Министерство науки и высшего образования Российской Федерации

Томский государственный университет систем управления и радиоэлектроники

Н.А. Ярушкина
О.В. Килина

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Методические указания для выполнения лабораторных и самостоятельных работ студентами очной формы обучения технических направлений подготовки бакалавриата

Томск
2022

УДК 519.2
ББК 22.17
Я-78

Рецензент:
Антипин М.Е., доцент кафедры управления инновациями, канд. физ.-мат. наук.

Ярушкина, Наталья Анатольевна
Килина, Ольга Владимировна
Я-78 Теория вероятностей и математическая статистика: методические указания для выполнения лабораторных и самостоятельных работ студентами очной формы обучения технических направлений подготовки бакалавриата / Н.А. Ярушкина, О.В. Килина – Томск: Томск. гос. ун-т систем упр. и радиоэлектроники, 2022. – 44 с.

Настоящие методические указания по дисциплине «Теория вероятностей и математическая статистика» содержат описание основных теоретических положений по курсу, примеры решения задач и их выполнения в ППП Excel, задачи для самостоятельного решения обучающимися.

Одобрено на заседании кафедры управления инновациями, протокол № 5 от 28.12.2022.

УДК 519.2
ББК 22.17

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ

Целью изучения дисциплины «Теория вероятностей и математическая статистика» является формирование у студентов математической культуры, позволяющей строить и анализировать модели технических и др. систем с помощью вероятностно-статистических методов.

В числе задач дисциплины: освоение основного понятийного аппарата теории вероятностей и математической статистики; развитие логического мышления и умения оперировать с конкретными выборками, привитие навыков корректного употребления вероятностных и статистических рассуждений; овладение способами решения вероятностных задач, освоение основных моделей и соответствующих программных средств обработки статистического материала; овладение знаниями в области вероятностных расчетов и анализа данных, необходимыми в практической и учебной деятельности.

По окончанию освоения дисциплины «Теория вероятностей и математическая статистика» обучающиеся должны демонстрировать следующие результаты обучения:

1) знать: основы теории вероятностей, характеристики случайных величин и законы их распределения, основы математической статистики;
2) уметь: анализировать случайные события, вычислять их вероятности, доверительные интервалы, решать стандартные задачи математической статистики;
3) владеть: навыками проверки статистических гипотез, способностью вычислять и анализировать числовые характеристики случайных величин, проводить математическое моделирование и теоретические исследования.

Вероятностно-статистический подход для обработки и интерпретации экспериментальных данных широко используется на всех этапах работы с различной информацией. Это обуславливается тем, что любые данные, полученные экспериментальным путем, являются случайными. Кроме того, из-за наложения помех, связанных с погрешностью приборов, различными неоднородностями, неучтенными вариациями физических объектов и ряда других причин, объект исследования реализуется случайным образом. Следовательно, если на практике исследователь имеет дело с данными, которые с большим основанием оцениваются случайными величинами и процессами, то для выделения полезной информации он обязательно должен использоваться вероятностно-статистический подход. В силу этого в настоящих методических указаниях основной упор делается на математические методы построения вероятностных моделей и реализацию этих методов на реальных задачах профессиональной деятельности посредством использования ППП Excel.

Материал методических указаний содержит семь лабораторных работ, каждая из которых включает: краткий теоретический материал; методические рекомендации по выполнению заданий; указания для выполнения лабораторных работ в пакете прикладных программ Excel и задания для самостоятельного выполнения.

1 ЛАБОРАТОРНАЯ РАБОТА
«ПРАКТИКА РАСЧЕТА ВЕРОЯТНОСТЕЙ СЛОЖНЫХ СОБЫТИЙ»

Цель работы: приобретение инструментального навыка расчета вероятности случайных событий посредством использования функций ППП Excel.

1.1 Основные теоретические положения

Вероятностью (классическое определение) события *A* называется отношение числа *m* элементарных исходов, благоприятствующих событию *A*, к общему числу исходов *n*

$$P(A)=\frac{m}{n}.$$

Правило сложения вероятностей

$$P(A+B)=P(A)+P(B)-P(AB).$$

Если события и несовместны, то $P(AB)=0$, следовательно

$$P(A+B)=P(A)+P(B).$$

Вероятность наступления события *A* при условии, что событие *B* произошло, называется ***условной вероятностью***

$$P(A/B)=\frac{P(AB)}{P(B)}.$$

Формула умножения вероятностей

$$P(AB)=P(A)\cdot P(B/A)=P(B)\cdot P(A/B).$$

Формула полной вероятности

$$P(A)=P(H_1)\cdot P(A/H_1)+P(H_2)\cdot P(A/H_2)+\ldots+P(H_n)\cdot P(A/H_n)=\sum_{i=1}^{n}P(H_i)\cdot P(A/H_i),$$

в предположении, что событие *A* может произойти с одним из событий $H_1,H_2,\ldots,H_n$, образующих полную группу несовместных событий (гипотез), а вероятности гипотез $P(H_1),P(H_2),\ldots,P(H_n)$ и условные вероятности $P(A/H_1),P(A/H_2),\ldots,P(A/H_n)$ известны.

Формула Байеса

$$P(H_i/A)=\frac{P(H_i)P(A/H_i)}{\sum\limits_{k=1}^{n}P(H_k)P(A/H_k)},\ i=1,2,\ldots,n,$$

в предположении, что в результате проведения эксперимента событие *A* наступило.

Формула Бернулли (вероятность того, что событие *A* наступит ровно *m* раз из *n*)

$$P_n(m)=C_n^m p^m q^{n-m},$$

в предположении, что проводятся *n* одинаковых независимых испытаний с двумя исходами *A* или $\bar{A}$, а вероятности $P(A)=p$ и $P(\bar{A})=1-p=q$ постоянны и не равны ни нулю, ни единице.

1.2 Примеры выполнения заданий

Пример 1. В партии из 10 деталей две бракованные. Найти вероятность того, что среди выбранных случайным образом четырех деталей окажется одна бракованная.

Решение:

Пространство элементарных исходов представляет собой в этом случае множество всевозможных упорядоченных наборов из четырех любых деталей. Общее число таких элементарных исходов равно $n = C_{10}^{4} = \dfrac{10!}{4!6!} = 210$. Пусть событие A состоит в том, что в выборку попадут три годных детали и одна бракованная. Три годные детали из восьми можно взять C_8^3 способами. Следовательно, число благоприятствующих исходов равно $m = C_8^3 C_2^1 = 112$. Искомая вероятность равна отношению числа исходов, благоприятствующих событию, к числу всех элементарных исходов $P(A) = \dfrac{m}{n} = \dfrac{112}{210} = \dfrac{8}{15} = 0{,}533$.

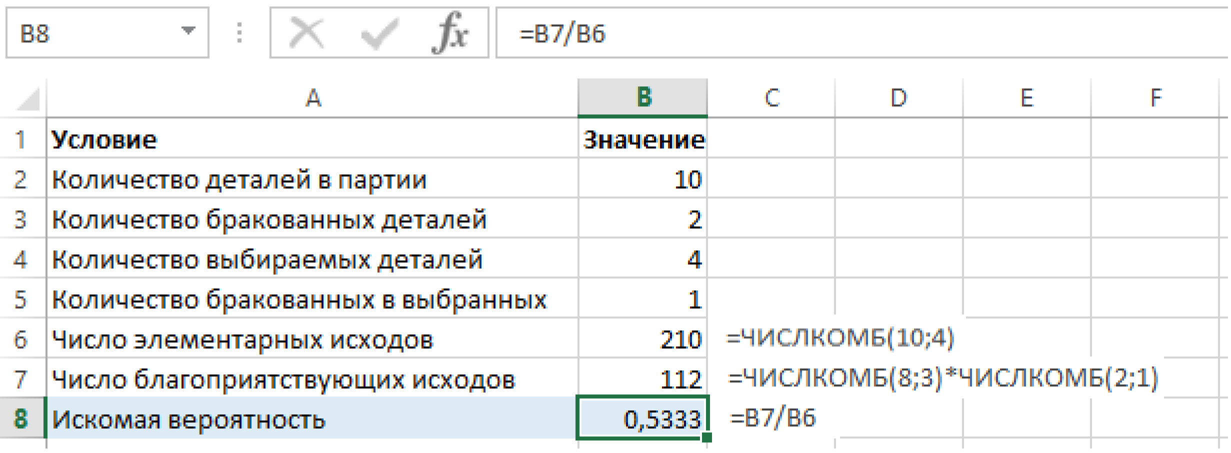
Для нахождения числа сочетаний (общее число элементарных исходов и число благоприятствующих исходов) в Excel используем функцию ЧИСЛКОМБ, возвращающую количество комбинаций для заданного числа элементов. Функция ЧИСЛКОМБ используется для определения общего числа всех групп, которые можно составить из элементов данного множества.

Синтаксис функции: ЧИСЛКОМБ(число;число_выбранных)

Аргументы функции ЧИСЛКОМБ:

– Число – обязательный аргумент. Количество элементов.

– Число_выбранных – обязательный аргумент. Количество элементов в каждой комбинации.



B8 =B7/B6

	A	B	C
1	**Условие**	**Значение**	
2	Количество деталей в партии	10	
3	Количество бракованных деталей	2	
4	Количество выбираемых деталей	4	
5	Количество бракованных в выбранных	1	
6	Число элементарных исходов	210	=ЧИСЛКОМБ(10;4)
7	Число благоприятствующих исходов	112	=ЧИСЛКОМБ(8;3)*ЧИСЛКОМБ(2;1)
8	Искомая вероятность	0,5333	=B7/B6

Рисунок 1 – Вычисление вероятности события *A* (пример 1)

Пример 2. По каналу связи передаются три сообщения, каждое из которых может быть передано верно или частично искажено. Вероятность того, что сообщение передано верно – 0,8. Считая, что сообщение искажается или передается верно независимо от количества передач и от результата предыдущей связи, найти вероятности следующих событий:

1) $B = \{\text{все три сообщения переданы верно}\}$;

2) $C = \{\text{ровно одно из трех сообщений искажено}\}$;

3) $D = \{\text{хотя бы одно из трех сообщений искажено}\}$.

Решение:

Обозначим через A_i событие, состоящее в том, что i-ое сообщение передано верно. Событие $B = A_1 A_2 A_3$. Применяя формулу умножения вероятностей для независимых событий и учитывая, что $P(A_i) = 0,8$, вычислим

$$P(B) = P(A_1 A_2 A_3) = P(A_1) P(A_2) P(A_3) = 0,8^3 = 0,512.$$

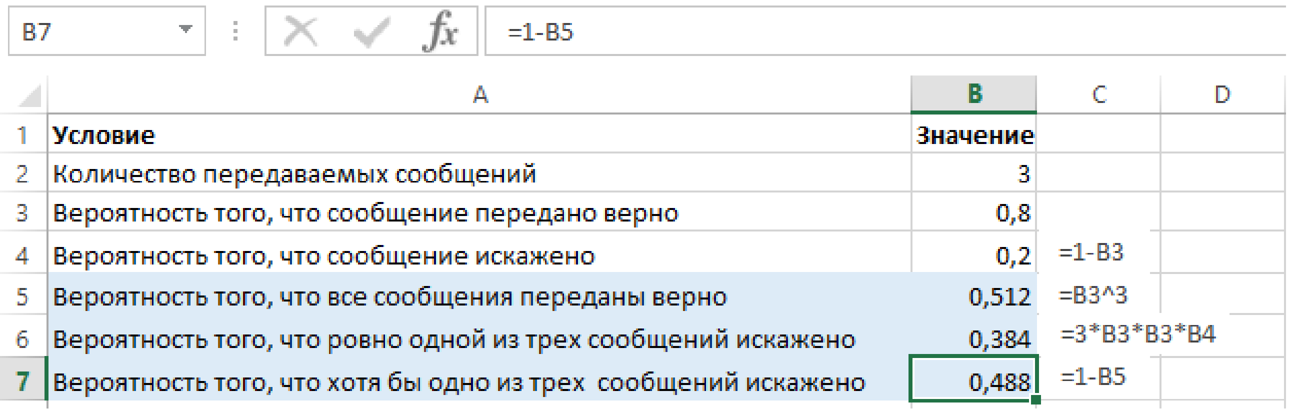
Событие C можно выразить через события A_1, A_2 и A_3 следующим образом: $C = \overline{A}_1 A_2 A_3 + A_1 \overline{A}_2 A_3 + A_1 A_2 \overline{A}_3$. Применяя формулу сложения вероятностей несовместных событий и формулу умножения вероятностей, найдем вероятность этого события:

$$P(C) = P(\overline{A}_1 A_2 A_3 + A_1 \overline{A}_2 A_3 + A_1 A_2 \overline{A}_3) = P(\overline{A}_1 A_2 A_3) + P(A_1 \overline{A}_2 A_3) + P(A_1 A_2 \overline{A}_3) =$$

$$= (1 - 0,8) \cdot 0,8 \cdot 0,8 + 0,8 \cdot (1 - 0,8) \cdot 0,8 + 0,8 \cdot 0,8 \cdot (1 - 0,8) = 0,384.$$

Событие $D = \overline{A}_1 + \overline{A}_2 + \overline{A}_3$. Теорему сложения для несовместных событий применить нельзя, так как события A_1, A_2 и A_3 совместны. Вероятность события D удобно вычислять через вероятность противоположного события $\overline{D} = B = A_1 A_2 A_3$. Вычислим

$$P(D) = 1 - P(B) = 1 - 0,512 = 0,488.$$



B7 =1-B5

	A	B	C	D
1	**Условие**	**Значение**		
2	Количество передаваемых сообщений	3		
3	Вероятность того, что сообщение передано верно	0,8		
4	Вероятность того, что сообщение искажено	0,2	=1-B3	
5	Вероятность того, что все сообщения переданы верно	0,512	=B3^3	
6	Вероятность того, что ровно одной из трех сообщений искажено	0,384	=3*B3*B3*B4	
7	Вероятность того, что хотя бы одно из трех сообщений искажено	0,488	=1-B5	

Рисунок 2 – Вычисление вероятности событий (пример 2)

Пример 3. В ящике содержится 12 деталей, изготовленных фирмой №1, 20 деталей – фирмой №2 и 18 деталей – фирмой №3. Вероятность того, что деталь, изготовленная фирмой №1, отличного качества, равна 0,9; для деталей, изготовленных фирмами №2 и №3, эти вероятности соответственно равны 0,6 и 0,9. Найти вероятность того, что извлеченная наудачу деталь окажется отличного качества.

Решение:

Введем группу гипотез:

1) $H_1 = \{\text{деталь изготовлена фирмой №1}\}$;

2) $H_2 = \{\text{деталь изготовлена фирмой №2}\}$;

3) $H_3 = \{\text{деталь изготовлена фирмой №3}\}$.

Найдем вероятности гипотез по классическому определению вероятности:

1) $P(H_1) = \dfrac{12}{12 + 20 + 18} = \dfrac{12}{50} = 0,24$;

2) $P(H_2) = \dfrac{20}{12 + 20 + 18} = \dfrac{20}{50} = 0,40$;

3) $P(H_3)=\dfrac{18}{12+20+18}=\dfrac{18}{50}=0,36$.

