009	-0,4433	-0,1774	-0,1162	-1,4048868	0,28098	-0,07784346	-1,48273026
-0,0775	-0,4268	-0,208	0,04407	0,02268	-0,911189884	0,18224	-0,32477759	-1,235967477
-0,4484	0,05791	-0,1553	-0,0409	0,02476	-0,400403562	0,08008	-0,08775805	-0,488161615
-0,2157	-0,0534	-0,2286	-0,1585	-0,1042	-1,92379741	0,38476	-0,10787852	-2,031675931
0,24966	-0,0816	-0,0228	0,20325	0,10085	2,784713049	0,55694	0,576714335	3,361427385
0,36533	0,03734	-0,0833	-0,1925	-0,2104	-0,715236976	0,14305	-0,09991036	-0,815147334
0,17216	0,03274	-0,4141	-0,2271	0,22789	0,017996743	0,0036	-0,00820328	0,009793464
-0,0164	0,03575	0,41142	0,32064	0,26985	5,942488886	1,1885	1,201660907	7,144149793
0,05333	0,11557	-0,0556	-0,323	-0,0587	-0,736710601	0,14734	0,133554004	-0,603156597

1 2 3 4 5 6 коэффициенты a0, a1, a2, a3, a4, a5, a6

коэффициенты корреляции между у и зависимыми признаками xi

0,071011799	0,12785372	0,79261	0,8157	0,72928

скорректированные коэффициенты корреляции

0,075389266	0,13480019	0,79505	0,81768	0,7331

t - статистика

0,2138416	0,38478452	3,70744	4,01744	3,04871

$t_{0,975}(8) = 2,31$ критическое значение t-статитсики

Значимые коэффициенты корреляции x3*y, x4*y, x5*y, так как t > 2,31

Задание 5.1.2 Множественный коэффициент корреляции

матрица корреляции

1	0,07101	0,12785	0,79261	0,8157	0,72928
0,07101	1	0,03524	-0,0583	-0,1385	-0,1821
0,12785	0,03524	1	-0,0803	-0,2778	0,02137
0,79261	-0,0583	-0,0803	1	0,66418	0,33129
0,8157	-0,1385	-0,2778	0,66418	1	0,58796
0,72928	-0,1821	0,02137	0,33129	0,58796	1

0,00032 определитель Н

0,29468 определитель Н1

$$R_{y/x_1,x_2,...,x_m} = \sqrt{1-\frac{|\rho|}{|\rho_1|}}$$

$$\tilde{R}_{y/x_1,x_2,...,x_k} = \sqrt{1-(1-R^2)\frac{n-1}{n-k-1}}$$

$$F = \frac{\frac{1}{2}R^2_{y/x_1,...,x_5}}{\frac{1}{n-6}\left(1-R^2_{y/x_1,...,x_5}\right)}$$

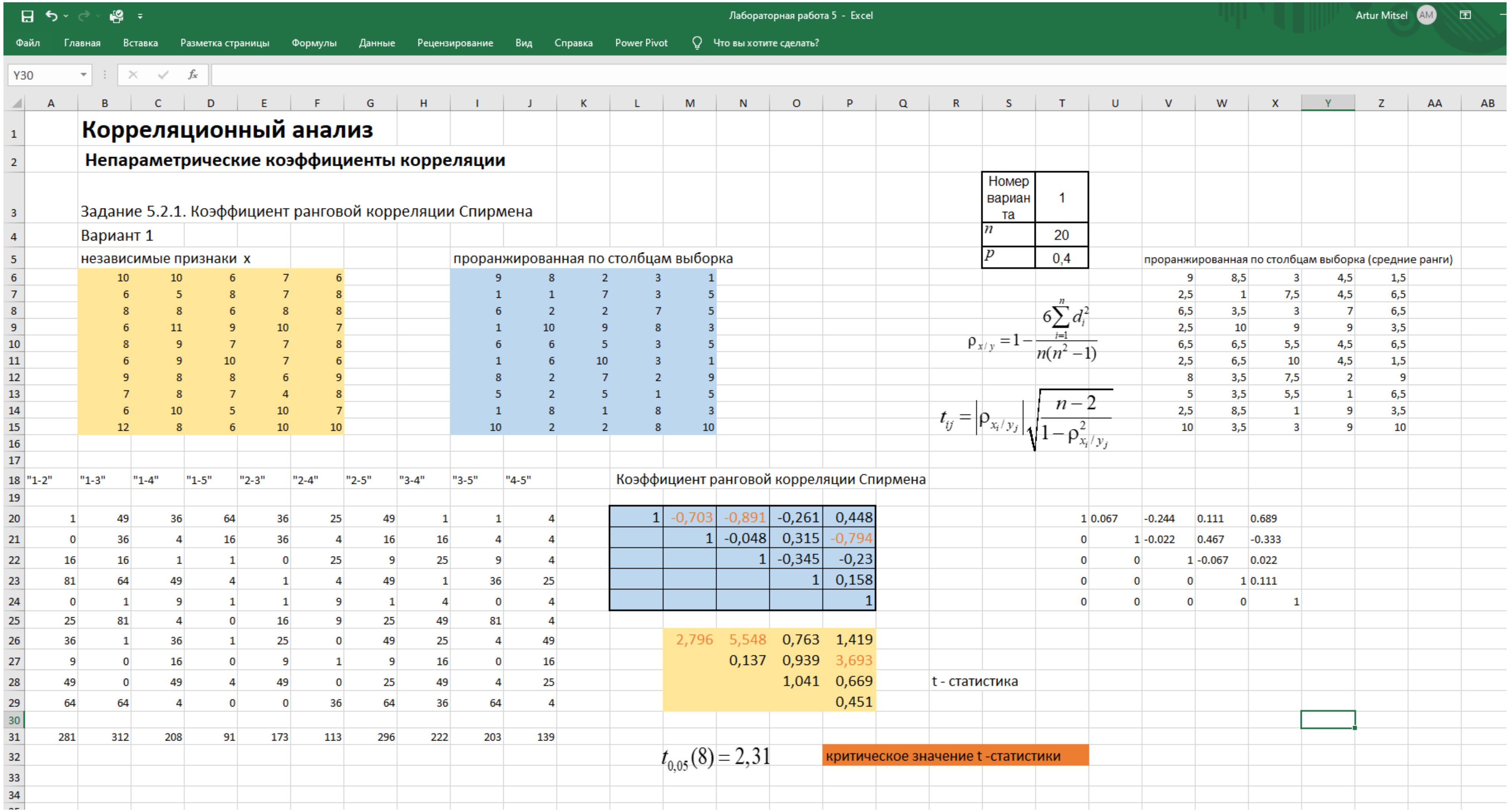
0,999452728 множественный коэффициент корреляции

0,998768217 скорректированный множественный коэффициент корреляции

810,3316785 статисика Фишера

6,256056502 критическое значение статистики Фишера

вывод: множественный коэффициент корреляции значим



Корреляционный анализ

Непараметрические коэффициенты корреляции

Задание 5.2.1. Коэффициент ранговой корреляции Спирмена

Вариант 1

Номер варианта	1
n	20
p	0,4

независимые признаки x

10	10	6	7	6
6	5	8	7	8
8	8	6	8	8
6	11	9	10	7
8	9	7	7	8
6	9	10	7	6
9	8	8	6	9
7	8	7	4	8
6	10	5	10	7
12	8	6	10	10

проранжированная по столбцам выборка

9	8	2	3	1
1	1	7	3	5
6	2	2	7	5
1	10	9	8	3
6	6	5	3	5
1	6	10	3	1
8	2	7	2	9
5	2	5	1	5
1	8	1	8	3
10	2	2	8	10

$$\rho_{x/y} = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

$$t_{ij} = \left|\rho_{x_i/y_j}\right| \sqrt{\frac{n-2}{1-\rho^2_{x_i/y_j}}}$$

проранжированная по столбцам выборка (средние ранги)

9	8,5	3	4,5	1,5
2,5	1	7,5	4,5	6,5
6,5	3,5	3	7	6,5
2,5	10	9	9	3,5
6,5	6,5	5,5	4,5	6,5
2,5	6,5	10	4,5	1,5
8	3,5	7,5	2	9
5	3,5	5,5	1	6,5
2,5	8,5	1	9	3,5
10	3,5	3	9	10

