

Работа 1
Знакомство с программой Microsoft Excel
в процессе заполнения электронной таблицы

Задача 1. Два токаря (Миша и Паша) поочередно работают на двух станках (станок 1 и станок 2). Они изготавливают втулки с отверстиями, диаметр которых должен превышать некоторый номинальный размер на величину допуска (минимум: +200 мкм, максимум: +225 мкм.). В Таблице 1 приведены измеренные превышения диаметра отверстий втулок, изготавливавшихся в течение 10 дней. Ежедневно с каждого станка случайным образом отбиралось и измерялось по 4 втулки.

1. Построить гистограммы: (а) общую; (б) для Миши и Паши; (в) для 1-ого и 2-ого станков.
2. Рассчитать выборочные средние и исправленные дисперсии для токарей и станков.
3. Построить доверительные интервалы для оценки генеральных средних, считая, что имеет место нормальный закон распределения.

Таблица 1

День	Токарь	Станок 1				Станок 2			
1	Миша	209.2	209.5	210.2	212.0	214.3	221.8	214.6	214.4
2	Миша	208.5	208.7	206.2	207.8	215.3	216.7	212.3	212.0
3	Миша	204.2	210.2	210.5	205.9	215.7	213.8	215.2	202.7
4	Паша	204.0	203.3	198.2	199.9	212.5	210.2	211.3	210.4
5	Паша	209.6	203.7	213.2	209.6	208.4	214.9	212.8	214.8
6	Миша	208.1	207.9	211.0	206.2	212.3	216.2	208.4	210.8
7	Миша	205.2	204.8	198.7	205.8	208	211.9	212.9	209.0
8	Паша	199.0	197.7	202.0	213.1	207.5	209.9	210.6	212.3
9	Паша	197.2	210.6	199.5	215.3	206.9	207.1	213.6	212.2
10	Паша	199.1	207.2	200.8	201.2	209.6	209.5	206.8	214.2

После запуска программы Excel 97 указатель ячейки, как правило, находится в ячейке A1. Данная ячейка является активной, т.е. пользователь может вводить в неё данные.

Введём заголовок таблицы: *Результаты работы двух токарей (Миши и Паши) на двух станках (1 и 2)*. Длина вводимого текста превышает ширину

Математическая статистика на персональном компьютере (на основе программы Microsoft Excel). Практикум для студентов факультета технологии и компьютеризации машиностроения. Составил Г.Д.Гефан.- Иркутск, 2000. - 20 с.

Практикум, рассчитанный на 12 часов занятий, включает в себя разбор задач и задания для самостоятельной работы по основным разделам математической статистики: выборка, её графическое представление, точечные и интервальные оценки числовых характеристик выборки, корреляционный и регрессионный анализ, проверка статистических гипотез. Предполагается, что студенты уже знакомы с теоретическими основами математической статистики и имеют минимальный опыт работы в операционной системе Windows 95. Изучение программы обработки электронных таблиц Excel 97 и необходимых функций рабочего листа проводится в контексте конкретных задач статистики.

Библиогр. 5 назв. Табл. 6 Рис. 5.

Зак.Б-100-2000. ИР 000863 от 30.12.00. ИРГТУ.
 #66074, Иркутск, ул.Лермонтова, 83.

ячейки A1. Поэтому текст будет выступать за правый край ячейки. После нажатия клавиши [Enter] указатель переместится в ячейку A2.

Нажмём клавишу [↓] для перемещения указателя ячейки из A2 в A3 или активизируем ячейку A3 посредством щелчка. Запишем в эту ячейку заголовок *День*, а в ячейки A4 - A13 - номера дней (с 1 по 10).

Переместимся в ячейку B3 (это можно сделать с помощью клавиш [↑] и [→], но проще - посредством щелчка. Поместим в этой ячейке заголовок *Токарь*, а в ячейках B4-B13 - имена токарей. Поскольку имена повторяются, можно воспользоваться следующими операциями. Активизируем ячейку B4 с внесённым туда текстом *Миша*. Установим курсор мыши на рамку ячейки (при этом крестик превратится в стрелку) и, удерживая левую кнопку мыши нажатой, перетащим курсор в ячейку B5 и отпустим кнопку. Текст *Миша* появился в ячейке B5, но...исчез из B4. Так получилось потому, что мы вместо операции **копирования** выполнили операцию **перемещения**. Для того, чтобы не перемещать, а копировать, нужно было во время операции удерживать нажатой клавишу [Ctrl]. Воспользовавшись этим приемом, скопируйте содержимое ячейки B5 (*Миша*) в ячейки B4, B6, B9 и B10.

В ячейку B7 поместим текст *Паша*. Для копирования этого текста в оставшиеся незаполненными ячейки данного столбца воспользуемся другим приёмом, который называется **Автозаполнение значений ячеек**. Как только Вы начнёте вводить с клавиатуры слово *Паша* в ячейку B8 (наберёте первую букву), полный текст появится сам собой! Программа «узнала» текст и выполнила работу за Вас. Опция **Автозаполнение значений ячеек** активизируется или отключается на панели **Правка** в диалоговом окне **Параметры** меню **Сервис**.