Введем событие $A=\{\text{деталь отличного качества}\}$.

Условные вероятности известны по условию задачи: $P(A/H_1)=0,9$, $P(A/H_2)=0,6$, $P(A/H_3)=0,9$.

Найдем вероятность события A по формуле полной вероятности:

$$P(A)=P(H_1)\cdot P(A/H_1)+P(H_2)\cdot P(A/H_2)+P(H_3)\cdot P(A/H_3)=$$
$$=0,24\cdot 0,9+0,4\cdot 0,6+0,36\cdot 0,9=0,78.$$

B11 =СУММПРОИЗВ(B5:B7;B8:B10)

	A	B	C	D	E	F
1	**Условие**	**Значение**				
2	Количество деталей, изготовленных фирмой №1	12				
3	Количество деталей, изготовленных фирмой №2	20				
4	Количество деталей, изготовленных фирмой №3	18				
5	Вероятность того, что детали изготовлены фирмой №1	0,9				
6	Вероятность того, что детали изготовлены фирмой №2	0,6				
7	Вероятность того, что детали изготовлены фирмой №3	0,9				
8	Вероятность гипотезы 1	0,24	=B2/СУММ(B2:B4)			
9	Вероятность гипотезы 2	0,4	=B3/СУММ(B2:B4)			
10	Вероятность гипотезы 3	0,36	=B4/СУММ(B2:B4)			
11	Искомая вероятность	0,78	=СУММПРОИЗВ(B5:B7;B8:B10)			

Рисунок 3 – Вычисление вероятности события (пример 3)

Пример 4. Прибор может собираться из высококачественных деталей и из деталей обычного качества. 40% приборов собирается из высококачественных деталей и их надежность за время *t* равна 95%. Приборы из обычных деталей за время *t* имеют надежность 0,7. Прибор испытан и за время *t* работал безотказно. Какова вероятность того, что он собран из высококачественных деталей?

Решение:

Введем группу гипотез:

1) $H_1=\{\text{прибор собран из высококачественных деталей}\}$;

2) $H_2=\{\text{прибор собран из обычных деталей}\}$;

До проведения опыта вероятности этих гипотез равны: $P(H_1)=0,4$, $P(H_2)=0,6$.

В ходе испытания наблюдалось событие

$$A=\{\text{прибор работал безотказно в течении времени } t\}.$$

Условные вероятности этого события при гипотезах H_1 и H_2 соответственно равны $P(A/H_1)=0,95$, $P(A/H_2)=0,7$.

Теперь переходим непосредственно к формуле Байеса и находим условную вероятность первой гипотезы – прибор собран из высококачественных деталей:

$$P(H_1/A)=\frac{P(H_1)P(A/H_1)}{P(H_1)P(A/H_1)+P(H_2)P(A/H_2)}=\frac{0,4\cdot 0,95}{0,4\cdot 0,95+0,6\cdot 0,7}=0,475.$$

B6 =B2*B4/(B2*B4+B3*B5)

	A	B
1	**Условие**	**Значение**
2	Вероятность гипотезы 1	0,4
3	Вероятность гипотезы 2	0,6
4	Условная вероятность события при гипотезе 1	0,95
5	Условная вероятность события при гипотезе 2	0,7
6	Искомая вероятность	0,475

Рисунок 4 – Вычисление вероятности события (пример 4)

Пример 5. Из 100 аккумуляторов за год хранения 8 выходит из строя. Наудачу выбирают 5 аккумуляторов. Определить вероятность того, что среди них 2 неисправных.

Решение:

Имеем схему Бернулли с параметрами $p=\frac{8}{100}=0,08$ (вероятность того, что аккумулятор выйдет из строя), $n=5$ (число испытаний), $m=2$ (число «успехов», неисправных аккумуляторов).

Получаем вероятность того, что из 5 аккумуляторов ровно 2 неисправных по формуле Бернулли:

$$P_5(2)=C_5^2(0,08)^2(1-0,08)^{5-2}=\frac{5!}{2!3!}(0,08)^2(1-0,08)^{5-2}=0,0498\,.$$

Для нахождения искомой вероятности в Excel может быть использована функция БИНОМ.РАСП, возвращающая отдельное значение биномиального распределения. Используется функция БИНОМ.РАСП в задачах с фиксированным числом тестов или испытаний, если результатом любого испытания может быть только успех или неудача, испытания независимы, а вероятность успеха остается постоянной в течение всего эксперимента.

Синтаксис функции:

БИНОМ.РАСП(число_успехов;число_испытаний;вероятность_успеха;интегральная)

Аргументы функции БИНОМ.РАСП:

– Число_успехов – обязательный аргумент. Количество успешных испытаний.

– Число_испытаний – обязательный аргумент. Количество независимых испытаний.

– Вероятность_успеха – обязательный аргумент. Вероятность успеха каждого испытания.

– Интегральная – обязательный аргумент. Логическое значение, определяющее форму функции. Если имеется истина, то БИНОМ. Функция РАСП возвращает накопительную функцию распределения, то есть вероятность того, что успешных результатов не будет.

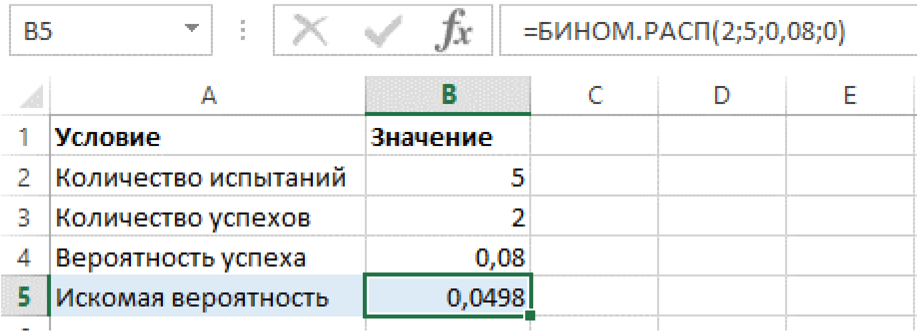


Рисунок 5 – Вычисление вероятности по схеме Бернулли (пример 5)

1.3 Задания для самостоятельного выполнения

Задание 1. В партии из 20 деталей пять бракованных. Найти вероятность того, что среди выбранных случайным образом шести деталей окажется две бракованных.

Задание 2. По каналу связи передаются пять сообщений, каждое из которых может быть передано верно или частично искажено. Вероятность того, что сообщение передано верно – 0,85. Считая, что сообщение искажается или передается верно независимо от количества передач и от результата предыдущей связи, найти вероятности следующих событий:

1) $A = \{\text{все пять сообщений переданы верно}\}$;
2) $B = \{\text{ровно одно из пяти сообщений искажено}\}$;
3) $C = \{\text{ровно два из пяти сообщений искажено}\}$;
4) $D = \{\text{хотя бы одно из пяти сообщений искажено}\}$.

Задание 3. В ящике содержится 15 деталей, изготовленных фирмой №1, 20 деталей – фирмой №2 и 20 деталей – фирмой №3. Вероятность того, что деталь, изготовленная фирмой №1, отличного качества, равна 0,8; для деталей, изготовленных фирмами №2 и №3, эти вероятности соответственно равны 0,9 и 0,6. Найти вероятность того, что извлеченная наудачу деталь окажется отличного качества.

Задание 4. Прибор может собираться из высококачественных деталей и из деталей обычного качества. 45% приборов собирается из высококачественных деталей и их надежность за время t равна 97%. Приборы из обычных деталей за время t имеют надежность 0,72. Прибор испытан и за время t работал безотказно. Какова вероятность того, что он собран из высококачественных деталей?

Задание 5. Вероятность выпуска бракованного изделия на станке равна 0,2. Определить вероятность того, что в партии из десяти выпущенных на данном станке деталей ровно k будут без брака. Решить задачу для k = 0, 1, 10.

2 ЛАБОРАТОРНАЯ РАБОТА
«ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ СЛУЧАЙНЫХ ВЕЛИЧИН»

Цель работы: изучение числовых характеристик одномерных случайных величин, приобретение инструментального навыка их вычисления в ППП Excel.

2.1 Основные теоретические положения

Случайной величиной называется величина, которая в результате испытания принимает только одно значение из возможного множества своих значение, наперед неизвестное и зависящее от случайных причин.

Законом распределения случайной величины называют соответствие между возможными значениями случайной величины и их вероятностями.

К основным числовым характеристикам случайных величин относятся: математическое ожидание (математическое ожидание случайной величины имеет смысл ее среднего значения), дисперсия, среднеквадратическое отклонение.

Случайная величина X называется ***дискретной***, если множество ее значений конечное или бесконечное, но счетное (возможные значения дискретной случайной величину можно перенумеровать).

Закон распределения (закон распределения вероятностей) дискретной случайной величины X может быть задан в виде таблицы, в первой строке которой указаны в порядке возрастания все возможные значения случайной величины, а во второй строке соответствующие вероятности этих значений).

Для задания дискретной случайной величины наряду с законом распределения используется ***функция распределения вероятностей*** $F(x)$ случайной величины X, определяемая по формуле

$$F(x) = P(X < x).$$

Математическим ожиданием дискретной случайной величины X называется выражение вида

$$M[X] = \sum_{i=1}^{n} x_i p_i = m_x.$$

Дисперсия характеризует величину отклонения случайной величины от ее среднего значения и определяется по формуле

$$D[X] = M\left[(X - m_x)^2\right] = \sum_{i=1}^{n} (x_i - m_x)^2 p_i = M\left[X^2\right] - m_x^2.$$

Для получения характеристики рассеивания случайной величины от ее среднего значения используют также ***среднеквадратическое отклонение***, определяемое по формуле

$$\sigma = \sqrt{D[X]}\,.$$

Наряду с перечисленными характеристиками случайной величины используют моду M_o и медиану M_e. ***Модой*** называется то значение случайной величины, вероятность появления которого максимальна. ***Медианой*** называется значение случайной величины, для которого вероятность появления случайной величины, меньшей или большей этого значения, является одинаковой. Рассчитываются также показатели формы распределения – коэффициенты ***асимметрии*** и ***эксцесса***.

Случайная величина X называется ***непрерывной***, если ее функция распределения $F(x) = P(X < x)$ непрерывна и имеет производную.

Функция $f(x)=F'(x)$ называется ***плотностью распределения***.

Так как $F(x)=P(X<x)=P(-\infty<X<x)$, то

$$F(x)=\int_{-\infty}^{x} f(x)dx.$$

Математическое ожидание непрерывной случайной величины X

$$M[X]=\int_{-\infty}^{\infty} x\cdot f(x)dx.$$

Дисперсия непрерывной случайной величины X

$$D[X]=\int_{-\infty}^{\infty} \left(x-M[X]\right)^2 f(x)dx.$$

Среднеквадратическое отклонение

$$\sigma=\sqrt{D[X]}.$$

2.2 Примеры выполнения заданий

Пример 1. В ящике содержится 7 стандартных и 3 бракованных детали. Вынимают детали последовательно до появления стандартной, не возвращая их обратно. Дискретная случайная величина X – число извлеченных бракованных деталей. Составить закон распределения дискретной случайной величины, вычислить ее математическое ожидание, дисперсию, среднеквадратическое отклонение, построить многоугольник распределения и график функции распределения.

Решение:

Дискретная случайная величина X – число извлеченных бракованных деталей может принимать значения 0, 1, 2, 3.

Если $X=0$, то первая же извлеченная деталь – стандартная; вероятность $P(X=0)=\frac{7}{10}$.

Если $X=1$, то первая извлеченная деталь – бракованная, вторая – стандартная; вероятность $P(X=1)=\frac{3}{10}\cdot\frac{7}{9}=\frac{7}{30}$.

Если $X=2$, то первая извлеченная деталь – бракованная, вторая – бракованная, третья – стандартная; вероятность $P(X=2)=\frac{3}{10}\cdot\frac{2}{9}\cdot\frac{7}{8}=\frac{7}{120}$.

Если $X=3$, то первая извлеченная деталь – бракованная, вторая – бракованная, третья – бракованная, четвертая – стандартная; вероятность $P(X=3)=\frac{3}{10}\cdot\frac{2}{9}\cdot\frac{1}{8}\cdot\frac{7}{7}=\frac{1}{120}$.

Закон распределения случайной величины представим в таблице:

x	0	1	2	3
p	$\frac{7}{10}$	$\frac{7}{30}$	$\frac{7}{120}$	$\frac{1}{120}$

Математическое ожидание случайной величины

$$M[X]=\sum_{i=1}^{n} x_i p_i = 0\cdot\frac{7}{10}+1\cdot\frac{7}{30}+2\cdot\frac{7}{120}+3\cdot\frac{1}{120}=\frac{3}{8}=0,375.$$

Дисперсия случайной величины

$$D[X]=\sum_{i=1}^{n}(x_i-m_x)^2 p_i=\left(0-\frac{3}{8}\right)^2\cdot\frac{7}{10}+\left(1-\frac{3}{8}\right)^2\cdot\frac{7}{30}+\left(2-\frac{3}{8}\right)^2\cdot\frac{7}{120}+\left(3-\frac{3}{8}\right)^2\cdot\frac{1}{120}=\frac{77}{192}=0,401$$

Среднеквадратическое отклонение $\sigma=\sqrt{D[X]}=\sqrt{\frac{77}{192}}=0,633$.

Составим функцию распределения случайной величины X: $F(x)=P(X<x)$.