"1-2"	"1-3"	"1-4"	"1-5"	"2-3"	"2-4"	"2-5"	"3-4"	"3-5"	"4-5"
1	49	36	64	36	25	49	1	1	4
0	36	4	16	36	4	16	16	4	4
16	16	1	1	0	25	9	25	9	4
81	64	49	4	1	4	49	1	36	25
0	1	9	1	1	9	1	4	0	4
25	81	4	0	16	9	25	49	81	4
36	1	36	1	25	0	49	25	4	49
9	0	16	0	9	1	9	16	0	16
49	0	49	4	49	0	25	49	4	25
64	64	4	0	0	36	64	36	64	4
281	312	208	91	173	113	296	222	203	139

Коэффициент ранговой корреляции Спирмена

1	-0,703	-0,891	-0,261	0,448
	1	-0,048	0,315	-0,794
		1	-0,345	-0,23
			1	0,158
				1

1	0.067	-0.244	0.111	0.689
0	1	-0.022	0.467	-0.333
0	0	1	-0.067	0.022
0	0	0	1	0.111
0	0	0	0	1

t - статистика

2,796	5,548	0,763	1,419
	0,137	0,939	3,693
		1,041	0,669
			0,451

$t_{0,05}(8) = 2,31$ критическое значение t -статистики

AC73

Задание 5.2.2. Коэффициент ранговой корреляции Кендалла

$$\tau_{x,y} = \frac{2(P_{x,y} - Q_{x,y})}{n(n-1)} = \frac{2S_{x,y}}{n(n-1)} \qquad |\tau_{x,y}| > \tau_\alpha = u_\alpha \sqrt{\frac{2(2n+5)}{9n(n-1)}} \qquad u_{0,95} = 1,645$$

проранжированная по столбцам выборка

					"1-2"		"1-3"		"1-4"		"1-5"		"2-3"		"2-4"		"2-5"	
9	8	2	3	9	1	1	1	7	1	3	1	1	1	7	1	3	1	1
1	1	7	3	1	1	10	1	9	1	8	1	1	2	2	2	7	2	6
6	2	2	7	6	1	6	1	10	1	3	1	1	2	7	2	2	2	8
1	10	9	8	1	1	8	1	1	1	8	1	1	2	5	2	1	2	5
6	6	5	3	6	5	2	5	5	5	1	5	5	2	2	2	8	2	10
1	6	10	3	1	6	2	6	2	6	7	6	6	6	5	6	3	6	6
8	2	7	2	8	6	6	6	5	6	3	6	6	6	10	6	3	6	1
5	2	5	1	5	8	2	8	7	8	2	8	8	8	2	8	3	8	9
1	8	1	8	1	9	8	9	2	9	3	9	9	8	1	8	8	8	1
10	2	2	8	10	10	2	10	2	10	8	10	10	10	9	10	8	10	1

"1-2"

1								
1	0							
1	0	1						
1	0	0	0					
1	0	0	0	1				
1	0	1	0	1	1			
1	0	0	0	1	1	0		
1	0	1	1	1	1	1	1	
1	0	0	0	1	1	0	1	0
							P12=	25
0								
0	1							
0	1	0						
0	1	1	1					
0	1	1	1	0				
0	1	0	1	0	0			
0	1	1	1	0	0	1		
0	1	0	0	0	0	0	0	
0	1	1	1	0	0	1	0	1
							Q12=	20
							tau12=	0,11111

"3-4"		"3-5"		"4-5"	
1	8	1	1	1	5
2	3	2	9	2	8
2	7	2	6	3	9
2	8	2	10	3	1
5	3	5	6	3	6
5	1	5	5	3	1
7	3	7	1	7	6
7	2	7	8	8	1
9	8	9	1	8	1
10	3	10	1	8	10

Коэффициент ранговой корреляции Кендала

1	0,111	-0,244	0,111	1
0	1	-0,022	0,467	0,022
0	0	1	-0,067	-0,2
0	0	0	1	0,067
0	0	0	0	1

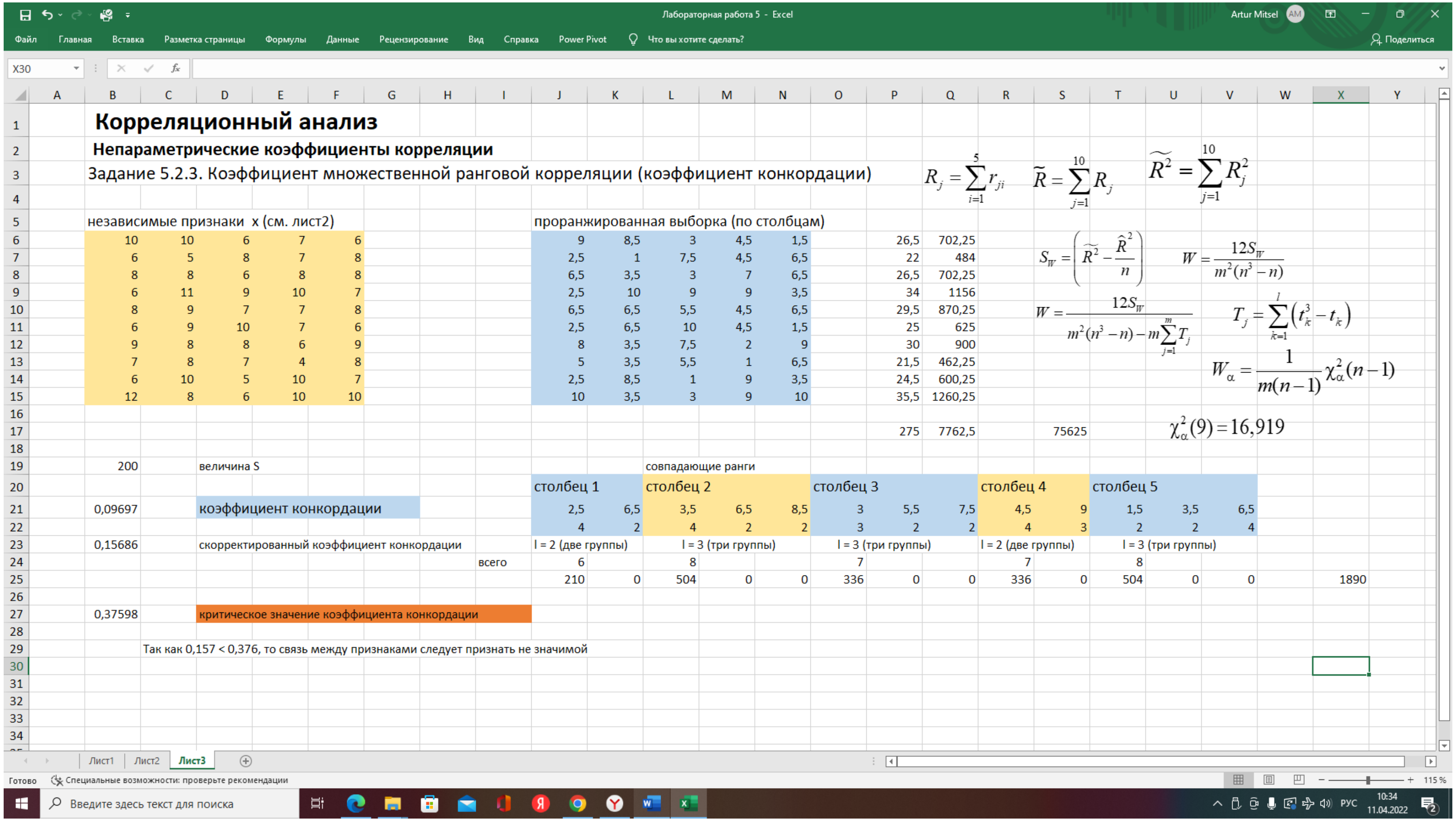
расчеты выполнены в маткаде

Вывод: коэффициенты корреляции tau15 и tau24 значимы
остальные коэффициенты корреляции не значимы

0,4087 критическая статистика

коэффициенты Кендалла по средним рангам

1	0.067	-0.244	0.111	0.689
0	1	-0.022	0.467	-0.333
0	0	1	-0.067	0.022
0	0	0	1	0.111
0	0	0	0	1



Корреляционный анализ

Непараметрические коэффициенты корреляции

Задание 5.2.3. Коэффициент множественной ранговой корреляции (коэффициент конкордации)

$$R_j = \sum_{i=1}^{5} r_{ji} \qquad \widetilde{R} = \sum_{j=1}^{10} R_j \qquad \widetilde{R^2} = \sum_{j=1}^{10} R_j^2$$

