

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное образовательное
учреждение высшего образования
«СЕВАСТОПОЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

Кафедра «Информационные технологии и компьютерные системы»

СТАТИСТИЧЕСКИЙ ПАКЕТ SPSS ДЛЯ СОЦИАЛЬНЫХ НАУК

Методические указания
для выполнения практических работ по дисциплине
для студентов очной формы обучения направления подготовки
39.03.01 «Социология»

В трех частях

Часть I

Математические методы в социологических исследованиях

Севастополь
СевГУ
2025

УДК 316:004.428(076.5)
ББК 60.5:32.973-018.2я73
С78

Р е ц е н з е н т :
Т. В. Волкова - доцент кафедры ИТиКС

Составитель: И. А. Балакирева

С78 Статистический пакет SPSS для социальных наук : методические указания к выполнению практических работ по дисциплине для студентов очной формы обучения направления подготовки 39.03.01 «Социология». – В 3-х частях. Ч. 1. Математические методы в социологических исследованиях / Севастопольский государственный университет, кафедра «Информационные технологии и компьютерные системы» ; сост. И. А. Балакирева. – Севастополь : СевГУ, 2025. – 68 с. – Текст : электронный.

Целью методических указаний является оказание помощи студентам в изучении понятий общей теории статистики и подготовка к работе со статистическим пакетом PSPP.

УДК 316:004.428(076.5)
ББК 60.5:32.973-018.2я73

Рассмотрены и утверждены на заседании кафедры ИТиКС, протокол № 8 от 11 марта 2025 г.

© Балакирева И. А., сост., 2025
© ФГАОУ ВО «Севастопольский
государственный университет», 2025

СОДЕРЖАНИЕ

1	Практическая работа № 1. Закон распределения и числовые характеристики случайных величин.....	4
2	Практическая работа № 2. Интервальный статистический ряд. Частотный анализ.....	14
3	Практическая работа № 3. Статистические оценки параметров распределения.....	19
4	Практическая работа № 4. Критерии согласия.....	22
5	Практическая работа № 5. Проверка гипотезы об однородности дисперсий.....	28
6	Практическая работа № 6. Сравнение средних показателей в выборках на основе критерия Стьюдента.....	32
7	Практическая работа № 7. Исследование взаимосвязи между данными на основе корреляционного анализа.....	40
8	Практическая работа № 8. Исследование взаимосвязи между данными на основе дисперсионного анализа.....	43
9	Практическая работа № 9. Анализ взаимосвязей между данными на основе регрессионного анализа.....	48
	Приложение А. Варианты для выполнения задания по практической работе № 1.....	53
	Приложение Б. Варианты для выполнения задания по практической работе № 2.....	57
	Приложение В. Критические точки для статистики Колмогорова.....	58
	Приложение Г. Варианты для выполнения задания по практической работе № 5.....	59
	Приложение Д. Варианты для выполнения задания по практической работе № 6.....	61
	Приложение Е. Варианты для выполнения задания по практической работе № 8.....	62

Практическая работа № 1

Закон распределения и числовые характеристики случайных величин

Цель: ознакомиться с законами распределения и числовыми характеристиками случайных величин.

1 Краткие теоретические сведения.

1.1 Понятие случайной величины.

Под **случайной величиной** (СВ) понимается величина, которая в результате опыта со случайным исходом принимает то или иное значение, причем до опыта, неизвестно, какое именно. СВ будем обозначать большими буквами: X, Y, Z ; их значения - соответствующими малыми буквами: x, y, z , Ω_x - множество возможных значений величины X .

СВ X называется **дискретной** (ДСВ), если множество Ω_x - счетное, т.е. его элементы можно расположить в определенном порядке и пронумеровать.

СВ X называется **непрерывной** (НСВ), если множество Ω_x - несчетное.

1.2 Способы описания СВ

Законом распределения СВ X называется любая функция (правило, таблица и т.п.), устанавливающая соответствие между значениями СВ и вероятностями их наступления и позволяющая находить вероятности всевозможных событий $p(a \leq x < b)$, $\forall a, b$, связанных с СВ.

Функцией распределения $F(x)$ СВ X называется вероятность того, что она примет значение меньшее, чем аргумент функции x $F(x) = P\{X < x\}$.

Функция распределения используется при рассмотрении как дискретных, так и непрерывных СВ.

Для описания дискретных СВ наряду с функцией распределения $F(x)$ используется ряд распределения вероятностей.

Рядом распределения ДСВ X называется таблица, в верхней строке которой перечислены все возможные значения СВ $x_1, x_2, \dots, x_n$, ($x_{i-1} < x_i$),

а в нижней - вероятности их появления $p_1, p_2, \dots, p_n$, где $p_i = P(X = x_i)$.

x_i	x_1	x_2	...	x_n
p_i	p_1	p_2	...	p_n

Функция распределения (ФР) любой дискретной СВ есть разрывная ступенчатая функция, скачки которой происходят в точках, соответствующих возможным значениям случайной величины, и равны вероятностям этих значений: $F(x) = \sum_{x_1 < x_2} p(X = x_i)$, где

суммирование распространяется на все значения x_i , которые меньше x (рис. 1.1а)).

Многоугольник вероятностей есть графическое изображение ряда распределения вероятностей. По оси абсцисс откладываются возможные значения СВ, а по оси ординат - вероятности этих значений (рис. 1.1б)). Для наглядности полученные точки соединяются отрезками прямых. Многоугольник распределения, как и ряд распределения, полностью характеризует СВ и является одной из форм закона распределения.

2 Числовые характеристики случайных величин.

В числовых характеристиках СВ выделяют:

- меры центральной тенденции;
- меры рассеяния;
- меры формы распределения.

2.1 Меры центральной тенденции случайных величин

К мерам центральной тенденции **дискретных** СВ относятся:

- математическое ожидание (среднее)

$$m_x = \sum_{i=1}^N x_i \cdot p_i$$

- модальное значение (мода),

$$Mo = \max\{x_i\}$$

которая является наиболее типичным (часто встречающимся) значением СВ.

Пример 1. СВ задана рядом распределений, представленным в таблице 1.1. Построить функцию распределения СВ и многоугольник частот, определить меры центральной тенденции - среднее и модальное значения.

Таблица 1.1 Исходный ряд распределений СВ

x	1	2	3	4	5
P	0,37	0,21	0,10	0,15	0,17

Для построения функции распределения прежде всего необходимо вычислить вероятности для каждого ее значения:

$$F(x < 1) = 0$$

$$F(x < 4) = F(x < 1) + 0,37$$

$$F(x < 6) = F(x < 1) + 0,21 = 0,58$$

$$F(x < 7) = F(x < 6) + 0,10 = 0,68$$

$$F(x < 8) = F(x < 7) + 0,15 = 0,83$$

$$F(x \geq 8) = F(x < 8) + 0,17 = 1.$$

Вычисленные значения записать в таблицу 1.2, которая имеет следующий вид:

Таблица 1.2 Ряд распределений СВ

x	1	2	3	4	5
P	0,37	0,21	0,10	0,15	0,17
F	0,37	0,58	0,68	0,83	1

Функция распределения строится по значениям 1-й и 3-й строк таблицы 1.2, а многоугольник вероятностей по 1-й и 2-й строкам (рис.1.1).

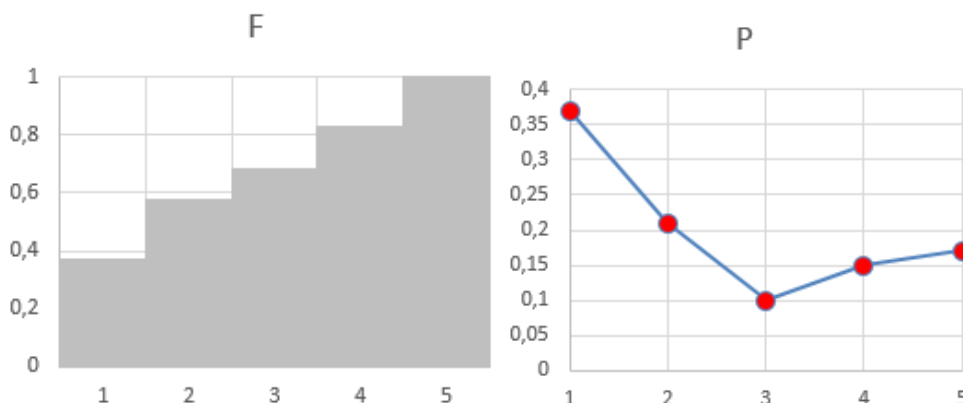


Рисунок 1.1 – графическое представление законов распределения ДСВ:
а) функция распределения; б) многоугольник вероятностей

Среднее значение ДСВ, заданной рядом распределения вычисляется по формуле (1.1). Рекомендуется для вычисления использовать функцию СУММПРОИЗВ:

	A	B	C	D	E	F
1	x	1	2	3	4	5
2	P	0,37	0,21	0,1	0,15	0,17
3	F	0,37	0,58	0,68	0,83	1
4						
5						
6	mx=	=СУММПРОИЗВ(B1:F1;B2:F2)				

а)

6	mx=	2,54
---	-----	------

б)

Среднее значение равно 2,54.

В заданном распределении наиболее часто встречающимся значением (модой) является $x=1$, вероятность его появления максимальная – 0,37.

Для **непрерывных** СВ в качестве мер центральной тенденции используются следующие характеристики: среднее, мода, медиана.

Среднее значение НСВ вычисляется по формуле:

$$m_x = \frac{\sum_{i=1}^n x_i}{n}$$

Для вычисления медианы необходимо упорядочить значения ряда, пронумеровать их и найти середину ряда. Если число значений в ряду нечетное, выбирается середина ряда, если четное, то два средних значения усредняются это значение и будет медианой. Кроме этого, можно использовать функцию МЕДИАНА.

Для определения наиболее часто встречающегося значения используется функция МОДА.

Пример 2. Непрерывная СВ задана значениями:

21	28	19	18	25
30	32	29	34	21
32	23	19	36	32
30	43	32	37	46
38	29	38	41	42
33	41	38	47	44
42	35	32	44	43

Вычислить меры центральной тенденции распределения: среднее, моду и медиану. Среднее значение вычислить: используя формулу (1.1) и функцию СРЗНАЧ:

B	C	D	E	F	G		B	C	D	E	F				
21	28	19	18	25			21	28	19	18	25				
30	32	29	34	21			30	32	29	34	21				
32	23	19	36	32			32	23	19	36	32				
30	43	32	37	46			30	43	32	37	46				
38	29	38	41	42			38	29	38	41	42				
33	41	38	47	44			33	41	38	47	44				
42	35	32	44	43			42	35	32	44	43				
mx=						=СУММ(B2:F8)/СЧЁТ(B2:F8)	mx=						=CP3HA4(B2:F8)	mx=	33,5

	A	B	C	D	E	F	G	H	I	J	K
1											
2		21	28	19	18	25					
3		30	32	29	34	21		Мо=	32		
4		32	23	19	36	32		Мо=	=МОДА(B2:F8)		
5		30	43	32	37	46					
6		38	29	38	41	42					
7		33	41	38	47	44					
8		42	35	32	44	43					

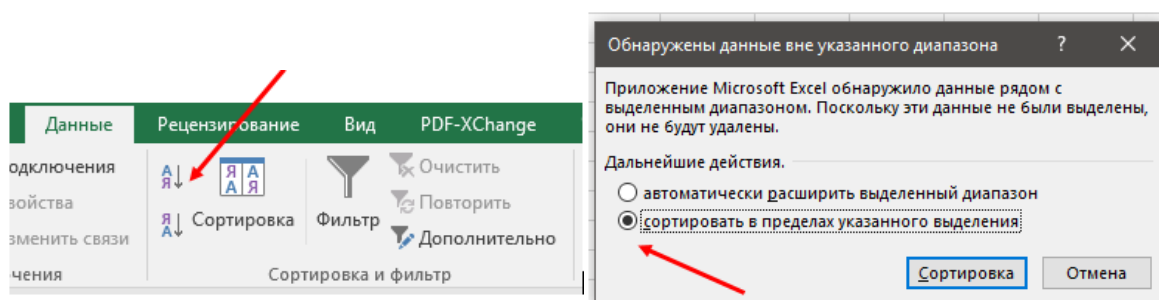
Модальное значение ряда равно 32, то есть это наиболее часто встречающееся значение признака или наиболее типичное его значение.

Для вычисления **медианы** преобразуем матрицу со случайными значениями в столбец, упорядочим по возрастанию и пронумеруем значения.

18	1
19	2
19	3
21	4
21	5
23	6
25	7
28	8
29	9
29	10
30	11
30	12
32	13
32	14
32	15
32	16
32	17
33	18
34	19
35	20

Для нумерации по порядку в ячейке набрать номер 1 для первого значения, установить маркер заполнения и при нажатой клавише CTRL протянуть указатель мыши до последнего значения.

Для упорядочения необходимо выделить интервал со значениями, выбрать команду **Данные** и указанную стрелкой кнопку. Затем в ДО установите указанный переключатель



Пронумеруйте упорядоченную выборку. Наблюдение с номером 18, то есть 33, и есть медиана распределения. Если эти значения интерпретировать как возраст респондентов, то ровно половина их не старше 33 лет. Поскольку не всегда возможно визуально определять медиану, для ее вычисления используется функция МЕДИАНА:

Me=	33
Me=	=МЕДИАНА(B2:F8)

Поскольку меры центральной тенденции (среднее, медиана и мода) незначительно отличаются друг от друга, можно сделать предположение, что распределение является симметричным.

Кроме медианы в статистике вычисляют **квартили**. Каждый квартиль представляет четверть распределения.



Для вычисления значений квартилей применяется функция КВАРТИЛЬ.ВКЛ. Аргументами функции является интервал с распределением и номер квартиля (части распределения) – 1, 2, 3. Если необходимо выбрать наименьшее и наибольшее значение распределения, то в качестве второго аргумента соответственно указывается 0 или 4

1 кварт	29
2 кварт	33
3 кварт	41
3 кварт	=КВАРТИЛЬ(В2:F8;3)

В статистике интерес представляют значения первого Q1 и третьего Q3 квартилей. Эти значения называют точками Тьюки. Смысл их в том, что ровно половина распределения заключена между этими точками. Если эти значения интерпретировать как возраст респондентов, то ровно половина их имеет возраст от 29 лет до 41 года.

2.2 Меры рассеяния случайных величин

К мерам рассеяния относят следующие характеристики распределения:

- дисперсию;
- среднее квадратическое отклонение (стандартное отклонение);
- размах вариации.

Дисперсия является основной характеристикой, характеризующей разброс значений распределения и вычисляется по формуле:

$$D_x = \frac{\sum_{i=1}^n (m_x - x_i)^2}{n},$$

где m_x – среднее значение распределения.

Пример 3. Вычислить меры рассеяния для выборки, представленной в примере 2.

Для вычисления дисперсии в ЭТ Excel используются следующие функции СУММКВ, СУММКВРАЗН и функции, непосредственно вычисляющие дисперсию.

Для использования функций СУММКВ и СУММКВРАЗН предварительно необходимо вычислить среднее значение распределения - m_x .

Для применения функции СУММКВ необходимо создать вспомогательный интервал со значениями разности $m_x - x_i$. Для формирования интервала разностей необходимо использовать абсолютную адресацию ячейки с вычисленным средним значением \$D\$10, иначе при заполнении интервала эта ячейка будет смещаться по рабочему листу.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		исходный интервал						интервал с разностями					
2		21	28	19	18	25		-12,5	-5,5	-14,5	-15,5	-8,5	
3		30	32	29	34	21		-3,5	-1,5	-4,5	0,5	-12,5	
4		32	23	19	36	32		-1,5	-10,5	-14,5	2,5	-1,5	
5		30	43	32	37	46		-3,5	9,5	-1,5	3,5	12,5	
6		38	29	38	41	42		4,5	-4,5	4,5	7,5	8,5	
7		33	41	38	47	44		-0,5	7,5	4,5	13,5	10,5	
8		42	35	32	44	43		8,5	1,5	-1,5	10,5	9,5	
9													
10			mx= 33,5							Dx= 66,9			
11										Dx= =СУММКВ(H2:L8)/СЧЁТ(H2:L8)			

Для применения функции СУММКВРАЗН необходимо создать интервал со средними значениями. Вспомогательный интервал по размерам должен соответствовать исходному интервалу. Для заполнения вспомогательного интервала средними значениями необходимо использовать команду **Специальная вставка**, а в ДО установить переключатель **Значения**.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
13		исходный интервал						интервал со средним							
14		21	28	19	18	25		33,5	33,5	33,5	33,5	33,5			
15		30	32	29	34	21		33,5	33,5	33,5	33,5	33,5			
16		32	23	19	36	32		33,5	33,5	33,5	33,5	33,5			
17		30	43	32	37	46		33,5	33,5	33,5	33,5	33,5			
18		38	29	38	41	42		33,5	33,5	33,5	33,5	33,5			
19		33	41	38	47	44		33,5	33,5	33,5	33,5	33,5			
20		42	35	32	44	43		33,5	33,5	33,5	33,5	33,5			
21															
22			mx= 33,5							Dx= 66,9					
23										Dx= =СУММКВРАЗН(B14:F20;H14:L20)/СЧЁТ(H14:L20)					

Дисперсию можно вычислить непосредственно, применяя функцию рабочего листа. В ЭТ Excel имеется несколько функций для вычисления дисперсии. В данном случае нужно выбирать функцию ДИСП.Г (генеральная дисперсия). Эта функция вычисляет дисперсию для всех значений интервала – по всей совокупности.

	A	B	C	D	E	F	G	H	I	J	K
26		исходный интервал									
27		21	28	19	18	25					
28		30	32	29	34	21		Dx= 66,9			
29		32	23	19	36	32		Dx= =ДИСП.Г(B27:F33)			
30		30	43	32	37	46					
31		38	29	38	41	42					
32		33	41	38	47	44					
33		42	35	32	44	43					

Среднее квадратическое отклонение (стандартное отклонение) вычисляется по формуле

$$s = \sqrt{D_x}.$$

В ЭТ Excel для этого используется функция КОРЕНЬ.

Dx=	66,9
s=	8,18
s=	=КОРЕНЬ(I3)

Для характеристики разброса значений распределения используют, как правило, стандартное отклонение, поскольку оно имеет ту же размерность, что и распределение. Дисперсию применяют для проверки статистических гипотез.

Размах вариации (диапазон) представляет собой разность между наибольшим и наименьшим значениями распределения. Если значения в распределении упорядочены по возрастанию, то диапазон есть разность между последним (с номером n) значением и первым (с номером 1):

$$R = x_n - x_1.$$

Если распределение не упорядочено, как в рассматриваемом примере, то можно воспользоваться функциями ЭТ Excel:

	A	B	C	D	E	F	G	H	I	J	K	L
1		исходный интервал										
2		21	28	19	18	25						
3		30	32	29	34	21		R=	29			
4		32	23	19	36	32		R=	=МАКС(B2:F8)-МИН(B2:F8)			
5		30	43	32	37	46						
6		38	29	38	41	42						
7		33	41	38	47	44						
8		42	35	32	44	43						

Размах вариации обычно вычисляется для построения интервального ряда распределения.

2.3. Меры формы распределения.

Меры распределения характеризуют распределение с геометрической точки зрения. Сюда относятся коэффициент асимметрии и эксцесс.

Коэффициент асимметрии характеризует скошенность распределения и вычисляется по формуле Пирсона:

$$A_s = \frac{m_x - Mo}{s_x}.$$

В зависимости от значения коэффициента определяется смещение модального значения в ту или иную сторону (рис.1.2).

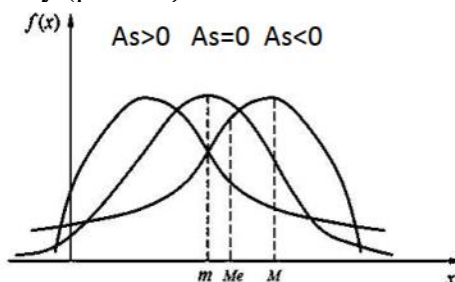


Рисунок 1.2 – Асимметрия распределения

Для вычисления коэффициента используем данные, вычисленные ранее:

Q	R
mx=	33,5
Mo=	33,0
s=	8,2
As=	0,066
As=	=(R1-R2)/R3

Коэффициент асимметрии гораздо меньше 1, поэтому можно считать распределение практически симметричным.

Коэффициент крутизны или эксцесс характеризует островершинность распределения. В зависимости от знака коэффициента распределение может быть, как плоско вершинным, так и островершинным. Если коэффициент равен нулю, то распределение считается нормальным (рис.1.2).

