МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

Автономная образовательная некоммерческая организация «Институт Менеджмента, Маркетинга и Финансов» (ИММиФ)

МЕТОДИЧЕСКИЕ УКАЗАНИЯ

К ВЫПОЛНЕНИЮ ЛАБОРАТОРНОГО ПРАКТИКУМА

по разделу

Теория вероятностей и математическая статистика

дисциплины «Математика»

(1 курс 2 семестр)

Кафедра: ***Математики и математических методов экономики***

Воронеж 2007

Составитель: Сергей Игоревич Моисеев

Моисеев С.И. Методические указания к выполнению лабораторного практикума по разделу «Теория вероятностей и математическая статистика» дисциплины «Математика», (1 курс, 2 семестр) // Воронеж, ИММиФ, 2007.- 36 с.

Методические указания предназначены для проведения лабораторного практикума по дисциплине «Математика», разделу «Теория вероятностей и математическая статистика», который изучается на первом курсе во втором семестре. Лабораторный практикум проводится в соответствии с учебными планами у студентов дневного отделения как полной, так и сокращенной формы обучения. Основная задача лабораторного курса состоит в обучении студентов методам решения экономических задач с использованием математического аппарата теории вероятностей, математической статистики, теории случайных процессов и теории массового обслуживания с помощью ЭВМ в среде MS Excel.

Лабораторный практикум включает в себя восемь лабораторных работ, каждая из которых содержит описательную часть, примеры решения поставленных задач на ЭВМ, задания для самостоятельного решения, содержащие 12 вариантов и список литературы.

Методические рекомендации утверждены на заседании кафедры Математики и математических методов экономики.

Протокол № ________ от «______»____________ 200__ г.

Заведующий кафедрой
д.ф.-м.н., профессор Свиридов В.В. _______________

ВВЕДЕНИЕ

В данных методических указаниях предлагается курс лабораторных работ, основная задача которых состоит в обучении студентов экономических специальностей методам решения стохастических (вероятностных) и статистических задач экономики и управления с помощью ЭВМ. Лабораторный практикум включает в себя восемь лабораторных работ по основным темам, предусмотренным учебной программой по дисциплине. Каждая лабораторная работа содержит описательную часть, примеры решения поставленных задач на ЭВМ и задания для самостоятельного решения. Для проведения лабораторного практикума необходимо программное обеспечение MS EXCEL, с возможностью подключения надстройки «Анализ данных» (Data Analysis), входящее в пакет прикладных программ MS OFFICE.

Лабораторные работы подготовлены в соответствии с темами, предусмотренными ГОС и учебными программами ИММиФ и содержат следующие разделы дисциплины:

- Статистические методы обработки информации (лабораторная работа № 1, время выполнения 2 часа);

- Статистическое оценивание параметров распределения генеральной совокупности (лабораторная работа № 2, время выполнения 2 часа);

- Проверка статистических гипотез (лабораторные работы № 3 и №4, время выполнения 4 часа);

- Регрессионный и корреляционный анализ (лабораторная работа № 5, время выполнения 2 часа);

- Теория случайных процессов (лабораторные работы № 6 и № 7, время выполнения 4 часа);

- Теория массового обслуживания (лабораторная работа № 8, время выполнения 2 часа).

В ходе выполнения каждой работы студент должен проделать на ЭВМ все примеры в соответствии с предложенным описанием, выполнить все задания, в том числе и дополнительные, выданные преподавателем, оформить отчет по лабораторной работе. Отчет должен содержать:

- название лабораторной работы;
- цель работы;
- постановку каждого задания;
- результаты решения на ЭВМ;
- анализ полученного решения, интерпретация результатов;
- выводы и заключения по заданию.

Лабораторная работа № 1

ИССЛЕДОВАНИЕ СТАТИСТИЧЕСКИХ ФУНКЦИЙ. СТАТИСТИЧЕСКИЕ МЕТОДЫ ОБРАБОТКИ ДАННЫХ.

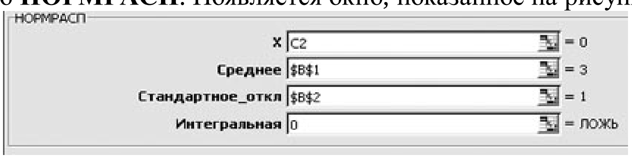
Цель: *Исследовать с помощью ЭВМ основные виды функций, применяемые в математической статистике: функции нормального распределения, распределения хи-квадрат, Стьюдента и Фишера. Научиться основным методам обработки данных, представленных выборкой. Изучить графические представления данных.*

Часть 1. Исследование статистических функций

При статических исследованиях широко используются специальные функции – законы нормального распределения, распределений хи-квадрат, Стьюдента и Фишера. Получим графики этих функций и исследуем их свойства.

Функция плотности **нормального распределения** имеет вид: $f(x) = \frac{1}{s\sqrt{2p}} \exp\left(-\frac{(x-m)^2}{2s^2}\right)$. Она зависит от двух параметров m и σ, которые имеют смысл математического ожидания и среднеквадратического отклонения. Построим график этой функции и исследуем влияние параметров m и σ на него.

Запускаем программу EXCEL и задаем значения параметров m и σ. Пусть, например, m=3 и σ=1. Для этого в ячейки A1 и A2 первого листа вводим подписи «m=» и «sig=» (кавычки здесь и далее вводить не надо), а в соседние B1 и B2 вводим значения 3 и 1. Для построения графика протабулируем в столбцах C и D функцию плотности нормального распределения на отрезке (0;6) с шагом 0,2. Для этого вводим в C1 подпись «X=», а в D1 подпись «f=». Вводим в C2 значение 0, в C3 значение 0,2, обводим, выделяя, ячейки C2 и C3 и захватив за нижний правый угол рамки вокруг ячеек C2 и C3, перетягиваем его вниз до ячейки C32, что позволит автоматически занести в столбец значения от 0 до 6 с шагом 0,2. Ставим курсор в ячейку D2 и вызываем функцию плотности нормального распределения. Для этого нажимаем кнопку мастера функций f_x выбираем категорию «**Статистические**» и функцию **НОРМРАСП**. Появляется окно, показанное на рисунке.



Вводим ссылкой на переменную X: «C2» (для ввода ссылки достаточно щелкнуть мышью по ячейке с данной адресацией), ссылкой на m и σ - «\$B\$1» и «\$B\$2». Эти ссылки абсолютные, т.к. ячейки со значениями m и σ всегда B1 и B2, поэтому пишется знак \$ (чтобы быстро относительную ссылку сделать абсолютной нужно после ввода ссылки нажать F4). В поле «Интегральное» ставим ноль, нажимаем «ОК». В ячейке D2 появился результат 0,004432, а в строке формул – запись «**=НОРМРАСП(C2;\$B\$1;\$B\$2;0)**». За нижний правый угол ячейки D2 автозаполняем результат на ячейки D2-D32.

Строим график по данным. Ставим курсор в любой свободной ячейке. Вызываем мастер диаграмм выбрав пункты меню ВСТАВКА/ДИАГРАММА. Выбираем тип диаграммы «График» и вид – левый график в верхнем ряду, нажимаем «Далее». Ставим курсор в поле «Диапазон» и обводим курсором ячейки D2-D32, переходим на закладку «Ряд», ставим курсор в поле «Подписи оси X» и обводим диапазон данных C2-C32, нажимаем «Готово». Получаем график плотности нормального распределения.