независимые признаки x (см. лист2)					проранжированная выборка (по столбцам)						
10	10	6	7	6	9	8,5	3	4,5	1,5	26,5	702,25
6	5	8	7	8	2,5	1	7,5	4,5	6,5	22	484
8	8	6	8	8	6,5	3,5	3	7	6,5	26,5	702,25
6	11	9	10	7	2,5	10	9	9	3,5	34	1156
8	9	7	7	8	6,5	6,5	5,5	4,5	6,5	29,5	870,25
6	9	10	7	6	2,5	6,5	10	4,5	1,5	25	625
9	8	8	6	9	8	3,5	7,5	2	9	30	900
7	8	7	4	8	5	3,5	5,5	1	6,5	21,5	462,25
6	10	5	10	7	2,5	8,5	1	9	3,5	24,5	600,25
12	8	6	10	10	10	3,5	3	9	10	35,5	1260,25
										275	7762,5

75625

$$S_W = \left(\widetilde{R^2} - \frac{\widehat{R}^2}{n}\right) \qquad W = \frac{12S_W}{m^2(n^3-n)}$$

$$W = \frac{12S_W}{m^2(n^3-n) - m\sum_{j=1}^{m} T_j} \qquad T_j = \sum_{k=1}^{l}\left(t_k^3 - t_k\right)$$

$$W_\alpha = \frac{1}{m(n-1)}\chi_\alpha^2(n-1)$$

$$\chi_\alpha^2(9) = 16,919$$

200 величина S

0,09697 коэффициент конкордации

0,15686 скорректированный коэффициент конкордации

0,37598 критическое значение коэффициента конкордации

совпадающие ранги

	столбец 1		столбец 2			столбец 3			столбец 4		столбец 5			
	2,5	6,5	3,5	6,5	8,5	3	5,5	7,5	4,5	9	1,5	3,5	6,5	
	4	2	4	2	2	3	2	2	4	3	2	2	4	
	l = 2 (две группы)		l = 3 (три группы)			l = 3 (три группы)			l = 2 (две группы)		l = 3 (три группы)			
всего	6		8			7			7		8			
	210	0	504	0	0	336	0	0	336	0	504	0	0	1890

Так как 0,157 < 0,376, то связь между признаками следует признать не значимой

Приложение 5.1

Варианты заданий по лабораторной работе №5

Варианты задания 5.1

Номер варианта	1	2	3	4	5	6	7	8	9	10
μ	0	0,5	1,0	2,0	2,5	3,0	3,5	4,0	4,5	5,0
σ	0,2	0,2	0,2	0,5	0,5	0,5	1	1	1	1,5
a_0	1	2	3	4	5	1	-2	3	-4	5
a_1	2	3	4	5	6	-2	3	-4	5	-6
a_2	3	4	5	6	7	3	-4	5	-6	7
a_3	4	5	6	7	8	-4	5	-6	7	-8
a_4	5	6	7	8	9	5	-6	7	-8	9
a_5	6	7	8	9	10	-6	7	-8	9	-10

Варианты задания 5.2

Номер варианта	1	2	3	4	5	6	7	8	9	10
n	20	30	40	35	25	100	90	80	70	50
p	0,4	0,5	0,7	0,4	0,8	0,3	0,6	0,7	0,1	0,5

ЛАБОРАТОРНАЯ РАБОТА №6

Тема. Линейная регрессия

Цель работы:

Оценка уравнения линейной регрессии на основе выборочных данных

6.1. Необходимые сведения из теории

6.1.1. Построение модели парной регрессии

Рассмотрим линейную по коэффициентам модель парной регрессии:

$$y = f(x) + \varepsilon = \beta_0 + \beta_1 f_1(x) + \beta_2 f_2(x) + \ldots + \beta_k f_k(x) + \varepsilon, \qquad (6.1)$$

где ε - случайная величина с математическим ожиданием равным нулю и дисперсией σ^2.

Полагая, $x_j = f_j(x), \quad j = \overline{1,k}$ перейдем к модели множественной линейной регрессии:

$$y = f(x) + \varepsilon = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon. \qquad (6.2)$$

Пусть для оценки неизвестных параметров $\beta_j,\ j = \overline{0,k}$ уравнения регрессии (5.2) взята выборка объемом n из значений величин $(Y, X_1, X_2, \ldots, X_k)$. Тогда

$$Y = XB + \varepsilon,$$

где $Y = (y_1, y_2 \ldots, y_n)^T$ - вектор значений переменной y;

$B = (\beta_0, \beta_1, \ldots, \beta_k)^T$ - вектор параметров модели;

$\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^T$ – вектор ошибок, где $\varepsilon_i \in \mathrm{N}(0, \sigma^2)$ и независимы;

X - матрица исходных данных переменных X_j размерами $n \times (k+1)$. Первый столбец матрицы X содержит единицы (значения фиктивной переменной x_0), остальные столбцы значения переменных $x_1, x_2, \ldots, x_k$:

$$X = \begin{pmatrix} 1 & x_1^1 & \cdots & x_k^1 \\ 1 & x_1^2 & \cdots & x_k^2 \\ & & \cdots & \\ 1 & x_1^n & \cdots & x_k^n \end{pmatrix}.$$

Для нахождения оценки B^* вектора параметров $B=\left(\beta_0,\beta_1,\ldots,\beta_k\right)^T$ используем метод наименьших квадратов, согласно которому в качестве оценок $\beta_0^*,\beta_1^*,\ldots,\beta_k^*$ берутся такие, которые минимизируют сумму квадратов Q отклонений значений y_i от $f(\vec{x}_i)$:

$$Q=\sum_{i=1}^{n}\left(y_i - f(\vec{x}_i)\right)^2 =\sum_{i=1}^{n}\varepsilon_i^2 =\varepsilon^T\varepsilon=\left(Y-XB\right)^T\left(Y-XB\right). \tag{6.3}$$

Оценка B^* метода наименьших квадратов имеет вид:

$$B^*=\left(X^TX\right)^{-1}X^TY. \tag{6.4}$$

6.1.2. Оценка погрешности регрессии

Качество регрессионной модели можно оценить, используя оценку s^2 дисперсии предсказания σ^2:

$$s^2=\frac{1}{n-k-1}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2=\frac{1}{n-k-1}\sum_{i=1}^{n}e^2=\frac{1}{n-k-1}e^Te,$$ где $\hat{y}_i=\beta_0^*+\beta_1^*x_i+\ldots\beta_k^*x_k$.

Качество модели также можно оценить с использованием оценки коэффициента детерминации: $R^2=\dfrac{\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}$.

Чем ближе значения R^2 к 1, тем большую долю дисперсии величины Y объясняет модель регрессии.

Оценка дисперсии коэффициента β_j находится по формуле: $s_j^2=s^2\left[(X^TX)^{-1}\right]_{jj}$, где $\left[(X^TX)^{-1}\right]_{jj}$ соответствующий диагональный элемент матрицы $(X^TX)^{-1}$.

Доверительные интервал для σ^2 находится с использованием статистики $\chi^2=(n-k-1)s^2/\sigma^2$, которая при нормальном распределении ε_i имеет распределение хи-квадрат с $n-k-1$ степенью свободы.

Для проверки значимости коэффициентов уравнения регрессии используем статистику $t_j=\dfrac{\beta_j^*}{\sqrt{s^2\left[(X^TX)^{-1}\right]_{jj}}}$, которая при истинности гипотезы H_0: $\beta_j=0$, имеет распределение Стьюдента с $n-k-1$ степенью свободы. Если для заданного уровня значимости α значение $|t_j|$ больше критического $t_{êðèò}=t_{1-\alpha/2}$, то нулевая гипотеза отвергается и коэффициент признается значимым. В противном случае коэффициент признается незначимым, и соответствующее слагаемое исключается из модели.

В пакете Excel рассчитывается также уровень значимости α статистики $|t_j|$, т.е. вероятность $P\left(x>\left|t_j\right|\right)$. Степень значимости параметров распределения качественно определяется по уровню значимости: не значимые (α ≥ 0,100), слабо значимые (0,100 > α ≥ 0,050), статистически значимые (0,050 > α ≥ 0,010), сильно значимые (0,010 > α ≥ 0,001), высоко значимые (0,001 > α).