При $x\le 0$: $F(x)=P(X<0)=0$; при $0<x\le 1$: $F(x)=P(X<1)=\frac{7}{10}$;

при $1<x\le 2$: $F(x)=P(X<2)=\frac{7}{10}+\frac{7}{30}=\frac{28}{30}=\frac{14}{15}$;

при $2<x\le 3$: $F(x)=P(X<3)=\frac{7}{10}+\frac{7}{30}+\frac{7}{120}=\frac{119}{120}$;

при $x>3$: $F(x)=\frac{7}{10}+\frac{7}{30}+\frac{7}{120}+\frac{1}{120}=1$ (событие достоверно):

$$F(x)=\begin{cases}0, & x\le 0,\\ \frac{7}{10}, & 0<x\le 1,\\ \frac{14}{15}, & 1<x\le 2,\\ \frac{119}{120}, & 2<x\le 3,\\ 1, & x>3.\end{cases}$$

B9 =КОРЕНЬ(B8)

	A	B	C	D	E
1	**Условие**	**Значение**			
2	Количество стандартных деталей	7			
3	Количество бракованных деталей	3			
4	Количество извлеченных деталей	0	1	2	3
5	Вероятность извлечения	0,7000	0,2333	0,0583	0,0083
6	Математическое ожидание	0,375	=СУММПРОИЗВ(B4:E4;B5:E5)		
7	Квадрат отклонения количества извлеченных деталей от математического ожидания	0,1406	0,3906	2,6406	6,8906
8	Дисперсия	0,4010	=СУММПРОИЗВ(B7:E7;B5:E5)		
9	Среднеквадратическое отклонение	0,6333	=КОРЕНЬ(B8)		

Рисунок 6 – Вычисление числовых характеристик случайной величины (пример 6)

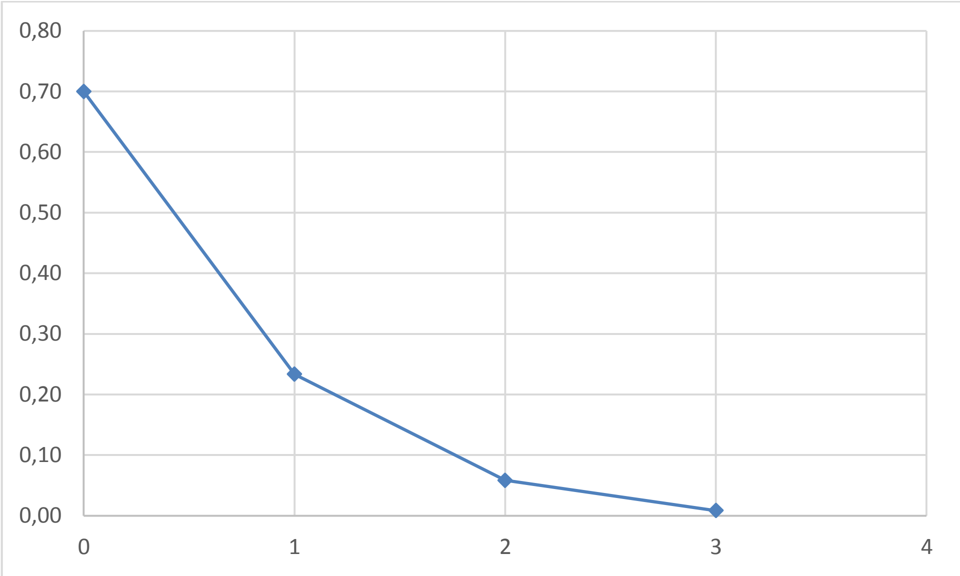


Рисунок 7 – Многоугольник распределения случайной величины

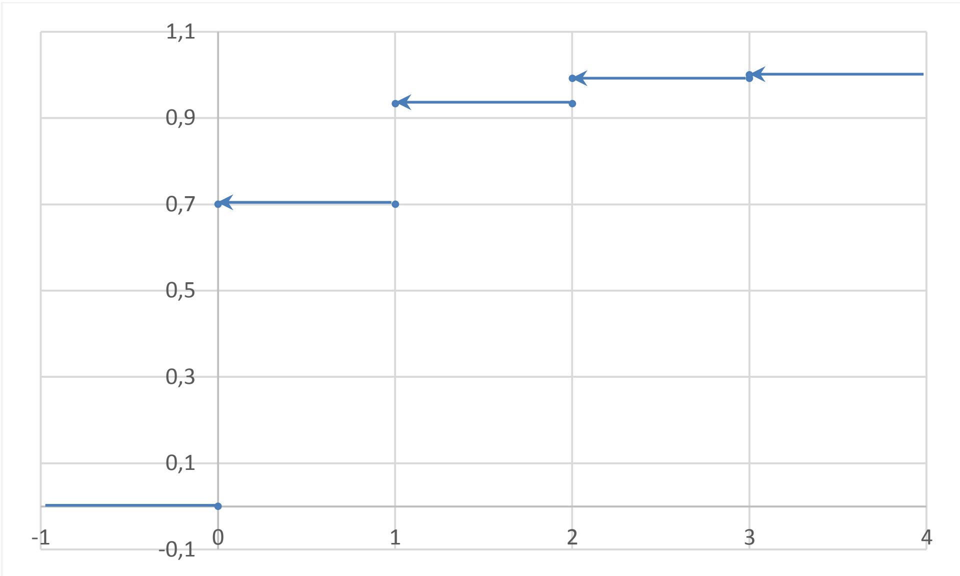


Рисунок 8 – Функция распределения случайной величины

Пример 2. Случайная величина задана плотностью распределения вероятностей:

$$f(x)=\begin{cases}0, & x<-\dfrac{\pi}{2},\\ \dfrac{1}{2}\cos x, & -\dfrac{\pi}{2}\le x\le\dfrac{\pi}{2},\\ 0, & x>\dfrac{\pi}{2}.\end{cases}$$

Найти функцию распределения случайной величины, математическое ожидание, дисперсию и среднеквадратическое отклонение.

Решение:

На интервале $\left(-\infty,\dfrac{\pi}{2}\right)$ функция распределения $F(x)=0$. Если $x\in\left[-\dfrac{\pi}{2},\dfrac{\pi}{2}\right]$, то

$$F(x)=\frac{1}{2}\int\limits_{-\frac{\pi}{2}}^{x}\cos x\,dx=\frac{1}{2}(1+\sin x).$$

При $x\in\left(\frac{\pi}{2},\infty\right)$ функция распределения $F(x)=\frac{1}{2}\int\limits_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\cos x\,dx=1$.

Таким образом,

$$F(x)=\begin{cases}0, & x<-\frac{\pi}{2},\\ \frac{1}{2}(1+\sin x), & -\frac{\pi}{2}\le x\le\frac{\pi}{2},\\ 1, & x>\frac{\pi}{2}.\end{cases}$$

Математическое ожидание

$$M[X]=\frac{1}{2}\int\limits_{-\frac{\pi}{2}}^{\frac{\pi}{2}}x\cdot\cos x\,dx=0.$$

Дисперсия

$$D[X]=\frac{1}{2}\int\limits_{-\frac{\pi}{2}}^{\frac{\pi}{2}}(x-0)^2\cdot\cos x\,dx=0,467.$$

Среднеквадратическое отклонение

$$\sigma=\sqrt{0,467}=0,683.$$

2.3 Задания для самостоятельного выполнения

Задание 1. В ящике содержится 10 стандартных и 5 бракованных деталей. Вынимают детали последовательно до появления стандартной, не возвращая их обратно. Дискретная случайная величина X – число извлеченных бракованных деталей. Составить закон распределения дискретной случайной величины, вычислить ее математическое ожидание, дисперсию, среднеквадратическое отклонение, построить многоугольник распределения и график функции распределения.

Задание 2. Случайная величина задана плотностью распределения вероятностей:

$$f(x)=\begin{cases}0, & x<0,\\ 2\sin x, & 0\le x\le\frac{\pi}{4},\\ 0, & x>\frac{\pi}{4}.\end{cases}$$

Найти функцию распределения случайной величины, математическое ожидание, дисперсию и среднеквадратическое отклонение.

3 ЛАБОРАТОРНАЯ РАБОТА «ОСНОВНЫЕ ЗАКОНЫ РАСПРЕДЕЛЕНИЯ ОДНОМЕРНЫХ СЛУЧАЙНЫХ ВЕЛИЧИН»

Цель работы: изучение основных законов распределения одномерных случайных величин, приобретение навыка вычисления параметров распределений и числовых характеристик, построения графиков посредством функционала ППП Excel.

3.1 Основные теоретические положения

Дискретная случайная величина X имеет ***биномиальный закон распределения*** с параметрами n и p, если она принимает значения 0, 1, 2, 3,…, m,…, n с вероятностями

$$P\left(X=m\right)=C_n^m p^m q^{n-m},$$

где $0<p<1$, $q=1-p$.

Для биномиального распределения:

– математическое ожидание $M\left[X\right]=n\cdot p$;

– дисперсия $D\left[X\right]=n\cdot p\cdot q$.

Дискретная случайная величина X имеет ***распределение Пуассона***, если она может принимать значения 0, 1, 2, 3,…, k,… с вероятностями

$$P\left(X=k\right)=\frac{\lambda^k}{k!}e^{-\lambda},$$

где $\lambda>0$ – параметр распределения.

Для распределения Пуассона $M\left[X\right]=D\left[X\right]=\lambda$.

Дискретная случайная величина $X=m$ имеет ***геометрическое распределение*** с параметром p, если она может принимать значения 0, 1, 2, 3,…, m,… (бесконечное, но счетное множество значений) с вероятностями

$$P\left(X=m\right)=p\cdot q^{m-1},$$

где $0<p<1$, $q=1-p$.

Для геометрического распределения:

– математическое ожидание $M\left[X\right]=\dfrac{1}{p}$;

– дисперсия $D\left[X\right]=\dfrac{q}{p^2}$.

Непрерывная случайная величина X имеет ***равномерный (прямоугольный) закон распределения*** на отрезке $\left[a,\,b\right]$, если ее плотность вероятности $f\left(x\right)$ постоянна на этом отрезке и равна нулю вне его

$$f(x)=\begin{cases}\dfrac{1}{b-a}, & a\le x\le b,\\ 0, & x<a,\ x>b.\end{cases}$$

Функция распределения случайной величины X, распределенной по равномерному закону

$$F(x)=\begin{cases}0, & x\le a,\\ \dfrac{x-a}{b-a}, & a<x\le b,\\ 1, & x>b.\end{cases}$$

Для равномерного распределения:

– математическое ожидание $M\left[X\right]=\dfrac{a+b}{2}$;

– дисперсия $D\left[X\right]=\dfrac{\left(b-a\right)^2}{12}$.

Непрерывная случайная величина X имеет ***показательный (экспоненциальный) закон распределения*** с параметром $\lambda>0$, если ее плотность вероятности имеет вид

$$f(x)=\begin{cases}\lambda e^{-\lambda x}, & x\ge 0,\\ 0, & x<0.\end{cases}$$

Функция распределения случайной величины *X*, распределенной по показательному (экспоненциальному) закону

$$F(x)=\begin{cases}1-e^{-\lambda x}, & x\ge 0,\\ 0, & x<0.\end{cases}$$

Для показательного распределения:

– математическое ожидание $M\left[X\right]=\dfrac{1}{\lambda}$;

– дисперсия $D\left[X\right]=\dfrac{1}{\lambda^2}$.

Непрерывная случайная величина X имеет ***нормальный закон распределения*** **(*закон Гаусса*)** с параметрами a и σ^2 $\left(X\sim N\left(a,\sigma\right)\right)$, если ее плотность распределения имеет вид

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{\left(x-a\right)^2}{2\sigma^2}}.$$

Нормальное распределение случайной величины X_0 называется ***стандартным***, если $a=0$, $\sigma=1$, плотность имеет вид

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}.$$

Взаимосвязь случайных величин $X_0=\dfrac{X-a}{\sigma}$.

Функция распределения случайных величин X и X_0:

$$F\left(x\right)=\int_{-\infty}^{x}f\left(t\right)dt=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{\left(t-a\right)^2}{2\sigma^2}}dt;$$

$$F_0\left(x\right)=\int_{-\infty}^{x}f_0\left(t\right)dt=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}dt.$$

Для определения значений функции нормального распределения используются табличные значения функций:

– $\varphi(x)=\dfrac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ (четная функция Гаусса);

– $\Phi\left(x\right)=\dfrac{1}{\sqrt{2\pi}}\displaystyle\int_{0}^{x}e^{-\frac{t^2}{2}}dt=F_0\left(x\right)-0,5$ (нечетная функция Лапласа).

Вероятность попадания нормальной случайной величины в заданный интервал

$$P(\alpha < X < \beta) = \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right).$$

3.2 Примеры выполнения заданий

Пример 1. Устройство состоит из трех независимо работающих элементов. Вероятность отказа каждого элемента в одном опыте равна 0,1. Составить закон распределения числа отказавших элементов в одном опыте. Найти функцию распределения случайной величины и построить график функции. Найти математическое ожидание, дисперсию и среднеквадратическое отклонение дискретной случайной величины.

Решение:

Дискретная случайная величина X – число отказавших элементов в одном опыте, имеет следующие возможные значения: 0 (ни один из элементов устройства не отказал), 1 (отказал один элемент), 2 (отказало два элемента) и 3 (отказали три элемента).

Отказы элементов независимы друг от друга, вероятности отказа каждого элемента равны между собой, поэтому применима формула Бернулли.

Учитывая, что, по условию, $n = 3$, $p = 0{,}1$, $q = 1 - p = 1 - 0{,}1 = 0{,}9$, определим вероятности:

$P_3(0) = C_3^0 \cdot 0{,}1^0 \cdot 0{,}9^{3-0} = 0{,}729$;

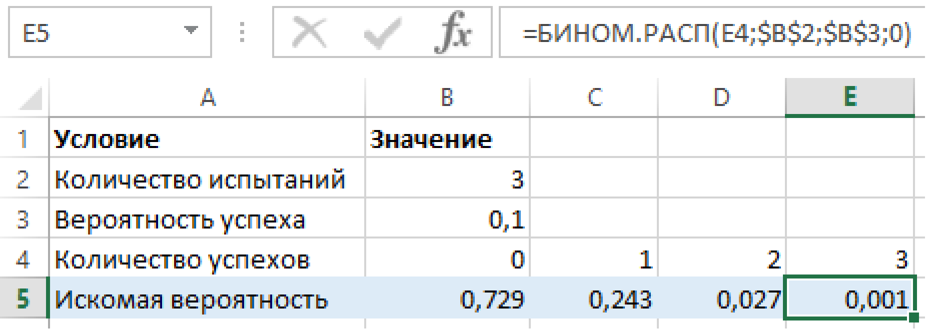
$P_3(1) = C_3^1 \cdot 0{,}1^1 \cdot 0{,}9^{3-1} = 0{,}243$;

$P_3(2) = C_3^2 \cdot 0{,}1^2 \cdot 0{,}9^{3-2} = 0{,}027$;

$P_3(3) = C_3^3 \cdot 0{,}1^3 \cdot 0{,}9^{3-3} = 0{,}001$.