Исследуем, как влияют параметры на вид графика. Для этого изменяем в ячейке B1 значение 3 на значение 4, нажимаем Enter. Видим, что график сместился вправо, изменяем на 2, график сместился влево. Возвращаем в B1 значение 3, и изменяем в B2 значение 1 на 2. График растянулся. Изменяем в B2 на 0,5 – график сжался. Делаем вывод: Параметр m изменяет положение графика, с увеличением параметра график смещается вправо. Параметр σ влияет на ширину графика, с увеличением параметра график растягивается.

Рассмотрим теперь другие виды законов распределений.

1. Распределение хи – квадрат определяется как сумма k независимых стандартных нормальных величин. Число k называется числом степеней свободы. Когда k =1 случайная величина равна квадрату стандартной нормальной величины. Хи – квадрат распределение имеет только один параметр – число степеней свободы k, являющийся целым положительным числом. Функция, возвращающая значение плотности распределения хи-квадрат находится в категории «**Статистические**» и называется «**ХИ2РАСП**».

Задание 1. Построить график плотности распределения хи-квадрат, протабулировав эту функцию на отрезке от 0 до 10 с шагом 0,2 и взяв степень свободы k=5. Проанализировать зависимость параметра распределения k на график.

2. t - распределение Стьюдента важно в тех случаях, когда рассматриваются оценки среднего, оценки коэффициентов регрессионного уравнения, оценки параметров временных рядов. t - распределение Стьюдента с единственным параметром k**,** называемым степенью свободы сосредоточено на всей действительной оси, симметрично относительно начала координат. Функция, возвращающая значение плотности распределения Стьюдента находится в категории «**Статистические**» и называется «**СТЬЮДРАСП**». Функция имеет дополнительный чисто вычислительный параметр «Хвосты», который не связан с распределением Стьюдента, а связан с выводом полученных результатов программой EXCEL. Его всегда задаем равным 1.

Задание 2. Построить график плотности распределения Стьюдента, протабулировав эту функцию на отрезке от 0 до 7 с шагом 0,2 и взяв степень свободы k=4. Проанализировать зависимость параметра распределения k на график.

3. F- распределение Фишера возникает в регрессионном, дисперсионном, дискриминантном анализе, а также в других видах многомерного анализа данных. Случайная величина, имеющая F- распределение с парой степеней свободы m и n**,** определяется как отношение двух независимых случайных величин, имеющих распределение хи – квадрат со степенями свободы m и n с умножением на нормированный сомножитель n/m**.** F- распределение сосредоточено на положительной полуоси. Это распределение несимметрично. Функция, возвращающая значение плотности распределения Фишера находится в категории «**Статистические**» и называется «**FРАСП**». Она имеет два параметра m и n, называемых степенями свободы.

Задание 3. Построить график плотности распределения Фишера, протабулировав эту функцию на отрезке от 0 до 5 с шагом 0,2 и взяв степени свободы m=4 и n=5. Проанализировать зависимость параметров распределения m и n на график.

Часть 2. Обработка опытных данных

Основным объектом исследования в математической статистике является выборка. Выборкой объема n называются числа $x_1, x_2, \dots, x_n$, 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. 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 $\boxed{f_x}$, 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 C2 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 n_i используют функцию ЧАСТОТА, находящуюся в категории «Статистические». Введем ее в ячейку E1. В строке «Массив данных» введем диапазон выборки (ячейки A1-A25). 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 X» ввести интервалы в ячейках D1-D9, нажать «ДАЛЕЕ» ввести название «ГИСТОГРАММА», подписи осей: ось X - «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 «**=E1**», 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 4. Дана выборка выручки магазина за последние 30 дней. Составить статистический ряд, построить гистограмму, полигон, кумуляту.

Вариант	**Выборка**														
1.	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.	22	23	23	22	21	20	21	18	16	22	18	25	13	23	17
	24	21	17	19	27	26	25	21	26	19	24	20	18	23	18
3.	37	32	29	32	28	32	33	35	30	36	32	28	34	32	32
	27	32	38	38	32	29	30	39	39	31	30	31	39	29	33
4.	46	43	36	44	39	47	41	47	41	50	50	49	41	40	50
	45	46	47	44	48	46	48	46	51	41	47	51	52	40	47
5.	72	74	69	71	73	68	73	77	76	77	76	76	76	64	65
	75	70	75	71	69	72	69	78	72	67	72	81	75	72	69
6.	52	51	46	43	50	50	53	57	48	55	56	45	55	51	55
	41	54	60	52	52	59	49	51	50	47	49	57	54	54	42
7.	44	44	46	45	49	44	47	47	36	37	35	40	35	39	41
	34	38	42	44	42	35	43	45	39	33	39	45	47	41	45
8.	59	60	65	50	55	64	66	63	55	62	60	58	67	58	65
	63	59	57	65	56	66	59	59	60	61	65	59	50	64	63
9.	55	71	66	74	71	70	68	76	75	73	65	75	73	70	67
	59	63	68	65	65	81	69	64	57	58	68	70	71	71	71
10.	65	72	69	68	62	71	74	74	70	67	76	73	79	77	70
	65	70	66	75	66	74	75	84	87	71	69	67	67	75	60
11.	68	63	72	62	58	77	67	67	71	72	75	73	70	66	73
	70	69	78	73	64	71	69	73	71	71	68	65	66	69	74
12.	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

Лабораторная работа № 2

ТОЧЕЧНОЕ И ИНТЕРВАЛЬНОЕ ОЦЕНИВАНИЕ ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЙ

Цель: Овладеть навыками расчета с помощью ЭВМ основных числовых характеристик выборки. Научиться строить доверительные интервалы для математического ожидания и дисперсии.

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

Часть 1. Точечное оценивание.

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. 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 С1 вводим текст «Функция» и ниже определяем функции, соответствующие названию (из второй колонки таблицы). Все функции вызываются нажатием на кнопку $\boxed{f_x}$, находятся в категории «Статистические» и в качестве массива данных (поле «ЧИСЛО 1»), указывается ссылка на А1-А25. Например, для ввода первой из них ставим курсор в С2, нажимаем $\boxed{f_x}$, выбираем категорию «Статистические» и функцию «Счет», в открывшемся окне ставим курсор в поле «Число 1» и обводим курсором ячейки А1-А25, нажимаем «ОК». Также поступаем и с другими функциями.

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 «Сервис» подменю «Анализ данных» (Data Analysis[1]). Если в меню «Сервис» отсутствует этот пункт, то в меню «Сервис» нужно выбрать пункт «Надстройки» и в нем поставить флажок напротив пункта «Пакет анализа» (Analysis ToolPak). После этого в меню «Сервис» появится «Анализ данных» (Data Analysis). В окне «Анализ данных» нужно выбрать пункт «Описательная статистика» (Descriptive Statistics). В появившемся окне в поле «Входной интервал» (Input Range) делаем ссылку на выборку А1-А25, помещая курсор в поле и обводя эти ячейки. Оставляем группирование «По столбцам» (Columns). В разделе «Параметры вывода» (Output Options) ставим флажок на «Выходной интервал» (Output Range) и в соседнем поле задаем ссылку на верхнюю левую ячейку области вывода (напри-

[1] Здесь и далее приводятся английские названия подписей для непереведенных версий надстройки

мер D1), ставим флажок напротив «Описательная статистика» (Summary Statistics), нажимаем «ОК». Результат – основные характеристики выборки (сделайте шире столбец D, переместив его границу в заголовке).

Задание 1. Для данных из задания 4 лабораторной работы № 1 вычислить основные числовые характеристики выборки обоими способами.