Для нахождения доверительных интервалов для коэффициентов β_j используют статистики $\tilde{t}_j=\dfrac{\beta_j^*-\beta_j}{\sqrt{s^2\left[(X^TX)^{-1}\right]_{jj}}}$, имеющие распределение Стьюдента с $n-k-1$ степенью свободы. Для уровня значимости α доверительный интервал рассчитывается по формуле $\beta_j^*\pm t_\alpha\sqrt{s^2\left[(X^TX)^{-1}\right]_{jj}}$, где t_α – квантиль распределение Стьюдента с $n-k-1$ степенью свободы.

Доверительный интервал для условного среднего $\tilde{y}=M(Y\,|\,X=x)$ в многомерной точке $X_0=(1,x_1^0,\dots,x_k^0)^T$ определяется по формуле: $\left[\left(X_0^TB^*\right)\pm t_{1-\alpha/2}s\sqrt{\left(X_0^T\left[(X^TX)^{-1}\right]X_0\right)}\right]$, где t_α – квантиль распределение Стьюдента с $n-k-1$ степенью свободы. Соответственно доверительный интервал для значений y в точке $X_0=(1,x_1^0,\dots,x_k^0)^T$ будет иметь вид: $\left[X_0^TB^*\pm t_{1-\alpha/2}s\left(1+\sqrt{X_0^T\left[(X^TX)^{-1}\right]X_0}\right)\right]$, так как погрешность $y=f(x)+\varepsilon$ будет определяться двумя источниками: погрешностью $(\Delta f)^2=s^2\left(X_0^T\left[(X^TX)^{-1}\right]X_0\right)$, связанной с погрешностями параметров модели, и погрешностью собственно модели $\varepsilon^2=s^2$.

Если выборочная регрессия y удовлетворительно описывает истинную зависимость между y и x, остатки e_i должны быть независимыми нормально распределенными случайными величинами с нулевым средним и в значениях e_i должен отсутствовать тренд. Нормальность распределения остатков e_i может быть установлена одним из критериев согласия (см. раздел 4.3 в лекционном пособии).

Независимость в последовательности значений e_i $(i=1,\dots,n)$ может быть проверена с помощью сериального коэффициента корреляции Дарбина-Ватсона. Статистика сериального коэффициента корреляции Дарбина-Ватсона имеет вид

$$D=\frac{\sum_{i=2}^{n}\left(e_i-e_{i-1}\right)^2}{\sum_{i=1}^{n}e_i^2}. \qquad (6.5)$$

Если $D>D_1(\alpha)$ или $D>4-D_1(\alpha)$, то с достоверностью α принимается гипотеза о наличии соответственно отрицательной или положительной корреляции остатков.
Если $D_2(\alpha)>D>D_1(\alpha)$ или $4-D_1(\alpha)>D>4-D_2(\alpha)$, то критерий не позволяет принять решение по гипотезе о наличии или отсутствии корреляции остатков. Если $D_2(\alpha)<D<4-D_2(\alpha)$, то гипотеза корреляции остатков отклоняется. Критические значения $D_1(\alpha)$ и $D_2(\alpha)$ для различных α и числа k коэффициентов в регрессии приведены в табл. 11 (см. статистические таблицы).

6.2. Пример выполнения задания

Имеется выборка значений совместно наблюдаемых величин X и Y:

X	0,5	1	1,5	2	2,5	3	3,5	4	4,5	5
Y	2,96	0,61	4,63	2,44	2,23	4,89	4,98	3,89	6,74	8,07
X	5,5	6	6,5	7	7,5	8	8,5	9	9,5	10
Y	8,34	9,56	9,30	12,35	11,46	11,09	7,91	8,16	6,54	7,88

Требуется подобрать подходящую модель регрессии, характеризующую зависимость Y от X, если известно, что ошибка $\sigma^2=1,3$.

Нанесем точки (X,Y) на координатную плоскость – построим корреляционное поле, соответствующее нашей выборке (рис. 5.1)

Рис. 6.1. Исходные данные

Видим, что существует зависимость, между значениями X и Y, причем зависимость явно нелинейная. Попробуем аппроксимировать эту зависимость для начала полиномами различных порядков. Возьмем в качестве уравнения регрессии квадратное уравнение:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2$$

Чтобы воспользоваться МНК для оценки коэффициентов, проведем линеаризацию модели, положив $x_1 = x$, $x_2 = x^2$, получим

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

Тогда оценку вектора параметров, согласно МНК, найдем как

$$B^* = \left(X^T X\right)^{-1} X^T Y$$

Здесь X - матрица, первый столбец которой содержит единицы, а второй и последующий значения x_1 и x_2.

Для облегчения подбора модели можно воспользоваться встроенными функциями пакета EXCEL (для выбранной модели все равно потом потребуется провести все вычисления вручную, чтобы построить доверительные интервалы). В пакете анализа необходимо выбрать функцию “регрессия”, задать столбец значений Y и матрицу, соответствующую X (единичный столбец в этом случае задавать не надо). Если выбрать вывод остатков, то помимо регрессионной статистики, будут выведены и предсказанные значения Y, т.е.

$$y^* = \beta_0^* + \beta_1^* x + \beta_2^* x^2$$

Для нашей модели регрессионная статистика, полученная пакетом Fxcel будет иметь следующий вид (приведена лишь часть):

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	0,852379622
R-квадрат	0,72655102
Нормированный R-квадрат	0,694380552
Стандартная ошибка	1,820336831
Наблюдения	20

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*
Y-пересечение	-0,963028418	1,354297293	-0,711090853	0,486670901
Переменная X 1	2,604940094	0,594036759	4,385149663	0,000403854
Переменная X 2	-0,167559332	0,054953956	-3,049085893	0,007253372

Здесь в первой таблице:

- Множественный R – корень квадратный из коэффициента детерминации;
- R-квадрат – коэффициент детерминации;
- Нормированный R-квадрат – это скорректированная величина коэффициента детерминации, вычисляемая по формуле $R^2 = 1 - (1 - R^2)\dfrac{n-1}{n-k-1}$;
- Стандартная ошибка – значение $s = \sqrt{s^2}$, где s^2 - оценка дисперсии предсказания σ^2;
- Наблюдения – объем выборки

Во второй таблице:

- Коэффициенты – значения оценок коэффициентов $\beta_0^*, \beta_1^*, \beta_2^*$;
- Стандартная ошибка – значения оценок среднеквадратичных отклонений оценок коэффициентов $\beta_0^*, \beta_1^*, \beta_2^*$;
- t-статистика – наблюдаемые значения статистик критерия проверки значимости коэффициентов для соответственно коэффициентов $\beta_0, \beta_1, \beta_2$;

- Р-значения – достигнутые значения уровня значимости $P\left(x>\left|t_j\right|\right)$.

Соответствующий график предсказанных значений в сравнении с исходными данными имеет вид (рис. 6.2):

