МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное
учреждение высшего образования
«Воронежский государственный технический университет»

Кафедра управления

СТАТИСТИКА

МЕТОДИЧЕСКИЕ УКАЗАНИЯ

к выполнению лабораторных работ для студентов
направления 38.03.02 «Менеджмент» (профиль «Менеджмент организаций»)
всех форм обучения

Воронеж 2022

УДК 657(07)
ББК 65.052.9(2)2я7

Составители: канд. физ.-мат. наук, доц. С. И. Моисеев,
канд. техн. наук О .С. Перевалова

Статистика: методические указания к выполнению лабораторных работ для студентов направления 38.03.02 «Менеджмент» (профиль «Менеджмент организаций») всех форм обучения / ФГБОУ ВО «Воронежский государственный технический университет»; сост.: С. И. Моисеев, О. С. Перевалова. Воронеж: Изд-во ВГТУ, 2022. 29 с.

В методических указаниях приводятся сведения, разъясняющие студентам процесс выполнения статистических расчетов с применением программ для работы с электронными таблицами MS Excel.

Предназначены для студентов, обучающихся по направлению 38.03.02 «Менеджмент» (профиль «Менеджмент организаций») всех форм обучения.

Методические указания подготовлены в электронном виде и содержатся в файле МУ_ЛР_Стат_МО.pdf.

Ил. 2. Табл. 24. Библиогр.: 11 назв.

УДК 657(07)
ББК 65.052.9(2)2я7

Рецензент – Л. В. Шевченко, канд. техн. наук, доц.
кафедры управления ВГТУ

Издается по решению редакционно-издательского совета Воронежского государственного технического университета

ВВЕДЕНИЕ

Целями дисциплины «Статистика» являются:

– овладение студентами статистической методологией и ее применение при всестороннем исследовании социально-экономических процессов, протекающих в организациях, на предприятиях, фирмах и в отраслях национальной экономики; овладение совокупностью математических методов, используемых для количественной оценки экономических явлений и процессов;

– обучение эконометрическому моделированию, т. е. построению экономико-математических моделей, параметры которых оцениваются средствами математической статистики;

– обучение эмпирическому выводу экономических законов; подготовку к прикладным исследованиям в области экономики.

Задачи изучения дисциплины «Статистика» перечислены ниже:

– получение студентами знаний и навыков формирования статистической информации, ее использования для получения обоснованной системы показателей, с помощью которых выявляются имеющиеся резервы роста эффективности производства и прогноз тенденций его развития;

– научить студентов использовать данные наблюдения для построения количественных зависимостей для экономических соотношений, для выявления связей, закономерностей и тенденций развития экономических явлений;

– выработать у студентов умение формировать экономические модели, основываясь на экономической теории или на эмпирических данных, оценивать неизвестные параметры в этих моделях, делать прогнозы и оценивать их точность, давать рекомендации по экономической политике и хозяйственной деятельности.

Результатом освоения дисциплины является освоение компетенций предусмотренных учебным планов направлению подготовки 38.03.02 «Менеджмент» профиль «Менеджмент организаций» всех форм обучения.

Лабораторные занятия – это одна из разновидностей практического занятия, являющаяся эффективной формой учебных занятий в организации высшего образования. Лабораторные занятия имеют выраженную специфику в зависимости от учебной дисциплины, углубляют и закрепляют теоретические знания. На этих занятиях студенты осваивают конкретные методы изучения дисциплины. Лабораторные занятия дают наглядное представление об изучаемых явлениях и процессах, студенты осваивают постановку и ведение эксперимента, учатся умению наблюдать, оценивать полученные результаты, делать выводы и обобщения. Следовательно, ведущей целью лабораторных работ является научить студентов решать практические задачи.

ЛАБОРАТОРНАЯ РАБОТА 1
ДИСПЕРСИОННЫЙ АНАЛИЗ

Цель работы: изучение основ различных видов дисперсионного анализа с помощью программы для работы с электронными таблицами MS Excel.

Время выполнения работы: 4 часа.

Ход работы

Дисперсионный анализ – статистический метод, позволяющий анализировать влияние различных факторов на исследуемую переменную. *Однофакторный дисперсионный анализ* выявляет влияние некоторого фактора, под действием которого общая выборка делится на m частных выборок (групп). Рассмотрим пример.

ПРИМЕР 1. В институте 6 факультетов. Ставиться задача, на уровне значимости α=0,05 определить, различаются или нет уровни подготовки по иностранным языкам на факультетах, то есть влияет ли фактор «Факультет» на уровень подготовки. Для этой цели были протестированы студенты всех факультетов по иностранному языку, результаты тестирования (баллы) приведены в табл. 1.

Таблица 1

Исходные данные

Факультет 1	43	81	75	94	64	51	34	84	39	69	59	83
Факультет 2	63	82	61	78	94	58	84	79	43	96	89	
Факультет 3	99	50	54	92	33	90	56	46	75	55	35	85
Факультет 4	78	57	52	31	69	51	48	36	94	75		
Факультет 5	37	74	33	66	33	79	81	75	33	87	83	92
Факультет 6	80	58	70	93	56	82	35	53	99	59	48	

Вводим эту таблице вместе с подписями в Excel в ячейки от A1 до шестой строки и столбца M. Затем вызываем надстройку «Анализ данных».

В окне «Анализ данных» нужно выбрать пункт «Однофакторный дисперсионный анализ». В открывшемся окне в поле «Входной интервал» делаем ссылку на диапазон A1-M6, группирование – «по строкам», ставим галочку напротив «Метки в первом столбце», альфа – 0,05, в разделе «Параметры вывода» ставим точку рядом с «Выходной интервал» и в поле рядом делаем ссылку на ячейку, с которой начнется вывод данных, например с A9, нажимаем ОК.

На рабочем листе появится таблица с результатами однофакторного дисперсионного анализа. Она состоит из двух частей. В первой «Итоги» приведены основные статистические показатели по выборкам: их объем, сумма элементов, среднее и дисперсия. Во второй «Дисперсионный анализ» - непосредственно расчет критерия. В строках «Между группами» и «Внутри групп» приведены остаточные суммы ($\boldsymbol{SS}$), степени свободы ($\boldsymbol{df}$) и дисперсии ($\boldsymbol{MS}$). Далее приведена F-статистика (большая дисперсия на меньшую), критическое значение уровня значимости (*P-значение*) и F-критическое. Если F-статистика больше F-критического, или критическое значение уровня значимости меньше заданного α, то средние в выборках (группах) различаются и фактор влияет на показатель. У на этого не наблюдается, следовательно, уровень подготовки на всех факультетах можно считать одинаковым.

Задание 1. В школе 6 пятых классов. Психологу ставиться задача, определить, одинаковый ли средний уровень ситуативной тревожности в классах. Для этого были проведены тесты тревожности среди школьников, результаты которых следующие (табл. 2).

Таблица 2

Исходные данные

5 «А» класс	54	32	24	33	49	24	26	32	50	52	33	32	36
5 «Б» класс	64	44	33	50	35	40	66	38	46	42	44	68	43
5 «В» класс	30	39	48	30	31	21	39	19	23	40	43	40	
5 «Г» класс	36	24	46	19	47	47	34	40	32	22	17	20	
5 «Д» класс	51	44	38	55	28	28	29	41	39	25	61		
5 «Е» класс	27	56	49	37	21	28	47	30	44	48	28	25	21

Проверить на уровне значимости α=0,02 предположение, что средняя ситуативная тревожность в классах не различается.

Задание 2. Студентов разделили на 8 групп. При этом предполагается, что средний уровень знаний в группах равный. Результаты тестирования по дисциплине представлены в табл. 3.

Таблица 3

Исходные данные

Группа	**Обучающийся**														
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
Группа 1	51	51	51	51	51	51	51	51	51	51	51	51	51	51	51
Группа 2	63	63	63	63	63	63	63	63	63	63	63	63	63	63	
Группа 3	49	49	49	49	49	49	49	49	49	49	49	49	49	49	49
Группа 4	54	54	54	54	54	54	54	54	54	54	54	54	54		
Группа 5	56	56	56	56	56	56	56	56	56	56	56	56	56	56	
Группа 6	57	57	57	57	57	57	57	57	57	57	57	57	57	57	57
Группа 7	65	65	65	65	65	65	65	65	65	65	65	65			
Группа 8	42	42	42	42	42	42	42	42	42	42	42	42	42	42	

Проверить на уровне значимости α=0,05 предположение, что средняя подготовленность обучающихся в группах не различается.