Часть 2. 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 рисунку:

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

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 2. Для данных из задания 4 лабораторной работы № 1 вычислить доверительные интервалы для матожидания и дисперсии при $\alpha = 0{,}01$. Изменяя значение уровня значимости α сделать вывод о его влиянии на ширину интервала.

Лабораторная работа № 3

ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ О ВИДЕ РАСПРЕДЕЛЕНИЯ

Цель: Ознакомиться с методами проверки статистических гипотез о принадлежности генеральной совокупности, представленной выборочными данными, к тому или иному типу распределений, используя критерий согласия Пирсона (хи-квадрат) с помощью ЭВМ.

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. Необходимо проверить статистическую гипотезу о том, что прибыль данной фирмы распределена по нормальному закону распределения. Взять уровень значимости $\alpha = 0{,}05$.

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 A2-A41, делаем подписи для расчетных параметров в соответствии с рисунком:

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

Вычисляем параметры по выборке. Для этого вводим в ячейку B3: **«=СЧЁТ(A2:A41)»** (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 B5 вводим: «**=МАКС(A2:A41)**», в B7: «**=МИН(A2:A41)**», в B9: «**=СРЗНАЧ(A2:A41)**», в B11: «**=СТАНДОТКЛОН(A2:A41)**».

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 C2-C11 границы интервалов:

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(A2:A41;C3:C11)**». 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 E3 разницу между функциями нормального распределения (функция НОРМРАСП категории «Статистические») с параметрами: «X» – значение границы интервала, «Среднее» - ссылка на ячейку B9, «Стандартное_откл» - ссылка на B11, «Интегральная» - 1. В результате в E3 будет формула:

=НОРМРАСП(C3;B9;B11;1)-НОРМРАСП(C2;B9;B11;1)

Автозаполняем эту формулу на E3-E10 перемещая нижний правый угол E3 до ячейки E10. В последней ячейке столбца E11 для соблюдения условия нормировки вводим дополнение предыдущих вероятностей до единицы. Для этого вводим в E11: «**=1-СУММ(E3:E10)**»

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

Для вычисления элементов суммы $\dfrac{(n-n')^2}{n'}$ 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: «**=ХИ2ОБР(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. Проверить по критерию Пирсона на уровне значимости $\alpha = 0{,}02$ статистическую гипотезу о том, что генеральная совокупность, представленная выборкой, имеет нормальный закон распределения.

Вариант	**Выборка**														
1.	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
2.	48	43	52	42	38	57	47	47	51	52	55	53	50	46	53
	50	49	58	53	44	51	49	53	51	51	48	45	46	49	54
3.	65	81	76	84	81	80	78	86	85	83	75	85	83	80	77
	69	73	78	75	75	91	79	74	67	68	78	80	81	81	81
4.	75	82	79	78	72	81	84	84	80	77	86	83	89	87	80
	75	80	76	85	76	84	85	94	97	81	79	77	77	85	70
5.	78	73	82	72	68	87	77	77	81	82	85	83	80	76	83
	80	79	88	83	74	81	79	83	81	81	78	75	76	79	84
6.	70	59	57	62	49	63	59	60	57	66	64	57	59	58	59
	56	62	56	57	63	59	55	58	62	61	60	59	59	61	63
7.	39	41	35	41	42	38	41	41	36	45	40	39	41	41	40
	42	45	39	39	35	41	36	36	39	41	43	40	41	38	44
8.	15	31	26	34	31	30	28	36	35	33	25	35	33	30	27
	19	23	28	25	25	41	29	24	17	18	28	30	31	31	31
9.	25	32	29	28	22	31	34	34	30	27	36	33	39	37	30
	25	30	26	35	26	34	35	44	47	31	29	27	27	35	20
10.	59	60	65	50	55	64	66	63	55	62	60	58	67	58	65
	63	59	57	65	56	66	59	59	60	61	65	59	50	64	63
11.	40	41	37	37	40	42	39	43	38	41	45	44	48	43	28
	39	41	39	38	44	37	41	42	45	40	43	35	44	44	44
12.	54	59	55	57	44	42	52	55	49	53	51	50	61	59	53
	46	47	44	52	49	48	56	40	52	46	46	45	52	59	57

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 А1 заголовок «НОРМА» и ниже в А2-А5 показатели – числа 600, 200, 100, 100. В ячейку В1 введем заголовок «НАБЛЮДЕНИЯ» и ниже в В2-В5 наблюдаемые показатели 550, 220, 120, 110. В третьем столбце вводятся формулы для критерия: в С1 заголовок «КРИТЕРИЙ», в С2 формулу «=(А2-В2)*(А2-В2)/А2». Автозаполнением размножим эту формулу на С3-С5. В ячейку С6 запишем общее значение критерия – сумму столбца С2-С5. Для этого поставим курсор в С6 и вызвав функции в категории «Математические» найдем СУММ и в аргументе «Число 1» укажем ссылку на С2-С5. 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 Е1 вводим функцию ХИ2ОБР (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 (вариант соответствует номеру ЭВМ):

Вариант	0 дефектов	1 дефект	2 дефекта	3 дефекта	более 3
1.	489	144	135	122	110
2.	491	145	134	125	105
3.	489	155	133	123	100
4.	483	153	132	130	102
5.	516	148	131	110	95
6.	508	152	129	111	100
7.	494	147	136	121	102
8.	492	155	128	120	105
9.	471	160	137	122	110
10.	471	159	135	127	108
11.	489	156	131	117	107
12.	486	153	136	119	106

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

Лабораторная работа № 4

ПРОВЕРКА ГИПОТЕЗ О РАВЕНСТВЕ ДИСПЕРСИЙ И МАТЕМАТИЧЕСКИХ ОЖИДАНИЙ

Цель: Используя F-критерии Фишера и t-критерий Стьюдента научиться проверять гипотезы о равенстве дисперсий и математических ожиданий (средних) с помощью ЭВМ.

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. 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:

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 A4 подпись «Уровень значимости», а в B4 функцию ФТЕСТ, аргументами которой должны быть ссылки на ячейки A1-L1 и A2-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» (Data Analysis). Для ее подключения нужно в меню «СЕРВИС» выбрать «НАДСТРОЙКИ» и поставить флажок напротив «Пакет анализа» (Analysis ToolPak). После этого в меню «СЕРВИС» появится пункт «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 (A1-L1 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) (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 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. 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. 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 $\alpha = 0{,}02$.

Вар.	Выборки размеров деталей										
1, 6, 11	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
2, 7, 12	1 станок	36,6	34,3	33,9	30,3	30,0	31,4	29,9	26,8	24,7	32,5
	2 станок	28,4	32,5	31,5	28,2	33,9	24,7	31,7	29,7	30,1	28,0
	3 станок	33,1	30,4	33,4	29,6	27,7	33,2	28,3	31,6	31,6	29,1
	4 станок	30,6	31,6	29,3	26,3	33,8	29,1	26,1	32,3	32,4	31,3
3, 8	1 станок	34,1	35,1	30,7	30,4	35,6	29,9	28,0	32,7	30,0	33,1
	2 станок	30,8	34,4	30,3	26,6	25,8	30,6	32,9	25,5	28,2	31,6
	3 станок	30,7	30,6	30,0	26,3	30,7	30,4	32,3	27,8	31,8	30,7
	4 станок	30,6	31,3	27,0	27,4	31,4	30,4	28,4	30,3	27,2	27,3
4, 9	1 станок	28,1	27,1	33,6	32,8	24,8	33,8	29,4	26,6	24,4	27,5
	2 станок	31,8	27,1	32,6	34,3	27,8	29,1	26,0	34,1	33,1	30,6
	3 станок	27,1	34,6	26,5	28,8	26,1	34,8	30,1	31,0	32,9	35,8
	4 станок	28,1	32,6	27,5	29,7	29,3	34,6	26,0	27,2	29,5	26,8
5, 10	1 станок	29,7	30,4	35,2	28,5	27,6	27,8	31,8	33,9	25,7	32,9
	2 станок	30,0	33,0	27,0	32,3	33,7	26,5	31,2	24,7	30,2	33,0
	3 станок	28,8	30,7	35,5	22,8	30,1	29,6	33,0	33,7	34,9	24,5
	4 станок	25,0	31,3	30,6	32,0	29,5	32,5	34,0	35,7	26,1	31,9

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.