Таким образом, искомый биномиальный закон распределения X имеет вид, представленный в таблице:

x	0	1	2	3
p	0,729	0,243	0,027	0,001


E5 =БИНОМ.РАСП(E4;$B$2;$B$3;0)

	A	B	C	D	E
1	Условие	Значение			
2	Количество испытаний	3			
3	Вероятность успеха	0,1			
4	Количество успехов	0	1	2	3
5	Искомая вероятность	0,729	0,243	0,027	0,001

Рисунок 9 – Биномиальный закон распределения случайной величины

Найдем функцию распределения

При $x \le 0$: $F(x) = P(X < 0) = 0$;

при $0 < x \le 1$: $F(x) = P(X < 1) = 0{,}729$;

при $1 < x \le 2$: $F(x) = P(X < 2) = 0{,}729 + 0{,}243 = 0{,}972$;

при $2 < x \le 3$: $F(x) = P(X < 3) = 0{,}729 + 0{,}243 + 0{,}027 = 0{,}999$;

при $x > 3$: $F(x) = 0{,}729 + 0{,}243 + 0{,}027 + 0{,}001 = 1$ (событие достоверно):

$$F(x)=\begin{cases}0, & x\le 0,\\ 0{,}729, & 0<x\le 1,\\ 0{,}972, & 1<x\le 2,\\ 0{,}999, & 2<x\le 3,\\ 1, & x>3.\end{cases}$$

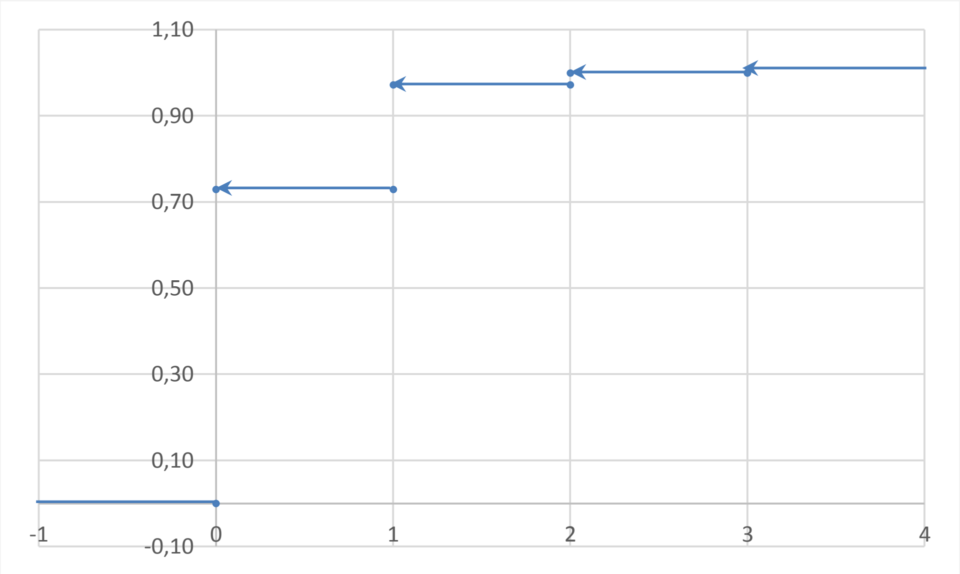


Рисунок 10 – Функция распределения случайной величины

Пример 2. Автомобиль проходит технический осмотр и обслуживание. Число неисправностей, обнаруженных во время техосмотра, распределяется по закону Пуассона с параметром 0,63. Если неисправностей не обнаружено, техническое обслуживание автомобиля продолжается в среднем 2 ч. Если обнаружены одна или две неисправности, то на устранение каждой из них тратится в среднем еще полчаса. Если обнаружено больше двух неисправностей, то автомобиль становится на профилактический ремонт, где он находится в среднем 4 ч. Определить закон распределения среднего времени T обслуживания и ремонта автомобиля и его математическое ожидание.

Решение:

Пусть T – дискретная случайная величина, равная времени обслуживания и ремонта автомобиля. Она может принимать значения 2, 2,5, 3 или 4 (часа). Найдем соответствующие вероятности. Для вычисления вероятности обнаружения k неисправностей, используем формулу Пуассона $P(T=k)=\dfrac{\lambda^k}{k!}e^{-\lambda}=\dfrac{0{,}63^k}{k!}e^{-0{,}63}$.

Если неисправностей не обнаружено, то $T=2$, $k=0$:

$$P(T=2)=\frac{0{,}63^0}{0!}e^{-0{,}63}\approx 0{,}533.$$

Если обнаружена одна неисправность, то $T=2{,}5$, $k=1$:

$$P(T=2{,}5)=\frac{0{,}63^1}{1!}e^{-0{,}63}\approx 0{,}336.$$

Если обнаружены две неисправности, то $T=3$, $k=2$:

$$P(T=3)=\frac{0{,}63^2}{2!}e^{-0{,}63}\approx 0{,}106.$$

Если обнаружено больше двух неисправностей, то $T=4$:

$$P(T=4)=1-0{,}533-0{,}336-0{,}106=0{,}026.$$

Получили закон распределения

T	2	2,5	3	4
p	0,533	0,336	0,106	0,026

Математическое ожидание $M[T]=2\cdot 0{,}533+2{,}5\cdot 0{,}336+3\cdot 0{,}106+4\cdot 0{,}026=2{,}328$.

Для получения закона распределения в Excel может быть использована функция ПУАССОН.РАСП, возвращающая распределение Пуассона.

Синтаксис функции:

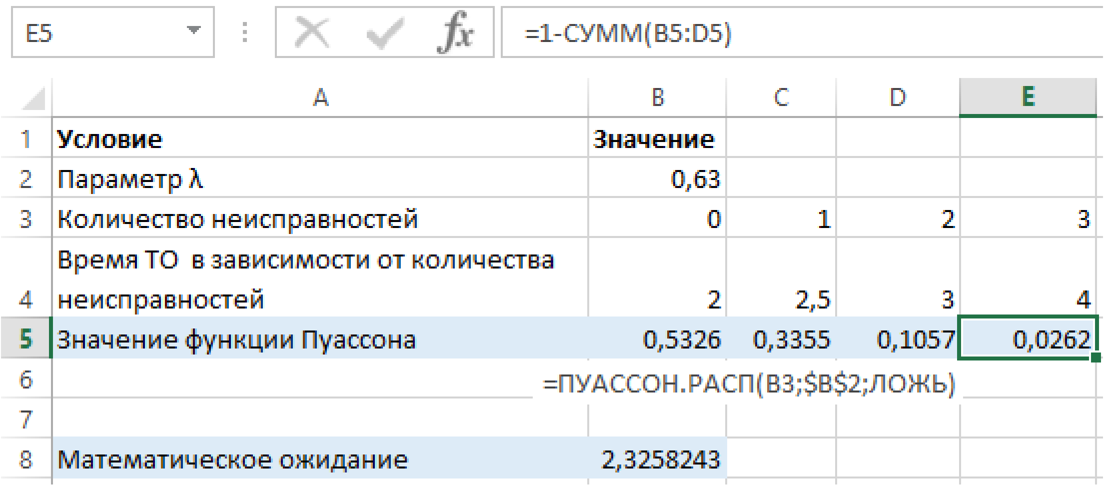
ПУАССОН.РАСП(x;среднее;интегральная)

Аргументы функции ПУАССОН.РАСП:

– x – обязательный аргумент. Количество событий.

– Среднее – обязательный аргумент. Ожидаемое числовое значение.

– Интегральная – обязательный аргумент. Логическое значение, определяя которое определяет форму возвращаемого распределения вероятности. Если значение ИСТИНА, то ПУАССОН.РАСП возвращает совокупное значение вероятности того, что число случайных событий включительно будет от нуля до *x*. Если ЛОЖЬ, возвращается вероятность того, что количество произошедших событий будет точно *x*.



	A	B	C	D	E
1	Условие	Значение			
2	Параметр λ	0,63			
3	Количество неисправностей	0	1	2	3
4	Время ТО в зависимости от количества неисправностей	2	2,5	3	4
5	Значение функции Пуассона	0,5326	0,3355	0,1057	0,0262
6		=ПУАССОН.РАСП(B3;$B$2;ЛОЖЬ)			
7					
8	Математическое ожидание	2,3258243			

Рисунок 11 – Закон распределения Пуассона

Пример 3. Проводится проверка большой партии деталей до обнаружения бракованной (без ограничения числа проверенных деталей). Составить закон распределения числа проверенных деталей. Найти его математическое ожидание и дисперсию, если известно, что вероятность брака для каждой детали равна 0,1.

Решение:

Случайная величина X – число проверенных деталей до обнаружения бракованной детали имеет геометрическое распределение с параметром $p=0{,}1$. Поэтому ряд распределения имеет вид $X=m$:

x_i	1	2	3	4	…	m	…
p_i	0,1	0,09	0,081	0,0729	…	$0{,}9^m\cdot 0{,}1$	…

Математическое ожидание $M\left[X\right]=\frac{1}{p}=\frac{1}{0{,}1}=10$.

Дисперсия $D\left[X\right]=\frac{q}{p^2}=\frac{0{,}9}{0{,}1^2}=90$.

Для получения закона распределения в Excel может быть использована функция ОТРБИНОМ.РАСП, возвращающая отрицательное биномиальное распределение – вероятность возникновения определенного числа неудач до указанного количества успехов при заданной вероятности успеха.

Эта функция аналогична биномиальному распределению и отличается от него тем, что количество успехов – фиксированное, а количество испытаний – переменное. Как и в случае биномиального распределения, испытания считаются независимыми.

Синтаксис функции:

ОТРБИНОМ.РАСП(число_неудач;число_успехов;вероятность_успеха;интегральная)

Аргументы функции ОТРБИНОМ.РАСП:

– Число_неудач – обязательный. Количество неудачных испытаний.
– Число_успехов – обязательный. Пороговое значение числа успешных испытаний.
– Вероятность_успеха – обязательный. Вероятность успеха.
– Интегральная – обязательный аргумент. Логическое значение, определяющее форму функции. Если аргумент "интегральная" имеет значение ИСТИНА, функция ОТРБИНОМ.РАСП возвращает интегральную функцию распределения; если этот аргумент имеет значение ЛОЖЬ, возвращается функция плотности распределения.

F4 =ОТРБИНОМ.РАСП(F3;1;$B$2;ЛОЖЬ)

	A	B	C	D	E	F
1	**Условие**	**Значение**				
2	Вероятность брака каждой детали	0,1				
3	Число проверенных деталей до обнаружения бракованной	0	1	2	3	4
4	Значение функции геометрическогораспределения	0,100	0,090	0,081	0,073	0,066
5		=ОТРБИНОМ.РАСП(B3;1;$B$2;ЛОЖЬ)				
6	Математическое ожидание	10	=1/B2			
7	Дисперсия	90	=(1-B2)/B2^2			

Рисунок 12 – Геометрическое распределение

Пример 4. Установлено, что время ремонта электронного прибора есть случайная величина *X*, распределенная по показательному закону. Определить вероятность того, что на ремонт прибора потребуется не менее 20 дней, если среднее время ремонта электронных приборов данной модели составляет 15 дней. Найти плотность вероятности, функцию распределения и среднеквадратическое отклонение случайной величины *X*.

Решение:

По условию математическое ожидание $M\left[X\right]=\frac{1}{\lambda}=15$, откуда параметр $\lambda=\frac{1}{15}$.

Тогда плотность вероятности и функция распределения имеют вид

$$f\left(x\right)=\frac{1}{15}e^{-\frac{1}{15}x}; \quad F\left(x\right)=1-e^{-\frac{1}{15}x} \quad \left(x\ge 0\right).$$

Вероятность того, что на ремонт прибора потребуется не менее 20 дней

$$P(X \ge 20) = 1 - P(X < 20) = 1 - F(20) = 1 - \left(1 - e^{-\frac{20}{15}}\right) = 0,264\,.$$

Среднеквадратическое отклонение $\sigma_X = M[X] = 15$ дней.

Для расчета значений плотности вероятности и функции показательного (экспоненциального) распределения в Excel может быть использована функция ЭКСП.РАСП, возвращающая экспоненциальное распределение.

Синтаксис функции: ЭКСП.РАСП(x;лямбда;интегральная)

Аргументы функции ЭКСП.РАСП:

– x – обязательный аргумент. Значение функции.

– Лямбда – обязательный аргумент. Значение параметра.

– Интегральная – обязательный аргумент. Логическое значение, определяющее форму экспоненциальной функции, которую следует использовать. Если аргумент "интегральная" имеет значение ИСТИНА, функция ЭКСП.РАСП возвращает интегральную функцию распределения; если имеет значение ЛОЖЬ, возвращается функция плотности распределения.

B6 =1/B2

	A	B	C	D	E
1	**Условие**	**Значение**			
2	Значение параметра λ = 1/15	0,0667			
3	Аргумент функции (20 дней)	0			
4	Значение функции экспоненциального распределения	0,7364	=ЭКСП.РАСП(20;1/15;ИСТИНА)		
5	Вероятность того, что на ремонт прибора потребуется не менее 20 дней	0,2636	=1-B4		
6	Математическое ожидание	15			
7	Среднеквадратическое отклонение	15			

Рисунок 13 – Экспоненциальное распределение

Пример 5. Автоматический токарный станок настроен на выпуск деталей со средним диаметром 2,00 см и со среднеквадратическим отклонением 0,005 см. Действует нормальный закон распределения. Компания технического сервиса рекомендует остановить станок для технического обслуживания и корректировки в случае, если образцы деталей, которые он производит, имеют средний диаметр более 2,01 см, либо менее 1,99 см.

Найти вероятность остановки станка, если он настроен но инструкции на 2,00 см.

Решение:

Пусть X – диаметр деталей – нормально распределенная случайная величина с параметрами $a = 2$ и $\sigma = 0,005$. Компания технического сервиса рекомендует остановить станок для технического обслуживания и корректировки в случае, если образцы деталей, которые он производит, имеют средний диаметр более 2,01 см либо менее 1,99 см. Найдем вероятность того, что станок будет работать:

$$P(1,99 < X < 2,01) = \Phi\left(\frac{2,01-2}{0,005}\right) - \Phi\left(\frac{1,99-2}{0,005}\right) = \Phi(2) - \Phi(-2) = \Phi(2) + \Phi(2) =$$

$$= 2\Phi(2) = 2 \cdot 0,4772 = 0,9544.$$

Тогда вероятность остановки станка $P = 1 - 0,9544 = 0,0456$.

Для расчета значений функции Лапласа в Excel может быть использована функция НОРМРАСП, возвращающая нормальную функцию распределения для указанного среднего и стандартного отклонения. Эта функция очень широко применяется в статистике, в том числе при проверке гипотез.

Синтаксис функции: НОРМРАСП(x;среднее;стандартное_откл;интегральная)

Аргументы функции НОРМРАСП:

– x – обязательный. Значение, для которого строится распределение.
– Среднее – обязательный. Среднее арифметическое распределения.
– Стандартное_откл – обязательный. Стандартное отклонение распределения.
– Интегральная – обязательный аргумент. Логическое значение, определяющее форму функции. Если аргумент "интегральная" имеет значение ИСТИНА, функция НОРМРАСП возвращает интегральную функцию распределения; если этот аргумент имеет значение ЛОЖЬ, возвращается весовая функция распределения.

B9 =1-B8

	A	B	C	D	E	F
1	**Условие**	**Значение**				
2	Средний диаметр	2,00				
3	Среднеквадратическое отклонение	0,005				
4	Верхнее предельное значение диаметра	2,01				
5	Нижнее предельное значение диаметра	1,99				
6	Значение функции Лапласа на верхнем предельном значении	0,97725	=НОРМ.РАСП(B4;$B$2;$B$3;ИСТИНА)			
7	Значение функции Лапласа на верхнем предельном значении	0,02275	=НОРМ.РАСП(B5;$B$2;$B$3;ИСТИНА)			
8	Вероятность работы станка (диаметр не выходит за пределы)	0,9545	=B6-B7			
9	Вероятность остановки станка	0,0455	=1-B8			

Рисунок 14 – Нормальное распределение

3.3 Задания для самостоятельного выполнения

Задание 1. Компьютер состоит из трех независимо работающих элементов: системного блока, монитора и клавиатуры. При однократном резком повышении напряжения вероятность отказа каждого элемента равна 0,1. Составить и охарактеризовать закон распределения числа отказавших элементов при скачке напряжения в сети.