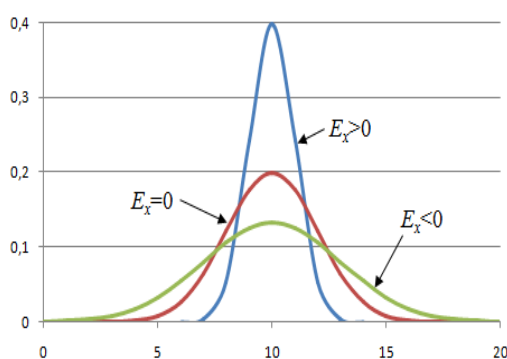


Рисунок 1.3 – Эксцесс распределения

Эксцесс вычисляем с помощью функции ЭТ Excel ЭКСЦЕСС. В качестве аргумента функции используется исходный интервал:

E=	-0,82788
E=	=ЭКСЦЕСС(B2:F8)

Рисунок 2.8 – Вычисление эксцесса распределения.

Эксцесс отрицательный, значит можно говорить о плосковершинности распределения.

2.4 Описательная статистика

ЭТ Excel предоставляет пользователю вычислить основные характеристики распределения СВ на основе использования **Пакета анализа**.

Чтобы воспользоваться возможностями **Пакета анализа** необходимо в меню выбрать команду **Данные**, в разделе **Анализ** должна быть команда **Анализ данных**.

Если в меню **Данные** анализ данных отсутствует, необходимо его загрузить. Для этого:

1. Открыть меню **Файл**.
2. Выбрать команду **Параметры**.
3. Выбрать меню **Надстройки**.
4. В списке Неактивные надстройки приложений выбрать **Пакет анализа**.
5. Кнопка **Перейти ...**
6. В ДО **Надстройки** установить флажок **Пакет анализа**. ОК.

Когда **Пакет анализа** загрузится, выбрать команду **Анализ данных**. В ДО **Анализ данных** из списка выбрать **Описательная статистика**. Заполнить ДО. Для этого:

- указать исходный анализируемый интервал с данными; если анализ осуществляется по всему интервалу, то интервал должен быть представлен либо одним столбцом, либо одной строкой; иногда бывает необходимость вычислить статистики одновременно для нескольких распределений, тогда указывают, тогда указывают, как сгруппированы данные, устанавливая соответствующий переключатель;
- если вместе с данными выделяются их метки, также нужно установить соответствующий флажок – указать начальную ячейку выходного интервала;
- установить флажок **Итоговая статистика**. В результате на рабочий лист выводятся статистик:

	I	J	K	L
1	18,187			
2	18,571			
3	18,947	Среднее		33,4253
4	20,656	Стандартная ошибка		1,40378
5	20,720	Медиана		33,3065
6	23,429	Мода		#Н/Д
7	24,661	Стандартное отклонение		8,30487
8	27,995	Дисперсия выборки		68,9709
9	28,788	Экссесс		-0,8119
10	29,208	Асимметричность		-0,2839
11	29,696	Интервал		28,7663
12	30,150	Минимум		18,1873
13	31,511	Максимум		46,9536
14	31,589	Сумма		1169,89
...	...	Счет		35

Для нескольких переменных с метками результат будет представлен для каждой из переменных:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1								Var1		Var2		Var3		Var4		Var5	
2																	
3								Средне	32,286	Средне	33	Средне	29,571	Средне	36,714	Средне	36,143
4								Стандар	2,5139	Стандар	2,7168	Стандар	2,9989	Стандар	3,5705	Стандар	3,8137
5		Var1	Var2	Var3	Var4	Var5		Медиан	32	Медиан	32	Медиан	32	Медиан	37	Медиан	42
6		21	28	19	18	25		Мода	30	Мода	#Н/Д	Мода	19	Мода	#Н/Д	Мода	#Н/Д
7		30	32	29	34	21		Стандар	6,6512	Стандар	7,188	Стандар	7,9343	Стандар	9,4466	Стандар	10,09
8		32	23	19	36	32		Диспер	44,238	Диспер	51,667	Диспер	62,952	Диспер	89,238	Диспер	101,81
9		30	43	32	37	46		Экссесс	0,8592	Экссесс	-1,0306	Экссесс	-1,2619	Экссесс	2,6318	Экссесс	-1,5761
10		38	29	38	41	42		Асимме	-0,2836	Асимме	0,2073	Асимме	-0,5311	Асимме	-1,3897	Асимме	-0,6725
11		33	41	38	47	44		Интерв	21	Интерв	20	Интерв	19	Интерв	29	Интерв	25
12		42	35	32	44	43		Миним	21	Миним	23	Миним	19	Миним	18	Миним	21
13								Максим	42	Максим	43	Максим	38	Максим	47	Максим	46
14								Сумма	226	Сумма	231	Сумма	207	Сумма	257	Сумма	253
15								Счет	7	Счет	7	Счет	7	Счет	7	Счет	7

Следует обратить внимание, что дисперсия и стандартное отклонение в Описательной статистике вычисляются для выборочных данных. Если необходимо вычислить генеральную дисперсию, то нужно использовать функцию ДИСП.Г.

3. Задание для выполнения практической работы.

1. Выполнить **примеры 1-3** и сравнить результаты с представленными в данных методических указаниях.

2. **Задание 1.** Задан ряд распределения дискретной случайной величины. Построить функцию распределения СВ и многоугольник частот, определить меры центральной тенденции - среднее и модальное значения.

Индивидуальные варианты для выполнения задания 1 приведены в Приложении А.1. Номер варианта совпадает с номером студента в списке группы.

3. **Задание 2.** Задан ряд распределения дискретной случайной величины. Построить функцию распределения СВ и многоугольник частот, определить меры центральной тенденции - среднее и модальное значения.

Индивидуальные варианты для выполнения задания 2 приведены в Приложении А.2. Номер варианта совпадает с номером студента в списке группы.

3. **Задание 3.** Заданы значения непрерывной СВ. Вычислить числовые характеристики СВ.

Индивидуальные варианты для выполнения задания 3 приведены в Приложении А.3.

4. **Задание 4.** Для непрерывной СВ (задание 2) вычислить итоговые характеристики на основе Пакета анализа табличного процессора.

4 Содержание отчета.

1. Титульный лист.
2. Цель работы.
3. Постановка задачи.
4. Порядок выполнения работы.
5. Анализ результатов и выводы.

Практическая работа № 2

Интервальный статистический ряд. Частотный анализ

Цель: получить навыки статистической обработки выборочных данных, ознакомиться с функциями и графическими средствами ЭТ Excel, применяемыми для статистической обработки данных.

1. КРАТКИЕ ТЕОРЕТИЧЕСКИЕ СВЕДЕНИЯ

Математическая статистика — это наука, разрабатывающая методы обработки опытных данных, полученных в результате наблюдений над случайными явлениями.

Основу статистического исследования составляет множество данных, полученных по результатам измерения одного или нескольких признаков. Результаты наблюдений или измерений при этом интерпретируются совокупностью значений, принятых в результате n опытов некоторой случайной величиной X . Наблюдаемые значения $x_1, x_2, \dots, x_n$ называют выборкой наблюдений объема n из генеральной совокупности.

Статистическая обработка данных выполняется поэтапно. Основные этапы статистической обработки данных приводятся ниже.

1 этап. Выборочные данные представляются в виде вариационного ряда, ($x_1=x_{min}, x_2, x_3, \dots, x_n=x_{max}$) т.е. упорядочиваются в порядке возрастания. Далее определяются значения минимальной и максимальной вариант x_{min} и x_{max} и вычисляется размах вариации:

$$R = x_{max} - x_{min}. \quad (2.1)$$

2. По упорядоченной статистической совокупности строится статистическая функция распределения. Статистической функцией распределения называется функция $F^*(x)$, определяющая для каждого выборочного значения x относительную частоту события $\{X < x\}$. Тогда по определению

$$F^*(x) = l_x / n,$$

где l_x - число наблюдений в выборке, меньших x . Если x изменяется, то изменяется и относительная частота, т.е. относительная частота l_x / n есть функция от x . Функция $F^*(x)$ имеет число скачков равное числу различных выборочных значений. Величина скачка равна «кратности» l_i значения x_i в выборке, деленной на n . Функция $F^*(x)$ записывается в аналитической форме

3. Построение группированного статистического ряда. Для этого необходимо весь участок оси абсцисс, на котором расположены выборочные значения случайной величины X , разделить на интервалы или карманы. Число интервалов k обычно определяется эмпирическим путем. В частности, $k \approx \sqrt{n}$, если объем выборки $n \leq 100$. При этом длина одного интервала вычисляется по формуле:

$$h = \frac{R}{k}. \quad (2.2)$$

Далее вычисляются границы интервалов $x_{i+1} = x_i + h$ и v_i^* - статистическая абсолютная частота (количество значений переменной в каждом интервале). В соответствии с выражением (3.3) определяется p_i^* - статистическая относительная частота

$$p_i^* = \frac{v_i^*}{n}, \quad (2.3)$$

где n – объем выборки.

В результате получаем группированный статистический ряд (табл. 3.1), представляет собой частотное распределение.

Таблица 2.1 - Группированный статистический ряд представляется в таблице вида

X	$x_1 \div x_2$	$x_2 \div x_3$	...	$x_i \div x_{i+1}$	...	$x_{k-1} \div x_k$
N	ν_1^*	ν_2^*	...	ν_i^*	...	ν_k^*
P	p_1^*	p_2^*	...	p_i^*	...	p_k^*

По данным группированного статистического ряда можно построить эмпирическую функцию распределения и графически представить частотное распределение случайной величины X в виде гистограммы.

2. Пример выполнения задания.

Наблюдается время, затраченное на дорогу от дома до университета в течении 20 дней (рис.2.1):

№	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
T	31	41	49	54	38	53	39	47	38	43	44	38	37	43	30	47	36	36	38	42

Рисунок 2.1 – Результаты наблюдений

Требуется провести частотный анализ и описать закон распределения выборки.

Порядок выполнения задания.

1. Ряд наблюдений упорядочивается по возрастанию и определяются максимальное и минимальное значения переменной (рис.2.2):

T	30	31	36	36	37	38	38	38	38	39	41	42	43	43	44	47	47	49	53	54
---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

Рисунок 2.2 - Вариационный ряд

2. Определяется размах вариации:

$$R = x_{\max} - x_{\min} = 54 - 30 = 24.$$

3. Определяется количество интервалов: $k \approx 5$.

4. Вычисляется длина интервала $h = R/k = 24/5 = 4,8$.

5. Вычисляются границы интервалов $x_1 = x_{\min}, \dots, x_{i+1} = x_i + h, \dots, x_{k+1} = x_{\max}, i = 1, k$:

№ инт	x_n	x_{ε}
1	30	34,8
2	34,8	39,6
3	39,6	44,4
4	44,4	49,2
5	49,2	54

Рисунок 2.3 – Границы интервалов

6. Подсчитывается количество переменных, попадающих в каждый интервал – абсолютная частота (рис. 2.4):

30	31	36	36	37	38	38	38	38	39	41	42	43	43	44	47	47	49	53	54
1 инт		2 инт								3 инт					4 инт			5 инт	
2		8								5					3			2	

Рисунок 2.4 – Абсолютные частоты

7. Абсолютные частоты записываются в таблицу 2.2 и следующим столбцом, в эту же таблицу, заносятся значения статистической вероятности, вычисленные в соответствии с выражением (2.3).

Таблица 2.2 – Частотное распределение выборки

№ инт	x_n	x_e	v	p^*
1	30	34,8	2	0,1
2	34,8	39,6	8	0,4
3	39,6	44,4	5	0,25
4	44,4	49,2	3	0,15
5	49,2	54	2	0,1
				1

Сумма абсолютных частот должна быть равна объему выборки, а сумма вероятностей – 1.

8. Графическая интерпретация частотного распределения выборки представлена гистограммой и эмпирической функцией распределения. Для построения гистограммы можно использовать как значения абсолютных частот, так и статистической вероятности. Для построения графика функции распределения необходимо заранее вычислить ее значения.

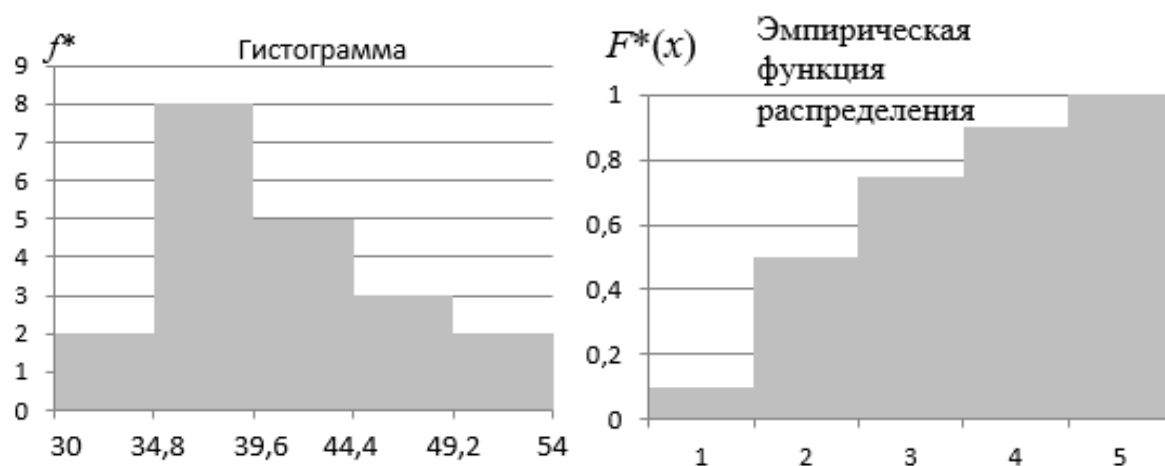


Рисунок 2.5 – Графическая интерпретация частотного анализа

Часто в статистике для представления частотного распределения используют круговую диаграмму. Ее можно построить как по значениям абсолютных частот, так и по статистической вероятности.

9. Частотный анализ в ЭТ Excel.

Для проведения частотного анализа в ЭТ Excel необходимо предварительно подготовить данные. На рабочем листе нужно разместить выборку с наблюдениями и интервал карманов. Интервал карманов представляет собой массив с верхними границами интервалов (рис. 2.6). Поскольку верхняя граница последнего интервала не включает максимальное значение, оно будет включено в последующий интервал. Поэтому верхнюю границу нужно увеличить на любое, сколь угодно малое число, пусть это будет 0,01. В этом случае максимальное значение выборки попадет в последний интервал (рис. 2.6):

	A	B	C	D	E
1	№	T			
2	1	30	Xmin=	30	
3	2	31	xmax=	54	
4	3	36	R=	24	
5	4	36	h=	4,8	
6	5	37			
7	6	38	xв		
8	7	38		34,8	
9	8	38		39,6	
10	9	38		44,4	
11	10	39		49,2	
12	11	41		54,01	
13	12	42			

Рисунок 2.6 – Исходные данные для проведения частотного анализа в ЭТ Excel

Далее необходимо выбрать команды: **Данные. Анализ данных. Гистограмма** и заполнить соответствующее диалоговое окно.

В диалоговом окне указать:

- **Входной интервал** – выделить интервал с выборкой по данным в примере B2:B21;
- **Интервал карманов** – выделить интервал с верхними границами карманов D8:D12;
- **Параметры вывода**

- установить переключатель **Выходной интервал** и в поле ввода указать ячейку на свободном месте рабочего листа G2,

- установить флажки **Интегральный процент**, **Вывод графика**. **ОК**

В результате на рабочем листе (рис.2.7) будет представлена таблица с абсолютными частотами и графики эмпирической функции распределения (интегральный %) и гистограммы.

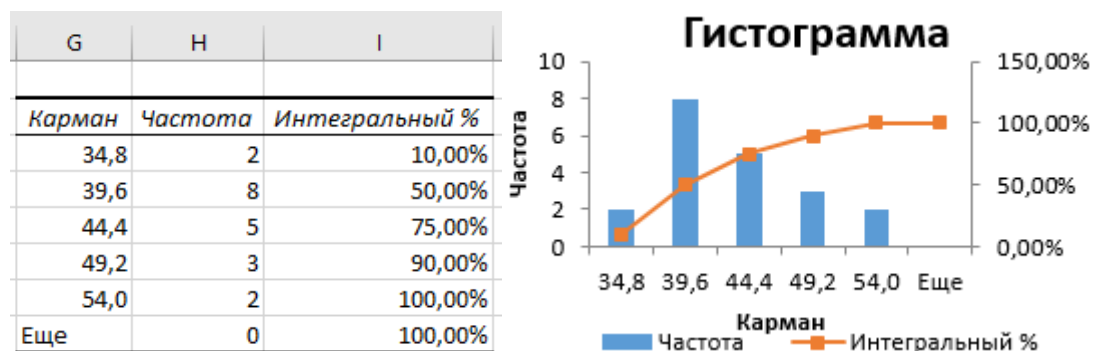


Рисунок 2.7 – Частотный анализ в ЭТ Excel

При желании диаграмму можно редактировать и форматировать, придавая ей желаемый вид. Можно сделать копию построенной диаграммы и, удалив один из рядов, изобразить гистограмму и интегральный процент на разных диаграммах.

3. Задание для выполнения практической работы.

1. Выполнить задание, представленное в п.2 и удостовериться в правильности расчетов.

2. Выбрать из Приложения Б индивидуальный вариант задания в соответствии с номером в списке группы. Требуется провести частотный анализ и описать закон распределения выборки.

4. Содержание отчета

1. Постановка задачи.
2. Ход выполнения задания с пояснениями.
3. Анализ результатов и выводы.

Практическая работа № 3

Статистические оценки параметров распределения

Цель: получить навыки оценки параметров генеральной совокупности по выборочным данным средствами ЭТ Excel.

1. КРАТКИЕ ТЕОРЕТИЧЕСКИЕ СВЕДЕНИЯ

Статистической оценкой $\hat{\Theta}$ параметра Θ распределения называется приближенное значение параметра, вычисленное по результатам эксперимента (по выборке). Статистические оценки делят на точечные и интервальные.

Точечной называется оценка, определяемая одним числом. Точечные оценки должны отвечать следующим требованиям:

1. Оценка $\hat{\Theta}$ называется **состоятельной**, если при увеличении объема выборки n она сходится к значению параметра Θ .

2. Оценка $\hat{\Theta}$ называется **несмещенной**, если ее среднее точно равно параметру Θ для любого объема выборки.

3. Несмещенная оценка $\hat{\Theta}$ является **эффективной**, если ее дисперсия минимальна по отношению к дисперсии любой другой оценки этого параметра: $D[\hat{\Theta}] = \min$.

Оценка генеральной средней по выборочной средней. Из генеральной совокупности (ГС) влечена выборка объема n со значениями признака $x_1, x_2, \dots, x_n$.

Пусть генеральная средняя $\bar{x}_Г$ неизвестна и требуется оценить ее по данным выборки.

Тогда в качестве оценки генеральной средней принимают выборочную среднюю:

$$\bar{x}_B = \frac{1}{n} \sum_{i=1}^n x_i. \quad (3.1)$$

В ЭТ Excel для оценки генеральной средней используется функция СРЗНАЧ.

Генеральная дисперсия характеризует рассеяние значений количественного признака X ГС вокруг своего среднего значения $\bar{x}_Г$. **Генеральной дисперсией** $D_Г$ называют среднее арифметическое квадратов отклонений значений признака ГС от их среднего значения и вычисляют по формуле:

$$D_Г = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_Г)^2. \quad (3.2)$$

Пусть требуется по данным выборки оценить неизвестную генеральную дисперсию $D_Г$. Выборочная дисперсия D_B является смещенной оценкой генеральной $D_Г$. Если в качестве оценки генеральной дисперсии принять D_B , то эта оценка будет приводить к систематическим ошибкам, давая заниженное значение $D_Г$. Поэтому для оценки генеральной дисперсии по выборке используют несмещенную оценку:

$$D_B = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_B)^2. \quad (3.3)$$

В ЭТ Excel для оценки генеральной дисперсии используется функция ДИСП.В.

Среднее квадратическое отклонение (стандартное отклонение) генеральной совокупности можно оценить как:

$$s_B = \sqrt{D_B}. \quad (3.4)$$

В ЭТ Excel для оценки генерального стандартного отклонения используется функция СТАНДОТКЛОН.В.

Интервальные оценки числовых характеристик. Пусть для параметра Θ получена из опыта несмещенная оценка $\hat{\Theta}$. Оценим возможную ошибку, возникающую при замене

параметра Θ его оценкой $\hat{\Theta}$. Зададимся достаточно большой вероятностью γ , и найдем такое значение ε , для которого

$$P(|\hat{\Theta} - \Theta| < \varepsilon) = \gamma. \quad (3.5)$$

Диапазон практически возможных значений ошибки, возникающей при замене Θ на $\hat{\Theta}$, будет равен $\pm \varepsilon$.

Большие по абсолютной величине ошибки будут появляться только с малой вероятностью $\alpha = 1 - \gamma$.

Равенство (4.4) означает, что с вероятностью γ неизвестное значение параметра Θ попадает в интервал

$$I_\gamma = (\hat{\Theta} - \varepsilon; \hat{\Theta} + \varepsilon). \quad (3.6)$$

Выражение (4.5) представляет собой доверительный интервал.