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 ТТЕСТ (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. Проверить на уровне значимости 0,01 статистическую гипотезу о том, что среднее число продаж товара в городах различно.

Первый город (одинаково для всех вариантов)													
23	25	23	22	23	24	28	16	18	23	29	26	31	19

Вариант	Второй город (по вариантам)											
1.	22	29	36	24	28	24	30	24	34	24	29	27
2.	27	27	20	22	28	21	29	36	19	21	32	20
3.	28	20	31	20	28	30	16	28	25	21	28	22
4.	37	25	30	30	21	23	32	27	25	28	18	32
5.	22	28	26	26	35	20	27	24	22	21	26	29
6.	29	34	30	23	33	19	21	28	28	26	22	19
7.	24	25	23	22	28	22	25	23	31	37	26	13
8.	21	28	18	30	27	24	28	28	22	26	27	26
9.	16	27	29	24	17	24	30	33	23	26	20	34
10.	32	32	29	25	17	27	21	22	26	25	28	28
11.	24	25	24	21	25	22	35	20	26	29	37	21

Лабораторная работа № 5

ОСНОВЫ РЕГРЕССИОННОГО И КОРРЕЛЯЦИОННОГО АНАЛИЗА

Цель: *Освоить методы построения линейного уравнения парной регрессии с помощью ЭВМ, научиться получать и анализировать основные характеристики регрессионного уравнения.*

Уравнение регрессии строится для анализа статистических зависимостей между двумя или более показателей. Если показателей два, то регрессия называется парной. Если зависимость между показателями X и Y пропорциональная, то регрессия будет линейной и описывается уравнением вида $y = ax + b$. Рассмотрим методику построения регрессионного уравнения на примере.

ПРИМЕР. Торговая организация желает выяснить, как влияет количество вложенных в рекламную акцию денег - X (тыс.руб.) на количество проданного товара – Y (тыс. шт.). Для этого проводились наблюдения в разных городах региона и были получены следующие данные.

X	12	15	17	19	20	22	25	27	28	30	33	33
Y	34	42	45	49	53	55	61	68	67	71	75	74

Введем эту таблицу в ячейки A1-M2 электронной книги Excel. Просмотрим предварительно, как лежат точки на графике и какое уравнение регрессии лучше выбрать. Для этого строим график. Вызвав мастер диаграмм и выбрав тип диаграммы «Точечная» нажимаем «Далее» и поместив курсор в поле «Диапазон» обводим курсором данные Y (ячейки B2-M2). Переходим на закладку «Ряд» и в поле

«Значения X» делаем ссылку на ячейки B1-M1, обводя их курсором. Нажимаем «Готово» Как видно из графика, точки хорошо укладываются на прямую линию, поэтому будем находить уравнение линейной регрессии вида $y = ax + b$.

Для нахождения коэффициентов a и b уравнения регрессии служат функции НАКЛОН и ОТРЕЗОК. категории «Статистические». Вводим в A5 подпись «*a*=» а в соседнюю ячейку B5 вводим функцию НАКЛОН, ставим курсор в поле «Изв_знач_у» задаем ссылку на ячейки B2-M2, обводя их мышью. Аналогично в поле «Изв_знач_х» даем ссылку на B1-M1. Результат 1,923921. Найдем теперь коэффициент b. Вводим в A6 подпись «b=», а в B6 функцию ОТРЕЗОК с теми же параметрами, что и у функции НАКЛОН. Результат 12,78151. Следовательно, уравнение линейной регрессии есть $y = 1{,}92x + 12{,}78$.

Построим график уравнения регрессии. Для этого в третью строчку таблицы введем значения функции регрессии в заданных точках X (первая строка) - $y(x_i)$. Для получения этих значений используется функция ТЕНДЕНЦИЯ категории «Статистические». Вводим в A3 подпись «Y(X)» и, поместив курсор в B3, вызываем функцию ТЕНДЕНЦИЯ. В полях «Изв_знач_у» и «Изв_знач_х» даем ссылку на B2-M2 и B1-M1. В поле «Нов_знач_х» вводим также ссылку на B1-M1. В поле «Константа» вводят 1, если уравнение регрессии имеет вид $y = ax + b$, и 0, если $y = ax$. В нашем случае вводим единицу. Функция ТЕНДЕНЦИЯ является массивом, поэтому для вывода всех ее значений выделяем область B3-M3 и нажимаем F2 и Ctrl+Shift+Enter. Результат – значения уравнения регрессии в заданных точках. Строим график. Ставим курсор в любую свободную клетку, вызываем мастер диаграмм, выбираем категорию «Точечная», вид графика – линия без точек (в нижнем правом углу), нажимаем «Далее», в поле «Диапазон» вводим ссылку на B3-M3. Переходим на закладку «Ряд» и в поле «Значения X» вводим ссылку на B1-M1, нажимаем «Готово». Результат – прямая линия регрессии. Посмотрим, как различаются графики опытных данных и уравнения регрессии. Для этого ставим курсор в любую свободную ячейку, вызываем мастер диаграмм, категория «График», вид графика – ломаная линия с точками (вторая сверху левая), нажимаем «Далее», в поле «Диапазон» вводим ссылку на вторую и третью строки B2-M3. Переходим на закладку «Ряд» и в поле «Подписи оси X» вводим ссылку на B1-M1, нажимаем «Готово». Результат – две линии (Синяя – исходные данные, красная – уравнение регрессии). Видно что линии мало различаются между собой.

Для вычисления коэффициента корреляции r_{xy} служит функция ПИРСОН. Размещаем графики так, чтобы они располагались выше 25 строки, и в A25 делаем подпись «Корреляция», в B25 вызываем функцию ПИРСОН, в полях которой «Массив 1» и «Массив 2» вводим ссылки на исходные данные B1-M1 и B2-M2. Результат 0,993821. Коэффициент детерминации R_{xy} – это квадрат коэффициента корреляции r_{xy}. В A26 делаем подпись «Детерминация», а в B26 – формулу «=B25*B25». Результат 0,987681.