Продолжим формирование первой строки таблицы (строки заголовков). Поскольку с каждого станка отбиралось 4 втулки, заголовок *Станок 1* должен располагаться на столбцах C - F, а *Станок 2* - на столбцах G - J. Для этого вводим заголовок *Станок 1* в ячейку C3 и выделяем ячейки C3-F3, перемещая указатель мыши при нажатой левой кнопке (в результате ячейки D3-F3 должны предстать в инверсном изображении). Теперь для центрирования заголовка *Станок 1* нажимаем кнопку **Объединить и поместить в центре** на панели форматирования. Остаток щёлкнуть по левой кнопке мыши. Совершенно аналогично добиваемся правильного расположения текста *Станок 2*.

Приступим к заполнению таблицы числовыми данными. Прежде всего определим формат. Для этого выделим диапазон ячеек, в которые мы будем заносить данные. Активизируем ячейку, находящуюся в левом верхнем углу диапазона (C4) и переместим указатель мыши при нажатой левой кнопке в нужном направлении так, чтобы весь диапазон (C4 : J13) предстал в инверсном изображении. Теперь кнопку мыши необходимо отпустить и открыть меню **Формат**, а в нём выбрать **Ячейки...** В открывшемся диалоговом окне установим числовой формат с одним десятичным знаком (это соответствует характеру наших данных). Теперь можно ввести данные в таблицу. В последствии формат любой

ячейки или всего диапазона может быть изменён с помощью кнопок увеличения и уменьшения разрядности, расположенных на панели форматирования. Столбцы цифр могут быть выровнены по левому краю, по центру, по правому краю. Соответствующие кнопки расположены на панели форматирования.

Если расположение полученной таблицы на экране Вас не устраивает, Вы можете улучшить её внешний вид, используя изменение масштаба (панель инструментов) и шрифта (панель форматирования). Для того, чтобы таблица полностью помещалась на экране, установим масштаб 75%.

Окончательный вид (такой же, как у Таблицы 1) наша таблица приобретёт за счёт **обрамления**. Во-первых, уберём сетку (сочетание сетки и рамки не рекомендуется). Для этого в меню **Сервис** в диалоговом окне **Параметры** откроем панель **Вид** и отменим опцию **Сетка**. Для внешнего оформления таблицы выделим её, в меню **Формат** откроем окно **Ячейки...** и активизируем панель **Граница**. Установим двойную линию (**Тип линии**), выберем опцию **Внешние**. Внутреннее оформление таблицы (строка заголовка и три разграничительные вертикальные линии) сделайте самостоятельно. Для этого можно воспользоваться и другим способом: специальной кнопкой **Границы** на панели форматирования.

Работа 2

Нахождение частотного распределения выборки. Использование функций рабочего листа

Продолжим решение Задачи 1. Поскольку превышение диаметра отверстия втулки - непрерывный количественный признак, нужно строить не полигон частот, а гистограмму.

Прежде всего, найдём минимальное и максимальное значения количественного признака. Конечно, когда данных не много, это можно сделать и «вручную». Однако, как правило, статистические ряды довольно велики, и имеет смысл воспользоваться так называемыми **функциями рабочего листа**. Мы будем делать это неоднократно, поскольку смысл компьютерного практикума по математической статистике как раз и заключается в использовании средств (в том числе, функций рабочего листа), позволяющих практически полностью освободиться от рутинной обработки статистических данных, доверив её программе Excel.

Итак, введём в ячейку A15 текст *Минимум*, а в ячейку A16 - *Максимум*. Активизируем ячейку C15. Тем самым, мы готовим её для помещения минимального значения из списка. Щёлкнем по кнопке **Вставка функций**, находящейся на панели инструментов, откроется соответствующее диалоговое окно. Теперь необходимо выбрать нужную нам функцию. В списке категорий выбираем **Статистические**, в списке функций находим **МИН**. После щелчка на кнопке **ОК** открывается диалоговое окно с описанием функции **МИН**, где должны быть заданы аргументы. Курсор ввода находится в поле ввода первого

аргумента. Задать аргументы можно различными способами. Можно ввести адрес диапазона ячеек, содержащих значения, среди которых мы ищем минимальное (в нашем случае это С4:J13). Можно просто выделить этот диапазон в таблице (напомним, что эта операция выполняется с помощью удерживания нажатой левой кнопки мыши). В результате, выделенный диапазон будет окружён движущейся штриховой рамкой, а в поле ввода появится адрес С4:J13. Наконец, можно просто ввести сами значения в поле ввода (Конечно, в данном случае это будет совершенно не рациональным способом, т.к. придётся заново набрать весь список. Однако, когда нужно вводить всего 1-2 числа, такой способ удобен.) Так или иначе, после ввода аргументов в нижней части окна появится результат (197.2). Щёлкнув на кнопке ОК, мы перенесём его в ячейку С15.

Имейте в виду, что теперь эта ячейка всегда будет показывать минимальное значение из диапазона С4:J13, «отслеживая» все происходящие в нём изменения. Если Вы забудете о содержимом этой ячейки, Вы всегда можете щелчком активизировать её и прочесть в Строке формул (над таблицей) формулу или функцию, по которой вычислено значение, хранящееся в ячейке.