Итак, **доверительным** называется интервал I_γ , в который с заданной вероятностью (**надежностью**) γ попадают значения параметра Θ . Вероятность γ выбирается близкой к единице: 0,9; 0,95; 0,975; 0,99. Если объем выборки не превышает

Если объем выборки не превышает $n \leq 30$, то доверительный интервал для **среднего** с надежностью γ определяется на основании вычисления t-статистики Стьюдента:

$$\bar{x}_B - \frac{s \cdot t_{\gamma, n-1}}{\sqrt{n}} < \bar{x}_T < \bar{x}_B + \frac{s \cdot t_{\gamma, n-1}}{\sqrt{n}}, \quad (3.7)$$

где n – объем выборки; s – стандартное отклонение; $n-1$ – степени свободы, $t_{\gamma, n-1}$ – табличное критическое значение распределения Стьюдента (t-статистика Стьюдента) выбирается из соответствующих таблиц, либо вычисляется с применением функции ЭТ Excel СТЬЮДРАСПОБР с аргументами вероятность $1-\gamma$ и $n-1$. Ширину интервала можно вычислить, используя функцию ЭТ Excel ДОВЕРИТ.СТЮДЕНТ, аргументами функции являются вероятность $1-\gamma$, стандартное отклонение и объем (размер) выборки).

Если объем выборки $n > 30$, то для вычисления ширины интервала используют функцию Лапласа. В ЭТ Excel используется функция ДОВЕРИТ.НОРМ с теми же аргументами, что и в функции ДОВЕРИТ.СТЮДЕНТ.

2. Задание для выполнения практической работы.

Материалы для выполнения заданий содержатся по ссылке:

<https://do.sevsu.ru/mod/assign/view.php?id=514619>

1. **Задание 1.** Задана генеральная совокупность (Лист 1 файла **ПР_3.xlsx**). Из генеральной совокупности получить 2 выборки объемом $n=20$ и $n=40$. По каждой из выборок определить:

- Точечные оценки генеральной совокупности - среднее, дисперсию и стандартное отклонение.

- Интервальную оценку среднего с надежностью 0,95.

Рекомендации: задания выполнить для примера, представленного в методических указаниях и для выборок, полученных самостоятельно.

2. **Задание 2.** Задана выборка (Практическая работа №2, Приложение 2.А). Вычислить точечные и интервальные характеристики выборки.

3. **Задание 3.** Задана генеральная совокупность (Лист 2 файла **ПР_3.xlsx**). Из генеральной совокупности получить выборку объемом $n=40$ и вычислить интервальную оценку среднего с надежностью 0,95.

3. Пример выполнения задания.

Из генеральной совокупности (файл ПР_3) получить случайную выборку можно, используя Пакет анализа ЭТ Excel. Для этого выполнить команды **Данные. Анализ данных**, далее из списка в диалоговом окне выбрать **Выборка** и установить параметры в диалоговом окне:

- В поле **Входной интервал** указать интервал с генеральной совокупностью.
- Установить переключатель **Случайный**.
- В поле **Число выборок** ввести объем выборки, например, 20.
- Установить переключатель **Выходной интервал** и в поле ввода указать одну ячейку на свободном месте рабочего листа. ОК.

Итак, исходные данные сформированы и можно начинать выполнять задание.

1. Используя выражения (4.1)-(4.4) оценить по выборке параметры генеральной совокупности:

$$\text{ср.Выб.} = 41,7 \quad \text{дисп. Выб.} = 154 \quad s \text{ выб.} = 12,4$$

2. Используя выражения (4.4) - (4.7) и функции ЭТ Excel построить для генеральной средней интервальную оценку (рис. 3.1):

n=	20	n=	20
γ =	0,95	γ =	0,95
s=	12,41	s=	12,41
$t_{\gamma, n-1}$ =	2,09	ϵ =	5,81
ср.В=	41,7	ϵ =	<code>=ДОВЕРИТ.СТЮДЕНТ(0,05;Y27;Y25)</code>
ϵ =	5,81		<code>ДОВЕРИТ.СТЮДЕНТ(альфа; станд_откл; размер)</code>
ϵ =	<code>=V26*V27/КОРЕНЬ(V24)</code>		

а)

б)

Рисунок 3.1 – Вычисление доверительного интервала для генеральной средней:

а) на основании расчетов; б) с использованием функций ЭТ Excel

Ширина доверительного интервала в обоих случаях одинакова и равна 5,81. Теперь на основании (4.6) можно вычислить границы интервала $I_{0,95}=(37,64; 44,07)$.

Таким образом, интервальная оценка генеральной средней равна $35,86 < \bar{x}_T < 47,47$ при уровне надежности 0,95.

4. Содержание отчета

1. Постановка задачи.
2. Ход выполнения задания с пояснениями.
3. Анализ результатов и выводы.

Практическая работа № 4

Критерии согласия

Цель: получить навыки проверки гипотез о нормальном законе распределения выборочных данных средствами ЭТ Excel.

1. КРАТКИЕ ТЕОРЕТИЧЕСКИЕ СВЕДЕНИЯ

Критериями согласия называются критерии, используемые для проверки гипотез о предполагаемом законе распределения.

В ходе выполнения практической работы рассматриваются критерии Пирсона (χ^2) и Колмогорова-Смирнова.

1.1 Алгоритм вычисления критерия согласия Пирсона (χ^2).

1. Построить интервальный статистический ряд и гистограмму.
2. По виду гистограммы выдвинуть гипотезу:
 - H_0 – СВ X распределена по **нормальному** закону: $f(x)=f_0(x)$,
 - H_1 – СВ X не распределена по **нормальному** закону: $f(x) \neq f_0(x)$, $f_0(x)$ – плотность гипотетического закона распределения.
3. Определить оценки $\hat{\theta}_1, \dots, \hat{\theta}_m$ неизвестных параметров *гипотетического* закона распределения. Для нормального распределения это среднее выборочное и стандартное отклонение.
4. Вычислить значение критерия по формуле:

$$\chi^2 = \sum_{i=1}^k \frac{(v_i^* - v_i)^2}{v_i} = n \cdot \sum_{i=1}^k \frac{(p_i^* - p_i)^2}{p_i}, \quad (4.1)$$

где p_i – теоретическая вероятность попадания случайной величины в i -й интервал при условии, что гипотеза H_0 верна.

При вычислении вероятностей попадания случайной величины для первого и последнего интервалов x_{min} и x_{max} следует использовать теоретические границы гипотетического распределения. Для нормального закона $x_{min} = -\infty$, а $x_{max} = \infty$. Для всех внутренних интервалов ряда верхняя и нижняя границы, полученные эмпирическим путем, совпадают. После вычисления теоретических частот полезно проверить контрольное соотношение:

$$|1 - \sum_{i=1}^k p_i| \leq 0,01. \quad (4.2)$$

5. Из таблицы распределения χ^2 выбирается значение $\chi^2(\alpha, r)$. Распределение χ^2 зависит от параметра r – числа степеней свободы и уровня значимости α . Количество степеней свободы равно $r = k - 1 - m$.

Здесь m – число неизвестных параметров гипотетического закона распределения для нормального закона $m=2$, k – количество интервалов.

6. Если значение, вычисленное по формуле (5.1), больше, чем критическое значение, т.е.

$$\chi^2 > \chi^2(\alpha, r), \quad (4.3)$$

то гипотеза H_0 отклоняется, в противном случае нет оснований ее отклонить.

Определение теоретических частот для нормального закона распределения $N(m_x, \sigma_x)$.

Для вычисления теоретических частот можно использовать функцию ЭТ Excel НОРМРАСП, аргументы которой (рис. 4.1):

- x – значение, для которого определяется вероятность;
- среднее – выборочное среднее;
- стандартное отклонение – определяется по выборке;

- ИСТИНА – используется интегральная функция.

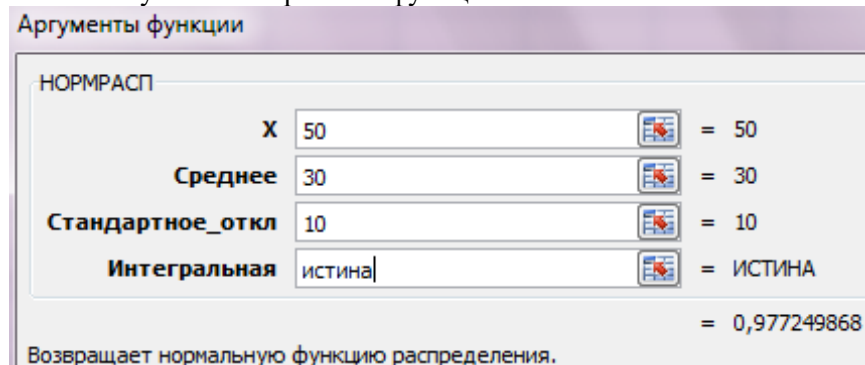


Рисунок 4.1– Диалоговое окно функции НОРМРАСП

Например, необходимо вычислить вероятность попадания в интервал $P[10, 50]$ случайной величины, распределенной по нормальному закону с параметрами $N(30, 10)$.

Верхняя граница интервала – 50, а нижняя – 10. Вероятность равна разности $P(50)-P(10)$.

=НОРМРАСП(50;30;10;ИСТИНА)	0,97725
НОРМРАСП(х; среднее; стандартное_откл; интегральная)	
=НОРМРАСП(10;30;10;ИСТИНА)	0,02275
НОРМРАСП(х; среднее; стандартное_откл; интегральная)	

Тогда $P[10, 50] = P(50)-P(10) = 0,9545$.

1.2 Алгоритм вычисления критерия согласия Колмогорова-Смирнова (D)

1. Выполнить частотный анализ выборки (рис. 4.2) и вычислить значения эмпирической функции распределения $F^*(x)$.

2. Выдвинуть гипотезу:

- $H_0: F^*(x)=F_0(x)$ – эмпирическая функция распределения выборки совпадает с функцией распределения нормального закона;
- $H_1: F^*(x) \neq F_0(x)$ эмпирическая функция распределения выборки не соответствует нормальному закону. $F_0(x)$ - ФР нормального закона распределения.

3. Определить оценки среднего и стандартного отклонения выборки.

4. Рассчитать значения функции $F_0(x)$ для каждой СВ в выборке и построить ее график в одной системе координат с функцией $F^*(x)$.

5. Определить максимальное по модулю отклонение между функциями $F^*(x)$ и $F_0(x)$:

$$D = \max_{i=1}^n |F^*(x_i) - F_0(x_i)|, \quad (4.4)$$

максимальная абсолютная разность значений теоретической и наблюдаемой функций распределения используется в качестве критерия.

6. В таблице «**Критические точки для статистики Колмогорова D**» (Приложение В) по уровню значимости α и объему выборки n выбирается критическое значение.

7. Если

$$D_{расч} < D_{крит}(\alpha, n), \quad (4.5)$$

то гипотеза H_0 о соответствии выборки нормальному закону распределения принимается. В противном случае принимается альтернативная гипотеза H_1 .

2. Задание для выполнения практической работы.

1. **Задание 1.** Проверить гипотезу о нормальном распределении на основе критерия согласия Пирсона для выборочной совокупности (практическая работа № 2, Приложение Б) при уровне значимости $\alpha=0,05$.

2. **Задание 2.** Проверить гипотезу о нормальном распределении на основе критерия Колмогорова для выборочной совокупности (практическая работа № 2, Приложение Б) при уровне значимости $\alpha=0,05$.

Рекомендации: задания выполнить для примера, рассмотренного в данных методических указаниях (п.3) и в соответствии со своим вариантом.

3. Пример выполнения задания.

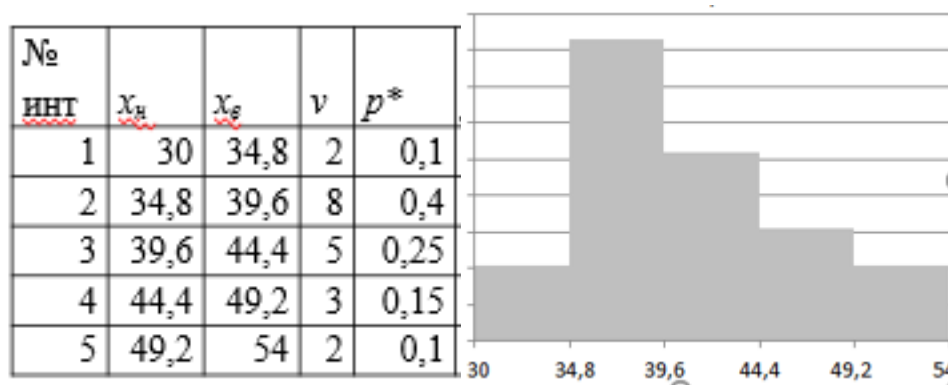
Имеется выборка

№	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
T	31	41	49	54	38	53	39	47	38	43	44	38	37	43	30	47	36	36	38	42

Проверить соответствие выборки нормальному закону распределения.

3.1 Критерий согласия Пирсона.

Шаг 1. Построить интервальный статистический ряд и гистограмму



Шаг 2. По виду гистограммы выдвинуть гипотезу:

H_0 – СВ X распределена по нормальному закону,

H_1 - СВ X не распределена по нормальному закону.

Шаг 3. Определить оценки математического ожидания и СКО: $\bar{x}_B = 41,011$; $s = 6,5$.

Шаг 4. Для каждого интервала вычислить значения теоретических частот:

Для первого интервала: $p_1 = \text{НОРМРАСП}(x_2; x_{cp}; s; \text{ИСТИНА}) - 0$

	A	B	C	D	E	F	G	H
1	хср выб=	41,011						
2	с выб=	6,52						
3	№ инт	хн	хк	v	p*	p	χ ² расч	
4	1	30	34,8	2	0,1	=НОРМРАСП(C4;B1;B2;ИСТИНА)		
5	2	34,8	39,6	8	0,4	НОРМРАСП(x; среднее; стандартн		
6	3	39,6	44,4	5	0,25	НОРМРАСП(x; среднее; стандартн		
7	4	44,4	49,2	3	0,15	НОРМРАСП(x; среднее; стандартн		
8	5	49,2	54	2	0,1	НОРМРАСП(x; среднее; стандартн		

Для последнего интервала: $p_k = 1 - \text{НОРМРАСП}(x_k; x_{cp}; s; \text{ИСТИНА})$

	A	B	C	D	E	F	G	H
1	хср выб=	41,011						
2	с выб=	6,52						
3	№ инт	хн	хв	v	p*	p	χ ² расч	
4	1	30	34,8	2	0,1	0,170395	0,029082	
5	2	34,8	39,6	8	0,4	0,243939	0,09984	
6	3	39,6	44,4	5	0,25	0,28406	0,004084	
7	4	44,4	49,2	3	0,15	0,197045	0,011232	
8	5	49,2	54	2	0,1	=1-НОРМРАСП(B8;B1;B2;ИСТИНА)		
9						НОРМРАСП(х; среднее; стандартн		

Для внутренних интервалов:

$p_i = \text{НОРМРАСП}(\chi_{i+1}; x_{cp}; s; \text{ИСТИНА}) - \text{НОРМРАСП}(\chi_i; x_{cp}; s; \text{ИСТИНА}), i=1, k.$

	A	B	C	D	E	F	G	H	I	J	K	L
1	хср выб=	41,011										
2	с выб=	6,52										
3	№ инт	хн	хв	v	p*	p	χ ² расч					
4	1	30	34,8	2	0,1	0,170395	0,029082					
5	2	34,8	39,6	8	0,4	=НОРМРАСП(C5;B\$1;B\$2;ИСТИНА)-НОРМРАСП(B5;B\$1;B\$2;ИСТИНА)						
6	3	39,6	44,4	5	0,25	НОРМРАСП(х; среднее; стандартное_откл; интегральная)						

В результате получим:

№ инт	хн	хв	v	p*	p
1	30	34,8	2	0,1	0,170395
2	34,8	39,6	8	0,4	0,243939
3	39,6	44,4	5	0,25	0,28406
4	44,4	49,2	3	0,15	0,197045
5	49,2	54	2	0,1	0,104561
					1

Для вычисленных теоретических частот проверить условие (5.2). В рассматриваемом примере все выполнено правильно, сумма теоретических частот равна 1.

Шаг 5. В соответствии с выражением (5.1) вычислить значение критерия Пирсона (Хи-квадрат). Расчетное значение критерия равно $\chi^2_{\text{расч}}=0,7222$.

№ инт	хн	хв	v	p*	p	норм
1	30	34,8	2	0,1	0,170395	0,029082
2	34,8	39,6	8	0,4	0,243939	0,09984
3	39,6	44,4	5	0,25	0,28406	0,004084
4	44,4	49,2	3	0,15	0,197045	0,011232
5	49,2	54	2	0,1	0,104561	0,000199
					1	0,722187

Критическое значение критерия вычисляем используя функцию ЭТ Excel ХИ2ОБР:

$\chi^2(0,05;2) = 6$
$\chi^2(0,05;2) = 5,99146455$
$\chi^2(0,05;2) = \text{ХИ2ОБР}(0,05;2)$
ХИ2ОБР(вероятность; степени_свободы)

Степень свободы (см. п.5 алгоритма) равна $r=k-1-m=5-1-2=2$.

Поскольку выполняется соотношение (4.3) нулевую гипотезу о нормальном распределении выборки принимаем при уровне значимости $\alpha=0,05$.

3.2 Критерий согласия Колмогорова.

Шаг 1. Выполнить частотный анализ выборки (рис. 4.2) и вычислить значения эмпирической функции распределения $F^*(x)$.

T	30	31	36	37	38	39	41	42	43	44	47	49	53	54	
ni	1	1	2	1	4	1	1	1	2	1	2	1	1	1	20
pi	0,05	0,05	0,1	0	0,2	0,05	0,05	0,1	0,1	0,05	0,1	0,05	0,05	0,05	1
F*(x)	0,05	0,1	0,2	0	0,45	0,5	0,55	0,6	0,7	0,75	0,85	0,9	0,95	1	

Рисунок 4.2 – Фрагмент рабочего листа с расчетами

Здесь n_i – абсолютные частоты значений выборки, p_i – относительные частоты значений выборки.

Шаг 2. Выдвинуть гипотезу:

- $H_0: F^*(x) = F_0(x)$ – эмпирическая функция распределения выборки совпадает с функцией распределения нормального закона;
- $H_1: F^*(x) \neq F_0(x)$ эмпирическая функция распределения выборки не соответствует нормальному закону. $F_0(x)$ – ФР нормального закона распределения.

Шаг 3. Определить оценки параметров нормального закона распределения – среднего и стандартного отклонения выборки:

Используя выражения (3.1)-(3.4) практической работы №1.3 получить значения оценок: выборочное среднее равно 41,011, а среднее квадратическое отклонение (стандартное отклонение) – 6,52.

Шаг 4. Рассчитать значения функции $F_0(x)$ для каждой СВ в выборке, используя функцию НОРМРАСП:

выборка	$F^*(x)$	$F_0(x)$	Z
30	0,05	=НОРМРАСП(B63;41,011;6,52;ИСТИНА)	
31	0,1	(НОРМРАСП(x; среднее; стандартное_откл; интегральная)	
36	0,2	0,236	0,0365

Шаг 5. Определить максимальное по модулю отклонение между функциями $F^*(x)$ и $F_0(x)$ на основании выражения (4.4):

выборка	$F^*(x)$	$F_0(x)$	D
30	0,05	0,046	=ABS(C63-D63)
31	0,1	0,065	ABS(число)
36	0,2	0,236	0,0365

В столбце D выбрать максимальное значение. Это и будет расчетное значение критерия согласия Колмогорова (рис.4.3) $D_{расч} = 0,1211$

x	$F^*(x)$	$F_0(x)$	D
30	0,05	0,046	0,0044
31	0,1	0,065	0,0345
36	0,2	0,236	0,0365
37	0,25	0,287	0,0373
38	0,45	0,343	0,1067
39	0,5	0,379	0,1211
41	0,55	0,472	0,0777
42	0,6	0,56	0,0396
43	0,7	0,627	0,0726
44	0,75	0,671	0,0789
47	0,85	0,817	0,0325
49	0,9	0,899	0,0008
53	0,95	0,965	0,0151
54	1	0,98	0,0196

Функции распределения

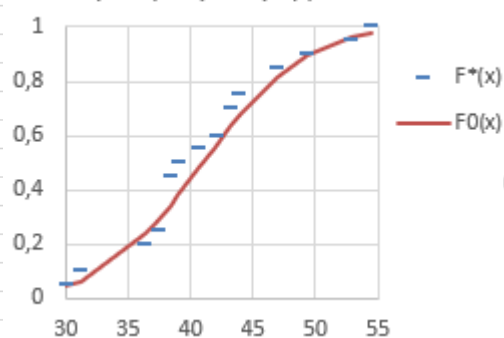


Рисунок 4.3 – Расчеты и графики

Шаг 6. Выбрать критическое значение D по уровню значимости $\alpha=0,05$ и объему выборки $n=20$ из таблицы Приложения В: $D_{крит}=0,29$.