Рассмотрим теперь пример *двухфакторного дисперсионного анализа*. В этом случае методика расчета та же, но общая выборка распадается на группы под влиянием двух факторов, влияние одного из них сгруппируем в строках, а второго – в столбцах. Рассмотрим решение на примере.

ПРИМЕР 2. Предположим теперь, что на качество освоения иностранного языка может влиять не только факультет (которых шесть), но и сам язык (которых три: английский, немецкий и французский). Тогда с каждого факультета были взяты по 5 студентов, изучающих разные языки. Выборки результатов тестов (баллы) приведены в табл. 4.

Таблица 4

Исходные данные

ЯЗЫК	Факультет 1	Факультет 2	Факультет 3	Факультет 4	Факультет 5	Факультет 6
Английский	54	56	50	48	89	85
	33	97	98	82	34	72
	78	84	54	34	46	78
	93	65	67	44	50	40
	85	86	72	31	91	49
Немецкий	81	89	82	67	77	35
	44	39	85	35	70	73
	91	39	55	38	56	47
	72	43	69	69	49	82
	89	97	49	78	67	68
Французский	92	74	56	86	57	71
	91	83	88	85	73	89
	84	97	89	65	80	85
	65	55	69	54	60	75
	40	89	95	88	83	49

Нужно проверить следующие гипотезы (на уровне значимости α=0,05):

1. Влияет ли факультет на уровень подготовки.
2. Влияет ли язык на уровень подготовки.
3. Влияют ли друг на друга (коррелируют) язык и факультет.

Переходим на новый рабочий лист. Вводим данные из таблицы вместе с подписями в ячейки A1-G16. В первом столбце группировать ячейки не нужно, просто введите подписи «английский» в A2, «немецкий» в A7 и «французский» в A12. Вызывает надстройку «Анализ данных» и в ней – «Двухфакторный дисперсионный анализ с повторениями». В открывшемся окне в поле «Входной интервал» делаем ссылку на диапазон A1-G16, в поле «Число строк для выборки» вводим 5, альфа – 0,05, в разделе «Параметры вывода» ставим точку рядом с «Выходной интервал» и в поле рядом делаем ссылку на ячейку, с которой начнется вывод данных, например с A18, нажимаем ОК.

На рабочем листе появиться область с результатами двухфакторного дисперсионного анализа. Она состоит из нескольких таблиц, в которых приведены основные статистические показатели для факультетов и языков. В последней таблице «Дисперсионный анализ» приведены результаты расчета критерия. Первая строка «Выборка» отображает результаты по языкам. Видно, что F-статистика больше, чем F-критическое и критический уровень значимости P-значение меньше заданного 0,05. Следовательно, средние результаты теста для языков значимо различаются и фактор «Язык» влияет на качество освоения языков.

Во второй строке «Столбцы» отображаются результаты по факультетам. Видно, что F-статистика меньше, чем F-критическое и критический уровень значимости P-значение больше заданного 0,05. Следовательно, средние результаты теста для факультетов не различаются и фактор «Факультет» не влияет на качество освоения языков.

В третьей строке «Взаимодействие» отображаются результаты выявления зависимости факторов «Язык» и «Факультет» друг на друга. Видно, что F-статистика меньше, чем F-критическое и критический уровень значимости P-значение больше заданного 0,05. Следовательно, факторы «Язык» и «Факультет» не влияют по качеству освоения языков друг на друга.

Задание 3. Ставиться задача, определить, влияет ли тип темперамента и образование на производительность труда рабочих предприятия. Для этого были отобраны по 4 рабочих каждого типа темперамента образования и измерены производительности труда. Проверить на уровне значимости α=0,02 влияние темперамента и образование на производительность труда и друг на друга (табл. 5).

Таблица 5

Исходные данные

Образование / Темперамент	Высшее	Средне-профес.	Среднее общее	Среднее полное	Образование / Темперамент	Высшее	Средне-профес.	Среднее общее	Среднее полное
Холерик	386	498	400	347	Флегматик	387	383	304	335
	240	334	218	456		263	492	231	422
	209	389	230	258		338	295	237	262
	403	327	247	420		397	315	351	452
Сангвиник	371	251	368	270	Меланхолик	366	370	460	359
	339	352	469	327		261	231	421	229
	298	258	260	422		357	343	283	345
	322	299	440	372		275	241	372	206

Задание 4. Менеджер распределяет группы из 5 сотрудников по 6 районам с целью выполнения 4 видов работ. Имеются данные о предыдущих показателях эффективности (процент прибыли от работы) этих сотрудников в районах по каждой работе. Они указаны в таблице. Проверить на уровне значимости α=0,05 влияние района и вида работ на эффективность, а также влияние друг на друга района и эффективности (табл. 6).

Таблица 6

Исходные данные

Работа / Район	Работа1	Работа2	Работа3	Работа4	Работа / Район	Работа1	Работа2	Работа3	Работа4
Район 1	17,7	27,3	24,9	30,3	Район 4	36,7	23,2	24,5	25,7
	31,2	28,7	25,4	33,9		27,0	26,1	33,2	21,8
	25,0	16,1	21,5	26,9		22,8	24,4	31,0	27,5
	19,3	23,5	22,1	27,2		26,0	18,9	15,4	29,0
	19,5	29,5	17,6	21,3		19,0	25,3	22,4	32,0
Район 2	17,9	16,7	19,9	21,3	Район 5	12,8	16,0	16,2	29,9
	14,3	18,3	12,8	21,0		10,9	4,0	23,6	11,5
	13,4	23,1	16,1	15,9		21,5	30,4	11,8	15,4
	17,6	15,2	20,5	18,0		27,8	24,3	16,0	15,9
	14,3	17,6	15,9	20,7		19,4	21,7	14,2	24,4
Район 3	34,6	25,0	27,0	20,6	Район 6	29,6	26,7	27,5	33,1
	34,5	27,5	10,3	31,1		28,1	23,9	29,5	31,1
	27,1	21,6	21,4	13,2		32,1	30,8	28,4	23,5
	26,7	18,2	32,8	20,3		29,0	36,2	24,6	35,3
	18,2	37,9	15,6	26,9		31,1	30,9	32,4	27,0

Содержание отчета о выполнении работы:

1. Название и цель работы;
2. Результаты выполненных заданий;
3. Выводы и рекомендации (при необходимости).

ЛАБОРАТОРНАЯ РАБОТА 2
РАНГОВЫЙ КРИТЕРИЙ ВИЛКОКСОНА

Цель работы: приобретение практических навыков сравнения средних значений показателя в двух группах.

Время выполнения работы: 2 часа.

Ход работы

Критерий Вилкоксона, который еще называют критерием Манна и Уитни, является аналогом критерия Стьюдента и позволяет сравнить средние значения показателя в двух группах. Однако, данный критерий не требует, чтобы распределение показателя было нормальным и его можно использовать для любых выборок.

Пример 1. Психолог разработал методику, увеличивающую скорость реакции и как следствие производительность труда рабочих на сборочном конвейере крупного машиностроительного предприятия. Для обоснования эффективности своей методики им были отобраны 2 группы рабочих численностью 12 и 13 человек. В первой группе методика, повышающая скорость реакции не проводилась, а во второй проводилась. Затем путем тестирования были измерены скорости реакции в обеих группах. Результаты представлены в табл. 7.

Таблица 7

Исходные данные

1 группа	24	26	22	24	20	23	21	27	23	25	28	25	
2 группа	28	31	26	24	32	29	30	32	24	29	33	24	31

Необходимо проверить гипотезу об однородности уровня скорости реакции в обоих группах, то есть об одинаковости характеристик положения на уровне значимости α=0,05.