Теперь поместите в ячейку С16 максимальное значение из того же списка. Единственное отличие в Ваших действиях будет заключаться в том, что вместо функции МИН Вы должны использовать функцию МАКС. Результатом будет число 221.8.

Введём частичные интервалы так, чтобы весь диапазон наблюдаемых значений был охвачен: 196 - 198; 198 - 200; 200 - 202;...220 -222. Наша цель - найти статистическое распределение количественного признака по этим интервалам. Проще говоря, мы должны определить, сколько раз в данной выборке превышение диаметра отверстия втулки попадает в заданные интервалы, т.е. определить частоту каждого интервала. Подготовьте Таблицу 2, расположив её

Таблица 2

Интервалы			Частоты				
Нач.	Сер.	Кон.	Сумма	Миша	Паша	Ст-к 1	Ст-к 2
196	197	198					
...
220	221	222					

так, чтобы ячейка А17 располагалась в её левом верхнем углу.

Здесь под строкой, заполненной многоточиями, подразумевается 11 строк в соответствии с заданным разбиением на частичные интервалы (общее число интервалов - 13). Первые три столбца таблицы - начала, середины и кон-

цы интервалов. Правая часть таблицы будет заполнена эмпирическими частотами, для нахождения которых мы воспользуемся одной из функций рабочего листа.

Выделим столбец Сумма (ячейки D19:D31). Щёлкнем на кнопке **Вставка функций**. В открывшемся диалоговом окне выберем: **Статистические, ЧАСТОТА**. Мы выбрали функцию, которая, как Вы видите, «возвращает распределение частот в виде вертикального массива». Эта функция, в отличие от МИН и МАКС, обязательно должна иметь два аргумента: массив данных и двоичный массив. Последний называют также множеством карманов (т.е. интервалов в математическом смысле). В первое окно ввода мы вводим все наблюдавшиеся значения (диапазон С4:J13), поскольку сейчас мы определяем суммарное распределение частот. Переместив с помощью мыши курсор во второе окно ввода, мы выделяем ячейки С19:С31, в которых находятся концы интервалов. (Обратите внимание, что все эти действия отражаются в строке формул). В правой части окна появился результат: {2,6,3,4,5,9,14,13,11,10,2,0,1}. Это и есть искомые частоты. Для того, чтобы поместить их в таблицу, щёлкнем на кнопке ОК. В ячейке D19 появилось значение 2, но остальные ячейки столбца остались незаполненными. Дело в том, что результатом обращения к функции ЧАСТОТА является не просто одно число, как это было с функциями МИН или МАКС, а массив. Это требует от нас следующих дополнительных действий. Нажмём клавишу [F2]. В строке состояния (она находится в нижней части экрана) появится надпись **Правка**, а в активной ячейке - информация о её содержимом. Теперь нажмём на комбинацию клавиш [Ctrl]+[Shift] и, не отпуская, на клавишу [Enter]. Все ячейки столбца должны заполниться числами.

Вычисление частотных распределений превышения диаметра для токарей несколько более сложно, т.к. в исходной таблице относящиеся к каждому из них данные находятся в несмежных диапазонах. (Если бы каждый токарь проработал пять дней подряд, этой сложности не было бы.) Поэтому выполним преобразование Таблицы 1 в следующей последовательности:

- Выделим диапазоны С4:J6 и С9:J10 одновременно (это делается при нажатой клавише [Ctrl]).
- Скопируем их содержимое в буфер при помощи кнопки **Копировать** на панели инструментов.
- Активизируем ячейку К4.
- Щёлкнем на кнопке **Вставить** на панели инструментов. Ячейки К4:R8 заполнятся данными, относящимися к первому токарю.
- Аналогичные операции сделаем с диапазонами С7:J8 и С11:J13, копируя их содержимое в ячейку К9. Данные, относящиеся ко второму токарю, попадут в ячейки К9:R13.

Теперь Вы сможете найти частотные распределения для токарей (а затем и для станков) и тем завершить формирование таблицы (Таблица 3).

Таблица 3

Интервалы			Частоты				
Нач.	Сер.	Кон.	Сумма	Миша	Паша	Ст-к 1	Ст-к 2
196	197	198	2	0	2	2	0
198	199	200	6	1	5	6	0
200	201	202	3	0	3	3	0
202	203	204	4	1	3	3	1
204	205	206	5	5	0	5	0
206	207	208	9	4	5	5	4
208	209	210	14	8	6	7	7
210	211	212	13	8	5	6	7
212	213	214	11	4	7	2	9
214	215	216	10	6	4	1	9
216	217	218	2	2	0	0	2
218	219	220	0	0	0	0	0
220	221	222	1	1	0	0	1

Работа 3

Построение гистограмм

В программе Excel 97 существует широкий круг средств графического представления числовых данных. Гистограмма является частным случаем (типом) диаграммы, часто используемым в статистике. Изучим основные стадии процесса создания гистограммы по данным, представленным в Таблице 3.