Шаг 7. Поскольку $D_{расч}=0,1211$ меньше $D_{крит}=0,29$, нулевую гипотезу о соответствии выборки нормальному закону принимаем.

4. Содержание отчета

1. Постановка задачи.
2. Ход выполнения задания с пояснениями.
3. Анализ результатов и выводы.

Практическая работа № 5

Проверка гипотезы об однородности дисперсий

Цель: получить навыки проверки гипотезы об однородности дисперсий на основе параметрических и непараметрических критериев.

1. КРАТКИЕ ТЕОРЕТИЧЕСКИЕ СВЕДЕНИЯ

1.1 Проверка гипотезы о равенстве дисперсий (Двухвыборочный F-тест для дисперсии).

F-критерий (Критерий Фишера) служит для проверки гипотезы о равенстве дисперсий двух генеральных совокупностей. Из каждой генеральной совокупности производятся выборки объема n_1 и n_2 . Выборки должны быть распределены нормально, иначе *Двухвыборочный F-тест для дисперсии* не применим, и в этом случае используются непараметрические тесты.

Проверяемая гипотеза H_0 утверждает, что $D_1=D_2$, то есть, дисперсии двух генеральных совокупностей однородны.

Альтернативная гипотеза H_1 утверждает, что $D_1 \neq D_2$, то есть, дисперсии двух генеральных совокупностей неоднородны.

Из каждой генеральной совокупности производятся выборки объема n_1 и n_2 и по выборочным данным вычисляются дисперсии S_1^2 и S_2^2 .

Далее вычисляется критерий Фишера (5.1), причем, в числителе находится большая из дисперсий, а в знаменателе – меньшая:

$$F_{расч} = \frac{\max\{S_1^2, S_2^2\}}{\min\{S_1^2, S_2^2\}}. \quad (5.1)$$

Критическое значение $F_{кр}$ выбирается из таблицы распределения Фишера (Приложение Г.1)

$F_{кр}(\alpha; n_1-1; n_2-1)$, где α – уровень значимости, а n_1-1 , n_2-1 степени свободы. Это же значение можно вычислить, применяя функцию ЭТ Excel FРАСП с аргументами α , n_1-1 и n_2-1 .

Если $F_{расч} < F_{кр}$, то гипотезу H_0 можно принять при уровне значимости α , иначе гипотеза должна быть отклонена.

1.2 Ранговый критерий Вилкоксона об однородности двух выборок.

Критерий Вилкоксона служит для проверки однородности двух независимых выборок: $(x_1, x_2, \dots, x_{n_1})$ и $(y_1, y_2, \dots, y_{n_2})$. В общем случае выборки могут иметь различную размерность, то есть $n_1 \neq n_2$. Достоинство этого критерия состоит в том, что он применим к случайным величинам, распределения которых неизвестны; требуется лишь, чтобы величины были непрерывными.

Если выборки однородны, то считается, что они извлечены из одной и той же генеральной совокупности и, следовательно, имеют одинаковые, хотя и неизвестные функции распределения (ФР) $F_1(x)$ и $F_2(x)$.

Алгоритм проверки гипотезы об однородности дисперсий.

Шаг 1. Формулируется гипотеза об однородности двух независимых выборок n_1 и n_2 ($n_1 \leq n_2$).

$$H_0: F_1(x) = F_2(x)$$

$$H_1: F_1(x) \neq F_2(x).$$

Шаг 2. Варианты обеих выборок располагаются в возрастающем порядке. Каждой вариане присваивается ранг r_i , $i=1, n_1+n_2$.

Вычисляется в этом ряду наблюдаемое значение критерия $W_{набл}$

$$W_{набл} = \sum_{r_i \in X} r_i \quad (5.2)$$

для одной из выборок, обычно выбирают первую из полученных выборок.

Шаг 3. По таблице «Критические точки критерия Вилкоксона» определяется нижняя критическая точка $w_{н\text{кр}}(Q; n_1, n_2)$, где $Q=\alpha/2$;

Шаг 4. Определяется верхняя критическая точка по формуле

$$w_{вкр} = (n_1 + n_2 + 1)n_1 - w_{н\text{кр}}$$

Шаг 5. Если $W_{набл} < w_{н\text{кр}}$ или $W_{набл} > w_{вкр}$, то нулевую гипотезу отвергают.

Если $w_{н\text{кр}} < W_{набл} < w_{вкр}$, то нет оснований отклонить нулевую гипотезу. В этом случае две выборки считаются однородными.

2. Задание для выполнения практической работы.

Проверить гипотезу об однородности дисперсий двух генеральных совокупностей, распределенных нормально при уровне значимости $\alpha=0,05$. Для каждого варианта предлагается 2 выборки, представленные в Приложении Г.2. Выполнить проверку однородности дисперсий для каждой пары выборок АВ и АС. Также эти варианты представлены в файле «Задания ПР_5.xlsx». Для каждой пары АВ и АС проводить проверку:

- 1) на основе расчета критерия Фишера;
- 2) с помощью *Пакета Анализа* ЭТ Excel;
- 3) на основании рангового критерия Вилкоксона.

3. Пример выполнения задания.

Даны две выборки. Определить, принадлежат ли эти выборки одной генеральной совокупности на основе критериев однородности выборок.

3.1 Проверка гипотезы о равенстве дисперсий на основе критерия Фишера.

1	5,04	6,64
2	5,95	6,55
3	5,56	9,20
4	3,58	8,81
5	3,32	7,56
6	3,66	6,18
7	3,83	7,31
8	8,03	8,13
9	5,25	7,36
10	4,12	6,40

1. Для каждой выборки вычисляем дисперсию, используя функцию ДИСП.В.

1	5,04	6,64
2	5,95	6,55
3	5,56	9,20
4	3,58	8,81
5	3,32	7,56
6	3,66	6,18
7	3,83	7,31
8	8,03	8,13
9	5,25	7,36
10	4,12	6,40
Дисп=	2,10	1,06

2. Определяем расчетное значение критерия Фишера в соответствии с (5.1), при этом нужно помнить, что в числитель дроби записываем большую дисперсию:

Дисп=	2,10	1,06
Фрасч=	1,98	
Фрасч=	=W24/X24	

3. Определяем критическое значение критерия Фишера, используя функцию ФРАСП:

Фкрит=	3,18
Фкрит=	=FРАСПОБР(0,05;9;9)
FРАСПОБР(вероятность; степени_свободы1; степени_свободы2)	

Аргументы функции: вероятность – уровень значимости $\alpha=0,05$; поскольку размеры выборок равны $n_1=n_2=10$, то и степени свободы тоже равны $n_1-1=n_2-1=9$.

Поскольку $F_{расч} < F_{кр}$, то гипотезу H_0 можно принять при уровне значимости 0,05. Это означает, что дисперсии двух выборок однородны, и они принадлежат одной генеральной совокупности.

3.2 Двухвыборочный F-тест для дисперсии в Пакете анализа ЭТ Excel.

Исходными данными для проверки однородности дисперсий являются две выборки, не обязательно равного размера. Задание выполняется для исходных данных п.3.1.

Порядок действий:

Шаг 1. Выполнить команду **Данные. Анализ данных**.

Шаг 2. В диалоговом окне выбрать из списка **Двухвыборочный F-тест** для дисперсии.

Шаг 3. Заполнить поля ввода в открывшемся диалоговом окне:

- Интервал выборки 1 – интервал выборки с **большой** дисперсией;
- Интервал выборки 2 – интервал выборки с **меньшей** дисперсией;
- Выходной интервал – установить переключатель и в поле ввода ввести одну ячейку на свободном месте рабочего листа.

В результате на рабочем листе появится таблица:

Двухвыборочный F-тест для дисперсии			
	Переменная 1	Переменная 2	
Среднее	4,833981178	7,413604312	
Дисперсия	2,104765127	1,063253096	
Наблюдения	10	10	
df	9	9	
F	1,979552313		
P(F<=f) односто	0,161786475		
F критическое с	3,178893104		

Получили результат: $F_{расч} < F_{кр}$, поэтому гипотезу H_0 можно принять при уровне значимости 0,05. Это означает, что дисперсии двух выборок однородны, и они принадлежат одной генеральной совокупности.

Кроме того, Р-значение больше 0,1. Это также говорит о том, что нулевую гипотезу можно принять с достаточно большой долей уверенности.

Замечание. Если расчетное значение F-критерия получилось меньше единицы $F_{расч} < 1$, это означает, что в **Интервал выборки 1** внесена выборка с меньшей из двух дисперсий. В этом случае для принятия нулевой гипотезы должно выполняться соотношение $F_{расч} > F_{кр}$, а Р-значение все равно должно быть больше 0,1.

3.2 Ранговый критерий Вилкоксона об однородности дисперсий.

Задание выполняется для исходных данных п.3.1 (рис.5.1а).

Порядок действий:

Шаг 1. Для визуального удобства рекомендуется раскрасить выборки разными цветами (рис.5.1а).

Шаг 2. Объединить выборки в одну (рис.5.1б)).

Шаг 3. Отсортировать объединенную выборку по возрастанию (рис 5.1в)).

Шаг 4. Отметить цветом ранги выборки А (рис. 5.1г)) и сложить их. Ранги подсчитывают для выборки, со значения которой начинается объединенная выборка. Сумма рангов (5.2) и есть наблюдаемое значение критерия Вилкоксона $W_{набл} = 62$.

	A	5,04		A	3,32		A	3,32	1
	A	5,95		A	3,58		A	3,58	2
	A	5,56		A	3,66		A	3,66	3
	A	3,58		A	3,83		A	3,83	4
	A	3,32		A	4,12		A	4,12	5
	A	3,66		A	5,04		A	5,04	6
	A	3,83		A	5,25		A	5,25	7
	A	8,03		A	5,56		A	5,56	8
	A	5,25		A	5,95		A	5,95	9
	A	4,12		B	6,18		B	6,18	10
	B	6,64		B	6,40		B	6,40	11
	B	6,55		B	6,55		B	6,55	12
	B	9,20		B	6,64		B	6,64	13
	B	8,81		B	7,31		B	7,31	14
	B	7,56		B	7,36		B	7,36	15
	B	6,18		B	7,56		B	7,56	16
	B	7,31		A	8,03		A	8,03	17
	B	8,13		B	8,13		B	8,13	18
	B	7,36		B	8,81		B	8,81	19
	B	6,40		B	9,20		B	9,20	20

а)

б)

в)

г)

Рисунок 5.1 – Этапы вычисления наблюдаемого значения критерия Вилкоксона:

а) исходные данные; б) объединенная выборка; в) упорядоченная выборка;

г) подсчет рангов

Шаг 5. Определение критических значений критерия Вилкоксона. Значения нижней и верхней границы области принятия гипотезы определяются исходя из размеров выборок и уровня значимости (шаги 3 и 4 алгоритма) и для рассматриваемого примера равны $w_{н\text{кр}} = 78$ и $w_{в\text{кр}} = 132$, то есть для принятия нулевой гипотезы наблюдаемое значение критерия должно находиться в интервале (78, 132).

Поскольку наблюдаемое значение критерия находится за пределами области принятия гипотезы, нулевую гипотезу отвергаем и принимаем альтернативную. Следовательно, рассматриваемые выборки неоднородны и принадлежат разным генеральным совокупностям.

Итак, для одних и тех же выборок получены разными методами различные результаты. Какой из полученных результатов принять, решает лицо, принимающее решение.

4. Содержание отчета

1. Постановка задачи.
2. Ход выполнения задания с пояснениями.
3. Анализ результатов и выводы.

Практическая работа № 6

Сравнение средних показателей в выборках на основе критерия Стьюдента

Цель: получить навыки проверки сравнения средних показателей в группах на основе Т-теста (критери Стьюдента).

1. КРАТКИЕ ТЕОРЕТИЧЕСКИЕ СВЕДЕНИЯ

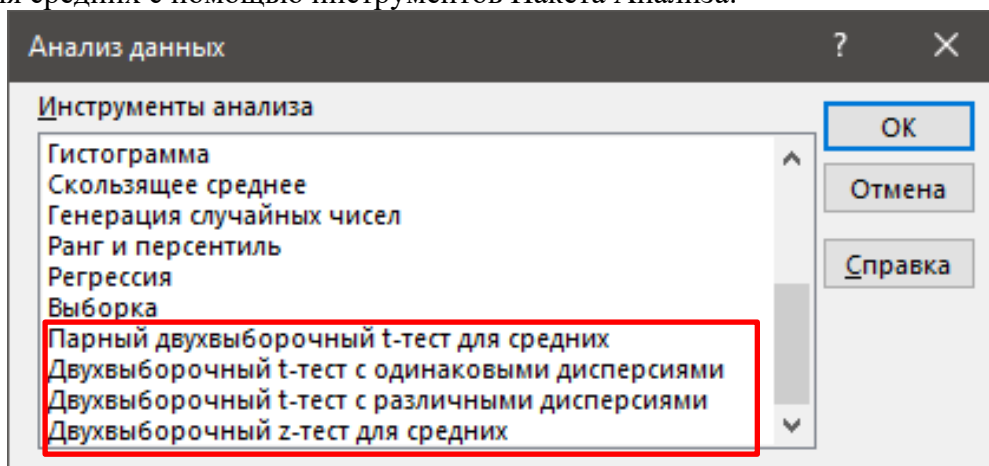
Методы сравнения средних величин часто используются в социологических, маркетинговых и медицинских исследованиях для выявления взаимосвязи между исследуемыми переменными. К таким методам относятся Т-тесты и дисперсионный анализ.

Метод Т-теста используется для проверки достоверности различия двух выборок по количественной переменной.

Нулевая гипотеза **H₀** формулируется следующим образом:

«Взаимосвязи между исследуемыми переменными нет». Например, формулируя исходную (нулевую) гипотезу предполагаем равенство среднего чека покупателей разного возраста, это говорит о том, что люди разного возраста в среднем тратят одинаковую сумму на покупки. В результате проведенного исследования данная гипотеза либо подтверждается, либо опровергается.

ЭТ Excel предоставляет исследователю возможность выполнить процедуру сравнения средних с помощью инструментов Пакета Анализа:



1.1 Парный двухвыборочный t-тест для средних. *Т-критерий для парных выборок* позволяет сравнить средние значения двух измерений одного признака для одной и той же выборки. Например, результаты опроса потребителей в начале и в конце месяца.

1.2 Двухвыборочный t-тест с одинаковыми дисперсиями и двухвыборочный t-тест с различными дисперсиями. *Т-критерий для независимых выборок* используется для сравнения средних значений двух выборок. Для данных выборок должны быть определены значения одной и той же переменной. Например, удовлетворенность покупкой мужчин и женщин, семейных и несемейных и т.д. Эти инструменты предполагают предварительное исследование однородности дисперсий выборок на основе инструмента **Пакета Анализа** «Двухвыборочный F-тест для дисперсии».

1.3 Двухвыборочный z-тест для средних. Используется также для независимых выборок, требует предварительный расчет выборочных дисперсий.

1.4 Функция ЭТ Excel Z.ТЕСТ позволяет проверять *Одновыборочный t-критерий*, который используется для сравнения средней величины изучаемого признака в выборке с некоторой эталонной величиной. Например, отличается ли средняя цена на конкретный

товар выбранных торговых предприятий от средней цены, рассчитанной официальным источником.

Т-критерий для независимых выборок используется для сравнения средних значений двух выборок. Для данных выборок должны быть определены значения одной и той же переменной. Например, удовлетворенность покупкой мужчин и женщин, семейных и несемейных и т.д.

Т-критерий для парных выборок позволяет сравнить средние значения двух измерений одного признака для одной и той же выборки. Например, результаты опроса потребителей в начале и в конце месяца.

В случае, когда необходимо сравнивать величины более чем в двух группах, применяется *однофакторный дисперсионный анализ*.

Результат сравнения средних значений с применением Т-теста оценивается еще и по уровню значимости (*p-значение*). Значимость является мерой достоверности вычисленных результатов. Если «Значимость» не превышает 0,05, это означает, что нулевая (исходная) гипотеза может быть отклонена с вероятностью ошибки меньше 5%, т.е. ее можно отклонить и, следовательно, она неверна.

2. Пример выполнения задания

Пример 1. Продуктовый магазин провел исследование о частоте покупок мужчинами и женщинами в течении месяца. Определить, действительно ли частота покупок у мужчин статистически отличается от частоты покупок у женщин. Исходные данные представлены в таблице 6.1.

Таблица 6.1 – Частота покупок у мужчин и женщин

№	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Ж	2	3	3	6	2	1	6	3	4	13	6	2	6	4	12
М	7	10	5	9	8	5	9	4	10	9	5	12	11	4	3

H₀: различия между средними статистически неразличимы (частота покупок у мужчин и женщин одинакова).

H₁: есть существенные различия между выборками.

1. Выборки являются независимыми, поэтому выбираем инструмент Пакета Анализа **Двухвыборочный t-тест ...**

2. Проверяем дисперсии выборок на однородность на основе инструмента Пакета Анализа **Двух выборочный F-тест для дисперсии**. Выдвигаем гипотезу:

H₀ – выборки однородны,

H₁ – выборки неоднородны. В результате получаем:

Двухвыборочный F-тест для дисперсии		
	Переменная 1	Переменная 2
Среднее	4,866666667	7,4
Дисперсия	12,40952381	8,257142857
Наблюдения	15	15
df	14	14
F	1,502883506	
P(F<=f) одност	0,227787325	
F критическое	2,483725741	

Поскольку значения критериев таковы, что $F_{крит} > F_{расч}$ ($2,48 > 1,50$), принимаем гипотезу H_0 и считаем, что выборки однородны.

3. Так как дисперсии выборок однородны, выбираем инструмент **Двухвыборочный t-тест с одинаковыми дисперсиями** и заполняем диалоговое окно:

Поле «Гипотетическая средняя разность» заполнять нет необходимости. Интервалы для переменных можно выделять с заголовками, но потом обязательно выставить флажок **Метки**. В результате:

Степени свободы: $df=28$
t-статистика: $T_{расч}=2,16$

t критическое
двухстороннее: $T_{крит}=2,05$.
 $T_{расч} > T_{крит}$ – выборки
различны.

$P(T \leq t)$ двухстороннее:
P-значение $p=0,04$.

Двухвыборочный t-тест с одинаковыми дисперсиями		
	Переменная 1	Переменная 2
Среднее	4,866666667	7,4
Дисперсия	12,40952381	8,257142857
Наблюдения	15	15
Объединенная дисп	10,33333333	
Гипотетическая разн	0	
df	28	
t-статистика	-2,15825497	
$P(T \leq t)$ односторон	0,019815814	
t критическое одно	1,701130934	
$P(T \leq t)$ двухсторон	0,039631629	
t критическое двухс	2,048407142	

Поскольку значения критериев $T_{расч} > T_{крит}$ ($2,16 > 2,05$) гипотезу **H_0** отклоняем и принимаем альтернативную гипотезу **H_1** : выборки статистически различимы, то есть **частота покупок у мужчин и женщин различна**. Кроме того, p-значение $p=0,04 < 0,05$ есть подтверждение того, что **H_0** надо отклонить.

Пример 2. По исходным данным **Примера 1** определить, действительно ли частота покупок у мужчин статистически отличается от частоты покупок у женщин на основе инструмента **Двухвыборочный z-тест для средних**. Для применения z-теста необходимо заранее вычислить выборочные дисперсии, а затем ввести их значения в соответствующие поля ввода.

№	Ж	М
1	2	7
2	3	10
3	3	5
4	6	9
5	2	8
6	1	5
7	6	9
8	3	4
9	4	10
10	13	9
11	6	5
12	2	12
13	6	11
14	4	4
15	12	3
дисп.в	12,40952	8,257143

Двухвыборочный z-тест для средних

Входные данные

Интервал переменной 1:

Интервал переменной 2:

Гипотетическая средняя разность:

Дисперсия переменной 1 (известная) } Поля для ввода дисперсий.

Дисперсия переменной 2 (известная)

☐ Метки

Альфа:

Параметры вывода

☒ Выходной интервал:

☐ Новый рабочий диск:

☐ Новая рабочая книга

OK Отмена Справка

В результате:

Двухвыборочный z-тест для средних		
	Переменная 1	Переменная 2
Среднее	4,866666667	7,4
Известная дисперсия	12,41	8,26
Наблюдения	15	15
Гипотетическая разность	0	
z	-2,158080938	
P(Z<=z) односторонний	0,01546077	
z критическое одностороннее	1,644853627	
P(Z<=z) двухсторонний	0,030921539	
z критическое двухстороннее	1,959963985	

Поскольку значения критериев $Z_{расч} > Z_{крит}$ ($2,16 > 1,95$) делаем вывод, что выборки различны. Кроме того, р-значение $p = 0,031$, что меньше $\alpha = 0,05$. Это обстоятельство также говорит о том, что гипотезу H_0 следует отклонить и принять альтернативную H_1 .

Итак, частота покупок у мужчин и женщин различна и составляет в среднем соответственно 4,87 и 7,4 покупки.

Пример 3. Необходимо установить влияние занятий спортивной гимнастикой на число подтягиваний на перекладине.