Задание 2. Устройство состоит из 1000 элементов, работающих независимо один от другого. Вероятность отказа любого элемента в течение времени t равна 0,002. Необходимо: а) составить закон распределения отказавших за время t элементов; б) найти математическое ожидание и дисперсию этой случайной величины; в) определить вероятность того, что за время t откажет хотя бы один элемент.

Задание 3. Цена деления шкалы измерительного прибора равна 0,2. Показания прибора округляют до ближайшего целого числа. Полагая, что при отсчете ошибка округления распределена по равномерному закону, найти: 1) математическое ожидание, дисперсию и среднеквадратическое отклонение этой случайной величины; 2) вероятность того, что ошибка округления: а) меньше 0,04; б) больше 0,05.

Задание 4. Среднее время безотказной работы прибора равно 80 ч. Полагая, что время безотказной работы прибора имеет показательный закон распределения, найти: а) выражение его плотности вероятности и функции распределения; б) вероятность того, что в течение 100 ч прибор не выйдет из строя.

Задание 5. Ошибка измерителя дальности подчинена нормальному закону с систематической ошибкой 20 м и среднеквадратическим отклонением 60 м. Найти вероятность того, что измеренное значение дальности будет отклоняться от истинного не более, чем на 30 м.

4 ЛАБОРАТОРНАЯ РАБОТА «ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ ДВУМЕРНОЙ СЛУЧАЙНОЙ ВЕЛИЧИНЫ»

Цель работы: изучение числовых характеристик двумерных случайных величин, приобретение инструментального навыка их вычисления в ППП Excel.

4.1 Основные теоретические положения

Двумерной называется случайная величина (X,Y), возможные значения которой есть пары чисел (x,y). Составляющие X и Y, рассматриваемые одновременно, образуют систему двух случайных величин.

Законом распределения системы случайных величин называется соотношение, устанавливающее связь между областями возможных значений системы случайных величин и вероятностями появления системы в этих областях.

Функцией распределения системы двух случайных величин называется функция двух аргументов $F(x,y)$ равная вероятности совместного выполнения двух неравенств

$$F(x,y) = P(X < x, Y < y).$$

Дискретной называется двумерная величина, составляющие которой дискретны.

Пусть составляющие X и Y дискретны и имеют соответственно следующие возможные значения: $x_1, x_2, \ldots, x_n$, $y_1, y_2, \ldots, y_m$.

Закон распределения дискретной двумерной случайной величины может быть задан:

а) в виде таблицы, где вероятности $p_{ij} = P\left(X = x_i, Y = y_j\right)$, $i = 1,\ldots,n$, $j = 1,\ldots,m$:

(X,Y)	y_1	y_2	$\cdots$	y_m
x_1	p_{11}	p_{12}	$\cdots$	p_{1m}
x_2	p_{21}	p_{22}	$\cdots$	p_{2m}
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
x_n	p_{n1}	p_{n2}	$\cdots$	p_{nm}

б) аналитически, например, в виде функции распределения $F(x,y) = P(X < x, Y < y)$.

Так как события составляют полную группу попарно несовместных событий, то $\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m} p_{ij} = 1$.

Зная закон распределения двумерной дискретной случайной величины, можно найти законы каждой из составляющих. В общем случае, для того чтобы найти вероятность $P(X = x_i)$, нужно просуммировать вероятности столбца x_i

$$P(X = x_i) = \sum_j p_{ij} .$$

Аналогично сложив вероятности строки y_j, получим вероятность $P(Y = y_j)$.

$$P(Y = y_j) = \sum_i p_{ij} .$$

Условным распределением составляющей X при $Y = y_j$ называют совокупность условных вероятностей:

$$p(x_1/y_j),\ p(x_2/y_j),\ \ldots,\ p(x_n/y_j).$$

Аналогично определяется условное распределение *Y*.

Условные вероятности составляющих X и Y вычисляются соответственно по формулам:

$$p(x_i/y_j) = \frac{p(x_i, y_j)}{p(y_j)}, \quad p(y_j/x_i) = \frac{p(x_i, y_j)}{p(x_i)} .$$

Математические ожидания дискретных случайных величин X и Y вычисляются по формулам:

$$M[X] = \sum_{i=1}^{n} \sum_{j=1}^{m} x_i p_{ij} , \quad M[Y] = \sum_{i=1}^{n} \sum_{j=1}^{m} y_j p_{ij} .$$

Дисперсии дискретных случайных величин X и Y вычисляются по формулам:

$$D[X] = \sum_{i=1}^{n} \sum_{j=1}^{m} (x_i - M[X])^2 p_{ij} = M[X^2] - M^2[X],$$

$$D[Y] = \sum_{i=1}^{n} \sum_{j=1}^{m} (y_j - M[Y])^2 p_{ij} = M[Y^2] - M^2[Y].$$

Ковариацией (корреляционным моментом) дискретных случайных величин X и Y называется математическое ожидание произведения их отклонений

$$\operatorname{cov}(X,Y) = M[(X - M[X]) \cdot (Y - M[Y])] = M[X \cdot Y] - M[X] \cdot M[Y],$$

$$\operatorname{cov}(X,Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} [x_i - M(X)][y_j - M(Y)] p(x_i, y_j) .$$

Коэффициент корреляции – отношение ковариации двухмерной случайной величины к произведению среднеквадратических отклонений

$$r(X,Y) = \frac{\operatorname{cov}(X,Y)}{\sqrt{D[X] \cdot D[Y]}} .$$

4.2 Примеры выполнения заданий

Пример 1. С целью нарушения информационной безопасности двумя злоумышленниками независимо друг от друга были запущены вредоносные программы в корпоративные информационные системы (КИС) разных компаний – первым злоумышленником в две компании, вторым – в две другие компании. Вероятность поражения КИС вредоносными программами первого и второго злоумышленников $p_1 = 0,7$ и $p_2 = 0,4$ соответственно. Найти матрицу распределения и функцию распределения двумерной случайной величины (X,Y), где X – количество поражений КИС вирусом первого злоумышленника, Y – количество поражений КИС вирусом второго злоумышленника.

Решение:

Найдем множество значений случайных величин X и Y: число возможных поражений КИС вирусом каждого злоумышленника $\{0, 1, 2\}$. Тогда случайный вектор (X,Y) принимает 9 значений (x_i, y_j), $0 \le x_i, y_j \le 2$. Вычислим вероятность одного из них, учитывая независимость событий:

$$p(X=0, Y=0) = (1-p_1)^2 (1-p_2)^2 = 0,3^2 \cdot 0,6^2 = 0,0324.$$

Вероятность каждого из 9 значений случайного вектора вычисляется аналогично, например:

$$p(X=1, Y=0) = p_1(1-p_1)p_2^2 + (1-p_1)p_1p_2^2 = 0,7 \cdot 0,3 \cdot 0,6^2 \cdot 2 = 0,1512$$

Сводим вычисления в таблицу распределения:

(X,Y)	0	1	2
0	0,0324	0,0432	0,0144
1	0,1512	0,2016	0,0672
2	0,1764	0,2352	0,0784

Проверка: $\sum_{i=1}^{3}\sum_{j=1}^{3} p_{ij} = 1$.

По таблице находим законы распределения составляющих: вероятности значений X есть построчные суммы чисел p_{ij}, а для Y – суммы по столбцам таблицы:

(X,Y)	0	1	2	X
0	0,0324	0,0432	0,0144	0,09
1	0,1512	0,2016	0,0672	0,42
2	0,1764	0,2352	0,0784	0,49
Y	0,36	0,48	0,16	1

Найдем функцию распределения как функцию накопленных вероятностей. Значение функции $F(x,y)$ зависит от того, где находятся аргументы x и y по отношению к числам 0, 1, 2. Например, если $x \le 0$ или $y \le 0$, то $F(x,y) = 0$; если $1 < x \le 2$ и $0 < y \le 1$, то $F(x,y) = p((X=0, Y=0)$ или $(X=1, Y=0)) = p_{11} + p_{21} = 0,0324 + 0,1512 = 0,1836$.

Продолжая вычисления аналогично, получаем ступенчатую функцию накопленных вероятностей, задаваемую таблицей:

(x,y)	$y \le 0$	$0 < y \le 1$	$1 < y \le 2$	$y > 2$
$x \le 0$	0	0	0	0
$0 < x \le 1$	0	0,0324	0,0756	0,09
$1 < x \le 2$	0	0,1836	0,4284	0,51
$x > 2$	0	0,36	0,84	1

Пример 2. Дан закон распределения двумерной случайной величины (X,Y):

(X,Y)	-2	-1	0	1
-1	0,1	0,2	0,1	0,1
0	0,0	0,1	0,0	0,2
2	0,1	0,0	0,1	0,0

Записать одномерные законы распределения случайных величин *X* и *Y*, вычислить их математические ожидания и дисперсии, найти ковариацию и коэффициент корреляции.

Решение:

Найдем одномерный закон *X*:

X	-1	0	2	Итого
p	0,5	0,3	0,2	1

Математическое ожидание

$$M[X]=\sum_{i=1}^{n} x_i p_i = -1\cdot 0,5+0\cdot 0,3+2\cdot 0,2=-0,1$$

Дисперсия

$$D[X]=M\left[X^2\right]-M^2[X]=(-1)^2\cdot 0,5+0^2\cdot 0,3+2^2\cdot 0,2-(-0,1)^2=1,29\,.$$

Найдем одномерный закон *Y*:

Y	-2	-1	0	1	Итого
p	0,2	0,3	0,2	0,3	1

Математическое ожидание

$$M[Y]=\sum_{j=1}^{m} y_j p_j = -2\cdot 0,2-1\cdot 0,3+0\cdot 0,2+1\cdot 0,3=-0,4\,.$$

Дисперсия

$$D[Y]=M\left[Y^2\right]-M^2[Y]=(-2)^2\cdot 0,2+(-1)^2\cdot 0,3+0^2\cdot 0,2+1^2\cdot 0,3-(-0,4)^2=1,24\,.$$

Ковариация

$$\operatorname{cov}(X,Y)=M[X\cdot Y]-M[X]\cdot M[Y]=-0,1-(-0,1)\cdot(-0,4)=-0,14\,,$$

$$M[X\cdot Y]=-2\cdot(-1\cdot 0,1+0\cdot 0+2\cdot 0,1)-1\cdot(-1\cdot 0,2+0\cdot 0,1+2\cdot 0)+$$
$$+0\cdot(-1\cdot 0,1+0\cdot 0+2\cdot 0,1)+1\cdot(-1\cdot 0,1+0\cdot 0,2+2\cdot 0)=-0,1.$$

Коэффициент корреляции

$$r(X,Y)=\frac{\operatorname{cov}(X,Y)}{\sqrt{D[X]\cdot D[Y]}}=\frac{-0,14}{\sqrt{1,29\cdot 1,24}}=-0,111\,.$$

D22 =D21/КОРЕНЬ(D17*D19)

	A	B	C	D	E	F
1	(X,Y)	-2	-1	0	1	
2	-1	0,1	0,2	0,1	0,1	
3	0	0	0,1	0	0,2	
4	2	0,1	0	0,1	0	
5		Одномерный закон *X*				
6	X	-1	0	2	итог	
7	P	0,5	0,3	0,2	1	
8		=СУММ(B2:E2)			=СУММ(B9:D9)	
9	X*X	1	0	4		
10		Одномерный закон *Y*				
11	Y	-2	-1	0	1	итог
12	P	0,2	0,3	0,2	0,3	1
13		=СУММ(B2:B4)				=СУММ(B15:E15)
14	Y*Y	4	1	0	1	
15						
16	Математическое ожидание *X*			-0,1	=СУММПРОИЗВ(B8:D8;B9:D9)	
17	Дисперсия *X*			1,29	=СУММПРОИЗВ(B11:D11;B9:D9)-(СУММПРОИЗВ(B8:D8;B9:D9))^2	
18	Математическое ожидание *Y*			-0,4	=СУММПРОИЗВ(B14:E14;B15:E15)	
19	Дисперсия *Y*			1,24	=СУММПРОИЗВ(B17:E17;B15:E15)-(СУММПРОИЗВ(B14:E14;B15:E15))^2	
20	Математическое ожидание *XY*			-0,1		
21	Ковариация			-0,14	=D20-D16*D18	
22	Коэффициент корреляции			-0,111	=D21/КОРЕНЬ(D17*D19)	

Рисунок 15 – Числовые характеристики двумерной случайной величины

4.3 Задание для самостоятельного выполнения

Закон распределения дискретной двумерной случайной величины (X,Y) задан таблицей. Требуется:

– определить одномерные законы распределения случайных величин *X* и *Y*;
– вычислить математические ожидания случайных величин *X* и *Y*;
– вычислить дисперсии *X* и *Y*;
– вычислить ковариацию и коэффициент корреляции.

(X,Y)	3	5	8	10	12
-1	0,04	0,04	0,03	0,03	0,01
1	0,04	0,07	0,06	0,05	0,03
3	0,05	0,08	0,09	0,08	0,05
6	0,03	0,04	0,04	0,06	0,08

5 ЛАБОРАТОРНАЯ РАБОТА «ВАРИАЦИОННЫЕ РЯДЫ И ИХ ХАРАКТЕРИСТИКИ»

Цель работы: приобретение инструментального навыка построения вариационных рядов, изучение числовых характеристик и их вычисление в ППП Excel, приобретение навыка визуализации распределений.

5.1 Основные теоретические положения

Различные значения признака (случайной величины *X*) называются ***вариантами*** (обозначаем их через *x*).

Вариационным рядом называется ранжированный в порядке возрастания (или убывания) ряд вариантов с соответствующими им весами (частотами или частностями).

Числа, показывающие, сколько раз встречаются варианты из данного интервала, называются ***частотами*** n_i, а отношение их к общему числу наблюдений – ***частостями***, или ***относительными частотами***, т.е. $w_i = \dfrac{n_i}{n}$. Частоты и частости называются ***весами***.

Вариационный ряд называется ***дискретным***, если любые его варианты отличаются на постоянную величину, и ***непрерывным*** (***интервальным***), если варианты могут отличаться один от другого на сколь угодно малую величину.

Согласно формуле Стерджесса, рекомендуемое число интервалов

$$m = 1 + 3,322 \cdot \lg n,$$

величины интервалов $h = \dfrac{x_{\max} - x_{\min}}{m}$.

Для графического изображения вариационных рядов наиболее часто используются полигон, гистограмма, кумулятивная кривая.

Полигон, как правило, служит для изображения дискретного вариационного ряда и представляет собой ломаную, в которой концы отрезков имеют координаты $\left(x_i, n_i\right)$, $i = 1, \ldots, m$.

Гистограмма служит только для изображения интервальных вариационных рядов и представляет собой ступенчатую фигуру из прямоугольников с основаниями, равными интервалам значений признака и высотами, равными частотам (частостям) интервалов. Если соединить середины верхних оснований прямоугольников отрезками прямой, то можно получить *полигон* того же распределения.