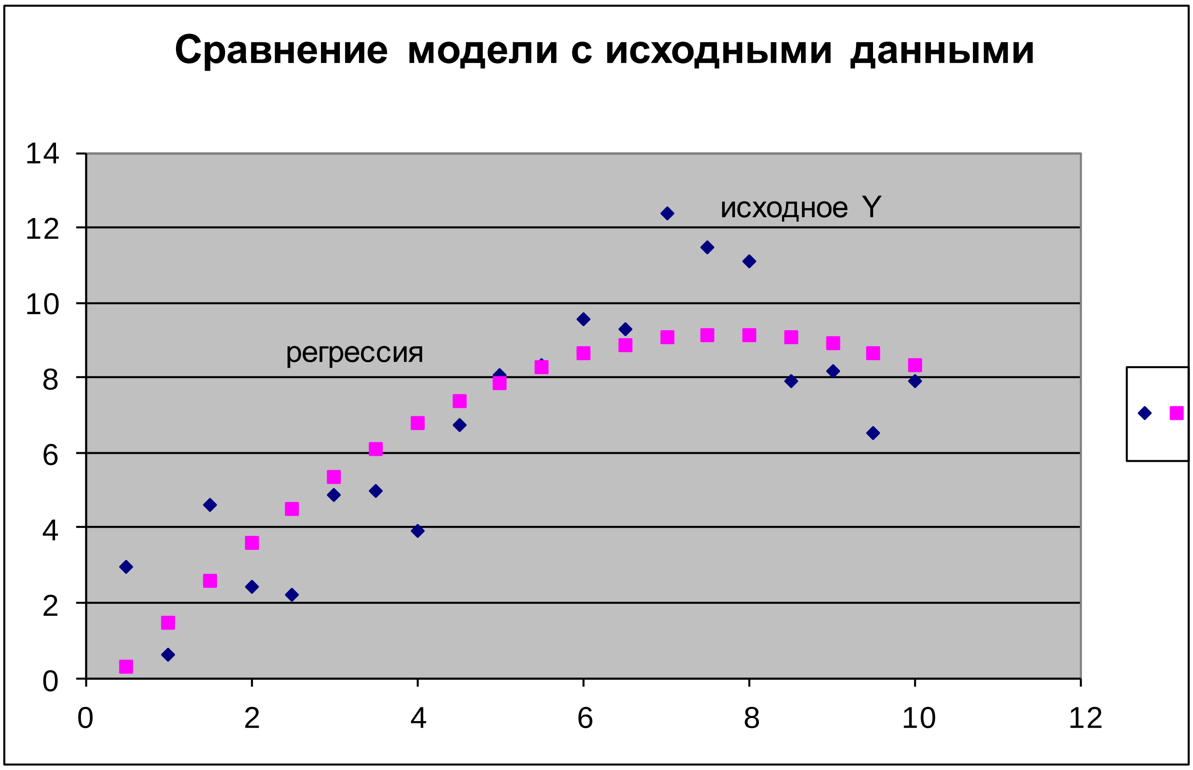


Рис. 6.2. Сравнение модели $y=\beta_0+\beta_1 x+\beta_2 x^2$ с исходными данными

Отметим, что полученная оценка значения σ велика: $s=1{,}82$. Что касается коэффициентов модели, то кроме β_0 все они значимо отличаются от нуля (достигнутый уровень значимости достаточно мал, поэтому можно отвергнуть гипотезу о равенстве коэффициентов нулю). Попробуем улучшить модель, увеличим порядок полинома, пусть

$$y=\beta_0+\beta_1 x+\beta_2 x^2+\beta_3 x^3$$

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	0,925507816
R-квадрат	0,856564717
Нормированный R-квадрат	0,829670602
Стандартная ошибка	1,358958106
Наблюдения	20

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*Р-Значение*
Y-пересечение	3,176148907	1,484433212	2,13963746	0,048142193
Переменная X 1	-1,619155418	1,194561911	-1,355438679	0,194104204
Переменная X 2	0,814063749	0,261005991	3,118946605	0,006612096
Переменная X 3	-0,062325275	0,016365816	-3,808259635	0,001545564

Проводим линеаризацию, полагая $x_1 = x$, $x_2 = x^2$, $x_3 = x^3$, и оцениваем коэффициенты новой модели.

График показан на рис. 6.3

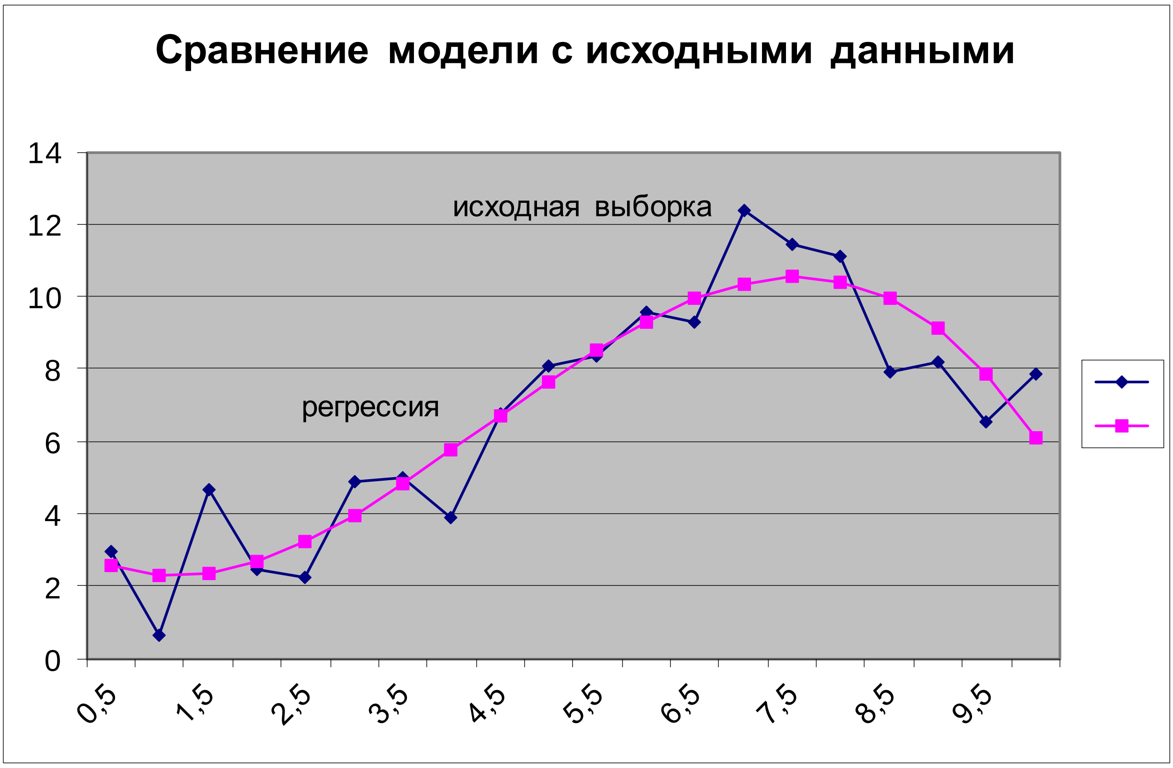


Рис. 6.3. Сравнение модели $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$ с исходными данными

Заметим, что коэффициент детерминации увеличился, а оценка σ (стандартная ошибка) уменьшилась, что говорит о лучшем качестве модели по сравнению с предыдущей. Причем значение этой оценки близко к значениям $\sigma = 1{,}3$, указанному в задании. Из коэффициентов можно считать, что β_1 не значимо отличается от нуля (достигнутый уровень значимости $\alpha = 0{,}194$, говорит о том, что при истинности гипотезы $H_0 : \beta_1 = 0$, такое или большее значение t-статистики критерия могло наблюдаться с вероятностью 0,194). Поэтому, можно положить $\beta_1 = 0$ и модель соответственно примет вид:

$$y = \beta_0 + \beta_2 x^2 + \beta_3 x^3$$

Результаты регрессионного анализа для этой модели:

Регрессионная статистика	
Множественный R	0,916566765
R-квадрат	0,840094635
Нормированный R-квадрат	0,82128224
Стандартная ошибка	1,39201886
Наблюдения	20

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*
Y-пересечение	1,356030259	0,648133062	2,092209668	0,051733578
Переменная X 1	0,469599546	0,060941974	7,70568321	6,05722E-07
Переменная X 2	-0,041727702	0,006223514	-6,70484584	3,6971E-06