Мальчики (8 человек), принятые в ДЮСШ на отделение спортивной гимнастики, через месяц проходили испытания в контрольном упражнении – подтягивании на перекладине.

Позволяют ли приведенные данные (табл. 6.2) утверждать, что в процессе второго испытания все мальчики улучшили общую физическую подготовленность? Необходимо доказать, что это улучшение достоверно.

Таблица 6.2 – Результаты испытаний

Испытуемый		1	2	3	4	5	6	7	8
Число подтягиваний	1 испытание	4	7	6	5	8	5	8	6
	2 испытание	7	10	10	8	12	9	11	11

Выборки, соответствующие результатам испытаний, являются зависимыми (парными), поскольку сравниваются средние значения двух измерений одного и того же признака – учащийся спортивной школы. Поэтому для решения поставленной задачи необходимо выбрать инструмент Пакета Анализа **Парный двухвыборочный t-тест для средних**.

Выдвигаем статистические гипотезы:

– нулевая гипотеза H_0 – отсутствие влияния занятий спортивной гимнастикой на число подтягиваний на перекладине;

– альтернативная гипотеза H_1 – наличие влияния занятий спортивной гимнастикой на число подтягиваний на перекладине.

В результате применения соответствующего инструмента получаем:

Парный двухвыборочный t-тест для средних		
	Переменная 1	Переменная 2
Среднее	6,125	9,75
Дисперсия	2,125	2,785714286
Наблюдения	8	8
Корреляция Пирсон	0,895414351	
Гипотетическая разн	0	
df	7	
t-статистика	-13,78053794	
P(T<=t) односторонн	1,25013E-06	
t критическое одно	1,894578605	
P(T<=t) двухсторонн	2,50026E-06	
t критическое двухс	2,364624252	

Поскольку значение критериев Трасч>Ткрит ($13,78 > 2,36$) и р-значение практически равно нулю, гипотезу **H0** отвергаем и принимаем гипотезу **H1** о наличии влияния занятий на результаты. Более того, коэффициент корреляции Пирсона близок к 1 (0,89), что говорит о **значительном** влиянии тренировок на результаты испытаний. Количество подтягиваний увеличилось в среднем с 6,1 до 9,75.

Пример 4. Средняя высота растений известна и равна 40,5 см. Определить, совпадает ли высота растений, выборочно измеренная на поле (табл. 6.3), с эталонным значением. Исследование проводится по одной выборке и его можно выполнить в ЭТ Excel, используя функцию Z.ТЕСТ:

=Z.ТЕСТ(D2:D11;E6)
Z.ТЕСТ(массив; x; [сигма])

Аргументами функции являются:

- массив – интервал с выборкой;
- x – «эталонное значение»;
- [сигма] – необязательный параметр (стандартное отклонение).

Функция возвращает р-значение. Если $p > 0,05$, гипотеза H_0 принимается, иначе – отвергается.

Таблица 6.3 – Высота растений

№	1	2	3	4	5	6	7	8	9	10
высота	42	39	40	45	41	43	44	41	43	42

Для решения поставленной задачи выдвигаем гипотезу:

H_0 : среднее выборки статистически не отличается от «эталонного» значения.

H_1 : среднее выборки значимо отличается от «эталонного» значения.

Заполняя соответствующие аргументы функции Z.ТЕСТ, получаем результат:

р-знач=	0,008232
р-знач=	=Z.ТЕСТ(B2:B11;C6)

Поскольку р-значение равно 0,008, что меньше 0,05, гипотезу H_0 отвергаем и принимаем альтернативную гипотезу H_1 : высота растений на поле значимо отличается от эталонной высоты 40,5 см.

Пример 5. Отдел кадров исследует возрастной состав сотрудников организации. Определить, действительно ли статистически отличается возрастной состав отдела № 1 и отдела № 2 (табл. 6.4).

Таблица 6.4 – Возрастной состав отделов

Отд1	51	48	37	43	44	43	37	43	40	39
Отд2	42	49	38	51	31	59	47	41	35	36

H_0 : различия между средними статистически неразличимы (возрастной состав отделов одинаковый).

H_1 : есть существенные различия между выборками.

1. Выборки являются независимыми, поэтому выбираем инструмент Пакета Анализа **Двухвыборочный t-тест ...**

2. Проверяем дисперсии выборок на однородность на основе инструмента Пакета Анализа **Двух выборочный F-тест для дисперсии**. Выдвигаем гипотезу:

H_0 – выборки однородны,

H_1 – выборки неоднородны. В результате получаем:

Двухвыборочный F-тест для дисперсии		
	Переменная 1	Переменная 2
Среднее	42,9	42,5
Дисперсия	73,21111111	20,5
Наблюдения	10	10
df	9	9
F	3,571273713	
P(F<=f) одност	0,035828666	
F критическое	3,178893104	

Расчетное значение критерия Фишера $F=3,57$, а расчетное – $F_{расч}=3,18$. Поскольку расчетное значение критерия превышает критическое, то считаем, что дисперсии выборок неоднородны - различны. Для проверки равенства выборок выбираем **Двухвыборочный t-тест с различными дисперсиями**.

В результате применения соответствующего инструмента получаем:

Двухвыборочный t-тест с различными дисперсиями		
	Переменная 1	Переменная 2
Среднее	42,5	42,9
Дисперсия	20,5	73,21111111
Наблюдения	10	10
Гипотетическая дисперсия	0	
df	14	
t-статистика	-0,130666557	
P(T<=t) односторонний	0,448949054	
t критическое	1,761310136	
P(T<=t) двусторонний	0,897898108	
t критическое	2,144786688	

Расчетное значение t-статистики равно $t=0,13$, критическое значение равно $t_{крит}=2,14$. Поскольку критическое значение превышает расчетное, принимаем гипотезу H_0 и считаем, что выборки статистически не различимы. Кроме того, Р-значение значительно превышает уровень значимости, что тоже говорит о том, что гипотезу H_0 надо принять. Таким образом, возрастной состав в исследуемых отделах статистически неразличим.

3. Задание для выполнения практической работы.

Материалы для выполнения практических заданий содержатся по ссылке:

<https://do.sevsu.ru/mod/assign/view.php?id=518183>

Индивидуальные варианты заданий представлены в Приложении Д и в файле **Задания_ПР6**)

1. **Задание 1.** Выполнить примеры 1-5 (файл **Задания_ПР6**, лист Задания п.1) и сравнить результаты самостоятельных исследований с представленными в п. 2. Результаты выполнения представить в виде файла .xlsx.

2. **Задание 2.** Даны выборки X_1 , X_2 , Y и эталонное значение μ_0 (Приложение А, лист в файле **Задания п.2**):

- определить равенство средних в выборках X_1 и X_2 .
- определить, являются ли зависимыми выборки X_1 и Y .
- исследовать соответствие среднего выборки Y с эталонным значением.

3. **Задание 3.** Имеются данные выборочного наблюдения (лист **Задание п.3**), характеризующие товарооборот за месяц однотипных торговых организаций двух микрорайонов города (обследовалось по 20 предприятий каждого микрорайона).

Проверить, является ли различие по двум микрорайонам в величине товарооборота предприятия статистически значимым.

4. **Задание 4.** Компания, работающая на рынке, хочет оценить влияние обучения персонала на уровень продаж. Кадровая служба компании осуществляет обучение сотрудников отдела продаж. Результаты продаж десяти сотрудников после обучения приведены в таблице на Листе **Задание п.4**. Необходимо выяснить, влияет ли обучения персонала на уровень продаж.

5. **Задание 5.** Имеются следующие данные (Лист **Задание п.5**) о величине дневной выработки водителей городских автобусов двух автотранспортных предприятий (АТП), расположенных в двух микрорайонах города (обследовалось по 20 водителей каждого АТП).

Проверить, является ли различие в дневной выработке водителей по двум АТП статистически значимым.

6. Задание 6. Исследуется эффективность обучения двумя различными методами. Студентам дается задание изучить определенную тему. Для этого 10 студентов конспектируют первоисточник, а 10 изучают ее по электронному учебнику. По окончании их уровень знаний проверяется с помощью теста, состоящего из 100 вопросов. Результаты представлены на Листе **Задание п.6**.

Определить, влияет ли методика изучения темы на результат.

7. Задание 7. Средний рейтинг подготовки в группе оздоровления составляет 100 единиц. После применения комплекса оздоровительных упражнений установлен рейтинг (Лист **Задание п.7**) для каждого члена группы. Определить, есть ли влияние нового комплекса оздоровительных упражнений состояние здоровья.

4. Содержание отчета

1. Выполнить задания 1-7. Для студентов, присутствовавших на лекции, задание 1 в отчет не включать.

Каждое задание оформлять по следующей схеме:

- Постановка задачи.
- Ход выполнения заданий с пояснениями и **интерпретацией** результатов исследования.

2. Общие выводы по практической работе.

Практическая работа № 7

Исследование взаимосвязи между данными на основе корреляционного анализа

Цель работы: научиться выявлять наличие взаимосвязи между данными на основе корреляционного анализа.

1. Краткие теоретические сведения.

Исследуя явления в самых различных областях, статистика неизбежно сталкивается с зависимостями, как между количественными, так и качественными признаками.

Задача статистики – выявить такие зависимости и дать их количественную характеристику.

Взаимосвязанные признаки (переменные) делятся на 2 группы:

- *независимые (факторные)* - определяющие факторы, влияющие на изменение других признаков;

- *зависимые (результативные)* - результат влияния независимых признаков.

Зависимости между признаками могут быть двух видов:

- функциональные (детерминированные);

- статистические (недетерминированные, стохастические).

Зависимость между двумя признаками x и y называется *функциональной*, если определенному значению признака x однозначно соответствует значение признака y .

Зависимость называется *статистической*, если зависимый признак y определяется одновременным действием многих факторов и его значение при фиксированном значении x может варьироваться.

Статистические зависимости можно обнаружить только при массовом наблюдении. Исследование статистических зависимостей и есть суть анализа взаимосвязей.

Анализ взаимосвязей включает в себя:

- Корреляционный анализ
- Дисперсионный анализ
- Регрессионный анализ.

Основные показатели корреляционного анализа (КА) – *коэффициент корреляции (количественная оценка)* и *диаграмма рассеяния (наглядность)*.

Если зависимость существует, то коэффициент корреляции позволяет также установить, является она прямой или обратной.

Прямая зависимость означает, что факторная и результативная переменная возрастают или убывают одновременно.

Обратная зависимость присутствует тогда, когда факторная переменная возрастает, а результативная - убывает и наоборот.

Коэффициент корреляции Пирсона (r -Пирсона) характеризует существование *линейной* связи между двумя величинами.

Условия применения корреляционного анализа:

1. Исследуемые переменные x и y должны быть распределены нормально.
2. Исследуемые переменные x и y должны быть измерены в числовой или порядковой шкале.
3. Количество значений в исследуемых переменных x и y должно быть одинаковым.

Коэффициент корреляции изменяется в диапазоне от -1 до +1. Считается, что если коэффициент корреляции по абсолютной величине больше $>0,5$, то связь между признаками заметная, иначе – связь отсутствует (рис. 7.1).

Сильная	Нет	Сильная
обратная	линейной	прямая
связь	связи	связь
-1	0	+1

Рисунок 7.1 – Характеристика связей между данными.

Пример.1 Вычислить коэффициент корреляции Пирсона, построить диаграмму рассеяния и определить тип зависимости для двух признаков – возраста пациента и его артериального давления. Исходные данные приведены в таблице 7.1.

Таблица – 7.1 Результаты наблюдения пациентов.

Субъект	Возраст x_i	Давление y_i
1	43	128
2	48	120
3	56	135
4	61	143
5	67	141
6	70	152

Для решения задачи необходимо выполнить команду **Данные. Анализ данных. Корреляция**. В результате имеем (рис. 7.2):

	Возраст x_i	Давление y_i
Возраст x_i	1	
Давление y_i	0,896672815	1

Коэффициент Пирсона

Рисунок 7.2 – Результаты корреляционного анализа

Поскольку коэффициент Пирсона $r=0,897>0$, связь между переменными *Возраст пациента* и *Давление* прямая и достаточно сильная. По результатам анализа можно сделать вывод, что артериальное давление сильно зависит от возраста, при этом с увеличением возраста увеличивается и давление.

2. Задания для выполнения практической работы

Материалы для выполнения практического задания содержатся по ссылке:

<https://do.sevsu.ru/mod/assign/view.php?id=518687>

1. **Задание 1.** Выполнить пример 1 пункта 1. Представить выполненное задание в файле.xlsx.

2. **Задание 2.** Построить диаграмму рассеяния, вычислить коэффициент корреляции Пирсона и определить тип зависимости для двух признаков – количество пропущенных занятий и итоговая оценка.

Число пропусков	6	2	15	9	12	5	8
Итоговая оценка	82	86	43	74	58	90	78

3. **Задание 3.** Исходные данные размещены в файле **Задания_ПР7.xlsx**, Лист **Задание 3**. Рассматриваются две переменные: «Продолжительность подготовки (в часах)» студентов перед экзаменом и «Итоговая оценка». Построить диаграмму рассеяния, вычислить коэффициент корреляции Пирсона и определить тип зависимости для данных переменных.

4. **Задание 4.** Исходные данные размещены в файле **Задания_ПР7.xlsx**, Лист **Задание 4**. В выборку вошли данные о 20 женщинах в возрасте от 25 до 30 лет. В наборе данных указаны следующие переменные:

- у - количество жира в организме;
- x1 - толщина кожной складки трицепса;
- x2 - окружность бедра;
- x3 - окружность предплечья.

На основе корреляционного анализа определить, есть ли зависимость между переменными:

- а) толщина кожной складки трицепса и количество жира в организме;
- б) окружность бедра и количество жира в организме;
- в) окружность предплечья и количество жира в организме;
- г) толщина кожной складки трицепса и окружность бедра;
- д) толщина кожной складки трицепса и окружность предплечья;
- е) окружность бедра и окружность предплечья.

4. Содержание отчета

1. Цель работы.
2. Постановка задачи
3. Ход выполнения работы.
4. Анализ результатов (для каждого задания).
5. Общие выводы по работе.

Практическая работа № 8

Исследование взаимосвязи между данными на основе дисперсионного анализа

Цель работы: научиться выявлять наличие взаимосвязи между данными на основе корреляционного и однофакторного дисперсионного анализа.

1. Краткие теоретические сведения.

Исследуя явления в самых различных областях, статистика неизбежно сталкивается с зависимостями, как между количественными, так и качественными признаками.

Задача статистики – выявить такие зависимости и дать их количественную характеристику.

Взаимосвязанные признаки (переменные) делятся на 2 группы:

- *независимые (факторные)* - определяющие факторы, влияющие на изменение других признаков;

- *зависимые (результативные)* - результат влияния независимых признаков.

Зависимости между признаками могут быть двух видов:

- функциональные (детерминированные);
- статистические (недетерминированные, стохастические).

Зависимость между двумя признаками x и y называется *функциональной*, если определенному значению признака x однозначно соответствует значение признака y .

Зависимость называется *статистической*, если зависимый признак y определяется одновременным действием многих факторов и его значение при фиксированном значении x может варьироваться.

Статистические зависимости можно обнаружить только при массовом наблюдении. Исследование статистических зависимостей и есть суть анализа взаимосвязей.

Анализ взаимосвязей включает в себя:

- Корреляционный анализ
- Дисперсионный анализ
- Регрессионный анализ.

Дисперсионный анализ используют для изучения взаимосвязей, исследования влияния одной или нескольких качественных переменных на одну зависимую количественную переменную. Основной целью подобного исследования является установление значимости различия между средними значениями зависимой количественной переменной по группам фактора. Достигается это с помощью разложения общей дисперсии зависимой переменной на составляющие: общая дисперсия; межгрупповая дисперсия; средняя внутригрупповая дисперсия.

Величина *общей* дисперсии σ_o^2 характеризует вариацию признака под влиянием всех факторов, формирующих уровень признака у единиц данной совокупности, и определяется по формуле

$$\sigma_o^2 = \frac{1}{k \cdot n} \sum_{j=1}^k \sum_{i=1}^n (y_{ij} - \bar{y})^2 \quad (8.1)$$

где $\bar{y}$ - общая средняя арифметическая для всей изучаемой совокупности.

Межгрупповая дисперсия (дисперсия групповых средних - δ^2) отражает систематическую вариацию, т.е. различия в величине изучаемого признака, которые возникают под влиянием фактора, положенного в основу группировки.

$$\delta^2 = \frac{1}{k} \sum_{j=1}^k (\bar{y}_j - \bar{y})^2, \quad (8.2)$$

где $\bar{y}_j$ - средняя арифметическая по каждой группе, $j=1, k$.

Средняя внутригрупповая дисперсия $\bar{\sigma}^2$ характеризует случайную вариацию, возникающую под влиянием других, неучтенных факторов, и не зависит от признака-фактора, положенного в основу группировки:

$$\bar{\sigma}^2 = \frac{1}{k} \sum_{j=1}^k \sigma_j^2, \quad (8.3)$$

где дисперсия в каждой группе рассчитывается по формуле:

$$\sigma_j^2 = \frac{1}{n_j} \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2, \quad (8.4)$$

Указанные дисперсии взаимосвязаны. Их связь между собой выражена равенством, которое называется правилом сложения дисперсий: **величина общей дисперсии равна сумме межгрупповой дисперсии и средней внутригрупповой**. Математическое выражение для правила сложения дисперсий представлено ниже:

$$\sigma_o^2 = \delta^2 + \bar{\sigma}^2. \quad (8.5)$$

Опираясь на это правило, можно определить, какая доля общей дисперсии складывается под влиянием признака-фактора, положенного в основу группировки δ^2 , то есть дисперсию за счет разбиения на группы и дисперсию за счет остальных факторов. Анализируя эти компоненты дисперсии, можно оценить долю воздействия каждого фактора на исследуемую величину.

В зависимости от количества факторов, определяющих вариацию признака-результата, дисперсионный анализ подразделяют на однофакторный и многофакторный.

Простейшим случаем дисперсионного анализа является однофакторный анализ, в котором n единиц совокупности распределены на m групп по значениям одного фактора. Основная идея однофакторного дисперсионного анализа заключается в сравнении дисперсии исследуемого признака, вызванной действием фактора, с дисперсией ошибок измерения этого признака. Если различие между ними значимо, то фактор оказывает существенное влияние на исследуемый признак.

В ходе расчетов вариации значений используются суммы квадратов отклонений от средних:

- **общая** вариация **SSE** (*the Sum of Squared estimate of Errors*);
- вариация, объясняющая различия **между группами** **SSB** (*between groups*) - **межгрупповую сумму квадратов отклонений**
- вариация, объясняющую различия между единицами совокупности внутри группы **SSW** (*within groups*) - **внутригрупповую сумму квадратов отклонений**.

Поскольку суммы квадратов отклонений также описывают вариацию результирующего признака, можно представить выражением:

$$SSE = SSW + SSB.$$

Далее рассчитывается межгрупповой средний квадрат отклонений (на одну степень свободы) **MSB**, а также внутригрупповой средний квадрат отклонений (на одну степень свободы) **MSW**:

$$S_1^2 = \frac{SSB}{(m-1)} = MSB, \quad S_2^2 = \frac{SSW}{n-m} = MSW,$$

где n - число единиц анализируемой совокупности;

m - число групп, по которым распределены единицы совокупности.

Если различия внутригрупповых средних носят случайный характер, т.е. не связаны с изменением факторного признака, то $S_1^2 = S_2^2$.

Это положение соответствует нулевой гипотезе **$H_0: S_1^2 = S_2^2$** .

Для выявления степени влияния изучаемого фактора X и сопоставления ее с разбросом вычисляют наблюдаемое значение критерия Фишера:

$$F_{расч} = \frac{S_1^2}{S_2^2}.$$

Все эти величины объединяются в таблицу дисперсионного анализа, содержание которой иллюстрирует таблица 8.2.

Таблица 8.2. - Однофакторный дисперсионный анализ

Источник вариации	df	SS	MS	F	Значимость F
Межгрупповая (between groups)	$m-1$	SSB	$SSB/(m-1)$	$SSB/(m-1) / SSW / (n-m)$	
Внутригрупповая (within groups)	$n-m$	SSW	$SSW/(n-m)$		
Общая(total)	$n-1$	$SSB+SSW$			

Наблюдаемое значение критерия сравнивают с критическим $F_{табл} = F(\alpha, df_1, df_2)$, которое находят по таблицам F -распределения для уровня значимости α , числа степеней свободы $df_1 = m-1$ и $df_2 = n-m$.

Если $F_{расч} \leq F_{табл}$, то влияние фактора X незначительно, и следует принять нулевую гипотезу. Следовательно, все полученные результаты измерений принадлежат одной генеральной совокупности с общим средним и дисперсией.

При $F_{расч} > F_{табл}$ влияние фактора принимают существенным и нулевую гипотезу отвергают. В этом случае полагают, что есть m нормально распределенных совокупностей, каждая из которых имеет свои среднее и групповую дисперсию.