Эмпирической функцией распределения $F_n\left(x\right)$ называется относительная частота (частость) того, что признак (случайная величина *X*) примет значение, меньшее заданного *x*, т.е.

$$F_n\left(x\right) = w\left(X < x\right) = w_x^{\text{накопл}}.$$

Вариационный ряд является статистическим аналогом (реализацией) распределения признака (случайной величины *X*). В этом смысле полигон (гистограмма) аналогичен кривой распределения, а эмпирическая функция распределения – функции распределения случайной величины *X*.

Средняя арифметическая вариационного ряда

$$\overline{x} = \frac{\sum\limits_{i=1}^{m} x_i \cdot n_i}{n} = \sum_{i=1}^{m} x_i \cdot w_i$$

Медианой вариационного ряда называется значение признака, приходящееся на середину ранжированного ряда наблюдений.

Модой вариационного ряда называется вариант, которому соответствует наибольшая частота.

Дисперсия вариационного ряда (эмпирическая, выборочная дисперсия)

$$s^2 = \frac{\sum_{i=1}^{m} \left(x_i - \overline{x}\right)^2 \cdot n_i}{n} = \sum_{i=1}^{m} \left(x_i - \overline{x}\right)^2 \cdot w_i \ .$$

Среднеквадратическое отклонение

$$s = \sqrt{\frac{\sum_{i=1}^{m} \left(x_i - \overline{x}\right)^2 \cdot n_i}{n}} = \sqrt{\sum_{i=1}^{m} \left(x_i - \overline{x}\right) \cdot w_i} \ .$$

5.2 Примеры выполнения заданий

Пример 1. В результате наблюдения за 45 образцами электронного оборудования получены данные о времени работы до первого отказа, результаты представлены интервальным рядом распределения

№ группы	Интервал времени работы, ч	Количество образцов, ед.	№ группы	Интервал времени работы, ч	Количество образцов, ед.
1	0-10	8	4	30-40	3
2	10-20	13	5	40-50	1
3	20-30	19	6	50-60	1

Рассчитать выборочное среднее, дисперсию и среднеквадратическое отклонение. Построить гистограмму и полигон распределения.

Решение:

Строим в Excel таблицу относительных частот, по ним рассчитываем накопленные частоты, выборочное среднее, дисперсию и среднеквадратическое отклонение.

По значениям относительных частот строим гистограмму и полигон частот.

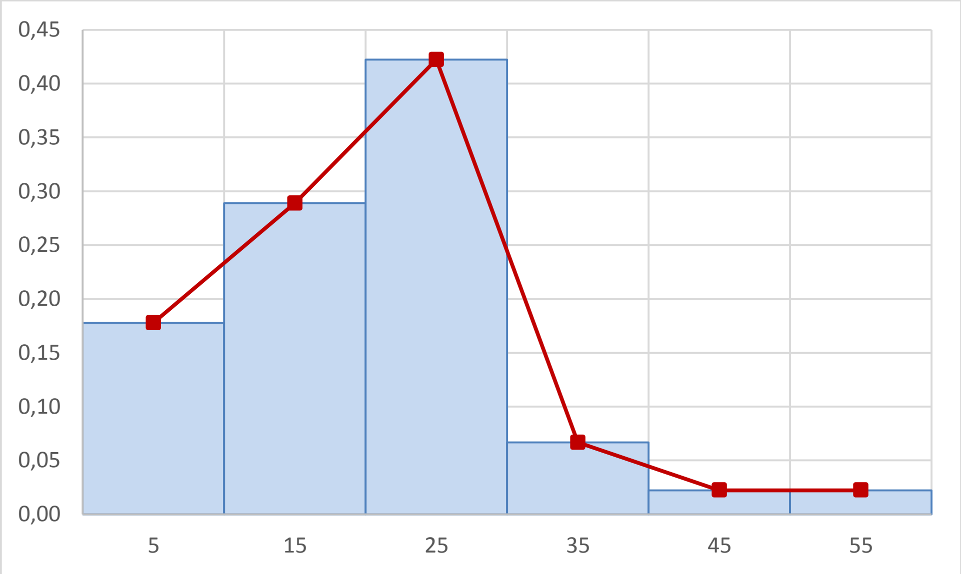


Рисунок 16 – Полигон и гистограмма частот

E11 | =КОРЕНЬ(E10)

	A	B	C	D	E	F	G
1	№ группы	Нижняя граница интервала	Верхняя граница интервала	Середина интервала	Количество образцов (частота)	Относительная частота	Накопленная частота
2	1	0	10	5	8	0,178	0,178
3	2	10	20	15	13	0,289	0,467
4	3	20	30	25	19	0,422	0,889
5	4	30	40	35	3	0,067	0,956
6	5	40	50	45	1	0,022	0,978
7	6	50	60	55	1	0,022	1,000
8			=СРЗНАЧ(B7:C7)		=E7/СУММ($E$2:$E$7)		=G6+F7
9	Выборочное среднее				20,333		
10	Выборочная дисперсия				113,778		
11	Выборочное среднеквадратическое отклонение				10,667		

Рисунок 17 – Расчет выборочных характеристик

5.3 Задания для самостоятельного выполнения

Задание 1. В результате наблюдения за 45 образцами электронного оборудования получены данные о времени работы до первого отказа, результаты представлены интервальным рядом распределения:

№ группы	Интервал времени работы, ч	Количество образцов, ед.	№ группы	Интервал времени работы, ч	Количество образцов, ед.
1	0-5	1	8	35-40	3
2	5-10	5	9	40-45	3
3	10-15	8	10	45-50	3
4	15-20	2	11	50-55	2
5	20-25	5	12	55-60	1
6	25-30	6	13	65-70	1
7	30-35	4	14	70-75	1

Построить гистограмму, полигон и эмпирическую функцию распределения. Рассчитать выборочное среднее, моду, медиану, дисперсию и среднеквадратическое отклонение.

Задание 2. При выборочной проверке по схеме собственно случайной бесповторной выборки из 200 микросхем было отобрано 100. Их распределение по задержке распространения сигнала дано в таблице:

Задержка сигнала, мкс	Менее 2900	2900-3000	3000-3100	3100-3200	3200-3300	3300-3400
Количество микросхем, ед.	80	50	270	160	300	140

Необходимо определить абсолютные и относительные частоты, накопленные частоты для статистических данных. Изобразить их в виде гистограммы, полигона и кумулятивной кривой. Вычислить среднюю арифметическую, вариационный размах, эмпирическую дисперсию, эмпирическое среднеквадратическое отклонение.

6 ЛАБОРАТОРНАЯ РАБОТА «ПРОВЕРКА ГИПОТЕЗ О ЗАКОНЕ РАСПРЕДЕЛЕНИЯ»

Цель работы: приобретение навыка проверки статистических гипотез о законах распределения дискретных и непрерывных случайных величин посредством использования функционала ППП Excel.

6.1 Основные теоретические положения

Универсальным критерием для проверки гипотезы о законе распределения является ***критерий Хи-квадрат Пирсона***. Данный критерий можно применять, как к количественным, так и категориальным данным.

Пусть $\{X_1, X_2, \ldots, X_n\}$ выборка из генеральной совокупности, причем предполагается, что данные измерены в количественной шкале. Проверяется гипотеза $H_0 : F = F_0$ против альтернативы $H_1 : F \neq F_0$ (H_0: выборочные значения получены из генеральной совокупности с функцией распределения $F(x)$, зависящей от ряда параметров)

Представим выборку в виде интервального ряда, разбив наблюдаемую область значений случайной величины на m интервалов (число интервалов группирования определяют, например, по формуле Стерджесса). Пусть n_i, $i = 1, \ldots, m$ – число элементов выборки попавших в i-ый интервал, а p_i – теоретическая вероятность попадания в этот интервал, вычисленная при условии истинности H_0. Заметим, что если параметры закона распределения F_0 неизвестны, необходимо предварительно найти оценки этих параметров, чтобы вычислить вероятности p_i (правильнее говорить в этом случае не о вероятностях p_i, а об их оценках). Составим статистику

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

которая характеризует сумму квадратов отклонений наблюдаемых значений n_i от ожидаемых np_i по всем интервалам группирования.

Если H_0 верна, то при фиксированном m и $n \to \infty$ статистика χ^2 стремится к распределению Хи-квадрат с $m - k - 1$ степенью свободы, где k – число неизвестных параметров распределения, оцениваемых по выборке. Очевидные, что малые наблюдаемые значения статистики χ^2 в пользу H_0, большие же значения в пользу H_1, поэтому критическую область значений статистики следует выбрать правостороннюю. Выберем в качестве $\chi^2_{\text{кр}}$ квантиль распределения Хи-квадрат уровня $1 - \alpha$, тогда критерий уровня значимости α для проверки H_0 будет иметь вид:

$$\delta = \begin{cases} H_0, & \chi^2_{\text{набл}} < \chi^2_{\text{кр}}, \\ H_1, & \chi^2_{\text{набл}} \ge \chi^2_{\text{кр}}. \end{cases}, \qquad \chi^2_{\text{набл}} = \sum_{i=1}^{m} \frac{\left(n_i - np_i\right)^2}{np_i},$$

где $\chi^2_{\text{кр}}$ – квантиль распределения Хи-квадрат уровня $1-\alpha$ с $m-k-1$ степенью свободы, k – число неизвестных параметров распределения, оцениваемых по выборке. Наблюдаемый уровень значимости критерия: $\alpha_{\text{набл}} = P\left(\chi^2 \ge \chi^2_{\text{набл}}\right)$.

Для корректного применения критерия Пирсона, важно чтобы ожидаемое значение np_i для каждого интервала было не слишком мало. Считается, минимально допустимое значение $np_i = 5$. Если данное условие для каких то интервалов не выполняется, следует либо объединять интервалы, либо по другому осуществлять разбивку на интервалы, в том числе, используя интервалы разной длины.

6.2 Примеры выполнения заданий

Пример 1. Известны данные об отказах аппаратуры за фиксированный промежуток времени 10000 часов:

Число отказов	0	1	2	3	4	5	≥ 6
Число случаев, в которых наблюдались отказы	35	33	20	10	1	1	0

При уровне значимости $\alpha = 0,05$ проверить нулевую гипотезу о том, что число отказов имеет распределение Пуассона $P\left(X = k\right) = \frac{\lambda^k}{k!} e^{-\lambda}$.

Решение:

Заносим исходные данные в таблицу Excel. Вычисляем по выборочным данным выборочную среднюю, принимаемую в качестве параметра λ распределения Пуассона. Используя функцию ПУАССОН.РАСП, вычисляем ожидаемые частоты. Задаем уровень значимости $\alpha = 0,05$ и число степеней свободы $m-k-1=6-1-1=4$. Вычисляем критериальную статистику $\chi^2_{\text{набл}}$, используя формулу массива.

Значение квантиля распределения распределения χ^2 вычисляем с помощью функции ХИ2ОБР.

Синтаксис функции ХИ2.ОБР(вероятность;степени_свободы)

Аргументы функции ХИ2.ОБР:

– Вероятность – обязательный аргумент. Вероятность, связанная с распределением хи-квадрат.

– Степени_свободы – обязательный аргумент. Число степеней свободы.

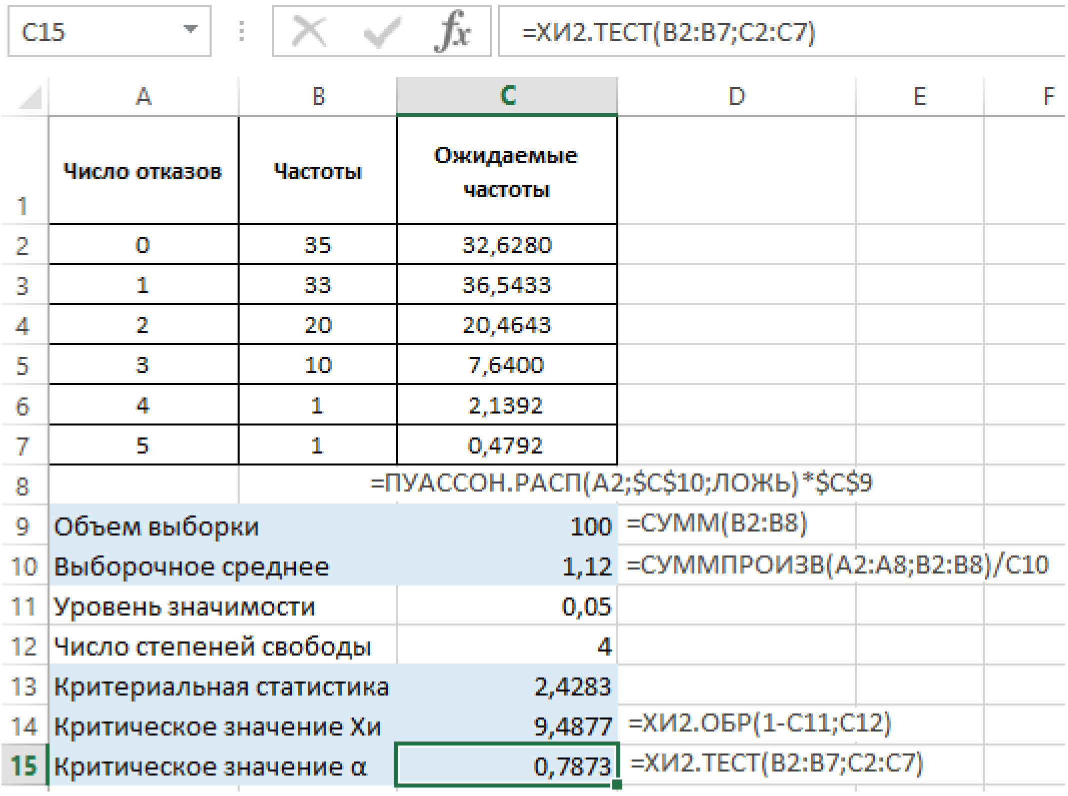
Функция ХИ2.ТЕСТ возвращает значение статистики для распределения хи-квадрат и соответствующее число степеней свободы.

Синтаксис функции ХИ2.ТЕСТ(фактический_интервал;ожидаемый_интервал)

Аргументы функции ХИ2.ТЕСТ:

– Фактический_интервал – обязательный аргумент. Интервал данных, который содержит результаты наблюдений, подлежащие сравнению с ожидаемыми значениями.

– Ожидаемый_интервал – обязательный аргумент. Интервал данных, который содержит отношение произведений итогов по строкам и столбцам к общему итогу.