- Прежде всего необходимо выделить в рабочем листе диапазон ячеек, данные из которого должны быть представлены в гистограмме. В соответствии с Задачей 1 начнём с построения общей гистограммы. В Таблице 3 удалим название *Сер.* из ячейки B18 (чтобы программа не воспринимала этот столбец таблицы как ряд значений функции). Выделим несмежные диапазоны ячеек: B18:B31 и D18:D31.
- Активируем кнопку **Мастер диаграмм** на панели инструментов. На экране появится первое диалоговое окно Мастера диаграмм, в котором можно

выбрать тип диаграммы. Для нашего примера мы выберем **Гистограмму** (обычную) и нажмём кнопку **Далее**.

- Во втором окне необходимо подтвердить, что ряды данных находятся в столбцах.
- В третьем диалоговом окне задаётся описание диаграммы. Можно добавить легенду (сделаем это), дать название диаграмме (введём в это окно текст *Общая гистограмма*), и дать названия по осям (назовём категории по оси X - *Превышение диаметра отверстий*, по оси Y - *Частоты*).
- В последнем окне определим размещение диаграммы. Вставим её в имеющийся лист.

Рис.1 Общая гистограмма



В рабочем листе гистограмма может находиться в закрытом состоянии (для этого нужно щёлкнуть мышью вне области диаграммы) или в активном, выделенном состоянии (щёлкнуть мышью в области диаграммы). Удерживая левую кнопку мыши нажатой, можно позиционировать диаграмму в любом месте листа. Обратите также внимание на маркеры, расположенные на границе выделенной диаграммы. Перемещая эти маркеры с помощью мыши, можно изменить размеры окна диаграммы. Разместим гистограмму в диапазоне ячеек A34 - H49.

Для обработки гистограммы с помощью специальных средств следует активизировать её с помощью двойного щелчка мышью. Форматирование диаграммы может включать в себя вставку или изменение легенды, задание цвета и узора, помещение текста на осях, форматирование осей и т.п. Интересующиеся могут ознакомиться с этими средствами самостоятельно.

Продолжим выполнение Задачи 1. Оставшиеся 4 гистограммы (для двух токарей и двух станков) имеет смысл сгруппировать и представить на каждом графике по две гистограммы: отдельно для токарей и отдельно для станков.

Выделим в Таблице 3 столбцы: *Сер.*, *Миша*, *Паша*. Активируем кнопку **Мастер диаграмм** и выделим под диаграмму диапазон A51:H66. В третьем диалоговом окне Мастера выберем образец №1 (ведь на данном графике должно быть две гистограммы, а не одна). Дадим данному графику название:

Гистограммы для токарей. Все остальные операции выполняются точно так же, как при построении предыдущей диаграммы.

Допустим, что нас не устраивают какие-то элементы графика, например, зазор между отдельными столбиками гистограммы. Установив курсор на любом из столбиков, с помощью щелчка на правой кнопке можно открыть контекстное меню и выбрать в нём **Формат данных**. В диалоговом окне активизируем кнопку **Параметры**, установим ширину зазора равную нулю и щелкнем на кнопке **ОК**. Желаемое изменение в график внесено.

Рис.2 Гистограммы для токарей



Создайте самостоятельно *Гистограммы для станков* в области ячеек А68:Н83.

Рис.3 Гистограммы для станков



Итак, первый пункт Задачи 1 выполнен. Анализ частотных распределений превышения диаметра отверстий втулок с помощью гистограмм показывает, что имеют место случаи, когда это превышение недостаточно велико (менее 200 мкм). При этом Миша работает явно лучше, чем Паша, а 2-ой станок - лучше, чем 1-ый.

Работа 4

Точечные оценки числовых характеристик генеральной совокупности. Интервальная оценка генеральной средней

До сих пор мы ограничивались качественным анализом частотных распределений. Перейдём к расчётам числовых характеристик выборки. Начнём с выборочной средней

$$\bar{x}_g = \frac{\sum_{i=1}^n x_i}{n}, \quad (1)$$

которая, как известно, является несмещённой оценкой генеральной средней. Формула (1) предполагает, что отдельные эмпирические значения x_i не объединяются в группы и их частоты не определяются. Конечно, такой подход является более точным (хотя и менее наглядным), чем введение интервалов и соответствующих им частот, как это мы делали на предыдущих занятиях. Для расчёта выборочной средней по формуле (1) в Excel 97 служит функция СРЗНАЧ, относящаяся, как и все используемые нами функции, к списку статистических.

Другой важной числовой характеристикой выборки служит выборочная дисперсия:

$$D_g = \frac{\sum_{i=1}^n (x_i - \bar{x}_g)^2}{n}, \quad (2)$$

являющаяся смещённой оценкой дисперсии генеральной совокупности. Теоретически более точной (несмещённой) оценкой генеральной дисперсии является исправленная дисперсия:

$$S = \frac{n}{n-1} D_g, \quad (3)$$

для вычисления которой в Excel 97 предусмотрена функция ДИСП.

Несмещённой оценкой генеральной средней квадратического отклонения (с.к.о.) служит «исправленное» с.к.о.:

$$s = \sqrt{S}. \quad (4)$$

Несмотря на то, что его можно вычислять по формуле (4), в Excel 97 предусмотрена специальная функция для расчёта этой величины: СТАНДОТКЛОН. Её мы и будем использовать, заодно контролируя правильность наших действий формулой (4).