Открываем новый рабочий лист Excel и вводим в A1 подпись «Группа 1», в B1-M1 результаты теста для первой группы, в A2 вводим «Группа 2» и в B2-N2 результаты теста для второй группы. Находим порядковые номера в общей, смешанной группе каждого значения, если расположить их в порядке возрастания, то есть ***ранги***. Для этого служит функция РАНГ, категория «Статистические». В ячейку A3 делаем подпись «Порядок 1». Затем ставим курсор в B3, вызываем мастер функций $\boxed{f_x}$, выбираем категорию «Статистические» и функцию «РАНГ», в открывшемся окне ставим курсор в поле «Число», обводим курсором ячейки B1-M1, ставим курсор в поле «Массив» (или «Ссылка в других версиях Excel), обводим курсором ячейки B1-N2, ставим курсор в поле «Порядок» и вводим 1, чтобы указать, что элементы упорядочены по возрастанию и нажимаем «ОК». В ячейке B3 появился порядковый номер первого числа первой группы 6. Но нам нужно вывести порядковые номера всех чисел из первой группы. Для этого обводим мышкой по центру ячеек B3-M3, выделяя их и нажимаем клавишу F2, затем нажимаем и удерживаем три клавиши в следующей последовательности: “Ctrl”, “Shift” и “Enter”. Получили порядковые номера всех элементов первой группы в общем вариационном ряду.

Проделываем ту же процедуру для второй группы. В ячейку A4 делаем подпись «Порядок 2». Ставим курсор в B4, вызываем функцию «РАНГ», в открывшемся окне в поле «Число», обводим курсором ячейки B2-N2, переводим курсор в поле «Массив» («Ссылка»), обводим курсором ячейки B1-N2, в поле «Порядок» и вводим 1 и нажимаем «ОК». Затем обводим мышкой ячейки B4-N4, выделяя их и нажимаем клавишу F2, затем “Ctrl”, “Shift” и “Enter”.

Согласно методики расчета критерия, если несколько элементов вариационного ряда равны по величине, то каждый элемент имеет один и тот же ранг, равный среднеарифметическому их порядковых номеров. Однако Excel при расчете ранга это правило не выполняет. Для устранения этой проблемы вводим поправочный коэффициент, который рассчитывается по формуле $\frac{n+1-R^{+}-R^{-}}{2}$, где n – число элементов в группе, R^+ - порядковый номер при упорядочении по возрастанию а R^- - порядковый номер при упорядочении по убыванию.

Ставим курсор в A5 и вводим подпись «Поправка 1», затем в B5 вводим **=(СЧЁТ(B1:N2)+1-РАНГ.СР(B1:M1;B1:N2;0)-РАНГ.СР(B1:M1;B1:N2;1))/2**. При вводе формулы ссылки на диапазон ячеек B1:N2 и B1:M1 вводятся в английской раскладке клавиатуры, причем при их вводе можно просто обвести соответствующий диапазон от B1 до N2 или от B1 до M1 мышью. Затем обводим мышкой ячейки B5-M5 и нажимаем клавишу F2, затем “Ctrl”+“Shift”+“Enter”.

Ставим курсор в A6 и вводим подпись «Поправка 2», затем в B6 вводим **=(СЧЁТ(B1:N2)+1-РАНГ.СР(B2:N2;B1:N2;0)-РАНГ.СР(B2:N2;B1:N2;1))/2**. Затем обводим мышкой ячейки B6-N6 и нажимаем клавишу F2, затем “Ctrl”+“Shift”+“Enter”.

Теперь находим ранги элементов, прибавляя к порядковому номеру поправку. Вводим в A7 подпись «Ранг 1», а в соседнюю B7 формулу **=B3+B5**, автозаполняем на ячейки A7-M7. Вводим в A8 подпись «Ранг 2», а в соседнюю B8 формулу **=B4+B6**, автозаполняем на ячейки A8-N8.

На следующем этапе вводим итоговые характеристики критерия. Записываем объемы выборок и суммы рангов для каждой группы. Вводим объемы выборок. Ставим курсор в A9, вводим «n1=», а в соседнюю B9 вводим 12, в A10, вводим «n2=», а в соседнюю B10 вводим 13. Рассчитываем суммы рангов. В C9 вводим «R1=», в D9 вводим формулу **=СУММ(B7:M7)**, в C10 вводим «R2=», в D10 вводим формулу **=СУММ(B8:N8)**.

Рассчитываем теперь статистики критерия:

$$\omega_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1,\ \omega_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2,\ W = \min(\omega_1, \omega_2).$$

Вводим в E9 подпись «w1=», а в E10 подпись «w2=», в E11 подпись «W=».

В F9 вводим формулу **=B9*B10+B9*(B9+1)/2-D9**, в F10 формулу **=B9*B10+B10*(B10+1)/2-D10**, в E11 формулу **=МИН(F9:F10)**.

Полученное значение критерия Вилкоксона находится в ячейке E11. Согласно методике критерия полученное значение нужно сравнить с критическим. Но, к сожалению, в Excel нет функции, возвращающей обратное распределение Вилкоксона. Поэтому воспользуемся приближенной формулой. Рассчитаем другую статистику $Z = \frac{n_1 n_2/2 - W}{\sqrt{n_1 n_2 (n_1 + n_2 + 1)/12}}$. Для этого вводим в E12 вводим подпись «Z=», а в соседнюю ячейку F12 вводим формулу статистики Z: **=(B9*B10/2-F11)/КОРЕНЬ(B9*B10*(B9+B10+1)/12)**. Результат 3,100391. Критическое значение находим из обратного нормального распределения. Вводим в G12 подпись «Zкр=», а в соседней H12 вызываем мастер функции и в категории «Статистические» находим функцию НОРМСТОБР, аргументом которой будет доверительная вероятность $p = 1 - \alpha = 1 - 0{,}05 = 0{,}95$. Вводим 0,95 в поле «Вероятность» вызванной функции. Видно, что Z-статистика критерия больше критического значения 1,644854, следовательно скорости реакции в группах значимо различаются, методика разработанная психологом действительно повышает скорость реакции и производительность труда.

Задание 1. Имеются данные о количествах продаж товара в двух городах (табл. 8). Проверить на уровне значимости 0,01 статистическую гипотезу о том, что среднее число продаж товара в городах различно.

Таблица 8

Исходные данные

Первый город													
23	25	23	22	23	24	28	16	18	23	29	26	31	19

Второй город											
22	29	36	24	28	24	30	24	34	24	29	27

Содержание отчета о выполнении работы:

1. Название и цель работы;
2. Результаты выполненных заданий;
3. Выводы и рекомендации (при необходимости).

ЛАБОРАТОРНАЯ РАБОТА 3 СТАТИСТИЧЕСКИЕ МЕТОДЫ ОБРАБОТКИ ДАННЫХ

Цель работы: приобретение практических навыков обработки собранных статистических данных.

Время выполнения работы: 4 часа.

Ход работы

1. Предварительная обработка данных

Основным объектом исследования в математической статистике является выборка. Выборкой объема n называются числа x_1 , x_2 , …, x_n, получаемые на практике при n – кратном повторении эксперимента в неизменных условиях. На практике выборку чаще всего представляют статистическим рядом. Для этого вся числовая ось, на которой лежат значения выборки, разбивается на k интервалов (это число выбирается произвольно от 5 до 10), которые обычно равны, вычисляются середины интервалов z_i, и считается число элементов выборки, попадающих в каждый интервал n_i. Статистическим рядом называется последовательность пар (z_i , n_i). Рассмотрим решение задачи на ЭВМ в программе EXCEL на следующем примере.

ПРИМЕР 1. Дана выборка числа проданных автомобилей торговой фирмой за 25 недель:

14, 18, 16, 21, 12, 19, 27, 19, 15, 20, 27, 29, 22, 28, 19, 17, 18, 24, 23, 22, 19, 20, 23, 21, 19.