C15 =ХИ2.ТЕСТ(B2:B7;C2:C7)

	A	B	C	D	E	F
1	Число отказов	Частоты	Ожидаемые частоты			
2	0	35	32,6280			
3	1	33	36,5433			
4	2	20	20,4643			
5	3	10	7,6400			
6	4	1	2,1392			
7	5	1	0,4792			
8			=ПУАССОН.РАСП(A2;$C$10;ЛОЖЬ)*$C$9			
9	Объем выборки		100	=СУММ(B2:B8)		
10	Выборочное среднее		1,12	=СУММПРОИЗВ(A2:A8;B2:B8)/C10		
11	Уровень значимости		0,05			
12	Число степеней свободы		4			
13	Критериальная статистика		2,4283			
14	Критическое значение Хи		9,4877	=ХИ2.ОБР(1-C11;C12)		
15	Критическое значение α		0,7873	=ХИ2.ТЕСТ(B2:B7;C2:C7)		

Рисунок 18 – Критерий Пирсона (распределение Пуассона)

Как видим из расчетов, $\chi^2_{\text{набл}} < \chi^2_{\text{кр}}$ (значение критериальной статистики существенно меньше критического), что говорит о том, что выборочные данные не противоречат гипотезе о том, что их распределение является распределением Пуассона с параметром $\lambda = 0,6$.

Кроме того, критическое значение $\alpha = 0,7873$ значительно больше уровня значимости $\alpha = 0,05$, что также свидетельствует о необходимости принятия гипотезы H_0.

Пример 2. В результате испытания 200 элементов на длительность работы построен интервальный вариационный ряд. Величина X – длительность работы в часах:

Номер интервала	Границы интервала	Частота
1	0 – 5	133
2	5 – 10	45
3	10 – 15	15
4	15 – 20	4
5	20 – 25	2
6	25 – 30	1

Требуется, при уровне значимости $\alpha = 0,05$ проверить гипотезу о том, что время работы элементов распределено по показательному закону.

Решение:

Записываем исходные данные в таблицу, определяем середины интервалов, абсолютные частоты и относительные частоты. Находим среднее время работы элементов $\bar{x} = 5$.

Выдвигаем гипотезу H_0 – генеральная совокупность имеет показательное распределение. Задаем уровень значимости $\alpha = 0{,}05$ и вычисляем теоретические вероятности попадания в интервалы. Для расчета значений плотности вероятности и функции показательного (экспоненциального) распределения в Excel может быть использована функция ЭКСП.РАСП, возвращающая экспоненциальное распределение.

Синтаксис функции: ЭКСП.РАСП(x;лямбда;интегральная)

Аргументы функции ЭКСП.РАСП:

– x – обязательный аргумент. Значение функции.

– Лямбда – обязательный аргумент. Значение параметра.

– Интегральная – обязательный аргумент. Логическое значение, определяющее форму экспоненциальной функции, которую следует использовать. Если аргумент "интегральная" имеет значение ИСТИНА, функция ЭКСП.РАСП возвращает интегральную функцию распределения; если имеет значение ЛОЖЬ, возвращается функция плотности распределения.

Ожидаемые частоты определяем как произведение объема выборки на теоретическую вероятность.

C16 | =ХИ2.ТЕСТ(D2:D7;F2:F7)

	A	B	C	D	E	F	G
1	Нижняя граница интервала	Верхняя нраница интервала	Середина интервала	Частоты	Теоретическая вероятность	Ожидаемые частоты	
2	0	5	2,5	133	0,6321	126,424	=$C$9*E2
3	5	10	7,5	45	0,2325	46,509	
4	10	15	12,5	15	0,0855	17,110	
5	15	20	17,5	4	0,0315	6,294	
6	20	25	22,5	2	0,0116	2,316	
7	25	30	27,5	1	0,0043	0,852	
8		=ЭКСП.РАСП(B7;$C$11;ИСТИНА)-ЭКСП.РАСП(A7;$C$11;ИСТИНА)					
9	Объем выборки		200	=СУММ(D2:D7)			
10	Выборочное среднее		5,00	=СУММПРОИЗВ(C2:C7;D2:D7)/C9			
11	Параметр распределения λ		0,2	=1/C10			
12	Уровень значимости		0,05				
13	Число степеней свободы		4				
14	Критериальная статистика		1,5562				
15	Критическое значение Хи		9,4877	=ХИ2.ОБР(1-C12;C13)			
16	Критическое значение α		0,9065	=ХИ2.ТЕСТ(D2:D7;F2:F7)			

Рисунок 19 – Критерий Пирсона (показательное распределение)

Как видим из расчетов, $\chi^2_{набл} < \chi^2_{кр}$ (значение критериальной статистики существенно меньше критического), что говорит о том, что выборочные данные не противоречат гипотезе о том, что их распределение является показательным с параметром $\lambda = 0{,}2$.

Кроме того, критическое значение $\alpha = 0{,}9065$ значительно больше уровня значимости $\alpha = 0{,}05$, что также свидетельствует о необходимости принятия гипотезы H_0.

6.3 Задания для самостоятельного выполнения

Задание 1. Известны данные об отказах аппаратуры за фиксированный промежуток времени 10000 часов:

Число отказов	0	1	2	3	4	5	≥ 6
Число случаев, в которых наблюдались отказы	35	29	18	11	6	1	0

При уровне значимости $\alpha = 0,05$ проверить нулевую гипотезу о том, что число отказов имеет распределение Пуассона $P\left(X = k\right) = \frac{\lambda^k}{k!} e^{-\lambda}$.

Задание 2. Распределение 200 диодных электрических лампочек по времени безотказной работы представлено в таблице:

Время безотказной работы, ч	1000-1100	1100-1200	1200-1300	1300-1400	1400-1500	1500-1600
Число устройств	133	45	15	4	2	1

Проверить гипотезу о том, что время безотказной работы диодных электрических лампочек имеет показательный закон распределения. Уровень значимости принять равным $\alpha = 0,1$.

Задание 3. Даны результаты измерения некоторой числовой характеристики микропроцессоров, которые поступают на предприятие для дальнейшей установки на изделия бытовой техники:
36, 39, 43, 45, 26, 34, 50, 33, 36, 57, 29, 40, 31, 34, 17, 47, 39, 35, 41, 28, 25, 30, 39, 36, 49, 42, 24, 27, 20, 52, 36, 33, 18, 32, 56, 37, 40, 29, 31, 46, 38, 19, 28, 33, 42, 26, 35, 37, 34, 48, 44, 22, 36, 49, 30, 27, 40, 32, 41, 43, 45, 38, 24, 37, 46, 36, 29, 25, 39, 52, 50, 21, 38, 34, 41, 47, 29, 31, 28, 35, 44, 55, 39, 30, 27, 32, 34, 40, 54, 36, 25, 53, 45, 33, 43, 37, 26, 42, 28, 51.

1. Построить интервальный ряд распределения.
2. Построить гистограмму и полигон относительных частот.
3. Выдвинуть гипотезу о возможном законе распределения.
4. Найти частоты закона распределения и проверить справедливость гипотезы о законе распределения с помощью критерия Пирсона при уровне значимости $\alpha = 0,05$.

7 ЛАБОРАТОРНАЯ РАБОТА

«ПАРНАЯ ЛИНЕЙНАЯ РЕГРЕССИЯ И КОРРЕЛЯЦИЯ»

Цель работы: приобретение навыка вычисления и интерпретации линейного коэффициента корреляции, построения уравнения парной линейной регрессии и характеристики его качества с использованием надстройки Excel «Анализ данных».

7.1 Основные теоретические положения

Парная регрессия – уравнение связи двух переменных *y* и *x*:

$$y = \hat{f}(x),$$

где y – зависимая переменная (результативный признак); x – независимая переменная (факторный признак).

Различают ***линейные*** и ***нелинейные*** регрессии. ***Линейная регрессия***: $y = a + b \cdot x + \varepsilon$.

Коэффициентом корреляции случайных величин X и Y (коэффициентом корреляции Пирсона) называется отношение корреляционного момента к произведению среднеквадратических отклонений этих величин

$$r_{XY} = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y}.$$

Коэффициент корреляции является безразмерной величиной.

Если две случайные величины X и Y имеют в отношении друг друга линейные функции регрессии, то говорят, что величины X и Y связаны ***линейной корреляционной зависимостью***.

Выборочный коэффициент линейной корреляции определяется по выборке (x_i, y_i), $i = 1, 2, \ldots, n$ из двумерной нормально распределенной случайной величины объема n равенством

$$r_{xy} = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\sqrt{\overline{y^2} - \left(\overline{y}\right)^2} \sqrt{\overline{x^2} - \left(\overline{x}\right)^2}}.$$

Абсолютная величина коэффициента корреляции не превышает единицы $|r_{xy}| \le 1$. Чем ближе величина $|r_{xy}|$ к единице, тем теснее линейная связь и тем лучше линейная зависимость согласуется с данными наблюдений. При $r_{xy} > 0$ связь является прямой, при $r_{xy} < 0$ – обратной. Для оценки тесноты связи применяют шкалу Чеддока:

Показатели	Характеристика
0,1-0,3	слабая
0,3-0,5	умеренная
0,5-0,7	заметная
0,7-0,9	высокая
0,9-0,99	весьма высокая

Для оценки ***значимости линейного коэффициента корреляции*** r_{xy} применяется *t*-критерий Стьюдента, согласно которому выдвигается «нулевая» гипотеза о статистической незначимости коэффициента линейной корреляции.

Эта гипотеза отвергается при выполнении условия $t_r > t_{\text{табл}}$, где t_r – фактическое значение *t*-критерий Стьюдента, определяется как

$$t_r = r_{xy} \cdot \sqrt{\frac{n-2}{1 - r_{xy}^2}}$$

$t_{\text{табл}}$ – критическое (табличное) значение, определяется из таблицы значений *t*-критерий Стьюдента с учетом заданного уровня значимости α и числа степеней свободы $k = n - 2$.

Уровнем значимости (обозначается α) в статистических гипотезах называется вероятность отвергнуть верную гипотезу (это так называемая ошибка первого рода). Уровень значимости α обычно принимает значения 0,05 и 0,01, что соответствует вероятности совершения ошибки первого рода 5% и 1%.

Таким образом, если $t_r > t_{\text{табл}}$, то величина линейного коэффициента корреляции признается существенной (статистически значимой), в противном случае гипотеза о статистической незначимости коэффициента корреляции принимается.

Построение уравнения регрессии сводится к оценке ее параметров. Для оценки параметров регрессии, линейной по параметрам, используют ***метод наименьших квадратов*** (МНК). МНК позволяет получить такие оценки параметров, при которых сумма квадратов отклонений фактических значений результативного признака y от теоретических $\hat{y}_x$ минимальна, т.е.

$$\sum (y - \hat{y}_x)^2 \to \min .$$

Для линейных уравнений, а также для уравнений, приводимых к линейным, решается следующая система уравнений относительно параметров a и b:

$$\begin{cases} na + b\sum x = \sum y, \\ a\sum x + b\sum x^2 = \sum xy. \end{cases}$$

Решая систему, получаем:

$$a = \bar{y} - b \cdot \bar{x}, \qquad b = \frac{\overline{yx} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2}.$$

F-тест, используемый для оценивания качества уравнения регрессии, состоит в проверке гипотезы H_0 о статистической незначимости уравнения регрессии. Для этого выполняется сравнение фактического F_r и критического (табличного) $F_{\text{табл}}$ значений F-критерия Фишера. Расчетное значение F_r определяется из соотношения значений факторной и остаточной дисперсий, рассчитанных на одну степень свободы:

$$F_r = \frac{\sum (\hat{y} - \bar{y})^2 / m}{\sum (y - \hat{y})^2 / (n - m - 1)} = \frac{r_{xy}^2}{1 - r_{xy}^2}(n-2),$$

где n – число единиц совокупности; m – число параметров при переменных x.

$F_{\text{табл}}$ – это максимально возможное значение критерия под влиянием случайных факторов при данных степенях свободы и уровне значимости α. Уровень значимости α – вероятность отвергнуть правильную гипотезу при условии, что она верна. Обычно α принимается равной 0,05 или 0,01. Если $F_r > F_{\text{табл}}$, то гипотеза о случайной природе оцениваемых характеристик отклоняется и признается их статистическая значимость и надежность, в противном случае признается статистическая незначимость, ненадежность уравнения регрессии.

Для оценки ***статистической значимости коэффициентов регрессии*** рассчитываются ***t-критерий Стьюдента*** и ***доверительные интервалы*** каждого из показателей. Выдвигается гипотеза H_0 о случайной природе показателей, т.е. о незначимом их отличии от нуля. Оценка значимости коэффициентов регрессии и корреляции с помощью t-критерия Стьюдента проводится путем сопоставления их значений с величиной случайной ошибки:

$$t_a = \frac{a}{m_a}, \quad t_b = \frac{b}{m_b}.$$

Стандартные ошибки параметров линейной регрессии определяются по формулам:

$$m_b = \sqrt{\frac{\sum (y - \hat{y}_x)^2 / (n-2)}{\sum (x - \bar{x})^2}} = \sqrt{\frac{S_{ocm}^2}{\sum (x - \bar{x})^2}} = \frac{S_{ocm}^2}{\sigma_x \sqrt{n}};$$

$$m_a = \sqrt{\frac{\sum(y-\hat{y}_x)^2}{(n-2)} \cdot \frac{\sum x^2}{n\sum(x-\bar{x})^2}} = \sqrt{S_{ост}^2 \frac{\sum x^2}{n^2\sigma_x^2}} = S_{ост}^2 \frac{\sqrt{\sum x^2}}{n\sigma_x}.$$

Сравнивая фактическое и критическое значения t-статистики, принимаем или отвергаем нулевую гипотезу. Если $t_r > t_{табл}$, то параметры a и b не случайно отличаются от нуля и сформировались под влиянием систематически действующего фактора x, в противном случае гипотеза о статистической незначимости коэффициента корреляции принимается.

Для расчета доверительного интервала определяем ***предельную ошибку Δ*** для каждого показателя:

$$\Delta_a = t_{табл} m_a, \qquad \Delta_b = t_{табл} m_b.$$

Формулы для расчета доверительных интервалов имеют вид:

$$\gamma_a = a \pm \Delta_a, \qquad \gamma_b = b \pm \Delta_b.$$

Если в границы доверительного интервала попадает ноль, т.е. Нижняя граница отрицательна, а верхняя положительна, то оцениваемый параметр принимается нулевым, так как он не может одновременно принимать и положительное, и отрицательное значения.

7.2 Пример выполнения задания

На производственном предприятии, упаковывающем свои изделия в тонкую полиэтиленовую пленку перед складированием уровень дефектности в апреле существенно вырос. Были высказаны несколько гипотез, одна из которых – влияние влажности воздуха в помещении цеха. Для проверки гипотезы были собраны данные за неделю:

Относительная влажность воздуха, %	78	84	72	90	94	85	80
Доля дефектов, %	2,2	2,6	2,1	2,7	2,8	2,6	2,2

Построить корреляционное поле. Рассчитать линейный коэффициент корреляции, сформулировать вывод о зависимости доли дефектов от относительной влажности воздуха. Определить параметры линейной модели регрессии методом наименьших квадратов, охарактеризовать статистическую значимость параметров модели и уравнения регрессии в целом. Построить график линейной зависимости доли дефектов от влажности воздуха.