Однако, в Excel существует одна функция, которая рассчитывает все основные характеристики линейной регрессии. Это функция ЛИНЕЙН. Ставим курсор в B28 и вызываем функцию ЛИНЕЙН, категории «Статистические». В полях «Изв_знач_y» и «Изв_знач_x» даем ссылку на B2-M2 и B1-M1. Поле «Константа» имеет тот же смысл, что и в функции ТЕНДЕНЦИЯ, у нас она равна 1. Поле «Стат» должно содержать 1, если нужно вывести полную статистику о регрессии. В нашем случае ставим туда единицу. Функция возвращает массив размером 2 столбца и 5 строк. После ввода выделяем мышью ячейки B28-C32 и нажимаем F2 и Ctrl+Shift+Enter. Результат – таблица значений, числа в которой имеют следующий смысл:

Коэффициент a	Коэффициент b
Стандартная ошибка m_a	Стандартная ошибка m_b
Коэффициент детерминации R_{xy}	Среднеквадратическое отклонение y
F – статистика	Степени свободы n-2
Регрессионная сумма квадратов $S_в^2$	Остаточная сумма квадратов S_a^2

Анализ результата: в первой строчке – коэффициенты уравнения регрессии, сравните их с рассчитанными функциями НАКЛОН и ОТРЕЗОК. Вторая строчка – стандартные ошибки коэффициентов. Если одна из них по модулю больше чем сам коэффициент, то коэффициент считается нулевым. Коэффициент детерминации характеризует качество связи между факторами. Полученное значение 0,987681 говорит об очень хорошей связи факторов. F – статистика проверяет гипотезу о адекватности регрессионной модели. Данное число нужно сравнить с критическим значением. для его получения вводим в E33 подпись «F-критическое», а в F33 функцию FРАСПОБР, аргументами которой вводим соответственно «0,05» (уровень значимости), «1» (число факторов X) и «10» (степени свободы). Видно, что F – статистика больше, чем F– критическое, значит регрессионная модель адекватна. В последней строке приведены регрессионная сумма квадратов $S_в^2 = \sum_{i=1}^{n}(\tilde{y}(x_i) - \bar{y})^2$ и остаточные суммы квадратов $S_в^2 = \sum_{i=1}^{n}(\tilde{y}(x_i) - y_i)^2$. Важно, чтобы регрессионная сумма (объясненная регрессией) была намного больше остаточной (не объясненная регрессией, вызванная случайными факторами). В нашем случае это условие выполняется, что говорит о хорошей регрессии.

Задание. Даны выборки факторов x_i и y_i. По этим выборкам найти уравнение линейной регрессии $\tilde{y} = ax + b$. Найти коэффициент парной корреляции. Проверить на уровне значимости $\alpha = 0{,}05$ регрессионную модель на адекватность.

Вариант	Значения фактора x_i (одинаковое для всех вариантов)									
	0	1	2	3	4	5	6	7	8	9
	Значения фактора y_i (по вариантам)									
1.	-3,7	-3,1	-4,4	-6,5	-4,6	-4,4	-8,4	-4,1	-5,5	-7,5
2.	12,1	12,1	10,7	12,1	9,6	11,2	12,8	12,5	10,0	16,6
3.	-2,3	-2,7	-2,9	-2,8	-2,1	3,0	2,2	4,7	4,5	3,2
4.	3,8	3,0	3,5	3,1	1,0	-0,6	0,1	-2,5	2,6	-1,2
5.	6,7	6,3	4,4	9,5	5,2	4,3	7,7	7,1	7,1	7,9
6.	11,3	7,4	10,7	9,0	7,4	6,2	3,9	5,8	13,4	9,1
7.	3,2	3,1	3,7	1,4	3,5	4,3	0,6	-3,5	-2,4	-2,3
8.	15,1	11,0	12,3	10,3	9,6	6,2	8,0	10,6	8,3	6,4
9.	0,0	-0,8	1,9	3,5	2,4	5,4	8,7	11,2	10,8	12,7
10.	1,9	5,4	10,0	9,1	12,5	16,6	13,9	17,0	21,0	20,2
11.	0,0	3,7	4,6	3,0	0,2	5,3	5,0	6,2	9,2	14,5
12.	10,3	11,0	10,6	12,0	11,3	13,7	12,7	14,7	16,5	14,2

Лабораторные работы № 6-7

МОДЕЛИРОВАНИЕ ЭКОНОМИЧЕСКИХ СИСТЕМ С ПОМОЩЬЮ СЛУЧАЙНЫХ ПРОЦЕССОВ

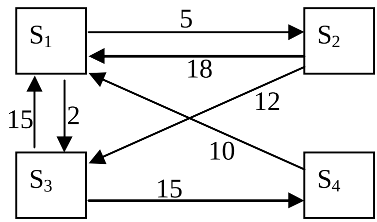
Цель: Научиться методам моделирования стохастических экономических систем с помощью дискретных Марковских процессов с непрерывным временем на ЭВМ.

Многие экономические явления и процессы зависят от множества случайных показателей, поведение которых заранее непредсказуемо. Для описания и анализа таких систем необходимо использовать математические методы и модели теории вероятностей. Такие методы называются *вероятностными или стохастическими*. Для моделирования многих стохастических систем применяется математических аппарат, основанный на дискретных Марковских случайных процессах. Случайный процесс называется Марковским, если в нем отсутствует последействие, то есть будущее поведение системы определяет только настоящее и на него никак не влияет прошлое, когда неважно, в каких состояниях и в какие моменты времени система находилась ранее. Случайные процессы называются дискретными, если возможные состояния системы счетные и четко определены. Если система может сменить состояние в любой момент времени, то такой процесс называется процессом с непрерывным временем.

Рассмотрим стохастическую систему, которая имеет n состояний, которые обозначим $S_1, S_2, ..., S_n$. Основными характеристиками такой системы являются вероятности состояний $P_i,\ i = 1,2,...,n$, равные вероятности того, что в случайные моменты времени система находилась в состоянии S_i. Вероятности состояний также можно интерпретировать как среднюю долю времени, которую проводит система в состоянии S_i при долгом функционировании. Для нахождения вероятностей состояний необходимо знать интенсивности переходных потоков λ_{ij}, равную среднему числу переходов $S_i \to S_j$ за единицу времени, которую система проводит в состоянии S_i. Для нахождения вероятностей состояний необходимо решать систему уравнений: $\sum_{i=1}^{n} P_i \lambda_{ij} - P_j \sum_{i=1}^{n} \lambda_{ji} = 0;\ \ j = 1,2,...,n$, в котором одно из уравнение заменено на условие нормировки: $P_1 + P_2 + ... + P_n = 1$.

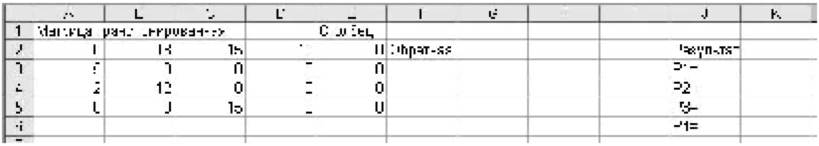
ПРИМЕР. Центральный пульт управления лаборатории обрабатывает поступающие запросы с помощью Супер-ЭВМ. Периодически, в среднем 5 раз в месяц ЭВМ проходит тестирование, которое продолжается в среднем 1 день. В результате такого тестирования в среднем в 2-х случаях из 5-и обнаруживаются проблемы, которые требуют перенастройки ЭВМ, которая длится в среднем 1 день. Кроме того, в среднем 2 раза в месяц ЭВМ производит сбой и требуется перенастройка. После перенастройки в 50 % случаев требуется ремонт, который длится в среднем 3 дня. Необходимо определить сколько в среднем дней в месяц ЭВМ работает, тестируется, перенастраивается и ремонтируется. Сколько нужно времени в среднем тратить на ремонт, чтобы ЭВМ в рабочем состоянии в среднем находилась 70 % времени?