Если есть основания считать, что в генеральной совокупности распределение случайной величины подчинено нормальному закону, то можно дать не только точечную оценку генеральной средней (1), но и построить доверительный интервал для математического ожидания случайной величины (при заданной надёжности γ):

$$\bar{x}_g - \frac{t\sigma}{\sqrt{n}} < a < \bar{x}_g + \frac{t\sigma}{\sqrt{n}}, \quad (5)$$

где σ - известное генеральное с.к.о., а t - значение аргумента функции Лапласа

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_0^t e^{-x^2/2} dx$$

при её значении, равном $\gamma/2$. Если σ не известно, то оно может быть заменено в формуле (5) точечной оценкой s (4). Параметр t в этом случае должен быть заменён величиной $t(\gamma;n)$, которая ищется в таблице. Однако при достаточно большом объёме выборки ($n > 30$) можно не делать разницы между σ и s , а также между $t(\gamma;n)$ и t (например, при $\gamma = 0.95$ и $n=40$ имеем: $t=1.96$, $t(\gamma;n)=2.02$). Мы будем придерживаться именно этой методики, что позволит нам использовать функцию Excel 7.0 под названием ДОВЕРИТ, которая даёт значение полуширины доверительного интервала для генеральной средней по считающемуся известным генеральному с.к.о. σ (5).

Для вычисления числовых характеристик выборки нам необходимо будет снова пользоваться Таблицей 1, которая сейчас находится в самом верху рабочего листа. Для большего удобства скопируем таблицу в ту область рабочего листа, где мы сейчас находимся. Копирование целой таблицы выполняется подобно копированию отдельной ячейки (см. Занятие 1). Таблица выделяется и «перетаскивается» при нажатой клавише [Ctrl]. Поместим копию Таблицы 1 в область A85:J95. Ниже будет располагаться Таблица 4 с числовыми характеристиками выборки (отдельно для суммы, для каждого токаря и каждого станка).

Для нахождения выборочной средней пользуемся функцией СРЗНАЧ, задавая в качестве аргумента соответствующие диапазоны ячеек в Таблице 1. Если необходимо задавать несмежные диапазоны (для токарей), делаем это при нажатой клавише [Ctrl].

Для нахождения исправленных дисперсий и с.к.о. пользуемся функциями ДИСП и СТАНДОТКЛОН, соответственно. Контролируем результаты формулой (4).

Для построения доверительных интервалов необходимо задаться надёжностью. Примем $\gamma = 0.95$. Функция ДОВЕРИТ имеет 3 аргумента: уровень значимости «альфа» ($\alpha = 1 - \gamma$), «стандартное отклонение» (исправленное с.к.о.) и

«размер» (объём выборки). В нашем случае $\alpha = 0.05$, объём выборки составляет 80 для Суммы и 40 для токарей и станков. В качестве второго аргумента («стандартное отклонение») удобнее задавать не конкретные числа, полученные с помощью функции СТАНДОТКЛОН, а адреса ячеек, куда эти значения помещены.

Таблица 4

Числовые характеристики	Сумма	Миша	Паула	Ст-к 1	Ст-к 2
Выборочная средняя	208.86	210.23	207.49	205.87	211.85
Исправленная дисперсия	26.79	19.67	30.76	23.72	12.22
Исправленное с.к.о.	5.18	4.43	5.55	4.87	3.50
Полуширина дов. интервала	1.13	1.37	1.72	1.51	1.08
Нижняя граница дов. интервала для генеральной средней	207.72	208.85	205.77	204.36	210.76
Верхняя граница дов. интервала для генеральной средней	209.99	211.60	209.21	207.38	212.93

Для определения границ доверительных интервалов воспользуемся заданием **Формул** в соответствующих ячейках рабочего листа. Например, мы хотим поместить в ячейку F103 формулу для вычисления нижней границы доверительного интервала. Активизируем эту ячейку, введём с клавиатуры знак =. Допустим теперь, что соответствующая выборочная средняя находится в ячейке F98, а полуширина доверительного интервала - в ячейке F102. Мы можем либо ввести с клавиатуры F98-F102, либо воспользоваться выделением нужных ячеек с помощью мыши (арифметические знаки всё равно придётся вводить с клавиатуры). Так или иначе, после этого в **Строке формул** и в самой ячейке F103 должна появиться надпись: =F98-F102. Это и есть **Формула**, которая будет автоматически «отслеживать» все изменения, происходящие в ячейках F98 и F102. После нажатия на клавишу [Enter] вместо формулы в ячейке появится результат вычисления. Естественно, в дальнейшем вы всегда можете, активизировав ячейку, прочесть (и при желании, отредактировать) хранящуюся в данной ячейке формулу.

Отметим, что формулу можно задать и с помощью функций. Например, мы могли бы использовать для нижней границы доверительного интервала формулу =СРЗНАЧ(...)-ДОВЕРИТ(...), где в скобках должны присутствовать аргументы функций.