Если установлено влияние, то связь между фактором, положенным в основу группировки можно измерить на основании эмпирического корреляционного отношения:

$$\eta_{yx} = \sqrt{\frac{SSB}{SSE}}$$

Этот показатель принимает значения в интервале от 0 до 1. Связь тем теснее, чем ближе значение η к 1, и наоборот (о направлении связи здесь речь не идет).

Для вывода (заключения) о тесноте связи можно воспользоваться следующими рекомендациями, сведенными в таблицу 8.3:

Таблица 8.3. – Шкала изменения тесноты связи результата с фактором

Значение эмпирического корреляционного	Теснота связи
0,3 и меньше	слабая
0,3 - 0,5	средняя
0,5 - 0,7	умеренно тесная
0,7 и больше	тесная

Пример 1. Требуется проверить нулевую гипотезу о влиянии профессии рабочих на продолжительность заболеваний хроническим бронхитом, повлекшим временную нетрудоспособность. В ходе эксперимента были выделены следующие профессиональные группы: станочники, слесари, прочие. Выборки сделаны на предприятии из 4 цехов с близкими условиями и характером труда работников. Предполагается, что распределение подчиняется нормальному закону.

Таким образом, в роли фактора (имеющего три уровня) выступает профессия рабочих, а в роли зависимого признака (результатирующего) – длительность заболевания в днях. Исходные данные представлены на рисунке 8.3.

Цех	Длительность заболевания		
	станочники	Слесари	Прочие
1	10	5	4
2	14	11	6
3	9	7	11
4	12	8	4

Рисунок 8.3 – Результаты исследования длительности заболевания у рабочих.

Формулируем гипотезу **H₀**: дисперсии по группам равны, влияния фактор, положенный в основу группировки не оказывает; **H₁**: Влияние фактора, положенного в основу группировки, существенно.

Для выполнения задания необходимо:

- выполнить команду **Данные. Анализ данных. Однофакторный дисперсионный анализ**;
- заполнить поля ввода диалогового окна (входной интервал выделен на рис. 8.3):

В результате имеем:

1. **Итоговую** таблицу с описательной статистикой по группам.

Однофакторный дисперсионный анализ				
ИТОГИ				
Группы	Счет	Сумма	Среднее	Дисперсия
станочники	4	45	11,25	4,916667
Слесари	4	31	7,75	6,25
Прочие	4	25	6,25	10,91667

2. Результаты дисперсионного анализа.

Дисперсионный анализ						
Источник вариации	SS	df	MS	F	P-Значение	F критическое
Между группами	52,66667	2	26,33333	3,577358	0,071902535	4,256494729
Внутри групп	66,25	9	7,361111			
Итого	118,9167	11				

В столбце SS представлены суммы квадратов разностей, то есть вариация межгрупповая и внутри групповая; df – соответствующие степени свободы; MS – средние суммы квадратов разностей, F – расчетное значение критерия Фишера, p -значение, $F_{критическое}$ – критическое значение того же критерия.

3. Интерпретация результатов.

Поскольку $F_{кр} > F$, то принимается нулевая гипотеза об отсутствии влияния фактора на значение зависимого признака. Кроме того, p -значение $p > 0,05$, что тоже свидетельствует о том, что нулевую гипотезу надо принять.

Таким образом выяснено, что профессия рабочих не влияет на длительность заболевания.

2. Задания для выполнения практической работы

Задания для выполнения практических заданий содержатся по ссылке:

<https://do.sevsu.ru/mod/assign/view.php?id=567352>

1. **Задание 1.** Выполнить пример 1 пункта 1.

2. **Задание 2.** Требуется определить, зависит ли время работы мобильных телефонов после полной зарядки от производителя. Рассматривается по 5 телефонов от каждого из 4-х различных производителей. Исходные данные приведены в таблице:

Порядковый номер телефона	Производитель 1	Производитель2	Производитель3	Производитель4
1	21,2	25,4	22,4	26
2	19,5	24,3	24	21,7
3	23,1	25,1	21,6	19
4	19,9	24,1	21,5	22,1
5	20,5	24,7	22,6	24,2

3. **Задание 3.** Исходные данные содержатся в файле **Задания_ПР8.xlsx**.

Данные состоят из экзаменационных баллов 42 студентов, посещающих занятия в некотором университете.

Выяснить, отличаются ли оценки студентов по разным экзаменам.

4. **Задание 4.** Выяснить влияние фактора, положенного в основу группировки на значения средних в группах. Вариант задания выбрать из Приложения Е в соответствии с номером студента в списке группы.

4. Содержание отчета

1. Цель работы.
2. Постановка задачи
3. Ход выполнения работы.
4. Анализ результатов (для каждого задания).
5. Общие выводы по работе.

Практическая работа № 9

Анализ взаимосвязей между данными на основе регрессионного анализа

Цель: изучить правила построения линейной регрессионной модели для выявления связей между данными и прогнозирования.

Регрессионный анализ (РА) применяется для выявления влияния одной или нескольких независимых количественных переменных на одну зависимую также количественную переменную.

РА позволяет дать математическое описание зависимости между переменными, построить модель (тренд), отражающий данную зависимость и позволяющий строить прогнозы изменения зависимой переменной.

В практике статистических исследований с помощью РА можно найти ответы на следующие вопросы:

Имеется ли зависимость между успеваемостью студента и количеством пропущенных занятий? Построить вид этой зависимости.

В какой зависимости находятся результаты обучения учащихся от затрат на повышение квалификации преподавателя?

В какой зависимости находятся сумма расходов на отдых в целом и сумма расходов на питание и развлечения?

Чему равен вклад расходов на персонал в изменение объема продаж при фиксированных ценах и расходах на продвижение?

Какие объемы продаж можно ожидать при заданных ценах, расходах на продвижение и расходах на персонал?

В РА могут быть включены только количественные переменные.

Регрессионная модель (РМ) представляет собой функциональную зависимость наблюдаемой зависимой переменной от одной или нескольких независимых переменных (факторов). Функциональная зависимость может быть представлена функцией любого вида, но чаще всего РМ представляется в виде линейного многочлена:

- простая линейная регрессия: $\hat{y} = a_0 + a_1x$;

- множественная линейная регрессия: $\hat{y} = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n$.

Таким образом, целью РА является определение свободного члена (a_0) и коэффициентов регрессии (a_i). Обязательным условием проведением РА является расчет показателей, характеризующих практическую применимость и статистическую значимость результатов анализа и построенной РМ.

К данным показателям относятся:

Коэффициент корреляции (R) — отражает связь между значениями зависимой и независимыми переменными. Значения коэффициента корреляции могут изменяться в интервале от нуля до единицы. При приближении коэффициента к единице, взаимосвязь усиливается. Если R превышает 0,5, то говорят о наличии связи между переменными, и наоборот.

Коэффициент R-квадрат — коэффициент детерминации равен той доле результатов наблюдений, которая объясняется уравнением регрессии.

Значения *R-квадрат* изменяются в интервале от нуля до единицы. Если *R-квадрат* равен 0,5, то только для 50% всех наблюдений справедлива построенная регрессионная модель. Для значимых результатов исследования значение коэффициента *R-квадрат* должно превышать 0,5.

Показатель значимости p - характеризует значимость коэффициентов регрессионной модели, построенной на основе данных опроса респондентов, попавших в выборку, для всей генеральной совокупности. Если $p < 0,05$, то можно говорить о

справедливости регрессионной модели для всей генеральной совокупности. Обратный результат говорит о том, что связь между переменными слабая или не обнаружена.

1. Правила построения РМ.

Используя результаты оценки потребителями 15 магазинов с точки зрения уровня обслуживания и приемлемости цен, проведите регрессионный анализ и выявите зависимость в предпочтениях от приемлемости цен и уровня обслуживания. При оценке магазинов использовалась 10-балльная шкала. 10 - соответствует максимальной оценке, 1 - минимальной.

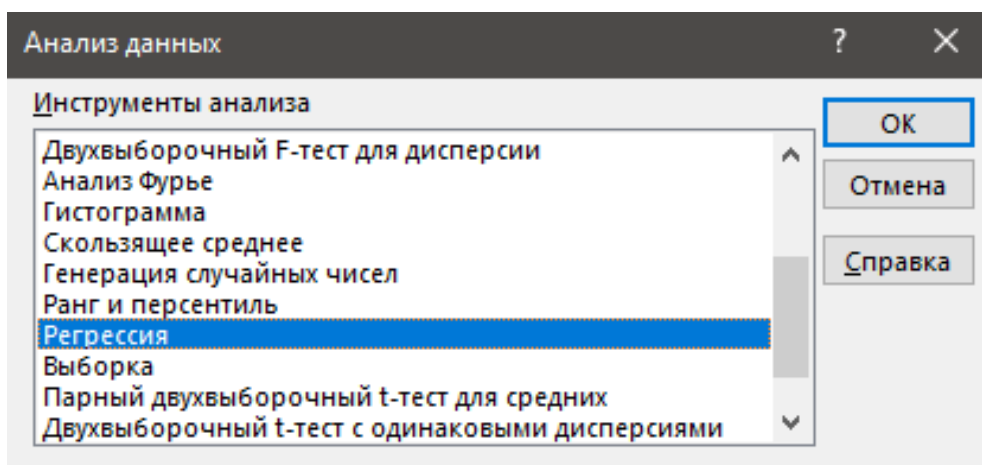
№	Предпочтение	Качество обслуживания	Цена
1	6	5	3
2	9	6	10
3	8	6	4
4	3	2	1
5	10	6	10
6	4	3	1
7	5	4	7
8	2	1	4

№	Предпочтение	Качество обслуживания	Цена
9	10	9	8
10	9	5	10
11	10	8	8
12	2	1	5
13	9	8	5
14	5	3	2
15	3	8	3

1.1 Простая линейная регрессия

Пример 1. Исходные данные содержатся в файле **Задания_ПР9.xlsx**. Построить регрессионную модель зависимости **Предпочтения** магазина от переменной **Качество обслуживания**. Порядок действий при построении РМ представлен на рисунке 9.1.

Шаг 1. Выполнить команду **Данные. Анализ данных. Регрессия**.



Шаг 2. В диалоговом окне **Регрессия** заполнить поля ввода интервалами с входными и выходными значениями, указать ячейку для результата – **Выходной интервал**.

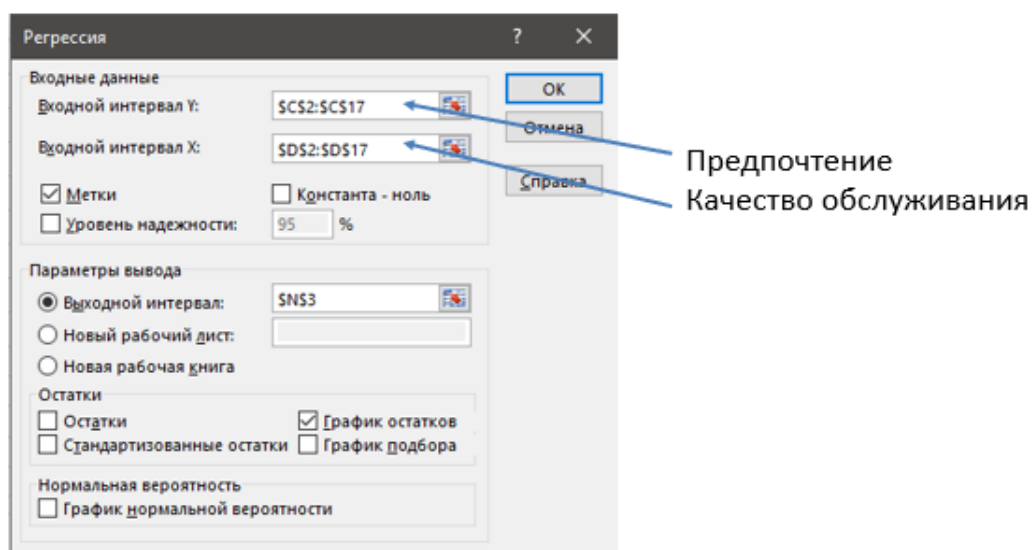


Рисунок 9.1 – Построение РМ в Пакете Анализа ЭТ Excel

В результате получим (рис. 9.2):

ВЫВОД ИТОГОВ	
<i>Регрессионная статистика</i>	
Множественный R	0,751300955
R-квадрат	0,564453125
Нормированный R-квадрат	0,530949519
Стандартная ошибка	2,113562735
Наблюдения	15

Коэффициент корреляции $R=0,75>0,5$ – связь XY – существенная;
 R -квадрат=0,56 модель точно описывает 56% всех наблюдений

Дисперсионный анализ					
	df	SS	MS	F	Значимость F
Регрессия	1	75,26041667	75,26041667	16,84753363	0,00124245
Остаток	13	58,07291667	4,467147436		
Итого	14	133,3333333			

$F_{расч} > F_{крит}$
Значимость 0,001 < 0,05
Модель адекватна

	Коэффициенты	стандартная ошибка	t-статистика	P-Значение	Нижние 95%	Верхние 95%	Нижние 95%	Верхние 95%
Y-пересечение	1,90625	1,208771887	1,577013844	0,138806621	-0,705142897	4,517643	-0,70514	4,517643
Качество обслуживания	0,885416667	0,215714593	4,10457477	0,00124245	0,41939362	1,35144	0,419394	1,35144

Уравнение регрессии: $Y=1,91+0,89x$ p -значение < 0,05
 $t_{расч} > t_{крит}$

Гипотезу о равенстве коэффициента 0 отвергаем. Коэффициент значим.

Рисунок 9.2 – Итоги построения РМ

Построенная модель описывается уравнением регрессии: $Y=1,91+0,89x$, где зависимая переменная Y представляет собой предпочтение покупателей в выборе магазина, а независимая переменная x – качество обслуживания. Свободный член уравнения регрессии, отличный от 0 и равный 1,91, говорит о том, что кроме качества обслуживания имеются и другие факторы, влияющие на предпочтение покупателей. Увеличение коэффициента при переменной **Качество** обслуживания на 1 показывает, что примерно на столько же увеличится и значение переменной **Предпочтение**.

Графическое изображение РМ и правила построения представлены на рисунке 9.3.

Вставка. Диаграммы. Тип «Точечная»

$$Y = 1,91 + 0,89 \cdot X$$

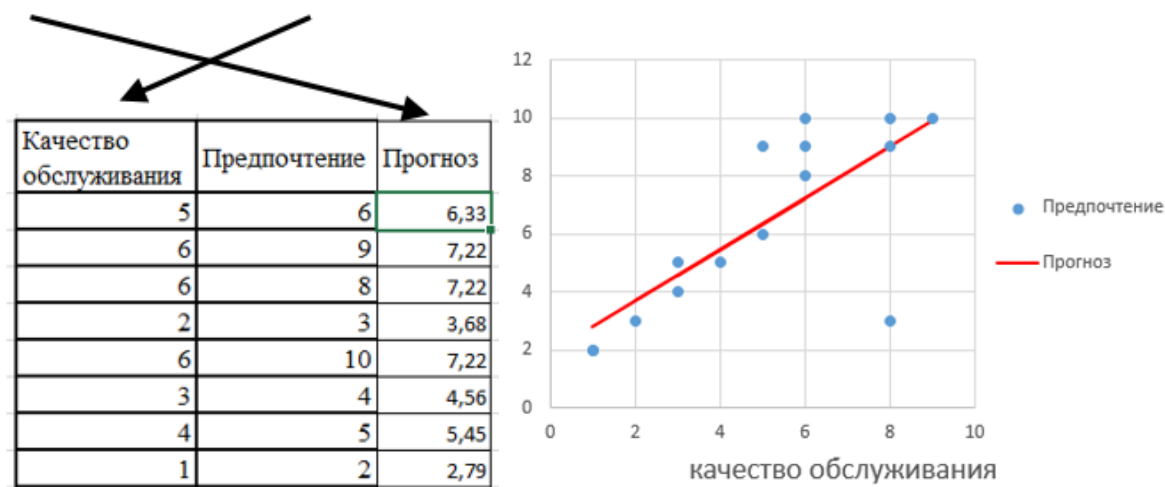


Рисунок 9.3 - Графическое представление РМ

1.2 Множественная линейная регрессия.

Множественная регрессия исследует влияние двух и более факторов на зависимую переменную.

Пример 2. Построить зависимость предпочтения магазина от уровня цен и качества обслуживания одновременно.

Для построения модели, как и в простом случае используются команды Данные. Анализ данных. Регрессия. При заполнении диалогового окна следует выделить входной интервал X, который в данном случае занимает 2 столбца – x1 - уровень цен и x2 - качество обслуживания. В результате получим следующие итоги (рис. 9.4):

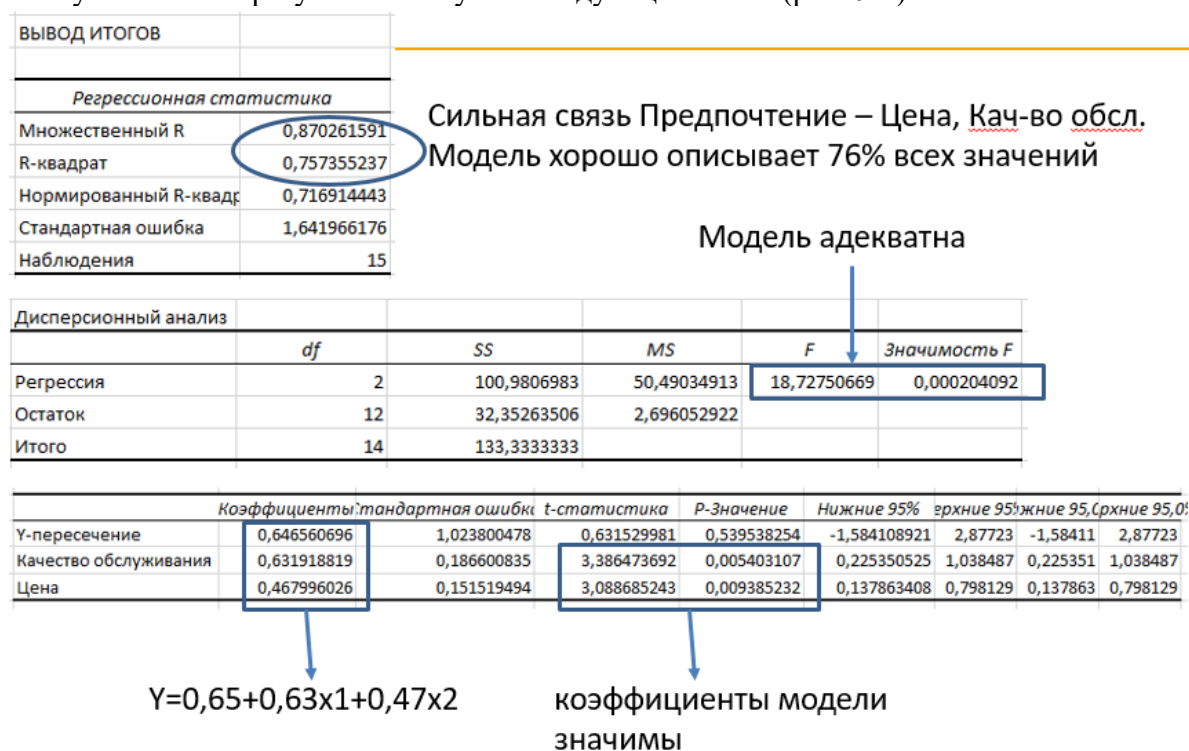


Рисунок 9.4 – Множественная регрессионная модель

При анализе уравнения регрессии $Y=0,65+0,63x_1+0,47x_2$ можно установить, что, кроме рассматриваемых 2-х факторов, существуют и другие факторы, которые оказывают влияние на предпочтения покупателей, но это влияние не велико. Уровень предпочтения цены оказывает меньшее влияние на предпочтения покупателей при выборе магазина (коэффициент при x_1 больше, чем при x_2).

2. Задание для выполнения практической работы.

Материалы для выполнения практических заданий содержатся по ссылке:

<https://do.sevsu.ru/mod/assign/view.php?id=567356>

1. **Задание 1.** Выполнить примеры 1 и 2 методических указаний. Данные наблюдений представлены в файле **Задания_ПР9.xlsx**, Лист **Задания 1_2**.
2. **Задание 2.** На основании регрессионной модели определить влияние уровня цен на предпочтения покупателей при выборе магазина (таблица из методических указаний).
3. **Задание 3.** Определить, как влияют расходы на рекламу спортивной секции на доход от ее деятельности. Данные наблюдений представлены в файле **Задания_ПР9.xlsx**, Лист **Задание 3**.
4. **Задание 4.** На основе регрессионного анализа определить зависимость количества жира в организме – y от переменных x_1 - толщины кожной складки трицепса; x_2 - окружности бедра и x_3 - окружности предплечья. Данные наблюдений представлены в файле **Задания_ПР9.xlsx**, Лист **Задание 4**.