Введем состояния: S_1 – ЭВМ работает; S_2 – ЭВМ тестируется; S_3 – ЭВМ перенастраивается; S_4 – ЭВМ в ремонте. Построим граф состояний. Для этого находим интенсивности переходных вероятностей. Возьмем за единицу времени один месяц. Тогда тестирование проводится по условию задачи 5 раз в месяц, поэтому указываем над стрелкой между 1-м и 2-м состоянием интенсивность 5. Тестирование длится 1 день, то есть 30 раз в месяц. При этом, в 2-х случаях их 5-и, то есть в 12-и случаях из 30-и обнаруживается неисправность и требуется перенастройка, а в 18-ти случаях, соответственно производится возврат в рабочее состояние. По этой причине, ставим над стрелкой 2→3 интенсивность 12, а над стрелкой 2→1 интенсивность 18. Перенастройка длится также 1 день, то есть 30 раз в месяц, в половине случаев происходит выход в рабочее состояние, в половине в ремонт. Поэтому над 3→1 и 3→4 ставим по 15. Ремонт длится 3 дня, это 10 раз в месяц, над 4→1 ставим 10.   Построим теперь матрицу переходных интенсивностей, которая полностью описывает граф состояний. Если из состояния с номером i в состояние с номером j идет стрелка с интенсивностью λ_{ij}, то в i-й строке и j-м столбце будет стоять эта интенсивность λ_{ij}. Если между состояниями перехода нет, то в соответствующей позиции матрицы стоит ноль. Для данной задачи матрица переходных интенсивностей имеет вид:



$$\begin{pmatrix} 0 & 5 & 2 & 0 \\ 18 & 0 & 12 & 0 \\ 15 & 0 & 0 & 15 \\ 10 & 0 & 0 & 0 \end{pmatrix}$$

Открываем электронную таблицу EXCEL. В ячейки A1 и E1 делаем подписи «Матрица транспонированная» и «Столбец». В диапазон A2-D5 вводим транспонированную матрицу переходных вероятностей, то есть первый столбец вводится первой строкой, второй столбец это вторая строка и т.д. В столбец E2-E5, число ячеек которого равно размеру матрицы, всегда вводятся нули. Также подготовим подписи для обратной матрицы и вывода результата (см. рисунок).



На втором этапе необходимо в транспонированную матрицу ***на место диагональных элементов ввести сумму всех остальных элементов данного столбца со знаком «минус».*** Для этого в А2 вводим «**=-СУММ(А3:А5)**» (здесь и далее кавычки не надо). В ячейку В3 вводим «**=-В2-В4-В5**», в С4 вводим «**=-С2-С3-С5**», в D5 вводим «**=-СУММ(D2:D4)**». Полученная матрица будет вырожденной и для получения единственного решения системы уравнений нужно одно любое уравнение заменить условием нормировки $P_1 + P_2 ... + P_n = 1$, которому будет соответствовать строка из единиц в расширенной матрице. Вводим во все ячейки диапазона А6-Е6 числа «1».

На третьем этапе находим обратную матрицу. Ставим курсор в ячейку F3 и вызываем мастер функций кнопкой $\boxed{f_x}$, в категории «Математические» выбираем функцию МОБР. Ставим курсор в поле «Массив» и задаем ссылку на расширенную матрицу, исключая первую строку и последний столбец, обводя мышкой ячейки от А3 до D6, нажимаем «ОК». Обводим мышкой 4 строки и 4 столбца, то есть ячейки от F3 до I6, нажимаем F2 а затем одновременно Ctrl+Shift+Enter. Находим результат решения задачи – вероятности состояний, матрица которых есть результат перемножения обратной матрицы и столбца свободных членов системы уравнений. Ставим курсор в К3, вызываем мастер функций и в категории «Математические» выбираем функцию МУМНОЖ. В поле «Массив 1» даем ссылку на диапазон ячеек от F3 до I6, обводя эти ячейки. В поле «Массив 2» даем ссылку на диапазон ячеек от Е3 до Е6. Видим, что функция выдает только одно значение: 0,66667. Для вывода всего массива обводим данную ячейку и три ниже К3-К6, выделяя их, нажимаем F2 а затем одновременно Ctrl+Shift+Enter. В результате получаем вероятности состояний:

$$P_1 = 0{,}667;\ P_2 = 0{,}111;\ P_3 = 0{,}089;\ P_4 = 0{,}133.$$

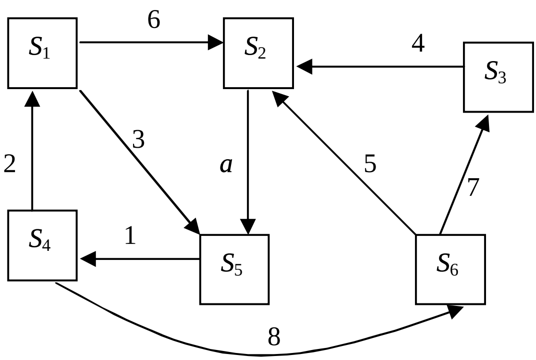
Умножив эти вероятности на 30 дней можно рассчитать, сколько дней в среднем в месяц система находится в каждом состоянии: ЭВМ работает 0,667·30 = 20 дней, ЭВМ тестируется 0,111·30 = 3,33 дня, ЭВМ перенастраивается 0,089·30 = 2,67 дней, ЭВМ в ремонте 0,133·30 = 4 дня.

Ответим теперь на второй вопрос: сколько нужно времени в среднем тратить на ремонт, чтобы ЭВМ в рабочем состоянии в среднем находилась 70 % времени? Ставим курсор в любой свободной ячейке и выбираем пункт меню «Сервис», а в нем подменю «Подбор параметра». В открывшемся окне в поле «Установить в ячейке» даем ссылку на ячейку К3, соответствующую доли времени нахождения ЭВМ в рабочем состоянии. Затем вводим в поле «Значение» число «0,7», а в поле «Изменяя значение ячейки» даем ссылку на D2 (ставим курсор в данное поле и щелкаем мышкой по D2). Нажимаем ОК и видим в D2 результат 15,46, что означает, что ремонт должен продолжаться 30/15,46=1,94 дня.

Задание 1. Найти вероятности состояний системы, которая описывается Марковским случайным процессом с дискретными состояниями, матрица переходных вероятностей которой имеет вид

$$\begin{pmatrix} 0 & 0 & 11 & 5 & 0 \\ 9 & 0 & 9 & 0 & 1 \\ 0 & 0 & 0 & 8 & 4 \\ 5 & 15 & 0 & 0 & 0 \\ 0 & 7 & 12 & 3 & 0 \end{pmatrix}.$$

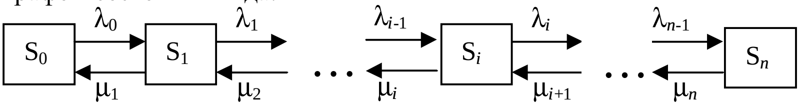
Задание 2. Найти вероятности состояний системы, граф состояний которой приведен ниже (величину *a* взять равную номеру варианта):



Задание 3. Предприниматель работает таксистом на своем автомобиле. Однако в среднем 3 раза в месяц автомобиль проходит техосмотр, который длится в среднем 1 сутки и в среднем в каждом третьем случае обнаруживается неисправность, ремонт которой в среднем длится 3 суток. Определить, сколько в среднем в месяц автомобиль работает, проходит техосмотр, ремонтируется?