Построим статистический ряд, полигон, гистограмму и кумулятивную кривую. Откроем книгу программы EXCEL, Введем в первый столбец (ячейки A1-A25) исходные данные. Определим область чисел, на какой лежат данные. Для этого найдем максимальный и минимальный элементы выборки. Введем в B1 подпись «Максимум», а в B2 - подпись «Минимум». В соседних ячейках C1 и C2 определим функции «MAX» и «MIN». Для этого ставим курсор в C1 и вызываем мастер функций, нажав на кнопку fx, в открывшемся окне в поле «Категория» выбираем «Статистические», и ниже ищем функцию МАКС и вызываем ее двойным щелчком мыши по названию. В качестве аргумента функции (в графе «Число 1») обведем область данных (ячейки A1-A25). Поле «Число 2»

оставляем пустым. Нажимаем «ОК». Результатом будет число 29. Ставим курсор в ячейку С2 и аналогично вводим функцию МИН. Результат – число 12. Видно, что все данные укладываются на отрезке [12;30]. Разделим его на девять (выбирается произвольно от 5 до 10) интервалов по 2 единицы каждый:
12-14, 14-16, 16-18, 18-20, 20-22, 22-24, 24-26, 26-28, 28-30. В ячейки D1-D9 вводим верхние границы интервалов группировки – числа 14, 16, 18, 20, 22, 24, 26, 28, 30. Для вычисления частот ni используют функцию ЧАСТОТА, находящуюся в категории «Статистические». Введем ее в ячейку Е1. В строке «Массив данных» введем диапазон выборки (ячейки А1-А25). В строке «Массив интервалов» введем диапазон верхних границ интервалов группировки (ячейки D1-D9). Результат функции является массивом и выводится в ячейках E1-E9. Для полного вывода (не только первого числа в E1) нужно выделить ячейки E1-E9, обведя их мышью, и нажать F2, а далее одновременно CTRL+SHIFT+ENTER. Результат – частоты интервалов 2,2,3,7,4,3,0,3,1.

Для построения гистограммы нужно выбрать ВСТАВКА/ДИАГРАММА или нажать на соответствующий значок на основной панели (при этом курсор должен стоять в свободной ячейке). Далее выбрать тип: ГИСТОГРАММА, вид по выбору, нажать «ДАЛЕЕ», в строке «ДИАПАЗОН» обвести частоты E1-E9, перейти на вкладку «РЯД», в строке « ПОДПИСИ ОСИ Х» ввести интервалы в ячейках D1-D9, нажать «ДАЛЕЕ» ввести название «ГИСТОГРАММА», подписи осей: ось Х - «ИНТЕРВАЛЫ» и ось Y - «ЧАСТОТА», нажать «ГОТОВО». Для создания полигона перейти на пустую ячейку и сделать то же самое, только вместо типа диаграммы «ГИСТОГРАММА», выбрать «ГРАФИК». Для построения кумулятивной кривой нужно посчитать накопленные частоты. Для этого в ячейку F1 вводим «=Е1», в F2 – вводим «=F1+E2» и автозаполнением перетаскиваем эту ячейку до F9. Далее строим график как и в случае полигона, но в строке «ДИАПАЗОН» вводим накопленные частоты, ссылаясь на F1-F9, а на вкладке «РЯД», в строке « ПОДПИСИ ОСИ Х» вводим интервалы в ячейках D1-D9.

Задание 1. Дана выборка выручки магазина за последние 30 дней (табл. 9). Составить статистический ряд, построить гистограмму, полигон, кумуляту.

Таблица 9

Исходные данные

Выборка														
18	19	21	18	16	19	18	16	17	18	15	22	18	17	22
14	19	16	14	14	22	14	21	18	16	12	19	18	18	15

2. Точечное оценивание

Для исследования основных свойств явления или объекта, представленного выборкой вычисляют точечные и интервальные оценки.

Точечные оценки параметров распределения это оценки, полученные по выборке и приближенно равные оцениваемым параметрам. Основными точечными оценками являются:

Объем выборки n – количество элементов в выборке.

В*ыборочное среднее* $\bar{x}$ – оценка математического ожидания, среднеарифметическое элементов выборки.

Выборочная дисперсия s^2 – среднее квадратов отклонения элементов выборки от выборочного среднего, является оценкой дисперсии, характеризует разброс выборочных значений.

Стандартное отклонение S – корень из дисперсии.

Медиана h – средний элемент вариационного ряда или полусумма двух средних элементов, если объем выборки четный.

Мода d – наиболее часто повторяющийся элемент.

Коэффициент эксцесса δ - характеризует «островерхость» гистограммы или полигона по сравнению с кривой Гаусса нормального распределения.

Коэффициент асимметрии γ - характеризует степень симметричности гистограммы или полигона.

Перцентиль на уровне p - значение t_p, меньше которого $p \cdot 100\%$ элементов выборки.

ПРИМЕР 2. Имеется выборка числа автомобилей, проданных автосалоном за 25 недель: 43, 38, 34, 51, 47, 45, 41, 52, 50, 38, 43, 44, 39, 46, 49, 42, 42, 38, 53, 55, 48, 45, 41, 49, 47. Найти основные числовые характеристики выборки.

Запускаем программу EXCEL, первый лист. Вводим исходные данные в ячейки A1-A25. Находим числовые характеристики. Для ввода функций выделяем два столбца, например B и C, в первом вводим название характеристики, во втором – функцию. В ячейки B1-B11 вводим подписи числовых характеристик, то есть вписываем в эти ячейки первый столбец таблицы приведенной ниже. В C1 вводим текст «Функция» и ниже определяем функции, соответствующие названию (из второй колонки таблицы). Все функции вызываются нажатием на кнопку fx, находятся в категории «Статистические» и в качестве массива данных (поле «ЧИСЛО 1»), указывается ссылка на A1-A25. Например, для ввода первой из них ставим курсор в C2, нажимаем fx, выбираем категорию «Статистические» и функцию «Счет», в открывшемся окне ставим курсор в поле «Число 1» и обводим курсором ячейки A1-A25, нажимаем «ОК». Также поступаем и с другими функциями (табл. 10).

Таблица 10

Виды функций

Характеристика	Функция
Объем выборки	СЧЁТ(массив данных)
Выборочное среднее	СРЗНАЧ(массив данных)
Дисперсия	ДИСП(массив данных)
Стандартное отклонение	СТАНДОТКЛОН(массив данных)
Медиана	МЕДИАНА(массив данных)
Мода	МОДА(массив данных)
Коэффициент эксцесса	ЭКСЦЕСС(массив данных)
Коэффициент асимметрии	СКОС(массив данных)
Перцентиль 40%	ПЕРСЕНТИЛЬ(массив данных; 0,4)
Перцентиль 80%	ПЕРСЕНТИЛЬ(массив данных; 0,8)

Существует другой способ вычисления числовых характеристик выборки. Для этого ставим курсор в свободную ячейку (например, D1). Затем вызываем в меню «Данные» подменю «Анализ данных». Если в меню «Данные» отсутствует этот пункт, то в обратитесь к преподавателю для его подключения. В окне «Анализ данных» нужно выбрать пункт «Описательная статистика». В появившемся окне в поле «Входной интервал» делаем ссылку на выборку A1-A25, помещая курсор в поле и обводя эти ячейки. Оставляем группирование «По столбцам». В разделе «Параметры вывода» ставим флажок на «Выходной интервал» и в соседнем поле задаем ссылку на верхнюю левую ячейку области вывода (например, D1), ставим флажок напротив «Описательная статистика», нажимаем «ОК». Результат – основные характеристики выборки (сделайте шире столбец D, переместив его границу в заголовке).

Задание 2. Для данных из задания 1 вычислить основные числовые характеристики выборки обоими способами.

3. Интервальное оценивание

Рассмотрим теперь методы интервального оценивания. Доверительным интервалом называется интервал $(a;b)$, в который с заданной вероятностью p попадает оцениваемый параметр. Вероятность p называется доверительной. Вместо нее часто задают величину $\alpha = 1 - p$, называемую уровнем значимости. Если выборка объема n представляет случайную величину, распределенную нормально, то доверительные интервалы для матожидания и дисперсии равны

$$m \in \left(\bar{x} - \frac{S \cdot t_{1-\alpha/2}(n-1)}{\sqrt{n}};\ \bar{x} + \frac{S \cdot t_{1-\alpha/2}(n-1)}{\sqrt{n}} \right),$$

$$\sigma^2 \in \left(\frac{S^2 \cdot (n-1)}{\chi^2_{1-\alpha/2}(n-1)}; \frac{S^2 \cdot (n-1)}{\chi^2_{\alpha/2}(n-1)} \right),$$

где $t_p(n)$ и $\chi^2_p(n)$ - квантили распределения Стьюдента и хи-квадрат, $\alpha = 1 - p$.