Анализ Табл.4 подтверждает качественные выводы, сделанные нами на Занятии 3. Действительно, токарь Миша и 2-ой станок имеют лучшие показатели. Их средние находятся ближе к середине интервала допущенных превы-

шений диаметра (212.5), а значения дисперсий - значительно меньше, чем у Паши и у 1-ого станка (это показатель стабильной работы).

Работа 5

Корреляционный и регрессионный анализ статистических данных

Задача 2. Исследовано 30 образцов некоторого сорта стали с целью определения зависимости прочности на разрыв y , 10^7 Н/м², от содержания углерода x , %. Получена Таблица 5.

Таблица 5

№	x	y	№	x	y
1	2.0	43	16	2.7	47
2	2.4	46	17	2.1	42
3	2.2	45	18	2.6	48
4	2.3	44	19	2.4	45
5	2.5	45	20	2.1	43
6	2.8	48	21	2.3	45
7	2.2	43	22	2.2	43
8	2.7	47	23	2.3	46
9	2.4	44	24	2.4	47
10	2.3	45	25	2.3	44
11	2.0	42	26	2.4	45
12	2.2	44	27	2.6	46
13	2.6	47	28	2.5	42
14	2.1	44	29	2.6	46
15	2.5	46	30	2.4	46

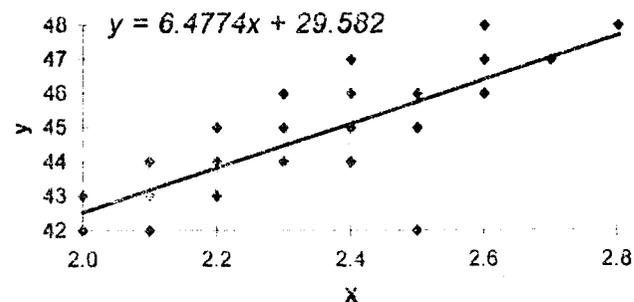
1. Построить диаграмму рассеивания и выполнить её визуальный анализ.
2. Найти уравнение прямой линии регрессии Y на X .
3. Найти коэффициент линейной корреляции.

Для построения диаграммы рассеивания расположите ряды данных x и y в два столбца. В диалоговом окне **Мастера диаграмм** выберите тип диаграммы **Точечная**, вид первый. Поскольку значения x лежат в интервале от 2 до 3, отформатируйте ось абсцисс, щёлкнув на ней дважды мышью и активизировав

кнопку **Шкала** в диалоговом окне **Формат оси** (для этого необходимо отключить режим **Авто** при определении некоторых параметров шкалы).

Визуальный анализ построенной диаграммы рассеивания позволяет сделать вывод о наличии положительной корреляции. Форма корреляции похожа на линейную (прямая пропорциональность между функцией и аргументом).

Рис.4. Диаграмма рассеивания и линия тренда



В Excel 7.0 имеется возможность интерпретировать ряды данных в виде **Линий тренда**, представляющих собой аппроксимации, полученные на основе регрессионного анализа. Активизируем диаграмму двойным щелчком, а затем установим курсор мыши на любую точку графика и сделаем щелчок на правой кнопке. В открывшемся контекстном меню выберем **Добавить линию тренда**, а в соответствующем диалоговом окне - тип аппроксимации **Линейная**. Теперь в этом же окне активизируем кнопку **Параметры** и установим опцию **Показывать уравнение на диаграмме**. После нажатия на кнопку [OK] на диаграмме появится линия тренда и соответствующее ей уравнение линейной регрессии.

Конечно, для того, чтобы получить уравнение регрессии в Excel 97, вовсе не обязательно пользоваться графическими средствами. Например, функция **ПРЕДСКАЗ** даёт значение линейного тренда (иначе говоря, значение Y в уравнении регрессии) для любого заданного значения X . В качестве первого аргумента задаётся значение $X = x$, а в качестве второго и третьего аргументов - массивы известных значений Y и X . Обратившись к функции **ПРЕДСКАЗ** дважды - для $x = 2$ и для $x = 3$ - получим: $Y(x = 2) = 42.537$; $Y(x = 3) = 49.014$. Получаем уравнение прямой по двум точкам:

$$\frac{Y - 42.537}{49.014 - 42.537} = \frac{x - 2}{1} \quad \text{или} \quad Y = 6.477x + 29.582,$$

что совпадает с полученным ранее результатом (Рис.4).

Наконец, имеется функция НАКЛОН, которая позволяет найти коэффициент регрессии по массивам известных значений X и Y . Обратитесь к ней, и Вы получите значение $a_{yx} = 6.477$.

Теперь перейдём к количественному исследованию тесноты корреляционной связи. Выборочный коэффициент корреляции может быть найден по формулам

$$r_g = (\sigma_x / \sigma_y) a_{yx} \quad (6)$$

или

$$r_g = \frac{\sum xy - n\bar{x}\bar{y}}{n\sigma_x\sigma_y} \quad (7)$$

Последняя формула используется тогда, когда данные не сгруппированы, т.е. не объединены в корреляционную таблицу.