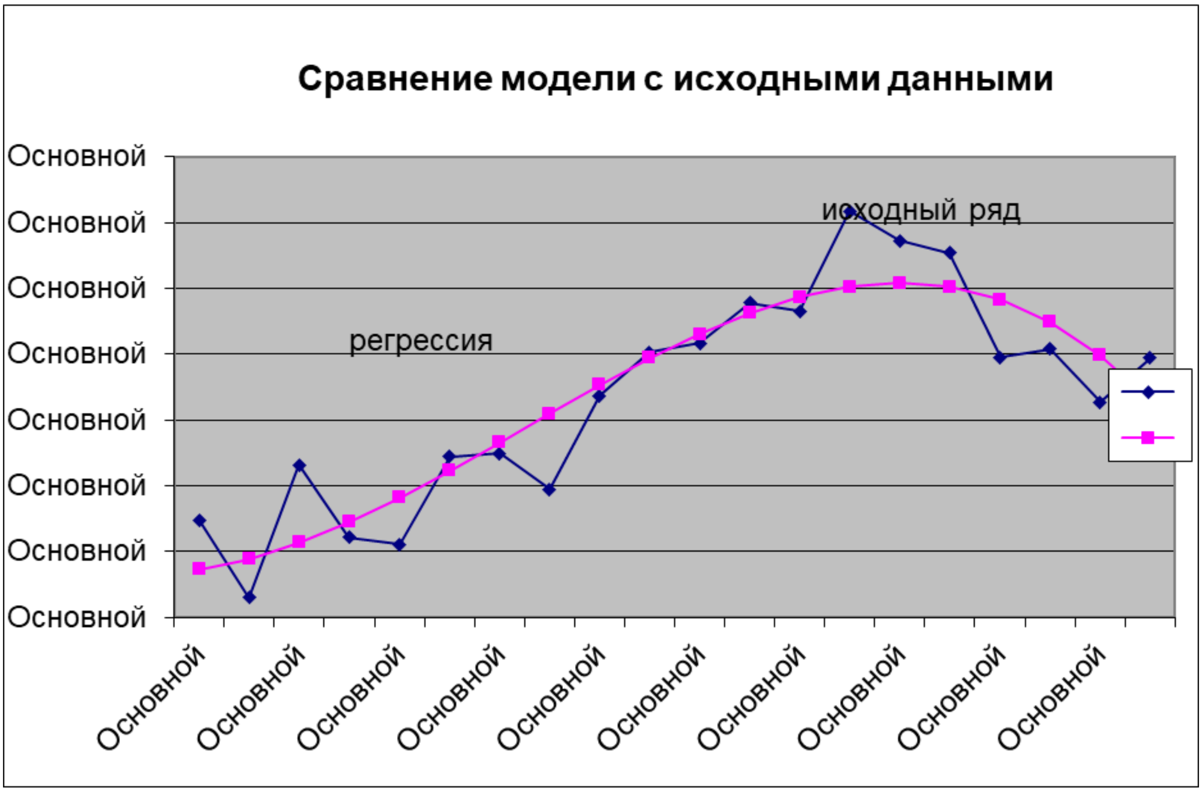


Рис. 6.4. Сравнение модели $y = \beta_0 + \beta_2 x^2 + \beta_3 x^3$ с исходными данными

И хотя, параметры модели немного ухудшились (сравните R-квадрат и стандартную ошибку!), тем не менее все коэффициенты получились значимыми, поэтому данная модель, предпочтительнее предыдущей.

Можно ли повысить еще качество модели? В классе полиномов это сделать не удастся. Повышение порядка полинома (можно проверить!), уже больше не понижает стандартную ошибку. Следовательно, улучшение нужно искать, используя иные классы функций. Например, можно предположить, что существует периодическая зависимость значений Y от X, тогда надо добавить в модель гармонические составляющие вида

$\beta_1 cos(\omega x) + \beta_2 sin(\omega x)$. Частоты этих составляющих ω придется подбирать отдельно. Можно предположить, анализируя зависимость, что у нас присутствует периодическая составляющая с частотой порядка 1 (одно колебание за весь интервал изменений X). Построим модель вида

$$y = \beta_0 + \beta_1 x + \beta_2 cos(2\pi x/10) + \beta_3 sin(2\pi x/10).$$

Получим оценки для этой модели:

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	0,934315451
R-квадрат	0,872945362
Нормированный R-квадрат	0,849122618
Стандартная ошибка	1,279008211
Наблюдения	20

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*
Y-пересечение	5,080728355	0,88616922	5,733361352	3,08053E-05
Переменная X 1	0,308759559	0,159762044	1,932621493	0,0711836
Переменная X 2	-1,299656278	0,412270758	-3,152433814	0,006163641
Переменная X 3	-3,032825609	0,646493647	-4,691191664	0,000245237

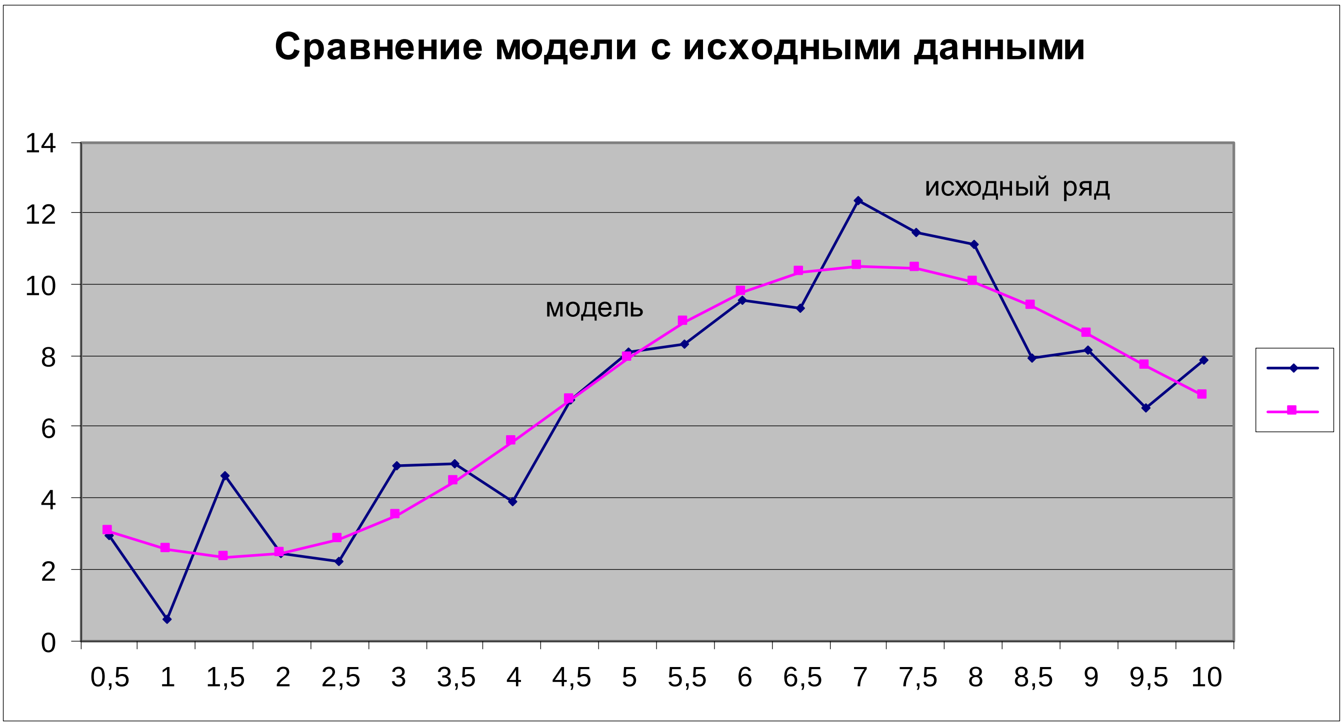


Рис. 6.5. Сравнение модели $y = \beta_0 + \beta_1 x + \beta_2 cos(2\pi x/10) + \beta_3 sin(2\pi x/10)$ с исходными данными

Заметим, что мы получили значение стандартной ошибки меньше заданой величины с. Это говорит о том, что дальше улучшать модель бессмысленно. Если мы продолжим, то будем по сути аппроксимировать случайные ошибки, а не реальную существующую зависимость Y от X. Качество модели выше, чем у модели вида $y = \beta_0 + \beta_2 x^2 + \beta_3 x^3$. Однако, данная модель содержит на один коэффициент больше, и кроме того содержит параметр, значение которого по сути определяется вручную. Поскольку модели принципиально различные, то предпочесть следует ту, которая более соответствует физике явления (если она известна). Мы остановимся на последней модели $y = \beta_0 + \beta_1 x + \beta_2 \mathrm{Cos}(2\pi x/10) + \beta_3 \mathrm{Cos}(2\pi x/10)$.

Найдем доверительный интервал для σ^2 соответствующий уровню 0,95. Находим квантили распределения хи-квадрат уровней 0,025 и 0,975 соответственно для числа степеней свободы $\nu = n - k - 1 = 20 - 3 - 1 = 16$ (k – число коэффициентов модели, не считая β_0): $\tau_{0,025} = 6{,}91$, $\tau_{0,975} = 28{,}85$. Тогда $\frac{(n-k-1)s^2}{\tau_{0,975}} < \sigma^2 < \frac{(n-k-1)s^2}{\tau_{0,025}}$ $\Rightarrow$ $0{,}91 < \sigma^2 < 3{,}79 \quad \Rightarrow 0{,}95 < \sigma < 1{,}95$.

Доверительные интервалы для коэффициентов уравнения регрессии можно найти в Итоговой статистике.

Доверительный интервал для условного среднего $\tilde{y} = M(Y \mid X = x)$ для тех же значений X, что приведены в выборке, найдем по формуле $\left[\left(X^T B^* \right)_j \pm t_\alpha s \sqrt{\left(X (X^T X)^{-1} X^T \right)_{jj}} \right]$ (см. рис 5.6), где t_α квантиль распределение Стьюдента с $n-k-1$ степенью свободы (доверительный уровень возьмем 0,67, тогда $\alpha = 0{,}33$ и $t_\alpha = 1{,}0047$).