3. Содержание отчета

1. Титульный лист.
2. Цель работы.
3. Постановка задачи
4. Ход выполнения работы с выводами после выполнения каждого задания.
5. Анализ результатов (для каждого задания).
6. Общие выводы по работе.

Библиографический список

1. Толстова Ю. Н. Математическая статистика для социологов : учебник и практикум для вузов / Ю. Н. Толстова. — 2-е изд., испр. и доп. — Москва : Юрайт, 2025. — 196 с. — (Высшее образование). — URL: <https://urait.ru/bcode/558884> (дата обращения: 03.03.2025) — ISBN 978-5-534-20860-3. — Текст : электронный.
2. Макарова Н. В. Статистика в Excel : учеб. пособие / Н. В. Макарова, В. Я. Трофимец. — М. : Финансы и статистика, 2002. — 368 с. — ISBN 5-279-02282-9. — Текст : непосредственный.

ПРИЛОЖЕНИЕ А.1

Ряд распределения дискретной случайной величины

Значения случайной величины	1	2	3	4	5
№ варианта	Вероятность появления случайной величины				
1	0,12	0,23	0,21	0,14	0,3
2	0,21	0,13	0,25	0,15	0,26
3	0,12	0,21	0,11	0,32	0,24
4	0,15	0,32	0,12	0,27	0,14
5	0,26	0,14	0,17	0,26	0,17
6	0,11	0,43	0,13	0,10	0,23
7	0,32	0,14	0,17	0,25	0,12
8	0,18	0,19	0,39	0,09	0,15
9	0,22	0,33	0,17	0,15	0,13
10	0,42	0,05	0,10	0,20	0,23
11	0,16	0,13	0,23	0,16	0,32
12	0,13	0,22	0,14	0,30	0,21
13	0,28	0,12	0,15	0,20	0,25
14	0,31	0,30	0,14	0,10	0,15
15	0,08	0,20	0,25	0,35	0,12
16	0,15	0,27	0,20	0,18	0,20
17	0,42	0,11	0,20	0,15	0,12
18	0,31	0,29	0,12	0,20	0,08
19	0,14	0,23	0,16	0,30	0,17
20	0,21	0,29	0,16	0,15	0,19
21	0,30	0,20	0,15	0,18	0,17
22	0,27	0,31	0,15	0,20	0,07
23	0,13	0,19	0,22	0,20	0,26
24	0,06	0,25	0,31	0,08	0,30
25	0,37	0,21	0,10	0,15	0,17
26	0,24	0,16	0,19	0,32	0,09
27	0,07	0,25	0,16	0,34	0,18
28	0,27	0,14	0,17	0,30	0,12
29	0,10	0,32	0,12	0,27	0,19
30	0,21	0,19	0,17	0,26	0,17

ПРИЛОЖЕНИЕ А.2

1	Xi	8	9	10	11	12	13	14	15	16	17	18	19	20	
	Pi	0,005	0	0,02	0,035	0,045	0,11	0,16	0,22	0,16	0,16	0,07	0,02	0,01	
2	Xi	4	5	6	7	8	9	10	11	12	13				
	Pi	0,01	0,06	0,095	0,2	0,23	0,2	0,095	0,09	0,01	0,01				
3	Xi	4	5	6	7	8	9	10	11	12	13	14			
	Pi	0,015	0,005	0,03	0,075	0,13	0,25	0,155	0,185	0,1	0,04	0,02			
4	Xi	6	7	8	9	10	11	12	13	14	15	16	17		
	Pi	0,02	0,035	0,055	0,145	0,165	0,15	0,22	0,105	0,065	0,03	0,01	0,01		
5	Xi	8	9	10	11	12	13	14	15	16	17	18			
	Pi	0,005	0,015	0,035	0,095	0,14	0,18	0,215	0,19	0,1	0,005	0,02			
6	Xi	4	5	6	7	8	9	10	11	12	13	14	15	16	17
	Pi	0,005	0	0	0	0,02	0,03	0,075	0,145	0,195	0,18	0,19	0,12	0,04	0,01
7	Xi	11	12	13	14	15	16	17	18	19					
	Pi	0,01	0,015	0,04	0,1	0,19	0,27	0,18	0,14	0,055					
8	Xi	15	16	17	18	19	20	21	22						
	Pi	0,025	0,05	0,12	0,205	0,22	0,23	0,135	0,015						
9	Xi	6	7	8	9	10	11	12	13	14	15	16	17		
	Pi	0,005	0	0,01	0,01	0,08	0,11	0,175	0,21	0,205	0,105	0,08	0,02		
10	Xi	12	13	14	15	16	17	18	19	20	21				
	Pi	0,025	0,06	0,11	0,1	0,18	0,26	0,135	0,085	0,035	0,01				
11	Xi	13	14	15	16	17	18	19	20						
	Pi	0,02	0,03	0,07	0,195	0,22	0,28	0,145	0,04						
12	Xi	9	10	11	12	13	14	15	16	17	18				
	Pi	0,015	0,05	0,07	0,11	0,225	0,17	0,21	0,1	0,045	0,005				
13	Xi	9	10	11	12	13	14	15	16	17	18				
	Pi	0,005	0	0,03	0,085	0,085	0,22	0,245	0,245	0,075	0,01				
14	Xi	6	7	8	9	10	11	12	13	14	15	16	17		
	Pi	0,01	0,005	0,045	0,045	0,105	0,23	0,23	0,13	0,095	0,08	0,02	0,01		
15	Xi	4	5	6	7	8	9	10	11	12	13				
	Pi	0,035	0,055	0,08	0,13	0,23	0,215	0,11	0,085	0,05	0,01				
16	Xi	2	3	4	5	6	7	8	9	10	11	12	13	14	
	Pi	0,005	0	0,01	0,015	0,03	0,055	0,165	0,24	0,175	0,145	0,1	0,06	0,01	
17	Xi	6	7	8	9	10	11	12	13	14	15	16	17		
	Pi	0,005	0,015	0,045	0,075	0,165	0,145	0,18	0,14	0,115	0,08	0,03	0,01		
18	Xi	10	11	12	13	14	15	16	17	18	19				
	Pi	0,005	0,02	0,03	0,13	0,145	0,17	0,225	0,14	0,09	0,045				
19	Xi	4	5	6	7	8	9	10	11	12	13	14	15	16	
	Pi	0,005	0,005	0,035	0,04	0,09	0,135	0,165	0,15	0,185	0,08	0,07	0,04	0,01	
20	Xi	3	4	5	6	7	8	9	10	11	12	13	14		
	Pi	0,005	0	0,01	0,01	0,08	0,11	0,175	0,21	0,205	0,105	0,08	0,02		

ПРИЛОЖЕНИЕ А.3

1

33	22	31	21	29
32	25	24	25	33
23	29	18	34	32
26	20	22	32	21
25	20	26	27	22
29	29	26	36	30
19	24	24	23	15

2

40	26	39	39	15
24	11	19	41	13
15	20	17	23	44
37	19	19	28	40
34	34	27	34	38
31	23	15	16	37
38	45	12	37	16

3

28	23	36	39	26
28	20	21	37	32
32	22	21	26	24
30	28	36	31	26
30	27	40	36	27
22	37	24	28	33
28	36	24	37	29

4

24	22	30	26	26
18	23	21	32	31
32	28	17	24	40
20	28	13	17	17
15	26	22	30	17
25	23	28	27	22
27	25	27	34	23

5

18	35	25	44	10
18	21	43	22	17
23	33	34	27	31
21	22	32	31	30
24	40	42	20	29
42	15	31	37	39
32	35	28	18	37

6

35	29	26	21	21
33	23	35	21	21
36	30	28	25	25
22	26	35	38	31
23	35	35	32	31
23	28	35	39	35
25	20	20	38	26

7

27	20	27	23	25
26	23	23	28	38
20	22	17	34	17
17	34	13	22	26
20	24	27	28	21
29	16	27	24	33
26	24	29	23	30

8

26	29	35	31	17
13	13	43	10	32
23	29	36	26	43
36	15	20	31	35
13	28	29	15	28
25	28	27	24	11
40	16	29	14	44

9

30	21	22	33	35
25	30	39	25	31
27	22	39	30	34
24	40	38	23	35
37	24	33	37	39
29	34	33	38	32
27	24	22	38	36

10

23	25	22	29	25
27	25	21	28	20
26	31	27	34	31
31	27	29	26	33
21	25	26	31	21
28	28	31	19	19
24	29	16	21	25

11

22	11	22	27	22
21	45	13	34	24
39	44	18	13	11
15	45	43	32	26
43	16	20	19	11
23	35	16	15	40
42	27	16	35	30

12

37	34	29	23	34
21	30	26	36	24
40	28	31	20	31
29	28	35	35	28
38	37	38	33	30
35	37	21	33	25
32	22	35	38	22

13

27	22	25	29	34
22	29	33	29	23
37	18	22	33	27
21	28	30	27	33
25	23	19	18	27
22	22	17	21	27
23	17	23	30	25

14

28	36	30	14	38
39	43	31	28	44
39	23	26	12	19
14	23	30	40	36
15	19	41	43	16
11	37	18	36	11
34	34	30	44	23

15

23	32	23	35	32
20	22	28	25	37
37	31	31	26	36
27	25	20	21	37
23	34	28	33	33
39	23	23	28	32
37	39	22	24	27

16

23	24	23	13	12
17	24	22	21	33
24	23	26	24	25
20	30	19	28	17
32	28	28	21	29
22	28	24	26	23
16	23	29	18	28

17

26	29	14	18	26
36	34	30	16	24
26	27	33	26	36
28	18	17	25	23
34	19	41	18	23
12	16	30	14	28
34	44	23	20	10

18

20	35	33	38	37
21	38	28	26	25
27	29	32	31	32
33	40	29	26	35
27	23	27	29	30
37	38	21	31	21
34	26	38	20	20

19

28	24	17	19	22
35	31	21	23	25
25	18	27	24	25
24	30	27	32	21
35	26	31	20	30
23	18	27	28	29
33	27	25	24	19

20

19	32	20	39	13
16	39	14	23	25
21	12	35	16	19
14	26	25	16	30
41	33	33	11	13
41	33	23	21	24
40	15	18	41	13

21

25	20	21	25	25
28	24	36	28	33
39	35	32	38	38
35	39	39	20	34
32	33	33	31	28
35	28	23	22	20
33	32	35	22	38

22

27	30	23	26	20
32	28	24	23	20
38	31	27	25	24
19	22	14	22	21
25	28	21	18	24
29	22	25	28	30
18	27	17	34	23

23

34	41	20	23	14
35	19	40	39	37
13	17	15	39	17
15	38	11	26	37
26	16	28	33	17
44	10	10	29	25
28	29	20	12	21

24

26	32	28	21	22
30	26	33	28	29
27	35	32	24	38
21	37	33	36	37
32	34	22	39	27
33	22	37	24	26
38	32	22	29	37

25

20	21	14	23	29
33	31	18	20	24
29	14	21	24	15
17	18	23	32	29
32	33	36	22	33
31	36	32	16	22
16	21	20	31	24

26

35	41	28	24	28
16	30	25	34	32
25	29	37	43	32
16	24	26	41	12
31	16	17	35	24
24	23	20	28	26
24	19	21	15	41

27

25	39	35	35	21
24	26	20	20	38
38	25	30	25	26
38	28	29	23	31
39	37	29	38	33
38	36	31	33	32
25	27	25	29	27

28

24	20	25	24	22
26	29	24	30	21
20	36	22	31	11
19	23	19	28	33
19	29	26	31	21
29	21	27	23	28
19	28	23	17	23

29

33	12	44	33	39
22	45	20	28	12
18	20	18	28	24
25	13	44	15	42
13	32	22	33	30
17	39	23	44	32
40	15	25	33	41

30

36	24	33	20	32
35	36	35	28	21
23	39	24	22	36
38	33	40	21	33
24	40	22	30	39
38	28	24	32	27
37	31	34	27	21

Приложение Б
Индивидуальные варианты для выполнения задания

№	Выборочные значения
1	95,6,21, 77, 62, 46,3,71 ,93 ,68 ,84 ,55 ,64 ,89 ,66 ,26 ,18, 91,7,50
2	68, 11, 65, 92, 94, 27, 76,1,65 ,67, 92, 93, 10, 52, 25, 60,9,33 ,66 ,62
3	70, 15, 8, 25, 42, 43, 35, 27, 45, 18, 9, 40, 34, 61, 32, 5, 38,31, 28, 87
4	18, 85, 81, 60, 53, 27, 54, 50, 40, 7, 68, 34, 9, 64, 65, 5, 11, 31, 61, 62
5	22, 87, 27, 14, 65, 6, 92, 73, 9, 62, 51, 75, 27, 56, 95, 47, 8, 50, 38, 98
6	81, 86, 89, 75, 70,0,73, 86, 52, 43, 38,0,29, 81 ,75, 70, 87, 36 ,46 ,41
7	18, 47,2,46, 48, 50, 23, 94, 59, 23, 22, 98,7,26, 99, 18, 77, 37, 27, 95
8	47, 79,2,5,23, 69, 12, 75, 77, 43, 81, 20, 16, 40, 6,9,41, 46, 95, 37
9	86, 30, 63, 33, 45, 22, 57, 43, 23, 25,6,2,80, 47, 28, 35, 34,1,28, 67
10	61, 14, 76, 88,3,33, 50, 78, 64, 30, 81,9,63, 36, 34, 51, 74, 55,9,74
11	68, 44, 56, 17, 37, 62, 29, 46, 74, 11, 28, 56, 21, 35, 7, 65, 29,48, 85,8
12	79, 97, 8 ,36, 64, 4, 20, 70, 25, 66, 91, 81, 95, 10, 84, 69,4,47, 57, 86
13	26, 14, 29, 19, 87, 50,10,20, 65, 41, 27, 6, 50, 48, 63, 41, 7, 34, 23, 80
14	79, 52, 31, 78 ,18, 6, 63,9,72, 32,4,1,55, 10, 65,6,42, 61, 32, 65
15	12,0,55,22,39,92,93,41,9,72, 5 40, 71, 67, 40, 25, 74, 99, 29,8
16	28, 56,0,30, 98, 81, 80, 37, 20, 26, 24, 57, 97, 47, 14, 96, 14, 48,1,49
17	80,6, 48, 79, 86, 27, 25, 29, 39, 16, 85, 55, 67, 66, 16 ,10, 71, 90, 37,9
18	8,63,2,0,72, 36, 10, 13, 42, 41, 53, 74, 31, 18, 41, 97, 59,1,76, 39
19	74, 63, 3, 63, 80, 15, 59, 71, 66, 86,9,56, 42, 47, 76, 42, 19, 52. 16, 10
20	92,4,52, 19, 16, 71, 91, 64, 80, 18, 77, 52, 27, 20. 83, 63, 62, 67, 7, 40
21	51, 9, 84, 35, 13, 10, 48, 7, 44, 17, 67, 63, 35, 9, 40, 82, 66, 12, 71 21
22	88, 66, 7, 10, 34, 9, 48,6,11, 90, 57, 38, 72, 41, 66, 38, 59, 44,9,43
23	78, 82, 63, 55, 43, 10, 9, 59, 9, 70, 63, 79, 75, 88, 96, 8, 65, 37, 45, 65
24	62, 26, 81, 50,8,72, 36, 75, 96, 38, 28, 83, 33, 69, 5, 90, 59, 46, 39,65
25	75, 57, 67, 18, 84, 87, 24, 50, 18, 19, 5, 32, 2,70, 47, 83, 84, 8, 31, 70
26	36, 10, 13, 42, 41, 53, 68, 34, 9, 64, 65, 5, 11, 31, 61, 62, 26
27	81, 60, 53, 27, 54, 50, 40, 7, 68, 34, 9, 37, 62, 9, 46, 45, 22, 57, 43, 23
28	27, 56, 95, 47, 41, 66, 38, 59, 44,9, 70, 25, 66, 91, 69, 5, 90, 59
29	10, 13, 42, 12, 75, 7,41, 53, 74, 31, 18, 41, 97, 93 ,68 ,4 ,55 ,64 ,89,66
30	9, 48,6,11, 90, 57, 38, 34, 61, 32, 87, 24, 50, 18, 19, 5, 32, 6, 63,7

ПРИЛОЖЕНИЕ В

Критические точки для статистики Колмогорова D_n

Объем выборки n	Уровень значимости α			
	0,10	0,05	0,02	0,01
1	0,95	0,98	0,99	0,995
2	0,78	0,84	0,90	0,93
3	0,64	0,71	0,78	0,83
4	0,57	0,62	0,69	0,73
5	0,51	0,56	0,62	0,67
6	0,47	0,52	0,58	0,62
7	0,44	0,48	0,54	0,58
8	0,41	0,45	0,51	0,54
9	0,39	0,43	0,48	0,51
10	0,37	0,41	0,46	0,49
11	0,35	0,39	0,44	0,47
12	0,34	0,38	0,42	0,45
13	0,33	0,36	0,40	0,43
14	0,31	0,35	0,39	0,42
15	0,30	0,34	0,38	0,40
16	0,29	0,33	0,37	0,39
17	0,29	0,32	0,36	0,38
18	0,28	0,31	0,34	0,37
19	0,27	0,30	0,34	0,36
20	0,26	0,29	0,33	0,35

ПРИЛОЖЕНИЕ Г.1

Таблица значений F-критерия Фишера при уровне значимости $\alpha=0,05$

Уровень значимости $\alpha = 0,05$											
v_1 v_2	1	2	3	4	5	6	7	8	9	10	12
1	161	200	216	225	230	234	237	239	241	242	244
2	18,51	19,00	19,16	19,25	19,30	19,33	19,36	19,37	19,38	19,39	19,41
3	10,13	9,55	9,28	9,12	9,01	8,94	8,88	8,84	8,81	8,78	8,74
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,90
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74	4,68
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63	3,57
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34	3,28
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13	3,07
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97	2,91
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,79
12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,76	2,69
13	4,67	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,60
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,60	2,53
15	4,54	3,68	3,29	3,06	2,90	2,79	2,70	2,64	2,59	2,55	2,48
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42
17	4,45	3,59	3,20	2,96	2,81	2,70	2,62	2,55	2,50	2,50	2,38

ПРИЛОЖЕНИЕ Г.2

Проверка однородности дисперсий

	A	B	C		A	B	C		A	B	C		A	B	C		A	B	C
1	48	39	40	2	40	29	42	3	20	38	33	4	41	30	34	5	36	34	35
	45	47	38		35	42	36		41	35	36		29	44	41		35	34	39
	37	39	39		42	37	40		49	33	31		43	42	50		33	35	37
	39	46	28		45	29	47		30	27	41		43	40	45		36	33	38
	31	41	27		48	28	35		45	38	41		35	39	45		35	38	37
	39	31	44		38	46	42		31	33	42		37	26	36		36	34	39
	34	34	42		39	35	44		30	46	29		47	41	45		34	36	36
	43	39	47		39	44	48		35	38	30		46	39	35		34	35	40
	31	37	51		33	39	38		32	37	43		26	38	41		35	36	37
	43	45	42		32	38	43		43	36	41		42	40	47		35	34	36
	A	B	C		A	B	C		A	B	C		A	B	C		A	B	C
6	45	28	45	7	36	42	44	8	39	35	41	9	42	31	29	10	33	38	33
	36	39	34		43	37	33		41	38	44		42	43	40		34	36	36
	37	35	32		47	40	46		46	39	34		41	48	23		35	36	38
	39	39	44		28	30	35		34	40	38		33	41	21		34	32	36
	33	46	32		43	35	36		51	36	29		33	45	22		35	33	38
	47	40	35		41	31	45		40	32	49		37	43	38		37	37	33
	35	37	23		40	43	38		36	49	40		35	45	33		38	36	29
	37	50	31		36	34	48		25	32	33		34	42	34		38	33	32
	44	33	29		39	36	40		38	46	48		36	33	23		32	30	35
	45	34	25		41	23	36		35	39	35		46	38	30		36	34	36
	A	B	C		A	B	C		A	B	C		A	B	C		A	B	C
11	32	38	35	12	51	47	42	13	39	39	48	14	41	45	32	15	34	32	33
	30	30	41		38	35	51		38	41	42		39	28	50		36	30	36
	43	25	34		37	35	45		35	26	50		38	46	55		33	30	39
	39	37	38		41	42	47		49	34	30		30	39	44		34	29	38
	31	28	33		51	46	39		28	34	45		36	37	39		37	37	38
	43	48	38		32	35	43		49	42	51		39	35	38		37	30	35
	39	42	27		39	40	36		38	39	49		24	40	27		36	39	34
	40	31	37		37	46	41		30	33	39		38	32	43		33	23	33
	27	33	32		41	33	33		28	29	41		41	36	36		33	33	33
	47	34	45		45	34	28		38	26	36		37	42	47		34	40	33
	A	B	C		A	B	C		A	B	C		A	B	C		A	B	C
16	28	39	31	17	31	44	44	18	43	44	35	19	56	34	44	20	34	34	31
	42	32	54		36	46	33		37	43	29		42	44	37		36	37	35
	44	42	37		35	45	39		52	29	42		36	43	37		34	46	39
	33	41	25		43	54	44		42	33	39		34	35	43		35	32	41
	38	39	43		36	49	51		44	46	41		29	39	34		30	32	35
	41	27	37		45	41	30		36	39	36		46	30	31		34	39	35
	48	44	40		41	28	49		43	38	47		42	42	43		35	32	37
	43	37	32		38	44	38		41	41	33		44	31	42		35	45	39
	34	35	35		46	45	52		36	37	38		38	45	34		35	33	36
	28	38	39		26	40	46		40	30	39		29	42	38		34	31	38
	A	B	C		A	B	C		A	B	C		A	B	C		A	B	C
21	36	43	40	22	37	42	50	23	35	40	35	24	37	41	28	25	34	34	31
	38	39	42		46	47	36		36	36	39		30	33	38		36	37	35
	40	41	37		30	43	43		34	22	50		36	37	37		34	46	39
	39	34	54		39	34	48		32	41	41		43	39	41		35	32	41
	29	38	27		42	41	45		40	31	43		35	38	37		30	32	35
	37	39	37		30	38	39		43	44	46		37	47	36		34	39	35
	36	33	37		30	47	41		37	39	33		34	48	39		35	32	37
	39	41	36		31	37	44		30	44	45		43	35	45		35	45	39
	34	39	31		36	45	25		44	27	28		34	42	44		35	33	36
	34	27	34		35	38	48		35	17	42		35	36	37		34	31	38