Пользователю Excel 97 нет нужды в применении этих формул. Достаточно обратиться к функции КОРРЕЛ, задав в качестве аргументов два соответствующих друг другу массива данных, и Вы получите значение выборочного коэффициента корреляции. В нашей задаче он равен 0.789.

Замечание. Мы рассматриваем здесь только корреляционный и регрессионный анализ данных, представленных в виде перечня. Группирование данных и их объединение в корреляционную таблицу с нахождением соответствующих частот обычно применяется для того, чтобы процедура расчётов не была слишком сложной. Однако, если мы располагаем таким мощным средством статистической обработки данных, как Excel 97, тяжесть расчётов берёт на себя компьютер. Поэтому составление и использование корреляционных таблиц теряет смысл.

Работа 6

Проверка статистической гипотезы о нормальном распределении генеральной совокупности

Задача 3. В результате серии наблюдений получена выборка значений случайной величины X (Табл.6). Проверить гипотезу о том, что случайная величина X распределена нормально.

Таблица 6

x_i	5	7	9	11	13	15	17	19	21
n_i	15	26	25	30	26	21	24	20	13

Метод решения этой задачи состоит в сравнении эмпирических (т.е. полученных в опыте) частот с теоретическими частотами, которые рассчитываются по формуле

$$n_i' = nP(x_i - h/2 < x < x_i + h/2) = nhf(x_i), \quad (8)$$

где n - объём выборки, h - длина интервалов (расстояние между равноотстоящими вариантами, которые интерпретируются как середины интервалов), $f(x_i)$ - плотность распределения вероятностей случайной величины X в точке x_i в предположении, что она подчиняется гипотетическому закону.

В случае, когда выдвигается гипотеза о нормальном распределении, плотность определяется законом Гаусса:

$$f(x) = \frac{1}{\sqrt{2\pi}s} e^{-\frac{(x-\bar{x}_g)^2}{2s^2}} \quad (9)$$

где \bar{x}_g - выборочная средняя, s - «исправленное» с.к.о. случайной величины.

Создайте в новом рабочем листе Excel 7.0 таблицу, состоящую из одного столбца и занимающую ячейки A1:A200 (объём выборки равен 200). Заполните их по следующему принципу (сравните с Табл 6): 15 «пятёрки», 26 «семерки», 25 «девятки» и т.д. Легче всего это сделать с помощью автоматического заполнения. Введя в ячейку A1 число 5, поместите курсор мыши на маркер заполнения в правом нижнем углу ячейки. Курсор приобретает форму чёрного крестика. Нажав на левую кнопку мыши, переместите курсор на ячейку A15. Как только Вы отпустите кнопку, выделенные ячейки автоматически заполнятся одним и тем же числом: 5. Прибегать к этой процедуре можно для заполнения ряда ячеек не только одинаковыми числами, но и одинаковыми формулами, чем мы скоро и воспользуемся.

Имея таблицу, состоящую из единственного столбца, убедитесь с помощью уже известных Вам функций, что $\bar{x}_g = 12.63$, $s = 4.707$, поместив соответствующие формулы в ячейки F1 и F2, соответственно. Теперь воспроизведите в диапазоне C4:L5 Таблицу 6. Добавьте к этой таблице третью строку, каждая ячейка которой (D6:L6) есть копия ячейки F1. Если Вы попытаетесь сделать это с помощью обычного перетаскивания, то потерпите неудачу. Причина состоит в том, что копируемая ячейка содержит ссылку на диапазон. Не вдаваясь в подробности (желающие могут прочесть в книгах по Excel 7.0 об относительных и абсолютных ссылках), укажем выход из этой ситуации: необходимо в строке формул во всех адресах ссылки перед буквами, указывающими строки, и цифрами, указывающими столбцы, поставить знак \$. Иначе говоря, строка формул для копируемой ячейки F1 должна иметь вид не

=СРЗНАЧ(A1:A200),

а

=СРЗНАЧ(\$A\$1:\$A\$200).

Теперь можно скопировать F1 в D6, а затем с помощью курсора заполнения - во все остальные ячейки третьей строки таблицы. Аналогично, скопируйте ячейку F2 в четвертую строку (D7:L7).

В ячейках пятой строки таблицы должны быть помещены формулы для расчёта теоретических частот. Для этого воспользуемся функцией НОРМРАСП, которая вычисляет значение $f(x)$ по формуле (9). Как Вы видите из описания синтаксиса этой функции, в качестве первых трёх её аргументов должны быть введены значения x , \bar{x}_0 , s , а на месте четвертого аргумента необходимо набрать Ложь (мы ищем не интегральную функцию распределения, а функцию плотности распределения).

Итак, введём в ячейку D8 формулу

=НОРМРАСП(D4;D6;D7;ЛОЖЬ)*2*200

(последние два множителя есть шаг h и объём выборки n (см. формулу (8)). После нажатия на клавишу [Enter] в D8 появится значение частоты: 9,11.

Казалось бы, теперь нам придётся повторить последнее действие для всех ячеек пятой строки, каждый раз изменяя адреса, что, конечно заняло бы немало времени. На самом деле всё гораздо проще: достаточно использовать маркер заполнения всех ячеек строки от «источника» D8. Программа «поняла», что мы хотим посчитать каждый элемент строки по одной и той же формуле, но с разными параметрами, расположенными в соответствующих столбцах (реально в нашем случае отличаются только значения x).