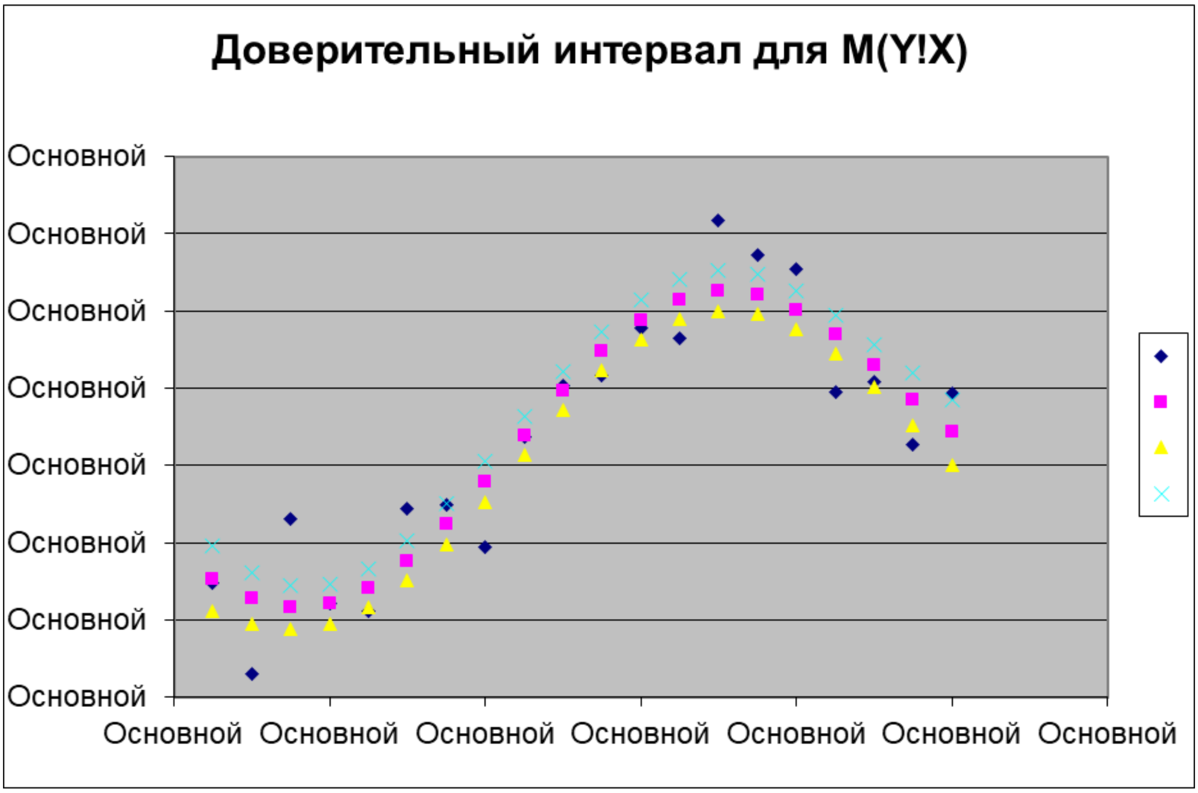


Рис. 6.6. Доверительный интервал для M(Y|X)

Доверительный интервал для значений y рассчитываем по формуле $y_j = \left[\left(X^T B^* \right)_j \pm t_\alpha s \left(1 + \sqrt{\left(X (X^T X)^{-1} X^T \right)_{jj}} \right) \right]$. Доверительный интервал для значений y, полученный по этим формулам, отображен на рис 5.7.

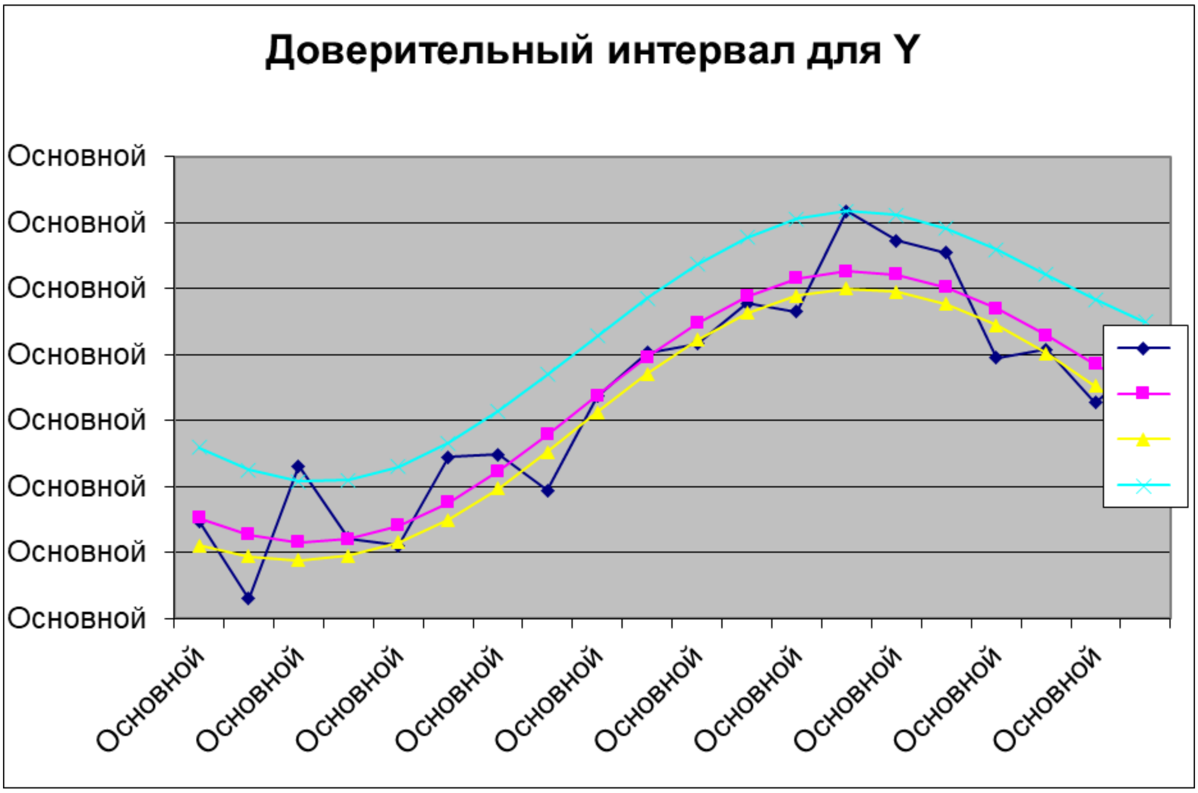


Рис. 6.7. Доверительный интервал для Y

Выполним анализ остатков модели.