Решение:

По исходным данным, приведенным в таблице строим корреляционное поле, по которому визуально определяем характер линейной зависимости доли дефектов от относительной влажности воздуха.

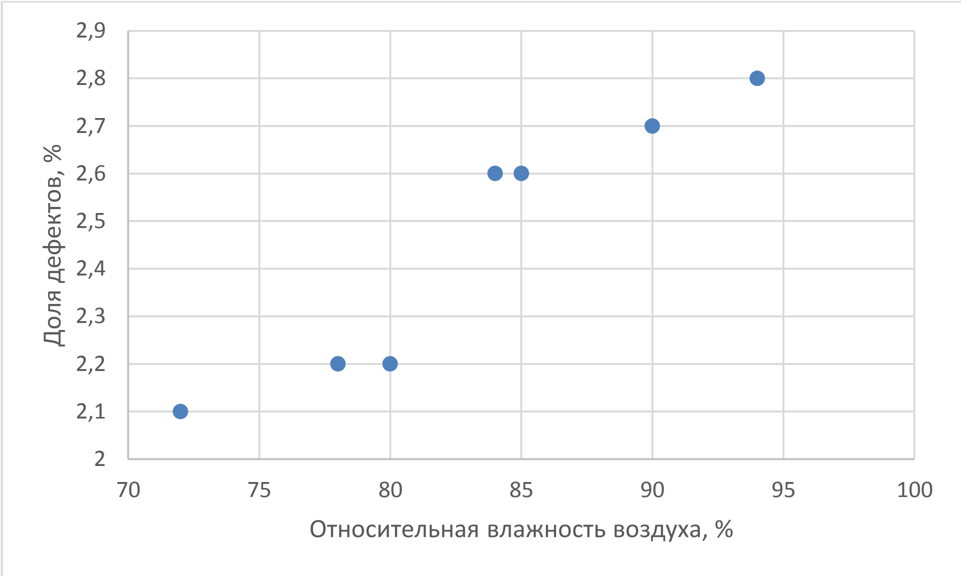


Рисунок 20 – Корреляционное поле

Сформируем расчетную таблицу $(n=7)$

x	y	xy	x^2	y^2
78	2,2	171,6	6084	4,84
84	2,6	218,4	7056	6,76
72	2,1	151,2	5184	4,41
90	2,7	243,0	8100	7,29
94	2,8	263,2	8836	7,84
85	2,6	221,0	7225	6,76
80	2,2	176,0	6400	4,84
$\sum x = 583$	$\sum y = 17,2$	$\sum xy = 1444,4$	$\sum x^2 = 48885$	$\sum y^2 = 42,74$
$\overline{x} = 83,286$	$\overline{y} = 2,457$	$\overline{xy} = 206,343$	$\overline{x^2} = 6983,571$	$\overline{y^2} = 6,106$

Линейный коэффициент корреляции

$$r_{xy} = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\sqrt{\overline{y^2} - \left(\overline{y}\right)^2} \sqrt{\overline{x^2} - \left(\overline{x}\right)^2}} = \frac{206,343 - 83,286 \cdot 2,457}{\sqrt{6,106 - 2,457^2} \sqrt{6983,571 - 83,286^2}} = \frac{1,709}{0,263 \cdot 6,857} = 0,95,$$

по шкале Чеддока степень линейной зависимости доли дефектов от влажности воздуха весьма высокая, по направлению – прямая.

Для расчета линейного коэффициента парной корреляции в Excel может быть использована функция КОРРЕЛ, возвращающая коэффициент корреляции двух диапазонов ячеек.

Синтаксис функции: КОРРЕЛ(массив1;массив2)

Аргументы функции КОРРЕЛ:

– массив1 – обязательный аргумент. Диапазон значений ячеек.
– массив2 – обязательный аргумент. Второй диапазон значений ячеек.

D12 | =КОРРЕЛ(A2:A8;B2:B8)

	A	B	C	D	E	F	G
1	x	y	xy	xx	yy		
2	78	2,2	171,6	6084	4,84		
3	84	2,6	218,4	7056	6,76		
4	72	2,1	151,2	5184	4,41		
5	90	2,7	243	8100	7,29		
6	94	2,8	263,2	8836	7,84		
7	85	2,6	221	7225	6,76		
8	80	2,2	176	6400	4,84		
9	583	17,2	1444,4	48885	42,74	=СУММ(E2:E8)	
10	83,286	2,457	206,343	6983,571	6,106	=СРЗНАЧ(E2:E8)	
11							
12	Коэффициент корреляции			0,948	=КОРРЕЛ(A2:A8;B2:B8)		

Рисунок 21 – Коэффициент корреляции

Для оценки значимости линейного коэффициента корреляции r_{xy} применяем *t*-критерий Стьюдента, определяем расчетное значение критерия

$$t_r = r_{xy} \cdot \sqrt{\frac{n-2}{1-r_{xy}^2}} = 0,95\sqrt{\frac{7-2}{1-0,95^2}} = 6,803,$$

$t_{\text{табл}} = 2,571$ – критическое (табличное) значение, определяем с помощью функции СТЬЮДЕНТ.ОБР.2Х, возвращающей двустороннее обратное *t*-распределение Стьюдента с учетом заданного уровня значимости $\alpha = 0,05$ и числа степеней свободы $k = n - 2 = 7 - 2 = 5$.

Синтаксис функции СТЬЮДЕНТ.ОБР.2Х(вероятность;степени_свободы)

Аргументы функции СТЬЮДЕНТ.ОБР.2Х:

– Вероятность – обязательный. Вероятность, соответствующая *t*-распределению Стьюдента.
– Степени_свободы – обязательный. Число степеней свободы, характеризующее распределение.

14	Уровень значимости	0,05
15	Число степеней свободы	5
16	Расчетное значение критерия Стьюдента	2,5706
17		=СТЬЮДЕНТ.ОБР.2Х(E14;E15)

Рисунок 22 – Критическое значение критерия Стьюдента

Так как $t_r > t_{\text{табл}}$, то величина линейного коэффициента корреляции признается существенной (статистически значимой).

Для проведения регрессионного анализа используем надстройку «Анализ данных», инструмент анализа «Регрессия».

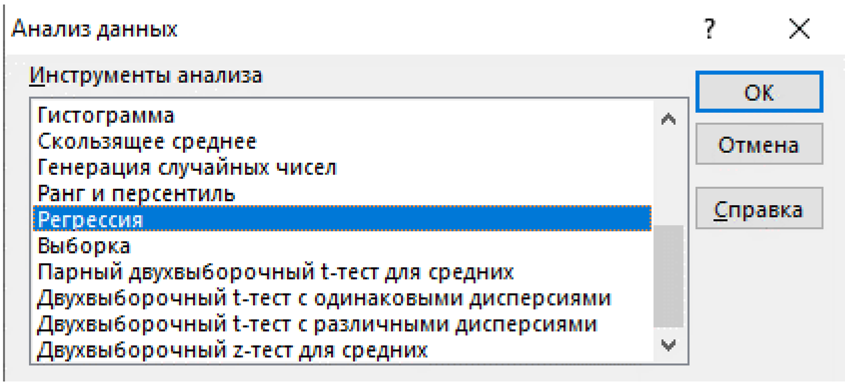


Рисунок 23 – Инструмент «Регрессия» надстройки «Анализ данных»

В открывшемся диалоговом окне задаем диапазоны значений, указываем начальную ячейку выходного интервала данных.

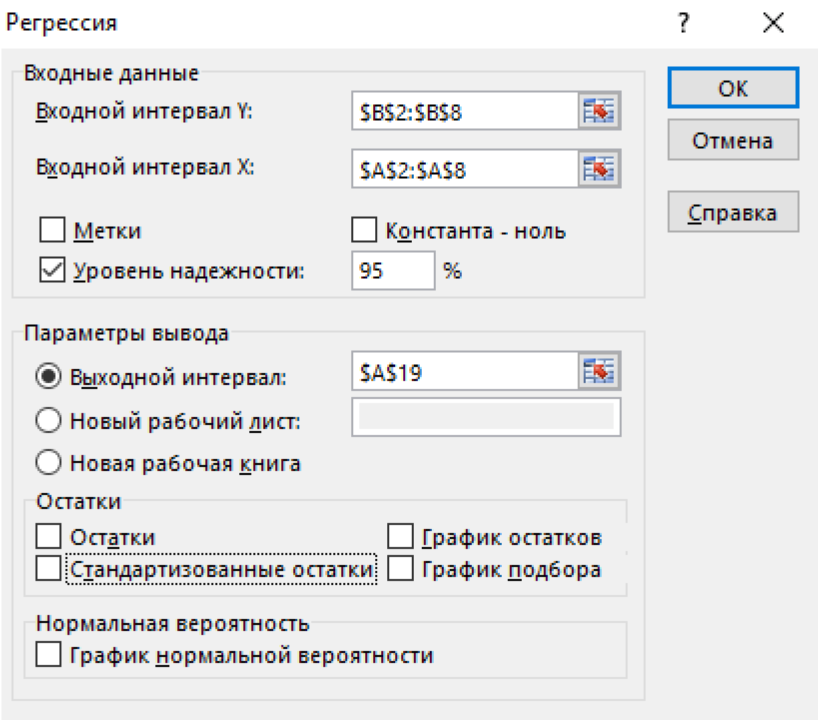


Рисунок 24 – Окно «Регрессия»

В разделе «Дисперсионный анализ» вывода итогов, выбираем значение F-критерия Фишера и сравниваем его с табличным, которое находим с учетом заданного уровня значимости $\alpha = 0{,}05$ и числа степеней свободы $k = n - 2 = 7 - 2 = 5$. Так как $F_r > F_{табл}$ существенно превышает табличное, делаем вывод о статистическое значимости уравнения регрессии.

28	Дисперсионный анализ					
29		*df*	*SS*	*MS*	*F*	*Значимость F*
30	Регрессия	1	0,4288	0,4288	44,3847	0,0012
31	Остаток	5	0,0483	0,0097		
32	Итого	6	0,4771			

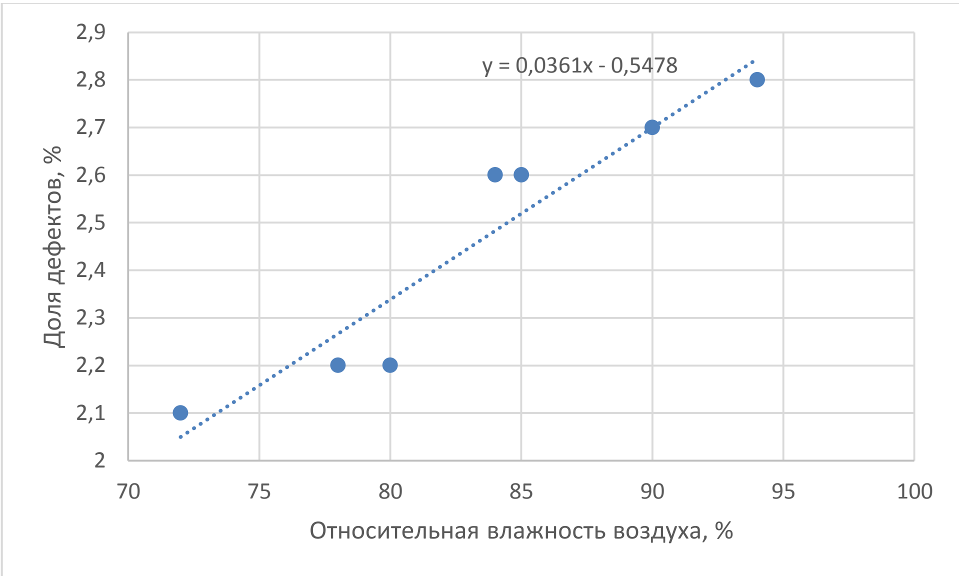
Рисунок 25 – Дисперсионный анализ

В следующем разделе вывода итогов выбираем значения параметров уравнения регрессии и фактические значения *t*-статистик для параметров.

	Коэффициенты	*Стандартная ошибка*	*t-статистика*
Y-пересечение	-0,547788378	0,452570352	-1,210393866
Переменная X 1	0,036079792	0,005415609	6,662185379

Рисунок 26 – Оценка параметров уравнения регрессии

Сравниваем расчетные значения *t*-статистик с табличным. Так как $t_r > t_{\text{табл}}$ для коэффициента регрессии существенно превышает табличное, делаем вывод о статистической значимости данного параметра. Свободный член уравнения признается не значимым.



7.3 Задание для самостоятельного выполнения

Суть эксперимента состоит в изучении зависимости основных характеристик лампы накаливания – светового потока и цветовой температуры от напряжения в питающей сети. На источник света (лампа накаливания) подавалось напряжение 220 В. Устанавливался светоприемник на расстояние от лампы, при котором освещенность равнялась 1000 лк. С помощью лабораторного трансформера устанавливались значения напряжения и регистрировались значения освещенности и цветовой температуры:

Напряжение, В	230	220	210	200	190	180	170	160	150	110
Световой поток, лк	1140	1000	861	764	635	538	452	379	308	101
Цветовая температура, К	3040	3000	2950	2900	2840	2790	2740	2680	2650	2350

Для характеристики линейной зависимости светового потока и цветовой температуры от напряжения в сети рассчитать линейные коэффициенты корреляции, сделать выводы. Построить корреляционные поля. Определить параметры линейных моделей зависимостей светового потока и цветовой температуры от напряжения в сети методом наименьших квадратов, определить для них доверительные интервалы, охарактеризовать статистическую значимость параметров моделей и уравнений регрессии в целом. Построить графики.

СПИСОК РЕКОМЕНДУЕМОЙ ЛИТЕРАТУРЫ

1. Кремер, Н. Ш. Теория вероятностей и математическая статистика : учебник и практикум для вузов / Н. Ш. Кремер. – 5-е изд., перераб. и доп. – Москва : Издательство Юрайт, 2023. – 538 с. – (Высшее образование). – ISBN 978-5-534-10004-4. – Текст : электронный // Образовательная платформа Юрайт [сайт]. – URL: https://urait.ru/bcode/517540 (дата обращения: 10.06.2023).

2. Энатская, Н. Ю. Теория вероятностей и математическая статистика для инженерно-технических направлений : учебник и практикум для вузов / Н. Ю. Энатская, Е. Р. Хакимуллин. – Москва : Издательство Юрайт, 2023. – 399 с. – (Высшее образование). – ISBN 978-5-534-02662-7. – Текст : электронный // Образовательная платформа Юрайт [сайт]. – URL: https://urait.ru/bcode/511231 (дата обращения: 13.05.2023).

3. Яковлев, В. Б. Статистика. Расчеты в Microsoft Excel : учебное пособие для вузов / В. Б. Яковлев. – 2-е изд., испр. и доп. – Москва : Издательство Юрайт, 2023. – 353 с. – (Высшее образование). – ISBN 978-5-534-01672-7. – Текст : электронный // Образовательная платформа Юрайт [сайт]. – URL: https://urait.ru/bcode/514005 (дата обращения: 29.05.2023).