Возвращаемся на лист 1 электронной таблицы с данными примера и для них вычислим доверительные интервалы при p=0,05. Вводим данные согласно рис. 1.

	F	G	H	I
1		Уровень значимости		0,05
2		Интервал	Левая граница	Правая граница
3		Матожидание		
4		Дисперсия		

Рис. 1. Лист введенных данных

Для вычисления величины $\frac{S \cdot t_{1-\alpha/2}(n-1)}{\sqrt{n}}$ служит функция «ДОВЕРИТ» категории «Статистические» с тремя параметрами «Альфа» - уровень значимости $\alpha = 1 - p$, «Станд_откл» - среднеквадратическое отклонение S, «Размер» - объем выборки n. Таким образом, вводим в H3 функцию:

=СРЗНАЧ(A1:A25)-ДОВЕРИТ(I1;СТАНДОТКЛОН(A1:A25);25)

а в ячейку I3 функцию:

=СРЗНАЧ(A1:A25)+ДОВЕРИТ(I1;СТАНДОТКЛОН(A1:A25);25)

Для вычисления доверительного интервала для дисперсии следует отметить, что функция вычисления квантили распределения хи-квадрат (обратного распределения хи-квадрат) называется «ХИ2ОБР» (категория «Статистические») и имеет два параметра: первый «Вероятность» содержит доверительную вероятность p, второй – степень свободы n-1. Вводим в соответствии с данными условиями и формулой для доверительного интервала в ячейку H4 запись:

=ДИСП(A1:A25)*24/ХИ2ОБР(0,025;24)

а в ячейку I4 запись: **=ДИСП(A1:A25)*24/ХИ2ОБР(0,975;24)**.

Получаем значения границ доверительных интервалов.

Задание 3. Для данных из задания 1 вычислить доверительные интервалы для матожидания и дисперсии при $\alpha = 0{,}01$. Изменяя значения уровня значимости α сделать вывод о его влиянии на ширину интервала.

Содержание отчета о выполнении работы:

1. Название и цель работы;
2. Результаты выполненных заданий;
3. Выводы и рекомендации (при необходимости).

ЛАБОРАТОРНАЯ РАБОТА 4
ПРОВЕРКА ГИПОТЕЗ О РАВЕНСТВЕ ДИСПЕРСИЙ И МАТЕМАТИЧЕСКИХ ОЖИДАНИЙ

Цель работы: приобретение практических навыков проверки гипотез о равенстве дисперсий и математических ожиданий.

Время выполнения работы: 4 часа.

Ход работы

Часть 1. Критерий Фишера сравнения дисперсий

Используется в случае, если нужно проверить различается ли разброс данных (дисперсии) у двух выборок. Это может использоваться, например, при сравнении точностей обработки деталей на двух станках, равномерности продаж товара в течении некоторого периода в двух городах и т.д. Для проверки статистической гипотезы о равенстве дисперсий служит F- критерий Фишера. Основной характеристикой критерия является уровень значимости α, который имеет смысла вероятности ошибиться, предполагая, что дисперсии и, следовательно, точность, различаются. Вместо α в задачах также иногда задают доверительную вероятность $p=1-\alpha$, имеющую смысл вероятности того, что дисперсии и в самом деле равны. Обычно выбирают критическое значение уровня значимости, например 0,05 или 0,1, и если α больше критического значения, то дисперсии считаются равными, в противном случае, различны. При этом критерий может быть односторонним, когда нужно проверить, что дисперсия конкретной выделенной выборки больше, чем у другой, и двусторонним, когда просто нужно показать, что дисперсии не равны. Существует два способа проверки таких гипотез. Рассмотрим их на примерах.

Пример 1. Два автомата расфасовывают муку по мешкам, емкостью 50 кг (табл. 11). Необходимо проверить, можно ли с вероятностью не менее 0,95 считать, что точность расфасовки на обоих автоматах одинакова. Для проверки гипотезы отбираются две выборки весов муки, расфасованной на первом и втором автомате.

Таблица 11

Исходные данные

1 автом.	47,5	52,9	51,3	48,1	52,6	49,4	48,0	52,3	45,9	52,6	46,8	49,0
2 автом.	52,5	50,5	48,4	48,6	50,6	50,0	50,1	49,5	49,7	51,1	49,2	49,7

По условию задачи критерий двусторонний, так как требуется проверить различие дисперсий (точностей). Доверительная вероятность задана p=0.95, следовательно, уровень значимости $\alpha=1-p=1-0{,}95=0{,}05$. Вводим данные выборок (без подписей) в две строчки в ячейки A1-L1 и A2-L2 соответственно. Для вычисления уровня значимости двустороннего критерия служит функция ФТЕСТ(массив1;массив2). Вводим в А4 подпись «Уровень значимости», а в В4 функцию ФТЕСТ, аргументами которой должны быть ссылки на ячейки А1-L1 и А2-L2 соответственно. Результат 0,011591293 говорит о том, что вероятность ошибиться, приняв гипотезу о различии дисперсий, около 0,01, что меньше критического значения, заданного в условии задачи 0,05. Следовательно, можно говорить что опытные данные с большой вероятностью подтверждают предположение о том, что дисперсии разные и точность расфасовки автоматов различна.

Другой способ решения задачи – использовать надстройку «Анализ данных», которая находится в меню «Данные» подменю «Анализ данных». Если в меню «Данные» отсутствует этот пункт, то в меню «Файл» нужно выбрать пункт «Параметры», в нем пункт «Надстройки», нажать на кнопку «Перейти» внизу окна, и в нем поставить флажок напротив пункта «Пакет анализа» (Analysis ToolPak). После этого в меню «Данные» появится «Анализ данных» (Data Analysis). Вызвав его, откроется окно, в котором нужно выбрать «Двухвыборочный F-тест для дисперсий» (F-test Two-Sample for Variances). В открывшемся окне в полях «Интервал переменной 1» (Variable 1 Range) и «Интервал переменной 2» (Variable 1 Range) вводят ссылки на данные (А1-L1 и А2-L2, соответственно), если имеются подписи данных, то ставят флажок у надписи «Метки» (Label) (у нас их нет, поэтому флажок не ставится). Далее вводят уровень значимости в поле «Альфа» (Alpha) (по условия это 0,05, и данное значение уже указано по умолчанию). В разделе «Параметры вывода» (Output Options) ставят метку около «Выходной интервал» (Output Range) и поместив курсор в появившееся поле напротив надписи, щелкают левой кнопкой в ячейке В7. Вывод результата будет осуществляться начиная с этой ячейки. Нажав на «ОК» появляется таблица результата. Сдвиньте границу между столбцами В и С, С и D, D и E, увеличив ширину столбцов В, С и D так, чтобы умещались все надписи. В таблице указаны средние и дисперсии каждой выборки, значение F-критерия, односторонний критический уровень значимости в строке «P(F<=f) одностороннее» («P(F<=f) one-tail») и критическое значение F-критерия (F critical one tail). Если значение F-критерия ближе к единице, чем F-критическое, то с заданной вероятностью можно считать, что дисперсии равны. Об этом же говорит и то, что критический уровень значимости «P(F<=f) одностороннее» больше заданного значения α. В нашем случае F-критерий равен 5,128330184 а F-критическое 2,817927225, то есть F-критерий дальше от единицы, чем критическое значение. Это говорит о том, что дисперсии различны и автоматы имеют разную точность расфасовки.

Задание 1. Четыре станка в цеху обрабатывают детали (табл. 12). Для проверки точности обработки. взяли выборки размеров деталей у каждого станка. Необходимо сравнить с помощью F-теста попарно точности обработки всех станков (рассмотреть пары 1-2, 1-3, 1-4, 2-3, 2-4, 3-4) и сделать вывод, для каких станков точности обработки (дисперсии) равны, для каких нет. Взять уровень значимости $\alpha = 0{,}02$.