Процессы гибели и размножения

Во многих экономических системах, в которых функционирует случайный процесс, возникают ситуации, когда возможен переход из любого состояния S_i в соседние состояния S_{i+1} и S_{i-1}. такие процессы называются процессами гибели и размножения и они описываются графом состояний вида:



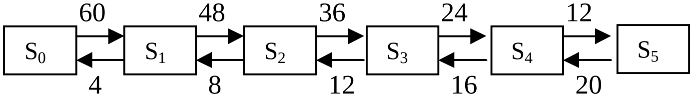
Интенсивности l_i называются интенсивностями размножения, а μ_i – интенсивности гибели. Для нахождения вероятности каждого состояния используются формулы:

$$P_0 = \frac{1}{1 + \frac{\lambda_0}{\mu_1} + \frac{\lambda_0\lambda_1}{\mu_1\mu_2} + \frac{\lambda_0\lambda_1\lambda_2}{\mu_1\mu_2\mu_3} + \ldots + \frac{\lambda_0\lambda_1\lambda_2\ldots\lambda_{n-1}}{\mu_1\mu_2\mu_3\ldots\mu_n}}; \qquad (1)$$

$$P_1 = \frac{\lambda_0}{\mu_1}\cdot P_0; \quad P_2 = \frac{\lambda_1}{\mu_2}\cdot P_1; \ \ldots \ P_i = \frac{\lambda_{i-1}}{\mu_i}\cdot P_{i-1}. \qquad (2)$$

ПРИМЕР. В автохозяйстве 5 автомобилей. Каждый из них в среднем 4 раза в год ломается и ремонт длиться с среднем 1 месяц. Определить, какую долю времени *i* автомобилей исправны ($i = 0,1,2,3,4,5$) и среднее число исправных автомобилей в произвольный момент времени.

Введем следующие состояния: S_0 – все автомобили сломаны; S_1 – один исправен; S_2 – 2 исправны; S_3 – 3 исправны; S_4 – 4 исправны; S_5 – все автомобили исправны. Граф состояний будет иметь вид:



Переходим на новый лист Excel и вводим исходные данные в соответствии с рисунком.

	A	B	C	D	E	F	G
1	Состояния	S0-S1	S1-S2	S2-S3	S3-S4	S4-S5	
2	Лямда	60	48	36	24	12	
3	Мю	4	8	12	16	20	
4	Вероятности	P0	P1	P2	P3	P4	P5

Для расчета суммы в знаменателе формулы (1) выделим строку для промежуточных вычислений. Вводим в A7 цифру 1, а в соседнюю B7 вводим формулу **=A7*B2/B3**. Автозаполняем результат на ячейки B7-F7. Для расчета вероятности P_0 по формуле (1) вводим в ячейку B5 формулу **=1/СУММ(A7:F7)**. Для расчета других вероятностей по формулам (2) вводим в C5 формулу **=B5*B2/B3** и автозаполняем результат на ячейки C5-G5. Полученные в ячейках B5-G5 числа и есть доли времени того, что *i* автомобилей исправны.

Для расчета среднего числа исправных автомобилей в произвольный момент времени ставим курсор в любую свободную ячейку и вводим формулу **=0*B5+1*C5+2*D5+3*E5+4*F5+5*G5**. Результат – 3,75. Задача решена.

Задание 4. В парикмахерской работают 2 мастера. В среднем за час приходят 5 клиентов и каждый обслуживается свободным мастером. Среднее время обслуживания клиента 30 минут. Если оба мастера заняты, то лишь 50 % остаются ожидать в очереди. Если мастера заняты и один человек ожидает в очереди, то лишь 20 % остаются ожидать. Если мастера заняты и двое ожидают в очереди, то клиенты уходят. Определить вероятности нахождения в парикмахерской *i* клиентов ($i = 0,1,2,3,4$). Сколько в среднем клиентов находятся в парикмахерской?

УКАЗАНИЕ. Если состояние S_i имеет смысл нахождения в парикмахерской *i* клиентов ($i = 0,1,2,3,4$), то граф состояний будет иметь вид:

$$\begin{tikzcd} \boxed{S_0} \arrow[r, shift left, "5"] & \boxed{S_1} \arrow[l, shift left, "2"] \arrow[r, shift left, "5"] & \boxed{S_2} \arrow[l, shift left, "4"] \arrow[r, shift left, "2{,}5"] & \boxed{S_3} \arrow[l, shift left, "4"] \arrow[r, shift left, "1"] & \boxed{S_4} \arrow[l, shift left, "4"] \end{tikzcd}$$

Задание 5. Среднее число посетителей газозаправочной станции равно 20 в час. На станции два заправщика, каждый обслуживает клиента в среднем 5 минут. Если оба заправщика заняты то лишь каждый третий клиент становится в очередь. Если оба заправщика заняты и имеется очередь, то все клиенты покидают заправку. Определить среднее число обслуженных клиентов за час.

Лабораторная работа № 8

ТЕОРИЯ МАССОВОГО ОБСЛУЖИВАНИЯ

Цель: *Исследовать с помощью ЭВМ основные виды систем массового обслуживания (СМО): одноканальные и многоканальные, с ограниченной и неограниченной очередью. Проанализировать влияние времени обслуживания на основные характеристика СМО.*

1. Одноканальная СМО с ограниченной очередью

Предположим, что система массового обслуживания имеет один канал обслуживания. Входящий поток заявок на обслуживание имеет интенсивность λ. Интенсивность потока обслуживания равна μ (т. е. в среднем непрерывно занятый канал будет выдавать μ обслуженных заявок). Заявка, поступившая в момент, когда канал занят, становится в очередь и ожидает обслуживания. Рассмотрим систему с ограниченной очередью. Предположим, что независимо от того, сколько требований поступает на вход обслуживающей системы, данная система (очередь + обслуживаемые клиенты) не может вместить более N-требований (заявок), из которых одна обслуживается, а (N-1) ожидают, Клиенты, не попавшие в ожидание, вынуждены обслуживаться в другом месте и такие заявки теряются. Наконец, источник, порождающий заявки на обслуживание, имеет неограниченную (бесконечно большую) емкость.

Обозначим P_k - вероятность того, что в системе находится k заявок. Эта величина вычисляется по формуле:

$$P_k = \frac{1-\rho}{1-\rho^{N+1}} \cdot \rho^k, \ \rho \neq 1, \ k = 0,1,2,\ldots,N \text{, где } \rho = \lambda / \mu \ . \qquad (1)$$

Рассмотрим данную модель на примере.

ПРИМЕР: На станции техобслуживания в ГИБДД имеется одна компьютерная станция диагностики, проверяющая в рамках технического осмотра автомобилей их ходовые характеристики. В среднем за час на станцию пребывает λ=20 автомобилей. Среднее время обслуживания автомобиля 2 минуты. В случае если станция диагностики занята, имеется стоянка для ожидания, рассчитанная на 19 мест (плюс одно для обслуживания). Если все места заняты, то вновь прибывающие автомобили не обслуживаются и заявки теряются.

а) Определить вероятности P_k того, что в системе будут находиться k автомобилей (k =0,1,…,20).

б) Проанализировать зависимость вероятности нахождения в системе k автомобилей P_k от времени обслуживания автомобиля

Открываем электронную книгу EXCEL. Ставим курсор в ячейку A1 и вводим подпись: «Одноканальная СМО с ограниченной очередью». Входящий поток имеет интенсивность λ=20. В ячейку A2 вводим подпись «лямбда=», а в соседнюю ячейку B2 вводим число 20. Так, как автомобили обслуживаются 2 минуты, то в среднем за час в среднем будет обслужено μ=30 автомобилей. Вводим в A3 подпись «Мю=», а в B3 число 30. Далее рассчитываем параметр $\rho = \lambda / \mu$. Вводим в A4 подпись «Ро=», а в B4 формулу **=B2/B3**. Так как система может содержать в себе максимум 20 автомобилей, то N=20. Вводим в A5 подпись «N=», а в B5 число 20.