ПРИЛОЖЕНИЕ Д
Сравнение средних показателей в выборках

	X1	X2	Y	μ_0		X1	X2	Y			X1	X2	Y	μ_0		X1	X2	Y	μ_0
1	48	39	40		2	40	29	42		3	20	38	33		4	41	30	34	
	45	47	38			35	42	36			41	35	36			29	44	41	
	37	39	39			42	37	40			49	33	31			43	42	50	
	39	46	28			45	29	47			30	27	41			43	40	45	
	31	41	27	40		48	28	35	38		45	38	41	38		35	39	45	40
	39	31	44			38	46	42			31	33	42			37	26	36	
	34	34	42			39	35	44			30	46	29			47	41	45	
	43	39	47			39	44	48			35	38	30			46	39	35	
	31	37	51			33	39	38			32	37	43			26	38	41	
	43	45	42			32	38	43			43	36	41			42	40	47	
	X1	X2	Y	μ_0		X1	X2	Y	μ_0		X1	X2	Y	μ_0		X1	X2	Y	μ_0
5	45	28	45		6	36	42	44		7	39	35	41		8	42	31	29	
	36	39	34			43	37	33			41	38	44			42	43	40	
	37	35	32			47	40	46			46	39	34			41	48	23	
	39	39	44			28	30	35			34	40	38			33	41	21	
	33	46	32			43	35	36			51	36	29			33	45	22	
	47	40	35	39		41	31	45	38		40	32	49	37		37	43	38	40
	35	37	23			40	43	38			36	49	40			35	45	33	
	37	50	31			36	34	48			25	32	33			34	42	34	
	44	33	29			39	36	40			38	46	48			36	33	23	
	45	34	25			41	23	36			35	39	35			46	38	30	
	X1	X2	Y	μ_0		X1	X2	Y	μ_0		X1	X2	Y	μ_0		X1	X2	Y	μ_0
9	32	38	35		10	51	47	42		11	39	39	48		12	41	45	32	
	30	30	41			38	35	51			38	41	42			39	28	50	
	43	25	34			37	35	45			35	26	50			38	46	55	
	39	37	38			41	42	47			49	34	30			30	39	44	
	31	28	33			51	46	39			28	34	45			36	37	39	
	43	48	38	36		32	35	43	40		49	42	51	35		39	35	38	38
	39	42	27			39	40	36			38	39	49			24	40	27	
	40	31	37			37	46	41			30	33	39			38	32	43	
	27	33	32			41	33	33			28	29	41			41	36	36	
	47	34	45			45	34	28			38	26	36			37	42	47	
	X1	X2	Y	μ_0		X1	X2	Y	μ_0		X1	X2	Y	μ_0		X1	X2	Y	μ_0
13	28	39	31		14	31	44	44		15	43	44	35		16	56	34	44	
	42	32	54			36	46	33			37	43	29			42	44	37	
	44	42	37			35	45	39			52	29	42			36	43	37	
	33	41	25			43	54	44			42	33	39			34	35	43	
	38	39	43			36	49	51			44	46	41			29	39	34	
	41	27	37	38		45	41	30	39		36	39	36	40		46	30	31	38
	48	44	40			41	28	49			43	38	47			42	42	43	
	43	37	32			38	44	38			41	41	33			44	31	42	
	34	35	35			46	45	52			36	37	38			38	45	34	
	28	38	39			26	40	46			40	30	39			29	42	38	
	X1	X2	Y	μ_0		X1	X2	Y	μ_0		X1	X2	Y	μ_0		X1	X2	Y	μ_0
17	36	43	40		18	37	42	50		19	35	40	35		20	37	41	28	
	38	39	42			46	47	36			36	36	39			30	33	38	
	40	41	37			30	43	43			34	22	50			36	37	37	
	39	34	54			39	34	48			32	41	41			43	39	41	
	29	38	27			42	41	45			40	31	43			35	38	37	
	37	39	37	38		30	38	39	40		43	44	46	35		37	47	36	40
	36	33	37			30	47	41			37	39	33			34	48	39	
	39	41	36			31	37	44			30	44	45			43	35	45	
	34	39	31			36	45	25			44	27	28			34	42	44	
	34	27	34			35	38	48			35	17	42			35	36	37	

ПРИЛОЖЕНИЕ Е

Задания для самостоятельной работы

Задача 1. Четыре бригады выполняли работы по посадке декоративных кустарников. Объем работы, выполненный каждой бригадой за каждый из четырех рабочих дней, приведен в таблице. При уровне значимости $\alpha = 0,05$ требуется выяснить, зависит ли объем выполненных работ от работающей бригады?

День	Бригада (фактор)			
	Бригада 1	Бригада 2	Бригада 3	Бригада 4
1-й	140	150	148	150
2-й	144	149	149	155
3-й	142	152	146	154
4-й	145	150	147	152

Задача 2. Группой из пяти экспертов проведена оценка качества объекта при трех значениях уровня фактора, влияние которого на качество объекта изучается. Значения уровня фактора во время оценки экспертам не известны. При уровне значимости $\alpha = 0,05$ проверить нулевую гипотезу о влиянии фактора на качество объекта. Экспертные оценки уровня качества объекта приведены в таблице:

№	Уровень фактора (Ф)		
	Ф1	Ф2	Ф3
1	18	24	36
2	28	36	12
3	12	28	22
4	14	40	45
5	32	16	40

Задача 3. Используя анализ однофакторной модели, проверить гипотезу о влиянии методик на оценку качества признака при уровне значимости $\alpha = 0,05$.

Исследуется эффективность обучения тремя различными методами. Студентам дается задание изучить тему «Денежная система в Англии». Для этого 10 студентов конспектируют первоисточник, 10 изучают ее по электронному учебнику, 10 - с помощью обучающих компьютерных программ. По окончании их уровень знаний проверяется с помощью теста, состоящего из 100 вопросов. Результаты представлены в таблице:

№	Баллы по тесту для трех методик		
	Первоисточник	Учебник	Компьютер
1	28	39	41
2	33	52	49
3	42	53	56
4	47	54	62
5	48	56	63
6	50	58	64
7	50	59	65
8	51	63	72
9	60	64	77
10	71	77	87

Как в результате проведенного однофакторного дисперсионного анализа можно ответить на следующие вопросы:

- 1) Влияет ли методика изучения темы на результат? Есть ли значимые различия между тремя выборками по уровню усвоения материала?
- 2) Есть ли статистически значимая тенденция возрастания показателей в порядке «первоисточник» - «учебник» - «компьютер»?

Задача 4. В течение шести лет в хозяйстве использовались четыре различных технологии по выращиванию сельскохозяйственной культуры. Необходимо установить при уровне значимости $\alpha = 0,05$ влияние различных технологий на урожайность культуры. Данные наблюдений представлены в таблице:

Год	Технология (фактор А)			
	A1	A2	A3	A4
1	0,9	1	1,3	1,8
2	0,8	0,7	1,5	1,9
3	1	1,3	1,6	1,1
4	1,3	1	1,4	1,4
5	1,4	1,3	1,2	1,3
6	0,8	1,3	1,1	1,9

Задача 5. Требуется оценить влияние квалификации наладчиков (фактор К) на рассеяние диаметров шариков, которые они изготавливают, при уровне значимости $\alpha = 0,05$. Замеры отклонения диаметра от номинала для каждого из пяти наладчиков проводились по 6 раз. Результаты представлены в таблице:

№	Квалификация наладчика				
	K1	K2	K3	K4	K5
1	1,2	0,6	0,9	1,7	1,0
2	1,1	1,1	0,6	1,4	1,4
3	1,0	0,8	0,8	1,3	1,1
4	1,3	0,7	1,0	1,6	0,9
5	1,1	0,7	1,0	1,2	1,2

Проверяется нулевая гипотеза о равенстве математических ожиданий отклонения для всех пяти наладчиков, то есть предполагается, что квалификация наладчика не влияет на точность изготовления шариков.

Задача 6. Компания, работающая на рынке, хочет оценить влияние обучения персонала на уровень продаж. Кадровая служба компании осуществляет обучение сотрудников отдела продаж в группах (групповое обучение) и организует индивидуальное обучение. Результаты продаж восьми сотрудников после группового и индивидуального обучения приведены в таблице А.8. Используйте однофакторный дисперсионный анализ, задав уровень значимости $\alpha = 0,05$, чтобы выяснить, влияет ли обучение персонала на уровень продаж.

Объем продаж за различные периоды			
Статистический период	Объем продаж до обучения (млн. руб.)	Объем продаж после группового обучения	Объем продаж после индивидуального обучения
1	203	217	213
2	223	230	231
3	206	221	221
4	205	227	222
5	234	218	229
6	204	231	235
7	219	224	213
8	210	225	228
9	233	211	214
10	223	229	225

Задача 7. Используя анализ однофакторной модели, проверить гипотезу о влиянии наглядности задания на оценку качества признака при уровне значимости $\alpha = 0,05$.

Исследуется эффективность методики выдачи задания тремя различными способами. Студентам дается задание разного уровня наглядности. Для этого 10 студентов получают задание от преподавателя устно, 10 - письменно, 10 - в виде текста с иллюстрациями. По окончании их уровень знаний проверяется с помощью теста, состоящего из 100 вопросов. Результаты представлены в таблице:

№	Баллы по тесту для трех методик		
	Устно	Письменно	Текст с иллюстрациями
1	28	39	41
2	33	52	49
3	42	53	56
4	47	54	62
5	48	56	63
6	50	58	64
7	50	59	65
8	51	63	72
9	60	64	77
10	71	77	87

Как в результате проведенного однофакторного дисперсионного анализа можно ответить на вопрос: влияет ли методика выдачи задания на результат? Есть ли значимые различия между тремя выборками по уровню усвоения материала?

Задача 8. Используя анализ однофакторной модели, проверить гипотезу о влиянии технологии чистовой обработки детали на точность ее изготовления при уровне значимости $\alpha = 0,05$.

Возможно применение трех видов технологии чистовой обработки детали. Исследуется влияние технологии чистовой обработки детали на точность ее изготовления. Проводятся замеры точности обработки детали, по 4 замера отклонения размера детали от номинала в мкм (для каждого вида технологии). Результаты представлены в таблице:

Номер замера	Вид технологии		
	1	2	3
1	1	2	3
2	2	1	2
3	2	3	2
4	1	2	3

Задача 9. Четыре группы студентов на уборке картофеля за четыре дня выполнили объем работы, приведенный в таблице. При уровне значимости $\alpha = 0,05$ требуется выяснить, зависит ли объем выполненных работ от работающей группы.

Дата	Количество убранных рядков картофеля за рабочий день			
	Группа 1	Группа 2	Группа 3	Группа 4
20 августа	140	150	148	150
21 августа	144	149	149	155
22 августа	142	152	146	154
23 августа	145	150	147	152

Задача 10. Достоверны ли при уровне значимости $\alpha = 0,05$ различия приростов бычков в четырех экспериментальных группах по пять животных, данные о которых приведены в таблице?

Группа	№ животного				
	1	2	3	4	5
1	40	24	46	20	35
2	29	27	20	39	45
3	11	31	17	37	39
4	17	21	38	33	21

Задача 11. На одной из опытных станций испытывались шесть местных сортов пшеницы. Результаты учета урожайности приведены в таблице. Используя однофакторный дисперсионный анализ, выясните, достоверны ли различия в урожайности между сортами при уровне значимости $\alpha = 0,05$.

№ сорта	Повторение			
	1	2	3	4
1	26,1	29,2	30,0	27,3
2	25,0	24,3	28,5	29,0
3	27,2	26,4	31,0	26,4
4	23,6	27,2	25,2	24,8
5	30,0	33,0	36,0	29,8
6	23,0	26,0	26,0	24,8

Задача 12. Результаты оценки влияния различных норм минеральных удобрений на урожайность яровой пшеницы в полевом опыте приведены в таблице. Используя однофакторный дисперсионный анализ, выясните достоверность и степень влияние минеральных удобрений на урожайность при уровне значимости $\alpha = 0,05$.

Норма удобрений, кг д. в./га	Повторение			
	1	2	3	4
0 (Без удобрений)	16,2	15,8	14,0	15,1
Удобрение 1	17,5	18,9	16,7	18,2
Удобрение 2	18,9	18,7	19,3	20,6
Удобрение 3	26,3	24,5	24,3	25,5
Удобрение 4	23,0	23,9	24,2	23,9

Задача 13. Изучается, как влияет тип магазина на товарооборот. Данные о зафиксированном товарообороте (в млн руб.) в магазинах трех типов за каждые 8 месяцев работы приведены в таблице. Используя однофакторный дисперсионный анализ, выясните достоверность влияния тип магазина на товарооборот при уровне значимости $\alpha = 0,05$.

Магазин 1	Магазин 2	Магазин 3
19	20	16
23	20	15
26	32	18
18	37	26
20	40	19
20	24	17
18	22	19
35	18	18

Задача 14. Крупной организации для принятия решения о закупках лицензионного программного обеспечения (ПО) необходимо выяснить значимость выбора поставщика с точки зрения возможности сокращения при этом издержек. Требуется выяснить, насколько выбор поставщика комплекта лицензионного ПО влияет на издержки предприятия. В шести классах градации фактора П (1-й поставщик, 2-й поставщик и т.д.) собраны данные о потенциальных затратах на покупку 100 единиц продукции в тысячах рублей в течение 4 лет. Данные о затратах за год (в тыс. руб.) приведены в таблице. Используйте однофакторный дисперсионный анализ, задав уровень значимости $\alpha = 0,05$, чтобы выяснить, влияет ли выбор поставщика на размер потенциальных затрат.

Поставщик	2018	2019	2020	2021
1-й	72,1	75,5	72,9	76,0
2-й	78,9	82,7	73,9	81,8
3-й	72,5	70,1	73,7	75,3
4-й	77,5	74,1	72,7	74,2
5-й	77,0	82,8	85,5	86,0
6-й	75,6	80,5	80,7	78,4

Задача 15. Менеджер по продажам в сети супермаркетов хочет знать, влияет ли расположение рекламных щитов на объем продаж товара. Для каждого из трех видов щитов отобрано случайным образом по 6 магазинов, расположенных в соответствующем районе. Объемы продаж за месяц (в млн руб.) приведены в таблице. Используйте однофакторный дисперсионный анализ, задав уровень значимости $\alpha = 0,05$, чтобы выяснить, влияет ли расположение рекламных щитов на объем продаж товара.

Расположение щитов		
1	2	3
8,1	3,9	4,8
7,5	4,4	6,0
5,8	2,8	4,4
6,6	3,4	5,8
5,9	4,8	6,2
4,8	3,6	4,9

Задача 16. Деталь может изготавливаться из трех видов материалов. Производитель проверяет влияние на износостойкость детали материала, из которого она изготовлена. Получены данные по износостойкости пяти деталей для каждого материала: время работы детали до износа, тыс. час. Результаты испытаний приведены в таблице. Используйте однофакторный дисперсионный анализ, задав уровень значимости $\alpha = 0,05$, чтобы проверить влияние на износостойкость детали материала (три вида), из которого она изготовлена.

Материал 1	1,25	1,32	1,28	1,26	1,29
Материал 2	1,12	1,15	1,26	1,19	1,21
Материал 3	1,32	1,33	1,34	1,29	1,30

Задача 17. Требуется оценить эффективность нескольких различных способов действий, направленных на достижение одной цели - получение максимально дешевого кредита на покупку автомобиля. Для этого рассмотрим данные о годовой процентной ставке на кредиты для новых автомобилей в девяти самых крупных банках шести городов США. Данные представлены в таблице. Проверяется нулевая гипотеза о независимости в среднем ставки от города при уровне значимости $\alpha = 0,05$. Выполнить однофакторный дисперсионный анализ. Проверяется нулевая гипотеза о независимости в среднем ставки от города.

Данные					
Atlanta	Chicago	Houston	Memphis	New York	Philadelphia
13,75	14,25	14	15	14,5	13,5
13,75	13	14	14	14	12,25
13,5	12,75	13,51	13,75	14	12,25
13,5	12,5	13,5	13,59	13,9	12
13	12,5	13,5	13,25	13,75	12
13	12,4	13,25	12,97	13,25	12
13	12,3	13	12,5	13	12
12,75	11,9	12,5	12,25	12,5	11,9
12,5	11,9	12,5	11,89	12,45	11,9

Задача 18. Используя анализ однофакторной модели, проверить гипотезу о влиянии методик на оценку качества признака при уровне значимости $\alpha = 0,05$.

Исследуется эффективность обучения тремя различными методами. Студентам дается задание изучить тему «Денежная система в Англии». Для этого 10 студентов конспектируют первоисточник, 10 изучают ее по электронному учебнику, 10 - с помощью обучающих компьютерных программ. По окончании их уровень знаний проверяется с помощью теста, состоящего из 100 вопросов. Результаты представлены в таблице:

№	Баллы по тесту для трех методик		
	Первоисточник	Учебник	Компьютер
1	28	39	41
2	33	52	49
3	42	53	56
4	47	54	62
5	48	56	63
6	50	58	64
7	50	59	65
8	51	63	72
9	60	64	77
10	71	77	87

Задача 18. Требуется оценить влияние квалификации наладчиков (фактор К) на рассеяние диаметров шариков при уровне значимости $\alpha = 0,05$. Замеры отклонения диаметра от номинала для каждого из шести наладчиков проводились по 6 раз. Результаты представлены в таблице:

№	Квалификация наладчика					
	K1	K2	K3	K4	K5	K6
1	1,2	0,6	0,9	1,7	1,0	1,4
2	1,1	1,1	0,6	1,4	1,4	1,3
3	1,0	0,8	0,8	1,3	1,1	1,1
4	1,3	0,7	1,0	1,6	0,9	1,6
5	1,1	0,7	1,0	1,2	1,2	1,3
6	0,8	0,9	1,1	1,3	1,5	1,5

Проверяется нулевая гипотеза о равенстве математических ожиданий отклонения для всех шести наладчиков, то есть предполагается, что квалификация наладчика не влияет на точность изготовления шариков.

Задача 19. Изучается, как влияет тип магазина на товарооборот. Данные о зафиксированном товарообороте (в млн руб.) в магазинах трех типов за каждые 8 месяцев работы приведены в таблице. Используя однофакторный дисперсионный анализ, нужно

выяснить достоверность влияния тип магазина на товарооборот при уровне значимости $\alpha = 0,05$.

Магазин 1	Магазин 2	Магазин 3
19	20	19
23	20	23
26	32	18
18	37	26
20	40	20
20	24	20
18	22	18
35	18	35

Задача 20. Используя анализ однофакторной модели, проверить гипотезу о влиянии технологии чистовой обработки детали на точность ее изготовления при уровне значимости $\alpha = 0,05$.

Возможно применение трех видов технологии чистовой обработки детали. Исследуется влияние технологии чистовой обработки детали на точность ее изготовления. Проводятся замеры точности обработки детали, по 4 замера отклонения размера детали от номинала в мкм (для каждого вида технологии). Результаты представлены в таблице:

Номер замера	Вид технологии		
	1	2	3
1	1	2	3
2	2	1	2
3	2	3	2
4	1	2	3

СТАТИСТИЧЕСКИЙ ПАКЕТ SPSS ДЛЯ СОЦИАЛЬНЫХ НАУК

Методические указания

Часть I

Математические методы в социологических исследованиях

Составитель: Балакирева Ирина Аркадьевна

В авторской редакции

Изд. № 106/2025. Объем 4,25 п. л. Усл. печ. л. 3,95. Уч.-изд. л. 4,17.
Издательский центр ФГАОУ ВО «Севастопольский государственный университет»