Таблица 12

Исходные данные

Выборки размеров деталей										
1 станок	29,1	26,2	30,7	33,8	33,6	35,2	23,4	29,3	33,3	26,7
2 станок	29,0	28,9	34,0	29,7	29,4	28,5	35,9	32,6	37,1	28,0
3 станок	25,7	27,5	25,4	28,9	29,9	30,1	29,0	36,6	24,8	27,8
4 станок	32,1	31,0	27,2	29,3	30,4	31,7	30,4	27,3	35,7	31,5

Часть 2. Критерий Стьюдента сравнения средних

Используется для проверки предположения о том, что средние значения двух показателей, представленных выборками, значимо различаются. Существует три разновидности критерия: один – для связанных выборок, и два для несвязанных выборок (с одинаковыми и разными дисперсиями). Если выборки не связаны, то предварительно нужно проверить гипотезу о равенстве дисперсий, чтобы определить, какой из критериев использовать. Так же как и в случае сравнения дисперсий имеются 2 способа решения задачи, которые рассмотрим на примере.

Пример 2. Имеются данные о средненедельных количествах продаж товара (тыс. шт.) до и после смены производителем оформления упаковки (табл. 13).

Таблица 13

Исходные данные

до смены	16	19	14	15	17	16	19	16	19	14	15	19	13
после смены	18	19	21	15	19	18	15	20	17	16	21	15	

Можно ли с вероятностью 0,99 считать, что смена упаковки привела к среднему увеличению количества продаж?

По условию p=0,99, α=0,01, выборки не связаны, критерий односторонний, т.к. нужно показать, что средние показателя, представленного второй выборкой, больше чем у первой. Вводим в ячейки A1-M1 и A2-L2 исходные данные. Т.к. выборки несвязаны, то предварительно сравниваем дисперсии (сделать это самостоятельно аналогично предыдущему примеру из п. 2 любым способом). В результате проверки дисперсии оказываются равными.

Первый способ решения задачи, как и в случае дисперсий, использовать стандартную функцию. Ею является ТТЕСТ(массив1;массив2;хвосты;тип), решающий задачу по t-критерию Стьюдента. В ячейке B4 вводим подпись «t-критерий», а в соседнюю C4 функцию TTEСТ (категория «Статистические») Аргументы функции:

- **массив1, массив2** – исходные данные (ссылки на A1-M1 и A2-L2);
- **хвосты** – вид критерия: если 1 – односторонний критерий, если 2 – двусторонний (в нашем случае ставится единица);
- **тип** – тип критерия: если выборки связаны, то 1, для несвязанных выборок с равными дисперсиями – ставим 2, для несвязанных выборок с неравными дисперсиями ставим 3. В нашем случае дисперсии равны, поэтому выбираем 2.

Функция возвращает критическое значение уровня значимости, имеющего смысл ошибиться, приняв гипотезу о различии средних. Если критическое значение больше заданного, то средние нужно считать равными. Результат в нашем случае 0,0476828 больше заданного $\alpha = 0{,}01$. Следовательно, смена производителем упаковки не привела к среднему увеличению продаж и изменения в количествах продаж, вероятнее всего, связано с какими-то случайными факторами.

Второй способ – использовать пакет «Анализ данных» (Data Analysis). Способ вызова и подключения его был описан в п.2. В зависимости от типа критерия выбирается один из трех: «Парный двухвыборочный t-тест для средних» (t-Teat: Paired Two Sample for Means) – для связанных выборок, и «Двухвыборочный t-тест с одинаковыми дисперсиями» (t-Teat: Two Sample Assuming Equal Variances) или «Двухвыборочный t-тест с разными дисперсиями» (t-Teat: Two Sample Assuming Unequal Variances) - для несвязанных выборок. Вызовите тест с одинаковыми дисперсиями, в открывшемся окне в полях «Интервал переменной 1» (Variable 1 Range) и «Интервал переменной 2» (Variable 2 Range) вводят ссылки на данные (A1-M1 и A2-L2, соответственно), если имеются подписи данных, то ставят флажок у надписи «Метки» (Label) (у нас их нет, поэтому флажок не ставится). Далее вводят уровень значимости в поле «Альфа» (Alpha) - 0,01. Поле «Гипотетическая средняя разность» (Hypothesized Mean Difference) оставляют пустым. В разделе «Параметры вывода» (Output Options) ставят метку около «Выходной интервал» (Output Range) и поместив курсор в появившееся поле напротив надписи, щелкают левой кнопкой в ячейке B7. Вывод результата будет осуществляться начиная с этой ячейки. Нажав на «ОК» появляется таблица результата. Сдвиньте границу между столбцами B и C, C и D, D и E, увеличив ширину столбцов B, C и D так, чтобы умещались все надписи. Процедура выводит основные характеристики выборок, t-статистику (t-stat), критические значения этих статистик и критические уровни значимости «P(T<=t) одностороннее» (P(T<=t) one-tail) и «P(T<=t) двухстороннее» (P(T<=t) two-tail). Если по модулю t-статистика меньше критического, то средние показатели с заданной вероятностью равны. В нашем случае |-1,739215668| < 2,499873517, следовательно, среднее число продаж значимо не увеличилось. Следует отметить, что если взять уровень значимости α=0,05, то результаты исследования будут совсем иными.

Задание 2. Имеются данные о количествах продаж товара в двух городах (табл. 14-15). Проверить на уровне значимости 0,01 статистическую гипотезу о том, что среднее число продаж товара в городах различно.

Таблица 14

Исходные данные

Первый город													
23	25	23	22	23	24	28	16	18	23	29	26	31	19

Таблица 15

Исходные данные

Второй город											
22	29	36	24	28	24	30	24	34	24	29	27

Задание 3. Инженер Иванов придумал новый способ изготовления микросхем. При этом, как утверждает он, количество брака должно упасть. Для проверки этого предположения были взяты несколько партий микросхем до (x, %) и после (y, %) внедрения нового способа изготовления. Проценты браков в этих выборках имеют вид (табл. 16).

Таблица 16

Исходные данные

x, %	5,1	5,3	4,7	4,9	5,7	5,2	5,1	5,0	4,9	4,6	4,9	4,8	5,2	
y, %	4,0	4,9	5,1	5,0	3,9	5,5	3,1	3,1	4,6	3,5	5,2	3,1	2,9	4,0

Можно ли с 95 % вероятностью утверждать, что новый способ производства в среднем уменьшает процент брака, если считать, что генеральные совокупности распределены по неизвестному, отличному от нормального закону?

Задание 4.

Две ценные бумаги А и В имели курсовую стоимость в течении 18 недель, как показано в табл. 17.

Таблица 17

Исходные данные

Бумага А	62	44	60	55	54	44	54	60	60	56	40	41	45	53	49	51	54	59
Бумага В	51	58	50	55	50	51	58	58	40	53	49	55	48	56	41	43	55	52

Проверить предположение о том, что риск этих бумаг не различается при $\alpha = 0{,}02$.

Указание: Риском ценных бумаг является показатель дисперсии.

Задание 5. Имеется выборка твердости вещества, измеренные на 2 приборах (табл. 18). На уровне значимости $\alpha = 0{,}05$ проверить гипотезу о том, что второй прибор завышает твердость, если генеральные совокупности, представленные выборками имеют нормальный закон распределения. Проверить также предположение о равенстве точности (дисперсий) для приборов на том же уровне значимости.

Таблица 18

Исходные данные

Образец	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.
1 прибор	21	32	26	34	25	33	31	32	28	33	28	34	27	26
2 прибор	27	26	35	32	34	33	32	19	25	31	25	30	30	28

Содержание отчета о выполнении работы:

1. Название и цель работы;
2. Результаты выполненных заданий;
3. Выводы и рекомендации (при необходимости).

ЛАБОРАТОРНАЯ РАБОТА 5
ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ О ВИДЕ РАСПРЕДЕЛЕНИЯ

Цель работы: приобретение практических навыков проверки статистических гипотез о виде распределения.