Рассчитаем теперь по формуле (1) зависимости вероятностей P_k того, что в системе будет находиться k автомобилей от числа автомобилей k, которое может принимать значения от 0 до 20. Для этого вводим в D2 подпись «k=», а в E2 подпись «Pk=». В ячейки D3-D23 вводим целые числа от 0, 1, 2 и так до 20. В E3 вводим формулу (1) в виде:

=(1-B4)/(1-СТЕПЕНЬ(B4;B5+1))*СТЕПЕНЬ(B4;D3).

Автозаполнением переносим формулу на ячейки E3-E23. Построим по полученным данным график зависимости вероятности того, что в системе будет k автомобилей. Для этого ставим курсор в любую свободную ячейку и вызываем мастер диаграмм, выбирая в меню «ВСТАВКА» пункт «ДИАГРАММА». Выбираем тип диаграммы «График», нажимаем «Далее». Ставим курсор в поле «Диапазон» и обводим мышкой ячейки от E3 до E23. Переходим на закладку «Ряд», ставим курсор в поле «Подписи оси X» и обводим ячейки от D3 до D23, нажимаем «Далее». В поле «Ось X (категорий)» вводим текст «Число автомобилей», в поле «Ось Y (значений)» вводим текст «Вероятность», нажимаем «Готово». График построен.

Исследуем теперь зависимость вероятности нахождения в системе k автомобилей от времени обслуживания автомобиля, то есть от параметра μ. Возьмем для определенности k=5. Вводим в A6 подпись «К=», а в соседнюю ячейку B6 вводим 5. Зададим разные значения параметра μ. Вводим в F2 подпись «Мю=», а в ячейки F3-F22 значения 21, 22, …, 40. Рассчитаем теперь параметр $\rho = \lambda / \mu$. Вводим в G2 подпись «Ро=», а в ячейку G3 формулу **=B2/F3**. Автозаполняем этой формулой ячейки от G3 до G22. Находим далее вероятность P_k по формуле (1). Вводим в ячейку H2 подпись «Вероятность», а в H3 формулу **=(1-G3)/(1-СТЕПЕНЬ(G3;B5+1))*СТЕПЕНЬ(G3;B6).**

Автозаполняем формулу на ячейки H3-H22. Построим по полученным данным график. Для этого ставим курсор в любую свободную ячейку и вызываем мастер диаграмм (ВСТАВКА/ДИАГРАММА). Выбираем тип диаграммы «График», нажимаем «Далее». Ставим курсор в поле «Диапазон» и обводим мышкой ячейки от H3 до H22. Переходим на закладку «Ряд», ставим курсор в поле «Подписи оси X» и обводим ячейки от F3 до F22, нажимаем «Далее». В поле «Ось X (категорий)» вводим текст «Скорость обслуживания», в поле «Ось Y (значений)» вводим текст «Вероятность», нажимаем «Готово». График построен.

Задание 1. Проанализировать, как влияет число автомобилей в системе на полученную зависимость. Для этого, нужно менять число k в ячейке B6 на 0, 1, 5, 10, 20. Сделать вывод о влиянии параметра k на вид графика.

2. Одноканальная СМО с неограниченной очередью

Перейдем теперь к рассмотрению одноканальной СМО с ожиданием без ограничения на вместимость блока ожидания (т.е. $N \to \infty$). Остальные условия функционирования СМО остаются без изменений.

Устойчивое решение в такой системе существует только тогда, когда $\lambda<\mu$, то есть заявки должны обслуживаться с большей скоростью, чем поступают, в противном случае очередь может разрастись до бесконечности.

Вероятность того, что в системе находится k заявок, вычисляется по формуле:

$$P_k=(1-\rho)\cdot\rho^k, \quad k=0,1,2,\ldots \qquad (2)$$

Задание 2. Предположим теперь, что в условиях предыдущего примера длина очереди на обслуживание автомобилей не ограничена. Определить вероятности P_k того, что в системе будут находиться k автомобилей (k =0,1,…,20).

Проанализировать зависимость вероятности нахождения в системе k автомобилей P_k от времени обслуживания автомобиля

Указания. Для выполнения задания перейти на новый лист Excel, в ячейку A1 вводим подпись: «Одноканальная СМО с неограниченной очередью», ячейки A2, A3, A4, A6, B2, B3, B4, B6, D2-D23, E2, F2-F22, G2-G22, H2 заполнить так же как и в предыдущем примере. Для ввода формулы (2) в E3 вводим формулу **=(1-B4)*СТЕПЕНЬ(B4;D3)**, а в H3 вводим формулу **=(1-G3)*СТЕПЕНЬ(G3;B6)**, автозаполняем результаты, строим по полученным данным графики. Меняем параметр k и делаем вывод о его влиянии на вид графика.

3. Многоканальная СМО с неограниченной очередью

Рассмотрим ситуацию, когда в условиях предыдущего примера на станции техобслуживания открыли вторую компьютерную станцию диагностики. Такая ситуация будет описываться многоканальной (двухканальной) моделью СМО. Вероятности того, что в системе находятся k заявок (2 обслуживаются, остальные ожидают в очереди), для случая наличия очереди равны:

$$P_k = \frac{\rho^k}{2^{k-1} \cdot \left(1 + \rho + \frac{\rho^2}{2-\rho}\right)}, \quad k = 2,3,... \qquad (3)$$

Задание 3. Определить вероятности P_k того, что в системе в очереди будут находиться k автомобилей (k =2,3,…,20).

Проанализировать зависимость вероятности нахождения в системе k автомобилей P_k от времени обслуживания автомобиля.

ЛИТЕРАТУРА

1. Афифи А., Эйзен С. Статистический анализ. Подход с использованием ЭВМ. — М.: Мир, 1982.
2. Дюк В. Обработка данных на ПК в примерах - СПб: Питер, 1997.
3. Енюков И. С. Методы, алгоритмы, программы многомерного статистического анализа. — М.: Финансы и статистика, 1986.
4. Кулаичев А.П. Пакеты для анализа данных/ МИР ПК, №1, 1995.
5. Статистические методы для ЭВМ / Под ред. К. Эйнслейна, Э.Рэлстона, Г.С.Уолфа — М.: Наука, 1986.
6. Тюрин Ю.Н., Макаров А.А. Анализ данных на компьютере / Под ред. В. Э. Фигурнова. — М.: ИНФРА-М и Финансы и статистика, 1995.
7. Тюрин Ю.Н., Макаров А.А. Статистический анализ данных на компьютере / Под ред. В.Э. Фигурнова. — М.: ИНФРА-М, 1998.
8. Булдык Г.Д. Теория вероятностей и математическая статистика. - М.: Высшая школа, 1989.
9. Вентцель Е.С. Теория вероятностей: Учеб. для вузов.- М.: Высш. шк., 1999.
10. Гмурман В.Е. Руководство к решению задач по теории вероятности и математической статистики. М.: Высш. шк., 1998.
11. Гмурман В.Е. Теория вероятностей и математическая статистика. – Изд-во «Высшая школа», 1998.

СОДЕРЖАНИЕ

® МЕТОДИЧЕСКИЕ УКАЗАНИЯ к выполнению лабораторного практикума по разделу «**Теория вероятностей и математическая статистика**» дисциплины «**Математика**», (1 курс, 2 семестр)