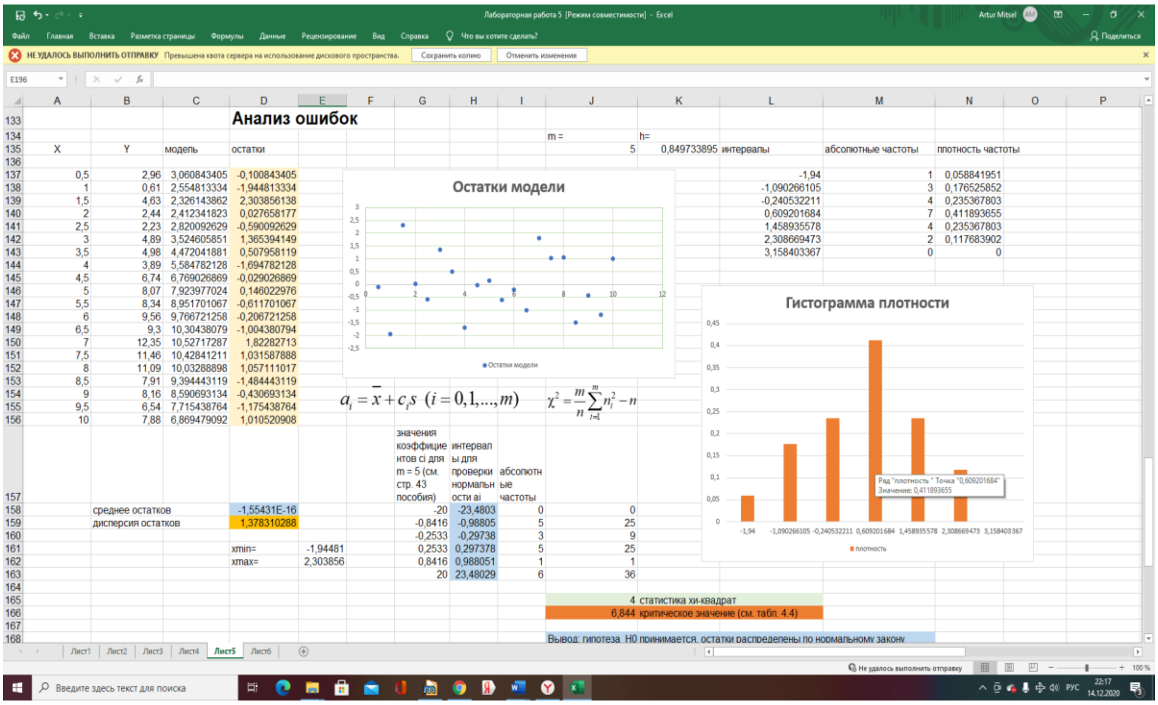


Рис. 6.8. Анализ остатков

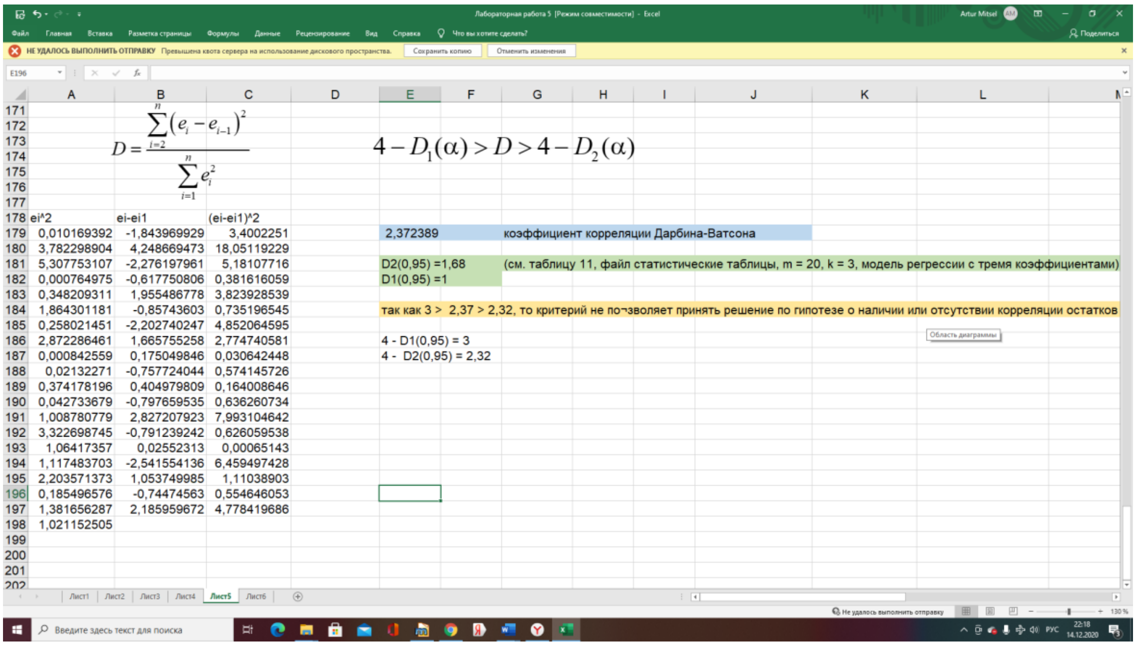


Рис. 6.9. Проверка корреляции остатков

6.3. Задание по работе №6

Имеется выборка значений совместно наблюдаемых величин X и Y (приложение).

Требуется:

1) Отобразить графически поле наблюдаемых значений величин X и Y.

2) Подобрать 2-3 аппроксимирующих зависимости для уравнения регрессии, провести линеаризацию моделей.

Методом наименьших квадратов найти оценки параметров каждой модели. Оценить для каждой модели остаточную дисперсию (дисперсию предсказания). Сравнить графически каждую кривую регрессии с наблюдаемыми значениями Y и выбрать модель регрессии. Желательно, чтобы для выбранной модели оценка дисперсии s^2 была в пределах указанной в задании величины σ^2 (Величина σ^2 в расчетах нигде не используется и указана только лишь для сопоставления с s^2!!!).

3) Для выбранной модели найти оценки дисперсий коэффициентов уравнения регрессии и проверить значимость коэффициентов уравнения регрессии в предположении нормальности распределения ошибок (доверительный уровень принять равным 0,95).

4) Определить доверительные интервалы для остаточной дисперсии и коэффициентов уравнения регрессии.

5) Определить границы доверительного интервала для средних значений $\tilde{y} = M(Y \mid X = x)$ и значений y при каждом наблюдаемом значении x (отобразить графически).

6) Выполнить анализ остатков модели.

Приложение 6.1

Варианты заданий по лабораторной работе №6

Вариант 1.

X	1	2	3	4	5	6	7	8	9	10
Y	0,53	-0,64	0,21	0,38	-1,85	0,69	5,11	4,12	6,09	4,01
X	11	12	13	14	15	16	17	18	19	20
Y	2,85	8,70	6,28	4,31	4,76	-0,82	3,07	-6,32	-6,74	-3,15

$\sigma = 2$

Вариант 2.

X	1	2	3	4	5	6	7	8	9	10
Y	7,81	6,36	9,07	8,12	7,01	5,28	5,23	7,42	5,72	-2,11
X	11	12	13	14	15	16	17	18	19	20
Y	-3,57	0,06	-4,96	-3,88	-3,04	-11,18	-4,93	0,58	3,99	5,55

$\sigma = 2{,}2$

Вариант 3.

X	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
Y	0,47	0,38	0,56	0,70	0,75	0,86	0,53	0,80	0,99	0,84
X	1	1,1	1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9
Y	0,93	0,88	0,89	0,99	1,01	0,85	0,96	0,89	0,84	0,64

$\sigma = 0{,}1$

Вариант 4.

X	1	2	3	4	5	6	7	8	9	10
Y	5,98	7,43	0,98	-0,75	0,34	-2,84	1,79	6,04	5,09	8,78
X	11	12	13	14	15	16	17	18	19	20
Y	10,39	6,73	4,39	6,53	10,08	12,51	8,75	21,91	26,93	27,73

$\sigma = 2$

Вариант 5.

X	0,2	0,4	0,6	0,8	1	1,2	1,4	1,6	1,8	2
Y	2,21	1,55	2,53	3,21	3,01	3,23	3,93	2,91	2,99	4,33
X	2,2	2,4	2,6	2,8	3	3,2	3,4	3,6	3,8	4
Y	3,91	3,70	4,16	3,68	3,73	3,92	2,94	2,91	2,29	1,09

$\sigma = 0{,}5$

Вариант 6.

X	0,5	1	1,5	2	2,5	3	3,5	4	4,5	5
Y	7,49	4,79	4,72	3,99	4,65	4,70	5,85	4,38	5,30	4,16

X	5,5	6	6,5	7	7,5	8	8,5	9	9,5	10
Y	5,37	7,75	8,63	12,04	11,12	7,24	9,22	15,11	15,54	17,70

$\sigma = 1$

Вариант 7.

X	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
Y	1,78	1,65	0,96	1,04	1,26	1,09	1,35	0,95	0,93	0,77
X	1,1	1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9	2
Y	0,86	0,82	1,22	1,25	1,29	1,50	1,35	1,52	1,97	2,40

$\sigma = 0{,}15$

Вариант 8.

X	1	2	3	4	5	6	7	8	9	10
Y	7,37	10,35	9,69	10,25	12,23	14,91	15,06	14,27	11,89	13,31
X	11	12	13	14	15	16	17	18	19	20
Y	13,51	12,68	12,83	14,38	13,33	12,34	11,86	14,71	13,56	12,52

$\sigma = 1{,}3$

Вариант 9.

X	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
Y	-8,74	-5,33	-6,17	-5,17	-2,22	1,73	2,43	1,94	-0,61	1,82
X	1,1	1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9	2
Y	2,69	2,23	3,07	5,73	5,01	4,37	4,41	8,80	7,97	7,29

$\sigma = 1{,}7$

Вариант 10.

X	1	2	3	4	5	6	7	8	9	10
Y	2,26	3,82	3,50	3,47	3,90	4,56	4,24	3,56	2,29	2,51
X	11	12	13	14	15	16	17	18	19	20
Y	2,32	1,77	1,67	2,16	1,74	1,42	1,39	2,73	2,63	2,69

$\sigma = 0{,}5$