Время выполнения работы: 4 часа.

Ход работы

Методы проверки статистических гипотез занимают центральное место в исследованиях математической статистики. Одной из важнейших групп критериев проверки статгипотез являются критерии проверки гипотез о виде распределений (критерии согласия). Они по выборочным данным проверяют предположение о принадлежности генеральной совокупности к тому или иному виду распределений. Одним из наиболее мощных критериев согласия является критерий Пирсона, называемый еще критерием хи-квадрат. Его суть заключается в сравнении теоретических частот элементов выборки n_i (для дискретных распределений) с теоретическими частотами $n_i' = np_i$, где p_i- вероятность принять это значение, рассчитанное по исследуемому закону распределения. Если распределение непрерывное, то строится группированный статистический ряд из k интервалов и $p_i = F(b_i) - F(a_i)$ есть вероятность попасть в i-й интервал группировки (здесь $F(x)$ - функция распределения проверяемого закона). Статистикой критерия является величина $\chi^2 = \sum_{i=1}^{k} \frac{(n-n')^2}{n'}$. Критическое значение критерия равно обратному распределению хи-квадрат со степенями свободы (k-r-1): $\chi^2_{kr} = \chi^2_{1-\alpha}(k-r-1)$, где r – число оцениваемых параметров закона распределения. Распределение можно считать соответствующим теоретическому если выполняется условие $\chi^2 < \chi^2_{kr}$. Рассмотрим решение данной задачи на примере.

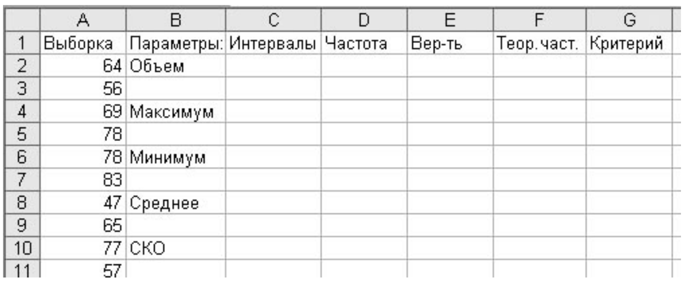
Пример 1. Имеется выборка прибыли (тыс. руб.) коммерческой фирмы за 40 дней (табл. 19). Необходимо проверить статистическую гипотезу о том, что прибыль данной фирмы распределена по нормальному закону распределения. Взять уровень значимости $\alpha = 0{,}05$.

Таблица 19

Исходные данные

Выборка прибыли коммерческой фирмы за 40 дней (тыс. руб.)																			
64	56	69	78	78	83	47	65	77	57	61	52	50	58	60	48	62	63	68	64
64	64	79	66	65	62	85	75	88	61	82	52	72	75	84	66	62	73	64	74

Для проверки гипотезы о принадлежности генеральной совокупности нормальному виду распределений необходимо строить группированный статистический ряд, т.к. нормальное распределение является непрерывным. Для этого нужно знать размах выборки, который равен разнице между максимальным и минимальным элементами выборки. Кроме того, нужно рассчитать точечные оценки математического ожидания и среднеквадратического отклонения (СКО). Открываем электронную таблицу и вводим данные выборки в нее в ячейки А2-А41, делаем подписи для расчетных параметров в соответствии с рис. 2.



	A	B	C	D	E	F	G
1	Выборка	Параметры:	Интервалы	Частота	Вер-ть	Теор.част.	Критерий
2	64	Объем					
3	56						
4	69	Максимум					
5	78						
6	78	Минимум					
7	83						
8	47	Среднее					
9	65						
10	77	СКО					
11	57						

Рис. 2. Вид оформления рабочего листа

Вычисляем параметры по выборке. Для этого вводим в ячейку В3: «**=СЧЁТ(А2:А41)**» (здесь и далее кавычки вводить не надо, функции можно вводить с помощью мастера функций из категории «Статистические», как в лабораторной работе № 2, ссылки на ячейки можно ввести щелкнув мышью по ячейке). В В5 вводим: «**=МАКС(А2:А41)**», в В7: «**=МИН(А2:А41)**», в В9: «**=СРЗНАЧ(А2:А41)**», в В11: «**=СТАНДОТКЛОН.В(А2:А41)**».

Видно, что весь диапазон значений элементов лежит на интервале от 47 до 88. Разобьем этот интервал на интервалы группировки: [0; 50], (50; 55], (55; 60], (60; 65], (65; 70], (70; 75], (75; 80], (80; 85], (85; 90]. Для этого вводим в ячейки С2-С11 границы интервалов (табл. 20).

Таблица 20

Вводимые данные

Ячейка	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
Число	0	50	55	60	65	70	75	80	85	90

Для вычисления частот n используем функцию ЧАСТОТА. Для этого в D3 вводим формулу «**=ЧАСТОТА(А2:А41;С3:С11)**». Затем обводим курсором ячейки D3-D11, выделяя их и нажимаем F2, а затем одновременно Ctrl+Shift+Enter. В результате в ячейках D3-D11 окажутся значения частот.

Для расчета теоретической вероятности $p_i = F(b_i) - F(a_i)$ вводим в ячейку Е3 разницу между функциями нормального распределения (функция НОРМ-РАСП категории «Статистические») с параметрами: «Х» – значение границы интервала, «Среднее» - ссылка на ячейку В9, «Стандартное_откл» - ссылка на В11, «Интегральная» - 1. В результате в Е3 будет формула:

=НОРМ.РАСП(C3;B9;B11;1)-НОРМ.РАСП(C2;B9;B11;1)

Автозаполняем эту формулу на Е3-Е10 перемещая нижний правый угол Е3 до ячейки Е10. В последней ячейке столбца Е11 для соблюдения условия нормировки вводим дополнение предыдущих вероятностей до единицы. Для этого вводим в Е11: «**=1-СУММ(Е3:Е10)**»

Для расчета теоретической частоты $n_i' = np_i$ вводим в F3 формулу: «**=E3*B3**», автозаполняем ее на F3-F11.

Для вычисления элементов суммы $\frac{(n-n')^2}{n'}$ критерия Пирсона вводим в G3 значение «**=(D3-F3)*(D3-F3)/F3**» и автозаполняем его на диапазон G3-G11.

Находим значение критерия χ^2 и критическое значение χ^2_{kr}. Для этого вводим в F12 подпись «Сумма», а в F13 подпись «Критич.». Вводим в соседние ячейки формулы – в G12: «**=СУММ(G3:G11)**», а в G13: «**=ХИ2.ОБР(0,05;6)**», здесь параметр $\alpha = 0{,}05$ взят из условия, а степень свободы (k-r-1)=(9-2-1)=6, так как k=9 – число интервалов группировки, а r=2, т.к. были оценены два параметра нормального распределения: математическое ожидание и СКО. Видно, что $\chi^2 < \chi^2_{kr}$, то есть можно считать, что прибыль данной фирмы распределена по нормальному закону распределения.

Проверим это, построив графики плотностей эмпирического и теоретического распределений. Ставим курсор в любую свободную ячейку и вызываем мастер диаграмм (Вставка/Диаграмма). Выбираем тип диаграммы «График» и вид «График с маркерами» самый левый во второй строке, нажимаем «Далее». Ставим курсор в поле «Диапазон» и удерживая кнопку CTRL обводим мышью область ячеек D3-D11 а затем F3-F11. Переходим на закладку «Ряд» и в поле «Подписи оси X» обводим область C3-C11. Нажимаем «Готово». Видно, что графики достаточно хорошо совпадают, что говорит о соответствии данных нормальному закону.

Задание 1. Дана выборка числа посетителей Интернет – сайта за 30 дней (табл. 21). Проверить по критерию Пирсона на уровне значимости $\alpha = 0{,}02$ статистическую гипотезу о том, что генеральная совокупность, представленная выборкой, имеет нормальный закон распределения.