Недостатком описанной процедуры является необходимость создания дополнительных строк (в нашем примере - третьей и четвертой), заполненных одинаковыми значениями (формулами). Другой путь заключается в использовании формулы массива. Рассмотрим и его. Выделим диапазон D9:L9. Введём в ячейку D9 формулу массива:

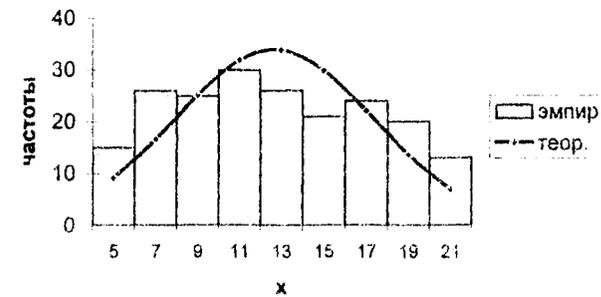
=НОРМРАСП(D4:L4;D6;D7;ЛОЖЬ)*2*200

Далее поступаем точно так, как мы делали ранее (Занятие 2) при обращении к функции ЧАСТОТА, когда результатом являлся массив (через клавишу [F2] и комбинацию клавиш [Ctrl]+[Shift]+[Enter]). В результате строка D9:L9 заполнится теми же числами, что и ранее строка D8:L8.

Вы можете пользоваться одним из двух описанных способов, который кажется Вам более простым.

Итак, теоретические частоты найдены. Постройте диаграмму, показывающую распределение эмпирических и теоретических частот. Среди нестандартных диаграмм выберите График / гистограмма и отформатируйте. Результат должен быть примерно таким, как на Рис.5.

Рис. 5. Эмпирические и теоретические частоты



Вид полученной диаграммы внушает серьезные сомнения в том, что случайная величина X в генеральной совокупности имеет нормальное распределение. Строгий количественный подход к этому вопросу заключается в рассмотрении гипотезы о типе распределения. Критерием принятия гипотезы является выполнение неравенства

$$\chi_{n \text{ аб. } \bar{x}}^2 \leq \chi_{k \text{ p.}}^2(\alpha, k), \quad (10)$$

где

$$\chi_{n \text{ аб. } \bar{x}}^2 = \sum \frac{(n_i - n_i')^2}{n_i'}, \quad (11)$$

а $\chi_{k \text{ p.}}^2(\alpha, k)$ - критическая точка распределения Хи-квадрат при заданном уровне значимости гипотезы α и числе степеней свободы k , которое в случае нормального распределения на 3 единицы меньше, чем число групп выборки (в нашей задаче $k = 9 - 3 = 6$).

Для вычисления слагаемых суммы (11) можно воспользоваться одним из предложенных выше двух способов заполнения очередной строки формируемой таблицы, например, ввести в неё формулу массива

=(D5:L5-D8:L8)*(D5:L5-D8:L8)/D8:L8

Найдя сумму ячеек этой строки с помощью функции СУММ, получаем: $\chi_{n \text{ аб. } \bar{x}}^2 = 22,14$. Задав уровень значимости гипотезы $\alpha = 0,05$, в таблице находим значение $\chi_{k \text{ p.}}^2(0,05,6) = 12,6$. Поскольку неравенство (10) не выпол-

няется, гипотезу о нормальном распределении случайной величины X следует отвергнуть. Различие эмпирических и теоретических частот значимо.

Список литературы

1. К.А.Браунли. Статистическая теория и методология в науке и технике. - М.: Наука, 1977.
2. Гмурман В.Е. Теория вероятностей и математическая статистика. - М.: Высшая школа, 1977.
3. Колесников А. Excel 97 для Windows 95. - Киев: ВНУ, 1998.
4. Статистические методы повышения качества. Под ред. Х.Куме. - М.: Финансы и статистика, 1990.
5. Власов В.Л., Гефан Г.Д. Семь лекций по математической статистике. - Иркутск: Издательство ИрГТУ, 1999.

ОГЛАВЛЕНИЕ

Работа 1	Знакомство с программой Microsoft Excel в процессе заполнения электронной таблицы.	3
Работа 2	Нахождение частотного распределения выборки. Использование функций рабочего листа.	5
Работа 3	Построение гистограмм.	8
Работа 4	Точечные оценки числовых характеристик генеральной совокупности. Интервальная оценка генеральной средней.	11
Работа 5	Корреляционный и регрессионный анализ статистических данных.	14
Работа 6	Проверка статистической гипотезы о нормальном распределении генеральной совокупности.	16
	Список литературы	20

Министерство образования РФ

ИРКУТСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

ВЫСШАЯ МАТЕМАТИКА

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА НА ПЕРСОНАЛЬНОМ КОМПЬЮТЕРЕ (НА ОСНОВЕ ПРОГРАММЫ MICROSOFT EXCEL)

Практикум для студентов факультета
технологии и компьютеризации машиностроения

Иркутск 2000