Таблица 21

Исходные данные

Выборка														
45	52	49	48	42	51	54	54	50	47	56	53	59	57	50
45	50	46	55	46	54	55	64	67	51	49	47	47	55	40

Критерий Пирсона также можно использовать для проверки предположения о том, что полученные в результате наблюдений данные соответствуют нормам. Пусть имеются некоторые показатели, которые должны соответствовать стандартным нормам. Для проверки из генеральной совокупности получается выборка значений данных показателей. Рассматривается гипотеза о том, что отклонения от норм невелики, и ими можно пренебречь. Рассмотрим проверку гипотезы на примере.

Пример 2. На консервном заводе принимаемое зерно горошка считается высшего сорта, если в нем не менее 60 % зерна размером более 7 мм в диаметре, не менее 20 % зерна размером 5-7 мм, 10 % зерна 4-5 мм и 10 % зерна менее 4 мм в диаметре. На завод привезли партию зерна, из которой отобрали одну тонну для проверки. В результате оказалась, что размером более 7 мм в диаметре 550 кг, зерна размером 5-7 мм 220 кг, зерна 4-5 мм 120 кг и зерна размером менее 4 мм 110 кг. Можно ли с вероятностью 0,95 ($\alpha = 0{,}05$) говорить о том, что привезенное зерно высшего сорта?

Если бы зерно точно бы соответствовало норме, то его количество из одной тонны распределялось бы по размерам как 600 кг, 200 кг, 100 кг и 100 кг. Введем в A1 заголовок «НОРМА» и ниже в A2-A5 показатели – числа 600, 200, 100, 100. В ячейку B1 введем заголовок «НАБЛЮДЕНИЯ» и ниже в B2-B5 наблюдаемые показатели 550, 220, 120, 110. В третьем столбце вводятся формулы для критерия: в C1 заголовок «КРИТЕРИЙ», в C2 формулу «=(A2-B2)*(A2-B2)/A2». Автозаполнением размножим эту формулу на C3-C5. В ячейку C6 запишем общее значение критерия – сумму столбца C2-C5. Для этого поставим курсор в C6 и вызвав функции в категории «Математические» найдем СУММ и в аргументе «Число 1» укажем ссылку на C2-C5. Получится результат критерия Z=11,16667. для ответа на вопрос, соответствуют ли опытные показатели нормам, Z сравнивают с критическим значением Zкр. Вводим в D1 текст «критическое значение» в E1 вводим функцию **ХИ2.ОБР** (категория «Статистические») у которой два аргумента: «Вероятность» – вводится уровень значимости $\alpha = 1 - p$ (в нашем случае 1-0,95=0,05) и «Степени_свободы» – вводят число *n*-1, где *n* – число норм (в нашем случае 4-1=3). Результат 7,814725. Видно, что критическое значение меньше критерия, следовательно опытные данные не соответствует стандартам и зерно с заданной вероятностью нельзя отнести к высшему сорту.

Задание 2. При производстве микросхем процессоров используются кристаллы кварца. Стандартом предусмотрено, чтобы у 50 % образцов не было обнаружено ни одного дефекта кристаллической структуры, у 15% - один дефект, у 13 % - 2 дефекта, у 12 % - 3 дефекта, у 10 % более 3 дефектов. При анализе выборочной партии оказалось, что из 1000 экземпляров распределение по дефектам следующее (табл. 22).

Таблица 22

Исходные данные

0 дефектов	1 дефект	2 дефекта	3 дефекта	более 3
489	144	135	122	110

Можно ли с вероятностью 0,99 (при $\alpha = 0{,}01$) считать, что партия соответствует стандарту?

Задание 3. При производстве зеленого горошка консервный комбинат в г. Острогожске скупает горох у близлежащих сельскохозяйственных артелей. Технологическим процессом предусмотрено, чтобы у 30 % заказываемого гороха диаметр горошин был более 8 мм, у 40 % - от 6 до 8 мм, у 15 % - от 5 до 6 мм, у 10 % от 4 до 5 мм, у 5 % - менее 4 мм. При анализе выборочной партии оказалось, что из 1000 килограммов гороха распределение по диаметру горошин следующее (табл. 23).

Таблица 23

Исходные данные

более 8 мм	от 6 до 8 мм	от 5 до 6 мм	от 4 до 5 мм	менее 4 мм
288	402	176	91	43

Можно ли с вероятностью 0,99 считать, что партия соответствует стандарту техпроцесса?

Задание 4. Среднее число посетителей сайта по 7 дням недели имеет вид (табл. 24).

Таблица 24

Исходные данные

День недели	Пн.	Вт.	Ср.	Чт.	Пт.	Сб.	Вс.
Число посетителей	67	86	77	80	75	68	64

Можно ли с вероятностью 0,95 считать, что среднее число посетителей не зависит от дня недели?

Содержание отчета о выполнении работы:

1. Название и цель работы;
2. Результаты выполненных заданий;
3. Выводы и рекомендации (при необходимости).

ЗАКЛЮЧЕНИЕ

Методические указания разработаны с целью единого подхода к организации и проведению лабораторных работ. В предлагаемых материалах даны понятия лабораторным занятиям, рассмотрены их основные положения, ход работы, приведены примеры. Методические указания предназначенные для студентов, обучающихся по направлению подготовки 38.03.03 «Управление персоналом» профиль «Технологии управления персоналом» всех форм обучения являются удобным инструментом для организации деятельности студентов на занятиях по выполнению лабораторных работ.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. С.А. Баркалов, П.Н. Курочка, Е.Ю. Шмелева Практикум по статистике. Воронеж, ВГАСУ, 2006
2. Баркалов С.А., Курочка П.Н., Курносов В.Б. Статистика. УМК. Воронеж: «Научная книга», 2010 - 728 с.
3. Баркалов С.А., Курочка П.Н., Шмелева Е.Ю. Практикум по статистике. Воронеж, ВГАСУ, 2010.
4. Баркалов С.А., Курочка П.Н., Перевалова О.С. Статистика. Практикум. Воронеж, ВГАСУ, 2016.
5. Елисеева И.И. Общая теория статистики. М.: Финансы и статистика, 2006
6. Практикум по теории статистики - под ред. проф. Р. А. Шмайловой. М.: Финансы и статистика, 2001.
7. Гмурман В.И. Теория вероятностей и математическая статистика, М.: Высшая школа, 2003.
8. Вентцель Е.С. Теория вероятностей - 4-е изд. - М.: Наука, 1969. - 576 с.
9. Цыпин, А. П. Статистика в табличном редакторе Microsoft Excel [Электронный ресурс] : лабораторный практикум / А. П. Цыпин, Л. Р. Фаизова. — Электрон. текстовые данные. — Оренбург : Оренбургский государственный университет, ЭБС АСВ, 2016. — 289 c. — 978-5-600-01401-5. — Режим доступа: http://www.iprbookshop.ru/71328.html
10. Васильева, Э. К. Статистика [Электронный ресурс] : учебник для студентов вузов, обучающихся по специальностям экономики и управления (080100) / Э. К. Васильева, В. С. Лялин. — Электрон. текстовые данные. — М. : ЮНИТИ-ДАНА, 2017. — 398 c. — 978-5-238-01192-9. — Режим доступа: http://www.iprbookshop.ru/21058.html
11. Гусаров, В. М. Статистика [Электронный ресурс] : учебное пособие для студентов вузов, обучающихся по экономическим специальностям / В. М. Гусаров, Е. И. Кузнецова. — 2-е изд. — Электрон. текстовые данные. — М. : ЮНИТИ-ДАНА, 2017. — 429 c. — 928-5-238-01226-1. — Режим доступа: http://www.iprbookshop.ru/21166.html

ОГЛАВЛЕНИЕ

СТАТИСТИКА

МЕТОДИЧЕСКИЕ УКАЗАНИЯ

к выполнению лабораторных работ для студентов направления
38.03.02 «Менеджмент» (профиль «Менеджмент организаций»)
всех форм обучения

Составители:
Моисеев Сергей Игоревич
Перевалова Ольга Сергеевна

Компьютерный набор О. С. Перевалова

Издается в авторской редакции

Подписано к изданию 04.02.2022.
Уч.-изд. л. 1,8.

ФГБОУ ВО «Воронежский государственный технический университет»
394006 Воронеж, ул. 20-летия Октября, 84